

# Constructing feed-forward artificial neural networks to fit potential energy surfaces for molecular simulation of high-temperature gas flows

Paolo Valentini,<sup>\*,†</sup> Maninder S. Grover,<sup>†</sup> and Eswar Josyula<sup>‡</sup>*Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433, USA*

(Received 4 August 2020; accepted 8 October 2020; published 9 November 2020)

Kinetic rates for thermochemical nonequilibrium models are generally computed from quasiclassical trajectory (QCT) calculations on accurate *ab initio* potential energy surfaces (PES). In this article, we use a feed-forward artificial neural network (ANN) to fit existing single-point energies for  $N_2 + N_2$  interactions [Bender *et al.*, *J. Chem. Phys.* **143**, 054304 (2015)] to construct a PES suitable for molecular simulation of high-temperature gas flows. We then perform detailed comparisons with a widely used  $N_4$  PES that was built using the permutation invariant polynomials (PIP) method. Specific physical considerations in the construction of the ANN for this application are detailed. Translation, rotation, and permutation invariance are precisely satisfied by mapping the interatomic distances onto a set of permutation invariant inputs, known as fundamental invariants (FI) that generate the permutation invariant polynomial ring. The diatomic energy is imposed by decomposing the total potential energy into a sum of a two-body and a many-body energy contribution. To obtain the correct dynamical behavior with the most basic, yet computationally efficient ANN, spurious long-distance interactions must be removed to avoid incorrect physical behavior at the dissociation threshold. We use a simple apodization function to smoothly taper off to zero any residual many-body interaction at large separations. Both accuracy and performance of the FI-ANN PES are assessed. QCT calculations are used to compute dissociation probabilities and vibrational energy distributions at various equilibrium temperatures. Excellent agreement with the results obtained from the PIP PES is found. For our test case, the ANN PES is also significantly more computationally efficient than the PIP PES at comparable root-mean-square error levels.

DOI: [10.1103/PhysRevE.102.053302](https://doi.org/10.1103/PhysRevE.102.053302)

## I. INTRODUCTION

The accurate description of internal energy transfer mechanisms and dissociation kinetics in high-speed flows is important to accurately predict the surface heat flux on hypersonic vehicles. At Mach numbers above 5, shock layer temperatures reach tens of thousands of degrees K. Molecular collisions in this environment are extremely energetic and lead to excitation of internal energy modes, dissociation, and radiation. When flow and chemical characteristic time scales become comparable, some regions of the flow field may be in thermochemical nonequilibrium.

Current thermochemical models used in computational fluid dynamics (CFD) are often based on equilibrium distributions of molecular internal energies and are calibrated with experimental data, which are typically characterized by a large uncertainty. Often flight conditions cannot be reproduced with ground testing, and the extrapolation of low temperature laboratory measurements to high temperatures typical of hypersonic conditions is problematic, due to strong nonlinearities that characterize these phenomena. For these reasons, these models may fail altogether in predicting some key physics.

Higher fidelity is obtained from state-to-state kinetic models [1–10]. This approach consists in precomputing an *ab initio* potential energy surface (PES) by interpolating an extensive set of single-point energies that best describe the interaction between either two molecules or an atom and a molecule. Then, quasiclassical (or semiclassical [4]) trajectory calculations (QCT) are used to obtain reaction probabilities, collision cross sections, and transition rates [11]. State-resolved approaches are very powerful and, in principle, free of any empiricism.

An alternative to state-to-state kinetic models is the direct molecular simulation (DMS) method. Similar to the direct simulation Monte Carlo (DSMC) method of Bird [12], particle free flight and particle-particle collisions are decoupled, thus achieving an enormous computational gain in a dilute gas simulation compared to brute force molecular dynamics. However, unlike standard DSMC, DMS does not rely on simplified models for the collisions or precomputed collision cross-sections; instead, a full trajectory calculation is performed for each colliding pair of molecules to determine the post-collision states [13].

Both QCT and DMS calculations generally integrate an extremely large number of molecular trajectories (up to billions, depending on temperature) on a PES with femtosecond-sized time steps. The computational cost of molecular-dynamics-like methods [14] is directly proportional to the cost of evaluating the gradient of the potential energy. Trajectories are then obtained by a numerical integration of Newton's second law, where the negative of the gradient provides the interatomic forces. Depending on the temporal accuracy of

<sup>\*</sup>Aerothermodynamic Research Engineer, University of Dayton Research Institute, Dayton, OH, USA; corresponding author: [paolo.valentini@udri.udayton.edu](mailto:paolo.valentini@udri.udayton.edu)

<sup>†</sup>Chemical Kinetics Engineer, University of Dayton Research Institute, Dayton, OH, USA.

<sup>‡</sup>Senior Research Aerospace Engineer, AFRL/RQHF.

the numerical integration method (e.g., Verlet, Runge-Kutta, etc. [14]), multiple force evaluations are needed per time step. Such expensive computations can only be justified if the trajectories are obtained on an accurate PES to produce benchmark results for validation of reduced-order, simplified descriptions. Clearly, the DMS method could be more easily extended to increasingly realistic flows beyond simple zero-dimensional relaxations [13] with more computationally efficient algorithms to evaluate interatomic forces (i.e., PES gradient) needed in trajectory integration.

In general, it is desirable that the *accuracy* of the PES should be primarily determined by the quality of the underlying quantum mechanical calculations, and to a lesser degree by the interpolation methodology. Different *ab initio* data sets for the same system are often obtained with different quantum mechanical methods. For example, the  $N_2 + N_2$  PES has been independently obtained by researchers at NASA Ames [15], the University of Minnesota [16], and the University of Perugia [17]. But “each research group has their own recipe for devising the geometric grid, computing the electronic energy [...], and defining the analytic expression used to represent the PES.” Furthermore, “the process of determining the functional form and values for the parameters that reproduce these energies is quite tedious and time-consuming [...]” [18]. Therefore, it is also desirable for the interpolation scheme to require as little human intervention as possible and to be computationally efficient to be repeated as needed if the training data set requires additional energies to improve the description of the system energetics. Ultimately, the main effort in the PES construction should be in the actual calculation of the single-point energies.

Quite recently, artificial neural networks (ANN) have emerged as a class of algorithms that have shown a number of advantages over conventional PES construction methods [19]. It has been shown formally [20] that feed-forward neural networks are universal approximators, i.e., they are capable of approximating any real-valued function from one finite dimensional space to another to any desired degree of accuracy, provided a sufficient number of neurons in the hidden layer is used (see Sec. II for a brief description of the ANN architecture). Therefore, a feed-forward ANN is a natural choice for PES construction. In fact, artificial neural networks have been previously applied to several systems, from small molecules (e.g.,  $H_2O$ ) to condensed systems (Si or C) [19].

The objective of this work is to assess the physical accuracy of the simplest feed-forward ANN architecture to construct relatively low-dimensional PESs for application to molecular simulation of high-temperature flows, where strong coupling between internal energy excitation and chemical reactivity occurs. In doing so, we also evaluate the computational efficiency of an ANN PES obtained from fitting an existing  $N_2 + N_2$  *ab initio* single-point energy data set [16,21] compared to previous interpolations [21]. Recent works have extended the application of ANNs to more complex systems, e.g., condensed phases [22–24]. In these applications, the PES dimensionality is very large (hundreds or thousands of dimensions) and permutation invariance is obtained by utilizing descriptors of the local atomic environment. These strategies, although quite general, greatly increase the sophistication of the ANN, and therefore likely complicate its training and in-

crease the computational cost of evaluating energy and forces compared to the simplest feed-forward ANN architecture. Moreover, for air thermochemistry, molecular collisions only involve a small number of atoms, thus resulting in a low dimensional PES. For example, a three-dimensional PES describes  $N + N_2$  interactions, whereas a six-dimensional one is required for  $N_2 + N_2$  collisions. Hence, we think that the application of sophisticated, but more general strategies, is not necessary when the objective is to minimize computational cost without loss of accuracy compared to polynomial interpolators.

The advantages of using ANN for PES construction are well-known and, in summary, include the following:

(1) The accuracy of the fit is generally determined by the *architecture of the network*, i.e., number of hidden layers and corresponding constituent neurons. Unlike predetermined functional forms, the ANN architecture can be easily varied to obtain a desired accuracy. Furthermore, a change of architecture (e.g., number of neurons in a layer) does not require any modification to the algorithm to compute its gradient with respect to input variables.

(2) Neural networks are efficient at fitting large data sets. Generally, larger data sets require more neurons and hidden layers, but there is no need to modify functional forms or adding functional terms like in the standard approach described earlier.

(3) The training of neural networks does not require much human intervention.

Because of these reasons, neural networks are very suitable to construct a PES for the applications of interest here. For example, different *ab initio* data sets on the same system could be trained with the same network architecture to the same level of accuracy (e.g., same root-mean-square error, RMSE). This would enable comparisons between different ANN-based surfaces, thus potentially highlighting differences due to the underlying quantum mechanical data rather than the specific selected fitting function. During the preparation of this article, Li and co-workers [25] published a PIP neural network fit of an *extended*  $N_2 + N_2$  *ab initio* energy data set. Several similarities and differences between our approach and theirs will be briefly highlighted where appropriate.

The paper is organized as follows. In Sec. II, we briefly illustrate the mathematical form of feed-forward artificial neural networks, with a particular focus on physical requirements that must be satisfied for application to trajectory calculations, namely translation, rotation, and permutation invariance (Sec. II B) and long-range asymptotic behavior (Sec. II C). We select the  $N_2 + N_2$  system as a test case. The results are presented in Sec. III. First, we discuss the fitting performance of ANN in Sec. III A. Then, we analyze results from quasi-classical trajectories on the ANN PES and compare them to those obtained with a widely used PES for  $N_2 + N_2$  [26] in Sec. III B. We draw the conclusions in Sec. IV.

## II. MATHEMATICAL FORMULATIONS

### A. Feed-forward neural network: Mathematical formulation

The comprehensive mathematical treatment of fully connected feed-forward neural network is well-established (for

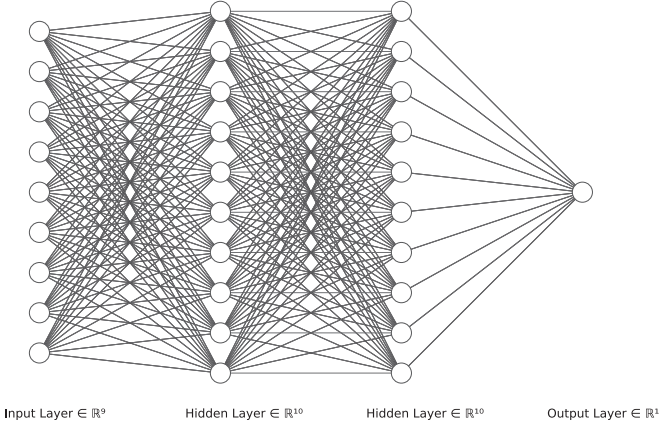


FIG. 1. Schematics of a fully connected feed-forward ANN with two hidden layers.

example, an introductory description is contained in the book of Goodfellow, Bengio, and Courville [27] among many other references). Here, we provide a short description of the basic ingredients.

