THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

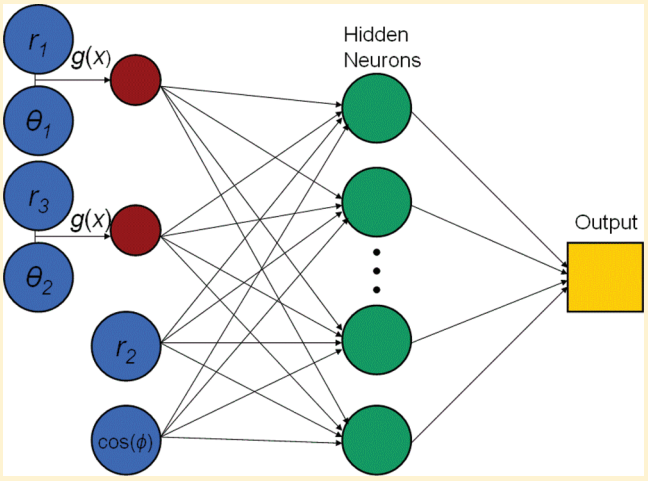
Modified Feed-Forward Neural Network Structures and Combined-Function-Derivative Approximations Incorporating Exchange Symmetry for Potential Energy Surface Fitting

Hieu T. T. Nguyen and Hung M. Le*

Faculty of Materials Science, College of Science, Vietnam National University, Ho Chi Minh City, Vietnam

Supporting Information

ABSTRACT: The classical interchange (permutation) of atoms of similar identity does not have an effect on the overall potential energy. In this study, we present feed-forward neural network structures that provide permutation symmetry to the potential energy surfaces of molecules. The new feed-forward neural network structures are employed to fit the potential energy surfaces for two illustrative molecules, which are H_2O and ClOOCl. Modifications are made to describe the *symmetric* interchange (permutation) of atoms of similar identity (or mathematically, the permutation of *symmetric* input parameters). The combined-function-derivative approximation algorithm (*J. Chem. Phys.* **2009**, *130*, 134101) is also implemented to fit the neural-network potential energy surfaces accurately. The combination of our symmetric neural networks and the function-derivative fitting effectively produces PES fits using fewer numbers of training data points. For H_2O, only 282 configurations are employed as the training set; the testing root-mean-squared and mean-absolute energy errors are respectively reported as 0.0103 eV (0.236 kcal/mol) and 0.0078 eV (0.179 kcal/mol). In the ClOOCl case, 1693 configurations are required to construct the training set; the root-mean-squared and mean-absolute energy errors for the ClOOCl testing set are 0.0409 eV (0.943 kcal/mol) and 0.0269 eV (0.620 kcal/mol), respectively. Overall, we find good agreements between ab initio and NN prediction in term of energy and gradient errors, and conclude that the new feed-forward neural-network models advantageously describe the molecules with excellent accuracy.



I. INTRODUCTION

Artificial neural network[1] (NN) is a powerful tool in function fitting and pattern classification. The method has been applied to many research areas during the last two decades. The terminology "neural network" derives from the superficial resemblance of the mathematical network present in a NN to that present in the human brain.[2] To date, several NN models with different mathematical structures are suggested. It has been found that the feed-forward NN model[1] is particularly robust, and it has been vastly employed in function fitting and data processing. The simple feed-forward NN constructions provide easy manipulation and utilization; hence they are applied in many chemical and biological research aspects.[3] Nearly two decades ago, Gasteiger and Zupan suggested several specific uses of NNs in analysis of spectroscopy, chemical reaction, process examinations, and electrostatic potentials.[4]

For a long time, the applications of feed-forward NNs in theoretical reaction dynamics have been proposed and utilized, in which the NN models have been employed to produce analytic fits for potential energy surfaces (PES) that allow rapid reproduction of energy and analysis of gradients. By adopting the NN technique, the fitted PESs for various systems have been developed with different levels of complexity depending upon the molecular systems of interest. Those systems include condensed-phase and gas-phase molecular systems. Two detailed reviews about NN methodology and applications in analytical PES construction are available for consulting in the literature.[5]

The first effort that employed the NN method to produce analytic PESs for solid system interactions was presented by Blank et al.,[6] in which the NN potentials described the absorption of CO on Ni(111) surface and interaction between H_2 and Si(100)-2×1 surface. Investigations of surface reaction dynamics of H_2 on the potassium-covered (and sulfur-covered in a subsequent study) Pd(100) surface were conducted by Lorenz and Scheffler,[7] in which the NN method is employed to construct six-dimensional PESs of the investigated systems. A variety of studies conducted by Behler and co-workers that involved NN PES construction and molecular dynamics (MD) simulations, i.e., dissociation of O_2 at Al(111) in consideration

Received: March 1, 2012
Revised: April 24, 2012
Published: April 25, 2012

   dx.doi.org/10.1021/jp3020386 | *J. Phys. Chem. A* 2012, 116, 4629−4638

of spin selection rules,[8] pressure-induced phase transition of silicon,[9] interatomic potential for high pressure, and high temperature sodium liquid and crystal.[10] The PES of zinc oxide bulk material was developed using the NN method,[11] and it was found that the NN energies were in excellent agreement with the DFT energies whereas the NN function allowed more rapid access of energies and gradients. In a recent work, a NN PES of energetic interaction of water dimer was reported, and this effort was devoted to be an intermediate step to produce NN potentials that describes water system with higher complexity.[12]

For isolated gas-phase systems, the NN method has been a popular tool and widely applied for years. Prudente and Neto reported an investigation of HCl^+ photodissociation that involved NN fitting of the PES.[13] Several other systems with higher complexity have been reported and recognized to date, including a chemical reaction that involve multiplicity switch (surface hopping) like SiO_2,[14] the complicated dissociation schemes of vinyl bromide (CH_2CHBr),[15] HONO,[16] HOOH,[17] BeH + H_2,[18] and ozone (O_3).[19] In those reported problems, the NN method has been proved to be a powerful and robust method that can be employed to reproduce ab initio potential energies rapidly and accurately.

Since the rigorous development of NN PESs, accuracy in numerical fitting has become a leading context, especially for MD simulations. It is significant to have both energies and gradients accurately predicted to perform MD trajectories. In an earlier work, the combined-energy-gradient fitting algorithm in feed-forward NNs has been proposed and testified successfully in the illustrating H + HBr problem.[20] In terminology, this technique is referred to as combined-function-derivative approximation (CFDA). It is also reported elsewhere that the approximation of a function and its derivatives was numerically achieved using radial-basis NN,[21] and the fitting results were measured with superior accuracy. In our work, besides proposing a new feed-forward NN structure, we also implement the CFDA algorithm for accurate energy and gradient fitting, which would further help to interpolate data points and better resemble function curvatures based on the numerical fitting of function derivatives. Such CFDA implementation is based on the referenced study,[20] and the algorithm is implemented to work properly for our modified NN training.

