Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface

Sönke Lorenz and Matthias Scheffler
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin-Dahlem, Germany

Axel Groß
Abteilung Theoretische Chemie, Universität Ulm, D-89069 Ulm, Germany
(Dated: October 6, 2005)

A neural network (NN) approach is proposed for the representation of six-dimensional *ab initio* potential-energy surfaces (PES) for the dissociation of a diatomic molecule at surfaces. We report tests of NN representations that are fitted to six-dimensional analytical PESs for H_2 dissociation on the clean and the sulfur covered Pd(100) surfaces. For the present study we use high-dimensional analytical PESs as the basis for the NN training, as this enables us to investigate the influence of phase space sampling on adsorption rates in great detail. We note, however, that these analytical PESs were obtained from detailed DFT calculations. When information about the PES is collected only from a few high-symmetric adsorption sites, we find that the obtained adsorption probabilities are not reliable. Thus, intermediate configurations need to be considered as well. However, it is not necessary to map out complete elbow plots above non-symmetric sites. Only a few energies in the region where the molecular bond breaks need to be considered. With this understanding, the required number of NN training energies for obtaining a high-quality PES that provides a reliable description of the dissociation and adsorption dynamics is orders of magnitude smaller than the number of total-energy calculations needed in traditional *ab initio* on the fly molecular dynamics. Our analysis also demonstrates the importance of a reliable, high-dimensional PES to describe reaction rates for dissociative adsorption of molecules at surfaces.

PACS numbers: 68.35.Ja, 82.20.Kh, 68.43.-h

I. INTRODUCTION

Theoretical studies of reaction dynamics at surfaces, like the dissociative adsorption of diatomic molecules on metal surfaces, require knowledge of the potential energy of the moving nuclei taking part in the process [1]. Density-functional theory (DFT) total-energy calculations have proven to be a powerful tool to calculate such properties [2–5]. *Ab initio* MD simulations, where the potential and the forces are determined by DFT, are computationally very elaborate and costly. A theoretical simulation of the dissociation probability of a molecule on a surface for different initial energies might require up to 10^7 evaluations of the potential-energy and the forces. Due to the high computational task, *ab initio* molecular dynamics hardly allow the determination of reaction probabilities and so far are limited to dynamical studies of only a few trajectories [6–9].

In order to reduce the computational burden and make a simulation of the sticking probably feasible, Gross and Scheffler had proposed and implemented a three step approach [10, 11]. First, one determines the *ab initio* potential energy surface (PES) on a mesh of several hundred configurations using DFT. In a second step an analytical function is fitted to these points. The last step consists of molecular dynamics calculations on this continuous representation of the *ab initio* PES. The crucial part of this approach is choosing the appropriate analytical function for the interpolation of the total energies. The interaction of a diatomic molecule with a well-defined surface is at least six-dimensional, corresponding to the six degrees of freedom of the molecule and a fixed substrate. The latter assumption is often fulfilled for densely packed metal surfaces. However, on Si(100) for instance, the rearrangement upon adsorption is indeed crucial for the adsorption and desorption mechanism and one easily arrives at 12 and more dimensions [7].

The fitting of a mesh of *ab initio* energies to a continuous representation is a non-trivial task. A high-dimensional, flexible, accurate, reliable and fast interpolation scheme is needed. Ideally this method should be general to allow its application to a wide range of problems. Various approaches to fit a PES can be found in the literature [12–21]. All of the proposed methods have some advantages and some drawbacks. For instance, the fitting of *ab initio* data using analytical functions [10, 22–24] requires an appropriate choice of an analytical form, which is very cumbersome to find in high dimensions. It is therefore fair to say that despite its importance a general, fast, and accurate interpolation tool for PESs is still lacking.

Six-dimensional molecular dynamics calculations based on an analytical interpolation of total energies had shown that dynamical effects as well as a proper statistical sampling can be crucial and differences from a static theory can be significant (see [25, 26] and references therein). Such studies have advanced the understanding of the dissociation dynamics and caused the modification of established concepts. Some phenomena, like the so-called *steering effect* [10], can only be modeled in a theoretical simulation including a sufficiently large number of degrees of freedom. Thus, high-dimensional dynamical studies lead to progress not only in the quantitative, but also in the qualitative understanding of processes on surfaces.

As an alternative to the hitherto proposed fitting schemes we will introduce an interpolation method based on neural networks (NNs) [27–29]. A brief account of this approach has been given elsewhere [30]. Some ideas along these lines had been used earlier by Doren et al. [31]. NNs can in principle approximate any continuous function to arbitrary accuracy [32, 33]. They do not require any assumptions about the functional form of the underlying problem. Once the NN representation of the PES has been determined, the evaluation of the potential-energy is cheap and the derivatives, the forces, are obtainable. Therefore, provided that the number of parameters and required data for a good fit scale favorably with dimension, NNs will be suitable for molecular dynamics applications.

In order to learn more about the advantages, difficulties, and limitations of a NN representation of a high-dimensional PES it is important to analyze realistic test problems. Analytical PESs provide ideal test cases for various reasons. They are fast to evaluate and therefore allow us to study the influence of the data sampling on the quality of the NN fit in great detail. Furthermore, they have been successfully used for the *ab initio* description of the hydrogen dissociation on metal surfaces using a six-dimensional PES [10, 22, 24, 34–41]. Moreover, as an additional check of the accuracy of the obtained NN model we are able to compare the NN-MD results to calculations performed on the analytical PES. We have chosen analytical PESs for the clean [10] as well as the sulfur covered [42] Pd(100) surface as test problems.

The structure of this paper is as follows. Section II introduces the concept of artificial NNs. Section III describes the test of the NN interpolation ability for the dissociation of hydrogen on the clean palladium surface, and Sec. IV reports the interpolation of the second test problem, the analytical PES for the dissociation of hydrogen on a sulfur covered Pd(100) surface. The paper concludes with a summary in Sec. V.

II. ARTIFICIAL NEURAL NETWORKS

Neural Networks can be considered as general, nonlinear fitting functions that do not require any assumptions about the functional form of the underlying problem [27–29]. The main area of research in neural computing is devoted to classification or pattern recognition problems which is a profoundly different task from the interpolation of a multidimensional function. However, NNs have also been applied to problems involving function approximation in general [43, 44] and more recently to the interpolation of potential-energy surfaces [31, 45–50]. These works had concentrated on low-dimensional studies of the PES of molecules in the gas phase. We will extend the approach to study chemical reactions of molecules on surfaces on a high-dimensional PES and in particular to employ it in extended molecular dynamics simulations.

It is important to notice that there is no such thing like "the neural network". NNs are rather a class of algorithms inspired from neuro-science. Different architectures exist. A number of optimization algorithms are applicable for the optimization of the NN parameters. Different basis functions can be used in the interpolation. The number of parameters necessary to obtain a satisfactory representation and how to sample the points used for the interpolation in order to obtain a satisfactory representation are *a priori* unknown.

A. Neural Network Structure

An artificial NN consists of a number of artificial neurons or *nodes*, typically arranged in layers, and interconnected via a set of links. A schematic representation of such a net is plotted in Fig. 1. Each link multiplies its input by a parameter, the weight, before supplying it to a new node. Each node sums over its inputs and applies a function to the resulting value. In the *input layer* the identity function is used to distribute the information to the second layer. This layer is called the *hidden layer* because its input and output is not visible from to the outside world. The hidden layer is the core of the nonlinear fitting of the data set. The *output layer* collects the information from the hidden layer and transforms it again. This network design, in which every node is connected to every node in the adjacent layers but nodes in the same layer are not connected and the information is transmitted only in one direction, is called a *multilayer feed-forward NN*.