A typical schematic of a fully connected feed-forward ANN is shown in Fig. 1. Neurons are organized in layers. For the input layer, the number of neurons equals the domain dimensionality of the function to approximate. The number of hidden layers is a measure of the network's *depth*. Each hidden layer comprises several neurons. The number of output neurons is determined by the dimensionality of the co-domain of the function to interpolate. In this work, there is one output, the potential energy of the system.

All nodes in each layer are connected to the nodes in the next layer (feed-forward ANN) by weights, the fitting parameters of the ANN. Furthermore, each node gets an activation function whose input is the biased weighted sum of the activation state of all the neurons in the previous layer and that are connected to that node. We denote  $\omega_{i,j}^{(k,l)}$  the weight connecting neuron  $i$  in layer  $k$  with neuron  $j$  in layer  $l = k + 1$ . Further, hidden and output nodes are connected to a *bias* node by the bias weight  $b_j^{(l)}$ . The input layer is denoted with  $k = 0$  and has no bias weights. Each neuron in the first hidden layer gets a biased weighted sum of the input coordinates  $G_i$  (shown later to be a function of the interatomic distances, see Sec. II B),

$$y_m^{(1)} = b_m^{(1)} + \sum_i^{N_0} \omega_{i,m}^{(0,1)} G_i, \quad (1)$$

where  $N_0$  represents the number of input neurons. Then, each neuron in the first hidden layer gets an *activation*  $h_m^{(1)}$ ,

$$h_m^{(1)} = f(y_m^{(1)}), \quad (2)$$

where  $f$  is the activation function. Similarly, a generic neuron  $m$  in the hidden layer  $n$  is activated to a state,

$$h_m^{(n)} = f(y_m^{(n)}) = f\left(b_m^{(n)} + \sum_i^{N_{n-1}} \omega_{i,m}^{(n-1,n)} h_i^{(n-1)}\right), \quad (3)$$

where  $N_{n-1}$  is the number of neurons in the  $n - 1$  layer. In this work, the sigmoid function is the activation function

in all hidden layers. For a simple network, several choices for the activation function  $f$  are possible. We note that, for simplicity, the same activation function was used in all hidden layers. Although this is not a general requirement, we do not expect any advantage in using different activation functions in different layers (for two or more hidden layers). The output node has a linear activation function for mapping inputs to a real-valued output in all of  $\mathbb{R}$ .

Weights are determined by minimizing an error function,

$$\Delta = \frac{1}{2} \sum_{i \in \Gamma} (V_i^{\text{QM}} - V_i^{\text{ANN}})^2, \quad (4)$$

where  $V_i^{\text{QM}}$  represents the single-point energy obtained with a quantum mechanical calculation and  $V_i^{\text{ANN}}$  the corresponding ANN prediction.  $\Gamma$  is the set of training points. In this work, we use the Levenberg-Marquardt minimizer.

By differentiating the negative of the potential energy with respect to atomic coordinates, one obtains the forces required to integrate the equations of motion according to Newton's second law. We find that a relatively shallow ANN is sufficient to accurately fit the QM energy points (see Sec. III A). Therefore, for simplicity, we present the gradient expressions with respect to input variables for a network with one hidden layer and for a network with two hidden layers.

In the simplest case (one hidden layer), consisting of  $N_0$  inputs and  $N_1$  neurons in the hidden layer, the output can be written as

$$o_k = \sum_i^{N_1} \omega_{i,k}^{(1,2)} h_i^{(1)} + b_k^{(2)}, \quad (5)$$

where  $o_k = V^{\text{ANN}}$  as  $k = 1$ . Furthermore,

$$h_i^{(1)} = f\left(\sum_m^{N_0} \omega_{m,i}^{(0,1)} G_m + b_m^{(1)}\right). \quad (6)$$

We seek to compute

$$\frac{\partial o_k}{\partial G_j} = \frac{\partial V^{\text{ANN}}}{\partial G_j}. \quad (7)$$

Because of the linearity of the sum operator, it follows that

$$\frac{\partial o_k}{\partial G_j} = \sum_i^{N_1} \omega_{i,k}^{(1,2)} \frac{\partial h_i^{(1)}}{\partial G_j}, \quad (8)$$

and using the chain rule in Eq. (6),

$$\frac{\partial h_i^{(1)}}{\partial G_j} = \omega_{i,j}^{(0,1)} f'\left(\sum_m^{N_0} \omega_{m,i}^{(0,1)} G_m + b_m^{(1)}\right). \quad (9)$$

By combining Eq. (9) with Eq. (8), we obtain

$$\frac{\partial o_k}{\partial G_j} = \sum_i^{N_1} \omega_{i,j}^{(0,1)} \omega_{i,k}^{(1,2)} f'\left(\sum_m^{N_0} \omega_{m,i}^{(0,1)} G_m + b_m^{(1)}\right). \quad (10)$$

A similar derivation can be done for a network with two hidden layers. The ANN now consists of  $N_0$  inputs,  $N_1$  neurons in the first hidden layer, and  $N_2$  neurons in the second hidden layer. The output can be written as

$$o_k = \sum_i^{N_2} \omega_{i,k}^{(2,3)} h_i^{(2)} + b_k^{(3)}, \quad (11)$$

where

$$h_i^{(2)} = f\left(\sum_m^{N_1} \omega_{m,i}^{(1,2)} h_m^{(1)} + b_i^{(2)}\right) \quad (12)$$

and

$$h_m^{(1)} = f\left(\sum_l^{N_0} \omega_{l,m}^{(0,1)} G_l + b_m^{(1)}\right). \quad (13)$$

By a similar procedure to that described earlier for one hidden layer,

$$\frac{\partial o_k}{\partial G_j} = \sum_i^{N_2} \omega_{i,k}^{(2,3)} \frac{\partial h_i^{(2)}}{\partial G_j}, \quad (14)$$

where

$$\frac{\partial h_i^{(2)}}{\partial G_j} = \sum_m^{N_1} \omega_{m,i}^{(1,2)} \frac{\partial h_m^{(1)}}{\partial G_j} f'\left(\sum_i^{N_1} \omega_{m,i}^{(1,2)} h_m^{(1)} + b_i^{(2)}\right) \quad (15)$$

and

$$\frac{\partial h_m^{(1)}}{\partial G_j} = \omega_{l,m}^{(0,1)} f'\left(\sum_l^{N_0} \omega_{l,m}^{(0,1)} G_l + b_m^{(1)}\right). \quad (16)$$

By substituting Eq. (16) into Eq. (15), and then Eq. (15) into Eq. (14), one obtains the desired result, namely  $\frac{\partial o_k}{\partial G_j} = \frac{\partial V^{\text{ANN}}}{\partial G_j}$ .

## B. Permutation invariance

Arguably, the main difficulty with the application of ANN to PES construction is the correct inclusion of symmetry. Because we are interested in low-dimensional PESs relevant for air thermochemistry, we will not discuss here possible strategies based on local atomic environment energy decompositions [22]. As stated previously, the main advantage of such approaches is their generality, in that they do not depend on the particular type of system, but they are more suited to condensed phases or biomolecules, whose PESs have much higher dimensionality.

In general, rotational and translational invariance are satisfied in a straightforward way by setting the input variables  $G_j$  as a function relative distances between atoms ( $r_{ij}$ ). However, the total energy invariance with respect to the exchange of like-atoms is much more problematic to satisfy for the simple feed-forward ANN, due to the fact that weights connecting the input layer to the first hidden layer are *hardwired* to each particular input variable  $G_j$ . For polynomial-based interpolators, the review article of Braams and Bowman [28] provides a summary of the computational mathematics underlying the construction of PESs based on permutationally invariant polynomials. We remark that several concepts presented there are used later in the work illustrated here.

The simplest approach to account for permutation invariance in a neural network is to explicitly include all equivalent structures in the training data set [29]. In this way, the ANN is said to learn the symmetry. For example, for each  $N_2 + N_2$  geometry, there are 24 equivalent arrangements that must produce the same energy, corresponding to all possible permutation of N atoms. This approach has two obvious drawbacks:

(1) the training set for highly symmetric systems becomes quite large, thus making the ANN training computationally expensive;

(2) permutation invariance is satisfied only approximately.

An alternative solution is to use symmetric neurons in the first hidden layer, as suggested by Prudente *et al.* [30]. Although this ANN modification does enforce permutation symmetry by construction, it requires *ad hoc* changes to the network structure and the training algorithm, and it still remains system-specific. Another approach involves *symmetrizing* atomic coordinates to obtain transformed input variables that satisfy the required symmetry [31]. Once again, this approach does enforce symmetry by construction, but it is highly problem-specific and it may introduce discontinuities in the gradient.

A more general and rigorous method is that of Guo and coworkers [32,33] who use permutation invariant polynomials to map the input variables (interatomic distances) onto a set of permutation invariant polynomials containing all the polynomials truncated at a given degree (the highest degree of the primary and secondary invariants). This is the approach of Li and coworkers [25]. Despite its generality and mathematical rigor, the method becomes impractical as the number of polynomials increases nonlinearly with the degree bound. A simpler and more computationally efficient approach is that of Shao *et al.* [34] who use a set of invariant polynomials, called fundamental invariants (FI), as the input of the ANN. This is the approach that we use in this work.

In the following, we will use the following nomenclature:

(1) the polynomial PES fit based on permutation invariant polynomials using mixed-exponential Gaussian (MEG) bond order variables will be denoted by MEG-PIP [26];

(2) the neural network PES fit based on fundamental invariants inputs using MEG bond order variables will be denoted by FI-ANN (present work);

(3) the neural network PES fit based on permutation invariant polynomials inputs using exponential bond order variables will be denoted by PIP-ANN [25].

