universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

"Artificial Intelligence in Theoretical Chemistry"

verfasst von / submitted by

Michael Gastegger, MSc

angestrebter akademischer Grad /
in partial fulfillment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 419
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Chemie
Betreut von / Supervisor:	Priv.-Doz. Dr. Philipp Marquetand

ABSTRACT

The application of methods originating in artificial intelligence research – and machine learning in particular – to problems in theoretical chemistry offers exciting new possibilities. However, due to the relatively young age of this field and the complexity of the employed models, various difficulties complicate a routine use of these methods. The main goal of this thesis is to develop improved techniques and establish new protocols in order to overcome these limitations. These developments consist of: 1) A new algorithm for training high-dimensional neural network potentials (HDNNPs) – a machine learning technique especially well suited for modeling chemical systems. This training algorithm, termed the element-decoupled Kalman filter, yields HDNNPs of improved accuracy and is also able to incorporate molecular forces into the training process. 2) A HDNNP based fragmentation approach is introduced in order to model extended molecular systems. With this approach, macromolecular properties can be recovered using only the information contained in small fragments of a molecule. 3) These developments are supplemented by an improved adaptive sampling scheme, which identifies molecular configurations required to construct an accurate HDNNP model for a system in a highly automated fashion. 4) A machine learning approach capable of modeling molecular dipole moments is obtained by extending the structure of HDNNPs. The proficiency of all these novel strategies is investigated based on a variety of chemical systems. The performance of the element-decoupled training algorithm is analysed by modeling the Claisen rearrangement reaction of allyl vinyl ether to 4-pentenal. Linear all-trans alkanes serve as a simple test system to assess the HDNNP-based fragmentation approach. A final study is dedicated to the adaptive sampling scheme and the dipole moment model, where these techniques are applied to the molecular dynamics simulation of infrared spectra of methanol, n-alkanes of different lengths and the protonated alanine tripeptide.

ZUSAMMENFASSUNG

Strategien aus dem Forschungsbereich der künstlichen Intelligenz und insbesondere dem Gebiet des maschinellen Lernens eröffnen vielversprechende alternative Lösungswege für zahlreiche Probleme aus der theoretischen Chemie. Unglücklicherweise wird eine routinemäßige Anwendung dieser Methoden zur Zeit noch durch zahlreiche Komplikationen erschwert, welche Aufgrund des relativ jungen Alters dieses Forschungsgebietes sowie der hohen Komplexität der verwendeten Modelle auftreten. Das Ziel dieser Arbeit ist es daher, diese Hürden sowohl durch die Verbesserung von vorhandenen als auch die Entwicklung von neuartigen Herangehensweisen zu überwinden. Zu den zu diesem Zwecke eingeführten Neuerungen zählen: 1) Ein verbesserter Algorithmus zum Trainieren von hochdimensionalen neuronalen Netzwerkpotentialen (HDNNPs), ein Ansatz aus dem Bereich des maschinellen Lernens, welcher besonders für die Beschreibung von Molekülen geeignet ist. Die Verwendung dieses Trainingsalgorithmus – des sogenannten „element-decoupled“ Kalman Filters – macht es möglich, hochqualitative HDNNPs zu konstruieren und darüber hinaus auch molekulare Kräfte im Konstruktionsvorgang zu berücksichtigen. 2) Ein Fragmentationsansatz basierend auf HDNNPs wurde einführt, um große Moleküle beschreiben zu können. Mittels dieses Ansatzes können die Eigenschaften großer Moleküle allein anhand kleiner molekularer Fragmente vorhergesagt und so enorme Einsparungen an Rechenzeit erzielt werden. 3) Eine adaptive Selektionsstrategie stellt eine Ergänzung zu den beiden vorherigen Entwicklungen dar und ist in der Lage vollautomatisch jene Molekülgeometrien zu identifizieren, welche notwendig sind um ein chemisch aussagekräftiges HDNNP zu erhalten. 4) Eine Modifizierung der HDNNP Struktur ermöglicht es, ein generelles Modell für die Vorhersage von molekularen Dipolmomenten zu erhalten. Die Brauchbarkeit dieser neu eingeführten Methoden wird anhand einer Reihe von unterschiedlichen chemischen Testsystemen untersucht. Eine Analyse der Leistungsfähigkeit des „element-decoupled“ Kalman Filters erfolgt anhand der Claisen-Umlagerung von Allyl-vinylether zu 4-Pentenal. Lineare Alkanketten dienen der Untersuchung des HDNNP-basierten Fragmentationsansatzes. Eine weitere Studie beschäftigt sich mit der adaptiven Selektionsstrategie und dem Modell zur Vorhersage von Dipolmomenten. Im Rahmen dieser Studie werden die Infrarotspektren von Methanol, n-Alkanen unterschiedlicher Länge sowie des Alanintripeptidkations mittels Moleküldynamik simuliert.

CONTENTS

ACRONYMS

ACSF Atom-centered Symmetry Function

A-GEKF Atomic Global Extended Kalman Filter

AIMD *Ab initio* Molecular Dynamics

DFT Density Functional Theory

ED-GEKF Element-decoupled Global Extended Kalman Filter

E-GEKF Elemental Global Extended Kalman Filter

FF Force Field

GEKF Global Extended Kalman Filter

HDNNP High-dimensional Neural Network Potential

IR Infrared

MAE Mean Absolute Error

MD Molecular Dynamics

ML Machine Learning

NHC Nóse-Hoover Chain

NH Nóse-Hoover

NN Neural Network

PES Potential Energy Surface

QSAR Quantitative Activity Structure Relationship

RMSE Root Mean Squared Error

SGD Stochastic Gradient Descent

SMF Systematic Molecular Fragmentation

1 INTRODUCTION

The prospect of creating intelligent automata has fascinated humankind since ancient times. Fueled by the advent of modern computer architectures and sophisticated algorithms, the realization of this dream has nowadays left the realms of pure fiction and become a realistic possibility. The last decade in particular has seen several astounding developments in the field of artificial intelligence research. Especially the sub-field of machine learning (ML) – the science of autonomously learning complex relationships from past experience and accumulated data – has undergone dramatic progress, as is attested by the plethora of ML based applications which now permeate our everyday life:[1,2] Speech and handwriting recognition, spam filters, reverse image search, recommendation engines and self-driving cars are only a few examples which make heavy use of ML techniques.

It is therefore hardly surprising, that modern ML algorithms have also proven to be an invaluable addition to the fields of Theoretical Chemistry and Chemoinformatics.[3–5] Here, their ability to model highly complicated relationships is utilized in a variety of applications. In computer aided drug design and materials design for example, ML methods can be used to relate properties of compounds (e.g. solubility, biological activity, toxicity, melting points) directly to their structure (e.g. chemical graphs) in the form of quantitative activity structure relationship (QSAR) and similar models.[3,4,6] Due to the predictive power and computational efficiency of the underlying ML algorithms, the resulting models can then be used to screen vast databases for promising new compounds without the explicit need for experiments. Hence, ML based strategies can provide important guidance in the development of new drugs and materials, as is attested by their use to e.g. design special metal organic frameworks[7] and discover efficient organic photovoltaic materials[8]. Another application where ML shows promise is the prediction of organic reaction outcomes[9], a task which is highly relevant for efficient synthetic planning and design. Similar to an organic chemist, ML algorithms can learn to predict the products of reactions based on a set of reagents and reactants.

However, perhaps one of the most fascinating contributions of ML to Theoretical Chemistry is with respect to computational chemistry methods.[3,5] While these methods constitute a central tool of Theoretical Chemistry, they are hampered by intrinsic limitations. The most accurate of them aim to find a numerically exact solution to the electronic time-independent Schrödinger equation (full configuration interaction[10], quantum Monte Carlo[11]), thus providing an exact description of a molecular system within the non-relativistic Born–Oppenheimer picture. Unfortunately, the computations necessary to solve the resulting eigenvalue problem and integrals scale very unfavorably with the size of the system and quickly become prohibitively expensive. As a consequence, only extremely small molecules (up to a few atoms) can be treated with these methods on a routine basis. In order to model larger systems, physical rigor and in turn predictive accuracy need to be sacrificed in favor of computational efficiency. This necessity has lead to a variegated spectrum of computational chemistry methods, introducing varying degrees of physical and empirical approximations. High-level electronic structure methods, such as coupled cluster based approaches[12] (e.g. CCSD(T)), retain a relatively high accuracy, but are still restricted in the size of systems they can treat (up to a hundred atoms). More approximate methods, such as

density functional theory[13] (DFT), are able to describe much larger systems, albeit at a reduction in the reliability of their predictions. Situated at the far end of this spectrum are empirical force fields[14,15] (FF), which substitute the laws of quantum mechanics by simple, physically motivated functions fit to experimental and/or theoretical data. Consequently, FF can be used to simulate systems containing hundreds of thousands of atoms. However, due to the simplicity of the functions employed, the predictive accuracy of FF is limited, making most of them e.g. unable to account for changing bonding patterns. In general, this unfortunate relation between accuracy and computational efficiency inherent to all computational chemistry methods leads to something akin to a tightrope walk. One has to carefully choose a method which is able to describe the system to be investigated reasonably well, while at the same time being affordable from a computational perspective. Yet, based on the chemical complexity and size of the system at hand, a good choice of method might not be readily apparent if it is available at all.

ML offers the tantalizing possibility to overcome this precarious balancing act: By exploiting the ability of ML techniques to model highly complex relationships, approximate models of high-level electronic structure methods can be created based on only a handful of reference data points. Due to the powerful statistical machinery at the core of modern ML algorithms, these approximate ML models exhibit the same accuracy as the underlying electronic structure method, but can be evaluated at only a fraction of the original computational cost. This combination of accuracy and computational efficiency makes it possible to study problems typically beyond the reach of conventional theoretical chemistry methods.

ML models of electronic structure methods have undergone rapid development in the last decade, leading to the emergence of a variety of different strategies. Depending on their mode of application and the ML algorithms they are based on, it is possible to group these strategies into different classes. From an application based point of view, three main types can be differentiated: First, models which retain the basic formalism of electronic structure methods and only use ML techniques to approximate partial aspects. Examples include the use of ML techniques to substitute Gaussian basis sets[16] and attempts to approximate the exact exchange-correlation functional in DFT[17]. The second type of models employs ML in order to augment basic electronic structure calculations. Here, ML algorithms are used to model the difference between high-level electronic structure methods and a cheap approximate baseline method.[18] For subsequent predictions, only low level computations have to be performed and the ML model is used as a correction to recover the high-level results at virtually no extra cost. This approach has e.g. been used to accurately predict various physicochemical properties of a database containing 134 000 small organic compounds based on only 10 000 reference samples[18]. The final type of ML models completely forgoes any kind of electronic structure formalism or baseline and instead relies purely on the power of ML algorithms to model the target properties. In this case, electronic structure calculations are only used to generate suitable reference data points. As a consequence, this class of ML models is extremely efficient from a computational point of view but the high flexibility can complicate the creation process. Examples for this class can e.g. be found in References 5,19–22. A special case of the latter models are so-called ML potentials, which aim to accurately predict molecular potential energy surfaces (PESs) and the associated forces (first derivatives).[23–27] ML potentials constitute the ML analogue to empirical FFs. However, unlike FFs, these potentials are based on the flexible functional forms provided by ML algorithms and can therefore account for situations where classical

FFs fail (e.g bond breaking and formation events). Moreover, ML potentials exhibit the same chemical accuracy as the electronic structure method they are based on, while at the same time retaining the excellent computational speeds of conventional FFs. Due to this favorable combination of high accuracy and computational efficiency, ML potentials can be used to study chemical systems and problems inaccessible with conventional methods. A wide range of applications have been reported for ML potentials, with a particular focus on solid state systems[28–33], reactions on surfaces[34–42] and molecules[21,43–63].

From a method-centered perspective, the ML techniques employed in the above applications can be grouped into two main classes – Kernel methods[64] and artificial neural networks (NNs)[1]. In Kernel methods, predictions of an unknown sample are performed by measuring its similarity to known examples. As a similarity measure, Kernel methods use so-called Kernel functions, which represent an implicit mapping to high dimensional feature spaces. Due to their underlying structure, Kernel algorithms are well described by statistical learning theory and can be trained relatively fast. However, as a new sample needs to be compared against all reference points during prediction, the prediction speed is slower than other methods and grows linearly with the size of the reference data set. Moreover, appropriate Kernel functions – which can be thought of as basis functions used during fitting – have to be specified in advance and are not adapted during training, which can potentially limit the expressive power of Kernel methods. Notable examples of Kernel algorithms include Kernel ridge-regression, Gaussian processes (Kriging) and support vector machines (for a detailed discussion of these different methods, see References 64 and 2). The second class of algorithms – NNs – is inspired by the central nervous system. Similar to their biological counterpart, NNs are built from small subunits, which are connected in elaborate patterns. This arbitrarily complex structure can make NNs difficult to train and methods to successfully construct large NN models rank amongst the most important discoveries in ML in the last decade.[65] NNs also lack a sound theoretical foundation aiding in the interpretation of their predictions. However, the immense flexibility arising due to this complex structure is at the same time also the greatest strength of NN methods: Instead of being dependent on predefined basis functions, NNs are able to directly learn a set of appropriate basis functions during training. This property makes them a powerful tool for modeling abstract tasks, such as e.g. pattern recognition.[1] In addition, unlike Kernel methods, NNs do not use observed reference data during prediction, but instead store the relevant information directly in their connections. As a consequence, NNs are generally much faster to evaluate during prediction than Kernel methods. A wide variety of NN models are employed nowadays, with feed-forward NNs, convolutional NNs and recurrent NNs only being a few examples (an in depth description of all models can be found in Reference 1). In general, the choice between Kernel methods and NNs strongly depends on the task to be modeled and whether interpretability and fast training times or high flexibility and efficient prediction times of the final model are desired.

As is attested by the various examples reported above, ML models of electronic structure methods are a rapidly developing and diverse field of research. Yet, despite their apparent potential, approximate ML models are still far removed from being used as a routine tool in Theoretical Chemistry. While this situation has improved dramatically during the last few years due to the ongoing research efforts by different groups, several issues remain which limit the general applicability of ML models significantly: First, special adaptations to standard ML algorithms are necessary

in order to deal with the three dimensional structure of molecular systems and various invariances arising from the laws of quantum mechanics. Consequently, conventional training methods can no longer be used out of the box and need to be modified in appropriate ways. These circumstances can complicate the process of generating ML approximations of electronic structure methods substantially. A second issue concerns the economic use of the available reference data. Due to the computational cost of high-level electronic structure calculations, the construction of extensive reference data sets is not feasible in most cases. Hence, it is highly desirable to limit the number of reference calculations and instead effectively utilize additional information during the construction of the ML models (e.g. molecular forces). However, doing so once again requires elaborate modifications of existing training methods. Third, when constructing reference data sets, no straightforward strategies exist to determine which reference points need to be selected in order to obtain accurate and reliable ML models. Optimal selection criteria depend strongly on the chemical problem to be investigated. In addition, unproductive reference calculations should be kept to a minimum in accordance to the previous issue. These requirements render the generation of reference data tedious and difficult to automatize. Finally, the computational cost of high-level electronic structure methods is not only problematic with regards to the number of reference calculations but also with respect to the maximum size of the studied molecules. Approximate ML models are in principle able to describe molecular systems far larger than what is possible with conventional electronic structure methods. However, in order to train these models, expensive reference computations still need to be carried out for the whole system. This requirement imposes a severe limit on the maximum system size which can be treated with ML models.

The central goal of this thesis is to address the above issues and improve the general applicability and accessibility of ML methods for problems in Theoretical Chemistry. Due to the wide spectrum of possible applications and ML techniques, the current study will focus on a particular type of NN based ML potentials, so-called high-dimensional NN potentials (HDNNPs)[66]. HDNNPs are an adaptation of standard NNs and possess several properties which make them especially well suited for modeling molecular systems.

In the context of these HDNNPs, a new training algorithm was developed to deal with the first two issues described previously. This algorithm produces ML models of improved accuracy compared to conventional training methods. Moreover, it is also able to utilize additional information in the form of molecular forces during the training process in a consistent manner. Concerning issue number three, the possibility to simplify and automatize the construction of reference data sets was investigated by introducing a special sampling scheme. This scheme selects relevant data points based directly on the predictive uncertainty of the respective ML models and greatly reduces the number of reference data points required to obtain accurate HDNNPs. Finally, in order to overcome the limitations with respect to molecular size, the ability of HDNNPs to operate in a manner akin to fragmentation methods[67] was studied. By pursuing this strategy, HDNNPs are capable to reconstruct macromolecular properties based only on the information contained in small fragments of the original molecular system. In addition, it was found, that the special structure of HDNNPs is not only well-suited for modeling PESs, but can also be extended to describe molecular dipole moments.

A general overview of the theory underlying HDNNPs and the associated ML techniques, as well as of the simulation methods employed in this thesis is given

in Section 2. The strategies developed to address the issues highlighted above are described in detail in Section 3. Section 4 deals with the application of these strategies to concrete chemical systems and problems. The effectiveness of the employed methods is studied based on a Claisen rearrangement reaction, n-alkanes of varying lengths, the methanol molecule and a small tripeptide. To investigate the effectiveness of the HDNNP-based dipole moment model, it is used to simulate infrared (IR) spectra of methanol, the n-alkanes and the tripeptide. Finally, a conclusion and summary of the topics covered in this thesis is provided in Section 5.

2 THEORY

The purpose of the following chapter is to provide a short overview of the different methods and techniques forming the basis for the research conducted in this thesis. After giving a summary of artificial neural networks (NNs), high dimensional NN potentials (HDNNPs) are introduced along with atom-centered symmetry functions (ACSFs). This is followed by a discussion of different methods to train these potentials and NNs in general, as well as their respective advantages and drawbacks. The chapter is concluded with a brief description of the molecular dynamics (MD) simulation technique.

2.1 ARTIFICIAL NEURAL NETWORKS (NNS)

Like the central nervous system, NNs are an arrangement of interconnected subunits – so-called artificial neurons.[1,68–70] These neurons collect, process and transmit incoming signals based on the strength and pattern of the network connections. Typically, NNs are structured into layers – groups of neurons performing similar functions: The input layer of the NN collects signals from the environment. These signals are then transformed by one or more hidden layers. Finally, the signal processed in this manner is returned by the output layer.

Due to their modular structure, NNs are extremely flexible ML models. As a consequence, a wide variety of NN architectures suitable for different tasks exists nowadays, differing greatly e.g. in their connectivity patterns or arrangement of layers. One of the first and most frequently used architectures are feed-forward NNs. An example of such a NN with one hidden layer is shown in Figure 1. In a

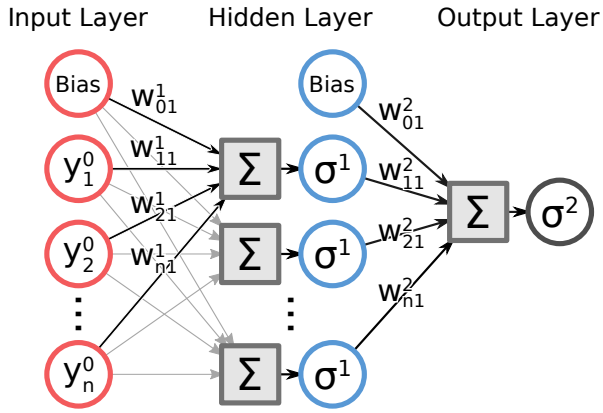


Figure 1: An example for a feed-forward NN with a single hidden layer and one output node. The nodes of two neighboring layers are connected via the weight parameters $\{w^l_{\alpha\beta}\}$, which have to be determined during training. The bias nodes provide an adjustable offset to the transfer functions σ^l.

feed-forward NN, just the neurons between adjacent layers are connected and signals are only transmitted into one direction. Beginning from the input layer, a vector of

inputs $\mathbf{y^0}$ is propagated through the layers l of the network by recursively applying the relation

$$y_\beta^l = \sigma^l\left(w_{0\beta}^l + \sum_\alpha^{n^{l-1}} w_{\alpha\beta}^l y_\alpha^{l-1}\right) \tag{2.1}$$

until the output layer is reached. The signals $\{y_\alpha^{l-1}\}$ obtained in the previous layer are scaled by the weight parameters $\{w_{\alpha\beta}^l\}$, which encode the strength and patterns of the connections within the NN. The weight $w_{\alpha\beta}^l$ connects the neuron α of the previous layer to the neuron β in the current layer, while the so-called bias weights $\{w_{0\beta}^l\}$ provide an adjustable offset to the transfer function σ. These weights are collected into the vector of weights $\mathbf{w}$ for the whole NN and have to be determined during the training process. Following this weighting step, the sum over the signals is formed and a transfer function σ^l is applied. This transfer function computes a nonlinear transformation (e.g. hyperbolic tangens) of the incoming signals and is the reason NNs can in principle fit any continuous function to arbitrary accuracy.[71–73]

Because of this property, as well as their computational efficiency and the availability of analytic derivatives, feed-forward NNs are used as the basis of many ML potentials. Unfortunately, standard feed-forward NNs are plagued by several issues when they are applied to model molecular PES directly. First, the architecture and hence the size of the input layer of a NN is fixed. As a consequence, only molecules containing a specific number of atoms predefined by the size of the input layer can be described. Second, the values of the weights remain constant after training. This means, that the changing the order the inputs (e.g. atomic coordinates) are presented to the NN, also changes its predictions, a behavior which is unphysical in the context of PES. Moreover, if Cartesian or related coordinates are used to describe the molecular geometries, the resulting ML potential is no longer invariant with respect to translations and rotations of the molecule.

2.2 HIGH DIMENSIONAL NEURAL NETWORK POTENTIALS (HDNNPS)

HDNNPs[66,74] are able to overcome the above problems by pursuing a two-pronged approach. First, in a HDNNP, the total energy E of a molecule containing N atoms is expressed as a sum of individual atomic energy contributions E_i:

$$E = \sum_i^N E_i. \tag{2.2}$$

These contributions depend on the chemical environment around each atom i and are modeled by feed-forward NNs, where one NN is used for atoms belonging to the same element. An example for a HDNNP is shown in Figure 2. Due to this structure, HDNNPs no longer depend on the order of the input coordinates and can now describe molecules containing a varying number of atoms. If the size of the molecule changes, the corresponding terms simply have to be added or removed from the sum in Equation 2.2. Second, the local chemical environment of an individual atom i is described by a set of special atom-centered symmetry functions[75] (ACSFs) $\{G_i\}$. These ACSFs are many-body functions obtained from the Cartesian coordinates $\{R_i\}$ and by construction invariant with respect to rotations and translations of the molecule.

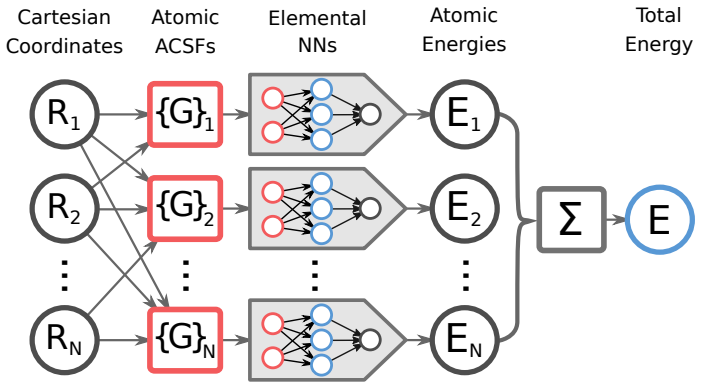


Figure 2: Schematic representation of a high-dimensional neural network potential. Each Cartesian coordinate is transformed into a set of atom-centered symmetry functions (ACSFs) $\{G_i\}$ for every atom i, describing the atoms chemical environment. Using these ACSFs as inputs, the energy contribution E_i of the atom is predicted by the corresponding elemental neural network (NN). The total molecular energy E is obtained as the sum over these contributions.

Analytic expressions for molecular forces, which are required for e.g. molecular dynamics simulations, are readily available within the HDNNP formalism.[74] The force with respect to the Cartesian coordinates $\mathbf{R}_i$ of an atom i is obtained via the relation

$$\mathbf{F}_i = -\sum_{j}^{N}\sum_{\varsigma}^{S_j}\frac{\partial E_j}{\partial G_{j\varsigma}}\frac{\partial G_{j\varsigma}}{\partial \mathbf{R}_i}, \tag{2.3}$$

where the first term in the sum is the partial derivative of the elemental NNs with respect to the ACSFs and the second term the partial derivative of the ACSFs with respect to the Cartesian coordinates $\mathbf{R}_i$. S_j is the number of ACSFs used to describe the environment of atom j.

2.3 ATOM-CENTERED SYMMETRY FUNCTIONS (ACSFS)

The ACSFs[75] $\{G_i\}$ representing the chemical environment of an atom i are computed from the Cartesian coordinates $\{\mathbf{R}_j\}$ of all neighboring atoms within a sphere around i. This sphere is defined via the cutoff function

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi r_{ij}}{r_c}\right)+1\right], & r_{ij} \leq r_c \\ 0, & r_{ij} > r_c, \end{cases} \tag{2.4}$$

where r_{ij} is the distance between the central atom i and its neighbor j and r_c is a predefined cutoff radius. By introducing this cutoff, the description of the local chemical environment is focused on the chemically relevant regions. As a consequence, the computational cost of HDNNPs scales linearly with system size.

In order to describe the atomic environment within this sphere, a combination of radial and angular ACSFs is employed. Radial ACSFs are distribution functions constructed from Gaussian functions according to

$$G_i^{\mathrm{rad}} = \sum_{j\neq i}^{N} e^{-\eta(r_{ij}-r_s)^2} f_c(r_{ij}), \tag{2.5}$$

where η controls the width and r_s the offset of the Gaussians (see Figure 3). Different radial ACSFs are defined for every chemical element present in the environment of the central atom.

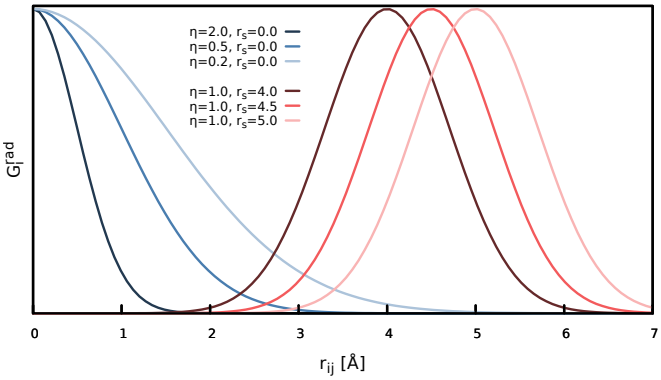


Figure 3: Radial ACSF terms $e^{-\eta(r_{ij}-r_s)^2}$ with different values for the width η (given in Å^{-2}) and the offset r_s (in Å) of the Gaussian function. No cutoff function is used in this example.

ACSFs describing the angular environment take the form

$$G_i^{\text{ang}} = 2^{1-\zeta} \sum_{j,k\neq i}^{N} \left(1 + \phi\cos(\theta_{ijk})\right)^{\zeta} e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)} \times f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}). \quad (2.6)$$

Here, θ_{ijk} is the angle spanned by atom i and its neighbors j and k. The first term in the sum describes the distribution of angles around the central atom. The parameter ϕ can take the values $\phi = 1$ and $\phi = -1$ and shifts the maximum of the angular term between 0° and 180°, while ζ regulates its width (Figure 4). η once again controls the width of a Gaussian function describing the radial arrangement of the atoms. As is the case for the radial functions, individual angular ACSFs are constructed for every possible pair of chemical elements present among the adjacent atoms.

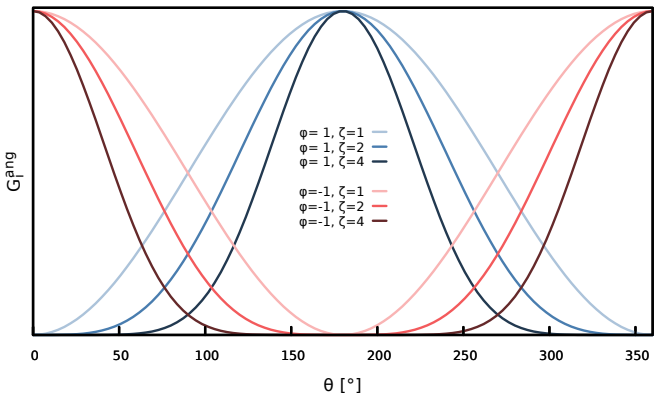


Figure 4: Angular term $2^{1-\zeta}\left(1 + \phi\cos(\theta_{ijk})\right)^{\zeta}$ of an ACSF, varying in the parameters chosen for the phase factors ϕ and widths ζ.

Typically, a set of ACSFs $\{G_i\}$ consisting of several radial and angular functions using a range of different values for η, r_s, ϕ and ζ and covering all possible elemental

combinations is used to describe the local environment of an atom. Appropriate values for the individual parameters have to be determined empirically. A detailed description on various aspects of ACSFs can be found in Reference [75].

2.4 NN TRAINING

In order to obtain valid ML models of molecular PESs, a NN potential must first learn to reproduce the electronic structure energies of a set of reference geometries in a process called training. During training, the weight parameters $\mathbf{w}$ of a NN are adapted iteratively in order to minimize a loss function of the type

$$\mathcal{L}(\mathbf{w}) = \frac{1}{M} \sum_{\iota}^{M} \mathcal{L}_{\iota}(\mathbf{w}). \tag{2.7}$$

Here, M is the number of molecules in the reference data set. $\mathcal{L}_{\iota}(\mathbf{w})$ is the error between the NN prediction and the electronic structure energy value of molecule ι which should be minimized. One commonly used loss function is the mean squared error (MSE), which measures the prediction error of the NN as

$$\mathcal{L}_{\iota}(\mathbf{w}) = \frac{1}{2} \left(E_{\iota} - \widetilde{E}_{\iota}(\mathbf{w}) \right)^2 . \tag{2.8}$$

E_{ι} is the electronic structure energy computed for molecule ι, while $\widetilde{E}_{\iota}(\mathbf{w})$ is the energy predicted by the NN using the current set of weights $\mathbf{w}$. Since different strategies can be used to adjust the weights in a manner that minimizes the MSE, a variety of algorithms suitable for training NNs exist.

2.4.1 *Stochastic Gradient Descent (SGD)*

Perhaps the most widely used NN training algorithm is SGD [1]. In SGD, the weight vector is adapted every iteration κ according to

$$\mathbf{w}_{\kappa+1} = \mathbf{w}_{\kappa} - \gamma \nabla \mathcal{L}_{\iota}(\mathbf{w}_{\kappa}), \tag{2.9}$$

where $\nabla \mathcal{L}_{\iota}(\mathbf{w}_{\kappa})$ is the gradient of the prediction error for sample ι with respect to the network weights $\mathbf{w}_{\kappa}$ and determines the direction of the update. γ is the so-called learning rate, a scalar factor which controls the size of the update step. Typically, several passes through the reference data set are performed in order to minimize the loss function. One such a pass over all samples is referred to as an epoch. In SGD, the weights are updated for every training example, using the associated gradient $\nabla \mathcal{L}_{\iota}(\mathbf{w}_{\kappa})$. This procedure is referred to as stochastic or online learning and has several advantages when training NNs over batch learning, where the gradient is first accumulated for every sample and the weights are adjusted once per epoch (see e.g. References 1 and 2). In batch or standard gradient descent, the weight update takes the form:

$$\mathbf{w}_{\kappa+1} = \mathbf{w}_{\kappa} - \frac{\gamma}{M} \sum_{\iota}^{M} \nabla \mathcal{L}_{\iota}(\mathbf{w}_{\kappa}), \tag{2.10}$$

using an average of the individual sample gradients $\nabla \mathcal{L}_{\iota}(\mathbf{w}_{\kappa})$. Consequently, batch optimization is advantageous if the loss function surface is relatively smooth or

convex. However, if many local minima and maxima are present, as is the case for NNs, online learning algorithms are generally faster, handle redundancy in the reference data better and converge to better minima.

When using a standard feed-forward NN to model a PES, the gradient of the prediction error $\mathcal{L}_\iota(\mathbf{w}_\kappa)$ becomes

$$\nabla \mathcal{L}_\iota(\mathbf{w}_\kappa) = \nabla \frac{1}{2} \left(E_\iota - \widetilde{E}_\iota(\mathbf{w}_\kappa) \right)^2 \tag{2.11}$$

$$= -\nabla \widetilde{E}_\iota(\mathbf{w}_\kappa) \left(E_\iota - \widetilde{E}_\iota(\mathbf{w}_\kappa) \right) \tag{2.12}$$

after substituting Expression 2.8 and applying the chain rule. E_ι and $\widetilde{E}_\iota(\mathbf{w}_\kappa)$ are once again the electronic structure and NN energies. The term $\nabla \widetilde{E}_\iota(\mathbf{w}_\kappa)$ is the matrix of first derivatives of the NN with respect to the current weights $\mathbf{w}_\kappa$, which is also commonly referred to as the Jacobian $\mathbf{J}_\kappa$. $\mathbf{J}_\kappa$ can be computed easily via the backpropagation algorithm[76], which propagates the local derivatives of each network node backward through the NN by recursively applying the chain rule starting from the output layer. The bracketed expression in Equation 2.12 is the difference between the reference value and NN prediction ν_κ obtained with the weights $\mathbf{w}_\kappa$. By using these two conventions and substituting Expression 2.12 into Equation 2.9, the SGD update for a feed-forward NN potential becomes

$$\mathbf{w}_{\kappa+1} = \mathbf{w}_\kappa + \gamma \mathbf{J}_\kappa \nu_\kappa. \tag{2.13}$$

SGD is a extremely versatile and robust algorithm and works well in practice for a wide range of NN architectures. Unfortunately, standard SGD can exhibit pathological behavior.[77] The learning rate γ regulating the size of the update step in SGD is the same for every individual parameter of the weight vector $\mathbf{w}$. However, using the same step size for every update direction can be problematic, e.g. when the surface spanned by the loss function exhibits narrow valleys. An example is shown in Figure 5. In cases like this, different learning rates would be required for every

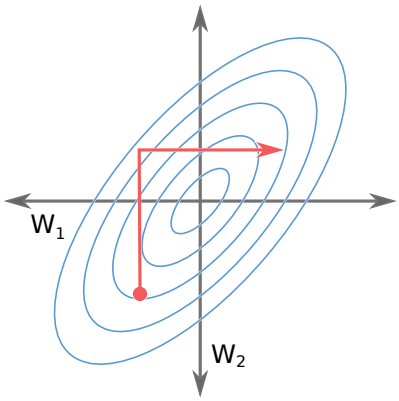


Figure 5: Pathological behavior of stochastic gradient descent training in a narrow valley of the loss surface. Since the update step uses the same magnitude γ in all directions, it overshoots the minimum and achieves only suboptimal convergence.

weight in order to achieve good convergence. Using a fixed learning rate instead, the training process might either diverge or converge very slowly, depending on the initial choice of γ. To overcome this problem, several adaptations to standard SGD have been developed, where e.g. the learning rate is annealed or varied for

every weight. Some of the most widely used SGD variants are Nesterov's accelerated gradient descent[78], AdaGrad[79] and ADAM[80].

2.4.2 *Global Extended Kalman Filter (GEKF)*

An elegant solution to the above problem of determining optimal update magnitudes are training algorithms, which incorporate information on the second-order derivatives – the curvature – of the error surface into the weight update. One such second-order training algorithm for NNs is the GEKF[81,82]. This algorithm is an adaptation of the Kalman filter originally used in signal processing. The regression for a GEKF training step is given by

$$\mathbf{w}_{\kappa+1} = \mathbf{w}_\kappa + \mathbf{K}_\kappa \nu_\kappa \tag{2.14}$$

$$\mathbf{K}_{\kappa+1} = \mathbf{P}_\kappa \mathbf{J}_\kappa {\mathbf{A}_\kappa}^{-1} \tag{2.15}$$

$$\mathbf{A}_\kappa = \lambda_\kappa \mathbf{I} + \mathbf{J}_\kappa^{\mathrm{T}} \mathbf{P}_\kappa \mathbf{J}_\kappa \tag{2.16}$$

$$\mathbf{P}_{\kappa+1} = \lambda_\kappa^{-1} \left[\mathbf{I} - \mathbf{K}_\kappa \mathbf{J}_\kappa^{\mathrm{T}}\right] \mathbf{P}_\kappa. \tag{2.17}$$

Here, $\mathbf{K}_\kappa$ is the so-called Kalman gain matrix and $\mathbf{P}_\kappa$ is the filter covariance matrix. $\mathbf{A}_\kappa$ is a global scaling matrix, where $\mathbf{I}$ represents the identity matrix. A time varying forgetting schedule is introduced into the GEKF regression, in order to avoid premature convergence to local minima. The associated forgetting factor λ_κ is computed as $\lambda_\kappa = \lambda_{\kappa-1}\lambda_0 + 1 - \lambda_0$, where the initial values for λ_κ and λ_0 are chosen close to unity.

Similar to the SGD algorithm (compare Equations 2.13 and 2.14), the update direction of $\mathbf{w}_\kappa$ is given by the current Jacobian $\mathbf{J}_\kappa$. However, instead of using the same scaling factor γ for every direction, the magnitude of the GEKF update is instead controlled via the filter covariance matrix $\mathbf{P}_\kappa$. $\mathbf{P}_\kappa$ is a weighted history of Gauss–Newton approximations to the inverse Hessian of the loss function and hence introduces second-order information into the update step. As a consequence, the GEKF algorithm exhibits superior training and convergence properties compared to SGD.

However, this improved performance of second order algorithms comes at an increased computational cost due to the additional matrix multiplications which have to be performed during every training step. In the case of GEKFs, this additional cost can be reduced significantly by introducing a so-called adaptive filter limit. If the NN error ν_κ is below this threshold for a molecule, no significant improvement of the ML model will be attained and the weight update step is skipped. Hence, the GEKF algorithm focuses computational resources on productive update steps and a significant speed up of the training procedure is achieved. Typically, a fraction of the overall RMSE computed during the previous training epoch is used as the filter limit.

While GEKF training can be applied out of the box to standard NN potentials, additional modifications are required for it to work with HDNNPs. This need arises due to the special structure of HDNNPs: The term $\widetilde{E}_\iota(\mathbf{w}_\kappa)$ in the prediction error (Equation 2.8) is now a sum of atomic contributions depending on the weight vectors of different NNs. Moreover, HDNNPs employ one individual NN for every chemical element. Further adaptations are necessary if molecular forces should be incorporated into the training process.

2.4.3 *Weight Initialization*

Before the training process can be started, a set of initial values for the weight vector $\mathbf{w}_{\kappa=0}$ needs to be chosen. This initial choice of weights can influence the training process dramatically[83]: If the weights are too small, the gradients backpropagated through the NN shrink uncontrollably and the training process slows down. If the weights are too large, the gradients explode and the training procedure begins to diverge. To counteract this phenomenon, it is desirable to select the weights in a manner, which allows the overall distributions of signals and gradients to maintain a variance close to unity and a mean close to zero as they pass through the NN layers.

