

Accelerating vibrational free energy calculations for anharmonic crystals

Mark Steven Johnson

Theory of Condensed Matter Group
Cavendish Laboratory



This dissertation is submitted for the degree of
Doctor of Philosophy
at the University of Cambridge

January 2021
Darwin College

Supervised by
Prof. Richard J. Needs

Declaration

This dissertation describes work carried out between October 2016 and January 2021 in the Theory of Condensed Matter group at the Cavendish Laboratory, Cambridge, under the supervision of Prof. R. J. Needs.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements and specified in the text. This dissertation has not been submitted in whole or in part for any other degree or diploma at this or any other university. This dissertation does not exceed the word limit of 60,000 words.

Mark Johnson
Cambridge, January 2021

Abstract

Accelerating vibrational free energy calculations for anharmonic crystals

Mark Johnson

This thesis presents a study of vibrational modelling of periodic solids, focussing in particular on vibrational self-consistent field theory (VSCF).

Firstly, the mathematical framework within which vibrational modelling takes place is detailed. The approximations which must be made are introduced, and the limitations imposed by these approximations are explored. Then the existing approaches to vibrational modelling are reviewed, and the modelling process is separated into three interrelated problems: representing the vibrational potential energy surface, fitting the free parameters of that representation using electronic structure theory, and using the fitted representation to calculate the vibrational wavefunctions and the vibrational free energy.

A number of developments to the VSCF method are described. These focus on exploiting the notions of symmetry and size consistency to reduce the computational cost of the method and to improve its accuracy and reliability. The discussions of symmetry and size consistency also reveal a number of fundamental problems with the VSCF method. Solutions are proposed to several of these problems, but other problems remain unsolved, and it is proposed that these problems may imply that VSCF is not well suited for modelling crystals.

A software package which implements the ideas contained within this thesis is introduced, and the methods used to encode VSCF as accurate and efficient software are detailed. The results of an example application of this software package are also presented.

Acknowledgements

I would like to thank the principle investigators in TCM who have supported me along the way: Richard Needs for being my supervisor and making everything possible; Mike Payne for helping me navigate the world of being a PhD student, for proofreading my thesis, and for continually finding additional funding whenever it was needed; Chris Pickard for keeping the group together and providing direction when the pandemic tried to shut everything down; and Tomeu Monserrat for providing his version of `caesar`, from which my version of `caesar` was born, and for giving plentiful help and advice as I learned the theory.

I would also like to thank everyone who used `caesar` and provided much-needed testing and feedback, often long before the code was remotely ready for real use. In particular, I would like to thank Martin Schlegel, who also helped write the interface between `caesar` and QUIP and who provided the titanium MEAM potential used in this work, and Alice Shipley, Michael Hutcheon and Jack Whaley-Baldwin, who, in addition to their roles as `caesar` testers also collectively made the Needs group a thought-provoking and entertaining place to work.

I would also like to thank my office-mates; Will Grant, Matthew Evans and Matthew Smith; who somehow kept me sane through four-and-a-half years of PhD work, and who provided many many hours of entertainment, discussion and other procrastination. Thanks also go to the motley cast and crew of the TCM coffee area, without whom life would have been so much duller and less unpredictable. Thanks in particular to Mike Payne, again, for running the show.

Last, but certainly not least, thanks must go to the incomparable Michael Rutter, whose many roles include being the group's sys-admin and archivist, the ever-available hardware maestro and Fortran guru, the bearer of the weekly news, and the bottomless pool of esoteric wisdom. What would TCM do without you, Michael?

Glossary

Crystallographic Labels and Co-ordinates

- R** An R-vector, a sum of lattice vectors.
- G** A G-vector, a sum of reciprocal lattice vectors.
- q** A q-point, a vector in reciprocal space.
- L* The lattice matrix, whose rows are the primitive R-vectors.
- ρ A displacement index, a directional index in real space, combining an R-vector, a primitive cell atom index and a cartesian direction.
- χ A normal-mode index, a directional index in reciprocal space, combining a q-point and a band index.
- d_ρ The mass-weighted displacement along the ρ direction.
- u_χ The displacement along the χ normal mode.
- $U_{\rho\chi}$ The transformation matrix between d_ρ and u_χ co-ordinates.
- ξ A subspace of degenerate modes.
- ζ A combination of subspaces.

Operators and Thermodynamic Variables

- \mathcal{H} The Hamiltonian.
- \mathcal{T} The kinetic energy.
- V The potential energy.
- T The temperature.
- S The entropy.
- Ω The volume.
- U The internal energy.
- F The Helmholtz free energy.
- G The Gibbs free energy.
- Z The partition function.

Acronyms

PES	Born-Oppenheimer potential energy surface.
PAA	The principal axes approximation.
HF	Hartree-Fock theory.
VSCF	Vibrational self-consistent field theory.
MP	Møller-Plesset perturbation theory.
VMP n	Vibrational Møller-Plesset theory at n th order.
CI	Configuration interaction.
VCI	Vibrational configuration interaction.
CC	Coupled cluster.
VCC	Vibrational coupled cluster.
DFT	Density-functional theory.
PIMD	Path integral molecular dynamics.
RPMD	Ring-polymer molecular dynamics.
DMC	Diffusion Monte-Carlo.
PIMC	Path integral Monte-Carlo.

Mathematical Symbols

$:O:$	The normal-ordering of operator O .
$n!!$	The double factorial of n . $(2n - 1)!!$ is equal to the product of the first n odd numbers.
c.c.	The complex conjugate of the preceding term.
$\hat{\mathbf{x}}$	A unit vector or basis vector along the direction of vector \mathbf{x} .
$\mathbb{1}$	The identity matrix.
$\mathbf{x} \otimes \mathbf{y}$	The outer product of vectors \mathbf{x} and \mathbf{y} .
\dot{x}	The first derivative of x with respect to time.

Contents

1	Introduction	12
1.1	Periodic crystals	13
1.2	Electronic and vibrational structure	15
1.3	Finite temperature anharmonic modelling	15
1.4	Material properties	16
1.5	Thesis outline	17
2	Theoretical groundwork	19
2.1	Necessary approximations	20
2.2	Periodic co-ordinates	26
2.3	Supercells	33
2.4	Symmetry	37
2.5	Summary	41
3	The problem of free energy	43
3.1	Framing the problem	44
3.2	The harmonic approximation	46
3.3	Sampling the potential energy surface	51
3.4	Representing the potential energy surface	56
3.5	Calculating free energy	61
3.6	Calculating other properties	78
4	Mapping the potential energy surface	81
4.1	Symmetry and the potential energy surface	81
4.2	Size consistency and the potential energy surface	87
4.3	Exploiting the finite range of interactions	95
4.4	Fitting the potential energy surface	96

5	Vibrational self-consistent field theory	105
5.1	Symmetry and vibrational self-consistent field theory	105
5.2	Size consistency and vibrational self-consistent field theory	109
5.3	The dispersionless anharmonic oscillator	114
5.4	Calculating free energy	120
5.5	Converging the self-consistent coefficients	126
6	Introducing caesar	134
6.1	Electronic structure as a black box	138
6.2	Interchangeable calculation elements	139
6.3	Exact symmetry representation	140
6.4	Snap to symmetry	140
6.5	Generating single-subspace potentials	140
7	Calculating the phase diagram of elemental titanium	142
7.1	The phases of titanium	143
7.2	Calculating the phase diagram from the free energy	144
7.3	Existing calculations	146
7.4	Calculating properties under the harmonic approximation	146
7.5	Calculating anharmonic properties	152
8	Conclusions	168
8.1	Representations of the potential energy surface	169
8.2	Vibrational self-consistent field theory	170
8.3	Final thoughts	171
A	Harmonic eigenstates	185
B	Thermodynamic addition and subtraction	191
C	Copyright permissions	193

Chapter 1

Introduction

The quest to understand how materials behave and to exploit this behaviour is one of the pillars of modern science, and one of the driving forces behind modern technology [1, 2].

Historically, there were two branches of materials science: physical experiments, in which new materials were thoroughly tested; and thought experiments, in which the underlying physics was investigated and its consequences examined [3]. However, experiments can be costly endeavours, and require the extended efforts of many experts in order to make progress [4]. Moreover, thought experiments become infeasible as the systems being considered become more complex [5], and many materials are very complex indeed [6].

With the advent of modern computing, a third branch of materials science has been created [7, 8]: computational physics, where the properties of materials are investigated using computer simulations of large and complex material models. Such simulations bridge the gap between theory and experiment, allowing the real-world consequences of new theories to be calculated so that they can be compared to experiment, and providing a fast and accurate method of identifying which new systems are likely to prove fruitful for experimentalists to study [1, 2, 4, 9].

Computational physics also dramatically accelerates the discovery and selection of new materials for technological applications [10]. Fast and accurate computer simulations allow millions of candidate materials to be investigated for their suitability to a given role, in the time it would take an experimental lab to investigate only a handful of materials and at a fraction of the cost.

In order for materials simulations to be useful, they must be accurate. This means that they must be based on models which are predictive rather than simply explanatory, and which are transferable between different materials. The models which are maximally accurate, predictive and transferrable are those which are

based entirely on the underlying physics of quantum mechanics, without reference to artificially generated model parameters. These models are described as ab-initio, and are considered to be the gold standard of materials modelling [1, 2]. Unfortunately, ab-initio models are also generally the most computationally demanding models. This has previously limited their application to simple systems where few calculations are needed, but as computing power increases and algorithms are improved they are becoming appropriate for an increasing number of tasks [2, 9].

Computer models are not only ideally suited to studying known materials, they are also capable of predicting entirely new ones [10]. Computational materials discovery is a rapidly developing field, capable of generating vast numbers of candidate materials in a very short space of time, a feat which is revolutionising materials science [9]. In order to make maximal use of these new materials, computational models must be capable of accurately and autonomously processing many structures very quickly. This is a paradigm called high-throughput computing, and is of increasing relevance to materials science [10].

The desire for highly accurate methods which are suitable for high-throughput computing has created a need to accelerate ab-initio methods without compromising their accuracy. This work aims to make some little progress in this endeavour, presenting a number of improvements to the ab-initio modelling of periodic crystals, attempting to improve their speed, accuracy and stability to bring them a little closer to that goal.

1.1 Periodic crystals

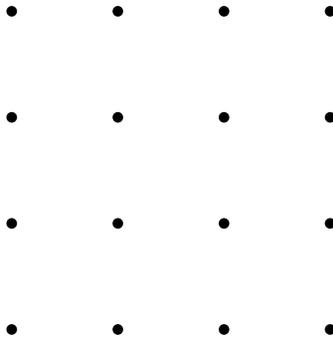
Crystals are a widely-occurring type of solid matter, distinguished in that they are composed of a microscopic motif of atoms which is repeated with regular periodicity in all directions throughout a macroscopic volume [11]. This is depicted in figure 1.1, which shows the atomic motif, the repeating lattice of points which defines the periodicity of the crystal, and the crystal itself which is composed of a motif of atoms at each point in the lattice. The periodic nature of crystals means that the properties of a crystal are usually highly homogeneous, and these properties are directly determined by the arrangement of the atomic motif which repeats to form the crystal.

This has direct implications for scientific and technological purposes; the sharp dependence of the properties of the macroscopic crystal on the structure of the microscopic motif means that experiments can probe microscopic structure while only having to interact with macroscopic quantities of material, and materials with given macroscopic properties can be designed by only varying microscopic atomic arrangements.

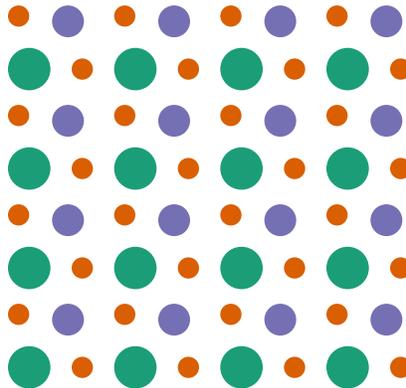
This is also convenient for computational modelling, since many of the properties of the whole crystal can be captured by a model which only includes a small section of the crystal [12].



(a) The motif of atoms which is repeated throughout a crystal. Each circle represents an atom, with atoms of different species shown with different colours and sizes.



(b) A section of a crystal lattice. Each black dot represents a single point in the lattice, at which a motif of atoms will be situated. In an idealised crystal, the lattice repeats infinitely in all directions.



(c) A section of a crystal, composed of a copy of the atomic motif at every point in the lattice.

Figure 1.1

There are a number of details of modelling crystals which are different from the modelling of other materials.

Firstly, every crystal has *crystal symmetries* [13]. Each crystal symmetry is a combination of a translation, a rotation and a reflection which leaves the crystal unchanged. Every crystal has purely translational symmetries in all spatial dimensions, and many crystals also have other symmetries.

Secondly, crystal models should be *size-consistent* [14]. This means that models of differently sized sections of the crystal must give results which scale correctly with the size of the crystal; *extensive* quantities such as the total energy must scale linearly with the number of atoms simulated, and *intensive* quantities such as pressure must not scale with the number of atoms.

These properties can be a help or a hinderance; symmetry and size consistency can often be exploited to duplicate information and reduce the amount of computation necessary, but they also complicate the model,

and any model which does not respect symmetry and size consistency is likely to be inaccurate. This work aims to develop methods which guarantee symmetry invariance and size consistency, and which use these properties to accelerate computation.

1.2 Electronic and vibrational structure

Crystals are composed of two kinds of particle: massive nuclei, and light electrons¹. When modelling a crystal, the differing masses of these particles allows their motions to be separated. From the perspective of the fast-moving electrons the nuclei are effectively stationary, while from the perspective of the slow-moving nuclei the electrons are only seen as a fast-moving blur [15].

This leads to two distinct fields of materials modelling: *electronic structure* calculates the properties of electrons in a stationary potential defined by fixed nuclei, while *vibrational structure* calculates the properties of the nuclei in a potential which is defined by the average properties of the electronic motion.

Modelling electronic structure is a mature field, with many widely-used implementations of many different methods available [16–20]. In contrast, modelling vibrational structure is much less developed. This work will treat electronic structure as a solved problem, and focus only on vibrational structure.

1.3 Finite temperature anharmonic modelling

Vibrational motion is routinely neglected from calculations. This is equivalent to treating nuclei as classical particles at zero temperature, which simply remain stationary at the bottom of the potential well defined by the electrons. This is a gross approximation, although there are many applications for which this is sufficient.

In order to move beyond the classical approximation, the nuclei must be treated as quantum particles. This means that the crystal properties are defined in terms of vibrational wavefunctions, which are in turn defined with respect to a Hamiltonian which depends on the potential felt by the nuclei [21]. Neither the Hamiltonian nor the vibrational wavefunctions can be calculated or stored exactly, and so both must be approximated. Even coarse approximations require significant computation, and highly accurate approximations can be extremely computationally demanding [22].

It is common practice to take the *harmonic approximation*, which approximates the vibrational potential as a quadratic function of the nuclear co-ordinates. This leads to a particularly simple Hamiltonian, and to wavefunctions and crystal properties which can be calculated with relative ease [2]. While the harmonic

¹At the conditions which most crystal models are concerned with, it can be assumed that nuclear physics can be neglected, and that the effects of quantum electrodynamics can be simplified to electromagnetic forces between electrons and nuclei.

approximation is significantly more accurate than neglecting vibrational motion entirely, it still has many limitations [23]. Methods which go beyond the harmonic approximation are *anharmonic*, and will be the focus of this work.

Nuclear motion is often modelled at a temperature of absolute zero, where each nucleus vibrates as little as possible. This simplifies calculations, since then only the low-energy regions of the potential need to be modelled accurately, and the nuclear motion can be described using only the ground-state wavefunction. However, this brings limitations; the properties of many materials are qualitatively different at finite temperature than at zero temperature, and a large proportion of these materials adopt completely different phases as temperature changes [24]. This work will focus on methods which are capable of explicitly modelling crystals at finite temperatures.

1.4 Material properties

The point of materials modelling is to calculate the physical properties of materials. But which properties are of interest varies, and some properties are easier to calculate than others.

The most fundamental property of a material is its stability [2]. Different phases of a material have different stabilities, and the most stable phase will occur predominantly over the others. The calculation of material stability of different phases at different temperatures and pressures gives the phase diagram of that material, and calculating such phase diagrams accurately often requires anharmonic calculation [24, 25].

The electronic properties of a crystal are affected by its vibrational properties, and anharmonic calculations have been shown to improve the accuracy of electronic band structure calculations [26–28], along with more exotic properties including ferroelectricity [29], thermoelectricity [30], exciton energies [31], charge-density waves [29] and superconductivity [30].

A number of crystals have been found which are superconductors at high temperatures, where anharmonic effects are more likely to be important [32]. With increased understanding of these crystals, it may be possible to predict novel crystals which are superconductors at even higher temperatures, with the holy grail being a crystal which is a superconductor in ambient conditions [29].

Understanding the optical properties of crystals enables the design of new photovoltaic and optoelectronic devices [33]. Optical properties also play a critical role in experiments designed to identify unknown materials, and to study the dynamics of known materials [34]. Different materials can only be distinguished by their optical properties if the computational models of these properties are in good agreement with experimental data [33, 35]. Accurately calculating optical spectra, including for infra-red, Raman, THz and

X-ray spectroscopy, requires finite-temperature calculation, and including anharmonicity has been shown to improve the accuracy of a variety of results [2, 33, 35–39].

In addition to optical spectroscopy, vibrational properties are relevant when calculating other spectroscopic signatures, including inelastic neutron scattering spectroscopy [2], electron energy loss spectroscopy [40] and nuclear magnetic resonance spectroscopy [38].

Understanding the thermal properties of crystals is vital for many applications [41]. Modelling thermal properties clearly requires finite-temperature modelling [42], and this can be improved using anharmonic calculations. Thermal properties of interest include the thermal conductivity [43–45], thermo-elastic properties [2, 46] and thermal expansion [2, 47, 48]. Of particular topical relevance is the phenomenon of negative thermal expansion, describing materials which shrink when heated. Negative thermal expansion is entirely driven by anharmonic effects [47, 49].

The vibrational properties of materials are also of interest in and of themselves. The burgeoning field of phononics, which seeks to harness atomic vibrations, offers many exciting possibilities [50].

Vibrational effects also play an important role in how a crystal changes over time, affecting atomic transport and defect formation within crystals [51].

The accurate calculation of all of these properties depends on the accurate calculation of the vibrational Hamiltonian and the vibrational wavefunctions. This work will focus on improving the calculation of the Hamiltonian, wavefunctions and free energy, and will leave the calculation of other properties from these variables to other authors.

1.5 Thesis outline

Chapter 2 details the framework in which vibrational modelling takes place, introducing the approximations which must be made, and the notation and definitions which will be used throughout the rest of this work.

Chapter 3 introduces the problem of calculating free energy, which is decomposed into three interrelated problems: representing the vibrational potential, fitting the free parameters of that representation using electronic structure calculations, and using this information to calculate the vibrational wavefunctions and the vibrational free energy. The existing methods in each of these areas are reviewed, with an eye for their relative advantages and disadvantages.

Chapter 4 explores a number of improvements to the representation and fitting of the vibrational potential energy surface. Methods are proposed which guarantee the symmetry invariance and size consistency of

this potential, and which exploit these properties and others in order to reduce the cost of calculating the potential.

Chapter 5 explores a number of improvements to the method of vibrational self-consistent field theory. The method is adapted to respect symmetry and size consistency, and a number of accelerations to the method are proposed.

Chapter 6 introduces `caesar`, the software package which accompanies this work. This code implements many of the features described in this work.

Chapter 7 presents an example application of vibrational self-consistent field theory, using `caesar` to study the solid phase diagram of a classical potential which describes elemental titanium.

And finally, chapter 8 concludes the work, summarising the progress made and the problems which remain unsolved, and offering thoughts on some avenues of research which might prove fruitful in light of the work presented here.

Chapter 2

Theoretical groundwork

Any computational model requires a mathematical description of the system it is modelling. As with many systems, there are many possible descriptions of periodic crystals, and each of these descriptions inherently makes certain assumptions about the behaviour of the crystal.

In addition, a model of an arbitrary crystal in full generality would be extremely complicated, and too intractable to solve at a reasonable computational cost. As such, further assumptions and approximations must be made so that the model becomes tractable to solve.

This chapter details the formalism of the specific description of crystals used throughout this work, and the assumptions and approximations which are made. It will also give an overview of the strengths and limitations of this description.

Nothing of note in this chapter is new, and indeed much of it has been the subject matter of textbooks for many years [12, 52, 53], but it still bears repeating as it forms the groundwork on which the rest of this work is based, and much of the notation used varies between sources.

A note on units

Throughout this work, Hartree atomic units [54] will be used, such that the reduced Planck constant \hbar , the electron charge e , the Bohr radius a_0 and the electron mass m_e are all unity. In addition, temperature and entropy will be measured in Planck units [55], such that Boltzmann's constant k_B is unity.

2.1 Necessary approximations

2.1.1 Thermal equilibrium

This work is only concerned with modelling systems in thermal equilibrium. That is to say, any given system will have a well-defined temperature T , which will be assumed to be constant and unchanging with time. Under this approximation, the system will be unchanging, and its physical properties will not be functions of time.

There are a number of non-equilibrium quantities, such as transport and scattering coefficients, which can be calculated using equilibrium modelling [45]. However, there are also a number of quantities which cannot be calculated in such a manner, including the details of phase transitions [56], for which other models are needed.

2.1.2 The Born-Oppenheimer approximation

A 3-dimensional system containing \mathcal{N}^n nuclei and \mathcal{N}^e electrons is described in terms of the $3\mathcal{N}^n$ -dimensional nuclear co-ordinate \mathbf{r}^n and the $3\mathcal{N}^e$ -dimensional co-ordinate \mathbf{r}^e .

The system is governed by the *non-relativistic time-independent Schrödinger equation*¹, which in its most general form can be written in these co-ordinates as [18]

$$\mathcal{H}(\mathbf{r}^n, \mathbf{r}^e)\psi_i(\mathbf{r}^n, \mathbf{r}^e) = E_i\psi_i(\mathbf{r}^n, \mathbf{r}^e), \quad (2.1)$$

where \mathcal{H} is the non-relativistic Hamiltonian of the system, ψ_i is the wavefunction of the i 'th eigenstate of the system, and E_i is the energy of state i .

The Hamiltonian can be separated into a number of parts,

$$\mathcal{H}(\mathbf{r}^n, \mathbf{r}^e) = \mathcal{T}^n(\mathbf{r}^n) + V^n(\mathbf{r}^n) + \mathcal{T}^e(\mathbf{r}^e) + V^e(\mathbf{r}^e) + V^{e-n}(\mathbf{r}^n, \mathbf{r}^e) \quad (2.2)$$

where \mathcal{T}^n and \mathcal{T}^e are the nuclear and electronic kinetic energy operators respectively, V^n and V^e are the potential energy operators corresponding to purely nuclear interactions and purely electronic interactions respectively, and V^{e-n} is the potential energy operator corresponding to interactions between electrons and nuclei.

¹In some cases, particularly when dealing with very heavy elements, it is necessary to treat electrons relativistically. This changes the derivation of the Born-Oppenheimer Hamiltonian slightly, but since the nuclei can always be treated non-relativistically the end result is unchanged.

With full generality, the wavefunction of any state can be expressed in terms of nuclear and electronic parts as

$$\psi_i(\mathbf{r}^n, \mathbf{r}^e) = \psi_i^n(\mathbf{r}^n) \sum_j c_{ij\mathbf{r}^n} \psi_{j\mathbf{r}^n}^e(\mathbf{r}^e), \quad (2.3)$$

where ψ_i^n is a nuclear wavefunction which depends purely on the nuclear co-ordinate, $\{\psi_{j\mathbf{r}^n}^e\}$ are electronic wavefunctions which are each defined for a given value of \mathbf{r}^n but which depend purely on the electronic co-ordinate, and $\{c_{ij\mathbf{r}^n}\}$ are complex scalar coefficients.

It is convenient to consider nuclear motion independent of electronic motion. This becomes possible if the *Born-Oppenheimer approximation* is taken [15, 18].

Approximation 1 (The Born-Oppenheimer approximation):

Changes in energy due to changing the electronic state are far larger than those due to changing the nuclear state.

Under this approximation, electronic relaxation happens far faster than nuclear motion, and so the coefficients $c_{ij\mathbf{r}^n}$ become independent of the nuclear state, and can be replaced with universal coefficients $c_{j\mathbf{r}^n}$, which are simple thermal weightings. As such, the eigenstates become

$$\psi_i(\mathbf{r}^n, \mathbf{r}^e) = \left(\psi_i^n(\mathbf{r}^n) \right) \left(\sum_j c_{j\mathbf{r}^n} \psi_{j\mathbf{r}^n}^e(\mathbf{r}^e) \right). \quad (2.4)$$

This allows the Schrödinger equation to be decoupled and solved in two parts. Firstly, the electronic Schrödinger equation is solved at fixed \mathbf{r}^n ,

$$\langle \mathbf{r}^n | \mathcal{H}(\mathbf{r}^n, \mathbf{r}^e) | \mathbf{r}^n \rangle \psi_{j\mathbf{r}^n}^e(\mathbf{r}^e) = E_{j\mathbf{r}^n}^e \psi_{j\mathbf{r}^n}^e(\mathbf{r}^e), \quad (2.5)$$

which gives the electronic states and the corresponding energies. The energies can then be used to calculate $\{c_{j\mathbf{r}^n}\}$.

Taking the expectation of the Hamiltonian with respect to the electronic states gives the *Born-Oppenheimer Hamiltonian* $\mathcal{H}^{\text{B-O}}(\mathbf{r}^n)$, defined as

$$\mathcal{H}^{\text{B-O}}(\mathbf{r}^n) = \sum_j c_{j\mathbf{r}^n} \langle \psi^e |_{j\mathbf{r}^n} \mathcal{H}(\mathbf{r}^n, \mathbf{r}^e) | \psi^e \rangle_{j\mathbf{r}^n}. \quad (2.6)$$

Secondly, the nuclear Schrödinger equation is solved using $\mathcal{H}^{\text{B-O}}$,

$$\mathcal{H}^{\text{B-O}}(\mathbf{r}^n) \psi_i^n(\mathbf{r}^n) = E_i^n \psi_i^n(\mathbf{r}^n). \quad (2.7)$$

There are many ways of finding or approximating the solution to equation 2.5, collectively known as electronic structure methods. A number of these methods will be detailed in section 3.3.1. This work will focus on equation 2.7 only, and it will simply be assumed that equation 2.5 can be solved to some approximation. The accuracy of that approximation will affect the accuracy of the overall calculation, but no further consideration of this will be made.

While there exist methods which do not assume the Born-Oppenheimer approximation [57–60], these are generally significantly more computationally expensive than the methods which do assume this approximation, and should only be used for systems where the Born-Oppenheimer approximation is not appropriate.

The Born-Oppenheimer potential energy surface

The Born-Oppenheimer Hamiltonian can be separated into two parts,

$$\mathcal{H}^{\text{B-O}} = \mathcal{T}^{\text{n}} + V^{\text{B-O}}. \quad (2.8)$$

The nuclear kinetic energy \mathcal{T}^{n} is unchanged from its previous definition, but the operator $V^{\text{B-O}}$ is new. This is the *Born-Oppenheimer potential energy surface* (PES). It defines the effective nuclear potential which includes the mean field effect of the electrons.

The remainder of this work will focus on the Born-Oppenheimer Schrödinger equation, and so the superscripts ⁿ and ^{B-O} will be dropped, and all references to co-ordinates, Hamiltonians and potentials will refer to nuclear co-ordinates, Born-Oppenheimer Hamiltonians and Born-Oppenheimer potentials respectively, unless stated otherwise.

The Born-Oppenheimer approximation assumes that there is no correlation between the electronic and nuclear motion in the system. This is always an approximation, but there are systems where this approximation holds better than others [61]. In many systems, in particular most wide-gap insulators, the Born-Oppenheimer approximation is highly accurate, and the corrections to this approximation can generally be neglected. In some systems, including many conventional superconductors [30], there is some electron-phonon correlation, but this can be accounted for using perturbative methods. In other systems the electron-phonon correlation is so strong that perturbative methods are not sufficient to produce accurate results [57]. In these systems, methods based on the Born-Oppenheimer approximation cannot be used, and other techniques must be used instead.

2.1.3 The static-lattice approximation

Having taken the Born-Oppenheimer approximation, there are many ways of modelling the motion of the nuclei in the PES. The simplest of these is the *static-lattice approximation*, which assumes that the nuclei are entirely motionless classical particles at zero temperature.

Under this approximation, the energy U and free energy F of the system are simply the value of the PES at the nuclear configuration, $U = F = V(\mathbf{r})$. The structures which are stable under the static-lattice approximation are those for which the free energy is minimised, and so the static-lattice configurations of these structures are those which minimise $V(\mathbf{r})$. Structures which are stabilised by vibrational effects may be at PES minima, but may also be at saddle points of the PES [62, 63].

The static-lattice approximation is a gross approximation, but it is also the approximation which is the easiest to calculate. As such, many computational methods use this approximation as a first guess for the free energy [21]. Since the contribution to the free energy from nuclear motion is typically much smaller than that of the electronic structure, this approximation is often good enough to separate systems into those with a low enough free energy that they might be stable, and those whose free energy is so high that they will be unstable regardless of their vibrational properties [10].

There are many limitations to the static-lattice approximation. Since the nuclear motion is modelled as entirely stationary, independent of temperature, many quantities cannot be calculated, including heat capacities, thermal conductivities, transport coefficients, scattering coefficients and many others. Additionally, since the nuclei are modelled as classical particles, their zero-point motion is not taken into account. This means that the modelled free energy is typically an underestimate of the true value, even at zero temperature. Finally, and perhaps most problematically, under the static-lattice approximation the nuclear entropy is always zero, meaning that this approximation cannot meaningfully be used for any thermodynamic calculations.

The static-lattice approximation provides a useful theoretical tool: the *static-lattice co-ordinate* \mathbf{r}^0 . This is the $3\mathcal{N}$ -dimensional co-ordinate describing the nuclei in their static-lattice configuration, which is always a stationary point of the PES. This co-ordinate is one of the building blocks of the notation which is used to describe all models of nuclear motion. The energy of the system at the static-lattice co-ordinate, $V(\mathbf{r}^0)$, is the *static-lattice energy*, and is the energy relative to which all other energies in this work will be defined.

2.1.4 The vibrational approximation

While the static-lattice approximation is overly restrictive, it is still necessary to make a number of assumptions about the nature of nuclear motion before it becomes possible to model that motion. The first such approximation is that crystals are perfectly elastic solids. There are a number of equivalent definitions of this approximation, but the most natural in this context is

Approximation 2 (Perfect elasticity):

The nuclear co-ordinate \mathbf{r} is confined to a bounded region of phase space.

This region of space will be assumed to include the static-lattice co-ordinate \mathbf{r}^0 . There are a number of consequences to this approximation, including that there can be no flux of material over time.

Defining the minimum value of the PES within this region as $V_{\min} = \min_{\mathbf{r}} V(\mathbf{r})$, the elastic approximation is reasonable provided that the region is bounded by a closed contour $V(\mathbf{r}) = V_{\text{boundary}}$, where $V_{\text{boundary}} - V_{\min}$ is much larger than the temperature T and the zero-point energy of the system. This is because the bounded region is then a well-defined potential well, and, neglecting quantum tunnelling, the system in equilibrium will have such a small nuclear density in the region beyond the boundary that this can be neglected.

For systems which do not have a well-defined potential well, the elastic approximation is a bad one. This includes all fluids, and also includes solid systems with mobile nuclei, such as host-guest structures with mobile ions [64] and some quasiperiodic structures [65]. Such systems should be modelled using other methods. For reasons of practicality, this limitation also applies to systems where the potential well contains a large region of phase space, for example systems containing large flexible molecules which are free to contort.

Even for systems where the elastic approximation is reasonable, it is never a perfect approximation. The PES is modelled as a potential well around a single minimum, where in reality by translational symmetry the PES must consist of an infinite lattice of identical wells about identical minima. Nevertheless, this approximation is sufficient to calculate many properties of crystals to high accuracy [45].

2.1.5 The perfectly periodic approximation

In addition to modelling crystals as perfectly elastic, this work will also model crystals as perfectly periodic. This means each crystal is composed of infinite copies of a fixed motif of atoms, tiled in a regularly repeating pattern throughout space.

Each atom in the motif is labelled with an index j , running from 1 to n , where n is the number of atoms in

the motif. Each atom is defined by its atomic species, z_j , and its static-lattice co-ordinate, \mathbf{r}_j^0 .

The position of each atom is described by a 3-dimensional co-ordinate \mathbf{r}_j , and the vector of all such co-ordinates is the nuclear co-ordinate \mathbf{r} . As a consequence of approximation 2, each co-ordinate \mathbf{r}_j is confined to a bounded region of 3-dimensional cartesian space, and that bounded region contains \mathbf{r}_j^0 .

In three dimensions, the tiling is defined by three vectors, $\widehat{\mathbf{R}}_1$, $\widehat{\mathbf{R}}_2$ and $\widehat{\mathbf{R}}_3$, collectively known as the *primitive lattice vectors* of the crystal. These vectors are linearly independent, and form a basis for the *lattice vectors*, or *R-vectors*, of the crystal, defined as

Definition 1 (Lattice vectors):

The lattice vectors \mathbf{R} are linear combinations of the primitive lattice vectors, $\mathbf{R} = R_1\widehat{\mathbf{R}}_1 + R_2\widehat{\mathbf{R}}_2 + R_3\widehat{\mathbf{R}}_3$, with integer coefficients $\{R_i\}$.

It is useful to define the *crystal lattice matrix* L , which is a real-valued 3×3 matrix whose rows are the primitive lattice vectors of the crystal. This matrix is useful for defining various properties of the crystal.

Given the definition of the lattice vectors, it is natural to define the periodic approximation as

Approximation 3 (Perfect periodicity):

For every lattice vector \mathbf{R} and every atom j with static-lattice position \mathbf{r}_j^0 there is an atom, k , of the same species, $z_k = z_j$, with the static-lattice position $\mathbf{r}_k^0 = \mathbf{r}_j^0 + \mathbf{R}$.

The periodic approximation is also always wrong, since physical crystals all have finite extent. However, the differences between the properties of finite physical crystals and infinite theoretical crystals, known as *finite size effects*, are usually negligible for macroscopic crystals [66], and calculating finite size effects is often much harder than calculating properties under the periodic approximation.

Primitive cells

Starting at an arbitrary origin, the R-vectors describe a lattice of points throughout space. These points can be used to divide all of space into smaller volumes, each of which contains a single motif of atoms. These volumes of space are called *primitive cells*, defined as

Definition 2 (Primitive cells):

The primitive cell with label \mathbf{R} is the space $\mathbf{r} = \mathbf{R} + r_1\widehat{\mathbf{R}}_1 + r_2\widehat{\mathbf{R}}_2 + r_3\widehat{\mathbf{R}}_3$ with coefficients satisfying $0 \leq r_i < 1$.

Each primitive cell is a parallelepiped containing n nuclei, which form a single motif. The volume of a

primitive cell can be calculated as the determinant of the lattice matrix, $|L|$. The R-vectors and unit cells in a section of a crystal are depicted in figure 2.1.

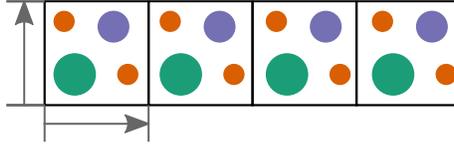


Figure 2.1: A section of a crystal, showing the unit cells as black boxes and the primitive R-vectors as grey arrows.

2.2 Periodic co-ordinates

2.2.1 Labelling directions

An arbitrary atom is defined by its index within the motif, j , and the R-vector \mathbf{R} of the primitive cell which contains the static lattice position \mathbf{r}_j^0 . Variables associated with a given atom will be subscripted with these labels. For example, the nuclear position \mathbf{r} of the atom at R-vector \mathbf{R} with index j is written as the three-dimensional vector $\mathbf{r}_{\mathbf{R},j}$.

When discussing vector quantities, it is convenient to also include the cartesian direction α [67] along with the atomic labels \mathbf{R} and j , so that the labelled quantity becomes a scalar. For example, the nuclear co-ordinate of atom j at R-vector \mathbf{R} in the α direction is written $r_{\mathbf{R},j,\alpha}$. To simplify the notation, these three quantities; atomic label, R-vector and cartesian direction; will be bundled into a single label vector, $\rho = (j, \mathbf{R}, \alpha)$ [68].

For convenience, this label will be used even when only some of its components are needed. For example, the species of an atom might be labelled z_ρ , even though the labels \mathbf{R} and α are superfluous. In addition, it can be useful to label the components of a given label, where for example the R-vector \mathbf{R}_ρ is the R-vector in the label ρ . To avoid cluttering equations, this label will regularly be implied, for example in the sum $\sum_\rho e^{2\pi i \mathbf{R} \cdot \mathbf{q}}$ it should be assumed that $\mathbf{R} = \mathbf{R}_\rho$.

2.2.2 Mass-weighted displacement co-ordinates

Rather than describing the nuclei using the origin-dependent co-ordinate \mathbf{r} , it is convenient to define a co-ordinate system relative to the static-lattice co-ordinate \mathbf{r}^0 .

For reasons which will be detailed in section 2.2.10, it is also convenient to weight this co-ordinate system by the atomic masses, to give

Definition 3 (Mass-weighted displacement co-ordinates):

The mass-weighted co-ordinate $d_\rho = \sqrt{m_\rho}(r_\rho - r_\rho^0)$, where m_ρ is the mass of the atom with label ρ .

Each displacement d_ρ has a corresponding unit vector $\hat{\mathbf{d}}_\rho$. The cartesian co-ordinate is related to the displacement co-ordinate as

$$\mathbf{r} = \mathbf{r}^0 + \mathbf{d} = \mathbf{r}^0 + \sum_{\rho} d_\rho \hat{\mathbf{d}}_\rho. \quad (2.9)$$

The displacement co-ordinate for a single atom is shown in figure 2.2.

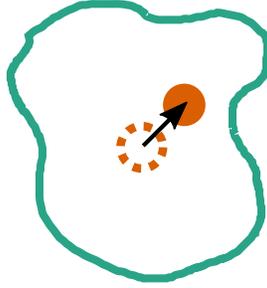


Figure 2.2: An atom moving in a bounded potential well. The dashed outline represents the static-lattice position of the atom, \mathbf{r}_ρ^0 . The solid circle represents the position of the atom itself, \mathbf{r}_ρ . The black arrow shows the displacement vector of the atom, $\mathbf{r}_\rho - \mathbf{r}_\rho^0$. The turquoise curve represents the boundary contour $V(\mathbf{r}_\rho) = V_{\text{boundary}}$.

2.2.3 Plane-wave co-ordinates

A crystal can be described in terms of the motions of single nuclei using the displacement co-ordinates d_ρ , but for many purposes it can be more convenient to describe the crystal in terms of the collective motions of many nuclei at once. A particularly natural description, which takes advantage of the periodicity of the crystal, is that of *periodic plane-wave co-ordinates*.

There are two possible definitions of plane-wave co-ordinates. The definition adopted here is that of *complex plane-wave co-ordinates*, defined such that the atomic displacements due to motion along a single periodic co-ordinate are described by

$$d_\rho \propto e^{-2\pi i \mathbf{q} \cdot \mathbf{R}_\rho}, \quad (2.10)$$

where \mathbf{q} is a vector which labels the periodicity of the periodic co-ordinate, the *wavevector* of the plane wave.

The other definition of periodic co-ordinates, which will not be used here, is that of *real plane-wave co-ordinates*, which replace $e^{\pm 2\pi i \mathbf{q} \cdot \mathbf{R}_\rho}$ with $\sin(2\pi \mathbf{q} \cdot \mathbf{R}_\rho)$ and $\cos(2\pi \mathbf{q} \cdot \mathbf{R}_\rho)$. The two choices of co-ordinates have different benefits in different situations. Complex plane-wave co-ordinates will be used here because they interact more naturally with symmetry [13] and size consistency [14], as detailed in sections 4.1 and 4.2

respectively.

In order to form a complete basis for a system containing N primitive cells, it is necessary to define a plane-wave basis containing N distinct wavevectors. This set of wavevectors are the q -points of the system. The choice of wavevectors for a given system will be discussed further in section 2.3.

2.2.4 Normal-mode co-ordinates

In a 3-dimensional system containing n atoms in the primitive cell, there are $3n$ plane waves at each q -point. These waves are labelled by an integer k , known either as the *mode index* or as the *phonon band index*. Just as the \mathbf{R} , j and α labels for real-space displacements can be combined into a single label ρ , the \mathbf{q} and k labels for plane-wave displacements can be combined into a single label χ .

Each plane-wave can be defined in terms of a directional unit vector $\hat{\mathbf{u}}_\chi$ and the corresponding displacement u_χ . These are typically chosen to diagonalise some harmonic potential, as described in section 3.2.1. As such, the u_χ co-ordinates describe the normal modes of the harmonic potential, and are named as such.

Definition 4 (Normal-mode co-ordinates):

The normal-mode co-ordinates u_χ are defined as $u_\chi = \frac{1}{N} \sum_\rho U_{j\alpha\mathbf{q}k}^* e^{2\pi i \mathbf{q} \cdot \mathbf{R}} d_\rho$, where U is a unitary matrix.

The reverse transformation is $d_\rho = \sum_\chi U_{j\alpha\mathbf{q}k} e^{-2\pi i \mathbf{q} \cdot \mathbf{R}} u_\chi$, and the matrix U satisfies $\sum_{j\alpha} U_{j\alpha\mathbf{q}k} U_{j\alpha\mathbf{q}k'} = \delta_{kk'}$ and $\sum_k U_{j\alpha\mathbf{q}k} U_{j'\alpha'\mathbf{q}k} = \delta_{jj'} \delta_{\alpha\alpha'}$.

This choice of normalisation means that displacements in normal-mode co-ordinates are independent of the number of primitive cells being simulated, N . If normal-mode co-ordinates were otherwise normalised then the values u_χ would change as the system size changed. That displacements are invariant to changes of N does not imply that all functions of $\{u_\chi\}$ are likewise invariant. This will be discussed in detail in section 4.2.3.

Just as with real-space displacements, the collective nuclear co-ordinates of the crystal can be uniquely defined in normal-mode co-ordinates as

$$\mathbf{r} = \mathbf{r}^0 + \mathbf{u} = \mathbf{r}^0 + \sum_\chi u_\chi \hat{\mathbf{u}}_\chi. \quad (2.11)$$

2.2.5 Reciprocal space

Crystals are periodic in real space. One consequence of this is that if all displacements $d_{\mathbf{r},j,\alpha}$ are transformed to $d_{\mathbf{r}+\mathbf{x},j,\alpha}$ for some fixed vector \mathbf{x} , then the crystal is invariant if \mathbf{x} is an R-vector of the crystal.

Crystals are also periodic in the space of plane waves, known as *reciprocal space* [12]. If all displacements $u_{\mathbf{q},k}$ are transformed to $u_{\mathbf{q}+\mathbf{x},k}$ for some fixed vector \mathbf{x} , there exists a set of values of \mathbf{x} for which the crystal is again invariant. These values of \mathbf{x} are labelled \mathbf{G} , and are known as the *reciprocal lattice vectors*, or *G-vectors*, of the crystal.

Just as all R-vectors can be described as sums of primitive R-vectors with integer coefficients, $\mathbf{R} = R_1\hat{\mathbf{R}}_1 + R_2\hat{\mathbf{R}}_2 + R_3\hat{\mathbf{R}}_3$, so too can G-vectors be described as sums of *primitive reciprocal lattice vectors* with integer coefficients,

Definition 5 (Reciprocal lattice vectors):

The reciprocal lattice vectors \mathbf{G} are linear combinations of the primitive reciprocal lattice vectors, $\mathbf{G} = G_1\hat{\mathbf{G}}_1 + G_2\hat{\mathbf{G}}_2 + G_3\hat{\mathbf{G}}_3$, with integer coefficients $\{G_i\}$.

Where the primitive R-vectors are the rows of the crystal lattice matrix L , the primitive G-vectors are the rows of the matrix L^{-T} , known as the *reciprocal lattice matrix* of the crystal. The primitive R- and G-vectors satisfy the relation $\hat{\mathbf{G}}_i \cdot \hat{\mathbf{R}}_j = \delta_{i,j}$.

2.2.6 Fractional co-ordinates

Real space co-ordinates can either be expressed in cartesian co-ordinates defined in terms of fixed unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$, or in *fractional co-ordinates* defined in terms of the primitive lattice vectors $\hat{\mathbf{R}}_1$, $\hat{\mathbf{R}}_2$ and $\hat{\mathbf{R}}_3$. Equally, co-ordinates in reciprocal space can either be expressed reciprocal co-ordinates in terms of fixed reciprocal unit vectors, or in *fractional reciprocal co-ordinates* in terms of the primitive reciprocal lattice vectors $\hat{\mathbf{G}}_1$, $\hat{\mathbf{G}}_2$ and $\hat{\mathbf{G}}_3$.

These two co-ordinate systems can be transformed between using the crystal lattice matrix, L . If \mathbf{r}_c is a cartesian co-ordinate and \mathbf{r}_f is the fractional co-ordinate corresponding to the same location then

$$\mathbf{r}_c = L^T \mathbf{r}_f. \quad (2.12)$$

Equivalently, in reciprocal space if a \mathbf{q}_r is a q-point in reciprocal space and \mathbf{q}_f is the same q-point in fractional reciprocal co-ordinates then

$$\mathbf{q}_r = L^{-1} \mathbf{q}_f. \quad (2.13)$$

2.2.7 The reciprocal primitive cell

The invariance of the crystal to reciprocal-space translations under arbitrary \mathbf{G} -vectors leads to a redundancy in the definition of a \mathbf{q} -point. Namely, displacing the crystal by the plane wave $u_{\mathbf{q},k}$ is equivalent to displacing the crystal by the plane wave $u_{\mathbf{q}+\mathbf{G},k}$ for any \mathbf{G} -vector \mathbf{G} .

As such, all \mathbf{q} -points which differ only by a \mathbf{G} -vector are exactly equivalent, and it is only necessary to consider one \mathbf{q} -point from each equivalent set. There are many ways to select a set of \mathbf{q} -points such that every \mathbf{q} -point is accounted for but without selecting two or more equivalent \mathbf{q} -points. A natural choice for this is the set of \mathbf{q} -points within a region of reciprocal space called the *reciprocal primitive cell* [69]

Definition 6 (The reciprocal primitive cell):

The reciprocal primitive cell is the set of \mathbf{q} -points whose coefficients in fractional reciprocal co-ordinates satisfy $-\frac{1}{2} < q \leq \frac{1}{2}$.

Just as the primitive cell forms a parallelepiped in cartesian space, the reciprocal primitive cell forms a parallelepiped in reciprocal space. A primitive cell and a reciprocal primitive cell are depicted in figure 2.3.



Figure 2.3: A primitive cell and a reciprocal primitive cell.

Left: a primitive cell in fractional co-ordinates, outlined in grey. The black dots represent \mathbf{R} -vectors.

Right: a reciprocal primitive cell in fractional reciprocal co-ordinates, shown in grey. The black dots represent \mathbf{G} -vectors, and the turquoise dots highlight the \mathbf{q} -points which are their own pairs.

2.2.8 Paired \mathbf{q} -points

The nuclear co-ordinate of a crystal whose atoms have been displaced along a single plane wave co-ordinate is in general complex. This is of course unphysical, since atomic co-ordinates can only take real values.

In order to recover real-valued co-ordinates, it is first necessary to define, for each \mathbf{q} -point \mathbf{q} , the \mathbf{q} -point $\bar{\mathbf{q}}$ such that $\mathbf{q} + \bar{\mathbf{q}}$ is a \mathbf{G} -vector. This is known as the *paired \mathbf{q} -point* of \mathbf{q} . Most \mathbf{q} -points have a distinct paired \mathbf{q} -point, but there are eight \mathbf{q} -points for which $\bar{\mathbf{q}} = \mathbf{q}$. The \mathbf{q} -points which are paired with themselves are those whose coefficients in fractional co-ordinates are all 0 or $\frac{1}{2}$, i.e. the eight vertices of the positive octant of the reciprocal primitive cell. Four of these \mathbf{q} -points are depicted in figure 2.3, and the other four \mathbf{q} -points lie out of the plane of that figure.

Once paired q-points have been defined, the mode transformations $U_{j\alpha\mathbf{q}k}$ can then be constrained such that $U_{j\alpha\bar{\mathbf{q}}k} = U_{j\alpha\mathbf{q}k}^*$. Then the k th mode at \mathbf{q} is the *paired mode* to the k th mode at $\bar{\mathbf{q}}$. The paired mode to mode χ is labelled $\bar{\chi}$. Modes at q-points for which $\bar{\mathbf{q}} = \mathbf{q}$ are their own pairs, $\bar{\chi} = \chi$.

Real-valued co-ordinates can be recovered by insisting that for every displacement u_χ there is a matching displacement along the paired mode of $u_{\bar{\chi}} = u_\chi^*$. This has consequences for modes which are their own pairs, since displacements along such modes must then be real.

2.2.9 Transforming between displacement and plane-wave co-ordinates

With the definition of paired modes, the transformation between real-space displacements and plane-wave displacements can be simplified. Defining the transformation matrix

$$U_{\rho\chi} = U_{j\alpha\mathbf{q}k} e^{-2\pi i\mathbf{q}\cdot\mathbf{R}}, \quad (2.14)$$

which obeys $U_{\rho\bar{\chi}} = U_{\rho\chi}^*$, $\sum_\chi U_{\rho\chi}U_{\rho'\bar{\chi}} = N\delta_{\rho\rho'}$ and $\sum_\rho U_{\rho\chi}U_{\rho\chi'} = N\delta_{\chi\bar{\chi}'}$, the co-ordinate transformations simplify to

$$d_\rho = \sum_\chi U_{\rho\chi}u_\chi \quad (2.15)$$

and

$$u_\chi = \frac{1}{N} \sum_\rho U_{\rho\bar{\chi}}d_\rho. \quad (2.16)$$

2.2.10 Working with normal-mode co-ordinates

When working with normal-mode co-ordinates, it is necessary to express physical quantities in terms of these co-ordinates. For simple functions of the co-ordinates $f(\mathbf{r})$ which do not depend on the derivatives of \mathbf{r} , this is a simple case of transforming from \mathbf{r} co-ordinates to \mathbf{d} co-ordinates and then to \mathbf{u} co-ordinates.

For quantities which depend on derivatives, in particular the kinetic energy, this is a little more complex. Additionally, normal-mode co-ordinates are not a unitary transformation of cartesian co-ordinates, and so a non-trivial Jacobian must be included when performing integration in these co-ordinates.

The kinetic energy

In cartesian co-ordinates, the kinetic energy operator is given by [26]

$$\mathcal{T} = \sum_{\rho} -\frac{1}{2m_{\rho}} \frac{\partial^2}{\partial r_{\rho}^2}, \quad (2.17)$$

where m_{ρ} is the mass of the atom indexed by ρ .

Transforming into mass-reduced co-ordinates, the mass dependence of the kinetic energy is absorbed by the co-ordinate system, and so the kinetic energy becomes

$$\mathcal{T} = \sum_{\rho} -\frac{1}{2} \frac{\partial^2}{\partial d_{\rho}^2}. \quad (2.18)$$

This is the reason for mass-weighting the co-ordinates: the mass dependence of the kinetic energy is automatically accounted for.

Transforming into normal-mode co-ordinates, the kinetic energy becomes

$$\mathcal{T} = \sum_{\substack{\chi \\ \bar{\chi}=\chi}} -\frac{1}{2N} \frac{\partial^2}{\partial u_{\chi}^2} + \sum_{\substack{\chi \\ \bar{\chi} \neq \chi}} -\frac{1}{2N} \frac{\partial^2}{\partial u_{\chi} \partial u_{\bar{\chi}}}, \quad (2.19)$$

where N is the number of primitive cells in the simulation.

In each term in the kinetic energy where a mode u_{χ} appears, its paired mode $u_{\bar{\chi}}$ also appears, and the two cannot be separated. As a result, whenever a mode is considered its pair must also be considered and the two cannot truly be regarded as independent. If real plane-wave co-ordinates had been used instead of complex plane-wave co-ordinates then the kinetic energy would separate into a sum of single-mode terms. However, this would not imply that the Hamiltonian was separable, as in general the potential energy is not separable in either co-ordinate system. This will be detailed further in section 4.1.

Integration and the Jacobian

When performing integration in cartesian co-ordinates \mathbf{r} , no Jacobian is needed.

Transforming to mass-weighted displacement co-ordinates \mathbf{d} , the mass-weighting introduces a Jacobian factor of

$$J = \left| \frac{\partial \mathbf{r}}{\partial \mathbf{d}} \right| = \prod_{\rho} \frac{1}{\sqrt{m_{\rho}}}, \quad (2.20)$$

where m_{ρ} is the mass of the atom with label ρ .

The transformation from displacement co-ordinates to normal-mode co-ordinates introduces a factor of

$$J = \left| \frac{\partial \mathbf{d}}{\partial \mathbf{u}} \right| = \prod_{\rho} \sqrt{N}, \quad (2.21)$$

where N is the number of primitive cells being simulated. It should be noted that this does not depend on the unit vectors of the normal modes, as every choice of unit vectors is a unitary transformation of any other choice.

Combining the two transformations, the Jacobian for any integration in normal-mode co-ordinates is

$$J = \prod_{\rho} \sqrt{\frac{N}{m_{\rho}}}. \quad (2.22)$$

2.3 Supercells

When modelling crystals it is necessary to consider configurations in which the nuclei are in different positions in different primitive cells. In order to simulate such configurations, it is necessary to construct collections of multiple primitive cells called *supercells*.

Each supercell is defined by the supercell matrix S , whose elements are integers. A supercell has a lattice matrix L_S , defined in terms of the primitive cell lattice matrix L as $L_S = S \cdot L$. The rows of this matrix are the supercell lattice vectors. Fractional supercell co-ordinates can be defined analogously to primitive fractional co-ordinates, as the coefficients of a vector whose basis vectors are the *supercell lattice vectors*.

Definition 7 (Supercells):

A supercell is the region of space with co-ordinates in fractional supercell co-ordinates with coefficients r satisfying $0 \leq r < 1$.

Just like the primitive cells, each supercell describes a parallelepiped in cartesian space.

Just as the primitive lattice vectors can be used to define the R-vectors of the primitive cell, the supercell lattice vectors can be used to define the set of supercell R-vectors, which are linear combinations of the supercell lattice vectors with integer coefficients. Every supercell R-vector is also an R-vector of the primitive cell.

When performing calculations using a given supercell, each point in space within the supercell is treated identically to an infinite set of points called the *images* of that point,

Definition 8 (Supercell images):

The images of a point \mathbf{r}_1 are those points \mathbf{r}_2 satisfying $\mathbf{r}_2 - \mathbf{r}_1 = \mathbf{R}$ for some supercell \mathbf{R} -vector \mathbf{R} .

The images of a given point are defined with respect to a given supercell. When no supercell is specified, it should be assumed that the images of a given point refer to the images with respect to the primitive cell.

Each supercell contains a number of \mathbf{R} -vectors equal to the *size of the supercell*, N , which can be calculated as the determinant of the supercell matrix, $N = |S|$. The volume of the supercell is exactly N times the volume of the primitive cell, and the supercell contains exactly N images of every point in the primitive cell.

2.3.1 Reciprocal subcells

The reciprocal lattice of each supercell is a subcell of the reciprocal primitive cell. Where the supercell is bounded by the supercell lattice vectors, the subcell is bounded by the *subcell reciprocal lattice vectors*, which are the rows of the supercell reciprocal lattice matrix $L_S^{-T} = S^{-T}L^{-T}$.

While the supercell is N times larger than the primitive cell, the subcell is N times smaller than the reciprocal primitive cell. For every point in reciprocal space within the subcell, the reciprocal primitive cell contains N images of that point with respect to the subcell, where subcell images in reciprocal space are defined analogously to supercell images in real space.

Since all of the images of a given supercell are treated identically, it is only possible to simulate the crystal at a given displacement u_χ using a given supercell if the q -point \mathbf{q}_χ takes one of a finite number of values, known as the q -points which are commensurate with the supercell [70],

Definition 9 (Commensurate q -points):

A q -point \mathbf{q} is commensurate with a supercell S if \mathbf{q} can be written as a linear combination of the subcell reciprocal lattice vectors with integer coefficients.

For each supercell, there are N q -points within the reciprocal primitive cell which are commensurate with that supercell. This always includes the Γ -point, $\mathbf{q} = 0$.

When simulating a displacement $\mathbf{u} = \sum_\chi u_\chi \hat{\mathbf{u}}_\chi$, the supercell used must be commensurate with the entire set of q -points \mathbf{q}_χ corresponding to non-zero displacements u_χ [70]. Figure 2.4 shows two supercells and their corresponding Brillouin zone subcells.

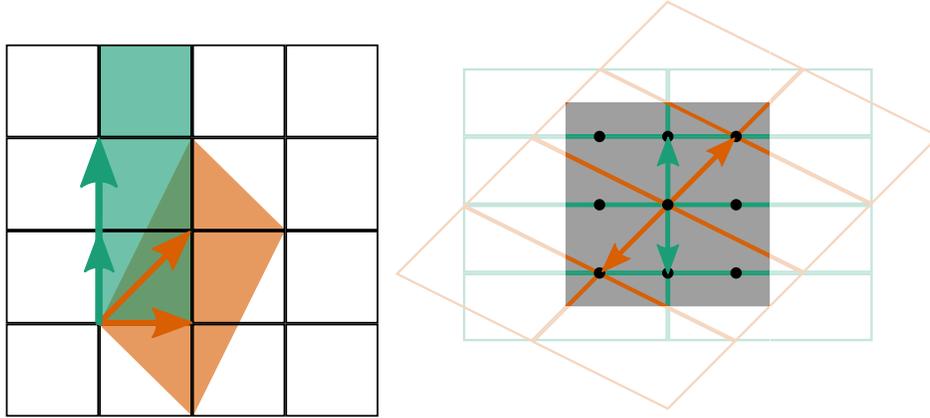


Figure 2.4: Supercells and subcells.

Left: Real space. Primitive unit cells are shown as black squares, and the supercells with supercell matrices $\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$ are shown as teal and orange parallelograms respectively. The distinct R-vectors within the supercells are shown as coloured arrows.

Right: Reciprocal space. The irreducible Brillouin zone is shown as a grey square, with the q-points which are commensurate with a supercell with supercell matrix $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$ are shown as black circles. The Brillouin zone subcells corresponding to the supercells as the left figure are shown, with coloured arrows showing the q-points which are commensurate with these supercells.

