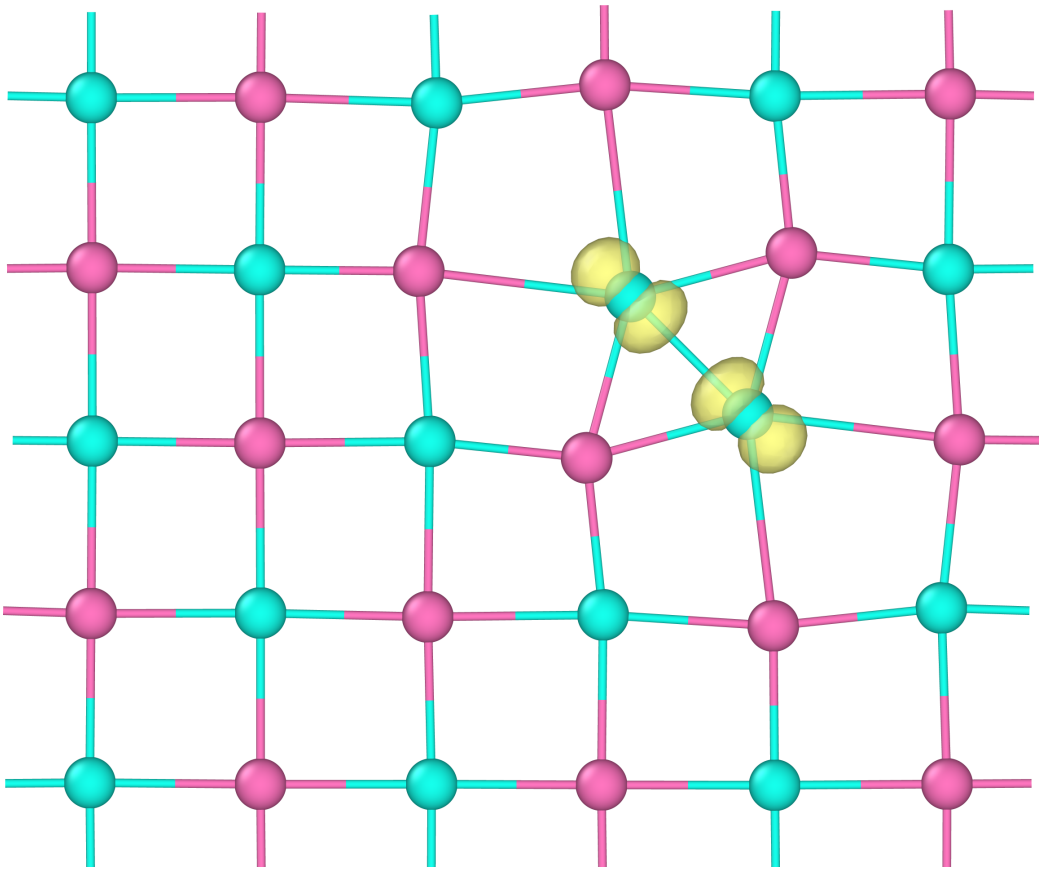


Machine-learning potentials for finite-temperature simulations of polarons

Applying neuroevolution potentials to the study of polarons in MgO and LiF at finite temperatures

Master's thesis in Physics

Sofia Cellini

DEPARTMENT OF PHYSICS AND ASTRONOMY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Machine-learning potentials for finite-temperature simulations of polarons

Applying neuroevolution potentials to the study of polarons in MgO and LiF at finite temperatures

Sofia Cellini

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics and Astronomy
Division of Condensed Matter and Materials Theory
Wiktor Group
Chalmers University of Technology
Gothenburg, Sweden 2026

Machine-learning potentials for finite-temperature simulations of polarons
Applying neuroevolution potentials to the study of polarons in MgO and LiF at finite temperatures
Sofia Cellini

Supervisor: Ethan Berger, Department of Physics and Astronomy
Examiner: Julia Wiktor, Department of Physics and Astronomy

Master's Thesis 2026
Department of Physics and Astronomy
Division of Condensed Matter and Materials Theory
Wiktor Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Isosurface of the V_k polaron state in LiF.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Machine-learning potentials for finite-temperature simulations of polarons
Applying neuroevolution potentials to the study of polarons in MgO and LiF at finite temperatures
Sofia Cellini
Department of Physics and Astronomy
Chalmers University of Technology

Abstract

The performance of functional materials in advanced energy technologies, such as photovoltaics and photocatalysis, can be strongly influenced by charge localization. Excess charges can cause distortions in the surrounding lattice and localize in a self-generated potential well, forming quasiparticles called polarons. Although first-principles methods based on density functional theory (DFT) offer accurate results for polaron calculations, these methods are limited at finite temperatures due to their high computational cost. In the present work, polarons are modelled in MgO and LiF using machine-learned interatomic potentials (MLIPs) to enable finite-temperature calculations at high speed. At 0 K, LiF exhibits two types of stable polarons. The transition between the two polaron states is first studied with hybrid density functional calculations using the string method. One of the polaron states is metastable and the energy barrier is found to be small, 27 meV, making it difficult to model the metastable polaron state with MLIPs. The more stable polaron state is then studied at finite temperatures, together with the polaron in MgO. Using hybrid density functional calculations as reference data, neuroevolution potential models are trained for each material. The models are used to investigate the temperature dependence of the charge transition level (CTL), a key energetic property of polarons. Temperatures up to 500 K are studied, and the models predict variations in the CTL of approximately 55 meV for MgO, and 380 meV for LiF.

Keywords: polarons, density functional theory, hybrid functionals, neuroevolution potentials, molecular dynamics, thermodynamic integration, charge transition level

Acknowledgements

I would like to thank my supervisor, Ethan Berger, for his invaluable help. Thank you for all the discussions and for answering endless questions throughout the project. I would also like to thank my examiner, Julia Wiktor, for the opportunity to do this thesis and for crucial input in the project. Lastly, I would like to thank the division of Condensed Matter and Materials Theory. Both working on the thesis and getting to know the division have been great fun.

Sofia Cellini, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

CB conduction band. 6

CBM conduction band minimum. xiv, 6, 7, 9, 36, 37

CTL charge transition level. xiii, xvi, 2, 3, 7, 8, 20–22, 28, 31, 33, 36, 45–47, 49, 50

DFT density functional theory. xiii–xv, xvii, 1, 2, 8–10, 12–15, 21–29, 31, 33–37, 39–42, 44, 45, 49, 50

DOS density of states. 29

GGA generalized gradient approximation. 12

LDA local density approximation. 12

MD molecular dynamics. xiii, 2, 16, 18, 20, 22, 24, 25, 28, 29, 37, 38, 40–42, 44, 49, 50

MEP minimum energy path. xiii, xiv, 14, 15, 21–23, 33, 34

MLIP machine learning interatomic potential. 1, 2, 15, 41

NEP neuroevolution potential. xiii–xv, 15, 16, 18, 20–29, 31, 34–45, 49, 50

PES potential energy surface. xiv, 9, 36, 37, 50

RDF radial distribution function. xii, xiii, xv, 18, 22, 28, 37–39, 49, 50

RMSE root mean square error. xiv, xvi, 24, 28, 34, 35, 49, I–III

SIE self-interaction error. 12–14

VB valence band. 6

VBM valence band maximum. xv, xvi, 6–9, 22, 28, 29, 31, 44–47, 49

Contents

List of Figures

List of Tables

1
Introduction

As the demand for renewable energy grows, it is becoming increasingly important to harness advanced energy technologies. Photovoltaics and photocatalysis are examples of such technologies, in which solar energy is used to generate electricity or to drive chemical reactions such as water-splitting for hydrogen fuel production. A thorough understanding of the functional materials used in these technologies is crucial for efficient implementation. In particular, it is of great importance to understand how the performance of these materials is affected by the behaviour of excess charges. A surplus of electrons or holes may be present in a material due to processes such as doping or photon excitation. In certain materials, excess charges can distort the surrounding lattice and create a self-generated potential well in which they become localized. Such localized charges, together with the induced lattice distortions, constitute quasiparticles known as polarons. Depending on whether the localized excess charge is a hole or an electron, the quasiparticle is called a hole polaron or an electron polaron. Additionally, a distinction is made between small and large polarons. In small polarons, the spatial extension of the localization is comparable to the lattice parameter. On the other hand, polarons that extend over many lattice sites are called large polarons. The presence of polaronic states strongly influences the electronic properties of a material, highlighting the importance of being able to model polarons and understand their effects in realistic conditions.

Prior studies of polarons have predominantly been performed at zero temperature, despite the fact that the behaviour of polarons is generally temperature-dependent. These studies are often based on first-principles calculations such as density functional theory (DFT) which require large computational resources. While zero-temperature simulations only require a single calculation, it is necessary to perform a great number of time-evolving calculations to capture finite-temperature effects. Consequently, methods such as DFT are limited at higher temperatures. To overcome this limitation, machine learning methods have become increasingly common. By training a machine learning interatomic potential (MLIP), atomistic simulations can be performed with similar accuracy but many times faster [1]. However, MLIPs are not generally able to perform atomistic simulations involving excess charges, since the potential is modelled as a function of the atomic positions without taking electrons into consideration. Nevertheless, by labelling the atomic site of a localized excess charge, MLIPs can learn to simulate small polarons [2]. This is referred to as the labelling method in this project, and is applied to the the training of MLIPs in order to analyse the finite-temperature effects of polarons.

New methods must be benchmarked before they can be applied to the relevant, complex materials that are of interest for the aforementioned energy applications.

This project aims to study polarons in the relatively simple materials magnesium oxide (MgO) and lithium fluoride (LiF). Both materials exhibit stable, small polarons, and consist of only two atomic species in a simple rock-salt lattice structure. Moreover, there are many prior studies of polarons in these materials allowing for comparison and validation, making them suitable materials for benchmarking new methods. In this project, MLIPs are trained for each material using DFT calculations as reference data. The MLIPs learn to emulate the physics of both the pristine and polaron systems, enabling high-speed finite-temperature calculations.

1.1 Thesis objective and limitations

The objective of the present work is to model small, hole polarons in MgO and LiF at finite temperatures by training MLIPs using the labelling method. LiF exhibits two types of small polarons. To better understand their stability and the possibility of modelling both polarons with a MLIP, the minimum energy path between the two polaron states is first studied at zero temperature with DFT calculations. Due to the low energy barrier between the two states, only the more stable polaron state is modelled with MLIPs, as well as the polaron in MgO. The MLIPs for each material are used to study the temperature dependence of the polaron distortions and of the charge transition level (CTL). The CTL is a key property of polarons which indicates the energetic level of the polaron state in the band gap.

A limitation of the project is that only small polarons can be modelled with the labelling method. It is difficult to label large polarons that spread over many lattice sites in a way that would allow MLIPs to learn their properties. For a similar reason, the polarons are assumed to always remain localized at the same atomic site and cannot transition between different polaron types. In the simulations, the labelled site cannot change or move through the crystal structure even though polarons generally can hop between different sites or transition to different polaron types in a real crystal. However, since the study is already limited to small polarons, the assumption is less restrictive when it comes to the hopping of polarons. Small polarons are more strongly localized and tend to have lower mobility than large polarons [3].

Another limitation, which is present in almost any atomic simulation, is that realistically sized crystals cannot be directly modelled. The computational cost of the simulations would become unreasonably high. Instead, supercells are simulated with periodic boundary conditions to approximate an infinite crystal. If the supercell is not large enough, finite-size effects may influence the results. When performing molecular dynamics (MD) with a trained potential, large supercells containing millions of atoms can be used. However, DFT calculations may be affected by finite-size effects since they can only be performed with a supercell containing up to a few hundred atoms. In polaron modelling, finite-size effects may arise from electrostatic interactions between periodically repeated images of the polarons. Since electrostatic interactions are long-ranged, the energy converges slowly with system size. Electrostatic corrections can be added to counteract the effects in smaller supercells. However, electrostatic corrections are neglected in this project. The supercells used in DFT calculations are large enough for the electrostatic corrections

to be relatively small. Additionally, the focus of the project lies on the temperature dependence of polaron properties, such as the CTL. The electrostatic corrections should not strongly depend on temperature and should only cause an overall shift.

2

Theory

Condensed matter physics aims to describe how the macroscopic properties of materials emerge from the underlying physics of the atomic nuclei and electrons that the material is composed of. Understanding how atomic-scale phenomena such as polarons affect a material's electronic properties is an example of this. However, predicting macroscopic properties from individual atoms is highly challenging. The immense number of particles, together with all of the interactions between them, constitute a system of such great complexity that although the underlying physical laws are well-known, their direct calculation is unfeasible. In order to model and analyse these systems, appropriate simplifications must be made. This chapter aims to cover the basic concepts and approximations that are necessary to understand many-body quantum systems and polarons, and particularly the relevant numerical methods that are implemented in this project to analyse their properties.

2.1 Electronic band structure

Solids can be modelled as systems composed of nuclei and electrons with the Hamiltonian

$$H = H_e + H_n + H_{en}, \tag{2.1}$$

where H_e describes the electronic subsystem, H_n describes the nuclei and H_{en} describes the interaction between them. However, a model describing the positions of all nuclei and electrons at the same time is incredibly complex. A common simplification is the Born-Oppenheimer approximation, which separates the two subsystems. Due to the significant mass difference between nuclei and electrons, their dynamics occur on different time scales. The electrons move much faster than the nuclei and can be thought of as instantaneously adjusting to the positions of the nuclei, rather than trailing behind them. In other words, the nuclei can be considered fixed in their positions during analysis of the electronic subsystem. When making this assumption, the Hamiltonian of the nuclei H_n can be disregarded. The nuclei have no kinetic energy, and their potential is a constant. Additionally, the many-body wavefunction becomes a function of only the electron coordinates and the interaction term between nuclei and electrons, H_{en}, can be seen as a potential in which the electrons exist, which we will denote $U(\boldsymbol{r})$.

The Hamiltonian for the electronic subsystem can be divided into $H_e = H_{e,\text{kin}} + H_{e,\text{int}}$, where $H_{e,\text{kin}}$ describes the kinetic energy of the electrons and $H_{e,\text{int}}$ describes the electron-electron interactions. The electronic states for a single electron in a periodic lattice potential can be found through the time-independent Schrödinger

equation,

$$\Big[H_{e,\text{kin}} + U(\boldsymbol{r})\Big]\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{2.2}$$

where $\psi(\boldsymbol{r})$ is the single-electron wavefunction and $U(\boldsymbol{r})$ is the periodic lattice potential from the nuclei fulfilling $U(\boldsymbol{r}) = U(\boldsymbol{r} + \boldsymbol{R}_{\text{lattice}})$, with $\boldsymbol{R}_{\text{lattice}}$ representing the lattice vector. Bloch's theorem states that solutions to equation 2.2 can be written in the form

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{\boldsymbol{k}}(\boldsymbol{r}), \tag{2.3}$$

where $\boldsymbol{k}$ is the wave vector, $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ is a Bloch wavefunction and $u_{\boldsymbol{k}}(\boldsymbol{r})$ is a function with the periodicity of the lattice, $u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R}_{\text{lattice}})$. Importantly, equation 2.3 shows that the probability density of an electron in a periodic lattice potential has the periodicity of the lattice. The electronic states are delocalized throughout the crystal.

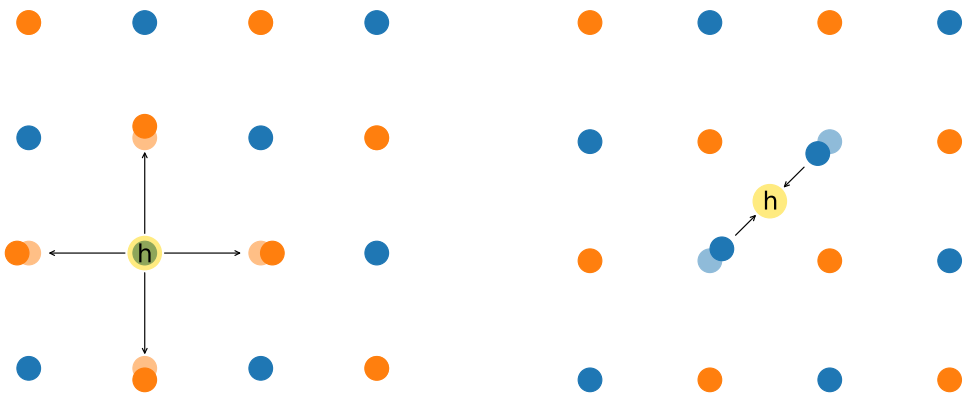
Another important property of periodic systems is that the allowed energetic levels for the electrons form energy bands, each band containing a continuum of energetic states. Together, they constitute the electronic band structure of the system. In crystalline semiconductors and isolators, the highest occupied band is the valence band (VB), and the highest energy state in the VB is the valence band maximum (VBM). On the other hand, the lowest unoccupied band is the conduction band (CB), and its lowest energy state is referred to as the conduction band minimum (CBM). The gap between the valence and conduction bands, where there are no available energy levels, is the band gap. The Fermi energy is the energy of the highest occupied state at zero temperature. In semiconductors and isolators, the Fermi energy lies in the band gap, meaning that the valence band is fully occupied while the conduction band is fully unoccupied [4].

2.2 Polarons

The polaron is a quasiparticle consisting of an excess charge, either an electron or a hole, together with the distortions it causes in the surrounding lattice. The lattice distortions create a self-generated potential well in which the excess charge becomes localized [5]. Polarons may be large or small depending on the spatial extension of the localization. When the charge carrier is strongly coupled to the lattice distortions, it becomes localized over a short length scale, comparable to the lattice parameter, and this corresponds to a small polaron. On the other hand, polarons that extend over many lattice sites are named large polarons [6]. In this project, the focus lies on small, hole polarons.

2.2.1 Polaron distortions

Rock-salt systems, such as MgO and LiF, can have different types of hole polarons, each with different lattice distortions. In one case, the hole is centred on a single atom, causing the neighbouring six atoms to be pushed away. In this project, we refer to this type of polaron as a V_c polaron, which is schematically shown in Figure 2.1a. On the other hand, the hole can be shared between two atoms, causing them to be pulled together and form a dimer. This type of polaron is commonly referred



(a) **(b)**

Figure 2.1: **(a)** Schematic illustration of a V_c polaron. **(b)** Schematic illustration of a V_k polaron. Blue and orange dots represent atoms of different species in a rock-salt lattice, and h represents an excess hole. The black arrows show the direction of the distortions caused by the excess charge.

to as the V_k polaron [7][8]. The lattice distortions are schematically shown in Figure 2.1b.

Both MgO and LiF exhibit stable, small hole polarons. In MgO, there are no previous reports of the V_k polaron. Therefore, only the V_c polaron is modelled in this project, which has been previously studied [6][9]. The excess hole in the V_c polaron in MgO is centred on an O-atom. In LiF, both V_c and V_k polarons are stable at 0 K [8][9][10][11]. The excess hole is either centred on one F-atom or shared between two F-atoms, leading to the distortions shown in Figure 2.1.

