



**HAL**  
open science

# Complete Nuclear Permutation Inversion Invariant Artificial Neural Network (CNPI-ANN) Diabatization for the Accurate Treatment of Vibronic Coupling Problems

David M. G. Williams, Wolfgang Eisfeld

► **To cite this version:**

David M. G. Williams, Wolfgang Eisfeld. Complete Nuclear Permutation Inversion Invariant Artificial Neural Network (CNPI-ANN) Diabatization for the Accurate Treatment of Vibronic Coupling Problems. *Journal of Physical Chemistry A*, 2020, 124 (37), pp.7608-7621. 10.1021/acs.jpca.0c05991 . hal-02984550

**HAL Id: hal-02984550**

**<https://hal.science/hal-02984550>**

Submitted on 31 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Complete Nuclear Permutation Invariant Artificial Neural Network (CNPI-ANN) Diabatization for the Accurate Treatment of Vibronic Coupling Problems

Published as part of *The Journal of Physical Chemistry virtual special issue "Machine Learning in Physical Chemistry"*.

David M. G. Williams\* and Wolfgang Eisfeld\*

Cite This: *J. Phys. Chem. A* 2020, 124, 7608–7621

Read Online

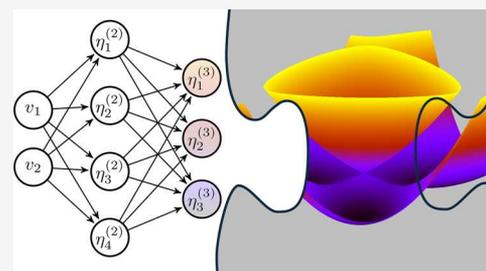
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** A recently developed scheme to produce accurate high-dimensional coupled diabatic potential energy surfaces (PESs) based on artificial neural networks (ANNs) [*J. Chem. Phys.* 2018, 149, 204106 and *J. Chem. Phys.* 2019, 151, 164118] is modified to account for the proper complete nuclear permutation inversion (CNPI) invariance. This new approach cures the problem intrinsic to the highly flexible ANN representation of diabatic PESs to account for the proper molecular symmetry accurately. It turns out that the use of CNPI invariants as coordinates for the input layer of the ANN leads to a much more compact and thus more efficient representation of the diabatic PES model without any loss of accuracy. In connection with a properly symmetrized vibronic coupling reference model, which is modified by the output neurons of the CNPI-ANN, the resulting adiabatic PESs show perfect symmetry and high accuracy. In the present paper, the new approach will be described and thoroughly tested. The test case is the representation and corresponding vibrational/vibronic nuclear dynamics of the low-lying electronic states of planar  $\text{NO}_3$  for which a large number of ab initio data is available. Thus, the present results can be compared directly with the previous studies.



## 1. INTRODUCTION

The fundamental theoretical understanding of chemical processes by quantum dynamics simulations depends critically on the availability of accurate models for the molecular potential energy, the so-called potential energy surfaces (PESs). The development of single, uncoupled PESs for systems of more than three atoms made impressive progress in the past couple of decades.<sup>1–6</sup> Particularly the use of artificial neural networks (ANNs) seems to be a very promising approach to achieve high accuracy even for high-dimensional problems.<sup>7–24</sup> The decisive advantage of ANNs may be that it can be shown mathematically that they are capable of uniformly representing any continuous real function of  $n$  dimensions up to arbitrary accuracy.<sup>25</sup> Thus, in principle, ANNs should be capable of representing arbitrary PESs up to the limitations of the underlying data. Besides the potential accuracy, they also can be evaluated very efficiently once trained, as their evaluation consists mostly of matrix-vector multiplications. This is an invaluable advantage for the use with quantum dynamics methods such as the multiconfiguration time-dependent Hartree (MCTDH) method,<sup>26,27</sup> especially when using the correlated discrete variable representation (CDVR)<sup>28</sup> scheme for which the evaluation of the PES is the most time-demanding part. Alternatively to the CDVR approach, a sum of products form of the PES model must be used, which is also possible in the ANN framework.<sup>16,24,29</sup> A

further advantage is that the training of ANNs can be done very efficiently and is much less demanding than other nonlinear optimization approaches. Of course there are also problems of which one of the most prominent is an issue commonly referred to as “over-fitting”. This algorithmic problem, related to the extremely high flexibility and nonlinearity of typical ANNs, limits the accuracy of the function representation and has to be accounted for in the training algorithm. However, the impressive results already obtained by the ANN representation of single molecular PESs based on high-level ab initio data demonstrate the power of this approach.<sup>14,23,24,29–32</sup>

The situation is way more complicated in the case that multiple and strongly interacting electronic states are involved in the process. For such nonadiabatic processes, not only the PESs corresponding to different states but also the interactions between those states have to be represented by a PES model. One fundamental problem is here that the Born–Oppenheimer

Received: June 30, 2020

Revised: July 31, 2020

Published: August 11, 2020



approximation will break down in regions in nuclear configuration space (NCS), where interactions among electronic states become significant or even singular as is the case for conical intersections.<sup>33</sup> It is well-established by now that the solution in such cases is to use a *quasi-diabatic* representation of the coupled electronic states and corresponding PES model.<sup>33–49</sup> For the sake of brevity, we will refer to quasi-diabatic representations as “diabatic” representations from here on. The diabatic representation removes singularities in the nonadiabatic couplings and nondifferentiable points on the PESs but is neither unique nor easily available. Many methods have been developed to find the unitary basis transformation between the adiabatic states and energies directly available from ab initio calculations and diabatic states and PES matrices.<sup>39,42,43,45,47–76</sup> One particularly successful and straightforward approach is the multimode linear vibronic coupling method by Köppel, Domcke, and Cederbaum<sup>50</sup> and similar approaches.<sup>70,72,77,78</sup> These models are sufficient to treat many ultrafast nonadiabatic processes but are not accurate enough for many problems that are more complicated and depend critically on more extended regions in the NCS. Thus, the vibronic coupling approach has been extended to higher coupling orders as well.<sup>54–66,68,69,74</sup> However, the use of higher-order Taylor expansions for the representation of the diabatic PES matrices is quite cumbersome, and therefore alternative representations have been developed as well.<sup>57,79–86</sup>

The latest development in the pursuit of finding better diabatic PES models is the use of ANNs.<sup>87–97</sup> Different strategies have been followed of which some are based on the use of complete nuclear permutation inversion (CNPI) invariant polynomials of the input coordinates.<sup>89,91,92</sup> Our recently developed approach presented in ref 90 and thoroughly analyzed in ref 95 is based on a different strategy and uses all symmetry-adapted valence coordinates of the molecule rather than CNPI invariants. The reason for this is that the approach combines the basic, simple structure of a low-order vibronic coupling *ansatz*, which is tuned by an ANN so that very accurate results are obtained. The high flexibility of the ANN introduces slight symmetry contaminations, which are removed by an a posteriori symmetrization. The new method presented here circumvents the problem of the symmetry contamination by using CNPI invariants as an input layer of the ANN in combination with a standard low-order vibronic coupling reference model.

To stay consistent with our previous studies,<sup>90,95</sup> the performance of the new approach was tested using planar (SD) NO<sub>3</sub>, for which plenty of data and experience are available in the group.<sup>64,65,74,98</sup> A 5 × 5 diabatic model for the ground and first two excited electronic states of <sup>2</sup>A<sub>2</sub>′, <sup>2</sup>E′, and <sup>2</sup>E′ symmetry is trained, and vibrational/vibronic eigenstates are computed. The latter can be compared to previous theoretical<sup>69,74,95,99–105</sup> and experimental results.<sup>106,107</sup>

## 2. THEORY

**2.1. Adiabatic and Diabatic Representation.** At the center of any diabaticization approach is the adiabatic-to-diabatic transformation that connects an adiabatic and a diabatic basis of electronic states. There are many established ways to do this, and for a deeper discussion we refer to the literature, for example, ref 33. The necessary information for this basis transformation usually is obtained from adiabatic energies only (“diabatization by *ansatz*”), by using the adiabatic wave function representation alone (“block diagonalization”), or by a combination of both

(“hybrid diabaticization”).<sup>76</sup> The method presented here is designed for the hybrid diabaticization approach, though in the example calculations, the wave function information is not used. The fundamental requirement utilized in this method is that the adiabatic and diabatic PES matrices are related by the unitary transformation **U** that diagonalizes the diabatic matrix, and the eigenvalues reproduce the adiabatic energies

$$\mathbf{U}^\dagger \mathbf{W}^d \mathbf{U} = \mathbf{W}^a = \text{diag}(W_j^a) \quad (1)$$

The eigenvectors are related to the adiabatic electronic wave functions and can also be used to determine the matrix elements of the diabatic PES matrix **W**<sup>d</sup>.

**2.2. Artificial Neural Networks.** The diabaticization approach presented here utilizes a multilayer perceptron feed-forward neural network. The input layer  $\boldsymbol{\eta}^{(1)}$  consists of the molecular coordinates (cf. Section 2.3), which are connected to the hidden layers  $\boldsymbol{\eta}^{(k)}$  according to

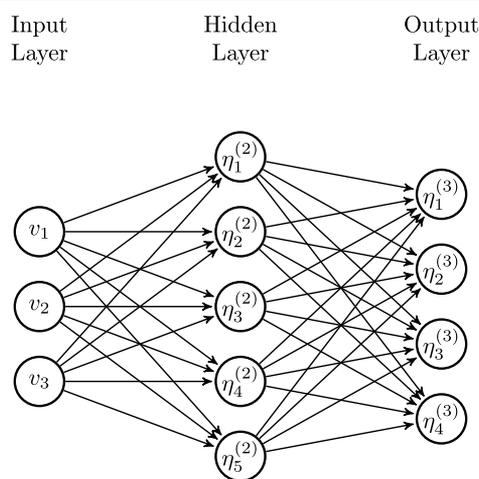
$$\eta_j^{(k)} = f^{(k)} \left( \underbrace{\beta_j^{(k)} + \sum_l \omega_{jl}^{(k)} \eta_l^{(k-1)}}_{f^{(k)}(\chi_j^{(k)})} \right) \quad (2)$$

The activation functions  $f^{(k)}$  used in the present version depend on the layer  $k$ . While the output layer ( $k = 3$ ) uses the identity Id, the activation function for the hidden layer reads

$$f^{(2)}(\chi) = \sqrt{3} \cdot \tanh(A \cdot \chi) \quad (3)$$

with  $A = \text{artanh}\left(\frac{1}{\sqrt{3}}\right)$ . This choice of factors rescales the sigmoid activation function in such a way that its “linear domain” resides in the interval  $[-1, 1]$ , approximating the identity, while the exponential plateaus reside outside of this interval. The output layer is used to parametrize the diabatic PES matrix (cf. Section 2.4). A simplified version of the neural network utilized (with a reduced number of neurons in each layer) can be found in Figure 1.

The ANN is trained by a modified Marquardt–Levenberg back-propagation method described in ref 90. The modification is necessary, because the output layer cannot be compared directly to the reference data. Only the eigenvalues (and eigenvectors) of the matrix parametrized by the output neurons



**Figure 1.** A neural network with one hidden layer. Arrows correspond to weighted contributions. Bias terms and activation functions are not visualized.

can be used to determine the root-mean-square error (RMSE), which is minimized by the method. Therefore, an additional numerical differentiation of the eigenvalues (and eigenvectors) with respect to the output neurons is required.

One notorious problem with the ANN training commonly is called overfitting and is related to the very high flexibility of the ANN. Since the training data are limited, an optimal representation of this reference data might result in a poor representation of data not contained in the reference set. This can be avoided, at least to some degree, by a method called “early stopping”. For this, the complete available data set is partitioned into a reference and a validation set. Only the reference set is used for the actual training to minimize the reference RMSE. In each step also the validation RMSE is computed, and if the latter rises, the training is stopped. The details of how the early stopping is implemented in the present method has been described in ref 90.

