UNIVERSIDAD DE INVESTIGACION DE TECNOLOGIA EXPERIMENTAL YACHAY

Escuela de Ciencias Físicas y Nanotecnología

TÍTULO: First-Principles and Machine Learning Investigations into the Atomic and Mechanical Properties of Cement Hydrates.

Trabajo de integración curricular presentado como requisito para la obtención del título de Físico

Autor:

Balarezo J. Gabriel

Tutor:

Ph.D. - Pinto Henry

Urcuquí, Octubre 2025

AUTORÍA

Yo, **Balarezo Balarezo Juan Gabriel**, con cédula de identidad 0106019219, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Octubre - 2025.

Juan Gabriel Balarezo Balarezo
C.I: 0106019219

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Balarezo Balarezo Juan Gabriel**, con cédula de identidad 0106019219, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, Octubre - 2025.

Juan Gabriel Balarezo Balarezo
C.I: 0106019219

To the younger self who dared to dream,
and to the present self who refuses to falter.
For every sleepless night,
and every quiet triumph along the way.

Balarezo J. Gabriel

Resumen

El concreto es el segundo material más utilizado en el mundo después del agua, con más de 35 mil millones de toneladas producidas anualmente. Sin embargo, comprender las propiedades atómicas y mecánicas de su componente principal, el silicato cálcico hidratado (C–S–H), la fase aglutinante compleja del concreto, sigue siendo un desafío considerable debido a su complejidad estructural y naturaleza desordenada. Este trabajo tiene como objetivo desarrollar y validar un campo de fuerza basado en aprendizaje automático (MLFF) capaz de reproducir con precisión el comportamiento estructural y mecánico del C–S–H. Para ello, se emplearon cálculos de teoría del funcional de la densidad (DFT) para estudiar la estructura electrónica, las características de enlace y la respuesta elástica del C–S–H a nivel atómico. Posteriormente, se entrenó un MLFF *on-the-fly* utilizando simulaciones de dinámica molecular *ab initio* (AIMD) a 400 K. Después de un exhaustivo proceso de ajuste y validación, los MLFF resultantes fueron evaluados mediante relajación estructural y simulaciones de dinámica molecular, obteniendo predicciones de energía, fuerza y tensor de esfuerzo comparables a resultados precisos de DFT. Luego, los MLFF se utilizaron para calcular propiedades termodinámicas y mecánicas clave, incluida la ecuación de estado (EOS) y el módulo volumétrico, obteniendo valores entre 55-58 GPa, en favorable acuerdo con los datos experimentales disponibles. Asimismo, se evaluó y confirmó la transferibilidad del MLFF en un rango de temperatura de 200-400 K, manteniendo la coherencia física a lo largo de dicho intevalo.

Palabras clave: Teoría del Funcional de la Densidad, Silicatos Cálcicos Hidratados, Aprendizaje Automático, Campos de Fuerza, Dinámica Molecular *Ab Initio*, Ecuación de Estado, Módulo Volumétrico.

Abstract

Concrete is the second most used substance in the world after water, with more than 35 billion tonnes produced annually. Yet, understanding the atomic and mechanical properties of its principal component, calcium-silicate-hydrate (C–S–H)—the complex binder phase of concrete—remains a considerable challenge owing to its structural complexity and disordered nature. This works aims to develop and validate a machine learning force field (MLFF) capable of accurately reproducing the structural and mechanical behaviour of C–S–H. To this end, density functional theory (DFT) calculations were employed to study the electronic structure, bonding characteristics, and elastic response of C–S–H at the atomic level. Subsequently, an *on-the-fly* MLFF was trained using *ab initio* molecular dynamics (AIMD) simulations at 400 K. After a thorough refitting and validation process, the resulting MLFFs were assessed through structure relaxation and molecular dynamics simulations, yielding energy, force, and stress tensor predictions comparable to accurate DFT results. The MLFFs were then used to compute key thermodynamic and mechanical properties, including the equation of state (EOS) and bulk modulus, obtaining values bwetween 55–58 GPa, in favourable agreement with available experimental data. Furthermore, the transferability of the MLFF was evaluated and confirmed across a temparature range of 200–400 K, maintaining physical consistency throughout.

Keywords: Density Functional Theory, Calcium Silicate Hydrates, Machine Learning, Force Fields, Ab Initio Molecular Dynamics, Equation of State, Bulk Modulus.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Concrete is the synthetic material currently produced in volumes larger than any other material on Earth. With an annual consumption of approximately 35 billion metric tonnes, it is only second to water in terms of global usage [1, 2]. It plays a pivotal role in the construction industry, serving as the backbone for buildings, roads, bridges, dams, and many other infrastructure elements central to modern society. Its widespread adoption is the result of its unique combination of properties, including high compressive strength, durability, versatility, and cost-effectiveness [1], rendering it an important asset that directly influences the quality of life and economic development worldwide [3, 4]. Nevertheless, the massive production and use of concrete come with significant environmental challenges. The production of its main constituent, Portland cement, is responsible for 8-9% of the global anthropogenic CO_2 emissions [2]. Additionally, around 40% of produced concrete is employed to repair and maintain existing infrastructure [1], which aggravates the environmental impact of concrete. Therefore, the development of more durable and sustainable concrete is of utmost importance, which requires a better understanding of concrete's composition and microstructure.

Concrete itself is a composite material and can be regarded as a two-phase system [1]—the aggregate phase, composed of particles of varying size and shape; and the binding medium, composed of hydrated cement paste. The latter is, in turn, a heterogeneous mixture of different cement hydration products, with calcium silicate hydrate (C–S–H)[1] being the most abundant and important phase. C–S–H makes up 50 to 60% of the volume of solids in a hydrated cement paste and is responsible for the majority of the long-term strength and durability of concrete [1]. Together, the aggregate and binding phases form a complex microstructure that bridges the nanoscale chemistry of hydration products with the properties of concrete at the engineering scale.

[1] Conventional cement chemistry notation: C = CaO, S = SiO_2, H = H_2O

Nonetheless, the underlying microstructure-property relationships in concrete are not yet fully understood, hindering our ability to manipulate and tailor its properties for specific applications.

In this context, numerous efforts have been made to understand and model the properties of C–S–H, owing to its central role in determining the properties of concrete [5, 6, 7]. Characterisation techniques—such as X-ray diffraction (XRD) [8, 9, 10], nuclear magnetic resonance (NMR) [11, 12], and small angle neutron scattering (SANS)—have provided valuable insights into the structure and composition of C–S–H upon which many molecular models have been developed. The pioneering work of Pellenq *et al.* [13]—which introduced a realistic molecular model of cement hydrates—paved the way for a wide range of molecular modelling techniques [14, 15, 16, 17, 18] intended to capture the nanoscale structure and properties of C–S–H accurately. Additionally, the advancement of computational power and the development of efficient molecular dynamics (MD) simulation methods have enabled the exploration of mechanical, thermal, and transport properties of C–S–H [19, 20, 21, 22] under conditions that are relevant to concrete applications but difficult to replicate experimentally. The upscaling of C–S–H properties can be viewed from a hierarchical multiscale perspective, as illustrated in Figure 1.1, which highlights the microstructure-property relationships of C–S–H at different scales and the relevance of nanoscale properties to the engineering scale of concrete.

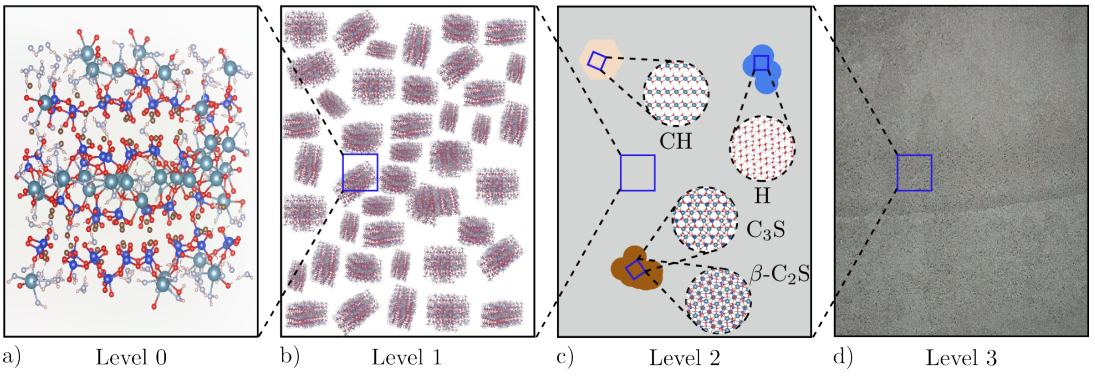

Figure 1.1: A four-level model representing the upscaling of C–S–H properties from the nanoscale to the engineering scale. (a) snapshot of C–S–H's nanostructure. (b) microstructure of C–S–H created by agglomeration of randomly oriented C–S–H nanoparticles. (c) microtexture of hardened paste composed of hydration products. (d) Macrotexture of cement paste at the engineering scale. Adapted from Ref. [19].

Despite the significant progress made in understanding C–S–H at the atomic scale, the inherent complexity of this material makes it challenging to model its structure and properties using classical methods realistically —primarily molecular dynamics (MD) simulations. Resorting to *ab initio* methods—such as density functional theory (DFT)—can hugely improve the accuracy of these models, but demand high computational resources, making it nearly intractable for real applications [23]. In this context, machine learning (ML) based concrete research has emerged as a promising approach to address these challenges [23, 24, 25]. Trained on large, high-quality *ab initio* datasets, ML models can capture the underlying physics of C–S–H and predict its

properties with high accuracy, comparable to that of first-principles methods, while requiring significantly less computational power.

1.2 Problem Statement

Concrete production is projected to increase by 50% annually by 2050 [2], and with no foreseeable alternatives to Portland cement, the urgent need for more sustainable concrete is evident. The development of advanced concrete with enhanced durability and performance could significantly lower the environmental burden. State-of-the-art methods such as machine learning show great potential to accelerate this transition, providing a powerful tool to advance our understanding of concrete's microstructure and properties.

In this regard, this report aims to investigate the performance of a machine learning force field (MLFF) in modelling and predicting the mechanical properties of C–S–H. Such MLFF will be trained and validated using *ab initio* data, ensuring it captures the complex atomic interactions and structural variability of C–S–H. Ultimately, our goal is to develop a reliable and computationally efficient model that can predict the mechanical properties of C–S–H, thereby supporting concrete research towards a sustainable future.

1.3 General and Specific Objectives

The main goal of this work is to use density functional theory (DFT), *ab initio* molecular dynamics (AIMD), and machine learning (ML) to develop a machine learning force field (MLFF) for calcium silicate hydrates (C–S–H) to study the mechanical properties of C–S–H. To achieve this, the following specific objectives are defined:

- To describe the theoretical foundations of DFT, AIMD, and MLFFs, including key concepts on exchange-correlation functionals such as PBE and PBEsol, and pseudopotentials.
- To perform geometric relaxation on bulk C–S–H model employing the VASP software with the PBEsol functional.
- To analyse the electronic structure of C–S–H by computing the density of states (DOS) of C–S–H.
- To train, refit, and test an MLFF using AIMD simulations.
- To compute the equation of state (EOS) and mechanical properties of C–S–H using the MLFF and simulated annealing methods.

- To evaluate the transferability of the MLFF by computing the thermal expansion coefficient of C–S–H.

1.4 Overview

The remainder of this report is organised in the following manner: Chapter 2 introduces the theoretical framework of the computational methods utilised in this work, including DFT, AIMD, and MLFFs. Chapter 3 presents the methodology employed to generate the results presented in this report, describing the molecular model of C–S–H, the computational setup, and the MLFF generation process. Then, Chapter 4 presents the results of our computational investigations. Ultimately, Chapter 5 finalises this report, the main conclusions about the work done are drawn, and the outlook for future work is discussed.

Chapter 2

Theoretical Background

This chapter presents the theoretical foundations and formalism of Density Functional Theory (DFT) and related methods necessary for the development of the results presented in this work. Starting with the many-body Schrödinger equation, this chapter covers the Born-Oppenheimer approximation, the Hartree-Fock approximation, Hohenberg-Kohn theorems, the Kohn-Sham equations, exchange-correlation functionals, and definitions on Ab initio molecular dynamics (AIMD) and machine learning force fields (MLFFs), along with implementation details in the Vienna Ab initio Simulation Package (VASP).

Ultimately, this chapter aims to provide a comprehensive understanding of the theoretical framework that underpins the computational methods utilised in this work, enabling the reader to grasp the principles and assumptions that govern the simulations and analysis performed throughout.

2.1 Many Body Schrödinger Equation

In our efforts to unravel the tapestry of materials, quantum mechanics guides our path towards describing the complex yet fundamental interactions that govern particle behavior at the atomic scale. Our journey shall commence by describing the physical laws that shape the interactions among particles constituting a system—electrons and nuclei alike.

2.1.1 The Coulomb Interaction

Materials may be thought of as complex assemblies of electrons and nuclei, held together by a delicate balance between attractive Coulomb interactions—primarily between electrons and nuclei—and repulsive interactions between like-charged particles, such as electron-electron and

nucleus-nucleus pairs, which govern the overall dynamics of the material system [26, 27, 28]. From classical electrostatics, these interactions can be mathematically expressed as follows:

- Electron-electron interactions

$$\hat{V}_{ee} = \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \tag{2.1}$$

- Electron-nucleus interactions

$$\hat{V}_{nn} = \frac{1}{2} \sum_{I \neq J} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{2.2}$$

- Electron-nuclei interactions

$$\hat{V}_{en} = -\sum_{i \neq I} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{2.3}$$

where e is the electronic charge, ϵ_0 is the vacuum permittivity, Z_I and Z_J are the atomic numbers of nuclei I and J, respectively, and $\mathbf{r}_i$ and $\mathbf{R}_I$ are the position vectors of electrons and nuclei, respectively. Moreover, we must also consider the kinetic energy of the collection of electrons and nuclei

$$\hat{T} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 \tag{2.4}$$

where $\hbar$ is the reduced Planck's constant, m_e is the electron mass, and M_I is the mass of the nucleus I.

2.1.2 The Time-Independent Schrödinger Equation

The Time-Independent Schrödinger Equation (TISE) lies at the heart of non-relativistic quantum mechanics, providing a mathematical framework to describe stationary electronic states of quantum systems. It takes the following form:

$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{2.5}$$

where $\hat{H}$ is the Hamiltonian of the system, incorporating both the kinetic and potential energies, $\psi(\mathbf{r})$ is an eigenstate of the system, and E is the energy eigenvalue associated with eigenstate $\psi(\mathbf{r})$. It is important to note that Equation 2.5 is only applicable to a single particle. However, a material system is composed of many electrons (N) and nuclei (M) with spatial coordinates

$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ and $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M$, respectively. Therefore, we must introduce a so-called many-body wavefunction given by:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M) \tag{2.6}$$

On this basis, the many-body version of Equation 2.5 shall be constructed by combining the kinetic (Equation 2.4) and potential (Equations 2.1, 2.2, and 2.3) energy contributions, leading to the following expression:

$$\begin{aligned}\Bigg[&-\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j|} + \\ &\frac{1}{2}\sum_{I\neq J}\frac{e^2}{4\pi\epsilon_0}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{i,I}\frac{e^2}{4\pi\epsilon_0}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}\Bigg]\Psi = E_{\text{tot}}\Psi\end{aligned} \tag{2.7}$$

Equation 2.7 provides a complete description of the stationary states of a non-relativistic many-body system, under time-independent conditions and in the absence of external fields. Additionally, we can achieve a more compact formulation by introducing the concept of atomic units. To this end, let us consider the simplest electron-nucleus system—the hydrogen atom—where the electron orbital has an average radius $a_0 \approx 0.529\,\text{Å}$. Thereby, the Coulomb energy for such a system is given by:

$$E_{\text{Ha}} = \frac{e^2}{4\pi\epsilon_0 a_0} \tag{2.8}$$

where 'Ha' stands for Hartree. Within this framework, the Hartree energy represents the Coulomb interaction between two fundamental charges separated by a distance of one Bohr radius (a_0). Moreover, the angular momentum quantisation condition for the electron in the hydrogen atom is given by

$$m_e v a_0 = \hbar \tag{2.9}$$

Additionally, we can express the equilibrium condition between the nuclear attraction and the electron's centrifugal force as:

$$\frac{e^2}{4\pi\epsilon_0 a_0^2} = \frac{m_e v^2}{a_0} \tag{2.10}$$

By combining Equations 2.8, 2.9, and 2.10, we derive the following relationships:

$$\frac{e^2}{4\pi\epsilon_0 a_0} = \frac{\hbar^2}{m_e a_0^2} \tag{2.11}$$

$$\frac{1}{2}m_e v^2 = \frac{1}{2}E_{\text{Ha}} \tag{2.12}$$

The latter relation showcases that the kinetic energy is of the same order as E_{Ha}, rendering it convenient to normalise Equation 2.7 by this quantity:

$$\begin{aligned}\Bigg[-\sum_i \frac{1}{2}a_0^2\nabla_i^2 - \sum_I \frac{1}{2}\frac{1}{(M_I/m_e)}\nabla_I^2 + \frac{1}{2}\sum_{i\neq j}\frac{a_0}{|\mathbf{r}_i-\mathbf{r}_j|} \\ +\frac{1}{2}\sum_{I\neq J} Z_I Z_J \frac{a_0}{|\mathbf{R}_I-\mathbf{R}_J|} - \sum_{i,I} Z_I \frac{a_0}{|\mathbf{r}_i-\mathbf{R}_I|}\Bigg]\Psi = \frac{E_{\mathrm{tot}}}{E_{\mathrm{Ha}}}\Psi\end{aligned} \tag{2.13}$$

A final simplification involves setting our energy units to Ha, distance units to a_0, and mass units to m_e. The last missing constant e is set to 1, leading to the following expression:

$$\begin{aligned}\Bigg[-\sum_i \frac{\nabla_i^2}{2} - \sum_I \frac{\nabla_I^2}{2M_I} + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} + \\ \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|}\Bigg]\Psi = E_{\mathrm{tot}}\Psi\end{aligned} \tag{2.14}$$

Ultimately, even though Equation 2.14 provides an exact method capable of yielding various properties of a material system—such as elastic, thermal, and electronic properties—a combination of mathematical complexity and computational limitations renders it intractable to solve for any realistic system. Moreover, the wavefunction contains vastly more information than is necessary to describe most observable properties of a material. Therefore, we must resort to alternative formulations that allow us to extract only the relevant information from the wavefunction while reducing the computational cost of the calculations. The remainder of this chapter is dedicated to presenting such alternative approaches that ultimately lead to the computational methods employed throughout this work.

2.2 The Born-Oppenheimer Approximation

For atoms in a solid, we can think of nuclei as being held immobile in a fixed position, while electrons instantaneously react to any nucleus's movement. This assumption is based on the fact that nuclei are much heavier than electrons—by three to four orders of magnitude—making the former behave like classical particles. Thereby, we can rewrite the many-body wavefunction as a product of two wavefunctions:

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{R}_M) = \psi_{\mathbf{R}}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\chi(\mathbf{R}) \tag{2.15}$$

where $\psi_{\mathbf{R}}$ is the electronic wavefunction parametrised by the nuclear positions $\mathbf{R}$, and χ is the nuclear wavefunction. Furthermore, this significant mass disparity enables a systematic

approximation scheme, wherein the electronic wavefunction is solved for fixed nuclei, and its solution is used as an effective potential for the nuclear dynamics afterwards. First, nuclei' kinetic energy is neglected, as their positions are assumed to be fixed:

$$\sum_I \frac{\nabla_I^2}{2M_I} = 0 \quad \text{and} \quad E = E_{\text{tot}} - \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{2.16}$$

Following, we define the Coulomb potential of the nuclei experienced by the electrons as:

$$V_{\text{n}}(\mathbf{r}) = -\sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{2.17}$$

Then, Equation 2.14 can be rewritten as:

$$\left[-\sum_i \frac{\nabla_i^2}{2} + \sum_i V_{\text{n}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \Psi = E\Psi \tag{2.18}$$

Finally, by using Equation 2.15, we can define the electronic and nuclear Schrödinger equations as follows:

$$\left[-\sum_i \frac{\nabla_i^2}{2} + \sum_i V_{\text{n}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \psi_{\mathbf{R}} = E_{\mathbf{R}} \psi_{\mathbf{R}} \tag{2.19}$$

$$\left[-\sum_I \frac{\nabla_I^2}{2M_I} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} + E(\mathbf{R}_1, \ldots, \mathbf{R}_M) \right] \chi(\mathbf{R}) = E_{\text{tot}} \chi(\mathbf{R}) \tag{2.20}$$

where $E_{\mathbf{R}} = E(\mathbf{R}_1, \ldots, \mathbf{R}_M)$ is the electronic surface energy, which is a function of the nuclear positions, and serves as an effective potential shaping the nuclear dynamics.

2.3 Hartree-Fock Approximation

The essence of the Hartree-Fock approximation (HFA) is to approximate the interacting many-electron system (Equation 2.18) by a set of non-interacting single-particle problems subject to an effective mean-field potential [29, 30, 31]. As a means to this, we first rewrite the total wavefunction for a system with N electrons as the product of single-electron wavefunctions—often referred to as the Hartree approximation [32]—as showcased in Equation 2.21.

$$\Psi^H(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \prod_{i=1}^{N} \phi_i(\mathbf{r}_i) \tag{2.21}$$

Following, we construct the total energy functional as the expectation value of the Hamiltonian operator:

$$E^H[\{\phi_i\}] = \left\langle \Psi^H \middle| \hat{H} \middle| \Psi^H \right\rangle \tag{2.22}$$

Expanding Equation 2.22:

$$E^H[\{\phi_i\}] = \sum_{i=1}^{N} \int \phi_i^*(\mathbf{r}) \left(-\frac{\nabla^2}{2} + V_{\mathrm{n}}(\mathbf{r}) \right) \phi_i(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \sum_{i \neq j} \int \int \frac{|\phi_i(\mathbf{r})|^2 |\phi_j(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{2.23}$$

where the first term sums the kinetic energy and electron-nuclear attraction for all electrons, while the second accounts for the classical electron-electron repulsion energy averaged over the electron density distribution. In order to find the set of orbitals $\{\phi_i\}$ that minimises the total energy functional, we use the variational principle, where we shall impose the orthonormality condition:

$$\int \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} = \delta_{ij} \tag{2.24}$$

for what we introduce the Lagrange multipliers λ_{ij} to enforce these constraints and define the Lagrangian:

$$\mathcal{L} = E^H[\{\phi_i\}] - \sum_{i=1}^{N} \lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \tag{2.25}$$

which ultimately simplifies to:

$$\mathcal{L} = E^H[\{\phi_i\}] - \sum_{i=1}^{N} \varepsilon_i (\langle \phi_i | \phi_i \rangle - 1) \tag{2.26}$$

Then, we need to compute the derivative of $\mathcal{L}$ with respect to ϕ_i^* and set it to zero:

$$\frac{\delta \mathcal{L}}{\delta \phi_i^*}(\mathbf{r}) = 0 \tag{2.27}$$

which yields the Hartree equation:

$$\left[-\frac{\nabla^2}{2} + V_{\mathrm{n}}(\mathbf{r}) + V_i^H(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{2.28}$$

where $V_n(\mathbf{r})$ represents the electrostatic interaction between electrons and nuclei, and the Hartree potential

$$V_i^H(\mathbf{r}) = \sum_{j \neq i} \int \frac{|\phi_j(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{2.29}$$

accounts for the average electrostatic interaction experienced by the i-th electron due to all other electrons in the system. This effective mean-field potential replaces the electron-electron interactions, effectively simplifying the many-body problem into single-particle problems.

2.3.1 The Pauli Exclusion Principle

So far, we have introduced the Hartree approximation, which assumes that the many-electron wavefunction can be expressed as a product of single-particle wavefunctions. However, this approach does not account for the indistinguishability of electrons and the Pauli exclusion principle, which states that no two fermions—half-spin particles, such as electrons—can reside in the same quantum state simultaneously. In doing so, it imposes a restriction on the possible configurations of electrons in a system that shall be accounted for.

In order to achieve this, V. Fock [33] introduced a different approximation to the wavefunction by using a Slater determinant—a mathematical construct that combines one-electron wavefunctions in such a way that satisfies the antisymmetry principle. This is done by expressing the overall wavefunction as the determinant of a matrix of single-electron wavefunctions

$$\Psi^{HF}(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \ldots & \phi_1(\mathbf{r}_N) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \ldots & \phi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1) & \phi_N(\mathbf{r}_2) & \ldots & \phi_N(\mathbf{r}_N) \end{vmatrix} \tag{2.30}$$

where $1/\sqrt{N!}$ is a normalisation factor. To illustrate this, consider a two-electron system with single-particle wavefunctions $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$. The Slater determinant for this system would be

$$\Psi^{HF}(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}\begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) \end{vmatrix} = \frac{1}{\sqrt{2}}\left[\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) - \phi_1(\mathbf{r}_2)\phi_2(\mathbf{r}_1))\right] \tag{2.31}$$

Evidently, $\Psi^{HF}(\mathbf{r}_1,\mathbf{r}_2) = -\Psi^{HF}(\mathbf{r}_2,\mathbf{r}_1)$, which satisfies the antisymmetry principle.

