


RESEARCH ARTICLE | NOVEMBER 28 2018

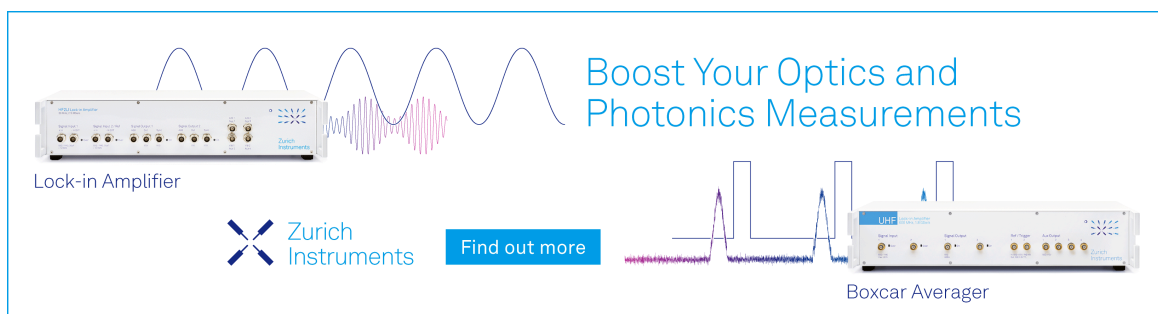
## Neural network diabaticization: A new *ansatz* for accurate high-dimensional coupled potential energy surfaces

David M. G. Williams; Wolfgang Eisfeld 




*J. Chem. Phys.* 149, 204106 (2018)

<https://doi.org/10.1063/1.5053664>



Boost Your Optics and Photonics Measurements

Lock-in Amplifier

 Zurich Instruments

[Find out more](#)

Boxcar Averager

# Neural network diabaticization: A new *ansatz* for accurate high-dimensional coupled potential energy surfaces

David M. G. Williams and Wolfgang Eisfeld<sup>a)</sup>

*Theoretische Chemie, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany*

(Received 24 August 2018; accepted 23 October 2018; published online 28 November 2018)

A new diabaticization method based on artificial neural networks (ANNs) is presented, which is capable of reproducing high-quality *ab initio* data with excellent accuracy for use in quantum dynamics studies. The diabatic potential matrix is expanded in terms of a set of basic coupling matrices and the expansion coefficients are made geometry-dependent by the output neurons of the ANN. The ANN is trained with respect to *ab initio* data using a modified Marquardt-Levenberg back-propagation algorithm. Due to its setup, this approach combines the stability and straightforwardness of a standard low-order vibronic coupling model with the accuracy by the ANN, making it particularly advantageous for problems with a complicated electronic structure. This approach combines the stability and straightforwardness of a standard low-order vibronic coupling model with the accuracy by the ANN, making it particularly advantageous for problems with a complicated electronic structure. This novel ANN diabaticization approach has been applied to the low-lying electronic states of NO<sub>3</sub> as a prototypical and notoriously difficult Jahn-Teller system in which the accurate description of the very strong non-adiabatic coupling is of paramount importance. Thorough tests show that an ANN with a single hidden layer is sufficient to achieve excellent results and the use of a “deeper” layering shows no clear benefit. The newly developed diabatic ANN potential energy surface (PES) model accurately reproduces a set of more than 90 000 Multi-configuration Reference Singles and Doubles Configuration Interaction (MR-SDCI) energies for the five lowest PES sheets. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5053664>

## I. INTRODUCTION

The fundamental understanding of chemical processes and spectroscopy is one of the main goals in chemical physics and theoretical chemistry. A key component in the strive for such understanding is the study of the underlying molecular quantum dynamics. With the ongoing advances of theoretical and experimental treatments of chemical systems alike, fundamental insights regarding chemical dynamics become increasingly available. However, the complexity of experimental data and the intricacy of theoretical treatments vastly increase with the size of the system. Therefore, extending the fundamental insights to larger systems is of great interest and an active field of current research.

One of the core issues in the theoretical treatment of systems beyond triatomics is the development of accurate high-dimensional potential energy surfaces (PESs), which provide the basis for quantum-dynamical investigations. In the case of a single adiabatic electronic state, several strategies for developing higher-dimensional PESs of sufficient accuracy have been established. One class of approaches is based on local interpolation techniques,<sup>1–5</sup> and another one uses invariant polynomials and least-squares fitting to utilize the permutation symmetry of indistinguishable nuclei.<sup>6</sup> However, extending these methods to PESs of multiple

excited states proves difficult due to the added complexity of accounting for state-state interactions. Recent attempts to tackle this issue have been formulated by utilizing the modified Shepard interpolation<sup>7–9</sup> and by using invariant polynomials and complete nuclear permutation-inversion (CNPI) symmetry.<sup>10,11</sup>

The proper inclusion of excited states requires a number of PESs and their couplings to be represented in an appropriate form. In general, there will be at least some region in nuclear configuration space (NCS) with a high density of electronic states, such that interactions among electronic states become significant, rendering the Born-Oppenheimer (BO) approximation invalid. For such a case, a *quasi-diabatic* representation of the coupled electronic states has been found to be of great advantage.<sup>12–25</sup> Quasi-diabatic, which will be referred to as “diabatic” from here on for the sake of simplicity, means that the state basis to represent the electronic Hamiltonian is required to preserve the character of the electronic states as much as possible, reducing the nonadiabatic (or derivative) coupling to a degree that it can be neglected safely. Thus, the electronic Hamiltonian will be represented by a non-diagonal PES matrix encoding all state energies and couplings. One advantage is that the matrix elements of the electronic Hamiltonian become simple, well-behaved functions of the nuclear coordinates, reducing the complexity of finding analytic expressions for them. The simplest (and most commonly used) such approach utilizing the diabatic representation is the multi-mode linear vibronic coupling method by

<sup>a)</sup>wolfgang.eisfeld@uni-bielefeld.de

Köppel, Domcke, and Cederbaum<sup>26</sup> as well as similar methods developed by Yarkony and co-workers.<sup>27–30</sup> These treatments, while suitable for the simulation of certain ultra-fast nonadiabatic processes, are too limited to represent more extended regions of the coupled PESs as required for the treatment of more complicated dynamical processes. Therefore, extensions of vibronic coupling models needed for the treatment of dynamical processes of higher complexity have been developed.<sup>31–36</sup> Furthermore, several other approaches exist utilizing properties of the adiabatic electronic wave functions for diabaticization.<sup>18,21,22,24,37–42</sup> Most of these methods do not require a model diabatic matrix, meaning that they do not yield a set of PESs in closed mathematical form. Therefore, a second step would be required to represent the diabatic matrix elements provided by a point-wise diabaticization scheme by some external mathematical model. Representing the diabatic energies and couplings accurately and in closed form as diabatic PESs beyond simple models is still a major problem for systems with more than three atoms.<sup>43–46</sup> This has been attempted by the aforementioned methods of extending vibronic coupling models in various ways,<sup>31–36</sup> using modified Shepard interpolation<sup>7–9</sup> or invariant polynomials and CNPI symmetry<sup>10</sup> as well as by choosing elaborate functional forms for the diabatic matrix elements.<sup>43–45</sup> However, the number of accurate diabatic surfaces for such systems in the literature remains rather scarce.

Recently an alternative approach for the representation of PESs evolved utilizing artificial neural networks (ANNs),<sup>47–65</sup> which are trained by various methods usually based on analytic reference data from other models or *ab initio* data. It has been shown mathematically that ANNs are capable of uniformly representing any continuous real function of  $n$  dimensions up to arbitrary accuracy.<sup>66</sup> Therefore, in principle, it should be possible to train an appropriate neural network to represent any PES accurately with respect to the underlying data given. Apart from the high accuracy that can be achieved, ANNs have the advantage that they can be evaluated very efficiently once trained, which plays a key role in quantum dynamics methods such as multi-configuration time-dependent Hartree (MCTDH) where the evaluation of the PES is the most time demanding part.<sup>67,68</sup> While ANNs are in principle capable of reproducing arbitrary continuous functions, this is in practice of course limited by the acquisition of data, training algorithm, chosen network architecture, and an issue commonly referred to as “over-fitting.” However, ANNs have already been used with impressive results to represent a single PES based on high-level *ab initio*-data.<sup>63,65</sup> Some first attempts to extend the use of ANNs to diabatic PESs are also known.<sup>69,70</sup>

In the present work, a novel ANN based diabaticization approach is used to accurately represent the coupled PESs of electronically excited states with strong Jahn-Teller (JT) couplings. For this purpose, an ANN-based diabatic model has been developed and trained to represent the low-lying electronic states of NO<sub>3</sub>. The new method is described here for the first time and the overall quality, the resulting PES model, and the stability of the fitting procedure is analyzed depending on various factors.