2.3.2 Wigner-Seitz cells

While a given primitive cell or supercell is useful definition of a particular set of points in real space, the choice of a parallelepipedal volume is for mathematical convenience rather than for physical meaning. In order to run calculations using supercells, it is important to define another kind of cell, the *Wigner-Seitz cell*.

A Wigner-Seitz cell is defined with reference to a given supercell and a given origin in cartesian co-ordinates.

Definition 10 (Wigner-Seitz cells):

The Wigner-Seitz cell contains exactly one image of every point in the reference cell, which is chosen as the image which is closest in cartesian co-ordinates to the reference origin.

For points which lie on the boundary of the Wigner-Seitz cell, such that multiple images are equally close to the origin, the choice of which images are included in the cell will depend on how the Wigner-Seitz cell is being used. In calculations which involve summing or integrating across the Wigner-Seitz cell, the contribution from each point on the boundary of the cell is typically taken as the average contribution from all images of that point which lie on the boundary. A Wigner-Seitz cell is depicted in figure 2.5.

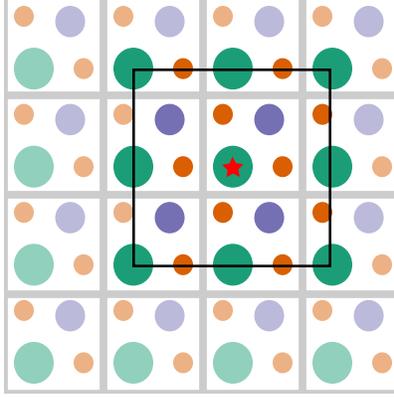


Figure 2.5: The black box depicts the Wigner-Seitz cell for the supercell with supercell matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, centred on the starred atom. The atoms which lie outside the Wigner-Seitz cell are partially transparent. Each grey box is a single primitive unit cell.

2.3.3 Born-von Karman boundary conditions

It is not possible to model the entirety of an extended system, as this would require the simulation of an intractable number of atoms [13]. Instead, such systems are modelled by exploiting the fact that inter-atomic interactions weaken with increasing distance [71]. This means that pairs of atoms which are sufficiently separated from one another can be approximated as non-interacting, which in turn means that simulating the properties of a finite section of the system will give an approximation of the properties of the entire system. As the size of the finite section is increased this approximation will improve, tending asymptotically to being correct in the limit of an infinitely large supercell [72].

The natural choice of finite sections in a periodic crystal are the supercells. A supercell can be simulated in isolation using *Born-von Karman periodic boundary conditions* [12, 72],

$$|\psi(\mathbf{r} + \mathbf{R})\rangle = |\psi(\mathbf{r})\rangle, \quad (2.23)$$

which must hold for every supercell \mathbf{R} -vector \mathbf{R} . Using periodic boundary conditions ensures that the local environment of every atom in the supercell is the same as it would be in the extended crystal.

A simulation of a supercell using periodic boundary conditions can be converted to a simulation of the whole system by approximating the inter-atomic interactions as finite ranged [13],

Approximation 4 (Finite ranged interactions):

Each atom only interacts with other atoms in the Wigner-Seitz supercell centred on that atom. For atoms on the boundary of the supercell, the interaction is divided equally between all images of that atom on the boundary of the supercell.

It should be noted that simulations using periodic boundary conditions are only appropriate for systems which are electrically neutral and which have no net electric or magnetic dipole moment. For systems which are not electrically neutral, the long-ranged nature of the Coulomb interactions means that the size of the finite section which must be simulated in order to accurately approximate the full system is too large for calculation to be tractable. For systems with a finite dipole moment, the long-ranged nature of the dipole-dipole interaction is again a problem, although less of a problem than for the Coulomb interaction, and in addition the periodicity of the system is broken. For some materials, the effect of this is small enough to be treated as a perturbative correction to a periodic simulation [27, 68].

2.3.4 Hierarchical supercell simulations

In order to simulate the properties of a crystal in a given Born-von Karman supercell, it is not always necessary to simulate the full supercell explicitly. Rather, the properties of the full supercell can often be re-constructed from the simulation of multiple smaller supercells [26]. Methods which take advantage of this are known as *hierarchical supercell methods* [13]. Such methods will be used extensively in this work.

The advantage of simulating a large supercell implicitly via many explicit simulations of smaller supercells is that electronic structure calculations have a computational cost which increases with the number of electrons in the simulation, and in some cases this cost scaling can be very unfavourable [20]. As such, the increased cost of running multiple calculations is often more than compensated for by the lower cost of running smaller calculations.

2.4 Symmetry

One of the most fundamental properties of a crystal is its symmetry. Every crystal is invariant under at least translational symmetry, and many crystals are also invariant under a variety of symmetries, consisting of combinations of rotations, reflections and translations.

2.4.1 Crystal symmetries

Crystal symmetries are operators, \mathcal{S} , defined in terms of a unitary rank 2 tensor S and a vector \mathbf{s} , which act in 3-dimensional space as [73]

$$\mathcal{S}\mathbf{r} = S \cdot \mathbf{r} + \mathbf{s}. \tag{2.24}$$

The tensor and vector are referred to as the *rotation* and *translation* of the symmetry respectively, although it is worth noting that the tensor may describe an improper rotation (a rotation followed by an inversion). The crystal symmetries of a given crystal are those symmetries which map the static lattice co-ordinate \mathbf{r}_ρ^0 of every atom onto the static lattice co-ordinate of an atom of the same species,

$$\mathcal{S}\mathbf{r}_\rho^0 = \mathbf{r}_{\rho'}^0 \text{ s.t. } z_\rho = z_{\rho'} \quad \forall \rho. \quad (2.25)$$

In cartesian co-ordinates, \mathcal{S} is a unitary tensor. In fractional co-ordinates, \mathcal{S} is a tensor with integer elements. The set of rotations of all of the crystal symmetries forms a finite group. One consequence of this is that for every crystal symmetry \mathcal{S} , the *inverse* symmetry $\bar{\mathcal{S}} = \mathcal{S}^{-1}$ is also a crystal symmetry.

Crystal symmetries are relevant because the Hamiltonian of a given crystal is *invariant* under every symmetry of that crystal, i.e. $[\mathcal{H}, \mathcal{S}] = 0$. Among other consequences, this means that if the symmetry operator acts on a crystal in thermal equilibrium, then all observables are unchanged.

Only a subset of the crystal symmetries have reasonable representations in displacement co-ordinates. In general, a symmetry operator which acts on a crystal whose atoms are all close to their static lattice positions will not produce a crystal whose atoms are close to those same static lattice positions. Rather, each atom will now be close to the static lattice position of some other atom of the same species.

This work will only consider symmetries which preserve the absolute displacement of the nuclei from the static-lattice configuration, $|\mathbf{u}|$. While all other symmetries are neglected, their potential future applications will be discussed in section 8.1.

2.4.2 Re-labelling atoms

In addition to being invariant under crystal symmetries, crystals are also invariant under the *re-labelling* of their atoms, provided each atom is re-labelled as an atom of the same species [68]. For example, if the atoms with labels ρ and ρ' are of the same species, then the crystal with $\mathbf{r}_\rho = \mathbf{r}_1$ and $\mathbf{r}_{\rho'} = \mathbf{r}_2$ is identical to the crystal with $\mathbf{r}_\rho = \mathbf{r}_2$ and $\mathbf{r}_{\rho'} = \mathbf{r}_1$. This is of course a mathematical operation rather than a physical one.

This re-labelling can be combined with a symmetry operator which does not preserve $|\mathbf{u}|$ in order to construct an operator which does preserve $|\mathbf{u}|$. These combined operators first act as a crystal symmetry, before re-labelling the atoms in such a way that the combined operator maps each static-lattice position onto itself. The action of these combined operators is shown in figure 2.6.

In order to work with such operators, it is necessary to keep track of the re-labelling of atoms associated with each symmetry. For this, it is necessary to briefly drop the ρ notation and return to (\mathbf{R}, j) notation for

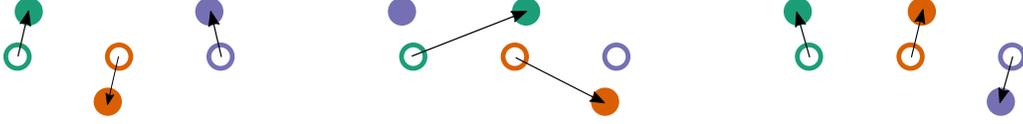


Figure 2.6: A symmetry which translates each atom to the right. The solid circles represent atomic positions, and the hollow circles represent static-lattice positions. The black arrows are the displacement vectors between the static-lattice position and the current position of each atom.

Left: atoms before the action of the symmetry. Each atom is at a small displacement from its static-lattice position.

Centre: atoms after the action of the symmetry but before re-labelling. The atoms are now at large displacements from their static-lattice positions.

Right: atoms after the action of the symmetry and re-labelling. The atoms have been returned to small displacements from their static-lattice positions.

atomic labelling.

Prior to re-labelling, symmetry \mathcal{S} maps atom (\mathbf{R}', j') onto atom (\mathbf{R}, j) . It is convenient to define $(\mathbf{R}', j') \equiv \bar{\mathcal{S}}(\mathbf{R}, j)$.

The mapping of atom indices j is independent of \mathbf{R} , and so it is further convenient to define $j' \equiv \bar{\mathcal{S}}j$.

The R-vector \mathbf{R}' is simply the vector between the static-lattice co-ordinate $\mathcal{S}\mathbf{r}_{\mathbf{R},j}^0 \equiv \mathbf{r}_{\bar{\mathcal{S}}(\mathbf{R},j)}^0$ and the static-lattice co-ordinate of the image of that atom with $\mathbf{R} = 0$, $\mathbf{r}_{(0,\bar{\mathcal{S}}j)}^0$,

$$\mathbf{r}_{\bar{\mathcal{S}}(\mathbf{R},j)}^0 = \mathbf{R}' + \mathbf{r}_{(0,\bar{\mathcal{S}}j)}^0. \quad (2.26)$$

By definition,

$$\mathbf{r}_{(\mathbf{R},j)}^0 = \mathcal{S}\mathbf{r}_{\bar{\mathcal{S}}(\mathbf{R},j)}^0 = S \cdot \mathbf{r}_{\bar{\mathcal{S}}(\mathbf{R},j)}^0 + \mathbf{s}, \quad (2.27)$$

and combining the two equations gives

$$\mathbf{r}_{(\mathbf{R},j)}^0 = S \cdot \mathbf{R}' + S \cdot \mathbf{r}_{(0,\bar{\mathcal{S}}j)}^0 + \mathbf{s}. \quad (2.28)$$

It is convenient to denote the quantity $\mathbf{R}_j = \mathbf{r}_{\mathcal{S}(0,\bar{\mathcal{S}}j)}^0 - \mathbf{r}_{0,j}^0$. This is the R-vector from atom $(0, j)$ to the image of that atom which is reached by acting with symmetry \mathcal{S} on atom $(0, j')$.

With this definition, noting that $\mathbf{r}_{(\mathbf{R},j)}^0 = \mathbf{r}_{(0,j)}^0 + \mathbf{R}$, the R-vector \mathbf{R}' can conveniently be written as $\mathbf{R}' = S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j)$. This vector is shown in figure 2.7.

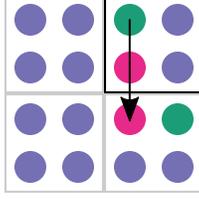


Figure 2.7: The R-vector \mathbf{R}_j for a clockwise rotation of 90° about the origin.

The origin $\mathbf{r} = 0$ is at the centre of the figure, and the primitive unit cell at $\mathbf{R} = 0$ is the top-right cell, outlined in black. The coloured circles represent atoms, all of which have the same species regardless of colour. The atoms $(0, j)$ and $\mathcal{S}(0, j)$ are shown in turquoise, in the top-right and bottom-right cells respectively. Atoms $(0, \bar{\mathcal{S}}j)$ and $\mathcal{S}(0, \bar{\mathcal{S}}j)$ are shown in pink, again in the top-right and bottom-right cells respectively. The R-vector \mathbf{R}_j is shown as the black arrow.

2.4.3 Symmetries in normal-mode co-ordinates

When combined with the appropriate re-labelling, each symmetry acts in cartesian co-ordinates as

$$\mathcal{S}\mathbf{r}_{\mathbf{R},j} = S \cdot \mathbf{r}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)} + \mathbf{s}. \quad (2.29)$$

Expanding the cartesian co-ordinates into displacement co-ordinates gives

$$\mathcal{S}(\mathbf{r}_{\mathbf{R},j}^0 + \mathbf{d}_{\mathbf{R},j}) = S \cdot (\mathbf{r}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)}^0 + \mathbf{d}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)}) + \mathbf{s}. \quad (2.30)$$

Noting that every combined symmetry leaves the static-lattice co-ordinate invariant, $\mathcal{S}\mathbf{r}_{(\mathbf{R},j)}^0 = \mathbf{r}_{(\mathbf{R},j)}^0$, this can be simplified to give

$$\mathcal{S}\mathbf{d}_{\mathbf{R},j} = S \cdot \mathbf{d}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)}. \quad (2.31)$$

This is the reason for combining symmetries with re-labelling; rather than acting on atoms, the combined symmetries can be thought of as acting directly on atomic displacements. Under the action of the symmetry, the displacement of each atom is simply replaced with the displacement of a different atom, transformed by S .

Moving to periodic co-ordinates,

$$\begin{aligned} \mathcal{S}\mathbf{x}_{\mathbf{q},j} &= \frac{1}{N} \sum_{\mathbf{R}} e^{2\pi i \mathbf{q} \cdot \mathbf{R}} \mathcal{S}\mathbf{d}_{\mathbf{R},j} \\ &= \frac{1}{N} \sum_{\mathbf{R}} e^{2\pi i \mathbf{q} \cdot \mathbf{R}} S \cdot \mathbf{d}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)} \\ &= \frac{1}{N} \sum_{\mathbf{R}} e^{2\pi i \mathbf{q} \cdot \mathbf{R}_j} e^{2\pi i \mathbf{q} \cdot S \cdot S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j)} S \cdot \mathbf{d}_{(S^{-1} \cdot (\mathbf{R} - \mathbf{R}_j), \bar{\mathcal{S}}j)} \\ &= e^{2\pi i \mathbf{q} \cdot \mathbf{R}_j} S \cdot \mathbf{x}_{\mathbf{q},S,\bar{\mathcal{S}}j}, \end{aligned} \quad (2.32)$$

and so the symmetry simply maps a periodic displacement at atom j and q-point \mathbf{q} onto a periodic dis-

placement at atom $\bar{S}j$ and q-point $\mathbf{q} \cdot S$, transformed by S and with a phase factor which depends only on $\mathbf{q} \cdot \mathbf{R}_j$.

Including the unitary mapping from periodic displacement co-ordinates to normal-mode co-ordinates, the symmetry becomes

$$\mathcal{S}u_\chi = \sum_{\chi'} S_{\chi\chi'} u_{\chi'}, \quad (2.33)$$

where $S_{\chi\chi'}$ is a unitary tensor whose only non-zero elements are between modes at q-points with $\mathbf{q}' = \mathbf{q} \cdot S$ [73].

2.4.4 Pure translations

Pure translations are a special case of crystal symmetry, with $S = \mathbb{1}$. In order to be a symmetry of a crystal, a translation must have $\mathbf{s} = \mathbf{R}_S$, where \mathbf{R}_S is an R-vector of the crystal. One consequence of this is that $\mathbf{R}_j = \mathbf{R}_S$, independent of j .

Thus, a pure translation acts in cartesian co-ordinates as

$$\mathcal{S}\mathbf{r}_{\mathbf{R},j} = \mathbf{r}_{\mathbf{R} - \mathbf{R}_S,j}, \quad (2.34)$$

which in displacement co-ordinates becomes

$$\mathcal{S}\mathbf{d}_{\mathbf{R},j} = \mathbf{d}_{\mathbf{R} - \mathbf{R}_S,j}. \quad (2.35)$$

This simple form continues into periodic co-ordinates,

$$\mathcal{S}\mathbf{x}_{\mathbf{q},j} = e^{2\pi i \mathbf{q} \cdot \mathbf{R}_S} \mathbf{x}_{\mathbf{q},j}, \quad (2.36)$$

and into normal-mode co-ordinates,

$$\mathcal{S}\mathbf{u}_\chi = e^{2\pi i \mathbf{q} \cdot \mathbf{R}_S} \mathbf{u}_\chi. \quad (2.37)$$

This means that a pure translation is simply a \mathbf{q} -dependent phase shift in normal-mode co-ordinates.

2.5 Summary

The approximations made of crystals; that they are perfectly periodic, perfectly elastic solids in which atoms vibrate in equilibrium about fixed static-lattice co-ordinates; make it possible to describe them using the

formalism of normal-mode co-ordinates and finite supercells with periodic boundary conditions. These approximations limit the systems to which the resulting models can be applied, but offer an accurate description of the wide range of systems which can be modelled.

This formalism lays the groundwork for the remainder of this work. The Born-Oppenheimer Hamiltonian is very naturally described in terms of normal-mode co-ordinates, and Born-von Karman supercells provide an ideal method for constructing models of the PES from the underlying physical system. This sets the stage for constructing, approximating and solving the nuclear Schrödinger equation, in order to calculate free energies and other properties as desired.

Chapter 3

The problem of free energy

Free energy is a thermodynamic potential which defines the amount of energy which is available to do work at constant temperature [74]. Free energy is a relative quantity, and the free energy of a given system must be defined with respect to some other reference system. If two different phases of the same material have different free energies then the phase with the larger free energy can spontaneously decay into the other phase, doing work equal to the free energy difference between the phases as it does so.

This means that the stability of a given phase is defined by its free energy; the lower the phase's free energy, the more stable it is [2, 22]. Calculating the free energy of all competing phases over an appropriate range of thermodynamic conditions is thus sufficient to calculate the phase diagram of the system [24].

There are two kinds of free energy which will be considered here. The *Helmholtz free energy*, F , defines the energy available at constant volume, and the *Gibbs free energy*, G , defines the energy available at constant pressure. The Gibbs free energy can be calculated in terms of F and its derivatives with respect to the crystal lattice vectors [22], and so this work will focus only on the calculation of F . Whenever free energies are mentioned in the remainder of this work it should be assumed that this refers to F and not G , unless otherwise specified.

The calculation of the free energy of a given crystal is a three-stage process [13]. Firstly, a representation of the PES must be chosen. Secondly, the free parameters of this representation must be fitted using electronic structure calculations. Finally, the free energy must be calculated from the PES representation.

The three stages are somewhat independent, although the choice of method for each stage places limitations on the other stages. However, as each stage has associated costs and accuracies it is important to pick methods which complement one another so as to avoid having an overall calculation with higher computational cost

than necessary for its accuracy [21].

This chapter will first outline the underlying theory of calculating free energy exactly, along with reasons why this theory must be approximated in practice. Secondly the simplest approximation to the free energy, the *harmonic approximation*, will be detailed, along with a description of why this approximation is insufficient for many purposes. Finally, the existing methods for each stage of the calculation beyond the harmonic approximation will be reviewed.

3.1 Framing the problem

3.1.1 Calculation without approximation

Without making approximations, the free energy is calculated from the PES using quantum mechanics and statistical mechanics [75].

Once the PES, V , has been mapped, it can be used to construct the Hamiltonian, \mathcal{H} ,

$$\mathcal{H} = \mathcal{T} + V, \quad (3.1)$$

where \mathcal{T} is the nuclear kinetic energy.

The eigenstates of the system can then be calculated using the non-relativistic, time-independent Schrödinger equation,

$$\mathcal{H} |\psi\rangle = E_\psi |\psi\rangle, \quad (3.2)$$

where each $|\psi\rangle$ is an eigenstate of the Hamiltonian, and each E_ψ is the corresponding eigenenergy.

Once the eigenstates and eigenenergies have been calculated, the thermal weightings of the states, P_ψ , are calculated as Boltzmann factors,

$$P_\psi = \frac{e^{-E_\psi/T}}{\sum_\psi e^{-E_\psi/T}}. \quad (3.3)$$

This allows the calculation of the internal energy, which is the weighted expectation of the Hamiltonian,

$$U = \langle \mathcal{H} \rangle = \sum_\psi \langle \psi | P_\psi \mathcal{H} | \psi \rangle, \quad (3.4)$$

and the entropy, given by the Gibbs formula,

$$S = - \sum_{\psi} P_{\psi} \ln(P_{\psi}). \quad (3.5)$$

The free energy can then be calculated from U and S as

$$F = U - TS. \quad (3.6)$$

Finally, if the pressure p and volume Ω of the system are known, the Gibbs free energy is calculated as

$$G = F + p\Omega. \quad (3.7)$$

3.1.2 Calculation with approximation

The PES and vibrational eigenstates of a macroscopic crystal are functions with massive dimensionality, and so calculating the free energy is intractable without making approximations [13].

As detailed in section 2.1.4, under the vibrational approximation it is only necessary to map a region of the PES about the static-lattice configuration. As such, approximations to the PES can focus on accurately capturing the PES in this region, while neglecting the features of the PES at larger distances from the static-lattice configuration.

Similarly, the calculation of the free energy is dominated by the eigenstates of the Hamiltonian which have the lowest energy, and the densities of these eigenstates are concentrated in the low-energy regions of the PES, in the vicinity of the static-lattice configuration. As such, approximate free-energy methods focus only on accurately capturing the behaviour of low-energy states in the low-energy region of the PES, neglecting the states' details outside of the low-energy region of the PES, and neglecting states with higher energies entirely.

Different approximations lead to results which are applicable to different systems, and to results of varying accuracy. For the approximations which are widely used, there is generally a trade-off where increasing accuracy incurs increasing computational cost.

3.1.3 Vibrational free energy

The PES is a relative quantity, defined only up to a constant offset. The thermodynamic potentials are defined in terms of the PES, and so they inherit this offset. The offset is arbitrary, but must be defined consistently between different systems.

It is natural to separate the thermodynamic potentials into their static-lattice values and the contribution from the PES [34], as

$$F = V(\mathbf{r}^0) + F[V(\mathbf{u})]. \quad (3.8)$$

Under this separation, it is natural to define arbitrary offset in the PES such that $V(\mathbf{r}^0)$, the *static-lattice energy*, is zero.

This work is only concerned with the calculation of $F[V(\mathbf{u})]$, the *vibrational free energy*. Defining $V(\mathbf{r}^0) = 0$ removes the arbitrary offset from the vibrational free energy, and so this is an absolute quantity. The calculation of $V(\mathbf{r}^0)$, including its arbitrary offset, is left to the electronic structure method.

For the remainder of this work, all references to the calculation of the free energy will refer to the vibrational free energy. The only time the static-lattice energy will be considered will be in the results section, when the choice of electronic structure method has been made and it becomes necessary to calculate the full free energy differences between different structures.

It should be noted that the separation of the free energy into a static-lattice component and a vibrational component is only valid under the assumption that the Born-Oppenheimer approximation holds perfectly. Small deviations from the Born-Oppenheimer approximation can be corrected for by adding temperature-dependent electronic terms and electron-phonon coupling terms to the free energy [21, 76]. The details of these terms depend on the electronic structure method, and will not be considered further here.

3.2 The harmonic approximation

3.2.1 The harmonic potential

The harmonic approximation is perhaps the simplest, and certainly the most widely used, approximation to the PES. Under the harmonic approximation, the PES is expanded as a Taylor series in cartesian displacement co-ordinates, and this Taylor series is truncated at the second order. This gives a very simple form for the PES [2],

$$V = \sum_{\rho_1 \rho_2} V_{\rho_1 \rho_2}^{(2)} d_{\rho_1} d_{\rho_2}, \quad (3.9)$$

where $V_{\rho_1\rho_2}^{(2)}$ are real coefficients.

Transforming to normal-mode co-ordinates, the potential becomes

$$V = \sum_{\chi_1\chi_2} V_{\chi_1\chi_2}^{(2)} u_{\chi_1} u_{\chi_2}, \quad (3.10)$$

where $V_{\chi_1\chi_2}^{(2)}$ are complex coefficients, related to $V_{\rho_1\rho_2}^{(2)}$ by [67]

$$V_{\chi_1\chi_2}^{(2)} = \sum_{\rho_1\rho_2} V_{\rho_1\rho_2}^{(2)} U_{\rho_1\chi_1} U_{\rho_2\chi_2}. \quad (3.11)$$

The normal-mode co-ordinates are chosen such that they diagonalise the harmonic potential, i.e. such that $V_{\chi_1\chi_2}^{(2)}$ is zero unless $\chi_1 = \bar{\chi}_2$. As such, the PES becomes

$$V = \sum_{\chi} \frac{1}{2} N \omega_{\chi}^2 u_{\chi} u_{\bar{\chi}}, \quad (3.12)$$

where $\omega_{\chi} = \sqrt{\frac{2}{N} V_{\chi\bar{\chi}}^{(2)}}$ is the frequency of the χ mode, and N is the size of the supercell used to define the normal mode co-ordinates. This potential is simply a sum of single- and double-mode harmonic oscillators, where the single-mode oscillators correspond to modes with $\chi = \bar{\chi}$, and the double-mode oscillators correspond to modes with $\chi \neq \bar{\chi}$.

3.2.2 Fitting the harmonic potential

The only free parameters in the harmonic potential are the coefficients $V_{\rho_1\rho_2}^{(2)}$. These can be fitted simply by calculating the matrix of second derivatives of the PES with respect to the atomic displacements. There are many standard ways of doing this [29], including *linear response* methods which use perturbation theory to calculate the derivatives directly, and *finite displacement* methods which use calculations of forces or energies at small values of \mathbf{u} to calculate the derivatives indirectly.

3.2.3 Free energy under the harmonic approximation

A major advantage of the harmonic potential is that the corresponding Schrödinger equation can be solved analytically. The details of this are outlined in appendix A, and only the results will be presented here.

The partition function for the harmonic Hamiltonian is [67]

$$Z = \prod_{\chi} \frac{2}{\sinh\left(\frac{\omega_{\chi}}{2T}\right)}, \quad (3.13)$$

and so the internal energy is

$$U = \sum_{\chi} \frac{\omega_{\chi}}{2} \coth\left(\frac{\omega_{\chi}}{2T}\right) \quad (3.14)$$

and free energy is [2]

$$F = \sum_{\chi} T \ln\left(\frac{1}{2} \sinh\left(\frac{\omega_{\chi}}{2T}\right)\right). \quad (3.15)$$

3.2.4 Interpolating the harmonic potential

Another advantage of the harmonic approximation is that it is straightforward to interpolate from a calculation of a finite Born-von Karman supercell to a calculation of an infinite crystal [13], under the approximations discussed in section 2.3.3.

Interpolation in displacement co-ordinates

The harmonic potential in displacement co-ordinates in the simulated supercell S is

$$V = \sum_{\substack{\rho_1, \rho_2 \\ \rho \in S}} V_{\rho_1, \rho_2}^{(2)} d_{\rho_1} d_{\rho_2}, \quad (3.16)$$

where the indices ρ_1 and ρ_2 run over all of the atoms in S .

Moving to a different supercell, S' , the potential becomes

$$V = \sum_{\substack{\rho'_1, \rho'_2 \\ \rho' \in S'}} V_{\rho'_1, \rho'_2}^{(2)} d_{\rho'_1} d_{\rho'_2}, \quad (3.17)$$

which has the same functional form as equation 3.16, but now the indices ρ'_1 and ρ'_2 run over all of the atoms in S' .

Under the minimum image approximation [13], the coefficients $V_{\rho'_1, \rho'_2}$ can be constructed as a linear combination of the coefficients V_{ρ_1, ρ_2} , as

$$V_{\rho'_1, \rho'_2} = \sum_{\substack{\rho_1, \rho_2 \\ \rho \in S}} \theta_{\rho'_1, \rho'_2, \rho_1, \rho_2} V_{\rho_1, \rho_2}, \quad (3.18)$$

where $\theta_{\rho'_1 \rho'_2 \rho_1 \rho_2}$ is the *pairwise minimum image function*, defined as

$$\theta_{\rho'_1 \rho'_2 \rho_1 \rho_2} = \begin{cases} 0 & \text{if } \rho'_1 \text{ is not an image of } \rho_1 \text{ in } S. \\ 0 & \text{if } \rho'_2 \text{ is not a minimum image of } \rho_2 \text{ with respect to } \rho'_1 \text{ in } S. \\ 1/n_{\rho_1 \rho_2} & \text{otherwise} \end{cases} \quad (3.19)$$

where $n_{\rho_1 \rho_2}$ is the number of images of ρ'_2 which are minimum images with respect to ρ'_1 . It should be noted that $\sum_{\rho'_1 \rho'_2} \theta_{\rho'_1 \rho'_2 \rho_1 \rho_2} = 1$ for all values of ρ_1 and ρ_2 .

As mentioned in section 2.3.3, this corresponds to approximating the inter-atomic interactions as having finite range, with each atom only interacting with other atoms which lie within the Wigner-Seitz cell of S centred on that atom [13].

Interpolation in normal-mode co-ordinates

Transforming from real-space co-ordinates to normal-mode co-ordinates, the interpolation becomes

$$N' \omega_{\chi'}^2 = \sum_{\substack{\rho'_1 \rho'_2 \\ \rho' \in S'}} U_{\rho'_1 \chi'} U_{\rho'_2 \bar{\chi}'} \sum_{\substack{\rho_1 \rho_2 \\ \rho \in S}} \theta_{\rho'_1 \rho'_2 \rho_1 \rho_2} \sum_{\chi} \frac{1}{N^2} U_{\rho_1 \bar{\chi}} U_{\rho_2 \chi} N \omega_{\chi}^2. \quad (3.20)$$

Since for the non-zero terms the ρ' displacements must be images of the corresponding ρ displacements, the sum $\sum_{\rho'_1 \rho'_2 \rho_1 \rho_2}$ can be reduced to $\sum_{\mathbf{R}'_1 \mathbf{R}'_2 j_1 \alpha_1 \mathbf{R}_1 j_2 \alpha_2 \mathbf{R}_2}$.

In order to simplify equation 3.20, it is convenient to introduce the *q-point overlap function*

$$f_{j_1 j_2}(\mathbf{q}' - \mathbf{q}) = \frac{1}{NN'} \sum_{\mathbf{R}'_1 \mathbf{R}'_2 \mathbf{R}_1 \mathbf{R}_2} \theta_{\rho'_1 \rho'_2 \rho_1 \rho_2} e^{2\pi i(\mathbf{q}' - \mathbf{q}) \cdot (\mathbf{R}_2 - \mathbf{R}_1)}, \quad (3.21)$$

which describes the overlap between \mathbf{q}' and \mathbf{q} with respect to the atomic pair j_1 and j_2 . The q-point overlap is equal to one if $\mathbf{q}' = \mathbf{q}$, zero if \mathbf{q}' is commensurate with S but is not \mathbf{q} , and in most cases¹ rapidly decreases towards zero with increasing $|\mathbf{q}' - \mathbf{q}|$.

It is also convenient to introduce the *mode overlap function*

$$U_{j \chi' \bar{\chi}} = \sum_{\alpha} U_{j \alpha \mathbf{q}' k'} U_{j \alpha \bar{\mathbf{q}} k}, \quad (3.22)$$

which describes the overlap between modes χ' and χ with respect to atom j . The total mode overlap between χ' and χ is described by $\sum_j U_{j \chi' \bar{\chi}}$, which may take any complex value with a norm of at most one.

¹The q-point overlap function only changes with $|\mathbf{q}' - \mathbf{q}|$ if $(\mathbf{q}' - \mathbf{q}) \cdot (\mathbf{R}_2 - \mathbf{R}_1)$ takes multiple values.

Using the overlap functions, the interpolation simplifies to

$$\omega_{\chi'}^2 = \sum_{\chi} \sum_{j_1, j_2} U_{j_1 \chi' \bar{\chi}} U_{j_2 \bar{\chi} \chi} f_{j_1 j_2}(\mathbf{q}' - \mathbf{q}) \omega_{\chi}^2. \quad (3.23)$$

The frequency $\omega_{\chi'}$ varies smoothly with changes in \mathbf{q}' and $U_{j\alpha\mathbf{q}'k'}$, and whenever a modes in S' is identical to a mode in S , $\chi' = \chi$, its frequency is also the same, $\omega_{\chi} = \omega_{\chi'}$.

Interpolating the harmonic free energy

By equation 3.15 above, the free energy under the harmonic approximation is simply a sum of one-mode terms, where each one-mode term only depends on the frequency of that mode. By equation 3.23, the normal modes and corresponding frequencies at an arbitrary interpolated q-point can be calculated as a linear combination of the calculated modes. Combining the two equations means that the free energy can be calculated at an arbitrary q-point.

As such, the free energy can be calculated using an arbitrarily dense q-point grid, simply by calculating and adding together the contributions to the free energy from each q-point on that grid. This is computationally inexpensive, as each q-point can be calculated separately without making reference to the other q-points on the interpolated grid.

3.2.5 Advantages and disadvantages of the harmonic approximation

The harmonic approximation is widely used [21], and has a number of advantages. It is extremely simple to represent the harmonic PES, and the free parameters of the model are straightforward to fit and unambiguous. The calculation of the free energy is analytic and also extremely simple, and this can easily be interpolated to the large-supercell limit. All this means that calculations using the harmonic approximation are comparatively inexpensive, and can easily be made to be stable and robust to errors.

There are, however, a number of weaknesses and limitations inherent in the harmonic approximation [21, 23, 77]. The simple structure of the PES means that it is only a good representation for a limited number of systems [22]. The nature of the free energy also means that many of the details of the system are lost. For example, any heat capacity calculated under the harmonic approximation will artificially be independent of temperature.

The harmonic approximation also becomes increasingly inaccurate as temperature is increased. The PES of a periodic solid is itself periodic, and so it cannot be well approximated across its entire domain by a

harmonic potential. This does not always matter at low temperatures, where the PES need only be accurate for small displacements $|\mathbf{u}|$. However, as the temperature increases the relevant range of displacements $|\mathbf{u}|$ increases, and the deviation of the harmonic PES from the true PES causes the approximation to become inaccurate.

The harmonic approximation is also inappropriate for modelling structures about a static-lattice point which is a saddle-point of the PES rather than a minimum. For such structures, the harmonic frequencies associated with one or more normal modes will be imaginary. This means that the free energy of the system is undefined, and under the harmonic approximation the structure is unstable. This is particularly relevant when studying phases which are only stable at high temperatures, since the natural choice of static-lattice configuration for such structures is typically a saddle-point [62].

3.3 Sampling the potential energy surface

In order to move beyond the harmonic approximation it is necessary to choose an anharmonic representation of the PES, and then to fit the free parameters of that representation to the true PES. The “true” PES here refers to the Born-Oppenheimer potential energy surface as calculated by whichever electronic structure method is being used to simulate the system. This surface will generally differ from the PES of the physical crystal, but such differences are not relevant to the current work.

The choice of PES representation constrains how the PES may be sampled, as detailed in section 3.4 below. Before this choice can be made it is important to understand the methods by which each PES representation may be sampled and fitted, as it is these methods which generally determine the overall computational cost of the entire calculation, along with the stability, accuracy and reliability of the fit.

For all but the simplest of electronic structure methods, it is not possible to gather information about large regions of the PES at once. Instead, the PES is fitted to a set of *samples*, each of which is an individual electronic structure calculation at particular nuclear configuration \mathbf{u} . Each sample yields information about the PES at that sample, which always includes the energy $V(\mathbf{u})$ and, depending on the electronic structure method, may include the forces $\mathbf{f}(\mathbf{u})$, the Hessian $D(\mathbf{u})$, and other variables.

Once the PES has been sampled, it is then necessary to calculate the free parameters of the representation from the samples. The details of this calculation vary depending on the details of the representation, and on what information is available from the samples.

3.3.1 Electronic structure methods

There are many electronic structure methods available, each with its own set of strengths and weaknesses. For the purposes of this work, these methods can be treated as a black box, and so the details of how each works will not be discussed here. These details are discussed extensively in the papers cited below and elsewhere. In addition, many electronic structure methods have analogous vibrational methods, and the details of these are briefly discussed in section 3.5.

The only details which matter for the present work are the variables which can be calculated at each sample, and the accuracy and computational cost of each calculation.

Classical potentials

There are many ways of approximating electronic structure in a manner which is not ab-initio. These will be collectively referred to as *classical potentials*. These range from models where the PES is a simple function of the nuclear co-ordinates, through to advanced models using chemical orbitals or machine learning [78].

Classical potentials are broadly similar in concept to the representations of the PES discussed here, with the major difference that they attempt to represent most or all of the PES, rather than focussing on just the region around a given static-lattice configuration. This inevitably means that the accuracy of the potential in the region of interest is less than that of a bespoke representation of the PES.

Classical potentials are generally extremely fast to sample, massively reducing the computational cost of the overall method. In addition, many classical potentials are represented in a manner such that a representation of the PES about the static-lattice configuration can be calculated directly from the potential, without having to take individual samples.

However, classical potentials are rarely accurate enough to calculate observables of interest to the desired accuracy [2, 23].

Quantum chemistry

There are a number of variational approximations which can be made to the electronic wavefunction in order to make its calculation tractable. These are collectively known as quantum chemical methods:

- Under *Hartree-Fock theory* (HF) [18], the many-body wavefunction is approximated as a single Slater determinant of single-body wavefunctions, approximating the Hamiltonian as a sum of single-body

Hamiltonians.

- *Møller-Plesset perturbation theory* (MP) [16, 79] treats the difference between the Hartree-Fock Hamiltonian and the true Hamiltonian perturbatively, calculating perturbative corrections to the results of Hartree-Fock theory.
- *Configuration interaction* (CI) [80] and *coupled cluster* (CC) [20] take the Hartree-Fock eigenstates and construct many-body eigenstates which are combinations of these eigenstates, using variational methods to minimise the free energy with respect to the free parameters of these variations.

Taken together, the quantum chemical methods provide a hierarchy of methods of increasing accuracy and computational cost. The most accurate forms of CI and CC can achieve arbitrary accuracy, but have costs which scale very poorly with the size of the system. For example, the cost of the most accurate form of CC in widespread use scales as $\mathcal{O}(N_e^7)$ in the number of electrons N_e [81].

Quantum chemical methods allow the direct calculation of forces, Hessians and other properties using linear response theory [82]. However, such calculations are more expensive than energy calculations, and so whether they are cost-efficient must be evaluated on a case-by-case basis.

Density-functional theory

Density-functional theory (DFT) [16–19, 83, 84] exploits the fact that at zero temperature the electronic free energy is purely a functional of the ground-state three-dimensional electron density, rather than depending on the full details of the ground-state many-body wavefunction. Although the form of this functional is not known, there are a variety of functionals which can be substituted for the true density functional which produce results to a reasonable accuracy.

DFT is not systematically improvable. There are density functionals which are generally more or less accurate for given systems, but there is no formal hierarchy of functionals, and there is no guarantee that a functional which is supposed to be more accurate will always increase the accuracy of a given calculation [19].

Despite this, DFT is widely used. This is because its accuracy, while bounded, is generally good enough to produce acceptable results. In addition, the computational cost of DFT is very favourable, and the scaling of DFT with the number of electrons in the system, N_e , is at worst $\mathcal{O}(N_e^3)$ [83], and some DFT implementations can reduce this to $\mathcal{O}(N_e)$ for a range of systems [85].

There are several methods which extend the DFT formalism in order to achieve higher accuracies than can be achieved by DFT alone. These methods include *Green's function perturbation theory* [17, 86, 87] and *dynamical mean-field theory* [88, 89]. These methods are significantly more accurate than DFT methods

but are also significantly more computationally demanding, generally making them impractical for use with vibrational calculations.

The calculation of forces using DFT has a broadly similar computational cost to the calculation of energies, and so forces should always be used [90]. Higher derivatives of the PES, including the Hessian, can be calculated using the related method *density-functional perturbation theory* (DFPT) [91]. The calculation of second derivatives using DFPT is widely used, and more recently the calculations of third derivatives have been considered [41].

Calculating the Hessian using DFPT is more computationally expensive than calculating energies and forces using DFT, both in terms of the cost of simulating a small system, and in terms of the scaling of the cost with the number of electrons in the system. The implementation of DFPT is also significantly more complicated than that of DFT, meaning that it is only available for a subset of the available DFT methods [45].

Diffusion Monte-Carlo

Diffusion Monte-Carlo (DMC) stochastically samples the ground-state electronic wavefunction using a number of *walkers*, which diffuse through phase space, repeatedly sampling the electronic energy, and are created and destroyed according to the imaginary-time Schrödinger equation [17, 92].

The diffusion of the walkers depends on a *nodal surface*, which defines the sign of the wavefunction as a function of the electronic co-ordinate. This nodal surface is an approximation, and limits the accuracy of the method [17]. Despite this, DMC is generally significantly more accurate than DFT, although it is also typically more computationally expensive than DFT. In addition, DMC treats electron correlation explicitly, and so it cannot exactly represent periodic systems using periodic boundary conditions. This means that DMC calculations typically require calculations to be performed using supercells which are larger than would be necessary for other methods, which can significantly increase their computational cost [17].

DMC can calculate forces and Hessians [93], at increased cost compared to energy calculations. These calculations are not exact, but rather are estimators of the true forces and Hessians [17].

3.3.2 Choice of samples

Fitting a given representation of the PES requires a minimum amount of information, and the accuracy of this fit will generally improve with additional information. Information about the PES can be gained by increasing the number of samples, improving the locations of the samples, and increasing the amount of information calculated at each sample. Each gain of information comes with a corresponding increase in

computational cost, and so the overall strategy for sampling the PES must be chosen with care, so as to reach the desired accuracy at the minimal cost.

The PES can be sampled using a regular configuration of samples, or a less regular method. Using regular configurations of samples can lead to greater stability than other methods [42], and makes it possible to take advantage of the expected smoothness of the PES by first interpolating the PES samples before fitting the PES representation [94]. On the other hand, less regular sampling methods are more flexible, and make it easier to adaptively sample the PES, such that more important regions are sampled more thoroughly [34].

A PES sample at $\mathbf{u} = \sum_{\chi} u_{\chi} \hat{\mathbf{u}}_{\chi}$ generally yields more information as the number of non-zero displacements u_{χ} increases [13]. However, as described in section 2.3, PES samples can only be calculated using supercells which are commensurate with all of the q-points corresponding to their non-zero displacements [70]. This means that samples generally require larger supercells as the number of non-zero displacements increases, with correspondingly increased computational costs. This leads to another tradeoff between reduced computational cost and increased information.

There is a similar tradeoff with the number of crystal symmetries broken by a given sample. Breaking symmetries generally increases the information yielded by a sample [13], but many electronic structure methods use symmetries to accelerate calculation, and so become more expensive as symmetries are broken [95].

As previously mentioned, many electronic structure methods offer a range of variables which can be calculated at each given sample. The higher-information variables, such as the Hessian, are generally more expensive to calculate than the lower-information variables, such as the energy. Depending on how the cost increases with the increased information, it again may or may not be worth calculating these additional variables [34].

The optimal choice for each tradeoff depends on the details of the electronic structure method, and can also depend on the details of the system being simulated. There are many proposed sampling methods for particular electronic structure methods in the literature [13, 34, 42, 70]. However, it is desirable to be able to calculate bespoke sampling methods on a method-by-method and system-by-system basis. Some developments towards this goal will be detailed in section 4.4.

3.3.3 Multi-level calculations

In the context of molecular modelling, it is common practice to use two or more different electronic structure methods for sampling the PES, where the methods are hierarchically arranged in order of increasing accuracy but also increasing computational cost [34].

This is useful because the cheaper, less accurate methods can be used for sampling the PES in regions where the exact details of the PES are expected to contribute less to the free energy, such as further from the static-lattice configuration [96]. Additionally, the results of less accurate methods can be used as a guide to which features of the PES should be sampled using more accurate methods [97, 98].

This is harder to achieve when using electronic structure methods which are not systematically improvable, such as DFT and DMC. Unfortunately, these are the electronic structure methods of choice for modelling many crystals, and so multi-level calculations are not generally performed for crystals.

3.4 Representing the potential energy surface

The first step when constructing a model PES is to define how that PES will be represented. As detailed in section 3.3, the choice of representation constrains how the PES can be sampled and fitted. The PES representation also determines the range of systems for which accurate results can be calculated, and the accuracy of those results. Additionally, when calculating the free energy from the PES the cost and accuracy of the calculation is in part determined by how the PES is represented.

There are many ways of representing the PES [42], and these can be grouped into three main categories: *basis function methods*, which construct the PES as a linear combination of basis functions with functional forms which are defined across the entire region of interest; *grid-based methods*, where the PES is sampled on a regular grid of points, and properties are either calculated from these points directly or from a simple interpolation between these points; and *stochastic and dynamical methods*, where the PES is sampled via a stochastic or dynamical process until a representative set of points have been sampled, and where the PES is never explicitly stored but rather properties are calculated directly from the samples.

3.4.1 Grid-based representations

It is perhaps the most straightforward to represent the PES directly as a regular grid of samples, typically at co-ordinates which are determined in advance of any electronic structure calculations being performed.

Unlike other representations, grid-based representations make no assumptions about the PES. As such, if the grid is sufficiently dense and covers all relevant regions of the PES then these methods are likely to be more accurate and reliable than any other method.

The main disadvantage of grid-based representations is their computational cost. These representations provide little information about the PES in the regions which are not close to a grid point. As such, the features of the PES must be entirely covered by grid points, and the spacing between grid points must be

short enough to capture the variation of the PES. If the PES has large, high-dimensional features this makes these representations extremely costly [94].

Grid-based representations can be improved by interpolating the PES between the grid points, exploiting the expected smoothness of the PES. A popular interpolation scheme is Shepard interpolation [34, 77], which linearly interpolates between regions where the PES is known. This allows for the derivatives of the PES calculated at given points to be included in the interpolation of the values of the PES at those points.

While the set of samples required by grid-based representations is less flexible than for other representations, there are still some adaptations that can be made. The density of the grid points can be varied, such that the most important features of the PES are more thoroughly sampled [98]. It is also possible to use a grid of samples to fit a set of basis functions [96], although this is likely to combine many of the disadvantages of both grid-based representations and basis-function representations.

It is relatively trivial to use symmetry to reduce both the number of samples needed and the numerical error in the sampling process [98]. In contrast, it is not at all clear how size consistency could be exploited when using grid-based representations.

3.4.2 Stochastic and dynamical representations

Instead of constructing an explicit representation of the PES, it is also possible to proceed by collecting a set of samples of the PES which, with proper weighting, are representative of the whole PES, such that the expectation of properties across these samples converge to the equilibrium values of these properties.

The most common procedures for generating these samples are based on molecular dynamics [22] and Monte-Carlo methods [48, 99]. In both cases, the PES measurement at each sample is used to generate the next sample, and as the number of samples increases the set of samples converges towards being representative of the whole PES.

A recent innovation is the Gaussian approximation potential [100]. For a given set of samples this method generates an approximation to the PES, along with a measure of how accurate this approximation is throughout space. This measure can then be used to generate further samples, in a manner which improves the accuracy of the PES as optimally as possible. These methods were originally developed in the context of transferable potentials, but they have recently been used for vibrational calculations [77], with reported performance being at least as accurate as, and frequently more accurate than, equivalent Taylor-expansion methods.

Another important approach to mapping the PES is the adaptive density-guided approach [34], which uses

the result of a vibrational calculation to identify the regions of phase space where the magnitude of the vibrational density² is highest, and samples the PES more thoroughly there, before running a second vibrational calculation. This approach is particularly useful when more than one vibrational method is being employed; a rough mapping of the PES can be made using an inexpensive vibrational method to estimate the vibrational density, and this density can be used to make a much more accurate mapping of the PES, suitable for use with a more accurate and more expensive vibrational method. This method can also be used iteratively, repeatedly adding PES samples and calculating the free energy until the calculation converges [34].

The major advantage of stochastic and dynamical representations is that they generally have systematically improvable accuracy. If the method of selecting new points to sample is good enough then simply leaving the method running for longer and allowing it to sample more of the PES will continually improve the accuracy of the resulting PES. If this is paired with a metric for estimating the accuracy of the method then, in theory, any desirable accuracy can be achieved.

3.4.3 Basis function representations

Rather than representing the PES directly by the set of samples, it can instead be represented as a sum of basis functions, where the basis function coefficients are calculated from the PES samples. This somewhat separates the PES representation from the PES samples, which can have both advantages and disadvantages.

Unlike the other PES representations discussed in this work, there is no unique way of sampling the PES for a given basis function representation. This means there is a large degree of flexibility in how the PES is sampled, allowing for different sampling schemes to be used in different circumstances, raising the possibility of methods with variable accuracy, stability and speed as required [42].

The main limitation of basis function methods is that they inherently make assumptions about the features of the PES. Any given set of basis functions is limited to representing a subset of PES features, and so there will only be a subset of systems for which these basis functions can be a good representation. However, if the set of basis functions is a good fit to the PES of a given system then representing the PES using basis functions will generally be faster than using other representations.

Basis function representations are, in theory, systematically improvable, as simply adding more basis functions will improve the accuracy of the result. However, this is not always practically possible. The increase in computational cost with the number of basis functions can be prohibitive, and the increase in accuracy is generally quite slow [99].

²The vibrational density $\rho(\mathbf{r})$ is the thermal average of the magnitude of the vibrational wavefunctions, $\rho(\mathbf{r}) = \sum_{\psi} P_{\psi} |\langle \mathbf{r} | \psi \rangle|^2$. This is relevant because the expectation of the potential can be written as $\langle V \rangle = \int V(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r}$.

A major advantage of basis function representations is that once they are constructed they are much easier to work with than other PES representations. They are often amenable to analytic integration, which is much faster and more reliable than the numerical integration required for other PES representations. Further, the decomposition of the PES into explicit basis functions means that notions like symmetry and size consistency can be applied one basis function at a time, considerably simplifying their use.

Taylor series representations

For vibrational models, where the only relevant section of the PES is that in the vicinity of the static-lattice configuration, it is natural to expand the PES as a Taylor series in cartesian displacement co-ordinates [21],

$$V = \sum_{n=2}^{\infty} V_{\rho_1 \dots \rho_n}^{(n)} \prod_{j=1}^n d_{\rho_j}, \quad (3.24)$$

where each $V_{\rho_1 \dots \rho_n}^{(n)}$ is a real constant coefficient.

The integer n is the *expansion order* of the term. There is no $n = 0$ term because the PES is defined relative to the static-lattice energy, and there is no $n = 1$ term because the static-lattice configuration is always a stationary point of the PES.

The Taylor series can be transformed into normal-mode co-ordinates as

$$V = \sum_{n=2}^{\infty} V_{\chi_1 \dots \chi_n}^{(n)} \prod_{j=1}^n u_{\chi_j}, \quad (3.25)$$

where each $V_{\chi_1 \dots \chi_n}^{(n)}$ is a complex constant coefficient, related to the coefficients in cartesian displacement co-ordinates by

$$V_{\chi_1 \dots \chi_n}^{(n)} = \sum_{\rho_1 \dots \rho_n} V_{\rho_1 \dots \rho_n}^{(n)} \prod_{j=1}^n U_{\rho \chi_j} \quad (3.26)$$

and

$$V_{\rho_1 \dots \rho_n}^{(n)} = \frac{1}{N^n} \sum_{\chi_1 \dots \chi_n} V_{\chi_1 \dots \chi_n}^{(n)} \prod_{j=1}^n U_{\rho \chi_j}. \quad (3.27)$$

The simple structure of the Taylor series representation makes it comparatively simple to work with, both in terms of mapping the PES and working with the PES once it has been mapped [21]. This representation is also particularly amenable to notions of symmetry and size consistency. This will be discussed further in sections 4.1 and 4.2 respectively. As detailed in appendix A, Taylor series can often be integrated analytically with respect to analytic basis states, rather than relying on costly numerical integration. Additionally, Taylor series are the starting point for many formulations of quantum mechanics, so no conversion of the PES is

needed before further calculations can be run.

Further, the harmonic approximation is simply a Taylor series truncated to second order, and so a Taylor series PES is simply the sum of a harmonic PES and an anharmonic correction. This makes it easy to measure the anharmonicity of the potential, and allows the harmonic and anharmonic components of the PES to be treated separately. This will be exploited in section 4.3.

Taylor series are naturally iteratively improvable. Any model using a given subset of a Taylor series can be improved simply by including more terms from the Taylor series. Which terms are included or neglected can be decided based on physical reasoning, as each term corresponds to an interaction between a given set of modes at a given order.

However, there are a number of disadvantages to Taylor-series representations. Even though Taylor series are iteratively improvable, there is no guarantee that the series will converge to the underlying PES [101]. Additionally, as discussed in section 8.1, Taylor series potentials cannot be made invariant under those symmetries which do not preserve $|\mathbf{u}|$. This means that truncated Taylor series can only accurately represent a finite region of the PES, and such representations are not appropriate for systems where atoms can travel large distances from their static-lattice positions.

Taylor series can be tuned to capture the PES at small $|\mathbf{u}|$ by limiting the value of n [96]. It is usually only feasible to include terms up to a small, even expansion order, typically $n = 4$ [42] or $n = 6$ [45]. This comes at the cost of being unable to capture large-displacement behaviour, where the PES may flatten with displacement as bonds weaken or may spike sharply as atoms are forced into proximity with one another [101].

Taylor series representations using normal-mode co-ordinates can also provide a poor description of systems which have freely-rotating torsional degrees of freedom [22, 34]. This problem may be overcome by using *curvilinear co-ordinates*, but identifying the correct curvilinear co-ordinates for a given crystal is non-trivial, and the details of fitting the PES and calculating free energy become significantly more complicated when using these co-ordinates [98].

Another choice of co-ordinates which has previously been successfully used to map the PES are *cluster co-ordinates* [102]. These co-ordinates are specifically constructed to behave well under the action of the crystal symmetry operators. However, cluster co-ordinates are localised in real space, with the consequence that interpolating between supercells and evaluating size consistency are likely to be challenging.

In spite of these limitations, the simplicity of the Taylor expansion makes it a natural default representation for the PESs of many systems [21], and the option is always available to use a different representation when a Taylor series is not appropriate.

The principal axes approximation

In addition to truncating the Taylor expansion at a given order, it is common practice to restrict the included terms to those which contain a limited number of modes. For example, with a limit of two coupled modes, terms of the form $u_1^{n_1}u_2^{n_2}$ would be included but terms of the form $u_1^{n_1}u_2^{n_2}u_3^{n_3}$ would not. This restriction is known by different sources as the principle axes approximation (PAA) [26] or the n-mode representation [42]. The rationale behind this approximation is the idea that the coefficients of the terms in the PES will decrease in magnitude as the number of modes on which those terms depend increases [103].

The PAA is widely used in the simulation of molecular systems, where it is broadly found to accelerate calculations at little loss of accuracy [23]. It has more recently been applied to the simulation of periodic crystals [26]. However, it will be demonstrated in sections 4.1.2 and 4.2.4 that the PAA is neither symmetry-invariant nor size-consistent, and as such must be heavily modified before its application to extended systems can become physically reasonable.

It should be noted that the PAA is usually formulated in terms of real periodic co-ordinates rather than the complex periodic co-ordinates used in this work, as defined in section 2.2.3. This co-ordinate change does not solve the problems with the approximation, and ceases to be relevant once the PAA is amended as detailed in section 4.1.2.

3.4.4 Summary

The only representation of the PES which will be considered in this work is a basis-function representation based on a truncated Taylor series. This is not because this is considered the best representation, indeed it has a large number of drawbacks, but rather because it is the representation which is most amenable to acceleration using the methods presented in chapter 4. It is hoped that this acceleration can be transferred to other PES representations in future work.

3.5 Calculating free energy

Once the PES of a crystal has been mapped, all that remains of the calculation is to calculate the eigenstates of the vibrational Hamiltonian, and to use these eigenstates to calculate the free energy of the crystal and any other properties of interest. A completely accurate calculation of the free energy would require working with an infinite number of eigenstates, each of which is defined in a massively-dimensional space, and so instead this calculation must be approximated. There are many methods which calculate the free energy approximately, and a number of these will be presented below.

3.5.1 Diagrammatic perturbation theory

For Hamiltonians which are only weakly anharmonic, it is possible to calculate the anharmonic corrections to the free energy and other quantities using diagrammatic perturbation theory [1, 45].

As the Hamiltonian becomes increasingly anharmonic, diagrammatic perturbation theory requires the calculation of increasing numbers of diagrams in order to remain accurate. This makes the method increasingly intractable, both in terms of computational cost and in terms of the difficulty of implementing software which can calculate all of the diagrams.

In the strongly anharmonic limit, where the anharmonic correction to the free energy is comparable to the free energy calculated under the harmonic approximation, a point is reached where the coefficients of the higher-order diagrams no longer tend to zero sufficiently rapidly for the method to converge, and the whole method fails [29, 45].

3.5.2 Variational methods

The free energy F of a system is entirely defined by the Born-Oppenheimer Hamiltonian \mathcal{H} , an orthonormal basis of states $\{|\psi\rangle\}$, and the corresponding thermal occupations of these states $\{P_\psi\}$ [75],

$$F[\mathcal{H}, \{|\psi\rangle, P_\psi\}] = \sum_{\psi} P_\psi \langle \psi | \mathcal{H} | \psi \rangle + T \sum_{\psi} P_\psi \ln(P_\psi). \quad (3.28)$$

For a given Hamiltonian, the equilibrium free energy will be that which minimises F [29],

$$F[\mathcal{H}] = \min_{\{|\psi\rangle, P_\psi\}} F[\mathcal{H}, \{|\psi\rangle, P_\psi\}]. \quad (3.29)$$

If this minimisation is performed without constraints on $\{|\psi\rangle, P_\psi\}$ beyond $\{|\psi\rangle\}$ being a valid basis then the states which minimise F will be those which diagonalise the Hamiltonian and satisfy the Schrödinger equation [75],

$$\mathcal{H} |\psi\rangle = E_\psi |\psi\rangle, \quad (3.30)$$

and the corresponding thermal occupations of these states will be the standard Boltzmann factors of the energies E_ψ ,

$$P_\psi = \frac{e^{-E_\psi/T}}{\sum_{\phi} e^{-E_\phi/T}}. \quad (3.31)$$

In practice, for systems with many degrees of freedom it is infeasible to perform a global minimisation in

the space of all $|\psi\rangle$ and P_ψ . The *variational methods* instead perform a constrained minimisation, first applying constraints to $\{|\psi\rangle, P_\psi\}$ and only then minimising the free energy within these constraints. Since the problem is variational, this will lead to the calculation of an upper bound to the true value of $F[\mathcal{H}]$. In general, the looser the constraints placed on $\{|\psi\rangle, P_\psi\}$, the more accurate the method and the closer the upper bound will be to the true value of F , but the more computationally expensive the method [104].

