FIRST-PRINCIPLES CALCULATION OF DYNAMICAL PROPERTIES OF INSULATORS IN FINITE ELECTRIC FIELDS AND ANOMALOUS HALL CONDUCTIVITY OF FERROMAGNETS BASED ON BERRY PHASE APPROACH

BY XINJIE WANG

A Dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

Written under the direction of

Professor David Vanderbilt

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

First-principles calculation of dynamical properties of insulators in finite electric fields and anomalous Hall conductivity of ferromagnets based on Berry phase approach

by Xinjie Wang

Dissertation Director: Professor David Vanderbilt

We present first-principles methods for calculating two distinct types of physical quantities within the framework of density functional theory: the response properties of an insulator to finite electric fields, and the anomalous Hall conductivity of a ferromagnet. Both of the methods are closely related to the same ingredient, namely the Berry phase, a geometric phase acquired by a quantum system transporting in parameter space. We develop gauge-invariant formulations in which the random phases of Bloch functions produced by numerical subroutines are irrelevant.

First, we provide linear-response methods for calculating phonon frequencies, Born effective charge tensors and dielectric tensors for insulators in the presence of a finite electric field. The starting point is a variational total-energy functional with a field-coupling term that represents the effect of the electric field. This total-energy functional is expanded with respect to both small atomic displacements and electric fields within the framework of density-functional perturbation theory. The linear responses of field-polarized Bloch functions to atomic displacements and electric fields are obtained by minimizing the second-order derivatives of the total-energy functional. The

desired second-order tensors are then constructed from these optimized first-order field-polarized Bloch functions.

Next, an efficient first-principles approach for computing the anomalous Hall conductivity is described. The intrinsic anomalous Hall conductivity in ferromagnets depends on subtle spin-orbit-induced effects in the electronic structure, and recent *ab-initio* studies found that it was necessary to sample the Brillouin zone at millions of k-points to converge the calculation. We start out by performing a conventional electronic-structure calculation including spin-orbit coupling on a uniform and relatively coarse k-point mesh. From the resulting Bloch states, maximally localized Wannier functions are constructed which reproduce the *ab-initio* states up to the Fermi level. With inexpensive Fourier and unitary transformations the quantities of interest are interpolated onto a dense k-point mesh and used to evaluate the anomalous Hall conductivity as a Brillouin-zone integral. The present scheme, which also avoids the cumbersome summation over all unoccupied states in the Kubo formula, is applied to bcc Fe, giving excellent agreement with conventional, less efficient first-principles calculations.

Finally, we consider another *ab-initio* approach for computing the anomalous Hall conductivity based on Haldane's Fermi-surface formulation. Working in the Wannier representation, the Brillouin zone is sampled on a large number of equally spaced parallel slices oriented normal to the total magnetization. On each slice, we find the intersections of the Fermi surface sheets with the slice by standard contour methods, organize these into a set of closed loops, and compute the Berry phase of the Bloch states as they are transported around these loops. The anomalous Hall conductivity is proportional to the sum of the Berry phases of all the loops on all the slices.

Acknowledgements

I would like to express deep gratitudes to my Ph.D. supervisor Prof. David Vanderbilt for his careful guidance, great patience, wonderful inspiration and insightful thought. Working with David lets me not only learn a great deal of physics but also gain valuable research experience.

I would also like to thank Prof. Ivo Souza and Dr. Jonathan Yates for their great help on the project of anomalous Hall effect. It is very pleasant to work with them.

I would like to thank Dr. Don Hamann for his kind help on the ABINIT project and many others.

I would also like to thank the other members of the group: Prof. Karin Rabe, Prof. David Langreth, Dr. Oswaldo Diéguez, Dr. Ping Sun, Dr. Timo Thonhauser, Dr. Valentino Cooper, Dr. Shen Li, Dr. Nishimatsu Takeshi, Dr. Viktor Oudovenko and Andrei Malashevich.

Thanks to my family who have been unconditionally understanding and supportive of my studies.

I would like to thank my dear fiancee Junyi who has given me great support during the past four years.

Finally thanks to my friends Chen Xun, Qu Qiang, Wei Zhou, Xifan Wu, Chenglin Zhang, Weihua Zheng, Tian Feng, Sung Po Chao, and Marianna Maltseva.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Response properties of an insulator

The response properties of an insulator to atomic displacements, homogeneous electric fields, or both of these perturbations include the phonon force-constant matrix, the dielectric tensor, and the Born effective charge tensor. These tensors, defined as second derivatives of the total energy, are among the most extensively studied properties of an insulator. Phonon properties are very helpful to explain the occurrence of peaks in Raman and infrared spectroscopy [1], and to identify structure instability and the resulting phase transition [2]. Dielectric properties are of fundamental and practical interest. In theory, many models have been developed to calculate dielectric constants of various materials, and in practice, high-κ dielectric materials have critical applications in next-generation high-performance integrated circuits. The Born effective charge is important in determining the response to an electric field. It provides a transparent picture for understanding and modeling how atoms respond to an applied electric field.

The early theoretical models for calculating phonon properties involved development of empirical interatomic potentials and the parameter fitting to experimental phonon spectra. Valence-force-field models [3, 4], bond charge models [5] and shell models [6] are examples. As the result of the complexity of atomic interactions, these models usually have complicated parameter dependence, low accuracy and poor transferability.

The Born effective charge and the dielectric constant are two elementary responses of an insulator to electric fields. The electrostatic potential of a homogeneous electric field is linear in space

and unbounded from below, which breaks the translational symmetry. Previous methods circumvent this problem by treating electric fields as perturbations and performing calculations in the slowly-varying limit of periodically-varying fields [7]. The main drawback of these methods is that they only work when the reference unperturbed system has a vanishing electric field.

Density-functional theory provides a powerful framework for performing calculations on realistic materials. Within this framework, density-functional perturbation theory has been developed by several authors [8, 9, 10] to facilitate response calculations. The response properties can now be calculated very conveniently at high accuracy without using empirical parameters.

Recently, the effect of finite electric fields attracted both theoretical and practical attention. The tuning of the response properties of an insulator by applying finite electric fields is of importance in a variety of practical applications in electronics, such as for capacitors and oscillators. It is highly desirable to develop a theoretical approach for calculating response properties of a periodic system in the presence of finite electric fields.

1.2 Anomalous Hall effect

In 1880, Edwin R. Hall discovered the ordinary Hall effect, and just one year later he discovered anomalous Hall effect [11, 12]. In non-ferromagnets, it is well-known that the flow of electrons in a perpendicular magnetic field is affected by the Lorentz force. However, in ferromagnets such as Fe, Ni and Co, a Hall current is still present even in the absence of a magnetic field. Since the magnetic field is not necessary, this effect is called the anomalous Hall effect (AHE).

In addition to simple ferromagnetic metals, many other types materials of different crystal structures have been found to exhibit AHE, such as $SrRuO_3$ (perovskite structure) [13], $CuCr_2Se_4$ (spinel structure) [14], diluted magnetic semiconductors [15], and so on. The AHE has become a useful tool to characterize and investigate magnetic properties of nano-scale systems. Its fundamental physics

as well as potential technical applications in memory devices and sensors motivate intensive theoretical and experimental studies.

Unlike the ordinary Hall effect, the AHE is strongly related to the spin-orbit coupling of electrons. It is not produced by the internal field associated with magnetization. Historically, the mechanism of the AHE has been controversial for many years. There are two distinct approaches to explain the AHE. Karplus and Luttinger first showed that it can arise in a perfect crystal as a result of the spin-orbit interaction of polarized conduction electrons [16]. Later, two alternative mechanisms, skew scattering [17] and side-jump scattering [18], were proposed by Smit and Berger, respectively. In skew scattering, the spin-orbit interaction gives rise to an asymmetric scattering cross section even if the defect potential is symmetric, and in side-jump scattering it causes the scattered electron to acquire an extra transverse translation after the scattering event. These two mechanisms involve scattering from impurities or phonons, while the Karplus-Luttinger contribution is a scattering-free band-structure effect.

In recent years, new insights into the Karplus-Luttinger contribution have been obtained by several authors, who reexamined it in the modern language of Berry phases. The term $\Omega_n(\mathbf{k})$ in the equations below was recognized as the Berry curvature of the Bloch states in reciprocal space, a quantity which had previously appeared in the theory of the integer quantum Hall effect, and which is also closely related to the Berry-phase theory of polarization.

1.3 Berry phase

In 1984, M.V. Berry introduced a geometric phase $e^{i\gamma(C)}$ acquired by a quantum system in an eigenstate adiabatically transporting along a closed path C in parameter space [19]. Here, the parameter space is simply a set of values on which Hamiltonian depends. Consider the Hamiltonian

$H(\mathbf{R})$ as a function of parameters $\mathbf{R} = \{a_1, a_2, ...\}$. The Berry phase is defined as

$$\gamma_n(C) \equiv i \oint_C \langle n(\mathbf{R})|\nabla_{\mathbf{R}} n(\mathbf{R})\rangle \cdot d\mathbf{R} \ , \tag{1.1}$$

where $\nabla_{\mathbf{R}}$ means that the derivative is with respect to $\mathbf{R}$, and $|n(\mathbf{R})\rangle$ is the eigenstate of $H(\mathbf{R})$ satisfying

$$H(\mathbf{R})|n(\mathbf{R})\rangle = E_n|n(\mathbf{R})\rangle \ .$$

It is worth to note that while the Berry phase $\gamma_n(C)$ itself is gauge-invariant (a change of wave functions of the form $|n(\mathbf{R})\rangle \to e^{i\theta(\mathbf{R})}|n(\mathbf{R})\rangle$ is called a gauge transformation), the inner product $\langle n(\mathbf{R})|\nabla_{\mathbf{R}} n(\mathbf{R})\rangle$ is *not* gauge-invariant, i.e., adding $e^{i\theta(\mathbf{R})}$ to $|n(\mathbf{R})\rangle$

$$\langle n(\mathbf{R})e^{-i\theta(\mathbf{R})}|\nabla_{\mathbf{R}}|e^{i\theta(\mathbf{R})}n(\mathbf{R})\rangle = \langle n(\mathbf{R})|\nabla_{\mathbf{R}} n(\mathbf{R})\rangle + i\nabla_{\mathbf{R}}\theta(\mathbf{R}) \ .$$

In general, $\nabla_{\mathbf{R}}\theta(\mathbf{R})$ is not zero. However, its integral is zero, that is

$$\oint_C \nabla_{\mathbf{R}}\theta(\mathbf{R}) d\mathbf{R} = 0 \ ,$$

which clearly shows the $\gamma_n(C)$ gauge-invariant.

If C lies close to a degeneracy of $H(\mathbf{R})$, the Berry phase takes a particularly simple and meaningful form as described below. Consider a degeneracy that involves only two states. The Hamiltonian near this degeneracy can generally be described by a 2×2 Hermitian matrix

$$H = (X \cdot \sigma_x + Y \cdot \sigma_y + Z \cdot \sigma_z) + K \cdot \sigma_0$$

where $\sigma_{x,y,z}$ are Pauli matrices, σ_0 is the unit matrix, and X, Y, Z, K are real coefficients. If X, Y, Z are viewed as coordinates in a 3-dimensional space $\mathbf{F}$, the Berry phase is given by

$$\gamma_{\pm}(C) = \mp\frac{1}{2}\Omega(C') \ ,$$

where C' in F-space is the image of C in $\mathbf{R}$ and $\Omega(C')$ is the solid angle subtended by C' in $\mathbf{F}$-space.

In our work, the parameters $\mathbf{R}$ are the three Cartesian components of the reciprocal vector $\mathbf{k}$, and $|n(\mathbf{R})\rangle$ is the periodic part of Bloch function $|u_{\mathbf{k}n}\rangle$. For practical calculations, one normally works on a discretized k-mesh. The integration in Eq. (1.1) can then be performed over the k-mesh [20], using

$$\gamma_n(C) = \mathrm{Im}\ln\prod_{j=0}^{J-1}\langle u_{\mathbf{k}_j n}|u_{\mathbf{k}_{j+1} n}\rangle \tag{1.2}$$

where Im means imaginary part and j indicates the discretized consecutive k points on the closed path C. It is straightforward to verify that when $J \to \infty$, Eq. (1.2) goes back to its continuous form in Eq. (1.1). It is also easy to confirm that Eq. (1.2) is independent of the arbitrary phase $e^{i\theta(\mathbf{k})}$: both $\langle u_{\mathbf{k}_j n}|$ and $|u_{\mathbf{k}_j n}\rangle$ appear in Eq. (1.2), so the arbitrary phase cancels out.

In the past twenty years, Berry phases have been found to manifest themselves in many solid-state physics observables such as the Aharonov-Bohm effect [19], electronic polarization [20], anomalous Hall conductivity [21] and quantum Hall effect [22]. Our work is related to the electronic polarization and anomalous Hall conductivity.

1.4 Outline of the present work

The present work is devoted to developing first-principles methods for calculating response properties of an insulator to external perturbations in finite electric fields, and efficient *ab-initio* methods for the computing anomalous Hall effect of ferromagnets.

Chapter 2 describes the theoretical basis for our study, including density-functional theory and density-functional perturbation theory. In addition to the fundamental theories, approximations that are necessary to carry out practical applications and some important numerical algorithms are also discussed in this chapter.

In Chapter 3, we present a perturbative method for calculating response properties of an insulator in the presence of a finite electric field. We expand a variational total-energy functional with

a field-coupling term that represents the effect of the electric field with respect to small atomic displacements and electric fields within the framework of density-functional perturbation theory. The first-order response of field-polarized Bloch functions is obtained by minimizing the second-order derivatives of the total-energy functional. We implement the method in the `ABINIT` code and perform illustrative calculations of the field-dependent phonon frequencies for III-V semiconductors.

In Chapter 4, we report an efficient first-principles approach for computing the anomalous Hall conductivity. This approach has three steps: (1) a conventional electronic-structure calculation is performed including spin-orbit coupling on a uniform and relatively coarse k-point mesh; (2) from the resulting Bloch states, maximally-localized Wannier functions are constructed which reproduce the *ab-initio* states up to the Fermi level; (3) the quantities of interest are interpolated onto a dense k-point mesh and used to evaluate the anomalous Hall conductivity as a Brillouin-zone integral. The present scheme, which also avoids the cumbersome summation over all unoccupied states in the Kubo formula, is applied to bcc Fe, giving excellent agreement with conventional, less efficient first-principles calculations. Remarkably, we find that about 99% of the effect can be recovered by keeping a set of terms depending only on the Hamiltonian matrix elements, not on matrix elements of the position operator.

In Chapter 5, we present an *ab-initio* approach for computing the anomalous Hall conductivity by converting the integral over the Fermi sea into a more efficient integral on the Fermi surface only. First, a conventional electronic-structure calculation is performed with spin-orbit interaction included. Maximally-localized Wannier functions are then constructed by a post-processing step in order to convert the *ab-initio* electronic structure around the Fermi level into a tight-binding-like form. Working in the Wannier representation, the Brillouin zone is sampled on a large number of equally spaced parallel slices oriented normal to the total magnetization. On each slice, we find the intersections of the Fermi-surface sheets with the slice by standard contour methods, organize these

into a set of closed loops, and compute the Berry phases of the Bloch states as they are transported around these loops. The anomalous Hall conductivity is proportional to the sum of the Berry phases of all the loops on all the slices. Illustrative calculations are performed for Fe, Co and Ni.

Chapter 2

First-principles methods

In this chapter, the fundamental theory of our first-principles tools for calculating electronic structures of a system is introduced. Since its birth [23, 24], density-functional theory (DFT) has became a most powerful, widely-used theoretical tool for investigating properties of various materials. The basic idea of DFT is that it exactly maps an electron many-body electron problem onto an equivalent non-interacting single-electron problem, keeping the complexity of the problem in an unknown functional. Mathematically, DFT exactly transforms the many-body Schroedinger equation of $3N$ degrees of freedom into an equivalent set of Schroedinger-like equations of 3 degrees of freedom.

2.1 Density-functional theory

2.1.1 Hohenberg-Kohn theorem

There had been many attempts to express the total energy of systems in terms of the electron charge density. The real breakthrough came in 1964. In that year, Hohenberg and Kohn proposed an exact formal variational principle of the ground state of an inhomogeneous interacting electron gas in an external potential. The charge density $n(\mathbf{r})$ is the basic variable. They proved the existence of a universal functional of density $F[n(\mathbf{r})]$ which is independent of the external potential $v(\mathbf{r})$, such that the minimum value of the total energy $E \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})]$ is equal to the ground-state energy of the system.

For the non-degenerate ground state, $v(\mathbf{r})$ is a unique functional of $n(\mathbf{r})$ up to a constant potential. Since $v(\mathbf{r})$ fixes the Hamiltonian, the full many-body ground state is a unique functional of the

charge density $n(\mathbf{r})$.

If the form of the universal functional $F[n(\mathbf{r})]$ were known, the true ground-state energy can be obtained by minimizing the total energy functional

$$E \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})] \tag{2.1}$$

with respect to a three-dimensional density function $n(\mathbf{r})$.

2.1.2 Kohn-Sham equation

The exact form of the functional $F[n(\mathbf{r})]$ is unknown. Approximation method have been developed. In 1965, Kohn and Sham developed a set of self-consistent equations including correlation effects.

The Kohn-Sham equation can be derived by applying the variational principle to the Kohn-Sham energy functional

$$E \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})] \ . \tag{2.2}$$

It is obvious that the universal functional $F[n(\mathbf{r})]$ includes the classical Coulomb energy

$$F[n(\mathbf{r})] = \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + G[n(\mathbf{r})] \ , \tag{2.3}$$

where $G[n(\mathbf{r})]$ is a universal functional consisting of the kinetic energy of non-interacting electrons and the exchange-correlation energy

$$G[n(\mathbf{r})] = T[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] \ . \tag{2.4}$$

Under the constraints of $\int \delta n(\mathbf{r})d\mathbf{r} = 0$, the stationary property is

$$\int \delta n(\mathbf{r}) \left\{ v + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \frac{\delta T[n(\mathbf{r})]}{\delta n(\mathbf{r})} + \frac{E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \right\} d\mathbf{r} = 0 \ . \tag{2.5}$$

This is equivalent to solving the following Schrödinger-like partial differential equation,

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V_{eff}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \ , \tag{2.6}$$

where m is the mass of electron, $v_{eff}(\mathbf{r})$ is an effective and *local* potential, $\psi_i(\mathbf{r})$ is the ith Kohn-Sham orbital, and ϵ_i is the eigenvalue of ith Kohn-Sham orbital. The effective potential $V_{eff}(\mathbf{r})$, the most important component in the KS equation, is defined as follows:

$$v_{eff}(\mathbf{r}) \equiv v(\mathbf{r}) + v_{\mathrm{H}}(\mathbf{r}) + v_{xc}(\mathbf{r}) \tag{2.7}$$

$$v(\mathbf{r}) \equiv -\sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{2.8}$$

$$v_{\mathrm{H}}(\mathbf{r}) \equiv \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2.9}$$

$$v_{xc}(\mathbf{r}) \equiv \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \tag{2.10}$$

$$n(\mathbf{r}) \equiv \sum_i |\psi_i(\mathbf{r})|^2 \tag{2.11}$$

In an intuitive picture, the KS equation describes an electron moving in an effective potential produced by other electrons. It is worth noting that the density contains the contribution from the electron itself. Thus the electron seems to interact with itself as shown in $v_H(\mathbf{r})$ and $v_{xc}(\mathbf{r})$ terms. The point is that if the exchange energy is treated exactly (as in the Hartree-Fock equation), the self-interaction is canceled out in the Hartree term and exchange term. Unfortunately, in the most popular local-density approximation (LDA), the cancellation is not exact. The proper treatment of the self-interaction correction is an important but difficult challenge that is the subject of current research. In this thesis, we shall restrict ourselves to working with the LDA only.

2.1.3 Approximations for exchange-correlation energy

In Hohenberg and Kohn's paper, they considered the case of a slowly varying charge density. If $n(\mathbf{r})$ is sufficiently slowly varying, the exchange-correlation energy functional can be approximated by

$$E_{xc}[n(\mathbf{r})] = \int n(\mathbf{r}) \epsilon_{xc}^{\mathrm{unif}}(n) d\mathbf{r} \ , \tag{2.12}$$

where $\epsilon_{xc}(n)$ is the exchange-correlation energy per electron of a uniform electron gas of density $n(\mathbf{r})$. This approximation is called the local-density approximation (LDA). It becomes exact in two limiting cases: the slowly varying case $r_s/r_0 << 1$, or the high-density case $r_s/a_0 << 1$, where $r_s = (3/4\pi n)^{1/3}$ is the Wigner-Seitz radius and r_0 is a typical length over which charge density changes considerably. Despite the fact that neither condition is well satisfied for most real materials, this simple approximation turned out to be very successful in practical calculation of metals, alloys and semiconductors.

In the local spin-density approximation (LSDA), the spin-orbit coupling effect is ignored and thus the spin degrees of freedoms are decoupled from the spatial coordinates, that is the (2×2) Hamiltonian matrix is diagonal in spin space (the spin-up component of the spinor wavefunction is decoupled from the spin-down component). The Kohn-Sham energy functional is then written as

$$E_{\rm ex}^{\rm LSDA}[n_\uparrow, n_\downarrow] = \int d\mathbf{r} n \epsilon_{xc}^{\rm unif}(n_\uparrow, n_\downarrow) \; , \tag{2.13}$$

where

$$n(\mathbf{r}) = n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r}) \; , \tag{2.14}$$

$$n_\uparrow(\mathbf{r}) = \sum_i |\psi_i^\uparrow(\mathbf{r})|^2 \; , \tag{2.15}$$

$$n_\downarrow(\mathbf{r}) = \sum_i |\psi_i^\downarrow(\mathbf{r})|^2 \; . \tag{2.16}$$

In comparison with the LDA, the generalized gradient approximation (GGA) [25, 26, 27] makes use of both charge density n and its gradient ∇n, improving total energies, atomization energies and structural energy differences. The GGA exchange-correlation energy functional takes the form

$$E_{\rm xc}^{\rm GGA}[n_\uparrow, n_\downarrow] = \int d\mathbf{r} f(n_\uparrow, n_\downarrow, \nabla n_\uparrow, \nabla n_\downarrow) \; . \tag{2.17}$$

In our study of bcc Fe, the GGA gives the correct ground-state structure whereas the LDA gives the wrong result.

The Kohn-Sham method can easily be extended to treat spin-orbit coupling. The treatment of spin degrees of freedom requires spinor wavefunctions. The DFT is then expressed in terms of a 2×2 density matrix

$$n(\mathbf{r}) = \begin{pmatrix} n_{11}(\mathbf{r}) & n_{12}(\mathbf{r}) \\ n_{21}(\mathbf{r}) & n_{22}(\mathbf{r}) \end{pmatrix}$$

The electron density is given by the trace of $n(\mathbf{r})$. The exchange-correlation energy becomes [28],

$$E_{xc}^{LDA}[\bar{n}(\mathbf{r})] = \int \bar{n}(\mathbf{r}) \cdot \epsilon_{xc}[\bar{n}(\mathbf{r}), m(\mathbf{r})] d\mathbf{r}$$

where $\epsilon_{xc}[\bar{n}(\mathbf{r}), m(\mathbf{r})]$ is the exchanged-correlation energy functional of spin-polarized homogeneous electron gas with charge density $\bar{n}(\mathbf{r}) = n_{11}(\mathbf{r}) + n_{22}(\mathbf{r})$ and magnetization density $m(\mathbf{r}) = |n_+(\mathbf{r}) - n_-(\mathbf{r})|$. Here, $n_+(\mathbf{r})$ and $n_-(\mathbf{r})$ are the eigenvalues of the matrix $n(\mathbf{r})$.

2.1.4 Plane-wave pseudopotential approach

To facilitate solving the Kohn-Sham equations, good basis functions must be used. The most commonly used basis functions are plane waves. The Bloch theorem states that the Bloch function of band n at $\mathbf{k}$ takes the form

$$\psi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \ . \tag{2.18}$$

In the plane-wave basis, the cell-periodic part $u_{n\mathbf{k}}$ can be expanded as

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \ , \tag{2.19}$$

where $\mathbf{G}$ runs over reciprocal lattice vectors. The summation over $\mathbf{G}$ is infinite, but in practice it is truncated to include only plane waves with kinetic energies smaller than some cutoff energy, i.e., $\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2 < E_{\text{cut}}$. It should be kept in mind that the truncated set of $\{\mathbf{G}\}$'s varies with wave vector $\mathbf{k}$. For example, the inner product between two wave functions $\langle u_{\mathbf{k}} | u_{\mathbf{k}+\delta\mathbf{k}} \rangle$ is computed on the common set of $\{\mathbf{G}\}$'s of $\mathbf{k}$ and $\mathbf{k}+\delta\mathbf{k}$.

As a result of its delocalized nature, the plane-wave basis is not efficient for expanding localized core states. In order to use a considerably smaller number of plane waves, the pseudopotential approximation has been intensively developed by many authors [29, 30, 31, 32]. The basic idea of the pseudopotential approximation is to remove the core electrons from the Kohn-Sham orbitals and replace them and the bare ionic potential by a smoother pseudopotential which acts on the corresponding pseudo wavefunctions. In practice, the pseudopotential has two parts, the long-range local part and a short-range nonlocal part, i.e.,

$$v_{\rm pp} = v_{\rm local}(\mathbf{r}) + \sum_{lm} |lm\rangle v_{{\rm nonlocal},l}(\mathbf{r}) \langle lm| \tag{2.20}$$

where the $|lm\rangle$ are spherical harmonics and $v_{{\rm nonlocal},l}(\mathbf{r})$ is the nonlocal pseudopotential for angular momentum l.

The pseudopotential approach has been extended to include spin-orbit coupling by taking the solutions of a fully relativistic radial atomic Dirac-like equation as the reference all-electron wave functions. The nonlocal part of such a pseudopotential is a 2×2 matrix of operators acting on spinor wave functions. In a simplified form, it looks like

$$v^{\alpha\beta}_{\rm nonlocal} = \sum_n f_n |Y^\alpha_n\rangle \langle Y^\beta_n| \ ,$$

where Y^α_n are projector functions and α, β are spin indices.

2.1.5 Solving the Kohn-Sham equation by minimizing the total energy

Since $v_{eff}(\mathbf{r})$ in Eq. (2.6) depends on the charge density $n(\mathbf{r})$ which is then computed from the Kohn-Sham orbitals ψ_i, the Kohn-Sham equation should be solved self-consistently. Typically, one starts from a trial charge density $n^{(0)}(\mathbf{r})$ and then calculates the effective potential $v^{(0)}_{eff}(\mathbf{r})$. In the next step, the Kohn-Sham Hamiltonian matrix $H_{\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}'}$ in the plane-wave basis is diagonalized to obtain orbitals $\psi^{(1)}_i(\mathbf{r})$. From these orbitals, a new charge density $n^{(1)}(\mathbf{r})$ is computed. This

procedure repeats until the total-energy difference between iteration i and $i+1$ is smaller than a predetermined tolerance. In practice, the number of plane waves in the basis set is of the order from 10^2 to 10^4 per atom. In this range, iterative diagonalizing methods such as the conjugate-gradient (CG) method are more efficient than direct methods.

2.2 Berry-phase polarization and finite electric fields

The development of first-principles methods for treating the effect of an electric field $\mathcal{E}$ in a periodic system has been impeded by the presence of the electrostatic potential $\mathcal{E} \cdot \mathbf{r}$ in the Hamiltonian. This potential is linear in real space and unbounded from below, and thus is incompatible with periodic boundary conditions. The electronic bandstructure becomes ill-defined after application of a potential of this kind. Many attempts have been made to overcome this difficulty. For example, linear-response approaches have been used to treat the electric field as a perturbation [8, 7]. It is possible to formulate these approaches so that only the off-diagonal elements of the position operator

$$\langle \psi_m | \mathbf{r} | \psi_n \rangle = \frac{\langle \psi_m | [H, \mathbf{r}] | \psi_n \rangle}{\epsilon_m - \epsilon_n} \ ,$$

which remain well defined, are needed, thus allowing for the calculation of Born effective charges, dielectric constants, etc. Since it is a perturbative approach, a finite electric field cannot be introduced.

2.2.1 Berry-phase polarization

According to the modern theory of polarization [20], the electronic contribution to the macroscopic polarization is given by

$$\mathbf{P}_{\mathrm{mac}} = \frac{ief}{(2\pi)^3} \sum_{n=1}^{M} \int_{\mathrm{BZ}} d\mathbf{k} \langle u_{\mathbf{k}n} | \nabla_k | u_{\mathbf{k}n} \rangle, \tag{2.21}$$

where e is the charge of an electron ($e < 0$), f=2 for spin degeneracy, M is the number of occupied bands, $u_{\mathbf{k}n}$ are the cell-periodic Bloch functions, and the integral is over the Brillouin zone (BZ). Making the transition to a discretized k-point mesh, this can be written, following Eq. (1.2), in a form

$$\mathbf{P}_{\mathrm{mac}} = \frac{ef}{2\pi\Omega} \sum_{i=1}^{3} \frac{\mathbf{a}_i}{N_\perp^{(i)}} \sum_{l=1}^{N_\perp^{(i)}} \mathrm{Im} \ln \prod_{j=1}^{N_i} \det S_{\mathbf{k}_{lj},\mathbf{k}_{l,j+1}} \tag{2.22}$$

that is amenable to practical calculations. In this expression, for each lattice direction i associated with primitive lattice vector $\mathbf{a}_i$, the BZ is sampled by $N_\perp^{(i)}$ strings of k-points, each with N_i points spanning along the reciprocal lattice vector conjugate to $\mathbf{a}_i$, and

$$(S_{\mathbf{k}\mathbf{k}'})_{mn} = \langle u_{m\mathbf{k}} | u_{n\mathbf{k}'} \rangle \tag{2.23}$$

are the overlap matrices between cell-periodic Bloch vectors at neighboring locations along the string. Because Eqs. (2.21-2.22) can be expressed in terms of Berry phases, this is sometimes referred to as the "Berry-phase theory" of polarization.

2.2.2 Finite electric field

Recently, a total-energy method for treating insulators in nonzero electric fields has been proposed [33, 34]. In this approach, an electric enthalpy functional is defined as a sum of the usual Kohn-Sham energy and an $\mathcal{E} \cdot \mathbf{P}$ term expressing the linear coupling of the electric field to the polarization $\mathbf{P}$:

$$F[\mathbf{R}; \psi; \mathcal{E}] = E_{\mathrm{KS}}[\mathbf{R}; \psi] - \Omega \mathcal{E} \cdot \mathbf{P}_{\mathrm{mac}}[\psi] \ . \tag{2.24}$$

The enthalpy functional is minimized with respect to field-polarized Bloch states, and the information on the response to the electric field is contained in these optimized Bloch states. Using this approach, it is possible to carry out calculations of dynamical effective charges, dielectric susceptibilities, piezoelectric constants, etc., using finite-difference methods [33, 34]. It would also be

possible to study properties of phonons at $\mathbf{q} = 0$, but can be extended to phonons at $\mathbf{q} \neq 0$ only with difficulty by using supercell methods that are restricted to highly commensurate $\mathbf{q}$-vectors.

2.3 Density-functional perturbation theory (DFPT)

2.3.1 Non-self-consistent approach

A large variety of physical properties of solids are based on the lattice-dynamical properties, such as infrared and Raman spectra, ferroelectric transitions, thermal expansions and superconductivity. Density-functional perturbation theory provides a powerful and accurate theoretical tool for calculating phonon frequencies, dielectric constants, Born effective charges and piezoelectric tensors which are second derivatives of the total energy with respect to atomic displacements, electric field, and cell size and shape.

DFPT can be viewed as a combination of DFT and perturbation theory. Consider a small perturbation characterized by a parameter λ applied to the Kohn-Sham potential. Expanding the Kohn-Sham potential in terms of λ gives

$$v_{eff}(r,\lambda) = v_{eff}^{(0)}(\mathbf{r}) + \lambda v_{eff}^{(1)}(\mathbf{r}) + ... \tag{2.25}$$

Due to the perturbation, the Kohn-Sham orbitals and Kohn-Sham energies can also be expanded in terms of λ as

$$\psi_i(\mathbf{r},\lambda) = \psi_i^{(0)}(\mathbf{r}) + \lambda\psi_i^{(1)}(\mathbf{r})... \tag{2.26}$$

$$\epsilon_i(\lambda) = \epsilon_i^{(0)} + \lambda\epsilon_i^{(1)} + ... \tag{2.27}$$

Inserting the expansions of $v_{eff}(r,\lambda)$, $\psi_i(\mathbf{r},\lambda)$ and $\epsilon_i(\lambda)$ into Eq. (2.6) and keeping only the first term yields

$$\left[-\frac{\hbar^2\nabla^2}{2m} + v_{eff}^{(0)}(\mathbf{r}) - \epsilon_i^{(0)}\right]\psi_i^{(1)}(\mathbf{r}) = \left[v_{eff}^{(1)}(\mathbf{r}) - \epsilon_i^{(1)}\right]\psi_i^{(0)}(\mathbf{r})\ , \tag{2.28}$$

which is known as the Sternheimer equation [35]. Multiplying both sides of Eq. (2.28) by $[\psi_i^{(0)}(\mathbf{r})]^*$ and integrating over the unit cell, we have

$$\epsilon_i^{(1)} = \langle \psi_i^{(0)}(\mathbf{r}) | v_{eff}^{(1)}(\mathbf{r}) | \psi_i^{(0)}(\mathbf{r}) \rangle \ . \tag{2.29}$$

Similarly, we can obtain

$$|\psi_i^{(1)}\rangle = \sum_{j \neq i} \frac{\langle \psi_i^{(0)} | v_{eff}^{(1)} | \psi_j^{(0)} \rangle}{\epsilon_j^{(0)} - \epsilon_i^{(0)}} |\psi_j^{(0)}\rangle \ . \tag{2.30}$$

The drawback of the above equation is the summation over all states. This requires the computation of all eigenvalues and eigenstates of the Hamiltonian matrix, which is not efficient at all.

2.3.2 Self-consistent and variational approach

To avoid the drawback mentioned above, one can solve Eq. (2.28) for $\psi_i^{(1)}(\mathbf{r})$ self-consistently under the constraints,

$$\langle \psi_i^{(1)} | \psi_j^{(0)} \rangle = 0 \ . \tag{2.31}$$

All zero-order quantities in Eq. (2.28) are calculated in advance. The first-order charge density can be computed from

$$n^{(1)}(\mathbf{r}) = 2\mathrm{Re} \sum_i \psi_i^{(0)*}(\mathbf{r}) \psi_i^{(1)}(\mathbf{r}) \ . \tag{2.32}$$

The first-order charge density in turn can be used to calculate $v_{eff}^{(1)}(\mathbf{r})$.

Instead of solving the self-consistent equations, one could use the variational method to obtain the first-order wavefunctions $\psi_i^{(1)}(\mathbf{r})$ [8, 9, 10]. Similar to the expansion of the potential and wavefunctions, the total energy can be expanded in terms of λ as

$$E^{\lambda} = E^{(0)} + E^{(1)}\lambda + E^{(2)}\lambda^2 + ... \ . \tag{2.33}$$

The first-order term $E^{(1)}$ is usually vanishing, while the second-order term depends only on the zero-order and the first-order wavefunctions as a consequence of the "2n+1" theorem [36]. The

first-order wavefunctions are obtained by minimizing the second-order energy,

$$E^{(2)} = \min_{\psi^{(1)}} E^{(2)}[\psi^{(0)}, \psi^{(1)}]. \tag{2.34}$$

In fact, the solution $\psi_i^{(1)}(\mathbf{r})$ of Eq. (2.28) is identical to that from Eq. (2.34).

In the next chapter, this approach will be developed further and applied to the cases in the presence of finite electric fields.

Chapter 3

DFPT in the presence of finite electric fields

The understanding of ferroelectric and piezoelectric materials, whose physics is dominated by soft phonon modes, has benefited greatly from the availability of first-principles methods for calculating phonon properties. In general, these methods can be classified into two main types, the direct or frozen-phonon approach [37, 38] and the linear-response approach [10, 9]. In the former approach, the properties of phonons at commensurate wavevectors are obtained from supercell calculations of forces or total-energy changes between between equilibrium and distorted structures. In the latter approach, based on density-functional perturbation theory (DFPT), expressions are derived for the second derivatives of the total energy with respect to atomic displacements, and these are calculated by solving a Sternheimer equation [10] or by using minimization methods [9, 8]. Compared to the direct approach, the linear-response approach has important advantages in that time-consuming supercell calculations are avoided and phonons of arbitrary wavevector can be treated with a cost that is independent of wavevector. However, existing linear-response methods work only at zero electric field.

The development of first-principles methods for treating the effect of an electric field $\mathcal{E}$ in a periodic system has been impeded by the presence of the electrostatic potential $\mathcal{E} \cdot \mathbf{r}$ in the Hamiltonian. This potential is linear in real space and unbounded from below, and thus is incompatible with periodic boundary conditions. The electronic bandstructure becomes ill-defined after application of a potential of this kind. Many attempts have been made to overcome this difficulty. For example, linear-response approaches have been used to treat the electric field as a perturbation [8, 7]. It is

possible to formulate these approaches so that only the off-diagonal elements of the position operator, which remain well defined, are needed, thus allowing for the calculation of Born effective charges, dielectric constants, etc. Since it is a perturbative approach, a finite electric field cannot be introduced.

Recently, a total-energy method for treating insulators in nonzero electric fields has been proposed [33, 34]. In this approach, an electric enthalpy functional is defined as a sum of the usual Kohn-Sham energy and an $\mathcal{E} \cdot \mathbf{P}$ term expressing the linear coupling of the electric field to the polarization $\mathbf{P}$. The enthalpy functional is minimized with respect to field-polarized Bloch states, and the information on the response to the electric field is contained in these optimized Bloch states. Using this approach, it is possible to carry out calculations of dynamical effective charges, dielectric susceptibilities, piezoelectric constants, etc., using finite-difference methods [33, 34]. It would also be possible to use it to study phonon properties in finite electric field, but with the aforementioned limitations (large supercells, commensurate wavevectors) of the direct approach.

In this work, we build upon these recent developments by showing how to extend the linear-response methods so that they can be applied to the finite-field case. That is, we formulate DFPT for the case in which the unperturbed system is an insulator in a finite electric field. Focusing on the case of phonon perturbations, we derive a tractable computational scheme and demonstrate its effectiveness by carrying out calculations of phonon properties of polar semiconductors in finite electric fields.

3.1 Linear response methods for calculating phonon frequencies, dielectric tensor and Born effective charge tensor in finite electric fields.

3.1.1 Effect of electric field on phonon frequencies

Exact theory

We work in the framework of a classical zero-temperature theory of lattice dynamics, so that quantum zero-point and thermal anharmonic effects are neglected. In this context, the phonon frequencies of a crystalline insulator depend upon an applied electric field in three ways: (i) via the variation of the equilibrium lattice vectors (i.e., strain) with applied field; (ii) via the changes in the equilibrium atomic coordinates, even at fixed strain; and (iii) via the changes in the electronic wavefunctions, even at fixed atomic coordinates and strain. Effects of type (i) (essentially, piezoelectric and electrostrictive effects) are beyond the scope of the present work, but are relatively easy to include if needed. This can be done by computing the relaxed strain state as a function of electric field using the approach of Ref. [33], and then computing the phonon frequencies in finite electric field for these relaxed structures using the methods given here. Therefore, in the remainder of the paper, the lattice vectors are assumed to be independent of electric field unless otherwise stated, and we will focus on effects of type (ii) ("lattice effects") and type (iii) ("electronic effects").