In most reported works regarding NN construction for PES, one disadvantage of the method is that it requires a large amount of data points to train the NNs. In the vinyl bromide (CH_2CHBr) problem,[15a] nearly 72 000 points were required to fit the PES for such a six-body system with 15 internal coordinates. Several other works for four-body systems (with 6 internal coordinates) were also reported with the PESs constructed by fitting more than 20 000 data points.[16−18] To construct the PESs for three-atom molecules such as SiO_2 and O_3, it was reported that about 6000 configurations were employed.[14,19] With the implementation of derivative fitting in the CFDA algorithm, the NN can better interpolate data points and thereby reproduce the approximating functions within a requirement of fewer configurations. We look forward to maintaining the fitting quality and reducing the number of training data points in the fitting process as presented in the two illustrative problems (the vibrational PES for H_2O and the reactive PES for ClOOCl).

In molecules such as H_2O and ClOOCl, when we interchange two or multiple atoms of similar identity, the potential energy is not affected, and we term such input variables to be *symmetric*. One limitation can be pointed out clearly from many NN studies; i.e., the symmetric property of variables is understood by neither general feed-forward NN construction nor automatic machine-learning algorithm. In several previous studies, this circumstance was roughly handled by duplicating the existing database (with the *symmetric* variables being interchanged).[17−19] However, this treatment would result in big extension of the database, hence cause lower fitting accuracy and high computational cost. Consequently, it is not realistic to adopt the above treatment to deal with molecules with high complexity (with multiple pairs of *symmetric* variables). Therefore, the main objective in this research is to develop a new feed-forward NN construction that can automatically and effectively handle permutation of *symmetric* input variables in the two case studies.

The handlings of *symmetry* have been demonstrated using different approaches in a numerous NN studies. The potential energy surface of the $H_2O-Al^{3+}-H_2O$ system was constructed as a symmetric function that allowed interchange of atoms of similar identity. In such work, the symmetry of O and H atoms was handled by initially processing the inputs, which employed some "symmetrization functions" to destroy the individuality of initial symmetric variables and thus produce a new set of linear variables in the NN.[22] The PES of the H_3^+ system was developed by Prudente and co-workers, in which all permutations of three distance variables were introduced into the generalized NN.[23] Lorenz et al.[24] employed several symmetry functions to produce a set of eight symmetry-adapted coordinates, which sufficiently described the interaction of H_2 and (2×2) potassium-covered Pd(100) surface. In another work, symmetry functions similar to empirical potentials were employed by Behler and Parrinello[25] to manipulate the input signals, and constraints were put on the weights of the NN function to produce symmetry. The modifications on neural network structures in our study is distinctive from those treatment reported in the literature; i.e., modifications are made directly on the first neural layer of feed-forward NN structures and effectively incorporate exchange *symmetry* to NN functions. Such modifications are made on the weight values of the first neural layer and consequently results in a smaller number of NN parameters, which is an advantage of this method.

Two objectives are proposed and executed in this NN research. In the first objective, we present a modified designation for two-layer feed-forward NNs that effectively handle the molecules in which some input variables can be symmetrically permutated (1). The CFDA back-propagation fitting algorithm developed by Pukrittayakamee et al.[20] to train both energy and derivatives is implemented to train our *symmetric* neural networks (2). The presented techniques are applied to construct two PESs for two case studies, which are H_2O vibration and ClOOCl molecular dissociation.

II. TRADITIONAL TWO-LAYER FEED-FORWARD NEURAL NETWORK CONSTRUCTION

The mathematical formation of a traditional two-layer feed-forward NN is presented in this section. The structure of an artificial NN somewhat resembles the structure of real human-brain NN, in which information is transformed at one layer of neurons, and transmitted to the following layer for next-level processing. Adopting this phenomenon, in the artificial NN, the initial numerical input information is transmitted into the very first artificial neural layer, transformed by some predefined mathematical functions, and converted to be the input signal for the next neural layer. The activity of a typical two-layer NN is illustrated in Figure 1.

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

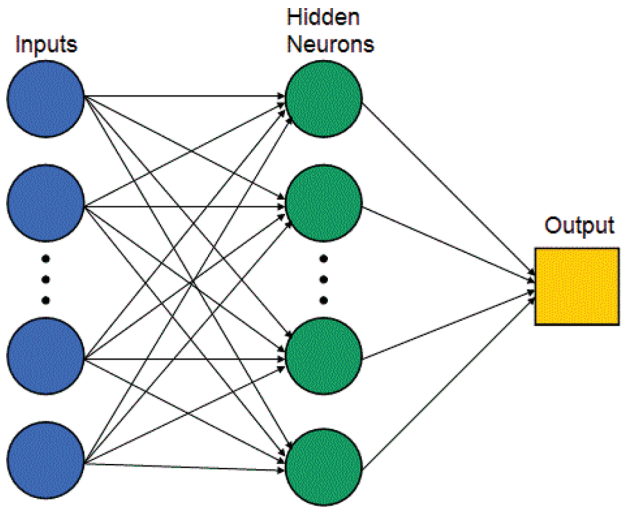


Figure 1. Traditional two-layer feed-forward NN designation.

Let us assume that the input signal comprises N real (and dimensionless) numbers, and we denote them as $(r_1, r_2, ..., r_N)$. If there are M neurons in the hidden layer, the input signals $(r_1, r_2, ..., r_N)$ are processed in the first neural layer to produce M output values a_i^1 as follows:

$$a_i^1 = f(n_i^1) = f\left(\sum_{j=1}^{N} w_{i,j}^1 r_j + b_i^1\right) \qquad i = 1, ..., M \tag{1}$$

where $w_{i,j}^1$ and b_i^1 are the weight and bias values of the first layer, respectively. f, the transfer function, is utilized to convert the sum signal to an output value, which is later adopted by the next neural layer as an input signal. In some earlier studies, it has been witnessed that the hyperbolic tangent function (tanh) and log-sigmoid function $((1 + e^{-x})^{-1})$ result in excellent fitting accuracy when they are employed as transfer functions in artificial NNs for global approximations of analytic functions.[5a,14−19,26]

The numerical outputs from the initial neural layer are then transmitted into the second layer (the output layer in our case) as input signals, and the final NN output a is calculated as shown in the following equation:

$$a = \sum_{i=1}^{M} w_i^2 a_i^1 + b^2 \tag{2}$$

In this equation, w_i^2 and b^2 are the weight and bias values of the second layer, respectively.

Usually, the NN-approximating function to a PES is achieved by training 90% of data, whereas 5% of data serve as a testing set, and the remaining 5% of data is used as a validation set. To prevent overfitting, the training procedure is terminated when the mean-squared error of validation set increases consecutively in a predefined number of training iterations (chosen by users). Such a technique is termed "early stopping",[1] and it is widely adopted in many NN training processes.

III. MODIFIED NEURAL NETWORK STRUCTURES FOR MOLECULES WITH SYMMETRIC INPUT VARIABLES

In this paper, we present NN fitting for two molecules in which input variables can be symmetrically interchanged (permutated) without affecting the potential energy. Those two molecules are H_2O and ClOOCl. For the H_2O system with C_{2v} symmetry, we do not construct a global PES that fully covers long-range atomic interaction nor H_2O dissociation.[27] In fact, we only consider a simple PES of molecular vibration as an illustrative problem.

Chlorine peroxide (ClOOCl) is a highly reactive compound that can dissociate easily to give radical products, which include $ClO^•$, $ClOO^•$, and $Cl^•$. It has been mentioned in several previous studies that this compound is an environmental hazard reagent that causes ozone depletion.[26,28] In this second case study, we construct a reactive PES for the complex four-body molecule on the basis of the available ClOOCl database to testify the effectiveness of our symmetry treatment and the energy-gradient fitting algorithm.