2.2.2 Charge transition level

A key energetic property of polarons is the trapping energy, which is the energy gained when the excess charge becomes localized, distorting the surrounding lattice and forming a polaron. The trapping energy is a measure of the stability of polarons, and corresponds to the energy difference between the polaron state and the band edge, CBM for electron polarons and VBM for hole polarons. The energy difference is directly related to the CTL, which is defined as the Fermi energy at which the formation energies of different charge states are equal [12]. For hole polarons, it is relevant to study the CTL between the charge states $q = +1$ and $q = 0$, corresponding to systems with and without a hole polaron. The formation energy of a hole polaron of charge $q = +1$ is given by

$$E_F = G^f[\text{polaron}] - G^f[\text{bulk}] + q(E_{\text{VBM}} + \varepsilon_f) + E_{\text{corr}}, \tag{2.4}$$

where $G^f[\text{polaron}]$ is the Gibbs free energy of the system with a hole polaron, $G^f[\text{bulk}]$ of the bulk system without polarons, and where $q(E_{\text{VBM}}+\varepsilon_f)$ describes the

energy required to add a hole. This quantity depends on the Fermi energy ε_f, and is given relative to the energy of the VBM of the bulk system, E_{VBM}. E_{corr} is the correction term for finite-size effects. As mentioned in Section 1.1, the correction term is neglected in DFT calculations in the present work. For the charge state $q = 0$, there is no polaron and the formation energy is zero. Figure 2.2 shows a

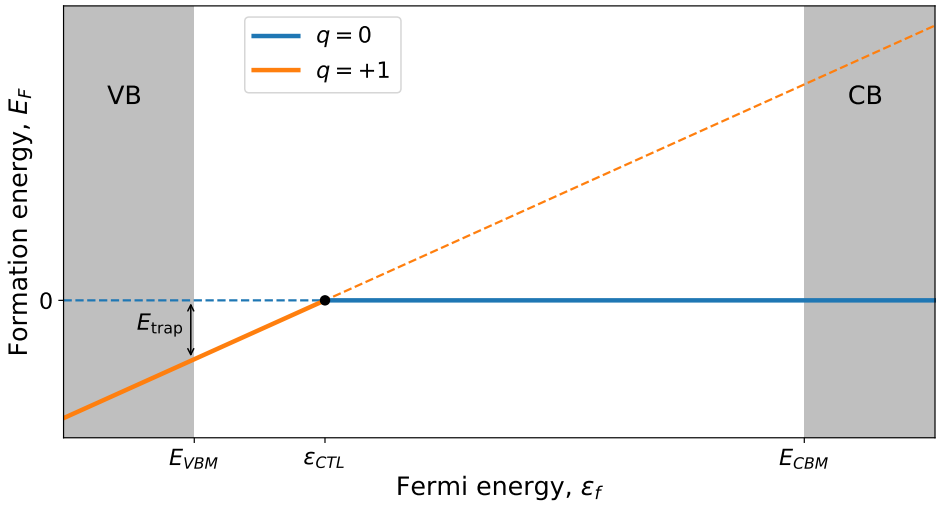


Figure 2.2: Schematic figure of the formation energies of charge states $q = +1$ and $q = 0$ as a function of Fermi energy. The CTL between the two states is shown with a black marker at the intersection. The solid lines indicate the thermodynamically favourable charge state for a given Fermi energy. E_{trap} is the trapping energy.

schematic example of the formation energy of a polaron of charge $q = +1$ and the CTL between charge states $q = +1$ and 0. The CTL is the Fermi energy at which the thermodynamically favourable charge state changes. The trapping energy is also indicated in the Figure 2.2. Note that the distance between the CTL and the E_{VBM} is the same value as E_{trap}, since the slope of the formation energy of the charge state $q = +1$ is one. Equating the formation energy in equation 2.4 with zero and inserting $q = +1$, the CTL is obtained as

$$\varepsilon_{\text{CTL}} = G^f[\text{bulk}] - G^f[\text{polaron}] - E_{\text{VBM}}. \tag{2.5}$$

To calculate the temperature dependence of the CTL, both the free energy difference, $G^f[\text{bulk}] - G^f[\text{polaron}]$, and the energy of the VBM, E_{VBM} must be evaluated at finite temperatures.

Physically, the CTL corresponds to the energetic position of the polaron state in the band gap. The CTL relative to the VBM, obtained through equation 2.5, is important in applications where excess holes move between the VBM and the polaron state and where the energy emitted/absorbed in the process is of interest. The absolute CTL can also be important. For example in photocatalysis, the absolute energetic position of the excess charge should match the energy required to drive the chemical reactions as closely as possible. The temperature dependence of the absolute CTL can be calculated by studying the free energy difference $G^f[\text{bulk}] - G^f[\text{polaron}]$ at finite temperatures.

2.2.3 Potential energy surfaces

The potential energy surface (PES) of the system can be described as a function of the configuration coordinate Q_{dist}, which describes the atomic configuration. In Figure 2.3, $Q_{\text{dist}} = 0$ corresponds to the undistorted pristine configuration and $Q_{\text{dist}} = 1$ corresponds to a polaron distorted atomic configuration, such as those shown in Figures 2.1. The ground state has a minimum at $Q_{\text{dist}} = 0$, as shown in Figure 2.3. When an electron-hole pair is excited, the system is elevated to an excited state. The PES for the excited state with a delocalized hole and delocalized electron is shown in blue, where the minimum is at potential energy E_g above the ground state, E_g being the band gap. The PES for the excited state with a localized hole is shown in red, with a minimum at $Q_{\text{dist}} = 1$. At $Q_{\text{dist}} = 0$, the hole has the energy of the VBM and at $Q_{\text{dist}} = 1$, the energy is lowered by the trapping energy E_{trap}. The electron remains delocalized and with an energy corresponding to the energy of the CBM. Since the electron remains unchanged, it does not need to be included in calculations. The excited state is modelled as the ground state with an excess hole, and the resulting energies as a function of Q_{dist} are simply shifted by the energy of the CBM, E_{CBM}. The shift corresponds to adding the excess electron. We refer to the addition of the excess electron through a shift of the potential energy surface by E_{CBM} as the addition of a "virtual" electron.

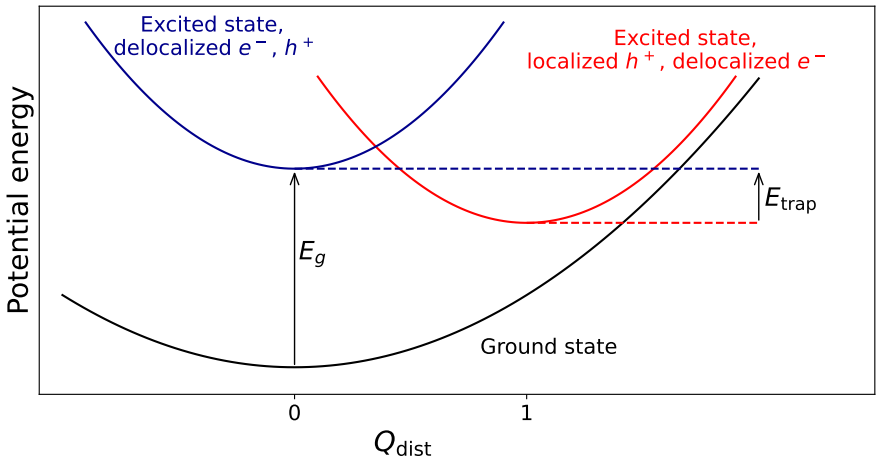


Figure 2.3: Schematic of potential energy surfaces for ground and excited state. The ground state is shown in black, while the delocalized and localized excited states are shown in blue and red, respectively. Q_{dist} is the configuration coordinate, where $Q_{\text{dist}} = 0$ corresponds to the undistorted pristine system and $Q_{\text{dist}} = 1$ corresponds to the polaron distorted system. E_g is the band gap and E_{trap} is the trapping energy.

2.3 Density functional theory

Density functional theory DFT is a widely used first-principles method for calculating ground state properties of many-body quantum systems. In this section, the basic theory of DFT is covered, starting with the Hohenberg-Kohn theorems and the Kohn-Sham equations. Different exchange correlation functionals are then

discussed, as well as the importance of considering the self-interaction error when modelling polarons. Lastly, hybrid functionals are described, which offer a method for taking the self-interaction error into account.

2.3.1 Hohenberg-Kohn theorems

According to the Born-Oppenheimer approximation, the nuclei can be considered fixed, as described in Section 2.1. In that case, the many-body wavefunction depends only on the electron coordinates. For a system with N electrons, the time independent Schrödinger equation can then be written as

$$\hat{H}\Psi(\boldsymbol{r}_1, ..., \boldsymbol{r}_N) = E\Psi(\boldsymbol{r}_1, ..., \boldsymbol{r}_N), \tag{2.6}$$

where $\boldsymbol{r}_i$ is the coordinate of electron i for $i = 1, ..., N$. The Hamiltonian $\hat{H}$ of the many-body system is given by the sum of the electron kinetic energy, the electron-electron interaction term and the the external potential caused by the electron's interaction with the nuclei. In atomic units, the Hamiltonian is given by

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \hat{V}_{\text{ext}}. \tag{2.7}$$

Although the terms are known, directly solving equation 2.6 for the $3N$-dimensional wavefunction is not feasible for almost any system of interest. A large reduction of dimensions can be achieved through DFT, where the central quantity is the 3-dimensional electron density rather than the many-body wavefuntion. Hohenberg and Kohn proved two theorems that are central to DFT in 1964 [13], which are referred to as the Hohenberg-Kohn theorems.

Theorem 1 *The electron density $n(\boldsymbol{r})$ uniquely determines the external potential $\hat{V}_{ext}$ up to a constant.*

Theorem 1 implies that given the electron density, the Hamiltonian in equation 2.7 is determined. Consequently, all ground and excited eigenstates, as well as all properties of the system, are determined by the electron density. In particular, the energy of the system must be a functional of the electron density, $E = E[n(\boldsymbol{r})]$.

Theorem 2 *For any positive-definite trial electron density $\widetilde{n}(\boldsymbol{r})$ fulfilling $\int \widetilde{n}(\boldsymbol{r})d\boldsymbol{r} = N$, we have $E_0 \leq E[\widetilde{n}(\boldsymbol{r})]$, where E_0 is the ground state energy.*

In other words, Theorem 2 states that there exists a universal functional $E[n(\boldsymbol{r})]$, that does not depend on the external potential $\hat{V}_{\text{ext}}(\boldsymbol{r})$, which could be minimized in order to obtain the ground state energy and electron density, if its form were known [14].

2.3.2 Kohn-Sham equations

From the form of the Hamiltonian in equation 2.7, we can see that the form of the energy functional must be the sum of a kinetic term, the electron-electron interaction term and the external potential term.

$$E[n] = T[n] + V_{\text{ee}}[n] + V_{\text{ext}}[n]. \tag{2.8}$$

In terms of the electron density, the last term is given by

$$V_{\text{ext}}[n] = \int \hat{V}_{\text{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r}. \tag{2.9}$$

However, the first two functionals in equation 2.8 are unknown. To address this, the Kohn-Sham ansatz can be used, which replaces the many-body system with an auxiliary non-interacting system of particles. In this fictitious system, the wavefunction is the antisymmetric product of single-particle wavefunctions ϕ_i, which are referred to as the Kohn-Sham orbitals and represent non-interacting electrons. The ground state electron density of the non-interacting system,

$$n(\boldsymbol{r}) = \sum_i |\phi_i|^2, \tag{2.10}$$

reproduces the ground state of the true system if an effective external potential is introduced to compensate for the error that is made by assuming that the particles are non-interacting. The kinetic energy functional of the non-interacting system is given by

$$T_0[n] = -\frac{1}{2} \sum_i \langle \phi_i | \nabla^2 | \phi_i \rangle . \tag{2.11}$$

Introducing also the energy corresponding to classical Coulomb interaction,

$$V_{\text{cc}}[n] = \frac{1}{2} \int \frac{n(\boldsymbol{r}_1) n(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} d\boldsymbol{r}_1 d\boldsymbol{r}_2, \tag{2.12}$$

we can write the total energy functional as

$$E[n] = T_0[n] + V_{\text{cc}}[n] + V_{\text{ext}}[n] + E_{\text{xc}}[n], \tag{2.13}$$

where $E_{\text{xc}}[n]$ is the sum of the error made by using the kinetic energy of the non-interacting system and the classical Coulomb energy instead of the true electron-electron interaction.

$$E_{\text{xc}}[n] = (T[n] - T_0[n]) + (V_{\text{ee}}[n] - V_{\text{cc}}[n]) \tag{2.14}$$

The single particle wavefunctions, ϕ_i, that minimize the energy can be found through the Kohn-Sham equations [15];

$$\left[-\frac{1}{2} \nabla^2 + \int \frac{n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r}' + \hat{V}_{\text{ext}}(\boldsymbol{r}) + \hat{V}_{\text{xc}}(\boldsymbol{r}) \right] \phi_i = \epsilon_i \phi_i. \tag{2.15}$$

The energies ϵ_i are the Kohn-Sham eigenenergies, and the exchange correlation potential $v_{\text{xc}}(\boldsymbol{r})$ is given by the functional derivative

$$\hat{V}_{\text{xc}}(\boldsymbol{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n}. \tag{2.16}$$

We notice that the $3N$-dimensionl many-body Schrödinger equation 2.6 has been reduced the 3-dimensional Kohn-Sham equations. If the exchange correlation potential were known, the ground state density would be reproduced exactly. Unfortunately, this is not the case. However, there are various methods for approximating the exchange correlation functional [14]. These methods vary in accuracy and computational cost, as is described in Sections 2.3.4 and 2.3.6.

2.3.3 Self-consistency cycle

Performing DFT calculations consists of solving the Kohn-Sham equations 2.15 consistently with equation 2.10. The equations constitute a coupled, non-linear system of equations, which cannot generally be solved directly. When performing DFT in practice, the electron density is calculated iteratively in a self-consistency cycle. The process starts with an initial guess for the electron density $n^{(0)}(\boldsymbol{r})$, which is used to calculate the Hamiltonian in the Kohn-Sham equations 2.15, using a suitable approximation for the exchange correlation functional. The Kohn-Sham orbitals ϕ_i are obtained by solving the equations 2.15, and the corresponding electron density $n(\boldsymbol{r})$ is then evaluated with the Kohn-Sham orbitals using equation 2.10. If the initial electron density $n^{(0)}(\boldsymbol{r})$ was correct, it should be consistent with the newly calculated electron density $n(\boldsymbol{r})$. If it is not consistent, the process is repeated with a new electron density. The cycle continues until a self-consistent electron density $n^{(k)}(\boldsymbol{r})$ is obtained, fulfilling $||n^{(k)}(\boldsymbol{r}) - n(\boldsymbol{r})|| < \delta$, where δ is the accepted margin of error. It may also be required that the total energy of the system is converged before interrupting the self-consistency cycle. Energies and forces of the system can then be calculated using the final electron density. The energy is obtained via the energy functional in equation 2.8, while the forces are obtained as the first derivatives of the energy with respect to the nuclear coordinates.

2.3.4 Local and semi-local approximations

While DFT is a powerful tool that has become the standard method for electronic structure calculations, the quality of the results is greatly dependent on the method of approximation for the exchange correlation functional. The simplest approach is the local density approximation (LDA). Define the exchange correlation energy per electron, $\varepsilon_{\text{xc}}(\boldsymbol{r})$, by

$$E_{\text{xc}}[n] = \int d\boldsymbol{r}\, n(\boldsymbol{r})\, \varepsilon_{\text{xc}}(\boldsymbol{r}). \tag{2.17}$$

In LDA, it is assumed that $\varepsilon_{\text{xc}}(\boldsymbol{r})$ only depends on the electron density at point $\boldsymbol{r}$. Then, we have $\varepsilon_{\text{xc}}(\boldsymbol{r}) = \varepsilon_{\text{xc}}(n(\boldsymbol{r}))$. Using the exchange correlation energy of a homogeneous electron gas, which is accurately known and parametrized, we obtain

$$E_{\text{xc}}^{\text{LDA}}[n] = \int d\boldsymbol{r}\, n(\boldsymbol{r})\, \varepsilon_{\text{xc}}^{\text{hom}}(n(\boldsymbol{r})), \tag{2.18}$$

where $\varepsilon_{\text{xc}}^{hom}(n(\boldsymbol{r}))$ is the exchange correlation energy of a homogeneous electron gas of constant density $n(\boldsymbol{r})$. The LDA is reasonable for systems with slowly varying densities.

The generalized gradient approximation (GGA) offers a more accurate way of calculating the exchange correlation energy using a semi-local approximation. Both the electron density itself and its gradient at point $\boldsymbol{r}$ are used to approximate ε_{xc}.

$$E_{\text{xc}}^{\text{GGA}}[n] = \int d\boldsymbol{r}\, n(\boldsymbol{r})\, \varepsilon_{\text{xc}}^{\text{GGA}}(n(\boldsymbol{r}), \nabla n(\boldsymbol{r})). \tag{2.19}$$

An example of a GGA functional is the PBE functional [16]. While this approximation is a substantial improvement over LDA, there are still problems such as the self-interaction error (SIE) that are not corrected in this method [17].

2.3.5 Self-interaction error

In order to accurately describe polarons, it is crucial to choose an exchange correlation functional that takes the SIE into consideration. Conceptually, the SIE arises from the substitution of individual electrons with the total density $n(\boldsymbol{r})$ in the electron-electron interaction term V_{cc} (see equation 2.12). If the electrons both give rise to the potential through $n(\boldsymbol{r})$, and also interact with it, they are essentially interacting with themselves. The SIE is the error that arises from this spurious inclusion of electrons interacting with themselves, and leads to many quantitative and qualitative errors in DFT, one such error being excessive charge delocalization [18]. Polaron localization and formation energies are sensitive to the description of this error [19].

Two types of self-interactions can be defined: the one-body and the many-body self-interaction. In a system with only one electron, there are no electron-electron interactions and the system can be defined as free from the one-body SIE if the classical coulomb energy in equation 2.13 is cancelled exactly by the exchange correlation functional, $V_{\text{cc}}[n] + E_{xc}[n] = 0$. This is fulfilled in Hartree-Fock theory, which is briefly described in Section 2.3.6. On the other hand, the self-interaction in a system with many electrons is more difficult to formulate mathematically in an intuitive way. The many-body self-interaction has been defined as the deviation from piecewise linearity of the system's total energy as a function of electron occupation. Exchange correlation functionals that yield energies fulfilling this condition of piecewise linearity are free from the many-body SIE [19]. Figure 2.4 shows the total energy as a function electron occupation. An exact calculation would yield total energy that is piecewise linear with electron occupation. However, both Hartree-Fock theory and semi-local functionals suffer from the many-body SIE, which can be seen from their erroneous curvature. Hybrid functionals offer a relatively simple method for the correction of the many-body SIE.

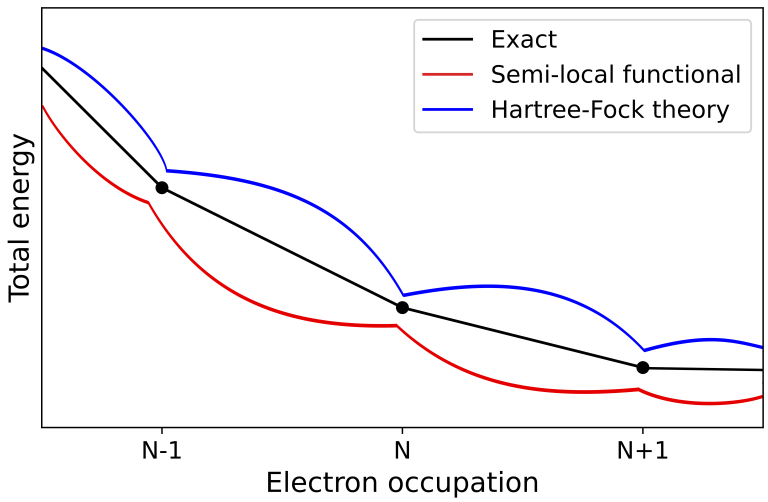


Figure 2.4: Schematic figure of the total energy as a function of electron occupation. An exact calculation would yield piecewise linear total energy, while semi-local functionals and Hartree-Fock theory yield erroneous curvature.

2.3.6 Hybrid functionals

Hybrid functionals include a portion of exact exchange from Hartree-Fock theory into the exchange correlation functional. The proportion of exact exchange is chosen such that the total energy of the system is piecewise linear as a function of electron occupation, allowing for calculations that are free from the many-body SIE. Although hybrid functionals offer more accurate calculations, they also require more computational resources.