Several heuristic initialization schemes have been developed to satisfy this criterion, drawing weights from specially modified random distributions. One scheme that works particularly well in practice was suggested by Glorot and Bengio in Reference 83. Here, the weights $\{w^l_{\alpha\beta}\}$ connecting the neurons in layer l to those in the previous layer $l-1$ are drawn from the uniform random distribution

$$\{w^l_{\alpha\beta}\} = \mathcal{U}\left[-\sqrt{\frac{6}{n^{l-1}+n^l}}, \sqrt{\frac{6}{n^{l-1}+n^l}}\right] \tag{2.18}$$

where n^{l-1} and n^l are the numbers of neurons in the previous and current layer. The bias weights $\{w^l_{0\beta}\}$ are initialized as zero in this scheme.

2.4.4 *Early Stopping*

An important issue encountered during the training of NNs is overfitting. When overfitting, a ML model no longer describes the relationships underlying the data, but instead memorizes minor variations in the training samples. The resulting model exhibits poor generalization performance: While it is able to reproduce the reference data with almost perfect accuracy, its predictions fail for new data not encountered during training. NNs are especially prone to overfitting, due to their large number of internal parameters in the form of the weights.

A simple strategy to avoid overfitting is early stopping (see Figure 6).[1] In early

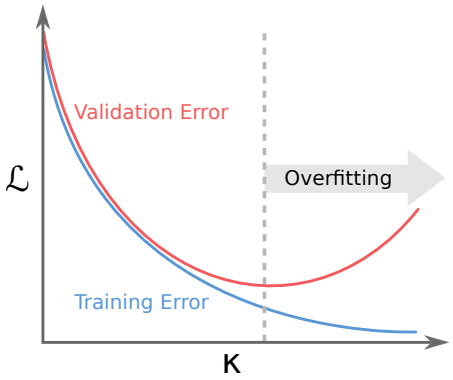


Figure 6: Evolution of the loss function for the training and validation set during training. While the training error decreases continuously, the validation error begins to deteriorate past a certain number of iterations. This behavior marks the onset of overfitting and in the early stopping scheme training is stopped at this point.

stopping, the reference data set is split into a training and validation set. The NN is

then trained based on the data in the training set, while error measures (e.g. root mean squared error) are monitored for both – training and validation set. The error computed for the validation set serves as an approximation to the generalization performance of the model. Initially, both errors should decrease during training, as the NN learns to model the relations encoded in the samples. However, past a certain point the validation error begins to deteriorate, while the training error continues to improve. This behavior indicates the onset of overfitting, the training is stopped and the weight vector $\mathbf{w}_\kappa$ resulting in the minimum validation error is returned.

2.5 MOLECULAR DYNAMICS SIMULATIONS

One field of application for ML potentials are molecular dynamics (MD) simulations. MD is a simulation technique used to describe the dynamical evolution of a molecular system.[84,85] In MD, the nuclei of a molecule move classically on the associated PES. This is achieved by numerically solving Newton's equations of motion[86] based on the forces acting on the individual particles. The required forces, as well as potential energies, can be obtained using different potentials. MD simulations provide invaluable insights into molecular motion at an atomic resolution and are employed to study a wide range of phenomena. Examples include the folding of proteins, solvent effects or sampling molecular configurations.

2.5.1 *Potentials*

As stated previously, the molecular forces and energies necessary in MD simulations can be obtained in various ways. Perhaps the most commonly used potentials are empirical FFs[14,15], which have been parametrized for certain classes of systems in order to reproduce relevant experimental and/or theoretical properties. Due to the simple functions used at the core of FFs, forces and potential energies can be obtained extremely efficiently, making it possible to simulate systems containing several hundreds of thousands of atoms. This high computational efficiency comes at the cost of overall accuracy and disadvantages of classical FF include their inability to e.g. describe bond breaking and bond formation or the coordination geometries of transition metal complexes.

More reliable descriptions of a chemical system can be obtained by computing molecular forces and energies via electronic structure methods. The practice of using quantum mechanics to simulate the forces acting on the nuclei while modeling their overall motion classically is referred to as *ab initio* MD (AIMD).[87] AIMD can be used to study phenomena not accessible with conventional FFs, such as chemical reactions or excitation processes.[88–92] Unfortunately, AIMD simulations are very costly due to the underlying need for electronic structure computations. As a consequence, the range of problems which can be modeled with this technique is limited.

ML potentials and HDNNPs in particular represent a promising alternative to the above potentials. Since HDNNPs employ ML based techniques in order to closely reproduce the results of electronic structure computations, they inherit the excellent computational efficiency of these techniques, exhibiting speeds on par with classical FFs and thus overcoming one of the inherent limitations of AIMD simulations. At the same time, the powerful functional forms provided by ML empower HDNNPs with the ability to describe effects not accounted for by standard FFs, e.g. chemical reactions. The combination of these properties, as well as the availability of analytic

forces (see Equation 2.3), makes them well suited for MD simulations, offering the possibility to achieve the accuracy of AIMD simulations at only fraction of the original cost.

2.5.2 *Thermostats*

In some cases it is desirable to perform MD simulations at constant temperatures, e.g. if experimental conditions should be reproduced. This can be achieved by introducing thermostat algorithms, which couple a system to an external heat bath.[93]

Various thermostat algorithms exist (e.g. the Berendsen[94] or Andersen thermostats[95]), however not all of them are able to model realistic constant temperature conditions. A MD trajectory should ultimately be ergodic, meaning all energetically feasible regions of a molecule's phase space are actually accessed during long enough simulations. This condition of ergodicity is fulfilled by the Nóse–Hoover chain (NHC) thermostat.[96] The NHC algorithm is an extension of the Nóse–Hoover (NH) thermostat[97,98]. The NH thermostat expands the equations of motion by an additional degree of freedom, which corresponds to an external heat bath. Since the basic form of this thermostat is not ergodic for small and rigid molecular systems, the NHC extends the NH equations by coupling the bath variable to additional degrees of freedom, thus resembling a chain of baths. The resulting thermostat is robust as well as physically accurate and hence frequently used in MD and AIMD simulations.

2.5.3 *Metadynamics*

However, even if MD simulations fulfill the criterion of ergodicity and all relevant PES regions can in principle be accessed, problems can still arise if it becomes necessary to sample events encountered only infrequently. For example, chemical reactions or rotations along a peptide backbone occur only rarely and thus long simulation times are needed in order to sample them properly. For this purpose, different methods have been developed to accelerate the sampling process of these rare events.[99,100]

One such technique is metadynamics.[101] During metadynamics simulations, Gaussian bias potentials are deposited along a set of collective variables at every point of the PES visited by the MD trajectory. This procedure can be likened to filling the valleys present in the PES, thus making it easier to overcome adjacent energy barriers. As a consequence, regions which have already been explored previously are visited less frequently and the exploration of new parts of the PES is encouraged.

3 METHOD DEVELOPMENT

This chapter introduces the different new strategies and methods developed in this thesis. An initial discussion is dedicated to an improved training algorithm for HDNNPs, termed the element decoupled global extended Kalman filter (ED-GEKF). Subsequently, a HDNNP-based fragmentation method is explored with the aim of reducing the computational effort of obtaining electronic structure reference data for large molecular systems. This is followed by the description of an adaptive sampling scheme for the fully automated construction of reference data sets. Finally, a ML model for molecular dipole moments based on the HDNNP architecture is introduced. Practical applications of these techniques and the associated results are presented in Chapter 4. The published articles detailing the above developments are reprinted in the corresponding sections of the Appendix (A.1-A.3).

3.1 ELEMENT-DECOUPLED GLOBAL EXTENDED KALMAN FILTER (ED-GEKF)

In order to successfully train HDNNPs, the basic GEKF algorithm needs to be modified to suit their special structure. Instead of using a single NN, HDNNPs model the PES of a molecule as a sum of NNs, where the same NN is used for all atoms belonging to one chemical element. Hence, the HDNNP prediction error for a molecule ι, which needs to be minimized during training, takes the form

$$\mathcal{L}_\iota(\mathbf{w}_\kappa) = \frac{1}{2}\left(E_\iota - \sum_i^{N_\iota} \widetilde{E}_i^{(\iota)}(\mathbf{w}_\kappa^{(Z_i)})\right)^2, \tag{3.1}$$

where $\mathbf{w}_\kappa^{(Z_i)}$ are the weights of the elemental subnet corresponding to the element Z_i of atom i. The term in brackets is the difference v_κ between the predicted and reference energies. The overall HDNNP weight vector $\mathbf{w}_\kappa$ is the combination of all elemental weight vectors $\mathbf{w}_\kappa^{(Z)}$.

The Jacobians $\mathbf{J}_\kappa^{(Z)}$ associated with the individual elemental NNs can be easily derived in this framework by applying the rules of differential calculus:

$$\mathbf{J}_\kappa^{(Z)} = \frac{\partial \sum_i^N \widetilde{E}_i}{\partial \mathbf{w}_\kappa^{(Z)}} \tag{3.2}$$

$$= \sum_i^N \frac{\partial \widetilde{E}_i}{\partial \mathbf{w}_\kappa^{(Z)}} \tag{3.3}$$

$$= \sum_i^N \mathbf{J}_\kappa^{(i)} \delta_{Z_i,Z}. \tag{3.4}$$

The expression $\delta_{Z_i,Z}$ is one if the element of the current atom Z_i is the same as the elemental index Z of the Jacobian and zero otherwise. Hence, the elemental Jacobian $\mathbf{J}^{(Z)}$ is simply the sum of the Jacobians associated with all atoms sharing the same element. However, it is not straightforward on how to implement GEKF training based on these Jacobians. While different schemes applying independent GEKF optimizations to the elemental subnets seem feasible at a first glance, they

all suffer from the same disadvantage. Since no reference values for the individual atomic energy contributions $\widetilde{E}_i$ are known, all schemes of this type would need to introduce some kind of *ad hoc* partitioning of the energy error ν_κ.

A potential solution to this problem can be found by viewing the HDNNP scheme as a large composite NN (see Figure 7). Depending on the molecule to be modeled,

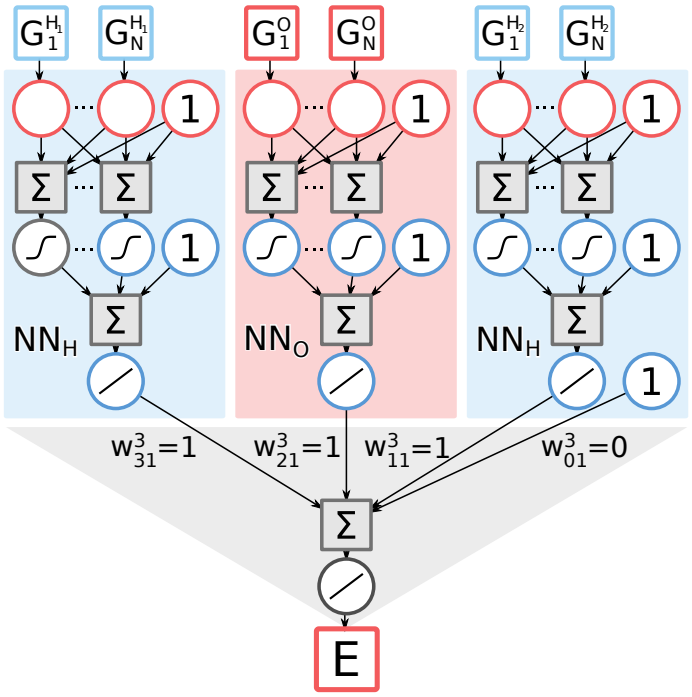


Figure 7: A high-dimensional neural network potential (HDNNP) for the water molecule represented as one composite neural network (NN). The final summation in HDNNPs corresponds to an output layer with a linear transfer function using constant weights of one and a bias weight of zero. The elemental NNs predicting the energy contribution of the two hydrogens share the same weights.

the HDNNP can vary in size, by adding or removing groups of neurons in the form of the elemental subnets. The last layer of this composite NN – corresponding to the summation step – can simply be seen as an output layer with linear activation functions, a bias of zero and all weights equal to one. The most important feature of this composite picture, is the fact that the groups of neurons belonging to different atoms are disconnected and can hence be assumed to be independent of each other. In the context of the GEKF algorithm, this means that the error covariances between these terms can be neglected and the covariance matrix $\mathbf{P}$ of the composite NN takes a block diagonal form.

This bears striking parallels to the decoupled Kalman filter[102,103], a variant of the GEKF initially proposed to reduce its overall computational cost. In the decoupled Kalman filter, sets of weights are assumed to be independent, leading to a similar block diagonal structure of the covariance matrix $\mathbf{P}$. As a consequence, the matrix multiplications necessary for a GEKF update can be broken down into multiplications of smaller matrices, leading to a significant improvement in computational efficiency.

By introducing a similar structure into the HDNNP covariance, the decoupled Kalman equations can be applied, thus making it possible to derive a suitable training

algorithm. In doing so, elemental updates arise naturally and the Kalman filter equations for HDNNPs become:

$$\mathbf{w}_{\kappa+1}^{(Z)} = \mathbf{w}_{\kappa}^{(Z)} + \mathbf{K}_{\kappa}^{(Z)} \nu_{\kappa} \tag{3.5}$$

$$\mathbf{K}_{\kappa+1}^{(Z)} = \mathbf{P}_{\kappa}^{(Z)} \mathbf{J}_{\kappa}^{(Z)} {\mathbf{A}_{\kappa}}^{-1} \tag{3.6}$$

$$\mathbf{A}_{\kappa} = \lambda_{\kappa} \mathbf{I} + \sum_{Z}^{\varepsilon} \left(\mathbf{J}_{\kappa}^{(Z)} \right)^{\mathrm{T}} \mathbf{P}_{\kappa}^{(Z)} \mathbf{J}_{\kappa}^{(Z)} \tag{3.7}$$

$$\mathbf{P}_{\kappa+1}^{(Z)} = \lambda_{\kappa}^{-1} \left[\mathbf{I} - \mathbf{K}_{\kappa}^{(Z)} \left(\mathbf{J}_{\kappa}^{(Z)} \right)^{\mathrm{T}} \right] \mathbf{P}_{\kappa}^{(Z)}. \tag{3.8}$$

This scheme is termed the "element-decoupled" GEKF or ED-GEKF for short. $\mathbf{K}_{\kappa}^{(Z)}$ is the Kalman matrix associated with element Z. $\mathbf{P}_{\kappa}^{(Z)}$ is the error covariance of the weights $\mathbf{w}_{\kappa}^{(Z)}$ of an elemental subnet and corresponds to one diagonal block of the covariance matrix of the compound NN. ε is the number of different elements in the current sample. An important feature of the decoupled scheme is how the global scaling matrix $\mathbf{A}_{\kappa}$ is computed (Equation 3.7). In the ED-GEKF, $\mathbf{A}_{\kappa}$ intrinsically depends on all subnetworks. This makes it possible to perform training on the whole HDNNP model using the total energy error ν_{κ} in the update, thus eliminating the need for any partition schemes.

The ED-GEKF can be adapted to incorporate molecular forces into the training process. This is achieved by expanding the loss function for the energies in Equation 3.1 by a term measuring the error in the force predictions of the HDNNP

$$\mathcal{L}_{\iota}(\mathbf{w}_{\kappa}) = \frac{1}{2} \left[\left(E_{\iota} - \sum_{i}^{N_{\iota}} \widetilde{E}_{i}^{(\iota)}(\mathbf{w}_{\kappa}^{(Z_i)}) \right)^2 + \frac{\vartheta}{3N_{\iota}} \sum_{i}^{N_{\iota}} \left| \mathbf{F}_{i}^{(\iota)} - \widetilde{\mathbf{F}}_{i}^{(\iota)}(\mathbf{w}_{\kappa}) \right|^2 \right]. \tag{3.9}$$

Here, $\mathbf{F}_{i}^{(\iota)}$ are the reference forces acting on atom i of molecule ι, while $\widetilde{\mathbf{F}}_{i}^{(\iota)}(\mathbf{w}_{\kappa})$ are the corresponding HDNNP forces, computed via Relation 2.3. ϑ is a factor used to control the influence of the force errors relative to the energy errors during training. The resulting force version of the ED-GEKF only changes in the weight update step, which becomes:

$$\mathbf{w}_{\kappa+1}^{(Z)} = \mathbf{w}_{\kappa}^{(Z)} + \mathbf{K}_{\kappa}^{(Z)} \nu_{\kappa} + \frac{\vartheta}{3N_{\iota}} \mathbf{P}_{\kappa}^{(Z)} \sum_{i}^{N_{\iota}} \frac{\partial \widetilde{\mathbf{F}}_{i}^{(\iota)}}{\partial \mathbf{w}_{\kappa}^{(Z)}} \mathbf{B}_{\kappa}^{(i)} \xi_{\kappa}^{(i)}. \tag{3.10}$$

$\frac{\partial \widetilde{\mathbf{F}}_{i}^{(\iota)}}{\partial \mathbf{w}_{\kappa}^{(Z)}}$ is the derivative of the HDNNP forces of atom i with respect to the weights of the NN describing element Z. $\xi_{\kappa}^{(i)}$ is the difference between the HDNNP and reference forces $\mathbf{F}_{i}^{(\iota)} - \widetilde{\mathbf{F}}_{i}^{(\iota)}$ acting on atom i. Finally, $\mathbf{B}_{\kappa}^{(i)}$ is a scaling matrix, computed according to the relation

$$\mathbf{B}_{\kappa}^{(i)} = \left[\lambda_{\kappa} \mathbf{I} + \sum_{Z}^{\varepsilon} \left(\frac{\partial \widetilde{\mathbf{F}}_{i}^{(\iota)}}{\partial \mathbf{w}_{\kappa}^{(Z)}} \right)^{T} \mathbf{P}_{\kappa}^{(Z)} \frac{\partial \widetilde{\mathbf{F}}_{i}^{(\iota)}}{\partial \mathbf{w}_{\kappa}^{(Z)}} \right]^{-1}. \tag{3.11}$$

This scaling matrix is introduced for similar reasons as $\mathbf{A}_{\kappa}$: The derivatives of the forces with respect to the weights possess a component vector for every elemental subnet, where the exact contribution of each component to the atomic force error $\xi_{\kappa}^{(i)}$ is unknown.

The ED-GEKF exhibits superior convergence behavior and accuracy of the trained HDNNPs compared to other training algorithms. Since it is specially tailored to the structure of HDNNPs, it performs equally well for different chemical systems with a wide range of elemental compositions. If just one chemical element is present, the ED-GEKF reduces to the standard GEKF algorithm. Similar to the GEKF, the ED-GEKF can make use of an adaptive update strategy in order to reduce the overall training time. As was demonstrated above, forces can be included in the training process in a consistent manner, which is especially useful for molecular dynamics applications. The use of forces increases the overall computational cost due to the additional derivatives and matrix multiplications which need to be computed. However, the number of electronic structure reference computations required in order to construct a valid HDNNP is also reduced significantly as the additional information contained in the $3N$ force components of every molecule can be leveraged in addition to the energy. Moreover, a similar adaptive scheme as for the energies can be introduced in the force ED-GEKF in order to accelerate the training procedure. Further details on the ED-GEKF training scheme and its derivation in general, as well as on the choice of individual parameter settings can be found in the reprint of the original article provided in Section A.1.

3.2 HDNNP FRAGMENTATION

An important feature of ML potentials and HDNNPs in particular is their ability to model system sizes otherwise beyond the reach of the electronic structure methods they are based on. However, in order to obtain HDNNPs suitable for simulations, reference computations using the original electronic structure method are still necessary during training. For sufficiently large molecular systems, even a few of these reference calculations can pose an insurmountable bottleneck, thus ultimately restricting the utility of HDNNPs.

HDNNPs have the potential to overcome this problem by means of their special structure, as it imbues them with the ability to operate as a fragmentation method. Fragmentation methods partition the original molecule into several smaller fragments, for which electronic structure computations can still be carried out.[67,104] Subsequently, the energy of the unfragmented molecule is obtained through recombination of the fragment energies. In HDNNPs, small local fragments of the molecule are introduced implicitly via the cutoff spheres of the symmetry functions. The total energy of the molecule is then recovered as the sum of these fragment energies, modeled via NNs. As a consequence, HDNNPs are able to reconstruct the properties of large molecular systems based only on the information contained in small fragments in the same way as standard fragmentation methods. Thus, expensive reference computations never need to be performed for the whole molecule and the process of generating reference data essentially becomes linear scaling with respect to system size. This property of HDNNPs has already been exploited to model several solid state problems, such as surfaces[105] and metal oxides[28], but is relatively unexplored in the context of molecules.

A HDNNP fragmentation procedure typically proceeds along the following steps: First, the target molecule is partitioned into smaller fragments. The way these fragments are generated is an important aspect of every fragmentation method. While different alternatives are in principle possible, the most natural way to generate fragments in the context of HDNNPs is directly inspired by the cutoff spheres employed

by the symmetry functions (see Figure 8). Initially, the central atom of the fragment

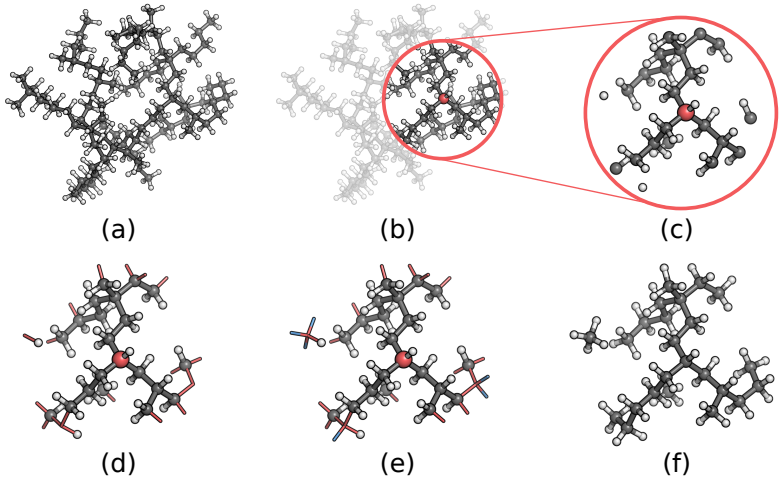


Figure 8: Generation of a molecular fragment using an alkane (a) as example. First, a cutoff sphere of a predefined radius is placed around the central atom (b). All atoms beyond this cutoff sphere are removed from the molecule (c). Afterwards, free valencies are saturated: Free single valencies are saturated with hydrogens. In the case of overlapping hydrogen caps or if the free valency itself is located on a hydrogen, the next heavy atom present at this position in the original molecule is included in the fragment (d). This procedure is repeated iteratively (e) until the final fragment (f) is obtained.

is chosen. Then, all atoms whose distance from the central atom exceed a predefined cutoff radius are removed. Usually, the same cutoff radius as in the symmetry functions is used. Finally, free valencies are saturated in an iterative procedure: I) Free single valencies are saturated with hydrogen atoms. II) If the free valency is situated on a hydrogen atom or the cut bond is a double or triple bond, the next heavy atom bonded to this position in the original molecule is included in the fragment. III) If two hydrogen caps would overlap in the final fragment, the heavy atom occupying their position in the unfragmented structure is instead included in the fragment. These steps are repeated until no free valencies remain. After a fragment has been obtained in this manner, the next atom is selected and the same procedure is applied. This results in one atom-centered fragment for every atom present in the molecule. Electronic structure computations are then performed for these fragments. Using the results of these calculations as a reference data set, a HDNNP is trained. Finally, the resulting HDNNP is used to predict the energy of the unfragmented molecule. Since no electronic structure computation has to be performed for the whole molecule in this way, significant speedups can be achieved using the HDNNP-based fragmentation method. A reprint of the study first introducing HDNNP-based fragmentation in the context of molecules is provided in Section A.2, while the further refinement of the method to its present form is discussed in the preliminary manuscript given in Section A.3.

3.3 ADAPTIVE SELECTION SCHEME

A central aspect of training HDNNPs is the selection of suitable electronic structure reference data. In order for a HDNNP to provide a reliable description of a chemical system, the reference data set needs to be representative for all relevant regions of the

associated PES. While it would in principle be possible to perform extensive sampling using only electronic structure methods (e.g via AIMD), this strategy is rarely feasible in a practical setting. First, high-level electronic structure computations can quickly become prohibitive due to their computational cost and should hence be kept to a minimum. Second, the overall advantage of introducing HDNNPs compared to directly simulating a problem with electronic structure methods decreases with the number of required reference calculations. Hence, selection schemes which can automatically generate a reference data set spanning all relevant regions of the PES with as few samples as possible are highly desirable.

A core component of any automatic selection scheme is a so-called uncertainty measure, providing an estimate of how well a sample is described by the current ML model and whether it should be included in the training procedure or not. Various uncertainty measures have been employed to guide the selection of reference data in the past, with geometry fingerprints[106] and Bayesian inference[61] being only a few examples. In the context of HDNNPs, it is possible to formulate a simple uncertainty measure, which exploits the high flexibility of NNs. This measure is based on the direct comparison of different HDNNPs and leads to a selection procedure of the following form: An initial reference data set is used to train several HDNNPs, differing in their initial sets of weights and/or their architectures. Using these preliminary HDNNPs, new molecular configurations are sampled (e.g. with MD). Subsequently, the predictions of the different ML models are compared to each other. If the differences between predicted values are small, the corresponding PES regions are represented well by the HDNNPs. Diverging predictions, however, indicate PES regions where the HDNNPs extrapolate and hence fail at providing an accurate description of the sample. The afflicted configuration is then computed using the electronic structure method of choice and added to the reference data set. Based on this expanded reference set, the HDNNPs are retrained and the whole procedure is repeated until the quality of the HDNNPs reaches a desired level.

While this procedure has been employed successfully for solid state systems, surfaces and metal clusters (see 74 and references within), additional modifications are necessary to make it suitable for molecules and expensive reference methods. One important adaptation introduced in this thesis is to combine the previously independent HDNNPs into an ensemble. New configurations are then sampled using the energies and forces of the HDNNP ensemble, which are computed according to

$$\overline{E} = \frac{1}{\mathfrak{N}} \sum_{\mathfrak{n}=1}^{\mathfrak{N}} \widetilde{E}_{\mathfrak{n}}, \tag{3.12}$$

$$\overline{\mathbf{F}} = \frac{1}{\mathfrak{N}} \sum_{\mathfrak{n}=1}^{\mathfrak{N}} \widetilde{\mathbf{F}}_{\mathfrak{n}} \tag{3.13}$$

and simply represent the average of $\mathfrak{N}$ HDNNP predictions. Based on these expressions, the predictive uncertainty of an ensemble of HDNNP takes the form

$$E_\sigma = \sqrt{\frac{1}{\mathfrak{N}-1} \sum_{\mathfrak{n}}^{\mathfrak{N}} \left(\widetilde{E}_{\mathfrak{n}} - \overline{E}\right)^2}. \tag{3.14}$$

The introduction of ensembles in the adaptive sampling scheme brings several distinct advantages. The uncertainty measure in Equation 3.14 grows more reliable as larger ensembles are used, since it is increasingly unlikely for a large number of

HDNNPs to show a similar behavior in regions where they extrapolate. Moreover, ensembles provide a significant increase in predictive accuracy compared to isolated HDNNPs. This is a direct consequence of a cancellation of errors between the individual HDNNPs and can lead to a reduction of the ensemble prediction error by a factor of $\frac{1}{\sqrt{\mathfrak{N}}}$. By exploiting this property, ensemble based simulations are more robust and reliable, which is especially important in the early exploratory stages of the adaptive sampling scheme, where only very rudimentary HDNNPs would be available otherwise. With ensembles, only a few initial electronic structure reference points are required to dynamically grow a suitable reference data set. The resulting procedure is highly automated and progresses in the following manner (see Figure 9): An initial HDNNP ensemble is trained on a few starting configurations, which can

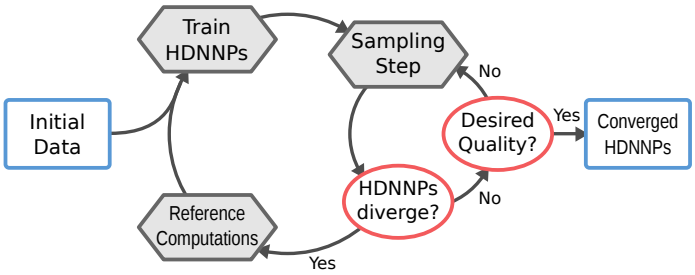


Figure 9: In the adaptive selection scheme, a preliminary HDNNP ensemble is first trained on a small number of initial configurations. Using this ensemble, new molecular configurations are sampled via e.g. molecular dynamics configurations. For every sampled configuration, the predictions of the ensemble HDNNPs are compared. If these predictions diverge, the sampling is stopped and reference calculations are performed for the afflicted configuration. The reference data set is then expanded by the new electronic structure data and a next generation of HDNNPs is trained. This procedure is repeated in an iterative fashion, until the divergence stays below a predefined threshold.

e.g. be obtained via a short AIMD simulation. Sampling runs are then carried out with the resulting ensemble using a suitable simulation method. In principle any sampling method can be used during this step, such as MD, metadynamics or Monte–Carlo algorithms[107]. During sampling, the uncertainty measure E_σ is monitored. If a predefined threshold is exceeded for a configuration, the simulation is stopped and electronic structure calculations are carried out. The reference data set is expanded by the additional data and the ensemble is retrained. Starting from this point, the simulations are continued and the process is repeated in a self-consistent fashion until convergence.

While the above procedure is effective for generating suitable reference data sets, it operates in a highly sequential manner. Two different strategies can be pursued to circumvent this limitation. One is to use only cheap electronic structure methods during the iterative refinement and then recompute the reference configurations at a higher level of theory. This so-called “up-scaling” is based on the assumption that the shape of both PESs are sufficiently similar and hence additional refinement steps might be necessary at the higher level of theory. Another strategy is to use multiple replicas of the molecular system during the sampling stage. In this case, independent simulations are carried out on copies of the system, using the same HDNNP ensemble, but e.g. different initial conditions. After sampling, all insufficiently described configurations are recomputed and added to the reference set. In this way, different problematic regions of the PES can be explored simultaneously.

A more detailed presentation of the adaptive sampling scheme is available in the original article reprinted Section A.3.

3.4 DIPOLE MOMENT MODEL

Although the primary area of application of HDNNPs is the description of molecular PES, their unique structure is also well suited to model a wide range of other properties. An excellent example for this versatility are HDNNP-based ML models of molecular dipole moments, which can for example be used to simulate infrared absorption spectra.[90]

In a classical system of point charges, the total dipole moment μ is defined as

$$\tilde{\mu} = \sum_{i}^{N} q_i \mathbf{r}_i, \tag{3.15}$$

where q_i is the partial charge of atom i and $\mathbf{r}_i$ is the distance vector from the molecular center of mass to the atom. Equation 3.15 bears a striking resemblance to the expression for the HDNNP energy (see Equation 2.2): In both cases, a global property is expressed as the sum of atomic properties, weighted by $\mathbf{r}_i$ in the case of μ. While both – energy and dipole moment – are quantum mechanical observables, the same does not hold for the individual atomic properties – atomic energies and charges. Therefore, it makes sense on an intuitive level to also use environment dependent NNs to model the atomic charges $\tilde{q}$ in Equation 3.15 in direct analogy to HDNNPs. The resulting dipole moment model is trained by minimizing the loss function:

$$\mathcal{L}_{\iota}^{\mathrm{Q}}(\mathbf{w}) = \frac{1}{2}\left(Q_\iota - \widetilde{Q}_\iota(\mathbf{w})\right)^2 + \frac{1}{3}\sum_{m}^{3}\frac{1}{2}\left(\mu_{\iota m} - \tilde{\mu}_{\iota m}(\mathbf{w})\right)^2 + \dots \tag{3.16}$$

Here, Q_ι is the total charge of the molecule ι and $\widetilde{Q}_\iota$ is the charge of the NN model, computed as the sum of NN charges $\widetilde{Q}_\iota = \sum_i^N \tilde{q}_i$. $\mu_{\iota m}$ is the mth Cartesian component of the dipole moment as computed with electronic structure methods, while $\tilde{\mu}_{\iota m}$ is computed according to Equation 3.15 using the NN charges $\tilde{q}_i$. The first term on the right-hand side serves as an additional constraint, driving the dipole moment model to reproduce the overall molecular charge. In principle, higher multipole moments can also be modeled via the above approach by including them in Expression 3.16. The resulting HDNNP-based dipole moment model is transferable. Once trained, it can be used to predict the molecular dipole moments of other systems differing e.g. in their size or composition, provided they are sufficiently similar from a chemical perspective.

Perhaps one of the most interesting properties of this model is related to the NN charges $\tilde{q}_i$. During training, only physical observables in the form of the total molecular charge and the dipole moment are used and the individual atomic NN charges are inferred based purely on statistical principles. Hence, the NN dipole model represents a charge partitioning scheme[108] offering access to environment dependent charges. Potential uses for these charges include the description of electrostatic interactions or to augment classical FFs, which typically employ charges that do not change based on the environment. If used in the latter manner, the above NN charges represent a potential alternative to polarizable FFs[109]. For further details on the dipole moment model, see the reprint provided in Section A.3.

4 RESULTS AND DISCUSSION

In this chapter, the procedures and algorithms developed in Section 3 will be applied to a range of chemical problems in order to investigate their overall efficacy. The Claisen rearrangement reaction of allyl vinyl ether to 4-pentenal serves as a test case for the ED-GEKF and its force variant (Figure 10a). HDNNP fragmentation will be

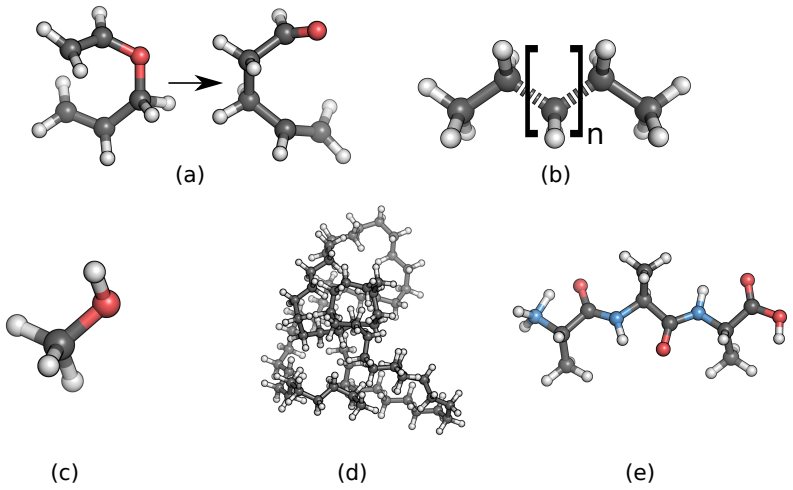


Figure 10: The chemical systems studied in the present thesis: The Claisen rearrangement reaction of allyl vinyl ether to 4-pentenal (a), linear all-trans alkanes of various lengths (b), the methanol molecule (c), different n-alkanes (d) and the protonated alanine tripeptide (e).

studied based on all-trans linear alkanes (Figure 10b), where a direct comparison to a conventional fragmentation method is carried out. Finally, the adaptive sampling scheme and dipole moment model are used in combination with the above developments in order to model the dynamic infrared (IR) spectra of different organic molecules. Here, an isolated methanol molecule serves as a simple test system to probe the general accuracy of the employed ML methods (Figure 10c). The applicability of the HDNNP fragmentation scheme for molecular dynamics simulations and the dipole moment model is analyzed using n-alkanes of different lengths as examples (Figure 10d). A last study uses the protonated alanine tripeptide (Figure 10e) to investigate the performance of the dipole moment model for species with a complicated charge distribution pattern. The articles containing the respective studies are reprinted in Sections A.1 to A.3 of the appendix.

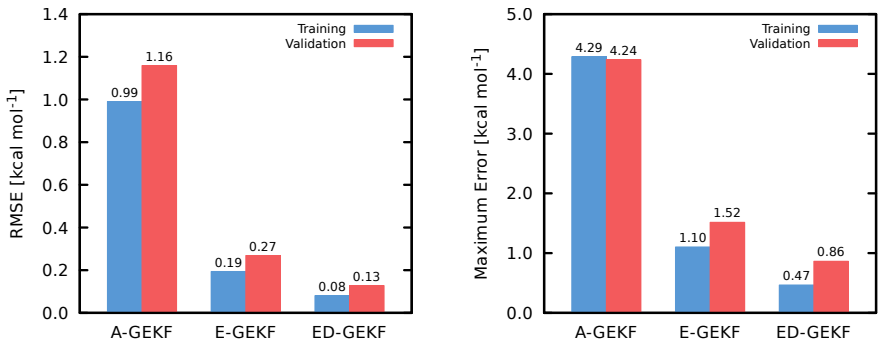
4.1 ED-GEKF TRAINING AND FORCES

The aliphatic Claisen rearrangement of allyl vinyl ether to 4-pentenal (Figure 10a) constitutes an excellent example to compare the performance of the ED-GEKF training algorithm to alternative GEKF variants. The reaction proceeds via bond breaking and formation events, which have to be modeled accurately by the different ML models. In addition, reference computations can be carried out with little computational effort, due to the relatively small size of the system. At the same time, different

chemical motifs are present, making it possible to investigate the applicability of HDNNPs to organic systems in general. Before the development of the ED-GEKF, two other variants of the Kalman filter were used to train HDNNPs, both of which rely on *ad hoc* partitioning of the HDNNP prediction error: In the first variant, the atomic GEKF or A-GEKF for short, the error is distributed evenly between all atoms and weight updates are performed for every atom individually. The second filter, an elemental variant called E-GEKF, splits the error between the different elements present in the molecule and introduces elemental updates of the weight vectors in a similar manner as the ED-GEKF. These two adaptations of the GEKF algorithm are now compared with the newly developed ED-GEKF. A metadynamics trajectory consisting of 17100 configurations computed at the BP86[110–114] level served as a common reference data set to investigate the performance of the ED-GEKF in relation to the two other variants. An in depth discussion of the different filter algorithms as well as the results and general setup of this study can be found in the original article reproduced in Section A.1 of the appendix.

4.1.1 *Accuracy*

The prediction accuracy with respect to the molecular energies of the final HDNNP models obtained with the different GEKF variants is given in Figure 11a. The root



(a) **(b)**

Figure 11: RMSEs (a) and maximum errors (b) associated with the predictions of HDNNPs trained with the different GEKF variants.

mean squared error (RMSE) is used as a performance measure. The training set contains 90 percent of the reference data points, while the validation set is formed by the remaining samples. In order to account for random effects introduced during the training procedure, five HDNNPs differing in their initial weights and partitioning of the data were trained with each GEKF algorithm and only the averaged results are shown.