The output of a fully connected three layer NN with one input, one hidden and one output layer and n_0, n_1 and n_2 nodes in each layer, respectively, can be written as:

$$y_k^{(2)} = f_2 \left(w_{0k}^{(2)} + \sum_{j=1}^{n_1} w_{jk}^{(2)} f_1 \left(w_{0j}^{(1)} + \sum_{i=1}^{n_0} w_{ij}^{(1)} y_i^{(0)} \right) \right), \tag{1}$$

$\forall\ k = 1, 2, \ldots, n_2$. The *bias* weights like $w_{0k}^{(2)}$ act as an adjustable offset of the activation function. In the case of fitting a PES with a NN, the $\{y_i^{(0)}\}$ represent the coordinates of the reactants. For a diatomic molecule the input layer of the net consist e.g. of six units corresponding to $\{y_i^{(0)}\} = \{X_c, Y_c, Z_c, r, \theta, \phi\}$. In the output layer we will have just one output node, the potential-energy $V_{pot}(\{y_i^{(0)}\})$.

For convenience we have presented in Fig. 1 a feed-forward network with only one hidden layer. However, it should be noted that the number of hidden layers is not

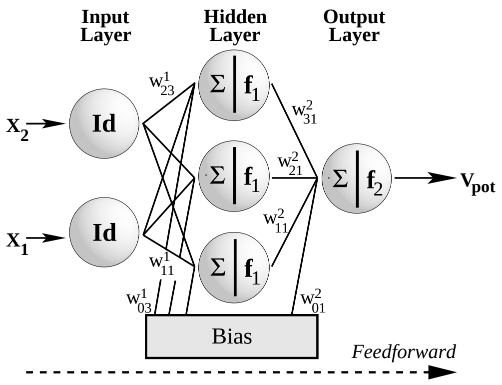

FIG. 1: Schematic architecture of a feed-forward NN. The neurons are arranged in layers. The output function of this network with non-linear basis functions $f_{1,2}(x)$ is: $V_{pot}(x_1, x_2) = f_2\left(w^{(2)}_{01} + \sum_{j=1}^{3} w^{(2)}_{j1}\; f_1\left(w^{(1)}_{0j} + \sum_{i=1}^{2} w^{(1)}_{ij} x_i\right)\right)$.

restricted. In fact, we will most often use networks with two hidden-layers.

Non-linear activation or basis functions are what give NNs their non-linear capabilities. The function must be differentiable for the optimization of the parameters and we normally want it to saturate at both extremes. For example, Gaussian functions, sinusoidal functions or Fermi-like functions can be used. The most common forms of the so-called activation functions are the monotonically increasing sigmoidal or Fermi-like functions, like the sigmoid $f(x) = 1/(1+e^{-x})$ or the hyperbolic tangent function $f(x) = tanh(x)$. Tests on different activation functions are presented in the Section III.

In order to describe the network architecture in a simple way the following notation is used: the number of nodes in the layers, followed by letters denoting the activation function, with s for sigmoid, l for linear, and t for the hyperbolic tangent. In this notation, the network in Fig. 1 in conjunction with a hyperbolic tangent function in the hidden layer and a linear function in the output layer has a $\{2-3-1\; tl\}$ structure.

B. Optimization of the Network Weights

The training of a feed-forward NN is equivalent to performing a non-linear optimization of the network parameters, the weights. In so-called supervised learning the optimization is done by comparing the output with known correct answers. In the case of fitting a PES, the known answers are the energies obtained from *ab initio* total energy calculations. The optimization of the network weights is performed by some iterative optimization scheme until a desirable solution measured by the sum of the squared residuals between the true or targeted value and the actual output of the network depending on the inputs and the weights is reached. In order to minimize the costs the network cycles repeatedly through the following steps of the learning process: (1) present the network one example of the set of data, (2) measure the response of the output layer of the net, (3) calculate the mean squared error between the output and the target value, (4) adjust the weights to minimize the cost function, (5) if the root mean squared error (RMSE) reaches a desired lower bound, stop the iteration, otherwise go back to (1).

Two different update schemes of the parameters in (4) exist. One can first present the network the whole set of examples, called an *epoch*, and only then change the weights accordingly, known as *batch* or *off-line* learning, or the update is performed after the presentation of every single example, chosen randomly. This is called *stochastic* or *on-line* learning. There are several advantages of stochastic over batch learning. It results most likely in better solutions, because updating the weights after each example increases the probability of getting out of a local minimum of the error surface before the iteration gets stuck [27–29]. Optimization methods like steepest descent and conjugate gradients can be used for off-line learning.

We tested several optimization algorithms like steepest descend, conjugate gradients and the Extended Kalman Filter (EKF) [51]. The EKF can be viewed as an iterative second-order or quasi-Newton optimization algorithm and has recently been applied to the optimization of the weights in NNs [52–54]. We found that the EKF algorithm is clearly superior to the other methods. It leads to smaller values of the error function which are in addition reached faster than in other algorithms.

We will therefore use on-line learning with the Kalman filter as the optimization scheme. We employ its adaptive version [53]. Since some parts of the PES may be more interesting than others we have altered the Kalman Filter algorithm to allow individual weighting of each training example.

For training a NN we split the data set into a training and a test set. We optimize the weights only with respect to the training set but monitor the error on the test set during optimization. The error on the training set will decrease, whereas the error on the test set will first decrease and then increase. We stop training as soon as the error on the validation set is higher than it was before. It is here that the network weights provide the best generalization ability, i.e. the network does not only represent the fitted data set very well but is also able to predict new data points reliably.

III. NEURAL NETWORK TEST: 6-D ANALYTICAL PES FOR H_2/Pd(100)

Analytical PESs for the sticking of H_2 on metal surfaces provide ideal test cases for the NN approach for different reasons. First of all, the energy of an analytical PES is fast to evaluate. This allows us to study the influence of the sampling of the data points on the quality of the NN-approximation as measured by the root mean squared error (RMSE) in great detail. Secondly, analytical PESs have proven to describe such adsorption events reliably [10, 24, 34]. Furthermore, as an additional check of the accuracy of the obtained NN-model besides the RMS-error, we are able to compare the results of classical molecular dynamics (MD) calculations using the NN representation to MD-calculations performed on the analytical PES. Namely, we can use the sticking probability - calculated with the NN and the analytical PES - as a further, and most important, test of the accuracy of the approximation.

A. *Ab initio* and analytical PES

The PES of hydrogen dissociation on the clean palladium surface, H_2/Pd(100), has been calculated by Wilke and Scheffler by DFT [55, 56]. The dissociation is non-activated, i.e. pathways to dissociation exist with no energy barrier, and the molecule can freely dissociate above certain sites. The *ab initio* PES has been mapped out following the usual approach of calculating 2D cuts through configuration space above high-symmetric geometries. The equilibrium position of a hydrogen atom is the surface hollow site with a small adsorption height of 0.1 Å above the topmost palladium layer. The minimum pathway for the dissociation of H_2 molecules is above the bridge site with the H-atoms oriented towards the hollow site.

The *ab initio* PES has been fitted with analytical functions by Groß, Wilke, and Scheffler [10], and expressed as a function of the six degrees of freedom of the hydrogen molecule, keeping the surface geometry fixed: $V(X_c, Y_c, Z_c, d, \theta, \phi)$, where X_c, Y_c and Z_c are the centre of mass coordinates of the hydrogen molecule, d is the distance between the two hydrogen atoms, θ and ϕ are the polar and azimuthal angles of the molecule. The potential in the Zd plane is described in reaction path coordinates s along the reaction path and r perpendicular to it [10, 57]. The fit has been performed by a least square method such that the difference between the analytical potential $V(X_c, Y_c, s, r, \theta, \phi)$ and the *ab initio* total energies, which have been calculated for more than 250 configurations, on the average is smaller than 25 meV. In Fig. 2 two cuts through the six-dimensional configuration space of the analytical interpolation have been plotted.