For an  $A_4$  system (e.g.,  $N_2 + N_2$  or  $O_2 + O_2$ ), the PES can be written as a function of the six interatomic distances:

$$V(\mathbf{x}) = V(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}), \quad (17)$$

where we define the vector  $\mathbf{x} = (r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$ . Similar to a previous study [26], we used MEG functions of interatomic distances:

$$X_m := e^{-\frac{x_j - r_{eq}}{a} - \frac{(x_j - r_{eq})^2}{b}} \quad (18)$$

for  $m = 1, 2, \dots, 6$ ,  $r_{eq}$  is the equilibrium bond length of the diatomic molecule, while  $a$  and  $b$  are arbitrary scaling parameters. For  $A_4$  systems, the FI polynomials are the following [34]:

$$G_1 = X_1 + X_2 + X_3 + X_4 + X_5 + X_6,$$

$$G_2 = X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2,$$

$$G_3 = X_1X_2 + X_1X_3 + X_1X_4 + X_1X_5 + X_2X_3 + X_2X_4 + X_2X_6 + X_3X_5 + X_3X_6 + X_4X_5 + X_4X_6 + X_5X_6,$$

$$G_4 = X_1^3 + X_2^3 + X_3^3 + X_4^3 + X_5^3 + X_6^3,$$

$$G_5 = X_1^2X_2 + X_1^2X_3 + X_1X_3^2 + X_3^2X_5 + X_3^2X_6$$



$$\begin{aligned}
& + X_2 X_3^2 + X_2^2 X_3 + X_1 X_2^2 + X_2^2 X_4 + X_2^2 X_6 \\
& + X_2 X_6^2 + X_3 X_6^2 + X_5 X_6^2 + X_4 X_6^2 + X_4^2 X_6 \\
& + X_2 X_4^2 + X_1 X_4^2 + X_4^2 X_5 + X_4 X_5^2 + X_5^2 X_6 \\
& + X_3 X_5^2 + X_1 X_5^2 + X_1^2 X_5 + X_1^2 X_4,
\end{aligned}$$

$$G_6 = X_1 X_2 X_3 + X_3 X_5 X_6 + X_2 X_4 X_6 + X_1 X_4 X_5,$$

$$G_7 = X_1^4 + X_2^4 + X_3^4 + X_4^4 + X_5^4 + X_6^4,$$

$$\begin{aligned}
G_8 = & X_1^3 X_2 + X_1^3 X_3 + X_1 X_3^3 + X_3^3 X_5 + X_3^3 X_6 + X_2 X_3^3 \\
& + X_2^3 X_3 + X_1 X_2^3 + X_2^3 X_4 + X_2^3 X_6 + X_2 X_6^3 + X_3 X_6^3 \\
& + X_5 X_6^3 + X_4 X_6^3 + X_4^3 X_6 + X_2 X_4^3 + X_1 X_4^3 + X_4^3 X_5 \\
& + X_4 X_5^3 + X_5^3 X_6 + X_3 X_5^3 + X_1 X_5^3 + X_1^3 X_5 + X_1^3 X_4,
\end{aligned}$$

$$G_9 = X_1^5 + X_2^5 + X_3^5 + X_4^5 + X_5^5 + X_6^5. \quad (19)$$

We note that the input space is now nine-dimensional, at variance with the PIP-ANN PES of Li *et al.* [25] which instead has 276 permutation invariant input polynomials. The expressions for the derivatives of the FI-ANN output with respect to input variables are unchanged, as detailed in Sec. II. The gradient of the ANN output with respect to the MEG inputs  $X_i$  is given by

$$\nabla_{\mathbf{X}} V = (\mathcal{J} \nabla_{\mathbf{G}} V)^T, \quad (20)$$

where  $\mathcal{J}$  is the Jacobian of the transformation in Eq. (19) and  $\mathbf{G} = (G_1, G_2, \dots, G_9)$  are the FI polynomials that are now the input set of the ANN. The Jacobian is given by

$$\mathcal{J} = \begin{bmatrix} \frac{dG_1}{dX_1} & \frac{dG_2}{dX_1} & \cdots & \frac{dG_9}{dX_1} \\ \frac{dG_1}{dX_2} & \frac{dG_2}{dX_2} & \cdots & \frac{dG_9}{dX_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dG_1}{dX_6} & \frac{dG_2}{dX_6} & \cdots & \frac{dG_9}{dX_6} \end{bmatrix}. \quad (21)$$

The chain rule of calculus gives us the gradient with respect to  $r_{ij}$ , with components

$$\frac{\partial V}{\partial r_{ij}} = \frac{\partial V}{\partial X_m} \frac{\partial X_m}{\partial r_{ij}}, \quad (22)$$

i.e., the forces needed to integrate the molecular trajectories (Sec. III B). As stated, we adapt this description to the  $N_2 + N_2$  system. Therefore, we set  $r_{eq} = 1.098 \text{ \AA}$  and select  $a = 1 \text{ \AA}$  and  $b = 1.5 \text{ \AA}^2$  as done by Bender and coworkers [26].

To conclude this section, we note that the FI polynomials in Eq. (19) are specific to systems of four identical interacting atoms (e.g.,  $O_2 + O_2$ ,  $F_2 + F_2$ , etc.). For systems of comparable dimensionality (e.g.,  $O_2 + O$  or  $N_2 + O_2$ ), a different set of inputs  $\mathbf{G}$  is required, as discussed by Shao *et al.* [34]. However, the procedure just described to construct the FI-ANN PES and compute its gradient is identical. Finally, a significant advantage of the FI-ANN PES is its lower input dimensionality compared to the PIP-ANN PES [25]. In fact, although rather tedious, the analytic evaluation of the FI-ANN PES gradient is fairly straightforward and can be coded quite compactly. This allows us to avoid using a numerical evaluation of the gradient in trajectory integration, as instead done in the work of Li *et al.* [25], likely due to the cost of evaluating the transformation Jacobian [Eq. (21)] for the 276-dimensional polynomial input vector. This removes a

discretization error associated with the numerical evaluation of the PES gradient, which could be significant in regions of the surface characterized by steep variations, which may be explored during high-energy collisions typical of high-temperature gas conditions.

### C. Long-range asymptotic behavior

The robust application of neural networks (but in general both parametric [35] and nonparametric techniques [36]) to trajectory calculations requires further considerations on the long-range behavior of the resulting energy fit. It is convenient to decompose the total energy as a sum of pair-wise energies,  $V_2(x_i)$ , and a many-body contribution  $V_{MB}(\mathbf{x})$ , as in previous works [15,16],

$$V(\mathbf{x}) = V_0 + \sum_i^6 V_2(x_i) + V_{MB}(\mathbf{x}), \quad (23)$$

where  $V_0$  is the total energy required to break the molecular bonds of two isolated molecules. There are several reasons to apply the decomposition in Eq. (23):

(1) For many diatomic molecules, the bond energy curve  $V_2(x_i)$  is very well-known from highly accurate quantum mechanical calculations (e.g.,  $N_2$  in the work of Gdanitz [37]).

(2) A stand-alone function to compute  $V_2(x_i)$  makes the calculation of molecular internal states using the Wentzel-Kramers-Brillouin [11] method more efficient than directly using the full PES  $V(\mathbf{x})$  function.

(3) Provided that  $V_{MB}(\mathbf{x})$  tends to zero at large connected distances (i.e.,  $r_{13}$ ,  $r_{14}$ ,  $r_{23}$ , and  $r_{24}$ ), the total energy of the system reduces to  $V(\mathbf{x}) = V_0 + V_2(x_1) + V_2(x_6)$ .

Therefore, we fit the FI-ANN PES to the many-body term  $V_{MB}^{ANN}(\mathbf{x})$  and use the diatomic energy curve  $V_2(x_i)$  for  $N_2$  as parametrized by Pauku *et al.* [16].

In general,  $V_{MB}(\mathbf{x})$  is a function of all distances, including unconnected distances ( $r_{12}$  and  $r_{34}$ ), even when MEG variables [Eq. (18)], that decay to zero at infinite separation, are used. For example,  $X_1$  and  $X_6$  will not be zero for two bound molecules separated by an infinite distance. For this reason, terms containing  $X_1$  and  $X_6$  that are not multiplied by terms decaying to zero ( $X_i$  with  $i \in \{2, 3, 4, 5\}$ ) cause a small, spurious interaction even when two diatoms are infinitely far away, as shown later in Fig. 3. This spurious coupling remains because of the inability of the optimizer to fit with extreme precision the near-zero energy of interaction of diatoms at large separations.

Unlike in polynomial descriptions (see Supplemental Material of the work of Pauku *et al.* [16]), where such terms can be explicitly removed from the functional form, this is not straightforward, or even possible at all, in other types of interpolators. To correct for this nonphysical behavior, we use a tapering function  $\omega = \omega[V_{MB}(\mathbf{x})]$  to smoothly bring  $V_{MB}(\mathbf{x})$  to zero at large connected distances where  $V_{MB}(\mathbf{x}) \rightarrow 0$ :

$$V(\mathbf{x}) = V_0 + \sum_i^6 V_2(x_i) + \omega V_{MB}(\mathbf{x}), \quad (24)$$

where  $\omega$  takes the following expression:

$$\omega[V_{MB}(\mathbf{x})] = \frac{1}{2} \{1 + \tanh[B_0(V_{MB}(\mathbf{x})^2 - D_0^2)]\}. \quad (25)$$

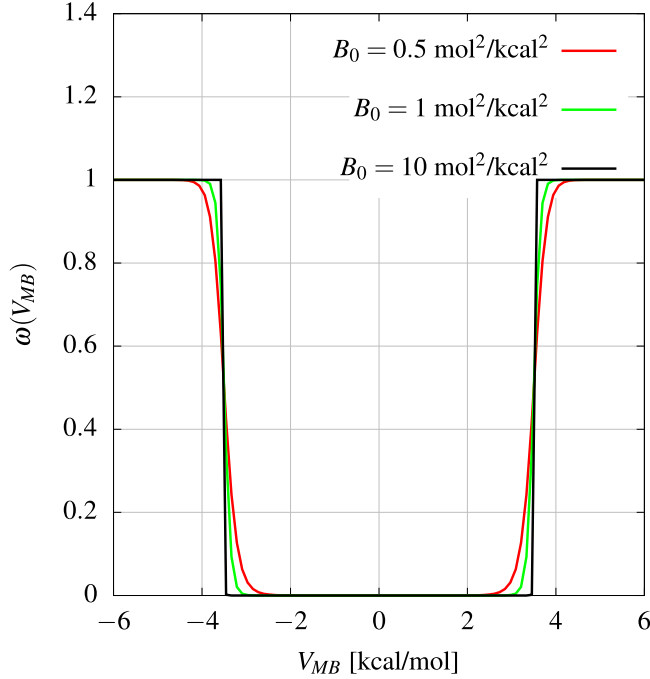


FIG. 2. Tapering function to remove spurious long-distance interactions between diatoms.