1. Water (H_2O) Molecule. There are three internal variables that fully describe the geometric configuration of water molecule, which are two O−H bonds and an H−O−H bending angle, as shown in Figure 2. For simplicity, let us

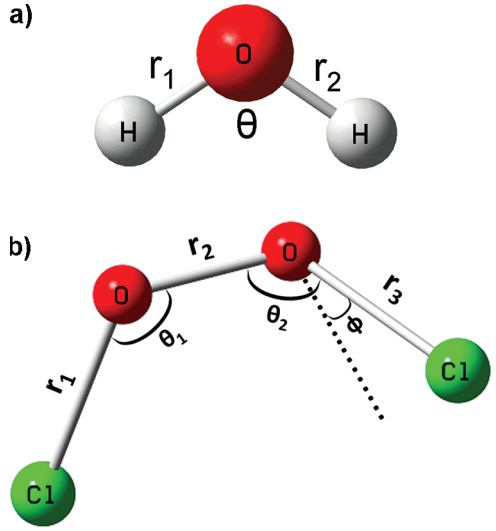


Figure 2. (a) Molecular structure of H_2O with the definition of three input variables. Note that r_1 and r_2 are two *symmetric* variables, and their permutation does not affect the overall energy. (b) Molecular structure of ClOOCl with the definition of six input variables. In this illustration, r_1 and r_3 are two *symmetric* variables, θ_1 and θ_2 are the other two *symmetric* variables, and the simultaneous interchanges $r_1 \leftrightarrow r_3$ and $\theta_1 \leftrightarrow \theta_2$ do not affect the overall potential energy.

denote those three variables as (r_1, r_2, r_3) where r_3 is the HOH bending angle, r_1 and r_2 variables are the two symmetric O−H bonds that can be permutated without affecting the overall potential energy of the system. Initially, inputs r_1 and r_2 are mapped in the range of $[0; 1]$ to give dimensionless input signal p_i using the equation below:

$$p_k = \frac{(r_k - r_{12,\min})}{(r_{12,\max} - r_{12,\min})} \qquad k = 1, 2 \tag{3}$$

In eq 3, r_{12_min} and r_{12_max} are the minimum and maximum values of r_1 (and r_2), respectively. Because r_1 and r_2 are two symmetric variables that can be interchanged, the scaled input variables p_1 and p_2 also share the interchangeable property, or in other words, they can be interchanged in the analytic NN function without affecting the output (energy). Similarly to r_1 and r_2, input parameter r_3 is scaled in the range $[0; 1]$ using the below equation:

$$p_3 = \frac{(r_3 - r_{3\_\min})}{(r_{3\_\max} - r_{3\_\min})} \tag{4}$$

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

The output value (energy) is also scaled in the range of [0; 1] by adopting a similar mathematical formula. In the first neural layer, function $g(x) = k_1x + k_2 \sin(x)$ is defined as the distinction function, and the log-sigmoid function $(1 + e^{-x})^{-1}$ is defined as the transfer function $f(x)$. For simplicity, we choose k_1 and k_2 to be unity, and function $g(x)$ simply becomes $x + \sin(x)$. The first and second derivatives of function $g(x)$ are therefore $\dot{g}(x) = 1 + \cos(x)$ and $\ddot{g}(x) = -\cos(x)$, respectively. The symmetric-variable problem is technically handled by modifying the weight values of the first layer. Inputs p_k are introduced into the first neural layer with M neurons and processed as below:

$$a_i^1 = f(n_i) = f(g(w_{i,1}^1 p_1) + g(w_{i,1}^1 p_2) + w_{i,3}^1 p_3 + b_i^1) \quad \text{for } i = 1...M \tag{5}$$

It should be noted in eq 5 that p_1 and p_2 share the same weight parameters $w_{i,1}^1$, and this modification provides *symmetry* to the NN function. Matrix w^1 thereby reads

$$w^1 = \begin{pmatrix} w_{11}^1 & w_{11}^1 & w_{13}^1 \\ w_{21}^1 & w_{21}^1 & w_{23}^1 \\ \vdots & \vdots & \vdots \\ w_{M1}^1 & w_{M1}^1 & w_{M3}^1 \end{pmatrix} \tag{6}$$

The use of such w^1 matrix provides *symmetry* to the NN. However, we further need to apply the distinction function $g(x)$ to distinct the signals p_1 and p_2. The purpose of such $g(x)$ utilization is to provide different identity to the gradients with respect to p_1 and p_2, i.e., without the use of $g(x)$, the derivatives of the NN function with respect to p_1 and p_2 are always identical. In a previous work reported by Behler and Parrinello,[25] the symmetry function G_i^μ is employed to transform the input signals and describe the "local geometric environment" of atom i in accordance with the remaining atoms. The use of our distinction function in our case adopts somewhat similar concepts. There is, however, a different purpose of using $g(x)$, which is providing different identity to gradients with respect to p_1 and p_2 as discussed above.

The output signal, a^1, is an M-dimension vector that presents M outputs of the first neural layer. The NN final output (produced in the second neural layer) is computed as shown in eq 2. In this simple case study of H_2O, we employ a 25-neuron NN ($M = 25$) to construct the PES for H_2O ground-state vibrations.

The NN training process is executed for both energy and derivatives with respect to inputs using the back-propagation algorithm.[1,29] In this H_2O illustrating problem, the energies and corresponding sets of gradients (with respect to three input parameters) are calculated using the second-order Moller–Plesset perturbation theory[30] (MP2) with the 6-31G* basis set[31] implemented in the Gaussian 03 suite of programs.[32] According to our ab initio calculations, the zero-point vibrational energy of H_2O molecule is approximately 0.584 eV; therefore, we believe that it is appropriate to choose the PES upper limit to be 1.500 eV. Hence, our goal is to develop a PES for H_2O that accurately reproduces energies of those configurations that are below 1.500 eV.

Suppose that a is the NN-predicted output energy, whereas t is the true target energy provided by MP2 calculations. During the NN training process, the linear combination of energy and gradient squared errors is denoted as P:

$$P = F + \rho D = \sum \left[(t - a)^2 + \rho \sum_{i=1}^{3} \left(\frac{\partial t}{\partial p_i} - \frac{\partial a}{\partial p_i} \right)^2 \right] \tag{7}$$

In the above equation, ρ is the scale factor that appears before the gradient errors and determines the significance of gradients. This factor may be adjusted to give the best optimal fitting result. Depending upon the training data set, it can be predetermined empirically as follows:

$$\rho = \left(\frac{\max|t|}{\max|\partial t / \partial p|} \right)^2 \tag{8}$$

Because all inputs and outputs are scaled using the scaling equation, all physical parameters (configuration inputs and output) are dimensionless (unitless), and such linear combination of energy and gradients in eq 7 are physically appropriate.