In order to separate these two types of effects, we first write the change in phonon frequency resulting from the application of the electric field as

$$\Delta\omega(\mathbf{q};\mathcal{E}) = \omega(\mathbf{q};\mathbf{R}_{\mathcal{E}},\mathcal{E}) - \omega(\mathbf{q};\mathbf{R}_0,0), \tag{3.1}$$

where $\omega(\mathbf{q};\mathbf{R},\mathcal{E})$ is the phonon frequency extracted from the second derivative of the total energy of Eq. (2.24) with respect to the phonon amplitude of the mode of wavevector $\mathbf{q}$, evaluated at displaced coordinated $\mathbf{R}$ and with electrons experiencing electric field $\mathcal{E}$. Also, $\mathbf{R}_{\mathcal{E}}$ are the relaxed

atomic coordinates at electric field $\mathcal{E}$, while $\mathbf{R}_0$ are the relaxed atomic coordinates at zero electric field. Then Eq. (3.1) can be decomposed as

$$\Delta\omega(\mathbf{q};\mathcal{E}) = \Delta\omega_{\rm el}(\mathbf{q};\mathcal{E}) + \Delta\omega_{\rm ion}(\mathbf{q};\mathcal{E}) \tag{3.2}$$

where the electronic part of the response is defined to be

$$\Delta\omega_{\rm el}(\mathbf{q};\mathcal{E}) = \omega(\mathbf{q};\mathbf{R}_0,\mathcal{E}) - \omega(\mathbf{q};\mathbf{R}_0,0) \tag{3.3}$$

and the lattice (or "ionic") part of the response is defined to be

$$\Delta\omega_{\rm ion}(\mathbf{q};\mathcal{E}) = \omega(\mathbf{q};\mathbf{R}_\mathcal{E},\mathcal{E}) - \omega(\mathbf{q};\mathbf{R}_0,\mathcal{E}). \tag{3.4}$$

In other words, the electronic contribution reflects the influence of the electric field on the wavefunctions, and thereby on the force-constant matrix, but evaluated at the zero-field equilibrium coordinates. By contrast, the ionic contribution reflects the additional frequency shift that results from the field-induced ionic displacements.

The finite-electric-field approach of Refs. [33]-[34] provides the methodology needed to compute the relaxed coordinates $\mathbf{R}_\mathcal{E}$, and the electronic states, at finite electric field $\mathcal{E}$. The remainder of this work is devoted to developing and testing the techniques for computing $\omega(\mathbf{q};\mathbf{R},\mathcal{E})$ for given $\mathbf{q}$, R, and $\mathcal{E}$, needed for the evaluation of Eq. (3.1). We shall also use these methods to calculate the various quantities needed to perform the decomposition of Eqs. (3.2-3.4), so that we can also present results for $\Delta\omega_{\rm el}$ and $\Delta\omega_{\rm ion}$ separately in Sec. 3.5.

Approximate theory

Our approach above is essentially an exact one, in which Eq. (3.1) is evaluated by computing all needed quantities at finite electric field. However, we will also compare our approach with an approximate scheme that has been developed in the literature over the last few years [39, 40, 41, 42],

in which the electronic contribution is neglected and the lattice contribution is approximated in such a way that the finite-electric-field approach of Refs. [33]-[34] is not needed.

This approximate theory can be formulated by starting with the approximate electric enthalpy functional [39]

$$F[\mathbf{R};\mathcal{E}] = E^{(0)}_{\rm KS}[\mathbf{R}] - \Omega\mathcal{E}\cdot\mathbf{P}^{(0)}_{\rm mac}[\mathbf{R}], \tag{3.5}$$

where $E^{(0)}_{\rm KS}[\mathbf{R}]$ is the *zero-field* ground-state Kohn-Sham energy at coordinates $\mathbf{R}$, and $\mathbf{P}^{(0)}_{\rm mac}$ is the corresponding *zero-field* electronic polarization. In the presence of an applied electric field $\mathcal{E}$, the equilibrium coordinates that minimize Eq. (3.5) satisfy the force-balance equation

$$-\frac{dE^{(0)}_{\rm KS}}{d\mathbf{R}} + \mathbf{Z}^{(0)}\cdot\mathcal{E} = 0 \tag{3.6}$$

where $\mathbf{Z}^{(0)} = \Omega\, d\mathbf{P}^{(0)}_{\rm mac}/d\mathbf{R}$ is the *zero-field* dynamical effective charge tensor. That is, the sole effect of the electric field is to make an extra contribution to the atomic forces that determine the relaxed displacements; the electrons themselves do not "feel" the electric field except indirectly through these displacements. In Ref. [39], it was shown that this theory amounts to treating the coupling of the electric field to the electronic degrees of freedom in linear order only, while treating the coupling to the lattice degrees of freedom to all orders. Such a theory has been shown to give good accuracy in cases where the polarization is dominated by soft polar phonon modes, but not in systems in which the electronic and lattice polarizations are comparable [39, 40, 41, 42, 43].

In this approximate theory, the effect of the electric field on the lattice dielectric properties [42] and phonon frequencies [41] comes about through the field-induced atomic displacements. Thus, in the notation of Eqs. (3.1-3.4), the frequency shift (relative to zero field) is

$$\Delta\omega'_{\rm ion}(\mathbf{q};\mathcal{E}) = \omega(\mathbf{q};\mathbf{R}'_{\mathcal{E}},0) - \omega(\mathbf{q};\mathbf{R}_0,0) \tag{3.7}$$

in this approximation, where $\mathbf{R}'_{\mathcal{E}}$ is the equilibrium position according to Eq. (3.6). We will make comparisons between the exact $\mathbf{R}_{\mathcal{E}}$ and the approximate $\mathbf{R}'_{\mathcal{E}}$, and the corresponding frequency shifts

$\Delta\omega_{\rm ion}(\mathbf{q},\mathcal{E})$ and $\Delta\omega'_{\rm ion}(\mathbf{q},\mathcal{E})$ later in Sec. 3.5.

3.1.2 Background and definitions

We start from the electric enthalpy functional [33]

$$F[\mathbf{R};\psi;\mathcal{E}] = E_{\rm KS}[\mathbf{R};\psi] - \Omega\mathcal{E}\cdot\mathbf{P}_{\rm mac}[\psi], \tag{3.8}$$

where $E_{\rm KS}$ has the same form as the usual Kohn-Sham energy functional in the absence of an electric field. Here Ω is the cell volume, $\mathbf{P}_{\rm mac}$ is the macroscopic polarization, $\mathcal{E}$ is the homogeneous electric field, $\mathbf{R}$ are the atomic positions, and ψ are the field-polarized Bloch functions. Note that $\mathbf{P}_{\rm mac}$ has both ionic and electronic contributions. The former is an explicit function of $\mathbf{R}$, while the latter is an implicit function of $\mathbf{R}$ through the Bloch functions, which also depend on the atomic positions. When an electric field is present, a local minimum of this functional describes a long-lived metastable state of the system rather than a true ground state (indeed, a true ground state does not exist in finite electric field) [33].

3.1.3 Perturbation expansion of the electric enthalpy functional

We consider an expansion of the properties of the system in terms of small displacements λ of the atoms away from their equilibrium positions, resulting in changes in the charge density, wavefunctions, total energy, etc. We will be more precise about the definition of λ shortly. We adopt a notation in which the perturbed physical quantities are expanded in powers of λ as

$$Q(\lambda) = Q^{(0)} + \lambda Q^{(1)} + \lambda^2 Q^{(2)} + \lambda^3 Q^{(3)} + ... \tag{3.9}$$

where $Q^{(n)} = (1/n!)d^nQ/d\lambda^n$. The immediate dependence upon atomic coordinates is through the external potential $v_{\rm ext}(\lambda)$, which has no electric-field dependence and thus depends upon coordinates and pseudopotentials in the same way as in the zero-field case. The changes in electronic

wavefunctions, charge density, etc. can then be regarded as being induced by the changes in $v_{\rm ext}$.

3.2 Phonon perturbation with zero q wavevector

The nuclear positions can be expressed as

$$\mathbf{R}_{n\nu} = \mathbf{t}_n + \mathbf{d}_\nu + \mathbf{b}_{n\nu}, \tag{3.10}$$

where $\mathbf{t}_n$ is a lattice vector, $\mathbf{d}_\nu$ is a basis vector within the unit cell, and $\mathbf{b}_{n\nu}$ is the instantaneous displacement of atom ν in cell n. We consider in this section a phonon of wavevector $\mathbf{q} = \mathbf{0}$, so that the perturbation does not change the periodicity of the crystal, and the perturbed wavefunctions satisfy the same periodic boundary condition as the unperturbed ones. To be more precise, we choose one sublattice ν and one Cartesian direction α and let $b_{n\nu\alpha} = \lambda$ (independent of n), so that we are effectively moving one sublattice in one direction while while freezing all other sublattice displacements. Since the electric enthalpy functional of Eq. (2.24) is variational with respect to the field-polarized Bloch functions under the constraints of orthonormality, a constrained variational principle exists for the second-order derivative of this functional with respect to atomic displacements [36]. In particular, the correct first-order perturbed wavefunctions $\psi^{(1)}_{m\mathbf{k}}$ can be obtained by minimizing the second-order expansion of the total energy with respect to atomic displacements,

$$\begin{aligned} F^{(2)}[\psi^{(0)}_{m\mathbf{k}};\mathcal{E}] \;=\; & \min_{\psi^{(1)}} \Big(E_{\rm KS}[\psi^{(0)}_{m\mathbf{k}};\psi^{(1)}_{m\mathbf{k}}] \\ & -\Omega \mathbf{P}_{\rm mac}[\psi^{(0)}_{m\mathbf{k}};\psi^{(1)}_{m\mathbf{k}}]\cdot\mathcal{E} \Big)^{(2)}, \end{aligned} \tag{3.11}$$

subject to the constraints

$$\langle \psi^{(0)}_{m\mathbf{k}} | \psi^{(1)}_{n\mathbf{k}} \rangle = 0 \tag{3.12}$$

(where m and n run over occupied states). The fact that only zero-order and first-order wavefunctions appear in Eq. (3.11) is a consequence of the "2n+1 theorem."[36]

Recalling that $|\psi^{(1)}_{n\mathbf{k}}\rangle$ is the first-order wavefunction response to a small real displacement λ of basis atom ν along Cartesian direction α, we can expand the external potential as

$$v_{\rm ext}(\mathbf{r}) = v^{(0)}_{\rm ext}(\mathbf{r}) + v^{(1)}_{{\rm ext},\nu\alpha}(\mathbf{r})\lambda + v^{(2)}_{{\rm ext},\nu\alpha}(\mathbf{r})\lambda^2 + ... \tag{3.13}$$

where

$$v^{(1)}_{{\rm ext},\nu\alpha}(\mathbf{r}) = \sum_n \frac{\partial v_{\rm ext}(\mathbf{r})}{\partial R_{n\nu\alpha}} \ , \tag{3.14}$$

$$v^{(2)}_{{\rm ext},\nu\alpha}(\mathbf{r}) = \sum_n \frac{\partial^2 v_{\rm ext}(\mathbf{r})}{\partial R^2_{n\nu\alpha}} \ , \tag{3.15}$$

etc. From this we shall construct the second-order energy $F^{(2)}$ of Eq. (3.11), which has to be minimized in order to find $|\psi^{(1)}_{n\mathbf{k}}\rangle$. The minimized value of $F^{(2)}$ gives, as a byproduct, the diagonal element of the force-constant matrix associated with displacement $\nu\alpha$. Once the $|\psi^{(1)}_{n\mathbf{k}}\rangle$ have been computed for all $\nu\alpha$, the off-diagonal elements of the force-constant matrix can be calculated using a version of the $2n+1$ theorem as will be described in Sec. 3.2.2.

3.2.1 Discretized k mesh

In practice, we always work on a discretized mesh of k-points, and we have to take into account the orthogonality constraints among wavefunctions at a given k-point on the mesh. Here, we are following the "perturbation expansion after discretization" (PEAD) approach introduced in Ref. [44]. That is, we write down the energy functional in its discretized form, and then consistently derive perturbation theory from this energy functional. Introducing Lagrange multipliers $\Lambda_{\mathbf{k},mn}$ to enforce the orthonormality constraints

$$\langle\psi_{m\mathbf{k}}|\psi_{n\mathbf{k}}\rangle = \delta_{mn}, \tag{3.16}$$

where $\psi_{n\mathbf{k}}$ are the Bloch wavefunctions, and letting N be the number of k-points, the effective total-energy functional of Eq. (2.24) can be written as

$$F = F_{\rm KS} + F_{\rm BP} + F_{\rm LM} \tag{3.17}$$

where $F_{\rm KS} = E_{\rm KS}$, $F_{\rm BP} = -\Omega \mathbf{P}_{\rm mac} \cdot \mathcal{E}$, and $F_{\rm LM}$ are the Kohn-Sham, Berry-phase, and Lagrange-multiplier terms, respectively. The first and last of these are given by

$$F_{\rm KS} = \frac{f}{N} \sum_{\mathbf{k}n}^{\rm occ} \langle \psi_{n\mathbf{k}} | T + v_{\rm ext} | \psi_{n\mathbf{k}} \rangle + E_{\rm Hxc}[n], \tag{3.18}$$

and

$$F_{\rm LM} = -\frac{f}{N} \sum_{\mathbf{k},mn}^{\rm occ} \Lambda_{\mathbf{k},mn} (\langle \psi_{m\mathbf{k}} | \psi_{n\mathbf{k}} \rangle - \delta_{mn}), \tag{3.19}$$

where N is the number of k-points in the BZ. As for the Berry-phase term, we modify the notation of Eq. (2.22) slightly to write this as

$$F_{\rm BP} = -\frac{ef}{2\pi} \sum_{i=1}^{3} \frac{\mathcal{E} \cdot \mathbf{a}_i}{N_\perp^{(i)}} \sum_{\mathbf{k}} D_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} \tag{3.20}$$

where

$$D_{\mathbf{k}\mathbf{k}'} = {\rm Im} \ln \det S_{\mathbf{k}\mathbf{k}'} \tag{3.21}$$

and $\mathbf{g}_i$ is the reciprocal lattice mesh vector in lattice direction i. (That is, $\mathbf{k}$ and $\mathbf{k}+\mathbf{g}_i$ are neighboring k-points in one of the $N_\perp^{(i)}$ strings of k-points running in the reciprocal lattice direction conjugate to $\mathbf{a}_i$.) Recall that the matrix of Bloch overlaps was defined in Eq. (2.23).

We now expand all quantities in orders of the perturbation, e.g., $\Lambda(\lambda) = \Lambda^{(0)} + \lambda\Lambda^{(1)} + \lambda^2\Lambda^{(2)} + ...$, etc. Similarly, we expand $S_{\mathbf{k}\mathbf{k}'}(\lambda) = S^{(0)}_{\mathbf{k}\mathbf{k}'} + \lambda S^{(1)}_{\mathbf{k}\mathbf{k}'} + \lambda^2 S^{(2)}_{\mathbf{k}\mathbf{k}'} + ...$ where

$$\begin{aligned} S^{(1)}_{\mathbf{k}\mathbf{k}',mn} &= \langle u^{(0)}_{m\mathbf{k}} | u^{(1)}_{n\mathbf{k}'} \rangle + \langle u^{(1)}_{m\mathbf{k}} | u^{(0)}_{n\mathbf{k}'} \rangle \ , & (3.22) \\ S^{(2)}_{\mathbf{k}\mathbf{k}',mn} &= \langle u^{(0)}_{m\mathbf{k}} | u^{(2)}_{n\mathbf{k}'} \rangle + \langle u^{(1)}_{m\mathbf{k}} | u^{(1)}_{n\mathbf{k}'} \rangle \\ &\quad + \langle u^{(2)}_{m\mathbf{k}} | u^{(0)}_{n\mathbf{k}'} \rangle \ , & (3.23) \end{aligned}$$

and we also define

$$Q_{\mathbf{k}'\mathbf{k}} = [S^{(0)}_{\mathbf{k}\mathbf{k}'}]^{-1} \tag{3.24}$$

to be the inverse of the zero-order S matrix. Applying the $2n+1$ theorem to Eq. (3.17), the variational second-order derivative of the total-energy functional is

$$F^{(2)} = F^{(2)}_{\rm KS} + F^{(2)}_{\rm BP} + F^{(2)}_{\rm LM} \tag{3.25}$$

where

$$\begin{aligned} F^{(2)}_{\rm KS} &= \frac{1}{N}\sum_{\mathbf{k},m}^{\rm occ}\Big[\langle\psi^{(1)}_{m\mathbf{k}}|T^{(0)}+v^{(0)}_{\rm ext}|\psi^{(1)}_{m\mathbf{k}}\rangle + \langle\psi^{(0)}_{m\mathbf{k}}|v^{(1)}_{\rm ext}|\psi^{(1)}_{m\mathbf{k}}\rangle \\ &\qquad + \langle\psi^{(1)}_{m\mathbf{k}}|v^{(1)}_{\rm ext}|\psi^{(0)}_{m\mathbf{k}}\rangle\Big] + E^{(2)}_{\rm Hxc}[n]\ , \end{aligned} \tag{3.26}$$

$$F^{(2)}_{\rm BP} = -\frac{ef}{4\pi}\sum_{i=1}^{3}\frac{\mathcal{E}\cdot a_i}{N^{(i)}_{\perp}}\sum_{\mathbf{k}} D^{(2)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i}\ , \tag{3.27}$$

$$\begin{aligned} F^{(2)}_{\rm LM} &= \frac{1}{N} - \sum_{\mathbf{k},mn}\Big[\Lambda^{(1)}_{\mathbf{k},mn}\left(\langle\psi^{(1)}_{m\mathbf{k}}|\psi^{(0)}_{n\mathbf{k}}\rangle + \langle\psi^{(0)}_{m\mathbf{k}}|\psi^{(1)}_{n\mathbf{k}}\rangle\right) \\ &\qquad + \Lambda^{(0)}_{\mathbf{k},mn}\langle\psi^{(1)}_{m\mathbf{k}}|\psi^{(1)}_{n\mathbf{k}}\rangle\Big]\ . \end{aligned} \tag{3.28}$$

In the Berry-phase term, Eq. (3.27), we use the approach of Ref. [44] to obtain the expansion of $\ln\det S_{\mathbf{k}\mathbf{k}'}$ with respect to the perturbation. It then follows that

$$D^{(2)}_{\mathbf{k}\mathbf{k}'} = {\rm Im\,Tr}\left[\,2S^{(2)}_{\mathbf{k}\mathbf{k}'}Q_{\mathbf{k}'\mathbf{k}} - S^{(1)}_{\mathbf{k}\mathbf{k}'}Q_{\mathbf{k}'\mathbf{k}}S^{(1)}_{\mathbf{k}\mathbf{k}'}Q_{\mathbf{k}'\mathbf{k}}\,\right] \tag{3.29}$$

where $S^{(2)}$, $S^{(1)}$ and Q are regarded as $L\times L$ matrices (L is the number of occupied bands), matrix products are implied, and Tr is a matrix trace running over the occupied bands. Finally, in the Lagrange-multiplier term, Eq. (3.28), a contribution $\Lambda^{(2)}_{\mathbf{k},mn}(\langle\psi^{(0)}_{m\mathbf{k}}|\psi^{(0)}_{n\mathbf{k}}\rangle - \delta_{mn})$ has been dropped from Eq. (3.28) because the zero-order wavefunctions, which have been calculated in advance, always satisfy the orthonormality constraints $\langle\psi^{(0)}_{m\mathbf{k}}|\psi^{(0)}_{n\mathbf{k}}\rangle = \delta_{mn}$. Moreover, the zero-order Lagrange multipliers are made diagonal by a rotation among zero-order wavefunctions at each k point, and the first-order wavefunctions are made orthogonal to the zero-order ones at each iterative step, so

that Eq. (3.28) simplifies further to become just

$$F^{(2)}_{\rm LM} = -\epsilon_{m\mathbf{k}} \langle \psi^{(1)}_{m\mathbf{k}} | \psi^{(1)}_{m\mathbf{k}} \rangle \ . \tag{3.30}$$

Here, we have restored the notation $\epsilon_{m\mathbf{k}} = \Lambda^{(0)}_{\mathbf{k},mm}$ for the diagonal zero-order Lagrange multipliers.

3.2.2 Conjugate-gradient minimization

The second-order expansion of the electric enthalpy functional in Eq. (3.25) is minimized with respect to the first-order wavefunctions using a "band-by-band" conjugate-gradient algorithm [8, 45]. For a given point $\mathbf{k}$ and band m, the steepest-descent direction at iteration j is $|\zeta_{m\mathbf{k},j}\rangle = \partial F^{(2)}/\partial \langle u^{(1)}_{m\mathbf{k}}|$, where $F^{(2)}$ is given by Eqs. (3.26-3.27) and (3.30). The derivatives of $F^{(2)}_{\rm KS}$ and $F^{(2)}_{\rm LM}$ are straightforward; the new element in the presence of an electric field is the term

$$\frac{\partial E^{(2)}_{\rm BP}}{\partial \langle u^{(1)}_{m\mathbf{k}}|} = -\frac{ief}{4\pi} \sum_{i=1}^{3} \frac{\mathcal{E}\cdot\mathbf{a}_i}{N^{(i)}_{\perp}} \left(|\mathcal{D}_{m\mathbf{k},\mathbf{k}+\mathbf{g}_i}\rangle - |\mathcal{D}_{m\mathbf{k},\mathbf{k}-\mathbf{g}_i}\rangle \right) \tag{3.31}$$

where

$$\mathcal{D}_{m\mathbf{k}\mathbf{k}'} = \left(|u^{(1)}_{\mathbf{k}'}\rangle Q_{\mathbf{k}'\mathbf{k}} - |u^{(0)}_{\mathbf{k}'}\rangle Q_{\mathbf{k}'\mathbf{k}} S^{(1)}_{\mathbf{k}\mathbf{k}'} Q_{\mathbf{k}'\mathbf{k}} \right)_m \ . \tag{3.32}$$

In this equation, $|u^{(1)}_{\mathbf{k}'}\rangle$ and $|u^{(0)}_{\mathbf{k}'}\rangle$ are regarded as vectors of length L (e.g., $|u^{(1)}_{m\mathbf{k}'}\rangle$, $m = 1, L$), and vector-matrix and matrix-matrix products of dimension L are implied inside the parentheses. The standard procedure translates the steepest-descent directions $|\zeta_{m\mathbf{k},j}\rangle$ into preconditioned conjugate-gradient search directions $|\varphi_{m\mathbf{k},j}\rangle$. An improved wavefunction for iteration $j+1$ is then obtained by letting

$$|u^{(1)}_{m\mathbf{k},j+1}\rangle = |u^{(1)}_{m\mathbf{k},j}\rangle + \theta |\varphi_{m\mathbf{k},j}\rangle \ , \tag{3.33}$$

where θ is a real number to be determined. Since the θ-dependence of $F^{(2)}(\theta)$ is quadratic, the minimum of $F^{(2)}$ along the conjugate-gradient direction is easily determined to be

$$\theta_{\rm min} = -\frac{1}{2} \left. \frac{dF^{(2)}}{d\theta} \right|_{\theta=0} \left(\left. \frac{d^2F^{(2)}}{d\theta^2} \right|_{\theta=0} \right)^{-1} \ . \tag{3.34}$$

Construction of the force-constant matrix

To calculate phonon frequencies, we have to construct the force-constant matrix

$$\Phi_{\nu\alpha,\mu\beta} = \frac{\partial^2 F}{\partial R_{\nu\alpha}\partial R_{\mu\beta}} \ . \tag{3.35}$$

Each diagonal element $\Phi_{\mu\beta,\mu\beta}$ has already been obtained by minimizing the $F^{(2)}$ in Eq. (3.25) for the corresponding perturbation $\mu\beta$. The off-diagonal elements $\Phi_{\nu\alpha,\mu\beta}$ can also be determined using only the first-order wavefunctions $u^{(1)}_{m\mathbf{k},\mu\beta}$ using the (non-variational) expression

$$\begin{aligned}\Phi_{\nu\alpha,\mu\beta} &= \frac{2\Omega}{(2\pi)^3}\int_{\mathrm{BZ}}\sum_{m}^{\mathrm{occ}}\Big(\langle u^{(0)}_{m\mathbf{k}}|v^{(1)}_{\mathrm{ext},\nu\alpha}+v^{(1)}_{\mathrm{Hxc},\nu\alpha}|u^{(1)}_{m\mathbf{k},\mu\beta}\rangle \\ &\quad +\langle u^{(0)}_{m\mathbf{k}}|v^{(2)}_{\mathrm{ext},\nu\alpha,\mu\beta}|u^{(0)}_{m\mathbf{k}}\rangle\Big)d\mathbf{k}+\frac{1}{2}E^{(2)}_{\mathrm{Hxc},\nu\alpha,\mu\beta}\end{aligned} \tag{3.36}$$

where $v^{(1)}_{\mathrm{ext},\nu\alpha} = \partial v_{\mathrm{ext}}/\partial R_{\nu\alpha}$ etc.

3.3 Phonon perturbation with nonzero wavevector

In the case of a phonon of arbitrary wavevector $\mathbf{q}$, the displacements of the atoms are essentially of the form $b_{n\nu\alpha} = \lambda\ \exp(i\mathbf{q}\cdot\mathbf{t}_n)$, where λ is a complex number. However, a perturbation of this form does not lead by itself to a Hermitian perturbation of the Hamiltonian. This is unacceptable, because we want the second-order energy to remain real, so that it can be straightforwardly minimized. Thus, we follow the approach of Ref. [8] and take the displacements to be

$$b_{n\nu\alpha} = \lambda\, e^{i\mathbf{q}\cdot\mathbf{t}_n} + \lambda^*\, e^{-i\mathbf{q}\cdot\mathbf{t}_n} \ , \tag{3.37}$$

leading to

$$\begin{aligned}v_{\mathrm{ext}}(\mathbf{r}) &= v^{(0)}_{\mathrm{ext}}(\mathbf{r}) + \lambda\, v^{(1)}_{\mathrm{ext},\nu\alpha,\mathbf{q}}(\mathbf{r}) + \lambda^*\, v^{(1)}_{\mathrm{ext},\nu\alpha,-\mathbf{q}}(\mathbf{r}) \\ &+ \lambda^2\, v^{(2)}_{\mathrm{ext},\nu\alpha,\mathbf{q},\mathbf{q}}(\mathbf{r}) + \lambda^{*2}\, v^{(2)}_{\mathrm{ext},\nu\alpha,-\mathbf{q},-\mathbf{q}}(\mathbf{r}) \\ &+ \lambda\lambda^*\, v^{(2)}_{\mathrm{ext},\nu\alpha,\mathbf{q},-\mathbf{q}}(\mathbf{r}) + \lambda^*\lambda\, v^{(2)}_{\mathrm{ext},\nu\alpha,-\mathbf{q},\mathbf{q}}(\mathbf{r}) \\ &+ \ ...\end{aligned} \tag{3.38}$$

where

$$v^{(1)}_{\mathrm{ext},\nu\alpha,\pm\mathbf{q}}(\mathbf{r}) = \sum_n \frac{\partial v_{\mathrm{ext}}(\mathbf{r})}{\partial R_{n\nu\alpha}} e^{\pm i\mathbf{q}\cdot\mathbf{t}_n} \ , \tag{3.39}$$

$$v^{(2)}_{\mathrm{ext},\nu\alpha,\pm\mathbf{q},\pm\mathbf{q}}(\mathbf{r}) = \sum_{nm} \frac{\partial^2 v_{\mathrm{ext}}(\mathbf{r})}{\partial R_{n\nu\alpha}\partial R_{m\nu\alpha}} \, e^{\pm i\mathbf{q}\cdot\mathbf{t}_n} \, e^{\pm i\mathbf{q}\cdot\mathbf{t}_m} \ , \tag{3.40}$$

etc. Similarly, the field-dependent Bloch wavefunctions ψ and enthalpy functional F can also be expanded in terms of λ and its hermitian conjugate as

$$\psi_{m\mathbf{k}}(\mathbf{r}) = \psi^{(0)}_{m\mathbf{k}}(\mathbf{r}) + \lambda\, \psi^{(1)}_{m\mathbf{k},\mathbf{q}}(\mathbf{r}) + \lambda^*\, \psi^{(1)}_{m\mathbf{k},-\mathbf{q}}(\mathbf{r}) + ... \tag{3.41}$$

and

$$\begin{aligned} F[\mathcal{E}] &= \lambda F^{(0)}[\mathcal{E}] + F^{(1)}_{\mathbf{q}}[\mathcal{E}] + \lambda^* F^{(1)}_{-\mathbf{q}}[\mathcal{E}] \\ &+ \lambda^2 F^{(2)}_{\mathbf{q},\mathbf{q}}[\mathcal{E}] + 2\lambda\lambda^* F^{(2)}_{\mathbf{q},-\mathbf{q}}[\mathcal{E}] \\ &+ \lambda^{*2} F^{(2)}_{-\mathbf{q},-\mathbf{q}}[\mathcal{E}] + ... \ . \end{aligned} \tag{3.42}$$

The first-order wavefunctions in response to a perturbation with wavevector $\mathbf{q}$ have translational properties

$$\psi^{(1)}_{m\mathbf{k},\mathbf{q}}(\mathbf{r}+\mathbf{R}) = e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{R}} \psi^{(1)}_{m\mathbf{k},\mathbf{q}}(\mathbf{r}) \tag{3.43}$$

that differ from those of the zero-order wavefunctions

$$\psi^{(0)}_{m\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi^{(0)}_{m\mathbf{k}}(\mathbf{r}) \ . \tag{3.44}$$

As a result, we cannot simply work in terms of perturbed Bloch functions or use the usual Berry-phase expression in terms of strings of Bloch functions. Also, in contrast to the $\mathbf{q}$=0 case, in which only one set of first-order wavefunctions was needed, we now need to solve for two sets $\psi^{(1)}_{m\mathbf{k},\pm\mathbf{q}}$ corresponding to the non-Hermitian perturbation at wavevector $\mathbf{q}$ and its Hermitian conjugate at wavevector $-\mathbf{q}$ [8].

We now proceed to write out the second-order energy functional $F^{(2)}[\psi^{(0)}_{m\mathbf{k}};\psi^{(1)}_{m\mathbf{k},\pm\mathbf{q}};\mathcal{E}]$, corresponding to the sum of the quadratic terms in Eq. (3.42), and minimize it simultaneously with respect to $\psi^{(1)}_{m\mathbf{k},\mathbf{q}}$ and $\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}$.

First, making the same decomposition as in Eq. (3.25), we find that the Kohn-Sham part is

$$F^{(2)}_{\rm KS} = E^{(2)}_{\mathbf{q},-\mathbf{q}}[\psi^{(0)}_{m\mathbf{k}};\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}] + E^{(2)}_{-\mathbf{q},\mathbf{q}}[\psi^{(0)}_{m\mathbf{k}};\psi^{(1)}_{m\mathbf{k},\mathbf{q}}] \;, \tag{3.45}$$

where

$$\begin{aligned} E^{(2)}_{-\mathbf{q},\mathbf{q}} \;=\; & \frac{2\Omega}{(2\pi)^3}\int_{\rm BZ}\sum_{m}^{\rm occ}\Big(\langle u^{(1)}_{m\mathbf{k},\mathbf{q}}|v^{(0)}_{{\rm ext},\mathbf{k}+\mathbf{q},\mathbf{k}+\mathbf{q}}|u^{(1)}_{m\mathbf{k},\mathbf{q}}\rangle \\ & + \langle u^{(1)}_{m\mathbf{k},\mathbf{q}}|v^{(0)}_{{\rm Hxc},\mathbf{k}+\mathbf{q},\mathbf{k}+\mathbf{q}}|u^{(1)}_{m\mathbf{k},\mathbf{q}}\rangle \\ & + \langle u^{(1)}_{m\mathbf{k},\mathbf{q}}|v^{(1)}_{{\rm ext},\mathbf{k}+\mathbf{q},\mathbf{k}} + v^{(1)}_{{\rm Hxc},\mathbf{k}+\mathbf{q},\mathbf{k}}|u^{(0)}_{m\mathbf{k}}\rangle \\ & + \langle u^{(0)}_{m\mathbf{k}}|v^{(1)}_{{\rm ext},\mathbf{k},\mathbf{k}+\mathbf{q}} + v^{(1)}_{{\rm Hxc},\mathbf{k},\mathbf{k}+\mathbf{q}}|u^{(1)}_{m\mathbf{k},\mathbf{q}}\rangle \\ & + \langle u^{(0)}_{m\mathbf{k}}|v^{(2)}_{{\rm ext},\mathbf{k},\mathbf{k}}|u^{(0)}_{m\mathbf{k}}\rangle\Big)d\mathbf{k} + \frac{1}{2}E^{(2)}_{\rm Hxc} \;. \end{aligned} \tag{3.46}$$

Note that terms $E^{(2)}_{\mathbf{q},\mathbf{q}}$ and $E^{(2)}_{-\mathbf{q},-\mathbf{q}}$ vanish, essentially because such terms transform like perturbations of wavevector $\pm 2\mathbf{q}$ which, except when $2\mathbf{q}$ equals a reciprocal lattice vector, are inconsistent with crystal periodicity and thus cannot appear in the energy expression. (If $2\mathbf{q}$ is equal to a reciprocal lattice vector, $E^{(2)}_{\mathbf{q},\mathbf{q}}$ and $E^{(2)}_{-\mathbf{q},-\mathbf{q}}$ still vanish, as can be shown using time-reversal symmetry.)

Second, we consider the Berry-phase coupling term. The treatment of this term is rather subtle because, as mentioned above, the perturbed wavefunctions are now admixtures of parts with periodicity as in Eq. (3.43) and as in Eq. (3.44), so that the usual Berry-phase formula for the polarization [20] cannot be used. A different approach is needed now in order to express the polarization in terms of the perturbed wavefunctions. For this purpose, we consider a virtual supercell in which the wavevectors $\mathbf{k}$ and $\mathbf{q}$ would be commensurate, and make use of the definition introduced by Resta [46] specialized to the non-interacting case. The details of this treatment are deferred to the

Appendix, but the results can be written in the relatively simple form

$$F^{(2)}_{\rm BP} = -\frac{ef}{2\pi}\sum_{i=1}^{3}\frac{\mathcal{E}\cdot a_i}{N^{(i)}_{\perp}}\sum_{\mathbf{k}} D^{(2)}_{\mathbf{k}}(\mathbf{g}_i) \tag{3.47}$$

where

$$\begin{aligned} D^{(2)}_{\mathbf{k}}(\mathbf{g}) = \mathrm{Tr}\,\big[\, & S^{(1,1)}_{\mathbf{k},\mathbf{k}+\mathbf{g}} Q_{\mathbf{k}+\mathbf{g},\mathbf{k}} - S^{(1,0)}_{\mathbf{k},\mathbf{k}+\mathbf{g}-\mathbf{q}} \\ & \times Q_{\mathbf{k}+\mathbf{g}-\mathbf{q},\mathbf{k}-\mathbf{q}} S^{(0,1)}_{\mathbf{k}-\mathbf{q},\mathbf{k}+\mathbf{g}} Q_{\mathbf{k}+\mathbf{g},\mathbf{k}}\,\big] \end{aligned} \tag{3.48}$$

with $Q_{\mathbf{k}'\mathbf{k}}$ given by Eq. (3.24) and the superscript notation $S^{(s,t)} = \partial^{s+t}S/\partial(\lambda^*)^s\partial\lambda^t$. From Eqs. (2.23) and (3.43), we can write these explicitly as

$$\begin{aligned} S^{(1,0)}_{\mathbf{k}\mathbf{k}',mn} &= \langle\psi^{(0)}_{m\mathbf{k}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(1)}_{n\mathbf{k}',\mathbf{q}}\rangle \\ &\quad + \langle\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(0)}_{n\mathbf{k}'}\rangle\ , \end{aligned} \tag{3.49}$$

$$\begin{aligned} S^{(0,1)}_{\mathbf{k}\mathbf{k}',mn} &= \langle\psi^{(0)}_{m\mathbf{k}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(1)}_{n\mathbf{k}',-\mathbf{q}}\rangle \\ &\quad + \langle\psi^{(1)}_{m\mathbf{k},\mathbf{q}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(0)}_{n\mathbf{k}'}\rangle\ , \end{aligned} \tag{3.50}$$

$$\begin{aligned} S^{(1,1)}_{\mathbf{k}\mathbf{k}',mn} &= \langle\psi^{(1)}_{m\mathbf{k},\mathbf{q}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(1)}_{n\mathbf{k}',\mathbf{q}}\rangle \\ &\quad + \langle\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}|e^{-i\mathbf{g}\cdot\mathbf{r}}|\psi^{(1)}_{n\mathbf{k}',-\mathbf{q}}\rangle\ . \end{aligned} \tag{3.51}$$

Third, the treatment of the Lagrange-multiplier term is straightforward; in analogy with Eq. (3.30), we obtain

$$F^{(2)}_{\rm LM} = -\epsilon_{m\mathbf{k}}\big(\langle\psi^{(1)}_{m\mathbf{k},\mathbf{q}}|\psi^{(1)}_{m\mathbf{k},\mathbf{q}}\rangle + \langle\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}|\psi^{(1)}_{m\mathbf{k},-\mathbf{q}}\rangle\big)\ . \tag{3.52}$$

If we look closely at Eq. (3.48), we see that the second term involves not simply pairs of k-points separated by the mesh vector $\mathbf{g}$, but *quartets* of k-points, as illustrated in Fig. 3.1. Reading from left to right in the second term of Eq. (3.48), the k-point labels are $\mathbf{k}$, then $\mathbf{k}+\mathbf{g}-\mathbf{q}$, then $\mathbf{k}-\mathbf{q}$, then $\mathbf{k}+\mathbf{g}$, and finally back to $\mathbf{k}$. This is the loop illustrated in Fig. 3.1. Each dark arrow represents a

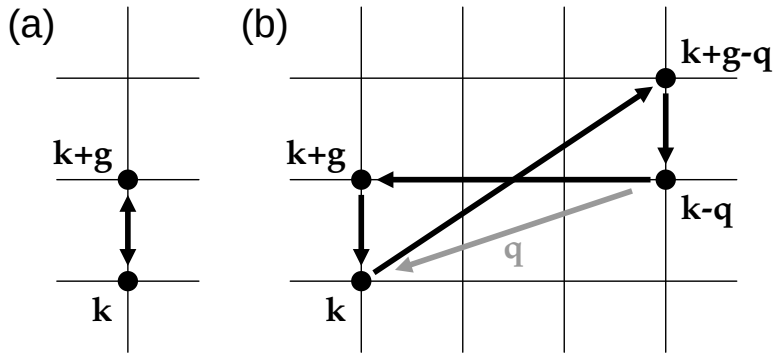

Figure 3.1: Pattern of couplings between k-points arising in (a) the first term, and (b) the second term, of Eq. (3.48). Reciprocal vector $\mathbf{q}$ is the phonon wavevector, while $\mathbf{g}$ is a primitive vector of the k-point mesh (indicated by thin horizontal and vertical lines).

matrix element of $S^{(1,0)}$, $S^{(0,1)}$, or Q; the gray arrow indicates the phonon $\mathbf{q}$-vector. These loops arise because there are two kinds of coupling between k-points entering into the present theory. First, even in the absence of the phonon perturbation, wavevectors at neighboring k-points separated by mesh vector $\mathbf{g}$ are coupled by the $\mathcal{E}\cdot\mathbf{P}$ term in the energy functional. Second, the phonon introduces a perturbation at wavevector $\mathbf{q}$. It is the interplay between these two types of inter-k-point coupling that is responsible for the appearance of these four-point loops in the expression for $F^{(2)}_{\rm BP}$.