We manually optimized  $B_0 = 10 \text{ mol}^2/\text{kcal}^2$  and  $D_0 = 3.5 \text{ kcal/mol}$ . The function  $\omega$  is shown in Fig. 2. Because of the way  $\omega[V_{\text{MB}}(\mathbf{x})]$  is defined, the switching to zero occurs smoothly whenever  $|V_{\text{MB}}| < D_0$ , with  $D_0 \simeq \text{RMSE}$ .

In Fig. 3, we show the internal separations  $r_{12}(t)$  and  $r_{34}(t)$  of two isolated molecules that vibrate and rotate without any

relative motion. By introducing the tapering function  $\omega$ , it can be seen in Figs. 3(a) and 3(b) that  $r_{12}(t)$  is unaffected by  $r_{34}(t)$  and viceversa. However, when  $\omega = \text{const} = 1$ , it is shown in Figs. 3(c) and 3(d) that  $r_{12}(t)$  and  $r_{34}(t)$  are nonphysically coupled by the residual  $V_{\text{MB}}(\mathbf{x}) \neq 0$  at large separations. Although such an effect is rather small, about 4% in Figs. 3(c) and 3(d), it must be eliminated as it may produce incorrect dissociation of molecules at the threshold of dissociation (e.g., quasibound states). Because dissociation of diatomic molecules is strongly localized at high rovibrational levels [26,38,39], a lack of such a correction would predict a much greater (and incorrect) dissociation probability due to this numerical artifact rather than the potential surface morphology.

The correction suggested in this work for the robust application of ANNs to PES construction is different than the approach of Li and coworkers [25] in their PIP-ANN PES, where terms containing unconnected distances are removed from the input vector. We tested a similar approach in the FI-ANN method, but observed a considerably worse fitting, which resulted in unacceptable levels for the global RMSE. Although we did not investigate this in detail, we believe that is caused by the fact that at close range (which is the vast majority of the geometries in the training set), all distances are required to describe the energetics of the system, including *unconnected distances*.

In addition to removing long-distance spurious coupling, the use of a tapering function mathematically guarantees that the interaction energy becomes precisely zero at large separations. This is at variance with the approach taken by Li and coworkers [25] where the decay to zero is obtained by including several single-point energies at very large separations (well beyond the dispersive force range) and by relying

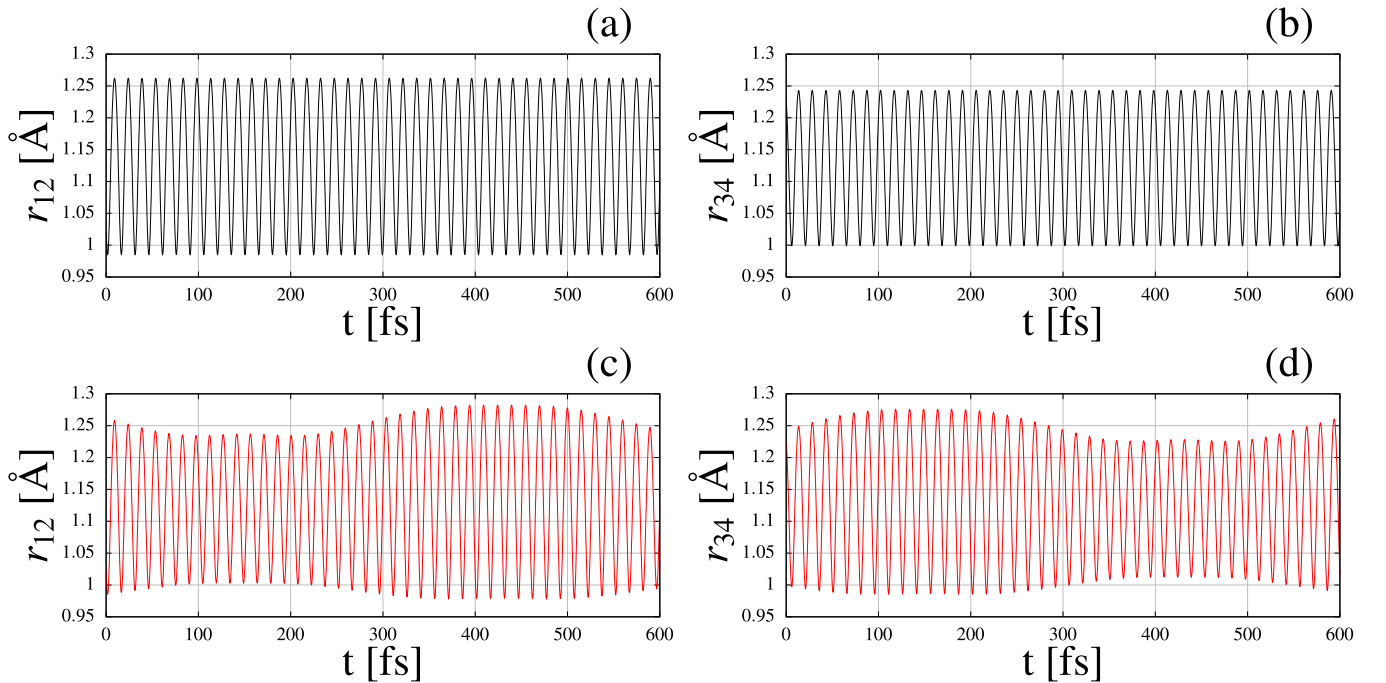


FIG. 3. Rovibrational motion of two isolated diatoms with no relative speed: (a, b) with  $\omega = \omega[V_{\text{MB}}(\mathbf{x})]$  (no long-distance coupling); (c, d)  $\omega = \text{const} = 1$  (spurious long-distance coupling).

on the optimization procedure to reduce the fitting error on those configurations to nearly zero. While the two approaches could be equivalent for standard QCT methodologies, the decay of the long-range interaction energy precisely to zero could be advantageous in DSMC-like methods, such as DMS [13], where molecules undergo thousands, or even millions, of successive collisions. Therefore, a small, but finite interaction energy at large distances could introduce a small kinetic energy loss after the termination of each trajectory. Such a small energy deficit could potentially accumulate over physical time (or, equivalently, collisions) to produce a sizable energy loss in an adiabatic system.

In previous works using a polynomial PES [40,41], geometry-based switching functions were utilized to activate or deactivate the relevant physical portions of the PES (e.g., switching from the close-range interaction to the very long range interaction [41]). Such switching functions were based on a geometric criterion, generally based on interatomic distances. Although similar, our approach is directly based on the many-body energy, which does not decay precisely to zero at large separations. This allows us to avoid any geometric constraints (e.g., identifying a range over which the switching occurs  $[r_{\min}, r_{\max}]$ ) that are system-specific. With the definition of  $\omega$ , the tapering occurs independently of the particular chemical system (i.e.,  $N_4$ ,  $O_4$ ,  $N_2 + O_2$ , etc.). Furthermore, for a generic chemical system, the many-body interaction energy is not necessarily a strictly positive (or negative) term asymptoting to zero at large separations, but can change sign in the close-range interaction region, thus crossing zero. If that happens, then the tapering function would reduce smoothly the term to zero within  $\pm D_0$ , which is set to be of the order of the global RMSE. This is to ensure that the overall quality of the fit would not be significantly compromised.

Finally, we remark that the choice of the tapering function is not unique and has no physical justification, but only a mathematical one. At very low temperatures, where the long-range attraction becomes important, a further investigation of an appropriate choice for  $\omega[V_{MB}(\mathbf{x})]$  may become necessary. In this work, we are only concerned with kinetic processes at high temperatures (hypersonic shock layer conditions), where dispersive attractive forces are not important. At high collisional energies, bound-bound transitions are primarily affected by the repulsive wall of the PES, which is well-reproduced by the FI-ANN fit, as shown later.

### III. RESULTS

#### A. Fit to $N_2 + N_2$ *ab initio* energy data

To assess the physical accuracy of the FI-ANN PES, we select the  $N_2 + N_2$  system, for which an extensive data set of *ab initio* single point energies was obtained by Paukku *et al.* [16], and later slightly expanded by Bender *et al.* [26]. This data set contains 16534 geometries, most of which describe various interactions of two  $N_2$  molecules. A small sub-set of geometries are for  $N_2 + N$  interactions, thus making the resulting PES transferable to the nitrogen atom-diatom system. The energies were calculated by the complete active space second-order perturbation theory (CASPT2) quantum mechanical electronic structure method with the

maug-cc-pVTZ58 basis set [16,26]. As this work was being prepared, Li *et al.* [25] further extended the data set with an additional 4859 single-energy points using the CASPT2 method and 13 single-energy points using restricted coupled-cluster calculations with single and double excitations and quasiperturbative connected triple excitations, CCSD(T).

For a detailed description of the geometries, we refer the reader to the work of Paukku *et al.* [16]. Here, however, we briefly describe the main features of the selected geometries. Most  $N_2 + N_2$  atomic configurations in the data set [16,26] include a *spectator* molecule at its equilibrium bond distance ( $r_A = r_{eq}$ ), while the other molecule undergoes dissociation, i.e.,  $r_B$  is varied, at various separations between their centers of mass (from 1 Å to 10 Å). Paukku and coworkers [16] further considered an elongated ( $r_A = r_{eq} + 0.2$  Å) and a compressed ( $r_A = r_{eq} - 0.2$  Å) spectator molecule, thus, in practice, describing interactions where only one molecule may be significantly rovibrationally excited. Unlike Paukku *et al.* [16], another commonly used PES for  $N_2 + N_2$ , developed at NASA Ames Research Center [15], is instead based on geometries where also the spectator molecule's bond length is significantly varied beyond its equilibrium value (between  $r_{eq} - 0.2$  Å to 5 Å). A detailed description of the differences between these two surfaces, both from the perspective of geometry selection and quantum chemistry methods, is presented by Jaffe and coworkers [18].

This clearly demonstrates how the selection of atomic configurations in the training set is somewhat arbitrary and relies on physical intuition more than a systematic approach. For example, restricting the spectator bond's stretch to 0.2 Å may not be ideal to describe interactions between highly excited molecules in the shock layer, where, depending on conditions, temperatures often exceed tens of thousands of K. This also exemplifies a scenario where a portion of the function domain might require some refining, and the fitting algorithm should be robust enough to a possibly significant expansion of the training set. Quite recently, Li and coworkers [25] have extended the data set for this system to include geometries where both interacting molecules have bond distances significantly different from the equilibrium value.