2.3.2 The Hartree-Fock Equations

The Hartree-Fock equations are derived similarly to how we addressed the Hartree equations. We first define the total energy with the Hartree-Fock wavefunction (Equation 2.30)

$$\begin{aligned} E^{HF}[\{\phi_i\}] &= \left\langle \Psi^{HF} \middle| \hat{H} \middle| \Psi^{HF} \right\rangle \\ &= \sum_i \langle \phi_i | \frac{\nabla^2}{2} + V_{\mathrm{n}}(\mathbf{r}) | \phi_i \rangle \\ &+ \frac{1}{2}\sum_{i\neq j} \langle \phi_i\phi_j | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \phi_i\phi_j \rangle \\ &- \frac{1}{2}\sum_{i\neq j} \langle \phi_i\phi_j | \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \phi_j\phi_i \rangle \end{aligned} \tag{2.32}$$

Consequently, using the variational principle, we derive the Hartree-Fock equations:

$$\left[-\frac{\nabla^2}{2} + V_{\mathrm{n}}(\mathbf{r}) + V_i^H(\mathbf{r}) +\right] \phi_i(\mathbf{r}) - \sum_{j\neq i} \langle\phi_j| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} |\phi_i\rangle \phi_j(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{2.33}$$

Noticeably, Equation 2.33 has an extra term compared with the Hartree equation (Equation 2.28). This term is called the "exchange" term [28], and describes the effects of exchange between electrons. It is convenient to try to express the Hartree-Fock equations in a more compact form, so we define the single-particle and total densities as

$$\rho_i(\mathbf{r}) = |\phi_i(\mathbf{r})|^2 \tag{2.34}$$

$$\rho(\mathbf{r}) = \sum_i \rho_i(\mathbf{r}) \tag{2.35}$$

thus, the Hartree potential can be expressed as

$$V_i^H(\mathbf{r}) = \sum_{j\neq i} \int \frac{\rho_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = \int \frac{\rho(\mathbf{r}') - \rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{2.36}$$

Therefore, we can construct the single-particle exchange density as

$$\rho_i^X(\mathbf{r}, \mathbf{r}') = \sum_{j\neq i} \frac{\phi_i(\mathbf{r}')\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')}{\phi_i(\mathbf{r})\phi_i^*(\mathbf{r})} \tag{2.37}$$

Finally, the Hartree-Fock equations take the form

$$\left[-\frac{\nabla^2}{2} + V_{\mathrm{n}}(\mathbf{r}) + V_i^H(\mathbf{r}) + V_i^X(\mathbf{r})\right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{2.38}$$

where V_i^X stands for the exchange potential

$$V_i^X(\mathbf{r}) = -\int \frac{\rho_i^X(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{2.39}$$

2.4 Density Functional Theory

So far, we have acknowledged that determining the state of a system with N electrons remains a formidable challenge, for it involves a wavefunction defined in a 3N-dimensional space. We also recognise that it is possible—heuristically speaking—to simplify such representation by utilising products of single-particle wavefunctions. Nevertheless, such an independent electron approximation necessitates the wavefunctions to be explicitly specified, thereby yielding a rather drastic approximation for the behaviour of the system. Thus, it is natural to consider a different

approach to develop an exact single-particle framework, onto which approximations can be introduced afterwards.

We hereby introduce the Density Functional Theory (DFT), which draws upon the insight that any property of a system of many electrons can be viewed as a functional of the ground-state density $n(\mathbf{r})$ [34] (Equation 2.40)—a scalar function defined over three spatial coordinates.

$$n(\mathbf{r}) = N \int \Psi^*(\mathbf{r}, \ldots, \mathbf{r}_N)\Psi(\mathbf{r}, \ldots, \mathbf{r}_N)d\mathbf{r}_2 \ldots d\mathbf{r}_N \tag{2.40}$$

The foundational principles of DFT were established in the original papers by Hohenberg, Kohn, and Sham [35, 36], where they present two theorems that establish the theoretical framework of DFT. However, for this discussion, we will base our exposition on explanatory texts [34, 26, 28, 27].

2.4.1 First Hohenberg-Kohn Theorem

Theorem 2.1 (First Hohenberg-Kohn Theorem)**.** *The ground-state electron density $n(\mathbf{r})$ uniquely determines the external potential $V(\mathbf{r})$ and, consequently, the ground-state energy E_0 of a many-electron system.*

Proof. Suppose two different external potentials, $V(\mathbf{r})$ and $V'(\mathbf{r})$ (different ionic potentials) yield the same ground-state electron density $n(\mathbf{r})$. Given that $V(\mathbf{r})$ and $V'(\mathbf{r})$ are different in a non-trivial way, we will show that this statement leads to a contradiction. Let E and Ψ be the total energy and wavefunction and E' and Ψ' be the total energy and wavefunction corresponding to the systems with hamiltonians $\hat{H}$ and $\hat{H}'$, respectively, with the first hamiltonian containing $V(\mathbf{r})$ and the second containing $V'(\mathbf{r})$ as an external potential:

$$\hat{H} = \hat{T} + \hat{U} + V, \quad \hat{H}' = \hat{T} + \hat{U} + V', \quad E = \langle\Psi|\hat{H}|\Psi\rangle, \quad E' = \langle\Psi'|\hat{H}'|\Psi'\rangle$$

Here, $\hat{T}$ and $\hat{U}$ correspond to the kinetic and interaction energy operators, thereby being common for both Hamiltonians. Now, we assume that the ground states of the two Hamiltonians are different because the external potentials are different. Then, according to the variational principle:

$$\begin{aligned} E < \langle\Psi'|\hat{H}|\Psi'\rangle &= \langle\Psi'|\hat{T} + \hat{U} + V + V' - V'|\Psi'\rangle \\ &= \langle\Psi'|\hat{H}' + V - V'|\Psi'\rangle \\ &= E' + \langle\Psi'|(V - V')|\Psi'\rangle \end{aligned} \tag{2.41}$$

Following the same reasoning, we can prove that

$$E' < E - \langle\Psi|(V - V')|\Psi\rangle \tag{2.42}$$

Adding Equations 2.41 and 2.42, we obtain:

$$E + E' < E' + E - \langle \Psi | (V - V') | \Psi \rangle + \langle \Psi' | (V - V') | \Psi' \rangle \tag{2.43}$$

where the last two terms result in

$$\int n'(\mathbf{r})(V - V')d\mathbf{r} - \int n(\mathbf{r})(V - V')d\mathbf{r} = 0 \tag{2.44}$$

since $n(\mathbf{r}) = n'(\mathbf{r})$ by assumption. Finally, we arrive at the following expression:

$$E + E' < E + E' \tag{2.45}$$

This is a contradiction, which implies that our initial assumption about the densities being the same ought to be false, thereby proving there is a one-to-one correspondence between an external potential $V(\mathbf{r})$ and the electron density $n(\mathbf{r})$. Moreover, since $V(\mathbf{r})$ determines the wavefunction, the wavefunction must be a unique functional of the density. So we conclude that the expression

$$\mathcal{F}[n(\mathbf{r})] = \langle \Psi | \hat{T} + \hat{U} | \Psi \rangle \tag{2.46}$$

must be a universal functional of the electronic density—*i.e.*, common to all solids—and that the ground state energy is a functional of the density:

$$E[n(\mathbf{r})] = \mathcal{F}[n(\mathbf{r})] + \int V(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \tag{2.47}$$

□

2.4.2 Second Hohenberg-Kohn Theorem

Theorem 2.2 (Second Hohenberg-Kohn Theorem)**.** *The ground-state energy E can be obtained through the variation of trial densities $\tilde{n}(\mathbf{r})$ instead of trial wavefunctions $\tilde{\Psi}$.*

Proof. First, we fix a trial density $\tilde{n}(\mathbf{r})$ and define the trial wavefunctions $\tilde{\Psi}^{\alpha}_{\tilde{n}(\mathbf{r})}$. Therefore, the constrained energy minimum is defined as

$$\begin{aligned} E[\tilde{n}(\mathbf{r})] &= \min_{\alpha} \langle \tilde{\Psi}^{\alpha}_{\tilde{n}(\mathbf{r})} | \hat{H} | \tilde{\Psi}^{\alpha}_{\tilde{n}(\mathbf{r})} \rangle \\ &= \mathcal{F}[\tilde{n}(\mathbf{r})] + \int V(\mathbf{r}) \tilde{n}(\mathbf{r}) d\mathbf{r} \end{aligned} \tag{2.48}$$

Secondly, we minimise Equation 2.48 over all n

$$E = \min_{\tilde{n}(\mathbf{r})} E[\tilde{n}(\mathbf{r})] = \min_{\tilde{n}(\mathbf{r})} \left\{ \mathcal{F}[\tilde{n}(\mathbf{r})] + \int V(\mathbf{r}) \tilde{n}(\mathbf{r}) d\mathbf{r} \right\} \tag{2.49}$$

For a non-degenerate ground state, the minimum corresponds to the ground-state $n(\mathbf{r})$, or to one of the ground-state densities otherwise. □

Finally, we have managed to map the formidable challenge of finding the minimum of $\langle\Psi|\hat{H}|\Psi\rangle$ involving a 3N-dimensional wavefunction into a much simpler problem of finding the minimum of $E[n(\mathbf{r})]$ involving a 3-dimensional function. This is the essence of DFT, which allows us to compute the ground-state properties of many-electron systems without explicitly solving the many-body Schrödinger equation.

2.4.3 Kohn-Sham Equations

Even though the Hohenberg-Kohn theorems provide a rigorous foundation for DFT, they offer no guidance whatsoever for constructing the functional $\mathcal{F}[n(\mathbf{r})]$. As such, density functional theory would lack practical utility if it were not for the auxiliary system proposed by Kohn and Sham [34]. It consists of replacing the many-electron problem by an auxiliary independent-particle problem that yields the same ground-state density, incorporating the many-body effects into a so-called exchange-correlation functional. To this end, we first define the density of the auxiliary system as

$$n(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{2.50}$$

where $\phi_i(\mathbf{r})$ are the single-particle wavefunctions of the auxiliary system. Next, we define the independent-particle kinetic energy functional as

$$T_s[n(\mathbf{r})] = \frac{1}{2}\sum_i \int \phi_i^*(\mathbf{r})(-\nabla^2)\phi_i(\mathbf{r})d\mathbf{r} \tag{2.51}$$

and we can redefine the Hartree energy functional as

$$E_H[n(\mathbf{r})] = \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' \tag{2.52}$$

yielding the following expression for the total energy functional:

$$E^{\mathrm{KS}}[n(\mathbf{r})] = T_s[n(\mathbf{r})] + \int V_{\mathrm{n}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] \tag{2.53}$$

where $E_{xc}[n(\mathbf{r})]$ is the exchange-correlation energy functional is defined as

$$E_{xc}[n(\mathbf{r})] = \langle\hat{T}\rangle - T_s[n(\mathbf{r})] + \langle\hat{U}\rangle - E_H[n(\mathbf{r})] \tag{2.54}$$

with $\langle \hat{T} \rangle$ and $\langle \hat{U} \rangle$ denoting the exact kinetic energy and electron–electron interaction energy, respectively. Finally, we choose a variation in the density to be

$$\delta n(\mathbf{r}) = \delta \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) \tag{2.55}$$

along with the following constraint

$$\int \delta n(\mathbf{r}) d\mathbf{r} = \int \delta \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) d\mathbf{r} = 0 \tag{2.56}$$

and by applying the Kohn-Sham variational principle, we arrive at the Kohn-Sham equations:

$$\left[-\frac{\nabla^2}{2} + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{2.57}$$

where $V_{\text{ext}}(\mathbf{r})$ is the external potential, $V_H(\mathbf{r})$ is the Hartree potential, and $V_{xc}(\mathbf{r})$ is the exchange-correlation potential defined as

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{2.58}$$

Our task now shall focus on constructing appropriate approximations for the exchange-correlation functional $E_{xc}[n(\mathbf{r})]$.

2.4.4 Exchange-Correlation Functionals

The usefulness of DFT relies entirely on whether we can construct reliable approximations for the exchange-correlation functional $E_{xc}[n(\mathbf{r})]$ with sufficient accuracy and computational efficiency. Therefore, we shall now present the most commonly used approximations for $E_{xc}[n(\mathbf{r})]$, as well as their strengths and weaknesses.

2.4.4.1 Local Density Approximation

The simplest—and remarkably effective—approximation for the exchange-correlation functional is the local-density approximation (LDA) [37]. It approximates the exchange-correlation energy of an inhomogeneous electron system by that of a homogeneous electron gas (HEG) having the same electron density.

$$E_{xc}^{LDA}[n(\mathbf{r})] = \int n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r})) d\mathbf{r} \tag{2.59}$$

where $\epsilon_{xc}(n)$ is the exchange and correlation energy per electron of a uniform electron gas of density n [35, 36]. This quantity depends solely on the local density and surrounding electrons in the vicinity of $\mathbf{r}$, for example, a sphere of radius $\sim \lambda_F(\mathbf{r})$—the local Fermi wavelength [31] $\lambda_F(\mathbf{r}) \equiv [3\pi^2 n(\mathbf{r})]^{-1/3}$.

The exchange and correlation contributions to E_{xc}^{LDA} can be separated into two terms

$$E_{xc}^{LDA}[n(\mathbf{r})] = E_x^{HEG}[n(\mathbf{r})] + E_c^{HEG}[n(\mathbf{r})] \tag{2.60}$$

The first term corresponds to the exchange energy density contribution

$$\begin{aligned} \epsilon_x^{HEG}(n) &= -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3} \\ &= -\frac{0.458128}{r_s} \end{aligned} \tag{2.61}$$

where r_s is the Wigner-Seitz radius—the radius of a sphere containing one electron and given by $(4\pi/3)r_s^3 = n^{-1}$.

The second term corresponds to the correlation energy, which was computed by Ceperley and Alder [38] using quantum Monte Carlo methods. Subsequently, the extracted data were parameterised by Perdew and Zunger [39], yielding the following expression:

$$\epsilon_c^{HEG}(r_s) = \begin{cases} 0.0311 \ln r_s - 0.0480 + 0.002 r_s \ln r_s - 0.0116 r_s & r_s < 1 \\ \dfrac{-0.1423}{1 + 1.0529\sqrt{r_s} + 0.3334 r_s} & r_s \geq 1 \end{cases} \tag{2.62}$$

LDA has demonstrated remarkable success for most applications involving systems with slowly varying densities or systems with high densities. Nevertheless, LDA—and its spin-polarised version (LSDA)— breaks down in systems governed by strong correlation effects, losing any resemblance to non-interacting electron gases.

2.4.4.2 Generalised Gradient Approximation

In contrast to LDA—which considers the electronic density to be locally uniform—the generalised gradient approximation (GGA) systematically improves upon LDA by incorporating not only the local density $n(\mathbf{r})$, but also its gradient $\nabla n(\mathbf{r})$, thereby accounting for the inhomogeneities in the electron distribution.

Within this framework, the exchange-correlation energy functional is expressed as a function of the local density and its gradient

$$E_{xc}^{GGA}[n(\mathbf{r})] = \int n(\mathbf{r}) \epsilon_{xc}^{HEG}(n(\mathbf{r})) f(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r}), \nabla n_\uparrow(\mathbf{r}), \nabla n_\downarrow(\mathbf{r})) d\mathbf{r} \tag{2.63}$$

where $n_\uparrow(\mathbf{r})$ is the spin-up electron density and $n_\downarrow(\mathbf{r})$ is the corresponding spin-down density. Additionally, the exact form of f—a parametrised analytic function—depends on the GGA under consideration.

In this regard, one of the most prominent and widely adopted GGA functionals is the Perdew-Burke-Ernzerhof (PBE) [40] functional, which was proposed as a solution for the drawbacks of previously proposed GGAs, such as the Perdew-Wang (PW91) functional. Within PBE, all parameters—other than those in $\epsilon_{xc}^{HEG}(n(\mathbf{r}))$—are fundamental constants. Consequently, the exchange energy term in the PBE functional is given by

$$E_x^{PBE} = \int n(\mathbf{r})\epsilon_x^{HEG}(n(\mathbf{r}))\left[1+\kappa-\frac{\kappa}{1+\mu s^2/\kappa}\right]d\mathbf{r} \tag{2.64}$$

where $\kappa = 0.804$ and $\mu = 0.219$ are the parameters of the PBE functional, $k_F = (3\pi^2 n(\mathbf{r}))^{1/3}$ is the local Fermi wavevector and $s = |\nabla n(\mathbf{r})|/(2k_F n(\mathbf{r}))$ is a dimensionless density gradient.

The correlation energy term in the PBE functional is expressed as

$$E_c^{PBE} = \int n(\mathbf{r})\left[\epsilon_c^{HEG} + \gamma\phi^3 \ln\left\{1+\frac{\beta}{\gamma}t^2\left[\frac{1+At^2}{1+At^2+A^2t^4}\right]\right\}\right] \tag{2.65}$$

where $\gamma = 0.031091$, $\beta = 0.066725$, ϕ is a spin-scaling factor, and A and t are defined as

$$A = \frac{\beta}{\gamma}\left[\exp\left\{\frac{-\epsilon_c^{HEG}}{\gamma\phi^3}\right\}-1\right]^{-1}, \quad t(\mathbf{r}) = \frac{|\nabla n(\mathbf{r})|}{2\phi k_s n(\mathbf{r})} \tag{2.66}$$

where $k_s = \sqrt{4k_F/\pi}$ is the Thomas-Fermi screening wavenumber.

Even though PBE holds a significant advantage over LDA in terms of accuracy, it is not exempt from certain limitations and shortcomings. PBE tends to overestimate equilibrium lattice constants by about 1%—LDA underestimates them by the same amount—which is detrimental for accurate calculations of other equilibrium properties, such as bulk moduli, phonon frequencies, and magnetism.

To address this issue, a revised version of the PBE—the PBEsol functional [41]—was developed, which improves equilibrium properties of densely-packed solids and their surfaces, reducing the overestimation of lattice constants by a factor of ~ 4. Nonetheless, PBEsol does not perform well for semiconductors and can lead to less accurate total energy calculations compared to the PBE method. Ultimately, the choice of GGA functional depends on the specific system and properties being investigated, as well as the desired balance between accuracy and computational efficiency.

2.4.4.3 Hybrid Functionals

Functionals mentioned above and their inherent limitations motivated the exploration of hybrid functionals. They offer improved accuracy by incorporating a fraction of the exact nonlocal Hartree-Fock exchange energy into the exchange-correlation functional, allowing efficient yet accurate calculations.

One of these functionals includes PBE0 [42]—a combination of the PBE functional and the exact Hartree-Fock exchange energy. It is defined as

$$E_{xc}^{PBE0} = \frac{1}{4}E_x^{HF} + \frac{3}{4}E_x^{PBE} + E_c^{PBE} \tag{2.67}$$

Another prominent example is the HSE (Heyd-Scuseria-Ernzerhof) functional [43], which splits the exchange energy into short-range (SR) and long-range (LR) contributions

$$E_{xc}^{HSE06} = \frac{1}{4}E_x^{HF,SR}(\omega) + \frac{3}{4}E_x^{PBE,SR}(\omega) + E_c^{PBE,LR}(\omega) + E_c^{PBE}(\omega) \tag{2.68}$$

where ω is an adjustable parameter that controls the short and long-range separation in the decomposed Coulomb operator

$$\frac{1}{r} = SR_\omega(\mathbf{r}) + LR_\omega(\mathbf{r}) = \frac{erfc(\omega r)}{r} + \frac{erf(\omega r)}{r} \tag{2.69}$$

Ultimately, hybrid functionals provide a considerable improvement in the accuracy of electronic structure calculations over LDA and GGA. Nonetheless, this advantage comes at a higher computational cost. This hierarchy of exchange-correlation functionals is better summarized in the so-called Jacob's ladder [44] (Figure 2.1), which depicts the trade-off between accuracy and computational cost as we ascend the ladder towards more sophisticated functionals.

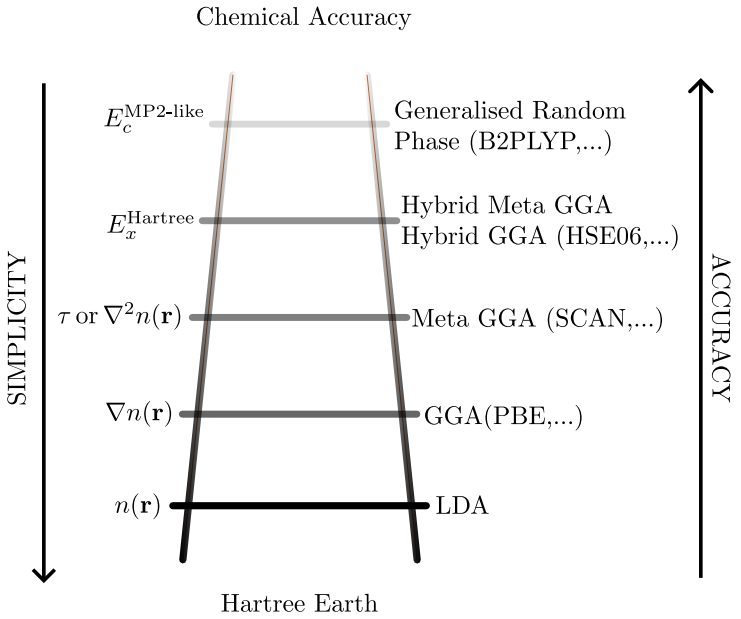

Figure 2.1: Jacob's ladder of exchange-correlation functionals as proposed by J. Perdew [44]. The ladder categorises the functionals into rungs, from the simplest approximation (LDA) at the bottom, progressing to more sophisticated and accurate approximations (Generalised Random Phase) at the uppermost rung.

2.5 Ab initio Molecular Dynamics

Molecular dynamics (MD) [45, 46] is a widely used computational method that allows us to simulate a many-body condensed matter system and compute its thermodynamic and dynamical properties. In MD, a system is modelled as a collection of particles—atoms or molecules—whose trajectories evolve under the influence of interatomic forces following Newton's equations of motion.

One of the most challenging yet crucial aspects of MD is calculating the interatomic forces. In classical MD, these forces are typically computed using predefined potential functions or force fields, either constructed upon empirical data or from independent electronic structure calculations that have been parameterised to reproduce experimental or *ab initio* data for small reference systems. Despite their fair success—which we acknowledge but shall not discuss in detail—these empirical interatomic potentials are often limited in their accuracy and transferability. Certain atoms, molecules, or even large systems may give rise to highly complex interatomic interactions that, if attempted to be modelled with empirical potentials, would require a significant amount of effort. Likewise, these potentials are very often limited to a narrow range of configurations, making them ill-suited for processes involving significant structural changes, such as phase transitions or large deformations.

Therewith, classical MD can be extended by a first-principles approach, where the interatomic forces are computed on-the-fly from accurate electronic structure calculations, ultimately leading to *ab initio* molecular dynamics (AIMD). This approach enables us to overcome the limitations outlined for classical MD, albeit at a significant computational cost. AIMD employs electronic structure methods, such as DFT, to compute the interatomic forces at each time step, without relying on predefined interatomic potentials, thereby providing AIMD with improved predictive power and flexibility.

2.5.1 Hellmann-Feynman Theorem

The Hellmann-Feynman theorem [47, 48] establishes a relation between the derivative of the total energy E of a system concerning a parameter λ and the expectation value of the derivative of the Hamiltonian with respect to that same parameter. While the proof of this theorem is relatively straightforward—as shown later in this section—it plays a pivotal role in computing interatomic forces in AIMD.

To this end, let us consider the total energy of a system

$$E = \langle \Psi \big| \hat{H} \big| \Psi \rangle \tag{2.70}$$

If λ is a parameter that appears explicitly in the Hamiltonian, then

$$\begin{aligned}\frac{\partial E}{\partial \lambda} &= \frac{\partial}{\partial \lambda}\langle\Psi|\hat{H}|\Psi\rangle \\ &= \left\langle \Psi \middle| \frac{\partial \hat{H}}{\partial \lambda} \middle| \Psi \right\rangle + \left\langle \frac{\partial \Psi}{\partial \lambda} \middle| \hat{H} \middle| \Psi \right\rangle + \left\langle \Psi \middle| \hat{H} \middle| \frac{\partial \Psi}{\partial \lambda} \right\rangle\end{aligned} \tag{2.71}$$

Provided that $\hat{H}$ is hermitian and Ψ is an eigenstate of the hamiltonian, $\hat{H}|\Psi\rangle = E|\Psi\rangle$, Equation 2.71 simplifies to

$$\begin{aligned}\frac{\partial E}{\partial \lambda} &= \left\langle \Psi \middle| \frac{\partial \hat{H}}{\partial \lambda} \middle| \Psi \right\rangle + \left\langle \frac{\partial \Psi}{\partial \lambda} \middle| \hat{H} \middle| \Psi \right\rangle + \left\langle \frac{\partial \Psi}{\partial \lambda} \middle| \hat{H} \middle| \Psi \right\rangle \\ &= \left\langle \Psi \middle| \frac{\partial \hat{H}}{\partial \lambda} \middle| \Psi \right\rangle + E\frac{\partial}{\partial \lambda}\langle\Psi|\Psi\rangle + E\frac{\partial}{\partial \lambda}\langle\Psi|\Psi\rangle\end{aligned} \tag{2.72}$$

where the last two terms add up to zero, so that Equation 2.72 becomes,

$$\frac{\partial E}{\partial \lambda} = \left\langle \Psi \middle| \frac{\partial \hat{H}}{\partial \lambda} \middle| \Psi \right\rangle \tag{2.73}$$

yielding the Hellmann-Feynman theorem.

When λ corresponds to the position of a nucleus, the negative derivative of the total energy with respect to λ results in the force acting on that nucleus. More generally, the force on nucleus I can be expressed as the negative gradient of the total energy with respect to its position R_I,

$$\mathbf{F}_I = -\nabla_{R_I}\langle\Psi|\hat{H}|\Psi\rangle \tag{2.74}$$

This result—referred to as the Hellmann-Feynman force—provides a fundamental connection between the quantum mechanical energy landscape and the classical nuclei motion in AIMD.

2.5.2 Born-Oppenheimer Molecular Dynamics

As previously discussed, AIMD relies on solving the static electronic structure problem at each time step, given a set of fixed nuclear positions at a certain instant in time. One approach to achieve this is the so-called Born-Oppenheimer molecular dynamics (BOMD) [46], wherein the potential energy $E[\{\psi_i\};\mathbf{R}]$ is minimised at each time step with respect to the single-electron wavefunctions $\{\psi_i(\mathbf{r})\}$ under the orthonormality constraint $\langle\psi_i(\mathbf{r})|\psi_j(\mathbf{r})\rangle = \delta_{ij}$. The result is a potential energy surface on which the nuclei evolve classically.

