A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits†

Sergei Manzhos,‡ Xiaogang Wang,§ Richard Dawes,⊥ and Tucker Carrington Jr.*

Département de chimie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada

Received: September 16, 2005; In Final Form: November 7, 2005

It is shown that neural networks (NNs) are efficient and effective tools for fitting potential energy surfaces. For H_2O, a simple NN approach works very well. To fit surfaces for HOOH and H_2CO, we develop a nested neural network technique in which we first fit an approximate NN potential and then use another NN to fit the difference of the true potential and the approximate potential. The root-mean-square error (RMSE) of the H_2O surface is 1 cm^{-1}. For the 6-D HOOH and H_2CO surfaces, the nested approach does almost as well attaining a RMSE of 2 cm^{-1}. The quality of the NN surfaces is verified by calculating vibrational spectra. For all three molecules, most of the low-lying levels are within 1 cm^{-1} of the exact results. On the basis of these results, we propose that the nested NN approach be considered a method of choice for both simple potentials, for which it is relatively easy to guess a good fitting function, and complicated (e.g., double well) potentials for which it is much harder to deduce an appropriate fitting function. The number of fitting parameters is only moderately larger for the 6-D than for the 3-D potentials, and for all three molecules, decreasing the desired RMSE increases only slightly the number of required fitting parameters (nodes). NN methods, and in particular the nested approach we propose, should be good universal potential fitting tools.

I. Introduction

To understand ro-vibrational spectra, intramolecular relaxation processes, chemical reactions, etc., theorists use potential energy surfaces (PESs).[1−8] Spectra, cross sections, and rate constants are computed by solving classical or quantum equations for the motion of the nuclei on a PES. The PES is a consequence of the Born−Oppenheimer approximation. To calculate a PES, one must solve the electronic Schroedinger equation for a large set of geometries of the molecule or reacting system. A great deal of effort has been devoted to the development of good (accurate and efficient) quantum chemistry and dynamics methods, but to compute a spectrum, a rate constant etc., it is not enough to be able to compute electronic energies at selected points: one needs a PES.

The generation of PESs from ab initio data is the middle ground between quantum chemistry and dynamics. When classical mechanics is adequate for the purpose of studying the motion of the nuclei, it is possible to use "direct dynamics": the potential is computed at points on a classical trajectory and no potential function is necessary.[9−12] This is wasteful, because information is thrown away as it is acquired. It seems obvious that it would be good to retain and exploit potential surface information as it is obtained. If one uses direct dynamics, the quality of the results is determined by the quantum chemistry method used, and it is not possible to use high level methods due to the number of required calculations. To use quantum mechanics to compute a spectrum, a cross section etc., one must know the potential at geometries that correspond to the quadrature (or discrete variable representation (DVR))[13,14] points required to obtain converged results. It is usually not possible to compute the electronic energy at all of the quadrature points (optimal points are not known before the surface is obtained), and it is therefore important to find a function that reproduces calculated electronic energies well.

PESs are sometimes obtained by choosing the parameters of a fitting function so that the function very nearly reproduces a set of ab initio data points.[15−17] This works well for many molecules, but it has several important disadvantages. First, to develop a good function, one needs much experience and intuition. The functional form reflects the nature of the most important interactions.[18−21] In practice, most potentials generated in this way are the work of a few groups that have invested years of experience. Using a physically motivated function has the potential advantage of reducing the number of required parameters, but it also means that each potential is a new project. The functions one would use for a semirigid molecule such as H_2O or a double-well molecule like H_2O_2 would be rather different. Second, systematically improving the potential by adding more parameters is not always easy. Third, correlation between the parameters can plague the fitting process.

Black-box potential fitting routines that make no attempt to exploit knowledge of the important interactions and coupling terms greatly simplify obtaining potentials. Fitting procedures that are not physically motivated have the virtue of simplicity. The most popular procedures of this type are spline,[22−25] interpolating moving least squares (IMLS),[26,27] reproducing kernel Hilbert space (RKHS),[28,29] modified Sheppard interpolation (MSI),[30−31] distributed approximating functionals (DAF),[34−37] and neural network (NN) algorithms.[38−45] These approaches are systematically improvable, work well even if coupling is large, and they are easy to use; e.g., most parameter values are completely determined by the numerical algorithm and do not need to be estimated on the basis of experimental results or intuition. However, in all of these methods, there are parameters that must

† Part of the special issue "John C. Light Festschrift".

* Corresponding author. E-mail: Tucker.Carrington@umontreal.ca. Fax: 514-343-7586.

‡ E-mail: Sergei.Manzhos@umontreal.ca.

§ E-mail: Xiaogang.Wang@umontreal.ca.

⊥ E-mail: r.dawes@umontreal.ca

10.1021/jp055253z CCC: $33.50
Published on Web 12/16/2005

be chosen by the user. The popular Taylor-expansion based MSI method requires potential derivatives and is used almost exclusively with second-order expansions, as higher derivatives are difficult to obtain from ab initio calculations. The RKHS method is easiest to use with tensor product grids[28,29,46] which make it difficult to fit surfaces for systems with more than three atoms.[34,36,47−50] With a trajectory-based point selection scheme, the MSI approach can fit dynamically relevant parts of PESs of four-atom molecules[32,51−54] and some five- and six-atom systems.[55−57] For the cited four-atom systems, the potential fitting errors of these methods are of the order of 10^1-10^2 cm^{-1}. Similar errors are obtained with physically motivated fitting functions.[18,19,58] Recently, permutation invariant polynomials and interatomic distances have been used to fit several potentials. See, for example, ref 59. This approach appears very promising.

In this paper, we present improved NN methods for fitting potentials. The ease of use of NN fitting methods is independent of the point distribution: there is no reason to use a tensor product grid. The NN procedure employs an analytic expression that depends on parameters.[60,61] Parameter values are chosen so that (1) potential values at a set of fitting points are nearly reproduced and (2) the NN PES gives reasonable values at other points (not those used to fit). Due to the fact that parameters are chosen, it is legitimate to describe the NN procedure as a fitting algorithm. However, if one uses NNs, it is particularly easy to vary the number of parameters as the fit is improved. NNs have all of the advantages of the black-box methods mentioned in the previous paragraph. To use them, one need not have a priori knowledge of the potential shape, and it is not necessary to have potential derivatives. In addition, the NN fitted potential is inexpensive to evaluate, which could be important, e.g., if it is used with classical trajectories. NNs are universal fitting functions.[62,63] In the NN literature,[60,64,65] the process of finding the best parameter values is called learning. One begins with a set of coordinate values and their associated energies and chooses weights and biases (fitting parameters) to get a good fit. NN have been used for curve fitting[66] and, in particular for PES fitting,[38−45] but also for modeling kinetic equations[67] and interpolation of solutions of the Schroedinger equation.[68−70] The usefulness of NNs is, of course, not limited to fitting.[71−74] In section II, we briefly explain the standard NN approach and discuss some of its advantages.

We use NNs to fit potentials for H_2O (3-D), HOOH (6-D), and H_2CO (6-D). The method (see section III) we use for choosing data points has not yet been applied to potential fitting. For H_2O, a straightforward application of NN fitting software (we use Matlab) works very well.[75] To make good NN fitting functions for HOOH and H_2CO, we develop a hierarchical NN fitting procedure. This is explained in subsection IIIb. Most previous applications of NN fitting methods are for systems with fewer than 4 atoms and achieve only moderate accuracy. We demonstrate that it is possible to obtain accurate fits. The root-mean-square error (RMSE) of our H_2O fit is 1 cm^{-1}. For HOOH and H_2CO, we achieve a 2 cm^{-1} RMSE. Some discussions of NNs give the impression that they work well only if low accuracy is sufficient.[64,65,76] Most previous NN fits have RMSEs larger than 10 cm^{-1}. For example: a RMSE of 15−26 cm^{-1} for H_3^+ potentials;[41] a RMSE of 878 cm^{-1} for an intermolecular potential of $H_2O-Al^{3+}-H_2O$;[40] a RMSE of about 8 cm^{-1} (actually a mean absolute deviation of 7.7 cm^{-1}) for a 3-D slice of a CO/Ni(111) potential;[44] a RMSE of 770 cm^{-1} for a 12-D potential for H_2 on Si(100);[44] a RMSE of 7.7 cm^{-1} for regions of a potential surface of vinyl bromide required to describe dissociation;[38] a RMSE of 30 cm^{-1} for a potential surface of Si_5.;[38] a RMSE of 25 cm^{-1} for a 4-D model of a $(HF)_2$ potential below 5000 cm^{-1}.[43] Our 6-D NN-based potentials reported below have a RMSE of 2 cm^{-1} and extend up to more than 11000 cm^{-1} above the zero point energy (ZPE). Our RMSEs are also smaller than those typically obtained with methods such as IMLS, MSI, RKHS.