While overfitting is a crucial aspect, it is naturally not the only limitation to consider. ANNs are generally rather sensitive toward the data set with which they are trained. A simple, yet important example of this sensitivity lies in the sheer amount of data required, as neural networks tend to extrapolate data rather poorly. However, the use of a robust low-level reference model tuned by an ANN as presented here at least ensures that the extrapolated regions behave reasonably without showing pathologies. Another, more subtle problem related to the input data and the ANN training arises from the range of the input data. Since the present training method relies on derivatives of the activation function with respect to (wrt) its argument, initial convergence will be slowed tremendously if the derivatives are very small. Given the present activation eq 3, this means that  $\chi$  should, on average, reside in the interval  $[-1, 1]$ . For this purpose, the molecular coordinates  $\eta^{(1)}$  are scaled and shifted in such a manner that all  $\eta_j^{(1)}$  have a mean of 0 and a standard deviation of 1 across the presented data set. This is achieved by first computing the mean  $\bar{\eta}_j$  and standard deviation  $\sigma(\eta_j)$  of the coordinates  $\eta_j^{(1)}$  over the training data; in other words

$$\bar{\eta}_j = \frac{1}{N_D} \sum_{k=1}^{N_D} \eta_j^{(1)}(\mathbf{Q}_k) \quad (4a)$$

$$\sigma(\eta_j) = \sqrt{\frac{1}{N_D} \sum_{k=1}^{N_D} (\eta_j^{(1)}(\mathbf{Q}_k) - \bar{\eta}_j)^2} \quad (4b)$$

where  $N_D$  is the number of ab initio data points, and  $\eta_j^{(1)}(\mathbf{Q}_k)$  is the explicit value of  $\eta_j^{(1)}$  at the given data point. The normalized input coordinates  $\tilde{\eta}_j$  actually used in the training process are then obtained by

$$\tilde{\eta}_j = \frac{\eta_j^{(1)} - \bar{\eta}_j}{\sigma(\eta_j)} \quad (5)$$

which by definition has an average of 0 and a standard deviation of 1 over the given set of ab initio data. Once the training is completed, the optimized parameters corresponding to  $\omega_{jl}^{(2)}$  and  $\beta_j^{(2)}$  from Equation 2 are modified to obtain an equivalent neural network depending on the original coordinates  $\eta_j^{(1)}$ , eliminating the need to compute the means and variances when evaluating the PES model during dynamics calculations. The equivalent parameters  $\tilde{\omega}_{jl}^{(2)}$  and  $\tilde{\beta}_j^{(2)}$  follow from Equation 2 for  $k = 2$  by

substituting  $\eta_j^{(1)}$  with its normalized version of Equation 5 and read

$$\tilde{\omega}_{jl}^{(2)} = \frac{\omega_{jl}^{(2)}}{\sigma(\eta_l)} \quad (6a)$$

$$\begin{aligned} \tilde{\beta}_j^{(2)} &= \beta_j^{(2)} - \sum_l \frac{\omega_{jl}^{(2)} \bar{\eta}_l}{\sigma(\eta_l)} \\ &= \beta_j^{(2)} - \sum_l \tilde{\omega}_{jl}^{(2)} \bar{\eta}_l \end{aligned} \quad (6b)$$

In the present work this preconditioning is combined with an appropriate choice of random initial weights and biases to ensure better convergence.

**2.3. Coordinates and CNPI Invariants.** The *ansatz* diabatic PES matrix has to fulfill strict symmetry requirements for the matrix elements, because the represented Hamiltonian must be invariant under all symmetry transformations. However, the off-diagonal elements are not necessarily totally symmetric, which has to be taken into account. In standard vibronic coupling approaches the matrix elements are usually expressed as polynomials in terms of symmetry-adapted coordinates  $Q$  according to

$$W_{kj}^d(Q) = \sum_{\alpha} p_{\alpha}^{kj} \prod_l Q_l^{n_{\alpha l}} \quad (7)$$

The expansion coefficients  $p$  are strictly related by symmetry, and the coordinates transform according to different irreducible representations of the molecular point group. Alternatively, the diabatic matrix can be constructed from CNPI invariants of the actual CNPI group and the invariants of its permutation subgroups.<sup>91,108</sup>

The present approach is illustrated for the  $D_{3h}$  symmetric  $\text{NO}_3$  radical in the following. In this case it is convenient to use symmetry-adapted coordinates for the representation of the vibronic coupling model, because the groups  $D_{3h}$  and  $S_3 \times C_i$  are isomorphic. The five symmetry-adapted coordinates excluding the umbrella motion read

$$a = \sqrt{\frac{1}{3}} (m_1 + m_2 + m_3) \quad (8a)$$

$$x_s = \sqrt{\frac{1}{6}} (2m_1 - m_2 - m_3) \quad (8b)$$

$$y_s = \sqrt{\frac{1}{2}} (m_2 - m_3) \quad (8c)$$

$$x_b = \sqrt{\frac{1}{6}} (2\alpha'_1 - \alpha'_2 - \alpha'_3) \quad (8d)$$

$$y_b = \sqrt{\frac{1}{2}} (\alpha'_2 - \alpha'_3) \quad (8e)$$

in which the transformed distances  $m_i$  and angles  $\alpha'_i$  are

$$m_i = 1 - \exp(-\gamma(r_i - r_0)) \quad (9a)$$

$$\alpha'_i = \frac{\alpha_i - \alpha_0}{r_i r_k}, \quad i \neq j \neq k \quad (9b)$$

The coordinate  $a$  transforms totally symmetric as  $a_1'$ , and the vectors  $(x, y)$  transform as  $e'$ . The resulting invariants are determined as<sup>64,65,74,109,110</sup>

$$v^{(0)} = 1 \quad (10a)$$

$$v^{(1)} = 0 \quad (10b)$$

$$v^{(2)} = (x^2 + y^2) \quad (10c)$$

$$v^{(3)} = (2x^3 - 6xy^2) \quad (10d)$$

$$v^{(4)} = (x^4 + 2x^2y^2 + y^4) \quad (10e)$$

for the first few orders, and the mixed terms of two different  $e'$  modes read

$$v_{ee}^{(2,1)} = 2(x_1x_2 + y_1y_2) \quad (11a)$$

$$v_{ee}^{(3,1)} = 2(x_1x_2^2 - x_1y_2^2 - 2x_2y_1y_2) \quad (11b)$$

$$v_{ee}^{(3,2)} = 2(x_1^2x_2 - x_2y_1^2 - 2x_1y_1y_2) \quad (11c)$$

$$v_{ee}^{(4,1)} = 2(x_1^2x_2^2 - x_1^2y_2^2 - x_2^2y_1^2 + y_1^2y_2^2 + 4x_1x_2y_1y_2) \quad (11d)$$

$$v_{ee}^{(4,2)} = 2(x_1^3x_2 + x_1^2y_1y_2 + x_1x_2y_1^2 + y_1^3y_2) \quad (11e)$$

$$v_{ee}^{(4,3)} = 2(x_2^3x_1 + x_1x_2y_2^2 + x_2^2y_1y_2 + y_2^3y_1) \quad (11f)$$

$$v_{ee}^{(4,4)} = x_1^2x_2^2 + x_1^2y_2^2 + x_2^2y_1^2 + y_1^2y_2^2 \quad (11g)$$

Each of the above invariants may be multiplied by an arbitrary power of the  $a$  mode and still remain an invariant. Similarly, if we were to include the umbrella motion  $\theta$ , we could also multiply each term with arbitrary powers of  $\theta^2$ . Since for  $\text{NO}_3$  the CNPI group and the  $D_{3h}$  point group are isomorphic, the above point group invariants can be used for convenience. In cases in which point group invariants are not sufficient, the CNPI invariants may be constructed following the techniques developed by Braams and Bowman.<sup>6</sup> Finally, it may be worth noting that, in the present planar case, the CNPI and CNP symmetries are equivalent.

**2.4. Diabatic Model (Ansatz).** The new approach will be demonstrated using a  $5 \times 5$  diabatic PES model to represent the electronic Hamiltonian corresponding to the first few electronic states of planar  $\text{NO}_3$ , namely,  $\tilde{X}^2A_2'$ ,  $\tilde{A}^2E''$ , and  $\tilde{B}^2E'$ . The diabatic PES model is based on a simple vibronic reference model to simulate the basic physics of the problem, which then is tuned by the ANN to achieve high accuracy. This reference model is expressed in terms of a set of coupling matrices depending on the symmetry-adapted coordinates  $Q$  parametrized by expansion coefficients  $\lambda$  and reads

$$\mathbf{W}^d(\mathbf{Q}) = \mathbf{W}_{\text{diag}}^d(\mathbf{Q}) + \begin{pmatrix} \lambda_1 & 0^T & 0^T \\ 0 & \lambda_2 \mathbf{1} & 0 \\ 0 & 0 & \lambda_3 \mathbf{1} \end{pmatrix} + \begin{pmatrix} 0 & 0^T & 0^T \\ 0 & \lambda_4 \mathbf{e}_s + \lambda_5 \mathbf{e}_b & 0 \\ 0 & 0 & \lambda_6 \mathbf{e}_s + \lambda_7 \mathbf{e}_b \end{pmatrix} + \begin{pmatrix} 0 & 0^T & \lambda_8 \boldsymbol{\rho}_s^T + \lambda_9 \boldsymbol{\rho}_b^T \\ 0 & 0 & 0 \\ \lambda_8 \boldsymbol{\rho}_s + \lambda_9 \boldsymbol{\rho}_b & 0 & 0 \end{pmatrix} \quad (12)$$

$\mathbf{1}$  in the above supermatrix is a  $2 \times 2$  unit matrix, the matrix blocks  $\mathbf{e}_{s,b}$  are the well-known linear Jahn–Teller coupling matrices, and the vectors  $\boldsymbol{\rho}_{s,b}$  represent the linear pseudo-Jahn–Teller couplings, which are given explicitly as

$$\mathbf{e}_{s,b} = \begin{pmatrix} x_{s,b} & y_{s,b} \\ y_{s,b} & -x_{s,b} \end{pmatrix} \text{ and } \boldsymbol{\rho}_{s,b} = \begin{pmatrix} y_{s,b} \\ x_{s,b} \end{pmatrix} \quad (13)$$

The symmetry-adapted coordinates corresponding to the degenerate asymmetric bending and stretching modes are labeled  $x_{s,b}$ ,  $y_{s,b}$ , respectively. The diabatic PES matrix eq 12 can be recast in a more compact form as

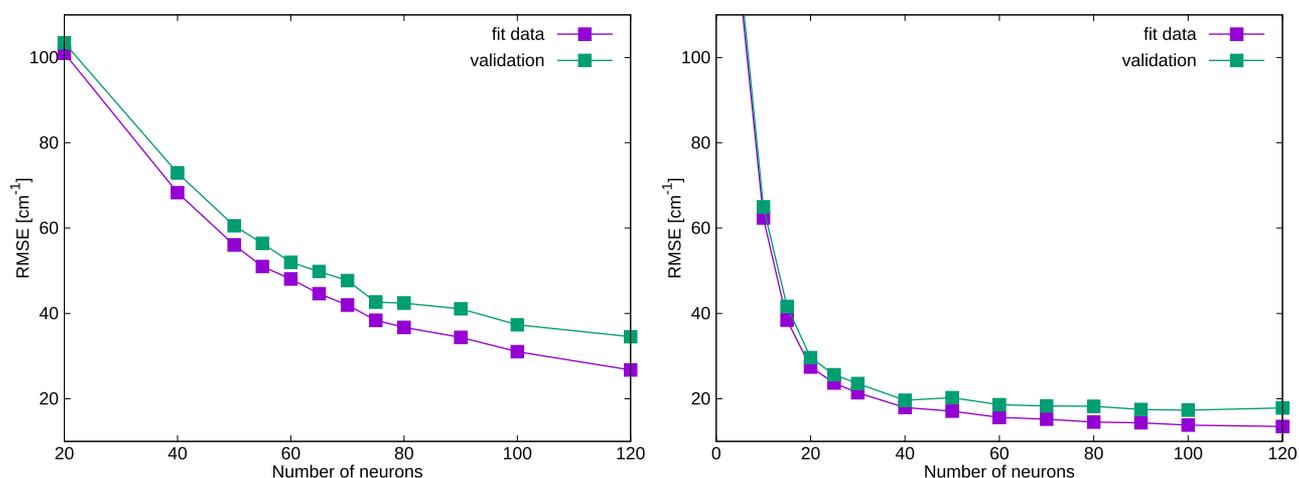
$$\mathbf{W}^d(\mathbf{Q}) = \sum_L \lambda_L(\mathbf{v}) \cdot \mathbf{M}_L(\mathbf{Q}) \quad (14)$$

utilizing explicit coupling matrices  $\mathbf{M}_L$ . The ANN is introduced into this model through the expansion coefficients

$$\lambda_L(\mathbf{v}) = \begin{cases} \lambda_L^0 & \text{for uncorrected terms} \\ \lambda_L^0 \cdot (1 + \eta_L^{(f)}(\mathbf{v})) & \text{otherwise} \end{cases} \quad (15)$$

The constants  $\lambda_L^0$ , obtained from standard nonlinear fitting procedures, are tuned by the outputs  $\eta_L^{(f)}(\mathbf{v})$  of a trained ANN.