The ED-GEKF algorithm outperforms the other variants significantly, achieving the lowest HDNNP prediction errors for both training and validation set (RMSEs of 0.08 kcal mol^{-1} and 0.13 kcal mol^{-1} respectively). Even compared to the E-GEKF, which already yields excellent results (0.19 kcal mol^{-1} and 0.27 kcal mol^{-1}), the ED-GEKF still manages to reduce all errors by more than half. Since both algorithms employ elemental updates, the main reason for this difference in performance can be found in the special way the ED-GEKF update is formulated. Using the decoupled

Kalman filter formalism as a basis, it is possible to model a HDNNP as one large composite NN (see Section 3). Hence, the individual weight updates in the ED-GEKF are no longer independent and no empirical partitioning of the error is required, as opposed to the E-GEKF. By far the worst HDNNPs are obtained with the A-GEKF algorithm, exhibiting significantly larger deviations from the electronic structure reference (0.99 kcal mol^{-1} and 1.16 kcal mol^{-1} respectively) than those trained with the other variants. This sub par prediction quality is caused by a combination of two effects. First, the A-GEKF treats every element independently, making it necessary to distribute the total error in the energy predictions evenly between all atoms during training. Moreover, since a weight update of the elemental subnets is performed for every atom individually in the A-GEKF, a bias in favor of more abundant elements is introduced. Based on the chemical composition of the system under investigation, the subnets of common elements, such as e.g. hydrogen, are therefore updated more frequently, leading to an imbalanced training algorithm. By introducing elemental updates in the E-GEKF and ED-GEKF algorithms, the above bias is eliminated and HDNNPs of increased quality can be trained. The use of the element-decoupled formalism in the ED-GEKF improves predictive accuracy further and a training algorithm especially well suited for the HDNNP architecture is obtained.

The above observations are even more pronounced with respect to the maximum deviations from the electronic structure reference obtained with the different filters shown in Figure 11b. Here, the largest deviations of the HDNNPs trained with the ED-GEKF are 0.47 kcal mol^{-1} for the training set and 0.86 kcal mol^{-1} for the validation set, which lie below the commonly accepted threshold for chemical accuracy of 1 kcal mol^{-1} and far below the uncertainty introduced by the electronic structure reference method.

4.1.2 *Filter Convergence*

In addition to the increased accuracy compared to the other GEKF versions, the ED-GEKF training algorithm also exhibits superior convergence behavior. The evolution of the training set RMSE for the HDNNPs trained with the different filter variants over 100 epochs is shown in Figure 12.

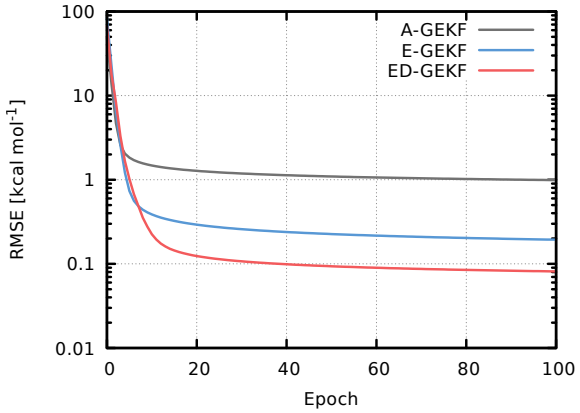


Figure 12: Logarithmic plot of the training set RMSEs yielded by the three different Kalman filter variants over the course of the training process.

Once again, the the ED-GEKF outclasses the two other filters and the corresponding HDNNP can be considered fully trained after 40 epochs. In contrast, the E-GEKF only reaches converge after approximately 60 epochs, while the A-GEKF shows no signs of converging even after 100 epochs, as is indicated by a noticeable slope present even towards the end of the training procedure. Moreover, the ED-GEKF already achieves significantly better results after only 10 epochs of training than the other variants at the end of a full run. This excellent behavior is a direct consequence of the decoupled formalism, which leads to an optimal scaling of the individual update steps during HDNNP training. Another implication of these results is, that training times can be reduced significantly by using the ED-GEKF, making it possible to obtain competitive HDNNP models after only a handful of iterations. Combined with the accuracies observed above, these findings serve to demonstrate the potential of this algorithm in general.

4.1.3 *Force Training*

An important feature of the ED-GEKF algorithm is the possibility to incorporate molecular forces into the training procedure in a consistent manner. Although the previously constructed HDNNPs can be used to predict forces (see Equation 2.3), only energies were used in their training. In order to investigate the effects of including forces, an additional HDNNP is trained on molecular forces and energies with the force variant of the ED-GEKF – termed ED-GEKF+F for short. Apart from the different training procedure, the new HDNNP is identical to the other models.

Subsequently, the RMSEs of the energies and forces predicted by the four different models are computed over the whole reference data set (Table 1). As can be seen in

Table 1: RMSEs of energies (kcal mol^{-1}) and forces (kcal mol^{-1} Å^{-1}) over the whole reference data set.

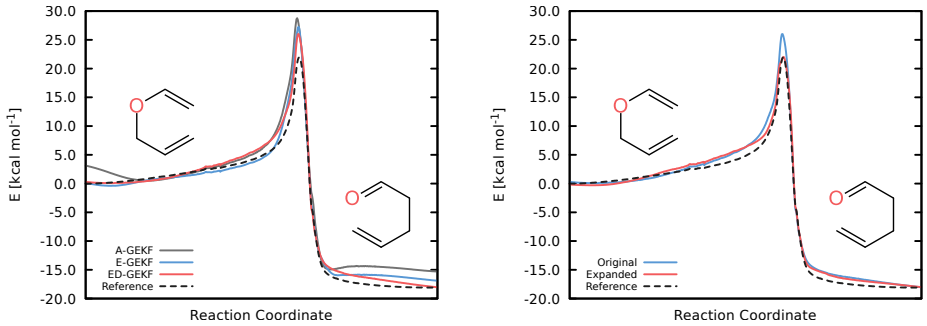
	RMSE	
Filter type	Energies	Forces
A-GEKF	0.86	18.66
E-GEKF	0.20	12.40
ED-GEKF	0.08	11.84
ED-GEKF+F	0.17	6.79

Table 1, the inclusion of forces has a noticeable effect on the overall shape of the PES predicted by the different HDNNPs. Compared to the standard ED-GEKF yielding a RMSE of 11.84 kcal mol^{-1} Å^{-1}, the ED-GEKF+F algorithm reduces the error in the HDNNP forces almost by a factor of two to 6.79 kcal mol^{-1} Å^{-1}. Hence, the force variant of the ED-GEKF training algorithm is especially useful for applications, where accurate molecular forces are required, such as MD simulations. The improved quality of the HDNNP forces, however, comes at a reduced accuracy with respect to the energy predictions, dropping from 0.08 kcal mol^{-1} (ED-GEKF) to 0.17 kcal mol^{-1} (ED-GEKF+F). The reason for this deterioration of the energy RMSE can be found in the additional fitting criterion introduced in the ED-GEKF+F (see Equation 3.9). When using the force variant of the ED-GEKF, the different influences exerted by the energy and force components of the update step need to be balanced. However, the present study does not yet use such a balanced implementation. Consequently, for a single

configuration containing N atoms, the magnitude of the force update outweighs the energy updates by a factor of $3N$ and much more emphasis is put on reducing the force RMSE during training. This observation has lead to the introduction of a scaling factor of $\frac{\vartheta}{3N}$ in all future applications, which yields much more balanced potentials, as can be seen in Section 4.3.

4.1.4 *Interpolation*

The main purpose of HDNNPs and other ML potentials is to reliably interpolate PESs based on electronic structure reference samples. To study the interpolation capabilities of HDNNPs trained with the different filter variants and the ED-GEKF in particular, they are use to predict the energies encountered along the reaction profile of the Claisen rearrangement. While the reference data set sampled via metadynamics provides a general representation of the reaction, no reference configuration lies exactly on the reaction path, making it an excellent test case. Figure 13a shows the potential energy curves predicted by the different HDNNP models for the 500 configurations sampled along the reaction profile. The associated overall deviations



(a) **(b)**

Figure 13: Energies along the reaction path connecting the allyl vinyl ether to 4-pentenal as predicted by HDNNPs trained with the three filter variants on the standard reference data set (a) and by a HDNNP trained with the ED-GEKF on an expanded data set (b) with the original ED-GEKF values shown in blue.

from the electronic structure energies, as well as the errors associated with the optimized ether, aldehyde and transition state structures can be found in Table 1. The least reliable model is produced by the A-GEKF algorithm, exhibiting an overall

Table 2: RMSEs of energies (kcal mol^{-1}) along the reaction path and deviations from the reference energy ΔE (kcal mol^{-1}) for reactant (ether), transition state (TS) and product (aldehyde).

		Ether	TS	Aldehyde
Filter type	RMSE		ΔE	
A-GEKF	2.48	3.12	6.81	2.88
E-GEKF	1.43	0.15	5.32	1.20
ED-GEKF	1.24	0.30	4.06	0.10

RMSE of 2.48 kcal mol^{-1}. Besides a significantly overestimated barrier height (ΔE of 6.81 kcal mol^{-1}), the associated HDNNP also introduces artificial minima in the

vicinity of the ether region and close to the barrier in the aldehyde. At the same time, the energies of the reactant and product are overestimated by approximately 3 kcal mol^{-1}. A more faithful reproduction of the reaction profile is achieved by the E-GEKF (RMSE of 1.43 kcal mol^{-1}). Especially the regions corresponding to the ether are modeled well and the error of the barrier is reduced to 5.32 kcal mol^{-1}. However, a much larger error is associated with the configurations leading to the aldehyde and a similar artificial minimum as in the case of the A-GEKF is found. By far the most accurate description of the reaction path is provided by the HDNNP trained with the ED-GEKF algorithm. Both – reactant and product sides – are described equally well and no artificial minima are introduced. In addition, the ED-GEKF potential exhibits the lowest error in the barrier height (4.06 kcal mol^{-1}). Although the regions close to the barrier are slightly overestimated, the ED-GEKF once again provides the best overall results in general with a RMSE of 1.24 kcal mol^{-1}.

While the profile produced with the ED-GEKF can be considered sufficiently accurate for most practical purposes, deviations are still present, especially close to the barrier. This finding indicates, that the associated regions of the PES are not represented well in the reference data set. In order to validate this assumption, the optimized structures of reactant and product, as well as the transition state, are included in the reference set and a new HDNNP is trained on the expanded data. The inclusion of these three points improves the reaction profile significantly (see Figure 13b), reducing the overall RMSE to 0.91 kcal $mol^{-1}$and thus confirming the previous suspicion. Since the use of a single metadynamics trajectory constitutes a rather naïve approach to the construction of a well balanced data set, it is little surprising to find undersampled regions of the PES. As can be seen above, this problem can in principle be solved by manually expanding the reference data in an appropriate manner using chemical intuition as a guidance. However, the ultimate goal is to completely automatize this potentially tedious procedure. This aspiration has lead to the development of the fully automated sampling scheme presented in Section 3.3 and evaluated in Section 4.3.

4.1.5 *Extrapolation*

A last study explores the behavior of HDNNPs in regions of the PES absent from the reference data set. This investigation is based on the dissociation of one of the hydrogen atoms bound to the sp^3 carbon atom of the allyl vinyl ether. The associated potential energy curves computed with the electronic structure reference and the different HDNNP models is depicted in Figure 14. Similarly as in the previous experiments, it is found that the HDNNP trained with the A-GEKF algorithm shows the worst performance, thus confirming that this GEKF variant is indeed ill suited for HDNNPs. The obtained potential energy curve only remotely resembles the electronic structure reference. For bond lengths ranging from 1.02 to 1.24 Å, both elemental filters – E-GEKF and ED-GEKF – produce scans of similar quality, with the RMSEs being in favor of the ED-GEKF (0.36 kcal mol^{-1} and 0.53 kcal mol^{-1}respectively). These regions correspond to the thermal fluctuations of the C-H bond encountered during the metadynamics sampling and are hence accounted for in the reference data set, explaining the good agreement. However, when leaving these regions, all HDNNPs show chaotic behavior and exhibit RMSEs close to 60 kcal mol^{-1} for the entire curve, independent of the training algorithm used in their construction. This behavior is an excellent demonstration for the high flexibility of HDNNPs and NNs

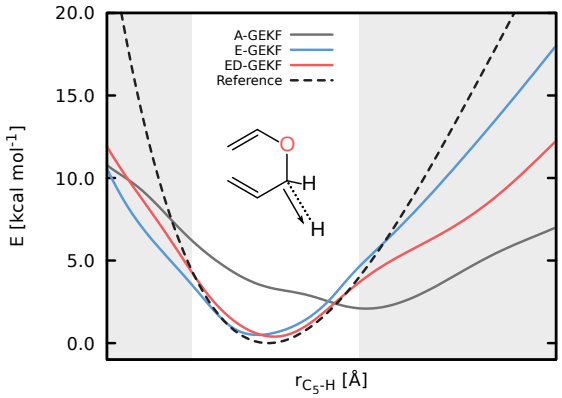


Figure 14: Reaction profiles predicted for the dissociation of one of the hydrogens bound to the sp^3 carbon atom of the allyl vinyl ether. The regions corresponding to bond lengths not sampled during the initial metadynamics simulation and hence not represented in the reference data set are highlighted in gray.

in general. In the regions of the PES represented in the reference data, properly trained HDNNPs show excellent agreement with the underlying electronic structure method, due to their ability to accurately model the complex relationships between configurations and potential energies. For configurations not sufficiently similar to those encountered during training, the HDNNPs begin to extrapolate wrongly and unphysical predictions are obtained. This behavior is essential for the adaptive sampling algorithm described in Section 3.3, as it can be exploited as an uncertainty measure for HDNNPs.

4.2 HDNNP FRAGMENTATION

The accuracy of HDNNP-based fragmentation is investigated using linear all-trans alkanes (Figure 10b) containing up to 10 000 carbon atoms as a model system. To this end, HDNNP fragmentation is compared to a conventional fragmentation method, the systematic molecular fragmentation (SMF) approach developed by Collins and co-workers[104]. An analysis of the deviations of both methods with respect to a highly accurate reference model provides insights into their general accuracy. This reference model is based on the CCSD(T)[12,115] energies for alkane chains containing up to 11 carbon atoms. While conventional electronic structure computations at the CCSD(T) level are impossible for the largest systems addressed in this study, an accurate reference model for chains of arbitrary length can be derived due to the systemic behavior SMF exhibits for this particular model system. Further details on this study and the associated results are provided in the reprint in Section A.2.

4.2.1 *Reference Model*

The SMF approach constructs fragments based on the functional groups present in a molecule – CH_3 and CH_2 groups in the case of alkanes (for a general description of SMF, see Reference 104). A predefined number of these functional groups are collected into overlapping fragments. The potential energy of the unfragmented molecule is then obtained by summing the energies computed for the individual

fragments and subtracting the energies of the "doubly counted" overlap regions. The exact number of the functional groups contained in these overlap regions corresponds to the SMF fragmentation level X. Typically, the error between the SMF approach and the original electronic structure method decreases as the fragmentation level and thus the size of the molecular fragments increase. Moreover, this error behaves in an especially systematic manner if linear all-trans alkane chains are used as a model system. This behavior can easily be recognized in Figure 15, which shows the SMF errors for different chain lengths (containing 5 to 11 carbon atoms) and fragmentation levels ranging from $X = 3$ to $X = 7$. Here, the errors associated with

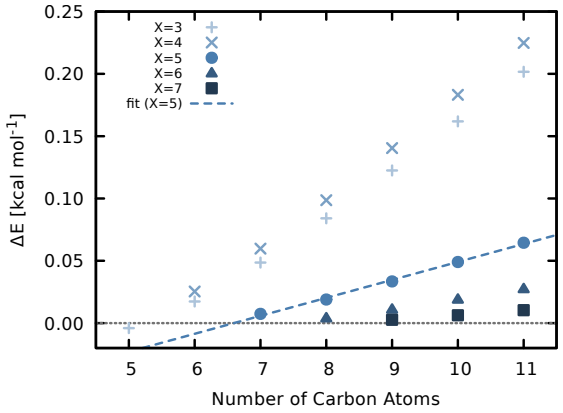


Figure 15: Deviation ΔE of the systematic fragmentation approach from the electronic structure reference in dependence on the length of the alkane chains and the employed fragmentation level ($X = 3 - 7$). For fragmentation $X = 5$, the linear fit used to derive the correction term is shown. The null line indicates the reference method.

a given fragmentation level X grow approximately linearly with respect to the size of the molecule. The reason for this phenomenon lies in the way the total energy of a molecule is recovered in the SMF approach: Due to the regular structure of the alkanes, longer chains are basically constructed by repeating small, chemically identical fragments. Hence, every additional CH_2 group contributes approximately the same error to the overall energy, leading to the observed linear behavior. This relation can be exploited to derive an empirical error correction for the SMF energies via a simple linear fit, as is shown for $X = 5$ in Figure 15. By applying this correction to standard SMF results, it is possible to recover almost the exact energies for all chains studied directly with the CCSD(T) method and it is expected that this trend also holds for the longer chains. An indicator for the validity of this assumption is the close agreement between the corrected energies obtained for the different fragmentation levels (within 1.1 kcal mol^{-1} for the longest alkane). In the present application, the model constructed from the level 5 fragmentations and the corresponding energy corrections is used as a reference. Note that this correction procedure is only possible in this case due to the linear nature of the alkane model systems.

4.2.2 *HDNNP Fragmentation*

Having obtained a reliable reference model for alkanes of arbitrary length with the above procedure, i.e., using SMF with the correction term, the fragmentation capabilities of HDNNPs are studied. To this end, a set of HDNNPs is trained on the

electronic structure energies of all chains ranging from C_5H_{12} to $C_{11}H_{24}$ and their corresponding fragments. The resulting HDNNP models are then used to predict the potential energies of alkane chains containing up to 10 000 carbon atoms. The minimum and maximum deviations achieved with HDNNP fragmentation compared to the reference model can be found in Figure 16 in addition to the deviations obtained for different levels of the original SMF approach without the correction term. Once again, a dependence of the accuracy of SMF on the fragment size is

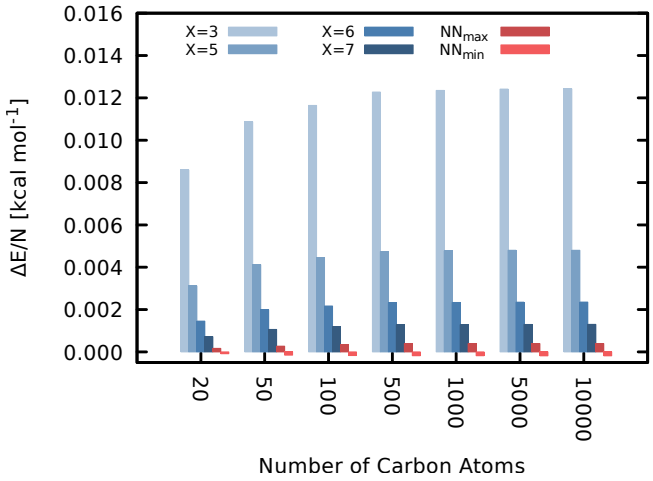


Figure 16: Energy deviations ΔE of the different fragmentation approaches from the reference model for alkane chains of various lengths. Shown are the uncorrected systematic molecular fragmentation levels $X = 3$, $X = 5$, $X = 6$ and $X = 7$, as well as the minimum (NN_{min}) and maximum (NN_{max}) deviations achieved with HDNNP-based fragmentation. ΔE is normalized by the total number of atoms N contained in each chain.

observed, with the best results being achieved with a fragmentation level of $X = 7$. However, all HDNNP models exhibit a significantly better agreement with the reference energies and even the HDNNP showing the largest deviations from the reference outperforms the highest SMF level fragmentation. The HDNNP with the smallest deviations closely reproduces the reference model. The high accuracy of the HDNNPs is remarkable insofar, as the ACSFs used in their construction employ a cutoff of 5 Å. As a consequence, the HDNNP models essentially operate on fragments containing seven carbon atoms, which corresponds to an SMF fragmentation level of $X = 6$. Yet significantly smaller deviations are found for the HDNNPs than for the equivalent SMF procedure. These observations indicate, that HDNNPs are able to exploit the local chemical information contained in small fragments to a much greater extent than the SMF method. Hence, HDNNPs constitute an interesting alternative to conventional fragmentation approaches. While the above results for HDNNPs are encouraging, it is not clear whether they hold for more realistic chemical systems. Hence, the fragmentation capabilities of HDNNPs with regards to more realistic systems and applications are investigated in Section 4.3.

4.3 ADAPTIVE SELECTION SCHEME AND DIPOLE MOMENT MODEL

An important application, where the selection scheme and the dipole moment model render efficient ML simulations possible is the prediction of molecular IR spectra.

One of the most accurate simulation techniques to model IR spectra is AIMD.[90] In AIMD, IR spectra are obtained via the Fourier transform of the time autocorrelation function of the time derivative of the molecular dipole moment $\dot{\mu}$:

$$I_{IR} \propto \int_{-\infty}^{+\infty} \langle \dot{\mu}(\tau)\dot{\mu}(\tau + t)\rangle_{\tau}\ e^{-i\omega t}\,\mathrm{d}t, \tag{4.1}$$

where ω is the vibrational frequency, τ is a delay time and t is the time. Since AIMD inherently accounts for the temporal evolution of a chemical system, it is able to model several effects which are typically neglected in conventional static approaches for computing IR spectra[90,116]. Among these phenomena, vibrational anharmonicities and temperature effects are of particular importance. The ability to provide accurate descriptions of both effects, makes AIMD an indispensable tool for the interpretation of experimental IR spectra of e.g. biological systems, where temperature and anharmonic phenomena play a significant role.

However, even a single AIMD trajectory requires a large number of electronic structure calculations (see Section 2.5). As a consequence, AIMD simulations suffer from a high computational cost and are usually limited to small systems and/or short timescales. By replacing the individual electronic structure calculations by significantly cheaper ML models, it is possible to overcome these limitations. This endeavor represents an excellent challenge to the different approaches developed previously, since several different aspects need to be considered when modeling IR spectra via Equation 3.15: First of all, an accurate HDNNP model is required to describe the time evolution of the investigated system. Such a model can only be constructed, if the reference data set is representative of all relevant regions of the molecular PES and hence constitutes an exemplary test case for the adaptive sampling scheme. Moreover, if IR spectra for large molecules should be predicted, HDNNP-based fragmentation is indispensable in order to carry out the required reference computations efficiently. With an appropriate reference data set at hand, ED-GEKF training would be employed to construct HDNNPs providing access to reliable molecular forces and energies. Finally, an accurate description of the molecular dipole moments is necessary, making it possible to assess the performance of dipole moment model. These different aspects are studied by modeling the IR spectra of three different chemical systems: The general accuracy and validity of the different developments is assessed based on the methanol molecule (Figure 10d). Different n-alkanes (Figure 10e) serve as a test case for the fragmentation method , which goes beyond the linear all-trans alkanes studied previously (see Section 4.2). Subsequently, the proficiency of the sampling scheme and the dipole moment model is analyzed by modeling the protonated alanine tripeptide (Figure 10e). The submitted manuscript containing a detailed description of the above study and its results is reprinted in Section A.3.

4.3.1 *Methanol*

The ML model used for methanol consists of an ensemble of two HDNNPs and a dipole moment model, which were trained on 245 reference configurations identified with the adaptive sampling scheme. The necessary electronic structure energies, forces and dipole moments were computed with the BP86 density functional. In order to study the accuracy of the individual components of this model, a conventional AIMD simulation spanning a time interval of 30 ps is performed. The ML model is then used to predict energies, forces and dipole moments for the 60 000 configurations

produced by this simulation. Figure 17 depicts the distribution of errors between ML predictions and the BP86 reference in blue.

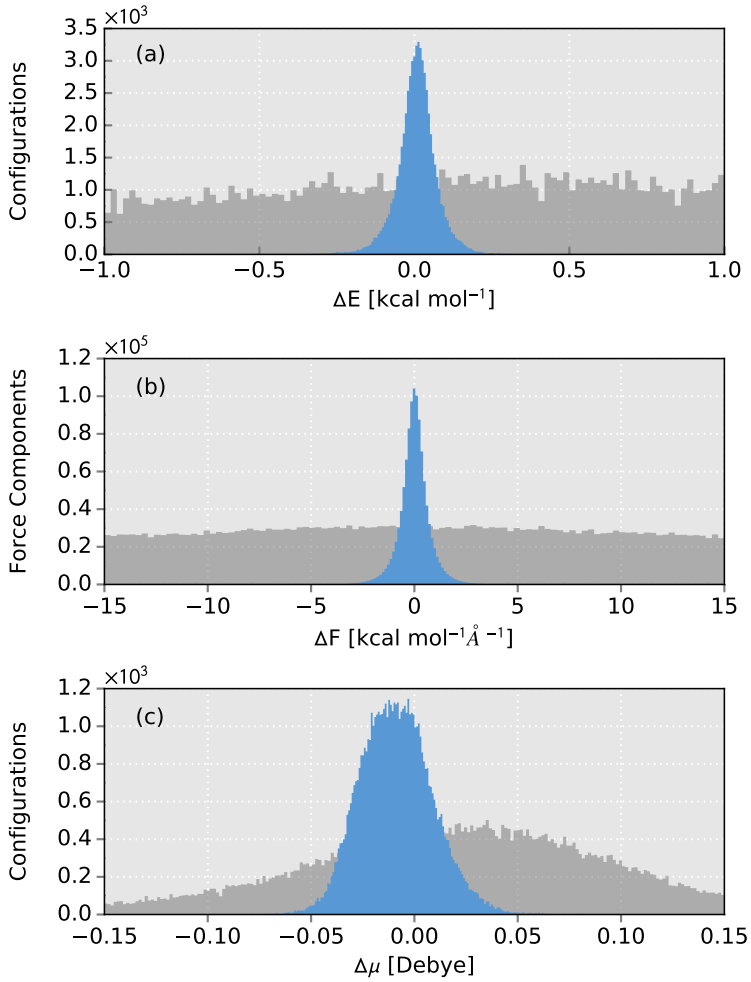


Figure 17: Distribution of the error between the ML predictions and the BP86 reference computed for the energies (a), force components (b) and dipole moments (c) of the 60 000 configurations sampled by the AIMD trajectory. The deviations associated with the ML model based on the adaptive sampling scheme are shown in blue, while those obtained for a ML model trained on configurations drawn randomly from a classical molecular dynamics simulation are depicted in gray.

In all instances, the ML model is able to reproduce the electronic structure results with excellent fidelity. With a mean absolute error (MAE) of only 0.048 kcal mol^{-1} the deviations observed for the energies (Figure 17a) are significantly smaller than the postulated limit of chemical accuracy (1.0 kcal mol^{-1}), as well as the error inherent to the BP86 method. A similar trend is observed for the molecular forces (Figure 17b), where a MAE of 0.533 kcal mol^{-1} Å^{-1} is achieved. These results are an excellent demonstration for the ability of the ED-GEKF training algorithm to construct high quality HDNNPs suitable for MD simulations based only on a small number of reference energies and forces. The molecular dipole moment model shows an equally satisfactory performance, exhibiting an overall MAE of 0.016 D for the magnitude of

the dipole moments and MAEs ranging between 0.0173 D and 0.0200 D with respect to the spatial orientation of the dipole vector.

After the analysis of its individual components, the ability of the composite model to simulate molecular IR spectra is investigated. A comparison of the IR spectra obtained via the ML approach (red) and the BP86 AIMD simulation (blue) is shown in Figure 18, together with an experimental spectrum recorded between 600 cm^{-1} and 4100 cm^{-1} for a methanol molecule in the gas phase[117] (gray). The ML spec-

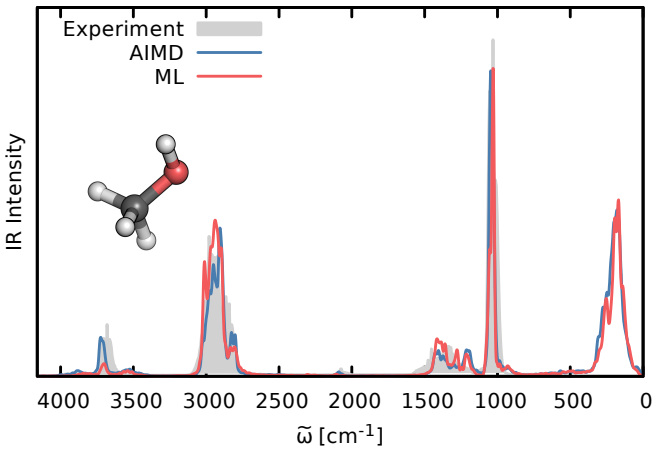


Figure 18: Gas phase IR spectrum of methanol, as predicted by the BP86 AIMD simulation (blue) and the ML model (red). The ML spectrum agrees closely with the BP86 reference and both methods accurately reproduce the experimental spectrum recorded in the range between 600 cm^{-1} and 4100cm^{-1} (gray).

trum shows excellent agreement with the AIMD spectrum, accounting for all peak positions and intensities in a highly accurate fashion. A minor deviation from the reference can be observed for the intensity of the O-H stretching vibration at 3700 cm^{-1}, which is underestimated in the ML spectrum. A possible reason for this discrepancy are small fluctuations in the dipole moment model. The general utility of both approaches – traditional AIMD and the ML based model – is attested by how closely they reproduce the experimental spectrum.

On a final note, it should once again be stressed, that the current ML model is only based on a small set of 245 BP86 reference points, while 60 000 single points were used in the AIMD simulation. Nevertheless, the ML model is able to accurately describe the PES and dipole moments of the methanol system, serving as a testimony to the effectiveness of the adaptive sampling scheme. Additional insights into the performance of the newly developed scheme are gained by comparing the accuracy of the current methanol model to one constructed from 245 configurations selected by a random selection scheme. As can be seen in Figure 17, the random ML model suffers from a significant reduction in accuracy, indicated by the wide error distributions associated with the different properties. This observation highlights the importance of a representative reference data set and, at the same time, serves as a confirmation for the high efficacy of the adaptive sampling scheme employed in this thesis.

4.3.2 *n-Alkanes*

The applicability of the HDNNP-based fragmentation approach to structurally diverse systems, as well as the ability of HDNNPs and the dipole moment model to describe molecules significantly larger than methanol is studied by using these three techniques to predict the IR spectrum of the $C_{69}H_{140}$ n-alkane (Figure 10d). In this study, the B2PLYP double hybrid density functional[118] is used as electronic structure reference method. B2PLYP provides highly accurate predictions of a molecule's electronic structure, albeit at considerable computational cost. As such, conventional AIMD simulations of the $C_{69}H_{140}$ IR spectrum using the B2PLYP method are currently close to impossible, demonstrating the potential of the ML approach pursued in this thesis. The composite ML model for this n-alkane – consisting of two HDNNPs and a dipole moment model – is constructed from the energies, forces and dipole moments of 534 molecular fragments. These fragments are generated from configurations selected by the adaptive sampling scheme via the fragmentation procedure described in Section 3.2, using a cutoff radius of 4 Å. In this manner, fragments containing an average of 37 atoms and with a maximum size of 70 atoms are obtained.

The $C_{69}H_{140}$ IR spectrum simulated with the composite ML model is depicted in Figure 19. All features typical for the IR spectrum of an alkane are present in

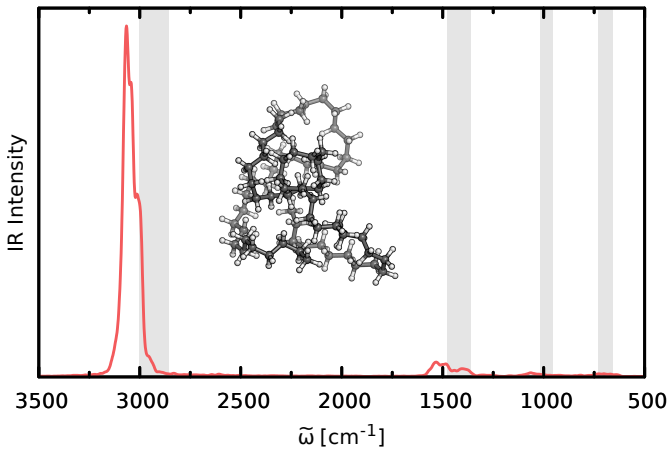


Figure 19: Gas phase IR spectrum predicted by the composite ML model for the $C_{69}H_{140}$ molecule. The regions of the expected experimental frequencies for alkanes are highlighted in gray.

the ML spectrum: The bands close to 3100 cm^{-1} correspond to the C-H stretching vibrations. The scissoring vibrations of the CH_2 groups are responsible for the peaks in the vicinity of 1500 cm^{-1}. Features indicating the C-C bond stretches are found at 1000 cm^{-1} and the peak close to 600 cm^{-1} is caused by the CH_2 rocking vibrations.

While the ML model is able to account for the general structure of the IR spectrum in a reliable manner, a blue shift compared to the expected experimental frequencies is observed for all peaks. For instance, the C-H stretching bands are shifted from a typical value of 2900 cm^{-1} to 3100 cm^{-1}in the ML spectrum. Although a direct analysis of this phenomenon is not possible due to the high computational cost of the B2PLYP method, its source can still be investigated in an indirect manner by exploiting the transferability of the composite model. Since the HDNNPs and

the dipole moment model are also valid for systems chemically similar to $C_{69}H_{140}$, the composite model can be used to simulate the IR spectrum of the much smaller n-Butane molecule for which electronic structure computations at the B2PLYP level are still feasible. The ML spectrum of n-butane, along with a static B2PLYP spectrum and an experimental spectrum recorded in the gas phase [117], is shown in Figure 20. As can be seen, the spectral bands in the B2PLYP spectrum are shifted to the same

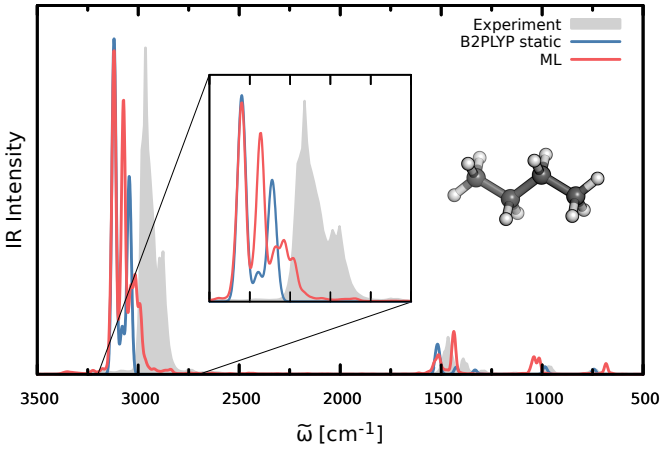


Figure 20: Comparison of the n-Butane IR spectrum simulated with the ML model (red) and the static IR spectrum calculated with the B2BLYP method (blue) within the harmonic oscillator approximation. The bands of the static spectrum are convoluted with Gaussian functions. An experimental gas phase spectrum is shown in gray.

extent as those in the ML spectrum. This finding supports the conclusion, that the observed blue shift is not an artifact introduced by the ML model, but indeed caused by the underlying electronic structure method. In general, the peak positions predicted by the ML approach show excellent agreement with the reference method. Moreover, in contrast to the static spectrum, the ML accelerated AIMD approach is able to reproduce the vibrational structure found for the experimental C-H bands to a high degree of accuracy (see insert Figure 20), highlighting the importance of including dynamic and anharmonic effects even when modeling the IR spectra of small molecules.

The above results serve not only as a demonstration for the accuracy of the composite ML approach, but also for its transferability, as well as the efficacy of HDNNP-based fragmentation. Further insights with regards to the computational efficiency of the ML models paired with fragmentation can be gained by studying the timings obtained for $C_{69}H_{140}$: The search for all relevant configurations with the sampling scheme takes approximately 7 days on a single Xeon E5-2650 v3 CPU. The computation of the energies and forces of all fragments at the B2PLYP level requires 0.9 days if one core is used for every fragment. Training of the dipole moment model, as well as the HDNNPs takes 0.3 days. The energies, forces and dipole moments of the 110 000 configurations (5 ps equilibration and 50 ps production) encountered during the ML-accelerated AIMD simulation can be predicted in 3.5 hours. Hence, the total time necessary to obtain the ML spectrum is approximately 8.3 days. Using the B2PLYP method, computing the energies and forces of a single $C_{69}H_{140}$ configuration

would take 30 days. Since a full AIMD trajectory requires 110 000 such evaluations, the total simulation time becomes 3.3 million days or 9 041 years, respectively.

4.3.3 *Protonated Alanine Tripeptide*

As a final test for the new developments presented in this thesis, the IR spectrum of the protonated alanine tripeptide (Ala_3^+, Figure 10e) is simulated. The tripeptide ML model is composed of two HDNNPs and a dipole model trained on the BLYP density functional[110–112,119] energies and forces of 717 reference points selected by the adaptive sampling scheme.

The experimental spectrum of Ala_3^+ is composed of the individual spectra of three different conformers (Figure 21):[120] First, an elongated chain protonated at the amine group of the N-terminus, henceforth referred to as the NH_3 conformer. A second conformer – termed NH_2 for short – differs in its protonation site, with the proton located at the oxygen of the N-terminal carbonyl group. The final species is a folded chain which carries its proton at the same position as the NH_3 conformer. To emulate these circumstances, the overall IR spectrum is computed by performing ML accelerated AIMD simulations for the three individual conformers and averaging the resulting spectra afterwards. The ML spectrum based on BLYP, as well as the spectra of the NH_3, NH_2 and folded species are depicted in blue in Figure 21. Since the experimental spectrum[120] (shown in gray) was recorded for the region between 2700 $cm^{-1}$and 3700 cm^{-1}, the following analysis will focus primarily on the stretching vibrations involving hydrogen atoms.

On the whole, the composite ML model provides an accurate description of the different features encountered in the experimental spectrum. The ML spectrum faithfully reproduces the peak caused by the O-H stretching vibrations of the C-terminal carboxylic acid at 3570 cm^{-1}. N-H stretching modes not involved in any hydrogen bonds to neighboring atoms are found in the spectral regions from 3300 cm^{-1} to 3500 cm^{-1}. The two free hydrogens of the N-terminal amine group in the NH_3 and folded conformers give rise to the particularly intense signal at 3420 cm^{-1}. Deviations from the experimental spectrum are observed in the region between 3250 cm^{-1} to 3350 cm^{-1}, which are underpopulated in the ML spectrum. This difference is caused by the employed BLYP method and – to a certain extent – temperature effects (see discussion in Appendix A.3). The N-H vibrations participating in hydrogen bonds are correctly shifted to the region ranging from 3100 cm^{-1} to 3300 cm^{-1}, where the ML model is able to resolve several experimental subpeaks. The regions between 2800 cm^{-1} and 3100 cm^{-1} correspond to the vibrations of the C-H groups. Here, the ML spectrum is able to differentiate the peaks due to the C_α groups of the peptide at 2930 cm^{-1} and the methyl groups at 2970 cm^{-1}.