The implementation of the conjugate-gradient minimization algorithm proceeds in a manner very similar to that outlined in Sec. 3.2.2. Naively, one would have to work simultaneously with the two search-direction vectors

$$\begin{aligned} |\zeta_{m\mathbf{k},\mathbf{q}}\rangle &= \partial F^{(2)}/\partial\langle u^{(1)}_{m\mathbf{k},\mathbf{q}}| \,, \\ |\zeta_{m\mathbf{k},-\mathbf{q}}\rangle &= \partial F^{(2)}/\partial\langle u^{(1)}_{m\mathbf{k},-\mathbf{q}}| \,, \end{aligned} \tag{3.53}$$

where $u^{(1)}_{m\mathbf{k},\pm\mathbf{q}}$ are the periodic parts of $\psi^{(1)}_{m\mathbf{k},\pm\mathbf{q}}$. However, minimizing the second-order energy $F^{(2)}$ with respect to two sets of first-order wavefunctions $u_{n\mathbf{k},\pm\mathbf{q}}$ would double the computational cost and would involve substantial restructuring of existing computer codes. We can avoid this by using the fact that the second-order energy is invariant under time reversal to eliminate one set of

first-order wavefunctions $\psi^{(1)}_{n\mathbf{k},-\mathbf{q}}$ in favor of the other set $\psi^{(1)}_{n\mathbf{k},\mathbf{q}}$ following the approach given in Ref. [8]. Specifically, the two sets of first-order wavefunctions are related by

$$\psi^{(0)}_{n\mathbf{k}}(\mathbf{r}) = e^{i\theta_{n\mathbf{k}}}\psi^{(0)*}_{n-\mathbf{k}}(\mathbf{r}) \,, \tag{3.54}$$

$$\psi^{(1)}_{n\mathbf{k},\mathbf{q}}(\mathbf{r}) = e^{i\theta_{n\mathbf{k}}}\psi^{(1)*}_{n-\mathbf{k},-\mathbf{q}}(\mathbf{r}) \,, \tag{3.55}$$

where $\theta_{n\mathbf{k}}$ is an arbitrary phase independent of $\mathbf{r}$. The arbitrary phase $\theta_{n\mathbf{k}}$ cancels out in the expression of $F^{(2)}$ since every term in $F^{(2)}$ is independent of the phase of the first-order wavefunctions. Thus, we choose $\theta_{n\mathbf{k}} = 0$ for simplicity and write the second-order energy functional in terms of wavefunctions $\psi_{n\mathbf{k},\mathbf{q}}$ only.

The minimization procedure now proceeds in a manner similar to the zero-wavevector case, except that the calculation of the Berry-phase part involves some vector-matrix-matrix products as in Eq. (3.32), but circulating around three of the sides of the loop in Fig. 3.1. Since $F^{(2)}$ remains in a quadratic form, the minimum of $F^{(2)}$ is again easily searched along the conjugate-gradient direction. Wavefunctions are updated over k-points one after another, and the first-order wavefunctions are updated. This procedure continues until the self-consistent potential is converged. Once the first-order responses of wavefunctions are obtained, the diagonal elements of the dynamical matrix are obtained by evaluating $F^{(2)}$, and the off-diagonal elements are obtained from a straightforward generalization of Eq. (3.36),

$$\begin{aligned}\Phi_{\nu\alpha,\mu\beta} = \frac{2\Omega}{(2\pi)^3}\int_{\mathrm{BZ}}\sum_{m}^{\mathrm{occ}}\Big(&\langle u^{(0)}_{m\mathbf{k}}|v^{(1)}_{\mathrm{ext},\nu\alpha,\mathbf{k},\mathbf{k}+\mathbf{q}}|u^{(1)}_{m\mathbf{k},\mu\beta,\mathbf{q}}\rangle \\ &+\langle u^{(0)}_{m\mathbf{k}}|v^{(1)}_{\mathrm{Hxc},\nu\alpha,\mathbf{k},\mathbf{k}+\mathbf{q}}|u^{(1)}_{m\mathbf{k},\mu\beta,\mathbf{q}}\rangle \\ &+\langle u^{(0)}_{m\mathbf{k}}|v^{(2)}_{\mathrm{ext},\nu\alpha,\mu\beta}|u^{(0)}_{m\mathbf{k}}\rangle\Big)d\mathbf{k} \\ &+\frac{1}{2}E^{(2)}_{\mathrm{Hxc},\nu\alpha,\mu\beta} \,.\end{aligned} \tag{3.56}$$

3.4 Electric field perturbation

We start from the electric enthalpy functional [44, 33]

$$F[\mathbf{R};\psi;\mathcal{E}] = E_{\rm KS}[\mathbf{R};\psi] - \Omega\mathcal{E}\cdot\mathbf{P}[\psi] \ , \tag{3.57}$$

where $\mathbf{R}$, $\mathcal{E}$, Ω and $\mathbf{P}$ are, respectively, the atomic coordinates, the electric field, the cell volume, and the macroscopic polarization, $E_{\rm KS}$ is Kohn-Sham energy functional at zero electric field, and atomic units are used throughout. After minimizing this functional, the field-polarized Bloch functions ψ may be regarded as depending implicitly on the electric field $\mathcal{E}$. Our treatment of this functional will parallel the treatment given in Sec. 3.3.

In the present case, we take the electric field $\mathcal{E}$ to consist of two parts, a finite part $\mathcal{E}^{(0)}$ and a small variation $\delta\mathcal{E}$. In the following, we consider the perturbation expansion of the functional in Eq. (2.24) with respect to the small variation $\delta\mathcal{E}$ under the orthonormality constraints

$$\langle\psi_{m\mathbf{k}}|\psi_{n\mathbf{k}}\rangle = \delta_{mn} \ . \tag{3.58}$$

The wave functions are to be relaxed, subject to these constraints, in such a way as to minimize the electric enthalpy functional

$$F = F_{\rm KS} + F_{\rm BP} + F_{\rm LM} \ , \tag{3.59}$$

where $F_{\rm KS} = E_{\rm KS}$ is the Kohn-Sham energy (as it would be calculated at $\mathcal{E} = 0$), $F_{\rm BP} = -\Omega\mathcal{E}\cdot\mathbf{P}$ contains the coupling of the Berry-phase polarization $\mathbf{P}$ to the electric field, and the constraint is implemented by the inclusion of the Lagrange-multiplier term $F_{\rm LM}$. The first and last of these terms are given by

$$F_{\rm KS} = \frac{f}{N_k}\sum_{\mathbf{k}n}^{\rm occ}\langle\psi_{n\mathbf{k}}|T + v_{\rm ext}|\psi_{n\mathbf{k}}\rangle + E_{\rm Hxc}[n] \tag{3.60}$$

and

$$F_{\rm LM} = -\frac{f}{N_k}\sum_{\mathbf{k},mn}^{\rm occ}\Lambda_{\mathbf{k},mn}(\langle\psi_{m\mathbf{k}}|\psi_{n\mathbf{k}}\rangle - \delta_{mn}) \tag{3.61}$$

where f is the spin degeneracy (normally f=2), N_k is the number of k-points, and $\Lambda_{\mathbf{k},mn}$ is the matrix of Lagrange multipliers. In a notation similar to that of Sec. 3.2, the second term may be written as

$$F_{\rm BP} = -\frac{ef}{2\pi}\sum_{i=1}^{3}\frac{\mathcal{E}\cdot\mathbf{a}_i}{N_\perp^{(i)}}\sum_{\mathbf{k}} D_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} \ . \tag{3.62}$$

Here $\mathbf{a}_i$ are the three primitive real-space lattice vectors, and the mesh of N_k k-points is defined by mesh vectors $\mathbf{g}_i = \mathbf{b}_i/N^{(i)}$ where $\mathbf{b}_i$ is the reciprocal lattice vector dual to $\mathbf{a}_i$. Thus, $N_k = N^{(1)}N^{(2)}N^{(3)}$, and we also define $N_\perp^{(i)} = N_k/N^{(i)}$ as the number of k-point strings running in direction i. Finally,

$$D_{\mathbf{k}\mathbf{k}'} = \mathrm{Im}\ln\det S_{\mathbf{k}\mathbf{k}'} \tag{3.63}$$

where the overlap matrix is defined as

$$(S_{\mathbf{k}\mathbf{k}'})_{mn} \ = \ \langle u_{m\mathbf{k}}|u_{n\mathbf{k}'}\rangle \ . \tag{3.64}$$

In order to obtain the desired response properties, we now wish to expand the finite-field enthalpy functional $F_{\rm KS}$ up to second order in the electric field. We shall assume for the moment that the electric field is applied in Cartesian direction α only. The expansion of $F_{\rm KS}$ with respect to atomic displacements was already obtained in Sec. 3.2, and the expansion with respect to electric field can be carried through in a very similar way. Indeed, the second-order expansions of $F_{\rm KS}$ and $F_{\rm LM}$ can essentially be transcribed from Sec. 3.2 with the first-order wave functions with respect to displacement replaced here by the first-order wave functions with respect to electric field, giving

$$\begin{aligned} F_{\rm KS}^{(2)} \ &= \ \frac{1}{2}\frac{\partial^2 F_{\rm KS}}{\partial\mathcal{E}_\alpha^2} \\ &= \ \frac{f}{N_k}\sum_{\mathbf{k}}\sum_{n=1}^{occ}\langle u_{n\mathbf{k}}^{\mathcal{E}_\alpha}|T+v_{ext}|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle \\ &\qquad\qquad + E_{\rm Hxc}^{\mathcal{E}_\alpha\mathcal{E}_\alpha} \end{aligned} \tag{3.65}$$

and

$$F^{(2)}_{\rm LM} \;=\; -\frac{f}{N_k}\sum_{\mathbf{k},n}^{\rm occ}\Lambda^{(0)}_{\mathbf{k},nn}\langle u^{\mathcal{E}_\alpha}_{n\mathbf{k}}|u^{\mathcal{E}_\alpha}_{n\mathbf{k}}\rangle\ . \tag{3.66}$$

As in Sec. 3.2, terms that can be eliminated by use of the "$2n+1$ theorem" (e.g., $\langle u^{\mathcal{E}_\alpha\mathcal{E}_\alpha}_{n\mathbf{k}}|T+v_{ext}|u^{(0)}_{n\mathbf{k}}\rangle$) have been dropped. The the first-order wave functions are

$$|u^{\mathcal{E}_\alpha}_{n\mathbf{k}_j}\rangle = \frac{\partial|u_{n\mathbf{k}_j}\rangle}{\partial\mathcal{E}_\alpha} \tag{3.67}$$

and the second-order $E_{\rm Hxc}$ are

$$E^{\mathcal{E}_\alpha\mathcal{E}_\alpha}_{\rm Hxc} = \frac{\partial^2 E_{\rm Hxc}}{2\partial\mathcal{E}_\alpha\partial\mathcal{E}_\alpha}\ . \tag{3.68}$$

In these and subsequent equations, the partial derivatives indicate that the *structural* coordinates $\mathbf{R}$ are being held fixed (while, however, the *wave functions* $|u_{n\mathbf{k}}\rangle$ are allowed to vary).

The second-order expansion of $F_{\rm BP}$ with respect to electric field requires somewhat more care. We find

$$\begin{aligned} F^{(2)}_{\rm BP} &= \frac{1}{2}\frac{\partial^2 F_{\rm BP}}{\partial\mathcal{E}_\alpha^2} \\ &= -\frac{\Omega}{2}\frac{\partial^2(\mathcal{E}\cdot\mathbf{P})}{\partial\mathcal{E}_\alpha^2} \\ &= -\Omega(\hat{\mathbf{e}}_\alpha\cdot\mathbf{P}^{\mathcal{E}_\alpha}+\mathcal{E}^{(0)}\cdot\mathbf{P}^{\mathcal{E}_\alpha\mathcal{E}_\alpha})\ , \end{aligned} \tag{3.69}$$

where $\hat{\mathbf{e}}_\alpha$ is the unit vector along Cartesian direction α. The first term in the last line of Eq. (3.69) is special to the case of the electric-field perturbation, while the second term can be derived in close correspondence to the case of displacement perturbations in Sec. 3.2. The first-order variation of $\mathbf{P}$ with field $\mathcal{E}_\alpha$ is

$$\mathbf{P}^{\mathcal{E}_\alpha} = -\frac{\mathbf{ef}}{\mathbf{2\pi\Omega}}\sum_{\mathbf{i=1}}^{\mathbf{3}}\frac{\mathbf{a_i}}{\mathbf{N}^{(\mathbf{i})}_\perp}\sum_{\mathbf{k}}\mathbf{D}^{(\mathbf{1})}_{\mathbf{k},\mathbf{k}+\mathbf{g_i}} \tag{3.70}$$

and its second-order variation is

$$\mathbf{P}^{\mathcal{E}_\alpha\mathcal{E}_\alpha} = -\frac{\mathbf{ef}}{\mathbf{4\pi\Omega}}\sum_{\mathbf{i=1}}^{\mathbf{3}}\frac{\mathbf{a_i}}{\mathbf{N}^{(\mathbf{i})}_\perp}\sum_{\mathbf{k}}\mathbf{D}^{(\mathbf{2})}_{\mathbf{k},\mathbf{k}+\mathbf{g_i}}\ , \tag{3.71}$$

where

$$D^{(1)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} = \mathrm{ImTr}\left[S^{(1)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} Q_{\mathbf{k}+\mathbf{g}_i,\mathbf{k}}\right] \tag{3.72}$$

and

$$\begin{aligned} D^{(2)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} &= \mathrm{ImTr}\Big[2S^{(2)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} Q_{\mathbf{k}+\mathbf{g}_i,\mathbf{k}} \\ &\quad - S^{(1)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} Q_{\mathbf{k}+\mathbf{g}_i,\mathbf{k}} S^{(1)}_{\mathbf{k},\mathbf{k}+\mathbf{g}_i} Q_{\mathbf{k}+\mathbf{g}_i,\mathbf{k}}\Big] \ . \end{aligned} \tag{3.73}$$

In these equations, 'Tr' indicates a trace of the bracketed matrix over band indices, and Q, $S^{(1)}$, and $S^{(2)}$ are defined with respect to the series expansion of the overlap matrix via

$$S_{\mathbf{k}\mathbf{k}'}(\mathcal{E}_\alpha) = S^{(0)}_{\mathbf{k}\mathbf{k}'} + \mathcal{E}_\alpha S^{(1)}_{\mathbf{k}\mathbf{k}'} + \mathcal{E}^2_\alpha S^{(2)}_{\mathbf{k}\mathbf{k}'} + ... \tag{3.74}$$

and

$$Q_{\mathbf{k}\mathbf{k}'} = [S^{(0)}_{\mathbf{k}'\mathbf{k}}]^{-1} \ . \tag{3.75}$$

The first- and second-order expansions of the overlap matrix take the form

$$S^{(1)}_{\mathbf{k},\mathbf{k}',mn} = \langle u^{\mathcal{E}_\alpha}_{m\mathbf{k}} | u^{(0)}_{n\mathbf{k}'} \rangle + \langle u^{(0)}_{m\mathbf{k}} | u^{\mathcal{E}_\alpha}_{n\mathbf{k}'} \rangle \tag{3.76}$$

and

$$S^{(2)}_{\mathbf{k},\mathbf{k}',mn} = \langle u^{\mathcal{E}_\alpha}_{m\mathbf{k}} | u^{\mathcal{E}_\alpha}_{n\mathbf{k}'} \rangle \ . \tag{3.77}$$

In the last equation above, terms like $\langle u^{\mathcal{E}_\alpha\mathcal{E}_\alpha}_{m\mathbf{k}} | u^{(0)}_{n\mathbf{k}'} \rangle$ have again been dropped by virtue of the "$2n+1$ theorem."

3.4.1 First-order wave functions with respect to electric-field perturbation

The second-order term in the expansion of the energy functional, given by the sum $F^{(2)} = F^{(2)}_{\mathrm{KS}} + F^{(2)}_{\mathrm{BP}} + F^{(2)}_{\mathrm{LM}}$ of the expressions in Eqs. (3.65), (3.69), and (3.66) respectively, is minimized with

respect to the first-order wave functions $|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle$ using standard conjugate-gradient methods. The steepest-descent direction is obtained from the gradient of $F^{(2)}$ with respect to $\langle u_{n\mathbf{k}}^{\mathcal{E}_\alpha}|$, whose contributions take the form

$$\frac{\delta F_{\rm KS}^{(2)}}{\delta u_{n\mathbf{k}}^{\mathcal{E}_\alpha *}} = \frac{f}{N_k}\left[\left(T+v_{ext}^{(0)}\right)|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle + \frac{\delta E_{\rm Hxc}^{\mathcal{E}_\alpha\mathcal{E}_\alpha}}{\delta u_{n\mathbf{k}}^{\mathcal{E}_\alpha *}}\right], \tag{3.78}$$

$$\begin{aligned}\frac{\delta F_{\rm BP}^{(2)}}{\delta u_{n\mathbf{k}}^{\mathcal{E}_\alpha *}} &= \frac{ief}{4\pi}\sum_{i=1}^{3}\frac{\mathcal{E}^{(0)}\cdot\mathbf{a}_i}{N_\perp^{(i)}}\left(|\mathcal{C}_{m\mathbf{k},\mathbf{k}+\mathbf{g}_i}\rangle - |\mathcal{C}_{m\mathbf{k},\mathbf{k}-\mathbf{g}_i}\rangle\right)\\ &+ \frac{ief}{4\pi}\sum_{i=1}^{3}\frac{\hat{\mathbf{e}}_\alpha\cdot\mathbf{a}_i}{N_\perp^{(i)}}\left(|\mathcal{D}_{m\mathbf{k},\mathbf{k}+\mathbf{g}_i}\rangle - |\mathcal{D}_{m\mathbf{k},\mathbf{k}-\mathbf{g}_i}\rangle\right),\end{aligned} \tag{3.79}$$

and

$$\frac{\delta F_{\rm LM}^{(2)}}{\delta u_{n\mathbf{k}}^{\mathcal{E}_\alpha *}} = \frac{f}{N_k}\epsilon_{n\mathbf{k}}^{(0)}|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle\,. \tag{3.80}$$

Here

$$\mathcal{C}_{m\mathbf{k}\mathbf{k}'} = \left(|u_{\mathbf{k}'}^{\mathcal{E}_\alpha}\rangle Q_{\mathbf{k}'\mathbf{k}} - |u_{\mathbf{k}'}^{(0)}\rangle Q_{\mathbf{k}'\mathbf{k}} S_{\mathbf{k}\mathbf{k}'}^{(1)} Q_{\mathbf{k}'\mathbf{k}}\right)_m, \tag{3.81}$$

$$\mathcal{D}_{m\mathbf{k}\mathbf{k}'} = \left(|u_{\mathbf{k}'}^{(0)}\rangle Q_{\mathbf{k}'\mathbf{k}}\right)_m, \tag{3.82}$$

and $\epsilon_{n\mathbf{k}}^{(0)}$ is the diagonal zero-order matrix of Lagrange multipliers. Convergence of the conjugate-gradient procedure yields a set of first-order wave functions $|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle$. These then become the essential ingredients for constructing the dielectric and Born charge tensors as discussed below.

3.4.2 Dielectric permittivity tensor

The dielectric permittivity tensor can be written as

$$\epsilon_{\alpha\beta}^{\infty} = \delta_{\alpha\beta} + 4\pi\chi_{\alpha\beta} \tag{3.83}$$

where the electric susceptibility tensor $\chi_{\alpha\beta}$ at a *finite* electric field is defined as

$$\begin{aligned} \chi_{\alpha\beta} &= -\frac{1}{\Omega}\frac{\partial^2 F(\mathcal{E})}{\partial\mathcal{E}_\alpha \partial\mathcal{E}_\beta}\bigg|_{\mathcal{E}=\mathcal{E}^{(0)}} \\ &= \frac{\partial P_\alpha}{\partial\mathcal{E}_\beta}\bigg|_{\mathcal{E}=\mathcal{E}^{(0)}} = \hat{\mathbf{e}}_\alpha \cdot \mathbf{P}^{\mathcal{E}_\beta} \ . \end{aligned} \tag{3.84}$$

The derivative $\mathbf{P}^{\mathcal{E}_\beta}$ of the polarization with respect to electric field is already given by Eq. (3.70). Since the first-order wave functions $|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle$ have already been obtained in Sec. 3.4.1, it is straightforward to evaluate Eq. (3.84) and thus obtain the polarizability and permittivity.

The dielectric responses above are the static responses computed with atomic coordinates frozen. That is, they correspond to the dielectric response that would be measured at frequencies low compared to electronic frequencies but high compared to any infrared-active phonon modes. The true static susceptibility could be computed by including the lattice displacements (and, if appropriate, the piezoelectric strains) using, e.g., the methods of Ref. [47].

3.4.3 Born effective charge tensor

The electronic contribution to the Born effective charge tensor at finite electric field takes the form

$$Z^*_{\kappa,\alpha\beta} = -\frac{\partial^2 F(\mathcal{E})}{\partial\mathcal{E}_\alpha \partial\tau_{\kappa,\beta}}\bigg|_{\mathcal{E}=\mathcal{E}^{(0)}} . \tag{3.85}$$

This expression can be calculated equivalently in two different ways. First, introducing the force $f_{\kappa,\alpha} = -\partial F(\mathcal{E})/\partial\tau_{\kappa,\alpha}$ acting on atom κ in direction α, it can be written as

$$Z^*_{\kappa,\alpha\beta} = \frac{\partial f_{\kappa,\beta}}{\partial\mathcal{E}_\alpha} \ . \tag{3.86}$$

Using the Hellmann-Feynman theorem, the expression for the force is given as

$$f_{\kappa,\beta} = \frac{f}{N_k}\sum_{\mathbf{k}}\sum_{n=1}^{occ}\langle u_{n\mathbf{k}}^{(0)}|(T+v_{ext})^{\tau_{\kappa,\beta}}|u_{n\mathbf{k}}^{(0)}\rangle \ , \tag{3.87}$$

and taking an additional derivative with respect to electric field yields

$$Z^*_{\kappa,\alpha\beta} = \frac{2f}{N_k}\sum_{\mathbf{k}}\sum_{n=1}^{occ}\langle u_{n\mathbf{k}}^{(0)}|(T+v_{ext})^{\tau_{\kappa,\beta}}|u_{n\mathbf{k}}^{\mathcal{E}_\alpha}\rangle \ . \tag{3.88}$$

This has essentially the same form as Eq. (43) in Ref. [9], except that here the zero-order wave functions are already polarized by the preexisting finite electric field.

Alternatively, Eq. (3.85) can be computed as the derivative of the polarization with respect to the displacement,

$$Z^*_{\kappa,\alpha\beta} = \Omega \frac{\partial P_\alpha}{\partial \tau_{\kappa,\beta}} = \Omega \hat{\mathbf{e}}_\alpha \cdot \mathbf{P}^{\tau_{\kappa,\beta}} \ . \tag{3.89}$$

Here $\mathbf{P}^{\tau_{\kappa,\beta}}$ takes a form very similar to that of Eq. (3.70), except that the first-order changes $|u^{\mathcal{E}_\alpha}_{n\mathbf{k}_j}\rangle$ in the wave functions in response to an electric field are replaced by the corresponding changes $|u^{\tau_{\kappa,\beta}}_{n\mathbf{k}_j}\rangle$ in response to a sublattice displacement. The computation of the $|u^{\tau_{\kappa,\beta}}_{n\mathbf{k}_j}\rangle$ has already been described in detail in Sec. 3.2.

The computation of the first-order derivatives of the wave functions is typically the most time-consuming step of the linear-response calculation. Therefore, for a complicated unit cell with many atoms M per cell, the computation of the three derivatives $|u^{\mathcal{E}_\alpha}\rangle$ will be much cheaper than that of the $3M$ derivatives $|u^{\tau_{\kappa,\beta}}\rangle$, and the method of Eq. (3.88) will therefore be significantly faster than the method of Eq. (3.89). In the special case that the displacement derivatives $|u^{\tau_{\kappa,\beta}}\rangle$ have already been computed for some other reason (e.g., for the purpose of computing the phonon frequencies in finite field), the use of the latter method may be advantageous. In any case, a comparison of the two methods should provide a useful check on the internal consistency of the theory and its computational implementation.

3.5 Test calculations for III-V semiconductors

3.5.1 Phonon frequencies

In order to test our method, we have carried out calculations of the frequency shifts induced by electric fields in two III-V semiconductors, AlAs and GaAs. We have chosen these two materials because they are well-studied systems both experimentally and theoretically, and because the symmetry allows some phonon mode frequencies to shift linearly with electric field while others shift quadratically. Since our main purpose is to check the internal consistency of our theoretical approach, we focus on making comparisons between the shifts calculated using our new linear-response method and those calculated using standard finite-difference methods. Moreover, as mentioned at the start of Sec. 3.1.1, we have chosen to neglect changes in phonon frequencies that enter through the electric-field induced strains (piezoelectric and electrostrictive effects), and we do this consistently in both the linear-response and finite-difference calculations. For this reason, our results are not immediately suitable for comparison with experimental measurements.

Our calculations are carried out using a plane-wave pseudopotential approach to density-functional theory. We use the ABINIT code package [48], which incorporates the finite electric field method of Souza *et al.* [33] for the ground-state and frozen-phonon calculations in finite electric field. We then carried out the linear-response calculations with a version of the code that we have modified to implement the linear-response formulas of the previous section.

The details of the calculations are as follows. We use Troullier-Martins norm-conserving pseudopotentials [49], the Teter Pade parameterization [50] of the local-density approximation, and a plane-wave cutoff of 16 Hartree. A $10\times10\times10$ Monkhorst-Pack [51] k-point sampling was used, and we chose lattice constants of 10.62 Å and 10.30 Å for AlAs and GaAs, respectively. The crystals are oriented so that the vector $(a/2)(1,1,1)$ points from a Ga or Al atom to an As atom.

Table 3.1: Calculated frequency shifts, in cm^{-1}, induced by an electric field of 5.14×10^8 V/m applied along x in GaAs and AlAs (This field is larger than the breakdown field of GaAs 1.8×10^7 V/m). 'FD' are the results of finite-difference (frozen-phonon) calculations in which atoms are displaced by hand and restoring forces are calculated, while 'LR' refers to the use of the linear-response developed here. The L and X points are at $(2\pi/a)(1,1,1)$ and $(2\pi/a)(1,0,0)$ respectively.

	GaAs		AlAs	
Mode	FD	LR	FD	LR
Γ O1 [1]	-3.856	-3.856	-5.941	-5.941
Γ O2 [1]	-0.282	-0.281	-0.300	-0.299
Γ O3 [1]	3.548	3.548	5.647	5.647
L LO	2.701	2.703	4.282	4.282
L TO1	-3.749	-3.749	-5.663	-5.663
L TO2	0.567	0.564	0.952	0.952
X LO	0.050	0.050	-0.243	-0.243
X TO1	-3.953	-3.953	-6.083	-6.083
X TO2	3.753	3.753	5.919	5.919

Table 3.1 shows the changes in phonon frequencies resulting from an electric field applied along a Cartesian direction at several high-symmetry q-points in GaAs and AlAs. Both the electronic and ionic contributions, Eqs. (3.3-3.4), are included. We first relaxed the atomic coordinates in the finite electric field until the maximum force on any atom was less than 10^{-6} Hartree/Bohr. We then carried out the linear-response calculation, and in addition, to check the internal consistency of our linear-response method, we carried out a corresponding calculation using a finite-difference frozen-phonon approach. For the latter, the atoms were displaced according to the normal modes obtained from our linear-response calculation, with the largest displacement being 0.0025 Bohr. (Because the electric field lowers the symmetry, the symmetry-reduced set of k-points is not the same as in the absence of the electric field.) The agreement between the finite-different approach and the new linear-response implementation can be seen to be excellent, with the small differences visible for some modes being attributable to truncation in the finite-difference formula and the finite density of the k-point mesh.

In Table 3.2, we decompose the frequency shifts into the ionic contribution $\Delta\omega_{\text{ion}}(\mathbf{q};\mathcal{E})$ and the

electronic contribution $\Delta\omega_{\rm el}(\mathbf{q};\mathcal{E})$ defined by Eqs. (3.4) and (3.3), respectively, calculated using the linear-response approach. It is clear that the largest contributions are ionic in origin. For example, the large, roughly equal and opposite shifts of the O1 and O3 modes at Γ arise from the ionic terms. However, there are special cases (e.g., O2 at Γ and LO at X) for which the ionic contribution happens to be small, so that the electronic contribution is comparable in magnitude.

The pattern of ionic splittings appearing at Γ can be understood as follows. Because the non-analytic long-range Coulomb contribution is not included, the three optical modes at Γ are initially degenerate with frequency ω_0 in the unperturbed lattice. A first-order electric field along x induces a first-order relative displacement u_x of the two sublattices, also along x. By symmetry considerations, the perturbed dynamical matrix is given, up to quadratic order in u_x, as

$$D(\Gamma) = \omega_0^2 \begin{pmatrix} 1+\mu u_x^2 & 0 & 0 \\ 0 & 1+\nu u_x^2 & \kappa u_x \\ 0 & \kappa u_x & 1+\nu u_x^2 \end{pmatrix} . \tag{3.90}$$

The off-diagonal κ term arises from the E_{xyz} coupling in the expansion of the total energy in displacements; this is the only third-order term allowed by symmetry. The μ and ν terms arise from fourth-order couplings of the form E_{xxxx} and E_{xxyy} respectively. The eigenvalues of this matrix are proportional to $1+\mu u_x^2$ and $1 \pm \kappa u_x + \nu u_x^2$. Thus, two of the modes should be perturbed at first order in the field-induced displacements with a pattern of equal and opposite frequency shifts, while all three modes should have smaller shifts arising from the quadratic terms. This is just what is observed in the pattern of frequency shifts shown in Table 3.2. (The symmetry of the pattern of electronic splittings is the same, but it turns out that the linear shift is much smaller in this case, so that for the chosen electric field, the linear and quadratic contributions to the electronic frequency shift have similar magnitudes.) A similar analysis can be used to understand the patterns of frequency shifts at the L and X points.

Table 3.2: Same as in Table 3.1, but with the frequency shifts decomposed into ionic and electronic contributions as defined in Eqs. (3.4) and (3.3) respectively.

	GaAs		AlAs	
	Ion	Elec.	Ion	Elec.
Γ O1 [1]	−3.659	−0.198	−5.684	−0.257
Γ O2 [1]	−0.146	−0.135	−0.123	−0.177
Γ O3 [1]	3.655	−0.107	5.589	0.058
L LO	2.341	0.362	3.633	0.649
L TO1	−3.486	−0.262	−5.628	−0.034
L TO2	1.181	−0.617	1.658	−0.707
X LO	0.122	−0.073	−0.033	−0.209
X TO1	−3.411	−0.543	−5.658	−0.424
X TO2	3.388	0.365	5.609	0.310

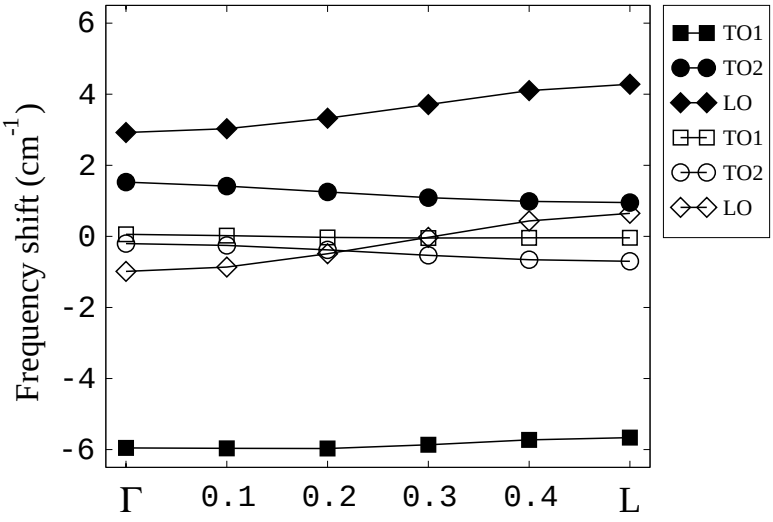

Figure 3.2: Frequency shifts induced by an electric field of 5.14×10^8 V/m along x in AlAs, plotted along Γ to L. Filled and open symbols indicate the total shift $\Delta\omega_{\rm el} + \Delta\omega_{\rm ion}$ and the electronic contribution $\Delta\omega_{\rm el}$ respectively.

We have also plotted, in Fig. 3.2, the calculated total frequency shift $\Delta\omega_{\rm el}(\mathbf{q}) + \Delta\omega_{\rm ion}(\mathbf{q})$ and its electronic contribution $\Delta\omega_{\rm el}(\mathbf{q})$ along the line from Γ to L for the case of AlAs. (The 'LO' and 'TO' symmetry labels are not strictly appropriate here because the electric field along x mixes the mode eigenvectors; the notation indicates the mode that would be arrived at by turning off the field.) In contrast to the results presented in Tables 3.1-3.2, the frequencies at Γ in Fig. 3.2 were computed by including the long-range non-analytic Coulomb contribution for $\hat{q} \parallel (111)$ in order to extend the curves to $q = 0$. (Because the direct linear-response calculation of the dynamical effective charge and dielectric susceptibility tensors had not yet been developed and implemented in the presence of a finite electric field, the needed tensor elements were computed by finite differences.) It is clearly evident that the electronic terms remain much smaller than the ionic ones for all three optical modes over the entire branch in q-space.

Returning now to the comparison between our exact theory of Sec. 3.1.1 and the approximate theory of Sec. 3.1.1, we compare the equilibrium positions and phonon frequencies predicted by these theories in Table 3.3. Recall that $R_{\mathcal{E}}$ is calculated in the approximate theory by using Eq. (3.6). Using this force, the ion coordinates were again relaxed to a tolerance of 10^{-6} (Hartree/Bohr) on the forces. It can be seen that $R_{\mathcal{E}}$ is predicted quite well by the approximate theory, with errors of only $\sim$2%, confirming that the displacements can be calculated to good accuracy using a linearized

Table 3.3: Comparison of ionic displacements and frequency shifts at the L point in GaAs as computed by the approximate and exact approaches of Sec. 3.1.1 and 3.1.1 respectively, again for an electric field of 5.14×10^8 V/m along x. $R_{\mathcal{E}}$ is the induced displacement of the cation sublattice along x, and the $\Delta\omega_{\rm ion}$ are ionic contributions to the frequency shifts as defined in Eq. (3.4).

		$R_{\mathcal{E}}$	$\Delta\omega_{\rm ion}$(L) (cm^{-1})		
		(10^{-3} Å)	LO	TO1	TO2
GaAs	Approx.	5.07	2.63	-3.89	1.37
	Exact	4.95	2.34	-3.49	1.18
AlAs	Approx.	5.69	3.75	-5.66	1.65
	Exact	5.62	3.63	-5.63	1.66

theory for this magnitude of electric field. The changes in the phonon frequencies resulting from these displacements (evaluated at zero and non-zero field for the approximate and exact theories respectively) are listed in the remaining columns of Table 3.3. The discrepancies in the phonon frequencies are now somewhat larger, approaching 15% in some cases. This indicates that the approximate theory is able to give a moderately good description of the phonon frequency shifts of GaAs in this field range, but the exact theory is needed for accurate predictions. (Also, recall that the approximate theory does not provide any estimate for the electronic contributions, which are not included in Table 3.3.)

Finally, we illustrate our ability to calculate the nonlinear field dependence of the phonon frequencies by presenting the calculated optical L-point phonon frequencies of AlAs in Fig. 3.3 as a function of electric field along x. These are again the results of our exact theory, obtained by including both ionic and electronic contributions. The two TO modes are degenerate at zero field, as they should be. All three modes show a linear component that dominates their behavior in this range of fields. However, a quadratic component is also clearly evident, illustrating the ability of the present approach to describe such nonlinear behavior.

3.5.2 Born effective charge and dielectric constant

In order to check our method, we have performed test calculations on two prototypical III-V semiconductors, AlAs and GaAs, for which the electronic contribution to the polarization is typically comparable to the ionic contribution [43]. The calculation is carried out using the planewave-pseudopotential method based on density-functional theory with local-density approximation (LDA). We use Troullier-Martins norm-conserving pseudopotentials [49] in which the $3d$ states on the Ga and As atoms are treated as core states. (The omission of the semicore $3d$ states from the valence on the Ga atom may limit the accuracy of the Ga pseudopotential somewhat.) A $16 \times 16 \times 16$

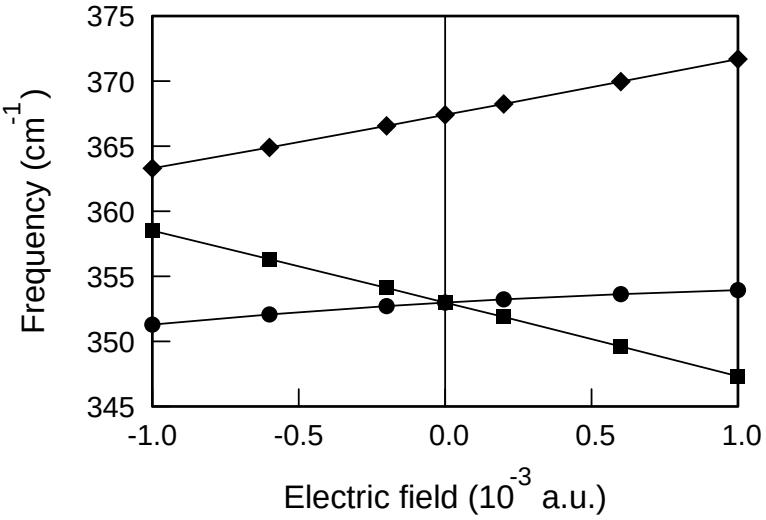

Figure 3.3: Frequencies of LO and TO modes at L in AlAs as a function of electric field (where 10^{-3} a.u. = 5.14×10^{8} V/m) applied along x. The symbols have the same interpretation as in Fig. 3.2.

Table 3.4: Calculated electronic dielectric constants of AlAs and GaAs at zero field, and changes resulting from an electric field of 3.08×10^{8} V/m along the [100] direction. 'LR' and 'FD' denote the results of linear-response [Eq. (3.84)] and finite-difference calculations, respectively.

		ϵ_∞	$\Delta\epsilon_{\infty,23}$	$\Delta\epsilon_{\infty,11}$	$\Delta\epsilon_{\infty,33}$
AlAs	LR	9.681	0.039	0.027	0.013
	FD	9.681	0.040	0.027	0.013
GaAs	LR	13.315	0.202	0.211	0.104
	FD	13.319	0.203	0.207	0.098

Monkhorst-Pack mesh is used for the k-point sampling. More computational details can be found in Sec. 3.2.

The calculation of the dielectric permittivity tensor and the Born effective charge tensor is carried out in three steps. First, a ground-state calculation at finite electric field is performed using the Berry-phase approach [33] implemented in the `ABINIT` code, and the field-polarized Bloch functions are stored for the later linear response calculation. Second, the linear response calculation is carried out to obtain the first-order response of Bloch functions. Third, the matrix elements of the dielectric and Born effective charge tensors are computed using these first-order responses.

The first column of Table 3.4 shows the calculated electronic dielectric constants of AlAs and GaAs at zero electric field, and the remaining ones show the nonzero changes in the dielectric tensor elements after the application of an electric field $\mathcal{E}^{(0)}$ of 3.08×10^8 V/m along the [100] direction. The results obtained with the linear-response approach of Eq. (3.84) are compared with those calculated by finite differences. In the latter case, polarizations are computed at several values of the electric field in steps of 3.08×10^5 V/m, and the dielectric tensor is calculated using a finite-difference version of Eq. (3.84). It can be seen that the agreement between the linear-response and the finite-difference results is excellent, demonstrating the internal consistency between the two approaches.

In Table 3.5 we present similar results for the cation Born effective charges of the same two materials, first at zero field and then again under application of a field of $\mathcal{E}^{(0)}$ of 3.08×10^8 V/m along the [100] direction. The linear-response results were obtained using Eq. (3.88), but we also computed the corresponding values using Eq. (3.89) and found agreement between the two linear-response approaches with a maximum fractional error smaller than 10^{-6} for all values reported. For the finite-difference comparison, the polarizations were computed at several values of the atomic displacements in steps of 10^{-3} Bohr and the Born charge tensors were calculated using a finite-difference version of Eq. (3.89). It can again be seen the agreement between the linear-response and the finite-difference results is excellent.