In Hartree-Fock theory, the many-body wavefunction is approximated as the antisymmetric product of single electron wavefunctions. Using this approximation and the Hamiltonian in equation 2.7, a term for the exchange energy of the system can be derived [14]. This term is the exact exchange energy, $E_{\mathrm{x}}^{\mathrm{HF}}[n]$. Hybrid functionals combine exact exchange $E_{\mathrm{x}}^{\mathrm{HF}}[n]$ with a DFT functional, typically a GGA functional, and can be expressed as

$$E_{\mathrm{xc}}^{\mathrm{hyb}} = \alpha E_{\mathrm{x}}^{\mathrm{HF}} + (1-\alpha)E_{\mathrm{x}}^{\mathrm{DFT}} + E_{\mathrm{c}}^{\mathrm{DFT}}, \tag{2.20}$$

where $E_{\mathrm{x}}^{\mathrm{DFT}}$ and $E_{\mathrm{c}}^{\mathrm{DFT}}$ are the exchange and correlation energies from the DFT functional, and α is the proportion of exact exchange from Hartree-Fock theory. In this project, the hybrid functional PBE0 is used with mixing proportions $\alpha = 0.34$ for MgO and $\alpha = 0.50$ for LiF. These values have been found to satisfy the piecewise linearity condition described in Section 2.3.5 for MgO [6] and LiF [20] in previous studies.

2.4 String method

DFT can be combined with the string method to find the minimum energy path (MEP) between two stable states, each corresponding to a local minimum in the energy landscape. The MEP is defined as the path in configuration space along which the potential force is always parallel to the path. The energy barrier between the two states is obtained by calculating the highest energy along the MEP. In this project, the MEP between the $\mathrm{V_c}$ and $\mathrm{V_k}$ polaron states in LiF is calculated in order to assess the stability of the $\mathrm{V_c}$ polaron and determine whether the $\mathrm{V_c}$ polaron is likely to transition to a $\mathrm{V_k}$ polaron at finite temperatures.

The string method is a chain-of-states method. A number of images are created along a path between the initial and final states, and their positions are optimized in order to find the MEP. The most intuitive approach is to simply move the images according to the potential force. However, this leads to problems such as the accumulation of images in low-energy regions and difficulties obtaining saddle points along the path. Various methods can be used to improve convergence to the correct path and to enforce a suitable distribution of the images along the path. In the nudge-elastic-band method, which is another common method for finding the MEP, the images are connected with springs of a certain spring constant to enforce their distribution. In the string method, the distribution of the images is controlled by enforcing a particular parametrization of the path [21].

The string method can also be combined with the climbing image technique for accurate calculation of saddle points [22]. In order to calculate the energy barrier

between the initial and final state, one of the images must finish at the point of highest energy along the MEP. This is achieved by identifying the image of highest energy after a few iterations and forcing it to maximize its energy along the path. The chosen image is referred to as the climbing image. The energy barrier is calculated as the difference in energy between the initial image and the climbing image.

2.5 Machine learned interatomic potentials

MLIPs have become increasingly popular as a way of overcoming the high computational costs of first-principles methods such as DFT. The MLIP used in this project is named neuroevolution potential (NEP), and the basic theory behind it is explained in Section 2.5.1. Generally, a NEP cannot take excess charges into account. However, the labelling method allows for a NEP to identify atoms with localized excess charges and learn to emulate the site energies and forces around them. The labelling method is described in Section 2.5.2.

2.5.1 Neuroevolution potential

The NEP model is a simple feedforward neural network that is trained with GPUMD [23]. Given an atomic structure, the NEP calculates the potential energy of the total system by summing the site energies, U_i, of all atoms in the system.

$$U_{tot} = \sum_i U_i \tag{2.21}$$

The site energy can generally be expressed as a function of the local environment surrounding the atom,

$$U_i = U_i(\{\boldsymbol{r}_{ij}\}_{j \in N_i}), \tag{2.22}$$

where $\boldsymbol{r}_{ij}$ is the position difference between atoms i and j, and N_i are the atoms surrounding atom i [24]. In GPUMD, the local environment is described by assigning each atom a descriptor vector, $\boldsymbol{q^i}$, characterizing the structural and chemical environment within a certain cutoff radius. The descriptor consists of a number of radial and angular components, and has a total of N_{des} components. The neural network has a single hidden layer with N_{neu} neurons. The site energy of atom i is a function of the descriptor vector, given by

$$U_i = \sum_{\mu=1}^{N_{\text{neu}}} w_\mu^{(1)} \tanh\left(\sum_{\nu=1}^{N_{\text{des}}} w_{\mu\nu}^{(0)} q_\nu^i - b_\mu^{(0)}\right) - b^{(1)}, \tag{2.23}$$

where $\boldsymbol{w}^{(0)}$ is the connection weight matrix from the descriptor vector to the hidden layer and $\boldsymbol{w}^{(1)}$ is the connection weight vector from the hidden layer to the output node. The activation function in the hidden layer is $\tanh x$, whereas $\boldsymbol{b}^{(0)}$ and $b^{(1)}$ are the bias vectors for the hidden layer and for the output node, respectively.

The total number of trainable parameters is the sum of the neural network parameters and the trainable parameters in the descriptor vectors. The parameters

are optimized by minimizing a loss function consisting of energy, force and virial loss terms, as well as two regularization terms for the $\mathcal{L}_1$ and $\mathcal{L}_2$ parameter norms. The loss function is shown in equation 2.24, where $\boldsymbol{z}$ represents all the parameters in the NEP model and $\varepsilon_i(\boldsymbol{z})$ for $i = e, f, v, 1, 2$ is a term that gives a measure of the errors on the dataset for energies, forces, virials and $\mathcal{L}_1$ and $\mathcal{L}_2$ parameter norms.

$$L(\boldsymbol{z}) = \lambda_e\varepsilon_e(\boldsymbol{z}) + \lambda_f\varepsilon_f(\boldsymbol{z}) + \lambda_v\varepsilon_v(\boldsymbol{z}) + \lambda_1\varepsilon_1(\boldsymbol{z}) + \lambda_2\varepsilon_2(\boldsymbol{z}) \tag{2.24}$$

Each term in the loss function is weighted by a hyperparameter λ_i. Their values can be varied to improve the optimization, as is discussed in Section 3.4.3.

To calculate the forces on each atom with NEP, the derivative of the potential energy is used [24],

$$\boldsymbol{F}_i = -\frac{\partial}{\partial \boldsymbol{r}_i}\sum_j U_j, \tag{2.25}$$

where $\boldsymbol{r}_i$ is the position of atom i. Equation 2.25 can be rewritten as

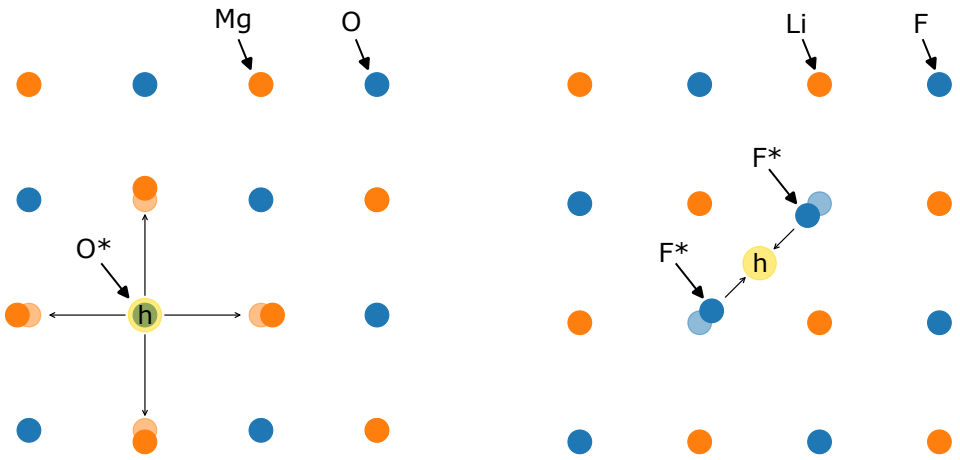
$$\boldsymbol{F}_i = \sum_{i\neq j}\frac{\partial U_i}{\partial \boldsymbol{r}_{ij}} - \frac{\partial U_j}{\partial \boldsymbol{r}_{ji}} = \sum_{i\neq j}\boldsymbol{F}_{ij}, \tag{2.26}$$

where $\boldsymbol{r_{ij}} = \boldsymbol{r_j} - \boldsymbol{r_i}$, and $\boldsymbol{F}_{ij}$ is the force experienced by atom i from atom j. In this form, it is clear that $\boldsymbol{F}_{ij} = -\boldsymbol{F}_{ji}$ is fulfilled.

2.5.2 Labelling method

The input structures for a NEP consist of a list of atomic positions and species. In order to learn the physics around a polaron, the model must be able to differentiate between atoms with and without localized excess charges. Since the NEP does not take charges as input, the differentiation is done by labelling atoms with localized excess charges in the input structures for the model. In this way, the model is able to learn that labelled atoms lead to different site energies and forces compared to unlabelled atoms. The labelling method is enabled by the fact that the spatial extension of the effects of a small polaron is limited to its close surroundings. Using the method, a single NEP can model both pristine and polaron systems [2].

For V_c polarons, the atom at the centre of the distortions is labelled. For V_k polarons, the two atoms that share the localized excess charge are labelled. Figure 2.5 shows how the atoms are labelled for the two cases that are relevant in this project. In MgO, the excess hole in the V_c polaron is centred an oxygen atom, which is labelled O*. In the case of the V_k polaron in LiF, the excess hole is shared between two fluoride atoms, which are both labelled as F*. As mentioned in Section 1.1, NEPs are not trained for the V_c polaron in LiF, due to the low energy barrier that was found between the V_c and V_k polaron states. At finite temperatures, the V_c polaron is likely to transition to the V_k polaron. However, the transition cannot be modelled with the labelling method. This is a limitation of the method. While performing MD simulations, the labelling of an atom cannot change. For this reason, the labelling method is unable to describe both the transition between different polaron types and the hopping of polarons between different atomic sites.



(a) (b)

Figure 2.5: **(a)** Labelling method applied to the V_c polaron in MgO. The atom where the excess hole is localized is labelled as O*. **(b)** Labelling method applied to the V_k polaron in LiF. The excess hole is shared between two fluoride atoms which are both labelled as F*.

2.6 Molecular dynamics

Given initial positions and velocities for the atoms in a structure, and a way to calculate the forces acting on them, one can find the structure's phase trajectory through MD simulations. In GPUMD, the velocity-Verlet integration method is implemented [24]. The method is based on the Taylor expansion for the position of atom i at time $t + \Delta t$ in terms of velocity $\boldsymbol{v}_i(t)$ and forces $\boldsymbol{F}_i(t)$.

$$\boldsymbol{r}_i(t + \Delta t) \approx \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t)\Delta t + \frac{\boldsymbol{F}_i(t)}{2m_i}(\Delta t)^2 \tag{2.27}$$

In order to propagate further, the velocity also needs to be calculated at $t + \Delta t$. Taylor expanding the position backwards in time yields

$$\boldsymbol{r}_i(t - \Delta t) \approx \boldsymbol{r}_i(t) - \boldsymbol{v}_i(t)\Delta t + \frac{\boldsymbol{F}_i(t)}{2m_i}(\Delta t)^2. \tag{2.28}$$

By moving one step forward in time we obtain

$$\boldsymbol{r}_i(t) \approx \boldsymbol{r}_i(t + \Delta t) - \boldsymbol{v}_i(t + \Delta t)\Delta t + \frac{\boldsymbol{F}_i(t + \Delta t)}{2m_i}(\Delta t)^2. \tag{2.29}$$

Combining equation 2.27 with equation 2.29, we obtain

$$\boldsymbol{v}_i(t + \Delta t) \approx \boldsymbol{v}_i(t) + \frac{\boldsymbol{F}_i(t) + \boldsymbol{F}_i(t + \Delta t)}{2m_i}\Delta t. \tag{2.30}$$

Equations 2.27 and 2.30 constitute the velocity-Verlet algorithm. Propagating an MD simulation with this algorithm corresponds to an NVE ensemble, where the

number of particles, N, the volume, V, and the energy, E, are constant. In most practical applications, it is uncommon to use the NVE ensemble, since this would correspond to an isolated system. To simulate an NVT or NPT ensemble instead, various thermostats and/or barostats can be implemented. In the present work, MD is used to evaluate the NEP, calculate the RDF and perform free energy calculations with thermodynamic integration.

2.6.1 Radial distribution function

In MD simulations, bond lengths and lattice distortions can be studied with the radial distribution function (RDF) [25], which describes the probability of finding an atomic pair with distance r between the atoms. The software GPUMD, which is used in this project, defines the RDF as

$$g(r) = \frac{V}{N^2} \frac{dN(r, dr)}{4\pi r^2 dr}, \tag{2.31}$$

where V is the volume of the system, N is the number of particles, and $N(r, dr)$ is the number of atomic pairs with distances within r and $r + dr$ between the atoms. In many cases, it is relevant to calculate the partial RDF between specific atomic species. The partial RDF $g_{ab}(r)$ between atom types a and b is given by

$$g_{ab}(r) = \frac{V}{N_a N_b} \frac{dN_{ab}(r, dr)}{4\pi r^2 dr}, \tag{2.32}$$

where N_a and N_b are the number of atoms of type a and type b respectively, and $N_{ab}(r, dr)$ is the number of a-b pairs with distances within r and $r + dr$ between the atoms.

2.7 Thermodynamic integration

Thermodynamic integration is used in this project to calculate the free energy difference between the pristine and polaron systems. The free energy difference is calculated in terms of the irreversible work performed on the system along a path connecting the two systems, and can be obtained through MD simulations using NEPs. This section describes the theory of the method.

The Boltzmann density describes the probability that a system in the canonical ensemble (NVT) is in a certain state. It is defined as

$$\rho(p, q) = \frac{e^{-\beta H(p,q)}}{h^{3N} N! Z}, \tag{2.33}$$

where p and q are the positions and momenta of the N particles, and H is the Hamiltonian [26]. The normalisation factor Z is the canonical partition function, given by the following integral over phase space

$$Z(T, V, N) = \int d^{3N}p \, d^{3N}q \frac{e^{-\beta H(p,q)}}{h^{3N} N!}. \tag{2.34}$$

From the partition function, the free energy of the system can be defined as

$$G = -\beta \ln Z. \tag{2.35}$$

The Boltzmann density and partition function can be generalized to other ensembles, yielding the corresponding free energies. To calculate the free energy difference between two systems, the free energy can be integrated along a path connecting the two systems. An example of one such path is

$$H(\lambda) = \lambda H_f + (1-\lambda)H_i, \tag{2.36}$$

where H_i and H_f are the Hamiltonians of the initial and final systems, and λ is varied from 0 to 1. This corresponds to the Frenkel-Ladd path when one of the Hamiltonians is that of an Einstein crystal [27]. The derivative of the free energy with respect to λ can be written as [28]

$$\frac{\partial G(\lambda)}{\partial \lambda} = -\frac{1}{\beta}\frac{\partial}{\partial \lambda}(\ln Z) = -\frac{1}{\beta Z}\frac{\partial Z}{\partial \lambda}. \tag{2.37}$$

Inserting the definition of the partition function, we obtain

$$\frac{\partial G(\lambda)}{\partial \lambda} = \frac{1}{Z}\int d^{3N}p\, d^{3N}q \frac{e^{-\beta H(\lambda)}}{h^{3N}N!}\frac{\partial H}{\partial \lambda}, \tag{2.38}$$

which is the Boltzmann-weighted average of $\partial H/\partial\lambda$, denoted with $\langle ... \rangle_\lambda$.

$$\frac{\partial G(\lambda)}{\partial \lambda} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \tag{2.39}$$

Using the Hamiltonian in equation 2.36, we obtain the free energy difference between two systems as

$$\Delta G = G(\lambda = 1) - G(\lambda = 0) = \int_0^1 \left\langle H_f - H_i \right\rangle_\lambda d\lambda. \tag{2.40}$$

When λ is changed infinitely slowly along the path so that the system is always in equilibrium, the total work performed on the system is equal to ΔG [29]. The work along such a path is known as the reversible work, W_{rev}. In equation 2.40, it is assumed that the process is reversible. If the system were to undergo irreversible changes along the path, it would exhibit hysteresis, causing the equality in equation 2.40 to become incorrect [28]. This can happen if λ is changed at a finite rate. The work performed on the system then becomes a stochastic variable which differs for each realization and is known as the irreversible work, W_{irr}. These quantities are related by

$$\Delta G = W_{\text{rev}} = \overline{W_{\text{irr}}} - \overline{E_{\text{diss}}}, \tag{2.41}$$

where E_{diss} is the dissipated heat along the non-equilibrium path and the overline denotes the average over an ensemble of realizations. However, if the non-equilibrium path is close enough to a reversible, infinitely slow, equilibrium path, the dissipated

heat is the same for the irreversible processes that are carried out in opposite directions. We then have $\overline{E^{\text{diss}}_{i\to f}} = \overline{E^{\text{diss}}_{f\to i}}$. In this case, the dissipated heat cancels itself out, and the free energy difference can be calculated as

$$\Delta G = \frac{1}{2}\left[W^{\text{rev}}_{i\to f} - W^{\text{rev}}_{f\to i}\right] = \frac{1}{2}\left[\overline{W^{\text{irr}}_{i\to f}} - \overline{W^{\text{irr}}_{f\to i}}\right]. \tag{2.42}$$

The convergence of the free energy as a function of the rate by which λ is varied can be used to determine whether the path is slow enough for the equation to hold true.

To calculate the irreversible work, the path is made explicitly time dependent. The path then becomes $\lambda = \lambda(t)$, where the value of $\lambda(t)$ goes from 0 to 1, as t goes from 0 to t_s, the switching time. The irreversible work is given by

$$W^{\text{irr}}_{i\to f} = \int_0^{t_s} dt \frac{d\lambda}{dt}\left(\frac{\partial H}{\partial \lambda}\right)_{\mathbf{\Gamma}(t)}, \tag{2.43}$$

where $\mathbf{\Gamma}(t)$ is the phase space trajectory taken along the path. In numerical calculations, the integral becomes the following sum, where the Hamiltonian in equation 2.36 has been inserted. We have

$$W^{\text{irr}}_{i\to f} = \sum_k \Delta\lambda_k \left(H_f - H_i\right)_{\mathbf{\Gamma}(k\Delta t)}, \tag{2.44}$$

where Δt is the MD time step and $\Delta\lambda_k$ is the amount by which λ is varied between steps k and $k+1$ [27]. Combining equations 2.42 and 2.44, the free energy is obtained.

In the case of the polaron, the initial and final Hamiltionians H_i and H_f correspond to two NEPs, the first being a NEP for the pristine system and the second for the polaron system. Using MD, the system is propagated between the two systems using the Hamiltonian in equation 2.36. The free energy is calculated for different switching times t_s, in order to find the rate of λ-variation at which equation 2.42 converges. A number of simulations are performed in order to evaluate the averages in equation 2.42. Repeating the process at different temperatures allows for analysis of the temperature dependence of the free energy difference, which is used in equation 2.5 for analysis of the CTL.

3
Methods

In this project, hole polarons were modelled in MgO and LiF at finite temperatures with NEPs that are trained on reference data produced with hybrid DFT calculations. The transition between the V_k and V_c polaron states in LiF was first studied by calculating the MEP between the two polaron states with the string method. The calculations are used to evaluate the energy barrier between the two polaron states. NEP models were then trained for the V_c polaron in MgO and for the V_k polaron in LiF. The models were used to calculate the temperature dependence of the CTL and of the polaron distortions. In this chapter, the methods used to obtain these results are discussed. An overview of the workflow used throughout the project is first given in Section 3.1, while more detailed descriptions of the methods are given in the remaining sections.

3.1 Project workflow

The steps necessary to analyse polaron properties at finite temperatures can be categorised into relaxation, NEP training and production of results. The general workflow is shown in Figure 3.1. The initial structures for cubic rock-salt MgO and LiF were taken from the Materials Project database, with IDs mp-1265 and mp-1138, respectively. Relaxation calculations were first performed with DFT using the FHI-aims software to obtain relaxed structures for each material. Relaxation was performed for the pristine system in both materials, for the V_c and V_k polaron structures in LiF, and for the V_c polaron structure in MgO. The relaxed V_c and V_k polaron structures in LiF were then used for string method calculations, enabling the study of the MEP between the two states. The relaxed V_c polaron structure in MgO and the V_k polaron structure in LiF were used to train NEP models, which were then used to study the two polarons at finite temperatures.