## II. THEORY

### A. Adiabatic and diabatic representation

As pointed out above, we aim for an accurate diabatic representation of the electronic Hamiltonian of a given molecular system. Therefore, the theoretical background of adiabatic and diabatic representations is briefly summarized in the following. The starting point is the total molecular Hamiltonian with electronic degrees of freedom  $\mathbf{q}$  and nuclear degrees of freedom  $\mathbf{Q}$  reading

$$\hat{H}(\mathbf{q}, \mathbf{Q}) = \hat{T}_{nuc}(\mathbf{Q}) + \underbrace{\hat{T}_{el}(\mathbf{q}) + V_C(\mathbf{q}, \mathbf{Q})}_{\hat{H}_{el}(\mathbf{q}, \mathbf{Q})}. \quad (1)$$

If the nuclear motions are frozen, the nuclear kinetic energy  $\hat{T}_{nuc}(\mathbf{Q})$  vanishes and what remains is the electronic Hamiltonian  $\hat{H}_{el}$  consisting of the electronic kinetic energy  $\hat{T}_{el}$  and the Coulomb potential  $V_C$ . The eigenvalues  $E_k^a(\mathbf{Q})$  and eigenfunctions  $\psi_k^a(\mathbf{q}; \mathbf{Q})$  of  $\hat{H}_{el}$  can be evaluated by suitable *ab initio* methods for selected points  $\mathbf{Q}$  in the nuclear configuration space. The  $E_k^a(\mathbf{Q})$  depend parametrically on the nuclear coordinates  $\mathbf{Q}$  which yields the adiabatic potential energy surfaces (PESs) for the electronic states in question. The set of all electronic eigenfunctions  $\psi_k^a$  forms the *adiabatic basis* for the complete molecular wave functions. The total wave function thus can be expanded as

$$\Psi_j^a(\mathbf{q}, \mathbf{Q}) = \sum_k \phi_k^a(\mathbf{Q}) \cdot \psi_k^a(\mathbf{q}; \mathbf{Q}). \quad (2)$$

Expanding  $\hat{H}$  in terms of  $\psi_k^a(\mathbf{q}; \mathbf{Q})$  yields the *adiabatic representation*  $\hat{H}^a(\mathbf{Q})$  with all electronic degrees of freedom integrated out

$$\hat{H}_{kj}^a = (\hat{T}_{nuc}(\mathbf{Q}) + E_k^a(\mathbf{Q})) \cdot \delta_{kj} - \hat{\Lambda}_{kj}(\mathbf{Q}). \quad (3)$$

In case that the derivative coupling terms  $\hat{\Lambda}(\mathbf{Q})$  are small, they can be ignored which is widely known as Born-Oppenheimer (BO) approximation. However, this condition is often not fulfilled for electronically excited states and  $\hat{\Lambda}(\mathbf{Q})$  may even become singular or very large for conical intersections or avoided crossings, respectively.<sup>12</sup>

By contrast, a *diabatic basis*  $\{\psi_k^d(\mathbf{q}; \mathbf{Q})\}$  is chosen such that  $\hat{\Lambda}$  remains negligibly small by construction so that in the resulting Hamiltonian

$$\hat{H}_{kj}^d = \hat{T}_{nuc} \cdot \delta_{kj} + W_{kj}^d(\mathbf{Q}) \quad (4)$$

the derivative coupling terms can be ignored just like in the BO approximation. The kinetic (derivative) coupling of the adiabatic representation is transformed into a potential coupling in the diabatic representation since the diabatic potential matrix  $W^d(\mathbf{Q})$  is not diagonal anymore. Unfortunately, a diabatic basis cannot be defined unambiguously and can only be determined *ab initio* through the computation of adiabatic eigenstates and energies. However, it can be shown that the diabatic matrix elements  $W_{kj}^d(\mathbf{Q})$  must be simple and smooth functions of the nuclear coordinates. Therefore, the approximate mathematical representation of  $W^d(\mathbf{Q})$  is usually much easier than that of  $E^a(\mathbf{Q})$ . Finally, the adiabatic basis functions

can be expanded in terms of diabatic basis functions as

$$\psi_l^a(\mathbf{q}; \mathbf{Q}) = \sum_k u_{kl}(\mathbf{Q}) \cdot \psi_k^d(\mathbf{q}; \mathbf{Q}). \quad (5)$$

This relates to a basis transformation  $\mathbf{U}(\mathbf{Q})$  which would be exact if both bases would be complete or span the same vector space. In this case, the unitary matrix  $\mathbf{U}$  diagonalizes  $\mathbf{W}^d$

$$\mathbf{U}^\dagger \mathbf{W}^d \mathbf{U} = \mathbf{W}^a = \text{diag}(W_j^a) \quad (6)$$

and the eigenvalues exactly reproduce the adiabatic energies  $E_k^a(\mathbf{Q})$ .

Up to this point no approximations were necessary as all bases were assumed to be complete. However, for any practical application, all bases involved are finite and only a small number of states can be handled. Thus, the adiabatic states are approximately represented in a finite  $N^d$ -dimensional diabatic basis by

$$\psi_l^a(\mathbf{q}; \mathbf{Q}) \approx \sum_k^{N^d} u_{kl}^d(\mathbf{Q}) \cdot \psi_k^d(\mathbf{q}; \mathbf{Q}). \quad (7)$$

$N^d$  can be equal or greater than the number of adiabatic states required.

## B. Artificial neural networks

In the present work, multilayer perceptron feed-forward neural networks are utilized for the diabaticization of adiabatic molecular energy data. A feed-forward neural network is a function taking a vector  $\boldsymbol{\eta}^{(1)}$  as the input layer and processing it via intermediate results  $\boldsymbol{\eta}^{(k)}$ , the so-called hidden layers, to a final output vector  $\boldsymbol{\eta}^{(f)}$  called the output layer. The vector elements  $\eta_j^{(k)}$  of the  $k^{\text{th}}$  layer are the neurons (perceptrons). Each intermediate  $\boldsymbol{\eta}^{(k)}$  depends solely on the previous layer  $\boldsymbol{\eta}^{(k-1)}$  by

$$\eta_j^{(k)} = f^{(k)} \left( \underbrace{\beta_j^{(k)} + \sum_l \omega_{jl}^{(k)} \eta_l^{(k-1)}}_{\chi_j^{(k)}} \right). \quad (8)$$

Here,  $f^{(k)}$  is a function of one variable  $\chi_j^{(k)}$  called the activation function,  $\chi_j^{(k)}$  being a weighted sum of the values of the neurons  $\eta_l^{(k-1)}$  of the previous layer with an added bias term  $\beta_j^{(k)}$ . The resulting ANN can be visualized as shown in Fig. 1.

## C. Diabatic model (ansatz)

Since the diabatic matrix elements are by nature slowly varying functions of the nuclear coordinates, it is straightforward to expand them as multi-dimensional polynomials as

$$W_{kj}^d(\mathbf{Q}) = \sum_\alpha p_\alpha^{kj} \prod_l Q_l^{n_{\alpha l}}. \quad (9)$$

The order of the polynomials is given by the sum of the exponents  $n = \sum_l n_{\alpha l}$ . Special boundary conditions like asymptotic behaviour or periodicity may be incorporated into the definition of the nuclear coordinates  $Q_l$ . The symmetry of the system is conveniently accounted for by using symmetry-adapted

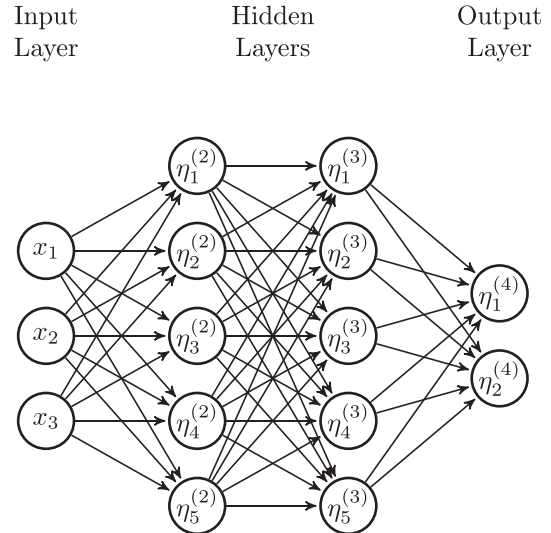


FIG. 1. A neural network with 2 hidden layers. Arrows correspond to weighted contributions. Bias terms and activation functions are not visualized.

coordinates in which case certain expansion parameters  $p_\alpha^{kj}$  may vanish or be strictly related to other expansion parameters. In the latter case, it is very convenient to rewrite this expansion in terms of basis matrices as

$$\mathbf{W}^d(\mathbf{Q}) = \sum_L \lambda_L \cdot \mathbf{M}_L(\mathbf{Q}). \quad (10)$$

The symmetry relations are encoded in the basis matrices  $\mathbf{M}_L(\mathbf{Q})$  here. This approach is extremely successful and straightforward in the case of *linear vibronic coupling*<sup>26</sup> or for other low-order expansions. The order of the expansion terms needs to be increased if higher accuracy of the PES model is required over more extended regions of the nuclear configuration space (NCS). However, the higher-order terms cause several technical problems for the required nonlinear fitting as well as possibly causing artifacts in the shape of the PESs. Hence, the choice of terms included in the model usually requires a trade-off between lower accuracy with higher stability and higher (local) accuracy at the cost of more erratic behavior outside the sampling space. In the present work, we therefore propose a new approach based on an expansion in the basis matrices of lowest orders. The expansion coefficients are first obtained from a standard nonlinear least-squares fit and then are modified by *corrections* provided by ANNs. As will be shown below, this new method overcomes the characteristic oscillation features plaguing typical higher-order polynomial expansions. Furthermore, this approach does not produce the erratic behavior outside of the sampling space, often observed in higher-order polynomial expansions. The new ANN approach, thus, leads to very robust diabatic ANN-PES models.