Additional insights into the accuracy of the present ML approach are offered by a comparison of the BLYP based model to another model using the BP86 functional as electronic structure reference (the corresponding BP86 IR spectrum is shown in red in Figure 21). Although both methods are closely related and should produce similar spectra, striking deviations can be found between BLYP and BP86. Compared to the BLYP and experimental spectra, the peak corresponding to the O-H stretching modes of the C-terminal carboxylic acid group is blue shifted by almost 80 $cm^{-1}$and split into two smaller peaks. Moreover, significant differences can also be found in the regions of the N-H stretching vibrations. Whether these discrepancies are caused by the ML models or the respective electronic structure methods can be ascertained using he

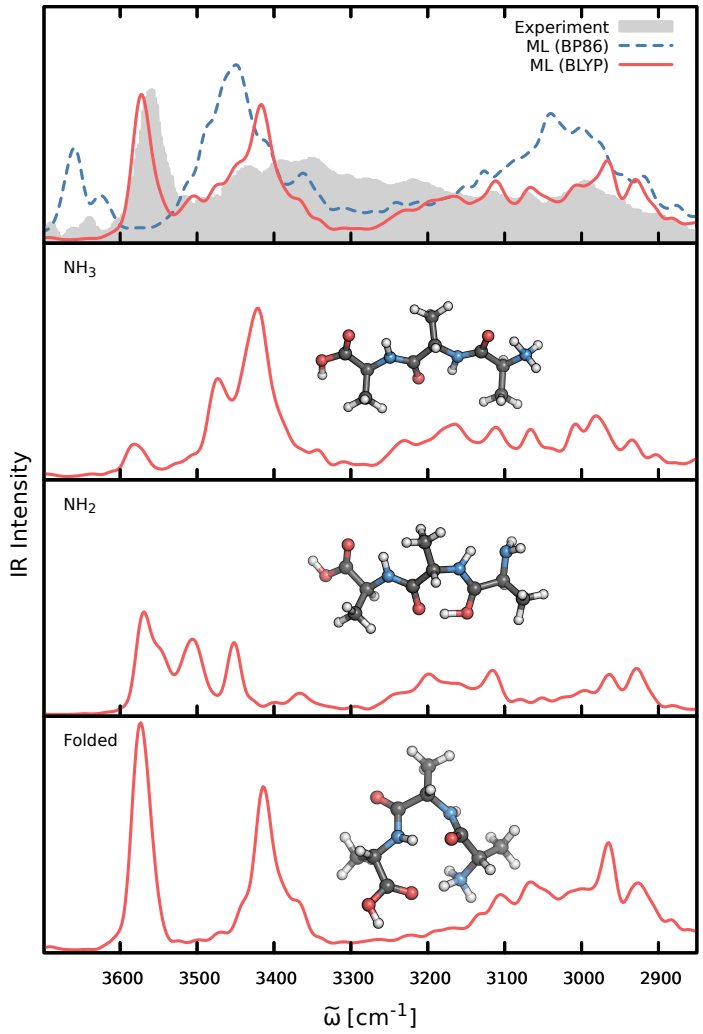


Figure 21: Gas phase IR spectra of the protonated alanine tripeptide. The spectra predicted by the ML models based on the BLYP (red) and BP86 (blue) electronic structure methods are shown in the top panel, alongside the experimental spectrum (gray). The individual BLYP based ML spectra of the different conformers are given in the subsequent panels.

vibrations of the N-terminal amine group in the NH_3 conformer as an example. The hydrogens in this moiety participate in a proton transfer reaction to the oxygen of the adjacent carbonyl group. As can be seen in Figure 22, the energy barrier computed for this reaction differs significantly between the BLYP and BP86 method, while the ML models manage to closely reproduce their respective reference. The small changes in the barrier height, which are present in the electronic structure calculations as well as the ML models, lead to different rates for the proton transfer event, which in turn gives rise to the disparate spectral features in the final spectra. Based on these findings, it is safe to assume that the deviations between the BLYP and BP86 spectra are not caused by erroneous behavior of the ML models, but instead by the intrinsic differences in the underlying electronic structure reference. Faced with this significant disagreement between two closely related methods, the faithfulness with which the ML models reproduce their respective reference methods is particularly impressive: Compared to the error between BLYP and BP86, the deviation of the ML predictions from their respective reference methods is negligible (see Figure 22), once

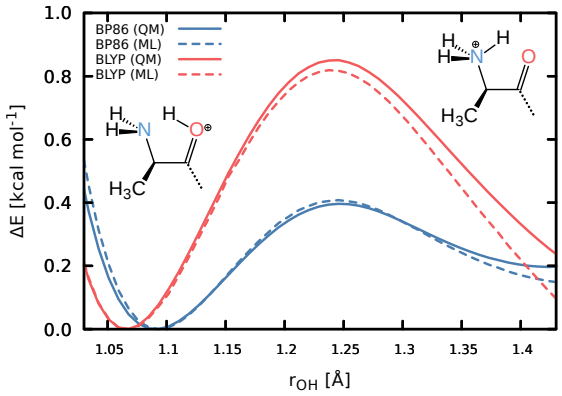


Figure 22: Energy barrier of the proton transfer from the N-terminal amine to the adjacent carbonyl group in the NH_3 conformer of the alanine tripeptide. The energy profiles calculated with the BLYP (red) and BP86 (blue) electronic structure methods are shown as solid lines. The barriers predicted by the ML models based on these methods are depicted as dashed lines. The distance r_{OH} between the carbonyl oxygen and the transfered proton is used as the reaction coordinate.

again demonstrating the potential of the ML models constructed with the strategies introduced in this thesis.

At the same time, the above example emphasizes an important feature of the adaptive sampling scheme, which is its ability to select molecular configurations important for the characterization of a chemical system. Proton transfer events play a crucial role in the experimental IR spectrum of Ala_3^+.[120] Although no information about this chemical transformation is present in the initial reference data set, the sampling scheme automatically chooses samples in such a manner, that the final ML model is able to provide an accurate description of these events (see Figure 22). The relatively small reference set size needed to obtain high quality ML spectra throughout this study serves as an indicator, that this property of the sampling scheme also holds in the general case.

5 SUMMARY AND CONCLUSION

The central objective of this thesis was the development of new strategies and protocols for machine learning (ML) methods in theoretical chemistry. ML techniques – and ML potentials in particular – offer tantalizing new possibilities for the simulation of molecular systems, combining the excellent computational efficiency of empirical force fields with the accuracy of high-level electronic structure methods. However, several issues complicate a routine application of these methods: First, standard ML algorithms need to be adapted in special ways in order to accommodate the three dimensional structure of molecules. Due to the complexity of the resulting ML architectures, new training strategies need to be developed. Second, although ML models can treat far larger molecular systems than conventional electronic structure methods, electronic structure reference computations are still required in order to train accurate models. As a consequence, the prohibitive scaling of these methods imposes an indirect limit on the range of systems which can be modeled with ML. Third, in order to obtain a ML model providing a reliable description of a chemical problem, all relevant regions of the PES need to be accounted for in the reference data set. However, no clear criteria on how to select relevant configurations exist, making it difficult to automate this process. Finally, the simulation of various molecular properties imposes further requirements on ML architectures apart from the adaptations necessary to model molecular structures, presenting an additional challenge for the construction of valid ML models. In order to overcome the above limitations, a range of new strategies is introduced in this thesis and their efficacy tested through the practical application to different chemical systems.

Among the main developments is an improved training algorithm for high dimensional neural network potentials (HDNNPs), a ML technique which is especially well suited for modeling molecular potential energy surfaces (PESs). This training algorithm – termed the element decoupled global extended Kalman filter (ED-GEKF) – makes it possible to create HDNNPs of unprecedented accuracy, reducing the overall errors of these ML models significantly, especially compared to alternative training algorithms. Moreover, the ED-GEKF also exhibits superior convergence behavior. As a consequence, highly accurate HDNNPs can now be obtained at only a fraction of the training time required by other methods. In addition to energies, molecular forces can also be included into ED-GEKF training in a consistent manner. The use of molecular forces results in a twofold advantage: 1) Fewer reference computations are required, since the additional information contained in the $3N_{\text{atoms}}$ molecular force components of every configuration can now be utilized during training and 2) the overall accuracy, and especially the accuracy of predicted forces, of the HDNNPs is improved, a property which is highly desirable in such applications as e.g. molecular dynamics simulations.

Another important innovation is the introduction of a HDNNP-based fragmentation approach capable of bypassing the system size restrictions imposed by the need for electronic structure reference data. Using this fragmentation scheme, properties of a macromolecular system (e.g. energies) can be reconstructed given only the information contained in small fragments of the original molecule. Hence, prohibitive electronic structure computations of macromolecular systems are no longer necessary, since these molecules can now be treated in a divide and conquer approach.

Moreover, due to the special structure of HDNNPs, the effective computational cost for obtaining electronic structure reference data using HDNNP fragmentation scales linearly with the size of the system. Besides a high computational efficiency, the present fragmentation approach is furthermore characterized by an excellent accuracy, easily outperforming a conventional fragmentation scheme by achieving significantly lower deviations from the electronic structure reference. This accuracy also holds for properties other than energies, such as molecular forces and dipole moments. In addition, the HDNNP-based fragmentation scheme can easily be interfaced with molecular dynamics simulations.

The two above developments are supplemented by introducing a special adaptive selection scheme. Driven by ensembles of HDNNPs, this scheme makes it possible to incrementally grow reference data sets in a highly automated fashion based on only a small initial number of electronic structure reference data points. Using this scheme, the number of reference computations required to construct HDNNPs can be reduced significantly, while at the same time maintaining a high predictive power. This is achieved through the ability of the adaptive selection scheme to automatically infer the chemistry underlying the investigated systems, selecting representative configurations in a highly reliable manner. When combined with the ability of the ED-GEKF to incorporate forces during HDNNP training, the overall size of the electronic structure reference sets can be decreased even further. As a consequence, even HDNNPs based on expensive high level electronic structure methods can be constructed with ease, thus extending the potential applications for this ML technique significantly.

As a final step towards improving the general applicability of ML methods for theoretical chemistry simulations, the architecture of HDNNPs is modified in order to provide them with the ability to model molecular dipole moments. The resulting dipole moment model is able to reproduce electronic structure dipole moments with high fidelity, making it for example suitable for the simulation of molecular infrared (IR) spectra. However, the present model does not only offer access to molecular dipole moments, but also to environment dependent atomic partial charges. These charges are determined from quantum mechanical observables in the form of the molecular multipole moments based purely on statistical principles. Hence, the dipole moment model constitutes a novel, ML based charge partitioning scheme. As such, the dipole model is expected to be only the first example of a class of HDNNP-based models, serving as partition schemes to access various molecular properties or even modeling long range interactions.

To obtain insights into the proficiency of the newly introduced strategies, they are applied to a variety of chemical problems. The performance of ED-GEKF training was analyzed using the Claisen rearrangement of allyl vinyl ether to 4-pentenal as an example. Linear all-trans alkanes served as a simple test case for a proof of concept study of the HDNNP-based fragmentation approach. A final study was dedicated to probing the efficacy of the adaptive sampling scheme and the dipole moment model, as well as exploring the overall synergies between the individual developments. To this end, the above techniques were used to simulate the molecular IR spectra of methanol, n-alkanes of different length and the protonated alanine tripeptide via ML accelerated *ab initio* molecular dynamics (AIMD). In all cases, excellent results were achieved, attesting to the effectiveness of the methods introduced in this thesis.

Moreover, these practical applications also provide a broader perspective on the potential inherent to the ML approaches used in this work: The combination of accuracy and speed of the obtained ML models offers an elegant way to overcome

the inherent limitations of electronic structure methods. This was demonstrated by using a ML model based on a high-level electronic structure method to perform a molecular dynamics simulation of the $C_{69}H_{140}$ n-alkane, a feat close to impossible with the reference method itself. Furthermore, the successful description of a Claisen rearrangement, as well as the proton transfer reactions occurring in the tripeptide, highlights the ability of the present ML models to operate in situations where conventional force fields encounter difficulties, the prime example being bond breaking and bond formation events. However, the application with perhaps the most immediate implications is the use of the introduced ML techniques to predict molecular IR spectra. Since these simulations employ ML accelerated AIMD to obtain IR spectra, they account for several effects typically neglected in alternative approaches, such as temperature effects and vibrational anharmonicities. As a result, highly accurate spectra can be obtained at only a fraction of the cost required for conventional AIMD simulations. This finding is especially relevant for the interpretation of experimental IR spectra of biomolecules (e.g. proteins), where the large system size and the high importance of anharmonic and temperature effects renders a direct analysis with electronic structure methods impractical.

The entirety of the research conducted in this thesis – encompassing the development of new strategies, as well as their practical applications – represents a large step towards the goal of establishing ML techniques as a routine tool in theoretical chemistry. However, in the same manner as concurrent ML algorithms are still far from achieving the dream of true artificial intelligence, several unsolved problems also remain in the field of theoretical chemistry, providing a rich substrate for future research. Among the most interesting issues to address is the design of ML models, which are capable of automatically learning suitable representations of molecular geometries. The process of determining an optimal set of representations currently involves a significant amount of trial and error, thus rendering it a time consuming task. An equally fascinating endeavor is the interpretation and rationalization of the predictions of ML models, since as of now little is understood about the inner workings at their at their core.

A

APPENDIX: REPRINTED PUBLICATIONS

The reprints of the different publications forming the basis of this thesis can be found in the following appendix. Section A.1 contains the paper detailing the training algorithm introduced in Section 3.1 and the associated computational studies discussed in Section 4.1. HDNNP based fragmentation (Section 3.2), as well as the associated test studies on linear all-trans alkanes (Section 4.2) are covered in the publication reprinted in Section A.2. The dipole moment model (Section 3.4) and adaptive sampling scheme (Section 3.3) are part of the submitted paper given in Section A.3, which also details the application of all techniques developed in this thesis to the simulation of molecular IR spectra (Section 4.3).

appendix a.1 high-dimensional neural network potentials for organic reactions and an improved training algorithm

Michael Gastegger and Philipp Marquetand

J. Chem. Theory Comput., **11**, 2187–2198 (2015).
http://dx.doi.org/10.1021/acs.jctc.5b00211

Contributions:

Michael Gastegger conceived and implemented the training algorithm, performed and evaluated the test simulations, and contributed to the initial draft and final version of the manuscript.

Philipp Marquetand supervised the method developments and simulations, and contributed to the final manuscript.

JCTC Journal of Chemical Theory and Computation

Article

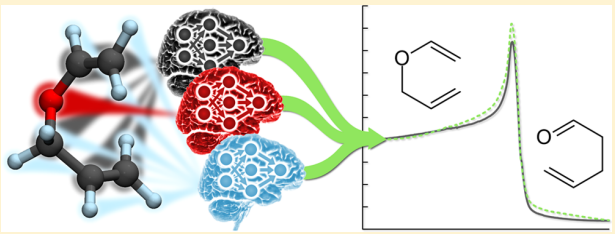
pubs.acs.org/JCTC

High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm

Michael Gastegger and Philipp Marquetand*

Institute of Theoretical Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria

ABSTRACT: Artificial neural networks (NNs) represent a relatively recent approach for the prediction of molecular potential energies, suitable for simulations of large molecules and long time scales. By using NNs to fit electronic structure data, it is possible to obtain empirical potentials of high accuracy combined with the computational efficiency of conventional force fields. However, as opposed to the latter, changing bonding patterns and unusual coordination geometries can be described due to the underlying flexible functional form of the NNs. One of the most promising approaches in this field is the high-dimensional neural network (HDNN) method, which is especially adapted to the prediction of molecular properties. While HDNNs have been mostly used to model solid state systems and surface interactions, we present here the first application of the HDNN approach to an organic reaction, the Claisen rearrangement of allyl vinyl ether to 4-pentenal. To construct the corresponding HDNN potential, a new training algorithm is introduced. This algorithm is termed "element-decoupled" global extended Kalman filter (ED-GEKF) and is based on the decoupled Kalman filter. Using a metadynamics trajectory computed with density functional theory as reference data, we show that the ED-GEKF exhibits superior performance − both in terms of accuracy and training speed − compared to other variants of the Kalman filter hitherto employed in HDNN training. In addition, the effect of including forces during ED-GEKF training on the resulting potentials was studied.



1. INTRODUCTION

Computational chemistry is a tightrope walk between accuracy and efficiency. Evidently, a sufficiently accurate potential energy surface (PES) is required in order to provide a reasonable description of a molecular system. However, when dealing with large systems or long simulation times, computational efficiency becomes an additional concern, and a compromise has to be found.

The most accurate PESs are provided by electronic structure methods, based on first-principles quantum mechanics, albeit at considerable computational expense, due to the explicit treatment of the many-electron problem. As a result, only systems of limited size and relatively short time-scales are accessible on a routine basis, when using these methods.[1,2]

Empirical force fields are several orders of magnitude faster to evaluate, since the PES is described as a sum of physically motivated analytic functions fitted to experimental or computational reference data. However, as a consequence of the fitting procedure involved, empirical potentials are only accurate for limited regions of the PES. Another drawback is the inability to describe bond breaking and bond formation events, as well as unusual coordination geometries and bonding situations, due to predefined atom types and the form of the analytic functions employed.[3,4] While so-called reactive force fields are capable of addressing this kind of problems, the accuracy of these approaches is ultimately still limited by the underlying physical approximations.[5]

A promising alternative is the use of machine learning (ML) techniques to construct PESs from electronic structure data. The highly flexible functional form of ML potentials allows for very accurate interpolation even of complicated PESs, as well as the description of complex and changing bonding patterns. Moreover, paired with a computational efficiency on par with empirical potentials, ML potentials can offer an accuracy comparable to high level electronic structure methods at the computational cost of force fields.[6]

Up to date, various ML approaches have been successfully applied to either the construction of PESs or the description of contributing terms, with Polynomial Fitting procedures,[7] Gaussian Processes,[8,9] Modified Shepard Interpolation,[10−13] Interpolating Moving Least Squares,[14,15] and Support Vector Machines[16] only being some of the most prominent examples. Particularly promising in this field are Neural Networks (NNs), a machine-learning technique inspired by the central nervous system.

The number of neural network potential energy surfaces (NN-PESs) has steadily increased over the course of the past decade, with applications ranging from surface science[17−25] to molecular systems.[26−45] For an exhaustive list of NN-PESs, see reviews.[46−48] These applications were accompanied by several advances to NN training algorithms and architectures, such as the inclusion of gradients in the training procedure.[42,49] Another example is the use of symmetrized NNs[43] or symmetry functions[18,50,51] to account for permutational invariance in the

Received: March 4, 2015
Published: April 16, 2015

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

input data and high dimensional representation schemes,[52,53] which employ multiple NNs for the construction of the PES.

An important step toward the routine use of NN-PESs in computational chemistry is the introduction of the high-dimensional neural network (HDNN) scheme by Behler and Parinello.[52] In the HDNN scheme, the PES is constructed from a sum of individual atomic energy contributions. These contributions depend on the atomic environment and are computed by a set of NNs, where a single NN is used for atoms of the same element. In order to describe the environment of the different atoms, a special set of atom centered symmetry functions (ACSFs) is employed.[51] The combination of these features allows HDNN potentials to overcome several limitations of standard NNs, the most important ones being limited system size, transferability of parameters, fixed amount of atoms and elements, and dependence of the NN-PES on translations and rotations in input space. The utility of HDNN potentials is demonstrated by a wide range of applications to solid state systems,[54−59] surface interactions,[21,22,60] and water clusters.[61−63]

Our work is based on these foundations and goes beyond in two ways: First, a novel training algorithm for HDNNs is presented. This "element-decoupled" Kalman filter is based on the decoupled formulation of the standard global extended Kalman filter (GEKF)[64] and exhibits superior performance in terms of training speed and accuracy of the final PES fits compared to other variants of the Kalman filter commonly used for HDNN training. Second, this algorithm is applied to the construction of HDNN-PESs for the Claisen rearrangement of allyl vinyl ether to 4-pentenal, which presents, to the best knowledge of the authors, the first application of the HDNN scheme to an organic reaction involving bond breaking and formation. Based on the resulting HDNN-PESs, the overall performance of the element-decoupled filter and its interpolation capabilities as well as the effect of the inclusion of forces during the training process are studied.

2. THEORETICAL BACKGROUND

This section is divided into two parts: The first part provides an outline of the NN architecture employed, with special focus on Behler−Parinello HDNNs. In the second part, the standard Kalman filter training algorithm will be reviewed, and adaptations thereof with regard to HDNNs will be discussed.

2.1. Neural Network Architecture. Similar to biological nervous systems, NNs consist of an arrangement of simple subunits, so-called neurons, often also referred to as nodes. These artificial neurons collect, process, and transmit incoming signals according to a predefined connection pattern, making NNs highly parallel, flexible, and robust nonlinear universal approximators capable of accurately fitting any continuous real function.[65−67] Due to these properties, NNs are typically employed in classification tasks, pattern recognition, and for the approximation of complicated functional relations, e.g. PESs.

One of the most frequently used NN architectures is the feed-forward multilayer perceptron.[68] An example for a NN of this type with one hidden layer and a single neuron in the output layer is depicted in Figure 1. In a feed-forward NN, signals are only transmitted into one direction. Starting with the vector of inputs $\mathbf{y}^0$ in the input layer, the output of a node j in layer l can be computed according to the regression

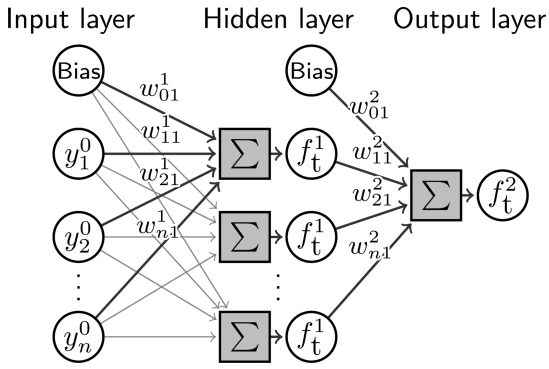


Figure 1. Feed-forward neural network with a single output node and one hidden layer. Bias nodes provide a constant offset to the transfer functions f_t^l. Nodes in adjacent layers are connected by the weights $\{w_{ij}^l\}$, the fitting parameters of the neural network.

$$y_j^l = f_t^l\left(w_{0j}^l + \sum_i^{n^{l-1}} w_{ij}^l y_i^{l-1}\right) \tag{1}$$

until the output layer is reached. The signals of the previous layer $\{y_i^{l-1}\}$ are scaled by a set of weight parameters $\{w_{ij}^l\}$, where w_{ij}^l is the weight connecting node i of the previous layer to node j in the current layer and w_{0j}^l is a bias weight, which provides an adjustable offset to the transfer function, hence adding additional flexibility. These weights are the adjustable parameters of the NN and are collected in the vector of weights $\mathbf{w}$, which has to be determined in the training process. After weighting and summation, a transfer function f_t^l is applied to the nodes in layer l. Typically, sigmoidal functions (e.g., hyperbolic tangents) are employed in the hidden layers, as they provide the NN with the ability to model any continuous, real function to arbitrary accuracy with a finite number of nodes in the hidden layer.[65−67] In the output layer, usually linear transfer functions are used. The output of the NN can be described as a function $f_{NN}(\mathbf{y}^0,\mathbf{w})$ depending on the input parameters and the vector of all weights.

Due to their robustness, computational efficiency, simplicity of training, and the availability of analytic gradients, feed-forward NNs are the type of NNs most commonly used in the construction of NN-PESs. However, several problems limit their applicability when employed in the interpolation of multidimensional PESs. Due to the predetermined structure of the NN after training, the output of the NN depends on the ordering of the inputs, in the case of a NN-PES the molecular coordinates. In a similar manner, the number of atoms that can be treated with such simple NNs must be proportional to the number of input nodes, making it impossible to treat molecules of different size with the same NN. In addition, the NN shows an unwanted variance with respect to translation and rotation of the molecular geometry if Cartesian coordinates are used.

One strategy to overcome these limitations is the HDNN scheme introduced by Behler and Parinello in 2007.[52] In this approach, the total energy E of a molecular system with N atoms is constructed from atomic energy contributions E_i according to

$$E = \sum_i^N E_i \tag{2}$$

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

Journal of Chemical Theory and Computation Article

The atomic energies depend on the chemical environment of atom i and are computed by individual NNs. Figure 2 shows an example for a Behler−Parinello HDNN.

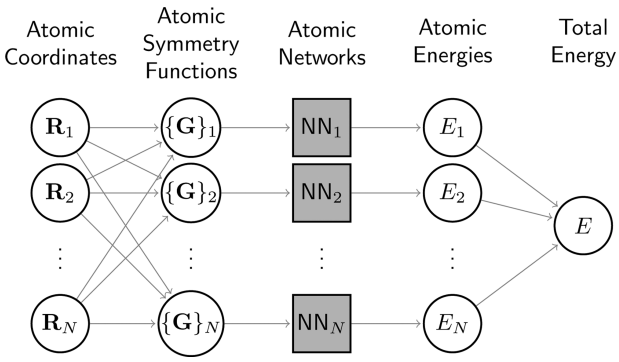


Figure 2. Scheme of a high-dimensional neural network of the Behler−Parinello type. The Cartesian coordinates $\{\mathbf{R}_i\}$ of a molecule are transformed into a set of atomic symmetry functions $\{\mathbf{G}_i\}$ describing an atom subject to its chemical environment. With these symmetry functions as inputs, the atomic contributions E_i to the total energy E are computed with atomic neural networks.

Since the total energy is now expressed as a sum of NN contributions E_i, any dependence on the ordering of inputs is eliminated. Moreover, by using one NN for nuclei of the same element, the HDNN can model molecular systems of arbitrary size. For every new atom, the sum in eq 2 is simply augmented by the energy contribution of the corresponding elemental NN. To ensure invariance with respect to translation and rotation of the molecule, the Cartesian coordinates $\{\mathbf{R}_i\}$ are transformed to a set of many-body symmetry functions $\{\mathbf{G}_i\}$, depending on the internuclear distances and angles between all atoms. These ACSFs describe the local chemical environment of atom i via radial and angular distributions of the surrounding nuclei.[51] An example for a radial ACSF is

$$G_i^{\mathrm{rad}} = \sum_{j\neq i}^{N_{\mathrm{atoms}}} e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij}) \tag{3}$$

while an angular distribution can be characterized by

$$G_i^{\mathrm{ang}} = 2^{1-\zeta} \sum_{j,k\neq i}^{N_{\mathrm{atoms}}} \left(1 + \lambda \frac{\mathbf{R}_{ij}\mathbf{R}_{ik}}{R_{ij}R_{ik}}\right)^{\zeta} e^{-\eta(R_{ij}^2+R_{ik}^2+R_{jk}^2)} \times f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \tag{4}$$

Here, R_{ij} is the distance between atoms i and j, $\mathbf{R}_{ij}$ is the vector $\mathbf{R}_j - \mathbf{R}_i$, and R_s, η, ζ, and λ are parameters which determine the shape of the ACSFs. Typically, a combination of radial and angular ACSFs $\{\mathbf{G}_i\}$, differing in the parameters R_s, η, ζ, and λ, is employed in the description of every atom. In addition, a radial cutoff R_c is introduced around the central atom in the form of a cutoff function

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right], & R_{ij} \leqslant R_c \\ 0, & R_{ij} > R_c \end{cases} \tag{5}$$

This reduces the local chemical environment to the energetically relevant regions, leading to a linear scaling of the HDNN computation cost with system size.

Due to the well-defined functional form of the individual elemental NNs, an analytic expression for the HDNN forces can be derived. The force with respect to the Cartesian coordinates $\mathbf{R}_j$ of atom j is given by the relation

$$\mathbf{F}^{(j)} = -\sum_{i}^{N} \sum_{\alpha}^{M_i} \frac{\partial E_i}{\partial G_{i,\alpha}} \frac{\partial G_{i,\alpha}}{\partial \mathbf{R}_j} \tag{6}$$

where M_i is the number of ACSFs used to describe nucleus i.[51] The first term in eq 6 is the derivative of the elemental NNs with respect to the ACSFs, and the second term is the derivative of the ACSFs with respect to the Cartesian coordinates of atom j.

2.2. HDNN Training. A NN is defined by its structure and its weights $\mathbf{w}$. Once the structure − the number of hidden layers and nodes in the hidden layers − has been determined empirically, the initially random NN weights need to be optimized in order to reproduce the desired function, a process also called "training" of the NN.

Several different NN training algorithms exist, from gradient-based approaches, such as the so-called backpropagation algorithm,[69] to second-order methods, e.g. the Levenberg−Marquardt optimization.[70] One second-order training method that is particularly promising for the construction of NN-PESs is the GEKF.[71,72] In the GEKF, every sample (e.g., molecular geometry) is presented to the algorithm sequentially, immediately followed by an update of the weight vector $\mathbf{w}$. The correction of $\mathbf{w}$ is based on the current error and a weighted history of previous update steps. In the case of PES fitting, the error ν_k of the current update step k, also referred to as filter innovation, is

$$\nu_k = E_k - \tilde{E}_k \tag{7}$$

with E_k being the reference energy (from a quantum-chemical calculation), and $\tilde{E}_k$ being the corresponding NN-PES value. The direction in which $\mathbf{w}$ is changed in each update step is based on the current Jacobian matrix $\mathbf{J}_k$ − the NN gradient with respect to the weights − while the magnitude of the update is given by the filter covariance matrix $\mathbf{P}_k$. In short, $\mathbf{P}_k$ is the weighted history of Gauss−Newton approximations to the inverse Hessian of the modeling error surface and the main reason for the excellent training properties of the Kalman filter. For a detailed explanation of the GEKF for standard NNs as well as the associated recursion scheme and derivation thereof, we refer to refs 49, 71, and 72.

For a HDNN of the Behler-Parinello type, $\tilde{E}_k$ in eq 7 becomes the HDNN total energy given in eq 2. Since an individual NN is now used for every element with atomic number Z, the different elemental weight vectors $\mathbf{w}_k^{(Z)}$ need to be optimized simultaneously with the help of their respective error covariances $\mathbf{P}_k^{(Z)}$. One possible approach to HDNN training is to use an "atomic" Kalman filter (A-GEKF), where an update of the corresponding $\mathbf{w}_k^{(Z)}$ is performed for every atom i in the molecular sample of step k according to the regression scheme:

$$\mathbf{w}_k^{(Z)} = \mathbf{w}_{k-1}^{(Z)} + \mathbf{K}_k^{(i)} N_{\mathrm{atom}}^{-1} \nu_k \tag{8}$$

$$\mathbf{K}_k^{(i)} = \mathbf{P}_{k-1}^{(Z)} \mathbf{J}_k^{(i)} \left[\lambda_k^{(Z)} \mathbf{I} + \left(\mathbf{J}_k^{(i)}\right)^T \mathbf{P}_{k-1}^{(Z)} \mathbf{J}_k^{(i)}\right]^{-1} \tag{9}$$

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

$$\mathbf{P}_k^{(Z)} = (\lambda_k^{(Z)})^{-1}\left[\mathbf{I} - \mathbf{K}_k^{(i)}\left(\mathbf{J}_k^{(i)}\right)^T\right]\mathbf{P}_{k-1}^{(Z)} \tag{10}$$

Here, $\mathbf{K}_k^{(i)}$ is the Kalman gain matrix computed for atom i, $\mathbf{I}$ is the identity matrix, and the expression in brackets on the right-hand side of eq 9 is a scaling matrix. The λ_k are time-varying forgetting factors computed according to $\lambda_k = \lambda_{k-1}\lambda_0 + 1 - \lambda_0$, which are introduced in order to reduce the risk of convergence to local minima early in the training process.[71] Both λ_k and λ_0 are typically initialized close to unity. For the update of the weights (eq 8), the total error is averaged over the number of atoms N_{atom}, because the individual atomic contribution to the error is unknown. This assumption is problematic, especially in combination with atomic weight updates, since the weights of some elements are updated more frequently, leading to a smaller relative error in the weights when compared to the ones of less abundant elements. This uneven distribution of the errors imposes a severe limitation to the quality of the PES fit obtained.

An alternative is to model the HDNN scheme as one large composite NN. Using the sum rule, it is possible to show that the Jacobian $\mathbf{J}_k^{(Z)}$ associated with every element can be computed as

$$\mathbf{J}_k^{(Z)} = \sum_i^{N_{\text{atom}}} \mathbf{J}_k^{(i)}\delta_{Z_i,Z} \tag{11}$$

where $\delta_{Z_i,Z}$ is one if the element of the current atom Z_i corresponds to the elemental index Z of the Jacobian and zero otherwise. Instead of individual updates for all atoms, one update per element is now performed for every molecule, according to

$$\mathbf{w}_k^{(Z)} = \mathbf{w}_{k-1}^{(Z)} + \mathbf{K}_k^{(Z)}N_{\text{elem}}^{-1}\nu_k \tag{12}$$

$$\mathbf{K}_k^{(Z)} = \mathbf{P}_{k-1}^{(Z)}\mathbf{J}_k^{(Z)}\left[\lambda_k\mathbf{I} + \left(\mathbf{J}_k^{(Z)}\right)^T\mathbf{P}_{k-1}^{(Z)}\mathbf{J}_k^{(Z)}\right]^{-1} \tag{13}$$

$$\mathbf{P}_k^{(Z)} = \lambda_k^{-1}\left[\mathbf{I} - \mathbf{K}_k^{(Z)}\left(\mathbf{J}_k^{(Z)}\right)^T\right]\mathbf{P}_{k-1}^{(Z)} \tag{14}$$

In this "elemental" GEKF (E-GEKF), the bias due to the frequency of elements is eliminated, as every element is only updated once. However, it is still assumed that the elemental NNs contribute to the total error in a similar fashion. Since no clear guidelines on how to treat the total error in eq 12 exist, ad hoc corrections have to be introduced e.g. by weighting ν_k with the elemental fraction or, as is the case here, by dividing by the number of elements present in the molecule.

Compared to gradient-based NN training methods such as backpropagation, second-order algorithms based on the GEKF offer superior accuracy and convergence behavior, albeit at an increased computational cost. For every sample in the training data, several matrix operations including inversion have to be performed, and several passes over the training data, so-called "epochs", are required in order to obtain a suitably trained NN. However, the computational effort associated with HDNN training is still negligible compared to the generation of the electronic structure reference data, which currently represents the bottleneck in NN-PES construction. Furthermore, the efficiency of the GEKF can be increased by employing an adaptive threshold to the filter updates.[71] In this approach, the error of the current sample is compared to a threshold defined as a fixed fraction of the total root-mean-square error (RMSE) of the previous epoch. An update of the weights is only performed, if the error exceeds the threshold. In this way, the number of unproductive updates can be kept to a minimum, leading to an efficiency almost on par with standard backpropagation, while retaining all advantages of the GEKF.

3. A NEW TRAINING ALGORITHM

In the following, we present a new GEKF variant for HDNN training, which requires neither partitioning of the total error, nor any assumptions regarding its distribution among the subnets. This approach exhibits superior fitting performance and filter convergence behavior compared to the atomic and elemental GEKF algorithms. This improved filter algorithm is based on a variant of the GEKF initially proposed to reduce its computational effort, the so-called "decoupled" Kalman filter.[64,73]

3.1. Element-Decoupled Kalman Filter. In the decoupled Kalman filter, sets of weights are treated as independent from each other in order to reduce the dimensionality of the matrix operations. Since it can be safely assumed that the weight vectors $\mathbf{w}_k^{(Z)}$ of the different elemental NNs in a HDNN fulfill this criterion, a decoupled Kalman scheme can be applied to HDNNs in the form of an "element-decoupled" GEKF (ED-GEKF). In this scheme, the computation of the Kalman gain in eq 13 is modified to

$$\mathbf{K}_k^{(Z)} = \mathbf{P}_{k-1}^{(Z)}\mathbf{J}_k^{(Z)}\left[\lambda_k\mathbf{I} + \sum_z^{N_{\text{elem}}}\left(\mathbf{J}_k^{(z)}\right)^T\mathbf{P}_{k-1}^{(z)}\mathbf{J}_k^{(z)}\right]^{-1} \tag{15}$$

where the expression in the brackets now depends on the subnets of all elements (indicated by z). In this way, the total error can be used in the weight update, without introducing additional assumptions regarding the distribution of individual atomic or elemental errors, making the "element-decoupled" GEKF an excellent training algorithm for the HDNN network architecture. Hence, eq 12 is modified to

$$\mathbf{w}_k^{(Z)} = \mathbf{w}_{k-1}^{(Z)} + \mathbf{K}_k^{(Z)}\nu_k \tag{16}$$

Moreover, it is also possible to extend the element-decoupled GEKF to include the forces in the training process. The force innovation associated with atom i is given by

$$\boldsymbol{\xi}_k^{(i)} = \mathbf{F}_k^{(i)} - \tilde{\mathbf{F}}_k^{(i)} \tag{17}$$

where $\mathbf{F}_k^{(i)}$ is the atomic reference force as computed by e.g. electronic structure methods, and $\tilde{\mathbf{F}}_k^{(i)}$ is the force obtained for the HDNN via eq 6. To incorporate forces in the element-decoupled scheme, only the weight update (eq 16) has to be changed to

$$\mathbf{w}_k^{(Z)} = \mathbf{w}_{k-1}^{(Z)} + \mathbf{K}_k^{(Z)}\nu_k + \mathbf{P}_{k-1}^{(Z)}\sum_i^{N_{\text{atom}}}\frac{\partial\tilde{\mathbf{F}}_k^{(i)}}{\partial\mathbf{w}_k^{(Z)}}\mathbf{B}_k^{(i)}\boldsymbol{\xi}_k^{(i)} \tag{18}$$

where $(\partial\tilde{\mathbf{F}}_k^{(i)})/(\partial\mathbf{w}_k^{(Z)})$ is the derivative of the HDNN forces with respect to the weights. $\mathbf{B}_k^{(i)}$ is a scaling matrix computed according to

$$\mathbf{B}_k^{(i)} = \left[\lambda_k\mathbf{I} + \sum_z^{N_{\text{elem}}}\left(\frac{\partial\tilde{\mathbf{F}}_k^{(i)}}{\partial\mathbf{w}_k^{(z)}}\right)^T\mathbf{P}_{k-1}^{(z)}\frac{\partial\tilde{\mathbf{F}}_k^{(i)}}{\partial\mathbf{w}_k^{(z)}}\right]^{-1} \tag{19}$$

This scaling matrix is required, since the derivation of the forces with respect to the weights produces a component vector for

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

every elemental NN and the contribution of each subnet to the overall atomic force error is unknown.

The ED-GEKF algorithm is straightforward to implement, and the associated computational performance is similar to or even better than the other filter variants. Compared to both the A-GEKF and E-GEKF, the presented algorithm exhibits a significant increase in the quality of the obtained PES fits, as well as improved convergence behavior (see Section 5). These properties open up new possibilities in the construction of HDNN potentials for a multitude of compound classes and reaction types. The ED-GEKF is expected to perform especially well for systems with several different elements (e.g., organic molecules, proteins), since the distribution of errors between the elemental subnets is inherently modeled by the algorithm. In addition, forces can be included in the training process in a consistent manner, albeit at an increased computational cost, due to the more expensive computation of the derivatives of the HDNN forces with respect to the weights compared to the standard Jacobian.