A first dataset for NEP training was obtained through rattling, which was done with the Python packages Atomic Simulation Environment (ASE) and hiPhive. Rattled structures were obtained by adding small, random displacements to the atoms in the relaxed structures. Rattling was performed for both the pristine structures and for the relevant polaron structures. By performing single-point DFT calculations on the rattled structures, properties such as energy and forces were obtained for each structure, allowing us to create a dataset for NEP training. Importantly, polaron structures were labelled according to the labelling method. The first generation (G1) of NEP models was trained with the rattled dataset, and the models were then improved iteratively through cycles of active learning, in which new structures for

the dataset were obtained through MD simulations. All NEP models were trained using the GPUMD software with the nep executable, while the gpumd executable was used for MD simulations. NEP training and active learning are described in more detail in Section 3.4.

Once the final generation of NEP models was obtained, production calculations were performed. In order to calculate the temperature dependence of the CTL, both free energy and VBM calculations at finite temperatures are necessary, as shown in equation 2.5. The temperature dependence of the free energy difference $G^f[\text{bulk}] - G^f[\text{polaron}]$ was obtained via thermodynamic integration, whereas the VBM was obtained at finite temperatures using a combination of MD simulations and single-point DFT calculations. Thermodynamic integration and VBM calculations are described in Section 3.5. In addition to the CTL, the final NEP models were also used to calculate the RDFs of the pristine and polaron systems through MD, which yielded polaron distortions at finite temperatures.

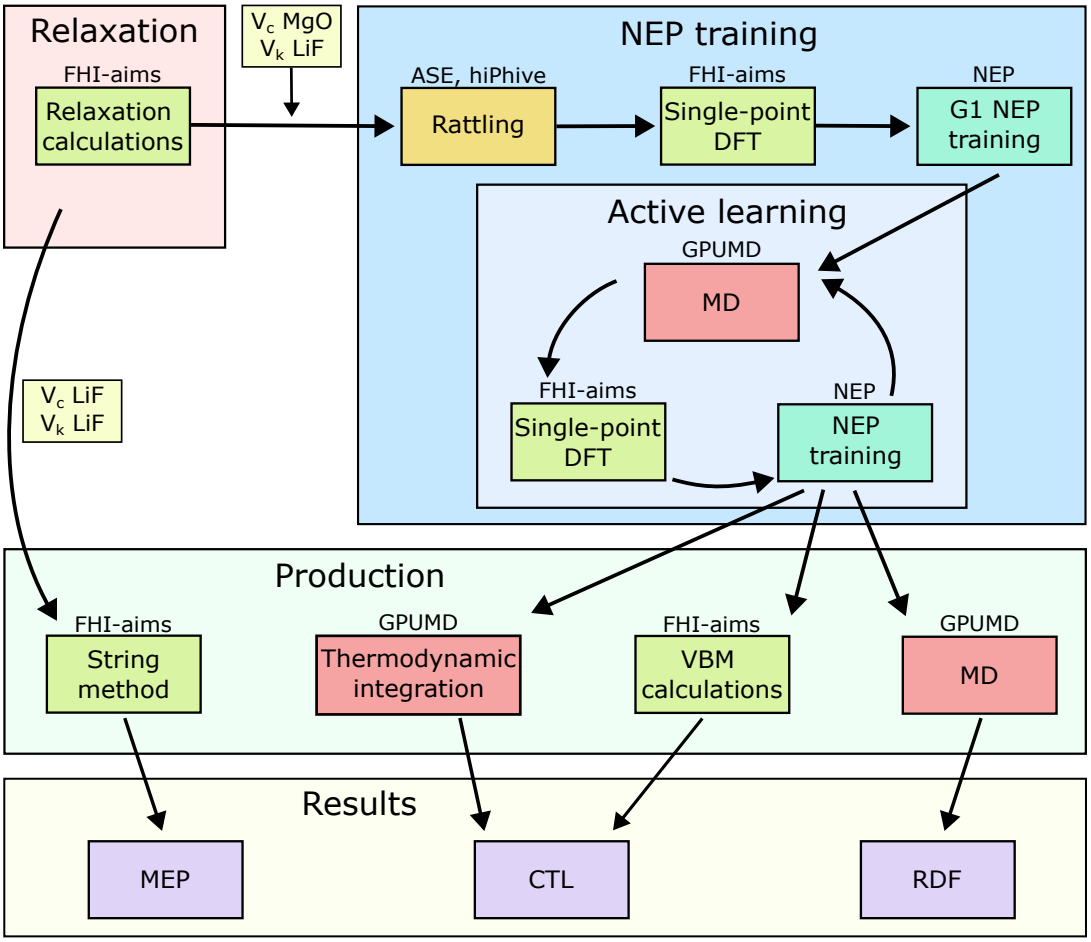


Figure 3.1: Workflow of the present project. NEP models are trained by active learning, enabling calculation of the CTL and RDF at finite temperatures. Additionally, the MEP between the two polaron states in LiF is obtained using string method calculations.

3.2 DFT calculations

All DFT calculations were performed using the FHI-aims software [30][31] with the hybrid functional PBE0, with mixing proportions $\alpha = 0.34$ for LiF and $\alpha = 0.50$ for MgO (see Section 2.3.6). Supercells containing 216 atoms were used for both materials, corresponding to a $3 \times 3 \times 3$ repetition of the unit cell. Only the Γ-point was sampled in reciprocal space, corresponding to the wavevector $\boldsymbol{k} = (0, 0, 0)$. The convergence criterion for the self-consistency cycle was based on both the electron density and the total energy. The cycle was interrupted when the difference in volume-integrated root-mean square change in electron density between the present and previous iterations was below 10^{-4} electrons, and when the change in energy was below 10^{-4} eV. Settings for the basis set were taken from the pre-constructed default definitions provided by FHI-aims. The basis set is the basis of functions used to represent the electron density and must be specified for each atomic species. FHI-aims provides different levels of settings ranging from "light" to "really tight". The "light" settings were used for all DFT calculations in this project. The same settings were used for pristine and polaron structures, apart from the charge of the system and the use of spin-polarized calculations. The pristine structures are neutral and the calculations were spin-restricted, meaning that electrons of opposite spin are paired together and all occupied states have exactly zero spin moment. On the other hand, polaron structures have an odd number of electrons. The net charge is $q = +1$ and spin-polarized calculations were performed, allowing occupied states to have non-zero spin moment.

Both relaxation and single-point calculations were performed with DFT. In a relaxation calculation, the structure is relaxed until there are no force components for any atoms that exceed 0.01 eV/Å. At this point, properties such as the energy and band structure of the system are evaluated, as well as atom-specific properties such as the spin moment or charge on each atom. To obtain atom-specific properties, the electron density must be partitioned among the atoms in order to define the fraction of it that can be associated with each atom. The partitioned electron density is then used to calculate atom-specific properties. Hirshfeld partitioning is a method for dividing the electron density into atomic contributions, which is available in the FHI-aims software. The atom-specific Hirshfeld spin moment was evaluated when performing DFT calculations for polaron structures to measure the localization of the excess hole. This was important for determining which structures could be included in the training set for the NEP models, as described in Section 3.4. In single-point calculations, all the same properties are calculated, but without relaxing the geometry of the structure.

3.3 String method calculations

In order to study the transition between the V_c and V_k polaron in LiF, string method calculations were performed with FHI-aims. Using the relaxed structures for each polaron type as initial and final structures, the MEP between the two polaron states was calculated, yielding the energy barrier between the two states as well as the

Hirshfeld spin moment on each atom along the path. Six images were used between the initial and final states, and the climbing image technique was implemented to ensure the correct calculation of the energy barrier.

The energy barrier between the two states was used to determine whether NEP models could be trained for both polaron types in LiF. If the energy barrier is small, V_c polarons are likely to transition to V_k polarons when the thermal energy is increased. However, the labelling method is not able to model the transition, since the label of an atom cannot change in the middle of a simulation. As shown in Section 4.2, the energy barrier was low, motivating the choice to only model the V_k polaron in LiF, together with the V_c polaron in MgO.

3.4 NEP training

As previously mentioned, the NEP models were trained in cycles of active learning. The process is illustrated in Figure 3.2. Rattled structures were first created by randomly drawing displacements from a normal distribution with standard deviation 0.075 Å, and applying the displacements to the atoms in the relaxed pristine structures and the relevant polaron structures. The first dataset for each material was obtained by performing single-point DFT calculations on the rattled structures, and labelling the atoms with localized excess charges according to the labelling method (see Section 3.4.1). The dataset, consisting of labelled rattled structures together with their corresponding energy, forces and stress tensor, was divided into five random splits. For each split, 50 structures were randomly extracted from the training set and instead used as a test set, enabling the evaluation of the models' performance on structures that it had not been trained on. The five NEP models, each trained on one split, constitute the first generation (G1). The models were then improved upon through the process of active learning, which entails creating more training data using structures of high uncertainty. The uncertainty of a structure is a measure of how well the five models agree upon a prediction, and is defined as

$$\sigma_f = \max_i \sqrt{\sigma_{i,x}^2 + \sigma_{i,y}^2 + \sigma_{i,z}^2}, \tag{3.1}$$

where σ_f is the uncertainty of a structure, and $\sigma_{i,k}^2$ is the variance of the predictions from the five models for the force on atom i in Cartesian direction $k \in x, y, z$ [32][33]. Structures of high uncertainty were obtained by performing MD simulations. While the propagation of the MD simulations was done by one of the models, the uncertainties were calculated along the trajectory using all five models. Structures of higher uncertainty were then extracted from the trajectory and single-point DFT calculations were performed on them to obtain energies, forces and stress tensors. Lastly, before adding the new structures to the dataset, the localization of the excess charge must be checked and the atoms labelled accordingly, as described in Section 3.4.1. In all generations, splits were created by randomly extracting 50 structures for the test set. The process of retraining and using the next generation of models to obtain more training data for the next cycle was repeated until acceptable models were obtained. The decision to interrupt the cycle and start production calculations with the models was based on performance indicators such as root mean square error

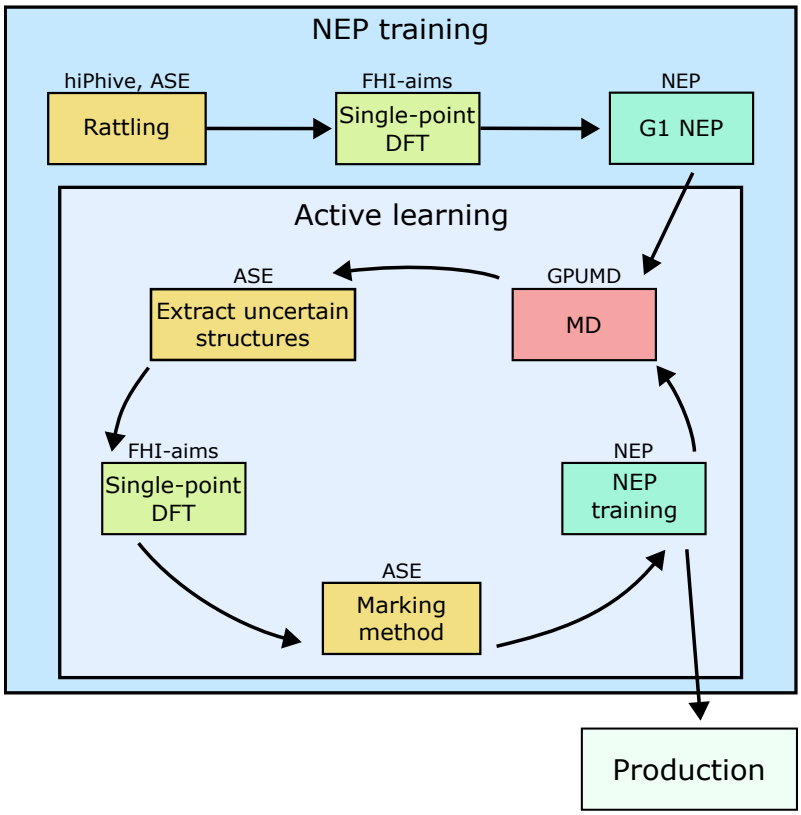


Figure 3.2: Workflow of active learning. Using the first generation (G1) of NEPs, MD is performed to obtain structures of high uncertainty which are added to the dataset after single-point DFT calculations. Using the labelling method, a new generation of NEPs is trained on the dataset, with which the process is repeated until acceptable models are obtained.

(RMSE) for the training and test set, convergence of the loss function, uncertainties in MD simulations and agreement with DFT calculations at low temperatures. When enough training data was obtained, a full split was also trained using the whole dataset.

The structures in the training set for the final model can be divided into three categories: rattled structures, active learning structures and path structures. Rattled structures refer to the structures with random displacements that are used to train the first generation of NEPs. Active learning structures are structures of high uncertainty from MD simulations that were added to the dataset during active learning. Lastly, path structures are structures that were included in the last generation of models to improve their behaviour at low temperatures. Path structures are described in Section 3.4.2. Table 3.1 in Section 3.4.3 shows the number of training structures of each type that were included in the final dataset. For both MgO and LiF, active learning structures were created with the first and second generations to obtain a third generation. A fourth generation was then trained, where path structures were also included, and used as the final model.

3.4.1 Labelling method

When adding new polaron structures to the dataset, it must be checked that the excess hole is localized correctly. As previously mentioned, the localization of the excess hole was measured in terms of the Hirshfeld spin moment. However, the localization can vary continuously across the atoms. Only structures with spin moment values that fulfil the following conditions were included in the dataset. In pristine structures, there are no excess charges and the spin moment is zero on all atoms. For V_c polaron structures in MgO, it was required that the labelled atom had spin moment $s_{\text{labelled}} > 0.3\,\mu_B$, where μ_B is the Bohr magneton, and that all other atoms in the same structure had spin moment $s_{\text{other}} < s_{\text{labelled}} - 0.25\,\mu_B$. For V_k polarons in LiF, it was required that both labelled atoms had spin moment $s_{\text{labelled}} > 0.33\,\mu_B$. Polaron structures fulfilling these conditions were labelled according to Figure 2.5 and added to the dataset for the NEP models.

3.4.2 Path structures

After training the first three generations, it was found that the models in both materials made inaccurate predictions at low temperatures. The predicted energies of relaxed bulk and polaron structures were consistently underestimated, and the models for MgO incorrectly relaxed the geometry of the V_c polaron. For this reason, specific low temperature structures along two paths were added to the dataset of the fourth generation. We refer to these paths as path A and path B.

Path A is the linear interpolation between the relaxed bulk and polaron structures from DFT, corresponding to the atomic positions $\mathbf{A}_{\text{bulk}}^{\text{DFT}}$ and $\mathbf{A}_{\text{pol}}^{\text{DFT}}$. Introduce $\boldsymbol{\eta}$ as a vector of 11 evenly spaced values in the interval $[-0.5, 1.5]$. Then, we have $\boldsymbol{\eta} = (-0.5, -0.3, ..., 1.3, 1.5)$. For $i = 1, ..., 11$, the following structure was added to the dataset.

$$\mathbf{A}_i^{\text{Path A}} = \mathbf{A}_{\text{bulk}}^{\text{DFT}} + \eta_i(\mathbf{A}_{\text{pol}}^{\text{DFT}} - \mathbf{A}_{\text{bulk}}^{\text{DFT}}) \tag{3.2}$$

The structures were added both with and without an excess charge. This corresponds to 11 bulk structures and 11 polaron structures. The structures were included for both LiF and MgO. Note that these structures correspond to the structures necessary for studying the potential energy surfaces in Figure 2.3. The potential energy surfaces were studied with both DFT and NEP models and are shown in Section 4.4.

Path B is the linear interpolation between the relaxed polaron structure from DFT, and the predicted relaxed polaron structure from the NEPs. Their atomic positions are denoted $\mathbf{A}_{\text{pol}}^{\text{DFT}}$ and $\mathbf{A}_{\text{pol}}^{\text{NEP}}$. Introduce the vector $\boldsymbol{\xi}$ as a vector of 13 evenly spaced values in the interval $[-1, 1]$. The structures taken from path B can be written as

$$\mathbf{A}_j^{\text{Path B}} = \mathbf{A}_{\text{pol}}^{\text{NEP}} + \xi_j(\mathbf{A}_{\text{pol}}^{\text{DFT}} - \mathbf{A}_{\text{pol}}^{\text{NEP}}), \tag{3.3}$$

for $j = 2, ..., 12$. Note that for $j = 13$, $\mathbf{A}_{13}^{\text{Path B}}$ corresponds to the relaxed polaron structure from DFT. The geometry of $\mathbf{A}_1^{\text{Path B}}$ was not included either, making this a set of 11 structures that were added to the dataset for MgO. For LiF, the NEPs of the third generation already predicted the correct geometry. For this reason, Path

B was only included in the training set for MgO. All 11 structures correspond to polaron structures and were therefore labelled accordingly.

3.4.3 Training overview

An overview of the types of structures included in the final dataset is shown in Table 3.1. For each material, the number of bulk and polaron structures are listed. However, single-point DFT calculations with polaron structures sometimes did not converge, or did not fulfil the localization conditions in Section 3.4.1. The number of structures that could be used in the training set is shown without parentheses. The total number of structures on which DFT calculations were performed is written in parentheses.

Table 3.1: Overview of structures included in the final dataset. For polaron structures, the number of structures that could be used in the dataset is listed without parentheses, while the total number of structures on which DFT calculations were performed is listed in parentheses.

Structure type	MgO		LiF	
	Bulk	Polaron	Bulk	Polaron
Rattled	100	63 (100)	80	39 (40)
MD, G1	50	29 (50)	60	60 (60)
MD, G2	-	63 (80)	-	88 (94)
MD, G3	20	-	60	-
Path A	11	6 (11)	11	8 (11)
Path B	-	11 (11)	-	-
Relaxed structures	1	1	1	1
Total	182	173 (253)	212	196 (206)

For both materials, the first three generations were trained with default GPUMD parameters. Default parameters include using 100 000 generations and an energy coefficient $\lambda_e = 1$ in the loss function. In the fourth generation, the training was altered in three ways to improve performance at low temperatures. First, the relaxed pristine and polaron structures were repeated in the dataset. In MgO, the relaxed structures were repeated 30 times, while in LiF they were repeated 40 times. Second, when training NEPs on different splits of data, all path structures were forced to be in the training set instead of splitting the whole dataset randomly. Finally, the training was done in a two-step process. The NEPs were trained with 50 000 generations in each step. In the first step, the energy coefficient in the loss function was low, $\lambda_e = 0.1$, allowing the models to improve the performance on forces. In the second step, the energy coefficient was high, $\lambda_e = 10$, to improve performance on

energies. While the RMSE for the forces may increase slightly in the second step, the RMSE values for energies are significantly improved compared to the original one-step training.

3.5 Finite-temperature calculations

The final NEP models were used to calculate the RDF and CTL at finite temperatures. The RDF was calculated by performing MD simulations at temperatures up to 500 K. In MgO, the partial RDF for Mg-O* in the polaron system was compared to the partial RDF for the Mg-O bond in the pristine system. In LiF, the same was done for the F*-F* bond in the polaron system and the pristine F-F bonds. The increase in bond length with temperature was calculated by taking the weighted average of the RDF. Bond lengths at low temperatures were compared to the 0 K bond lengths in the corresponding DFT relaxed structures.

To calculate the temperature dependence of the CTL, the free energy difference $G^f[\text{bulk}] - G^f[\text{polaron}]$ and the VBM were computed at temperatures up to 500 K. The CTL relative to the VBM was obtained as the difference between the two quantities, as shown in equation 2.5, while the temperature dependence of the absolute CTL was studied by calculating the variation of $G^f[\text{bulk}] - G^f[\text{polaron}]$ with temperature. The free energy differences were computed via thermodynamic integration with the GPUMD fork TI-NEP. However, both convergence with the number of MD steps used during thermodynamic integration and with the number of atoms were tested before the final free energy calculations could be performed. Thermodynamic integration and convergence tests are described in Section 3.5.1. The temperature dependence of the VBM was calculated using a combination of MD simulations and single-point DFT calculations, as described in Section 3.5.2.