The fundamental idea of this method is to render the expansion coefficients of Eq. (10) coordinate dependent reading

$$\mathbf{W}^d(\mathbf{Q}) = \sum_L \lambda_L(\mathbf{Q}) \cdot \mathbf{M}_L(\mathbf{Q}), \quad (11)$$

where each coefficient is of the form

$$\lambda_L(\mathbf{Q}) = \begin{cases} \lambda_L^0 & \text{for uncorrected terms} \\ \lambda_L^0 \cdot (1 + \eta_{k_L}^{(f)}(\mathbf{Q})) & \text{otherwise.} \end{cases} \quad (12)$$

The constants  $\lambda_L^0$  obtained from standard nonlinear fitting procedures are modified by the coordinate dependent outputs of a trained ANN. This way, one can combine the stability and basic qualitative description of low-order expansions while still introducing flexibility to achieve excellent accuracy. Of course, the ANN must be trained properly with respect to *ab initio* reference data and this training also involves a necessary adiabatic-to-diabatic transformation. This requires a modification of standard ANN training algorithms described next.

#### D. Backpropagation of generalized outputs

All special properties of ANNs aside, they can be viewed as parametrized functions modeled for the specific purpose of closely reproducing an expected output  $\mathbf{t}$  for each point in the function's domain. In the most simple case, the function output  $\boldsymbol{\eta}^{(f)}$  is modeled to *directly* reproduce the desired output  $\mathbf{t}$ , in which case the error can be measured in terms of the difference  $\mathbf{e}$  between the two. The problem of finding an optimal set of parameters can then be expressed in terms of minimizing the sum of squares error

$$V = \frac{1}{2} |\mathbf{e}^2|. \quad (13)$$

Applying a (gradient descent) step in parameter space to improve the current parameter set requires derivatives of  $V$  with respect to all parameters  $\omega_{jl}^{(k)}$  and  $\beta_j^{(k)}$ . However, thanks to the mathematical structure of feed-forward networks, the derivatives can be easily expressed in terms of derivatives with respect to the weighted sums  $\chi_j^{(k)}$  by defining

$$\delta_j^{(k)} := \frac{\partial V}{\partial \chi_j^{(k)}} \quad (14)$$

which yields

$$\frac{\partial V}{\partial \omega_{jl}^{(k)}} = \delta_j^{(k)} \cdot \eta_l^{(k-1)}, \quad (15a)$$

$$\frac{\partial V}{\partial \beta_j^{(k)}} = \delta_j^{(k)}. \quad (15b)$$

These equations generalize naturally for sums over multiple data points by introducing a sum over all points  $\mathbf{Q}_p$  which is omitted for brevity. These  $\delta_j^{(k)}$  are in return analytically known from the recurrence relation

$$\delta_j^{(k)} = \underbrace{f_j^{(k)}(\chi_j^{(k)})}_{f_j^{(k)}} \cdot \sum_k \omega_{kj}^{(k+1)} \delta_k^{(k+1)}, \quad (16)$$

which terminates at the final layer, yielding

$$\delta_j^{(f)} = -e_j \cdot f_j^{(f)}. \quad (17)$$

Evaluating and applying gradient descent steps by exploiting the recursive nature of the required derivatives seen in Eq. (16) is referred to as *backpropagation*. This scheme is generally applicable as long as the error can be *directly* expressed in

terms of the ANN output. However, in principle, the neural network output  $\boldsymbol{\eta}^{(f)}$  can relate to the desired output  $\mathbf{t}$  in arbitrarily *indirect* ways. For example, the present case uses the individual output values as coefficients for a diabatic expansion, which in turn produces a matrix that is diagonalized to obtain adiabatic energies comparable to the actual *ab initio* data. This processed form of  $\boldsymbol{\eta}^{(f)}$ , denoted  $\mathcal{D}(\boldsymbol{\eta}^{(f)})$ , is now what is actually comparable to  $\mathbf{t}$ . Hence we developed a modification allowing for arbitrary differentiable relations  $\mathcal{D}(\boldsymbol{\eta}^{(f)})$  between the output layer and physical reference data. While the actual backpropagation remains the same, the final layer now terminates to

$$\begin{aligned} \delta_j^{(f)} &= \frac{\partial V}{\partial \chi_j^{(f)}} = - \sum_l e_l \cdot \frac{\partial \mathcal{D}_l(\boldsymbol{\eta}^{(f)})}{\partial \chi_j^{(f)}} \\ &= - \sum_l e_l \cdot \frac{\partial \mathcal{D}_l(\boldsymbol{\eta}^{(f)})}{\partial \eta_j^{(f)}} \frac{\partial \eta_j^{(f)}}{\partial \chi_j^{(f)}} \\ &= - \sum_l e_l \cdot \frac{\partial \mathcal{D}_l}{\partial \eta_j^{(f)}} \cdot f_j^{(f)}. \end{aligned} \quad (18)$$

In the present case, where  $\mathcal{D}(\boldsymbol{\eta}^{(f)})$  refers to the eigenvalues of a diabatic matrix which is parametrized by the output layer  $\boldsymbol{\eta}^{(f)}$  as seen in Eqs. (11) and (6), partial derivatives are obtained from numerical differentiation. This is always possible as long as the target  $\mathbf{t}$  refers to a continuously differentiable property. Since the diagonalization is carried out numerically, the partial derivatives must be computed numerically as well.

#### E. (Adapted) Marquardt-Levenberg method

The most basic training algorithm using the backpropagation would be a simple gradient decent approach. However, due to the extreme non-linearity of the fitting problem, it is desirable to choose a more sophisticated algorithm to optimize a given parameter set  $\boldsymbol{\Omega}$  (containing all weights and biases) than a gradient descent method can provide. A widely used improvement is to utilize a Marquardt-Levenberg algorithm<sup>71</sup> which approximates the exact Hessian using the Jacobian

$$J_{lk} = \frac{\partial e_l}{\partial \Omega_k} \quad (19)$$

and a damping term  $\lambda$ . The working equation to produce the correction step  $\Delta \boldsymbol{\Omega}$  in parameter space then reads

$$\Delta \boldsymbol{\Omega} = (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{1})^{-1} \mathbf{J}^T \mathbf{e}. \quad (20)$$

It can be shown, in analogy to the derivatives of  $V$  with respect to individual parameters, that  $\mathbf{J}$  can be evaluated analytically. By using backpropagation with a modified termination condition,<sup>71</sup> a single backpropagation yields all  $J_{lk}$  for a fixed row index  $l$ , meaning one requires as many backpropagations to construct  $\mathbf{J}$  as there are physical outputs per data point (e.g., adiabatic energies). In the standard case, each output neuron corresponds to one physical output and the final backpropagation layer  $\Delta^{(f,l)}$  reads

$$\Delta_j^{(f,l)} = -\delta_{lj} \cdot f_j^{(f)} \quad (21)$$

for a given  $l$ . We again derived a modified version for the case that there is no one to one correspondence of neural network output and physical output but an arbitrary differentiable relation  $\mathcal{D}(\boldsymbol{\eta}^{(f)})$  instead. Equation (21) then generalizes to

$$\Delta_j^{(f,l)} = -\frac{\partial \mathcal{D}_l}{\partial \eta_j^{(f)}} \cdot f_j^{(f)}. \quad (22)$$

### III. APPLICATIONS

The new method should be widely applicable and can be used for conventional diabaticization-by-*ansatz* problems as well as for the more advanced approaches like the hybrid diabaticization<sup>72</sup> recently developed by us. A prototypical Jahn-Teller problem is chosen to demonstrate the novel ANN diabaticization. The first three electronic states of  $\text{NO}_3$ ,  $\tilde{X}^2A_2'$ ,  $\tilde{A}^2E''$ , and  $\tilde{B}^2E'$ , have been studied by us extensively<sup>35,36,73–75</sup> and thus  $\text{NO}_3$  is an ideal benchmark system for the present study. The results of the present work can be compared directly with our latest diabatic PES model.<sup>75</sup>