Auxilliary Hamiltonians

Working with $\{|\psi\rangle, P_\psi\}$ directly can be difficult due to the constraint that $\{|\psi\rangle\}$ must be an orthonormal basis and the simple fact that each $|\psi\rangle$ is a high-dimensional object which can be computationally expensive to store and manipulate. These difficulties can be somewhat circumvented by introducing the auxilliary Hamiltonian \mathcal{H}' . The states $\{|\psi\rangle\}$ are then chosen to be those states which diagonalise \mathcal{H}' ,

$$\mathcal{H}' |\psi\rangle = E'_\psi |\psi\rangle, \quad (3.32)$$

and the corresponding thermal occupations are the Boltzmann factors of the auxilliary energies E'_ψ ,

$$P_\psi = \frac{e^{-E'_\psi/T}}{\sum_\psi e^{-E'_\psi/T}}. \quad (3.33)$$

Since $\{|\psi\rangle, P_\psi\}$ is uniquely defined by \mathcal{H}' , the free energy minimisation can be re-cast as

$$F[\mathcal{H}] = \min_{\mathcal{H}'} F[\mathcal{H}, \mathcal{H}']. \quad (3.34)$$

Instead of placing constraints on $\{|\psi\rangle, P_\psi\}$, the constraints are placed on \mathcal{H}' , where they are often more intuitive. In order for the method to be useful, the auxilliary Hamiltonian is always constructed such that the states which diagonalise it can be calculated at a reasonable cost. Additionally, it is generally more computationally tractable to parameterise \mathcal{H}' than $\{|\psi\rangle, P_\psi\}$.

It should be noted that the inclusion of an auxilliary Hamiltonian implicitly introduces the constraint that $\{|\psi\rangle, P_\psi\}$ diagonalises a valid \mathcal{H}' . Although every \mathcal{H}' corresponds to a single $\{|\psi\rangle, P_\psi\}$, there may be values of $\{|\psi\rangle, P_\psi\}$ which correspond to multiple or no values of \mathcal{H}' . This is particularly relevant at zero temperature, where the discontinuity of $\{P_\psi\}$ with respect to $\{E_\psi\}$ means that many values of $\{|\psi\rangle, P_\psi\}$ have no corresponding \mathcal{H}' .

Varying free energy with respect to the auxilliary Hamiltonian

Once the free energy has been parameterised in terms of the auxilliary Hamiltonian, it must be minimised with respect to the free parameters of that Hamiltonian. This is done by calculating the derivative of the free energy with respect to the auxilliary Hamiltonian, $\frac{\delta F[\mathcal{H}, \mathcal{H}']}{\delta \mathcal{H}'}$, which can be calculated using perturbation theory.

Under first-order perturbation theory, the derivatives of the eigenstate energies are [79],

$$\delta E'_\psi = \langle \psi | \delta \mathcal{H}' | \psi \rangle \quad (3.35)$$

and the derivatives of the eigenstates themselves are

$$\delta |\psi\rangle = \sum_{\phi \neq \psi} \frac{\langle \phi | \delta \mathcal{H}' | \psi \rangle}{E'_\psi - E'_\phi} |\phi\rangle. \quad (3.36)$$

It follows that the derivatives of the thermal weightings are

$$\delta P_\psi = P_\psi \frac{\langle \delta \mathcal{H}' \rangle - \langle \psi | \delta \mathcal{H}' | \psi \rangle}{T}, \quad (3.37)$$

where $\langle X \rangle = \sum_\psi P_\psi \langle \psi | X | \psi \rangle$ is the thermal expectation of an operator.

Taking the derivative of the free energy as written in equation 3.28 gives

$$\begin{aligned} \delta F &= \sum_\psi \delta P_\psi (\langle \psi | \mathcal{H} | \psi \rangle + T \ln(P_\psi)) + P_\psi ((\delta \langle \psi |) \mathcal{H} | \psi \rangle + \langle \psi | \mathcal{H} (\delta | \psi \rangle)) + T \delta P_\psi \\ &= \sum_\psi P_\psi \left[\left(\frac{\langle \delta \mathcal{H}' \rangle - \langle \psi | \delta \mathcal{H}' | \psi \rangle}{T} \right) (\langle \psi | \mathcal{H} | \psi \rangle + T \ln(P_\psi)) + \left(\sum_{\phi \neq \psi} \frac{\langle \psi | \mathcal{H} | \phi \rangle \langle \phi | \delta \mathcal{H}' | \psi \rangle}{E'_\psi - E'_\phi} + \text{c.c.} \right) \right]. \end{aligned} \quad (3.38)$$

This can be used to minimise the free energy. As expected, this derivative is zero when $\mathcal{H}' = \mathcal{H}$.

3.5.3 Effective harmonic methods

Effective harmonic methods are a class of variational methods which limit the auxilliary Hamiltonian to being a harmonic potential. Unlike the harmonic approximation, the static-lattice co-ordinate $\mathbf{r}^{(0)}$, normal-mode vectors $\{u_\chi\}$ and frequencies $\{\omega_\chi\}$ are all treated as variational parameters, and the free energy is minimised with respect to these parameters [29, 105, 106].

With an effective harmonic auxiliary Hamiltonian \mathcal{H}' , the free energy can be found as [29]

$$F[\mathcal{H}, \mathcal{H}'] = F[\mathcal{H}'] + \langle \mathcal{H} \rangle' - \langle \mathcal{H}' \rangle' , \quad (3.39)$$

where $\langle X \rangle'$ refers to the expectation value of operator X with respect to the states of the \mathcal{H}' Hamiltonian. Since the auxiliary Hamiltonian \mathcal{H}' has a harmonic form, its eigenstates can be calculated analytically, and so $F[\mathcal{H}, \mathcal{H}']$ and $\frac{\delta F}{\delta \mathcal{H}'}$ can be calculated analytically.

There are a number of effective harmonic methods, with many different names in the literature [1, 48]. The most notable of these are the *quasi-harmonic approximation* and the *self-consistent harmonic approximation*. The methods are distinguished by the details of how the PES is represented, how the effective harmonic model is calculated, and by whether or not any assumptions are made about the variational parameters of the model.

The most general case of the effective harmonic method has been shown to be equivalent to diagrammatic perturbation theory using a specific subset of phonon-phonon interaction diagrams [45]. This means that, while effective harmonic theories are not subject to the same unfavourable cost increases with increasing anharmonicity, they are still bound by the fundamental limitations of perturbative methods, and will fail in the case where the anharmonic correction to the free energy is comparable to or larger than the free energy calculated under the harmonic approximation.

However, while effective harmonic methods may fail to accurately estimate the free energy, they are still variational methods, and as such they will still provide a reliable upper bound on the free energy, and this upper bound will be at least as accurate as, and generally significantly more accurate than, the upper bound provided by the wavefunctions calculated under the harmonic approximation.

3.5.4 Vibrational self-consistent field theory

For many systems, it is excessively limiting to constrain the auxiliary Hamiltonian to harmonic forms, and so it is desirable to find weaker constraints which are still computationally tractable. Many of the vibrational methods which go beyond the effective Harmonic methods derive these constraints as vibrational analogues of electronic structure methods, particularly the variational methods of quantum chemistry [21].

The simplest of these methods is to constrain the auxiliary Hamiltonian to being a sum of single-mode Hamiltonians,

$$\mathcal{H}'(\mathbf{u}) = \sum_{\chi} \mathcal{H}'_{\chi}(u_{\chi}). \quad (3.40)$$

This is known as *vibrational self-consistent field theory* (VSCF), and is analogous to Hartree-Fock theory

from quantum chemistry [21, 23, 107].

The eigenstates of the VSCF Hamiltonian are products of single-mode states [21, 23],

$$|\psi\rangle = \prod_{\chi} |\psi_{\chi}\rangle, \quad (3.41)$$

where the single-mode states diagonalise the corresponding single-mode Hamiltonian,

$$\mathcal{H}'_{\chi} |\psi_{\chi}\rangle = E'_{\chi} |\psi_{\chi}\rangle. \quad (3.42)$$

The energy corresponding to each product state is then simply the sum of the energies corresponding to each single-mode state in the product,

$$\mathcal{H}' \prod_{\chi} |\psi_{\chi}\rangle = \left(\sum_{\chi} E'_{\psi_{\chi}} \right) \prod_{\chi} |\psi_{\chi}\rangle. \quad (3.43)$$

Along with many of the other methods described here, VSCF was originally developed for simulating small molecules [108–111], and was only later adapted for crystal simulations [26].

The vibrational self-consistent field theory equations

The free energy $F[\mathcal{H}, \mathcal{H}']$ is minimised with respect to the auxilliary VSCF Hamiltonian \mathcal{H}' if and only if it is minimised with respect to variations in each single-mode Hamiltonian \mathcal{H}'_{χ} individually. As such, it is only necessary to consider variations $\delta\mathcal{H}'$ which vary a single \mathcal{H}'_{χ} at a time.

For such a variation, equation 3.38 becomes [23]

$$\delta F = \sum_{\psi_{\chi}} P_{\psi_{\chi}} \left[\left(\frac{\langle \delta\mathcal{H}' \rangle - \langle \psi_{\chi} | \delta\mathcal{H}' | \psi_{\chi} \rangle}{T} \right) \left(\langle \psi_{\chi} | \mathcal{H}_{\chi} | \psi_{\chi} \rangle + T \ln(P_{\psi_{\chi}}) \right) + \left(\sum_{\phi_{\chi} \neq \psi_{\chi}} \frac{\langle \psi_{\chi} | \mathcal{H}_{\chi} | \phi_{\chi} \rangle \langle \phi_{\chi} | \delta\mathcal{H}' | \psi_{\chi} \rangle}{E'_{\psi_{\chi}} - E'_{\phi_{\chi}}} + \text{c.c.} \right) \right], \quad (3.44)$$

where \mathcal{H}_{χ} is the single-mode Hamiltonian constructed as the thermal expectation of \mathcal{H} taken with respect to all modes apart from u_{χ} ,

$$\mathcal{H}_{\chi} = \langle \mathcal{H} \rangle_{\{\chi' \neq \chi\}} + c_{\chi}, \quad (3.45)$$

where c_{χ} is an arbitrary constant offset, and where the the expectation value of an operator X taken with respect to a given mode χ is given by

$$\langle X \rangle_{\chi} = \sum_{\psi_{\chi}} \langle \psi_{\chi} | P_{\psi_{\chi}} X | \psi_{\chi} \rangle. \quad (3.46)$$

Equation 3.44 yields $\delta F = 0$ when $\mathcal{H}'_\chi = \mathcal{H}_\chi$, for some value of the constant offset c_χ .

Setting $\mathcal{H}'_\chi = \mathcal{H}_\chi$, this gives two equations which must be solved self-consistently [23]: equation 3.45 defines the VSCF Hamiltonian \mathcal{H}' in terms of the full Hamiltonian \mathcal{H} and the set of states and thermal weights $\{|\psi\rangle, P_\psi\}$; and the Schrödinger equation and Boltzmann distribution define the set of states and thermal weights $\{|\psi\rangle, P_\psi\}$ in terms of the VSCF Hamiltonian \mathcal{H}' .

The constant term in the vibrational self-consistent field theory Hamiltonian

The previous section defines the VSCF Hamiltonian up to a constant offset. It is convenient to define this offset such that the expectation of the VSCF Hamiltonian, $\langle \mathcal{H}' \rangle$, is equal to the expectation of the full Hamiltonian, $\langle \mathcal{H} \rangle$ [21].

The expectation of each single-mode term in the VSCF Hamiltonian, $\langle \mathcal{H}'_\chi \rangle$ is equal to $\langle \mathcal{H} \rangle$. As such, the VSCF Hamiltonian can be made to have the same expectation as the full Hamiltonian by defining it as

$$\mathcal{H}^{\text{VSCF}} = \langle \mathcal{H} \rangle + \sum_\chi \left(\langle \mathcal{H} \rangle_{\{\chi' \neq \chi\}} - \langle \mathcal{H} \rangle \right). \quad (3.47)$$

Temperature-dependent vibrational self-consistent field theory

Several previous implementations of VSCF calculate the single-mode Hamiltonians using the ground-state expectation [21, 23]

$$\langle \mathcal{H} \rangle_\chi = \langle 0_\chi | \mathcal{H} | 0_\chi \rangle, \quad (3.48)$$

where $|0_\chi\rangle$ is the ground state along the χ normal mode. Thermodynamic quantities may be calculated from ground-state VSCF using a method known as virtual VSCF [112]

This ground-state VSCF differs from the temperature-dependent VSCF described in this work, which uses thermally weighted expectation values [113],

$$\langle \mathcal{H} \rangle_\chi = \sum_{\psi_\chi} e^{-\beta E_{\psi_\chi}} \langle \psi_\chi | \mathcal{H} | \psi_\chi \rangle \bigg/ \sum_{\psi_\chi} e^{-\beta E_{\psi_\chi}}. \quad (3.49)$$

Ground-state VSCF is computationally faster than the temperature-dependent VSCF because the VSCF Hamiltonian is then independent of temperature and so only needs to be calculated once. When this calculation is computationally expensive it is impractical to calculate the VSCF Hamiltonian more than once. However, with the accelerations to the VSCF method described in chapter 5 it is unnecessary to make

this cost saving, and in general the cost of the complete free energy calculation will depend negligibly on the calculation of the VSCF Hamiltonian, but rather will be dominated by electronic structure calculations required to map the PES.

It is preferable to use temperature-dependent VSCF when its cost is not prohibitory. Temperature-dependent VSCF is the formulation used in the derivation of VSCF as a variational method, and ground-state VSCF is only an approximation to this. As such, ground-state VSCF will yield less accurate free energies at each temperature, and the notion of temperature itself is muddled by not being applied consistently. Further, the mixing of zero-temperature expectation values with finite-temperature free-energy calculations makes it unclear how to interpolate the resulting Hamiltonian onto a larger supercell, making it difficult to exploit size consistency as detailed in section 5.2.

Advantages and disadvantages of vibrational self-consistent field theory

The calculation of free energies under VSCF is significantly more accurate than under the harmonic approximation [21] or by effective harmonic methods [113]. VSCF also overcomes a number of the limitations of simpler methods, including a lack of temperature dependence and the inability to represent configurations which are saddle-points of the PES.

VSCF is also inexpensive when compared to many other anharmonic methods, allowing it to be used for calculations involving large systems [23]. For most systems, the computational requirements of the VSCF method will be negligible compared to the requirements of mapping the PES [26].

VSCF still has several limitations however. Most notably, no account is made of correlations between different modes, and this limits the accuracy of the resulting calculations [21]. Additionally, the standard formulation of VSCF does not respect the symmetry of the system, and does not have the correct scaling with the size of the system. These issues will be addressed in sections 5.1 and 5.2 respectively.

The accuracy of VSCF is also strongly dependent on the choice of co-ordinate system. For example, constructing VSCF using normal-mode co-ordinates is a poor fit for systems with soft torsional degrees of freedom [23]. Further, constructing VSCF using system-specific co-ordinates adds significant complexity and computational cost to the method [23].

3.5.5 Vibrational Møller-Plesset theory

The second method for calculating free energy which can be lifted directly from electronic structure theory is Møller-Plesset perturbation theory [21, 79, 114]. This method starts by constructing the Hamiltonian as

a sum of a reference Hamiltonian which can be solved exactly, $\mathcal{H}^{(0)}$, plus a perturbation Hamiltonian $\mathcal{H}^{(1)}$ multiplied by a scalar parameter λ ,

$$\mathcal{H} = \mathcal{H}^{(0)} + \lambda\mathcal{H}^{(1)}. \quad (3.50)$$

Treating λ as a small parameter, the eigenstates and corresponding energies are expanded as series in terms of increasing powers of lambda,

$$|\psi\rangle = \sum_{i=0}^n \lambda^i |\psi^{(i)}\rangle \quad (3.51)$$

and

$$E_\psi = \sum_{i=0}^n \lambda^i E_\psi^{(i)}. \quad (3.52)$$

The Schrödinger equation can then be separated into several equations, each corresponding to a given power of λ . The first few equations are

$$\begin{aligned} \mathcal{H}^{(0)} |\psi^{(0)}\rangle &= E^{(0)} |\psi^{(0)}\rangle \\ \mathcal{H}^{(0)} |\psi^{(1)}\rangle + \mathcal{H}^{(1)} |\psi^{(0)}\rangle &= E^{(0)} |\psi^{(1)}\rangle + E^{(1)} |\psi^{(0)}\rangle \\ \mathcal{H}^{(0)} |\psi^{(2)}\rangle + \mathcal{H}^{(1)} |\psi^{(1)}\rangle &= E^{(0)} |\psi^{(2)}\rangle + E^{(1)} |\psi^{(1)}\rangle + E^{(2)} |\psi^{(0)}\rangle. \end{aligned} \quad (3.53)$$

These equations are solved in order of increasing power of λ , up to some power n , which is typically a small even number [21].

In the case where the reference Hamiltonian has degenerate or near-degenerate eigenstates, Møller-Plesset theory must be augmented with degenerate perturbation theory. This adds complications to the method, but does not dramatically increase its computational cost. The application of such theory is now routine [23, 115].

The reference Hamiltonian can take many forms, but it is common practice to use the VSCF Hamiltonian. With this choice of reference Hamiltonian, the theory has several names in the literature, including cc-VSCF and VSCF-PT2 [23], but the nomenclature which is most consistent with other methods is *n*th-order *Vibrational Møller-Plesset theory* (VMP*n*) [107].

If the full PES is available, VMP*n* generally offers improved accuracy over VSCF, at little additional computational cost. This accuracy gain tends to decrease as the number of modes in the system increases [23]. However, as described in section 5.1.3, there are ways of accelerating a VSCF calculation in which a number of terms in the PES are neglected, including terms which are relevant when calculating VMP*n*. It will then vary on a system-by-system basis as to whether the cost of mapping these additional PES terms is worth the increased accuracy afforded by VMP*n*.

As the order, n , of the VMP n calculation increases, the cost and complexity of the calculation increases. However, this ever-increasing cost is not in general matched by an ever-increasing improvement in accuracy [21]. This means that VMP n should only ever be used for small n , and that the methods have a maximum accuracy which is computationally tractable. If additional accuracy is required, other methods must be used.

3.5.6 Vibrational configuration interaction

A further method drawn from quantum chemistry is configuration interaction theory, which becomes *vibrational configuration interaction* (VCI) in the context of vibrational modelling [21, 107].

The VCI ground state is parameterised as a linear combination of basis states [21],

$$|0^{\text{VCI}}\rangle = \sum_{\psi} c_{\psi} |\psi\rangle . \tag{3.54}$$

The basis states $\{|\psi\rangle\}$ are typically drawn from the eigenstates of a harmonic, effective harmonic or VSCF Hamiltonian. These eigenstates are product states of single-mode eigenstates,

$$|\psi\rangle = |\{n_{\chi}\}\rangle = \prod_{\chi} |i_{\chi}\rangle , \tag{3.55}$$

where each index n_{χ} is a non-negative integer labelling the occupation of the χ mode. Each of these basis states has a total excitation, $n_{\psi} = \sum_{\chi} n_{\chi}$, which defines how many times the state is excited above the ground state.

VCI truncates the basis of states by only including states which satisfy $n_{\psi} \leq n_{\text{max}}$. This is an approximation, but in the limit as $n_{\text{max}} \rightarrow \infty$ this approximation becomes exact. This makes VCI systematically improvable, meaning that it is theoretically possible to achieve arbitrary accuracy by increasing n_{max} .

Vibrational states other than the ground state can also be calculated, allowing for finite-temperature calculations, although calculating these states to a given accuracy generally requires the use of more basis states than required by the calculation of the ground state alone. This in turn increases the cost of the method, and in general the cost will increase rapidly with temperature, as the calculation of more and more eigenstates is needed.

Additionally, VCI becomes computationally intractable very rapidly with increasing system size and as n_{max} is increased [23, 107, 116].

3.5.7 Vibrational coupled cluster

The most sophisticated method which is borrowed from quantum chemistry is coupled cluster theory, which becomes *vibrational coupled-cluster theory* (VCC) in the context of vibrational modelling [21].

The formulation of VCC is based on a second-quantization of the vibrational Hamiltonian [116]. Like VCI, VCC starts with a basis of states which are usually drawn from the eigenstates of an effective harmonic method or VSCF. Unlike VCI, where the the ground state is a linear combination of basis states, the VCC ground state $|0^{\text{VCC}}\rangle$ is parameterised in terms of the basis ground state $|0^{\text{basis}}\rangle$ as [116]

$$|0^{\text{VCC}}\rangle = e^X |0^{\text{basis}}\rangle, \tag{3.56}$$

where X is the *cluster operator*, defined as a sum of excitation operators [116]

$$X = \sum_{i=0}^{n_{\text{max}}} \sum_{\{\chi, n\}_i} c_{\{\chi, n\}_i} \prod_{j=1}^i E_{n_j}^{\chi_j}, \tag{3.57}$$

where each E_n^χ is the second-quantized excitation operator which excites the ground state of mode χ into the state $|n\rangle$,

$$E_n^\chi = |n_\chi\rangle \langle 0_\chi|. \tag{3.58}$$

Each set $\{\chi, n\}_i$ contains i distinct modes $\{\chi\}$ and i excitations $\{n\}$. The variables $c_{\{\chi, n\}_i}$ are scalar coefficients, and it is these which are varied when finding the ground state which minimises the free energy. The non-negative integer n_{max} limits the size of the basis, just as for VCI.

Just like VCI, VCC is systematically improvable, hypothetically allowing for calculations with arbitrary accuracy. In general, VCC is less computationally intensive for a given level of accuracy than VCI [21].

Unlike VCI, the non-linear parameterisation of VCC makes it difficult to calculate states other than the ground state, making it impractical for finite-temperature calculations [117]. The computational cost of VCC also scales very unfavourably with system size, and with the number of states included in the reference basis.

3.5.8 Molecular dynamics

Classical molecular dynamics

Molecular dynamics (MD) methods work by building up a set of representative samples of the PES, and using the energies of these samples to calculate the free energy directly [118].

Each MD sample is defined by a co-ordinate \mathbf{r} and a classical momentum \mathbf{p} . This momentum has no effect on the measurement of the PES $V(\mathbf{r})$, but is important for the calculation of subsequent samples and for the calculation of the kinetic energy.

In the simplest form of MD, the set of MD samples $\{(\mathbf{r}, \mathbf{p})_i\}$ are built up one at a time, with each sample being calculated from the last using classical equations of motion,

$$\dot{\mathbf{p}} = -\nabla V \quad (3.59)$$

and

$$\dot{\mathbf{r}} = M^{-1}\mathbf{p}, \quad (3.60)$$

where M is the mass tensor, which relates the classical momentum to the classical velocity as $\mathbf{p} = M\dot{\mathbf{r}}$.

There are many ways to integrate these equations, each of which has its own cost, accuracy, and rate of convergence [119].

The expectation values of the potential and kinetic energies can then be calculated directly from the samples, as

$$\langle V \rangle = \frac{1}{N} \sum_i V(\mathbf{r}_i) \quad (3.61)$$

and

$$\langle T \rangle = \frac{1}{N} \sum_i \frac{1}{2} \mathbf{p}_i \cdot M^{-1} \cdot \mathbf{p}_i. \quad (3.62)$$

Once sufficient samples have been taken, these values converge to their equilibrium values [120].

The algorithm described here is for simulations at constant energy and constant volume. In order to instead run calculations at constant temperature or constant pressure, a thermostat or barostat must be included in the sample generation, as detailed elsewhere [121, 122].

MD does not require an explicit representation of the PES, and does not require the system to be bound to a small region about the static-lattice configuration. Instead, all anharmonic effects are naturally included by the method [123]. This makes it a better choice than vibrational methods for systems which explore large regions of configuration space [123]. MD is also a fundamentally finite-temperature method, requiring no extra work to run calculations as the temperature increases [123].

However, MD is based on classical mechanics, not quantum mechanics, and so is unable to account for zero-point energy and other quantum effects. This makes it fundamentally inaccurate for many systems where these effects are relevant, and for any system under its Debye temperature [48, 118].

Additionally, MD can require many PES samples to converge, making it more computationally expensive than other methods of similar accuracy for small systems [107, 123], although the cost scaling with system size is more favourable than that of many quantum chemistry methods. The cost of MD is also variable, with the number of samples needed depending on the exact details of the PES.

Further, the lack of explicit PES representation means there is not enough information available to calculate some quantities of interest. In particular, some scattering calculations depend on mode-resolved state information, and this is not available from MD simulations [107].

Thermodynamic integration

Thermodynamic integration is a method for calculating the free energy difference between two different systems, by smoothly transforming the Hamiltonian from one system to the other.

Starting from a reference Hamiltonian $\mathcal{H}^{(0)}$ with known free energy $F^{(0)}$, thermodynamic integration finds the free energy $F^{(1)}$ of the full Hamiltonian $\mathcal{H}^{(1)}$ by constructing a series of intermediate Hamiltonians.

These intermediate Hamiltonians are parameterised as [22],

$$\mathcal{H}^{(\lambda)} = (1 - \lambda)\mathcal{H}^{(0)} + \lambda\mathcal{H}^{(1)}, \quad (3.63)$$

where λ is a real scalar parameter satisfying $0 \leq \lambda \leq 1$. When λ is equal to 0 or 1, this Hamiltonian is equal to $\mathcal{H}^{(0)}$ or $\mathcal{H}^{(1)}$ respectively, and by varying λ smoothly from 0 to 1 the Hamiltonian transforms smoothly from $\mathcal{H}^{(0)}$ to $\mathcal{H}^{(1)}$.

Thermodynamic integration gives an expression for $F^{(1)} - F^{(0)}$ in terms of a thermodynamic integral [22],

$$F^{(1)} - F^{(0)} = \int_0^1 d\lambda \left\langle \mathcal{H}^{(\lambda)} - \mathcal{H}^{(0)} \right\rangle_{\mathcal{H}^{(\lambda)}}, \quad (3.64)$$

where the expectation value $\langle X \rangle_{\mathcal{H}^{(\lambda)}}$ refers to the expectation of operator X with respect to the eigenstates of the Hamiltonian $\mathcal{H}^{(\lambda)}$.

Path-integral molecular dynamics

Path-Integral Molecular Dynamics (PIMD) calculates the vibrational free energy by first calculating the free energy of an analytically soluble model, typically the harmonic approximation, and then using thermodynamic integration to calculate the difference between the soluble model and a fully anharmonic model.

The thermodynamic integral is performed numerically, by using MD to calculate the expectation value

$\langle \mathcal{H}^{(\lambda)} - \mathcal{H}^{(0)} \rangle_{\mathcal{H}^{(\lambda)}}$ at a number of discrete values of λ . For a sufficiently dense discretization this converges to the true value [22].

PIMD shares many of the advantages and disadvantages of classical molecular dynamics; there is no need to represent or approximate the PES, finite-temperature effects are included automatically, and the method is computationally demanding for small systems but this cost scales favourably with system size [22, 122]. The main difference between the two is that PIMD includes quantum effects including zero-point motion, meaning that it is not limited to systems which behave classically. This comes at a significant increase in computational cost [1, 48, 124].

Just like VCI and VCC, PIMD is extremely accurate, and this accuracy is systematically improvable. For a given level of accuracy, PIMD is generally more expensive than VCI and VCC for small systems, but should be less expensive for larger systems [22].

Ring-polymer molecular dynamics

Another method for calculating the free energy based on MD is *Ring-polymer molecular dynamics* (RPMD) [125].

RPMD replaces the quantum simulation of a system which has the Hamiltonian

$$\mathcal{H} = \mathcal{T} + V(\mathbf{u}) \tag{3.65}$$

with a classical simulation of a system containing N copies of the original system arranged in a ring, where each system is connected to the systems on either side of it in the ring by a harmonic spring. This gives a Hamiltonian [126]

$$\mathcal{H} = \sum_{i=1}^N \left(\mathcal{T}_i + V_i(\mathbf{u}_i) + T^2 N^2 |\mathbf{u}_i - \mathbf{u}_{i+1}|^2 \right). \tag{3.66}$$

Since the systems are arranged in a ring, the system with $i = N + 1$ is equivalent to the system with $i = 1$.

This substitution is possible because in the limit $N \rightarrow \infty$ the partition function of the classical ring polymer becomes that of the quantum system, meaning that many of the observables of the quantum system can be calculated using the ring polymer [126, 127]. In practice, reasonable accuracy can be reached by using a finite N , typically on the order of thirty, although this must be converged for each system [127].

The computational cost of RPMD is simply N times that of classical MD. Just like PIMD, RPMD has systematically improvable accuracy.

3.5.9 Monte-Carlo sampling

Like MD methods, *Monte-Carlo sampling* (MC) methods work by building up a set of PES samples which cover the PES in a manner such that the density of samples is proportional to the vibrational density. This means that the expectation value of any observable can be calculated directly as the sum of the values of that operator at each sample.

Instead of using dynamical processes to sample the PES, MC methods use stochastic sampling, where each new sample is generated from previous samples in such a way that the low-energy regions of the PES are sampled more thoroughly than high-energy regions.

Just as MD can include quantum effects through PIMD, MC can include quantum effects through *path integral Monte-Carlo* (PIMC). PIMC proceeds identically to PIMD, except that the expectation values at each value of λ are calculated using MC methods rather than MD [59]. Just like PIMD, PIMC is systematically improvable, and has a computational cost which is larger than that of VCI and VCC for small systems but smaller for large systems.

Diffusion Monte-Carlo

The *Diffusion Monte-Carlo* (DMC) method takes advantage of the similarity between the Schrödinger equation and the diffusion equation [128].

The time-dependent Schrödinger equation for an arbitrary state $|\Psi\rangle$ which is a superposition of the eigenstates of the Hamiltonian $|\psi\rangle$ with energies E_ψ is [129]

$$|\Psi\rangle = \sum_{\psi} c_{\psi} e^{-iE_{\psi}t} |\psi\rangle, \quad (3.67)$$

where each c_{ψ} is a complex scalar coefficient, and t is the time.

If the time t is replaced with imaginary time $\tau = it$ then the Schrödinger equation becomes [129]

$$|\Psi\rangle = \sum_{\psi} c_{\psi} e^{-E_{\psi}\tau} |\psi\rangle. \quad (3.68)$$

Thus, advancing forwards in imaginary time will cause any state to transform towards the ground state of the system.

DMC takes advantage of this by taking a set of PES samples, which are referred to in this context as *walkers*. At each step of the algorithm, the walkers are first displaced by a random vector $\delta\mathbf{r}$ which is drawn

from a Gaussian distribution. Then PES samples are taken at each walker’s position. Finally, each walker is replaced by zero or more walkers, and the next step begins. The number of walkers which replace each walker is decided stochastically, in a manner which depends on the value of the PES at that walker’s position, $V(\mathbf{r})$ [129]. If V is large then the walker is likely to simply be removed, with no walker taking its place. As V decreases, the probability of the walker being replaced by another walker increases, and as V decreases further the probability increases that the walker is replaced by more than one walker.

In this manner, the walkers diffuse around the PES, and more walkers are generated where the value of the PES is lower. With sufficient steps, the set of samples generated by the walkers converges towards a representative sampling of the ground state of the Hamiltonian [129].

While DMC is fundamentally a ground-state method, and so a zero-temperature method, it is possible to calculate excited states through the imposition of *nodal surfaces* [129], which define the sign of the wavefunction of the state in different regions of space. However, these nodal surfaces are approximations, and this can add significant limitations and inaccuracies to the method.

For zero temperature calculations, DMC offers systematically improvable accuracy, and can reach accuracies comparable to those achievable by other systematically improvable methods. The computational cost of DMC is similar to that of PIMD, with a similar cost scaling with the size of the system. PIMD is typically more expensive than DMC for a given Born-von Karman supercell, but DMC typically requires a larger Born-von Karman supercell than other methods in order to converge [2]. Which method is faster for a given system will depend on the details of the system and the implementation of the two methods [99].

3.5.10 Size consistency

Following the definition of Hirata [130], when modelling an electrically neutral, periodic solid under periodic boundary conditions, a method is size-consistent if the internal energy is proportional to the volume. This condition is important, because without size consistency a method cannot be expected to converge as the size of the Born-von Karman supercell is increased [71, 131].

The harmonic approximation and effective harmonic methods are naturally size-consistent. It is possible to construct size-consistent formulations of VSCF [14, 132, 133] and VMP2 [14, 134] by restricting the allowed terms in the auxilliary Hamiltonian. DMC can also be made size-consistent [99]. VCC is also naturally size-consistent [14, 116]. Additionally, in the limit of arbitrary accuracy the systematically-improvable methods must all be size-consistent, although this does not guarantee size consistency when this limit is not reached.

It is not apparent whether the size consistency of PIMD and related methods has been evaluated.

A major downside of VCI is that, outside of the limit of arbitrary accuracy, it cannot be constructed to be size-consistent [14, 116].

3.5.11 Beyond Born-Oppenheimer methods

Methods exist which do not assume the Born-Oppenheimer approximation. These generally take an electronic structure method and its analogous vibrational structure method, and simply combine the two as a single calculation rather than two separate calculations.

In the context of DFT, this is known as *multicomponent density-functional theory* (MCDFT) [57]. MCDFT is significantly slower than DFT, since the nuclear density is a $3N_z$ -dimensional object, where N_z is the number of nuclear species in the system. Additionally, significantly less work has been done on finding good density functionals for representing nuclei, and as a result MCDFT tends to be significantly less accurate than DFT [135].

Other methods which have been extended to go beyond the Born-Oppenheimer approximation include DMC [58], PIMC [59], and many-body Green's function theory [60]. These methods can offer extreme accuracies, but this comes at a large computational cost. As such, both are currently limited to applications to very small systems and are not appropriate for high-throughput methods.

3.5.12 Summary

Beyond the harmonic approximation, this work will focus on VSCF rather than any of the other methods presented here. This is because the aim of this work is to find a method which is more accurate than the harmonic approximation and effective harmonic methods, but which is nonetheless inexpensive enough to be used in conjunction with automated high-throughput methods. VSCF appears to be the natural choice for a method which fits this description [23].

Methods which offer greater accuracy than VSCF will not be considered in this work due to their computational cost. However, in section 5.2 it will be demonstrated that VSCF may have flaws which make it unsuitable for the task at hand. In section 8.3, possible alternatives to VSCF which do not share these flaws will be considered for future work.

3.6 Calculating other properties

Once the vibrational free energy of a crystal has been calculated, it is relatively straightforward to calculate many other properties of interest of the crystal, including many of the properties listed in section 1.4. Some properties can be calculated directly from the electronic and vibrational wavefunctions. Other properties require the mapping of a *property surface*, which is a function of \mathbf{u} analogous to the PES, but for a variable other than the potential energy. These surfaces can generally be mapped in exactly the same manner as the PES is mapped. Once they are mapped, their expectation can be taken with respect to the vibrational wavefunctions, which generally requires very little additional calculation.

3.6.1 Gibbs free energy and stress

In order to simulate a crystal at constant pressure, rather than at constant volume, it is necessary to calculate the Gibbs free energy G . Calculating G requires the calculation of the free energy F and the stress tensor σ . The stress tensor σ can either be calculated directly as part of the vibrational calculation, or it can be calculated from the derivatives of the free energy.

The free energy F and the stress tensor σ are calculated as functions of the lattice matrix L and temperature T , as $F(L, T)$ and $\sigma(L, T)$ respectively. When working at constant isotropic pressure, it is necessary to find the values of L for which the stress tensor σ is a scalar multiple of the identity matrix. The pressure p can then be calculated as $p = -\frac{1}{3} \text{Tr}[\sigma]$, and finally the Gibbs free energy G can be calculated using the Legendre transform

$$G(p, T) = F(\Omega, T) + p\Omega. \quad (3.69)$$

Calculating stress from the free energy

The stress tensor σ can be calculated in terms of the derivatives of the free energy as [26, 48]

$$\sigma = - \left(\frac{\partial}{\partial \varepsilon} \right)_T \frac{F}{\Omega}, \quad (3.70)$$

where ε is the strain tensor, Ω is the volume of the primitive cell, equal to the determinant of the lattice matrix $|L|$, and $\left(\frac{\partial}{\partial \varepsilon} \right)_T$ is the tensor derivative with respect to the strain tensor at constant temperature. The strain tensor can be expressed in terms of the lattice matrix, to give

$$\sigma = -\frac{1}{2} \left[L^T \cdot \left(\frac{\partial}{\partial L} \right)_T \frac{F}{\Omega} + \text{transpose} \right]. \quad (3.71)$$

There are two methods of calculating the derivative $(\frac{\partial}{\partial L})_T \frac{F}{\Omega}$. The first method is to calculate the free energy F at a number of closely-spaced values of L , and to use finite-difference equations to calculate the derivative.

The second method is to fit the free energy as a function of the lattice matrix, $F(L)$, using the *Birch-Murnaghan equation of state* [136]. The Birch-Murnaghan equation of state expresses the free energy as a Taylor series in the *quadratic Euler strain tensor* E .

The quadratic Euler strain tensor describes how the length of a vector changes as the lattice matrix changes. Consider two copies of the same crystal, but with different lattice matrices, L_0 and L . A vector which has the same fractional co-ordinates in both crystals will have the cartesian co-ordinates \mathbf{r}_0 in the first crystal and \mathbf{r} in the second. The quadratic Euler strain tensor is then defined as

$$\mathbf{r} \cdot E \cdot \mathbf{r} = \frac{1}{2} \left(|\mathbf{r}|^2 - |\mathbf{r}_0|^2 \right). \quad (3.72)$$

This quantity is independent of the vector \mathbf{r} , and can be expressed in terms of the lattice matrix L and a reference lattice matrix L_0 as

$$E(L) = \frac{1}{2} \left(\mathbb{1} - L^{-1} \cdot L_0 \cdot L_0^T \cdot L^{-T} \right). \quad (3.73)$$

Once the Birch-Murnaghan equation of state has been fitted, the derivative $(\frac{\partial}{\partial L})_T \frac{F}{\Omega}$ can be calculated analytically, and this can be used to calculate the stress tensor.

Calculating stress directly

Instead of calculating the stress from the derivatives of the free energy, it is possible to calculate it directly [48]. In order to perform this calculation, the stress must first be expressed as the sum of the expectations of two operators: the *kinetic stress* $\sigma^{(T)}$ and the *potential stress* $\sigma^{(V)}$ [26].

The kinetic stress is the expectation of the operator

$$\sigma_{\alpha\beta}^{(T)} = \frac{1}{N\Omega} \sum_{\mathbf{r}_j} \frac{\partial}{\partial \mathbf{r}_{\mathbf{r}_j \alpha}} \frac{\partial}{\partial \mathbf{r}_{\mathbf{r}_j \beta}} \quad (3.74)$$

with respect to the VSCF eigenstates.

The potential stress is calculated analogously to the calculation of the free energy [101], by first fitting a stress surface $\sigma^{(V)}(\mathbf{u})$, analogous to the PES $V(\mathbf{u})$, and then calculating the expectation of this with respect to the VSCF eigenstates to give $\langle \sigma^{(V)}(\mathbf{u}) \rangle$. This is only possible if the chosen electronic structure method calculates the stress tensor in addition to the energy for each sample.

The process of fitting of the stress surface is broadly similar to the process of fitting the PES, in that first a set of symmetry-invariant basis functions are generated from monomial terms, then these are used to find an optimal set of samples, before using linear regression to fit the basis function coefficients. However, there are a couple of complications.

The first complication is that the stress surface is a tensor field. This means the stress surface has more free parameters than the PES, and so more information is required in order to fit these parameters. This also somewhat complicates the application of symmetry to the stress surface, as the action of the symmetries on the tensor components needs to be taken into account.

The second complication is that electronic structure methods do not generally yield the derivative of the potential stress tensor with respect to nuclear co-ordinates, in contrast to the PES derivatives, where forces are generally readily available. This can make fitting the stress surface significantly more computationally expensive than fitting the PES.

For these reasons, it can often be less expensive to calculate the stress tensor from several free energy calculations, and so avoid having to calculate the stress surface. This approach also avoids the inaccuracies inherent in approximating the stress surface.

Chapter 4

Mapping the potential energy surface

Mapping the PES is one of the computational bottlenecks of the entire vibrational free energy calculation, and one of the potential sources of inaccuracy. This chapter will explore a number of ways in which the process of representing and mapping the PES can be improved, both in terms of increasing the accuracy of the PES representation and minimising the computational cost of mapping the PES, by exploiting symmetry, size consistency and finite-range effects.

4.1 Symmetry and the potential energy surface

It is important to ensure that the PES of a crystal is invariant under the crystal symmetries. Failure to symmetrise the PES can lead to inaccurate calculations [137], while exploiting the symmetries of the crystal can dramatically accelerate the mapping of the PES [13].

4.1.1 Symmetry and subspaces

As detailed in section 2.4, in normal-mode co-ordinates, each symmetry acts as a unitary tensor,

$$S u_{\chi} = \sum_{\chi'} S_{\chi\chi'} u_{\chi'}. \quad (4.1)$$

Since each symmetry commutes with the crystal Hamiltonian, the elements of $S_{\chi\chi'}$ may only be non-zero if the modes χ and χ' are degenerate, i.e. their frequencies are the same.

In this manner, the normal modes of the crystal are grouped into *subspaces* of degenerate modes [73]. These

subspaces will be labelled with the index ξ . The crystal symmetry matrices within each subspace form an *irreducible representation* of the symmetry group. This means that when the modes are arranged in order of subspace, the crystal symmetry matrices are block diagonal, with non-zero elements between modes in the same subspace but no non-zero elements between modes in different subspaces.

Each subspace contains one or more degenerate modes at a given q-point \mathbf{q} , and the same number of modes at each q-point \mathbf{q}' which can be mapped to \mathbf{q} by a symmetry operator, $\mathbf{q}' \cdot S = \mathbf{q}$. The set of such q-points is *star* of \mathbf{q} . Since the matrix S is unitary, all q-points \mathbf{q}' in the star of \mathbf{q} will have $|\mathbf{q}'| = |\mathbf{q}|$. In addition, since in fractional co-ordinates all S matrices have purely integer elements, if \mathbf{q} has $n\mathbf{q} = \mathbf{G}$ for some positive integer n and some G-vector \mathbf{G} then all q-points in the star of \mathbf{q} will also have $n\mathbf{q}' = \mathbf{G}'$ for some other G-vector \mathbf{G}' , where n is the same for all q-points in the star.

As a consequence of the PES being real-valued for all real-valued displacements \mathbf{u} , every mode u_χ should be considered to be in the same subspace as its conjugate mode $u_{\bar{\chi}}$, regardless of whether or not there is a symmetry linking the two modes. This does not impose any additional restrictions on the PES.

In the infinite crystal there are an infinite number of R-vectors, and so there are an infinite number of purely translational symmetries. This means that the real-space representations of the symmetries form an infinite group. Within a subspace of normal-mode co-ordinates however, the pure translation \mathbf{R} becomes a phase change $e^{2\pi i \mathbf{q} \cdot \mathbf{R}}$. Provided that $n\mathbf{q} = \mathbf{G}$ for some finite integer n , this means that there are only a finite number of translational symmetries, and so the symmetries form a finite closed group [73].

One consequence of this is that for every symmetry operator \mathcal{S} there is some finite integer $n_{\mathcal{S}}$ which is the smallest positive integer such that $\mathcal{S}^{n_{\mathcal{S}}} = \mathbf{1}$. This is the *order* of the symmetry with respect to the given subspace.

4.1.2 Symmetry and the principal axes approximation

The PAA does not behave well under symmetry. In general, when the PES is constructed such that it is invariant under symmetry, there may be symmetries relating terms containing different numbers of normal modes. This leads to the coefficients of these terms to be the same or very similar, invalidating the PAA in its basic form.

For example, if there are two modes u_1 and u_2 which are related by a three-fold symmetry which acts as

$$\mathcal{S} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos(120^\circ) & \sin(120^\circ) \\ -\sin(120^\circ) & \cos(120^\circ) \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad (4.2)$$

then the symmetry-invariant third-order basis functions are $u_1^3 - 3u_1u_2^2$ and $u_2^3 - 3u_2u_1^2$. As such, since a

symmetry-invariant PES must contain only symmetry-invariant basis functions, the $u_1 u_2^2$ and $u_1^2 u_2$ terms cannot be neglected without also neglecting the u_1^3 and u_2^3 terms.

The PAA can be made symmetry-invariant by changing its definition. Under the old definition, taking the PAA at order n limited the terms in the PES to those containing up to n modes. Instead, taking the PAA at order n should limit the terms in the PES to those containing modes from up to n subspaces. For example, if modes u_1 and u_2 are in one subspace and mode u_3 is in another, then taking the PAA at the first order should mean the PES includes the terms $u_1^{n_1} u_2^{n_2}$ and $u_3^{n_3}$, but not $u_1^{n_1} u_3^{n_3}$ or $u_2^{n_2} u_3^{n_3}$.

Making the PAA symmetry invariant significantly increases the number of terms included in the PES for a given order of the PAA. However, there is a limit to the number of modes in a single subspace, and this limit does not change as the system size changes. This means that the number of terms in the PES scales the same with the system size regardless of whether or not the PAA is made symmetry invariant.

Splitting subspaces by q-point

While there is a constant limit to the number of modes in a subspace, this limit is still large, and this can mean that each subspace contains many dimensions. This in turn makes mapping the PES across an entire subspace impractical, even before subspace coupling is considered.

This can be avoided by noting that every symmetry \mathcal{S} transforms an arbitrary sum of modes at a given q-point \mathbf{q} onto a sum of modes at a single other q-point $\mathbf{q} \cdot \mathcal{S}$.

This means that when taking the PAA, each subspace can be divided into a set of single-q-point subspaces, each of which contains the degenerate modes at a q-point and its paired q-point.

For example, if modes u_1, u_2, u_3 and u_4 form a degenerate subspace, with modes u_1 and u_2 at \mathbf{q}_1 and modes u_3 and u_4 at \mathbf{q}_2 , such that $\mathbf{q}_2 \neq \bar{\mathbf{q}}_1$, then taking the PAA at the first order would include terms of the form $u_1^{n_1} u_2^{n_2}$ and $u_3^{n_3} u_4^{n_4}$, but not terms of the form $u_1^{n_1} u_2^{n_2} u_3^{n_3} u_4^{n_4}$.

This can dramatically reduce the number of basis functions at each order of the PAA, making calculation feasible.

4.1.3 Symmetric basis functions

If the PES is represented as a sum of basis functions,

$$V(\mathbf{u}) = \sum_i c_i \mu_i(\mathbf{u}), \quad (4.3)$$

where each c_i is a real scalar coefficient, then it is desirable to make each basis function invariant under the symmetry operators. This guarantees that the potential as a whole is invariant under the symmetry operators, and means that the coefficients can take any value without needing to reference the symmetries [13].

The construction of symmetric basis functions will be demonstrated here for polynomials in normal-mode co-ordinates.

Polynomial basis functions

The starting point for constructing polynomial basis functions is the set of normal-mode monomials,

$$\nu_j = \prod_{\chi} u_{\chi}^{n_{j\chi}}, \quad (4.4)$$

where each $n_{j\chi}$ is a non-negative integer power.

It is convenient to define the paired monomial of each monomial. Analogously to paired modes, the paired monomial of ν_j is

$$\nu_{\bar{j}} = \prod_{\chi} \bar{u}_{\chi}^{n_{j\chi}}. \quad (4.5)$$

Each basis function is then constructed as a sum of these monomials,

$$\mu_i = \sum_j c_{ij} \nu_j, \quad (4.6)$$

with complex coefficients c_{ij} .

Just as the whole PES can be made symmetry invariant by making each basis function symmetry invariant, it is possible to guarantee that the PES is real-valued for real-valued displacements by guaranteeing the same for each basis function. This is done by ensuring that the coefficient each monomial is the conjugate of the coefficient of the paired monomial, $c_{i\bar{j}} = c_{ij}^*$.

Translational symmetry

It can immediately be seen that the action of a pure translation on a monomial is

$$\mathcal{S}\nu_j = \left(\prod_{\chi} e^{2\pi i n_{j\chi} \mathbf{q}_{\chi} \cdot \mathbf{R} \mathbf{s}} \right) \nu_j. \quad (4.7)$$

Thus, the only monomials which may appear in the potential are those which are invariant under all pure translations, which must obey

$$\sum_x n_x \mathbf{q}_x = \mathbf{G} \quad (4.8)$$

for some G-vector \mathbf{G} . This condition is simply the conservation of pseudomomentum.

Subspace powers

Each monomial can be classified according to the total power of modes it contains from each subspace. For example, if modes u_1 and u_2 are in subspace 1 and mode u_3 is in subspace 2 then the monomial $u_1 u_2^2 u_3^4$ would have the *subspace power vector* (3, 4).

For each subspace power vector, the crystal symmetry matrices in the basis of monomials corresponding to that subspace power vector form an irreducible representation of the symmetry group. This means that each symmetry only maps each monomial onto other monomials with the same subspace power vector. As such, it is possible to construct a complete set of basis functions such that each basis function only contains monomials with a given subspace power vector.

Separating basis functions by subspace power vector also means that each basis function only contains monomials with a given Taylor expansion order, and with a given combination of modes. This means that the basis functions separate cleanly according to the truncation of the Taylor expansion and the order of the PAA.

Projection matrices

Given the representation of a symmetry in normal-mode co-ordinates, it is possible to construct the representation of that symmetry in the basis of monomials as a tensor product of the single-mode representations, to give

$$\mathcal{S}\nu_j = \sum_{j'} S_{jj'} \nu_{j'} . \quad (4.9)$$

In order to calculate the functions which are invariant under this operator, it is first necessary to define the projection operator,

$$\mathcal{P}(\mathcal{S}) = \sum_{i=1}^{n_S} \mathcal{S}^i . \quad (4.10)$$

This is a matrix whose eigenvalues are either 0 or 1. The eigenvectors corresponding to eigenvalues of 1 form a complete basis for the space of vectors which are invariant under the action of the symmetry.

Taking the product of all of the projection operators from each of the crystal symmetries gives the projection operator for the entire symmetry group. This matrix again has eigenvalues which are either 0 or 1. The eigenvectors corresponding to eigenvalues of 1 form a complete basis for the space of vectors which are invariant under any symmetry, and these can be used as the PES basis functions.

The basis function algorithm

The method proceeds as follows:

- The set of monomials which conserve pseudomomentum are generated up to the Taylor expansion truncation and the order of the PAA.
- These monomials are separated into sets according to their subspace power vectors.
- Within each set:
 - For monomials which are not their own pairs, superpositions are taken to give $\frac{1}{\sqrt{2}}(\nu_j \pm i\nu_{\bar{j}})$.
 - The representations of each of the symmetry generators are constructed in the basis of these superpositions of monomials.
 - The projection matrix for each generator is calculated.
 - The projection matrix for the symmetry group is calculated.
 - The projection matrix is diagonalised.
 - The eigenvectors of the projection matrix with eigenvalue 1 are the basis functions.

The space of basis functions corresponding to each subspace power vector is well defined, but individual basis functions are only defined up to a unitary transformation. Any set of basis functions which spans the correct space can be used, but it is convenient to reduce the number of monomials in each basis function. This can be done by taking an appropriate QR-decomposition of the c_{ij} matrix.

It should be noted that when splitting the subspaces by q-point, the splitting is only relevant when deciding which terms to include in the PES under the PAA. When constructing symmetric basis functions from the included terms it is in general necessary to consider entire subspaces rather than just the parts of subspaces at individual q-points.

Summary

This method directly calculates a basis of real-valued, symmetry-invariant basis functions. This method applies to Taylor series representations of the PES, but the process should easily extend to other basis

function representations of the PES.

In the context of Taylor series representations, these symmetry-invariant basis functions are known as the *space group irreducible derivatives* of the PES. These are discussed in depth by Fu et al. [13], and an alternative method for calculating these derivatives is presented in that work. The method of Fu et al. [13] will likely be faster and less reliant on numerical details than that presented here, although it is more conceptually intricate, and so it is hoped that the method here may still prove useful when extending these ideas to basis functions other than those based on Taylor series.

The method here will also be significantly faster and more reliable than the methods which prevailed before Fu et al. [13], which used each symmetry to construct a separate linear constraint on the space of monomials [45, 68].

Similar ideas can also be applied to PES representations which use co-ordinate systems other than normal-mode co-ordinates. Such methods have previously been applied to *cluster co-ordinates*, which are specifically constructed for their symmetry properties [102].

4.2 Size consistency and the potential energy surface

Ensuring that the PES is size-consistent is of similar importance to ensuring it is symmetry-invariant. If the PES is not size-consistent then the convergence of observables with the size of the Born-von Karman supercell may be very slow or may not happen at all [71]. On the other hand, size consistency can be exploited to accelerate the calculation, and to allow for interpolation between calculations performed in different supercells.

4.2.1 Interpolating the anharmonic potential

As the size of the modelled supercell increases, the free energy predicted by the model should converge to the large-supercell limit. It is possible to do this by simply calculating the free energy using a number of supercells with different sizes. However, this converges to the large-supercell limit relatively slowly.

As with the harmonic approximation, a finite supercell calculation can be interpolated to the large-supercell limit. This changes the model from a finite supercell with periodic boundary conditions to an infinite crystal but with inter-atomic interactions limited in range to those which fit within the Wigner-Seitz supercell corresponding to the finite supercell. By interpolating each supercell in this manner, convergence is improved [13].

In order to interpolate the PES from a coarse q-point grid (corresponding to a small supercell in real space) to a fine q-point grid (corresponding to a large supercell in real space), it is necessary to calculate the coefficients of the monomials on the fine grid from the coefficients of the monomials on the coarse grid. This process is a three-stage process, similar to Fourier interpolation under the harmonic approximation:

- First, the PES is transformed from the normal-mode co-ordinates of the coarse q-point grid to the displacement co-ordinates in the small supercell, using $u_\chi = \frac{1}{N} \sum_\rho U_{\rho\bar{\chi}} d_\rho$.
- Second, the PES is transformed from the displacement co-ordinates of the small supercell to the displacement co-ordinates of the large supercell, under the assumption that inter-atomic interactions are limited to the Wigner-Seitz cell of the small supercell.
- Finally, the PES is transformed from the displacement co-ordinates of the large supercell to the normal-mode co-ordinates of the fine q-point grid, using $d_\rho = \sum_\chi U_{\rho\chi} u_\chi$.

The transformations between normal-mode co-ordinates and displacement co-ordinates are well-defined multi-dimensional linear transforms. The transformation between supercell co-ordinates however is not uniquely defined, and requires an anharmonic analogue of the minimum image convention.

4.2.2 The anharmonic minimum image convention

There are a number of ways to generalise the minimum image approximation to higher-dimensional polynomials. Any such generalisation is useful provided that it is exact in the limit of the small supercell being large enough to fully contain all the relevant inter-atomic interactions.

One particularly mathematically tractable method is to take a sum of terms, where each term contains the pair-wise minimum images with respect to a single atom.

Formally, if the PES in the small supercell S is

$$V = \sum_{n=2}^{\infty} \sum_{\rho_1 \dots \rho_n} V_{\rho_1 \dots \rho_n}^{(n)} \prod_{j=1}^n d_{\rho_j} \quad (4.11)$$

and the PES in the large supercell S' is

$$V = \sum_{n=2}^{\infty} \sum_{\rho'_1 \dots \rho'_n} V_{\rho'_1 \dots \rho'_n}^{(n)} \prod_{j=1}^n d_{\rho'_j} \quad (4.12)$$

then the coefficients in the two supercells transform as

$$V_{\rho'_1 \dots \rho'_n}^{(n)} = \sum_{\substack{\rho_1 \dots \rho_n \\ \rho \in S}} \left(\frac{1}{n} \sum_{j=1}^n \prod_{\substack{k=1 \\ k \neq j}}^n \theta_{\rho'_j \rho'_k \rho_j \rho_k} \right) V_{\rho_1 \dots \rho_n}^{(n)}, \quad (4.13)$$

where $\theta_{\rho'_j \rho'_k \rho_j \rho_k}$ is the two-body minimum image function, as previously defined in section 3.2.4.

Normal-mode co-ordinates

Transforming to normal-mode co-ordinates, the equation for interpolating monomial coefficients becomes

$$V_{\chi'_1 \dots \chi'_n}^{(n)} = \sum_{\substack{\rho'_1 \dots \rho'_n \\ \rho' \in S'}} \left(\prod_{j=1}^n U_{\rho'_j \chi'_j} \right) \sum_{\substack{\rho_1 \dots \rho_n \\ \rho \in S}} \left(\frac{1}{n} \sum_{j=1}^n \prod_{\substack{k=1 \\ k \neq j}}^n \theta_{\rho'_j \rho'_k \rho_j \rho_k} \right) \sum_{\substack{\chi_1 \dots \chi_n \\ \chi \in S}} \left(\frac{1}{N^n} \prod_{j=1}^n U_{\rho_j \bar{\chi}_j} \right) V_{\chi_1 \dots \chi_n}^{(n)}. \quad (4.14)$$

This expression is somewhat unwieldy, and so it is convenient to simplify it by introducing the multidimensional q-point overlap

$$f_{j_1 \dots j_n}((\mathbf{q}' - \mathbf{q})_{1 \dots n}) = \frac{1}{N' N^{n-1}} \frac{1}{n} \sum_{j=1}^n \sum_{\substack{\mathbf{R}'_j \ \mathbf{R}_j \\ \mathbf{R}'_k \ \mathbf{R}_k \\ k=1 \\ k \neq j}} \prod_{k=1}^n \left(\sum_{\substack{\mathbf{R}'_k \ \mathbf{R}_k \\ k=1 \\ k \neq j}} \theta_{\rho'_j \rho'_k \rho_j \rho_k} e^{2\pi i (\mathbf{q}'_k - \mathbf{q}_k) \cdot (\mathbf{R}'_j - \mathbf{R}'_k)} \right), \quad (4.15)$$

which describes the overlap between the q-points $\{\mathbf{q}'\}$ of $\chi_1 \dots \chi_n$ and the q-points $\{\mathbf{q}\}$ of $\chi'_1 \dots \chi'_n$, with respect to the set of atoms $j_1 \dots j_n$. The q-point overlap has a number of important properties:

- If all $\mathbf{q}'_j = \mathbf{q}_j$ then $f = 1$.
- If any \mathbf{q}'_j is commensurate with S but $\mathbf{q}'_j \neq \mathbf{q}_j$ then $f = 0$.
- For large enough¹ supercells, f drops off rapidly as each $|\mathbf{q}'_j - \mathbf{q}_j|$ increases.
- If $\mathbf{q}'_{j_1} = \mathbf{q}_{j_1}$, then $f_{j_1 \dots j_n}$ simplifies to $f_{j_2 \dots j_n}$, and similarly for other j .

