Atom-centered symmetry functions for constructing high-dimensional neural network potentials

Jörg Behler[a)]
Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany

(Received 8 December 2010; accepted 21 January 2011; published online 16 February 2011)

Neural networks offer an unbiased and numerically very accurate approach to represent high-dimensional *ab initio* potential-energy surfaces. Once constructed, neural network potentials can provide the energies and forces many orders of magnitude faster than electronic structure calculations, and thus enable molecular dynamics simulations of large systems. However, Cartesian coordinates are not a good choice to represent the atomic positions, and a transformation to symmetry functions is required. Using simple benchmark systems, the properties of several types of symmetry functions suitable for the construction of high-dimensional neural network potential-energy surfaces are discussed in detail. The symmetry functions are general and can be applied to all types of systems such as molecules, crystalline and amorphous solids, and liquids. © *2011 American Institute of Physics*.
[doi:10.1063/1.3553717]

I. INTRODUCTION

Ab initio molecular dynamics (MD) simulations have become a standard approach to study a variety of chemical processes.[1,2] In each time step, the energy and forces are calculated on-the-fly by electronic structure methods, most importantly density-functional theory. However, in spite of the availability of fast computers and efficient implementations, there is still a large number of interesting problems, which are not directly accessible by *ab initio* MD due to the high computational costs. In such cases empirical potentials can be used, and the development of reliable potentials is a very active field of research. Most potentials start from a physically motivated functional form and contain some parameters, which are fitted to experimental or *ab initio* data. While the physical basis of the obtained potential-energy surfaces (PESs) is appealing and in many cases the transferability is reasonable, the numerical accuracy is often limited by the fixed functional forms underlying the potentials.

An alternative approach is to sacrifice the physically motivated functional form in favor of an unbiased, purely mathematical fitting procedure. Potentials of this type are often very accurate in reproducing known reference data, but the transferability must be tested very carefully. A variety of methods such as splines,[3] Taylor expansions,[4–6] Gaussian based methods[7] and many others have been proposed. Another class of very flexible functions is given by artificial neural networks (NNs).[8] It is known that in principle NNs are able to fit any real-valued function with arbitrary accuracy,[9,10] and NNs have been applied to the construction of PESs for more than a decade.[11,12]

In general, NNs are high-dimensional functions which can adapt to a known set of reference points very accurately in an iterative fitting process. In case of NN potentials this function is the PES, and the general goal of the NN is to construct a functional relation between the energy and the atomic configuration of a system.

In the present paper we discuss an approach to construct high-dimensional NN PESs, which is based on an expansion of the total energy in terms of environment-dependent atomic energy contributions.[13,14] The key step in this method, which has been applied successfully to a number of systems,[15–17] is a transformation of the Cartesian coordinates onto a set of symmetry functions describing the chemical environments of the atoms. These symmetry functions are crucial for the applicability of the method, and here we will introduce and discuss the functional forms and the properties of several suitable types of functions in detail. The presented functions are implemented in the NN package RuNNer,[18] which has also been used to demonstrate their performance for several simple model systems in the present work. Possible problems in the fitting procedure are identified and related to the properties of the employed symmetry function sets. The presented symmetry functions are very general and allow the construction of NN PESs for a wide range of systems, from small molecules to periodic bulk systems. They are equally well applicable to all types of bonding, from weak nonbonded interactions via covalent bonds to metals.

II. FEED-FORWARD NEURAL NETWORKS

In recent years, NNs have become a promising new tool to construct very accurate PESs. The structure, also called architecture, of a small NN is shown schematically in Fig. 1. It consists of a number of nodes, or neurons, arranged in layers. The node in the output layer provides the function value of the NN, the energy E. The energy is a function of the atomic positions, which are provided to the NN in the form of some suitable coordinates $\mathbf{G} = \{G_i\}$ in the nodes of the input layer. There is one input node per degree of freedom, i.e., Fig. 1 represents a four-dimensional PES. In between the input and the output layer there are one or more so-called hidden layers. The nodes in the hidden layers have no physical meaning, but define the functional form of the NN. The more nodes are used, the more flexible is the resulting NN. Each node in each

a)Electronic mail: joerg.behler@theochem.ruhr-uni-bochum.de.

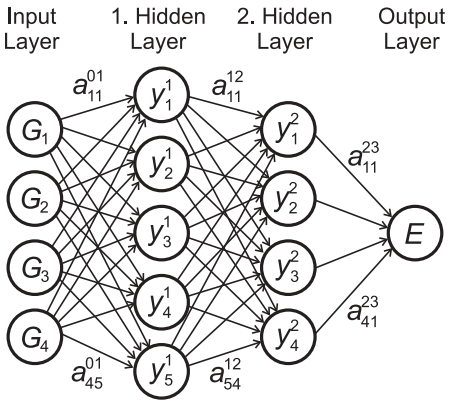

FIG. 1. Structure of a feed-forward neural network. In the output layer the energy E is obtained as a function of the four coordinates $\{G_i\}$ in the input layer. In between the input and the output layer, there are two hidden layers with five and four nodes, respectively, defining the functional form of the NN. The fitting parameters a_{ij}^{kl} are represented by the arrows. For clarity, the bias node and the bias weights are not shown.

layer is connected to the nodes in the previous layer and the subsequent layer by weight parameters, the fitting parameters of the NN. They are shown as arrows in Fig. 1. We use the notation that weight a_{ij}^{kl} is connecting node i in layer k with node j in layer l. The input layer corresponds to the layer superscript 0. Additionally, except for the input layer, each node i in layer j is connected to a bias node by the bias weight b_i^j (not shown). The bias node always has value one and the bias weights can be used as an adjustable offset to shift the input of the nodes. The value y_i^j of node i in layer j is then calculated as

$$y_i^j = f_i^j \left(b_i^j + \sum_k a_{ki}^{j-1,j} \cdot y_k^{j-1} \right). \tag{1}$$

In the special case $j - 1 = 0$, the y_k^{j-1} correspond to the G_k. The function f_i^j is called activation function of the NN. This is a nonlinear function providing the capability to fit arbitrary functions. Without the activation function, the energy would simply reduce to a linear combination of the coordinates. In the present work, a hyperbolic tangent is used as activation function for all nodes in the hidden layers, for the output node the linear function $f(x) = x$ is used.

The full analytic form of the feed-forward neural network, shown schematically in Fig. 1, is then given by

$$E = f_1^3 \left(b_1^3 + \sum_{l=1}^{4} a_{l1}^{23} \cdot f_l^2 \left(b_l^2 + \sum_{k=1}^{5} a_{kl}^{12} \cdot f_k^1 \left(b_k^1 + \sum_{j=1}^{4} a_{jk}^{01} \cdot G_j \right) \right) \right). \tag{2}$$

Typically, NN PESs contain two to three hidden layers with 25 to 40 nodes in each layer. Often, a notation describing the number of nodes in each layer is used, i.e., the NN in Fig. 1 is a 4-5-4-1 network.

The ability of the NN to accurately describe a given PES depends on the numerical values of the weight parameters. They are determined in an iterative optimization process. For this "training" of the NN, typically total energies and forces[19–21] from electronic structure methods are used. NN potentials contain a comparably large number of weight parameters. For a feed-forward NN with N_0 input nodes, N_1 nodes in the first hidden layer, and N_2 nodes in the second hidden layer a total number of $N_0 \cdot N_1 + N_1 \cdot N_2 + N_2$ connecting weights and $N_1 + N_2 + 1$ bias weights needs to be determined.

In recent years NN potentials have been constructed for a variety of systems. Most applications have been reported for small molecules in the gas phase,[22–32] but also PESs for the interaction of small molecules with (frozen) surfaces have been reported.[33–40]

III. HIGH-DIMENSIONAL NEURAL NETWORK POTENTIALS

Standard feed-forward NNs as discussed in Sec. II are useful tools to represent low-dimensional PESs. However, there are several problems if high-dimensional PESs for systems with thousands of degrees of freedom are to be constructed. First, since each degree of freedom is represented by an input node, the size of the NN increases significantly making the method less efficient. Second, the number of input nodes cannot be changed once a NN potential has been fitted. Therefore, in principle for each system size a separate potential would have to be constructed. This is not feasible. Finally, for a system containing a large number of atoms, exchanging the positions of any two atoms of the same chemical element does not change the structure. Obviously, this must be reflected by an invariance of the NN energy output with respect to such an exchange. However, this invariance is generally not present in feed-forward NNs, because the order of the coordinates in the NN input is not arbitrary. All input nodes are connected to the NN by numerically different weight parameters. Thus, in case of permutations in the input vector a different output is obtained. For small systems, it has been suggested to replicate the training data set to include all possible representations of a system,[26,41] but for large systems this cannot be done.