We test all our fits by computing vibrational spectra. This is a good way of detecting "holes"—isolated regions of the PES usually far from the equilibrium geometry or saddle points, where the surface is very sparsely sampled and where the fitted potential has large errors. It is somewhat dangerous to look only at 2-D potential cuts[40,41,43] or 1-D equilibrium slices.[38] In more than 3-D, we find that the standard NN method often produces a surface with holes. It is possible that some published NN potentials had holes or other unphysical regions that were not detected. Computing spectra also provides an unambiguous test of the quality of our fitted surfaces. NN fits have been used to compute a spectrum in only one other paper (for a three-atom molecule).[41] It is important to know how good the RMSE must be in order to give an accurate spectrum. Because we do fits for three different molecules with different levels of precision, we can study how the complexity of the fit (number of parameters) that yields a potential depends on the number of coordinates and on the required precision.

II. Neural Networks

A NN consists of nonlinear processing "nodes" organized into "layers" .[60,61,64] NNs can be used as fitting functions. The commonly used[38−45] single-hidden-layer NN fitting function can be written

$$V(x) = \sum_{p=1}^{n} c_p (1 + e^{\mathbf{w_p}\cdot\mathbf{x}+b_p})^{-1} \qquad (1)$$

The components of **x** are the coordinates of our fitting points. The $\mathbf{w_p}$ are weights and the b_p are called biases. A diagram of the NN that corresponds to eq 1 is shown in Figure 1. The neurons are functions of linear combinations of the coordinate (input) values. The neurons we use are sigmoid functions. The final output of the NN is a linear combination of the sigmoids. This is a NN with one layer of neurons. The layer of neurons is usually called a hidden layer.[60,64,65] NNs can have an arbitrary number of hidden layers.[60−65,77−79] The weights and biases are chosen so that the outputs are very close to the correct potential values. The NN we use is denoted a feed-forward network with a ridge activation function in the NN literature.[78] Other groups using a NN to fit potentials have also used this kind of NN.[38−45]

We choose to use NNs for two reasons. First, there are theorems proving that with the right choice of the **w**, **b**, and **c** parameters, it is possible to fit any function to arbitrary accuracy with a sigmoid NN with one hidden layer.[62,63] We are therefore certain that as we increase the number of fitting points and neurons (and therefore weight and bias parameters), the quality of our fit should improve. To demonstrate that a NN can be used to fit any function, one uses results from the field of functional representation theory,[77,78] and in particular, theorems of Kolmogorov[80] and Sprecher.[81,82] According to these theorems, it is always possible to represent a multidimensional continuous function of **x** in terms of a nonlinear function of a linear combination of continuous 1-D functions of new variables obtained from the **x**. Using these results and the idea that any continuous 1-D function can be expanded in terms of sigmoid functions,[83] one can prove that a NN can be used to fit any function. Second, mathematicians[78,84−86] argue that the number

of NN parameters required to obtain a good fit scales well as dimensionality is increased. There are theoretical estimates of upper bounds on the number of nodes required to fit particular classes of functions to a predefined accuracy.[78,79,87] Although the upper bounds are large, it seems clear that the number of required NN parameters does not scale as badly as the number of parameters of a direct product fit. In practice, modestly sized NNs have been used to fit diverse potentials.[39,44,88] An important advantage of NNs is the fact that although they make use of 1-D functions to fit, they do not fit with sums of products of functions. They therefore scale less badly with dimension than many fitting methods.

III. Using NNs To Fit Spectroscopically Relevant Parts of Potentials

In this section, we discuss the ideas we have used to obtain NN potentials for H_2O, HOOH, and H_2CO. The three potentials are very different (H_2O is quite rigid, HOOH is nonplanar and has a PES with two accessible minima separated by a tunneling barrier, and H_2CO is planar and has a single-well PES), but the NN procedures work well in all cases. Of course, the 6-D problems are harder. Although the real goal of NN potential fitting methods must be obtaining a potential from a set of ab initio points, in this paper, we develop and test NN methods by fitting a reference potential. We use the H_2O potential of ref 89, the HOOH potential of ref 18, and the H_2CO potential of ref 90. We do this so as to be able to assess directly the error introduced by the fitting procedure. Had we used ab initio energies, both the fitting error and the error of the ab initio method used would have influenced our results. This is a common strategy.[26,27,35,37,91−96] Because we start from a reference potential, it is possible to measure the effect of the fitting error on energy levels.

The first thing to do is to select a scheme for choosing points. We use the same scheme for all three molecules. We want a scheme that will put points in regions of the potential that are important for the purpose of computing low-lying vibrational energy levels. The lowest levels depend only on the shape of the bottom of the potential, and higher levels are, in general, also influenced by the shape of the bottom of the potential. We choose points from a distribution function proportional to $E_{cut} - V^0(\boldsymbol{x}) + \Delta$. The minimum value of all our potentials is zero. The same idea has been used previously for choosing centers of Gaussian basis functions.[97] In practice, we randomly sample configuration space in the smallest multidimensional box that encloses points with energies less than E_{cut} and retain a point $\boldsymbol{x_i}$ if $V^0(x_i) < E_{cut}$, where E_{cut} is a cutoff energy, and if

$$\frac{E_{cut} - V^0(\boldsymbol{x_i}) + \Delta}{E_{cut}+\Delta} > b_i \qquad (2)$$

where $b_i \in [0, 1]$ is a random number. The probability of accepting a point decreases as the energy increases and takes a nonzero value proportional to $\Delta \ll E_{cut}$ at E_{cut}. $V^0(\boldsymbol{x})$ is a zeroth-order separable potential. For $V^0(\boldsymbol{x})$, we use a sum of equilibrium slices. For each coordinate, we generate a set of points by fixing the other coordinates at their equilibrium values and taking 20 equally spaced values (one of which is the equilibrium value) for the coordinate in question. For each of the slice potentials, $V_{1D} \le E_{max}$. E_{max} is larger than the largest energy level we wish to compute and is the energy above which we do not wish to fit. In general, E_{max} and E_{cut} are close and $E_{max} \ge E_{cut}$. We use Excel to do the 1-D fits. Of course, if a NN were being used to fit ab initio data, it would be necessary to compute 1-D ab initio potentials, but for any molecule, this is easy and inexpensive. Instead of a separable V^0, it would also be possible to use any crude fit (perhaps an NN fit) for a predefined (sparse) grid of points. We find that the results we obtain do not depend sensitively on our choice of $V^0(\boldsymbol{x})$. With this point distribution scheme, the number of points per unit volume of configuration space is higher at low energies. There are many other ways to choose points. A widely used alternative is to choose points along classical trajectories.[32,38,96,98−101] We have also experimented with this idea but find the method of eq 2 simpler and good for our (spectroscopic) purposes. The points used to build the slice potentials are added to the points chosen according to eq 2 to make the fitting set. Outside the smallest multidimensional box which encloses all points with energy less than $E_{max} \ge E_{cut}$, we set the potential equal to a plateau value much larger than the energy of the fitting point with the largest energy. Inside this box, the potential is set equal to the value returned by the NN.

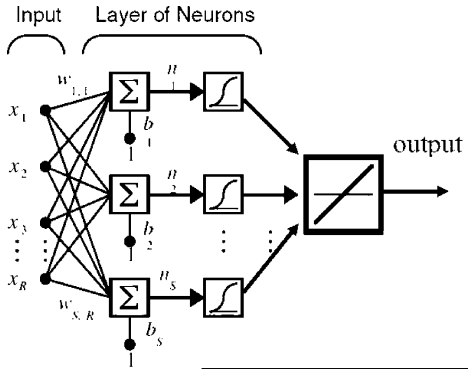

Figure 1. Classic single-hidden-layer back-propagation neural network used to fit the potential surfaces. The connection weights matrix **w** is used to form linear combinations of the input variables **x** via $\mathbf{w}\,\mathbf{x} + \mathbf{b}$. These are indicated by Σ in the figure and are arguments of the sigmoid nodes or neurons. A linear combination of the nodes outputs is the final network output.