#### A. Diabatic model for $\text{NO}_3$

For this first proof-of-principle study of the new ANN diabaticization method, a restricted-dimensional (5D) model of the lowest PESs corresponding to the  $\tilde{X}$ ,  $\tilde{A}$ , and  $\tilde{B}$  states of the planar  $\text{NO}_3$  radical has been developed. Due to the global  $D_{3h}$  symmetry of  $\text{NO}_3$ , two pairs of PESs become degenerate at symmetry points and are split for distorted geometries due to strong JT coupling. Thus, the symmetry and couplings have to be accounted for properly. Furthermore, the  $\tilde{A}$  state has a rather low dissociation threshold and thus the asymptotic behaviour of the PESs is relevant. To account for the basic asymptotic behavior and a correct description of the complete nuclear transformation symmetry in the underlying low-order model, a set of symmetry-adapted coordinates has been chosen as already described in previous work.<sup>33,36</sup> First, a set of primitive coordinates is defined comprising the three N–O distances  $r_i$  and a set of O–N–O angles  $\alpha_i$ . The latter are numbered according to the unique atom not involved in forming the angle. These primitive coordinates are first transformed nonlinearly as

$$m_i = 1 - \exp(-\gamma(r_i - r_0)), \quad (23a)$$

$$\alpha'_i = \frac{\alpha_i - \alpha_0}{r_j r_k}, \quad i \neq j \neq k \quad (23b)$$

to account for the asymptotic behaviour. Here,  $r_0$  and  $\alpha_0$  are the respective distances and angles at the reference point and  $\gamma$  is a chosen Morse-parameter. Then the transformed coordinates are symmetry-adapted to yield a breathing mode  $a$ , a degenerate pair of asymmetric stretching modes  $x_s, y_s$ , and a degenerate pair of asymmetric bending modes  $x_b, y_b$  by

$$a = \sqrt{\frac{1}{3}}(m_1 + m_2 + m_3), \quad (24a)$$

$$x_s = \sqrt{\frac{1}{6}}(2m_1 - m_2 - m_3), \quad (24b)$$

$$y_s = \sqrt{\frac{1}{2}}(m_2 - m_3), \quad (24c)$$

$$x_b = \sqrt{\frac{1}{6}}(2\alpha'_1 - \alpha'_2 - \alpha'_3), \quad (24d)$$

$$y_b = \sqrt{\frac{1}{2}}(\alpha'_2 - \alpha'_3). \quad (24e)$$

With these coordinates, all linear coupling matrices can be constructed with two non-trivial kinds of coupling matrix blocks

$$\boldsymbol{\varepsilon}_{s,b} = \begin{pmatrix} x_{s,b} & y_{s,b} \\ y_{s,b} & -x_{s,b} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\rho}_{s,b} = \begin{pmatrix} x_{s,b} \\ -y_{s,b} \end{pmatrix} \quad (25)$$

accounting for  ${}^2E'$  and  ${}^2E''$  Jahn-Teller coupling as well as *pseudo*-Jahn-Teller coupling between  ${}^2A_2'$  and  ${}^2E'$ . The diagonal potential  $\mathbf{W}^d(\mathbf{Q})$  consists of 12 coordinate-dependent terms up to second order listed in the [Appendix](#). The three constants defining the energy differences between the electronic states at the  $D_{3h}$  reference geometry are modified by the ANN as well as the expansion coefficients of the coupling matrices. This yields a total of 9 ANN-corrected basis matrices. In the following, the diabatic electronic basis states are used in order  ${}^2A_2'$ ,  ${}^2E''$ , and  ${}^2E'$ . This yields the diabatic matrix

$$\mathbf{W}^d(\mathbf{Q}) = \mathbf{W}_{\text{diag}}^d(\mathbf{Q}) + \begin{pmatrix} \lambda_1 & 0^T & 0^T \\ 0 & \lambda_2 \mathbf{1} & 0 \\ 0 & 0 & \lambda_3 \mathbf{1} \end{pmatrix} + \begin{pmatrix} 0 & 0^T & 0^T \\ 0 & \lambda_4 \boldsymbol{\varepsilon}_s + \lambda_5 \boldsymbol{\varepsilon}_b & 0 \\ 0 & 0 & \lambda_6 \boldsymbol{\varepsilon}_s + \lambda_7 \boldsymbol{\varepsilon}_b \end{pmatrix} + \begin{pmatrix} 0 & 0^T & \lambda_8 \boldsymbol{\rho}_s^T + \lambda_9 \boldsymbol{\rho}_b^T \\ 0 & 0 & 0 \\ \lambda_8 \boldsymbol{\rho}_s + \lambda_9 \boldsymbol{\rho}_b & 0 & 0 \end{pmatrix}, \quad (26)$$

where  $\mathbf{W}_{\text{diag}}^d$  contains all uncorrected diagonal terms. The first, second, and third matrix terms encode the ANN-corrected contributions to the diagonal potential terms, the Jahn-Teller coupling blocks, and the *pseudo*-Jahn-Teller coupling blocks, respectively. The corrected coefficients  $\lambda_i$  are of the general form

$$\lambda_j = \lambda_j^0 \cdot (1 + c_j \cdot \boldsymbol{\eta}_j^{(f)}(\mathbf{Q})). \quad (27)$$

Factors of  $c_j$  allow further flexibility if particular  $\lambda_i^0$  (*read*: the reference model terms) reside in a different order of magnitude than the other terms (the coupling terms) and hence require a different treatment.

#### B. *Ab initio* data and reference fit

The extremely intricate electronic structure of the  $\text{NO}_3$  radical requires a rigorous treatment in order to obtain stable results. The *ab initio* data points are taken from previous work<sup>75</sup> and were computed by Multi-configuration Reference Singles and Doubles Configuration Interaction (MR-SDCI) calculations based on Complete Active Space Self-Consistent Field (CASSCF) reference wave functions using a slightly adapted correlation consistent aug-cc-pVTZ standard basis.<sup>74</sup> For details of the *ab initio* computations,

see Refs. 36 and 73–75. The data points have been selected by a stochastic approach described in Ref. 76. To this end, normalised random vectors are generated which define linear cuts through the multi-dimensional PESs starting from the ground state equilibrium structure. Then, *ab initio* points are computed along these random directions. By this approach, we can check and ensure that the CASSCF/MRCI calculations converge to the correct result. This could not be achieved by computing single points at random positions due to the very complicated electronic structure of this system (as typical for radicals and excited states in general). The sampled nuclear configuration space is selected such that the resulting PES model is optimally suited for the simulation of spectroscopic properties. The 21 free parameters  $\lambda_j^0$  of the reference model were fitted against this data set using a Marquardt-Levenberg algorithm incorporated into a genetic algorithm with a resulting root mean square (rms) error of roughly  $1700\text{ cm}^{-1}$ .

The large data set of roughly 90 000 *ab initio* reference energies has been partitioned into training sets of 85% of the data and external validation sets of the remaining 15% of the energies. The validation set error is used as a convergence parameter and does not otherwise contribute to the neural network fit, but has been included in the polynomial reference model. This technique is referred to as *early stopping* in the literature.<sup>77</sup> In the present work, this early stopping technique is implemented in a “relaxed” fashion. Instead of immediately stopping the fit when the reference error does not improve along with the fitting error, the reference error must increase for a fixed number of consecutive cycles (default is 3) before early stopping is executed. The convergence of the fitting error does not suffer from this restriction, as discussed in Sec. IV below. The contributions to the squared error which is minimized during the training of the neural network are weighted with an exponential decay of energy differences of the form

$$\sigma_{ij} = \exp\left(-\alpha(E_j^a(\mathbf{Q}_i) - E_j^a(\mathbf{Q}_0))\right). \quad (28)$$

This scheme allows us to weigh regions in NCS more or less depending on their relevance in the nuclear dynamics and thus helps to improve the accuracy of the fit in the most relevant regions.

### C. Description of ANN setup

Apart from the actual fit, several parameters and options regarding the ANN have been taken into consideration to ensure an optimal setup. The final setup discussed in Sec. IV involves ANNs with only one hidden layer of varying sizes. The sigmoid  $\tanh(\chi)$  has been chosen as the hidden layer activation function. For the final layer, the identity  $\text{Id}(\chi)$  has been selected. Each ANN is trained starting from a set of 100 randomly generated initial guesses for weights and biases which are then optimized according to the working equations from Sec. II E. While modified backpropagation by itself has been tested, it has been found inferior for the current application compared to the modified Marquardt-Levenberg method. Similarly, “deeper” networks with more than one hidden layer have been tested with various size combinations, but no

improvement compared to single layer ANNs was observed. It has been shown by Cybenko that any continuous real function can be represented with arbitrary accuracy by a single layer ANN.<sup>66</sup> Since this is a consequence of the superposition of sigmoidal functions, this also holds for multi-layer ANNs and has been shown by Cybenko for ANNs with two hidden layers. In principle, it could be more efficient to achieve a certain accuracy with less optimization parameters by using multiple hidden layers of smaller size. Therefore, we tested this possibility by setting up ANNs with one through four hidden layers and layer sizes such that the total number of optimization parameters is roughly the same. Thus, a benefit of “deep layering” ANNs would manifest itself in lower rms errors. From a study involving various ANN topographies listed in greater detail in Sec. IV below, we conclude that increasing the number of hidden layers shows no clear advantage for the present case (see below). Increasing the number of hidden layers can yield slightly better as well as a significantly worse results compared to a single hidden layer. Larger numbers of initial guesses (up to 2000) were found to have no influence on the overall quality of the final model. Various splitting schemes for fitting reference and validation data, respectively, have been studied. The selection of 15% of the original data as validation set has been found to be sufficiently large to reduce over-fitting while not being detrimental to the overall quality of the final model by removing too many data from the training set. Table III provides a quick overview of various choices of data partitioning, which is discussed in Sec. IV. Furthermore, the validation data has been ensured to have no characteristic differences compared to the fitting data, as random reassignment of fitting and reference subsets had no significant effect on the fit quality.