These conceptual problems can be solved by replacing the NN representing the total energy by a set of atomic NNs.[13,14] Each NN then provides the contribution E_i of an atom to the total energy of the system,

$$E = \sum_i E_i. \tag{3}$$

The E_i depend on the chemical environments of the atoms, which are defined by the positions of all neighbors inside some sphere with cutoff radius R_c. The functional relation between the positions of the neighboring atoms and the atomic energy contribution is constructed using a NN. It is trained to assign individual energies to the atoms by "learning" the PES based on a set of reference energies. Typically, no atomic reference energies are available in electronic structure calculations. Consequently, the atomic energies are

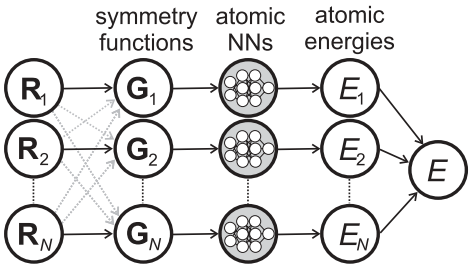

FIG. 2. Structure of a high-dimensional neural network potential. Each line represents one atom with the Cartesian coordinates $\mathbf{R}_i$. First, the coordinates are transformed to a set of symmetry function values $\mathbf{G}_i$ depending also on the Cartesian coordinates of all atoms in the local environment, as indicated by the dotted arrows. The symmetry function values represent the input vectors for atomic NNs yielding the atomic energy contributions E_i. The total energy E is the sum of all E_i.

constructed in order to reproduce the correct total energies according to Eq. (3).

The atom-based high-dimensional NN is shown schematically in Fig. 2 for a system containing N atoms. Each line represents one atom. Initially, the Cartesian coordinates $\mathbf{R}_i = (X_i, Y_i, Z_i)$ of all atoms are transformed to some suitable set of input coordinates for the NNs. These coordinates $\mathbf{G}_i = \{G_{i,j}\}$ are called symmetry functions and will be discussed in more detail in Sec. IV. The index $j = \{1, \ldots, M_i\}$ labels the symmetry function of atom i, whose environment is described by a total number of M_i functions. These functions are many-body functions depending on all atoms in the local environment, as indicated by the dotted arrows in Fig. 2. For each atom a set of symmetry function values is obtained, which is the input for its atomic NN. This yields the atomic energy contribution, and all contributions are finally added to obtain the total energy of the system.

For a given chemical element, the architectures and the weight parameters of the atomic NNs are identical, i.e., the sets of NN weights to be fitted correspond to the number of chemical species in the system. For example, in a binary system AB, two sets of NN weights need to be determined. Consequently, the high-dimensional NN scheme is applicable to systems with different numbers of atoms. If an atom is added, another atomic NN with the architecture and weights of the corresponding element is added to the total energy expression. If an atom is removed, its atomic NN is deleted. The sets of symmetry function values for two atoms of the same element must be identical, if the atomic environments are the same. Therefore, the symmetry functions represent structural fingerprints of the atoms. Then, exchanging the positions of any two atoms merely changes the order of the lines in Fig. 2, but the total energy is the same.

Finally, we note that several other advanced NN schemes for high-dimensional systems have been reported in the literature. If the number of atoms is still moderate, very systematic many-body expansions can be constructed and expressed by NNs.[27,29–31] For larger atom numbers it has been proposed to express terms of the Tersoff potential[42] by NNs.[44,45] Another high-dimensional NN approach expands the total energy in terms of structural fragments containing four to five atoms.[46,47]

IV. SYMMETRY FUNCTIONS

A. Properties of symmetry functions

It has been recognized more than a decade ago that Cartesian coordinates are not a good choice to describe the structural input of NN potentials. As NNs are numerical fitting methods, the output depends on the absolute values of the input coordinates. However, a translation or rotation of a molecule does not change its energy. Still, these operations change the Cartesian coordinates. Therefore, if the latter would be used as input for the NN, the energy would not be invariant with respect to translation and rotation. A straightforward solution is to describe molecular structures using internal coordinates such as interatomic distances. While this works well for small molecules, the choice is not unique and their number grows rapidly with system size. Thus, it has been recognized that for an efficient description of PESs it is necessary to take also the symmetry of the system into account. In particular the invariance of the total energy with respect to the exchange of any two atoms of the same chemical element should be included in the input coordinates of the NN by a proper symmetrization of the coordinates.[48] This approach has been further developed and other types of symmetry functions have been reported, e.g., to include the periodic potential of a crystalline surface experienced by adsorbing molecules.[34–36,38]

Apart from a correct description of the symmetry of the PES, symmetry functions have to fulfill several further requirements to become suitable inputs for NNs. They should be continuous, and analytic derivatives must be readily available. These derivatives are needed to calculate atomic forces. Further, for large interatomic distances the physical interactions between atoms decrease and approach zero. Therefore, it is numerically advantageous if the symmetry functions decay to zero in this case. While this is not strictly needed, it simplifies the fitting procedure significantly.

B. Symmetry functions for high-dimensional neural networks

For low-dimensional PESs, e.g., for small molecules, the number of degrees of freedom is typically fixed, if the number of atoms does not change. This is very convenient, because a necessary condition for setting up a NN potential is a constant number of input nodes. In general, for high-dimensional systems we need a NN potential that is applicable to arbitrary numbers of atoms. One step to achieve this goal has been discussed already in Sec. III, namely, the expression of the total energy as a sum of atomic energies and the introduction of a separate NN for each atom. This is equivalent to reducing the full-dimensional PES to a sum of lower-dimensional PESs of each atom, i.e., to the energetically relevant degrees of freedom.

However, the number of neighboring atoms in the local chemical environments cannot be fixed, since in the course of a MD simulation atoms can enter or leave the cutoff sphere. This represents an additional requirement for the symmetry functions. Their number must be fixed, i.e., it must be independent of the actual number of neighbors in the local

environment. Otherwise it would be necessary to train different NNs for each possible number of atoms, which is neither feasible nor desirable. A consequence is that a many-body expansion cannot be used, since the number of pair-terms, triples, etc., depends on the number of atoms in the sphere. This problem can be solved by constructing many-body symmetry functions simultaneously depending on the positions of all atoms inside the cutoff sphere.

We start the discussion of the functional forms of the symmetry functions by introducing the cutoff function

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c. \end{cases} \quad (4)$$

R_{ij} is the distance between atoms i and j, and if R_{ij} is larger than the cutoff radius R_c, the cutoff function and its derivative become zero. Similar cutoff functions have been used in other types of potentials, e.g., in the Tersoff potential.[42] A well-known problem of this cutoff function is that its second derivative has a discontinuity at the cutoff radius. If the cutoff chosen is sufficiently large, the effect of this discontinuity becomes small, as needed, e.g., in MD simulations. Other functional forms having continuous first and second derivatives can also be used.[43] Each symmetry function discussed below is multiplied by one or more cutoff functions to ensure that the total symmetry function decays to zero in value and slope at the cutoff radius. Consequently, atoms beyond the cutoff radius do not enter the atomic energy contributions.

Although all symmetry functions discussed here are true many-body functions, we call them radial and angular functions. Radial functions are constructed as sums of two-body terms, angular functions additionally contain sums of three-body terms. Three radial symmetry functions suitable for describing the radial environment of atom i are proposed,

$$G_i^1 = \sum_j f_c(R_{ij}), \quad (5)$$

$$G_i^2 = \sum_j e^{-\eta\left(R_{ij}-R_s\right)^2} \cdot f_c(R_{ij}), \quad (6)$$

and

$$G_i^3 = \sum_j \cos\left(\kappa R_{ij}\right) \cdot f_c(R_{ij}). \quad (7)$$

Function G^1 is simply the sum of the cutoff functions with respect to all neighboring atoms j. Function G^2 is a sum of Gaussians multiplied by cutoff functions. The width of the Gaussians is defined by a parameter η, and the center of the Gaussians can be shifted to a certain radial distance by the parameter R_s. These "shifted" G^2 functions then are suitable to describe a spherical shell around the reference atom. For small values of η and $R_s = 0$ function G^2 reduces to function G^1.

The radial distribution of neighbors can be described by using a set of radial functions with different spatial extensions, e.g., G^1 functions with different cutoff radii, or G^2 functions with different cutoffs and/or η parameters. Finally, radial function G^3 represents damped cosine functions with a period length adjusted by parameter κ. Like in a Fourier series expansion a suitable description of the radial atomic environment can be obtained by combining several G^3 functions with different κ values. However, function G^3 must be used with great care. Due to the existence of positive and negative function values, neighboring atoms at different distances can cancel each other's contributions to the summation. Therefore, it is recommended to use functions of type G^3 only in combination with other symmetry functions.

Typical forms of the radial symmetry functions are plotted in Fig. 3 for several parameter values. Since the radial functions are sums over all neighbors, physically they are related to effective coordination numbers, and a set of radial symmetry functions can be considered as a description of the coordination at various distances from the central atom.