In all cases, we use a NN with one hidden-layer as shown in Figure 1. The coordinate and energy ranges are scaled to [-1, 1].[102] The code is written in MatLab using MatLab's Neural Network Toolbox.[75] In principle, it is possible to use it to fit a potential of any dimensionality. The number of nodes, and hence the number of fitting parameters, is adjustable. In NN parlance, determining the fitting parameters is called training the network. We train the network using the Levenberg−Marquardt (LM) algorithm. After experimentation with different training methods including "resilient back-propagation", steepest descent, conjugate gradient, and quasi- Newton algorithms (all implemented in MatLab[75]) and genetic algorithms[103−106] (both hand-written and from MatLab[107]), we concluded that LM both converged most quickly and produced the best fit. The number of nodes is adjusted to achieve a fit with a predefined precision. If the number of nodes is too small, it is not possible to obtain a good fit. If the number of nodes is too large, there is a danger of "overfitting", i.e., of obtaining a function that does an excellent job for the points to which it is fit but, nevertheless, does not represent well the shape of the true potential (has spurious oscillations, for example). To prevent overfitting, we use early stopping. To this end, a second (independent) "validation" set of points was built using the same point selection scheme described above. The number of validation points was half the number of training points. The training algorithm returns the error on both sets of points. We accept a fit if the RMSE is less than the predefined limit for the training set and less than twice the same predefined limit for the validation set. All fits are initiated with random initial weights and biases. If the validation

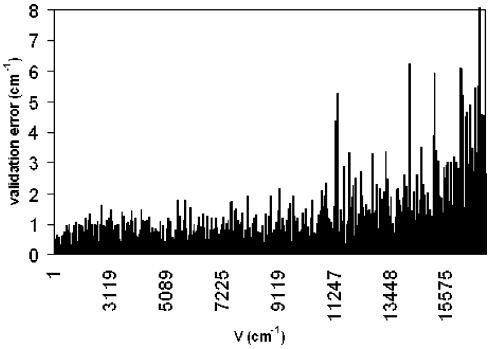

Figure 2. Absolute difference between the NN potential for H_2O fitted to 1 cm^{-1} RMSE and the Jensen potential[89] as a function of energy for the *validation* points.

set error begins to increase, we abort the fit and start again with a new set of random initial weights and biases. The fitted potential we obtain is symmetrized so that it is invariant under permutation (of the hydrogen atoms in H_2CO and H_2O and of the OHs in HOOH) Symmetrically equivalent points with respect to exchange are also added to the training and validation sets.

A. Water. It was easy to obtain a very good potential for H_2O. This experience leads us to believe that NNs provide an excellent black-box method for obtaining very good surfaces for semirigid three-atom molecules. NNs could be considered a method of first resort for such problems. We used Radau coordinates,[108] but there is no reason to believe that the NN surface obtained with any other set of coordinates would not be as good. Radau coordinates were used simply because it is easier to compute vibrational levels using Radau coordinates (there is no kinetic cross-term). As the reference potential, we selected the Jensen potential of ref 89. 1500 points were selected with $E_{cut} = 20\,000$ cm^{-1} and $\Delta = 1000$ cm^{-1}. Another 1500 points obtained by exchanging the two H atoms were added to the fitting set. The V^0 used with eq 2 is a sum of two Morse oscillators for the two stretches and a fourth order polynomial for the bend. The Morse parameters and the bend polynomial coefficients are obtained from simple least-squares fits. We use 20 points for each coordinate and $E_{max} = 20\,000$ cm^{-1}; with about 20 points, the V^0 parameters are almost independent of the actual number of points. The points along the 1-D slices are also included in the set of points used to fit the NN parameters. To test the NN algorithm and the cost of reducing the RMSE, we obtained one surface with a RMSE of 5 cm^{-1} and another with a RMSE of 1 cm^{-1}. It is important to have some measure of the cost of reducing the predefined required accuracy. If the predefined required accuracy is 5 cm^{-1}, only 17 nodes (corresponding to 3×17 input weights + 17 hidden layer biases + 17 output weights + 1 output bias = 86 fitting parameters) are necessary. Decreasing the predefined required accuracy to 1 cm^{-1} increases the number of nodes to 23 (116 fitting parameters). The largest errors for the points in the *validation* set are 32 and 8 cm^{-1}, for the 5 and 1 cm^{-1} RMSE fits, respectively. The majority of the *validation* points are reproduced to better than 2.5 and 0.5 cm^{-1} for the two fits. For the training set, the agreement is, obviously, even better.

In Figure 2, we show, for the 1 cm^{-1} RMSE fit, the error in the potential at the *validation* points. The corresponding plot for the 5 cm^{-1} RMSE fit has the same shape, but the errors are larger. We note that the bottom of the potential is best reproduced, and that the error grows consistently with energy; it only becomes noticeably larger than the target RMSE close to the highest energies. The NN fitting procedure works extremely well in this case. It is possible to obtain fits for which the validation set errors are about the same size as the fitting set errors. Significantly decreasing the target RMSE causes only a modest increase in the size of the network.

TABLE 1: Properties of Absolute Differences between Levels Computed on NN Potentials and Levels Computed on the Jensen Potential[a]

	error, cm^{-1}		
	NN 5 cm^{-1}	NN 1 cm^{-1}	Jensen[89]
mean absolute	0.681	0.113	2.109
median	0.561	0.071	1.377
minimum	0.000	0.004	0.005
maximum	4.212	1.180	9.785
no. of nodes	17	23	

[a]The errors for the 1 cm^{-1} RMSE fit are shown in Figure 3 as a function of energy. Deviations of the levels computed on Jensen's surface from their experimental counterparts are in the last column.

The ultimate test of a PES is to use it to compute observables that depend sensitively on its quality. We have computed a vibrational spectrum. Vibrational energy levels depend sensitively on the PES. It is obvious that a fit with a small RMSE is better than a fit with a large RMSE, but what really matters is the shape of the potential. It is not clear how small the RMSE must be in order to obtain energy levels of a given accuracy. We used a direct product basis and a Lanczos eigensolver to compute the spectrum. The stretch basis has 22 equilibrium slice PODVR[109,110] functions. The bend basis is 50 Legendre-DVR functions.[13] In Table 1, we present absolute errors for both fits, and in Figure 3, we plot the absolute energy level error as a function of energy for the 1 cm^{-1} RMSE fit. All errors are

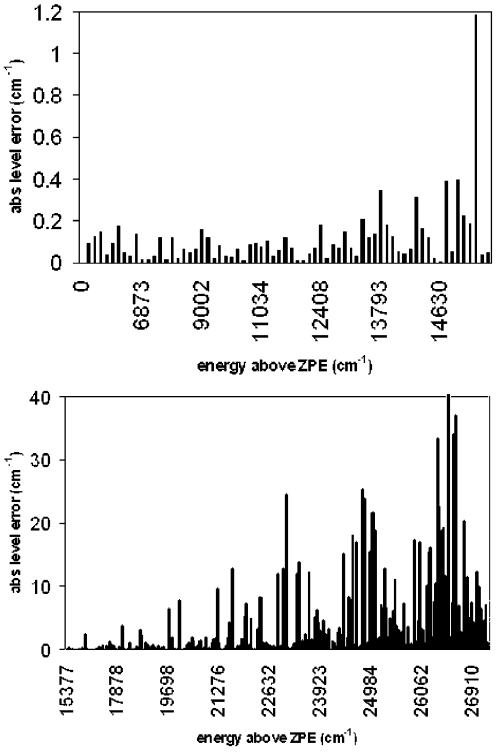

Figure 3. Absolute energy level error as a function of energy for the best H_2O NN potential. The origin is at ZPE = 4630.16 cm^{-1}, calculated on this surface. The largest energy training and validation point energies are ≈15370 cm^{-1} above the ZPE. The upper panel shows errors within the energy range covered by the training points. The lower panel shows level errors for energy levels larger than the energies of the training points.

differences between energies computed with the NN potential and with the original Jensen potential. Before computing differences, the appropriate ZPEs are subtracted. With one exception, levels in the figure below E_{cut} which is $\approx$15370 cm^{-1} (above the ZPE) have errors of less than 0.5 cm^{-1} and the majority of the levels have errors smaller than 0.08 cm^{-1}. Low-lying levels are more accurate, because we have chosen our point distribution to make the potential more accurate near the bottom of the well. The zero-point energy calculated on the NN potential is 4630.16 cm^{-1} vs 4630.35 cm^{-1} on Jensen's potential. The potential minimum is situated at r(OH) = 0.93804 Å, $\angle$HOH = 107.726° vs r(OH) = 0.93808 Å, $\angle$HOH = 107.706° for the original potential. The errors in the NN levels are substantially smaller than differences between levels computed on the Jensen potential and experimental energies (given in refs 89 and 111). See the right-most column of Table 1. In the lower panel of Figure 3, we have plotted the spectrum up to $\approx$27000 cm^{-1} (above the ZPE) to show that, in this case, the energy-level error increases smoothly outside the range of data used for training and validation. Not only does the NN fit do very well for levels below E_{cut} but, in this case, it also does pretty well even for higher levels.