4. COMPUTATIONAL DETAILS

In order to apply the new training algorithm in the simulation of an organic reaction, reference electronic structure calculations were carried out with ORCA[74] at the BP86/def2-SVP[75−80] level of theory employing the resolution of identity approximation.[81,82] The obtained energies and gradients were augmented by the empirical D3 dispersion correction of Grimme[83] with the Becke−Johnson damping scheme.[84] A minimum energy reaction path of 500 intermediate geometries describing the transition between substrate and product was generated and optimized at the same level of theory with the WOELFLING chain-of-states method[85] implemented in TURBOMOLE.[86,87] Energies and gradients of these geometries were then recomputed with ORCA in order to ensure consistency of the reference data.

Born−Oppenheimer dynamics[88] and metadynamics[89] were carried out using electronic-structure gradients. Newton's equations of motion for the nuclei were integrated using the Velocity-Verlet algorithm[90] with timesteps of 0.5 fs. A Berendsen thermostat was employed in all simulations.[91] Initial velocities were sampled from a Maxwell−Boltzmann distribution at the corresponding bath temperature.[92] Metadynamics simulations were performed according to the scheme of Laio and Parinello,[89] utilizing the interatomic distances r_{O-C_5} and $r_{C_2-C_3}$ (see Figure 3) as collective variables to describe the Claisen rearrangement reaction.

Behler-type ACSFs were used to represent the chemical environment of the individual atoms in the HDNN-scheme.[51] A set of 10 radial and 32 angular distribution functions was employed for hydrogen and carbon atoms, while 7 radial and 20 angular functions were used for the single oxygen atom. The different number of ACSFs for oxygen and the other elements is due to combinatorial reasons. Since there is only one oxygen atom present in the molecule, it can only have C and H atoms as neighbors, while a carbon or hydrogen atom can have the neighbors C, H, and O. For all ACSFs a radial cutoff of 10.0 Å was used.

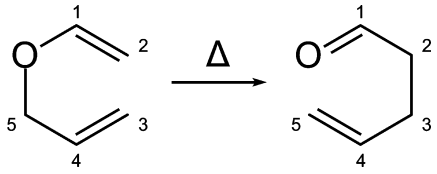


Figure 3. Thermal, aliphatic Claisen rearrangement reaction of allyl vinyl ether to 4-pentenal. Over the course of the reaction, the $O-C_5$ bond is broken, and a new bond is formed between C_2 and C_3.

Elemental subnets of different architectures were employed in the construction of the HDNN-PESs. The number of nodes in the input layer was constrained to the number of symmetry functions used in the description of the atomic environment (42 for hydrogen and carbon and 27 for oxygen), and a single node was used in the output layer. Hyperbolic tangent activation functions were used for the hidden layers, while a linear transformation was applied to the output layer of all networks. Since the elemental NN architectures can differ only in their respective hidden layers, a shorthand notation will be introduced, where e.g. C-40-40 refers to a subnet for the element carbon with two hidden layers of 40 nodes each. For the system at hand, elemental subnets of different size were tested, and it was found that two hidden layers of 40 nodes for every element offer the best compromise between accuracy and computational efficiency.

The weights of all NNs were initialized at random according to the scheme of Nguyen and Widrow.[93] In order to facilitate HDNN training, all symmetry functions derived from the reference data set and the corresponding reference energies were normalized to obtain an average of zero and a standard deviation of $\sigma = 1.0$ over the whole set. Overfitting of the reference data was detected by means of an early stopping procedure based on cross-validation.[94] In this approach, the reference data set is divided into a training and validation set. Only points in the training set are used in the construction of the HDNN potential, while the RMSEs of both data sets are monitored during training. At the onset of overfitting, the RMSE of the training set continues to decrease, while the RMSE of the validation set begins to increase. At this point, the fitting procedure is stopped, and the set of weights associated with the lowest validation RMSE is returned. In this work, the reference data was split into a training and validation set with a ratio of 9:1 in a random manner, and the training procedure was terminated after three successive training epochs with an increase in the validation RMSE. A ratio of 9:1 was chosen as a compromise between a sufficiently dense spacing of data points in the training set and a large enough validation set to reliably detect overfitting. The elemental covariance matrices required by the different Kalman filter variants were initialized as diagonal matrices $\mathbf{P}^{(Z)} = \delta^{-1}\mathbf{I}$, with a scaling factor $\delta = 0.01$. Values of $l_k = 0.99$ and $l_0 = 0.996$ were used for the time-varying forgetting schedule. The adaptive Kalman filter threshold was set to 0.9 times the RMSE of the training set in the previous epoch.[71]

Training of the HDNNs, as well as dynamics and metadynamics simulations, were carried out with a suite of programs developed in PYTHON[95] using the NUMPY package.[96]

5. RESULTS AND DISCUSSION

5.1. Model Reaction and Reference Data Generation.

To study the performance of the different Kalman filter variants and the effect of including forces into the training procedure, the aliphatic Claisen rearrangement of allyl vinyl ether was chosen as a model reaction (Figure 3). This reaction is a [3,3]-sigmatropic rearrangement, where the substrate allyl vinyl ether is converted thermally to the aldehyde 4-pentenal via bond

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

breaking and bond formation events. The transformation is irreversible and occurs at elevated temperatures. For an extensive review of this kind of rearrangement reaction, see ref 97 and references therein.

Several traits make this particular reaction an excellent subject of study for the construction of HDNN-PESs: The encountered molecules are comparatively small, rendering the generation of electronic structure reference data inexpensive. More than two different elements are present, leading to a multitude of chemical environments and allowing to test the applicability of the HDNN scheme to organic systems. Most importantly, different and changing bonding patterns have to be described in order to accurately model the rearrangement, which is one of the major advantages of ML potentials compared to standard empirical potentials.

The Claisen rearrangement reference data set for HDNN training was generated in a metadynamics run at 400 K using electronic structure energies and gradients. After an initial period of equilibration, Gaussians with a height of 6.28 kcal mol^{-1} and widths of 0.27 Å were deposited along the dimensions of the collective variables every 100 steps. The system was propagated for a total of 8.55 ps simulation time. Geometries, energies, and forces were collected every step, resulting in a total of 17100 molecules in the final reference data set.

It should be stressed at this point that the main focus of the present study is the accurate reproduction of a quantum chemically derived PES and not the accuracy of the underlying reference PES. Concerns regarding e.g. the character of the encountered species, quality of reaction barriers, and viability of the employed electronic structure method are therefore only of secondary importance. The BP86 functional was chosen in favor over more sophisticated methods primarily due to its overall computational cheapness and robustness, which further facilitated the metadynamics sampling procedure. All HDNN training methods presented in this work can be applied to any electronic structure method without additional adaptations.

5.2. ED-GEKF Performance. The performance of the different Kalman filter variants was studied based on HDNN-PESs constructed from HDNNs trained with the A-GEKF, E-GEKF, and the newly developed ED-GEKF algorithm, respectively. The employed HDNNs consisted of C-40-40, H-40-40, and O-40-40 subnets and were trained for 100 epochs. In order to account for the dependence of the final PES fit on the initial random partitioning of the data into training and validation sets, as well as the randomly initialized network weights, five HDNNs were trained with each filter, and the averaged results are reported here.

The RMSEs per molecule of the obtained HDNN-PES fits relative to the reference data set are shown in Figure 4. A clear improvement in the quality of the PESs is observed when going from the A-GEKF (0.99 kcal mol^{-1} RMSE for the training set and 1.16 kcal mol^{-1} for the validation set) over the E-GEKF (0.19 kcal mol^{-1} and 0.27 kcal mol^{-1}) to the ED-GEKF (0.08 kcal mol^{-1} and 0.13 kcal mol^{-1}). Especially the gain in accuracy obtained with the elemental and element-decoupled filter variants compared to the A-GEKF is pronounced. The reason for this behavior is most likely the elemental weight update in the A-GEKF algorithm, which is performed for each atom individually. This procedure introduces a bias in favor of more abundant elements, which – in combination with the assumption of an even distribution of the total error between

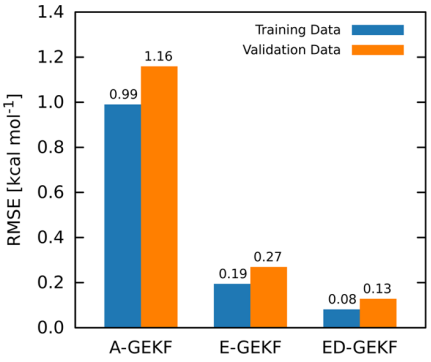


Figure 4. RMSEs (kcal mol^{-1}) for the potential energies predicted by HDNNs trained with different Kalman filter variants.

the individual subnets – ultimately limits the quality of PESs obtainable with this particular filter variant (see Section 2.2).

While the E-GEKF already shows superior performance compared to the A-GEKF, the HDNNs trained with the ED-GEKF reproduce the reference data even more accurately. Using the latter training algorithm, the RMSEs over the training and validation set are reduced to less than half of their corresponding E-GEKF values. Since the update frequency bias present in the A-GEKF is eliminated in both E-GEKF and ED-GEKF, this additional increase in fit quality can be solely attributed to the improved treatment of the error distribution between the subnets in the ED-GEKF. In contrast, no clear approach exists on how to distribute the total error between the respective subnets in the E-GEKF weight update (see eq 12), and one of the various imaginable empirical weighting schemes needs to be introduced. Several alternative schemes were investigated prior to this work, and the best results for the system at hand were achieved by averaging the total error evenly over the number of elements present in the current molecule N_{elem}. All data for the E-GEKF reported here was obtained with this weighting scheme. However, an unfortunate partitioning of the total error can easily lead to inferior fit accuracies on par with or even worse than those yielded by the A-GEKF algorithm, a problem not encountered with the ED-GEKF variant.

The aforementioned trends are even more pronounced, when the maximum deviations of the HDNN-PESs from the reference energies are compared (Figure 5). As above, changing from the A-GEKF to the other filters leads to a substantial improvement in the quality of the fit. The best accuracy is once

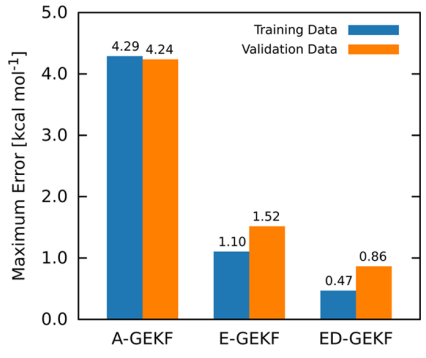


Figure 5. Maximum errors (kcal mol^{-1}) for the potential energies predicted by HDNNs trained with different Kalman filter variants.

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187–2198

again obtained with the ED-GEKF, exhibiting maximum errors smaller than 1.0 kcal mol^{-1} over the training as well as the validation data set. These results are an excellent demonstration for the high accuracy of HDNN-PESs produced by the ED-GEKF training algorithm.

With regards to computational efficiency, the A-GEKF algorithm performs worst, since the computation of the Kalman gain matrix, the weight update and the update of the covariance matrix are carried out for every individual atom present in the molecule. In contrast, these computations are only performed once per element in the E-GEKF and ED-GEKF variants, leading to a much better scaling behavior, especially for larger molecules. The major difference between the E-GEKF and ED-GEKF is that the scaling matrix (expression in the square brackets in eqs 13 and 15, respectively) has to be computed only once per molecular system in the ED-GEKF algorithm, while in the E-GEKF it is evaluated for every element. In order to calculate this expression, $n_{\mathrm{out}}^{(Z)} \times n_{\mathrm{out}}^{(Z)}$ matrices have to be inverted, where $n_{\mathrm{out}}^{(Z)}$ is the number of output nodes of the elemental subnets. Since a matrix inversion usually is an expensive operation in terms of computation time, the ED-GEKF is expected to perform even better than the E-GEKF in theory. In the special case of HDNNs, the elemental subnets possess only one output node (the atomic energy), and the matrix to be inverted is therefore a 1×1 matrix. Hence, the inversion operation reduces to a standard division, and virtually no difference in the computational efficiency between the E-GEKF and ED-GEKF variants is observed in praxis. In the implementation used in this work, both the ED-GEKF and E-GEKF algorithms are approximately 2.4 times faster than the A-GEKF for the system at hand from a purely computational perspective. It should be noted, however, that compared to the E-GEKF and A-GEKF variants, fewer iterations are needed for ED-GEKF to obtain accurate HDNN-PESs, making it the faster algorithm overall, as shown in the next section.

5.3. Filter Convergence. In addition to the quality of the fit, the evolution of the RMSEs over the course of the training procedure was investigated. The corresponding results for the training set are depicted in Figure 6. As before, the averages over five HDNNs for each filter variant are reported.

The atomic Kalman filter shows signs of convergence only at the end of the 100 training epochs, and even longer training periods would be required to fully converge the results. The E-GEKF can be considered converged after approximately 60 epochs, as only minor changes in the training RMSEs (smaller than 0.01 kcal mol^{-1}) are observed afterward. In case of the ED-GEKF algorithm, convergence is already reached after 40 epochs of training, and only small corrections to the network weights are performed after this point, once again demonstrating its excellent viability for HDNN training. Moreover, if training speed is of primary concern, sufficiently accurate HDNN-PESs with an average training RMSE of 0.12 kcal mol^{-1} can already be obtained after 20 training epochs. Even in this case, the accuracy of the ED-GEKF fit still exceeds the levels achieved by the A-GEKF and the E-GEKF variants after the full 100 epochs of training (0.99 kcal mol^{-1} and 0.19 kcal mol^{-1} training RMSE, respectively). Curves for the validation RMSEs, as well as the maximum deviations, are not shown here as they paint an identical picture.

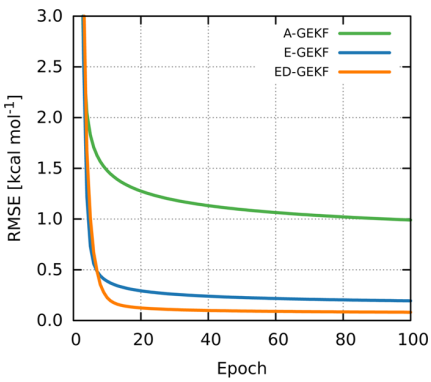


Figure 6. Evolution of the training set RMSEs (kcal mol^{-1}) during the training process for different GEKF variants.

5.4. Training including Forces. All HDNN-PESs presented in this work, up to this point, were created employing only energy data in the fitting procedure. It is, however, also possible to extend the ED-GEKF algorithm to perform a simultaneous fit of energies and forces (see Section 3.1). In order to study the effects of the inclusion of forces in the training procedure, a HDNN of the same dimensions as before (C-40-40, H-40-40, O-40-40) was trained with the force variant of the ED-GEKF (ED-GEKF+F). Due to the increased computational expense compared to the standard algorithm, only a single HDNN fit was performed. The required training time was further reduced by exploiting the fast convergence of the ED-GEKF and limiting the training to 40 epochs. The RMSEs over energies and forces obtained with the ED-GEKF+F and the other filter variants are given in Table 1. Since the trends observed for the training and validation data set are similar and the primary focus lies on the general performance of the different training algorithms, the RMSEs computed over the whole reference data set are reported here. In addition, due to the availability of only one HDNN for the ED-GEKF+F case, the values given for the A-GEKF, E-GEKF, and ED-GEKF are those obtained with the respective HDNNs exhibiting the smallest RMSE on the validation data set after training.

Table 1. RMSEs of Energies (kcal mol^{-1}) and Forces (kcal mol^{-1} Å^{-1}) over the Whole Reference Data Set

	RMSE	
filter type	energies	forces
A-GEKF	0.86	18.66
E-GEKF	0.20	12.40
ED-GEKF	0.08	11.84
ED-GEKF+F	0.17	6.79

For the HDNNs based on the energy data only, a similar pattern emerges for the forces as previously for the energies. While the A-GEKF is associated with the greatest deviation in the forces (18.66 kcal mol^{-1} Å^{-1}), the difference between the elemental and element-decoupled filters is less pronounced (12.40 kcal mol^{-1} Å^{-1} and 11.84 kcal mol^{-1} Å^{-1}, respectively) but still clearly in favor of the ED-GEKF, which produces the more accurate fit even in this case.

Compared to the standard element-decoupled algorithm, the force variant leads to a dramatic increase in the accuracy of the HDNN-PES forces. The RMSE of the forces over the whole reference data set is reduced to 6.79 kcal mol^{-1} Å^{-1}, almost half of the corresponding ED-GEKF value. However, while forces

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

are reproduced more accurately, a deterioration in the quality of the HDNN-PES energies is observed, with the RMSE increasing from 0.08 kcal mol^{-1} for the ED-GEKF HDNN to 0.17 kcal mol^{-1} for its ED-GEKF+F counterpart. This phenomenon originates from the second fit criterion introduced in the ED-GEKF+F algorithm in the form of the molecular forces. Since every molecule contributes only one energy but $3N_{atom}$ force components to the fitting process, a bias toward the fit of the forces over the energies is introduced. A similar strategy as in the combined function derivative approximation (CFDA) method by Pukrittayakamee and co-workers[42] can be employed to counteract this effect. By introducing a scaling factor for the force component of the weight update in eq 18, the relative importance of energies and forces during HDNN training can be regulated. However, an adaptation of this kind will be the subject of future research. Nevertheless, the deterioration in energies observed in this work is only minor (0.09 kcal mol^{-1}) compared to the greatly improved forces. This feature makes the force variant of the ED-GEKF algorithm especially attractive for cases where accurate forces are required, e.g. molecular dynamics simulations. In addition, since the ED-GEKF+F incorporates not only information on the energies but also information on the gradients (forces) in a similar manner to the CFDA method, no overfitting of the underlying PES should occur, thus eliminating the need for a cross-validation procedure. This implication will be tested in further studies.

5.5. HDNN Interpolation. One major feature of NNs is their capability to reliably interpolate data. To obtain insight into the interpolation performance of the HDNNs trained with the different Kalman filter variants, their ability to reproduce the electronic structure reaction profile of the Claisen rearrangement was studied. HDNN energies were computed for 500 intermediate geometries encountered along the reaction path (starting with allyl vinyl ether and ending in 4-pentenal) and compared to the corresponding reference values. The resulting reaction profiles are shown in Figure 7. RMSEs over

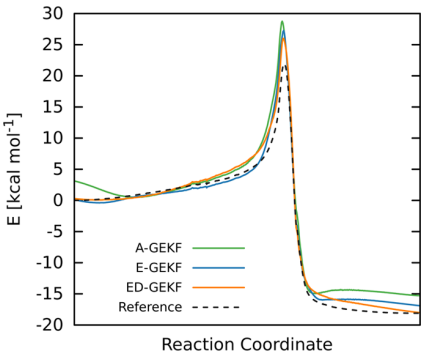


Figure 7. Reaction path of the Claisen rearrangement leading from allyl vinyl ether (left) to 4-pentenal (right) as obtained with the WOELFLING method implemented in TURBOMOLE. The curves for the different HDNNs correspond to the HDNN potential energies predicted for the geometries encountered along the profile.

the whole reaction path, as well as deviations of the HDNNs from the reference energies for the ether, the transition state (TS), and the aldehyde, are given in Table 2. As before, the HDNNs exhibiting the smallest validation RMSE for the respective filter algorithm were used.

Table 2. RMSEs of Energies (kcal mol^{-1}) along the Reaction Path and Deviations from the Reference Energy ΔE (kcal mol^{-1}) for Reactant (Ether), Transition State (TS), and Product (Aldehyde)

		ΔE		
filter type	RMSE	ether	TS	aldehyde
A-GEKF	2.48	3.12	6.81	2.88
E-GEKF	1.43	0.15	5.32	1.20
ED-GEKF	1.24	0.30	4.06	0.10

All investigated HDNN-PESs capture the general shape of the reaction profile, with a tendency to overestimate the barrier height. The HDNN trained with the A-GEKF algorithm yields the least accurate profile with a RMSE of 2.48 kcal mol^{-1}. In the region corresponding to allyl vinyl ether, the energy of the initial reactant is too high (ΔE = 3.12 kcal mol^{-1}), leading to an artificial minimum far from the original equilibrium structure of the ether. The onset of the A-GEKF's barrier, as well as the barrier itself, shows the greatest deviation from the reference energies out of all filter variants (ΔE = 6.81 kcal mol^{-1} for TS). The 4-pentenal region lies too high in energy (ΔE = 2.88 kcal mol^{-1}). In addition, a second artificial minimum is present immediately after the barrier. In case of the E-GEKF, better agreement with the reference curve is obtained (RMSE of 1.43 kcal mol^{-1}). The region preceding the barrier and the onset of the barrier are modeled accurately. Nevertheless, an artificial minimum can still be found close to the reactant. The barrier height is smaller than for the atomic filter variant (ΔE = 5.32 kcal mol^{-1} for TS). Although the region leading to the product is predicted closer to the reference (ΔE = 1.20 kcal mol^{-1}) by the E-GEKF, this part of the potential curve exhibits a similar shape as the one produced by the A-GEKF HDNN, showing an artificial minimum close to the barrier.

The most reliable reproduction of the original reaction profile is obtained by the ED-GEKF HDNN (RMSE of 1.24 kcal mol^{-1}). The region close to the ether is in good agreement with the reference. A small minimum is still present but less pronounced than in the other variants. The slightly larger error of the ED-GEKF (ΔE = 0.30 kcal mol^{-1}) compared to the E-GEKF (ΔE = 0.15 kcal mol^{-1}) observed directly at the equilibrium geometry of the reactant is due to the random nature of the initial HDNN weights and the partitioning of the reference data. Such small deviations from the overall trend that the ED-GEKF yields the highest quality fits are expected at individual points. The energies at the onset of the barrier are overestimated in the ED-GEKF, but the barrier height shows the smallest error compared to the reference energy (ΔE = 4.06 kcal mol^{-1} for TS). In the aldehyde region, energies are slightly overestimated, but near the product geometry, the HDNN energies closely resemble the reference values (ΔE = 0.10 kcal mol^{-1}). Unlike for the other filter variants, no additional minimum is present in the case of the ED-GEKF potential.

While the curve for the ED-GEKF is sufficiently accurate to demonstrate the interpolation capability of HDNNs trained with this algorithm, it still shows deviations from the electronic structure reaction profile, especially in the region of the barrier. The reason for this effect is that a single metadynamics trajectory was used in the generation of the reference data set. Due to this rather naïve sampling approach, no point in the reference data lies exactly on the reaction coordinate. Improved reference data sets could e.g. be generated with guided or self-consistent sampling procedures.[10,35,63]

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

A less sophisticated but yet extremely effective alternative is to include some critical points of the PESs into the training data. Here, three geometries corresponding to the equilibrium structures of allyl vinyl ether and 4-pentenal, as well as the TS structure of the rearrangement, were included in the reference data set to demonstrate the influence of additional training points on the interpolation performance. These geometries correspond to stationary points on the PES and are easily obtainable through routine electronic structure computations. HDNNs (C-40-40, H-40-40, O-40-40) were trained on this minimally expanded set for 100 epochs using the ED-GEKF. The reaction curve obtained for the HDNN with the smallest validation RMSE is depicted in Figure 8, along with the

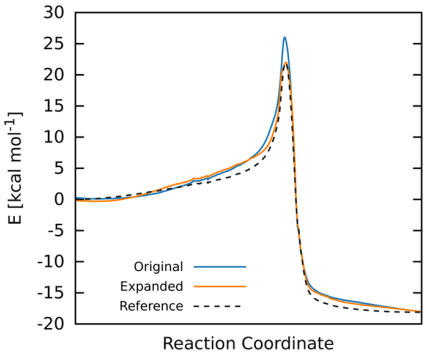


Figure 8. Comparison of the ED-GEKF HDNN reaction profiles obtained with the original reference data set and the data set expanded by the geometries of reactant, TS, and product.

reaction profile previously obtained for the ED-GEKF HDNN. Inclusion of only a few critical points greatly improves the accuracy of the HDNN reaction profile, reducing the RMSE to 0.91 kcal mol^{-1}. Especially the effect of incorporating the TS geometry is noticeable, as the barrier height is now reproduced accurately. The use of this extended reference data set also improves the performance of the other filter variants (not shown), although not to the same extent as in the ED-GEKF case.

Next, the distribution of the reference points in coordinate space is discussed. Although the 17100 points in the original reference data set are not situated directly on the reaction coordinate, the majority of them is nevertheless located in parts of the PES relevant for the transition. HDNN potentials trained on this data can therefore be expected to perform reasonably well in the description of the Claisen reaction, as demonstrated above.

In order to assess the performance of the HDNNs in regions of the PES not sampled during the initial metadynamics simulation, the bond dissociation of one of the hydrogen atoms bound to the C_5 carbon atom of allyl vinyl ether was studied exemplarily. The C−H bond length was varied from 0.8 to 4.5 Å in 500 equidistant steps. The resulting potential energy profiles obtained for the electronic structure reference and the different HDNNs trained on the unexpanded reference data set are shown in Figure 9 for bond lengths between 0.9 and 1.5 Å.

Similar to before, the A-GEKF HDNN produces the worst fit, in this case only remotely reproducing the shape of the reference curve. The E-GEKF and ED-GEKF HDNNs perform well between bond lengths from 1.02 to 1.24 Å (up to energies of approximately 4 kcal mol^{-1}), with the ED-GEKF exhibiting a slightly smaller RMSE (0.36 kcal mol^{-1}) compared to the E-

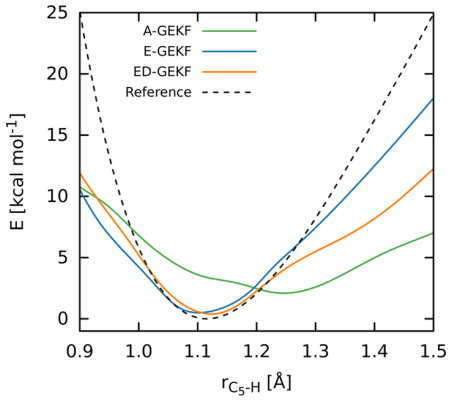


Figure 9. Potential energy profile for the dissociation of one of the hydrogen atoms bound to carbon atom C_5. The curves for the different HDNNs correspond to the HDNN potential energies predicted for the geometries encountered along the profile.

GEKF (0.53 kcal mol^{-1}) for this section of the PES. Such a behavior is expected, since this region of the reaction curve corresponds to the fluctuations of the C−H bond around the equilibrium bond length encountered during the metadynamics simulation and the E-GEKF and ED-GEKF HDNNs interpolate the reference data. Beyond this region, the HDNNs begin to extrapolate, and the quality of the HDNN-PESs deteriorates quickly, resulting in RMSEs close to 60 kcal mol^{-1} over the entire computed curve for all three HDNNs.

These observations are in accordance with the fact that NNs in general excel at interpolating data but perform poorly at extrapolation tasks. Because of this behavior, care should be taken to identify and avoid regions of the PES where extrapolation occurs. Approaches to address this problem exist, see e.g. ref 46. However, in those regions of the PES represented in the reference data set, an excellent fit can be obtained, provided a suitable training algorithm is chosen. Hence, while room for improvement still exists (e.g., addition of TS structure to the training set), the HDNN potential obtained with the ED-GEKF filter can be considered to give a reasonably accurate description of the degrees of freedom encountered in the Claisen reaction.

6. SUMMARY AND CONCLUSION

We report on a new training algorithm for high-dimensional neural networks (HDNNs) of the Behler−Parinello type and its application to the Claisen reaction of allyl vinyl ether. To the best of our knowledge, it is the first study of an organic reaction using HDNNs.

The training algorithm developed to generate the corresponding HDNNs has a substantially improved performance compared to other variants of the Kalman filter employed in HDNN training. In contrast to the latter, the new algorithm − termed element-decoupled global extended Kalman filter (ED-GEKF) − goes without the need for empirical weighting schemes or ad hoc assumptions. By employing the ED-GEKF during training, both root-mean-square errors and maximum fitting errors are reduced significantly. Moreover, fewer training periods and hence less time are required to obtain accurate HDNNs with the ED-GEKF in comparison to the other algorithms. The ED-GEKF can be extended to allow energies and forces to be fit simultaneously during training in a consistent manner, thus improving the description of the

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

HDNN forces tremendously. Especially applications which require highly accurate forces (e.g., molecular dynamics simulations) are expected to profit from this extension. The benefits of the ED-GEKF are expected to become even more pronounced for larger molecular systems or molecules involving many different elements. The ED-GEKF allows for the creation of HDNNs of unprecedented accuracy, putting the treatment of biological problems and complex organic reactions within reach.

In order to arrive at the aforementioned results, we used metadynamics and chain-of-states simulations to obtain the reaction path for the Claisen rearrangement of allyl vinyl ether to 4-pentenal. The potential energies were predicted by HDNNs trained on a set of reference points obtained with BP86/def2-SVP. Due to a naïve sampling scheme, none of the reference points lay on the reaction coordinate. We show that the accuracy of the predicted reaction profile can be drastically improved by including only a few critical points (reactant, transition state, and product geometries) in the reference data. This fact calls for the development of more efficient sampling procedures in the future in order to facilitate the automated generation of highly accurate potential energy surfaces for complicated systems and reactions with HDNNs.

AUTHOR INFORMATION

Corresponding Author

*Phone 43 1 4277 52764. Fax 43 1 4277 9527. E-mail: philipp.marquetand@univie.ac.at.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Leticia González, Christoph Flamm, and Jörg Behler for inspiring discussions. Allocation of computer time at the Vienna Scientific Cluster (VSC) is gratefully acknowledged.

REFERENCES

(1) Levine, I. N. *Quantum Chemistry*, 7th ed.; Prentice Hall: Boston, 2013.

(2) Friesner, R. A. *Ab initio* quantum chemistry: Methodology and applications. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6648−6653.

(3) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(4) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(5) Liang, T.; Shin, Y. K.; Cheng, Y.-T.; Yilmaz, D. E.; Vishnu, K. G.; Verners, O.; Zou, C.; Phillpot, S. R.; Sinnott, S. B.; van Duin, A. C. T. Reactive Potentials for Advanced Atomistic Simulations. *Annu. Rev. Mater. Res.* **2013**, *43*, 109−129.

(6) Handley, C. M.; Behler, J. Next generation interatomic potentials for condensed systems. *Eur. Phys. J. B* **2014**, *87*, 1−16.

(7) Brown, A.; Braams, B. J.; Christoffel, K.; Jin, Z.; Bowman, J. M. Classical and quasiclassical spectral analysis of CH_5^+ using an *ab initio* potential energy surface. *J. Chem. Phys.* **2003**, *119*, 8790−8793.

(8) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.

(9) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(10) Ischtwan, J.; Collins, M. A. Molecular potential energy surfaces by interpolation. *J. Chem. Phys.* **1994**, *100*, 8080−8088.

(11) Wu, T.; Werner, H.-J.; Manthe, U. First-Principles Theory for the H + CH_4 → H_2 + CH_3 Reaction. *Science* **2004**, *306*, 2227−2229.

(12) Takata, T.; Taketsugu, T.; Hirao, K.; Gordon, M. S. Ab initio potential energy surface by modified Shepard interpolation: Application to the CH_3 + H_2 → CH_4 + H reaction. *J. Chem. Phys.* **1998**, *109*, 4281−4289.

(13) Crespos, C.; Collins, M. A.; Pijper, E.; Kroes, G. J. Multi-dimensional potential energy surface determination by modified Shepard interpolation for a molecule-surface reaction: H_2 + Pt(111). *Chem. Phys. Lett.* **2003**, *376*, 566−575.

(14) Dawes, R.; Thompson, D. L.; Guo, Y.; Wagner, A. F.; Minkoff, M. Interpolating moving least-squares methods for fitting potential energy surfaces: Computing high-density potential energy surface data from low-density *ab initio* data points. *J. Chem. Phys.* **2007**, *126*, 184108.

(15) Dawes, R.; Wang, X.-G.; Carrington, T., Jr. CO Dimer: New Potential Energy Surface and Rovibrational Calculations. *J. Phys. Chem. A* **2013**, *117*, 7612−7630.

(16) Vitek, A.; Stachon, M.; Kromer, P.; Snael, V. Towards the Modeling of Atomic and Molecular Clusters Energy by Support Vector Regression. International Conference on Intelligent Networking and Collaborative Systems (INCoS). 2013; pp 121−126.

(17) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(18) Lorenz, S.; Scheffler, M.; Gross, A. Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface. *Phys. Rev. B* **2006**, *73*, 115431.

(19) Ludwig, J.; Vlachos, D. G. *Ab initio* molecular dynamics of hydrogen dissociation on metal surfaces using neural networks and novelty sampling. *J. Chem. Phys.* **2007**, *127*, 154716.

(20) Manzhos, S.; Yamashita, K.; Carrington, T., Jr. Fitting sparse multidimensional data with low-dimensional terms. *Comput. Phys. Commun.* **2009**, *180*, 2002−2012.

(21) Carbogno, C.; Behler, J.; Groß, A.; Reuter, K. Fingerprints for Spin-Selection Rules in the Interaction Dynamics of O_2 at Al(111). *Phys. Rev. Lett.* **2008**, *101*, 096104.

(22) Behler, J.; Reuter, K.; Scheffler, M. Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface. *Phys. Rev. B* **2008**, *77*, 115421.

(23) Latino, D. A. R. S.; Fartaria, R. P. S.; Freitas, F. F. M.; Aires-de Sousa, J.; Silva Fernandes, F. M. S. Mapping Potential Energy Surfaces by Neural Networks: The ethanol/Au(111) interface. *J. Electroanal. Chem.* **2008**, *624*, 109−120.

(24) Latino, D. A. R. S.; Fartaria, R. P. S.; Freitas, F. F. M.; Aires-De-Sousa, J.; Silva Fer-nandes, F. M. S. Approach to potential energy surfaces by neural networks. A review of recent work. *Int. J. Quantum Chem.* **2010**, *110*, 432−445.

(25) Liu, T.; Fu, B.; Zhang, D. H. Six-dimensional potential energy surface of the dissociative chemisorption of HCl on Au(111) using neural networks. *Sci. China Chem.* **2013**, *57*, 147−155.

(26) Tafeit, E.; Estelberger, W.; Horejsi, R.; Moeller, R.; Oettl, K.; Vrecko, K.; Reibnegger, G. Neural networks as a tool for compact representation of *ab initio* molecular potential energy surfaces. *J. Mol. Graphics Modell.* **1996**, *14*, 12−18.

(27) Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. Combining *ab initio* computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules. *J. Chem. Phys.* **1996**, *105*, 7597−7604.

(28) Houlding, S.; Liem, S. Y.; Popelier, P. L. A. A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer. *Int. J. Quantum Chem.* **2007**, *107*, 2817−2827.

(29) No, K. T.; Chang, B. H.; Kim, S. Y.; Jhon, M. S.; Scheraga, H. A. Description of the potential energy surface of the water dimer with an artificial neural network. *Chem. Phys. Lett.* **1997**, *271*, 152−156.

(30) Cho, K.-H.; No, K. T.; Scheraga, H. A. A polarizable force field for water using an artificial neural network. *J. Mol. Struct.* **2002**, *641*, 77−91.

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

(31) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. Representation of Intermolecular Potential Functions by Neural Networks. *J. Phys. Chem. A* **1998**, *102*, 4596−4605.
(32) Prudente, F. V.; Acioli, P. H.; Soares Neto, J. J. The fitting of potential energy surfaces using neural networks: Application to the study of vibrational levels of H_3^+. *J. Chem. Phys.* **1998**, *109*, 8801−8808.
(33) Rocha Filho, T. M.; Oliveira, Z. T.; Malbouisson, L. A. C.; Gargano, R.; Soares Neto, J. J. The use of neural networks for fitting potential energy surfaces: A comparative case study for the H_3^+ molecule. *Int. J. Quantum Chem.* **2003**, *95*, 281−288.
(34) Malshe, M.; Raff, L. M.; Rockley, M. G.; Hagan, M.; Agrawal, P. M.; Komanduri, R. Theoretical investigation of the dissociation dynamics of vibrationally excited vinyl bromide on an *ab initio* potential-energy surface obtained using modified novelty sampling and feedforward neural networks. II. Numerical application of the method. *J. Chem. Phys.* **2007**, *127*, 134105.
(35) Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. *Ab initio* potential-energy surfaces for complex, multichannel systems using modified novelty sampling and feedforward neural networks. *J. Chem. Phys.* **2005**, *122*, 084104.
(36) Agrawal, P. M.; Raff, L. M.; Hagan, M. T.; Komanduri, R. Molecular dynamics investigations of the dissociation of SiO_2 on an *ab initio* potential energy surface obtained using neural network methods. *J. Chem. Phys.* **2006**, *124*, 134306.
(37) Le, H. M.; Huynh, S.; Raff, L. M. Molecular dissociation of hydrogen peroxide (HOOH) on a neural network *ab initio* potential surface with a new configuration sampling method involving gradient fitting. *J. Chem. Phys.* **2009**, *131*, 014107.
(38) Manzhos, S.; Carrington, T., Jr. Using neural networks to represent potential surfaces as sums of products. *J. Chem. Phys.* **2006**, *125*, 194105.
(39) Le, H. M.; Raff, L. M. Cis → trans, trans → cis isomerizations and N-O bond dissociation of nitrous acid (HONO) on an *ab initio* potential surface obtained by novelty sampling and feed-forward neural network fitting. *J. Chem. Phys.* **2008**, *128*, 194310.
(40) Darley, M. G.; Handley, C. M.; Popelier, P. L. A. Beyond Point Charges: Dynamic Polarization from Neural Net Predicted Multipole Moments. *J. Chem. Theory Comput.* **2008**, *4*, 1435−1448.
(41) Le, H. M.; Dinh, T. S.; Le, H. V. Molecular Dynamics Investigations of Ozone on an Ab Initio Potential Energy Surface with the Utilization of Pattern-Recognition Neural Network for Accurate Determination of Product Formation. *J. Phys. Chem. A* **2011**, *115*, 10862−10870.
(42) Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *J. Chem. Phys.* **2009**, *130*, 134101.
(43) Nguyen, H. T. T.; Le, H. M. Modified Feed-Forward Neural Network Structures and Combined-Function-Derivative Approximations Incorporating Exchange Symmetry for Potential Energy Surface Fitting. *J. Phys. Chem. A* **2012**, *116*, 4629−4638.
(44) Chen, J.; Xu, X.; Xu, X.; Zhang, D. H. Communication: An accurate global potential energy surface for the OH + CO → H + CO_2 reaction using neural networks. *J. Chem. Phys.* **2013**, *138*, 221104.
(45) Li, J.; Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems. *J. Chem. Phys.* **2013**, *139*, 204103.
(46) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930−17955.
(47) Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys.: Condens. Matter* **2014**, *26*, 183001.
(48) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114*, 3371−3383.
(49) Witkoskie, J. B.; Doren, D. J. Neural Network Models of Potential Energy Surfaces:Prototypical Examples. *J. Chem. Theory Comput.* **2005**, *1*, 14−23.
(50) Behler, J.; Lorenz, S.; Reuter, K. Representing molecule-surface interactions with symmetry-adapted neural networks. *J. Chem. Phys.* **2007**, *127*, 014705.
(51) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.
(52) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(53) Manzhos, S.; Carrington, T., Jr. A random-sampling high dimensional model representation neural network for building potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, 084109.
(54) Artrith, N.; Hiller, B.; Behler, J. Neural network potentials for metals and oxides - First applications to copper clusters at zinc oxide. *Phys. Status Solidi B* **2013**, *250*, 1191−1203.
(55) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 153101.
(56) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. Metadynamics Simulations of the High-Pressure Phases of Silicon Employing a High-Dimensional Neural Network Potential. *Phys. Rev. Lett.* **2008**, *100*, 185501.
(57) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. Pressure-induced phase transitions in silicon studied by neural network-based metadynamics simulations. *Phys. Status Solidi B* **2008**, *245*, 2618−2629.
(58) Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. *Ab initio* quality neural-network potential for sodium. *Phys. Rev. B* **2010**, *81*, 184107.
(59) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Graphite-diamond phase coexistence study employing a neural-network mapping of the *ab initio* potential energy surface. *Phys. Rev. B* **2010**, *81*, 100103.
(60) Seema, P.; Behler, J.; Marx, D. Adsorption of Methanethiolate and Atomic Sulfur at the Cu(111) Surface: A Computational Study. *J. Phys. Chem. C* **2013**, *117*, 337−348.
(61) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections. *J. Phys. Chem. A* **2013**, *117*, 7356−7366.
(62) Morawietz, T.; Behler, J. A Full-Dimensional Neural Network Potential-Energy Surface for Water Clusters up to the Hexamer. *Z. Phys. Chem.* **2013**, *227*, 1559−1581.
(63) Morawietz, T.; Sharma, V.; Behler, J. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **2012**, *136*, 064103.
(64) Puskorius, G. V.; Feldkamp, L. A. Decoupled extended Kalman filter training of feedforward layered networks. IJCNN-91-Seattle International Joint Conference on Neural Networks. 1991; pp 771−777.
(65) Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Syst.* **1989**, *2*, 303−314.
(66) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359−366.
(67) Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **1991**, *4*, 251−257.
(68) Bishop, C. M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: New York, 2006.
(69) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533−536.
(70) Hagan, M.; Menhaj, M. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Networks* **1994**, *5*, 989−993.
(71) Blank, T. B.; Brown, S. D. Adaptive, global, extended Kalman filters for training feedforward neural networks. *J. Chemom.* **2005**, *8*, 391−407.