Figure 2(a) shows the analytic interpolation of the minimum path. The solid line marks the dissociation pathway, it exhibits no barrier towards dissociation. However, if we turn the molecule by 90°, keeping the molecular axis parallel to the surface, a distinct energy barrier of $E_{barr} > 0.5$ eV exists (Fig. 2(b)). Only one of the six coordinates has been changed and a qualitatively different dissociation behavior of the molecule has been obtained. Both elbow plots differ only in the small region of the PES where the bond of the hydrogen molecule breaks. The entrance channels with the center of mass of the

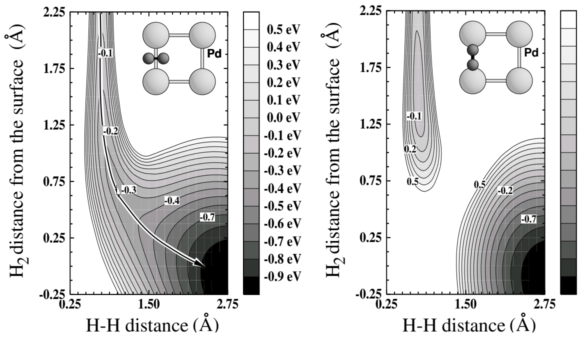

FIG. 2: (a) Hollow - Bridge - Hollow and (b) Top - Bridge - Top. Contour plots through the 6-D analytical PES of H_2/Pd(100) from [10]. Insets: Configuration of the dissociation pathways. Solid line in (a): minimum path towards dissociation. Energy spacing of the contour lines: 0.1 eV.

molecule more than $Z = 1$ Å above the surface are very similar, as well as the exit channels with a molecular bond length $r > 1.50$ Å. The crucial bond-breaking process of the molecule takes place in a relatively small region of the PES. Throughout the following, we will often refer to these two-dimensional cuts through the configuration space, but one should keep in mind that the complete PES is six-dimensional.

B. Tests of the activation functions

For a test of the different choices of activation functions like Gaussians, trigonometric functions or Fermi-like functions for the approximation of PESs a training set of 1,560 and a test set of 7,200 examples of the 6D-analytical PES have been used. Gaussian functions proved to be unsatisfactory for the given problem and structure of the NNs. For sine functions a number of 453 parameters was necessary to achieve a RMSE of the training set below 0.1 eV. However, a test set error of 0.49 eV reflected a poor generalisation ability. The output function was globally not smooth enough. A higher number of parameters lead to a further worsening of the generalization capability.

With sigmoidal or Fermi-like functions we were able to obtain a training error of 0.004 eV and a test error of around 0.16 eV. We will therefore use this group of activation functions throughout the following. However, the fit required the use of a large number of parameters, i.e. around $3,000$, leading to longer fitting times. This can be explained by the form of the dissociation PES. It consists of numerous local bumps in the bond-breaking region, one in each 2D cut of the 6D PES, and is rather smooth elsewhere. In order to form a peak with sigmoidals many of them - rotated around the centre of the hill - are necessary. Consequently, in order to properly describe the process of bond-breaking within a very localised region of a detailed PES and at the same time modelling a smooth function outside that region, a large number of Fermi-like basis functions is required. Both sigmoidal and hyperbolic tangent functions can be used. However, convergence of the RMSE in online learning with functions which are symmetric about the origin, like the hyperbolic tangent, is often faster [29] and therefore will be preferred. Furthermore, we found that a detailed PES can be fitted with less complexity if a network with two hidden layers is chosen for the NN architecture. In N dimensions $2N$ nodes in the first hidden layer and one node in the second hidden layer can form one bump [27].

In summary, as an optimization algorithm for the network weights we will employ the adaptive global extended Kalman filter (AGEKF) with two forgetting schedule parameters $\lambda(0)$, λ_0 and an adaptive threshold of $a_{th} \times RMSE$. The activation functions of the hidden layers are hyperbolic tangents and linear functions in the output layer. The NN structure will mainly consist of one input layer, two hidden layers and one output layer with a high number of parameters. Furthermore, the input data are pre-conditioned, i.e. we subtract the means and normalize the variances in order to improve ill-conditioning. In order to ensure a most accurate representation of the potential we use individual weighting of each energy. For instance, dissociation dynamics depend crucially on the region in which the bond of the molecule breaks, whereas the part where the potential is already elevated is of less importance. We will associate the former region with weights, which are up to ten times higher than the rest of the geometries.

C. Explicit consideration of the symmetry

The computational effort in DFT calculations can be reduced by taking advantage of the symmetry of the underlying problem. Since we know the surface symmetry beforehand it is clearly advantageous to include this knowledge prior to the optimization of the NN parameters. In this way we let the network concentrate on the crucial process, the bond-breaking of the molecule. In order to do so we pre-process the coordinates of the problem.

The original set of coordinates $X_c, Y_c, Z_c, d, \theta, \phi$ describe the six degrees of freedom of the molecule. Due to the high costs of *ab initio* calculations information on the clean Pd(100) has been determined only on the edges of the irreducible part of the unit cell [55, 56]. In order to represent the whole surface area the analytical fit assumed a certain set of symmetry operations to be valid [10]. We point out that the applied symmetry introduces artificial features into the PES. For instance, the molecule in the analytical PES does not have any ϕ dependency on the diagonals of the unit cell. However, since this PES serves as a test problem for our NN approach, we employed the same symmetry and transformed the original coordinates into a set of eight inputs to the NN:

$$
\begin{aligned}
\mathrm{X}_1 &= d, \\
\mathrm{X}_2 &= d^2, \\
\mathrm{X}_3 &= Z_c, \\
\mathrm{X}_4 &= \sin^2(\theta)\cos(2\phi)\,[\cos(G\cdot X_c) - \cos(G\cdot Y_c)], \\
\mathrm{X}_5 &= \sin^2(\theta)\cos(2\phi)\,[\cos(2G\cdot X_c) - \cos(2G\cdot Y_c)], \\
\mathrm{X}_6 &= \cos^2(\theta), \\
\mathrm{X}_7 &= \cos(G\cdot X_c) + \cos(G\cdot Y_c), \\
\mathrm{X}_8 &= \cos(2G\cdot X_c) + \cos(2G\cdot Y_c).
\end{aligned}
$$

The transformations are based on Fourier terms in the lateral coordinates X_c and Y_c up to a reciprocal lattice vector of $2\,G$ representing the periodicity of the surface, with $G = 2\pi/a$ and the lattice constant a. The term $\cos(2\phi)\,[\cos(G\cdot X_c) - \cos(G\cdot Y_c)]$ in the fourth coordinate reflects the four-fold symmetry of the surface. The factor $\sin^2(\theta)$ weights this term, since the energy of an upright molecule should not have any azimuthal dependency. It also reflects the internal symmetry of the diatomic molecule. From the theoretical *ab initio* calculations it has been found that the energy increases like $cos^2(\theta)$ [55], which we included as one input. There is no symmetry within the coordinates d and Z_c. However, the vibration of the molecule in the gas phase can be described by a harmonic oscillator and therefore we incorporated an additional coordinate d^2.

Instead of presenting the original six degrees of freedom of the molecule to the NN we now apply this new set of eight inputs representing the symmetry of the surface. The NN performs a *non-linear fit* on these new inputs. The transformation needs to be done only once per surface symmetry.

D. Neural Network PES

We will now present six dimensional continuous NN representations of the mesh of points created from the *analytical* PES for the dissociation of hydrogen over Pd(100). Open questions are the necessary number of training points and their sampling for obtaining a good description of the PES as well as the number of parameters of the NN description needed.