It is required in the back-propagation training algorithm that P is minimized during the training process by adjusting w_1, w_2, b_1, and b_2 on the basis of the derivatives of P with respect to those coefficients. The derivatives of P with respect to each coefficient of weight vector w^2 and bias b^2 read

$$\frac{\partial P}{\partial w^2} = -2(t - a)(a^1)^T - \frac{2\rho}{3} \left(\frac{\partial t}{\partial p} - \frac{\partial a}{\partial p} \right) \left(\frac{\partial a^1}{\partial p} \right)^T \tag{9}$$

$$\frac{\partial P}{\partial b^2} = -2(t - a) \tag{10}$$

Let us denote a new mathematical operator $\otimes$ as the array multiplication operator, which consequently produce an element-by-element product of two same-size arrays to give a new array of identical size. For convenience, we introduce a new vector d^1 of size $(M \times 1)$ and a new matrix H of size $(M \times 3)$ that will be used as an intermediate expression to back-propagate the derivatives of P with respect to w^1 and b^1.

$$d^1 = (1 - 2a^1) \otimes \left[\frac{\partial a^1}{\partial p} \left(-\frac{2}{3} \left(\frac{\partial t}{\partial p} - \frac{\partial a}{\partial p} \right)^T \right) \right] \otimes (w^2)^T \tag{11}$$

$$H = \frac{\partial n^1}{\partial w^1} = \begin{pmatrix} \dot{g}(w_{11}^1 p_1) p_1 + \dot{g}(w_{11}^1 p_2) p_2 & \dot{g}(w_{11}^1 p_1) p_1 + \dot{g}(w_{11}^1 p_2) p_2 & p_3 \\ \dot{g}(w_{21}^1 p_1) p_1 + \dot{g}(w_{21}^1 p_2) p_2 & \dot{g}(w_{21}^1 p_1) p_1 + \dot{g}(w_{21}^1 p_2) p_2 & p_3 \\ \vdots & \vdots & \vdots \\ \dot{g}(w_{M1}^1 p_1) p_1 + \dot{g}(w_{M1}^1 p_2) & \dot{g}(w_{M1}^1 p_1) p_1 + \dot{g}(w_{M1}^1 p_2) p_2 & p_3 \\ p_2 & & \end{pmatrix} \tag{12}$$

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629–4638

Matrix Y of size (N x 3) is defined as

$$Y = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} \\ y_{2,1} & y_{2,2} & y_{2,3} \\ \vdots & \vdots & \vdots \\ y_{M,1} & y_{M,2} & y_{M,3} \end{pmatrix} \tag{13}$$

where each value $y_{i,j}$ is computed as

$$y_{i,j} = -\frac{2}{3}\left(\frac{\partial t}{\partial p} - \frac{\partial a}{\partial p}\right)\begin{pmatrix} \ddot{g}(w^1_{i,1}p_1)w^1_{i,1}p_1 + \dot{g}(w^1_{i,1}p_1) \\ \ddot{g}(w^1_{i,1}p_2)w^1_{i,1}p_2 + \dot{g}(w^1_{i,1}p_2) \\ 0 \end{pmatrix} \quad \text{for } j = 1, 2$$

$$y_{i,j} = -\frac{2}{3}\left(\frac{\partial t}{\partial p} - \frac{\partial a}{\partial p}\right)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{for } j = 3$$

The derivatives of P with respect to w^1 and b^1 are

$$\frac{\partial P}{\partial w^1} = -2(t-a)\begin{pmatrix} a^1_1(1-a^1_1)w^2_1 & a^1_1(1-a^1_1)w^2_1 & a^1_1(1-a^1_1)w^2_1 \\ a^1_2(1-a^1_2)w^2_2 & a^1_2(1-a^1_2)w^2_2 & a^1_2(1-a^1_2)w^2_2 \\ \vdots & \vdots & \vdots \\ a^1_M(1-a^1_M)w^2_M & a^1_M(1-a^1_M)w^2_M & a^1_M(1-a^1_M)w^2_M \end{pmatrix} \otimes H + \rho\left[\begin{pmatrix} d^1 & d^1 & d^1 \end{pmatrix} \otimes h + \begin{pmatrix} a^1_1(1-a^1_1)w^2_1 & a^1_1(1-a^1_1)w^2_1 & a^1_1(1-a^1_1)w^2_1 \\ a^1_2(1-a^1_2)w^2_2 & a^1_2(1-a^1_2)w^2_2 & a^1_2(1-a^1_2)w^2_2 \\ \vdots & \vdots & \vdots \\ a^1_M(1-a^1_M)w^2_M & a^1_M(1-a^1_M)w^2_M & a^1_M(1-a^1_M)w^2_M \end{pmatrix} \otimes Y \right] \tag{14}$$

and

$$\frac{\partial P}{\partial b^1} = -2(t-a)\begin{pmatrix} a^1_1(1-a^1_1)w^2_1 \\ a^1_2(1-a^1_2)w^2_2 \\ \vdots \\ a^1_M(1-a^1_M)w^2_M \end{pmatrix} + \rho d^1 \tag{15}$$

At this point, we have successfully obtained the derivative of expression P with respect to each NN coefficient (w^1, b^1, w^2, and b^2) as shown in eqs 9, 10, 14, and 15. The scale factor ρ in this modified symmetric NN is predetermined as 0.0254, which is different from the value in a previous study.[20] The implemented back-propagation algorithm with modifications for combined-function-derivative approximation[1,20,29] is employed to train the modified symmetric NN on the basis of the computed analytic derivatives. To train the *symmetric* NN, the data points of H_2O nuclear configuration are sampled on the basis of a uniform distribution basis, and MP2/6-31G* calculations are executed to determine the potential energies and gradients.

2. Chlorine Peroxide (ClOOCl) Molecule. The configuration of ClOOCl requires a set of six geometric parameters for molecular definition, which are two Cl−O bonds, one O−O bond, two ClOO bending angles, and a dihedral angle. All of these parameters are denoted as (r_1, r_2, r_3, θ_1, θ_2, ϕ) as shown in Figure 3. Indeed, we use r_1, r_2, r_3, θ_1, θ_2, and $\cos(\phi)$ as the input signals for ClOOCl molecule.

There are two pairs of identical atoms in this molecular structure, i.e., two equivalent Cl atoms and two equivalent O atoms. When we interchange Cl^1 and Cl^4 (and/or O^2 and O^3), the potential energy remains unchanged. In a mathematical context, it is the simultaneous permutations of (r_1, θ_1) and (r_3, θ_2). Therefore, we need to modify the feed-forward NN structure in such a way that provides the mathematical equality $F(r_1,r_2,r_3,\theta_1,\theta_2,\cos(\phi)) = F(r_3,r_2,r_1,\theta_2,\theta_1,\cos(\phi))$.

Prior to the NN training process, the input parameters and energies in our database are all scaled in the range of [0; 1] using scaling expressions similar to that in eq 4. Instead of using the individual maxima and minima of r_1 and r_3 input parameters, the maximum and minimum of (r_1, r_3) are used in the scaling formulas for r_1 and r_3. Similarly, we also scale θ_1 and θ_2 using the maximum and minimum of (θ_1, θ_2). The scaling of all inputs and outputs guarantee that all parameters being processed in the NN are unitless.