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

(72) Shah, S.; Palmieri, F.; Datum, M. Optimal filtering algorithms for fast learning in feedforward neural networks. *Neural Networks* **1992**, *5*, 779−787.
(73) Murtuza, S.; Chorian, S. F. Node decoupled extended Kalman filter based learning algorithm for neural networks. Proceedings of the 1994 IEEE International Symposium on Intelligent Control. 1994; pp 364−369.
(74) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78.
(75) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(76) Dirac, P. A. M. Quantum Mechanics of Many-Electron Systems. *Proc. R. Soc., Ser. A* **1929**, *123*, 714−733.
(77) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.
(78) Slater, J. C. A Simplification of the Hartree-Fock Method. *Phys. Rev.* **1951**, *81*, 385−390.
(79) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(80) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(81) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283−290.
(82) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.
(83) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(84) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101−024101−7.
(85) Plessow, P. Reaction Path Optimization without NEB Springs or Interpolation Algorithms. *J. Chem. Theory Comput.* **2013**, *9*, 1305−1310.
(86) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 91−100.
(87) *TURBOMOLE* V6.6 2014, a development of the University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007, TURBOMOLE GmbH: since 2007. Available from http://www.turbomole.com (accessed 03.04.2015).
(88) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Reprint ed.; Cambridge University Press: Cambridge, 2012.
(89) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(90) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637−649.
(91) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(92) Tolman, R. C. *The Principles of Statistical Mechanics*, New ed.; Dover Publications: New York, 2010.
(93) Nguyen, D. H.; Widrow, B. Neural networks for self-learning control systems. *IEEE Control Systems Magazine* **1990**, *10*, 18−23.
(94) Plaut, D. C.; Nowlan, S. J.; Hinton, G. E. *Experiments on Learning by Back Propagation*; Technical Report; 1986.
(95) van Rossum, G.; Drale, F. L. *Python Reference Manual*; PythonLabs: Virginia, USA, 2001. http://www.python.org (accessed 03.04.2015).
(96) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(97) Ziegler, F. E. The thermal, aliphatic Claisen rearrangement. *Chem. Rev.* **1988**, *88*, 1423−1452.

DOI: 10.1021/acs.jctc.5b00211
J. Chem. Theory Comput. 2015, 11, 2187−2198

APPENDIX A.2 COMPARING THE ACCURACY OF HIGH-DIMENSIONAL NEURAL NETWORK POTENTIALS AND THE SYSTEMATIC MOLECULAR FRAGMENTATION METHOD: A BENCHMARK STUDY FOR ALL-TRANS ALKANES

MICHAEL GASTEGGER, CLEMENS KAUFFMANN, JÖRG BEHLER, AND PHILIPP MARQUETAND

J. Chem. Phys., **144**, 194110 (2016).
http://dx.doi.org/10.1063/1.4950815

Contributions:

MICHAEL GASTEGGER implemented the fragmentation scheme, supervised the exemplary computations, and contributed to the initial draft and final version of the manuscript.

CLEMENS KAUFFMANN performed the exemplary computations, analyzed the results, and contributed to the initial draft of the manuscript.

JÖRG BEHLER provided the RUNNER program[121], and participated in the preparation of the final manuscript.

PHILIPP MARQUETAND conceived, initiated and supervised the project, and contributed to the final manuscript.

THE JOURNAL OF CHEMICAL PHYSICS **144**, 194110 (2016)

Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes

Michael Gastegger,[1] Clemens Kauffmann,[1] Jörg Behler,[2] and Philipp Marquetand[1,a)]
[1]*Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, Vienna, Austria*
[2]*Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, Universitätsstraße 150, Bochum, Germany*

(Received 12 April 2016; accepted 5 May 2016; published online 20 May 2016)

Many approaches, which have been developed to express the potential energy of large systems, exploit the locality of the atomic interactions. A prominent example is the fragmentation methods in which the quantum chemical calculations are carried out for overlapping small fragments of a given molecule that are then combined in a second step to yield the system's total energy. Here we compare the accuracy of the systematic molecular fragmentation approach with the performance of high-dimensional neural network (HDNN) potentials introduced by Behler and Parrinello. HDNN potentials are similar in spirit to the fragmentation approach in that the total energy is constructed as a sum of environment-dependent atomic energies, which are derived indirectly from electronic structure calculations. As a benchmark set, we use all-trans alkanes containing up to eleven carbon atoms at the coupled cluster level of theory. These molecules have been chosen because they allow to extrapolate reliable reference energies for very long chains, enabling an assessment of the energies obtained by both methods for alkanes including up to 10 000 carbon atoms. We find that both methods predict high-quality energies with the HDNN potentials yielding smaller errors with respect to the coupled cluster reference. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4950815]

I. INTRODUCTION

Computer simulations of chemical processes rely on the potential energy surfaces (PESs) of the structures involved,[1] and consequently, the accuracy of these PESs defines the quality of the simulations. While highly accurate *ab initio* calculations are at hand for moderately sized systems, larger systems can only be addressed by employing an increasing number of empirical approximations in order to keep the computational effort feasible, which necessarily results in a reduced accuracy of the obtained energies. Thus, maintaining accuracy while enabling a fast evaluation is one of the main goals when constructing PESs. Many different approaches have been developed in past decades, which have either been based on physical considerations or on purely mathematical principles.

Within the latter subgroup, PESs derived from machine learning techniques,[2] and in particular, employing neural networks (NNs),[3–13] have made a lot of progress. NNs are nonlinear models inspired by the central nervous system, which are especially adept at interpolating trends in existing data. Their flexible and unbiased nature has lead to a variety of NN-based applications in many fields[14] and makes them a useful tool for fitting PESs for different types of chemical systems.[15,16] However, early NN potentials usually required system-specific adoptions and were limited to small numbers of atoms, which has been finally resolved in the high-dimensional NN (HDNN) approach by Behler and Parrinello.[4]

Still, the applicability of NN-based methods is limited by the need for large sets of *ab initio* reference calculations in order to construct a valid and accurate potential. Especially for large molecular systems—such as proteins—these reference calculations quickly become prohibitive, due to the scaling behavior of high-level *ab initio* methods. In the HDNN approach, the need for reference calculations comprising the full systems of interest is circumvented by the exploitation of the so-called[17] chemical locality. Consequently, it is possible to construct HDNNs based solely on fragments of the original molecular system, while the validity for the full system is retained. Hence, one costly reference computation can be replaced by several significantly cheaper calculations on smaller subsystems. This approach is well tested for solid state systems[18–20] as well as for molecular clusters[21] and liquid water[22] and has been used in numerous applications.[5]

Amongst the physically motivated approaches are fragmentation-based methods, where the original system is first divided into smaller independent subsystems. The properties of these fragments (e.g., energies) are then calculated with *ab initio* methods and recombined to obtain the composite properties of the whole molecular system. Several fragmentation schemes have been developed over the last 20 years, differing mainly in how the original system is divided into fragments and how the recombination step is carried out.[23,24] Here, we focus on the systematic molecular fragmentation approach (SMF) developed by Collins and coworkers.[25–27] The SMF approach generates overlapping fragments of a certain size by gathering bonded atoms into functional groups. The energy of the total system is calculated by summing the energies of

a)Electronic mail: philipp.marquetand@univie.ac.at

these fragments and subtracting the energy contributions of the overlap regions. SMF has been applied successfully to a wide range of molecular systems, including proteins,[27,28] water clusters,[28] SiO_2 crystals,[26] and organic molecules.[29]

The aim of the present study is to assess and compare the performance of the HDNN method and of the SMF approach in terms of the accuracy of the obtained potential energies. For this purpose, linear all-trans alkane chains of varying lengths containing up to 10 000 carbon atoms have been chosen as a model system. Reference computations for the shorter chains containing up to eleven carbon atoms have been carried out directly with the coupled cluster method including single, double, and perturbative triple excitations (CCSD(T)), while the reference energies of longer chains have been extrapolated using corrected energies of the fragmentation approach. Based on these reference calculations, we investigate how well HDNNs and the conventional fragmentation approach can predict the potential energies of large organic molecules if only the energies of the small fragments accessible by CCSD(T) are provided. The simplicity of our model system is motivated by the need for high-quality reference energies for very large molecules, which could not be obtained for more complex systems as detailed below, but our findings are general and not restricted to linear alkanes.

II. METHODS

A. High dimensional neural network potentials

Similar to their biological counterparts, NNs are assembled from several interconnected subunits, called neurons. These neurons collect and process incoming signals (e.g., molecular geometries) and assign an output (e.g., the potential energy). This processing is performed by computing a weighted sum and applying a nonlinear activation function, where the network weights control the magnitude of the incoming signals. If the input signals are the outputs of other neurons, a network structure is obtained and the weights represent the connections between the neurons in the network. In analogy to biological learning, the strength of these connections has to be determined in order to obtain NNs suitable for practical use and the weights are hence the important fitting parameters of a NN.

Unfortunately, the basic NN structure outlined above suffers from several drawbacks when applied to the interpolation of PESs. Once the weight parameters have been learned, the structure of the NN is fixed. As a consequence, the NN can only be used for molecules with the same number of atoms and elemental composition. Moreover, the output of the NN is not invariant with respect to translations and rotations of the molecule if standard Cartesian coordinates are used as inputs. One method to overcome these problems is the high-dimensional NN (HDNN) approach developed by Behler and Parrinello.[4,30]

In the HDNN approach, shown schematically in Figure 1, each atom of a system is characterized by its chemical environment. Depending on this environment, its energy contribution E_i to the total potential energy E_{pot} is then calculated as output of an individual atomic NN, which is

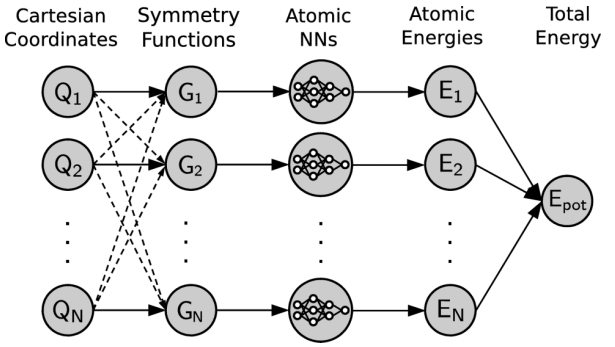


FIG. 1. Schematic structure of a high-dimensional neural network potential.[4] Each Cartesian atomic coordinate vector $\mathbf{Q}_i$ is transformed to a symmetry function vector $\mathbf{G}_i$, which is used as the input for the respective atomic NN. The resulting energy contributions E_i are summed to yield the molecule's potential energy.

usually a conventional feed-forward NN.[31] By summing these energy contributions, E_{pot} is obtained. The individual atomic NNs are identical for a given element to ensure the required permutation invariance of the final PES.

The local chemical environments of the atoms i are described via sets, i.e., vectors, of many-body atom-centered symmetry functions $\mathbf{G}_i$ (ACSFs), which depend on all Cartesian atomic position vectors $\mathbf{Q}_i$ within a predefined cutoff radius around the respective central atom. These symmetry functions resemble radial and angular distribution functions and are invariant to translations and rotations of the molecule, thus eliminating one of the problems of standard NNs. The introduction of a cutoff radius restricts the description of the atomic environments to the chemically relevant regions and facilitates exploiting chemical locality in the training and application of the HDNNs. An in-depth description of HDNNs and suitable symmetry functions can be found elsewhere.[5,30,32]

As stated above, the weights of the NNs have to be optimized in order to obtain meaningful potential energy predictions. This is done in a process called "training," where a reference set of geometries and corresponding energies is iteratively reproduced to minimize the root mean squared error (RMSE) of the energies predicted by the NN. This minimization can be achieved by a variety of algorithms, e.g., stochastic gradient descent[33] or Levenberg–Marquardt optimization.[34,35] In the present work, a special adaption of the global extended Kalman filter[36,37] for HDNNs, the element-decoupled Kalman filter,[38] has been used. This algorithm is well suited for the flexible structure of HDNNs and results in improved training speeds and an increased quality of the resulting PESs for molecular systems.

B. Systematic fragmentation method

The SMF method has been used for two purposes in the present work. First, its performance has been tested and compared to that of HDNNs. Second, it has been applied in combination with an energy correction scheme to provide very accurate reference energies, which enabled to test both methods for systems being inaccessible for direct coupled cluster calculations.

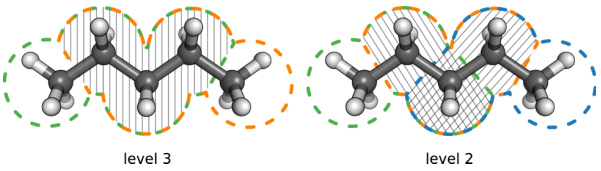


FIG. 2. Exemplary fragmentation of C_5H_{12} with fragmentation levels 3 and 2. The C–C bonds are broken homolytically, and hydrogen caps are added both on the respective colored fragments and shaded overlaps. The energies of the fragments are summed and the energies of the overlaps are subtracted, yielding an approximation for the molecule's potential energy.

The basic principle of the SMF method[25–27] is illustrated in Figure 2 using the fragmentation of C_5H_{12} as an example. The fragments (highlighted in green, red, and blue) are constructed from the functional groups of the molecule, in the case of alkanes CH_2 and CH_3 groups. The single-bonds are broken homolytically and hydrogen caps are added to maintain charge neutrality. The molecule's potential energy is then approximated by adding the fragment energies and subtracting the "double counted," shaded overlapping regions. By using larger fragments, thus increasing the overlap size, the approximation becomes more accurate and approaches the calculation results for the entire molecule. The overlap size is denominated by the fragmentation level X, where X indicates the number of functional groups (saturated C-atoms in our case) within the overlap.

III. COMPUTATIONAL DETAILS

All quantum mechanical reference calculations were carried out with ORCA.[39] Geometry optimizations were performed at the RI-MP2/cc-pVTZ level of theory.[40] MP2 correlation and Coulomb integrals were calculated employing the resolution of identity approximation[41,42] as well as the COSX numerical integration[43,44] for the Hartree-Fock exchange term. Single-point energies of the optimized structures and all of its fragments were obtained using implicitly correlated CCSD(T)-F12 with the resolution of identity approximation and the cc-pVTZ, cc-pVDZ-F12,[45] and cc-pVDZ-F12-CABS[46] basis sets.

For the SMF approach, C–C bonds were broken homolytically and hydrogen caps were added according to Collins,[28] using covalent radii of 0.31 Å for hydrogen and 0.76 Å for carbon. In total, 9 optimized alkanes containing between 3 and 11 carbon atoms, an additional alkane with 11 carbon atoms, and all 474 non-optimized fragments of these molecules have been calculated and included in the reference set irrespective of close structural similarities between many of these fragments.

The HDNN construction and training were carried out using the RuNNer code.[47] The atomic environments were characterized by ACSFs,[30] whose parameters are given in the supplementary material.[49] A combination of 8 radial and 24 angular functions was employed for both carbon and hydrogen. A cutoff radius of 5 Å was used for all ACSFs. Consequently each atomic NN contains 32 input nodes corresponding to the individual ACSFs, and one output node was used to obtain the atomic potential energy contribution. The architectures of the atomic NNs were determined by an initial training run using subnets with 1 or 2 hidden layers consisting of up to 35 nodes. Based on these preliminary training results (average error, standard deviation, minimal deviation), the five most promising architectures were chosen. The architectures are read as "first hidden layer"-"second hidden layer": 2-2, 3-4, 5-4, 10-2, and 15 for both carbon and hydrogen. Hyperbolic tangents were employed as activation function in the hidden layers, while a linear transformation was applied to the output layers.

The training process was performed using the "element-decoupled" global extended Kalman filter.[38] The weight parameters were adjusted over 150 epochs and an adaptive filter threshold of 0.9 times the RMSE of the previous epoch was used. Values of $\lambda_0 = 0.9995$ and $\lambda_k = 0.95$ were employed for the time-dependent forgetting schedule, and the network weights were initialized according to the scheme of Nguyen and Widrow.[48] In order to facilitate the training, the energies of the free atoms were subtracted from the reference energies of the studied molecules. Overfitting was controlled by early stopping using cross validation[5] with randomly chosen training and test sets with a ratio of 9:1. Five different random seeds were used to determine training and test set compositions of each HDNN architecture. In this way, the influence of the test set composition was ensured to be negligible. The HDNN with the lowest test set RMSE was then chosen for the subsequent calculations, a model with elemental NNs of size 15 and training set and test set RMSEs of 0.000 63 (kcal/mol)/atom and 0.001 26 (kcal/mol)/atom, respectively.

IV. RESULTS AND DISCUSSION

A. Fragmentation

The accuracy of the SMF approach for the model system used in this work is studied using short all-trans alkane chains with lengths ranging from 3 to 11 carbon atoms. After geometry optimization at the MP2 level, single point energies are computed with CCSD(T). Based on these optimized geometries, systematic fragmentation is carried out with fragmentation levels from 1 to the respective maximum level given by the chain length. The energies of the full alkane molecules obtained in this way are then compared to their respective CCSD(T) values.

Using alkanes from C_6H_{14} to $C_{10}H_{22}$ as an example, Figure 3 compares the fragmentation-derived potential energies E_{FragX} with the full-sized CCSD(T) calculations E_{CC}. With higher fragmentation levels, the potential energy approaches the coupled cluster result of the entire molecule, as higher fragmentation levels account for a larger overlap region between the fragmentation sites. Since a higher ratio of the entire molecule is included when calculating each fragment, the general convergence trend observed in Figure 3 can be expected. However, the increased potential energy difference of level 4 compared to 3 is interesting to note. We have not found a satisfactory explanation for this behavior.

Figure 4 shows the deviation of fragmentation energies obtained for alkanes of length 5 to 11 using fragmentation

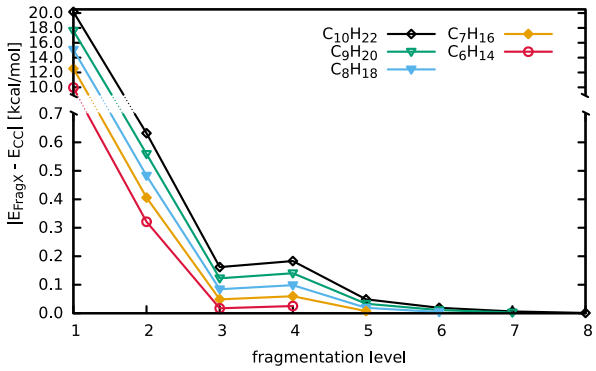


FIG. 3. Total energy deviations of the energies computed at different fragmentation levels from the coupled cluster reference calculations for the alkanes C_6H_{14} to $C_{10}H_{22}$.

levels starting with level 3 from the CCSD(T) results. Investigating the deviation as a function of the molecule size, the following trend can be observed: At a given fragmentation level, the energy difference increases with the number of carbon atoms (a trend which can also be observed in Figure 3). The reason for this behavior is the way the energy of the whole molecule is computed in the SMF approach. By using fragments of the same size to construct alkanes of different lengths, the corresponding error in energy is replicated with every additional C-atom, resulting in the approximately linear trend shown in Figure 4. Hence, the intrinsic error of the respective fragmentation level becomes visible.

This error is only small for short alkanes, but it increases with chain size. This linear increase in the error is a consequence of the chosen model system as each CH_2 group contributes approximately the same error with respect to the coupled cluster reference. This linear relation can therefore be employed to construct an energy correction for alkane chains of arbitrary length as shown for fragmentation level 5 in Figure 4. Taking this correction into account, the almost exact CCSD(T) values are recovered for chains with up to 11 C-atoms and it is reasonable to assume that this trend holds also for longer chains, where CCSD(T) calculations are unfeasible. The corrected energies obtained by adding the correction to the fragmentation energies are denoted as E_{corr}. While these energies can be calculated for every fragmentation

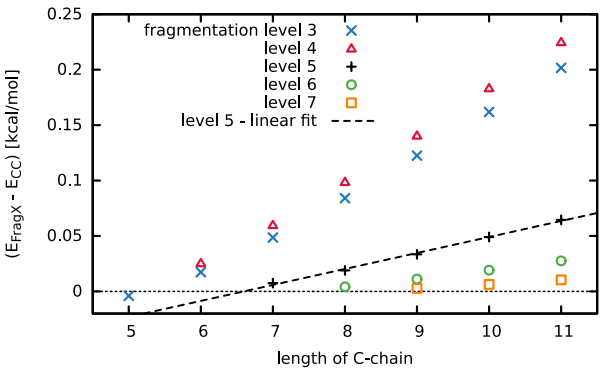


FIG. 4. Energy deviations between the fragmentation method and the coupled cluster calculations for different alkanes and fragmentation levels.

level, the E_{corr} values for the different fragmentation levels show only extremely small deviations from each other (within 1.1 kcal/mol for the 10 000 carbon chain), demonstrating the stability of the correction. In what follows, we use the E_{corr} derived from fragmentation level of 5 as it offered a sufficient amount of data points with reasonably small deviations from CCSD(T) results. By using this correction, we can go beyond the standard accuracy of the fragmentation method. However, this is possible only due to the linear nature of the chosen model system and such a scheme would not be applicable for arbitrary organic molecules, which is the reason why we have chosen linear alkanes for the present benchmark study.

B. Neural networks

In order to assess the ability of HDNNs to model the potential energy of large linear alkanes based on the information contained in small fragments, the five NN architectures introduced in Section III are used to predict potential energies of all-trans alkane chains with lengths up to 10 000 carbon atoms. Since the geometry optimization of alkanes of this size with *ab initio* methods is impossible, model geometries are used. These structures are obtained by replicating fragments based on the MP2 optimized bond lengths, angles, and dihedral angles calculated for $C_{10}H_{22}$ until the desired length is reached. In the present work, alkanes containing 11 to 10 000 carbon atoms are generated in this manner. The reference energies of these chains were computed using the level 5 fragmentation approach augmented by the previously derived correction.

The training set employed in the construction of the HDNN potentials contained the MP2 optimized alkanes (3 to 11 carbons) and a $C_{11}H_{24}$ structure generated as outlined in the previous paragraph, as well as all their respective fragments. Note that a training set with purely MP2 based geometries leads to strong fluctuations in the predicted potential energies for the long artificial chains. The reason for this effect is the highly regular nature of the linear alkane model system, which prevents a comprehensive sampling of the possible configuration space. As a consequence, the HDNNs are sensitive with respect to tiny differences between MP2-optimized and artificially generated geometries, which is not expected to happen in typical molecular systems if the PESs are based on more representative data sets of the relevant configuration space.

The deviations from E_{corr} of the potential energies computed with the different methods are illustrated in Figure 5. For the HDNNs, the predictions with the highest and lowest deviation ($E_{NN_{max}}$ and $E_{NN_{min}}$) are shown. They are compared with the E_{FragX} approximations, where X denominates the corresponding fragmentation level. In order to achieve a reasonable scale, the energy is normalized to the number of atoms N for demonstrative purposes.

Once again, the trend of the E_{FragX} energies to yield more accurate approximations with higher fragmentation levels can be observed. However, all HDNN approximations yield even smaller deviations from E_{corr}, with the maximum deviations still lying below the ones obtained for fragmentation level 7 and the best HDNNs ($E_{NN_{min}}$) performing significantly

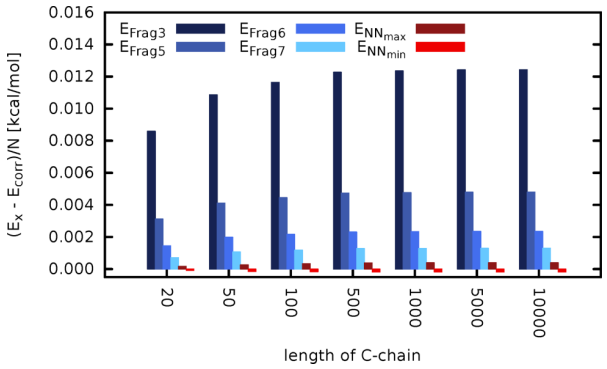


FIG. 5. Potential energies derived from fragmentation and NN approximations (E_{FragX} and E_{NN}) in comparison to the error-corrected level 5 fragmentation results E_{corr}. The energy is normalized to the number of atoms N of the alkanes.

better. This result is remarkable insofar, as the choice of a 5 Å cutoff radius used in the ACSFs limits the effective chemical environment seen by a HDNN to a maximum of 7 carbon atoms. Compared to the SMF method, this number of carbons corresponds to a fragmentation level of 6, which shows significantly larger deviations than the HDNNs. Apparently, the HDNNs are able to exploit chemical locality to a greater extent compared to the standard SMF method and hence utilize the information present in the molecular fragments in a more efficient manner.

The errors in the NN predictions in general exhibit the same linearity as the fragmentation derived values, which is to be expected as also for the HDNN potential, each additional CH_2 group contributes a certain energy error, but for the given reference the NN energies are notably more accurate. Thus, HDNNs represent a promising alternative to the fragmentation method, and we believe that this finding also holds for general organic molecules. A comparison between the SMF approach and HDNNs is more difficult in this case, as no simple corrections can be exploited and accurate reference data are hence more difficult or even impossible to obtain.

V. CONCLUSION

A comparison of the performance of high-dimensional neural network (HDNNs) potentials and of the SMF approach for the energies of linear all-trans alkanes has been presented. Due to the linearity of the energy error of the fragmentation approach with system size, an energy correction scheme could be implemented that enabled to assess the accuracy of both methods for systems containing up to 10 000 C-atoms. While both approaches provide very accurate energies close to the underlying coupled cluster data, the energy errors employing the HDNN approach have been found to be systematically smaller for all chain lengths. Unlike the fragmentation method, the purely mathematical structure of HDNNs is not restricted by underlying physical considerations. Another advantage of HDNN potentials is their transferability. Once trained, they can be used to obtain the energy of sufficiently similar molecules, without the need of additional *ab initio* calculations. This principal flexibility, accuracy, and efficiency illustrate the benefits of HDNNs for other chemical systems and applications. However, it should once again be stressed that the model system studied in this work is extremely well behaved and exhibits no significant long range electrostatic or dispersion interactions. Whether the results of our particular model system can be reproduced for more complex systems like proteins will be the subject of further studies.

ACKNOWLEDGMENTS

Allocation of computer time at the Vienna Scientific Cluster (VSC) is gratefully acknowledged. J.B. is grateful for financial support by the DFG.

[1]I. N. Levine, *Quantum Chemistry*, 7th ed. (Prentice Hall, Boston, 2013).
[2]C. M. Handley and J. Behler, "Next generation interatomic potentials for condensed systems," Eur. Phys. J. B **87**, 152 (2014).
[3]T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, "Neural network models of potential energy surfaces," J. Chem. Phys. **103**, 4129–4137 (1995).
[4]J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).
[5]J. Behler, "Representing potential energy surfaces by high-dimensional neural network potentials," J. Phys. Condens. Matter **26**, 183001 (2014).
[6]S. Manzhos and T. Carrington, Jr., "A random-sampling high dimensional model representation neural network for building potential energy surfaces," J. Chem. Phys. **125**, 084109 (2006).
[7]S. Manzhos and T. Carrington, Jr., "Using neural networks to represent potential surfaces as sums of products," J. Chem. Phys. **125**, 194105 (2006).
[8]A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, and R. Komanduri, "Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks," J. Chem. Phys. **130**, 134101 (2009).
[9]B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces," J. Chem. Phys. **139**, 054112 (2013).
[10]H. T. T. Nguyen and H. M. Le, "Modified feed-forward neural network structures and combined-function-derivative approximations incorporating exchange symmetry for potential energy surface fitting," J. Phys. Chem. A **116**, 4629–4638 (2012).
[11]G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, O. A. von Lilienfeld, and K.-R. Müller, "Learning invariant representations of molecules for atomization energy prediction," in *Advances in Neural Information Processing Systems*, edited by P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger (2012), Vol. 25, pp. 449–457.
[12]Q. Meng, J. Chen, and D. H. Zhang, "Communication: Rate coefficients of the H + CH_4 → H_2 + CH_3 reaction from ring polymer molecular dynamics on a highly accurate potential energy surface," J. Chem. Phys. **143**, 101102 (2015).
[13]S. Houlding, S. Y. Liem, and P. L. A. Popelier, "A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer," Int. J. Quantum Chem. **107**, 2817–2827 (2007).
[14]J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks **61**, 85–117 (2015).
[15]C. M. Handley and P. L. A. Popelier, "Potential energy surfaces fitted by artificial neural networks," J. Phys. Chem. A **114**, 3371–3383 (2010).
[16]J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," Phys. Chem. Chem. Phys. **13**, 17930–17955 (2011).
[17]X. He, T. Zhu, X. Wang, J. Liu, and J. Z. H. Zhang, "Fragment quantum mechanical calculation of proteins and its applications," Acc. Chem. Res. **47**, 2748–2757 (2014).
[18]J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, "Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential," Phys. Rev. Lett. **100**, 185501 (2008).
[19]N. Artrith and J. Behler, "High-dimensional neural network potentials for metal surfaces: A prototype study for copper," Phys. Rev. B **85**, 045439 (2012).
[20]N. Artrith, B. Hiller, and J. Behler, "Neural network potentials for metals and oxides — First applications to copper clusters at zinc oxide," Phys. Status Solidi B **250**, 1191–1203 (2013).

[21]T. Morawietz and J. Behler, "A density-functional theory-based neural network potential for water clusters including van der Waals corrections," J. Phys. Chem. A **117**, 7356 (2013).
[22]T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How Van der Waals Interactions Determine the Unique Properties of Water" (submitted).
[23]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: A route to accurate calculations on large systems," Chem. Rev. **112**, 632–672 (2012).
[24]M. A. Collins and R. P. A. Bettens, "Energy-based molecular fragmentation methods," Chem. Rev. **115**, 5607–5642 (2015).
[25]M. A. Collins and V. A. Deev, "Accuracy and efficiency of electronic energies from systematic molecular fragmentation," J. Chem. Phys. **125**, 104104 (2006).
[26]H. M. Netzloff and M. A. Collins, "*Ab initio* energies of nonconducting crystals by systematic fragmentation," J. Chem. Phys. **127**, 134113 (2007).
[27]M. A. Collins, "Systematic fragmentation of large molecules by annihilation," Phys. Chem. Chem. Phys. **14**, 7744–7751 (2012).
[28]M. A. Collins, M. W. Cvitkovic, and R. P. A. Bettens, "The combined fragmentation and systematic molecular fragmentation methods," Acc. Chem. Res. **47**, 2776–2785 (2014).
[29]M. A. Addicoat and M. A. Collins, "Accurate treatment of nonbonded interactions within systematic molecular fragmentation," J. Chem. Phys. **131**, 104103 (2009).
[30]J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," J. Chem. Phys. **134**, 074106 (2011).
[31]C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. (Springer, New York, 2006).
[32]J. Behler, "Constructing high-dimensional neural network potentials: A tutorial review," Int. J. Quantum Chem. **115**, 1032–1050 (2015).
[33]L. Bottou, "Stochastic gradient tricks," in *Neural Networks, Tricks of the Trade, Reloaded*, Lecture Notes in Computer Science, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, 2012), pp. 430–445.
[34]K. Levenberg, "A method for the solution of certain problems in least squares," Q. Appl. Math. **2**, 164–168 (1944).
[35]D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," SIAM J. Appl. Math. **11**, 431–441 (1963).
[36]S. Shah, F. Palmieri, and M. Datum, "Optimal filtering algorithms for fast learning in feedforward neural networks," Neural Networks **5**, 779–787 (1992).
[37]R. E. Kalman, "A new approach to linear filtering and prediction problems," J. Fluids Eng. **82**, 35–45 (1960).
[38]M. Gastegger and P. Marquetand, "High-dimensional neural network potentials for organic reactions and an improved training algorithm," J. Chem. Theory Comput. **11**, 2187–2198 (2015).
[39]F. Neese, "The ORCA program system," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73–78 (2012).
[40]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023 (1989).
[41]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, "Auxiliary basis sets to approximate Coulomb potentials," Chem. Phys. Lett. **240**, 283–290 (1995).
[42]O. Vahtras, J. Almlöf, and M. W. Feyereisen, "Integral approximations for LCAO-SCF calculations," Chem. Phys. Lett. **213**, 514–518 (1993).
[43]F. Neese, F. Wennmohs, A. Hansen, and U. Becker, "Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange," Chem. Phys. **356**, 98–109 (2009).
[44]R. Izsák and F. Neese, "An overlap fitted chain of spheres exchange method," J. Chem. Phys. **135**, 144105 (2011).
[45]K. A. Peterson, T. B. Adler, and H.-J. Werner, "Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, BNe, and AlAr," J. Chem. Phys. **128**, 084102 (2008).
[46]K. E. Yousaf and K. A. Peterson, "Optimized auxiliary basis sets for explicitly correlated methods," J. Chem. Phys. **129**, 184108 (2008).
[47]J. Behler, *RuNNer*—A program for constructing high-dimensional neural network potentials, Ruhr-Universität Bochum, 2007-2016.
[48]D. H. Nguyen and B. Widrow, "Neural networks for self-learning control systems," IEEE Control Syst. Mag. **10**, 18–23 (1990).
[49]See supplementary material at http://dx.doi.org/10.1063/1.4950815 for a listing of the symmetry functions and their respective parameters used to describe the local chemical environments in the present work.

APPENDIX A.3 MACHINE LEARNING MOLECULAR DYNAMICS FOR THE SIMULATION OF INFRARED SPECTRA

MICHAEL GASTEGGER, JÖRG BEHLER, AND PHILIPP MARQUETAND

Manuscript submitted to *Chemical Science* (19th May 2017).

Contributions:

MICHAEL GASTEGGER conceived and implemented the dipole moment model and adaptive sampling scheme, performed and analyzed the test simulations and contributed to the initial draft and final version of the manuscript.

JÖRG BEHLER provided the RUNNER program, and contributed to the final manuscript.

PHILIPP MARQUETAND conceived the scope of the manuscript, supervised the methodological developments, exemplary computations and data analysis, and contributed to the final manuscript.

Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra

Michael Gastegger,[a] Jörg Behler,[b] and Philipp Marquetand[a*]

Abstract

Machine learning has emerged as an invaluable tool in many research areas. In the present work, we harness this power to predict highly accurate molecular infrared spectra with unprecedented computational efficiency. To account for vibrational anharmonic and dynamical effects – typically neglected by conventional quantum chemistry approaches – we base our machine learning strategy on *ab initio* molecular dynamics simulations. While these simulations are usually extremely time consuming even for small molecules, we overcome these limitations by leveraging the power of a variety of machine learning techniques, not only accelerating simulations by several orders of magnitude, but also greatly extending the size of systems that can be treated. To this end, we develop a molecular dipole moment model based on environment dependent neural network charges and combine it with the neural network potentials of Behler and Parrinello. Contrary to the prevalent big data philosophy, we are able to obtain very accurate machine learning models for the prediction of infrared spectra based on only a few hundreds of electronic structure reference points. This is made possible through the introduction of a fully automated sampling scheme and the use of molecular forces during neural network potential training. We demonstrate the power of our machine learning approach by applying it to model the infrared spectra of a methanol molecule, n-alkanes containing up to 200 atoms and the protonated alanine tripeptide, which at the same time represents the first application of machine learning techniques to simulate the dynamics of a peptide. In all these case studies we find excellent agreement between the infrared spectra predicted via machine learning models and the respective theoretical and experimental spectra.

1 Introduction

Machine learning (ML) – the science of autonomously learning complex relationships from data – has experienced an immensely successful renaissance during the last decade.[1,2] Increasingly powerful ML algorithms form the basis of a wealth of fascinating applications, with image and speech recognition, search engines or even self-driving cars being only a few examples. In a similar manner, ML based techniques have lead to several exciting developments in the field theoretical chemistry.[3–6]

ML potentials are an excellent example for the benefits ML algorithms can offer when paired with theoretical chemistry methods.[7–13] These potentials aim to accurately reproduce the potential energy surface (PES) of a chemical system (and its forces) based on a number of data points computed with quantum chemistry methods. Due to the powerful non-linear learning machines at their core, ML potentials are able to retain the accuracy of the underlying quantum chemical method, but are several orders of magnitude faster to evaluate. This combination of speed and accuracy is especially advantageous in situations where a large number of costly quantum chemical calculations would be required.

