1-1-2013

Theory of the Bulk Photovoltaic effect in oxides, and First-Principles Computational Design of Materials with Bulk Dirac Points

Steve Young
University of Pennsylvania, syngularity@gmail.com

Theory of the Bulk Photovoltaic effect in oxides, and First-Principles Computational Design of Materials with Bulk Dirac Points

Abstract
Non-centrosymmetric crystals -- especially polar materials -- are capable of producing electric current in response to uniform illumination. This is called the bulk photovoltaic effect (BPVE), which we show can be identified with ``shift current'' theory. Shift currents exhibit unique physics, which are discussed and clarified. A discrete form of the expression required for numerical implementation is derived that allows for robust and efficient calculation from first-principles calculations. The response for $BaTiO_3$ and $BiFeO_3$

is calculated and found to agree well with experiment, and careful analysis of the computed response reveals how the magnitude depends on structural and chemical properties, providing criteria for the search for and design of

materials with large response. Additionally, the unique properties of shift currents allow for pure spin photocurrents in antiferromagnets with appropriate symmetry. We predict that these spin currents can be observed in $BiFeO_3$ and hematite (Fe_2O_3), and calculate the expected response. Topological insulators are a class of materials that are bulk insulators with metallic surface states that take the form of helical Dirac cones protected by time-reversal symmetry. Here we explore phenomena that occur near or at the transition between the trivial and topological insulating phase. In Bi_2Se_3, the relationship between the topological gap and material strain is investigated and used to explore the topological phase transition. At the critical strain, there exists a bulk 3D Dirac point that is analogous to the 2D Dirac points in graphene, and may be expected to exhibit similar properties. However, this 3D Dirac point is not robust and can be

easily gapped by perturbations. We propose that a 3D Dirac point marking a topological phase transition may be protected by spatial symmetries,

and outline the constraints under which symmetry groups may contain materials with such points. Based on first principles calculations, we propose BiO_2

in the β-cristobalite structure as a metastable 3D Dirac semimetal.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
Andrew M. Rappe

Keywords
ferroelectric, photovoltaic, topological insulator

Subject Categories
Chemistry | Mechanics of Materials | Physics

THEORY OF THE BULK PHOTOVOLTAIC EFFECT IN OXIDES, AND FIRST-PRINCIPLES COMPUTATIONAL DESIGN OF MATERIALS WITH BULK DIRAC POINTS

Steve M. Young

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2013

Supervisor Of Dissertation	Graduate Group Chairperson
Andrew M. Rappe	Tobias Baumgart
Professor of Chemistry	Associate Professor of Chemistry

Dissertation Committee

Joseph Subotnik	Assistant Professor of Chemistry
Charles Kane	Class of 1965 Endowed Term Chair Professor of Physics
Christopher Murray	Richard Perry University Professor of Chemistry and Materials Science and Engineering

To my mother, for all of the reasons.

Acknowledgements

First, I would like to thank my advisor, mentor, and friend, Andrew Rappe. His enthusiasm and kindness have made it a joy to work with him, and I cannot thank him enough for all the ways in which he has nurtured my development as a scientist and person.

I must also thank Charles Kane and Eugene Mele; it has been a great pleasure working with and learning from them, and I am especially grateful for their kindness, consideration, and incisive questions. Together with Andrew, they have provided stellar examples of scientific role models, both intellectually and personally.

I would like to thank my friends and colleagues Jeffrey Teo and Saad Zaheer. Both have been unfailingly generous with their time, and have taught me a great deal, some of which I have managed to learn.

I am very grateful to all the members of the Rappe group, past and present, as they have been invariably helpful and kind. In particular, I would like thank Myung-Won Lee, for (very) patiently teaching me Linux and how to run calculations.

Finally, I would like to thank my family. For what encompasses too much to mention, but I think they know anyway.

ABSTRACT

THEORY OF THE BULK PHOTOVOLTAIC EFFECT IN OXIDES, AND FIRST-PRINCIPLES COMPUTATIONAL DESIGN OF MATERIALS WITH BULK DIRAC POINTS

Steve M. Young

Andrew M. Rappe

Non-centrosymmetric crystals – especially polar materials – are capable of producing electric current in response to uniform illumination. This is called the bulk photovoltaic effect (BPVE), which we show can be identified with "shift current" theory. Shift currents exhibit unique physics, which are discussed and clarified. A discrete form of the expression required for numerical implementation is derived that allows for robust and efficient calculation from first-principles calculations. The response for $BaTiO_3$ and $BiFeO_3$ is calculated and found to agree well with experiment, and careful analysis of the computed response reveals how the magnitude depends on structural and chemical properties, providing criteria for the search for and design of materials with large response. Additionally, the unique properties of shift currents allow for pure spin photocurrents in antiferromagnets with appropriate symmetry. We predict that these spin currents can be observed in $BiFeO_3$ and hematite (Fe_2O_3), and calculate the expected response. Topological insulators are a class of materials that are bulk insulators with metallic surface states that take the form of helical Dirac cones protected by time-reversal symmetry. Here we explore phenomena that occur near or at the transition between the trivial and topological insulating phase. In Bi_2Se_3, the relationship between the topological gap and material strain is investigated and used to explore the topological phase transition. At the critical strain, there exists a bulk 3D Dirac point that is analogous to the 2D Dirac points in graphene, and may be expected to exhibit similar properties. However, this 3D Dirac point is not robust and can be easily gapped

by perturbations. We propose that a 3D Dirac point marking a topological phase transition may be protected by spatial symmetries, and outline the constraints under which symmetry groups may contain materials with such points. Based on first principles calculations, we propose BiO_2 in the β-cristobalite structure as a metastable 3D Dirac semimetal.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Bulk Photovoltaic Effect

The bulk photovoltaic effect (BPVE) refers to the generation of intrinsic photocurrents that can occur in single-phase materials lacking inversion symmetry [8, 9, 10, 11]. Ferroelectrics – materials that possess intrinsic, switchable polarization – exhibit this effect strongly, producing current in response to unpolarized, direct illumination. Traditionally, photovoltaic materials are heterogeneous, doped structures, relying on the electric field at a p-n junction to separate photoexcited electrons and holes. By contrast, the bulk photovoltaic effect can be observed even in pure homogeneous samples, as with $BaTiO_3$ [1]. Despite initial promise, several problems have discouraged efforts to apply it to the problem of solar energy collection. Early explorations revealed low energy conversion efficiency, in part due to the high band-gaps possessed by most know ferroelectrics. Additionally, despite several proposed mechanisms, the physical origin remained unclear [12, 13, 14, 15]. In combination with the relative abundance of cheap energy in the last decades of the twentieth century, interest in the phenomenon dissipated. However, recent emphasis on alternative energy technologies and the observation of the effect in novel visible-light-band-gap ferro-

electrics has renewed interest in the effect, though the origins of their photovoltaic properties are considered unresolved. Attention has been focused on interface effects, crystal orientation, and the influence of grain boundaries and defects, while any bulk photovoltaic contributions have been largely ignored [16, 17, 18, 19, 20, 21, 22, 23, 24, 25].

Recently, anomalous photovoltaic effects have been demonstrated in the multiferroic $BiFeO_3$, with reported efficiencies as high as 10% [26, 4, 5]. Furthermore, the availability of efficient *ab initio* methods and rapidly increasing computing power has opened up the possibility of designing and tuning materials to have more desirable properties, such as reduced band-gaps and stronger photovoltaic response. However, as long as the latter remains poorly understood, materials engineering efforts will be hampered. In Part I we explore the BPVE contribution. We provide a derivation of the theory of the so-called "shift current" with an emphasis on physical clarity, and provide strong evidence that it explains the BPVE. We highlight the material properties upon which BPVE most strongly depends, and using this insight propose several novel materials with significant predicted response. Finally, we show that the shift current mechanism is capable of generating pure spin photocurrents, and predict materials in which it is likely to be observed.

1.2 Topological Insulators

Phase transitions have long been identified with the breaking/reforming of symmetries. Two phases are distinguished by the value of an order parameter associated with the symmetry that distinguishes them; for the symmetry preserving phase the order parameter is zero, and for the symmetry-broken phase it becomes non-zero. In recent years, however, there has been an explosion of interest in phases of matter classified instead by topology The first known topological phase was the integer quantum hall state, properly identified in 1980 [27]. When a thin semiconductor was subjected to intense magnetic fields, the Hall

voltage became quantized. Furthermore, the resistivity was observed to be nonzero only for transitions between quantum levels. In 1982, Thouless, et al [28] demonstrated that the effect could be described by a winding around the bulk electronic states. Alternative (but equivalent) explanations of varying sophistication appeared [29, 30], but, put simply, each bulk electronic state can be characterized by some number of vortices, and the number of these vortices is related to the number of current-carrying channels that appear at the edge. Since the bulk of the material was insulating and channels only ran one direction, the carriers could not be backscattered, and no dissipation was possible. As a result, the voltage would be exactly quantized as an integer multiple of e^2/h. It was later pointed out using a graphene model that a magnetic field was not strictly necessary; merely breaking time-reversal symmetry was sufficient to create such gapless edge states [31]. In 2005 [32, 33] it was shown that in graphene spin-orbit coupling could effectively give rise to two pairs of Haldane states; as long as time-reversal symmetry was preserved, these states could not interact and a Dirac point would be required to exist at the Γ point. These states were topologically protected and could be described by a Z_2 invariant. Any time-reversal-preserving adiabatic deformation that did not close the gap in the bulk could not change the phase of a system. More specifically, the topological phase is reached when the bulk gap closes, the electronic states of the conduction band and valence band interchange, and the gap reopens with the states inverted. At the surface of the material, a spatial transition is made to the trivial state. This demands that the system become metallic at some point, yielding the requisite surface states. Since then topological insulators have been intensely studied, both theoretically and experimentally [33, 32, 7, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43], and new types of topological phases and topological phenomenon have been suggested and explored as well. While most investigations have focused on fundamental physical properties, the unique properties of the topological insulating phase suggest several practical applications, including spintronics and quantum computation [44, 45, 46, 47]. However,

significant progress towards technological applications will require deep understanding of the dependence of the fundamental physics on material structure and composition. In Part II, we explore the effects of strain on the topological phase, and consider the possibility of finding or engineering Dirac semimetals, materials held at the critical point between topological phases by crystallographic symmetry.

Chapter 2

Methods

Unless otherwise specified, all *ab initio* calculations were performed using plane-wave Density Functional Theory (DFT) with the Generalized Gradient Approximation (GGA) [48] and norm-conserving, designed non-local pseudopotentials produced using the OPIUM package [49, 50]. An energy cutoff for the wavefunction basis of 50 Ry was set. Calculations were performed including relativistic effects, including spin-orbit coupling, as appropriate. Crystal structure graphics were generated using VESTA [51].

Part I

The Bulk Photovoltaic Effect in Polar Oxides

Chapter 3

Background: Optical Response in Crystals

3.1 Position and Momentum Operators for Bloch States

The translational symmetry inherent to crystalline systems allows the energy eigenstates to be separated into a function $u_{n\mathbf{k}}(\mathbf{r})$ with the periodicity of the lattice modulated by a unitary scalar $\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}$

$$|\psi_{n\mathbf{k}}\rangle = \mathrm{e}^{i\mathbf{k}\cdot\hat{\mathbf{r}}}\,|n\mathbf{k}\rangle$$
$$\psi_{n\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$$

For a system containing N_i period units in dimension i, it is clear that k_i is constrained to be $2\pi\frac{m_i}{N_i}\mathbf{G}_i$, where m_i are integers and $\mathbf{G}_i$ are reciprocal lattice vectors. At each $\mathbf{k}$ the functions $u_{n\mathbf{k}}(\mathbf{r})$ are eigenstates, indexed by n of a Hamiltonian

$$\hat{\mathcal{H}}_{\mathbf{k}} = \frac{1}{2m_e}\,|\mathbf{k}+\hat{\mathbf{p}}|^2 + \hat{V}(\mathbf{r})$$

that lives in a Hilbert space encompassing the Wigner-Seitz Cell with periodic boundary conditions. We will often work in the "bulk" limit, where $N_i \to \infty$ and $\mathbf{k}$ becomes continuous, so that $u_{n\mathbf{k}} \to u_n(\mathbf{k})$. We draw attention to the behavior of the momentum and position operators in periodic systems.

$$\begin{aligned}
\langle \psi_{n'\mathbf{k}'} | \, \hat{\mathbf{p}} \, | \psi_{n\mathbf{k}} \rangle &= \int d^3\mathbf{r} \, d^3\mathbf{r}' \, \left\langle \psi_{n'\mathbf{k}'} \middle| \mathbf{r}' \right\rangle \langle \mathbf{r}' | \, \hat{\mathbf{p}} \, | \mathbf{r} \rangle \left\langle \mathbf{r} \middle| \psi_{n\mathbf{k}} \right\rangle \\
&= -\int d^3\mathbf{r} \, \left\langle \psi_{n'\mathbf{k}'} \middle| \mathbf{r} \right\rangle i\hbar\nabla_{\mathbf{r}} \left\langle \mathbf{r} \middle| \psi_{n\mathbf{k}} \right\rangle \\
&= -\int d^3\mathbf{r} \, \psi^{\dagger}_{n'\mathbf{k}'}(\mathbf{r}) i\hbar\nabla_{\mathbf{r}} \psi_{n\mathbf{k}}(\mathbf{r}) \\
&= -\int d^3\mathbf{r} \, \mathrm{e}^{-i\mathbf{k}'\cdot\mathbf{r}} u^{\dagger}_{n'\mathbf{k}'}(\mathbf{r}) i\hbar\nabla_{\mathbf{r}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \\
&= \hbar\mathbf{k} \int d^3\mathbf{r} \, \psi^{\dagger}_{n'\mathbf{k}'}(\mathbf{r}) \psi_{n\mathbf{k}}(\mathbf{r}) - \int d^3\mathbf{r} \, \mathrm{e}^{-i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}} u^{\dagger}_{n'\mathbf{k}'}(\mathbf{r}) i\hbar\nabla_{\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \\
&= \delta_{nn'}\delta_{\mathbf{k}\mathbf{k}'}\hbar\mathbf{k} - \delta_{\mathbf{k}\mathbf{k}'} \langle n'\mathbf{k}' | \, \hat{\mathbf{p}} \, | n\mathbf{k} \rangle
\end{aligned}$$

In the bulk limit

$$\langle \psi_{n'}(\mathbf{k}') | \, \hat{\mathbf{p}} \, | \psi_n(\mathbf{k}) \rangle = \delta_{nn'}\delta(\mathbf{k}-\mathbf{k}')\hbar\mathbf{k} - \delta(\mathbf{k}-\mathbf{k}') \langle n'\mathbf{k}' | \, \hat{\mathbf{p}} \, | n\mathbf{k} \rangle$$

We observe that only states at the same k-point have nonzero momentum coupling. Diagonal elements of the momentum matrix, which reflect the group velocity of a Bloch state, are determined purely by $\mathbf{k}$, while off-diagonal elements are determined by the momentum coupling of the bands' periodic components.

The position operator is less straightforward. Intuitively, we can see that defining the "position" of a periodic function is problematic. However, we are not without recourse. It is convenient here to begin in the bulk limit and then revert to the discrete case. Naively,

our position element is

$$\begin{aligned}\langle\psi_{n'}(\mathbf{k}')|\,\hat{\mathbf{r}}\,|\psi_n(\mathbf{k})\rangle &= \int d^3\mathbf{r}\,d^3\mathbf{r}'\,\left\langle\psi_{n'}(\mathbf{k}')\middle|\mathbf{r}'\right\rangle\langle\mathbf{r}'|\,\hat{\mathbf{r}}\,|\mathbf{r}\rangle\left\langle\mathbf{r}\middle|\psi_n(\mathbf{k})\right\rangle \\ &= \int d^3\mathbf{r}\,\left\langle\psi_{n'}(\mathbf{k}')\middle|\mathbf{r}\right\rangle\mathbf{r}\left\langle\mathbf{r}\middle|\psi_n(\mathbf{k})\right\rangle \\ &= \int d^3\mathbf{r}\,\psi_{n'}(\mathbf{k}',\mathbf{r})^\dagger\mathbf{r}\psi_n(\mathbf{k},\mathbf{r})\end{aligned}$$

By inspection it is apparent that the result will depend on our choice of cell boundary. Instead [52], we note that

$$\nabla_{\mathbf{k}}\psi_n(\mathbf{k},\mathbf{r}) = \nabla_{\mathbf{k}}\left[\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_n(\mathbf{k},\mathbf{r})\right] = i\mathbf{r}\psi_n(\mathbf{k},\mathbf{r}) + \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}\nabla_{\mathbf{k}}u_n(\mathbf{k},\mathbf{r})$$

and write the position element as

$$\begin{aligned}\langle\psi_{n'}(\mathbf{k}')|\,\hat{\mathbf{r}}\,|\psi_n(\mathbf{k},\mathbf{r})\rangle &= -i\int d^3\mathbf{r}\,\psi^\dagger_{n'}(\mathbf{k}',\mathbf{r})\nabla_{\mathbf{k}}\psi_n(\mathbf{k},\mathbf{r}) \\ &\qquad + \int d^3\mathbf{r}\,\mathrm{e}^{-i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}u^\dagger_{n'}(\mathbf{k}',\mathbf{r})i\nabla_{\mathbf{k}}u_n(\mathbf{k},\mathbf{r}) \\ &= -i\nabla_{\mathbf{k}}\int d^3\mathbf{r}\,\psi_{n'}(\mathbf{k}',\mathbf{r})^\dagger\psi_n(\mathbf{k},\mathbf{r}) \\ &\qquad + \delta(\mathbf{k}-\mathbf{k}')\int d^3\mathbf{r}\,u^\dagger_{n'}(\mathbf{k}',\mathbf{r})i\nabla_{\mathbf{k}}u_n(\mathbf{k},\mathbf{r}) \\ &= -i\delta_{nn'}\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}') \\ &\qquad + \delta(\mathbf{k}-\mathbf{k}')\,\langle n'\mathbf{k}'|\,i\nabla_{\mathbf{k}}\,|n\mathbf{k}\rangle \qquad (3.1)\end{aligned}$$

As with momentum, the expression is nonzero only for $\mathbf{k}' = \mathbf{k}$. If $n \neq n'$ then the result is well-defined, as only the second term obtains. However – consistent with our expectations when attempting to find the moment of a periodic state's density – the case of $n = n'$ appears problematic. For one, applying a unitary transformation $u_n(\mathbf{k},\mathbf{r}) \to \mathrm{e}^{i\theta(\mathbf{k})}(\mathbf{k},\mathbf{r})$

results in

$$\int d^3\mathbf{r}\, u^\dagger_{n'}(\mathbf{k},\mathbf{r}) i\nabla_{\mathbf{k}} u_n(\mathbf{k},\mathbf{r}) \to \int d^3\mathbf{r}\, u^\dagger_{n'}(\mathbf{k},\mathbf{r}) i\nabla_{\mathbf{k}} \left[\mathrm{e}^{i\theta(\mathbf{k})} u_n(\mathbf{k},\mathbf{r})\right]$$
$$\to \mathrm{e}^{i\theta(\mathbf{k})} \left[\int d^3\mathbf{r}\, u^\dagger_{n'}(\mathbf{k},\mathbf{r}) i\nabla_{\mathbf{k}} u_n(\mathbf{k},\mathbf{r}) - \nabla_{\mathbf{k}}\theta(\mathbf{k})\right]$$

Perhaps more alarming than this apparent loss of gauge invariance is the first term, which contains the derivative of a delta function and does not possess intrinsic meaning. However, we observe two things: the overall gauge invariance is actually restored by the first term, when we consider the precise expression that generated it, and the expression *can* be rendered meaningful if it appears in an integral with another function of $\mathbf{k}$. In particular, for an arbitrary function $g(\mathbf{k})$

$$\begin{aligned}
\int d^3\mathbf{k}'\, g(\mathbf{k}')\nabla_{\mathbf{k}'} \int d^3\mathbf{r}\, \psi^\dagger_n(\mathbf{k},\mathbf{r})\psi_n(\mathbf{k}',\mathbf{r}) & \\
&= -\int d^3\mathbf{k}' \left[\int d^3\mathbf{r}\, \psi^\dagger_n(\mathbf{k},\mathbf{r})\psi_n(\mathbf{k}',\mathbf{r})\right] \nabla_{\mathbf{k}'} g(\mathbf{k}') \\
&= -\int d^3\mathbf{k}'\, \delta(\mathbf{k}-\mathbf{k}')\nabla_{\mathbf{k}'} g(\mathbf{k}') \\
&= -\nabla_{\mathbf{k}} g(\mathbf{k})
\end{aligned}$$

so that

$$\int d^3\mathbf{k}'\, g(\mathbf{k}') \langle \mathbf{k}n|\, \hat{\mathbf{r}}\, |\mathbf{k}'n\rangle = i\nabla_{\mathbf{k}} g(\mathbf{k}) + g(\mathbf{k}) \langle n\mathbf{k}|\, i\nabla_{\mathbf{k}}\, |n\mathbf{k}\rangle \tag{3.2}$$

Similarly,

$$\int d^3\mathbf{k}'\, g(\mathbf{k}') \langle \mathbf{k}'n|\, \hat{\mathbf{r}}\, |\mathbf{k}n\rangle = -i\nabla_{\mathbf{k}} g(\mathbf{k}) + g(\mathbf{k}) \langle n\mathbf{k}|\, i\nabla_{\mathbf{k}}\, |n\mathbf{k}\rangle \tag{3.3}$$

To obtain the discrete limit, we note that

$$\frac{dA}{dk} = A\frac{d\ln A}{dk} = A\lim_{\Delta k\to 0}\frac{\ln A(k+\Delta k)-\ln A(k)}{\Delta k} = A\lim_{\Delta k\to 0}\ln\frac{A(k+\Delta k)}{\ln A(k)}\frac{1}{\Delta k}$$

allowing us to rewrite the above equations as

$$\begin{aligned}
&\lim_{\Delta\mathbf{k}'\to 0}\sum|\Delta\mathbf{k}'|g(\mathbf{k}')\left\langle\mathbf{k}n\right|\hat{\mathbf{r}}\left|\mathbf{k}'n\right\rangle \\
&\qquad= \lim_{\Delta\mathbf{k}\to 0}\frac{1}{|\Delta\mathbf{k}|}\Big[ig(\mathbf{k})\ln\frac{g(\mathbf{k}+\Delta\mathbf{k})}{\ln g(\mathbf{k})} \\
&\qquad\qquad + g(\mathbf{k})\int d^3\mathbf{r}\,u_n(\mathbf{k},\mathbf{r})iu_n(\mathbf{k},\mathbf{r})\frac{\ln\int d^3\mathbf{r}\,u_n^\dagger(\mathbf{k},\mathbf{r})u_n(\mathbf{k}+\Delta\mathbf{k},\mathbf{r})}{\ln\int d^3\mathbf{r}\,u_n^\dagger(\mathbf{k},\mathbf{r})u_n(\mathbf{k},\mathbf{r})}\Big] \\
&\qquad= \lim_{\Delta\mathbf{k}\to 0}\frac{1}{|\Delta\mathbf{k}|}\Big[ig(\mathbf{k})\ln\frac{g(\mathbf{k}+\Delta\mathbf{k})}{\ln g(\mathbf{k})} + ig(\mathbf{k})\ln\left\langle n'\mathbf{k}'\middle|n(\mathbf{k}+\Delta\mathbf{k})\right\rangle\Big]
\end{aligned}$$

We note that $\langle n'\mathbf{k}'|\, i\nabla_{\mathbf{k}}\,|n\mathbf{k}\rangle$ is the celebrated Berry connection, and for compactness will represent it by $\chi_n(\mathbf{k})$ in what follows.

Finally, we may relate position and momentum elements using Heisenberg's equations of motion

$$\begin{aligned}
\frac{d\hat{\mathbf{r}}}{dt} &= \frac{1}{\hbar}\left[\hat{\mathcal{H}},\hat{\mathbf{r}}\right] \\
\frac{1}{m_e}\hat{\mathbf{p}} &= \frac{1}{\hbar}\left(\hat{\mathcal{H}}\hat{\mathbf{r}}-\hat{\mathbf{r}}\hat{\mathcal{H}}\right) \\
\frac{1}{m_e}\left\langle\psi_{n'}(\mathbf{k})\right|\hat{\mathbf{p}}\left|\psi_n(\mathbf{k})\right\rangle &= \frac{1}{\hbar}\left\langle\psi_{n'}(\mathbf{k})\right|\hat{\mathcal{H}}\hat{\mathbf{r}}\left|\psi_n(\mathbf{k})\right\rangle - \frac{1}{\hbar}\left\langle\psi_{n'}(\mathbf{k})\right|\hat{\mathbf{r}}\hat{\mathcal{H}}\left|\psi_n(\mathbf{k})\right\rangle \\
\left\langle\psi_{n'}(\mathbf{k})\right|\hat{\mathbf{p}}\left|\psi_n(\mathbf{k})\right\rangle &= \frac{m_e}{\hbar}\left[E_{n'}(\mathbf{k})-E_n(\mathbf{k})\right]\left\langle\psi_{n'}(\mathbf{k})\right|\hat{\mathbf{r}}\left|\psi_n(\mathbf{k})\right\rangle
\end{aligned} \tag{3.4}$$

This expression is only well-defined for off-diagonal elements of the momentum and position operators.

3.2 Light-Matter Interaction

The interaction of electromagnetic (EM) radiation with matter is inherently time-dependent, and the evolution of the system will be governed by the time-dependent Schroedinger equation. For a two-state system we may solve this analytically, but in more complex cases, we must resort to approximation. From Schroedinger's equation

$$d\left|\psi\right\rangle = -\frac{i}{\hbar}\hat{\mathcal{H}}\left|\psi(t)\right\rangle dt$$
$$\left|\psi(t+dt)\right\rangle = \left(1-\frac{i}{\hbar}\hat{\mathcal{H}}\right)\left|\psi(t)\right\rangle$$

It is convenient at this point to leave the Schroedinger picture and apply the evolution due to the unperturbed Hamiltonian to the operators instead of the wavefunctions. This mix of Schroedinger and Heisenberg pictures is called the interaction picture, and allows us to think about and compute the evolution of the system due only to the interaction. In general, $\hat{A}^I = \mathrm{e}^{-\frac{i}{\hbar}\hat{\mathcal{H}}^{(0)}t}\hat{A}\mathrm{e}^{\frac{i}{\hbar}\hat{\mathcal{H}}^{(0)}t}$ for an operator $\hat{A}$. The operator $\hat{U}(t) = 1 - \frac{i}{\hbar}\hat{\mathcal{H}}^I dt$ evolves the system infinitesimally at time t; applying iteratively to effect finite-time evolution and collecting terms

$$\begin{aligned}
&\ldots\hat{U}(t_3')\hat{U}(t_2')\hat{U}(t_1')\left|\Psi^{(0)}\right\rangle \\
&\quad= \ldots(1-i\frac{\hat{\mathcal{H}}^I(t_3')}{\hbar}dt_3')(1-i\frac{\hat{\mathcal{H}}^I(t_2')}{\hbar}dt_2')(1-i\frac{\hat{\mathcal{H}}^I(t_1')}{\hbar}dt_1')\left|\Psi^{(0)}\right\rangle \\
&\quad= \left|\Psi^{(0)}\right\rangle - \sum_{t_1}\left(i\frac{\hat{\mathcal{H}}^I(t_1)}{\hbar}dt_1\right)\left|\Psi^{(0)}\right\rangle - \sum_{t_1,t_2}\left(\frac{\hat{\mathcal{H}}^I(t_2)}{\hbar}\frac{\hat{\mathcal{H}}^I(t_1)}{\hbar}dt_1dt_2\right)\left|\Psi^{(0)}\right\rangle\ldots \\
&\quad= \left|\Psi^{(0)}\right\rangle - \int_{-\infty}^{t}\left(i\frac{\hat{\mathcal{H}}^I(t_1)}{\hbar}dt_1\right)\left|\Psi^{(0)}\right\rangle \\
&\qquad\qquad - \int_{-\infty}^{t}\int_{-\infty}^{t_2}\left(\frac{\hat{\mathcal{H}}^I(t_2)}{\hbar}\frac{\hat{\mathcal{H}}^I(t_1)}{\hbar}dt_1dt_2\right)\left|\Psi^{(0)}\right\rangle\ldots
\end{aligned}$$

The finite time evolution is an infinite sum over different orders of interaction, with each term an integral over all possible sequences of events with proper time ordering. The first-order term accounts for the total effect of the perturbation on the unperturbed system, the second-order term describes the effect of the perturbation on states that have already been perturbed once, and so on.

The perturbing potential due to radiation-matter interaction is obtained by replacing the conjugate momentum with the canonical momentum, so that $\hat{\mathbf{p}} \to \hat{\mathbf{p}} - e\hat{\mathbf{A}}$. Under this substitution, representing the kinetic energy operator by $\hat{\mathcal{T}}$,

$$\begin{aligned}\hat{\mathcal{T}} &= \frac{1}{2m_e}\left|\hat{\mathbf{p}} - e\hat{\mathbf{A}}\right|^2 \\ &= \frac{1}{2m_e}\left|\hat{\mathbf{p}}\right|^2 - \frac{e}{2m_e}\left(\hat{\mathbf{p}}\cdot\hat{\mathbf{A}} + \hat{\mathbf{A}}\cdot\hat{\mathbf{p}}\right) + \frac{1}{2m_e}\left|\hat{\mathbf{A}}\right|^2\end{aligned}$$

The interaction with the EM field now conveniently appears as a perturbation. A common approximation that we shall adopt is the dipole approximation, wherein the last term is dropped. Under the Coulomb gauge, $\hat{\mathbf{p}}$ and $\hat{\mathbf{A}}$ commute, so that our perturbation is now

$$\hat{V} = -\frac{e}{m_e}\hat{\mathbf{A}}\cdot\hat{\mathbf{p}}$$

Quite frequently it is appropriate to use the classical description of the EM field. However, while our final results are not sensitive to this approximation, we will nonetheless move forward with a fully quantized description of radiation for clarity. Adopting second-quantization formalism and the interaction picture

$$\begin{aligned}\hat{V}^I(t) = -\sum_{n',n''}\left\langle n''\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n'\right\rangle \hat{c}^\dagger_{n''}\hat{c}_{n'} \\ \times\left[\mathrm{e}^{-i(-\omega-\omega_{n''}+\omega_{n'})t}\hat{a}^\dagger(\omega) + \mathrm{e}^{-i(\omega-\omega_{n''}+\omega_{n'})t}\hat{a}(\omega)\right]\end{aligned}$$

where $\hat{c}_n$ annihilates the electronic state indexed by n, and $\hat{a}(\omega)$ and $\hat{a}^\dagger(\omega)$ annihilate and create, respectively, a photon of frequency ω and polarization direction given by $\mathbf{n}_A(\omega)$ (where we set $|\mathbf{n}_A(\omega)| = \sqrt{\frac{\hbar}{2\epsilon_0\omega}}$). We have implicitly imposed the long-wavelength approximation, assuming that we will deal with radiation of wavelengths much longer than the atomic scale.

The infinitesimal-time evolution operator is then

$$\hat{U}(t) = 1 - i\frac{\hat{V}^I(t)}{\hbar}dt$$

3.3 Linear Response

We will first examine the linear response of a system to EM radiation. The linear response describes the dielectric behavior of the electronic system, including absorption. Furthermore, as will be seen later, the shift current response, though formally second-order, is in some sense a property of linear excitations.

Applying the infinitesimal-time-evolution operator once to a non-interacting ground state $\left|\Psi^{(0)}\right\rangle = \left|\Psi^{(0)}_{\mathrm{EM}}\right\rangle \left|\Psi^{(0)}_{\mathrm{el}}\right\rangle$,

$$\begin{aligned}
\left|\Psi^{(1)}\right\rangle &= \hat{U}(t_1)\left|\Psi^{(0)}\right\rangle \\
&= \left|\Psi^{(0)}\right\rangle + \frac{i}{\hbar}\sum_{n',n''}\langle n'|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1}\hat{a}(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\left|\Psi^{(0)}\right\rangle dt_1 \\
&\qquad + \frac{i}{\hbar}\sum_{n',n''}\langle n'|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_1}\hat{a}^\dagger(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\left|\Psi^{(0)}\right\rangle dt_1
\end{aligned}$$

The two terms correspond to emission and absorption. We will apply the rotating-wave

approximation and consider only absorption. The current is

$$\begin{aligned}
\mathbf{J}(\omega,t) =& \left\langle \Psi^{(0)} \right| \hat{U}(t_1)^\dagger \hat{\mathbf{J}}^I(t) \hat{U}(t_1) \left| \Psi^{(0)} \right\rangle \\
=& \frac{1}{\hbar^2} \sum_{n',n''} \Bigg[\left\langle \Psi^{(0)} \right| \hat{c}^\dagger_{n''} \hat{c}_{n'} \hat{a}^\dagger(\omega) \mathrm{e}^{i(\omega-\omega_{n'}+\omega_{n''})t_1} \left\langle n'' \right| \frac{e}{m_e} \mathbf{n}_A(\omega) \cdot \hat{\mathbf{p}} \left| n' \right\rangle \\
& \times \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega) \cdot \hat{\mathbf{p}} \left| n'' \right\rangle \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1} \hat{a}(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \left| \Psi^{(0)} \right\rangle \Bigg] dt_1^2 \\
&+ \frac{i}{\hbar} \sum_{n',n''} \Bigg[\left\langle \Psi^{(0)} \right| \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega) \cdot \hat{\mathbf{p}} \left| n'' \right\rangle \\
& \times \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1} \hat{a}(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \left| \Psi^{(0)} \right\rangle - c.c. \Bigg] dt_1
\end{aligned}$$

The first term vanishes, as it is simply the expected current of an eigenstate. Since

$$\hat{\mathbf{J}}^I(t) = -\frac{e}{m_e} \hat{\mathbf{p}}^I(t) = -\frac{e}{m_e} \sum_{m',m''} \left\langle m' \right| \hat{\mathbf{p}} \left| m'' \right\rangle \mathrm{e}^{i(\omega_{m'}-\omega_{m''})t} \hat{c}^\dagger_{m'} \hat{c}_{m''}$$

we have

$$\begin{aligned}
\mathbf{J}(\omega,t) =& -\frac{e}{m_e} \frac{i}{\hbar} \sum_{n',n''} \sum_{m',m''} \left\langle \Psi^{(0)} \right| \hat{c}^\dagger_{m'} \hat{c}_{m''} \left\langle m' \right| \hat{\mathbf{p}} \left| m'' \right\rangle \mathrm{e}^{i(\omega_{m'}-\omega_{m''})t} \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega) \cdot \hat{\mathbf{p}} \left| n'' \right\rangle \\
& \times \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1} \hat{a}(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \left| \Psi^{(0)} \right\rangle dt_1 + c.c. \\
=& -\frac{e}{m_e} \frac{i}{2\hbar} \sum_{n',n''} f_{n''} \left[1 - f_{n'} \right] \left\langle n'' \right| \hat{\mathbf{p}} \left| n' \right\rangle \left\langle n' \right| \frac{e}{m_e} \mathbf{A}^0 \cdot \hat{\mathbf{p}} \left| n'' \right\rangle \\
& \times \mathrm{e}^{i(\omega_{n''}-\omega_{n'})t} \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1} dt_1 + c.c.
\end{aligned}$$

where we have used $\left\langle \Psi^{(0)}_{\mathrm{EM}} \right| \mathbf{n}_A(\omega) \hat{a}(\omega) \left| \Psi^{(0)}_{\mathrm{EM}} \right\rangle \approx \frac{\mathbf{A}^0}{2}$. We will now integrate over all allowed events. For the integral to be well-behaved, we must multiply the integrand by an exponential $\mathrm{e}^{\eta t}$ and only allow $\eta \to 0$ (i.e., take the adiabatic limit) after performing the

integration.

$$
\begin{aligned}
\mathbf{J}(\omega,t) =& -\frac{e}{m_e}\frac{i}{2\hbar}\sum_{n',n''} f_{n''}\left[1-f_{n'}\right]\langle n''|\,\hat{\mathbf{p}}\,|n'\rangle\langle n'|\,\frac{e}{m_e}\mathbf{A}^0\cdot\hat{\mathbf{p}}\,|n''\rangle \\
&\qquad\times \mathrm{e}^{-i(\omega_{n''}-\omega_{n'})t}\lim_{\eta\to 0}\int_{-\infty}^{t}\mathrm{e}^{\eta t}\mathrm{e}^{-i(\omega-\omega_{n'}-\omega_{n''})t_1}dt_1 + c.c. \\
=& -\frac{e}{m_e}\frac{i}{2\hbar}\sum_{n',n''} f_{n''}\left[1-f_{n'}\right]\langle n''|\,\hat{\mathbf{p}}\,|n'\rangle\langle n'|\,\frac{e}{m_e}\mathbf{A}^0\cdot\hat{\mathbf{p}}\,|n''\rangle \\
&\qquad\times \mathrm{e}^{-i\omega t}\lim_{\eta\to 0}\mathrm{e}^{\eta t}\frac{1}{-i(\omega-\omega_{n'}+\omega_{n''})+\eta} + c.c. \\
=& -\frac{e}{m_e}\frac{i}{2\hbar}\sum_{n',n''} f_{n''}\left[1-f_{n'}\right]\langle n''|\,\hat{\mathbf{p}}\,|n'\rangle\langle n'|\,\frac{e}{m_e}\mathbf{A}^0\cdot\hat{\mathbf{p}}\,|n''\rangle \\
&\qquad\times \mathrm{e}^{-i\omega t}\left[i\mathcal{P}\frac{1}{\omega-\omega_{n'}+\omega_{n''}}+\pi\delta(\omega-\omega_{n'}+\omega_{n''})\right] + c.c.
\end{aligned}
$$

Since $\frac{d}{dt}\mathbf{A}^0\mathrm{e}^{-i\omega t} = -i\omega\mathbf{A}^0\mathrm{e}^{-i\omega t} = \mathbf{E}^0\mathrm{e}^{-i\omega t}$

$$
\begin{aligned}
\mathbf{J}(\omega,t) =& \left(\frac{e}{m_e}\right)^2\frac{1}{2\hbar\omega}\sum_{n',n''} f_{n''}\left[1-f_{n'}\right]\langle n''|\,\hat{\mathbf{p}}\,|n'\rangle\langle n'|\,\mathbf{E}^0\cdot\hat{\mathbf{p}}\,|n''\rangle \\
&\qquad\times \mathrm{e}^{-i\omega t}\left[i\mathcal{P}\frac{1}{\omega-\omega_{n'}+\omega_{n''}}+\pi\delta(\omega-\omega_{n'}+\omega_{n''})\right] + c.c.
\end{aligned}
$$

Polarization can be obtained by noting that $\mathbf{J} = \frac{d\mathbf{P}}{dt}$, so that

$$\begin{aligned}
\mathbf{P}(\omega, t) = &\left(\frac{e}{m_e}\right)^2 \frac{1}{2\hbar\omega^2} \sum_{n',n''} f_{n''}\left[1 - f_{n'}\right] \langle n''| \hat{\mathbf{p}} |n'\rangle \langle n'| \mathbf{E}^0 \cdot \hat{\mathbf{p}} |n''\rangle \\
&\times \mathrm{e}^{-i\omega t}\left[-\mathcal{P}\frac{1}{\omega - \omega_{n'} + \omega_{n''}} + i\pi\delta(\omega - \omega_{n'} + \omega_{n''})\right] + c.c. \\
P_i(\omega, t) = &\left(\frac{e}{m_e}\right)^2 \frac{1}{\hbar\omega^2} \sum_{n',n''} f_{n''}\left[1 - f_{n'}\right] \langle n''| p_i |n'\rangle \langle n'| p_j |n''\rangle \\
&\times \left[\mathcal{P}\frac{1}{\omega_{n'} - \omega_{n''} - \omega} + i\pi\delta(\omega - \omega_{n'} + \omega_{n''})\right] \frac{E_j^0}{2}\mathrm{e}^{-i\omega t} + c.c. \\
P_i(\omega, t) = &\epsilon_0 \Re\left[\chi_{ij}(\omega) E_j^0(\omega, t)\right]
\end{aligned}$$