3.5.1 Thermodynamic integration

As described in Section 2.7, the free energy difference is obtained as

$$\Delta G = \frac{1}{2}\left[\overline{W^{\text{irr}}_{i\to f}} - \overline{W^{\text{irr}}_{f\to i}}\right], \tag{2.42}$$

where the overline denotes the average over an ensemble of realizations. A single realization in the forwards direction is given by equation 2.44, which is repeated below.

$$W^{\text{irr}}_{i\to f} = \sum_k \Delta\lambda_k \left(H_f - H_i\right)_{\mathbf{\Gamma}(k\Delta t)}. \tag{2.44}$$

The energy difference $(H_f - H_i)_{\mathbf{\Gamma}(k\Delta t)}$ is calculated by performing MD, propagated by the Hamiltonian in equation 2.36 for $\lambda \in [0, 1]$ in steps of $\Delta\lambda_k$, with a structure that is labelled according to the labelling method. The energy H_f is the predicted energy from a NEP, corresponding to the energy of the polaron system. The energy H_i corresponds to the predicted energy from an inverted NEP. A NEP can be inverted by modifying the parameters so that all labelled atoms are interpreted as unlabelled atoms. Given a labelled structure, the inverted NEP will predict

energies corresponding to a pristine structure of the same geometry. NEP inversion is performed with the Python package calorine.

The irreversible work in equation 2.44 is calculated with energy predictions from a normal NEP and an inverted NEP along the MD trajectory using the time step Δt =1 fs. When t_s time steps are taken, λ will vary from 0 to 1 in t_s steps. The total number of steps along the path will be $2 \cdot t_s$, since the calculations must be performed both forwards and backwards. As described in Section 2.7, equation 2.42 only holds true if the path is slow enough. In other words, convergence of the free energy difference with t_s, the number of MD steps taken along the path, must be tested before performing the final free energy calculations. The convergence tests were performed at 500 K, since this is the highest temperature at which free energy calculations are performed for both materials. Five runs are performed for each t_s to obtain the averages $\overline{W^{\mathrm{irr}}}$. A value for t_s was then chosen by making a suitable compromise between convergence and computation time, and was used in the remaining free energy calculations. In addition to the convergence with t_s, convergence with the number of atoms in the simulation was also tested in the same way at 500 K. Afterwards, the free energy difference was calculated at temperatures $T = 50, 100, ..., 500$ K with the chosen value for t_s and number of atoms. Five runs were performed for each temperature.

The free energy of a system is dependent on the volume, which increases with temperature due to thermal expansion. Since thermodynamic integration is performed in the NVT ensemble, it was necessary to use the correct volume for each simulated temperature. The average volume for a given temperature was calculated by taking the mean of 1000 MD snapshots at that temperature in an NPT ensemble. The VBM is also affected by volume. The same average volumes for each temperature were used when preparing structures for VBM calculations with MD, as described in the following section.

3.5.2 Temperature dependence of the VBM

To calculate the VBM at finite temperatures, single-point DFT calculations were performed on MD snapshots. The MD simulations were performed in the NVT ensemble using the calculated average volume for the temperature. Five snapshots were taken at each temperature $T = 50, 100, ..., 500$ K.

The values for the VBM in each DFT calculation are relative to the internal zero, which may differ between calculations. The obtained VBM values were calibrated by aligning the density of states (DOS) of the core level from each calculation to the same value. This was done by fitting a Gaussian function to the core level DOS in each calculation and aligning the average of the Gaussian function to the Gaussian average from the relaxation calculation. The average and standard deviation of the calibrated VBM at each temperature were obtained from the five calculations, and a polynomial was fit to the data.

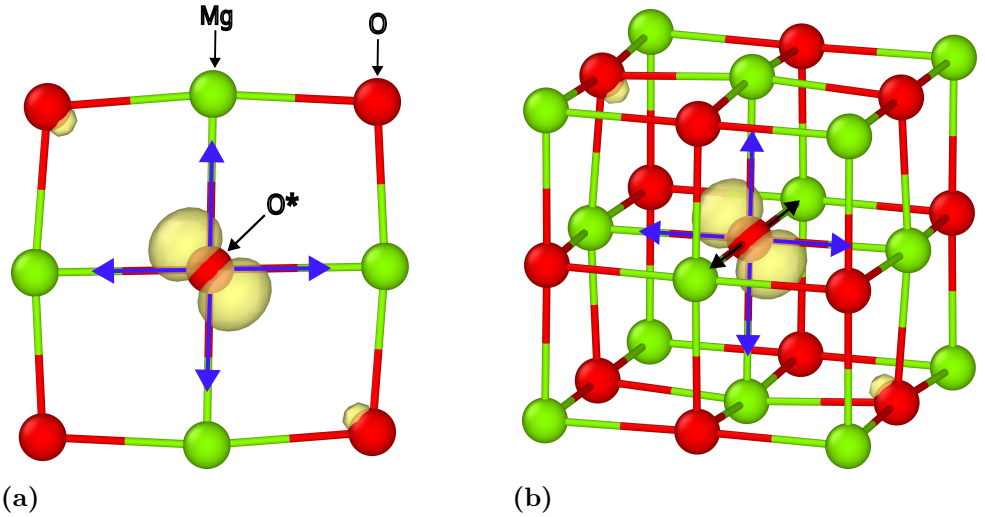
4

Results

The results obtained in the project are presented in this chapter. First, geometries and energetic properties at 0 K of pristine and polaron systems, obtained with DFT relaxation calculations, are presented. Also at 0 K, the transition between the V_c and V_k polaron states in LiF is shown. Subsequently, results from the NEP training are shown, as well as the NEP models' predictions for potential energy surfaces at 0 K and polaron distortions at finite temperatures. Finally, analysis of the CTL at finite temperatures is shown. This includes convergence tests and free energy calculations with thermodynamic integration, as well as the temperature dependence of the VBM.

4.1 Relaxation calculations

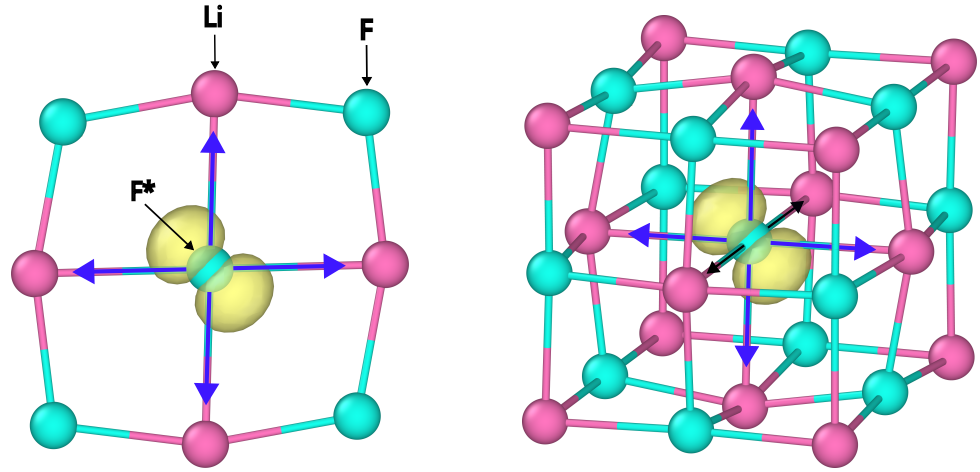
Relaxation calculations with PBE0(α) yielded the relaxed pristine systems for both MgO and LiF, as well as the three relaxed polaron structures. The V_c polaron distortions in MgO are shown in Figure 4.1. The distorted bond lengths are different depending on their direction relative to the orbital shown in yellow. The distortions



(a) (b)

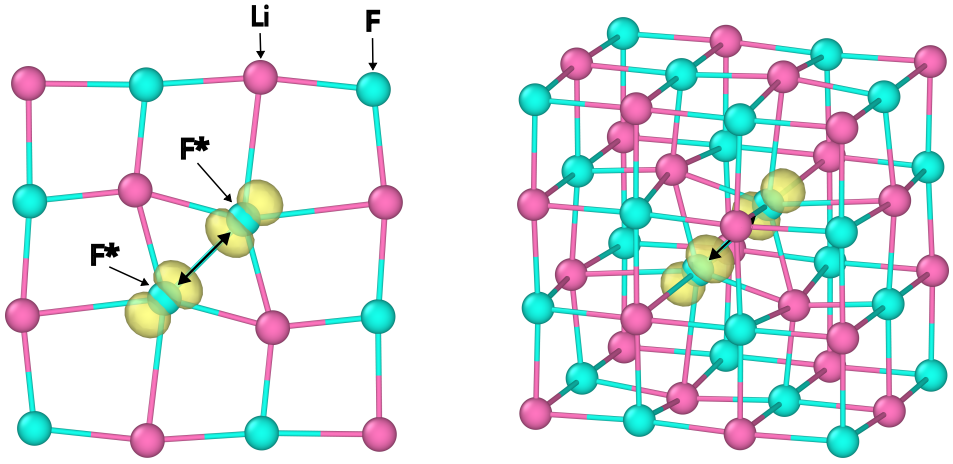
Figure 4.1: V_c polaron in MgO. Oxygen atoms are shown in red, while magnesium atoms are shown in green. The isosurface of the wavefunction of the polaron state is shown in yellow. Blue arrows correspond to a bond length of 2.20 Å while black arrows correspond to 2.16 Å. Rendered in Ovito [34].

shown in blue correspond to an Mg-O* bond length of 2.20 Å, while the distortions shown in black correspond to an Mg-O* bond length of 2.16 Å. Although the distortions are different, they both correspond to an elongation compared to the Mg-O bond length of 2.10 Å.



(a) (b)

Figure 4.2: V_c polaron in LiF. Lithium atoms are shown in pink, while fluoride atoms are shown in blue. The isosurface of the wavefunction of the polaron state is shown in yellow. Blue arrows correspond to a bond length of 2.22 Å while black arrows correspond to 2.18 Å. Rendered in Ovito [34].



(a) (b)

Figure 4.3: V_k polaron in LiF. Lithium atoms are shown in pink, while fluoride atoms are shown in blue. The isosurface of the wavefunction of the polaron state is shown in yellow. The F*-F* bond length is 1.93 Å. Rendered in Ovito [34].

The V_c polaron in LiF is shown in Figure 4.2. Similarly to the V_c polaron in MgO, the elongation of the bond lengths surrounding the labelled atom varies depending on their direction. Blue arrows correspond to a Li-F* bond length of 2.22 Å, while

black arrows correspond to 2.18 Å. The Li-F bond length in the pristine system is 1.98 Å. Finally, the V_k polaron in LiF is shown in Figure 4.3. In this case, the excess hole is shared between two fluoride atoms, causing the F*-F* bond length to be shortened to 1.93 Å. In the pristine system, the F-F bond length is 2.80 Å.

Relaxation calculations also yielded the necessary information for calculating the CTL of each polaron at 0 K, where the free energy in equation 2.5 simply reduces to the energy of the polaron and pristine systems. Table 4.1 shows the values of interest from the relaxation calculations. Energies for the polaron systems, E_{pol}, are relative to the energy of the pristine system. The CTL is calculated as the difference between E_{pol} and E_{VBM}. The band gap, E_g, for each material is also shown. The band gaps are used for the potential energy surfaces in Section 4.4.

Table 4.1: Values obtained from relaxation calculations. All values are given in eV. E_g is the band gap, and E_{pol} is the energy of the polaron system relative to the pristine system of the same material.

	E_g	E_{VBM}	E_{pol} (V_c)	E_{pol} (V_k)	CTL (V_c)	CTL (V_k)
MgO	8.18	-11.89	11.36	-	0.53	-
LiF	15.70	-16.60	14.30	14.03	2.30	2.58

4.2 String method calculations

The MEP between the V_c and V_k polaron states in LiF was obtained through string method calculations with DFT. Six images were used between the initial and final states, resulting in a chain of eight images with indices $i = 0, .., 7$. Image $i = 0$ corresponds to the V_c polaron state, while image $i = 7$ corresponds to the V_k polaron state. Figure 4.4 shows the potential energy along the MEP, where image $i = 2$ corresponds to the climbing image. The energy barrier is found to be 27 meV.

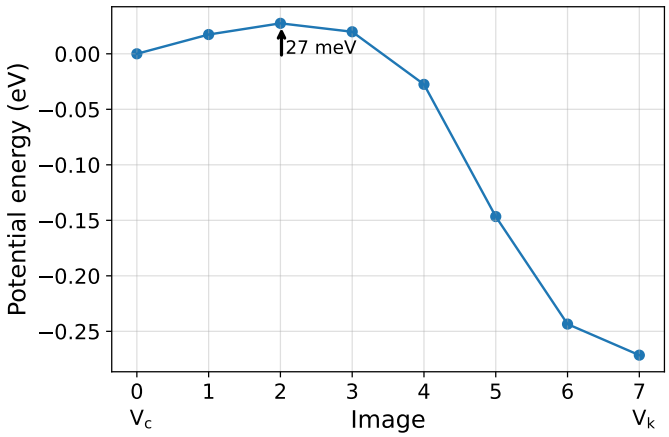


Figure 4.4: Potential energy along the MEP between V_c and V_k polaron states in LiF. The energy barrier is 27 meV.

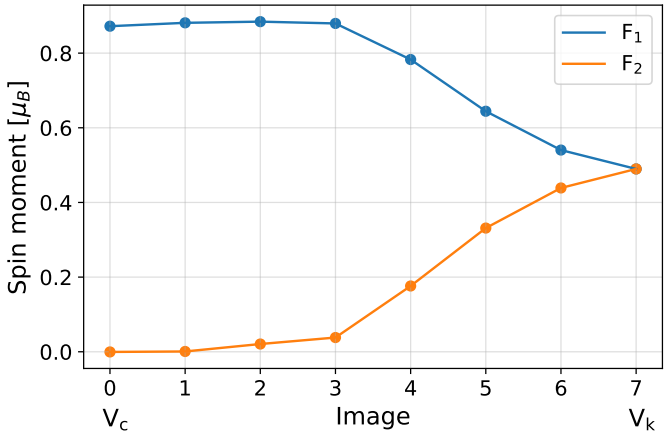


Figure 4.5: Spin moment on fluoride atoms F_1 and F_2 along the MEP. F_1 is the atom around which the excess hole is localized in the V_c polaron state, while the excess hole is shared between F_1 and F_2 in the V_k polaron state.

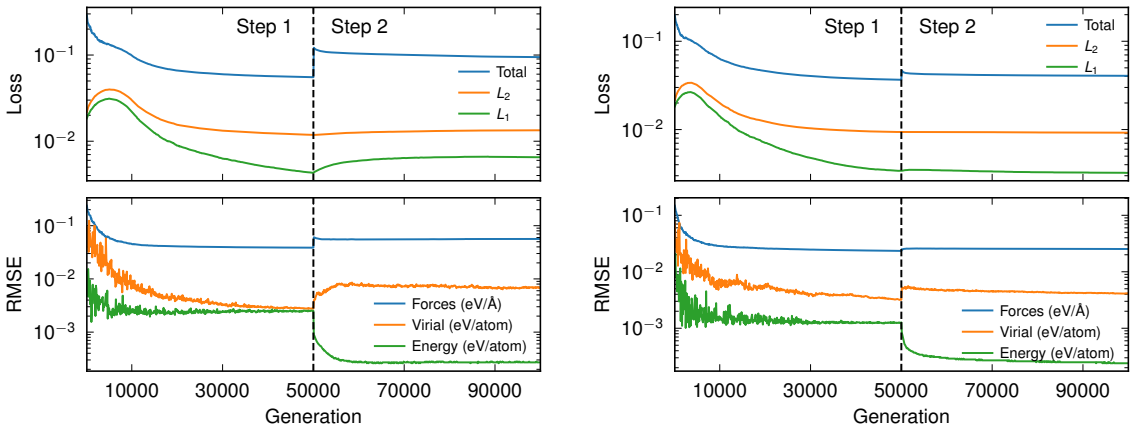
Figure 4.5 shows how the localization of the excess hole changes along the MEP in terms of the spin moment on the fluoride atoms F_1 and F_2. The fluoride atom F_1 corresponds to the atom where the excess hole is localized in the V_c polaron state. In the V_k polaron state, the excess hole is shared between the F_1 and F_2 atoms.

While Figure 4.4 shows that the V_c polaron is stable in LiF at zero temperature, the energy barrier is small. At higher temperatures, it is likely that the V_c polaron transitions to the V_k polaron, which cannot be modelled with the labelling method. For this reason, NEPs were only trained for V_c polarons in MgO and V_k polarons in LiF. In the remainder of this chapter, all LiF results are about the V_k polaron.

4.3 NEP training

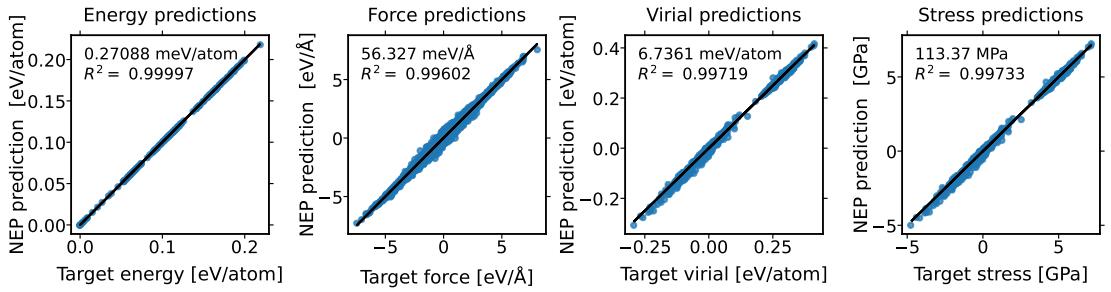
In the last generation of NEP training, full splits were trained for each material on the entire dataset. Figure 4.6 shows the convergence of the loss function and of RMSE values during training of the models. In the first half of the training, the energy coefficient in the loss function was low, $\lambda_e = 0.1$, while it was high in the second half of the training, $\lambda_e = 10$. The switch from the first step to the second step of training is shown with a black dashed line. While the RMSE for the forces increases slightly in the second step, the energy RMSE is significantly reduced.

Figure 4.7 shows parity plots for energy, forces, virial, and stress predictions from full split models for MgO and LiF, respectively. The NEP predictions are compared to the target data from DFT. R^2-scores and RMSE values are also shown for each quantity. R^2-scores are high for all quantities, indicating that the model parameters are well fitted to the training set. RMSE values are also low. However, the values shown in Figure 4.7 correspond to the models' performance on their own training set. For more realistic performance indicators, the models should be evaluated on a test set, which can be done for models that are trained on different splits of the dataset. Parity and convergence plots evaluated on training and test sets are shown in Appendix A.1 for one model from each material.

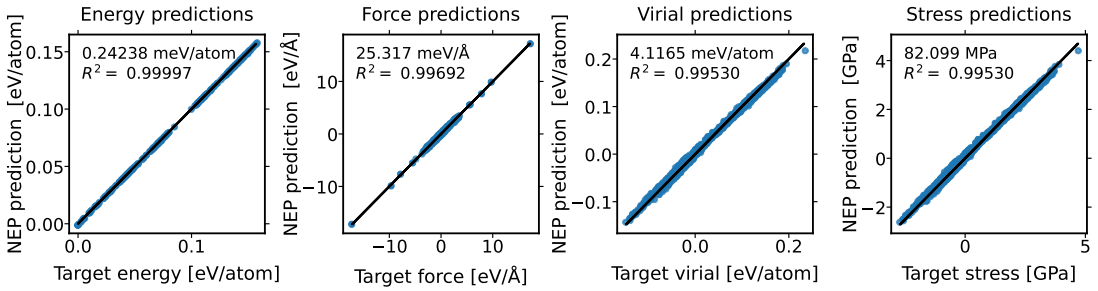


(a) (b)

Figure 4.6: **(a)** Convergence plots for MgO. **(b)** Convergence plots for LiF. Convergence of the L_1 and L_2 norms is shown together with the total loss function in the upper panels. Convergence of the RMSE values for energy, forces, and virials is shown in the lower panels. The switch from the first training step to the second is shown with a black, dashed line.