In terms of the multidimensional q-point overlap and the mode overlaps defined in section 3.2.4, the interpolated monomial coefficients become

$$V_{\chi'_1 \dots \chi'_n}^{(n)} = \frac{N'}{N} \sum_{\substack{\chi_1 \dots \chi_n \\ \chi \in S}} \left(\sum_{j_1 \dots j_n} f_{j_1 \dots j_n}((\mathbf{q}' - \mathbf{q})_{1 \dots n}) \prod_{k=1}^n U_{j_k \chi'_k \bar{\chi}_k} \right) V_{\chi_1 \dots \chi_n}^{(n)}. \quad (4.16)$$

¹Both supercells must contain enough R-vectors that $(\mathbf{q} - \mathbf{q}') \cdot \mathbf{R}$ takes multiple values.

To further simplify notation, it is convenient to denote the term in brackets as the interpolation tensor $f((\chi' - \chi)_{1\dots n})$, such that

$$V_{\chi'_1 \dots \chi'_n}^{(n)} = \frac{N'}{N} \sum_{\substack{\chi_1 \dots \chi_n \\ \chi \in S}} f((\chi' - \chi)_{1\dots n}) V_{\chi_1 \dots \chi_n}^{(n)}. \quad (4.17)$$

Just as with the q-point overlap, this reduces in dimensionality if \mathbf{q}'_j are commensurate with S for $k < j \leq n$, as

$$f((\chi' - \chi)_{1\dots n}) \rightarrow f((\chi' - \chi)_{1\dots k}) \prod_{j=k+1}^n \delta_{\chi'_j \chi_j}. \quad (4.18)$$

The calculation of f is conceptually relatively straightforward, and for the typical low-dimensional monomial it is also computationally inexpensive.

4.2.3 Size consistency

A necessary condition for size consistency is that if the system is interpolated to a supercell with N R-vectors, in the large-supercell limit the internal energy must scale as $\mathcal{O}(N^1)$ [14, 71].

An important result from section 4.2.2 above is that the coefficients of the monomials in normal-mode co-ordinates in a supercell S of size N all scale as

$$V^{(n)} \propto \mathcal{O}(N). \quad (4.19)$$

However, before the size consistency of the PES can be evaluated, it is necessary to evaluate the scaling of the expectation values of the monomials themselves, $\langle \prod_{j=1}^n u_{\chi_j} \rangle$, with N . In order to calculate this, it is first necessary to consider the scaling of the expectation values of monomials in cartesian displacement co-ordinates, $\langle \prod_{j=1}^n d_{\rho_j} \rangle$.

Scaling of monomials in displacement co-ordinates

The expectation value of each monomial $\langle \prod_{j=1}^n d_{\rho_j} \rangle$ is intensive, and so scales as $\mathcal{O}(N^0)$. This is because these monomials are local physical quantities, and so are independent of the details of the simulation. However, while these monomials do not vary with the supercell size, it is necessary to consider how they vary with the set of modes $\{\rho\}$ which are included in the term.

Under the assumption that inter-atomic interactions have a finite range, there is also a finite range to the correlation between atomic displacements. Since the motion of each atom is correlated with the motions of

only $\mathcal{O}(N^0)$ other atoms, and there are $\mathcal{O}(N)$ atoms in the supercell, this gives rise to $\mathcal{O}(N)$ pairs of atoms whose motion is correlated.

This does not mean that there are only $\mathcal{O}(N)$ non-zero expectation values $\langle \prod_{j=1}^n d_{\rho_j} \rangle$, as these expectation values can be non-zero even when including pairs of atoms whose motions are uncorrelated. For example, if atoms ρ_1 and ρ_2 are uncorrelated then the expectation value $\langle d_{\rho_1}^2 d_{\rho_2}^2 \rangle$ will simplify to $\langle d_{\rho_1}^2 \rangle \langle d_{\rho_2}^2 \rangle$, which is non-zero. Taking this into account, it is possible to count the number of monomials in the PES whose expectation values are non-zero, for each value of the monomial power n :

n	no. non-zero terms	
1	$\mathcal{O}(N^1)$	
2	$\mathcal{O}(N^1)$	
3	$\mathcal{O}(N^1)$	
4	$\mathcal{O}(N^2)$.
5	$\mathcal{O}(N^2)$	
...	...	
even $n > 1$	$\mathcal{O}(N^{n/2})$	
odd $n > 1$	$\mathcal{O}(N^{(n-1)/2})$	

(4.20)

Scaling of monomials in normal-mode co-ordinates

The scaling of the expectation value of a monomial in normal-mode co-ordinates $\langle \nu \rangle = \langle \prod_{j=1}^n u_{\chi_j} \rangle$ depends non-trivially on the power of the monomial n and on the set of normal-modes $\{\chi_j\}$. The simplest way of calculating this scaling is to factorise each monomial as $\nu = \prod_j f_j$, where each factor is itself a monomial in normal-mode co-ordinates, given by $f_j = \prod_{k=1}^{n_j} u_{\chi_k}$. This is useful because the expectation of the monomial $\langle \nu \rangle$ must scale at least as slowly as the product of the expectation values of its factors, $|\langle \nu \rangle| \geq \left| \prod_j \langle f_j \rangle \right|$. As such, the scaling of the monomial can be found by finding the factorisation which yields the slowest scaling.

Factorisations of ν are only relevant to the scaling of $\langle \nu \rangle$ if none of the expectation values of the factors $\langle f_j \rangle$ are zero. For the expectation value of a factor to be non-zero the factor must be translationally invariant, such that $\sum_{k=1}^{n_j} \mathbf{q}_{\chi_k} = \mathbf{G}$ for some G-vector \mathbf{G} , and there may not be any symmetry operators which map f_j onto $-f_j$.

The scaling of any factor f_j which can be further factorised into terms with non-zero expectation values $f_j = f'_j f''_j$ will be determined by the scaling of those factors, $|\langle f_j \rangle| \geq |\langle f'_j \rangle \langle f''_j \rangle|$. As such, it is only necessary to consider factorisations where each factor cannot be further factorised.

Expanding a factor f_j in cartesian co-ordinates gives

$$\langle f_j \rangle = \left\langle \prod_{k=1}^{n_j} \left(\sum_{\rho_k} \frac{1}{N} U_{\rho_k \bar{\chi}_k} d_{\rho_k} \right) \right\rangle. \quad (4.21)$$

The details of the matrices $U_{\rho_k \bar{\chi}_k}$ do not matter beyond the factor of $e^{2\pi i \mathbf{q} \cdot \mathbf{R}}$, and so

$$\langle f_j \rangle \propto \mathcal{O} \left(\frac{1}{N^{n_j}} \sum_{\mathbf{R}_1 \dots \mathbf{R}_{n_j}} \prod_{k=2}^{n_j} e^{2\pi i \mathbf{q}_k \cdot (\mathbf{R}_k - \mathbf{R}_1)} \left\langle \prod_{k=1}^{n_j} d_{\rho_k} \right\rangle \right). \quad (4.22)$$

As discussed above, if f_j is further factorised as $f_j = f'_j f''_j$ then $\langle f'_j \rangle = \langle f''_j \rangle = 0$. A consequence of this is that the contribution to the expectation value from any monomial in displacement co-ordinates $\prod_{k=1}^n d_{\rho_k}$ which can be factorised as $\langle \prod_{k=1}^n d_{\rho_k} \rangle = \langle \prod_{k=1}^m d_{\rho_k} \rangle \langle \prod_{k=m+1}^n d_{\rho_k} \rangle$ must also be zero. The only monomials which cannot be factorised in this way are those which include only combinations of atoms which directly interact with one another. These monomials must have small values of $|\mathbf{R}_k - \mathbf{R}_1|$ for all values of k , and so there are only $\mathcal{O}(N^1)$ such monomials which contribute to the $\sum_{\mathbf{R}_1 \dots \mathbf{R}_{n_j}}$ sum. Since each monomial is individually intensive, the total sum must scale as $\mathcal{O}(N^1)$.

Combining the $\mathcal{O}(N^1)$ scaling of the $\sum_{\mathbf{R}_1 \dots \mathbf{R}_{n_j}}$ sum with the $\frac{1}{N^{n_j}}$ prefactor gives the total scaling of $\langle f_j \rangle$ as $\mathcal{O}(N^{1-n_j})$, which depends only on the power of the factor, n_j . Since the powers of the factors must add up to the power of the original monomial, a monomial ν with power n which is factorised into n_f factors will scale as

$$\langle \nu \rangle \propto \mathcal{O}(N^{n_f - n}). \quad (4.23)$$

The scaling of $\langle \nu \rangle$ will be determined by the factorisation of ν which contains the most factors with non-zero expectation values n_f .

For example, the term $\langle u_1 u_2 u_3 \rangle$ with $\mathbf{q}_1 = \mathbf{q}_2 = \mathbf{q}_3 = 0$ will scale as $\langle u_1 u_2 u_3 \rangle \propto \langle u_1 \rangle \langle u_2 \rangle \langle u_3 \rangle \propto \mathcal{O}(N^0)$. Alternatively, if $\mathbf{q}_1 = \bar{\mathbf{q}}_2$ and $\mathbf{q}_3 = 0$ then $\langle u_1 u_2 u_3 \rangle \propto \langle u_1 u_2 \rangle \langle u_3 \rangle \propto \mathcal{O}(N^{-1})$, and if instead \mathbf{q}_1 , \mathbf{q}_2 and \mathbf{q}_3 are all non-zero then $\langle u_1 u_2 u_3 \rangle \propto \mathcal{O}(N^{-2})$.

This result is widely reported incorrectly. Fu et al. [13] incorrectly state that the u_χ quantities are intensive, which is only true at the Γ -point, and Hirata et al. [14] incorrectly state that $\langle V^{(n)} \prod_{j=1}^n u_{\chi_j} \rangle \propto (N^{1-\frac{n}{2}})$, which is only true for monomials composed entirely of pairs of modes $u_\chi u_{\bar{\chi}}$.

Scaling of the potential energy surface

Combining the scaling of the monomial coefficients and the scaling of the monomials themselves gives the scaling of the full PES with the size of the supercell N .

For a Taylor expansion truncated to order n but without taking the PAA, this means that the PES contains $\mathcal{O}(N^0)$ monomials which scale as $\mathcal{O}(N^1)$, contains $\mathcal{O}(N^1)$ monomials which scale as $\mathcal{O}(N^0)$, and in general contains $\mathcal{O}(N^x)$ monomials which scale as $\mathcal{O}(N^{1-x})$, up to $x = n - 1$. This means that the expectation of a truncated Taylor expansion of the PES is completely size-extensive, as expected.

The monomials in the potential up to $n = 4$ are presented in the table below, along with their scaling with N and the number of monomials of each kind which appear in the Taylor expansion PES.

scaling of monomial	no. of monomials	example monomials
$\mathcal{O}(N^1)$	$\mathcal{O}(N^0)$	$V^{(1)}u_{\Gamma}, V^{(2)}u_{\Gamma}u_{\Gamma}, V^{(3)}u_{\Gamma}u_{\Gamma}u_{\Gamma}, V^{(4)}u_{\Gamma}u_{\Gamma}u_{\Gamma}u_{\Gamma}$
$\mathcal{O}(N^0)$	$\mathcal{O}(N^1)$	$V^{(2)}u_{\mathbf{q}_1}u_{\bar{\mathbf{q}}_1}, V^{(3)}u_{\Gamma}u_{\mathbf{q}_1}u_{\bar{\mathbf{q}}_1}, V^{(4)}u_{\Gamma}u_{\Gamma}u_{\mathbf{q}_1}u_{\bar{\mathbf{q}}_1}$
$\mathcal{O}(N^{-1})$	$\mathcal{O}(N^2)$	$V^{(3)}u_{\mathbf{q}_1}u_{\mathbf{q}_2}u_{\bar{\mathbf{q}}_1+\bar{\mathbf{q}}_2}, V^{(4)}u_{\Gamma}u_{\mathbf{q}_1}u_{\mathbf{q}_2}u_{\bar{\mathbf{q}}_1+\bar{\mathbf{q}}_2}, V^{(4)}u_{\mathbf{q}_1}u_{\bar{\mathbf{q}}_1}u_{\mathbf{q}_2}u_{\bar{\mathbf{q}}_2}$
$\mathcal{O}(N^{-2})$	$\mathcal{O}(N^3)$	$V^{(4)}u_{\mathbf{q}_1}u_{\mathbf{q}_2}u_{\mathbf{q}_3}u_{\bar{\mathbf{q}}_1+\bar{\mathbf{q}}_2+\bar{\mathbf{q}}_3}$

(4.24)

In this table, $u_{\mathbf{q}}$ refers to any mode at q-point \mathbf{q} , and repeated $u_{\mathbf{q}}$ terms may refer to different modes at the same q-point. It should be assumed that all terms other than u_{Γ} are not at the Γ -point. However, the distinct non- Γ q-point labels may refer to the same q-point, so that e.g. $\mathbf{q}_1 = \mathbf{q}_2$ is valid.

The results presented in this table assume that there are no crystal symmetries other than the purely translational symmetries. If the crystal is invariant under additional symmetries then some of the monomials will have expectation values which are zero, and some will have reduced scaling with N , such that in the limit of large N the contribution of these monomials to the expectation of the PES can be neglected.

Section 5.3 presents explicit calculations of a number of monomial expectation values, using the example of the dispersionless anharmonic oscillator.

4.2.4 Size consistency and the principal axes approximation

There are two major problems with the PAA which mean it is not size-consistent and does not behave well under interpolation.

Firstly, the PES under the PAA includes sets of terms which are not size-consistent. For example, taking the PAA at the first order means that terms of the form $V^{(4)}u_{\mathbf{q}_1}u_{\bar{\mathbf{q}}_1}u_{\mathbf{q}_2}u_{\bar{\mathbf{q}}_2}$ are only included if $\mathbf{q}_1 = \mathbf{q}_2$. This constraint means that only $\mathcal{O}(N^1)$ such terms are included. Since these terms individually scale as $\mathcal{O}(N^{-1})$, the total contribution of these terms to the free energy scales as $\mathcal{O}(N^0)$, rather than as $\mathcal{O}(N^1)$ as required.

Secondly, the set of terms which are included or neglected changes unphysically as the size of the system changes. For example, if the PES is mapped in the primitive cell under the PAA at the first order, it will

contain terms of the form u_{Γ}^4 . Interpolating this PES to a supercell containing two primitive cells, and the q-points Γ and X , terms of the form $u_{\Gamma}^2 u_{\mathbf{x}}^2$ will be included. If however the PES was directly mapped in the larger supercell rather than interpolating, then under the PAA at the first order such terms would not be included. This can be seen in the example presented in section 5.3.

The inconsistency in interpolation arises because the representation of a given term $\prod_{\chi} u_{\chi}$ in cartesian coordinates is dependent on the Born-von Karman supercell in which the term is defined. This means that, unlike the harmonic terms, the anharmonic terms in the PES cannot simply be calculated in the smallest supercell which is commensurate with them and then transferred directly to the full Born-von Karman supercell.

There is also a more fundamental reason why the PAA is not size-consistent: the assumption upon which it is based, namely that the intra-mode terms in the PES are much larger than the inter-mode terms in the PES, breaks down in the large-supercell limit. Under the assumption of finite-ranged inter-atomic interactions, the PES is a smooth function of the q-points. This means that for q-points which are sufficiently close together, the inter-q-point terms must be very similar to the intra-q-point terms. This completely contradicts the PAA, and means the PAA must break down as the q-point grid becomes fine and the large-supercell limit is approached. The consequences of this contradiction are expanded upon in the context of VSCF in section 5.2.

Recovering size consistency under the principal axes approximation

Several changes must be made in order to make the PAA size-consistent.

Firstly, every basis function must be explicitly linked to the Born-von Karman supercell in which it is calculated. As such, the fundamental representation of each basis function should be that in cartesian displacement co-ordinates using a minimum image convention, rather than that in the normal-mode co-ordinates of a given supercell. The representation of the basis functions in normal-mode co-ordinates will then vary between different supercells.

Secondly, each set of basis functions calculated in a single supercell must be size-extensive. This is most easily achievable by constructing sets which contain $\mathcal{O}(N^n)$ basis functions which all scale as $\mathcal{O}(N^{1-n})$ for some n . If N and n are large then this can become extremely computationally intensive.

When fitting the basis functions in a given supercell, it is necessary to consider the terms generated in other supercells. This is best done by choosing a set of supercells which are hierarchically nested in order of size, such that every q-point in each supercell is commensurate with every larger supercell. This means that the basis functions calculated in each supercell do not interpolate to give additional terms in the smaller

supercells. This in turn means that the terms in each supercell can be calculated without needing to reference the larger supercells.

For example, if the desired supercell contains $4 \times 4 \times 4$ primitive cells, then calculations could be run in three cells: the $1 \times 1 \times 1$ primitive cell, an intermediate $2 \times 2 \times 2$ supercell, and the full $4 \times 4 \times 4$ supercell. The Γ -point basis functions would be calculated in the $1 \times 1 \times 1$ cell, without reference to the other cells. Then the calculations involving the eight q-points commensurate with the $2 \times 2 \times 2$ cell would be calculated, as a correction to the Γ -point basis functions. Finally, the calculations involving the q-points commensurate with the $4 \times 4 \times 4$ cell would be calculated, as a correction to both the Γ -point basis functions and the $2 \times 2 \times 2$ basis functions.

It should be noted that making the PAA both size-consistent and symmetry-invariant may require the calculation of additional terms which describe inter-subspace coupling which would be neglected by the PAA on their own, but which contribute to intra-subspace terms when interpolated.

While this treatment helps, it does not correct all of the problems with the PAA. The coefficients of the additional terms which arise when transferring terms from one supercell to another are calculated based on the approximation that the inter-atomic interactions are contained within the Born-von Karman supercell in which the electronic structure calculation is performed. This Born-von Karman supercell does not get larger as the Born-von Karman supercell of the full vibrational calculation gets larger, and so this approximation does not converge towards full accuracy as the system size is increased.

In order to regain convergence with supercell size, either all calculations must be performed using the full Born-von Karman supercell, or the coefficients of the terms which are introduced when transferring terms between supercells must be explicitly calculated. Both of these methods are significantly more computationally expensive than the size-inconsistent methods.

4.3 Exploiting the finite range of interactions

For many systems, the anharmonic corrections to the inter-atomic interactions have a much shorter effective range than the corresponding harmonic interactions [45].

Formally, this means that when constructing the PES in displacement co-ordinates,

$$V = \sum_j V_j \prod_{\rho} d_{\rho}^{n_{j\rho}}, \tag{4.25}$$

the anharmonic terms, with $\sum_{\rho} n_{j\rho} > 2$, have coefficients V_j which fall off faster with distance than the

corresponding coefficients of the harmonic terms, with $\sum_{\rho} n_{j\rho} = 2$. Distance is defined here in terms of the inter-atomic distances of the static-lattice structure, $|\mathbf{r}_{\rho}^0 - \mathbf{r}_{\rho'}^0|$. For terms containing more than two atoms, the associated distance can be taken as the maximum pair-wise distance.

Conceptually, this makes sense. Consider the inter-atomic energy between two atoms $V(\mathbf{r}_1 - \mathbf{r}_2)$. At large distances, this must be bounded by a decaying polynomial, $V = V_0|\mathbf{r}_1 - \mathbf{r}_2|^{-n}$ for some $n > 3$ [71]. The harmonic terms then scale as $|\mathbf{r}_1^0 - \mathbf{r}_2^0|^{-n-2}$, whereas e.g. the quartic terms scale as $|\mathbf{r}_1^0 - \mathbf{r}_2^0|^{-n-4}$.

This means that for a given accuracy of calculation, the anharmonic terms can generally be truncated at a shorter range than the harmonic terms. This in turn means that the anharmonic terms can be calculated using a smaller supercell than the harmonic terms.

There will be systems which do not obey this behaviour, for example due to the cancellation of interactions between different atoms, varying sensitivity to different terms in the Hamiltonian, or significantly larger anharmonic coefficients than harmonic coefficients. But many systems will exhibit this behaviour, and the cost saving of only calculating anharmonic terms in a smaller supercell is likely to be substantial.

4.4 Fitting the potential energy surface

Assuming a basis-function representation, the PES is modelled as

$$V(\mathbf{u}) = \sum_i c_i \mu_i(\mathbf{u}), \quad (4.26)$$

where each μ_i is a basis function, and each c_i is the corresponding scalar coefficient.

Once the set of basis functions has been identified and optimised, it is necessary to fit the basis function coefficients, such that the difference between the model PES and the true PES is as small as possible.

As detailed in section 3.3, the information available about the true PES is a set of *samples*, \mathbf{u} , at which the potential $V(\mathbf{u})$ and potentially other variables are calculated.

It is not trivial to calculate the error between the modelled PES and the true PES. Ideally, this would be defined in terms of the free energy difference between the two, but in practice it is far too computationally expensive to calculate the free energy for multiple model PESs, and the free energy of the true PES is not readily available. As such, the error between the modelled PES and the true PES must be calculated using a heuristic. Minimising this heuristic with respect to the basis function coefficients then gives the optimal values of those coefficients.

Sampling the true PES is typically the computational bottleneck of the entire free energy calculation, and the choice of samples is one of the factors determining the accuracy of the basis function coefficients and, consequently, the accuracy of the entire PES representation, which in turn determines the accuracy of the overall calculation. As such, it is desirable to choose an optimal set of samples which gives sufficient information for accurate calculation without adding unnecessary computational cost, and to have a method which optimally uses this information to calculate coefficients as accurately as possible.

4.4.1 Linear regression

Given a set of basis functions $\{\mu_i\}$ and a set of samples $\{\mathbf{u}_j\}$ with corresponding energies $\{V_j\}$ and possibly other values like forces $\{\mathbf{f}_j\}$ and Hessians $\{D_j\}$, one of the most straightforward ways of fitting the potential is to use linear regression.

The model PES is a sum of basis functions,

$$V^{\text{model}}(\mathbf{u}) = \sum_i c_i \mu_i(\mathbf{u}). \quad (4.27)$$

This can be differentiated to give forces,

$$\mathbf{f}^{\text{model}}(\mathbf{u}) = \sum_i c_i \mathbf{f}_i(\mathbf{u}) = \sum_i c_i \left(-\frac{\partial \mu_i}{\partial \mathbf{u}} \right) \quad (4.28)$$

and Hessians,

$$D(\mathbf{u})^{\text{model}} = \sum_i c_i D_i(\mathbf{u}) = \sum_i c_i \left(\frac{\partial^2 \mu}{\partial \mathbf{u} \partial \mathbf{u}'} \right). \quad (4.29)$$

The true PES is a sum of the model PES and an error function,

$$V^{\text{true}}(\mathbf{u}_j) = V^{\text{model}}(\mathbf{u}_j) + V^{\text{error}}(\mathbf{u}_j). \quad (4.30)$$

The error here comes from two sources: basis function coefficients not taking their true values, and the set of basis functions not being complete. The basis function coefficients are found by minimising the former error, while the latter error limits the region of the PES which can be described by a given set of basis functions.

One heuristic for the total error between the model PES and the true PES is a weighted sum of the difference between the true values and the model values at each sample,

$$\mathcal{L} = \sum_j w_j |\mathbf{x}^{\text{true}}(\mathbf{u}_j) - \mathbf{x}^{\text{model}}(\mathbf{u}_j)|^2, \quad (4.31)$$

where $\mathbf{x}(\mathbf{u}_j)$ is a vector representing all of the information available at sample \mathbf{u}_j , and w_j is a scalar weight which determines the importance of that sample to the fitting process. The choice of weights is discussed in section 4.4.3.

The vector $\mathbf{x}(\mathbf{u}_j)$ depends on the information available from the electronic structure method. For example, if only energies are available then

$$\mathbf{x} = \begin{pmatrix} V \end{pmatrix}. \quad (4.32)$$

Alternatively, with energies and forces

$$\mathbf{x} = \begin{pmatrix} V \\ w_{\mathbf{f}} f_{\sigma_1} \\ \vdots \\ w_{\mathbf{f}} f_{\sigma_N} \end{pmatrix}, \quad (4.33)$$

where $w_{\mathbf{f}}$ is a proportionally constant with dimensions of length, which should be chosen to reflect the accuracy of the energy information available compared to the accuracy of the force information available. This can be extended in a similar manner to include the elements of the Hessian or any other available information.

Minimising this heuristic with respect to the basis function coefficients gives the optimum basis function coefficients. Expanding out the individual basis functions, the heuristic becomes

$$\mathcal{L} = \sum_j w_j \left| \mathbf{x}^{\text{true}}(\mathbf{u}_j) - \sum_i c_i \mathbf{x}_i(\mathbf{u}_j) \right|^2. \quad (4.34)$$

Setting $0 = \frac{\partial \mathcal{L}}{\partial c_i}$ gives

$$0 = \sum_j w_j \left(\mathbf{x}^{\text{true}}(\mathbf{u}_j) - \sum_i c_i \mathbf{x}_i(\mathbf{u}_j) \right) \otimes \mathbf{x}_i(\mathbf{u}_j), \quad (4.35)$$

which can be re-arranged to give

$$c_i = \left(\sum_j w_j \mathbf{x}^{\text{true}}(\mathbf{u}_j) \otimes \mathbf{x}_i(\mathbf{u}_j) \right) \cdot \left(\sum_j w_j \mathbf{x}_i(\mathbf{u}_j) \otimes \mathbf{x}_i(\mathbf{u}_j) \right)^{-1}. \quad (4.36)$$

There are a number of benefits to this method. It is analytic, so the calculation is relatively straightforward. It is also holistic, calculating the entire coefficient vector from all available information. This makes it easy to include additional information that might be available, or to leave out any information which may be unavailable or unreliable.

There must be enough samples so that the matrix which must be inverted is not singular. Once this point is reached, additional samples can be added at will. Provided the choice of basis functions is appropriate, each

additional sample will reduce the error of the fit. This allows the PES coefficients to be repeatedly fitted, first using only the minimum number of samples and then including additional samples for each subsequent fit. The convergence of the coefficients $\{c_i\}$ and the error \mathcal{L} with increasing numbers of samples allows the verification both that the number of samples is sufficient and that the basis functions are appropriate.

4.4.2 Divide and conquer

The computational cost of linear regression is approximately cubic in the number of coefficients being fitted. This makes it impractical to fit all of the coefficients simultaneously. This limitation can be avoided by dividing the set of basis functions into a number of smaller sets, and fitting the basis functions in each set separately. A method for making this division will now be presented, followed by an example.

Division by subspace combination

By construction, each monomial in normal-mode co-ordinates,

$$\nu = \prod_{\chi} u_{\chi}^{n_{\chi}}, \quad (4.37)$$

only has non-zero powers n_{χ} in a small number of normal-mode subspaces ξ . This set of subspaces is the *subspace combination* corresponding to the monomial, $\zeta = \{\xi\}$ [101, 138]. All of the monomials in a given basis function μ_i correspond to the same subspace combination, and so each basis function corresponds to a single subspace combination. The contribution to the PES from each subspace combination is the *intrinsic potential* of that subspace combination [96], given by

$$V^{\zeta} = \sum_{i \in \zeta} c_i \mu_i. \quad (4.38)$$

Each sample also has a corresponding subspace combination. A sample corresponds to a given subspace combination if its co-ordinate,

$$\mathbf{u}_j = \sum_{\chi} u_{\chi} \hat{\mathbf{u}}_{\chi}, \quad (4.39)$$

only has non-zero values of u_{χ} along modes which are in the subspaces in that combination.

A subspace combination ζ_1 is a subset of another subspace combination ζ_2 if all of the subspaces in ζ_1 are also in ζ_2 , i.e. $\zeta_1 \subseteq \zeta_2$ iff $\xi \in \zeta_2 \forall \xi \in \zeta_1$. For example, the subspace combination $\{1, 2\}$ is a superset of the subspace combinations $\{\}$, $\{1\}$, $\{2\}$ and $\{1, 2\}$.

Importantly, if a basis function is evaluated at a sample then the result may only be non-zero if the subspace combination of the sample is a superset of the subspace combination of the basis function,

$$\mu_i^{\zeta_1}(\mathbf{u}_j^{\zeta_2}) = \begin{cases} \text{non-zero} & \text{if } \zeta_1 \subseteq \zeta_2 \\ 0 & \text{otherwise.} \end{cases} \quad (4.40)$$

This means that the value of the PES at a given sample depends only on the basis functions corresponding to subspace combinations which are subsets of the sample's subspace combination,

$$V(\mathbf{u}_j^{\zeta_2}) = \sum_{\zeta_1 \subseteq \zeta_2} V^{\zeta_1}(\mathbf{u}_j^{\zeta_2}) = \sum_{\zeta_1 \subseteq \zeta_2} \sum_{i \in \zeta_1} c_i \mu_i^{\zeta_1}(\mathbf{u}_j^{\zeta_2}). \quad (4.41)$$

This hierarchical dependence allows the subspace combinations to be fitted hierarchically, in increasing order of the number of subspaces in the subspace combination. This gives a set of equations for each subspace combination

$$\sum_{i \in \zeta_2} c_i \mu_i^{\zeta_2}(\mathbf{u}_j^{\zeta_2}) = V(\mathbf{u}_j^{\zeta_2}) - \sum_{\zeta_1 \subset \zeta_2} \sum_{i \in \zeta_1} c_i \mu_i^{\zeta_1}(\mathbf{u}_j^{\zeta_2}). \quad (4.42)$$

When fitting the basis functions corresponding to a given subspace combination, the coefficients c_i on the right hand side of the equations will already have been fitted, and so the equations can be solved to find the set of coefficients on the left hand side of the equation.

An example of dividing and conquering

Consider a system containing two normal modes: u_1 and u_2 . The modes are not degenerate, so the mode u_1 is in subspace 1 and the mode u_2 is in subspace 2.

The PES of this system is

$$V = \sum_i c_i \mu_i, \quad (4.43)$$

where the c_i are real coefficients to be fitted, and the μ_i are basis functions, given in this example by

	basis function	subspace combination	
μ_0	1	{}	
μ_1	u_1^2	{1}	
μ_2	u_1^4	{1}	
μ_3	u_2^2	{2}	
μ_4	u_2^4	{2}	
μ_5	$u_1^2 u_2^2$	{1, 2}	(4.44)

This PES is going to be fitted using a set of samples $\{\mathbf{u}_i\}$ given by

	co-ordinate	subspace combination
\mathbf{u}_0	0	$\{\}$
\mathbf{u}_1	$\hat{\mathbf{u}}_1$	$\{1\}$
\mathbf{u}_2	$2\hat{\mathbf{u}}_1$	$\{1\}$
\mathbf{u}_3	$\hat{\mathbf{u}}_2$	$\{2\}$
\mathbf{u}_4	$2\hat{\mathbf{u}}_2$	$\{2\}$
\mathbf{u}_5	$\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2$	$\{1, 2\}$

(4.45)

It is informative to calculate the value of each basis function at each sample,

		$\{\}$	$\{1\}$		$\{2\}$		$\{1, 2\}$
		\mathbf{u}_0	\mathbf{u}_1	\mathbf{u}_2	\mathbf{u}_3	\mathbf{u}_4	\mathbf{u}_5
$\{\}$	μ_0	1	1	1	1	1	1
$\{1\}$	μ_1	0	1	4	0	0	1
	μ_2	0	1	16	0	0	1
$\{2\}$	μ_3	0	0	0	1	4	1
	μ_4	0	0	0	1	16	1
$\{1, 2\}$	μ_5	0	0	0	0	0	1

(4.46)

where it should be noted that the value of a basis function evaluated at a sample is zero if the subspace combination corresponding to the sample is not a superset of the subspace combination corresponding to the basis function.

The first step of fitting the PES involves only the subspace combination $\{\}$. The only equation corresponding to this subspace combination is

$$c_0\mu_0(\mathbf{u}_0) = V(\mathbf{u}_0), \quad (4.47)$$

which is solved to give c_0 , the constant term of the PES.

The second step of fitting the PES is to fit the basis functions with subspace combinations $\{1\}$ and $\{2\}$. The equations for subspace combination $\{1\}$ are

$$c_1\mu_1(\mathbf{u}_1) + c_2\mu_2(\mathbf{u}_1) = V(\mathbf{u}_1) - c_0\mu_0(\mathbf{u}_1) \quad (4.48)$$

and

$$c_1\mu_1(\mathbf{u}_2) + c_2\mu_2(\mathbf{u}_2) = V(\mathbf{u}_2) - c_0\mu_0(\mathbf{u}_2), \quad (4.49)$$

which are solved to give c_1 and c_2 .

Similarly, the equations for subspace combination $\{2\}$ are

$$c_3\mu_3(\mathbf{u}_3) + c_4\mu_4(\mathbf{u}_3) = V(\mathbf{u}_3) - c_0\mu_0(\mathbf{u}_3) \tag{4.50}$$

and

$$c_3\mu_3(\mathbf{u}_4) + c_4\mu_4(\mathbf{u}_4) = V(\mathbf{u}_4) - c_0\mu_0(\mathbf{u}_4), \tag{4.51}$$

which are solved to give c_3 and c_4 .

Finally, the basis function in subspace combination $\{1, 2\}$ is fitted, using the equation

$$c_5\mu_5(\mathbf{u}_5) = V(\mathbf{u}_5) - \sum_{i=0}^4 c_i\mu_i(\mathbf{u}_5), \tag{4.52}$$

which is solved to give c_5 .

Caveats to dividing and conquering

It should be noted that this method is difficult to apply to basis functions which have been made size-consistent as described in section 4.2.4. When calculated using hierarchical supercell methods, such basis functions are no longer limited to a single subspace combination. There is still a certain amount of hierarchy between these terms, but in many cases it may still be necessary to fit basis functions corresponding to large numbers of subspace combinations at once. This fitting is likely to be costly, and is another point in favour of methods which run calculations using the whole Born-von Karman supercell rather than hierarchical supercells.

4.4.3 Choice of weights for linear regression

When fitting the basis function coefficients using linear regression, the weighting of the samples should be chosen so that quality of the fit is highest in the regions of the PES upon which the free energy most closely depends.

Ideally, this would be done by estimating the vibrational density at each sample, combining this with a term representing each sample's importance to the fit in terms of how similar it is to every other sample, and weighting proportionally to this value. However, such methods are computationally expensive, and provided sufficient samples are used the exact weighting scheme has little impact on the final result. In practice, it is sufficient to weight the samples in a manner which gives high weights to samples with low energies and

which monotonically reduces as the energy increases.

It is important to give samples with very high energies very low weights, as these regions have very low thermal occupation and contribute very little to the free energy, and in general they are very poorly represented by low-order polynomial basis functions, such that including them with higher weighting tends to cause problems with the fit.

4.4.4 Automated sample generation

Before the process of fitting the basis function coefficients from the PES samples can be performed, it is necessary to choose the set of PES samples $\{\mathbf{u}_j\}$. These should be chosen to minimise the total computational cost while keeping the error between the model PES and the real PES to within an acceptable tolerance.

There are many ways of selecting samples, as detailed in section 3.3. Importantly, aside from the methods where samples are iteratively generated after other samples have already been calculated, previous implementations all use a pre-determined sample generation scheme.

However, when symmetries are considered there are many possible combinations of PES basis functions, in addition to the many possible atomic configurations. Additionally, different electronic structure methods incur different computational costs for different kinds of calculation. This means that samples could be chosen more efficiently if the sampling scheme could be made bespoke to each combination of crystal structure and electronic structure method.

This would be possible if it were possible to evaluate the accuracy of a given set of samples for a given PES mapping in advance of performing any electronic structure calculations.

Estimating the accuracy of a set of samples

The true PES can be written as a sum of the model PES and an error term which cannot be captured by the chosen basis functions,

$$V^{\text{true}}(\mathbf{u}) = V^{\text{model}}(\mathbf{u}) + V^{\text{error}}(\mathbf{u}). \quad (4.53)$$

The separation of the error from the model can be extended to the vector of information at each sample,

$$\mathbf{x}^{\text{true}}(\mathbf{u}_j) = \mathbf{x}^{\text{model}}(\mathbf{u}_j) + \mathbf{x}^{\text{error}}(\mathbf{u}_j). \quad (4.54)$$

Running linear regression, this error will be propagated to the basis function coefficients as

$$c_i^{\text{true}} = c_i^{\text{model}} + c_i^{\text{error}}, \tag{4.55}$$

where

$$c_i^{\text{error}} = - \left(\sum_j w_j \mathbf{x}^{\text{error}}(\mathbf{u}_j) \otimes \mathbf{x}_i(\mathbf{u}_j) \right) \cdot \left(\sum_j w_j \mathbf{x}_i(\mathbf{u}_j) \otimes \mathbf{x}_i(\mathbf{u}_j) \right)^{-1}. \tag{4.56}$$

Importantly, this quantity does not depend on the \mathbf{x}^{true} vectors. Given an estimate of $\mathbf{x}^{\text{error}}(\mathbf{u})$, this can be used to estimate the error in the basis function coefficients for a given set of samples, without having to run any electronic structure calculations.

Optimising the choice of samples

Since the error estimate is not computationally costly to calculate, it can be calculated repeatedly to estimate the error of different choices of samples, allowing the choice of samples to be optimised prior to running calculations.

Given sufficient information about the scaling of the cost of the electronic structure method with supercell size and crystal symmetry, this estimator provides the basis for a fully automated method for identifying the samples which fit the PES to a given accuracy with the minimum possible expenditure of computational resources. This is simply a global optimisation problem, for which many solutions exist [10]. One such solution would be to vary the number and positions of the samples using a Monte-Carlo scheme.

This is believed to provide a generalisation of the sampling methods described by Fu et al. [13] and others [34, 42].

Chapter 5

Vibrational self-consistent field theory

5.1 Symmetry and vibrational self-consistent field theory

When calculating the vibrational properties of a system it is convenient to fix the symmetry of the system, such that the vibrational density is invariant under the same symmetries as the static-lattice structure. This allows the static-lattice symmetries to be exploited, in order to increase the accuracy of the calculation and reduce its computational cost.

Fixing the symmetry in this manner means that multiple values may be calculated for each of the properties of a given system; a different set of values will be calculated for each set of crystal symmetries imposed. The set of symmetries which give the lowest free energy are the ground state symmetries of the system, and the properties calculated using these symmetries are the true properties of the system. The other symmetries correspond to metastable and unstable structures. The ground state symmetries may be different at different temperatures, and may differ from the symmetries of the static-lattice structure with the lowest energy.

5.1.1 Symmetry-invariant vibrational self-consistent field theory

In previous works, the VSCF Hamiltonian is constructed as a sum of single-mode Hamiltonians,

$$\mathcal{H}^{\text{VSCF}} = \sum_{\chi} \mathcal{H}_{\chi}(u_{\chi}), \quad (5.1)$$

based on the assumption that the coupling between modes is small enough that treating the modes as uncorrelated yields a reasonable approximation to the free energy. This is equivalent to taking the PAA at

first order, but with coefficients which are calculated self-consistently rather than drawn directly from the PES.

Just like the general case of the PAA, the VSCF Hamiltonian does not behave well under the action of symmetry. If two or more modes at the same q-point are degenerate then any unitary transformation of those modes will also yield a valid basis of degenerate modes. As such, it is inappropriate to split the Hamiltonian into single-mode terms, and there are many cases where such a Hamiltonian cannot be made symmetry-invariant [137].

This applies even when the only crystal symmetries are the translational symmetries. Under a translation by an R-vector \mathbf{R} , the single-mode terms transform as

$$\mathcal{S}(u_\chi^{n_\chi}) = e^{2\pi i n_\chi \mathbf{q} \cdot \mathbf{R}} u_\chi^{n_\chi}, \quad (5.2)$$

and so these terms are only invariant under all translational symmetries if $n_\chi \mathbf{q}$ is a vector of integers. At q-points with irrational elements (formally, almost all q-points) this means that there are no symmetry-invariant single-mode terms, and for a given Taylor expansion truncation there will be no symmetry-invariant single-mode terms at the majority of rational q-points either. Clearly, this formulation of VSCF is inappropriate for constructing a symmetry-invariant Hamiltonian.

If VSCF is constructed using real co-ordinates $u_\pm = \frac{1}{\sqrt{2}}(u_\chi \pm u_{\bar{\chi}})$ rather than complex co-ordinates then more single-mode terms become symmetry invariant, including all of the harmonic terms. However, even with this formulation the majority of single-mode terms are not symmetry invariant.

In order to make the VSCF Hamiltonian invariant under symmetry, instead of constructing it as a sum of single-mode Hamiltonians, it should instead be constructed as a sum of single-subspace Hamiltonians, where each single-subspace Hamiltonian contains all of the degenerate modes at a given q-point and its conjugate q-point,

$$\mathcal{H}^{\text{VSCF}} = \sum_{\xi} \mathcal{H}_{\xi}(\{u_{\xi}\}) + \text{const}. \quad (5.3)$$

This leads to states which are products of single-subspace states,

$$|\psi\rangle = \prod_{\xi} |\psi_{\xi}\rangle, \quad (5.4)$$

and corresponding single-subspace Schrödinger equations,

$$\mathcal{H}_{\xi} |\psi_{\xi}\rangle = E_{\xi} |\psi_{\xi}\rangle, \quad (5.5)$$

analogous to the single-mode equations of standard VSCF.

The single-subspace Hamiltonians are naturally defined as

$$\mathcal{H}_\xi = \langle \mathcal{H} \rangle_{\{\xi' \neq \xi\}} - \langle \mathcal{H} \rangle, \quad (5.6)$$

where $\langle \mathcal{H} \rangle_{\{\xi' \neq \xi\}}$ denotes the thermal expectation of \mathcal{H} across all subspaces not including ξ . The constant in the VSCF Hamiltonian is simply $\langle \mathcal{H} \rangle$, as before.

Just as with the PAA, it is also possible to formulate VSCF in terms of all the modes in each subspace, rather than splitting by q-point. However, since the number of modes in a subspace can be large, this can make constructing and diagonalising the single-subspace Hamiltonians extremely computationally expensive. Further, since the eigenstates are Bloch wavefunctions there may be no mixing between states with different wavevectors, and since the wavevector of a state depends on its occupations at each q-point, this limits the mixing between states at different q-points.

5.1.2 Symmetry-invariant subspace Hamiltonians

The full Hamiltonian \mathcal{H} is symmetry invariant. If spontaneous symmetry breaking is neglected then the VSCF Hamiltonian $\mathcal{H}^{\text{VSCF}} = \sum_\xi \mathcal{H}_\xi$ must also be symmetry invariant. Since each symmetry operator only maps the modes in a given subspace onto other modes in that subspace, the VSCF Hamiltonian being symmetry invariant means that each single-subspace Hamiltonian \mathcal{H}_ξ must individually be symmetry invariant. This means that each single-subspace PES can be constructed as a sum containing only the single-subspace basis functions which are individually symmetry invariant. These basis functions can be constructed using the method presented in section 4.1.3.

The full PES can be expressed as a sum of products of single-subspace basis functions. Some of these products will be symmetry invariant but contain single-subspace basis functions which are not individually symmetry invariant. Such products are necessary for representing the full PES, but will integrate to zero during the construction of the VSCF Hamiltonian, and so have no effect on the VSCF Hamiltonian. It is convenient to call the basis functions which contribute to the VSCF Hamiltonian *VSCF basis functions*, and the basis functions which do not *non-VSCF basis functions*. It is convenient to separate the PES into VSCF and non-VSCF basis functions, and to remove the non-VSCF basis functions before the VSCF Hamiltonian is calculated.

For example, consider a system containing two non-degenerate modes u_1 and u_2 , with an inversion symmetry

\mathcal{S} which acts as $\mathcal{S}u_1 = -u_1$ and $\mathcal{S}u_2 = -u_2$. The full PES of this system might be

$$V = Au_1^2 + Bu_2^2 + Cu_1^2u_2^2 + Du_1^3u_2^3. \quad (5.7)$$

Every term in this PES is invariant under the symmetry \mathcal{S} . Since $\langle u_1^2 \rangle$ and $\langle u_2^2 \rangle$ may be non-zero, the term $Cu_1^2u_2^2$ will contribute to the coefficients of u_1^2 and u_2^2 in the single-subspace Hamiltonians \mathcal{H}_1 and \mathcal{H}_2 respectively, and so this term is a VSCF basis function. However, since $\langle u_1^3 \rangle$ and $\langle u_2^3 \rangle$ must both be zero as a result of the symmetry \mathcal{S} , the term $Du_1^3u_2^3$ will not contribute to the VSCF Hamiltonian, and so this term is a non-VSCF basis function.

5.1.3 Removing unneeded terms from potential fitting

If it were possible to remove the non-VSCF basis functions before the PES was constructed then the PES would contain fewer free coefficients, and so it would be possible to fit the PES coefficients using fewer electronic structure calculations. However, removing the non-VSCF basis functions entirely is not possible, as their contributions to the calculated PES need to be accounted for. Instead, it is possible to use the non-VSCF basis functions to reduce the number of samples at which electronic structure calculations must be run, as will now be detailed.

The result of the electronic structure calculation at each sample $V(\mathbf{u})$ can be separated into the sum of two contributions: the contribution from VSCF basis functions $V^{\text{vscf}}(\mathbf{u})$, and the contribution from non-VSCF basis functions $V^{\text{non-}}(\mathbf{u})$.

It is convenient to express the co-ordinate of each sample \mathbf{u} as a sum of single-subspace co-ordinates,

$$\mathbf{u} = \sum_{\xi} \mathbf{u}_{\xi}. \quad (5.8)$$

It is further to convenient to define single-subspace symmetry operators, \mathcal{S}_{ξ} , which act as

$$\mathcal{S}_{\xi}(\mathbf{u}) = \mathcal{S}(\mathbf{u}_{\xi}) + \sum_{\xi' \neq \xi} \mathbf{u}_{\xi'}, \quad (5.9)$$

i.e. which only act on the normal modes within a given subspace. Following the reasoning presented in section 4.1.3, from every symmetry operator \mathcal{S}_{ξ} it is possible to construct a projection operator $P_{\xi} = \sum_{j=1}^n \mathcal{S}_{\xi}^j$, where n is the order of the operator \mathcal{S}_{ξ} .

The VSCF basis functions are then those basis functions which are invariant under the action of every projection operator, $P_{\xi}V^{\text{vscf}} = V^{\text{vscf}}$. For every non-VSCF basis function, there will be at least one projection operator which maps that basis function to zero, $P_{\xi}V^{\text{non-}} = 0$. Thus, acting with every projection operator

in turn on the full PES will produce only the contribution to the PES from VSCF basis functions,

$$\left(\prod_P P_\xi\right) V = \left(\prod_P P_\xi\right) (V^{\text{vscf}} + V^{\text{non-}}) = V^{\text{vscf}}. \quad (5.10)$$

Acting with a projection operator P_ξ on the PES $V(\mathbf{u})$ is equivalent to averaging the PES across all coordinates \mathbf{u}' which can be reached by acting with the symmetry operator \mathcal{S}_ξ on \mathbf{u} ,

$$P_\xi V(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^n V(\mathcal{S}^j \mathbf{u}). \quad (5.11)$$

Thus, by sampling the PES at each value of \mathbf{u}' it is possible to isolate the contribution to the PES at \mathbf{u} from the VSCF basis functions only.

Since the VSCF basis functions are typically only a small subset of the total basis functions, this method generally offers a significant computational cost reduction over fitting the entire PES and then removing the non-VSCF basis functions afterwards.

There are a couple of disadvantages to this method. Firstly, the existence of non-VSCF basis functions with large coefficients is an indicator that the subspaces are correlated and that VSCF might be a poor approximation to the physical system. Without the calculation of these basis functions, this indicator is not available. Secondly as detailed in section 3.5.5, it is common practice to improve the results of a VSCF calculation using VMPn perturbation theory [107], since for most methods this adds a small improvement in accuracy at little computational cost. However, the non-VSCF basis functions make a contribution to the free energy correction calculated by VMPn, and if they are not present then VMPn will not offer the same improvements in accuracy.

5.2 Size consistency and vibrational self-consistent field theory

When VSCF is interpolated to the large-supercell limit, two of the underlying approximations of the theory interact in an unintended way.

The first approximation is the finite range of inter-atomic interactions, which leads to PES coefficients which vary smoothly with q-point. This means that, in the limit of closely spaced q-points, the coefficients of the inter-q-point terms in the PES must have the same values as the coefficients of the matching intra-q-point terms. This can be illustrated with an example. Consider the one-dimensional system with one atom per primitive cell. From this system, consider the quartic term $V^{(4)} d_{\mathbf{R}_1} d_{\mathbf{R}_2} d_{\mathbf{R}_3} d_{\mathbf{R}_4}$, where the inter-atomic separations $|\mathbf{R}_i - \mathbf{R}_j|$ are all smaller than some cutoff radius R^{max} . Transforming this term into normal-

mode co-ordinates gives inter-q-point terms $V_{\mathbf{q}_1\bar{\mathbf{q}}_1\mathbf{q}_2\bar{\mathbf{q}}_2}^{(4)}|u_{\mathbf{q}_1}|^2|u_{\mathbf{q}_2}|^2$ and intra-q-point terms $V_{\mathbf{q}_1\bar{\mathbf{q}}_1\mathbf{q}_1\bar{\mathbf{q}}_1}^{(4)}|u_{\mathbf{q}_1}|^4$. The ratio between the coefficients of these terms is given by

$$\frac{V_{\mathbf{q}_1\bar{\mathbf{q}}_1\mathbf{q}_2\bar{\mathbf{q}}_2}^{(4)}}{V_{\mathbf{q}_1\bar{\mathbf{q}}_1\mathbf{q}_1\bar{\mathbf{q}}_1}^{(4)}} = \frac{1}{6} \sum_{i=1}^4 \sum_{j=1, j>i}^4 \cos(2\pi(\mathbf{q}_2 - \mathbf{q}_1) \cdot (\mathbf{R}_1 - \mathbf{R}_2)). \quad (5.12)$$

As $|\mathbf{q}_2 - \mathbf{q}_1|$ decreases, this ratio tends to unity. Similar logic holds for all anharmonic terms. As such, in the limit of close q-points, the inter-q-point interactions are as strong as the intra-q-point interactions.

The second approximation is the VSCF approximation, that there is no correlation between nuclear motion along modes in different subspaces. This means that the VSCF Hamiltonian only contains intra-subspace terms, and neglects the inter-subspace terms. This contradicts the first approximation, and has the unfortunate consequence that in the large-supercell limit the VSCF Hamiltonian may only contain harmonic terms. Furthermore, since the anharmonic terms must be neglected in the large-supercell limit they must also be neglected outside of this limit if VSCF is to be made size-consistent. This will be demonstrated more rigorously in the following sections.

The problems faced when applying size consistency to VSCF are closely related to the lack of size consistency of the PAA, as detailed in section 4.2.4. Both VSCF and the PAA try to neglect inter-q-point interactions in favour of intra-q-point interactions, and both methods fail when, in the large-supercell limit, these interactions cannot be neglected.

5.2.1 Interpolated vibrational self-consistent field theory

Rather than interpolating the full PES onto a fine q-point grid and performing VSCF on this PES, it is preferable to calculate the single-subspace VSCF Hamiltonians on the fine q-point grid directly. This is possible by interpolating the coefficients of the VSCF potential rather than the full PES.

In terms of the full PES V and the kinetic energy operator \mathcal{T} , the VSCF Hamiltonian is

$$\mathcal{H}^{\text{VSCF}} = \mathcal{T} + \langle V \rangle + \sum_{\xi} \left(\langle V \rangle_{\{\xi' \neq \xi\}} - \langle V \rangle \right). \quad (5.13)$$

The kinetic energy operator is already a sum of single-subspace operators, and needs little work to interpolate. Additionally, the term $\langle V \rangle$ is a constant, and while this is relevant to calculating the free energy it has no effect on the VSCF eigenstates. As such, in order to calculate the eigenstates of the interpolated VSCF Hamiltonian on a given q-point grid, it is only necessary to calculate $\langle V \rangle_{\{\xi' \neq \xi\}}$ for the subspaces ξ on that grid.

It is convenient to denote the supercell for which the PES is initially calculated as S' , which has size N' and normal-mode co-ordinates $\{\chi'\}$. The VSCF Hamiltonian will be interpolated onto a supercell S , which has size N and normal-mode co-ordinates $\{\chi\}$.

The full PES is initially constructed as a Taylor series in supercell S' , as

$$V = \sum_{n=2}^{\infty} \sum_{\chi'_1 \dots \chi'_n} V_{\chi'_1 \dots \chi'_n}^{(n)} \prod_{j=1}^n u_{\chi'_j}. \quad (5.14)$$

Interpolating to the S supercell, this becomes

$$V = \sum_{n=2}^{\infty} \sum_{\chi_1 \dots \chi_n} V_{\chi_1 \dots \chi_n}^{(n)} \prod_{j=1}^n u_{\chi_j}, \quad (5.15)$$

where the interpolated monomial coefficients can be calculated from the original coefficients as

$$V_{\chi_1 \dots \chi_n}^{(n)} = \frac{N}{N'} \sum_{\chi'_1 \dots \chi'_n} f((\chi - \chi')_{1\dots n}) V_{\chi'_1 \dots \chi'_n}^{(n)}, \quad (5.16)$$

where $f((\chi - \chi')_{1\dots n})$ is the interpolation tensor as defined in section 4.2.2.

In order to calculate each single-subspace potential, the full PES must first be written in a manner which explicitly separates the modes in the subspace and the modes not in the subspace. Taking a binomial expansion of $\sum_{\chi} = \sum_{\chi \in \xi} + \sum_{\chi \notin \xi}$, and noting that the coefficients $V_{\chi_1 \dots \chi_n}^{(n)}$ are invariant under the exchange of any χ indices, equation 5.15 can be re-written as

$$V = \sum_{n=2}^{\infty} \sum_{l=0}^n \binom{n}{l} \sum_{\substack{\chi_1 \dots \chi_l \\ \chi \in \xi}} \sum_{\substack{\chi_{l+1} \dots \chi_n \\ \chi \notin \xi}} V_{\chi_1 \dots \chi_n}^{(n)} \prod_{j=1}^n u_{\chi_j}. \quad (5.17)$$

Taking the expectation of this potential across all subspaces other than ξ gives

$$\langle V \rangle_{\{\xi' \neq \xi\}} = \sum_{l=0}^{\infty} \sum_{\substack{\chi_1 \dots \chi_l \\ \chi \in \xi}} V_{\chi_1 \dots \chi_l}^{(l)} \prod_{j=1}^l u_{\chi_j}, \quad (5.18)$$

where the single-subspace PES coefficients $V_{\chi_1 \dots \chi_l}^{(l)}$ are given by

$$V_{\chi_1 \dots \chi_l}^{(l)} = \sum_{n=\max(2,l)}^{\infty} \binom{n}{l} \left\langle \sum_{\substack{\chi_{l+1} \dots \chi_n \\ \chi \notin \xi}} V_{\chi_1 \dots \chi_n}^{(n)} \prod_{j=l+1}^n u_{\chi_j} \right\rangle. \quad (5.19)$$

Following the reasoning laid out in section 4.2.3, the single-subspace PES coefficients scale in the same way

as the full PES coefficients,

$$V_{\chi_1 \dots \chi_l}^{(l)} \propto \mathcal{O}(N). \quad (5.20)$$

5.2.2 The large-supercell limit

The coefficients of the terms in the VSCF PES may scale in the same way as those in the full PES, but this does not mean that the VSCF Hamiltonian is size-consistent. In the full PES, the total contribution to the free energy from the terms which scale as $\mathcal{O}(N^{-x})$ with $x \geq 1$ will be size-extensive because there are $\mathcal{O}(N^{1+x})$ such terms. In the VSCF Hamiltonian, this is not the case.

The VSCF Hamiltonian contains $\mathcal{O}(N)$ single-subspace Hamiltonians, and each single-subspace Hamiltonian contains $\mathcal{O}(N^0)$ terms. As such, the total contribution to the free energy from all terms which scale as $\mathcal{O}(N^{-1})$ or faster will not be extensive, and the contribution to the specific free energy $\frac{F}{\Omega}$ from these terms will tend to zero in the large-supercell limit. This means that all such terms must be dropped in order to make the VSCF Hamiltonian size-consistent.

For subspaces at the Γ -point, the expectation value of every term in the single-subspace Hamiltonian scales as $\mathcal{O}(N^1)$, and so all terms are extensive. However, for subspaces not at the Γ -point, the expectation value of any term $V^{(n)} \prod_{j=1}^n u_{\chi_j}$ with $n > 2$ will decay with N as $\mathcal{O}(N^0)$ or faster, and so these terms must be dropped. This means that the size-consistent single-subspace Hamiltonians become harmonic.

The different properties at the Γ -point arise from the fact that $\langle u_\chi \rangle$ may be non-zero if and only if $\mathbf{q}_\chi = 0$. This can be accounted for by making a change of variables,

$$\mathbf{u} = \langle \mathbf{u} \rangle + \tilde{\mathbf{u}}. \quad (5.21)$$

This change of variables is useful because of the scaling of $\tilde{\mathbf{u}}$ with the supercell size. Since $\langle \tilde{\mathbf{u}} \rangle$ is zero by construction, any term $\langle \tilde{u}_\chi \rangle$ will be zero, any term $\langle \tilde{u}_{\chi_1} \tilde{u}_{\chi_2} \rangle$ will either be zero or will scale as $\mathcal{O}(N^{-1})$, and any term $\langle \prod_{j=1}^n \tilde{u}_{\chi_j} \rangle$ with $n > 2$ will decay with N as $\mathcal{O}(N^{-2})$ or faster, regardless of whether or not the normal modes are at the Γ -point. Taking this scaling with supercell size into account, and noting that the kinetic energy has no dependence on $\langle \mathbf{u} \rangle$, the functional form of the size-consistent VSCF Hamiltonian becomes

$$\mathcal{H}^{\text{VSCF}} = -\frac{1}{2N} \sum_{\chi} \frac{\partial^2}{\partial \tilde{u}_\chi \partial \tilde{u}_\chi} + \sum_{\chi_1 \chi_2} V_{\chi_1 \chi_2}^{(2)}(\langle \mathbf{u} \rangle) \tilde{u}_{\chi_1} \tilde{u}_{\chi_2} + V(\langle \mathbf{u} \rangle), \quad (5.22)$$

where $V^{(2)}(\langle \mathbf{u} \rangle)_{\chi_1 \chi_2}$ and $V(\langle \mathbf{u} \rangle)$ are functions of $\langle \mathbf{u} \rangle$ which do not depend on $\tilde{\mathbf{u}}$.

This functional form is simply a harmonic Hamiltonian whose origin has been offset from the static-lattice position by a constant offset $\langle \mathbf{u} \rangle$. As such, when size consistency is taken into account, VSCF is equivalent

to an effective harmonic model with a constant offset. Such models have been widely studied and there exist a variety of efficient methods for constructing and solving them [29].