We emphasize that the values of $\Delta\epsilon_\infty$ and ΔZ^* reported in Tables 3.4 and 3.5 are purely electronic or "frozen-ion" ones – that is, the sublattice displacements that would be induced by a truly static electric field $\mathcal{E}^{(0)}$ are not included. The results with ionic relaxations are presented in Table 3.6. It is evident that ionic relaxations have neglegible effects on the diagonal elements of dielectric tensors but moderate effects on the off-diagonal elements. For Born effective charge tensors, the effects of ionic relaxations are obvious for both diagonal and off-diagonal elements.

Table 3.5: Calculated cation Born effective charges of AlAs and GaAs at zero field, and changes resulting from an electric field of 3.08×10^8 V/m along the [100] direction. 'LR' and 'FD' denote the results of linear-response [Eq. (3.88)] and finite-difference calculations, respectively.

		Z^*	ΔZ^*_{23} $(\times 10^{-3})$	ΔZ^*_{11} $(\times 10^{-3})$	ΔZ^*_{33} $(\times 10^{-3})$
AlAs	LR	2.110	17.23	-0.06	-0.13
	FD	2.110	17.22	-0.05	-0.11
GaAs	LR	2.186	52.88	-3.42	-3.17
	FD	2.186	52.83	-3.36	-3.14

Table 3.6: Calculated changes in electronic dielectric constants and cation Born effective charges of AlAs and GaAs resulting from an electric field of 3.08 $\times 10^8$ V/m along the [100] direction. Here ions are fully *relaxed* to the electric field.

	AlAs	GaAs
$\Delta\epsilon_{\infty,23}$	0.024	0.145
$\Delta\epsilon_{\infty,11}$	0.027	0.209
$\Delta\epsilon_{\infty,33}$	0.013	0.101
$\Delta Z^*_{23}(\times 10^{-3})$	15.97	53.16
$\Delta Z^*_{11}(\times 10^{-3})$	-1.405	-1.653
$\Delta Z^*_{33}(\times 10^{-3})$	-0.431	-1.836

The values of ϵ_∞ and Z^* reported in Tables 3.4 and 3.5 are in good agreement with other theoretical values in the literature [52, 53, 54] and with experiment. The symmetry is such that the applied electric field along x breaks the degeneracy between the diagonal elements of the ϵ_∞ and Z^* tensors so that $\epsilon_{\infty,11} \neq \epsilon_{\infty,22} = \epsilon_{\infty,33}$ and $Z^*_{11} \neq Z^*_{22} = Z^*_{33}$, and introduces non-zero off-diagonal elements $\epsilon_{\infty,23} = \epsilon_{\infty,32}$ and $Z^*_{23} = Z^*_{32}$.

Symmetry considerations also imply that $\epsilon_{\infty,23}$ and Z^*_{23} should appear to first order in $\mathcal{E}^{(0)}$, while $\Delta\epsilon_{\infty,11}$, $\Delta\epsilon_{\infty,33}$, ΔZ^*_{11}, and ΔZ^*_{33} should be quadratic in $\mathcal{E}^{(0)}$. This is confirmed by our numerical calculations. Indeed, by repeating calculations like those shown in Tables 3.4 and 3.5 for several values of $\mathcal{E}^{(0)}$ and fitting to obtain the coefficients of the linear and quadratic dependence, we can extract information about the nonlinear dielectric response and the Raman tensor. The second-

Table 3.7: Values of second-order dielectric susceptibility and Raman matrix elements in AlAs, as defined by Eqs. (3.91) and (3.93) respectively, compared with previous theory and experiment.

	$\chi^{(2)}_{123}$ (pm/V)	$\lvert\alpha_{\rm TO}\rvert$ (Å^2)
Present work	62	8.0
Theory,[1] Ref. [33]	64	
Theory,[2] Ref. [55]	70	8.5
Theory,[1] Ref. [56]	79	9.0
Theory,[2] Ref. [57]		7.4
Experiment, Ref [58]	78$\pm$20	

and third-order nonlinear dielectric tensors are defined as

$$\chi^{(2)}_{123} = \frac{1}{2}\frac{\partial^2 P_2}{\partial\mathcal{E}_1\partial\mathcal{E}_3} = \frac{1}{2}\frac{\partial\chi_{23}}{\partial\mathcal{E}_1} \tag{3.91}$$

and

$$\chi^{(3)}_{1111} = \frac{1}{6}\frac{\partial^3 P_1}{\partial\mathcal{E}_1^3} = \frac{1}{6}\frac{\partial^2\chi_{11}}{\partial\mathcal{E}_1^2}\ , \tag{3.92}$$

while the Raman polarizability tensor is defined by

$$\alpha_{\rm TO} = \frac{\partial^2 f_2}{\partial\mathcal{E}_1\partial\mathcal{E}_3} = \frac{\partial Z_{23}}{\partial\mathcal{E}_1} \tag{3.93}$$

where $\mathbf{f}$ is the force on the cation sublattice induced by the electric field. In practice, we calculate χ_{23}, χ_{11}, and Z^*_{23} for a series of finite electric fields oriented along the x-axis with values of $\mathcal{E}^{(0)}$ ranging from zero to 5.14×10^8V/m in increments of one-fifth of the maximum value. Fitting these data to a polynomial in $\mathcal{E}^{(0)}$ then gives the values of $\chi^{(2)}_{123}$, $\chi^{(3)}_{1111}$, and $\alpha_{\rm TO}$. Note that $\alpha_{\rm TO}$ can alternatively be expressed as

$$\alpha_{\rm TO} = \Omega\frac{\partial\chi_{23}}{\partial\tau_1} \tag{3.94}$$

where τ_1 is a cation sublattice displacement and χ_{23} is computed at zero field. We have also computed $\alpha_{\rm TO}$ by fitting to a series of calculations of this type, and find values of $\alpha_{\rm TO}$ that agree with those obtained from Eq. (3.93) within 0.3%.

The results for the $\chi^{(2)}_{123}$ and $\alpha_{\rm TO}$ values as computed from Eqs. (3.91) and (3.93) are presented in Table 3.7 for AlAs, together with some previous theoretical and experimental values for comparison. In view of the fact that the calculation of higher-order tensor elements tends to be delicate, the agreement is generally quite good. In particular, Veithen et al. [55] have shown (see their Fig. 1) that the results for $\chi^{(2)}_{123}$ can be quite sensitive to the method of discretization in k-space and the fineness of the k-point mesh. For GaAs we find $\chi^{(2)}_{123} = 293$ pm/V and $\alpha_{\rm TO} = -24.1$ Å^2 (which is close to the value in Ref. [57]), but these numbers are of questionable accuracy because of our use of a Ga pseudopotential that does not include the $3d$ semicore orbitals in the valence. We obtain $\chi^{(3)}_{1111}$ values of 3.90 and 33.8$\times 10^{-11}$ esu for AlAs and GaAs, respectively. We are not aware of previous theoretical values of $\chi^{(3)}_{1111}$ with which to compare; this quantity is beyond the reach of the "$2n + 1$" theorem using first-order wave function responses only, and so is difficult to compute by pure DFPT methods. Experimental values ranging from 3.9 to 18$\times 10^{-11}$ esu for GaAs [59] can be found in the literature.

The discrepancies noted above between theory and theory, and between theory and experiment, may have many possible causes. In addition to some of the computational and convergence issues mentioned above, the adequacy of the LDA approximation itself is also a serious question. Because the LDA tends to underestimate gaps, some authors have included a so-called "scissors correction" in order to widen the gap artificially; this tends to decrease the magnitude of response tensors [60]. On the experimental side, the difficulty in obtaining reproducible results is surely also an issue. Nevertheless, we emphasize that the *relative* accuracy of the values reported in Tables I and II, which were done under the *same* computational conditions (same pseudopotentials, k-point meshes, etc.), demonstrates the correctness of our new finite-field linear-response formulation and the internal consistency of the computational framework that we employ.

3.6 Summary and discussion

We have developed a method for computing the phonon frequencies of an insulator in the presence of a homogeneous, static electric field. The extension of density-functional perturbation theory to this case has been accomplished by carrying out a careful expansion of the field-dependent energy functional $E_{\rm KS} + \Omega\mathcal{E} \cdot \mathbf{P}$, where $\mathbf{P}$ is the Berry-phase polarization, with respect to phonon modes both at $q = 0$ and at arbitrary q. In the general case of nonzero q, there is a subtle interplay between the couplings between neighboring k-points introduced by the electric field and the further-neighbor couplings introduced by the q-vector, so that terms arise that require the evaluation of four-sided loops in k-space. However, with the judicious use of time-reversal symmetry, the needed evaluations can be reduced to a form that is not difficult to implement in an existing DFPT code.

We have carried out test calculations on two III-V semiconductors, AlAs and GaAs, in order to test the correctness of our implementation. A comparison of the results of linear-response and finite-difference calculations shows excellent agreement, thus validating our approach. We also decompose the frequency shifts into "lattice" and "electronic" contributions and quantify these, and we find that the lattice contributions (i.e., those resulting from induced displacements in the reference equilibrium structure) are usually, but not always, dominant. We also evaluated the accuracy of an approximate method for computing the lattice contribution, in which only zero-field inputs are needed. We found that this approximate approach gives a good rough description, but that the full method is needed for an accurate calculation.

Our linear-response method has the same advantages, relative to the finite-difference approach, as in zero electric field. Even for a phonon at Γ, our approach is more direct and simplifies the calculation of the phonon frequencies. However, its real advantage is realized for phonons at arbitrary q,

because the frequency can still be obtained efficiently from a calculation on a single unit cell without the need for imposing commensurability of the q-vector and computing the mode frequencies for the corresponding supercell. We also emphasize that the method is not limited to infinitesimal electric fields. We thus expect the method will prove broadly useful for the study of linear and nonlinear effects of electric bias on the lattice vibrational properties of insulating materials.

We have also developed a linear-response method for computing dielectric constants and Born effective charges in the presence of a *finite* electric field. We have demonstrated the reliability of our approach by implementing it in the context of the `ABINIT` code package[48] and performing test calculations on two III-V semiconductors, AlAs and GaAs. We have confirmed that the results calculated using the new linear-response approach are consistent with those obtained from finite-difference calculations carried out within the same framework. In general, our results are also in good agreement with other theoretical calculations and with experiment.

A major advantage of the present approach is that, unlike the conventional long-wave linear-response method,[8] it can be applied to obtain response tensors in finite electric field. While it is possible to obtain similar information from a set of finite-difference calculations carried out for some chosen set of applied electric fields, the linear-response approach is more direct, and it avoids the troublesome truncation errors that may arise in a finite-difference approach. In the future, it may be of interest to extend the finite-field DFPT treatment not just to phonon perturbations and electric-field perturbations , but also to other perturbations such as those associated with strain or chemical composition. Taken together, these developments should allow for much greater flexibility in the calculation of materials properties of insulators under electrical bias and facilitate the study of higher-order nonlinear dielectric properties.

Chapter 4

Ab-initio calculation of the anomalous Hall conductivity by Wannier interpolation

The Hall resistivity of a ferromagnet depends not only on the magnetic induction, but also on the magnetization; the latter dependence is known as the anomalous Hall effect (AHE) [61]. The AHE is used for investigating surface magnetism, and its potential for investigating nanoscale magnetism, as well as for magnetic sensors and memory devices applications, is being considered [62]. Theoretical investigations of the AHE have undergone a revival in recent years, and have also led to the proposal for a spin counterpart, the spin Hall effect, which has subsequently been realized experimentally.

The first theoretical model of the AHE was put forth by Karplus and Luttinger [16], who showed that it can arise in a perfect crystal as a result of the spin-orbit interaction of polarized conduction electrons. Later, two alternative mechanisms, skew scattering [63] and side jump scattering [18], were proposed by Smit and Berger respectively. In skew scattering the spin-orbit interaction gives rise to an asymmetric scattering cross section even if the defect potential is symmetric, and in side-jump scattering it causes the scattered electron to acquire an extra transverse translation after the scattering event. These two mechanisms involve scattering from impurities or phonons, while the Karplus-Luttinger contribution is a scattering-free bandstructure effect. The different contributions to the AHE are critically reviewed in Ref. [64]. Perhaps because an intuitive physical picture was lacking, the Karplus-Luttinger theory was strongly disputed in the early literature. Attempts at estimating its magnitude on the basis of realistic bandstructure calculations were also rare [65].

In recent years, new insights into the Karplus-Luttinger contribution have been obtained by several authors [66, 67, 68, 15, 69], who reexamined it in the modern language of Berry's phases.

The term $\mathbf{\Omega}_n(\mathbf{k})$ in the equations below was recognized as the Berry curvature of the Bloch states in reciprocal space, a quantity which had previously appeared in the theory of the integer quantum Hall effect [22], and also closely related to the Berry-phase theory of polarization [20]. The dc anomalous Hall conductivity (AHC) is simply given as the Brillouin zone (BZ) integral of the Berry curvature weighted by the occupation factor of each state,

$$\sigma_{xy} = -\frac{e^2}{\hbar} \sum_n \int_{\mathrm{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} \, f_n(\mathbf{k}) \, \Omega_{n,z}(\mathbf{k}) \ , \tag{4.1}$$

where $\sigma_{xy} = -\sigma_{yx}$ is the antisymmetric part of the conductivity. While this can be derived in several ways, it is perhaps most intuitively understood from the semiclassical point of view, in which the group velocity of an electron wavepacket in band n is [70, 67]

$$\dot{\mathbf{r}} = \frac{1}{\hbar} \frac{\partial \mathcal{E}_{n\mathbf{k}}}{\partial \mathbf{k}} - \dot{\mathbf{k}} \times \mathbf{\Omega}_n(\mathbf{k}) \ . \tag{4.2}$$

The second term, often overlooked in elementary textbook derivations, is known as the "anomalous velocity." The expression for the current density then acquires a new term $e f_n(\mathbf{k}) \, \dot{\mathbf{k}} \times \mathbf{\Omega}_n(\mathbf{k})$ which, with $\dot{\mathbf{k}} = -e\mathbf{E}/\hbar$, leads to Eq. (4.1).

Recently, first-principles calculations of Eq. (4.1) were carried out for the ferromagnetic perovskite $SrRuO_3$ by Fang *et al.* [13], and for a transition metal, bcc Fe, by Yao *et al.* [21] In both cases the calculated values compared well with experimental data, lending credibility to the intrinsic mechanism. The most striking feature of these calculations is the strong and rapid variation of the Berry curvature in k-space. In particular, there are sharp peaks and valleys at places where two energy bands are split by the spin-orbit coupling across the Fermi level. In order to converge the integral, the Berry curvature has to be evaluated over millions of k-points in the Brillouin zone. In the previous work this was done via a Kubo formula involving a large number of unoccupied states; the computational cost was very high, even for bcc Fe, with only one atom in the unit cell.

In this chapter, we present an efficient method for computing the intrinsic AHC. Unlike the

conventional approach, it does not require carrying out a full *ab-initio* calculation for every k-point where the Berry curvature needs to be evaluated. The actual *ab-initio* calculation is performed on a much coarser k-point grid. By a post-processing step, the resulting Bloch states below and immediately above the Fermi level are then mapped onto well-localized Wannier-functions. In this representation it is then possible to interpolate the Berry curvature onto any desired k-point with very little computational effort and essentially no loss of accuracy.

4.1 Definitions and background

The key ingredient in the theory of the intrinsic anomalous Hall effect is the Berry curvature $\mathbf{\Omega}_n(\mathbf{k})$, defined as

$$\mathbf{\Omega}_n(\mathbf{k}) = \boldsymbol{\nabla} \times \mathbf{A}_n(\mathbf{k}) \ , \tag{4.3}$$

where $\mathbf{A}_n$ is the Berry connection,

$$\mathbf{A}_n(\mathbf{k}) = i\langle u_{n\mathbf{k}} | \boldsymbol{\nabla}_{\mathbf{k}} | u_{n\mathbf{k}} \rangle \ . \tag{4.4}$$

The integral of the Berry curvature over a surface bounded by a closed path in k-space is the Berry phase of that path [19]. In what follows it will be useful to write the Berry curvature as a second-rank antisymmetric tensor:

$$\Omega_{n,\gamma}(\mathbf{k}) = \epsilon_{\alpha\beta\gamma}\, \Omega_{n,\alpha\beta}(\mathbf{k}) \ , \tag{4.5}$$

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im} \left\langle \frac{\partial u_{n\mathbf{k}}}{\partial k_\alpha} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_\beta} \right\rangle , \tag{4.6}$$

where the Greek letters indicate Cartesian coordinates, $\epsilon_{\alpha\beta\gamma}$ is Levi-Civita tensor and $u_{n\mathbf{k}}$ are the cell-periodic Bloch functions.

With this notation we rewrite the quantity we wish to evaluate, Eq. (4.1), as

$$\sigma_{\alpha\beta} = -\frac{e^2}{\hbar} \int_{\mathrm{BZ}} \frac{d\mathbf{k}}{(2\pi)^3}\, \Omega_{\alpha\beta}(\mathbf{k}) \ , \tag{4.7}$$

where we have introduced the *total* Berry curvature

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_n f_n(\mathbf{k})\,\Omega_{n,\alpha\beta}(\mathbf{k}). \tag{4.8}$$

Direct evaluation of Eq. (4.6) poses a number of practical difficulties related to the presence of k-derivatives of Bloch states, as will be discussed in the next section. In previous work [13, 21] these were circumvented by recasting Eq. (4.6) as a Kubo formula [65, 22], where the k-derivatives are replaced by sums over states:

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2\mathrm{Im}\sum_{m\neq n}\frac{v_{nm,\alpha}(\mathbf{k})\,v_{mn,\beta}(\mathbf{k})}{(\omega_m(\mathbf{k})-\omega_n(\mathbf{k}))^2}\,, \tag{4.9}$$

where $\omega_n(\mathbf{k}) = \mathcal{E}_{n\mathbf{k}}/\hbar$ and the matrix elements of the Cartesian velocity operators $\hat{v}_\alpha = (i/\hbar)[\hat{H},\hat{r}_\alpha]$ are given by [71]

$$v_{nm,\alpha}(\mathbf{k}) = \langle\psi_{n\mathbf{k}}|\hat{v}_\alpha|\psi_{m\mathbf{k}}\rangle = \frac{1}{\hbar}\left\langle u_{n\mathbf{k}}\middle|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_\alpha}\middle|u_{m\mathbf{k}}\right\rangle\,, \tag{4.10}$$

where $\hat{H}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\hat{\mathbf{r}}}\hat{H}e^{i\mathbf{k}\cdot\hat{\mathbf{r}}}$. The merit of Eq. (4.9) lies in its practical implementation on a finite k-grid using only the wavefunctions at a single k-point. As is usually the case for such linear-response formulas, sums over pairs of occupied states can be avoided in the $T = 0$ version of Eqs. (4.8–4.9) for the total Berry curvature,

$$\Omega_{\alpha\beta}(\mathbf{k}) = -2\mathrm{Im}\sum_v\sum_c\frac{v_{vc,\alpha}(\mathbf{k})\,v_{cv,\beta}(\mathbf{k})}{(\omega_c(\mathbf{k})-\omega_v(\mathbf{k}))^2}\,, \tag{4.11}$$

where v and c subscripts denote valence (occupied) and conduction (unoccupied) bands, respectively. However, the evaluation of this formula requires the cumbersome summation over unoccupied states. Even if practical calculations truncate the summation to some extent, the computation could be time-consuming. Moreover, the time required to calculate the matrix elements of the velocity operator in Eq. (4.9) or (4.11) is not negligible.

4.2 Evaluation of the Berry curvature by Wannier interpolation

In view of the above-mentioned drawbacks of the Kubo formula for practical calculations, it would be highly desirable to have a numerical scheme based on the the "geometric formula" (4.6), in terms of the occupied states only. The difficulties in implementing that formula arise from the k-derivatives therein. Since in practice one always replaces the Brillouin zone integration by a discrete summation, an obvious approach would be to use a finite-difference representation of the derivatives on the k-point grid. However, this requires some care: a straightforward discretization will yield results which depend on the choice of phases of the Bloch states, even though Eq. (4.6) is in principle invariant under such "diagonal gauge transformations." The problem becomes more acute in the presence of band crossings and avoided crossings, because then it is not clear which two states at neighboring grid points should be taken as "partners" in a finite-differences expression. (Moreover, since the system is a metal, at $T = 0$ the occupation can be different at neighboring k-points.) Successful numerical strategies for dealing with problems of this nature have been developed in the context of the Berry-phase theory of polarization of insulators, and a workable finite-difference scheme which combines those ideas with Wannier interpolation is sketched in Appendix B.

We present here a different, more powerful strategy that also relies on a Wannier representation of the low-energy electronic structure. We will show that it is possible to express the needed derivatives analytically in terms of the Wannier functions, so that no finite-difference evaluation of a derivative is needed in principle. The use of Wannier functions allows us to achieve this while still avoiding the summation over all empty states which appears in the Kubo formula as a result of applying conventional $\mathbf{k} \cdot \mathbf{p}$ perturbation theory.

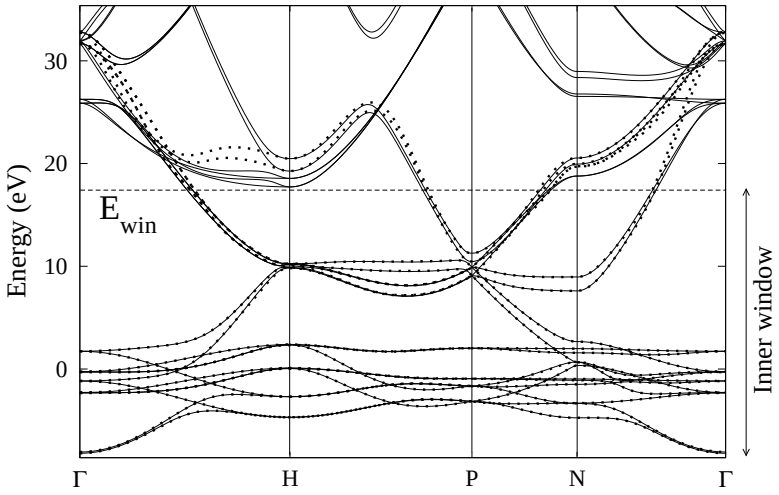

Figure 4.1: Band structure of bcc Fe with spin-orbit coupling included. Solid lines: original band structure from a conventional first-principles calculation. Dotted lines: Wannier-interpolated band structure. The zero of energy is the Fermi level.

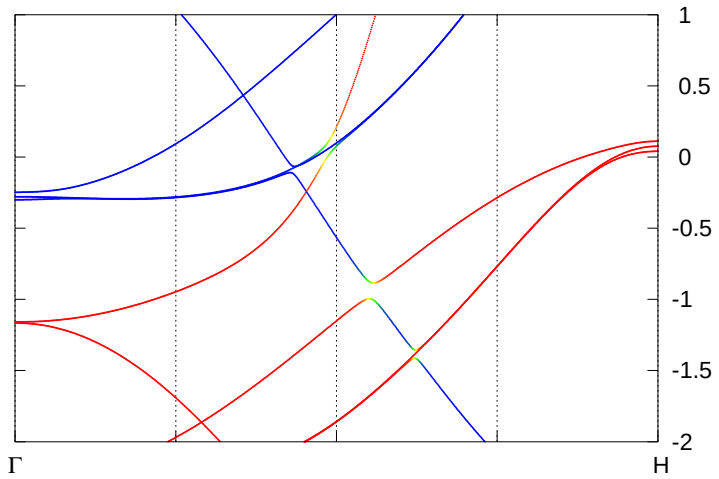

Figure 4.2: Wannier-interpolated bands of bcc Fe along Γ-H. The bands are colorcoded according to the value of the spin projection $\langle S_z \rangle$: red for spin-up and blue for spin-down. The energies are given in eV and the Fermi level is at 0 eV. The vertical dashed lines indicate k points on the ab initio mesh used for constructing the WFs.

4.2.1 Wannier representation

We begin by using the approach of Souza, Marzari, and Vanderbilt [72] to construct a set of Wannier functions (WFs) for the metallic system of interest. For insulators, one normally considers a set of WFs that span precisely the space of occupied Bloch states. Here, since we have a metallic system and we want to have well-localized WFs, we choose a number of WFs larger than the number $N_{\mathbf{k}}$ of occupied states at any $\mathbf{k}$, and only insist that the space spanned by the WFs should include, as a subset, the space of the occupied states, plus the first few empty states. Thus, these partially-occupied WFs will serve here as a kind of "exact tight-binding basis" that can be used as a compact representation of the low-energy electronic structure of the metal.

This is illustrated in Fig. 4.1, where the bandstructure of bcc Fe is shown. The details of the calculations will be presented later in Sec. 4.3. The solid lines show the full *ab-initio* bandstructure, while the dashed lines show the bands obtained within the Wannier representation using $M = 18$ WFs per cell (nine of each spin; see Sec. 4.3.2). In the method of Ref. [72], one specifies an energy $E_{\rm win}$ lying somewhat above the Fermi energy $E_{\rm f}$, and insists on finding a set of WFs spanning all the *ab-initio* states in an energy window up to $E_{\rm win}$. In the calculation of Fig. 4.1 we chose $E_{\rm win} \simeq 18\,$eV, and it is evident that there is an essentially perfect match between the fully *ab-initio* and the Wannier-represented bands up to, but not above, $E_{\rm win}$. Clearly, a Wannier-based calculation of any property of the occupied manifold, such as the intrinsic AHC, should be in excellent agreement with a direct *ab-initio* evaluation, provided that $E_{\rm win}$ is set above $E_{\rm f}$.

The Wannier-based method can even reproduce more fine but critical features as shown in Fig. 4.2. In order to capture the change of the spin of Bloch functions near Fermi surface, the energy bands are colorcoded according to the value of the spin projection $\langle S_z \rangle$. At several points along Γ-H, one state mixes with the crossing state to form an avoided crossing. The value of $\langle S_z \rangle$ changes rapidly when going through the avoided crossings, which implys that the derivative of

Bloch functions with respect to k could be orders of magnitude larger near these crossings than other regions.

We shall assume that we have M WFs per unit cell denoted as $|\mathbf{R}n\rangle$, where $n = 1, ..., M$ and $\mathbf{R}$ labels the unit cell. We shall also assume that the Bloch-like functions given by the phased sum of WFs

$$|u_{n\mathbf{k}}^{(\mathrm{W})}\rangle = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot(\hat{\mathbf{r}}-\mathbf{R})} |\mathbf{R}n\rangle \tag{4.12}$$

span the actual Bloch eigenstates $|u_{n\mathbf{k}}\rangle$ of interest ($n = 1, ..., N_{\mathbf{k}}$) at each $\mathbf{k}$ (clearly M must be $\geq N_{\mathbf{k}}$ everywhere in the BZ). It follow that, if we construct the $M \times M$ Hamiltonian matrix

$$H_{nm}^{(\mathrm{W})}(\mathbf{k}) = \langle u_{n\mathbf{k}}^{(\mathrm{W})}|\hat{H}(\mathbf{k})|u_{m\mathbf{k}}^{(\mathrm{W})}\rangle \tag{4.13}$$

and diagonalize it by finding an $M \times M$ unitary rotation matrix $U(\mathbf{k})$ such that

$$U^{\dagger}(\mathbf{k})H^{(\mathrm{W})}(\mathbf{k})U(\mathbf{k}) = H^{(\mathrm{H})}(\mathbf{k}) \tag{4.14}$$

where $H_{nm}^{(\mathrm{H})}(\mathbf{k}) = \mathcal{E}_{n\mathbf{k}}^{(\mathrm{H})}\delta_{nm}$, then $\mathcal{E}_{n\mathbf{k}}^{(\mathrm{H})}$ will be identical to the true $\mathcal{E}_{n\mathbf{k}}$ for all occupied bands. The corresponding Bloch states,

$$|u_{n\mathbf{k}}^{(\mathrm{H})}\rangle = \sum_{m} |u_{m\mathbf{k}}^{(\mathrm{W})}\rangle U_{mn}(\mathbf{k}), \tag{4.15}$$

will also be identical to the true eigenstates $|u_{n\mathbf{k}}\rangle$ for $\mathcal{E} \leq E_{\mathrm{f}}$. (In the scheme of Ref. [72], these properties will actually hold for energies up to E_{win}.) However, the band energies and Bloch states will *not* generally match the true ones at the energies higher than E_{win}, as shown in Fig. 4.1. We thus use the superscript 'H' to distinguish the projected band energies $\mathcal{E}_{n\mathbf{k}}^{(\mathrm{H})}$ and eigenvectors $|u_{n\mathbf{k}}^{(\mathrm{H})}\rangle$ from the true ones $\mathcal{E}_{n\mathbf{k}}$ and $|u_{n\mathbf{k}}\rangle$, keeping in mind that this distinction is only significant in the higher-energy unoccupied region ($\mathcal{E} > E_{\mathrm{win}}$) of the projected bandstructure.

The unitary rotation of states expressed by the matrix $U(\mathbf{k})$ is often referred to as a "gauge transformation," and we shall adopt this terminology here. We shall refer to the Wannier-derived

Bloch-like states $|u^{(\mathrm{W})}_{n\mathbf{k}}\rangle$ as belonging to the Wannier (W) gauge, while the eigenstates $|u^{(\mathrm{H})}_{n\mathbf{k}}\rangle$ of the projected bandstructure are said to belong to the Hamiltonian (H) gauge.

Quantities such as the Berry connection $\mathbf{A}_n(\mathbf{k})$ of Eq. (4.4) and the Berry curvature $\Omega_{n,\alpha\beta}(\mathbf{k})$ of Eq. (4.6) clearly depend upon the gauge in which they are expressed. (The curvature is actually invariant under the subset of gauge transformations of the diagonal form $U_{nm}(\mathbf{k}) = e^{i\phi_{n\mathbf{k}}}\delta_{nm}$, which is also the remaining gauge freedom within the Hamiltonian gauge.) The quantity that we wish to calculate, Eq. (4.8), is most naturally expressed in the Hamiltonian gauge, where it takes the form

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{M} f_n(\mathbf{k})\,\Omega^{(\mathrm{H})}_{n,\alpha\beta}(\mathbf{k})\ . \tag{4.16}$$

Here $\Omega^{(\mathrm{H})}_{n,\alpha\beta}(\mathbf{k})$ is given by Eq. (4.6) with $|u_{n\mathbf{k}}\rangle \to |u^{(\mathrm{H})}_{n\mathbf{k}}\rangle$. It is permissible to make this substitution because the projected bandstructure matches the true one for all occupied states. In practice one may take for the occupation factor $f_n(\mathbf{k}) = \theta(E_{\mathrm{f}} - \mathcal{E}_{n\mathbf{k}})$ (as done in the present work), or introduce a small thermal smearing.

Our strategy now is to see how the right-hand side of Eq. (4.16) can be obtained by starting with quantities that are defined and computed first in the Wannier gauge and then transformed into the Hamiltonian gauge. The resulting scheme can be viewed as a generalized Slater-Koster interpolation, which takes advantage of the smoothness in k-space of the Wannier-gauge objects, a direct consequence of the short range of the Wannier orbitals in real space.

4.2.2 Gauge transformations

Because the gauge transformation of Eq. (4.15) involves a unitary rotation among several bands, it is useful to introduce generalizations of the quantities in Eqs. (4.4) and (4.6) having two band indices instead of one. Thus, we define

$$A_{nm,\alpha}(\mathbf{k}) = i\langle u_n|\partial_\alpha u_m\rangle \tag{4.17}$$

and

$$\begin{aligned}\Omega_{nm,\alpha\beta}(\mathbf{k}) &= \partial_\alpha A_{nm,\beta} - \partial_\beta A_{nm,\alpha} \\ &= i\langle\partial_\alpha u_n|\partial_\beta u_m\rangle - i\langle\partial_\beta u_n|\partial_\alpha u_m\rangle \ , \end{aligned} \tag{4.18}$$

where every object in each of these equations should consistently carry either a (W) or (H) label. (We have now suppressed the $\mathbf{k}$ subscripts and introduced the notation $\partial_\alpha = \partial/\partial k_\alpha$ for conciseness.) In this notation, Eq. (4.16) becomes

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{M} f_n(\mathbf{k})\, \Omega^{(\mathrm{H})}_{nn,\alpha\beta}(\mathbf{k}) \ . \tag{4.19}$$

Note that when $\Omega_{\alpha\beta}$ appears without a (W) or (H) superscript, as on the left-hand side of this equation, it denotes the total Berry curvature on the left-hand side of Eq. (4.16).

The matrix representation of an ordinary operator such as the Hamiltonian or the velocity can be transformed from the Wannier to the Hamiltonian gauge, or vice-versa, just by operating on the left and right by $U^\dagger(\mathbf{k})$ and $U(\mathbf{k})$, as in Eq. (4.14); such a matrix is called "gauge-covariant." Unfortunately, the matrix objects in Eqs. (4.17–4.18) are not gauge-covariant, because they involve k-derivatives acting on the Bloch states. For example, a straightforward calculation shows that

$$A^{(\mathrm{H})}_\alpha = U^\dagger A^{(\mathrm{W})}_\alpha U + iU^\dagger\, \partial_\alpha U \tag{4.20}$$

where each object is an $M \times M$ matrix and matrix products are implied throughout. For every matrix object $\mathcal{O}$, we define

$$\overline{\mathcal{O}}^{(\mathrm{H})} = U^\dagger \mathcal{O}^{(\mathrm{W})} U \tag{4.21}$$

so that, by definition, $\overline{\mathcal{O}}^{(\mathrm{H})} = \mathcal{O}^{(\mathrm{H})}$ only for gauge-covariant objects.

The derivative $\partial_\alpha U$ may be obtained from ordinary perturbation theory. We adopt a notation in which $\|\phi_m\rangle\!\rangle$ is the m-th M-component column vector of matrix U, so that $\langle\!\langle\phi_n\|H^{(\mathrm{W})}\|\phi_m\rangle\!\rangle = \mathcal{E}_n\,\delta_{nm}$; the stylized bra-ket notation is used to emphasize that objects like $H^{(\mathrm{W})}$ and $\|\phi_n\rangle\!\rangle$ are

$M \times M$ matrices and M-component vectors, i.e., operators and state vectors in the "tight-binding space" defined by the WFs, not in the original Hilbert space. Perturbation theory with respect to the parameter $\mathbf{k}$ takes the form

$$\|\partial_\alpha \phi_n\rangle\!\rangle = \sum_{l \neq n} \frac{\langle\!\langle \phi_l \| H^{(\mathrm{W})}_\alpha \| \phi_n \rangle\!\rangle}{\mathcal{E}^{(\mathrm{H})}_n - \mathcal{E}^{(\mathrm{H})}_l} \, \|\phi_l\rangle\!\rangle \tag{4.22}$$

where $H^{(\mathrm{W})}_\alpha \equiv \partial_\alpha H^{(\mathrm{W})}$. In matrix notation this can be written

$$\partial_\alpha U_{mn} = \sum_l U_{ml} \, D^{(\mathrm{H})}_{ln,\alpha} = (U D^{(\mathrm{H})}_\alpha)_{mn} \tag{4.23}$$

where

$$D^{(\mathrm{H})}_{nm,\alpha} \equiv (U^\dagger \partial_\alpha U)_{nm} = \begin{cases} \dfrac{\overline{H}^{(\mathrm{H})}_{nm,\alpha}}{\mathcal{E}^{(\mathrm{H})}_m - \mathcal{E}^{(\mathrm{H})}_n} & \text{if } n \neq m \\ \\ 0 & \text{if } n = m \end{cases} \tag{4.24}$$

and $\overline{H}^{(\mathrm{H})}_{nm,\alpha} = (U^\dagger H^{(\mathrm{W})}_\alpha U)_{nm}$ according to Eq. (4.21). Note that while $\Omega_{\alpha\beta}$ and A_α are Hermitian in the band indices, $D^{(\mathrm{H})}_\alpha$ is instead antihermitian. The gauge choice implicit in Eqs. (4.22) and (4.24) is $\langle\!\langle \phi_n \| \partial_\alpha \phi_n \rangle\!\rangle = (U^\dagger \partial_\alpha U)_{nn} = 0$ (this is the so-called "parallel transport" gauge).

Using Eq. (4.23), Eq. (4.20) becomes

$$A^{(\mathrm{H})}_\alpha = \overline{A}^{(\mathrm{H})}_\alpha + i D^{(\mathrm{H})}_\alpha \tag{4.25}$$

and the derivative of Eq. (4.15) becomes

$$|\partial_\alpha u^{(\mathrm{H})}_n\rangle = \sum_m |\partial_\alpha u^{(\mathrm{W})}_m\rangle U_{mn} + \sum_m |u^{(\mathrm{H})}_m\rangle D^{(\mathrm{H})}_{mn,\alpha} \, . \tag{4.26}$$

Plugging the latter into Eq. (4.18), we finally obtain, after a few manipulations, the matrix equation

$$\begin{aligned} \Omega^{(\mathrm{H})}_{\alpha\beta} \;\; = \;\; & \overline{\Omega}^{(\mathrm{H})}_{\alpha\beta} - [D^{(\mathrm{H})}_\alpha, \overline{A}^{(\mathrm{H})}_\beta] \\ & + [D^{(\mathrm{H})}_\beta, \overline{A}^{(\mathrm{H})}_\alpha] - i [D^{(\mathrm{H})}_\alpha, D^{(\mathrm{H})}_\beta] \, . \end{aligned} \tag{4.27}$$

The band-diagonal elements $\Omega^{(\mathrm{H})}_{nn,\alpha\beta}(\mathbf{k})$ then need to be inserted into Eq. (4.19).

Eq. (4.27) can also be derived from Eq. (4.25), by combining it with the first line of Eq. (4.18):

$$\Omega^{(\mathrm{H})}_{\alpha\beta} = \partial_\alpha \overline{A}^{(\mathrm{H})}_\beta - \partial_\beta \overline{A}^{(\mathrm{H})}_\alpha - i[D^{(\mathrm{H})}_\alpha, D^{(\mathrm{H})}_\beta], \tag{4.28}$$

where we have used $i(\partial_\alpha U)^\dagger \partial_\beta U = -iD^{(\mathrm{H})}_\alpha D^{(\mathrm{H})}_\beta$. Invoking Eq. (4.21) we find

$$\begin{aligned} \partial_\alpha \overline{A}^{(\mathrm{H})}_\beta - \partial_\beta \overline{A}^{(\mathrm{H})}_\alpha = \; & - \; [D^{(\mathrm{H})}_\alpha, \overline{A}^{(\mathrm{H})}_\beta] + [D^{(\mathrm{H})}_\beta, \overline{A}^{(\mathrm{H})}_\alpha] \\ & + \; U^\dagger \big(\partial_\alpha A^{(\mathrm{W})}_\beta - \partial_\beta A^{(\mathrm{W})}_\alpha\big) U. \end{aligned} \tag{4.29}$$

The last term on the right-hand-side is $\overline{\Omega}^{(\mathrm{H})}_{\alpha\beta}$, and thus we recover Eq. (4.27).

4.2.3 Discussion

We expect, based on Eq. (4.9), that the largest contributions to the AHC will come from regions of k-space where there are small energy splittings between bands (for example, near spin-orbit-split avoided crossings) [13]. In the present formulation, this will give rise to small energy denominators in Eq. (4.24), leading to very large $D^{(\mathrm{H})}_\alpha$ values in those regions. These large and spiky contributions will then propagate into $A^{(\mathrm{H})}_\alpha$ and $\Omega^{(\mathrm{H})}_{\alpha\beta}$, whereas $A^{(\mathrm{W})}_\alpha$ and $\Omega^{(\mathrm{W})}_{\alpha\beta}$, and also $\overline{A}^{(\mathrm{H})}_\alpha$ and $\overline{\Omega}^{(\mathrm{H})}_{\alpha\beta}$, will remain with their typically smaller values. Thus, these spiky contributions will be present in the second and third terms, and especially in the fourth term, of Eq. (4.27). The contributions of these various terms are illustrated for the case of bcc Fe in Sec. 4.4.1, and we show there that the last term typically makes by far the dominant contribution, followed by the second and third terms, and then by the first term.