This leads to the following Lagrangian

$$\mathcal{L}_{\mathrm{BO}}(\{\psi_i\};\mathbf{R},\dot{\mathbf{R}}) = \frac{1}{2}\sum_I M_I\dot{\mathbf{R}}_I^2 - \min_{\{\psi_i\}} E[\{\psi_i(\mathbf{r})\};\mathbf{R}] + \sum_{ij}\lambda_{ij}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right) \tag{2.75}$$

where the first term on the right-hand side corresponds to the classical kinetic energy of the nuclei. The second term, $E[\{\psi_i\};\mathbf{R}] = E^{\mathrm{KS}}[\{\psi_i[n(\mathbf{r})]\};\mathbf{R}] + E_{II}$, is the total potential energy consisting of the Kohn-Sham ground-state energy and the nuclear-nuclear interaction energy.

By applying the Euler-Lagrange equations to Equation 2.75

$$\begin{aligned} \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} &= \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I} \\ \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \langle \dot{\psi}_i|} &= \frac{\partial \mathcal{L}}{\partial \langle \psi_i|} \end{aligned} \tag{2.76}$$

we obtain the equations of motion

$$M_I \ddot{\mathbf{R}}_I = -\nabla_{\mathbf{R}_I} \left[\min_{\{\psi_i\}} E[\{\psi_i(\mathbf{r})\};\mathbf{R}] \Big|_{\{\langle \psi_i(\mathbf{r})|\psi_j(\mathbf{r})\rangle = \delta_{ij}\}} \right] \tag{2.77}$$

$$0 = -\hat{H}^{\mathrm{KS}}|\psi_i\rangle + \sum_j \lambda_{ij}|\psi_j\rangle \tag{2.78}$$

where Equation 2.77 corresponds to the classical Newtonian equations of motion for the nuclei, and Equation 2.78 represents the Kohn-Sham eigenvalue problem.

Finally, AIMD simulations may also include temperature control, which can be achieved by coupling a thermostat to the system in order to simulate different thermodynamic ensembles—*e.g.*, canonical (NVT) or isothermal-isobaric (NPT) ensembles. However, the details of thermostats and different ensembles are beyond the scope of this report, and we shall refer the reader to the literature [49, 50] for a more detailed discussion of these topics.

2.6 Computational Implementation in VASP

The computational studies presented in this report were performed using the Vienna Ab initio Simulation Package (VASP). This is a software suited for performing *ab initio* quantum mechanical calculations [51]. It employs DFT as the basic method, albeit it also allows for beyond-DFT methods such as hybrid functionals, many-body perturbation theory (the GW method), and random phase approximation (RPA).

VASP is widely adopted in areas such as materials science, condensed matter physics, and quantum chemistry, owing to its demonstrated accuracy and state-of-the-art methods. Consequently, this section is devoted to explaining some important aspects of VASP, with great emphasis on the computational methods employed in this work.

2.6.1 Pseudopotentials

A crucial step in defining the wavefunctions that describe an atom is the effective modeling of the electron-ion potential. In this context, core electrons represent a challenge since their rapidly oscillating wavefunctions require a denser real-space grid and a larger number of basis functions to be accurately represented.

To circumvent this problem, we first need to consider the fact that core electrons are tightly bound and do not participate in chemical bonding. Therefore, we can treat them as frozen—*i.e.*, keeping them in their ground state and isolated from the solid-state environment—focusing the calculations on valence electrons, as they do participate in chemical bonding and determine the material's properties. Nonetheless, any adopted approximation must adequately account for the influence of core electrons on the valence states, such as screening and exchange interactions.

To this end, we shall introduce the concept of pseudopotentials [52]. This approximation aims to describe the valence electrons without explicitly treating the core states, replacing the strong Coulomb potential near the nucleus with a weaker one. We begin by separating the single-electron states into valence and core states, denoted as $|\psi_v\rangle$ and $|\psi_c\rangle$, respectively. We then define a new set of valence states $|\tilde{\psi}_v\rangle$ through the following relation

$$|\tilde{\psi}_v\rangle = |\psi_v\rangle + \sum_c |\psi_c\rangle\langle\psi_c|\tilde{\psi}_v\rangle \tag{2.79}$$

Applying the single-particle Hamiltonian H^{sp} to the pseudo-valence states, and taking into account that $|\psi_v\rangle$ and $|\psi_c\rangle$ are eigenstates of the single-particle Hamiltonian, we get

$$\begin{aligned} H^{sp}|\tilde{\psi}_v\rangle &= H^{sp}|\psi_v\rangle + \sum_c H^{sp}|\psi_c\rangle\langle\psi_c|\tilde{\psi}_v\rangle \\ &= \epsilon_v|\psi_v\rangle + \sum_c \epsilon_c|\psi_c\rangle\langle\psi_c|\tilde{\psi}_v\rangle \\ &= \epsilon_v|\psi_v\rangle + \sum_c (\epsilon_c - \epsilon_v)|\psi_c\rangle\langle\psi_c|\tilde{\psi}_v\rangle \end{aligned} \tag{2.80}$$

Rearranging the terms, we obtain the following expression

$$\left[H^{sp} + \sum_c (\epsilon_v - \epsilon_c)|\psi_c\rangle\langle\psi_c|\right]|\tilde{\psi}_v\rangle = \epsilon_v|\tilde{\psi}_v\rangle \tag{2.81}$$

Equation 2.81 describes the Hamiltonian of the pseudo-valence states, which we can define as

$$\hat{H}^{ps} = H^{sp} + \sum_c (\epsilon_v - \epsilon_c)|\psi_c\rangle\langle\psi_c| \tag{2.82}$$

The modified potential for these states is called the pseudopotential, given by

$$V^{ps}(\mathbf{r}) = V^{sp} + \sum_c (\epsilon_v - \epsilon_c)|\psi_c\rangle\langle\psi_c| \tag{2.83}$$

We emphasise that the pseudo-valence states associated with Equation 2.81 have the same single-particle energies as the valence states, and they remain orthogonal to the core states. However, they are not strictly normalised, which can introduce minor inaccuracies in practical calculations.

Various pseudopotentials have been proposed to achieve the desired trade-off between accuracy and computational cost. Those pseudopotentials requiring larger cut-off energies—hard pseudopotentials—are better suited for capturing the strong interactions inherent in core electrons. In contrast, soft pseudopotentials are often preferred because they require fewer basis functions, thereby reducing computational cost, albeit at the expense of accuracy.

2.6.2 Projector Augmented-Wave (PAW) Method in VASP

The projector augmented wave (PAW) method—implemented in VASP and used in this work—aims to address some limitations of pseudopotentials by introducing auxiliary functions called projectors, allowing for core and valence electrons to be better represented and enhancing computational efficiency.

However, before we introduce the PAW formalism—at least to the extent necessary for this report—we shall first introduce some important concepts that underpin the PAW method.

2.6.2.1 Key Concepts

- **Brillouin Zone and k-points**

  Crystalline solids are described in real space in terms of a primitive unit cell (PUC) [28], which is the smallest repeating unit capable of generating the entire solid through periodic boundary conditions. It is characterised by a periodic arrangement of lattice points, referred to as the Bravais lattice. All the lattice points are associated with a set of lattice vectors $\mathbf{R}$ formed by all the possible combinations of integer multiples of the primitive lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$

  $$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \quad n_1, n_2, n_3 \in \mathbb{Z} \tag{2.84}$$

  The primitive unit cell is then defined as the volume enclosed by the primitive lattice vectors

  $$\Omega_{\text{PUC}} = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) \tag{2.85}$$

Using this definition, it is possible to represent the entire crystal by modelling solely its unit cell. Additionally, to appropriately describe electronic properties, we must transition to reciprocal space, defined by the reciprocal vectors

$$\mathbf{b}_1 = \frac{2\pi}{\Omega_{\text{PUC}}}(\mathbf{a}_2 \times \mathbf{a}_3), \quad \mathbf{b}_2 = \frac{2\pi}{\Omega_{\text{PUC}}}(\mathbf{a}_3 \times \mathbf{a}_1), \quad \mathbf{b}_3 = \frac{2\pi}{\Omega_{\text{PUC}}}(\mathbf{a}_1 \times \mathbf{a}_2) \tag{2.86}$$

that satisfy the condition $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. Any reciprocal lattice vector $\mathbf{G}$ can then be expressed as

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \quad m_1, m_2, m_3 \in \mathbb{Z} \tag{2.87}$$

The reciprocal space is divided into regions called Brillouin zones (BZs). The first Brillouin zone is defined as the region in reciprocal space closest to the origin (Γ point). Moreover, the volume of the first BZ is given by

$$\Omega_{\text{BZ}} = \frac{(2\pi)^3}{\Omega_{\text{PUC}}} \tag{2.88}$$

which ensures consistency between the real and reciprocal spaces, and proper normalisation when integrating physical quantities over the BZ.

Finally, because electrons in a crystal are subjected to a periodic potential, single-electron wavefunctions follow Bloch's theorem and can be expressed as

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \tag{2.89}$$

where $\mathbf{k}$ is the wave vector inside the first BZ, and $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R})$ is a periodic function of the Bravais lattice. Many quantities—such as the total energy, electronic density and the total density of states—can be computed by integrating functions of $\mathbf{k}$ over the Brillouin zone,

$$\langle A \rangle = \frac{\Omega_{\text{PUC}}}{(2\pi)^3} \int_{BZ} A(\mathbf{k}) d\mathbf{k} \tag{2.90}$$

Attempting to compute this integral over the entire BZ would require an infinite number of $\mathbf{k}$ points, which is intractable in practice. Instead, we can discretise the BZ into a finite number of $\mathbf{k}$ points. In VASP, it is achieved by following the Monkhorst-Pack [53] scheme,

$$\mathbf{k} = \sum_{i=3}^{3} \frac{n_i + s_i + \frac{1-N_i}{2}}{N_1} \mathbf{b}_i \tag{2.91}$$

where n_i are indices representing the subdivisions along each reciprocal lattice vector, s_i is an optional shift in terms of subdivisions, and N_i is the total number of subdivisions along the $\mathbf{b}_i$ direction.

Ultimately, the Monkhorst-Pack scheme allows us to generate a uniform and symmetrical grid of $\mathbf{k}$ points in the BZ, and, consequently, to compute integrals over the BZ by summing

over a finite number of **k** points.

- **Plane-Wave Expansion and Cut-off Energy**

  Following from our previous discussion—where we expressed the single-electron wavefunction as a combination of a periodic function and a plane wave—we can expand $u_{\mathbf{k}}(\mathbf{r})$ in terms of a set of plane waves

  $$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{2.92}$$

  Combining Equations 2.89 and 2.92, gives

  $$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{2.93}$$

  where $C_{\mathbf{k}+\mathbf{G}}$ are the corresponding Fourier coefficients associated to the plane wave of wavevector $\mathbf{k}+\mathbf{G}$.

  Noticeably, Equation 2.93 introduces a new complexity in the calculations—the evaluation of a single k-point requires the summation over an infinite number of possible **G** vectors. However, the interpretation of Equation 2.93 as solutions to the Schrödinger equation allows for the kinetic energy to be expressed as

  $$E_{\text{kin}} = \frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \tag{2.94}$$

  In practice, lower energy solutions are more physically relevant, thereby allowing us to truncate the infinite summation by including only plane waves whose kinetic energy does not exceed a certain cut-off energy E_{cut}

  $$\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq E_{\text{cut}} \tag{2.95}$$

  This gives a finite expansion

  $$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{|\mathbf{k}+\mathbf{G}|^2/2 \leq E_{\text{cut}}} C_{\mathbf{k}+\mathbf{G}}\, e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{2.96}$$

  Finally, in VASP, we often set the cut-off energy based on a convergence criterion of the total energy of the system

  $$\Delta E_{tot} < 1\text{meV/atom} \tag{2.97}$$

  ensuring a good description of the electronic structure of the system and keeping the calculations computationally feasible.

2.6.2.2 Projector Augmented-Wave (PAW) Method

Having explored the base concepts of the Projector Augmented-Wave (PAW) method—regarding its implementation in VASP—we can transition to the PAW formalism itself. P. E. Blöchl [54] first introduced the PAW method as a generalisation of the pseudopotential method and linear augmented plane-wave method. Later on, it was further refined by G. Kresse and J. Joubert [55], providing a formal relationship between ultra-soft pseudopotentials and the PAW method.

The PAW method addresses the problem of describing the valence states with high accuracy while accounting for the large variations in the all-electron (AE) wavefunction near the atomic core. It does so by introducing a transformation that maps pseudo (PS) wavefunctions—smooth and computationally efficient—onto their corresponding AE counterparts. Thereby, any AE Kohn-Sham wavefunction Ψ_n can be recovered from a PS wavefunction $\tilde{\Psi}_n$ by means of a linear tranformation

$$|\Psi_n\rangle = |\tilde{\Psi}_n\rangle + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle\right) \langle\tilde{p}_i|\tilde{\Psi}_n\rangle \tag{2.98}$$

where important elements are to be highlighted:

- AE partial wavefunctions $|\phi_i\rangle$, obtained from a reference atom.
- PS partial wavefunctions $|\tilde{\phi}_i\rangle$, which match the AE wavefunctions outside a core radius r_c
- Projector functions $|\tilde{p}_i\rangle$, which quantify the contribution of the difference $|\phi_i\rangle - |\tilde{\phi}_i\rangle$ to be added to the PS wavefunction in order to recover the AE wavefunction. They fulfil the biorthonormality condition $\langle\tilde{p}_i|\tilde{\phi}_j\rangle = \delta_{ij}$ within the augmented region, ultimately leading to the completeness relation $\sum_i |\tilde{\phi}_i\rangle\langle\tilde{p}_i| = \mathbb{1}$.

Additionally, in the PAW method, the total AE electronic density is expressed as

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + n^1(\mathbf{r}) - \tilde{n}^1(\mathbf{r}) \tag{2.99}$$

where $\tilde{n}$ is the PS valence electronic density, n^1 is the AE one-center electronic density accounting for the contribution of the valence states inside the augmentation region, and $\tilde{n}^1$ is the corresponding PS one-center electronic density.

Finally, the PAW method helps us to recover the true AE electronic density by correcting the PS electronic density within the augmentation region. As a result, many electronic structure properties—such as the total energy, forces, and stresses—can be accurately evaluated.

2.6.3 Equation of State (EOS)

To study the mechanical properties of bulk materials, it is essential to establish a relationship between observables—such as the ground-state total energy—and macroscopic variables like volume or pressure. In this context, the third-order Birch-Murnaghan equation of state [56, 57] provides a means to relate the total energy to the change in volume in the system. By fitting energy-volume data to this equation, one can obtain important properties, including equilibrium volume and energy, the bulk modulus, and its pressure derivative. This equation is given by

$$E(V) = E_0 + \frac{9}{16} V_0 B_0 \left\{ \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^3 B_0' + \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{2/3} \right] \right\} \quad (2.100)$$

where E_0, V_0, B_0, and B_0' are the equilibrium energy, volume, bulk modulus, and its pressure derivative, respectively.

Equation 2.100 is utilised in this work to report the aforementioned mechanical properties, which are essential for understanding the material's mechanical response under compression or expansion.

2.6.4 Machine Learning Force Fields (MLFFs)

In an earlier discussion, we introduced the concept of *ab initio* molecular dynamics (AIMD) and highlighted its importance for determining the dynamical properties of a material. Nonetheless—even though VASP is highly optimised for these calculations—AIMD methods remain computationally expensive, restricting their applicability to small simulation cells and short simulation times.

In this regard, machine learning force fields (MLFFs) represent a promising alternative. These models are trained on accurate *ab initio* datasets, enabling them to learn the underlying potential energy and produce fast and accurate predictions of atomic energies, forces, and stresses. MLFFs can drastically reduce computational cost while minimising any human intervention and expertise required for the force field construction.

In VASP, MLFFs are implemented following an on-the-fly machine learning algorithm [58] (see Figure 2.2). To construct the machine-learning force field, several structure datasets are required. A structure dataset must define the Bravais lattice, atomic positions, total energy, forces, and the stress tensor computed from first principles (FP). Given the datasets, local configurations around an atom are identified and mapped onto a set of descriptors [59], describing the local environment around each atom in terms of an atomic distribution function.

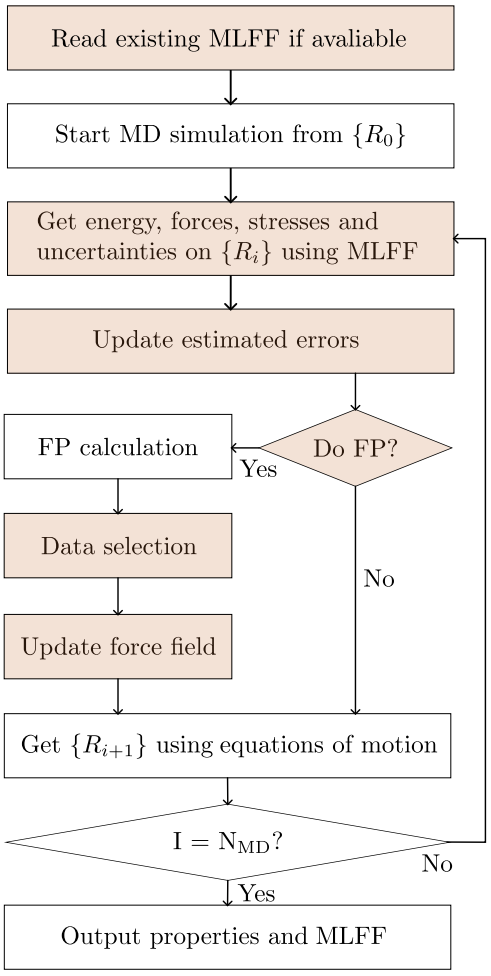

Figure 2.2: On-the-fly force field generation pipeline in VASP [58]. First, the algorithm reads the existing MLFF if available; otherwise, it generates a new one. If accurate enough, a new structure is generated using the force field; otherwise, a first-principles calculation is performed. If the predicted uncertainty is too large, the new structure is added to the dataset, and the force field is retrained. This oscillating process between training and prediction continues until the total number of ionic steps specified in the setup is reached.

$$\rho_i^{(2)}(r) = \frac{1}{4\pi} \int \rho(r\hat{\mathbf{r}}) d\hat{\mathbf{r}} \tag{2.101}$$

$$\rho_i^{(3)}(r, s, \theta) = \iint d\hat{\mathbf{r}} d\hat{\mathbf{s}}\, \delta(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}} - \cos\theta) \sum_{j\neq 1}^{N_a} \sum_{k\neq i,j}^{N_a} \tilde{\rho}_{ij}(r\hat{\mathbf{r}}) \tilde{\rho}_{ik}(s\hat{\mathbf{s}}) \tag{2.102}$$

Equation 2.101 defines the two-body descriptor as the probability of finding an atom $j (j \neq i)$ at a distance r from atom i. Conversely, Equation 2.102 is known as the three-body descriptor and describes the probability to find an atom $j (j \neq i)$ at a distance r from atom i and another atom $k (k \neq i, j)$ at a distance s from atom i spanning an angle $\angle kij$ between them. These descriptors are then used to construct the local potential energy functionals $U_i = F[\rho_i^{(2)}, \rho_i^{(3)}]$ upon which the force field is constructed.

The force field is then generated during AIMD simulations following the steps below:

1. The machine predicts the energy, forces, stress tensor, and uncertainty for the current atomic configuration using the available force field.
2. The algorithm decided whether to perform an FP calculation or not. This decision is based on the uncertainty of the prediction. If the uncertainty exceeds a predefined threshold, the machine proceeds to step 3; otherwise, it continues to step 5.
3. The FP calculation is performed for the current structure, and the obtained dataset is then stored as a candidate for the training dataset.
4. If the number of candidate structures reaches a certain threshold, or if the uncertainty becomes too large, the algorithm updates the training set and generates a new force field.
5. Update the atomic positions and velocities. If the force field is not accurate enough, the FP energy, forces, and stress tensor are used. Otherwise, those predicted by the force field are used. Afterwards, the algorithm returns to step 1 until the total number of ionic steps is reached.

Ultimately, MLFFs provide a robust and efficient method for performing MD simulations. It combines the accuracy of AIMD with the speed of classical MD, allowing for large and complex condensed matter systems to be studied with remarkable accuracy and efficiency.

2.7 Density Functional Tight Binding (DFTB+)

Density Functional Tight Binding (DFTB) implemented in the DFTB+ code [60] is a semi-empirical method derived from the Kohn-Sham DFT by performing a Taylor expansion of the total energy functional around a reference electronic density. This method is computationally efficient, allowing for simulations involving large systems and long timescales with reasonable accuracy, and is significantly faster compared to *ab initio* methods. In this work, the GFN1-xTB (Geometry, Frequency, Noncovalent-extended Tight Binding) method was used to perform an initial relaxation prior to the full structure relaxation using VASP.

Chapter 3

Methodology

This chapter outlines the methodology followed in this work to perform the computational simulations of calcium silicate hydrates (C–S–H). All the calculations were carried out using Density Functional Tight Binding (DFTB+)—primarily in the initial stages—and subsequently with the Vienna Ab-initio Simulation Package (VASP). All the computational parameters—`INCAR` files—used for core DFT calculations are detailed in Appendix B.

We begin by describing the initial C–S–H structure, followed by the details of the VASP workflow, emphasising the self-consistent field (SCF) cycle. We then present the main input and output files required for the simulations. Next, we discuss the structure relaxation procedure, which includes an initial relaxation using DFTB+ followed by a full structure relaxation with VASP. Finally, we discuss the generation of the machine learning force field (MLFF), covering the training, refinement, and testing phases.

3.1 Initial C–S–H Structure

The structure used for our investigations of calcium silicate hydrates (C–S–H) is the molecular model proposed by Pellenq *et al.* [13]. This model was constructed with the chemical composition as the overriding constraint. As such, the model has a calcium/silicon ratio (C/S) of 1.7, and a density of 2.6 g/cm^3; consistent with experimental observations.

It was derived from a monoclinic periodic cell of dry tobermorite, from which SiO_2 groups were removed to achieve an experimental C/S ratio. Thereafter, the structure was relaxed using a core-shell potential model at 0 K. Finally, a Grand Canonical Monte Carlo simulation of water adsorption was carried out at 300 K, reporting a chemical composition of $(CaO)_{1.65}(SiO_2)(H_2O)_{1.75}$.

The model, shown in Figure 3.1, contains 99 calcium (Ca), 60 silicon (Si), 323 oxygen (O), and 208 hydrogen (H) atoms, making a total of 690 atoms. It is noteworthy that the model is not

regarded as a perfect representation of C–S–H, but rather as a good approximation that captures the essential features of cement hydrates.

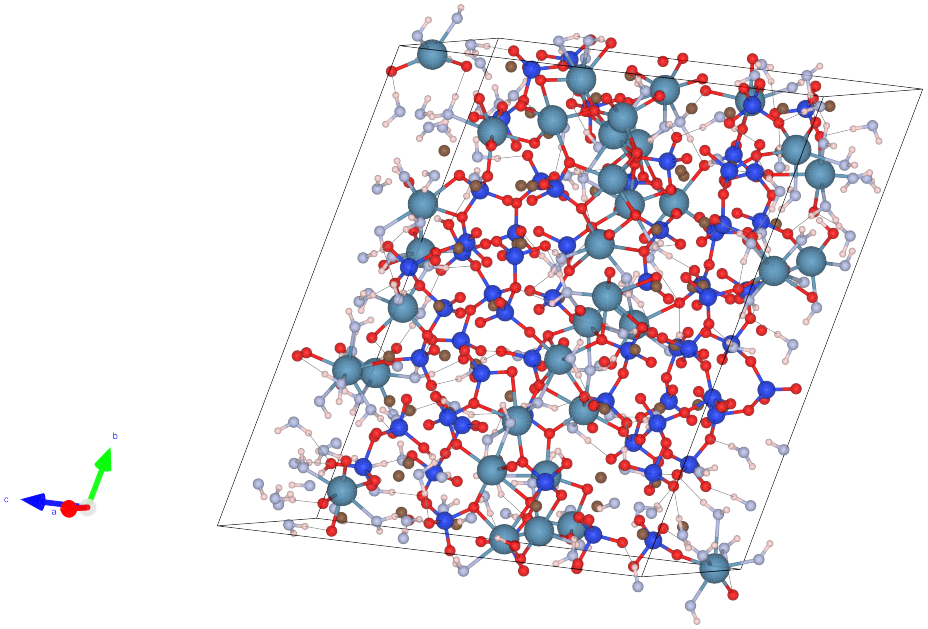

Figure 3.1: Molecular model of C–S–H proposed by Ref. [13]. Lavender and white spheres are oxygen and hydrogen from water molecules, respectively; light blue and brown spheres are inter- and intra-layer calcium ions, respectively; electric blue and red spheres are silicon and oxygen atoms from silica tetrahedra, respectively.

3.2 VASP Workflow

Most of our calculations were performed using the Vienna Ab Initio Simulation Package (VASP). A central part of the VASP workflow is the self-consistent field (SCF) cycle, illustrated in Figure 3.2. This cycle is essential for structure relaxation, *ab initio* molecular dynamics (AIMD) simulations, and other DFT calculations. We hereby present the main procedure of the SCF cycle:

- At the beginning of the cycle, a trial electronic density is generated—either from a previous calculation or from an initial guess.
- The algorithm then proceeds to construct the effective potential, defined as the sum of the Hartree, external, and exchange-correlation potentials. The latter is specified by the user (e.g., PBEsol, HSE06).
- VASP then solves the Kohn-Sham equation, generating a new set of single-electron wave-functions at each iteration.

- A new electronic density is calculated from the wavefunctions. This process repeats until self-consistency is achieved—*i.e.*, the total energy difference between consecutive iterations falls below a predefined tolerance. The user sets this value, and for our calculations, values of `EDIFF=1E-5` and `EDIFFG=-0.02` were used.