In general, an important question is how a PES performs outside of the training region. An example of how both the FI-ANN PES and the MEG-PIP PES extrapolate energies outside the fitting data set is shown in Fig. 4. For the selected PES cut (X-shape with  $d = 1$  Å and spectator molecule's bond  $r_A = 2.098$  Å), the difference between the MEG-PIP PES and the FI-ANN PES is revealed by the black lines. The solid black line denotes the FI-ANN PES while the black symbols are the MEG-PIP PES. A marked difference can be seen, particularly in the range  $r_B \in [2, 7]$  Å. We believe that the only way to resolve these differences is to compare to new quantum mechanical energies in the region of interest. Additionally, artificial neural networks are susceptible to poor performance when a significant portion of the functional domain is missing. This problem was also identified in the work of Li *et al.* [25], where more geometries were added to the training set to improve coverage of the PES domain, thus resulting in the addition of nearly 5000 more single-energy points. These new points were added based on physical intuition and, in part, guided by a less accurate polynomial fit, which is less

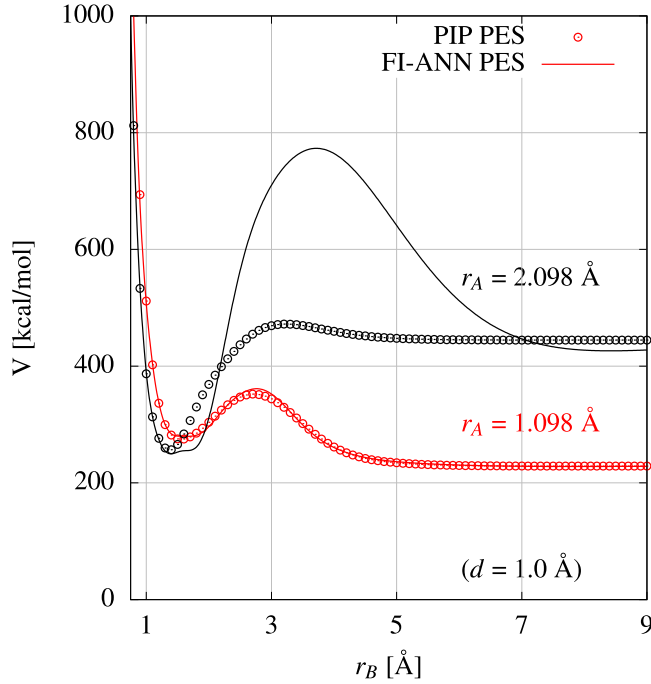


FIG. 4. FI-ANN PES (present work) and MEG-PIP[26] PES interpolation and extrapolation: X-shape with  $d = 1 \text{ Å}$  and (i) spectator molecule's bond  $r_A = 1.098 \text{ Å}$  (in the training set, red line is FI-ANN PES and symbols are MEG-PIP PES) and (ii) spectator molecule's bond  $r_A = 2.098 \text{ Å}$  (out of the training set, black line is FI-ANN PES and symbols are MEG-PIP PES).

susceptible to overtraining and issues with extrapolation in poorly resolved regions of the PES domain. In our study, however, the objective is not to refine or produce an alternative PES for the  $\text{N}_2 + \text{N}_2$  system with additional quantum mechanical single-point energies, but to demonstrate the correct application of artificial neural networks to fit *ab initio* energy data. For this reason, we *augment* the training set by including a few additional geometries with  $r_B$  stretched beyond  $\approx 1.2 \text{ Å}$ . Specifically, we added A-, BT-, HB-, H-, I-, X-, and T-shaped arrangements with  $r_B = \{1.498, 2.098, 2.498\} \text{ Å}$  to the training set (see the article of Paukku *et al.* [16] for a definition of the geometries) with corresponding *surrogate* energy values obtained from the MEG-PIP PES [26]. The *augmented* training set contains a total of about 33,000 geometries. We point out that it is not unusual for PES construction to compute tens of thousands of single point energies to populate the training data set (e.g., approximately 50 000 data points for the recent  $\text{N}_2 + \text{O}_2$  surface [42]). ANNs are very good at fitting large data sets by simply varying their architecture, as discussed earlier.

A feed-forward artificial neural network consists of several hidden layers, each containing several neurons (see Sec. II A). Hence, we test two network configurations, with one or two hidden layers. Each hidden layer is composed of a number of neurons ranging from 10 to 40. The input layer size is defined by the dimensionality of the problem (here,  $\mathbb{R}^9$  as explained in Sec. II B) and the output by the codomain of the function to interpolate (here,  $\mathbb{R}$ , i.e., the potential energy). We use the Levenberg-Marquardt optimizer implemented in

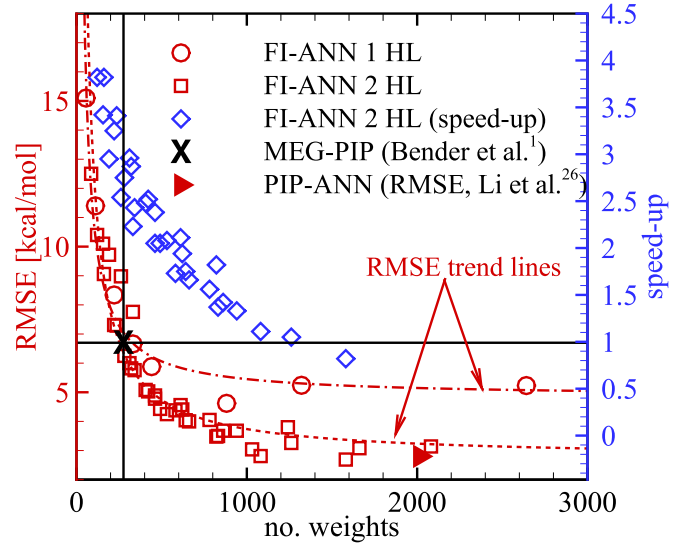


FIG. 5. Root-mean-square error (RMSE) as a function of number of network weights for a one hidden layer FI-ANN and a two hidden layer FI-ANN; speed-up of a two hidden layer FI-ANN compared to the MEG-PIP PES [26], including both function and gradient evaluation. The vertical line marks the number of fitting parameters in the MEG-PIP PES, i.e., 276; the horizontal line marks the overall RMSE of the MEG-PIP PES (6.7 kcal/mol) on the left and the reference speed-up of one on the right.

MATLAB R2018b [43] to determine the fitting parameters. Early stopping is used to prevent overfitting, with 20% of the randomized training set used as validation set. Because this optimization is highly nonlinear, the resulting RMSE depends on the initial guess for the network weights and biases. Therefore, we repeat each optimization 20 times and select the network with the lowest RMSE.

Figure 5 shows the RMSE on the  $\text{N}_2 + \text{N}_2$  original energy data set (16 534 geometries) as a function of the total number of ANN weights. For the FI-ANN, we restrict the computation of the RMSE only to the *ab initio* portion of the training set, thus excluding the surrogate energies used in the fitting. This is because we think that it is the RMSE (a global measure of error) on the first-principles data that has a greater significance when comparing different fits. For a given number of hidden layers, more weights corresponds to more neurons per hidden layer. As expected, the RMSE decreases as the total number of free parameters increases. When the number of weights becomes very large, the optimization algorithm is very likely to encounter a local minimum, due to the high-dimensionality of the optimization space. Therefore, no significant further reduction in RMSE is observed. However, for a given number of weights, the network depth (i.e., number of hidden layers) improves the fitting quality (reduction in RMSE), thus suggesting that increased depth enables the network to fit the nonlinearities of the data set more accurately.

In Fig. 5, the vertical solid line marks the total number of parameters in the MEG-PIP PES (276 free parameters). It can be seen that networks with around that number of parameters have comparable or better accuracy than the FI-PIP PES, while achieving a good computational speed-up. Here, the speed-up is defined as the ratio between the time



TABLE I. Comparison of the RMSE for the MEG-PIP [26] and selected network configurations with two hidden layers for the FI-ANN PES. The speed-up includes both function and gradient evaluation.

| Energy [kcal/mol] | MEG-PIP PES [26] | 10 × 10 | 20 × 20 | 30 × 30 |
|-------------------|------------------|---------|---------|---------|
| $V < 100$         | 1.3              | 2.7     | 1.6     | 1.2     |
| $100 < V < 228$   | 3.8              | 5.0     | 2.7     | 1.7     |
| $228 < V < 456$   | 4.9              | 6.1     | 3.6     | 2.7     |
| $456 < V < 1000$  | 12.2             | 12.8    | 7.4     | 5.3     |
| All               | 6.7              | 7.3     | 4.2     | 3.1     |
| Speed-up          | 1                | ~4      | ~3      | ~2      |

needed to integrate one time step on the MEG-PIP PES and on the FI-ANN PES, thus including both function and gradient evaluation. Remarkably, with networks that reduce the RMSE by roughly a half, a speed-up of about 2 is still obtained.

The results for the RMSE in kcal/mol for selected network configurations (two hidden layers with 10, 20, or 30 neurons each) used in trajectory calculations (Sec. III B) are listed in Table I and grouped by energy range. With the exception of the smallest network (10 × 10), relatively small networks (20 × 20 and 30 × 30) produce a similar or reduced RMSE compared to the MEG-PIP PES, both over the whole energy range and in each energy subset, while achieving a significant speed-up in the current implementation.

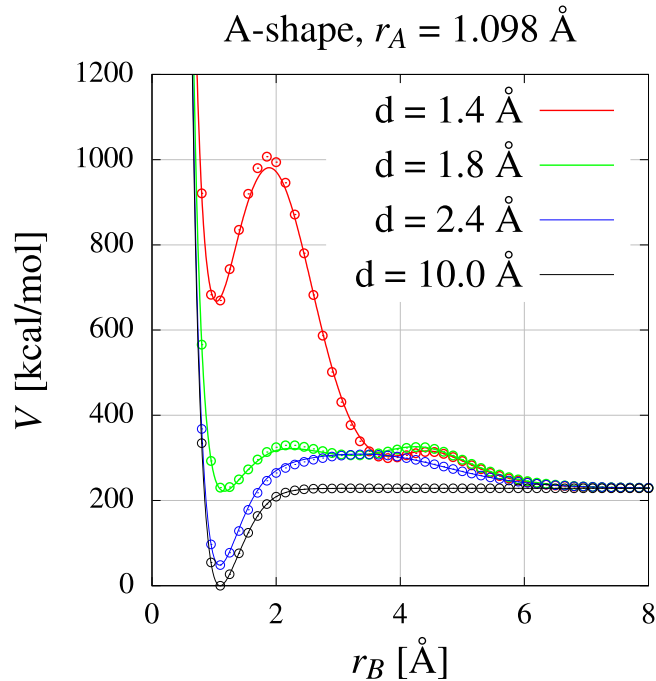


FIG. 6. Comparison between the 20 × 20 FI-ANN PES (symbols) and MEG-PIP[26] PES (solid line) for the A-shape geometry PES cut. The spectator molecules has a bond length of  $r_A = 1.098$  Å and the dissociating partner bond length  $r_B$  is varied between 0.5 Å and 8 Å. The distance between the centers of mass is denoted by  $d$ .