The dominant fourth term can be recast in the form of a Kubo formula as

$$\Omega^{DD}_{n,\alpha\beta} = -2\mathrm{Im} \sum_{m \neq n} \frac{\langle\!\langle \phi_n \| H^{(\mathrm{W})}_\alpha \| \phi_m \rangle\!\rangle \langle\!\langle \phi_m \| H^{(\mathrm{W})}_\beta \| \phi_n \rangle\!\rangle}{\big(\mathcal{E}^{(\mathrm{H})}_m - \mathcal{E}^{(\mathrm{H})}_n\big)^2} \, . \tag{4.30}$$

The following differences between this equation and the true Kubo formula, Eq. (4.9), should however be kept in mind. First, the summation in Eq. (4.30) is restricted to the M-band projected band

structure. Second, above E_{win} the projected bandstructure deviates from the original *ab-initio* one. Third, even below E_{win}, where they do match exactly, the "effective tight-binding velocity matrix elements" appearing in Eq. (4.30) differ from the true ones, given by Eq. (4.10). (The relation between them is particularly simple for energies below E_{win},

$$v^{(\mathrm{H})}_{nm,\alpha} = \frac{1}{\hbar}\overline{H}^{(\mathrm{H})}_{nm,\alpha} - \frac{i}{\hbar}\big(\mathcal{E}^{(\mathrm{H})}_{m} - \mathcal{E}^{(\mathrm{H})}_{n}\big)\overline{A}^{(\mathrm{H})}_{nm,\alpha}, \tag{4.31}$$

and follows from combining the identity [71] $A_{nm,\alpha} = i\langle\psi_n|\hat{v}_\alpha|\psi_m\rangle/(\omega_m - \omega_n)$, valid for $m \neq n$, with Eqs. (4.24–4.25).) All these differences are however exactly compensated by the previous three terms in Eq. (4.27). We emphasize that all terms in that equation are defined strictly within the projected space spanned by the Wannier functions.

We note in passing that it is possible to rewrite Eq. (4.27) in such a way that the large spiky contributions are isolated into a single term. This alternative formulation, which turns out to be related to a gauge-covariant curvature tensor, will be described in Appendix A.

4.2.4 Sum over occupied bands

In the above, we have proposed to evaluate $\Omega^{(\mathrm{H})}_{nn,\alpha\beta}$ from Eq. (4.27) and then insert it into the band sum, Eq. (4.19), in order to compute the AHC. However, this approach has the shortcoming that small splittings (avoided crossings) between a pair of *occupied* bands n and m lead to large values of $D^{(\mathrm{H})}_{nm,\alpha}$, and thus to large but canceling contributions to the AHC coming from $\Omega^{(\mathrm{H})}_{nn,\alpha\beta}$ and $\Omega^{(\mathrm{H})}_{mm,\alpha\beta}$. Here, we rewrite the total Berry curvature (4.19) in such a way that the cancellation is explicit.

Inserting Eq. (4.27) into Eq. (4.19) and interchanging dummy labels $n \leftrightarrow m$ in certain terms,

we obtain

$$\begin{aligned}\Omega_{\alpha\beta}(\mathbf{k}) &= \sum_n f_n \,\overline{\Omega}^{(\mathrm{H})}_{nn,\alpha\beta} \\ &+ \sum_{nm}(f_m - f_n)\left(D^{(\mathrm{H})}_{nm,\alpha}\overline{A}^{(\mathrm{H})}_{mn,\beta}\right. \\ &\quad \left. - D^{(\mathrm{H})}_{nm,\beta}\overline{A}^{(\mathrm{H})}_{mn,\alpha} + iD^{(\mathrm{H})}_{nm,\alpha}D^{(\mathrm{H})}_{mn,\beta}\right). \end{aligned} \tag{4.32}$$

The factors of $(f_m - f_n)$ insure that terms arising from pairs of fully occupied states give no contribution. Thus, the result of this reformulation is that individual terms in Eq. (4.32) have large spiky contributions only when avoided crossings or near-degeneracies occur across the Fermi energy. This approach is therefore preferable from the point of view of numerical stability, and it is the one that we have implemented in the current work.

As expected from the discussion in Sec. 4.2.3 and shown later in Sec. 4.4.2, the dominant term in Eq. (4.32) is the last one,

$$\Omega^{DD}_{\alpha\beta} = i\sum_{nm}(f_m - f_n)D^{(\mathrm{H})}_{nm,\alpha}D^{(\mathrm{H})}_{mn,\beta} \tag{4.33}$$

or, in a more explicitly Kubo-like form,

$$\Omega^{DD}_{\alpha\beta} = i\sum_{nm}(f_m - f_n)\,\frac{\overline{H}^{(\mathrm{H})}_{nm,\alpha}\overline{H}^{(\mathrm{H})}_{mn,\beta}}{\left(\mathcal{E}^{(\mathrm{H})}_m - \mathcal{E}^{(\mathrm{H})}_n\right)^2}\,. \tag{4.34}$$

In the zero-temperature limit, the latter can easily be cast into a form like Eq. (4.30), but with a double sum running over occupied bands n and unoccupied bands m, very reminiscent of the original Kubo formula in Eq. (4.11).

We remark that $(1/\hbar)\overline{H}^{(\mathrm{H})}_{nm,\alpha}$ coincides with the "effective tight-binding velocity operator" of Ref. [73]. This is an approximate tight-binding velocity operator. Comparison with Eq. (4.31) and Eq. (4.39) below shows that it is lacking the contributions which involve matrix elements of the position operator between the WFs [74]. We now recognize in Eq. (4.22) the standard result from $\mathbf{k}\cdot\mathbf{p}$ pertubation theory, but in terms of the approximate momentum operator. Using that equation,

Eq. (4.30) can be cast as the tight-binding-space analog of Eq. (4.6),

$$\Omega^{DD}_{n,\alpha\beta} = -2\,\mathrm{Im}\,\langle\!\langle \partial_\alpha \phi_{n\mathbf{k}} \| \partial_\beta \phi_{n\mathbf{k}} \rangle\!\rangle \,. \tag{4.35}$$

This allows to rewrite Eq. (4.34) in a form that closely resembles the total Berry curvature, Eq. (4.16):

$$\Omega^{DD}_{\alpha\beta} = \sum_{n=1}^{M} f_n\, \Omega^{DD}_{n,\alpha\beta} \,. \tag{4.36}$$

4.2.5 Evaluation of the Wannier-gauge matrices

Eq. (4.32) is our primary result. To review, recall that this is a condensed notation expressing the $M \times M$ matrix $\Omega^{(\mathrm{H})}_{nm,\alpha\beta}(\mathbf{k})$ in terms of the matrices $\overline{\Omega}^{(\mathrm{H})}_{nm,\alpha\beta}(\mathbf{k})$, etc. The basic ingredients needed are the four matrices $H^{(\mathrm{W})}$, $H^{(\mathrm{W})}_\alpha$, $A^{(\mathrm{W})}_\alpha$, and $\Omega^{(\mathrm{W})}_{\alpha\beta}$ at a given $\mathbf{k}$. Diagonalization of the first of them yields the energy eigenvalues needed to find the occupation factors f_n. It also provides the gauge transformation U which is then used to construct $\overline{H}^{(\mathrm{H})}_\alpha$, $\overline{A}^{(\mathrm{H})}_\alpha$, and $\overline{\Omega}^{(\mathrm{H})}_{\alpha\beta}$ from the other three objects via Eq. (4.21). Finally, $\overline{H}^{(\mathrm{H})}_\alpha$ is inserted into Eq. (4.24) to obtain $D^{(\mathrm{H})}_\alpha$, and all terms in Eq. (4.32) are evaluated.

In this section we explain how to obtain the matrices $H^{(\mathrm{W})}(\mathbf{k})$, $H^{(\mathrm{W})}_\alpha(\mathbf{k})$, $A^{(\mathrm{W})}_\alpha(\mathbf{k})$ and $\Omega^{(\mathrm{W})}_{\alpha\beta}(\mathbf{k})$ at an arbitrary point $\mathbf{k}$ for use in the subsequent calculations described above.

Fourier transform expressions

The four needed quantities can be expressed as follows:

$$H^{(\mathrm{W})}_{nm}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, \langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle \,, \tag{4.37}$$

$$H^{(\mathrm{W})}_{nm,\alpha}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, iR_\alpha\, \langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle \,, \tag{4.38}$$

$$A^{(\mathrm{W})}_{nm,\alpha}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle \,, \tag{4.39}$$

$$\Omega^{(\mathrm{W})}_{nm,\alpha\beta}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \Big(iR_\alpha \langle \mathbf{0}n|\hat{r}_\beta|\mathbf{R}m\rangle - iR_\beta \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle \Big) . \tag{4.40}$$

(The notation $|\mathbf{0}n\rangle$ refers to the n'th WF in the home unit cell $\mathbf{R} = \mathbf{0}$.) Eq. (4.37) follows by combining Eqs. (4.12) and (4.13), while Eq. (4.39) follows by combining Eqs. (4.12) and (4.17). Eqs. (4.38) and (4.40) are then obtained from (4.37) and (4.39) using $H_{nm,\alpha} = \partial_\alpha H_{nm}$ and Eq. (4.18), respectively.

It is remarkable that the only real-space matrix elements that are required between WFs are those of the four operators $\hat{H}$ and $\hat{r}_\alpha$ ($\alpha = x$, y, and z). Because the WFs are strongly localized, these matrix elements are expected to decay rapidly as a function of lattice vector $\mathbf{R}$, so that only a modest number of them need to be computed and stored once and for all. Collectively, they define our "exact tight-binding model" and suffice to allow subsequent calculation of all needed quantities. Furthermore, the short range of these matrix elements in real space insures that the Wannier-gauge quantities on the left-hand sides of Eqs. (4.37–4.40) will be smooth functions of $\mathbf{k}$, thus justifying the earlier discussion in which it was argued that these objects should have no rapid variation or enhancement in k-space regions where avoided crossings occur. (Recall that such large, rapidly-varying contributions only appear in the $D^{(\mathrm{H})}$ matrices and in quantities that depend upon them.) It should however be kept in mind that Eq. (4.32) is not written directly in terms of the smooth quantities (4.37–4.40), but rather in terms of those quantities transformed according to Eq. (4.21). The resulting objects are not smooth, since the matrices U change rapidly with $\mathbf{k}$. However, even while not smooth, they remain small.

Evaluation of real-space matrix elements

We conclude this section by discussing the calculation of the fundamental matrix elements $\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle$ and $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle$. There are several ways in which these could be computed, and the choice could

well vary from one implementation to another. One possibility would be to construct the WFs in real space, say on a real-space grid, and then to compute the Hamiltonian and position-operator matrix elements directly on that grid. In the context of a code that uses a real-space basis (e.g., localized orbitals or grids), this might be the best choice. However, in the context of plane-wave methods it is usually more convenient to work in reciprocal space if possible. This is in the spirit of the Wannier-function construction scheme [75, 72], which is formulated as a post-processing step after a conventional *ab-initio* calculation carried out on a uniform k-point grid. (In the following we will use the symbol $\mathbf{q}$ to denote the points of this *ab-initio* mesh, to distinguish them from arbitrary or interpolation-grid points denoted by $\mathbf{k}$.)

The end result of the Wannier-construction step are M Bloch-like functions $|u^{(\mathrm{W})}_{n\mathbf{q}}\rangle$ at each $\mathbf{q}$. The WFs are obtained from them via a discrete Fourier transform:

$$|\mathbf{R}n\rangle = \frac{1}{N_q^3}\sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot(\mathbf{R}-\hat{\mathbf{r}})}|u^{(\mathrm{W})}_{n\mathbf{q}}\rangle \ . \tag{4.41}$$

This expression follows from inverting Eq. (4.12). If the *ab initio* mesh contains $N_q \times N_q \times N_q$ points, the resulting WFs are really periodic functions over a supercell of dimensions $L \times L \times L$, where $L = N_q a$ and a is the lattice constant of the unit cell. The idea then is to choose L large enough that the rapid decay of the localized WFs occurs on a scale much smaller than L. This ensures that the matrix elements $\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle$ and $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle$ between a pair of WFs separated by more than $L/2$ are negligible, so that further refinement of the *ab-initio* mesh will have a negligible impact on the accuracy of Wannier-interpolated quantities. (In particular, the interpolated band structure, Fig. 4.1, is able to reproduce tiny features of the full bandstructure, such as spin-orbit-induced avoided crossings, even if they occur on a length scale much smaller than the *ab-initio* mesh spacing.) While the choice of reciprocal-space cell spanned by the vectors $\mathbf{q}$ is immaterial, because of the periodicity of reciprocal space, this is not so for the vectors $\mathbf{R}$. In practice we choose the $N_q \times N_q \times N_q$ vectors $\mathbf{R}$ to be evenly distributed on the Wigner-Seitz supercell of volume $N_q^3 a^3$

centered around $\mathbf{R} = \mathbf{0}$ [72]. This is the most isotropic choice possible, ensuring that the strong decay of the matrix elements for $|\mathbf{R}| \sim L/2$ is achieved irrespective of direction.

The matrix elements of the Hamiltonian are obtained from Eq. (4.41) as

$$\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle = \frac{1}{N_q^3}\sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} H_{nm}^{(\mathrm{W})}(\mathbf{q}) \ , \tag{4.42}$$

which is the reciprocal of Eq. (4.37), with the sum running over the coarse *ab-initio* mesh points. The position matrix is obtained similarly by inverting Eq. (4.39):

$$\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle = \frac{1}{N_q^3}\sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{q}) \ . \tag{4.43}$$

The matrix $A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{q})$ is then evaluated by approximating the k-derivatives in Eq. (4.17) by finite-differences on the *ab-initio* mesh using the expression [75]

$$A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{q}) \simeq i\sum_{\mathbf{b}} w_b b_\alpha \Big(\langle u_{n\mathbf{q}}^{(\mathrm{W})} | u_{m,\mathbf{q}+\mathbf{b}}^{(\mathrm{W})}\rangle - \delta_{nm} \Big) \ , \tag{4.44}$$

where $\mathbf{b}$ are the vectors connecting $\mathbf{q}$ to its nearest neighbors on the *ab-initio* mesh. This approximation is valid because in the Wannier gauge the Bloch states vary smoothly with $\mathbf{k}$. We note that the overlap matrices appearing on the right-hand side are available "for free" as they have already been computed and stored during the WF construction procedure. This is also the case for the matrices $H^{(\mathrm{W})}(\mathbf{q})$ needed in Eq. (4.42).

It should be kept in mind that the k-space finite-difference procedure outlined above entails an error of order $\mathcal{O}(\Delta q^2)$ in the values of the position operator matrix elements, where Δq is the *ab-initio* mesh spacing. The importance of such an error is easily assessed by trying denser q-point meshes; in our case, we find that it is not a numerically significant source of error for the $8 \times 8 \times 8$ mesh that we employ in our calculations. (In large measure this is simply because less than 2% of the total AHC comes from terms that depend on these position-operator matrix elements, as will be discussed in Sec. 4.4. Indeed, we find that the $\mathcal{O}(\Delta q^2)$ convergence of this small contribution

hardly shows in the convergence of the total AHC, which empirically appears to be approximately exponential in the *ab-initio* mesh density.) However, if the $\mathcal{O}(\Delta q^2)$ convergence is a source of concern, one could adopt the direct real-space mesh integration method mentioned at the beginning of this subsection, which should be free of such errors.

4.3 Computational details

In this section we present some of the detailed steps of the calculations as they apply to our test system of bcc Fe. First, we describe the first-principles bandstructure calculations that are carried out initially. Second, we discuss the procedure for constructing maximally localized Wannier functions for the bands of interest following the method of Souza, Marzari, and Vanderbilt [72]. Third, we discuss the variable treatment of the spin-orbit interaction within these first-principles calculations, which is useful for testing the dependence of the AHC on the spin-orbit coupling strength.

4.3.1 Band structure calculation

Fully relativistic band structure calculations for bcc Fe in its ferromagnetic ground state at the experimental lattice constant $a = 5.42$ Bohr are carried out using the `PWSCF` code [76]. A kinetic-energy cutoff of 60 Hartree is used for the planewave expansion of the valence wavefunctions (400 Hartree for the charge densities). Exchange and correlation effects are treated with the PBE generalized-gradient approximation [77].

The core-valence interaction is described here by means of norm-conserving pseudopotentials which include spin-orbit effects [78, 79] in separable Kleinman-Bylander form. (Our overall Wannier interpolation approach is quite independent of this specific choice and can easily be generalized to other kinds of pseudopotentials or to all-electron methods.) The pseudopotential was constructed using a reference valence configuration of $3d^7 4s^{0.75} 4p^{0.25}$. We treat the overlap of the valence states

with the semicore $3p$ states using the non-linear core correction approach [80]. The pseudopotential core radii for the $3d$, $4s$ and $4p$ states are 1.3, 2.0 and 2.2 Bohr, respectively. We find the small cut-off radius for the $3d$ channel to be necessary in order to reproduce the all-electron bandstructure accurately.

We obtain the self-consistent ground state using a $16\times16\times16$ Monkhorst-Pack [81] mesh of k-points and a fictitious Fermi smearing [82] of 0.02 Ry for the Brillouin-zone integration. The magnetization is along the [001] direction, so that the only non-zero component of the integrated Berry curvature, Eq. (4.7), is the one along z. The spin magnetic moment is found to be 2.22 μ_{B}, the same as that from an all-electron calculation [21] and close to the experimental value of 2.12 μ_{B}.

In order to calculate the Wannier functions, we freeze the self-consistent potential and perform a non-self-consistent calculation on a uniform $n \times n \times n$ grid of k-points (the "*ab-initio* mesh"). We tested several grid densities ranging from n=4 to n=10 and ultimately chose n=8 (see end of next subsection). Since we want to construct 18 WFs (s, p, and d-like for spin up and down), we need to include a sufficient number of extra bands to cover the orbital character of these intended WFs everywhere in the Brillouin zone. With this in mind, we calculate the first 28 bands at each k-point, and then exclude any bands above 58 eV, the "outer window" of Ref. [72]. (The choice of outer window is somewhat arbitrary as long as the number of bands it encloses is larger than the number of WFs, and we confirm that the calculated AHC has very little dependence upon this choice. The main effect of choosing a larger outer window is that one obtains slightly more localized WFs in real space, and thus slightly smoother bands in k-space.) The 18 WFs are then disentangled from the remaining bands using the procedure described in the next section.

4.3.2 Maximally-localized spinor Wannier functions for bcc Fe

The energy bands of interest (extending up to, and just above, the Fermi energy) have mainly mixed s and d character and are entangled with the bands at higher energies. In order to construct maximally-localized WFs to describe these bands, we use a modified version of the post-processing procedure of Ref. [72]. We start by reviewing the original two-step procedure from that work, as it applies to iron. In the first ("subspace selection") step, an 18-band subspace (the "projected space") is identified. This is done by minimizing a suitably defined functional, subject to the constraint of including the states within an inner energy window [72]. In the case of iron we choose this window to span an energy range of 30 eV from the bottom of the valence bands (up to $E_{\rm win}$ in Fig. 4.1). In the second ("gauge selection") step, the gauge freedom within the projected subspace is explored to obtain a set of Bloch-like functions $|u_{n\mathbf{k}}^{(\rm W)}\rangle$ which are optimally-smooth as a function of $\mathbf{k}$ [75]. They are related to the 18 maximally-localized WFs by Eq. (4.12). Although the method of Refs. [72] and [75] was formulated for the spinless case, it is trivial to adapt it to treat spinor wavefunctions, in which case the resulting WFs also have spinor character: each element of the overlap matrix, which is the key input to the WF-generation code, is simply calculated as the sum of two spin components,

$$S_{\mathbf{k},\mathbf{b}}^{nm} = \sum_{\sigma=\uparrow,\downarrow} \langle u_{n\mathbf{k}}^{\sigma} | u_{m,\mathbf{k}+\mathbf{b}}^{\sigma} \rangle \ . \tag{4.45}$$

In order to facilitate later analysis (e.g., of the orbital and spin character of various bands), we have used a modified three-step procedure. The initial subspace selection step remains unchanged. The new second step ("subspace division") consists of splitting the 18-dimensional projected space for each $\mathbf{k}$ on the *ab-initio* mesh into two 9-dimensional subspaces, as follows. At each k-point we form the 18×18 matrix representation of the spin operator $\hat{S}_z = (\hbar/2)\hat{\sigma}_z$ in the projected space and diagonalize it. The two 9-dimensional subspaces are then chosen as a mostly spin-up subspace spanned by the eigenstates having S_z eigenvalues close to $+1$, and a mostly spin-down subspace

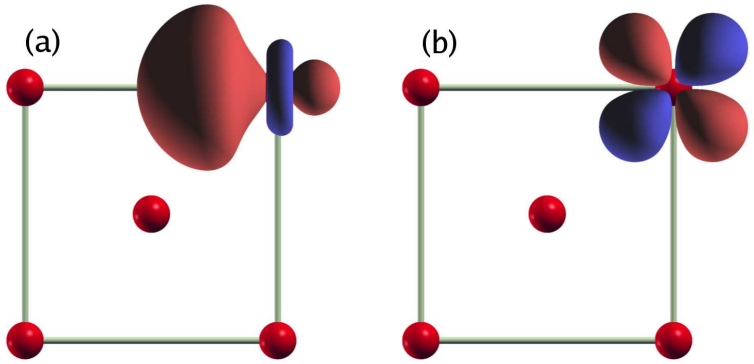

Figure 4.3: Isosurface contours of maximally-localized spin-up WF in bcc Fe (red for positive value and blue for negative value), for the $8 \times 8 \times 8$ k-point sampling. (a) sp^3d^2-like WF centered on a Cartesian axis; (b) d_{xy}-like WF centered on the atom.

associated with eigenvalues close to -1 (we will use units of $\hbar/2$ whenever we discuss S_z in the remainder of the manuscript). The third and final step is the gauge-selection step, which is now done separately for each of the two 9-dimensional subspaces. We thus emerge with 18 well-localized WFs divided into two groups: nine that are almost entirely spin-up and nine that are almost entirely spin-down (in practice we find $|\langle \hat{S}_z \rangle| > 0.999$ in all cases). While this procedure results in a total spread that is slightly greater than the original two-step procedure, we find that the difference is very small in practice, and the imposition of these rules makes for a much more transparent analysis of subsequent results. For example, it makes it much easier to track the changes in the WFs before and after the spin-orbit coupling is turned on, or to identify the spin character of various pieces of the Fermi surface.

The subspace-selection step can be initialized [72] by providing 18 trial functions having the form of s, p, and (e_g and t_{2g}) d-like Gaussians of pure spin character (nine up and nine down). In our first attempts at initializing the gauge-selection step, we used these same trial functions. However, we found that the iterative gauge-selection procedure [75], which projects the nine trial functions of each spin onto the appropriate band subspace and improves upon them, converted the three t_{2g}-like

trial functions into t_{2g}-like WFs, while it mixed the e_g, s, and p-like states to form six hybrid WFs of sp^3d^2-type [83]. Having discovered this, we have modified our procedure accordingly: henceforth, we choose three t_{2g}-like trial functions and six sp^3d^2-like ones in each spin channel. With this initialization, we find the convergence to be quite rapid, with only about 100 iterations needed to get a well-converged spread functional.

We have implemented the above procedure in the `WANNIER90` code [84]. The resulting WFs are shown in Fig. 4.3. The up-spin WFs are plotted, but the WFs are very similar for both spins. An example of an sp^3d^2-hybrid WF is shown in Fig. 4.3(a); this one extends along the $-x$ axis, and the five others are similarly projected along the $+x$, $\pm y$, and $\pm z$ axes. One of the t_{2g}-like WFs is shown in Fig. 4.3(b); this one has xy symmetry, while the others have xz and yz symmetry. The centers of the sp^3d^2-like WFs are slightly shifted from the atomic center along $\pm x$, $\pm y$, or $\pm z$, while the t_{2g}-like WFs remain centered on the atom.

We studied the convergence of the WFs and interpolated bands as a function of the density $n \times n \times n$ of the Monkhorst-Pack k-mesh used for the initial *ab-initio* calculation. We tested $n = 4$, 6, 8, and 10, and found that $n = 8$ provided the best tradeoff between interpolation accuracy and computational cost. This is the mesh that was used in generating the results presented in Sec. 4.4.

4.3.3 Variable spin-orbit coupling in the pseudopotential framework

Since the AHE present in ferromagnetic iron is a spin-orbit-induced effect, it is obviously important to understand the role of this coupling as thoroughly as possible. For this purpose, it is very convenient to be able to treat the strength of the coupling as an adjustable parameter. For example, by turning up the spin-orbit coupling continuously from zero and tracking how various contributions to the AHC behave, it is possible to separate out those contributions that are of linear, quadratic, or higher order in the coupling strength. Some results of this kind will be given later in Sec. 4.4.

Because the spin-orbit coupling is a relativistic effect, it is appreciable mainly in the core region of the atom where the electrons have relativistic velocities. In a pseudopotential framework of the kind adopted here, both the scalar relativistic effects and the spin-orbit coupling are included in the pseudopotential construction. For example, in the Bachelet-Hamann semilocal pseudopotential scheme [31], the construction procedure generates, for each orbital angular momentum l, a scalar-relativistic potential $V_l^{\rm sr}(r)$ and a spin-orbit difference potential $V_l^{\rm so}(r)$ which enter the Hamiltonian in the form

$$\hat{V}_{\rm ps} = \sum_l \hat{P}_l \, \left[V_l^{\rm sr}(r) + \lambda \, V_l^{\rm so}(r) \, \mathbf{L} \cdot \mathbf{S} \right] \, , \tag{4.46}$$

where $\hat{P}_l$ is the projector onto states of orbital angular momentum l and λ controls the strength of spin-orbit coupling (with λ=1 being the physical value). For the free atom, this correctly leads to eigenstates labeled by total angular momentum $j = l \pm 1/2$.

In our calculations, we employ fully non-local pseudopotentials instead of semilocal ones because of their computationally efficient form. In this case, controlling the strength of the spin-orbit coupling requires some algebraic manipulation. We write the norm-conserving non-local pseudopotential operator as

$$\hat{V}_{\rm ps} = |\beta_{lj\mu}\rangle \, D_{lj} \, \langle \beta_{lj\mu}| \tag{4.47}$$

where there is an implied sum running over the indices (orbital angular momentum l, total angular momentum $j = l \pm 1/2$, and $\mu = -j, ..., j$) and species and atomic position indices have been suppressed. The $|\beta_{lj\mu}\rangle$ are radial functions multiplied by appropriate spin-angular functions and the D_{lj} are the channel weights. We introduce the notation $\beta_l^{(+)}(r)$ and $\beta_l^{(-)}(r)$ for the radial parts of $|\beta_{l,l+1/2,\mu}\rangle$ and $\beta_{l,l-1/2,\mu}\rangle$, respectively, and similarly define $D_l^{(\pm)} = D_{l,l\pm1/2}$. Using this notation, we can define the scalar-relativistic (i.e., j-averaged) quantities

$$D_l^{\rm sr} = \frac{l+1}{2l+1} \, D_l^{(+)} + \frac{l}{2l+1} \, D_l^{(-)} \, , \tag{4.48}$$

$$\beta_l^{\rm sr}(r) = \frac{l+1}{2l+1}\sqrt{\frac{D_l^{(+)}}{D_l^{\rm sr}}}\ \beta_l^{(+)}(r) + \frac{l}{2l+1}\sqrt{\frac{D_l^{(-)}}{D_l^{\rm sr}}}\ \beta_l^{(-)}(r) \tag{4.49}$$

and the corresponding spin-orbit difference quantities

$$D_{lj}^{\rm so} = D_{lj} - D_l^{\rm sr}\ , \tag{4.50}$$

$$|\beta_{lj\mu}^{\rm so}\rangle = |\beta_{lj\mu}\rangle - |\beta_{lj\mu}^{\rm sr}\rangle\ . \tag{4.51}$$

where $|\beta_{lj\mu}^{\rm sr}\rangle$ is $\beta_l^{\rm sr}(r)$ multiplied by the spin-angular function with labels $(lj\mu)$. Then the non-local pseudopotential can be written as

$$\hat{V}_{\rm ps} = \hat{V}^{\rm sr} + \lambda\,\hat{V}^{\rm so} \tag{4.52}$$

where

$$\hat{V}^{\rm sr} = |\beta_{lj\mu}^{\rm sr}\rangle\, D_l^{\rm sr}\, \langle\beta_{lj\mu}^{\rm sr}| \tag{4.53}$$

and

$$\begin{aligned}\hat{V}_{\rm so} \;=\;& |\beta_{lj\mu}^{\rm sr}\rangle\, D_{lj}^{\rm so}\, \langle\beta_{lj\mu}^{\rm sr}| \\ &+ |\beta_{lj\mu}^{\rm so}\rangle\, (D_l^{\rm sr} + D_{lj}^{\rm so})\, \langle\beta_{lj\mu}^{\rm sr}| \\ &+ |\beta_{lj\mu}^{\rm sr}\rangle\, (D_l^{\rm sr} + D_{lj}^{\rm so})\, \langle\beta_{lj\mu}^{\rm so}| \\ &+ |\beta_{lj\mu}^{\rm so}\rangle\, (D_l^{\rm sr} + D_{lj}^{\rm so})\, \langle\beta_{lj\mu}^{\rm so}|\ .\end{aligned} \tag{4.54}$$

This clearly reduces to the desired results (4.47) for $\lambda = 1$ and (4.53) for $\lambda = 0$.

4.4 Results

In this section, we present the results of the calculations of the Berry curvature and its integration over the BZ using the formulas presented in Sec. 4.2, for the case of bcc Fe.

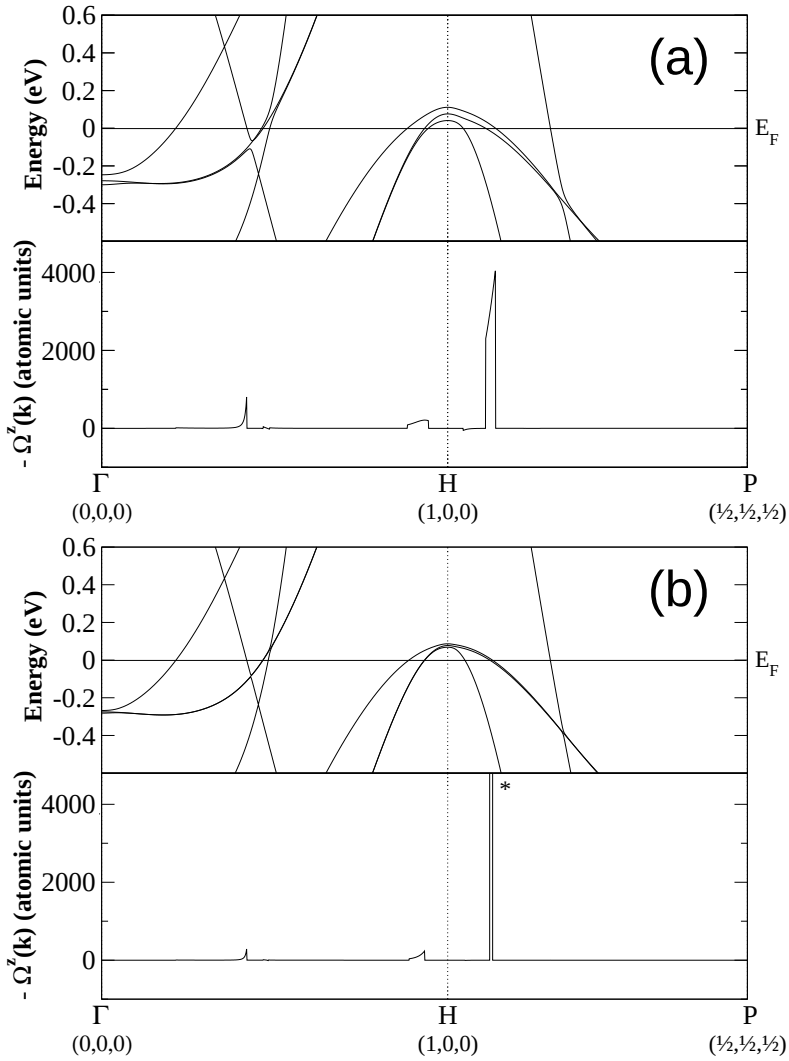

Figure 4.4: Band structure and total Berry curvature, as calculated using Wannier interpolation, plotted along the path Γ–H–P in the Brillouin zone. (a) Computed at the full spin-orbit coupling strength $\lambda = 1$. (b) Computed at the reduced strength $\lambda = 0.25$. The peak marked with a star has a height of $5{\times}10^4$ a.u.

4.4.1 Berry Curvature

We begin by illustrating the very sharp and strong variations that can occur in the total Berry curvature, Eq. (4.8), near Fermi-surface features in the bandstructure [13]. In Fig. 4.4(a) we plot the energy bands (top subpanel) and the total Berry curvature (bottom subpanel) in the vicinity of the zone-boundary point $\mathrm{H} = \frac{2\pi}{a}(1,0,0)$, where three states, split by the spin-orbit interaction, lie just above the Fermi level. The large spike in the Berry curvature between the H and P points arises where two bands, split by the spin orbit interaction, lie on either side of the Fermi level [21]. This gives rise to small energy denominators, and hence large contributions, mainly in Eq. (4.34). On reducing the strength of the spin-orbit interaction as in Fig. 4.4(b), the energy separation between these bands is reduced, resulting in a significantly sharper and higher spike in the Berry curvature. A second type of sharp structure is visible in Fig. 4.4, where one can see two smaller spikes, one at about 40% and another at about 90% of the way from Γ to H, which decrease in magnitude as the as the spin-orbit coupling strength is reduced. These arise from pairs of bands that straddle the Fermi energy even in the absence of spin-orbit interaction. Thus, the small spin-orbit coupling does not shift the energies of these bands significantly, but it does induce an appreciable Berry curvature that is roughly linear in the spin-orbit coupling.

The decomposition of the total Berry curvature into its various contributions in Eq. (4.32) is illustrated by plotting the first ("$\overline{\Omega}$") term, the second and third ("D–$\overline{A}$") terms, and the fourth ("D–D" or Kubo-like) term of Eq. (4.32) separately along the line Γ–H–P in Fig. 4.5. Note the logarithmic scale. The results confirm the expectations expressed in Secs. 4.2.3 and 4.2.4, namely, that the largest terms would be those reflecting large contributions to D arising from small energy denominators. Thus, the $\overline{\Omega}$ term remains small everywhere, the D–$\overline{A}$ terms become one or two orders of magnitude larger at places where small energy denominators occur, and the D–D term,

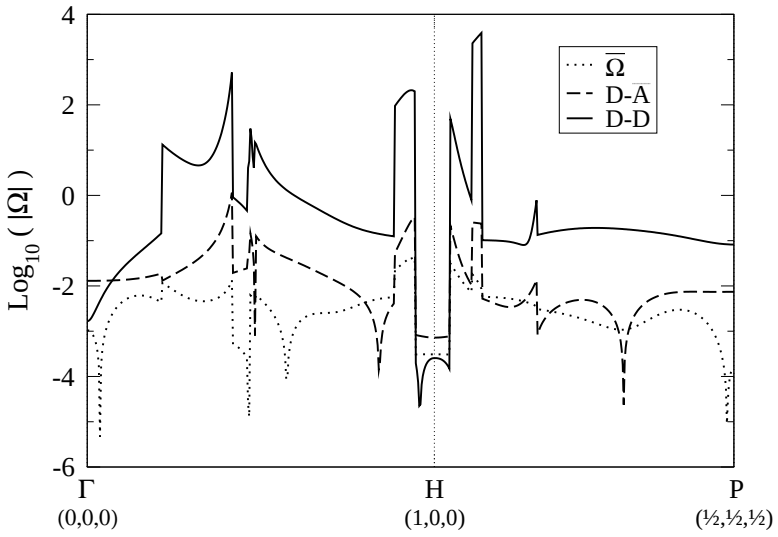

Figure 4.5: Decomposition of the total Berry curvature into contributions coming from the three kinds of terms appearing in Eq. (4.32). The path in k-space is the same as in Fig. 4.4. Dotted line is the first ($\overline{\Omega}$) term, dashed line is the sum of second and third (D–$\overline{A}$) terms, and solid line is the fourth (D–D) term of Eq. (4.32). Note the log scale on the vertical axis.

Eq. (4.34), is another one or two orders larger in those same regions. Scans along other lines in k-space reveal similar behavior. We may therefore expect that the D–D term will make the dominant overall contribution to the AHC. As we shall show in the next subsection, this is precisely the case.

In order to get a better feel for the connection between Fermi surface features and the Berry curvature, we next inspect these quantities on the $k_y = 0$ plane in the Brillouin zone, following Ref. [21]. In Fig. 4.6 we plot the intersection of the Fermi surface with this plane and indicate, using color coding, the S_z component of the spin carried by the corresponding wavefunctions. The good agreement between the shape of the Fermi surface given here and in Fig. 3 of Ref. [21] is further evidence that the accuracy of our approach matches that of all-electron methods. It is evident that the presence of the spin-orbit interaction, in addition to the exchange splitting, is sufficient to remove all degeneracies on this plane [85], changing significantly the connectivity of the Fermi surface.

The calculated Berry curvature is shown in Fig. 4.7. It can be seen that the regions in which

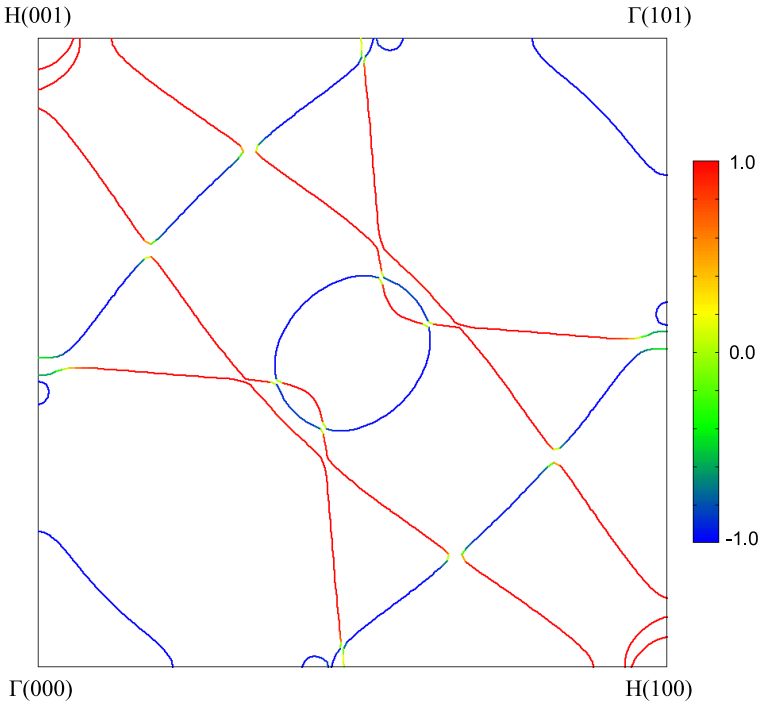

Figure 4.6: Lines of intersection between the Fermi surface and the plane $k_y = 0$. Colors indicate the S_z spin-component of the states on the Fermi surface (in units of $\hbar/2$).