One such case is *ab initio* molecular dynamics (AIMD), a simulation technique used to describe the evolution of chemical systems with time.[14] In AIMD, the motion of the nuclei is described classically according to Newton's equations of motion[15] and depends on the quantum mechanical force exerted by the electrons and nuclei. AIMD is a highly versatile tool and has been used to model a variety of phenomena like photodynamical processes or the vibrational spectra of molecules.[16–20]

The latter application is of particular interest in the field of vibrational spectroscopy. With the development of more and more sophisticated experimental techniques, it is now possible to use methods like infrared (IR) and Raman spectroscopy to obtain highly accurate spectra of macromolecular systems (e.g. proteins).[21,22] As a consequence, vibrational spectra have become increasingly complex and theoretical chemistry simulations are now an indispensable aid in their interpretation. Unfortunately, the standard approach to model vibrational spectra, static calculations based on the harmonic oscillator (HO) approximation, suffers from several inherent limitations.[18,23] Due to the HO approximation, anharmonic vibrational effects are neglected, which are of great importance in molecular systems with high degrees of flexibility and/or hydrogen bonding, such as biological systems. Moreover, HO based calculations are unable to account for conformational and dynamic effects, due to their restriction to one particular conformer. This also makes it hard to accurately model temperature effects, which have a large influence on conformational dynamics and are highly

[a] *University of Vienna, Faculty of Chemistry, Department of Theoretical Chemistry, Währinger Str. 17, 1090 Vienna, Austria.*
[b] *Universität Göttingen, Insititut für Physikalische Chemie, Theoretische Chemie, Tammanstr. 6, 37077 Göttingen, Germany.*
[*] *E-mail: philipp.marquetand@univie.ac.at*

relevant for spectra recorded at room temperature.[17] These deficiencies lead to disagreements between experimental and theoretical spectra, thus complicating a consistent analysis.

Different strategies, like the variational self-consistent field (VSCF) approach and its extensions[23], as well as quantum dynamics based methods[24,25], have been developed to account for these effects, but they either neglect dynamical effects or are computationally intractable for systems containing more than a few tens of atoms. Consequently, AIMD, which is able to describe anharmonicities, dynamic effects at manageable computational costs, is an invaluable tool for the practical simulation of vibrational spectra.[17,18]

Yet, standard AIMD is still comparatively expensive, placing severe restrictions on the maximum size of the systems under investigation (approximately 100 atoms) and on the quality of the quantum chemical method. However, AIMD simulations can be accelerated significantly without sacrificing chemical accuracy by replacing the individual electronic structure calculations with much cheaper ML computations. This opens the way for exciting new possibilities, making it possible to simulate larger systems and longer timescales in only a fraction of the original computer time.

The goal of the current work is to use ML accelerated AIMD calculations to simulate accurate IR spectra of different organic molecules. This is achieved by harnessing the synergies between established techniques, improvements to existing schemes and new developments: (I) A special kind of ML potential, called high-dimensional neural network potential (HDNNP), is used to model the PES.[26] (II) Molecular forces are employed in the construction of these HDNNPs, using a novel method based on the element decoupled Kalman filter.[27] (III) electronic structure reference data points are selected via an enhanced adaptive sampling scheme for molecular systems. (IV) A HDNNP based fragmentation method is used to accelerate reference computations for macromolecules.[28] Finally, (V) a new ML scheme to model dipole moments is introduced. A detailed description of all these individual components is given in the following section.

Three different molecular systems are studied using the strategies described above. First, a single methanol molecule serves as a test case to assess the overall accuracy of the HDNNP based simulations compared to spectra obtained with standard AIMD. Second, the ability of HDNNPs to efficiently deal with macromolecular systems is demonstrated by (a) constructing a HDNNP of a simple alkane chain based only on small fragments of the macromolecule and (b) then using the resulting model to predict the IR spectra of alkanes of varying chain lengths. In order to probe the suitability of HDNNPs for systems of biological relevance, a final study is dedicated to the protonated trialanine peptide. This also serves as an excellent test case for the ML based dipole moment model.

All HDNNPs are constructed using density functional theory (DFT) as electronic structure reference method. Generalized gradient functionals are used in for methanol and the tripeptide. In the case of alkanes, we demonstrate that in principle also highly accurate double-hybrid density functionals[29] can be used. The simulations carried out with these latter HDNNPs would be next to impossible using on-the-fly AIMD. In all cases, comparisons to experimental IR spectra are shown.

2 Theoretical Background

In AIMD, vibrational spectra are computed via the Fourier transform of time autocorrelation functions.[18] Different physical properties give rise to different types of spectra. IR spectra depend on the molecular dipole moments:

$$I_{IR} \propto \int_{-\infty}^{+\infty} \langle \dot{\mu}(\tau)\dot{\mu}(\tau+t)\rangle_\tau \; e^{-i\omega t} \mathrm{d}t, \qquad (1)$$

where $\dot{\mu}$ is the time derivative of the molecular dipole moment, ω is the vibrational frequency, τ is a time lag and t is the time.

Upon closer examination of Equation 1, several challenges to model AIMD quality IR spectra via ML become apparent: Reliable ML potentials (and especially forces) are required to describe the time evolution of a chemical system. Consequently, electronic structure reference points need to be selected from representative regions of the PES, while keeping the number of costly electronic structure calculations to a minimum. This also calls for efficient strategies to handle the reference calculations of large molecules. And finally, a method to accurately model molecular dipole moments is required.

2.1 High-Dimensional Neural Network Potentials.

In a HDNNP (shown in Figure 1), the total potential energy E_{pot} of a molecule is expressed as a sum of individual atomic energies.[26,30] The contribution E_i of every atom depends on its local chemical environment and is modeled by a neural network (NN). These atomic NNs are typically constrained to be the same for a given element and thus, also termed elemental NNs. Due to this unique structure, HDNNPs can easily adapt to molecules of different size and even be transferred between sufficiently similar molecular systems.

The chemical environment of an atom is represented by a set of many-body symmetry functions $\{G_i\}$, so-called atom-centered symmetry functions (ACSFs).[31] ACSFs depend on the positions $\{R_i\}$ of all neighboring atoms around the central atom, up to a predefined cutoff radius. By introducing a cutoff radius, an atom's environment is restricted to the chemically relevant regions. This brings two distinct advantages: the computational cost of HDNNPs now scales linearly with molecular size and chemical locality can be exploited in their construction and application[7], which has been demonstrated recently e.g. for alkanes[28]. In addition, HDNNPs are well suited for molecular dynamics simulations, since an analytic expression for molecular forces is

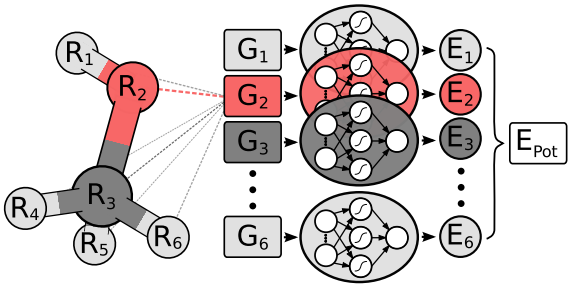


Figure 1: Schematic representation of a high-dimensional neural network potential (HDNNP). The Cartesian coordinates $\mathbf{R}$ are transformed into many-body symmetry functions $\{G_i\}$ describing an atom's chemical environment. Based on these functions, a NN then predicts the energy contribution E_i associated with atom i. The potential energy E_{Pot} of the whole molecule is obtained by summing over all individual atomic energies.

available due to their well-defined functional form. For a detailed discussion of HDNNPs and ACSFs, see reference [30].

In order for HDNNPs to yield reliable models of the PES, a set of optimal parameters needs to be determined for the elemental NNs. This is done in a process called training, where a cost function (typically the mean squared error) between reference data points (e.g. energies and forces) and the HDNNP predictions is minimized iteratively. Different algorithms can be used to carry out the minimisation. The current work uses the element-decoupled Kalman filter [27], a special adaptation of the global extended Kalman filter [32] for HDNNPs.

Besides the energies, it is also possible to include molecular forces in the training process, by minimizing the cost function [30]

$$\mathcal{C}_{\mathrm{E,F}} = \frac{1}{M}\sum_{m}^{M}\left(\tilde{E}_m - E_m\right)^2 + \frac{\eta}{M}\sum_{m}^{M}\frac{1}{3N_m}\sum_{\alpha}^{3N_m}\left(\tilde{F}_{m\alpha} - F_{m\alpha}\right)^2. \tag{2}$$

The first term on the right hand side corresponds to the mean squared error between reference energies E and HDNNP energies $\tilde{E}$. The second term describes the deviation between HDNNP ($\tilde{F}$) and quantum chemical forces (F). M is the number of molecules in the reference data set, N the number of atoms in a molecule, and α is an index running over the $3N$ Cartesian force components. η is a constant used to tune the importance of the force error on the update step. Including the forces in the training process leads to substantial improvements in the forces predicted by the HDNNP. Furthermore, instead of only one single energy, $3N$ points of additional information per molecule can now be utilized during training, thus greatly reducing the number of reference points required for a converged potential. An in-depth description of the element-decoupled Kalman filter and its extension to molecular forces can be found in reference [27].

2.2 Adaptive Selection Scheme.

Ultimately, the quality of a ML potential does not only depend on the underlying ML algorithm and the employed training procedure, but also on how well the reference data set represents the chemical problem under investigation. Ideally, the reference data spans all relevant regions of the PES with as few data points as possible to avoid costly electronic structure computations. To this end, different strategies – e.g. based on Bayesian inference [33] or geometric fingerprints [34] – have been developed in the past.

A simple, but relatively effective procedure to select data points is based on the use of multiple HDNNPs and is described for example in reference [30]. After choosing an initial set of reference data points, a set of preliminary HDNNPs is trained, differing in the initial parameters and/or architectures of their elemental NNs (Figure 2). These protopotentials are then used sample different molecular conformations, using e.g. molecular dynamics simulations. Afterwards, the predictions of the HDNNPs are compared to each other. Regions of the PES, where the different HDNNPs agree closely are assumed to be represented well, whereas conformations with diverging HDNNP predictions are modeled inaccurately. The inaccurately described conformations are recomputed with the electronic structure reference method and added to the reference data set. The HDNNPs are then retrained using the expanded data set and the process is repeated in a self consistent manner until the HDNNPs reach the desired quality.

The current work introduces small adaptations to this procedure in order to make it more suitable for the use with biomolecules and expensive electronic structure reference methods. Instead of performing independent sampling simulations with the individual HDNNPs, they are instead combined into an ensemble. In the ensemble, energy and forces are computed as the average of the J different HDNNP predictions:

$$\overline{E} = \tfrac{1}{J}\textstyle\sum_{j=1}^{J}\tilde{E}_j, \tag{3}$$

$$\overline{\mathbf{F}} = \tfrac{1}{J}\textstyle\sum_{j=1}^{J}\tilde{\mathbf{F}}_j. \tag{4}$$

Simulations are then carried out using these averaged properties. The prediction uncertainty of the HDNNP ensembles is defined as

$$E_\sigma = \sqrt{\frac{1}{J-1}\sum_{j}^{J}\left(\tilde{E}_j - \overline{E}\right)^2}. \tag{5}$$

Ensembles of HDNNPs are less susceptible to erratic behavior in their individual parts. Moreover, the error of ensemble methods is typically proportional to $\frac{1}{\sqrt{J}}$, leading to a significant improvement in accuracy at negligible extra cost. The combination of both effects leads to more reliable simulations, especially in the early stages of PES exploration, hence diminishing the number of electronic structure starting points needed to seed the self-consistent refinement procedure. As a consequence, HDNNPs can now be grown on

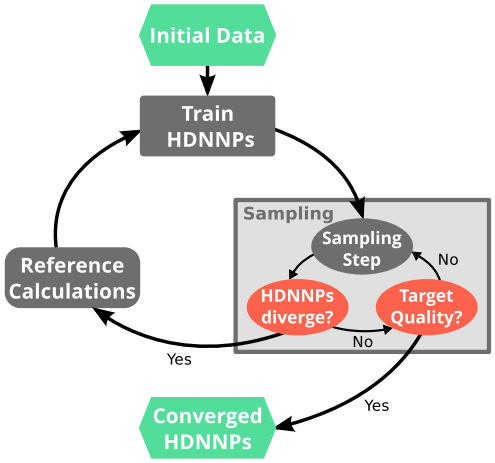


Figure 2: A typical run of the adaptive selection scheme starts by using a small set of initial reference data points to train a preliminary ensemble of HDNNPs. These HDNNPs are then used to sample new molecular conformations (e.g. via molecular dynamics simulations). During sampling, the predictions of the individual potentials are monitored and if divergence is detected, the sampling run is stopped. The conformation for which the HDNNPs disagree is computed with the electronic structure reference method and added to the set of reference points. Subsequently, the HDNNP ensemble is retrained on the expanded data set and sampling is continued with the new potential. This procedure is repeated in an iterative manner, until the divergence stops to exceed a predetermined threshold.

the fly from only a handful of data points in a highly automated manner: Starting from e.g. a few molecular dynamics steps, HDNNP ensemble simulations are run until E_σ of a visited structure exceeds a predefined threshold. The corresponding conformation is recomputed with the reference method and added to the training set. The HDNNPs are retrained and simulations are continued from the problematic conformation.

This procedure is effective, but highly sequential and calculations using expensive reference methods constitute a significant bottleneck. Under the assumption, that the approximate shape of PES is sufficiently similar for different electronic structure methods, an "upscaling" step is introduced. First, the iterative refinement is carried out using a low-level method until convergence of the HDNNPs. The conformations obtained in this manner are then recomputed using a high-level method. Since these high-level calculations can be done in parallel, the overall procedure is highly efficient. Afterwards, new HDNNPs are trained, now at the quality of the better method. The above assumption with regard to the similar shape of the PES at the different levels of theory is not necessarily valid, hence an upscaling step is typically followed by additional refinement steps at the higher level of theory.

A detailed discussion of the performance of the adaptive selection scheme and the convergence of the ML predictions with ensemble size can be found in the supporting information.

2.3 Fragmentation with High-Dimensional Neural Network Potentials.

Since the computational cost of electronic structure calculations scale very unfavorably with system size and the accuracy of the underlying method, individual reference computations can still be problematic. Hence, the required reference computations would quickly become intractable for highly accurate HDNNPs describing large molecular systems, despite the efficient sampling scheme.

It is possible to circumvent this problem by exploiting the special structure of HDNNPs. As a consequence of expressing the HDNNP energy as a sum of atomic contributions and introducing a cutoff radius, HDNNPs operate the same manner as fragmentation methods using a divide and conquer approach: Given only the energies of small molecular fragments, HDNNPs can reconstruct the energy of the total system.[7,28] Thus, expensive electronic structure calculations never have to be performed for the whole molecule, but only for small parts of it. The result is a linear scaling of the computational effort with system size.

In practice, a molecule is first divided into its individual fragments. Reference computations are then carried out for these fragments and the resulting data set is used to train a HDNNP. The ML potential is then applied to the geometry of the original molecule and the energy of the full system is recovered in this way. Different strategies can be used to partition the full molecular system. In the current work, every molecule is split into N atom-centered fragments (see Figure 3). The size and shape of these fragments are determined by a cutoff radius around the central atom. Atoms beyond the cutoff radius are removed and free valencies are saturated with hydrogen atoms. If a free valency is situated on a hydrogen atom or two capping hydrogens overlap, the heavy atom corresponding to this position is instead included in the fragment and the process is repeated iteratively. Typically, the same cutoff radius as in the ACSFs is used.

HDNNP fragmentation can easily be integrated into the adaptive sampling scheme. By using the deviations in atomic forces predicted by different HDNNPs as uncertainty measure, inaccurately modeled fragments can be identified. These fragments are then added to the reference data set.

2.4 Neural Network Dipole Moments and Charge Analysis.

A vital ingredient in the simulation of IR spectra with AIMD are molecular dipole moments (see Equation 1). While strategies to predict dipole moments using NNs exist[35], HDNNPs themselves have only been used to predict environment dependent charges in full analogy to the atomic energy

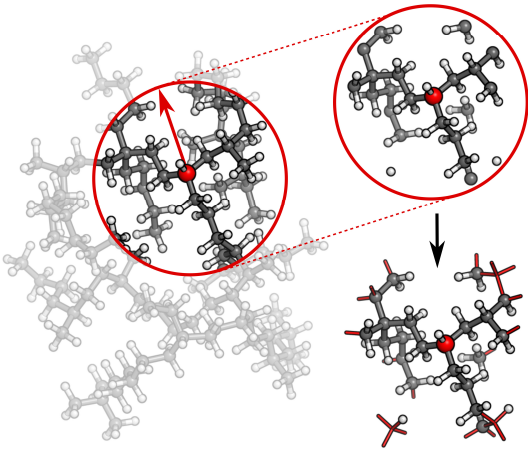


Figure 3: In order to generate molecular fragments, first all atoms beyond a predetermined cutoff radius from the central atom are removed. Afterwards, free valencies are saturated with hydrogen atoms, unless the valency itself is situated on a hydrogen or corresponds to a double bond in the unfragmented molecule. In this case, the heavy atom connected to this atom in the original molecule is included in the fragment and the process is repeated iteratively. This procedure is performed for the whole system, leading to one fragment per atom.

contributions with the aim to model electrostatic long range interactions.[36]

In this work, we extend this approach, by constructing molecular dipole moments as a sum of such environment dependent atomic partial charges:

$$\tilde{\mu} = \sum_{i}^{N} \tilde{q}_i \mathbf{r}_i, \tag{6}$$

where $\tilde{q}_i$ is the charge of atom i modeled by a NN and $\mathbf{r}_i$ is the distance vector of the atom from the molecule's center of mass.

While the elemental charge NNs could in principle be trained to reproduce charges computed with quantum chemical charge partitioning schemes (as was e.g. done in Reference[37] to model electrostatic interactions), this approach has the following problems: First, the charge of a given atom obtained with such a partitioning scheme can in principle change along a trajectory in a non-continuous manner. The resulting inconsistencies in the reference data can in turn lead to erratic predictions of the final model. Second, unlike molecular energies and forces, atomic partial charges are no quantum mechanical observable. Hence, there is no physically unique way to determine them and a variety of different partitioning schemes exists.[38] This complicates the choice of a suitable method to compute reference charges, since different schemes often exhibit vastly different behavior and sometimes fail to reproduce the molecular dipole moment accurately.[39]

Both problems can be avoided by training the elemental NNs to reproduce the molecular moments directly, while the environment dependent atomic charges $\tilde{q}_i$ are inferred in an indirect manner. In order to achieve this, a cost function of the form

$$\mathcal{C}_{\mathrm{Q}} = \frac{1}{M}\sum_{m}^{M}\left(\tilde{Q}_m - Q_m\right)^2 + \frac{1}{3M}\sum_{m}^{M}\sum_{l}^{3}\left(\tilde{\mu}_{lm} - \mu_{lm}\right)^2 + \ldots \tag{7}$$

is minimized. Here, Q_m and μ_{lm} are the reference total charge and dipole moment components of molecule m. The index l runs over the three Cartesian components of the dipole moment. $\tilde{Q}$ is the total charge of the composite NN model, computed as $\tilde{Q} = \sum_i^N \tilde{q}_i$, while $\tilde{\mu}$ is the NN dipole moment (Equation 6). While the cost function (from Equation 7) can be easily extended to include higher multipole moments, it was found that including only the total molecular charge and dipoles is sufficient for the purpose of modeling IR spectra. Since this scheme depends exclusively on molecular moments which are quantum mechanical observables, charge partitioning is no longer required. On the contrary, the trained NN model itself constitutes a new kind of partitioning scheme, where the atomic partial charges q_i depend on the chemical environment and are determined on a purely statistical basis. These charges can also be used for additional purposes, e.g. to compute electrostatic interactions. Another possible application would be to augment classical force fields[35], where partial charges typically do not change with the chemical environment.[40] As such, the NN charge scheme presented here constitutes an interesting alternative to static point charges or polarizable models.[41]

3 Computational Details

Electronic structure reference calculations were carried out with ORCA[42] at the BP86[43–47]/def2-SVP[48] (Methanol, Ala_3^+), BLYP[43–45,49]/def2-SVP (Ala_3^+) and B2PLYP[29]/def2-TZVPP[48] (n-alkanes) levels of theory. All calculations were accelerated using the resolution of identity approximation.[50,51]

All HDNNPs were constructed and trained with the RUNNER program.[52] The NN dipole models were implemented in `python`[53] using the `numpy`[54] and `theano`[55] packages. Reference data points were obtained with the adaptive selection scheme, employing molecular dynamics trajectories at a temperature of 500 K with a 0.5 fs timestep to sample relevant conformations. The final ML models are based on 245 (methanol), 534 (n-alkanes) and 718 (peptide) reference data points, with a maximum network size of 35-35-1 (two hidden layers with 35 nodes each and one node in the output layer) for the HDNNPs and 100-100-1 for the dipole moment model.

IR spectra were obtained with molecular dynamics simulations in the gas phase employing the same timestep as the sampling procedure. After a short initial equilibration period (3ps for methanol, 5ps otherwise), constant temperature molecular dynamics simulations were run for 30 ps

in the case of methanol and 50 ps for the other molecules. In addition to ML accelerated dynamics, AIMD simulations were carried out for methanol using the BP86 level of theory described above.

Detailed information regarding the setup of the electronic structure calculations and molecular dynamics simulations, as well as the ML models can be found in the supporting information.

4 Results and Discussion

4.1 Methanol.

Due to its small size, the methanol molecule constitutes an excellent test system, not only for the direct comparison between IR spectra obtained via standard AIMD and ML simulations, but also to investigate the overall accuracy of the ML approximations.

The final ML model for methanol consists of two HDNNPs and a NN dipole moment model trained on the BP86 data for 245 configurations. To assess the errors associated with the individual components of the model, a standard AIMD simulation is run 30 ps, producing 60 000 configurations. For the sampled geometries, energies, forces and dipoles are predicted with the ML model. These predictions are then compared to the respective electronic structure results. The distribution of errors between ML predictions and the BP86 method are shown in blue in Figure 4.

Excellent agreement between electronic structure calculations and the ML model is found for all investigated properties. In the case of energies (Figure 4a), the mean absolute error (MAE) of 0.048 kcal mol^{-1} (range of energies 13.620 kcal mol^{-1}) is well below the commonly accepted limit for chemical accuracy (1 kcal mol^{-1}) and is expected to be negligible compared to the intrinsic error of the electronic structure reference method in practical applications. The components of the force vectors are reproduced equally well (Figure 4b), with a MAE of 0.533 kcal mol^{-1} Å^{-1} (range 242.34 kcal mol^{-1} Å^{-1}). These findings are comparable with other state of the art ML learning strategies developed specifically for the modeling of forces [56] and demonstrate the excellent capabilities of HDNNPs to create potentials suitable for the dynamical simulation of molecules. This conclusion is also supported by a comparison of the normal mode frequencies obtained for the optimized methanol structure at the ML- and BP86-level (see Table 1). Although the HDNNP model was never explicitly trained to reproduce normal mode frequencies, its predictions agree well with the electronic structure frequencies, exhibiting a maximum deviation of only 31.38 cm^{-1} (0.090 kcal mol^{-1}). The NN dipole model is also found to provide an accurate description of the molecular dipole moments (Figure 4c). The total dipole moment shows an overall MAE of 0.016 D (over a range of 0.723 D) and the spatial orientation of the dipole vector is modeled equally reliable, with the MAEs of the individual Cartesian components ranging from 0.0173 D to 0.0200 D. The small shift of the dipole error distribution towards negative values is due to the fact that the atomic charges fluctuate around values other than zero. This effect is enhanced further, by the final summation to obtain the dipol moment model (see 6).

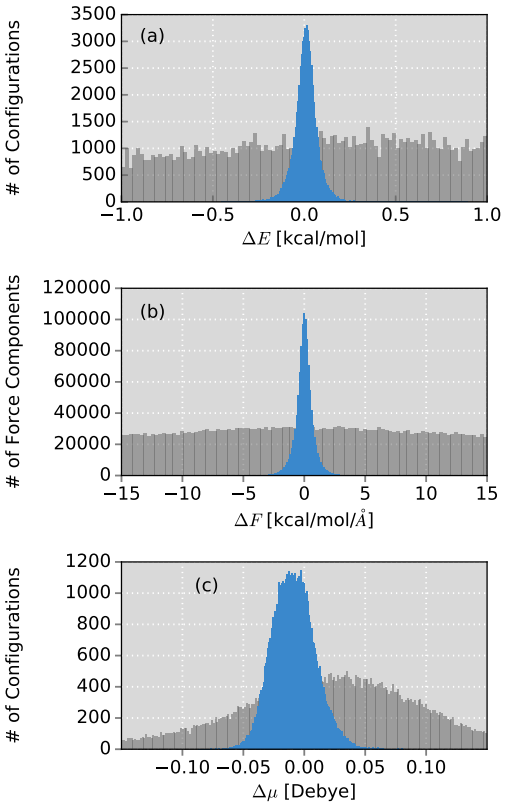


Figure 4: Distribution of errors between the ML model based on the adaptive sampling scheme and the BP86 reference (blue). The deviations were computed based on the electronic structure energies, forces and dipole moments (from top to bottom) of 60 000 configurations of methanol sampled with an AIMD simulation. The deviations obtained with a ML model trained on data points selected at random from a force field simulation are shown in grey (see supporting info).

In order to study the quality of the IR spectrum modeled with the composite ML model, it is compared directly to the spectrum obtained via the BP86 AIMD simulation. Figure 5 shows both IR spectra alongside an experimental spectrum of methanol recorded in the gas phase [57]. The overall shape of the ML spectrum, as well as the peak positions and intensities, show excellent agreement with the electronic structure reference. The most distinctive difference between QM and ML spectra is the intensity of the stretching vibration of the O-H bond observed at 3700 cm^{-1}. This relatively minor deviation is most likely caused by small deviations of the dipole moment model. Overall, the ML approach presented here is able to reproduce the AIMD IR spectrum of

Table 1: Comparison of the normal mode frequencies of methanol obtained with DFT and the ML model

#	DFT [cm^{-1}]	ML [cm^{-1}]	Δ [cm^{-1}]
1	331.70	346.94	-15.24
2	1037.82	1030.00	7.82
3	1080.46	1092.09	-11.63
4	1135.08	1138.21	-3.13
5	1328.95	1320.84	8.11
6	1420.02	1416.42	3.60
7	1427.64	1422.59	5.05
8	1449.79	1449.02	0.77
9	2880.76	2892.94	-12.18
10	2930.10	2961.48	-31.38
11	3034.15	3054.08	-19.93
12	3707.93	3707.73	0.20

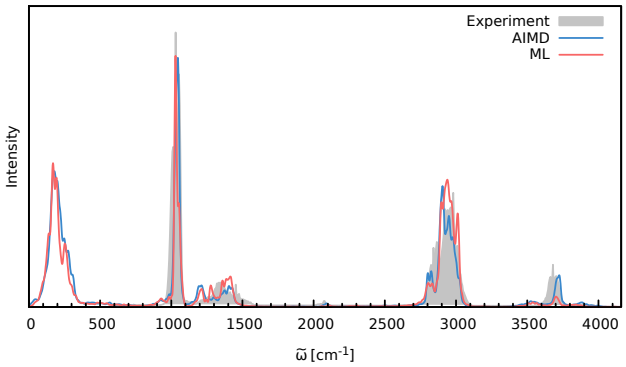


Figure 5: IR spectra of the methanol molecule. The ML spectrum (red) is able to reproduce the AIMD spectrum (blue) obtained with BP86 with high accuracy. In addition, both theoretical spectra agree well with the experimental one (grey).

methanol with high accuracy. These results are remarkable insofar, as the final ML model is based on only 245 electronic structure calculations. This demonstrates the effectiveness of the combination of HDNNPs and the NN dipole model, as well as the power of the improved sampling scheme.

Finally, both simulations agree well with experiment, serving as an example for the utility of AIMD and ML-accelerated AIMD for the prediction of accurate vibrational spectra.

4.2 n-Alkanes.

When constructing ML potentials for large molecular systems containing hundreds or thousands of atoms, the necessary electronic structure reference calculations can quickly become intractable, especially for high-level electronic structure methods. HDNNPs, as well as the dipole moment model presented in this work, can overcome this limitation via their implicit use of fragmentation (see Section 2.3). In order to demonstrate the potential of this approach, it is used to predict the IR spectrum of an n-alkane with the chemical formula $C_{69}H_{140}$ (depicted in Figure 6) via ML-accelerated AIMD simulations based on the B2PLYP double-hybrid density functional method.

The two HDNNPs and NN dipole moment model constituting the final ML model were trained on reference calculations for 534 fragments of the n-alkane. These fragments use a cutoff radius of 4.0 Å and contain 37 atoms on average and a maximum of 70 atoms. After initial adaptive sampling at the BP86/def2-SVP level, the final B2PLYP/TZVPP level ML-model is obtained via an upscaling step described in Section 2.2. Dispersion interactions, which are expected to play an important role in molecular systems of this size, are accounted for via a simple scheme: the HDNNPs are constructed from standard B2PLYP calculations and augmented with the empirical D3 dispersion correction using Becke–Johnson damping[58,59] in an *a posteriori* fashion.

The IR spectrum of the $C_{69}H_{140}$ n-Alkane predicted via ML is shown in Figure 6. It exhibits all spectroscopic features typical for simple hydrocarbons: The intense peak at 3000 cm^{-1} corresponds to symmetric and asymmetric C-H stretching vibrations. Deformations of the CH_2-groups give rise to the bands close to 1500 cm^{-1}, while the extremely weak signals in vicinity of 1000 cm^{-1} and 600 cm^{-1} are generated by C-C bond stretching and CH_2 rocking vibrations. Although the general shape and features of the IR spectrum

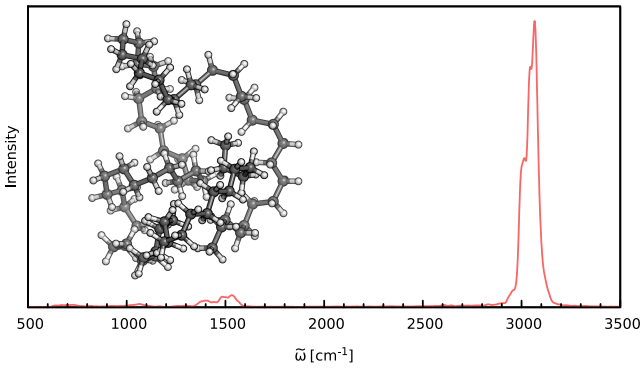


Figure 6: IR spectrum of the $C_{69}H_{140}$ alkane as predicted by the ML model based on the B2PLYP method.

are described well by the ML-model, some peak positions deviate from the expected experimental frequencies. This effect is especially pronounced for the C-H stretching vibrations, which are blue-shifted from the typical experimental value of 2900 cm^{-1} to 3040 cm^{-1}.

This blue shift is due to the electronic structure method (and not an artifact introduced by the ML approximations), as will be explained in the following. Direct AIMD simulations and even static frequency calculations are prohibitively expensive for the $C_{69}H_{140}$ molecule. Instead, we exploit the transferability of the combined HDNNP and dipole model and use it to simulate the IR spectrum of the much smaller n-butane, for which static theoretical and experimental spec-

tra can be obtained easily. Figure 7 shows the n-butane IR spectra obtained with ML-accelerated AIMD, static electronic structure calculations and experiment[57]. The strong

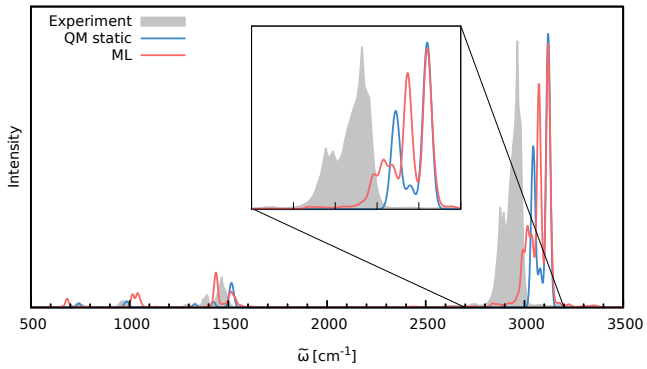


Figure 7: IR spectrum of n-Butane obtained via the ML model (red), compared to the static quantum mechanical spectrum computed at the B2PLYP level (blue) and convoluted with Gaussians. The peak positions in the ML and electronic structure spectra agree closely, suggesting that the observed deviations from experiment (grey) are due to the electronic structure method and not an artifact introduced by the ML approximation. The overall structure of the peaks is reproduced much better by the ML accelerated AIMD simulation, especially in the region of the C-H stretching vibrations (see insert).

blue shift of the C-H stretching vibrations present in the ML spectrum can also be found in the static electronic structure spectrum. Moreover, both spectra show good agreement with each other with respect to the overall positions of the spectral peaks. These findings support the conclusion, that the observed frequency shifts are indeed a consequence of the underlying electronic structure method and not an artifact of the ML approximation. Furthermore, the ML accelerated AIMD approach is found to accurately reproduce the structure of the experimental vibrational bands (especially the C-H stretching vibrations, see insert Figure 7). This is not the case for the static spectrum and shows, that even for relatively small molecules an accurate description of dynamic effects is important in order to obtain high-quality IR spectra. Both observations demonstrate the excellent accuracy of the HDNNP and NN dipole model, even for molecular systems not encountered during training.

Finally, to demonstrate the power the ML based approach in general and the fragmentation based approach in particular, a few exemplary timings are given for the $C_{69}H_{140}$ molecule (using a single core of an Intel Xeon E5-2650 v3 CPU): Obtaining the relevant molecular fragments using the iterative sampling scheme takes approximately 7 days. The reference calculations of the fragments on the B2PLYP level of theory can be carried out in a highly parallel manner within 1.2 days (using a single CPU per configuration), including the time necessary to construct the final ML model. ML-accelerated AIMD simulations for the $C_{69}H_{140}$ molecule which involve the calculation of 110 000 energies and forces (5ps equilibration and 50ps simulation) take 3 hours. The NN dipole moments can be obtained within half an hour. Including the generation of the model, the total time to obtain the ML based IR spectrum amounts to a little over 8 days. In contrast, the evaluation of a single energy and gradient at the B2PLYP level for the full n-alkane would require 30 days, extrapolating from the timings of the fragment reference calculations. Hence, performing the 110 000 calculations necessary for the AIMD simulation would require a total of 3.3 million days or 9 041 years.

4.3 Protonated Alanine Tripeptide.

Vibrational anharmonicities, as well as conformational and dynamic effects play a crucial role in the vibrational spectra of biomolecules. In order to investigate the ability of ML accelerated AIMD to account for these effects, the composite ML model is used to simulate the IR spectrum of the protonated alanine tripeptide molecule (Ala_3^+) in the gas phase. Modeling the Ala_3^+ molecule poses several challenges: An accurate description of the complicated PES depends crucially on the ability of the adaptive sampling scheme and the HDNNPs to reliably identify and interpolate relevant electronic structure data points. Moreover, the changing charge distribution and dipole moment of the protonated species need to be captured by the NN dipole model. Since the IR spectrum of Ala_3^+ has been studied extensively, both experimentally and theoretically[60,61], the quality of the ML approach can be assessed directly.

The composite Ala_3^+ ML model consists of two HDNNPs and a NN dipole model and was constructed from 658 reference geometries selected with the adaptive sampling scheme. The model exhibits overall RMSEs of 1.56 kcal mol^{-1}, 3.40 kcal mol^{-1} Å^{-1} and 0.26 Debye for energies, forces and dipoles respectively. This increase in the RMSEs and number of required data points compared to the previous systems is an indicator for the chemical complexity of the protein. Long range dispersion interactions were accounted for in the same manner as in the case of the n-alkanes.

Previous theoretical studies by Vaden and coworkers[61] have found, that the experimental IR spectrum of Ala_3^+ is primarily composed of the contributions of three different conformers: 1) An elongated Ala_3^+ chain with the proton situated at the N-terminal amine group, 2) a folded chain protonated at the same site and 3) a elongated form in which the proton is located a the carbonyl group of the N-terminus (see Figure 8), which will be referred to as the NH_3, folded and NH_2 families henceforth. In order to account for these effects, ML accelerated AIMD simulations were carried out for all three conformers at 350 K, the estimated experimental temperature. The final ML IR spectrum was then obtained by averaging. Figure 8 shows the overall spectrum, as well as the contributions of the individual conformations alongside the experimental spectrum.[61] Due to the range of the recorded spectrum and the high congestion of spectral

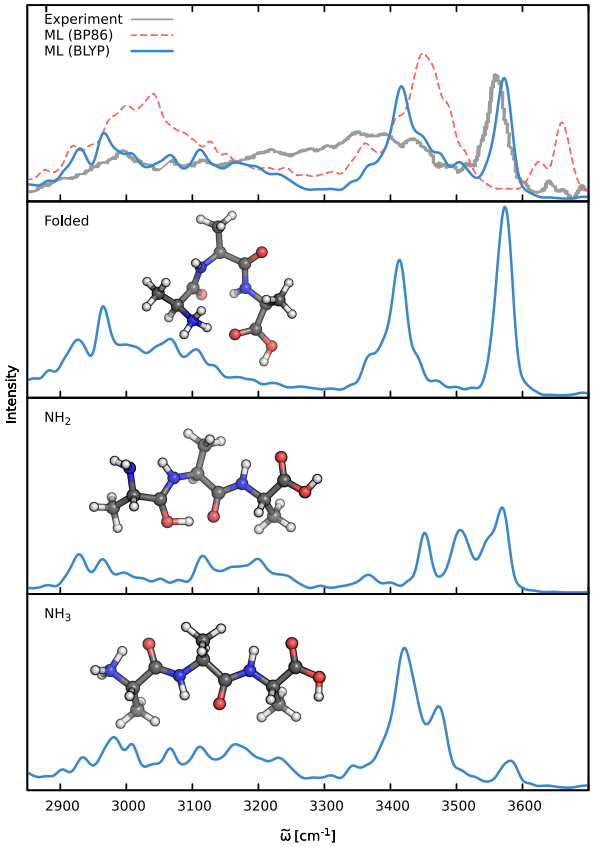


Figure 8: IR spectra of the protonated alanine tripeptide. The top panel shows the experimental spectrum (gray), as well as the ML spectra based on the BLYP (blue) and BP86 (red) reference methods. The lower panels depict the structures of the three main Ala_3^+ conformers, along with their respective contributions to the averaged BYLP ML spectrum.

bands in the regions of the lower vibrational modes, we restrict our discussion only to the stretching modes involving hydrogens (ca. 2700 cm^{-1} to 3700 cm^{-1}).

As can be seen, the ML model correctly captures the features present in the experimental spectrum. The intense peak at 3570 cm^{-1} is due to the O-H stretching vibrations of the carboxylic acid group of the C-terminus. The position as well as the slight asymmetry of this band are almost perfectly reproduced in the ML spectrum. The region from 3300 cm^{-1} to 3500 cm^{-1} is populated by signals arising from the stretching modes of N-H bonds not participating in hydrogen bonds (e.g. NH_2 terminus in the NH_2 family). The free N-terminal N-H groups of the NH_3 and folded family give rise to the intense feature at 3420 cm^{-1}. Vibrations associated with the N-H groups directly involved in hydrogen bonds are situated in the regions from 3100 cm^{-1} to 3300 cm^{-1}, where the ML spectrum captures several experimental subpeaks. Finally, the region from 2800 cm^{-1} to 3100 cm^{-1} corresponds to the C-H stretching vibrations. Here, the most distinct features are the peak at 2930 cm^{-1} due to C-H vibrations of the C_α groups and the peak at 2970 cm^{-1}, which is caused by the vibrations of the methyl group hydrogens. The generally good agreement between the ML and experimental spectrum and the ability to reliably resolve individual bands is a testament for the efficacy of the composite ML scheme introduced in this work: The dipole model is able to describe the charge distribution of Ala_3^+ accurately, while the HDNNP ensemble provides a reliable approximate PES.