Further, two types of angular functions are proposed. These functions are summations of cosine functions of the angles $\theta_{ijk} = \text{acos}(\mathbf{R_{ij}} \cdot \mathbf{R_{ik}}/R_{ij} \cdot R_{ik})$ centered at atom i. They are defined as

$$G_i^4 = 2^{1-\zeta} \sum_{j,k\neq i}^{\text{all}} (1 + \lambda \cos\theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2+R_{ik}^2+R_{jk}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}), \quad (8)$$

and

$$G_i^5 = 2^{1-\zeta} \sum_{j,k\neq i}^{\text{all}} (1 + \lambda \cos\theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2+R_{ik}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}). \quad (9)$$

Both functions share the same angular part plotted in Fig. 4, but they differ in the radial parts. In general, the angular part must be symmetric with respect to $\theta_{ijk} = 180^\circ$. The parameter λ can have the values $+1$ and -1 shifting the maxima of the cosine function to $\theta_{ijk} = 0^\circ$ and $\theta_{ijk} = 180^\circ$, respectively. The angular resolution is provided by the parameter ζ. High ζ values yield a narrower range of nonzero symmetry function values (cf. Fig. 4). Therefore, a set of angular functions with different ζ-values can be used to obtain the distribution of angles centered at each reference atom. This is similar to controlling the radial resolution of the radial functions G^2 by parameter η. Further, this angular distribution can be determined at various distances from the central atom by a suitable choice of η and R_c, which control the radial part.

For the special case of $R_{ij} = R_{ik} = R_{jk}$ the radial parts of G^4 and G^5 are compared for various values of η in Fig. 5. In function G^4 the product of three Gaussians ensures that only triples of atoms are included in the summation, for which all three interatomic distances are smaller than the cutoff radius. In function G^5 there is no constraint on R_{jk} resulting in a larger number of terms in the summation. Consequently, for given R_{ij} and R_{ik} function G^5 has nonzero values for a wider range of angles, as shown in Fig. 6.

Typically, in case of a single chemical element, about 50 symmetry functions with different parameters are used. The parameters define the symmetry functions and are not optimized during the training of the NN. The resulting set of function values is different for each atomic environment.

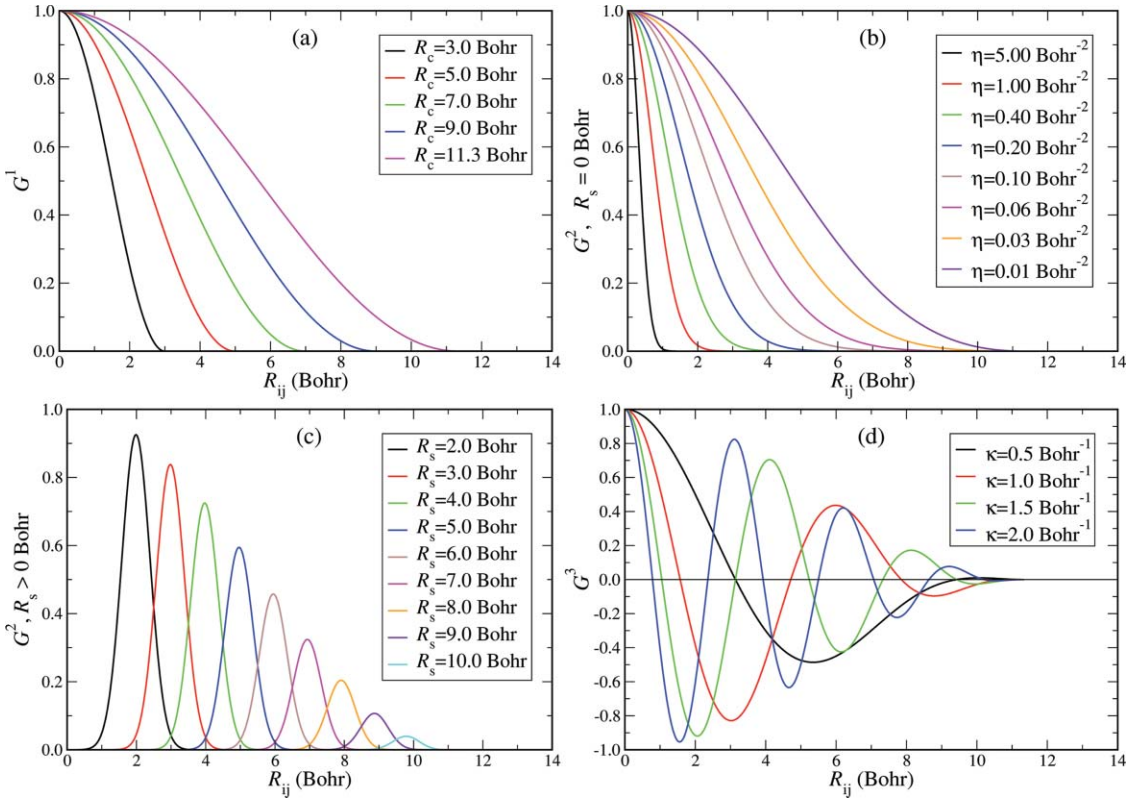

FIG. 3. Radial symmetry functions G^1, G^2, and G^3 for an atom with one neighbor only. In larger systems the radial symmetry function value of an atom is the sum of the contributions of each neighbor. For the G^2 and G^3 functions a cutoff $R_c = 11.3$ Bohr has been used, the Gaussian width of the shifted G^2 functions in panel (c) is $\eta = 3.0$ Bohr^{-2}.

Therefore, the total number of symmetry function values describing a given structure is clearly larger than the number of degrees of freedom of the system. This ensures that the full dimensionality of the system is captured. The resulting redundancy of the information is generally no problem for NNs.

C. Energy gradients: Forces and stress tensor

The analytic form of the neural network given in Eqs. (2) and (3) enables the calculation of analytic gradients, as needed e.g., for the forces in molecular dynamics simulations, and for the stress tensor required e.g., in Parrinello–Rahman MD[49,50] or metadynamics simulations for crystal structure prediction.[51]

The forces are the negative gradients of the total energy with respect to the atomic positions. In the present case, there is only an indirect relation due to the transformation of the atomic Cartesian coordinates into symmetry functions. Therefore, in order to calculate the force component $F_{k,\alpha}$ acting on atom k with respect to coordinate $R_{k,\alpha}$, $\alpha = (x, y, z)$, the

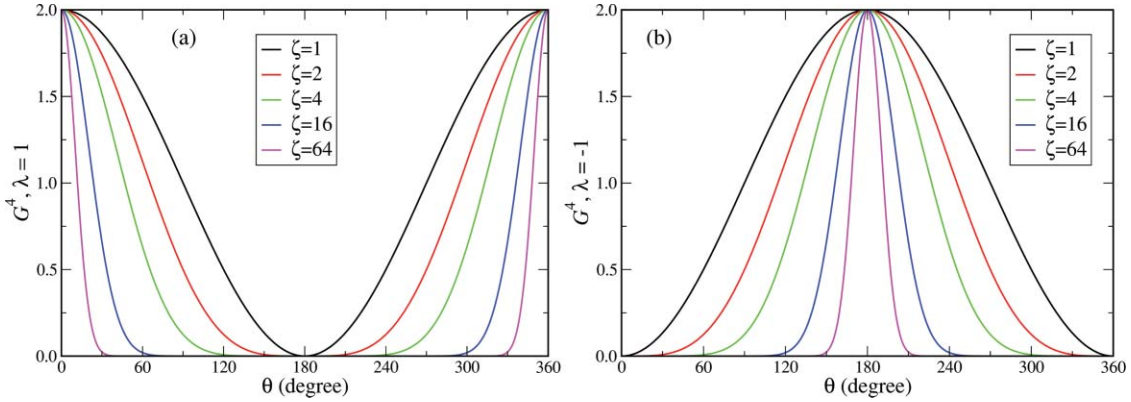

FIG. 4. Angular contributions of the angular symmetry function G^4. The angular parts of function G^5 are identical. The plotted angular part corresponds to a triatomic system. In many-atom systems the angular terms for all triples of atoms are added.