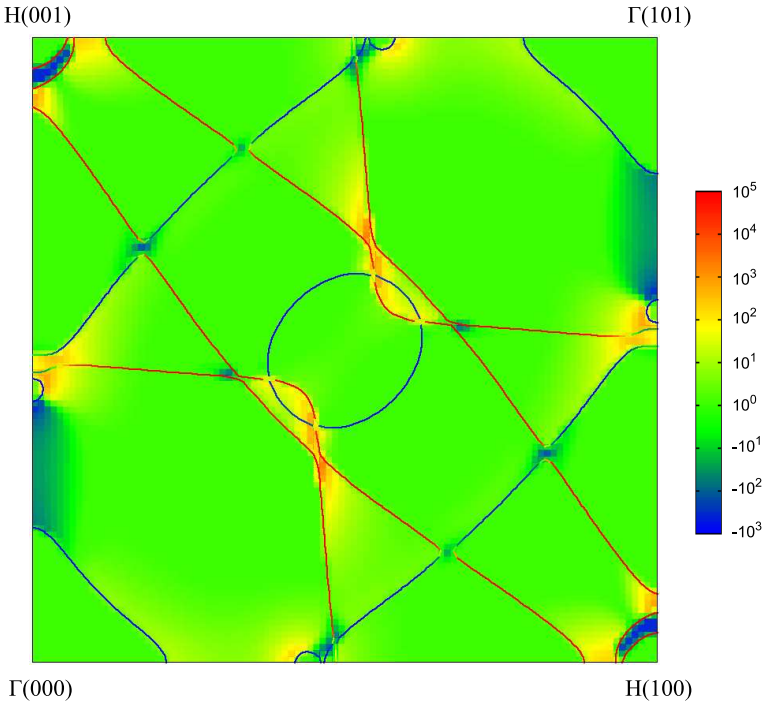

Figure 4.7: Calculated total Berry curvature Ω_z in the plane $k_y = 0$ (note log scale). Intersections of the Fermi surface with this plane are again shown.

the Berry curvature is small (light green regions) fill most of the plane. The largest values occur at the places where two Fermi lines approach one another, consistent with the discussion of Fig. 4.4. Of special importance are the avoided crossings between two bands having the same sign of spin, or between two bands of opposite spin. Examples of both kinds are visible in the figure, and both tend to give rise to very large contributions in the region of the avoided crossing. Essentially, the spin-orbit interaction causes the character of these bands to change extremely rapidly with $\mathbf{k}$ near the avoided crossing; this is the origin of the large Berry curvature. The large contributions near the H points correspond to the peaks that were already mentioned in the discussion of Fig. 4.4, resulting from mixing of nearly degenerate bands by the spin-orbit interaction.

4.4.2 Integrated anomalous Hall conductivity

We now discuss the computation of the AHC as an integral of the Berry curvature over the Brillouin zone, Eq. (4.7). We first define a nominal $N_0 \times N_0 \times N_0$ mesh that uniformly fills the Brillouin zone. We next reduce this to a sum over the irreducible wedge that fills $\frac{1}{16}$th of the Brillouin zone, using the tetragonal point-group symmetry (broken from cubic by the onset of ferromagnetism), and calculate Ω_z on each mesh point using Eq. (4.32). Finally, following Yao *et al.* [21], we implement an adaptive mesh refinement scheme in which we identify those points of the k-space mesh at which the computed Berry curvature exceeds a threshold value $\Omega_{\rm cut}$, and recompute Ω_z on an $N_a \times N_a \times N_a$ submesh spanning the original cell associated with this mesh point. The AHC is then computed as a sum of Ω_z over this adaptively refined mesh with appropriate weights.

The convergence of the AHC with respect to the choice of mesh is presented in Table 4.1. We have chosen $\Omega_{\rm cut} = 1.0 \times 10^2$ a.u., which causes the adaptive mesh refinement to be triggered at approximately 0.11% of the original mesh points. Based on the results of Table 4.1, we estimate the converged value to be $\sigma_{xy} = 756\,(\Omega\ {\rm cm})^{-1}$. This agrees to within 1% with the value of

Table 4.1: Convergence of AHC with respect to the density of the nominal k-point mesh (left column) and the adaptive refinement scheme used to subdivide the mesh in regions of large contributions (middle column).

k-point mesh	Adaptive refinement	σ_{xy} $(\Omega\ \mathrm{cm})^{-1}$
$200 \times 200 \times 200$	$3 \times 3 \times 3$	766.94
$250 \times 250 \times 250$	$3 \times 3 \times 3$	767.33
$320 \times 320 \times 320$	$3 \times 3 \times 3$	768.29
$200 \times 200 \times 200$	$5 \times 5 \times 5$	758.35
$250 \times 250 \times 250$	$5 \times 5 \times 5$	758.84
$320 \times 320 \times 320$	$5 \times 5 \times 5$	759.25
$200 \times 200 \times 200$	$7 \times 7 \times 7$	756.25
$250 \times 250 \times 250$	$7 \times 7 \times 7$	757.32
$320 \times 320 \times 320$	$7 \times 7 \times 7$	757.59
$320 \times 320 \times 320$	$9 \times 9 \times 9$	757.08
$320 \times 320 \times 320$	$11 \times 11 \times 11$	756.86
$320 \times 320 \times 320$	$13 \times 13 \times 13$	756.76

Table 4.2: Contributions to the AHC coming from different regions of the Brillouin zone, as defined in the text.

ΔE (eV)	like-spin (%)	opposite-spin (%)	smooth (%)
0.1	21	26	53
0.2	23	51	26
0.5	30	68	2

751 $(\Omega\ \mathrm{cm})^{-1}$ reported previously in Ref. [21], where an adaptive mesh refinement was also used. As discussed in Ref. [21], this value is in reasonable agreement with the available measurements [86, 87], which yield a value for σ_{xy} slightly above $1000\,(\Omega\ \mathrm{cm})^{-1}$.

It can be seen from Table 4.1 that a $200 \times 200 \times 200$ mesh with $3 \times 3 \times 3$ refinement brings us within $\sim$1% of the converged value. It is also evident that the level of refinement is more important than the fineness of the nominal mesh; a $200 \times 200 \times 200$ mesh with $5 \times 5 \times 5$ adaptive refinement yields a result that is within 0.2% of the converged value, better than a $320 \times 320 \times 320$ mesh with a lower level of refinement.

It is interesting to decompose the total AHC into contributions coming from different parts of

the Brillouin zone. For example, as we saw in Fig. 4.7, there is a smooth, low-intensity background that fills most of the volume of the Brillouin zone, and it is hard to know *a priori* whether the total AHC is dominated by these contributions or by the much larger ones concentrated in small regions. With this motivation, we have somewhat arbitrarily divided the Brillouin zone into three kinds of regions, which we label as 'smooth', 'like-spin', and 'opposite-spin'. To do this, we identify k-points at which there is an occupied band in the interval $[E_f - \Delta E, E_f]$ and an unoccupied band in the interval $[E_f, E_f + \Delta E]$, where ΔE is arbitrarily chosen to be a small energy such as 0.1, 0.2, or 0.5 eV. If so, the k-point is said to belong to the 'like-spin' or 'opposite-spin' region depending on whether the dominant characters of the two bands below and above the Fermi energy are of the same or of opposite spin. Otherwise, the k-point is assigned to the 'smooth' region. As shown in Table 4.2, the results depend strongly on the value of ΔE. Overall, what is clear is that the major contributions arise from the bands within ± 0.5 eV of E_f, and that neither like-spin nor opposite-spin contributions are dominant.

Next, we return to the discussion of the decomposition of the total Berry curvature in Eq. (4.32) into the $\overline{\Omega}$, D–$\overline{A}$, and D–D terms. We find that these three terms account for -0.39%, 1.36%, and 99.03%, respectively, of the total AHC. (Similarly, for the alternative decomposition of Appendix A, the second term of Eq. (4.58) is found to be responsible for more than 99% of the total.) Thus, if a 1% accuracy is acceptable, one could actually neglect the $\overline{\Omega}$ and D–$\overline{A}$ terms entirely, and approximate the total AHC by the D–D (Kubo-like) term alone, Eq. (4.34).

From a computational point of view, the fact that the D–D term is fully specified by the Hamiltonian matrix elements alone means that considerable savings can be obtained by avoiding the evaluation of the Fourier transforms in Eqs. (4.39–4.40) at every interpolation point (and avoiding the setup of the matrix elements $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle$, which can be costly in a real-space implementation). More importantly, this observation, if it turns out to hold for other materials as well, could prove

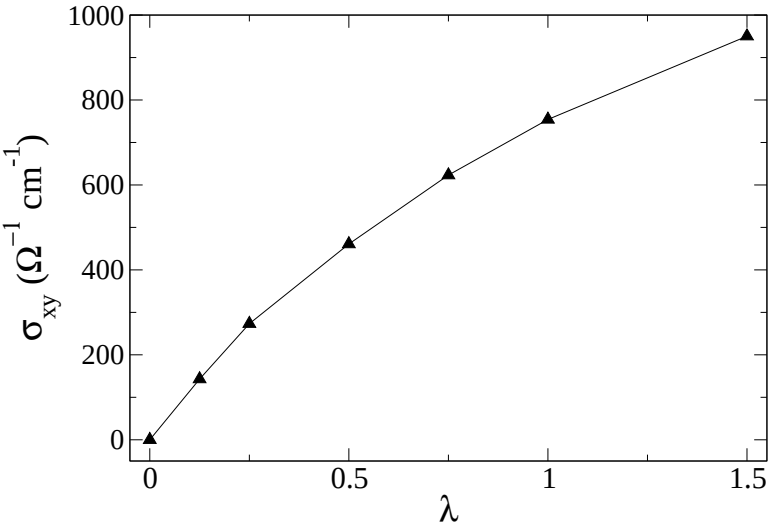

Figure 4.8: Anomalous Hall conductivity vs. spin-orbit coupling strength.

to be important for future efforts to derive approximate schemes capable of capturing the most important contributions to the AHC.

Finally, we investigate how the total AHC depends upon the strength of the spin-orbit interaction, following the approach of Sec. 4.3.3 to modulate the spin-orbit strength. The result is shown in Fig. 4.8. We emphasize that our approach is a more specific test of the dependence upon spin-orbit strength than the one carried out in Ref. [21]; there, the speed of light c was varied, which entails changing the strength of the various scalar relativistic terms as well. Nevertheless, both studies lead to a similar conclusion: the variation is found to be linear for small values of the spin-orbit coupling ($\lambda \ll 1$), while quadratic or other higher-order terms also become appreciable when the full interaction is included ($\lambda = 1$).

4.4.3 Computational considerations

The computational requirements for this scheme are quite modest. The self-consistent ground state calculation and the construction of the WFs takes 2.5 hours on a single 2.2GHz AMD-Opteron

processor. The expense of computing the AHC as a sum over interpolation mesh points depends strongly on the density of the mesh. On the same processor as above, the average CPU time to evaluate Ω_z on each k-point was about 14 msec. We find that the mesh refinement operation does not significantly increase the total number of k-point evaluations until the refinement level N_a exceeds $\sim$10. Allowing for the fact that the calculation only needs to be done in the irreducible $\frac{1}{16}$th of the Brillouin zone, the cost for the AHC evaluation on a $200\times200\times200$ mesh is about 2 hours.

The CPU time per k-point evaluation is dominated (roughly 90%) by the Fourier transform operations needed to construct the objects in Eqs. (4.37–4.40). The diagonalization of the 18×18 Hamiltonian matrix, and other operations needed to compute Eq. (4.32), account for only about 10% of the time. The CPU requirement for the Fourier transform step is roughly proportional to the number of $\mathbf{R}$ vectors kept in Eqs. (4.37–4.40); it is possible that this number could be reduced by exploring more sophisticated methods for truncating the contributions coming from the more distant $\mathbf{R}$ vectors.

Of course, the loop over k-points in the AHC calculation is trivial to parallelize, so for dense k-meshes we speed up this stage of the calculation by distributing across multiple processors.

4.5 Summary and Discussion

In summary, we have developed an efficient method for computing the intrinsic contribution to the anomalous Hall conductivity of a metallic ferromagnet as a Brillouin-zone integral of the Berry curvature. Our approach is based on Wannier interpolation, a powerful technique for evaluating properties that require a very dense sampling of the Brillouin zone or Fermi surface. The key idea is to map the low-energy first-principles electronic structure onto an "exact tight-binding model" in the basis of appropriately constructed Wannier functions, which are typically partially occupied. In the Wannier representation the desired quantities can then be evaluated at arbitrary k-points at

very low computational cost. All that is needed is to evaluate, once and for all, the Wannier-basis matrix elements of the Hamiltonian and a few other property-specific operators (namely, for the Berry curvature, the three Cartesian position operators).

When evaluating the Berry curvature in this way, the summation over all unoccupied bands and the expensive calculation of the velocity matrix elements needed in the traditional Kubo formula are circumvented. They are replaced by quantities defined strictly within the projected space spanned by the WFs. Our final expression for the total Berry curvature, Eq. (4.32), consists of three terms, namely, the $\overline{\Omega}$, D–$\overline{A}$, and D–D terms.

We have applied this approach to calculate the AHC of bcc Fe. While our Wannier interpolation formalism, with its decomposition (4.32), is entirely independent of the choice of an all-electron or pseudopotential method, we have chosen here a relativistic pseudopotential approach [76] that includes scalar relativistic effects as well as the spin-orbit interaction. We find that this scheme successfully reproduces the fine details of the electronic structure and of the Berry curvature. The resulting AHC is in excellent agreement with a previous calculation [21] that used an all-electron LAPW method [88].

Remarkably, we found that more than 99% of the integrated Berry curvature is concentrated in the D–D term of our formalism. This term, given explicitly in Eq. (4.34), takes the form of a Kubo-like Berry curvature formula for the "tight-binding states." Unlike the $\overline{\Omega}$ and D–$\overline{A}$ terms, it depends exclusively on the Hamiltonian matrix elements between the Wannier orbitals, and not on the position matrix elements. Thus we arrive at the very appealing result that a Kubo picture defined within the "tight-binding space" gives an excellent representation of the Berry curvature in the original *ab-initio* space. This result merits further investigation.

Several directions for future studies suggest themselves. For example, it would be desirable to obtain a better understanding of how the AHC depends on the weak spin-orbit interaction. As

we have seen, this weak interaction causes splittings and avoided crossings that give rise to very large Berry curvatures in very small regions of k-space. There is a kind of paradox here. Our numerical tests, as in Fig. 4.8, demonstrate that the AHC falls smoothly to zero as the spin-orbit strength λ is turned off, suggesting that a perturbation theory in λ should be applicable. However, in the limit that λ becomes small, the full calculation becomes *more difficult*, not less: the splittings occur in narrower and narrower regions of k-space, energy denominators become smaller, and Berry curvature contributions become larger (see Fig. 4.4), even if the *integrated* contribution is going to zero. It would be of considerable interest, therefore, to explore ways to reformulate the perturbation theory in λ so that the expansion coefficients can be computed in a robust and efficient fashion. Because the exchange splitting is much larger than the spin-orbit splitting, it may also be of use to introduce two separate couplings that control the strengths of the spin-flip and spin-conserving parts of the spin-orbit interaction respectively, and to work out the perturbation theory in these two couplings independently.

Another promising direction is to explore whether the AHC can be computed as a Fermi-surface integral using the formulation of Haldane [69] in which an integration by parts is used to convert the volume integral of the Berry curvature to a Fermi-surface integral involving Berry curvatures or potentials. Such an approach promises to be more efficient than the volume-integration approach, provided that a method can be developed for carrying out an appropriate sampling of the Fermi surface. This is likely to be a delicate problem, however, since the weak spin-orbit splitting causes Fermi sheets to separate and reattach in a complex way at short k-scales, and the dominant contributions to the AHC are likely to come from precisely these portions of the reconstructed Fermi surface that are the most difficult to describe numerically.

Finally, it would be of considerable interest to generalize the Wannier-interpolation techniques developed here for the dc anomalous Hall effect to treat finite-frequency magneto-optical effects.

In any case, even without such further developments, the present approach is a powerful one. It reduces the expense needed to do an extremely fine sampling of Fermi-surface properties to the level where the AHC of a material like bcc Fe can be computed on a workstation in a few hours. This opens the door to realistic calculations of the intrinsic anomalous Hall conductivity of much more complex materials. More generally, the techniques developed here for the AHE are readily applicable to other problems which also require a very dense sampling of the Fermi surface or Brillouin zone. For example, an extension of these ideas to the evaluation of the electron-phonon coupling matrix elements by Wannier interpolation is currently under way [89].

4.6 Appendix: Alternative expression for the Berry curvature

In this Appendix, we return to Eq. (4.27) and rewrite it in such a way that all of the large, rapidly varying contributions arising from small energy denominators in the expression for D_α, Eq. (4.24), are segregated into a single term. We do this by solving Eq. (4.25) for D_α and substituting into Eq. (4.27) to obtain

$$\Omega_{\alpha\beta}^{(\mathrm{H})} = \overline{\Omega}_{\alpha\beta}^{(\mathrm{H})} - i\left[\overline{A}_\alpha^{(\mathrm{H})}, \overline{A}_\beta^{(\mathrm{H})}\right] + i\left[A_\alpha^{(\mathrm{H})}, A_\beta^{(\mathrm{H})}\right] \ . \tag{4.55}$$

Then only the last term will contain the large, rapid variations. This equation could have been anticipated based on the fact that the tensor

$$\widetilde{\Omega}_{\alpha\beta} = \Omega_{\alpha\beta} - i[A_\alpha, A_\beta] \tag{4.56}$$

is well known to be a gauge-covariant quantity [90, 75]; applying Eq. (4.21) to $\widetilde{\Omega}_{\alpha\beta}$ then leads directly to Eq. (4.55).

This formulation provides an alternative route to the calculation of the matrix $\Omega_{\alpha\beta}^{(\mathrm{H})}$: evaluate $\widetilde{\Omega}_{\alpha\beta}^{(\mathrm{W})}$ in the Wannier representation using Eqs. (4.59-4.60) below, convert it to $\widetilde{\Omega}_{\alpha\beta}^{(\mathrm{H})}$ via Eq. (4.21),

compute $A_\alpha^{(\mathrm{H})}$ using Eq. (4.25), and assemble

$$\Omega_{\alpha\beta}^{(\mathrm{H})} = \widetilde{\Omega}_{\alpha\beta}^{(\mathrm{H})} + i[A_\alpha^{(\mathrm{H})}, A_\beta^{(\mathrm{H})}] \; . \tag{4.57}$$

The large and rapid variations then appear only in the last term involving commutators of the A matrices.

In Sec. 4.2.4, we showed how to write the total Berry curvature $\Omega_{\alpha\beta}(\mathbf{k})$ as a sum over bands in such a way that potentially troublesome contributions coming from small energy denominators between pairs of occupied bands are explicitly excluded, leading to Eq. (4.32). The corresponding expression based on Eq. (4.57) is

$$\begin{aligned} \Omega_{\alpha\beta}(\mathbf{k}) &= \sum_n f_n \, \widetilde{\Omega}_{nn,\alpha\beta}^{(\mathrm{H})} \\ &\quad + \sum_{nm} (f_n - f_m) \, A_{nm,\alpha}^{(\mathrm{H})} A_{mn,\beta}^{(\mathrm{H})} \; . \end{aligned} \tag{4.58}$$

Now, in addition to the four quantities given in Eqs. (4.37-4.40), we need a corresponding equation for $\widetilde{\Omega}_{\alpha\beta}$. After some manipulations, we find that

$$\widetilde{\Omega}_{nn,\alpha\beta}^{(\mathrm{W})}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \, w_{n,\alpha\beta}(\mathbf{R}) \tag{4.59}$$

where

$$\begin{aligned} w_{n,\alpha\beta}(\mathbf{R}) &= -i \sum_{\mathbf{R}'m} \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}'m\rangle\langle\mathbf{R}'m|\hat{r}_\beta|\mathbf{R}n\rangle \\ &\quad + i \sum_{\mathbf{R}'m} \langle \mathbf{0}n|\hat{r}_\beta|\mathbf{R}'m\rangle\langle\mathbf{R}'m|\hat{r}_\alpha|\mathbf{R}n\rangle \; . \end{aligned} \tag{4.60}$$

This formulation again requires the same basic ingredients as before, namely, the Wannier matrix elements of $\hat{H}$ and $\hat{r}_\alpha$. In some respects it is a little more elegant than the formulation of Eq. (4.32). However, the direct evaluation of $w_{n,\alpha\beta}$ in the Wannier representation, as given in Eq. (4.60), is not as convenient because of the extra sum over intermediate WFs appearing there; moreover, $w_{n,\alpha\beta}$ is

longer-ranged than the Hamiltonian and coordinate matrix elements. Also, one appealing feature of the formulation of Section 4.2, that more than 99% of the effect can be recovered without using the position-operator matrix elements, is lost in this reformulation. We have therefore chosen to base our calculations and analysis on Eq. (4.32) instead.

It is informative to obtain Eq. (4.57) in a different way: define the gauge-invariant band projection operator [75] $\hat{P}_{\mathbf{k}} = \sum_{n=1}^{M} |u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|$ and its complement $\hat{Q}_{\mathbf{k}} = 1 - \hat{P}_{\mathbf{k}}$. Inserting $\hat{1} = \hat{Q}_{\mathbf{k}} + \hat{P}_{\mathbf{k}}$ into Eq. (4.18) in the Hamiltonian gauge then yields directly Eq. (4.57) since, as can be easily verified, Eq. (4.56) may be written as

$$\widetilde{\Omega}_{nm,\alpha\beta} = i\langle\widetilde{\partial}_\alpha u_n|\widetilde{\partial}_\beta u_m\rangle - i\langle\widetilde{\partial}_\beta u_n|\widetilde{\partial}_\alpha u_m\rangle \ , \tag{4.61}$$

where $\widetilde{\partial}_\alpha \equiv \hat{Q}\partial_\alpha$. The gauge-covariance of $\widetilde{\Omega}_{\alpha\beta}$ follows directly from the fact that $\widetilde{\partial}_\alpha$ is a gauge-covariant derivative, in the sense that $|\widetilde{\partial}_\alpha u_n^{(\mathrm{H})}\rangle = \sum_{m=1}^{M} |\widetilde{\partial}_\alpha u_m^{(\mathrm{W})}\rangle U_{mn}$ is the same transformation law as Eq. (4.15) for the Bloch states themselves. It is apparent from this derivation that as the number M of WFs increases and $\hat{P}_{\mathbf{k}}$ approaches $\hat{1}$, the second term on the right-hand side of Eq. (4.58) increases at the expense of the first term. Indeed, in the large-M limit the entire Berry curvature is contained in the second term. For the choice Wannier orbitals described in the main text for bcc Fe, that term already accounts for 99.8% of the total AHC.

4.7 Appendix: Finite-difference approach

In this Appendix, we outline an alternative scheme for computing the AHC by Wannier interpolation. The essential difference relative to to the approaches described in Section 4.2 and in Appendix 4.6 is that the needed k-space derivatives are approximated here by finite differences instead of being expressed analytically in the Wannier representation.

This approach is most naturally applied to the zero-temperature limit where there are exactly $N_{\mathbf{k}}$

occupied states at a given $\mathbf{k}$. Instead of starting from the Berry curvature of each individual band separately, as in Eq. (4.6), we find it convenient here to work from the outset with the total Berry curvature

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{N_{\mathbf{k}}} \Omega_{nn,\alpha\beta}(\mathbf{k}) \tag{4.62}$$

of the occupied manifold at $\mathbf{k}$ (the zero-temperature limit of Eq. (4.19)). We now introduce a covariant derivative $\widetilde{\partial}_\alpha^{(N_{\mathbf{k}})} = \hat{Q}_{\mathbf{k}}^{(N_{\mathbf{k}})}\partial_\alpha$ designed to act on the occupied states only; here $\hat{Q}_{\mathbf{k}}^{(N_{\mathbf{k}})} = \hat{1} - \hat{P}_{\mathbf{k}}^{(N_{\mathbf{k}})}$ and $\hat{P}_{\mathbf{k}}^{(N_{\mathbf{k}})} = \sum_{n=1}^{N_{\mathbf{k}}} |u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|$. The only difference with respect to the definition of $\widetilde{\partial}_\alpha$ in Appendix 4.6 is that the projection operator here spans the $N_{\mathbf{k}}$ occupied states only, instead of the M states of the full projected space. Accordingly, terms such as "gauge-covariance" and "gauge-invariance" are to be understood here in a restricted sense. For example, the statement that $\widetilde{\partial}_\alpha^{(N_{\mathbf{k}})}$ is a gauge-covariant derivative means that under an $N_{\mathbf{k}} \times N_{\mathbf{k}}$ unitary rotation $\mathcal{U}(\mathbf{k})$ between the occupied states at $\mathbf{k}$ it obeys the transformation law

$$|\widetilde{\partial}_\alpha^{(N_{\mathbf{k}})} u_{n\mathbf{k}}\rangle \to \sum_{m=1}^{N_{\mathbf{k}}} |\widetilde{\partial}_\alpha^{(N_{\mathbf{k}})} u_{m\mathbf{k}}\rangle\, \mathcal{U}_{mn}(\mathbf{k}). \tag{4.63}$$

(We will use calligraphic symbols to distinguish $N_{\mathbf{k}} \times N_{\mathbf{k}}$ matrices such as $\mathcal{U}$ from their $M \times M$ counterparts such as U.) We now define a gauge-covariant curvature $\widetilde{\Omega}_{\alpha\beta}^{(N_{\mathbf{k}})}(\mathbf{k})$ by replacing $\widetilde{\partial}$ by $\widetilde{\partial}^{(N_{\mathbf{k}})}$ in Eq. (4.61). Since the trace of a commutator vanishes, it follows from Eq. (4.56) that Eq. (4.62) can be written as

$$\Omega_{\alpha\beta}(\mathbf{k}) = \mathrm{Tr}^{(N_{\mathbf{k}})} \left[\widetilde{\Omega}_{\alpha\beta}^{(N_{\mathbf{k}})}(\mathbf{k}) \right], \tag{4.64}$$

where the symbol $\mathrm{Tr}^{(N_{\mathbf{k}})}$ denotes the trace over the occupied states.

The advantage of this expression over Eq. (4.62) is that the covariant derivative of a Bloch state can be approximated by a very robust finite-differences formula [39, 91]:

$$\tilde{\partial}_{\mathbf{k}}^{(N_{\mathbf{k}})} \to \sum_{\mathbf{b}} w_b \mathbf{b} \hat{P}_{\mathbf{k},\mathbf{b}}^{(N_{\mathbf{k}})}\ , \tag{4.65}$$

where the sum is over shells of neighboring k-points [75], as in Eq. (4.44), and we have defined the gauge-invariant operator

$$\hat{P}^{(N_{\mathbf{k}})}_{\mathbf{k},\mathbf{b}} = \sum_{n=1}^{N_{\mathbf{k}}} |\widetilde{u}_{n,\mathbf{k}+\mathbf{b}}\rangle\langle u_{n\mathbf{k}}| \tag{4.66}$$

in terms of the gauge-covariant "dual states"

$$|\widetilde{u}_{n,\mathbf{k},\mathbf{b}}\rangle = \sum_{m=1}^{N_{\mathbf{k}}} |u_{m,\mathbf{k}+\mathbf{b}}\rangle \left(\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}}\right)_{mn} \ . \tag{4.67}$$

Here $\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}}$ is the inverse of the $N_{\mathbf{k}} \times N_{\mathbf{k}}$ overlap matrix,

$$\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}} = \left(\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\right)^{-1} \ , \tag{4.68}$$

where

$$\left(\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\right)_{nm} = \langle u_{n\mathbf{k}}|u_{m,\mathbf{k}+\mathbf{b}}\rangle \ . \tag{4.69}$$

The discretization (4.65) is immune to arbitrary gauge phases and unitary rotations among the occupied states. Because of that property, the occurrence of band crossings and avoided crossings does not pose any special problems.

Inserting Eqs. (4.65-4.69) into Eq. (4.64) and using $\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}} = \mathcal{Q}^{\dagger}_{\mathbf{k}+\mathbf{b},\mathbf{k}}$, we find that an appropriate finite-difference expression for the total Berry curvature is

$$\Omega^{(N_{\mathbf{k}})}_{\alpha\beta}(\mathbf{k}) = 2 \sum_{\mathbf{b}_1,\mathbf{b}_2} w_{b_1}\, w_{b_2}\, b_{1,\alpha}\, b_{2,\beta}\, \Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2}, \tag{4.70}$$

where

$$\Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2} = -\mathrm{Im}\,\mathrm{Tr}^{(N_{\mathbf{k}})}\left[\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}_1}\mathcal{S}_{\mathbf{k}+\mathbf{b}_1,\mathbf{k}+\mathbf{b}_2}\mathcal{Q}_{\mathbf{k}+\mathbf{b}_2,\mathbf{k}}\right] \ . \tag{4.71}$$

This expression is manifestly gauge-invariant, since both $\mathcal{S}$ and $\mathcal{Q}$ are gauge-covariant matrices, i.e., $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}} \to \mathcal{U}^{\dagger}(\mathbf{k})\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\mathcal{U}(\mathbf{k}+\mathbf{b})$, and the same transformation law holds for $\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$.

Eqs. (4.70-4.71) can be evaluated at an arbitrary point $\mathbf{k}$ once the overlap matrices $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$ are known. For that purpose we construct a uniform mesh of spacing Δk in the immediate vicinity of

$\mathbf{k}$, set up the needed shells of neighboring k-points $\mathbf{k}+\mathbf{b}$ on that local mesh, and then evaluate $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$ by Wannier interpolation. Since the WFs span the entire M-dimensional projected space, at this stage we revert to the full $M \times M$ overlap matrices $S_{\mathbf{k},\mathbf{k}+\mathbf{b}}$. In the Wannier gauge they are given by a Fourier transform of the form

$$\left(S^{(\mathrm{W})}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\right)_{nm} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}n | e^{i\mathbf{b}\cdot(\mathbf{R}-\hat{\mathbf{r}})} | \mathbf{R}m \rangle \ . \tag{4.72}$$

For sufficiently small Δk, this can be approximated as

$$\left(S^{(\mathrm{W})}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\right)_{nm} \simeq \delta_{nm} - i\mathbf{b} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}n | \hat{\mathbf{r}} | \mathbf{R}m \rangle \ . \tag{4.73}$$

Note that the dependence of the last expression on Δk is trivial, since it only enters as a multiplicative prefactor. In practice one chooses Δk to be quite small, $\sim 10^{-6}$ a.u.$^{-1}$, so as to reduce the error of the finite-differences expression.

In the Wannier gauge the occupied and empty states are mixed with one another, because the WFs are partially occupied. In order to decouple the two subspaces we perform the unitary transformation

$$S^{(\mathrm{H})}_{\mathbf{k},\mathbf{k}+\mathbf{b}} = U^{\dagger}(\mathbf{k}) S^{(\mathrm{W})}_{\mathbf{k},\mathbf{k}+\mathbf{b}} U(\mathbf{k}+\mathbf{b}) \ . \tag{4.74}$$

This produces the full $M \times M$ overlap matrix in the Hamiltonian gauge. The $N_{\mathbf{k}} \times N_{\mathbf{k}}$ submatrix in the upper left corner is precisely the matrix $\mathcal{S}^{(\mathrm{H})}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$ needed in Eq. (4.71).

Like the approach described in the main text, this approach still only requires the WF matrix elements of the four operators $\hat{H}$ and $\hat{r}_\alpha$ ($\alpha = x$, y, and z). We have implemented it, and have checked that the results agree closely with those obtained using using the method of the main text. Although not as elegant, this approach has the interesting feature of circumventing the evaluation of the matrix $D^{(\mathrm{H})}_{\alpha}$, Eq. (4.24). This may be advantageous in certain special situations. For example, if a parameter such as pressure is tuned in such a way that a k-space Dirac monopole [13] drifts to

the Fermi surface, the vanishing of the energy denominator in Eq. (4.24) may result in a numerical instability when trying to find the monopole contribution to the AHC.

We conclude by noting that Eq. (4.71) is but one of many possible finite-differences expressions, and may not even be the most convenient one to use in practice. By recalling that the Berry curvature is the Berry phase per unit area, one realizes that in the small-Δk limit of interest, the quantity $\Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2}$ in Eq. (4.70) can be viewed as the discrete Berry phase ϕ accumulated along the small loop $\mathbf{k} \to \mathbf{k}+\mathbf{b}_1 \to \mathbf{k}+\mathbf{b}_2 \to \mathbf{k}$. As is well-known, the Berry phase around a discrete loop is defined as [20]

$$\phi = -\,\mathrm{Im}\,\ln\det\left[\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}_1}\mathcal{S}_{\mathbf{k}+\mathbf{b}_1,\mathbf{k}+\mathbf{b}_2}\mathcal{S}_{\mathbf{k}+\mathbf{b}_2,\mathbf{k}}\right]\ . \tag{4.75}$$

It can be shown that $\phi = \Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2} + \mathcal{O}(\Delta k^2)$, so that for small loops the two formulas agree. Eq. (4.75) has the practical advantage over Eq. (4.71) that it does not require inverting the overlap matrix.

Chapter 5

Fermi-surface calculation of the anomalous Hall conductivity

It is by now well established that the intrinsic Karplus-Luttinger mechanism [16] plays a significant role in the anomalous Hall conductivity (AHC) of ferromagnets. This contribution can be expressed as an integral of the k-space Berry curvature over the occupied portions of the Brillouin zone (BZ) [66, 67, 68, 15]. First-principles calculations of the intrinsic AHC have been carried out by several authors, using either a Kubo linear-response formula [13, 21] or a direct "geometric" evaluation of the Berry curvature in Chapter 4, and achieving good agreement with experimental values for several ferromagnets. These studies revealed that the Berry curvature is very sharply peaked in certain regions of the BZ where spin-orbit splitting occurs near the Fermi level. As a result the calculations tend to be rather demanding; in the case of bcc Fe, for example, millions of k-points must be sampled to achieve convergence [21]. More efficient approaches are therefore highly desirable.

In the preceding chapter, we developed a strategy for calculating the AHC in which Wannier interpolation of the Bloch functions was used to circumvent the need to perform a full first-principles calculation for every k-point. Thus, while the required number of k-points was not reduced, the computational load per k-point was greatly reduced. In this approach, the actual first-principles calculations are performed on a comparatively coarse k-mesh. Then, in a postprocessing step, the calculated electronic structure is mapped onto an "exact tight-binding model" based on maximally-localized Wannier functions [72]. Working in the Wannier representation, the Berry curvature can then be evaluated very inexpensively at each of the k-points of the fine mesh needed for accurate evaluation of the AHC.

Recently, Haldane has shown that while the intrinsic AHC is usually regarded as a Fermi-sea

property of all the occupied states, it can alternatively, and in some ways more naturally, be regarded as a Fermi-surface property [69]. (More precisely, Haldane showed that these quantities are equal modulo the quantum of transverse conductivity that is well-known from the quantum Hall effect, since one cannot rule out the possibility that, e.g., some occupied bands carry non-zero Chern numbers [69]). By a kind of integration by parts, Haldane showed how the integral of the Berry curvature over the occupied portions of the BZ could be manipulated first into a Fermi-surface integral of a Berry connection, and then ultimately into a Fermi-surface integral of a Fermi-vector-weighted Berry curvature, augmented with some Berry-phase corrections for the case of non-simply-connected Fermi sheets.

In this chapter, we present a tractable and efficient computational scheme based on a Fermi-surface formulation of the AHC. While following the basic spirit of the Haldane idea, we proceed along a slightly different path. In our approach, the BZ is divided into a fine mesh of equally-spaced slices normal to the direction of the magnetization, and the integral of the Berry curvature over the occupied states of a given slice is transformed into a sum of Berry phases of Fermi loops lying in that slice. As a result, the three-dimensional BZ integration is avoided, and the method relies instead only on information calculated on the two-dimensional Fermi surface. As in Chapter 4, an important ingredient of our approach is the use of a Wannier interpolation scheme to lower the cost further by eliminating the need for a full first-principles evaluation at each point on the Fermi surface. Combining these two complementary strategies, we arrive at a robust and efficient method for computation of the AHC in ferromagnetic metals.

The chapter is organized as follows. In Sec. 5.1 we present the necessary formulas relating Berry phases on the Fermi surface to the AHC, as well as their evaluation in the Wannier representation. The details of the first-principles calculations and the determination of the Fermi loops are given in Sec. 5.2. In Sec. 5.3 the method is applied to the transition metals Fe, Co and Ni. A discussion of

issues of computational efficiency is given in Sec. 5.4, followed by a brief conclusion in Sec. 5.5.

5.1 Method

5.1.1 Fermi-loop formula

Our starting point is the AHC expressed as an antisymmetric Cartesian tensor in terms of the Berry curvature,

$$\sigma_{\alpha\beta} = -\frac{e^2}{\hbar} \sum_n \int_{\mathrm{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} \, f_n(\mathbf{k}) \, \Omega_{n,\alpha\beta}(\mathbf{k}) \ , \tag{5.1}$$

where the integration is over the three-dimensional BZ and the occupation function $f_n(\mathbf{k})$ restricts the sum to the occupied states (we work at zero temperature). $\Omega_{n,\alpha\beta}(\mathbf{k})$ is the Berry-curvature matrix of band n, defined as

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2 \, \mathrm{Im} \left\langle \frac{\partial u_{n\mathbf{k}}}{\partial k_\alpha} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_\beta} \right\rangle \tag{5.2}$$

where $u_{n\mathbf{k}}$ is the periodic part of the Bloch function $\psi_{n\mathbf{k}}$. Because $\Omega_{n,\alpha\beta}$ is antisymmetric, we can represent it instead in axial-vector notation as

$$\Omega_{n\gamma} = \frac{1}{2} \sum_{\alpha\beta} \epsilon_{\alpha\beta\gamma} \Omega_{n,\alpha\beta} \ , \tag{5.3}$$

or equivalently, $\Omega_{n,\alpha\beta} = \sum_\gamma \epsilon_{\alpha\beta\gamma} \Omega_{n\gamma}$, where $\epsilon_{\alpha\beta\gamma}$ is the antisymmetric tensor. The Berry curvature can also be written as

$$\mathbf{\Omega}_n(\mathbf{k}) = \nabla_{\mathbf{k}} \times \mathbf{A}_n(\mathbf{k}) \tag{5.4}$$

where the Berry connection is

$$\mathbf{A}_n(\mathbf{k}) = i \langle u_{n\mathbf{k}} | \nabla_{\mathbf{k}} | u_{n\mathbf{k}} \rangle \ . \tag{5.5}$$

Following Ref. [69], we rewrite Eq. (5.1) as

$$\sigma_{\alpha\beta} = \frac{-e^2}{\hbar} \frac{1}{(2\pi)^2} \sum_{n\gamma} \epsilon_{\alpha\beta\gamma} K_{n\gamma} \tag{5.6}$$

where

$$\mathbf{K}_n = \frac{1}{2\pi} \int_{\mathrm{BZ}} d\mathbf{k}\, f_n(\mathbf{k})\, \mathbf{\Omega}_n(\mathbf{k}) \ . \tag{5.7}$$

For the case of a completely filled band lying entirely below the Fermi level, Haldane has shown [69] that $\mathbf{K}_n$ is quantized to be a reciprocal lattice vector (the "Chern vector"), as will become clear in Sec. 5.1.2 below.

Let $\mathbf{a}_i$ and $\mathbf{b}_i$ be a conjugate set of primitive real-space and reciprocal-space lattice vectors respectively, $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, and let

$$c_{nj} = \frac{1}{2\pi}\, \mathbf{a}_j \cdot \mathbf{K}_n \tag{5.8}$$

so that

$$\mathbf{K}_n = \sum_j c_{nj}\, \mathbf{b}_j \ . \tag{5.9}$$

In order to compute c_{n3}, for example, we choose the BZ to be a prism whose base is spanned by $\mathbf{b}_1$ and $\mathbf{b}_2$ and whose height is $2\pi/a_3$, and convert the integral into one over slices parallel to the base. In general, separate calculations in which the slices are constructed parallel to the $\mathbf{b}_2$-$\mathbf{b}_3$ and $\mathbf{b}_1$-$\mathbf{b}_3$ planes are needed to compute c_{n1} and c_{n2} respectively [92]. However, this can be avoided in the common case that the magnetization lies parallel to a symmetry axis; one can then choose $\mathbf{b}_1$ and $\mathbf{b}_2$ perpendicular to this axis, and only c_{n3} needs to be computed.