1. NN-fit based on a dense grid of configurations

In order to test if NNs are able to fit PESs of surface reactions at all we first sampled the configurations from the analytical PES on a very dense mesh in all six dimensions. The corresponding training set consists of 80,685 energies evaluated above 55 adsorption sites. For the test set we collected 91,665 points. After 20 so-called epochs, i.e. 20 iterations through the whole set of training examples, the training root mean squared error measured 9 meV with a test error of 46 meV. Both errors lie well below the desired *ab initio* accuracy of 0.1 eV.

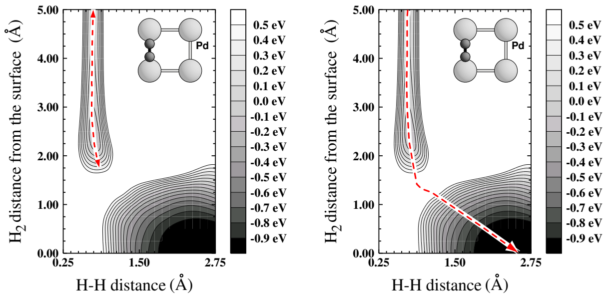

FIG. 3: Two classical molecular dynamics trajectories (dashed lines) on a NN-PES. The initial conditions of the molecules are the same except for their kinetic energy which are (a) $E_{\rm kin} = 0.5\,\rm eV$ and (b) $E_{\rm kin} = 0.9\,\rm eV$. The simulation time was (a) 52 fs and (b) 40 fs. Insets: configuration of the molecule.

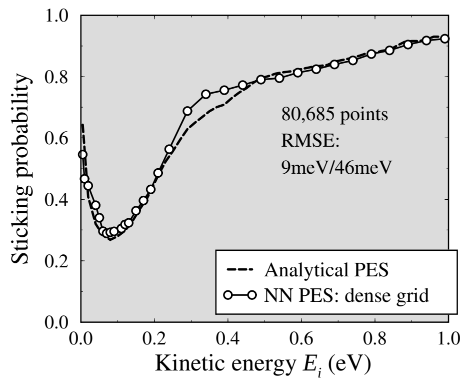

FIG. 4: Sticking probability versus kinetic energy for the dissociation of hydrogen on the Pd(100) surface. The training set consists of 80,685 training points sampled on a dense grid in all six dimensions, with e.g. 55 lateral configurations. The fit is based on a $\{8-50-50-1\ tl\}$ NN. Parameter set: $\lambda(0)=0.98$, $\lambda_0=0.99936$, $a\ _{rmth}=0.6$.

The obtained NN representation is then used as an input to extensive molecular dynamics calculations. Figure 3 illustrates the adsorption process over one particular site, the bridge site with the hydrogen atoms pointing towards the on-top sites. The molecule approaching the surface under normal incidence with its axis parallel to the surface at an energy of $0.5\,\rm eV$ is not able to overcome the barrier for dissociation. Due to the highly repulsive palladium top sites it is scattered back into the gas phase (Fig. 3(a)). With a kinetic energy of $0.9\,\rm eV$ the molecule has enough momentum to overcome the energy barrier and dissociates (Fig. 3(b)).

From the molecular dynamics simulations we evaluated the sticking probability of the impinging hydrogen molecule as a function of the initial kinetic energy. The dissociation process is highly site dependent which requires to consider a good statistical average over the initial configurations for the determination of the sticking coefficient. For each kinetic energy we need to calculate $500-1,000$ trajectories with randomly sampled initial configurations until convergence of the sticking coefficient is attained. The error of the sticking coefficient corresponds to $\sqrt{S(1-S)}/\sqrt{n}$, where S is the sticking probability and n is the number of trajectories. For each sticking curve the sticking probability has to be evaluated at a number of energies depending on the energy range of interest. For the presented adsorption coefficients we performed MD calculations with $10,000-30,000$ trajectories.

With the dense grid in all six degrees of freedom of the hydrogen molecule we were able to get excellent agreement between the analytical and neural sticking curve as displayed in Fig. 4. Both the initial high adsorption probability followed by a drop of sticking due to the steering effect and the increase with higher kinetic energies typical for dissociative adsorption are well reproduced. The differences between the analytical and neural sticking curve are smaller than 5% over the presented energy range.

However, in order to be applicable to the fitting of *ab initio* PESs, which are very time-consuming to evaluate and therefore only allow the calculation of a view hundred up to a view thousand configurations, we have to find a method to sample the configurations more efficiently.

2. *NN-fit based on high-symmetric configurations*

The usual approach in theoretical *ab initio* studies of dissociation processes is based on the calculation of 2D sections of the 6D energy surface, the elbow plots. For one such section, the orientation of the molecule (θ and ϕ) and the coordinates of the center of mass in the surface plane (X_c, Y_c) are kept fixed. Only the height Z_c and the bond length d vary. Commonly these elbow plots are evaluated with the molecule above high-symmetric sites. We will adopt this approach here as well and sample the points from the analytical PES in the same way. The NN is then used to interpolate between these sections.

We trained a $\{8–50–50–1\ tl\}$ NN with 1,560 examples calculated from the analytical PES. The elbow plots were evaluated above different high-symmetric sites, i.e. top, bridge, and hollow sites and one intermediate configuration at ($X_c = 0.25a, Y_c = 0.25a$, with a being the lattice constant of the (1x1) surface unit cell). At each site the energies were collected for five different angles ϕ with the molecule upright, 45° tilted, and parallel to the surface. For a single elbow plot we used 30 points along and perpendicular to the reaction path. The test set consists of $5,200$ energies sampled from the same elbow plots as the training set. The training error after 50 epochs and two hours runtime on an IBM-SP2 node measured $0.1\,\rm meV$ with a test error of $0.15\,\rm eV$. From the information of the root mean squared error alone we would judge this approximation as being satisfactory.

Also this NN representation was used as input to molecular dynamics simulations. Figure 5 compares the

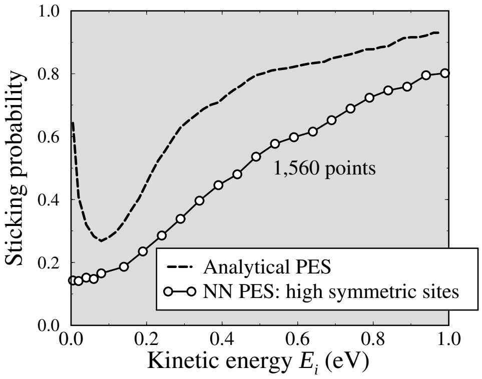

FIG. 5: Sticking probability versus kinetic energy for the system H_2/Pd(100). The sticking has been calculated by classical molecular dynamics on a six-dimensional analytical and neural PES, respectively. Training set: 1,560 examples.

sticking probability obtained from the NN-PES with the dynamical result from the underlying analytical PES. The NN-PES interpolating high-symmetric sites reproduces the increase of the sticking probability at energies larger than 0.2 eV qualitatively, but it fails to reproduce the high sticking probability at low kinetic energies.

It is now well understood that this behavior is a consequence of the corrugation and anisotropy of the multidimensional PES which lead to strong forces acting at the molecules upon adsorption. At low kinetic energies, these forces can either steer the molecule into a favorable configuration for direct dissociation [10, 41, 58, 59] or lead to the conversion of perpendicular kinetic energy into parallel kinetic energy and/or internal energy of the molecule so that they become temporarily dynamically trapped [18, 19, 60–63]. Both effects result in high adsorption probabilities at low kinetic energies but become suppressed at higher kinetic energies which causes the decrease in the adsorption probabilities. At even higher kinetic energies, molecules start to directly overcome the dissociation barriers.