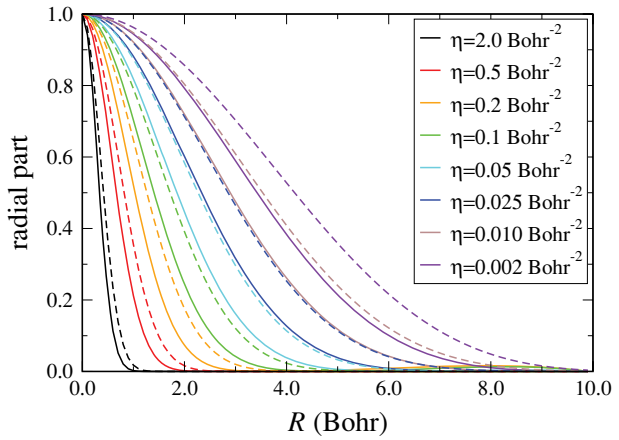

FIG. 5. Radial part of the angular symmetry functions shown for the case $R = R_{ij} = R_{ik} = R_{jk}$ and $R_c = 12$ Bohr. The solid curves represent the radial part of angular functions of type G^4, the dashed curves correspond to the radial parts of type G^5.

chain rule has to be applied,

$$F_{k,\alpha} = -\frac{\partial E}{\partial R_{k,\alpha}} = -\sum_{i=1}^{N} \frac{\partial E_i}{\partial R_{k,\alpha}} = -\sum_{i=1}^{N}\sum_{j=1}^{M_i} \frac{\partial E_i}{\partial G_{i,j}} \frac{\partial G_{i,j}}{\partial R_{k,\alpha}}. \tag{10}$$

N is the number of atoms, and M_i is the number of symmetry functions describing atom i. The first term, $\partial E_i/\partial G_{i,j}$, is given by the architecture of the NN, the second term, $\partial G_{i,j}/\partial R_{k,\alpha}$, is given by the definitions of the employed symmetry functions.

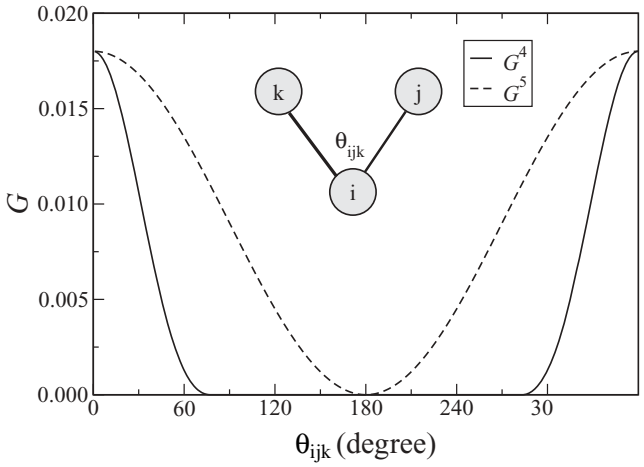

FIG. 6. Comparison of the total angular functions G^4 and G^5 (radial part × angular part) for atom i in a triatomic molecule. The interatomic distances R_{ij} and R_{ik} are fixed at $0.8R_c$ and the symmetry function values are plotted as a function of the angle θ_{ijk} centered at i. Function G^4 is zero for angles $70^\circ < \theta_{ijk} < 290^\circ$ because of the resulting large distance R_{jk} in the additional cutoff function $f_c(R_{jk})$.

In general, the stress tensor σ contains a kinetic contribution and a static contribution,

$$\sigma = \sigma^{\text{kin}} + \sigma^{\text{static}}. \tag{11}$$

The kinetic stress σ^{kin} is calculated from the atomic velocities v and masses m in MD simulations according to

$$\sigma^{\text{kin}}_{\alpha\beta} = \frac{1}{V}\sum_{k=1}^{N} m_k v_{k,\alpha} v_{k,\beta}, \tag{12}$$

V is the volume of the simulation cell.

The static stress σ^{static} can be calculated analytically for a given atomic configuration from the NN potential derivatives. We define

$$R_{ij,\alpha} = R_{i,\alpha} - R_{j,\alpha}. \tag{13}$$

We first calculate the atomic contributions to the stress tensor separately for the radial and the angular symmetry functions. The radial stress contribution of atom i to the static stress matrix element $\sigma^{\text{static}}_{i,\alpha\beta}$ is

$$\sigma^{\text{static,rad}}_{i,\alpha\beta} = \sum_{j=1}^{N} R_{ij,\alpha} \cdot F_{j,\beta}. \tag{14}$$

The angular stress contribution of atom i is

$$\sigma^{\text{static,ang}}_{i,\alpha\beta} = \sum_{j=1}^{N} R_{ij,\alpha} \cdot F_{j,\beta} + \sum_{k=1}^{N} R_{ik,\alpha} \cdot F_{k,\beta}. \tag{15}$$

The final matrix element of the static stress tensor is then obtained by adding all atomic contributions,

$$\sigma^{\text{static}}_{\alpha\beta} = \sigma^{\text{static,rad}}_{\alpha\beta} + \sigma^{\text{static,ang}}_{\alpha\beta} = \sum_{i=1}^{N}\left(\sigma^{\text{static,rad}}_{i,\alpha\beta} + \sigma^{\text{static,ang}}_{i,\alpha\beta}\right). \tag{16}$$

In Eqs. (14) and (15) we have to take into account the mapping of the Cartesian coordinates on the symmetry functions. For the radial stress we obtain

$$\begin{aligned}
\sigma^{\text{static,rad}}_{\alpha\beta} &= \sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha} \cdot F_{j,\beta} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha} \cdot \frac{\partial E}{\partial R_{j,\beta}} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha} \sum_{k=1}^{N} \frac{\partial E_k}{\partial R_{j,\beta}} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha} \sum_{k=1}^{N}\sum_{\mu=1}^{M_k} \frac{\partial E_k}{\partial G_{k,\mu}} \cdot \frac{\partial G_{k,\mu}}{\partial R_{j,\beta}} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\mu=1}^{M_k} R_{ij,\alpha} \frac{\partial E_k}{\partial G_{k,\mu}} \cdot \frac{\partial G_{k,\mu}}{\partial R_{j,\beta}} \\
&= -\sum_{k=1}^{N}\sum_{\mu=1}^{M_k} \frac{\partial E_k}{\partial G_{k,\mu}} \cdot \sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha} \frac{\partial G_{k,\mu}}{\partial R_{j,\beta}},
\end{aligned} \tag{17}$$

and for the angular stress correspondingly

$$\sigma_{\alpha\beta}^{\text{static,ang}} = \sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha}\cdot F_{j,\beta} + \sum_{i=1}^{N}\sum_{m=1}^{N} R_{im,\alpha}\cdot F_{m,\beta}$$
$$= -\sum_{k=1}^{N}\sum_{\mu=1}^{M_k}\frac{\partial E_k}{\partial G_{k,\mu}}\cdot\left(\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij,\alpha}\frac{\partial G_{k,\mu}}{\partial R_{j,\beta}} + \sum_{i=1}^{N}\sum_{m=1}^{N} R_{im,\alpha}\frac{\partial G_{k,\mu}}{\partial R_{m,\beta}}\right). \quad (18)$$

In principle, also analytic higher derivatives can be calculated.

V. RESULTS AND DISCUSSION

A. A one-dimensional model potential

As a first example for applying these symmetry functions we use a simple one-dimensional model function $E(R)$ exhibiting several local minima,

$$E/\text{eV} = \frac{\cos(5\cdot R/\text{Bohr}) + (R/\text{Bohr} - 4.5)^2}{5} - 1. \quad (19)$$

This function and its negative gradient $F(R)$, the force, are plotted in Fig. 7 ("E reference" and "F reference"). It is certainly more demanding to fit this function than a typical Lennard-Jones-like pair potential.

First, we fitted this function by a standard feed-forward NN using R as the only coordinate in a single input node. The function is represented by an equidistant grid of 71 points in the range 1.0 Bohr $< R <$ 8.0 Bohr. We then determined the required number of hidden layers and nodes per hidden layer and found a 1-25-25-1 NN to be sufficient to obtain a very high accuracy over the full range of R-values. Using only the function value for fitting we obtained a root mean squared error (RMSE) of 0.00010 eV for E and of 0.00449 eV/Bohr for F. If the derivatives of the function are also used for the fitting, the corresponding RMSEs are 0.00005 eV and 0.00023 eV/Bohr, respectively. In both cases, the fit is very close to the reference function.

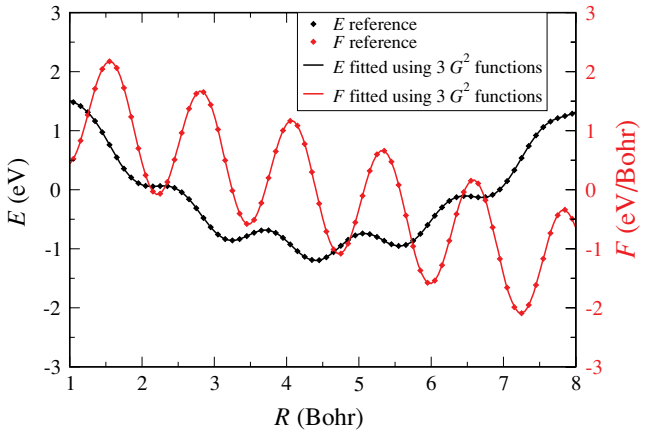

FIG. 7. Comparison of the one-dimensional model function defined in Eq. (19) (black diamonds) and its negative gradient (red diamonds) with the function (black line) and gradient (red line) obtained from a NN fit.