It should be noted that the coefficients $V^{(2)}(\langle \mathbf{u} \rangle)_{\chi_1 \chi_2}$ must obey translational invariance, and so the non-zero coefficients must satisfy $\mathbf{q}_1 = \bar{\mathbf{q}}_2$. However, the coefficients need not satisfy $\chi_1 = \bar{\chi}_2$. That is to say, the non-diagonal harmonic elements may be non-zero, meaning that the normal modes which diagonalise the VSCF Hamiltonian may not be the same as those which diagonalise the Hamiltonian of the system as calculated using the harmonic approximation.

This formulation of size-consistent VSCF is a generalisation of previously presented formulations. Keçeli and Hirata [132] present the correct behaviour for the non- Γ -point modes, and Keçeli et al. [139] present the correct behaviour for the Γ -point modes, but it is not believed that the two behaviours have been correctly combined in previous applications. This omission is likely a result of the incorrect analysis of the scaling of the expectation values of monomials with supercell size presented in previous works, as detailed in section 4.2.3.

5.2.3 Consequences of size consistency

That VSCF reduces from an anharmonic theory to a harmonic theory in the large-supercell limit is an important result, and a result which, since its first publication in 2010 by Hirata et al. [14], has not received the attention it warrants.

While harmonic theories are easier to work with than VSCF in general, they have more limited applicability to physical systems. As previously mentioned, the motivation for using VSCF is to provide a theory which is more accurate than harmonic theories. If VSCF reduces to a harmonic theory then this motivation is not satisfied.

Keçeli and Hirata [132] present size-consistent VSCF as an improvement to previous formulations of VSCF which were not size-consistent. However, the behaviour of VSCF in the large-supercell limit arises as a result of the fundamental inaccuracy of the VSCF approximation in that limit, and enforcing size consistency simply extends this inaccuracy to calculations performed using finite supercells. As such, calculations performed in finite supercells using formulations of VSCF which are not size-consistent are likely to be more accurate than calculations performed using size-consistent VSCF.

The accuracy of non-size-consistent VSCF calculations as a function of supercell size is subject to two competing effects: the approximation that inter-atomic interactions have finite range improves as the size of the supercell increases, improving the accuracy, but the VSCF approximation worsens as the size of the supercell increases, reducing the accuracy. This means that, rather than consistently improving with

supercell size, the accuracy of a given calculation will likely improve up to some supercell size, and then worsen again as the supercell size is increased further. The issue is that, without prior knowledge of the true free energy, there is no way to assess the accuracy of a given calculation, and so it is not possible to identify which supercell will yield an optimally accurate result.

5.3 The dispersionless anharmonic oscillator

The shortcomings of size-consistent VSCF can be made clearer by considering a toy model to which the VSCF method can be applied analytically. This will demonstrate many of the properties of the method without introducing the complexities inherent to applying VSCF to a real crystal.

Consider a one-dimensional crystal of N non-interacting diatomic molecules which has the Hamiltonian

$$\mathcal{H} = \sum_{\mathbf{R}=1}^N \left(-\frac{1}{2} \frac{\partial^2}{\partial d_{\mathbf{R}}^2} + V^{(2)} d_{\mathbf{R}}^2 + V^{(3)} d_{\mathbf{R}}^3 + V^{(4)} d_{\mathbf{R}}^4 \right), \quad (5.23)$$

where each $d_{\mathbf{R}}$ is the mass-weighted displacement from equilibrium of the inter-atomic separation of the molecule at \mathbf{R} -vector \mathbf{R} . The degrees of freedom corresponding to the free translation of each molecule have been neglected, such that the extended system is N -dimensional rather than $2N$ -dimensional.

It will be assumed that N is odd, such that the only \mathbf{q} -point with $q = \bar{q}$ is the Γ -point, $q_{\Gamma} = 0$. Relaxing this assumption does not meaningfully change any results, but does increase the complexity of the example unnecessarily.

Transforming into normal-mode co-ordinates, the Hamiltonian becomes

$$\mathcal{H} = \sum_{q_1} \left(-\frac{1}{2N} \frac{\partial^2}{\partial u_{q_1} \partial u_{\bar{q}_1}} + NV^{(2)} u_{q_1} u_{\bar{q}_1} \right) + \sum_{q_1 q_2} NV^{(3)} u_{q_1} u_{q_2} u_{\bar{q}_1 + \bar{q}_2} + \sum_{q_1 q_2 q_3} NV^{(4)} u_{q_1} u_{q_2} u_{q_3} u_{\bar{q}_1 + \bar{q}_2 + \bar{q}_3}. \quad (5.24)$$

5.3.1 Polynomial expectation values

Since the molecules are non-interacting, the expectation of a monomial in displacement co-ordinates can be written as a product of single-mode expectation values. Further, since every molecule is identical and indistinguishable, every expectation value will be identical. This will be denoted as $\langle d_{\mathbf{R}} \rangle = \langle d \rangle$, $\langle d_{\mathbf{R}}^2 \rangle = \langle d^2 \rangle$ etc. As such, the expectation values become

$$\left\langle \prod_{\mathbf{R}} d_{\mathbf{R}}^{n_{\mathbf{R}}} \right\rangle = \prod_{\mathbf{R}} \langle d^{n_{\mathbf{R}}} \rangle. \quad (5.25)$$

The expectation values $\langle d^n \rangle$ are intensive, and so are independent of N .

It is possible to calculate the expectation values of monomials in normal-mode co-ordinates directly from the expectation values of monomials in displacement co-ordinates. This calculation will now be performed explicitly, for the monomials of up to order three.

The only translationally invariant monomial of order one is u_Γ . The expectation of this is

$$\begin{aligned}
\langle u_\Gamma \rangle &= \left\langle \frac{1}{N} \sum_{\mathbf{R}} d_{\mathbf{R}} \right\rangle \\
&= \frac{1}{N} \sum_{\mathbf{R}} \langle d \rangle \\
&= \langle d \rangle \\
&\propto \mathcal{O}(N^0).
\end{aligned} \tag{5.26}$$

There are two kinds of translationally invariant monomials of order two. There is one term $u_\Gamma u_\Gamma$, whose expectation is

$$\begin{aligned}
\langle u_\Gamma u_\Gamma \rangle &= \left\langle \frac{1}{N^2} \sum_{\mathbf{R}_1 \mathbf{R}_2} d_{\mathbf{R}_1} d_{\mathbf{R}_2} \right\rangle \\
&= \frac{1}{N^2} \sum_{\mathbf{R}_1} \left(\langle d^2 \rangle + \sum_{\mathbf{R}_2 \neq \mathbf{R}_1} \langle d \rangle^2 \right) \\
&= \frac{1}{N} \left(\langle d^2 \rangle + (N-1) \langle d \rangle^2 \right) \\
&\propto \mathcal{O}(N^0).
\end{aligned} \tag{5.27}$$

There are also $\mathcal{O}(N)$ terms $u_q u_{\bar{q}}$ with $q \neq \Gamma$, whose expectations are

$$\begin{aligned}
\langle u_q u_{\bar{q}} \rangle &= \left\langle \frac{1}{N^2} \sum_{\mathbf{R}_1 \mathbf{R}_2} e^{2\pi i q(\mathbf{R}_2 - \mathbf{R}_1)} \right\rangle \\
&= \frac{1}{N^2} \sum_{\mathbf{R}_1} \left(\langle d^2 \rangle + \sum_{\mathbf{R}_2 \neq \mathbf{R}_1} e^{2\pi i q(\mathbf{R}_2 - \mathbf{R}_1)} \langle d \rangle^2 \right) \\
&= \frac{1}{N} \left(\langle d^2 \rangle - \langle d \rangle^2 \right) \\
&\propto \mathcal{O}(N^{-1}).
\end{aligned} \tag{5.28}$$

There are three kinds of translationally invariant monomials of order three. There is one term $u_\Gamma u_\Gamma u_\Gamma$, whose

expectation is

$$\begin{aligned}
\langle u_\Gamma u_\Gamma u_\Gamma \rangle &= \left\langle \frac{1}{N^3} \sum_{R_1 R_2 R_3} d_{R_1} d_{R_2} d_{R_3} \right\rangle \\
&= \frac{1}{N^3} \sum_{R_1} \left(\langle d^3 \rangle + \sum_{R_2 \neq R_1} 3 \langle d^2 \rangle \langle d \rangle + \sum_{R_2 \neq R_1} \sum_{R_3 \neq R_1, R_2} \langle d \rangle^3 \right) \\
&= \frac{1}{N^2} \left(\langle d^3 \rangle + 3(N-1) \langle d \rangle^2 \langle d \rangle + (N-1)(N-2) \langle d \rangle^3 \right) \\
&\propto \mathcal{O}(N^0).
\end{aligned} \tag{5.29}$$

There are also $\mathcal{O}(N)$ terms $u_\Gamma u_q u_{\bar{q}}$ with $q \neq \Gamma$, whose expectations are

$$\begin{aligned}
\langle u_\Gamma u_q u_{\bar{q}} \rangle &= \left\langle \frac{1}{N^3} \sum_{R_1 R_2 R_3} d_{R_1} d_{R_2} d_{R_3} e^{2\pi i q (R_3 - R_2)} \right\rangle \\
&= \frac{1}{N^3} \sum_{R_1} \left(\langle d^3 \rangle + \sum_{R_2 \neq R_1} \left(1 + 2e^{2\pi i q (R_2 - R_1)} \right) \langle d^2 \rangle \langle d \rangle + \sum_{R_2 \neq R_1} \sum_{R_3 \neq R_1, R_2} e^{2\pi i q (R_3 - R_2)} \langle d \rangle^3 \right) \\
&= \frac{1}{N^2} \left(\langle d \rangle^3 + (N-3) \langle d^2 \rangle \langle d \rangle - \langle d \rangle^3 \right) \\
&\propto \mathcal{O}(N^{-1}).
\end{aligned} \tag{5.30}$$

There are also $\mathcal{O}(N^2)$ terms $u_{q_1} u_{q_2} u_{\bar{q}_1 + \bar{q}_2}$ with $q_1 \neq \Gamma$, $q_2 \neq \Gamma$ and $q_1 \neq \bar{q}_2$, whose expectations are

$$\begin{aligned}
\langle u_{q_1} u_{q_2} u_{\bar{q}_1 + \bar{q}_2} \rangle &= \left\langle \frac{1}{N^3} \sum_{R_1 R_2 R_3} d_{R_1} d_{R_2} d_{R_3} e^{2\pi i (q_1 (R_2 - R_1) + q_2 (R_3 - R_1))} \right\rangle \\
&= \frac{1}{N^3} \sum_{R_1} \left(\langle d^3 \rangle + \sum_{R_2 \neq R_1} \left(e^{2\pi i q_1 (R_2 - R_1)} + e^{2\pi i q_2 (R_2 - R_1)} + e^{2\pi i (q_1 + q_2) (R_2 - R_1)} \right) \langle d^2 \rangle \langle d \rangle \right. \\
&\quad \left. + \sum_{R_2 \neq R_1} \sum_{R_3 \neq R_1, R_2} e^{2\pi i q (-R_2 + R_3)} \langle d \rangle^3 \right) \\
&= \frac{1}{N^2} \left(\langle d \rangle^3 - 3 \langle d^2 \rangle \langle d \rangle - \langle d \rangle^3 \right) \\
&\propto \mathcal{O}(N^{-2}).
\end{aligned} \tag{5.31}$$

To summarise, in the limit of large N the expectation values of the monomials are:

monomial	expectation value	scaling of monomial	no. of monomials
u_Γ	$\langle d \rangle$	$\mathcal{O}(N^0)$	$\mathcal{O}(N^0)$
$u_\Gamma u_\Gamma$	$\langle d \rangle^2$	$\mathcal{O}(N^0)$	$\mathcal{O}(N^0)$
$u_q u_{\bar{q}}$	$\frac{1}{N} (\langle d^2 \rangle - \langle d \rangle^2)$	$\mathcal{O}(N^{-1})$	$\mathcal{O}(N^1)$
$u_\Gamma u_\Gamma u_\Gamma$	$\langle d \rangle^3$	$\mathcal{O}(N^0)$	$\mathcal{O}(N^0)$
$u_\Gamma u_q u_{\bar{q}}$	$\frac{1}{N} \langle d^2 \rangle \langle d \rangle$	$\mathcal{O}(N^{-1})$	$\mathcal{O}(N^1)$
$u_{q_1} u_{q_2} u_{\bar{q}_1 + \bar{q}_2}$	$\frac{1}{N^2} (\langle d \rangle^3 - 3 \langle d^2 \rangle \langle d \rangle - \langle d \rangle^3)$	$\mathcal{O}(N^{-2})$	$\mathcal{O}(N^2)$

(5.32)

5.3.2 The principal axes approximation

If the PES is mapped in the primitive cell, then the Hamiltonian is

$$\mathcal{H} = \frac{1}{2} \frac{\partial^2}{\partial u_\Gamma^2} + V^{(2)} d_\Gamma^2 + V^{(3)} d_\Gamma^3 + V^{(4)} d_\Gamma^4. \quad (5.33)$$

Taking the principal axes approximation (PAA) at first order does not drop any terms, as the primitive cell of this system only contains one normal mode.

Since no terms have been dropped, if this PES is interpolated to a larger supercell then the full Hamiltonian is retained,

$$\mathcal{H} = \sum_{q_1} \left(-\frac{1}{2N} \frac{\partial^2}{\partial u_{q_1} \partial u_{\bar{q}_1}} + NV^{(2)} u_{q_1} u_{\bar{q}_1} \right) + \sum_{q_1 q_2} NV^{(3)} u_{q_1} u_{q_2} u_{\bar{q}_1 + \bar{q}_2} + \sum_{q_1 q_2 q_3} NV^{(4)} u_{q_1} u_{q_2} u_{q_3} u_{\bar{q}_1 + \bar{q}_2 + \bar{q}_3}. \quad (5.34)$$

In contrast, if the PES is mapped directly in a larger supercell then, taking the PAA at the first order, the Hamiltonian will instead be

$$\mathcal{H} = \sum_{q_1} \left(-\frac{1}{2N} \frac{\partial^2}{\partial u_{q_1} \partial u_{\bar{q}_1}} + NV^{(2)} u_{q_1} u_{\bar{q}_1} \right) + NV^{(3)} u_\Gamma^3 + \sum_{q_1} NV^{(4)} (u_{q_1} u_{\bar{q}_1})^2. \quad (5.35)$$

This inconsistency arises because the PAA is not size-consistent and does not behave well under interpolation. In addition, the expectation value $\langle (u_{q_1} u_{\bar{q}_1})^2 \rangle$ scales as $\mathcal{O}(N^{-2})$, but the sum $\sum_{q_1} NV^{(4)} (u_{q_1} u_{\bar{q}_1})^2$ contains only $\mathcal{O}(N^1)$ terms, so the contribution of this sum to the free energy is not extensive.

A purely quartic potential

The problems with the lack of size consistency can be emphasised by removing the harmonic and cubic terms from the potential of the dispersionless anharmonic oscillator, leaving a purely quartic potential. Taking the PAA at the first order, the Hamiltonian of this system is then

$$\mathcal{H} = \sum_q \left(-\frac{1}{2N} \frac{\partial^2}{\partial u_q \partial u_{\bar{q}}} + NV^{(4)}(u_q u_{\bar{q}})^2 \right). \quad (5.36)$$

Since the Hamiltonian is dispersionless, under the effective harmonic approximation all modes will have the same frequency, ω . Using the results derived in appendix A, the zero-temperature free energy under this approximation is then

$$\begin{aligned} F &= \langle \mathcal{H} \rangle \\ &= \sum_q \left(\frac{1}{2N} \frac{N\omega}{2} + NV^{(4)} \frac{2}{(2N\omega)^2} \right) \\ &= \frac{N\omega}{4} + \frac{V^{(4)}}{2\omega^2}. \end{aligned} \quad (5.37)$$

This is minimised when the effective frequency is $\omega = \left(\frac{V^{(4)}}{N} \right)^{1/3}$, giving a free energy $F = \frac{3}{4} N^{2/3} (V^{(4)})^{1/3}$. Since the effective frequency should be intensive and the free energy should be extensive, this demonstrates that the PAA is clearly unphysical.

5.3.3 The VSCF Hamiltonian

The VSCF Hamiltonian consists of the Γ -point Hamiltonian,

$$\begin{aligned} \mathcal{H}_\Gamma &= -\frac{1}{2N} \frac{\partial^2}{\partial u_\Gamma^2} + NV^{(2)} u_\Gamma^2 \\ &\quad + NV^{(3)} u_\Gamma^3 + \sum_q 3NV^{(3)} \langle u_q u_{\bar{q}} \rangle u_\Gamma \\ &\quad + NV^{(4)} u_\Gamma^4 + \sum_q 6NV^{(4)} \langle u_q u_{\bar{q}} \rangle u_\Gamma^2 \end{aligned} \quad (5.38)$$

and the q-point Hamiltonians,

$$\begin{aligned} \mathcal{H}_q &= -\frac{1}{2N} \frac{\partial^2}{\partial u_q \partial u_{\bar{q}}} + NV^{(2)} u_q u_{\bar{q}} \\ &\quad + 3NV^{(3)} \langle u_\Gamma \rangle u_q u_{\bar{q}} \\ &\quad + 3NV^{(4)} u_q^2 u_{\bar{q}}^2 + 6NV^{(4)} \langle u_\Gamma^2 \rangle u_q u_{\bar{q}} + \sum_{q_2 \neq q} 6NV^{(4)} \langle u_{q_2} u_{\bar{q}_2} \rangle u_q u_{\bar{q}}. \end{aligned} \quad (5.39)$$

It should be noted that all terms which cannot be defined purely as products of u_Γ and $u_q u_{\bar{q}}$ are non-VSCF terms and do not appear in the VSCF single-mode Hamiltonians.

Substituting in the polynomial expectation values, and defining $\mu = \langle d \rangle$ and $\sigma^2 = \langle d^2 \rangle - \langle d \rangle^2$, the single-mode Hamiltonians become

$$\begin{aligned} \mathcal{H}_\Gamma = & -\frac{1}{2N} \frac{\partial^2}{\partial u_\Gamma^2} + NV^{(2)} u_\Gamma^2 \\ & + NV^{(3)} u_\Gamma^3 + 3(N-1)V^{(3)} \sigma^2 u_\Gamma \\ & + NV^{(4)} u_\Gamma^4 + 6(N-1)V^{(4)} \sigma^2 u_\Gamma^2 \end{aligned} \quad (5.40)$$

and

$$\begin{aligned} \mathcal{H}_q = & -\frac{1}{2N} \frac{\partial^2}{\partial u_q \partial u_{\bar{q}}} + NV^{(2)} u_q u_{\bar{q}} \\ & + 3NV^{(3)} \mu u_q u_{\bar{q}} \\ & + 3NV^{(4)} u_q^2 u_{\bar{q}}^2 + 6V^{(4)} (\sigma^2 + N\mu^2) u_q u_{\bar{q}} + 6(N-3)V^{(4)} \sigma^2 u_q u_{\bar{q}}. \end{aligned} \quad (5.41)$$

Taking the large- N limit, these become

$$\mathcal{H}_\Gamma = -\frac{1}{2N} \frac{\partial^2}{\partial u_\Gamma^2} + 3NV^{(3)} \sigma^2 u_\Gamma + N \left(V^{(2)} + 6V^{(4)} \sigma^2 \right) u_\Gamma^2 + NV^{(3)} u_\Gamma^3 + NV^{(4)} u_\Gamma^4 \quad (5.42)$$

and

$$\mathcal{H}_q = -\frac{1}{2N} \frac{\partial^2}{\partial u_q \partial u_{\bar{q}}} + N \left(V^{(2)} + 3V^{(3)} \mu + 6V^{(4)} \mu^2 + 6V^{(4)} \sigma^2 \right) u_q u_{\bar{q}} + 3NV^{(4)} u_q^2 u_{\bar{q}}^2. \quad (5.43)$$

Changing co-ordinates to $\tilde{u}_\Gamma = u_\Gamma - \mu$, the Γ -point Hamiltonian becomes

$$\begin{aligned} \mathcal{H}_\Gamma = & -\frac{1}{2N} \frac{\partial^2}{\partial u_\Gamma^2} \\ & + N \left(2\mu V^{(2)} + 3V^{(3)} (\mu^2 + \sigma^2) + 4V^{(4)} \mu (\mu^2 + 3\sigma^2) \right) \tilde{u}_\Gamma \\ & + N \left(V^{(2)} + 3V^{(3)} \mu + 6V^{(4)} (\mu^2 + \sigma^2) \right) \tilde{u}_\Gamma^2 \\ & + N \left(V^{(3)} + 4V^{(4)} \mu \right) \tilde{u}_\Gamma^3 \\ & + NV^{(4)} \tilde{u}_\Gamma^4, \end{aligned} \quad (5.44)$$

The expectation of u_Γ will be that which minimises the potential term of \mathcal{H}_Γ . This will be the value which sets the coefficient of \tilde{u}_Γ to zero. As such,

$$2\mu V^{(2)} + 3V^{(3)} (\mu^2 + \sigma^2) + 4V^{(4)} \mu (\mu^2 + 3\sigma^2) = 0. \quad (5.45)$$

The scaling with N of the expectation values of the monomials involving \tilde{u}_Γ is different to that of the

monomials involving u_r . These are given by

$$\begin{aligned}
\langle \tilde{u}_r \rangle &= 0 \\
\langle \tilde{u}_r^2 \rangle &\propto \mathcal{O}(N^{-1}) \\
\langle \tilde{u}_r^3 \rangle &= 0 \\
\langle \tilde{u}_r^4 \rangle &\propto \mathcal{O}(N^{-2}).
\end{aligned}
\tag{5.46}$$

In the limit of large N , the contribution of any terms whose expectations decay as $\mathcal{O}(N^{-1})$ or faster will have no effect on the free energy calculation and should be neglected. As such, the single-subspace Hamiltonians will become

$$\mathcal{H}_r = -\frac{1}{2N} \frac{\partial^2}{\partial u_r^2} + N \left(V^{(2)} + 3V^{(3)}\mu + 6V^{(4)}(\mu^2 + \sigma^2) \right) \tilde{u}_r^2
\tag{5.47}$$

and

$$\mathcal{H}_q = -\frac{1}{2N} \frac{\partial^2}{\partial u_q \partial u_{\bar{q}}} + N \left(V^{(2)} + 3V^{(3)}\mu + 6V^{(4)}(\mu^2 + \sigma^2) \right) u_q u_{\bar{q}}.
\tag{5.48}$$

And so in the large- N limit the VSCF Hamiltonian becomes an effective harmonic Hamiltonian, where the equilibrium is offset from the static-lattice co-ordinate by $\langle u_r \rangle = \mu$ and where every mode has the same effective harmonic frequency, satisfying $\frac{1}{2}\omega^2 = V^{(2)} + 3V^{(3)}\mu + 6V^{(4)}(\mu^2 + \sigma^2)$.

5.4 Calculating free energy

Once the VSCF Hamiltonian has been calculated, the final steps of the free energy calculation are to calculate the eigenstates of the Hamiltonian, and to use statistical mechanics to calculate the free energy from the eigenstates.

As detailed in section 3.5.4, under VSCF the free energy of the whole crystal is the sum of the free energy in each subspace. The calculation of the free energy within each subspace is equivalent to a full vibrational free energy calculation, except that this calculation is contained within a single subspace and the single-subspace PES has already been calculated. There are many methods for calculating the vibrational free energy, as detailed in section 3.5. As each subspace has a much lower dimensionality than the entire crystal, it is possible to use highly accurate methods to calculate each single-subspace free energy, without these calculations becoming excessively computationally costly.

Out of the systematically improvable free energy methods, the most natural choice for calculating the free energy in each subspace is VCI, as detailed in section 3.5.6. VCI allows for finite-temperature calculations with significantly less complexity than VCC, and there is no stochastic element to VCI calculations, unlike

methods based on MD and MC.

Under VCI, the free energy in each subspace is calculated by first constructing the single-subspace Hamiltonian in an arbitrary basis of states, typically chosen as a subset of the states which diagonalise an effective harmonic Hamiltonian as described in section 3.5.3. The Hamiltonian in this representation is simply a matrix, which can be diagonalised numerically to give the single-subspace eigenstates and eigenenergies. The free energy can then be calculated from the eigenenergies using statistical mechanics. The number of states in the basis must be increased until the free energy converges.

In previous implementations of VSCF, where the VSCF Hamiltonian is a sum of single-mode Hamiltonians rather than single-subspace Hamiltonians, each VCI calculation is performed in the one-dimensional spaces described by the individual normal modes. These calculations are straightforward, and computationally inexpensive. However, when the VSCF Hamiltonian is constructed as a sum of single-subspace Hamiltonians, where each subspace contains the degenerate normal modes at a given q-point and its paired q-point, it is necessary to perform VCI in every subspace, and the dimensionality of the VCI calculation in a given subspace will be the same as the dimensionality of that subspace. These subspaces are regularly four- and six-dimensional, and eight-dimensional subspaces are possible.

The number of basis states in an n_d -dimensional subspace where n_s basis states are required along each mode scales as $\mathcal{O}(n_s^{n_d})$. The cost of performing VCI is cubic in the number of basis states, and so scales as $\mathcal{O}(n_s^{3n_d})$. The number of states which are required along each mode, n_s , varies with the temperature at which the calculation is being run, and with the details of the single-subspace Hamiltonian, but is typically a few tens of states. Thus, if the dimensionality of the subspace, n_d , is six or eight then this can become an extremely computationally costly calculation.

In order to make this calculation possible at reasonable computational expense, it is necessary to accelerate the method. A number of such accelerations will be detailed in the remainder of this section.

5.4.1 Analytic states and integrals

The eigenstates in a given subspace have a dimensionality equal to the dimensionality of that subspace, and the number of states which must be included in each subspace in order for the free energy calculation to converge scales exponentially with the dimensionality of that subspace. When the dimensionality of a subspace becomes large, it is possible that thousands or even millions of states are required. As such, it is necessary to represent the single-subspace eigenstates efficiently, such that the calculations required when constructing and working with these states do not become intractable.

This work represents the single-subspace eigenstates as linear combinations of the eigenstates of the effective

harmonic Hamiltonian as described in section 3.5.3. The expectation values of the Hamiltonian, and any other operators which can be represented using truncated Taylor expansions, can be calculated with respect to harmonic eigenstates using analytic results, rather than having to perform computationally expensive numerical integration. The details of calculating these analytic expectation values are presented in appendix A.

5.4.2 Symmetry and single-subspace states

Separation of states by symmetry eigenvalues

Since the Hamiltonian commutes with every symmetry operator, $\mathcal{H}\mathcal{S} = \mathcal{S}\mathcal{H}$, any two states ψ and ϕ which are both eigenstates of a given symmetry \mathcal{S} but with different eigenvalues must obey $\langle\psi|\mathcal{H}|\phi\rangle = 0$, since

$$\begin{aligned} \langle\psi|\mathcal{S}\mathcal{H}|\phi\rangle &= \langle\psi|\mathcal{H}\mathcal{S}|\phi\rangle \\ \implies \lambda_\psi \langle\psi|\mathcal{H}|\phi\rangle &= \lambda_\phi \langle\psi|\mathcal{H}|\phi\rangle . \end{aligned} \tag{5.49}$$

In general, the crystal symmetry operators do not all commute with one another, and so it is not possible to construct the basis states such that they diagonalise every symmetry operator. However, it is always possible to select a subset of the symmetry operators which do mutually commute, and construct the basis states so that they are eigenstates of these mutually commuting symmetry operators. The Hamiltonian in this basis is then block-diagonal, as it may only have non-zero off-diagonal elements between states which have the same eigenvalue with respect to every symmetry operator in the mutually commuting subset.

As the Hamiltonian is block-diagonal, it is only necessary to diagonalise the Hamiltonian within each block. If the Hamiltonian contains n approximately equally-sized blocks, this reduces the computational cost by a factor of $\mathcal{O}(n^2)$. A consequence of block diagonalisation is that the eigenstates calculated from each block of basis states will have the same eigenvalues with respect to the mutually commuting symmetry operators as the basis states in that block. This can be used to reduce the cost of subsequent calculations, including calculating the properties of the crystal from its eigenstates and performing VMP n .

The purely translational symmetries should always be included in the mutually commuting subset of symmetries. Doing this means that the basis states, and subsequently the single-subspace eigenstates, are Bloch states with well-defined wavevectors. Since the VSCF states are products of the single-subspace eigenstates, the VSCF states are also Bloch states. Conveniently, the harmonic eigenstates are already Bloch states,

satisfying

$$\mathcal{S}_{\mathbf{R}} |\{p_\chi\}\rangle = \left(\prod_{\chi} e^{2\pi i p_\chi \mathbf{q}_\chi \cdot \mathbf{R}} \right) |\{p_\chi\}\rangle, \quad (5.50)$$

where $|\{p_\chi\}\rangle$ is the harmonic eigenstate with occupation p_χ along each mode χ , and $\mathcal{S}_{\mathbf{R}}$ is the symmetry operator corresponding to a pure translation along \mathbf{R} -vector \mathbf{R} .

Separation of states by q-point occupations

As detailed in appendix A, the Hamiltonian can be written as

$$\mathcal{H} = \sum_{\{n_\chi^\dagger, n_\chi\}} h_{\{n_\chi^\dagger, n_\chi\}} (a_\chi^\dagger)^{n_\chi^\dagger} (a_\chi)^{n_\chi}, \quad (5.51)$$

where a_χ^\dagger and a_χ are the harmonic creation and annihilation operators respectively along mode χ , n_χ^\dagger and n_χ are non-negative integer powers, and $h_{\{n_\chi^\dagger, n_\chi\}}$ are scalar coefficients.

The Hamiltonian element between two harmonic eigenstates, $\langle \{p_\chi\} | \mathcal{H} | \{p'_\chi\} \rangle$, may only be non-zero if the Hamiltonian contains at least one non-zero coefficient $h_{\{n_\chi^\dagger, n_\chi\}}$ corresponding to a term which satisfies $n_\chi^\dagger - n_\chi = p_\chi - p'_\chi$ for all χ . This is relevant because there are two constraints on the Hamiltonian which make the Hamiltonian in this basis extremely sparse, and which at some q-points will significantly reduce the size of the blocks in the block-diagonal structure.

Firstly, if the PES is constructed from a Taylor expansion truncated at order n , then all terms in the Hamiltonian must satisfy $\sum_{\chi} (n_\chi^\dagger + n_\chi) \leq n$, and so $\langle \{p_\chi\} | \mathcal{H} | \{p'_\chi\} \rangle$ may only be non-zero if $\sum_{\chi} |p_\chi - p'_\chi| \leq n$. This means that in an n_d -dimensional subspace each row and column in the Hamiltonian will contain at most $\mathcal{O}(n^{n_d})$ non-zero elements. The total number of states in the basis scales as $\mathcal{O}(n_s^{n_d})$, where the number of states along each mode, n_s , is typically at least an order of magnitude larger than n . As such, the Hamiltonian will typically be extremely sparse.

Secondly, the single-subspace Hamiltonians may only contain terms which are invariant under translational symmetry, i.e. where $\sum_{\chi} (n_\chi^\dagger - n_\chi) \mathbf{q}_\chi$ is a \mathbf{G} -vector. As previously mentioned, every q-point has an integer $n_{\mathbf{q}}$ which is the smallest positive integer such that $n_{\mathbf{q}} \mathbf{q}$ is a \mathbf{G} -vector, and the star of q-points in a given subspace all have the same value of $n_{\mathbf{q}}$. If $n_{\mathbf{q}} > n$, then the only terms which may appear in the single-subspace Hamiltonian are those terms which are constructed as products of pairs of creation and annihilation operators where each pair is individually translationally invariant. These pairs either must contain two creation operators, $a_{\chi_1}^\dagger a_{\chi_2}^\dagger$, or two annihilation operators, $a_{\chi_1} a_{\chi_2}$, in either case satisfying $\mathbf{q}_1 = \bar{\mathbf{q}}_2$, or one creation operator and one annihilation operator, $a_{\chi_1} a_{\chi_2}^\dagger$ or $a_{\chi_1}^\dagger a_{\chi_2}$, satisfying $\mathbf{q}_1 = \mathbf{q}_2$.

Each state $|\{p_\chi\}\rangle$ can be characterised according to its occupation at each q-point, with the q-point occu-

pation $p_{\mathbf{q}}$ defined in terms of the single-mode occupations p_{χ} as

$$p_{\mathbf{q}} = \sum_{\substack{\chi \\ \mathbf{q}_{\chi}=\mathbf{q}}} p_{\chi} - \sum_{\substack{\chi \\ \mathbf{q}_{\chi}=\bar{\mathbf{q}}}} p_{\chi}. \quad (5.52)$$

The action of a translationally-invariant pair of creation and annihilation operators on such a state leaves the \mathbf{q} -point occupations invariant, and so if $n_{\mathbf{q}} > n$ the action of the VSCF Hamiltonian on such a state also leaves the \mathbf{q} -point occupations invariant. This means that the Hamiltonian is block-diagonal with blocks separated by the \mathbf{q} -point occupations of the basis states, in addition to the blocks being separated by the symmetry eigenvalues of those states. This further reduces the computational cost of calculating and working with the VSCF eigenstates.

5.4.3 Converging the basis of states

The free energy can be written as a sum of contributions from each eigenstate,

$$F = \sum_{\psi} P_{\psi} (\langle \psi | \mathcal{H} | \psi \rangle + T \ln(P_{\psi})) . \quad (5.53)$$

Increasing the number of states in the basis adds more terms to this sum, and also changes the values of the thermal weightings $\{P_{\psi}\}$.

The free energy is the sum of two terms. The first term, $\sum_{\psi} P_{\psi} \langle \psi | \mathcal{H} | \psi \rangle$, is the internal energy, and can generally be well-represented by a small basis of states. The second term, $\sum_{\psi} P_{\psi} T \ln(P_{\psi})$, is entropic, and requires a much larger basis of states to model accurately, as the sum depends heavily on the eigenstates with higher energies.

It is often not practical to include sufficient states in the VCI basis such that the free energy converges. This problem can be avoided by partially accounting for those states which are not included in the VCI basis, by treating these additional states implicitly, under the effective harmonic approximation. By including all states in this manner, the number of states which needs to be included in the VCI basis before the free energy converges is reduced. A method which allows for the implicit inclusion of these states is detailed below.

The explicit basis and the implicit basis

The harmonic eigenstates form a complete basis for the space. From this basis, it is only possible to consider a small number of the basis states explicitly. The remaining states can be considered, but since there are an infinite number of them they can only be considered implicitly. It is convenient to denote the explicitly

considered basis of states as $\{\psi\}^{\text{exp}}$, the implicitly considered basis as $\{\psi\}^{\text{imp}}$, and the complete basis which is the union of the two as $\{\psi\}^{\text{union}}$.

Treating the free energy as a functional of the Hamiltonian and the eigenstates, $F[\mathcal{H}, \{\psi_\lambda\}]$, the free energy F is that which minimises $F[\mathcal{H}, \{\psi_\lambda\}]$ with respect to the set of eigenstates $\{\psi_\lambda\}$. Performing this minimisation while only considering states in $\{\psi\}^{\text{exp}}$ thus gives an upper bound on F . Any inclusion of $\{\psi\}^{\text{imp}}$ will also give an upper bound on F , and so if this upper bound is lower than that which only considers $\{\psi\}^{\text{exp}}$ then it must be closer to the true value of F .

Since $\{\psi\}^{\text{imp}}$ contains an infinite number of states, it is not possible to diagonalise the Hamiltonian in this basis. However, it is possible to include the states in $\{\psi\}^{\text{imp}}$ without diagonalisation, approximating the eigenstates of \mathcal{H} as the effective harmonic eigenstates. This will in general be a poor approximation, but it will nonetheless be a better approximation than not including these states at all.

Thermodynamic calculus

There are two eigenbases which can be constructed in each of the three basis sets: the VSCF eigenbasis $\{\psi\}_{\text{vscf}}$ which diagonalises the VSCF Hamiltonian $\mathcal{H}_{\text{vscf}}$, and the harmonic eigenbasis $\{\psi\}_{\text{har}}$ which diagonalises a harmonic or effective harmonic Hamiltonian \mathcal{H}_{har} . Since the expectation of an operator can be calculated with respect to the complete harmonic eigenbasis, it is possible to perform calculations directly in both the explicit and complete bases $\{\psi\}_{\text{har}}^{\text{exp}}$ and $\{\psi\}_{\text{har}}^{\text{union}}$. However, when using the VSCF eigenbasis it is only possible to perform calculations directly in the explicit basis $\{\psi\}_{\text{vscf}}^{\text{exp}}$.

Appendix B details how to calculate any one of $F[\mathcal{H}, \{\psi\}^{\text{exp}}]$, $F[\mathcal{H}, \{\psi\}^{\text{imp}}]$ and $F[\mathcal{H}, \{\psi\}^{\text{union}}]$ given the values of the other two, under the approximation that there is no state mixing between the explicit and implicit bases. For the case of the harmonic eigenbases this approximation will be exact, since there is no such state mixing between these bases.

A basis which is likely to give the best upper bound on F is the union of the VSCF eigenbasis in the explicit basis and the harmonic eigenbasis in the implicit basis, $F[\mathcal{H}_{\text{vscf}}, \{\psi\}_{\text{vscf}}^{\text{exp}} + \{\psi\}_{\text{har}}^{\text{imp}}]$. However, this free energy cannot be calculated directly, and instead must be calculated using a series of thermodynamic additions and subtractions.

Firstly, it is necessary to calculate the partition function of the harmonic eigenbasis in the implicit basis, as

$$Z_{\text{har}}^{\text{imp}} = Z_{\text{har}}^{\text{union}} - Z_{\text{har}}^{\text{exp}}, \quad (5.54)$$

where Z is the partition function, which related to the free energy by

$$F[\mathcal{H}_{\text{har}}, \{\psi\}_{\text{har}}^{\text{basis}}] = -T \ln(Z_{\text{har}}^{\text{basis}}). \quad (5.55)$$

This can then be used to calculate the expectation values of the harmonic and VSCF Hamiltonians in the harmonic eigenbasis in the implicit basis,

$$\begin{aligned} \langle \mathcal{H}_{\text{har}} \rangle_{\text{har}}^{\text{imp}} &= \left(Z_{\text{har}}^{\text{union}} \langle \mathcal{H}_{\text{har}} \rangle_{\text{har}}^{\text{union}} - Z_{\text{har}}^{\text{exp}} \langle \mathcal{H}_{\text{har}} \rangle_{\text{har}}^{\text{exp}} \right) / Z_{\text{har}}^{\text{imp}} \\ \langle \mathcal{H}_{\text{vscf}} \rangle_{\text{har}}^{\text{imp}} &= \left(Z_{\text{har}}^{\text{union}} \langle \mathcal{H}_{\text{vscf}} \rangle_{\text{har}}^{\text{union}} - Z_{\text{har}}^{\text{exp}} \langle \mathcal{H}_{\text{vscf}} \rangle_{\text{har}}^{\text{exp}} \right) / Z_{\text{har}}^{\text{imp}}. \end{aligned} \quad (5.56)$$

The entropy of the harmonic eigenbasis in the implicit basis is calculated as

$$S_{\text{har}}^{\text{imp}} = \left(F[\mathcal{H}_{\text{har}}, \{\psi\}_{\text{har}}^{\text{imp}}] - \langle \mathcal{H}_{\text{har}} \rangle_{\text{har}}^{\text{imp}} \right) / T, \quad (5.57)$$

which can be combined with the expectation of the VSCF Hamiltonian in this basis to give the free energy of the VSCF Hamiltonian in this basis,

$$F[\mathcal{H}_{\text{vscf}}, \{\psi\}_{\text{har}}^{\text{imp}}] = \langle \mathcal{H}_{\text{vscf}} \rangle_{\text{har}}^{\text{imp}} - T S_{\text{har}}^{\text{imp}}. \quad (5.58)$$

Finally, this is combined with the directly-calculated free energy of the VSCF Hamiltonian in the VSCF basis in the explicit basis, to give the free energy of the VSCF Hamiltonian in the combined basis, via

$$\begin{aligned} \exp\left(-F[\mathcal{H}_{\text{vscf}}, \{\psi\}_{\text{vscf}}^{\text{exp}} + \{\psi\}_{\text{har}}^{\text{imp}}]/T\right) &= \exp(-F[\mathcal{H}_{\text{vscf}}, \{\psi\}_{\text{vscf}}^{\text{exp}}]/T) \\ &+ \exp\left(-F[\mathcal{H}_{\text{vscf}}, \{\psi\}_{\text{har}}^{\text{imp}}]/T\right). \end{aligned} \quad (5.59)$$

5.5 Converging the self-consistent coefficients

The VSCF Hamiltonian is defined by two equations. Firstly, the free energy of the full Hamiltonian under the eigenstates of the VSCF Hamiltonian is a minimum with respect to the coefficients of the VSCF Hamiltonian,

$$\frac{\delta F[\mathcal{H}, \mathcal{H}']}{\delta \mathcal{H}'} = 0. \quad (5.60)$$

Secondly, the VSCF Hamiltonian is self-consistent, i.e. the single-subspace VSCF Hamiltonians \mathcal{H}'_{ξ} are

equal to the single-subspace Hamiltonians generated when taking the expectation of the full Hamiltonian with respect to the eigenstates of the VSCF Hamiltonian,

$$\mathcal{H}'_{\xi} = \langle \mathcal{H} \rangle_{\{\xi' \neq \xi\}} + \text{constant} . \quad (5.61)$$

The VSCF Hamiltonian can be written as a linear combination of single-subspace basis functions,

$$\mathcal{H}'_{\xi} = \mathcal{T}_{\xi} + \sum_j x_{\xi j} \mu_j , \quad (5.62)$$

where \mathcal{T}_{ξ} is the kinetic energy operator in subspace ξ , each μ_j is a basis function, and each $x_{\xi j}$ is a real scalar coefficient. For a given set of basis functions, this means that the Hamiltonian can be parameterised entirely by the vector of basis function coefficients $\mathbf{x} = \{x_{\xi j}\}$. The free energy can also be calculated as a function of \mathbf{x} only; $F[\mathcal{H}] \equiv F(\mathbf{x})$.

Similarly, the self-consistent Hamiltonian can be parameterised in terms of the same basis functions to give

$$\mathcal{H}_{\xi} = T_{\xi} + \sum_j y_{\xi j} \mu_j . \quad (5.63)$$

The coefficient vector $\mathbf{y} = \{y_{\xi j}\}$ is again a function of \mathbf{x} only. For clarity, the coefficient vector \mathbf{x} will be referred to as the *input vector*, and \mathbf{y} will be referred to as the *output vector*. It is also convenient to define the *error vector*, $\mathbf{e}(\mathbf{x}) = \mathbf{y}(\mathbf{x}) - \mathbf{x}$.

Thus the conditions for finding the VSCF Hamiltonian can be simplified to the self-consistency condition

$$\mathbf{y}(\mathbf{x}) = \mathbf{x} \quad (5.64)$$

or equivalently

$$\mathbf{e}(\mathbf{x}) = 0 , \quad (5.65)$$

and the free energy minimisation condition

$$\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}} = 0 . \quad (5.66)$$

It should be noted that self-consistency is a necessary condition for free energy minimisation, but is not in general a sufficient condition.

Most methods for finding the value of \mathbf{x} which minimises the free energy will proceed iteratively, with each iteration starting from some value \mathbf{x}_k , calculating one or both of $\mathbf{y}_k = \mathbf{y}(\mathbf{x}_k)$ and $\frac{\partial F}{\partial \mathbf{x}_k}$, and using these to generate the next iteration \mathbf{x}_{k+1} . This continues until the conditions on \mathbf{x} are satisfied to within some

pre-determined tolerance.

5.5.1 Existing methods

There are a wide range of methods available for finding self-consistent solutions to free energy minimisation problems, many of which are primarily developed in the context of DFT [140]. These self-consistency methods can be categorised into two broad categories: *mixing schemes*, which calculate \mathbf{x}_{k+1} as a function of previous values of \mathbf{x}_k , and *direct minimisation schemes*, which use $\frac{\partial F}{\partial \mathbf{x}_k}$ and other details of the free energy landscape to move each iteration closer to the free energy minimum.

Direct minimisation schemes generally require the calculation of the gradient vector $\frac{\partial F}{\partial \mathbf{x}}$, and some direct minimisation schemes require higher derivatives of the free energy with respect to \mathbf{x} . In the context of VSCF this gradient vector can be calculated using perturbation theory, and the cost of this calculation is comparable to the cost of calculating the output vector \mathbf{y} . The calculation of higher derivatives of the free energy with respect to \mathbf{x} requires the intermediate calculation of two-subspace potentials, and so the computational cost, both in terms of processing time and memory requirements, is significant.

Mixing schemes do not require significant additional calculation at each iteration, beyond the calculation of \mathbf{y} from \mathbf{x} . As such, it is preferable to use a mixing scheme as the primary method of converging the VSCF coefficients. It is possible to automatically monitor the progress of the coefficient convergence, and so it is possible to have a direct minimisation scheme available as a fallback option should the rate of convergence of the mixing scheme fall below an acceptable threshold.

Damped iterative schemes

One of the simplest mixing schemes is the *damped iterative scheme*, where each input vector is constructed as a linear combination of the previous input and output vectors,

$$\mathbf{x}_{k+1} = \gamma \mathbf{x}_k + (1 - \gamma) \mathbf{y}_k, \quad (5.67)$$

where γ is a damping constant between 0 and 1.

For cases where the anharmonicity is small, this has previously been found to yield reasonable convergence [26]. However, for cases with larger anharmonic contributions, the function $\mathbf{y}(\mathbf{x})$ is likely to be highly non-linear and may not be particularly smooth. This can mean that a simple damped iterative scheme will converge slowly or fail to converge entirely. In particular, there are cases where $\frac{\partial(\mathbf{y}-\mathbf{x})}{\partial \mathbf{x}}$ is negative along one or more directions in the vicinity of the solution $\mathbf{y} = \mathbf{x}$, and in these cases any iterative scheme will actively

diverge from the correct solution with each iteration.

Pulay schemes

A more sophisticated and reliable mixing scheme, which has found widespread success in the self-consistency schemes used in DFT, is the Pulay scheme [141, 142].

Given k previous iterations, the Pulay method constructs the input vector \mathbf{x}_{k+1} as a linear combination of the previous input vectors,

$$\mathbf{x}_{k+1} = \sum_k c_k \mathbf{x}_k, \quad (5.68)$$

where the real coefficients $\{c_k\}$ satisfy the condition $\sum_k c_k = 1$.

The Pulay method assumes that the error vector is related to the input vector with the relationship,

$$\mathbf{e}_k = A\mathbf{x}_k + \mathbf{e}_0, \quad (5.69)$$

where A is an unknown real tensor.

Under this assumption, the error vector will be

$$\mathbf{e}_{k+1} = \sum_k c_k \mathbf{e}_k. \quad (5.70)$$

Minimising $|\mathbf{e}_{k+1}|$ with respect to $\{c_k\}$, subject to the condition $\sum_k c_k = 1$, yields the symmetric matrix equation

$$\begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1 & \dots & \mathbf{e}_1 \cdot \mathbf{e}_n & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{e}_n \cdot \mathbf{e}_1 & \dots & \mathbf{e}_n \cdot \mathbf{e}_n & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_n \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (5.71)$$

which can be inverted directly to give \mathbf{x}_{k+1} .

Pulay schemes have two problems. The first is that each iteration is a linear combination of the previous iterations, so if the solution is not a linear combination of the first few iterations then it will never be found. This problem can be overcome by mixing in additional vectors to break out of the linearly dependent subspace. These vectors can be generated using a damped iterative scheme, and if that is not sufficient then alternative vectors can be generated using a direct minimisation scheme or random noise.

The second problem is that the Pulay scheme assumes that the output vector is approximately a linear

function of the input vector. If this is not approximately true then the scheme will fail.

In DFT, the problem is parameterised in terms of the electron density. This makes convergence quite straightforward, as the output electron density is generally a smooth and slowly-varying functional of the input electron density. This in turn means that the output vector is almost always an approximately linear function of the input vector, and so the Pulay scheme is a good choice.

For VSCF, this is not the case. Due to the extremely high dimensionality of the eigenstates, the problem must be parameterised in terms of the auxiliary Hamiltonian. At low temperatures, small variations in the auxiliary Hamiltonian can lead to large changes in the thermal weightings of the eigenstates. This means that the output coefficients are able to vary rapidly as a function of the input coefficients.

This problem is exemplified by considering the case of linearly-dependent input vectors. These lead to eigenvectors of the Pulay matrix which correspond to zero-length vectors in the space of output vectors, and to which the Pulay scheme may allocate large prefactors. In the case of an almost-linear relationship between the input and output vectors, this is unimportant; the corresponding vectors in the input space will be very short, and so even with a large prefactor they will not have a large impact on the next iteration. However, with a highly non-linear relationship these vectors may well correspond to large displacements in the input space, causing the vectors to vary wildly with each iteration, and causing the Pulay scheme to fail to converge.

Local vs global minimisation.

To find the best possible estimate of the free energy of the system within the constraints of VSCF, the minimisation of $F[\mathcal{H}, \mathcal{H}']$ should be treated as a global optimisation problem in the space of the degrees of freedom of \mathcal{H}' .

However, global optimisation is in general far too computationally expensive to justify its use in this context. Since the free energy is variational, treating the problem as a local optimisation starting from the effective harmonic solution will always provide a better estimate of the free energy than that given by the effective harmonic solution.

In cases where the locally-optimised VSCF solution is not sufficiently accurate, it seems unlikely that the increase in accuracy from globally-optimised VSCF would justify the increased computational costs. Instead, it is likely that more accurate approximations than VSCF would prove a better use of computational resources.

5.5.2 New developments

Before a Pulay scheme can be used to find the self-consistent VSCF coefficients, it must be modified so that the possibility of a highly non-linear dependence of the error vector \mathbf{e} on the input vector \mathbf{x} does not cause the convergence scheme to fail. This problem can be avoided by explicitly calculating the relationship between the error vector and the input vector.

Linearised weighted Pulay

Rather than assuming that the error vectors are related to the input vectors by the matrix equation $\mathbf{e}_k(\mathbf{x}) = A \cdot \mathbf{x}_k + \mathbf{e}_0$, it is possible to instead use linear regression to find the values of A and \mathbf{e}_0 which come closest to satisfying this relationship for $\{\mathbf{x}_k\}$. Once A and \mathbf{e}_0 have been found, they can be used to calculate \mathbf{x}_{k+1} directly.

Firstly, the vector space containing \mathbf{x} can be extended with an additional fictitious dimension. Each \mathbf{x}_k vector is extended into this additional dimension by a constant amount, l , to give extended vectors

$$\mathbf{x}' = (x_1, \dots, x_n, l). \quad (5.72)$$

The real parameter l can take any value other than 0, although for reasons of numerical stability l should be chosen to have a similar order of magnitude to the other components in the vectors. Since the \mathbf{x} vectors are constrained to have a coefficient of l along the fictitious dimension, the addition of the fictitious dimension has the effect of removing the origin, $\mathbf{x} = 0$ from the space of \mathbf{x} . This means the relationship between \mathbf{e} and \mathbf{x} can be expressed as a linear matrix equation. Finding the value of \mathbf{x} which minimises $|\mathbf{e}|$ is then a constrained linear problem rather than an unconstrained non-linear problem. This is depicted in figure 5.1.

The relationship between the error vectors and the extended input vectors is then

$$\mathbf{e}_k = A' \cdot \mathbf{x}'_k, \quad (5.73)$$

where A' is an unknown real tensor.

Without regularisation, the tensor A' can be found by minimising the weighted sum

$$\mathcal{L} = \sum_k w_k |\mathbf{e}_k - A' \cdot \mathbf{x}'_k|^2, \quad (5.74)$$

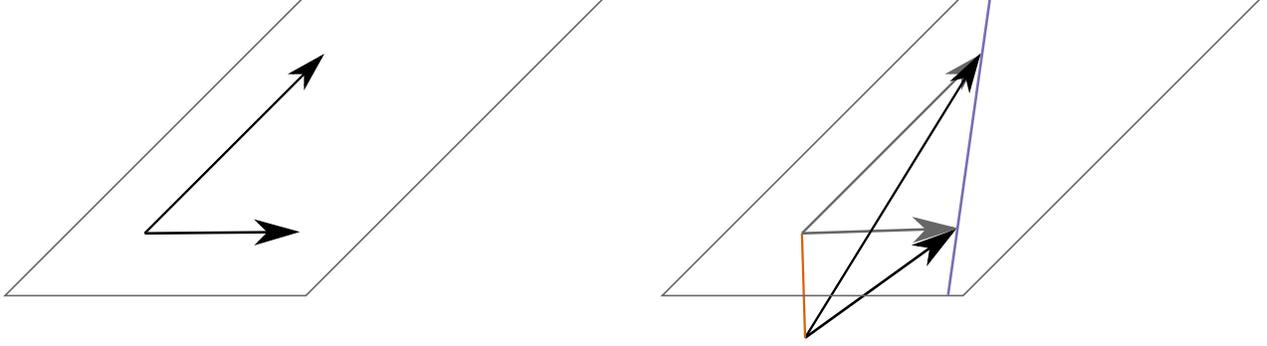


Figure 5.1: The arrows represent points \mathbf{x} . The grey rectangle represents the space in which \mathbf{x} may vary. Left: the origin lies in the space of \mathbf{x} . Any vector in the plane can be constructed as a linear combination of the \mathbf{x} vectors. The error vector \mathbf{e} cannot be a linear function of the \mathbf{x} vectors, as this function would be origin-dependent.

Right: the origin has been removed from the space of \mathbf{x} along the fictitious dimension, shown in orange. The only points in the plane which can be constructed as a linear combination of the \mathbf{x} vectors are those which lie along the turquoise line. The error vector \mathbf{e} may now be a linear function of the \mathbf{x} vectors as this function is origin-independent.

to give

$$A' \cdot \left(\sum_k w_k \mathbf{x}'_k \otimes \mathbf{x}'_k \right) = \sum_k w_k \mathbf{e}_k \otimes \mathbf{x}'_k. \quad (5.75)$$

The weights $\{w_k\}$ should be chosen such that the fit focusses on the region with small $|\mathbf{e}_k|$. A particularly natural choice of weights is $w_k = |\mathbf{e}_k|^{-2}$, as this means that the weighted norms of every error vector, $w_k |\mathbf{e}_k|^2$, will all be equal, and if the relationship between the error vectors \mathbf{e} and the extended input vectors \mathbf{x}' is approximately linear then the weighted norms of every extended input vector, $w_k |\mathbf{x}'_k|^2$, will be also all be approximately equal. This significantly improves the numerical stability of the algorithm.

Once A' is known, it can be used to find the value of \mathbf{x}' which minimises $|A' \cdot \mathbf{x}'|^2$ subject to the final component of \mathbf{x}' being l . The vector \mathbf{x}_{k+1} is then this value of \mathbf{x}' without the component in the fictitious dimension. This minimisation can be performed directly, but it is more convenient to construct the standard Pulay matrix equation but with each vector \mathbf{e}_k replaced by the corresponding vector $A' \cdot \mathbf{x}'_k$.

It should be noted that if the input vectors $\{\mathbf{x}_k\}$ are linearly independent, this method reduces to the standard Pulay scheme regardless of the weights chosen. In practice, where numerical noise almost always causes $\{\mathbf{x}_k\}$ to be linearly independent, regularisation must be used in order to bring any improvements beyond the standard Pulay scheme.

Regularisation

The linearised weighted Pulay scheme presented above has two related problems. The first is that the symmetric, positive semidefinite matrix $X = \sum_k w_k \mathbf{x}'_k \otimes \mathbf{x}'_k$ is not always invertible. The second is that the

fit can be extremely sensitive to small variations in \mathbf{x} .

This can be seen by taking an eigenvalue decomposition of X ,

$$X = \sum_l \lambda_l \mathbf{v}_l \otimes \mathbf{v}_l. \quad (5.76)$$

Since X is positive semidefinite, $\lambda_l \geq 0$.

The inverse matrix is then

$$X^{-1} = \sum_l \frac{1}{\lambda_l} \mathbf{v}_l \otimes \mathbf{v}_l, \quad (5.77)$$

which is ill-defined if $\lambda_l = 0$, and extremely sensitive to \mathbf{v}_l if λ_l is small.

Both of these problems can be addressed by introducing regularisation into the problem. Using L2-regularised ridge regression [143], the eigenvalues of X are replaced with

$$\lambda_l \rightarrow \lambda_l + \lambda, \quad (5.78)$$

where λ is a positive real regularisation parameter. This makes X positive definite, guaranteeing that it is invertible and removing the sensitivity to \mathbf{v}_l . However, this introduces new problems if $\frac{\partial \mathbf{e}}{\partial \mathbf{x}}$ has large elements, since in the absence of regularisation this would lead to large elements of A' , but these have been limited by the regularisation.

This can be mitigated by instead replacing the eigenvalues of X with

$$\lambda_l \rightarrow \lambda_l + \lambda e^{-\lambda_l/\lambda}. \quad (5.79)$$

In the limit of $\lambda_l \gg \lambda$, the eigenvalues are unchanged. In the limit of $\lambda_l \ll \lambda$, the eigenvalues are replaced with λ . In between these limits, the eigenvalue is replaced with a value which varies smoothly and monotonically from λ to λ_l . This has the effect of regularising those directions in which there is little projection of the \mathbf{x}'_k vectors, whilst leaving those directions with large projections from the \mathbf{x}'_k vectors unchanged. This in turn leads to stable inversion of the X matrix without artificially constraining its values.

Chapter 6

Introducing caesar

The improvements to the mapping of the PES and to the implementation of VSCF theory presented in chapters 4 and 5 respectively have been implemented in the software package `caesar`. `caesar` has been made publicly available with an open-source license, and is currently hosted at github.com/veryreverie/caesar.

Within the constraints detailed in chapter 2, `caesar` seeks to provide automated anharmonic free energy calculations for any static-lattice configuration, and to work with any electronic structure software which can sample the Born-Oppenheimer PES.

`caesar` performs calculations in two steps: a harmonic calculation, which is presented in figure 6.1, and an anharmonic calculation, which is presented in figure 6.2.

The inputs to the harmonic calculation are the crystal lattice matrix L , the static-lattice co-ordinate \mathbf{r}^0 and two sets of q-points $\{\mathbf{q}\}$; a coarse grid of q-points on which the PES will be mapped and a fine grid which will be used to calculate the free energy. From these, `caesar` uses the `spglib` package [144] to calculate the symmetry operators of the crystal $\{\mathcal{S}\}$, and the crystal is optionally snapped to this symmetry, a process which will be detailed in section 6.4.