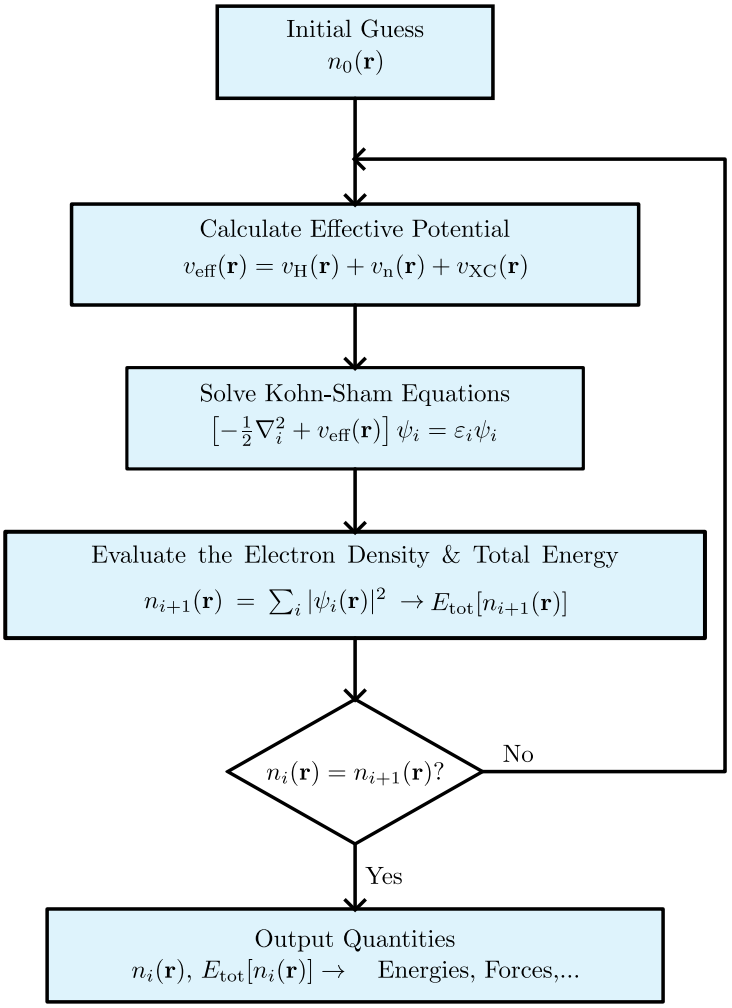

Figure 3.2: Self-consistent field (SFC) cycle in VASP for DFT calculations adapted from Ref. [27]. The entire cycle starts with an initial guess of the electronic density $n_0(\mathbf{r})$, which is then used to calculate the effective potential $v_{\text{eff}}(\mathbf{r})$. Then, the resulting potential is used to solve the Kohn-Sham equations, from which single-electron wavefunctions $\psi_i(\mathbf{r})$ are obtained. Consequently, the new electronic density $n_{i+1}(\mathbf{r})$ is calculated. Should the old and new densities be close enough—up to a predefined threshold—the cycle stops, and the final electronic density is used to calculate the energies, forces, and stress tensor of the system. Otherwise, the cycle repeats itself until convergence is achieved.

3.3 VASP Input & Output Files

The VASP input and output files are essential for our calculations. On the one hand, the input files contain necessary information—such as the initial structure, exchange-correlation functionals, PAW pseudopotentials, k-point grid, and convergence criteria—that guide the different simulations.

On the other hand, the output files contain the results of the different simulations, such as the total energy, forces, stress tensor, and the fully relaxed structure.

This section provides an overview of the required input files for VASP calculations, as well as some relevant output files. For a detailed and rather technical description of the files described herein, we refer the reader to the VASP manual [61].

3.3.1 Input Files

3.3.1.1 INCAR

The `INCAR` file (see Figure B.1) defines the computational parameters in VASP and specifies the type of calculation to be performed. Each simulation stage—such as structure optimisation, Density of States (DOS) calculations, AIMD simulations, or MLFF training—is defined by a specific set of INCAR tags.

```
                    GENERAL
SYSTEM    = C–S–H     # System name
PREC      = Accurate  # Precision level
              ELECTRONIC OPTIMIZATION
ENCUT     = 800       # Plane-wave cutoff (eV)
LREAL     = Auto      # Real-space projection
ISMEAR    = 0         # Smearing method
SIGMA     = 0.05      # Smearing width (eV)
ALGO      = F         # Electronic minimization algorithm
AMIX      = 0.1       # Charge density mixing parameter (damping)
        EXCHANGE-CORRELATION / FUNCTIONAL
GGA       = PS        # PBEsol functional
IVDW      = 11        # DFT-D3(zero) vdW correction
LASPH     = .TRUE.    # Non-spherical contributions
LMAXMIX   = 4         # Maximum l for charge mixing
              CHARGE & WAVEFUNCTION
LCHARG    = F         # Do not write CHGCAR
                IONIC RELAXATION
NELMIN    = 4         # Minimum SCF steps
MAXMIX    = 40        # Maximum mixing steps
IBRION    = 2         # Ionic relaxation algorithm
ISIF      = 3         # Relax ions + cell shape + volume
NSW       = 700       # Maximum ionic steps
EDIFFG    = -0.02     # Convergence criterion (eV/Å)
ADDGRID   = T         # Additional grid for accuracy
```

Figure 3.3: Example of an `INCAR` file used for structure relaxation of C–S–H. This file specifies the optimisation algorithm, force convergence criteria, exchange-correlation functional (PBEsol) and ionic relaxation parameters. Depending on the type of calculation to be performed, different tags may be added or removed.

These parameters control convergence thresholds, exchange-correlation functionals, long-range corrections, ensemble choices, and other simulation parameters. If not specified by the user, VASP uses default values. Nonetheless, for reliable and reproducible results, the main parameters must be tailored to the system and the type of calculation.

3.3.1.2 POSCAR

The `POSCAR` (see Figure 3.4) file provides the actual structure to be studied. It is subdivided into several sections, each one providing specific information about the system.

```
Ca Si O H
1.0
 13.18335946    0.18445997    0.00755401
-16.45244030   24.21622147   -0.00875423
  1.20664987   -0.82375620   23.18729854
Ca Si O H
99 60 323 208
Direct
 0.38821570    0.10613519    0.29312228
 0.37259751    0.56538816    0.26881558
 0.37469040    0.31600944    0.23882914
 0.35542617    0.78384113    0.38748504
 0.91347068    0.11233954    0.31634176
 0.90355777    0.57838562    0.23567872
 0.87836092    0.32109531    0.25338100
 0.84529168    0.81264469    0.29236748
 0.11769176    0.02588977    0.65372010
```

Figure 3.4: Unit cell structure in fractional coordinates for the C–S–H (Calcium Silicate Hydrates) system. The lattice vectors, atomic species (99 Ca, 60 Si, 323 O, 208 H), and the first 9 atomic positions are shown. All coordinates are expressed in direct (fractional) form.

The first line contains a comment specifying the name of the system or a brief description of it. Lines 2-4 provide the scaling factor and the corresponding lattice vectors. The actual lattice vectors are obtained by multiplying the scaling factor (line 2) by the numbers in lines 3-5. Lines 6-7 specify the atomic species as well as the number of ions of each species. Finally, lines 9-onwards provide the ionic positions in angstroms.

3.3.1.3 KPOINTS

Defining the k-point grid is an essential and one of the first steps when performing DFT calculations, as the accuracy and convergence of the results depend on it. Figure 3.5 illustrates the `KPOINTS` file used in this work. The specified mesh is a $1 \times 1 \times 1$ Gamma-centered grid, obtained after a convergence test.

```
C–S–H kpoints
0
Gamma
    1 1 1
    0 0 0
```

Figure 3.5: C–S–H k-point grid centered at the Gamma point. The values "1 1 1" define the grid dimensions in the x, y, and z directions. For large systems, a Gamma-centered grid is enough to achieve convergence.

3.3.1.4 POTCAR

The `POTCAR` file contains the PAW pseudopotentials for each atomic species in the system. As such, it defines how valence electrons interact with the atomic cores. It is constructed by concatenating the individual POTCAR files for each species into a single file. In our work, we employed the following PAW pseudopotentials from the PAW_PBE library:

- Ca: Ca_pv ([Ar] $4s^2$)
- Si: Si ([Ne] $3s^2$ $3p^2$)
- O: O ([He] $2s^2$ $2p^4$)
- H: H ($1s^1$)

3.3.2 Output Files

These are the primary output files generated upon finishing an FP calculation in VASP. They provide core information regarding the performance of the calculations, documenting the simulation and providing the basis for further analysis.

3.3.2.1 OUTCAR

The `OUTCAR` file is a comprehensive output file that contains detailed information about the VASP calculation. It includes a summary of the input parameters, the evolution of the SCF cycle, the total energy, forces on the atoms, and the stress tensor.

3.3.2.2 CONTCAR

The `CONTCAR` file records the final atomic positions and lattice vectors after a structure relaxation or optimisation. Additionally, this file may also contain atomic velocities and predictor-corrector information if it was written during an AIMD simulation. It has a compatible format with the `POSCAR` file, making it possible to use it as an input structure for subsequent calculations.

3.3.2.3 DOSCAR

The `DOSCAR` file stores the Density of States (DOS) and integrated DOS, expressed in states/eV and cumulative number of states, respectively. This data is beneficial for analysing the electronic properties of the system, and understanding features such as the band gap and the distribution of states across the valence and conduction bands.

3.3.2.4 OSZICAR

The `OSZICAR` file records a summary of the electronic and ionic iterations during a DFT calculation. It allows the user to monitor the progress of the SCF cycle convergence, visualise changes in the total energy, and follow the evolution of the ionic relaxation process.

3.3.2.5 ML_ABN

The `ML_ABN` file contains the training dataset collected during an on-the-fly MLFF training process. As previously described, the MLFF is trained together with an AIMD simulation, where atomic configurations are sampled. Representative configurations are then written to this file, which can be reused to continue the training by renaming it to a `ML_AB` file.

3.3.2.6 ML_FFN

The `ML_FFN` file is a binary file that stores the trained machine learning force field (MLFF) model at the end of the training phase. It contains the model parameters, such as weights and hyperparameters, that define the MLFF. The model can be used for prediction or further refinement by renaming it to an `ML_FF` file.

3.4 Strucure Relaxation

Relaxing the C–S–H structure is a crucial step towards obtaining equilibrium properties of this material. This process involves minimising both the forces on atoms and the total energy of the system, leading to a stable configuration. In this work, we performed structure relaxation in two stages: an initial relaxation was conducted using DFTB+, which provided a good starting point for VASP and reduced the computational cost of the full structure relaxation. The second stage consists of a full structure relaxation using VASP. Here we outline both stages of the structure relaxation process.

3.4.1 Initial Relaxation with DFTB+

Given the large size of the C–S–H structure, it was necessary to perform a preliminary relaxation using the GFN1-xTB method implemented in DFTB+. For this step, we first conducted a k-point convergence test. We then plugged this value into the relaxation script in order to run the structure relaxation. This trick allows us to take our structure closer to its equilibrium configuration, without spending too much time and computational power to do so. This approach is valid because we are not using the final structure as our actual optimised structure, but rather as a means to reduce the computation time required for full structure relaxation in VASP.

3.4.2 Full Structure Relaxation with VASP

After the rough approximation provided by DFTB+, a full structure relaxation is performed using VASP. To achieve this, we first perform a cut-off energy convergence test, followed by a k-point convergence test. Thereafter, we define the k-point mesh in the `KPOINTS` file, and the cut-off energy in the `INCAR` file, where we also specified the PBEsol functional and a force convergence criteria of 0.02 eV/Å. Once the structure has been fully relaxed, it is used to study the Density of States (DOS) and to train the machine learning force field.

3.5 Machine Learning Force Field Generation

This stage is subdivided into three main phases: training, refinement, and testing. In this section, we describe each one of them in detail.

3.5.1 Training

As previously discussed, the MLFF is generated on-the-fly during an AIMD simulation. To begin, we use the `CONTCAR` file, which contains our relaxed structure, as the input `POSCAR` file for this step. In the `INCAR` file some parameters need to be set: `IBRION=0`, indicates VASP to switch to an AIMD simulation; `NSW=50000` indicates the number of ionic steps; `POTIM=2.0` is the MD time step in fs; `MDALGO=3` tells VASP to use the Langevin thermostat; `TEBEG=400` sets the temperature (in K) at which the simulation is performed, and `ISIF=3` allows for positions, cell shape and volume to be updated.

Finally, `ML_LMLFF=T` and `ML_ISTART=0` tags govern the MLFF training process. The former enables the use of machine learning force fields, and the latter tells VASP to generate a new MLFF from scratch. Although the parameters described herein are the most important, additional tags may need to be set depending on the performance of the training phase.

3.5.2 Refinement

The refinement phase allows for improvements to be made in the generated MLFF by tuning the hyperparameters in the model. To this end, we first generate a set of 50000 structures using the force field, from which we uniformly sample 50 configurations. Then we compute the total energy, forces and stress tensor for each one of them in two separate runs. The first run utilises first principles, whereas the second run is performed using the MLFF model. The corresponding data is then postprocessed and the errors between DFT and MLFF results are evaluated using configuration-wise Root Mean Square Error (RMSE) defined as follows:

$$E^i_{\text{error}} = \frac{E_i^{\text{DFT}} - E_i^{\text{MLFF}}}{\text{N}_a} \tag{3.1a}$$

$$F^i_{\text{RMSE}} = \sqrt{\frac{1}{3\text{N}_a}\sum_{j=1}^{\text{N}_a}\sum_{k=1}^{3}\left(\mathbf{F}_{ijk}^{\text{DFT}} - \mathbf{F}_{ijk}^{\text{MLFF}}\right)^2} \tag{3.1b}$$

$$S^i_{\text{RMSE}} = \sqrt{\frac{1}{6}\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\left(\sigma_{i\alpha\beta}^{\text{DFT}} - \sigma_{i\alpha\beta}^{\text{MLFF}}\right)^2} \tag{3.1c}$$

where i indicates the configuration index, N_a is the number of atoms in the system, j is the atom index, k corresponds to the Cartesian components of the forces, and α, β are the Cartesian indices of the stress tensor. Note that the energy error does not correspond to a RMSE, but rather to a per-atom energy error. Likewise, global RMSE values for the entire set of configurations are

computed as follows:

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{\mathrm{N}_s}\sum_{i=1}^{\mathrm{N}_s}\left(E_i^{\mathrm{DFT}} - E_i^{\mathrm{MLFF}}\right)^2} \tag{3.2a}$$

$$F_{\mathrm{RMSE}} = \sqrt{\frac{1}{\mathrm{N}_s}\sum_{i=1}^{\mathrm{N}_s}\left(F_{\mathrm{RMSE}}^i\right)^2} \tag{3.2b}$$

$$S_{\mathrm{RMSE}} = \sqrt{\frac{1}{\mathrm{N}_s}\sum_{i=1}^{N_s}\left(S_{\mathrm{RMSE}}^i\right)^2} \tag{3.2c}$$

where N_s is the total number of sampled configurations. These results are significant as they provide the means to evaluate the performance of our force field.

Afterwards, we conduct a hyperparameter optimisation in order to improve the performance of the force field. It is noteworthy that VASP provides various hyperparameters that we can optimise; nevertheless, in this work, only two hyperparameters were considered—the two and three-body descriptors—as they directly affect the accuracy of the force field. In this regard, we set `ML_MODE=refit` and `ML_RCUT1=#` in the `INCAR` file. The latter parameter corresponds to the radial descriptor (given in Å), and is to be modified accordingly to a reasonable range. Thereafter, we use the resulting MLFF file to evaluate the performance of the refitted force field for the given `RCUT1`. We achieve this by setting `IBRION=-1` and `ML_MODE=run` in the `INCAR` file. This calculation will return the RMSE for energies, forces, and the stress tensor as a function of the descriptor. The same process is then applied to the angular descriptor `RCUT2`, and the optimal hyperparameters are chosen to minimise the errors.

3.5.3 Testing

Following the refinement process, we can utilise the MLFF model to conduct various simulations, including AIMD simulations, structure relaxation, and other DFT calculations. In this work, we used the generated force field to compute the equation of state (EOS) of C–S–H. Additionally, we also performed a simulated annealing process to obtain a more stable structure and computed its corresponding EOS as well. Finally, various MD simulations were conducted at temperatures of 200, 250, 300, 350, and 400 K to study the transferability of the force field as well as the expansion coefficient of C–S–H.

Chapter 4

Results & Discussion

We hereby present the results of the computational investigations of C–S–H performed throughout this work. The results are organised into the following main sections: (Section 4.1) Structure relaxation and Density of States (DOS) calculations, (Section 4.2) MLFF generation, (Section 4.3) Thermodynamic properties of C–S–H, and (Section 4.4) Transferability of MLFFs and thermal expansion coefficient of C–S–H.

4.1 Structure Relaxation and Density of States (DOS) calculations

The initial stage of our computational study on C–S–H focused on establishing optimal parameters for VASP calculations. In particular, we determined the optimal plane-wave cut-off energy and Brillouin zone sampling (k-point mesh) through convergence tests, ensuring a balance between accuracy and computational cost. Following structural relaxation with these parameters, the DOS was computed to assess the insulating (ceramic) nature of C–S–H. The results of these convergence analyses and DOS calculations are presented in this section.

4.1.1 Cut-off Energy

The cut-off energy is essential in VASP calculations as it determines the maximum kinetic energy of the plane-wave basis set. As such, it ensures the completeness of this basis set and the accurate description of the electronic structure of the system. The cut-off energy convergence test presented herein was conducted using the PBEsol functional. As shown in Figure 4.1, an optimal E_{cut} is achieved at 800 eV, where the convergence criteria of 1 meV/atom is satisfied. The notably high cut-off energy can be attributed to the presence of heavy elements in the C–S–H structure, such as Calcium (Ca) and Silicon (Si), which require a high cut-off energy value [62]. All the subsequent VASP calculations were performed using this cut-off energy.

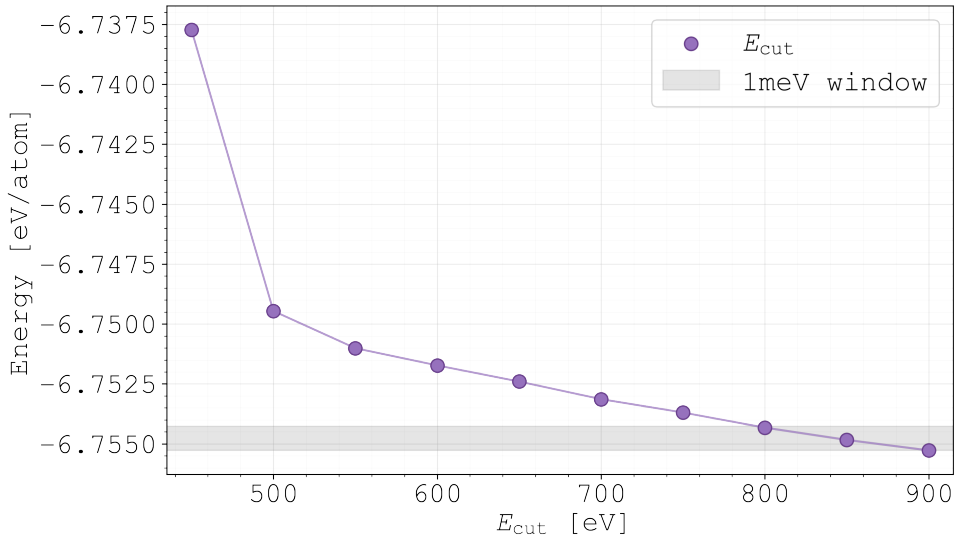

Figure 4.1: Cut-off energy convergence test performed in VASP employing the PBEsol functional for $300 \leq E_{\text{cut}} \leq 900$ eV. The $\Delta E = 1\text{meV/atom}$ convergence criteria is achieved at $E_{\text{cut}} = 800$ eV

4.1.2 k-point convergence

Two k-point convergence tests were performed: the first before the initial relaxation of the C–S–H structure using DFTB+ (GFN1-xTB method), and the second before the full relaxation in VASP (PBEsol functional). In both cases, the convergence criteria of $\Delta E = 1\text{meV/atom}$ was satisfied with a Γ-centered $(1 \times 1 \times 1)$ mesh, corresponding to a k-point spacing of $\Delta k = 0.06$ Å^{-1} (Figure 4.2 and 4.3).

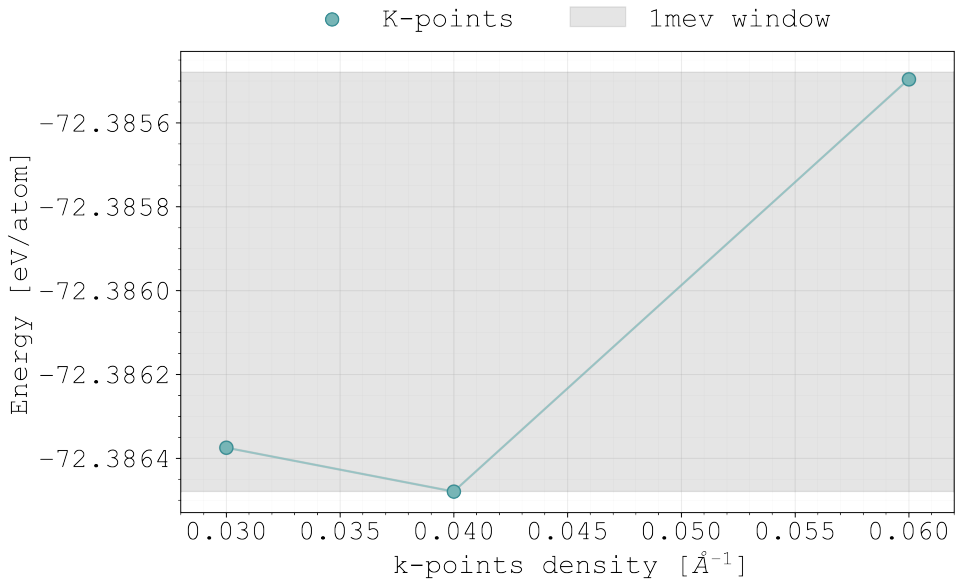

Figure 4.2: k-point convergence test performed in DFTB+ using the GFN1-xTB method for $0.03 \leq \Delta k \leq 0.06$. The $\Delta E = 1\text{meV/atom}$ convergence criteria is achieved at corresponding to a $(1 \times 1 \times 1)$ k-point grid.

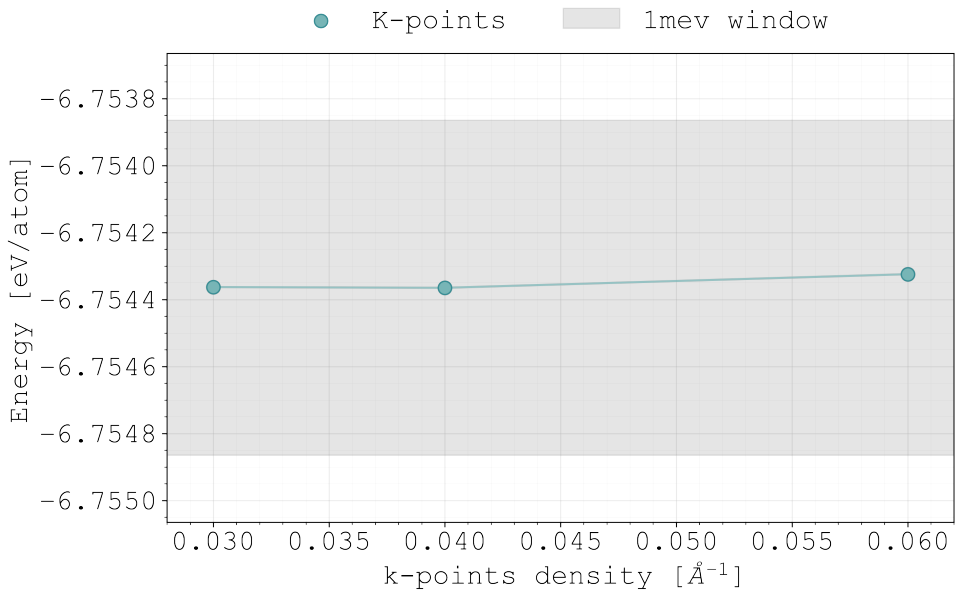

Figure 4.3: k-point convergence test performed in VASP using the PBEsol functional for $0.03 \leq \Delta k \leq 0.06$. The $\Delta E = 1\text{meV/atom}$ convergence criteria is achieved at $\Delta k = 0.06\,\text{Å}^{-1}$ corresponding to a $(1 \times 1 \times 1)$ k-point grid, in agreement with the DFTB+ results.

The coarse mesh is justified by the large simulation cell of C–S–H (690 atoms), which results in a small Brillouin zone, where fine sampling offers no significant improvement to the total energy accuracy [63]. The consistency between DFTB+ and VASP supports the reliability of our choice.

4.1.3 Density of States (DOS)

The electronic Density of St-ates (DOS) of C–S–H was calculated using the HSEsol functional—a hybrid functional combining features of PBEsol and HSE—[64] in VASP, with the relaxed structure obtained from the optimised cut-off energy and k-point mesh. The resulting DOS is presented in Figure 4.4.

The DOS profile exhibits a clear band gap of approximately 4.81 eV between the valence and conduction bands, with no electronic states at the Fermi level region. The absence of electronic states at E_F confirms the insulating nature of C–S–H, consistent with its classification as a ceramic material. The valence band is predominantly populated by O2p and Ca3p states, while the conduction band is mainly composed of Ca3d states in fair agreement with previous computational studies [65].