Now we investigate, if it is possible to fit this function using the high-dimensional NN scheme and a suitable set of symmetry functions instead of coordinate R. In the notation of the high-dimensional NN approach the function can be considered as some fictitious potential of a homonuclear dimer. The chemical environments of both atoms are the same and consist of a single atom at distance R. Using a set of three radial symmetry functions of type G^2 ($\eta_1 = 1.00$ Bohr^{-2}, $\eta_2 = 0.10$ Bohr^{-2}, and $\eta_3 = 0.01$ Bohr^{-2}; $R_c = 12.0$ Bohr and $R_s = 0.0$ Bohr for all three functions), we obtain an RMSE of 0.00027 eV for E and of 0.00506 eV/Bohr for F if only the function values are used for fitting. If the gradients are also included in the fit, the RMSEs are 0.00003 eV and 0.00075 eV/Bohr, respectively. The same NN architecture as above has been used.

The $E(R)$ and $F(R)$ curves obtained in the latter fit are plotted in Fig. 7 and compared to the reference function and derivative values. It can be seen that an excellent agreement has been obtained. Thus, although the three radial symmetry functions are certainly more complicated than necessary in the present example, we have shown that they provide the same fitting accuracy as the original variable R. Certainly, fitting a typical dimer potential with a more physical form, i.e., only one minimum, will be significantly simpler than the chosen model function. However, in contrast to the coordinate R, the symmetry functions are general and allow to describe much more complicated systems. This has been demonstrated by a number of successful applications to high-dimensional PESs of condensed systems containing large numbers of atoms.[13,14,16,17]

B. Fitting with inappropriate symmetry functions

Neural networks are able to represent high-dimensional functions with high accuracy. A necessary condition for the applicability of this method to PESs is, therefore, the existence of a unique energy value for any given point in coordinate space. If contradictory data are present, e.g., in the form of different energies assigned to the same set of coordinates, this condition is not fulfilled and the NN will not be able to fit the data properly. This has to be kept in mind when the symmetry functions are constructed, since they represent the only coordinates visible to the NN.

In this Section we will use two simple model systems, the Lennard-Jones dimer and the Lennard-Jones trimer (confined to equilateral triangles only, $R_{ij} = R_{ik} = R_{jk}$), to discuss the consequences of unsuitable symmetry function sets. The Lennard-Jones potential is given by

$$E = 4\epsilon\left[\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}\right]. \quad (20)$$

In the following tests we set the parameters $\sigma = 1.0$ Bohr and $\epsilon = 1.0$ eV. The potential then has a value of 0.0 eV at $R = 1.0$ Bohr, and a minimum of $E = -1.0$ eV at $R = 2^{1/6} \approx 1.12246$ Bohr. The well depth is of the typical order of magnitude of a chemical bond and will allow to assess the accuracy of the NN fit.

In Fig. 8 the LJ energies for the dimers and trimers are plotted as a function of the interatomic distance R (diamond

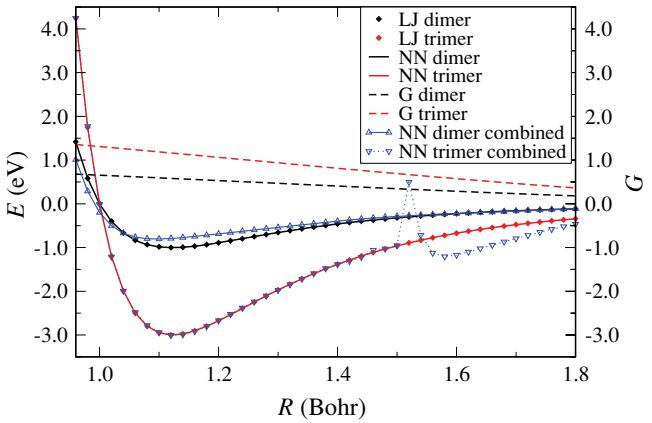

FIG. 8. Neural network fits of Lennard-Jones (LJ) dimer and trimer energies (diamond symbols) as a function of the interatomic distance R. The cluster structures are described by a single radial symmetry function only, whose value G is shown separately for the two cluster sizes (dashed curves). If just one cluster size is fitted, the NN can reproduce the energies very accurately, because for each symmetry function value there is a unique energy value. If the dimers and trimers are fitted simultaneously (blue triangles, "combined"), the trimers with $R < 1.5$ Bohr are reproduced well, because their symmetry function values are still unique ($G > 0.67$). All dimers in the data set have lower symmetry function values. For larger R there are dimers with the same symmetry function values ($0 < G < 0.67$) but different energies resulting in poor fits for these trimers and all dimers.

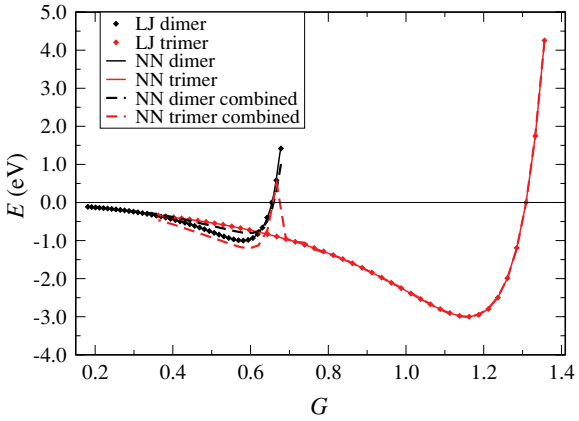

FIG. 9. Plot of the Lennard-Jones energies (diamond symbols) of dimers and trimers as a function of the radial symmetry function value G. The dimer and trimer curves can be fitted separately very well (lines). However, for $G < 0.67$ there are two different LJ energies for the same G. Consequently, the neural network energies are not accurate for this range if dimers and trimer are fitted together ("combined" curves), but represent an averaged energy per atom. In the combined fit, the NN dimer, and trimer energies differ by a factor 1.5 due to the different number of atoms [cf. Eq. (3)].

symbols). Now we use a single radial symmetry function (type G^1, $R_c = 2.5$ Bohr) to describe these structures. The symmetry function values are shown as dashed curves for both clusters in Fig. 8. They decrease monotonously as a function of the interatomic distance, so there is a unique relation between the distance and the symmetry function value. Separately, each data set, dimer or trimer, therefore can be fitted very well by NNs (solid curves). However, if the dimer and trimer data sets are combined and fitted simultaneously, the fitting accuracy is strongly reduced (blue triangles). Investigating this in more detail, we find that the trimer is still very accurately represented for $R < 1.5$ Bohr, but for larger interatomic distances there is a jump in the energy and the energy representation is poor beyond this point. For the dimer curve the effect is less drastic, but the precision of the energies is clearly reduced for the full range of bond lengths. The reason for this result is that for the well-represented trimers with short interatomic distances there are no dimer structures within the same range of symmetry function values. For the given dimer data set, the maximum symmetry function value being present is about 0.67. If also shorter distances would be included in the dimer data set, the maximum value G^1 could obtain is 1.0, because in the dimer each atom has just one neighbor. The theoretical range of values for G^1 in the trimer case is $[0, 2]$, because each atom has two neighbors. Therefore, there is a range of symmetry function values for the trimer, which does not overlap with dimer values. For the present example this is the case for $G^1 > 0.67$. For $G^1 < 0.67$ the NN has to fit dimer and trimer energies, which are different and thus contradictory, to the same symmetry function values. Therefore, the NN produces averaged atomic energies, and the errors of these points are large.

Reinvestigating the trimer curve of the combined fit in Fig. 8 shows that for $G^1 > 0.67$ the trimer curve resembles a compressed dimer curve shifted to larger interatomic distances. To further analyze this phenomenon, in Fig. 9 the dimer and trimer energies are plotted as a function of the symmetry function value. This corresponds to the structural description available to the NN, which does not obtain any information about R. For G^1 values larger than 0.67 there are only trimer energies, but below this value, different dimer and trimer energies are assigned to the same symmetry function value. In this plot it is clearly visible that the jump in the trimer energy around $G^1 = 0.67$ is caused by the attempt of the NN to simultaneously fit the shape of the dimer curve. For lower values of G^1 the dashed curves of the NN dimer energy curve and the NN trimer energy curve seem to be different, but this is only due to the different number of atoms. According to Eq. (3), the total energy is a sum of atomic energy contributions. These atomic energy contributions (depending just on G^1) are the quantities fitted by the NN and are exactly identical for the dimer and the trimer in this range. The effective factor of 1.5 between the dimer and the trimer energy curve is due to the ratio of 2 : 3 in the atom numbers.