We analyzed the data in more detail in order to determine the reason for the discrepancy of sticking probability in the NN fit. In particular, we compared the corrugation of the barrier heights calculated from the analytical and neural PES, respectively. We did this by fixing the hydrogen molecule at a height of $Z = 1.6$ Å above the surface with an intramolecular distance of $r = 1.0$ Å and angles $\phi = \pi/2$, $\theta = \pi/2$ while changing the lateral coordinates across the unit cell. The configuration of the molecule corresponds to the region where the bond already starts to break. In the underlying analytical corrugation a high barrier for dissociation is present if the molecule approaches the surface above the top site. Above the bridge and the hollow site the molecule is able to dissociate freely. Furthermore, the energy barrier decreases monotonically from the top site to the bridge site. A slow molecule is able to move from the top site where it experiences a high barrier to the favorable dissociation configuration above the bridge site. It is also able to reach the bridge site from the hollow site. However, we found that this is not true for the NN-PES interpolating the top, bridge, hollow and one intermediate site only. The PES exhibits additional barriers between bridge and top site and bridge and hollow site, respectively. These artificial barriers diminish the steering effect and thus cause a monotonically increasing sticking curve as shown in Fig. 5.

We conclude that for interpolations of PESs with NNs it is essential to include more than the usually calculated elbow plots above high-symmetric sites in the training and test sets. For instance, if we apply additional configurations in the *test set* of the above presented NN-approximation we get a test error of 0.32 eV, which is clearly above the desired accuracy. Hence, with the use of additional configurations also the RMSE reflects the unsatisfactory interpolation based on high-symmetric sites only. The results also demonstrate that the steering effect involves all six degrees of freedom and underlines the importance of high-dimensional studies in order to predict reaction probabilities.

3. NN-fit based on an enhanced lateral grid

In order to achieve a better representation of the steering effect with NNs we increased the number of training points in the lateral directions of the unit cell. Instead of applying only four lateral configurations we used ten different adsorption sites in the irreducible part of the unit cell.

We performed a number of interpolations of the analytical PES with different training sets using a $\{9-50-50-1\ tl\}$ NN. Figure 6 displays the dynamical results of two of them. For the interpolation with 3,270 training points with the above introduced enhanced lateral grid - while keeping the sampling of the other dimensions as described in the previous section - the sticking probability of the analytical PES is well reproduced. The training and test error after 20 epochs were 2 meV and 0.1 eV. In particular, the corrugation is now well represented, allowing the steering effect to become effective.

In Fig. 6 we also plotted the result obtained from a less good NN fit. The training and test errors with 6 meV and 0.15 eV based on a training set of 8,850 energies were slightly worse. The higher training and test errors lead to a larger deviation of the sticking probability from the analytical PES. We point out that it may always be possible that a better NN fit with a different set of Kalman filter parameters and a different number of weights exists. Yet, we like to emphasize that it is difficult to assess the quality of a PES without knowing the results of dynamical simulations.

We emphasize that the NN-PES with this finer lateral grid is not uniquely defined. An increase of the number of points in the other degrees of freedom as done for the training set with 8,850 energies, does not necessarily lead

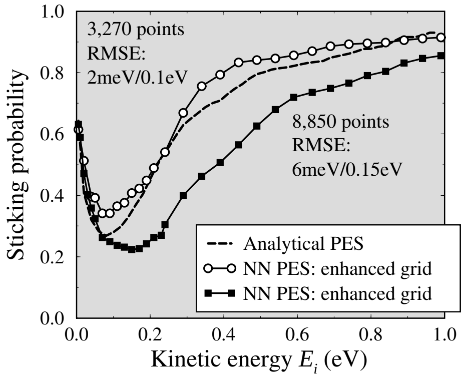

FIG. 6: Sticking probability versus kinetic energy for the system H_2/Pd(100) for two different training sets based on ten lateral configurations of the molecule. The training and test errors for each fit are indicated in the graph. Higher errors lead to larger discrepancies to the original sticking curve. NN: $\{8-50-50-1\ tl\}$. Extended Kalman Filter parameter: $\lambda(0)=0.98$, $\lambda_0=0.99936$, $a_{th}=0.6$.

to a better fit. All degrees of freedom play a role and the mesh is not sufficiently dense enough in order to give a quantitatively better dynamical result with a higher number of points. In any case, the NN-models based on an enhanced lateral mesh reproduced the steering effect in all cases qualitatively.

A detailed analysis of the accuracy of the NN-model based on the dense grid of points revealed that 94% of the test examples have an error smaller than 0.1 eV and already 99% do not exceed a threshold of 0.2 eV. Large errors occur only at values above 1 eV. This is the region far away from the valley of the elbow plots. The errors were influenced by the imposed higher weighting of the points close to the minimum dissociation pathway. However, the results support that indeed the regions of higher potential energies have almost no influence on the reaction probabilities as plotted in Fig. 4. This is an important issue for the fitting of PESs. Not all configurations are equally important for the determination of the sticking probability. For instance, at lower kinetic energies the adsorption dynamics depend crucially on whether there is a small energy barrier in the entrance channel, where the center of mass of the molecule is still far away from the surface, or not. Yet, the region where the potential is already elevated might have almost no influence on the dissociation probability. Consequently, the root mean squared error, which is usually used as a measurement of the accuracy of the fit, is of less significance. We will always weight the configurations close to the valley of each dissociation pathway up to 10 times higher than the other geometries.

IV. NN TEST: 6-D ANALYTICAL PES FOR $H_2/(2\times2)$S/PD(100)

In the previous section we discussed a system in which *activated and non-activated* pathways towards dissociation existed on the same surface, with the former ones being a minority but having important dynamic consequences. We showed that in order to obtain a very good agreement between the analytical and neural sticking probability a high number of training points and parameters were required. Still, in comparison to *direct ab initio molecular dynamics* where up to 10^7 energies need to be calculated, orders of magnitude fewer DFT calculations would be necessary to obtain a reliable description of the dynamical properties.

As a second test problem for the representation of the PES in dissociation reactions with NNs we will now investigate a system for which all reaction pathways are activated: The dissociation of H_2 over a sulfur covered Pd(100) surface. It is experimentally well known that sulfur adsorbates hinder the H_2 dissociation process on Pd(100) [64–66]. This observation was verified by DFT studies [42, 55, 67], and it was found that the poisoning effect of sulfur adatoms for H_2 dissociation at low sulfur coverages ($\Theta_S \leq 0.25$) is governed by the formation of energy barriers and not by blocking of adsorption sites.

A. *Ab initio* and analytical PES

DFT calculations of the system $H_2/(2\times2)$S/Pd(100) revealed that the PES is modified significantly compared to the dissociation on the clean Pd(100) surface [42, 55, 67]. While the process on the latter surface is non-activated, for a (2×2) sulfur adlayer corresponding to a coverage of $\Theta_S = 0.25$ it is inhibited by energy barriers. Their heights depend strongly on the distance between the hydrogen and the sulfur atoms leading to a highly corrugated PES. The minimum barrier towards dissociative adsorption has a height of 0.1 eV, while close to the adsorbate atoms the barriers become larger than 2.5 eV due to the strong repulsion between sulfur and hydrogen. The adsorption height of the sulfur atoms is 1.31 Å above the surface. The adsorption energy at all sites close to sulfur atoms is reduced in comparison to the clean surface. But still, H_2 adsorption into all hollow sites not occupied by sulfur remains an exothermic process.

For the theoretical investigation of the high-dimensional PES the common strategy of computing 2D cuts through the 6D configuration space has been followed, using an analytical representation that is similar to the form previously employed for the clean Pd(100) surface [42]. Due to the larger unit cell some higher Fourier coefficients have been included in the lateral directions, and in the azimuthal dependence a higher order term was introduced. Again, the coordinates in the (Zd) plane were transformed into reaction path coordinates. The parameters were determined such that the

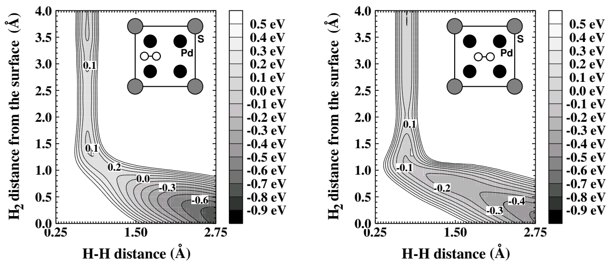

FIG. 7: Contour plots through the six-dimensional analytical PES of the dissociation of H_2 over (2×2)S/Pd(100) for (a) the Pd bridge site and (b) the fourfold hollow site. Insets: geometry of the dissociation pathways.