χ is the dielectric function, and the permittivity $\epsilon = 1 + \chi$. The expression above reveals several important features of χ. First, it obeys the Kramers-Kronig relations, which are a consequence of preserving causality. We observe that this arises automatically from properly performing the time integral. Second, the real and imaginary components correspond to virtual and resonant excitations, respectively. The latter may result in real absorption events, and therefore dissipation, while the former may not. We can confirm this by noting that work may only be done via the imaginary component:

$$\begin{aligned}
W = &\int \frac{dP}{dt} \cdot \mathbf{E}(\omega, t) dt \\
= &\int \epsilon_0 \omega \left|\mathbf{E}^0(\omega)\right|^2 \left[\chi^{\Im} \cos^2(\omega t) - \chi^{\Re} \sin(\omega t)\cos(\omega t)\right] dt \\
= &\frac{\epsilon_0}{2} \left|\mathbf{E}^0(\omega)\right|^2 \chi^{\Im} t
\end{aligned}$$

We may derive an expression for the absorption coefficient from Maxwell's Equations

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= 0 \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \times \mathbf{B} &= \frac{\mu}{c}\sigma^{\Re}\mathbf{E} + \frac{\mu}{c}\epsilon^{\Re}\frac{\partial \mathbf{E}}{\partial t}
\end{aligned}$$

where $\mathbf{E}$ remains the electric field, $\mathbf{B}$ is the magnetic field, μ is the permeability, c is the speed of light, $\epsilon^{\Re}$ is the real permittivity, and σ is the real conductivity. Since

$$\sigma\mathbf{E} = \mathrm{Tr}\left(\hat{\rho}\hat{\mathbf{J}}\right) = \mathrm{Tr}\left(\hat{\rho}\frac{\partial \hat{\mathbf{P}}}{\partial t}\right)$$

where $\mathbf{J}$ is current density, we have

$$\sigma^{\Re} = \Re\left(-i\omega\chi\right) = \omega\epsilon^{\Im}$$

Manipulating to produce second-order differential equations for E

$$\begin{aligned}
\nabla \times \nabla \times \mathbf{E} &= -\frac{1}{c}\frac{\partial \nabla \times \mathbf{B}}{\partial t} \\
\nabla \times \nabla \times \mathbf{E} &= -\frac{\mu}{c^2}\omega\epsilon^{\Im}\frac{\partial \mathbf{E}}{\partial t} - \frac{\mu}{c^2}\epsilon^{\Re}\frac{\partial^2 \mathbf{E}}{\partial t^2} \\
\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} &= -\frac{\mu}{c^2}\omega\epsilon^{\Im}\frac{\partial \mathbf{E}}{\partial t} - \frac{\mu}{c^2}\epsilon^{\Re}\frac{\partial^2 \mathbf{E}}{\partial t^2} \\
-\nabla^2\mathbf{E} &= -\frac{\mu}{c^2}\left(\omega\epsilon^{\Im}\frac{\partial \mathbf{E}}{\partial t} + \epsilon^{\Re}\frac{\partial^2 \mathbf{E}}{\partial t^2}\right)
\end{aligned}$$

Now, supposing a solution of the form $\mathbf{E}^0 \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}\mathrm{e}^{-i\omega t}$, where $\mathbf{k}$ is a complex wavevector, we

arrive at

$$(\mathbf{k}\cdot\mathbf{k}) = \frac{\mu}{c^2}\omega^2\left(i\epsilon^{\Im}+\epsilon^{\Re}\right)$$

Solving for $\mathbf{k}$ (assuming propagation direction Z)

$$\begin{aligned}
k_z^2 =& \frac{\mu}{c^2}\omega^2\left(i\epsilon^{\Im}+\epsilon^{\Re}\right)\\
k_z =& \pm\frac{\sqrt{\mu}}{c}\omega\left(\sqrt{\frac{\sqrt{\left(\epsilon^{\Re}\right)^2+\left(\epsilon^{\Im}\right)^2}+\epsilon^{\Re}}{2}}+i\sqrt{\frac{\sqrt{\left(\epsilon^{\Re}\right)^2+\left(\epsilon^{\Im}\right)^2}-\epsilon^{\Re}}{2}}\right)
\end{aligned}$$

From Beer's law

$$\begin{aligned}
I \propto |\mathbf{E}(z)|^2 &= \left|\mathbf{E}^0\right|^2 \mathrm{e}^{-2k_z^{\Im}z}\\
I = I^0\mathrm{e}^{-2k_z^{\Im}z} &= I^0\mathrm{e}^{-\alpha z}
\end{aligned}$$

Thus, the absorption coefficient is defined as

$$\alpha = 2k_z^{\Im} = 2\frac{\sqrt{\mu}}{c}\omega\sqrt{\frac{\sqrt{\left(\epsilon^{\Re}\right)^2+\left(\epsilon^{\Im}\right)^2}-\epsilon^{\Re}}{2}}$$

3.4 Conventional photovoltaics

Traditionally, photovoltaic effects rely on thermally excited carriers being separated and prevented from recombining by an electric field. In the most common case, a p-n junction is created by interfacing a hole-doped semiconductor with one that is electron-doped. The chemical potential of the electron-doped region lies above that of the hole-doped region, and electrons will flow into holes until the resulting electric field balances the chemical

potential gradient. Carriers excited in the region near the interfaces respond to this field and separate, reducing the likelihood of recombination. However, these carriers must obey detailed-balance; as a result the photovoltaic efficiency is fundamentally constrained, and the photovoltage cannot exceed the band-gap. As we shall see, shift current is a non-linear effect, and the carriers violate the principle of detailed-balance, allowing for photovoltages well above the band-gap.

Chapter 4

Shift Current

The theory of shift current was first introduced over 30 years ago. It is a second-order optical effect, and may be derived from time-dependent perturbation theory. Previous treatments, however, rely on a classical EM field and/or the density matrix formalism [13, 15, 53]. For the following we will use an explicitly quantized EM field, and operate at the level of the wavefunction; while the resulting expression is identical, we feel this approach more readily reveals the underlying physics.

4.1 Derivation and Physical Interpretation

For a non-interacting ground state $\left|\Psi^{(0)}\right\rangle = \left|\Psi^{(0)}_{\mathrm{EM}}\right\rangle \left|\Psi^{(0)}_{\mathrm{el}}\right\rangle$ we are interested in the result of second-order processes

$$\left|\Psi^{(2)}\right\rangle = \hat{U}(t_2)\hat{U}(t_1)\left|\Psi^{(0)}\right\rangle$$

We can expect two contributions to arise: one from terms that combine two positive or negative frequency waves, and one from terms with both positive and negative frequency components. The former constitutes second harmonic generation and generates emission

of doubled frequency. The latter, however, will result in zero-frequency – or dc-like – emission, rather than an oscillating dipole. While this contribution is usually classified as optical rectification, which results in a static polarization, a portion of such terms will conspire to produce current. Expanding,

$$
\begin{aligned}
\left|\Psi^{(2)}\right\rangle =& \hat{U}(t_2)\hat{U}(t_1)\left|\Psi^{(0)}\right\rangle \\
=& \left|\Psi^{(0)}\right\rangle \\
&- \frac{i}{\hbar}\sum_{n',n''}\left\langle n'\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n''\right\rangle\left[\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_1}\hat{a}^\dagger(\omega)\right. \\
&\qquad\left.+\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1}\hat{a}(\omega)\right]\hat{c}^\dagger_{n'}\hat{c}_{n''}\left|\Psi^{(0)}\right\rangle dt_1 \\
&- \frac{i}{\hbar}\sum_{n''',n''''}\left\langle n'''\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n''''\right\rangle\left[\mathrm{e}^{-i(-\omega-\omega_{n'''}+\omega_{n''''})t_2}\hat{a}^\dagger(\omega)\right. \\
&\qquad\left.+\mathrm{e}^{-i(\omega-\omega_{n'''}+\omega_{n''''})t_2}\hat{a}(\omega)\right]\hat{c}^\dagger_{n'''}\hat{c}_{n''''}\left|\Psi^{(0)}\right\rangle dt_2 \\
&- \frac{1}{\hbar^2}\sum_{n^{(i)}}\left\langle n'\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n''\right\rangle\left\langle n''''\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n'''\right\rangle \\
&\qquad\times\left[\mathrm{e}^{-i(-\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_2}\hat{a}^\dagger(\omega)\hat{a}^\dagger(\omega)\right. \\
&\qquad\left.+\mathrm{e}^{-i(\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_2}\hat{a}(\omega)\hat{a}(\omega)\right] \\
&\qquad\times\hat{c}^\dagger_{n'}\hat{c}_{n''}\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\left|\Psi^{(0)}\right\rangle dt_1dt_2 \\
&- \frac{1}{\hbar^2}\sum_{n^{(i)}}\left\langle n'\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n''\right\rangle\left\langle n''''\right|\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\left|n'''\right\rangle \\
&\qquad\times\left[\mathrm{e}^{-i(\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_2}\hat{a}^\dagger(\omega)\hat{a}(\omega)\right. \\
&\qquad\left.+\mathrm{e}^{-i(-\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_2}\hat{a}(\omega)\hat{a}^\dagger(\omega)\right] \\
&\qquad\times\hat{c}^\dagger_{n'}\hat{c}_{n''}\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\left|\Psi^{(0)}\right\rangle dt_1dt_2
\end{aligned}
$$

We see that there are terms corresponding to single photon absorption(emission), two photon absorption(emission), and net zero photon absorption, as expected. Taking the current,

and keeping only the terms $(\omega, -\omega)$

$$
\begin{aligned}
\mathbf{J}(\omega,t) =& \left\langle \Psi^{(0)} \right| \hat{U}(t_1)^\dagger \hat{U}(t_2)^\dagger \hat{\mathbf{J}}^I(t) \hat{U}(t_2)\hat{U}(t_1) \left| \Psi^{(0)} \right\rangle \\
=& \frac{1}{\hbar^2} \sum_{n^{(i)}} \Bigg[\left\langle \Psi^{(0)} \right| \hat{c}^\dagger_{n''''} \hat{c}_{n'''} \hat{a}(\omega) \mathrm{e}^{i(-\omega-\omega_{n'''}+\omega_{n''''})t_2} \\
& \qquad \times \left\langle n'''' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n''' \right\rangle \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n'' \right\rangle \\
& \qquad \times \mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_1} \hat{a}^\dagger(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \left| \Psi^{(0)} \right\rangle + c.c. \Bigg] dt_1 dt_2 \\
&+ \frac{1}{\hbar^2} \sum_{n^{(i)}} \Bigg[\left\langle \Psi^{(0)} \right| \hat{c}^\dagger_{n''''} \hat{c}_{n'''} \hat{a}^\dagger(\omega) \mathrm{e}^{i(\omega-\omega_{n'''}+\omega_{n''''})t_2} \\
& \qquad \times \left\langle n'''' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n''' \right\rangle \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n'' \right\rangle \\
& \qquad \times \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1} \hat{a}(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \left| \Psi^{(0)} \right\rangle + c.c. \Bigg] dt_1 dt_2 \\
&- \frac{1}{\hbar^2} \sum_{n^{(i)}} \Bigg[\left\langle \Psi^{(0)} \right| \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n'' \right\rangle \left\langle n'''' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n''' \right\rangle \\
& \qquad \times \mathrm{e}^{-i(\omega-\omega_{n''''}+\omega_{n'''})t_1} \mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_2} \\
& \qquad \times \hat{a}^\dagger(\omega) \hat{a}(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \hat{c}^\dagger_{n''''} \hat{c}_{n'''} \left| \Psi^{(0)} \right\rangle + c.c. \Bigg] dt_1 dt_2 \\
&- \frac{1}{\hbar^2} \sum_{n^{(i)}} \Bigg[\left\langle \Psi^{(0)} \right| \hat{\mathbf{J}}^I(t) \left\langle n' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n'' \right\rangle \left\langle n'''' \right| \frac{e}{m_e} \mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}} \left| n''' \right\rangle \\
& \qquad \times \mathrm{e}^{-i(-\omega-\omega_{n''''}+\omega_{n'''})t_1} \mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_2} \\
& \qquad \times \hat{a}(\omega) \hat{a}^\dagger(\omega) \hat{c}^\dagger_{n'} \hat{c}_{n''} \hat{c}^\dagger_{n''''} \hat{c}_{n'''} \left| \Psi^{(0)} \right\rangle + c.c. \Bigg] dt_1 dt_2
\end{aligned}
$$

Since $\hat{\mathbf{J}}^I(t) = -\frac{e}{m_e}\hat{\mathbf{p}}^I(t) = -\frac{e}{m_e}\sum_{m',m''}\langle m'|\,\hat{\mathbf{p}}\,|m''\rangle\, \mathrm{e}^{i(\omega_{m'}-\omega_{m''})t}\hat{c}^\dagger_{m'}\hat{c}_{m''}$

$$
\begin{aligned}
\mathbf{J}(\omega,t) = &-\frac{e}{m_e}\frac{1}{\hbar^2}\sum_{n^{(i)},m',m''}\Bigg[\mathrm{e}^{i(\omega_{m'}-\omega_{m''})t}\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\omega-\omega_{n''''}+\omega_{n'''})t_2}\\
&\qquad\times\langle n'|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\langle n''''|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n'''\rangle\,\langle m'|\,\hat{\mathbf{p}}\,|m''\rangle\\
&\qquad\times\left\langle\Psi^{(0)}\right|\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\hat{a}(\omega)\hat{c}^\dagger_{m'}\hat{c}_{m''}\hat{a}^\dagger(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\left|\Psi^{(0)}\right\rangle + c.c.\Bigg]dt_1dt_2\\
&-\frac{e}{m_e}\frac{1}{\hbar^2}\sum_{n^{(i)},m',m''}\Bigg[\mathrm{e}^{i(\omega_{m'}-\omega_{m''})t}\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(-\omega-\omega_{n''''}+\omega_{n'''})t_2}\\
&\qquad\times\langle n'|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\langle n''''|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n'''\rangle\,\langle m'|\,\hat{\mathbf{p}}\,|m''\rangle\\
&\qquad\times\left\langle\Psi^{(0)}\right|\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\hat{a}^\dagger(\omega)\hat{c}^\dagger_{m'}\hat{c}_{m''}\hat{a}(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\left|\Psi^{(0)}\right\rangle + c.c.\Bigg]dt_1dt_2\\
&+\frac{e}{m_e}\frac{1}{\hbar^2}\sum_{n^{(i)},m',m''}\Bigg[\mathrm{e}^{i(\omega_{m'}-\omega_{m''})t}\mathrm{e}^{-i(\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(-\omega-\omega_{n'}+\omega_{n''})t_2}\\
&\qquad\times\langle n'|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\langle n''''|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n'''\rangle\,\langle m'|\,\hat{\mathbf{p}}\,|m''\rangle\\
&\qquad\times\left\langle\Psi^{(0)}\right|\hat{c}^\dagger_{m'}\hat{c}_{m''}\hat{a}^\dagger(\omega)\hat{a}(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\left|\Psi^{(0)}\right\rangle + c.c.\Bigg]dt_1dt_2\\
&+\frac{e}{m_e}\frac{1}{\hbar^2}\sum_{n^{(i)},m',m''}\Bigg[\mathrm{e}^{i(\omega_{m'}-\omega_{m''})t}\mathrm{e}^{-i(-\omega-\omega_{n''''}+\omega_{n'''})t_1}\mathrm{e}^{-i(\omega-\omega_{n'}+\omega_{n''})t_2}\\
&\qquad\times\langle n'|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n''\rangle\,\langle n''''|\,\frac{e}{m_e}\mathbf{n}_A(\omega)\cdot\hat{\mathbf{p}}\,|n'''\rangle\,\langle m'|\,\hat{\mathbf{p}}\,|m''\rangle\\
&\qquad\times\left\langle\Psi^{(0)}\right|\hat{c}^\dagger_{m'}\hat{c}_{m''}\hat{a}(\omega)\hat{a}^\dagger(\omega)\hat{c}^\dagger_{n'}\hat{c}_{n''}\hat{c}^\dagger_{n''''}\hat{c}_{n'''}\left|\Psi^{(0)}\right\rangle + c.c.\Bigg]dt_1dt_2
\end{aligned}
$$

Each of the four terms is nonzero under only two sets of (nontrivial) conditions. In the first two terms, which arise from the superposition of two different absorptions, either $n'''' = n''$, $m' = n'''$, and $m'' = n'$, or $n'''' = m''$, $m' = n''$, and $n''' = n'$. In the last two terms, which arise from absorptions followed by emissions, either $n'''' = n''$, $m' = n'''$, and $m'' = n'$, or $n'''' = m''$, $m' = n''$, and $n''' = n'$. In both cases, the first set of

conditions correspond to electron excitations, and the second to hole excitations. Since $\langle \Psi^0_{\mathrm{EM}} |\, \mathbf{n}_A(\omega)\mathbf{n}_A(\omega)\hat{a}^\dagger(\omega)\hat{a}(\omega)\, |\Psi^0_{\mathrm{EM}}\rangle \approx \frac{\mathbf{A}^0\mathbf{A}^0}{4}$, we let $V^0 = \frac{e}{m_e}\mathbf{A}^0(\omega)\cdot\hat{\mathbf{p}}$, and the current becomes

$$
\begin{aligned}
\mathbf{J}(\omega,t) = \\
&-\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\Bigg[\mathrm{e}^{i(\omega_{n'''}-\omega_{n'})t}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n''}+\omega_{n'''})t_2} \\
&\qquad\times\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle\, f_{n''}\left(1-f_{n'''}\right)\left(1-f_{n'}\right)+c.c.\Bigg]dt_1dt_2 \\
&+\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n''''}\Bigg[\mathrm{e}^{i(\omega_{n''}-\omega_{n''''})t}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n''''}+\omega_{n'})t_2} \\
&\qquad\times\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''''|\,\hat{V}^0\,|n'\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n''''\rangle\,\left(1-f_{n'}\right)f_{n''}f_{n''''}+c.c.\Bigg]dt_1dt_2 \\
&+\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\Bigg[\mathrm{e}^{i(\omega_{n'''}-\omega_{n'})t}\mathrm{e}^{-i(\pm\omega-\omega_{n''}+\omega_{n'''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n'}+\omega_{n''})t_2} \\
&\qquad\times\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle\, f_{n'''}\left(1-f_{n''}\right)\left(1-f_{n'}\right)+c.c.\Bigg]dt_1dt_2 \\
&-\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n''''}\Bigg[\mathrm{e}^{i(\omega_{n''}-\omega_{n''''})t}\mathrm{e}^{-i(\pm\omega-\omega_{n''''}+\omega_{n'})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n'}+\omega_{n''})t_2} \\
&\qquad\times\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''''|\,\hat{V}^0\,|n'\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n''''\rangle\,\left(1-f_{n''}\right)f_{n'}f_{n''''}+c.c.\Bigg]dt_1dt_2
\end{aligned}
$$

The use of $\pm/\mp$ is meant to indicate that the two sign configurations are to be summed over. Relabeling indices to simplify

$$\begin{aligned}\mathbf{J}(\omega,t) = -\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\Bigg[&\mathrm{e}^{i(\omega_{n'''}-\omega_{n'})t}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n''}+\omega_{n'''})t_2}\\ &\times\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle\\ &\times\left[f_{n''}\left(1-f_{n'''}\right)\left(1-f_{n'}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right]+c.c.\Bigg]dt_1dt_2\\ +\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\Bigg[&\mathrm{e}^{i(\omega_{n''}-\omega_{n'''})t}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n'''}+\omega_{n'})t_2}\\ &\times\langle n'''|\,\hat{V}^0\,|n'\rangle\,\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n'''\rangle\\ &\times\left[f_{n''}\left(1-f_{n'}\right)\left(1-f_{n'''}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right]+c.c.\Bigg]dt_1dt_2\end{aligned}$$

It is now clear that each electron excitation is matched with a hole excitation moving in the opposite direction. However, in the above expression, the current appears to oscillate. The promised dc-like response is recovered only after performing the integration over all second-order processes.

The total current is

$$\begin{aligned}
\mathbf{J}(\omega) = &- \frac{e}{m_e}\frac{1}{4\hbar^2}\Bigg[\sum_{n',n'',n'''}\left[f_{n''}\left(1-f_{n'''}\right)\left(1-f_{n'}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right] \\
&\quad\times\left\langle n'\right|\hat{V}^0\left|n''\right\rangle\left\langle n''\right|\hat{V}^0\left|n'''\right\rangle\left\langle n'''\right|\hat{\mathbf{p}}\left|n'\right\rangle \\
&\quad\times \mathrm{e}^{-i(\omega_{n'}-\omega_{n'''})t}\int_{-\infty}^{t}\int_{-\infty}^{t_2}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n''}+\omega_{n'''})t_2}dt_1dt_2+c.c.\Bigg] \\
&+\frac{e}{m_e}\frac{1}{4\hbar^2}\Bigg[\sum_{n',n'',n'''}\left[f_{n''}\left(1-f_{n'}\right)\left(1-f_{n'''}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right] \\
&\quad\times\left\langle n'''\right|\hat{V}^0\left|n'\right\rangle\left\langle n'\right|\hat{V}^0\left|n''\right\rangle\left\langle n''\right|\hat{\mathbf{p}}\left|n'''\right\rangle \\
&\quad\times \mathrm{e}^{-i(\omega_{n'''}-\omega_{n''})t}\int_{-\infty}^{t}\int_{-\infty}^{t_2}\mathrm{e}^{-i(\pm\omega-\omega_{n'}+\omega_{n''})t_1}\mathrm{e}^{-i(\mp\omega-\omega_{n'''}+\omega_{n'})t_2}dt_1dt_2+c.c.\Bigg]
\end{aligned}$$

Letting $\hat{V}^0 \to \mathrm{e}^{\eta t}\hat{V}^0$ (i.e., a slowly-turned-on interaction), and performing the integration

$$\begin{aligned}
\mathbf{J}(\omega) = &\frac{e}{m_e}\frac{1}{4\hbar^2}\Bigg[\sum_{n',n'',n'''}\left[f_{n''}\left(1-f_{n'''}\right)\left(1-f_{n'}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right] \\
&\quad\times\left\langle n'\right|\hat{V}^0\left|n''\right\rangle\left\langle n''\right|\hat{V}^0\left|n'''\right\rangle\left\langle n'''\right|\hat{\mathbf{p}}\left|n'\right\rangle \\
&\quad\times \mathrm{e}^{2\eta t}\frac{1}{\left(\pm\omega-\omega_{n'}+\omega_{n''}\right)+i\eta}\frac{1}{\left(\omega_{n'''}-\omega_{n'}\right)+2i\eta}+c.c.\Bigg] \\
&-\frac{e}{m_e}\frac{1}{4\hbar^2}\Bigg[\sum_{n',n'',n'''}\left[f_{n''}\left(1-f_{n'}\right)\left(1-f_{n'''}\right)-\left(1-f_{n''}\right)f_{n'}f_{n'''}\right] \\
&\quad\times\left\langle n'''\right|\hat{V}^0\left|n'\right\rangle\left\langle n'\right|\hat{V}^0\left|n''\right\rangle\left\langle n''\right|\hat{\mathbf{p}}\left|n'''\right\rangle \\
&\quad\times \mathrm{e}^{2\eta t}\frac{1}{\left(\pm\omega-\omega_{n'}+\omega_{n''}\right)+i\eta}\frac{1}{\left(\omega_{n''}-\omega_{n'''}\right)+2i\eta}+c.c.\Bigg]
\end{aligned}$$

Each individual second-order process contributes both a constant phase factor and an evolving oscillation. Processes occurring at different times interfere with each other, resulting

in a net phase of zero and a nonzero factor proportional to the detuning frequency. To emphasize: the overall coherent state, and by extension the current, exists continuously for the duration of the illumination, and is formed by the temporal interference of individual interaction processes.

Manipulating indices

$$
\begin{aligned}
\mathbf{J}(\omega) =& \frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle \\
&\times \mathrm{e}^{2\eta t}\frac{1}{(\omega_{n'''}-\omega_{n'})+2i\eta}\frac{1}{(\pm\omega-\omega_{n'}+\omega_{n''})+i\eta} \\
&\times\Big[f_{n''}\,(1-f_{n'''})\,(1-f_{n'})-(1-f_{n''})\,f_{n'}\,f_{n'''} \\
&\qquad -f_{n'}\,(1-f_{n''})\,(1-f_{n'''})+(1-f_{n'})\,f_{n''}\,f_{n'''}\Big] \\
&-\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\langle n'''|\,\hat{V}^0\,|n'\rangle\,\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n'''\rangle \\
&\times \mathrm{e}^{2\eta t}\frac{1}{(\omega_{n''}-\omega_{n'''})+2i\eta}\frac{1}{(\pm\omega-\omega_{n'}+\omega_{n''})+i\eta} \\
&\times\Big[f_{n''}\,(1-f_{n'})\,(1-f_{n'''})-(1-f_{n''})\,f_{n'}\,f_{n'''} \\
&\qquad -f_{n'}\,(1-f_{n'''})\,(1-f_{n''})+(1-f_{n'})\,f_{n'''}\,f_{n''}\Big] \\
=& \frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle\,[f_{n''}-f_{n'}] \\
&\times \mathrm{e}^{2\eta t}\frac{1}{(\omega_{n'''}-\omega_{n'})+2i\eta}\frac{1}{(\pm\omega-\omega_{n'}+\omega_{n''})+i\eta} \\
&-\frac{e}{m_e}\frac{1}{4\hbar^2}\sum_{n',n'',n'''}\langle n'''|\,\hat{V}^0\,|n'\rangle\,\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n'''\rangle\,[f_{n''}-f_{n'}] \\
&\times \mathrm{e}^{2\eta t}\frac{1}{(\omega_{n''}-\omega_{n'''})+2i\eta}\frac{1}{(\pm\omega-\omega_{n'}+\omega_{n''})+i\eta}
\end{aligned}
$$

Taking the limit

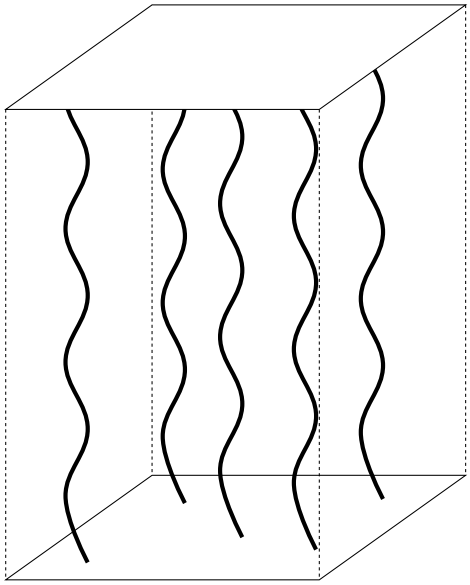
$$\begin{aligned}
\mathbf{J}(\omega) =& \lim_{\eta\to 0} \frac{e}{m_e}\frac{1}{4\hbar^2} \sum_{n',n'',n'''} \mathrm{e}^{-2\eta t} \left[f_{n'} - f_{n''}\right] \frac{1}{(\pm\omega - \omega_{n'} + \omega_{n''}) + i\eta} \\
&\times \left(\langle n'| \hat{V}^0 |n''\rangle \langle n''| \hat{V}^0 |n'''\rangle \langle n'''| \hat{\mathbf{p}} |n'\rangle \frac{1}{(\omega_{n'''} - \omega_{n'}) + 2i\eta} \right. \\
&\qquad \left. - \langle n'''| \hat{V}^0 |n'\rangle \langle n'| \hat{V}^0 |n''\rangle \langle n''| \hat{\mathbf{p}} |n'''\rangle \frac{1}{(\omega_{n''} - \omega_{n'''}) + 2i\eta} \right) \\
=& \frac{e}{m_e}\frac{1}{4\hbar^2} \sum_{n',n'',n'''} \left[f_{n'} - f_{n''}\right] \left[\mathcal{P}\frac{1}{\pm\omega - \omega_{n'} + \omega_{n''}} - i\pi\delta(\pm\omega - \omega_{n'} + \omega_{n''}) \right] \\
&\times \left(\langle n'| \hat{V}^0 |n''\rangle \langle n''| \hat{V}^0 |n'''\rangle \langle n'''| \hat{\mathbf{p}} |n'\rangle \left[\mathcal{P}\frac{1}{\omega_{n'''} - \omega_{n'}} - i\pi\delta(\omega_{n'''} - \omega_{n'}) \right] \right. \\
&\qquad \left. - \langle n'''| \hat{V}^0 |n'\rangle \langle n'| \hat{V}^0 |n''\rangle \langle n''| \hat{\mathbf{p}} |n'''\rangle \left[\mathcal{P}\frac{1}{\omega_{n''} - \omega_{n'''}} - i\pi\delta(\omega_{n''} - \omega_{n'''}) \right] \right)
\end{aligned}$$

If time reversal symmetry prevails, we may eliminate several terms. If, for the two terms of the summed-over expression, $\omega_{n'''} = \omega_{n'}$ or $\omega_{n''} = \omega_{n'''}$, then the corresponding term contains diagonal elements of the momentum matrix, whose contributions must vanish under time-reversal. If both the initial excitation and overall process are off-resonant, then the contribution to the current would come from the real parts of the products of momentum matrix elements; only the imaginary parts can be nonzero under time-reversal. However, a resonant initial excitation but off-resonant overall process provides a contribution due to

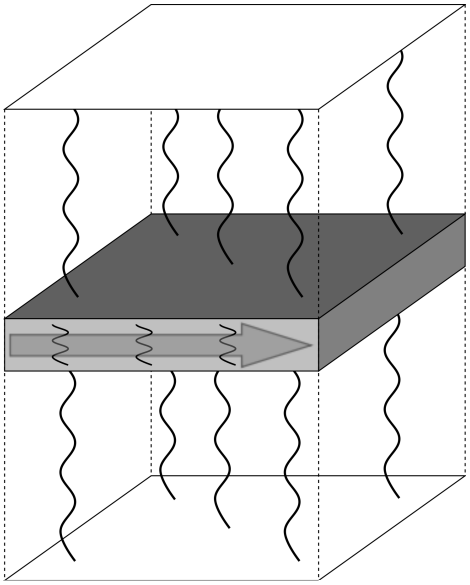
these imaginary parts, and may be nonzero. This leaves

$$\mathbf{J}(\omega) = \frac{e}{m_e}\frac{i\pi}{4\hbar^2}\sum_{n',n'',n'''}\left[f_{n'} - f_{n''}\right]\delta(\pm\omega - \omega_{n'} + \omega_{n''}) \times \left(\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{V}^0\,|n'''\rangle\,\langle n'''|\,\hat{\mathbf{p}}\,|n'\rangle\,\mathcal{P}\frac{1}{\omega_{n'''} - \omega_{n'}} - \langle n'''|\,\hat{V}^0\,|n'\rangle\,\langle n'|\,\hat{V}^0\,|n''\rangle\,\langle n''|\,\hat{\mathbf{p}}\,|n'''\rangle\,\mathcal{P}\frac{1}{\omega_{n''} - \omega_{n'''}} \right)$$

Thus, the shift current results from the modification of the superposition between excited and ground states that results from resonant linear absorption. The system proceeds to a virtual state via the second, off-resonant interaction; this state carries the current. Upon (adiabatic) removal of illumination, the system is constrained to relax to either of the states coupled by the initial excitation, in the same way that an off-resonant linear absorption must relax to the ground state.



(a)



(b)

Figure 4.1: In (a) a perfect cavity with periodic boundary conditions there is a steady-state population of photons associated with a standing EM wave of the kind assumed for the present treatment. If (b) a slab of materials is introduced, some portion of the photons will become entangled with electronic transitions, reducing the population of photons in the cavity and perturbing the electrons of the material from their ground-state-Hamiltonian eigenstates. If the material has appropriately-broken inversion symmetry, then these perturbed electronic states will have a net velocity, creating a current. In the absence of scattering and in the low-intensity limit, this is a constant current, and the coherent states persist indefinitely. In the event of scattering, the electron may return to the ground state, or occupy a thermalized excited state. In the former case, the electron will rapidly be re-excited into a coherent, current-carrying state. In the latter case, the return to the current-carrying state is delayed by the lifetime of the excited state; the current is not limited by scattering per se, but by the proportion of carriers that are in thermalized excited states.

While the illumination persists, the electron and hole are coherently coupled with the light, forming a quasiparticle with fixed net momentum. The current does not evolve from a field accelerating excited electrons through a medium; it is instead an intrinsic property of this quasiparticle. To see this, let us consider a situation very different from the context that photovoltaics are found in: the idealized case of an optical cavity containing a shift-current-active material without dissipation. Without the material, the electromagnetic field is described by a single-mode, standing – rather than traveling – wave, the condition under which our expression is actually derived (Fig. 4.1(a)). With the material introduced, the population of the mode in the cavity is reduced, with a portion of the photons bound into quasiparticles with the electrons and holes of the material (Fig. 4.1(b)). This is a steady state of the system, with the quasiparticles carrying a constant current. If scattering is introduced and a particular carrier is thermalized, the electron halts, and either returns to the time-independent ground state, with the photon returning to the cavity, or absorbs the photon and comes to rest in an excited state; this latter case provides the means for dissipation. If we assume that those returning to the ground state very rapidly reform quasiparticles, then the current is a function of the intensity and thermalized carrier population. This explains why the shift current expression provides a good description of current magnitude, despite including no scattering terms: the current is only mitigated to the extent that there are free excited carriers, the number of which are usually small compared to total number of electrons.

Since the current is due to the component of the wavefunction arising from the perturbation of the initial excited state by the field, it is useful to express the current as a property of this initial excitation. First, we shall explicitly adopt the Bloch formalism, so that our

states are now a function of $\mathbf{k}$. Our expression becomes

$$\begin{aligned}\mathbf{J}(\omega) =& \frac{e}{m_e}\frac{i\pi}{4\hbar^2}\sum_{n',n'',n'''}\int d^3\mathbf{k}'\int d^3\mathbf{k}''\int d^3\mathbf{k}''' \\ &\times\left[f_{n'}(\mathbf{k}')-f_{n''}(\mathbf{k}'')\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\ &\times\Bigg[\left\langle n'\mathbf{k}'\right|\hat{V}^0\left|n''\mathbf{k}''\right\rangle\left\langle n''\mathbf{k}''\right|\hat{V}^0\left|n'''\mathbf{k}'''\right\rangle\left\langle n'''\mathbf{k}'''\right|\hat{\mathbf{p}}\left|n'\mathbf{k}'\right\rangle\mathcal{P}\frac{1}{\omega_{n'''}(\mathbf{k}''')-\omega_{n'}(\mathbf{k}')} \\ &\quad-\left\langle n'''\mathbf{k}'''\right|\hat{V}^0\left|n'\mathbf{k}'\right\rangle\left\langle n'\mathbf{k}'\right|\hat{V}^0\left|n''\mathbf{k}''\right\rangle\left\langle n''\mathbf{k}''\right|\hat{\mathbf{p}}\left|n'''\mathbf{k}'''\right\rangle\mathcal{P}\frac{1}{\omega_{n''}(\mathbf{k}'')-\omega_{n'''}(\mathbf{k}''')}\Bigg]\end{aligned} \tag{4.1}$$

Noting the appearance of the expression

$$\left\langle n'\mathbf{k}'\right|\hat{\mathbf{p}}\left|n\mathbf{k}\right\rangle\mathcal{P}\frac{1}{\omega_{n'}(\mathbf{k}')-\omega_n(\mathbf{k})}$$

and that one of the indices appears nowhere else in the term, we find we may simplify the expression using Eq. (3.4). Since we wish to integrate over the principle part of the expression on the LHS of Eq. (3.4), which results in the exclusion of the pole occurring at $n=n'$ and $\mathbf{k}=\mathbf{k}'$, we must explicitly exclude the corresponding term from our integral on the RHS.

$$\begin{aligned}\sum_{n'}\int d^3\mathbf{k}'\ \left|n'\mathbf{k}'\right\rangle\mathcal{P}\frac{\left\langle n'\mathbf{k}'\right|\hat{\mathbf{p}}\left|n\mathbf{k}\right\rangle}{m_e i\left(\omega_{n'}(\mathbf{k}')-\omega_n(\mathbf{k})\right)} &= \sum_{n'}\int d^3\mathbf{k}'\ \left|n'\mathbf{k}'\right\rangle\left\langle n'\mathbf{k}'\right|\hat{\mathbf{r}}\left|n\mathbf{k}\right\rangle-\lim_{\epsilon\to\mathbf{0}}\int_{\mathbf{k}-\epsilon}^{\mathbf{k}+\epsilon}d^3\mathbf{k}'\ \left|n\mathbf{k}'\right\rangle\left\langle n\mathbf{k}'\right|\hat{\mathbf{r}}\left|n\mathbf{k}\right\rangle \\ \sum_{n'}\int d^3\mathbf{k}'\ \left|n'\mathbf{k}'\right\rangle\mathcal{P}\frac{\left\langle n'\mathbf{k}'\right|\hat{\mathbf{p}}\left|n\mathbf{k}\right\rangle}{m_e i\left(\omega_{n'}(\mathbf{k}')-\omega_n(\mathbf{k})\right)} &= \hat{\mathbf{r}}\left|n\mathbf{k}\right\rangle-\lim_{\epsilon\to\mathbf{0}}\int_{\mathbf{k}-\epsilon}^{\mathbf{k}+\epsilon}d^3\mathbf{k}'\ \left\langle n\mathbf{k}'\right|\hat{\mathbf{r}}\left|n\mathbf{k}\right\rangle\left|n\mathbf{k}'\right\rangle\end{aligned} \tag{4.2}$$

Inserting this into Eq. (4.1) we obtain

$$\begin{aligned}
\mathbf{J}(\omega) = -\Bigg\{ \frac{e\pi}{4\hbar^2} \sum_{n',n''} \int d^3\mathbf{k}' \int d^3\mathbf{k}'' \\
&\times \left[f_{n'}(\mathbf{k}') - f_{n''}(\mathbf{k}'') \right] \delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}'')) \\
&\times \Big[\langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \langle n''\mathbf{k}''| \hat{V}^0 \hat{\mathbf{r}} |n'\mathbf{k}'\rangle \\
&\qquad - \langle n''\mathbf{k}''| \hat{\mathbf{r}} \hat{V}^0 |n'\mathbf{k}'\rangle \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \Big] \Bigg\} \\
+\Bigg\{ \frac{e\pi}{4\hbar^2} \sum_{n',n''} \int d^3\mathbf{k}' \int d^3\mathbf{k}'' \\
&\times \left[f_{n'}(\mathbf{k}') - f_{n''}(\mathbf{k}'') \right] \delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}'')) \\
&\times \lim_{\epsilon \to \mathbf{0}} \Bigg[\int_{\mathbf{k}'-\epsilon}^{\mathbf{k}'+\epsilon} d^3\mathbf{k}''' \ \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \langle n''\mathbf{k}''| \hat{V}^0 |n'\mathbf{k}'''\rangle \langle n'\mathbf{k}'''| \hat{\mathbf{r}} |n'\mathbf{k}'\rangle \\
&\qquad - \int_{\mathbf{k}''-\epsilon}^{\mathbf{k}''+\epsilon} d^3\mathbf{k}''' \ \langle n''\mathbf{k}'''| \hat{V}^0 |n'\mathbf{k}'\rangle \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \langle n''\mathbf{k}''| \hat{\mathbf{r}} |n''\mathbf{k}'''\rangle \Bigg] \Bigg\}
\end{aligned}$$

Since

$$\begin{aligned}
&\langle n''\mathbf{k}''| \hat{\mathbf{r}} \hat{V}^0 |n'\mathbf{k}'\rangle \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle - \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \langle n''\mathbf{k}''| \hat{V}^0 \hat{\mathbf{r}} |n'\mathbf{k}'\rangle \\
&\qquad = \langle n''\mathbf{k}''| \left[\hat{\mathbf{r}}, \hat{V}^0 \right] |n'\mathbf{k}'\rangle \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \\
&\qquad = \langle n''\mathbf{k}''| \mathbf{A}^0 \cdot [\hat{\mathbf{r}}, \hat{\mathbf{p}}] |n'\mathbf{k}'\rangle \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle \\
&\qquad = i\hbar \mathbf{A}^0 \delta_{n'n''} \delta(\mathbf{k}' - \mathbf{k}'') \langle n'\mathbf{k}'| \hat{V}^0 |n''\mathbf{k}''\rangle
\end{aligned}$$

the first group of terms is zero for allowed transitions.