The solution for this problem is straightforward. Instead of using just one radial symmetry function, a set of several radial symmetry functions can be used, which allows to distinguish all dimers and trimers properly. The symmetry functions must be linearly independent, which is given by definition for all symmetry functions defined in Sec. IV B. In Fig. 10 three radial symmetry functions have been used (G^1, $R_c = 2.0$, 2.5, and 3.0 Bohr). They are shown as dashed curves. This vector of three symmetry function values is unique for each structure and allows to distinguish dimers from trimers. Consequently, both, the dimer and the trimer energy curves are now very accurately represented by the same NN fit. It should be noted, however, that while the

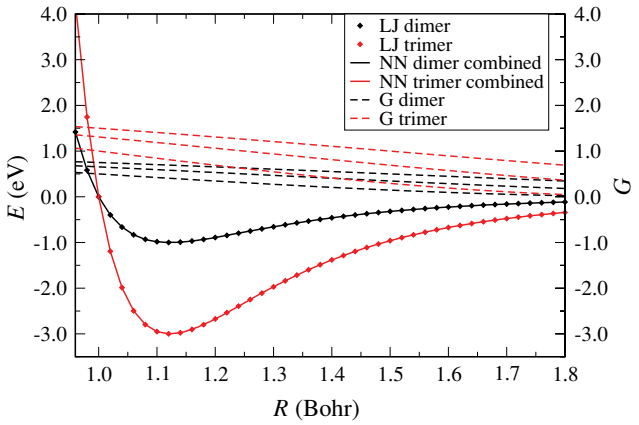

FIG. 10. Neural network fits of Lennard-Jones dimer and trimer energies as a function of the interatomic distance R. Three radial functions have been used to describe the atomic environments in the clusters. The resulting set of three symmetry function values is different for all dimers (black dashed curves) and trimers (red dashed curves). Therefore the NN is able to distinguish all structures and assign the energies correctly to both clusters over the full range of bond lengths.

solution for this model system is straightforward, for complex, high-dimensional systems involving many symmetry functions, identifying and solving problems of this kind is often a nontrivial task.

C. Identifying deficient sets of symmetry functions

Less accurate fits can have a number of reasons, e.g., numerical noise in the reference electronic structure data, too small NN architectures, convergence to poor local minima, an inappropriate choice of initial weights and many others. In Sec. V B we have seen that the choice of the set of symmetry functions is also of crucial importance to obtain good fits. The set of functions must provide a sufficient structural distinction of inequivalent atomic environments to avoid fitting contradictory data. In order to investigate if a set of symmetry functions is in principle appropriate, we need to establish a method to identify contradictory data. As the symmetry functions describe atomic environments, we need target properties, which can be assigned to individual atoms. Here, we use the absolute force acting on the atom i, $F_i = |\mathbf{F}_i|$. Consequently, we have to investigate now if for a given set of symmetry function values, different and thus contradictory, force values exist in the training data set. For this purpose we have to compare vectors of symmetry function values. Only if they are sufficiently different, the atomic forces may differ. We use the fact that the sets of symmetry function values of two atoms ($\{G_i\}$ and $\{G_j\}$) are the same, if the absolute value of the difference vector $\Delta G = |\mathbf{G}_i - \mathbf{G}_j|$ is zero. For a suitable set of symmetry function values, for $\Delta G = 0$ there should be a unique force F, or equivalently, $\Delta F = |\mathbf{F}_i - \mathbf{F}_j| = 0$. For $\Delta G > 0$, nonzero values of ΔF are acceptable, because the same positive ΔG values can be obtained for different atomic environments of atom j. In general, the larger ΔG, the larger is the acceptable scattering of ΔF values. A set of symmetry functions can now be constructed systematically by adding step by step more symmetry functions until this criterion is fulfilled.

We demonstrate this by fitting a set of 45 000 Lennard-Jones trimer structures. This set of structures has been splitted into a training set (36 000 structures) used to optimize the weight parameters of the NN, and an independent test set (9000 structures) to check on the generalization properties of the fits for structures not included in the training set. We use three symmetry functions to describe the atomic environments. All are of type G^2 (η=1.0, 0.5, and 0.2 Bohr^{-2}; R_s=0.0 Bohr; R_c=12.0 Bohr). Three symmetry function sets are used, the first contains just the first symmetry function, the second two symmetry functions, while set three contains all three functions. The fits have been done using the energies only and using the energies and forces. If only the energies are used for fitting, one piece of information is available per structure. In case of force fitting, each force component of each atom provides additional information so that in total 10 pieces of information per cluster are available. Two different NN architectures have been employed (X-10-10-1 and X-15-15-1, $X = \{1, 2, 3\}$). The RMSEs of the energies and forces of the training and test sets are listed in Table I. It can be seen that the fit using just one symmetry function is much worse than the fits employing the larger symmetry function sets. This holds for both NN architectures. Using the gradients generally improves the quality of the forces at the cost of slightly worse energies.

TABLE I. Root mean squared errors of 45 000 Lennard-Jones trimer structures obtained with different sets of symmetry functions. Two different NN architectures have been used.

Energy fit					
		Energy RMSE (eV/atom)		Force RMSE (eV/Bohr)	
Symmetry function set	Number of weights	train	test	train	test
		X-10-10-1 NN			
1	141	0.0354	0.0374	0.8057	0.7912
2	151	0.0040	0.0038	0.1059	0.1075
3	161	0.0025	0.0022	0.0564	0.0533
		X-15-15-1 NN			
1	286	0.0365	0.0377	1.1962	1.1009
2	301	0.0018	0.0016	0.0587	0.0556
3	316	0.0016	0.0013	0.0405	0.0388
Energy+Force fit					
		Energy RMSE (eV/atom)		Force RMSE (eV/Bohr)	
Symmetry function set	Number of weights	train	test	train	test
		X-10-10-1 NN			
1	141	0.1300	0.1299	0.5339	0.5344
2	151	0.0080	0.0062	0.0292	0.0277
3	161	0.0053	0.0047	0.0216	0.0212
		X-15-15-1 NN			
1	286	0.0741	0.0748	0.1755	0.1835
2	301	0.0052	0.0043	0.0245	0.0258
3	316	0.0029	0.0020	0.0129	0.0127

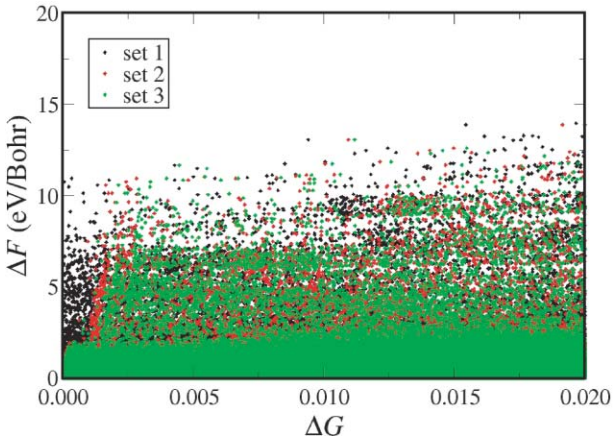

FIG. 11. Analysis of the symmetry functions of a set of 1000 representative Lennard-Jones trimers. The difference of the forces of two atoms i and j is plotted, $\Delta F = |\mathbf{F}_i - \mathbf{F}_j|$ vs. the difference in the vectors of symmetry function values $\Delta G = |\mathbf{G}_i - \mathbf{G}_j|$. If points with large ΔF but small ΔG are present, the structural description of the atomic environments is not sufficient and poor fits are obtained.

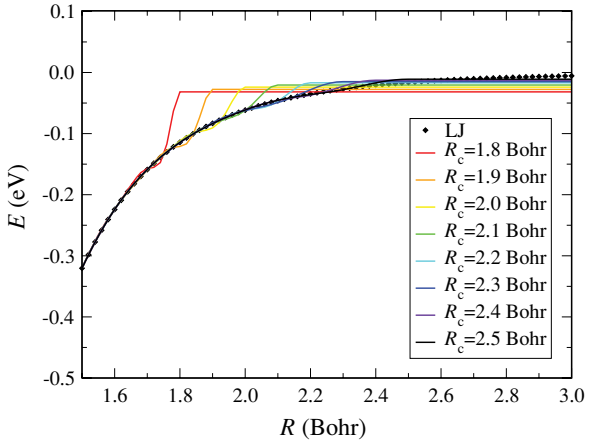

FIG. 12. Fit of the Lennard-Jones potential using a single G^1 symmetry function with a very small cutoff. Beyond the cutoff radius R_c the neural network is not able to represent the potential. Therefore, the predicted NN energy is an average of the reference energies beyond R_c. Just the energies have been used for fitting.

In order to understand why symmetry function set 1 yields poor fits, we compared the symmetry function values and forces of all atoms in all structures. The obtained $(\Delta G, \Delta F)$ data points of 1000 representative LJ timers are plotted in Fig. 11. For data set 1, there are many data points with large ΔF values at small values of ΔG. This indicates that a number of atoms have very similar symmetry function values, but clearly different forces. This contradictory information cannot be fitted by the NN and results in poor fits. If more symmetry functions are added, these contradictory data at low ΔG values disappear. In general, if more symmetry functions are used, the scattering of the ΔF values is more and more reduced. This improved capability do distinguish different structures is a necessary condition to obtain better fits, which are indeed found for large symmetry function sets (cf. Table I).

