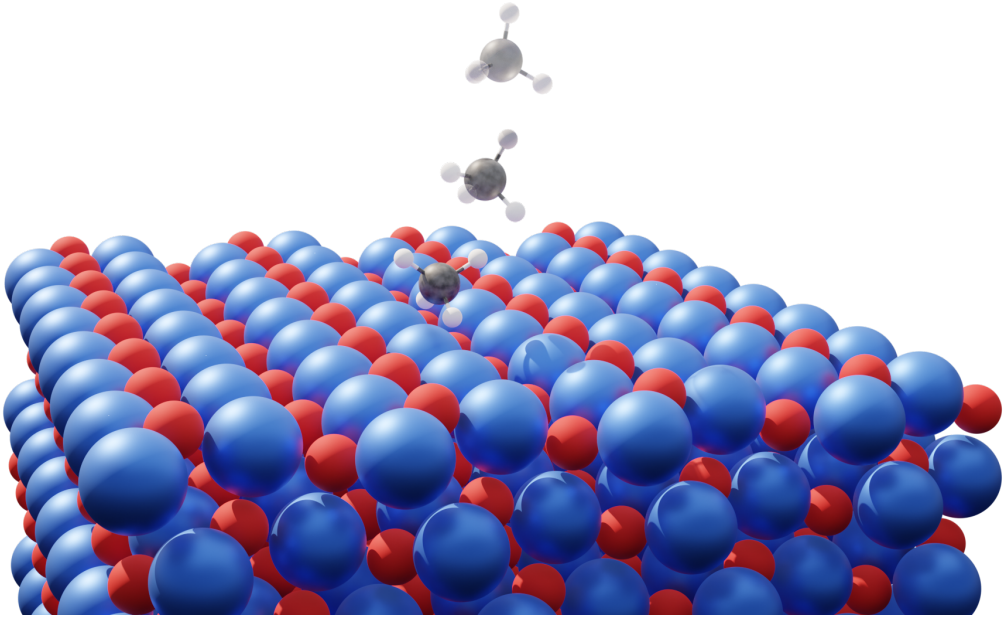


Sticking of Methane on Palladium Oxide: a Computational Approach

A combination of density functional theory, neural networks, and molecular dynamics to study the adsorption and dissociation of methane on a palladium oxide (101) surface.

Master's thesis in Physics

RASMUS SVENSSON

DEPARTMENT OF PHYSICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

Sticking of Methane on Palladium Oxide: a Computational Approach

A combination of density functional theory, neural networks, and molecular dynamics to study the adsorption and dissociation of methane on a palladium oxide (101) surface.

Rasmus Svensson

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics
Division of Chemical Physics
Chalmers University of Technology
Gothenburg, Sweden 2021

Sticking of Methane on Palladium Oxide: a Computational Approach
A combination of density functional theory, neural networks, and molecular dynamics to study the adsorption and dissociation of methane on a palladium oxide (101) surface.
Rasmus Svensson

Supervisor: Professor Anders Hellman, Department of Physics
Examiner: Professor Anders Hellman, Department of Physics

Master's Thesis 2021
Department of Physics
Division of Chemical Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone: +46 31 772 1000

Cover: A methane molecule approaching the active site of the palladium oxide (101) surface. The figure was modelled and rendered in Blender.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Sticking of Methane on Palladium Oxide: a Computational Approach
A combination of density functional theory, neural networks, and molecular dynamics to study the adsorption and dissociation of methane on a palladium oxide (101) surface.
Rasmus Svensson
Department of Physics
Chalmers University of Technology

Abstract

The catalytic properties of palladium oxide for the combustion of methane have been studied extensively in recent years. The rate-determining step of this reaction is believed to be the dissociation of methane on the surface. The rate of the event is dependent on both the active sites of the catalyst and the energy and orientation of the incoming methane molecules. The dependence of energy and orientation is often summarized in a sticking coefficient.

Here, we will address the challenge of calculating the sticking coefficient from first-principles. However, due to the large number of trials and large time scales required to study this event, ab initio molecular dynamics would be too computationally expensive to perform, and an alternative approach using neural networks is applied. The adsorption position on the active sites and the activation energy of the dissociation process are studied using density functional theory. To determine the probability of a sticking event, a neural network is trained to predict the multidimensional potential energy surface, which is used to perform molecular dynamics. The density functional theory calculations confirm that the active sites of the catalyst are the under-coordinated palladium atoms, with an apparent activation energy of 0.2 eV for the dissociation reaction. The neural network is able to predict the energies of the system five orders of magnitude faster than regular density functional theory calculations, with an MAE of 0.02 eV. The molecular dynamics suggest that the previously believed most probable transition path might be dominated by the sum of the other, less likely, transition paths. The hope is that the results and understanding obtained from this computational study can be used to assist in the discovery of more efficiently designed catalysts in the future.

Keywords: Density functional theory, neural networks, methane, palladium oxide, sticking, adsorption, dissociation, molecular dynamics, potential energy surface, catalysis, activation energy

Acknowledgements

I want to express my wholehearted gratitude to my supervisor Anders Hellman, who always manages to be incredibly positive and helpful. Thank you for your guidance, and our entertaining and informative discussions. I would also like to thank the members of the Chemical Physics division for being very welcoming of me. Special thanks to Henrik Grönbeck, Alvaro Posada Borbon, and Björn Wickman for sharing your expertise and making the work environment very pleasant.

Rasmus Svensson, Gothenburg, June 2021

Table of Contents

1. Introduction

The importance of sustainable energy sources is becoming more paramount each day. Two energy sources, which have been of great interest in recent years, are the combustion of natural gas and biogas. The advantage of natural gas and biogas, compared to, e.g., oil and gasoline, is that they are constituted mostly of methane gas, which when combusted has a low emission rate of carbon dioxide in relation to the amount of energy produced [1]. However, since methane is a potent greenhouse gas [2], the need for complete combustion is of utmost importance. This has sparked the need for effective catalysts. Some of the most efficient catalysts for the complete combustion of methane are based on palladium, and in particular palladium oxide [3], [4], of which, the latter will be considered in this report.

The combustion of methane on a palladium oxide catalyst is a very complex process [5]. The first part of this process is the adsorption and dissociation of methane on the surface, yielding methyl and a hydrogen atom. This part is believed to be the rate-limiting step of the reaction, due to the weak interactions between the palladium oxide surface and the methane molecules, and the large activation energy of the dissociation process [4]. Due to these conditions, the rate of dissociation is highly dependent on both the active sites of the catalyst, and the energy and orientation of the incoming methane molecules. The dependence of energy and orientation can be summarized in a sticking coefficient [5]. Previous, experimental, studies on noble metal catalysts have yielded sticking coefficient values between 10^{-4}–10^{-1} [6]. However, to computationally determine the sticking coefficient, and study the dissociation process from first-principles, a very large number of molecular dynamics simulations have to be performed. Due to the large number of simulations and large time scales required to study the process, regular ab initio molecular dynamics would be too computationally expensive to perform. Therefore, feed-forward neural networks, trained with density functional theory data, will be applied to determine the multi-dimensional potential energy surface and its gradients, relevant for, e.g., the molecular dynamics simulations.

1.1 Aim of the Project

The aim of this project is to study the active sites of the catalyst, the dissociation process, and the sticking coefficient, with different computational tools. The active site will be studied using density functional theory. The sticking coefficient will be determined, and the dissociation process will be studied, using a combination of density functional theory and neural networks.

1.2 Outline

The first chapter will include general theory about the catalyst, density functional theory, and feed-forward neural networks. In the second chapter, more specific theory about the methods, and computational tools, used in the project will be described. This will be followed by a description of the simulations, experiments, and work done in the project. This chapter will also include some results required to determine the computational setup of the simulations. In the succeeding chapter, the results will be presented, analysed, and interpreted. The report will be concluded with a reflection and an outlook.

2. Theory

"*In the case of science, I think that one of the things that make it very difficult, is that it takes a lot of imagination. It's very hard to imagine all the crazy things that things really are like.*"

— Richard Feynman, Fun to Imagine (1983)

In this chapter, the theory behind some of these crazy things will be explained. This will include theory about the catalyst, density functional theory calculations, feed-forward neural networks, and the Velocity Verlet algorithm.

2.1 The Catalyst

The complete combustion of methane results in carbon dioxide and water, i.e.

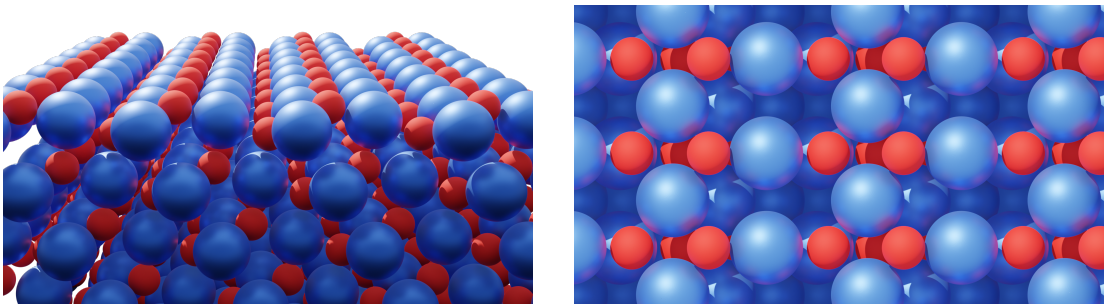
$$\mathrm{CH_4(g)} + 2\,\mathrm{O_2(g)} \rightleftharpoons \mathrm{CO_2(g)} + 2\,\mathrm{H_2O\,(l)} \quad \Delta H^\circ = -891\,\mathrm{kJ/mol}, \tag{2.1}$$

where ΔH° is the combustion enthalpy [7], [8]. Even though the reaction is highly exothermic, the activation energy is very large (see Figure 4.1). To solve this problem, a catalyst can be used, in order to enable an alternative reaction mechanism, with a lower activation energy [9]. It is important to note that a catalyst only affects the kinetics of a reaction, not the overall thermodynamics. Many different catalysts have been considered for reaction (2.1), including platinum [10], rhodium/zirconium oxide [11], different perovskite-type oxides [12], supported palladium structures [13], and palladium oxides [4], [5], [14], of which, the latter will be considered in this report.

The thermodynamically most stable crystal structure of palladium is Face Centered Cubic (FCC), with a lattice parameter of 3.89 Å at room temperature [15]. The experimental value of the cohesive energy for bulk palladium, i.e., the energy of a palladium atom in gas phase relative to the energy of a palladium atom in the FCC crystal, is 3.91 eV at 0 K and 3.92 eV at 298.15 K [16]. Palladium oxide, on the other hand, has a tetrogonal crystal structure with a base of 3.043 Å and a height of 5.336 Å [17]. The primitive cell contains two palladium atoms located at $(0, 0, 0)$ and $(0.5, 0.5, 0.5)$ and two oxygen atoms located at $(0.5, 0, 0.25)$ and $(0.5, 0, 0.75)$ [18]. The enthalpy of formation for palladium oxide is 1.22 eV at 298.15 K [19].

Previous calculations and X-ray diffraction have shown that the palladium oxide facet with the largest methane conversion rate is the (101) surface, either pure or grown on metallic palladium (100) [4], [20]. The structure of palladium oxide (101)

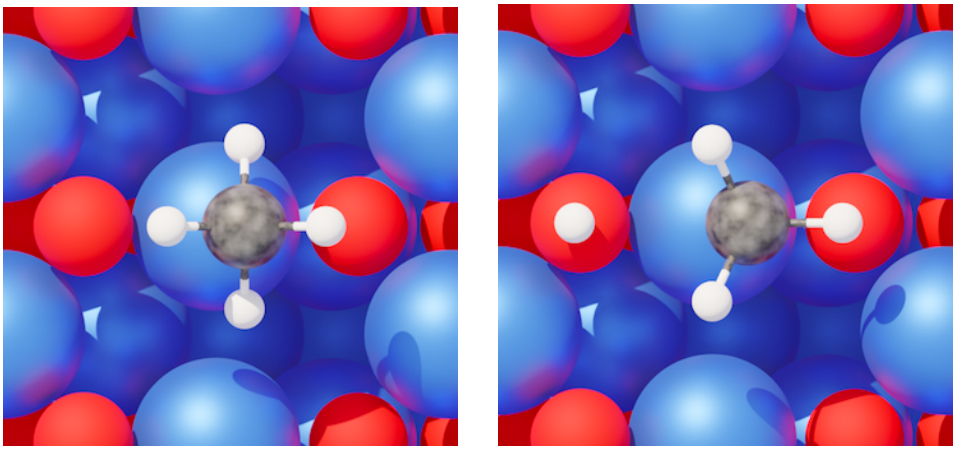
is shown in Figure 2.1. If palladium oxide (101) is grown on metallic palladium (100), the latter lattice is rotated 27° compared to the oxide, this is known as the $(\sqrt{5} \times \sqrt{5})$R27° oxide [21]. Even though (2.1) might look simple, the combustion of methane on a palladium oxide catalyst is a very complex reaction, with many different processes [5]. The first process in this reaction is the sticking and dissociation of methane on the surface, yielding methyl and a hydrogen atom. This step is believed to be the rate-limiting step, due to the weak interactions between the palladium oxide (101) surface and the methane molecule, and the large activation energy of the dissociation [4].



(a) Side view **(b)** Top view

Figure 2.1: The structure of palladium oxide (101).

In Figure 2.1 it can be seen that the palladium atoms in palladium oxide (101) exist in two different environments; alternating every second row. Previous calculations have shown that the active site for the dissociation of methane is over the 3-fold coordinated palladium atoms, which has one oxygen atom on each side and one directly below [4], [5], [14]. The initial state and the final state of the methane dissociation on the surface are shown in Figure 2.2. Previous calculations have shown that the adsorption energy when the methane molecule saddles over the 3-fold coordinated palladium atom, neglecting the van der Waals interactions, is between negative 0.14 – 0.17 eV [4], [14]. At the initial position, the H–C–H angle between the saddled hydrogen atoms has been determined to be approximately 114° – 115°, compared to 109.5° for the free methane molecule. It has been shown that the energy of the final state is 0.41 – 0.49 eV lower than the energy of the initial state. The activation energy of the process has also been proven to be very dependent on the computational tools, and usually ranges between 0.3 – 0.7 eV [4], [5], [14]. The apparent activation energies ranges between 0.35 – 0.5 eV. Experimentally, the value of the apparent activation energy for (2.1) has ranged between 0.3 – 1.3 eV, dependent on the experimental conditions [22].



(a) Initial State **(b)** Final State

Figure 2.2: The initial state and the final state in the dissociation reaction of methane on the palladium oxide (101)-surface.

2.2 Density Functional Theory

Properties of chemical systems, including e.g., molecules, metal-, and metal-oxide surfaces, can be described precisely from the solutions of the Schrödinger equation. The time-independent Schrödinger equation is given by

$$E\left|\Psi\right\rangle = H\left|\Psi\right\rangle, \tag{2.2}$$

where E is the energy eigenvalue, and H is the *Hamiltonian* [23]. Even though some systems can be solved numerically when exploiting symmetries, geometries, and boundary conditions, there are no analytic solutions to the Schrödinger equation when two or more electrons are considered. The reason being that for systems with more than one electron, the Hamiltonian gets more complex; more Coulomb interactions and the Pauli exclusion principle have to be considered [24]. This means that the particles are antisymmetric under the interchange of two particles, i.e.,

$$\left|\Psi(\mathbf{r}_1, ..., \mathbf{r}_i, ..., \mathbf{r}_n, ...)\right\rangle = -\left|\Psi(\mathbf{r}_1, ..., \mathbf{r}_n, ..., \mathbf{r}_i, ...)\right\rangle, \tag{2.3}$$

where $\mathbf{r}_j$ is the position vector for particle j. Generally, if the nuclei are seen as individual particles, the Hamiltonian can be expressed as

$$\begin{aligned} H = &-\sum_i \frac{\hbar^2}{2m_\mathrm{e}}\nabla_i^2 - \sum_k \frac{\hbar^2}{2M_k}\nabla_k^2 \\ &+ \frac{1}{8\pi\varepsilon_0}\sum_{i\neq j}\frac{e^2}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} - \frac{1}{4\pi\varepsilon_0}\sum_{i,k}\frac{Z_k e^2}{\left|\mathbf{r}_i - \mathbf{R}_k\right|} + \frac{1}{8\pi\varepsilon_0}\sum_{k\neq l}\frac{Z_k Z_l e^2}{\left|\mathbf{R}_k - \mathbf{R}_l\right|}, \end{aligned} \tag{2.4}$$

where the first and second terms are the kinetic energy for the electrons and the kinetic energy for the nuclei, respectively [24]. The last three terms are the Coulomb interactions between the electrons, between the nuclei and the electrons and between

the nuclei, in that order. From now on, Hartree atomic units will be used, where the reduced Planck constant, the Bohr radius, the electronic charge, the electronic mass, and the Coulomb force constant are all set to 1, i.e.,

$$\hbar = a_0 = e = m_e = \frac{1}{4\pi\varepsilon_0} = 1. \tag{2.5}$$

A common approximation to ease the computations of the Schrödinger equation is the Born-Oppenheimer approximation [25]. The mass of the nuclei is three to four orders of magnitude greater than the mass of an electron. The large difference in time scales entails the possibility to consider the wave functions of the nuclei and the electrons as decoupled [26]. This means that the electronic wave function can be solved from an electronic Schrödinger equation, with an external potential, dependent on the instantaneous positions of the nuclei. The Hamiltonian for the electronic Schrödinger equation is thus given by

$$H_{\mathrm{e}} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} - \sum_{i,k}\frac{Z_k}{|\mathbf{r}_i - \mathbf{R}_k|}. \tag{2.6}$$

However, even with the use of the Born-Oppenheimer approximation, the solutions to the electronic Schrödinger equation are still too computationally expensive to find. One way of dealing with this computationally costly problem is by using *Density Functional Theory* (DFT). Instead of dealing with the many-electron wave function, the energy of the system is evaluated using only the electron density [27].