Noting that $\langle n\mathbf{k}|\,\hat{\mathbf{p}}\,|n'\mathbf{k}'\rangle = 0$ when $\mathbf{k} \neq \mathbf{k}'$, we find that

$$\begin{aligned}\mathbf{J}(\omega) = \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k}) - f_{n''}(\mathbf{k})\right]\delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}''))\\ \times\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\lim_{\epsilon\to\mathbf{0}}\left[\int_{\mathbf{k}-\epsilon}^{\mathbf{k}+\epsilon} d^3\mathbf{k}'''\;\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}'''\rangle\,\langle n'\mathbf{k}'''|\,\hat{\mathbf{r}}\,|n'\mathbf{k}\rangle\right.\\ \left. - \int_{\mathbf{k}-\epsilon}^{\mathbf{k}+\epsilon} d^3\mathbf{k}'''\;\langle n''\mathbf{k}'''|\,\hat{V}^0\,|n'\mathbf{k}\rangle\,\langle n''\mathbf{k}|\,\hat{\mathbf{r}}\,|n''\mathbf{k}'''\rangle\right]\end{aligned} \tag{4.3}$$

Naively taking the limit would result in position expectation values of the Bloch states. The appropriate action is to use Eqs. (3.2) and (3.3)

Inserting these into Eq. (4.3), with the momentum elements serving as functions g, we obtain

$$\begin{aligned}\mathbf{J}(\omega) = \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k}) - f_{n''}(\mathbf{k})\right]\delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}''))\\ \times\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\left[\left(-i\nabla_{\mathbf{k}'''=\mathbf{k}}\,\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}'''\rangle + \langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\,\chi_{n'}(\mathbf{k})\right)\right.\\ \left. - \left(i\nabla_{\mathbf{k}'''=\mathbf{k}}\,\langle n''\mathbf{k}'''|\,\hat{V}^0\,|n'\mathbf{k}\rangle + \langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\,\chi_{n''}(\mathbf{k})\right)\right]\end{aligned}$$

Setting $\phi_{n''n'}(\mathbf{k})$ to be the phase of $\langle n''\mathbf{k}|\,\hat{\mathbf{p}}\,|n'\mathbf{k}\rangle$ (and consequently $\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle$)

$$
\begin{aligned}
\mathbf{J}(\omega) =& \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\Big(-\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\, i\nabla_{\mathbf{k}}\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle \\
&\qquad -\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\left[\chi_{n''}(\mathbf{k})-\chi_{n'}(\mathbf{k})\right]\Big) \\
=& \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\Big(-\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\, i\nabla_{\mathbf{k}}\mathrm{e}^{i\phi_{n''n'}(\mathbf{k})}\left|\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\right| \\
&\qquad -\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\left[\chi_{n''}(\mathbf{k})-\chi_{n'}(\mathbf{k})\right]\Big) \\
=& \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\Big(\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\Big[\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\,\nabla_{\mathbf{k}}\phi_{n''n'}(\mathbf{k}) \\
&\qquad -\mathrm{e}^{i\phi_{n''n'}(\mathbf{k})}i\nabla_{\mathbf{k}}\left|\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\right|\Big] \\
&\qquad -\langle n'\mathbf{k}|\,\hat{V}^0\,|n''\mathbf{k}\rangle\langle n''\mathbf{k}|\,\hat{V}^0\,|n'\mathbf{k}\rangle\left[\chi_{n''}(\mathbf{k})-\chi_{n'}(\mathbf{k})\right]\Big)
\end{aligned}
$$

Dropping the imaginary parts yields

$$
\begin{aligned}
\mathbf{J}(\omega) = & \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\;\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\langle n'\mathbf{k}|\,V^0\cdot\hat{\mathbf{p}}\,|n''\mathbf{k}\rangle\langle n''\mathbf{k}|\,V^0\cdot\hat{\mathbf{p}}\,|n'\mathbf{k}\rangle\left[\nabla_{\mathbf{k}}\phi_{n''n'}(\mathbf{k})-\chi_{n''}(\mathbf{k})+\chi_{n'}(\mathbf{k})\right]
\end{aligned}
$$

or, written as a response tensor, where we have again used the relation between the vector potential and electric field,

$$
\begin{aligned}
J_q(\omega) &= \sigma_{rsq}(\omega) E_r^0(\omega) E_s^0(\omega) \\
\sigma_{rsq}(\omega) &= e\pi \left(\frac{e}{2m_e \hbar \omega} \right)^2 \sum_{n',n''} \int d^3\mathbf{k} \, \left[f_{n'}(\mathbf{k}) - f_{n''}(\mathbf{k}) \right] \delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}'')) \\
&\quad \times \langle n'\mathbf{k}| \, \hat{p}_r \, |n''\mathbf{k}\rangle \, \langle n''\mathbf{k}| \, \hat{p}_s \, |n'\mathbf{k}\rangle \left[\{\nabla_{\mathbf{k}} \phi_{n''n'}(\mathbf{k})\}_q - \chi_{n'';q}(\mathbf{k}) + \chi_{n';q}(\mathbf{k}) \right]
\end{aligned}
\tag{4.4}
$$

We now have an expression that gives the current in terms of the states composing the superposed state of the initial excitation. The sum over current contributions from rotation of this linearly excited state into other states by the second interaction that appears in Eq. (4.3) has been shown equivalent to

$$
\mathcal{R}_{n'',n'}(\mathbf{k}) = \nabla_{\mathbf{k}} \phi_{n''n'}(\mathbf{k}) - \chi_{n''}(\mathbf{k}) + \chi_{n'}(\mathbf{k}) \tag{4.5}
$$

and now appears multiplied by the strength of linear absorption. This has been called the shift vector [13] and has units of distance. That it is a distance should not be taken to be physically significant. This arises since the current contributions of the sum it represents are weighted by the inverse of their detuning frequency; this weighting factor with units of time applies physically to the coefficient of the wavefunction components. More physically, one can consider the net velocity acquired by the excited component of the carrier's wavefunction, which may be loosely identified as the carrier velocity. The magnitude of

this component is given by the resonant part of the first-order term

$$\begin{aligned}
\left|\Psi^{(1)}\right\rangle &= \left|\Psi^{(0)}\right\rangle + \int_{-\infty}^{t} \left(i\frac{H^I(t_1)}{\hbar}dt_1\right)\left|\Psi^{(0)}\right\rangle \\
&= \left|\Psi^{(0)}\right\rangle + \int_{-\infty}^{t} \frac{i}{2\hbar}\sum_{n',n''} \left\langle n''\right| \frac{e}{m_e}\mathbf{n}_A\cdot\hat{\mathbf{p}}\left|n'\right\rangle \hat{c}^\dagger_{n''}\hat{c}_{n'} \\
&\qquad\qquad \times \left[\mathrm{e}^{-i(-\omega-\omega_{n''}+\omega_{n'})t}\hat{a}^\dagger(\omega) + \mathrm{e}^{-i(\omega-\omega_{n''}+\omega_{n'})t}\hat{a}(\omega)\right]\left|\Psi^{(0)}\right\rangle \\
&= \left|\Psi^{(0)}\right\rangle + \frac{i\pi}{2\hbar}\sum_{n',n''}\delta(\omega-\omega_{n''}+\omega_{n'})\left\langle n''\right| \frac{e}{m_e}\mathbf{n}_A\cdot\hat{\mathbf{p}}\left|n'\right\rangle \hat{c}^\dagger_{n''}\hat{c}_{n'}\hat{a}(\omega)\left|\Psi^{(0)}\right\rangle
\end{aligned}$$

For periodic systems this becomes

$$\begin{aligned}
\left|\Psi^{(1)}\right\rangle &= \left|\Psi^{(0)}\right\rangle + \frac{i\pi}{2\hbar}\sum_{n',n''}\int d^3\mathbf{k}\,\delta(\omega-\omega_{n''}(\mathbf{k})+\omega_{n'}(\mathbf{k})) \\
&\qquad\qquad \times \left\langle n''\mathbf{k}\right| \frac{e}{m_e}\mathbf{n}_A\cdot\hat{\mathbf{p}}\left|n'\mathbf{k}\right\rangle \hat{c}^\dagger_{n''}(\mathbf{k})\hat{c}_{n'}(\mathbf{k})\hat{a}(\omega)\left|\Psi^{(0)}\right\rangle \\
&= \left|\Psi^{(0)}\right\rangle + \frac{i\pi}{2}\sum_{n',n''}\int d\left|E_{n''}(\mathbf{k})-E_{n'}(\mathbf{k})\right| \\
&\qquad\qquad \times \delta(E_\omega - E_{n''}(\mathbf{k})+E_{n'}(\mathbf{k}))D_{n',n''}(E_{n''}(\mathbf{k})-E_{n'}(\mathbf{k})) \\
&\qquad\qquad \times \left\langle n''\mathbf{k}\right| \frac{e}{m_e}\mathbf{n}_A\cdot\hat{\mathbf{p}}\left|n'\mathbf{k}\right\rangle \hat{c}^\dagger_{n''}(\mathbf{k})\hat{c}_{n'}(\mathbf{k})\hat{a}(\omega)\left|\Psi^{(0)}\right\rangle \\
&= \left|\Psi^{(0)}\right\rangle + \frac{i\pi}{2}D_{n',n''}(E_\omega)\left\langle n''\mathbf{k}\right| \frac{e}{m_e}\mathbf{n}_A\cdot\hat{\mathbf{p}}\left|n'\mathbf{k}\right\rangle \hat{c}^\dagger_{n''}(\mathbf{k})\hat{c}_{n'}(\mathbf{k})\hat{a}(\omega)\left|\Psi^{(0)}\right\rangle
\end{aligned}$$

where $D_{n',n''}(E)$ is the density of transitions of energy E between bands n' and n''. In the last term, $\mathbf{k}$ is now the k-point where $E_{n''}(\mathbf{k}) - E_{n'}(\mathbf{k}) = E_\omega$. The magnitude of the first-order perturbation is then

$$\frac{\pi^2}{4}D_{n',n''}(E_\omega)^2\left\langle n'\mathbf{k}\right| V^0\left|n''\mathbf{k}\right\rangle\left\langle n''\mathbf{k}\right| V^0\left|n'\mathbf{k}\right\rangle$$

and the velocity, taking the integral for the same transition, is

$$\begin{aligned}\mathbf{v}_{n',n''}(\mathbf{k}) =& \frac{1}{e}\frac{\mathbf{J}(\omega,\mathbf{k})}{\frac{\pi^2}{4}D_{n',n''}(E_\omega)^2\left|\langle n''\mathbf{k}|\,V^0\,|n'\mathbf{k}\rangle\right|^2} \\ \mathbf{v}_{n',n''}(\mathbf{k}) =& \frac{\mathcal{R}_{n'',n'}(\mathbf{k})}{D_{n',n''}(E_\omega)\pi\hbar}\end{aligned} \tag{4.6}$$

We can see that the velocity is directly related to the shift vector, with a factor determined by the density of transitions.

4.2 Observables

In order to compare our calculations to experiment we must provide the same quantities that will be measured for real systems. These are primarily the current density as a function of light intensity (rather than field as produced by our derivation), and the Glass coefficient, which obtains for macroscopically thick samples.

To convert the current density to a response to intensity, we use the relation for energy density of an EM wave:

$$u(\omega) = \frac{\epsilon_0\epsilon^{\Re}}{2}\left|\mathbf{E}(\omega)\right|^2$$

Since the energy flux is depends on the speed of light in the sample

$$I(\omega) = u(\omega)\frac{c}{n} = \frac{\epsilon_0\epsilon^{\Re}c}{2n}\left|\mathbf{E}(\omega)\right|^2$$

where n is refractive index of the material in question.

The shift current response tensor becomes

$$\bar{\sigma}_{rsq}(\omega) = \frac{e\pi n}{2\epsilon_0 \epsilon^{\Re} c}\left(\frac{e}{m_e \hbar\omega}\right)^2 \sum_{n',n''} \int d^3\mathbf{k}\ \left[f_{n'}(\mathbf{k}) - f_{n''}(\mathbf{k})\right] \delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}''))$$
$$\times \langle n'\mathbf{k}|\, \hat{p}_r\, |n''\mathbf{k}\rangle \langle n''\mathbf{k}|\, \hat{p}_s\, |n'\mathbf{k}\rangle \left[\{\nabla_{\mathbf{k}} \phi_{n''n'}(\mathbf{k})\}_q - \chi_{q,n''}(\mathbf{k}) + \chi_{q,n'}(\mathbf{k})\right]$$

We note that this expression requires some care when $r \neq s$. For convenience we may write a transition intensity as

$$\mathcal{I}_{n'',n';rs}(\mathbf{k},\omega) = \pi \frac{n}{2\epsilon_0 \epsilon^{\Re} c}\left(\frac{e}{m_e \hbar\omega}\right)^2 \sum_{n',n''} \int d^3\mathbf{k}\ \langle n'\mathbf{k}|\, \hat{p}_r\, |n''\mathbf{k}\rangle \langle n''\mathbf{k}|\, \hat{p}_s\, |n'\mathbf{k}\rangle$$
$$\times \left[f_{n'}(\mathbf{k}) - f_{n''}(\mathbf{k})\right] \delta(\pm\omega - \omega_{n'}(\mathbf{k}') + \omega_{n''}(\mathbf{k}''))$$

so that

$$\bar{\sigma}_{rsq}(\omega) = e \sum_{n',n''} \int d^3\mathbf{k}\, \mathcal{I}_{n'',n';rs}(\mathbf{k},\omega) \mathcal{R}_{n'',n';q}(\mathbf{k})$$

Determining the total current in a sample is complicated by the attenuation of incident illumination as it travels through the material. In the limit of a thick sample that will completely absorb the illumination, the current can be obtained from the Glass coefficient [10]

$$J_q(\omega) = \frac{\bar{\sigma}_{rrq}(\omega)}{\alpha_{rr(\omega)}} I_r(\omega) \mathcal{W} = G_{rrq}(\omega) I_r(\omega) \mathcal{W} \tag{4.7}$$

where α is the absorption coefficient, and $\mathcal{W}$ is the sample width. Thus, the current density tensor and Glass coefficient describe the response in the regimes of infinitely thin and infinitely thick samples, respectively.

However, the Glass coefficient provides additional information about the response. In

the limit where $\epsilon^{\Im} << \epsilon^{\Re}$,

$$\alpha \approx \frac{\omega}{cn}\epsilon^{\Im} = 2\hbar\omega \sum_{n',n''} \int d\mathbf{k}\, \mathcal{I}_{n',n''}(\mathbf{k},\omega)$$

and the Glass coefficient becomes

$$G_{rrq}(\omega) = \frac{e}{2\hbar\omega} \frac{\sum_{n,n'} \int d\mathbf{k}\, \mathcal{I}_{n,n';rr}(\mathbf{k},\omega)\mathcal{R}_{n,n';q}(\mathbf{k})}{\sum_{nn'} \int d\mathbf{k}\, \mathcal{I}_{n,n';rr}(\mathbf{k},\omega)}$$

The Glass coefficient is therefore closely related to the weighted average shift vector, and allows us to estimate the contribution of both terms in the shift current expression.

4.3 Numerical Implementation

In the case of a degeneracy, the second-order resonances apply to degenerate blocks of states. Rather than the matrix elements that appear in (4.1), we seek to manipulate matrix blocks such that our expression is invariant under rotations of degenerate subspaces. Suppose that n and n' now refer to sets of degenerate bands, which are indexed by α, β, etc. Then we can write a block of an operator $\hat{B}$ as

$$\hat{B}_{nn'}(\mathbf{k},\mathbf{k}') = \begin{bmatrix} \langle n_\alpha(\mathbf{k})| \hat{B} |n'_\alpha(\mathbf{k}')\rangle & \langle n_\alpha(\mathbf{k})| \hat{B} \left|n'_\beta(\mathbf{k}')\right\rangle & \cdots \\ \langle n_\beta(\mathbf{k})| \hat{B} |n'_\alpha(\mathbf{k}')\rangle & \langle n_\beta(\mathbf{k})| \hat{B} \left|n'_\beta(\mathbf{k}')\right\rangle & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

and a block of the overlap matrix as

$$\hat{S}_n(\mathbf{k},\mathbf{k}') = \begin{bmatrix} \left\langle n_\alpha\mathbf{k}\middle|n_\alpha(\mathbf{k}')\right\rangle & \left\langle n_\alpha\mathbf{k}\middle|n_\beta(\mathbf{k}')\right\rangle & \cdots \\ \left\langle n_\beta\mathbf{k}\middle|n_\alpha(\mathbf{k}')\right\rangle & \left\langle n_\beta\mathbf{k}\middle|n_\beta(\mathbf{k}')\right\rangle & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

with this, Eq. (4.1) becomes

$$\mathbf{J} = \frac{e}{m_e}\frac{i\pi}{4\hbar^2}\sum_{n'''\neq n'',n'}\int d^3\mathbf{k}'\int d^3\mathbf{k}''\int d^3\mathbf{k}'''$$
$$\times\left(\mathrm{Tr}\left\{\hat{V}^0_{n'n''}(\mathbf{k}',\mathbf{k}'')\hat{V}^0_{n''n'''}(\mathbf{k}'',\mathbf{k}''')\hat{\mathbf{p}}_{n'''n'}(\mathbf{k}''',\mathbf{k}')\right\}\frac{1}{\omega_{n'''}(\mathbf{k}''')-\omega_{n'}(\mathbf{k}')}\right.$$
$$\left.-\mathrm{Tr}\left\{\hat{V}^0_{n'''n'}(\mathbf{k}''',\mathbf{k}')\hat{V}^0_{n'n''}(\mathbf{k}',\mathbf{k}'')\hat{\mathbf{p}}_{n''n'''}(\mathbf{k}'',\mathbf{k}''')\right\}\frac{1}{\omega_{n''}(\mathbf{k}'')-\omega_{n'''}(\mathbf{k}''')}\right)$$
$$\times\left[f_{n'}(\mathbf{k}')-f_{n''}(\mathbf{k}'')\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}''))$$

Eq. (4.2) becomes

$$\sum_{n'}\int d^3\mathbf{k}'\ |n'\mathbf{k}'\rangle\,\mathcal{P}\frac{\hat{\mathbf{p}}_{n'n}(\mathbf{k}',\mathbf{k})}{m_e i\left(\omega_{n'}(\mathbf{k}')-\omega_n(\mathbf{k})\right)} = \hat{\mathbf{r}}\,|n\mathbf{k}\rangle - \lim_{\epsilon\to\mathbf{0}}\int_{\mathbf{k}-\epsilon}^{\mathbf{k}+\epsilon} d^3\mathbf{k}'\,\hat{\mathbf{r}}_{nn}(\mathbf{k}',\mathbf{k})\,|n\mathbf{k}'\rangle$$

and Eqs. (3.2) and (3.3) are

$$\int d^3\mathbf{k}'\int d^3\mathbf{k}'\,g(\mathbf{k}')\hat{\mathbf{r}}_{nn}(\mathbf{k},\mathbf{k}') = i\nabla_{\mathbf{k}'=\mathbf{k}}\left[g(\mathbf{k}')S_n(\mathbf{k},\mathbf{k}')\right]$$
$$g(\mathbf{k}')\hat{\mathbf{r}}_{nn}(\mathbf{k}',\mathbf{k}) = -i\nabla_{\mathbf{k}'=\mathbf{k}}\left[g(\mathbf{k}')S_n(\mathbf{k}',\mathbf{k})\right]$$

With our generalized equations in hand we apply the same procedure as for the non-degenerate case. The final expression will be

$$\mathbf{J} = \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\,\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}''))$$
$$\times\nabla_{\mathbf{k}'''=\mathbf{k}}\Im\mathrm{Tr}\Big\{\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})S_{n'}(\mathbf{k},\mathbf{k}''')\hat{V}^0_{n'n''}(\mathbf{k}''',\mathbf{k})$$
$$+\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k}''')S_{n''}(\mathbf{k}''',\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\Big\}$$

In the continuous case we can evaluate this easily for any arbitrary selection of the degenerate states. In the discrete case, we are again faced with the problem of securing gauge invariance. The derivative of the logarithm of a (square) matrix can be taken by expanding the log as a Taylor series

$$\begin{aligned}
\mathrm{Tr}\left\{\hat{B}\frac{d\ln\hat{B}}{dk}\right\} &= \mathrm{Tr}\left\{\hat{B}\left(\frac{d\hat{B}}{dk}+\frac{1}{2}\frac{d\hat{B}}{dk}(\hat{B}-1)+\frac{1}{2}(\hat{B}-1)\frac{d\hat{B}}{dk}...\right)\right\} \\
&= \mathrm{Tr}\left\{\hat{B}\left(\frac{d\hat{B}}{dk}+\frac{d\hat{B}}{dk}(\hat{B}-1)...\right)\right\} \\
&= \mathrm{Tr}\left\{\hat{B}\frac{d\hat{B}}{dk}\frac{1}{\hat{B}}\right\} \\
&= \mathrm{Tr}\left\{\frac{d\hat{B}}{dk}\right\}
\end{aligned}$$

Thus, $\mathrm{Tr}\{\nabla B\} = \mathrm{Tr}\{B\nabla\ln B\}$ and

$$\begin{aligned}
\mathbf{J} =& \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\,\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\Im\mathrm{Tr}\Big\{\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\nabla_{\mathbf{k}'''=\mathbf{k}}\ln\left[\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})S_{n'}(\mathbf{k},\mathbf{k}''')\hat{V}^0_{n'n''}(\mathbf{k}''',\mathbf{k})\right] \\
&\qquad+\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\nabla_{\mathbf{k}'''=\mathbf{k}}\ln\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k}''')S_{n''}(\mathbf{k}''',\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\Big\} \\
\mathbf{J} =& \frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\,\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}'')) \\
&\times\lim_{\Delta\mathbf{k}\to0}\frac{1}{|\Delta\mathbf{k}|}\Im\mathrm{Tr}\Big\{\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\ln\Big[-\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k}) \\
&\qquad+\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})S_{n'}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\Big] \\
&\quad+\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\ln\Big[-\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k}) \\
&\qquad+\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})S_{n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\Big]\Big\}
\end{aligned}$$

Since $\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})$ will be real, this leaves

$$\begin{aligned}\mathbf{J} =&\frac{e\pi}{4\hbar^2}\sum_{n',n''}\int d^3\mathbf{k}\,\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}''))\\
&\times\lim_{\Delta\mathbf{k}\to 0}\frac{1}{|\Delta\mathbf{k}|}\Im\mathrm{Tr}\Big\{\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\\
&\qquad\times\ln\left[\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})S_{n'}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\right]\\
&\quad+\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\\
&\qquad\times\ln\left[\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})S_{n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\right]\Big\}\end{aligned}$$

The last term may appear to be problematic, as the log may be applied to a singular matrix. However, this only occurs when the matrix $\hat{V}^0_{n''n'}(\mathbf{k},\mathbf{k})\hat{V}^0_{n'n''}(\mathbf{k},\mathbf{k})$ has one or more zero eigenvalues; as this term multiplies the log, the contribution from the responsible states is zero. Such states can be removed during calculation. Dropping the limit, we obtain the response tensor

$$\begin{aligned}\sigma_{rsq} = e\pi\left(\frac{e}{m_e\hbar\omega}\right)^2&\sum_{n',n''}\sum_{\mathbf{k}}\mathcal{V}_{\mathbf{k}}\left[f_{n'}(\mathbf{k})-f_{n''}(\mathbf{k})\right]\delta(\pm\omega-\omega_{n'}(\mathbf{k}')+\omega_{n''}(\mathbf{k}''))\\
&\times\frac{1}{\Delta k_q}\Im\Big(\mathrm{Tr}\Big\{\hat{p}_{r;n''n'}(\mathbf{k},\mathbf{k})\hat{p}_{s;n'n''}(\mathbf{k},\mathbf{k})\\
&\qquad\ln\left[\hat{p}_{r;n''n'}(\mathbf{k},\mathbf{k})\hat{S}_{n'}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})\hat{p}_{s;n'n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\right]\Big\}\\
&\quad+\mathrm{Tr}\Big\{\hat{p}_{s;n'n''}(\mathbf{k},\mathbf{k})\hat{p}_{r;n''n'}(\mathbf{k},\mathbf{k})\\
&\qquad\ln\left[\hat{p}_{s;n'n''}(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})\hat{S}_{n''}(\mathbf{k}+\Delta\mathbf{k},\mathbf{k})\hat{p}_{r;n''n'}(\mathbf{k},\mathbf{k})\right]\Big\}\Big)\end{aligned}$$

where $\mathcal{V}_{\mathbf{k}}$ is a volume element of our discretized reciprocal space.

Chapter 5

Experimental Comparison and Validation

In order to test the hypothesis that BPVE is explained primarily by shift current, we compare calculated results with those in the literature. The material in this section appears in Refs. [54] and [55].

5.1 $BaTiO_3$

For bulk, single-crystal $BaTiO_3$, experimental spectra are available for energies near the band gap [56, 1]. The total current in a bulk crystal for light incident normal to the current direction can by computed using Eq. (4.7), which applies to samples of sufficient thickness to absorb all incident light. We obtained the light intensity and crystal dimensions from [1, 57], which were $\approx 0.35 - 0.6$ mW/cm^2 and 0.1-0.2 cm, respectively. In Fig. 5.1, the experimental current response from [1] is compared to the response computed using shift current theory. Despite the uncertainty in experimental parameters, the agreement is striking, in both magnitude and spectrum profile, for both tensor elements. This includes

the difference of sign between the majority of the transverse and longitudinal response, which is unusual [15], as well as the small positive region of the longitudinal response near the band edge.

We emphasize that these calculations not only reproduce the magnitude of response, but its idiosyncratic features as well. Because this theory reproduces all the salient features found in the experiments, this comparison provides strong evidence that shift current is the correct description of the bulk photovoltaic effect.

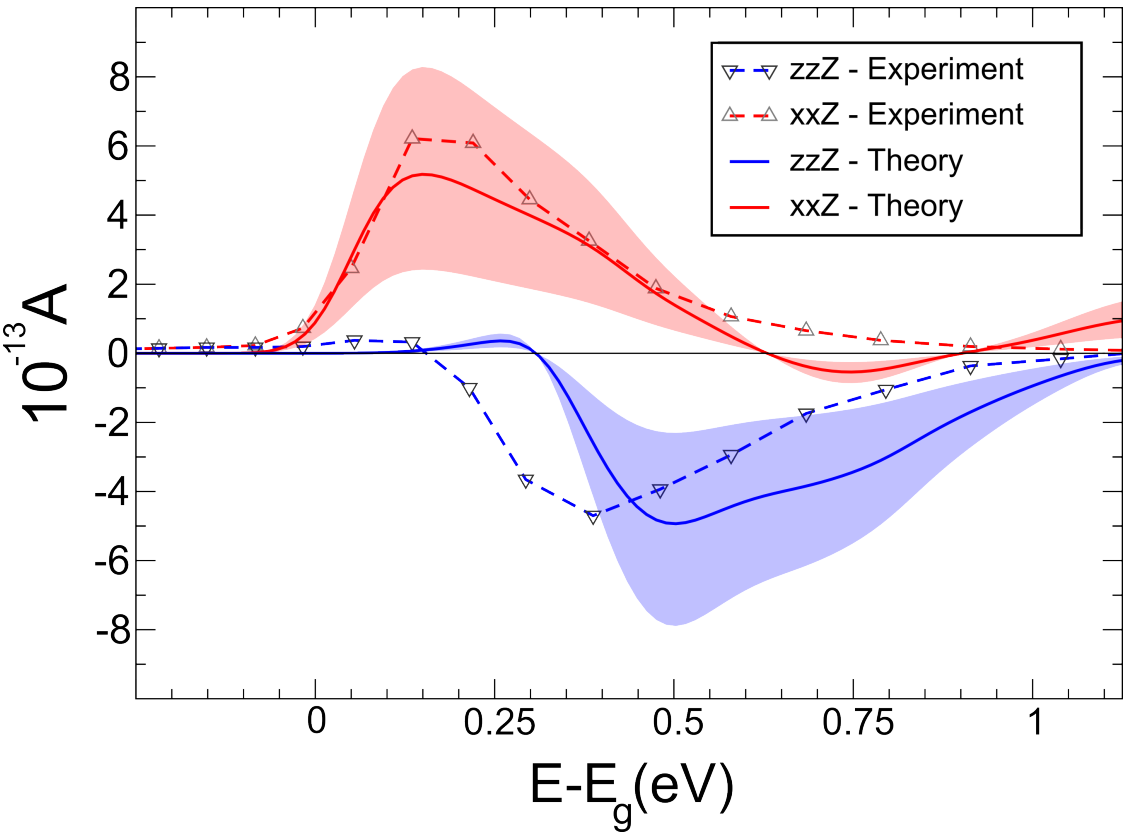

zzZ - Experiment
xxZ - Experiment
zzZ - Theory
xxZ - Theory
10^-13 A
E-E_g(eV)
8
6
4
2
0
-2
-4
-6
-8
0
0.25
0.5
0.75
1

Figure 5.1: For $BaTiO_3$, the experimental current [1] and computed current (this work), for transverse (xxZ) and longitudinal (zzZ) electric field orientation, as a function of energy above their respective bandgaps. The solid lines are calculated results for a choice of experimental parameters of 0.5 mW/cm^2 illumination intensity and 0.15 cm sample width. The shaded regions are bounded by the results using experimental parameters in the given range that provide the lowest and highest response.

5.2 $BiFeO_3$

$BiFeO_3$ (BFO) is a well-known multiferroic material that displays both large (90 $\mu C/cm^2$) polarization and weak ferromagnetism [58]. Furthermore, it possesses a relatively small band gap (2.74 eV) that falls in the visible spectrum [2]. This presents a major advantage over other ferroelectric oxides with large band gaps (typically $\approx$3.5 eV). There has been a flurry of interest in $BiFeO_3$ as a photovoltaic, highlighting it as a viable candidate for ferroelectric-based photovoltaic devices [26, 4, 5, 59, 3].

However, the origin of the observed photovoltaic properties is not quite clear. Recently, large photovoltages have been observed opposite the direction of material polarization in BFO [4, 5]. The photovoltage was attributed to the periodic domain walls acting to both separate and collect photoexcited carriers. The strong dependence of photovoltage on density of domain walls supports this hypothesis. However, while the bulk photovoltaic effect, as observed in crystals such as $LiNbO_3$, was discarded as a primary mechanism for the large photovoltages, the existence and impact of such an effect could not be ruled out or quantified. Indeed, other studies of BFO evaluated the photovoltaic response parallel [59] and perpendicular [3] to the ferroelectric polarization direction and observed substantial photocurrent generation, strongly suggesting that this effect should be significant even in samples with domain walls. The photocurrent in BFO has also been reported as being generated uniformly throughout the sample, consistent with a bulk effect [60]. Understanding the contributions of various mechanisms involved in the response collinear with material polarization is difficult, as the experiments cannot separate the bulk photovoltaic effect from polarization dependent mechanisms like those in Refs. [4] and [5]. Presently, we calculate the shift current response for BFO, showing good agreement with the results from Ref. [3] and providing estimates of the photocurrents collinear with the material polarization. Using these results we are able to determine the impact of the bulk photovoltaic effect

on the system studied in Refs. [4] and [5], explaining its apparent absence.

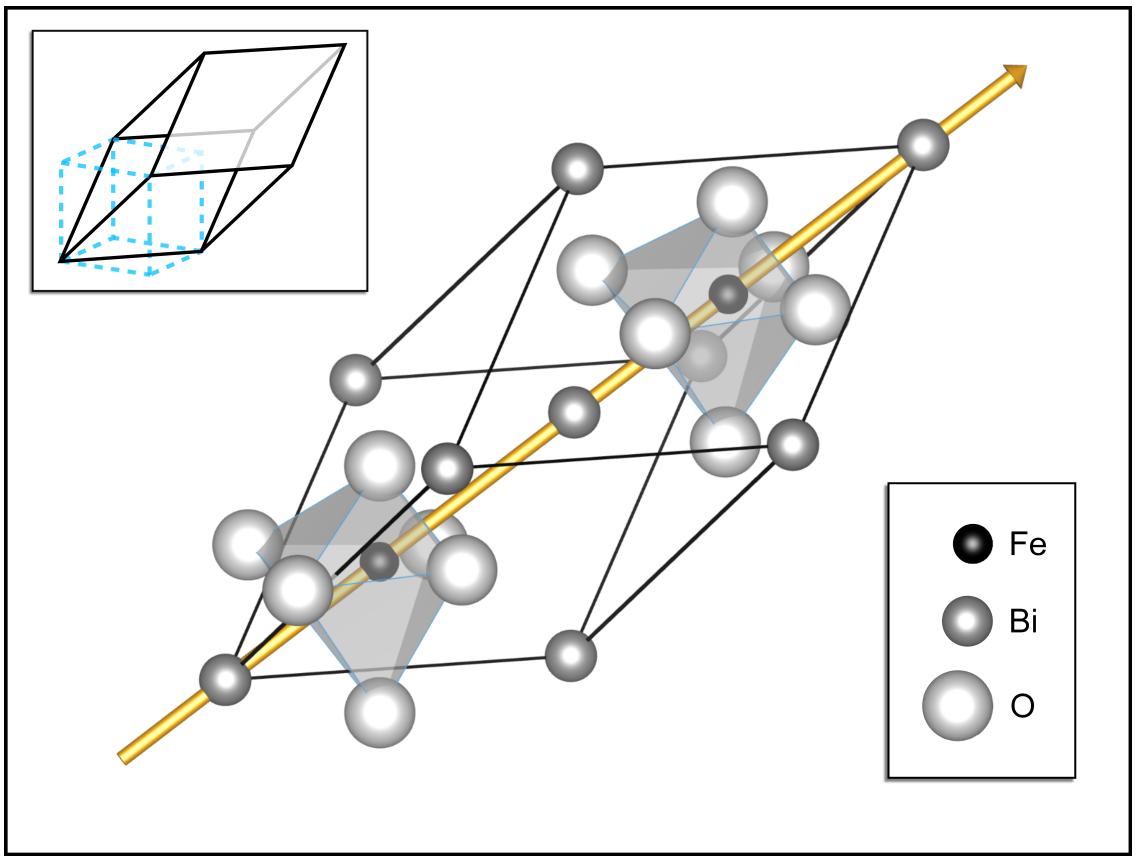

Fe
Bi
O

Figure 5.2: The rhombohedral unit cell of bismuth ferrite with the polarization direction denoted by the gold arrow. The two iron atoms are coordinated by distorted octahedral oxygen cages, rotated in opposite directions. The structure is only slightly distorted from the cubic perovskite structure, so the pseudo-cubic unit-cell, shown in the inset in relation to the rhombohedral unit cell, is often used to represent the structure.

BFO belongs to space group $R3c$. Shown in Fig. 5.2 is the primitive (10 atom) unit cell with the polarization direction along [0001] (z direction), as well as the relation to the pseudo-cubic unit cell sometimes used, for which the polarization is along the [111] direction. BFO is nearly a G-type anti-ferromagnetic (AFM) material; however, it is found that there is spin canting in the xy plane (perpendicular to the polarization direction) [61]. Because of the small magnitude ($\approx$0.1 μB per unit cell), this spin canting is ignored in the present work, and only spin-polarized calculations are performed. The experimental geometry is used for calculations throughout.

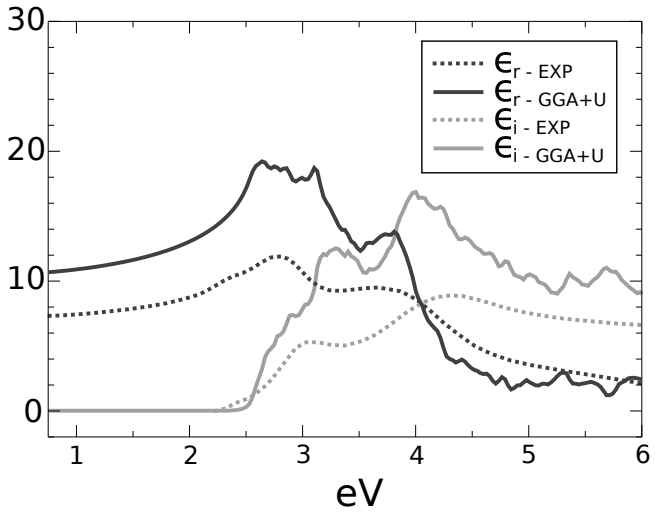
In the coordinates of this geometry, a third rank tensor, such as the shift current response tensor, must have the form

$$\sigma = \begin{bmatrix} 0 & 0 & 0 & 0 & \sigma_{yzY} & -\sigma_{yyY} \\ -\sigma_{yyY} & \sigma_{yyY} & 0 & \sigma_{yzY} & 0 & 0 \\ \sigma_{xxZ} & \sigma_{xxZ} & \sigma_{zzZ} & 0 & 0 & 0 \end{bmatrix} \tag{5.1}$$

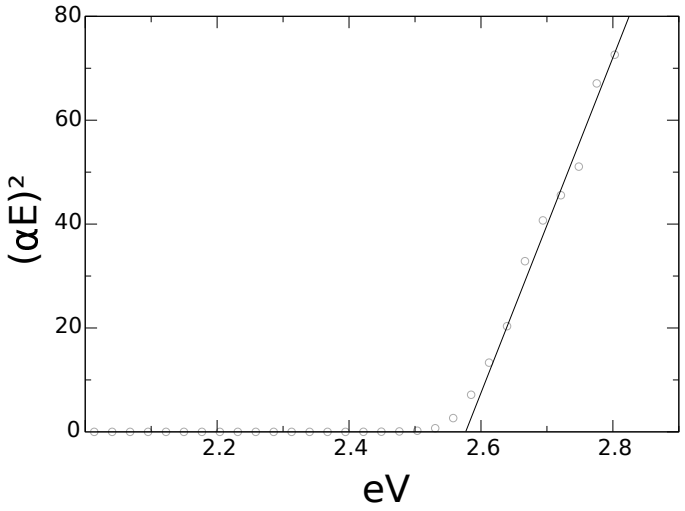
where the electric field degrees of freedom have been condensed to a single dimension, as in Voigt notation. For clarity, we show the tensor index corresponding to current direction in uppercase, while the indices giving the light polarization are in lowercase. We note that the above tensor is Cartesian, such that the rhombohedral lattice vectors in terms of Cartesian coordinates are, by convention, $\vec{A}_1 = \left(\frac{a}{2}, -\frac{a}{2\sqrt{3}}, \frac{c}{3}\right), \vec{A}_2 = \left(0, \frac{a}{\sqrt{3}}, \frac{c}{3}\right), \vec{A}_3 = \left(-\frac{a}{2}, -\frac{a}{2\sqrt{3}}, \frac{c}{3}\right)$.