The scaled input parameters are denoted as (p_1, p_2, p_3, p_4, p_5, p_6) where p_1 and p_3 represent the scaled value of r_1 and r_3, respectively, p_2 is the scaled value of r_2, p_4 and p_5 are the scaled values of θ_1 and θ_2, respectively, and p_6 represents the scaled value of $\cos(\phi)$. For simplicity, we will discuss our NN structure and training using the scaled input parameters (p_1, p_2, p_3, p_4, p_5, p_6) from this point. It can be easily seen that (p_1, p_3) and (p_4, p_5) are two symmetric pairs of input variables, and the simultaneous interchanges of $p_1 \leftrightarrow p_3$ and $p_4 \leftrightarrow p_5$ do not result in energy change.

The symmetry consideration of the ClOOCl molecule is more complicated than that of H_2O, and the previously proposed NN structure for H_2O cannot be employed in this problem. In fact, it is necessary to propose another modified NN structure that can account the simultaneous interchanges

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

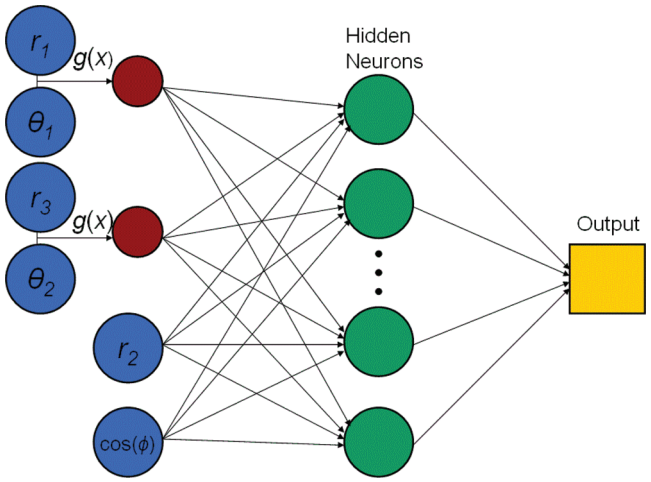


Figure 3. Two-layer feed-forward NN structure with symmetry modifications to deal with the *symmetric* property of the ClOOCl molecule. As illustrated, r_1 and θ_1 are combined as one signal using function $g(x)$, r_2 and θ_2 are combined as one signal also using function $g(x)$, and the switch of these two combined signals (red signals) would not affect the overall output (potential energy).

of $p_1 \leftrightarrow p_3$ and $p_4 \leftrightarrow p_5$. The input signals are processed in the first neural layer to give M output signals as following:

$$a_i = f(n_i) = f[g(w^1_{i,1}p_1 + w^1_{i,4}p_4) + g(w^1_{i,1}p_3 + w^1_{i,4}p_5) + w^1_{i,2}p_2 + w^1_{i,6}p_6 + b^1_i] \quad (16)$$

where $i = 1, ..., M$, matrix w^1 constitutes the first-layer weights and b^1 is the bias vector of the first neural layer. We employ a 55-neuron NN ($M = 55$) to fit the PES in this case. $g(x) = x + \sin(x)$ is again employed as the distinction function, and $f(x)$, a log-sigmoid function, is defined as the transfer function. It should be noted that p_1 and p_3 are connected to the same weight values $w^1_{i,1}$, and p_4 and p_5 are connected to the same weight values $w^1_{i,4}$. Also, function $g(x)$ is employed to distinct (p_1, p_4) and (p_3, p_5) to account for simultaneous *symmetric* interchange (permutation) of these two pairs of input variables. The final NN output is then calculated as previously shown in eq 2.

We also introduce P as the combination of energy and gradient squared errors. As shown in eq 6, the scale factor ρ is utilized to evaluate the significance of six gradients with respect to the input parameters in the fitting scheme. In the ClOOCl case, we again employ eq 8 to determine the value of ρ as 0.0013, which is much smaller than the value in the case of H_2O (0.0254). Using the provided scale coefficient ρ, the derivatives of P with respect to w^1, b^1, w^2, and b^2 can be analytically obtained and used to minimize the deviation of P in the back-propagation algorithm.

In this ClOOCl problem, real energies and gradients are obtained from a previous work[26] using MP2 calculations[30] with the 6-311 g(d,p) basis set.[33] In such work, it was reported that the two reaction channels, Cl−O and O−O dissociations, were very sensitive with the reaction barriers of 0.193 and 0.716 eV, respectively. Consequently, the energy upper limit for ClOOCl PES was selected as 1.200 eV. In this case study, we also look forward to reproducing energies with the same upper limit.

The availability of immediate access to 35 006 configurations in the database allows us to reduce efforts in ab initio calculations. Indeed, we have selected 1693 data points of ClOOCl to construct the training set on the basis of a uniform distribution basis. The maximum and minimum input parameters used in the scaling formulas are shown in Table 1. Subsequently, the back-propagation algorithm is employed to train the NN coefficients to give the best approximating function.[1,20,29]

Table 1. Minimum and Maximum Input Parameters for H_2O and ClOOCl Systems

	H_2O		ClOOCl			
	O−H bond (Å)	H−O−H angle (deg)	Cl−O bond (Å)	O−O bond (Å)	ClOO angle (deg)	$\cos(\phi)$
min	0.781	64.2	1.481	1.048	73.5	−1.00
max	1.293	164.6	2.448	2.823	179.7	1.00
	V_{max} = 1.500		V_{max} = 1.200 eV			

IV. RESULTS AND DISCUSSION

In our back-propagation fitting procedure, there are three sets of data, which include the training, validation, and testing sets. Unlike several earlier studies reported in the literature, in this work, we use a small training set to train the *symmetric* NN without affecting the fitting quality. For the case study of H_2O molecule, we first sample a training set of 191 configurations, a validation set of 180 configurations, and 5612 H_2O configurations constitute the testing set. As mentioned earlier, to construct the PES for H_2O vibrational dynamics, we employ a 25-neuron NN with modifications for symmetry fitting in its structure. The deviations ∂P of three data sets are examined simultaneously during the fitting process. Overfitting, a major concern in many NN studies,[14−19] is handled empirically by monitoring the fitting error of the validation set. If the fitting error of the validation set increases in n (defined by user depending upon the problem of interest) consecutive times, the training process is terminated, and the final NN coefficients and fitting result are reported.

In total, more than 60 000 epochs (fitting iterations) are executed to minimize the mean-squared deviation of P in the H_2O problem. During the training process, we conceive that the mean-squared deviations of three data sets (training, validation, and testing) drop rapidly during the first 400 epochs (as shown in Figure 4), and the dropping process becomes

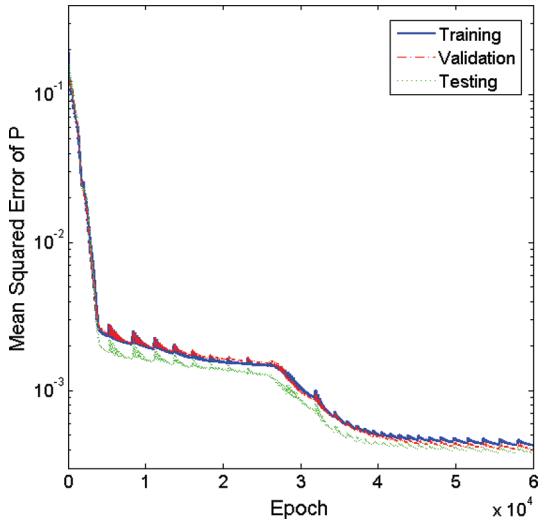


Figure 4. Training, validation, and testing deviations of expression P for H_2O molecule. After 40 000 epochs, the training deviations are stabilized (the deviations of P do not drop significantly for three sets).