B. Hydrogen Peroxide. For hydrogen peroxide, we fit the potential of Kuhn et al.[18] As coordinates, we use the O−O bond length $r(O_1O_2)$; two O−H bond lengths $r(O_1H_1)$, $r(O_2H_2)$; the angles $\angle H_1O_1O_2$ and $\angle H_2O_2O_1$ ($\theta_{1,2}$), and the angle τ between the $H_1-O_1-O_2$ and $H_2-O_2-O_1$ planes. As was the case for H_2O, there is nothing special about these coordinates. 5000 points with $180° < \tau < 360°$ were selected using E_{cut} = 11 000 cm^{-1} and $\Delta = 0$. The point set was augmented by adding points along equilibrium slices up to E_{max} = 18 000 cm^{-1} and by adding points by exchanging the two OH's. Previous fits for HOOH used a similar number of points.[18,91] We find that a better fit is obtained by choosing $E_{max} > E_{cut}$ because doing so increases the energy of points that are in the E_{max} box but not in the fitting region. If E_{cut} is not $\lesssim$15000 cm^{-1} the point density is too low. The V^0 used with eq 2 is a sum of Morse potentials for the stretch coordinates, cubic polynomials for the bending angles, and a fourth-order polynomial for the dihedral angle. The Morse parameters and the bend and dihedral coefficients are again determined by least-squares fits. We use 20 points for each coordinate.

We applied the same NN procedure that was used in the previous subsection. 36 nodes were required to obtain a surface with a RMSE of 20 cm^{-1}. Using a product Lanczos method, we then attempted to compute vibrational levels on the NN surface. We use PODVR functions for the stretches and spherical harmonic type functions for the bend coordinates. Several groups have used this type of angular functions with the Lanczos algorithm.[112,113] On attempting to converge the spectrum (by increasing the basis size and the quadrature grid size), we discovered that the NN surface has regions in which the energy is much less than the energy at the bottom of the HOOH well. This happens despite the fact that the RMSE for the validation set is good (less than twice the RMSE for the training set) and despite the fact that the 1-D cuts through the potential that we examined did not reveal holes. The holes appear at the outskirts of the fitted region, where the density of points is extremely low. In principle, one could find such holes by looking at potential values for a huge grid of points. Rather than looking at a huge multidimensional grid of potential values, one can compute energy levels. Holes in the surface are revealed by energy levels significantly lower than the expected ground state energy or levels that fail to converge. Clearly, a multidimensional NN fitted surface determined with a small fit tolerance and a small validation set error cannot be assumed to be free of unphysical regions, and merely looking at slice profiles is not guaranteed to reveal problems. Computing energy levels is a good way of determining whether there are holes, because the quadrature (DVR) grid used spans the entire surface.

How can one avoid making surfaces with holes? Of course, one can always ensure that holes on the outskirts do not affect energy levels by increasing E_{cut} and E_{max} and the number of points, but for a molecule with four atoms, this is a costly solution to the problem. A better method for preventing holes involves doing a two-step fit. A two-step fit was also used in ref 91. The final potential is the sum of a first-step function, which we call a base potential, and a difference potential. The difference potential is obtained by fitting the difference of the full and the base potentials. For HOOH, we used two different base potentials: one equal to V^0 and another that is an approximate NN fit for the full point set.

The separable base potential has qualitatively the right shape, which makes it less likely that the two-step NN potential obtained from it will have holes. The approximate NN base potential is unlikely to have holes, because it is fit with a small number of neurons and holes in regions with a low density of fitting points are caused by overfitting. The number of nodes used to construct the NN base potential is chosen so that it is small enough that the associated base potential does not have holes, but large enough that the difference potential is small in most regions of configuration space. In practice, we choose the number of neurons for the base potential so that the number of NN parameters is at least an order of magnitude less than the number of points and so that the target RMSE of the first-stage NN is about an order of magnitude larger than the target RMSE for the whole PES. To eliminate completely the possibility of having deep holes or false peaks in the final potential, we replace all potential values obtained from the network that are larger than a ceiling value with the ceiling value and all potential values obtained from the network that are less than a floor value with the floor value. The floor and ceiling values are the lowest and highest energies of the fitting points (after subtracting the base potential values).

With the uncoupled base potential, the difference potential could be fit to 20 cm^{-1} RMSE with 62 nodes and to 10 cm^{-1} RMSE with 100 nodes. The difference potential took values between $\pm\approx$7500 cm^{-1}. This large range of values is due to the very approximate nature of the base potential. Using a base potential with a smaller range of values would facilitate NN training at the second stage. It might be possible to choose more nearly decoupled coordinates to improve the quality of the sum-of-slices base potential, but we seek a more general solution. An NN-fitted potential with a fairly large RMSE does a good job. The final potential is therefore obtained from a *hierarchical NN fit*. The "nested" NN procedure is very easy to use. The base and difference NN potentials are determined using the same algorithm, the same points, and the same computer program. We have found that the quality of the final potential does not depend sensitively on the target RMSE for the base potential. The total number of fitting parameters is the *sum* of the number of parameters for the base potential and the number of parameters for the difference potential. Increasing the number of base potential nodes has a fairly small effect on the total cost of the fitting process. On the other hand, if no base potential were used and the potential were fitted with a one-step NN method, increasing the number of nodes would greatly increase the cost (due to the cubic scaling of the LM algorithm) and the

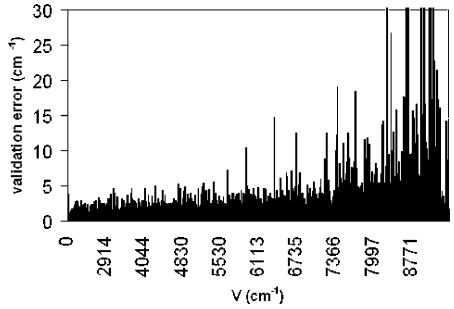

Figure 4. Absolute difference between the NN potential of HOOH fitted to 2 cm^{-1} RMSE and the reference potential by Kuhn et al.[18] as a function of energy for the *validation* set of points.

TABLE 2: Error in the NN Fit and NN Eigenvalues for HOOH[a]

	error, cm^{-1}			
	NN fit		NN eigenvalues	
	10 cm^{-1} RMSE[b]	2 cm^{-1} RMSE[c]	10 cm^{-1} RMSE[b]	2 cm^{-1} RMSE[c]
mean absolute	8.854	1.70	2.044	1.605
median	5.917	1.11	1.220	0.685
minimum	0.003	0.00	0.003	0.000
maximum	209.0	37.0	12.55	11.83

[a] Properties of the absolute differences of the HOOH NN potentials and the reference potential at the *validation* points are in the second and third columns. The fourth and fifth columns contain values that characterize the accuracy of the energy levels computed from the NN potentials. The ZPE is subtracted from all levels and then differences of the first 352 NN potential levels and the reference potential levels are computed. [b] 100 nodes, V^0 base potential. [c] Hierarchical NN with 57/175 nodes.

danger of holes. Note that there is no reason to use a separable model as the base potential. Indeed, the best fits presented in the HOOH and H_2CO tables are obtained with NN base potentials.