Note that, in our previous studies, the ANN was fed with all symmetry-adapted coordinates  $Q$ , while in the present approach the invariants  $v$  are used. This turns out to make a significant difference. Though the vibronic coupling model with all output neurons  $\eta_L^{(f)}$  set to zero fulfills the symmetry transformation invariance exactly, this is not the case anymore when the output neurons depend on the symmetry-adapted coordinates  $Q$ . One could say that the ANN simply does not know about the symmetry of the problem except for what it can “learn” from the reference data. The errors are very small ( $\sim 1 \text{ cm}^{-1}$ ) and have been removed by an a posteriori resymmetrization in our previous studies. However, the new approach presented here deletes this issue intrinsically by feeding the ANN with invariants rather than symmetry-adapted coordinates. Thus, on the one hand, regardless of the various equivalent realizations of a molecular geometry related by symmetry transformations, the values of the invariants are the same and thus enforce the proper symmetry of the model. On the other hand, it is not obvious whether or not the invariants contain all the coordinate information represented by the symmetry-adapted coordinates. The lowest-order invariants are  $a$ ,  $v^{(2)}(x_s, y_s)$ ,  $v^{(2)}(x_b, y_b)$ , and  $v_{ee}^{(2,1)}(x_s, y_s, x_b, y_b)$ . This means that there are only four instead of five coordinates available. Furthermore, these invariants do only represent information about the radial displacement of the



**Figure 2.** Weighted RMSE of the best neural network of a given set as a function of the hidden-layer size. Purple line represents the fitting set error; green line represents the validation set error. (left) Previous model. (right) New model.

molecular geometry with respect to the conical intersection but none about the tangential displacement with respect to the three equivalent positions around the conical intersection.<sup>110</sup> The missing information can be retrieved by utilizing the next higher order of invariants as will be demonstrated by numerical tests. It turns out that adding the four third-order invariants provides sufficient input information for an accurate representation of the full diabatic PES model.

Finally, the cubic terms in the invariants on the input layer of the ANN cause small numerical errors, which only show in the detailed results of the quantum dynamics calculations for the vibrational/vibronic levels of the system. This is avoided by damping the invariants according to

$$v'_k(Q) = c_k \tanh(d_k v_k(Q)) \quad (16)$$

where  $c_k$  and  $d_k$  are empirical scaling factors chosen according to the physical problem to be treated. A reasonable choice for this scaling appears to be  $c_k = d_k^{-1} = 10$ , and this assumption is tested numerically in Section 4.3.

This diabatic PES model is designed such that the stability and basic qualitative description of a low-order vibronic coupling expansion is combined with the flexibility of the ANN to achieve excellent accuracy. We found in previous studies that the basic reference model should be restricted to the lowest reasonable orders to avoid numerical problems in the ANN training. This also ensures that the model is free of pathologies. The accuracy is not limited by this reference model but by the ANN tuning the corresponding coupling parameters according to eq 15. What remains is to fit the reference parameters  $\lambda_L^0$  and to train the ANN with respect to ab initio reference data.

### 3. COMPUTATIONAL DETAILS

The present method is tested by representing the  $5 \times 5$  diabatic Hamiltonian for the  $\text{NO}_3$  radical. The training of the ANN is based on an extensive set of ab initio reference energies computed by Multiconfiguration 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>111</sup> All details of the ab initio computations can be found in refs 74, 98, 111, and 112. The sampling approach for the data points is based on the method described in refs 113 and 90.

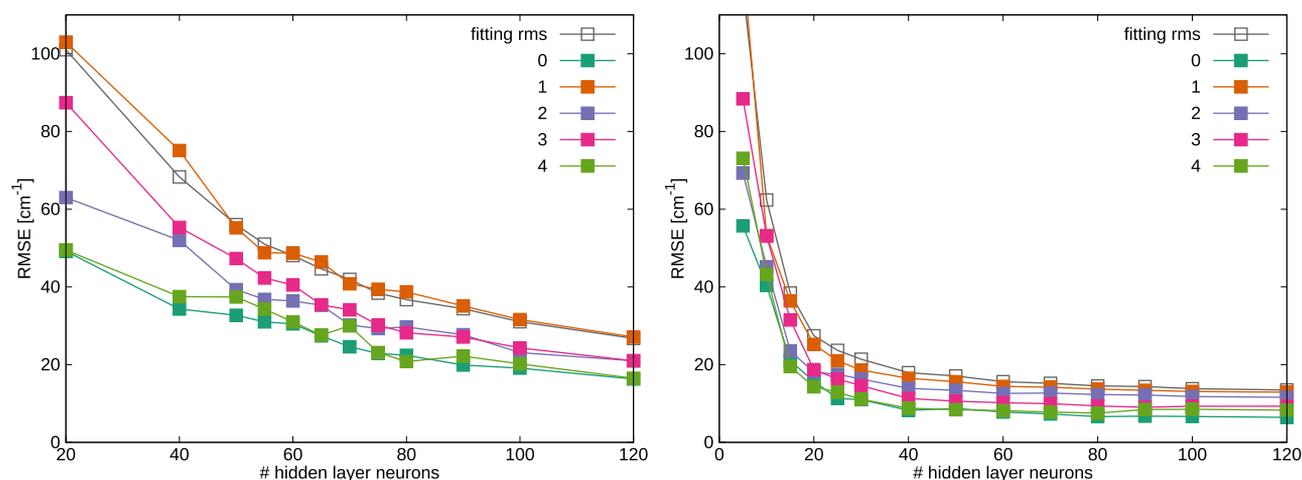
These reference data are used first to fit the 21 free parameters  $\lambda_j^0$  of the reference model using a Marquardt–Levenberg algorithm incorporated into a genetic algorithm with a resulting RMSE of  $\sim 1700 \text{ cm}^{-1}$ . This is the same reference model as used before. Then the reference data set is partitioned into a training set of 85% of the data and an external validation set of the remaining 15% of the energies. The training set is used to train CNPI-ANNs with a single hidden layer of various sizes. The validation set is used in the early stopping mechanism to prevent overfitting as suggested in ref 114. The specific way how early stopping is used here has been described in ref 90. All data are energy weighted as described before in order to get the best accuracy of the PES model in the region of the NCS most relevant for the nuclear dynamics.

The resulting PES models were tested by computing vibrational/vibronic energy levels on the adiabatic ground state as well as the first electronically excited state, which is separable for planar geometries. The second excited state of  ${}^2E'$  symmetry cannot be tested easily by the dynamics calculations, because it is coupled to the  ${}^2A_2'$  ground state by *pseudo*-JT coupling, and thus vibronic eigenstates would not be accessible easily and therefore are not computed in the present study. The computation of the vibrational/vibronic energy levels was performed using a time-independent Hermite discrete variable representation (Hermite-DVR) method,<sup>115</sup> in which the corresponding Hamiltonian was diagonalized by an exact short iterative Lanczos method. Further details can be found in refs 74, 90, 98, and 105.

### 4. RESULTS AND DISCUSSION

The new scheme is tested and compared to its predecessor laid out in refs 90 and 95. First, both are compared in terms of fitting performance and convergence with respect to the number of parameters needed. In a second step, both models are compared in terms of the dynamics results for both  ${}^2A_2'$  ground and  ${}^2E''$  excited states, respectively. Finally, the new approach will be assessed by its capability of reproducing the available experimental data,<sup>106,107</sup> in comparison to previous efforts.

While various ANN parametrizations will be used to provide deeper insights regarding the stability of the produced results (wrt the initial guess), each scheme will be primarily represented by one specific parametrization as a reference case, which turned out to be particularly suitable for the present case. The previous



**Figure 3.** State-resolved convergence behavior of unweighted RMSE for ab initio data points up to 1 eV above respective reference point energy. States are enumerated from the ground state to the energetically highest lying state. Gray: weighted fitting set error. (left) Previous model. (right) Current model.

method will be represented by an ANN model with a single hidden layer of 75 neurons (corresponding to 1134 formal parameters), and the new model has a single hidden layer of 30 neurons (corresponding to 549 formal parameters).

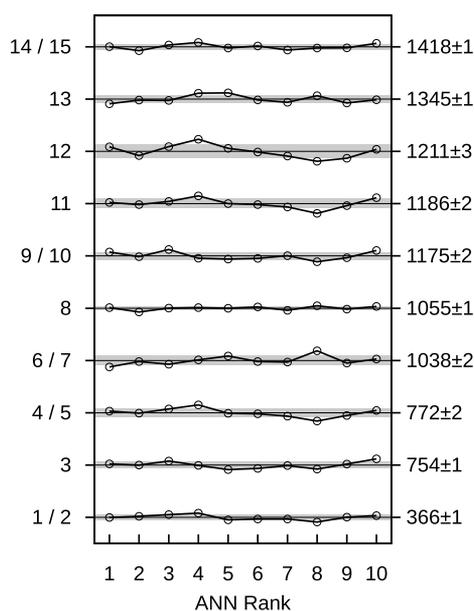
**4.1. Fitting Performance.** The first and somewhat surprising observation is that the new ANN setup outperforms the previous efforts significantly in terms of compactness and overall fitting results. Various neural network sizes between 5 and 120 hidden-layer neurons have been tested. The convergence behavior of the best obtained RMS error (as well as the corresponding validation error) is presented in Figure 2, together with the analogous graph from ref 90. Not only does the new scheme converge to roughly half of the previous fitting error, convergence itself occurs significantly faster with respect to the ANN size. As in previous work, a sudden increase of the difference between fitting error and validation error is observed at a certain number of hidden-layer neurons. This is a sign for overfitting, and ANN models with a larger hidden layer are considered less reliable. This point is reached for 50 hidden-layer neurons with the new approach compared to 80 neurons for the initial method. In the present study we will focus on the best fitted 30 neuron network, with a fitting RMS of  $21.4 \text{ cm}^{-1}$  and a validation error of  $23.6 \text{ cm}^{-1}$ , respectively. This was chosen instead of the 40 neuron network, because both result in almost indistinguishable results, and the smaller ANN is more efficient to test thoroughly. It also should be pointed out that the reference data are obtained from very intricate ab initio calculations rather than an analytic test model, and thus the remaining errors not only reflect possible shortcomings of the present model but also the errors intrinsic to the multireference configuration interaction (MRCI) calculations. We are inclined to estimate that an RMSE on the order of  $10\text{--}20 \text{ cm}^{-1}$  might be the lower limit of what is possible to achieve with the present data set.

While the total root-mean-square error provides a measure for how well the given ab initio data are reproduced overall, it provides no information regarding how deviations are distributed. For this purpose, unweighted RMS errors were computed for each adiabatic energy separately below a certain threshold. This threshold was chosen the same way it was in previous work,<sup>90</sup> 1 eV above the reference point energies, to obtain a measure of how well the model performs in the

physically relevant regions of the individual sheets. A comparison between the convergence behavior of both present and previous models can be found in Figure 3. Apart from inheriting the same smoother, more well-behaved manner of convergence, state-specific errors produced by the new method are significantly closer together, meaning the overall error is far more evenly distributed across the individual electronic states. This is in stark contrast to the previous model, where the error seems largely dominated by the error of the lower  ${}^2E''$  sheet. A reasonable explanation for this is that, while its intricate triple-well structure remains the topographically most demanding to reproduce, the inherent symmetry of the modified input coordinates greatly reduces the overall strain on the network. This also explains the significant reduction of required parameters (less than half of the previous reference ANN surface).