This can be done thanks to certain discoveries of Hohenberg, Kohn, and Sham. The first theorem of Hohenberg and Kohn states that the external potential can be precisely determined (up to a constant) by the electron density in the system [28]; i.e., there exists a functional S, such that

$$V_{\mathrm{ext}} = S[\rho(\mathbf{r})], \tag{2.7}$$

where $\rho(\mathbf{r})$ is the electron density, calculated from the total electronic wave function as

$$\rho(\mathbf{r}) = \left|\Psi(\mathbf{r})\right|^2. \tag{2.8}$$

This was a revolutionary discovery, considering that from here on, not only could the electron density be determined from the external potential, but the contrary as well. A consequence of this is that the energy of the system can be calculated from only the electron density. To this day, there does not exist any determined form of this functional, and approximations have to be made. Apart from the Born-Oppenheimer approximation, this is the first approximation required in order to continue.

It is by minimizing the energy functional, with respect to the electron density, that the ground state energy of the system can be determined, i.e.,

$$E_0 = E[\rho_0(\mathbf{r})] = \min_{\rho(\mathbf{r})} E[\rho(\mathbf{r})]. \tag{2.9}$$

The theorem states that the ground state energy is found when the electron density equals the ground state electron density [28].

However, it is extremely difficult to determine the ground-state electron density. An alternative approach to solve this was suggested by Kohn and Sham [29]. The approximation is known as the Kohn-Sham ansatz. It states that the many-particle problem can be replaced by non-interacting particles existing in a so called effective potential. The fact that the electrons are seen as individual, non-interacting, particles, makes it possible to express the total electron density as a sum of the individual densities, i.e.,

$$\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2, \quad \text{where} \quad \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} = N \tag{2.10}$$

is the number of electrons in the system and $\phi_i(\mathbf{r})$ are obtained by solving the wave equations for the individual non-interacting electrons. The energy functional (2.9) now takes the form

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] - \int \sum_k \frac{Z_k \rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R_k}|} \mathrm{d}\mathbf{r} + \frac{1}{2} \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}[\rho(\mathbf{r})], \tag{2.11}$$

where the functional T yields the kinetic energy of the non-interacting electrons, i.e.,

$$T[\rho(\mathbf{r})] = -\frac{1}{2} \sum_i \langle \phi_i | \nabla_i^2 | \phi_i \rangle . \tag{2.12}$$

The functional E_{xc} is known as the exchange-correlation functional, and captures quantum mechanical properties that are neglected with this approach. The exchange-correlation functional is described more precisely in Section 2.3.

With the use of Lagrange multipliers, and exploiting the fact that the number of electrons are constant, the Kohn-Sham equations can be derived [29]. These equations, on which most DFT calculation are dependent, are

$$\left(-\frac{1}{2}\nabla_i^2 + v_{\mathrm{eff}}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}). \tag{2.13}$$

The effective potential, $v_{\mathrm{eff}}(\mathbf{r})$, in which the electrons exist, is given by

$$v_{\mathrm{eff}} = -\sum_k \frac{Z_k}{|\mathbf{r} - \mathbf{R_k}|} + \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}' + \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}. \tag{2.14}$$

The eigenvalues obtained from (2.13) do not correspond exactly to the total energy of the system [29]. Instead, the eigenvalues and eigenvectors obtained from (2.13) are used to obtain the total energy of the system according to

$$E = \sum_i \varepsilon_i - \frac{1}{2} \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}} - \int \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \rho(\mathbf{r}) \mathrm{d}\mathbf{r}, \tag{2.15}$$

where the total electron density $\rho(\mathbf{r})$ is calculated from the individual electron densities according to (2.10). The equations in this section are all that is needed in order

to perform a so called self-consistency loop, with the attempt to find the ground state energy of a system. The loop can be seen in Figure 2.3.

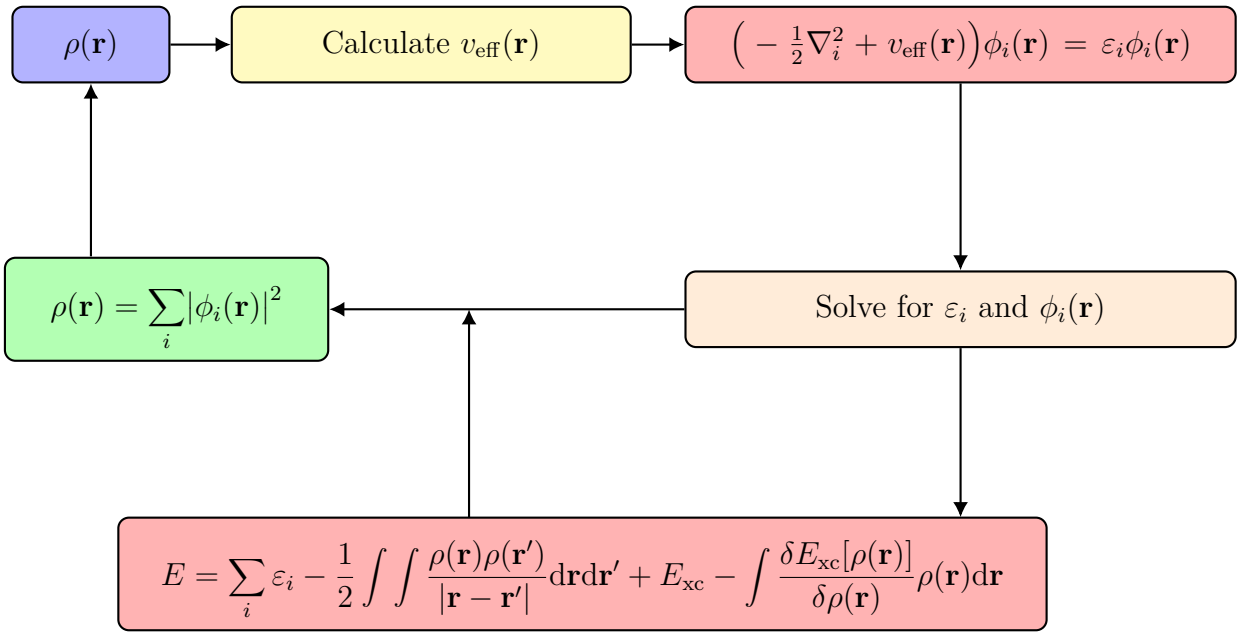


Figure 2.3: The self-consistency loop applied to obtain the ground state electron density and ground state energy of the electrons.

2.3 Exchange-Correlation Functional

When expressing the energy as in (2.11), an addition functional has been added, known as the *Exchange-Correlation Functional*, E_{xc}. This term captures certain quantum mechanical properties that are neglected when using the the Hartree-Fock approximation and the Kohn-Sham ansatz [30]. The exchange-correlation functional should, for example, capture

- the antisymmetry with respect to the interchange of two particles (2.3),
- corrections for self-interacting electrons in the density formulation,
- kinetic energy differences due to the Kohn-Sham ansatz [27].

Kohn and Sham proposed an exchange-correlation functional known as the Local Density Approximation (LDA), based on the homogeneous electron gas, which, of course, is a large simplification for most systems [29]. The functional is written as

$$E_{\text{xc}}^{\text{LDA}}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{\text{xc}}[\rho(\mathbf{r})]\mathrm{d}\mathbf{r}, \tag{2.16}$$

where $\varepsilon_{\text{xc}}[\rho(\mathbf{r})]$ is the energy due to exchange and correlation for a homogeneous electron gas. The exchange and correlation per electron is in this case known, and

can be expressed as

$$\varepsilon_{\text{xc}} = \underbrace{-\frac{3}{4}\left(\frac{9}{4\pi^2}\right)\frac{1}{r_s}}_{\varepsilon_x-\text{exchange}} + \underbrace{\begin{cases} 0.0311\ln r_s - 0.048 + 0.002 r_s \ln r_s - 0.0116 r_s, & \text{if } r_s < 1, \\ -0.1423/(1 + 1.0529\sqrt{r_s} + 0.3334 r_s), & \text{if } r_s \geq 1, \end{cases}}_{\varepsilon_c-\text{correlation}} \tag{2.17}$$

where

$$r_s = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{\frac{1}{3}}, \tag{2.18}$$

is known as the Wigner-Seitz radius. The exchange part in (2.17) is an exact, derived, expression [31], while the the correlation part has been obtained from computational experiments [32]. It has been shown that this approximation is adequate for systems where the electron density is relatively stationary [28]. However, this approximation is often not satisfactory, especially when calculating binding energies [30]. These limitations can, to an extent, be overcome with the introduction of a generalized gradient approximation (GGA), which has proven to improve, e.g., total energy calculations [33] and activation energy calculations [34]. The exchange-correlation energy is then determined from the electron density of the system, as well as the gradient of the electron density, in order to better approximate inhomogeneous electron densities [35]. This can be done in different ways. Perdew, Burke, and Ernzerhof chose to incorporate a scaling functional to (2.16), i.e.,

$$E_{\text{xc}}^{\text{GGA}}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{\text{xc}}[\rho(\mathbf{r})]F_{\text{xc}}\Big[\nabla\rho(\mathbf{r}), \rho(\mathbf{r})\Big]\mathrm{d}\mathbf{r}. \tag{2.19}$$

Both the local density approximation and the generalized gradient approximation can be refined to also capture the spin-polarization of the electrons [31], [35]. It is possible to also incorporate the kinetic energy densities in the exchange-correlation energy functional; these functionals are known as meta-GGAs [36].

The three exchange-correlation functionals considered in this report are PBE [35], CX [37], and BEEF [38]. The PBE exchange-correlation functional is on the form (2.19), with a scaling factor mostly dependent on the exchange part, on the form

$$F_{\text{x}} = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa}, \tag{2.20}$$

where κ and μ are parameters, and s is dependent of the electron density and Fermi wave vector as

$$s = \frac{|\nabla\rho(\mathbf{r})|}{2k_{\text{F}}\rho(\mathbf{r})} \quad [35]. \tag{2.21}$$

Both CX and BEEF incorporate van der Waals interactions, i.e., additional coulomb interactions, resulting from small shifts in the electron density. The CX functional includes plasmon-response descriptions for both the exchange and correlation. This exchange correlation functional has been proven to be very accurate when predicting binding energies and lattice parameters, especially for systems with non-covalent bonds [37]. The BEEF functional is a Bayesian error estimation functional, with the main advantage of providing computational error estimations [38].

2.4 Neural Networks

McCulloch and Pitts studied, in 1943, how neurons in a network could be used to process information [39]. They stated a list of assumptions, describing how a neural network should be designed, including that the activity of the neuron is an "all-or-none" process, the structure of the neural network does not change with time, and that the only significant delay within the nervous system is synaptic delay. The artificial neuron has later been developed to include a bias and have continuous values of the neuron output, making it possible to, with a sufficiently large network, approximate any continuous function [40]. Mathematically, the output of a McCulloch-Pitts neuron, and the output of a modern neuron, which is used in this work, are expressed as

$$O_{\text{McCulloch-Pitts}} = \theta\Big(\sum \text{input}\Big), \quad \theta - \text{Heaviside function} \tag{2.22}$$

and

$$O_{\text{modern}} = g\Big(b + \sum \text{input}\Big) \tag{2.23}$$

where b is the bias and g is an arbitrary activation function. Common activation functions are e.g. the sigmoid function, the ReLU function, and the tanh function [41], defined as

$$\text{Sigmoid}: \quad \sigma(x) = \frac{1}{1+e^{-x}} \tag{2.24}$$

$$\text{ReLU}: \quad f(x) = x\theta(x) \tag{2.25}$$

$$\tanh: \quad g(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{2.26}$$

The neural network used in this project is a feed-forward neural network consisting of neurons described in (2.23). A feed-forward neural network consists of an input layer, an output layer, and an arbitrary number of hidden layers [42]. There are no connections between neurons in the same layer, and information is only passed forward in the network, towards the output layer. A schematic layout of a feed-forward neural network is shown in Figure 2.4.

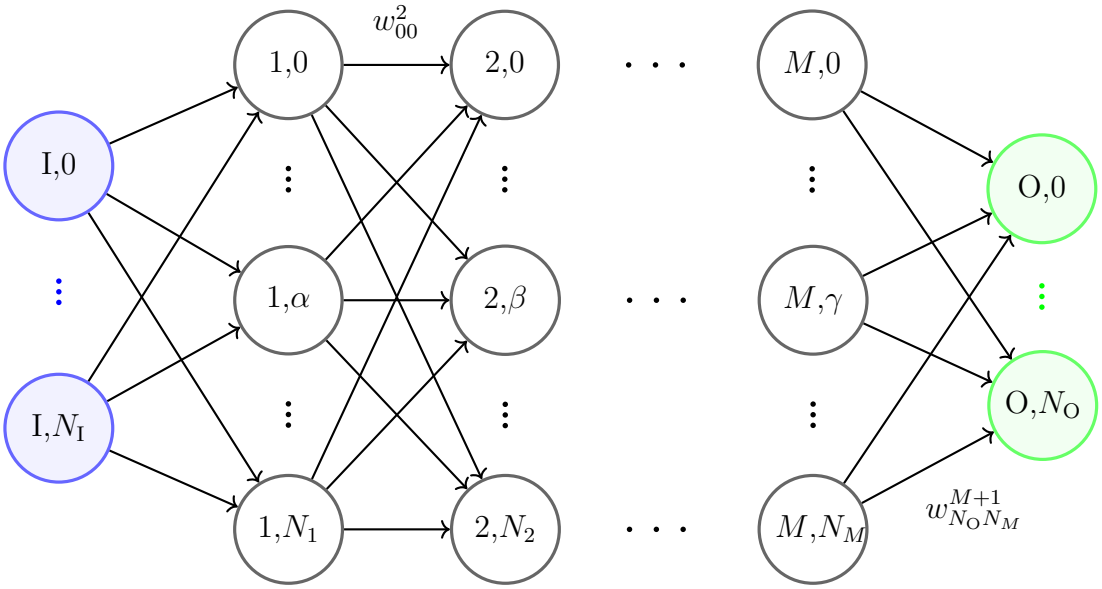


Figure 2.4: A schematic figure of a fully-connected feed-forward neural network.

The value of a neuron n_i^j (the i:th neuron in layer j), with this notation, is given by

$$x_i^j = g\left(b_i^j + \sum_{k=1}^{N_{j-1}} w_{ik}^j x_k^{j-1}\right). \tag{2.27}$$

The training of the neural network is performed with, so called, supervised learning, where the training set consists of labeled data, with input data and corresponding output data [42]. The predicted output of the (fully connected) feed-forward neural network is used to update the weights and biases in the network via back-propagation [43]. The aim is to reduce the *Loss Function* $\mathcal{L}$, which determines the correctness between the predicted output and the output data. The loss function is often chosen to be the squared errors of the complete training set, i.e.,

$$\mathcal{L} = \frac{1}{2}\sum_{i,\mu}\left(y^\mu_{i,\mathrm{predicted}} - y^\mu_{i,\mathrm{data}}\right)^2, \tag{2.28}$$

where the summation is performed over all output neurons, for all training data. Gradient descent is used to update the weights and biases, in order to minimize the loss function, according to

$$w^j_{ik,\mathrm{new}} = w^j_{ik} - \eta\frac{\partial\mathcal{L}}{\partial w^j_{ik}} \tag{2.29}$$

and

$$b^j_{i,\mathrm{new}} = b^j_i - \eta\frac{\partial\mathcal{L}}{\partial b^j_i}, \tag{2.30}$$

which can be solved numerically, or analytically via the chain rule, if the activation functions and the loss function are differentiable.