## IV. RESULTS AND DISCUSSION

The single layer ANN setup described in Sec. III C yields excellent fit results for a surprisingly low number of hidden layer neurons. The (weighted) rms error of the primitive reference model is fairly large and about the same for both fitting and validation data set ( $1730\text{ cm}^{-1}$  and  $1710\text{ cm}^{-1}$ , respectively). This is to be expected as both data sets have been included equally in the reference fit. The reference model is not flexible enough to represent the surface in a satisfying quantitative way but is sufficient to reproduce the reference data qualitatively. The remaining deviation is corrected by the ANN model yielding PESs of very high accuracy. Various neural network sizes between 20 and 120 hidden layer neurons have been tested as summarized in Table I. The convergence behavior of the best rms error (and the corresponding reference error) is presented in Fig. 2.

Considering the best ANN parametrization found in each set of 100 initial guesses, the fitting rms expresses satisfactory convergence for 60–75 neurons, the best fit from the 75 neuron set yielding a weighted rms error of  $38.4\text{ cm}^{-1}$ . While the fitting rms expectedly decreases further for ever increasing network sizes, the difference between fitting and validation error *increases* also. For this reason the 75 neuron network will be considered as the optimal result and primary

TABLE I. Fitting results of best and 10th best out of 100 trained networks for each hidden layer size, depending on the number of hidden layer neurons. Both fitting rms and rms of validation set are shown. All rms errors are weighted and are given in  $\text{cm}^{-1}$ .

Neurons	Best	Validation	10th best	Validation
20	101	103	107	111
40	68.3	72.9	69.9	74.3
50	56.0	60.5	58.5	64.5
55	51.0	56.4	54.4	59.8
60	48.1	52.0	49.5	54.1
65	44.6	49.8	46.6	51.2
70	41.9	47.7	43.2	49.0
75	38.4	42.6	41.1	45.3
80	36.7	42.4	38.6	44.0
90	34.4	41.1	35.0	41.0
100	31.0	37.3	32.0	38.7
120	26.8	34.5	27.8	34.0

example to alleviate potential concerns of over fitting. Interestingly, the closeness of the validation set and fitting set error (here  $38.4 \text{ cm}^{-1}$  and  $42.6 \text{ cm}^{-1}$ ) is consistent throughout many parametrizations in each set, as even the 10th percentile of each set of 100 ranked ANNs generally shows the same behavior as the best ANN itself. This is both a testament to the consistency of the model in areas unknown to the fit as it is to the stability of the method because a comparably small number of initial guesses already produces a lot of networks of similar quality. Knowing the basic capabilities of a 75 neuron single layer ANN, it may be interesting to consider comparable networks of “increased depth,” that is, ANNs with multiple hidden layers. To achieve a fair comparison one can increase the number of layers while keeping the number of corresponding formal parameters (weights and biases) approximately constant and keeping all other technical influences of the fit the same. Overall, the errors produced by networks of increasing depth do not show a straightforward improvement compared to the single-layer case, as seen in Table II. While the ANN with three hidden layers does produce slightly better fitting errors,

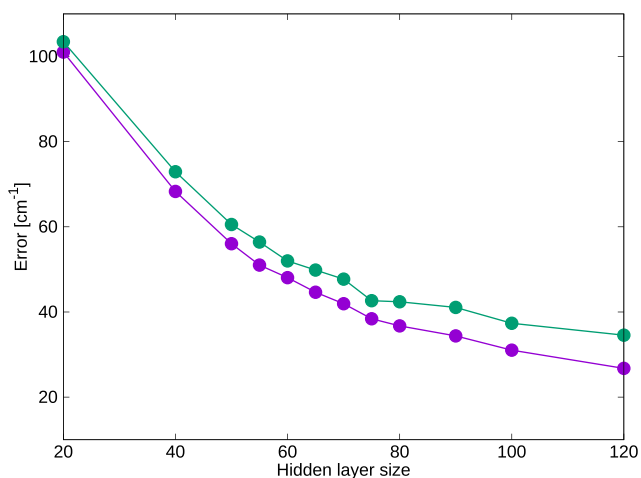


FIG. 2. Weighted root mean square error of the best neural network of a given set as a function of hidden layer size. Purple line represents fitting set error, and green line represents validation set error.

TABLE II. Fitting results of best out of 100 trained networks for each network topography. Both fitting rms and rms of validation set are shown. All rms errors are weighted and are given in  $\text{cm}^{-1}$ .

Hidden layers	Neurons/Layer	Parameters	Best	Validation
1	75	1134	38.4	42.6
2	27	1170	96.6	104
3	20	1149	26.4	30.2
4	17	1182	75.9	82.5

both two and four hidden layer networks perform significantly worse than the single layer model in this particular setup. Therefore, increasing the number of layers shows no particular advantage compared to the cost of increased complexity.

Another technical factor influencing the fit is that of *early stopping* as described above. In order to investigate potential (negative) impacts on the overall quality of the PES model, the single layer 75 neuron model was fitted again without early stopping and a maximum number of 1000 Marquardt-Levenberg iterations. Both fitting and validation error have been found to show no significant improvement ( $35.7 \text{ cm}^{-1}$  and  $40.3 \text{ cm}^{-1}$ ) at the cost of missing an additional safeguard against over-fitting. Furthermore, the training without early stopping is much less efficient in terms of computer time which might be a big disadvantage for the representation of larger systems and thus larger ANNs.

Apart from the validation set as used in this work, it is not uncommon to withhold an additional part of the data set which is not considered in the fit at all, the so-called *testing set*. Furthermore, choosing the relative sizes of these three data sets may be a point of concern. In particular, increasing the validation (and testing) set(s) may be a way to further decrease the risk of overfitting. The impact of these two factors has been studied for the present case, again using the single layer 75 neuron architecture as a benchmark. Table III shows the results of comparing the present case (85% fit: 15% validation: 0% testing) against a standard literature case (85%: 10%: 5%) as well as the extreme case of choosing a training and validation set of equal size, again with and without the 5% testing set. Due to the extreme similarity of validation and testing rms, no benefit has been observed in withholding data for a testing set. Instead, introducing a testing set at the cost of reducing the validation set has only been found to *increase* the difference between fit and validation for the present case. Similarly, an even split between validation and fitting data has not been found to have any beneficial effect compared to the cost of increased

TABLE III. Fitting results of best out of 100 trained networks (single layer, 75 neurons) for each choice of data split. The full data set is segmented into a set for fitting, a validation set for convergence testing (*early stopping*), and an independent test set. All rms errors are weighted and are given in  $\text{cm}^{-1}$ .

Split	Best	Validation	Testing
85%: 15%: 0%	38.4	42.6	...
85%: 10%: 5%	39.3	47.0	45.6
50%: 50%: 0%	37.7	53.2	...
50%: 45%: 5%	38.4	50.2	50.2



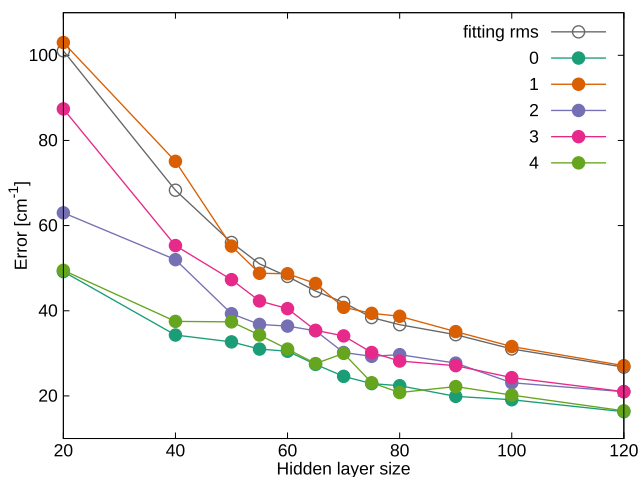


FIG. 3. State-resolved convergence behavior of unweighted root mean square error for *ab initio* data points up to 1 eV above respective reference point energy. States are enumerated from ground state to energetically highest lying state. Gray: weighted fitting set error.

disparity between the two. For this reason, our preferred distribution of the reference data is 85% fit: 15% validation: 0% testing.

Another useful measure for the accuracy of the surface is the unweighted rms error of the adiabatic energies below a certain threshold. Since we are particularly interested in the quality of the surface in the physically relevant region, the cutoff energies have been chosen 1 eV above the reference point energies, which results in about 15 000 *ab initio* values for each electronic state.