#### 4.2. Impact of the Hidden-Layer Size on Dynamics.

While representing ab initio data reliably is a central aspect and strong point of the presented method, it is equally essential for the resulting models to behave consistently in dynamics calculations. To this end, in a first step, fits of neural networks are considered, which differ only in which particular (randomly generated) initial guess was used in the training for the neural network parameters. As mentioned above, these tests are performed with 30 as the number of hidden-layer neurons for this particular study. The resulting surfaces can vary significantly, as the highly nonlinear nature of the model allows for a great deal of different realizations of nearly identical fitting errors. The study itself is performed in analogy to that of ref 95, taking the 10 best-fitted networks (in terms of the RMS error) from a set of 100. These networks only differ marginally in terms of the fitting error, ranging from  $21.4$  to  $22.2 \text{ cm}^{-1}$ , meaning they can be considered roughly identical in terms of quality. They are enumerated by their numerical “rank” (starting at 1 for the best network) for the sake of convenience in the following. Figure 4 shows an excerpt of the data resulting from computing vibrational transition energies for the  $\tilde{X} \text{ } ^2A_2'$  state, together with the obtained average transition energies as well as respective standard deviations  $\sigma_i$ . The latter are of particular interest, as they provide a simple measure for how dynamics results “spread” among samples of very similar networks. In other words, these  $\sigma_i$  values are a measure for what kind of



**Figure 4.** Comparison of the computed transition energies for the  $\tilde{X}^2A_2'$  state of  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) for the 10 best (in terms of the fitting error) ANNs from a set of 100 fitted neural networks, including the reference ANN. Level distances, not to scale, correspond to a deviation of  $20 \text{ cm}^{-1}$  from average. Gray: Standard deviations  $\sigma_i$  around average energies. Average energies are provided on the right-hand side together with the deviation  $\sigma_i$  in inverse centimeters.

deviations are a priori indistinguishable from competing networks of the same architecture as our reference, which is why excitation energies will be denoted with a  $\pm\sigma_i$  when appropriate.

Overall, the dynamics results obtained from the new 30 neuron model closely resemble those from the previous 75 hidden-layer neuron model both with respect to the average energies produced as well as with respect to the corresponding standard deviations. Standard deviations are close to  $2 \text{ cm}^{-1}$  for all 18 states considered in the present work. This more than likely confirms the previous conclusions that the previous model already exhausted the limits of what accuracy can be achieved within the confines of the ab initio data presently available. This indicates that the new approach inherits one of the most favorable aspects of its predecessor, in particular, the consistency of the dynamics results produced and thus the overall tendency of the method to produce robust results of excellent quality.

With the standard deviations  $\sigma_i$  as a measure for significant deviations from the reference network at hand, the findings presented in Table 1 can be contextualized.

The 20 neuron network, with its significantly smaller number of formal parameters, results in vibrational level energies deviating significantly for all states known to be affected by insufficiently flexible ANN models in the past (in particular, 4/5, 11, and 12), suggesting that these states are particularly sensitive regardless of the specifics of the model used. States 9/10, 13, and 14/15 are additionally affected, further suggesting that this network size is insufficient to reproduce the lowest adiabatic sheet accurately. The 25 neuron network, while being more consistent with the reference network overall, still shows significant deviations for states 12 and 13 as well as 6/7, 16/17, and 18, in particular, which were previously indistinguishable from the reference. Most interestingly, the 40 neuron network (corresponding to an increase of formal parameters by  $\sim 30\%$ )

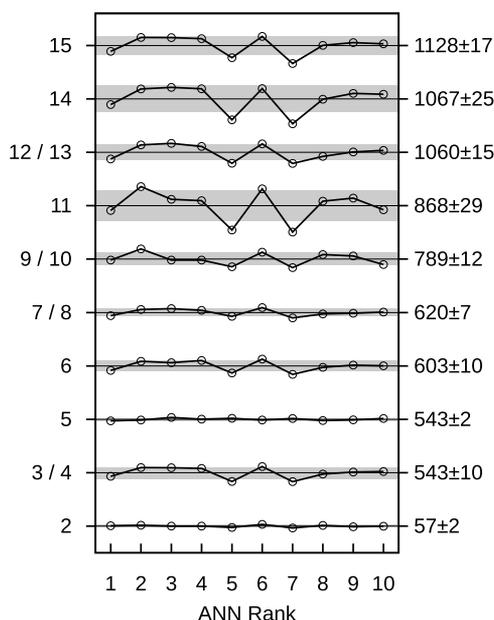
**Table 1.** Comparison of Computed Vibrational Transition Energies<sup>a</sup>

state	30	20	25	40	60	80
	$E_i - E_0 \pm \sigma_i$	$\delta(E_i - E_0)$				
1/2	$365.8 \pm 1.0$	-0.9	0.2	∴ 0.5	1.4	1.2
3	$754.6 \pm 1.3$	-1.5	-0.5	∴ -0.2	-0.2	0.9
4/5	$772.5 \pm 1.7$	-3.2	-0.8	∴ -0.5	1.2	1.5
6/7	$1035.5 \pm 1.7$	0.4	2.6	∴ 1.9	2.1	2.2
8	$1055.6 \pm 0.7$	-1.7	-1.0	∴ -1.2	-1.2	-1.1
9/10	$1176.6 \pm 1.5$	-2.8	-1.1	∴ -1.9	-1.8	-0.4
11	$1186.6 \pm 1.9$	-2.7	0.4	∴ -0.8	1.8	1.3
12	$1212.3 \pm 2.5$	-6.1	-3.1	∴ -2.3	0.6	1.7
13	$1342.7 \pm 1.5$	-2.3	2.5	∴ 0.5	1.2	1.9
14/15	$1417.6 \pm 1.0$	-2.5	-0.7	∴ -0.7	-0.2	0.0
16/17	$1488.1 \pm 0.8$	0.1	1.8	∴ -1.1	-0.1	0.0
18	$1489.8 \pm 2.1$	0.0	3.6	∴ 0.2	2.8	2.8

<sup>a</sup>Comparison of the computed vibrational transition energies for the  $\tilde{X}^2A_2'$  state of  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) for different hidden-layer sizes. The vibrational excitation energies of the reference (30) surface are reported together with the deviation  $\sigma_i$  as defined before (see text). For the five hidden-layer sizes tested,  $\delta(E_i - E_0)$  provides level shifts compared to reference network (30). The dotted vertical line indicates the separation between increased and decreased hidden-layer sizes with respect to the reference.

shows little to no significant change when compared to the reference network. This was the reason why the less flexible 30 neuron ANN was chosen as reference model, because it is the more rigorous test for the efficiency of the new model. It turns out that network sizes depending on more than twice the number of formal parameters than the reference network (such as 60 and 80 hidden-layer neurons) show very few significant deviations, with state 18 featuring the largest deviation overall. This is surprising, as overfitting effects were much more apparent in the previous model scheme, where the molecule's symmetry was not directly taken into account by construction. This change in behavior suggests that, at least for the sample of states considered, reproducing the adiabatic ground-state PES accurately is less indicative of the model's overall performance. Apparently, the new CNPI-ANN approach greatly reduces the strain on the network to reproduce the comparably simple, highly symmetric structure of this PES.

Moving on to the significantly more sensitive dynamics of the  $^2E''$  first excited state, again first the 10 best networks are considered for the reference hidden-layer size (30 neurons) to estimate the size of a "significant deviation". A visualization of the first 15 states is again provided in Figure 5. Despite the overall improvement of the new model in terms of reproducing ab initio data for this particular electronic state, no change in the overall sensitivity of the resulting transition energies could be observed. However, while the standard deviations  $\sigma_i$  remain rather similar to those produced by the analogous study using the previous model and 75 neuron network for each state considered, even the average energies of the first few states (such as 3/4, 6, and 8) deviate up to  $30 \text{ cm}^{-1}$ . On the one hand, the latter is to be expected, as the changes in fit performance were the most significant for the coupled surfaces of the  $^2E''$  state. The former, on the other hand, can easily be interpreted as the sensitivity of these states being intrinsic to the complexity of the particular system, combined with the limited availability of ab initio data. This interpretation is consistent with our findings in ref 95.



**Figure 5.** Comparison of the computed transition energies for  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) on the  $\tilde{A}$  state for the 10 best (in terms of fitting error) ANNs in analogy to Figure 4. Level distances, not to scale, correspond to a deviation of  $100 \text{ cm}^{-1}$  from average. Gray: Standard deviation  $\sigma_i$  around average energies. Excitation energies together with the deviation  $\sigma_i$  are given at the right-hand side.

The obtained standard deviations again provide a context in which the resulting vibronic excitation energies of different network sizes, summarized in Table 2, can be compared. In addition to the energies, this table also provides  $D_{3h}$  labels and state assignments based on symmetry information and projected wave function densities. As in the previous work,<sup>95</sup> state assignments in the degenerate stretch and bend coordinates are given in terms of radial  $r_{s,b}^n$  and tangential excitations  $t_{s,b}^n$ .<sup>110</sup> Excitations in the totally symmetric mode are denoted with  $a^n$ . While deviations are largely well below  $40 \text{ cm}^{-1}$ , the 80 neuron

network showed such immense overfitting artifacts that comparison was not deemed feasible, which is why it was omitted. Overall, deviations observed are rather large for different neural network sizes, in particular, for states corresponding to radial stretch excitations  $r_s$ . The fact that the radial stretch excitations remain sensitive despite the new model representing the related ab initio data significantly better than before further confirms the previous conclusions that this may be a limitation of the underlying data itself being insufficiently sampled in the dissociative domain of the nuclear configuration space. Apart from that, the overall size of the deviations is consistent with findings from the previous modeling scheme.

On the one hand, considering again the smaller networks first, the 20 hidden-layer neuron network exhibits the most significant deviations for the  $r_b$  fundamental and the  $t_b^2 a_1''$  state (state 20), with most other states showing deviations close to their respective  $\sigma_i$ . The 25 neuron network, on the other hand, shows expectedly fewer deviations than the 20 neuron network, as the flexibility of the network increases. Both networks remain comparably more similar to the reference ANN overall, at least relative to the deviations found for the same ANN sizes and the corresponding dynamics results for the ground-state calculations.

Intriguingly, this pattern was also observed in the previous model. A plausible explanation for this is that the adiabatic ground state is more sensitive to underfitting, requiring a minimal flexibility to be reproduced though having a simple overall structure. Conversely, the  $\tilde{A}$  state appears to be more sensitive to overfitting, having the most complicated structure. This sensitivity toward overfitting can be observed when looking at the 60 neuron network. In this case, deviations are again larger, reaching up to  $40 \text{ cm}^{-1}$ , with state 20 again deviating the most, relative to the standard deviation. The 40 neuron network, while significantly deviating for a few states (state 20, in particular), is overall very similar to the reference network and can be seen as an overall equivalent choice.

In conclusion, our findings show that the new modeling scheme is overall very similar to its predecessor in terms of how

**Table 2. Comparison of Computed Vibronic Transition Energies<sup>a</sup>**

state	assignment	30 $E_i - E_0 \pm \sigma_i$	20 $\delta(E_i - E_0)$	25 $\delta(E_i - E_0)$	40 $\delta(E_i - E_0)$	60 $\delta(E_i - E_0)$
2	$0 a_1''$	$58.1 \pm 2.0$	2.2	-2.5	:	2.8
3/4	$r_s^1 e''$	$536.7 \pm 10.5$	12.3	15.3	:	13.1
5	$t_b^1 a_2''$	$539.5 \pm 2.1$	2.4	3.5	:	-1.6
6	$r_s^1 a_1''$	$595.0 \pm 9.9$	13.4	17.0	:	14.8
7/8	$t_b^1 e''$	$614.3 \pm 6.5$	11.6	7.3	:	10.0
9/10	$r_b^1 (*) e''$	$787.3 \pm 11.6$	19.5	-0.3	:	3.3
11	$r_b^1 (*) a_1''$	$859.4 \pm 28.7$	36.5	17.2	:	25.7
12	$r_s^1 t_b^1 a_2''$	$1047.7 \pm 14.7$	20.4	23.9	:	17.1
13/14	$r_s^2 e''$	$1056.4 \pm 24.8$	25.0	29.2	:	22.7
15	$r_s^2 a_1''$	$1117.3 \pm 17.4$	26.2	20.7	:	19.3
16/17	$t_b^2 (*) e''$	$1121.9 \pm 18.5$	29.6	30.1	:	28.9
18/19	$r_s^1 t_b^1 (*) e''$	$1168.8 \pm 5.9$	5.3	13.5	:	3.5
20	$t_b^2 (*) a_1''$	$1176.2 \pm 13.6$	25.0	19.9	:	20.2
21/22	$a^1 e''$	$1176.3 \pm 10.3$	18.2	16.3	:	12.9
23	$t_s^1 a_2''$	$1271.4 \pm 6.7$	5.7	8.3	:	-2.6

<sup>a</sup>Comparison of computed transition energies for  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) on the  $\tilde{A}$  state for different hidden-layer sizes.  $\delta(E_i - E_0)$  provides level shifts compared to the reference network (30). All states are given labels in terms of excitations in the breathing mode ( $a$ ) as well as radial and tangential excitations in the degenerate stretch and bend modes ( $r_{s,b}$ ,  $t_{s,b}$ ). Assignments marked with (\*) display visible contamination of other  $r_s$  excitation(s). The dotted vertical line indicates the separation between increased and decreased hidden-layer sizes with respect to the reference.

it reacts to varying the hidden-layer size and overfitting, while being capable of reproducing ab initio reference data both more accurately and with less than half as many parameters. The primary difference to previous findings (in terms of behavior) is that the model shows a more clear-cut change in behavior as overfitting effects become dominant. However, it is important to note that this does not relevantly affect the stability of the method itself, as these effects occur only for more than twice the number of parameters needed to accurately reproduce the given data, and resides far beyond any suitable model choice. As a consequence, the 30 neuron model remains the focal point of this work, with the 40 neuron model yielding largely equivalent data at the cost of needing significantly more parameters.