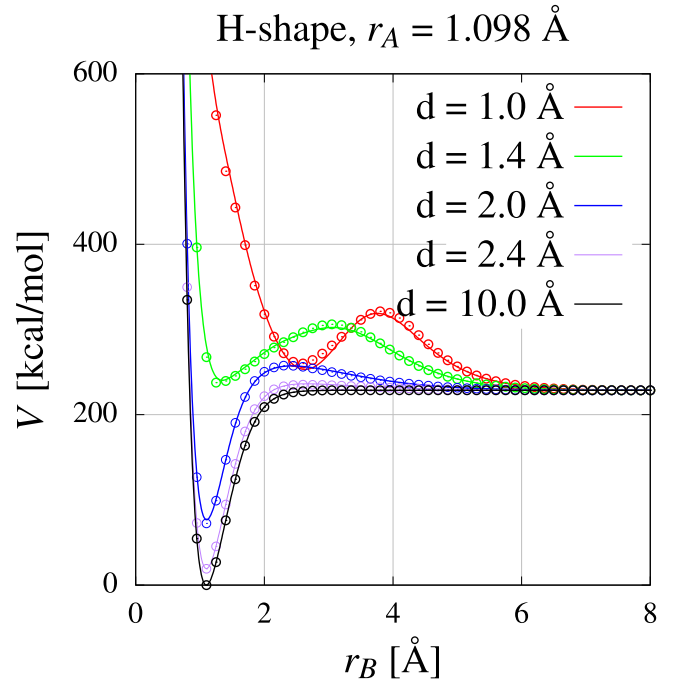


FIG. 7. Comparison between the 20 × 20 FI-ANN PES (symbols) and MEG-PIP [26] PES (solid line) for the H-shape geometry PES cut. The spectator molecules has a bond length of  $r_A = 1.098$  Å and the dissociating partner bond length  $r_B$  is varied between 0.5 Å and 8 Å. The distance between the centers of mass is denoted by  $d$ .

In the work of Li *et al.* [25], a network containing 2016 fitting parameters was selected and produced an overall RMSE of 2.8 kcal/mol on the complete training data set. Although on a different training set, our FI-ANN networks produced a relatively comparable RMSE (4.2 kcal/mol with 641 parameters for the 20 × 20 ANN and 3.1 kcal/mol with 1261 parameters for the 30 × 30 ANN). As shown in Fig. 5, FI-ANNs with around 2000 parameters could achieve an even greater reduction of the RMSE below 3 kcal/mol. This, however, was at the expense of computational efficiency. Furthermore, it is unclear how a further reduction of RMSE reflects in the error on the observables obtained from QCT or DMS techniques, for example. Finally, it is difficult to comment on the difference in computational performance between our FI-ANN PES and the PIP-ANN PES in the work of Li *et al.* [25]. While we generally observed a computational speed-up or similar cost compared to the MEG-PIP PES (Fig. 5), Li *et al.* [25] report that their ANN force calculation (with a numerical evaluation of the PES gradient) was about an order of magnitude slower than the MEG-PIP PES. Such differences could be due to the specific implementations, and, thus, should not be made general when comparing different methods of fitting *ab initio* data set for PES construction.

Selected cuts are shown in Figs. 6–10 at various  $N_2 + N_2$  separations ( $d$ ) and spectator molecule's bond length. In the figures, the FI-ANN 20 × 20 predictions (symbols) are compared to the MEG-PIP PES energy (solid line) for various

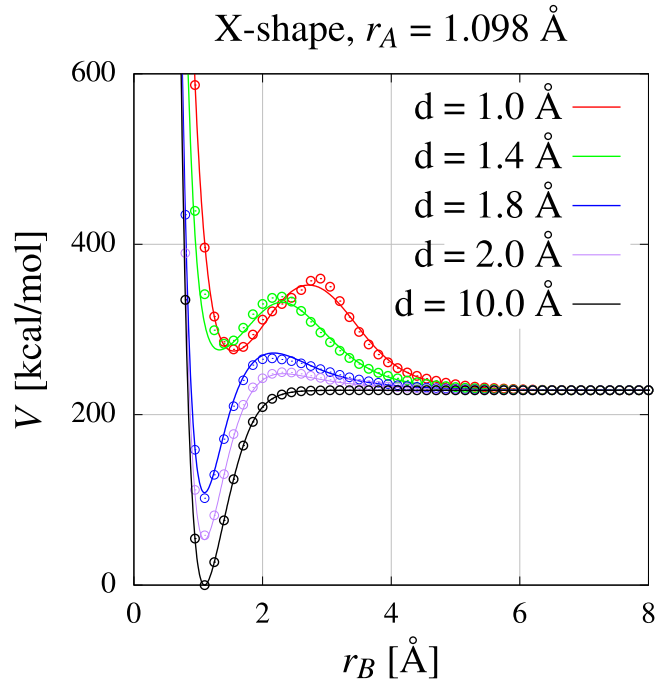


FIG. 8. Comparison between the  $20 \times 20$  FI-ANN PES (symbols) and MEG-PIP [26] PES (solid line) for the X-shape geometry PES cut. The spectator molecules has a bond length of  $r_A = 1.098$  Å and the dissociating partner bond length  $r_B$  is varied between 0.5 Å and 8 Å. The distance between the centers of mass is denoted by  $d$ .

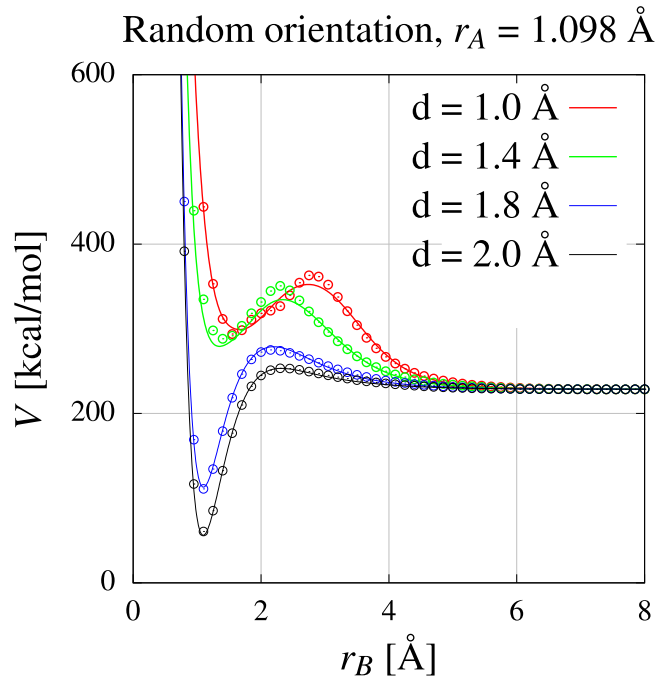


FIG. 9. Comparison between the  $20 \times 20$  FI-ANN PES (symbols) and MEG-PIP [26] PES (solid line) for a random orientation geometry PES cut. The spectator molecules has a bond length of  $r_A = 1.098$  Å and the dissociating partner bond length  $r_B$  is varied between 0.5 Å and 8 Å. The distance between the centers of mass is denoted by  $d$ .

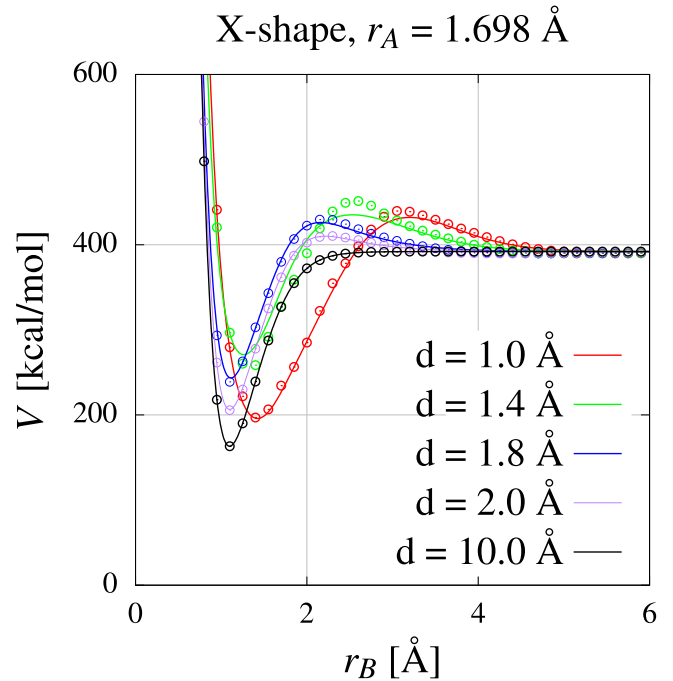


FIG. 10. Comparison between the  $20 \times 20$  FI-ANN PES (symbols) and MEG-PIP [26] PES (solid line) for an elongated X-shape geometry PES cut. The spectator molecules has a bond length of  $r_A = 1.698$  Å and the dissociating partner bond length  $r_B$  is varied between 0.5 Å and 6 Å. The distance between the centers of mass is denoted by  $d$ .

geometries in and out of the training set. Excellent agreement is seen in all cases.