2.5 Molecular Dynamics

One of the most common molecular dynamics techniques is the *Velocity Verlet* algorithm [44]. It is, in essence, a classical technique derived from Taylor expansions of Newton's equations of motion. For a chosen time step Δt, the algorithm can be described with the following scheme:

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + \frac{1}{2}\mathbf{a}(t)\Delta t \tag{2.31}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \Delta t/2)\Delta t \tag{2.32}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t/2) + \frac{1}{2}\mathbf{a}(t + \Delta t)\Delta t \tag{2.33}$$

Before (2.33), new accelerations are calculated from the forces at that time [45].

3. Methods

In this section, the theory and implementation of the methods used in the project are described in more detail. Here, it is revealed how neural networks can be used as a speed-amplifier, in order to perform a very large amount of molecular dynamics simulations. In addition to the theory and implementation descriptions, the work methodology and simulations are described. A few computational results are presented, required to motivate the choices of structure and computational setup.

3.1 Plane Waves, Cutoff Energies and k-points Sampling

When solving the Kohn-Sham equations (2.13) it is convenient to expand the electronic wave functions into a linear combination of plane waves. For structures with periodic boundary conditions, and periodic potentials, the *Bloch theorem* states that

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{3.1}$$

where $\mathbf{k}$ is the wave vector (a vector in the reciprocal space which can always be confined in the first Brillouin zone due to the periodicity), and $u_{\mathbf{k}}$ is a function with the same periodicity as the structure [46]. The function $\varphi_{\mathbf{k}}$ can be expressed as a Fourier series over the reciprocal space, entailing

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{k}+\mathbf{G}} \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \tag{3.2}$$

where $\mathbf{G}$ are integer linear combinations of the reciprocal base vectors. The expressions in (3.2) can then be substituted into (2.13), to solve the Kohn-Sham equations. The expansion of the Fourier series in (3.2) is truncated by the so called cutoff energy, defined as

$$E_{\text{cutoff}} = \frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 . \tag{3.3}$$

The software used in this project to perform the DFT calculations is *Vienna Ab initio Simulation Package* (VASP), which uses projector-augmented plane waves [47]. The method used in VASP, formulated and implemented by Blöchl, treats the electrons within a certain distance from the nucleus, known as the (semi-)core electrons, differently compared to the valence electrons [48]. The wave functions of the valence electrons are still described by plane waves, but the rapidly oscillating core electrons are described to exist in a smoother, more easily calculated pseudopotential. Since the valence electrons determine all chemical properties, this can be done without loss of accuracy, but at a reduced computational cost.

The sampling of the wave vectors, $\mathbf{k}$, when using plane waves (3.1), is performed with the original Monkhorst-Pack scheme [49]. The k-points are then sampled in the reciprocal space, with the base vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, according to

$$\mathbf{k}_{ijk} = \frac{2p_i - q_i - 1}{2q_i}\mathbf{e}_1 + \frac{2p_j - q_j - 1}{2q_j}\mathbf{e}_2 + \frac{2p_k - q_k - 1}{2q_k}\mathbf{e}_3, \tag{3.4}$$

where q is the number of k-points in each direction, and p ranges from 1 to q. Notice that the Γ-point[1] is included if there is an odd number of k-points in each direction.

3.2 Tools

In this section, the computational tools used in *Atomic Simulation Environment* (ASE) in order to find minimum energy states, vibrational modes, and activation energies between initial and final states, will be described. In many cases, it is necessary to not only determine the energy of the system, but also the forces on each atom, defined as the derivative of the energy with respect to the positions of the nuclei, i.e., the forces on atom i is given by

$$\boldsymbol{F}_i = -\nabla_i E. \tag{3.5}$$

BFGS is a minimization algorithm used for finding the positions and values of the global and local minima in potential energy surfaces [50]. The changes in energy, as a function of the positions of all nuclei $\mathbf{R}$, are calculated iteratively via the forces and the Hessian of the energy. The energy is expanded as a second order Taylor series as

$$E(\mathbf{R}_{\text{new}}) = E(\mathbf{R}_{\text{old}}) + \mathrm{d}\mathbf{R}^{\mathrm{T}}\boldsymbol{F}(\mathbf{R}_{\text{old}}) + \frac{1}{2}\mathrm{d}\mathbf{R}^{\mathrm{T}}H(\mathbf{R}_{\text{old}})\mathrm{d}\mathbf{R}, \tag{3.6}$$

where $\mathrm{d}\mathbf{R} = \mathbf{R}_{\text{new}} - \mathbf{R}_{\text{old}}$ and $\boldsymbol{F}$ is an array containing all forces according to (3.5). The choice of $\mathbf{R}_{\text{new}}$ is determined from the eigenvalues and eigenvectors of the Hessian. If the matrix containing all eigenvectors is notated V, the new position array $\mathbf{R}_{\text{new}}$, is given by

$$\mathbf{R}_{\text{new}} = \mathbf{R}_{\text{old}} + s\,VV^{\mathrm{T}}\boldsymbol{F}, \tag{3.7}$$

where s is a scaling factor, determined by the step length and the eigenvalues of the Hessian. The Hessian matrix can then be updated in several ways, depending on the problem [51]. For all calculations in this report, the minimization is considered converged when the largest force on an atom is below $0.05\,\mathrm{eV/Å}$.

When determining the vibrational modes, the energy of the molecule can be approximated to behave as a harmonic oscillator [52]. The vibrational modes are calculated using finite differences to calculate the Hessian of the energy, with respect to the Cartesian coordinates of all atoms in the molecule. The energy can then be expressed as

$$E = E_0 + \frac{1}{2}\mathrm{d}\mathbf{R}H\mathrm{d}\mathbf{R}, \tag{3.8}$$

[1] The Γ-point is the reciprocal space vector with coordinates $(0, 0, 0)$.

where E_0 is the equilibrium energy and d$\mathbf{R}$ is a vector containing all deviations from the equilibrium positions. If the deviations follow a sinusoidal time evolution Newton's equations of motion yields the eigenvalue equation

$$Hu_k = \omega_k^2 M u_k, \tag{3.9}$$

where M is a matrix, containing the masses of the particles. It now follows that the vibrational modes are the eigenvectors of (3.9). The vibrational frequencies are given by the square-root of the eigenvalues of (3.9). Calculations on the methane molecule yield 15 eigenmodes, of which 6 correspond to translations and rotations (number of degrees of freedom), while the other 9 correspond to vibrations. The eigenmodes corresponding to the translations and rotations have eigenvalues close to 0. When the methane molecule is positioned at a saddle point in the potential energy surface, one of the eigenvalues will be imaginary.

In order to find the transition paths and, in particular, the activation energy of a reaction, the *Nudged Elastic Band* (NEB) method was used [53]. More specifically, a refined version of this method, using a climbing image was utilized [54]. The initial state and the final state are both local minima, and are held fixed in the NEB method. It is between these two states that the elastic band is spanned. A number of images are interpolated between the initial and final states, as an initial guess of the minimum energy transition path. The large advantage with NEB, compared to previous methods, is that due to the elastic band, the path can be more uniformly sampled, entailing the entire transition path to be well described [53]. The activation energy is a saddle point in the potential energy surface; a maximum in the tangential direction of the band, and a minimum in the perpendicular direction. In the optimization process, the forces acting on each image in the interpolated path are the tangential spring force from the band and the force perpendicular to the band from the DFT calculations. In the climbing image case, the spring force from the band on the highest energy image is neglected after a certain number of iterations [54]. The force is then purely determined from the potential energy surface. Consequently, the activation energy is more converged, while the other images still describe the transition path.

3.3 Atomistic Machine-learning Package

The software used to build, train and apply the neural network was the *Atomistic Machine-learning Package* (AMP), developed by Alireza Khorshidi and Andrew A. Peterson at Brown University [55]. It is integrated with ASE in such a way that it can be used as a regular calculator. The neural network uses the methods described in Section 2.4; it is a feed-forward neural network, utilizing back-propagation to update the weights and biases. However, the input to the neural network is not whole structures, but feature vectors, generated for each atom in the system. This enables the network to be applied on different structures, of different sizes, as long as the elements are the same.

The output of the system is the energy of the system, and the forces on each atom,

equivalent to the definition in (3.5). The total energy of the system E_{system}, determined by the neural network, is calculated as the sum of the energy contributions of each atom, which in turn depends on the element and the local environment around the atom, i.e.,

$$E_{\text{system}} = \sum_{i=1}^{N} E^{\text{local}}(\mathbf{r}_i), \tag{3.10}$$

where N is the number of atoms in the system. To capture the features, and make it easier for the neural network to determine the energies of these local environments, radial symmetry functions based on Gaussian functions are used [56]. The radial symmetry functions for an atom are created by pair interactions and triplets interactions with the other atoms. The radial symmetry functions for atom i corresponding to pair interactions are given by

$$G_i^{\text{pair}} = \sum_{j \neq i} \mathrm{e}^{-\eta r_{ij}^2} f_c(r_{ij}), \tag{3.11}$$

where η is a parameter and r_{ij} is the distance between atom i and atom j. The function f_c is a cutoff function, determining how large of an impact the energy contribution has. It is limited by a cutoff radius R_c, creating a sphere, which is considered as the local environment. The cutoff function is defined as

$$f_c(r) = \begin{cases} \frac{1}{2}\Big(1 + \cos\frac{\pi r}{R_c}\Big), & \text{if } r \leq R_c, \\ 0, & \text{if } r > R_c. \end{cases} \tag{3.12}$$

The radial symmetry functions for atom i corresponding to triplet interactions are given by

$$G_i^{\text{triplet}} = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos\theta_{ijk})^{\zeta} \mathrm{e}^{-\eta(r_{ij}^2 + r_{ik}^2 + r_{jk}^2)} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}), \tag{3.13}$$

where θ_{ijk} is the angle between the three atoms, $\lambda = \pm 1$, and ζ is a parameter. These symmetry functions are known as fingerprints [57]. A feed-forward neural network is created for each element, where the inputs are the radial symmetry functions, i.e., the feature vectors. For methane on a palladium oxide surface, this corresponds to four different neural networks. The neurons in the neural network then regress the fingerprints to energies and the derivatives of the fingerprints to forces. The total energy of the system is then evaluated according to (3.10). A schematic overview of the whole process can be seen in Figure 3.1.

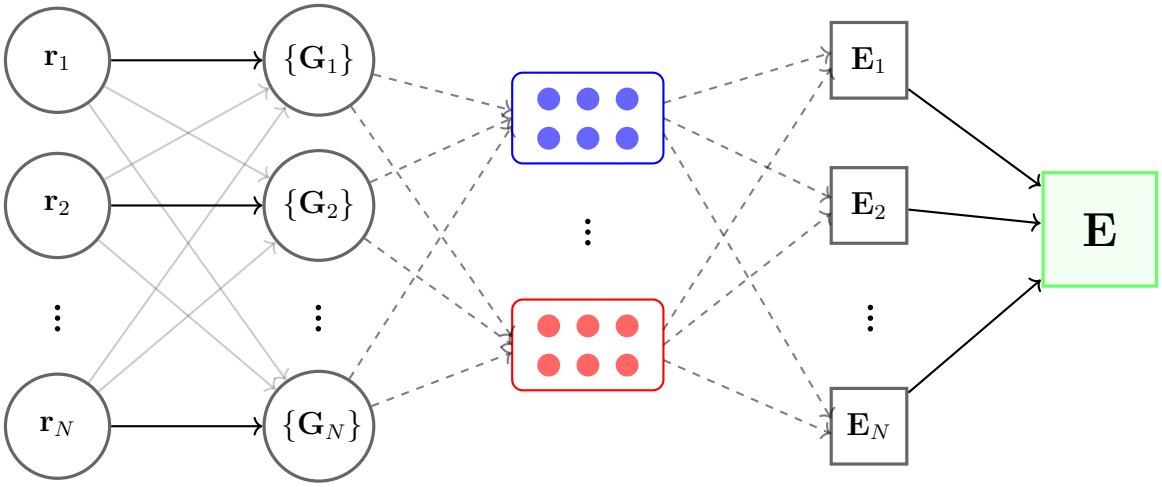


Figure 3.1: A schematic figure of the energy (and force) calculations using the neural network approach. For each atom, feature vectors (and their derivatives) are calculated and fed to the corresponding neural network. The output energy for all individual atoms are summarized to obtain the total energy of the system.

During the training, the energies and forces are compared to the true values generated by DFT, via a loss function. The loss function that was used is in the same form as (2.28), dependent on the energies and forces as

$$\mathcal{L} = \frac{1}{2}\sum_{\mu=1}^{N_\mu}\left(\frac{1}{N^2}\left(E_{\mu,\mathrm{Amp}} - E_{\mu,\mathrm{DFT}}\right)^2 + \frac{\alpha}{3N}\sum_{i=1}^{N}\sum_{k=1}^{3}\left(F_{\mu,i,k,\mathrm{Amp}} - F_{\mu,i,k,\mathrm{DFT}}\right)^2\right), \quad (3.14)$$

where N_μ is the number of training images, and α is a scaling factor to determine the relative impacts between the energy loss and the force loss.

3.4 Determination of Cutoff Energy and number of k-points for the calculations

To determine the required cutoff energy and amount of k-points in each direction (according to the Monkhorst-Pack scheme) for the DFT calculations, convergence tests for the lattice parameters were performed. The primitive cell of palladium (FCC) was built with periodic boundary conditions. In practice, this is equal to an infinitely large bulk structure of the metal in each direction. The energies of the metal with 10 different lattice parameters uniformly sampled between 3.7–4.1 Å were calculated for a range of different cutoff energies and number of k-points. The number of k-points ranged between 4 and 15, and the cutoff energy ranged between 200–650 eV. The energy of the bulk metal as a function of lattice parameter is believed to behave as a harmonic oscillator. Therefore, for a given value of the cutoff energy and the number of k-points, a second-order polynomial regression of the lattice parameter versus energy was performed, in order to find the lattice

parameter of the most stable structure. This was done for the three exchange-correlation functionals PBE, CX, and BEEF. The lattice parameters which minimize the energy as a function of cutoff energy and the number of k-points for PBE, CX and BEEF are shown in Figure 3.2, Figure 3.3 and Figure 3.4, respectively. The lattice parameter is considered converged when a cutoff energy of 450 eV and 12 k-points are used. In addition to the convergence of the lattice parameter, the cohesive energy of FCC palladium was calculated, in order to compare the three exchange-correlation functionals. The cohesive energy for palladium is defined as

$$E_{\text{coh}}^{\text{Pd}} = E_{\text{atom}}^{\text{Pd}} - E_{\text{FCC}}^{\text{Pd}}. \tag{3.15}$$

The energy of a vacuum box of size (10, 11, 12) Å, containing one palladium atom, was calculated using two k-points in each direction in the reciprocal space, and a cutoff energy of 450 eV.

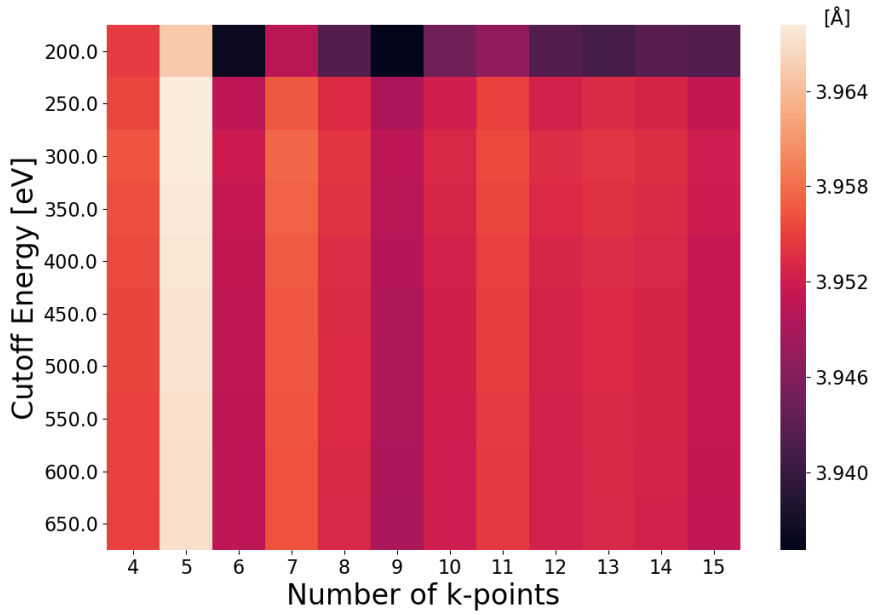


Figure 3.2: The lattice parameter as a function of the cutoff energy and the number of k-points for the DFT calculations using the PBE exchange-correlation functional.