With the hierarchical NN procedure, it is possible to fit the HOOH PES to 25/5 cm^{-1} and to 10/2 cm^{-1} RMSE with 30/100 and 57/175 nodes, respectively (first/second NN stage). The difference potential takes values in the range $\pm\approx$60 cm^{-1}. The second and third columns of Table 2 characterize the potential errors of the 10 and 2 cm^{-1} RMSE NN potentials. The corresponding distribution of absolute errors at the validation points is shown in Figure 4 for the 2 cm^{-1} RMSE fit. The majority of the *validation* points have errors below 1.2 cm^{-1} with the maximum at 37 cm^{-1}. The potential energy errors are smallest at the bottom of the potential and grow gradually with energy. Compared with the RMSE of other fitting methods, a 2 cm^{-1} RMSE is good.[18,36,51,91−93] Because different fits are designed for different regions of the surface and use different numbers of points, it is hard to compare the RMSEs. In any case, the NN surface error at the validation points is less than the error due to imperfections in most ab initio methods.[114−116]

Using this two-step NN potential, we now compute vibrational energy levels. We use diatom−diatom Jacobi coordinates and enough basis functions and quadrature points to converge all the levels reported in Table 2 to within 0.01 cm^{-1}. The basis size is similar to previous studies.[112,113] Unlike the calculation with the one-step NN surface, which had holes, the calculations with the two-step surfaces converged well. There are no holes. Results are shown in Figure 5. We subtract the appropriate ZPE from the energy levels computed on both the 2 cm^{-1} RMSE two-step NN potential and the original Kuhn et al. potential and then compute differences. The magnitude of the differences

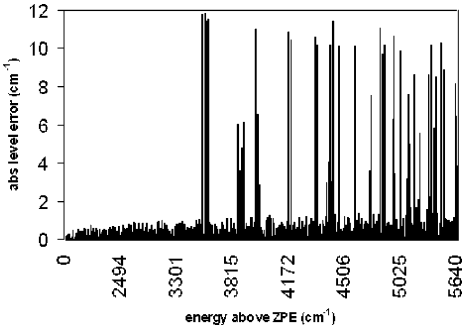

Figure 5. Absolute difference between vibrational levels (above the ZPE) of HOOH calculated on NN-interpolated potential (2 cm^{-1} RMSE) and levels computed on the reference potential[18] for the first 352 vibrational levels of both parities.

TABLE 3: Stationary Point Geometries and Torsional Barrier Heights for the NN-Fitted PES of HOOH and the PCPSDE Surface of Kuhn et al.[18] (in Brackets)

property	minimum	trans saddle	cis saddle
r(OH), a_0	1.818 [1.818]	1.817 [1.816]	1.818 [1.818]
r(OO), a_0	2.745 [2.745]	2.761 [2.761]	2.753 [2.754]
θ_1 (deg)	99.76 [99.8]	98.27 [98.3]	104.9 [104.9]
τ (deg)	114.4 [114.3]	180.0 [180.0]	0 [0]
E, cm^{-1}	0.21 [0]	359.6 [361]	2649 [2645]

for the first 352 vibrational levels (both parities) below 5665 cm^{-1} above the ZPE is plotted. The same level errors are characterized in the last column of Table 2. The majority of the levels are reproduced to better than 0.7 cm^{-1} with a maximum error of 11.8 cm^{-1}. The tunneling splitting is 11.025 cm^{-1} which is very close to the original surface's splitting of 11.031 cm^{-1} and to the experimental value of 11.44 cm^{-1}.[117] Given that we did not empirically adjust the surface nor choose the point distribution to favor dynamically important geometries (e.g., the top of the barrier) the agreement is very good. The stationary points on the NN surface are very close to those found on the original surface, as listed in Table 3.

The ZPE of our 2 cm^{-1} RMSE NN surface is 5695.79 cm^{-1}, whereas it is 5691.66 cm^{-1} [113] for the reference potential. The ZPE error (and the error in many of the energy levels if the ZPE is not subtracted before errors are computed) is therefore larger than the error in levels from which the ZPE has been subtracted. Energy level differences are more accurate than absolute energies. Subtracting the ZPE from the levels yields differences that are less sensitive to the potential than the energy levels themselves. An error of 0.7 cm^{-1} (subtracting the ZPE before computing the error) or 4.1 cm^{-1} (the ZPE difference) is small compared to the errors due to the inaccuracy of most ab initio calculations. For example, with reasonable basis set sizes, the error in CCSD(T) harmonic frequencies is generally larger than a few cm^{-1}.[118−120] This is probably a good measure of the extent to which the shape of a potential is distorted by ab initio error. Of course, one wishes the fitting error to shift energy levels as little as possible but, if ab initio error shifts levels by several inverse centimeters, our fitting errors can be considered small. To decrease further the energy level errors, it would be necessary to do a better job fitting the surface in the "corners" of the box, whose size is determined by E_{max}, that contains all our fitting points. Outside the box, the potential is equal to a large plateau value. Where the reference potential is less than E_{cut}, the fit is very good. At points for which the reference potential is larger than E_{cut} but inside the fitting box and for which the potential is therefore neither fit nor set equal to the plateau value, the NN function is sometimes poor. We have confirmed that these nonfitted regions are responsible for

a significant part of the error in the HOOH ZPE and absolute (no subtraction of the ZPE) levels. The shifts introduced by the NN fit are also smaller than differences between experimental levels and those computed on high-quality PESs.[112,117,121−126]

The HOOH NN surface can be improved by decreasing the RMSE from 10 to 2 cm^{-1} without drastically increasing the number of neurons (parameters). However, we find that to obtain an even better NN surface, we would need to increase the number of points. If we do not increase the number of points, we find that although we can improve the RMSE for the fitting points, we cannot find a surface for which the RMSE for the validation points is only a factor of 2 larger than the RMSE for the fitting points. A surface, for which the RMSE for the fitting points is very good but the RMSE for the validation points is not good, is "overfit".

C. Formaldehyde. To test the generality of the two-stage NN fitting procedure we developed to obtain a HOOH potential, we have also applied it to formaldehyde. We use as a reference potential a potential of Carter and Handy.[90] The HOOH and H_2CO potentials are both 6-D, but they are very different and fit to very different functional forms. In this subsection, we demonstrate that the nested NN approach that works well for HOOH also works well for H_2CO. This implies that the NN fitting procedure we used for HOOH is probably useful for a large class of molecules.

We select points according to eq 2 with $E_{cut} = 17\,000$ cm^{-1} and $\Delta = 500$ cm^{-1} and make V^0 from 1-D Morse functions for all stretch coordinates, $r(CO)$, $r(CH_{1,2})$; second-order polynomials in the bend angles, $\angle H_{1,2}CO$; and a fourth-order polynomial in the dihedral angle τ between the HCO planes. As for H_2O and HOOH, the V^0 parameters are determined by doing simple 1-D least-squares fits; for H_2CO, $E_{max} = E_{cut} = 17\,000$ cm^{-1}. At the equilibrium geometry, $\tau = 180°$. Because $V(180° + \delta) = V(180° - \delta)$, it is in principle possible to fit only half the potential. Instead, we choose a set of 2500 fitting points with $\tau < 180°$ and then add the corresponding points with $\tau > 180°$. For H_2CO, this improves the quality of the fit at the bottom of the well, but for HOOH, it is unnecessary owing to its out-of-plane equilibrium geometry. We also add points obtained by exchanging the two H atoms. Note that if we were fitting ab initio points, adding points obtained by exchanging the two hydrogen atoms and points with $\tau > 180°$ to the set of fitting points would not require extra ab initio calculations. The points along the 1-D slices up to E_{max} are also included in the set of points used to fit the NN parameters. Fits with a RMSE of 5 cm^{-1} and a RMSE of 2 cm^{-1} are obtained with 51/94 and 50/119 nodes (first/second stage).

In Figure 6, we show the potential error for the points in the *validation* set as a function of energy for the 2 cm^{-1} RMSE fit. The second and third columns of Table 4 characterize the error for the two fits. For the 2 cm^{-1} RMSE fit, most points are reproduced to better than 0.8 cm^{-1}. There are no low-energy points with significantly larger errors. Some points with large energies have much larger errors; the largest error is 32.9 cm^{-1}. Note that these errors give a realistic impression of the quality of the fit, because they are errors for points in the validation set and not in the fitting set. The minimum on the NN-fitted surface $r(CH) = 1.10069$ Å, $r(CO) = 1.20298$ Å, $\angle HCO = 121.645°$, and $\tau = 180°$ compares well with the reference surface minimum at $r(CH) = 1.10064$ Å, $r(CO) = 1.20296$ Å, $\angle HCO = 121.648°$, and $\tau = 180°$.[90]

As we did for the previous two molecules, we have computed vibrational energy levels to test for holes and to assess the effect of the RMSE of the surface on observables. The eigenvalues

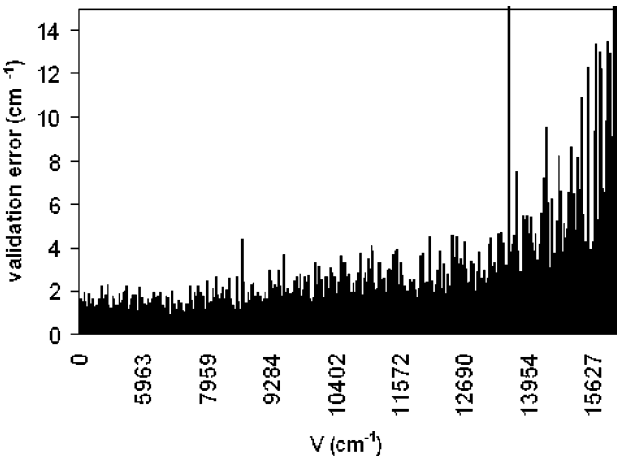

Figure 6. Absolute difference between the NN potential of H_2CO fitted to 2 cm^{-1} RMSE and the reference potential[90] as a function of energy for the *validation* set of points.