The symmetry-invariant harmonic basis functions $\{\mu\}$ are generated on the coarse q-point grid, using the algorithm detailed in section 4.1.3, and a set of samples $\{\mathbf{r}_j\}$ which can be used to optimally fit these basis functions are generated, using the algorithm detailed in section 4.4.4. Electronic structure calculations are performed at each sample, which provide the value of the PES at each sample $V(\mathbf{r}_j)$, optionally along with forces $\mathbf{f}(\mathbf{r}_j)$ and other details of the PES, as will be detailed in section 6.1. The harmonic basis function coefficients are then fitted from the PES samples, as detailed in section 4.4. This gives the harmonic PES, $V(\mathbf{r})$.

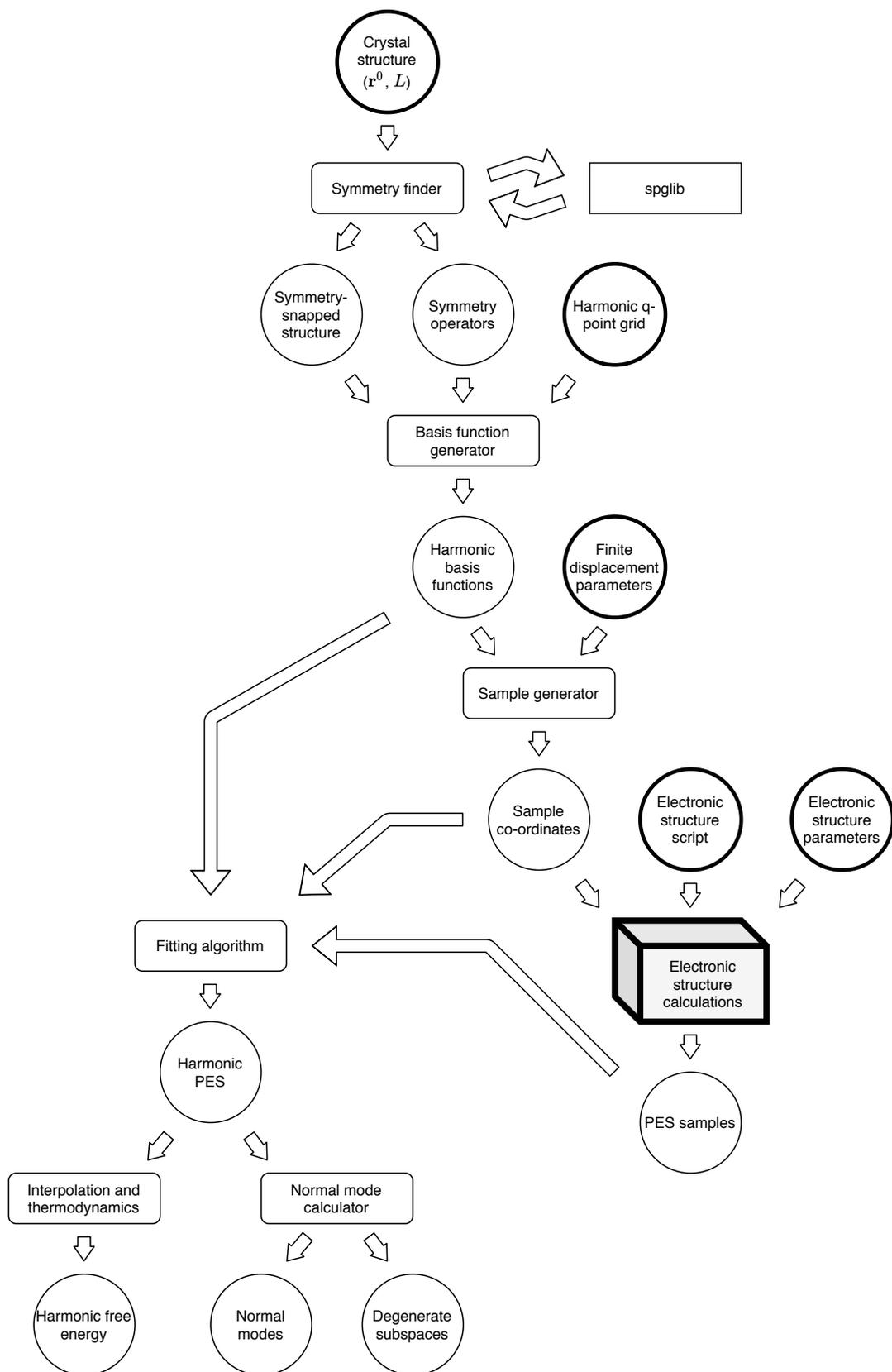


Figure 6.1: The harmonic calculation performed by **caesar**. Rectangles represent algorithmic processes, and circles represent data. Circles with bold borders represent the inputs to the algorithm. The electronic structure calculations are detailed in section 6.1.

The harmonic PES is then interpolated onto the fine q-point grid, where the harmonic free energy and other properties are calculated. If the anharmonic calculation is to be performed using a different coarse q-point grid to the harmonic calculation then the harmonic PES is interpolated onto this grid, and the normal modes $\{\hat{\mathbf{u}}\}$ on this grid are calculated.

The inputs to the anharmonic calculation include the inputs to the harmonic calculation; the static-lattice co-ordinate \mathbf{r}^0 , the crystal lattice matrix L and the q-point grid $\{\mathbf{q}\}$ on which the anharmonic PES will be mapped; and additionally the normal modes $\{\hat{\mathbf{u}}\}$ as calculated by the harmonic calculation. From these, the symmetry-invariant anharmonic basis functions $\{\mu\}$ are calculated using the algorithm presented in section 4.1.3. The non-VSCF basis functions are removed if they are not required, as detailed in section 5.1.3. The anharmonic calculation then proceeds identically to the harmonic calculation; optimal samples $\{\mathbf{u}\}$ are generated, electronic structure calculations are performed at each sample, and the results of these electronic structure calculations are used to fit the basis function coefficients, giving the anharmonic PES. The anharmonic PES is generally calculated using a coarser q-point grid than the harmonic PES, and the two can be interpolated and combined to give a more accurate PES, as described in section 4.3.

Before the VSCF calculation can be performed, it is necessary to perform an effective harmonic calculation, such that the eigenstates of the effective harmonic Hamiltonian can be used as basis states for the VSCF calculation¹. The effective harmonic calculation uses the self-consistency algorithm presented in section 5.5 to find the normal-mode frequencies which minimise the free energy of the crystal under the effective harmonic approximation, as detailed in section 3.5.3. This calculation does not require an explicit basis of states, as the expectation values of operators can be calculated with respect to an entire harmonic basis, as detailed in section 5.4.1. It should be noted that the effective harmonic calculation is constrained to keep the normal-mode unit vectors $\hat{\mathbf{u}}$ constant, as allowing these unit vectors to vary would greatly increase the number of terms which must be represented in the PES.

The VSCF calculation itself begins with the calculation of the single-subspace basis functions, which are calculated using the same method used to calculate the PES basis functions, as detailed in section 4.1.3. Then, using the basis of states which diagonalise the effective harmonic Hamiltonian, the self-consistency algorithm presented in section 5.5 is used to calculate the single-subspace Hamiltonians, and thus the VSCF Hamiltonian. The VSCF Hamiltonian is then interpolated onto a fine q-point grid, as detailed in section 5.2.1, and finally the thermodynamic calculations detailed in section 5.4.3 are performed in order to calculate the free energy and other properties under the VSCF approximation.

¹If the harmonic Hamiltonian contains normal modes with imaginary frequencies then the eigenstates of the harmonic Hamiltonian will not be well defined. In contrast, the effective harmonic Hamiltonian cannot contain normal modes with imaginary frequencies, as a result of how this Hamiltonian is defined.

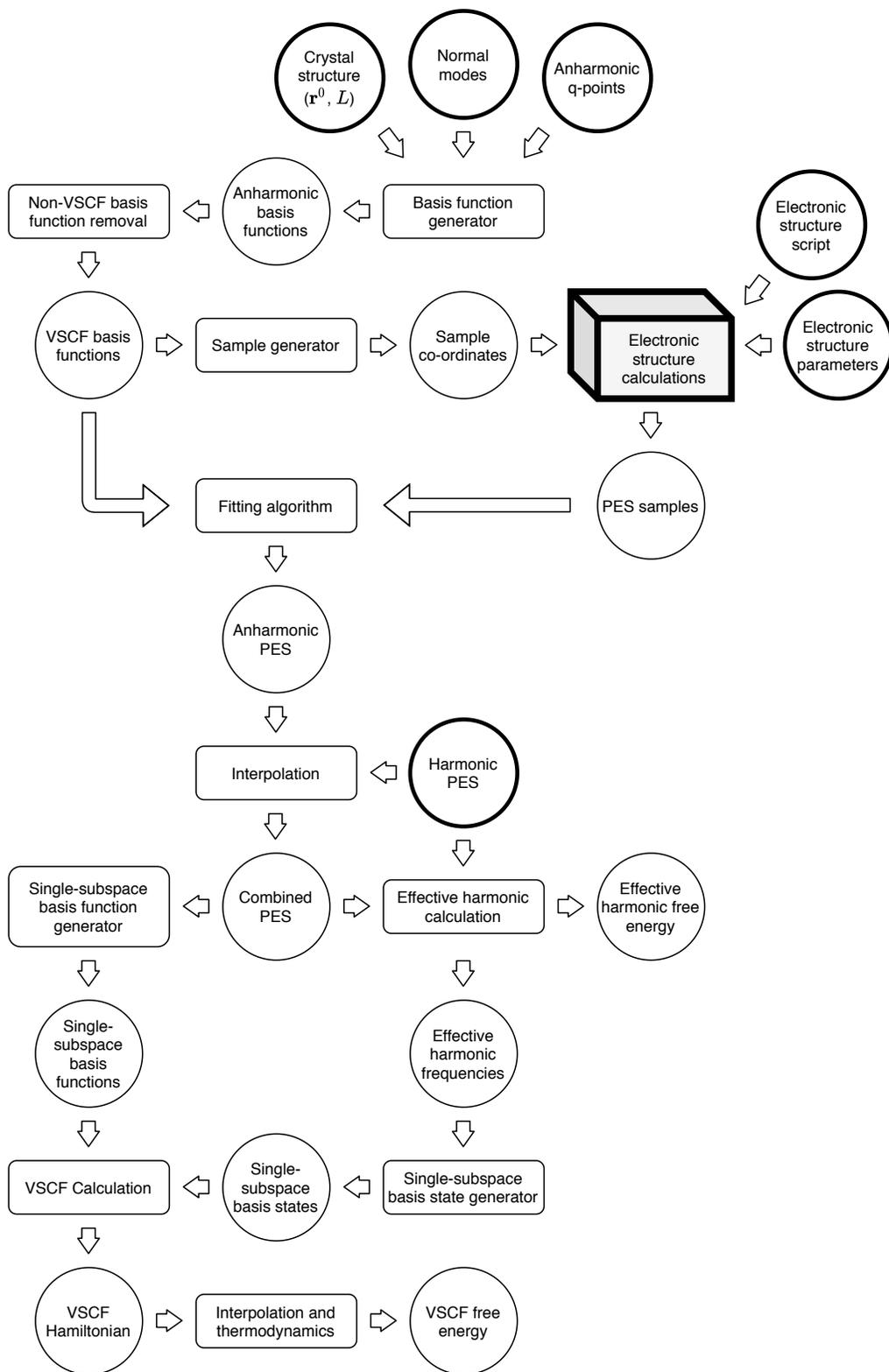


Figure 6.2: The anharmonic calculation performed by *caesar*. Rectangles represent algorithmic processes, and circles represent data. Circles with bold borders represent the inputs to the algorithm, including user inputs and the outputs of the harmonic calculation. The electronic structure calculations are detailed in section 6.1.

6.1 Electronic structure as a black box

In order to be compatible with as wide a range of electronic structure implementations as possible, `caesar` treats the electronic structure calculation as a black box. The only requirement on the electronic structure package is that, given a configuration of nuclear co-ordinates \mathbf{r} and the lattice vectors L , it must be able to calculate the Born-Oppenheimer potential V , and optionally any additional information such as the forces, Hessian or stress at that configuration.

In order to achieve this, `caesar` interacts with the electronic structure calculation purely through a user-supplied script. This script is required to read the nuclear configuration, run the electronic structure calculation, and write out the required calculation results. For many electronic structure methods, this simply entails the script calling the electronic structure package with specified input and output files, and pointing at a user-supplied settings file if needed.

In order to accommodate as wide a range of electronic structure packages as possible, `caesar` is able to work with a number of widely-used input and output electronic structure file formats, including formats compatible with the QUIP molecular dynamics package [145] and the Atomic Simulation Environment package [146], which provide compatibility with a wide range of electronic structure codes. Additional electronic structure codes can be interfaced using external programs which convert files from one format to another, such as the `c2x` file converter of Rutter [147].

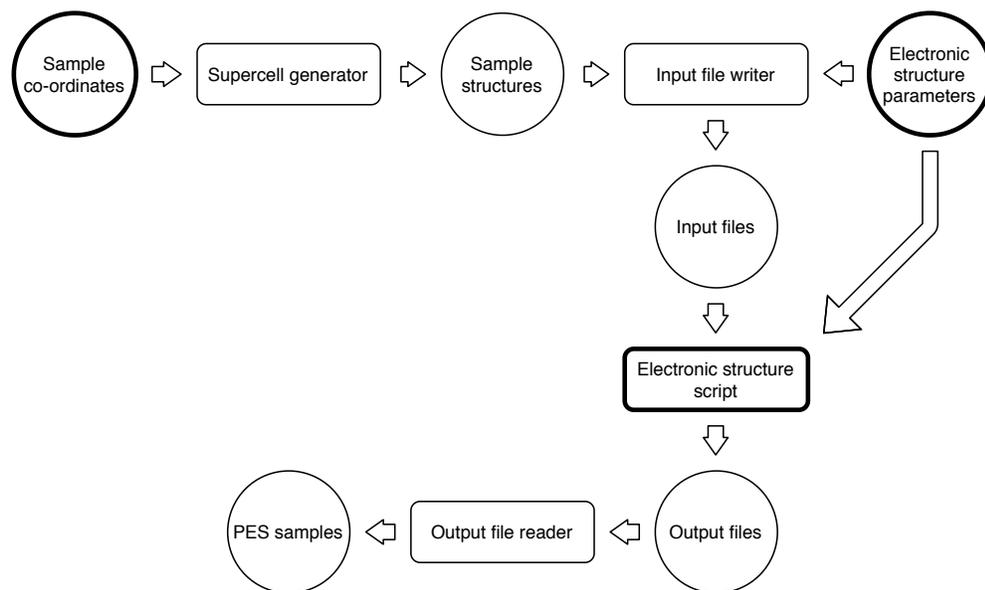


Figure 6.3: The electronic structure calculation wrapper implemented by `caesar`. Rectangles represent algorithmic processes, and circles represent data. Objects with bold borders represent the inputs to the algorithm.

The details of the electronic structure black box are presented in figure 6.3. Starting with a co-ordinate \mathbf{r} or \mathbf{u} , the algorithm first generates the commensurate supercell S which is most computationally inexpensive to simulate. The atomic co-ordinates are then calculated for every atom in this supercell. The details of this supercell, along with any necessary parameters, are converted into a format that the electronic structure method can process, and written to a file. The user-provided electronic structure script is called, which is expected to read the input file, and produce an output file containing the calculated PES data. This output file is then read, and converted into a format that `caesar` can process.

6.2 Interchangeable calculation elements

There are many parts of the VSCF method which can be represented in a wide range of different ways. For example, the PES can be represented as the interpolation of a grid of sampling points, or as a sum of basis functions, and the basis functions themselves can have many different functional forms. For another example, the free energy within each subspace can be calculated using a number of methods, ranging from an effective harmonic approximation through to methods based on configuration interaction or coupled cluster theories.

In order to allow for this variety of interchangeable parts, `caesar` is built in a modular manner, using object inheritance to separate the abstract notions of the VSCF methods from the concrete implementations of each method. Each building block of the method, such as the PES or the basis of states, is implemented as an abstract class, containing only the interfaces needed for the other parts of the method to interact with that block. Each abstract class then has one or more child classes, which each provide a particular concrete implementation for each of the inherited interfaces.

For example, the PES is represented by the abstract class `Potential`, which has a method `integrate()` which allows it to be integrated with respect to a set of basis states. A concrete implementation of this is the class `PolynomialPotential`, represented as a weighted sum of polynomial basis functions. The `PolynomialPotential` class implements `integrate()` by using the set of basis states to integrate each of its component monomials in turn, and summing the result. The integration of a monomial would in turn be an abstract method of the abstract `BasisStates` class, which would have a concrete implementation `HarmonicBasisStates`, which would integrate monomials analytically.

6.3 Exact symmetry representation

In cartesian co-ordinates, each symmetry \mathcal{S} is represented by a tensor S and a translation vector \mathbf{s} . In general, the elements of the tensor and the vector are real numbers, making this representation unsuitable for applications where exact comparisons are needed.

The symmetries can be represented exactly by storing their tensors in fractional co-ordinates, where they have purely integer elements, and by replacing the translation vector with a group which describes how the symmetry maps the static-lattice position of each atom at $\mathbf{R} = 0$ onto the static-lattice position of another atom, where the latter atom is represented by the index of its image in the primitive cell and its \mathbf{R} -vector in fractional co-ordinates.

This allows for calculations involving symmetry operators to be performed exactly and repeatedly, without any accumulation of error with repeated calculation.

6.4 Snap to symmetry

`caesar` uses the utility `spglib` [144] to identify the symmetries of each system it is given.

Often, static-lattice configurations are calculated through the use of numerical algorithms [9], and will only satisfy the given symmetry operations up to some pre-defined numerical tolerance. This can lead to numerical error, and it is desirable to find the static-lattice structure which is closest to the given structure, but which satisfies the symmetries exactly. This is known as snapping the structure to symmetry.

`caesar` performs the snap-to-symmetry algorithm in two stages, first snapping the lattice vectors, followed by snapping the nuclear co-ordinates. In both cases, snapping is done by using the symmetry operators to construct the set of linear equations which the elements of the given vectors must obey, before using linear regression to find the vectors which are closest to the original vectors given the constraints.

6.5 Generating single-subspace potentials

For a system with N subspaces, a naïve implementation of VSCF constructs each of the N single-subspace potentials by integrating each across the $N - 1$ other subspaces, requiring the calculation of $\mathcal{O}(N^2)$ integrals to calculate every single-subspace potential. The computational cost of this calculation can be reduced by replacing the naïve method with a bisection method.

First, the subspaces are split into two approximately equal sets, $\xi_1 \dots \xi_{\frac{N}{2}}$ containing the first $\frac{N}{2}$ subspaces, and $\xi_{\frac{N}{2}+1} \dots \xi_N$ containing the remaining subspaces. Then two multi-subspace potentials are constructed,

$$\mathcal{H}_{\xi_1 \dots \xi_{\frac{N}{2}}} = \langle \mathcal{H} \rangle_{\xi_{\frac{N}{2}+1} \dots \xi_N} \quad (6.1)$$

and

$$\mathcal{H}_{\xi_{\frac{N}{2}+1} \dots \xi_N} = \langle \mathcal{H} \rangle_{\xi_1 \dots \xi_{\frac{N}{2}}} . \quad (6.2)$$

Each multi-subspace Hamiltonian $\mathcal{H}_{\xi_i \dots \xi_j}$ is used to generate a further two Hamiltonians, as

$$\mathcal{H}_{\xi_i \dots \xi_{\frac{i+j}{2}}} = \langle \mathcal{H}_{\xi_i \dots \xi_j} \rangle_{\xi_{\frac{i+j}{2}+1} \dots \xi_j} \quad (6.3)$$

and

$$\mathcal{H}_{\xi_{\frac{i+j}{2}+1} \dots \xi_j} = \langle \mathcal{H}_{\xi_i \dots \xi_j} \rangle_{\xi_i \dots \xi_{\frac{i+j}{2}}} . \quad (6.4)$$

This process is repeated until the single-subspace Hamiltonians are reached.

By taking expectation values in this manner, the number of integrals required to calculate every single-subspace potential is reduced to $\mathcal{O}(N \log(N))$.

Chapter 7

Calculating the phase diagram of elemental titanium

The developments to VSCF and related methods presented in this work will now be illustrated by applying them to a system of interest. The chosen system is the modified embedded atom potential (MEAM) of Hennig et al. [148], as adapted for use with the QUIP molecular dynamics package [145] by Schlegel [149]. This potential is designed to model elemental titanium [148], although the accuracy of this potential for this purpose will not be evaluated. Instead, the aim is to accurately model the vibrational properties of the fictional system which is described exactly by the MEAM potential.

There are a number of reasons why this system provides a good illustration of the developments in the method, and of the limitations imposed by some of those developments:

- At low pressures, the solid phase diagram of the system contains only three phases [148], and so mapping this phase diagram requires relatively few free energy calculations.
- Each phase has a primitive unit cell which is invariant under the action of many symmetry operators [148]. This high-symmetry structure means that the potential reduction in computational cost from exploiting symmetry is high.
- Each phase has a small primitive unit cell, containing at most three atoms [148]. This means that the Born-von Karman supercell in which each vibrational calculation is performed must contain many primitive unit cells in order for the free energy to converge, and so it is important to correctly handle size consistency and interpolation.
- The MEAM potential as implemented in QUIP is extremely computationally inexpensive, allowing

the necessary electronic structure calculations containing hundreds of atoms to be performed in a few seconds each.

- One of the phases is unstable under the harmonic approximation, and contains a high proportion of phonon modes with imaginary frequencies [148]. This phase is stabilised at finite temperatures by anharmonic effects [62, 150]. This means that the free energy of this phase can only be accurately calculated using anharmonic methods.

7.1 The phases of titanium

At low pressures, elemental titanium exhibits three solid phases, labelled α , β and ω [148, 151]. A number of high pressure phases also exist [152], but these are not relevant to the current work.

The α phase has a hexagonal close-packed (hcp) crystal structure, with two atoms per primitive unit cell. There are two free parameters in this structure: the volume per atom, Ω , and the ratio of the lengths of the primitive R-vectors, c . The lattice matrix for this structure is

$$L = \sqrt[3]{\frac{4\Omega}{\sqrt{3}c}} \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & c \end{pmatrix}, \quad (7.1)$$

and the fractional co-ordinates of the atoms in the primitive unit cell are¹

$$\begin{aligned} \mathbf{r}_1 &= (0 \ 0 \ 0) \\ \mathbf{r}_2 &= \left(\frac{1}{3} \ \frac{2}{3} \ \frac{1}{2}\right). \end{aligned} \quad (7.2)$$

This is shown on the left of figure 7.1.

The β phase has a body-centred cubic (bcc) crystal structure, with one atom per primitive unit cell. The only free parameter in this structure is the volume per atom, Ω . The lattice matrix for this structure is

$$L = \sqrt[3]{\frac{\Omega}{4}} \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}, \quad (7.3)$$

¹Due to translational invariance, the choice of origin for the fractional co-ordinates is arbitrary. For convenience, this origin is consistently chosen so that one atom has a fractional co-ordinate of 0.

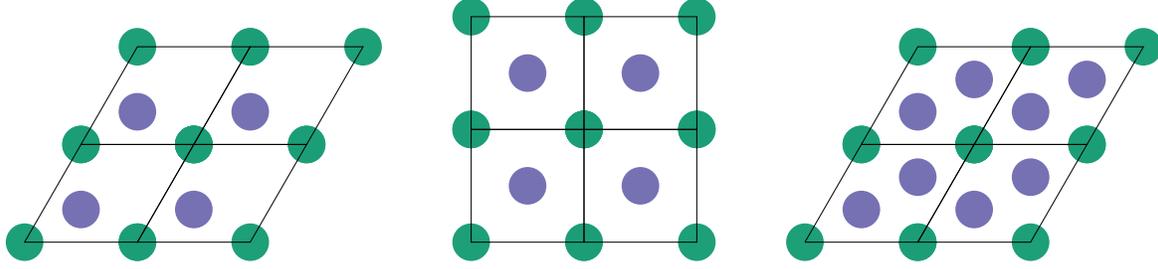


Figure 7.1: The crystal structures of the three phases of titanium, with α titanium on the left, β titanium in the centre and ω titanium on the right. The primitive unit cells of the α and ω phases, and the conventional unit cells of the β phase, are shown as black boxes. The third lattice vector for each structure lies directly out of the page. The atoms which are coloured turquoise have fractional co-ordinates of 0 along this lattice vector, while the atoms which are coloured blue have fractional co-ordinates of $\frac{1}{2}$ along this lattice vector.

and the fractional co-ordinate of the atom in the primitive unit cell is

$$\mathbf{r}_1 = (000). \quad (7.4)$$

This is shown in the centre of figure 7.1.

The ω phase has a hexagonal crystal structure, with three atoms per primitive unit cell. There are two free parameters in this structure: the volume per atom, Ω , and the ratio of the lengths of the primitive R-vectors, c . The lattice matrix for this structure is

$$L = \sqrt[3]{\frac{6\Omega}{\sqrt{3}c}} \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & c \end{pmatrix}, \quad (7.5)$$

and the fractional co-ordinates of the atoms in the primitive unit cell are

$$\begin{aligned} \mathbf{r}_1 &= (000) \\ \mathbf{r}_2 &= \left(\frac{1}{3} \frac{2}{3} \frac{1}{2}\right) \\ \mathbf{r}_3 &= \left(\frac{2}{3} \frac{1}{3} \frac{1}{2}\right). \end{aligned} \quad (7.6)$$

This is shown on the right of figure 7.1.

7.2 Calculating the phase diagram from the free energy

The output of `caesar` is the vibrational free energy surface $F(\Omega, c, T)$. This is used to calculate the Gibbs free energy surface $G(p, T)$, and then the phase diagram for the system, using the method described in

section 3.6.1.

When comparing multiple structures it is necessary to consider the total free energy, including the electronic and vibrational contributions. Since the system under consideration is described by a classical potential, the electronic contribution to the free energy is simply the energy of the static-lattice structure. All free energies and Gibbs free energies in the remainder of this chapter should be assumed to include this electronic contribution to the free energy.

Firstly, the calculated values of $F(\Omega, c, T)$ are used to fit a third-order Birch-Murnaghan equation of state [136] for each value of T . For β titanium, this is given by

$$F_T(\Omega) = \alpha_0 + \alpha_1 \Omega^{-\frac{2}{3}} + \alpha_2 \Omega^{-\frac{4}{3}} + \alpha_3 \Omega^{-\frac{6}{3}}, \quad (7.7)$$

and for α and ω titanium this is given by

$$\begin{aligned} F_T(\Omega, c) = & \alpha_0 \\ & + \alpha_1 \left(\frac{\Omega}{c}\right)^{-\frac{2}{3}} + \alpha_2 (\Omega c^2)^{-\frac{2}{3}} \\ & + \alpha_3 \left(\frac{\Omega}{c}\right)^{-\frac{4}{3}} + \alpha_4 \left(\frac{\Omega}{c}\right)^{-\frac{2}{3}} (\Omega c^2)^{-\frac{2}{3}} + \alpha_5 (\Omega c^2)^{-\frac{4}{3}} \\ & + \alpha_6 \left(\frac{\Omega}{c}\right)^{-\frac{6}{3}} + \alpha_7 \left(\frac{\Omega}{c}\right)^{-\frac{4}{3}} (\Omega c^2)^{-\frac{2}{3}} + \alpha_8 \left(\frac{\Omega}{c}\right)^{-\frac{2}{3}} (\Omega c^2)^{-\frac{4}{3}} + \alpha_9 (\Omega c^2)^{-\frac{6}{3}}. \end{aligned} \quad (7.8)$$

Secondly, the values of $c(\Omega, T)$ are found for which the stress tensor $\sigma(\Omega, c, T)$ corresponds to an isotropic pressure, i.e. for which $\sigma \propto \mathbb{1}$. For the β phase, the pressure is guaranteed to be isotropic by symmetry, and for the α and ω phases this corresponds to the values of c for which $\left(\frac{\partial}{\partial c}\right)_\Omega F_T(\Omega, c) = 0$. Substituting these values of $c(\Omega, T)$ back into $F(\Omega, c, T)$ gives the isotropic free energy surface $F(\Omega, T)$.

Thirdly, the pressure $p(\Omega, T)$ is calculated from $F(\Omega, T)$ via

$$p(\Omega, T) = - \left(\frac{\partial F(\Omega, T)}{\partial \Omega} \right)_T. \quad (7.9)$$

Finally, the Gibbs free energy $G(p, T)$ is calculated from $F(\Omega, T)$ and $p(\Omega, T)$, using the Legendre transform

$$G(p, T) = F(\Omega, T) + p(\Omega, T)\Omega. \quad (7.10)$$

Once the Gibbs free energy per atom has been calculated for all three phases, it is possible to construct the phase diagram of solid titanium. The stable phase at a given temperature and pressure is the phase with

the lowest Gibbs free energy per atom at those conditions.

7.3 Existing calculations

Hennig et al. [148] calculate the solid portion of the phase diagram of titanium using classical molecular dynamics, similar to the methods discussed in section 3.5.8 of this work. They do not calculate and compare the Gibbs free energy of the three phases. Instead, they calculate the most stable phase at a given temperature and pressure by performing a classical molecular dynamics simulation of the system under those conditions using a thermostat and barostat, and identify which phase is predominant in the simulation. The phase diagram is then constructed by repeating these simulations at a range of temperatures and pressures [148]. The phase diagram of Hennig et al. [148] is reproduced in figure 7.8 below. This phase diagram is qualitatively similar to the phase diagram of titanium as derived from experiments or calculated using other methods [151–155].

As discussed in section 3.5.8, calculating the phase diagram in this way has a number of advantages and disadvantages compared to the VSCF method as implemented in `caesar`. The molecular dynamics simulations sample the PES directly, and so no error is introduced through fitting processes, and no approximations need to be made to the PES itself. Unlike in vibrational methods, molecular dynamics places no restriction on the distance atoms may move from their static-lattice positions, and so the error arising from these restrictions is not incurred.

However, the use of classical mechanics means the simulation does not include zero-point effects, and so the phase diagram is likely to be inaccurate at low temperatures [152]. The method also requires a large number of electronic structure calculations, each of which simulates a large number of atoms. Hennig et al. [148] ran molecular dynamics simulations of systems containing 432 atoms, performing electronic structure calculations at time steps of 1 fs, over simulations lasting for up to 1 ns. This means that identifying the stable phase at each temperature and pressure required up to 1 million electronic structure calculations [148]. This scale of calculation may be feasible when using a MEAM potential, but is unlikely to be feasible in a high-throughput context when using more accurate ab-initio electronic structure methods.

7.4 Calculating properties under the harmonic approximation

The first stage of the vibrational calculation is a harmonic calculation. This must be performed for each value of Ω and c for each phase, and each calculation gives results for a range of temperatures T .

7.4.1 Converging the harmonic supercell

In order to get accurate results, the Born-von Karman supercell used for each harmonic calculation must be sufficiently large that the calculation result is converged. For consistency, the same supercell was used for all calculations of each phase. These supercells were converged using a single representative structure from each phase. These structures were all chosen to have primitive cell volumes of $\Omega = 17.5\text{\AA}^3$, and the α and ω structures were chosen to have R-vector ratios of $c = 1.6$ and $c = 0.6$ respectively.

The top panels of figures 7.2, 7.3 and 7.4 show the convergence of the free energy with respect to the q-point spacing of the harmonic supercell, for the chosen structure in the α , β and ω phases respectively.

When the free energy was converged to an accuracy of 10^{-5} Ha per cell, the α , β and ω phases had q-point spacings of $3.5 \times 10^{-2} \text{ bohr}^{-1}$, $4.6 \times 10^{-2} \text{ bohr}^{-1}$ and $2.6 \times 10^{-2} \text{ bohr}^{-1}$ respectively, corresponding to harmonic supercells with supercell matrices

$$S_\alpha = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad S_\beta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \text{ and } S_\omega = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{pmatrix}. \quad (7.11)$$

The bottom panels of figures 7.2, 7.3 and 7.4 show the phonon dispersion curves and phonon densities of states for the converged supercells for the α , β and ω phases respectively. It should be noted that the β phase has normal modes with imaginary frequencies, and that the contribution of these modes to the free energy must be neglected under the harmonic approximation.

7.4.2 The harmonic phase diagram

In order to allow for comparison with the results presented in Hennig et al. [148], the vibrational properties of each phase were calculated for values of Ω between 15 and 20 \AA^3 per atom, in steps of 0.5 \AA^3 per atom. For the α phase, the ratio c was varied between 1.55 and 1.70 and for the ω phase c was varied between 0.55 and 0.65, in steps of 0.01 in each case. Each vibrational calculation was performed at temperatures between 0 and 1500 K, in steps of 300 K.

The isotropic free energy surfaces $F(\Omega, T)$ for the three phases are shown in figure 7.5. As previously noted, the β phase is unstable under the harmonic approximation over a range of volumes. Figure 7.6 shows the proportion of normal modes of the β which have imaginary frequencies. As the volume increases, this proportion tends to one. Under the harmonic approximation, the vibrational free energy is a sum of single-mode contributions, and the modes with imaginary frequencies must be neglected from this sum. As such,

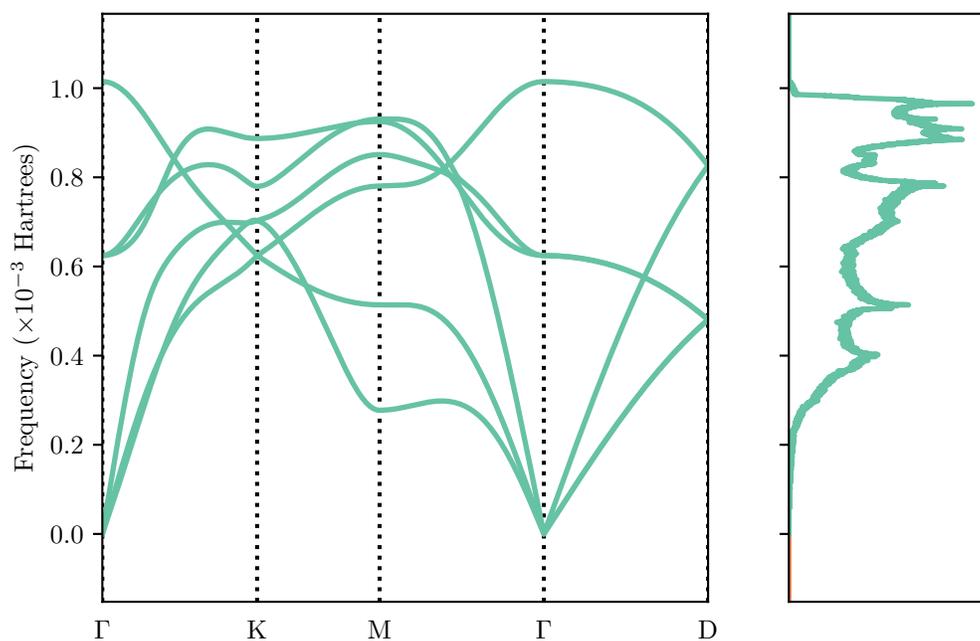
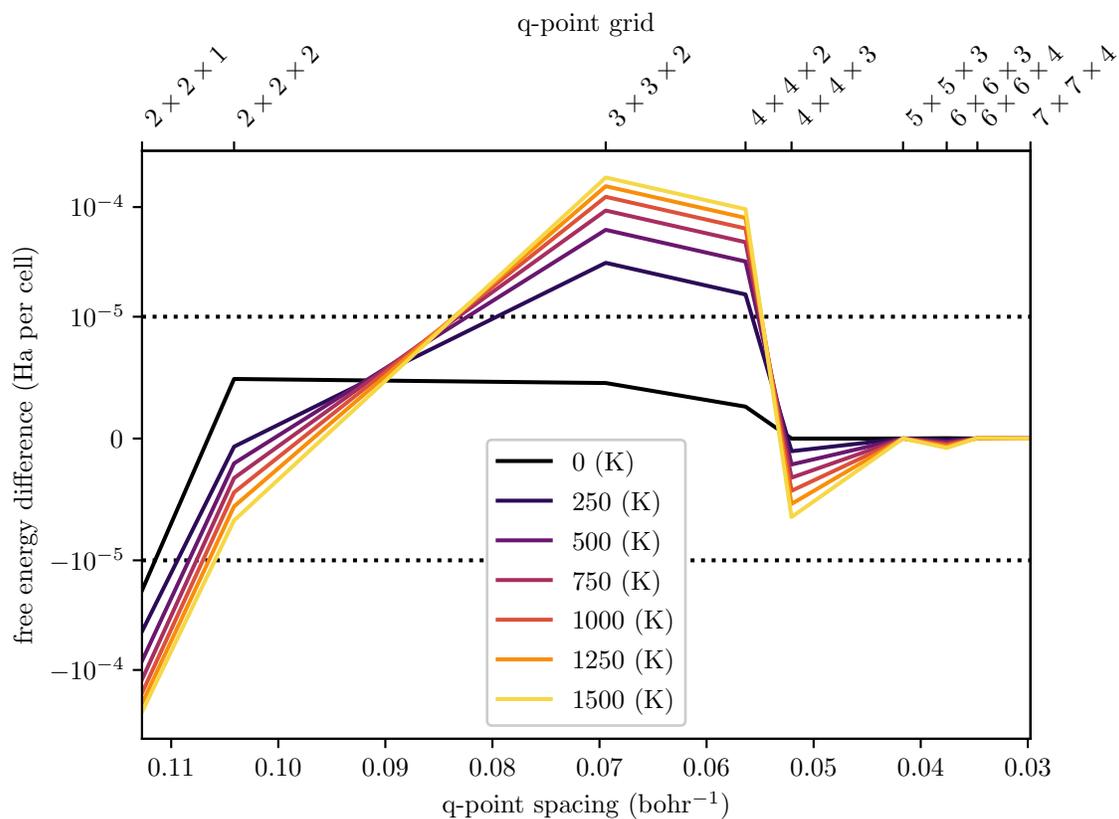


Figure 7.2: α titanium

Top: Convergence of harmonic free energy with respect to q-point spacing. Free energies are shown relative to the most-converged value for each temperature, on a semilog axis.

Bottom: Dispersion and density of states for harmonic phonons. The fractional co-ordinates of the high-symmetry q-points are $\Gamma = 0$, $K = (\frac{1}{3}, \frac{1}{3}, 0)$, $M = (\frac{1}{2}, 0, 0)$ and $D = (0, 0, \frac{1}{2})$.

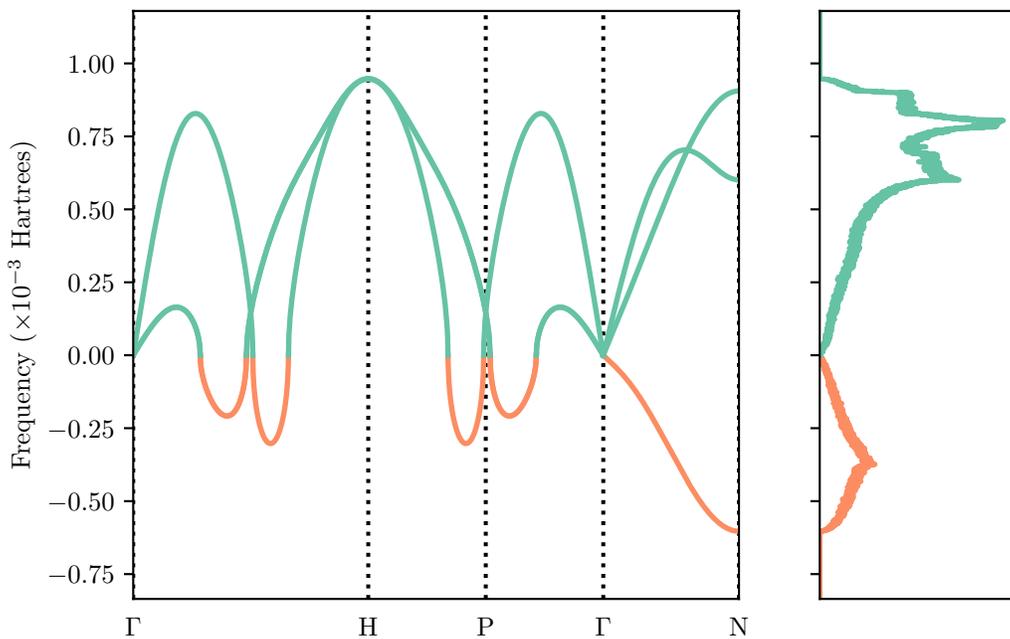
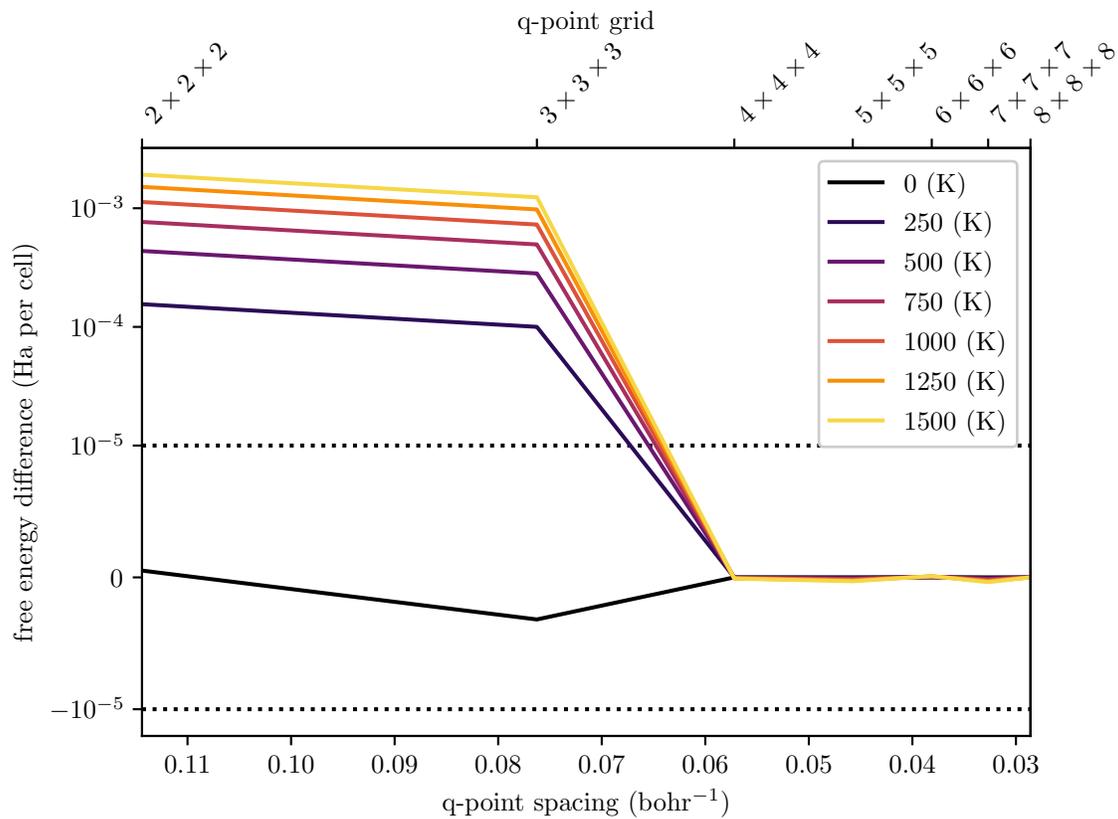


Figure 7.3: β titanium

Top: Convergence of harmonic free energy with respect to q-point spacing. Free energies are shown relative to the most-converged value for each temperature, on a semilog axis.

Bottom: Dispersion and density of states for harmonic phonons. Real frequencies are coloured turquoise and imaginary frequencies are coloured orange. The fractional co-ordinates of the high-symmetry q-points are $\Gamma = 0$, $H = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $P = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $N = (0, \frac{1}{2}, 0)$.

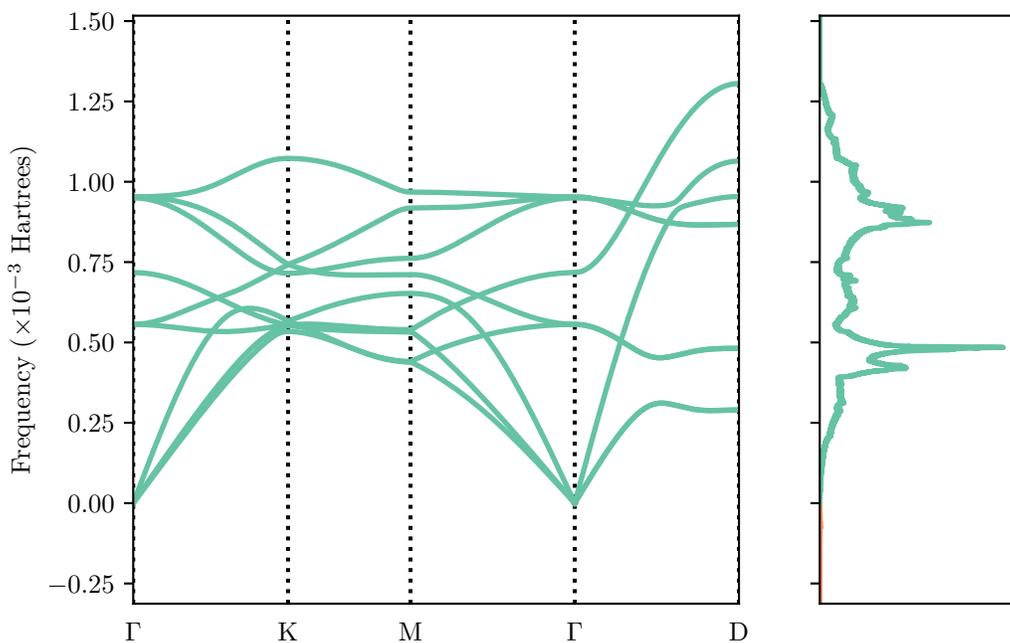
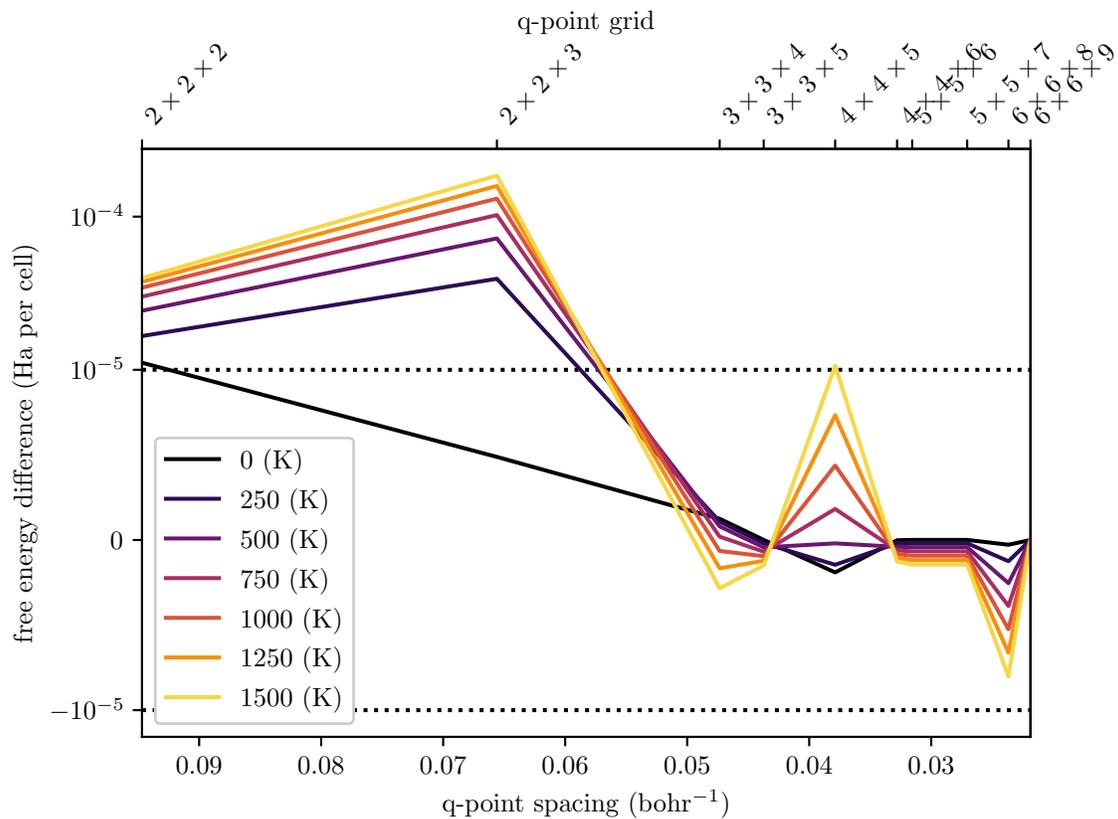


Figure 7.4: ω titanium

Top: Convergence of harmonic free energy with respect to q-point spacing. Free energies are shown relative to the most-converged value for each temperature, on a semilog axis.

Bottom: Dispersion and density of states for harmonic phonons. The fractional co-ordinates of the high-symmetry q-points are $\Gamma = 0$, $K = (\frac{1}{3}, \frac{1}{3}, 0)$, $M = (\frac{1}{2}, 0, 0)$ and $D = (0, 0, \frac{1}{2})$.

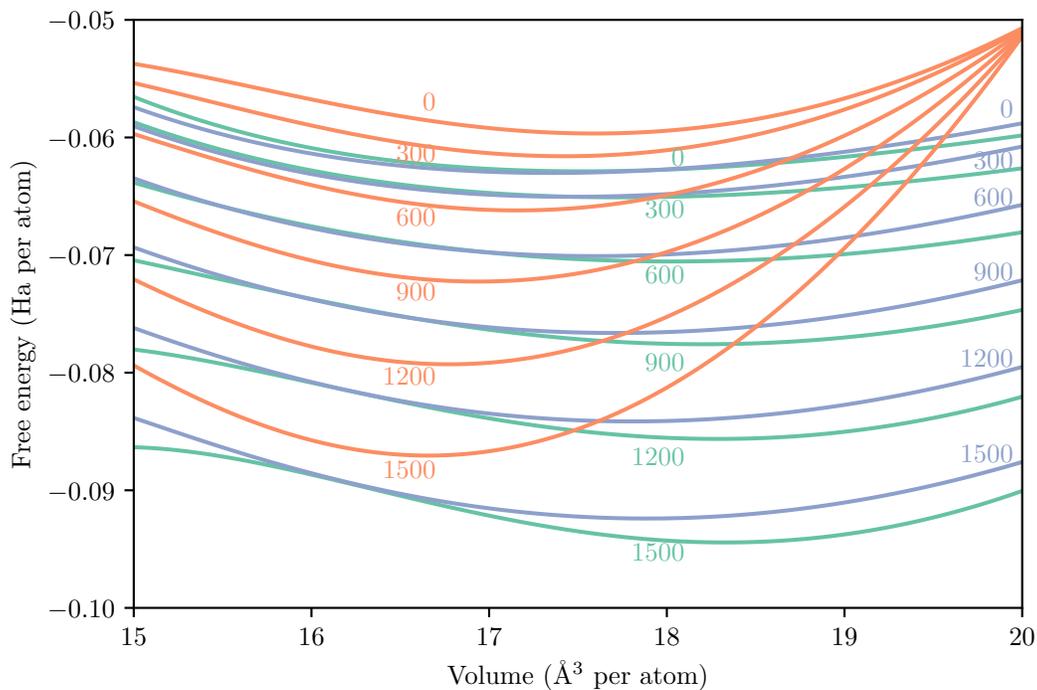


Figure 7.5: Free energy calculated under the harmonic approximation. The α , β and ω phases are coloured turquoise, orange and blue respectively. The coloured labels denote the temperature in Kelvin.

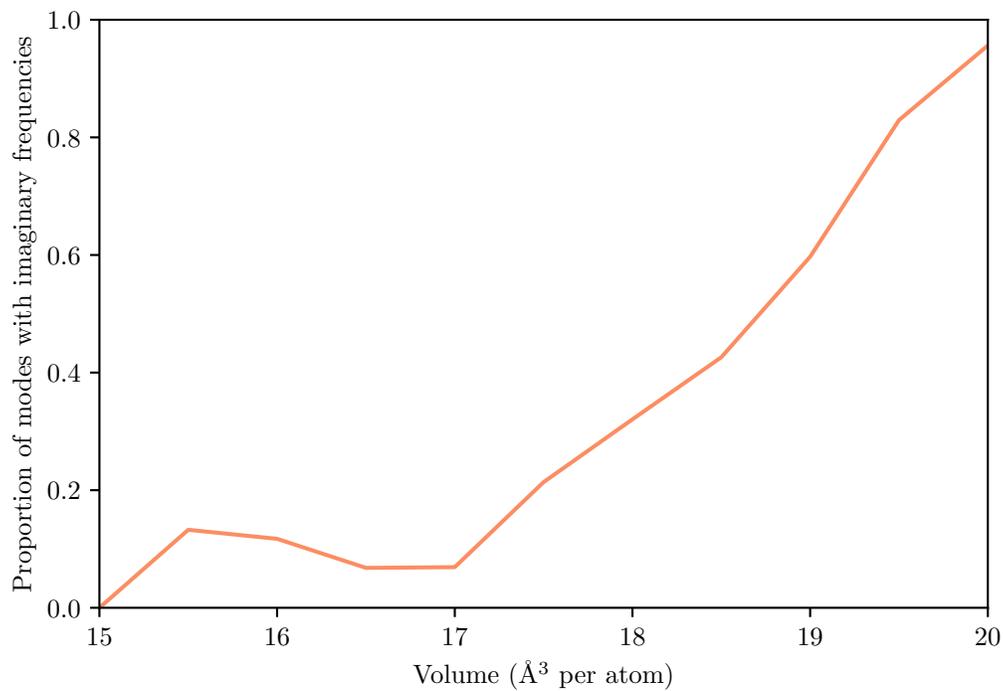


Figure 7.6: The proportion of interpolated normal modes of the β phase which have imaginary frequencies under the harmonic approximation.

as the volume increases the vibrational free energy under the harmonic approximation tends to zero and becomes an increasingly inaccurate estimate of the true vibrational free energy. This is why the free energy curves of the β phase at different temperatures all tend to the same value, the static-lattice energy, at the right of figure 7.5.

The phase diagram calculated under the harmonic approximation is shown in figure 7.7, and the phase diagram calculated by Hennig et al. [148] is shown in figure 7.8 for comparison.

The phase boundary between the α and ω phases as calculated under the harmonic approximation in figure 7.7 is qualitatively similar to that calculated by Hennig et al. [148] in figure 7.8. The two are not expected to be identical because, as previously noted, the calculations under the harmonic approximation do not consider the effects of anharmonicity, and the calculations of Hennig et al. [148] do not consider the effects of zero-point motion.

The phase boundary between the β phase and the other two phases is missing from figure 7.7, as the β phase is not predicted to be stable under the harmonic approximation at any values of temperature and pressure. This is because, as shown in figure 7.6, the β phase contains a high proportion of modes with imaginary frequencies. The contribution of these modes to the free energy must be neglected under the harmonic approximation, but an anharmonic calculation would be expected to show that the neglected contribution is negative. This means the harmonic approximation is expected to overestimate the free energy of the β phase.

The phase transition to a liquid phase is not shown in figure 7.7. As discussed in section 2.1.4, this transition cannot be calculated using vibrational methods. Further, as discussed in section 3.2.5, the accuracy of the harmonic approximation decreases as temperature increases, and so the region of the phase diagram close to the melting temperature of titanium will be inaccurate.

7.5 Calculating anharmonic properties

7.5.1 Mapping the anharmonic potential energy surface

Since the β phase has only one atom per unit cell, the supercell whose supercell matrix is the identity matrix $\mathbb{1}$ only contains the three normal modes which correspond to translation in space. As such, the smallest supercell which can be used to perform meaningful vibrational calculations is that with supercell matrix $2\mathbb{1}$.

This supercell contains eight q-points, arranged into three q-point stars. Every q-point in this supercell is its own pair, and so every normal mode in this supercell is real. The q-point star Γ^* contains only $\mathbf{q}_r = (0, 0, 0)$.

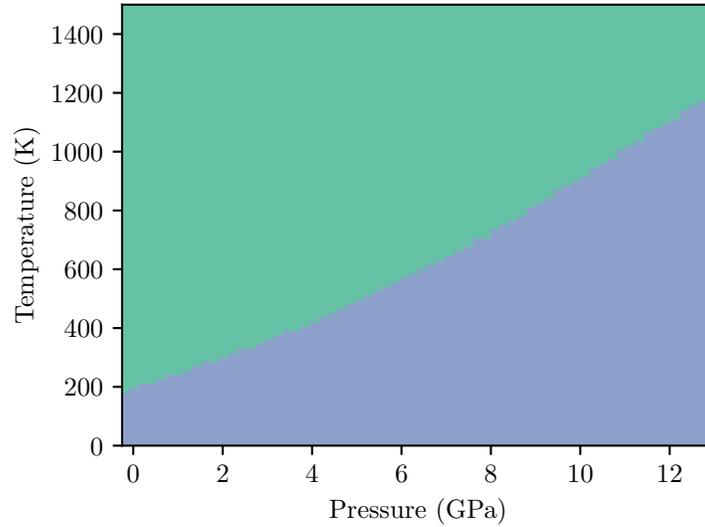


Figure 7.7: The solid phase diagram of titanium under the harmonic approximation. The α and ω phases of titanium are coloured turquoise and blue respectively. The β phase does not appear as it is not the stable phase at any combination of temperature and pressure.

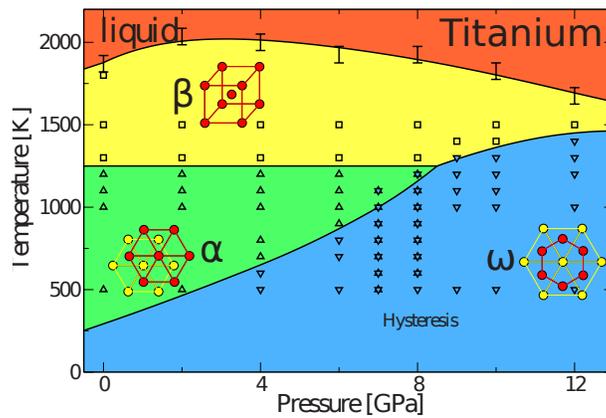


Figure 7.8: The phase diagram of titanium as calculated by Hennig et al. [148]. Reprinted figure with permission from R. G. Hennig, T. J. Lenosky, D. R. Trinkle, S. P. Rudin and J. W. Wilkins, Physical Review B, 78, 054121, (2008) [148]. Copyright (2008) by the American Physical Society. See appendix C for details.