**4.3. Invariant Input Coordinates and Numerical Stability.** So far, only the model's behavior with respect to the hidden-layer size of the neural network has been investigated. However, this is of course not the only free parameter related to the neural network that could be adapted. Most of these parameters, such as the number of hidden layers or particularities regarding the distribution of the data set into the reference and validation set already have been discussed at length in ref 90. What remains is to analyze the impact of the new parameters introduced in Equation 16. In the present case, these parameters were set in such a way that  $c_k = d_k^{-1} = 10$ . What this means is that  $\nu_k'(Q)$  approximates  $\nu_k(Q)$  for all values of  $|\nu_k|$  less than  $c_k$  while approaching  $\pm c_k$  as  $\nu_k \rightarrow \pm\infty$ . In other words,  $\nu_k'$  serves as an asymptotically damped version of  $\nu_k$ .

One of the great advantages of neural networks, as we have stressed before, is the homogeneous, bounded nature of the sigmoid functions they are ultimately composed of. This is in stark contrast to the terms of a polynomial expansion, which increase more rapidly in size for increasing polynomial orders, which can cause artifacts caused by rapid oscillations in the undersampled domains of a fit. However, some of this disadvantage is reintroduced by the use of higher-order invariants as input coordinates, which means that small changes in the primitive coordinates translate into large changes in the third-order coordinates. This appears to cause unnecessary numerical noise to the neural network, thus hindering (or even impeding) convergence. The above damping scheme was introduced to remove this problem and make the resulting ANN model more robust numerically. What remains is to address the question of how to choose the resulting parameters  $c_k$ . Since the goal of the damping scheme is to avoid artifacts caused by ill-behaved coordinates rapidly increasing in size, the present scheme aims to minimize the risk of numerical issues by keeping  $\nu_k'$  as limited as possible. This means that the dynamically relevant region of the surface must be represented perfectly and that the damping does not interfere with that region.

Table 3 provides a summary of how the computed  $\tilde{A}$  state transitions change for the reference network size (30) for various choices of  $c_k$ . Again, labels and deviations wrt the reference network are provided in the same format as that of Table 2, including standard deviations  $\sigma_i$  as a measure for the expected deviations from the reference for equivalent networks (see Section 4.2). No significant impact on the  $\tilde{X}$  state could be observed. For this reason the present study will focus solely on the  $\tilde{A}$  state. Choosing a significantly smaller value of  $c_k$  than 10 causes the resulting network to become incapable of reproducing ab initio reference data, as small  $c_k$  begin to dampen  $\nu_k$  for smaller and smaller displacements from the reference geometry. On the one hand, this loss of quality is so

Table 3. Comparison of Computed Transition Energies<sup>a</sup>

state	assignment	$E_i - E_0 \pm \sigma_i$				
		$c_k = 10$	$c_k = 15$	$c_k = 20$	$c_k = 30$	$c_k = 40$
2	$0 a_1''$	$58.1 \pm 2.0$	-0.9	-1.9	1.0	1.5
3/4	$r_s^1 e''$	$536.7 \pm 10.5$	-0.7	8.1	7.6	16.2
5	$t_b^1 a_2''$	$539.5 \pm 2.1$	-1.4	4.5	2.8	4.6
6	$r_s^1 a_1''$	$595.0 \pm 9.9$	2.3	12.5	10.4	17.6
7/8	$t_b^1 e''$	$614.3 \pm 6.5$	-3.0	6.1	6.9	13.6
9/10	$r_b^1 (*) e''$	$787.3 \pm 11.6$	5.8	-6.0	-2.1	-2.3
11	$r_s^1 (*) a_1''$	$859.4 \pm 28.7$	1.4	2.0	7.3	17.7
12	$r_s^1 t_b^1 a_2''$	$1047.7 \pm 14.7$	-3.5	13.9	11.3	23.4
13/14	$r_s^2 e''$	$1056.4 \pm 24.8$	-0.8	22.8	15.1	29.5
15	$r_s^2 a_1''$	$1117.3 \pm 17.4$	1.9	19.2	14.2	25.6
16/17	$t_b^2 (*) e''$	$1121.9 \pm 18.5$	-0.6	26.3	15.8	35.7
18/19	$r_s^1 t_b^1 (*) e''$	$1168.8 \pm 5.9$	-1.8	7.6	2.3	8.8
20	$t_b^2 (*) a_1''$	$1176.2 \pm 13.6$	2.4	14.5	11.0	21.5
21/22	$a^1 e''$	$1176.3 \pm 10.3$	-3.1	19.6	12.8	27.7
23	$t_s^1 a_2''$	$1271.4 \pm 6.7$	1.6	6.1	-2.0	-0.2

<sup>a</sup>Comparison of the computed transition energies for  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) on the  $\tilde{A}$  state for different damping thresholds  $c_k = d_k^{-1}$  used in Equation 16. Columns for  $c_k$  values larger than 10 provide level shifts compared to reference network (30 hidden-layer neurons,  $c_k = 10$ ). Labels and standard deviations  $\sigma_i$  taken from Table 2.

significant that obtained dynamics results would fail to provide relevant insight, as the resulting dynamics change too drastically to allow a proper comparison. On the other hand, the behavior of the network changes only marginally when increasing  $c_k$  from 10 to 15. This suggests that  $c_k = 10$  is indeed large enough to cover the dynamically relevant region, while minimizing the risk of artifacts. Networks with thresholds of 50 or greater (essentially meaning that  $\nu_k' \rightarrow \nu_k$ ) have been found to cause significant artifacts in the surface, rendering further comparisons unfeasible. One way in which these artifacts become noticeable is that dynamics results of formally degenerate level energies begin to show deviations of  $\sim 1 \text{ cm}^{-1}$  due to numerical instability. As the damping threshold  $c_k$  becomes higher and higher, the computed level energies change increasingly, with deviations from the reference calculation values exceeding  $\sigma_i$ , until a "tipping point" (close to  $c_k = 50$ ) is reached, rendering the obtained dynamics calculations unusable.

In summary, tuning the damping threshold  $c_k$  one observes a significant change in the dynamics results obtained, with one limiting case on either side. Each limiting case is accompanied by an abrupt change into pathological behavior. Our interpretation of the data presented in Table 3 is that very large values of the  $\nu_k$  outside the dynamically relevant region cause the numerical artifacts and thus need to be damped as quickly as possible. This means that the proper values of  $c_k$  can be estimated approximately by the inspection of the reference data and should be chosen reasonably small. For this reason, all further results are obtained with a value of  $c_k = 10 \forall k$ .

**4.4. Comparison with Experimental and Previous Theoretical Data.** Up to this point it was investigated primarily how the ANN choices like hidden-layer size affects the fit performance as well as quantum dynamics result in comparison with a reference ANN model and how these observations fit into the conclusions drawn from previous efforts. The comparisons aim at estimating errors intrinsic to the model with respect to the differences in excitation energies produced by the ANNs of (in terms of fitting error) indistinguishable quality. These estimates

in return allow to judge different networks in terms of their internal consistency among one another, without the need for an external reference. In this section, the focus will be on the quality of the surfaces based on external consistency with experimental data, considering  $\tilde{X}$  and  $\tilde{A}$  state vibrational/vibronic transition energies.

**4.4.1. Vibrational Eigenstates in the  $\tilde{X}^2A_2'$  Electronic State.** The results of the vibrational eigenstates of the  $\tilde{X}$  state are given in Table 4. The experimental frequencies are compared to the

**Table 4. Summary of the Comparison<sup>a</sup>**

assignment	$E_{\text{exp}}$	$E_{\text{theo}}^{\text{ANN,old}}$	$E_{\text{theo}}^{\text{ANN,new}}$	$\delta E_{\text{exp}}^{\text{ANN,old}}$	$\delta E_{\text{exp}}^{\text{ANN,new}}$
$4^1 e'$	365.5	366.8	365.8	-1.3	-0.3
$4^2 a_1'$	752.4	756.3	754.6	-3.9	-2.2
$4^2 e'$	771.8	774.6	772.5	-2.8	-0.7
$1^1 a_1'$	1051.2	1054.9	1055.6	-3.7	-4.4
$3^1 e'$	1055.3	1039.7	1035.5	15.6	19.8
$4^3 e'$	1173.6	1179.3	1176.6	-5.7	-3.
$4^3 a_1'$	1214	1215.1	1212.3	-1.1	1.7
$1^1 4^1 e'$	1413.6	1417.7	1417.6	-4.1	-4.
$3^1 4^1 a_2'$	1491	1344.9	1342.7	146.1	148.3
$3^1 4^1 e'$	1492.4	1490.0	1488.1	2.4	4.3
$3^1 4^1 a_1'$	1499.8	1498.2	1489.8	1.6	10.

<sup>a</sup>Summary of the comparison between the experimental and computed transition energies on the  $\tilde{X}^2A_2'$  state of  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ). The "assignment" column provides the corresponding  $D_{3h}$  labels as well as state assignments including dominant state contributions.  $E_{\text{theo}}^{\text{ANN,old}}$  provides excitation energies using the previous (ANN-based) model,<sup>90,95</sup> and  $E_{\text{theo}}^{\text{ANN,new}}$  relates to the ones obtained for the present model. The remaining two columns provide respective differences from the experimental data as listed under  $E_{\text{exp}}$ , reported from ref 106.

computed transitions for both the novel and previous ANN model. As becomes apparent quickly from the fifth and sixth columns of Table 4, deviations from experiment remain largely similar, with all but one change in deviation being most likely just coincidental and due to the model-intrinsic fluctuations discussed in Section 4.2. The only state to which a noteworthy change can be attributed is the  $3^1 4^1 a_1'$  state. However, this deviation by itself does not allow for any meaningful interpretation and still resides well within the limits of what can be expected of the model under present conditions. The new model also reconfirms the long-debated  $3^1$  fundamental, which can be assigned confidently to a state computed at  $1036 \text{ cm}^{-1}$  (previously  $1040 \text{ cm}^{-1}$ ), agreeing well with other theoretical treatments<sup>102–104,116</sup> and recent experimental assignments.<sup>106,117,118</sup> Similarly, the new model also provides further evidence that the measured energy level of  $1492.4 \text{ cm}^{-1}$  does not correspond to the  $3^1$  fundamental but to the combination band  $3^1 4^1 e'$ . Finally, the large deviation found for the  $3^1 4^1 a_2'$  state can only be explained by a misassignment of the experiment as a disagreement of almost  $150 \text{ cm}^{-1}$ , for a single state appears more than unlikely, especially with excitations even higher in energy not reaching one-tenth of that.