Inserting Eq. (5.7) into Eq. (5.8) yields

$$c_{nj} = \frac{a_j}{2\pi} \int_0^{\frac{2\pi}{a_j}} dk_\perp \frac{\phi_n(k_\perp)}{2\pi} \tag{5.10}$$

where

$$\phi_n(k_\perp) = \int_{\mathcal{S}_n(k_\perp)} d^2k\ \hat{a}_j \cdot \mathbf{\Omega}_n(\mathbf{k}) \ . \tag{5.11}$$

Here $k_\perp$ labels the slice and $\mathcal{S}_n(k_\perp)$ is the region of the slice in which band n is occupied. Recalling Eq. (5.4) and noting that $\hat{a}_j$ is the unit vector normal to the slice, the application of Stokes' theorem

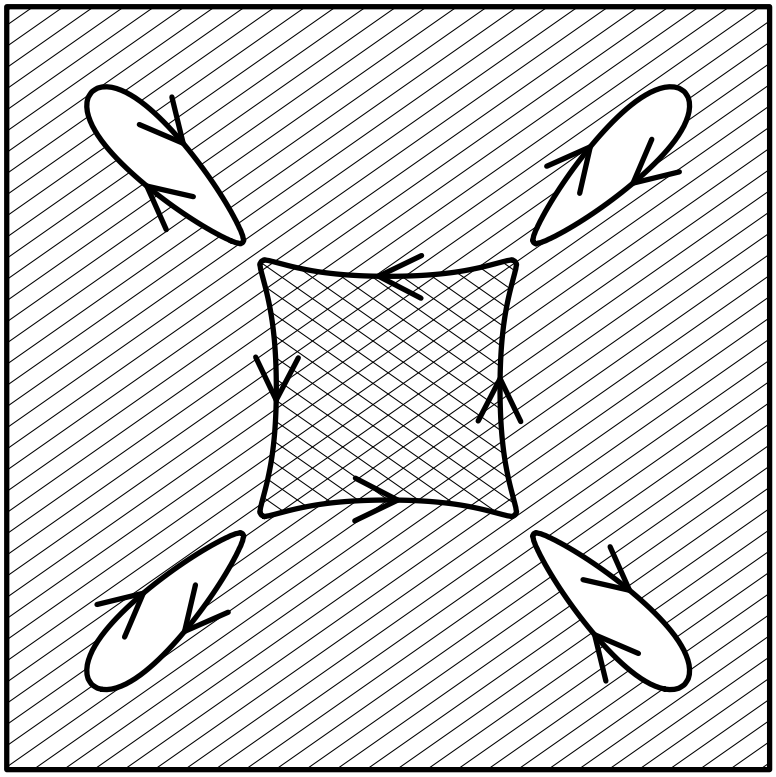

Figure 5.1: Sketch of intersections of the Fermi surface with a constant-$k_\perp$ plane. Open, hashed, and cross-hashed regions correspond to filling of zero, one, and two bands, respectively. The four small Fermi loops belong to the first band, while the large central one belongs to the second. Arrows indicate sense of circulation for performing the Berry-phase integration.

to Eq. (5.11) yields

$$\phi_n(k_\perp) = \oint_{\mathcal{C}_n(k_\perp)} \mathbf{A}_n(\mathbf{k}) \cdot d\mathbf{l} \tag{5.12}$$

where $\mathcal{C}_n(k_\perp)$ is the oriented curve bounding $\mathcal{S}_n(k_\perp)$ on the slice and $\phi_n(k_\perp)$ has the interpretation of a Berry phase. For later convenience we also define

$$\phi(k_\perp) = \sum_n \phi_n(k_\perp) \tag{5.13}$$

and similarly $c_j = \sum_n c_{nj}$ etc. The calculation of the AHC has thus been reduced to a calculation that is restricted to the Fermi surface only, in the spirit of Haldane [69] but using a somewhat different formulation.

In general, the occupied or unoccupied region of band n in slice $k_\perp$ need not be simply connected, in which case the boundary $\mathcal{C}_n(k_\perp)$ is really the union of several loops. Moreover, loops encircling hole pockets should be taken in the negative direction of circulation. This is illustrated in Fig. 5.1, where the first band exhibits four hole pockets and the second band has one electron pocket, so that $\mathcal{C}_1$ is the union of four countercirculating loops and $\mathcal{C}_2$ is a fifth loop of positive circulation. If higher bands are unoccupied, then $\phi(k_\perp)$ for this slice is just given by the sum of the Berry phases of these five loops. We shall assume for simplicity in the following that $\mathcal{C}_n(k_\perp)$ is simply connected, but the generalization to composite loops is straightforward.

5.1.2 The quantum of Hall conductivity

We claimed earlier that if band n is fully occupied, $\mathbf{K}_n$ in Eq. (5.7) is quantized to a reciprocal lattice vector. This can now be seen by noting that under those circumstances the integral in Eq. (5.11) runs over a two-dimensional BZ, which can be regarded as a closed two-dimensional manifold (two-torus), and for topological reasons [93] the integral of the Berry curvature over such a closed manifold must be an integer multiple of 2π (the Chern number). Then each c_{nj} is an integer, and $\mathbf{K}_n$ in Eq. (5.9) must be a reciprocal lattice vector as claimed. If the system is an insulator, then $\mathbf{K} = \sum_n \mathbf{K}_n$ (summed over occupied bands) is also guaranteed to be a reciprocal lattice vector, and if it is a nonzero one, the insulator would have a quantized Hall conductivity and could be regarded as a quantum Hall crystal (or "Chern insulator") [69, 94]. No physical realization of such a system is known experimentally, but the search for one remains an interesting challenge.

Let us consider again a slice for which band n is fully occupied but has a non-zero Chern number. If this slice is regarded as an open rectangle (or parallelogram) rather than a closed two-torus, and a continuous choice of gauge is made in its interior (i.e., $\mathbf{A}_n(\mathbf{k})$ is free of singularities), then the boundary $\mathcal{C}_n(k_\perp)$ is the perimeter of this rectangle and Eq. (5.12) will yield the same

integer multiple of 2π as Eq. (5.11). In the spirit of Fig. 5.1, however, we prefer to regard the slice as a closed two-torus and to exclude the perimeter from our definition of the boundary $\mathcal{C}_n(k_\perp)$. Then $\mathcal{C}_n(k_\perp)$ is null and Eq. (5.12) vanishes for the case at hand, in disagreement with Eq. (5.11). The disagreement arises because of the impossibility of making a continuous choice of gauge on a closed manifold having a non-zero Chern number [93]; the best that can be done is to make $\mathbf{A}_n(\mathbf{k})$ finite everywhere except at singularities ("vortices") which, when included, restore the missing contributions of 2π.

Returning to the general case of a partially occupied band n with $\mathcal{C}_n(k_\perp)$ defined to exclude the perimeter of the slice, we conclude that Eq. (5.12) is really only guaranteed to equal the true result of Eq. (5.11) modulo 2π. Moreover, the Berry phase will be evaluated in practice using a discretized Berry-phase formula [20] of the form

$$\phi_n(k_\perp) = -\mathrm{Im}\ln\prod_j \langle u_{n\mathbf{k}_j}|u_{n\mathbf{k}_{j+1}}\rangle \tag{5.14}$$

where $\mathbf{k}_j$ discretizes the loop $\mathcal{C}_n(k_\perp)$. (We will actually use a modified version, Eq. (5.25), of this formula.) The choice of branch cut is now arbitrary, and again the agreement with Eq. (5.12) or Eq. (5.11) is only guaranteed modulo 2π. By convention one normally restricts phases to lie in the interval $(-\pi, \pi]$, but then $\phi_n(k_\perp)$ would in general have unwanted discontinuities at some values of $k_\perp$. In practice we discretize the $k_\perp$ integration, so that using Eq. (5.10), $c_j = \sum_n c_{nj}$ becomes

$$c_j = \frac{1}{n_{\mathrm{slice}}}\sum_{i=1}^{n_{\mathrm{slice}}}\frac{\phi(i)}{2\pi}\ . \tag{5.15}$$

We then enforce continuity of the total phase $\phi(k_\perp)$ of Eq. (5.13) by choosing $\phi(i)$ such that $|\phi(i) - \phi(i-1)| \ll 2\pi$ for each slice $i = 2, 3, ...$ in sequence. Since the true phase given by the sum of contributions in Eq. (5.11) is also continuous, this guarantees that our calculated $\phi(k_\perp)$ differs from the true one by the same multiple of 2π for all $k_\perp$. Our computed AHC would then differ from the true one by a multiple of the quantum and could be said to give the "non-quantized part" of the

intrinsic AHC in the sense of Haldane [69]. However, it is straightforward to remove this overall ambiguity of branch choice by evaluating $\phi(k_\perp)$ from Eq. (5.11) on the first slice and then enforcing continuity for each subsequent slice, thus arriving at the correct AHC without any question of a quantum.

We note in passing that an isolated point of degeneracy ("Dirac point") between a pair of bands n and $n+1$ can generically occur in three-dimensional k-space in the absence of time-reversal symmetry [69]. If such a Dirac point occurs below the Fermi energy, then $\phi_n(k_\perp)$ and $\phi_{n+1}(k_\perp)$ will, when evaluated from Eq. (5.11), exhibit equal and opposite discontinuities of 2π at the $k_\perp$ of the Dirac point. However, the total phase $\phi(k_\perp)$ will remain continuous, so that the algorithm described in the previous paragraph will still work correctly.

We close this subsection by emphasizing that the discussion of possible non-zero Chern numbers or the presence of Dirac points is rather academic. In our calculations on Fe, Ni and Co, we have not encountered any indications of such anomalies; they presumably occur rarely or not at all in the materials studied here.

5.1.3 Evaluation of the Fermi-loop Berry phase

The essential problem now becomes the computation of the loop integral of Eq. (5.12). As is well known, the Berry connection $\mathbf{A}_n(\mathbf{k})$ of Eq. (5.5) is gauge-dependent, i.e., sensitive to the k-dependent choice of phase of the Bloch functions. If Eq. (5.12) is to be calculated by the direct evaluation of $\mathbf{A}_n(\mathbf{k})$ and its subsequent integration around the loop, this lack of gauge-invariance may present difficulties. For example, it means that there is no unique Kubo-formula expression for $\mathbf{A}_n(\mathbf{k})$. An alternative and more promising approach is to compute $\phi_n(k_\perp)$ by the discretized Berry-phase formula [20] of Eq. (5.14), where the inner products are computed from the full first-principles calculations at neighboring pairs of k-points around the loop. However, this may still

be quite time-consuming if it has to be done at very many k-points. We avoid this by using the technique of Wannier interpolation [72, 95, 96] to perform the needed loop integral inexpensively. In this formulation, the loop integral of Eq. (5.12) can be expressed as a sum of two terms, one in which a contribution to $\mathbf{A}_n(\mathbf{k})$ is evaluated and integrated explicitly, and a second that takes a form like that of Eq. (5.14).

The key idea of Wannier interpolation is to map the low-energy first-principles electronic structure onto an "exact tight-binding model" using a basis of appropriately constructed crystalline Wannier functions. For metallic systems like those considered here, the bands generated by these Wannier functions are only partially occupied. They are guaranteed by construction to reproduce the true first-principles bands in an energy window extending somewhat above the Fermi level, so that all valence and Fermi-surface states are properly described [72]. In the Wannier representation, the desired quantities such as band energies, eigenstates and the derivatives of eigenstates with respect to wavevector k can then be evaluated at arbitrary k-points at very low computational cost. All that is needed is to evaluate, once and for all, the Wannier-basis matrix elements of the Hamiltonian and position operators in Chapter 4. It is worth pointing out that it may sometimes be expedient to drop some lower occupied bands and construct the Wannier functions so that they correctly represent the Bloch functions only in some narrower energy window containing the Fermi energy; since the present formulation involves only Fermi-surface properties, the nonquantized part of the AHC will then still be given correctly.

The Wannier construction procedure of Ref. [72] provides us with a set of M Wannier functions $|\mathbf{R}n\rangle$ ($n = 1, ..., M$) in each cell labeled by lattice vector $\mathbf{R}$. From these the Bloch basis functions $|u_{n\mathbf{k}}^{(\mathrm{W})}\rangle$ are constructed according to the Fourier transform relation

$$|u_{n\mathbf{k}}^{(\mathrm{W})}\rangle = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})}|\mathbf{R}n\rangle \; . \tag{5.16}$$

Here the superscript (W) indicates that these are obtained from the Wannier representation, that is,

they are not yet Hamiltonian eigenstates. To obtain those, we construct the $M \times M$ Hamiltonian matrix

$$H^{(\mathrm{W})}_{nm}(\mathbf{k}) = \langle u^{(\mathrm{W})}_{n\mathbf{k}} | \hat{H}(\mathbf{k}) | u^{(\mathrm{W})}_{m\mathbf{k}} \rangle \tag{5.17}$$

via

$$H^{(\mathrm{W})}_{nm} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \, \langle \mathbf{0}n | \hat{H} | \mathbf{R}m \rangle \, . \tag{5.18}$$

At any given $\mathbf{k}$ this matrix can be diagonalized to yield an $M \times M$ unitary matrix $U_{nm}(\mathbf{k})$, i.e.,

$$U^{\dagger}(\mathbf{k}) H^{(\mathrm{W})}(\mathbf{k}) U(\mathbf{k}) = H^{(\mathrm{H})}(\mathbf{k}) \tag{5.19}$$

where $H^{(\mathrm{H})}(\mathbf{k}) = \mathcal{E}^{(\mathrm{H})}_n \delta_{mn}$ are the energy eigenvalues and

$$|u^{(\mathrm{H})}_{n\mathbf{k}}\rangle = \sum_m |u^{(\mathrm{W})}_{m\mathbf{k}}\rangle U_{mn}(\mathbf{k}) \tag{5.20}$$

are the corresponding band states. By the construction procedure of Ref. [72], $\mathcal{E}^{(\mathrm{H})}_n$ is identical to the true $\mathcal{E}_n$ (and similarly for the eigenvectors $u^{(\mathrm{H})}_{n\mathbf{k}}$) for all occupied states and low-lying empty states. This is strictly true only for k-points on the original *ab-initio* mesh. The power of this interpolation scheme lies in the fact that, by virtue of the spatial localization of the Wannier functions, the error remains extremely small even for points away from that grid [96].

The next step is to evaluate $\mathcal{E}^{(\mathrm{H})}_{n\mathbf{k}}$ on a two-dimensional mesh of k-points covering a single slice and then use a contour-finding algorithm to map out and discretize the Fermi loops therein. This part of our scheme will be described in more detail in Sec. 5.2.2. For now we can just assume that the output is a sequence of points $\mathbf{k}_j$ $(j = 0, \ldots, J-1)$ providing a fairly dense mapping of the contour. (As before, we assume for simplicity that the Fermi contour consists of a single loop; the extension to multiple loops is straightforward.)

Next we need to obtain the Berry connection $\mathbf{A}_n(\mathbf{k}) = i\langle u^{(\mathrm{H})}_{n\mathbf{k}} | \nabla_{\mathbf{k}} | u^{(\mathrm{H})}_{n\mathbf{k}} \rangle$ as in Eq. (5.5). Using

Eq. (5.20), this becomes

$$\begin{aligned}\mathbf{A}_n(\mathbf{k}) &= \sum_{lm} U^{\dagger}_{nl}(\mathbf{k})\, \mathbf{A}^{(\mathrm{W})}_{lm}(\mathbf{k}) U_{mn}(\mathbf{k}) \\ &\quad + i \sum_{m} U^{\dagger}_{nm}(\mathbf{k}) \nabla_{\mathbf{k}} U_{mn}(\mathbf{k}) \end{aligned} \tag{5.21}$$

where

$$\mathbf{A}^{(\mathrm{W})}_{nm}(\mathbf{k}) = i \langle u^{(\mathrm{W})}_{n\mathbf{k}} | \nabla_{\mathbf{k}} | u^{(\mathrm{W})}_{m\mathbf{k}} \rangle \tag{5.22}$$

is computed in practice from the expression

$$\mathbf{A}^{(\mathrm{W})}_{nm}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, \langle \mathbf{0} n | \hat{\mathbf{r}} | \mathbf{R} m \rangle \tag{5.23}$$

in a manner similar to Eq. (5.18). Details concerning the method of calculating Eqs. (5.18) and (5.23) can be found in Chapter 4.

The decomposition of $\mathbf{A}_n(\mathbf{k})$ into two terms in Eq. (5.21) is an artifact of the choice of Wannier functions; only the sum of the two terms is physically meaningful (upon a circuit integration). However, for a given choice of Wannier functions, the first term arises because the Bloch functions $|u^{(\mathrm{H})}_{n\mathbf{k}}\rangle$ acquire some of the Berry curvature attached to the full subspace of M Wannier functions used to represent them, whereas the second term represents the Berry curvature arising from changes of character of this Bloch state *within* the Wannier subspace. To clarify this viewpoint, we introduce a notation in Chapter 4 in which $\|v_{n\mathbf{k}}\rangle\!\rangle$ is defined to be the nth column vector of matrix U, so that the second term of Eq. (5.21) becomes $i\langle\!\langle v_{n\mathbf{k}} \| \nabla_{\mathbf{k}} \| v_{n\mathbf{k}} \rangle\!\rangle$. Plugging into Eq. (5.12), this yields

$$\begin{aligned}\phi_n(i) &= \oint \langle\!\langle v_{n\mathbf{k}} \| \mathbf{A}^{(\mathrm{W})}(\mathbf{k}) \| v_{n\mathbf{k}} \rangle\!\rangle \cdot d\mathbf{l} \\ &\quad + i \oint \langle\!\langle v_{n\mathbf{k}} \| \nabla_{\mathbf{k}} \| v_{n\mathbf{k}} \rangle\!\rangle \cdot d\mathbf{l} \end{aligned} \tag{5.24}$$

for the Berry phase of slice i appearing in Eq. (5.15). Note that the integrand in the first term is gauge-invariant (here "gauge" refers to the application of a phase twist $\|v_{n\mathbf{k}}\rangle\!\rangle \rightarrow e^{i\beta(\mathbf{k})}\, \|v_{n\mathbf{k}}\rangle\!\rangle$), while in the second term only the entire loop integral is gauge-invariant. Indeed, the second term is

just a Berry phase defined within the M-dimensional "tight-binding space" provided by the Wannier functions. Recalling that $\mathbf{k}_j$ for $j = 0, \ldots, J-1$ is our discretized description of the Fermi loop, and using standard methods for discretizing Berry phases [20] as in Eq. (5.14), our final result becomes

$$\begin{aligned} \phi_n(i) &= \sum_{j=0}^{J-1} \langle\!\langle v_{n\mathbf{k}} \| \mathbf{A}^{(\mathrm{W})}(\mathbf{k}) \| v_{n\mathbf{k}} \rangle\!\rangle \cdot \Delta\mathbf{k} \\ &\quad -\mathrm{Im}\ln \prod_{j=0}^{J-1} \langle\!\langle v_{n\mathbf{k}_j} \| v_{n\mathbf{k}_{j+1}} \rangle\!\rangle, \end{aligned} \tag{5.25}$$

where $\Delta\mathbf{k} = (\mathbf{k}_{j+1} - \mathbf{k}_{j-1})/2$.

As we shall see below, in practice we only encounter closed orbits, in which case it is clearly appropriate to set $\mathbf{k}_J = \mathbf{k}_0$ and close the phases with $\|v_{n,\mathbf{k}_J}\rangle\!\rangle = \|v_{n,\mathbf{k}_0}\rangle\!\rangle$. For lower-symmetry situations, however, open orbits with $\mathbf{k}_J = \mathbf{k}_0 + \mathbf{G}$ may be encountered. Even in this case, however, we would still set $\|v_{n,\mathbf{k}_J}\rangle\!\rangle = \|v_{n,\mathbf{k}_0}\rangle\!\rangle$; in contrast to the full Bloch states which obey [20] $u_{n,\mathbf{k}_J} = e^{-i\mathbf{G}\cdot\mathbf{r}}\, u_{n,\mathbf{k}_0}$, no extra phase factors are needed here because the Fourier-transform convention of Eq. (5.16) treats the Wannier functions as though they are all nominally located at the cell origin.

In summary, our strategy is to evaluate Eq. (5.15) by decomposing each generalized path $\mathcal{C}_n(i)$ into connected simple loops, and sum the loop integrals as computed using Eq. (5.25). The operations needed to evaluate Eq. (5.25) are inexpensive as they all involve vectors and matrices defined in the low-dimensional space of the Wannier representation.

5.2 Computational details

5.2.1 First-principles calculations

Fully relativistic band-structure calculations are carried out for the ferromagnetic transition metals Fe, Co and Ni at their experimental lattice constants (5.42, 4.73, and 6.65 bohr, respectively) using the `PWSCF` code [76]. Norm-conserving pseudopotentials with spin-orbit coupling [97] are generated using similar parameters as in Chapter 4. An energy cutoff of 60 Hartree is used for the

Table 5.1: Calculated spin magnetic moment per atom (in μ_{B}) for the three transition metals Fe, Ni and Co, with magnetization along [001], [111] and [001], respectively.

	bcc Fe	fcc Ni	hcp Co
Theory	2.22	0.62	1.60
Experiment[1]	2.13	0.56	1.59

planewave expansion of the valence wavefunctions (400 Hartree for the charge densities), and the PBE generalized-gradient approximation [77] is used for the exchange-correlation functional. The self-consistent ground state is obtained using a $16 \times 16 \times 16$ Monkhorst-Pack [81] mesh of k-points and a fictitious Fermi smearing [82] of 0.02 Ry for the Brillouin-zone integration.

The calculated spin magnetic moments are shown in Table 5.1. The effect of spin-orbit coupling on these moments is included in the calculation, since it is needed in any case to obtain a nonzero AHC. The agreement with experiment is rather good, confirming that our norm-conserving pseudopotentials are suitable for describing the ferromagnetic state of the transition metals.

The maximally-localized Wannier functions are generated using the `WANNIER90` code [84]; details are given in Secs. 5.3.1-5.3.3 below.

5.2.2 Mapping and sampling of Fermi loops

As discussed above, our basic strategy involves dividing the BZ into a series of parallel slices and finding the intersections of the Fermi surface with each of these slices. Each slice is sampled on a uniform $N \times N$ k-point mesh, with N ranging from 300 to 500, and the band energies are computed on the mesh using Wannier interpolation. A standard contour-finding algorithm of the kind used to make contour plots is then used to generate a list of Fermi loops and, for each loop, a list $\mathbf{k}_0, \ldots, \mathbf{k}_{J-1}$ of k-points providing a discretized representation of the loop.

As shown in Fig. 5.2(a), the Fermi contours in the first BZ are sometimes composed of multiple

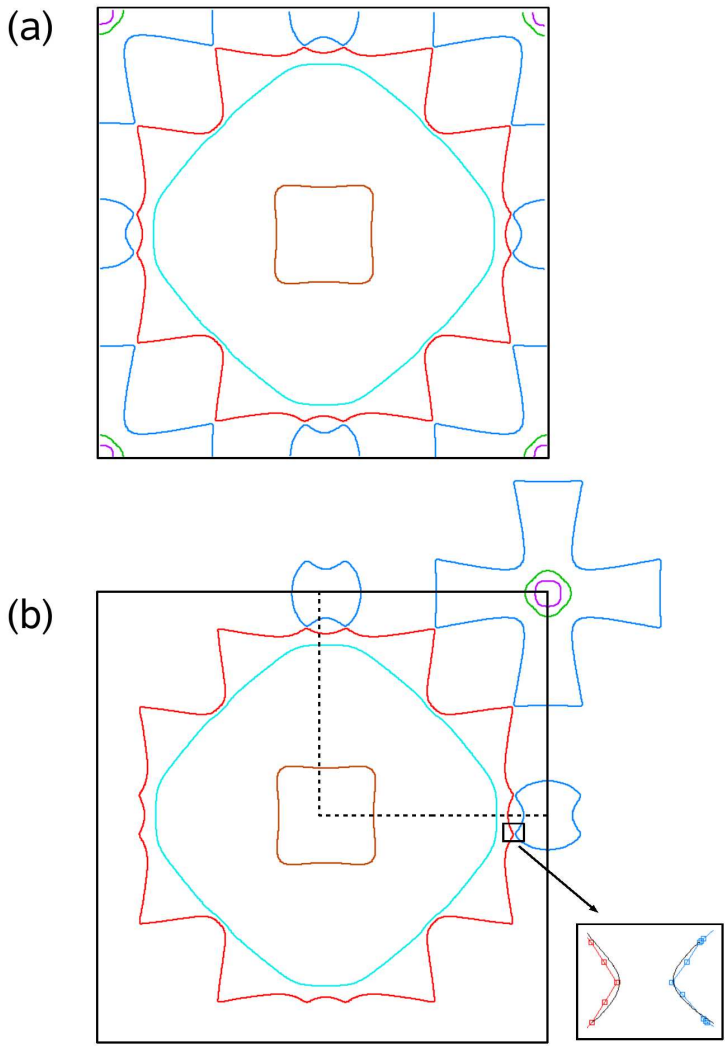

Figure 5.2: Calculated Fermi-surface intersections (Fermi loops) on the $k_z = 0.02$ plane for bcc Fe; different bands are color-coded for clarity. (a) Fermi contours within the first Brillouin zone. (b) Fermi contours after reassembly to form closed contours by translating some portions by a reciprocal lattice vector. Inset: enlargement showing part of an avoided crossing where a refined mesh (black lines) is used to obtain a more accurate representation of the Fermi loop. The actual calculation is performed within the dashed box.

segments terminating at the BZ boundary. To insure that we get closed loops suitable for the evaluation of Eq. (5.25), we actually do the initial contour-finding procedure in an extended zone with 3×3 times the size of the first BZ. We then select closed loops located near the central cell while identifying and discarding loops or portions of loops that correspond to periodic images of these chosen loops. The result is a set of closed loops that partially extend outside the first BZ as shown in Fig. 5.2(b). Of course, if there were open orbits on the Fermi surface, it would not always be possible to select closed loops in the above sense; one would have to accept a "loop" with $\mathbf{k}_J = \mathbf{k}_0 + \mathbf{G}$ as discussed following Eq. (5.25). However, we never encounter such open orbits in practice for the types of materials studied here, in which the magnetization is aligned with a three-fold, four-fold, or six-fold rotational symmetry axis. The slices are perpendicular to the symmetry axis, and the symmetry ensures that open orbits cannot occur on the slices.

A potential difficulty in applying the Fermi-loop method to real materials arises from the possible presence of degeneracies or near-degeneracies between bands. If two bands are degenerate at the Fermi energy, this means that two Fermi loops touch, and it is no longer straightforward to define and compute the Berry phases of these loops. Fortunately, the presence of ferromagnetic spin splitting and spin-orbit coupling removes almost all degeneracies. In our calculations we found no true degeneracies in hcp Co or fcc Ni, and the only degeneracies in bcc Fe were found to lie in the $k_z = 0$ plane. (In the latter case, we avoid the $k_z = 0$ plane by picking a $k_\perp$ mesh that is offset so that this plane is skipped over.) On the other hand, we do find numerous weakly avoided crossings induced by the spin-orbit interaction, and while these introduce no difficulty in principle, they do require special care in practice. Indeed, we find that it is important to sample the Fermi surface very accurately in the vicinity of these crossings. To do so, we calculate the Berry curvature at each $\mathbf{k}_j$ using Wannier interpolation, and if a large value is encountered, we introduce a refined mesh with 4×4 greater density in this region, repeat the contour-finding procedure there, and replace

the discretized representation of this portion of the loop with a denser one. We also take care to recompute $\mathcal{E}_{n\mathbf{k}}$ at each $\mathbf{k}_j$ and iteratively adjust the k-point location in the direction transverse to the loop in order to insure that $\mathcal{E}_{n\mathbf{k}}$ lies precisely at the Fermi energy. An example of a portion of a Fermi loop that has been refined in this way is illustrated in the inset to Fig. 5.2(b). Overall, the resulting number J of k-points per loop ranges from several hundreds to thousands, depending on the size and complexity of the Fermi loop.

In our current implementation, the entire procedure above is repeated independently on each of the slices. As already mentioned in Sec. 5.1.2, it is important to make a consistent choice of branch of the Berry phase $\phi(i)$ on consecutive slices. We do this by adding or subtracting a multiple of 2π to the Berry phase calculated from Eq. (5.15) such that $|\phi_n(i) - \phi_n(i-1)| \ll 2\pi$ is satisfied, always checking for consistency between the first and last slice.

5.2.3 Use of symmetry to reduce computational load

The presence of a net magnetization results in a considerable reduction in symmetry, but several symmetries still remain that can be exploited to reduce the computational cost. In the previous Fermi-sea-based methods [21, 95] the use of symmetries is straightforwardly implemented by restricting the k-point sampling to the irreducible wedge of the BZ. For the Fermi-loop method, the use of symmetries needs more careful treatment.

Here we discuss the difficulties, and point out their solution, using ferromagnetic bcc Fe as an example. We focus our attention on the mirror symmetries M_x and M_y. Since each slice lies in an x-y plane, we can use these to restrict the bandstructure calculation and the search for Fermi contours to a reduced BZ having one-fourth of the area of the full BZ, as shown by the dashed line in Fig. 5.2(b). However, a typical Fermi loop will no longer close within this reduced BZ. Because a Berry phase is a global property of a closed loop, one cannot just compute the Berry phase of

open segment lying inside the reduced BZ and multiply by four; the Berry phase of this segment is ill-defined unless the phases of the wavefunctions at its terminal points are specified.

Our solution to this difficulty is illustrated in Fig. 5.3. We make some arbitrary but definite choice of the phases of the Bloch functions in the upper-right segment, compute the open-path Berry phase following Eq. (5.25), and multiply by four. We then add corrections that take account of the phase jumps at the segment boundaries. For example, we let M_x acting on the Bloch states from 1 to 2 define the Bloch states from $1'$ to $2'$. The correction arising from the $1'$-1 boundary is then given by the phase of $\langle u_{1'}|u_1\rangle = \langle M_x u_1|u_1\rangle$. (Here M_x is defined in the spinor context and includes a complex conjugation component. Since the Bloch functions are expressed in the Wannier basis in our approach, information about the symmetries of the Wannier functions has to be extracted and made available for the application of the symmetry transformations.) Similar corrections, using also M_y, are obtained for the $2'$-$2''$, $1''$-$1'''$, and $2'''$-2 segment boundaries. By including these mismatch corrections, we are able to calculate the global Fermi-loop Berry phase in a correct and globally gauge-invariant manner.

We have tested this procedure and confirmed that the results obtained are essentially identical to those computed without the use of symmetry. The BZ could in principle be reduced further in bcc Fe using the diagonal mirror operations, but we have not tried to implement this.

5.3 Results

In this section we present the results of our calculations of the anomalous Hall conductivity using the Fermi-loop approach of Eq. (5.25) as applied to the three ferromagnetic transition metals Fe, Co and Ni.

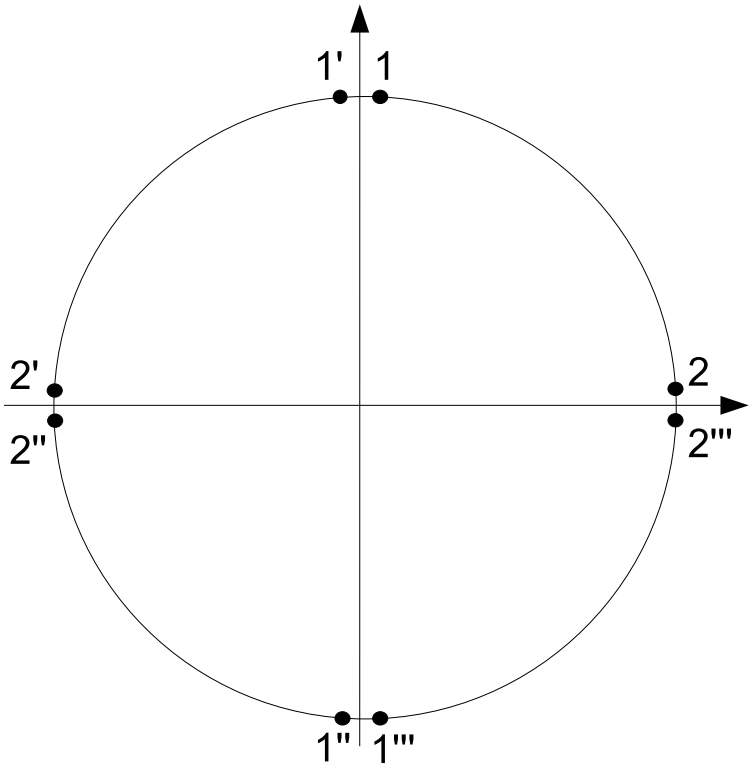

Figure 5.3: Illustration of use of M_x and M_y mirror symmetries on a slice of the Brillouin zone in bcc Fe. Only the segment of the Fermi loop from Point 2 to Point 1 is actually computed; the three other segments are included using symmetry operations.

5.3.1 bcc Fe

We have previously presented calculations of the AHC of bcc Fe computed using the Fermi-sea formulation in Chapter 4. Here we adopt the same choice of Wannier functions as in that work, namely 18 Wannier functions covering the s, p and d character and both spins. The orbitals of s, p, and e_g character are actually rehybridized into Wannier functions of sp^3d^2 type, and the Wannier functions are only approximate spin eigenstates because of the presence of spin-orbit interaction (see Chapter 4 for details).

Six bands cross the Fermi energy in bcc Fe. In Fig. 5.4 we show the Fermi-surface sheets for four of these, plotted using the `Xcrysden` package [99]. Some of these are quite complicated but, as expected, they all conform to the lattice symmetries. What is not clearly visible in these plots are the tiny spin-orbit-induced splittings, which change the connectivity of the Fermi surface. As mentioned earlier, such features play an important role in the AHC, and need to be treated with care.

We take the magnetization to lie along the $[001]$ axis. Choosing $\mathbf{b}_1 = (2\pi/a)(1\bar{1}0)$ and $\mathbf{b}_2 = (2\pi/a)(110)$ in the notation of Sec. 5.1.1, it follows that $\mathbf{a}_3 = 2\pi\,\mathbf{b}_1 \times \mathbf{b}_2 / V_{\rm recip} = (0, 0, a)$ where $V_{\rm recip}$ is the primitive reciprocal cell volume, and we only need to compute the c_{n3} in Eq. (5.10). The slices are square in shape, and $k_\perp = k_z$ is discretized into 500 slices.

In Fig. 5.5 we have plotted the total Berry phase Eq. (5.13) on each slice as computed from Eq. (5.25). The results are symmetric under mirror symmetry, so only half of the range of $k_\perp$ is shown. The sharp peaks and valleys in Fig. 5.5 are related to degenerate or near-degenerate bands that have been split by the spin-orbit interaction, as was illustrated, e.g., in the inset of Fig. 5.2. To validate the calculation, we compare it against a direct numerical integration of the Berry curvature over the occupied bands using Eq. (5.11), as indicated by the symbols in Fig. 5.5. In spite of rather complex and irregular Fermi surfaces, the agreement between the two methods in Fig. 5.5 is excellent.

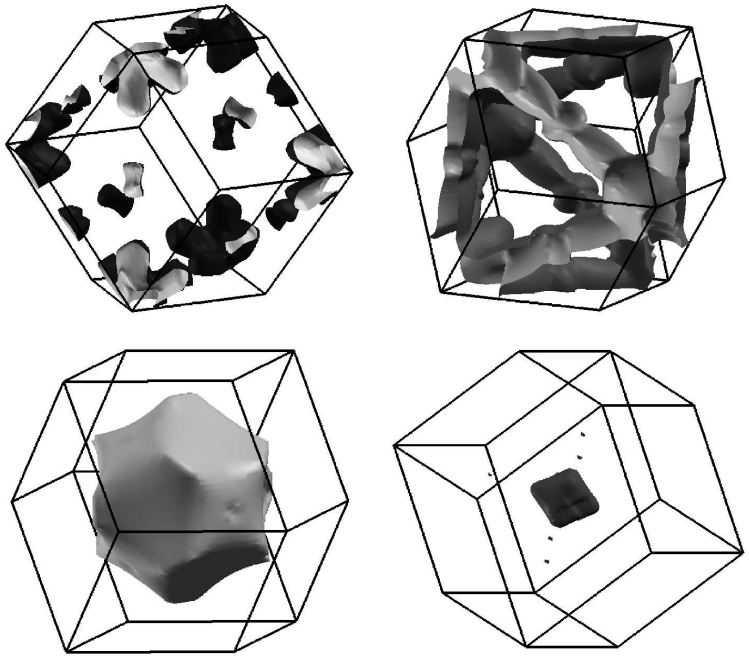

Figure 5.4: Calculated Fermi surfaces of bcc Fe. The outside frame is the boundary of the Brillouin zone.

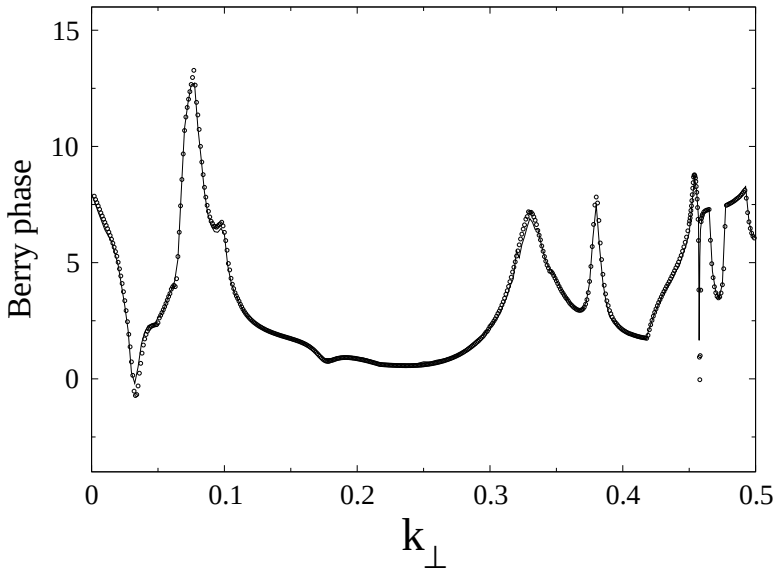

Figure 5.5: Calculated Berry phase $\phi(k_\perp)$ of bcc Fe (in radians) as a function of $k_\perp$ (in units of $2\pi/a$). Solid line shows results obtained from the Fermi-loop method of Eq. (5.25); circles indicate reference results obtained by the integration of the Berry curvature on each slice using Eq. (5.11).

Table 5.2: Anomalous Hall conductivity, in Ω-cm^{-1}. First three rows show values computed using Eqs. (5.6)–(5.10) together with Eq. (5.25), the first term only of Eq. (5.25), or Eq. (5.11), respectively. Results of previous theory and experiment are included for comparison.

	bcc Fe	fcc Ni	hcp Co
Fermi loop	750	-2275	478
Fermi loop (1st term)	7	0	-4
Berry curvature	753	-2203	477
Previous theory	751[1]	-2073[2]	492[2]
Experiment	1032[3]	-646[4]	480[5]

The values of the integrated anomalous Hall conductivity using the new approach and the reference approach are shown in the first and third lines of Table 5.2. The second line shows the contribution obtained from integrating only the first term of Eq. (5.25); clearly, this contribution is very small. The agreement with the previous theory of Yao *et al.*[21] is excellent, while the agreement with experiment is only fair. Table 5.2 will be discussed further in Sec. 5.3.4.

5.3.2 fcc Ni

For fcc Ni we chose 14 Wannier functions, seven each of approximately spin-up and spin-down character. These were comprised of five Wannier functions of d-like symmetry centered on the Ni atoms and two Wannier functions of tetrahedral symmetry located on the tetrahedral interstitial sites, similar to the choice that was made for Cu in Ref. [72]. The inner energy window was chosen to extend 21 eV above the bottom of the bands, thus extending 7.1 eV above the Fermi energy and including several unoccupied bands as well.

In Fig. 5.6 we show the Fermi sheets for four of the five bands that cross the Fermi energy in fcc Ni. The shapes of these Fermi surfaces are somewhat more spherical than those of bcc Fe. As expected, they again conform to the lattice symmetries.