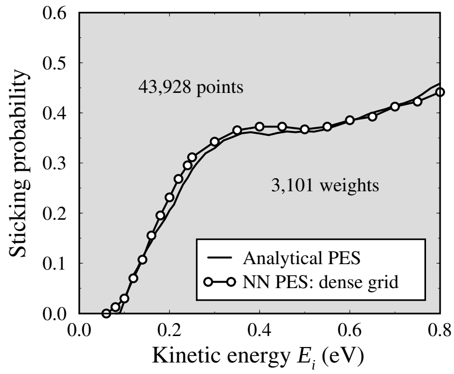

FIG. 8: Sticking probability versus kinetic energy for $H_2/(2\times2)$S/Pd(100) for a {9–50–50–1 tl} NN. The data sampling from the analytical PES is based on a dense mesh of configurations in all six degrees of freedom of the H_2 molecule. Parameter: $\lambda(0)=0.98$, $\lambda_0=0.99906$, $a_{th}=0.6$.

difference to the *ab initio* calculations on the average is smaller than 50 meV.

Figure 7 shows two 2D-cuts through the six-dimensional configuration space. Whereas on the clean surface the molecule over the palladium bridge site was able to dissociate freely, due to the presence of sulfur the molecule experiences a barrier of 0.16 eV. The minimum pathway is now over the fourfold hollow site with an energy barrier of 0.11 eV.

B. Incorporation of the symmetry

We incorporated the symmetry within the NN within the neural by using the same terms as on the clean Pd(100) surface but adding one higher order term for the azimuthal dependency. In analogy to the analytical PES we employed reaction path coordinates in the (Zd) plane. Furthermore, we did not employ the distance of the hydrogen molecule from the surface as an input to the NN, but rather an exponential decay of that coordinate. In reaction path coordinates this translated to the term $e^{(-s/2)}$, where s is the coordinate along the reaction path. The transformation reflects that far away from the surface the molecule is in the gas phase and any dependency on the distance from the substrate should vanish. Moreover, in the gas phase the potential-energy is isotropic. Only the bond length of the two hydrogen atoms should play a role, and therefore we weighted all other terms with the same factor $e^{(-s/2)}$.

The new set of nine coordinates, i.e. the inputs to the NN, are:

$X_1 = d$,
$X_2 = d^2$,
$X_3 = e^{(-s/2)}$,
$X_4 = \sin^2(\theta)\cos(2\phi)\,[\cos(GX_c) - \cos(GY_c)]\, e^{(-s/2)}$,
$X_5 = \sin^2(\theta)\cos(2\phi)\,[\cos(2GX_c) - \cos(2GY_c)]\, e^{(-s/2)}$,
$X_6 = \cos^2(\theta) e^{(-s/2)}$,
$X_7 = [\cos(GX_c) + \cos(GY_c)]\, e^{(-s/2)}$,
$X_8 = [\cos(2GX_c) + \cos(2GY_c)]\, e^{(-s/2)}$,
$X_9 = \sin^4(\theta)\cos(4\phi)\,[\cos(2GX_c) + \cos(2GY_c)]\, e^{(-s/2)}$.

C. NN-PES

On the clean Pd(100) surface it was necessary to use a high number of training points along with a high number of parameters to represent the detailed PES with activated and non-activated paths towards dissociation. Correspondingly, the first test of a NN approximation of the analytical PES for the sulfur covered Pd(100) will be based on a dense grid of points.

1. NN-fit based on a dense grid of configurations

We fitted 43, 928 data points from the analytical PES on a dense grid of configurations in all six degrees of freedom of the hydrogen molecule. The network consists of two hidden layers with fifty nodes in each of them $\{9-50-50-1\ tl\}$. For the test of the accuracy of the interpolation we used 5, 891 energies. After 40 epochs the training and test error were 0.033 eV and 0.043 eV. The NN-PES has subsequently been used in molecular dynamics calculations to determine the sticking probability. Figure 8 displays these results. The NN sticking curve agrees very well with the analytical sticking coefficient, their values differ by less than 3%. In comparison, for a good fit on the clean surface a number of examples twice as large was required.

Furthermore, for the sulfur poisoned surface the number of weights in the approximation can be greatly reduced without loosing much of the networks performance. The sticking probability for a {9–20–20–1 tl} network differs from the value based on the analytical PES by less than 5%. The training and test error (0.068 eV and 0.081 eV) were slightly higher than for the network with 3101 parameters, but still within the desired *ab ini-*

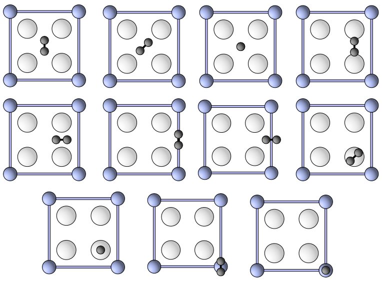

FIG. 9: Eleven adsorption configurations of the system $H_2/(2\times2)S/Pd(100)$. The corresponding elbow plots are used for the NN-fit.

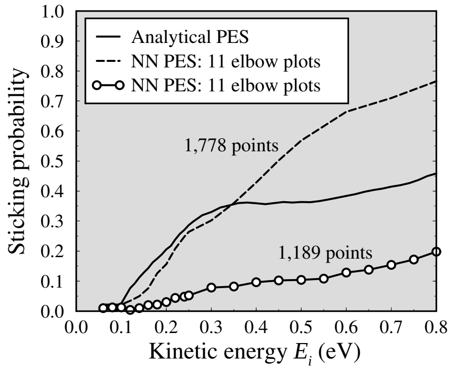

FIG. 10: Sticking probability versus kinetic energy for the system $H_2/(2\times2)S/Pd(100)$: Analytical PES and a NN-PES based on the eleven configurations in Fig. 9. NN: $\{9-20-20-1\ tl\}$. Parameter: $\lambda(0)=0.98$, $\lambda_0=0.99906$, $a_{th}=0.3$.

tio accuracy. The training time with such a high number of examples but only 641 weights reduces to seven hours on an IBM-SP2 node in comparison to several days for the 3101 parameter case.

On the clean surface a NN with such a small number of parameters was not able to describe the correct coexistence of activated and non-activated pathways. We conclude, with respect to the number of training points and the complexity of the appropriate NN, that fitting a strictly activated PES is a profoundly easier task.

2. *NN-fit based on eleven elbow plots*

Usually, a dense grid of energies as presented in the previous section will not be available due to the high numerical costs of *ab initio* calculations. Commonly, DFT studies of PESs concentrate on 2D cuts through the configuration space with the molecule above high-symmetric sites. In Fig. 9 we plotted eleven such configurations which have been used for the system $H_2/(2\times2)S/Pd(100)$. The molecule approaches the surface above the fourfold hollow site, the palladium bridge site, the sulfur bridge site, on-top of a palladium atom and on-top of a sulfur atom. The orientation of the molecule is either parallel or perpendicular to the surface.

We performed a $\{9-20-20-1\ tl\}$ NN interpolation based on $1,189$ training and 471 test energies obtained from the analytical PES in the configurations of Fig. 9. The test and training error after 100 epochs measured 0.078 eV and 0.096 eV, respectively. The resulting neural sticking coefficient in Fig. 10 exhibits the same increase in the sticking probability with kinetic energy as the corresponding analytical curve but its value is strongly reduced. A NN fit based on 1778 training examples from the same cuts resulted in a description of the PES which was too reactive at high kinetic energies (see Fig. 10).