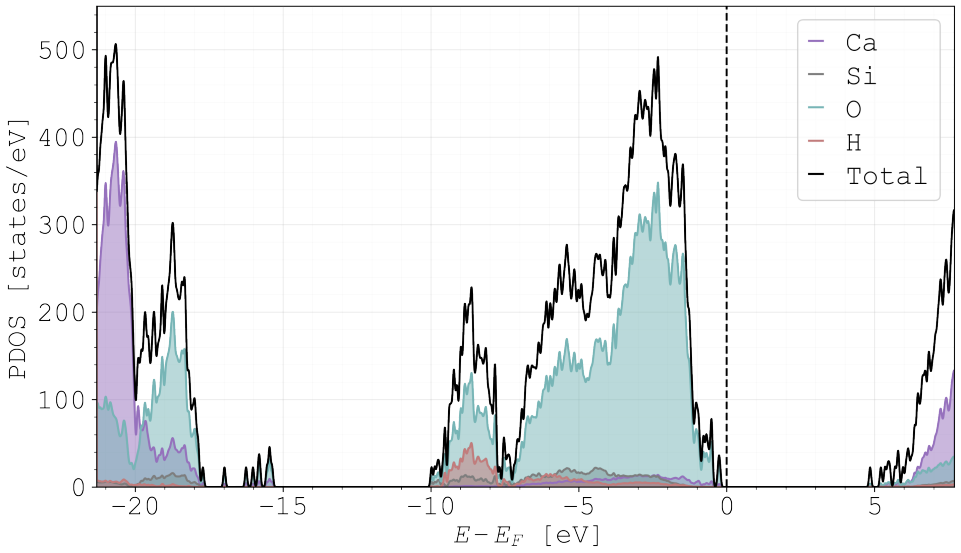

Figure 4.4: Electronic Density of States (DOS) of C–S–H calculated using the HSEsol functional in VASP after full structure relaxation. The Fermi level is set to 0 eV (dashed line), and a band gap of approximately 4.81 eV is observed.

4.2 Machine Learning Force Field (MLFF) Generation

This section is devoted to the core focus of this work: the training, testing and refinement of a machine learning force field (MLFF) for C–S–H. The MLFF was generated on-the-fly within VASP—which employs a Bayesian-learning algorithm to construct a force field on-the-fly during an AIMD simulation [66]. We hereby present the training and testing statistics, as well as the details of the final refined MLFF.

4.2.1 Training

The training phase of the force field consisted of an AIMD simulation performed in VASP using the PBEsol functional, with a time step of 2 fs for a total of 50000 steps (100 ps). Total energy, cell volume and Bayesian error were monitored during the simulation and are reported in Figure 4.5. Initial peaks are observed for all three quantities, primarily due to the thermalisation of the system.

After 10 ps, the total energy stabilises around -4830 eV, with regular fluctuations associated with an increased uncertainty in the force field, whereas the cell volume oscillates around 7500 $Å^3$. On the other hand, the Bayesian error exhibits a general decreasing trend, reflecting the progressive improvement of the force field as more configurations are explored and added to the training set. Regular spikes in the Bayesian error indicate the appearance of configurations that differ substantially from those previously explored. This prompts the algorithm to switch

from prediction to training mode, and the newly identified representative configurations are incorporated into the training set.

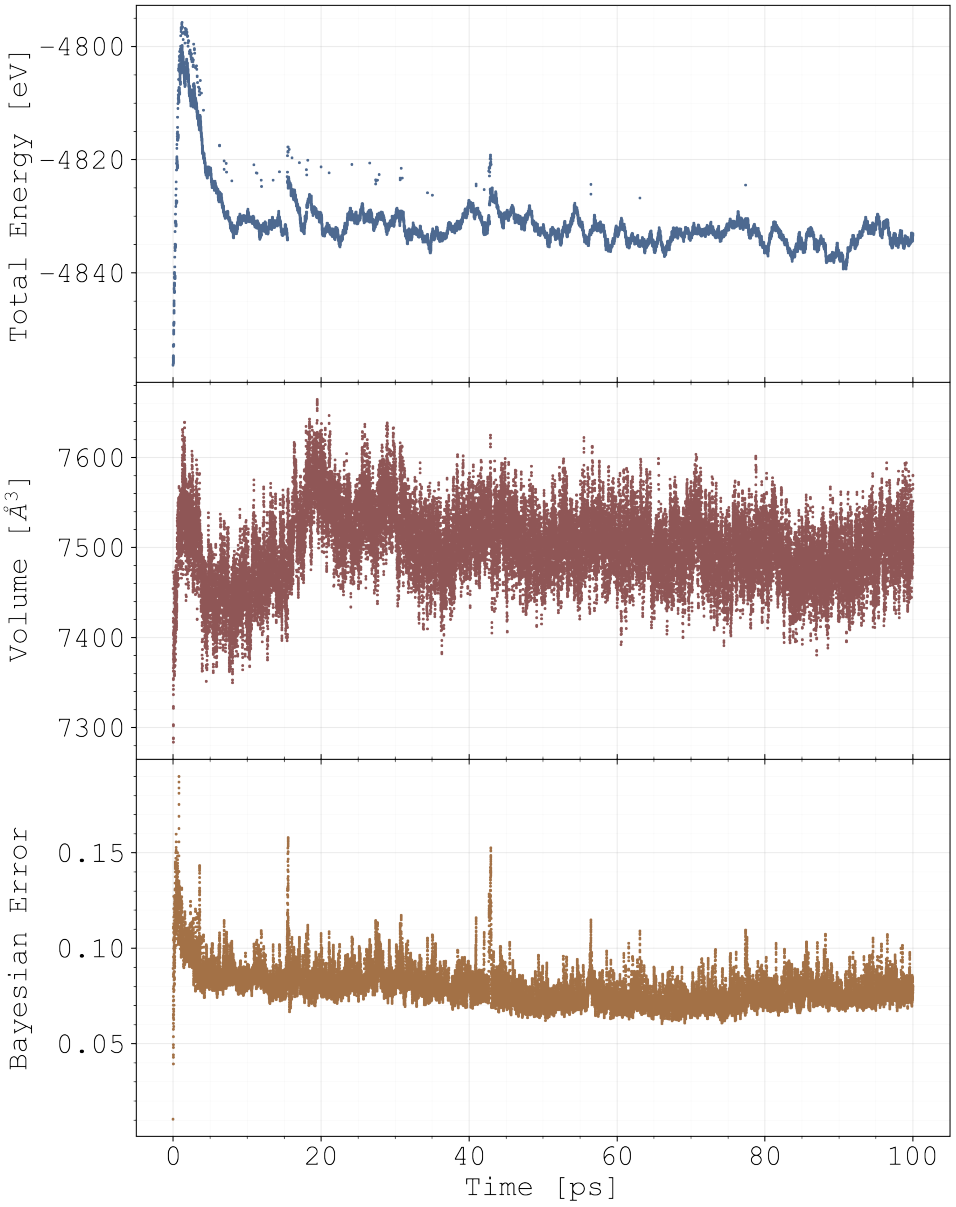

Figure 4.5: Training statistics of the MLFF generated on-the-fly during an AIMD simulation in VASP. The plots show the evolution of the total energy, cell volume and the Bayesian error over a total simulation time of 100 ps.

After 80 ps, we observe no significant fluctuations in the total energy, indicating that the force field has converged to a stable representation of the C–S–H system, as supported by the low Bayesian error. Notably, the final 20 ps of the simulation were performed entirely by the force field in prediction mode, further confirming the stability and reliability of the MLFF.

4.2.2 Evaluation

Following the training phase, we evaluated the performance of the force field in two steps. First, we carried out an MD simulation using the MLFF in prediction mode with a time step of 2 fs for a total simulation time of 100 ps. Second, we randomly selected 50 configurations from the generated trajectory and computed the total energy, forces and stress tensor using both DFT and the MLFF separately. We then computed the errors between the results obtained to quantify the performance of the force field.

Figure 4.6 shows the evolution of the total energy and the cell volume during the MD simulation in prediction mode. No abrupt fluctuations or drifts are observed, with the total energy oscillating around -4833 eV and the cell volume around 7494 Å^3. These results show consistency with the training statistics and reflect the ability of the MLFF to accurately represent the potential energy surface of C–S–H.

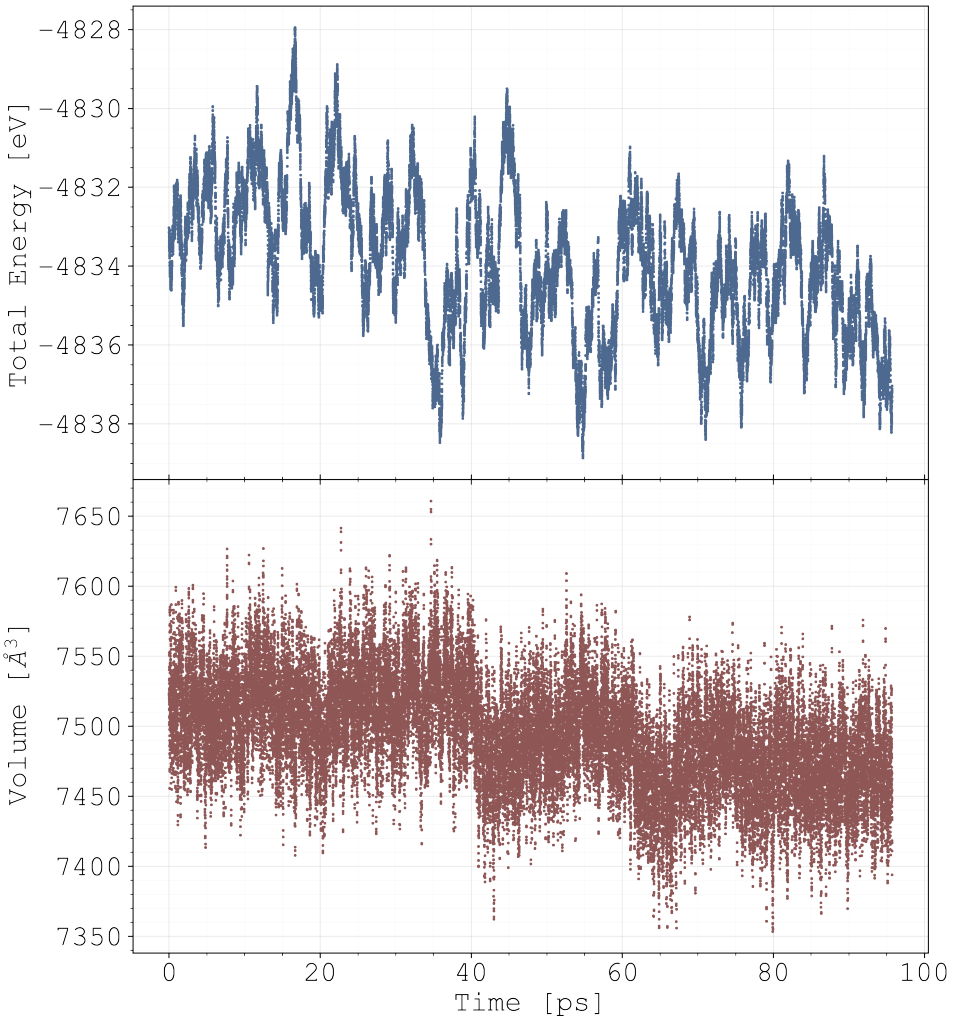

Figure 4.6: Evolution of the total energy and cell volume during an MD simulation of C–S–H using the MLFF in prediction mode over a total simulation time of 100 ps.

On the other hand, the errors between DFT and MLFF predictions for the total energy, forces and stress tensor are reported in Figure 4.7. An RMSE of 12.2 meV/atom is observed for the energy, 223.4 meV/Å for the forces, and 0.961 kbar for the stress tensor. Similar work on C–S–H (C/S=1.7) conducted by Zhu [25] reported an energy RMSE of 6 meV/atom and a force RMSE of 160 meV/Å, indicating that our MLFF comparable in terms of accuracy. As for the stress tensor RMSE, no reference data was found. Finally, the reported errors could be further reduced by refining the MLFF, as discussed in the next subsection.

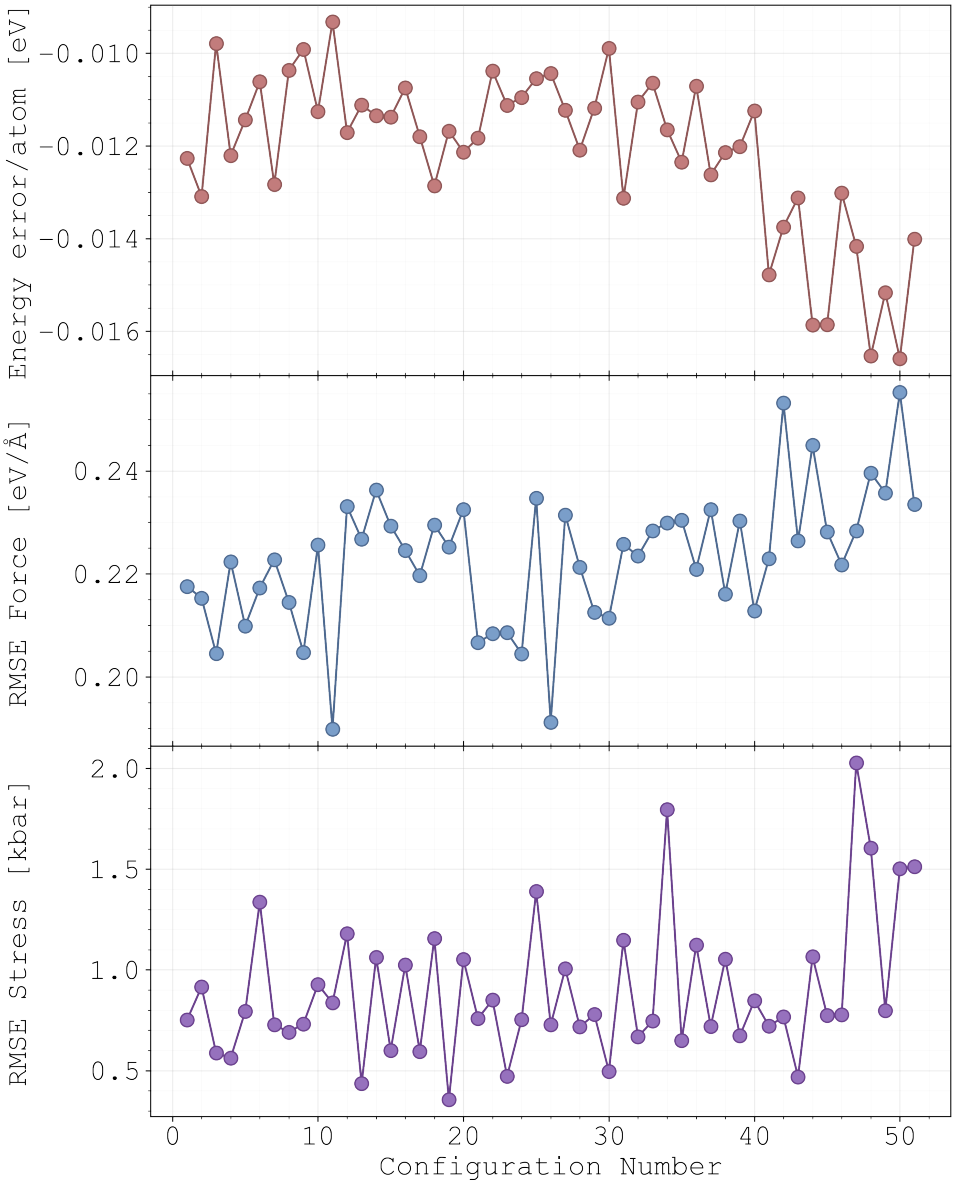

Figure 4.7: Energy error per atom and root-mean-square error (RMSE) for forces and stress tensor of C–S–H between DFT and MLFF predictions, evaluated on 50 configurations randomly selected from an independent set of 50000 configurations generated via MD simulation using the MLFF in prediction mode without refitting.

4.2.3 Refinement

The MLFF refinement involved varying hyperparameters of the force field, running a refitting procedure and evaluating the performance of the refined MLFF. In this work, we focused on the radial and angular descriptors, as they are crucial for the representation of the local interactions of atoms in C–S–H. Figure 4.8 and 4.9 show the error dependence on the radial and angular descriptors, respectively.

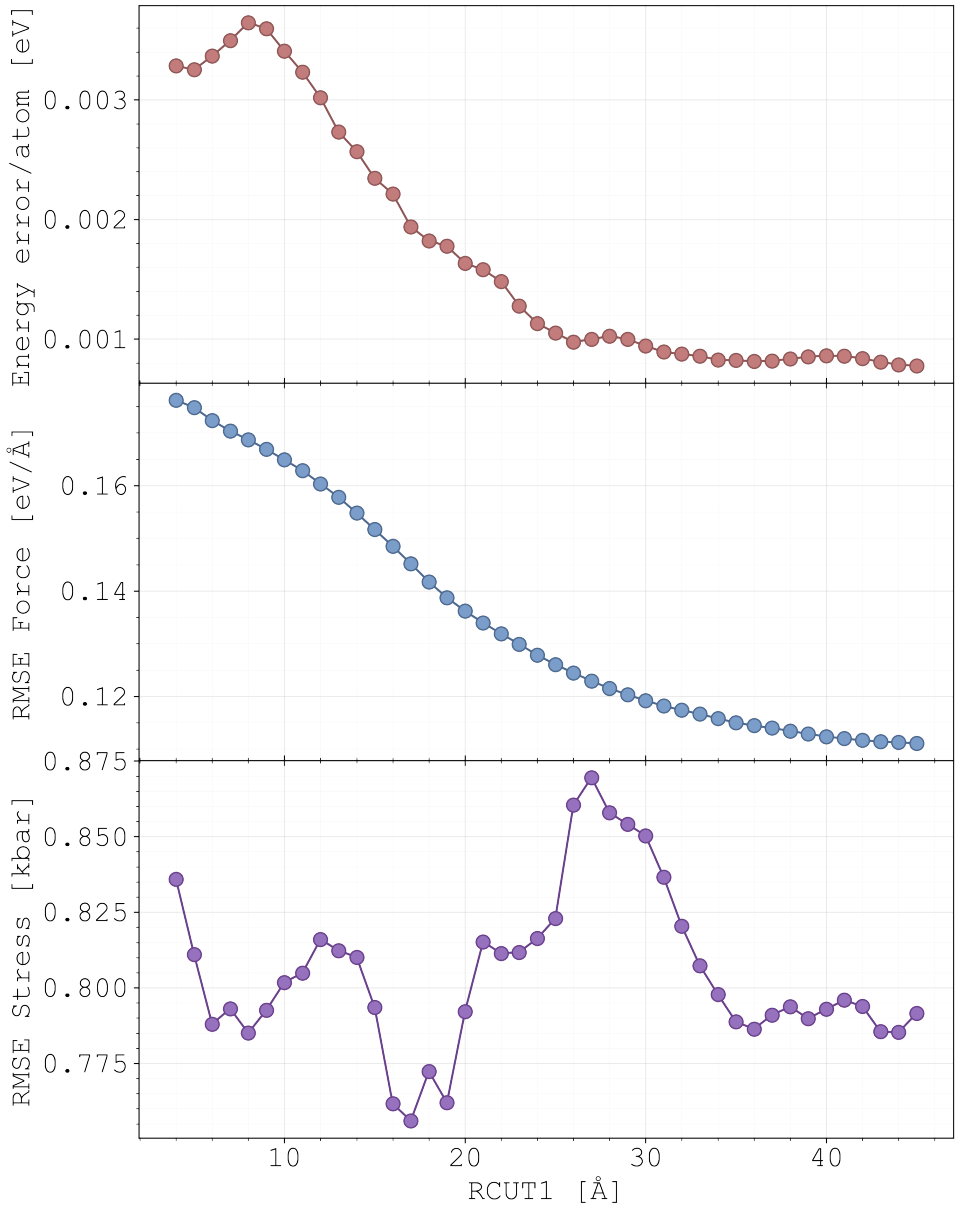

Figure 4.8: Root-mean-square error (RMSE) for the total energy, forces and stress tensor as a function of the radial descriptor (`RCUT1`).

We varied the radial descriptor (`RCUT1`) over the range of 4.0 to 45.0 Å, and the angular descriptor (`RCUT2`) over the range 2.0 to 10.0 Å. With the energy error and the force RMSE monotonically decreasing, no clear minimum is observed for `RCUT1` in the considered range; however, a minimum

stress RMSE occurs at 17.0 Å. Conversely, `RCUT2` exhibits a clear minimum for both the energy error and the force RMSE at 3.0 Å, and for the stress tensor RMSE at 4.0 Å.

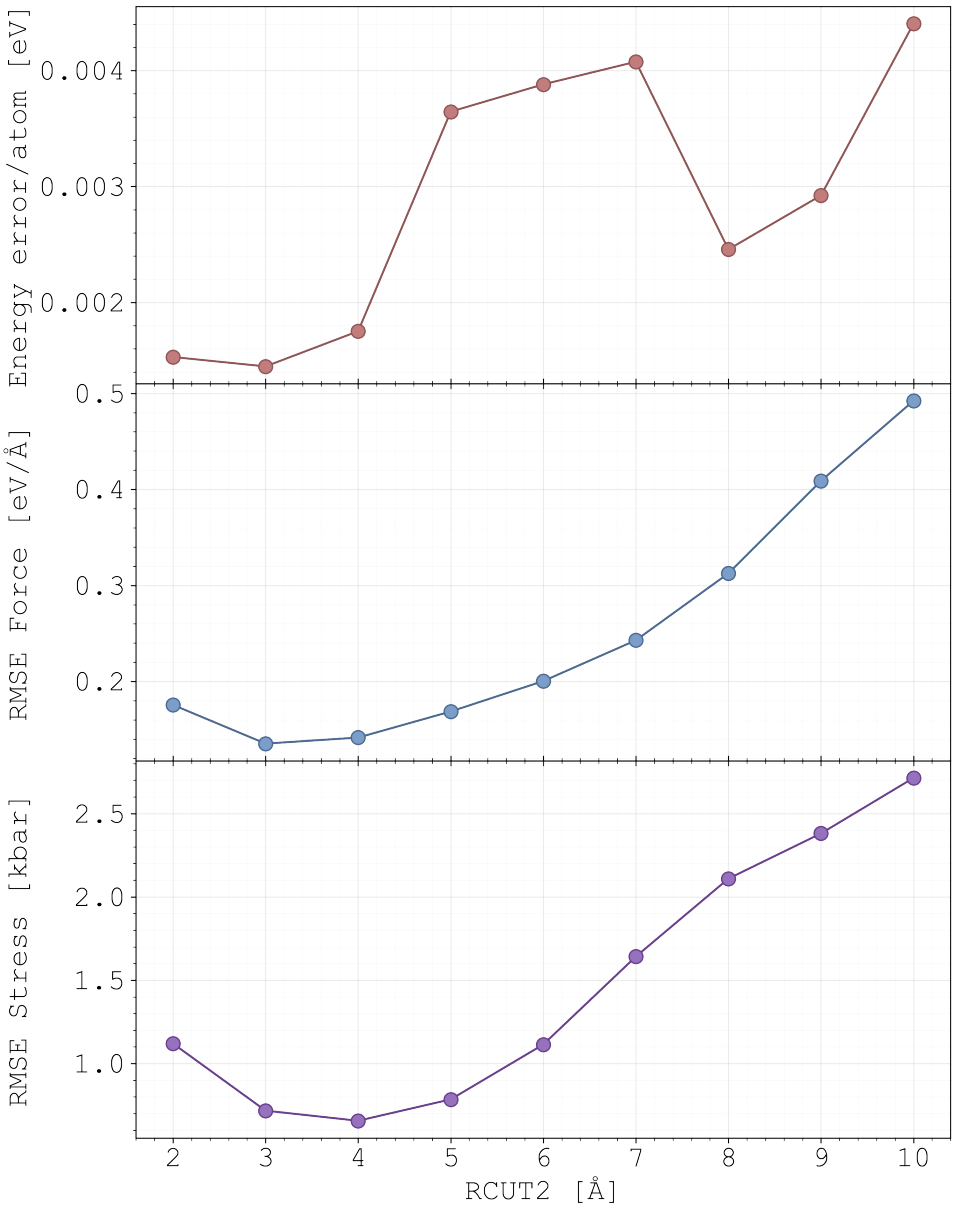

Figure 4.9: Root-mean-square error (RMSE) for the total energy, forces and stress tensor as a function of the angular descriptor (`RCUT2`).

Based on these results, two refined MLFFs were generated: **FF1** with `RCUT1`=17.0 Å and `RCUT2`=4.0 Å, and **FF2** with `RCUT1`=36.0 Å and `RCUT2`=4.0 Å. Additionally, we called **FF0** the original MLFF generated during the training phase, which used `RCUT1`=8.0 Å and `RCUT2`=5.0 Å. **FF2** was chosen to explore the effect of a larger radial descriptor, albeit no performance evaluation was carried out due to the computational cost. The performance of **FF1** was evaluated following the same procedure as before, and the results are presented in Figure 4.10. A significant improvement is observed, with an RMSE of 1.282 meV/atom for the total energy, 163 meV/Å for the forces, whereas the stress tensor RMSE increased to 1.24 kbar.