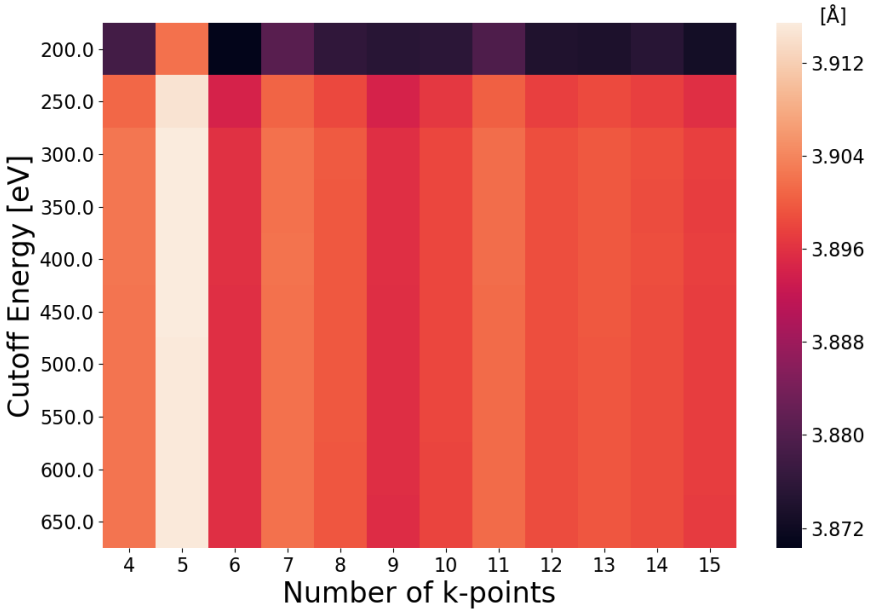


Figure 3.3: The lattice parameter as a function of the cutoff energy and the number of k-points for the DFT calculations using the CX exchange-correlation functional.

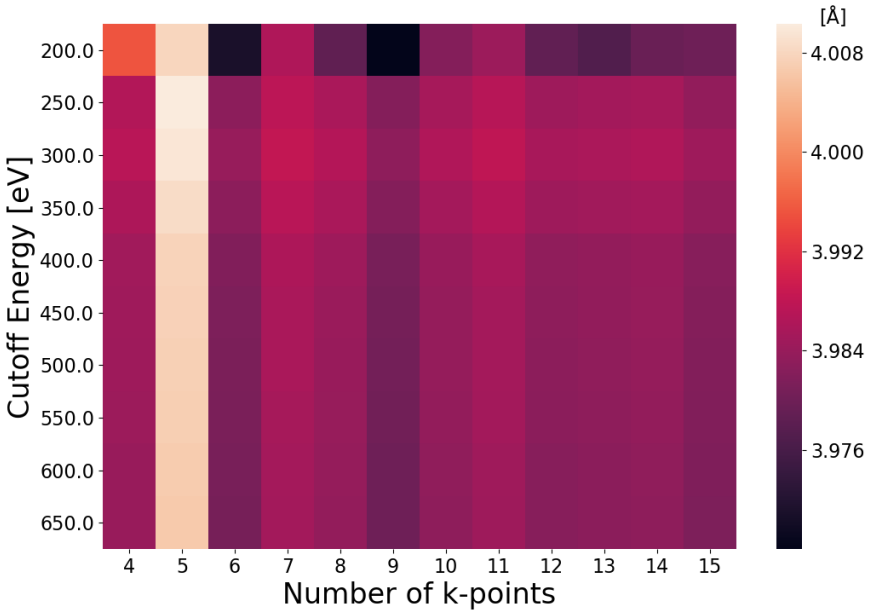


Figure 3.4: The lattice parameter as a function of the cutoff energy and the number of k-points for the DFT calculations using the BEEF exchange-correlation functional.

The converged lattice parameters, which minimized the energy, using a cutoff energy of 450 eV and 12 k-points, as well as the cohesive energy of FCC palladium, for the three exchange-correlation functionals, are shown in Table 3.1.

Table 3.1: The converged lattice parameters, as well as the cohesive energies, for FCC palladium using the three different exchange-correlation functionals.

XC-functional	Lattice Parameter [Å]	Cohesive Energy [eV]
PBE	3.953	3.75
CX	3.899	4.36
BEEF	3.983	3.21
Experimental	3.89	3.91

Analogous calculations were performed for palladium oxide. The primitive cell was constructed with 10 base parameters uniformly sampled between 2.5–3.5 Å, and 10 height parameters uniformly sampled between 4.8–5.8 Å, with periodic boundary conditions. The relative distances between the 4 atoms in the primitive cell were held constant. The number of k-points ranged between 4 and 15, and the cutoff energy ranged between 200–650 eV. In this case, 100 lattice parameter pairs, with corresponding energies, were calculated for each cutoff energy and number of k-points. The energy contribution from the two lattice parameters could not with safety be considered as decoupled; therefore, a six-dimensional polynomial regression (including the cross-term) was performed to find the parameter pair yielding the lowest energy. The lattice parameter pairs minimizing the energy of the palladium oxide structure for PBE, CX and BEEF are shown in Figure 3.5, Figure 3.6 and Figure 3.7, respectively. In this case, it is easier to see when the lattice parameters are converged, namely, when a cutoff energy of 450 eV and 12 k-points are used. For the tetragonal palladium oxide structure the formation energy, defined as

$$E_{\text{form}}^{\text{PdO}} = E_{\text{FCC}}^{\text{Pd}} + \frac{1}{2}E_{\text{gas}}^{\text{O}_2} - E_{\text{tetra}}^{\text{PdO}}, \tag{3.16}$$

was calculated with the different exchange-correlation functionals. The energy of a vacuum box of size (15,16,17) Å, containing an oxygen molecule, was calculated using two k-points in each direction, and a cutoff energy of 450 eV.

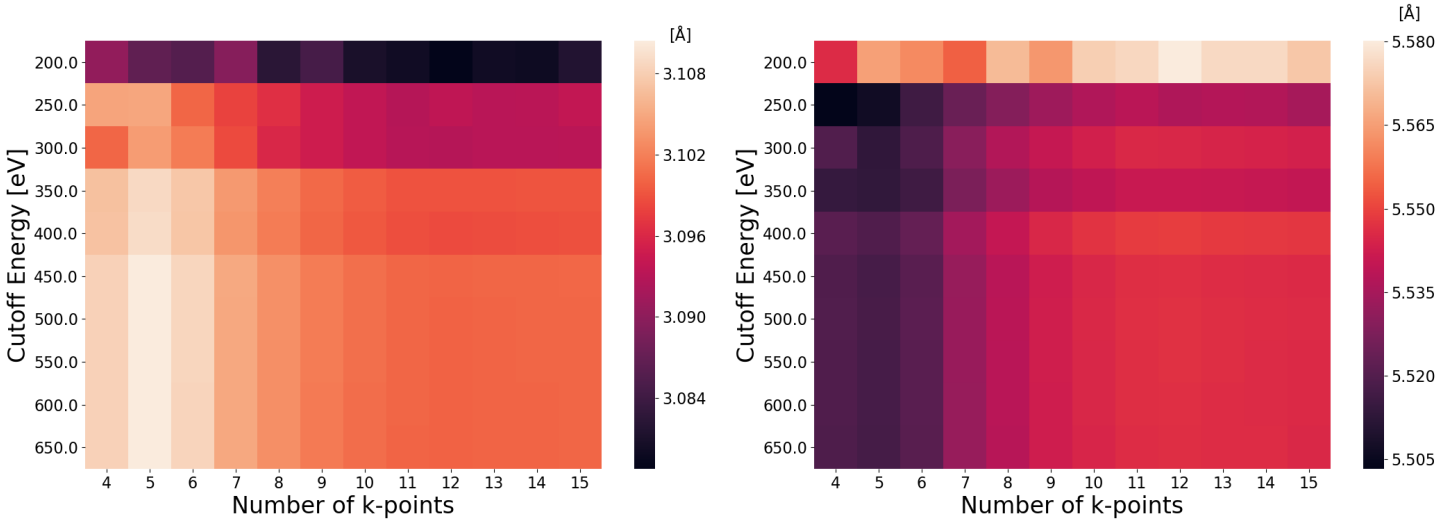


Figure 3.5: The base and the height of the tetragonal structure that minimize the energy, as a function of the cutoff energy and the number of k-points for the DFT calculations using the PBE exchange-correlation functional.

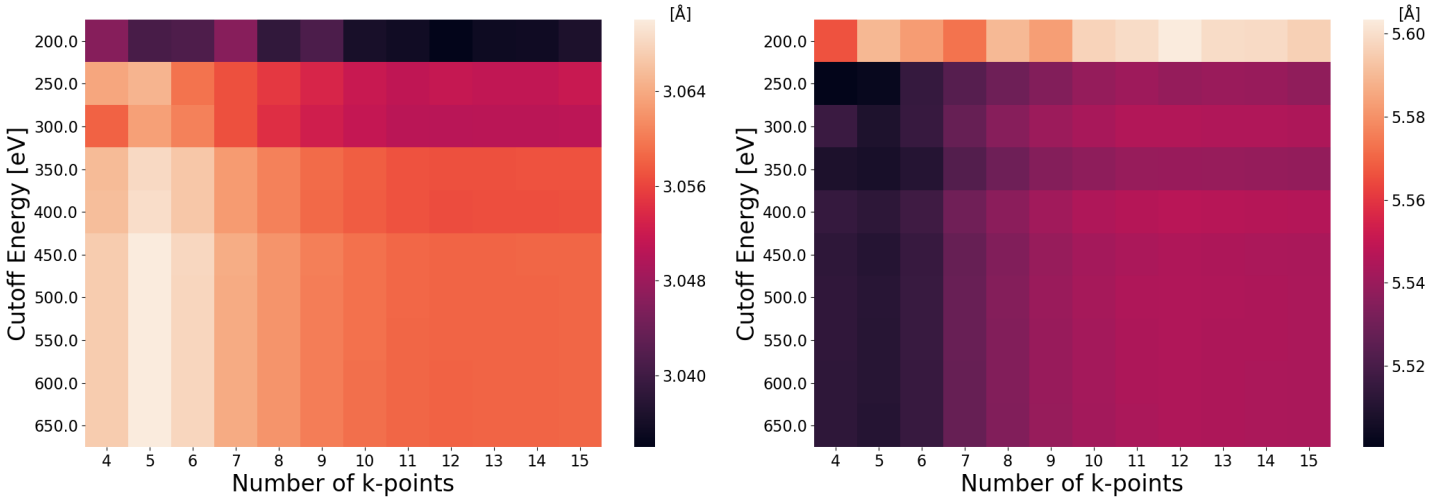


Figure 3.6: The base and the height of the tetragonal structure that minimize the energy, as a function of the cutoff energy and the number of k-points for the DFT calculations using the CX exchange-correlation functional.

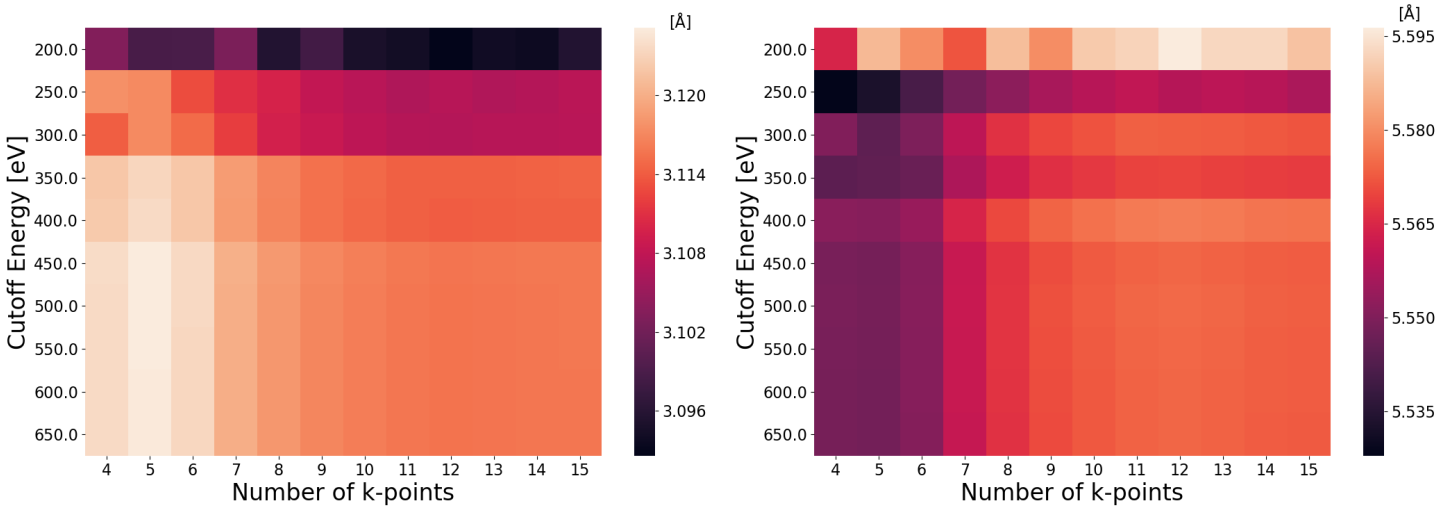


Figure 3.7: The base and the height of the tetragonal structure that minimize the energy, as a function of the cutoff energy and the number of k-points for the DFT calculations using the BEEF exchange-correlation functional.

The converged lattice parameter pairs, which minimized the energy, using a cutoff energy of 450 eV and 12 k-points, as well as the formation energy of palladium oxide, for the three exchange-correlation functionals are shown in Table 3.2.

Table 3.2: The converged lattice parameter pairs and the formation energies, for bulk palladium oxide using the three different exchange-correlation functionals.

XC-functional	Base [Å]	Height [Å]	Formation Energy [eV]
PBE	3.100	5.547	1.19
CX	3.058	5.545	1.31
BEEF	3.116	5.574	1.20
Experimental	3.043	5.336	1.22

3.5 Vibrational Modes of Methane

The vibrational modes of the methane molecule in gas form, (a (15,16,17) Å vacuum box containing one molecule), were studied using the three different exchange-correlation functionals. The energies and movements (rotations, translations and vibrations with corresponding symmetry group) are presented in Table 3.3. The vibrational modes 1–6, with energy eigenvalues close to 0, correspond to transla-

tions and rotations. Vibrational modes 7–15 correspond to different vibrational movements between the carbon atom and the hydrogen atoms.

Table 3.3: The vibrational modes for the methane molecule in a (15,16,17) Å vacuum box, calculated using the three different exchange-correlation functionals. The energies of the vibrational modes are expressed in cm^{-1}, with the movement described in the rightmost column.

Mode	Energy [cm^{-1}] (PBE)	Energy [cm^{-1}] (CX)	Energy [cm^{-1}] (BEEF)	Movement
1	135.4 i	82.62 i	39.24 i	Rotation
2	124.8 i	68.09 i	33.05 i	Rotation
3	111.4 i	58.29 i	15.27 i	Rotation
4	30.23 i	36.13 i	126.2	Translation
5	20.00 i	20.70 i	166.4	Translation
6	10.25	29.78	186.3	Translation
7	1281	1290	1336	Vib. f_2
8	1289	1299	1338	Vib. f_2
9	1291	1302	1343	Vib. f_2
10	1502	1503	1549	Vib. e
11	1503	1505	1549	Vib. e
12	2939	2891	2967	Vib. a_1
13	3092	3064	3114	Vib. f_2
14	3101	3065	3120	Vib. f_2
15	3101	3066	3123	Vib. f_2

3.6 Building the Structures

In order to determine the structure, five candidates were constructed in ASE, and certain properties were inspected. The five structures were

- 1 layer - $\sqrt{5}$: 5 layers of 2×2 primitive cells palladium (100) with 1 layer of 1×2 primitive cells palladium oxide (101) on top, according to the $(\sqrt{5} \times \sqrt{5})$R27° structure [21].

- 2 layers - $\sqrt{5}$: 5 layers of 2×2 primitive cells palladium (100) with 2 layer of 1×2 primitive cells palladium oxide (101) on top, according to the $(\sqrt{5} \times \sqrt{5})$R27° structure [21].

- 7 layers PdO: 7 layers of 1×2 primitive cells of palladium oxide (101).
- 5 layers PdO: 5 layers of 1×2 primitive cells of palladium oxide (101).
- 5 layers PdO - fixed: 5 layers of 1×2 primitive cells of palladium oxide (101) with the atoms in the bottom layer fixed in their positions.