Considering again all network sizes as shown in Fig. 3 and Table IV, one finds that the unweighted errors are generally well below the weighted fitting rms for all states except the first (and to a lesser degree third) excited state. This indicates that the physically relevant regions of the surface are reproduced even significantly better than the weighted rms over all *ab initio* data points would indicate. While the convergence of the individual states initially appears more erratic than the total weighted error for small networks, it becomes well-behaved

TABLE IV. Fitting results of best out of 100 trained networks of size 20 to 120, depending on the number of hidden layer neurons. Both fitting rms and rms of validation set are shown. All rms errors are unweighted with a cutoff 1 eV above each state's reference energy. All errors are given in  $\text{cm}^{-1}$ .

Neurons	State 0	State 1	State 2	State 3	State 4
20	49.2	103	63.0	87.4	49.5
40	34.3	75.1	52.0	55.3	37.5
50	32.7	55.2	39.3	47.3	37.4
55	31.0	48.8	36.8	42.3	34.3
60	30.5	48.7	36.4	40.5	31.0
65	27.4	46.4	35.3	35.4	27.6
70	24.6	40.8	30.2	34.1	30.0
75	22.9	39.4	29.3	30.2	23.1
80	22.4	38.7	29.7	28.2	20.8
90	19.9	35.1	27.7	27.1	22.2
100	19.1	31.6	23.1	24.3	20.2
120	16.3	27.1	21.0	21.0	16.5

for networks approaching acceptable sizes (55 onwards). What remains to be investigated further is the outlying error of the first excited state, which almost appears to dominate the total weighted error for large network sizes. This behavior is easily explained when investigating the topographical complexity of the individual sheets, as the first excited state has a far more complicated multi-minimum structure than the ground state, resulting in more strain for the network. Conversely, the ground state itself is extremely well reproduced due to its significantly simpler shape. With both convergences with respect to network size and initial guess number sufficiently expanded upon, what remains is a deeper analysis of the produced diabatic model.

Given the previously considered 75 neuron network, the cutoff-rms of  $30.0 \text{ cm}^{-1}$  turns out to be significantly smaller than the weighted rms of  $38.4 \text{ cm}^{-1}$ . The ground state by itself is reproduced with a corresponding cutoff-rms of only  $22.9 \text{ cm}^{-1}$ . The performance of the ANN model compared to the polynomial reference model is also demonstrated graphically in Figs. 4 and 5. Figure 4 shows that the ANN model consistently reproduces the *ab initio* data quantitatively over the complete energy range of 6 eV for an arbitrary cut through the 5D surfaces. While the reference model shows significant deviations of over 1 eV, the ANN fit is indistinguishable from the reference energies at the scale of the plot. This also applies to scans from the validation data set, as can be seen in Fig. 5,

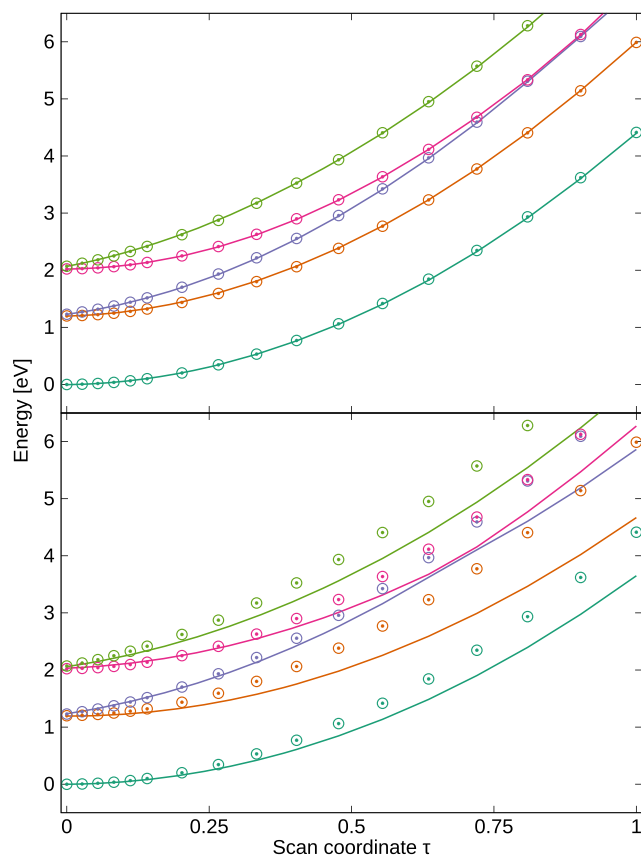


FIG. 4. Comparison of ANN model with underlying primitive reference model along a linear random cut through the 5D surfaces taken out of the training set. Above: energies produced by ANN model. Below: energies produced by polynomial reference model. Circled dots are *ab initio* fitting data.  $\tau$  is an arbitrary scan parameter.

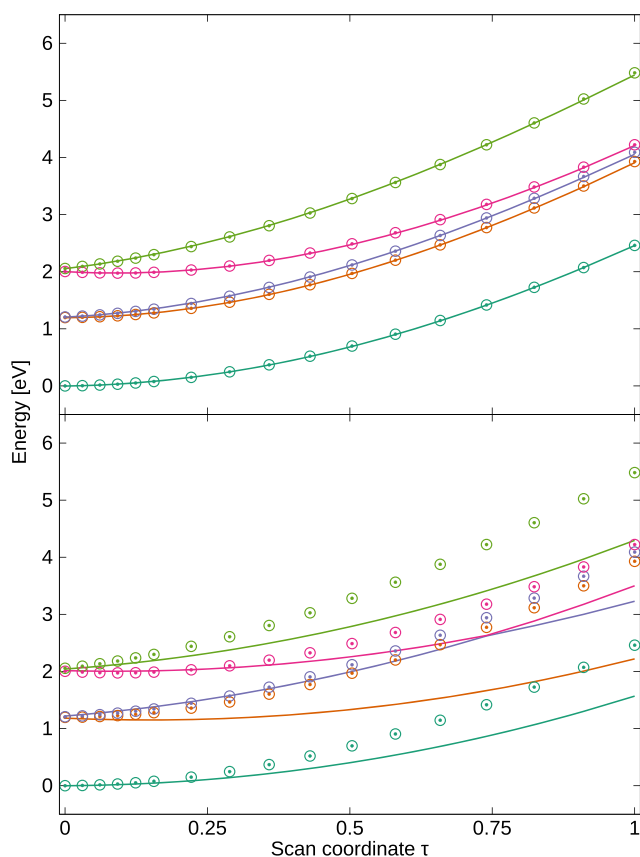


FIG. 5. Comparison of ANN model with underlying primitive reference model using validation set data. Above: energies produced by ANN model. Below: energies produced by polynomial reference model. Circled dots are *ab initio* validation data. Data points belong to a random cut through the 5D surfaces,  $\tau$  is an arbitrary scan parameter.

indicating consistent behavior beyond the fitting data. Looking at the two sheets of the  ${}^2E''$  state as a function of the  $e'$  stretch coordinates  $x_s$  and  $y_s$ , Fig. 6 displays the symmetry-induced three equivalent minima characteristic for the lower adiabatic sheet. While this behavior is to be expected of a quantitatively correct surface, it follows in no way from the underlying reference model. This is due to the fact that *linear* coupling

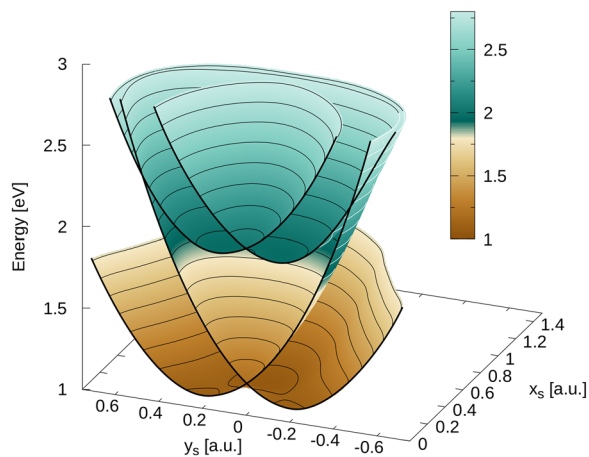


FIG. 6. Cross section of diabatic PESs of the two sheets of the  ${}^2E''$  and two sheets of the  ${}^2E'$  state of  $\text{NO}_3$  depending on the stretching modes  $x_s$  and  $y_s$ . Contour lines facing outwards are drawn white. The cross section's rim has been marked in black to highlight the present conical intersections.

terms as used in the reference model are only capable (without inclusion of higher order coupling terms) of producing rotationally symmetric “mexican hat”-like potentials. Hence, all characteristic features of the PESs are inherited from the ANN correction. The two adiabatic PES sheets belonging to the  $\tilde{B}^2E'$  state are also displayed in Fig. 6. The three equivalent PES wells in the  ${}^2E'$  manifold are less pronounced compared to the  ${}^2E''$  state and thus not so clearly visible in the figure. The figure also reveals additional intersections between the  ${}^2E''$  and  ${}^2E'$  PESs which certainly has an impact on the quantum dynamics of the photochemistry of  $\text{NO}_3$  after excitation into the  $\tilde{B}$  state. These additional conical intersections also become obvious from Fig. 7 in which the adiabatic energies of the five state components are displayed along the 1D bond dissociation coordinate  $r_{\text{NO}}$ .