Due to the well-known tendency of DFT calculations to underestimate the localization of the d-orbital electrons, the DFT+U method is used in the calculation, including an effective Hubbard U_{eff}=$U - J$ in the Hamiltonian. In order to choose the proper U_{eff} value, the imaginary permittivity was calculated in the long wavelength approximation with different U_{eff} values and compared to experiment. Shown in Fig. 5.3 (a) are the real and imaginary permittivity, with experimental data taken from Ref. [2]. We find that the calculation

with GGA+$U_{\mathrm{eff}} = 5$ eV best matches the experimental imaginary permittivity(dielectric loss) for energies near the band-gap, especially the energies of the t_{2g} and e_g peaks. As is commonly the case with DFT calculations, due in part to absence of quasiparticle corrections [62], the magnitude of the permittivity is substantially overestimated. This has been observed in hematite as well [63, 64], which is structurally and chemically similar to BFO. For additional calibration, we calculate the gap using the same method in Ref. [2], shown in Fig. 5.3(b), and find that the calculated value underestimates the experimental gap by 0.16 eV.

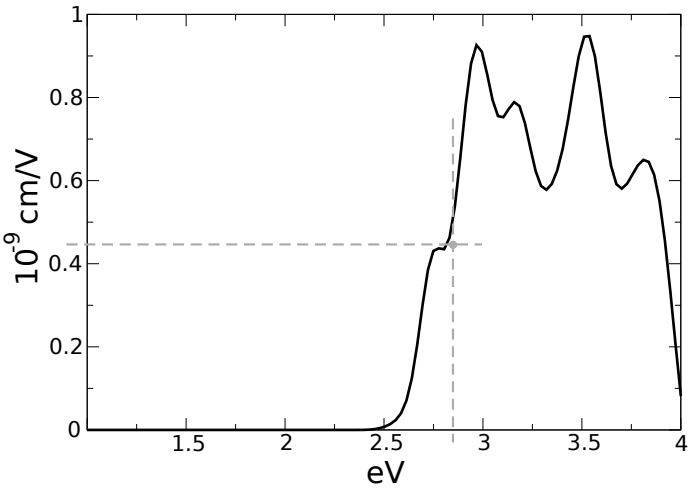

30
20
10
0
1
2
3
4
5
6
eV
ϵr - EXP
ϵr - GGA+U
ϵi - EXP
ϵi - GGA+U

(a)

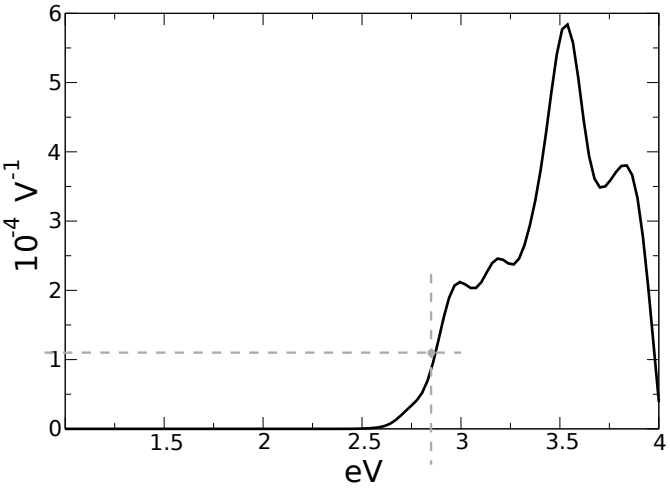

80
60
40
20
0
2.2
2.4
2.6
2.8
eV
(αE)²

(b)

Figure 5.3: In (a), the calculated components of the permittivity, ϵ, are compared to experiment. In (b), the calculated band-gap is shown to be 2.58 eV, 0.16 eV less than the experimentally determined value of 2.74 eV [2].


1
0.8
0.6
0.4
0.2
0
10^{-9} cm/V
1.5
2
2.5
3
3.5
4
eV

(a)

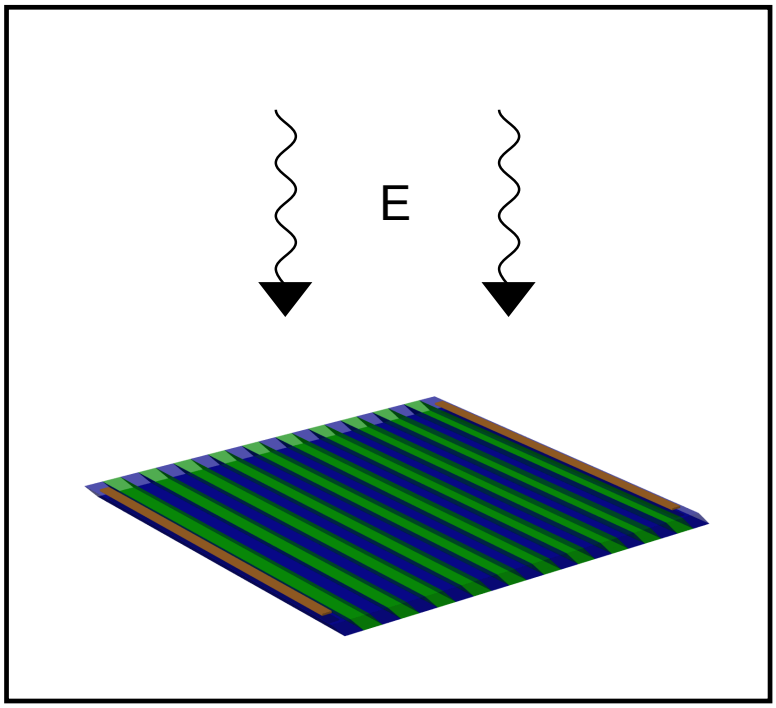

6
5
4
3
2
1
0
10^{-4} V^{-1}
1.5
2
2.5
3
3.5
4
eV

(b)

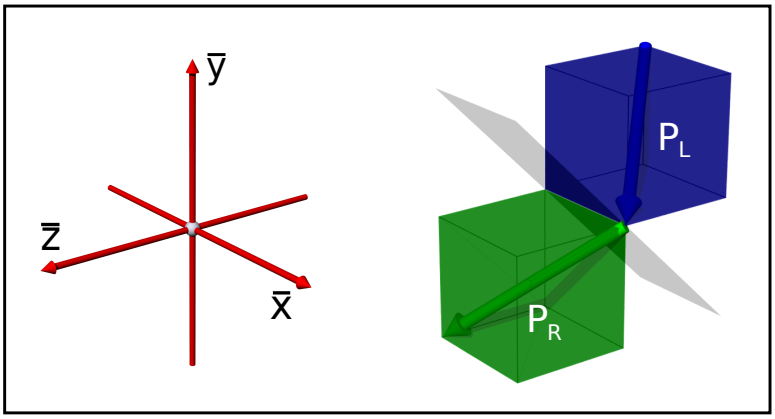
Figure 5.4: The calculated Glass coefficient (a), G_{yyY}, and bulk photovoltaic coefficient (b), σ_{yyY}, are shown, with the experimental values marked for comparison [3].

Shown in Fig. 5.4 are the calculated photovoltaic tensor elements and Glass coefficients that yield current in the Y direction in the plane normal to the polarization axis, compared with the experimental results from Ref. [3].

We have adjusted our results to account for the slightly lower band-gap, shifting them right by 0.16 eV. Compared to the experimental measurements of Glass coefficient and shift current tensor (G_{yyY}=4.48$\times 10^{-10}$ cm/V and σ_{yyY}=1.1$\times 10^{-4}$ V^{-1} around the photon energy of 2.85 eV in Ref. [3]), our results agree very well.



(a)



(b)

Figure 5.5: The experimental setup of Refs. [4] and [5]. The film shown in (a) is composed of alternating domains with polarizations of adjacent domains at 71° angles to one another. Presently we label one domain with the letter 'R' and the color green, and the other with the letter 'L' and the color blue. Large photovoltages and photocurrents are observed when electrodes, shown in gold, are placed parallel to the domains. No response is observed when the electrodes are perpendicular to the domains (not shown). In (b) the orientation of the pseudo-cubic unit cell and polarization for each domain is shown, along with the principal axes of the experimental coordinate frame.

In Ref. [4], the photovoltaic response to unpolarized light is measured in a thin film with parallel 71° domain walls. The orientation of domains in the experimental setup is shown in Fig. 5.5. The net material polarization is normal to the domain walls, and the photocurrent is measured both parallel and perpendicular to the in-plane component of net material polarization, while photovoltage is measured only parallel to the in-plane component of net material polarization. While high response in the antiparallel direction is observed for polydomain samples, with photovoltage scaling linearly with domain density, little or no photovoltage is detected for monodomain crystals, and negligible photocurrent is measured perpendicular to net polarization, supporting the proposed mechanism of domain-wall driven generation of large photovoltages. However, based on the results for photocurrent in the plane normal to the polarization obtained in Ref. [3], the authors speculate that the bulk photovoltaic response along the polarization direction should be significant. This view is supported by their studies of the photocurrent parallel to polarization [59].

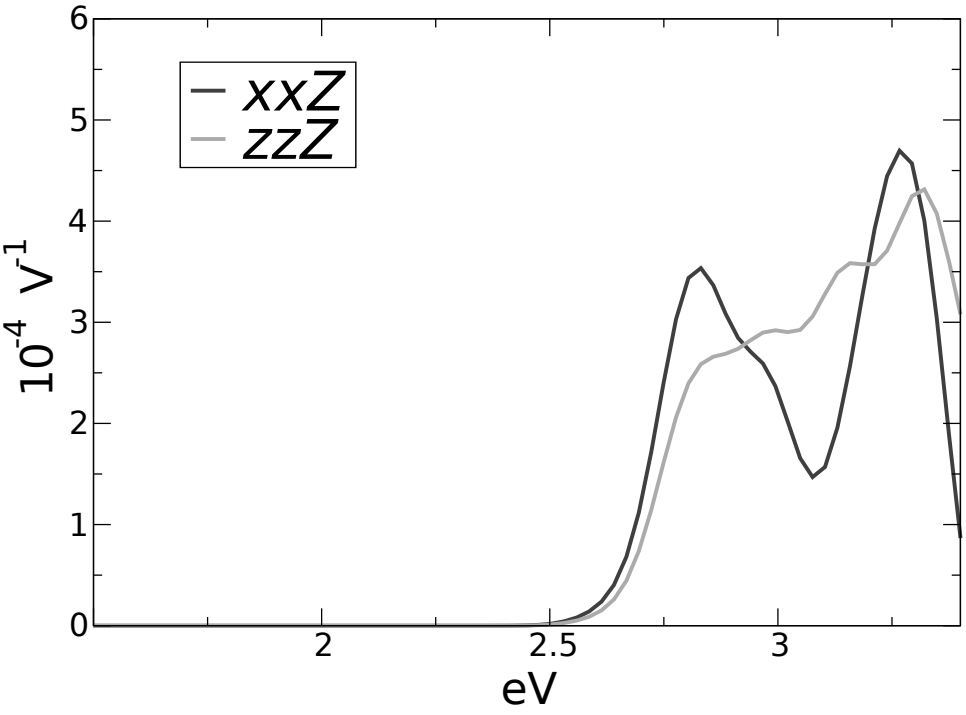

xxZ
zzZ
10^{-4} V^{-1}
eV
0
1
2
3
4
5
6
2
2.5
3

Figure 5.6: The photovoltaic response tensor elements for current collinear with the material polarization.

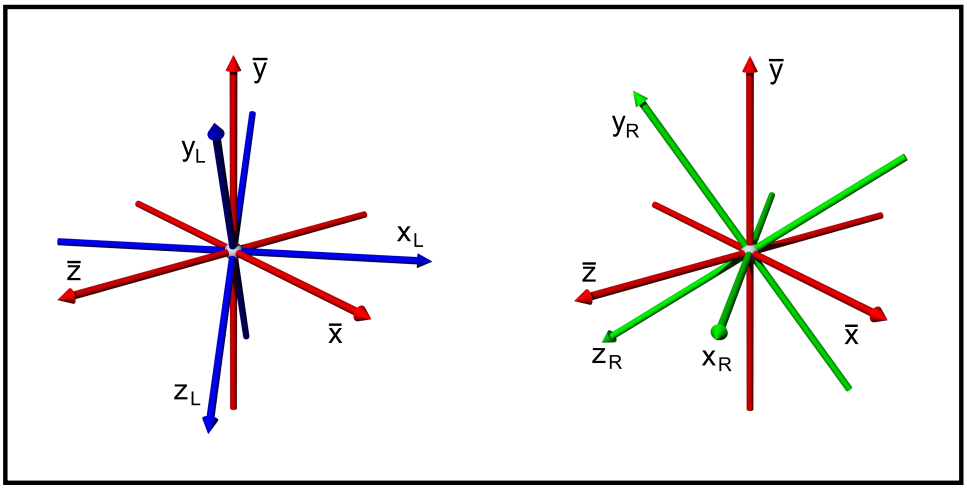

$\bar{y}$
y_L
x_L
$\bar{z}$
$\bar{x}$
z_L
$\bar{y}$
y_R
$\bar{z}$
z_R
x_R
$\bar{x}$

Figure 5.7: The relationship between the principal axes of the material coordinates and the lab coordinates $(\bar{x}, \bar{y}, \bar{z})$ in Fig. 5.5. The z_{R} and z_{L} axes are parallel to the polarization in their respective domains, the x_{R} and x_{L} are in the $\bar{x}\bar{z}$ plane, and y_{R} and y_{L} complete the orthogonal bases.

The calculated response tensor elements yielding current in the material polarization direction are shown in Fig. 5.6. The direction is uniformly parallel, rather than antiparallel, to the material polarization, and similar in magnitude to the in-plane response. These results appear to confirm that the bulk photovoltaic effect constitutes a meaningful contribution to the response observed in the experimental setup. However, the geometry of the system shown in Fig. 5.5, especially the orientation of the domain polarization to the incident illumination, suggests a more complicated picture, as many different tensor elements will contribute to the observed response. As evident from Eq. 5.1, some of these elements have opposite sign from one another, allowing for the possibility that significant cancellations may occur. To properly calculate the response, the field of the incoming radiation must first be expressed in the coordinate system of the material as used to express the response tensor. For both domains,

$$\mathbf{E}(\theta) = E_0 \begin{bmatrix} \cos(\theta) \\ \sin(\theta)\sqrt{\frac{1}{3}} \\ \sin(\theta)\sqrt{\frac{2}{3}} \end{bmatrix} \tag{5.2}$$

for light with wavevector parallel to $\bar{y}$ and polarization angle θ to x_L/x_R. In accordance with Malus's law, upolarized light can be decomposed into any two fields of perpendicular polarization; we select θ to be 0° and 90° for both domains. After computing the response in the material frame, we must rotate the current density vector back into the lab coordinate system. Thus, for unpolarized, overhead illumination, the total response can be calculated as

$$\begin{aligned} J_l = & R^{\mathrm{R}}_{lk} \left(\sigma_{ijk} \frac{E_i(0^\circ)E_j(0^\circ) + E_i(90^\circ)E_j(90^\circ)}{2} \right) \\ & + R^{\mathrm{L}}_{lk} \left(\sigma_{ijk} \frac{E_i(0^\circ)E_j(0^\circ) + E_i(90^\circ)E_j(90^\circ)}{2} \right) \end{aligned}$$

where R^{R} and R^{L} rotate from the principal axes of the two domain types, denoted as $(x_{\mathrm{R}}, y_{\mathrm{R}}, z_{\mathrm{R}})$ and $(x_{\mathrm{L}}, y_{\mathrm{L}}, z_{\mathrm{L}})$ and shown in green and blue respectively in Fig. 5.7, to experimental coordinates, denoted as $(\bar{x}, \bar{y}, \bar{z})$ and shown in red, where $\bar{z}$ is parallel to the in-plane component of net material polarization, $\bar{y}$ is normal to the film surface, and $\bar{x}$ is orthogonal to $\bar{y}$ and $\bar{z}$, forming a right-handed coordinate system. In the barred system

$$x_{\mathrm{R}} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) \qquad x_{\mathrm{L}} = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

$$y_{\mathrm{R}} = \left(-\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, \frac{1}{\sqrt{6}}\right) \qquad y_{\mathrm{L}} = \left(\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, \frac{1}{\sqrt{6}}\right)$$

$$z_{\mathrm{R}} = \left(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \qquad z_{\mathrm{L}} = \left(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

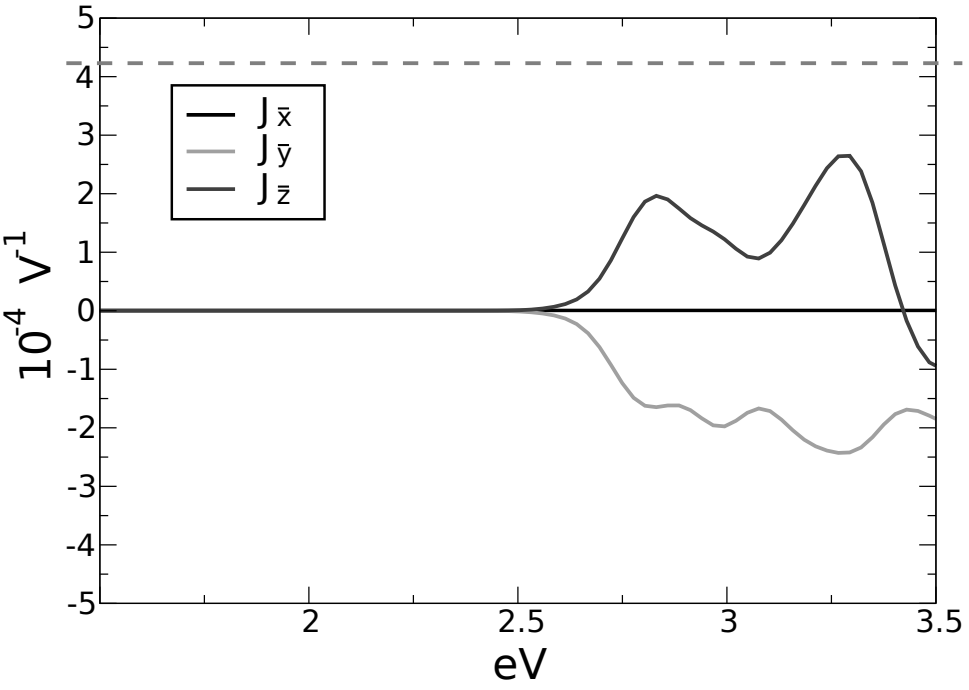

5
4
3
2
1
0
-1
-2
-3
-4
-5
2
2.5
3
3.5
eV
10^{-4} V^{-1}
J$_{\bar{x}}$
J$_{\bar{y}}$
J$_{\bar{z}}$

Figure 5.8: The calculated current densities per light intensity in the lab frame of the BFO film, with the experimental value from Ref. [4] marked by the red dashed line. The current in the $\bar{x}$ direction, parallel to the domain walls, vanishes, while much of the remaining current is directed toward the bottom of the sample.

The components of the total current in the lab frame are shown in Fig. 5.5(a). The current in the $\bar{x}$ direction – parallel to the domain walls – vanishes, in agreement with experiment. The current in the $\bar{z}$ direction is substantial, however, and proceeds in the opposite direction of the experimental response, indicating that the two photovoltaic effects partially cancel. The magnitude of this component is a large fraction of the experimentally measured current, suggesting that the reduction in current may be significant. In the single-domain case, where the impact of carrier separation at domain walls is suppressed, one might expect a photovoltage in the opposite direction, but experimentally negligible photovoltage was observed. However, a large portion of the current is directed to the lower surface, especially in a thin film geometry. Upon illumination, layers of carriers will form on the surfaces, allowing any charge imbalances between electrodes to rapidly equilibrate without domain wall traps in between. Thus, any shift current in the $\bar{z}$ direction will be prevented from sustaining a photovoltage in that direction, due to the conductivity of the surfaces resulting from the $\bar{y}$ directed photovoltage. This is also consistent with the observation of significant photocurrent obtained in Ref. [26], where the orientation of the crystal to the incident light is the same, but current is measured in the direction perpendicular to that of the setup in Ref. [4].

We have reconciled the large bulk photovoltaic response found in Refs. [3] and [5] with the apparent negligible contribution evinced in Ref. [4] through first principles calculations. Importantly, we find that the bulk photovoltaic effect will partially cancel domain-wall driven carrier separation, indicating that even higher efficiencies may be possible. Effective photovoltaic materials may be found which take advantage of a domain wall structure of the kind explored in Refs. [4] and [5], especially as a mechanism of trapping carriers, but relying on the bulk photovoltaic effect to contribute to carrier separation, rather than suppress it.

Chapter 6

Relationship to Material Properties

Having established with reasonable certainty that the shift current is the dominant component of BPVE, we turn our attention to the question of how it is influenced by the material's structure and chemistry. While Eq. (4.4) allows us to calculate the response, it contains few clues about the dependence on material properties. As demonstrated, the response depends on the strength of linear excitations, which are at least familiar, and the mysterious quantity of shift vector. Since the carrier velocity is proportional to the shift vector, as shown in Eq. (4.6), it is vital to understand what determines its magnitude. A natural place to start is clarifying the relationship to material polarization. It is apparent that there is no direct, mechanistic dependence of response on material polarization, as is the case for many mechanisms to which photovoltaic effects in ferroelectrics have been attributed. However, shift current requires broken inversion symmetry, which here derives from the lattice distortion that produces ferroelectric polarization, suggesting that the response may appear to depend on polarization in some fashion. Unfortunately, the mathematical form of the shift vector does not reveal a straightforward relationship between the magnitude of symmetry breaking, and the resulting shift current response.

6.1 Numerical Study

The content of this section appears in Ref. [54]. To investigate the connection to material polarization, we studied a systematic family of structures based on $PbTiO_3$. Starting with the cubic perovskite in the paraelectric structure, we rigidly displaced oxygen ions along a single Cartesian axis by amplitudes ranging from 0.01 to 0.09 lattice vectors, without otherwise altering the geometry. The spectra of shift current and aggregate shift vector are shown in Fig. 6.1 for several displacements. The results indicate a complex relationship between shift current and material polarization. As Fig. 6.1 shows, with soft mode amplitude 0.01, the shift current at 3.2 eV above band-gap is negative; with amplitude 0.07, the shift current reverses direction, resulting in a change of -200%. With amplitude 0.01, there is a negative peak at 3.8 eV; with amplitude 0.07, the peak shifts to 4.2 eV and is four times the size, for an increase of over 300%. However, in the intervening frequency range, the response is relatively small at all displacements.

Next, we turn our attention to the integrated shift vector $\bar{\mathcal{R}}_Z(\omega)$. The changes in shift vector are of special interest, since the symmetry constrains the overall shift current expression via the shift vector. The integrated shift vector spectrum echoes the overall current response, but contains some distinct features. The increase in current from 4-5 eV does not appear to result from increased shift vector length, but from stronger coincidence of high transition intensity and large shift vectors. In fact, the overall shift vector changes little with displacement. However, from 7.5-8.5 eV, the integrated shift vector changes dramatically, suggesting that at some points in the Brillouin zone the oxygen displacement substantially alters the shift vector. Changes to the overall response are thus a combination of changes both to shift vector and associated intensity.

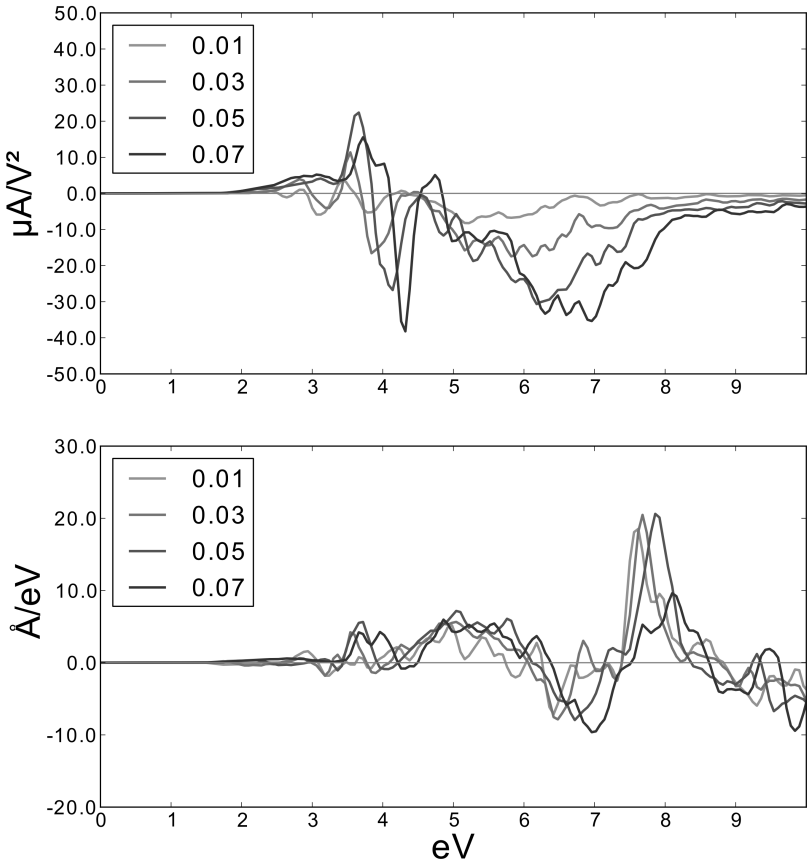
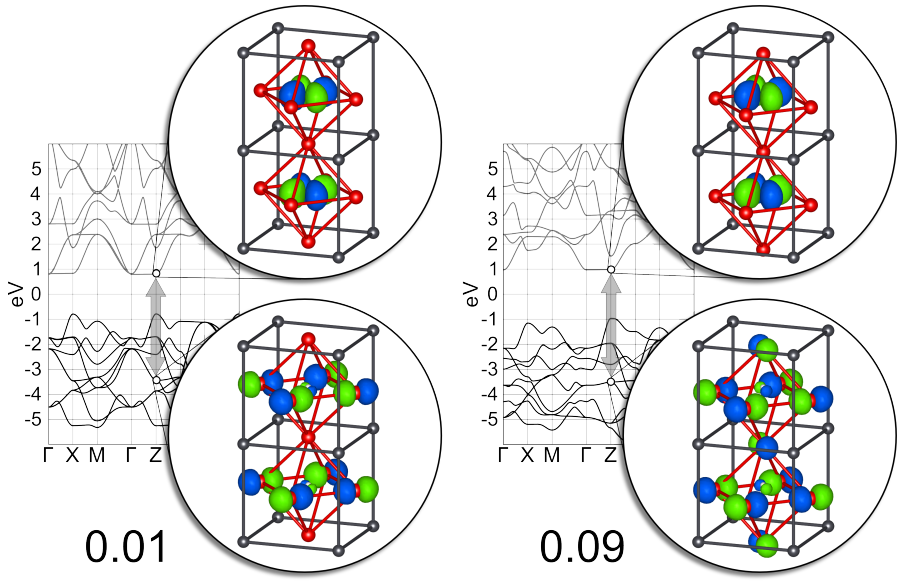

0.01
0.03
0.05
0.07
μA/V²
Å/eV
eV

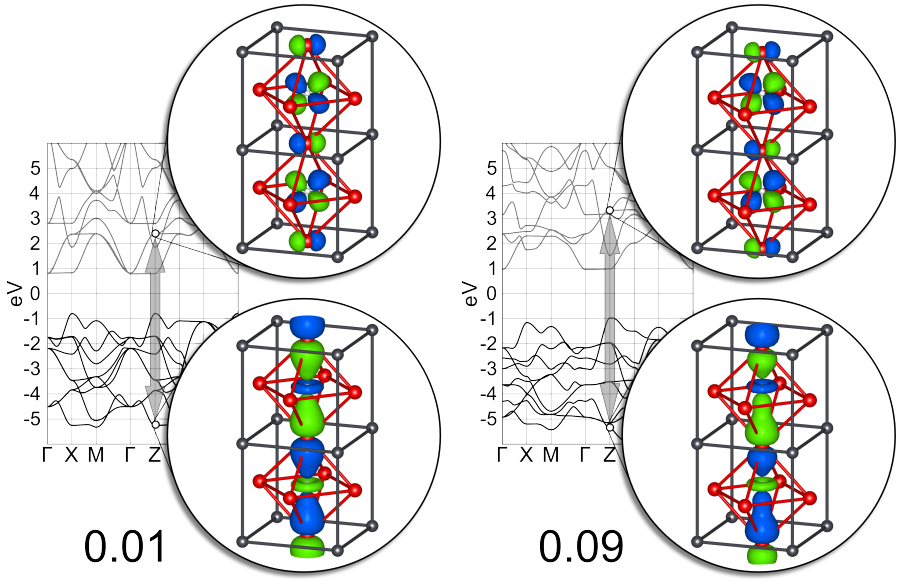
Figure 6.1: The overall current susceptibility and aggregated shift vector $\bar{\mathcal{R}}$ are shown for $PbTiO_3$ with varying polarization.

To understand these results, the electronic bands participating in transitions in these frequency ranges were examined directly. For the 4-5 eV range, examples of the transitions and associated Bloch states that dominate are shown in Fig. 6.2(a) at 0.01 and 0.09 lattice vector displacements. For this transition, the shift vector is 0.6 Å at displacement of 0.01, and 1.0 Å at 0.09. The valence state is largely composed of oxygen p-orbitals, while the conduction state is essentially a titanium d_{xy} state. The states, like the shift vectors, are largely unchanged by the oxygen displacement.

However, the transitions in the higher energy range are notably different. Shown in Fig. 6.2(b) are examples of the dominant transitions in the 7.5-8.5 eV range. The shift vector is large and positive (32.3 Å) for 0.01 lattice vector displacement, and large but negative (-22.7 Å) at 0.09 displacement. The participating valence state can be characterized as bonding between the Ti and O atoms collinear with polarization, while the conduction state features Ti-O anti-bonding. These results point not to a simple dependence on material polarization, but to a dependence of shift current on the extent of localization of the initial and final states, which in turn depends on atomic displacement. Transitions between states that do not experience bonding interactions in the direction of ferroelectric polarization manifest short shift vectors and insensitivity to oxygen displacement.



(a)



(b)

Figure 6.2: (a) The non-bonding Bloch states of $PbTiO_3$ are involved in a transition that is insensitive to material polarization, with a shift vector length change from 0.6 Å to 1.0 Å as O sublattice displacement increases from 0.01 to 0.09, and (b) a transition from bonding to antibonding gives a shift vector that is highly sensitive to material polarization, with shift vector length change from 32.4 Å to -22.7 Å for increasing O sublattice displacement.

6.2 Analytical Model

Using a simple tight-bonding model, we will derive an analytical expression for the shift current, illuminating its dependence on system properties and corroborating our empirical study. Since we are interested in quantities that depend on the phase of the Bloch functions, we must construct our eigenstates with care. In general, for a basis composed of orbital-like functions $\phi_j(\mathbf{r})$ centered at $\mathbf{r}_j$ in the unit cell

$$\psi_n(\mathbf{k},\mathbf{r}) = \sum_{lj} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R}_l} u_{nj}(\mathbf{k})\varphi_j(\mathbf{k},\mathbf{r}-\mathbf{R}_l)$$

where $\mathbf{R}_l$ are lattice vectors, and

$$\varphi_j(\mathbf{k},\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}\phi_j(\mathbf{r})$$

We can see that Eq. (3.1) give the correct expression for the periodic position elements

$$\begin{aligned}
\langle n\mathbf{k}'|\,\hat{\mathbf{r}}\,|n\mathbf{k}\rangle &= \sum_{l'lj'j} \mathrm{e}^{-i\mathbf{k}'\cdot\mathbf{R}_{l'}}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{R}_l} u^\dagger_{nj'}(\mathbf{k}')u_{nj}(\mathbf{k})\,\langle\varphi_{j'}(\mathbf{k},\mathbf{r}-\mathbf{R}_{l'})|\,\hat{\mathbf{r}}\,|\varphi_j(\mathbf{k},\mathbf{r}-\mathbf{R}_l)\rangle \\
&= \sum_{l'lj'j} u^\dagger_{nj'}(\mathbf{k}')u_{nj}(\mathbf{k})\,\langle\phi_{j'}(\mathbf{r}-\mathbf{R}_{l'})|\,\hat{\mathbf{r}}\,|\phi_j(\mathbf{r}-\mathbf{R}_l)\rangle \\
&= \sum_{lj} u^\dagger_{nj}(\mathbf{k}')u_{nj}(\mathbf{k})\,(\mathbf{R}_l+\mathbf{r}_j) \\
&= -i\,\langle n\mathbf{k}|\,\nabla_{\mathbf{k}}\,|n\mathbf{k}\rangle + i\delta(\mathbf{k}-\mathbf{k}')\sum_j u^\dagger_{nj}(\mathbf{k}')\nabla_{\mathbf{k}}u_{nj}(\mathbf{k}) \\
&= -i\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}') + i\delta(\mathbf{k}-\mathbf{k}')\sum_j u^\dagger_{n'j}(\mathbf{k}')\nabla_{\mathbf{k}}u_{nj}(\mathbf{k})
\end{aligned}$$

and our momentum elements are

$$\begin{aligned}\langle n'\mathbf{k}'|\,\hat{\mathbf{P}}\,|n\mathbf{k}\rangle &= \sum_{l'lj'j} u^{\dagger}_{n'j'}(\mathbf{k}')u_{nj}(\mathbf{k})\mathrm{e}^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}\,\langle\phi_{j'}(\mathbf{r}-\mathbf{R}_{l'})|\,\mathbf{k}-i\nabla_{\mathbf{r}}\,|\phi_j(\mathbf{r}-\mathbf{R}_l)\rangle \\ &= \hbar\mathbf{k}\delta(\mathbf{k}-\mathbf{k}')\delta_{n'n} \\ &\quad - \hbar\delta(\mathbf{k}-\mathbf{k}')\sum_{lj'j} u^{\dagger}_{n'j'}(\mathbf{k}')u_{nj}(\mathbf{k})\,\langle\phi_{j'}(\mathbf{r}-\mathbf{R}_l)|\,i\nabla_{\mathbf{r}}\,|\phi_j(\mathbf{r})\rangle\end{aligned}$$

The shift vector can then be written as

$$\begin{aligned}\mathcal{R}_{n'n}(\mathbf{k}) &= \chi_n(\mathbf{k}) - \chi_{n'}(\mathbf{k}) - i\frac{\nabla_{\mathbf{k}}\,\langle n'\mathbf{k}'|\,\hat{\mathbf{P}}\,|n\mathbf{k}\rangle}{\langle n'\mathbf{k}'|\,\hat{\mathbf{P}}\,|n\mathbf{k}\rangle} \\ &= \sum_j u^{\dagger}_{nj}(\mathbf{k})\nabla_{\mathbf{k}}u_{nj}(\mathbf{k}) - \sum_j u^{\dagger}_{n'j}(\mathbf{k})\nabla_{\mathbf{k}}u_{n'j}(\mathbf{k}) \\ &\qquad + \frac{\sum_{lj'j} i\nabla_{\mathbf{k}}\left[u^{\dagger}_{n'j'}(\mathbf{k})u_{nj}(\mathbf{k})\right]\,\langle\phi_{j'}(\mathbf{r}-\mathbf{R}_l)|\,i\nabla_{\mathbf{r}}\,|\phi_j(\mathbf{r})\rangle}{\sum_{lj'j} u^{\dagger}_{n'j'}(\mathbf{k})u_{nj}(\mathbf{k})\,\langle\phi_{j'}(\mathbf{r}-\mathbf{R}_l)|\,i\nabla_{\mathbf{r}}\,|\phi_j(\mathbf{r})\rangle}\end{aligned}$$

In particular, we can consider a very simple 1D model with two sites, similar to the Su-Schrieffer-Heeger model. We label the sites A and B, giving them different on-site potentials, and introduce asymmetry via the hopping terms. We include higher-order hopping terms between the two site types that we assume to decay exponentially

$$\begin{aligned}H(k) = \epsilon c^{\dagger}_A c_A - \epsilon c^{\dagger}_B c_B &+ t\left[\mathrm{e}^{\Delta}\mathrm{e}^{ik\frac{R}{2}}\sum_l \mathrm{e}^{-(\tau-ik)lR} + \mathrm{e}^{-\Delta}\mathrm{e}^{-ik\frac{R}{2}}\sum_l \mathrm{e}^{-(\tau+ik)lR}\right] c^{\dagger}_B c_A \\ &+ t\left[\mathrm{e}^{-\Delta}\mathrm{e}^{ik\frac{R}{2}}\sum_l \mathrm{e}^{-(\tau-ik)lR} + \mathrm{e}^{\Delta}\mathrm{e}^{-ik\frac{R}{2}}\sum_l \mathrm{e}^{-(\tau+ik)lR}\right] c^{\dagger}_A c_B\end{aligned}$$

We have set $\mathbf{R}_l \to lR$, where R is now the primitive lattice vector, and assigned Δ and τ to represent the asymmetry and hopping decay, respectively. The sums can be expressed in

closed form, yielding

$$
\begin{aligned}
H(k) = \epsilon c_A^\dagger c_A - \epsilon c_B^\dagger c_B \\
&+ 2t \frac{\cosh(\Delta)\cos\left(k\frac{R}{2}\right)\sinh\left(\tau\frac{R}{2}\right) - i\sinh(\Delta)\sin\left(k\frac{R}{2}\right)\cosh\left(\tau\frac{R}{2}\right)}{\mathrm{e}^{-\tau\frac{R}{2}}\left[\cosh(\tau R) - \cos(kR)\right]} c_B^\dagger c_A \\
&+ 2t \frac{\cosh(\Delta)\cos\left(k\frac{R}{2}\right)\sinh\left(\tau\frac{R}{2}\right) + i\sinh(\Delta)\sin\left(k\frac{R}{2}\right)\cosh\left(\tau\frac{R}{2}\right)}{\mathrm{e}^{-\tau\frac{R}{2}}\left[\cosh(\tau R) - \cos(kR)\right]} c_A^\dagger c_B
\end{aligned}
$$

Letting

$$
\begin{aligned}
\Omega(k) &= 2t \frac{\cosh(\Delta)\cos\left(k\frac{R}{2}\right)\sinh\left(\tau\frac{R}{2}\right) + i\sinh(\Delta)\sin\left(k\frac{R}{2}\right)\cosh\left(\tau\frac{R}{2}\right)}{\mathrm{e}^{-\tau\frac{R}{2}}\left[\cosh(\tau R) - \cos(kR)\right]} \\
\Theta(k) &= \arctan\left[\sin\left(k\frac{R}{2}\right)\frac{\tanh(\Delta)}{\tanh\left(\tau\frac{R}{2}\right)}\right]
\end{aligned}
$$