**4.4.2. Vibronic Eigenstates in the  $\tilde{A}^2E''$  Electronic State.** The available data for the electronically excited  $^2E''$  state are restricted to three experimentally measured transitions, as was the case with the previous model. While a greater deal of experimentally observed transitions is indeed available, they all involve umbrella excitations, which cannot be reproduced with the present model being restricted to planar geometries. A comparison of how both old and new ANN models reproduce

available experimental data is provided in Table 5. The two adiabatic PES sheets corresponding to the  $^2E''$  state show a

**Table 5. Summary of the Comparison<sup>a</sup>**

assignment	$E_{\text{exp}}$	$E_{\text{theo}}^{\text{ANN,old}}$	$E_{\text{theo}}^{\text{ANN,new}}$	$\delta E_{\text{exp}}^{\text{ANN,old}}$	$\delta E_{\text{exp}}^{\text{ANN,new}}$
$4_0^1 (t_b^1)$	539.5	539.3	539.5	0.2	0.0
$4_0^2 (r_s^1 t_b^1)$	1056.6	1024.8	1047.7	31.8	8.9
$3_0^1 (t_s^1)$	1270.5	1276.1	1271.4	-5.6	-0.9

<sup>a</sup>Summary of the comparison between the experimental and computed transition energies on the  $\tilde{A}^2E''$  state of  $\text{NO}_3$  (in  $\text{cm}^{-1}$ ). Assignments in the first column correspond to those of Table VI of ref 107, with assignments from Table 2 given in parentheses.  $E_{\text{theo}}^{\text{ANN,old}}$  provides excitation energies using the previous (ANN-based) model,<sup>90,95</sup> and  $E_{\text{theo}}^{\text{ANN,new}}$  relates to the ones obtained for the present ANN-based model. The remaining two columns provide the respective differences from the experimental data.

particularly strong Jahn–Teller coupling resulting in a pronounced triple-well structure on the lower adiabatic sheet. As a result, the assignment of the  $D_{3h}$  labels based on normal mode excitations are not straightforward. The intricacies of how tunneling and the geometric phase effect affect the resulting spectra as well as the structure of the vibronic eigenstates have been analyzed in detail in refs 105 and 110.

Of the three experimentally known excitations, two have already been reproduced within the intrinsic limitations of the previous model. This continues to be the case for the present method, excellently reproducing experimental data for the  $4_0^1 (t_b^1)$  and the  $3_0^1 (t_s^1)$  level. In addition, the previously far less well reproduced  $r_s^1 t_b^1$  state now is computed with unprecedented agreement with experimental data by the new CNPI-ANN model. While it is impossible to extrapolate too far from such a small data set, it may be concluded that it is a fair assumption that this improvement is indicative of the overall improved quality of the new model, in particular, with respect to the  $\tilde{A}$  state.

## 5. CONCLUSIONS AND OUTLOOK

The recently developed diabaticization method based on a combination of a simple vibronic coupling model and an ANN that tunes the coupling parameters has been improved and thoroughly tested in the present study. The use of symmetry transformation invariants as input coordinates for the ANN intrinsically removes symmetry contaminations. Apparently, the proper symmetry handling by the invariants reduces the stress on the ANN and leads to reduced RMS errors of the represented ab initio data. Furthermore, these improved results are obtained with significantly smaller ANNs. While with the original method an optimal ANN with 75 hidden-layer neurons was determined, the present CNPI-ANN method needs only 30 neurons to yield a PES model with noticeably improved RMSE. Thus, the new method not only is capable of reaching better accuracy but also is considerably more efficient. This can reduce the numerical effort in demanding quantum dynamics calculations considerably and thus can be of tremendous advantage.

The performance of the diabatic PES model obtained with the CNPI-ANN method has been tested by representing an extensive set of ab initio data for the few lowest electronic states  $^2A_2'$ ,  $^2E''$ , and  $^2E'$  of planar  $\text{NO}_3$ . The same data set has been used as in our previous studies, and thus the results can be compared directly. In addition to the residual errors in the representation of the ab initio data, vibrational- and vibronic-

level energies have been computed. This allows to assess the sensitivity of the nuclear dynamics results on the various parametrizations of the CNPI-ANN PES model. It turns out that the new CNPI-ANN model behaves very similarly to the previous plain ANN model. In particular, the onset of overfitting can be assessed from the development of the validation error compared to the reference error when increasing the hidden-layer size. For each ANN, many different parametrizations are obtained, which are very similar in terms of RMSE. The corresponding quantum dynamics calculations yield a certain scattering of values for each vibrational/vibronic level that is indicative of the sensitivity of the results on the minute details of the PES model. The observed standard deviations are only on the order of a couple of inverse centimeters for the  ${}^2A_2'$  ground-state vibrational levels. The standard deviations computed for the vibronic levels of the excited  ${}^2E''$  state, having a much more complicated PES topography, are approximately an order of magnitude larger. Comparison of these dynamics results with the available experimental values shows overall excellent agreement, and most level energies are reproduced more or less within the computed standard deviations. The agreement with experiment is slightly improved over the plain ANN model, which already yielded excellent results. The most noticeable improvements are observed for the  ${}^2E''$  state both in fitting RMSE and quantum dynamics results.

The present model limited to 5D planar  $\text{NO}_3$  will be extended to full 6D in the near future to be able to investigate the nuclear quantum dynamics and spectroscopy of this intriguing radical in full detail. Of course, the CNPI-ANN diabaticization model can be applied also to other vibronic coupling problems. The present results seem quite encouraging that accurate and reliable diabatic PES models can be obtained by the present method in an efficient way.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.0c05991>.

Apart from the provided materials, two further tables are available. These provide a full comparison of computed transition energies for the 10 best (in terms of fitting error) ANNs for the reference network on both  $\bar{X}$  and  $A$  state. Figures 4 and 5 provide a visual representation of these tables for the first 15 states (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

David M. G. Williams – *Theoretische Chemie, Universität Bielefeld, D-33501 Bielefeld, Germany*; Email: [d.williams@uni-bielefeld.de](mailto:d.williams@uni-bielefeld.de)

Wolfgang Eisfeld – *Theoretische Chemie, Universität Bielefeld, D-33501 Bielefeld, Germany*;  [orcid.org/0000-0002-3691-730X](https://orcid.org/0000-0002-3691-730X); Email: [wolfgang.eisfeld@uni-bielefeld.de](mailto:wolfgang.eisfeld@uni-bielefeld.de)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpca.0c05991>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Part of this work was generously supported by the Deutsche Forschungsgemeinschaft.

## ■ REFERENCES

- (1) Collins, M. A.; Parsons, D. F. Implications of Rotation-Inversion-Permutation Invariance for Analytic Molecular-Potential Energy Surfaces. *J. Chem. Phys.* **1993**, *99*, 6756–6772.
- (2) Ischtwan, J.; Collins, M. A. Molecular-Potential Energy Surfaces by Interpolation. *J. Chem. Phys.* **1994**, *100*, 8080–8088.
- (3) Ho, T. S.; Rabitz, H. A General Method for Constructing Multidimensional Molecular Potential Energy Surfaces from Ab Initio Calculations. *J. Chem. Phys.* **1996**, *104*, 2584–2597.
- (4) Maisuradze, G. G.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. Interpolating Moving Least-Squares Methods for Fitting Potential Energy Surfaces: Detailed Analysis of One-Dimensional Applications. *J. Chem. Phys.* **2003**, *119*, 10002–10014.
- (5) Maisuradze, G. G.; Kawano, A.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. Interpolating Moving Least-Squares Methods for Fitting Potential Energy Surfaces: Analysis of an Application to a Six-Dimensional System. *J. Chem. Phys.* **2004**, *121*, 10329–10338.
- (6) Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.
- (7) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural-network models of potential-energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129–4137.
- (8) Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. Combining ab initio computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules. *J. Chem. Phys.* **1996**, *105*, 7597–7604.
- (9) Tai No, K.; Ha Chang, B.; Yeon Kim, S.; Shik Jhon, M.; Scheraga, H. A. Description of the potential energy surface of the water dimer with an artificial neural network. *Chem. Phys. Lett.* **1997**, *271*, 152–156.
- (10) Prudente, F. V.; Acioli, P. H.; Neto, J. J. S. The fitting of potential energy surfaces using neural networks: Application to the study of vibrational levels of  $\text{H}_3^+$ . *J. Chem. Phys.* **1998**, *109*, 8801–8808.
- (11) Lorenz, S.; Gross, A.; Scheffler, M. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. *Chem. Phys. Lett.* **2004**, *395*, 210–215.
- (12) Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. Ab initio potential-energy surfaces for complex, multichannel systems using modified novelty sampling and feedforward neural networks. *J. Chem. Phys.* **2005**, *122*, No. 084104.
- (13) Lorenz, S.; Scheffler, M.; Gross, A. Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 115431.
- (14) Manzhos, S.; Wang, X. G.; Dawes, R.; Carrington, T. A nested molecule-independent neural network approach for high-quality potential fits. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.
- (15) Manzhos, S.; Carrington, T., Jr. A random-sampling high dimensional model representation neural network for building potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, No. 084109.
- (16) Manzhos, S.; Carrington, T., Jr. Using neural networks to represent potential surfaces as sums of products. *J. Chem. Phys.* **2006**, *125*, 194105.
- (17) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (18) Manzhos, S.; Carrington, T., Jr. Using redundant coordinates to represent potential energy surfaces with lower-dimensional functions. *J. Chem. Phys.* **2007**, *127*, No. 014103.
- (19) Malshe, M.; Narulkar, R.; Raff, L. M.; Hagan, M.; Bukkapatnam, S.; Komanduri, R. Parametrization of analytic interatomic potential functions using neural networks. *J. Chem. Phys.* **2008**, *129*, No. 044111.
- (20) Manzhos, S.; Carrington, T., Jr. Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface. *J. Chem. Phys.* **2008**, *129*, 224104.
- (21) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, No. 074106.