The novel ANN approach presented here has one further extremely positive feature beyond its general ability to reproduce the *ab initio* data with very high accuracy. The Taylor expansions of the diabatic matrix used so far tend to produce rapid oscillations and even *unbound areas* in configuration space where no *ab initio* reference data is available. These Taylor expansions require a considerable trade-off between *local accuracy* (higher orders, greater flexibility) and *stability* of the model (lower orders, fewer oscillations). By contrast, the ANN model is found to be completely robust against such oscillations and unbound regions. Considering, for example, a scan along an N–O distance as seen in Fig. 7 which has not been included in the fit, the model shows the ability to produce a complex coupling structure of the higher states without producing oscillations in the repulsive wall. The reason for

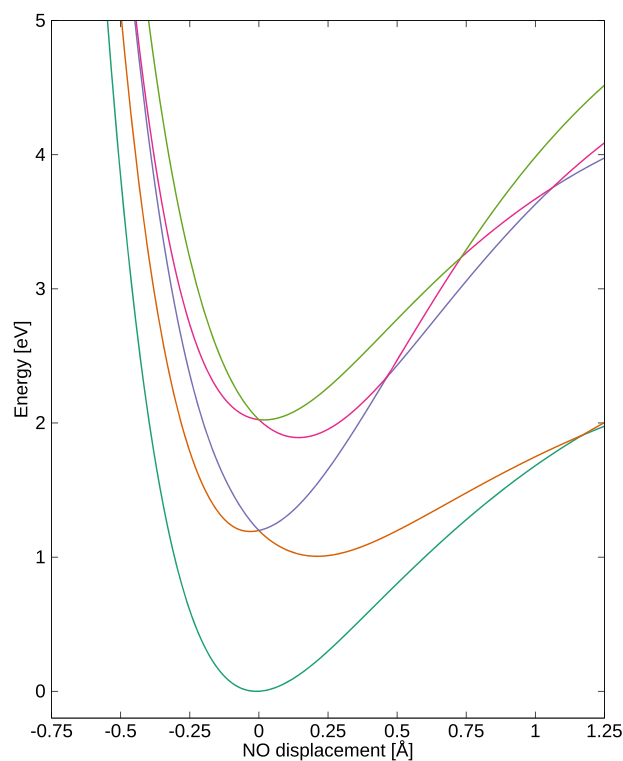


FIG. 7. Scan along NO distance around the reference point with all other coordinates frozen at the reference geometry. The ANN model reproduces conical intersections not included in the reference model and shows appropriate asymptotic behaviour.

this is that the used neural networks, unlike polynomials, are by nature composed of slowly varying sub-expressions (sigmoid functions) which merely increase in number but remain unchanged in character for larger networks while polynomial expansions introduce terms of increasingly erratic behavior which need to be compensated for by all other terms at every reasonable point in coordinate space. Furthermore, it is important to keep in mind that the neural network model does not need to reproduce the coupling terms directly. It only needs to correct a low-order polynomial model which is already intrinsically bound and free of oscillations. Each contribution is scaled in a way such that the neural network corrections reside in the order of magnitude of the expected terms. It turns out that this is an invaluable property of the novel ANN approach when used for quantum dynamics simulations. A demonstration of the strength of the ANN diabaticization method presented here will be given in a forthcoming quantum dynamics study.

## V. CONCLUSIONS AND OUTLOOK

A new diabaticization approach is presented to accurately represent the coupled PESs of strongly interacting electronic states. The core idea of this method is to express the diabatic model in terms of a primitive low-order polynomial expansion whose coefficients are then modulated by an artificial neural network (ANN). This approach combines the stability and straightforwardness of a standard low-order vibronic coupling model with the accuracy provided by the ANN as required for reliable quantum dynamics simulations. The new approach is applied to a 5-dimensional 5-state model of planar NO<sub>3</sub>. Special attention has been paid to the stability of the fitting algorithm, the overall consistency of the resulting PESs throughout the *ab initio* data, and the quality of the model with respect to the whole data set including points not used for the ANN training as well as the regions which are most relevant for future dynamics calculations, namely, the potential wells. The model has been found to excel in all of the above areas. Furthermore, extensive tests show that a single hidden layer is sufficient for this excellent performance and that “deeper” layering shows no clear benefit. While the primitive and easy to obtain reference model only provides a general bound shape of the potential roughly representing the reference data, the final model is of outstanding quality over an energy range of 6 eV. Despite being intrinsically non-linear, the algorithm produces many high-quality parameter sets even for a small number of initial guesses. Of the roughly 90 000 *ab initio* reference energies obtained from *ab initio* CASSCF/MR-SDCI calculations, a subset of 15% has been excluded from the training data, but used as validation set for a convergence test to prevent over-fitting. This validation data set has been found also to be represented with excellent accuracy by the final ANN model despite these data not being used to train the ANN.

The present study demonstrates that the proposed method is capable of producing stable, high-quality PESs based on a straightforward and easy to extend model. What remains is the extension of the model to a full-dimensional description of NO<sub>3</sub>, including the umbrella motion as well as corresponding couplings in the diabatic matrix. Furthermore, as a final study

of the resulting PESs, appropriate quantum dynamics calculations are already in progress. As the approach laid out in this paper is generic, applying it to different systems as well as different kinds of coupling (e.g., relativistic coupling) appears to be very promising.

## ACKNOWLEDGMENTS

We are grateful to the Deutsche Forschungsgemeinschaft (DFG) for generous financial support. Calculations leading to the results presented here were performed on resources provided by the Paderborn Center for Parallel Computing.

## APPENDIX: DIAGONAL MODEL TERMS

For the sake of simplicity, let  $r_{s,b}^2$  be given as

$$r_{s,b}^2 = x_{s,b}^2 + y_{s,b}^2. \quad (\text{A1})$$

The totally symmetric diagonal contributions of the reference model are expressed here in terms of three independent scalar functions  $V_i(\mathbf{Q})$

$$\mathbf{W}_{\text{diag}}^d(\mathbf{Q}) = \begin{pmatrix} V_1(\mathbf{Q}) & 0^T & 0^T \\ 0 & V_2(\mathbf{Q})\mathbf{1} & 0 \\ 0 & 0 & V_3(\mathbf{Q})\mathbf{1} \end{pmatrix} \quad (\text{A2})$$

Apart from the constant terms, referring to the vertical excitation energies at the reference point, an expansion of each  $V_i(\mathbf{Q})$  up to second order yields four (constant) coefficients  $\mu_k^i$  and the corresponding polynomial terms

$$V_i(\mathbf{Q}) = \mu_1^i \cdot a + \mu_2^i \cdot a^2 + \mu_3^i \cdot r_s^2 + \mu_4^i \cdot r_b^2. \quad (\text{A3})$$

- <sup>1</sup>M. A. Collins and D. F. Parsons, *J. Chem. Phys.* **99**, 6756 (1993).
- <sup>2</sup>J. Ischtwan and M. A. Collins, *J. Chem. Phys.* **100**, 8080 (1994).
- <sup>3</sup>T. S. Ho and H. Rabitz, *J. Chem. Phys.* **104**, 2584 (1996).
- <sup>4</sup>G. G. Maisuradze, D. L. Thompson, A. F. Wagner, and M. Minkoff, *J. Chem. Phys.* **119**, 10002 (2003).
- <sup>5</sup>G. G. Maisuradze, A. Kawano, D. L. Thompson, A. F. Wagner, and M. Minkoff, *J. Chem. Phys.* **121**, 10329 (2004).
- <sup>6</sup>B. J. Braams and J. M. Bowman, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- <sup>7</sup>C. R. Evenhuis and M. A. Collins, *J. Chem. Phys.* **121**, 2515 (2004).
- <sup>8</sup>C. R. Evenhuis, X. Lin, D. H. Zhang, D. Yarkony, and M. A. Collins, *J. Chem. Phys.* **123**, 134110 (2005).
- <sup>9</sup>O. Godsi, C. R. Evenhuis, and M. A. Collins, *J. Chem. Phys.* **125**, 104105 (2006).
- <sup>10</sup>X. Zhu, J. Y. Ma, D. R. Yarkony, and H. Guo, *J. Chem. Phys.* **136**, 234301 (2012).
- <sup>11</sup>C. Xie, J. Ma, X. Zhu, D. H. Zhang, D. R. Yarkony, D. Xie, and H. Guo, *J. Phys. Chem. Lett.* **5**, 1055 (2014).
- <sup>12</sup>*Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*, edited by W. Domcke, D. R. Yarkony, and H. Köppel (World Scientific, Singapore, 2004).
- <sup>13</sup>H. C. Longuet-Higgins, *Adv. Spectrosc.* **2**, 429 (1961).
- <sup>14</sup>W. Lichten, *Phys. Rev.* **131**, 229 (1963).
- <sup>15</sup>W. Lichten, *Phys. Rev.* **164**, 131 (1967).
- <sup>16</sup>F. T. Smith, *Phys. Rev.* **179**, 111 (1969).
- <sup>17</sup>M. Baer, *Chem. Phys.* **15**, 49 (1976).
- <sup>18</sup>H. Werner and W. Meyer, *J. Chem. Phys.* **74**, 5802 (1981).
- <sup>19</sup>C. A. Mead and D. G. Truhlar, *J. Chem. Phys.* **77**, 6090 (1982).
- <sup>20</sup>C. A. Mead, *J. Chem. Phys.* **78**, 807 (1983).
- <sup>21</sup>H.-J. Werner, B. Follmeg, and M. H. Alexander, *J. Chem. Phys.* **89**, 3139 (1988).
- <sup>22</sup>T. Pacher, L. S. Cederbaum, and H. Köppel, *J. Chem. Phys.* **89**, 7367 (1988).
- <sup>23</sup>T. Pacher, C. A. Mead, L. S. Cederbaum, and H. Köppel, *J. Chem. Phys.* **91**, 7057 (1989).