All structures were periodic in the x, y and z-directions, but separated by 30 Å of vacuum in the z-direction. Physically, this means that the surface is infinite in the x, y-plane; and there are an infinite amount of such surfaces in the z-direction. This is perfectly possible due to the periodicity of the projector augmented plane waves described in Section 3.1. All structures were built with the lattice parameters in Table 3.1 and Table 3.2. However, for all structures, and all exchange-correlation functionals, the structures were optimized to minimize the energy; this is especially important for the $(\sqrt{5} \times \sqrt{5})$R27° structures, due to the interface between palladium and palladium oxide. The structures were optimized with the BFGS algorithm, using a cutoff energy of 450 eV and $(6, 6, 1)$ k-points according to the Monkhorst-Pack scheme (3.4). The choice of cutoff energy and k-point sampling is based on the convergence from Section 3.4. Since the size of the structure in the x, y-plane is approximately two times that of the primitive cells of palladium and palladium oxide, half as many k-points are needed in the reciprocal space. Similarly, the reciprocal space is as well sampled for the structure when only one k-point is used in the z-direction.

3.7 Initial State

When the surface structures had been relaxed to their minimum-energy configuration, the methane molecule was added to the system. To find the adsorption position of the methane molecule on the surface, i.e., the global minimum of the potential energy surface, the BFGS algorithm was used once more. From above, and sideways from, the adsorption position, the potential energy surface is very flat, thus making it difficult for the minimization algorithm to converge to the correct energy. Therefore, the methane molecule was initialized close to the surface in the exponentially repelling region of the potential energy surface. The exponential wall in the 1-dimensional potential energy surface is shown in Figure 4.6. The strong repulsive force acting between the surface and the methane molecule makes it easier for the BFGS algorithm to find the global energy minimum. The adsorption energy and hydrogen-carbon-hydrogen angle for all structures and exchange-correlation functionals can be seen in Table 3.4.

Table 3.4: The Adsorption energy and H−C−H bond angle for the five potential structures and three potential exchange-correlation functionals.

Structure	XC-functional	Adsorption Energy [eV]	H-C-H Angle [°]
1 layer - $\sqrt{5}$	PBE	-0.022	109.6
1 layer - $\sqrt{5}$	CX	-0.201	110.6
1 layer - $\sqrt{5}$	BEEF	-0.0970	110.2
2 layers - $\sqrt{5}$	PBE	-0.130	114.4
2 layers - $\sqrt{5}$	CX	-0.429	115.4
2 layers - $\sqrt{5}$	BEEF	-0.177	112.2
7 layers PdO	PBE	-0.170	114.9
7 layers PdO	CX	-0.479	116.4
7 layers PdO	BEEF	-0.186	112.8
5 layers PdO	PBE	-0.159	115.3
5 layers PdO	CX	-0.448	116.2
5 layers PdO	BEEF	-0.182	112.8
5 layers PdO - fixed	PBE	-0.182	114.9
5 layers PdO - fixed	CX	-0.469	116.5
5 layers PdO - fixed	BEEF	-0.192	112.9

3.8 Choice of Structure and Exchange-Correlation Functional

Since the density functional theory calculations are very computationally expensive, it would not be possible to consider all structures and exchange-correlation functionals in a reasonable time frame. Therefore, this section will be dedicated to explain the reasoning behind the choice of structure and exchange-correlation functional.

In Table 3.4 it can be seen that the BEEF functional has problems with the activation of the methane molecule at the adsorption position. Even though van der Waals interactions are accounted for, the adsorption energies are relatively small. The angles between the two saddling hydrogen atoms are considerably smaller, indicating that the adsorption of methane on the surface was not satisfactorily calculated. Therefore, since the adsorption is the most important process, the BEEF functional was discarded. Both the PBE and CX exchange-correlation functionals show satisfactory activation of the methane molecule and reasonable activation energies. The

same can be concluded when studying the vibrational modes of methane, seen in Table 3.3. The lattice parameters in Table 3.1 and Table 3.2 are in better agreement with experimental values when using the CX functional, while the cohesive energy and the formation energy are in better agreement when using the PBE functional. The choice of exchange-correlation functional fell on the CX functional, due to the larger angle and adsorption energy of the methane molecule on the surfaces.

The choice of structure was solely based on the results from Table 3.4. The adsorption energy and activation of the methane molecule, the H−C−H angle, are considerably larger when studying the 2 layers - $\sqrt{5}$ structure compared to the 1 layer - $\sqrt{5}$ structure. This is not strange, since the 1 layer - $\sqrt{5}$ structure does not contain any 3-fold coordinated palladium atoms, i.e., a palladium atom with one oxygen atom on each side and one directly beneath, shown in Figure 2.1. 2 layers - $\sqrt{5}$, 7 layers PdO, 5 layers PdO, and 5 layers PdO - fixed all show very similar adsorption energies and H−C−H angles; with somewhat larger absolute values when any pure palladium oxide was studied. Since all structures seemed satisfactory, the structures with the lowest computational cost were prioritized, i.e., 5 layers PdO and 5 layers PdO - fixed. In order to make the surface more bulk like, the structure with fixed atomic positions in its bottom layer was chosen. This simplifies the calculation a bit, and also makes the surface more robust, since it will not move in the collision with incoming methane molecules. Therefore, from here on, only the exchange-correlation functional CX and the 5 layers PdO - fixed structure were used.

3.9 Activation Energy

In order to determine the activation energy of the methane dissociation and the minimum energy transition path (see Figure 2.2), the final state of the structure needed to be determined. The BFGS algorithm was used, similar to when the initial state was found. The energy of the final state was found to be 0.40 eV lower than that of the initial state. With the initial and final states of the dissociation reaction determined, the climbing image nudged elastic band method was applied in order to find the minimum energy transition path and the activation energy. Between the initial and final states, seven images were interpolated, over which the elastic band was spanned. The energies and forces in the nudged elastic band method were minimized with the BFGS algorithm, until the convergence criteria were met. The criteria were the same as for the systems; a converged result is obtained when the forces on each atom, in all images, are below 0.05 eV/Å. To check whether the correct transition path, and activation energy, was found, a vibrational analysis was performed on the methane molecule at the saddle point in the reaction path. If the methane molecule is, in fact, in a saddle point in the potential energy surface, the molecule should have only one imaginary vibrational mode. The results from the nudged elastic band calculations and the vibrational analysis are shown in Section 4.1

3.10 Training data

It is impossible to sample the whole 135-dimensional space to be used as training data for the neural networks. Therefore, the physically most probable, and important, positions were sampled first. Firstly, the atoms in the surface and methane molecule were thermally disordered from a Maxwell-Boltzmann distribution, entailing an artificial temperature around 800 K, after which the Velocity Verlet algorithm was applied. DFT simulations were performed to calculate the trajectories with the thermal vibrations captured. The positions of all atoms in a system, the energy of the system, and the 135 partial derivatives were stored for each time step in the Velocity Verlet algorithm. The same was done for the dissociated methane molecule, i.e., free methyl and hydrogen moving on the surface. Transition paths, including the minimum energy transition path, for the dissociation process were determined with the use of nudged elastic bands. Since the molecular dynamics and the transition paths are biased towards certain positions and configurations, methane molecules with random orientation were sampled in the regions with low sampling rate, in order to get a more well sampled space. The training data set consisted of a total of 20898 systems, with corresponding energies and forces. Each system was built with $1 \times 2 \times 5$ primitive cells of palladium oxide (101) and one methane molecule, i.e., 20 palladium atoms, 20 oxygen atoms, one carbon atom and four hydrogen atoms. The distribution of the carbon atom in the methane molecule in the training data set can be seen in Figure 3.8.

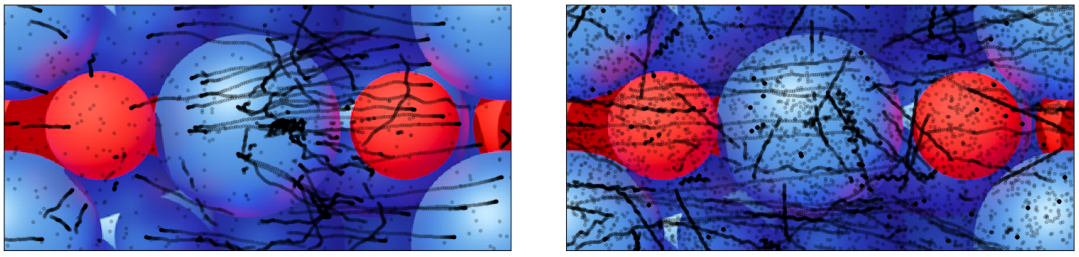


Figure 3.8: The position of the carbon atoms in the training data set. The left most figure shows the distribution in the x, y-plane for methane molecules 0–3 Å above the surface in the z-direction with respect to the centered palladium atom. The rightmost figure is the distribution of training data in the x, y-plane for methane molecules 3–6 Å above the surface in the z-direction, with respect to the centered palladium atom.

The leftmost figure in Figure 3.8 shows the close distance region between the methane molecule and the surface, between 0–3 Å in distance in the z-direction. Since the palladium oxide (101) surface has a wave-shaped structure (see Figure 2.1), the distance is ambiguous; hence the palladium atoms are set as the reference in the z-direction. Due to the reference, it can be seen that the number of sampled methane molecules is rather small above the left oxygen atom, since the position of that atom is above the palladium atoms, and the exponential wall, i.e., where the repelling forces are very large, becomes more prominent at larger z-positions. The rightmost figure

shows the carbon positions of the methane molecules with a distance of 3–6 Å to the reference palladium atoms in the surface in the z-direction. This space is heavier sampled, since all positions in this region are possible, unlike the close range region, where the methane molecule is repelled close to the surface. Even though the distance between the methane molecules and the surface is larger, this data is also important when the neural network predicts the energies in the close range regions, since interactions with palladium atoms and oxygen atoms in the inner layers of the surface are also accounted for.

3.11 Construction of the Neural Network

The neural network, used for predicting the potential energy surface and its gradients, was constructed using the AMP software. For each atom, a cosine cutoff function with a cutoff radius of 6.5 Å was used, according to (3.12). The number of symmetry functions was chosen to be 56, for each atom; 16 pair interaction functions (3.11) and 40 triplet interaction functions (3.13). The parameters of the Gaussian functions were the same for each element. For the four pair-interactions, the η values were 0.05, 0.23, 1.07 and 5, resulting in 16 total functions. For the triplet-interaction functions, η was always 0.005, and the four combinations of (ζ, γ) where $(1, \pm 1)$ and $(4, \pm 1)$. The different (ζ, γ), ranging over all possible combinations of triplets for a given element, equals 40 triplet-interaction functions. The feature vectors, and the gradients, were calculated for each system in the training data set.

For each element, a fully connected, feed-forward neural network was constructed. The number of input layers for each neural network is the same as the number of feature vectors. The energy output corresponds to one output neuron, and the gradient of the energy corresponds to three output neurons. To ease the interpolation between the feature vectors and the outputs, each network was constructed with two hidden layers, each containing 20 neurons. A few different architectures were also tested, including 2×10, 3×10, 3×20, 2×40 and 2×64 hidden neurons. However, when the number of hidden neurons increases, the number of weights increases approximately quadratically, and the training time increases rapidly. After training each of the network on approximately half the training data set, only including systems with temperatures between 0–425 K, the optimal architecture was chosen to be 2×20 hidden neurons for all four neural networks, due to the relatively fast training, decreased risk of overfitting, and large reduction of the loss function. The activation function for each neuron was chosen to be the tanh function. The loss function of the combined neural networks (i.e., sum of the four networks) is given in (3.17) as

$$\mathcal{L} = \frac{1}{2} \sum_{\mu=1}^{N_\mu} \left(\frac{1}{N^2} \Big(E_{\mu,\mathrm{Amp}} - E_{\mu,\mathrm{DFT}} \Big)^2 + \frac{\alpha}{3N} \sum_{i=1}^{N} \sum_{k=1}^{3} \Big(F_{\mu,i,k,\mathrm{Amp}} - F_{\mu,i,k,\mathrm{DFT}} \Big)^2 \right), \quad (3.17)$$

where $N_\mu = 20898$ is the number of training images, $\alpha = 0.05$, and $N = 45$ is the number of atoms in the system. The combined neural network was trained for 1200 epochs, after which, the RMSEs for the energies and forces, for the entire systems, were 0.103 eV, and 0.174 eV/Å with respect to the training data, respectively.

3.12 Molecular Dynamics Simulations

In order to predict the ratio between the number of dissociated methane molecules on the surface and the number of incoming methane molecules, a large number of molecular dynamics simulations had to be performed. The forces used to perform the Velocity Verlet algorithm, i.e., the gradients of the potential energy surface, were predicted with the trained neural network. However, in order to obtain a satisfactory number of trials for the whole process, i.e. the adsorption and the dissociation reaction, with as small a time step as possible, some approximation had to be made. The event was divided into two parts. Firstly, the probability of adsorption on the surface was determined. Secondly, the probability of a dissociation reaction was determined. This can be done without loss of accuracy, since the only positions where the methane molecules can be adsorbed on the surface is around the initial position of the dissociation reaction (see Figure 4.7). Since all dissociation reactions must follow an adsorption process, the probability that an incoming methane molecule is adsorbed, and dissociated, on the surface, can be determined by the product of the two probabilities. Additional predictions for the activation energy of the dissociation process and determination of the potential energy surface were also performed, partly to check the accuracy of the predictions, partly to obtain results that would otherwise be too computationally expensive to obtain.

4. Results

Results are sometimes very difficult to interpret; but sometimes a simple plot can explain everything one wants to know. Here, the main results will be presented and interpreted. It starts with results from the dissociation process, and continues with the performance and results obtained by the neural network. The probability of adsorption and dissociation, combined with a potential new dissociation event are presented. Lastly, a brief discussion of possible errors and corrections is presented.

4.1 The Dissociation Process

The energy required to dissociate the methane molecule into a methyl molecule and a free hydrogen atom without the use of a catalyst was calculated to be 6.63 eV. The potential energy as a function of bond length is shown in Figure 4.1. The bond distance in the methane molecule that minimizes the energy is 1.09 Å. The behaviour of the energy as a function of bond length also strengthens the argument that the energy can be expressed as a harmonic oscillator when the vibrational analysis is performed.

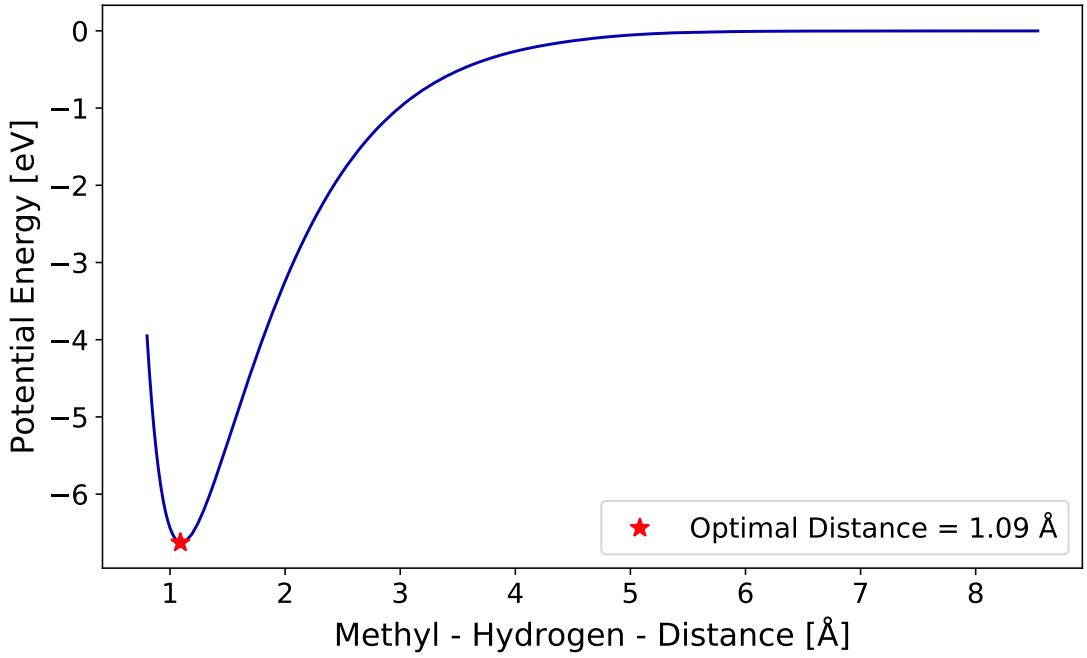


Figure 4.1: The potential energy of the methane molecule as a function of bond length between the methyl molecule and the hydrogen atom.