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

much slower in the later stage. After 40 000 epochs, the mean-squared deviation of P becomes almost stabilized. The value of mean-squared deviation of P is, however, not meaningful for determination of fitting accuracy. In fact, we evaluate the root-mean-squared errors (rmse) and mean-absolute errors (mae) of energy, which are respectively revealed as 0.0142 eV (0.328 kcal/mol) and 0.0108 eV (0.249 kcal/mol) for the training set, and 0.0141 eV (0.325 kcal/mol) and 0.0107 eV (0.246 kcal/mol) for the testing set when the training progress is terminated. Note that the maximum potential energy for H_2O system is about 1.5 eV. The rmse and mae of energies for the H_2O and ClOOCl cases are summarized in Table 2.

Table 2. Root-Mean-Squared and Mean-Absolute Errors for the Training, Validation, and Testing Sets of H_2O and ClOOCl

		root-mean-squared error		mean-absolute error	
		eV	kcal/mol	eV	kcal/mol
H_2O	Training	0.0106	0.244	0.0079	0.182
	Validation	0.0100	0.231	0.0077	0.178
	Testing	0.0103	0.236	0.0078	0.179
ClOOCl	Training	0.0313	0.722	0.0222	0.512
	Validation	0.0340	0.784	0.0239	0.552
	Testing	0.0409	0.943	0.0269	0.620

Although good fitting accuracy is reported in the H_2O problem, the testing error can be further improved by introducing additional data points into the training set. From the original training set (191 data points), we construct a new training set of 282 data points and perform a new NN fit using the same NN method. Consequently, the fitting accuracy is improved as the rmse and mae for the training set are reported as 0.0106 and 0.0079 eV, respectively, whereas the rmse and mae for the testing set are 0.0103 and 0.0078 eV, respectively. Compared to the previous fitting errors for H_2O, we conceive that the fitting errors for both training (282 data points) and testing sets decrease. Hence, it can be concluded that with an addition with a small number of data points, the accuracy of the *symmetric* NN is better improved.

It is previously stated that the CFDA algorithm is employed to reproduce the PES with high accuracy in terms of numerical error and function curvature. Thus, the fitting errors for gradients are also reported to illustrate the advantage of CFDA-symmetric-NN combination. In the H_2O problem, the testing rmse for the gradient with respect to r_1 (and the equivalent r_2) is 0.335 eV/Å, which is approximately 1.794% of the maximum absolute value of the corresponding force. The relative percent error of gradient with respect to the bending angle θ is about 0.675%, which is better than the prediction of forces with respect to O−H bond. Overall, we still see that those two reported rmse present good gradient prediction by the *symmetric* NN. For convenience, we summarize the force errors for both H_2O and ClOOCl cases and report them in Table 3.

For illustration, a small testing set of 50 configurations is chosen, the NN gradients with respect to r_1 are subsequently computed and compared to the corresponding gradients resulted from MP2 calculations. The plot of this comparison is shown in Figure 5.

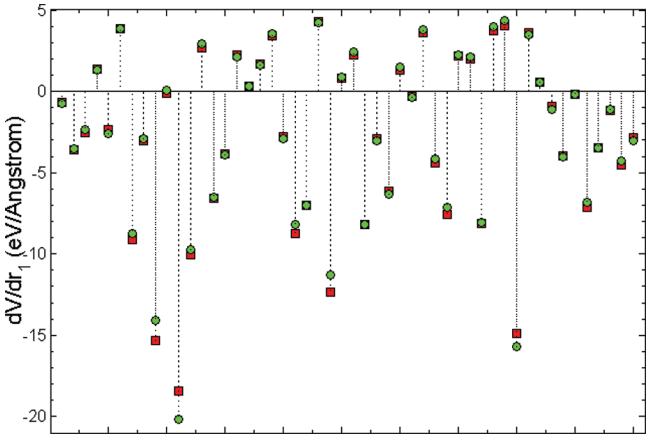


Figure 5. Analysis of NN and MP2 gradients with respect to r_1 for the H_2O case. This analysis is conducted on a small testing set of 50 configurations.

The error of validation set is examined to prevent overfitting. However, it is empirically observed in this study that the use of a validation set is unnecessary in a CFDA fitting scheme. When derivative fitting is incorporated in the fitting process, our *symmetric* NN follows the potential energy function curvature and prevents inappropriate variations of the derivatives. Thus, the CFDA algorithm would automatically prevent overfitting. As shown in Figure 4, the validation error drops consistently with the training error during the training process, and we do not observe epochs at which the validation error rapidly increases.

We should imply that the testing errors of H_2O energies and gradients reported in Table 3 are both conducted on a set of 5612 configurations, which is much larger than the training set (282 configurations). On the basis of those reported results, we consequently conclude that excellent fitting accuracy is obtained when only 282 configurations are employed to train the PES NN function. As a result, it can be concluded with certainty that the modified NN structure provides high accuracy (relatively small fitting errors for energies and gradients when we perform error evaluation on a large testing set) and statistical consistency (the fitting errors of testing set drops consistently in accordance with fitting error of the training set during the training process, as shown in Figure 4) for the PES of H_2O. For the simple case of H_2O vibrational PES, we can conclude that our modified NN construction and function-derivative-approximation back-propagation fitting algorithm is highly advantageous.

Not only enhancing fitting accuracy and handling *symmetric* properties, the improvement of the NN method in this work

Table 3. Root-Mean-Squared Testing Errors for Gradients of H_2O and ClOOCl Systems

	H_2O		ClOOCl			
gradient with respect to	r_1 (eV/A)	θ (eV/rad)	r_1 (eV/A)	r_2 (eV/A)	θ (eV/rad)	ϕ (eV/rad)
maxlforcel	18.678	4.299	11.731	24.427	5.708	5.043
rms error	0.335	0.029	0.368	0.581	0.246	0.174
percent error (%)	1.794	0.675	3.137	2.379	4.310	3.450

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

also includes reducing the number of required fitting data points by implementing CFDA algorithm for the modified NN structure. As we mention earlier in this paper, for three-atom PESs (like SiO_2 and O_3), about 6000 nuclear configurations are required to produce NN fits.[14,19] The PES construction for H_2O vibrational dynamics, on the other hand, requires only 282 training data points to produce an analytic function with excellent fitting accuracy (with the upper-limit energy being 1.5 eV). Although the levels of complexity for reactive PESs (for SiO_2 and O_3 systems) are higher than that for a non-reactive PES (for H_2O vibrational dynamics in this case study), the effort of reducing data points is noticeable when only 282 data points are required to constitute the training set instead of thousands of data points if we approach the H_2O problem using the traditional feed-forward NN structure without implementing the modified CFDA algorithm.

The construction of ClOOCl reactive PES requires a lot of computational efforts on data fitting and ab initio calculations for more than 35 000 data points in a recently reported study.[26] The objective in this work is to reconstruct the reactive NN PES using fewer training data points without affecting the fitting quality.