All three normal modes at \mathbf{q}_T are degenerate, and these are the three purely translational modes, which do not contribute to vibrational calculations. The q-point star H^* contains only $\mathbf{q}_H = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. All three normal modes at \mathbf{q}_H are also degenerate, forming a single degenerate subspace, ξ_H . The q-point star N^* contains six q-points, $\mathbf{q}_{N_a} = (\frac{1}{2}, 0, 0)$, $\mathbf{q}_{N_b} = (0, \frac{1}{2}, 0)$, $\mathbf{q}_{N_c} = (0, 0, \frac{1}{2})$, $\mathbf{q}_{N_{ab}} = (\frac{1}{2}, \frac{1}{2}, 0)$, $\mathbf{q}_{N_{ac}} = (\frac{1}{2}, 0, \frac{1}{2})$ and $\mathbf{q}_{N_{bc}} = (0, \frac{1}{2}, \frac{1}{2})$. The normal modes at each q-point in N^* are not related to one another by symmetry, and so they form three degenerate subspaces, ξ_{N_1} , ξ_{N_2} and ξ_{N_3} , each of which contains six normal modes, with a single mode at each q-point in N^* .

The PES has been fitted using a Taylor expansion truncated at sixth order, as this is the highest order for which computation is currently feasible. Without using the PAA, and neglecting symmetry, the sixth PES representation would contain more than $\binom{21}{6} = 54264$ unique monomial terms. Reducing this using translational and space-group symmetry as described in section 4.1.3, and removing the non-VSCF basis functions as described in section 5.1.1, gives a PES representation containing only 230 distinct basis functions, and which correspondingly has 230 real coefficients which must be fitted.

Including forces in the fitting process, as described in section 4.4, it is possible to accurately map the PES using a small number of sampling points per basis function. For these calculations, four sampling points per basis function were used. In order to remove the non-VSCF basis functions, some of these sampling points had to be duplicated up to 30 times, as described in section 5.1.1. This gave a total of 5324 sampling points. Since the calculation included terms involving three or more q-points, the majority of these sampling points could not take advantage of the non-diagonal supercell method, and had to be performed using the full supercell containing eight atoms.

Figure 7.9 shows a one-dimensional slice through the PES along one mode from each of the four degenerate subspaces. Each plot shows the true value of the PES along each mode, taken directly from electronic structure calculations, along with the values of the sixth anharmonic PES and the harmonic PES. The true PES is clearly better represented by the anharmonic PES than by the harmonic PES, but the anharmonic PES is still far from perfect. As can be seen in figure 7.5, in order to distinguish phases the free energy calculation must have a better accuracy than 0.01 Hartree per atom, and the anharmonic PES does not reliably achieve this.

The main reason for this lack of accuracy is that the sixth form of the anharmonic PES does not have the flexibility needed to capture all of the features of the true PES. Increasing the truncation order of the Taylor expansion would increase the flexibility of the representation, but as the complexity of the PES increases exponentially with the truncation order this is not computationally feasible.

A number of the problems with Taylor series PES representations discussed in section 3.4.3 can be seen in figure 7.9. The true PES of β titanium softens with increasing displacement, i.e. the gradient of the

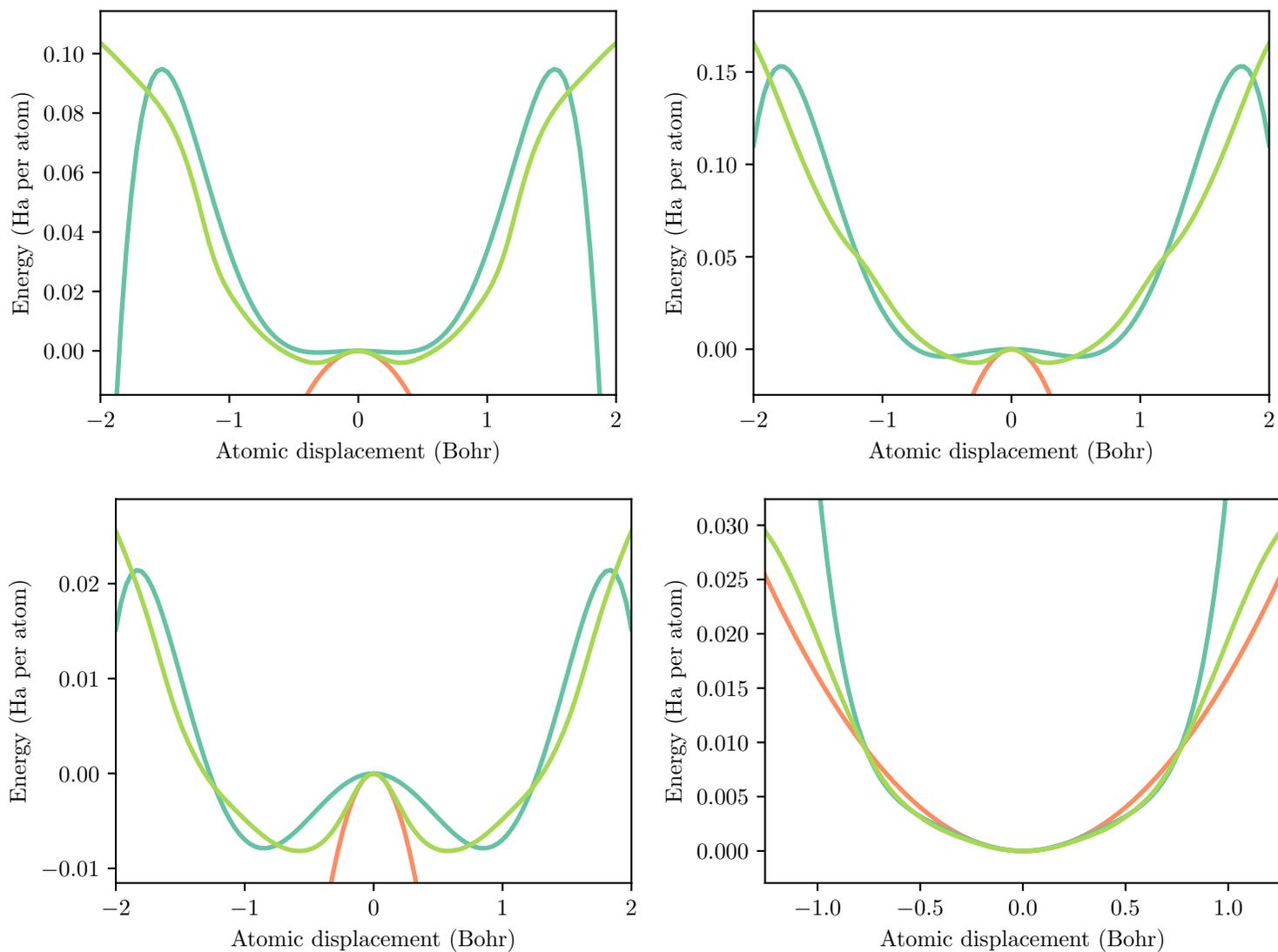


Figure 7.9: One-dimensional slices of the PES of β titanium along one normal mode from each subspace in the 21 supercell. The subspaces, in reading order, are ξ_H , ξ_{N_1} , ξ_{N_2} and ξ_{N_3} . The true potential, sextic anharmonic potential and harmonic potential are shown in green, blue and orange respectively.

PES $|\frac{\partial}{\partial \mathbf{u}} V(\mathbf{u})|$ decreases as the displacement $|\mathbf{u}|$ increases. However, the functional form of a Taylor series truncated at low order means that beyond a certain displacement the gradient must increase with increasing displacement. This causes the Taylor series PES to reach an artificial maximum near the edge of where the PES is sampled, and then to drop rapidly to large negative values. This can most clearly be seen for the PES mapped along the mode in the ξ_H subspace shown in the top left panel of figure 7.9.

As discussed in section 2.1.4, the mapped region of the PES must contain a potential well which is sufficiently deep to contain the vibrational wavefunctions. The error associated with only mapping the potential well up to a depth of V_{\max} at temperature T_{\max} is expected to scale approximately as $e^{-V_{\max}/k_B T_{\max}}$. As can be seen in figure 7.8, in order to model the region of stability of the β phase of titanium, it is necessary to perform accurate calculations at $T_{\max} = 1500$ K, and so the depth of the potential well must satisfy $V_{\max} \gg 4.8 \times 10^{-3}$ Hartree. As can be seen in the bottom left panel of figure 7.9, this means the PES must be mapped up to a displacement of about 2 bohr per atom. This large displacement has consequences for the accuracy of the fitted PES, as for a fixed set of basis functions there must be a trade-off between the size of the region being mapped and the accuracy of that mapping.

7.5.2 The self-consistent harmonic method

As described in section 5.4, the full VSCF calculation relies on an effective harmonic calculation to provide the basis of states from which the VSCF wavefunctions will be constructed. This effective harmonic calculation is performed using the same self-consistent convergence method as VSCF, as described in section 5.5, except that where in VSCF the free energy is minimised with respect to the VSCF wavefunctions coefficients, in the effective harmonic calculation the free energy is minimised with respect to the effective harmonic frequencies. This free energy minimisation is performed at multiple temperatures, giving temperature-dependent effective frequencies.

The effective frequencies of the modes in the $2\mathbb{1}$ supercell can be interpolated to arbitrary q-points, as described in section 3.2.4, and this can be used to calculate an approximation to the free energy of the system. This can be improved by combining the effective harmonic calculation in the $2\mathbb{1}$ supercell with the harmonic calculation in the $5\mathbb{1}$ supercell, giving a Hamiltonian

$$\mathcal{H}_{\text{combined}}^{(5\mathbb{1})} = \mathcal{H}_{\text{effective}}^{(2\mathbb{1})} - \mathcal{H}_{\text{harmonic}}^{(2\mathbb{1})} + \mathcal{H}_{\text{harmonic}}^{(5\mathbb{1})}. \quad (7.12)$$

This is equivalent to modelling the short-range interactions as anharmonic, and the longer-range interactions as harmonic.

For a given set of effective frequencies, the free energy is calculated using the analytic integrals described in

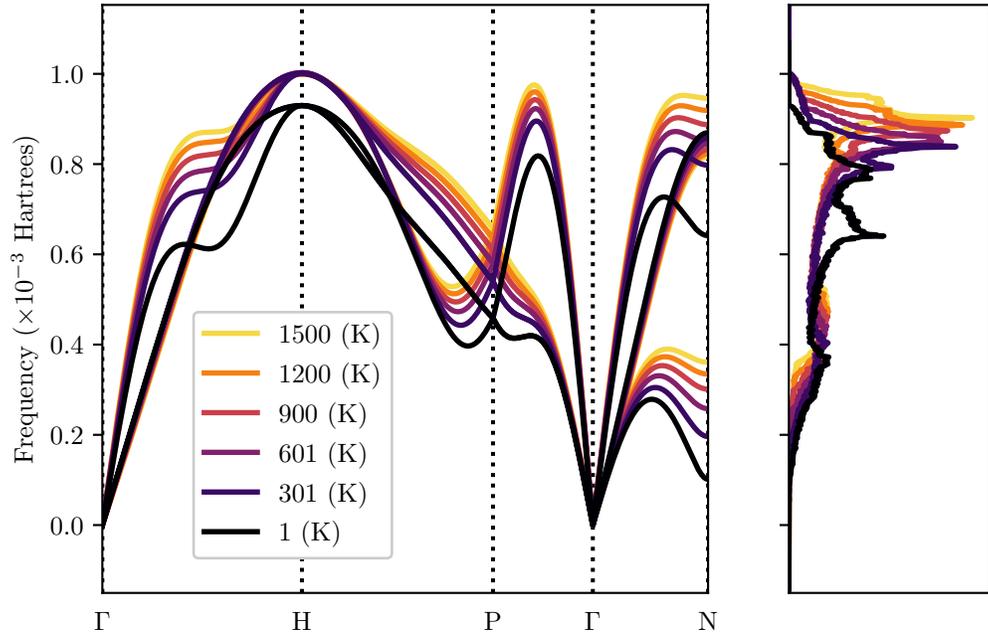


Figure 7.10: The phonon dispersion and density of states of β Titanium with $\Omega = 17.5$ (\AA^3 per atom), calculated under the self-consistent harmonic approximation at a range of temperatures. The fractional co-ordinates of the high-symmetry q-points are $\Gamma = 0$, $H = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $P = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $N = (0, \frac{1}{2}, 0)$.

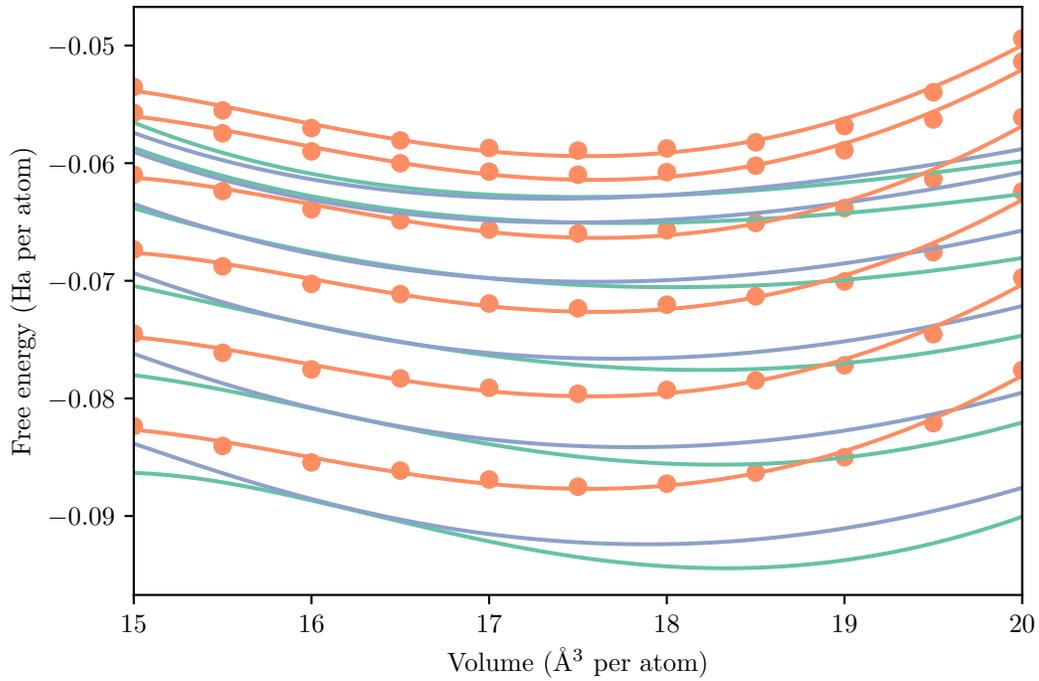


Figure 7.11: Free energy of the three phases of titanium. The free energies of the α and ω phases, coloured turquoise and blue respectively, are calculated under the self-consistent harmonic approximation. The free energy of the β phase, coloured orange, is calculated using an effective harmonic method. The coloured dots show the results of individual calculations, and the coloured lines show the fitted third-order Birch-Murnaghan equation of state.

For each phase, the different lines correspond to different temperatures, with the top line showing $0(K)$ and the bottom showing $1500(K)$, and the intermediate lines spaced evenly between these temperatures in intervals of $300(K)$.

appendix A. Using such analytic integrals allows each free energy calculation to be performed quickly and accurately, but also imposes a limitation on the frequencies which may be accurately calculated. Since the domain of the analytic integrals is across all space rather than only across the mapped region of the PES, these integrals will only be accurate if the nuclear density outside the mapped region of the PES is small. Along a given mode with effective frequency ω , the nuclear density beyond some displacement u increases as ω decreases, and so if the effective frequency becomes too small then a significant fraction of the nuclear density will lie outside the mapped region of the PES, and the integral will become inaccurate.

This limitation becomes a problem when the PES erroneously drops to large negative values beyond the mapped region, as the sextic PES of β titanium does as shown in figure 7.9. This allows the free energy minimisation to produce free energies which are artificially large and negative, simply by reducing the effective frequency until the nuclear density escapes the mapped potential well. This meant that in a number of the calculations there was no acceptable set of effective frequencies which minimised the free energy, and so at least one effective frequency had to be artificially set to a lower bound of 10^{-4} Hartree.

The phonon dispersion curve and density of states of β Titanium with $\Omega = 17.5 \text{ \AA}^3$ per atom, calculated using the effective harmonic method, is shown in figure 7.10, and can be compared to the same data calculated under the harmonic approximation shown in the bottom panel of figure 7.3. Importantly, in the effective harmonic phonon spectrum there are no modes with imaginary frequencies. This is guaranteed by construction at the q-points commensurate with the $2\mathbb{1}$ supercell, but is not guaranteed at the interpolated q-points. This suggests that including only the anharmonic interactions which can be contained within the $2\mathbb{1}$ supercell is sufficient to ensure that the β phase lies within a well-defined potential well.

Figure 7.10 shows that the effective frequencies are strongly temperature-dependent. This is expected, as the PES is very strongly anharmonic. The increase in frequency with temperature shown by most of the modes is characteristic of potentials which are dominated by terms with higher order than harmonic; as the temperature increases, the average vibrational displacements increase, and the average second derivative of the potential increases.

The free energy of the β phase calculated using the effective harmonic method is shown in figure 7.11, alongside the free energies of the α and ω phases calculated under the harmonic approximation for comparison. This can be compared with the free energy calculated under the harmonic approximation shown in figure 7.5. The free energies calculated using the effective harmonic method are generally lower than those calculated using the harmonic approximation. This is consistent with the expectation that the average contribution to the free energy from the modes with imaginary frequencies is negative.

The reduction in the free energy of the β phase using the effective harmonic method causes a corresponding reduction in the Gibbs free energy, but this is not sufficient to make the β phase stable at any combination

of temperature and pressure, and as such the phase diagram calculated using the effective harmonic method is identical to that calculated under the harmonic approximation as shown in figure 7.7.

7.5.3 The vibrational self-consistent field method

Symmetry and the potential energy surface of titanium

Once the effective frequencies have been calculated, it is possible to perform the full VSCF calculation.

As previously noted, the ξ_H subspace contains three degenerate normal modes, and so provides a good illustration of the effects of symmetry on VSCF discussed in section 5.1. It is important to note that the directions of these modes are not uniquely defined by the calculation under the harmonic approximation; for any given choice of the three unit vectors $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$, any orthogonal transformation of these unit vectors will give a different, equally valid choice of unit vectors.

It is informative to consider the two symmetry-invariant single-subspace basis functions in the ξ_H subspace at quartic order. The first basis function, V_1 , obeys full spherical symmetry, and can be written as

$$V_1 = c_1(u_1^2 + u_2^2 + u_3^2)^2 = c_1(u_1^4 + u_2^4 + u_3^4 + 2u_1^2u_2^2 + 2u_1^2u_3^2 + 2u_2^2u_3^2), \quad (7.13)$$

regardless of the choice of normal modes. The second basis function, V_2 , has the symmetries of a cube. For a specific choice of normal modes, this basis function can be written as

$$V_2 = c_2(u_1^4 + u_2^4 + u_3^4), \quad (7.14)$$

but this functional form will be different for a different choice of normal modes, and in general will also contain terms of the form $u_i u_j^3 - u_i^3 u_j$.

This demonstrates the importance of performing VSCF using single-subspace potentials rather than single-mode potentials; V_1 cannot be approximately separated into single-mode terms, and separating V_2 into single-mode terms would result in a different set of terms, and so a different free energy, for each choice of normal modes.

However, using single-subspace potentials comes at a cost. The computational cost and memory requirements of the calculation scale unfavourably with the number of states in each basis of states. In order to achieve the same accuracy from the basis of states that would be achieved with N basis states in each single-mode basis, an n -dimensional single-subspace basis must contain $\mathcal{O}(N^n)$ states. For the present simulation, the basis of states had to be limited to only include states with total occupation numbers of 15 or less. This

meant that the basis of states in the three-dimensional ξ_H subspace contained 816 states.

Free energy calculations

The free energy of the β phase calculated using VSCF is shown in the top panel of figure 7.12. This calculation was performed using the $2\mathbb{1}$ supercell, and without using the PAA. The free energies calculated using VSCF are lower again than those calculated using the effective harmonic method shown in figure 7.11, particularly at larger volumes. However, the resulting Gibbs free energies are still not low enough to make the β phase stable at any combination of temperature and pressure, and so the phase diagram calculated using VSCF is again identical to that calculated under the harmonic approximation as shown in figure 7.7.

There are several reasons which might explain why the VSCF calculation disagrees with the results of Hennig et al. [148]. As previously discussed, the anharmonic representation of the PES shows large deviations from the true PES, and the effective harmonic frequencies had to be artificially limited to prevent them from becoming too small. Both of these problems are caused primarily by the lack of flexibility inherent to the sextic Taylor expansion representation of the PES. Additionally, the $2\mathbb{1}$ supercell is small, and likely does not capture all of the important anharmonic interactions in the PES. The seven q-points at which non-translational modes are found also all lie on the boundary of the Brillouin zone, and so the normal modes at these q-points are not expected to be representative of the Brillouin zone as a whole. Rather, as can be seen in 7.10, the majority of these normal modes have effective frequencies which are much higher than the average frequency.

7.5.4 The principal axes approximation

The second order principal axes approximation

It is not computationally feasible to perform a VSCF calculation using a sextic PES in a supercell larger than $2\mathbb{1}$ without making approximations. As discussed in section 3.4.3, it is standard practice to take the PAA in order to make progress.

Taking the PAA at second order, the number of basis functions in the PES of β titanium in the $2\mathbb{1}$ supercell is reduced from 230 to 102. This also reduces the number of samples per basis function required in order to remove the non-VSCF basis functions, reducing the total number of samples required to map the PES from 5324 to 1016 samples. Further, since taking the PAA at second order removes all of the terms which include coupling between three or more q-points, the non-diagonal supercell method can be used for all electronic structure calculations, such that each calculation contains at most four atoms.

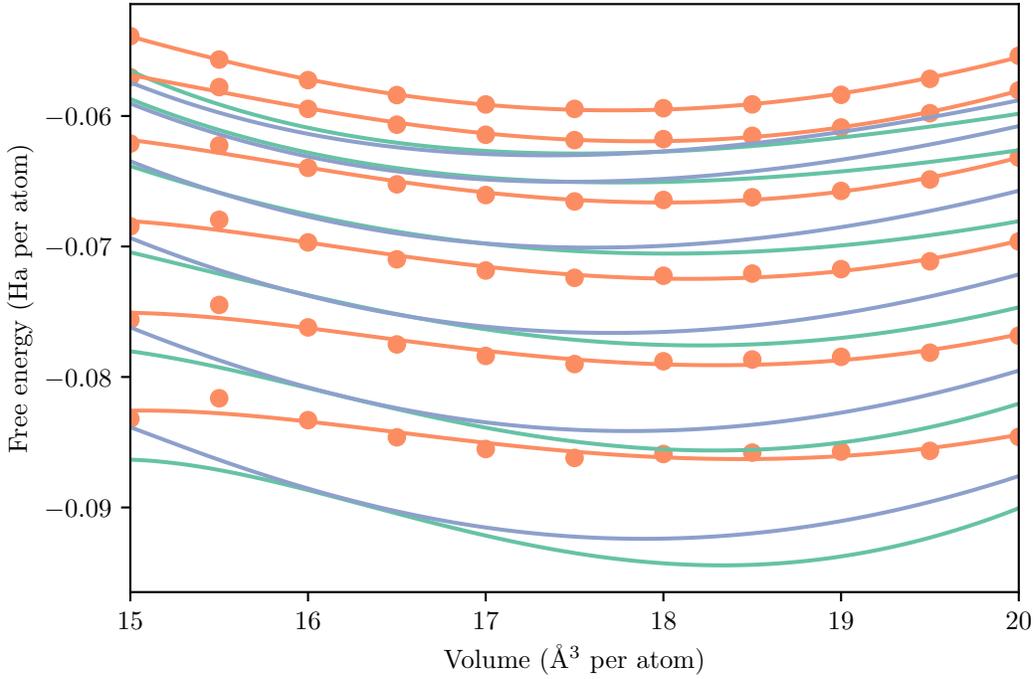
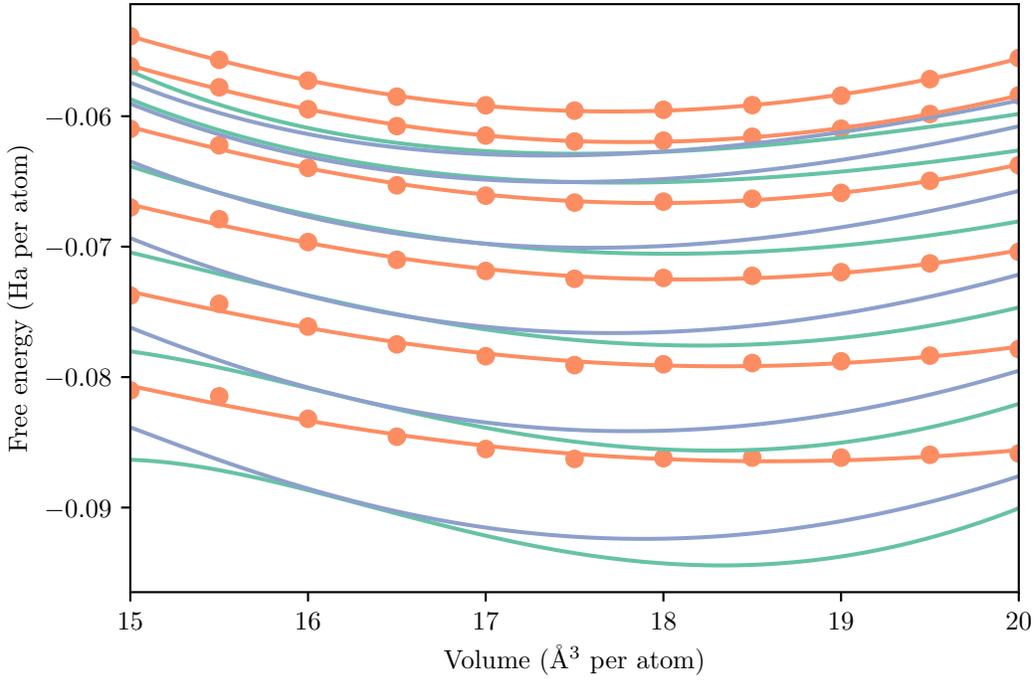


Figure 7.12: Free energy of the three phases of titanium. The free energies of the α and ω phases, coloured turquoise and blue respectively, are calculated under the harmonic approximation. The free energy of the β phase, coloured orange, is calculated using VSCF. The coloured dots show the results of individual calculations, and the coloured lines show the fitted third-order Birch-Murnaghan equation of state. For each phase, the different lines correspond to different temperatures, with the top line showing $0(K)$ and the bottom showing $1500(K)$, and the intermediate lines spaced evenly between these temperatures in intervals of $300(K)$.

Top: The PES of the β phase mapped on the 21 supercell without using the PAA.

Bottom: The PES of the β phase mapped on the 21 supercell using the PAA at second order.

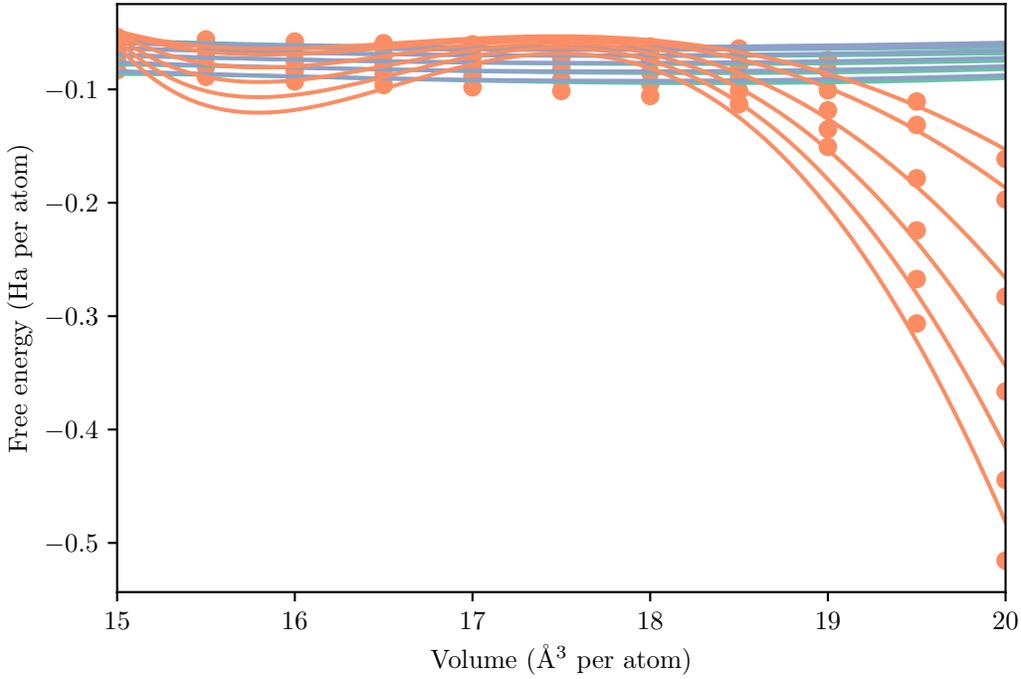
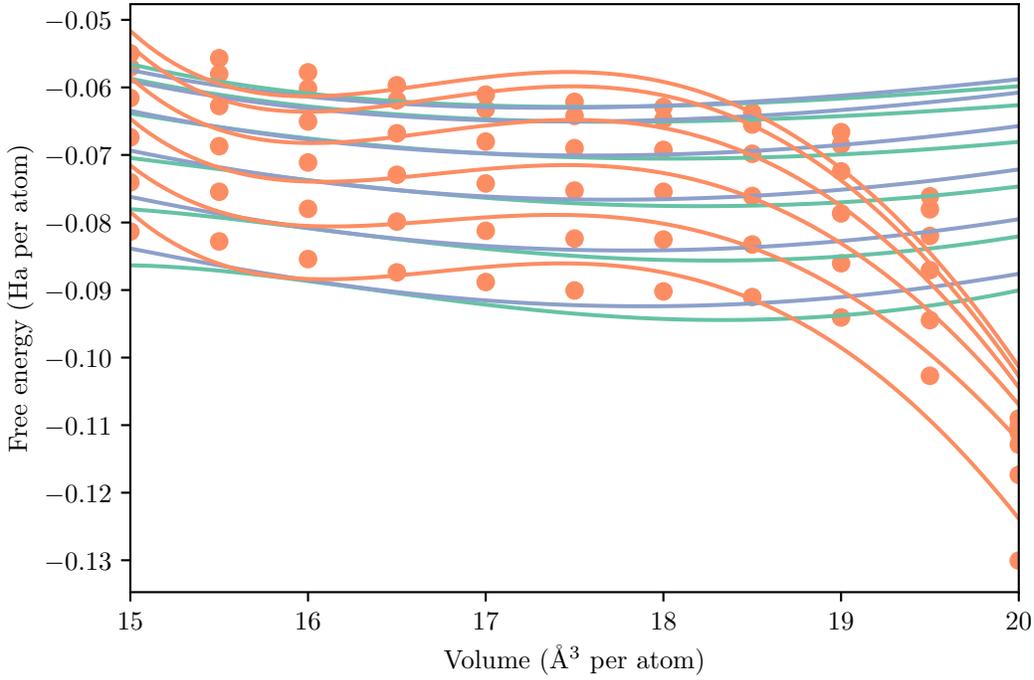


Figure 7.13: Free energy of the three phases of titanium. The free energies of the α and ω phases, coloured turquoise and blue respectively, are calculated under the harmonic approximation. The free energy of the β phase, coloured orange, is calculated using VSCF. The coloured dots show the results of individual calculations, and the coloured lines show the fitted third-order Birch-Murnaghan equation of state. For each phase, the different lines correspond to different temperatures, with the top line showing $0(K)$ and the bottom showing $1500(K)$, and the intermediate lines spaced evenly between these temperatures in intervals of $300(K)$.

Top: The PES of the β phase mapped on the 21 supercell using the PAA at first order.

Bottom: The PES of the β phase mapped on the 51 supercell using the PAA at first order.

The free energy of β titanium calculated using the PAA at second order is shown in the bottom panel of figure 7.12. It can be seen that there is little difference between this and the same free energy calculated without using the PAA as shown in the top panel of figure 7.12, despite the PAA not being size-consistent as discussed in section 4.2. This suggests that the terms which are neglected under the PAA at second order do not make a large contribution to the free energy calculation.

The success of the PAA at second order may be a result of the specific form of the anharmonic PES. As can be seen in figure 7.9, if β titanium is displaced from the static-lattice configuration along a typical normal mode, the value of the true PES of β titanium initially decreases, then comes to a minimum and starts increasing again, before starting to soften as the displacement approaches the edge of the potential well. Using a sextic potential, this behaviour is naturally represented as a negative harmonic term, a positive quartic term, and a small negative sextic term.

The free energy calculation depends mostly on the details of the bottom of the potential well, and with this combination of terms these details are primarily defined by the harmonic and quartic terms. These are also the terms which are largely unaffected by taking the PAA at second order. As such, it is possible that a more accurate representation of the PES, for which more terms were relevant to the bottom of the potential well, would be more strongly affected by taking the PAA at second order.

The first order principal axes approximation

Going further, and taking the PAA at first order, the number of basis functions in the sextic PES in the 21 supercell is reduced to just 15. There is also no need to duplicate sampling points to remove non-VSCF basis functions, and so the full PES can be sampled using just 60 sampling points. The electronic structure calculations at these sampling points can also take full advantage of the non-diagonal supercell method, and so each calculation only contains two atoms.

The free energy of β titanium calculated using the 21 supercell and taking the PAA at first order is shown in the top panel of figure 7.13. When compared to the same free energy calculated without taking the PAA shown in the top panel of figure 7.12, it can be seen that taking the PAA at first order introduces a large negative error to the calculated free energy, particularly at larger volumes, where a larger proportion of the normal modes are unstable under the harmonic approximation.

Under the PAA at first order there are only $\mathcal{O}(N)$ quartic terms in the PES of a supercell of size N . There are no relevant normal modes at the Γ -point and so, as discussed in section 4.2.3, the expectation value of each of the quartic term scales as $\mathcal{O}(N^{-1})$ or faster, and so there are not enough quartic terms in the PES for the sum of their expectations to be extensive. This means that the effect of the quartic terms tends to zero as the supercell size is increased. As discussed above, when modelled using a sextic potential, β titanium is

largely stabilised by the quartic terms, and so as the effect of these terms decreases the free energy becomes increasingly negative.

This can be demonstrated more clearly by performing the same calculation using a larger supercell. The bottom panel of figure 7.13 shows the free energy of β titanium calculated using the 51 supercell and again taking the PAA at first order. As expected, increasing the size of the supercell increases the error associated with using the PAA, and the erroneous reduction in the free energy at large volumes becomes more pronounced.

7.5.5 Interpolating the anharmonic potential

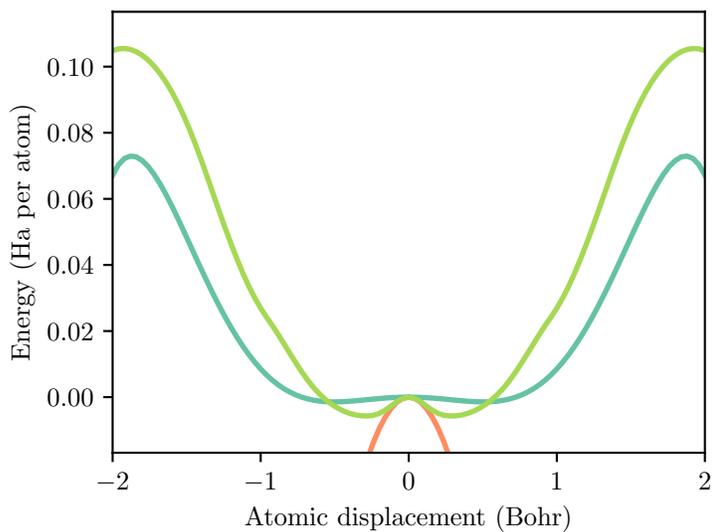
As discussed in section 4.3, rather than calculating the anharmonic PES of a given supercell directly, it is possible to reconstruct this PES by interpolate the anharmonic PES calculated using a smaller supercell, and combining this with a calculation of the harmonic PES of the original supercell. If the anharmonic interactions can be well represented using the smaller supercell, then this process should accurately reproduce the PES of the original supercell at a fraction of the computational cost.

Figures 7.14 and 7.15 each show a one-dimensional slice of the PES of β titanium with $\Omega = 10 \text{ \AA}^3$ per atom, along the normal modes of the 51 supercell with the largest imaginary and real harmonic frequencies respectively. Each figure shows the PES calculated using the 51 supercell directly, along with a number of PES interpolated from smaller supercells. Specifically, the PES was interpolated from PES calculated using the 21, 31 and 41 supercells, taking the PAA at first order, and from the PES calculated using the 21 supercell without the use of the PAA.

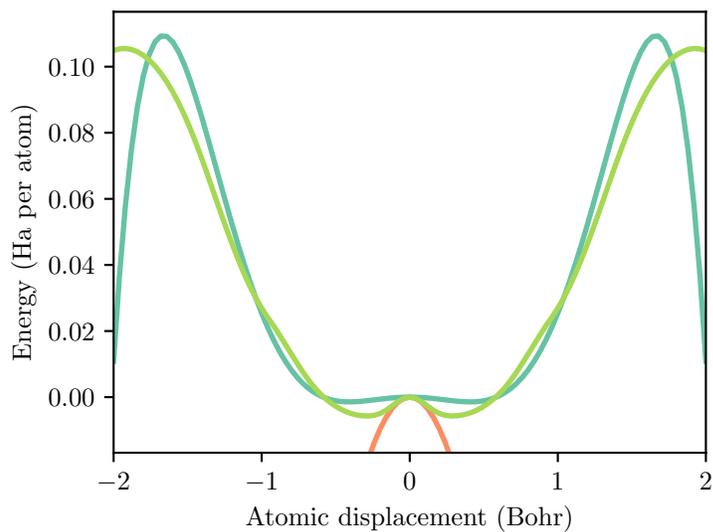
It can be seen that interpolating the PES from that calculated using the 21 and the PAA produces a PES which is only qualitatively similar to the 51 PES. The PES interpolated from that calculated using the 21 supercell without taking the PAA is a much closer match to the directly calculated PES.

However, if the PAA is taken then increasing the size of the supercell where the PES is calculated does not systematically improve the interpolated PES. This can be seen most clearly in panel (d) of figure 7.14 and in panel (c) of figure 7.15, where the interpolated PES does not even qualitatively resemble the directly calculated PES. This, again, is because the PES calculated using the PAA is not size-consistent.

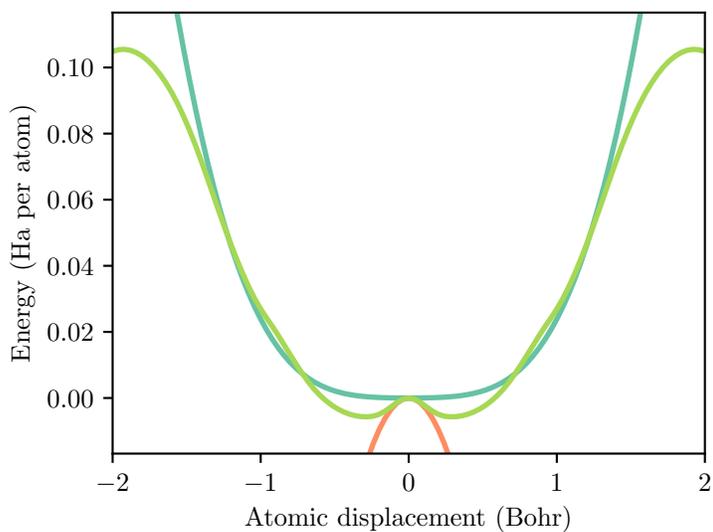
If the PES calculated using the 51 supercell and the PAA were interpolated onto one of the smaller supercells, the resulting PES would contain a number of terms which did not obey the PAA in the smaller supercell. As such, these non-PAA terms are needed when interpolating from the smaller supercell back to the 51 supercell, and the fact that they are not included because of the PAA means that the interpolated PES is wrong.



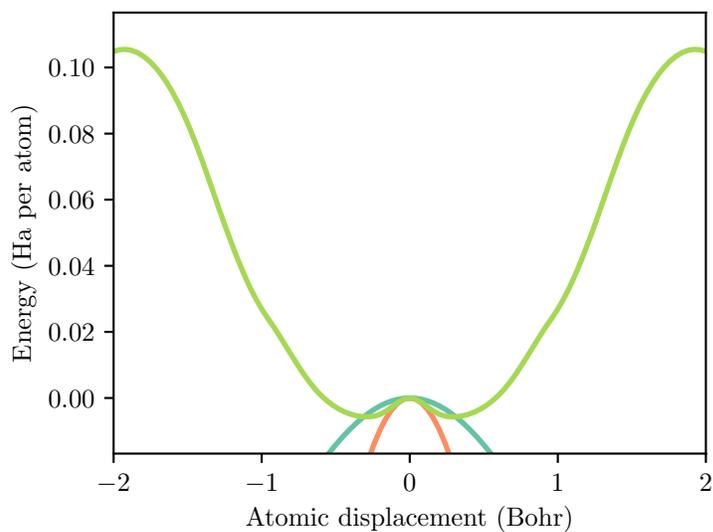
(a) The anharmonic PES interpolated from the 21 supercell, taking the PAA at first order.



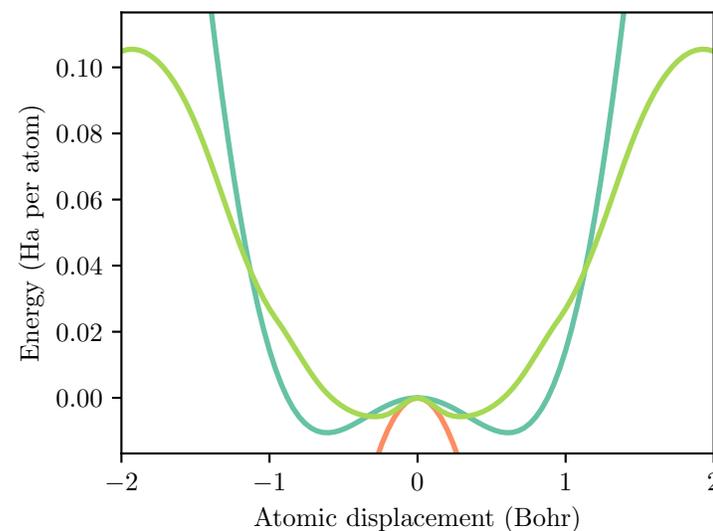
(b) The anharmonic PES interpolated from the 21 supercell, without taking the PAA.



(c) The anharmonic PES interpolated from the 31 supercell, taking the PAA at first order.



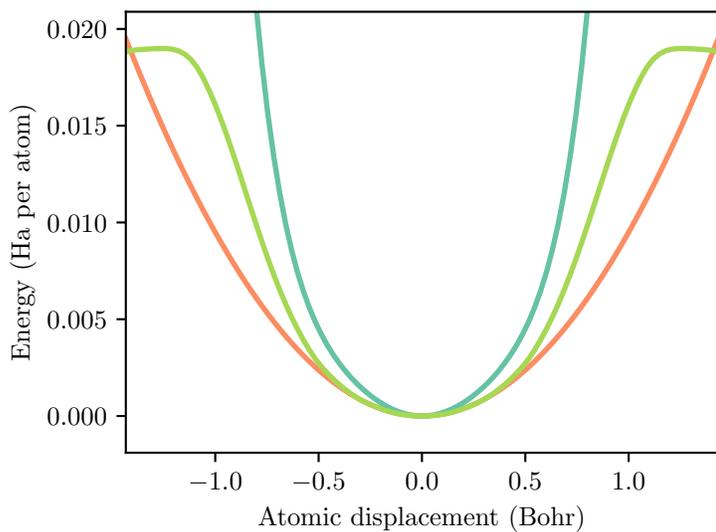
(d) The anharmonic PES interpolated from the 41 supercell, taking the PAA at first order.



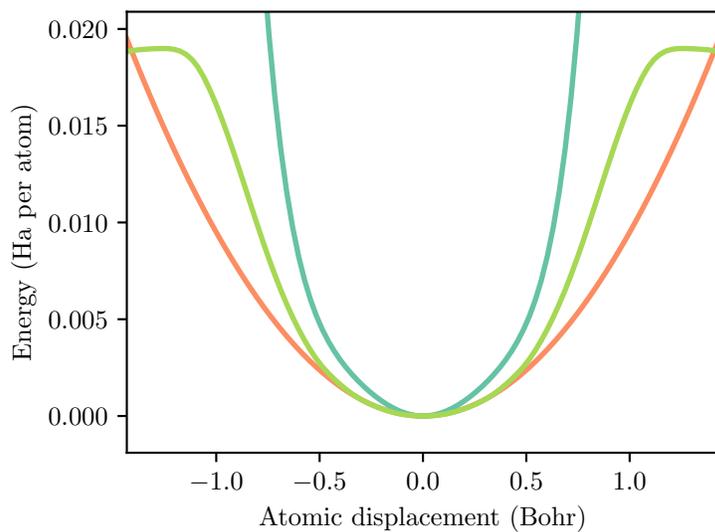
(e) The anharmonic PES calculated using the 51 supercell directly.

Figure 7.14: One-dimensional slices of the PES of β titanium, along the mode on the 51 supercell with the largest imaginary frequency.

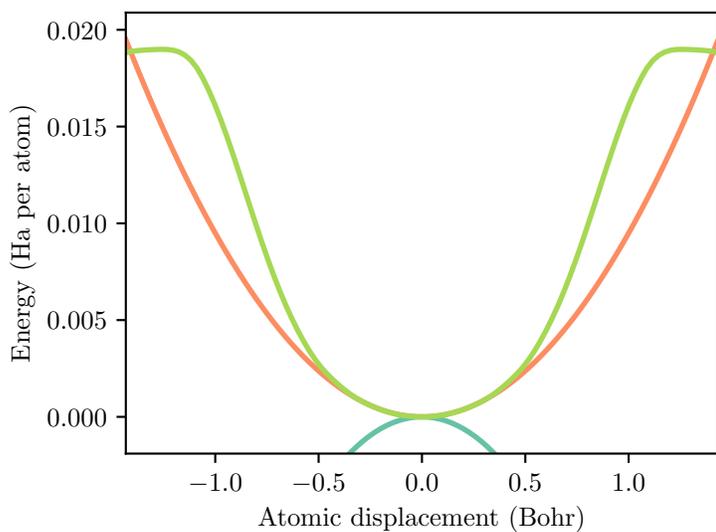
Each figure shows the true PES in green, the harmonic PES in orange, and the sixthic anharmonic PES in blue.



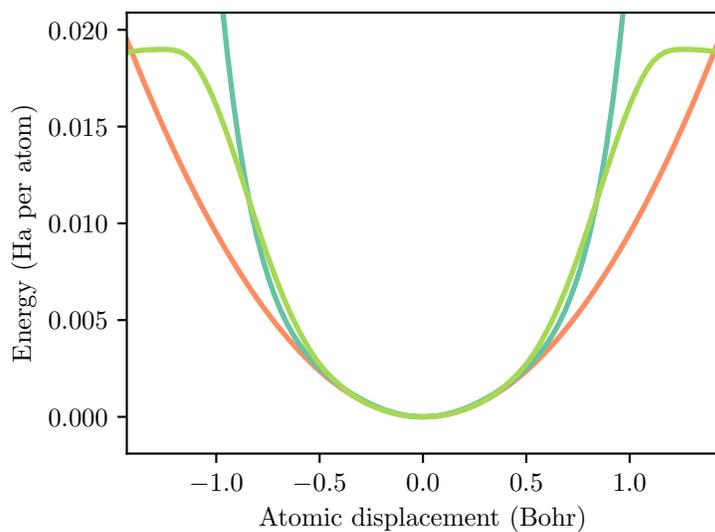
(a) The anharmonic PES interpolated from the 21 supercell, taking the PAA at first order.



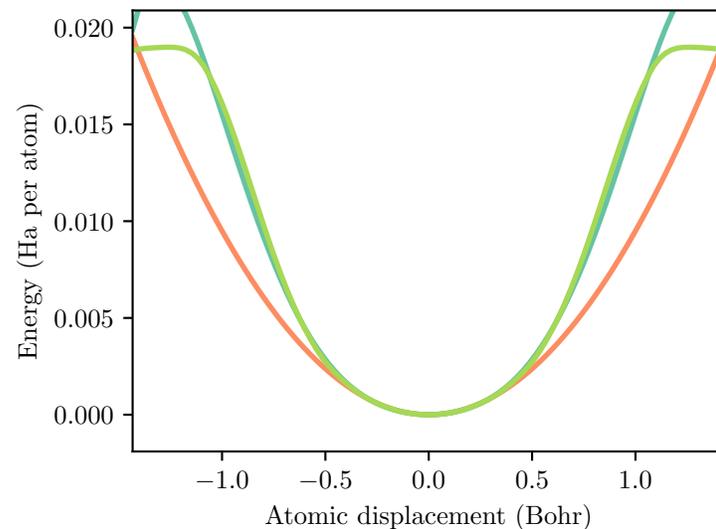
(b) The anharmonic PES interpolated from the 21 supercell, without taking the PAA.



(c) The anharmonic PES interpolated from the 31 supercell, taking the PAA at first order.



(d) The anharmonic PES interpolated from the 41 supercell, taking the PAA at first order.



(e) The anharmonic PES calculated using the 51 supercell directly.

Figure 7.15: One-dimensional slices of the PES of β titanium, along the mode on the 51 supercell with the largest real frequency.

Each figure shows the true PES in green, the harmonic PES in orange, and the sixth-order anharmonic PES in blue.

7.5.6 Limitations of `caesar`

It is unfortunately not possible to use `caesar` to perform VSCF calculations using larger supercells than those presented above, due to computational limitations. The current version of `caesar` suffers from a memory leak thought to be caused by a known bug in the compiler. This memory leak means that any calculation lasting more than about a day, depending on hardware, runs into memory limits and has to be terminated.

Due to these limitations, and to the inaccuracies of the anharmonic PES representation discussed above, it was not possible to calculate the free energy of the β phase of titanium with sufficient accuracy to make meaningful comparisons with the results of Hennig et al. [148].

Chapter 8

Conclusions

This thesis has presented a study of the vibrational modelling of periodic solids. These models have been presented as having two separate stages; first constructing a representation of the Born-Oppenheimer nuclear potential energy surface (PES) in the region of phase space around some static-lattice nuclear configuration, and then using this PES to calculate the free energy of the solid.

The investigation into representations of the PES has focussed on methods which represent the PES as a sum of independent basis functions, and in particular on methods which use the polynomial terms of a truncated Taylor expansion to construct these basis functions. An approximation to this representation, the principal axes approximation (PAA), has also been investigated.

The investigation into the calculation of the free energy has focussed on mean-field methods, which model a system with many dimensions as a set of smaller-dimensional subsystems, where each subsystem only interacts with the averaged configuration of each other subsystem. In particular, the method of vibrational self-consistent field theory (VSCF) has been investigated.

Two important properties of these methods have been analysed. The first property, symmetry invariance, applies to methods whose results do not change unphysically when the system is transformed by a symmetry. The second property, size consistency, applies to methods whose results are consistent as the size of the system being simulated is increased.

8.1 Representations of the potential energy surface

In section 4.1.3, a method was presented by which the number of free parameters in a PES representation can be dramatically reduced by exploiting the symmetries of the PES. The advantages of this method are system-dependent, but for many systems this method offers an improvement in accuracy and a significant reduction in computational cost, and this has been demonstrated in section 7.5.1. This result matches that presented by Fu et al. [13], which was published while the investigations presented here were ongoing.

The method for exploiting the symmetry of the PES presented in this work can currently only be applied to PES which are constructed from polynomial terms. However, the principles behind the method are more widely applicable, and should be applied to other PES representations in future work.

In section 4.2, the requirements of size consistency on PES representations have been presented. If a PES representation is size-consistent then calculations performed using a large number of atoms can be approximated by interpolating calculations performed using a smaller number of atoms, potentially offering another dramatic reduction in computational cost. If a PES representation is not size-consistent, then the results calculated using that PES will change depending on the size of the system being simulated, and it may not be possible to converge the results of the calculation with respect to the size of the simulation. These results have been demonstrated in section 7.5.4.

The main conclusions of the analysis of size consistency of the PES presented here match those presented in Hirata et al. [71] and papers cited therein, but the details of these conclusions differ. Where the analysis of Hirata et al. [71] and related papers suggests that the scaling of a term $\prod_i u_i^{n_i}$ with the size of the simulation depends only on the power $\sum_i n_i$ of that term, the symmetry analysis in section 4.2.3 of this work demonstrates that different terms with the same power can scale differently, and that this scaling is determined by how the term transforms under the symmetries of the system. This means that a size-consistent PES can contain terms which are excluded by the analysis of Hirata et al. [71] and related papers.

This work has demonstrated that the PAA is not size-consistent. This suggests that the PAA in its current form should not be used when modelling periodic systems. Other approximations should be found which fill the same roll as the PAA but which do not violate size consistency. Possible avenues of research include potentials which consider all coupling between q-points within a certain distance of one another within the Brillouin zone, and potentials which consider all coupling between normal modes with harmonic frequencies within a certain separation of one another.

The use of polynomial basis functions in this work enabled much of the investigation into symmetry invariance and size consistency, but in chapter 7 it was demonstrated that these polynomials may not be well suited to capturing the functional forms of the PES of real systems. Future work should investigate the symmetry

invariance and size consistency of other forms of basis functions, and methods should be developed which can use a range of basis functions as best suited to the problem. In particular, the adaptive representations mentioned in section 3.4 appear to offer significant improvements over polynomials, potentially allowing for sophisticated and highly-automated representations of the PES.

One constraint which would need to be considered when improving the behaviour of the PES at large displacements is invariance under the action of the symmetries which do not preserve absolute displacement, as discussed in section 2.4.1. For a PES to be invariant under these symmetries, it would need to have a periodic dependence on the normal mode co-ordinates. This suggests that the PES should be constructed in terms of Fourier-transformed co-ordinates of the form $\nu = \int e^{2\pi i\nu \cdot \mathbf{u}} \mathbf{u}$. This Fourier transform is distinct from the discrete Fourier transform which transforms R-vectors to q-points and vice versa.

8.2 Vibrational self-consistent field theory

In section 5.1.1, a symmetry-invariant formulation of the VSCF method has been presented, which approximates the Hamiltonian of the system as a sum of single-subspace Hamiltonians rather than as a sum of single-mode Hamiltonians. In future work, this formulation should be used in place of VSCF formulations which are not symmetry-invariant. Unfortunately, this formulation dramatically increases the computational cost of the method for some systems.

In section 5.2, it has been demonstrated that most formulations of VSCF are not size-consistent. In the large-supercell limit, VSCF becomes formally equivalent to an effective harmonic theory, and any size-consistent formulation of VSCF must also be equivalent to an effective harmonic theory outside of this limit. This result matches that presented in Hirata et al. [71] and works cited therein, although the differences with those works discussed in the previous section mean that the size-consistent formulation of VSCF presented here differs from the size-consistent formulations of VSCF presented in Hermes et al. [133], in that additional terms are included in the PES. An important consequence of this difference is that the VSCF method presented here may include a non-zero offset $\langle \mathbf{u} \rangle$ at the Γ -point, which is missing from the methods of Hermes et al. [133].

The fact that size-consistent VSCF is equivalent to an effective harmonic theory has major consequences for the usefulness of the method. VSCF is significantly more computationally expensive than effective harmonic methods, and if the results of VSCF are only as accurate as those calculated using effective harmonic methods then the VSCF method should not be used for periodic systems.

While VSCF in its current form cannot be made size-consistent without being reduced to an effective harmonic theory, it may be possible to construct a mean-field theory which does not have this property. In particular, it may be possible to explicitly include the correlation between neighbouring q-points when cal-

culating mean-field expectations, such that a single-subspace potential is constructed as

$$V_{\xi}(\mathbf{u}_{\xi}) = \sum_{\psi \in \bar{\xi}} \langle \psi | P_{\psi}(\mathbf{u}_{\xi}) V(\mathbf{u}) | \psi \rangle . \tag{8.1}$$

This would potentially avoid the problems with the VSCF approximation, although it is unclear what the computational cost of this method would be or how this method would correspond to a system with physically-meaningful states.

8.3 Final thoughts

In light of the shortcomings of the VSCF method presented in this work, attention should be turned to other methods which can provide more accurate results than effective harmonic methods. As discussed in section 3.5, the methods which appear to offer increased accuracy for the least computational cost are those based on molecular dynamics and Monte-Carlo simulations, including ring-polymer molecular dynamics (RPMD) and diffusion Monte-Carlo (DMC).

These methods are traditionally combined with direct sampling of the PES, a process which can require huge numbers of electronic structure calculations, and these calculations are often required to contain large numbers of atoms.

There is no reason in principle why RPMD and DMC cannot be used with the PES representations presented in this work. Such a combination might be slightly less accurate than sampling the true PES directly, but should be significantly more accurate than effective harmonic methods. This combination would also enable the use of the developments to the representation of the PES presented in this work, possibly allowing for dramatic reductions in the computational cost of these methods.

It is hoped that the ideas presented in this work can be applied by other authors to other forms of vibrational calculations, and that this will help advance the field towards the ultimate goal of producing a vibrational method which is both accurate and efficient.

Bibliography

- [1] Ion Errea. Approaching the strongly anharmonic limit with ab initio calculations of materials' vibrational properties – a colloquium*. *The European Physical Journal B*, 89(11):237, Nov 2016. ISSN 1434-6036.
- [2] Johannes Hoja, Anthony M. Reilly, and Alexandre Tkatchenko. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *WIREs Computational Molecular Science*, 7(1):e1294, 2017. doi: 10.1002/wcms.1294.
- [3] W. Kohn. An essay on condensed matter physics in the twentieth century. *Rev. Mod. Phys.*, 71:S59–S77, Mar 1999. doi: 10.1103/RevModPhys.71.S59.
- [4] Emily A. Carter. Challenges in modeling materials properties without experimental input. *Science*, 321(5890): 800–803, 2008. ISSN 0036-8075. doi: 10.1126/science.1158009.
- [5] Sang Wook Yi. The nature of model-based understanding in condensed matter physics. *Mind & Society*, 3(1): 81–91, Mar 2002. ISSN 1860-1839. doi: 10.1007/BF02511868.
- [6] Andrea Cavagna and Irene Giardina. Bird flocks as condensed matter. *Annual Review of Condensed Matter Physics*, 5(1):183–207, 2014. doi: 10.1146/annurev-conmatphys-031113-133834.
- [7] Tao Pang. An introduction to computational physics. *American Journal of Physics*, 67(1):94–95, 1999. doi: 10.1119/1.19198.
- [8] D. Raabe. Challenges in computational materials science. *Advanced Materials*, 14(9):639–650, 2002. doi: 10.1002/1521-4095(20020503)14:9<639::AID-ADMA639>3.0.CO;2-7.
- [9] Richard J. Needs and Chris J. Pickard. Perspective: Role of structure prediction in materials discovery and design. *APL Materials*, 4(5):053210, 2016. doi: 10.1063/1.4949361.
- [10] Artem R. Oganov, Chris J. Pickard, Qiang Zhu, and Richard J. Needs. Structure prediction drives materials discovery. *Nature Reviews Materials*, 4:331–348, 2019. doi: 10.1038/s41578-019-0101-8.
- [11] Richard JD Tilley. *Crystals and crystal structures*. John Wiley & Sons, 2020.
- [12] Max Born and Kun Huang. *Dynamical theory of crystal lattices*. Clarendon press, 1954.