Finally, we note that of course one coordinate is not sufficient to describe the structure of a trimer, since in general $3N - 6 = 3$ coordinates are needed to describe the positions of the neighbors of each atom. In the special case of pure pair potentials this further reduces to two degrees of freedom. However, in general an appropriate number of symmetry functions is not a sufficient criterion, since they also need to allow for a reliable structural discrimination by a proper choice of parameters.

D. The symmetry function cutoff

The choice of the cutoff radius of the symmetry functions is very important. If a very large cutoff is chosen, many symmetry functions are needed to describe the chemical environments of the atoms within the cutoff. If the cutoff is too small, large fitting errors may arise, because atoms outside the cutoff radius by construction have no effect on the atomic energy contribution. If such structural features are energetically important, their contributions will be treated as noise and only an averaged energy will be obtained. As a consequence, the fitting errors will be high. In general, it is recommended to use a cutoff of at least 5–6 Å to include the relevant part of van der Waals interactions in the potential.

We demonstrate this using a Lennard-Jones dimer potential. In Fig. 12 the LJ potential is fitted using a single symmetry function of type G^1 with different cutoffs. Only the energy is used for fitting. By construction, beyond the cutoff the NN produces a constant energy independent of R. This constant energy corresponds to the average energy of all training points beyond the cutoff to minimize the overall RMSE. Close to the cutoff, these constant energies are connected to the rather good energy fit for $R < R_c$ by a sudden drop in the potential. In Fig. 13 the same fits are presented, but this time also the forces have been included in the fitting. As in the energy fit, beyond the cutoff the energies are constant. Within the cutoff region the energies are worse than for the energy-only fit over a wide range of R values. The reason for this is the competition of the energies and forces in the fit. If in addition to the

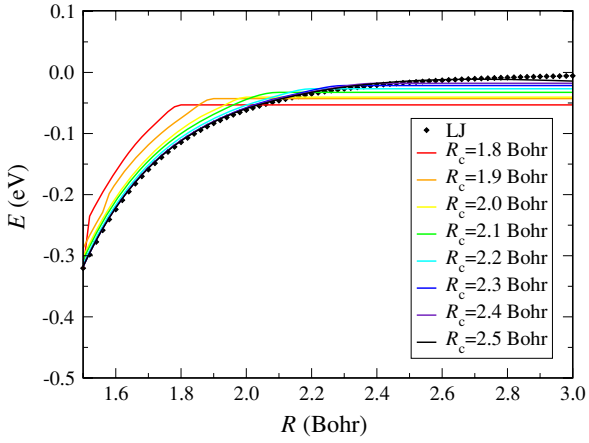

FIG. 13. Fit of the Lennard-Jones potential using a single G^1 symmetry function with a very small cutoff. Beyond the cutoff radius R_c the neural network is not able to represent the potential. The forces and energies have been used for fitting.

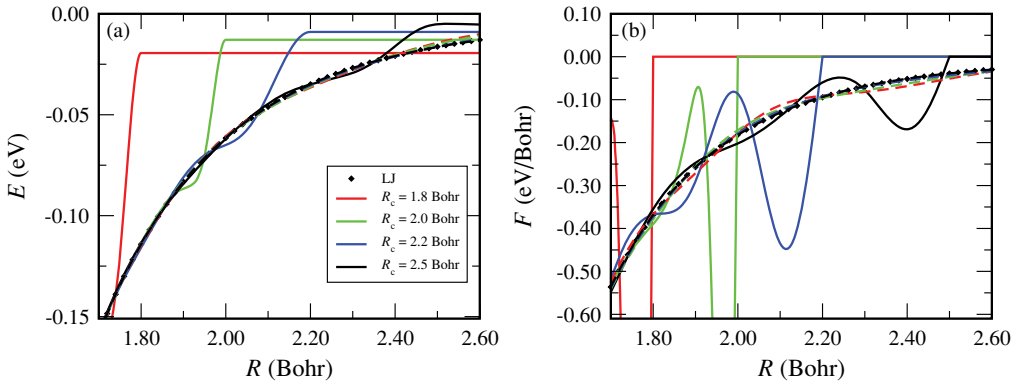

FIG. 14. Comparison of the energy (a) and force (b) curves obtained fitting the Lennard-Jones potential using radial symmetry functions with different effective spatial extensions. The solid curves represent fits using functions of type G^1. The spatial extension is controlled by employing different cutoff radii R_c. The dashed curves represent fits using functions of type G^2 with a comparable spatial extension, which in this case is controlled by parameter η. For the G^2 functions a cutoff of 6.0 Bohr has been used in all cases, R_s is set to 0 Bohr. G^1 and G^2 functions with a comparable spatial extension are given in the same color. The strict cutoff in the G^1 functions results is a jump of the energy to a constant value at the cutoff radius. This jump is also reflected in a poor representation of the forces close to R_c. The results obtained using the G^2 functions are of much better quality and hardly distinguishable from the reference data (black diamonds).

energy RMSE also the force RMSE has to be minimized, the energies are necessarily less accurate. Consequently, the energy curve shows a parallel offset with respect to the training points. This effect is also visible in increased RMSEs for the energy if gradient information is included in the fit. If only energies are used for the fitting, the opposite effect occurs. Then, the forces have a higher error, because the gradients are wrong close to the cutoff radius, as can be clearly seen in Fig. 12 by the large slope. This effect is reduced if the cutoff is increased.

These results show that symmetry functions with small cutoff radii should not be used. The discussed problems can be avoided by replacing the symmetry functions of type G^1 by symmetry functions of type G^2. For the G^2 functions a large cutoff can be used to ensure smooth energy and force curves, while at the same time the effective range of significantly nonzero values can be controlled by a suitable choice of parameter η. This way it is possible to construct a set of radial symmetry function values with different effective radial extensions, but without the discussed cutoff-related problems. In Fig. 14 the energies and forces for fits using G^1 and G^2 functions are compared. Only the energies have been used for the fitting. In each fit just one symmetry function has been used, and the effective range has been controlled by parameter R_c for the G^1 functions and by η for the G^2 functions. The formal cutoff of all G^2 functions has been set to 6.0 Bohr. The solid curves represent the G^1 fits, the dashed curves the G^2 fits employing η values of 1.0, 0.8, 0.6, and 0.38 Bohr^{-2}, respectively. The different colors represent comparable effective spatial extensions, i.e., significant nonzero values of the symmetry functions. In the energy plot in particular for very short effective ranges, the G^2 fits are clearly superior to the G^1 fits, and their energy curves are very close to the training data. In the force plot in Fig. 14 the G^1 function fits show strong oscillations, because matching the constant energies at the cutoff radius enforces strong changes in the gradients. For the G^2 fits, the forces are much more accurate for all effective ranges, although they have not been used for the fitting, because the long tails of the G^2 functions prevent sudden changes in the energy curve.

E. Preconditioning the symmetry function values

The individual symmetry functions used to describe the atomic environments can have a very different range of values. For numerical reasons, it is therefore advantageous to precondition the range of values for each symmetry function.

The most basic operation is to shift the mean value of each symmetry function to zero. Accordingly, the symmetry function number k at atom i, $G_{i,k}$ changes as

$$G'_{i,k} = G_{i,k} - \frac{1}{N}\sum_{j=1}^{N} G_{j,k}. \tag{21}$$

Here, N is the overall number of atoms in the training set. Further, the relative importance of all symmetry functions can be balanced by rescaling the range of values for each G_i to a predefined interval $[S_{\min}, S_{\max}]$ according to

$$G'_i = \frac{G_i - G_{i,\min}}{G_{i,\max} - G_{i,\min}} \cdot (S_{\max} - S_{\min}) + S_{\min}. \tag{22}$$

Here, $G_{i,\min}$ is the smallest value of G_i occurring in any atomic environment, $G_{i,\max}$ is the largest value. The centering and rescaling operations can also be combined.

F. Extrapolation

NN potentials provide a very accurate method to interpolate a given set of reference data from electronic structure calculations. However, due to the very flexible functional form, they are not always reliable if energies and forces for structures are requested, which are very different from the structures in the training set. For example, a NN potential constructed for bulk materials is very likely to fail when applied to small clusters, if no clusters are included in the training set and vice versa.