Instead of fitting more than 35 000 configurations, we randomly select data points from the available database on the basis of a uniform distribution basis. The training data set is then constituted by choosing 1693 configurations, and the CFDA procedure is subsequently executed for more than 60 000 epochs (illustrated in Figure 6). In the ClOOCl case,

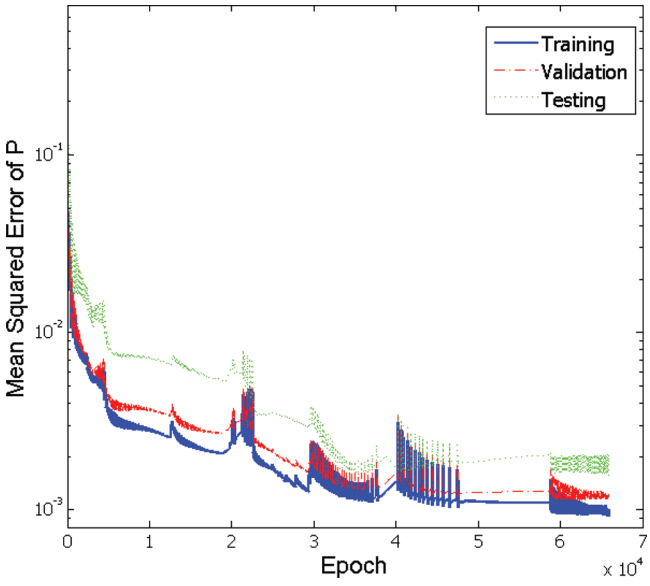


Figure 6. Training, validation, and testing deviations of expression P for ClOOCl molecule. After more than 60 000 epochs, the training process reaches convergence and is terminated.

we also choose a large number of configurations (15 374 configurations) to construct the testing set.

When the training process is terminated, we obtain the rmse and mae of energy for the training set as 0.0313 eV (0.722 kcal/mol) and 0.0222 eV (0.512 kcal/mol), respectively. For the large testing set (15 374 configurations), the rmse and mae are respectively reported as 0.0409 eV (0.943 kcal/mol) and 0.0269 eV (0.620 kcal/mol), which are highly consistent with the errors reported for the training set and reveals excellent statistical fitting quality. The gradient errors are analyzed on the testing set of 15 374 configurations, and we conceive good agreements between MP2 gradients and NN gradients for the ClOOCl case, as shown in Table 3.

Overall, we have developed a reactive PES for ClOOCl dissociation using a *symmetric* NN with high fitting accuracy for both energies and gradients. The fitting accuracy in this work is somewhat lower than that reported in a previous work (with the errors being almost twice), but still, the magnitude of errors is in an acceptable range, and can be employed for molecular dynamics simulations. When this level of accuracy is compared to the fitting quality reported in other NN PES works in the literature,[14−19] it can be revealed that we have obtained an excellent NN fit by only fitting 1693 data points in a six-dimensional configuration hyperspace. The distribution of energy errors for the testing set is presented in Figure 7, which

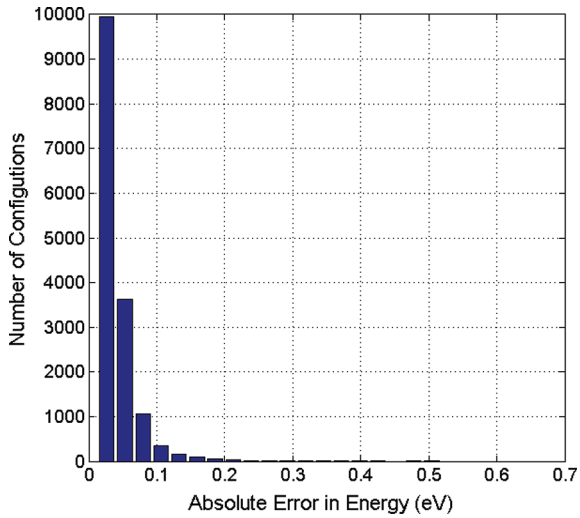


Figure 7. Distribution of energy errors for the ClOOCl NN PES. This distribution plot is made by examining the absolute errors of the testing set.

shows a domination of small fitting errors and provides a valid ground for fitting consistency.

The testing errors in this study are not high compared to the reported fitting errors previously reported in the literature. Although improving fitting quality is a major concern in most PES studies, we would like to imply that the main objective in this research is to produce *symmetric* NN fits with appropriate accuracy using smaller numbers of fitting parameters. To effectively improve the fitting accuracy, we believe that the distinction function needs to be re-examined more carefully. For illustrative purposes, we simply use $g(x) = k_1x + k_2 \sin(x)$ with k_1 and k_2 being unity in this work.

From the two testing cases, we have clearly demonstrated two advantages of the new *symmetric* neural network model, which are reducing the number of NN weight parameters in the first neural layer (because of symmetry incorporating for input variables) and reducing the number of training data points (as discussed above). We believe that our *symmetric* NN is a good alternative option that deals with molecules with *symmetric* variables. Compared to the existing methods that deal with symmetry, in our NN structure, it is neither obligated to perform permutation to every *symmetric* variable pair[23] nor required symmetry functions to initially process all input signals.[22,24,25] One disadvantage is also pointed out from the presented NN structures in this study; i.e., the *symmetric* NN

dx.doi.org/10.1021/jp3020386 | J. Phys. Chem. A 2012, 116, 4629−4638

structures are somewhat difficult to generalize for all molecules, and it requires significant efforts to design a *symmetric* NN model for the particular molecule of interest.

In fact, reducing the numbers of NN fitting parameters is a nice effort that is promising in fitting larger molecules with higher complexity, such as DNA fragments and polypeptides. However, the potential complications of designing the *symmetric* NNs for those cases are some major difficulties that we must take into account. Therefore, the potential applications of this *symmetric* NN model in DNA fragments and polypeptides fitting remain questionable. Such concerns can be better explained after the *symmetric* NN models presented in this study are tested with systems with lower complexity, such as bulk periodic materials in the condensed phase.

In conclusion, we have developed modifications to NN structures that can be successfully employed to treat molecules that have *symmetric* configuration input variables, and we also implement CFDA back-propagation algorithm[1,20] for symmetric NN fitting. The NN modifications in this study have been successfully utilized to deal with two different molecules (H_2O and ClOOCl) with very good accuracy in term of both energies and gradients.

V. SUMMARY AND CONCLUSIONS

In this work, we have successfully developed new NN designations that effectively deal with molecules in which multiple pairs of *symmetric* variable may be interchanged without affecting the potential energy. The major modifications are made in the first layer of two-layer feed-forward NNs. Such modifications are capable of handling the interchange of input variables (configuration parameter permutation) effectively. Moreover, the CFDA fitting algorithm is implemented to train the modified NNs with very good fitting accuracy.

Two molecules are tested in this study, which include water monomer (H_2O) and chlorine peroxide (ClOOCl). For H_2O, we only consider constructing a PES that sufficiently describes its vibrational dynamics. ClOOCl, on the other hand, is tested using an existing configuration database in a previous study of ClOOCl dissociation.[26] After about 60 000 epochs, the training process for H_2O NN PES is terminated, and it is reported that the rmse and mae for H_2O training set are 0.0106 eV (0.244 kcal/mol) and 0.0079 eV (0.182 kcal/mol), respectively. For the large testing set (5612 configurations), the rmse and mae are 0.0103 eV (0.231 kcal/mol) and 0.0078 eV (0.179 kcal/mol), respectively.