- <sup>24</sup>T. Pacher, H. Köppel, and L. S. Cederbaum, *J. Chem. Phys.* **95**, 6668 (1991).
- <sup>25</sup>T. Pacher, L. S. Cederbaum, and H. Köppel, *Adv. Chem. Phys.* **84**, 293 (1993).
- <sup>26</sup>H. Köppel, W. Domcke, and L. S. Cederbaum, *Adv. Chem. Phys.* **57**, 59 (1984).
- <sup>27</sup>M. S. Schuurman and D. R. Yarkony, *J. Chem. Phys.* **127**, 094104 (2007).
- <sup>28</sup>B. N. Papas, M. S. Schuurman, and D. R. Yarkony, *J. Chem. Phys.* **129**, 124104 (2008).
- <sup>29</sup>X. Zhu and D. R. Yarkony, *J. Chem. Phys.* **130**, 234108 (2009).
- <sup>30</sup>X. Zhu and D. R. Yarkony, *J. Chem. Phys.* **132**, 104101 (2010).
- <sup>31</sup>A. Viel and W. Eisfeld, *J. Chem. Phys.* **120**, 4603 (2004).
- <sup>32</sup>W. Eisfeld and A. Viel, *J. Chem. Phys.* **122**, 204317 (2005).
- <sup>33</sup>A. Viel, W. Eisfeld, S. Neumann, W. Domcke, and U. Manthe, *J. Chem. Phys.* **124**, 214306 (2006).
- <sup>34</sup>A. Viel, W. Eisfeld, C. R. Evenhuis, and U. Manthe, *Chem. Phys.* **347**, 331 (2008).
- <sup>35</sup>S. Faraji, H. Köppel, W. Eisfeld, and S. Mahapatra, *Chem. Phys.* **347**, 110 (2008).
- <sup>36</sup>W. Eisfeld, O. Vieuxmaire, and A. Viel, *J. Chem. Phys.* **140**, 224109 (2014).
- <sup>37</sup>R. Cimraglia, J. P. Malrieu, M. Persico, and F. Spiegelmann, *J. Phys. B: At. Mol. Phys.* **18**, 3073 (1985).
- <sup>38</sup>W. Domcke and C. Woywod, *Chem. Phys. Lett.* **216**, 362 (1993).
- <sup>39</sup>G. J. Atchity and K. Ruedenberg, *Theor. Chem. Acc.* **97**, 47 (1997).
- <sup>40</sup>H. Nakamura and D. G. Truhlar, *J. Chem. Phys.* **115**, 10353 (2001).
- <sup>41</sup>H. Nakamura and D. G. Truhlar, *J. Chem. Phys.* **117**, 5576 (2002).
- <sup>42</sup>H. Nakamura and D. G. Truhlar, *J. Chem. Phys.* **118**, 6816 (2003).
- <sup>43</sup>P. Cattaneo and M. Persico, *Theor. Chem. Acc.* **103**, 390 (2000).
- <sup>44</sup>S. Nangia and D. G. Truhlar, *J. Chem. Phys.* **124**, 124309 (2006).
- <sup>45</sup>Z. H. Li, R. Valero, and D. G. Truhlar, *Theor. Chem. Acc.* **118**, 9 (2007).
- <sup>46</sup>X. Zhu and D. R. Yarkony, *J. Chem. Phys.* **140**, 024112 (2014).
- <sup>47</sup>T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, *J. Chem. Phys.* **103**, 4129 (1995).
- <sup>48</sup>D. F. R. Brown, M. N. Gibbs, and D. C. Clary, *J. Chem. Phys.* **105**, 7597 (1996).
- <sup>49</sup>K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, and H. A. Scheraga, *Chem. Phys. Lett.* **271**, 152 (1997).
- <sup>50</sup>F. V. Prudente, P. H. Acioli, and J. J. S. Neto, *J. Chem. Phys.* **109**, 8801 (1998).
- <sup>51</sup>S. Lorenz, A. Gross, and M. Scheffler, *Chem. Phys. Lett.* **395**, 210 (2004).
- <sup>52</sup>L. M. Raff, M. Malshe, M. Hagan, D. I. Doughan, M. G. Rockley, and R. Komanduri, *J. Chem. Phys.* **122**, 084104 (2005).
- <sup>53</sup>S. Lorenz, M. Scheffler, and A. Gross, *Phys. Rev. B* **73**, 115431 (2006).
- <sup>54</sup>S. Manzhos, X. G. Wang, R. Dawes, and T. Carrington, *J. Phys. Chem. A* **110**, 5295 (2006).
- <sup>55</sup>S. Manzhos and T. Carrington, Jr., *J. Chem. Phys.* **125**, 084109 (2006).
- <sup>56</sup>S. Manzhos and T. Carrington, Jr., *J. Chem. Phys.* **125**, 194105 (2006).
- <sup>57</sup>J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- <sup>58</sup>S. Manzhos and T. Carrington, Jr., *J. Chem. Phys.* **127**, 014103 (2007).
- <sup>59</sup>M. Malshe, R. Narulkar, L. M. Raff, M. Hagan, S. Bukkapatnam, and R. Komanduri, *J. Chem. Phys.* **129**, 044111 (2008).
- <sup>60</sup>S. Manzhos and T. Carrington, Jr., *J. Chem. Phys.* **129**, 224104 (2008).
- <sup>61</sup>J. Behler, *J. Chem. Phys.* **134**, 074106 (2011).
- <sup>62</sup>H. T. T. Nguyen and H. M. Le, *J. Phys. Chem. A* **116**, 4629 (2012).
- <sup>63</sup>B. Jiang and H. Guo, *J. Chem. Phys.* **139**, 054112 (2013).
- <sup>64</sup>W. Koch and D. H. Zhang, *J. Chem. Phys.* **141**, 021101 (2014).
- <sup>65</sup>M. Majumder, S. E. Hegger, R. Dawes, S. Manzhos, X.-G. Wang, C. Tucker, Jr., J. Li, and H. Guo, *Mol. Phys.* **113**, 1823 (2015).
- <sup>66</sup>G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- <sup>67</sup>H. D. Meyer, U. Manthe, and L. S. Cederbaum, *Chem. Phys. Lett.* **165**, 73 (1990).
- <sup>68</sup>U. Manthe, H. D. Meyer, and L. S. Cederbaum, *J. Chem. Phys.* **97**, 3199 (1992).
- <sup>69</sup>Y. Guan, B. Fu, and D. H. Zhang, *J. Chem. Phys.* **147**, 224307 (2017).
- <sup>70</sup>T. Lenzen and U. Manthe, *J. Chem. Phys.* **147**, 084105 (2017).
- <sup>71</sup>M. T. Hagan and M. B. Menhaj, *IEEE Trans. Neural Networks* **5**, 989 (1994).
- <sup>72</sup>N. Wittenbrink, F. Venghaus, D. Williams, and W. Eisfeld, *J. Chem. Phys.* **145**, 184108 (2016).
- <sup>73</sup>W. Eisfeld and K. Morokuma, *J. Chem. Phys.* **113**, 5587 (2000).
- <sup>74</sup>W. Eisfeld and K. Morokuma, *J. Chem. Phys.* **114**, 9430 (2001).
- <sup>75</sup>A. Viel and W. Eisfeld, *Chem. Phys.* **509**, 81 (2018).
- <sup>76</sup>W. Eisfeld, *J. Chem. Phys.* **134**, 054303 (2011).
- <sup>77</sup>M. Hagan, H. Demuth, M. Beale, and O. De Jesús, *Neural Network Design*, 2nd ed. (Martin Hagan, 2014), ISBN: 9780971732117, URL: <https://books.google.de/books?id=4EW9oQEACAAJ>.