Therefore, it is essential to detect situations, in which predicted energies and forces might be unreliable. This is likely to happen if the transformation of the atomic positions to symmetry functions yields a set of symmetry function values far from the sets used for training the NN. A straightforward way to detect this extrapolation is to determine for each symmetry

function G_i the minimum value $G_{i,\min}$ and maximum value $G_{i,\max}$ present in the training set. These values define an interval for which no fundamental problems are to be expected, if the training set is sufficiently dense. Then, whenever in the application of the NN potential an energy is requested for a new structure, the symmetry function values of all atoms are calculated and compared with the $G_{i,\min}$ and $G_{i,\max}$. If for at least one symmetry function the value falls outside this interval, an automatic warning can be given by the program that the NN prediction may not be reliable. This warning can also be used to extend the range of validity of the NN potential. Problematic structures can be searched systematically, and then added to the training set to increase the intervals $[G_{i,\min}, G_{i,\max}]$.

Finally, it should be noted that the presence of extrapolation also depends on the choice of coordinates, and a good choice of coordinates can strongly reduce the frequency of extrapolation. To illustrate this, we assume that a periodic function $f(x)$ with period length 2π is fitted as a function of x using training points in the interval $[0, 2\pi]$. Then, extrapolation will occur, whenever function values are requested for $x < 0$ or $x > 2\pi$. However, it is possible to define two symmetry functions $G^1 = \cos(x)$ and $G^2 = \sin(x)$ and fit the training points as a function of the two-dimensional vector $\mathbf{G} = (G^1, G^2)$ instead of x. Once the function $f(G^1, G^2)$ has been constructed, correct function values can be obtained for arbitrary x. First, x is transformed to $\mathbf{G}$, then the function values $f(\mathbf{G}(x))$ are calculated. No extrapolation for any x will occur in this case, since the $\mathbf{G}$ vector for any x is within the known training interval of G^1 and G^2. Note, that using just one function, either G^1 or G^2, is not sufficient for fitting general periodic functions, because each symmetry function value appears twice in the interval, and most likely different values of the target function will be present for both x. Only the combined set of G^1 and G^2 uniquely determines x and is suitable for constructing the NN fit.

VI. CONCLUSIONS

In summary, the properties of symmetry functions suitable for constructing high-dimensional NN potentials have been discussed. Several functional forms for radial and angular symmetry functions describing the local atomic environments have been suggested. All functions are many-body functions depending explicitly on the positions of all atoms in the local chemical environment. They can be used for all types of periodic and nonperiodic systems and are independent of the bonding nature. They are equally applicable to very different systems such as bulk metals, clusters or molecules. Using simple model systems, it has been shown that a careful analysis of the symmetry functions allows to understand possible problems in the fitting process and to obtain NN potentials of improved quality.

Compared to simple analytic potentials, the presented symmetry functions as well as the NN itself have a rather complicated form. While this form is the very reason for the numerical accuracy of the method, its performance is lower than for typical classical force fields. However, the accuracy is close to the reference calculations, which are typically electronic structure methods such as density-functional theory. With respect to these methods the NN potential can be evaluated about 4 to 5 orders of magnitude faster. Therefore, NN potentials enable to extend the time and length scales of MD simulations significantly. This is true in particular if complex bonding situations are present, which cannot be expressed by simple potentials and otherwise would require to carry out *ab initio* MD simulations. In particular, the breaking and formation of bonds poses no restriction to NN potentials, as they depend just on the atomic positions. This opens applications to a variety of interesting systems in materials science, surface science and in solution, which will strongly benefit from the availability of accurate and fast potentials.

ACKNOWLEDGMENTS

Financial support by the DFG (Emmy Noether program), the FCI and the Academy of Sciences of NRW is gratefully acknowledged.

[1] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[2] D. Marx and J. Hutter, Ab initio *Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge, 2009).
[3] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and P. Flannert, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[4] J. Ischtwan and M. A. Collins, J. Chem. Phys. **100**, 8080 (1994).
[5] M. J. T. Jordan, K. C. Thompson, and M. A. Collins, J. Chem.Phys. **102**, 5647 (1995).
[6] M. A. Collins, Theor. Chem. Acc. **108**, 313 (2002).
[7] A. P. Bartok, M. C. Payne, R. Kondor, and G. Csanyi, Phys. Rev. Lett. **104**, 136403 (2010).
[8] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, 1995).
[9] G. Cybenko, Math. Control, Signals Syst. **2**, 303 (1989).
[10] K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
[11] C. M. Handley and P. L. A. Popelier, J. Phys. Chem. A **114**, 3371 (2010).
[12] J. Behler, Chem. Modelling **7**, 1 (2010).
[13] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[14] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, Phys. Status Solidi (b) **245**, 2618 (2008).
[15] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, Phys. Rev. Lett. **100**, 185501 (2008).
[16] H. Eshet, R. Z. Khaliullin, T. Kühne, J. Behler, and M. Parrinello, Phys. Rev. B **81**, 184107 (2010).
[17] R. Z. Khaliullin, H. Eshet, T. Kühne, J. Behler, and M. Parrinello, Phys. Rev. B **81**, 100103 (2010).
[18] RuNNer—A Neural Network Code for High-Dimensional Potential-Energy Surfaces, Jörg Behler, Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, Germany.
[19] J. B. Witkoskie and D. J. Doren, J. Chem. Theory Comput. **1**, 14 (2005).
[20] H. M. Le and L. M. Raff, J. Phys. Chem. A **114**, 45 (2010).
[21] A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, and R. Komanduri, J. Chem. Phys. **130**, 134101 (2009).
[22] F. V. Prudente and J. J. Soares Neto, Chem. Phys. Lett. **287**, 585 (1998).
[23] F. V. Prudente, P. H. Acioli, and J. J. Soares Neto, J. Chem. Phys. **109**, 8801 (1998).
[24] A. C. P. Bittencourt, F. V. Prudente, and J. D. M. Vianna, Chem. Phys. **297**, 153 (2004).
[25] T. M. Rocha Filho, Z. T. Oliveira, Jr., L. A. C. Malbouisson, R. Gargano, and J. J. Soares Neto, Int. J. Quantum. Chem. **95**, 281 (2003).
[26] S. Manzhos, X. Wang, R. Dawes, and T. Carrington, Jr., J. Phys. Chem. A **110**, 5295 (2006).
[27] S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **125**, 84109 (2006).
[28] S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **125**, 194105 (2006).

[29]S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **127**, 014103 (2007).
[30]S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **129**, 224104 (2008).
[31]M. Malshe, R. Narulkar, L. M. Raff, M. Hagan, S. Bukkapatnam, P. M. Agrawal, and R. Komanduri, J. Chem. Phys. **130**, 184102 (2009).
[32]M. Malshe, R. Narulkar, L. M. Raff, M. Hagan, S. Bukkapatnam, and R. Komanduri, J. Chem. Phys. **129**, 044111 (2008).
[33]T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129 (1995).
[34]S. Lorenz, A. Groß, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).
[35]S. Lorenz, M. Scheffler, and A. Groß, Phys. Rev. B **73**, 115431 (2006).
[36]J. Behler, S. Lorenz, and K. Reuter, J. Chem. Phys. **127**, 014705 (2007).
[37]J. Behler, K. Reuter, and M. Scheffler, Phys. Rev. B **77**, 115421 (2008).
[38]J. Ludwig and D. G. Vlachos, J. Chem. Phys. **127**, 154716 (2007).
[39]D. A. R. S. Latino, R. P. S. Fartaria, F. F. M. Freitas, J. Aires-de-Sousa, and F. M. S. Silva Fernandes, J. Electroanal. Chem. **624**, 109 (2008).
[40]C. Carbogno, J. Behler, K. Reuter, and A. Groß, Phys. Rev. B **81**, 035410 (2010).
[41]P. M. Agrawal, L. M. Raff, M. T. Hagan, and R. Komanduri, J. Chem. Phys. **124**, 124306 (2006).
[42]J. Tersoff, Phys. Rev. B **39**, 5566 (1989).
[43]We thank an unknown referee for suggesting the following cutoff function with continuous first and second derivatives:

$$f_c(R_{ij}) = \begin{cases} \tanh^3\left[1 - \frac{R_{ij}}{R_c}\right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c. \end{cases}$$

[44]S. Hobday, R. Smith, and J. Belbruno, Modell. Simul. Mater. Sci. Eng. **7**, 397 (1999).
[45]S. Hobday, R. Smith, and J. BelBruno, Nucl. Instrum. Methods Phys. Res. B **153**, 247 (1999).
[46]A. Bholoa, S. D. Kenny, and R. Smith, Nucl. Instrum. Methods Phys. Res. B **255**, 1 (2007).
[47]E. Sanville, A. Bholoa, R. Smith, and S. D. Kenny, J. Phys. Condens. Matter **20**, 285219 (2008).
[48]H. Gassner, M. Probst, A. Lauenstein, and K. Hermansson, J. Phys. Chem. A **102**, 4596 (1998).
[49]M. Parrinello and A. Rahman, Phys. Rev. Lett. **45**, 1196 (1980).
[50]M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[51]R. Martoňák, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 75503 (2003).