- (22) Nguyen, H. T. T.; Le, H. M. Modified Feed-Forward Neural Network Structures and Combined-Function-Derivative Approximations Incorporating Exchange Symmetry for Potential Energy Surface Fitting. *J. Phys. Chem. A* **2012**, *116*, 4629–4638.
- (23) Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys.* **2013**, *139*, No. 054112.
- (24) Koch, W.; Zhang, D. H. Communication: Separable potential energy surfaces from multiplicative artificial neural networks. *J. Chem. Phys.* **2014**, *141*, No. 021101.
- (25) Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **1989**, *2*, 303–314.
- (26) Meyer, H. D.; Manthe, U.; Cederbaum, L. S. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Lett.* **1990**, *165*, 73.
- (27) Manthe, U.; Meyer, H. D.; Cederbaum, L. S. Wave-packet dynamics within the multiconfigurational Hartree framework: General aspects and application to NOCl. *J. Chem. Phys.* **1992**, *97*, 3199–3213.
- (28) Manthe, U. A time-dependent discrete variable representation for (multiconfiguration) Hartree methods. *J. Chem. Phys.* **1996**, *105*, 6989–6994.
- (29) Brown, A.; Pradhan, E. Fitting potential energy surfaces to sum-of-products form with neural networks using exponential neurons. *J. Theor. Comput. Chem.* **2017**, *16*, 1730001.
- (30) Majumder, M.; Hegger, S. E.; Dawes, R.; Manzhos, S.; Wang, X.-G.; Carrington, T., Jr; Li, J.; Guo, H. Explicitly correlated MRCl-F12 potential energy surfaces for methane fit with several permutation invariant schemes and full-dimensional vibrational calculations. *Mol. Phys.* **2015**, *113*, 1823–1833.
- (31) Jiang, B.; Li, J.; Guo, H. Potential energy surfaces from high fidelity fitting of ab initio points: the permutation invariant polynomial neural network approach. *Int. Rev. Phys. Chem.* **2016**, *35*, 479–506.
- (32) Pradhan, E.; Brown, A. A ground state potential energy surface for HONO based on a neural network with exponential fitting functions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 22272–22281.
- (33) Domcke, W.; Yarkony, D. R.; Köppel, H., Eds. *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*; World Scientific: Singapore, 2004.
- (34) Longuet-Higgins, H. C. Some Recent Developments in the Theory of Molecular Energy Levels. In *Advances in Spectroscopy*; Interscience Publishers, Ltd.: New York, 1961; Vol. 2, p 429.
- (35) Lichten, W. Resonant charge exchange in atomic collisions. *Phys. Rev.* **1963**, *131*, 229.
- (36) Lichten, W. Molecular wave functions and inelastic atomic collisions. *Phys. Rev.* **1967**, *164*, 131.
- (37) Smith, F. T. Diabatic and adiabatic representations for atomic collision problems. *Phys. Rev.* **1969**, *179*, 111.
- (38) Baer, M. Adiabatic and diabatic representations for atom-diatom collisions - Treatment of 3-dimensional case. *Chem. Phys.* **1976**, *15*, 49–57.
- (39) Werner, H.; Meyer, W. MCSCF study of the avoided curve crossing of the two lowest  $^1\Sigma^+$  states of LiF. *J. Chem. Phys.* **1981**, *74*, 5802–5807.
- (40) Mead, C. A.; Truhlar, D. G. Conditions for the Definition of a Strictly Diabatic Electronic Basis for Molecular Systems. *J. Chem. Phys.* **1982**, *77*, 6090–6098.
- (41) Mead, C. A. *J. Chem. Phys.* **1983**, *78*, 807.
- (42) Werner, H.-J.; Follmeg, B.; Alexander, M. H. Adiabatic and diabatic potential energy surfaces for collisions of CN with He. *J. Chem. Phys.* **1988**, *89*, 3139.
- (43) Pacher, T.; Cederbaum, L. S.; Köppel, H. Approximately Diabatic States from Block Diagonalization of the Electronic Hamiltonian. *J. Chem. Phys.* **1988**, *89*, 7367–7381.
- (44) Pacher, T.; Mead, C. A.; Cederbaum, L. S.; Köppel, H. Gauge-Theory and Quasidiabatic States in Molecular Physics. *J. Chem. Phys.* **1989**, *91*, 7057–7062.
- (45) Pacher, T.; Köppel, H.; Cederbaum, L. S. Quasidiabatic States from Ab Initio Calculations by Block Diagonalization of the Electronic Hamiltonian - Use of Frozen Orbitals. *J. Chem. Phys.* **1991**, *95*, 6668–6680.
- (46) Pacher, T.; Cederbaum, L. S.; Köppel, H. *Adv. Chem. Phys.* **1993**, *84*, 293.
- (47) Xu, X.; Yang, K. R.; Truhlar, D. G. Diabatic Molecular Orbitals, Potential Energies, and Potential Energy Surface Couplings by the 4-fold Way for Photodissociation of Phenol. *J. Chem. Theory Comput.* **2013**, *9*, 3612–3625.
- (48) Shu, Y.; Kryven, J.; Sampaio de Oliveira-Filho, A. G.; Zhang, L.; Song, G.-L.; Li, S. L.; Meana-Pañeda, R.; Fu, B.; Bowman, J. M.; Truhlar, D. G. Direct diabaticization and analytic representation of coupled potential energy surfaces and couplings for the reactive quenching of the excited  $^2\Sigma^+$  state of OH by molecular hydrogen. *J. Chem. Phys.* **2019**, *151*, 104311.
- (49) Parker, K. A.; Truhlar, D. G. Semiglobal diabatic potential energy matrix for the N–H photodissociation of methylamine. *J. Chem. Phys.* **2020**, *152*, 244309.
- (50) Köppel, H.; Domcke, W.; Cederbaum, L. S. Multimode Molecular Dynamics Beyond the Born-Oppenheimer Approximation. *Adv. Chem. Phys.* **1984**, *57*, 59–246.
- (51) Cimiraaglia, R.; Malrieu, J. P.; Persico, M.; Spiegelmann, F. Quasi-Diabatic States and Dynamical Couplings from Ab Initio CI Calculations - A New Proposal. *J. Phys. B: At. Mol. Phys.* **1985**, *18*, 3073–3084.
- (52) Gadea, F. X.; Pelissier, M. *J. Chem. Phys.* **1990**, *93*, 545.
- (53) Domcke, W.; Woywod, C. Diabatic CASSCF Orbitals and Wave-Functions. *Chem. Phys. Lett.* **1993**, *216*, 362–368.
- (54) Yarkony, D. *R. Rev. Mod. Phys.* **1996**, *68*, 985.
- (55) Atchity, G. J.; Ruedenberg, K. Determination of diabatic states through enforcement of configurational uniformity. *Theor. Chem. Acc.* **1997**, *97*, 47–58.
- (56) Thiel, A.; Köppel, H. Proposal and numerical test of a simple diabaticization scheme. *J. Chem. Phys.* **1999**, *110*, 9371–9383.
- (57) Cattaneo, P.; Persico, M. Diabatic and adiabatic potential-energy surfaces for azomethane photochemistry. *Theor. Chem. Acc.* **2000**, *103*, 390–398.
- (58) Köppel, H.; Gronki, J.; Mahapatra, S. Construction scheme for regularized diabatic states. *J. Chem. Phys.* **2001**, *115*, 2377–2388.
- (59) Nakamura, H.; Truhlar, D. G. The direct calculation of diabatic states based on configurational uniformity. *J. Chem. Phys.* **2001**, *115*, 10353–10372.
- (60) Nakamura, H.; Truhlar, D. G. Direct diabaticization of electronic states by the fourfold way. II. Dynamical correlation and rearrangement processes. *J. Chem. Phys.* **2002**, *117*, 5576–5593.
- (61) Abrol, R.; Kuppermann, A. An optimal adiabatic-to-diabatic transformation of the  $1^2A'$  and  $2^2A'$  states of  $H_3$ . *J. Chem. Phys.* **2002**, *116*, 1035–1062.
- (62) Nakamura, H.; Truhlar, D. G. Extension of the fourfold way for calculation of global diabatic potential energy surfaces of complex, multiarrangement, non-Born-Oppenheimer systems: Application to HNC(O)(S-0,S-1). *J. Chem. Phys.* **2003**, *118*, 6816–6829.
- (63) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Conical intersections and their impact on molecular dynamics. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127.
- (64) Viel, A.; Eisfeld, W. Effect of Higher-Order Jahn-Teller Coupling on the Nuclear Dynamics. *J. Chem. Phys.* **2004**, *120*, 4603–13.
- (65) Eisfeld, W.; Viel, A. Higher Order  $(A + E) \otimes e$  Pseudo Jahn-Teller Coupling. *J. Chem. Phys.* **2005**, *122*, 204317.
- (66) Viel, A.; Eisfeld, W.; Neumann, S.; Domcke, W.; Manthe, U. Photoionization-Induced Dynamics of the Ammonia Cation: Ab Initio Potential-Energy Surfaces and Time-Dependent Wave-Packet Calculations. *J. Chem. Phys.* **2006**, *124*, 214306.
- (67) Viegas, L. P.; Alijah, A.; Varandas, A. J. C. Accurate ab initio based multisheeted double many-body expansion potential energy surface for the three lowest electronic singlet states of  $H_3^+$ . *J. Chem. Phys.* **2007**, *126*, No. 074309.
- (68) Viel, A.; Eisfeld, W.; Evenhuis, C. R.; Manthe, U. Photoionization-Induced Dynamics of the Ammonia Cation Studied by Wave

Packet Calculations Using Curvilinear coordinates. *Chem. Phys.* **2008**, *347*, 331–339.

(69) Faraji, S.; Köppel, H.; Eisfeld, W.; Mahapatra, S. Towards a Higher Order Description of Jahn-Teller Coupling Effects in Molecular Spectroscopy: The  $\tilde{A}^2E''$  State of  $\text{NO}_3$ . *Chem. Phys.* **2008**, *347*, 110–119.

(70) Papas, B. N.; Schuurman, M. S.; Yarkony, D. R. Determining quasideiabatic coupled electronic state Hamiltonians using derivative couplings: A normal equations based method. *J. Chem. Phys.* **2008**, *129*, 124104.

(71) Subotnik, J. E.; Yeganeh, S.; Cave, R. J.; Ratner, M. A. Constructing diabatic states from adiabatic states: Extending generalized Mulliken-Hush to multiple charge centers with Boys localization. *J. Chem. Phys.* **2008**, *129*, 244101.

(72) Zhu, X.; Yarkony, D. R. On the determination of optimized, fully quadratic, coupled state quasideiabatic Hamiltonians for determining bound state vibronic spectra. *J. Chem. Phys.* **2009**, *130*, 234108.

(73) Opalka, D.; Domcke, W. High-order expansion of  $T(2) \times t(2)$  Jahn-Teller potential-energy surfaces in tetrahedral molecules. *J. Chem. Phys.* **2010**, *132*, 154108.

(74) Eisfeld, W.; Vieuxmaire, O.; Viel, A. Full-dimensional diabatic potential energy surfaces including dissociation: The  $^2E$  state of  $\text{NO}_3$ . *J. Chem. Phys.* **2014**, *140*, 224109.

(75) Venghaus, F.; Eisfeld, W. Block-diagonalization as tool for the robust diabaticization of high-dimensional potential energy surfaces. *J. Chem. Phys.* **2016**, *144*, 114110.

(76) Wittenbrink, N.; Venghaus, F.; Williams, D.; Eisfeld, W. A new approach for the development of diabatic potential energy surfaces: Hybrid block-diagonalization and diabaticization by ansatz. *J. Chem. Phys.* **2016**, *145*, 184108.

(77) Schuurman, M. S.; Yarkony, D. R. On the vibronic coupling approximation: A generally applicable approach for determining fully quadratic quasideiabatic coupled electronic state Hamiltonians. *J. Chem. Phys.* **2007**, *127*, No. 094104.

(78) Zhu, X.; Yarkony, D. R. Toward eliminating the electronic structure bottleneck in nonadiabatic dynamics on the fly: An algorithm to fit nonlocal, quasideiabatic, coupled electronic state Hamiltonians based on ab initio electronic structure data. *J. Chem. Phys.* **2010**, *132*, 104101.

(79) Evenhuis, C. R.; Collins, M. A. Interpolation of Diabatic Potential Energy Surfaces. *J. Chem. Phys.* **2004**, *121*, 2515–2527.

(80) Evenhuis, C. R.; Lin, X.; Zhang, D. H.; Yarkony, D.; Collins, M. A. Interpolation of Diabatic Potential-Energy Surfaces: Quantum Dynamics on Ab Initio Surfaces. *J. Chem. Phys.* **2005**, *123*, 134110.

(81) Godsi, O.; Evenhuis, C. R.; Collins, M. A. Interpolation of Multidimensional Diabatic Potential Energy Matrices. *J. Chem. Phys.* **2006**, *125*, 104105.

(82) Nangia, S.; Truhlar, D. G. Direct calculation of coupled diabatic potential-energy surfaces for ammonia and mapping of a four-dimensional conical intersection seam. *J. Chem. Phys.* **2006**, *124*, 124309.

(83) Li, Z. H.; Valero, R.; Truhlar, D. G. Improved direct diabaticization and coupled potential energy surfaces for the photodissociation of ammonia. *Theor. Chem. Acc.* **2007**, *118*, 9–24.

(84) Zhu, X.; Ma, J. Y.; Yarkony, D. R.; Guo, H. Computational Determination of the  $\tilde{A}$  State Absorption Spectrum of  $\text{NH}_3$  and of  $\text{ND}_3$  Using a New Quasi-Diabatic Representation of the X and A States and Full Six-Dimensional Quantum Dynamics. *J. Chem. Phys.* **2012**, *136*, 234301.

(85) Xie, C.; Ma, J.; Zhu, X.; Zhang, D. H.; Yarkony, D. R.; Xie, D.; Guo, H. Full-Dimensional Quantum State-to-State Nonadiabatic Dynamics for Photodissociation of Ammonia in its A-Band. *J. Phys. Chem. Lett.* **2014**, *5*, 1055–1060.

(86) Zhu, X.; Yarkony, D. R. Fitting coupled potential energy surfaces for large systems: Method and construction of a 3-state representation for phenol photodissociation in the full 33 internal degrees of freedom using multireference configuration interaction determined data. *J. Chem. Phys.* **2014**, *140*, No. 024112.

(87) Guan, Y.; Fu, B.; Zhang, D. H. Construction of diabatic energy surfaces for LiFH with artificial neural networks. *J. Chem. Phys.* **2017**, *147*, 224307.

(88) Lenzen, T.; Manthe, U. Neural network based coupled diabatic potential energy surfaces for reactive scattering. *J. Chem. Phys.* **2017**, *147*, No. 084105.

(89) Xie, C.; Zhu, X.; Yarkony, D. R.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. IV. Coupled diabatic potential energy matrices. *J. Chem. Phys.* **2018**, *149*, 144107.

(90) Williams, D. M. G.; Eisfeld, W. Neural network diabaticization: A new ansatz for accurate high-dimensional coupled potential energy surfaces. *J. Chem. Phys.* **2018**, *149*, 204106.

(91) Lenzen, T.; Eisfeld, W.; Manthe, U. Vibronically and spin-orbit coupled diabatic potentials for  $X(^2P) + \text{CH}_4 \rightarrow \text{HX} + \text{CH}_3$  reactions: Neural network potentials for  $X = \text{Cl}$ . *J. Chem. Phys.* **2019**, *150*, 244115.

(92) Guan, Y.; Guo, H.; Yarkony, D. R. Neural network based quasideiabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections. *J. Chem. Phys.* **2019**, *150*, 214101.