The results from the nudged elastic band calculations, minimized with the BFGS algorithm, are shown in Figure 4.2. The minimum energy transition path with corresponding activation energy calculated with DFT is shown in the leftmost figure. Note that the potential energy is expressed with the initial state as reference. The activation energy was determined to be 0.657 eV. Since the adsorption energy was −0.469 eV, the apparent activation energy of the dissociation reaction was 0.188 eV. The energy of the final state was 0.401 eV lower than the energy of the initial state. The middle figure shows the neural network predicted energies and forces for the minimum energy transition path determined from the DFT calculations. The neural network predicted an activation energy of 0.750 eV, and an energy difference between the initial and final state of 0.384 eV. In the rightmost figure, the minimum energy transition path with corresponding activation energy determined purely with the neural network is shown. The activation energy of the minimum energy transition path was determined to be 0.693 eV. The neural network predicted an adsorption energy of −0.434 eV, and a final state energy 0.384 eV lower than the initial state energy. This means that the apparent activation energy predicted by the neural network was 0.259 eV. The neural network also predicted the H−C−H angle to be 116.3°. This agrees well with the DFT calculated angle 116.5°.

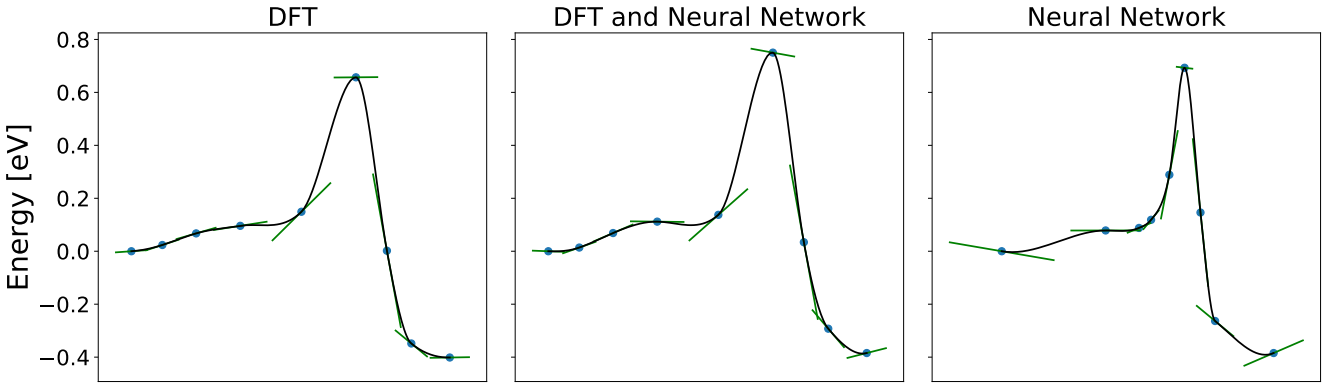


Figure 4.2: The minimum energy transition paths and activation energies calculated with nudged elastic bands. The leftmost figure shown the transition path calculated with DFT. The middle figure used the same transition path as the left figure, but the energies and forces from the band are calculated with the neural network. The rightmost figure shows the minimum energy transition path calculated purely with the neural network.

To ensure that the determined activation energy was the true saddle point in the potential energy surface, a vibrational analysis was done for the methane molecule. The vibrational modes of the methane molecule at the assumed saddle point are shown in Table 4.1. One vibrational mode has a large imaginary energy eigenvalue, which in large captures the reaction path. The second vibrational mode, with a low imaginary energy eigenvalue, has captured parts of the movement in the reaction path. However, the results from the vibrational analysis ensure that the state is positioned very close to a saddle point in the potential energy surface.

Table 4.1: The vibrational modes with corresponding energies for the methane molecule positioned at the transition state.

Mode	Energy [cm^{-1}]
1	1266 i
2	73.27 i
3	106.4
4	209.3
5	468.9
6	569.8
7	687.2
8	1024
9	1144
10	1375
11	1390
12	1598
13	2913
14	2988
15	3029

4.2 Neural Network Performance

The neural network, with 2×20 hidden neurons and 56 feature vectors per element, predicted the energies approximately five orders of magnitude faster than DFT calculations, and the forces approximately three order of magnitude faster than DFT calculations. The difference in time gain is due to the computational differences between DFT and the neural network. When using DFT, the obtained electron density, used to determine the energy can, according to the Hellmann-Feynman theorem, be used to determine the forces as well, without much computational effort. However, when using neural networks, the derivatives of the feature vectors, with respect to all atoms and coordinates, need to be determined to obtain the forces in the system. For a test data set, containing 4387 systems at 0 K, the energy RMSE was determined to be 0.0254 eV, the energy MAE was determined to be 0.0196 eV, and the Pearson correlation coefficient was determined to be 0.996. It is common to see the energy errors expressed per atom, rather than the whole system. In that case the energy RMSE per atom was determined to be 0.564 meV per atom, and the

energy MAE 0.436 meV per atom. Note that this is for systems at 0 K, and that the energy RMSE tends to increase with a factor 3-4 when the temperature of the systems is increased to 800 K.

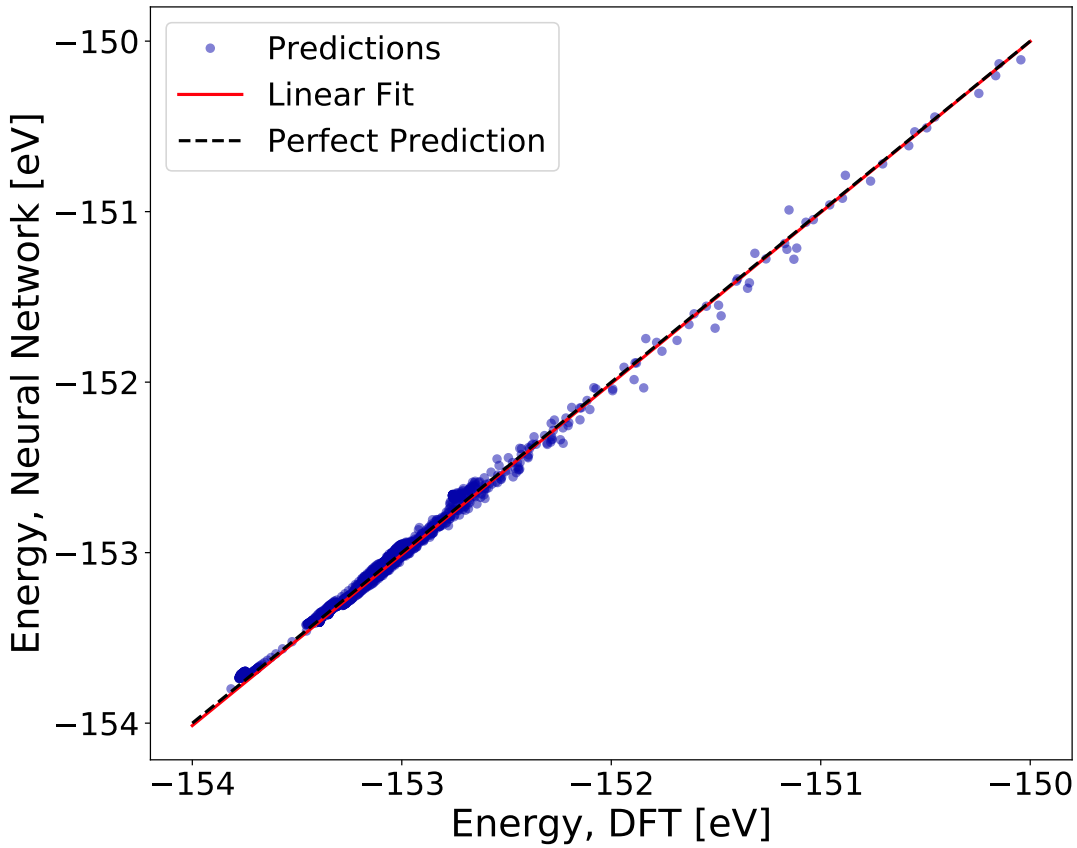


Figure 4.3: The neural network predicted energies compared to the DFT-calculated energies.

Table 4.2: The root-mean-square error, the mean absolute error and the Pearson correlation coefficient for the test set. The errors are given for the whole system, and the errors per atom is hence obtained if the values are divided by 45.

RMSE	MAE	r
0.0254 [eV]	0.0196 [eV]	0.996

An additional test to determine the accuracy of the neural network was done by comparing the energies of DFT calculated trajectories of the system. In the trajectories, the methane molecule and the surface were initially disordered following a Maxwell-Boltzmann distribution corresponding to 600 K. The methane molecules were initialized at random positions 4.827 Å above the surface, relative to the cen-

tered palladium atoms. 300-400 steps were performed with the Velocity Verlet algorithm, with the time step 1 fs. Six different trajectories were studied, and the energy errors between the DFT calculations and the neural network predictions were calculated. In Figure 4.4 the potential energies, relative to the free methane molecule and the surface at 0 K, for the six trajectories are shown. The total energy MAE and the energy RMSE, for all 2076 points in the six trajectories, were determined to be 0.101 eV and 0.131 eV, respectively.

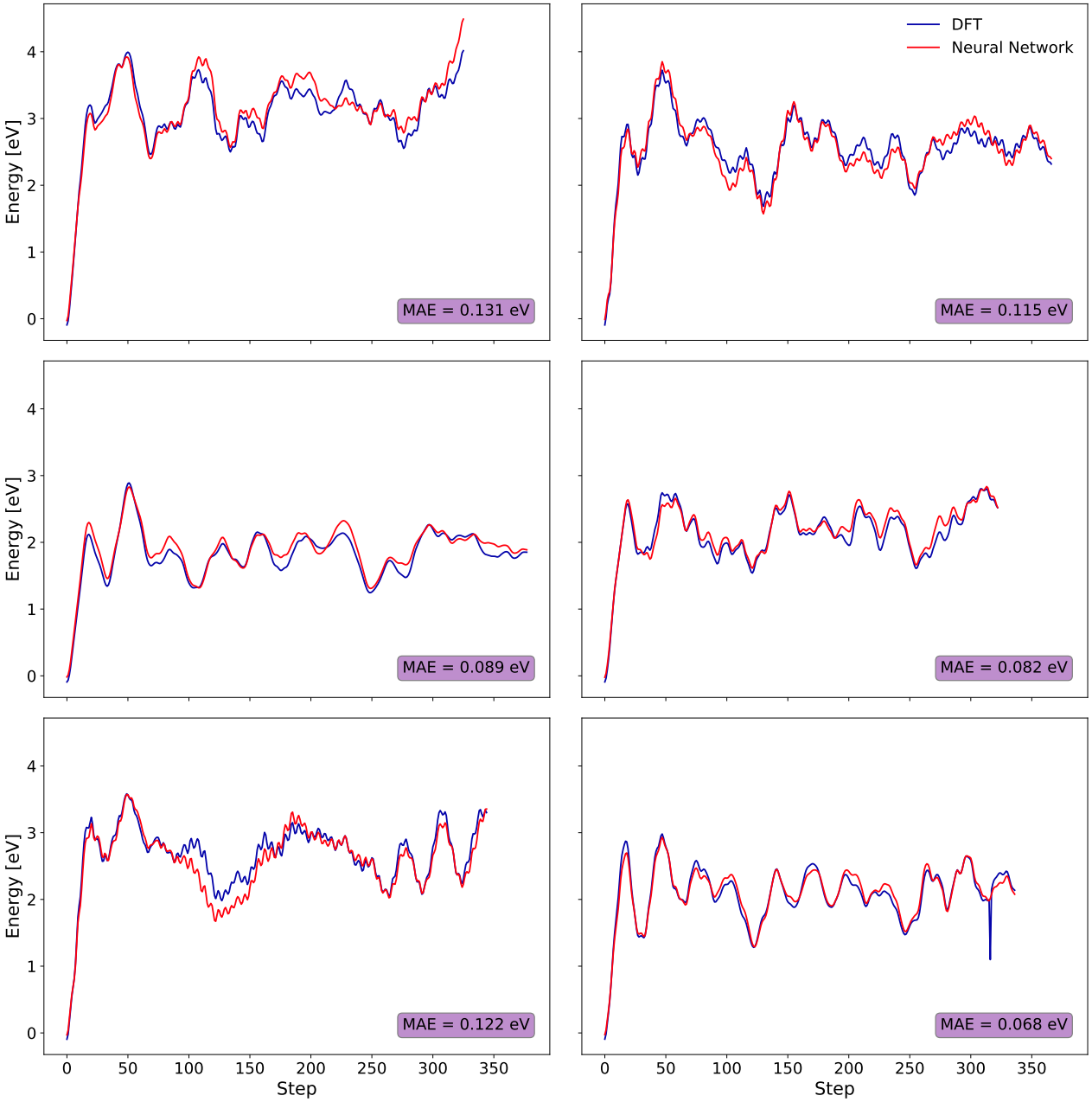


Figure 4.4: The DFT calculated energies and the neural network predicted energies, relative to the free methane molecule and the surface at 0 K. The total energy MAE and energy RMSE were determined to be 0.101 eV and 0.131 eV, respectively.

The temperature distribution for the six trajectories can be seen in Figure 4.5. It is clear that even though the neural network was only trained for temperatures up to 800 K, it could still predict the energies of systems with higher temperatures with good precision. This can be explained by the feature vectors the neural network used. Since they are based on distances between the atoms, translation and rotational energies do not affect the energy predictions. There could, however, be problems when the vibrational energies are too large, and the distances between the hydrogen atoms and the carbon atom are too large. If only one carbon-hydrogen distance is increased, it would not be a problem, since the neural network is trained to predict energies of systems containing methyl and hydrogen. Problems could arise for vibrational modes where more than one carbon-hydrogen bond length is increased, since the training data did not contain systems at these temperatures, nor systems with a methylene molecule and two free hydrogen atoms. The energy RMSE for the trajectories can be compared with the energy RMSE for the training data set, which was determined to be 0.103 eV. The relatively small difference, partly influenced by the increased temperatures in the six trajectories, shows that there was no overfitting in the training of the neural network.

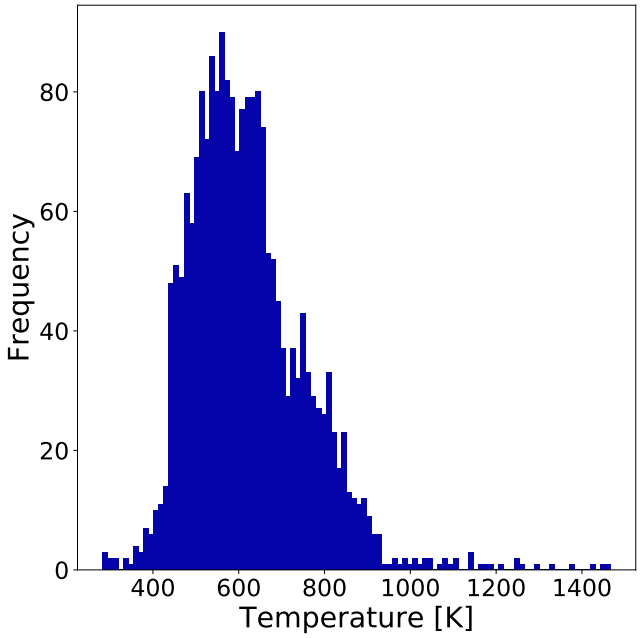


Figure 4.5: The temperature distribution for the 6 trajectories shown in Figure 4.4.

4.3 Potential Energy Surfaces

The potential energy surface, on which all molecular dynamics simulations were dependent, was predicted with the neural network. In Figure 4.6 the 1-dimensional potential energy surface is shown for the adsorbed methane molecule on the surface. The adsorption energy is expressed as a function of the distance between the carbon

atom and the palladium atom, on which the molecule is adsorbed. As a sanity check, a few DFT calculations were performed to compare with the neural network predicted potential energy surface. The neural network predictions and the DFT calculations agree very well, both in the repelling close range region, the adsorption region, and in the long-range region. The energy MAE and energy RMSE are in this region 0.0194 eV and 0.0230 eV, respectively.