TABLE 4: Error in NN Fit and NN Eigenvalues for H_2CO[a]

	error, cm^{-1}			
	NN fit		NN eigenvalues	
	5 cm^{-1} RMSE	2 cm^{-1} RMSE	5 cm^{-1} RMSE	2 cm^{-1} RMSE
mean absolute	3.01	1.05	0.915	0.728
median	2.12	0.72	0.715	0.440
minimum	0.00	0.00	0.000	0.000
maximum	80.1	32.9	7.947	8.338

[a] Properties of the absolute differences of the H_2CO NN potentials and the reference potential at the *validation* points are in the second and third columns. The fourth and fifth columns contain values that characterize the accuracy of the energy levels computed from the NN potentials. The ZPE is subtracted from all levels and then differences of the NN potential levels and the reference potential levels are computed

were calculated with a product basis Lanczos method using Radau vectors for the CH_2 group and a Jacobi vector for the O atom.[127] In Figure 7, we give the magnitude of the difference between the levels (from which we have subtracted the ZPE) computed on the NN and the reference potentials as a function of energy up to ≈ 8000 cm^{-1} above the ZPE for the first 260 vibrational levels of both parities. These errors are characterized in the last column of Table 4. The majority of the levels are reproduced to better than 0.5 cm^{-1} with a maximum error of 8.3 cm^{-1}. Our ZPE is 5775.46 vs 5775.28 cm^{-1} for the reference potential.[90] Both the RMSE (2 cm^{-1}) and the level errors are small compared with typical fitting and ab initio errors.[120,128−130] Of course, there is still room for improvement, but the fitting

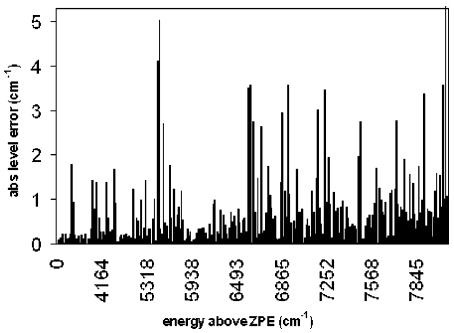

Figure 7. Absolute difference between vibrational levels (above the ZPE) of H_2CO calculated on NN-interpolated potential (2 cm^{-1} RMSE) and levels computed on the reference potential[90] for the first 260 vibrational levels of both parities.

error we have achieved is good. To get an even better fit, it would probably be necessary to increase the number of points. Without increasing the number of points, the level errors are improved only modestly by reducing the RMSE from 5 to 2 cm^{-1} (Table 4).

IV. Conclusion

As computer power increases and new theoretical methods are developed, it becomes increasingly easy to do both ab initio electronic structure calculations at preselected geometries[131−133] and quantum dynamics calculations on potential energy surfaces in 3 to 9 dimensions.[113,134−139] However, it is not possible to do quantum dynamics without a potential: it is not enough to know the potential at a large set of points. Methods for obtaining surfaces from points are therefore of paramount importance. If we can find general tools for making surfaces, it will be possible to do much interesting chemistry. Quantum dynamics techniques are at the point where they can be used to produce results of real interest to many chemists, but, of course, this is possible only if PES's are available. In this paper, we argue that NN's, if used cleverly, can be used to fit potentials for many small molecules. NN's are black-box, universal fitting methods that seem very promising. There are good pre-programmed NN tools readily available, and it is not hard to use them.[75] To use the NN procedure, there is no need for physical insight or experimental data (dipole moments, polarizabilities etc.). All of the parameters are determined solely by the NN algorithm. The structure of the algorithm does not depend on how coordinates are chosen. Of course, this universality is an advantage and a disadvantage. A disadvantage of NNs is that there is no simple way to build the correct asymptotic behavior or symmetry into the potential.

For H_2O, the NN fit was easy to obtain and accurate. For HOOH and H_2CO, we have developed a nested NN approach. The key elements of the nested NN fitting procedure are: (1) points are chosen so that the density of points is larger in low-lying regions of the potential; (2) an NN base potential is fitted; (3) the difference between the base potential and the true potential is fit using another NN; (4) a validation set is used to prevent overfitting; (5) we include symmetrically equivalent points in the fitting set and symmetrize the network output. We have shown that good fits can be obtained in this fashion. If the nesting idea is used, the size of the networks required to achieve a predefined accuracy increases only slowly as the dimensionality of the problem increases or the predefined RMSE decreases. Of course, the idea is readily generalized to more than two levels of NN fit.

Some authors have found that when using NNs, it is hard to avoid local minima.[140] We found that this was not a serious problem. If, during the course of the fit, the validation error begins to increase, we abort the fit and start again with a new set of random initial weights. The local minimum problem was slightly worse (i.e., it was necessary to restart more often) with fewer coordinates and small networks ($<20-30$ nodes). For the potential surfaces discussed in section III, it was virtually inexistent − if the number of nodes was large enough to reach the desired precision, the fit converged within two or three attempts. Of course, we have no way of knowing whether a slightly better fit might be obtained by restarting with another set of random initial weights, but any fit that satisfies the error target is good enough.

To obtain good fits, we need to use as many as 100 or more nodes. For a four-atom molecule, 100 nodes correspond to ≈800 fitting parameters (weights and biases). This may seem excessive, but it is the price we pay for generality. Note in comparison that for a 3-D spline PES, it is necessary to use orders of magnitude more parameters.[24,141,142] Other black-box fitting methods such as IMLS, MSI, DAF, or RKHS also require a large number of parameters. Using a functional form designed for a particular molecule, it is possible to obtain a good fit with fewer parameters, but having many parameters is not a significant disadvantage. Of course, the larger the number of parameters the more costly are the matrix manipulations that must be done to determine the best parameter values. For our calculations, we used the LM algorithm, but other options exist.[39,44,60,88 143] Even using the LM algorithm, it took only hours using a MatLab program on a Windows PC at 3 GHz to converge the fits.

We were able to decrease the fit RMSE to reasonable limits without drastically increasing the network sizes. For HOOH and H_2CO, with about 10^4 fitting points in the energy ranges within which we fit, it was not possible to reduce the RMSE below 2 cm^{-1} without increasing the error for the validation points. Getting even better surfaces would necessitate more points. In molecules with four atoms or more, this would require sequential training of the NN lest the training process become too slow. Such algorithms exist.[44,88,143]

Perhaps the greatest strength of an NN-based approach is its adaptability. It is common to fit a potential by choosing a functional form and varying parameters to achieve a good fit. NN methods are more flexible because the actual functional form is varied during the fitting process (as neurons are added to the network). It is this feature that makes them good universal fitting tools. In the language of the NN community, this is the ability of the NN to "learn". We have exploited this advantage of NN methods to develop a nested NN potential fitting approach. A key component of the approach is the use of a set of validation points to ensure that the NN function we obtain not only accurately reproduces the potential at the fitting points but also gives the correct shape. It is encouraging that by using NNs in a hierarchical fashion the total number of parameters required to fit a 6-D surface is quite small. Using the NN procedures we describe, or methods derived from them, it should be possible to develop black-box, general potential fitting tools. These would make it possible to bridge the gap between single point quantum chemistry and quantum dynamics calculations and to compute spectra and maybe cross sections, rate constants etc. in a routine fashion.

Acknowledgment. This work has been supported by the Natural Sciences and Engineering Research Council of Canada. Some of the calculations were done on a computer of the Réseau québécois de calcul de haute performance (RQCHP).