A good perspective on the accuracy of the ML approach can also be gained by comparing the current ML model to one based on a different electronic structure reference method. The top panel of Figure 8 shows the averaged IR spectrum predicted by a ML model based on the BP86 density functional next to the previously discussed BLYP spectrum. Although one would expect the closely related BLYP and BP86 methods to give similar results, significant differences can be found: Besides a strong blue shift of the signal caused by the C-terminal COOH group by almost 80 cm^{-1}compared to the BLYP spectrum and experiment, large deviations are also found in the shape and positions of the bands corresponding to N-H stretching vibrations. Here, we investigate the cause of the latter effect by closer examination of the NH_3 conformer. Since the hydrogens of the N-terminal NH_3 group can be involved in a proton transfer event to the neighboring carbonyl group, different spectra can arise depending on how often this transfer occurs. The transfer rate is directly correlated to the energy barrier associated with the transfer, suggesting that BLYP and BP86 differ significantly in the description of this event, which in turn leads to differences in the ML spectra. Whether this phenomenon is caused by the ML approximations or due to the BP86 method itself, can easily be verified by computing the proton transfer barriers with both electronic structure methods and ML models. As can be seen in Figure 9, the barrier height is indeed underestimated by the BP86 functional compared to BLYP, giving rise to the observed behavior. At the same time, the ML models faithfully reproduce the barriers found with their respective electronic structure methods. This is an excellent demonstration for the reliability of the ML approach, since the deviation between ML model and reference method is actually negligible compared to the differences between two closely related electronic structure methods. The ease with which ML of different QM methods can be generated, also suggests a potential use of the ML approach presented here as an efficient tool for extensively comparing and thus benchmarking electronic structure methods. Additional ML models can simply be constructed by recomputing in a parallel fashion the representative conformations selected by the sampling scheme with a different method and subsequent retraining of the new model (see Section2.2). Possible applications of this finding will be explored in the future.

The above observations also serve to highlight the ability of the ML model to automatically infer the chemistry underlying the Ala_3^+ system. Proton transfer events are essen-

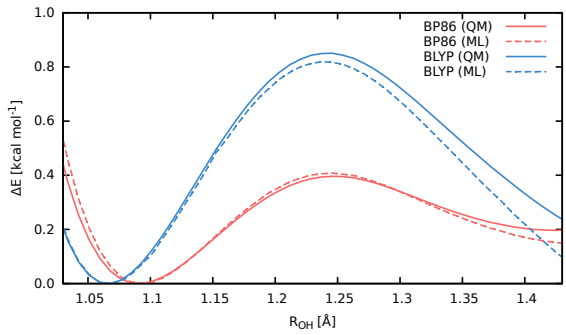


Figure 9: Reaction barriers associated with the proton transfer from the N-terminal NH_3 group in the NH_3 conformer of Ala_3^+ to the neighboring carbonyl. The reaction coordinate is the distance between the transferred NH_3 hydrogen and the carbonyl oxygen. The barriers computed with the electronic structure reference methods are shown as solid lines colored red for the BYLP method and blue in case of BP86. The dashed curves correspond to the predictions of the respective ML models, maintaining the above color scheme.

tial in characterizing the experimental spectrum.[60] Driven by the automated sampling scheme, the composite ML approach gradually learns to describe these relevant chemical events, as is nicely demonstrated based on the reaction barrier previously obtained for the NH_3 transfer (Figure 9): Although the description of this event was never explicitly targeted in the training procedure, the barrier is nevertheless reproduced to an excellent degree of accuracy. This feat is impressive insofar, as the ML model is based on an relatively small set of *ab intio* computations. These findings also serve to highlight an important advantage of HDNNPs over typical classical force fields, which is the ability to describe bond breaking and formation reactions.

Once again, the excellent computational efficiency of the composite ML model should be stressed: While the computational chemistry method employed for Ala_3^+ is already considered to be relatively cheap, the speedup gained is still significant. A single step in the BP86 simulation takes approximately 1.5 minutes (on a single Intel Xeon E5-2650 v3 CPU). The dynamics of every Ala_3^+ conformer are simulated for 55 ps, requiring a total of 110 000 steps. This amounts to a simulation time of 114 days for full AIMD. In contrast, using the ML model one can perform the same simulation in only one hour.

5 Conclusions

Here, we present the first application of machine learning (ML) techniques to the dynamical simulation of molecular infrared spectra. We find that our ML approach is able to predict infrared spectra of various chemical systems in a highly reliable manner, correctly describing anharmonicities, as well as dynamic effects, such as proton transfer events. The excellent accuracy – which is only limited by the underlying computational chemistry method – is paired with high computational efficiency, reducing the overall computation time by several orders of magnitude. This makes it possible to treat molecular systems usually beyond the scope of standard electronic structure methods. As a proof of principle, we have simulated n-alkanes containing several hundreds of atoms, as well as the protonated alanine tripetide. However, even larger systems can in principle be handled easily by our ML approach. To realize the above simulations, we combined neural network potentials (NNPs) of the Behler–Parrinello type[26] with a newly developed ML model for molecular dipole moments. This neural network based model constitutes a new form of charge partitioning scheme based purely on statistical principles and offers access to environment dependent atomic charges. For the efficient selection of electronic structure data points, a new adaptive sampling scheme is introduced. By employing this scheme, it is possible to incrementally grow ML potentials for specific applications in a highly automated manner based on only a small initial seed of reference data. When combined with the ability of NNPs to include molecular forces in their training procedure, the amount of electronic structure data points required to construct a ML potential is reduced tremendously (e.g. 700 conformations are sufficient for a converged potential of the tripeptide). Furthermore, we demonstrate the ability of NNPs to model macromolecules based only on the information contained in small fragments, making it possible to treat even these systems with highly accurate electronic structure methods in a divide and conquer fashion. The above findings are not only restricted to the simulation of infrared spectra via dynamics simulations, but apply to ML potentials in a broader sense. The ML approach presented here thus constitutes an alternative to the currently prevailing trend of fitting potentials to more and more reference data points. The latter strategy suffers from the disadvantage, that electronic structure reference calculations become prohibitively expensive for highly accurate methods and/or large molecular systems. Here we show that these problems can be overcome through the efficient use of data, bringing the dream of simulating the dynamics of e.g. enzymatic reactions with highly accurate methods one step closer.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 1st edn, 2006.

[2] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[3] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller and A. Tkatchenko, *Nat. Commun.*, 2017, **8**, 13890.

[4] J. N. Wei, D. Duvenaud and A. Aspuru Guzik, *ACS Central Science*, 2016, **2**, 725–732.

[5] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *arXiv:1702.05532*, 2017.

[6] R. Gómez Bombarelli, J. Aguilera Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. Aspuru Guzik, *Nat. Mater.*, 2016, **15**, 1120–1127.

[7] J. Behler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930–17955.

[8] C. M. Handley and P. L. A. Popelier, *J. Phys. Chem. A*, 2010, **114**, 3371–3383.

[9] J. Behler, *J. Phys. Condens. Matter*, 2014, **26**, 183001.

[10] J. Behler, *J. Chem. Phys.*, 2016, **145**, 170901.

[11] V. Botu, R. Batra, J. Chapman and R. Ramprasad, *J. Phys. Chem. C*, 2017, **121**, 511–522.

[12] A. P. Bartók and G. Csányi, *Int. J. Quantum Chem.*, 2015, **115**, 1051–1057.

[13] Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 096405.

[14] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge, Reprint edn, 2012.

[15] I. Newton, *Philosophiae naturalis principia mathematica*, J. Societatis Regiae ac Typis J. Streater, 1687.

[16] M. Barbatti, *WIREs Comput. Mol. Sci.*, 2011, **1**, 620–633.

[17] M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.

[18] M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.

[19] S. Mai, P. Marquetand and L. González, *Int. J. Quantum Chem.*, 2015, **115**, 1215–1231.

[20] P. Marquetand, J. Nogueira, S. Mai, F. Plasser and L. González, *Molecules*, 2016, **22**, 49.

[21] J. Simons, *Mol. Phys.*, 2009, **107**, 2435–2458.

[22] de Vries. Mattanjah S. and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585–612.

[23] T. K. Roy and R. B. Gerber, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9468–9492.

[24] H.-D. Meyer, *WIREs Comput. Mol. Sci.*, 2012, **2**, 351–374.

[25] P. S. Thomas and T. Carrington Jr., *J. Phys. Chem. A*, 2015, **119**, 13074–13091.

[26] J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

[27] M. Gastegger and P. Marquetand, *J. Chem. Theory Comput.*, 2015, **11**, 2187–2198.

[28] M. Gastegger, C. Kauffmann, J. Behler and P. Marquetand, *J. Chem. Phys.*, 2016, **144**, 194110.

[29] S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.

[30] J. Behler, *Int. J. Quantum Chem.*, 2015, **115**, 1032–1050.

[31] J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.

[32] T. B. Blank and S. D. Brown, *J. Chemom.*, 1994, **8**, 391–407.

[33] J. Li, B. Jiang and H. Guo, *J. Chem. Phys.*, 2013, **139**, 204103.

[34] V. Botu and R. Ramprasad, *Int. J. Quantum Chem.*, 2015, **115**, 1074–1083.

[35] M. G. Darley, C. M. Handley and P. L. A. Popelier, *J. Chem. Theory Comput.*, 2008, **4**, 1435–1448.

[36] N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B*, 2011, **83**, 153101.

[37] T. Morawietz, V. Sharma and J. Behler, *J. Chem. Phys.*, 2012, **136**, 064103.

[38] S. M. Bachrach, in *Population Analysis and Electron Densities from Quantum Mechanics*, John Wiley & Sons, Inc., 2007, pp. 171–228.

[39] K. B. Wiberg and P. R. Rablen, *J. Comput. Chem.*, 1993, **14**, 1504–1518.

[40] A. D. Mackerell, *J. Comput. Chem.*, 2004, **25**, 1584–1604.

[41] C. M. Baker, *WIREs Comput. Mol. Sci.*, 2015, **5**, 241–254.

[42] F. Neese, *WIREs Comput. Mol. Sci.*, 2012, **2**, 73–78.

[43] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

[44] P. A. M. Dirac, *Proc. R. Soc. A*, 1929, **123**, 714–733.

[45] J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.

[46] S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

[47] J. C. Slater, *Phys. Rev.*, 1951, **81**, 385–390.

[48] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

[49] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

[50] K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283–290.

[51] O. Vahtras, J. Almlöf and M. W. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514–518.

[52] J. Behler, *RuNNer – A program for constructing high-dimensional neural network potentials*, 2017, Universiät Göttingen.

[53] G. van Rossum and F. L. Drake (eds), *Python Reference Manual*, PythonLabs, Virginia, USA, 2001; Available at `http://www.python.org` (accessed date 06.04.2017).

[54] S. van der Walt, S. C. Colbert and G. Varoquaux, *Comput. Sci. Eng.*, 2011, **13**, 22–30.

[55] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde Farley and Y. Bengio, Proceedings of the Python for Scientific Computing Conference (SciPy), 2010.

[56] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt and K.-R. Müller, *Sci. Adv.*, 2017, In press.

[57] P. Chu, F. Guenther, G. Rhoderick and W. Lafferty, in *NIST Chemistry WebBook NIST Standard Reference Database Number 69*, ed. P. Linstrom and W. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, doi:10.18434/T4D303, (retrieved April 24, 2017).

[58] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[59] S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

[60] A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek and M.-P. Gaigeot, *J. Chem. Theory Comput.*, 2009, **5**, 1068–1078.

[61] T. D. Vaden, T. S. J. A. de Boer, J. P. Simons, L. C. Snoek, S. Suhai and B. Paizs, *J. Phys. Chem. A*, 2008, **112**, 4608–4616.

BIBLIOGRAPHY

[1] I. Goodfellow, Y. Bengio, A. Courville: *Deep Learning*, MIT Press (2016), `http://www.deeplearningbook.org`.

[2] C. M. Bishop: *Pattern Recognition and Machine Learning*, Springer, New York, 1st edition (2006).

[3] G. B. Goh, N. O. Hodas, A. Vishnu: Deep learning for computational chemistry, *J. Comput. Chem.*, **38**, 1291 (2017).

[4] J. B. O. Mitchell: Machine learning methods in chemoinformatics, *WIREs Comput. Mol. Sci.*, **4**, 468 (2014).

[5] R. Ramakrishnan, A. O. von Lilienfeld: *Reviews in Computational Chemistry*, chapter Machine Learning, Quantum Mechanics, and Chemical Compound Space, 225–256, John Wiley & Sons, Inc. (2017).

[6] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik: Deep neural nets as a method for quantitative structure–activity relationships, *J. Chem. Inf. Model.*, **55**, 263 (2015).

[7] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist: Machine-learning-assisted materials discovery using failed experiments, *Nature*, **533**, 73 (2016).

[8] E. O. Pyzer-Knapp, K. Li, A. Aspuru-Guzik: Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery, *Adv. Funct. Mater.*, **25**, 6495 (2015).

[9] J. N. Wei, D. Duvenaud, A. Aspuru Guzik: Neural Networks for the Prediction of Organic Chemistry Reactions, *ACS Central Science*, **2**, 725 (2016).

[10] I. N. Levine: *Quantum Chemistry*, Prentice Hall, Boston, 7th edition (2013).

[11] B. M. Austin, D. Y. Zubarev, W. A. Lester: Quantum monte carlo and related approaches, *Chem. Rev.*, **112**, 263 (2012).

[12] J. Čížek: On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, *J. Chem. Phys.*, **45**, 4256 (1966).

[13] W. Kohn, A. D. Becke, R. G. Parr: Density Functional Theory of Electronic Structure, *J. Phys. Chem.*, **100**, 12974 (1996).

[14] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus: CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.*, **4**, 187 (1983).

[15] A. D. Mackerell: Empirical force fields for biological macromolecules: Overview and issues, *J. Comput. Chem.*, **25**, 1584 (2004).

[16] S. Manzhos, T. Carrington, Jr.: An improved neural network method for solving the Schrödinger equation, *Can. J. Chem.*, **87**, 864 (2009).

[17] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke: Finding density functionals with machine learning, *Phys. Rev. Lett.*, **108**, 253002 (2012).

[18] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld: Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach, *J. Chem. Theory Comput.*, **11**, 2087 (2015).

[19] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, A. O. von Lilienfeld, K.-R. Müller: Learning Invariant Representations of Molecules for Atomization Energy Prediction, in *Advances in Neural Information Processing Systems 25*, 449–457 (2012).

[20] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko: Quantum-chemical insights from deep tensor neural networks, *Nat. Commun.*, **8**, 13890 (2017).

[21] E. Tafeit, W. Estelberger, R. Horejsi, R. Moeller, K. Oettl, K. Vrecko, G. Reibnegger: Neural networks as a tool for compact representation of *ab initio* molecular potential energy surfaces, *J. Mol. Graphics Modell.*, **14**, 12 (1996).

[22] J. Behler: Perspective: Machine learning potentials for atomistic simulations, *J. Chem. Phys.*, **145**, 170901 (2016).

[23] J. Behler: Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, *Phys. Chem. Chem. Phys.*, **13**, 17930 (2011).

[24] C. M. Handley, P. L. A. Popelier: Potential Energy Surfaces Fitted by Artificial Neural Networks, *J. Phys. Chem. A*, **114**, 3371 (2010).

[25] V. Botu, R. Batra, J. Chapman, R. Ramprasad: Machine Learning Force Fields: Construction, Validation, and Outlook, *J. Phys. Chem. C*, **121**, 511 (2017).

[26] A. P. Bartók, G. Csányi: Gaussian approximation potentials: A brief tutorial introduction, *Int. J. Quantum Chem.*, **115**, 1051 (2015).

[27] Z. Li, J. R. Kermode, A. De Vita: Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces, *Phys. Rev. Lett.*, **114**, 096405 (2015).

[28] N. Artrith, B. Hiller, J. Behler: Neural network potentials for metals and oxides – First applications to copper clusters at zinc oxide, *Phys. Status Solidi B*, **250**, 1191 (2013).

[29] N. Artrith, T. Morawietz, J. Behler: High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide, *Phys. Rev. B*, **83**, 153101 (2011).

[30] J. Behler, R. Martoňák, D. Donadio, M. Parrinello: Metadynamics Simulations of the High-Pressure Phases of Silicon Employing a High-Dimensional Neural Network Potential, *Phys. Rev. Lett.*, **100**, 185501 (2008).

[31] J. Behler, R. Martoňák, D. Donadio, M. Parrinello: Pressure-induced phase transitions in silicon studied by neural network-based metadynamics simulations, *Phys. Status Solidi B*, **245**, 2618 (2008).

[32] H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler, M. Parrinello: *Ab initio* quality neural-network potential for sodium, *Phys. Rev. B*, **81**, 184107 (2010).

[33] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, M. Parrinello: Graphite-diamond phase coexistence study employing a neural-network mapping of the *ab initio* potential energy surface, *Phys. Rev. B*, **81**, 100103 (2010).

[34] T. B. Blank, S. D. Brown, A. W. Calhoun, D. J. Doren: Neural network models of potential energy surfaces, *J. Chem. Phys.*, **103**, 4129 (1995).

[35] S. Lorenz, M. Scheffler, A. Gross: Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface, *Phys. Rev. B*, **73**, 115431 (2006).

[36] J. Ludwig, D. G. Vlachos: *Ab initio* molecular dynamics of hydrogen dissociation on metal surfaces using neural networks and novelty sampling, *J. Chem. Phys.*, **127**, 154716 (2007).

[37] S. Manzhos, K. Yamashita, T. Carrington, Jr.: Fitting sparse multidimensional data with low-dimensional terms, *Comput. Phys. Commun.*, **180**, 2002 (2009).

[38] C. Carbogno, J. Behler, A. Gross, K. Reuter: Fingerprints for Spin-Selection Rules in the Interaction Dynamics of O_2 at Al(111), *Phys. Rev. Lett.*, **101**, 096104 (2008).

[39] J. Behler, K. Reuter, M. Scheffler: Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface, *Phys. Rev. B*, **77**, 115421 (2008).

[40] D. A. R. S. Latino, R. P. S. Fartaria, F. F. M. Freitas, J. Aires-de Sousa, F. M. S. Silva Fernandes: Mapping Potential Energy Surfaces by Neural Networks: The ethanol/Au(1 1 1) interface, *J. Electroanal. Chem.*, **624**, 109 (2008).

[41] D. A. R. S. Latino, R. P. S. Fartaria, F. F. M. Freitas, J. Aires-De Sousa, F. M. S. Silva Fernandes: Approach to potential energy surfaces by neural networks. A review of recent work, *Int. J. Quantum Chem.*, **110**, 432 (2010).

[42] T. Liu, B. Fu, D. H. Zhang: Six-dimensional potential energy surface of the dissociative chemisorption of HCl on Au(111) using neural networks, *Sci. China Chem.*, **57**, 147 (2013).

[43] D. F. R. Brown, M. N. Gibbs, D. C. Clary: Combining *ab initio* computations, neural networks, and diffusion Monte Carlo: An efficient method to treat weakly bound molecules, *J. Chem. Phys.*, **105**, 7597 (1996).

[44] S. Houlding, S. Y. Liem, P. L. A. Popelier: A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer, *Int. J. Quantum Chem.*, **107**, 2817 (2007).

[45] K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, H. A. Scheraga: Description of the potential energy surface of the water dimer with an artificial neural network, *Chem. Phys. Lett.*, **271**, 152 (1997).

[46] K.-H. Cho, K. T. No, H. A. Scheraga: A polarizable force field for water using an artificial neural network, *J. Mol. Struct.*, **641**, 77 (2002).

[47] H. Gassner, M. Probst, A. Lauenstein, K. Hermansson: Representation of Intermolecular Potential Functions by Neural Networks, *J. Phys. Chem. A*, **102**, 4596 (1998).

[48] F. V. Prudente, P. H. Acioli, J. J. Soares Neto: The fitting of potential energy surfaces using neural networks: Application to the study of vibrational levels of H_3^+, *J. Chem. Phys.*, **109**, 8801 (1998).

[49] T. M. Rocha Filho, Z. T. Oliveira, L. A. C. Malbouisson, R. Gargano, J. J. Soares Neto: The use of neural networks for fitting potential energy surfaces: A comparative case study for the H_3^+ molecule, *Int. J. Quantum Chem.*, **95**, 281 (2003).

[50] M. Malshe, L. M. Raff, M. G. Rockley, M. Hagan, P. M. Agrawal, R. Komanduri: Theoretical investigation of the dissociation dynamics of vibrationally excited vinyl bromide on an *ab initio* potential-energy surface obtained using modified novelty sampling and feedforward neural networks. II. Numerical application of the method, *J. Chem. Phys.*, **127**, 134105 (2007).

[51] L. M. Raff, M. Malshe, M. Hagan, D. I. Doughan, M. G. Rockley, R. Komanduri: *Ab initio* potential-energy surfaces for complex, multichannel systems using modified novelty sampling and feedforward neural networks, *J. Chem. Phys.*, **122**, 084104 (2005).

[52] P. M. Agrawal, L. M. Raff, M. T. Hagan, R. Komanduri: Molecular dynamics investigations of the dissociation of SiO_2 on an *ab initio* potential energy surface obtained using neural network methods, *J. Chem. Phys.*, **124**, 134306 (2006).

[53] H. M. Le, S. Huynh, L. M. Raff: Molecular dissociation of hydrogen peroxide (HOOH) on a neural network *ab initio* potential surface with a new configuration sampling method involving gradient fitting, *J. Chem. Phys.*, **131**, 014107 (2009).

[54] S. Manzhos, T. Carrington, Jr.: Using neural networks to represent potential surfaces as sums of products, *J. Chem. Phys.*, **125**, 194105 (2006).

[55] H. M. Le, L. M. Raff: Cis → trans, trans → cis isomerizations and N-O bond dissociation of nitrous acid (HONO) on an *ab initio* potential surface obtained by novelty sampling and feed-forward neural network fitting, *J. Chem. Phys.*, **128**, 194310 (2008).

[56] M. G. Darley, C. M. Handley, P. L. A. Popelier: Beyond Point Charges: Dynamic Polarization from Neural Net Predicted Multipole Moments, *J. Chem. Theory Comput.*, **4**, 1435 (2008).

[57] H. M. Le, T. S. Dinh, H. V. Le: Molecular Dynamics Investigations of Ozone on an *Ab Initio* Potential Energy Surface with the Utilization of Pattern-Recognition Neural Network for Accurate Determination of Product Formation, *J. Phys. Chem. A*, **115**, 10862 (2011).

[58] A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, R. Komanduri: Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks, *J. Chem. Phys.*, **130**, 134101 (2009).

[59] H. T. T. Nguyen, H. M. Le: Modified Feed-Forward Neural Network Structures and Combined-Function-Derivative Approximations Incorporating Exchange Symmetry for Potential Energy Surface Fitting, *J. Phys. Chem. A*, **116**, 4629 (2012).

[60] J. Chen, X. Xu, X. Xu, D. H. Zhang: Communication: An accurate global potential energy surface for the OH + CO → H + CO_2 reaction using neural networks, *J. Chem. Phys.*, **138**, 221104 (2013).

[61] J. Li, B. Jiang, H. Guo: Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems, *J. Chem. Phys.*, **139**, 204103 (2013).

[62] J. S. Smith, O. Isayev, A. E. Roitberg: ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci.*, **8**, 3192 (2017).

[63] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller: Machine Learning of Accurate Energy-Conserving Molecular Force Fields, *Sci. Adv.*, **3**, e1603015 (2017).

[64] T. Hofmann, B. Schölkopf, A. J. Smola: Kernel methods in machine learning, *Ann. Statist.*, **36**, 1171 (2008).

[65] G. E. Hinton, S. Osindero, Y.-W. Teh: A fast learning algorithm for deep belief nets, *Neural Comput.*, **18**, 1527 (2006).

[66] J. Behler, M. Parrinello: Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.*, **98**, 146401 (2007).

[67] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko: Fragmentation methods: A route to accurate calculations on large systems, *Chem. Rev.*, **112**, 632 (2012).

[68] W. S. McCulloch, W. Pitts: A logical calculus of the ideas immanent in nervous activity, *B. Math. Biophys.*, **5**, 115 (1943).

[69] M. Minsky, S. Papert: *Perceptrons*, M.I.T. Press, Oxford, England (1969).

[70] F. Rosenblatt: The perceptron: A probabilistic model for information storage and organization in the brain, *Psychol. Rev.*, **65**, 386 (1958).

[71] G. Cybenko: Approximation by superpositions of a sigmoidal function, *Math. Control Signal Syst.*, **2**, 303 (1989).

[72] K. Hornik, M. Stinchcombe, H. White: Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359 (1989).

[73] K. Hornik: Approximation capabilities of multilayer feedforward networks, *Neural Networks*, **4**, 251 (1991).

[74] J. Behler: Constructing high-dimensional neural network potentials: A tutorial review, *Int. J. Quantum Chem.*, **115**, 1032 (2015).

[75] J. Behler: Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, **134**, 074106 (2011).

[76] D. E. Rumelhart, G. E. Hinton, R. J. Williams: Learning representations by back-propagating errors, *Nature*, **323**, 533 (1986).

[77] Y. LeCun, L. Bottou, G. B. Orr, K. R. Müller: Efficient backprop, in *Neural Networks: Tricks of the Trade*, Springer Berlin Heidelberg, Berlin, Heidelberg (1998).

[78] Y. Nesterov: A method of solving a convex programming problem with convergence rate O(1/k2), in *Soviet Mathematics Doklady*, volume 27, 372–376 (1983).

[79] J. Duchi, E. Hazan, Y. Singer: Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.*, **12**, 2121 (2011).

[80] D. P. Kingma, J. Ba: Adam: A Method for Stochastic Optimization, *CoRR*, **abs/1412.6980**, 0000 (2014).

[81] T. B. Blank, S. D. Brown: Adaptive, global, extended Kalman filters for training feedforward neural networks, *J. Chemometrics*, **8**, 391 (2005).

[82] S. Shah, F. Palmieri, M. Datum: Optimal filtering algorithms for fast learning in feedforward neural networks, *Neural Networks*, **5**, 779 (1992).

[83] X. Glorot, Y. Bengio: Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256 (2010).

[84] A. Rahman: Correlations in the Motion of Atoms in Liquid Argon, *Phys. Rev.*, **136**, A405 (1964).

[85] B. J. Alder, T. E. Wainwright: Studies in Molecular Dynamics. I. General Method, *J. Chem. Phys.*, **31**, 459 (1959).

[86] I. Newton: *Philosophiae naturalis principia mathematica*, J. Societatis Regiae ac Typis J. Streater (1687).

[87] D. Marx, J. Hutter: *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge, Reprint edition (2012).

[88] M. Barbatti: Nonadiabatic dynamics with trajectory surface hopping method, *WIREs Comput. Mol. Sci.*, **1**, 620 (2011).

[89] M.-P. Gaigeot: Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase, *Phys. Chem. Chem. Phys.*, **12**, 3336 (2010).

[90] M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, B. Kirchner: Computing vibrational spectra from ab initio molecular dynamics, *Phys. Chem. Chem. Phys.*, **15**, 6608 (2013).

[91] S. Mai, P. Marquetand, L. González: A General Method to Describe Intersystem Crossing Dynamics in Trajectory Surface Hopping, *Int. J. Quantum Chem.*, **115**, 1215 (2015).

[92] P. Marquetand, J. Nogueira, S. Mai, F. Plasser, L. González: Challenges in Simulating Light-Induced Processes in DNA, *Molecules*, **22**, 49 (2016).

[93] P. H. Hünenberger: *Thermostat Algorithms for Molecular Dynamics Simulations*, 105–149, Springer Berlin Heidelberg, Berlin, Heidelberg (2005).

[94] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak: Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, **81**, 3684 (1984).

[95] H. C. Andersen: Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.*, **72**, 2384 (1980).

[96] G. J. Martyna, M. L. Klein, M. Tuckerman: Nosé–Hoover chains: The canonical ensemble via continuous dynamics, *J. Chem. Phys.*, **97**, 2635 (1992).

[97] S. Nosé: A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.*, **52**, 255 (1984).

[98] W. G. Hoover: Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A*, **31**, 1695 (1985).

[99] P. Tiwary, A. van de Walle: *A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics*, 195–221, Springer International Publishing, Cham (2016).

[100] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, A. Shehu: Principles and overview of sampling methods for modeling macromolecular structure and dynamics, *PLoS Comput. Biol.*, **12**, 1 (2016).

[101] A. Laio, M. Parrinello: Escaping free-energy minima, *Proc. Natl. Acad. Sci. USA*, **99**, 12562 (2002).

[102] S. Murtuza, S. F. Chorian: Node decoupled extended Kalman filter based learning algorithm for neural networks, in *Proceedings of the 1994 IEEE International Symposium on Intelligent Control*, 364–369 (1994).

[103] G. V. Puskorius, L. A. Feldkamp: Decoupled extended Kalman filter training of feedforward layered networks, in *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume i, 771–777 (1991).

[104] M. A. Collins, M. W. Cvitkovic, R. P. A. Bettens: The Combined Fragmentation and Systematic Molecular Fragmentation Methods, *Acc. Chem. Res.*, **47**, 2776 (2014).

[105] P. Seema, J. Behler, D. Marx: Adsorption of Methanethiolate and Atomic Sulfur at the Cu(111) Surface: A Computational Study, *J. Phys. Chem. C*, **117**, 337 (2013).

[106] V. Botu, R. Ramprasad: Adaptive machine learning framework to accelerate ab initio molecular dynamics, *Int. J. Quantum Chem.*, **115**, 1074 (2015).

[107] D. Frenkel, B. Smit: Chapter 3 - Monte Carlo Simulations, in *Understanding Molecular Simulation*, 23 – 61, Academic Press, San Diego, 2nd edition (2002).

[108] S. M. Bachrach: *Population Analysis and Electron Densities from Quantum Mechanics*, 171–228, John Wiley & Sons, Inc. (2007).

[109] C. M. Baker: Polarizable force fields for molecular dynamics simulations of biomolecules, *WIREs Comput. Mol. Sci.*, **5**, 241 (2015).

[110] A. D. Becke: Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A*, **38**, 3098 (1988).

[111] P. A. M. Dirac: Quantum Mechanics of Many-Electron Systems, *Proc. R. Soc. A*, **123**, 714 (1929).

[112] J. P. Perdew: Density-functional approximation for the correlation energy of the inhomogeneous electron gas, *Phys. Rev. B*, **33**, 8822 (1986).

[113] S. H. Vosko, L. Wilk, M. Nusair: Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, *Can. J. Phys.*, **58**, 1200 (1980).

[114] J. C. Slater: A Simplification of the Hartree-Fock Method, *Phys. Rev.*, **81**, 385 (1951).

[115] G. D. I. Purvis, R. J. Bartlett: A full coupled-cluster singles and doubles model: The inclusion of disconnected triples, *J. Chem. Phys.*, **76**, 1910 (1982).

[116] T. K. Roy, R. B. Gerber: Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications, *Phys. Chem. Chem. Phys.*, **15**, 9468 (2013).

[117] P. Chu, F. Guenther, G. Rhoderick, W. Lafferty: *NIST Chemistry WebBook NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg MD, 20899 (doi:10.18434/T4D303, (retrieved April 24, 2017)).

[118] S. Grimme: Semiempirical hybrid density functional with perturbative second-order correlation, *J. Chem. Phys.*, **124**, 034108 (2006).

[119] C. Lee, W. Yang, R. G. Parr: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B*, **37**, 785 (1988).

[120] T. D. Vaden, T. S. J. A. de Boer, J. P. Simons, L. C. Snoek, S. Suhai, B. Paizs: Vibrational Spectroscopy and Conformational Structure of Protonated Polyalanine Peptides Isolated in the Gas Phase, *J. Phys. Chem. A*, **112**, 4608 (2008).

[121] J. Behler: RuNNer – A program for constructing high-dimensional neural network potentials, Universität Göttingen (2017).

[122] M. Gastegger, J. Behler, P. Marquetand: Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra, In review (2017).

[123] M. Gastegger, C. Kauffmann, J. Behler, P. Marquetand: Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes, *J. Chem. Phys.*, **144**, 194110 (2016).

[124] M. Gastegger, P. Marquetand: High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm, *J. Chem. Theory Comput.*, **11**, 2187 (2015).

ACKNOWLEDGMENTS

My sincerest thanks go to all the people who made this thesis possible!

Zuallererst möchte ich meiner Familie und insbesondere meinen Eltern Herta und Johann sowie meiner Schwester Julia danken. Danke, dass ihr mir mein Studium ermöglicht und mich in allen denkbaren Belangen stets unterstützt habt, auch wenn ihr euch manchmal nicht sicher wart, was genau ich jetzt wirklich treibe. Ein herzliches Danke gilt auch meinen Großeltern Ottilie, Franz, Hermine und Johann.

A central pillar of this thesis was Dr. Philip Marquetand, who shares my fascination with this research topic and provided invaluable support over the last years, starting out as my inofficial and rising to the position of my official PHD supervisor in the process.

My deepest gratitude also goes to Prof. Dr. Leticia González, in whose group this thesis was conducted. Thank you for not only giving me the opportunity to carry out my PHD and providing me with the necessary funding, but even going so far as allowing me and Philipp to "play around" and start this new research topic.

Special thanks also go to my once and present office colleagues, Dr. Leon Freitag, Dr. Sebastian Mai and Clemens Rauer. I doubt these last years would have been even half as enjoyable without the many discussions, games and out of work collaborations arising in this inner circle.

A significant contribution to the relaxed atmosphere of my PHD research was also provided by the many, many members of the core and extended González family: Dr. Andrew Atkins, Sandra Gómez-Rodríguez, Dr. Boris Maryasin, Dr. Juan Jose Nogueira-Peréz, Dr. Pedro Sanchez-Murcia, Dr. Vera Krewald, Davide Avagliano, Christoph Bauer, David Ferro, Dr. Andrea Fülöpová, Lucy Grosvenor, Moritz Heindl, Dr. Daniel Kinzel, Dr. Federico Latorre, Maximilian Meixner, Maximilian Menger, Prof. Antonio Mota, Dr. Aurora Muñoz-Losa, Dr. Rana Obaid, Solène Oberli, Dr. Felix Plasser, Dr. Martin Richter, Dr. Matthias Ruckenbauer, Dr. Stefan Ruiders, Isolde Sandler, Ludwig Schwiedrzik, Martina de Vetta, David Weichselbaum, Julia Westermayr and J. Patrick Zobel. Thanks for all the cake, singing, numerous fruitful and not especially fruitful discussions (in particular during the seminars) and the many funny and enjoyable moments even out of work.

I also want to thank my opera ally and Konzerthaus colleague from the 3rd floor Roman Ochsenreiter for the many hours of suffering through Wagner and Mozart together just to get to the good parts. There never have been better conditions to think deeply about research than five hours of *Parsifal*.

My deep gratitude goes to our system administrator Dr. Markus Oppel and the unbroken chains of institute technicians Jackie Klaura, Markus Hickel, and Dimitri Robl and secretaries Kathrine Baumann, Edith Steinwidder, Nicole Irmler, Irina Stumpf, and Monika Schett, who kept the real life problems from interfering with my research.

It was a pleasure to collaborate with Prof. Dr. Jörg Behler, who gave me the opportunity to implement aspects of my research in his `RuNNer` program and supported me with his expansive expertise in the field.

Likewise, I would like thank the students Florian Berzsenyi, Alexander Fischer, Marius Bittermann, Clemens Kauffmann, Isolde Sandler, Ludwig Schwiedrzik

and Julia Westermayr, who had the dubious pleasure of being supervised by me, for their patience and help in my research.

Finally, I want to express my sincere gratitude the University of Vienna for providing financial support over these last years, as well as to the Österreichische Akademie der Wissenschaften (ÖAW), Springer and the Gesellschaft Österreichischer Chemiker (GÖCH) for awarding me the Stipendium der Monatshefte für Chemie to further my PHD research.

And on a preemptive note, I wish to thank our future virtual overlords. Without your humble predecessors this research would not have been possible.

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne Zuhilfenahme weiterer als der aufgeführten Quellen angefertigt habe. Alle wörtlich oder sinngemäß übernommenen Textstellen anderer Verfasser wurden als solche gekennzeichnet.

Michael Gastegger Wien, 2017

Michael Gastegger

Education

January 2014 – present	Ph. D. in Chemistry **University of Vienna** Advisor: Priv.-Doz. Dr. Philipp Marquetand
December 2013	M. Sc. in Chemistry **University of Vienna** Thesis: "*De-novo* design of an enzyme for olefin metathesis" Advisor: Univ.-Prof. Dr. Leticia González
August 2011	B. Sc. in Chemistry **University of Vienna**
June 2006	Matura **Bundesrealgymnasium Lilienfeld**

Research Interests and Expertise

- Machine learning in theoretical chemistry
- Fragmentation methods for large molecular systems
- *De novo* enzyme design

Publications

3. M. Gastegger, J. Behler, P. Marquetand: Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra, In review (2017).

2. M. Gastegger, C. Kauffmann, J. Behler, P. Marquetand: Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes, *J. Chem. Phys.*, **144**, 194110 (2016).

1. M. Gastegger, P. Marquetand: High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm, *J. Chem. Theory Comput.*, **11**, 2187 (2015).

Conference Talks

1. M. Gastegger and P. Marquetand: Dynamical Simulation of Infrared Spectra with Neural Network Potentials, International Workshop on Machine Learning for Materials Science, 08.03.2017–09.03.2017, Espoo, Finland.

Conference Posters

5. M. Gastegger, C. Kauffmann, J. Behler and P. Marquetand: High-Dimensional Neural Network Potentials: Inherent Fragmentation Capa-

bilities and Impact of Structural Descriptors, CECAM-workshop: Exploring Chemical Space with Machine Learning and Quantum Mechanics, 30.05.2016–03.06.2016, Zürich, Switzerland.

4. M. Gastegger and P. Marquetand: An Improved Training Algorithm for High-Dimensional Neural Network Potentials Applied to a Claisen Reaction, 51st Symposium on Theoretical Chemistry (STC 2015), 20.09.2015–24.09.2015, Potsdam, Germany.

3. M. Gastegger and P. Marquetand: An Improved Training Algorithm for High-Dimensional Neural Network Potentials Applied to a Claisen Reaction, CECAM-workshop: From Many-Body Hamiltonians to Machine Learning and Back, 01.05.2015–13.05.2015, Berlin, Germany.

2. M. Gastegger, P. Marquetand, L. González, and C. Flamm: A Genetic Algorithm for the Creation of Olefin Metathesis Active Sites in De-Novo Enzyme Design, 50th Symposium on Theoretical Chemistry (STC 2014), 14.09.2014–18.09.2014, Vienna, Austria.

1. M. Gastegger, P. Marquetand, L. González, and C. Flamm: An Evolutionary Approach to Computational De-Novo Enzyme Design for Olefin Metathesis, MDMM 2014 - Modelling and Design of Molecular Materials, 29.06.2014 - 03.07.2014, Kudowa Zdrój, Poland.

Languages

German: mother tongue
English: fluent