(93) Guan, Y.; Zhang, D. H.; Guo, H.; Yarkony, D. R. Representation of coupled adiabatic potential energy surfaces using neural network based quasi-diabatic Hamiltonians: 1,2 ( $^2A'$ ) states of LiFH. *Phys. Chem. Chem. Phys.* **2019**, *21*, 14205–14213.

(94) Yin, Z.; Guan, Y.; Fu, B.; Zhang, D. H. Two-state diabatic potential energy surfaces of  $\text{ClH}_2$  based on nonadiabatic couplings with neural networks. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20372–20383.

(95) Williams, D. M. G.; Viel, A.; Eisfeld, W. Diabatic neural network potentials for accurate vibronic quantum dynamics-The test case of planar  $\text{NO}_3$ . *J. Chem. Phys.* **2019**, *151*, 164118.

(96) Guan, Y.; Yarkony, D. R. Accurate Neural Network Representation of the Ab Initio Determined Spin-Orbit Interaction in the Diabatic Representation Including the Effects of Conical Intersections. *J. Phys. Chem. Lett.* **2020**, *11*, 1848–1858.

(97) Guan, Y.; Guo, H.; Yarkony, D. R. Extending the Representation of Multistate Coupled Potential Energy Surfaces To Include Properties Operators Using Neural Networks: Application to the 1,2( $^1A$ ) States of Ammonia. *J. Chem. Theory Comput.* **2020**, *16*, 302–313.

(98) Viel, A.; Eisfeld, W.  $\text{NO}_3$  full-dimensional potential energy surfaces and ground state vibrational levels revisited. *Chem. Phys.* **2018**, *509*, 81–90.

(99) Mayer, M.; Cederbaum, L. S.; Köppel, H. Ground state dynamics of  $\text{NO}_3$ : Multimode vibronic borrowing including thermal effects. *J. Chem. Phys.* **1994**, *100*, 899.

(100) Okumura, M.; Stanton, J.; Deev, A.; Sommar, J. New insights into the Jahn-Teller effect in  $\text{NO}_3$  via the dark  $\tilde{A}^2E''$  state. *Phys. Rev. Lett.* **2006**, *73*, C64–C70.

(101) Mahapatra, S.; Eisfeld, W.; Köppel, H. Effects of multimode Jahn-Teller coupling on the photodetachment spectrum of nitrate anion ( $\text{NO}_3^-$ ). *Chem. Phys. Lett.* **2007**, *441*, 7–15.

(102) Stanton, J. F. On the vibronic level structure in the  $\text{NO}_3$  radical: II. Adiabatic calculation of the infrared spectrum. *Mol. Phys.* **2009**, *107*, 1059–1075.

(103) Simmons, C. S.; Ichino, T.; Stanton, J. F. The  $v(3)$  Fundamental in  $\text{NO}_3$  Has Been Seen Near  $1060 \text{ cm}^{-1}$ , Albeit Some Time Ago. *J. Phys. Chem. Lett.* **2012**, *3*, 1946–1950.

(104) Homayoon, Z.; Bowman, J. M. Communication: MULTI-MODE calculations of low-lying vibrational states of  $\text{NO}_3$  using an adiabatic potential energy surface. *J. Chem. Phys.* **2014**, *141*, 161104.

(105) Eisfeld, W.; Viel, A. Vibronic eigenstates and the geometric phase effect in the  $^2E''$  state of  $\text{NO}_3$ . *J. Chem. Phys.* **2017**, *146*, No. 034303.

(106) Kawaguchi, K.; Narahara, T.; Fujimori, R.; Tang, J.; Ishiwata, T. Infrared spectroscopy of  $2v(4)$  and  $v(3)+2v(4)$  bands of the  $\text{NO}_3$  radical. *J. Mol. Spectrosc.* **2017**, *334*, 10–21.

(107) Codd, T.; Chen, M.-W.; Roudjane, M.; Stanton, J. F.; Miller, T. A. Jet cooled cavity ringdown spectroscopy of the  $\tilde{A}^2E'' \leftarrow X^2A_2'$  transition of the  $\text{NO}_3$  radical. *J. Chem. Phys.* **2015**, *142*, 184305.

(108) Lenzen, T.; Manthe, U. Vibronically and spin-orbit coupled diabatic potentials for  $X(P) + \text{CH}_4 \rightarrow \text{HX} + \text{CH}_3$  reactions: General

theory and application for  $X(P) = F(^2P)$ . *J. Chem. Phys.* **2019**, *150*, No. 064102.

(109) Zeng, T.; Seidu, I. Revisiting the  $(E + A) \otimes (e + a)$  problems of polyatomic systems with trigonal symmetry: general expansions of their vibronic Hamiltonians. *Phys. Chem. Chem. Phys.* **2017**, *19*, 11098–11110.

(110) Weike, T.; Williams, D. M. G.; Viel, A.; Eisfeld, W. Quantum dynamics and geometric phase in  $E \otimes e$  Jahn-Teller systems with general  $C_{nv}$  symmetry. *J. Chem. Phys.* **2019**, *151*, No. 074302.

(111) Eisfeld, W.; Morokuma, K. *Ab initio* investigation of the vertical and adiabatic excitation spectrum of  $\text{NO}_3$ . *J. Chem. Phys.* **2001**, *114*, 9430–9440.

(112) Eisfeld, W.; Morokuma, K. A detailed study on the symmetry breaking and its effect on the potential surface of  $\text{NO}_3$ . *J. Chem. Phys.* **2000**, *113*, 5587.

(113) Eisfeld, W. Highly accurate determination of the electron affinity of  $\text{SF}_6$  and analysis of structure and photodetachment spectrum of  $\text{SF}_6^-$ . *J. Chem. Phys.* **2011**, *134*, No. 054303.

(114) Hagan, M.; Demuth, H.; Beale, M.; De Jesús, O. *Neural Network Design*, 2nd ed.; Martin Hagan, 2014.

(115) Light, J. C.; Carrington, T. Discrete-variable representations and their utilization. *Adv. Chem. Phys.* **2007**, *114*, 263–310.

(116) Stanton, J. F. On the vibronic level structure in the  $\text{NO}_3$  radical. I. The ground electronic state. *J. Chem. Phys.* **2007**, *126*, 134309.

(117) Jacox, M. E.; Thompson, W. E. The infrared spectroscopy and photochemistry of  $\text{NO}_3$  trapped in solid neon. *J. Chem. Phys.* **2008**, *129*, 204306.

(118) Kawaguchi, K.; Fujimori, R.; Tang, J.; Ishiwata, T. On the vibrational assignment in the ground electronic state of  $\text{NO}_3$ . *J. Mol. Spectrosc.* **2015**, *314*, 73–78.

**Complete nuclear permutation inversion invariant  
artificial neural network (CNPI-ANN)  
diabatization for the accurate treatment of  
vibronic coupling problems**

David M. G. Williams\* and Wolfgang Eisfeld\*

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

E-mail: [d.williams@uni-bielefeld.de](mailto:d.williams@uni-bielefeld.de); [wolfgang.eisfeld@uni-bielefeld.de](mailto:wolfgang.eisfeld@uni-bielefeld.de)

## Supporting Information

**Table S1:** Comparison of computed transition energies for  $\tilde{X}^2A_2'$  state of  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) for the 10 best (in terms of fitting error) ANNs from a set of 100 fitted neural networks, including the reference ANN. Standard deviation is provided w.r.t. *average* energy levels  $\mu_i$ .

	best	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>	$\mu_i$	$\sigma_i$
state	$E_i - E_0$	deviations from best network									mean	std. dev.
1/2	365.8	0.5	1.1	1.7	-0.8	-0.5	-0.5	-1.7	0.1	0.8	365.9	1.0
3	754.6	-0.4	1.1	-0.6	-2.1	-1.7	-0.6	-1.9	-0.1	2.0	754.2	1.3
4/5	772.5	-0.8	0.8	2.3	-0.8	-1.0	-2.0	-3.8	-1.7	0.2	771.9	1.7
6/7	1035.5	2.0	1.1	2.7	4.2	2.0	1.9	6.2	1.5	3.0	1038.0	1.7
8	1055.6	-1.7	-0.2	0.0	-0.2	0.3	-1.1	0.7	-0.6	0.4	1055.3	0.7
9/10	1176.6	-1.8	1.0	-2.4	-2.7	-2.4	-1.4	-3.7	-2.2	0.6	1175.1	1.5
11	1186.6	-0.8	0.4	2.5	-0.4	-0.8	-1.7	-4.2	-1.2	1.8	1186.2	1.9
12	1212.3	-3.3	0.1	2.9	-0.6	-2.0	-3.5	-5.5	-4.3	-1.0	1210.6	2.5
13	1342.7	1.4	1.3	4.1	4.2	1.5	0.6	3.1	0.3	1.5	1344.5	1.5
14/15	1417.6	-1.6	0.6	1.6	-0.5	0.3	-1.3	-0.5	-0.5	1.3	1417.6	1.0
16/17	1488.1	0.4	0.4	0.9	1.2	0.3	0.1	1.5	-0.4	2.2	1488.8	0.8
18	1489.8	0.7	-0.2	4.0	6.3	1.8	1.3	3.7	0.2	2.5	1491.9	2.1

**Table S2:** Comparison of computed transition energies for  $^{14}\text{NO}_3$  (in  $\text{cm}^{-1}$ ) on the  $\tilde{A}$  state for the 10 best (in terms of fitting error) ANNs in analogy to Table S1. All states are given labels in terms of excitations in the breathing mode ( $a$ ) as well as radial and tangential excitations in the degenerate stretch and bend modes ( $r_{s,b}$ ,  $t_{s,b}$ ). Assignments marked with (\*) display visible contamination of other  $r_s$  excitation(s).

		best	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>	$\mu_i$	$\sigma_i$	
state	assignment	$E_i - E_0$	deviations from best network										mean	std. dev.
2	0 $a_1''$	58.1	1.0	-0.6	-0.6	-3.2	2.6	-4.2	0.8	-2.0	-0.9	57.4	2.0	
3 / 4	$r_s^1 e''$	536.7	16.4	16.3	14.9	-9.5	18.5	-9.9	4.3	8.3	9.0	543.5	10.5	
5	$t_b^1 a_2''$	539.5	1.6	6.7	3.3	4.9	1.6	4.7	0.7	2.2	4.6	542.5	2.1	
6	$r_s^1 a_1''$	595.0	17.0	14.4	18.6	-4.8	20.9	-7.6	5.6	9.8	8.7	603.2	9.9	
7 / 8	$t_b^1 e''$	614.3	11.6	13.1	9.9	-1.6	14.9	-4.2	3.2	4.3	6.5	620.1	6.5	
9 / 10	$r_b^1 (*) e''$	787.3	20.8	-0.1	0.1	-12.6	14.7	-14.2	10.3	7.7	-8.4	789.1	11.6	
11	$r_b^1 (*) a_1''$	859.4	44.8	20.9	18.4	-36.4	40.7	-40.3	17.4	23.4	1.3	868.4	28.7	
12	$r_s^1 t_b^1 a_2''$	1047.7	26.3	29.2	23.5	-8.2	28.2	-8.5	4.6	13.1	16.1	1060.2	14.7	
13 / 14	$r_s^2 e''$	1056.4	29.1	32.0	29.7	-28.7	29.8	-36.0	10.0	21.0	19.2	1067.1	24.8	
15	$r_s^2 a_1''$	1117.3	26.0	25.5	23.6	-12.0	27.8	-23.0	11.2	16.2	14.2	1128.3	17.4	
16 / 17	$t_b^2 (*) e''$	1121.9	34.5	32.6	32.8	-8.1	38.5	-11.6	10.4	22.3	18.1	1138.8	18.5	
18 / 19	$r_s^1 t_b^1 (*) e''$	1168.8	8.9	9.8	12.0	-0.3	12.8	-4.4	0.1	5.7	6.6	1173.9	5.9	
20	$t_b^2 (*) a_1''$	1176.2	26.2	26.8	26.4	-2.1	30.6	-6.4	6.5	14.0	13.9	1189.9	13.6	
21 / 22	$a^1 e''$	1176.3	18.4	22.9	20.4	-2.0	23.2	-2.2	3.4	9.0	11.7	1186.8	10.3	
23	$t_s^1 a_2''$	1271.4	5.7	7.0	3.8	-10.0	3.0	-12.7	1.4	6.0	3.2	1272.2	6.7	