References and Notes

(1) Papousek, D.; Aliev, M. R. *Molecular Vibrational−Rotational Spectra*; Elsevier: Amsterdam, 1982.
(2) Bacic, Z.; Light, J. C. *Annu. Rev. Phys. Chem.* **1989**, *40*, 469.
(3) Carrington, T., Jr. *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons: New York, 1998.
(4) *Dynamics of Molecules and Chemical Reactions*, Wyatt, R. E., Zhang, J. Z. H.; Eds.; Dekker: New York, 1996.
(5) Schinke, R. *Photodissociation Dynamics*; Cambridge University Press: Cambridge, U.K., 1993.
(6) Bowman, J. M. *J. Phys. Chem.* **1998**, *102*, 3006.
(7) Bowman, J. M.; Gazdy, B.; Bentley, J. A.; Lee, T. J.; Dateo, C. E. *J. Chem. Phys.* **1993**, *99*, 308.
(8) Gruebele, M.; Bigwood, R. *Int. Rev. Phys. Chem.* **1988**, *17*, 91.
(9) Bolton, K.; Hase, W. L.; Schlegel, H. B.; Song, K. *Chem. Phys. Lett.* **1998**, *288*, 621.
(10) Chen, W.; Hase, W. L.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *228*, 436.

(11) Steckler, R.; Thurman, G. M.; Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1997**, *106*, 3926.
(12) Doubleday, C.; Bolton, K.; Peslherbe, G. H.; Hase, W. L. *J. Am. Chem. Soc.* **1996**, *118*, 9922.
(13) Light, J. C.; Carrington, T., Jr. *Adv. Chem. Phys.* **2000**, *114*, 263.
(14) Lill, J. V.; Parker, G. A.; Light, J. C. *Chem. Phys. Lett.* **1982**, *89*, 483.
(15) Schatz, G. C. *Rev. Mod. Phys.* **1989**, *61*, 669.
(16) Truhlar, D. G.; Steckler, R.; Gordon, M. S. *Chem. Rev.* **1987**, *87*, 217.
(17) Hirst, D. M. *Potential Energy Surfaces*; Taylor and Francis: London, 1985.
(18) Kuhn, B.; Rizzo, T. R.; Luckhaus, D.; Quack, M.; Suhm, M. A. *J. Chem. Phys.* **1999**, *111*, 2565.
(19) Quack, M.; Suhm, M. A. *J. Chem. Phys.* **1991**, *95*, 28.
(20) Koput, J.; Carter, S.; Handy, N. C. *J. Phys. Chem. A.* **1998**, *102*, 6325.
(21) Carter, S.; Handy, N. C. *J. Mol. Spectrosc.* **1996**, *179*, 65.
(22) *Approximation theory*; Lorentz, G. G., Chui, C. K., Shumaker, L. L., Eds.; Academic: New York, 1976; Vol. II.
(23) *Approximation theory and spline functions*; Singh, S. P., Barry, J. H. W., Watson, B., Eds.; Reidel: Dordrecht, The Netherlands, 1984.
(24) Bowman, J. M.; Bittman, J. S.; Harding, L. B. *J. Chem. Phys.* **1986**, *85*, 911.
(25) Chapman, S.; Dupuis, M.; Green, S. *Chem. Phys.* **1983**, *78*, 93.
(26) Maisuradze, G. G.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2003**, *119*, 10002.
(27) Guo, Y.; Kawano, A.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2004**, *121*, 5091.
(28) Hollebeek, T.; Ho, T.-S.; Rabitz, H. *Annu. Rev. Phys. Chem.* **1999**, *50*, 537.
(29) Ho, T.-S.; Rabitz, H. *J. Chem. Phys.* **2003**, *119*, 6433.
(30) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080.
(31) Jordan, M. J. T.; Thompson, K. C.; Collins, M. A. *J. Chem. Phys.* **1995**, *102*, 5647.
(32) Betetens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 816.
(33) Collins, M. A. *Theor. Chem. Acc.* **2002**, *108*, 313.
(34) Frishman, A. M.; Hoffman, D. K.; Rakauskas, R. J.; Kouri, D. J. *Chem. Phys. Lett.* **1996**, *252*, 62.
(35) Hoffman, D. K.; Frishman, A.; Kouri, D. J. *Chem. Phys. Lett.* **1996**, *262*, 393.
(36) Frishman, A. M.; Hoffman, D. K.; Kouri, D. J. *J. Chem. Phys.* **1997**, *107*, 804.
(37) Szalay, V. *J. Chem. Phys.* **1999**, *111*, 8804.
(38) Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. *J. Chem. Phys.* **2005**, *122*, 084104.
(39) Lorenz, S.; Gross, A.; Scheffler, M. *Chem. Phys. Lett.* **2004**, *395*, 210.
(40) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. *J. Phys. Chem. A* **1998**, *102*, 4596.
(41) Prudente, F. V.; Acioli, P. H.; Neto, J. J. S. *J. Chem. Phys.* **1998**, *109*, 8801.
(42) Prudente, F. V.; Neto, J. J. S. *Chem. Phys. Lett.* **1998**, *287*, 585.
(43) Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. *J. Chem. Phys.* **1996**, *105*, 7597.
(44) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. *J. Chem. Phys.* **1995**, *103*, 4129.
(45) Sumpter, B. G.; Noid, D. W. *Chem. Phys. Lett.* **1992**, *192*, 455.
(46) Hollebeek, T.; Ho, T.-S.; Rabitz, H. *J. Chem. Phys.* **1997**, *106*, 7223.
(47) Ho, T.-S.; Rabitz, H. *J. Chem. Phys.* **1996**, *104*, 2584.
(48) Ho, T.-S.; Hollebeek, T.; Rabitz, H.; Harding, L. B.; Schatz, G. C. *J. Chem. Phys.* **1996**, *105*, 10472.
(49) Hollebeek, T.; Ho, T.-S.; Rabitz, H.; Harding, L. B. *J. Chem. Phys.* **2001**, *114*, 3945.
(50) Ho, T.-S.; Hollebeek, T.; Rabitz, H.; Chao, S. D.; Skodji, R. T.; Zyubin, A. S.; Mebel, A. M. *J. Chem. Phys.* **2002**, *116*, 4124.
(51) Collins, M. A.; Bettens, R. P. A. *Phys. Chem. Chem. Phys.* **1999**, *1*, 939.
(52) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.
(53) Bettens, R. P. A.; Hansen, T. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 6322.
(54) Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 2424.
(55) Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302.
(56) Collins, M. A.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 6222.
(57) Morzano, G. E.; Collins, M. A. *J. Chem. Phys.* **2003**, *119*, 5510.
(58) Brown, A.; McCoy, A. B.; Braams, B. J.; J. Z., Bowman, J. M. *J. Chem. Phys.* **2004**, *121*, 4105.
(59) Huang, X.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2005**, *122*, 044308.
(60) Hassoun, M. H. *Fundamentals of artificial neural networks*; MIT Press: Cambridge, MA, 1995.
(61) Widrow, B.; Lehr, M. A. *Proc. IEEE* **1990**, *78*, 1415.
(62) Hornik, K.; Stinchcombe, M.; White, H. *Neural Networks* **1989**, *2*, 359.
(63) Hornik, K. *Neural Networks* **1991**, *4*, 251.
(64) Sumpter, B. G.; Getino, C.; Noid, D. W. *Annu. Rev. Phys. Chem.* **1994**, *45*, 439.
(65) Zupan, J.; Gastreiger, J. *Anal. Chim. Acta* **1991**, *248*, 1.
(66) Bishop, C. M.; Roach, C. M. *Rev. Sci. Instrum.* **1992**, *63*, 4450.
(67) Shenvi, N.; Geremia, J. M.; Rabitz, H. *J. Chem. Phys.* **2004**, *120*, 9942.
(68) Darsey, J. A.; Noid, D. W.; Upadhyaya, B. R. *Chem. Phys. Lett.* **1991**, *177*, 189.
(69) E Lagaris, I.; Likas, A.; Fotiadis, D. I. *Comput. Phys. Comm.* **1997**, *104*, 1.
(70) Sugawara, M. *Comput. Phys. Comm.* **2001**, *140*, 366.
(71) Wahab, A.; Ng, G. S.; Dickiyanto, R. *Neurocomputing* **2005**, *68*, 13.
(72) Valle-Lisboa, J. C.; Reali, F.; Anastasia, H.; Mizraji, E. *Neural Networks* **2005**, *18*, 863.
(73) Kanter, I.; Kinzel, W.; Kanter, E. *Europhys. Lett.* **2002**, *57*, 141.
(74) Redko, V. G.; Mosalov, O. P.; Prokhorov, D. V. *Neural Networks* **2005**, *18*, 738.
(75) Demuth, H.; Beale, M. *Neural Network Toolbox Users Guide*; The MathWorks, Inc.: 1992−2004.
(76) Duch, W.; Diercksen, G. H. F. *Comput. Phys. Comm.* **1994**, *82*, 91.
(77) Anthony, M. A.; Bartlett, P. L. *Neural network learning: theoretical foundations*; Cambridge University Press: Cambridge, U.K. 1999.
(78) Scarselli, F.; Tsoi, A. C. *Neural Networks* **1998**, *11*, 15.
(79) Kurkova, V. *Neural Networks* **1992**, *5*, 501.
(80) Kolmogorov, A. N. *Dokl. Akad. Nauk SSSR* **1957**, *114*, 369.
(81) Sprecher, D. A. *Proc. Am. Math. Soc.* **1965**, *16*, 200.
(82) Sprecher, D. A. *Trans. Am. Math. Soc.* **1965**, *115*, 340.
(83) Funahashi, K. *Neural Networks* **1989**, *2*, 183.
(84) Barron, A. *IEEE Trans. Inf. Theor.* **1993**, *39*, 930.
(85) Mhaskar, H.; Miccheli, C. *Adv. Appl. Math.* **1992**, *13*, 350.
(86) Mhaskar, H.; Miccheli, C. *IBM J. Res. Dev.* **1994**, *38*, 277.
(87) Chui, C.; Li, X. *J. Approx. Theor.* **1992**, *70*, 131.
(88) Blank, T. B.; Brown, S. D. *Chemometrics* **1994**, *8*, J.; 391.
(89) Jensen, P. *J. Mol. Spectrosc.* **1989**, *133*, 438.
(90) Carter, S.; Handy, N. C.; Demaison, J. *Mol. Phys.* **1997**, *90*, 729.
(91) Kawano, A.; Guo, Y.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2004**, *120*, 6414.
(92) Maisuradze, G. G.; Thompson, D. L. *J. Phys. Chem.* **2003**, *107*, 7118.
(93) Maisuradze, G. G.; Kawano, A.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2004**, *121*, 10329.
(94) Thompson, K. C.; Collins, M. A. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 871.
(95) Collins, M. A.; Zhang, D. H. *J. Chem. Phys.* **1999**, *111*, 9924.
(96) Crittenden, D. L.; Thompson, K. C.; Chebib, M.; Jordan, M. J. T. *J. Chem. Phys.* **2004**, *121*, 9844.
(97) Garashchuk, S.; Light, J. C. *J. Chem. Phys.* **2001**, *114*, 3929.
(98) Fuller, R. O.; Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **2001**, *114*, 10711.
(99) Moyano, G. E.; Collins, M. A. *J. Chem. Phys.* **2003**, *119*, 5510.
(100) Moyano, G. E.; Collins, M. A. *J. Chem. Phys.* **2004**, *121*, 9769.
(101) Evenhuis, C. R.; Collins, M. A. *J. Chem. Phys.* **2004**, *121*, 2515.
(102) LeCu, Y.; Bottou, L.; Orr, G. B.; Muller, K.-R. Efficient backprop. In *Neural networks: tricks of the trade*; Orr, G. B.; Muller, K.-R.; Eds., Lecture notes in computer science; Springer: New York, 1998; Chapter 1, pp 9−50.
(103) Sexton, R. S.; Dorsey, R. E.; Johnson, J. D. *Decision Support Syst.* **1998**, *22*, 171.
(104) Sexton, R. S.; Dorsey, R. E.; Johnson, J. D. *Eur. J. Operational Res.* **1999**, *114*, 589.
(105) Sexton, R. S.; Dorsey, R. E. *Decision Support Syst.* **2000**, *30*, 11.
(106) Sexton, R. S.; Sriram, R. S.; Etheridge, H. *Decision Sci.* **2003**, *34*, 421.
(107) *Genetic Algorithm and Direct Search Toolbox Users Guide*; The MathWorks, Inc.: 2004−2005.
(108) Smith, F. T. *Phys. Rev. Lett.* **1980**, *45*, 1157.
(109) Wei, H.; Carrington, T., Jr. *J. Chem. Phys.* **1992**, *97*, 3029.
(110) Echave, J.; Clary, D. C. *Chem. Phys. Lett.* **1992**, *190*, 225.
(111) Halonen, L.; Carrington, T., Jr. *J. Chem. Phys.* **1988**, *88*, 4171.
(112) Chen, R.; Ma, G.; Guo, H. *J. Chem. Phys.* **2001**, *114*, 4763.
(113) Wang, X.-G.; Carrington, T., Jr. *J. Chem. Phys.* **2002**, *117*, 6923.
(114) Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6648.
(115) Abrams, M. L.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 5611.