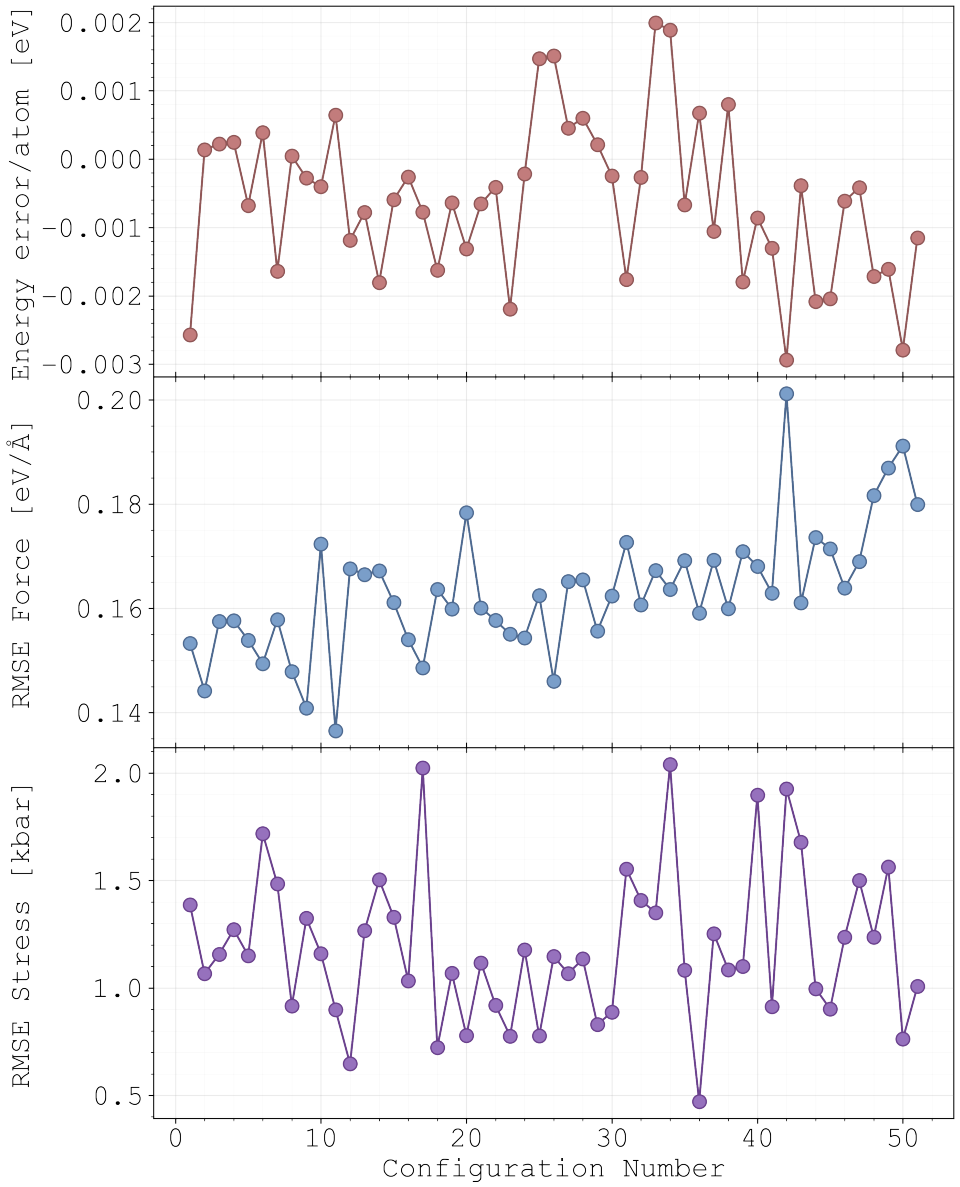

Figure 4.10: Energy error per atom and root-mean-square error (RMSE) for forces and stress tensor of C–S–H between DFT and MLFF predictions, evaluated on 50 configurations using the refined MLFF **FF1**.

Additionally, scatter parity plots comparing the reference DFT results and the predictions made by our refitted force field **FF1** are presented in Figure 4.11. This visual representation allows us to assess the accuracy of the force field predictions. We observe a good agreement between the *ab initio* data and the MLFF predictions for the total energy, forces and stresses. The energy predictions present a high density of points towards the centre of the diagonal and a moderate spread, indicating a good overall accuracy. On the other hand, the force predictions seem to be lower than the DFT values, as indicated by the spread of points below the diagonal. Finally, the stress predictions are equally distributed around the diagonal, indicating a good overall accuracy despite the increased RMSE after refinement.

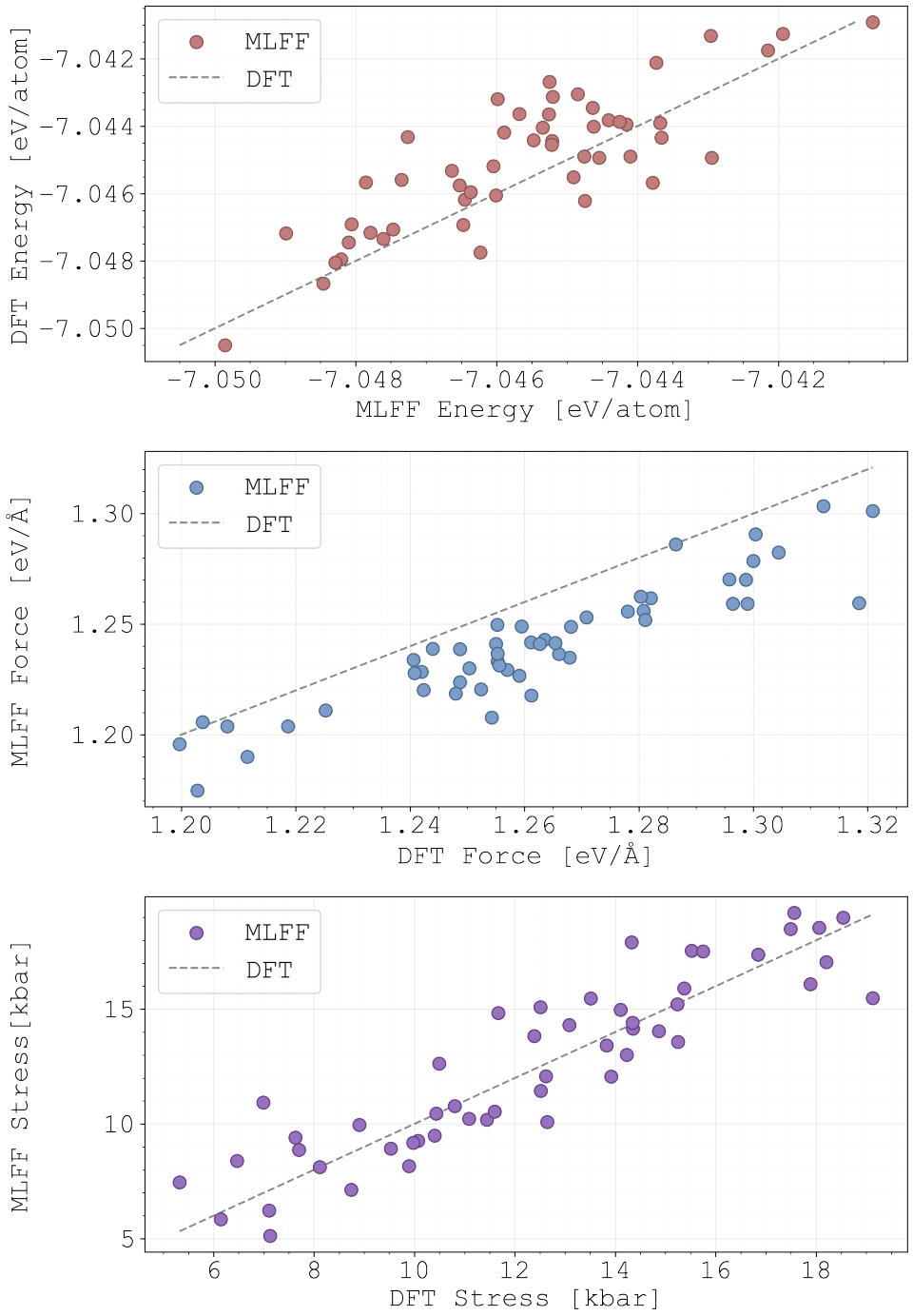

Figure 4.11: The predicted results for the total system energy, average force magnitude and average stress magnitude of C–S–H per configuration. The gray dashed line represents the results of DFT calculations, while the coloured dots represent the predictions made by **FF1**. Closer proximity to the gray dashed line indicates more accurate predictions.

4.3 Thermodynamic Properties of C–S–H

In this section, we present the thermodynamic properties of C–S–H, including the equation of state (EOS) and optimal bulk parameters. To this end, the three MLFFs generated in the previous section were employed to perform a series of energy-volume calculations and fit the

Birch-Murnaghan equation of state (EOS) to the results. The obtained results are summarised in Table 4.1.

4.3.1 Equation of State (EOS) and Bulk Parameters

Figure 4.12, 4.13, and 4.14 show the Birch-Murnaghan equation of state (EOS) obtained by fitting the energy-volume data obtained with **FF0**, **FF1** and **FF2**, respectively. For the three cases, we employed the same input structure—the relaxed C–S–H structure obtained in Section 4.1—and performed a series of energy-volume calculations by varying the cell volume within a range of 1% around the relaxed volume.

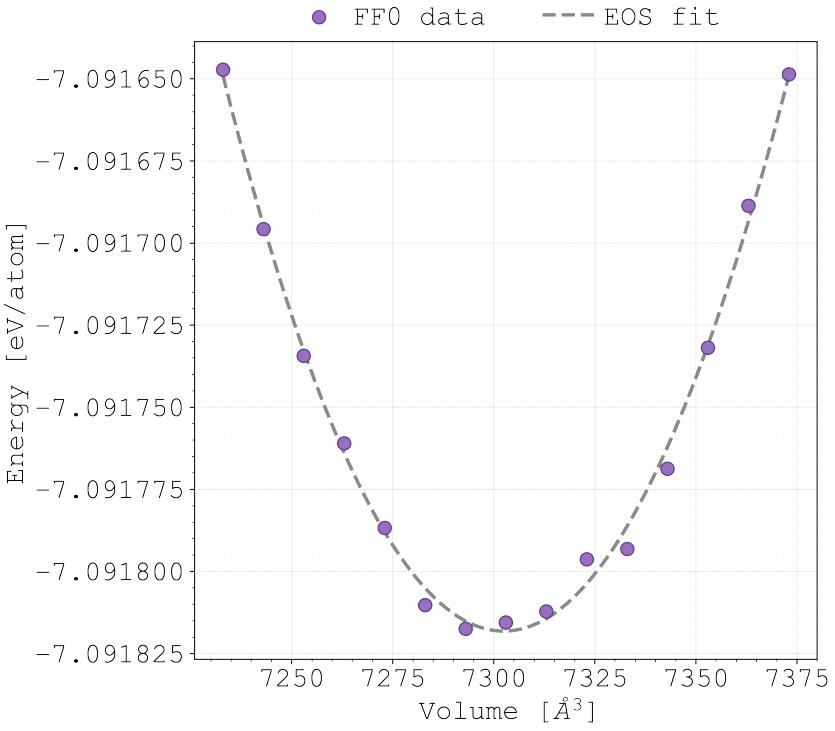

Figure 4.12: Birch-Murnaghan equation of state (EOS) obtained by fitting the energy-volume obtained with **FF0** (purple dots) for C–S–H. Optimal volume $V_0 = 7302.49$ Å^3 and bulk modulus $B_0 = 55.72$ GPa are reported.

We observe a good agreement in the optimal energy E_0 and bulk modulus B_0 values obtained with the three models; however, the optimal volume V_0 predicted by **FF2** deviated significantly from the other models. This discrepancy is likely caused by the large radial descriptor used in **FF2**, which appears to cause the force field to overfit the training data, degrading generalisation of the energy-volume relationship. Moreover, **FF2** yields a negative pressure derivative of the bulk modulus, B_0', indicating that C–S–H would become more compressible under increasing pressure. This is counterintuitive and non-physical, as materials typically become less compressible under higher pressures owing to the increased overlap of electronic wavefunctions and Pauli repulsion [67].

Consequently, we excluded **FF2** from further analysis, as it provides no meaningful physical representation of C–S–H.

Experimental values of B_0 and B_0' are retrieved from the literature for comparison. Oh *et al.* [10] reported a bulk modulus of 47 ± 3 GPa for a 14 Å tobermorite—a crystalline analogue of C–S–H—using high-pressure synchrotron X-ray diffraction, and assumed a pressure derivative of the bulk modulus $B_0' = 4$ based on previous studies. Additionally, Oh *et al.* [68] reported a bulk modulus of 34 ± 7 GPa for C–S–H(I) (C/S=0.96)—a more crystalline form of C–S–H—using high-pressure X-ray diffraction, whereas the pressure derivative of the bulk modulus was calculated to be 7 ± 7.

Notably, our predicted values for B_0 are consistent with the experimental values reported in [10]. On the other hand, our values are notably higher than the experimental value reported in [68], likely due to the different C/S ratio used in that study. Finally, even though our predicted values for B_0' are within the uncertainty range reported in [68], more accurate experimental data is needed to validate our results.

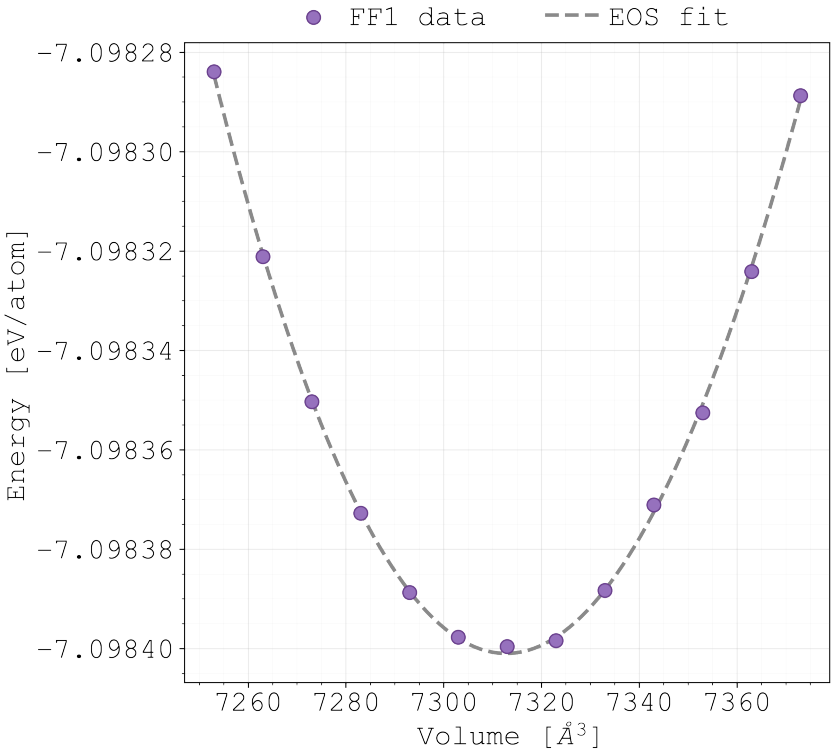

Figure 4.13: Birch-Murnaghan equation of state (EOS) obtained by fitting the energy-volume obtained with **FF1** (purple dots) for C–S–H. Optimal volume $V_0 = 7312.8$ Å^3 and bulk modulus $B_0 = 51.14$ GPa are reported.

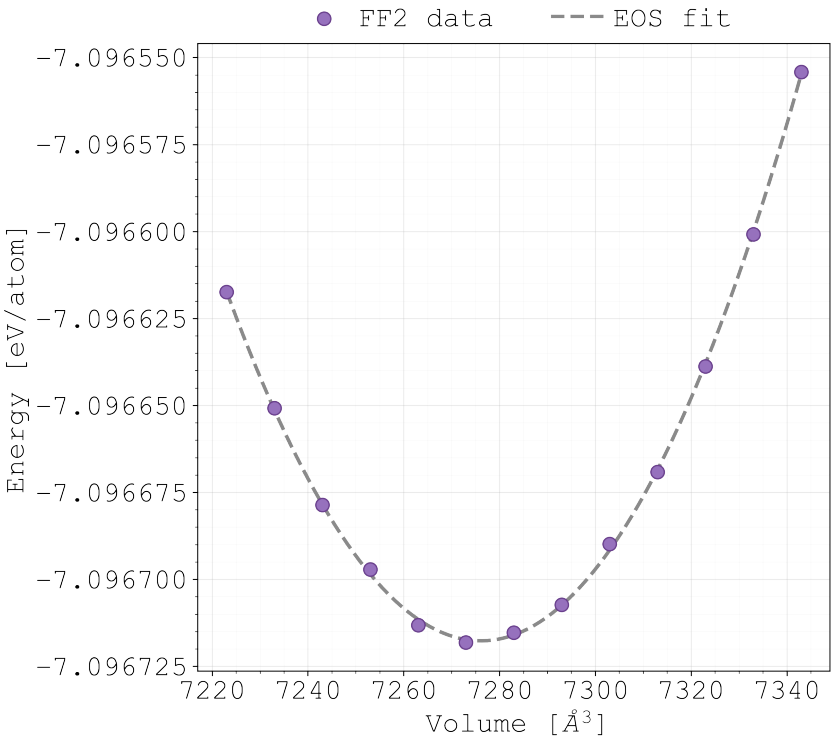

Figure 4.14: Birch-Murnaghan equation of state (EOS) obtained by fitting the energy-volume obtained with **FF2** (purple dots) for C–S–H. Optimal volume $V_0 = 7276.16$ Å^3 and bulk modulus $B_0 = 57.88$ GPa are reported.

4.3.2 Simulated Annealing (SA) and EOS

A simulated annealing (SA) procedure was applied to further optimise the C–S–H structure and potentially improve the bulk parameters obtained previously. During this process, the system is gradually cooled down, and as it does so, it reaches a low-energy stable configuration. Starting from the last configuration generated during the evaluation phase—which was already thermalised and stable—the SA was performed using the **FF0** force field, lowering the temperature from 400 K down to 0 K. This approach allows the system to explore a broader range of configurations, potentially leading to a more favorable state that improves the optimal bulk parameters.

Thereafter, the optimised structure was used to obtain the energy-volume data and compute the EOS, which is shown in Figure 4.15. The optimal parameters are reported in Table 4.1. We observe a significant increase in the optimal volume, whereas E_0 and B_0 are close to the values obtained with **FF0** and **FF1**. Finally, B_0' is comparable to the value obtained with **FF1** and is within the uncertainty range of the experimental value.

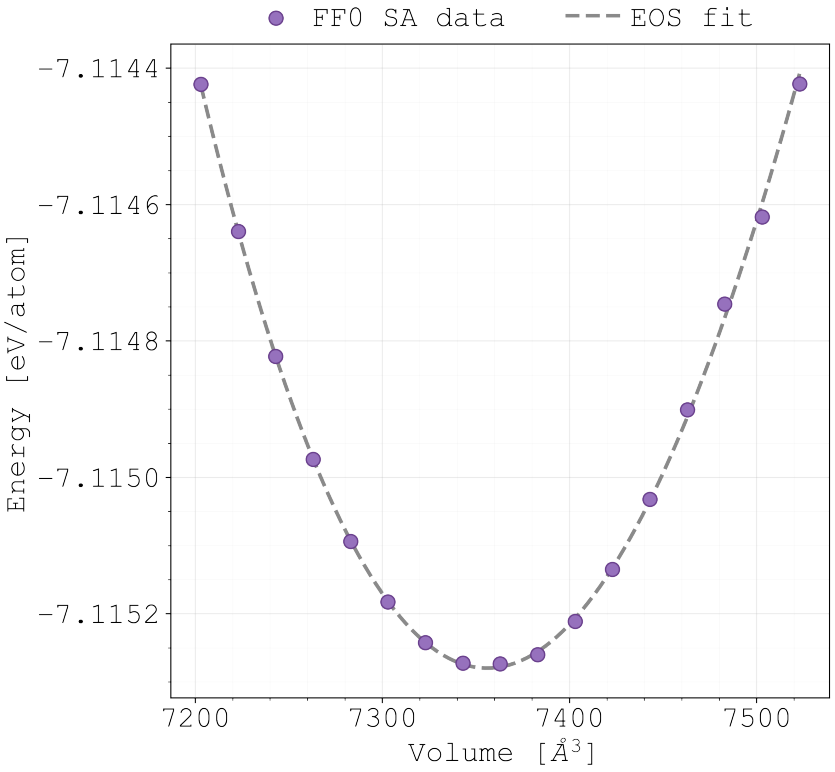

Figure 4.15: Birch-Murnaghan equation of state (EOS) obtained by fitting the energy-volume obtained with **FF0** (purple dots) for C–S–H after the simulated annealing (SA) procedure. Optimal volume $V_0 = 7356.09$ Å^3 and bulk modulus $B_0 = 55.12$ GPa are reported.

Model	V_0 (Å^3)	E_0 (eV)	B_0 (GPa)	B_0'
FF0	7302.5	-4893.3	55.72	3.96
FF1	7312.8	-4897.9	51.14	9.80
FF2	7276.2	-4896.7	57.88	-5.45
FF0 SA	7356.1	-4909.5	55.12	9.61
Literature	N/A	N/A	47±3 (tobermorite 14 Å) [10] 34±7 (Ca/Si=0.96) [68]	4.00 (assumed) [10] 7±7 (calculated) [68]

Table 4.1: Optimal energy E_0, volume V_0, bulk modulus B_0 and bulk modulus derivative B_0' are reported for the three MLFFs and the simulated annealing (SA) procedure. Experimental values from the literature are also included for comparison.

4.4 Transferability of MLFFs and Thermal Expansion Coefficient of C–S–H

The transferability of a machine learning force field is critical to its practical utility, as it determines the model's ability to accurately predict material properties under conditions not explicitly included in the training data. In this work, we assess the transferability of **FF0** by applying it to MD simulations across a range of temperatures (200 K to 400 K) and evaluating

the thermal expansion behaviour. To this end, the initial structure was thermalised at each temperature for 10 ps, followed by a 20 ps MD simulation at constant temperature. The average cell volume was computed for each temperature, and then a linear fit [69] was applied to the data (see Figure 4.16), yielding the following expression:

$$V(T) = 0.4724\,T + 7267.785, \quad (R^2 = 0.9435) \tag{4.1}$$

Afterwards, the thermal expansion coefficient $\alpha_v = (1/V_0)(\partial V/\partial T)$ was computed—where V_0 was set to be the average cell volume at 200 K—yielding a value of $\alpha_v = 6.4 \times 10^{-5}$ K^{-1}, whereas a value of $4.5(\pm 0.9) \times 10^{-5}$ K^{-1} was reported by Qomi *et al.* [19] in a numerical study of a 11-Å tobermorite C–S–H model. Notably, our obtained value is slightly higher, mainly due to the different model used in the literature. The obtained agreement with literature values indicates that **FF0** maintains physical consistency in reproducing structural responses to moderate thermal variations, demonstrating reasonable transferability within the temperature range of 200 K to 400 K, where our MLFF remains valid without further refinement requirements.

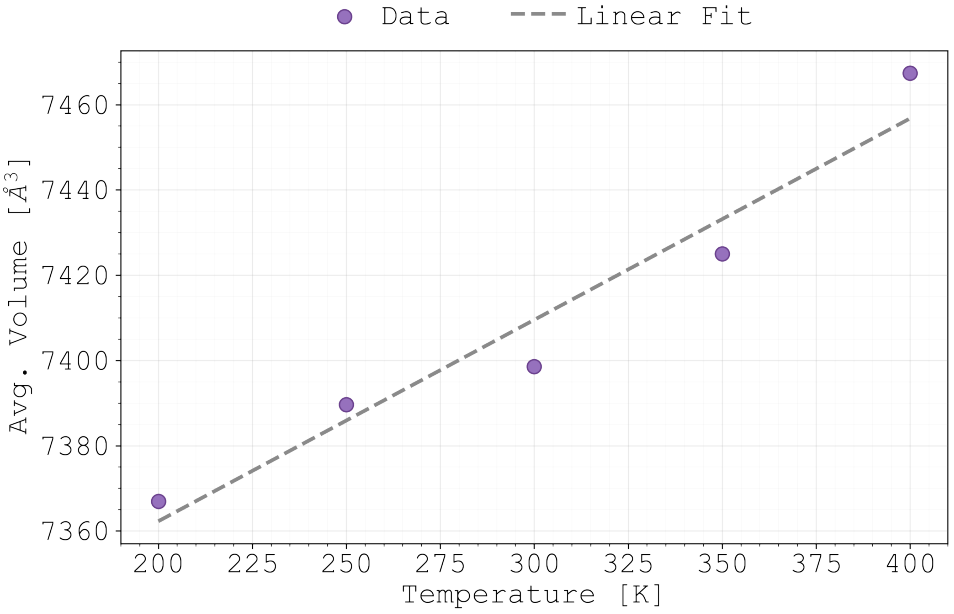

Figure 4.16: Average cell volume (purple dots) computed from MD simulations using **FF0** at 200, 250, 300, 350 and 400 K ran for 10000 steps (20 ps) each. A linear fit (dashed line) is applied to the data, and the thermal expansion coefficient α_v is computed from the slope of the fit.

Nevertheless, transferability to more different conditions, such as phase transitions or structure transitions where not assessed in this work, and would likely require further refinement of the MLFF. Finally, evaluating the transferability of **FF1** was not possible due to instability issues encountered during the MD simulations, which need to be addressed in future work.

Chapter 5

Conclusions & Outlook

In this work, first-principles calculations and machine learning techniques were successfully integrated to construct a robust and efficient machine learning force field (MLFF) tailored for calcium silicate hydrates (C–S–H). The model was developed through an on-the-fly training scheme within *ab initio* molecular dynamics (AIMD) simulations, enabling accurate representation of the atomic-scale interactions governing the structure and mechanics of C–S–H. Beginning with the relaxation of the C–S–H unit cell using VASP and the PBEsol exchange-correlation functional, a systematic workflow was established to train, validate, and refine the MLFF.

The results of molecular dynamics and relaxation simulations confirm that the developed MLFF faithfully reproduces the energetics and mechanical response of C–S–H, achieving close agreement with first-principles data in terms of total energy, atomic forces, and stress tensors. Moreover, the force field demonstrates excellent predictive capability for macroscopic mechanical properties—such as the bulk modulus—yielding values consistent with experimental findings. Finally, transferability tests across a temperature range of 200–400 K revealed that the MLFF reliably captures the thermal behaviour of C–S–H, offering a computationally efficient tool for exploring its structural and thermodynamic properties without requiring further training.