The training process for ClOOCl reactive PES requires more computational efforts to minimize the energy and gradient errors. At termination, the rmse and mae for the training set are reported as 0.0313 eV (0.722 kcal/mol) and 0.0222 eV (0.512 kcal/mol), respectively. The rmse and mae for ClOOCl testing set are respectively reported as 0.0409 eV (0.943 kcal/mol) and 0.0269 eV (0.620 kcal/mol). In both case studies, we witness good statistical agreement between the training and testing errors (recall that in both cases, we have selected large numbers of data points to construct the testing sets). Those reported errors (for energies and gradients) reveal excellent fitting quality among the gas-phase NN PESs that have been developed.[14−20]

The training sets for two illustrating problems are constituted with noticeably few numbers of nuclear configurations (282 points for H_2O and 1693 points for ClOOCl). The NN improvements in this study (*symmetric* modification and combined-function-derivative fitting) are believed to better interpolate data points in multidimensional hyperspaces and significantly contribute to the small numbers of required training data points, which is a noticeable advantage. The NN structure designed in this study is potentially capable of describing more complex molecules, such as bulk periodic materials. It is in our future objective to use such *symmetric* NN models to deal with more-complex molecules with higher level of symmetry (multiple pairs of input variables being permutated) and evaluate its capability in biological molecular applications.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The training codes in Matlab for two *symmetric* NNs are provided together with the training, validation, and testing data for both H_2O and ClOOCl (in ascii). All files are contained in one compressed file. This information is available for free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-lectronic mail: hung.m.le@hotmail.com. Phone: 84 838350831.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Faculty of Materials Science, College of Science, Vietnam National University, for their computing supports.

■ REFERENCES

(1) Hagan, M. T.; Demuth, H. B.; Beal, M. H. *Neural Network Design*; Colorado Bookstore: Boulder, CO, 1996.

(2) Hopfield, J. J. *Proc. Nl. Acad. Sci.* **1982**, *79*, 2554−2558.

(3) (a) Burns, J. A.; Whitesides, G. M. *Chem. Rev.* **1993**, *93*, 2583−2601. (b) Zupan, J.; Novič, M.; Ruisánchez, I. *Chemometr. Intell. Lab* **1997**, *38*, 1−23. (c) Zupan, J.; Gasteiger, J. *Anal. Chim. Acta* **1991**, *248*, 1−30. (d) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Paša-Tolić, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y.; Zhao, R.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 1039−1048. (e) Khan, J.; Wei, J. S.; Ringner, M.; Saal, L. H.; Ladanyi, M.; Westermann, F.; Berthold, F.; Schwab, M.; Antonescu, C. R.; Peterson, C.; Meltzer, P. S. *Nat. Med.* **2001**, *7*, 673−679. (f) So, S.-S.; Karplus, M. *J. Med. Chem.* **1996**, *39*, 1521−1530.

(4) Gasteiger, J.; Zupan, J. *Angew. Chem., Int. Ed.* **1993**, *32*, 503−527.

(5) (a) Behler, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930−17955. (b) Handley, C. M.; Popelier, P. L. A. *J. Phys. Chem. A* **2010**, *114*, 3371−3383.

(6) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(7) (a) Lorenz, S.; Groß, A.; Scheffler, M. *Chem. Phys. Lett.* **2004**, *395* (4−6), 210−215. (b) Lorenz, S.; Scheffler, M.; Gross, A. *Phys. Rev. B* **2006**, *73*, 115431.

(8) (a) Behler, J.; Delley, B.; Lorenz, S.; Reuter, K.; Scheffler, M. *Phys. Rev. Lett.* **2005**, *94*, 036104. (b) Behler, J.; Reuter, K.; Scheffler, M. *Phys. Rev. B* **2008**, *77*, 115421.

(9) (a) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 185501. (b) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. *Phys. Status Solidi B* **2008**, *245*, 2618−2629.

(10) Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. *Phys. Rev. B* **2010**, *81*, 184107.

(11) Artrith, N.; Morawietz, T.; Behler, J. *Phys. Rev. B* **2011**, *83*, 153101.

(12) Morawietz, T.; Sharma, V.; Behler, J. *J. Chem. Phys.* **2012**, *136*, 064103.

(13) Prudente, F. V.; Soares Neto, J. J. *Chem. Phys. Lett.* **1998**, *287*, 585−589.
(14) Agrawal, P. M.; Raff, L. M.; Hagan, M. T.; Komanduri, R. *J. Chem. Phys.* **2006**, *124*, 134306.
(15) (a) Malshe, M.; Raff, L. M.; Rockley, M. G.; Hagan, M.; Agrawal, P. M.; Komanduri, R. *J. Chem. Phys.* **2007**, *127*, 134105. (b) Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. *J. Chem. Phys.* **2005**, *122*, 084104. (c) Manzhos, S.; Carrington, T. J. *J. Chem. Phys.* **2008**, *129*, 224104.
(16) Le, H. M.; Raff, L. M. *J. Chem. Phys.* **2008**, *128*, 194310.
(17) Le, H. M.; Huynh, S.; Raff, L. M. *J. Chem. Phys.* **2009**, *131*, 014107.
(18) Le, H. M.; Raff, L. M. *J. Phys. Chem. A* **2009**, *114*, 45−53.
(19) Le, H. M.; Dinh, T. S.; Le, H. V. *J. Phys. Chem. A* **2011**, *115*, 10862−10870.
(20) Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. *J. Chem. Phys.* **2009**, *130*, 134101.
(21) Mai-Duy, N.; Tran-Cong, T. *Appl. Math. Model.* **2003**, *27*, 197−220.
(22) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. *J. Phys. Chem. A* **1998**, *102*, 4596−4605.
(23) Prudente, F. V.; Acioli, P. H.; Neto, J. J. S. *J. Chem. Phys.* **1998**, *109*, 8801−8808.
(24) Lorenz, S.; Groß, A.; Scheffler, M. *Chem. Phys. Lett.* **2004**, *395*, 210−215.
(25) Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(26) Le, A. T. H.; Vu, N. H.; Dinh, T. S.; Cao, T. M.; Le, H. M. *Theor. Chem. Acc.* **2012**, *131*, 1158.
(27) (a) Brandao, J.; Mogo, C.; Silva, B. C. *J. Chem. Phys.* **2004**, *121*, 8861−8868. (b) Brandao, J.; Rio, C. M. A. *J. Chem. Phys.* **2003**, *119*, 3148−3159.
(28) (a) Avallone, L. M.; Toohey, D. W. *J. Geophys. Res.* **2001**, *106*, 10411−10421. (b) Huang, W.-T.; Chen, A. F.; Chen, I. C.; Tsai, C.-H.; Lin, J. J.-M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8195−8203. (c) Stimpfle, R. M.; Wilmouth, D. M.; Salawitch, R. J.; Anderson, J. G. *J. Geophys. Res.* **2004**, *109*, D03301.
(29) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *323*, 533−536.
(30) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.
(31) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comput. Chem.* **2001**, *22*, 976−984.
(32) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; et al. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(33) (a) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639−5648. (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.