(116) Dutta, A.; Sherrill, C. D. *J. Chem. Phys.* **2003**, *118*, 1610.
(117) Helminger, P.; Bowman, W. C.; DeLucia, F. C. *J. Mol. Spectrosc.* **1981**, *85*, 120.
(118) Kaldor, U. *Chem. Phys. Lett.* **1990**, *170*, 17.
(119) Pak, Y.; Woods, R. C. *J. Chem. Phys.* **1995**, *103*, 9304.
(120) Martin, J. M. L. *Chem. Phys. Lett.* **1998**, *292*, 411.
(121) Camy-Peyret, C.; Flaud, J.-M.; Johns, J. W. C.; Noel, M. *J. Mol. Spectrosc.* **1992**, *155*, 84.
(122) Hillman, J. J.; Jennings, D. E.; Olson, W. B.; Goldman, A. *J. Mol. Spectrosc.* **1986**, *117*, 46.
(123) Flaud, J.-M.; Camy-Peyret, C.; Johns, J. W. C.; Carli, B. *J. Chem. Phys.* **1989**, *91*, 1504.
(124) Olson, W. B.; Hunt, R. H.; Young, B. W.; Maki, A. G.; Brault, J. W. *J. Mol. Spectrosc.* **1988**, *127*, 12.
(125) Perrin, A.; Valentin, A.; Fraud, J.-M.; Camy-Peyret, C.; Schriver, L.; Schriver, A.; Arcas, P. *J. Mol. Spectrosc.* **1995**, *171*, 358.
(126) Cook, W. B.; Hunt, R. H.; Shelton, W. N.; Flaherty, F. A. *J. Mol. Spectrosc.* **1995**, *171*, 91.
(127) Branley, M. J.; Carrington, T., Jr. *J. Chem. Phys.* **1993**, *99*, 8519.
(128) Martin, J. M. L.; Lee, T. J.; Taylor, P. R. *J. Mol. Spectrosc.* **1993**, *160*, 105.
(129) Carter, S.; Pinnavaia, N.; Handy, N. C. *Chem. Phys. Lett.* **1995**, *240*, 400.
(130) Zhang, X.; Zou, S.; Harding, L. B.; Bowman, J. M. *J. Phys. Chem. A* **2004**, *108*, 8980.
(131) Grant, G.; Richards, W. G. *Computational Chemistry*; Oxford University Press: Oxford, U.K., 1995.
(132) Hehre, W. J.; Radom, L.; Schleyer, P.; Pople, J. *Ab Initio Molecular Orbital Theory*; John Wiley and Sons: New York, 1986.
(133) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Pubs: London.
(134) Wang, X.-G.; Carrington, T., Jr. *J. Chem. Phys.* **2003**, *119*, 101.
(135) Mladenovic, M. *Spectrochim. Acta Part A* **2002**, *58*, 809.
(136) Luckhaus, D. *J. Chem. Phys.* **2000**, *113*, 1329.
(137) Chen, R.; Guo, H. *J. Chem. Phys.* **1998**, *108*, 6068.
(138) Lehoucq, R. B.; Gray, S. K.; Zhang, D.-H.; Light, J. C. *Comput. Phys. Commn.* **1998**, *109*, 15.
(139) Keutsch, F. N.; Braly, L. B.; Brown, M. G.; Harker, H. A.; Petersen, P. B.; Leforestier, C.; Saykally, R. J. *J. Chem. Phys.* **2003**, *119*, 8927.
(140) Witkoskie, J. B.; Doren, D. J. *J. Chem. Theor. Comput.* **2005**, *1*, 14.
(141) Raff, L. M.; Sathyamurthy, N. *J. Chem. Phys.* **1975**, *63*, 464.
(142) Xu, C.; Xiea, D.; Zhang, D. H.; Lin, S. Y.; Guo, H. *J. Chem. Phys.* **2005**, *122*, 244305.
(143) Shah, S.; Palmieri, F.; Datum, M. *Neural Networks* **1992**, *5*, 779.