The energy and wavefunctions can be written as

$$
\begin{aligned}
E^+(k) &= -E^-(k) = \sqrt{\epsilon^2 + |\Omega(k)|^2} \\
\psi^\pm(k) &= \sum_l \mathrm{e}^{iklR}\left[u_A^\pm(k)\varphi_A(r - lR) + u_B^\pm(k)\varphi_B(r - lR)\right] \\
u_A^+(k) &= \frac{\mathrm{e}^{i\Theta(k)}}{\sqrt{2}}\sqrt{1 + \frac{\epsilon}{E(k)}}\, u_A^-(k) = \frac{\mathrm{e}^{i\Theta(k)}}{\sqrt{2}}\sqrt{1 - \frac{\epsilon}{E(k)}} \\
u_B^+(k) &= \frac{1}{\sqrt{2}}\sqrt{1 - \frac{\epsilon}{E(k)}}, u_B^-(k) = -\frac{1}{\sqrt{2}}\sqrt{1 + \frac{\epsilon}{E(k)}}
\end{aligned}
$$

We can now write the momentum as

$$
\begin{aligned}
P^{\pm}(k) =& \sum_{l} u_A^{+\dagger}(k) u_A^{-}(k) \left\langle \phi_A(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_A(r) \right\rangle \\
&+ \sum_{l} u_B^{+\dagger}(k) u_B^{-}(k) \left\langle \phi_B(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_B(r) \right\rangle \\
&+ \sum_{l} u_A^{+\dagger}(k) u_B^{-}(k) \left\langle \phi_A(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_B(r) \right\rangle \\
&+ \sum_{l} u_B^{+\dagger}(k) u_A^{-}(k) \left\langle \phi_B(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_A(r) \right\rangle \\
P^{\pm}(k) =& \frac{1}{2}\sqrt{1 - \left(\frac{\epsilon}{E(k)}\right)^2} \\
&\times \sum_{l} \left[\left\langle \phi_A(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_A(r) \right\rangle - \left\langle \phi_B(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_B(r) \right\rangle \right] \\
&- \frac{\mathrm{e}^{-i\Theta(k)}}{2} \left(1 + \frac{\epsilon}{E(k)}\right) \sum_{l} \left\langle \phi_A(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_B(r) \right\rangle \\
&+ \frac{\mathrm{e}^{i\Theta(k)}}{2} \left(1 - \frac{\epsilon}{E(k)}\right) \sum_{l} \left\langle \phi_B(r - lR) \right| i\frac{\partial}{\partial r} \left| \phi_A(r) \right\rangle
\end{aligned}
$$

Assuming that the magnitude of the momentum overlaps are governed by Δ and τ with strength proportional to a constant C, we obtain

$$
\begin{aligned}
P^{\pm}(k) &= i\hbar C \frac{\sinh(\Delta)}{\mathrm{e}^{-\tau\frac{R}{2}} \sinh\left(\tau\frac{R}{2}\right)} \left[\frac{\mathrm{e}^{-i\Theta(k)}}{2} \left(1 + \frac{\epsilon}{E(k)}\right) - \frac{\mathrm{e}^{i\Theta(k)}}{2} \left(1 - \frac{\epsilon}{E(k)}\right) \right] \\
P^{\pm}(k) &= \hbar C \frac{\sinh(\Delta)}{\mathrm{e}^{-\tau\frac{R}{2}} \sinh\left(\tau\frac{R}{2}\right)} \left[\sin(\Theta) + i\frac{\epsilon}{E(k)} \cos(\Theta) \right]
\end{aligned}
$$

and the shift vector is (taking only the real part)

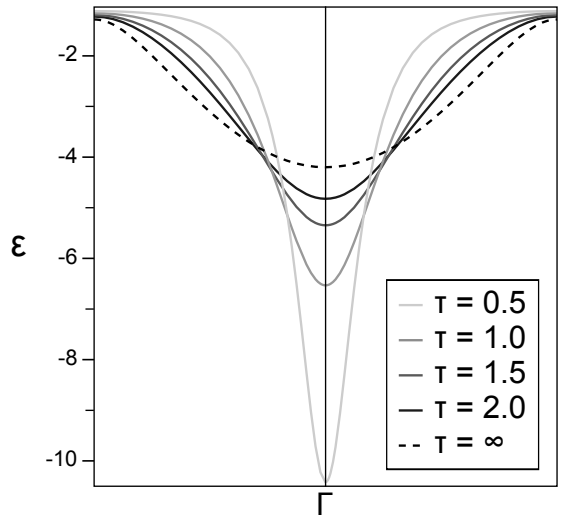
$$\mathcal{R}^{\pm}(k) = -\left[\frac{1}{2}\left(1-\frac{\epsilon}{E(k)}\right)\frac{\partial\Theta}{\partial k} - \frac{1}{2}\left(1+\frac{\epsilon}{E(k)}\right)\frac{\partial\Theta}{\partial k}\right] + i\frac{1}{P^{\pm}(k)}\frac{\partial P^{\pm}}{\partial k}$$

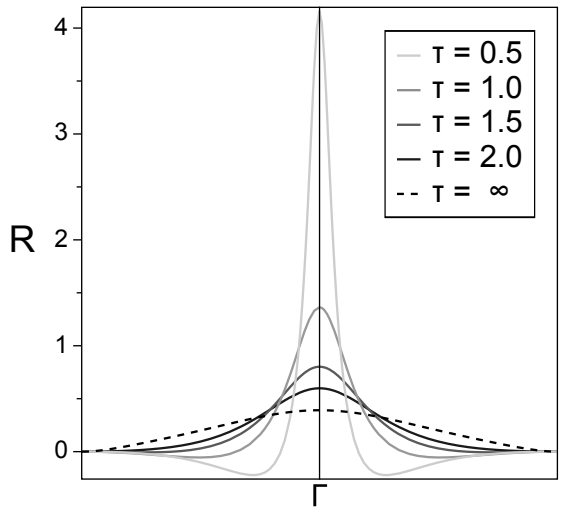
$$\mathcal{R}^{\pm}(k) = \frac{\epsilon}{E(k)}\frac{\partial\Theta}{\partial k} + i\frac{P^{\dagger\pm}(k)}{|P^{\pm}(k)|^2}\left(\frac{\partial P^{\pm}}{\partial\Theta}\frac{\partial\Theta}{\partial k} + \frac{\partial P^{\pm}}{\partial E}\frac{\partial E}{\partial k}\right)$$

$$\mathcal{R}^{\pm}(k) = \left[1 - \frac{\sin^2(\Theta)+\cos^2(\Theta)}{\sin^2(\Theta)+\frac{\epsilon}{E(k)}^2\cos^2(\Theta)}\right]\frac{\epsilon}{E(k)}\frac{\partial\Theta}{\partial k} + \left[\frac{\sin(\Theta)\cos(\Theta)-\sin(\Theta)\cos(\Theta)\frac{\epsilon}{E(k)}^2}{\sin^2(\Theta)+\frac{\epsilon}{E(k)}^2\cos^2(\Theta)}\right]\frac{\partial E}{\partial k}$$

$$\mathcal{R}^{\pm}(k) = \left[1 - \frac{1}{\sin^2(\Theta)+\frac{\epsilon}{E(k)}^2\cos^2(\Theta)}\right]\frac{\epsilon}{E(k)}\frac{\partial\Theta}{\partial k} + \frac{\sin(\Theta)\cos(\Theta)}{\sin^2(\Theta)+\frac{\epsilon}{E(k)}^2\cos^2(\Theta)}\left[1-\frac{\epsilon}{E(k)}^2\right]\frac{\partial E}{\partial k}$$

The first term ensures gauge invariance; the effect of an arbitrary gauge only appears in this term; thus both terms are gauge independent.


-2
-4
-6
-8
-10
ε
Γ
τ = 0.5
τ = 1.0
τ = 1.5
τ = 2.0
τ = ∞


4
3
2
1
0
R
Γ
τ = 0.5
τ = 1.0
τ = 1.5
τ = 2.0
τ = ∞

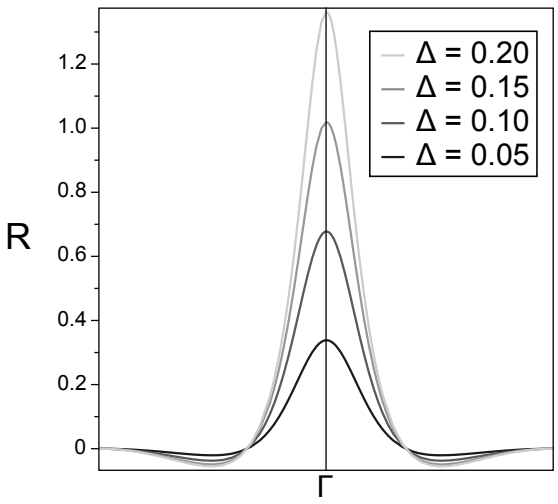

1.2
1.0
0.8
0.6
0.4
0.2
0
R
Γ
Δ = 0.20
Δ = 0.15
Δ = 0.10
Δ = 0.05

Figure 6.3: The (a) lower energy band for different values of the higher-order-hopping decay constant τ, and the shift vector in units of primitive lattice vector for different values of (b) τ and (c) asymmetry coefficient Δ. In all cases, the hopping constant t is taken to be twice the on-site energy ϵ. The effect of significant higher-order hopping is to increase and decrease the band dispersion at Γ and the Brillouin zone edge, respectively, and to increase the shift vector at Γ. The shift vector has a roughly linear dependence on the asymmetry.

In Fig. 6.3 we have plotted the energy band and shift vector for different values of Δ and τ. As seen in 6.3(a), the energy dispersion increases with larger τ as one expects, but interestingly the shape also changes, resulting in a sharp peak at Γ for small τ. This sharp dispersion is associated with increased shift vectors, seen in 6.3(b), which are increase considerably as higher-order hopping becomes significant. The effect is most pronounced at Γ since long-range hopping benefits favorably from a phase that only varies slowly in space.

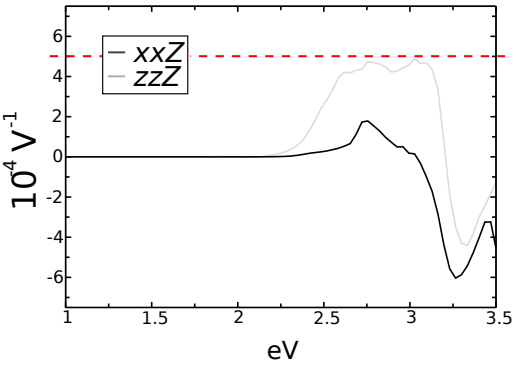
It is clear that there are two main drivers of high carrier velocity: delocalization and asymmetry of electronic states involved in a transition. The highest shift vectors and greatest potential for significant current occur when electronic states are relatively diffuse, and symmetry breaking of the lattice is able to induce corresponding asymmetry of the wavefunctions. While material polarization will be correlated with shift vector length, it will be only weakly. Materials design efforts should focus on materials with strongly covalent bonding between diffuse orbitals that lie on atoms experiencing large displacements.

Chapter 7

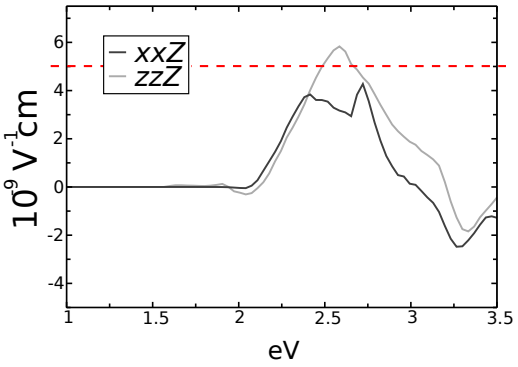
Design of Materials with Large BPVE Response

7.1 Design Strategy

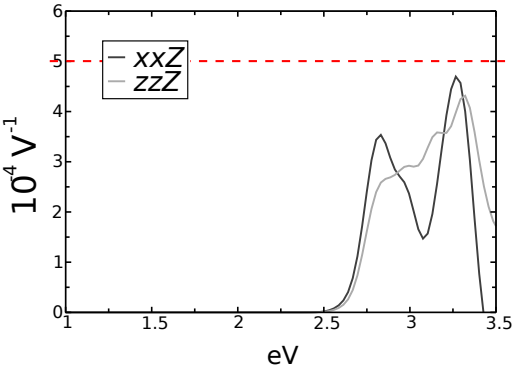
In the previous chapter we revealed the dependence of shift vector magnitude on the chemical and structural properties of materials. Large shift vectors were characterized by valence and/or conduction states that are both strongly asymmetric and delocalized in the current direction. In this regard, many distorted perovskite (ABO_3) ferroelectrics are crippled by the presence of d^0 cations enclosed in octahedral oxygen cages. The conduction states are dominated by t_{2g}-like d-states that are largely non-bonding. Coupled with the tendency for d states to localize, the result is that both shift vectors and transition response are very weak near the band-gap.


xxZ
zzZ
10^4 V^-1
eV

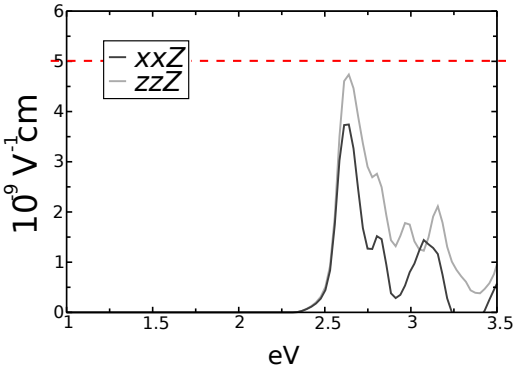
(a)


xxZ
zzZ
10^9 V^-1cm
eV

(b)


xxZ
zzZ
10^4 V^-1
eV

(c)


xxZ
zzZ
10^9 V^-1cm
eV

(d)

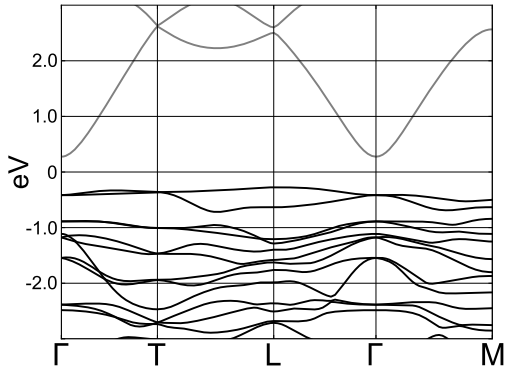
Figure 7.1: The current density response for LNO and BFO are shown in (a) and (c), respectively. The Glass coefficients of LNO and BFO appear in (b) and (d), respectively. In all cases, only the response in the direction of material polarization is shown, for both perpendicular (xxZ) and parallel (zzZ) light polarization.

We investigated systems that involve both large distortions to oxygen cages, reducing the non-bonding character of any d^0 states, as well as d^{10} cations with s and/or p states in near the band edge. It has already been noted that d^{10} cations can dramatically improve the activity of photocatalysts [65]. We found the most promising candidates to be polar oxides taking the distorted ilmenite structure, with $d^{10}s^0$ cations Pb^{4+} and Bi^{5+}. This structure can also be obtained by distorting the perovskite structure rhombohedrally, and allowing polar distortions along and oxygen-cage rotations about $\langle 111 \rangle$. Notable ferroelectrics with this structure include $LiNbO_3$ (LNO) and $BiFeO_3$ (BFO), which we discussed previously. LNO is known for its large non-linear optical response, and was one of the first materials in which the bulk photovoltaic effect was observed and studied [9, 66, 67]; however, its band-gap is well outside the visible spectrum [68]. BFO has garnered much attention recently for its multiferroic behavior [58] and low band-gap of about 2.74eV [2], which has led to explorations of its photovoltaic response [26, 4, 5, 59, 3]. We have used these as benchmarks for the present study. In all of these cases, as with the archtypical ferroelectrics $BaTiO_3$ and $PbTiO_3$, the LUMO are dominated by cation d-states, and have very similar response magnitudes.

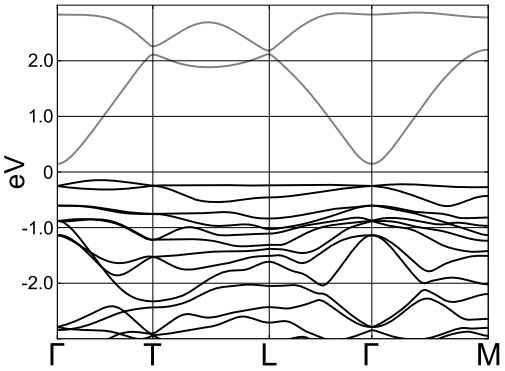
Wavefunctions were generated using QUANTUM ESPRESSO [69]. For $PbNiO_3$, a Hubbard U of 4.6eV was used, as in [6]. For band-gaps, self-consistent GW calculations using ABINIT were performed on a $4 \times 4 \times 4$ grid. The present results are for the experimental structure in the case of $PbNiO_3$, and computationally relaxed structures for the other materials.

We consider only current response in the direction of material polarization for both perpendicular (xxZ) and parallel (zzZ) light polarization, as these are the only tensor elements that can contribute to response to unpolarized light. For ease of comparison we mark baseline values reflecting the magnitude of response of our benchmark materials, shown in Fig. 7.1. These are, for the current density and Glass coefficient, respectively, $5 \times 10^{-4} \mathrm{V}^{-1}$

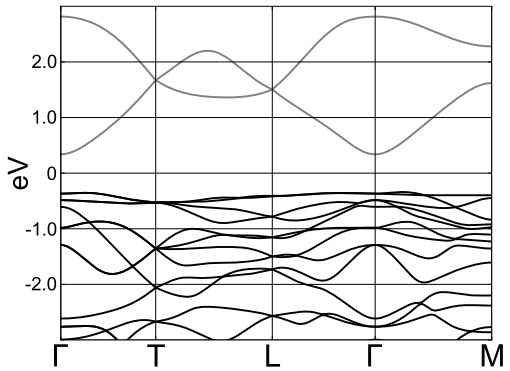
and 5×10^{-9}cm/V.



(a)



(b)

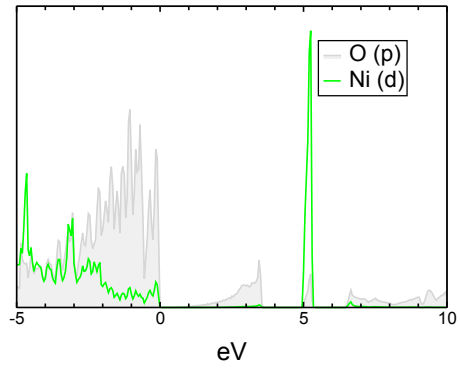


(c)

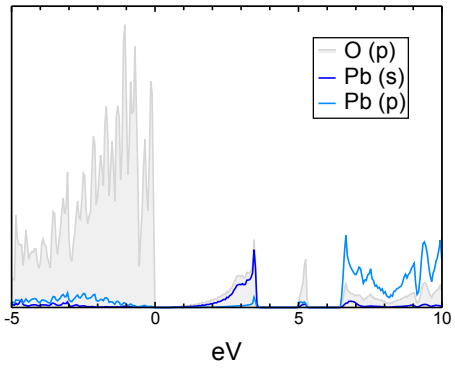
Figure 7.2: The band structures of (a) $PbNiO_3$, (a) $PbMg_{1/2}Zn_{1/2}O_3$, and (a) $BiLiO_3$.

7.2 First-principles calculations of $PbNiO_3$, $PbMg_{1/2}Zn_{1/2}O_3$, and $BiLiO_3$

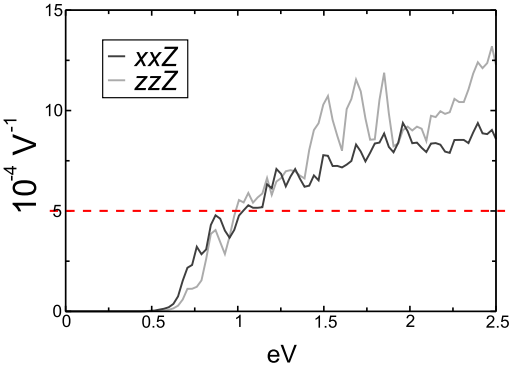
We present the calculated results of three materials: $PbNiO_3$, $PbMg_{1/2}Zn_{1/2}O_3$, and $BiLiO_3$. The first has been synthesized [70], and the latter two are similar in composition to known materials. All three satisfy our requirements of low band-gap, d^{10} cations, and large polar distortions. Furthermore, as seen in Fig. 7.2, all three have qualitatively the same band structure, featuring highly dispersive conduction bands, in contrast to the usual case of d^0 perovskite derivatives. As we will show, this arises due to unfilled s-like – rather than d-like – states composing the conduction band, and has profound consequences for the bulk photovoltaic response.


O (p)
Ni (d)
-5
0
5
10
eV

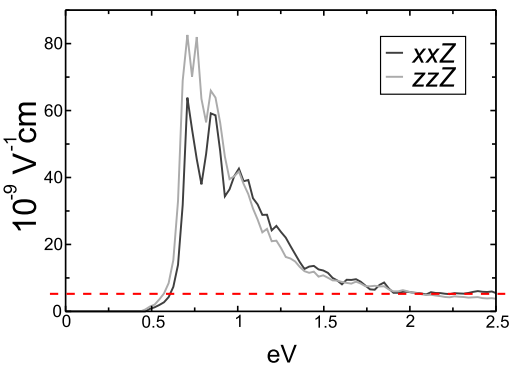
(a)


O (p)
Pb (s)
Pb (p)
-5
0
5
10
eV

(b)


xxZ
zzZ
10-4 V-1
15
10
5
0
0
0.5
1
1.5
2
2.5
eV

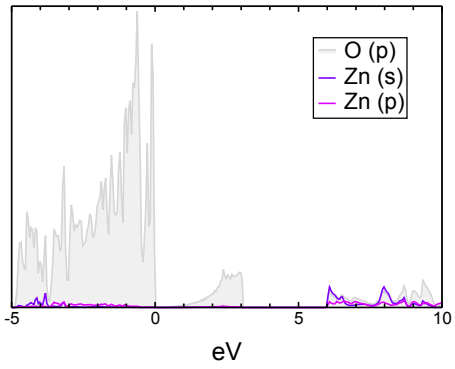
(c)


xxZ
zzZ
10-9 V-1cm
80
60
40
20
0
0
0.5
1
1.5
2
2.5
eV

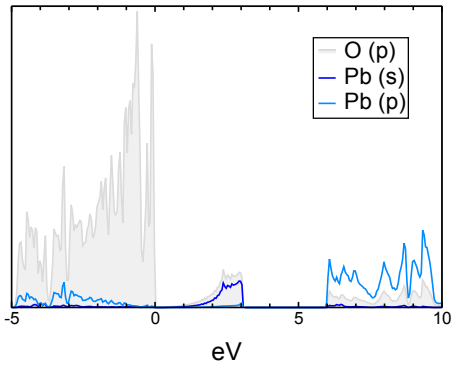
(d)

Figure 7.3: (a) and (b) give the projected density of states for $PbNiO_3$. The unfilled half of e_g of the high-spin d^8 nickel appears as a sharp peak *above* the unfilled lead s-orbitals, which have strongly hybridized with oxygen p-orbitals, resulting in a low band-gap (1.2eV in HSE [6]). This also effects a large (c) current density response, and a very large (d) Glass coefficient.

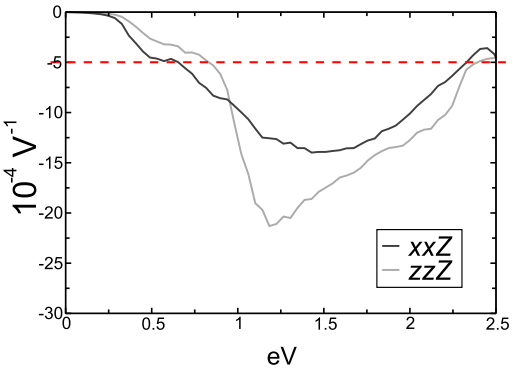
$PbNiO_3$ has recently been synthesized [70] and explored theoretically [6]. Like BFO, it is antiferromagnetic with weak spin-canting, and possesses an even larger polarization, calculated at 100 μC/cm^2 [6]. Its band-gap is even lower than BFO, with HSE predicting 1.2eV [6]. In BFO, bismuth has oxidation state $3+$, so that its s orbital is filled, and the exchange splitting of iron determines the gap. However, in $PbNiO_3$, lead is $4+$, and its s-states appear lower in energy than the nickel exchange splitting, resulting in a distinct electronic profile. This can be clearly seen in the projected density of states (Fig. 7.3(a)): the lowest conduction band is almost entirely Pb s and O p states, while the d-states only appear in the valence band and higher in the conduction manifold. While this serves to dramatically lower the band-gap, a further result of this is a Glass coefficient over an order of magnitude larger than the baseline. The current density is modest by comparison, though it still exceeds the baseline, indicating large shift vectors with relatively low absorption.


O (p)
Zn (s)
Zn (p)
-5
0
5
10
eV

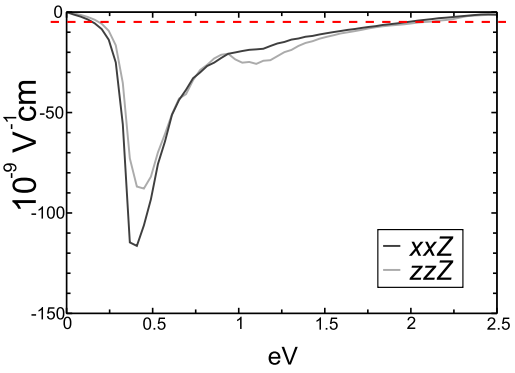
(a)


O (p)
Pb (s)
Pb (p)
-5
0
5
10
eV

(b)


0
-5
-10
-15
-20
-25
-30
10-4 V-1
xxZ
zzZ
0
0.5
1
1.5
2
2.5
eV

(c)


0
-50
-100
-150
10-9 V-1cm
xxZ
zzZ
0
0.5
1
1.5
2
2.5
eV

(d)

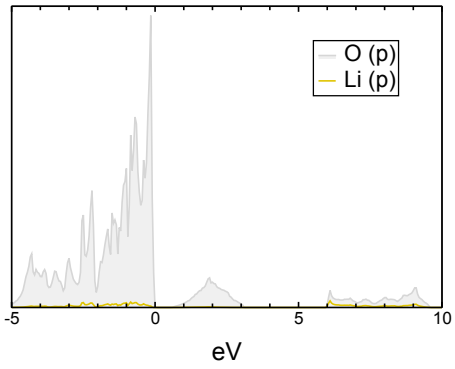
Figure 7.4: Relevant projected densities of states for $PbMg_{1/2}Zn_{1/2}O_3$ are shown in (a) and (b). The valence band is formed almost entirely from oxygen p-orbitals, and the conduction band is hybridized Pb s and O p-states. This results in a band-gap that is quite low, (c) high current density response, and (d) a very large Glass coefficient. Significantly, the response is antiparallel to material polarization.

$HgPbO_3$ [71] and $ZnSnO_3$ [72, 73] are known to take the ilmenite and $LiNbO_3$ structures, respectively, however, the first is metallic and the second has too high of a band-gap and only modest photovoltaic response. We first calculated the response of the structure $ZnPbO_3$, but found it to be borderline metallic, despite promising response; to raise the gap we substituted Mg for half of the Zn. Once again, as seen in Fig. 7.4(b), hybridized Pb s states compose the lowest unfilled band. The magnitude of the response is quite high, but interestingly, in the opposite direction of most materials, including our benchmark materials and the aforementioned $PbNiO_3$. We recall from Chapter 3 that in striped-domain BFO [4, 5] the BPVE response is in the opposite direction of the current generated by the polarization at the domain walls, partially canceling it. However, $PbMg_{1/2}Zn_{1/2}O_3$ in the same configuration would generate responses that add rather than cancel, potentially allowing for significant overall response. While the band-gap is actually lower than is preferable for solar energy collection, we emphasize that by tuning the composition, as done here in a coarse way, it may be possible to raise the band-gap without substantially diminishing the response.

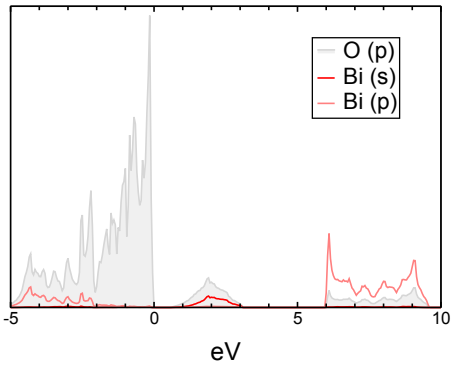
$BiLiO_3$ is known to exist in a structure with edge-sharing oxygen octahedra [74]. However, our calculations place the $LiNbO_3$-type structure nearby in energy, at only about 0.01eV per atom higher; additionally, $NaBiO_3$ is known to take the closely-related ilmenite structure [74]. In light of this we consider it highly possible that the proposed structure can be synthesized.

As shown in Fig. 7.5(a) and Fig. 7.5(b), the electronic structure is very similar to the previous two materials. As with Pb-containing compounds, the low-lying hybridized Bi s-states form the lowest unfilled bands, though the Bi s proportion is lower than that of Pb s in the aforementioned materials. Possibly as consequence, the dispersion of the conduction band is reduced compared to $PbNiO_3$ and $PbMg_{1/2}Zn_{1/2}O_3$ (Fig. 7.2), and the BPVE response is somewhat different: while the Glass coefficient is not as large as the two lead-

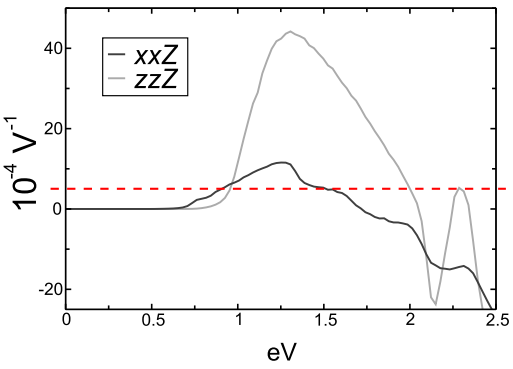
containing materials, the photocurrent density is higher, indicating increased absorption. Additionally, the band gap is larger, with GW predicting 1.7-1.8eV, positioned almost perfectly with respect to the visible spectrum.


O (p)
Li (p)
-5
0
5
10
eV

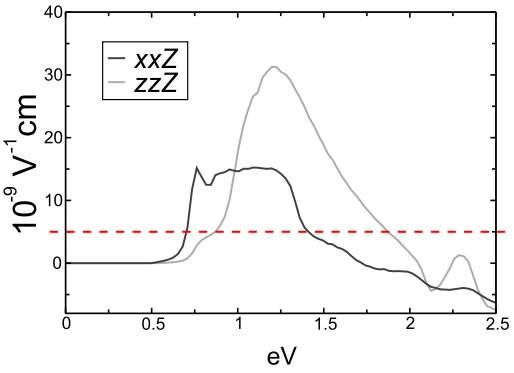
(a)


O (p)
Bi (s)
Bi (p)
-5
0
5
10
eV

(b)


xxZ
zzZ
10-4 V-1
40
20
0
-20
0
0.5
1
1.5
2
2.5
eV

(c)


xxZ
zzZ
10-9 V-1cm
40
30
20
10
0
0
0.5
1
1.5
2
2.5
eV

(d)

Figure 7.5: The density of states for $BiLiO_3$, shown in (a) and (b), is dominated by bismuth and oxygen (c). The band gap is dictated by transitions from O p to hybridized Bi s states. The band gap is modest (1.7eV in GW). The current density response, shown in (c) is quite high, with a somewhat high (d) Glass coefficient, indicating strong absorption in addition to long shift vectors.

	Band gap	Maximum Glass coefficient $\times 10^{-9}$cm/V	Maximum current density $\times 10^{-4}\mathrm{V}^{-1}$
$LiNbO_3$	3.7eV [68]	6	5
$BiFeO_3$	2.74eV [2]	5	5
$PbNiO_3$	1.2eV(HSE) [6]	80	13
$PbMg_{1/2}Zn_{1/2}O_3$	-	115	20
$BiLiO_3$	1.7eV(GW)	30	45

Table 7.1: The band gap and response characteristics of the presented materials, along with $LiNbO_3$ and $BiFeO_3$ for comparison.

In all three cases, the transitions occur between states dominated by cation s-states and oxygen cage p-states. If we view the oxygen cage as a single site, then along the axis of polarization these structures resemble a doubled cell of the tight-binding model of the last chapter. Folding the model band structure appropriately and superimposing it on the conduction band of $PbNiO_3$ (Fig.7.6), we see that the two match excellently for a significant amount of higher-order hopping. This provides strong evidence that the above model, especially the long-range hopping, captures the important physics governing the shift vector, in agreement with our empirical study, and the strength of the present results validates our materials design approach.

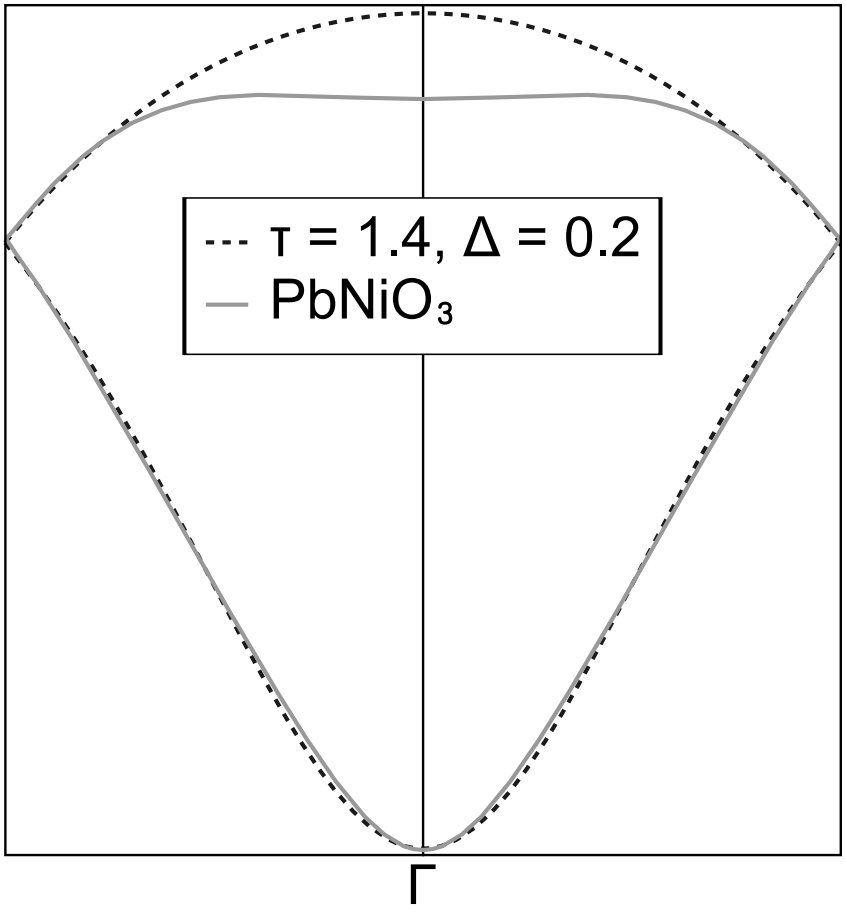

τ = 1.4, Δ = 0.2
PbNiO3
Γ

Figure 7.6: The model band structure folded back to reflect a supercell, with the *ab initio* calculated valence band of $PbNiO_3$. The value of τ represented is 1.4, reflecting a significant amount of higher-order-hopping.

We have presented several polar oxides in the $LiNbO_3$ structure with strong BPVE response and low band gaps, summarized in Table 7.1. One is already known to exist, and two we propose as targets for synthesis. The compositions, featuring Pb^{4+} or Bi^{5+}, were chosen for the absence of d-states at the band edge, and instead have conduction bands formed by low-lying s-states hybridized with oxygen p-states. In addition to creating significantly lower band gaps, this makes for large, diffuse orbitals and results in strongly delocalized states; combined with large polar distortions, they effect significant shift current response that is over an order of magnitude higher than that previously known. Given the minimal contributions from the other cations, the possibility of tuning the response via composition without altering its fundamental character is strongly suggested. Additionally, one material, $PbMg_{1/2}Zn_{1/2}O_3$, has response anti-parallel to material polarization, making it an excellent candidate for use in the structure described in Refs [4] and [5].

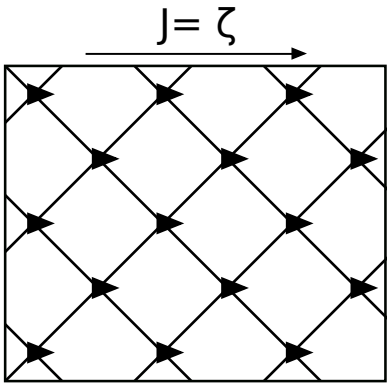
Chapter 8

Spin Bulk Photovoltaic Effect

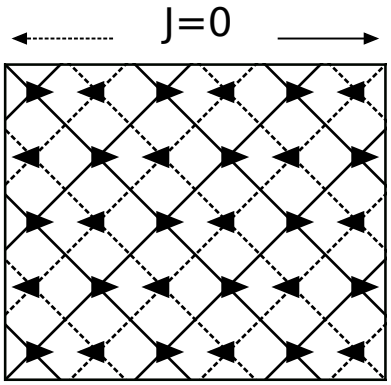
Spintronics - the use of electronic devices relying on the manipulation of spin rather than charge - promises to play an important role in the development of future electronic and computing devices [75]. However, precise control of electron spin, including the generation of spin-filtered currents, presents a difficult challenge. There are four main mechanisms for spin current generation currently known: spin-Hall effects [76, 77, 78], illumination with circularly polarized light [15, 79, 80, 81], subband splitting due to spin-orbit coupling [82, 83, 84, 85, 86], and, recently, the spin-Seebeck effect [87]. While pure spin current generation has been achieved using linearly polarized light, the subband splitting created by spin-orbit effects is required, along with strong inversion symmetry breaking, which constrains the strength of the response. In this chapter we add a new mechanism: spin separation in antiferromagnets by linearly polarized light. Neither spin-orbit coupling nor inversion symmetry breaking is required, making entirely distinct classes of materials candidates for application. The contents of this chapter were published as Ref. [88]

8.1 Symmetry Constraints

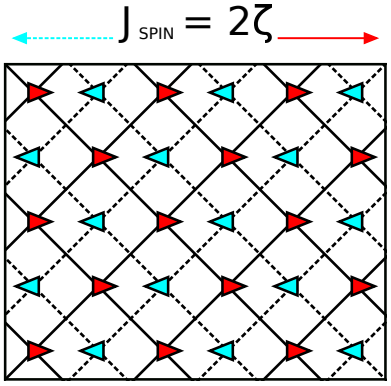
So far, we have only considered the charge current generated by the shift current mechanism. One can consider up and down spin electrons separately, but in the presence of time-reversal symmetry and negligible spin-orbit interaction, these are required to respond identically, and only charge currents are generated. However, when antiferromagnetic materials are considered, a new possibility emerges. The spin centers may produce opposite responses to the illumination, generating a net charge current of zero, and a net spin current.