- [13] Lyuwen Fu, Mordechai Kornbluth, Zhengqian Cheng, and Chris A. Marianetti. Group theoretical approach to computing phonons and their interactions. *Phys. Rev. B*, 100:014303, Jul 2019. doi: 10.1103/PhysRevB.100.014303.
- [14] So Hirata, Murat Keçeli, and Kiyoshi Yagi. First-principles theories for anharmonic lattice vibrations. *The Journal of Chemical Physics*, 133(3):034109, 2010. doi: 10.1063/1.3462237.
- [15] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927. doi: 10.1002/andp.19273892002.
- [16] Galina M. Chaban and R. Benny Gerber. Anharmonic vibrational spectroscopy calculations with electronic structure potentials: comparison of MP2 and DFT for organic molecules. *Theoretical Chemistry Accounts*, 120(1):273–279, May 2008. ISSN 1432-2234. doi: 10.1007/s00214-007-0299-1.
- [17] Patrick Huang and Emily A. Carter. Advances in correlated electronic structure methods for solids, surfaces, and nanostructures. *Annual Review of Physical Chemistry*, 59(1):261–290, 2008. doi: 10.1146/annurev.physchem.59.032607.093528. PMID: 18031211.
- [18] Yousef Saad, James R. Chelikowsky, and Suzanne M. Shontz. Numerical methods for electronic structure calculations of materials. *SIAM Review*, 52(1):3–54, 2010. doi: 10.1137/060651653.
- [19] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, 2017. doi: 10.1080/00268976.2017.1333644.
- [20] Igor Ying Zhang and Andreas Grüneis. Coupled cluster theory in materials science. *Frontiers in Materials*, 6:123, 2019. ISSN 2296-8016. doi: 10.3389/fmats.2019.00123.
- [21] Ove Christiansen. Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties. *Phys. Chem. Chem. Phys.*, 9:2942–2953, 2007. doi: 10.1039/B618764A.
- [22] Venkat Kapil, Edgar Engel, Mariana Rossi, and Michele Ceriotti. Assessment of approximate methods for anharmonic free energies. *Journal of Chemical Theory and Computation*, 15(11):5845–5857, 2019. doi: 10.1021/acs.jctc.9b00596. PMID: 31532997.
- [23] Tapta Kanchan Roy and R. Benny Gerber. Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications. *Phys. Chem. Chem. Phys.*, 15:9468–9492, 2013. doi: 10.1039/C3CP50739D.
- [24] Robert J. N. Baldock, Livia B. Pártay, Albert P. Bartók, Michael C. Payne, and Gábor Csányi. Determining pressure-temperature phase diagrams of materials. *Phys. Rev. B*, 93:174108, May 2016. doi: 10.1103/PhysRevB.93.174108.
- [25] Göran Grimvall, Blanka Magyari-Köpe, Vidvuds Ozoliņš, and Kristin A. Persson. Lattice instabilities in metallic elements. *Rev. Mod. Phys.*, 84:945–986, Jun 2012. doi: 10.1103/RevModPhys.84.945.

- [26] Bartomeu Monserrat, N. D. Drummond, and R. J. Needs. Anharmonic vibrational properties in periodic systems: energy, electron-phonon coupling, and stress. *Phys. Rev. B*, 87:144302, Apr 2013. doi: 10.1103/PhysRevB.87.144302.
- [27] S. Ponc e, Y. Gillet, J. Laflamme Janssen, A. Marini, M. Verstraete, and X. Gonze. Temperature dependence of the electronic structure of semiconductors and insulators. *The Journal of Chemical Physics*, 143(10):102813, 2015. doi: 10.1063/1.4927081.
- [28] Marios Zacharias, Matthias Scheffler, and Christian Carbogno. Fully anharmonic nonperturbative theory of vibronically renormalized electronic band structures. *Phys. Rev. B*, 102:045126, Jul 2020. doi: 10.1103/PhysRevB.102.045126.
- [29] Ion Errea, Matteo Calandra, and Francesco Mauri. Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: Application to platinum and palladium hydrides. *Phys. Rev. B*, 89:064302, Feb 2014. doi: 10.1103/PhysRevB.89.064302.
- [30] Ion Errea, Matteo Calandra, and Francesco Mauri. Huge anharmonic effects in superconducting hydrides and transition metal dichalcogenides. *physica status solidi (b)*, 251(12):2556–2562, 2014. ISSN 1521-3951. doi: 10.1002/pssb.201451164.
- [31] Antonios M. Alvertis, Raj Pandya, Loreta A. Muscarella, Nipun Sawhney, Malgorzata Nguyen, Bruno Ehrler, Akshay Rao, Richard H. Friend, Alex W. Chin, and Bartomeu Monserrat. Impact of exciton delocalization on exciton-vibration interactions in organic semiconductors. *Phys. Rev. B*, 102:081122, Aug 2020. doi: 10.1103/PhysRevB.102.081122.
- [32] Miguel Borinaga, Ion Errea, Matteo Calandra, Francesco Mauri, and Aitor Bergara. Anharmonic effects in atomic hydrogen: Superconductivity and lattice dynamical stability. *Phys. Rev. B*, 93:174308, May 2016. doi: 10.1103/PhysRevB.93.174308.
- [33] Cristina Puzzarini, Julien Bloino, Nicola Tasinato, and Vincenzo Barone. Accuracy and interpretability: The devil and the holy grail. new routes across old boundaries in computational spectroscopy. *Chemical Reviews*, 119(13):8131–8191, 2019. doi: 10.1021/acs.chemrev.9b00007. PMID: 31187984.
- [34] Ove Christiansen. Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations. *Phys. Chem. Chem. Phys.*, 14:6672–6687, 2012. doi: 10.1039/C2CP40090A.
- [35] Marco De La Pierre, Raffaella Demichelis, and Roberto Dovesi. *Vibrational Spectroscopy of Minerals Through Ab Initio Methods*, chapter 10, pages 341–374. John Wiley & Sons, Ltd, 2016. ISBN 9781118845226. doi: 10.1002/9781118845226.ch10.
- [36] Anna Putrino and Michele Parrinello. Anharmonic Raman spectra in high-pressure ice from ab initio simulations. *Phys. Rev. Lett.*, 88:176401, Apr 2002. doi: 10.1103/PhysRevLett.88.176401.
- [37] Peter Seidler, Jacob Kongsted, and Ove Christiansen. Calculation of vibrational infrared intensities and Raman activities using explicit anharmonic wave functions. *The Journal of Physical Chemistry A*, 111(44):11205–11213, 2007. doi: 10.1021/jp070327n. PMID: 17474728.

- [38] Ruidy Nemausat, Delphine Cabaret, Christel Gervais, Christian Brouder, Nicolas Trcera, Amélie Bordage, Ion Errea, and Francesco Mauri. Phonon effects on x-ray absorption and nuclear magnetic resonance spectroscopies. *Phys. Rev. B*, 92:144310, Oct 2015. doi: 10.1103/PhysRevB.92.144310.
- [39] Ian H. Godtlielsen and Ove Christiansen. Calculating vibrational spectra without determining excited eigenstates: Solving the complex linear equations of damped response theory for vibrational configuration interaction and vibrational coupled cluster states. *The Journal of Chemical Physics*, 143(13):134108, 2015. doi: 10.1063/1.4932010.
- [40] R. Egoavil, N. Gauquelin, G.T. Martinez, S. Van Aert, G. Van Tendeloo, and J. Verbeeck. Atomic resolution mapping of phonon excitations in stem-eels experiments. *Ultramicroscopy*, 147:1 – 7, 2014. ISSN 0304-3991. doi: <https://doi.org/10.1016/j.ultramic.2014.04.011>.
- [41] Lorenzo Paulatto, Francesco Mauri, and Michele Lazzeri. Anharmonic properties from a generalized third-order ab initio approach: Theory and applications to graphite and graphene. *Phys. Rev. B*, 87:214303, Jun 2013. doi: 10.1103/PhysRevB.87.214303.
- [42] Alessandro Erba, Jefferson Maul, Matteo Ferrabone, Philippe Carbonnière, Michel Rérat, and Roberto Dovesi. Anharmonic vibrational states of solids from DFT calculations. part I: Description of the potential energy surface. *Journal of Chemical Theory and Computation*, 15(6):3755–3765, 2019. doi: 10.1021/acs.jctc.9b00293. PMID: 31038943.
- [43] T Tadano, Y Gohda, and S Tsuneyuki. Anharmonic force constants extracted from first-principles molecular dynamics: applications to heat transfer simulations. *Journal of Physics: Condensed Matter*, 26(22):225402, may 2014. doi: 10.1088/0953-8984/26/22/225402.
- [44] Atsushi Togo, Laurent Chaput, and Isao Tanaka. Distributions of phonon lifetimes in brillouin zones. *Phys. Rev. B*, 91:094306, Mar 2015. doi: 10.1103/PhysRevB.91.094306.
- [45] Terumasa Tadano and Shinji Tsuneyuki. First-principles lattice dynamics method for strongly anharmonic crystals. *Journal of the Physical Society of Japan*, 87(4):041015, 2018. doi: 10.7566/JPSJ.87.041015.
- [46] Jieqiong Zhang, Junzhi Cui, Zihao Yang, and Shangkun Shen. Thermodynamic properties and thermoelastic constitutive relation for cubic crystal structures based on improved free energy. *Computational Mechanics*, 65: 989–1003, 2020. doi: 10.1007/s00466-019-01804-1.
- [47] Connor A. Occhialini, Gian G. Guzmán-Verri, Sahan U. Handunkanda, and Jason N. Hancock. Negative thermal expansion near the precipice of structural stability in open perovskites. *Frontiers in Chemistry*, 6:545, 2018. ISSN 2296-2646. doi: 10.3389/fchem.2018.00545.
- [48] Lorenzo Monacelli, Ion Errea, Matteo Calandra, and Francesco Mauri. Pressure and stress tensor of complex anharmonic crystals within the stochastic self-consistent harmonic approximation. *Phys. Rev. B*, 98:024106, Jul 2018. doi: 10.1103/PhysRevB.98.024106.

- [49] Martin T Dove and Hong Fang. Negative thermal expansion and associated anomalous physical properties: review of the lattice dynamics theoretical foundation. *Reports on Progress in Physics*, 79(6):066503, may 2016. doi: 10.1088/0034-4885/79/6/066503.
- [50] Francis VanGessel, Jie Peng, and Peter W. Chung. A review of computational phononics: the bulk, interfaces and surfaces. *Journal of Materials Science*, 53:5641–5683, 2018. doi: 10.1007/s10853-017-1728-8.
- [51] Erik Fransson and Paul Erhart. Defects from phonons: Atomic transport by concerted motion in simple crystalline metals. *Acta Materialia*, 196:770 – 775, 2020. ISSN 1359-6454. doi: <https://doi.org/10.1016/j.actamat.2020.06.040>.
- [52] Alexei A Maradudin, Elliott Waters Montroll, George Herbert Weiss, and IP Ipatova. *Theory of lattice dynamics in the harmonic approximation*, volume 3. Academic press New York, 1963.
- [53] Neil W Ashcroft, N David Mermin, et al. *Solid state physics*. New York: Holt, Rinehart and Winston, 1976.
- [54] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part I. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, 1928. doi: 10.1017/S0305004100011919.
- [55] Max Planck. Ueber irreversible strahlungsvorgänge. *Annalen der Physik*, 306(1):69–122, 1900. doi: 10.1002/andp.19003060105.
- [56] Bo Li, Di Zhou, and Yilong Han. Assembly and phase transitions of colloidal crystals. *Nature Reviews Materials*, 1(2):15011, Jan 2016. ISSN 2058-8437. doi: 10.1038/natrevmats.2015.11.
- [57] T. Kreibich and E. K. U. Gross. Multicomponent density-functional theory for electrons and nuclei. *Phys. Rev. Lett.*, 86:2984–2987, Apr 2001. doi: 10.1103/PhysRevLett.86.2984.
- [58] Norm M. Tubman, Ilkka Kylänpää, Sharon Hammes-Schiffer, and David M. Ceperley. Beyond the Born-Oppenheimer approximation with quantum Monte Carlo methods. *Phys. Rev. A*, 90:042507, Oct 2014. doi: 10.1103/PhysRevA.90.042507.
- [59] Grigory Kolesov, Efthimios Kaxiras, and Efstratios Manousakis. Density functional theory beyond the Born-Oppenheimer approximation: Accurate treatment of the ionic zero-point motion. *Phys. Rev. B*, 98:195112, Nov 2018. doi: 10.1103/PhysRevB.98.195112.
- [60] Ville J. Härkönen, Robert van Leeuwen, and E. K. U. Gross. Many-body Green’s function theory of electrons and nuclei beyond the Born-Oppenheimer approximation. *Phys. Rev. B*, 101:235153, Jun 2020. doi: 10.1103/PhysRevB.101.235153.
- [61] Takehiro Yonehara, Kota Hanasaki, and Kazuo Takatsuka. Fundamental approaches to nonadiabaticity: Toward a chemical theory beyond the Born-Oppenheimer paradigm. *Chemical Reviews*, 112(1):499–542, 2012. doi: 10.1021/cr200096s. PMID: 22077497.
- [62] Sara Kadkhodaei, Qi-Jun Hong, and Axel van de Walle. Free energy calculation of mechanically unstable but dynamically stabilized bcc titanium. *Phys. Rev. B*, 95:064101, Feb 2017. doi: 10.1103/PhysRevB.95.064101.

- [63] Bartomeu Monserrat, Neil D. Drummond, Philip Dalladay-Simpson, Ross T. Howie, Pablo López Ríos, Eugene Gregoryanz, Chris J. Pickard, and Richard J. Needs. Structure and metallicity of phase V of hydrogen. *Phys. Rev. Lett.*, 120:255701, Jun 2018. doi: 10.1103/PhysRevLett.120.255701.
- [64] Can P. Koçer, Kent J. Griffith, Clare P. Grey, and Andrew J. Morris. Lithium diffusion in niobium tungsten oxide shear structures. *Chemistry of Materials*, 32(9):3980–3989, 2020. doi: 10.1021/acs.chemmater.0c00483.
- [65] Philip Brown, Konstantin Semeniuk, Diandian Wang, Bartomeu Monserrat, Chris J. Pickard, and F. Malte Grosche. Strong coupling superconductivity in a quasiperiodic host-guest structure. *Science Advances*, 4(4), 2018. doi: 10.1126/sciadv.aao4793.
- [66] Marvin Hass. Lattice vibrations in ionic crystals of finite size. *Phys. Rev. Lett.*, 13:429–431, Oct 1964. doi: 10.1103/PhysRevLett.13.429.
- [67] M.L. Klein and G.K. Horton. The rise of self-consistent phonon theory. *Journal of Low Temperature Physics*, 9:151–166, 1972. doi: 10.1007/BF00654839.
- [68] Keivan Esfarjani and Harold T. Stokes. Method to extract anharmonic force constants from first principles calculations. *Phys. Rev. B*, 77:144112, Apr 2008. doi: 10.1103/PhysRevB.77.144112.
- [69] Leon Brillouin. Wave propagation in periodic structures: electric filters and crystal lattices. 1953.
- [70] Jonathan H. Lloyd-Williams and Bartomeu Monserrat. Lattice dynamics and electron-phonon coupling calculations using nondiagonal supercells. *Phys. Rev. B*, 92:184301, Nov 2015. doi: 10.1103/PhysRevB.92.184301.
- [71] So Hirata, Murat Keçeli, Yu-ya Ohnishi, Olaseni Sode, and Kiyoshi Yagi. Extensivity of energy and electronic and vibrational structure methods for crystals. *Annual Review of Physical Chemistry*, 63(1):131–153, 2012. doi: 10.1146/annurev-physchem-032511-143718. PMID: 22224701.
- [72] Jean-Louis Calais and Wolf Weyrich. Finite and infinite Born-von Kármán regions. *International Journal of Quantum Chemistry*, 63(1):223–227, 1997. doi: 10.1002/(SICI)1097-461X(1997)63:1<223::AID-QUA23>3.0.CO;2-B.
- [73] A. A. Maradudin and S. H. Vosko. Symmetry properties of the normal vibrations of a crystal. *Rev. Mod. Phys.*, 40:1–37, Jan 1968. doi: 10.1103/RevModPhys.40.1.
- [74] Hermann von Helmholtz. *Über die Erhaltung der Kraft*. Number 1. Engelmann, 1907.
- [75] Robert Eisberg and Robert Resnick. *Quantum physics of atoms, molecules, solids, nuclei, and particles*. 1985.
- [76] Guglielmo Mazzola, Andrea Zen, and Sandro Sorella. Finite-temperature electronic simulations without the Born-Oppenheimer constraint. *The Journal of Chemical Physics*, 137(13):134112, 2012. doi: 10.1063/1.4755992.
- [77] Jonathan P. Alborzpour, David P. Tew, and Scott Habershon. Efficient and accurate evaluation of potential energy matrix elements for quantum dynamics using gaussian process regression. *The Journal of Chemical Physics*, 145(17):174112, 2016. doi: 10.1063/1.4964902.

- [78] Susan B. Sinnott and Donald W. Brenner. Three decades of many-body potentials in materials research. *MRS Bulletin*, 37(5):469–473, 2012. doi: 10.1557/mrs.2012.88.
- [79] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, Oct 1934. doi: 10.1103/PhysRev.46.618.
- [80] C. David Sherrill and Henry F. Schaefer. The configuration interaction method: Advances in highly correlated approaches. volume 34 of *Advances in Quantum Chemistry*, pages 143 – 269. Academic Press, 1999. doi: [https://doi.org/10.1016/S0065-3276\(08\)60532-8](https://doi.org/10.1016/S0065-3276(08)60532-8).
- [81] Andrew G. Taube and Rodney J. Bartlett. Frozen natural orbital coupled-cluster theory: Forces and application to decomposition of nitroethane. *The Journal of Chemical Physics*, 128(16):164101, 2008. doi: 10.1063/1.2902285.
- [82] Henrik Koch, Hans Jørgen Aa. Jensen, Poul Jørgensen, Trygve Helgaker, Gustavo E. Scuseria, and Henry F. Schaefer. Coupled cluster energy derivatives. analytic hessian for the closed-shell coupled cluster singles and doubles wave function: Theory and applications. *The Journal of Chemical Physics*, 92(8):4924–4940, 1990. doi: 10.1063/1.457710.
- [83] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.
- [84] Guo-Xu Zhang, Anthony M Reilly, Alexandre Tkatchenko, and Matthias Scheffler. Performance of various density-functional approximations for cohesive properties of 64 bulk solids. *New Journal of Physics*, 20(6):063020, jun 2018. doi: 10.1088/1367-2630/aac7f0.
- [85] E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, and J. M. Soler. Linear-scaling ab-initio calculations for large and complex systems. *physica status solidi (b)*, 215(1):809–817, 1999. doi: 10.1002/(SICI)1521-3951(199909)215:1<809::AID-PSSB809>3.0.CO;2-0.
- [86] A. R. Williams, Peter J. Feibelman, and N. D. Lang. Green’s-function methods for electronic-structure calculations. *Phys. Rev. B*, 26:5433–5444, Nov 1982. doi: 10.1103/PhysRevB.26.5433.
- [87] Hong Jiang. Electronic band structure from first-principles Green’s function approach: theory and implementations. *Frontiers of Chemistry in China*, 6(4):253–268, Dec 2011. ISSN 1673-3614. doi: 10.1007/s11458-011-0261-6.
- [88] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.*, 78:865–951, Aug 2006. doi: 10.1103/RevModPhys.78.865.
- [89] J. M. Tomczak, P. Liu, A. Toschi, G. Kresse, and K. Held. Merging GW with DMFT and non-local correlations beyond. *The European Physical Journal Special Topics*, 226(11):2565–2590, Jul 2017. ISSN 1951-6401. doi: 10.1140/epjst/e2017-70053-1.

- [90] Joseph C. A. Prentice and R. J. Needs. Using forces to accelerate first-principles anharmonic vibrational calculations. *Phys. Rev. Materials*, 1:023801, Jul 2017. doi: 10.1103/PhysRevMaterials.1.023801.
- [91] Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, Jul 2001. doi: 10.1103/RevModPhys.73.515.
- [92] Michael Hutcheon. Stochastic nodal surfaces in quantum Monte Carlo calculations. *Phys. Rev. E*, 102:042105, Oct 2020. doi: 10.1103/PhysRevE.102.042105.
- [93] Y. Y. F. Liu, B. Andrews, and G. J. Conduit. Direct evaluation of the force constant matrix in quantum Monte Carlo. *The Journal of Chemical Physics*, 150(3):034104, 2019. doi: 10.1063/1.5070138.
- [94] D. Toffoli, J. Kongsted, and O. Christiansen. Automatic generation of potential energy and property surfaces of polyatomic molecules in normal coordinates. *The Journal of Chemical Physics*, 127(20):204106, 2007. doi: 10.1063/1.2805085.
- [95] Charles W. Bauschlicher and Peter R. Taylor. Symmetry and equivalence restrictions in electronic structure calculations. *Theoretica chimica acta*, 74(1):63–73, Jul 1988. ISSN 1432-2234. doi: 10.1007/BF00570553.
- [96] Guntram Rauhut. Efficient calculation of potential energy surfaces for the generation of vibrational wave functions. *The Journal of Chemical Physics*, 121(19):9313–9322, 2004. doi: 10.1063/1.1804174.
- [97] David M. Benoit. Fast vibrational self-consistent field calculations through a reduced mode-mode coupling scheme. *The Journal of Chemical Physics*, 120(2):562–573, 2004. doi: 10.1063/1.1631817.
- [98] Dominik Oschetzki, Michael Neff, Patrick Meier, Florian Pfeiffer, and Guntram Rauhut. Selected aspects concerning the efficient calculation of vibrational spectra beyond the harmonic approximation. *Croatica Chemica Acta*, 85(4):379–390, 2012.
- [99] Matthew R. Hermes and So Hirata. Stochastic many-body perturbation theory for anharmonic molecular vibrations. *The Journal of Chemical Physics*, 141(8):084105, 2014. doi: 10.1063/1.4892614.
- [100] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, Apr 2010. doi: 10.1103/PhysRevLett.104.136403.
- [101] Jacob Kongsted and Ove Christiansen. Automatic generation of force fields and property surfaces for use in variational vibrational calculations of anharmonic vibrational energies and zero-point vibrational averaged properties. *The Journal of Chemical Physics*, 125(12):124108, 2006. doi: 10.1063/1.2352734.
- [102] John C. Thomas and Anton Van der Ven. Finite-temperature properties of strongly anharmonic and mechanically unstable crystal phases from first principles. *Phys. Rev. B*, 88:214111, Dec 2013. doi: 10.1103/PhysRevB.88.214111.

- [103] Stuart Carter, Susan J. Culik, and Joel M. Bowman. Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of co adsorbed on cu(100). *The Journal of Chemical Physics*, 107(24):10458–10469, 1997. doi: 10.1063/1.474210.
- [104] Michel Krizek, Pekka Neittaanmaki, and Rolf Stenberg. *Finite element methods: fifty years of the Courant element*. CRC Press, 2016.
- [105] D.J. Hooton. LI. a new treatment of anharmonicity in lattice thermodynamics: I. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 46(375):422–432, 1955. doi: 10.1080/14786440408520575.
- [106] Liang-Feng Huang, Xue-Zeng Lu, Emrys Tennesen, and James M. Rondinelli. An efficient ab-initio quasiharmonic approach for the thermodynamics of solids. *Computational Materials Science*, 120:84 – 93, 2016. ISSN 0927-0256. doi: <https://doi.org/10.1016/j.commatsci.2016.04.012>.
- [107] Alessandro Erba, Jefferson Maul, Matteo Ferrabone, Roberto Dovesi, Michel Rérat, and Philippe Carbonnière. Anharmonic vibrational states of solids from DFT calculations. part II: Implementation of the VSCF and VCI methods. *Journal of Chemical Theory and Computation*, 15(6):3766–3777, 2019. doi: 10.1021/acs.jctc.9b00294. PMID: 31038948.
- [108] Joel M. Bowman. Self-consistent field energies and wavefunctions for coupled oscillators. *The Journal of Chemical Physics*, 68(2):608–610, 1978. doi: 10.1063/1.435782.
- [109] Grady D. Carney, Ludwig L. Sprandel, and C. William Kern. *Variational Approaches to Vibration-Rotation Spectroscopy for Polyatomic Molecules*, pages 305–379. John Wiley & Sons, Ltd, 2007. ISBN 9780470142561. doi: 10.1002/9780470142561.ch6.
- [110] M. Cohen, Susanne Greita, and R.P. McEarchran. Approximate and exact quantum mechanical energies and eigenfunctions for a system of coupled oscillators. *Chemical Physics Letters*, 60(3):445 – 450, 1979. ISSN 0009-2614. doi: [https://doi.org/10.1016/0009-2614\(79\)80609-0](https://doi.org/10.1016/0009-2614(79)80609-0).
- [111] R.B. Gerber and M.A. Ratner. A semiclassical self-consistent field (SC SCF) approximation for eigenvalues of coupled-vibration systems. *Chemical Physics Letters*, 68(1):195 – 198, 1979. ISSN 0009-2614. doi: [https://doi.org/10.1016/0009-2614\(79\)80099-8](https://doi.org/10.1016/0009-2614(79)80099-8).
- [112] Mikkel Bo Hansen, Ove Christiansen, Daniele Toffoli, and Jacob Kongsted. A virtual vibrational self-consistent-field method for efficient calculation of molecular vibrational partition functions and thermal effects on molecular properties. *The Journal of Chemical Physics*, 128(17):174106, 2008. doi: 10.1063/1.2912184.
- [113] Tapta Kanchan Roy and M. Durga Prasad. A thermal self-consistent field theory for the calculation of molecular vibrational partition functions. *The Journal of Chemical Physics*, 131(11):114102, 2009. doi: 10.1063/1.3213568.
- [114] Lawrence S. Norris, Mark A. Ratner, Adrian E. Roitberg, and R. B. Gerber. Møller-plesset perturbation theory applied to vibrational problems. *The Journal of Chemical Physics*, 105(24):11261–11267, 1996. doi: 10.1063/1.472922.

- [115] Nikita Matsunaga, Galina M. Chaban, and R. Benny Gerber. Degenerate perturbation theory corrections for the vibrational self-consistent field approximation: Method and applications. *The Journal of Chemical Physics*, 117(8):3541–3547, 2002. doi: 10.1063/1.1494978.
- [116] Ove Christiansen. A second quantization formulation of multimode dynamics. *The Journal of Chemical Physics*, 120(5):2140–2148, 2004. doi: 10.1063/1.1637578.
- [117] Felix Hummel. Finite temperature coupled cluster theories for extended systems. *Journal of Chemical Theory and Computation*, 14(12):6505–6514, 2018. doi: 10.1021/acs.jctc.8b00793. PMID: 30354112.
- [118] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, Nov 1985. doi: 10.1103/PhysRevLett.55.2471.
- [119] Mark E. Tuckerman and Michele Parrinello. Integrating the Car-Parrinello equations. I. basic integration techniques. *The Journal of Chemical Physics*, 101(2):1302–1315, 1994. doi: 10.1063/1.467823.
- [120] Rolf Lustig. Statistical thermodynamics in the classical molecular dynamics ensemble. I. fundamentals. *The Journal of Chemical Physics*, 100(4):3048–3059, 1994. doi: 10.1063/1.466446.
- [121] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984. doi: 10.1063/1.447334.
- [122] Mark E. Tuckerman, Dominik Marx, Michael L. Klein, and Michele Parrinello. Efficient and general algorithms for path integral Car-Parrinello molecular dynamics. *The Journal of Chemical Physics*, 104(14):5579–5588, 1996. doi: 10.1063/1.471771.
- [123] Marie-Pierre Gageot. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys. Chem. Chem. Phys.*, 12:3336–3359, 2010. doi: 10.1039/B924048A.
- [124] Michele Ceriotti, Joshua More, and David E. Manolopoulos. i-PI: A python interface for ab initio path integral molecular dynamics simulations. *Computer Physics Communications*, 185(3):1019 – 1026, 2014. ISSN 0010-4655. doi: <https://doi.org/10.1016/j.cpc.2013.10.027>.
- [125] Jianshu Cao and Gregory A. Voth. The formulation of quantum statistical mechanics based on the feynman path centroid density. I. equilibrium properties. *The Journal of Chemical Physics*, 100(7):5093–5105, 1994. doi: 10.1063/1.467175.
- [126] Ian R. Craig and David E. Manolopoulos. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *The Journal of Chemical Physics*, 121(8):3368–3373, 2004. doi: 10.1063/1.1777575.
- [127] Mariana Rossi, Michele Ceriotti, and David E. Manolopoulos. How to remove the spurious resonances from ring polymer molecular dynamics. *The Journal of Chemical Physics*, 140(23):234116, 2014. doi: 10.1063/1.4883861.
- [128] James B. Anderson. A random-walk simulation of the Schrödinger equation: H+3. *The Journal of Chemical Physics*, 63(4):1499–1503, 1975. doi: 10.1063/1.431514.

- [129] Anne B. McCoy. Diffusion Monte Carlo approaches for investigating the structure and vibrational spectra of fluxional systems. *International Reviews in Physical Chemistry*, 25(1-2):77–107, 2006. doi: 10.1080/01442350600679347.
- [130] So Hirata. Thermodynamic limit and size-consistent design. *Theoretical Chemistry Accounts*, 129(6):727–746, Aug 2011. ISSN 1432-2234. doi: 10.1007/s00214-011-0954-4.
- [131] Matthew R. Hermes and So Hirata. Diagrammatic theories of anharmonic molecular vibrations. *International Reviews in Physical Chemistry*, 34(1):71–97, 2015. doi: 10.1080/0144235X.2014.1001220.
- [132] Murat Keçeli and So Hirata. Size-extensive vibrational self-consistent field method. *The Journal of Chemical Physics*, 135(13):134108, 2011. doi: 10.1063/1.3644895.
- [133] Matthew R. Hermes, Murat Keçeli, and So Hirata. Size-extensive vibrational self-consistent field methods with anharmonic geometry corrections. *The Journal of Chemical Physics*, 136(23):234109, 2012. doi: 10.1063/1.4729602.
- [134] Matthew R. Hermes and So Hirata. Second-order many-body perturbation expansions of vibrational dyson self-energies. *The Journal of Chemical Physics*, 139(3):034111, 2013. doi: 10.1063/1.4813123.
- [135] Arindam Chakraborty, Michael V. Pak, and Sharon Hammes-Schiffer. Properties of the exact universal functional in multicomponent density functional theory. *The Journal of Chemical Physics*, 131(12):124115, 2009. doi: 10.1063/1.3236844.
- [136] Francis Birch. Finite elastic strain of cubic crystals. *Phys. Rev.*, 71:809–824, Jun 1947. doi: 10.1103/PhysRev.71.809.
- [137] Joseph C. A. Prentice, Bartomeu Monserrat, and R. J. Needs. First-principles study of the dynamic Jahn-Teller distortion of the neutral vacancy in diamond. *Phys. Rev. B*, 95:014108, Jan 2017. doi: 10.1103/PhysRevB.95.014108.
- [138] Carolin König and Ove Christiansen. Linear-scaling generation of potential energy surfaces using a double incremental expansion. *The Journal of Chemical Physics*, 145(6):064105, 2016. doi: 10.1063/1.4960189.
- [139] Murat Keçeli, So Hirata, and Kiyoshi Yagi. First-principles calculations on anharmonic vibrational frequencies of polyethylene and polyacetylene in the Γ approximation. *The Journal of Chemical Physics*, 133(3):034110, 2010. doi: 10.1063/1.3462238.
- [140] N D Woods, M C Payne, and P J Hasnip. Computing the self-consistent field in Kohn–Sham density functional theory. *Journal of Physics: Condensed Matter*, 31(45):453001, aug 2019. doi: 10.1088/1361-648x/ab31c0.
- [141] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393 – 398, 1980. ISSN 0009-2614. doi: [https://doi.org/10.1016/0009-2614\(80\)80396-4](https://doi.org/10.1016/0009-2614(80)80396-4).
- [142] P. Pulay. Improved scf convergence acceleration. *Journal of Computational Chemistry*, 3(4):556–560, 1982. doi: <https://doi.org/10.1002/jcc.540030413>.

- [143] Andrei Nikolajevits Tihonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, 4:1035–1038, 1963.
- [144] Atsushi Togo and Isao Tanaka. **Spglib**: a software library for crystal symmetry search. *arXiv cond-mat.mtrl-sci*, 2018. doi: <https://arxiv.org/abs/1808.01590>.
- [145] URL <https://libatoms.github.io/QUIP/>.
- [146] URL <https://wiki.fysik.dtu.dk/ase/>.
- [147] M.J. Rutter. C2x: A tool for visualisation and input preparation for castep and other electronic structure codes. *Computer Physics Communications*, 225:174 – 179, 2018. ISSN 0010-4655. doi: <https://doi.org/10.1016/j.cpc.2017.12.008>.
- [148] R. G. Hennig, T. J. Lenosky, D. R. Trinkle, S. P. Rudin, and J. W. Wilkins. Classical potential describes martensitic phase transformations between the α , β , and ω titanium phases. *Phys. Rev. B*, 78:054121, Aug 2008. doi: 10.1103/PhysRevB.78.054121.
- [149] Martin Schlegel. Fitting interatomic potentials to reproduce phase transitions. *Doctoral thesis*, 2020. doi: <https://doi.org/10.17863/CAM.50769>.
- [150] Sara Kadkhodaei and Ali Davariashiyani. Phonon-assisted diffusion in bcc phase of titanium and zirconium from first principles. *Phys. Rev. Materials*, 4:043802, Apr 2020. doi: 10.1103/PhysRevMaterials.4.043802.
- [151] A.I. Kartamyshev, A.G. Lipnitskii, V.N. Saveliev, V.N. Maksimenko, I.V. Nelasov, and D.O. Poletaev. Development of an interatomic potential for titanium with high predictive accuracy of thermal properties up to melting point. *Computational Materials Science*, 160:30–41, 2019. ISSN 0927-0256. doi: <https://doi.org/10.1016/j.commatsci.2018.12.044>.
- [152] Zhi-Gang Mei, Shun-Li Shang, Yi Wang, and Zi-Kui Liu. Density-functional study of the thermodynamic properties and the pressure–temperature phase diagram of ti. *Phys. Rev. B*, 80:104116, Sep 2009. doi: 10.1103/PhysRevB.80.104116.
- [153] Cui-E Hu, Zhao-Yi Zeng, Lin Zhang, Xiang-Rong Chen, Ling-Cang Cai, and Dario Alfè. Theoretical investigation of the high pressure structure, lattice dynamics, phase transition, and thermal equation of state of titanium metal. *Journal of Applied Physics*, 107(9):093509, 2010. doi: 10.1063/1.3407560.
- [154] M. I. Mendeleev, T. L. Underwood, and G. J. Ackland. Development of an interatomic potential for the simulation of defects, plasticity, and phase transformations in titanium. *The Journal of Chemical Physics*, 145(15):154102, 2016. doi: 10.1063/1.4964654.
- [155] Alberto Ferrari, Malte Schröder, Yury Lysogorskiy, Jutta Rogal, Matous Mrovec, and Ralf Drautz. Phase transitions in titanium with an analytic bond-order potential. *Modelling and Simulation in Materials Science and Engineering*, 27(8):085008, oct 2019. doi: 10.1088/1361-651x/ab471d.
- [156] So Hirata and Matthew R. Hermes. Normal-ordered second-quantized hamiltonian for molecular vibrations. *The Journal of Chemical Physics*, 141(18):184111, 2014. doi: 10.1063/1.4901061.

- [157] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge University Press, 2018.
- [158] Alexander Wurm and Marcus Berg. Wick calculus. *American Journal of Physics*, 76(1):65–72, 2008. doi: 10.1119/1.2805232.

Appendix A

Harmonic eigenstates

Under the harmonic approximation, the Hamiltonian separates into a sum of single- and double-mode harmonic oscillators¹, as

$$\mathcal{H} = \sum_{\substack{\chi \\ \bar{\chi}=\chi}} \mathcal{H}_{\chi}(u_{\chi}) + \sum_{\substack{\chi \\ \bar{\chi}\neq\chi}} \mathcal{H}_{\chi}(u_{\chi}, u_{\bar{\chi}}). \quad (\text{A.1})$$

The single-mode oscillators correspond to modes which are their own pairs, with $\chi = \bar{\chi}$, and the double-mode oscillators correspond to modes which are not their own pairs, with $\chi \neq \bar{\chi}$.

The eigenstates of this Hamiltonian are product states of the eigenstates of the single- and double-mode oscillators, which can in turn be calculated analytically and have well-defined functional forms [42].

When working with these eigenstates, it is necessary to calculate many integrals of the form $\langle \psi_1 | X | \psi_2 \rangle$, where ψ_1 and ψ_2 are eigenstates and X is an operator.

There are several ways of calculating these integrals. The simplest, but most time consuming, is numerical integration. Alternatively, the operator X can first be converted into a function purely of harmonic creation and annihilation operators [156]. Another method is to use the Hermite recurrence relations to calculate each integral in terms of a sum of integrals of states with lower occupations [107].

All of these methods are more computationally intensive than is necessary. It is preferable to instead have analytic expressions for the results of these integrals, so that integration can be performed by simply consulting a lookup table.

This appendix presents the derivations and results of various analytic single- and double-mode integrals, including the expectation values of various operators between two harmonic eigenstates, and the thermal expectation values of

¹By changing co-ordinates to real periodic co-ordinates it is possible to separate the harmonic Hamiltonian further into a sum of purely single-mode oscillators. However, the eigenstates of these oscillators do not have well-defined wavevectors, and it is more straightforward to calculate the expectation values of operators defined in complex co-ordinates using states which are also defined in complex co-ordinates.

the same operators with respect to an entire basis of harmonic eigenstates.

A.1 Real modes

For modes which are their own pair, $\bar{\chi} = \chi$, the single-mode harmonic Hamiltonian is given by

$$\mathcal{H}(u) = -\frac{1}{2N} \frac{\partial^2}{\partial u^2} + \frac{N}{2} \omega^2 u^2. \quad (\text{A.2})$$

This is the quantum harmonic oscillator, a Hamiltonian whose solution exists in nearly every textbook on quantum mechanics [157]. For completeness, the derivation will be briefly presented here.

Introducing the creation and annihilation operators,

$$a^\dagger = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} u - \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial u} \right) \quad (\text{A.3})$$

and

$$a = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} u + \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial u} \right) \quad (\text{A.4})$$

respectively, the Hamiltonian can be written as

$$\mathcal{H} = \left(a^\dagger a + \frac{1}{2} \right) \omega. \quad (\text{A.5})$$

Then the harmonic states are defined as

$$|n\rangle = \frac{1}{\sqrt{n!}} a^{\dagger n} |0\rangle, \quad (\text{A.6})$$

where $|0\rangle$ is the harmonic ground state,

$$\langle u|0\rangle = \left(\frac{m\omega}{4} \right)^{\frac{1}{4}} e^{-\frac{1}{2}N\omega u^2}. \quad (\text{A.7})$$

These states satisfy the Schrödinger equation

$$\mathcal{H} |n\rangle = \left(n + \frac{1}{2} \right) \omega |n\rangle. \quad (\text{A.8})$$

A.1.1 Calculating $\langle p|u^n|q\rangle$

The state $u^n|q\rangle$ is a superposition of states $|q'\rangle$ satisfying $q-n \leq q' \leq q+n$, and for which $q+n-q'$ is even. As such, $\langle p|u^n|q\rangle$ is zero if $n+p+q$ is odd or if $n < |p-q|$.

If $n+p+q$ is even, $\langle p|u^n|q\rangle$ can be calculated by first writing u^n in terms of creation and annihilation operators, as

$$\langle p|u^n|q\rangle = \frac{1}{\sqrt{2N\omega}^n} \langle p|(a^\dagger + a)^n|q\rangle, \quad (\text{A.9})$$

and then using Wick calculus [158] to expand $(a^\dagger + a)^n$ as a series of normal-ordered operator products, to give

$$\langle p|u^n|q\rangle = \frac{1}{\sqrt{2N\omega^n}} \langle p| \left(\sum_{j=0}^{\frac{n}{2}} \frac{n!}{2^j j! (n-2j)!} : (a^\dagger + a)^{n-2j} : \right) |q\rangle. \quad (\text{A.10})$$

Taking a binomial expansion of $(a^\dagger + a)^{n-2j}$ gives

$$\langle p|u^n|q\rangle = \frac{1}{\sqrt{2N\omega^n}} \sum_{j=0}^{\frac{n}{2}} \sum_{k=0}^{n-2j} \frac{n!}{2^j j! k! (n-2j-k)!} \langle p|a^{\dagger k} a^{n-2j-k}|q\rangle. \quad (\text{A.11})$$

Removing one of the sums by noting that $\langle p|a^{\dagger k} a^{n-2j-k}|q\rangle$ is only non-zero if $p-k = q - (n-2j-k)$, introducing $l = \frac{n-|p-q|}{2} - j$, and simplifying a little, the final result is

$$\langle p|u^n|q\rangle = \frac{n! \sqrt{2}^{|p-q|}}{2^n \sqrt{N\omega^n} \frac{n+|p-q|}{2}!} \sqrt{\frac{\max(p,q)!}{\min(p,q)!}} \sum_{l=0}^{\min(p,q, \frac{n-|p-q|}{2})} 2^l \binom{\frac{n+|p-q|}{2}}{|p-q|+l} \binom{\min(p,q)}{l}. \quad (\text{A.12})$$

While this result is complicated, it can be expressed as $\langle p|u^n|q\rangle = \frac{1}{\sqrt{N\omega^n}} f(n,p,q)$, where $f(n,p,q)$ has no dependence on N or ω . As such, the values of $f(n,p,q)$ need only be calculated once for each value of n , p and q .

A.1.2 Calculating $\langle u^{2n} \rangle$ for a harmonic basis

Under the harmonic approximation, the thermal expectation of u^{2n} is given by

$$\langle u^{2n} \rangle = (1 - e^{-\beta\omega}) \sum_{p=0}^{\infty} \langle p|u^{2n}|p\rangle e^{-\beta p\omega}. \quad (\text{A.13})$$

The expression $\langle p|u^{2n}|p\rangle$ simplifies to give

$$\langle p|u^{2n}|p\rangle = \frac{(2n-1)!!}{(2N\omega)^n} \sum_{l=0}^{\min(n,p)} \frac{2^l}{l!} \binom{n}{l} \prod_{k=1}^l (p-l+k). \quad (\text{A.14})$$

Combining the two expressions, re-ordering the sums as $\sum_{p=0}^{\infty} \sum_{l=0}^{\min(n,p)} = \sum_{l=0}^n \sum_{p=l}^{\infty}$ and then making the transformation $p \rightarrow p+l$, this becomes

$$\langle u^{2n} \rangle = (1 - e^{-\beta\omega}) \frac{(2n-1)!!}{(2N\omega)^n} \sum_{l=0}^n \frac{2^l}{l!} \binom{n}{l} \sum_{p=0}^{\infty} \left(\prod_{k=1}^l (p+k) \right) e^{-\beta(p+l)\omega}. \quad (\text{A.15})$$

This can be simplified by replacing p with $-\frac{\partial}{\partial(\beta\omega)}$, moving $\prod_{k=1}^l (-\frac{\partial}{\partial(\beta\omega)} + k)$ through the sum across p , and taking

the sum of the exponential to give

$$\langle u^{2n} \rangle = (1 - e^{-\beta\omega}) \frac{(2n-1)!!}{(2N\omega)^n} \sum_{l=0}^n \frac{2^l}{l!} \binom{n}{l} e^{-\beta l\omega} \left(\prod_{k=1}^l \left(-\frac{\partial}{\partial(\beta\omega)} + k \right) \right) \frac{1}{1 - e^{-\beta\omega}}. \quad (\text{A.16})$$

Noting that $(-\frac{\partial}{\partial x} + l) \frac{1}{(1-e^{-x})^l} = \frac{l}{(1-e^{-x})^{l+1}}$, this becomes

$$\langle u^{2n} \rangle = \frac{(2n-1)!!}{(2N\omega)^n} \sum_{l=0}^n \binom{n}{l} \left(\frac{2e^{-\beta\omega}}{1 - e^{-\beta\omega}} \right)^l. \quad (\text{A.17})$$

Noting that $\sum_{l=0}^n \binom{n}{l} x^l = (1+x)^n$, this finally simplifies to

$$\langle u^{2n} \rangle = \frac{(2n-1)!!}{(2N\omega)^n} \left(\frac{1 + e^{-\beta\omega}}{1 - e^{-\beta\omega}} \right)^n. \quad (\text{A.18})$$

A.2 Complex modes

For modes which are not their own pair, $\bar{\chi} \neq \chi$, the double-mode harmonic Hamiltonian is given by

$$\mathcal{H}(u, \bar{u}) = -\frac{1}{N} \frac{\partial^2}{\partial u \partial \bar{u}} + N\omega^2 u \bar{u}, \quad (\text{A.19})$$

where it should be noted that the identical \mathcal{H}_χ and $\mathcal{H}_{\bar{\chi}}$ terms from equation A.1 have been added together.

It is necessary to introduce two creation operators,

$$a^\dagger = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} u - \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial \bar{u}} \right) \quad (\text{A.20})$$

and

$$\bar{a}^\dagger = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} \bar{u} - \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial u} \right), \quad (\text{A.21})$$

and two annihilation operators,

$$a = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} \bar{u} + \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial u} \right) \quad (\text{A.22})$$

and

$$\bar{a} = \frac{1}{\sqrt{2}} \left(\sqrt{N\omega} u + \frac{1}{\sqrt{N\omega}} \frac{\partial}{\partial \bar{u}} \right), \quad (\text{A.23})$$

and then the Hamiltonian can be written as

$$\mathcal{H} = \left(a^\dagger a + \bar{a}^\dagger \bar{a} + 1 \right) \omega. \quad (\text{A.24})$$

Then the harmonic states are defined as

$$|n, \bar{n}\rangle = \frac{1}{\sqrt{n! \bar{n}!}} a^\dagger{}^n \bar{a}^\dagger{}^{\bar{n}} |0, 0\rangle, \quad (\text{A.25})$$

where $|0, 0\rangle$ is the harmonic ground state,

$$\langle u, \bar{u} | 0, 0 \rangle = \sqrt{\frac{2m\omega}{\pi}} e^{-N\omega u \bar{u}}. \quad (\text{A.26})$$

These states satisfy the Schrödinger equation

$$\mathcal{H} |n, \bar{n}\rangle = (n + \bar{n} + 1) \omega |n, \bar{n}\rangle. \quad (\text{A.27})$$

It should be noted that while the Hamiltonian can be separated as $\mathcal{H} = (a^\dagger a + \frac{1}{2}) \omega + (\bar{a}^\dagger \bar{a} + \frac{1}{2}) \omega$, and the operators a^\dagger and a commute with \bar{a}^\dagger and \bar{a} , it is not possible to separate each eigenstate into a product of a u_χ state and a $u_{\bar{\chi}}$ state, and so the modes cannot be considered separately.

A.2.1 Calculating $\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle$

The state $u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle$ is a superposition of states $|q', \bar{q}'\rangle$, satisfying $q' - \bar{q}' = q - \bar{q} + n - \bar{n}$ and $q + \bar{q} - n - \bar{n} \leq q' + \bar{q}' \leq q + \bar{q} + n + \bar{n}$. As such, $\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle$ is zero if $p - \bar{p} \neq q - \bar{q} + n - \bar{n}$ or if $n + \bar{n} < |p + \bar{p} - q - \bar{q}|$.

If $\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle$ is non-zero then it can be evaluated by first taking binomial expansions of $u = \frac{a^\dagger + \bar{a}}{\sqrt{2}}$ and $\bar{u} = \frac{\bar{a}^\dagger + a}{\sqrt{2}}$, noting that $[a^\dagger, \bar{a}] = [\bar{a}^\dagger, a] = 0$ and so normal-ordering these expansions does not introduce commutators, to give

$$\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle = \frac{1}{\sqrt{2N\omega}^{n+\bar{n}}} \sum_{k=0}^{\min(n,p)} \sum_{\bar{k}=\max(0,\bar{n}-q)}^{\bar{n}} \binom{n}{k} \binom{\bar{n}}{\bar{k}} \sqrt{\frac{p!q!(\bar{q}+\bar{k})!(\bar{p}+n-k)!}{\bar{p}!\bar{q}!(p-k)!(q-\bar{n}+\bar{k})!}} \langle p-k, \bar{p}+n-k | q-\bar{n}+\bar{k}, \bar{q}+\bar{k} \rangle. \quad (\text{A.28})$$

Noting that $\langle p-k, \bar{p}+n-k | q-\bar{n}+\bar{k}, \bar{q}+\bar{k} \rangle$ is only non-zero for $p-k = q-\bar{n}+\bar{k}$, after a little simplification this becomes

$$\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle = \frac{1}{\sqrt{2N\omega}^{n+\bar{n}}} \sqrt{\frac{p!q!}{\bar{p}!\bar{q}!}} \sum_{k=\max(0,p-n,q-\bar{n})}^{\min(p,q)} \binom{n}{p-k} \binom{\bar{n}}{q-k} \frac{(n+\bar{p}-p+k)!}{k!}. \quad (\text{A.29})$$

This can again be written as $\langle p, \bar{p} | u^n \bar{u}^{\bar{n}} | q, \bar{q} \rangle = \frac{1}{\sqrt{N\omega}^{n+\bar{n}}} f(p, \bar{p}, n, \bar{n}, q, \bar{q})$, such that once the values of f have been precomputed the overlap integrals can be calculated quickly.

A.2.2 Calculating $\langle (u\bar{u})^n \rangle$ for a harmonic basis

Under the harmonic approximation, the thermal expectation of $(u\bar{u})^n$ is given by

$$\langle (u\bar{u})^n \rangle = (1 - e^{-\beta\omega})^2 \sum_{p=0}^{\infty} \sum_{\bar{p}=0}^{\infty} \langle p, \bar{p} | (u\bar{u})^n | p, \bar{p} \rangle e^{-\beta(p+\bar{p})\omega}. \quad (\text{A.30})$$

The expression for $\langle p, \bar{p} | (u\bar{u})^n | p, \bar{p} \rangle$ can be simplified by introducing $l = p - k$, to give

$$\langle p, \bar{p} | (u\bar{u})^n | p, \bar{p} \rangle = \frac{1}{(2N\omega)^n} \sum_{l=0}^{\min(n,p)} \binom{n}{l} \binom{n}{l} \left(\prod_{j=1}^l (p-l+j) \right) \left(\prod_{j=1}^{n-l} (\bar{p}+j) \right). \quad (\text{A.31})$$

Combining the two expressions, re-arranging the sums as $\sum_{p=0}^{\infty} \sum_{l=0}^{\min(n,p)} = \sum_{l=0}^n \sum_{p=l}^{\infty}$, and making the transformation $p \rightarrow p+l$ this becomes

$$\langle (u\bar{u})^n \rangle = \frac{(1 - e^{-\beta\omega})^2}{(2N\omega)^n} \sum_{l=0}^n \binom{n}{l} \binom{n}{l} e^{-\beta l\omega} \left(\sum_{p=0}^{\infty} \left(\prod_{j=1}^l (p+j) \right) e^{-\beta p\omega} \right) \left(\sum_{\bar{p}=0}^{\infty} \left(\prod_{j=1}^{n-l} (\bar{p}+j) \right) e^{-\beta \bar{p}\omega} \right). \quad (\text{A.32})$$

Using the same identities as for the $\bar{\chi} = \chi$ case, this becomes

$$\langle (u\bar{u})^n \rangle = \frac{(1 - e^{-\beta\omega})^2}{(2N\omega)^n} \sum_{l=0}^n \binom{n}{l} \binom{n}{l} e^{-\beta l\omega} \frac{l!}{(1 - e^{-\beta\omega})^{l+1}} \frac{(n-l)!}{(1 - e^{-\beta\omega})^{n-l+1}}, \quad (\text{A.33})$$

which in turn simplifies to

$$\langle (u\bar{u})^n \rangle = \frac{n!}{(2N\omega)^n} \left(\frac{1 + e^{-\beta\omega}}{1 - e^{-\beta\omega}} \right)^n. \quad (\text{A.34})$$

Appendix B

Thermodynamic addition and subtraction

B.1 Thermal addition

Consider two bases of states, 1 and 2, taken to be non-overlapping such that every state in 1 is orthogonal to every state in 2. In isolation, both bases independently diagonalise the projection of the Hamiltonian in that basis. The i 'th basis will have internal energy U_i , entropy S_i and free energy F_i . The union of basis 1 and basis 2 is denoted $1 + 2$.

If these bases are brought into thermal equilibrium with one another, there will in general be mixing between the two bases. This is impossible to calculate without knowledge of the individual states. Instead, the equilibrium properties of basis $1 + 2$ can be approximated by bringing basis 1 and basis 2 into thermal equilibrium under the constraint that their states cannot mix.

Under this approximation, the density matrix of the $1 + 2$ basis, ρ_{1+2} , is simply a weighted sum of the density matrices of the 1 and 2 bases, $\rho_{1+2} = P_1\rho_1 + P_2\rho_2$, where the weights P_i must satisfy $P_1 + P_2 = 1$. The weights P_i are those which minimise the free energy of the $1 + 2$ basis.

In this case, the internal energy, entropy and free energy of the $1 + 2$ basis are

$$\begin{aligned} U_{1+2} &= P_1U_1 + P_2U_2 \\ S_{1+2} &= P_1S_1 + P_2S_2 - P_1\ln(P_1) - P_2\ln(P_2) \\ F_{1+2} &= P_1F_1 + P_2F_2 + TP_1\ln(P_1) + TP_2\ln(P_2). \end{aligned} \tag{B.1}$$

Minimising the free energy with respect to $\{P_i\}$ gives

$$P_i = \frac{e^{-F_i/T}}{e^{-F_1/T} + e^{-F_2/T}}. \quad (\text{B.2})$$

It is convenient to define the partition function, $Z_i = e^{-F_i/T}$. With this definition, the thermodynamic potentials of the 1 + 2 basis are

$$\begin{aligned} Z_{1+2} &= Z_1 + Z_2 \\ U_{1+2} &= \frac{Z_1 U_1 + Z_2 U_2}{Z_{1+2}} \\ S_{1+2} &= \frac{Z_1 U_1 + Z_2 U_2}{Z_{1+2} T} + \ln(Z_{1+2}) \\ F_{1+2} &= -T \ln(Z_{1+2}). \end{aligned} \quad (\text{B.3})$$

B.2 Thermal subtraction

The process of thermal addition can be reversed in order to remove one system from another. If basis 1 is entirely contained within basis 2, i.e. the states in 2 form a complete basis for every state in 1, then the properties of the complement of basis 1 in basis 2, denoted basis 2 - 1, can be calculated as

$$\begin{aligned} Z_{2-1} &= Z_2 - Z_1 \\ U_{2-1} &= \frac{Z_2 U_2 - Z_1 U_1}{Z_{2-1}} \\ S_{2-1} &= \frac{Z_2 U_2 - Z_1 U_1}{Z_{2-1} T} + \ln(Z_{2-1}) \\ F_{2-1} &= -T \ln(Z_{2-1}). \end{aligned} \quad (\text{B.4})$$

Appendix C

Copyright permissions

Figure 7.8 of this work is reproduced from figure 5 of Hennig et al. [148]. Reprinted figure with permission from R. G. Hennig, T. J. Lenosky, D. R. Trinkle, S. P. Rudin and J. W. Wilkins, *Physical Review B*, 78, 054121, (2008) [148]. Copyright (2008) by the American Physical Society.

The license from the American Physical Society to reproduce this figure follows, and written permission has been given by R. G. Hennig.



American Physical Society Reuse and Permissions License

14-Jun-2021

This license agreement between the American Physical Society ("APS") and Mark Johnson ("You") consists of your license details and the terms and conditions provided by the American Physical Society and SciPris.

Licensed Content Information

License Number: RNP/21/JUN/041119
License date: 14-Jun-2021
DOI: 10.1103/PhysRevB.78.054121
Title: Classical potential describes martensitic phase transformations between the α , β , and δ titanium phases
Author: R. G. Hennig et al.
Publication: Physical Review B
Publisher: American Physical Society
Cost: USD \$ 0.00

Request Details

Does your reuse require significant modifications: No
Specify intended distribution locations: UK & Commonwealth (excluding Canada)
Reuse Category: Reuse in a thesis/dissertation
Requestor Type: Student
Items for Reuse: Figures/Tables
Number of Figure/Tables: 1
Figure/Tables Details: Figure 5. Equilibrium phase diagram of Ti.
Format for Reuse: Electronic and Print
Total number of print copies: Up to 1000

Information about New Publication:

University/Publisher: University of Cambridge
Title of dissertation/thesis: Accelerating vibrational free energy calculations for anharmonic crystals
Author(s): Mark Johnson
Expected completion date: Jul. 2021

License Requestor Information

Name: Mark Johnson
Affiliation: Individual
Email Id: mj464@cam.ac.uk
Country: United Kingdom

TERMS AND CONDITIONS

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to Electronic and Print format, provided all criteria outlined below are followed.

1. You must also obtain permission from at least one of the lead authors for each separate work, if you haven't done so already. The author's name and affiliation can be found on the first page of the published Article.
2. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, <http://dx.doi.org/{DOI}>. The hyperlink may be embedded in the copyright credit line.
3. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."
4. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.
5. Use of the material does not and must not imply any endorsement by APS.
6. APS does not imply, purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestor's sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate it was reprinted by APS with permission from another source.
7. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.
8. You may not alter, edit or modify the material in any manner.
9. You may translate the materials only when translation rights have been granted.
10. APS is not responsible for any errors or omissions due to translation.
11. You may not use the material for promotional, sales, advertising or marketing purposes.
12. The foregoing license shall not take effect unless and until APS or its agent, Aptara, receives payment in full in accordance with Aptara Billing and Payment Terms and Conditions, which are incorporated herein by reference.
13. Should the terms of this license be violated at any time, APS or Aptara may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.
14. APS reserves all rights not specifically granted herein.
15. This document, including the Aptara Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.