(a)

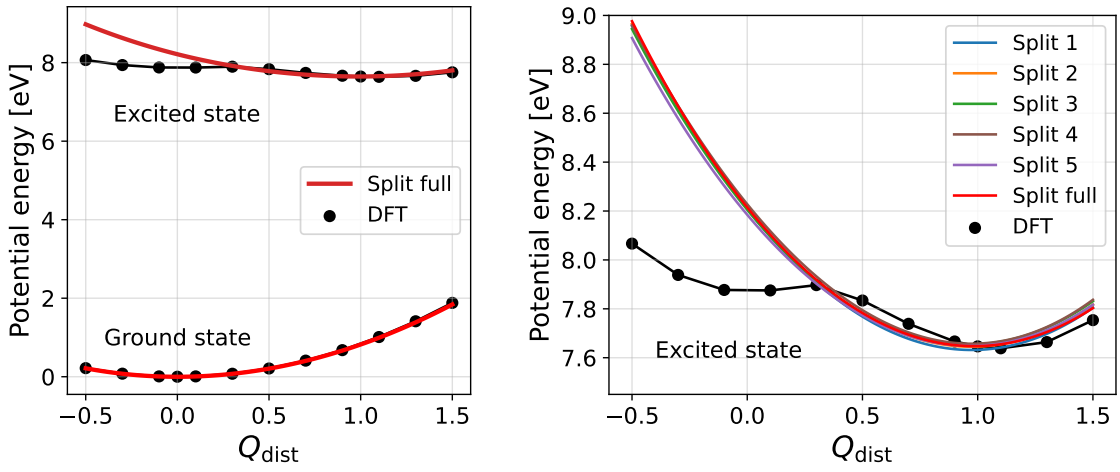


(b)

Figure 4.7: **(a)** Parity plots for MgO. **(b)** Parity plots for LiF. Target values for energy, force, virial and stress from DFT calculations are compared to NEP predictions. RMSE values and R^2-scores are shown for each quantity.

4.4 Potential energy surface

Potential energy surfaces (PES) for the ground and excited states at 0 K were predicted with both DFT calculations and NEP models. These correspond to the PESs that are schematically illustrated in Figure 2.3. The predicted PESs for MgO are shown in Figure 4.8. Energies are shown in reference to the minimum of the ground state. The energies of the excited state are shifted by the energy of the CBM, which can be obtained from table 4.1 as $E_g + E_{\text{VBM}}$. This corresponds to adding the "virtual" electron (see Section 2.2.3). Then, the energy of the excited state at $Q_{\text{dist}} = 0$ should theoretically correspond to E_g, and the difference between the energy of the excited state at $Q_{\text{dist}} = 0$ and at $Q_{\text{dist}} = 1$ should be the trapping energy, which is the same value as the CTL reported in Table 4.1. In Figure 4.8b, the energy of the excited state at $Q_{\text{dist}} = 0$ is significantly lower than E_g, and the trapping energy is smaller than the CTL of 0.53 eV in Table 4.1. This may be due to the sampling of reciprocal space in the DFT calculations. As described in section 3.2, only the Γ-point is sampled. Note that the structures used for the DFT calculations correspond to the structures in path A described in Section 3.4.2.



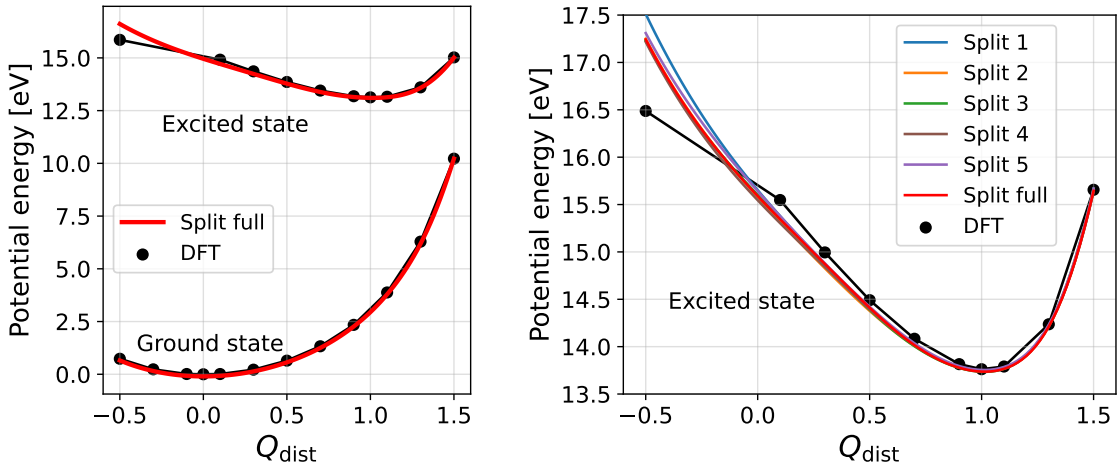
(a) **(b)**

Figure 4.8: **(a)**PES for ground and excited states in MgO predicted by the full split NEP (red) and from DFT calculations (black). **(b)** PES for the excited state in MgO predicted by all splits and by DFT calculations. The excited state is shifted by the energy of the CBM, E_{CBM}. Q_{dist} is the configuration coordinate, where $Q_{\text{dist}} = 0$ corresponds to the undistorted pristine system and $Q_{\text{dist}} = 1$ to the polaron distorted system as predicted from DFT relaxation calculations.

The NEP models are able to reproduce the results from DFT quite well. Since the models are not trained on structures with delocalized holes, it is natural that they are not able to predict the energies around $Q_{\text{dist}} = 0$ correctly for the excited state. Figure 4.8a shows the full split predictions for ground and excited states. Figure 4.8b shows predictions for the excited state from all splits, showing that they produce nearly identical results. We notice that they do not reproduce the exact

shape of the excited state PES, and that the minimum of the PES is slightly shifted to the left. A possible explanation for difficulties in learning the shape of the excited state PES in MgO is the fact that the PES is relatively flat. It is more difficult for the models to learn small differences in energy. In LiF, the excited state PES has significantly larger variations in energy, as shown in Figure 4.9.

Full split predictions for LiF are shown in Figure 4.9a for ground and excited states. The models are in good agreement with DFT results. Figure 4.9b shows predictions from all splits for the excited state. As in the case of MgO, the different splits produce nearly identical results. Three of the DFT calculations around $Q_{\text{dist}} = 0$ did not converge for the excited state. For this reason, the PES for the delocalized solution is missing. However, the models agree well with DFT results for the localized solution PES. As previously mentioned, the energy variations are larger in LiF than in MgO.



(a) **(b)**

Figure 4.9: **(a)**PES for ground and excited states in LiF predicted by the full split NEP (red) and from DFT calculations (black). **(b)** PES for the excited state in LiF predicted by all splits and by DFT calculations. The excited state is shifted by the energy of the CBM, E_{CBM}. Q_{dist} is the configuration coordinate, where $Q_{\text{dist}} = 0$ corresponds to the undistorted pristine system and $Q_{\text{dist}} = 1$ to the polaron distorted system as predicted from DFT relaxation calculations.

4.5 Temperature dependence of RDF

To study polaron distortions at finite temperatures, the RDF was calculated from MD simulations propagated using the full split NEP for each material. The size of the supercell was 13824 atoms, corresponding to a $12 \times 12 \times 12$ repetition of the 8-atom unit cell. While there are a great number of pristine bonds in the supercell, it is significantly more difficult to obtain abundant statistics for the polaron bonds, since there are only one or two labelled atoms in a polaron structure. For this reason,

a large number of MD steps, more precisely 2×10^5 MD steps, were taken at each temperature during RDF calculations. Figure 4.10 shows the normalised RDFs for MgO at temperatures between 100 K and 500 K. The partial RDF between Mg-O* in the polaron system is compared to the partial RDF between Mg-O in the pristine system. The RDFs from each system are normalised by the maximum value of their respective RDFs at 100 K.

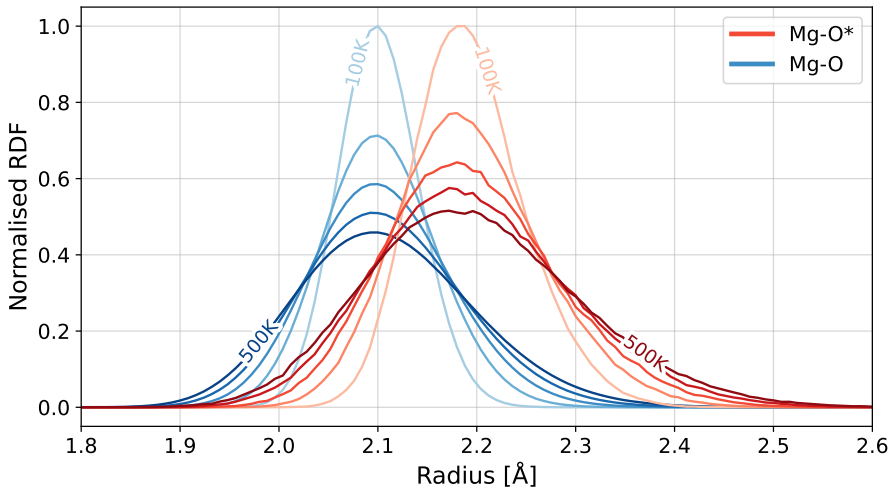
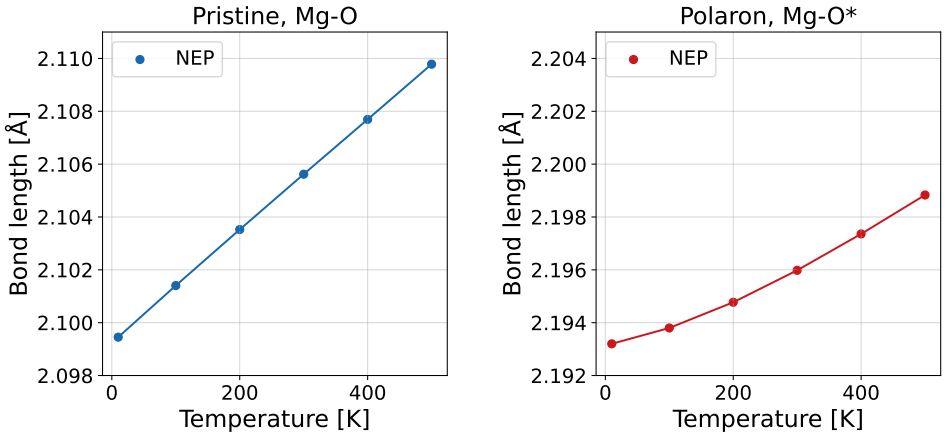


Figure 4.10: Partial RDF between Mg-O in the pristine system (blue) and between Mg-O* in the polaron system (red) for temperatures $T = 100, 200, 300, 400, 500$ K. The RDFs from each system are normalised by the maximum value of their respective RDFs at 100 K. The colours range from light to dark with increasing temperature.



(a) **(b)**

Figure 4.11: **(a)** Average Mg-O bond length in the pristine system for temperatures between 10 K and 500 K predicted by the full split NEP model. **(b)** Average Mg-O* bond length in the polaron system predicted by the full split NEP model.

Figure 4.10 clearly shows the polaron-induced bond elongation around the O*-atom, where the excess hole is localized. The figure also shows broadening of the curves with temperature. As the atoms gain more kinetic energy, they oscillate with higher amplitude around their equilibrium positions. Although the RDFs broaden with temperature, there are no large shifts in their equilibrium positions with temperature. This can be seen more clearly in Figure 4.11, which shows the pristine and polaron bond lengths, calculated by taking the weighted average of the corresponding RDFs. The model predicts an increase in the pristine Mg-O bond length of approximately 0.01 Å over 500 K. An increase in bond length is expected due to thermal expansion. The Mg-O* bond length is found to increase more slowly. Figure 4.12 shows the RDFs for LiF. Partial RDFs between F-F in the pristine system and between F*-F* in the polaron system are both shown. In this case, the Figure clearly shows that the bond length is significantly shortened by the polaron. Similarly to the case of MgO, broadening is seen with temperature. However, a clear shift can also be seen in the pristine F-F bond length. Figure 4.13 shows that an increase of approximately 0.05 Å over 500 K is predicted by the model. The increase in the F*-F* bond length is slower.

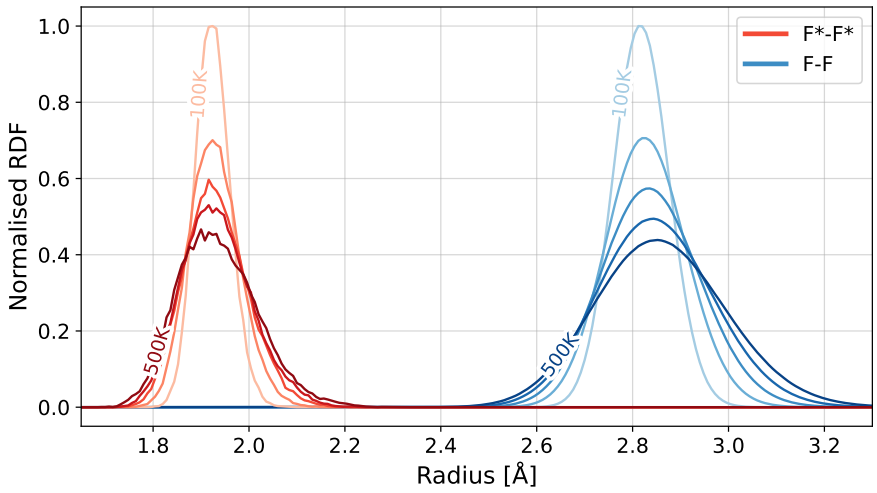
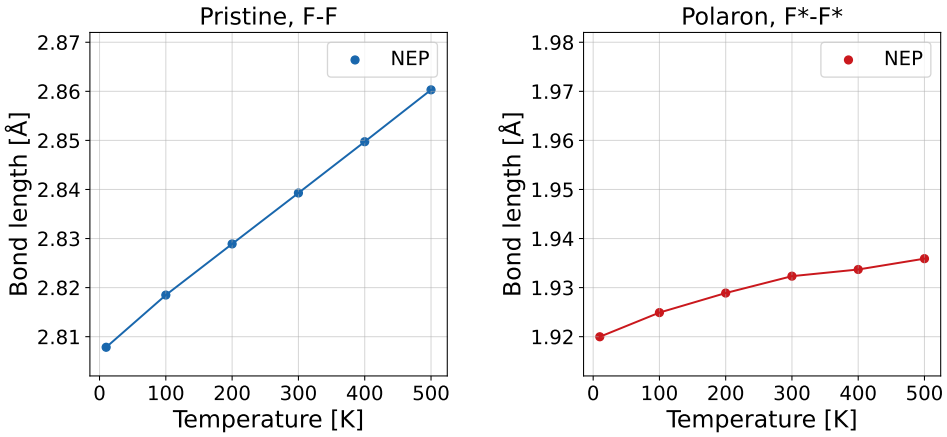


Figure 4.12: Partial RDF between F-F in the pristine system (blue) and between F*-F* in the polaron system (red) for temperatures $T = 100, 200, 300, 400, 500$ K. The RDFs from each system are normalised by the maximum value of their respective RDFs at 100 K. The colours range from light to dark with increasing temperature.

Values at low temperatures were compared to the bond lengths found in the DFT relaxations. The difference in bond length between NEP predictions at 10 K and DFT relaxed structures were less than 0.01 Å for polaron bonds and less than 0.005 Å for pristine bonds in both MgO and LiF. While the models are trained on a large number of pristine bonds, there are only one or two labelled atoms in each polaron structure, making it more difficult for the models to learn the polaron distortions compared to the pristine bond lengths.



(a) **(b)**

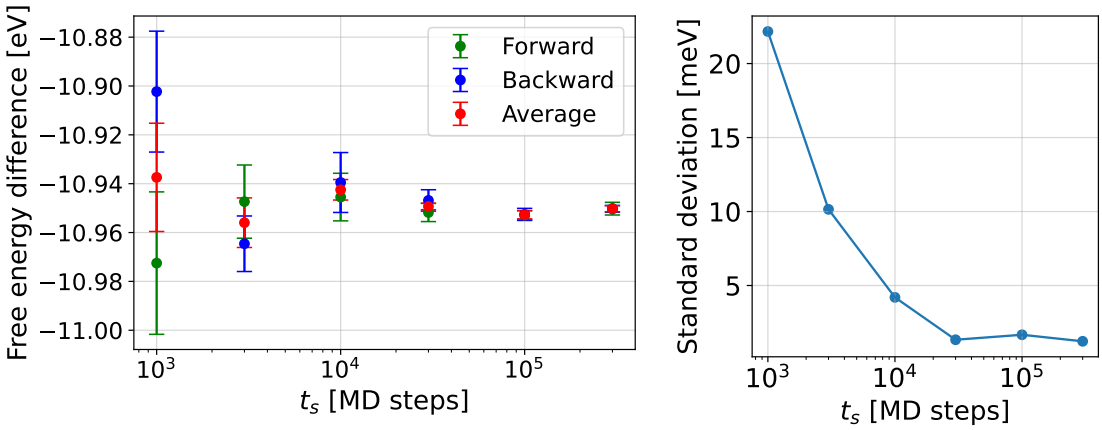
Figure 4.13: **(a)** Average F-F bond length in the pristine system for temperatures between 10 K and 500 K predicted by the full split NEP model. **(b)** Average F*-F* bond length in the polaron system for temperatures between 10 K and 500 K predicted by the full split NEP model.

4.6 Thermodynamic integration

The free energy difference between the pristine and polaron systems was calculated with thermodynamic integration. However, convergence with the number of MD steps taken along the path, t_s, and with the number of atoms must be tested before the free energy can be calculated. In this Section, convergence tests are first shown to motivate the choice of both t_s and the number of atoms. All convergence tests were performed at 500 K using the full split NEP models. The results from the convergence tests were then used to perform free energy calculations, which are shown in Section 4.6.2.

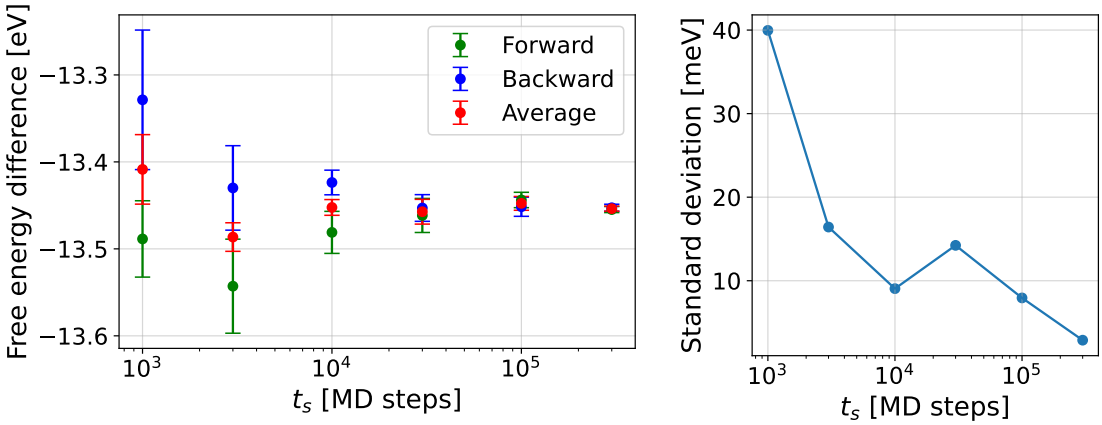
4.6.1 Convergence tests

The convergence of the free energy difference with the number of MD steps taken along the path is shown in Figures 4.14 and 4.15 for MgO and LiF, respectively. The free energy difference is calculated in both the forward and backward direction. The obtained values and their averages are shown in Figures 4.14a and 4.15a. The standard deviations for each quantity, calculated from five runs, are also shown by error bars. Convergence is seen by the decrease in variation between the forward and backward values, as well as by the decrease in standard deviation for each quantity. Figures 4.14b and 4.15b show the standard deviations of the average free energy differences in Figures 4.14a and 4.15a, respectively. By $t_s = 10^5$ MD steps, well-converged values are obtained in both materials with standard deviation for average values below 10 meV. Convergence with the number of atoms is shown in Figures 4.16a and 4.16b. Five runs were performed for each cell size and their standard deviation is shown by error bars. The number of atoms used in DFT calculations



(a) **(b)**

Figure 4.14: **(a)** Free energy difference between pristine and polaron systems in MgO at 500 K calculated in the forward (green) and backward (blue) directions, as well as their averages (red), as a function of t_s. Error bars show the standard deviation from five runs. **(b)** Standard deviation of average free energy difference.