(a)



(b)



(c)

Figure 8.1: A non-centrosymmetric lattice, like the one shown in (a), will generally exhibit the bulk photovoltaic effect. When a copy of the lattice related by mirror symmetry is added, shown in (b), the total current will be zero. However, if the two sublattices have opposite spin, represented dichromatically in (c), a pure spin current will result.

This is illustrated by the two-dimensional toy system in Fig. 8.1. Shown in (a) is a square lattice decorated by triangles. The lattice breaks inversion symmetry, and in general will produce a bulk photovoltaic response. However, suppose we add as a sublattice a duplicate of the original lattice, related to it by a symmetry operation. In (b) this is shown for a mirror symmetry. The additional sublattice will produce a bulk photovoltaic response that is the mirror of the response of the first lattice, canceling it. If, however, we turn on opposite spins for the two sublattices, as indicated by the coloring in (c), the currents produced by the two lattices will have opposite spin, resulting in pure spin current.

The procedure for determining the crystal classes that allow for this effect is similar to that for the charge bulk photovoltaic effect; however, the Shubnikov group – specifically, the black-and-white, or dichromatic group [89] – must be used instead of the space group. Shubnikov groups consist of the space group operations, a subset of which are multiplied by an additional operation of antisymmetry. It is important to note that these are distinct from double groups. The unit cell is divided into sections of two types, often denoted as "black" or "white", which interchange upon application of antisymmetry. In this case, our black/white are spin up/down, so the antisymmetry operation can be identified with time reversal. As seen in Fig. 8.1 above, the crystal may be antisymmetric under a given symmetry operation (*e.g.* inversion), but if the time reversal operator is applied, the combined operation is a member of the symmetry group. Formally,

$$M = H + \theta(G - H)$$

Where M is the magnetic group, θ is the time reversal operation, G is the space group of the lattice, and H is the invariant subgroup of G that respects spin symmetry.

Each magnetic group has a principal representation analogous to the operation possessing the full symmetry of the crystal when magnetic ordering is excluded. Only tensor

elements or linear combinations thereof that belong to this principal representation are allowed to be nonzero. For a third-rank tensor, this requires that the representation generated by taking the cube of the vector representation contain the principal representation.

Since the symmetry of a tensor is dependent only on a space group's isogonal point group, we restrict our analysis to the point groups. The magnetic groups that derive from a given point group can be determined from the parent point group's character table: for each invariant subgroup H there is a one-dimensional representation that has positive character for the operations in H only and becomes the principal representation of the magnetic group. The character tables for these child magnetic groups can be determined, but since we are only interested in the principal representation, we need only the monochromatic group tables to identify the representation associated with reduction of symmetry to H. However, one additional consideration must be made: the magnetic group must also be able to host antiferromagnetism. In some cases, the magnetic point group will not admit antiferromagnetism, but a non-symmorphic space group for which the point group is isogonal can. Using this we can identify all the dichromatic groups that allow the spin photovoltaic effect. Further analysis can reveal which tensor elements belong to the principal representation. Fortunately, this has already been performed for the piezomagnetic effect, which has identical symmetry properties [89].

We propose that these spin currents will be generated by the shift current mechanism. It is important to remember that the phenomenon is distinct from other photovoltaic effects; rather than excited carriers being split by an electric field, current is produced by coherent excitations that have themselves a non-zero net momentum. This momentum is a function of the reciprocal lattice vector, and therefore must reflect the symmetry of the Brillouin zone. Thus, while the preceding symmetry argument demonstrates that a spin photovoltaic effect may exist in principle, the unique properties of the shift current suggest it as a mechanism by which such an effect can physically manifest.

In the case of a spin-polarized system, the calculation is performed for spin up and spin down bands separately, so that

$$\begin{aligned}
\sigma^{\mathrm{S}}_{rsq}(\omega) =& \sigma_{rsq,\uparrow}(\omega) - \sigma_{rsq,\downarrow}(\omega) \\
\sigma_{rsq,\uparrow/\downarrow}(\omega) =& e\pi \left(\frac{e}{m_e \hbar \omega}\right)^2 \sum_{n',n''} \int d^3\mathbf{k} \, \left\langle n'_{\uparrow/\downarrow}\mathbf{k}\right| \hat{p}_r \left|n''_{\uparrow/\downarrow}\mathbf{k}\right\rangle \left\langle n''_{\uparrow/\downarrow}\mathbf{k}\right| \hat{p}_s \left|n'_{\uparrow/\downarrow}\mathbf{k}\right\rangle \\
& \times \left[f_{n'_{\uparrow/\downarrow}}(\mathbf{k}) - f_{n''_{\uparrow/\downarrow}}(\mathbf{k})\right] \delta\left(\pm\omega - \omega_{n'_{\uparrow/\downarrow}}(\mathbf{k}) + \omega_{n''_{\uparrow/\downarrow}}(\mathbf{k})\right) \\
& \times \left[\frac{\partial \phi_{n''_{\uparrow/\downarrow} n'_{\uparrow/\downarrow}}(\mathbf{k},\mathbf{k})}{\partial k_q} - \chi_{n''_{\uparrow/\downarrow} q}(\mathbf{k}) + \chi_{n'_{\uparrow/\downarrow} q}(\mathbf{k})\right]
\end{aligned} \tag{8.1}$$

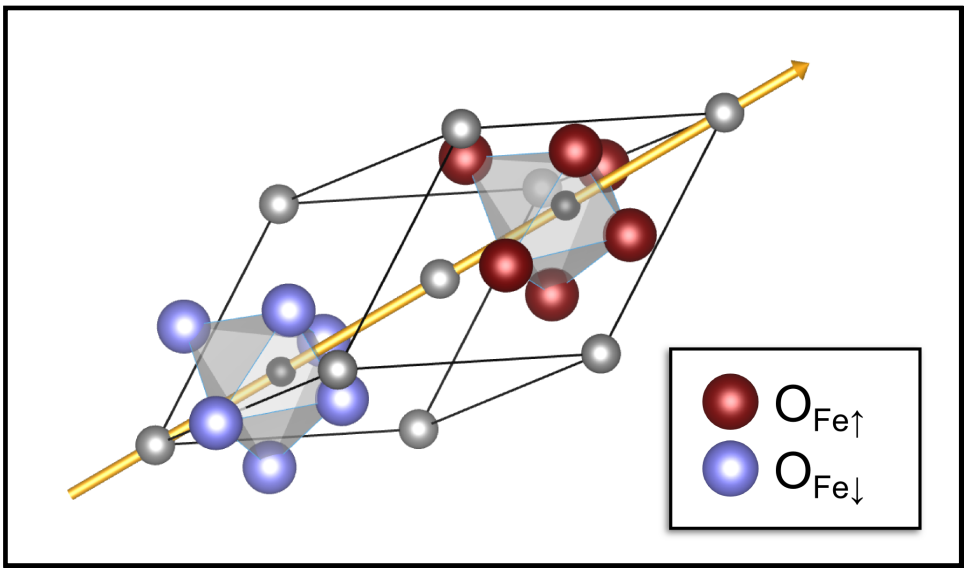
It is evident that the symmetry effects above are introduced through the intrinsic symmetry of the supplied electronic states, so that Eq. (8.1) is general; with the addition of time reversal symmetry it reduces to Eq. (4.4).

8.2 First-principles calculations of BFO and hematite

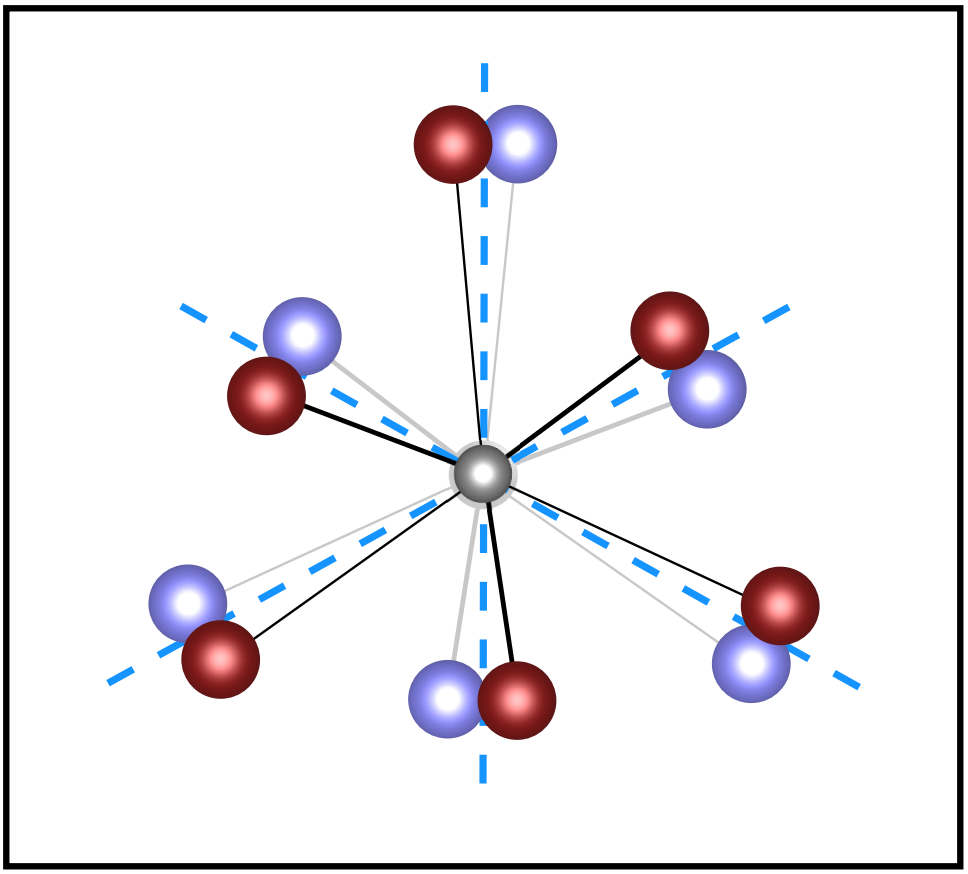
We have computed the spin photovoltaic response for the well-known antiferromagnets NiO, Fe_2O_3 (hematite), and the multiferroic $BiFeO_3$ (BFO). The wavefunctions used for the response calculations were generated using the Quantum ESPRESSO package [69]. Due to the well-known inability of DFT to model Mott insulator systems correctly, Hubbard U terms were added for hematite [90] and BFO [91]. Charge densities were generated on $8\times8\times8$ k-point grids and used to generate wavefunctions on finer grids as necessary.

The magnetic group for NiO derives from the $A_{2\mathrm{u}}$ representation of point group O_h. There are no third rank tensor elements that belong to this representation, so the crystal will have no spin bulk photovoltaic effect. Calculations were performed and confirm the absence of any response.

Hematite [90] has space group 167, with point group $D_{3\mathrm{d}}$, while BFO has space group 161, with point group $C_{3\mathrm{v}}$. The two materials both take the ilmenite structure, with BFO, shown in Fig. 8.2(a), experiencing a ferroelectric distortion. It is worth noting that inversion symmetry will kill any charge bulk photovoltaic effect in hematite, whereas BFO has been demonstrated to have a large bulk photovoltaic effect [3, 91]. In both cases the magnetic group is associated with the reduction to C_3 symmetry, deriving from the representations $A_{2\mathrm{g}}$(hematite) and A_2(BFO), so that a glide plane relates the up and down spins. As is evident in Fig. 8.2(b), which shows the oxygen cages viewed along the material polarization direction, the environments of these two spin centers differ by the direction of distortion of the coordinating oxygen atoms, converting what would otherwise be a mirror symmetry to a glide plane, and introducing a chirality into the structure. This is crucial, as it ensures that flipping the spins switches chirality, allowing a spin current to exist.



(a)



(b)

Figure 8.2: (a) shows the primitive unit cell for BFO, with the oxygen cages colored according to the spin of the iron atoms they enclose. Hematite takes a very similar structure, with iron in place of bismuth and no ferroelectric distortion. (b) shows the oxygen cages viewed along the polarization direction. The mirror components of the glide planes are shown by the blue dashed lines. From this view it is clear that reversing the distortion of the oxygen cages has the same effect as inverting the spins; the current generated under one oxygen cage distortion is the mirror of that generated by the opposite distortion, leading to spin current along the X axis. There may also be charge current in other directions depending on the symmetry, as in BFO.

We note that bismuth ferrite possesses significant spin-orbit coupling which introduces spin canting and weak ferromagnetism. While the photovoltaic response calculation can be performed with the full spinorial wavefunctions without much difficulty, in the presence of large spin-orbit interaction the result no longer conforms to a rigorous definition of spin current [92]. However, in the present context the effect is relatively small, so for our calculation we impose antiferromagnetic ordering, and compute the spin current for this approximation to the spin structure.

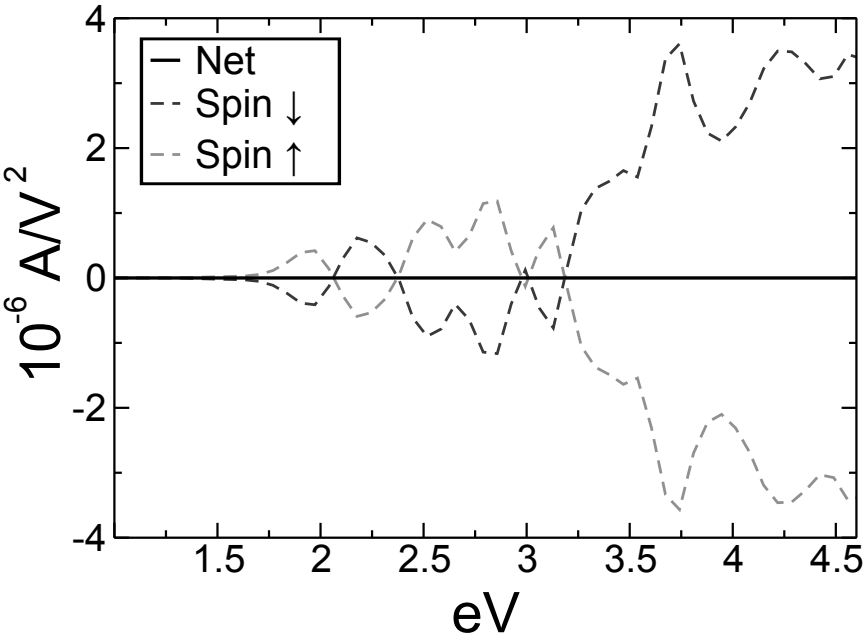
Tensor elements that are antisymmetric under the glide plane operation survive, and are

$$\sigma_{\text{hematite}} = \begin{bmatrix} \sigma_{11}^{\text{S}} & -\sigma_{11}^{\text{S}} & 0 & \sigma_{41}^{\text{S}} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sigma_{41}^{\text{S}} & -\sigma_{11}^{\text{S}} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

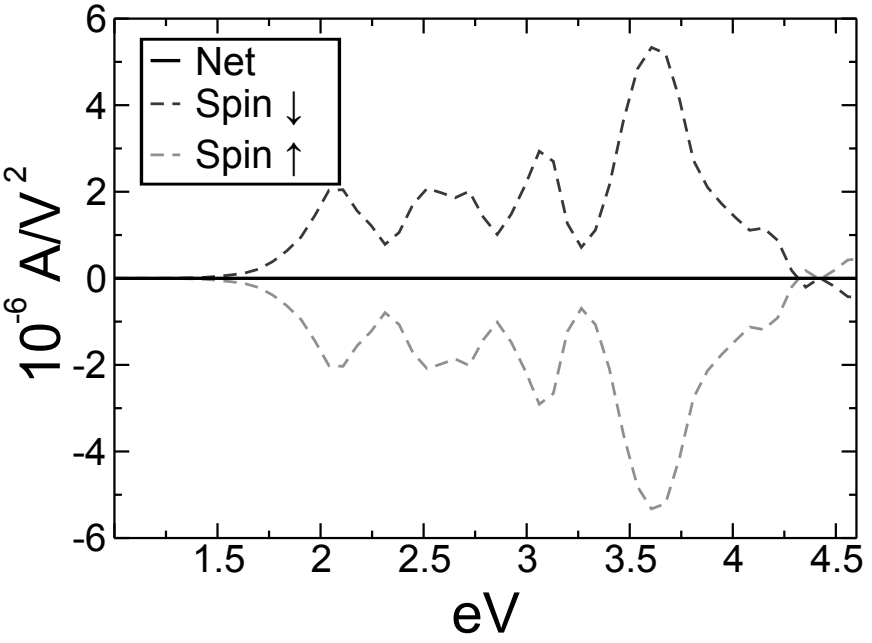
for hematite, and

$$\sigma_{\text{BFO}} = \begin{bmatrix} \sigma_{11}^{\text{S}} & -\sigma_{11}^{\text{S}} & 0 & \sigma_{41}^{\text{S}} & \sigma_{52} & -\sigma_{22} \\ -\sigma_{22} & \sigma_{22} & 0 & \sigma_{52} & -\sigma_{41}^{\text{S}} & -\sigma_{11}^{\text{S}} \\ \sigma_{13} & \sigma_{13} & \sigma_{33} & 0 & 0 & 0 \end{bmatrix}$$

for BFO, with charge photovoltaic response elements included for completeness.

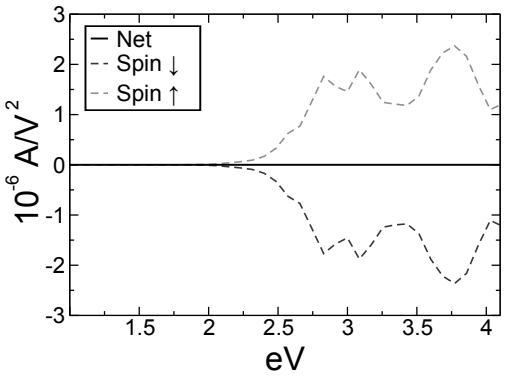

Net
Spin ↓
Spin ↑
10^-6 A/V^2
eV
4
2
0
-2
-4
1.5
2
2.5
3
3.5
4
4.5

(a)

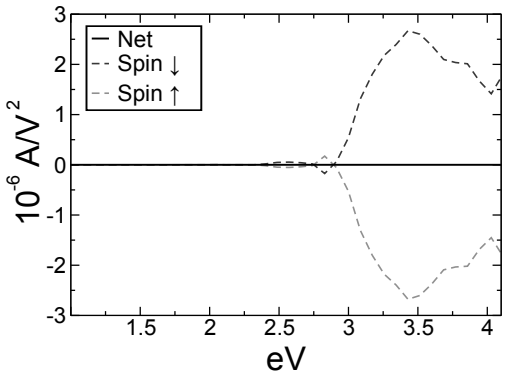

Net
Spin ↓
Spin ↑
10^-6 A/V^2
eV
6
4
2
0
-2
-4
-6
1.5
2
2.5
3
3.5
4
4.5

(b)

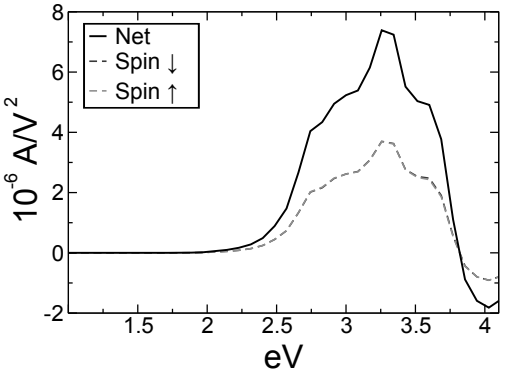
Figure 8.3: (a) displays the spin and charge current spectra for hematite in direction xxX (σ^{S}_{11}) and (b) shows the spectra in zxY (σ^{S}_{14}). The total charge currents vanish in all directions for hematite.


Net
Spin ↓
Spin ↑
10^-6 A/V^2
eV

(a)


Net
Spin ↓
Spin ↑
10^-6 A/V^2
eV

(b)


Net
Spin ↓
Spin ↑
10^-6 A/V^2
eV

(c)

Figure 8.4: Spin and charge photovoltaic tensor elements for $BiFeO_3$ in the xxX direction (σ^{S}_{11}) and the zxY direction (σ^{S}_{14}) are shown in (a) and (b). Compared with them is the charge current in yyY direction (σ_{22}).

The spectra for the unique elements are shown for hematite in Fig. 8.3, and for BFO in Fig. 8.4, with the charge photovoltaic response for comparison. The spin response for both materials is of a similar magnitude to the charge response of BFO, indicating that it should be easily observable.

We consider hematite to be the preferred material for measuring the spin bulk photovoltaic effect, as it cannot produce charge photocurrents, is uncomplicated by spin-orbit effects, has a lower band-gap, and is more readily available than BFO.

Part II

Computational Design of Materials at or near Topological Phase Transitions

Chapter 9

Background: Topological Phases

Here we provide a brief summary of the classification of topological phases, followed by a discussion of topological phase transitions.

9.1 Classification of Topological phases

The integer quantum hall effect can be understood theoretically by considering a two-dimensional system in the strong magnetic field limit as described by Landau [28, 29]. The bulk system is then described by degenerate sets of bands called Landau levels; the degeneracy and separation of which is determined by the strength of the field. It is within the bulk gaps that the quantized hall voltage and dissipationless edge currents appear.

As mentioned previously, the response of these states depends on a term which must take an integer value. In particular, it can be shown that [28]

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{2\pi i}\sum_n \int\int dk_x dk_y \left\{\nabla_{\mathbf{k}} \times \chi_n(\mathbf{k})\right\}_z = \frac{e^2}{h}\frac{1}{2\pi i}\oint d\mathbf{k}\Re\chi_{n;z}(\mathbf{k}) \tag{9.1}$$

where χ_n is the Berry connection of the nth band. Eq. (9.1) is known as the Chern number, and is a topological invariant of the $U(1)$ space in which the quantum hall states live [30].

From Stokes theorem, we can see that the Chern number is only nonzero when the "flow" of the Berry connection over the magnetic BZ is rotational; i.e., it cannot be a global single-valued analytic function. It can be shown that these obstructions are related to those existing in Landau levels above and below. At the edge, these obstructions must be repaired, manifesting as gapless states running between Landau levels [30].

The topological insulating state is not described by a nonzero Chern number; however, similar physics obtains. We follow closely the description appearing in Ref. [93], though the notation is slightly different. The relevant topological number – the Z_2 invariant – may be described by a "time-reversal" polarization, which is just the difference in the Berry phase polarization of two Kramers degenerate bands. We consider a pair of such bands (in one dimension for now), related by time-reversal as

$$|1,-k\rangle = -\mathrm{e}^{i\theta(k)}\Theta\,|2,k\rangle$$
$$|2,-k\rangle = \mathrm{e}^{i\theta(-k)}\Theta\,|1,k\rangle$$

where Θ is the time-reversal operator. It is important to note that time-reversal symmetry is not fully enforced at this point, due to the presence of the phase relation $\theta(k)$. The time-reversal polarization is then

$$P = -\frac{1}{2\pi}\int_0^\pi dk \nabla_k \left[-\theta(-k)+\theta(k)\right] \mod 2$$

It can be shown that this expression is gauge invariant. It is evident that $P = 0$ for $\theta(k) = 0$. However, if we choose any continuous function θ that winds around 2π, then $P = 1$. Now, suppose we choose the following shape for $\theta(k)$, which can be smoothly deformed to or

from any choice with the same P:

$$\theta(k) = \begin{cases} -2k & -\pi < k \leq 0 \\ 0 & 0 < k \leq \pi \end{cases}$$

this clearly has $P = 1$. However, we now slowly turn on the full time-reversal symmetry constraint

$$|1, -k\rangle = -\Theta\,|2, k\rangle$$
$$|2, -k\rangle = \Theta\,|1, k\rangle$$

Now, $\theta(k)$ must be a multiple of 2π everywhere in the BZ. As shown in Fig. 9.1, this cannot be achieved while preserving P without making $\theta(k)$ discontinuous. As an aside, we also note that for any gauge transformation, which may wind by a multiple of 2π, that is applied to one of our wavefunctions must now also introduce a gauge transformation in its counterpart, restricting the winding of θ to a multiple of 4π. This means that P cannot be changed by any such gauge transformation, and that the two possible values of P are distinct.

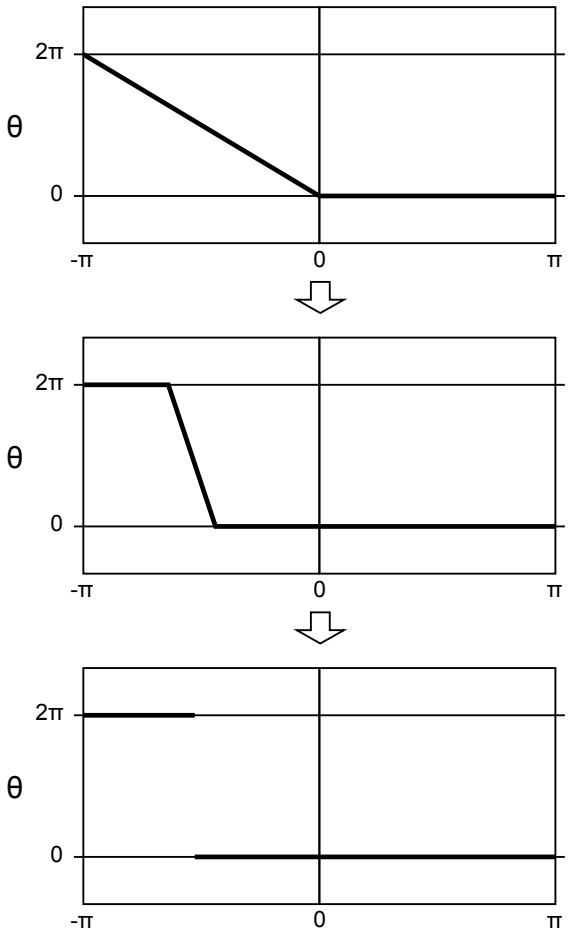

2π
θ
0
-π
0
π
2π
θ
0
-π
0
π
2π
θ
0
-π
0
π

Figure 9.1: The phase relating two Kramers degenerate bands is shown evolving as it is adiabatically transformed to obey the constraint from time-reversal symmetry that the phase be an integer multiple of 2π. As seen, if the phase winds by 2π when going around the BZ, there must be a gauge discontinuity when the wavefunctions are fully time-reversal symmetric.

We may construct a Z_2 invariant by extending our system to two dimensions, and considering P for the two pairs of paths of the kind above that now exist between time-reversal invariant coordinates. Letting our previous coordinates be along an x-axis, and our new dimension be the y-axis

$$Z_2 = (-1)^{\left[P_{k_y=0}+P_{k_y=\pi}\right]}$$

$Z_2 = -1$ marks a topological insulator. We note that if both time-reversal polarizations are 1, they may annihilate one another via gauge transformation. Thus, a nontrivial Z_2 index results from time-reversal symmetry obstructing a globally-defined gauge, in a similar fashion to the Chern number.

Since a two-dimensional material has four time-reversal symmetric points in its BZ, it has one Z_2 invariant. A three-dimensional material has eight such points, and is characterized by four Z_2 invariants: three can be obtained from the three planes of four points (excluding Γ) that may be constructed, and the fourth using all eight points [7]. Finally, we note that there are additional ways of understanding the topological insulating state and the Z_2 invariant that are not discussed here. Some of these can be found in Refs. [94, 34, 95, 96].

A third topological state that we consider is a Weyl point [97]. This is characterized by a single point degeneracy of two spinful bands, with linear dispersion in all three directions. Near this degeneracy it may be described by the Hamiltonian

$$\mathcal{H} = v_{ij}\sigma_i k_j \tag{9.2}$$

where σ_i are the Pauli matrices. We note that it spans all three Pauli matrices, such that any additional term will simply shift the degeneracy, rather than break it. This property arises because the degeneracy constitutes a discontinuity that prevents a single-valued gauge from

being defined for either band. As such, it may be characterized by a Chern number, where in this case the path of Eq. (9.1) becomes a surface enclosing the degeneracy.

9.2 Topological Phase Transition

We now consider the topological insulator transition. A topological insulator is created by a spin-orbit-driven exchange of states between appropriate valence and conduction bands in a portion of the BZ. This introduces the obstruction to a global gauge as mentioned above. Since the gap must close for this to happen (it is an adiabatic process), the transition between the topological and trivial insulating phases must be marked by a metallic state. In the simplest case, where the system is inversion-symmetric, this will appear as a single Dirac point. Near the phase transition, one may construct a low energy theory that takes the form of a Dirac Hamiltonian

$$\mathcal{H} = v_{ij}\gamma_i k_j + m\gamma_m \tag{9.3}$$

We note that this is a four-band model where the γ_i are three mutually anti-commuting Dirac matrices. For a mass m of zero, it is clear that at $\mathbf{k} = 0$ the system will be characterized by a four-fold degeneracy, with linear dispersion in all directions. Unlike the two-band Weyl Hamiltonian, however, there are γ_m that can gap the system. In particular, there will be two additional Dirac matrices that anticommute with γ_i. Thus, for such an appropriate Dirac matrix γ_m, a nonzero mass-term m will introduce a gap; however, the sign of m determines which states become the conduction or valence band. Therefore, for a proper γ_m, there is a topological insulator transition that is characterized by a change in sign of the mass and concomitant inversion of the conduction and valence bands. At the interface of a topological and trivial insulator (of which vacuum is an example), the mass necessarily

changes sign, and at some point there must be a Dirac-like state where the bands invert and the obstruction a single-valued phase is removed. It is worth emphasizing that, as with the integer quantum Hall effect, these surface states follow from the bulk electronic structure, and cannot be gapped without destroying the bulk state, either by allowing the value of the mass term in the bulk to change sign, or by removing the protection of time-reversal symmetry [7].

As an aside, it is worth considering what happens if we introduce a time-reversal-symmetry breaking term into Eq. (9.3). Such a term will necessarily commute with at least one of the γ_i and cannot totally gap this system; instead, when m is small compared to the strength of time-reversal violation, the Dirac point splits into (at least) two Weyl points at some $\mathbf{k}$ and $-\mathbf{k}$. Since there are no further Dirac matrices that anticommute with all of those present in the augmented Hamiltonian, the system cannot be perturbatively gapped, reflecting the nontrivial topology of the Weyl points. The system will only gap when m grows large enough; as the mass increases the Weyl points approach one another and annihilate, allowing a gap to open.

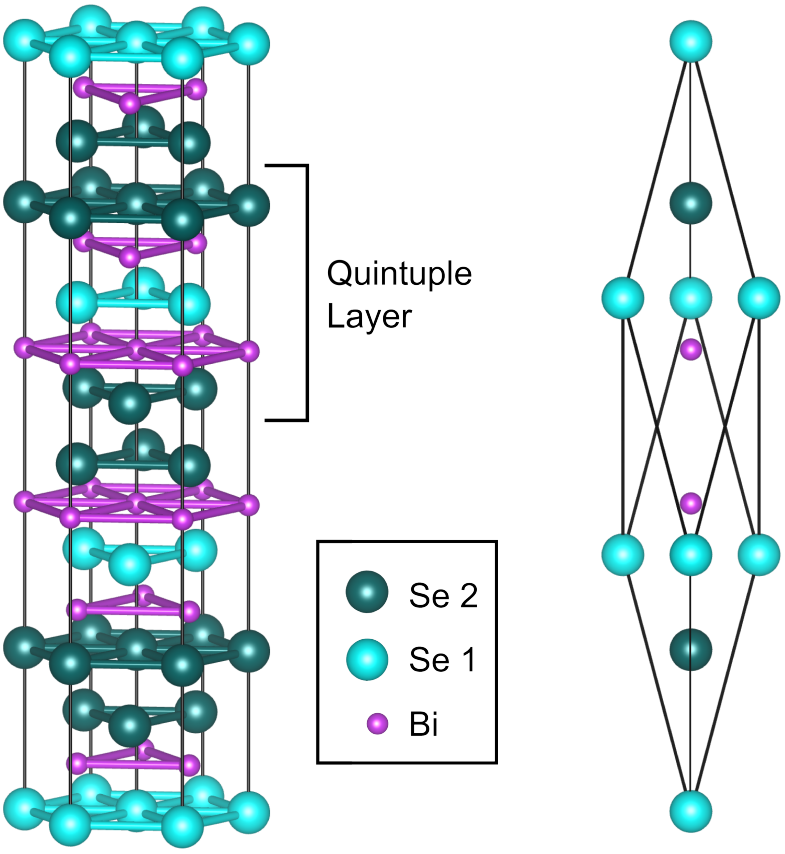
In what follows we explore systems at or near the type of bulk transition where time-reversal symmetry is preserved. We may use strain as an external parameter for tuning the effective mass term, and we investigate the conditions under which the transition state having effective mass zero can be protected by crystallographic symmetry.

Chapter 10

The Effects of Strain on the Topological Gap

10.1 Bismuth Selenide

In this chapter, we investigate the use of strain as a means of tuning the topological insulating phase of bismuth selenide. Though not without flaws, Bi_2Se_3 is relatively chemically stable, easy to synthesize, and exhibits a robust topological phase. Combined with the existing theoretical and experimental studies, [41, 42, 98, 99, 100, 101, 102, 103, 104] it has emerged as the prototypical topological insulator. Using *ab initio* methods, we have evaluated the elastic properties of Bi_2Se_3 and their connection to the properties of its topological insulating phase; the direct band-gap at the Γ point, where band inversion occurs, responds to elastic deformation in a way that can be described by adapting the formalism of continuum mechanics. The critical strain at which the topological phase transition occurs was predicted using the derived band-gap stress and stiffness tensors and observed in the computed band structure as a Dirac cone. The content of this section appears in Ref. [105]



(a) (b)

Figure 10.1: The (a) crystal structure of Bi_2Se_3 consists of hexagonal planes of Bi and Se stacked on top of each other along the z-direction. A quintuple layer with Se1-Bi-Se2-Bi-Se1 is indicated by the square bracket, where (1) and (2) refer to different lattice positions. The (b) primitive cell of Bi_2Se_3 is rhombohedral.

The bulk crystal structure of Bi_2Se_3 is rhombohedral with space group D^5_{3d} ($R\bar{3}m$, No. 166) [106], shown in Fig. 10.1. The primitive unit cell has two Bi and three Se atoms, and the atomic plane arrangement is Se(1)-Bi-Se(2)-Bi-Se(1), where Se(1) and Se(2) indicate the two different types of selenium atom in the crystal. In the hexagonal supercell, the structure can be described as quintuple layers (QL) (square region in Fig. 10.1 of atoms stacked along the trigonal axis (three-fold rotational axis).

DFT calculations were performed using Quantum-Espresso [69]. The lattice parameters were taken from experiments (a = 4.138 Å and c = 28.64 Å) [106]. After fixing the lattice parameters to their experimental values, the atomic coordinates were relaxed to generate the reference structure. It should be noted that the calculations give significant, nonzero stress for the crystal in this geometry.

Various strains were applied relative to the reference structure, including positive and negative uniaxial and shear strains up to 2%, as well as several combinations thereof. The atomic lattice coordinates were relaxed for each strain configuration, and the total and band-gap energies were computed both with and without spin-orbit coupling. Multiple regression analysis was performed to find the linear and quadratic dependence of the energy on strain tensor components. This yielded the elastic stiffness and stress tensors.

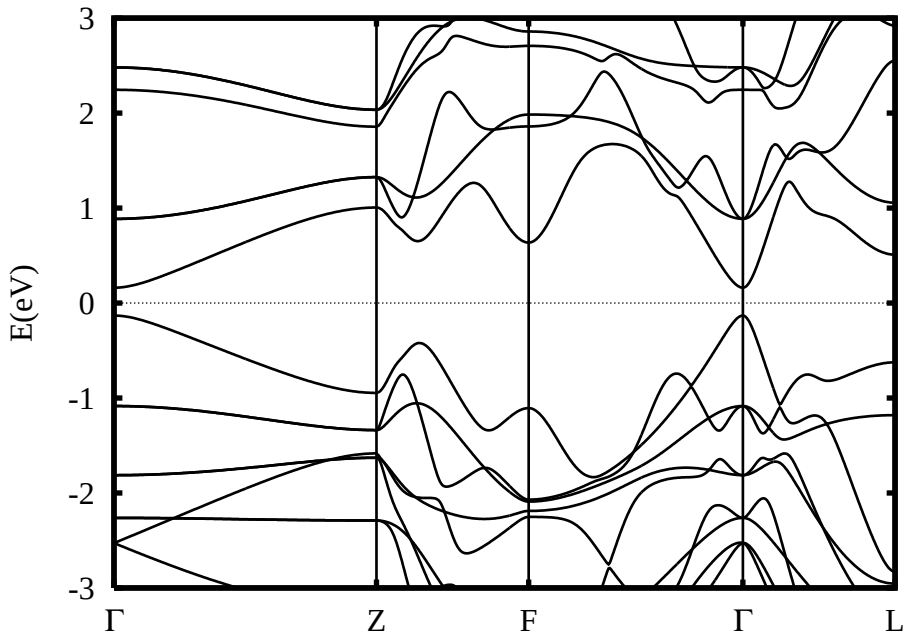
It is known that in bismuth selenide the topological index distinguishing ordinary insulating from topological insulating behavior is controlled by band inversion at the Γ point. Thus, a band-gap stress σ^Γ and band-gap stiffness c^Γ were defined as the linear and quadratic coefficients relating the Γ point band-gap to strain, by the same procudure used to determine the elastic tensors.

$$\Delta E_g^\Gamma(\epsilon) \quad = \tfrac{1}{2} c^\Gamma_{ijkl}\epsilon_{ij}\epsilon_{kl} + \sigma^\Gamma_{ij}(0)\epsilon_{ij} \tag{10.1}$$

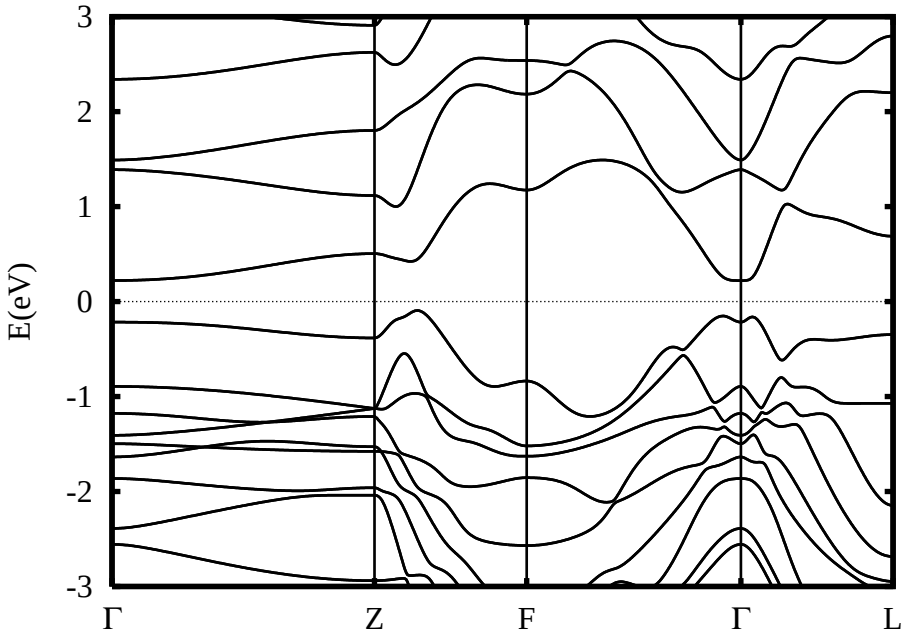
In both cases, only the tensor elements unique under the symmetry operations of the space group of bismuth selenide ($R\bar{3}m$, No. 166) were allowed as degrees of freedom. The stiffness and stress tensors, in Voigt notation, must have the forms

$$c = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & 0 & 0 \\ c_{12} & c_{11} & c_{13} & -c_{14} & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ c_{14} & -c_{14} & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & c_{14} \\ 0 & 0 & 0 & 0 & c_{14} & c_{66} \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_1 \\ \sigma_1 \\ \sigma_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{10.2}$$

The band structure of Bi_2Se_3 and related compounds have been theoretically predicted [42, 43, 107, 108] and experimentally observed [40, 41]. In our calculations, the band-gap of the unrelaxed, experimental structure of Bi_2Se_3 is 0.3 eV, which is consistent with the experimental data and other calculations. [40, 98, 99] Fig. 10.2 shows the band structure with and without spin-orbit interaction, and both are in excellent agreement with previous results. [43]



(a)



(b)

Figure 10.2: Band structure in the reference strain state of Bi_2Se_3 (a) excluding spin-orbit effect (NSO) and (b) including spin-orbit effects (SO). The dashed line indicates the Fermi level.