In the case of fcc Ni, the magnetization lies along the $[111]$ axis. Choosing $\mathbf{b}_1 = (2\pi/a)(0\bar{2}\bar{2})$

and $\mathbf{b}_2 = (2\pi/a)(20\bar{2})$ in the notation of Sec. 5.1.1, it follows that $\mathbf{a}_3 = 2\pi\,\mathbf{b}_1 \times \mathbf{b}_2/V_{\rm recip} = (a,a,a) = a\sqrt{3}\,\hat{\mathbf{e}}_{(111)}$, and we only need to compute the c_{n3} in Eq. (5.10). The slices are hexagonal in shape, and $k_\perp = \mathbf{k}\cdot\hat{\mathbf{e}}_{(111)}$ is discretized into about 100 slices.

The results are plotted in Fig. 5.7, along with symbols denoting the reference calculation by an integration of the Berry curvature over the slice. Once again, the agreement is very satisfactory. The values of the integrated AHC are again summarized in Table 5.2.

5.3.3 hcp Co

Co in the hcp structure has two atoms per unit cell. We choose 18 Wannier functions per Co atom, nine for each spin, in a very similar manner as was done for Fe in Sec. 5.3.1. We therefore have 36 Wannier functions per cell.

Seven bands cross the Fermi energy in hcp Co. We show the four largest Fermi-surface sheets in Fig. 5.8. The Fermi surfaces can be seen to respect the 6-fold crystal symmetry, and none of them touch each other.

The magnetization of hcp Co lies along the $[001]$ axis. We thus choose $\mathbf{b}_1 = (2\pi/a)(1/\sqrt{3},-1,0)$ and $\mathbf{b}_2 = (2\pi/a)(1/\sqrt{3},1,0)$ in the notation of Sec. 5.1.1, and it follows that $\mathbf{a}_3 = 2\pi\,\mathbf{b}_1 \times \mathbf{b}_2/V_{\rm recip} = (0,0,c)$. The slices are hexagonal in shape, and $k_\perp = k_z$ is discretized into about 200 slices.

The results are plotted in Fig. 5.9, along with the symbols denoting the reference calculation by integration of the Berry curvature. Once again, the peaks and valleys correspond to the places where two loops approach one another closely. Some pieces of the Fermi surfaces of hcp Co are nearly parallel to the slices (see the bottom right panel of Fig. 5.8), so that the number and shapes of the Fermi loops sometimes change rapidly from one slice to another. In particular, we found it difficult to enforce continuity of the branch choice of Eq. (5.25) as a function of $k_\perp$ near the sharp

features at $k_\perp a/2\pi = 0.18$ and 0.42 in Fig. 5.9. We therefore redetermined the correct branch choice by comparing with the result of the Berry-curvature integration at slices just outside these difficult regions. Despite these difficulties, it can still be seen that the Fermi-loop method works well for this case.

5.3.4 Discussion

The second row of Table 5.2 shows the results computed using only the first term of Eq. (5.25). In each case, its contribution is less than 1% of the total, and would therefore be negligible for most purposes. Actually, it can be shown that the inclusion of the first term only in Eq. (5.25) of the present method is equivalent to carrying out the Berry-curvature integration approach of Chapter 4 with the D–D term omitted in Eq. (32) of that work (that is, only the D–A and $\overline{\Omega}$ terms included). We have carried out this comparison and find values of 7, -0.5 and $-2\,\Omega\text{-cm}^{-1}$ for bcc Fe, fcc Ni, and hcp Co, respectively, in very good agreement with the values reported in Table 5.2. The physical interpretation for the small terms in the second row of Table 5.2 is basically that the full set of Bloch-like states constructed from the Wannier functions (e.g., the manifold of 18 Bloch-like states in bcc Fe) has some small Berry curvature of its own, and the projection of this curvature onto the occupied subspace gives the small first term of Eq. (5.25). On the other hand, spin-orbit induced splittings across the Fermi level *between* Bloch-like states built from these Wannier functions give large, sharply peaked contributions to the Berry curvature of the occupied subspace, and make a very much larger contribution to the total AHC. Of course, the precise decomposition between the first and second term of Eq. (5.25) depends on the exact choice of Wannier functions, but the present results seem to indicate that the dominance of the second term is probably a general feature, at least for systems in which the Wannier functions are well localized and the spin-orbit splitting is not very strong.

As mentioned in the previous section, the overall agreement seen in Table 5.2 between the results computed using the Fermi-loop approach and those computed using the Berry-curvature integration indicate the internal consistency of our theory and implementation. The agreement with the results of Yao and coworkers, which were obtained by a Berry-curvature integration using an all-electron approach [21], also demonstrates the robustness of our pseudopotential implementation, including its ability to represent spin-orbit interactions correctly.

In the last row of Table 5.2 we show comparison with some experimental values for the AHC of Fe, Ni, and Co. The agreement is not very good, but it was not expected to be. First, the theoretical values in Table 5.2 are all computed by including only the intrinsic Karplus-Luttinger contribution to the AHC; extrinsic skew scattering and side-jump scattering contributions are not included, and there is no reason *a priori* to expect these to be negligible. Second, the experimental values themselves have considerable uncertainty. For example, Ref. [101] gives values for Fe and Ni films of about 970 and $-480\,\Omega\text{-cm}^{-1}$ respectively (the value quoted in Table 5.2 for Co is also a film value) whereas it gives a value of about $2000\,\Omega\text{-cm}^{-1}$ for single-crystal Fe. Ref. [17] gives a value for Ni of $-753\,\Omega\text{-cm}^{-1}$ and Ref. [102] reports a value for Co of $500\,\Omega\text{-cm}^{-1}$. It could well be that different kinds of experimental samples have different impurity and defect populations, leading to different extrinsic contributions to the AHC. Finally, on the theoretical side, not much is yet known about the accuracy of common exchange-correlation functionals, such as the PBE functional used here [77]. for the AHC. Clearly, there is much work to be done on both the experimental and theoretical side before close agreement can be expected. Nevertheless, the very rough agreement, at the level of signs and general trends, between the theoretical and experimental values in Table 5.2 suggests that the intrinsic Karplus-Luttinger contribution is probably an important, and sometimes a dominant, one.

5.3.5 AHC anisotropy of Ni

Rather little is known, either experimentally or theoretically, about the dependence of the AHC on the crystallographic orientation of the magnetization. In crystal with a small magneto-crystalline anisotropy, it may be possible to rotate the direction of the magnetization with a small applied magnetic field, and to test, for example, whether the AHC pseudovector remains aligned with the magnetization (isotropic behavior). In at least one case, a highly anisotropic AHE has been observed experimentally in Ag-doped $FeCr_2S_4$ (ferromagnetic diamond lattice) [103]. In the following, we explicitly calculate the dependence of AHE on the angle between z axis and the direction of magnetization for the case of fcc Ni.

The easy axis of fcc Ni is determined experimentally to be along the [111] direction [98]. Along this direction, the calculated spin magnetization is 0.62 Bohr magneton, which is in good agreement with the experimental value 0.56. In Table 5.3, we present the results of the calculations of the anisotropy energy and anisotropic anomalous Hall conductivity along several directions between the [001] direction and [111] directions. For each of these directions, a noncollinear *ab-initio* calculation is performed with the orientation of the magnetization fixed to that direction and its magnitude is allowed to relax. The differences in total energy between different magnetization directions are not significant. It is worth noting that our noncollinear magnetization calculation gives the wrong prediction, namely that the [001] direction is the lowest-energy configuration. The spin magnetization does not change dramatically and it ranges from 0.61 to 0.63 Bohr magneton. It is clear that the largest change occurs for the anomalous Hall conductivity, which differs by about 12.8% between the [111] direction and the [001] direction. This is illustrated in Fig. 5.11. As the angle of the magnetization increases from the z-axis, the change in anomalous Hall conductivity increases, reaching its maximum at about 45°. In addition to the change in the magnitude of the

Table 5.3: Calculated magnitude of the magnetic moment and the anisotropic anomalous Hall conductivity, in units of $(\Omega\ \mathrm{cm})^{-1}$, as the magnetization is rotated in the $(01\bar{1})$ plane. θ_z is the angle between the magnetization and the z axis.

θ_z	AHC	M_{spin}	$E - E_{111}(\mu eV)$
0.0 (001)	-2366	0.63	-0.27
9.00	-2323	0.63	0.0
18.00	-2219	0.63	0.0
27.00	-2137	0.63	0.27
36.00	-2110	0.63	0.27
45.00	-2091	0.62	0.27
54.5 (111)	-2097	0.61	0.0
60.0	-2098	0.62	-0.27
70.0	-1986	0.63	0.27
80.0	-1931	0.63	0.0
90.0	-1910	0.62	0.0

magnetization, the pseudo-vector $\mathbf{K}$ defined in Eq. (5.7) is not parallel to the direction of magnetization, as shown in Fig. 5.12. It is evident that fcc Nickel has a clearly anisotropic anomalous Hall effect.

The anisotropic anomalous Hall effect can be understood in terms of the relation of the band structure and the Berry curvature. The calculated band structure around the Fermi surface and the Berry curvature along high-symmetry k points are shown in Fig. 5.10. The largest value of Berry curvature occurs at the place where the Fermi energy lies between a pair of bands which have a small gap (for example, at 75% of the distance from G to K). The value of the Berry curvature is very sensitive to the gap. The small gap is induced by spin-orbit coupling. Different directions of magnetization induce different gaps, which dramatically change the Berry curvature.

5.4 Computational efficiency

The motivation for developing a method for computing the AHC that relies only on information computed on the Fermi surface is, to some degree, aesthetic and philosophical: Haldane argued that

the AHC is physically most naturally regarded as a Fermi-surface property [69], and as such should be computed using a method that does not make use of extraneous information in arriving at the desired quantity. However, a much more important motivation from the practical point of view is the idea that the computational effort might be drastically reduced by having to compute quantities only on the two-dimensional Fermi surface rather than on a three-dimensional mesh of k-points.

In the present implementation as it stands, unfortunately, the computational savings gained through the use of the Fermi-loop Berry-phase approach is quite modest. After taking advantage of the symmetry as discussed in Sec. 5.2.3, the total computational time of our AHC calculation for bcc Fe is about 1.7 hours using a 200×200 k-mesh on each of 500 slices, to be compared with about 2 hours using our previous method of Chapter 4. (These timings are on a 2.2 GHz AMD-Opteron PC, and neither includes the Wannier construction step, which takes about 2.5 hours.) Roughly, the work on each slice can be divided into three phases: Step 1, computing the energy eigenvalues on the 200×200 k-mesh; Step 2, executing the contour-finding algorithm; and Step 3, evaluating Eq. (5.25) on the discretized Fermi loops. We find that less than 1% of the computer time goes to Step 2, while the remainder is roughly equally split between Step 1 and 3. The operations in these steps have been greatly accelerated by making use of Wannier interpolation methods, but this is also the case for the comparison method of Chapter 4. (We emphasize that, for this reason, both the method of Chapter 4 and the present one are orders of magnitude faster than methods based on direct first-principles calculations at every k-point.)

Many opportunities for further reduction of the computer time are worthy of further exploration. Regarding Step 1, for example, at the moment the contour-finding is done independently on each slice; it might be much more efficient to step from slice to slice and use a local algorithm to determine the deformation of the Fermi contours on each step. It may also be possible to do a first cut at the contour-finding using a coarser k-mesh (say 50×50) and then refine it in regions where the

loops approach one another or have sharp bends. It may also be possible to take larger steps between slices in most regions of $k_\perp$, and fall back to fine slices only in delicate regions. In implementing all such strategies, however, one should be careful to avoid missing any small loops that might appear suddenly from one slice to the next, or which might be missed on an initial coarse sampling of the slice. It may also be interesting to explore truly three-dimensional algorithms for finding contour surfaces, and then derive two-dimensional loops from these.

As for Step 3, it should be possible to use a lower density of k-points in the portions of the loop discretization where the character of the wavefunctions is changing slowly. The time for this step will also obviously benefit from taking larger steps between slices in regions where this is possible. Finally, a reduction by a factor of two or more may be possible by making use of symmetries not considered in Sec. 5.2.3, such as the diagonal mirror symmetries ($x \leftrightarrow y$ etc.) in bcc Fe.

The exploration of these issues is somewhat independent from the quantum-mechanical formulation of the underlying theory, which is the main focus of the present work, and we have therefore left the exploration of these possibilities for future investigations.

Finally, it should be emphasized that the computational load scales strongly with the dimension of the Wannier space used to represent the wavefunctions. In our calculations, this was 18, 14, and 36 for Fe, Ni, and Co, respectively. In some materials, there may be only a few bands crossing the Fermi energy, and it might be possible to represent them using a much smaller number of Wannier functions. This is the case in many transition-metal oxides such as Sr_2RuO_4, cuprate superconductors, etc. In ferromagnetic materials of this kind, it should be possible to choose an inner window in the Wannier disentanglement procedure [72] that brackets the Fermi energy but does not extend to the bottom of the occupied valence band, and to generate just a handful of Wannier functions (e.g., three t_{2g} orbitals times two for spin) to be used in the Wannier interpolation procedure. Then all matrices used in that procedure would be very much smaller (e.g., 6×6) and

the computation would go considerably faster.

5.5 Summary

In summary, we have developed a first-principles method for computing the intrinsic AHC of ferromagnets as a Fermi-surface property. Unlike conventional methods that are based on a k-space volume integration of the Berry curvature over the occupied Fermi sea, our method implements the Fermi-surface philosophy by dividing the Brillouin zone into slices normal to the magnetization direction and computing the Berry phases of the Fermi loops on these slices. While Haldane has pointed out that only the non-quantized part of the AHC can be determined in principle from a knowledge of Fermi-surface properties only, we find in practice that it is straightforward to make the correct branch choice and resolve the quantum of uncertainty by doing a two-dimensional Berry-curvature integration on just one or a few of the slices. Our method also makes use of methods of Wannier interpolation to minimize the number of calculations that have to be done using a full first-principles implementation; almost all the operations needed to compute the AHC are actually done by working with small matrices (e.g., 18×18 for bcc Fe) in the Wannier representation.

We have tested and validated our new method by comparing with our earlier implementation of a Fermi-sea Berry-curvature integration for bcc Fe, fcc Ni, and hcp Co. The different crystal structures and magnetization orientations in these three materials also allow us to demonstrate the flexibility of the method in dealing with these different cases. We find excellent agreement between the two approaches in all cases.

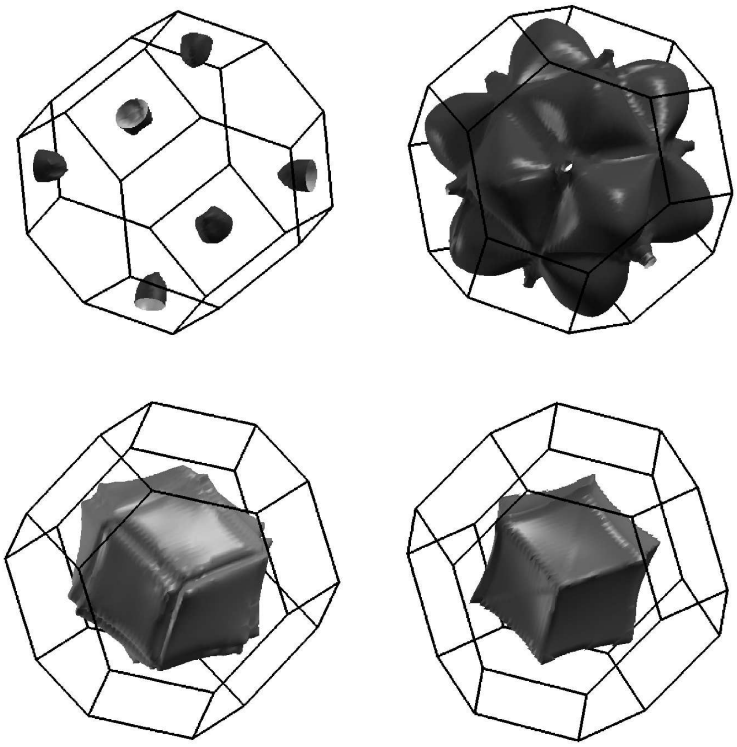

Figure 5.6: Calculated Fermi surfaces of fcc Ni. The outside frame is the boundary of the Brillouin zone.

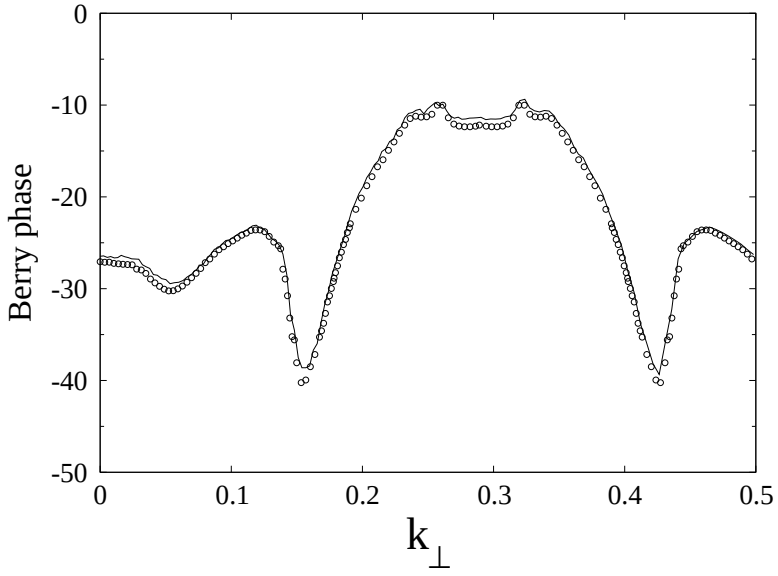

Figure 5.7: Calculated Berry phase $\phi(k_\perp)$ of fcc Ni (in radians) as a function of $k_\perp$ (in units of $2\pi/\sqrt{3}a$). Solid line shows results obtained from the Fermi-loop method of Eq. (5.25); circles indicate reference results obtained by the integration of the Berry curvature on each slice using Eq. (5.11).

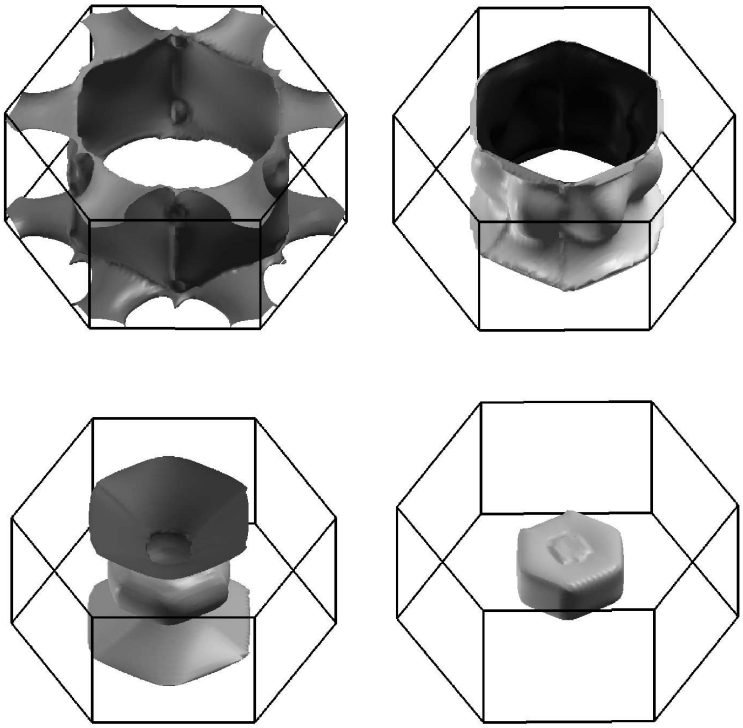

Figure 5.8: Calculated Fermi surfaces of hcp Co. The outside frame is the boundary of the Brillouin zone.

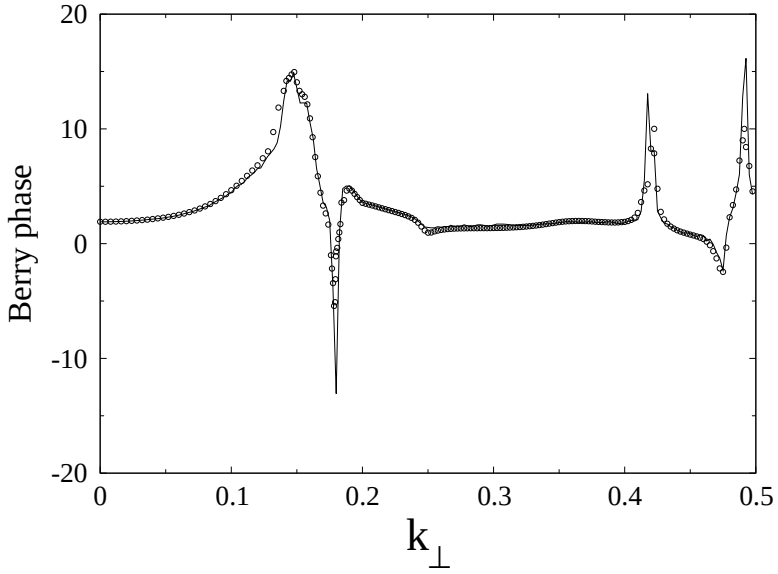

Figure 5.9: Calculated Berry phase $\phi(k_\perp)$ of hcp Co (in radians) as a function of $k_\perp$ (in units of $2\pi/c$). Solid line shows results obtained from the Fermi-loop method of Eq. (5.25); circles indicate reference results obtained by the integration of the Berry curvature on each slice using Eq. (5.11).

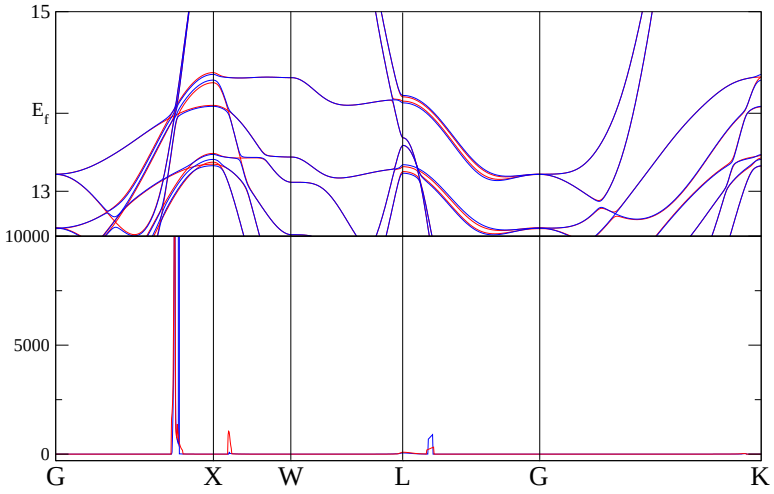

Figure 5.10: Top panel: calculated band structure of Ni (energy in units of eV)with magnetization along [001] (blue) and [111] (red) directions. Bottom panel: calculated Berry curvature in atomic units with the corresponding magnetization.

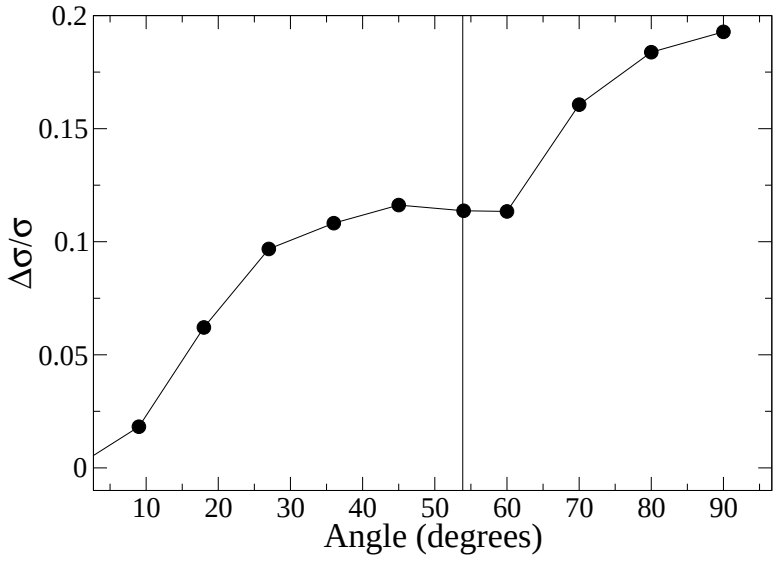

Figure 5.11: Calculated change in the magnitude of the anomalous Hall conductivity as a function of the angle between the z axis and the direction of magnetization, as the magnetization is rotated in the $(01\bar{1})$ plane. The vertical line indicates the [111] direction.

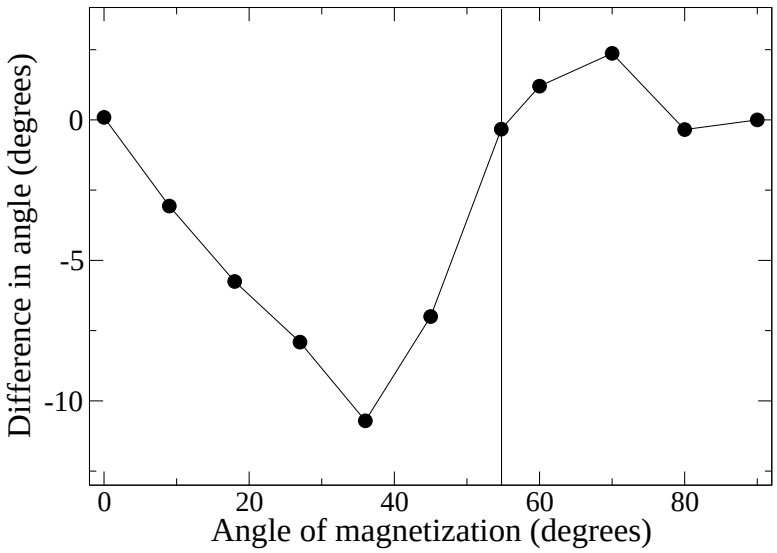

Figure 5.12: Calculated difference in angle between the magnetization and the pseudovector of the anomalous Hall conductivity, as the magnetization is rotated in the $(01\bar{1})$ plane. The vertical line indicates the [111] direction.

References

[1] A. Mooradian and A.L. McWhorter. *Phys. Rev. Lett.*, 19:849, 1967.

[2] Craig J. Fennie and Karin M. Rabe. *Phys. Rev. B*, 72:100103, 2005.

[3] H.L. McMurray, A.W. Solbrig, J.K. Boyter, and C. Noble. *J. Phys. Chem. Solids*, 28:2359, 1967.

[4] B.D. Singh and B. Dayal. *Phys. Status Solidi B*, 38:141, 1970.

[5] W. Weber. *Phys. Rev. Lett.*, 33:371, 1974.

[6] W. Cochran. *Rep. Prog. Phys.*, 26:1, 1963.

[7] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. *Rev. Mod. Phys.*, 73:515, 2001.

[8] X. Gonze. *Phys. Rev. B*, 55:10337, 1997.

[9] X. Gonze and C. Lee. *Phys. Rev. B*, 55:10355, 1997.

[10] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni. *Phys. Rev. B*, 43:7231, 1991.

[11] E. H. Hall. *Philos. Mag.*, 10:301, 1880.

[12] E. H. Hall. *Philos. Mag.*, 12:157, 1881.

[13] Z. Fang, N. Nagaosa, K.S. Takahashi, A. Asamitsu, R. Mathieu, T. Ogasawara, H. Yamada, M. Kawasaki, Y. Tokura, and K. Terakura. *Science*, 302:92, 2003.

[14] Wei-Li Lee, Satoshi Watauchi, R. J. Cava, and N. P. Ong. *Science*, 303:1647, 2004.

[15] T. Jungwirth, Q. Niu, and A.H. MacDonald. *Phys. Rev. Lett.*, 88:207208, 2002.

[16] R. Karplus and J.M. Luttinger. *Phys. Rev.*, 95:1154, 1954.

[17] J. Smit. *Physica*, 21:877, 1955.

[18] L. Berger. *Phys. Rev. B*, 2:4559, 1970.

[19] M.V. Berry. *Proc. R. Soc. Lond. A*, 45:392, 1984.

[20] R.D. King-Smith and David Vanerbilt. *Phys. Rev. B*, 47:1651, 1993.

[21] Y. Yao, L. Kleinman, A. H. MacDonald, J. Sinova, T. Jungwirth, D. S. Wang, E. Wang, and Q. Niu. *Phys. Rev. Lett.*, 92:037204, 2004.

[22] D.J. Thouless, M. Kohmoto, M.P. Nightingale, and M. den Nijs. *Phys. Rev. Lett.*, 49:405, 1982.

[23] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136:B864, 1964.

[24] W. Kohn and L. J. Sham. *Phys. Rev.*, 140:A1133, 1965.

[25] D. C. Langreth and M. J. Mehl. *Phys. Rev. B*, 28:1809, 1983.

[26] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. *Phys. Rev. B*, 46:6671, 1992.

[27] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77:3865, 1996.

[28] D. Hobbs, G. Kresse, and J. Hafner. *Phys. Rev. B*, 62:11556, 2000.

[29] J. C. Phillips. *Phys. Rev.*, 112:685, 1958.

[30] M. T. Yin and M. L. Cohen. *Phys. Rev. Lett.*, 25:7403, 1982.

[31] G. B. Bachelet and M. Schluter. *Phys. Rev. B*, 25:2103, 1982.

[32] D. Vanderbilt. *Phys. Rev. B*, 41:7892, 1990.

[33] I. Souza, J. Í niguez, and D. Vanderbilt. *Phys. Rev. Lett.*, 89:117602, 2002.

[34] P. Umari and A. Pasquarello. *Phys. Rev. Lett.*, 89:157602, 2002.

[35] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni. *Phys. Rev. B*, 43:7231, 1991.

[36] X. Gonze. *Phys. Rev. B*, 52:1096, 1995.

[37] Z-.Q. Li K. Parlinski and Y. Kawazoe. *Phys. Rev. Lett.*, 78:4063, 1997.

[38] K. Kunc and R. M. Martin. *Phys. Rev. Lett.*, 48:406, 1982.

[39] N. Sai, K.M. Rabe, and D. Vanderbilt. *Phys. Rev. B*, 66:104108, 2002.

[40] H. Fu and L. Bellaiche. *Phys. Rev. Lett.*, 91:57601, 2003.

[41] Ivan I. Naumov and Huaxiang Fu. *Phys. Rev. B*, 72:012304, 2005.

[42] A. Antons, J.B. Neaton, K.M. Rabe, and D. Vanderbilt. *Phys. Rev. B*, 71:024102, 2005.

[43] O. Diéguez and D. Vanderbilt. *Phys. Rev. Lett.*, 96:056401, 2006.

[44] R.W. Nunes and X. Gonze. *Phys. Rev. B*, 63:155107, 2001.

[45] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, and J.D. Joannopoulos. *Rev. Mod. Phys.*, 64:1045, 1992.

[46] R. Resta. *Phys. Rev. Lett.*, 80:1800, 1998.

[47] X. Wu, D. Vanderbilt, and D.R. Hamann. *Phys. Rev. B*, 72:035105, 2005.

[48] The ABINIT code is a common project of the Université Catholique de Louvain, Corning Incorporated, and other contributors (www.abinit.org). See: X. Gonze and others, Comp. Mat. Science **25**, 478 (2002).

[49] N. Troullier and J.L. Martins. *Phys. Rev. B*, 43:1993, 1991.

[50] S. Goedecker, M. Teter, and J. Hutter. *Phys. Rev. B*, 54:1703, 1996.

[51] H.J. Monkhorst and J.D. Pack. *Phys. Rev. B*, 13:5188, 1976.

[52] S. Baroni, P. Giannozzi, and A. Testa. *Phys. Rev. Lett.*, 58:1861, 1987.

[53] S. Gironcoli, S. Baroni, and R. Resta. *Phys. Rev. Lett.*, 62:2853, 1989.

[54] T. Sengstag, N. Binggeli, and A. Baldereschi. *Phys. Rev. B*, 52:R8613, 1995.

[55] M. Veithen, X. Gonze, and Ph. Ghosez. *Phys. Rev. B*, 71:125107, 2005.

[56] E. Roman, J.R. Yates, M. Veithen, D. Vanderbilt, and I. Souza. *Phys. Rev. B*, 74:245204, 2006.

[57] G. Deinzer and D. Strauch. *Phys. Rev. B*, 66:100301, 2002.

[58] I. Shoji, T. Kondo, and R. Ito. *Optical and Quantum Electronics*, 34:797, 2002.

[59] W.Y. Ching and Ming-Zhu Huang. *Phys. Rev. B*, 47:9479, 1993.

[60] W.G. Aulbur, L. Jönsson, and J.W. Wilkins. *Phys. Rev. B*, 54:8540, 1996.

[61] C.M. Hurd. *The Hall effect in metals and alloys. (Plenum, New York, 1972).*

[62] A. Gerber *et al. J. Magn. Magn. Mat.*, 242:90, 2002.

[63] J. Smit. *Physica*, 24:39, 1958.

[64] J. Sinova, T. Jungwirth, and J.Černe. *Int. J. Mod. Phys. B*, 18:1083, 2004.

[65] H.R. Leribaux. *Phys. Rev.*, 150:384, 1966.

[66] M.-C. Chang and Q. Niu. *Phys. Rev. B*, 53:7010, 1996.

[67] G. Sundaram and Q. Niu. *Phys. Rev. B*, 59:14915, 1999.

[68] M. Onoda and N. Nagaosa. *J. Phys. Soc. Jpn.*, 71:19, 2002.

[69] F.D.M. Haldane. *Phys. Rev. Lett.*, 93:206602, 2004.

[70] E.N. Adams and E.I. Blount. *J. Phys. Chem. Solids*, 10:286, 1959.

[71] E.I. Blount. *Solid State Phys.*, 13:305, 1962.

[72] I. Souza, N. Marzari, and D. Vanderbilt. *Phys. Rev. B*, 65:035109, 2001.

[73] M. Graf and P. Vogl. *Phys. Rev. B*, 51:4940, 1995.

[74] T.G. Pedersen, K. Pedersen, and T.B. Kriestensen. *Phys. Rev. B*, 63:201101, 2001.

[75] N. Marzari and D. Vanderbilt. *Phys. Rev. B*, 56:12847, 1997.

[76] S. Baroni, A. Dal Corso, S. de Gironcoli, P. Giannozzi, C. Cavazzoni, G. Ballabio, S. Scandolo, G. Chiarotti, P. Focher, A. Pasquarello, K. Laasonen, A. Trave, R. Car, N. Marzari, and A. Kokalj. `http://www.pwscf.org/`.

[77] J.P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77:3865, 1996.

[78] G. Theurich and N.A. Hill. *Phys. Rev. B*, 64:073106, 2001.

[79] A. Dal Corso and A.M. Conte. *Phys. Rev. B*, 71:115106, 2005.

[80] S.G. Louie, S. Froyen, and M.L. Cohen. *Phys. Rev. B*, 26:1738, 1982.

[81] H.J. Monkhorst and J.D. Pack. *Phys. Rev. B*, 13:5188, 1976.

[82] N. Marzari, D. Vanderbilt, A. De Vita, and M.C. Payne. *Phys. Rev. Lett.*, 82:3296, 1999.

[83] L. Pauling. *J. Am. Chem. Soc.*, 53:1367, 1931.

[84] A.A. Mostofi, J.R. Yates, N. Marzari, I. Souza, and D. Vanderbilt. `http://www.wannier.org/`.

[85] M. Singh, C.S. Wang, and J. Callaway. *Phys. Rev. B*, 11:287, 1975.

[86] P.N. Dheer. *Phys. Rev.*, 156:637, 1967.

[87] R.W Klaffky and R.V. Coleman. *Phys. Rev. B*, 10:2915, 1974.

[88] D.J. Singh. *Planewaves, Pseudopotentials and the LAPW Method (Kluwer Academic, Boston, 1994)*.

[89] F. Giustino, J.R. Yates, I. Souza, M.L. Cohen, and S.G. Louie. *Phys. Rev. Lett.*, 98:047005, 2007.

[90] C.A. Mead. *Rev. Mod. Phys.*, 64:51, 1992.

[91] I. Souza, J. Í niguez, and D. Vanderbilt. *Phys. Rev. B*, 69:085106, 2004.

[92] This is reminiscent of the procedure used to calculate the Berry-phase polarization $\mathbf{P} = \sum_j p_j \mathbf{a}_j$ in which each p_j is computed separately by choosing the BZ to be composed of strings of k-points oriented along $\mathbf{b}_j$. More details can be found in Ref. [20].

[93] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. *Phys. Rev. Lett.*, 49:405, 1982.

[94] F. D. M. Haldane. *Phys. Rev. Lett.*, 61:2015, 1988.

[95] X. Wang, J.R. Yates, I. Souza, and D. Vanderbilt. *Phys. Rev. B*, 74:195118, 2006.

[96] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza. *Phys. Rev. B*, 75:195121, 2007.

[97] A. Dal Corso and A. M. Conte. *Phys. Rev. B*, 71:115106, 2005.

[98] G.H.O. Daalderop, P.J. Kelly, and M.F.H. Schuurmans. *Phy. Rev. B*, 41:11919, 1990.

[99] A. Kokalj. *Comp. Mater. Sci.*, 28:155, 2003.

[100] J. M. Lavine. *Phys. Rev.*, 123:1273, 1961.

[101] T. Miyasato, N. Abe, T. Fujii, A. Asamitsu, S. Onoda, Y. Onose, N. Nagaosa, and Y. Tokura. *arXiv:cond-mat/0610324v1*, 2006.

[102] W. Jellinghaus and M. P. de Anderes. *Ann. Physik.*, 7:189, 1961.

[103] K. Ohgushi, S. Miyasaka, and Y. Tokura. *J. Phys. Soc. Jpn.*, 75:013710, 2006.

[104] J.C. Slater and G.F. Koster. *Phys. Rev.*, 94:1498, 1954.

[105] G.H. Wannier. *Phys. Rev.*, 52:191, 1937.

[106] E. Kruger. *Phys. Status Solidi B*, 52:215, 1972.

[107] http://www.semiconductors.co.uk/propiiiv5653.htm.

[108] J.M. Worlock and P.A. Fleury. *Phys. Rev. Lett.*, 19:1176, 1967.

[109] T. Fukui, Y. Hatsugai, and H. Suzuki. *J. Phys. Soc. Jpn.*, 74:1674, 2005.

[110] X. Wu, O. Diéguez, K.M. Rabe, and D. Vanderbilt. *Phys. Rev. Lett.*, 97:107602, 2006.

[111] X. Wang and D. Vanderbilt. *Phys. Rev. B*, 74:054304, 2006.

[112] X. Gonze, D.C. Allan, and M.P. Teter. *Phys. Rev. Lett.*, 68:3603, 1992.

[113] X. Gonze. *Phys. Rev. A*, 52:1086, 1994.

Curriculum Vitae

Xinjie Wang

2002-2007 Ph.D. in Physics, Rutgers University, Piscataway, NJ, USA

1998-2002 BS in Physics, Peking University, Beijing, China

2004-2007 Graduate Assistant, Department of Physics and Astronomy, Rutgers University, NJ

2002-2004 Teaching Assistant, Department of Physics and Astronomy, Rutgers University, NJ

Publications

Xinjie Wang and David Vanderbilt, *First-principles perturbative computation of phonon properties of insulators in finite electric fields*, Phys. Rev. B **74**, 054304 (2006).

Xinjie Wang, Jonathan Yates, Ivo Souza, and David Vanderbilt, *Ab initio calculation of the anomalous Hall conductivity by Wannier interpolation* , Phys. Rev. B **74**, 195118 (2006).

Xinjie Wang and David Vanderbilt, *First-principles perturbative computation of dielectric and Born charge tensors in finite electric fields*, Phys. Rev. B **75**, 115116 (2007).

Jonathan Yates, Xinjie Wang, David Vanderbilt, and Ivo Souza, *Low-field Hall coefficient and magneto-optical absorption of metals from Wannier interpolation*, Phys. Rev. B **75**, 195121 (2007).

Xinjie Wang, David Vanderbilt, Jonathan R. Yates, and Ivo Souza, *Fermi-surface calculation of the anomalous Hall conductivity*, submitted to Physical Review B.