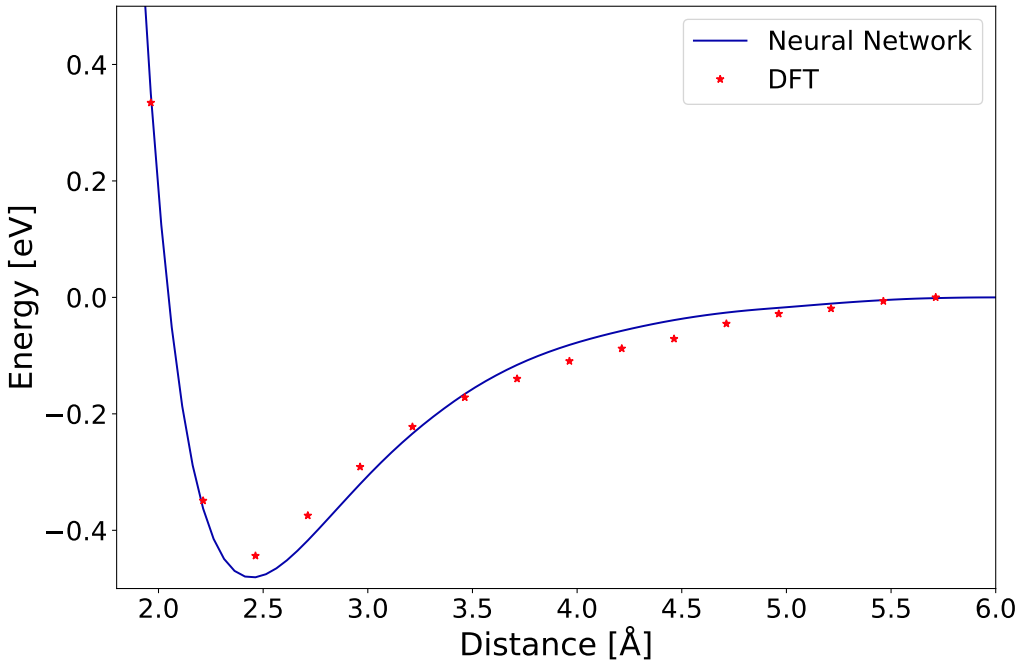


Figure 4.6: The one-dimensional potential energy surface for the methane molecule at its adsorption position, and at translations in the z-direction.

The 2-dimensional potential energy surfaces for a methane molecule at different heights above the surface at 0 K are shown in Figure 4.7. Of course, there exist an infinite amount of such representations, dependent of the orientation on the methane molecule. Therefore, two extreme cases were chosen. The left row in Figure 4.7 shows the potential energy surfaces for methane molecules with one hydrogen atom directed straight towards the surface and one straight down in the y-direction, and the right row shows the potential energy surfaces for the methane molecule with one hydrogen atom directed straight away from the surface and one straight up in the y-direction. The distance between the carbon atom and the surface, with the centered palladium atom as reference, were chosen to be 2.46 Å (equal to the adsorption position distance to the centered palladium atom), 3.0 Å, 4.0 Å and 5.0 Å, shown in the first, second, third and fourth row, respectively. The potential energy surfaces are above the primitive cell shown in Figure 4.8. It can be noted, that for the equilibrium structure of methane, i.e., all angles 109.5°, the largest adsorption energy in any of the two extreme cases is −0.345 eV, compared to the global minimum with

an adsorption energy of $-0.434\,\text{eV}$, with more favourable rotation and angles in the molecule. Since the methane molecule is very inert, and a large part of the potential energy is due to van der Waals interactions, the distance between the surface and the atoms in the molecule has a higher energy contribution than the orientation and activation of the methane molecule. However, to reach the global minimum in the potential energy surface, both orientation, activation and distance must be optimal.

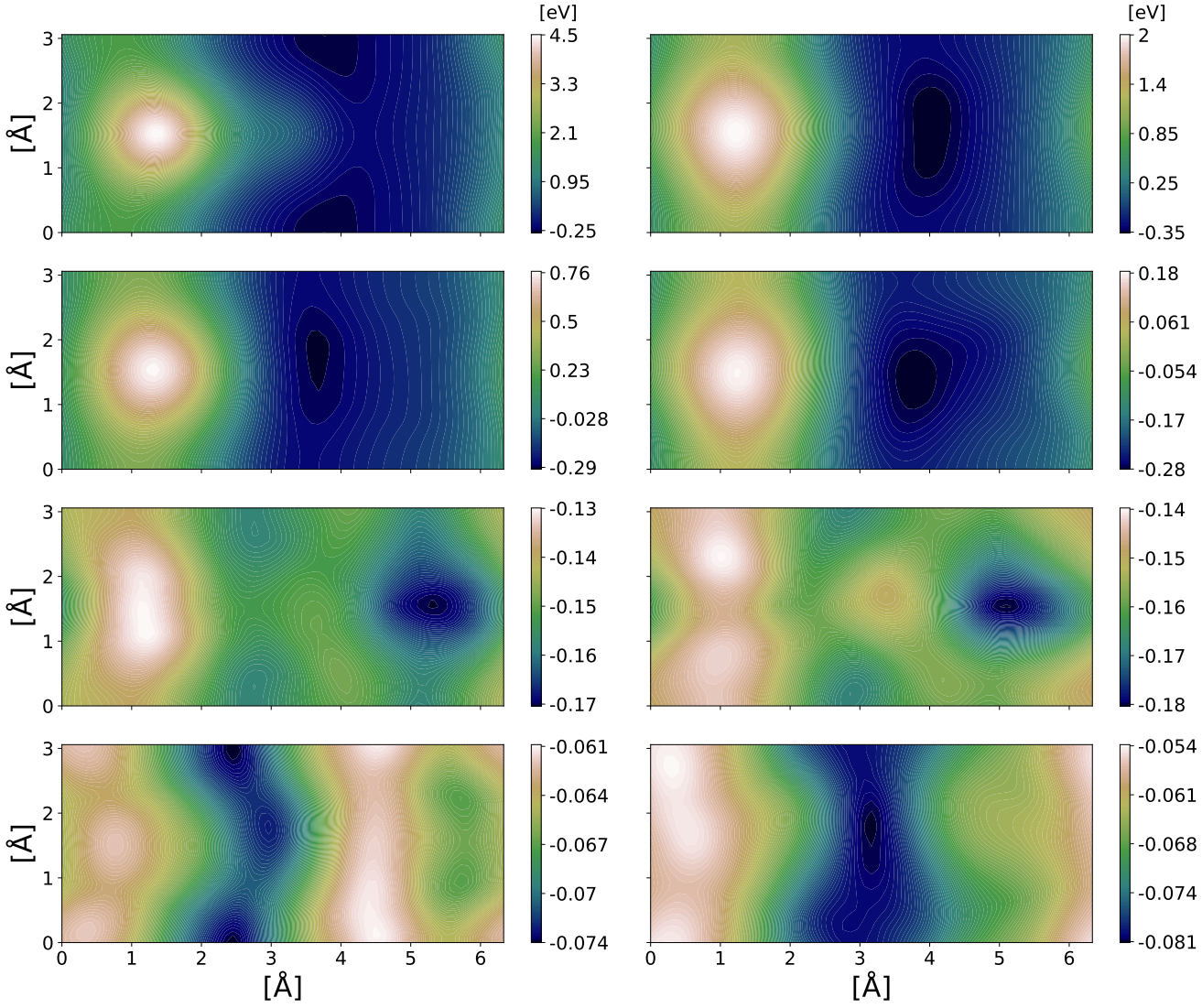


Figure 4.7: The potential energy surface for the methane molecule above the primitive cell of the surface, shown in Figure 4.8, when one hydrogen atom is directed straight toward the surface (left figures), and one hydrogen directed straight away from the surface. The distances above the surface are 2.46 Å (first row), 3.0 Å (second row), 4.0 Å (third row), 5.0 Å (fourth row).

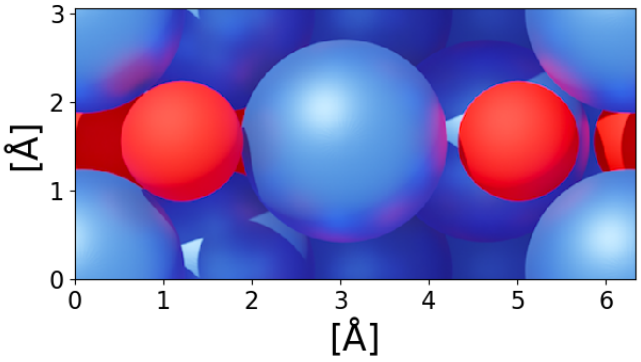


Figure 4.8: The primitive cell of the surface.

4.4 Sticking

To determine the probability of adsorption on the surface, methane molecules were placed 3.827 Å above the surface, with random orientations and random x and y-coordinates. The molecules were given an initial velocity of 0.0219 Å/fs corresponding to 0.4 eV, directed in the negative z-direction. The surface was thermally disordered with a Maxwell-Boltzmann distribution corresponding to 100 K. The Velocity Verlet algorithm was then applied with a time step of 2 fs and 200 steps. The number of incoming methane molecules was 12658. The methane molecule was considered adsorbed if it was within 3 Å of the surface, in the z-direction, relative to the palladium atoms. This is motivated by Figure 4.7, since the within 3 Å of the surface, the only positions of the surface that can adsorb a molecule are in close vicinity of the initial position of the dissociation reaction. Due to the low temperature of the surface, the adsorbed methane molecules stayed adsorbed on the surface. The number of methane molecules that were adsorbed on the surface was 570. Therefore, the probability of adsorption is 0.045. The initial positions of the adsorbed (blue) and repelled (red) molecules can be seen in Figure 4.9. When the methane molecule is initialized straight above the active site, the chance of an adsorption event is large. However, the chance of adsorption is also very large if the molecule is initialized above the rightmost region of the primitive cell, even though this region is relatively repelling (see Figure 4.7). This region, however, seems to be relatively soft, so that a large fraction of the translation energy of the methane molecule is transferred to the phonons in the surface. The repelled molecule then has a relatively small translation energy directed from the surface, and can therefore be attracted back towards the active site of the surface.

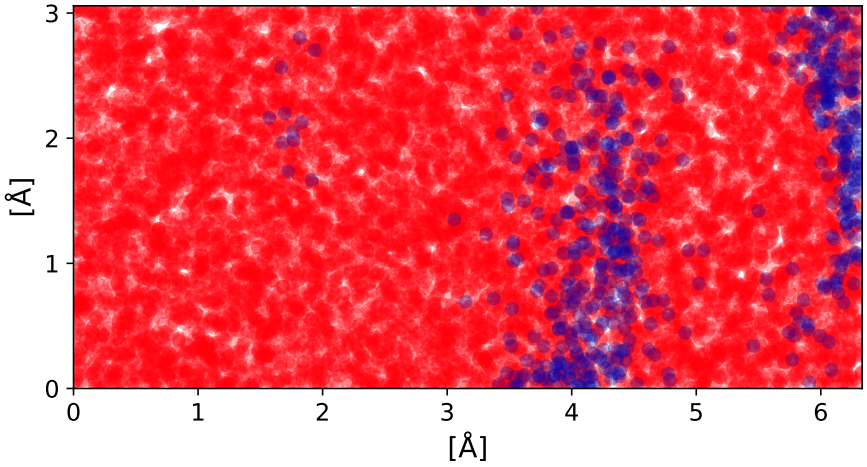


Figure 4.9: The initial positions of the methane molecules in the trajectories (with random orientations). The initial positions that resulted in an adsorption and in a rejection are plotted as blue dots and red dots, respectively.

To study the adsorption in more detail, the initialization of the methane molecules was constrained in the x, y-plane. The initialization was done analogous to previous tests, but the only allowed positions in the x, y-plane were within the rectangle [3.22, 3.82; 0.679, 2.279] Å in the primitive cell. The Velocity Verlet algorithm was applied for 1 ps, with a time step of 2 fs. The total number of incoming methane molecules was in this case 1062, and the number of adsorbed molecules was 76. The probability of adsorption is, in this case, 0.072. If one compares previous probability, this emphasises the importance of favourable collision on the surface in order for a methane molecule to be adsorbed. In some cases, if a methane molecule collided with an unfavourable position on the surface (see Figure 4.7), the repelled molecule could be attracted by the favourable positions and drawn back to the surface, if a large portion of the translation energy was transferred to vibrational and rotational energy or vibrations in the surface.

To study the dissociation process, the methane molecule was initialized at the global minimum of the potential energy surface, i.e., the initial position of the dissociation process. The surface was left unchanged, thermally disordered with a Maxwell-Boltzmann distribution corresponding to 100 K. The atoms in the methane molecule were given completely random momenta, via a Maxwell-Boltzmann distribution corresponding to 0.85 eV. This value is the sum of the initial kinetic energy of the methane molecule and the average neural network/DFT calculated adsorption energy value. Due to the high energies and large velocities, a time step of 2 fs for the Velocity Verlet algorithm was not satisfactory to capture the correct vibrations of the methane molecules and their interactions with the surface, hence not conserving the total energy of the system. The time step was therefore chosen to be 1 fs applied for 200 iterations. The methane molecule was considered dissociated if the distance between the carbon atom and any hydrogen atom was larger than 2 Å. If one con-

siders Figure 4.1, this would require an energy of approximately 3 eV without a catalyst. For 8415 dissociation trials, only one of the methane molecules dissociated on the surface, yielding a probability of 0.00012. In summation, this implies that the probability of an incoming particle being both adsorbed and dissociated on the surface, under these conditions, is $5.4 \cdot 10^{-6}$.

In order to study the temperature dependence of the dissociation reaction, the atoms in the methane molecule were also given completely random momenta, via a Maxwell-Boltzmann distribution corresponding to 1.0 eV, with an otherwise completely analogous environment. In this case, three out of 2567 trails resulted in a dissociation event, corresponding to a probability of 0.0012. Interestingly, none of the four dissociation events, either in the 0.85 eV-case or the 1.0 eV-case, resulted in a dissociation reaction path described in Figure 2.2 and Figure 4.2, where the hydrogen atom approaches the high-positioned oxygen atom. Instead, the methyl molecule was hybridized as suspected, but the hydrogen atom approached the interstitial, right between the centered palladium atoms, which can be seen in Figure 4.10. Note, that due to the symmetry of the surface, this is equivalent to the state obtained if the system is reflected in the horizontal axis.

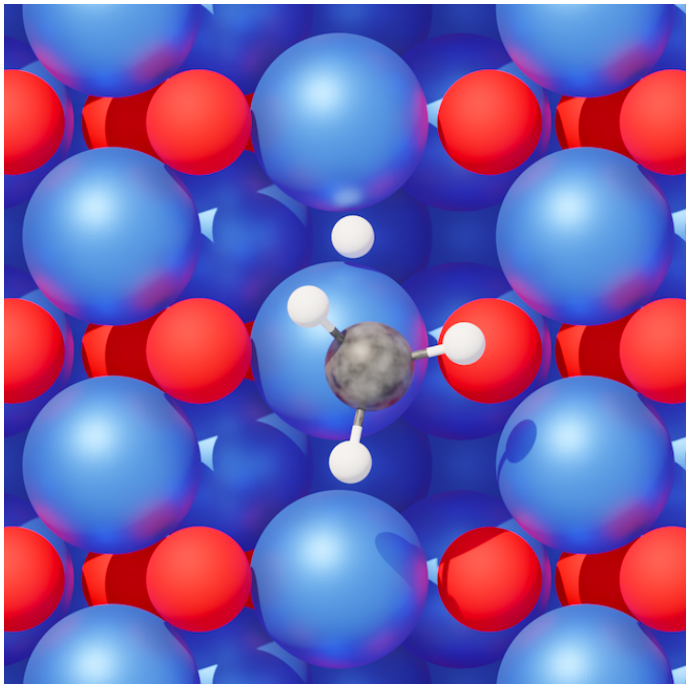


Figure 4.10: The methyl group and the interstitial hydrogen atom after the dissociation process that occurred.

The nudged elastic band calculation performed with the neural network showed that the activation energy to reach the state shown in Figure 4.10 was actually lower than the activation energy yielded in Figure 4.2. In Figure 4.11 the minimum energy transition path between the initial state and the newly found local minimum is shown. The activation energy of this dissociation process is 0.578 eV, corresponding

to an apparent activation energy of 0.109 eV. However, the final state of this process is 0.276 eV higher in energy than the initial state, suggesting that the hydrogen atom might continue towards the global minimum, showed in Figure 2.2, in later stages of the process. The energy difference between this local minimum and the initial state was also calculated with the BFGS algorithm using DFT. The energy difference between the two stages was determined to be 0.317 eV, which is in good agreement with the neural network prediction. This dissociation process, might be followed by the reverse reaction, since the activation energy is relatively small, and the energy difference favours the reverse reaction. However, it might also work as an alternative reaction path, in order to reach the final state, where the hydrogen atom is bonded to the high-positioned oxygen atom; either the previously believed, or the oxygen atom in another row. The existence of alternative reaction paths is a logical explanation to why so few methane molecules are dissociated for molecular dynamics simulations during 200 fs.

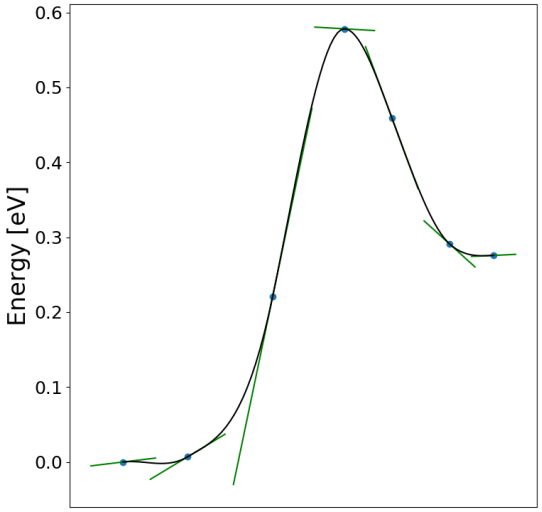


Figure 4.11: The minimum energy transition path and activation energy calculated with nudged elastic bands for a possible, alternative, dissociation path, predicted by the neural network.