Nonetheless, there remain opportunities for further improvement of the methodology presented herein. Notable directions for future work include validating the bulk parameters obtained with the MLFFs against DFT calculations to provide a better assessment of their accuracy. Additional improvements could involve the use of more advanced exchange–correlation functionals along with van der Waals dispersion corrections to enhance the accuracy of the force field, evaluating the performance of the MLFF using a larger supercell, and expanding the training dataset to include a wider range of C–S–H compositions to improve the model's generalisability.

Finally, machine learning-based approaches, as presented in this work, hold great promise for advancing materials research, particularly in the context of large and complex disordered systems

like C–S–H, where first-principles methods can be computationally prohibitive. In this regard, machine learning-based concrete research could significantly accelerate the development of more durable and sustainable concrete, addressing the pressing environmental challenges associated with its production and use.

Appendix A

Projected Density of States of C–S–H

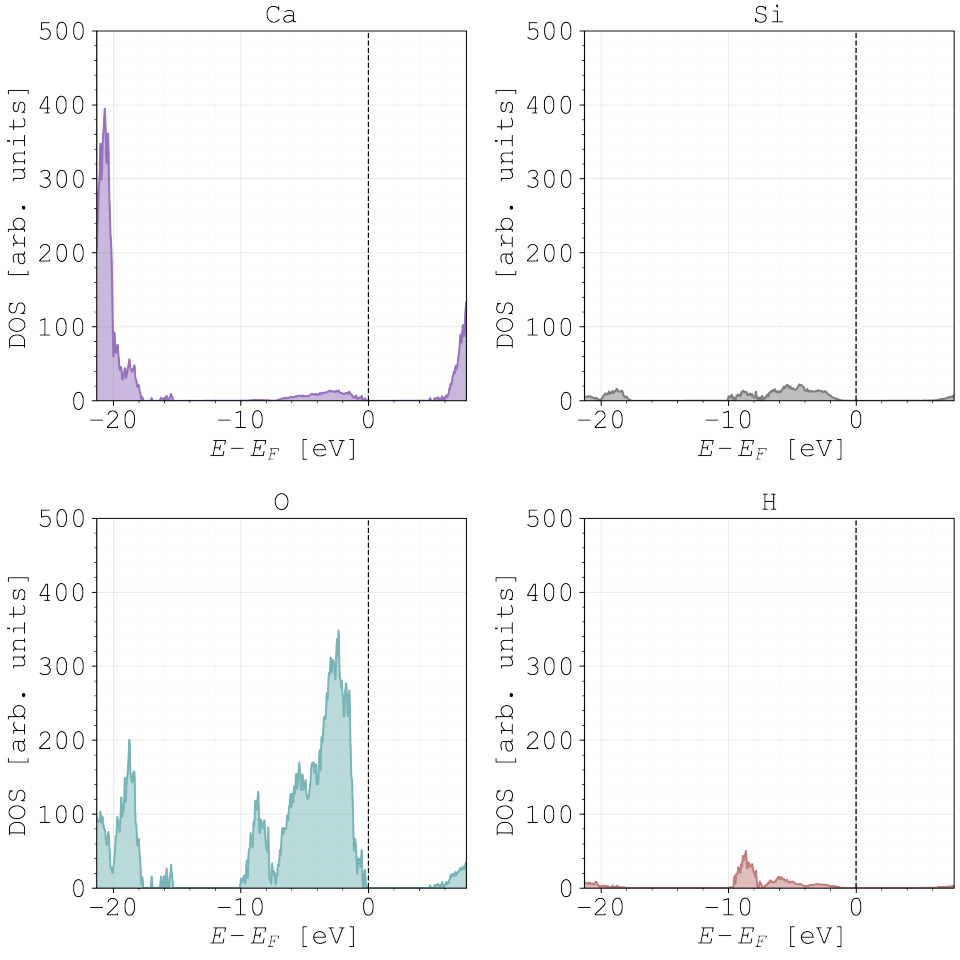

Figure A.1: Detailed electronic density of states (DOS) of C–S–H computed using the HSEsol hybrid functional. Element-resolved contributions from Ca, Si, O, and H are shown. The energy axis (x-axis) is referenced to the Fermi level, indicated by the dashed vertical line at 0 eV, while the y-axis represents the density of states (in states/eV).

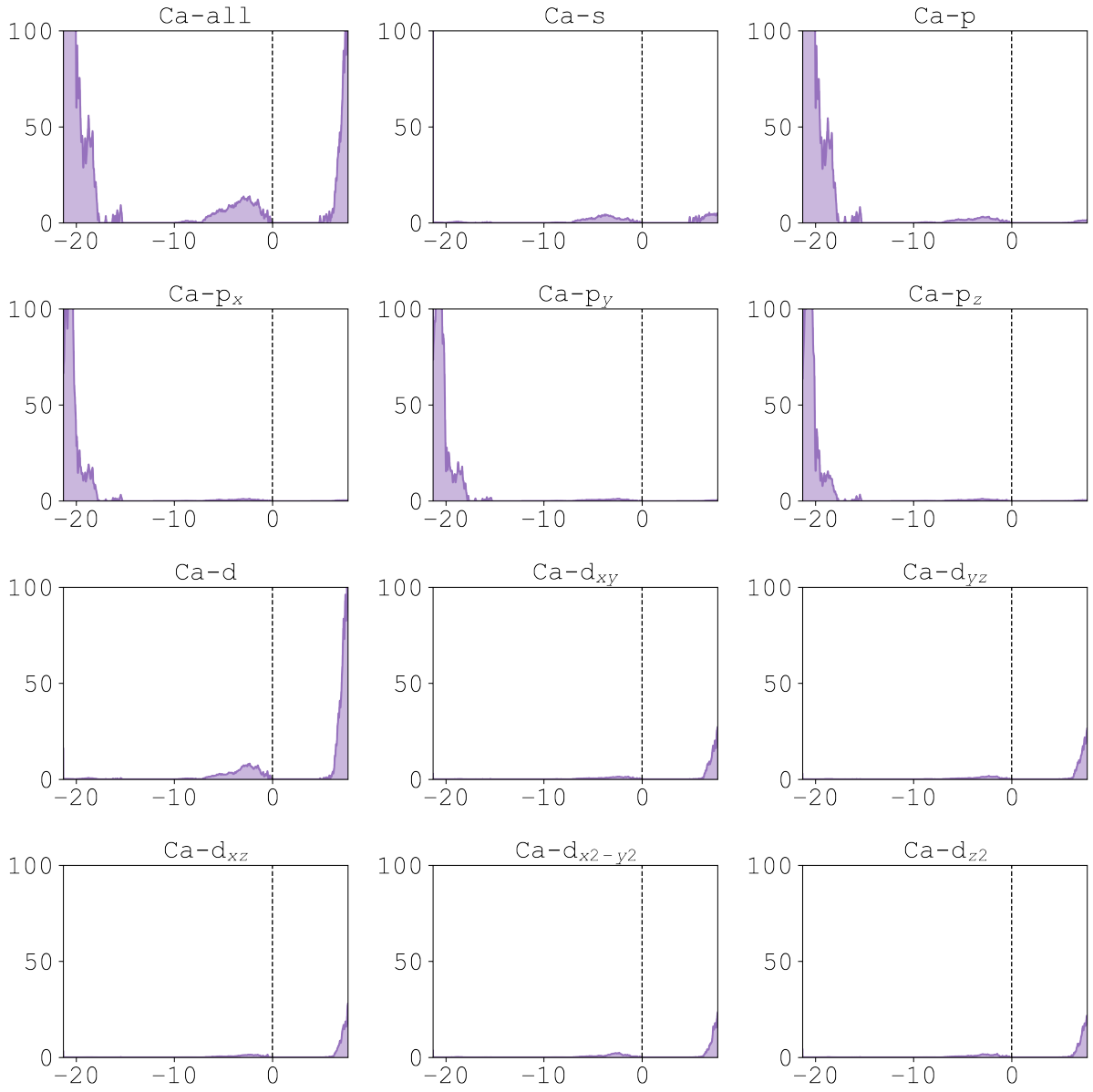

Figure A.2: Orbital-resolved density of states (DOS) for Ca atoms in C–S–H computed employing the HSEsol hybrid functional. The plots show the total Ca contribution and its decomposition into s, p (p_x, p_y, p_z), and d (d_{xy}, d_{yz}, d_{xz}, $d_{x^2-y^2}$, d_{z^2}) orbitals. The x-axis reports the energy (in eV) relative to the Fermi level (dashed vertical line at 0 eV), while the y-axis shows the density of states (in states/eV).

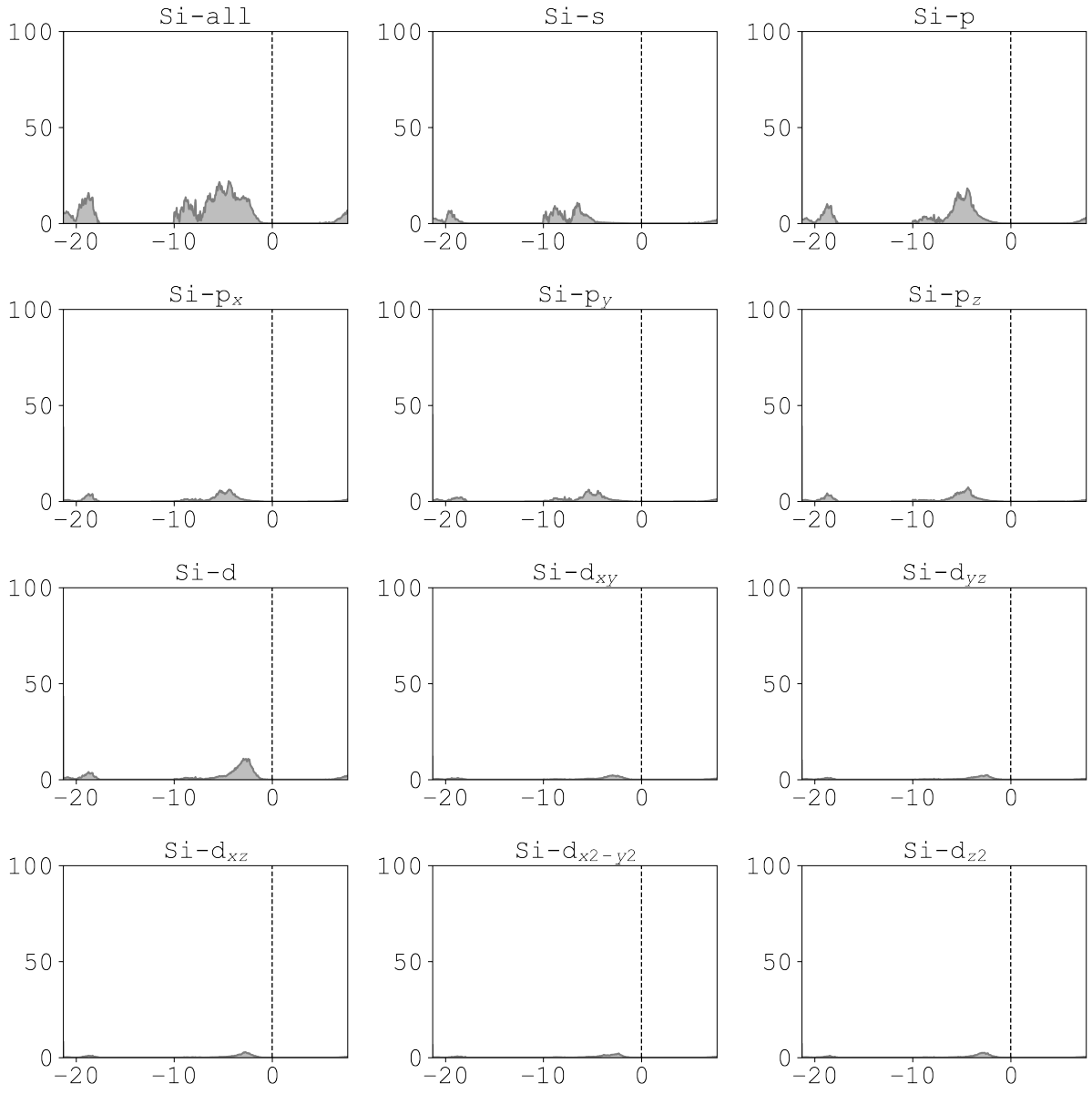

Figure A.3: Orbital-resolved density of states (DOS) for Si atoms in C–S–H computed employing the HSEsol hybrid functional. The plots show the total Si contribution and its decomposition into s, p (p_x, p_y, p_z), and d (d_{xy}, d_{yz}, d_{xz}, $d_{x^2-y^2}$, d_{z^2}) orbitals. The x-axis reports the energy (in eV) relative to the Fermi level (dashed vertical line at 0 eV), while the y-axis shows the density of states (in states/eV).

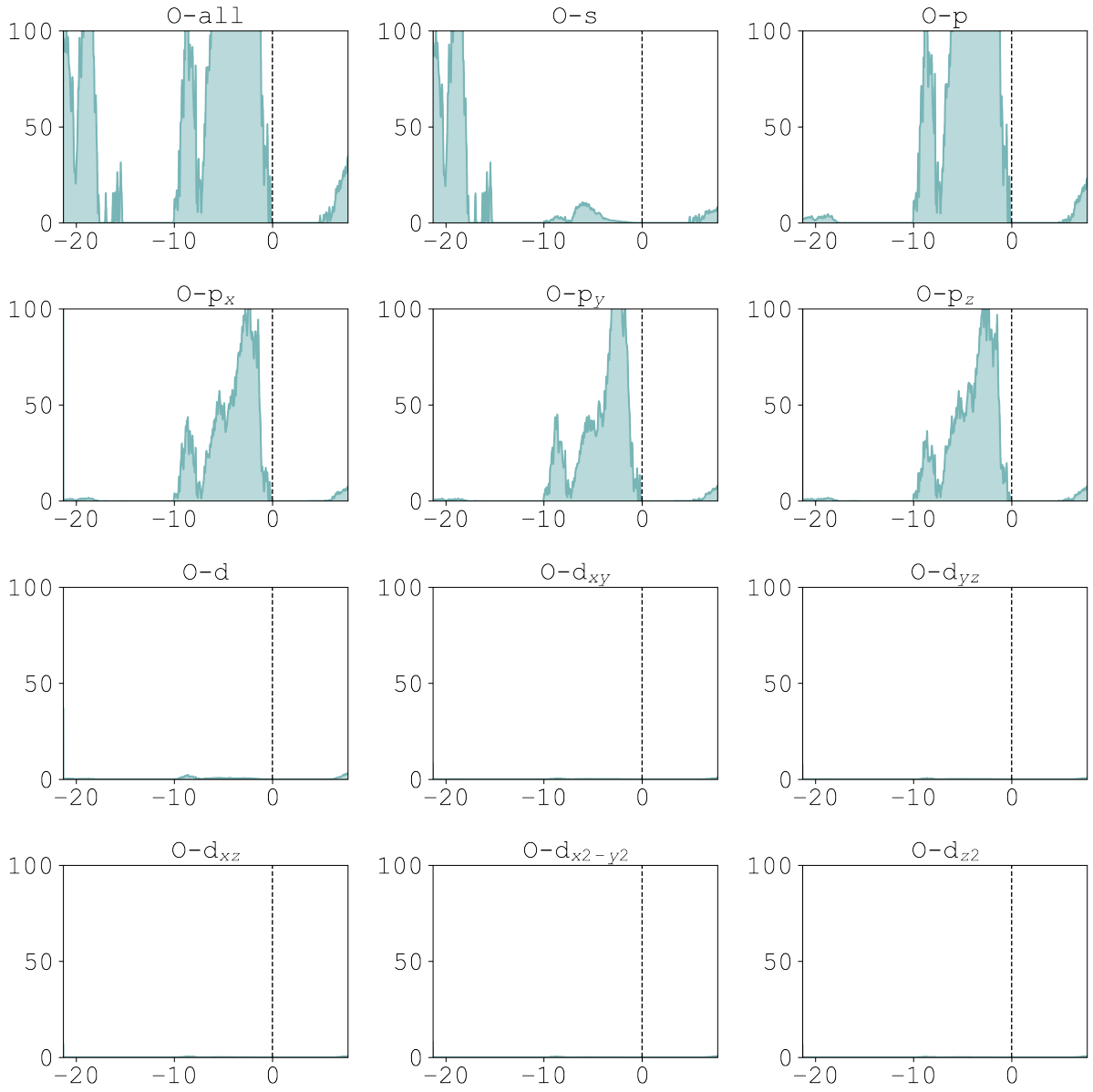

Figure A.4: Orbital-resolved density of states (DOS) for O atoms in C–S–H computed employing the HSEsol hybrid functional. The plots show the total O contribution and its decomposition into s, p (p_x, p_y, p_z), and d (d_{xy}, d_{yz}, d_{xz}, $d_{x^2-y^2}$, d_{z^2}) orbitals. The x-axis reports the energy (in eV) relative to the Fermi level (dashed vertical line at 0 eV), while the y-axis shows the density of states (in states/eV).

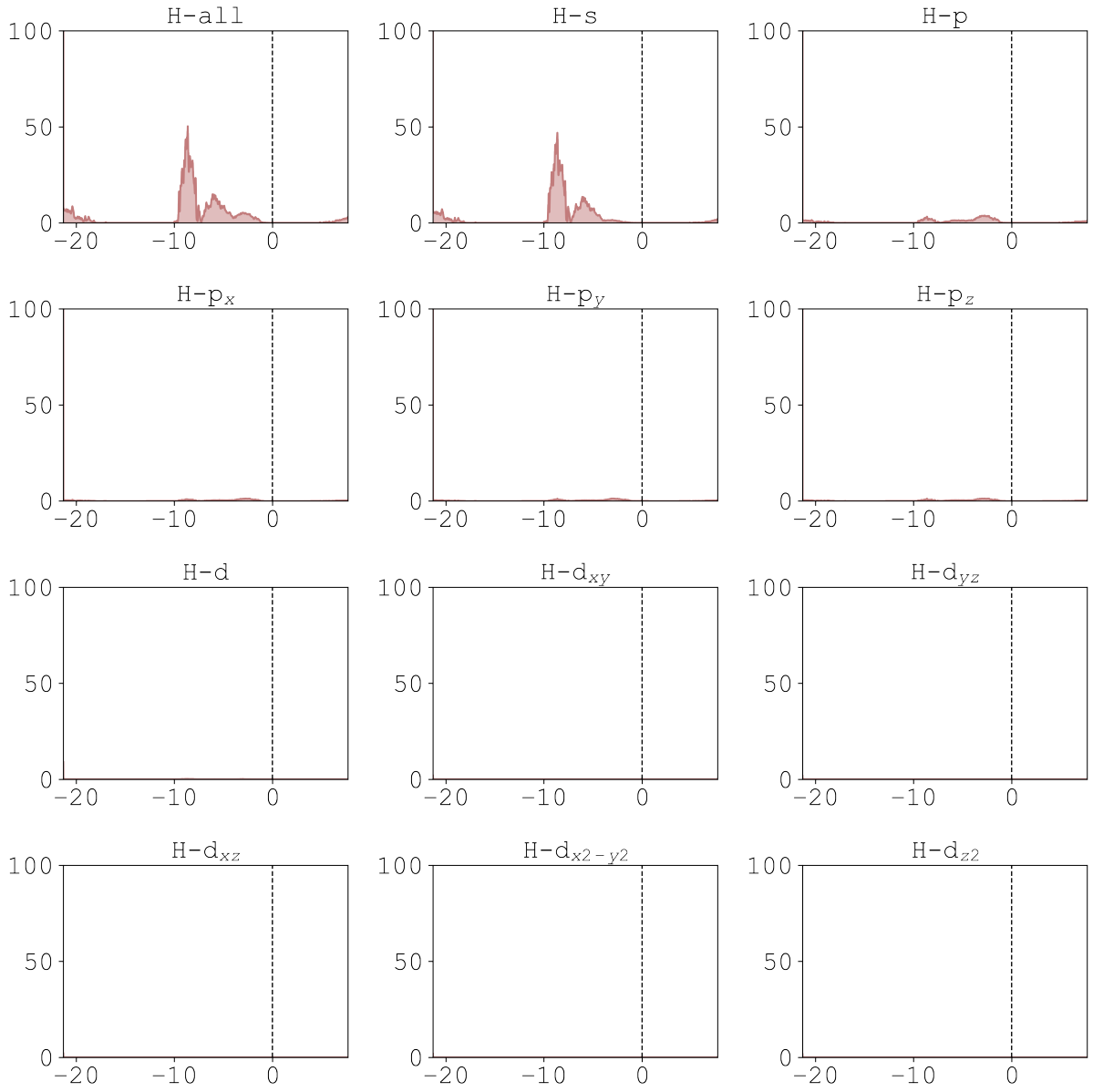

Figure A.5: Orbital-resolved density of states (DOS) for H atoms in C–S–H computed employing the HSEsol hybrid functional. The plots show the total H contribution and its decomposition into s, p (p_x, p_y, p_z), and d (d_{xy}, d_{yz}, d_{xz}, $d_{x^2-y^2}$, d_{z^2}) orbitals. The x-axis reports the energy (in eV) relative to the Fermi level (dashed vertical line at 0 eV), while the y-axis shows the density of states (in states/eV).

Appendix B

Computational Parameters

```
                              GENERAL
SYSTEM      = C–S–H       # System name
PREC        = Accurate    # Precision level
                        ELECTRONIC OPTIMIZATION
ENCUT       = 800         # Plane-wave cutoff (eV)
LREAL       = Auto        # Real-space projection
ISMEAR      = 0           # Smearing method
SIGMA       = 0.05        # Smearing width (eV)
ALGO        = F           # Electronic minimization algorithm
AMIX        = 0.1         # Charge density mixing parameter (damping)
              EXCHANGE-CORRELATION / FUNCTIONAL
GGA         = PS          # PBEsol functional
IVDW        = 11          # DFT-D3(zero) vdW correction
LASPH       = .TRUE.      # Non-spherical contributions
LMAXMIX     = 4           # Maximum l for charge mixing
                        CHARGE & WAVEFUNCTION
LCHARG      = F           # Do not write CHGCAR
                          IONIC RELAXATION
NELMIN      = 4           # Minimum SCF steps
MAXMIX      = 40          # Maximum mixing steps
IBRION      = 2           # Ionic relaxation algorithm
ISIF        = 3           # Relax ions + cell shape + volume
NSW         = 700         # Maximum ionic steps
EDIFFG      = -0.02       # Convergence criterion (eV/Å)
ADDGRID     = T           # Additional grid for accuracy
```

Figure B.1: Complete `INCAR` configuration used for C–S–H structure relaxation. Electronic optimisation is performed with a plane-wave cutoff of 800 eV (`ENCUT=800`), Gaussian smearing (`ISMEAR=0`, `SIGMA=0.05 eV`), and the RMM-DIIS algorithm (`ALGO=F`) with a charge mixing parameter of 0.1 (`AMIX=0.1`). Exchange-correlation is treated with the PBEsol functional (`GGA=PS`) including DFT-D3 zero-damping van der Waals corrections (`IVDW=11`) and non-spherical contributions (`LASPH=.TRUE.`). Ionic relaxation uses the conjugate-gradient method (`IBRION=2`) with full cell relaxation (`ISIF=3`), a maximum of 700 steps (`NSW=700`), and a force convergence criterion of 0.02 eV/Å (`EDIFFG=-0.02`). Additional grid refinement (`ADDGRID=.TRUE.`) is enabled for improved accuracy.

```
                    GENERAL SETTINGS
SYSTEM      = CSH-DOS      # System name
                    ELECTRONIC RX
ISMEAR      = 0            # Gaussian smearing (insulators)
SIGMA       = 0.05         # Smearing width (eV)
LREAL       = Auto         # Projection in real space
PREC        = Accurate     # Precision level
ENCUT       = 800          # Plane-wave cutoff energy
ALGO        = N            # Electronic minimization algorithm
NELM        = 300          # Max SCF steps
EDIFF       = 5E-5         # Electronic energy convergence
LORBIT      = 11           # Projected DOS output
                    FUNCTIONAL
GGA         = PS           # PBEsol base for HSEsol
LHFCALC     = .TRUE.       # Enable hybrid Hartree-Fock exchange
HFSCREEN    = 0.2          # HSE screening parameter
AEXX        = 0.25         # Fraction of exact exchange
              DISPERSION & PAW SETTINGS
IVDW        = 11           # D3 dispersion correction
LASPH       = .TRUE.       # Non-spherical PAW contributions
LMAXMIX     = 4            # Max l quantum number
VDW_S8      = 0.7220
VDW_SR      = 1.5810
              CHARGE & WAVEFUNCTIONS
ICHARG      = 2            # Full SCF calculation
LCHARG      = .TRUE.       # Write CHGCAR
LWAVE       = .FALSE.      # Do not write WAVECAR
                    DOS SETTINGS
NEDOS       = 3001         # Number of DOS points
EMIN        = -19.0        # Minimum energy (eV)
EMAX        = 10.0         # Maximum energy (eV)
                  IONIC RELAXATION
NSW         = 0            # No ionic relaxation
IBRION      = -1           # No ionic steps
                    OTHER SETTINGS
BANDGAP     = COMPACT
NCORE       = 24           # Cores per node
NSIM        = 6            # Parallelization parameter
```

Figure B.2: INCAR configuration used for density of states (DOS) calculations of C–S–H. Electronic optimisation uses a plane-wave cutoff of 800 eV (`ENCUT=800`), Gaussian smearing for insulators (`ISMEAR=0`, `SIGMA=0.05 eV`), up to 300 SCF steps (`NELM=300`), and the Normal blocked-Davidson algorithm (`ALGO=N`). The HSEsol hybrid functional is applied (`LHFCALC=.TRUE.`) with 25% exact exchange (`AEXX=0.25`) and screening parameter (`HFSCREEN=0.2`). DFT-D3 dispersion corrections are included (`IVDW=11`) with parameters (`VDW_S8=0.7220`) and (`VDW_SR=1.5810`). DOS output is defined by (`NEDOS=3001`) points in the energy range from -19.0 to 10.0 eV (`EMIN=-19.0`, `EMAX=10.0`). Ionic relaxation is disabled (`NSW=0`, `IBRION=-1`).

GENERAL	
SYSTEM	**= C–S–H**
ELECTRONIC OPTIMIZATION	
ENCUT	**= 800**
LREAL	**= auto**
ISMEAR	**= 0**
SIGMA	**= 0.05**
ALGO	**= N**
EDIFF	**= 1E-5**
EXCHANGE-CORRELATION / FUNCTIONAL	
GGA	**= PS**
LASPH	**= .TRUE.**
LMAXMIX	**= 4**
CHARGE & WAVEFUNCTION	
LWAVE	**= F**
LCHARG	**= F**
MOLECULAR DYNAMICS	
IBRION	**= 0**
NSW	**= 50000**
POTIM	**= 2.0**
MDALGO	**= 3**
LANGEVIN_GAMMA	**= 1 1 1 1**
LANGEVIN_GAMMA_L	**= 10**
PMASS	**= 10**
TEBEG	**= 400**
POMASS	**= 40.078 28.085 16.000 8.00**
ISIF	**= 3**
K-POINTS	
KSPACING	**= 1**
MACHINE LEARNING	
ML_LMLFF	**= .TRUE.**
ML_MODE	**= TRAIN**
ML_WTSIF	**= 2**

Figure B.3: INCAR configuration used for ab initio molecular dynamics (AIMD) simulations of C–S–H, simultaneously for machine learning force field (MLFF) training. Ionic dynamics employ the velocity-Verlet algorithm (`IBRION=0`) for 50,000 steps (`NSW=50000`) with a timestep of 2 fs (`POTIM=2.0`). A Langevin thermostat is applied with damping (`LANGEVIN_GAMMA=1`) and (`LANGEVIN_GAMMA_L=10`), targeting an initial temperature of 400 K (`TEBEG=400`). Electronic optimization uses a plane-wave cutoff of 800 eV (`ENCUT=800`), Gaussian smearing for insulators (`ISMEAR=0`, `SIGMA=0.05 eV`), and the Normal blocked-Davidson algorithm (`ALGO=N`) with convergence (`EDIFF=1E-5`). The PBEsol functional is applied with non-spherical contributions (`LASPH=.TRUE.`) and (`LMAXMIX=4`). Periodic cell relaxation is allowed (`ISIF=3`). Machine learning force field training is enabled (`ML_LMLFF=.TRUE.`, `ML_MODE=TRAIN`).

GENERAL	
SYSTEM	= **C–S–H**
MOLECULAR DYNAMICS	
IBRION	= **0**
NSW	= **50000**
POTIM	= **2.0**
MDALGO	= **3**
LANGEVIN_GAMMA	= **1 1 1 1**
LANGEVIN_GAMMA_L	= **10**
PMASS	= **10**
TEBEG	= **400**
POMASS	= **40.078 28.085 16.000 8.00**
ISIF	= **3**
K-POINTS	
KSPACING	= **1**
MACHINE LEARNING	
ML_LMLFF	= **.TRUE.**
ML_ISTART	= **2**
PARALLELIZATION	
NCORE	= **48**
NSIM	= **4**
ISYM	= **0**
LSCALU	= **.TRUE.**
LSCALAPACK	= **.TRUE.**

Figure B.4: INCAR configuration used for molecular dynamics (MD) simulations of C–S–H using a machine learning force field (MLFF). Ionic dynamics are performed with the velocity-Verlet algorithm (`IBRION=0`) for 50,000 steps (`NSW=50000`) with a timestep of 2 fs (`POTIM=2.0`). A Langevin thermostat is applied (`MDALGO=3`) with damping parameters (`LANGEVIN_GAMMA=1 1 1 1` and `LANGEVIN_GAMMA_L=10`), targeting an initial temperature of 400 K (`TEBEG=400`). Full ionic and cell relaxation is enabled (`ISIF=3`). Machine learning force field usage is controlled via (`ML_LMLFF=.TRUE.`, `ML_ISTART=2`). Electronic structure parameters are not included since the MD is performed solely with the MLFF.

GENERAL	
SYSTEM	= **cement**
MOLECULAR DYNAMICS	
IBRION	= **0**
NSW	= **20000**
POTIM	= **2.0**
MDALGO	= **3**
LANGEVIN_GAMMA	= **1 1 1 1**
LANGEVIN_GAMMA_L	= **10**
PMASS	= **10**
TEBEG	= **400**
TEEND	= **0**
POMASS	= **40.078 28.085 16.000 8.00**
ISIF	= **3**
INITIAL CONDITIONS	
ISTART	= **1**
ICHARG	= **1**
K-POINTS	
KSPACING	= **1**
MACHINE LEARNING	
ML_LMLFF	= **.TRUE.**
ML_ISTART	= **2**
PARALLELIZATION	
NCORE	= **48**
NSIM	= **4**
ISYM	= **0**
LSCALU	= **.TRUE.**
LSCALAPACK	= **.TRUE.**

Figure B.5: INCAR configuration used for simulated annealing molecular dynamics (MD) of C–S–H using a machine learning force field (MLFF). Ionic dynamics are performed with the velocity-Verlet algorithm (`IBRION=0`) for 20,000 steps (`NSW=20000`) with a timestep of 2 fs (`POTIM=2.0`). A Langevin thermostat (`MDALGO=3`) is applied with damping parameters (`LANGEVIN_GAMMA=1 1 1 1`, `LANGEVIN_GAMMA_L=10`), starting at `TEBEG=400 K` and cooling to `TEEND=0 K`. Full ionic and cell relaxation is enabled (`ISIF=3`). Machine learning force field usage is controlled via (`ML_LMLFF=.TRUE.`, `ML_ISTART=2`). Initial conditions correspond to a fresh start of the system (`ISTART=1`).

```
                    GENERAL
SYSTEM      = C–S–H     # System name
                    K-POINTS
KSPACING    = 1         # K-point spacing
                  MACHINE LEARNING
ML_LMLFF    = .TRUE.    # Enable machine learning force field
ML_MODE     = refit     # Machine learning mode
ML_RCUT1    = 60.0      # Cutoff radius for rcut1
```

Figure B.6: INCAR configuration used for machine learning refitting of the `RCUT1` descriptor for C–S–H. K-point spacing is defined by `KSPACING=1`. Machine learning force field usage is enabled (`ML_LMLFF=.TRUE.`) with the refitting mode (`ML_MODE=refit`). The `ML_RCUT1` parameter defines the cutoff radius for the `RCUT1` descriptor and is evaluated over a range of values during the refitting process.

```
                    GENERAL
SYSTEM      = C–S–H     # System name
                    K-POINTS
KSPACING    = 1         # K-point spacing
                  MACHINE LEARNING
ML_LMLFF    = .TRUE.    # Enable machine learning force field
ML_MODE     = refit     # Machine learning mode
ML_RCUT2    = 10.0      # Cutoff radius for rcut2
```

Figure B.7: INCAR configuration used for machine learning refitting of the `RCUT2` descriptor for C–S–H. K-point spacing is defined by `KSPACING=1`. Machine learning force field usage is enabled (`ML_LMLFF=.TRUE.`) with the refitting mode (`ML_MODE=refit`). The `ML_RCUT2` parameter defines the cutoff radius for the `RCUT2` descriptor and is evaluated over a range of values during the refitting process.

Bibliography

[1] P.K. Mehta and P.J.M. Monteiro. *Concrete: Microstructure, Properties, and Materials*. McGraw-hill's AccessEngineering. McGraw-Hill Education, 2014. ISBN: 978-0-07-179787-0. URL: https://books.google.com.ec/books?id=X84TAgAAQBAJ.

[2] Paulo J. M. Monteiro, Sabbie A. Miller, and Arpad Horvath. "Towards Sustainable Concrete". In: *Nature Materials* 16.7 (July 2017), pp. 698–699. ISSN: 1476-4660. DOI: 10.1038/nmat4930. (Visited on 03/16/2025).

[3] Henri Van Damme. "Concrete Material Science: Past, Present, and Future Innovations". In: *Cement and Concrete Research*. SI : Digital Concrete 2018 112 (Oct. 2018), pp. 5–24. ISSN: 0008-8846. DOI: 10.1016/j.cemconres.2018.05.002. (Visited on 03/16/2025).

[4] Joseph J. Biernacki et al. "Cements in the 21st Century: Challenges, Perspectives, and Opportunities". In: *Journal of the American Ceramic Society* 100.7 (2017), pp. 2746–2773. ISSN: 1551-2916. DOI: 10.1111/jace.14948. (Visited on 03/16/2025).

[5] Qing Ji, Roland J. M. Pellenq, and Krystyn J. Van Vliet. "Comparison of Computational Water Models for Simulation of Calcium–Silicate–Hydrate". In: *Computational Materials Science* 53.1 (Feb. 2012), pp. 234–240. ISSN: 0927-0256. DOI: 10.1016/j.commatsci.2011.08.024. (Visited on 09/24/2024).

[6] Styliani Papatzani, Kevin Paine, and Juliana Calabria-Holley. "A Comprehensive Review of the Models on the Nanostructure of Calcium Silicate Hydrates". In: *Construction and Building Materials* 74 (Jan. 2015), pp. 219–234. ISSN: 0950-0618. DOI: 10.1016/j.conbuildmat.2014.10.029. (Visited on 03/16/2025).

[7] Mohammad Javad Abdolhosseini Qomi, Mathieu Bauchy, and Roland J.-M. Pellenq. "Nanoscale Composition-Texture-Property Relation in Calcium-Silicate-Hydrates". In: *Handbook of Materials Modeling: Applications: Current and Emerging Materials*. Ed. by Wanda Andreoni and Sidney Yip. Cham: Springer International Publishing, 2020, pp. 1761–1792. ISBN: 978-3-319-44680-6. DOI: 10.1007/978-3-319-44680-6_128. (Visited on 09/24/2024).

[8] Andrew J. Allen, Jeffrey J. Thomas, and Hamlin M. Jennings. "Composition and Density of Nanoscale Calcium–Silicate–Hydrate in Cement". In: *Nature Materials* 6.4 (Apr. 2007), pp. 311–316. ISSN: 1476-4660. DOI: 10.1038/nmat1871. (Visited on 08/10/2025).

[9] Jacqueline R. Houston, Robert S. Maxwell, and Susan A. Carroll. "Transformation of Meta-Stable Calcium Silicate Hydrates to Tobermorite: Reaction Kinetics and Molecular Structure from XRD and NMR Spectroscopy". In: *Geochemical Transactions* 10.1 (Jan. 2009), p. 1. ISSN: 1467-4866. DOI: 10.1186/1467-4866-10-1. (Visited on 08/10/2025).

[10] Jae Eun Oh et al. "Experimental Determination of Bulk Modulus of 14 Å Tobermorite Using High Pressure Synchrotron X-ray Diffraction". In: *Cement and Concrete Research* 42.2 (Feb. 2012), pp. 397–403. ISSN: 0008-8846. DOI: 10.1016/j.cemconres.2011.11.004. (Visited on 04/10/2025).

[11] Emmy M. Foley, Jung J. Kim, and M. M. Reda Taha. "Synthesis and Nano-Mechanical Characterization of Calcium-Silicate-Hydrate (C-S-H) Made with 1.5 CaO/SiO2 Mixture". In: *Cement and Concrete Research* 42.9 (Sept. 2012), pp. 1225–1232. ISSN: 0008-8846. DOI: 10.1016/j.cemconres.2012.05.014. (Visited on 06/05/2025).

[12] Riccardo Maddalena et al. "Direct Synthesis of a Solid Calcium-Silicate-Hydrate (C-S-H)". In: *Construction and Building Materials* 223 (Oct. 2019), pp. 554–565. ISSN: 0950-0618. DOI: 10.1016/j.conbuildmat.2019.06.024. (Visited on 06/05/2025).

[13] Roland J.-M. Pellenq et al. "A Realistic Molecular Model of Cement Hydrates". In: *Proceedings of the National Academy of Sciences* 106.38 (Sept. 2009), pp. 16102–16107. DOI: 10.1073/pnas.0902180106. (Visited on 09/11/2024).

[14] M. J. Abdolhosseini Qomi et al. "Combinatorial Molecular Optimization of Cement Hydrates". In: *Nature Communications* 5.1 (Sept. 2014), p. 4960. ISSN: 2041-1723. DOI: 10.1038/ncomms5960. (Visited on 09/20/2024).

[15] I.G. Richardson. "Model Structures for C-(A)-S-H(I)". In: *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* 70.6 (2014), pp. 903–923. DOI: 10.1107/S2052520614021982.

[16] M. Bauchy et al. "Order and Disorder in Calcium-Silicate-Hydrate". In: *Journal of Chemical Physics* 140.21 (2014). DOI: 10.1063/1.4878656.

[17] Goran Kovačević et al. "Revised Atomistic Models of the Crystal Structure of C–S–H with High C/S Ratio". In: *Zeitschrift für Physikalische Chemie* 230.9 (Sept. 2016), pp. 1411–1424. ISSN: 2196-7156. DOI: 10.1515/zpch-2015-0718. (Visited on 09/20/2024).

[18] Aslam Kunhi Mohamed et al. "An Atomistic Building Block Description of C-S-H - Towards a Realistic C-S-H Model". In: *Cement and Concrete Research* 107 (May 2018), pp. 221–235. ISSN: 0008-8846. DOI: 10.1016/j.cemconres.2018.01.007. (Visited on 09/17/2024).

[19] Mohammad Javad Abdolhosseini Qomi, Franz-Josef Ulm, and Roland J.-M. Pellenq. "Physical Origins of Thermal Properties of Cement Paste". In: *Physical Review Applied* 3.6 (June 2015), p. 064010. DOI: 10.1103/PhysRevApplied.3.064010. (Visited on 09/11/2024).

[20] Ashraf A. Bahraq et al. "Molecular Simulation of Cement-Based Materials and Their Properties". In: *Engineering* 15 (Aug. 2022), pp. 165–178. ISSN: 2095-8099. DOI: 10.1016/j.eng.2021.06.023. (Visited on 09/17/2024).

[21] Byoung Hooi Cho, Wonseok Chung, and Boo Hyun Nam. "Molecular Dynamics Simulation of Calcium-Silicate-Hydrate for Nano-Engineered Cement Composites—A Review". In: *Nanomaterials* 10.11 (Nov. 2020), p. 2158. ISSN: 2079-4991. DOI: 10.3390/nano10112158. (Visited on 03/17/2025).

[22] Salim Barbhuiya and Bibhuti Bhusan Das. "Molecular Dynamics Simulation in Concrete Research: A Systematic Review of Techniques, Models and Future Directions". In: *Journal of Building Engineering* 76 (Oct. 2023), p. 107267. ISSN: 2352-7102. DOI: 10.1016/j.jobe.2023.107267. (Visited on 02/08/2025).

[23] *Machine Learning in Concrete Science: Applications, Challenges, and Best Practices | Npj Computational Materials*. URL: https://www.nature.com/articles/s41524-022-00810-x (visited on 11/21/2024).

[24] Keita Kobayashi et al. "Machine Learning Potentials for Tobermorite Minerals". In: *Computational Materials Science* 188 (Feb. 2021), p. 110173. ISSN: 0927-0256. DOI: 10.1016/j.commatsci.2020.110173. (Visited on 10/08/2024).

[25] Keming Zhu. "Performance Comparisons of NequIP and DPMD Machine Learning Interatomic Potentials for Tobermorites". In: *Computational Materials Science* 244 (Sept. 2024), p. 113212. ISSN: 0927-0256. DOI: 10.1016/j.commatsci.2024.113212. (Visited on 10/15/2024).

[26] F. Giustino. *Materials Modelling Using Density Functional Theory: Properties and Predictions*. Oxford University Press, 2014. ISBN: 978-0-19-966244-9. URL: https://books.google.com.ec/books?id=FzOTAwAAQBAJ.

[27] D.S. Sholl and J.A. Steckel. *Density Functional Theory: A Practical Introduction*. Wiley, 2023. ISBN: 978-1-119-84086-2. URL: https://books.google.com.ec/books?id=BRgEAAAQBAJ.

[28] E. Kaxiras. *Atomic and Electronic Structure of Solids*. Cambridge University Press, 2003. ISBN: 978-0-521-81010-4. URL: https://books.google.com.ec/books?id=WTL_vgbWpHEC.

[29] R.M. Martin, L. Reining, and D.M. Ceperley. *Interacting Electrons*. Cambridge University Press, 2016. ISBN: 978-0-521-87150-1. URL: https://books.google.com.ec/books?id=ch1CDAAAQBAJ.

[30] T. Helgaker, P. Jorgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, 2014. ISBN: 978-1-119-01955-8. URL: https://books.google.com.ec/books?id=lNVLBAAAQBAJ.

[31] D. Feng and G. Jin. *Introduction to Condensed Matter Physics, Volume 1*. World Scientific Publishing Company, 2005. ISBN: 978-981-310-219-4. URL: https://books.google.com.ec/books?id=IM47DQAAQBAJ.

[32] D. R. Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 24.1 (Jan. 1928), pp. 89–110. ISSN: 1469-8064, 0305-0041. DOI: 10.1017/S0305004100011919. (Visited on 07/10/2025).

[33] V. Fock. "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems". In: *Zeitschrift für Physik* 61.1 (Jan. 1930), pp. 126–148. ISSN: 0044-3328. DOI: 10.1007/BF01340294. (Visited on 07/10/2025).

[34] R.M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2020. ISBN: 978-1-108-42990-0. URL: https://books.google.com.ec/books?id=wvXvDwAAQBAJ.

[35] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136.3B (Nov. 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864. (Visited on 07/12/2025).

[36] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133. (Visited on 07/12/2025).

[37] W. Kohn. "Nobel Lecture: Electronic Structure of Matter—Wave Functions and Density Functionals". In: *Reviews of Modern Physics* 71.5 (1999), pp. 1253–1266. DOI: 10.1103/RevModPhys.71.1253.

[38] D. M. Ceperley and B. J. Alder. "Ground State of the Electron Gas by a Stochastic Method". In: *Physical Review Letters* 45.7 (Aug. 1980), pp. 566–569. DOI: 10.1103/PhysRevLett.45.566. (Visited on 07/19/2025).

[39] J. P. Perdew and Alex Zunger. "Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems". In: *Physical Review B* 23.10 (May 1981), pp. 5048–5079. DOI: 10.1103/PhysRevB.23.5048. (Visited on 07/19/2025).

[40] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Physical Review Letters* 77.18 (Oct. 1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865. (Visited on 07/19/2025).

[41] John P. Perdew et al. "Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces". In: *Physical Review Letters* 100.13 (Apr. 2008), p. 136406. DOI: 10.1103/PhysRevLett.100.136406. (Visited on 07/21/2025).

[42] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. "Hybrid Functionals Based on a Screened Coulomb Potential". In: *The Journal of Chemical Physics* 118.18 (May 2003), pp. 8207–8215. ISSN: 0021-9606. DOI: 10.1063/1.1564060. (Visited on 07/19/2025).

[43] Jonathan E. Moussa, Peter A. Schultz, and James R. Chelikowsky. "Analysis of the Heyd-Scuseria-Ernzerhof Density Functional Parameter Space". In: *The Journal of Chemical Physics* 136.20 (May 2012). ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.4722993. arXiv: 1205.4999 [cond-mat]. (Visited on 07/22/2025).

[44] John P. Perdew and Karla Schmidt. "Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy". In: *AIP Conference Proceedings* 577.1 (July 2001), pp. 1–20. ISSN: 0094-243X. DOI: 10.1063/1.1390175. (Visited on 07/19/2025).

[45] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009. ISBN: 978-1-139-47719-2. URL: https://books.google.com.ec/books?id=VRZUw8Wk4CIC.

[46] Thomas D. Kühne. "Ab-Initio Molecular Dynamics". In: *WIREs Computational Molecular Science* 4.4 (July 2014), pp. 391–406. ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.1176. arXiv: 1201.5945 [physics]. (Visited on 07/26/2025).

[47] R. P. Feynman. "Forces in Molecules". In: *Physical Review* 56.4 (1939), pp. 340–343. DOI: 10.1103/PhysRev.56.340.

[48] Peter Politzer and Jane S. Murray. "The Hellmann-Feynman Theorem: A Perspective". In: *Journal of Molecular Modeling* 24.9 (Aug. 2018), p. 266. ISSN: 0948-5023. DOI: 10.1007/s00894-018-3784-7. (Visited on 07/28/2025).

[49] M.E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts. Oxford University Press, 2023. ISBN: 978-0-19-882556-2. URL: https://books.google.com.ec/books?id=EEPJEAAAQBAJ.

[50] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2023. ISBN: 978-0-323-91318-8. URL: https://books.google.com.ec/books?id=jyipEAAAQBAJ.

[51] Jürgen Hafner. "Ab-Initio Simulations of Materials Using VASP: Density-functional Theory and Beyond". In: *Journal of Computational Chemistry* 29.13 (2008), pp. 2044–2078. DOI: 10.1002/jcc.21057. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21057.

[52] H. Hellmann. "A New Approximation Method in the Problem of Many Electrons". In: *The Journal of Chemical Physics* 3.1 (Jan. 1935), p. 61. ISSN: 0021-9606. DOI: 10.1063/1.1749559. (Visited on 07/30/2025).

[53] Hendrik J. Monkhorst and James D. Pack. "Special Points for Brillouin-zone Integrations". In: *Physical Review B* 13.12 (June 1976), pp. 5188–5192. DOI: 10.1103/PhysRevB.13.5188. (Visited on 08/01/2025).

[54] P. E. Blöchl. "Projector Augmented-Wave Method". In: *Physical Review B* 50.24 (Dec. 1994), pp. 17953–17979. DOI: 10.1103/PhysRevB.50.17953. (Visited on 07/30/2025).

[55] G. Kresse and D. Joubert. "From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method". In: *Physical Review B* 59.3 (Jan. 1999), pp. 1758–1775. DOI: 10.1103/PhysRevB.59.1758. (Visited on 07/30/2025).

[56] Francis Birch. "Finite Elastic Strain of Cubic Crystals". In: *Physical Review* 71.11 (June 1947), pp. 809–824. DOI: 10.1103/PhysRev.71.809. (Visited on 07/30/2025).

[57] J.P. Poirier. *Introduction to the Physics of the Earth's Interior*. Introduction to the Physics of the Earth's Interior. Cambridge University Press, 2000. ISBN: 978-0-521-66392-2. URL: https://books.google.com.ec/books?id=_YOCl86XPHoC.

[58] Ryosuke Jinnouchi, Ferenc Karsai, and Georg Kresse. "On-the-Fly Machine Learning Force Field Generation: Application to Melting Points". In: *Physical Review B* 100.1 (July 2019), p. 014105. DOI: 10.1103/PhysRevB.100.014105. (Visited on 07/31/2025).

[59] Ryosuke Jinnouchi et al. "Descriptors Representing Two- and Three-Body Atomic Distributions and Their Effects on the Accuracy of Machine-Learned Inter-Atomic Potentials". In: *The Journal of Chemical Physics* 152.23 (June 2020). ISSN: 0021-9606. DOI: 10.1063/5.0009491. (Visited on 07/31/2025).

[60] B. Hourahine et al. "DFTB+, a Software Package for Efficient Approximate Density Functional Theory Based Atomistic Simulations". In: *The Journal of Chemical Physics* 152.12 (Mar. 2020). ISSN: 0021-9606. DOI: 10.1063/1.5143190. (Visited on 07/31/2025).

[61] *The VASP Manual*. URL: https://www.vasp.at/wiki/index.php/The_VASP_Manual (visited on 07/31/2025).

[62] *Available Pseudopotentials - VASP Wiki*. URL: https://www.vasp.at/wiki/index.php/Available_pseudopotentials (visited on 08/05/2025).

[63] G. Kresse. "Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set". In: *Physical Review B* 54.16 (1996), pp. 11169–11186. DOI: 10.1103/PhysRevB.54.11169.

[64] Laurids Schimka, Judith Harl, and Georg Kresse. "Improved Hybrid Functional for Solids: The HSEsol Functional". In: *The Journal of Chemical Physics* 134.2 (Jan. 2011), p. 024116. ISSN: 1089-7690. DOI: 10.1063/1.3524336.

[65] C. C. Dharmawardhana, A. Misra, and Wai-Yim Ching. "Theoretical Investigation of C-(A)-S-H(I) Cement Hydrates". In: *Construction and Building Materials* 184 (Sept. 2018), pp. 536–548. ISSN: 0950-0618. DOI: 10.1016/j.conbuildmat.2018.07.004. (Visited on 08/11/2025).

[66] *Category:Machine-learned Force Fields - VASP Wiki*. URL: https://www.vasp.at/wiki/index.php/Category:Machine-learned_force_fields (visited on 08/11/2025).

[67] J.J. Gilman. *Electronic Basis of the Strength of Materials*. Cambridge University Press, 2003. ISBN: 978-1-139-43518-5.

[68] Jae Eun Oh, Simon M. Clark, and Paulo J. M. Monteiro. "Does the Al Substitution in C–S–H(I) Change Its Mechanical Property?" In: *Cement and Concrete Research* 41.1 (Jan. 2011), pp. 102–106. ISSN: 0008-8846. DOI: 10.1016/j.cemconres.2010.09.010. (Visited on 10/06/2025).

[69] H. Xu et al. "Anisotropic Thermal Expansion and Hydrogen Bonding Behavior of Portlandite: A High-Temperature Neutron Diffraction Study". In: *Journal of Solid State Chemistry* 180.4 (Apr. 2007), pp. 1519–1525. ISSN: 0022-4596. DOI: 10.1016/j.jssc.2007.03.004. (Visited on 08/12/2025).