### B. Trajectory calculations

Trajectory calculations are performed using the MEG-PIP PES [26] and three selected network configurations, namely two hidden layers each with 10, 20, or 30 neurons. Reactants preparation and product analysis are conducted according to the standard QCT procedure. The work of Bender *et al.* [26] contains the details. Briefly, the effective potential [11] is used to enumerate all allowed rovibrational states for the  $N_2$  molecule via the semiclassical Wentzel-Kramers-Brillouin (WKB) approximation [11]. Then, the rovibrational partition function [26] is used to generate the distribution of the initial energy states for thermal equilibrium corresponding to the desired temperature. The total molecular internal energy is partitioned between rotational and vibrational energy using a vibration-prioritized splitting scheme [5]. Rotational and vibrational energies are used to assign the internal phase space coordinates of the two nitrogen atoms [26]. Finally, a relative speed is assigned to the colliding pair from a Maxwell-Boltzmann distribution corresponding to the selected temperature and product states are analyzed using the WKB approximation. Whereas reactants states are assigned at *integer* values for rotational and vibrational states, product number states are *real* valued, i.e., no binning procedure is used.

The velocity Verlet time integration scheme is used to advance atomic positions in time [14]. A time step of 0.01 fs is used to ensure excellent energy conservation, as was

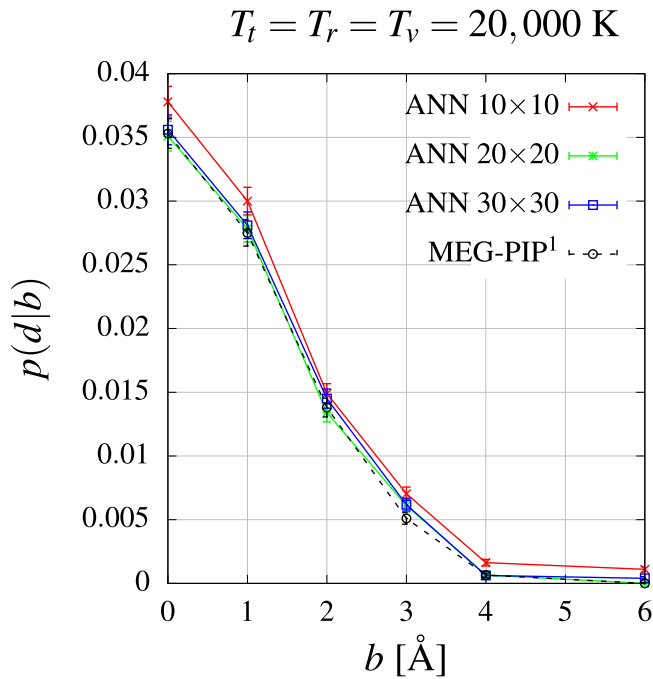


FIG. 11. Comparison of dissociation probabilities at 20 000 K obtained from equilibrium quasiclassical trajectory calculations on MEG-PIP PES [26] and FI-ANN PES.

demonstrated previously [26]. The selected temperatures are 10 000 K, 20 000 K, and 30 000 K. We only consider thermal equilibrium for the preparation of the reactant molecules, i.e.,  $T_t = T_r = T_v$ , where the subscripts  $t, r, v$  denote translation, rotation, and vibration, respectively. Dissociation probabilities are given a statistical uncertainty according to the Wald method [44].

The thermally averaged dissociation probabilities at a given impact parameter  $b$  are shown in Figs. 11, 12, and 13. In Fig. 11, we compare the results obtained from three different FI-ANN architectures ( $10 \times 10$ ,  $20 \times 20$ , and  $30 \times 30$ ) with those obtained from the MEG-PIP PES. Excellent agreement between dissociation probabilities obtained from trajectory calculations on the MEG-PIP PES and on the FI-ANN PES can be seen, particularly when the number of neurons for each hidden layer is 20 and 30. The  $10 \times 10$  FI-ANN appears to overpredict the dissociation probabilities compared to the MEG-PIP PES (about 20% higher overall dissociation rate). This is likely due to the poorer fitting performance (see Table I) compared to the larger networks.

Similar quantitative agreement is shown in Fig. 12 at 30 000 K. The  $20 \times 20$  FI-ANN PES reproduces the MEG-PIP PES probabilities very well over the whole impact parameter range. We expect similar results from the other networks like in the 20 000 K case.

In Fig. 13, we show the results at 10 000 K for  $b = 0$  Å and  $b = 1$  Å. This case is considerably more challenging than the previous ones because the reaction probabilities are about three orders of magnitude smaller. Each probability required over 3 million trajectories. Despite the statistical uncertainty, we again see excellent agreement between the  $20 \times 20$  FI-ANN PES and the MEG-PIP PES.

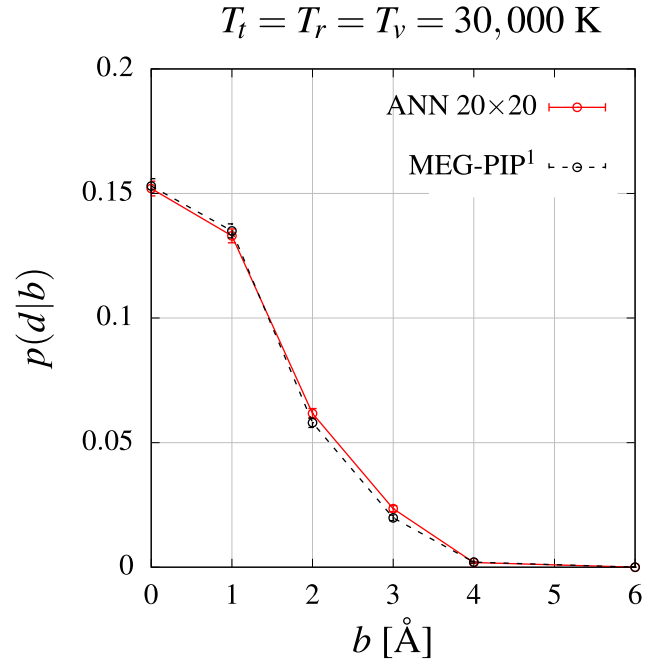


FIG. 12. Comparison of dissociation probabilities at 30 000 K obtained from equilibrium quasiclassical trajectory calculations on MEG-PIP PES [26] and FI-ANN PES.

Finally, we compare energy distributions obtained from the MEG-PIP PES and the FI-ANN PES trajectories. In Fig. 14, we plot the distribution of vibrational states of the product molecules that do not undergo a dissociation in collisions with  $b = 0$  Å. Also shown is the initial Boltzmann distribution of reactants. MEG-PIP PES and FI-ANN PES trajectories

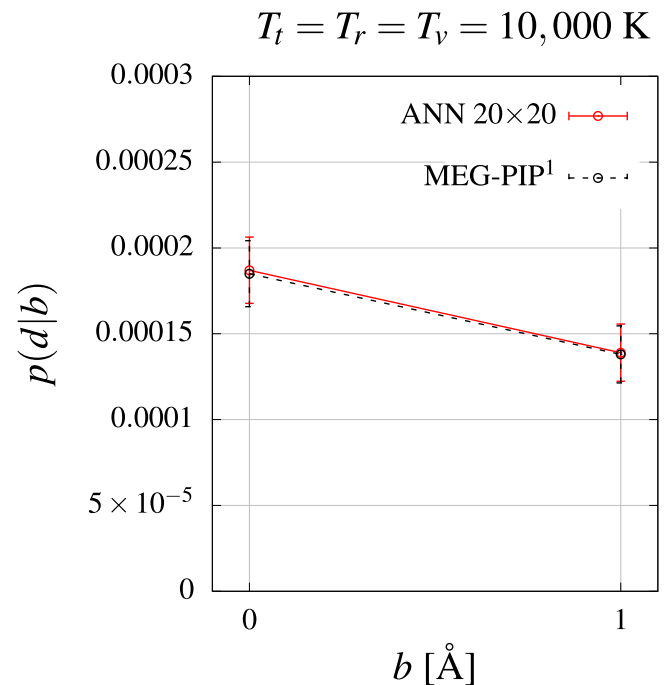


FIG. 13. Comparison of dissociation probabilities at 10 000 K obtained from equilibrium quasiclassical trajectory calculations on MEG-PIP PES [26] and FI-ANN PES.

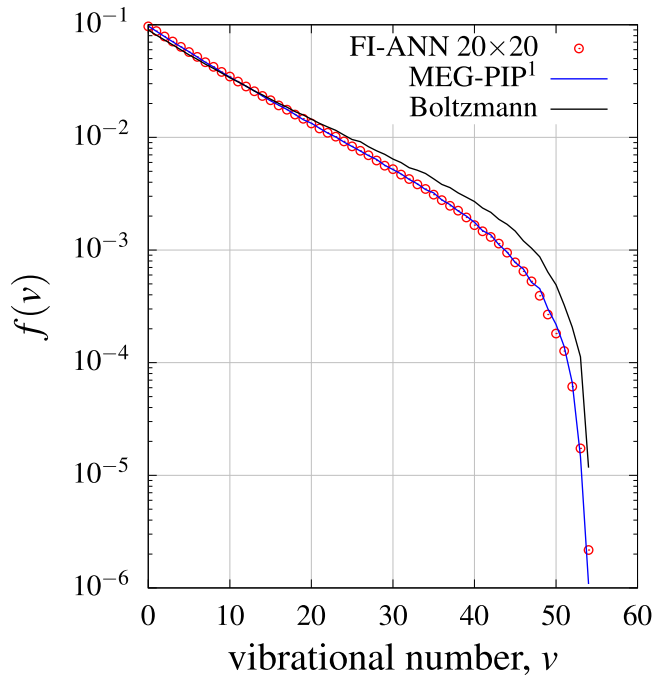


FIG. 14. Comparison of vibrational distributions for products undergoing nondissociative collisions at 30 000 K as obtained from equilibrium quasiclassical trajectory calculations on MEG-PIP [26] PES and FI-ANN PES.

predict nearly identical distributions, characterized by the well-known depletion of high- $v$  states [38,45].

#### IV. CONCLUSIONS

In this article, we have described the application of a simple feed-forward ANN architecture to fit the  $N_2 + N_2$  PES. For the applications of interest, namely molecular simulation of high-temperature gases, there are specific considerations on the physical accuracy of the resulting fit, due to the strong coupling of internal energy excitation and vibrational dissociation. Moreover, application to large-scale direct simulations requires computational efficiency as well. In our implementation, translation, rotation, and permutation invariance are satisfied by mapping the interatomic distances onto a set of

permutation invariant inputs, known as fundamental invariants, that generate the permutation invariant polynomial ring. With an appropriate energy decomposition, we impose the correct two-body energy contribution and we remove spurious long-distance interactions between diatoms with a tapering function.

The ANN PES is shown to be as accurate or to have superior accuracy compared to the standard description based on permutation invariant polynomials at about the same number of free fitting parameters. The RMSE on the training set geometries (16 534 data points) is generally lower than the MEG-PIP PES, with the exception of the smallest network considered ( $10 \times 10$ ). The RMSE is also broken down by energy range and similar improvements are observed. A relatively small network ( $20 \times 20$ ) considerably outperforms the MEG-PIP PES.