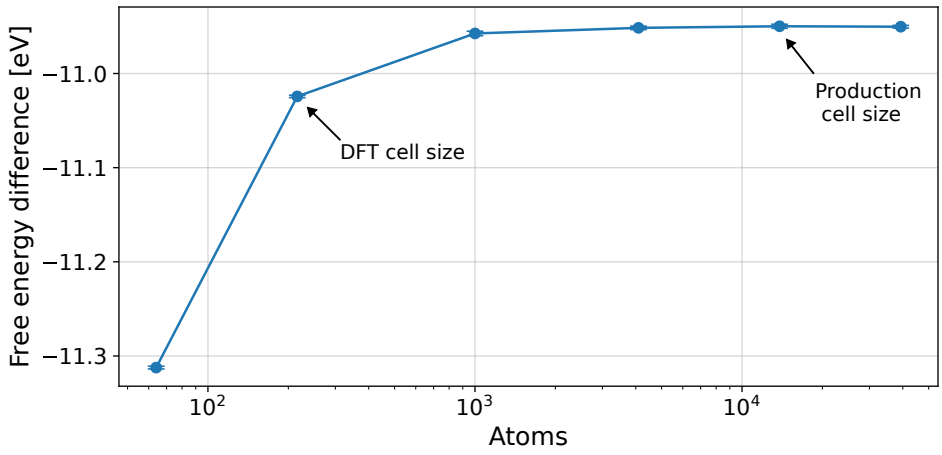
(a) **(b)**

Figure 4.15: **(a)** Free energy difference between pristine and polaron systems in LiF at 500 K calculated in the forward (green) and backward (blue) directions, as well as their averages (red), as a function of t_s. Error bars show the standard deviation from five runs. **(b)** Standard deviation of average free energy difference.

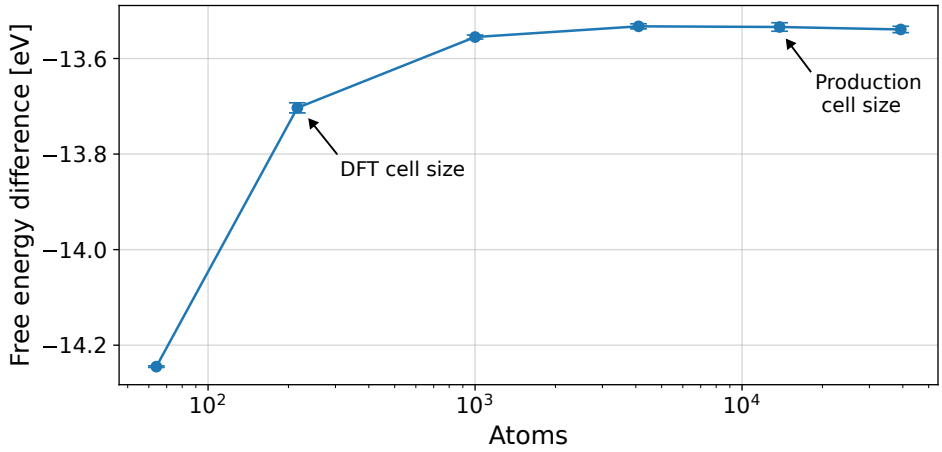
for production of training data is 216 atoms, while the number of atoms used for production calculations with NEP models is 13 824 atoms. The Figures show that converged free energy differences are obtained in both materials by using 13 824 atoms.

While the convergence tests are used to determine the necessary number of MD steps and atoms, they also motivate the use of MLIPs for these types of calculations. Performing 10^5 MD steps in a DFT simulation requires large computational resources, while it can be done with the NEP models in less than 15 minutes. The

size convergence also illustrates this point. DFT is generally only performed with up to a few hundred atoms, but Figures 4.16a and 4.16b show that converged values are not reached before cell sizes of 1000 atoms. Especially when combining the large cell sizes with a great number of MD steps, it becomes evident that the calculations are not feasible with DFT.



(a)



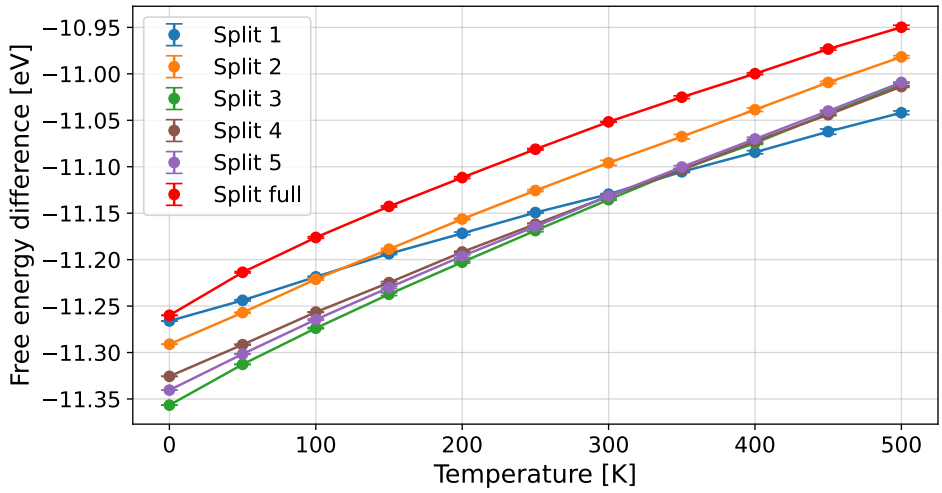
(b)

Figure 4.16: Convergence of free energy difference in **(a)** MgO and **(b)** LiF with cell size. Error bars show the standard deviation from five runs. The cell size used in DFT calculations for production of training data is 216 atoms, while production calculations with NEP models use a cell of 13 824 atoms.

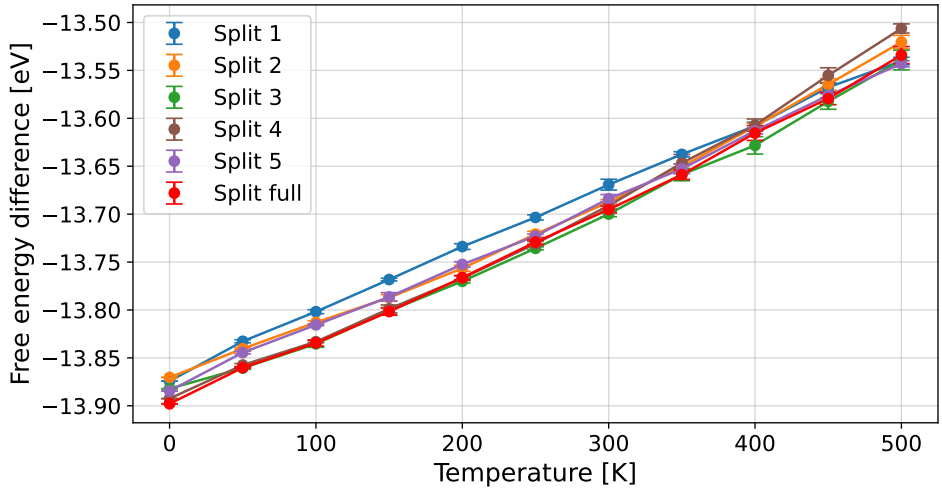
4.6.2 Free energy calculations

Free energy calculations were performed with all NEP splits using $t_s = 10^5$ and a cell of 13 824 atoms for temperatures between 50 K and 500 K. The results for MgO and LiF are shown in Figures 4.17a and 4.17b, respectively. Five runs were performed for each temperature and model, the standard deviation of which is shown by error

bars. The value at 0 K for each model was obtained by relaxing pristine and polaron structures of 13 824 atoms with the model, and predicting the energy difference between the two structures with the same model.



(a)



(b)

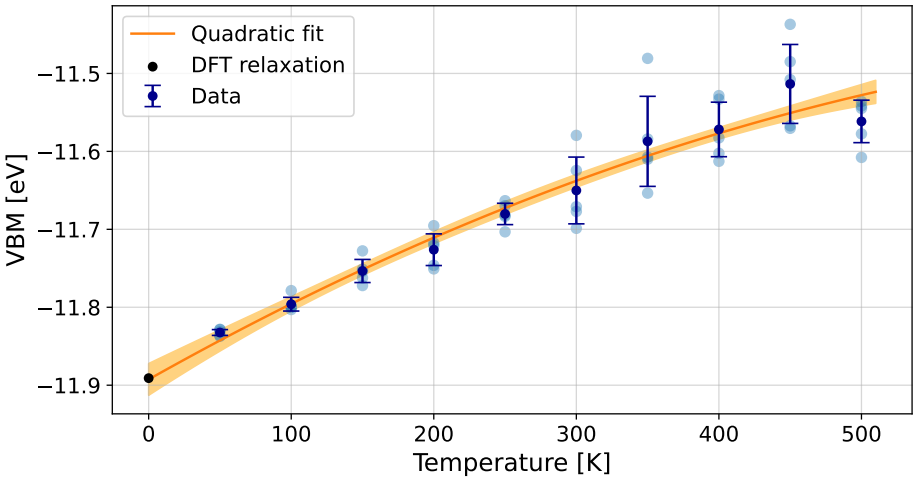
Figure 4.17: Free energy difference in **(a)** MgO and **(b)** LiF between pristine and polaron systems predicted by all NEP splits. Error bars show the standard deviation from five runs. Values at 0 K show the energy difference from relaxation calculations performed with the corresponding NEP split.

In the case of MgO, the results from different splits are shifted by values up to 0.1 eV, but show similar temperature dependencies. All splits, apart from split 1, have nearly identical slopes. Although no problems could be found in the training of split 1 in terms of parity and convergence plots (see Appendix A.1), it is still possible that the training process was not optimal. To obtain more consistent results, one could repeat the training. If the problem remains, it may indicate that the specific split of data causes the model to predict a different slope. In this case, more training

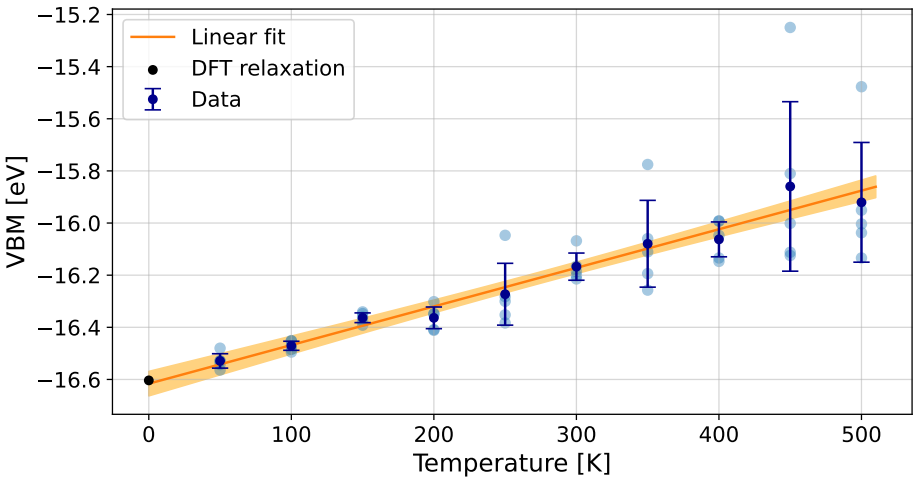
data should be obtained until differences between the splits are reduced. In the case of LiF, all NEP splits are in good agreement. Similar slopes are predicted by all models, and the curves are only shifted by values below 0.03 eV.

4.7 VBM calculations

The temperature dependence of the VBM was calculated by performing single-point DFT calculations on structures that were obtained from MD simulations at temperatures between 50 K and 500 K. Five structures were taken at each temperature and



(a)



(b)

Figure 4.18: VBM in **(a)** MgO and **(b)** LiF, with light blue markers showing VBM values from DFT calculations at each temperature. Their average and standard deviation are shown in dark blue, and the linear fit in orange. The shaded area shows the standard deviation of the fit. The black marker shows the VBM obtained from DFT relaxation.

a polynomial was fit to the obtained values. Only values at finite temperatures were used for the polynomial fit. VBM values at 0 K from DFT relaxation calculations were not used for the fit, since the NEP models can predict slightly different volumes compared to DFT at low temperatures, which can in turn affect the VBM. When performing the fit, the lowest possible order was chosen for the polynomial that still allowed the function to fit the data reasonably. The fit gives both the optimal parameters and the covariance matrix for the parameters. The covariance matrix was used to calculate standard deviations for the polynomial fits.

Figure 4.18a shows the obtained VBM values for MgO, as well as a quadratic fit to the data. For LiF, a linear fit was used. Results for LiF are shown in Figure 4.18b. In both Figures, VBM values from all five structures are shown at each temperature in light blue, while their averages and standard deviations are shown in dark blue. The polynomial fit is shown in orange, together with a shaded area indicating the standard deviation of the fit. In both MgO and LiF, the polynomial fits overlap with the error bars for all data points. They also closely reproduce the VBM at 0 K from DFT relaxations.

4.8 Temperature dependence of CTL

Combining free energy and VBM calculations according to equation 2.5, the CTL is obtained at finite temperatures. Since the temperature dependence of different splits was found to be similar in Section 4.6.2, CTL results are only shown for the full split NEP models. Figure 4.19 shows the free energy differences between pristine and polaron structures in MgO, as well as the VBM fit. The energy scale is shifted such that the VBM at 0 K is set to zero. The free energy difference increases by approximately 310 meV over 500 K, demonstrating a notable temperature dependence in the absolute level of the CTL. However, the VBM moves at a similar rate, 360 meV over 500 K, causing the CTL relative to the VBM to be almost constant. This is seen in 4.20, which shows the CTL relative to the VBM. The blue shaded area shows the standard deviation from the VBM fit, while the smaller, red shaded area shows the standard deviation from the five runs of free energy calculations. The relative CTL moves by approximately 55 meV over 500 K.

The corresponding results are shown for LiF in Figures 4.21 and 4.22. The free energy difference increases by approximately 360 meV over 500 K, showing that the absolute CTL is sensitive to temperature and increases at a similar rate as in MgO. However, the VBM in LiF has a strong temperature dependence and increases by 740 meV over 500 K, causing the CTL relative to the VBM to decrease by approximately 380 meV over 500 K. While only the relative CTL was found to have strong temperature dependence in MgO, both the relative and absolute CTL are sensitive to temperature in LiF.

In both materials, notable shifts are predicted in the VBM and the free energy difference over the studied temperature range. While DFT calculations can give accurate predictions for these values at 0 K, Figures 4.19 and 4.21 indicate that in general, those same results cannot be assumed to be accurate at realistic temperatures, showing the importance of taking finite-temperature effects into consideration.

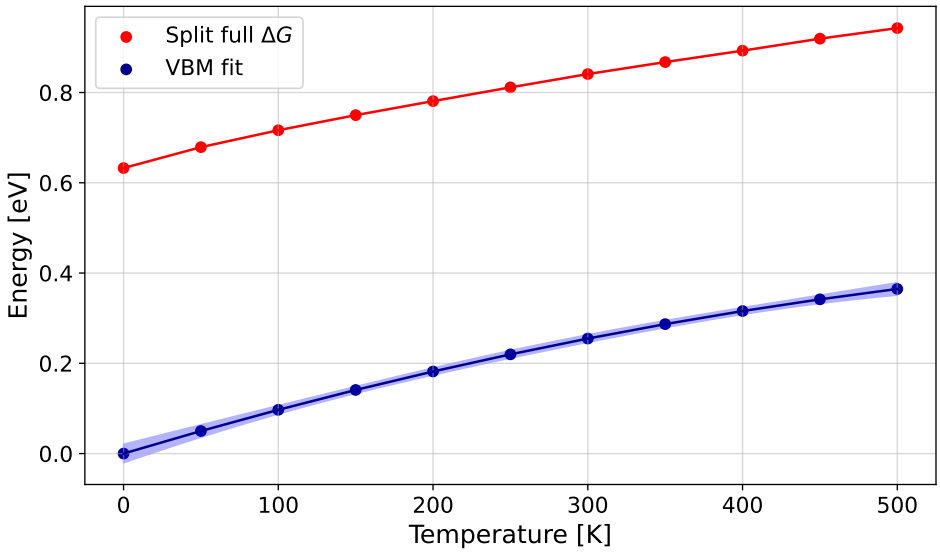


Figure 4.19: Free energy difference between pristine and polaron systems in MgO (red) and VBM polynomial fit (blue). The blue shaded area shows the standard deviation of the VBM polynomial fit.

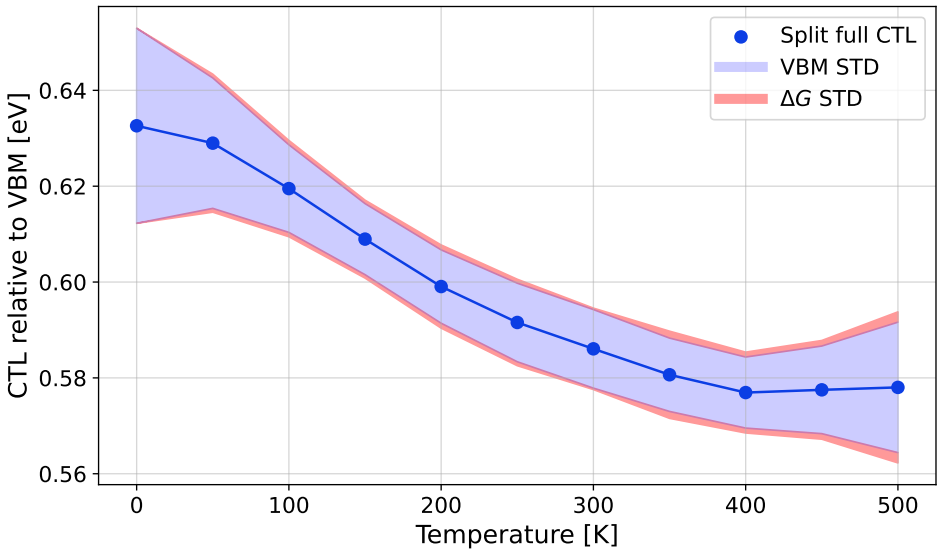


Figure 4.20: Temperature dependence of the CTL relative to the VBM in MgO. The blue shaded area shows the standard deviation of the VBM polynomial fit, while the red shaded area shows the standard deviation from the five runs of free energy calculations.

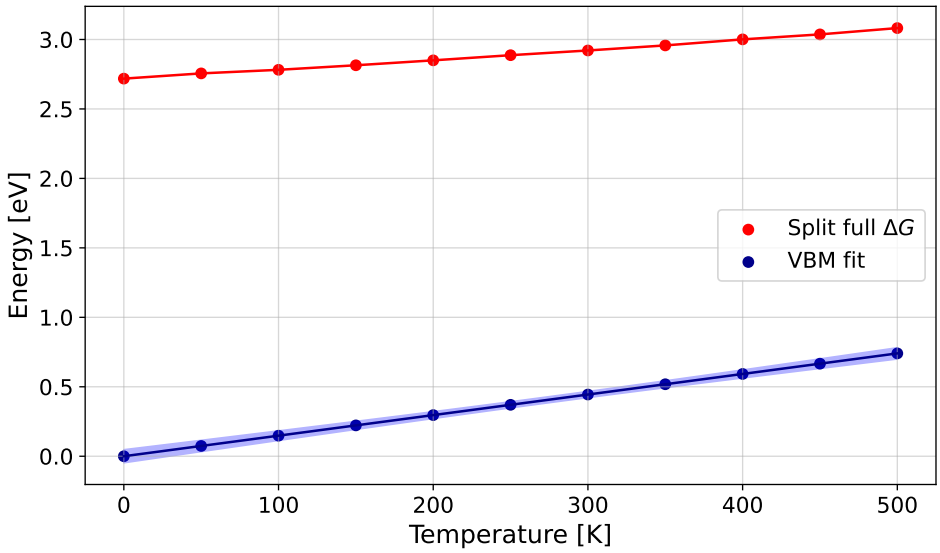


Figure 4.21: Free energy difference between pristine and polaron systems in LiF (red) and VBM polynomial fit (blue). The blue shaded area shows the standard deviation of the VBM polynomial fit.