3. *NN-fit based on eleven elbow plots and corrugation*

In order to get a reliable description of dynamical properties for dissociation processes with NNs it is not sufficient to follow the usual approach of restricting the calculations to 2D-cuts above high-symmetric sites. We need to add information about the PES, which is not present in the elbow plots. Numerical calculations based on the analytical PES revealed that the steering effect is not only present on the clean Pd(100) surface, but also on the sulfur covered sample [24]. Molecules approaching the surface above sites with a high barrier to dissociation can be reoriented by the forces to more favorable adsorption configurations. We showed in the discussion of the first test problem that the distribution of the barriers within the unit cell is important for the reproduction of the steering effect and we have therefore tested how the incorporation of the energetic corrugation improves the interpolations.

The H_2 molecule dissociates with its axis oriented parallel to the surface; the minimum path is located above the fourfold hollow site. Figure 11 (a) and (b) display the variation of the energy barriers the H_2 molecule experiences during adsorption for two angular orientations above different lateral positions. In Fig. 11(a) the molecule is oriented parallel to the surface above the sulfur bridge site with the H_2 atoms pointing towards the fourfold hollow site. In order to scan the barriers we fixed the (Z_c, d, θ, ϕ) configuration for two different bond lengths d and heights Z_c and moved the molecule from the sulfur bridge site to the fourfold hollow site. The same is done in Fig. 11(b) but now the H-atoms point initially in the direction of the sulfur atoms.

The configuration of the H_2 molecule for the solid lines in Fig. 11 correspond to the position of the maximum barrier in the entrance channel above the fourfold hollow site. In Fig. 11(a) the energy barrier decreases monotonically from a value of 0.3 eV above the sulfur bridge site at $(X_c, Y_c)=(0\,a, 0.5\,a)$ to 0.1 eV above the fourfold hollow site at $(0.5\,a, 0.5\,a)$, where a defines the length of the (2×2) unit cell. The monotonic decrease of the energy barriers enables the molecule to be redirected to the most favorable dissociation configuration above the hollow site

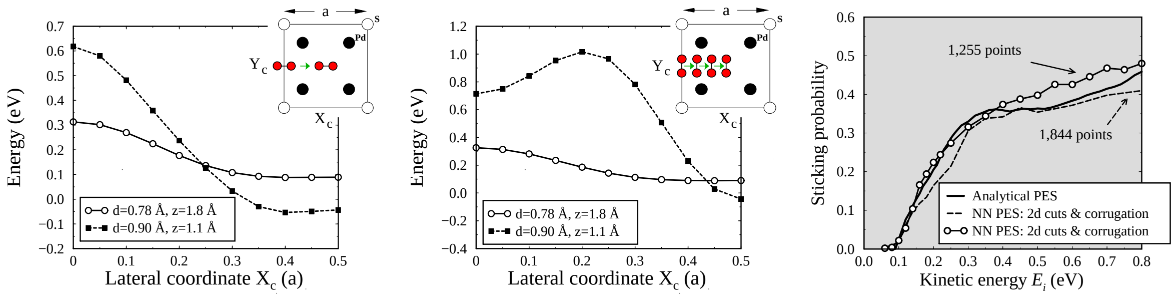

FIG. 11: (a) H_2 bond axis parallel to X_c, (b) H_2 bond axis perpendicular to X_c, and (c) Sticking probability. (a) and (b): Corrugation of the energy barriers for the system $H_2/(2\times2)S/Pd(100)$ with H_2 in two different orientations and its axis parallel to the surface. The energies are calculated for two different heights and bond lengths of the molecule. In both plots the molecule is moved from the sulfur bridge site to the fourfold hollow site. (c) Sticking probability versus kinetic energy for $H_2/(2\times2)S/Pd(100)$ calculated from the analytical PES and two NN-PESs. The neural PESs are based on the eleven configurations in Fig. 9 and the corrugation in (a) and (b).

even when it approaches the surface above, say, the palladium bridge site at $(X_c, Y_c) = (0.25\,a, 0.5\,a)$. If we further stretch the bond length of the hydrogen molecule and decrease the distance to the surface we obtain again a monotonic decrease of the energy (see the dashed line in Fig. 11(a)). However, the energy barrier at the sulfur bridge site has significantly increased due to the shorter distance to the repulsive sulfur atoms. Above the fourfold hollow site the energy is now negative, the molecule has started to dissociate. Note, that the energy zero relates to the situation where the molecule is located far away from the surface ($Z_c > 5$ Å) having its equilibrium bond length. If we let the bond length stretch further and allow the atoms to approach the surface the energy at the fourfold hollow site would further decrease reflecting that the dissociation process on the sulfur covered Pd(100) surface is exothermic.

In Fig. 11(b) the H_2 molecule has been rotated by 90° in the azimuthal direction. For the configuration corresponding to the solid line again the energy barriers decreases monotonically as a function of the distance from the sulfur atoms. With a stretched bond length of $d = 0.9$ Å and a distance from the surface of $z = 1.1$ Å the picture has changed. Now the barriers are at its highest value above the Pd bridge site. This is due to the repulsive character of the palladium atoms which the H-atoms point at in this configuration. At the fourfold hollow site the potential-energy is again negative.

To improve the NN description we included the information about the variation of the energy barriers within the unit cell from Fig. 11 in the training examples. We added 66 potential-energies related to the corrugation of the barriers to the information governed from the previously discussed eleven 2D cuts. Namely, instead of optimizing the NN with $1,189$ and $1,778$ training examples based on the elbow plots only as shown previously in Fig. 10, we now use $1,255$ and $1,844$ points, respectively. The sticking probability for both training sets in Fig. 11(c) agrees now semi-quantitatively with the underlying analytical PES. Thus, incorporating only a small number of additional information to the calculated elbow plots can lead to significant improvement of the dynamical result. Figure 11(c) demonstrates that the incorporation of available physical knowledge about the system of investigation improves the interpolation considerably. It is well known that steering in dissociation dynamics is present and can be essential for the calculation of a dynamical property like the sticking probability. It is clear that the reorientation of the molecule is affected by the distribution of the energy barriers on the surface. Together with the knowledge about the favorable dissociation configuration of the studied molecule which can be gained from DFT-calculations we were able to calculate a small number of additional energies. With this new information the NN was able to reproduce the adsorption coefficient with an error of less than 6%.

V. CONCLUSION

We have shown that NNs can represent *ab initio* PESs of several degrees of freedom accurately. The computational cost of training a NN is small and just a fraction of the costs of the DFT calculations. The resulting NN output function, the potential-energy, and its derivatives, the forces, are very efficient to evaluate and allow molecular dynamics calculations with extensive statistics.

Concerning the amount of training data required to obtain a reliable representation it is not sufficient to perform a NN fit based on the usually calculated top, bridge and hollow sites only. Intermediated configurations need to be considered. An *equidistant* sampling results in a number of $10^4 - 10^5$ total energies for an accurate interpolation. The required number of training energies for dissociation processes can be further reduced by an *efficient* sampling of the configurations.

Model calculations on the systems H_2/Pd(100) and H_2/S(2×2)/Pd(100) revealed, that the form of the energetic corrugation can significantly influence the dynamical result. We therefore proposed a modification of the usually applied sampling of total energies in DFT calculations of dissociation processes. In addition to elbow plots above high-symmetric sites we recommend to calculate the corrugation of the barrier heights in more detail by collecting information of the potential-energy as a function of the lateral coordinates within the surface unit cell. The modified sampling scheme allows to calculate dynamical results with NNs based on 10^3-10^4 *ab initio* energies. The costs for a description of dissociation reactions with NNs are orders of magnitude smaller than those of "on the fly" *ab initio* dynamics were up to 10^7 energies might be necessary.

[1] A. Groß, Surf. Sci. **500**, 347 (2002).
[2] P. Blaha, K. Schwarz, P. Sorantin, and S. B. Trickey, Comp. Phys. Comm. **59**, 399 (1990).
[3] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[4] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[5] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comp. Phys. Comm. **107**, 187 (1997).
[6] A. De Vita, I. Štich, M. J. Gillan, M. C. Payne, and L. J. Clarke, Phys. Rev. Lett. **71**, 1276 (1993).
[7] A. Groß, M. Bockstedte, and M. Scheffler, Phys. Rev. Lett. **79**, 701 (1997).
[8] V. Musolino, A. Selloni, and R. Car, Phys. Rev. Lett. **83**, 3242 (1999).
[9] M. Moseler, H. Häkkinen, and U. Landman, Phys. Rev. Lett. **89**, 176103 (2002).
[10] A. Groß, S. Wilke, and M. Scheffler, Phys. Rev. Lett. **75**, 2718 (1995).
[11] A. Groß and M. Scheffler, Phys. Rev. B **57**, 2493 (1998).
[12] J. H. McCreery and G. W. Jr., J. Chem. Phys. **63**, 2340 (1975).
[13] A. Forni, G. Wiesenekker, E. J. Baerends, and G. F. Tantardini, J. Phys.: Condens. Matter **7**, 7195 (1995).
[14] A. Forni and G. F. Tantardini, Surf. Sci. **352-354**, 142 (1996).
[15] U. Kleinekathöfer, K. T. Tang, J. P. Toennies, and C. L. Yiu, J. Chem. Phys. **111**, 3377 (1999).
[16] D. E. Makarov and H. Metiu, J. Chem. Phys. **108**, 590 (1998).
[17] A. Groß, M. Scheffler, M. J. Mehl, and D. A. Papaconstantopoulos, Phys. Rev. Lett. **82**, 1209 (1999).
[18] H. F. Busnengo, A. Salin, and W. Dong, J. Chem. Phys. **112**, 7641 (2000).
[19] H. F. Busnengo, W. Dong, and A. Salin, Chem. Phys. Lett. **320**, 328 (2000).
[20] G. Kresse, Phys. Rev. B **62**, 8295 (2000).
[21] C. Crespos, M. A. Collins, E. Pijper, and G.-J. Kroes, Chem. Phys. Lett. **376**, 566 (2003).
[22] G. Wiesenekker, G. J. Kroes, and E. J. Baerends, J. Chem. Phys. **104**, 7344 (1996).
[23] J. Dai and J. Light, J. Chem. Phys. **107**, 1676 (1997).
[24] A. Groß, C. M. Wei, and M. Scheffler, Surf. Sci. Lett. **416**, 1095 (1998).
[25] A. Groß, Surf. Sci. Rep. **32**, 291 (1998).
[26] G. J. Kroes, Prog. Surf. Sci. **60**, 1 (1999).
[27] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, 1996).
[28] R. Rojas, *Theorie der neuronalen Netze* (Springer Verlag, Heidelberg, 1996).
[29] Y. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, in *Neural Networks: Tricks of the Trade, G. B. Orr and K. R. Müller (eds.)* (Springer Verlag, Heidelberg, 1998).
[30] S. Lorenz, A. Groß, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).
[31] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129 (1995).
[32] G. Cybenko, Mathematics of Control, Signals, and Systems **2**, 303 (1989).
[33] K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
[34] A. Groß, B. Hammer, M. Scheffler, and W. Brenig, Phys. Rev. Lett. **73**, 3121 (1994).
[35] G. J. Kroes, E. J. Baerends, and R. C. Mowrey, Phys. Rev. Lett. **78**, 3583 (1997).
[36] G. J. Kroes, E. J. Baerends, and R. C. Mowrey, J. Chem. Phys. **107**, 3309 (1997).
[37] G.-J. Kroes, Prog. Surf. Sci. **60**, 1 (1999).
[38] G.-J. Kroes, A. Groß, E. J. Baerends, M. Scheffler, and D. A. McCormack, Acc. Chem. Res. **35**, 193 (2002).
[39] H. F. Busnengo, E. Pijper, M. F. Somers, G. J. Kroes, A. Salin, R. A. Olsen, D. Lemoine, and W. Dong, Chem. Phys. Lett. **356**, 515 (2002).
[40] R. T. van Willigen, M. F. Somers, H. F. Busnengo, and G. J. Kroes, Chem. Phys. Lett. **393**, 166 (2004).
[41] A. Dianat and A. Groß, J. Chem. Phys. **120**, 5339 (2004).
[42] C. M. Wei, A. Groß, and M. Scheffler, Phys. Rev. B **57**, 15572 (1998).
[43] F. Scarselli and A. C. Tsoi, Neural Networks **12**, 15 (1998).
[44] J. G. Attali and G. Pages, Neural Networks **10**, 1069 (1997).
[45] D. F. R. Brown, M. N. Gibbs, and D. C. Clary, J. Chem. Phys. **105**, 7597 (1996).
[46] K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, and H. A. Scheraga, Chem. Phys. Lett. **271**, 152 (1997).
[47] H. Gassner, M. Probst, A. Lauenstein, and K. Hermansson, J. Phys. Chem. A **102**, 4596 (1998).
[48] F. V. Prudente and J. J. S. Neto, Chem. Phys. Lett. **287**, 585 (1998).
[49] J. B. Witkoskie and D. J. Doren, J. Chem. Theory Comput. **1**, 14 (2005).
[50] J. Behler, B. Delley, S. Lorenz, K. Reuter, and M. Scheffler, Phys. Rev. Lett. **94**, 036104 (2005).
[51] A. Gelb, *Applied Optimal Estimisation* (MIT Press, Cambridge, MA, 1974).
[52] S. Shah, F. Palmieri, and M. Datum, Neural Networks **5**, 779 (1992).
[53] T. B. Blank and S. D. Brown, J. Chemometrics **8**, 391 (1994).
[54] D. P. Bertsekas, SIAM J. Optimization **6**, 807 (1996).

[55] S. Wilke and M. Scheffler, Surf. Sci. **329**, L605 (1995).
[56] S. Wilke and M. Scheffler, Phys. Rev. B **53**, 4926 (1996).
[57] W. Brening and H. Kasai, Surf. Sci. **213**, 170 (1989).
[58] M. Kay, G. Darling, S. Holloway, J. White, and D. Bird, Chem. Phys. Lett. **245**, 311 (1995).
[59] A. Dianat and A. Groß, Phys. Chem. Chem. Phys. **4**, 4126 (2002).
[60] A. Groß and M. Scheffler, J. Vac. Sci. Technol. A **15**, 1624 (1997).
[61] C. Crespos, H. F. Busnengo, W. Dong, and A. Salin, J. Chem. Phys. **114**, 10954 (2001).
[62] M. A. Di Cesare, H. F. Busnengo, W. Dong, and A. Salin, J. Chem. Phys. **118**, 11226 (2003).
[63] A. Groß, A. Eichler, J. Hafner, M. J. Mehl, and D. A. Papaconstantopoulos, Surf. Sci. **539**, L542 (2003).
[64] K. Rendulic, G. Anger, and A. Winkler, Surf. Sci. **208**, 404 (1989).
[65] M. Rutkowski, D. Wetzig, and H. Zacharias, Phys. Rev. Lett. **87**, 246101 (2001).
[66] M. Rutkowski, D. Wetzig, H. Zacharias, and A. Groß, Phys. Rev. B **66**, 115405 (2002).
[67] S. Wilke and M. Scheffler, Phys. Rev. Lett. **76**, 3380 (1996).