10.2 Band-gap response to strain

The computed elastic and band-gap tensor components are given in Table 10.1 and Table 10.2, respectively. First, we note that the gap stress shown in Table 10.2 for strain normal to the plane of the quintuple layers(σ_3^Γ) is much higher than for strain in the plane(σ_1^Γ), which is consistent with the notion that inter-layer interactions are more important in determining the band-gap than intra-layer interactions. Second, we observe the change in sign of σ^Γ when spin-orbit interactions are turned off or on. This is to be expected because the spin-orbit interaction leads to an inversion of the conduction and valence bands. Thus, strain reduces the bandgap for the trivial (un-inverted) phase and increases the gap for the topological (inverted) phase. Third, the magnitude of the gap stress is larger when spin-orbit interactions are present. From a tight-binding perspective, compressive strain not only strengthens the Coulombic interaction between sites, increasing the associated hopping coefficient and reducing the conventional gap, but also magnifies the spin-orbit effect and its hopping coefficient, increasing the topological gap. Thus, comparing the gap stress with and without spin-orbit interactions provides some insight into the effects of strain on the essential physics of the system.

Element	Coefficient (GPa)
c_{11}	91.8 $\pm$ 1.0
c_{33}	57.4 $\pm$ 1.4
c_{44}	45.8 $\pm$ 1.0
c_{66}	56.2 $\pm$ 1.8
c_{12}	36.6 $\pm$ 1.2
c_{13}	38.6 $\pm$ 2.0
c_{14}	24.2 $\pm$ 1.8
σ_1	-3.447 $\pm$ 0.007
σ_3	-1.977 $\pm$ 0.010

Table 10.1: The unique elements of the elastic stiffness and stress tensors of Bi_2Se_3. Spin-orbit coupling has been included.

Element	NSO Coefficient (eV)	SO Coefficient (eV)
c_{11}^{Γ}	-67.8 $\pm$ 2.6	35.4 $\pm$ 5.6
c_{33}^{Γ}	55.4 $\pm$ 3.6	-60.1 $\pm$ 7.6
c_{44}^{Γ}	-58.0 $\pm$ 2.6	23.4 $\pm$ 5.6
c_{66}^{Γ}	-126.6 $\pm$ 4.4	69.4 $\pm$ 9.6
c_{12}^{Γ}	60.2 $\pm$ 3.2	-33.8 $\pm$ 6.8
c_{13}^{Γ}	6.4 $\pm$ 4.2	-12.6 $\pm$ 10.8
c_{14}^{Γ}	-70.0 $\pm$ 4.2	45.6 $\pm$ 9.0
σ_{1}^{Γ}	0.16 $\pm$ 0.017	-1.67 $\pm$ 0.037
σ_{3}^{Γ}	4.33 $\pm$ 0.023	-5.27 $\pm$ 0.051

Table 10.2: The unique elements of the Γ band-gap stiffness and stress tensors excluding (NSO) or including (SO) spin-orbit coupling

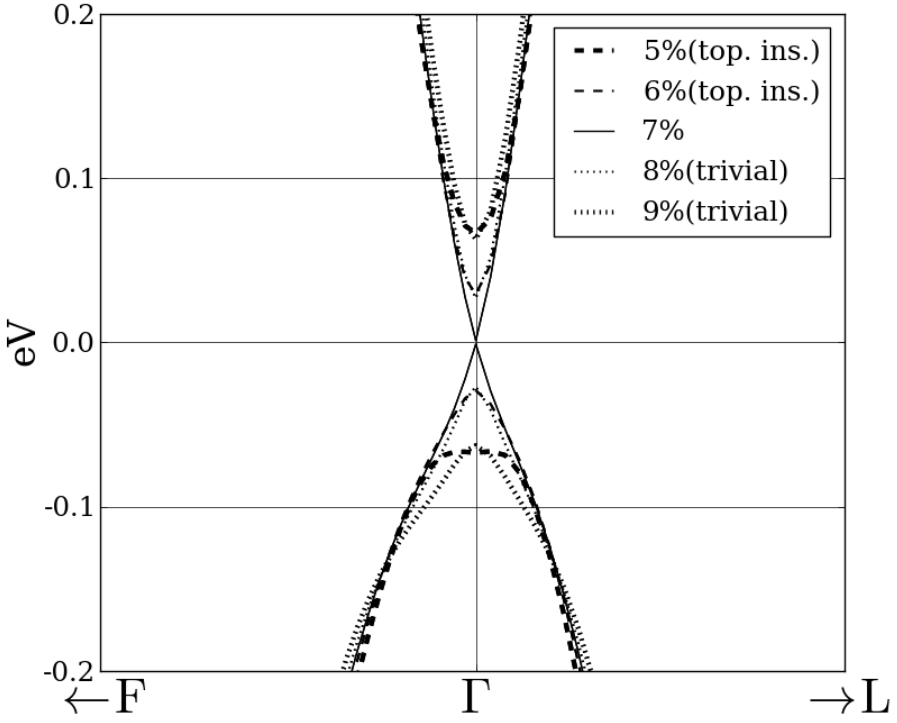

0.2
0.1
0.0
-0.1
-0.2
eV
←F
Γ
→L
5%(top. ins.)
6%(top. ins.)
7%
8%(trivial)
9%(trivial)

Figure 10.3: Band structure of Bi_2Se_3 near the Γ point as $\langle 111 \rangle$ uniaxial strain from 5% to 9% drives the topological phase transition.

Using the above tensors, we can predict that the topological phase transition will occur at 6.4% uniaxial strain in the $\langle 111 \rangle$ direction. In Fig. 10.3, the onset of the topological insulating phase at 7% strain can be observed through changes in the band structure as strain increases. At the transition point, the Dirac cone characteristic of the phase transition is distinctly observable. Of course, such strains are difficult to achieve experimentally. According to the computed elastic tensors, around 2 GPa of uniaxial tensile stress would be required to drive the phase transition, well past the yield stress. However, large strains may be possible by introducing internal stress through chemical substitution.

	Bi_2Se_3(reference)	Bi_2Se_3(strained)	Bi_2Te_3(reference)
Lattice parameters(Å)	a = 4.138 , c = 28.64	a = 4.358 c = 30.46	a = 4.358 c = 30.46
Anion radius (Å)	1.98	1.98	2.21
NSO Gap (eV)	0.02	0.31	0.20
SO Gap (eV)	0.42	-0.06	0.63

Table 10.3: Comparison of reference Bi_2Se_3, Bi_2Se_3 strained to match reference Bi_2Se_3 lattice, and reference Bi_2Te_3. Spin-Orbit(SO) and Non-Spin-Orbit(NSO) Γ-point gaps for all three structures are calculated.

Bi_2Te_3 is a very similar compound to Bi_2Se_3, differing only in substitution of the larger tellurium in place of selenium, which increases the size of the lattice by about 6%. It is also a topological insulator with band inversion occurring at the Γ point. Given the similarity, one may ask if it is reasonable to view Bi_2Te_3 as intrinsically strained Bi_2Se_3. To test this hypothesis, we performed a comparison of Bi_2Te_3 to strained Bi_2Se_3.

In order to generate an appropriate reference structure for comparison, Bi_2Te_3 was relaxed under identical external stress as the reference Bi_2Se_3. The results are shown in Table 10.3. The computed lattice parameters of Bi_2Te_3 are in good agreement with experiment. The Bi_2Se_3 lattice was then strained to match that of reference Bi_2Te_3. Using the gap stiffness and stress tensors, the band-gap of this strained Bi_2Se_3 was calculated and compared to the computed band-gap of Bi_2Te_3. Without spin-orbit interaction, the strained Bi_2Se_3 band-gap is similar to the band-gap of reference Bi_2Te_3. However, with spin-orbit interaction turned on, the gaps are dramatically different: in the strained Bi_2Se_3, the topological gap closes, but in Bi_2Te_3 the topological gap is quite large. This suggests that spin-orbit effects are strongly dependent on the chemical identity of the anion, and that treating Bi_2Te_3 as strained Bi_2Se_3 fails to capture the essential physics.

Strain is therefore an important parameter for influencing the topological insulating phase and can, in principle, drive the system through the topological insulating phase transition. While it may be possible to tune the band-gap with external stress, more interesting is the potential for inducing strain via chemical substitution. Viewing bismuth telluride as chemically strained bismuth selenide, however, fails dramatically, hinting at a complex relationship between chemical composition, material structure, and the physics underlying the topological insulating phase.

Chapter 11

Three-dimensional Dirac Semimetals

As we have seen, the transition between topological and trivial insulating phases is marked by a three-dimensional Dirac point. In addition to driving systems through this state, it is also of interest itself. We expect that a material at such a critical point, with a Dirac point comprising the only Fermi surface, would exhibit a wide range of interesting and possibly useful phenomena. In graphene, a number of striking properties, including high carrier mobilities, have been observed to arise from its two-dimensional Dirac points, and we may ask if a material with one or more three-dimensional Dirac points – which we shall call a Dirac semimetal – would display the same or analogous behavior. To be specific, in a Dirac semimetal, the conduction and valence bands contact only at discrete (Dirac) points in the Brillouin zone (BZ), dispersing linearly in all directions around these critical points. In this chapter we determine the preconditions for a symmetry protected Dirac point, and propose several materials that would have Dirac points as the sole Fermi surface. Much of this chapter's content appears in [109].

In general, a Dirac point is described by the Hamiltonian of Eq. (9.3) with $m = 0$. A useful way to think about a Dirac point is as two coincident Weyl points, which are described by Eq. (9.2). The Weyl Hamiltonian describes two linearly dispersing bands,

rather than the four of the massless Dirac Hamiltonian, that are degenerate at a point. As mentioned, provided $\det[v_{ij}] \neq 0$, the Weyl point is robust against perturbations because it uses all three Pauli matrices; any additional term will merely shift the location of the degeneracy. It is worth noting that graphene, when treated as spinless (as it often is), is actually a two-dimensional Weyl point described by only two Pauli matrices, and as such may be gapped by a term involving the unused Pauli matrix.

The Chern number describing this Weyl point takes values $\mathrm{sgn}(\det[v_{ij}]) = \pm 1$. If a Weyl point occurs at some BZ momentum $\mathbf{k}$, time reversal (T) symmetry requires that another Weyl point occur at $-\mathbf{k}$ with equal Chern number. However, the total Chern number associated with the entire Fermi surface must vanish, and there must exist *two more* Weyl points of opposite Chern number at $\mathbf{k}'$ and $-\mathbf{k}'$. Inversion (I) symmetry requires that Weyl points at $\mathbf{k}$ and $-\mathbf{k}$ have opposite Chern number. Hence, under both T and I symmetries, $\mathbf{k} = \mathbf{k}'$ and the effective Hamiltonian involves four linearly dispersing bands around $\mathbf{k}$. This is the Dirac Hamiltonian, and it is not robust against perturbations because there are additional 4×4 Dirac matrices that can be used to open a gap at the Dirac point.

The Fermi surface of a Dirac semimetal consists entirely of such point-like (Dirac) degeneracies. The phase transition between a topological and a normal insulator with inversion symmetry is identified with a single Dirac point [110, 105] (Ref. [111] demonstrates such a Dirac point degenerate with massive bands.) If either inversion or time-reversal symmetry is broken, the Dirac point separates into Weyl points and one obtains a Weyl semimetal (Fig. 11.5(c)). These Weyl points switch partners and combine to form another Dirac point before combining to form another Dirac point. The topological nature of Weyl points gives rise to interesting properties such as Fermi-arc surface states [97] and pressure induced anomalous Hall effect [112]. Recent proposals to design a Weyl semimetal have been predicated upon the existence of a parent Dirac semimetal which splits into a Weyl semimetal by breaking I [96] or T-symmetry [113]. Ref. [114] demonstrates the existence

of bulk chiral fermions due to crystal symmetry in single space-groups.

11.1 Symmetry Constraints

As discussed, Dirac points that arise in a topological phase transition exist at single points in parameter space and are not robust. In general, two Weyl points with opposite Chern numbers annihilate each other unless their degeneracy is otherwise protected. It is natural to ask, then, if we can stabilize the bulk Dirac point, both to exploit its interesting features, and as a starting point for obtaining topological phases. In particular, can we protect the Dirac point(s), and the Dirac semimetal phase, via crystallographic symmetry? In this section, we outline the conditions for such a protected Dirac point to exist, and describe the physical and chemical nature of this state in a Dirac semimetal.

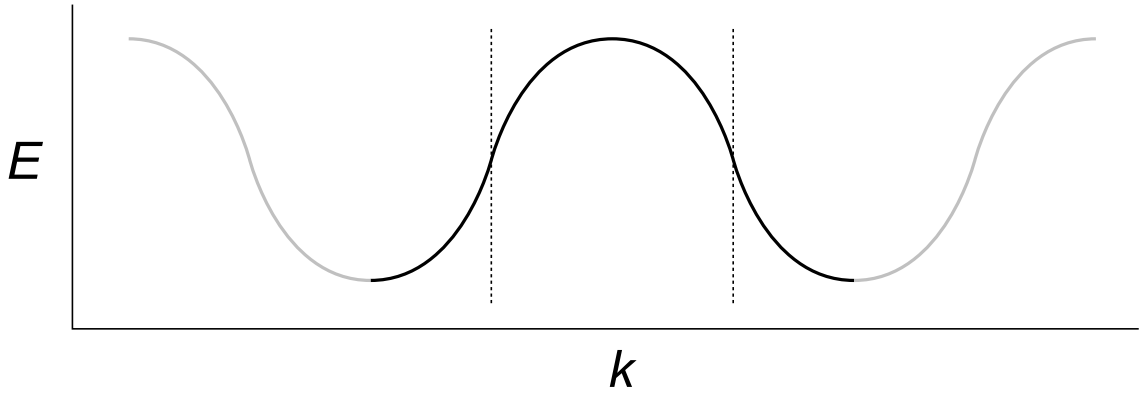
First, we shall describe the requirements for the existence of four-dimensional irreducible representations (FDIR), a necessary (but not sufficient) condition for a symmetry-protected Dirac point. Formally, we are interested in FDIRs of double space groups at points $\mathbf{k}$ that carry a Chern number of zero. The latter is necessary since these are to be the only Fermi surface and time-reversal symmetry is preserved. The Chern number of a degenerate representation can be determined up to an integer by the rotation eigenvalues of the valence bands. Electron states spanning an FDIR are equivalent to a $p_{\frac{3}{2}}$ quadruplet that exhibits eigenvalues $e^{\pm i3\pi/n}, e^{\pm i\pi/n}$ for a $2\pi/n$ rotation symmetry. Rotation eigenvalues of states at time-reversed momenta about the degenerate point are complex conjugates. Therefore the FDIR will carry Chern numbers $\pm 1 \mod n$ for one valence band and $\pm 3 \mod n$ for the other with total Chern number $\pm 4 \mod n$ or $\pm 2 \mod n$ for the FDIR. This is zero only for $n = 1, 2, 4$. If the conduction and valence bands are distinct in a small region around $\mathbf{k}$, the Chern number of the FDIR will be non-zero if the little group $G_{\mathbf{k}}$ contains a $2\pi/3$ or $2\pi/6$ rotation symmetry. The cubic groups all contain FDIRs that exist at the Γ

point; however, these are all protected by $2\pi/3$ rotation symmetry, eliminating them from consideration. As an example, HgTe, which has zincblende structure, has such a FDIR at Γ at the Fermi energy.

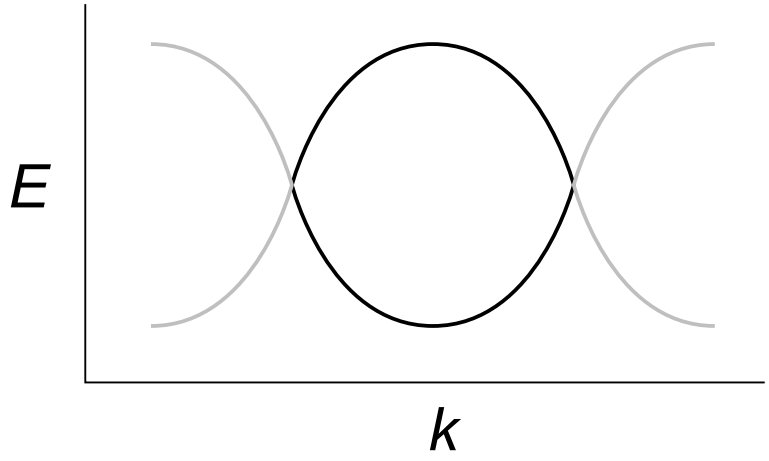
The only other FDIRs that may occur do so on BZ boundaries of non-symmorphic space groups. Non-symmorphic space-groups contain point group operations coupled with non-primitive lattice translations. For example, inversion interchanges the FCC sub-lattices in the diamond space-group. Representations of non-symmorphic space-groups at momenta inside the BZ momenta are obtained from regular representations, while those at the *surface* BZ momenta are obtained from *projective* representations of the associated crystal point group. The factor system of the projective representation is chosen to implement the required non-primitive translation corresponding to the non-symmorphic point group operation [115]. A theorem by Schur guarantees that projective representations of a group can be obtained by restricting to the group elements the regular representations of a larger group called the central extension group [115]. The central extension of a group is obtained by taking its product with another finite Abelian group. The important point to emphasize is that representations of non-symmorphic space-groups are obtained from representations of central extensions of the 32 point groups. Central extension groups exhibit FDIRs even without three-fold rotations in the original point group. This allows for Dirac points to exist in three dimensions as symmetry-allowed degeneracies.

While the rigorous justification for FDIRs at BZ boundaries, summarized above, is quite abstract, here we provide a simple conceptual picture. A non-symmorphic crystal may be thought of as a supercell of a symmorphic crystal in which the pure translation symmetry of the component primitive cells has been broken. The translational symmetry only holds when combined with one or more point group symmetries. When one creates a supercell from a primitive cell, the BZ is reduced; if one has the band structure of the primitive cell (Fig. 11.1(a)), the band structure of a supercell may be constructed by "folding" it in

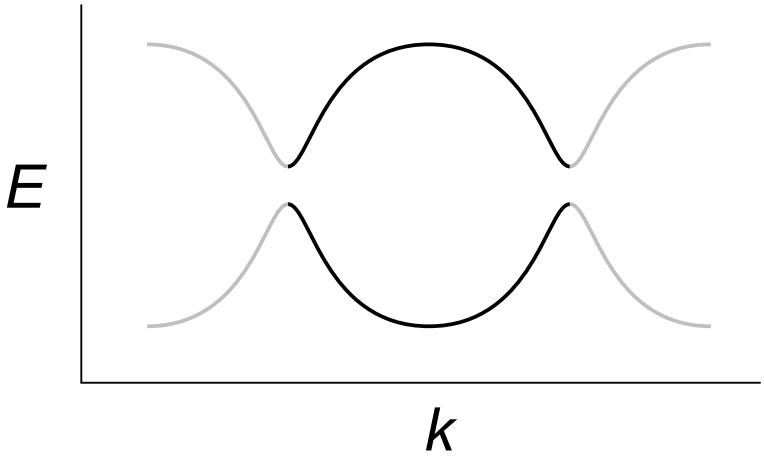
reciprocal space to lie within the new BZ (Fig. 11.1(b)). This yields a degeneracy where the folding occurred, at the new BZ boundary. Breaking translational symmetry removes this degeneracy, typically introducing a gap (Fig. 11.1(c)). However, if a point $\mathbf{k}$ has, as a symmetry, a non-symmorphic operation that includes the translational symmetry, it essentially does not see that it has been broken, and the degeneracy remains.



(a)



(b)



(c)

Figure 11.1: If a material with band structure shown in (a) has its unit cell doubled, the Brillouin zone is folded back along the dotted lines, and becomes (b). If the translational symmetry is broken (c), the degeneracy at the boundaries is gapped. However, if the translational symmetry is preserved as part of a non-symmorphic operation, then at the k that has the point group symmetry of the operation, it is as though the translational symmetry still holds, and the degeneracy is protected.

To realize a Dirac-like dispersion in the vicinity of an FDIR, some of the matrix elements $\langle\psi_i|\hat{\mathbf{p}}|\psi_j\rangle$, where $|\psi_i\rangle$ span the FDIR, must be non-zero. This is guaranteed if the symmetric Kronecker product of the FDIR with itself contains the vector representation of the central extension group to which the FDIR belongs [116]. We restrict to the symmetric part of the Kronecker product because matrix elements $\langle\psi_i|\hat{\mathbf{p}}|\psi_j\rangle$ correspond to level transitions between states spanning the same representation [117]. Finally, the allowed representations in the vicinity of an FDIR should be such that each band disperses with non-zero slope in all directions. This is possible only if the valence band is distinct from the conduction band everywhere except at the Dirac point. Fig. 11.2 illustrates the various possible ways in which an FDIR can split linearly.

Thus, a 3D double space-group must satisfy the following criteria to allow a Dirac point: It must admit four-dimensional irreducible representations (FDIRs) at some point $\mathbf{k}$ in the BZ such that the four bands degenerate at $\mathbf{k}$ disperse linearly in all directions around $\mathbf{k}$ and the two valence bands carry zero total Chern number. If the little group $G_{\mathbf{k}}$ at $\mathbf{k}$ contains a three-fold or a six-fold rotation symmetry and the valence and conduction bands around $\mathbf{k}$ are non-degenerate, the Chern number of the FDIR is guaranteed to be non-zero. This rules out symmorphic space-groups with FDIRs because they contain three-fold rotations. This also rules out interior BZ momenta because non-symmorphic little groups without three-fold rotations exhibit FDIRs only on the boundary of the BZ [115]. To guarantee linear dispersion of bands around $\mathbf{k}$, the symmetric kronecker product $[R_{\mathbf{k}} \times R_{\mathbf{k}}]$ of the FDIR with itself must contain the vector representation of $G_{\mathbf{k}}$ [116]. Finally, away from $\mathbf{k}$, the FDIR must split so that the valence and conduction bands are non-degenerate everywhere except at $\mathbf{k}$ (Fig. 11.2).

We apply the above criteria to two important space-groups. The space-group of diamond (227, Fd3m) exhibits FDIRs R_Γ at Γ and R_X at X. G_Γ contains three-fold rotation symmetry and $[R_\Gamma \times R_\Gamma]$ does not contain the vector representation of G_Γ. Therefore, the

Γ point in a diamond lattice cannot host a Dirac point. R_X is a projective representation of G_X which does not have any three-fold rotations because all the point group operations in G_X are those of the group D_{4h}. $[R_X \times R_X]$ contains the vector representation of G_X. Finally, R_X splits into either two doublets or four singlets away from X (Figs. 11.2(a) and 11.2(b)). Therefore, the X point in space-group 227 is a candidate to host a Dirac semimetal if its FDIR can be elevated to the Fermi level. The Dirac point at X in the FKM model is spanned by states belonging to R_X (Fig. 11.4(d)).

The zincblende lattice (space-group 216, F$\bar{4}$3m) has an FDIR R'_Γ at Γ and the little group G'_Γ has a three-fold rotation symmetry. $[R'_\Gamma \times R'_\Gamma]$ contains the vector representation of G'_Γ. Mirror symmetry in G'_Γ requires R'_Γ to split into a two-fold degenerate representation and two non-degenerate representations along the (111) axis, which is also the symmetry axis for the three-fold rotation. Time reversal symmetry requires that the two-fold degenerate band remain flat along the (111) axis, Fig. 11.2(d). Thus the lowest band carries Chern number 0, while the two flat bands carry 1 and -1. Therefore the dispersion of R'_Γ is not Dirac-like along (111).

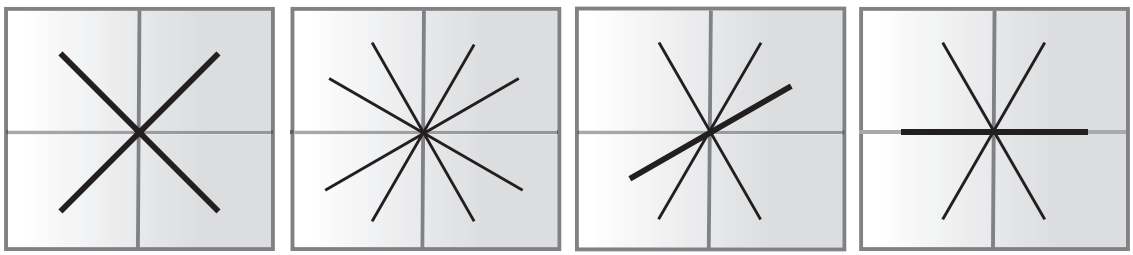

(a)
(b)
(c)
(d)

Figure 11.2: Linear splitting of four-fold degenerate irreducible representations (FDIRs). If the symmetric kronecker product of an FDIR with itself contains the vector representation of the group to which the FDIR belongs, it will split in one of the four possible ways displayed above. (a) The FDIR splits into two two-fold degenerate bands. This situation is realized at the X point of the FCC Brillouin zone in a diamond lattice. (b) The FDIR splits into four non-degenerate bands. This situation arises at the Γ point in zincblende if mirror symmetry is broken (although the FDIR in zincblende develops a non-zero Chern number due to three-fold rotation symmetry at Γ). (c) The FDIR splits into two non-degenerate and one two-fold degenerate band with linear dispersion. (d) The splitting of the FDIR at Γ in zincblende. The two-fold degenerate band is constrained to be flat, implying quadratic dispersion along that direction. The Chern number of this representation is zero in spite of a three-fold rotation symmetry because the conduction and valence bands are degenerate away from Γ.

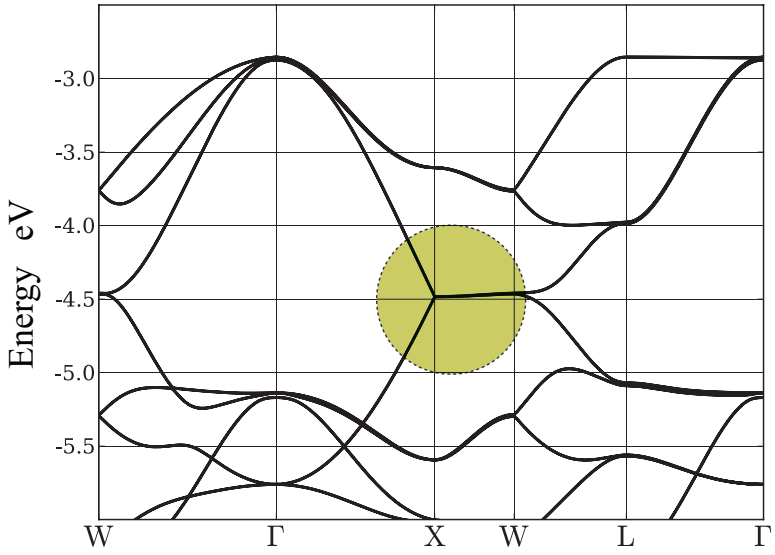
In HgTe, which takes the zincblende lattice, the degenerate valence and conduction states at Γ span R'_Γ and constitute the entire Fermi surface. It is known that in HgTe the valence and conduction bands disperse linearly in two directions around Γ and quadratically in a third (Fig. 11.2(d) and Ref. [117]). One might ask if a perturbation might turn HgTe into a Dirac semimetal. However, the zincblende lattice does not satisfy the criteria for 3D Dirac points as outlined above, so HgTe cannot host a Dirac semimetal. (a) Γ is an interior point of the BZ and the little group at Γ contains a three-fold rotation. (b) Mirror symmetry requires two bands to be degenerate along the $\langle 111 \rangle$, axis but since the Chern number must vanish, the degenerate bands must be flat and consist of a conduction and a valence band. This is why we see quadratic dispersion along the $\langle 111 \rangle$ axis. (c) Breaking mirror symmetry splits the degenerate flat band but then the Fermi surface develops other non-Dirac like pockets to compensate for the non-zero Chern number. (d) Breaking three-fold rotation symmetry splits the degeneracy at Γ entirely and the material becomes a topological insulator [39].

11.2 Proposed Materials and Physical Mechanism

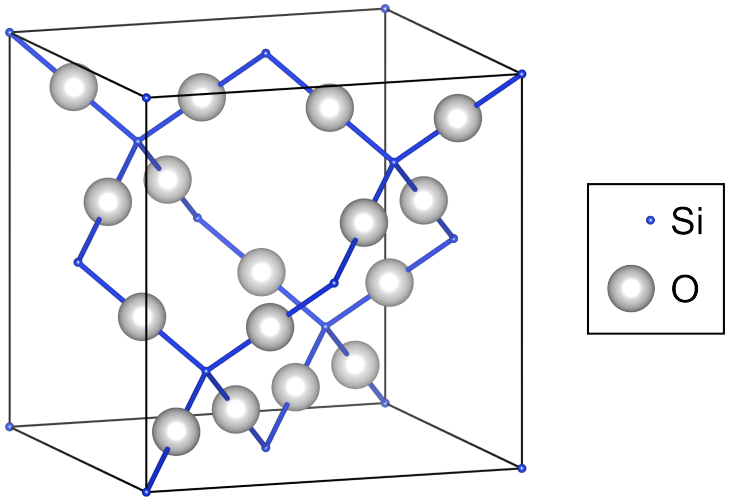
Although crystallographic symmetries determine whether 3D Dirac points can exist, physical and chemical considerations dictate whether they arise at the Fermi level without additional non-Dirac like pockets in the Fermi surface. In the FKM model, the Dirac point at X appears at the Fermi energy. However, in known materials on a diamond lattice the s-states appear below the Fermi energy. In realistic systems, additional orbitals hybridize with these s-states and bands cross the Fermi level at other points besides X. The problem is especially severe in space-group 227: without spin, the line V from X to W is two-fold degenerate. With spin-orbit coupling, this line splits weakly for lighter atoms so the bands dispersing along this line can hybridize and introduce additional Fermi surface. Forcing

species with s^1 valence states on the diamond lattice would fail to realize the FKM model. Indeed, *ab initio* calculations with group I elements and gold show that the splitting along V is insufficient to overcome this dispersion. In some cases, additional bands crossed the Fermi level.

We consider derivatives of the diamond lattice that remain in space-group 227. We place additional atoms in the lattice such that the configuration of added species allows its valence orbitals to either belong to the FDIR of interest, or appear away from the Fermi energy of the final structure. If the new species can split the nearby p states of the existing atoms away from the s levels, band crossing at the Fermi level can be avoided.



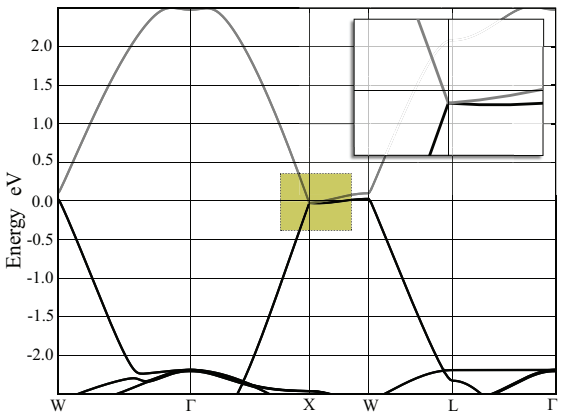
(a)



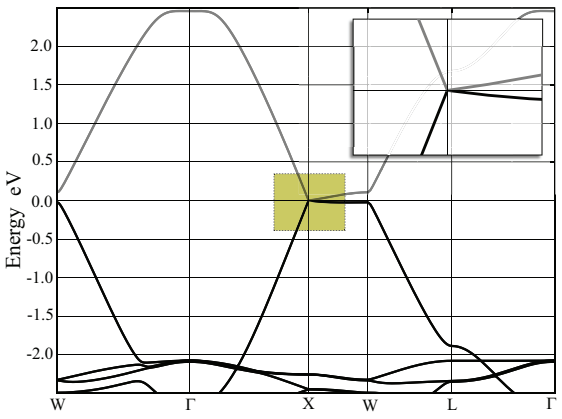
(b)

Figure 11.3: (a) Band structure of β-cristobalite SiO_2. Energy bands are plotted relative to the Fermi level. Each band is two-fold degenerate due to inversion symmetry. The FDIR (highlighted) at -4.5 eV is split into two linearly dispersing bands between X and Γ while the two degenerate bands along X and W are weakly split. This FDIR is buried deep below the Fermi level. (b) The β-cristobalite structure of SiO_2. Silicon atoms are arranged on a diamond lattice, with oxygen atoms sitting midway between pairs of silicon.

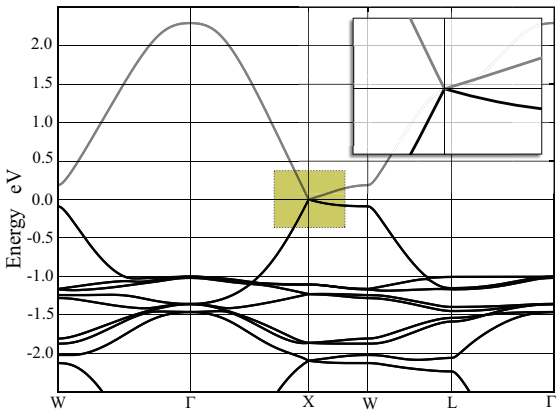
One such structure is β-cristobalite SiO_2 (Fig. 11.3(b)), which consists of silicon atoms on a diamond lattice with oxygen atoms midway between each pair of silicon atoms [118]. Oxygen atoms have two consequences: part of the O p-shell strongly hybridizes with the Si p-states, moving them away from the Si s-states, while the remaining O p-states hybridize with the Si s-states. A Dirac point can be realized by an Si $s-$O p bonding/anti-bonding set of states. Fig. 11.3(a) shows that the Si $s-$O p bands are present and take a configuration similar to the valence and conduction bands in the FKM model, but appear well below the Fermi energy. Additionally, the bands are nearly degenerate along the line V from X to W due to weak spin-orbit coupling.


Energy eV
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
W
Γ
X
W
L
Γ

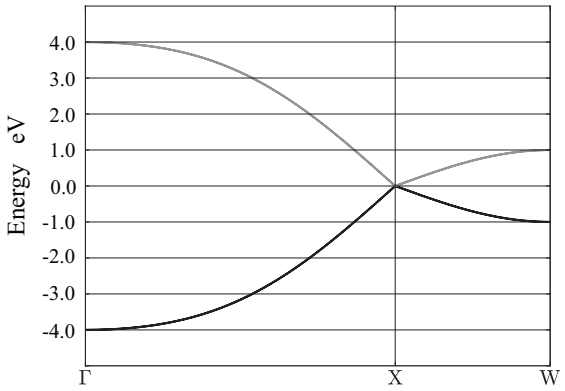
(a)


Energy eV
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
W
Γ
X
W
L
Γ

(b)


Energy eV
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
W
Γ
X
W
L
Γ

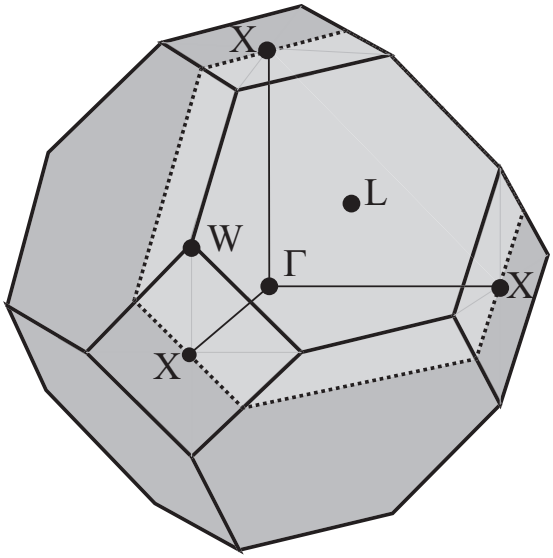
(c)


Energy eV
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
Γ
X
W

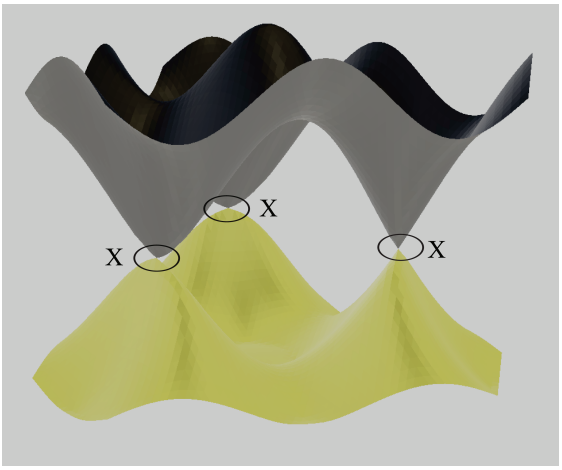
(d)

Figure 11.4: Band structures of (a) AsO_2, (b) SbO_2, and (c) BiO_2 in the β-cristobalite structure, and (d) s-states on a diamond lattice in the tight-binding model of Ref. [7]. Energy bands are plotted relative to the Fermi level. Each band is two-fold degenerate due to inversion symmetry. Insets: with increasing atomic number of the cation, spin-orbit coupling widens the gap along the line V from X to W. In $\mathrm{BiO_2}$ and $\mathrm{SbO_2}$, the dispersion around the X point is linear in all directions indicating the existence of Dirac points at X. $\mathrm{BiO_2}$ and $\mathrm{SbO_2}$ are Dirac semimetals because their Fermi surface consists entirely of Dirac points.

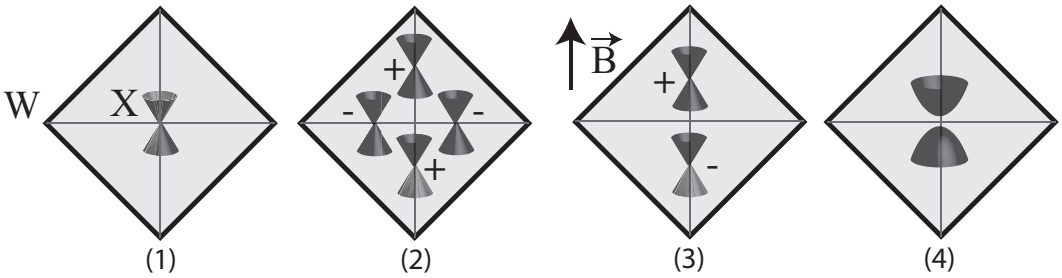
Heavier atoms substituting Si both widen this gap and bring the FDIR of interest at X to the Fermi level. Fig. 11.4 shows the band structures of compounds β-cristobalite XO_2 where X = As/Sb/Bi. The change in chemical identity promotes the X $s-$O p four-fold degeneracy at X to the Fermi level, and stronger spin-orbit coupling widens the gap along V. BiO_2 bears striking similarity to the FKM model, with linearly dispersing bands in a large energy range around a Dirac point at the Fermi level. Our calculations show that the phonon frequencies for β-cristobalite BiO_2 at Γ are positive, so it is a metastable structure. Further calculations reveal that it becomes unstable under uniform compression exceeding 2GPa, which represents a stability barrier of approximately 0.025eV per atom. On this basis, the possibility of synthesis appears promising. However, Bi_2O_4 is also likely to take the cervantite structure (after Sb_2O_4, which has similar stoichiometry [119]) which is 0.5 eV per atom lower in energy as compared to β-cristobalite and 60$\%$ smaller in volume. Therefore, we conclude that β-cristobalite BiO_2 would be metastable if synthesized, although preventing it from directly forming the cervantite structure would be challenging.



(a)

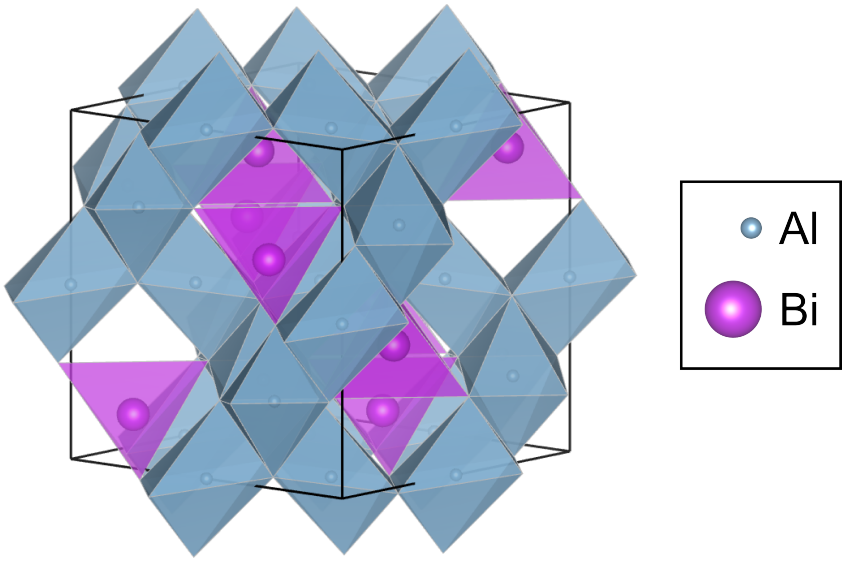


(b)

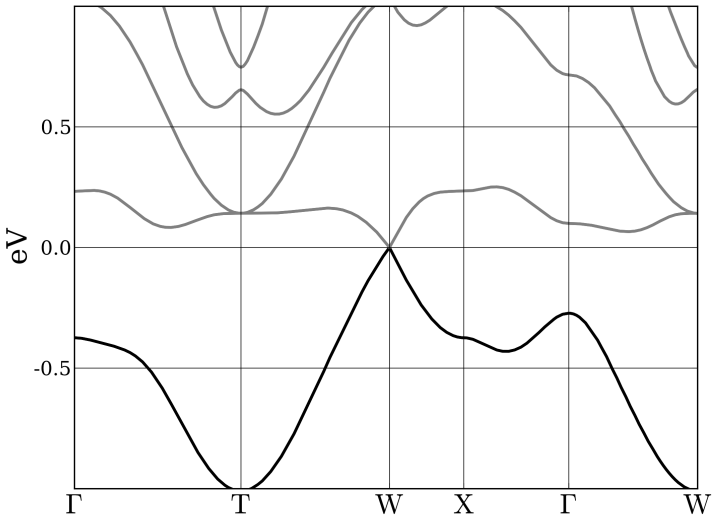


(c)

Figure 11.5: 3D Dirac semimetal in β-cristobalite BiO_2. (a) Brillouin zone (BZ) of the FCC lattice. The plane highlighted in gray joins the three symmetry-related X points. Other high symmetry points are also indicated. (b) Conduction and valence bands of β-cristobalite BiO_2 are plotted as functions of momentum on the plane highlighted in gray on the left. Each band is two-fold degenerate due to inversion symmetry. Dirac points appear at the center of the three zone faces of the BZ. (c) Dirac, Weyl and insulating phases in the diamond lattice. (1) The states at the Dirac point at X span a four-dimensional projective representation of the little group at X which contains a four-fold rotation accompanied by a sub-lattice exchange operation. (2) Four Weyl points on the zone face due to a small inversion breaking perturbation. The Chern number of each Weyl point is indicated. (3) Two Weyl points appear on the line from X to W for a T-breaking Zeeman field $\mathbf{B}$ oriented along that direction. $\mathbf{B}$ oriented along other directions gaps all the Dirac points by breaking enough rotational symmetry that no two-dimensional representations are allowed. (4) Gapped phase obtained by breaking the four-fold rotation symmetry or by applying a magnetic field in any direction except along $\hat{x}$, $\hat{y}$, or $\hat{z}$. The insulating phase can be a normal, strong or weak topological insulator [7].



(a)

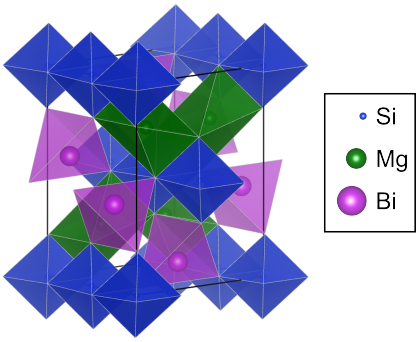


(b)

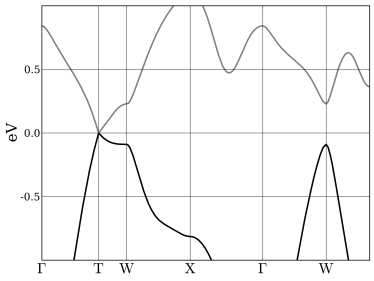
Figure 11.6: $BiAl_2O_4$ in the spinel structure is shown in (a). The polyhedra represent the oxygen cages coordinating the cations, with the oxygen atoms located at the vertices. For clarity, the atoms themselves are not shown explicitly. The (b) band structure reveals a clear Dirac point at X with high dispersion in all directions.

Next, we consider the spinel structure, also in space group 227. The spinel structure, shown in Fig. 11.6(a), has the composition AB_2O_4, where A=Mg and B=Al for the eponymous mineral. The species A sits on a diamond lattice and is tetrahedrally coordinated by oxygen; the species B sit at octahedrally coordinated sites. Starting from the mineral spinel and proceeding as in the case with the β-cristobalite structure, we find that $BiAl_2O_4$ provides a high-quality Dirac point at X (Fig. 11.6(b)). However, this structure is highly unstable. In fact, despite several cation choices, we were unable to identify a meta-stable Dirac semimetal with this structure.

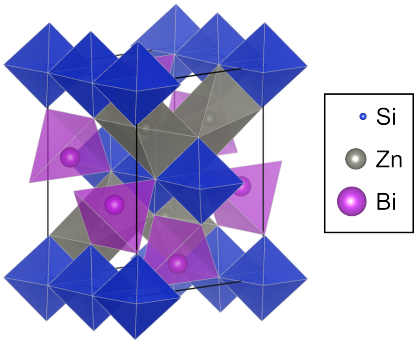
However, there is a lower-symmetry derivative of the spinel structure that can manifest a Dirac semimetal. Orthorhombic space group 74 is a descendant of 227 by way of 141. Two of the degenerate FDIRs of 227 are gapped, leaving one point capable of hosting a Dirac point. The composition of this distorted spinel structure is $ABB'O_4$, where the symmetry between the B-sites has been broken. Known examples are $VLiCuO_4$ [120] and $GeMnMnO_4$ [121]. We have identified three metastable compositions with this structure: $BiMgSiO_4$, $BiZnSiO_4$, and $BiInAlO_4$. Their crystal and band structures are shown in Fig. 11.7.


Si
Mg
Bi

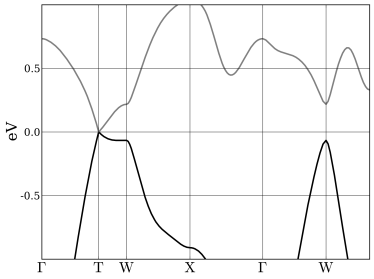
(a)


eV
0.5
0.0
-0.5
Γ
T
W
X
Γ
W

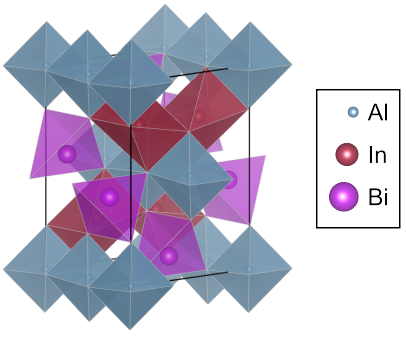
(b)


Si
Zn
Bi

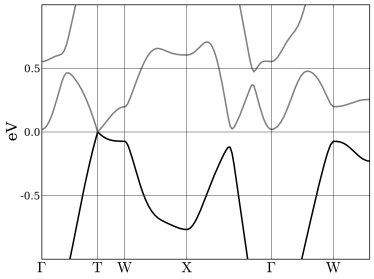
(c)


eV
0.5
0.0
-0.5
Γ
T
W
X
Γ
W

(d)


Al
In
Bi

(e)


eV
0.5
0.0
-0.5
Γ
T
W
X
Γ
W

(f)

Figure 11.7: The crystal structures, with oxygen atoms implied at polyhedral vertices, and band structures for $BiMgSiO_4$ ((a) and (b)), $BiZnSiO_4$ ((c) and (d)), and $BiInAlO_4$ ((e) and (f)). We note that the geometries and electronic structures of $BiMgSiO_4$ and $BiZnSiO_4$ appear to be almost identical. In all cases, there is high dispersion away from the Dirac point, which comprises the only Fermi surface.

Interestingly, all three of the structures indicate a nominal oxidation state of $2+$ for bismuth, in contrast to $4+$ in the case of BiO_2. Density of states calculations confirm that in the latter case it is the bismuth p-orbitals that are contributing to the Dirac point. To understand this difference, we inspect the wavefunctions directly. For clarity, cartoon representations of the Dirac point states are shown in Fig. 11.9.

In both cases, there exists a pair of states (neglecting spin) that are degenerate and related by the non-symmorphic symmetry operation(s), as symmetry dictates. These states lie on zig-zag chains of bismuth atoms. In cubic BiO_2, such chains can be identified running in all three dimensions (corresponding to each X point), and are visible by inspection of Fig. 11.3(b), and each orientation hosts a state of the kind shown in Fig. 11.9(a). In the orthorhombic distorted spinel, only one of the chains retains the appropriate symmetry; these can be seen more clearly in the view of the $BiZnSiO_4$ cyrstal structure in Fig. 11.8. For BiO_2, oxygen atoms lie along the Bi-Bi bonds, mediating the interaction between their s-states. However, in the distorted spinel structures, the oxygen atoms lie *between* the Bi-Bi bonds. The bismuth atoms must interact directly, and the s-states do so only weakly, flattening the band that they form, and compromising the quality of the associated Dirac point. The oriented p-orbitals are able to interact strongly, as shown in Fig. 11.9(b), resulting in significant dispersion away from the Dirac point.

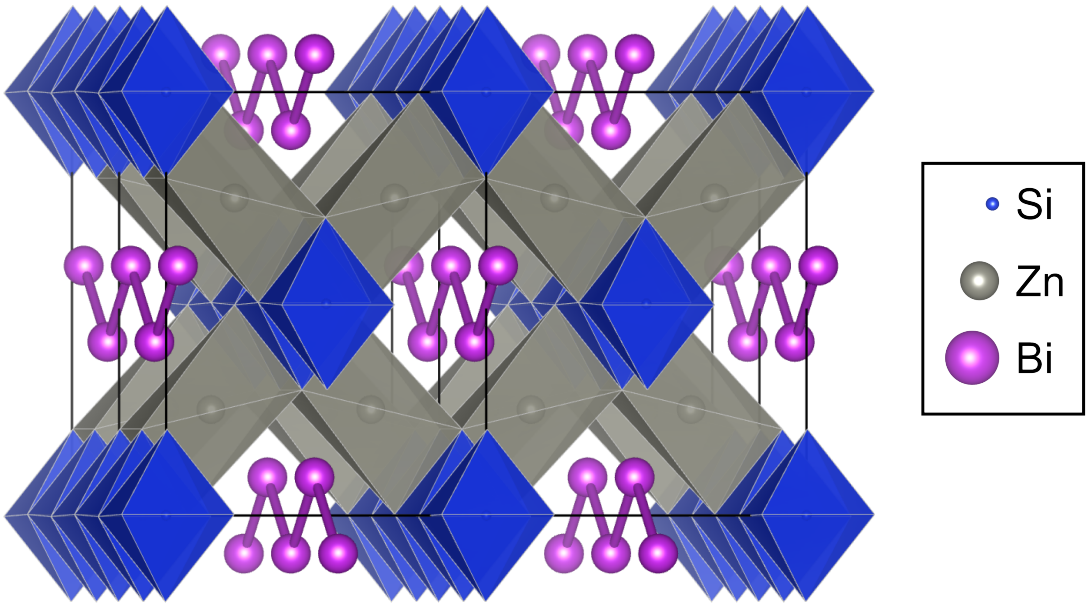
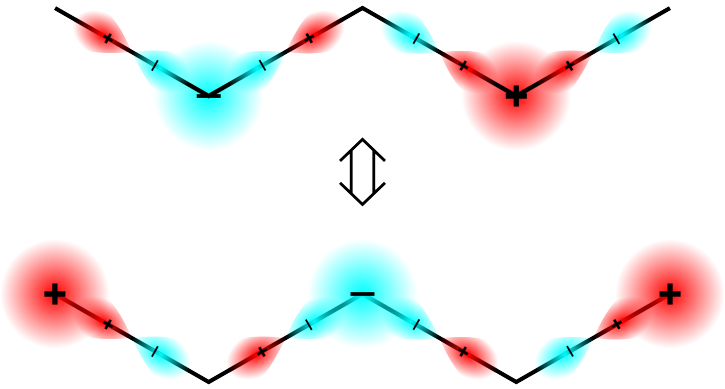
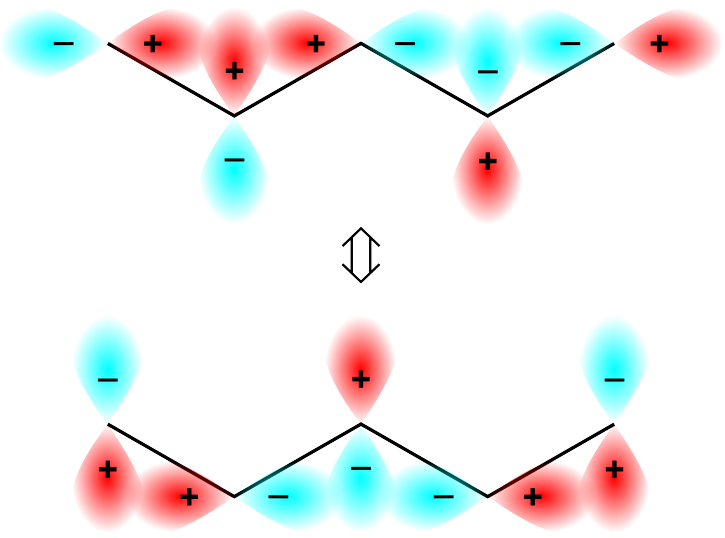

Si
Zn
Bi

Figure 11.8: A $2\times2\times1$ supercell of $BiZnSiO_4$. Zig-zag chains of bismuth are clearly seen running through channels created by surrounding oxygen cages of Zn and Si. The Bi-Bi bonds shown are quite short at 3.2Å



(a)



(b)

Figure 11.9: The Dirac point wavefunctions for (a) BiO_2 and (b) spinel-derived structures. In BiO_2 the Bi s-states strongly hybridize with O p-states, which mediate the Bi-Bi interaction. In the spinel structure, the oxygen atoms are arranged differently, and the Bi atoms must interact directly. This is most effectively accomplished by the Bi p-orbitals. In both cases, there are two degenerate states related by the non-symmorphic symmetry operation that enables the FDIR.

We may gain further insight by constructing a Lewis dot structure for the distorted spinel case, shown in Fig. 11.10. There is a lone pair for each s-orbital, and a single p-electron participating in bonding. Since there is no preferred direction, the bonding is essentially characterized by two resonance structures, as in a conjugated diene, or, in two dimensions, graphene.

The physical – rather than just the mathematical – nature of the Dirac semimetal state is now clear. Like in the Su-Schrieffer-Heeger model [122], our Dirac point is described by two degenerate states that gap if the symmetry is broken by biasing the bismuth bonds so that they tend towards dimers. No Peierls instability occurs because this is not particularly favorable; in the insulating limit bismuth must take an oxidation state of $2+$. The metallic-like bonding is actually preferred, as it avoids this awkward oxidation state. Thus, the bismuth chains can be thought of as charged metallic wires running through what is otherwise an insulator. The unusual nominal oxidation state of bismuth is *required*; otherwise the system will be unstable relative to a fully insulating state.

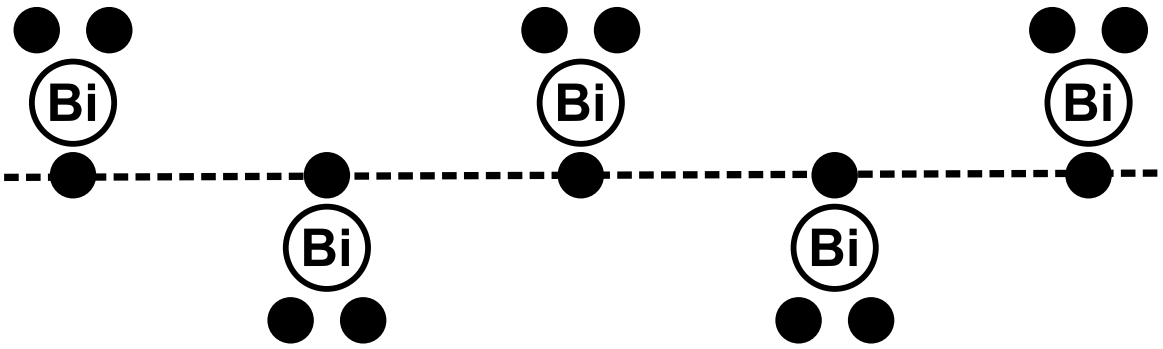

Bi
Bi
Bi
Bi
Bi

Figure 11.10: The s-states of bismuth behave essentially as lone pairs, while the unpaired p-electron of each bismuth participates in a delocalized bonding state similar to those in conjugated carbon chains.

We assert that these features are indicative of general rules for Dirac semimetals. First, symmetry must allow for an isolated Dirac point. Second, the structure must contain at least one high-Z atom with an unpaired electron. Third, the other atoms of the lattice should be as small as possible and should be arranged to allow or mediate strong interactions between the high-Z atoms. Fourth, the nominal oxidation state implied for the high-Z atom must be unfavorable, and the oxidation states of the other atoms should be well-defined. Finally, the high-Z atoms should form an extended structure within the lattice. These conditions allow for a quasi-metallic state associated with the Dirac point on the high-Z sub-structure, while minimizing metallic behavior elsewhere in the BZ.

Bibliography

[1] Koch, W. T. H., Munser, R., Ruppel, W., and Wurfel, P. *Ferroelectrics* **13**(1-4), 305–307 (1976).

[2] Ihlefeld, J. F., Podraza, N. J., Liu, Z. K., Rai, R. C., Xu, X., Heeg, T., Chen, Y. B., Li, J., Collins, R. W., Musfeldt, J. L., Pan, X. Q., Schubert, J., Ramesh, R., and Schlom, D. G. *Appl. Phys. Lett.* **92**(14), 142908 April (2008).

[3] Ji, W., Yao, K., and Liang, Y. C. *Phys. Rev. B* **84**(9), 094115 September (2011).

[4] Yang, S. Y., Seidel, J., Byrnes, S. J., Shafer, P., Yang, C. H., Rossell, M. D., Yu, P., Chu, Y. H., Scott, J. F., Ager, J. W., Martin, L. W., and Ramesh, R. *Nat. Nanotechnol.* **5**(2), 143–147 February (2010).

[5] Seidel, J., Fu, D. Y., Yang, S. Y., Alarcon-Llado, E., Wu, J. Q., Ramesh, R., and Ager, J. W. *Phys. Rev. Lett.* **107**(12), 126805 September (2011).

[6] Hao, X. F., Stroppa, A., Picozzi, S., Filippetti, A., and Franchini, C. *Phys. Rev. B* **86**(1), 014116 Jul (2012).

[7] Liang, F. Kane, C. L. and Mele, E. J. *Phys. Rev. Lett.* **98**, 106803–1–4 (2007).

[8] Chynoweth, A. G. *Phys. Rev.* **102**(3), 705–14 (1956).

[9] Chen, F. S. *J. Appl. Phys.* **40**(8), 3389–96 (1969).

[10] Glass, A. M., von der Linde, D., and Negran, T. J. *Appl. Phys. Lett.* **25**(4), 233–5 (1974).

[11] Fridkin, V. M. *Crystallog. Rep.* **46**, 654–8 (2001).

[12] Dalba, G., Soldo, Y., Rocca, F., Fridkin, V. M., and Sainctavit, P. *Phys. Rev. Lett.* **74**(6), 988–91 February (1995).

[13] von Baltz, R. and Kraut, W. *Phys. Rev. B* **23**(10), 5590–6 (1981).

[14] Tonooka, K., Poosanaas, P., and Uchino, K. volume 3324 of *Smart Materials Technologies*, 224–232. International Society for Optics and Photonics, (1994).

[15] Sturman, B. I. and Fridkin, V. M. *The Photovoltaic and Photorefractive Effects in Noncentrosymmetric Materials*, volume 8 of *Ferroelectricity and Related Phenomena*. Gordon and Breach Science Publishers, (1992).

[16] Qin, M., Yao, K., and Liang, Y. C. *Appl. Phys. Lett.* **95**, 022912–1–4 (2009).

[17] Qin, M., Yao, K., and Liang, Y. C. *Appl. Phys. Lett.* **93**(12), 122904–1–3 September (2008).

[18] Pintilie, L., Vrejoiu, I., Le Rhun, G., and Alexe, M. *J. Appl. Phys.* **101**(6), 064109–1–8 March (2007).

[19] Basu, S. R., Martin, L. W., Chu, Y. H., Gajek, M., Ramesh, R., Rai, R. C., Xu, X., and Musfeldt, J. L. *Appl. Phys. Lett.* **92**, 091905–1–3 (2008).

[20] Ichiki, M., Furue, H., Kobayashi, T., Maeda, R., Morikawa, Y., Nakada, T., and Nonaka, K. *Appl. Phys. Lett.* **87**(22), 222903–1–3 November (2005).

[21] Yue, Z. J., Zhao, K., Zhao, S. Q., Lu, Z. Q., Li, X. M., Ni, H., and Wang, A. J. *J. Phys. D-Appl. Phys.* **43**(1), 015104–1–4 January (2010).

[22] Yuan, G. L. and Wang, J. L. *Appl. Phys. Lett.* **95**(25), 252904–1–3 December (2009).

[23] Yang, S. Y., Martin, L. W., Byrnes, S. J., Conry, T. E., Basu, S. R., Paran, D., Reichertz, L., Ihlefeld, J., Adamo, C., Melville, A., Chu, Y. H., Yang, C. H., Musfeldt, J. L., Schlom, D. G., Ager III, J. W., and Ramesh, R. *Appl. Phys. Lett.* **95**, 062909–1–3 (2009).

[24] Pintilie, L., Stancu, V., Vasile, E., and Pintilie, I. *J. Appl. Phys.* **107**(11), 114111 June (2010).

[25] Cao, D. W., Zhang, H., Fang, L. A., Dong, W., Zheng, F. G., and Shen, M. R. *Appl. Phys. Lett.* **97**(10), 102104 September (2010).

[26] Choi, T., Lee, S., Choi, Y., Kiryukhin, V., and Cheong, S.-W. *Science* **324**, 63–6 (2009).

[27] Klitzing, K., Dorda, G., and Pepper, M. *Phys. Rev. Lett.* **45**(6), 494–497 Aug (1980).

[28] Thouless, D., Kohmoto, M., Nightingale, M., and den Nijs, M. *Phys. Rev. Lett.* **49**(6), 405–408 Aug (1982).

[29] Laughlin, R. *Phys. Rev. B* **23**(10), 5632–5633 May (1981).

[30] Hatsugai, Y. *Phys. Rev. Lett.* **71**(22), 3697–3700 Nov (1993).

[31] Haldane, F. D. M. *Phys. Rev. Lett.* **61**(18), 2015–2018 Oct (1988).

[32] Kane, C. L. and Mele, E. J. *Phys. Rev. Lett.* **95**, 146802 (2005).

[33] Kane, C. L. and Mele, E. J. *Phys. Rev. Lett.* **95**(22), 226801 November (2005).

[34] Moore, J. E. and Balents, L. *Phys. Rev. B* **75**(12), 121306 March (2007).

[35] Roy, R. *Phys. Rev. B* **79**(19), 195322 May (2009).

[36] Hasan, M. Z. and Kane, C. L. *Rev. Mod. Phys.* **82**(4), 3045–3067 November (2010).

[37] Bernevig, B. A., Hughes, T. L., and Zhang, S.-C. *Science* **314**, 1757–61 (2006).

[38] Fu, L. and Kane, C. L. *Phys. Rev. B* **76**(4), 045302 July (2007).

[39] Konig, M., Wiedmann, S., Brune, C., Roth, A., Buhmann, H., Molenkamp, L. W., Qi, X. L., and Zhang, S. C. *Science* **318**(5851), 766–770 November (2007).

[40] Hsieh, D., Qian, D., Wray, L., Xia, Y., Hor, Y. S., Cava, R. J., and Hasan, M. Z. *Nature* **452**(7190), 970–U5 April (2008).

[41] Chen, Y. L., Analytis, J. G., Chu, J. H., Liu, Z. K., Mo, S. K., Qi, X. L., Zhang, H. J., Lu, D. H., Dai, X., Fang, Z., Zhang, S. C., Fisher, I. R., Hussain, Z., and Shen, Z. X. *Science* **325**(5937), 178–181 July (2009).

[42] Xia, Y., Qian, D., Hsieh, D., Wray, L., Pal, A., Lin, H., Bansil, A., Grauer, D., Hor, Y. S., Cava, R. J., and Hasan, M. Z. *Nat. Phys.* **5**(6), 398–402 June (2009).

[43] Zhang, H. J., Liu, C. X., Qi, X. L., Dai, X., Fang, Z., and Zhang, S. C. *Nat. Phys.* **5**(6), 438–442 June (2009).

[44] Fu, L. and Kane, C. L. *Phys. Rev. Lett.* **100**(9), 096407 March (2008).

[45] Qi, X. L., Hughes, T. L., and Zhang, S. C. *Phys. Rev. B* **78**(19), 195424 November (2008).

[46] Qi, X. L., Li, R. D., Zang, J. D., and Zhang, S. C. *Science* **323**(5918), 1184–1187 February (2009).

[47] Essin, A. M., Moore, J. E., and Vanderbilt, D. *Phys. Rev. Lett.* **102**(14), 146805 April (2009).

[48] Perdew, J. P., Burke, K., and Ernzerhof, M. *Phys. Rev. Lett.* **77**, 3865–8 (1996).

[49] Rappe, A. M., Rabe, K. M., Kaxiras, E., and Joannopoulos, J. D. *Phys. Rev. B Rapid Comm.* **41**, 1227–30 (1990).

[50] Ramer, N. J. and Rappe, A. M. *Phys. Rev. B* **59**, 12471–8 (1999).

[51] Momma, K. and Izumi, F. *J. Appl. Crystallogr.* **41**(3), 653–658 Jun (2008).

[52] Blount, E. I. In *Solid State Physics: Advances in Research and Applications,* Seitz, F. and Turnbull, D., editors, volume 13, 305–73. Academic Press, (1962).

[53] Sipe, J. E. and Shkrebtii, A. I. *Phys. Rev. B* **61**, 5337–52 (2000).

[54] Young, S. M. and Rappe, A. M. *Phys. Rev. Lett.* **109**(11), 116601 Sep (2012).

[55] Young, S. M., Zheng, F., and Rappe, A. M. *Phys. Rev. Lett.* **109**(23), 236601 Dec (2012).

[56] Brody, P. S. *J. Solid State Chem.* **12**(3-4), 193–200 (1975).

[57] Koch, W. T. H., Munser, R., Ruppel, W., and Wurfel, P. *Solid State Commun.* **17**(7), 847 – 850 (1975).

[58] Catalan, G. and Scott, J. F. *Adv. Mater.* **21**(24), 2463–2485 June (2009).

[59] Ji, W., Yao, K., and Liang, Y. C. *Adv. Mater.* **22**(15), 1763–+ April (2010).

[60] Alexe, M. and Hesse, D. *Nat. Comms.* **2**, 256 (2011).

[61] Ederer, C. and Spaldin, N. A. *Phys. Rev. B* **71**(6), 060401 February (2005).

[62] Lebegue, S., Arnaud, B., and Alouani, M. *Phys. Rev. B* **72**(8), 085103 Aug (2005).

[63] Chen, C. T. and Cahan, B. D. *J. Opt. Soc. Am.* **71**(8), 932 Aug (1981).

[64] Yang, H., Mi, W., Bai, H., and Cheng, Y. *RSC Adv.* **2**, 10708–10716 (2012).

[65] Inoue, Y. *Energy Environ. Sci.* **2**, 364–386 (2009).

[66] Reznik, L. G., Anikiev, A. A., Umarov, B. S., and Scott, J. F. *Ferroelectrics* **64**(1), 215–219 Jun (1985).

[67] Anikiev, A., Reznik, L. G., Umarov, B. S., and Scott, J. F. *Ferroelectr. Lett.* **3**(4), 89–96 Feb (1985).

[68] Dhar, A. and Mansingh, A. *J. Appl. Phys.* **68**(11), 5804 (1990).

[69] Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Corso, A. D., de Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A. P., Smogunov, A., Umari, P., and Wentzcovitch, R. M. *J. Phys.: Condens. Matter* **21**, 395502–20 (2009).

[70] Inaguma, Y., Tanaka, K., Tsuchiya, T., Mori, D., Katsumata, T., Ohba, T., Hiraki, K.-i., Takahashi, T., and Saitoh, H. *J. Am. Chem. Soc.* **133**(42), 16920–16929 Oct (2011).

[71] Sleight, A. and Prewitt, C. *J. Solid State Chem.* **6**(4), 509–512 Apr (1973).

[72] Inaguma, Y., Yoshida, M., and Katsumata, T. *J. Am. Chem. Soc.* **130**(21), 6704–6705 May (2008).

[73] Gou, H., Gao, F., and Zhang, J. *Comput. Mater. Sci.* **49**(3), 552–555 Sep (2010).

[74] Takei, T., Haramoto, R., Dong, Q., Kumada, N., Yonesaki, Y., Kinomura, N., Mano, T., Nishimoto, S., Kameshima, Y., and Miyake, M. *J. Solid State Chem.* **184**(8), 2017–2022 Aug (2011).

[75] Awschalom, D. D. and Flatte, M. E. *Nat. Phys.* **3**(3), 153–159 March (2007).

[76] Murakami, S., Nagaosa, N., and Zhang, S. C. *Science* **301**(5638), 1348–1351 September (2003).

[77] Kato, Y. K., Myers, R. C., Gossard, A. C., and Awschalom, D. D. *Science* **306**(5703), 1910–1913 December (2004).

[78] Wunderlich, J., Kaestner, B., Sinova, J., and Jungwirth, T. *Phys. Rev. Lett.* **94**(4), 047204 February (2005).

[79] Ganichev, S. D., Ivchenko, E. L., Danilov, S. N., Eroms, J., Wegscheider, W., Weiss, D., and Prettl, W. *Phys. Rev. Lett.* **86**(19), 4358–4361 May (2001).

[80] Zhao, H., Pan, X. Y., Smirl, A. L., Bhat, R. D. R., Najmaie, A., Sipe, J. E., and van Driel, H. M. *Phys. Rev. B* **72**(20), 201302 November (2005).

[81] Bhat, R. D. R. and Sipe, J. E. *Phys. Rev. Lett.* **85**(25), 5432–5435 December (2000).

[82] Ganichev, S. D., Ivchenko, E. L., Bel'kov, V. V., Tarasenko, S. A., Sollinger, M., Weiss, D., Wegscheider, W., and Prettl, W. *Nature* **417**, 153–146 (2002).

[83] Ganichev, S. D., Bel'kov, V. V., Tarasenko, S. A., Danilov, S. N., Giglberger, S., Hoffmann, C., Ivchenko, E. L., Weiss, D., Wegscheider, W., Gerl, C., Schuh, D., Stahl, J., De Boeck, J., Borghs, G., and Prettl, W. *Nat. Phys.* **2**(9), 609–613 September (2006).

[84] Najmaie, A., Sherman, E. Y., and Sipe, J. E. *Phys. Rev. Lett.* **95**(5), 056601 July (2005).

[85] Yang, C. L., He, H. T., Ding, L., Cui, L. J., Zeng, Y. P., Wang, J. N., and Ge, W. K. *Phys. Rev. Lett.* **96**(18), 186605 May (2006).

[86] Bhat, R. D. R., Nastos, F., Najmaie, A., and Sipe, J. E. *Phys. Rev. Lett.* **94**(9), 096603 March (2005).

[87] Uchida, K., Takahashi, S., Harii, K., Ieda, J., Koshibae, W., Ando, K., Maekawa, S., and Saitoh, E. *Nature* **455**(7214), 778–781 (2008).

[88] Young, S. M., Zheng, F., and Rappe, A. M. *Phys. Rev. Lett.* **110**(5), 057201 Jan (2013).

[89] Cracknell, A. *Magnetism in Crystalline Materials: Applications of the Theory of Groups of Cambiant Symmetry*. Pergamon Press, Oxford, (1975).

[90] Rollmann, G., Rohrbach, A., Entel, P., and Hafner, J. *Phys. Rev. B* **69**(16), 165107 April (2004).

[91] Zheng, F., Young, S., and Rappe, A. (2012).

[92] Shi, J., Zhang, P., Xiao, D., and Niu, Q. *Phys. Rev. Lett.* **96**, 076604 (2006).

[93] Fu, L. and Kane, C. *Phys. Rev. B* **74**(19), 195312 Nov (2006).

[94] Soluyanov, A. A. and Vanderbilt, D. *Phys. Rev. B* **83**(3), 035108 Jan (2011).

[95] Kitaev, A. and Preskill, J. *Phys. Rev. Lett.* **96**(11), 110404 Mar (2006).

[96] Jiang, H.-C., Wang, Z., and Balents, L. *Nat. Phys.* **8**(12), 902–905 Nov (2012).

[97] Wan, X., Turner, A. M., Vishwanath, A., and Savrasov, S. Y. *Phys. Rev. B* **83**(20), 205101 May (2011).

[98] Mooser, E. and Pearson, W. B. *Phys. Rev.* **101**(1), 492–493 (1956).

[99] Black, J., Conwell, E. M., Seigle, L., and Spencer, C. W. *J. Phys. Chem. Solids* **2**(3), 240–251 (1957).

[100] Zhang, G. H., Qin, H. J., Teng, J., Guo, J. D., Guo, Q. L., Dai, X., Fang, Z., and Wu, K. H. *Appl. Phys. Lett.* **95**(5), 053114 August (2009).

[101] Peng, H. L., Lai, K. J., Kong, D. S., Meister, S., Chen, Y. L., Qi, X. L., Zhang, S. C., Shen, Z. X., and Cui, Y. *Nat. Mater.* **9**(3), 225–229 March (2010).

[102] Li, Y., Wang, G., Zhu, X., Liu, M., Ye, C., Chen, X., Wang, Y., He, K., Wang, L., Ma, X., Zhang, H., Dai, X., Fang, Z., Xie, X., Liu, Y., Qi, X., Jia, J., Zhang, S., and Xue, Q. *ArXiv e-prints* December (2009).

[103] Zhang, Y., He, K., Chang, C. Z., Song, C. L., Wang, L. L., Chen, X., Jia, J. F., Fang, Z., Dai, X., Shan, W. Y., Shen, S. Q., Niu, Q., Qi, X. L., Zhang, S. C., Ma, X. C., and Xue, Q. K. *Nat. Phys.* **6**(8), 584–588 August (2010).

[104] Zhang, T., Cheng, P., Chen, X., Jia, J. F., Ma, X. C., He, K., Wang, L. L., Zhang, H. J., Dai, X., Fang, Z., Xie, X. C., and Xue, Q. K. *Phys. Rev. Lett.* **103**(26), 266803 December (2009).

[105] Young, S. M., Chowdhury, S., Walter, E. J., Mele, E. J., Kane, C. L., and Rappe, A. M. *Phys. Rev. B* **84**(8), 085106 Aug (2011).

[106] Wiese, J. R. and Muldawer, L. *J. Phys. Chem. Solids* **15**(1-2), 13–16 (1960).

[107] Mishra, S. K., Satpathy, S., and Jepsen, O. *J. Phys.: Condens. Matter* **9**(2), 461–470 January (1997).

[108] Zhang, W., Yu, R., Zhang, H. J., Dai, X., and Fang, Z. *New J. Phys.* **12**, 065013 June (2010).

[109] Young, S. M., Zaheer, S., Teo, J. C. Y., Kane, C. L., Mele, E. J., and Rappe, A. M. *Phys. Rev. Lett.* **108**, 140405 Apr (2012).

[110] Murakami, S. *New J. Phys.* **9**, 356 September (2007).

[111] Smith, J. C., Banerjee, S., Pardo, V., and Pickett, W. E. *Phys. Rev. Lett.* **106**(5), 056401 Feb (2011).

[112] Yang, K.-Y., Lu, Y.-M., and Ran, Y. *Phys. Rev. B* **84**(7), 075129 Aug (2011).

[113] Burkov, A. A. and Balents, L. *Phys. Rev. Lett.* **107**(12), 127205 September (2011).

[114] Manes, J. L. *Phys. Rev. B* **85**(15), 155118 Apr (2012).

[115] Bradley, C. J. and Cracknell, A. P. *The Mathematical Theory of Symmetry in Solids*. Clarendon Press Oxford, (1972).

[116] Hamermesh, M. *Group Theory and its Applications to Physical Problems*. Addison-Wesley Publishing Company Inc., (1964).

[117] Dresselhaus, G. *Phys. Rev.* **100**, 580–586 (1955).

[118] Villars, P. and Calvert, L. *Pearsons Handbook of Crystallographic Data for Intermetallic Phases*. ASM International, Materials Park, Ohio, 2nd edition, (1991).

[119] Thornton, G. *Acta Crystallogr., Sect. B: Struct. Sci.* **33**(4), 1271–1273 Apr (1977).

[120] Lafontaine, M. A., Leblanc, M., and Ferey, G. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **45**(8), 1205–1206 Aug (1989).

[121] Morimoto, N., Akimoto, S.-I., Koto, K., and Tokonami, M. *Science* **165**(3893), 586–588 Aug (1969).

[122] Su, W. P., Schrieffer, J. R., and Heeger, A. J. *Phys. Rev. Lett.* **42**, 1698–1701 Jun (1979).