Then, we conduct a QCT study on the dissociation probability, conditioned on the impact parameter. The results show excellent agreement between the probabilities obtained from trajectories performed on the FI-ANN PES and on the MEG-PIP PES. Similar agreement is seen at the level of vibrational energy distributions. The most significant discrepancy is observed with the smallest network used in this study, namely the  $10 \times 10$  ANN. However, the error on the overall dissociation rate does not exceed about 20%.

In our implementation, FI-ANN PESs are generally more computationally efficient than the MEG-PIP PES at a comparable RMSE. For the  $20 \times 20$  ANN, which has an overall lower RMSE than the MEG-PIP PES, a considerable speed-up of nearly 3 is observed.

In conclusion, our study shows that fundamental invariant artificial neural networks represent an attractive alternative to traditional permutation invariant polynomials, provided that physical constraints and asymptotes are correctly enforced. We demonstrate that a methodology for reliable PES fitting should include both PES training and molecular dynamics trajectories to identify potential issues with extrapolation or long-range asymptotic behavior (e.g., spurious coupling). Furthermore, the fundamental invariant input method requires much smaller network sizes compared to formulations based on polynomial invariant inputs [25], thus making this approach computationally efficient for large-scale direct simulations of gas flows [13].

- [1] K. Haug, D. G. Truhlar, and N. C. Normand, Monte Carlo trajectory and master equation simulation of the nonequilibrium dissociation rate coefficient for  $Ar + H_2 \rightarrow Ar + 2H$  at 4500 K, *J. Chem. Phys.* **86**, 2697 (1987).
- [2] J. E. Dove and S. Raynor, *Ab initio* calculation of the rate of vibrational relaxation and thermal dissociation of hydrogen by helium at high temperatures, *J. Phys. Chem.* **1**, 127 (1979).
- [3] F. Esposito, I. Armenise, and M. Capitelli, N- $N_2$  state to state vibrational-relaxation and dissociation rates based on quasiclassical calculations, *Chem. Phys.* **331**, 1 (2006).
- [4] G. D. Billing and E. R. Fisher, VV and VT rate coefficients in  $N_2$  by a quantum-classical model, *Chem. Phys.* **43**, 395 (1979).
- [5] M. Panesi, R. L. Jaffe, D. W. Schwenke, and T. E. Magin, Rovibrational internal energy transfer and dissociation of  $N_2(^1\Sigma_g^+)-N(^4S_u)$  system in hypersonic flows, *J. Chem. Phys.* **138**, 044312 (2013).
- [6] A. Lagana, E. Garcia, and L. Ciccarelli, Deactivation of vibrationally excited nitrogen molecules by collision with nitrogen atoms, *J. Phys. Chem.* **91**, 312 (1987).
- [7] A. Lagana and E. Garcia, Temperature dependence of nitrogen atom-molecule rate coefficients, *J. Phys. Chem.* **98**, 502 (1994).
- [8] J. G. Kim and I. D. Boyd, State-resolved master equation analysis of thermochemical nonequilibrium of nitrogen, *Chem. Phys.* **415**, 237 (2013).
- [9] E. Josyula, W. F. Bailey, and S. M. Ruffin, Reactive and nonreactive vibrational energy exchanges in nonequilibrium hypersonic flows, *Phys. Fluids* **15**, 3223 (2003).



- [10] M. Capitelli, I. Armenise, D. Bruno, M. Cacciatore, R. Celiberto, G. Colonna, O. De Pascale, P. Diomede, F. Esposito, C. Gorse, K. Hassouni, A. Laricchiuta, S. Longo, D. Pagano, D. Pietanza, and M. Rutigliano, Non-equilibrium plasma kinetics: A state-to-state approach, *Plasma Sources Sci. Technol.* **16**, S30 (2007).
- [11] D. G. Truhlar and J. T. Muckerman, *Atom-Molecule Collision Theory: A Guide for the Experimentalist* (Plenum, New York, NY, 1979), p. 505.
- [12] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon, Oxford, 1994).
- [13] T. E. Schwartzentruber, M. S. Grover, and P. Valentini, Direct molecular simulation of nonequilibrium dilute gases, *J. Thermophys. Heat Transfer*, **32**, 892 (2017).
- [14] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, CA, 2002).
- [15] R. Jaffe, D. Schwenke, and G. Chaban, Vibrational and rotational excitation and dissociation in  $N_2$ - $N_2$  collisions from accurate theoretical calculations, Number AIAA 2010-4517, in *Proceedings of the 10th ASME/AIAA Joint Thermophysics Conference* (AIAA, Reston, VA, 2010).
- [16] Y. Paukku, Y. R. Ke, Z. Varga, and D. G. Truhlar, Global *ab initio* ground-state potential energy surface of  $N_4$ , *J. Chem. Phys.* **139**, 044309 (2013).
- [17] L. Pacifici, M. Verdicchio, N. F. Lago, A. Lombardi, and A. Constantini, A high-level *ab initio* study of the  $N_2 + N_2$  reaction channel, *J. Comput. Chem.* **34**, 2667 (2013).
- [18] R. L. Jaffe, M. Grover, S. Venturi, D. W. Schwenke, P. Valentini, T. E. Schwartzentruber, and M. Panesi, Comparison of potential energy surface and computed rate coefficients for  $N_2$  dissociation, *J. Thermophys. Heat Transfer* **32**, 869 (2018).
- [19] J. Behler, Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations, *Phys. Chem. Chem. Phys.* **13**, 17930 (2011).
- [20] K. Hornik, M. Stinchcombe, and H. White, Multilayer feed-forward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
- [21] J. D. Bender, S. Doraiswamy, D. G. Truhlar, and G. V. Candler, Potential energy surface fitting by a statistically localized, permutationally invariant, local interpolating moving least squares method for the many-body potential: Method and application to  $N_4$ , *J. Chem. Phys.* **140**, 054302 (2014).
- [22] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [23] N. Artrith, A. Urban, and G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, *Phys. Rev. B* **96**, 014112 (2017).
- [24] G. P. Purja Pun, R. Batra, R. Ramprasad, and Y. Mishin, Physically informed artificial neural networks for atomistic modeling of materials, *Nat. Commun.* **10**, 2339 (2019).
- [25] J. Li, Z. Varga, D. G. Truhlar, and H. Guo, Many-body permutationally invariant polynomial neural network potential energy surface for  $N_4$ , *J. Chem. Theory Comput.* **8**, 4822 (2020).
- [26] J. D. Bender, P. Valentini, I. Nompelis, Y. Paukku, Z. Varga, D. G. Truhlar, T. E. Schwartzentruber, and G. V. Candler, An improved potential energy surface and multi-temperature quasiclassical trajectory calculations of  $N_2 + N_2$  dissociation reactions, *J. Chem. Phys.* **143**, 054304 (2015).
- [27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016), <http://www.deeplearningbook.org>.
- [28] B. J. Braams and J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- [29] H. P. Le., A. R. Karagozian, and J.-L. Cambier, Complexity reduction of collisional-radiative kinetics for atomic plasma, *Phys. Plasma* **20**, 123304 (2013).
- [30] F. V. Prudente, P. H. Acioli, and S. J. J. Neto, The fitting of potential energy surfaces using neural networks: Application to the study of vibrational levels of  $H_3^+$ , *J. Chem. Phys.* **109**, 8801 (1998).
- [31] H. Gassner, M. Probst, A. Lauenstein, and K. Hermansson, Representation of intermolecular potential functions by neural networks, *J. Phys. Chem. A* **102**, 4596 (1998).
- [32] B. Jiang and H. Guo, Permutation invariant polynomial neural network approach to fitting potential energy surfaces, *J. Chem. Phys.* **139**, 054112 (2013).
- [33] J. Li, B. Jiang, and H. Guo, Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems, *J. Chem. Phys.* **139**, 204103 (2013).
- [34] K. Shao, J. Chen, Z. Zhao, and D. H. Zhang, Communication: Fitting potential energy surfaces with fundamental invariant neural network, *J. Chem. Phys.* **145**, 071101 (2016).
- [35] J. M. Bowman, G. Czako, and B. Fu, High-dimensional *ab initio* potential energy surfaces for reaction dynamics calculations, *Phys. Chem. Chem. Phys.* **13**, 8094 (2011).
- [36] J. Cui and R. V. Krems, Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with gaussian processes, *J. Phys. B: At. Mol. Phys.* **49**, 224001 (2016).
- [37] R. J. Gdanitz, Accurately solving the electronic Schrödinger equation of atoms and molecules using explicitly correlated (r12-)MR-CI: The ground state potential energy curve of  $N_2$ , *Chem. Phys. Lett.* **283**, 253 (1998).
- [38] P. Valentini, T. E. Schwartzentruber, J. D. Bender, I. Nompelis, and G. V. Candler, Direct molecular simulation of nitrogen dissociation based on an *ab initio* potential energy surface, *Phys. Fluids* **27**, 086102 (2015).
- [39] P. Valentini, T. E. Schwartzentruber, J. D. Bender, and G. V. Candler, Dynamics of nitrogen dissociation from direct molecular simulation, *Phys. Rev. Fluids* **1**, 043402 (2016).
- [40] Z. Xie, B. J. Braams, and J. M. Bowman, *Ab initio* global potential-energy surface for  $H_5^+ \rightarrow H_3^+ + H_2$ , *J. Chem. Phys.* **122**, 224307 (2005).
- [41] Z. Jin, B. J. Braams, and J. M. Bowman, An *ab initio* based global potential energy surface describing  $CH_5^+ \rightarrow CH_3^+ + H_2$ , *J. Phys. Chem. A* **110**, 1569 (2006);.
- [42] Z. Varga, R. Meana-Pañeda, G. Song, Y. Paukku, and D. G. Truhlar, Potential energy surface of triplet  $N_2O_2$ , *J. Chem. Phys.* **144**, 024310 (2016).
- [43] *MATLAB Release 2018b*, The MathWorks, Inc.
- [44] A. Agresti and B. A. Coull, Approximate is better than “exact” for interval estimation of binomial proportions, *Amer. Stat.* **52**, 119 (1998).
- [45] P. Valentini, P. Norman, C. Zhang, and T. E. Schwartzentruber, Rovibrational coupling in molecular nitrogen at high temperature: An atomic-level study, *Phys. Fluids* **26**, 056103 (2014).