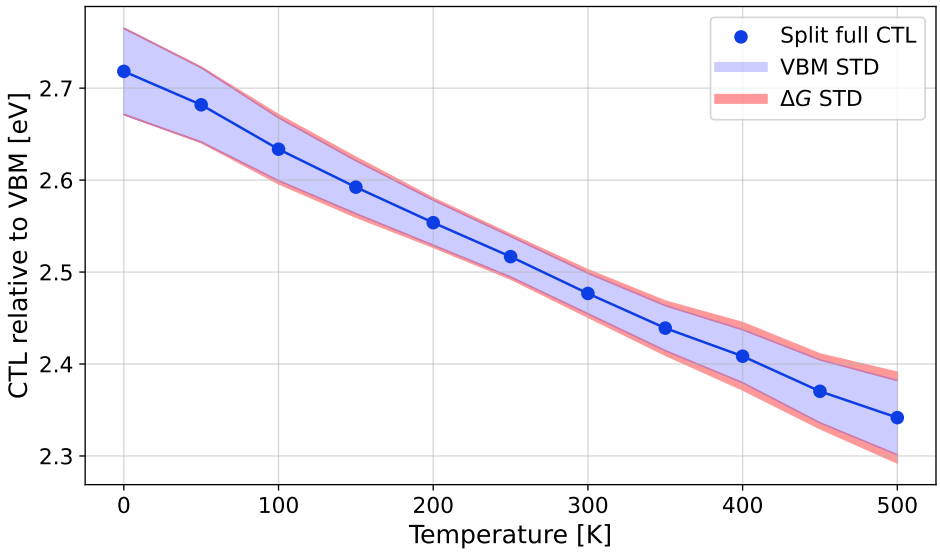


Figure 4.22: Temperature dependence of the CTL relative to the VBM in LiF. The blue shaded area shows the standard deviation of the VBM polynomial fit, while the red shaded area shows the standard deviation from the five runs of free energy calculations.

5
Conclusion

The main objective of this thesis was to study small polarons with NEP models in order to analyse polaron properties at finite temperatures. The NEP models were trained in cycles of active learning using DFT calculations with PBE0(α) and used to predict the temperature dependence of bond lengths, VBM and CTL. In both materials, the polaron bond lengths were found to increase more slowly than the pristine bond lengths and temperature was found to significantly influence the VBM and the free energy difference between pristine and polaron structures. In MgO, the absolute position of the CTL was found to have notable temperature dependence. The free energy difference increased by approximately 310 meV over 500 K. However, the CTL relative to the VBM only shifts by 55 meV due to the similar rate of increase of the VBM. In LiF, both the absolute and relative positions of the CTL are sensitive to temperature. The free energy difference increases at a similar rate as in MgO. The CTL relative to the VBM decreases by approximately 380 meV over 500 K due to the strong temperature dependence of the VBM. The results in both materials clearly show that relative and absolute CTL positions at 0 K cannot be assumed to be accurate at higher temperatures, demonstrating the importance of taking finite-temperature effects into consideration.

Calculations based solely on DFT cannot feasibly be used for finite-temperature simulations, especially with the computationally costly hybrid functionals that are important for accurate description of the polarons. At variance, the computational efficiency of the NEP models allowed us to perform the lengthy MD simulations that were necessary to obtain converged values and abundant statistics. The convergence tests in Section 4.6.1 showed that both large cells with more than 1000 atoms and a thermodynamic integration path of 10^5 MD steps were necessary for well-converged free energy differences. For the RDF calculations in section 4.5, 2×10^5 MD steps were taken at each temperature to obtain abundant statistics for the polaron bond length. Both examples demonstrate the usefulness of the high-speed NEP calculations, which enable both the use of large cells and a great number of MD steps.

The NEP models were validated through parity plots and RMSE/R^2-scores, as well as through robustness between NEP models that had been trained on different splits of the data. While the full split NEP model did not have a test set, the five NEP models that did performed well on both training and test sets, indicating accurate models that are not overfitted to the training data. Agreement between models trained on different splits of the data can be seen in the free energy calculations in Section 4.6.2, where robust results for the slope of the free energy with temperature were generally obtained between the different splits. As previously mentioned, the predictions from one of the splits in MgO deviate in slope from the others. However,

this may be due to an error in the training of the model. Even though the predictions for the slope of the free energy are generally in good agreement, the curves are shifted relative to each other. The separation between the curves is especially notable in MgO. However, the aim of the project is to predict the temperature dependence of the CTL. With a reliable temperature dependence, the values could be calibrated with, for example, experimental results, yielding accurate CTL values for all temperatures. Finally, the models were also validated through comparison with DFT results. The comparison between PESs predicted by NEP and DFT in Section 4.4 shows good agreement and the RDFs at low temperatures in Section 4.5 reproduce DFT bond lengths quite well.

While the V_c polaron in MgO and the V_k polaron in LiF could be studied at finite temperatures through the use of NEP models and the labelling method, the project also exemplifies a limitation of the labelling method. Due to the low energy barrier between the V_c and V_k polaron states in LiF, the V_c polaron could not be studied with NEP models. At higher temperatures, the V_c polaron would likely transition to a V_k polaron and the transition cannot be modelled with the labelling method since the labelling of an atom cannot change in the middle of an MD simulation. While one may argue that if it is not stable at the realistic operating temperatures for the relevant applications, it does not need to be modelled, the fact that the labelling of an atom cannot change also limits the possibility of modelling other effects, such as polaron hopping. Following the work in this thesis, it would be of interest to investigate a more dynamic way of modelling the polarons that does not constrain the localization of the excess charge to remain unchanged. An extension of the NEP model called qNEP has been developed [35], in which explicit, environment-dependent charges for each ion are predicted by the neural network in addition to the site energies. Extending these methods to the study of polarons could be useful for developing a more dynamic way of modelling of the localization of the excess charge.

On a broader note, the outlook of this project also includes combining polarons at finite temperatures with more advanced systems, such as more complex materials, interfaces and surfaces or studying their interactions with other defects or polarons. The ability to model polarons in realistic conditions would enable more accurate predictions of material properties, facilitating informed material design and the development of sustainable energy technologies.

Bibliography

[1] Volker L Deringer, Miguel A Caro, and Gábor Csányi. "Machine learning interatomic potentials as emerging tools for materials science". In: *Advanced Materials* 31.46 (2019), p. 1902765.

[2] Christopher Linderälv et al. "Optical line shapes of color centers in solids from classical autocorrelation functions". In: *npj Computational Materials* 11.1 (2025), p. 101.

[3] Dibyajyoti Ghosh et al. "Polarons in halide perovskites: a perspective". In: *The journal of physical chemistry letters* 11.9 (2020), pp. 3271–3286.

[4] Philip Hofmann. *Solid state physics: an introduction*. John Wiley & Sons, 2022.

[5] Cesare Franchini et al. "Polarons in materials". In: *Nature Reviews Materials* 6.7 (2021), pp. 560–586.

[6] Stefano Falletta and Alfredo Pasquarello. "Polarons free from many-body self-interaction in density functional theory". In: *Physical Review B* 106.12 (2022), p. 125119.

[7] R.D Popp and R.B Murray. "Diffusion of the Vk-polaron in alkali halides: Experiments in NaI and RbI". In: *Journal of Physics and Chemistry of Solids* 33.3 (1972), pp. 601–610. ISSN: 0022-3697. DOI: https://doi.org/10.1016/0022-3697(72)90069-8. URL: https://www.sciencedirect.com/science/article/pii/0022369772900698.

[8] Alexander S Tygesen, Jin Hyun Chang, and Juan María García-Lastra. "Dependence of polaron migration barriers on the fraction of Fock exchange in hybrid functionals: A systematic study of V k centers in alkali halides". In: *Physical Review B* 108.4 (2023), p. 045120.

[9] Zhenbang Dai et al. "Comparison between first-principles supercell calculations of polarons and the ab initio polaron equations". In: *arXiv preprint arXiv:2511.01764* (2025).

[10] Jon Lafuente-Bartolome et al. "Ab initio self-consistent many-body theory of polarons at all couplings". In: *Physical Review B* 106.7 (2022), p. 075119.

[11] Toraya Fernández Ruíz et al. "Dynamics of the electron-hole separation of the fundamental exciton in LiF". In: *Physical Review B* 112.5 (2025), p. 054308.

[12] Dahvyd Wing et al. "Role of long-range exact exchange in polaron charge transition levels: The case of MgO". In: *Phys. Rev. Mater.* 4 (8 Aug. 2020), p. 083808. DOI: 10.1103/PhysRevMaterials.4.083808. URL: https://link.aps.org/doi/10.1103/PhysRevMaterials.4.083808.

[13] Pierre Hohenberg and Walter Kohn. “Inhomogeneous electron gas”. In: *Physical review* 136.3B (1964), B864.

[14] NM Harrison. “An introduction to density functional theory”. In: *Nato Science Series Sub Series III Computer and Systems Sciences* 187 (2003), pp. 45–70.

[15] Walter Kohn and Lu Jeu Sham. “Self-consistent equations including exchange and correlation effects”. In: *Physical review* 140.4A (1965), A1133.

[16] John P Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized gradient approximation made simple”. In: *Physical review letters* 77.18 (1996), p. 3865.

[17] Roberto Car. “Introduction to Density-Functional Theory and ab-Initio Molecular Dynamics”. In: *Quantitative Structure-Activity Relationships* 21.2 (2002), pp. 97–104.

[18] Paula Mori-Sánchez, Aron J Cohen, and Weitao Yang. “Many-electron self-interaction error in approximate density functionals”. In: *The Journal of chemical physics* 125.20 (2006).

[19] Stefano Falletta and Alfredo Pasquarello. “Many-body self-interaction and polarons”. In: *Physical Review Letters* 129.12 (2022), p. 126401.

[20] Jing Yang, Stefano Falletta, and Alfredo Pasquarello. “One-shot approach for enforcing piecewise linearity on hybrid functionals: Application to band gap predictions”. In: *The Journal of Physical Chemistry Letters* 13.13 (2022), pp. 3066–3071.

[21] Weiqing Ren, Eric Vanden-Eijnden, et al. “Simplified and improved string method for computing the minimum energy paths in barrier-crossing events”. In: *The Journal of chemical physics* 126.16 (2007).

[22] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. “A climbing image nudged elastic band method for finding saddle points and minimum energy paths”. In: *The Journal of chemical physics* 113.22 (2000), pp. 9901–9904.

[23] Zheyong Fan et al. “GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations”. In: *The Journal of Chemical Physics* 157.11 (2022).

[24] Ke Xu et al. “GPUMD 4.0: a high-performance molecular dynamics package for versatile materials simulations with machine-learned potentials”. In: *Materials Genome Engineering Advances* 3.3 (2025), e70028.

[25] Aneesur Rahman and Frank H Stillinger. “Molecular dynamics study of liquid water”. In: *The Journal of Chemical Physics* 55.7 (1971), pp. 3336–3359.

[26] Josef Honerkamp. *Statistical physics: An advanced approach with applications*. Springer Science & Business Media, 2012. DOI: 10.1007/978-3-642-28684-1.

[27] Rodrigo Freitas, Mark Asta, and Maurice De Koning. “Nonequilibrium free-energy calculation of solids using LAMMPS”. In: *Computational Materials Science* 112 (2016), pp. 333–341.

[28] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[29] Christopher Jarzynski. “Nonequilibrium equality for free energy differences”. In: *Physical Review Letters* 78.14 (1997), p. 2690.

[30] Volker Blum et al. “Ab initio molecular simulations with numeric atom-centered orbitals”. In: *Computer Physics Communications* 180.11 (2009), pp. 2175–2196.

[31] Xinguo Ren et al. “Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions”. In: *New Journal of Physics* 14.5 (2012), p. 053020.

[32] Christoph Schran, Krystof Brezina, and Ondrej Marsalek. “Committee neural network potentials control generalization errors and enable active learning”. In: *The Journal of Chemical Physics* 153.10 (2020).

[33] Jesús Carrete et al. “Deep ensembles vs committees for uncertainty estimation in neural-network force fields: Comparison and application to active learning”. In: *The Journal of Chemical Physics* 158.20 (2023).

[34] Alexander Stukowski. “Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool”. In: *Modelling and simulation in materials science and engineering* 18.1 (2010), p. 015012.

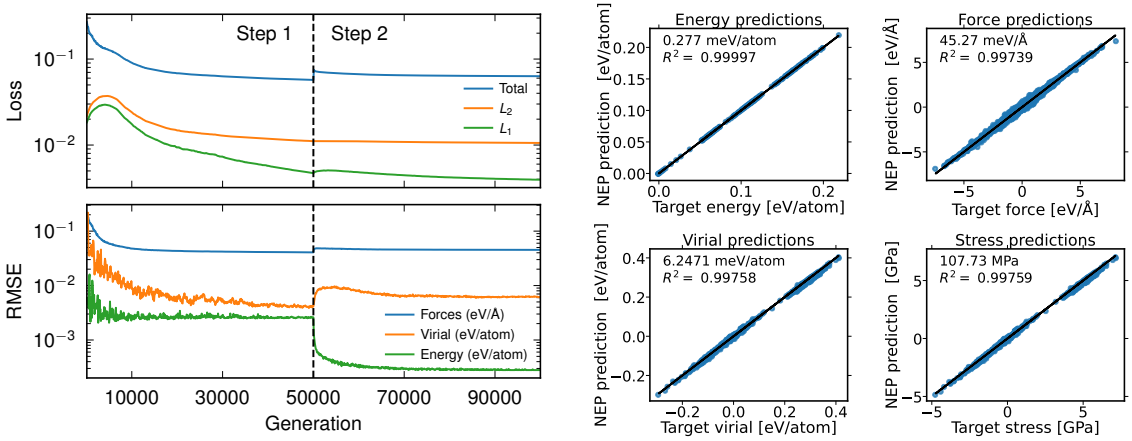
[35] Zheyong Fan et al. “qNEP: A highly efficient neuroevolution potential with dynamic charges for large-scale atomistic simulations”. In: *Journal of Chemical Theory and Computation* (2026).

A

Appendix

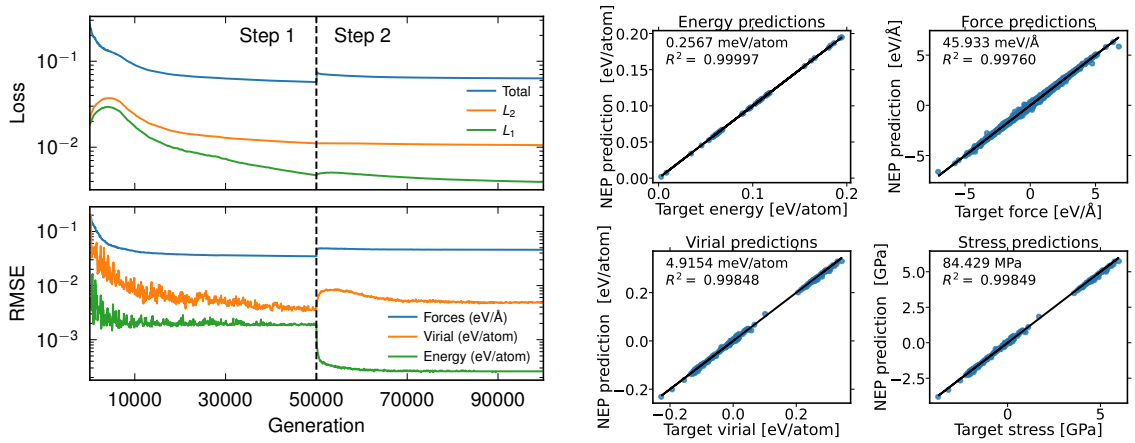
A.1 NEP training

Convergence of the loss function and parity plots are shown below for one split from each material. The performance on both the training and the test sets are shown.



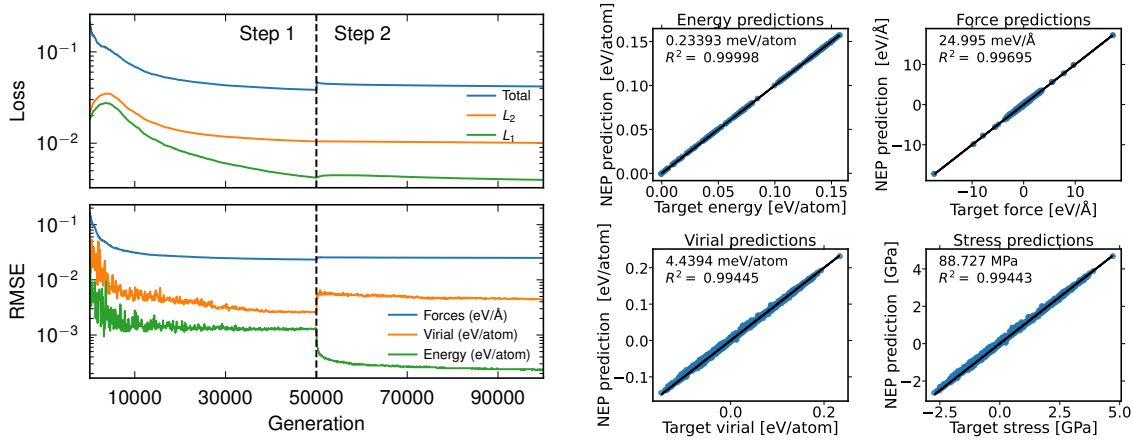
(a) **(b)**

Figure A.1: Split 1 in MgO. **(a)** Convergence of the loss function. RMSE values are evaluated on the training set. **(b)** Parity plots for the training set. RMSE and R^2-scores are shown for each quantity.



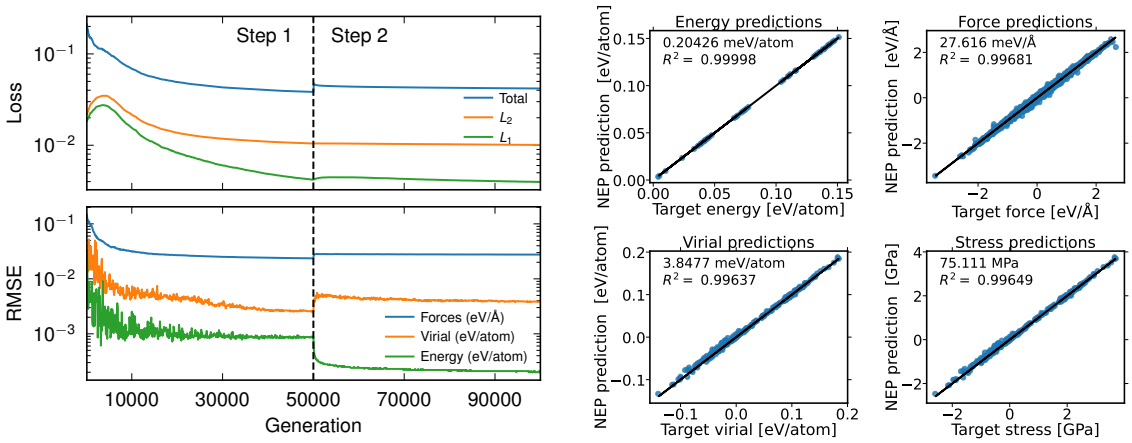
(a) **(b)**

Figure A.2: Split 1 in MgO. **(a)** Convergence of the loss function. RMSE values are evaluated on the test set. **(b)** Parity plots for the test set. RMSE and R^2-scores are shown for each quantity.



(a) **(b)**

Figure A.3: Split 1 in LiF. **(a)** Convergence of the loss function. RMSE values are evaluated on the training set. **(b)** Parity plots for the training set. RMSE and R^2-scores are shown for each quantity.



(a) **(b)**

Figure A.4: Split 1 in LiF. **(a)** Convergence of the loss function. RMSE values are evaluated on the test set. **(b)** Parity plots for the test set. RMSE and R^2-scores are shown for each quantity.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY