2022-05-01

Phonon Thermodynamics of bcc Zirconium with Machine Learning

Vanessa Judith Meraz
The University of Texas at El Paso

Phonon Thermodynamics of bcc Zirconium with Machine Learning

Vanessa Judith Meraz

Master's Program in Physics

APPROVED:

Jorge A. Muñoz, Ph.D., Chair

Wibe A. de Jong, Ph.D.

Ramon Ravelo, Ph.D.

Olac Fuentes, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

Phonon Thermodynamics of bcc Zirconium with Machine Learning

by

Vanessa Judith Meraz, B.S.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Physics

THE UNIVERSITY OF TEXAS AT EL PASO

May 2022

Abstract

First principles-based simulations have allowed us to explore emerging phenomena in a variety of systems. Its steadfast practicality has led to an increase in molecular and materials design ranging from drug discovery to planetary formation. However ubiquitous in its field, one of its biggest drawbacks is its computational cost, notably so in molecular dynamics simulations. To counter this setback, there have been many leading efforts in machine learning methods, whether it be in algorithms or network architectures. Our contribution uses an active learning algorithm paired with a tensor field network, e3nn. By steadily feeding new data points to our model, based on data it struggles with, our model gains a better idea of the system. First, using Born-Oppenheimer molecular dynamics, we simulate body-centered cubic zirconium with a $(4 \times 4 \times 4)$ supercell of 128 atoms at 1500K for 800 time steps. This DFT model is subsequently fed into the network where we define energies as our input and forces as our output. The active learning runs that follow reshuffle the top 5% errors of the test set into the train set. Several runs later, we analyze the accuracy of our trained model by generating phonon dispersions and comparing their structure to that of the original DFT data.

Table of Contents

List of Figures

List of Tables

Chapter 0

Motivation

Over the last century, there have been considerable efforts in researching arising phenomena in condensed matter, much of which has been spent on designing new materials. By exploiting known molecular properties, we have successfully redesigned generations of electronic devices. Because of this, it is fundamental to understand the interatomic interactions of a system to gain a sense of its underlying mechanisms. In doing so, this process has been made easier by calculating the properties of an atomic structure, whose current framework relies on density functional theory (DFT). Even so, this quantum mechanical approximation of molecular properties requires corrections. Historically, for accurate calculations, the Hartree-Fock (HF) method was used to solve for the electronic wavefunction, albeit it faltered for large enough systems. Although time has improved upon both DFT and HF, there are still glaring limitations in their application. For large molecular systems, this would be the computational complexity: As the number of atoms in the system increases so does the computational cost.

To accurately calculate the properties of a system, it is necessary to use a many-body approach. Yet for a good amount of systems, using this approach to model critical phenomena with high accuracy is still a ways away. In spite of this relative low inaccuracy, the more pressing issue is the scalability of a system, which is where machine learning offers a solution [1]. Currently, there is an abundance of methods yielding highly accurate results [2, 3]. For molecular systems, it is most common to use graph neural networks [4] to represent a system. Even though it's known that the arrangement of atoms in a system affects its properties, it is not very commonly modeled in neural network architecture. Remedying this are some symmetry aware networks in Euclidean space [5, 6, 7].

As the physical properties of a system have deterministic transformations, where conserved quantities are invariant, it has become increasingly apparent that equivariance is a necessity for the prediction of geometrical objects [8, 9]. Though it is common to interchange invariance and equivariance, they do have distinct meanings. Mainly, equivariance takes on a distortion with respect to the symmetric operation performed on an object e.g., if an image input is rotated, an invariant function wouldn't notice a difference but an equivariant function would know the shift difference. In molecular systems, this is an advantage as the recognition of a rotated object can cut down on data augmentation for training data. For the system investigated in this thesis, the equivariant Euclidean (3) Neural Networks [10] are used with a custom algorithm to accurately perform predictions at a low cost.

0.1 Structure

This thesis is split into two parts. The first builds upon the background of our simulations. Chapters 1 and 2 include a detailed description of the theoretical background of the project. Chapter 1's format is written with the introduction to groups and group representation as the basis for explaining the symmetric significance in crystals and later on its application for neural networks. Chapter 3 discusses the results of DFT simulations with phonon dispersions. The second part introduces the practicality of using machine learning for materials science. Chapter 4 provides the background of the algorithm we're using as well as the neural network. Chapter 5 discusses the results of trained models, errors, predictions, and the comparison to DFT phonon dispersions.

Chapter 1

Geometric Symmetries and Lattice Dynamics

The innate nature of particles forming symmetric patterns that are translationally invariant is the foundation of condensed matter physics. In 1912, after the discovery of crystal X-ray diffraction by M. Laue [11], W. L. and W. H. Bragg observed diffraction peaks to construct crystal structures [12][13]. The resulting so-called Bragg diffraction pattern is a consequence of the periodic arrangement of atoms, or group of atoms, in a crystal. Moreover, the (unusual) electronic properties of said crystal are accentuated and able to be studied.

1.1 Groups

The concept and consequence of symmetry in physical systems can be explained through tensor algebra and group theory. In solids, these symmetries can be used to determine the properties of a system. In turn, these crystal structure symmetries can be defined in terms of symmetric operations. Represented by individual elements, these operations form a *group*. Formally, a group is defined as a mathematical set which combines any two elements with an operation to form a third element.

Definition 1. *A group, G, is a set of elements g_n which satisfy the following [14]:*

Closure *If $g_i, g_j \in G$ then $g_k = g_i g_j \in G$*

Associativity *For $g_i,\ g_j,\ g_k \in G,\ g_i(g_j g_k) = (g_i g_j)g_k$*

Identity $\exists\ E : \forall\ g_n \in G,\ Eg_n = g_nE = g_n$

Inverse *Every element* $g_n \in G$ *has an inverse,* $g_n^{-1} : g_n^{-1}g_n = g_ng_n^{-1} = E$

Generally, as part of the closure condition, it is not required for the multiplication of elements to be commutative. However, such an instance describes a group as abelian; and a noncommutative composition as nonabelian.

Subgroups

A subset H of group G is considered a subgroup if H forms a group.

1.1.1 Group Representations

A representation D of group G is a mapping of the group elements g onto a set with the following properties [14]:

Group product $D(g_i)D(g_j) = D(g_ig_j)\ \forall\ g_i, g_j \in G$

Identity $D(Eg_i)D(g_iE) = D(g_i)$

Irreducible Representations

Irreducible representations, or irreps, are the smallest possible representations of an algebraic structure. If representation $D'(g)$ is related to $D(g)$ through a similarity transformation:

$$D'(g) = P^{-1}D(g)P \tag{1.1}$$

then a block-diagonal matrix can be constructed such that:

$$D'(g) = \begin{bmatrix} D^{(1)}(g) & 0 & \dots & 0 \\ 0 & D^{(2)}(g) & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & D^{(k)}(g) \end{bmatrix} \tag{1.2}$$

In the case of $D'(g)$, it is reducible and can be written as a sum of its representations:

$$D'(g) = D^{(1)}(g) \oplus D^{(2)}(g) \oplus \cdots \oplus D^{(k)}(g) \tag{1.3}$$

If representations $D^{(k)}(g)$ are no longer able to be reduced they are said to be irreducible.

1.1.2 Invariance and Equivariance

A function $f : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{Y}}$ is said to be equivariant $\forall\ g \in G$ and $x \in \mathcal{X}$ for matrix representation D(g) if:

$$f(D^{\mathcal{X}}(g)) = D^{\mathcal{Y}}(g) f(x) \tag{1.4}$$

where the function f commutes with group elements g; i.e., the symmetry operation(s) can be applied before or after evaluating the function.

The more commonly studied property of invariance is a type of equivariance where $D^{\mathcal{Y}}(g) = D^{\mathcal{X}}(g) = \mathbb{1}$, or more abstractly: a function whose value is unchanged under an operation.

1.1.3 Euclidean Space

A point in three-dimensional space can be most simply described as a tuple of three numbers. In real coordinate space, this point is referred to as being in three-dimensional *Euclidean* space, wherein there are three fundamental motions: rotation, translation, and reflection. Known as Euclidean isometries, these transformations preserve the Euclidean distance between any two points.

The rigid transformations of Euclidean space form a group called the Euclidean group, or E(n) for n-dimension(s).

3D Rotations

A group of proper rotations in E(3) is referred to as SO(3), a **s**pecial **o**rthogonal group. The orthonormal condition of this subgroup can be expressed as:

$$Q^T Q = Q Q^T = I \tag{1.5}$$

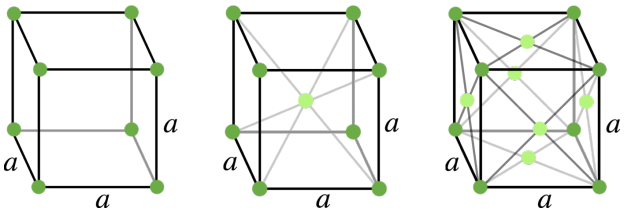
where I is the identity matrix and Q is an orthogonal matrix. It follows from this condition that the determinant of matrix Q is $\pm$ 1. Matrices with determinant 1 form the subgroup SO(n) of direct isometries which preserve chirality and are called proper. The direct isometries themselves form another subgroup called SE(n), the **s**pecial **E**uclidean group which consist of rotations, translations, and combinations thereof. It should be noted that the preservation of chirality excludes reflections from this group.

1.2 Crystal Structure

Symmetry groups

An operation on a geometric object is considered symmetric if it is invariant. The group of transformations in which this is true is called the symmetry group.

Crystal structures in three dimensions are classified by the following symmetry groups: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic. These seven Bravais lattice symmetry groups combine to construct the 14 known Bravais lattices. Each of these lattices can be represented by a point group which is defined as a set of symmetric transformations with at least one lattice point in common. From the crystallographic restriction of n-fold symmetries, there are 32 point groups that leave the crystal invariant.



Cubic symmetry group. From left to right: simple cubic, body-centered cubic, face-centered cubic.

Figure 1.1:

Lattice group

As a geometric term, a lattice is a group of points on which a set of operations produces another point. More conceptually, it can be described as a symmetric group of (discrete) translational symmetry.

As a crystalline structure basis, a lattice is an n-dimensional array with a periodic arrangement of atoms produced by primitive translation vectors: $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$. Where in three dimensions, the translation operations are described by:

$$\mathbf{R} = n_1\mathbf{a_1} + n_2\mathbf{a_2} + n_3\mathbf{a_3} \tag{1.6}$$

When viewed from point $\mathbf{r}$, such an arrangement will look the same as when viewed from point $\mathbf{r}'$:

$$\mathbf{r}' \longrightarrow \mathbf{r} + \mathbf{R} \tag{1.7}$$

As with translation, any operation performed on the crystal that is isomorphic to itself is part of the symmetry group for that crystal.

Primitive Unit Cell

The smallest possible representation of an ideal crystal is the (primitive) unit cell. Known also as a minimum-volume cell, the primitive unit cell contains one lattice point and the smallest basis for repetition. Its volume is defined by the primitive translation vectors as:

$$\mathbf{a_1} \cdot \mathbf{a_2} \times \mathbf{a_3} \tag{1.8}$$

In the case of a body-centered cubic lattice, the primitive translation vectors $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$ can be represented as:

$$\mathbf{a_1} = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}}) \qquad \mathbf{a_2} = \frac{1}{2}a(-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}) \qquad \mathbf{a_3} = \frac{1}{2}a(\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}) \tag{1.9}$$

or in matrix form:

$$\mathrm{A} = \frac{a}{2}\begin{bmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

where a represents the side of the cube.

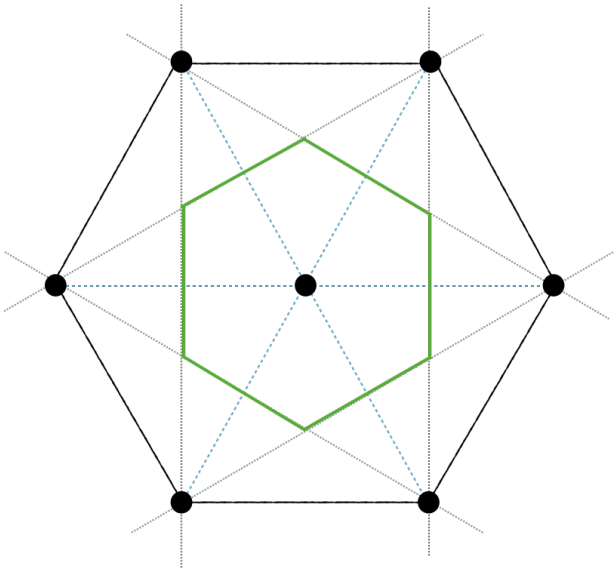


Figure 1.2: Wigner-Seitz cell of a hexagonal lattice.

One such case of a primitive unit cell, shown in figure 1.2, is the Wigner-Seitz cell. Picking the lattice point to be in the center of a hexagonal lattice, lines are then drawn to the nearest lattice point (in blue). A normal to these lines connecting two lattice points (in grey) then encloses the minimum-volume cell (in green).

Reciprocal lattice

In reciprocal space, there exists a reciprocal lattice represented by the Fourier transform of an original lattice. It is a mathematical representation of visual Bragg diffraction patterns caused by wave scattering.

As the crystal is invariant under translation operations, this is also required for its individual properties. Such that:

$$f(\mathbf{r} + \mathbf{R}) = f(\mathbf{r}) \tag{1.10}$$

where the fourier series $f(\mathbf{r})$ is a periodic function and its expansion with reciprocal wavevectors $\mathbf{G}$ (1.11), must be translationally invariant.

$$f(\mathbf{r}) = \sum_m f_m e^{i\mathbf{G}_m \cdot \mathbf{r}} \tag{1.11}$$

Similarly to translation vector $\mathbf{R}$ (1.6), the reciprocal lattice vector can be represented as:

$$\mathbf{G}_m = m_1\mathbf{b_1} + m_2\mathbf{b_2} + m_1\mathbf{b_2} \tag{1.12}$$

where reciprocal axis vectors $\mathbf{b_i}$ can be constructed using the primitive axis vectors satisfying the property: $\mathbf{b_i} \cdot \mathbf{a_j} = 2\pi\delta_{ij}$.

$$\mathbf{b_1} = 2\pi\frac{\mathbf{a_2} \times \mathbf{a_3}}{\mathbf{a_1} \cdot \mathbf{a_2} \times \mathbf{a_3}} \qquad \mathbf{b_2} = 2\pi\frac{\mathbf{a_3} \times \mathbf{a_1}}{\mathbf{a_1} \cdot \mathbf{a_2} \times \mathbf{a_3}} \qquad \mathbf{b_3} = 2\pi\frac{\mathbf{a_1} \times \mathbf{a_2}}{\mathbf{a_1} \cdot \mathbf{a_2} \times \mathbf{a_3}} \tag{1.13}$$

Brillouin Zone

The Brillouin zone is defined as a Wigner-Seitz cell of the reciprocal lattice. In figure 1.2 the minimum-volume cell enclosed in green is also known as the first Brillouin zone.

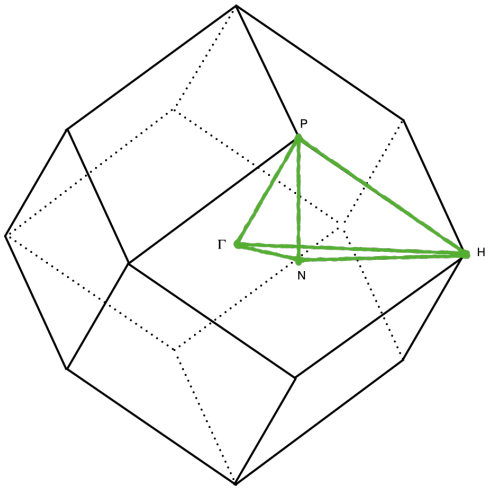
As the reciprocal lattice is the representation of Bragg diffraction, the first Brillouin zone is the set of points from the origin that don't cross a Bragg plane. Although there are n Brillouin zones, the first is sufficient to analyze the electronic structure of a system.

Scattering condition 1.16, provides an alternative for Bragg condition $n\lambda = 2d\sin\theta$ where the projection of wavevector $\mathbf{k}$ along the reciprocal lattice vector $\mathbf{G}$ lies in the Brillouin zone.

$$\mathbf{k} = \frac{2\pi}{\lambda}\hat{n} \tag{1.14}$$

$$\mathbf{k}' = \frac{2\pi}{\lambda}\hat{n}' \tag{1.15}$$

$$\mathbf{k} = \mathbf{k}' - \mathbf{G} \tag{1.16}$$



Brillouin zone of a body-centered cubic lattice with labeled high symmetry points. (Γ = zone center, N = face center, H = corner, P = point with three adjoining faces)

Figure 1.3:

High symmetry points

There are several points of interest in the Brillouin zone of a lattice structure. Their grouping is known as an irreducible Brillouin zone since it's a wedge that can be duplicated to replicate the reciprocal lattice. The main interest of this manuscript lies in the body-centered cubic structure as seen in figure 1.3. The symmetry points for a bcc structure are as follows:

$$\Gamma = (0,0,0) \qquad H = \frac{1}{2}(-1,1,0) \qquad P = \frac{1}{4}(1,1,1) \qquad N = \frac{1}{2}(0,1,0) \tag{1.17}$$

1.2.1 Phonons and Anharmonicity

When disturbed, the equilibrium interatomic distances for a system of particles produce vibrations. In the simple one-dimensional realm, a monatomic lattice can be modeled as follows:

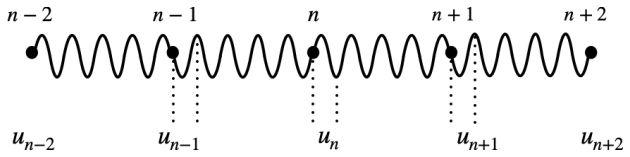


Figure 1.4: Linear monatomic chain with periodicity

where u represents the displacements of nth atoms of m mass a distance a apart. The imposed condition here is the Born-von Kármán boundary such that the wavefunction is invariant to a translation (1.18).

$$\psi(\mathbf{r} + \mathbf{R}) = \psi(\mathbf{r}) \tag{1.18}$$

Then, the motion of the nth atom can be described by Newton's second law and Hooke's Law as:

$$m\frac{d^2u_n}{dt^2} = C[(u_{n+1} - u_n) + (u_{n-1} - u_n)] \tag{1.19}$$

where C is the nearest neighbor force constant. The solution for this equation then follows a form:

$$u_n = \varepsilon e^{i(kx-\omega t)} \tag{1.20}$$

which is a general solution function to a wave. For $x = na$, where a is the spacing between planes n, 1.19 and 1.20 combine and reduce to:

$$m\omega^2 = 2C[\cos(ka) - 1] \tag{1.21}$$

where rewriting in terms of frequency, we have the dispersion relation ω:

$$\omega^2 = \frac{2C}{m}[\cos(ka) - 1] \tag{1.22}$$

Note here that the solution modes are either longitudinal or transversal, and hence, a (1×1) matrix. For systems of higher dimensions, solution 1.22 can be rewritten in terms of dynamical matrix D as an eigenvalue problem:

$$\begin{aligned} \omega^2 \varepsilon &= \frac{2C}{m}[\cos(ka) - 1]\epsilon \\ &= D(k)\varepsilon \end{aligned}$$

where dynamical matrix $D(k)$ contains the force constant matrices for every atomic interaction.

In the boundary of the first Brillouin zone, $k = \pm\pi/a$, the dispersion curve shows clear translational symmetry as seen in figure 1.5. While it is possible to analyse points beyond this range, the subsequent lattice motion would only replicate previously seen values. Thus any displacement u can be written in terms of k within the first zone.

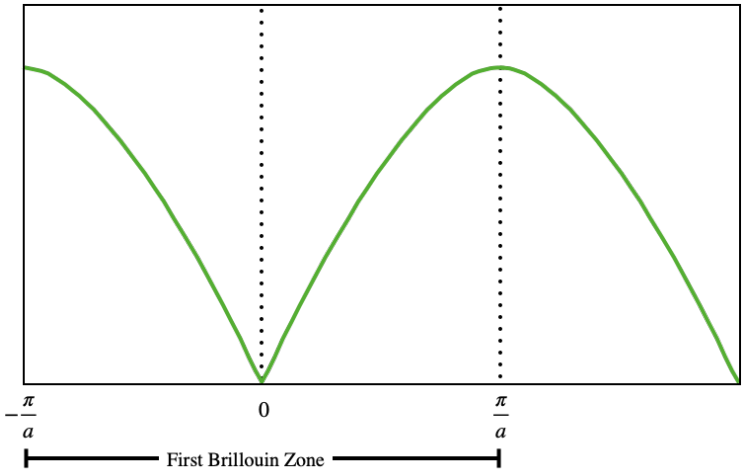


Figure 1.5: Brillouin zone boundary of dispersion relation $\omega(k)$ at k $= \pm\pi/a$

Harmonic approximation

Given a crystal structure, a pair of atoms contribute an energy $\phi(\mathbf{R})$ for a separation $\mathbf{r}$. Its potential energy is then defined as a sum of all unique pair contributions.

$$U = \frac{1}{2}\sum_{i,j}\phi(\mathbf{r}_i - \mathbf{r}_j) \tag{1.23}$$

If an atom in its equilibrium position is disturbed, the imposed forces surrounding that atom will attempt to push it back into equilibrium. For a small enough perturbation, a Taylor expansion of the potential energy about its equilibrium is a useful approximation.

$$U = \frac{1}{2}\sum_{i,j}\phi(\mathbf{r}_i - \mathbf{r}_j) + \frac{1}{2}\sum_{i,j}(\mathbf{u}_i - \mathbf{u}_j)\cdot\nabla\phi(\mathbf{r}_i - \mathbf{r}_j) + \frac{1}{4}\sum_{i,j}[(\mathbf{u}_i - \mathbf{u}_j)\cdot\nabla]^2\phi(\mathbf{r}_i - \mathbf{r}_j) + \mathcal{O}(\mathbf{u}^3) \tag{1.24}$$

The first term in equation 1.24 is a vanishing term as it's the equilibrium potential energy. The harmonic term is then reduced to:

$$\begin{aligned} U_{harm} &= \sum_{\substack{i,j \\ \alpha,\beta=x,y,x}} (\mathbf{u}_i - \mathbf{u}_j)_\alpha \phi_{\alpha,\beta}(\mathbf{r}_i - \mathbf{r}_j)(\mathbf{u}_i - \mathbf{u}_j)_\beta \\ &= \sum_{\substack{i,j \\ \alpha,\beta=x,y,x}} (\mathbf{u}_i - \mathbf{u}_j)_\alpha \frac{\partial^2(\mathbf{r}_i - \mathbf{r}_j)}{\partial(\mathbf{r}_i - \mathbf{r}_j)_\alpha \partial(\mathbf{r}_i - \mathbf{r}_j)_\beta}(\mathbf{u}_i - \mathbf{u}_j)_\beta \end{aligned}$$

where neglecting higher order terms, prevents us from analysing the anharmonic effects of a system. It should be noted that equation above, while perhaps unfamiliar, is akin to elastic potential energy:

$$U = \frac{1}{2}\sum_i kx_i^2 \tag{1.25}$$

where:

$$U_{harm} = \frac{1}{2}\sum_{\substack{i,j \\ \alpha,\beta=x,y,x}} (\mathbf{u}_i - \mathbf{u}_j)_\alpha \mathbf{k}(\mathbf{u}_i - \mathbf{u}_j)_\beta \tag{1.26}$$

for a force constant matrix $\mathbf{k}$. In general, the equations of motion are written as:

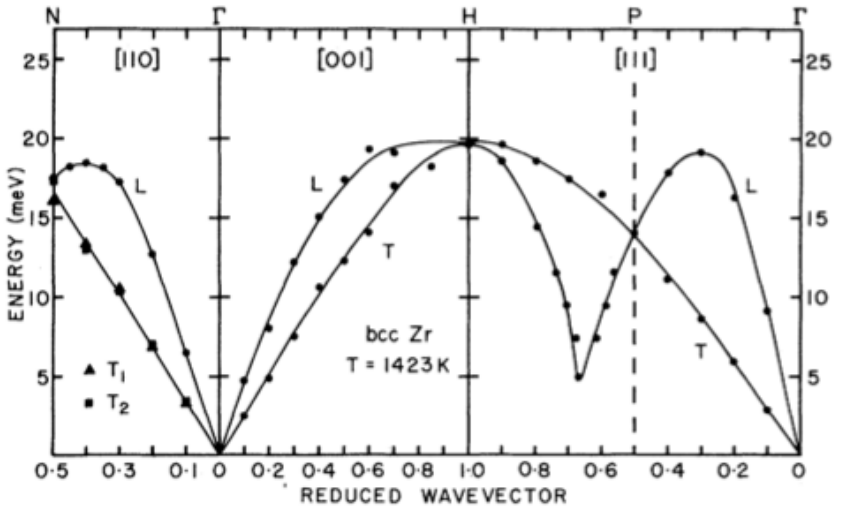
$$m\ddot{u_\alpha}(\mathbf{i},\mathbf{j}) = -\sum_j \phi_{\alpha,\beta}(\mathbf{r}_i - \mathbf{r}_j)u_\beta(\mathbf{i},\mathbf{j}) \tag{1.27}$$

which follow the same solution of the one-dimensional problem from 1.19.

Phonon Dispersions

The study of phonons give an insight into the thermodynamic behavior of a material. Most often, inelastic neutron scattering experiments provide phonon dispersions, ω, that when plotted as band structures, provide a wealth of information. Simply put, these structures describe the allowed energy levels of a crystalline solid.

The idea of the electronic band structure centers on the overlapping orbitals of nearby atoms. This interaction in a tightly-bound model causes the formation of bands from several discrete energy levels. As the main cause are outer electrons, the bands can reflect on electronic behavior such as conductivity and binding.



Electronic band structure for bcc Zr from inelastic neutron scattering [15].

Figure 1.6:

In the case of bcc Zr as shown in figure 1.6, there are three acoustic branches in the $\mathrm{N}-\Gamma$ section. The two lowest energy branches are called transverse and the highest energy branch, longitudinal.

Anharmonicity

In the harmonic approximation, an assumption is made that a lattice vibrates about its equilibrium at relatively small displacements, which is valid at low temperatures. While this model produces reasonable results, it neglects the finer details of temperature dependent phenomena. Thus, the previously labeled *small* anharmonic effect is necessary for understanding a temperature dependent system; or rather any real system, since it's inaccurate to model a solid with static ions and non-interacting phonons. Experimentally, this is also seen in inelastic neutron scattering where the expected sharp phonon peaks appear to have some width, strengthening the indication that an anharmonic treatment of a crystal is significant.

Realistically, at a finite temperature, the effects of thermal expansion and conductivity cannot be calculated with a harmonic approximation (which doesn't include temperature in calculating the thermal quantities of a system). To do this, the most common approach is using an adiabatic approximation, which uses the same principle as the Born-Oppenheimer approximation (2.5).

Chapter 2

Density Functional Theory

A computational modelling method for molecules and condensed-matter, density functional theory (DFT) relays predictions on fundamental properties of materials. As the basis for an accurate model of electronic behavior, DFT is used heavily. Its starting point relies on the non-relativistic time-dependent Schrödinger equation:

$$\hat{H}\,|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}\,|\psi(t)\rangle \tag{2.1}$$

Typically, an atomistic system (barring hydrogen) consists of multiple electrons and depending on the model, multiple nuclei. A system with N electrons and M nuclei is accurately described by the many-body Schrödinger equation as:

$$\hat{H}\boldsymbol{\Psi} = \left\{ -\sum_{i}^{M}\frac{1}{2}\nabla_{\alpha}^{2} - \sum_{i}^{N}\frac{1}{2}\nabla_{i}^{2} - \sum_{i,\alpha}^{N,M}\frac{Z_{\alpha}}{|\mathbf{r}_i - \mathbf{R}_{\alpha}|} + \sum_{i,j\neq i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha,\beta\neq\alpha}^{M}\frac{1}{2}\frac{Z_{\alpha}Z_{\beta}}{|\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}|} \right\}\boldsymbol{\Psi} \tag{2.2}$$

or:

$$\hat{H}\boldsymbol{\Psi} = \left\{ \hat{T}_n + \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \right\}\boldsymbol{\Psi} \tag{2.3}$$

where $\hat{T}_n$ is the kinetic energy due to the nuclei, $\hat{T}_e$ is due to the electrons, $\hat{V}_{en}$ is the Coulomb attraction, $\hat{V}_{ee}$ is the electron-electron repulsion, and $\hat{V}_{nn}$ is the nuclei-nuclei repulsion. As a formalism, equation 2.2 has been written with atomic units (2.4) in mind.

$$\hbar = e = a_0 = m_e = 1 \tag{2.4}$$

The main spectrum of this work focuses on simulated dynamics which employ the Born-Oppenheimer approximation. In order to scale-down the complexity of a system, we make a few assumptions: (1) The nuclei of a system is considerably heavier than its electrons and

(2) The relative dynamics of a nuclei are stable and virtually nonexistent in comparison to electrons. These assumptions will allow us to treat the nuclei and electrons as two separate systems. Effectively, the many-body Schrödinger equation will be reduced to:

$$\hat{H}\mathbf{\Psi} = \left\{ -\sum_{i}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i,\alpha}^{N,M} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i,j\neq i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha,\beta\neq\alpha}^{M} \frac{1}{2}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \right\}\mathbf{\Psi} \tag{2.5}$$

or:

$$\hat{H}\mathbf{\Psi} = \left\{ \hat{T} + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \right\}\mathbf{\Psi} \tag{2.6}$$

where we no longer have kinetic operator $\hat{T}_n$ to describe the nuclei.

Hohenberg-Kohn theorems

A main feature of DFT is the calculation of electronic density $n(\mathbf{r})$ instead of the wave-function ψ. This framework relies on two theorems by Hohenberg and Kohn [16]. First, for a system, the external potential $\hat{V_{en}}$ is a functional of the (ground state) electron density such that:

$$E[n(\mathbf{r})] = \int n(\mathbf{r})\hat{V_{en}} d\mathbf{r} + F[n(\mathbf{r})] \tag{2.7}$$

where $F[n(\mathbf{r})]$ is the Hohenberg-Kohn functional, although its form is not known. Second, the density functional of the system is the ground state density if it minimizes the total energy of the system. Thus finding the ground state energy of a system is possible by the ground state electron density.

As it wasn't possible to explicitly write $F[n(\mathbf{r})]$, Kohn and Sham [17] developed a method in which a fictitious system was introduced to solve this problem. In this approach, the ground state electron density of a non-interacting system is the same as that of an interacting system leading to the ground state energy of a many-body interacting system. The ground state energy functional is then rewritten to:

$$E_{KS}[n(\mathbf{r})] = T_s[n(\mathbf{r})] + \int n(\mathbf{r})\hat{V_{en}} d\mathbf{r} + E_{Hartree}[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] \tag{2.8}$$

where $T_s[n(\mathbf{r})]$ is the kinetic energy of the non-interacting system, $E_{Hartree}[n(\mathbf{r})]$ is the coulomb energy of electronic density, and $E_{xc}[n(\mathbf{r})]$ is the exchange correlation energy.

Exchange correlation functionals

The Kohn-Sham equations, while a better attempt at solving the ground state energy of a system, are not yet solvable without a form for the exchange correlation, $E_{xc}[n(\mathbf{r})]$. There are a variety of density functionals, all of which have varying terms of accuracy and computational efficiency. The most common, Local-density Approximation (LDA), is expressed as:

$$E_{xc}^{LDA}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}))d\mathbf{r} \tag{2.9}$$

where ϵ_{xc} is the exchange correlation energy for a homogeneous electron gas of charge density n. More sophisticated methods include the generalized gradient approximation (GGA) whose functional uses charge density and its gradient as follows:

$$E_{xc}^{GGA}[n(\mathbf{r})] = \int \epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}))d\mathbf{r} \tag{2.10}$$

The Perdew, Burke, Ernzerhof (PBE) functional [18] used as an improvement to GGA is widely used for its accuracy. Many hybrid functional methods improve upon non-hybrid functionals. However, there is still an under and overestimation of electronic properties, most notably in the band gap.

Pseudopotentials

The use of pseudopotentials is introduced to approximate the complicated behavior of all electrons. By rewriting the Schrödinger equation to have an effective potential instead of the complicated coulombic interaction between non-valence electrons, one can use a plane wave basis set. This method deals with valence electrons directly and considers non-valence electrons as part of the nuclei as their contribution to chemical binding is limited.

A common type of pseudopotential is that of the projector augmented wave method (PAW)[19] used for *ab initio* calculations. As wavefunctions tend to be highly oscillatory before a cut-off radius $\mathbf{r}_c$, PAW smoothens them as shown in figure 2.1.

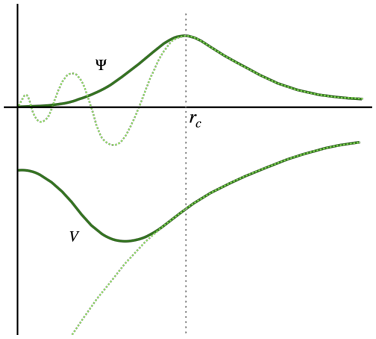


Figure 2.1: Smoothening of dashed wavefunction and potential before cutoff radius

2.1 Molecular Dynamics

Despite design restrictions, molecular dynamics (MD) has been the pinnacle of analyzing atomistic properties. It is a computational method made to simulate the time evolution and interaction of a (many-body) system. Historically, solving such a system is a known impracticality and thus there are several approximations to be taken. Nonetheless, such a solution to a system provides a remarkable amount of information.

Fundamentally, there needs to be a well parametrized potential or force field for the dynamics to reproduce electronic behavior. Albeit accurate, one of the biggest shortcomings of MD is its scalability: as the simulation time increases, the errors accumulating from its trajectory can no longer be ignored. The simplest model for interatomic interactions is the Lennard-Jones pair potential [20]:

$$V_{LJ}(\mathbf{r}) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.11}$$

Although it comes with severe limitations stemming from particle interactions and trajectories, accurate enough parameters can produce adequate results. As such, MD requires a many body potential. For greater accuracy, widely used potentials include embedded atom models [21] which use a function of the electronic density along with a pair potential (2.12).

$$V_i^{eam} = F_\alpha\left(\sum_{j\neq i}\rho_\beta(r_{ij})\right) + \frac{1}{2}\sum_{j\neq i}\phi_{\alpha\beta}(r_{ij}) \tag{2.12}$$

Statistical mechanics for an ideal system

As a system of N particles for $N > 3$ can scale beyond current computational capabilities, there are several numerical methods to constrain and preserve the system. Large enough and complex systems lend themselves to the use of statistical mechanics. For a system in thermal equilibrium, ensembles are used to idealize a system of particles. The simulation box can then be part of a microcanonical (NVE), canonical (NVT), or isothermal-isobaric (NPT) ensemble; where the quantities within the parentheses are conserved. In each of these cases, there are thermostats which control the motion and scaling of the particles. Since the object of study for this thesis is body-centered cubic (bcc) zirconium, it is essential to be able to control the temperature, as will be discussed in Chapter 3.

2.1.1 Ab initio molecular dynamics

While it is fast and easy to compute dynamics with a classical treatment, it is not reliable. Hence, merging classical MD and *ab initio* electronic structure theory created the framework for *ab initio* MD (AIMD) [22].

Despite the fact that AIMD is exceedingly accurate, it is much more computationally expensive. Thus, there are several methods, or approximations, that have followed (BOMD, QMD, FPMD)[23]. Nonetheless, the goal is the same: to compute atomic forces (2.13) as the trajectory is generated. In many (if not all) cases, this is done via the Hellmann-Feynman theorem:

$$F_{R_\gamma} = -\left\langle \psi \middle| \frac{\partial \hat{H}}{\partial R_\gamma} \middle| \psi \right\rangle \tag{2.13}$$

where the force on ion γ at position R_γ is equal to the negative expectation value of the derivative of the Hamiltonian; where the Hamiltonian in our case is that of the Born-Oppenheimer approximation (2.5). Hence, the name Born-Oppenheimer molecular dynamics.

Chapter 3

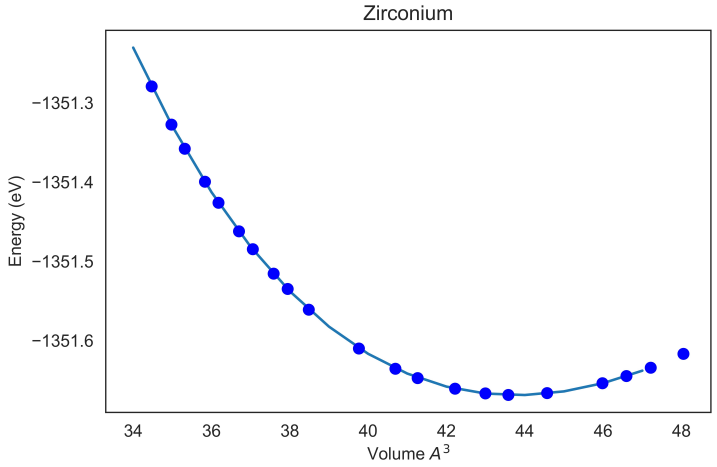
Phonon Thermodynamics of bcc Zirconium

Zirconium (Zr) is a group four transition metal whose characteristics and thermodynamic properties are of great interest. Like its naturally occurring counterparts, titanium (Ti) and hafnium (Hf), it has an $\alpha \rightarrow \beta$ phase transition at high temperature. Moreover, they all have a coinciding high pressure ω phase associated with this transition [24, 25]. Strikingly, while mechanically unstable at low temperatures, they become thermodynamically stable at high temperatures. The stability in the body-centered cubic (bcc) phase is attributed to either the anharmonic contribution from lattice vibrations [26, 27] or alloying in the ω phase [28]. In this thesis, we will analyse the anharmonic effects of high temperature bcc Zr by *ab initio* molecular dynamics simulations.

3.1 Methods

At zero Kelvin, zirconium has a hexagonal close-packed (α-phase) crystal structure which undergoes a martensitic transformation to a body-centered cubic structure (β-phase) at 1135K [15]. Hence, plane-wave self consistent field (scf) calculations were performed to find the equilibrium lattice parameter for a bcc structure. Using an ultrasoft PAW potential, we start with $a = 3.60$ Å [24] and perform scf calculations for several points above and below a. For the several points calculated, a fit to the Murnaghan equation of state 3.1 calculated the lattice parameter to be $a = 3.52$ Å, as shown by figure 3.1.

$$E(V) = E_0 + B_0 V_0 \left[\frac{1}{B_0(B_0' - 1)} \left(\frac{V}{V_0} \right)^{1-B_0'} + \frac{1}{B_0'} \frac{V}{V_0} - \frac{1}{B_0' - 1} \right] \tag{3.1}$$



Murnaghan equation of state fit for zirconium lattice parameter. Blue circles are scf calculations for lattice parameters, a. The energy minimization lattice parameter is $a = 3.52$ Å.

Figure 3.1:

Born-Oppenheimer Molecular Dynamics

Our DFT simulations were performed with Quantum Espresso [29] using a plane wave basis, an ultrasoft PAW potential, a wavefunction kinetic energy cutoff of 40 Ry, and a stochastic-velocity-rescaling (svr) thermostat. The crystal lattice used was a $(4 \times 4 \times 4)$ bcc Zr supercell of 128 atoms with an input temperature of 1500K. Since our goal is to analyze the thermal effects of our structure we simulate an NVT ensemble.

Before starting the MD simulations, we relaxed our supercell from positions slightly displaced from equilibrium using steepest-descent and damped dynamics. From this step forward, we simulated 800 trajectory steps at a high temperature range. To converge our steps faster, the electronic self-consistent mixing factor, `mixing_beta`, was set to 0.01. For every 20 steps, either the temperature or parameter `nraise` was analyzed and slightly changed to shorten the big range in temperature.

Phonon Dispersions

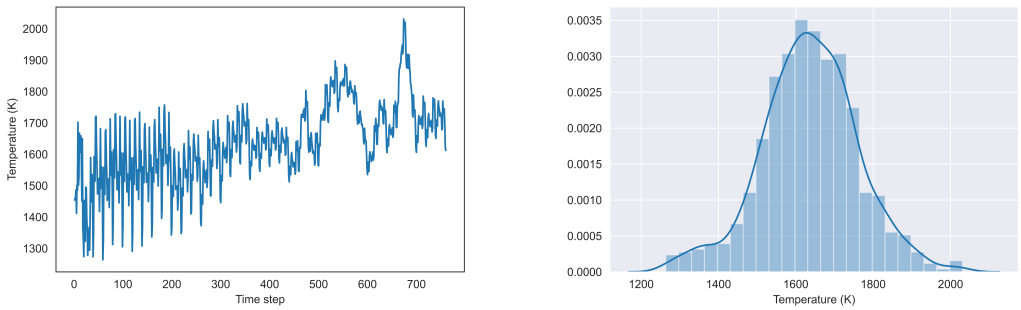
With our dataset, we use a Born-von Kármán (BvK) fit to calculate force constants. In the body-centered cubic lattice, there are 14 unique force constants for fifth nearest neighbors. While these results are our main interest, we also analyze first through fourth nearest neighbors.

Using positions and forces, the BvK fit can output each force constant matrix for unique pair interactions in our crystal. The fit output is formatted in such a way that a phonon calculator such as Phonopy [30] can use it as an input to calculate dispersions. For bcc zirconium, the high symmetry direction used is the following: $\Gamma - H - P - \Gamma - N$

3.2 Results

Thermostat

As zirconium changes phase below ∼1135K and melts above ∼2100K, controlling the thermostat is of utmost importance. As shown in figure 3.2, our temperature (using the stochastic-velocity-rescaling thermostat) ranged from 1250K to 2050K with most points being around 1600K.



(a) Temperature for every time-step

(b) Temperature distribution

Figure 3.2: BOMD Temperature

Phonon Dispersions

As seen in figure 1.6, bcc zirconium has non-degenerate transverse acoustic branches in the [110] direction. Across all time steps, the T_1 branch oscillates between instability and near degeneracy with the T_2 branch. It also has a vast frequency range when considering different nearest neighbors (3.8). Conversely, the T_2 branch has a lower oscillating range, mostly hovering about 3 THz. In the [111] direction, the T_1 branch also oscillates towards instability, however not as often as the N-phonon mode.

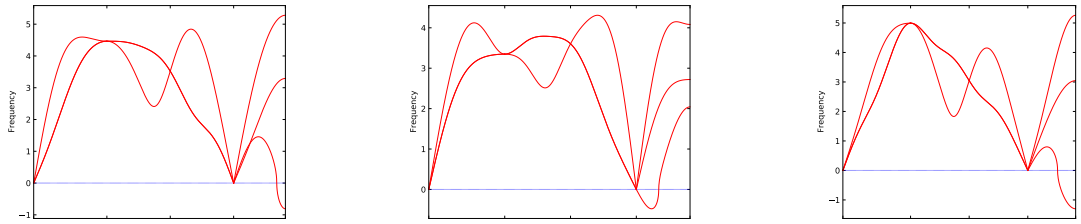


Figure 3.3: Single time step unstable phonon modes

The anharmonic effect on bcc zirconium results in the general instability of phonon modes for any given time step (3.3). Visually, this effect is more oscillatory (3.4 - 3.6). At some time step, the dispersion may have some unstable phonon mode which will eventually be stable for some other number of time steps before it becomes unstable again.

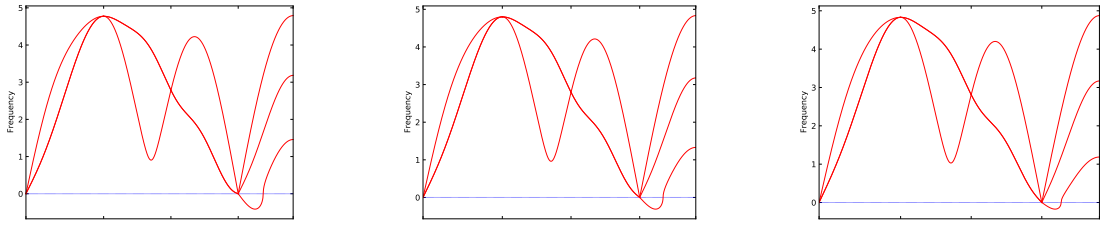


Figure 3.4: Fluctuations: 324-326

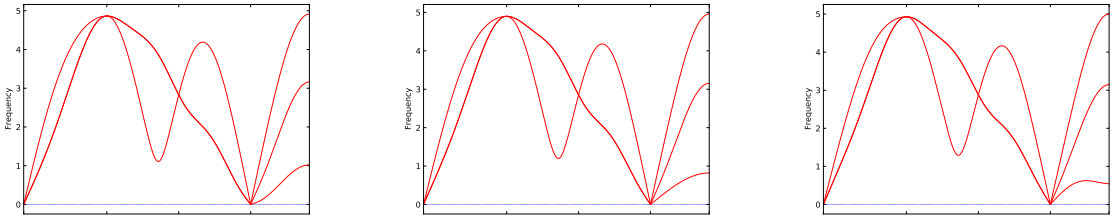


Figure 3.5: Fluctuations: 327-329

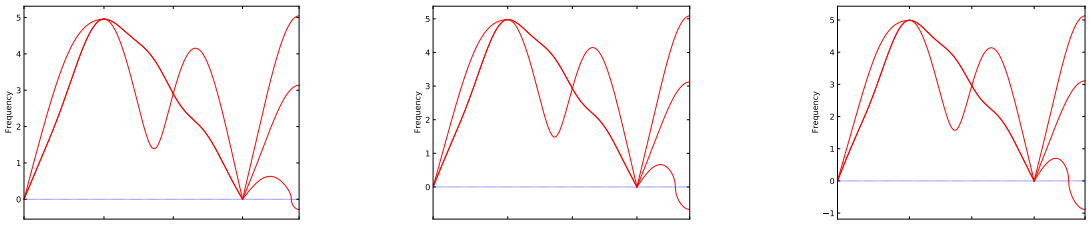


Figure 3.6: Fluctuations: 330-332

Following the fact that, on average, a bcc high temperature system is stabilised by the anharmonicity in lattice vibrations [31], we compute an average phonon dispersion from all time steps (3.8); with the exception of fifth nearest neighbors, this is seen quite well.

The heat maps in figure 3.9 were created by layering all time steps in the simulation. It is much easier to see the phonon broadening along different modes as well as its increase with the inclusion of more nearest neighbors, of which a side effect is the lowered intensity.

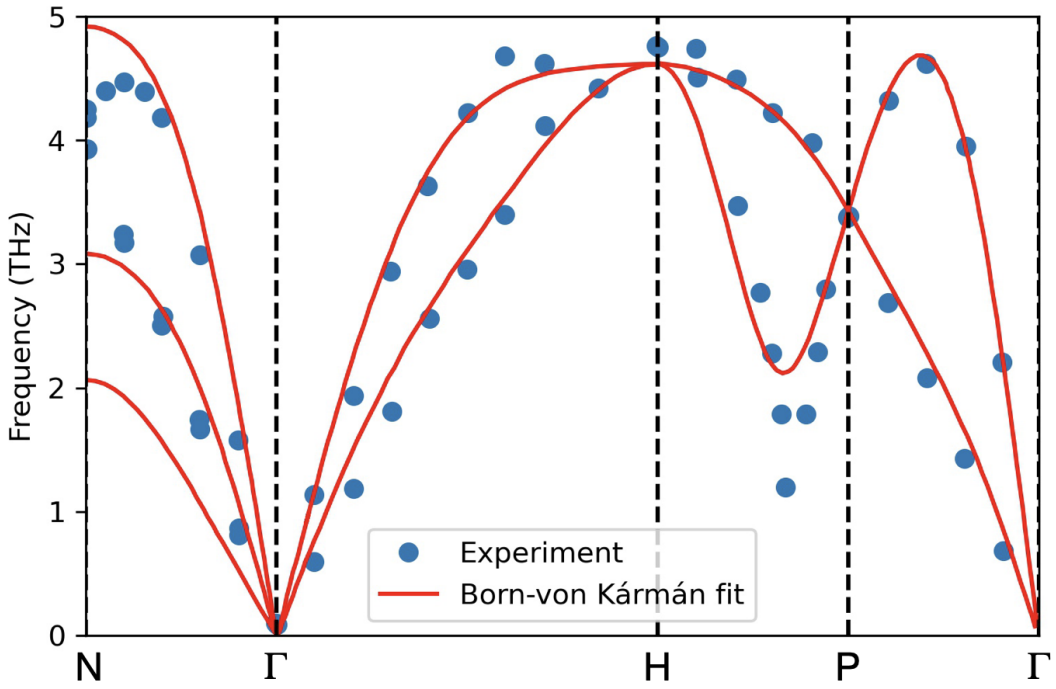
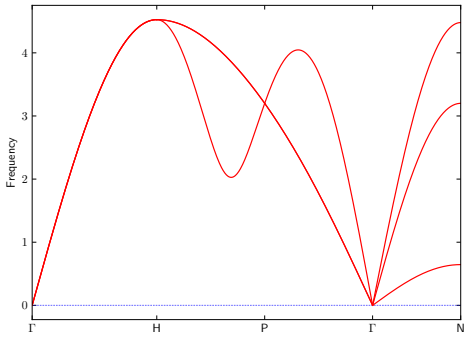


Figure 3.7: Overlay of figure 1.6 and our fit (3.8)

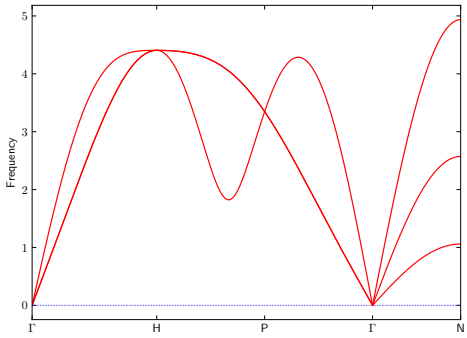
To compare with experimental data, we overlay points from figure 1.6 to a fourth nearest neighbors fit from our data. The visual guide to the experimental data is no longer there but our results match up well enough[1] (with the exception of the transverse branches in the N-Γ direction, where our results clearly show degenerate branches, and the low dip in the H-P direction).

[1]Note here the temperature discrepancies from the experiment (1423K) and our data (1300K-2000K).

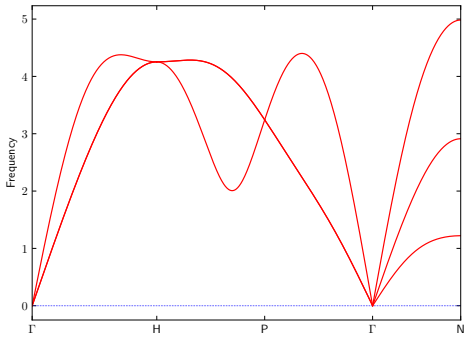
Phonon dispersions: 1^{st} through 5^{th} nearest neighbors



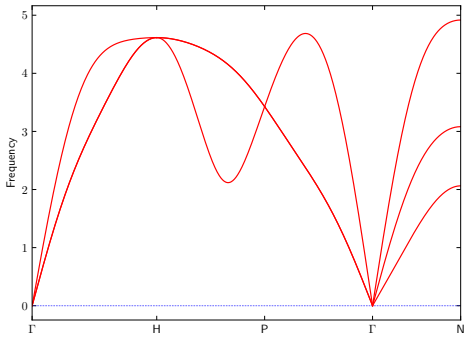
(a) 1^{st} nearest neighbors



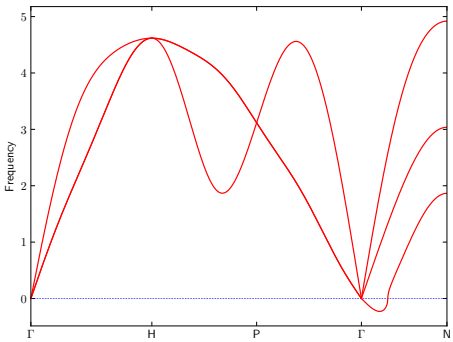
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors



(d) 4^{th} nearest neighbors



(e) 5^{th} nearest neighbors

Figure 3.8: Average phonon dispersions for all points

Heat maps

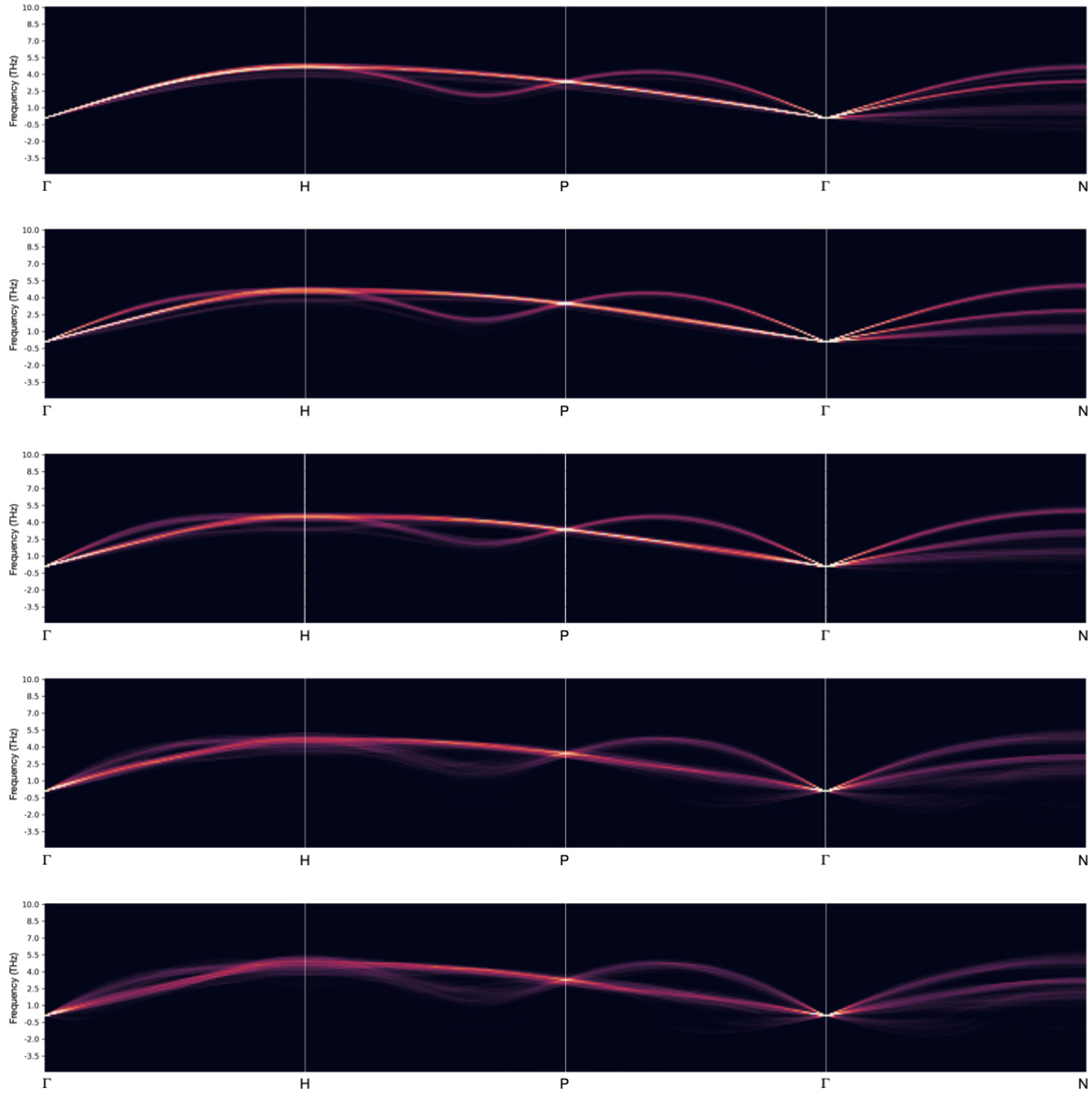


Figure 3.9: Heat maps of all time step phonon dispersions. From top to bottom: 1^{st} nearest neighbors to 5^{th} nearest neighbors

Chapter 4

Machine Learning and Deep Learning

Machine learning and curated neural networks have currently been at the forefront for solving computational problems in materials science, amongst a plethora of other topics. From its statistical beginnings, the field has enjoyed great success, most considerably in the last ten years.

The main goal of machine learning essentially lies in pattern recognition. Given a set of known information, a network will try to predict a possible outcome. Perhaps the most well known example is the classification of the MNIST dataset [32]. Several methods and algorithms have been tried and developed to accurately classify handwritten digital images of single digits, of which the best performing has been a convolutional neural network.

Handwritten digits (0-9) tested by different machine learning methods to study their accuracy

Figure 4.1:

As a whole, machine learning algorithms can be classified in terms of how supervised the training set data is. There are mainly three categories: supervised, unsupervised, and reinforcement learning. Occasionally, a fourth, semi-supervised, can be thrown into the mix. As the representation of data is essential to a successful model in the category of atomic structures, this thesis will focus on supervised learning methods.

Neural Networks

In 1943, W. McCullogh and W. Pitts proposed the idea of mathematically modeling neural activity in what would become the start of artificial neural networks (ANNs) [33]. Based on the biological system, ANNs are made up of nodes (neurons) and edges (connections between neurons) which have an associated *weight*. As a model learns, the weights on both nodes and edges update to lower the error on an output. Typically, in between the input and output layer there is one or more hidden layers in which nodes are organized. Note here that having more layers and more nodes per layer does not mean the model will learn faster. This concept of having multiple layers to extract different features is the basis of deep learning. The most common model is that of convolutional neural networks (CNNs).

CNNs are based on the mathematical operation of the same name. A convolution is defined as an operation on two functions $f,\ g$ that results in a third, $(f * g)$.

$$(f * g)(t) := \int_{-\infty}^{\infty} f(a)g(t-a)da \tag{4.1}$$

As convolution is commutative, equation 4.1 can also be written as:

$$(f * g)(t) := \int_{-\infty}^{\infty} g(a)f(t-a)da \tag{4.2}$$

A beneficial aspect of convolutions is their translational invariance. CNNs are very commonly used for image recognition and any translation operation on an image does not change it. As part of the network architecture, this is reflected in parameter sharing; across different pixels, a hidden layer output (feature map) is used on all of them.

Supervised learning

Supervised learning is defined as a function-learning task that uses its' input data to predict a desirable output. As it requires *labeled* data, the representation of the input data is an important facet to consider. Typically, the input is transformed into a feature vector which should describe as much of an object as possible in its dimensional limit. In materials science, there have been several approaches in defining the important features of a system as well as its machine learning method [34, 2, 3].

In considering an atomic system, as mentioned in Chapter 1, the structure is an important factor in determining physical properties. While using an existing neural network architecture is convenient for lowering errors in molecular predictions, it is perhaps more powerful to generalize the problem. It follows then that the data representation of the input should be functional across all atomic systems.

Active learning

A supervised learning method, active learning lets the user interact with and label new data points as desired. The fundamental idea is that fewer selected data points will allow a model to train faster than usual supervised learning methods [35]. As a typical dataset can span hundreds or thousands of data points, of which labeling is not a feasibility, active learning algorithms query or prompt the user to label new data. As is the case with all machine learning methods, the performance mileage may vary across different problems. There runs the risk of training with data that doesn't necessarily improve performance or let the model learn anything of significance. Thus, there are several methods within active learning to reduce errors [36, 37, 38].

4.1 Euclidean (3) Neural Networks: e3nn

The tensor field networks, e3nn [10], are a library of networks that enjoy the three dimensional symmetries of Euclidean space. A defining characteristic is their 3D translational- and rotational- equivariance [6]. As a consequence of this constraint, the rotation-invariant output is less expensive than traditional data augmentation. An advantage of Euclidean space is the freedom to pick any coordinate system for an atomic system. In rotating said system, the vector quantities on each atom is preserved (as a consequence of equivariance). Subsequently, the network is able to recognize features in different locations and orientations.

The architecture of the network itself is that of a modified convolutional neural network. Where CNNs operate on pixel data for 2D images, e3nn uses point clouds for three dimensions. The convolution filter: $W(\overrightarrow{r}) = R(r)Y_l^m(\hat{r})$, a product of a learnable radial function and a spherical harmonic, is chosen as a spherical basis which steers the structure of the network to be compatible with algebraic formalism, namely tensors. Hence, every layer of the network inputs and outputs geometric tensors [6]:

$$\mathcal{L}(\mathbf{r}_i, x_i) = (\mathbf{r}_i, y_i) \tag{4.3}$$

where layer $\mathcal{L}$ operates on point coordinates $x_i \in \mathbb{R}^{\mathcal{X}},\ y_i \in \mathbb{R}^{\mathcal{Y}}$. For group representations $D^{\mathcal{X}}(g),\ D^{\mathcal{Y}}(g)$ where g is an element of some group G, the condition for equivariance (1.4) is expressed as follows [6]:

$$\mathcal{L} \circ D^{\mathcal{X}}(g) = D^{\mathcal{Y}}(g) \circ \mathcal{L} \tag{4.4}$$

As the composition of two equivariant functions $\mathcal{L}_1,\ \mathcal{L}_2$ results in another equivariant function, $\mathcal{L}_1 \circ \mathcal{L}_2$, it's easy to follow that the layers across the entire network (4.3) are equivariant themselves.

The tensor fields of each layer in the network act on a set of points and a vector in a representation of $SO(3)$. They are implemented as a string of the following parameters: multiplicity, rotation-order, and parity. For example, the string `1x0e` is one instance of an

even scalar. The first half of the string `1x` corresponds to the number of copies needed. The second half, `0e`, can be expressed as `l p` where l, the rotation-order, is an integer depending on the tensor form (0 = scalar, 1 = vector, 2 = symmetric traceless matrix, .., etc.) and p is the parity. Since they are irreducible representations of $SO(3)$, they have dimensions $2l+1$. The irreducible representations of $SO(3)$ are Wigner-D matrices.

4.1.1 Spherical Harmonics

The spherical harmonics Y_l^m are a set of special functions that define points on a sphere. They are defined as follows:

$$Y_{lm} = \begin{cases} \frac{i}{\sqrt{2}}(Y_l^m) - (-1)^m Y_l^{-m} & m < 0 \\ Y_l^0 & m = 0 \\ \frac{1}{\sqrt{2}}(Y_l^{-m}) + (-1)^m Y_l^m & m > 0 \end{cases} \tag{4.5}$$

or visually:

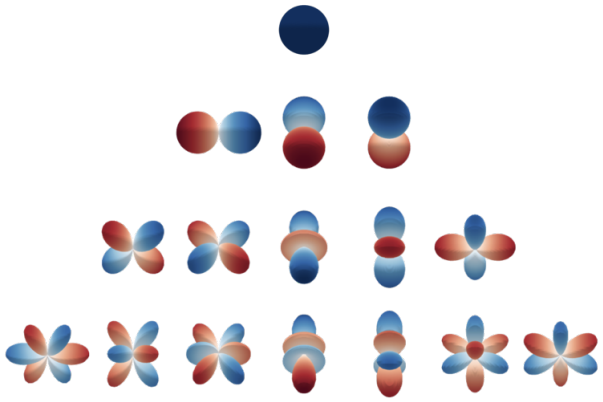


Figure 4.2: Spherical harmonics for l=0,1,2,3 from top to bottom

Equivariant to SO(3), they are considered basis functions for irreducible representations. Given a geometric input, E3NNs will transform this into features by spherical harmonic projections.

Under a rotation, $\mathcal{R}$, a spherical harmonic can be expressed as:

$$Y_m^{(l)}(\mathbf{r}') = \sum_{m'=-l}^{l} D_{mm'}^{(l)}(\mathcal{R}) Y_{m'}^{(l)}(\mathbf{r}) \tag{4.6}$$

However, because 3D rotations are not abelian like rotations in 2D, a fully rotation-equivariant filter is restricted to the following [6]:

$$F_{cm}^{(l_f,l_i)}(\mathbf{r}) = R_c^{(l_f,l_i)}(r) Y_m^{(l_f)}(\hat{\mathbf{r}}) \tag{4.7}$$

where r is the magnitude of vector $\mathbf{r}$, and $\hat{\mathbf{r}}$ is its normalized unit length vector. $R_c^{(l_f,l_i)} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ are weighted learned radial functions corresponding to input and filter rotation orders: $l_i,\ l_f$.

Since the layer inputs and filters are expressed in irreps of $SO(3)$, they must combine in a way that the output is equivariant. In e3nn, this is done with a tensor product, an equivariant operation, where the combination of two representations, $D^{\mathcal{X}},\ D^{\mathcal{Y}}$ can be expressed as: $D^{\mathcal{X}} \otimes D^{\mathcal{Y}} = D^{\mathcal{X} \otimes \mathcal{Y}}$. As previously mentioned, the irreps of $SO(3)$ are Wigner-D matrices whose product can be calculated using Clebsch-Gordan coefficients. Thus, a pointwise convolution layer is defined as [6]:

$$\mathcal{L}_{acm_\circ}^{l_\circ}(\mathbf{r}_a, V_{acm_i}^{(l_i)}) := \sum_{m_f,m_i} C_{(l_f,m_f)(l_i,m_i)}^{(l_\circ,m_\circ)} \sum_{b \in S} F_{cm_f}^{(l_f,l_i)}(\mathbf{r}_{ab}) V_{bcm_i}^{(l_i)} \tag{4.8}$$

Chapter 5

Phonon Thermodynamics of Neural Network Trained bcc Zirconium

As experienced by our calculations in Chapter 3, computational costs from density functional theory molecular dynamics simulations are extremely costly. Here, we explore the use of machine learning methods to cut down on those costs. Using the e3nn library with an active learning algorithm, we reduce our output error to $\sim$0.008 eV/Å.

5.1 Methods

e3nn

It is more common and practical to exploit the invariance of scalar quantities in a network not inherently built to recognize higher-order (geometric) tensors. By using an equivariant model, the features that describe the system are only limited by invariant features and symmetries [39]. Physical systems themselves inherently use equivariance to preserve their quantities, as they transform deterministically.

To train our model, we use the DFT data generated by Quantum Espresso [29]. Since the capabilities of e3nn are limited only by data structure, it was decided that energy would be used as an input and the forces would be outputted directly, as opposed to energies being outputted and forces calculated thereafter.

We use a modified gate point network to train our model since the distance between positions in our lattice are necessary to calculate the edge shift. Using e3nn's formalism for identifying the structure of our inputs and outputs, we set `irreps_in` to `1x0e` for energy

(scalar) and `irreps_out` to `1x1o` for force (vector). The rest of our parameters were chosen on the basis that drastically changing them does not create a faster or better output. Thus, `num_neighbors` is set to `30.0` and `max_radius` is set to 5.0 Å, where the latter is restricted by the lattice parameter, but any value well above that can be chosen.

```
net = SimpleModNetwork(
    irreps_in='1x0e',
    irreps_out='1x1o',
    max_radius=radial_cutoff,
    num_neighbors=num_neighbors,
    num_nodes=1.0,
    pool_nodes=False,
    )
```

Figure 5.1: Sample net definition

In our training cell, while the input is an energy, the loss of our force is minimized with respect to the original force data:

```
loss_forces = loss_fn(x['forces'], forces)
```

Figure 5.2: Loss minimization

where the loss function is defined as the mean squared error.

Active Learning

Using an expected error reduction [35, 36] query for our active learning method, we follow the form of Y.-H. Tang's protocol [40]:

- **repeat**
    1. $G \leftarrow Train(S)$
    2. $\alpha \leftarrow argmax_{i \in T} \mathcal{Q}(G, i)$
    3. $S \leftarrow S \cup \alpha$
    4. $T \leftarrow T \setminus \alpha$
- **until** criteria is met

For a training set S and test set T, an acquisition function $\mathcal{Q}(G, i)$ will determine data points to remove from T and shuffle back in to S. Our choice of acquisition function is the predicted error of our model above a certain threshold.

Using DFT data to train our model with, we first start with 100 randomized time steps in the train set, leaving the rest as part of the test set. Subsequent runs converge the loss function to a set value or shuffle the top 5% of test set errors back into the train set.

Force Predictions

Regarding our DFT data as the ground truth, the calculation of force prediction errors are as follows for a set of ground truth data points G and a set of prediction points P:

1. $A_{(128\times3)} = G - P$
2. $\mathbf{a}_{(128\times1)} = ||A||_F = \sqrt{\sum_{i,j}^{m,n} |A_{ij}|^2}$
3. $a_{(1\times1)} = \frac{\sum_i^n \mathbf{a}_i}{n}$

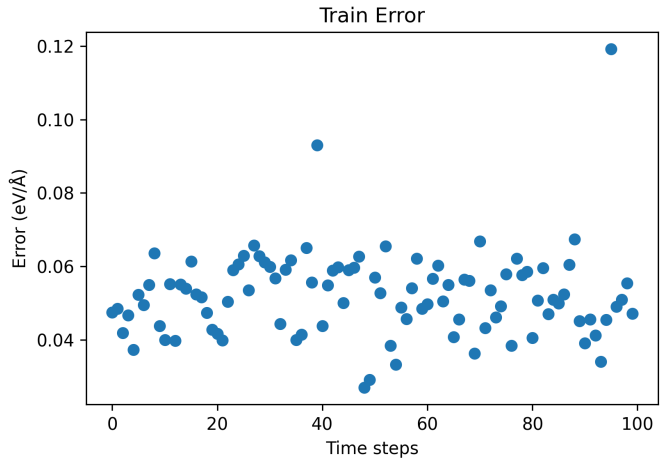


Figure 5.3: Force error plot

In figure 5.3, every dot corresponds to step 3 above as the average error of the entire crystal structure.

Point Sampling

Since the temperature distribution of our dataset has a wide range (3.2), we sample points in several ranges and analyze the phonon dispersions for their accuracy. Recalling that zirconium's bcc phase stabilises with the help of lattice vibrations, the number of points used in some of these ranges is not enough to provide stable phonon modes. Hence, several points at each temperature range is necessary; for this purpose, 1000 points were generated.

Table 5.1: Points sampled for network validation

Temperature	Points Generated	Original # of Points
1200K-1300K	1000	6
1300K-1400K	1000	26
1400K-1500K	1000	58
1500K-1600K	1000	195
1600K-1700K	1000	259

To generate 1000 points for each temperature range in table 5.1, the original points in their respective range were first averaged. Given one average range, a normal distribution dataset of width $1/\sigma$ (where σ is the standard deviation) was created with respect to the original points.

5.2 Results

Active Learning

The first couple of runs in our active learning method show an oscillatory motion in the test set (5.4 - 5.11) and a scatter-like pattern in the train set; with later runs seemingly gaining a structure of sorts. This effect is more than likely due to the fact that our time-steps are very close to each other as the DFT trajectory was generated and there is very little difference between them. Later runs, 5.16 - 5.19, are mostly void of the structure-like gains seen in earlier runs. They are also mostly true scatter-like plots with some slight oscillatory motion as seen in figures 5.17 and 5.18.

Interesting to note is the loss function in later runs as well as runs that optimized the loss to a set value. The loss function in run 1 (5.4) is what is usually observed when training a model. The subsequent run (2, 5.5) was a loss optimization run converging the loss value to 0.0025. Consequently, throughout the run, the model had a difficult time training as its loss heavily fluctuated. Run 4 (5.7) was also an optimization run where the models loss was even more volatile than run 2. In later runs, since the loss nears the initial conversion value it was given, it is then eventually lowered to 0.00005. Since the model hovers around this value relatively quickly, the last couple of runs are trained on 500 epochs, with no consideration to the loss value.

In its initial goal to lower the train and test set error, even before 30 runs, it is clearly visible how successful it was, as seen in table 5.2 and figure 5.20. With 621 training points in the 30^{th} run, by error standards, the model is sufficiently trained.

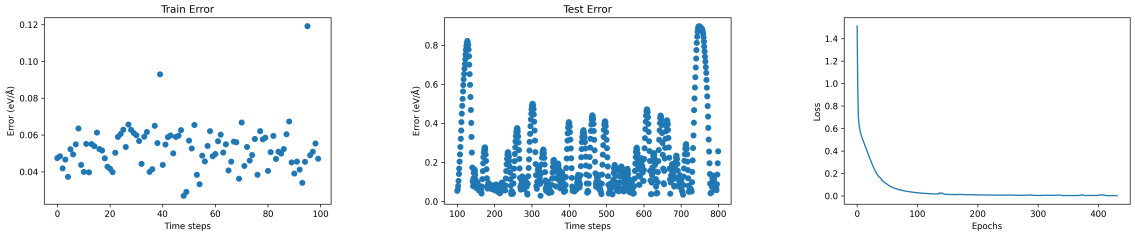


Figure 5.4: AL run 1

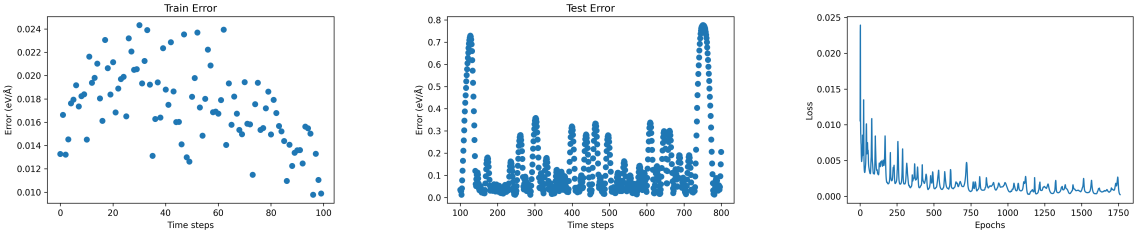


Figure 5.5: AL run 2

Figure 5.6: AL run 3

Figure 5.7: AL run 4

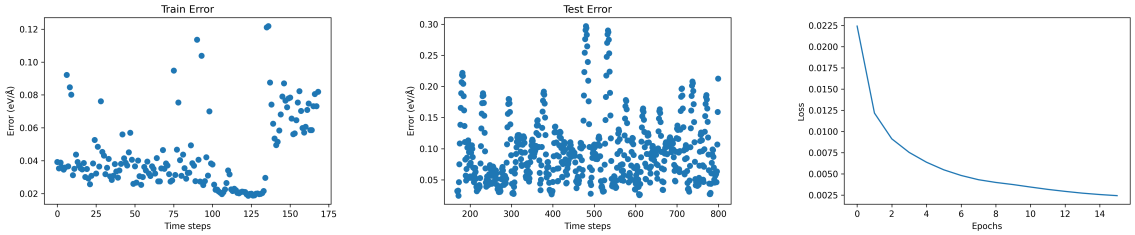


Figure 5.8: AL run 5

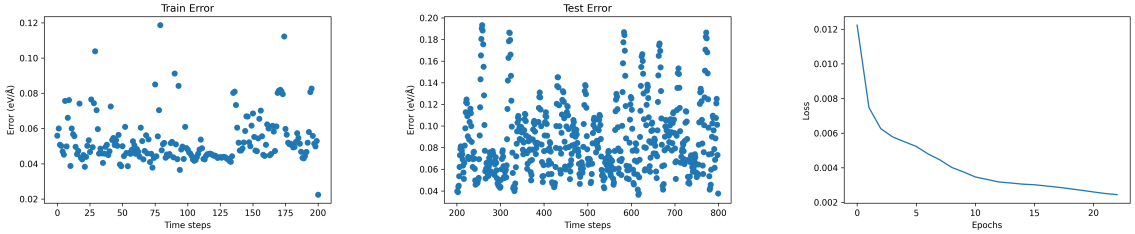


Figure 5.9: AL run 6

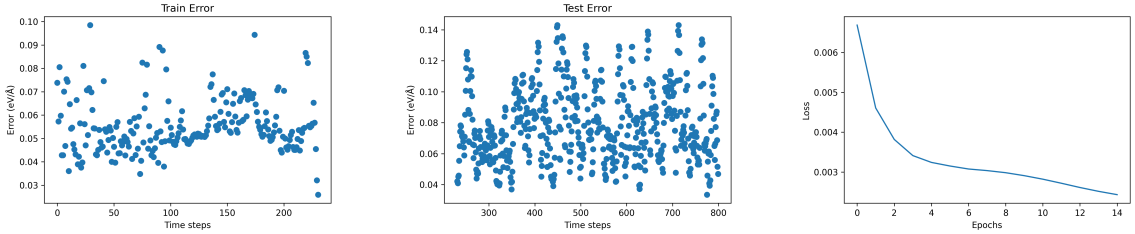


Figure 5.10: AL run 7

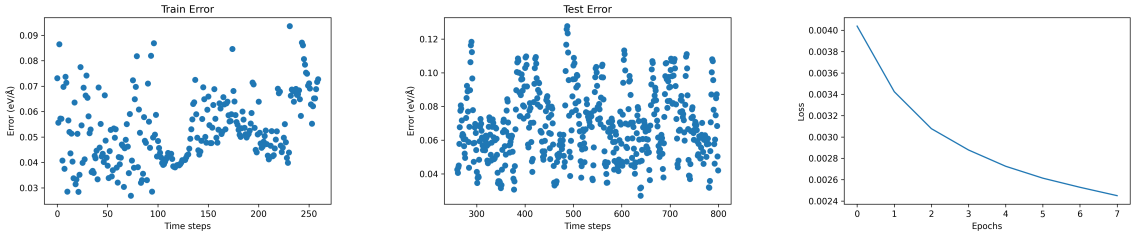


Figure 5.11: AL run 8

Figure 5.12: AL run 15

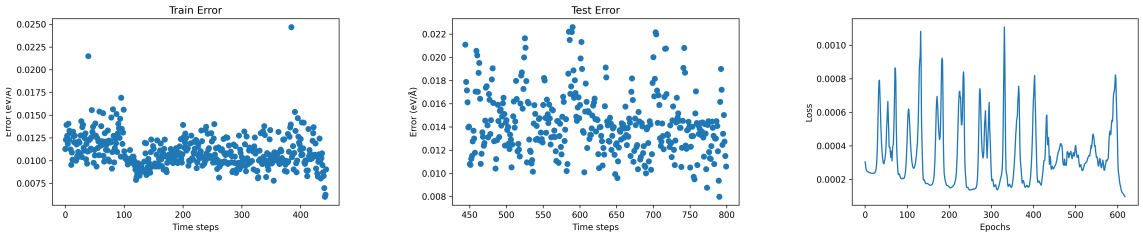


Figure 5.13: AL run 17

Figure 5.14: AL run 19

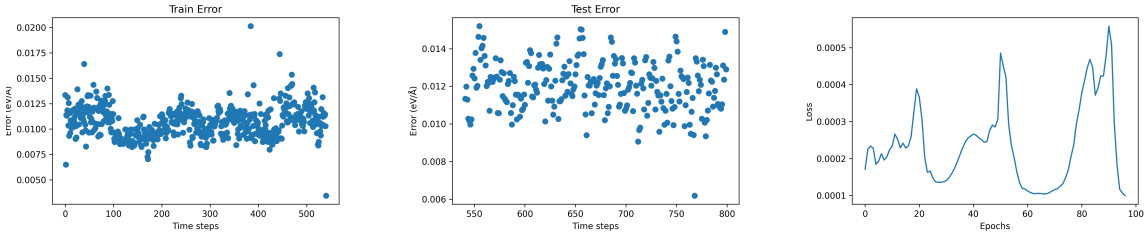


Figure 5.15: AL run 23

Figure 5.16: AL run 27

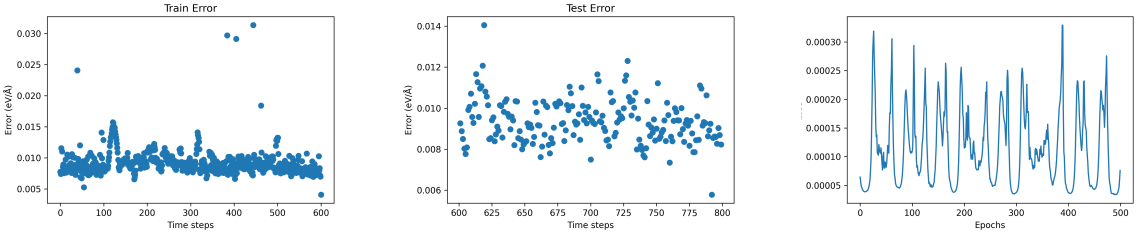


Figure 5.17: AL run 28

Figure 5.18: AL run 29

Figure 5.19: AL run 30

Table 5.2: Median force errors from train and test sets

Median errors (eV/Å)			
Runs	Train	Test	% difference
1	0.0519733	0.1471809	187.87
2	0.0172319	0.0882041	412.05
3	0.0397450	0.0968479	143.67
4	0.0313330	0.0847395	170.45
5	0.0362359	0.0869364	139.92
6	0.0494234	0.0789891	59.82
7	0.0531087	0.0739138	39.17
8	0.0509865	0.0660957	29.66
9	0.0463458	0.0586207	26.49
10	0.0373292	0.0430832	15.41
11	0.0283718	0.0331065	16.69
12	0.0223621	0.0293039	30.91
⋮	⋮	⋮	⋮
24	0.0093847	0.0105580	12.51
25	0.0093876	0.0105878	12.78
26	0.0105597	0.0111711	5.79
27	0.0083136	0.0091024	9.46
28	0.0088553	0.0092719	4.70
29	0.0080277	0.0087100	8.50
30	0.0055986	0.0063097	12.70

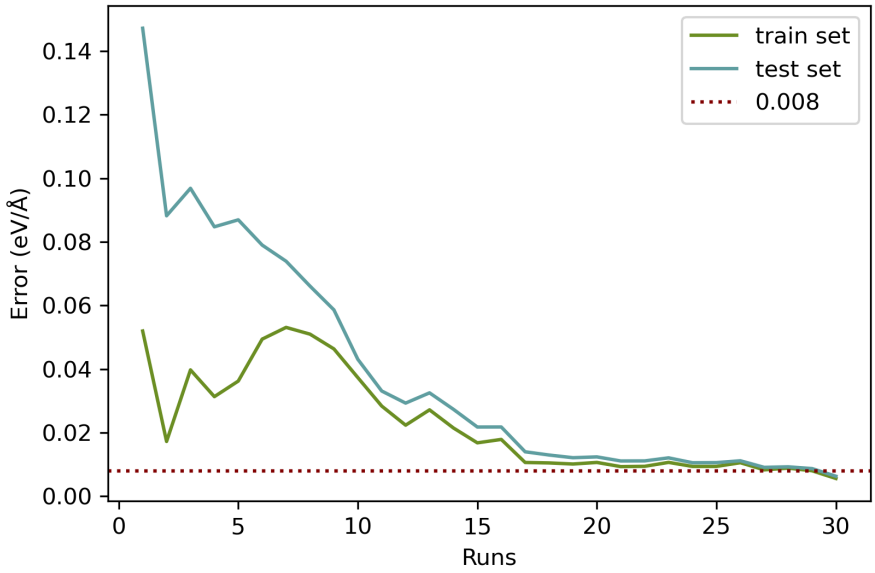


Figure 5.20: Median errors across all runs

Table 5.2 highlights the error difference between train and test sets across several runs. The median value for the prediction errors was picked because it provides a better depiction of errors in all time steps. Occasional outliers and deviations from the median are points that are re-shuffled into the train set.

Starting from the first run, the train error is already relatively low, but the test set error indicates that it's grossly overfit and not at all able to generalize the remaining data; clearly seen by figure 5.4. In runs 1-12, the table shows the test set error trending lower, fluctuating slightly. Somewhat contrary to this, the train set error gives a slight increasing trend as the result of new data points being added. In the following runs, both set errors decrease with minimal fluctuations. As shown by figure 5.20, both train and test set errors seem to converge towards $\sim$0.008 eV/Å.

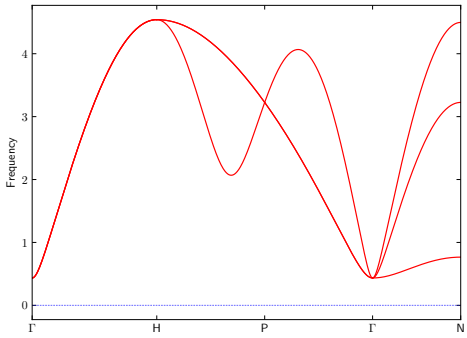
Phonon Dispersions: Network Accuracy

Of the trained models in all our runs, three were picked to analyze their accuracy. Models twenty-nine and thirty were picked because it was assumed that their training with a greater number of data points would allow the network to learn more about our system and therefore be more accurate. Their stark difference lies in train and test set errors as shown in table 5.2, or less directly in figures 5.18 and 5.19. Taking into account the risk of overfitting, it would be expected for run twenty-nine to perform better than run thirty. Run twelve was picked in an effort to see the difference in accuracy for a model with less training points and some slight overfit like run thirty.

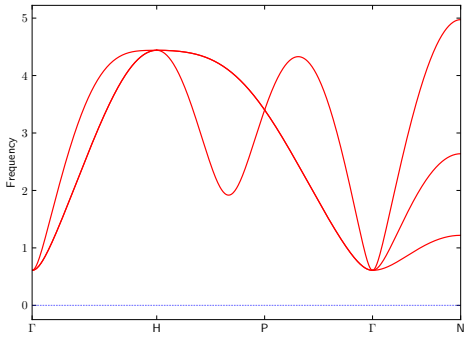
As seen in figures, 5.21, 5.22, and 5.23, there are vast differences in the average phonon dispersions spanning all nearest neighbors. For run twelve and thirty, dispersions for third through fifth nearest neighbors show an instability in the Γ-phonon mode. Remarkably, although run twelve has fewer points and is more overfit than run thirty, its dispersions are near-interchangeable, save for some extremely slight differences in the $\Gamma - \mathrm{N}$ direction. Run twenty-nine, with an overfit error just under ten percent, produced the most reliable dispersions. In fact, in comparing to the original DFT data (3.8), there is virtually no difference between them. Thus, we can conclusively say run twenty-nine, trained on just 611 data points, is able to accurately predict the forces of a system given an energy.

As with the original data, heat maps generated by overlayering all time steps were produced (A). Interestingly, at a glance, the heat maps don't seem to provide a clear distinct difference between models, although there are a few minute details. Overall, they are also very similar to the original data (3.9).

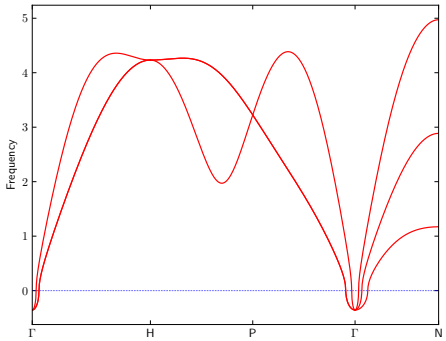
Network phonon dispersions



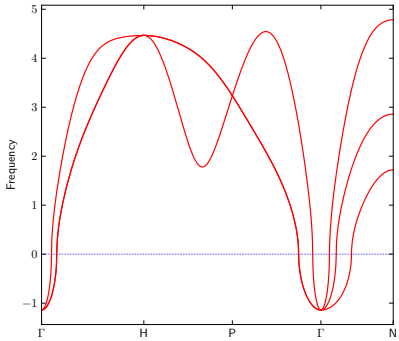
(a) 1^{st} nearest neighbors



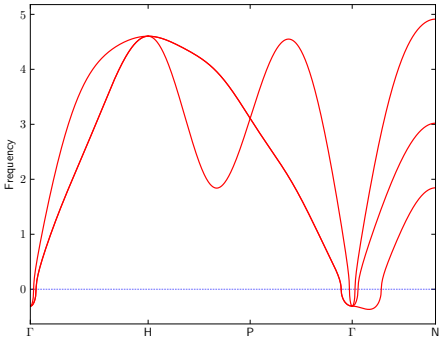
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors

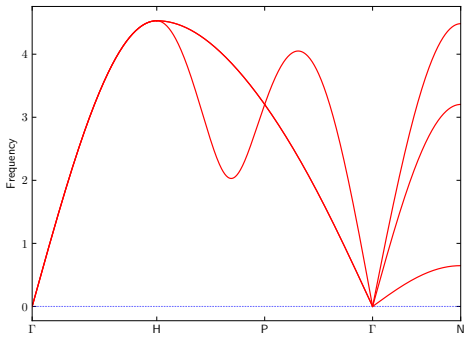


(d) 4^{th} nearest neighbors

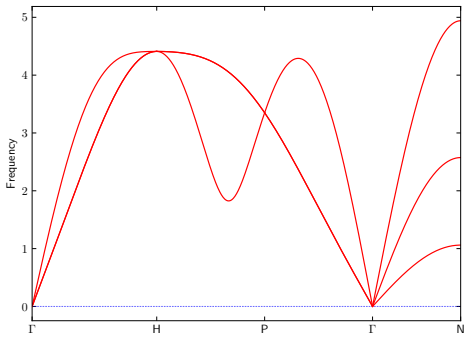


(e) 5^{th} nearest neighbors

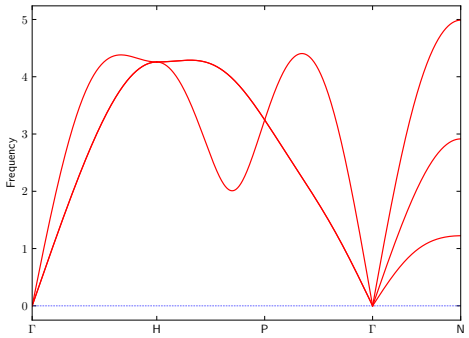
Figure 5.21: Run 12 average phonon dispersion



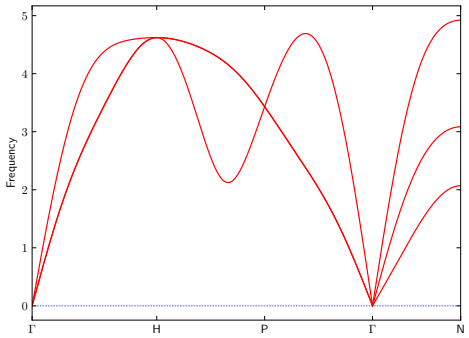
(a) 1^{st} nearest neighbors



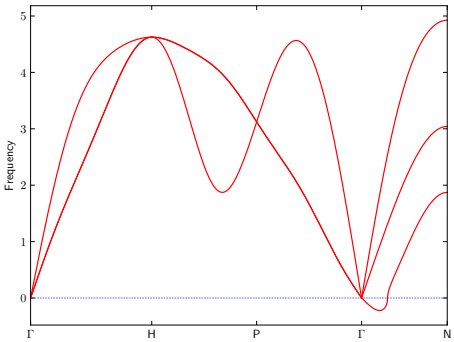
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors

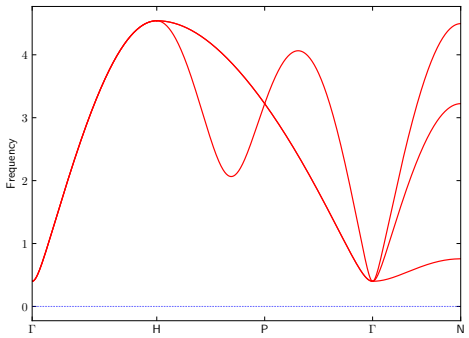


(d) 4^{th} nearest neighbors

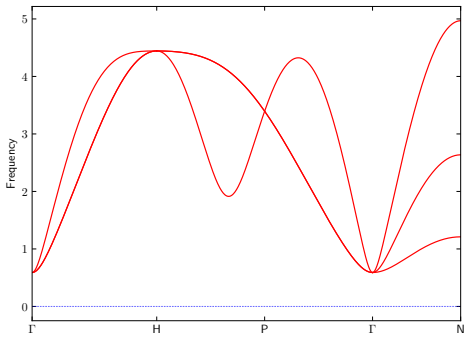


(e) 5^{th} nearest neighbors

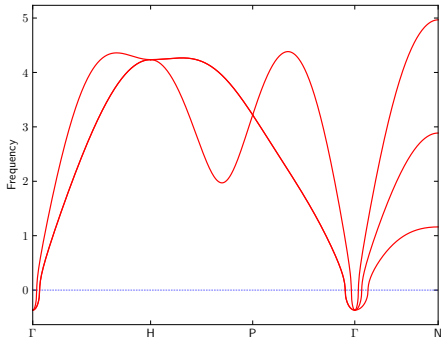
Figure 5.22: Run 29 average phonon dispersion



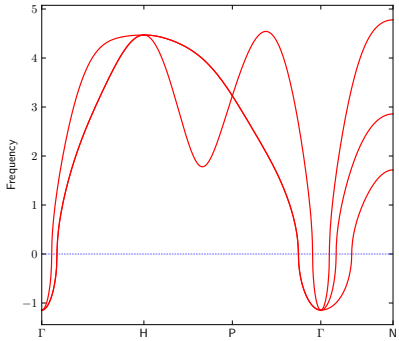
(a) 1^{st} nearest neighbors



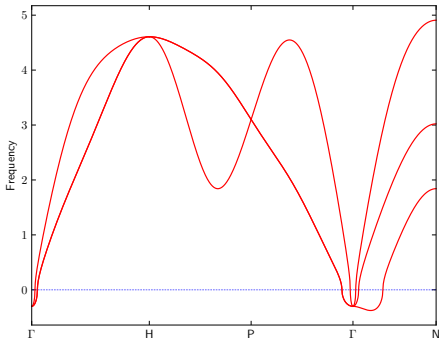
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors



(d) 4^{th} nearest neighbors



(e) 5^{th} nearest neighbors

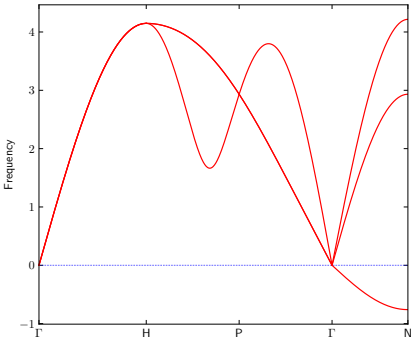
Figure 5.23: Run 30 average phonon dispersion

Phonon Dispersions: Temperature sample

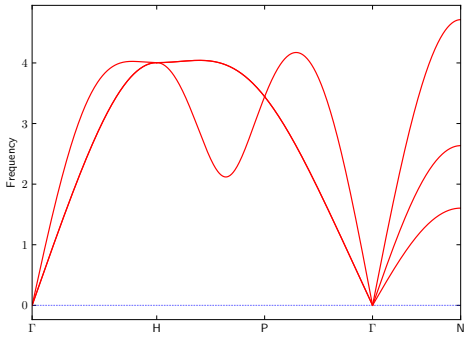
As seen in figure 3.2, because our DFT models' temperature range spans several hundred Kelvins, we decided to take sample increments of 100K (5.1) and generate fictional data. Using the best trained model, run twenty-nine, we feed the 1000 generated data points for force predictions. Similarly to the network models and DFT data, phonon dispersion averages are computed. With the exception of the first nearest neighbors, figures 5.24 to 5.27 show stable dispersions for all selected ranges.

The most significant difference from these selected ranges and the original data is the $\Gamma - \mathrm{H} - \mathrm{P}$ direction. In the fifth nearest neighbors plots, the longitudinal branch starts off with a bit of an elbow crease at 1200K-1300K that generally smoothens out by 1500K. This trend is generalizable as second through fifth nearest neighbors depict a similar pattern as temperature increases. In comparison to the DFT averages, these temperature range dispersions have similar features. For first nearest neighbors, the transverse and longitudinal branches in the $\Gamma - \mathrm{H}$ direction are degenerate and its peak softens as temperature increases. For second and third nearest neighbors, these branches are only degenerate at the H-phonon mode. In fourth and fifth nearest neighbors where a crease is present, the branches aren't completely degenerate and they shift to a higher frequency. The $\Gamma - \mathrm{N}$ direction in these samples behave similarly to the original data in terms of its frequency shift. Collectively, all these ranges produce reasonable dispersions associated with their temperature [24].

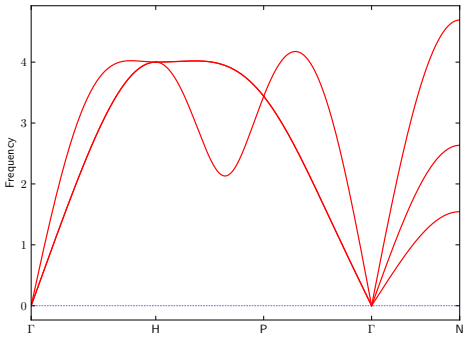
Point sampling phonon dispersions



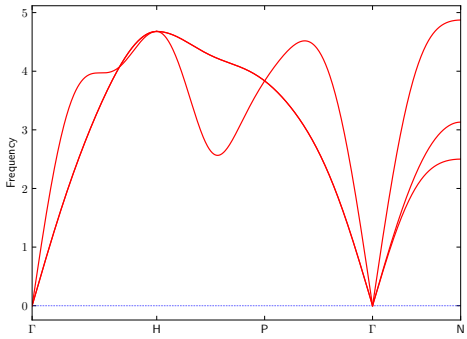
(a) 1^{st} nearest neighbors



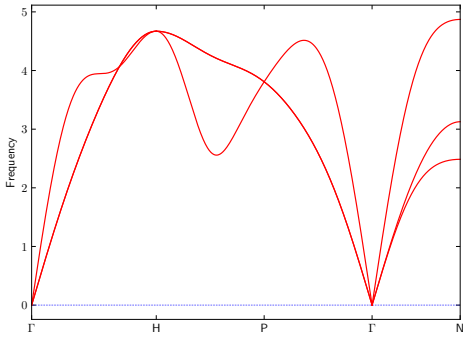
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors

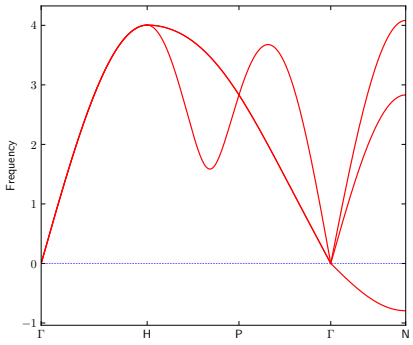


(d) 4^{th} nearest neighbors

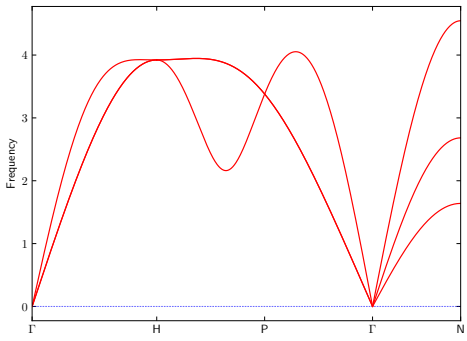


(e) 5^{th} nearest neighbors

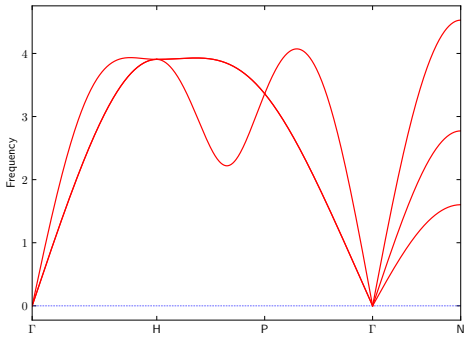
Figure 5.24: 1200K-1300K average phonon dispersion



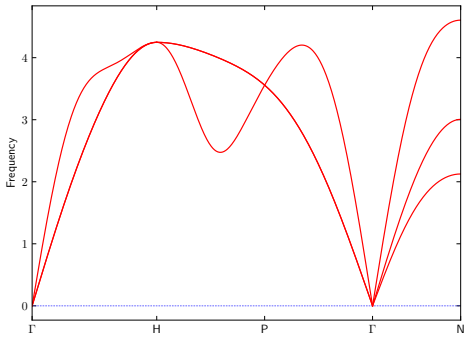
(a) 1^{st} nearest neighbors



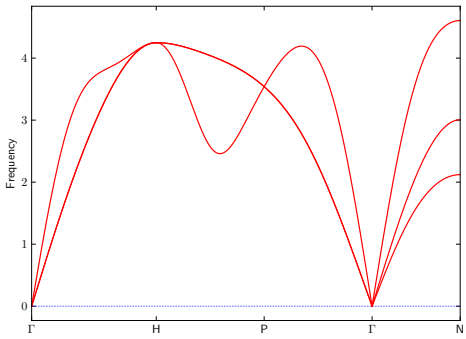
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors

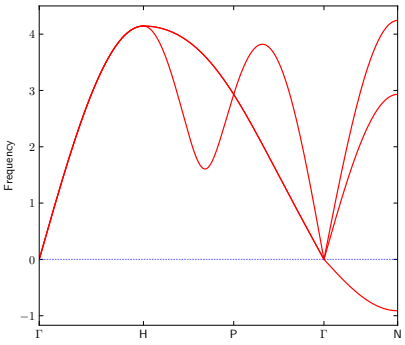


(d) 4^{th} nearest neighbors

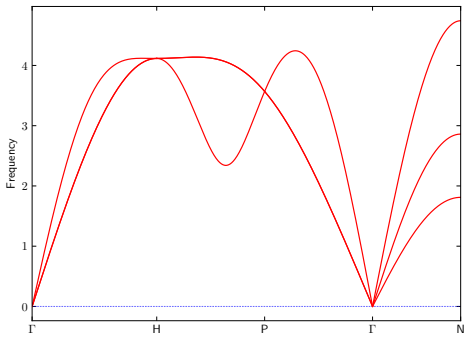


(e) 5^{th} nearest neighbors

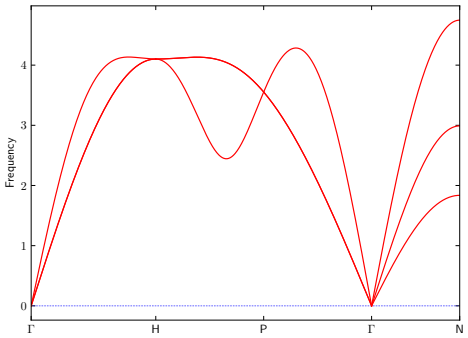
Figure 5.25: 1300K-1400K average phonon dispersion



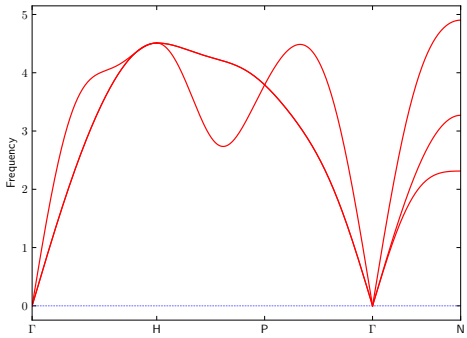
(a) 1^{st} nearest neighbors



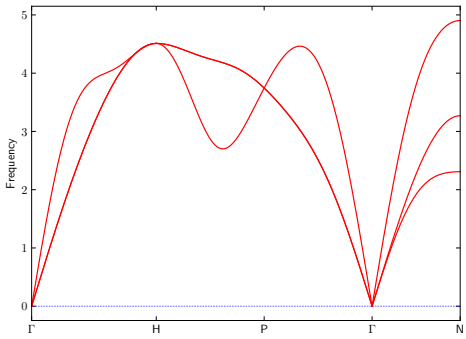
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors

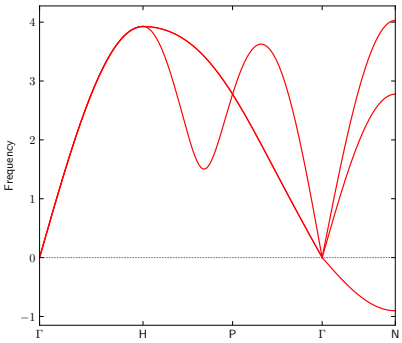


(d) 4^{th} nearest neighbors

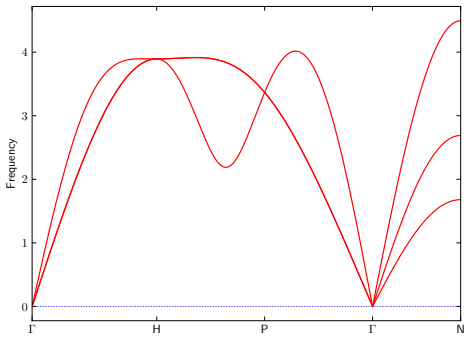


(e) 5^{th} nearest neighbors

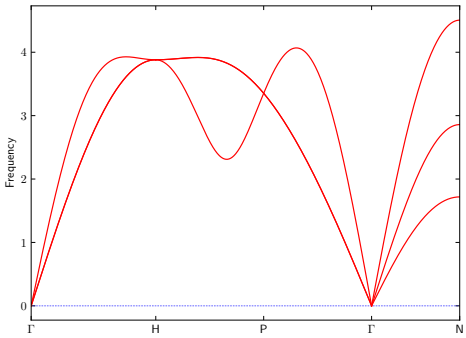
Figure 5.26: 1400K-1500K average phonon dispersion



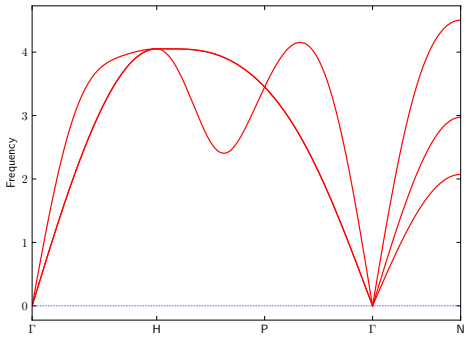
(a) 1^{st} nearest neighbors



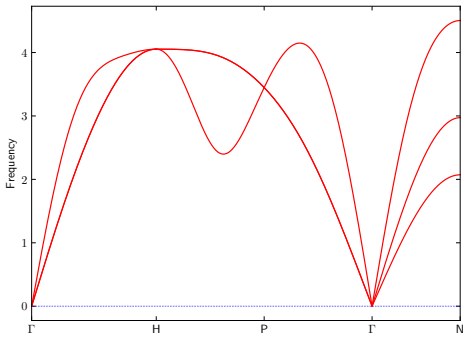
(b) 2^{nd} nearest neighbors



(c) 3^{rd} nearest neighbors



(d) 4^{th} nearest neighbors



(e) 5^{th} nearest neighbors

Figure 5.27: 1500K-1600K average phonon dispersion

Chapter 6

Conclusions

Atomic lattice vibrations play a non-negligible role for temperature-dependent phenomena in thermodynamic processes. We primarily analyzed its effects in phase stability for the high temperature system of body-centered cubic zirconium. The main work presented in this thesis demonstrates the viability of accurately predicting interatomic forces to derive the same effects from DFT simulations. The paring of an active learning algorithm with a Euclidean (3) neural network effectively reduced the needs to use a greater dataset than was initially provided. With just 611 data points our model was able to replicate the electronic band structure for bcc zirconium at a high degree of accuracy.

6.1 Future Work

The main direction for this work is in the molecular dynamics simulation of the neural network trained model. In order to completely remove the costly DFT dynamics, a pathway for neural network dynamics needs to be established. The successful simulation of a trained model would lead to the generalization for other materials as well as its widespread availability.

As part of the goal to generalize machine learning potentials, the zirconium dataset should be expanded to include more configurations. A larger sample space in a trained model would bypass the need to simulate different structures.

6.2 Notes

To the future contributor: the setting of the thermostat in dynamics simulations is paramount. It is by far a simple task to set the system to be in thermal equilibrium. Make little to no changes in an input file. Prior to this main task where the lattice parameter is to be derived, choose a suitable pseudopotential. The work in this thesis utilized a PBESOL ultrasoft (USPP) potential for its entirety, and there's much to be said about its accuracy (but beware the raised flags from software). In any case, a PAW PBE potential is a safe bet when concerned with reliability.

As for the machine learning aspect, not every model is created equal. Some may be more difficult to train than others (though not dependent on said models complexity). The criteria of the active learning algorithm may look arbitrary. Although a model can train well and reduce its loss over a couple thousand epochs, it is much faster or simpler to iteratively add small amounts of data. A bigger number of data points shuffled back into the train set can cause latency in error lowering. It's advised to base the criteria on the lowest amount of points that can make a difference in training (our case made that 5% of top errors). The input-output (scalar-vector) used is widely unconventional but useful. It might also interest the reader to use a scalar input and output to compare behavioral differences. In the vein of generating force-fields, it might also be prudent to try different configurations.

Appendix A

Heat map phonon dispersions

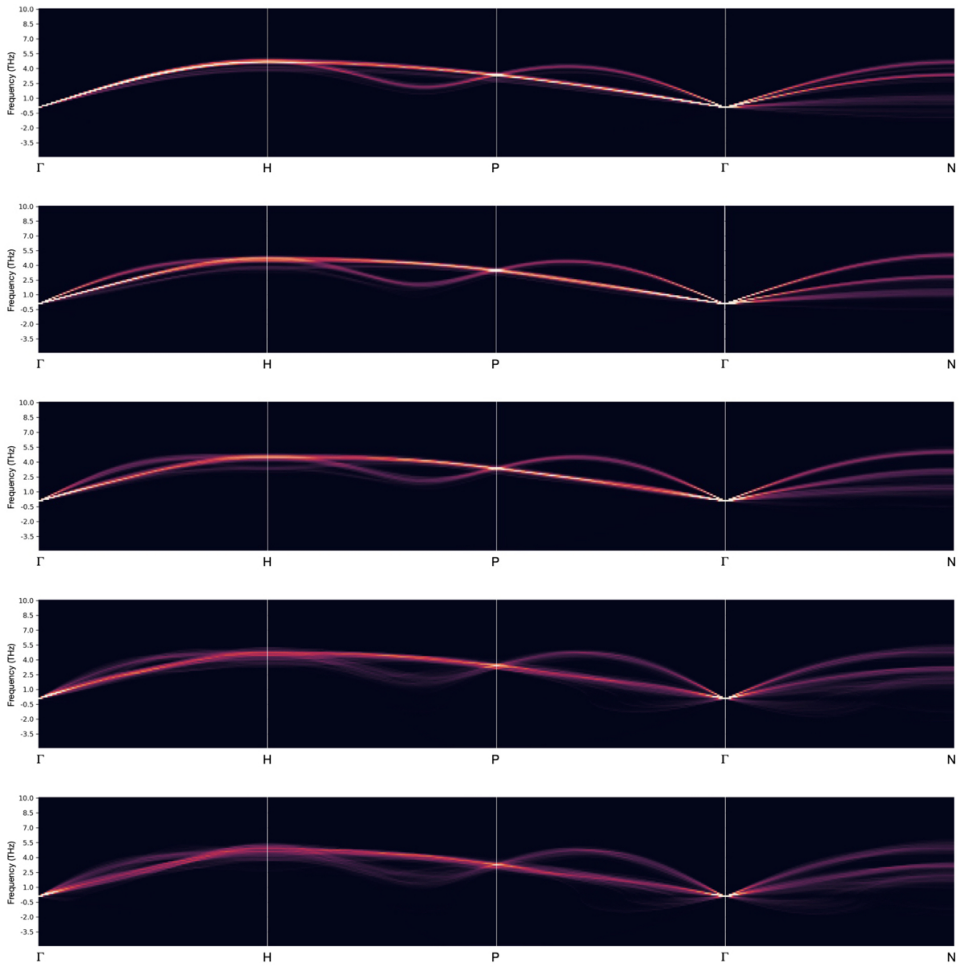


Figure A.1: Heat maps of active learning run 12. From top to bottom: 1^{st} nearest neighbors to 5^{th} nearest neighbors

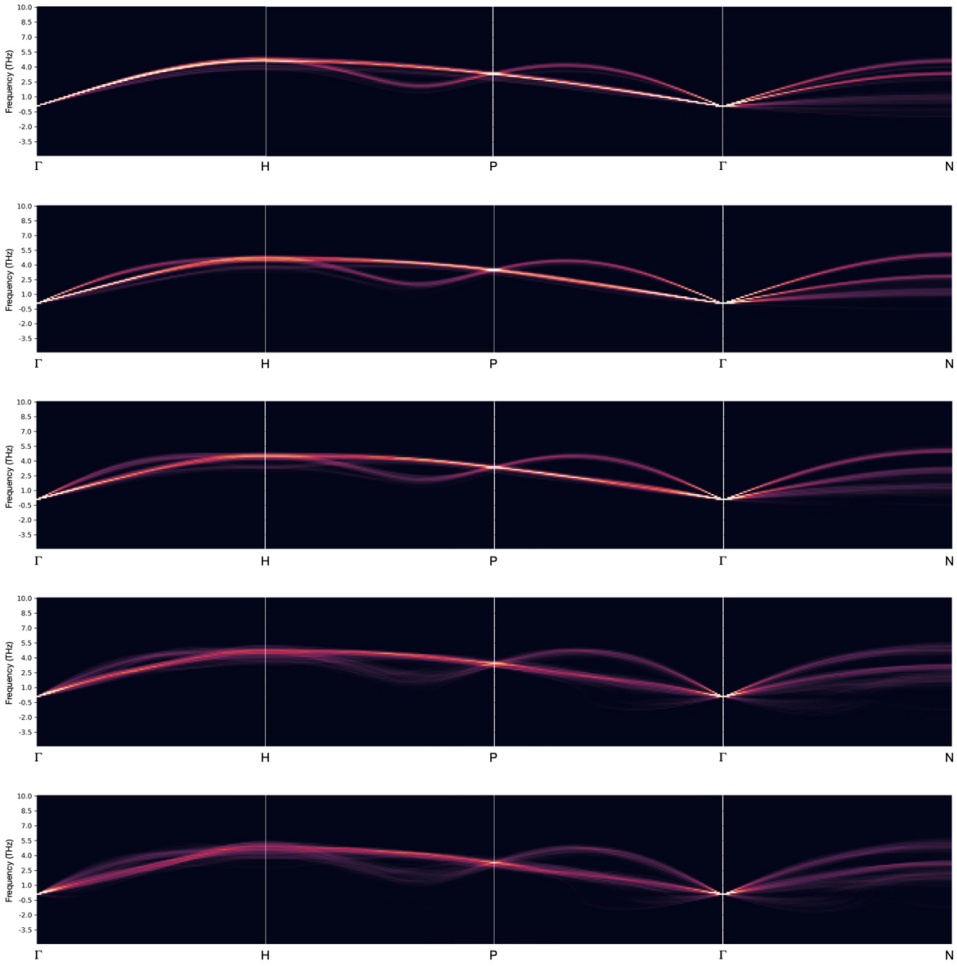


Figure A.2: Heat maps of active learning run 29. From top to bottom: 1^{st} nearest neighbors to 5^{th} nearest neighbors

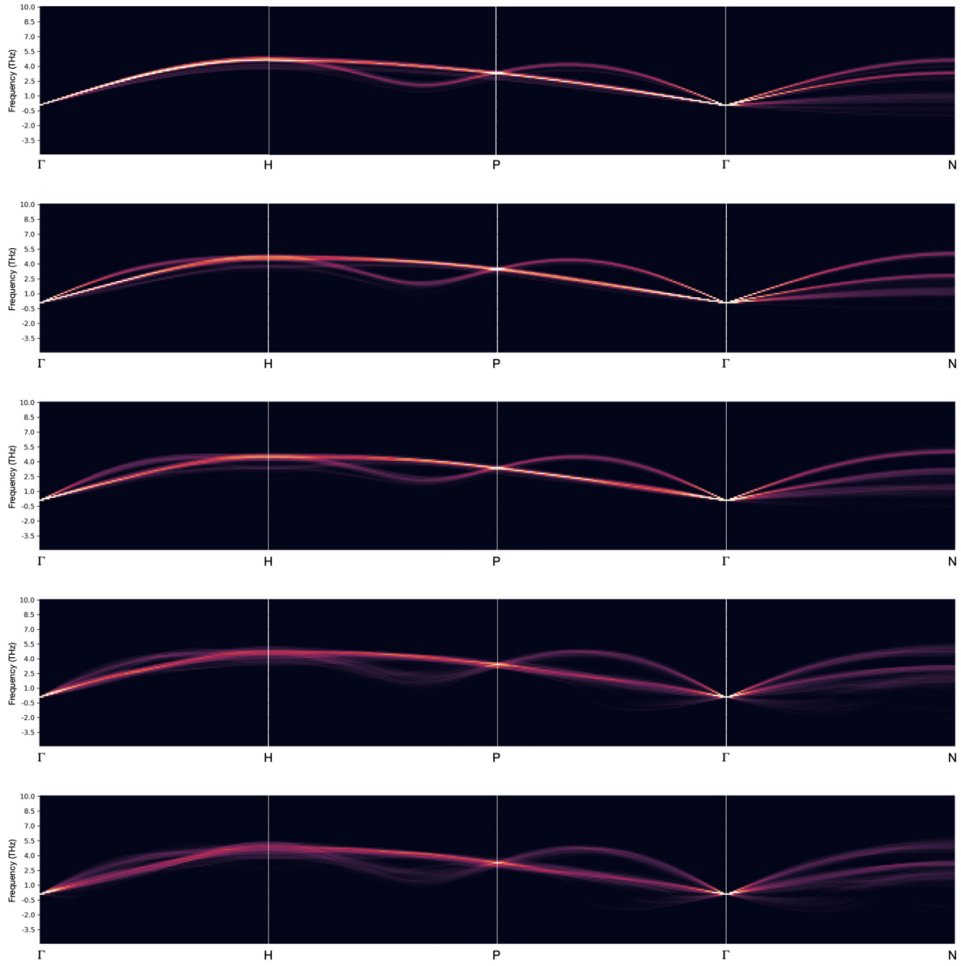


Figure A.3: Heat maps of active learning run 30. From top to bottom: 1^{st} nearest neighbors to 5^{th} nearest neighbors

References

[1] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical Review Letters*, vol. 98, no. 14, p. 14640, 2007.

[2] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "Schnet: A continuous-filter convolutional neural network for modeling quantum interactions," *Advances in neural information processing systems*, vol. 30, 2017.

[3] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: Moving beyond fingerprints," *Journal of Computer-Aided Molecular Design*, vol. 30, no. 8, pp. 595–608, 2016.

[4] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[5] R. Kondor, Z. Lin, and S. Trivedi, "Clebsch-gordan nets: a fully fourier space spherical convolutional neural network," *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, vol. 31, 2018.

[6] N. Thomas, T. Smidt, S. M. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, "Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds," *CoRR*, vol. abs/1802.08219, 2018.

[7] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. Cohen, "3d steerable cnns: Learning rotationally equivariant features in volumetric data," *CoRR*, vol. abs/1807.02547, 2018.

[8] R. J. L. Townshend, B. Townshend, S. Eismann, and R. O. Dror, "Geometric prediction: Moving beyond scalars," *CoRR*, vol. abs/2006.14163, 2020.

[9] K. Schütt, O. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.), vol. 139 of *Proceedings of Machine Learning Research*, pp. 9377–9388, PMLR, 18–24 Jul 2021.

[10] M. Geiger, T. Smidt, A. M., B. K. Miller, W. Boomsma, B. Dice, K. Lapchevskyi, M. Weiler, M. Tyszkiewicz, S. Batzner, M. Uhrin, J. Frellsen, N. Jung, S. Sanborn, J. Rackers, and M. Bailey, "Euclidean neural networks: e3nn," 2020.

[11] M. Laue, "Eine quantitative prüfung der theorie für die interferenzerscheinungen bei röntgenstrahlen," *Annalen der Physik*, vol. 346, no. 10, pp. 989–1002, 1913.

[12] W. L. Bragg, "The specular reflection of x-rays," *Nature*, vol. 90, p. 410, 1912.

[13] W. H. Bragg and W. L. Bragg, "The reflection of x-rays by crystals," *Proc. R. Soc. Lond.*, vol. 88, pp. 428–438, 1913.

[14] H. Georgi, *Lie Algebras In Particle Physics from Isospin To Unified Theories*. Westview Press, 2 ed., 1999.

[15] C. Stassis, J. Zarestky, and N. Wakabayashi, "Lattice dynamics of bcc zirconium," *Physical Review Letters*, vol. 41, no. 25, p. 1726, 1978.

[16] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review*, vol. 136, no. 3B, pp. B864–B871, 1964.

[17] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, no. 4A, pp. A1133–A1138, 1965.

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

[19] P. E. Blöchl, "Projector augmented-wave method," *Physical Review B*, vol. 50, no. 24, pp. 17953–17979, 1994.

[20] J. E. Jones, "On the determination of molecular fields. ii. from the equation of state of a gas," *Proc. R. Soc. Lond.*, vol. 106, pp. 463–477, 1924.

[21] S. D. Murray and M. I. Baskes, "Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals," *Physical Review B*, vol. 29, no. 12, p. 6443–6453, 1983.

[22] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Phys Rev Lett.*, vol. 55, pp. 2471–2474, 1985.

[23] D. Marx and J. Hutter, *Ab initio molecular dynamics basic theory and advanced methods*. Cambridge University Press, 1 ed., 2009.

[24] A. Heiming, W. Petry, J. Trampenau, M. Alba, C. Herzig, H. R. Schober, and G. Vogl, "Phonon dispersion of the bcc phase of group-iv metals. ii. bcc zirconium, a model case of dynamical precursors of martensitic transitions," *Physical Review B*, vol. 43, p. 10948, 1991.

[25] K. Masuda-Jindo, S. R. Nishitani, and V. V. Hung, "hcp-bcc structural phase transformation of titanium: Analytic model calculations," *Physical Review B*, vol. 70, p. 184122, 2004.

[26] Y.-Y. Ye, Y. Chen, K.-M. Ho, B. N. Harmon, and P.-A. Lindgård, "Phonon-phonon coupling and the stability of the high-temperature bcc phase of zr," *Physical Review Letters*, vol. 58, no. 17, p. 1769, 1987.

[27] S. Kadkhodaei and A. Davariashtiyani, "Phonon-assisted diffusion in bcc phase of titanium and zirconium from first principles," *Physical Review Materials*, vol. 4, no. 4, p. 043802, 2020.

[28] B. S. Hickman, "The formation of omega phase in titanium and zirconium alloys: A review," *Journal of Materials Science*, vol. 4, pp. 554–563, 1969.

[29] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," *Journal of Physics: Condensed Matter*, vol. 21, p. 395502, sep 2009.

[30] A. Togo and I. Tanaka, "First principles phonon calculations in materials science," *Scr. Mater.*, vol. 108, pp. 1–5, Nov 2015.

[31] S. Kadkhodaei, Q.-J. Hong, and A. van de Walle, "Free energy calculation of mechanically unstable but dynamically stabilized bcc titanium," *Physical Review B*, vol. 95, no. 6, p. 064101, 2017.

[32] Y. LeCun, C. Cortes, and C. J. Burges, "The mnist database of handwritten digits." `http://yann.lecun.com/exdb/mnist/`.

[33] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.

[34] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nature communications*, vol. 8, no. 1, p. 13890, 2017.

[35] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[36] N. Roy and A. McCallum, "Toward optimal active learning through monte carlo estimation of error reduction," *Proc. of the 18th Int. Conf. on Machine Learning (ICML)*, p. 441–448, 2001.

[37] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," *CoRR*, vol. abs/cmp-lg/9407020, 1994.

[38] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *CoRR*, vol. cs.AI/9603104, 1996.

[39] T. Smidt, "Euclidean symmetry and equivariance in machine learning," *Trends in Chemistry*, 2021.

[40] Y. Tang and W. A. de Jong, "Prediction of atomization energy using graph kernel and active learning," *CoRR*, vol. abs/1810.07310, 2018.

Vita

Vanessa Judith Meraz graduated from the University of Texas at El Paso in 2020 with a Bachelor of Science in Physics.