4.5 Potential Errors and Corrections

The three most probable causes of errors are due to the DFT calculations, the training data, and the architecture of the neural network. The surface, containing $1\times2\times5$ primitive cells of palladium oxide (101), i.e. $2\times2\times5$ palladium atoms, might not be large enough to describe all interactions and surface vibrations. However, since the computational cost increases rapidly with the number of atoms in the system, and the neural network had to be trained with a large number of systems, this compromise is deemed acceptable. The cutoff energy and number of k-points used were determined solely by the energies and lattice parameters of the bulk structures. To avoid undesirable energy fluctuations, more plane waves could have

been used, in a more heavily sampled reciprocal space. However, this would also be computationally more costly; and since the results of the adsorption energy, H–C–H angle, and activation energy are consistent with previous results, the cutoff energy and number of k-points were most likely enough. Compared to the neural network, the errors caused by the DFT calculations are considered negligible.

Due to memory storage problems, the size of the training data set was limited to roughly 20 000 systems. If more training data were used, more hidden neurons could have been used with a lower chance of overfitting. However, the network used in this project, with 2×20 hidden neurons, required approximately 10 days to train 1200 epochs of the training data set, although parallelized over 20 cores. More hidden neurons, with a larger training data set might require other sacrifices, such as fewer feature vectors and a smaller number of epochs during the training. The cutoff radius, used to determine the local environment around each atom was chosen to be 6.5 Å, to capture surface vibrations and adsorption energies at larger distances. However, since this cutoff radius is slight larger than the cell size, carbon-carbon interactions are, to a small degree, captured in the feature vectors. This could have been avoided with the use of a larger cell size.

It has been explained how the vibrational energies of methane molecules of higher temperatures could be predicted, despite not being part of the training set. However, to be able to use even larger energies of the incoming molecules, and obtain more accurate predictions during the collisions, the training data set could either include systems of higher energies, or, apart from methyl-hydrogen configurations, also include methylene-hydrogen-hydrogen configurations. This would capture higher vibrational energy states of the methane molecule, as well as higher energy states of the methyl-hydrogen configuration, after the dissociation process.

5. Conclusion

5.1 Reflections

The global minimum in the potential energy surface, obtained from the DFT calculations, confirmed that the active sites of the palladium oxide (101) surface are the 3-fold coordinated palladium atoms. As suspected, the $(\sqrt{5} \times \sqrt{5})$R27° structure of palladium (100) with one palladium oxide (101) layer on top did not yield a sufficient adsorption energy or methane activation, due to the absence of 3-fold coordinated palladium atoms. However, if two layers are used, the activation and adsorption energy are comparable to those of bulk palladium oxide (101). The adsorption energy and the activation of the methane molecule are dependent on the exchange-correlation functional, with the functionals yielding the most reliable results being PBE and CX.

The activation energy for the dissociation of methane on the surface, using the exchange-correlation functional CX, was determined to be 0.657 eV. The apparent activation energy was determined to be 0.188 eV, which is slightly lower than previous computational and experimental results. The final state of the dissociation, where methyl is bonded to the 3-fold coordinated palladium atom, and the hydrogen atom is bonded to the oxygen atom, was 0.401 eV lower in energy than the initial state of the dissociation.

The trained neural network was able to predict the energies and forces approximately 100 000 and 3000 times faster than DFT, respectively. The mean absolute energy error was determined to be 0.436 meV per atom at 0 K, and 2.24 meV per atom for a Maxwell-Boltzmann distributed temperature distribution centered at 600 K. Note that this includes temperatures outside the trained interval. Due to the use of feature vectors, rather than atomic positions, the neural network could, potentially, be applied on systems of other sizes and structures.

Thanks to the amplified computational speed when using neural networks, it was possible to predict the trajectories of a large number of incoming methane molecules approaching the surface. From 12658 incoming methane molecules, with completely random positions and orientations, and a translation energy of 0.4 eV, directed towards the surface with a temperature of 100 K, it was possible to determine the rate of adsorption. The ratio of adsorption was determined to be approximately 4.5%, which is in good agreement with previous experimental studies on other noble metal catalysts. It can also be concluded that the possibility of adsorption was heavily dependent on the initial positions of the methane molecules; the probability was the greatest when the initialization occurred right above the active site, or at the

rightmost part of the primitive cell.

The probability of dissociation of methane on the surface, was considerably lower than the probability of adsorption, for this choice of energy and surface temperature. During 200 fs, only one dissociation event occurred, out of 8415 trials. However, when the methane molecule was given an additional 0.15 eV, the probability of dissociation was increased by a factor of ten. The dissociation did, however, not occur the same way as the previously believed most probable transition path. Instead, the hydrogen atom took an interstitial position between the palladium atoms. The activation energy of this transition path was approximately 0.08 eV lower than if the hydrogen atom moved directly to the oxygen atom. However, the energy of the interstitial position was higher than both the initial state and oxygen-hydrogen bonded state, which indicates that the hydrogen atom might continue towards either of the high-positioned oxygen atoms. If the energy required to do so is relatively small, this newly discovered dissociation event might be the dominating one.

5.2 Outlook

There are two very clear directions for further research. The newly discovered dissociation event can be studied in more detail. The neural network predictions have been able to suggest the transition path, but nudged elastic bands, calculated with DFT would have to be used to study both the transition path between the initial state and the interstitial state, and the transition path between the interstitial state and possible final states in more detail.

The adsorption and dissociation, or more precisely the sticking of methane on the palladium oxide (101) surface, could be determined at different surface temperatures and energies of the incoming methane molecule. It is believed that the adsorption would decrease and the dissociation would increase, as the temperature increases. However, at higher temperatures, the dissociation would need to happen instantaneously after adsorption, since the methane coverage on the surface at these temperatures is extremely low. Hopefully, the temperature yielding the maximum sticking can be found from the perfect compromise between adsorption and dissociation.

References

[1] D. Ciuparu, M. R. Lyubovsky, E. Altman, L. D. Pfefferle, and A. Datye, "Catalytic combustion of methane over palladium-based catalysts," *Catalysis Reviews*, vol. 44, no. 4, pp. 593–649, 2002. DOI: 10.1081/CR-120015482.

[2] D. T. Shindell, G. Faluvegi, D. M. Koch, G. A. Schmidt, N. Unger, and S. E. Bauer, "Improved attribution of climate forcing to emissions," *Science*, vol. 326, no. 5953, pp. 716–718, 2009.

[3] J. H. Lee and D. L. Trimm, "Catalytic combustion of methane," *Fuel processing technology*, vol. 42, no. 2-3, pp. 339–359, 1995.

[4] A. Hellman, A. Resta, N. M. Martin, *et al.*, "The active phase of palladium during methane oxidation," *The Journal of Physical Chemistry Letters*, vol. 3, no. 6, pp. 678–682, 2012. DOI: 10.1021/jz300069s.

[5] M. V. den Bossche and H. Grönbeck, "Methane oxidation over PdO(101) revealed by first-principles kinetic modeling," *Journal of the American Chemical Society*, vol. 137, 2015. DOI: 10.1021/jacs.5b06069.

[6] H. L. Abbott and I. Harrison, "Methane dissociative chemisorption on ru (0001) and comparison to metal nanocatalysts," *Journal of Catalysis*, vol. 254, no. 1, pp. 27–38, 2008.

[7] J. A. Manion, "Evaluated enthalpies of formation of the stable closed shell c1 and c2 chlorinated hydrocarbons," 2002. DOI: 10.1063/1.1420703.

[8] Malcolm W and Chase, Jr., *NIST-JANAF Thermochemical Tables*, 4th ed. Journal of Physical and Chemical Reference Data, 1998, ISBN: 1563968312.

[9] I. Chorkendorff and J. W. Niemantsverdriet, *Concepts of modern catalysis and kinetics*. John Wiley & Sons, 2017, ISBN: 9783527305742.

[10] R. F. Hicks, H. Qi, M. L. Young, and R. G. Lee, "Structure sensitivity of methane oxidation over platinum and palladium," *Journal of Catalysis*, vol. 122, no. 2, pp. 280–294, 1990, ISSN: 0021-9517. DOI: https://doi.org/10.1016/0021-9517(90)90282-O.

[11] G. Pecchi, P. Reyes, R. Gómez, T. López, and J. Fierro, "Methane combustion on rh/zro2 catalysts," *Applied Catalysis B: Environmental*, vol. 17, no. 1, pp. L7–L13, 1998, ISSN: 0926-3373. DOI: https://doi.org/10.1016/S0926-3373(98)00012-5.

[12] H. Arai, T. Yamada, K. Eguchi, and T. Seiyama, "Catalytic combustion of methane over various perovskite-type oxides," *Applied Catalysis*, vol. 26, pp. 265–276, 1986, ISSN: 0166-9834. DOI: https://doi.org/10.1016/S0166-9834(00)82556-7.

[13] F. Ribeiro, M. Chow, and R. Dallabetta, "Kinetics of the complete oxidation of methane over supported palladium catalysts," *Journal of Catalysis*, vol. 146, no. 2, pp. 537–544, 1994. DOI: https://doi.org/10.1006/jcat.1994.1092.

[14] A. Trinchero, A. Hellman, and H. Grönbeck, "Metal-oxide sites for facile methane dissociation," *Physica Status Solidi*, vol. 8, 2014. DOI: 10.1002/pssr.201409090.

[15] C. Nordling and J. Österman, *Physics Handbook for Science and Engineering*. Studentlitteratur, 2006, ISBN: 9789144044538.

[16] D. D. Wagman, W. H. Evans, V. B. Parker, *et al.*, "The nbs tables of chemical thermodynamic properties," *Journal of Physical and Chemical Reference Data*, vol. 11, 1982.

[17] D. Rogers, R. Shannon, and J. Gillson, "Crystal growth and semiconductivity of palladium oxide," *Journal of Solid State Chemistry*, vol. 3, no. 2, pp. 314–316, 1971, ISSN: 0022-4596. DOI: https://doi.org/10.1016/0022-4596(71)90045-4.

[18] J. Waser, H. A. Levy, and S. Peterson, "The structure of pdo," *Acta Crystallographica*, vol. 6, no. 7, pp. 661–663, 1953.

[19] J. Nell and H. O'Neill, "Gibbs free energy of formation and heat capacity of pdo: A new calibration of the pd pdo buffer to high temperatures and pressures," *Geochimica et Cosmochimica Acta*, vol. 60, no. 14, pp. 2487–2493, 1996. DOI: https://doi.org/10.1016/0016-7037(96)00115-9.

[20] N. M. Martin, M. Van den Bossche, A. Hellman, *et al.*, "Intrinsic ligand effect governing the catalytic activity of pd oxide thin films," *ACS Catalysis*, vol. 4, no. 10, pp. 3330–3334, 2014. DOI: 10.1021/cs5010163.

[21] M. Todorova, E. Lundgren, V. Blum, *et al.*, "The pd(100)–(5×5)r27°-o surface oxide revisited," *Surface Science*, vol. 541, no. 1, pp. 101–112, 2003, ISSN: 0039-6028. DOI: https://doi.org/10.1016/S0039-6028(03)00873-2.

[22] G. Zhu, J. Han, D. Y. Zemlyanov, and F. H. Ribeiro, "Temperature dependence of the kinetics for the complete oxidation of methane on palladium and palladium oxide," *The Journal of Physical Chemistry B*, vol. 109, no. 6, pp. 2331–2337, 2005. DOI: 10.1021/jp0488665.

[23] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics, Second Edition*. Cambridge University Press, 2017, ch. 2, ISBN: 9781108422413.

[24] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics, Third Edition*. Cambridge University Press, 2018, ch. 5, ISBN: 9781107189638.

[25] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," *Annalen der physik*, vol. 389, no. 20, pp. 457–484, 1927.

[26] E. Hasselbrink and B. Lundqvist, *Handbook of Surface Science*. Elsevier B.V., 2008, ch. 10, ISBN: 9780444520562.

[27] J. Thijssen, *Computational Physics*, 2nd ed. Cambridge University Press, 2007. DOI: 10.1017/CBO9781139171397.

[28] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, B864–B871, 3B 1964-11. DOI: 10.1103/PhysRev.136.B864.

[29] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, A1133–A1138, 4A 1965-11. DOI: 10.1103/PhysRev.140.A1133.

[30] R. O. Jones and O. Gunnarsson, "The density functional formalism, its applications and prospects," *Rev. Mod. Phys.*, vol. 61, pp. 689–746, 3 1989-07. DOI: 10.1103/RevModPhys.61.689.

[31] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, A1133–A1138, 4A 1965-11. DOI: 10.1103/PhysRev.140.A1133.

[32] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," *Phys. Rev. B*, vol. 23, pp. 5048–5079, 10 1981-05. DOI: 10.1103/PhysRevB.23.5048.

[33] J. P. Perdew, J. A. Chevary, S. H. Vosko, *et al.*, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Phys. Rev. B*, vol. 46, pp. 6671–6687, 11 1992-09. DOI: 10.1103/PhysRevB.46.6671.

[34] B. Hammer, K. W. Jacobsen, and J. K. Nørskov, "Role of nonlocal exchange correlation in activated adsorption," *Phys. Rev. Lett.*, vol. 70, pp. 3971–3974, 25 1993-06. DOI: 10.1103/PhysRevLett.70.3971.

[35] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical review letters*, vol. 77, no. 18, p. 3865, 1996.

[36] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," *Phys. Rev. Lett.*, vol. 91, p. 146 401, 14 2003-09. DOI: 10.1103/PhysRevLett.91.146401.

[37] K. Berland and P. Hyldgaard, "Exchange functional that tests the robustness of the plasmon description of the van der waals density functional," *Phys. Rev. B*, vol. 89, p. 035 412, 3 2014-01. DOI: 10.1103/PhysRevB.89.035412.

[38] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, *et al.*, "Density functionals for surface science: Exchange-correlation model development with bayesian error estimation," *Phys. Rev. B*, vol. 85, p. 235 149, 23 2012-06. DOI: 10.1103/PhysRevB.85.235149.

[39] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, no. 5, pp. 115–133, 1943. DOI: 10.1007/BF02478259.

[40] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993, ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80131-5.

[41] S. Sharma and S. Sharma, "Activation functions in neural networks," *Towards Data Science*, vol. 6, no. 12, pp. 310–316, 2017.

[42] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and intelligent laboratory systems*, vol. 39, no. 1, pp. 43–62, 1997.

[43] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[44] L. Verlet, "Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Physical review*, vol. 159, no. 1, p. 98, 1967.

[45] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of chemical physics*, vol. 76, no. 1, pp. 637–649, 1982.

[46] C. Kittel, *Introduction to Solid State Physics*, 8th ed. Wiley, 2005, ISBN: 9780471415268.

[47] J. Hafner, "Ab-initio simulations of materials using vasp: Density-functional theory and beyond," *Journal of Computational Chemistry*, vol. 29, no. 13, pp. 2044–2078, 2008. DOI: https://doi.org/10.1002/jcc.21057.

[48] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, pp. 17 953–17 979, 24 1994-12. DOI: 10.1103/PhysRevB.50.17953.

[49] H. J. Monkhorst and J. D. Pack, "Special points for brillouin-zone integrations," *Phys. Rev. B*, vol. 13, pp. 5188–5192, 12 1976-06. DOI: 10.1103/PhysRevB.13.5188.

[50] "A broyden—fletcher—goldfarb—shanno optimization procedure for molecular geometries," *Chemical Physics Letters*, vol. 122, no. 3, pp. 264–270, 1985, ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(85)80574-1.

[51] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.

[52] D. Sholl and J. A. Steckel, *Density functional theory: a practical introduction*. John Wiley & Sons, 2011.

[53] H. Jónsson, G. Mills, and K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions," 1998.

[54] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A climbing image nudged elastic band method for finding saddle points and minimum energy paths," *The Journal of chemical physics*, vol. 113, no. 22, pp. 9901–9904, 2000.

[55] A. Khorshidi and A. A. Peterson, "Amp: A modular approach to machine learning in atomistic simulations," *Computer Physics Communications*, vol. 207, pp. 310–324, 2016, ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2016.05.010.

[56] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146 401, 14 2007-04. DOI: 10.1103/PhysRevLett.98.146401.

[57] J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *The Journal of chemical physics*, vol. 134, no. 7, p. 074 106, 2011.

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY