Characterization and Control of the Surface of
the Topological Insulator Bi$_2$Se$_3$

by
Avery James Green

A Dissertation
Submitted to the State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

SUNY Polytechnic Institute Colleges
of Nanoscale Science and Engineering
2017
Acknowledgments

Any success I’ve achieved in graduate school, and in the field of nanoscience research, must be attributed to the support of my advisor, Prof. Alain C. Diebold. His guidance and attention gave me the direction and motivation I needed when I would otherwise be lost in the complex and difficult field of topological insulators, and the combination of scientific rigor and kindness he constantly displayed make him all that one could hope for in an academic role model. He helped me grow from a student of physics to a researcher capable of implementing various laboratory and analytic skills toward solving a difficult problem. This work, and all of my other efforts in graduate school, would have been impossible without him, and I am vastly grateful for his mentorship.

I would also like to thank my committee, and all of the other members of the scientific community who have helped me along the way. Prof. Carl Ventrice (CNSE), Prof. Vince LaBella (CNSE), Prof. Hassaram Bakhru (CNSE), and Prof. Robert Hull (RPI) all provided essential feedback on my thesis work, and were instrumental in helping me produce my best possible dissertation. The samples and discussion provided by Prof. Chris Hinkle (UT-Dallas) and Dr. Lee Walsh (formerly UT-Dallas) allowed me to address a problem in the topological insulator field that I would have otherwise had no access to. Dr. George Orji of NIST was a highly valuable collaborator, enabling me (and all AFM users at CNSE) to make sub-nm topographical measurements with precision. I would also like to thank Prof. Ken Burch (Boston College) and Prof. Bradley Thiel (CNSE) for discussions of topological insulators and general materials science techniques. I must also acknowledge Dr. Brendan O’Brien for his assistance and teaching in TEM, and Westley Nolting for his STM characterization efforts.

The current and prior members of the Diebold lab group – Dr. Raja Muthinti, Dr. Manasa Medikonda, Dr. Sam O’Mullane, Dr. Dhairya Dixit, and Dr. Florence Nelson – were both peers
and counselors. They taught me the metrological techniques I needed, ensured that I was never alone in research, and showed me the path of a successful graduate student.

Finally, I would like to acknowledge my family and friends. I am fortunate be surrounded by people who are both warmhearted and brilliant. It’s only with the comfort and encouragement they provided that I had success in Albany after living in California for 22 years. My friends helped me feel at home away from home, and the solace and inspiration their companionship provided cannot be overstated. My mother, Donna Alkadri, my father, Richard Green, and my brother, Emerson Green, have always supported me in my decisions and pursuits. As the youngest of the family, I have always had three stellar examples to aspire to in life and in work, and have had access to a wealth of advice and unconditional love. This journey would have been impossible without them.
Abstract

The field of topological insulator (TI) materials is new. The ideal TI contains surface states in helical Dirac cones that can be used for spintronics or interconnect applications. Of the TI class, Bi$_2$Se$_3$ is the most promising for applications due to its stoichiometric composition, its relatively large band gap (0.3 eV), and the central (Γ-point) position of the Dirac cone in its 2D surface band structure. Although the theoretical solid-state models that the TI field has produced are powerful and unique, their novel emergent physical properties are not universally observed in every sample. These materials are difficult to grow and maintain under ambient conditions. Growths tend to either not be applicable to wafer-scale production or produce high polycrystallinity, and all samples experience natural oxidation, band bending, and intrinsic $n$-doping, which generates spin-degenerate or bulk conduction.

This thesis contains a primer on topologically non-trivial materials, and two studies aimed at understanding and minimizing defects at the surface of Bi$_2$Se$_3$. In the first, the aging process of Bi$_2$Se$_3$ when exposed to air at room temperature is investigated. The time scale and topographic changes of the oxidation process at micromechanically exfoliated surfaces are measured, and an optical model of the bulk and oxide layers are developed. The surface appears to oxidize starting at 2 hours after exfoliation, and continuing through 1.5 weeks, by which time, the oxide layer growth has reached an asymptote of 1.9 nm. New optical characterization methods are developed to monitor the orientation of the crystal (via second harmonic generation) and to measure the oxide growth at the surface (using spectroscopic ellipsometry and the derived dielectric functions of the bulk and oxide layers). The goal of the second study is to assess the use of Se capping and subsequent thermal decapping to preserve a pristine surface and maintain a constant Fermi level. This was measured by annealing samples in a UHV environment to successively higher
temperatures until the Bi$_2$Se$_3$ film decomposed, and measuring the surface crystallinity, topography, surface chemistry, and Fermi level between each anneal. Thermally decapping samples has no measurable effect on crystallinity, minimal effect on surface topography, reveals the expected Bi-Se surface bonds, and retains a mid-gap Fermi level. This may serve as a reference to improve the fabrication process of devices that include Bi$_2$Se$_3$. 
# Table Of Contents

ABSTRACT 2

LIST OF FIGURES 5

CHAPTER 1: Introduction 10

1.1: Moore’s Law 11
1.2: Spintronics 12
1.3: Graphene and High Mobility 2D Materials 16
1.4: Topological Insulators and Bi$_2$Se$_3$ 17
1.5: Defects in Bi$_2$Se$_3$ 20

CHAPTER 2: Topologically Non-Trivial Systems 23

2.1: The Hall Effect and Anomalous Hall Effect 24
2.2: The Integer Quantum Hall Effect and the Anomalous Quantum Hall Effect 26
2.3: Z$_2$ Topological Insulators 41
2.4: Bulk-Boundary Correspondence 48

CHAPTER 3: Bismuth Selenide, its Topological Properties, and its Defects 53

3.1 The Physical and Electronic Structure of Bi$_2$Se$_3$ 55
3.2: Characterizations and Applications of Bi$_2$Se$_3$ 60
3.3: Defects in Bi$_2$Se$_3$ and their Effects on Surface States 68

CHAPTER 4: Characterization Method Description 79

4.1: X-Ray Diffraction 79
4.2: Spectroscopic Ellipsometry 89
4.3: X-Ray Photoelectron Spectroscopy 99
4.4: Transmission Electron Microscopy 105
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5: Atomic Force Microscopy</td>
<td>109</td>
</tr>
<tr>
<td>4.6: Temperature Programmed Desorption</td>
<td>111</td>
</tr>
<tr>
<td>4.7: Low Energy Electron Diffraction</td>
<td>118</td>
</tr>
<tr>
<td>4.8: Scanning Tunneling Microscopy and Scanning Tunneling Spectroscopy</td>
<td>122</td>
</tr>
<tr>
<td>CHAPTER 5: Surface Oxidation in an Ambient Environment</td>
<td>129</td>
</tr>
<tr>
<td>5.1: Sample Preparation</td>
<td>129</td>
</tr>
<tr>
<td>5.2: Measurement Results</td>
<td>130</td>
</tr>
<tr>
<td>5.2: Analysis</td>
<td>154</td>
</tr>
<tr>
<td>CHAPTER 6:</td>
<td>156</td>
</tr>
<tr>
<td>6.1: Sample Preparation</td>
<td>157</td>
</tr>
<tr>
<td>6.2: Measurement Results</td>
<td>158</td>
</tr>
<tr>
<td>6.3: Analysis</td>
<td>175</td>
</tr>
<tr>
<td>CHAPTER 7: Discussion and Conclusions</td>
<td>178</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>186</td>
</tr>
</tbody>
</table>
List of Figures

Ch. 1: Introduction

1-1: Trends in transistors per microprocessor and price per gigabyte of storage.
1-2: Schematic of GMR and spin Hall effects.
1-3: Schematic of a spin-transfer torque device.
1-4: Demonstration of conduction of Dirac fermions in graphene.
1-5: Color map of the spatial variations in charge density of a micromechanically exfoliated graphene flake.
1-6: A transition between the insulating phase and QSH state, occurring at the predicted critical thickness.
1-7: Evidence of the first three dimensional topological insulator

Ch. 2: Topologically Non-Trivial Systems

2-1: A survey of topologically trivial and non-trivial states.
2-2: The first measurement of a topological surface state in the quantum hall phase.
2-3: Laughlin’s flux cylinder, used in his derivation of the quantum hall effect.
2-4: A topological view of the simple insulating state and quantum Hall state.
2-5: Hofstadter’s butterfly.
2-6: The physical configuration and final phase diagram of Haldane’s model.
2-7: Band structures and phase diagram of the Kane-Mele model.
2-8: Schematic of a one dimensional, TRI band structure showing Kramers degeneracy at and between $\pm \Gamma_v$.
2-9: Contour integrations around half the BZI in the Kane-Mele model in order to count the zeros of $P(\tilde{k})$.

viii
2-10: Z\textsubscript{2} index calculations, performed by multiplying the various \textit{i}, at TRI momenta.

2-11: Topological spin orientations in momentum space, as derived from the modified Dirac equation.

2-12: Surface conduction and band schematics of topological materials.

\textbf{Ch. 3: Bismuth Selenide, its Topological Properties, and its Defects}

3-1: Schematic of the Bi\textsubscript{2}Se\textsubscript{3} hexagonal unit cell.

3-2: Inversion of bands with opposite parity in DFT calculations of Bi\textsubscript{2}Se\textsubscript{3}.

3-3: Band structure of Bi\textsubscript{2}Se\textsubscript{3}, including all modifications to the Hamiltonian.

3-4: Spin-polarized Dirac cones, experimentally measured with ARPES.

3-5: Measured and theoretical LDOS distributions on the surface of Bi\textsubscript{0.92}Sb\textsubscript{0.08}.

3-6: Weak antilocalization in TRI materials.

3-7: SdH oscillations and analysis.

3-8: Predicted and measured Kerr rotation in thin Bi\textsubscript{2}Se\textsubscript{3}.

3-9: MBE growth of Bi\textsubscript{2}Se\textsubscript{3} on SiC (0001) with an epitaxial graphene layer.

3-10: Band bending-induced 2DEG, growing over time in Bi\textsubscript{2}Se\textsubscript{3}, as shown by ARPES.

3-11: Generation of the Rashba spin-split 2DEG, as shown by ARPES.

3-12: Some attempts at extrinsically p-doping Bi\textsubscript{2}Se\textsubscript{3} in order to move \(E_F\) toward \(E_D\).

3-13: Bi and Se oxides growing in air at the surface of nanoribbon and bulk Bi\textsubscript{2}Se\textsubscript{3} samples, as measured in XPS at the Bi 4f and Se 3d energies.

3-14: Similar Bi\textsubscript{2}Se\textsubscript{3} surface band response from exposure to different environments, as shown by ARPES.

3-15: A chemically inert Bi\textsubscript{2}Se\textsubscript{3} surface.

\textbf{Ch. 4: Characterization Method Description}
4-1: The Bragg condition indicates at what angles x-rays may be diffracted from parallel crystal planes.

4-2: Demonstration of the relationship between real space planes and reciprocal lattice vectors, using the (231) plane of an orthorhombic lattice.

4-3: Forbidden (002) diffraction in Si from one diamond cubic unit cell.

4-4: Demonstration of omega-2theta experimental configuration and reciprocal space scan direction for a Si sample with a (001) surface.

4-5: The Ewald sphere and the allowed diffraction points of Si.

4-6: Demonstration of planes with a sample that must be tilted with the sample in order for diffraction to be measured.

4-7: Band structure and imaginary dielectric function of germanium.

4-8: Polarization of reflected light.

4-9: Schematic of a dual rotating compensator ellipsometer.

4-10: Emission process for photoelectrons and Auger electrons.

4-11: Example of shake-up lines (s) of the Cu 2p observed in CuO and CuSO₄, but not pure Cu.

4-12: Intensity of electrons emerging from depth $d_{01}+d_{12}$ as a function of their emission and detection angle.

4-13: Schematic of STEM operation. Images can be produced from the BF, ADF, or HAADF detectors.

4-14: AFM operation.

4-15: Schematic of TPD system.

4-16: Typical temperature calibration and full array of TPD-measured molecules.
4-17: Universal mean free path curve of electrons traveling through solids.

4-18: Schematic of crystal diffractions in LEED.

4-19: Diagram of the LEED experimental system.

4-20: Schematic of an STM.

4-21: STM Topography measurement of Si(111) 7×7.

Ch. 5: Surface Oxidation in an Ambient Environment

5-1: Aberration-corrected HAADF STEM images of bulk Bi₂Se₃.

5-2: Omega-2Theta XRD scan of Bi₂Se₃.

5-3: Pole figure of exfoliated Bi₂Se₃ sample used in oxidation study.

5-4: Time-dependent surface topography of Bi₂Se₃.

5-5: Reference topography scan of a contaminated surface.

5-6: XPS and ARXPS results.

5-7: XPS scans of the Bi 4f region taken over various times.

5-8: ARXPS spectra of the Se 3d and Bi 5d region for the sample after ~1.5 weeks of exfoliation.

5-9: ARXPS analysis.

5-10: Variable angle spectroscopic ellipsometry data and results of optical modeling.

5-11: Contrast maps for Bi₂Se₃ on a plain Si substrate and on a Si substrate with 300 nm thermal SiO₂.

5-12: Measured rotational-anisotropy SHG (RA-SHG) scans in a dry air environment immediately, and at 1, 2, 3, 4, 12, and 24 hours after Bi₂Se₃ sample cleaving.

5-13: XRD orientation of the (0115) plane for SHG analysis.

5-14: Schematic of SHG at the surface of Bi₂Se₃ samples.
Ch. 6: Characterization of Bi$_2$Se$_3$ with Selenium protective layer and after Selenium layer removal

5-15: Schematic of 2-step surface aging process.

6-1: SEM images of Se-capped Bi$_2$Se$_3$ films, showing the crystallites on the cap surface.

6-2: Omega-2Theta scan of capped sample SB1.

6-3: Pole figures along the (01\bar{1}5) planes taken on the three rounds of MBE-grown Bi$_2$Se$_3$ samples prior to any thermal treatment, while the Se cap remained.

6-4: XPS results from a capped Bi$_2$Se$_3$ sample.

6-5: Decomposition and desorption of the Bi$_2$Se$_3$ film.

6-6: Effects of UHV thermal decapping procedure on the surface crystallinity of sample Rd3_2, as measured via LEED.

6-7: AFM images of the surface of the first set of samples taken over 5×5 μm$^2$ areas.

6-8: STM images of sample Rd3_1, thermally decapped in situ prior to measurement.

6-9: XPS spectra taken on sample Rd3_2 after TPD decapping.

6-10: Typical STS results and analytical fitting used to derive $E_F$ - $E_D$.

6-11: Histograms of Fermi level shifts in sample Rd3_1 after annealing temperatures of 150 (a), 180 (b), 210 (c), 240 (d), 265 (e), and 295 °C.

6-12: Statistical trends of Fermi level shifts over various anneals and time.
Chapter 1: Introduction

In order to outline the potential impact that $\text{Bi}_2\text{Se}_3$ and other topological insulators may have in future electronic devices, it is useful to examine historical trends. These explain why Si devices are increasingly difficult to improve, and why alternative materials are the subject of intense research. Many of these alternative materials have electrical properties that are increasingly divergent from those seen in conventional semiconductors, and may be applied to devices that do not operate on electrostatic principles. Graphene began a paradigm shift toward materials that host ballistic electron transport states. Topological insulators are a spiritual successor to graphene which maintain the linear Dirac cones that lead to ultra-high electron velocities and ‘massless’ carriers while coupling the spin of the electron to its momentum. This helicity causes a wide variety of physical properties that are not seen in ordinary, topologically trivial (as defined in Ch. 2) materials, such as forbidden backscattering and negative magnetoresistance. While these materials have a series of potentially promising device applications, they have proven to be fragile. Many studies report measurements where the signal comes from defect induced electronic states instead of from the Dirac cones. While the literature has described some of the defects (which include, but are not limited to, surface oxidation and atomic vacancy-induced doping), few studies have focused on understanding the defects themselves. This is a necessary step to deducing the conditions in which topological insulators can behave as theory predicts. Finding the right conditions, or the right protective measures, is a key step toward using $\text{Bi}_2\text{Se}_3$ in new devices.
1.1: Moore’s Law

The economically driven goal of the semiconductor industry has been to adhere to Moore’s law\(^1\), which states that the number of transistors in a microprocessor will double every two years, while cost per device will decrease. This was possible for decades that followed this 1965 prediction. During this time, advancements in device design and architecture improved transistor design, and new memory circuits were coupled with advancements in chip processing such as the development of chemically amplified resists, advanced UV light sources, and chemical mechanical polishing. More recently, attempts to improve scaling and performance have included the implementation of strained silicon, raised source/drain, high-K/metal gates, and a switch from planar devices to 3D fins with multiple gates. These improvements have driven transistor counts per microprocessor from the lower thousands in early 1970s (Intel 4004) to upward of 15 billion in 2016 (NVIDIA GP100 Pascal GPU), with accompanying exponential decreases in price per transistor (Fig. 1-1).

However, lithographic advancement beyond 193 nm light sources have slowed, and gate and source-drain leakage in Si devices impose an inherent limit on geometric scaling. Consequently, advancements beyond the 22-nm node have slowed, and the “double every two years” model has lengthened to every 2.5 or 3 years. In order to surmount the challenges created by parasitic tunneling effects in the front end of the line, researchers and companies are looking to alternative materials and physical phenomena, such as ferroelectricity, functionalized tunneling, and spintronics. The International Technology Roadmap for Semiconductors introduced a reorganization of initiatives and research directives in 2014, dubbed ITRS 2.0, which includes ‘Beyond CMOS’ as a major industry topic.
Spintronics is a field of device research that focuses on spin carriers and the magnetization of materials, rather than charge carriers and electrically polarizable or conductive materials. Some recent work has focused on giant magnetoresistance (GMR) (Fig. 1-2(a)) and tunnel magnetoresistance, the latter of which often takes the form of magnetic tunnel junctions (MTJs). GMR devices modulate conductivity through magnetic materials based on the alignment of the carrier spin to the magnetic moment of its host material, i.e. the resistance a carrier experiences during transport is low if its spin is parallel to the magnetic moment of the material, and high if its spin is antiparallel to the magnetic moment of the material. MTJs are non-volatile memory.

**Figure 1-1:** Trends in transistors per microprocessor (red) and price per gigabyte of storage (blue). The number of transistors per microprocessor increased exponentially for the majority of time since Moore’s law was predicted, and the price per gigabyte of storage has exponentially decreased at a similar rate. These trends have changed in recent years, due to inherent limits in silicon semiconductor manufacturing and performance. Fit lines are proportional to $1.4196^\text{year}$ and $0.5635^\text{-year}$ for counts and price, respectively, giving just over a doubling of counts and a decrease to just under $1/3$ price every 2 years during the given years.

1.2: Spintronics

Spintronics is a field of device research that focuses on spin carriers and the magnetization of materials, rather than charge carriers and electrically polarizable or conductive materials. Some recent work has focused on giant magnetoresistance (GMR) and tunnel magnetoresistance, the latter of which often takes the form of magnetic tunnel junctions (MTJs). GMR devices modulate conductivity through magnetic materials based on the alignment of the carrier spin to the magnetic moment of its host material, i.e. the resistance a carrier experiences during transport is low if its spin is parallel to the magnetic moment of the material, and high if its spin is antiparallel to the magnetic moment of the material. MTJs are non-volatile memory.
structures in which two ferromagnetic layers are separated by an insulator, and spin carriers may tunnel between the magnetic layers depending on their relative alignments. Switching is controlled by spin-transfer torque-induced magnetization reversal of the ‘free’ ferromagnetic layer, causing low tunneling amplitudes for the antiparallel configuration, and high tunneling amplitudes for the parallel configuration. This can be seen in rectangular H-field hysteresis loops, where a critical field amplitude is necessary to switch the device. The tunnel magnetoresistance ratio, defined as the difference ratio of resistances of parallel and antiparallel magnetization configurations \((R_{AP} - R_P)/R_P\) has been reported as high as 600% at room temperature. While this does not yet achieve the on/off ratio required in modern integrated circuits, it demonstrates a promising alternative paradigm for logic and memory devices.

A significant obstacle to generating MTJ devices is the difficulty in generating high amplitude directed magnetic switching fields in nanoscale geometries. A standard example of this process is electron conduction through a ferromagnetic or antiferromagnetic material, from which the gathered output electron has a spin aligned to the magnetic moment of its conducting material.
These spin-oriented carriers may then be collected and their collective moments may suffice to switch the write layer of the MTJ stack (Fig. 1-3). More recently, efforts have been made to harness the giant spin Hall effect (Fig. 1-2(b)), allowing more efficient conversion from electrical to spin conduction. Spin Hall angles ($\Theta_{\text{SH}} = J_s/J_e$) have improved from reports of 0.0037 in 2007 to 0.5 in $\beta$-W in 2016. This > 2 orders of magnitude increase indicates that novel physics, e.g. giant spin Hall effect and quantum spin Hall effect, may provide solutions to ongoing efforts in spintronics device development.

**Figure 1-3:** Schematic of a spin-transfer torque device. Spin current from a reference layer with magnetic moment $\mathbf{M}_r$ along $+\hat{y}$ provides spin torque, which may flip the magnetization of the free layer if the damping force is overcome. Weaker spin injections may result in magnetization precession around the effective magnetic field ($\mathbf{H}_{\text{eff}}$). By changing the current polarity, the magnetization can be switched between parallel and antiparallel with respect to the reference layer. Adapted from [8].
1.3: Graphene and 2D High Mobility Materials

Another relatively recent advent in device physics came in the form of graphene in 2004. This marked the beginning of a shift toward research in purely quantum effects implemented in electric field effect devices. Importantly, it displayed the first instance of Dirac cones in solid electronic

Figure 1-4: Demonstration of conduction of Dirac fermions in graphene. (a) Dependence of fundamental Shubnikov-de Haas oscillation (SdHO) field density on the gate voltage-modulated carrier density of the graphene film. This follows the expected trend for 2D systems as $B_F = f^2$, where $f$ is the degeneracy (equals 4 in graphene for spin and sublattice), and $\Phi_0$ is the magnetic flux quantum. (b) Tracking the inverse applied field as a function of SdHO index for graphene (red) and 5 nm graphite (black). The horizontal shift in graphene can be explained by its non-trivial Berry phase. (c) Relative change in longitudinal conductivity as a function of temperature shows a steeper drop-off for high gate voltage and high $n$ (lower) than for lower $n$ (upper), indicating higher mobilities for states near the Dirac point energy. (d) Carrier mass as a function of doping level or $E$. Data points conform to $m_L = E/c^2$, and from the function $m_L = (\hbar^2/2\pi)dS(E)/dE$, where $S$ is the area of Fermi energy orbits, the relationship $E = \hbar kc$, may be derived. This proves the existence of a linear cone structure of graphene, and fits to the data yield $c_s = 1 \times 10^6$ m/s. (e) Schematic of Dirac cone, where the Dirac point is the center of the linear dispersion, and where $n = 0$. Adapted from [10].
systems. These Dirac cones are local regions in the first Brillouin zone (BZ) where valence and conduction energy bands are connected at a point, and the energy of the upper and lower bands depends linearly on the momentum distance from the cone center, \( i.e. \, E(\vec{k}) - E_0 \propto |\vec{k} - \vec{K}_i| \), where \( E_0 \) is an offset and \( \vec{K}_i \) is the Dirac point location (Fig. 1-4). This semimetallic connection is nontrivially different from typical materials with parabolic bands, in that the linear dispersion gives rise to a “massless” electron state, with early Fermi velocities measured at \( 1 \times 10^6 \, \text{m/s} \), and mobilities measured at \( 230,000 \, \text{cm}^2/\text{Vs} \).

Despite the exciting physical properties of graphene, its implementation in devices has been difficult. Perhaps predictably, growth of 2D carbon with controlled doping levels, few structural defects, and minimal topography has been met with challenges. Though exfoliation has provided a reliable means to obtaining high quality flakes that are utilizable at the lab scale, efforts to grow graphene at wafer scale (on hexagonal SiC, Ru and Cu, (110)-Ge, hBN, etc.) have not

**Figure 1-5:** Color map of the spatial variations in charge density of a micromechanically exfoliated graphene flake. These may be caused by minor disorder in the graphene topography, trapped charge in the substrate oxide, doping via exposure to photoresists chemicals, atmospheric contaminants, etc. Adapted from [12]
produced large pristine domains consistently enough for implementation in current integrated circuits. Further, methods to transfer epitaxial graphene typically result in heterogeneous doping of the film, and can lead to localized charge ‘puddles’ \(^{12}\) (Fig. 1-5), which negatively impact control of electrons at or near the Dirac point. Research into production is ongoing, but it behooves the field of device physics to try alternative routes, where true 2D materials may be avoided. This option presented itself in 2009 in the form of Bi\(_2\)Se\(_3\), a topological insulator (TI) material. \(^{18}\)

1.4: Topological Insulators and Bi\(_2\)Se\(_3\)

Novel research that occurred soon after the discovery and characterization of graphene came in 2005, with the theoretical prediction of the quantum spin hall (QSH) state, and a topological classification to describe it. \(^{19,20}\) The topological nature of this state refers to a Hamiltonian that generates the same band gap as ordinary dielectric materials, but cannot be smoothly transformed into any “topologically trivial” states without closing the band gap, which results in exotic interface states (explained in detail in Ch. 2). These topological surface states (TSSs), like the K\(_i\) states of graphene, occur in Dirac cones, but unlike graphene their spins are locked perpendicular to their momentum direction, and they are protected against weak perturbations by the topology of the bulk. The central portions of the theory behind topological classification of solid systems had been developed decades earlier (and have recently won a Nobel prize) \(^{21-23}\), and some of its consequences had been demonstrated via the quantum Hall effect. \(^{24}\) The more recent development of the Z\(_2\) topological theory of Kane and Mele contains insights on time reversal invariant (TRI) systems, which lead to explosive growth in the field of topological insulators (TIs). The first physical realization of this effect was shown in HgTe quantum wells that displayed the QSH effect above a predicted critical thickness (Fig. 1-6). \(^{25,26}\) Quantum wells thinner than 6.3 nm contain
trivial insulator state, but when the thickness increases above that level, the band structure of the system changes, and the system shifts into a QSH state. Another major advance in the TI field from Charles Kane came with the development of $Z_2$ theory for 3D systems (another notable difference from graphene physics) with inversion symmetry. This significantly simplified topological calculations, and facilitated efforts to find and grow 3D TI materials, but initial efforts were largely focused on non-stoichiometric compounds such as $\text{Bi}_{1-x}\text{Sb}_x$. These materials demonstrated the first examples of Dirac cones on 3D surfaces, as predicted by $Z_2$ theory and shown in angle resolved photoemission spectroscopy (ARPES) in 2008, but had complex band structures with small gaps. In 2009, calculations of $\text{Bi}_2\text{Se}_3$ using the same inversion-symmetric $Z_2$ theory revealed that it had a simple band structure and larger gap than any previously discovered TI. Its surfaces were predicted to contain a single Dirac cone at the gamma point, and a (relatively)

**Figure 1-6:** A transition between the insulating phase and the QSH phase, occurring at the predicted critical thickness. Curve I shows resistance data from a thin quantum well, below the critical phase transition thickness of 6.3 nm. The low conductivity level of this film indicates its insulating phase. Curves III and IV are from measurements of films above the critical thickness, and show quantized conductivity levels equal to the predicted values of the quantum hall state. This structure was the first manifestation of the topological QSH effect. Adapted from [25].
large band gap of 0.3 eV. These were quickly realized experimentally\textsuperscript{30}, and research into the ‘second generation’\textsuperscript{31} of TI materials was born.

\(\text{Bi}_2\text{Se}_3\) contains the surface spin-polarized Dirac cones inherent to 3D TI materials. At these surfaces electron conduction is helical, \textit{i.e.} the momentum and spin are locked at a perpendicular angle. These band structures generate multiple interesting properties (see Ch. 3 for more complete descriptions). As in the case of graphene, these Dirac cone states have ultra high mobility due to the linear energy dispersion. The spin-polarized nature of these Dirac cones makes backscattering impossible, as an electron’s spin must to flip in order for it to inhabit a state with the opposite momentum. Carriers in these systems display an unusual property called weak antilocalization (WAL) that results in negative magnetoconductivity.\textsuperscript{32} This emerges from the phase difference

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1-7.png}
\caption{Evidence of the first three dimensional topological insulator, \textit{Bi}_{1-x}\textit{Sb}_x. (a) ARPES measurements show the graphene-like Dirac cone at the surface of this material. (b) A phase diagram demonstrates the critical Sb concentration that generates the topological insulator phase. The intersection of bands gives rise to a ‘negative band gap’, a common trait among topological materials. Adapted from \cite{28}.}
\end{figure}
that they acquire during diffusive transport, wherein traveling around a closed loop leads to destructive interference. These states also generate unusual Shubnikov-de Haas data.\textsuperscript{33} These arise in high field magnetoconductivity tests that measure resistance changes while the incident field shifts Landau level energies through the Fermi level. They provide information regarding the electrical mobility and density of the system, but also generate a clear sign of the system’s topology in the phase of the oscillations. The surfaces of Bi\textsubscript{2}Se\textsubscript{3} demonstrate an unusual half-integer quantized Hall conductivity, although this is difficult to measure as it requires isolating conduction to a single surface.\textsuperscript{34} They create an unusual optical behavior called the topological magnetoelectric effect that results in giant Kerr rotations. Ultimately, they may be used to generate Majorana fermions for quantum computing purposes, although research in this direction is at a nascent stage.\textsuperscript{35}

This array of unusual physical behaviors in Bi\textsubscript{2}Se\textsubscript{3} makes for promising potential applications. Its spin polarized Dirac cones are of particular interest to the spintronics community, where spin-polarized carriers are needed en masse for many device architectures.

1.5: Defects in Bi\textsubscript{2}Se\textsubscript{3}

Although the TSSs of Bi\textsubscript{2}Se\textsubscript{3} are necessary if these materials are going to be applied to spintronics, crystal growths have so far been plagued by defects. Samples produced for academic purposes are often grown from liquid melts. These are of high quality and large volume, displaying flat cleaved surfaces of > 1 mm\textsuperscript{2}.\textsuperscript{36} However, this growth methodology is not compatible with modern semiconductor fabrication techniques. More recently, thin film deposition has been attempted with molecular beam epitaxy (MBE).\textsuperscript{37} These films are more polycrystalline and have rougher surfaces than their bulk-grown counterparts. Based on available information, we believe
that all MBE films have triangular pyramidal surface structures. These topographic features are composed of vertically stacked QLs, and usually have heights on the order of > 10 nm above background. This roughness is a complication in the fabrication process, and may additionally serve as a detriment to TSS conduction, as these electrons are confined to the surface and may not be able to smoothly traverse a complex topography. Surfaces are also chemically fragile. The vapor pressure of Se is high, and Se vacancy formations have been shown to cause \( n \)-doping in Bi\(_2\)Se\(_3\) crystals.\(^\text{38}\) When degenerately doped, Bi\(_2\)Se\(_3\) may conduct metallic carriers in the bulk and on the surface. These non-TSSs produce Shubnikov-de Haas oscillations at a different phase, and exhibit spin-degenerate conduction. Moreover, since the density of states of a bulk band is exponentially higher than that of a 2D system, these may overwhelm the electrical conduction signals created by TSSs. Perhaps the most detrimental effect of doping emerges from the energy dependence of the Dirac cone helicity. Carriers that inhabit states below the Dirac point in the surface dispersion have opposite helicity to the carriers that inhabit states above the Dirac point. Thus, whether \( E_F \) moves away from \( E_D \) (the Dirac point energy) in the negative direction due to \( p \)-doping and causes fewer TSSs to be filled, or moves away from \( E_D \) in the positive direction and causes the spin polarizations of occupied states in the upper and lower cone halves to cancel, doping poses a severe threat to the helicity of TSS conduction. These crystals also naturally oxidize. The first QL of Bi\(_2\)Se\(_3\) and other similar van der Waals gapped materials\(^\text{39}\) has been shown to expand to roughly double their initial thickness during this oxidation process. This oxide exacerbates the surface roughness, correlates with a surface two dimensional electron gas (2DEG) formation\(^\text{40}\) (which creates similar spin-degenerate signals to bulk states), and may create an electrically insulating overlayer on top of the desired TSSs. While it has been frequently measured in attempts to characterize the Bi\(_2\)Se\(_3\) crystal, few reports contain information about its electrical or optical properties.
These defects pose a clear challenge to harnessing the unique properties of Bi$_2$Se$_3$ (and other structurally similar TIs such as Bi$_2$Te$_3$ and Sb$_2$Te$_3$)$^{18}$, and must be studied and mitigated before TI devices can be realized. However, as it is not practicable to fully comprehend the physical phenomena that topologically non-trivial systems generate or the impact that defects have on their unique properties, the following two chapters contain detailed explication of the theoretical and experimental foundations and recent advances in the field of TIs.
Chapter 2: Topologically Non-Trivial Systems

In order to understand the physics of the topologically protected surface states in Bi$_2$Se$_3$, it is useful to examine the early discoveries and theoretical developments in the field of topologically nontrivial systems. There are many systems that display topological distinction, each with their own physical origins, and they each are informative about the physics of modern topological insulator materials such as Bi$_2$Se$_3$. The 2D electron gas (2DEG), subject to a voltage bias and a perpendicular magnetic field, creates the classical Hall effect in conductors and, at low temperature, the quantum Hall (QH) effect in semiconductor quantum wells. Although these systems both show transverse conduction in response to longitudinal bias, but result from wildly discrepant physics. The anomalous quantum Hall (AQH) effect was theoretically predicted within a decade of the development of QH theory, and the quantum spin-Hall (QSH) came two decades later. These two systems are topologically distinct from vacuum, as well as most dielectric materials, and require no external potentials to generate topological surface states in linear electronic bands. These are all useful examples of unusual two-dimensional electronic behaviors and phases (Fig. 2-1) that result from topological physics, as well as its importance, in modern solid-state physics. Each theoretical description builds on the theory that explains the previous, leading to the $Z_2$ theory, which was initially developed to explain the QSH effect in graphene but underlies the topological nature of many 3D materials such as Bi$_2$Se$_3$ as well. In this way, the discussion in this chapter will proceed from the theoretical foundations describing the integer QHE for a 2DEG in a magnetic field at low temperature to topological insulators where the protected surface states are not a result of an external magnetic field.
2.1: The Hall Effect and Anomalous Hall Effect

In the classical Hall effect, the Lorentz force causes the conducting 2DEG carriers to arch in their paths according to \( \vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \). For a sample with electrons drifting with \( \vec{v} = v_x \hat{x} \) at volume density \( n \), subject to a magnetic field \( \vec{B} = B_z \hat{z} \), equilibrium is reached where the separated carriers generate a \( \hat{y} \)-direction bias \( v_x B_z = E_y = \frac{V_H}{w} \). Here, \( w \) is the sample width and \( V_H \) is the generated Hall voltage

\[
V_H = w v_x B_z = \frac{n q v_x A B_z}{n q d} = \frac{I_x B_z}{n q d} = R_H \frac{I_x B_z}{d} \tag{Eq. 2-1}
\]

This has provided a method for measuring net carrier density (where negative \( R_H \) indicates majority electron conduction and positive indicates majority hole conduction) for samples of known thickness \( d \) and measured longitudinal current \( I_x \), in the form of \( R_H = \frac{1}{qn} \). Transverse resistivity in this system is a continuous and linear function of \( \vec{B} \), and can be simply derived from a Drude model:

\[
\frac{d^2}{dt^2} \vec{x} = -\Gamma \frac{d}{dt} \vec{x} + \frac{q}{m_{\text{eff}}} \left( \vec{E} + \vec{v} \times \vec{B} \right) \tag{Eq. 2-2}
\]

where \( m_{\text{eff}} \) is the effective carrier mass and \( \Gamma \) is the average time between collisions, \( 1/t \). Using the relationship \( \rho \vec{J} = \vec{E} \), the resulting resistivity is

\[
\begin{pmatrix}
 1 & c & 0 \\
 c & 1 & 0 \\
 0 & 0 & 1
\end{pmatrix}
\tag{Eq. 2-3}
\]

where \( \rho = \frac{m_{\text{eff}}}{n q^2} \) and \( c = \frac{e B_z}{m_{\text{eff}}} \). This is a useful conceptual precursor to the topologically distinct IQHE, wherein transverse resistivity in the 2DEG of a semiconductor quantum well has a similar dependence on \( \vec{B} \), but is quantized step function rather than a continuous one.
The AHE is similar, in that the Lorentz force generates arching carrier paths, and with them, transverse voltages and resistivities. However, this effect takes place in the absence of an externally applied magnetic field. It is instead caused by conduction through a material with an inherent

**Figure 2-1:** A survey of topologically trivial and non-trivial states. (a) The insulating state is characterized by an energy gap separating the occupied and empty electronic states, which is a consequence of the quantization of the energy of atomic orbitals. (b) In the quantum Hall effect, at the edge, electrons execute “skipping orbits” as shown, ultimately leading to perfect conduction in one direction along the edge. (c) The edge of the “quantum spin Hall effect state” or 2D topological insulator contains left-moving and right-moving modes that have opposite spin and are related by time-reversal symmetry. (d) The surface of a 3D topological insulator supports electronic motion in any direction along the surface, but the direction of the electron’s motion uniquely determines its spin direction and vice versa. Adapted from [41].
magnetic moment \( \vec{M} \), such as a ferromagnet. This is often a much stronger effect than the classical Hall effect.\(^{42}\)

2.2: The Integer Quantum Hall and Anomalous Quantum Hall Effect

The integer QH effect (IQHE) is a physically complex phenomenon, and a simple preemptive conceptual encapsulation may help the reader. In addition to the localized states found in common insulators, where the Fermi level, \( E_F \), inside the band gap ensures that all valence states below the gap are filled and non-local scattering effects are prevented, the QH effect displays simple harmonic oscillator (SHO) states when a magnetic flux \( \Phi = \vec{B} \cdot \vec{A} \) is applied to the sample. Electrons in these states can be understood as having a circular motion with direction given by the right-hand rule, circumscribed with radius \( l_s = \frac{h}{eB} \). Because of the required uniform directionality of these emergent SHO states, even when \( E_F \) is in a gap, the effect of edge scattering is to push an electron into an adjacent circular orbit (Fig. 2-1 (b)), thus creating robust edge conduction in an otherwise insulating material.

The integer QH effect was first predicted for a 2DEG by Ando et al. in 1975\(^{43}\) and demonstrated in 1981 by von Klitzing et al.\(^{24}\). Ando’s theoretical work painted the picture of a 2DEG with robust quantized Hall conductivity in the face of short-ranged scatterers, and von Klitzing’s experiment in a metal-oxide-semiconductor field-effect transistor at liquid helium temperature (1.5 K), and with a strong external magnetic field (15 T) confirmed it. In this experiment, raising the gate voltage while maintaining the strong external magnetic field lead to discrete changes in Hall voltage (Fig 2-2). This transition occurred at gate bias levels coincident with the Landau level (LL) energy spacing \( E_n = \hbar \omega_c (n + \frac{1}{2}) = \frac{\hbar eB}{m_c} (n + \frac{1}{2}) \), and generated quantized
conductivity changes when $E_F$ crossed through LLs

$$\sigma_{xy} = \frac{N_L e}{B} = \frac{e}{B} \left( \frac{neB}{h} \right) = \frac{ne^2}{h}$$

(LL degeneracy $N_L = \frac{eB}{h}$ will be derived later). While von Klitzing’s original work described the conductivity, resistivity, and Hall resistance accurately (and successfully provided an accurate measurement of the fine structure constant $\frac{eC}{2nR_H}$ to a new level of precision), a novel quantum mechanical description by Laughlin$^{44}$ and Thouless et al.$^{23}$ was required to fully elucidate the quantum mechanical behavior of the 2DEG, i.e. to explain the quantized plateaus of Hall conductivity and resistivity, to explain the edge transport generated by dissipationless “skipping” states, and most importantly to classify the IQHE by a distinct topology of electronic bands in the first Brillouin

Figure 2-1: The first measurement of a topological surface state in the quantum hall phase. Recordings of the Hall voltage $U_H$, and the voltage drop between the potential probes, $U_{pp}$, are shown as a function of the gate voltage $V_g$ at $T=1.5$ K and constant magnetic field of 18 T in the work of von Klitzing et al. The inset shows the experimental setup. The integers on the gate voltage axis give Landau level energies, between which longitudinal voltage is 0 and Hall voltage is at a plateau. At the Landau levels, $E_F$ is effectively in a small continuous energy band, thus allowing longitudinal conduction. As $E_F$ increases inside LLs, the QHE increases in strength. Adapted from $[24]$. 
zone. As this provides a sound basis for TI physics, it is valuable to show some of the details of Thouless’ and others’ theoretical work.

The integer QH effect occurs in an initially free electron gas in a magnetic field can be quantumlectron gas in a magnetic field can be quantum mechanically expressed as:

\[ E\psi = \hat{H}\psi = \frac{\hat{p}^2}{2m} \psi = \frac{1}{2m} (\hat{p} + e\hat{A})^2 = \frac{1}{2m} (\hat{p}_z^2 + (\hat{p}_x + eBx)^2) \psi \]

\[ \text{Eq. 2-4} \]

The \( \hat{z} \)-component is ignored, as it generates a trivial kinetic energy contribution, and additional scalar potentials are not considered. Here, a Landau gauge choice of \( \hat{A} = Bx\hat{y} \) (for a vertical magnetic field \( \nabla \times \hat{A} = -B = B\hat{z} \)) results in a Hamiltonian that only includes a momentum operator in the \( \hat{y} \)-direction. Thus, a wave function of the form \( \psi = u(x)e^{iy} \) is valid, and generates

\[ E\psi(x) = \left( \frac{\hat{p}_z^2}{2m} + \frac{c}{2} x^2 + \frac{eB}{m} x + \frac{2iqB}{m} \right) u(x), \text{ where } c = \frac{qB}{m} \text{ and } \ell_n^2 = \frac{\hbar}{eB}, \]

which is a familiar quantum harmonic oscillator equation with an oscillation center of \( x_0 = k\ell_n^2 \). The energy levels of this 2D system are \( E_n = \hbar\omega_c (n + \frac{1}{2}) \), which are separated into discrete Landau levels (LLs). Each LL contains states with degeneracy \( N \) that can be calculated from the Gaussian spatial distribution of a SHO with wavefunction \( u(x) = H_n(x - x_0)e^{-\left(\frac{(x - x_0)^2}{2\ell_n^2}\right)} \), where the exponential denominator \( 2\ell_n^2 \) is taken to be the oscillation range:

\[ N = \frac{L_y}{2\pi} \int _{-\ell_n^2/2}^{\ell_n^2/2} g_s dk = \frac{L_yL_x}{2\pi\ell_n^2} \frac{eBA}{2\pi\hbar} = \frac{AB}{\Phi_0} \]

\[ \text{Eq. 2-5} \]

where \( g_s = 2 \) is the electron spin degeneracy and \( \frac{\hbar}{e} \) is called the quantum flux. The third quantity shows that degeneracy can be seen as single state taking up the area of a circle with radius \( \ell_n^2 \), and the last shows that each degenerate state is created by a single flux quantum.
The addition of an electric field in the $\hat{x}$-direction to the Hamiltonian,

$$H = \frac{p_x^2}{2m} + \frac{m}{2} \left( x + kl_B^2 \right)^2 e x,$$

causes carriers to drift. Completing the square with the external term,

$$H = \frac{p_x^2}{2m} + \frac{m}{2} \left( x + \left( \frac{kl_B^2}{B} \right) \right)^2 + e kl_B^2 \frac{m}{2B^2},$$

a new SHO center can be defined as $x_i = \frac{kl_B^2}{B} + \frac{eEx}{mE_B^2}$. In addition to this shift from the SHO center, the energy spectrum $E_{nk} = \hbar \omega - eEx + \frac{mE_B^2}{2B^2}$ results in a drift velocity of

$$v = \frac{1}{\hbar} \frac{\partial E_{nk}}{\partial k} = \frac{eE}{B}$$

Thus, analogously to the transverse arcing trajectories in the classical case caused by the Lorentz force, the quantum Hall effect results in transverse drift when a longitudinal field is applied.

The question of the origin of quantized conductivity was answered by Laughlin in a physically valid thought experiment. He imagined a Hall configuration in which a cylinder that contained a 2DEG was penetrated by a perpendicular magnetic field (Fig. 2-3). For this geometry, when the
applied magnetic field is perturbed by an additional magnetic vector potential \( A\hat{v} \), the SHO center shifts linearly as \( x_2 = kl_2 + \frac{A}{eB} \frac{A}{B} = x_1 \frac{A}{B} \). This carries through to the energy dispersion, 
\[
E_{n\lambda} = \hbar\omega_c \left(n + \frac{1}{2}\right) - eE \left(x_1 - \frac{\Delta A}{B}\right) + \frac{me^2}{2B^2},
\]
thus generating a linear relationship to the perturbation. Due to the geometry of this experiment, however, there is a gauge restriction on values of \( A \), which through the Aharonov-Bohm effect, must return the phase of the conducting electrons to their original values upon completion of a loop around the cylinder, \( \text{i.e. } e^{i2\pi nL/\hbar} = e^{i2\pi n} \), or 
\[
A = \frac{nh}{eL} = n\frac{a}{L}.
\]
Thus, the perturbations added must be integer multiples of the magnetic flux quantum divided by the circumference of the loop, \( L \). An addition of \( \alpha \) maps the system to its original state, and results in the transfer of \( n \) electrons from one edge to the other. This \( n \) is a valid quantum variable for defining states of particles looping the cylinder, and mirrors the degeneracy found in LLs. The current can therefore be related to the induced transverse voltage as \( I = \frac{n e V_H}{\alpha} \), resulting in the quantized hall conductivity, \( \sigma_{xy} = \frac{I}{V_H} = n \frac{e}{\alpha} \). Comparing to the resistivity of classical Hall effect \( \sigma_{xy} = \frac{B}{e} = \frac{N e}{N e} \) (temporarily using \( N = \frac{N}{A} \) as areal carrier density for clarity) this resultant quantized resistivity \( \sigma_{xy} = \frac{1}{\sigma_{xy}} = \frac{n}{ne} \) is a very reasonable and apparent quantum mechanical analogy.

The most important aspect of the IQHE to note, however, is the topological nature of its electronic states. This can be shown through the Berry phase of the system, which is the phase generated by adiabatic transformations in a parameterized Hamiltonian system \( H(\vec{R}(t)) \), where the wave function and phase changes are given by \( |\psi(t)\rangle = e^{i\int_{t_0}^t H(\vec{r}(t'),\vec{R}(t)) dt'} |n(\vec{R}(t))\rangle \). The time-dependent Schrödinger equation, \( \hat{H}\psi = i\hbar\partial_t \psi \) (where \( \partial_t \), is the partial time differential) produces:
where $\bar{A}_n(\bar{R})$ is the Berry connection, and is not gauge independent (as can be checked by computing $\bar{A}(\bar{R})$ for $i = e^{i\theta}$), where as the Berry curvature $\bar{\Omega}_n(\bar{R})$ is gauge independent, and thus physically measurable. This Berry phase is a purely geometric parameter that describes the phase acquired by a state that is translated around the entire surface extended by Hamiltonian parameters $\bar{R}$, such as the momentum generated by Bloch waves, $\bar{k}$. This value is only calculable for a surface that’s closed over the first Brillouin zone for a given $E_F$, as is the case for insulators, but is not the case for metals, wherein the Fermi level intersects band boundaries. The Berry curvature can also be expressed as

$$-i\bar{\Omega}_n(\bar{R}) = \left\langle \bar{\nabla}_\nu n(\bar{R}) \times \bar{\nabla}_\mu n(\bar{R}) \right\rangle = \left\langle \frac{\partial n(\bar{R})}{\partial R_a} \frac{\partial n(\bar{R})}{\partial R_b} \right\rangle - \left\langle \frac{\partial n(\bar{R})}{\partial R_b} \frac{\partial n(\bar{R})}{\partial R_a} \right\rangle$$

where Einstein summation notation is implied for all elements of $\bar{k}$, and from which its identity as a surface curvature is more apparent. The integral of this value over a compact surface, as seen from the Gauss-Bonnet formula $\iint_S K \, d\mathcal{A} = 2\pi \chi$ (where $K$ is the Gaussian curvature of the surface, $K \, d\mathcal{A}$ gives the magnitude of the curvature at a surface point, and $\chi$ is an even integer called the Euler character of the surface), is a quantized value that depends on the topology of the surface.
Using this Gauss-Bonnet formulation, one can classify smooth shapes distinctly, as in the 2-sphere and all of its smooth deformations, which have an Euler character of $c = 2$, whereas the torus and all of its smooth deformations have an Euler character of $c = 0$. This integrated curvature, which results in quantized Euler characters, can be physically visualized as differentiating the surfaces by the number of different ‘holes’ or ‘handles’ they contain (Fig. 2-4), or by the required number of points where vector fields on the surface of the shape must converge (singularities, as demonstrated at the poles of a sphere predicted by the “Hairy Sphere” theorem, or by the smooth field lines that cover a torus). The interaction of spaces that have differing topologies, and the resultant interface states, is the essence of the conductive states that are generated in topological insulators.

Figure 2-4: A topological view of the simple insulating state and quantum Hall state ((a) and (d)). Although both are separated by gaps ((b) and (e)), the curvatures of their entire magnetic Brillouin zones are categorically different. In the same way that the orange ($\chi = 2$) and donut ($\chi = 0$) ((c) and (f)) cannot be smoothly morphed into one another without closing the donut hole, which is a discontinuous process that creates undefined curvature, the Hilbert spaces that define the insulating state and quantum hall state cannot be smoothly connected without a sudden band gap closure. Because BZ1 can’t be as easily shown to have holes or handles, it is sometimes easier to note the singularities (or lack there of for the torus (h)) in the wave function, as shown at the poles of the sphere (d). This is the essence of topological differentiation and the Chern number. Adapted from [34] and [45].
Returning to the IQHE, while origin of the quantization of the Hall conductivity has been answered, the quantum Hall state’s behavior in true periodic lattice conditions and general topological character have still not been accounted for. For these properties, we turn to the work of Thouless et al.\textsuperscript{23}, who used a Bloch wave in both $x$- and $y$-dimensions, and rather than perturbing the vector potential, used the Kubo formula to derive the quantum Hall state. Accordingly, Thouless’ Bloch Hamiltonian is

$$\hat{H} = \frac{1}{2m}(\hat{p}_x + i \hbar k_x)^2 + \frac{1}{2m}(\hat{p}_y + i \hbar k_y + eBx)^2 + U(x,y) \tag{Eq. 2-9}$$

where $U(x,y)$ is the periodic potential of a lattice, and the conductivity of a sample can be computed as follows:

$$\sigma_{xy} = \frac{i e^2 \hbar}{4 \pi} \sum_{n \neq 0} \frac{\langle \psi_n \mid \nabla_{k_x} \mid \psi_n \rangle \langle \psi_n \mid \nabla_{k_y} \mid \psi_n \rangle - \langle \psi_n \mid \nabla_{k_y} \mid \psi_n \rangle \langle \psi_n \mid \nabla_{k_x} \mid \psi_n \rangle}{(E_n - E_0)^2} \tag{Eq. 2-10^6}$$

Here, the summation of excited states can be taken to represent the bands in a typical solid-state system. The velocity operators can be substituted by $v = \frac{1}{\hbar} \frac{\partial H}{\partial k}$, which produce expected value terms in the numerator that are in the form equal of the derivative of a first-order perturbation theory state: $\sum_{n \neq 0} \langle \psi_n \mid \frac{i \hbar \mu}{E_n - E_0} \mid \psi_n \rangle = \frac{\delta \psi}{\delta k_x}$. The inner product of these $k$ perturbation terms results in

$$\sigma_{xy} = \frac{i e^2}{\hbar \alpha} \left( \frac{\delta \psi}{\delta k_x} \frac{\delta \psi}{\delta k_y} - \frac{\delta \psi}{\delta k_y} \frac{\delta \psi}{\delta k_x} \right),$$

and this is exactly the form of the Berry curvature generated by the quantum variables $k_x$ and $k_y$.

$$\tilde{\Omega}(\vec{k}) = \frac{i}{\hbar} \left( \frac{\delta \psi}{\delta k_x} \frac{\delta \psi}{\delta k_y} - \frac{\delta \psi}{\delta k_y} \frac{\delta \psi}{\delta k_x} \right) \tag{Eq. 2-11}$$

While this is the correct form for all bands, it is useful to separate contributions from individual bands $n$, as in a real system, only some bands are filled (additionally, if all bands were filled, the
system would necessarily be topologically trivial\textsuperscript{46}). Thus, the Hall conductivity in a periodic lattice is directly proportional to the sum of Berry curvatures for filled bands,

\[ \sigma_y = \frac{e^2}{h\hbar_0} \sum_n \tilde{\Omega}_n(k) \]  

Eq. 2-12

This is the Hall conductivity for a given state \( k \). This readily becomes topologically significant when we look at the expected Hall conductivity, which requires averaging over all \( k \) values:

\[ \sigma_y = \frac{4\pi^2 e^2}{h} \sum_n \int d^2 k \tilde{\Omega}_n(k) = \frac{e^2}{h} \sum n \chi_n \]  

Eq. 2-13

This is more commonly written in terms of the Chern number, \( C_n = 2 \), named after Shiing-Shen Chern for his work on winding fiber bundles\textsuperscript{47}, giving the exact expected form for the IQHE,

\[ \sigma_y = \frac{e^2}{h} C_n \]  

Eq. 2-14

where each \( C_n \) is itself an integer (although it may be 0 in many cases) and the sum is consequently quantized. This is the result found by von Klitzing. Completing the calculation of the Chern number for a periodic potential is difficult and complex, and can be found in the work of Thouless \textit{et al.}\textsuperscript{23}

The distinct phases given by these topological distinctions may be elegantly plotted with the Hofstadter butterfly\textsuperscript{48} (Fig. 2-5).\textsuperscript{49} In an insulator, quantum hall phase emerges from any non-zero magnetic flux and \( E_F \).

The next logical step in describing topological systems is to look at the AQHE through the work of Duncan Haldane\textsuperscript{22}. His is a slightly generalized\textsuperscript{50} graphitic model\textsuperscript{51}, containing distinct atomic species on the typical diatomic graphene lattice, threaded by periodic antiparallel magnetic fluxes, such that the net flux in any unit cell equals zero (Fig. 2-6(a)). The 2D unit cell is hexagonal,
given by the lattice parameters \( \vec{a}_i = a(1,0) \) and \( \vec{a}_z = a(-\frac{1}{2}, \frac{\sqrt{3}}{2}) \), with the reciprocal lattice \( \vec{b}_i = 2\pi \frac{\vec{a}_i \times \vec{a}_j}{\vec{a}_i \cdot (\vec{a}_j \times \vec{a}_k)} \). This results in important first Brillouin zone (BZ1) corners given by \( \vec{K}_i = \frac{1}{2}(-2\vec{b}_i + \vec{b}_j) \) and \( \vec{K}_z = -\vec{K}_z \), as will be seen in subsequent calculations.

This system, modeled in the second quantized scheme, receives contributions from differential on-site energies \( M \), nearest neighbor (NN) hops \( t_i \), and next nearest neighbor (NNN) hops \( t_z \) that receive a directionally dependent phase contribution from the Aharonov-Bohm effect:

\[
H(\vec{r}) = M \sum_i c_i^\dagger(\vec{A}_i) c_i(\vec{A}_i) - c_i^\dagger(\vec{B}_i) c_i(\vec{B}_i) + t_i \sum_{i,j} c_i^\dagger(\vec{R} + \vec{A}_j) c_i(\vec{R} + \vec{B}_j) e^{i\phi} + t_z \sum_{i,j} c_i^\dagger(\vec{R} + \vec{A}_j) c_i(\vec{R} + \vec{B}_j) e^{i\phi} \tag{2-15}
\]

where \( \vec{a}_i \) and \( \vec{b}_i \) refer to NN and NNN vectors, \( \vec{A} \) and \( \vec{B} \) are atomic sites in the unit cell, and \( \vec{R} \)
may refer to either lattice site (for sites given by $i$). The creation and annihilation operators $c_i^+$ and $c_i$ may be Fourier transformed so that the Hamiltonian may be written in terms of $\vec{k}$:

$$H(\vec{k}) = \begin{pmatrix} c_a^+ & c_b^+ \\ c_a & c_b \end{pmatrix} \begin{pmatrix} M + t_2 \sum_j e^{-i\vec{k} \cdot \vec{a}_j} & t_1 \sum_j e^{-i\vec{k} \cdot \vec{a}_j} \\ t_1 \sum_j e^{i\vec{k} \cdot \vec{a}_j} & -M + t_2 \sum_j e^{i\vec{k} \cdot \vec{a}_j} \end{pmatrix} \begin{pmatrix} c_a \\ c_b \end{pmatrix}$$

Eq. 2-16

The mathematics simplifies with a substitution of Pauli matrices $\vec{\sigma}$. The Hamiltonian and energy may be written as $H = \vec{e}(\vec{k}) I + \vec{a}(\vec{k}) \cdot \vec{\sigma}$, where $\vec{e}$ and $\vec{a}$ are Hamiltonian matrix energy coefficients:

$$H(\vec{k}) = 2t_2 \sum_{j=1,3,5} \left( \cos \phi \cos(k \cdot \vec{\beta}_j) I - \sin \phi \sin(k \cdot \vec{\beta}_j) \sigma_z \right) + t_1 \sum_j \left( \cos(k \cdot \vec{\alpha}_j) \sigma_x + \sin(k \cdot \vec{\alpha}_j) \sigma_y \right) + M \sigma_z$$

Eq. 2-17

The two bands may only touch at the $\vec{k}_+$ and $\vec{k}_-$ points in BZ1, as in graphene (which is equivalent to this model for $t_2 = M = 0$), where $\vec{k}_+ \cdot \vec{\beta}_j = \frac{2\pi}{3}$ and $\vec{k}_- \cdot \vec{\alpha}_j = (0, -\frac{2\pi}{3}, \frac{2\pi}{3})$ (for the three $\vec{\alpha}_j$ hops). These are the only topologically significant points in BZ1, as the rest of the dispersion.

Figure 2-6: The physical configuration and final phase diagram of Haldane’s model. (a) The graphitic lattice used has 0 net flux due to its periodic distribution in the hexagonal Wigner-Seitz unit cell. (b) Phase diagram of the 2D system, in which topologically trivial phases occur for $\frac{M}{t_2} > |\frac{3\sqrt{3} \sin \phi}{3}|$ or $\frac{M}{t_2} < |\frac{3\sqrt{3} \sin \phi}{3}|$, and AQHE occurs for the central regions labeled $n = \pm 1$. Adapted from [22].
resembles a trivially gapped insulator. A Taylor expansion near the \( \vec{K} \) points results in a Hamiltonian of

\[
H(\vec{K} + \delta \vec{k}) = \epsilon I + \hbar v (\delta k_x \sigma_x - \delta k_y \sigma_y) + \vec{m}_x v^2 \sigma_z
\]

where \( \epsilon = 3t_c \cos \phi \) can largely be ignored, \( v = \frac{1}{2} t_c a / \hbar \), and \( \vec{m}_x v^2 = M \mp 3\sqrt{3} t_c \sin \phi \). \( H(\vec{K} + \delta \vec{k}) \) is a Dirac Hamiltonian with linear \( \delta \vec{k} \) dependence, similar to graphene, as shown by the simplified equation \( E_z = \pm \sqrt{d_z^2 + d_y^2 + d_x^2} \), where \( d_z(\vec{K}) = d_y(\vec{K}) = 0 \). However, the \( d_z \) term can open gaps at \( \vec{K} \), and the signs of these gaps (a bit of a misnomer given that either sign generates separation between bands) will be shown to determine the topology of the system. These signs are then completely determined by the \( \vec{m}_x v^2 \) term.

Among the ways to determine topology and the quantized Hall conductivity are to utilize the Kubo formula and integrate the Berry curvature over the whole BZ1 as Thouless did for the QHE, to use the Streda formula\(^{53}\) for quantized Hall conductivity as Haldane did, to integrate the solid angle generated by the Pauli coefficient unit vector \( \hat{d} = \frac{\hat{d}}{|\hat{d}|} \) as shown by Berry\(^{54}\)

\[ C_n = \frac{1}{2\pi} \int_{\text{BZ1}} d^3k \left( \hat{\partial}_x \hat{d} \times \hat{\partial}_y \hat{d} \right) \cdot \hat{d} \]

or to monitor whether a single well defined Bloch wave function may or may not be defined over the whole of BZ1 in order to perceive topological distinction. The latter method, which can indicate the presence of singularities in \( \hat{A} \) and thus topological distinction and non-zero Hall conductivity\(^{46}\), is useful here. Because singularities do not manifest physically, they cannot be observed in the energy dispersion of a system. Rather, they are only directly calculable through complex, non-Hermitian operations such as the Berry connection, or indirectly via integrated, similarly non-physical parameters over regions of Hilbert space.
In order to show the singularities of the Haldane system, it is sufficient to show that any viable wave function for the system has 0 amplitude at either of the $\vec{K}_s$ points, or that it must travel through 0 between them. The Schrodinger equation may be solved for the conduction and valence bands as

$$
\psi'_s = \begin{pmatrix} d_x - id_y \\ \bar{d} - d_z \end{pmatrix}, \quad \psi''_s = \begin{pmatrix} d_x + |\bar{d}| \\ d_x + id_y \end{pmatrix}, \quad \psi'_v = \begin{pmatrix} d_x - id_y \\ -\bar{d} + |\bar{d}| \end{pmatrix}, \quad \text{and} \quad \psi''_v = \begin{pmatrix} d_x - |\bar{d}| \\ d_x + id_y \end{pmatrix}
$$

**Eq. 2-19**

Because $d_x(\vec{K}_s) = d_y(\vec{K}_s) = 0$, these simplify significantly, and can be represented by the coefficients $b'_v = -a''_v = |d_x(\vec{K}_s) - d_y(\vec{K}_s)|$ and $a''_v = -b'_v = d_x(\vec{K}_s) + |d_y(\vec{K}_s)|$. Taking the first wave function in the valence band, for example, the sign of $b'_v$ is negative for $d_z > 0$ and $b'_v = 0$ for $d_z < 0$. Thus, if $d_z$ switches signs between the $\vec{K}_s$ points, then $b'_v$ must be 0 at one of the $\vec{K}_s$ points. If we temporarily ignore the differential onsite energy contributions, then $d_z = 3\sqrt{3} t_2 \sin \gamma$, where $\gamma = \pm 1$ indicates the $\vec{K}_s$ point, and $d_z$ switches signs between $\vec{K}_s$ points. Thus, $b'_v$ is not definable over BZ1. This calculation can be repeated to show that $a''_v$, $b''_v$, and $a''_v$, are also not definable over BZ1. There is no single wave function that can cover the entirety of either the valence or conduction band, and thus at least one singularity emerges. This guarantees that the system is topologically non-trivial.

The more general requirement for $d_z$ to switch signs between $\vec{K}_s$ points is $|M| < 3\sqrt{3} t_2 \sin \gamma$, and the equality of these factors is what separates the AQHE phase from the trivial insulator (Fig. 2-6(b)).

Because the two wave functions of the valence band are only differentiated by a gauge shift ($\psi''_v(\vec{k}) = \psi'_v(\vec{k}) e^{in(\vec{k})}$), it is easy to calculate the Chern number with a contour integration of the Berry connection:
Here, the integration over the whole BZ1 is separated into contours surrounding the domains which are nonzero for $\psi'(\vec{k})$, $D_i$ and for $\psi''(\vec{k})$, $D_{\parallel}$. The gauge shift of the Berry connection $A_{\parallel}(\vec{k}) = A_i(\vec{k}) + \vec{V}_i \theta(\vec{k})$ was used for simplification. In this case, the first domain can be taken to be a small circle surrounding a $\vec{K}_s$ point, thus allowing an angular substitution for $\vec{k}$:

$$2C = \frac{2}{0} \quad d = (\ = 0) \quad (\ = 2 ) = 2 \quad n \quad \text{Eq. 2-21}$$

The actual values of $(\ = 0)$ and $(\ = 2 )$ depend on the specific gauge shift. In this case,

$$e^{i\theta(\vec{k})}\psi' = \frac{d_z - |\vec{d}|}{d_z - |\vec{d}|} \cdot \left( \begin{array}{c} d_z - |\vec{d}| \\ d_z - |\vec{d}| \\ |\vec{d}| \end{array} \right) = \left( \begin{array}{ccc} \frac{d_z - |\vec{d}|}{d_z - |\vec{d}|} & \frac{d_z - |\vec{d}|}{d_z - |\vec{d}|} & \frac{d_z - |\vec{d}|}{d_z - |\vec{d}|} \end{array} \right) \psi'' .$$

Near a $\vec{K}_s$ point, the Dirac Hamiltonian parameters suffice, and produce $C = \pm 1$, and thus $C = \pm 1$. In the case that $|M| < 3\sqrt{3} t_2 \sin | and $\sin > 0$, $C = 1$, in the case that $|M| > 3\sqrt{3} t_2 \sin | and $\sin < 0$, $C = 1$, and in the case that $|M| > 3\sqrt{3} t_2 \sin |$, $C = 0$, as shown in the Fig. 2-7 phase diagram.

The final theory step toward generating a topological model for 3D crystals, including $\text{Bi}_2\text{Se}_3$, is to look at the QSHE and $Z_2$ theory developed by Charles Kane, staring in 2004.\textsuperscript{19,27,55} The QSHE model for graphene is remarkably similar to Haldane’s AQHE model, wherein spin orbit coupling (SOC) in Kane’s model takes the place of the periodic flux in Haldane’s model.

$$H = t_c c^\dagger_i c_j + i_{\text{SO}} v_{\psi} c^\dagger_i s^z c_j + v_i c^\dagger_i c_j$$

In this case, the system is topologically classified by $v > 3\sqrt{3} s_0$, which generates the insulating phase, or $v < 3\sqrt{3} s_0$, which generates the QSH phase (Fig. 2-7). A full derivation is omitted here.
A significant difference of note is that because there are no interactions between the spin states, the Hamiltonian is separable by spins, and we get a decoupled pair of Hamiltonians with opposite topological indices:

\[
H = H_t + H_l = \begin{pmatrix} H'' & 0 \\ 0 & H'' \end{pmatrix}
\]

Eq. 2-23

Here, the creation and annihilation vectors have a 4-component basis for both spins and sublattices,

\[
\begin{pmatrix} a & b & a & b \end{pmatrix},
\]

and the \( H'' \) is a Haldane-like model for a positive contribution from SOC.

The sign of the SOC contribution only changes the sign of the \( d_z \) term, which has the effect of negating the Chern number of the system. Because these Hamiltonians do not interact, the Kane-Mele model for SOC in a graphitic geometry contains two spin-polarized copies of the AQHE system. Each of these contribute equal charge conductivity in opposite directions at the edges.

**Figure 2-7:** Band structures and phase diagram of the Kane-Mele model. This model takes into account differential on site energies, NN and NNN hopping with SOC, and a Rashba term. Along the axes \( \lambda / \lambda_{SO} \) and \( \kappa / \lambda_{SO} \), which are ratios of on site and Rashba energies to the SOC energy and along which axes only one is considered, the system is in a QSH state for \( | \lambda / \lambda_{SO} | < 3\sqrt{3} \) and \( | \kappa / \lambda_{SO} | < 2\sqrt{3} \), but is a trivial insulator otherwise (inset). (a) and (b) show the band structure in the bulk, along with surface dispersions for the QSH phase and insulator phase. The surface bands are chiral, leading to spin current in the QSH phase, but the insulator phase does not conduct because the surface bands are filled. Adapted from [19].
but because they are spin polarized, electrons with spin up moving in the opposite direction as electrons with spin down generates a net spin current. This is what gives the QSHE its name.

There are two complications to this model that are worthy of consideration, and are integral to the understanding of topological insulators as a general class. The first is an additional term in the complete Kane-Mele model that involves interactions between spin states. This is the Rashba term,

$$H_\text{R} = i\lambda_\text{R} \sum c_i^\dagger (\vec{\sigma} \times \hat{d}_i^\dagger) c_j,$$

and its addition requires a model that goes beyond chiral Haldane pair model. The second complication is that for time reversal invariant (TRI) systems, each band has a partner with the opposite Chern number (as can be seen from $\tilde{\Omega}(-\vec{k}) = -\tilde{\Omega}(\vec{k})$), resulting in a net Chern number of zero. Thus, while it is possible to use a similar method to previous one to calculate wave function singularities and individual Chern numbers for the general system, this is a useful situation in which to introduce the more general $Z_2$ theory, which may characterize the system as a whole. Further, the $Z_2$ theory is generalizable to 3D systems with both TRI and inversion symmetry, and is thus extremely useful to Bi$_2$Se$_3$, which is a system that contains those symmetries.

2.3: $Z_2$ Topological Insulators

The $Z_2$ theory of topological materials is more general than the Chern classification of bulk topologies, although the price for this additional usefulness is paid in abstractness of the theory. However, while the Chern classification is rendered moot by TRI, the $Z_2$ theory simplifies (somewhat). This is largely due to the emergence of Kramers degeneracy. This concept refers to spin $\frac{1}{2}$ systems, in which TRI guarantees the existence of an isoenergetic state at $\pm \hat{p}$. In Bloch systems, this refers to momenta $\pm \vec{k}$, and thus Bloch functions are related by $|u'_{\vec{k}}\rangle = e^{i\vec{k} \cdot \vec{a}} |u_{\vec{k}}\rangle$.
and \( |u^I_k \rangle = e^{-i \theta} |u_k' \rangle \), where \( k \) indicates the Kramers degenerate band pair, and \( I \) or \( II \) indicates the band within the pair (Fig. 2-8).\(^{57}\) In general, these states are separated only by the direction of momenta and a gauge shift, but they belong to different bands of a pair. There are various points in Bloch systems at which \( H(\vec{k}) = H(-\vec{k}) \) \( e.g. \) \( \vec{k} = \vec{\Gamma} = \frac{1}{2} (n_1 \vec{b}_1 + n_2 \vec{b}_2 + n_3 \vec{b}_3) \) for any reciprocal lattice vectors \( \vec{b}_i \) in 3D, where \( n \in \{0,1\} \). Thus, the system is required to be doubly degenerate at these momenta.

The band doubling effect of TRI is what nullifies the net Chern number. Instead, the \( Z_2 \) theory looks at individual bands of each pair, and uses TRI implications in order to assess topology. The main tool in this assessment is the interaction matrix between the span of eigenstates and the time-reversed eigenstates,

\[
w_{mn}(k) = \langle u_m( -k) | u_n(k) \rangle \quad \text{Eq. 2-24}\]

Here, there is only overlap between the Kramers band pairs. Thus, \( w_{mn}(k) \) is a matrix composed of

---

**Figure 2-8:** Schematic of a one dimensional, TRI band structure showing Kramers degeneracy at and between \( \pm \vec{\Gamma} \). At the \( \pm \vec{\Gamma} \) points, labeled \( \pm \), the states are doubly degenerate in momentum and energy because the momenta are equal. Between the points, a state is only guaranteed to exist at \( \pm \vec{k} \). In general, these states belong to different bands. Adapted from \(^{57}\).
\[
\begin{pmatrix}
0 & e^{i \phi} \\
e^{-i \phi} & 0
\end{pmatrix}
\] blocks along the diagonal, and \( w_{mn}(k) \) becomes antisymmetric at the Kramers momenta. For antisymmetric matrices \( M \), the Pfaffian \( \text{Pf}(M)^2 = \det(M) \) is an apt descriptor of the matrix. For this specific TRI overlap matrix, \( \text{Pf}(w_{mn}(k)) = \pm e^{i \phi} \), where the sign of the product is positive for even numbers of band pairs, and ambiguous for odd numbers of band pairs. For one 2 \( \times 2 \) matrix, the Pfaffian is simply \( \text{Pf}(w(k)) = \pm e^{i \phi} \), and \( |\text{Pf}(w(k))| = 1 \).

The Kane-Mele QSH model was initially topologically characterized with a slightly different overlap matrix, \( P(\tilde{k}) = \text{Pf} \left[ \langle u_i(\tilde{k}) | \Theta | u_j(\tilde{k}) \rangle \right] \). The heart of this topological argument is what while in general, time reversed pairs are only differentiated by a momentum and gauge shift, there exist some pairs of points in BZ1 where \( P(\pm \tilde{k}^*) = 0 \), and that these cannot reconnect with one another because their opposite positions ensure that they would have to do so at a point where, much like in the case of the prior overlap matrix, \( |P(\tilde{k})| = 1 \). Thus, an individual pair is unable to reconnect, and these \( P(\pm \tilde{k}^*) = 0 \) are permanent singularities in the Bloch space, and serve to topologically characterize the space. These singularities will later be shown to be connected to the Berry parameters previously discussed. Because the Pfaffian is a product of wave functions, the Pfaffian zero is sensitive to a singularity in any band.

TRI ensures that a search over half of BZ1 is sufficient to find the Pfaffian zeroes. Around each, the phase of \( P(\tilde{k}) \) propagates in a given direction, giving it each a certain vorticity (Fig. 2-9(a,d)). It is possible for multiple pairs of zeros \( P(\pm \tilde{k}_{12}^*) = 0 \) to exist. In this case, singularities of either pair that have opposite vorticity may rejoin and annihilate each other. In that case, the topology of the system is trivial. If the system has an odd number of singularity pairs, however, it
is inherently topological regardless of some potential annihilations. The calculation of the number of Pfaffian singularities takes the form of a contour integration

$$I = \frac{1}{2\pi i} \oint_{\Gamma} d\theta \cdot \vec{V}_i \log \left[ P(\vec{k}) + i\theta \right]$$

Eq. 2-25

This tracks the total phase change along the contour, and results in an integer. In the Kane-Mele model, this encloses a singularity at ‘M’ point of the graphitic BZ1 (Fig. 2-9 (a,b)). The QSH phase contains singularities, while the insulating phase does not. The $Z_2$ index is $I \mod 2$. Because $I$ is the number of singularities, the $Z_2$ index can also be written $= \frac{1}{2} (C - \bar{C}) \mod 2$,\textsuperscript{58} where opposite Chern numbers for opposite spins in the TRI system result in a nonzero $Z_2$ index.

**Figure 2-9:** Contour integrations around half the BZ1 in the Kane-Mele model in order to count the zeros of $P(\vec{k})$. This process evinces the topological order of the system. (a) The familiar QSH system in graphene, which contains singularities in either half of BZ1. The vorticity of $+\vec{k}^*$ generates a topological index for the system. (b) The same system as (a) for equal onsite energies, where the Pfaffian zero is smeared into a line and the system is topologically undefined. (c) A plot of $P(\vec{k})$ along the $\alpha_2$ vector (d) for insulating (dashed) and QSH phases (solid). (d) A more generalized unit cell with TRI points labeled , at which $|P(\vec{k})| = 1$ act as a barrier to topological shifts for individual pairs of singularities. Adapted from [19].
Another important tool is the time reversal polarization parameter, $P$. While the Chern number is related to the charge polarization, which is a sum of partial polarizations $P^p = P^I + P^\mu = \sum_{\gamma} \frac{1}{2\pi} \int d\vec{k} \cdot \vec{A}^\gamma(\vec{k})$ (for one dimension) over bands $\gamma \{I, II\}$, the $Z_2$ index is related to

$$P = P^I - P^\mu$$

Eq. 2-26

Because of the gauge dependence of the Berry connection discussed previously, and the gauge transformations generated by TRI, a partial polarization may be expressed as

$$P^I = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\vec{k} \cdot [\vec{A}^I(\vec{k}) + \vec{A}^\mu(\vec{k})] - \frac{1}{2\pi} \sum_{\alpha} (\chi_{\gamma,\alpha} - \chi_{\epsilon,\alpha})$$. The gauge portion of the partial polarization can be rewritten with the matrix $w_{\gamma\alpha}(k)$ tool as $P^I = \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} d\vec{k} \cdot [\vec{A}^I(\vec{k}) + \vec{A}^\mu(\vec{k})] - i \log \left( \frac{\text{Pf}[w(\pi)]}{\text{Pf}[w(0)]} \right) \right)$, and the time reversal polarization is $P_s = \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} d\vec{k} \cdot \vec{A}(\vec{k}) - i \log \left( \frac{\text{Pf}[w(\pi)]}{\text{Pf}[w(0)]} \right) \right)$. This can be recompiled through properties of $w_{\gamma\alpha}(k)$, $A(k) + A(-k) = i \nabla_k \chi_{\gamma,\alpha} - i \nabla_k \chi_{\epsilon,\alpha} = \text{Tr}(w' \nabla_k w) = \text{Tr}(\tilde{\nabla}_k \ln(w)) = \tilde{\nabla}_k \ln(\det(w))$, into

$$P_s = \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} d\vec{k} \cdot \tilde{\nabla}_k \log(\det(w(\vec{k}))) - 2i \log \left( \frac{\text{Pf}[w(\pi)]}{\text{Pf}[w(0)]} \right) \right)$$. This easily integrates into

$$P = \frac{1}{2i} \left[ \log \left( \frac{\det[w(\vec{k})]}{\det[w(0)]} \right) - 2 \log \left( \frac{\text{Pf}[w(\pi)]}{\text{Pf}[w(0)]} \right) \right]$$, which can be rewritten to show

$$\left( 1 \right)^P = \sqrt[\text{Pf}[w(0)]]{\sqrt[ \text{Pf}[w(0)] ]{ \frac{\det[w(\vec{k})]}{\det[w(0)]} } \frac{\text{Pf}[w(\pi)]}{\text{Pf}[w(0)]} }$$

Eq. 2-27

Thus, the only tool necessary to show topologically significant time-reversal polarization is the antisymmetric overlap matrix $w_{\gamma\alpha}(k)$. For two dimensions, this takes the form

$$\left( 1 \right)^p = \prod_{i=1}^4 \sqrt{ \frac{\det[w(\pi)]}{\text{Pf}[w(\pi)]} } = \prod_{i=1}^4 \sqrt{ \frac{\det[w(\vec{k})]}{\text{Pf}[w(\vec{k})]} }$$

Eq. 2-28

where $\{i, \mu\}$ is the $Z_2$ index, which is 0 for trivial insulators and 1 for topological insulators.
There are only a few extra considerations for 3D and inversion symmetric systems, and the latter simplifies calculations considerably. The calculation for the \( Z_2 \) index naturally evolves into

\[
(1)^s = \prod_{i}^{8} \delta_{i}. \]

Here, \( \delta_{i} \) is the ‘strong’ \( Z_2 \) index. There are three other ‘weak’ \( Z_2 \) indices, \( \delta_1, \delta_2, \delta_3 \), for each dimensional axis. These are calculated by taking the product of four \( \delta_1 \), at

\[
\delta_i = \frac{1}{2} (n_{x,b} + n_{y,b} + n_{z,b}) \quad \text{for one} \quad n=1 \quad \text{and the other two are} \quad 0 \quad \text{or} \quad 1 \quad (\text{Fig. 2-10}).
\]

The topological systems in which \( \delta_0 = 0 \) but at least one weak index equals one are referred to as ‘weak’ TIs, and can be interpreted as stacked 2D QSH states, which each have \( \delta = 1 \) but sum to \( \mod 2 = 0 \). These are not robust in the face of weak disorder, and thus weak TIs are undesirable for applications. This stems from the fact that weak TI Fermi surfaces enclose more than one TRI momentum, thus allowing singularities to reconnect between them. The strong TI is the only system that requires Dirac cones at each surface. Weak TIs have surface bands that reach the boundary of the BZ1, and include sides with no surface bands at all.\(^{27}\)

**Figure 2-10:** \( Z_2 \) index calculations, performed by multiplying the various \( \delta_i \) at TRI momenta. The first ‘strong’ index is the product of all \( \delta_i \), and the latter three are the ‘weak’ indices, calculated by the product of \( \delta_i \) at the (100) face, the (010) face, and the (001) face, respectively. The bottom row shows the (001) surface projections, which clarify the calculation of the third weak index, and show possible Fermi surface intersections for each distinct topology. Adapted from [\(^{27}\)].
Inversion symmetry simplifies $Z_2$ calculations. The inversion operator is defined as $P\tilde{r}, s_\tau = -\tilde{r}, s_\tau$ for Bloch Hamiltonian functions. The spin conservation distinguishes this from the time reversal operator. Again, an important tool for these simplified calculations is the overlap matrix

$$u_{m\tilde{k}} = \langle u_{m\tilde{k}} | P\Theta | u_{n\tilde{k}} \rangle$$

**Eq. 2-29**

Because the $Z_2$ indices only depend on the product of time reversal polarizations at $i$, the previous overlap matrix can be expressed as $w_{m\tilde{i}} = \langle u_{m\tilde{i}} | P(P| a_{\tilde{i}} \rangle = a_{m}(i) a_{m}(i)$, where it should be noted that the wave vectors are complete Bloch functions, and not their periodic components $| u_{n\tilde{k}} \rangle$. These are still inversion eigenstates, with eigenvalues $a(i) = \pm 1$ (as is required for $P^2 = 1$).

Then the relevant Pfaffian at the TRI momenta is $\text{Pf}[w]^2 = \det[w] = \det(\begin{array}{c} 2N \end{array})$, where $2N$ is the number of occupied bands. Kramers degeneracy guarantees that each parity eigenvalue for the band of a Kramers pair is equal which generates a eigenvalue product of 1. The Pfaffian, however, may be rewritten $\text{Pf}[w] = \text{Pf} \begin{array}{c} M \end{array}$, where $M$ is the number of Kramers degenerate band pairs.

In order to address $\text{Pf}[w]$, it is useful to note that because the Berry curvature is odd under inversion and even under TRI, it must be zero, and thus there must be a gauge in which the Berry connection $A(\tilde{k}) = 0$. Again, utilizing some properties of the overlap matrix produces $A(\tilde{k}) = -i\tilde{\nabla}_k \log[\text{Pf}[\nu(\tilde{k})]]$, and so the Pfaffian must be constant, i.e. $\text{Pf}[\nu(\tilde{k})] = 1$. Thus, the time reversal polarization for at each TRI point for inversion symmetric systems can easily be calculated as
and the strong topological index for 3D, TRI, inversion symmetric systems is

\[ i = \frac{\sqrt{\det(w(i))}}{\text{Pf}[w(i)]]} = \prod_{n=1}^{M} 2^m(n) \]

Eq. 2-30

and the strong topological index for 3D, TRI, inversion symmetric systems is

\[ (1)^n = \prod_{i=1}^{8} \prod_{m=1}^{M} 2^m(n) \]

Eq. 2-31

Interestingly, all that is necessary to calculate the topology of such a system is knowledge of the parity eigenvalue of occupied bands. This simplification is an essential facilitator for computations of topology in real systems, including Bi$_2$Se$_3$.

2.4: Bulk-Boundary Correspondence, Surface States

While it is valuable to understand the physics that define topological insulators in their bulks, the potential application value of these materials lies in the states that emerge at their surfaces. This is where topological transitions take place, where exotic states emerge (barring Weyl semimetal physics$^{59-63}$ or other systems that contain topological nodes or lines in the bulk$^{64,65}$), and where the promising conductivity properties of topological insulators reside. The only way to transition between systems containing differing topological variants is to close the band gap (as can be seen in topological transitions within the Haldane model), and likewise, the only way to transition between two media with different topologies is to close the band gap at the interface. This generates linear dispersions in 2D (Fig. 2-1(c)), and Dirac cones in 3D (Fig. 2-1(d)), as will now be shown.

Solutions involving a topological transition at an interface resemble the solution of the Dirac equation generated by Jackiw and Rebbi$^{66}$ for a system involving mass inversion at an interface. The Dirac equation is a relativistic quantum mechanical equation containing linear momentum.
terms. In solid systems, parabolic bands depend on quadratic momentum, and thus the Dirac equation applied to topological solids can be written

$$H = v\hat{p} \cdot \alpha + (m\hat{v}^2 - B\hat{p}^2)\beta$$

Eq. 2-32

where $\alpha$ and $\beta$ are Pauli (or Dirac for 3+ dimensions) matrices and $B$ has $\frac{\hbar}{e}$ units. This is already similar to the Hamiltonian of the Haldane model for $\alpha = (\sigma_x, \sigma_y, \sigma_z)$, $v\hat{p}_z = t\sum \cos(k \cdot \alpha_i)$, and other similar momentum terms. This system has a band gap of $m\hat{v}^2$, but without the quadratic momentum term, the energy spectrum of this system is symmetric under $m \rightarrow m$, and topological information is lost. The quadratic term thus behaves as a vacuum “reference”, giving topological significance to the trivial case, and to the non-trivial case by distinction. With the quadratic term, topological information is encoded in the sign of $mB$, i.e., $mB > 0$ is a topologically non-trivial system, and $mB < 0$ is a topologically trivial system (like vacuum), as shown in Fig. 2-11.

In one dimension, the Dirac-like system can be written as $h(x) = v\hat{p}_x + (m\hat{v}^2 - B\hat{p}^2)\hat{z}$. The only viable edge solutions to this Hamiltonian have zero energy. For the Dirilecht condition

**Figure 2-11**: Topological spin orientations in momentum space, as derived from the modified Dirac equation. (a) For $mB < 0$, spin is the same for $p = 0$ and $p = \pm \infty$, and tracking the spin orientation along $p_x$, for example, shows a rotation toward $\hat{x}$ and then back to $\hat{z}$ (twice). (b) Whereas for $mB > 0$, spin makes a complete rotation from corner to corner. The topological distinction is seen in the relative spin orientations at $p = 0$ and $p = \pm \infty$. Adapted from [67].
(0) = ( ) = 0, topological solutions come in the form \[ s_{\mu B}(x) = \frac{C}{\sqrt{2}} \begin{pmatrix} \text{sgn}(B) \\ i \end{pmatrix} \left( e^{ix} - e^{-ix} \right), \] where \( \xi_0 = \frac{\nu}{2|B|\hbar} \left( 1 + \sqrt{1 - 4mB} \right) \), and \( C \) is a normalization constant. This solution extends readily to higher dimensional systems with four-component wave functions:

\[
\Psi_1 = \frac{C}{\sqrt{2}} \begin{pmatrix} \text{sgn}(B) \\ 0 \\ i \end{pmatrix} \left( e^{ix} - e^{-ix} \right) e^{iy \hbar} \\
\Psi_2 = \frac{C}{\sqrt{2}} \begin{pmatrix} 0 \\ \text{sgn}(B) \\ i \\ 0 \end{pmatrix} \left( e^{ix} - e^{-ix} \right) e^{iy \hbar}
\]

Eq. 2-33

In the case of 2D at the \( x \)-boundary, \( \xi_0 = \frac{\nu}{2|B|\hbar} \left( 1 + \sqrt{1 - 4mB + 4B^2 \rho_\perp^2 / \nu^2} \right) \). The Hamiltonian for this system is \( N = \text{mod} 2 \), and the energy dispersion is \( E_{p_x} = \pm v p_y \text{sgn}(B) \) where the sign indicates both the spin and velocity direction. This energy dispersion shows helical conduction. Dirac cones at 3D system boundaries have the same solution, but with a gauge shift can be written as

\[
\Psi_+ = C \begin{pmatrix} \cos \frac{\nu}{2} \text{sgn}(B) \\ -i \sin \frac{\nu}{2} \text{sgn}(B) \\ \sin \frac{\nu}{2} \\ i \cos \frac{\nu}{2} \end{pmatrix} \left( e^{ix} - e^{-ix} \right) e^{iy \hbar} \\
\Psi_- = C \begin{pmatrix} \sin \frac{\nu}{2} \text{sgn}(B) \\ i \cos \frac{\nu}{2} \text{sgn}(B) \\ -\cos \frac{\nu}{2} \\ i \sin \frac{\nu}{2} \end{pmatrix} \left( e^{ix} - e^{-ix} \right) e^{iy \hbar}
\]

Eq. 2-33

where \( \tan = \frac{\rho_\perp}{\rho_\parallel} \), and conduction remains spin-momentum locked. Predictably, the energy dispersion here is \( E_{p_z} = \pm v p_y = \pm v \sqrt{\rho_\parallel^2 + \rho_\perp^2} \) in the case of a \( z \)-boundary. These edge solutions are all products of mass inversion, which can be seen in all of the topological models, which have at least one negative mass singularity. As previously discussed, these inverted mass singularities are topologically protected, thus resulting in topologically protected (and TRI protected in the \( Z_2 \) case) surface states.

All of these topological systems display linear surface states (Fig 2-1(b-d)). Further, the topology of these systems can be seen directly from the surface bands, which cross the Fermi
energy between Kramers degenerate momenta at an odd number of points in the bulk band gap, whereas topologically trivial states contain even numbers of $E_F$ band crossings at the surface (Fig. 2-12). Topological transformations can be directly assessed from these surface band crossings as $N = \mod 2$, where $N$ is the number of band crossings between Kramers momenta and is the difference between $Z_2$ invariants at the interface. In the case of the QH state, there is only a single linear surface band. In the case of the QSH state, the bands are required to cross at a Kramers momentum point, thus only allowing one Fermi level intersection in the gap. This applies to 3D $Z_2$ TI materials, such as Bi$_2$Se$_3$, as well.
Figure 2-12: Surface conduction and band schematics of topological materials. (a,b) Generic band structures of topologically trivial and non-trivial TRI materials, showing surface bands extending between Kramers degenerate momenta. (a) Topologically trivial systems host pairs of surface bands between the $\Gamma$ and $\Gamma_b$ points, and a mid-gap $E_F$ intersects leads to an even number of surface states. This pairwise connection can be eliminated by pushing the bound states out of the gap. (b) Topological materials host single surface bands between $\Gamma$ and $\Gamma_b$, resulting in odd numbered surface states at $E_F$, which are topologically protected. (c) “ Skipping” conduction states at the QH state interface, and (d) its linear, spin degenerate, unidirectional surface band. (e) Charge neutral spin conduction in the QSH state stemming from (f) its linear, spin-oriented surface bands. Adapted from [34].
Chapter 3: Bismuth Selenide, its Topological Properties, and its Defects

Bismuth selenide has been described as the “ideal topological insulator” due to its large band gap (0.3 eV), the central location of its single surface Dirac cone, and its stoichiometric chemistry.\(^{18,30}\) The large band gap diminishes the effects that doping levels have on the generation of metallic states in the surface or on the bulk. For normal doping conditions, this band gap allows Bi\(_2\)Se\(_3\) to display its unique topological surface state behaviors at room temperature, which is not the case for all other TI materials. The Dirac cone at the center of the momentum dispersion allows TSS conduction in any direction on the surface. Other TI materials have Dirac cones on the sides of their BZ1, such as Bi\(_{1-x}\)Sb\(_x\) in which the linear surface states are projected from the bulk L point.\(^{29}\) In such systems, isotropic behaviors are not expected, and attempts to utilize the surface states must account for the orientation of the lattice. The Bi\(_2\)Se\(_3\) Dirac cone shows minimal deviation from a circular shape (at any given mid-gap energy), allowing precisely isotropic surface conduction, or equal conductivity at arbitrary in-plane surface directions. In general, stoichiometric compounds are easier to fabricate than alloys, which require significant control and tuning efforts to achieve the desired composition. These factors make Bi\(_2\)Se\(_3\) a promising TI candidate for applications as they lead to minimized metallic conductivity and facilitate fabrication.

Its uses extend beyond typical implementations of conductive layers in transistor or memory devices. Its topology and helical surface states create some highly unusual phenomena that can have potential uses in spintronics or optical phenomena. Most of these make use of the Berry phase or surface helicity. They are discussed in detail in Ch. 3.3.

Bi\(_2\)Se\(_3\) is still at the nascent stage of its development toward practical purpose, however. There are a variety of issues that plague the crystal, and inhibit measurements of the TSSs. The main
barrier toward implementation of this material is the metallic states that tend to emerge on the surface and in the bulk of the crystal. There are many examples in the literature that show degenerate $n$-doping.$^{28,30,38,68,69}$ When $E_F$ rises to the conduction band in the bulk, spin-unpolarized carriers dominate transport measurements due to the higher density of states of bulk bands, and the TSSs are unobservable. At the surface, band bending causes 2D electron gas formation. This has a similar effect to degenerate $n$-doping, as it creates conductive metallic states at the surface with higher DOS levels than the TSSs. While TSSs still exist, their behaviors are not detectable above the higher densities of other filled states. Therefore, none of the phenomena discussed in Ch. 3.3 are observable under these conditions. Surface oxidation is another distinct defect of interest. This has been shown to correlate with 2DEG formation$^{40}$ as well as decreased TSS signals in Shubnikov-de Haas oscillations.$^{70}$ Further, the oxide layer of Bi$_2$Se$_3$ creates an insulating barrier to electrical contacts, as will be discussed in Ch. 5. Before any progress can be made toward producing devices that contain Bi$_2$Se$_3$ for the purpose of using its topological properties, these defects need to be addressed in further detail than is currently available in the literature.

3.1: The Physical and Electronic Structure of Bi$_2$Se$_3$

Bi$_2$Se$_3$ is a three-dimensional crystal with a hexagonal unit cell containing 6 bismuth and 9 selenium atoms (Fig. 3-1). As such, this unit cell is three times the volume of the primitive cell, which belongs to the trigonal space group $R\bar{3}m$. Its lattice parameters are 4.14 Å in the in-plane directions and 28.64 Å perpendicular to the plane. Each unit cell is composed of three vertically (along the out-of-plane $c$-axis) stacked quintuple layers (QLs). These QLs each contain five vertically stacked atomic layers, Se$_2^A$-Bi$_8^B$-Se$_1^C$-Bi$_5^A$-Se$_2^B$, which are covalently bonded in an octahedral configuration and rhombohedrally stacked at (A, B, C atomic sites. Each QL is terminated by a
selenium atom, separated by a van der Waals (vdW) gap of 2.42 Å\(^71\), and also rhombohedrally stacked. This continuity of this rhombohedral stacking results in a screw axis-like structure of atomic layers. These vdW-gapped QLs lead to a natural cleavage plane at the (0001) hexagonal face, terminated by a Se atomic layer. The Se atoms are relatively light, and play the role of electron acceptors (as will be addressed in Ch. 2.7), while the Bi atoms are heavy, and are electron donors. The heaviness of the Bi atoms has another major effect in this system. It generates strong spin-orbit coupling, which is the genesis of the topological distinction of Bi\(_2\)Se\(_3\).

Spin-orbit coupling (SOC) is a relativistic atomic effect, in which the electron in its own rest frame experiences a dynamic electric field produced by the moving nucleus, as well as a magnetic

![Figure 3-1: Schematic of the Bi\(_2\)Se\(_3\) hexagonal unit cell. Each unit cell, which contains six Bi atoms and nine Se atoms, is three times the size of a primitive cell. Its lattice parameters are 4.14 Å in the in-plane directions and 28.64 Å perpendicular to the plane. Each unit cell is composed of three QLs](image)
field $\vec{B} = -\frac{\vec{v} \times \vec{E}}{c^2}$. Substituting momentum for velocity, using the radial dependence of a central electric field, and expressing that field as a potential results in $\vec{B} = \frac{1}{m_e c^2 r} \frac{\partial U}{\partial r}$. Using the angular momentum provides $\vec{B} = \frac{1}{m_e c^2 r} \frac{\partial U}{\partial r}$, and the first order perturbation Hamiltonian contribution

$$H_{so} = -\mu \cdot \vec{B} = \lambda \vec{L} \cdot \vec{S} = \lambda \frac{\hbar^2}{2} \vec{L} \cdot \vec{\sigma},$$

where $\vec{\sigma}$ is the Pauli matrix spin indicator in a 2-band system and is the spin order parameter, which describes the strength of the SOC. In solid systems, this has the effect of splitting some bands with opposing spins. For bands with orbital angular momentum quantum number $L > 0$ (e.g. $p$, $d$, $f$, etc.), this effect produces significant splitting, but has no effect on the $s$-like bands. The magnitude of the energy split is proportional to the atom’s atomic number to the fourth power, $Z^4$. Looking at some relevant atoms, and comparing the $Z^4$ ratios of their splitting strength shows that before finer considerations, $\gamma_u = 36$, $\gamma_s = 1,235$, and $\gamma_L = 36,619$. This large SOC is the basis for the broad presence of Bi in prospective TIs (BiSb, Bi$_2$Se$_3$, Bi$_2$Te$_3$, Bi$_2$Se$_2$Te, etc.) This interaction maintains TRI ($L \rightarrow L$ and $P \rightarrow P$), and is symmetric under inversion ($PL \rightarrow L$ and $P \rightarrow P$), which means that the $Z_2$ topological categorization with the inversion consideration holds (as was the case for the Kane-Mele QSH model). Thus, systems with TRI, inversion symmetry, and SOC can still be categorized according to their band parities.

The bands of Bi$_2$Se$_3$ can be calculated from density functional theory by initially organizing the atoms in their sites, and subsequently turning on chemical bonding (stage I), crystal-field splitting (II), and finally applying SOC (III) (Fig. 3-2). The focus of Zhang et al.$^{18}$ was to do these calculations near the point, where $E_g$ is at a minimum. This optimizes a search for SOC-induced band inversion, in which bands of opposite parity may exchange positions around $E_F$, thus resulting
in a sign switch in the parity of occupied bands and a topologically nontrivial band structure. As the net parity of core bands is zero, only the states near the Fermi surface need to be considered. Prior to chemical interactions, these are Bi $6s^2 6p^3$ and Se $4s^2 4p^4$, and the $s$-orbitals are ignored. Organizing the orbitals of a primitive cell (2 Bi and 3 Se) by parity results in two odd and one even state from each Se $p$ orbital, and one odd and one even state from each Bi $p$ orbital. Chemical bonding (I) hybridizes these states, pushing down ‘acceptor’ Se and lifting up ‘donor’ Bi orbitals. The applied crystal field (II) splits the in-plane from out-of-plane states according to Bi$_2$Se$_3$’s point group symmetries, and the $p_z$ states are left closest to $E_F$. Finally, the $H_{SO} = \lambda \mathbf{L} \cdot \mathbf{\Sigma}$ energy is turned on (III). Because the SOC mixes spin and orbital angular momenta while preserving the total angular momentum, there is a repelling effect between $|P_{1z}^+\rangle$ and $|P_{1x+iy}^+\rangle$. Opposite repulsion effects on neighboring $|P_{1z}^+\rangle$, $|P_{2z}^+\rangle$ states results in band inversion at the point, and the opposite parities of these bands results in $\mu = \sum_{n=1}^{2n}(-1)^n = 1$. However, because there is no band

\[ \mu = \sum_{n=1}^{2n}(-1)^n = 1 \]
inversion at any other Kramers momentum \( \langle i \rangle = \frac{M}{M_{m=1}^{m=n}} \), and the total product of time-reversal symmetries at is negative \( \langle 1 \rangle = \sum_{m=1}^{M} 2m \langle i \rangle = 1 \), and thus \( \langle \rangle = 1 \) and Bi\(_2\)Se\(_3\) is a strong topological insulator. The magnitude of the SOC that generates band inversion can be seen in Figure 3-2(b), where the SOC is increased from zero to its empirical magnitude as \((\text{Bi}) = x_0(\text{Bi})\) and \((\text{Se}) = x_0(\text{Se})\), where \( x_0(\text{Bi}) = 1.25 \) eV and \( x_0(\text{Se}) = 0.22 \) eV. The level crossing occurs where \( x \approx 0.58 \), and through the process the gap energy changes from \( E_g = 175 \) meV with non-inverted bands to \( E_g = 300 \) meV with inverted bands. The strong SOC, therefore, produces a mean band deviation of \( -\approx 300+175 = 237.5 \) meV, and generates the largest band gap of the 3D TI material class discovered so far.

The surface states of Bi\(_2\)Se\(_3\) can be calculated from a low-energy model, taking the Hamiltonian up to \( k^2 \). Considering a four-state system for two spins and two bands, centered near the point,

\[
H = E_0(\vec{k}) I_{4 \times 4} + \begin{pmatrix}
M(\vec{k}) & A_1 k_x & 0 & A_2 k_y \\
A_1 k_x & -M(\vec{k}) & A_2 k_y & 0 \\
0 & A_2 k_y & M(\vec{k}) & -A_1 k_x \\
A_2 k_y & 0 & -A_1 k_x & -M(\vec{k})
\end{pmatrix}
\]

where \( k_x = k_x \pm i k_y \), \( E_0(\vec{k}) = C + D_j k_j^2 + D_x k_x^2 \), and \( M(\vec{k}) = M - B_j k_j^2 - B_x k_x^2 \). All parameters except \( \vec{k} \) are fit to the appropriate energies via an ab initio calculation. This system is representable by the appropriate choice of Dirac matrices in the as-mentioned modified Dirac equation,

\[
H = E_0(\vec{k}) s_0 \otimes \sigma_0 + M(\vec{k}) s_0 \otimes \sigma_z + A_1 k_x s_z \otimes \sigma_z + A_2 (k_x s_x + k_y s_y) \otimes \sigma_z
\]

Given \( (M,B_i,B_j) > 0 \), one can see from the equation for \( M(\vec{k}) \) that the sign of the Hamiltonian of this system inverts from \( k = 0 \) to \( k = \), as expected from a TI. In this formulation, the required
time \( ( = K \hat{s}_y \hat{\sigma}_0 \), where \( K \) is the complex conjugate operator), inversion \( ( P = s_0 \hat{\sigma}_3 \), and three-fold rotation \( ( C_3 = \exp(i \hat{s}_z \hat{\sigma}_0) \) symmetries are apparent, and the Hamiltonian has uniaxial anisotropy along \( k_z \). The surface states can be calculated by keeping the quantum numbers \( k_x \) and \( k_y \), but substituting momentum for \( k_z \). The solution here is to the 1D \( ( \hat{z} \) modified Dirac equation projected to the surface boundary, with the solutions given in Sec. 2.4, and the previously given power expansion of \( k_x \) and \( k_y \) around the point. The low energy surface Hamiltonian becomes

\[
H_{\text{surf}} = \begin{pmatrix}
0 & A_2 k \\
A_2 k & 0
\end{pmatrix}
\]

where the two states are for opposite spins, and each comprised of a superposition of \( |P_1, \rangle \) and \( |P_2, \rangle \). The energy dispersion is \( E = \pm A_2 \sqrt{k_x^2 + k_y^2} \), and the states are spin-momentum locked (Fig. 3-3). The velocity at these Dirac cones is \( v_F = \frac{A_2}{\hbar k} = \gamma \frac{\hbar}{k} \approx 6.2 \times 10^3 \text{ ms}^{-1} \).

There is an apparent issue with this surface dispersion. This comes from the fermion doubling theorem, which states that having Hall conductivity in integers of the magnetic flux quantum
requires an even number of Dirac points in TRI systems.\textsuperscript{72} When a magnetic field is applied to a strong TI such as Bi\textsubscript{2}Se\textsubscript{3}, the surface Hall conductivity should be quantized in half integers, $\sigma_{xy} = (n + \frac{1}{2}) e^2 / h$. However, for experiments with physical leads on the top and bottom of a slab sample, both surface Dirac cones are inherently measured, and the conductivity ‘doubles’, and appears as an integer again. This coupling can only be surmounted with an experiment that can decouple these surface measurements. This ability to generate a single Dirac fermion without violating TRI is a unique property of the strong TI transition.

3.2: Characterizations and Applications of Bi\textsubscript{2}Se\textsubscript{3}

The topology and surface states of Bi\textsubscript{2}Se\textsubscript{3} create some very interesting physical phenomena, although they can be difficult to measure. Among these are high mobility and backscatter protection in surface transport, weak antilocalization, unusual Shubnikov-de Haas oscillations, a surface quantum Hall effect, topological magnetoelectric effects, and Majorana fermion emergence at superconductor interfaces. These lend to a variety of applications, some of which were discussed in Ch. 1. However, before these can be harnessed and manipulated, experimental confirmation of topology and its accompanying surface states is necessary.

Various research groups have measured the surface Dirac cones of Bi\textsubscript{2}Se\textsubscript{3} with angle resolved photoemission spectroscopy (ARPES) (\textbf{Fig. 3-4}). This has become the standard method of observing topological surface states (TSS) in physical samples, due to its surface sensitivity and the wealth of band structure information it provides. In this experiment, (further explained in Ch. 4) photoemitted electrons are detected from their respective energy bands at the surface as a function of momentum and energy, thus effectively displaying the surface band structure of the
In Bi$_2$Se$_3$, while the Dirac point is not necessarily located at the intrinsic $E_F$ level, it is in the band gap, and doping (intentional or not) can shift $E_F$ to the Dirac point. The more common Fermi level shift is toward the conduction band, as will be explained in Ch. 3.3. In general, the bulk bands, which contain high electron densities, emit a large number of photoelectrons, and are seen at the high- and low-energy limits of most published ARPES spectra. The surface bands, which have a significantly lower DOS (although the ratio of detected densities is a function of the surface sensitivity of the measurement), emerge from the bulk bands, connecting the bulk valence and conduction bands through the Dirac cone.

The band structure measured in ARPES lends significant credibility to the theory that predicts the rest of the emergent physical phenomena. Among the most experimentally challenging of the remaining characteristics is the lack of direct observation of the absence of backscattering. In theory, there is no backscattering in conducting Dirac cone states $|\vec{k},\uparrow\rangle$ because the state traveling in the opposite direction, but with the same spin, $|\vec{-k},\uparrow\rangle$, does not exist, and as primary scattering effects do not flip carrier spin, the transition between these states is impossible.$^{34}$ In an effort to

**Figure 3-4:** Spin-polarized Dirac cones, experimentally measured with ARPES. (a, b) Dirac cones with $t_{s\uparrow} \psi_{r\uparrow} - t_{s\downarrow} \psi_{r\downarrow}$ at 300 and 10 K, indicating the robustness of these surface states caused by SOC and band inversion up to room temperature. (c) An $n$-doped surface dispersion, where the Fermi level ($\nu_{\uparrow\downarrow} \psi_{r\uparrow\downarrow}$) is in the conduction band. (d) A cartoon schematic of the spin-polarized Dirac cone. Spin polarization was measured to 90° momentum locking ± 5°. Adapted from $^{28}$. 

---

$^{34}$ In an effort to
infer the existence of this effect, an experiment from the Cavagroup showed that the measured spatial distribution of carriers on the surface of Bi$_{0.92}$Sb$_{0.08}$ matched with the distribution predicted by the spin-dependent scattering probability.\textsuperscript{73} As the Bi$_{0.92}$Sb$_{0.08}$ system is subject to the same band inversion physics as Bi$_2$Se$_3$, this indicates that the physics that drives both of these topological phases can cause Bi$_2$Se$_3$ to exhibit backscatterless transport. This experiment investigated the periodicity of the spatial distribution of the local local density of states (LDOS) on the Bi$_{0.92}$Sb$_{0.08}$ sample surface with dI/dV scanning tunneling spectroscopy (STS) (Fig. 3-5). Fourier transformed LDOS maps were then compared to spin-dependent scattering probability predictions,

$$SSP(\bar{q}) = \int I(k)T(\bar{q},\bar{k})I(\bar{k} + \bar{q})d^2\bar{k} \quad \text{Eq. 3-4}$$

where the transfer matrix $T$ is just the overlap matrix of spin states with different momenta. This decreases as spins variance increases. When integrated over the whole BZ1, the SSP indicates which states should be filled with given probability, in a manner that’s highly sensitive to spin
polarized modes. The control for this experiment is the spin-independent joint density of states, 
\[ JDOS(\vec{q}) = \int I(\vec{k})I(\vec{k} + \vec{q}) d^2\vec{k}, \]
which is the same as SSP for a transfer matrix of unity. The FT-STS spectra match the SSP prediction and not the JDOS, allowing the control to be rejected, and providing strong evidence that backscattering in these 3D topological systems is suppressed.

Weak antilocalization (WAL) is a purely quantum effect in the diffusive transport regime in which magnetic field-dependent scattering leads to decreased conductivity (Fig. 3-6). In most topologically trivial TRI materials without significant SOC, carriers that are scattered in opposite direction along a closed contour are subject to opposite phase acquisitions over the loop. However, the overlap of these TRI contours creates constructive interference of the oppositely traveling states, and thus these scattered carriers have increased wave function amplitudes in localized areas, and conductivity is lowered. When a magnetic field is applied, this phase overlap is complicated by the TRI-breaking magnetic field phase factor, constructive interference is lowered, and the localization effect is negated. However, the presence of a Berry phase in topological materials

![Figure 3-6: Weak antilocalization in TRI materials. (a) In the diffusive transport regime, carriers may be scattered around a closed loop. (b) In the case of most materials, the overlap of the contours that propagate oppositely creates constructive interference, weak localization, and low conductivity. This effect is mitigated by an applied magnetic field. (c) For systems with a Berry phase such as the surface of Bi$_2$Se$_3$, this phase change is destructive rather than constructive, leading to weak antilocalization and high conductivity absent a magnetic field. (d) Experimental measurements of the weak antilocalization signature in the magnetoconductivity of Bi$_2$Se$_3$ as a function of applied field angle. Adapted from [67] and [74].](image)
causes an additional phase change of \( \frac{\pi}{4} \) along the scatter contour for oppositely directed carriers. Thus, these same TRI scattering contours generate *destructive* interference in the absence of a magnetic field. This is the weak antilocalization effect. In the local area of the scattering contour, the deconstructively interfering states have lowered amplitudes, and conductivity increases accordingly. The magnitude of this effect can be expressed as

\[
\Delta \sigma(B) = \alpha \frac{e^2}{h} \left[ \Psi\left(\frac{\hbar c}{2e^2B} + \frac{1}{2}\right) - \ln\left(\frac{\hbar c}{2e^2B}\right)\right],
\]

where \( \Psi \) is the digamma function and \( L \) is the phase coherence length. This effect is dependent on the angle of the applied magnetic field relative to the TI surface, as the phase change of traversing carriers is dependent on the relative angle between the contour normal and the applied field. In real systems, in which bulk is often a factor, thicker samples and higher applied fields can often result in positive magnetoconductivity.

Among the more useful measurements of carrier transport in topological insulators is high field magnetoconductivity and Shubnikov-de Haas (SdH) oscillation analysis. Incident high magnetic fields on these systems induce Landau levels around the Dirac point, and varying the field strength spreads the levels as \( E_n = \pm v_F \sqrt{2eBN} \), where the 0\(^{th}\) level is pinned at the Dirac point,\(^{75}\) barring very high field strengths that can open a band gap. When \( E_N \) crosses \( E_F \), the system’s carrier concentration changes, and its transverse (\( s_y \)) and longitudinal (\( s_x \)) conductivities oscillate, as seen in the integer QHE. The difference between the energies in this quantization scheme (and the Landau levels in the integer QHE), \( \Delta E_{\text{sho}} = \hbar \omega_c \), may give early indication of a measured system’s topology. Additionally, when \( B \) increases to push \( E_N \) through \( E_F \), SdH oscillations occur (Fig. 3-7). These can be modeled with the Lifshitz-Kosevich theory\(^{77}\) as

\[
s_x = A_y R_y R_x R_y \cos\left(2 \frac{F}{B} \left( \frac{1}{2} + \frac{1}{2} \right) \right),
\]

where \( F \) is the frequency of oscillations in \( \omega_c \), \( L \) is the phase offset \( R_y = \frac{2\pi^2 (k_B T / \hbar \omega_c)}{\sinh(2\pi^2 (k_B T / \hbar \omega_c))} \),
56

Figure 3-7: SdH oscillations and analysis. (a) Oscillations in $R_y$ (transverse resistance) show the energetic shift and angle dependence of Landau levels. (b) Plotting the LL trend against $\sqrt{B}$ shows the frequency $F$ as a linear trend, and the $x$-intercept gives the expected Berry phase. The insets show the $\omega_{\perp}$ and its Fourier transform for $F$. (c) Dingle analysis of SdH oscillation amplitude gives $T_D$, from which the surface mobility is calculated. The inset shows the temperature dependence of the oscillation amplitude and calculations for the cyclotron parameters. Adapted from [76].

$$R_y = \exp\left[-2\pi^2 (k_y T_D / \hbar \omega_c)\right], \quad \text{and} \quad R_s = \cos\left(\frac{1}{2} g m_e / m_c\right)$$

are the temperature, Dingle, and spin damping factors. $g$, $\omega_c$, and $m_c$ are the electron g-factor, cyclotron frequency, and cyclotron mass, and the Dingle temperature $T_D = \frac{\hbar}{2\pi k_B} \frac{1}{\sqrt{\nu}}$ gives the scattering energy from which the carrier lifetime can be inferred. The frequency of SdH oscillations $F = \frac{\hbar}{2\pi e} \left(\pi k_F^2\right) = \frac{\hbar}{e} n_c$ can be used to measure the surface carrier concentration and Fermi wave number, the phase offset $2 \pi n_c$ can be used to calculate the Berry phase, the temperature trend of SdH oscillation amplitude can be used to calculate $\omega_c = \frac{e B (\nu_c)}{\hbar (k_F)}$ for comparison against theoretical and ARPES band structures, and the magnetic field strength trend can be used to calculate the Dingle temperature $T_D = \frac{\hbar \omega_c / 2\pi B k_B}{\mu_e^{\text{surf}}}$ to extract the surface mobility. These experiments have been conducted by multiple groups in order to show well-behaved Dirac fermions, and to measure the insulating quality of the bulk. Because
the magnetic field may only induce states perpendicular to its direction, varying the angle between the magnetic field and the surface can decouple the contributions from 2D surface and 3D bulk carriers.

Bi$_2$Se$_3$, perhaps unsurprisingly, displays an unusual surface QHE. In relatively low magnetic fields, surface Hamiltonian $H = \psi_x (p_y, p_y, z)$ produces a half integer quantized Hall conductivity, $\nu = \frac{e^2}{\hbar} (N+\frac{1}{2})$. The half-integer portion can be understood to be a result of the Berry phase. Experimentally, it is very difficult to measure this half quantized property because each surface contributes equally to Hall measurements, which doubles the effect and results in whole integer conductivities. In a high enough magnetic field, Zeeman splitting opens a band gap and the Hamiltonian gets a massive contribution, $H = \psi_x (p_y, p_y, z) + \psi_y$. The Kubo formula can again be used to give half quantized hall conductivity at zero temperature $\nu = \frac{\text{sgn}(\theta)}{2} \frac{e^2}{\hbar}$ at the surface.

This is, of course, subject to the same experimental difficulties, as the top and bottom surfaces of a slab geometry often combine to produce integer results.

From the topological field theory, an unusual magnetoelectric term appears in the TI Lagrangian, $\nu = \frac{1}{8} \left( eE^2 - \frac{1}{4} B^2 \right) \left( 4 \frac{1}{4} \right) E \cdot B$. In ordinary materials is zero, and all of the electric and magnetic behaviors can be described by the common Maxwell equations. However for Z$_2$ TIs, the term $= \psi$, and the equations for electric polarization and magnetization become

$$\tilde{D} = \tilde{E} - 4\pi \tilde{P} - \frac{\mu_0}{\alpha} \tilde{B}$$

$$\tilde{H} = \tilde{B} - 4\pi \tilde{M} + \frac{\mu_0}{\alpha} \tilde{E}$$

and thus an electric field can induce magnetization $4\pi \tilde{M} = \alpha \tilde{E}$ and a magnetic field can induce electric polarization $4\pi \tilde{P} = \alpha \tilde{B}$ in an otherwise unaffected system. Unfortunately the ballistic
surface states tend to foil any attempts at static polarization, although using Zeeman splitting to open a gap in the surface states may surmount this issue. This magnetoelectric signature can also be measured optically, via giant Faraday or Kerr rotation. These effects cause the polarization of transmitted (Faraday) or reflected (Kerr) light to rotate from its incident polarization direction to a significant degree. Defining, for example, the Kerr rotation as \[ \theta_K = \frac{\arg \hat{E}_r - \arg \hat{E}_l}{2}, \] where + (-) indicates the right-handed (left-handed) polarized light component, and using \[ b = 4 \pi \text{Re} \ \varepsilon_{xy} \] and the reflected field expected from a thin TI film such as Bi$_2$Se$_3$: \[ \hat{E}' = \frac{\beta}{1 + \beta^2} \left[ -\beta \ 1 \right]^T, \] the Faraday rotation is predicted to become \[ \theta_K = \tan^{-1} \frac{\beta}{1 - \beta^2} \] at low frequencies and given a magnetic field-induced surface band gap \( \Delta > 0 \).\(^{81}\) This has been measured to as high as 65° (Fig. 3-8), which is larger by an order of magnitude than rotations typically found on high mobility GaAs heterostructures.\(^{83}\)

One of the strangest effects predicted to occur in $Z_2$ TI materials is the emergence of Majorana fermions at a TI/superconductor interface.\(^{35}\) Majorana fermions are their own antiparticles, and
their existence has never been confirmed, despite being predicted in 1937. They may only exist in particle-hole symmetric, spin non-degenerate systems, and the former criterion is readily produced in superconductors, but the latter has proven difficult to achieve simultaneously. The rigorous physics of these states, as well as their applications, are far beyond the scope of this work, but researchers have considered promising for quantum computation for more than a decade.\textsuperscript{84}

3.3: Defects in Bi\textsubscript{2}Se\textsubscript{3} and their Effects on Surface States

Despite all of these promising emergent phenomena the surface of Bi\textsubscript{2}Se\textsubscript{3} and the potential applications they should allow, the measurement and manipulation of topological surface states (TSSs) has proven very difficult at nearly every step, even at the lab scale. Growth and maintenance conditions of Bi\textsubscript{2}Se\textsubscript{3} need to be strictly controlled in order to minimize structural and chemical defects. Even in the most tightly controlled environments, there have been no reports of films devoid of polycrystallinity or doping, and they are all subject to oxidation and band bending, as discussed in Ch. 1.

The main methods of growing Bi\textsubscript{2}Se\textsubscript{3} are by the Bridgman method, chemical vapor transport (CVT), and molecular beam epitaxy (MBE). The former two produce large, high quality chunks, but are not applicable to wafer deposition, and are therefore only useful for research purposes. The latter growth method can be applied to wafers (given an appropriate substrate), but shows polycrystallinity with large grain sizes being on the order of 10 µm grains, and more often < 500
nm in length. The grain boundaries may act as scattering walls, accumulating bulk charge, changing the bulk gap, and imposing a limit on the phase coherence length of topological carriers, limiting the desired ballistic transport and high mobility properties of Bi$_2$Se$_3$. These grain boundaries may appear as twins such that the angle between them given by their normalized hexagonal in-plane lattice vectors $\cos^{-1}(\hat{a}_1 \times \hat{a}_2) = \pi/3$, generate ‘zipper’ boundaries for the angles in between 0 and $\pi/3$, or may cause a vertical edge dislocation and laterally truncated QLs. MBE also shows screw dislocations caused by spiral growths, and can induce sufficient strain alter the surface states, depending on the substrate and growth quality (Fig. 3-9). These strain

![Figure 3-9](image_url)

**Figure 3-9:** MBE growth of Bi$_2$Se$_3$ on SiC (0001) with an epitaxial graphene layer. QLs are intact, but grains are grown below 100 nm, have rounded edges, and exhibit clear screw dislocations. (a) A SiC miscut step causes grain structure irregularities. (b) Step edges appear to generate spiraling vertical growth. (c) These screw dislocations can occur at significant densities. (d) Domains are slightly rotated between neighboring domains of a spiral growth. Adapted from [37].
effects may raise or lower the Dirac point relative to the Fermi level, and could thus either help or hinder attempts to lower the carrier density.\textsuperscript{90}

Band bending is another major complication in attempts to measure the TSSs. The bulk bands tend to lower toward Bi\textsubscript{2}Se\textsubscript{3} surfaces and interfaces. This causes the conduction band to intersect $E_F$, generating an electron gas within 10 nm\textsuperscript{91} of the surface which is thus referred to as the two-dimensional electron gas (2DEG). In micromechanically exfoliated samples, the bands decrease over time after cleavage, and the 2DEG moves to lower energy and appears to increase in occupation density (Fig. 3-10). Rashba effects may subsequently remove the spin degeneracy of the 2DEG\textsuperscript{92}, potentially generating confusion between the spin-polarized TSSs and 2DEG.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3-10.png}
\caption{Band bending-induced 2DEG, growing over time in Bi\textsubscript{2}Se\textsubscript{3}, as shown by ARPES. (a) After cleavage of a degenerately $n$-doped sample, the Dirac point is measured at 0.3 eV below $E_F$, and the diffuse ARPSE signal above the Dirac cone is due to the filled bottom of the conduction band. (b) The sample is remeasured after three hours. During the interim, the bands have bend downward, lowering the Dirac point to 0.5 eV, and creating a lower rim of significant intensity below the conduction band. This is growth of the 2DEG near the surface, existing simultaneously with the TSSs. Adapted from \cite{91}.}
\end{figure}
states. Low energy electron diffraction shows that in these samples, there is no change in crystallography at the surface, implying that chemical changes in the bulk, defect formation, or surface impurity adhesion drives the 2DEG formation.\textsuperscript{91} Intentional band bending has been conducted in these materials by exposing the surface to carbon monoxide at a constant pressure of $8 \times 10^{-9}$ mbar, resulting in 2DEG emergence state after 20 minutes.\textsuperscript{93} A similar experiment was performed with applied water vapor, producing a 2DEG with between 18 and 54 Langmuir of exposure.\textsuperscript{40} In thin films, this can break the inversion symmetry of the bulk and destroy the TSSs.\textsuperscript{94} In 2D-sensitive conductivity measurements (usually for TSSs) such as SdH oscillations, it provides a source of confusion as the additional source of 2D surface states makes it difficult to filter between the TSS from the bulk signals.\textsuperscript{70} Because one of the especially promising uses of Bi$_2$Se$_3$ is in spintronics, wherein spin conduction of surface states is desired, the 2DEG, spin degenerate or not, poses a major threat to the spintronics performance of Bi$_2$Se$_3$. This is because even in Rashba split surface states, bands of spin up and spin down occur on either side of the gamma point (Fig. 3-11), and cannot produce spin-filtered carriers for spintronics applications. In the case

![Figure 3-11: Generation of the Rashba spin-split 2DEG, as shown by ARPES. (a) The degenerately n-doped Bi$_2$Se$_3$ spectrum is initially devoid of the 2DEG. (b) With time, the surface bands shift down and the 2DEG forms. (c) After further doping, Rashba splitting moves 2DEG with opposite spin orientations into different bands. (d) Plan view diagram of Rashba-induced spin polarization. Adapted from [92].](image)
of the spin degenerate 2DEG, any attempts to gather spin-filtered TSS carriers are overwhelmed by the higher current densities of the 2DEG. While in specific samples the 2DEG can be counteracted an electric field effect, the inconsistency of the usually incidental 2DEG emergence threatens the viability of Bi$_2$Se$_3$ usage in integrated circuits.

Somewhat connected to the issue of the 2DEG is natural bulk $n$-doping of the crystal. This is largely caused by Se vacancies ($V'_{\text{Se}}$), which act as double electron donors, and to a lesser extent by Se anti-site defects ($\text{Se}_{\text{Bi}}^-$), which act as single electron donors. Both of these formations are thermodynamically favorable, and doping is difficult to prevent unless growth is done with excess Se under appropriate conditions. Degenerate $n$-doping (or $p$-doping, which is the worst case scenario as no Dirac cone states are filled, although this is rarely an issue) in Bi$_2$Se$_3$ can move the bulk $E_F$ to move into the conduction band. When this happens, the entire material behaves metallically. Quantum Hall effects are overwhelmed by other contributions, and SdH oscillations are almost entirely generated by 3D states causing Berry phase offsets to be zero. High concentration spin-degenerate metallic states dominate TSS spin conduction, and WAL is replaced by weak localization from the bulk, etc. Once the bulk behaves as a metal, Bi$_2$Se$_3$ is rendered useless as a host to TSSs. $n$-doping is also problematic in slighter, but nondegenerate levels, as this can enhance the surface 2DEG and shift the $E_F$ upward along the Dirac cone. When the entire material’s $E_F$ is shifted up, less band-bending is required to produce the quantum well for the 2DEG states. Further, slight $n$-doping can eliminate the spin-polarized nature of Dirac cone conduction, as the top and bottom halves of the Dirac cone contain inverted helicities. The surface spin conduction is likely to drop off as $E_F$ moves away from $E_D$ proportionally to the Dirac cone spin-polarized DOS,

$$n_1 - n_1 = \frac{1}{4\pi^2} \left( \iint_{E_{\text{bulk}}} 2\pi k \, dk - \iint_{E_0 \leq k \leq E_F} 2\pi k \, dk \right) = \frac{E_0^2 - E_F^2}{4\pi^2 \hbar^2 v_F^2}$$
\[ \frac{\partial}{\partial E_F} (n_t - n_i) = - \frac{E_F}{2\pi h v_F} \]

Eq. 3-6

where \( E_o \) represents the surface area of the lower half of the cone, and \( E_F = 0 \) at the Dirac point. Because TSSs are the only helical and ballistic states, bulk conduction hinders the potential of Bi\(_2\)Se\(_3\) for spintronics use. In order to mitigate \( n \)-doping and its deleterious effects, epitaxy of Bi\(_2\)Se\(_3\) often involves significant overpressure of Se vapor in order to minimize vacancies, but bulk carrier densities are still typically measured to \( 10^{19} \) cm\(^{-3}\). Other attempts to counteract \( n \)-doping have included the introduction of Ca\(^{38,99}\) and Cd\(^{100}\) dopants to the bulk, and growing doped, non-stoichiometric alloys such as Pb-doped (Bi\(_{2-x}\)Sb\(_x\))Se\(_3\) (Fig. 3-12). These extrinsic dopants may introduce extra issues in the form of point defects that act as scattering centers, reducing the electron mobility.\(^{99}\)

Surface oxidation in Bi\(_2\)Se\(_3\) is another complicating factor. While some of the proof-of-concept TSS measurements, such as ARPES, allow \textit{in situ} cleavage and measurement to be completed in vacuum and at low temperatures, many others, such as electrical conductivity measurements that require leads to be attached after sample deposition or cleavage, require or are facilitated by measurement at room temperature and in air. It beehoves researchers to consider the effects of these ambient conditions, as the specific materials science involved is of interest to academia, and the potential consequences to device performance as a result of ambient exposure are of interest to industry. Surface oxidation has is an expected consequence of exposure to air, as it is a natural and thermodynamically favorable process. It has been called “the likely material origin of surface degradation and environmental doping.”\(^{70}\) Kong \textit{et al.} proposed that on both nanoribbons and bulk crystals, moderate doping (\( n \approx 5 \times 10^{18} \) cm\(^{-3}\)) should cause SdH oscillations to come from the bulk
Their SdH measurements resulted in oscillations that indicated in electron densities of $n_{3D}^{SdH}=4.7\times10^{18}$ cm$^{-3}$, indicating only moderate doping. However, their Hall measurements indicated bulk concentrations of $n_{3D}^{Hall}=1.7\times10^{19}$. The discrepancy was explained by the inability of low-mobility carriers to react to and participate in SdH measurements, and that these slow carriers were a result of environmental doping. Subsequent x-ray photoelectron spectroscopy (XPS) investigations into the chemical changes that occur on the surface reveal oxidation to be the primary source alterations in Be$_2$Se$_3$ surface chemistry (Fig. 3-13). The reported growth of BiO$_x$ and SeO$_x$, as seen in the Bi 4f and Se 3d XPS spectra, respectively, was observed

Figure 3-12: Some attempts at extrinsically p-doping Bi$_2$Se$_3$ in order to move $E_F$ toward $E_D$. (a) Plot of some recorded dopants and their effect on the bulk resistivity of Bi$_2$Se$_3$, measured at 4 K. The most insulating film had a Bi replacement level of 0.125% Ca. (b) Hall-measured carrier level in variously doped Bi$_2$Se$_3$ samples. The lowest carrier concentration was measured in the same Ca-doped sample. (c-e) ARPES measurements of variously Ca-doped samples, taken 20 minutes after cleavage. Higher dopants cause an n- to p- shift, and $E_F$ drops below $E_D$. (f-h) Momentum distribution curves of the ARPES measurements. (i-j) Diagram of the Ca doping effect. (l, m) Typical ARPES spectrum and momentum distribution taken 18 hours after cleavage. (n) Band-bending diagram responsible for the changes shown in (l, m). Adapted from [28].

Figure 3-13: Bi (a,c) and Se (b,d) oxides growing in air at the surface of nanoribbon (a,b) and bulk (c,d) Bi$_2$Se$_3$ samples, as measured in XPS at the Bi 4f and Se 3d energies. This report shows Bi oxide growing immediately after cleavage, although this specific process is debated by other reports. This oxidation process is suspected to be the cause of extrinsic doping in Bi$_2$Se$_3$. Adapted from [70].

contributions as $n_{3D}^{Hall}$. Their SdH measurements resulted in oscillations that indicated in electron densities of $n_{3D}^{SdH}=4.7\times10^{18}$ cm$^{-3}$, indicating only moderate doping. However, their Hall measurements indicated bulk concentrations of $n_{3D}^{Hall}=1.7\times10^{19}$. The discrepancy was explained by the inability of low-mobility carriers to react to and participate in SdH measurements, and that these slow carriers were a result of environmental doping. Subsequent x-ray photoelectron spectroscopy (XPS) investigations into the chemical changes that occur on the surface reveal oxidation to be the primary source alterations in Be$_2$Se$_3$ surface chemistry (Fig. 3-13). The reported growth of BiO$_x$ and SeO$_x$, as seen in the Bi 4f and Se 3d XPS spectra, respectively, was observed
in both bulk and nanoribbon samples. In both systems, BiO_x was shown to form immediately after cleavage, and SeO_x was shown after 2 days of exposure to air. The universal applicability of the oxidation behavior reported in reference [70] is the subject of debate. Other measurements, including those included in this thesis, taken on the Bi 5d XPS peaks, show oxidation occurring at a different time scale. These reported BiO_x peaks are more likely to be Se 3p peaks, which overlap with Bi 4f. Regardless, because oxidation has been shown to contribute to environmental doping in semiconductors such as Ge and GaN, it is expected that oxide formation may have the same effect in Bi_2Se_3. The mechanisms here may include surface charge trapping in the oxide or at the Bi_2Se_3/oxide interface, or that oxide formation causes interfacial defect states that pin the Fermi level, and induce band bending. Oxide formation may be generated by a variety of deposited oxygen-containing species. Pure O oxidation may occur, and is likely to be a p-doping process.\(^{101}\)

Following the controlled deposition of water vapor, Bi_2Se_3 may react to form H_2Se gas and a Bismuth hydroxide surface:

\[
\text{Bi}_2\text{Se}_3 + 6\text{H}_2\text{O} \rightarrow 3\text{H}_2\text{Se} \uparrow + 2\text{Bi(OH)}_3
\]

Eq. 3-7

This reaction was presented to explain Se vacancy formation, the resultant n-doping, and surface band bending.\(^{40}\) However, the same band-bending (and Rashba splitting) seem to occur in samples
that were left in residual vacuum gas at $3 \times 10^{10}$ mbar, and samples that were cleaved in air (Fig. 3-14). If oxidation is cause of Se vacancy formation and band bending, it is essential to that the oxidation process and the oxide layer be well understood. Some studies report that cleaved surfaces are highly resistive to oxide growth.\textsuperscript{36,102,103} These studies show differences in growth, cleavage area, and surface stability. The growth method in these reports was to melt a Bi$_2$Se$_3.01$ mixture in an evacuated fused quartz ampule, as to avoid a sequence of peritectic reactions in Bi-heavy phases. Subsequently, they were annealed and recrystallized with the Bridgman method, resulting in $\sim 4 \text{ cm}^3$ sized Bi$_2$Se$_3$ rods. These samples were left in an air environment over two weeks and subjected to a variety of measurements that showed the surface to be inert. XPS showed correct stoichiometry and no oxidation with XPS at the Bi 5d, Se 3d, and Bi 4f energy ranges (Fig. 3-15), AFM measured near atomic flatness (barring few QL steps), STM topography showed the expected interatomic spacing for a Se-terminated surface, and ellipsometry showed strong

\textbf{Figure 3-14:} Similar Bi$_2$Se$_3$ surface band response from exposure to different environments, as shown by ARPES. (a) When exposed to residual UHV gases, (b) when cleaved in air, or (c) when subjected to 324 Langmuir of water vapor, the surface bands seem to bend and form a Rashba-split 2DEG in similar fashions. Thus, while surface exposure to water vapor may, in part, be a source this band bending, it is likely not the only cause. Adapted from [40].
Figure 3-15: A chemically inert Bi$_2$Se$_3$ surface. (a) A cleaved surface shows expected peaks in the Bi 4f (also Se 3p) range, the Se 3d range, and nothing at the O 1s range. This spectrum is consistent for a surface that has been exposed to air for 1 month (b), as well as a surface that has been dosed with NO$_2$ and O$_2$ (c). The trend changes when the surface is etched with argon and exposed to NO$_2$ and O$_2$ (d). The resultant XPS spectra show oxidized Bi as well as a O 1s peak. Adapted from [102].

Subsequent experiments showed inertness after cleavage in the face of air exposure as well as NO$_2$ and O$_2$ dosing, and were only oxidized by a combination of Ar$^+$ etching and NO$_2$ + O$_2$ dosing. This was explained with DFT calculations that show a minimal enthalpy change by oxidation reactions that are not initiated at Se vacancies. The differences in reported oxidation behavior are surprising. One reference attributes this to insufficient mixing during growth of bulk samples. Some studies also claim that cleavage techniques can affect the surface quality, and possibly induce defects. It is clear that oxidation has not yet been sufficiently studied, that conflict remains in the literature, and that the dynamics of the oxidation process are not well understood.
Chapter 4: Characterization Method Description

In order to get a complete picture of the changes at the surface of a material (Bi$_2$Se$_3$ in this case), it is almost always necessary to use a variety of experimental techniques. A combination of optical, chemical, electronic, and topographically sensitive measurements are used in this study in order to characterize the oxidation and Se decapping process of Bi$_2$Se$_3$.

4.1: X-Ray Diffraction

X-ray diffraction (XRD) is a well-developed method of gathering information about periodic materials, such as crystals. It has been an essential tool in materials science since its discovery and development by von Laue in 1912, and is used as a primary tool in measuring the lattice parameters, point group, and chemical properties of unknown crystals, nanostructures, and composite materials. Because absorption of x-ray wavelengths is relatively low, with typical attenuation lengths of ~65 μm in Si and ~4 μm in lead, XRD is largely a bulk measurement method, and provides structural data for large volumes (typical probe size $\gg$ crystal unit cells) and density of defects. Compton scattering, an inelastic process, provides information about the attenuation length of materials, but the main use of XRD is in measuring elastically scattered rays, which exit the material at the same wavelength at which they were emitted. The periodicity of crystalline samples allows XRD analysis to provide information about the chemical composition, as the intensity of reflections is a function of the electron density and distribution in a material, and geometry of a unit cell. Likewise, deviations from expected single crystal XRD patterns may be used to measure perturbations in crystallinity. These deviations may include extra diffraction spots or unexpected spot intensities, and may be caused by polycrystallinity or atomic defects such as interstitials, vacancies, contamination, or other effects.
4.1.1: Theory of XRD

XRD is caused by coherent elastic scattering from periodic structures. The most essential concept underlying XRD is Bragg’s law. Bragg’s law describes the conditions under which parallel atomic planes can reflect x-rays while maintaining phase coherence. As the phase acquired by an x-ray traveling between planes is dependent on its angle relative to the planes, as well as the distance between the planes, this diffraction requirement also determines the angles at which x-rays at given wavelengths may be diffracted by crystals with given interplanar distances (Fig. 4-1). For adjacent planes, a ray must be able to travel some distance to the top plane for reflection, and also be able to travel that distance plus a distance of

\[ n = 2dsin \theta \]  

Eq. 4-1

to be reflected from the bottom plane and return to its original separation from the top reflection while having its phase changed by \( \theta \). This is the required condition for coherent scattering, and is the starting place for many XRD measurements.

Figure 4-1: The Bragg condition indicates at what angles x-rays may be diffracted from parallel crystal planes. \( d \) is the interplanar distance, and \( \theta \) is the angle that the incident beam makes with the planes. In this figure, in order to maintain coherent rays, the phase of the top beam must match the phase of the bottom beam upon reflection. The extra distance travelled by the bottom beam, \( 2dsin\theta \), must result in an integer number of oscillations \( n \) of wavelength \( \lambda \). Thus, diffractions occur for \( n = 2dsin \theta \).
In order to determine the complete structure, including the space group, of a crystal, it is essential to consider the three-dimensional periodicity of the sample, and not just simple surface oriented planes. This is simplified with considerations of the unit cell and its symmetries, which may be used to generate a reciprocal lattice, which is an essential map to determining sample orientations that correctly orient crystal planes for diffraction. The unit cell is described in real space by its lattice parameters, \( a_1 \), \( a_2 \), and \( a_3 \), where the sum of arbitrary integers of lattice parameter results in a symmetric translation, and leaves the system unchanged. A reciprocal lattice is a representation of crystal planes in reciprocal space, each of which may be represented by a vector (Fig. 4-2). Plane \((hkl)\) has coordinates in reciprocal space

\[
\vec{g}_{\text{rel}} = h\vec{b}_1 + k\vec{b}_2 + l\vec{b}_3,
\]

Eq. 4-2

where \( \vec{g}_{\text{rel}} \) is the reciprocal vector, \( hkl \) are the miller indices denoting that this plane has real space nodes at \( \frac{h}{a_1}, \frac{k}{a_2}, \) and \( \frac{l}{a_3} \), and \( \vec{b}_i = -\frac{\vec{a}_i \times \vec{a}_j}{\vec{a}_i \cdot (\vec{a}_j \times \vec{a}_k)} \). Each plane in the unit cell has a reciprocal lattice vector that gives diffraction direction and Bragg angle, but further considerations must be made to

**Figure 4-2:** Demonstration of the relationship between real space planes and reciprocal lattice vectors, using the (231) plane of an orthorhombic lattice. The plane is given in real space by connecting \( \frac{2}{a_1}, \frac{3}{a_2}, \) and \( \frac{1}{a_3} \). Its reciprocal vector is \( \vec{g}_{\text{rel}} = h\vec{b}_1 + k\vec{b}_2 + l\vec{b}_3 \). \( \vec{g}_{\text{rel}} \) is an essential tool to calculating sample orientations for diffraction measurements.
calculate diffraction intensities and forbidden diffraction planes. This is largely accomplished with the structure factor.

The structure factor gives the amplitude of diffraction that a crystal plane produces. It is a weighted Fourier transform of the unit cell, calculated as:

$$S_h = \sum_i f_i(\theta, \lambda) \exp(i2\pi \mathbf{h} \cdot \mathbf{\bar{u}}) \quad \text{Eq. 4-3}$$

where atom $i$ is at real space coordinates $\bar{x}_i = u_i \bar{a}_1 + v_i \bar{a}_2 + w_i \bar{a}_3$, $\mathbf{\bar{u}} = (u, v, w)$ is the normalized position of an atom in the unit cell, and $\mathbf{h} = (h, k, l)$ is the plane being considered. $f_i(\lambda, \theta)$ is the atomic scattering factor of atom $i$ that scatters light with wavelength $\lambda$ at angle $\theta$, and can be calculated as $f_i(\lambda, \theta) = \int \rho(\bar{r}) e^{2\pi i \mathbf{\bar{r}} \cdot \mathbf{\bar{Q}}} dV$, where $\rho(\bar{r})$ is the electron probability density function and $\mathbf{\bar{Q}} = \mathbf{\bar{Q}}_i - \mathbf{\bar{Q}}_0 = \frac{2\sin\theta}{\lambda} \mathbf{\hat{Q}}$ is the scattering vector, pointing along the bisection of the incident and reflected beams. The intensity of a scattered wave can be calculated from the structure factor of the plane as $I = |S_h|^2$. Therefore, the structure factor is essential in predicting and analyzing diffraction patterns. It contains all information about a crystal lattice, but because only intensity is measured, phase information is lost in conventional diffraction measurements. This causes confusion in back-calculating lattices from diffraction patterns, because while distances between atoms may be accurately measured, atomic positional offsets are lost with phase information. For crystals with known lattices, the structure factor provides essential information about expected diffraction intensities.

Because the structure factor is zero for many planes, caused by destructive interference of diffracted waves, diffraction is forbidden. This is only the case for non-primitive lattices. Primitive lattices may diffract for a plane with any Miller indices. However, for example, the structure factor predicts that the diamond cubic lattice will not produce diffractions at planes where the Miller
indices have a mix of even and odd numbers or where sum of the miller indices $h+k+l = 2(2n+1)$ for any integer $n$. In the (002) diffraction, x-rays do reflect coherently from every other plane, but the ones between also coherently reflect with a phase offset of exactly $\pi$, thus causing destructive interference and eliminating the diffraction signal (Fig. 4-3). Rules like these greatly can simplify structural measurements. Further, measurements involving forbidden diffraction planes are convenient because they don’t require any calculations for electron density-dependent intensities, and can still be highly informative about the symmetry of a crystal sample. They may obviate full Patterson analysis in tests for single crystallinity. If significant diffraction is measured where a family of forbidden planes is predicted, it is likely that point defects are not the cause. Polycrystallinity is a common cause of supposedly forbidden plane measurement, due to the potential variety of orientations of grains in the sample.
4.1.1.1: XRD in Practice

XRD data was acquired with a Bede Metrix-L diffractometer for this work. It uses a stationary copper K-α source, rotating a point detector, three goniometers controlling sample Euler angles, and has 3-axis high-resolution XRD (HRXRD) capabilities. This tool is designed to measure 300 mm wafers, but more irregular samples may be mounted to a 300 mm wafer and measured. Because the tool is effectively *ex situ*, few other sample requirements exist. The main measurements used in this work are ω-2θ and pole figure XRD scans. Omega-2theta scans are performed by changing the detector angle at twice the rate of source angle such that the reflection

![Diagram of XRD setup](image)

**Figure 4-4:** Demonstration of ω-2θ experimental configuration and reciprocal space scan direction for a Si sample with a (001) surface. (a) The detector is scanned at twice the rate of the source, $2\frac{d}{d\omega} = \frac{d}{d2\theta}$, such that the bisectrix of the reflection is maintained. If they are at the same position, as is shown, then the scan runs perpendicular to the sample surface. (b) ω-2θ reciprocal scans are representable in reciprocal space by linear vectors anchored at the origin. Therefore, ω-2θ scans of sufficient range measure plane $(hkl)$ and planes that are linearly scaled from its reciprocal vector, $\vec{g}_{\text{max}} \propto \vec{g}_{(hkl)}$. 
bisectrix is maintained (Fig. 4-4(a)). They measure a straight path in reciprocal space, and thus can gather signal from many sample planes with the same orientation (Fig. 4-4(b)).

ω-2θ scans are used to measure planes and reciprocal vectors with the same orientation. For scans of sufficient range, the interplanar distance can be calculated with the Bragg condition, and the AOI that produces diffraction peaks. However, because of forbidden planes, Bragg angle information can be incomplete, and thus extra information in the form of the sample space group or other diffraction data may be required to correctly index the measured planes and to calculate interatomic distances. However, because angular information is often accurate to within 100 arcsec, measurements of interplanar distance are as accurate as $d_{\omega\ell} = \frac{n}{2 \sin (\frac{\theta_B}{2} \pm \ell)}$, which for copper K-α measurements the Si (004) peak, (Bragg angle $\theta_B = 34.56^\circ$), results in accuracy of $d_{004}^{Si} = 1.357 \pm 0.001$ Angstroms, which gives the lattice parameter of Si to within 0.5 picometers. This accuracy makes XRD an invaluable tool for the measurement of crystal structures.

The second important scan type used in this work is the XRD pole figure. This is a natural complement to ω-2θ scans, because while the data it produces does not contain information about interplanar distances (although a proper experimental configuration where interplanar distances are taken into account is required to produce useful pole figure data), it is completely sensitive to the direction of the reciprocal space vectors producing diffraction. It can simultaneously show planes in any direction given that they have the same interplanar distance. This is an invaluable tool for scans of polycrystalline samples, in which certain grain orientations may be preferred, and each of which produce diffraction spot patterns in different directions. This scan is easily demonstrated with the Ewald sphere (Fig. 4-5).

In order to measure a diffraction spot with the reciprocal space vector $\vec{g}_{\omega\ell}$, the x-ray configuration must satisfy
Figure 4-5: The Ewald sphere and the allowed diffraction points of Si. The radius of the sphere is taken to coincide with the \( \langle 224 \rangle \) family of planes, which are plotted red. (a.) Three-dimensional representation, showing that all \( \langle 224 \rangle \) reciprocal vectors intersect the sphere. Because \( |\bar{g}_{\langle 224 \rangle}| \) is constant for the family, any XRD scan that maintains a constant angle between the source and detector while rotating the sample can, in theory, measure all \( \langle 224 \rangle \) diffraction spots. (b.) Tomograph of (a.), showing the x-ray configuration required to measure the \( 224 \) plane of Si.

\[
\bar{k}' - \bar{k} = \bar{g}_{\text{me}} \tag{Eq. 4-4}
\]

(Fig. 4-5(b)), where \( \bar{k} \) and \( \bar{k}' \) are the ingoing and outgoing x-ray momenta respectively, and \( \bar{g} \) is the reciprocal vector of the measured plane. This equation is valid for all elastic scattering events.

A pole figure is a scan at constant values of \( |\bar{k}' - \bar{k}| = |\bar{g}| \), only varying the Euler angles of the sample with respect to the fixed source and detector. In Fig. 4-5 this could be seen by plotting a vector from the origin to any point on the Ewald sphere, and rastering it both on azimuthal and elevation axes to generate a solid angle of data. It thus produces 2D maps of diffraction spots with the same magnitudes of reciprocal vector, and can measure all planes of any symmetric family given enough scanning range.

For samples with expected plane magnitudes and directions, it is possible to calculate the orientation of the sample, as well as the x-ray source and detector angles, to detect any diffraction
plane(s). The formula for interplanar distance for a completely generic (trigonal) crystal is somewhat complex:

\[
\frac{1}{d^2} = \frac{1}{a^2}(S_{11}h^2 + S_{22}k^2 + S_{33}l^2 + 2(S_{12}hk + S_{23}kl + S_{13}hl))
\]

\[
S_{11} = b^2c^2 \sin^2 \alpha \quad S_{12} = abc^2(\cos \alpha \quad \cos \beta \quad \cos \gamma)
\]

\[
S_{22} = a^2c^2 \sin^2 \beta \quad S_{23} = a^2bc(\cos \alpha \quad \cos \beta \quad \cos \gamma)
\]

\[
S_{33} = a^2b^2 \sin^2 \gamma \quad S_{13} = ab^2c(\cos \alpha \quad \cos \beta \quad \cos \gamma)
\]

Eq. 4-5

Here \(a, b, \) and \(c\) are the crystal lattice magnitudes, \(V\) is the unit cell volume \(\vec{a} \cdot (\vec{b} \times \vec{c})\), and \(\alpha, \beta, \gamma\), are the angles \(\angle(\vec{b}, \vec{c}), \angle(\vec{a}, \vec{c}), \) and \(\angle(\vec{a}, \vec{b})\). However, this is significantly simplified for crystals with more symmetry. For cubic and hexagonal crystals, it simplifies to the following:

\[
\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2}
\]

\[
\frac{1}{d^2} = \frac{4}{3}(\frac{h^2 + hk + k^2}{a^2}) + \frac{l^2}{c^2}
\]

Eq. 4-6

This interplanar distance informs the user of the x-ray angular separation, via the Bragg condition. The only other step is to predict the necessary orientation of the sample relative to the surface normal, in order to measure the desired diffraction spot. This requires knowledge of the azimuth and elevation (tilt) angles necessary to orient the diffraction planes normal to the desired x-ray reflection bisectrix (Fig. 4-6(a)) so that the Ewald sphere condition is satisfied.

In actual XRD experiments, goniometer axes are fixed, and so to align a reciprocal vector with \(\vec{k} - \vec{k} = \vec{k}\), two separate rotations are often needed (Fig. 4-6(b))). Fortunately, because pole figures are often performed with 360° range, only the tilt of the desired diffraction plane is required. Given the common surface plane direction \(s = \vec{g}_{001}\), the tilt angle required is easily calculated as

\[
\theta = \cos^{-1} \frac{\vec{g}_{001} \cdot \vec{b}}{\|\vec{g}_{001}\| \|\vec{b}\|} = \cos^{-1} \frac{\vec{b} \cdot \vec{b} \cdot \vec{b}}{\|\vec{b}\| \|\vec{g}_{001}\|} = \cos^{-1} \left( \frac{\vec{b}}{\|\vec{b}\|} \right)
\]

With the Bede Metrix-L tool, the x-ray angle given by
the Bragg condition and the tilt angle given by the Ewald sphere suffice to produce hemispherical pole figures, as will be shown later for Bi$_2$Se$_3$.

4.2: Spectroscopic Ellipsometry

Spectroscopic ellipsometry (SE) is an optical technique measures the changes in light polarization that occur upon reflection from (or transmission through) a sample surface as a function of the energy of the light. Changes in polarization that occur at interfaces are caused by interactions between the incident EM wave and the electrically polarizable sample surfaces, and therefore, ellipsometry is sensitive to the dielectric properties of the media at the interface(s). Further modifications to the incident light polarization arise in experiments that involve transmission through materials, where light absorption and phase procession need to be accounted for as the polarized light propagates through the sample. But with sufficient supplementary
information (usually coming in the form of at least a partial knowledge of the optical properties of the sample materials), these experiments can give fast (< 10 seconds), accurate (often within 1 nm), and non-destructive information about dielectric or semiconducting layers, as well as thin metallic films, including their thicknesses, roughness, electronic, and optical properties. Ellipsometry has become an invaluable tool in the semiconductor industry for measuring deposited thin films in rapid production environments.

4.2.1: Theory of Spectroscopic Ellipsometry:

In order to understand changes in polarized light that is reflected from a sample surface, it is useful to understand the electronic response of the reflecting material, and to explore the physics from which the dielectric function of the sample is derived. In optical experiments, the most consequential physical process is light absorption, and accompanying direct excitation of electrons to higher energy states. The band structure provides information about intra- and inter-band transitions. For gapped materials, the number of available direct transitions as a function of transition energy can be expressed via the joint density of states (JDOS)

\[ D_j(E_{cv}) = \frac{1}{8} \sqrt{\frac{dS_k}{V_k(E_{cv})}} \]

where \( E_{cv} = E_c - E_v \) is the energy difference between the valence and conduction bands. Fermi’s Golden rule for a continuous state distribution (as in band theory),

\[
R_{FGR} = \frac{2\pi}{\hbar} \left( \frac{e}{m_0} \right)^2 \left| \frac{E(\omega)}{2} \right|^2 \sum_k |P_{cv}|^2 \delta(E_{cv}(k) - \hbar\omega)
\]

\[
R_{FGR} = \frac{2\pi}{\hbar} \left( \frac{e}{m_0} \right)^2 \left| \frac{E(\omega)}{2} \right|^2 |P_{cv}|^2 D_j(E_{cv}) \]

where \( E(\ ) \) is the incident electric field, \( P_{cv} \) is a quantum mechanical transfer matrix constant,
and \( E_{\omega} = \hbar \omega \), uses the band structure to give the transition rate at a given light energy. In the first formulation, the Kroneker delta implies that the incident light must have the same energy as the band difference, but as \( k \) is treated as continuous in band structures, the discrete sum is replaced by the continuous JDOS integral in the second formulation. Multiplying that rate by energy \( \hbar \omega \) to get power loss, and comparing to power loss in a simple Beer-Lambert rule,

\[
\frac{dI}{dt} = \frac{d}{dt} \left( I_0 e^{-\alpha t} \right) = \frac{c}{n} I = \left( \frac{i}{n^2} \right) \left( \frac{n^2}{8} |E(\omega)|^2 \right) = \frac{i}{8} |E(\omega)|^2
\]

Eq. 4-9

one gets a formulation for the imaginary portion of the dielectric function versus JDOS:

\[
i = \left( \frac{2}{m} \right) P_{cv} |D_j(E_{cv})|
\]

Eq. 4-10

This formulation demonstrates that the imaginary portion of the dielectric function is a largely a function of the number of available direct electronic transitions between valence and conduction bands. Further, because \( i(\omega) = 0 \) and the dielectric function does not diverge, the Kramers-Kronig (KK) relations may be used to compute the real portion of the dielectric function,

\[
e = 1 + \frac{2}{P} \int_0^{\infty} \frac{i(\omega)'}{\omega^2} d\omega'
\]

Eq. 4-11

where \( P \) is the Cauchy principle value. Thus, a complete reconstruction of the dielectric function is possible through some knowledge of the band structure. Further, these dielectric properties are directly connected to the complex refractive index of the material as

\[
e = \tilde{n}^2
\]

Eq. 4-12

and therefore the band structure analysis provides a prediction methodology for electronic behavior, and light behavior, during optical/solid interactions.

There are places in the band structure where many momentum-adjacent transitions may occur at the same energy. These are called critical points, or van Hove singularities, because they occur where $D_j$ diverges, which is where

$$\nabla_j(E_{cv}) = 0$$

Eq. 4-13

At these critical points, $\nabla_j$ reaches a local maximum, and appears as peak in the $(\ )$ spectrum (or more rigorously, a local minimum in $\frac{d^2}{d\omega^2}$ due to other potential contributions to $(\ )$). Because these are highly visible features in ellipsometric analysis, they are very useful for comparison to theory. In systems with parabolic band transitions, critical points are categorized according to the curvature along their $\vec{k}$ axes. The transition energy can be written

$$E_{cv}(\vec{k}) = \alpha_1 k_1^2 + \alpha_2 k_2^2 + \alpha_3 k_3^2$$

Eq. 4-14

where the signs of the alpha coefficients determine the curvature directions of the critical point. All positive curvatures are classified as $M_0$, mixed sign curvatures are classified as $M_1$ or $M_2$, and all negative curvatures are classified as $M_3$, for van Hove singularities in three dimensions. Near $M_0$ points $E_{cv}$ increases in all $\vec{k}$ directions, near $M_1$ and $M_2 E_{cv}$ may increase or decrease depending on the direction, and near $M_3 E_{cv}$ decreases in all $\vec{k}$ directions. Typically, and as in Ge, there is a $M_0$ critical point at the center of the band structure (point). This is often the lowest direct transition. Due to the cubic symmetries of Ge and alike crystals, the maxima in $\omega$ are often generated by $M_2$ critical points along the momentum-degenerate $\langle 100 \rangle$ and $\langle 110 \rangle$ directions (Fig. 4-7). The critical points transitions in these directions are called $E_1$, distinct from $E_0$ transitions and the $\langle 111 \rangle$-direction $E_2$ transitions. Knowledge of these trends is helpful in analyzing band structures based on dielectric functions extracted from ellipsometric measurements.
However, because information about electron momentum is lost in linear optical measurements (in that for optical processes that don’t include a 2-step phonon-mediated process, which are rare, are independent of electron momentum), any feature in the measured dielectric function can only be taken to represent an integration of JDOS contributions at a given transition energy over all momenta. For complex band structures with multiple critical points at similar energies, assignment of to one or more of these particular critical points is often impossible, as is the case in Bi$_2$Se$_3$. Thus, while physical theory may support measurements of , complete confirmation of the band structure based on is impossible, and changes to the band structure can’t feasibly be parsed based purely on ellipsometric measurements of .

Another important parameter to consider is the optical conductivity of a material, . This gives the AC conductivity of the impinged material at the light frequency, $J(\omega) = \epsilon(\omega)E(\omega)$. This parameter is above nonzero even for electrically insulating materials, which have DC
conductivities of 0 below band gap biases. The relationship between the optical conductivity and the dielectric function is

\[
\sigma(\omega) = i \omega \epsilon(\omega)
\]

Eq. 4-15

Polarization changes measured in ellipsometry are given by

\[
\rho = \frac{r_p}{r_s} = \tan \Phi
\]

Eq. 4-16

Here, \( r_p \) and \( r_s \) are the complex reflectivities of \( p \)- and \( s \)-polarized light,

\[
\rho = \tan \left( \frac{r_p}{r_s} \right) = \rho_s = \arg \left( \frac{r_p}{r_s} \right)
\]

Eq. 4-17

are the angle given by the ratio of reflected amplitudes, and the difference between phase changes incurred by \( p \)- and \( s \)-polarized light upon reflection from a sample surface, respectively.

\[\Delta = \frac{d_p - d_s}{\tan \Phi} = \frac{\arg \left( \frac{r_p}{r_s} \right)}{\tan \Phi}\n\]

\[= \frac{\omega (E_{f,p}^{\max} - E_{f,s}^{\max})}{\tan \Phi}\n\]

\[\text{Figure 4-8: Polarization of reflected light. This light, traveling in the } +\hat{z} \text{ direction (toward the reader), has a left-handed ellipticity. } \rho = \tan \left( \frac{r_p}{r_s} \right) \text{ is the angle given by the ratio of reflected } p \text{ and } s \text{ amplitudes, and } \Delta = \arg \left( \frac{r_p}{r_s} \right) \text{ is the difference of phase changes incurred by } p \text{- and } s \text{-polarized light. For a linearly polarized incidence (in which } E_{f,p}^{\max} \text{ and } E_{f,s}^{\max} \text{ occur simultaneously), } \rho \text{ can be computed from the time between } E_{f,p}^{\max} \text{ and } E_{f,s}^{\max} \text{ as } \Delta = \frac{\omega (E_{f,p}^{\max} - E_{f,s}^{\max})}{\tan \Phi}\n\]
The reflection and transmission that occur at a sample surface can be determined by the Fresnel equations (as derived in the Appendix):

\[
\begin{align*}
    r_p &= \frac{N_1 \sqrt{N_2^2 - N_1^2 \sin^2 \theta}}{N_1 \sqrt{N_2^2 - N_1^2 \sin^2 \theta} + N_2^2 \cos \theta}, \\
    t_p &= \frac{2N_1 N_2 \cos \theta}{N_1 \sqrt{N_2^2 - N_1^2 \sin^2 \theta} + N_2^2 \cos \theta} \\
    r_s &= \frac{N_1 \cos \theta - \sqrt{N_2^2 - N_1^2 \sin^2 \theta}}{N_1 \cos \theta + \sqrt{N_2^2 - N_1^2 \sin^2 \theta}}, \\
    t_s &= \frac{2N_1 \cos \theta}{N_1 \cos \theta + \sqrt{N_2^2 - N_1^2 \sin^2 \theta}}
\end{align*}
\]

Here, \( N_1 \) and \( N_2 \) are the complex refractive indices of the interface materials, and \( \theta \) is the angle of incidence. For an experiment involving a flat substrate in air, \( N_1 = 1 \), and the Fresnel reflection equations may be directly solved for \( N_2 \). The dielectric function may then be analyzed for information about the critical points of the substrate’s band structure.

### 4.2.2: Spectroscopic Ellipsometry Experiments

The measurements for this work were performed with a J.A. Woollam RC2 dual rotating compensator ellipsometer (Fig. 4-9). It is a spectroscopic (210 to 1690 nm acquisition range), variable angle (25 to 90° AOI) system, in which complete changes in polarization may be measured with four optical components. Incident light may be set to any polarization with a rotating polarizer and first rotating compensator, and after reflection from the sample, a second rotating compensator and rotating analyzer select specific polarization intensities to be measured. The dual rotating compensator design provides significantly higher precision in the determination of than rotating polarizer models. Further, depolarization may be assessed, thus providing information about sample roughness or thickness inhomogeneity.\(^{105,106}\)

In standard SE, reflection from a sample is represented by a Jones matrix \( S \):
In standard ellipsometry, it is assumed that no cross polarization (i.e. transferrance between \( p \)- and \( s \)-polarized light) takes place, which is valid for most unpatterned samples. SE is generally performed near the Brewster’s angle of the sample, which is for transparent samples. At this angle, \( p \)-polarized light is completely transmitted through the interface, and all reflected light is \( s \)-polarized, resulting in high precision ellipsometric data.

Once and data are collected, a model may be generated to match the optical response of the sample. This process was done with J.A. Woollam CompleteEASE for all of the ellipsometric analysis in this thesis. CompleteEASE allows the user to construct a model sample composed of epitaxial layers on a substrate, and to fit the model layers’ physical parameters (e.g. thickness, refractive index or dielectric function, interlayer mixing or surface roughness, etc.) to match the collected data. Because physical materials have KK-consistent optical responses, it is useful to use

\[
\mathbb{E}_r = S\mathbb{E}_i = r \begin{pmatrix} \tan \Psi e^{i\Delta} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} E_{i,p} \\ E_{i,s} \end{pmatrix}
\]

Eq. 4-19

\[
\theta_B = \tan^{-1} \frac{n_2}{n_1}
\]

Eq. 4-20

Figure 4-9: Schematic of a dual rotating compensator ellipsometer. Light from a bright Xe source is polarized at a given angle and ellipticity by the rotating polarizer and first rotating compensator components. After reflection from the sample, polarization direction and ellipticity are filtered by the second rotating compensator and analyzer before the intensity is measured at all wavelengths by the detector.
KK-consistent oscillators to model the electronic response of materials. Two common physical models used in ellipsometric analysis are the Drude model and the Lorentz oscillator. The Drude function takes into account metallic behavior, wherein electrons are not localized, but experience resistance in their medium. It may be derived with simple Newtonian considerations, taking into account the impinging EM wave, and a drag coefficient: \( m\ddot{x} = -e\vec{E} - m\Gamma \dot{x} \). Here, we use an oscillatory \( \vec{E} = \vec{E}_0 e^{i\omega t} \), and take stationary solutions of \( x \) to match its frequency. The amplitude of \( x \) is easily solved, and using the classical definition of the dielectric function, \( D = E = \varepsilon_0 E + P = \varepsilon_0 E + nEc_n \), The Drude dielectric function may be written as

\[
( \varepsilon ) = 1 + \left( \frac{4\pi e^2}{\omega m} \right) \left( \frac{1}{i} \right) = 1 + \left( \frac{\rho}{i} \right)^2
\]

Eq. 4-21

where \( \rho \) is the plasma frequency. As seen in the denominator, the dielectric function is complex, and results in light absorption. The Lorenz oscillator has similar. It is used for locally bound states, and thus has a “spring” center contribution: \( m\ddot{x} = -e\vec{E} - m\Gamma \dot{x} - m\omega_0^2 \dot{x} \). This results in a dielectric function

\[
( \varepsilon ) = 1 + \left( \frac{4\pi e^2}{\omega m} \right) \left( \frac{1}{i} \right)^2 = 1 + \left( \frac{\omega_0^2}{i} \right)^2
\]

Eq. 4-22

where \( \omega_0 \) is the oscillation frequency of the bound state. Real samples have a variety of interband and intraband transitions, each of which often produce their own contribution to the dielectric function, and are therefore modeled with sums of oscillators that contribute to the complete complex dielectric function as \( ( \varepsilon ) = \sum_{m,n} f_D^m + g_L^n \), where \( m \) and \( n \) index individual band transitions, and \( f_D \) and \( g_L \) are the Drude and Lorentz dielectric functions, respectively. These two oscillator types are often all that’s needed to accurately represent the electronic and optical behavior of a sample, and are used to model various transitions in Bi\textsubscript{2}Se\textsubscript{3}. 

86
Once a model is made, in order to attempt a forward solution in which there are more model parameters (e.g. $m$, $\Gamma$, $\omega_0$, and others for a Lorentz oscillator) than there are data points to fit it ($\Psi$ and $\Delta$), it is necessary to perform error-minimizing procedures to match the parameterized model to the acquired data. This is done with the Levenberg-Marquardt algorithm in CompleteEASE:

$$\hat{b} = \text{min}_b S(b) = \text{min}_i (y_i - f(x_i, b))^2$$  \hspace{2cm} \text{Eq. 4-23}$$

where $\hat{b}$ and $x_i$ are the set of parameters (as a vector) that define the model $f$ and the independent variable that produces a signal to match the measured output $y_i$, and $\hat{b}$ is the optimal parameters that minimize the error $S$. This is solved by perturbing the model with a Jacobian $J_i = \frac{f(x_i, b)}{\partial b}$, which differentiates the model with respect to the various parameters, and setting the resultant differential error to zero:

$$S(b + d) \approx (y - f(b))^T (y - f(b)) - 2(y - f(b))^T J (y - f(b)) + (J^T J + l \text{diag}(J^T J)) d = (y - f(b))^T J (y - f(b))$$  \hspace{2cm} \text{Eq. 4-24}$$

In order to vary parameters to match data without changing them excessively, a damping function is required. With it, a solution for an error-minimized function is reached efficiently:

$$(J^T J + l \text{diag}(J^T J)) d = (y - f(b))^T J (y - f(b))$$  \hspace{2cm} \text{Eq. 4-25}$$

4.3: X-ray Photoelectron Spectroscopy

X-ray photoelectron spectroscopy (XPS) is a highly surface sensitive technique that uses x-rays to excite and eject core electrons in order to measure surface chemistry. Because sets of binding energies of core electrons are unique to elements (Table 4-1), and bonding states produce predictable core energy shifts, the detected electron energy spectra can completely identify all surface species and their molecular states.
Because the energy of emitted electrons is low, their extinction distance is quite low as well (the inelastic mean free path of Si $2p_{\text{Si}}$ 1 nm). Thus, the majority of XPS signal is produced from the top few nanometers of the sample surface, and signal intensity decays exponentially with depth. This surface sensitivity is advantageous for samples with epitaxial layers, as composition as a function of depth is measured fairly easily. This is facilitated with angle-resolved XPS measurements (ARXPS), which make use of detector angle, and signal intensity to determine path length and signal origin depth. Spectra are fit with appropriate Bremsstrahlung background models and oscillator-like XPS and auger peaks, and fit to minimize error between the complete model and acquired data. Quantitative analysis is facilitated with libraries of information about XPS energies, chemical states, extinction distances, etc. from NIST.

**4.3.1: Theory of XPS**

XPS uses high-energy photons to excite and eject, and subsequently measure, electrons from materials. The absorption of x-rays provides more than enough energy to allow core electrons to escape from their nucleus, and sometimes, to travel sufficient distances through the sample to escape the surface. The amount of energy required to eject a core electron from a sample (non-bonding $s, p, d...$) is equal to the energy difference between its core state and the Fermi level, called the binding energy, plus the energy difference between $E_F$ and vacuum, called the work function. If any kinetic energy from a state $i$ excited by an x-ray remains after it escapes from the surface, it is measured by the XPS system.
This effect was originally described as the photoelectric effect by Albert Einstein, which won him the Nobel Prize in Physics in 1921.

For a given element, the energy of a state at a given principle quantum number \( n \) can be calculated to first approximation in a Bohr model form, but significant tuning is required to calculate the correct, empirically measured energies. Sources of shifts in state energy include the effects of the orbital and spin quantum numbers, \( j = l + s \), as well as finer terms such as Zeeman splitting, etc. Hydrogenic levels up through fine structure considerations are given by the Sommerfield expression,

\[
E^j_n = \frac{\hbar^2}{8\pi^2m^2c^2} \left[ 1 + \frac{a_j^2}{n^2} \left( \frac{n}{j+\frac{1}{2}} \right)^2 \right] \left( \frac{3}{4} \right) 
\]

where \( \frac{\hbar^2}{8\pi^2m^2c^2} \) is the reduced mass of an electron traveling around its nucleus, and \( \frac{a_j^2}{n^2} \) is the fine structure constant. Fortunately, typical XPS peaks are only split by \( n, l, \) and \( s \), producing individual peaks such as Au 4f\(_{7/2} \), Si 2p\(_{3/2} \), C 1s, etc. Further, a well-documented history of XPS usage has provided

---

**Figure 4-10:** Emission process for photoelectrons and Auger electrons. (a) Photoelectrons come from core states that directly absorb an incoming x-ray and are subsequently ejected from the atom. (b) Auger electrons are produced by a secondary process, in which after a photoelectron is ejected, a higher state relaxes to fill its place, and releases energy for nearby electrons to absorb and excite out of their core state. Adapted from [107].

\[
E^j_k = \hbar \omega - (E_{\text{vac}} - E_F) - (E_p - E^j_{\text{core}}) = \hbar \omega - \phi - E_F
\]

Eq. 4-26

This effect was originally described as the photoelectric effect by Albert Einstein, which won him the Nobel Prize in Physics in 1921.
reliable maps of emitted photoelectron energies in a multitude of materials, so quantum mechanical calculations are not needed to identify peaks. Because core states in given elements are not drastically changed by the bonding states of the host material, and because core states in different elements occur at different energies, the binding energies detected are highly indicative of the elements present in the material.

Core electron energies are not completely independent of bond states, however. The relative electronegativities of bound atoms either acts to attract or repel nearby electrons. When a valence electron is attracted (repelled) to another atom, the core electrons below it experience less (more) screening, which causes an increase (decrease) in their binding energies. This effect can shift the energy levels of core electrons of up to a few eV. For example, the binding energy of Si in a pure crystal is often recorded at ~99.5 eV, but the binding energy of Si in SiO\textsubscript{2} is ~103.5 eV. These “chemical shifts” are too little to confuse different photoelectron peaks, but are easily large enough to be measured, and thus can be used to identify the chemical state of each constituent element.

Core emissions are not the only ones that take place in XPS experiments (Fig. 4-10). Signals from Auger electrons, x-ray satellites, x-ray ghost lines, shake-up peaks (Fig. 4-11), multiplet splitting, energy loss lines, and valence states may also be used to identify the chemistry of a sample.\textsuperscript{107} On the other hand, secondary peaks such as these may also be a source of confusion in the identification of elements in a sample, and can greatly hinder efforts to quantify the concentrations in a given material, due to peak overlapping that can be difficult to decouple.
The energies of collected electrons are not the only information gathered in XPS. The intensity of electron emission at each measured energy peak can also be used to calculate the concentration of the elements in the material. These intensities are a function of the x-ray brilliance, the absorptive properties of the given electron orbitals, the ability of an electron to travel to the sample surface, and the sensitivity of the detector. For a homogeneous sample, the signal intensity is a substantial, but relatively simple function

\[ I = nf \gamma AT \]

Here, \( I \) is the measured intensity, \( n \) is the volumetric density of atoms of the element, \( f \) is the x-ray photon flux, \( \gamma \) is the photoelectric cross section for the orbital of interest, \( A \) is an angular efficiency factor based on the experimental configuration, and \( T \) is the efficiency of the photoelectric

**Figure 4-11:** Example of shake-up lines (s) of the Cu 2p observed in CuO and CuSO₄, but not pure Cu. These can cause misidentification in peaks and elements, as they may overlap with other photoelectron, Auger, etc. peaks. Adapted from [107].
process for formation of photoelectrons of the normal photoelectron energy, \( I \) is the mean free path of the electron in its host material, \( A \) is the area of the sample illuminated, and \( T \) is the detection efficiency for electrons emitted from the sample. The majority of these factors can be grouped together such that \( I = nS \), where \( S \) is called the atomic sensitivity factor. In experiments, these sensitivities are a required part of the quantitative calibration process. They may change from system to system and from material to material. However, well-known sensitivity factors allow a relatively simple calculation to be made for the atomic fractions within a sample

\[
C_i = \frac{n_i}{\sum_j n_j (I_j/S_j) / (I_i/S_i)}
\]

Eq. 4-29

There is, however, an additional problem of separating a photoelectron (or Auger, etc.) peak from the background spectrum. This superimposed continuous signal is from Bermsstrahlung, or braking radiation, and contributes to all XPS signals. It is caused by electrons that are being slowed down and emitting photons as their kinetic energy is lost. Power loss density of traveling electrons is modeled with the quantum electrodynamic Thomas-Fermi function, but the total background spectrum in XPS can simply, generally, and empirically be described by Kramers’ law:

\[
I(l) \propto K \left( \frac{1}{l_{\min}} \right)^{1/2} d
\]

Eq. 4-30

The low wavelength (high energy) cutoff is given by \( l_{\min} = \frac{hc}{eV} \), where \( V \) is the x-ray tube accelerating voltage. There are a variety of methods to decouple this background spectrum, so that true peak intensities and areas may be measured, and quantitative analysis may be performed accurately.

Many XPS systems are equipped with angle-resolved detectors. This enables angle-resolved XPS (ARXPS), which provides information about samples with compositions that change with depth. Electrons are able travel through their host material approximately according to a Beer-
Lambert trend, \( I = I_0 e^{-x_l} \). Here, \( x_l \) is the mean free path of the electron, which is specific to its energy and the medium through which it’s traveling. Because an electron has to travel up to the surface in order to be detected, the intensity of electrons that escape is dependent on the angle they take toward the surface as \( I = I_0 \exp\left(-\frac{d}{\lambda_{x_l} \cos \theta}\right) \), where \( d \) is the depth from which the electron is initially emitted, and \( \theta \) is the angle that the electron’s path makes with the surface normal (Fig. 4-12). It is important to examine this effect on multilayered systems. The intensity that results from a buried layer \( n \) is

\[
I = I_0 \exp\left(\frac{1}{\cos \sum_{i=1}^{n} \frac{d_i}{\lambda_i(E)}}\right)
\]

Eq. 4-31

Here, \( d_i \) is the vertical distance to the layer (or sample surface) above (Fig. 4-12). This is a bit of an oversimplification, as the initial intensity itself can be a function of emission angle. However, for known mean free paths, these relationships can be used to calculate the depth from which an elemental signal originates, and can thus be used to non-destructively find the thicknesses of
distinct thin layers. For substrates with single epitaxial layers, the low angles often have far higher bulk intensities, and the high angles have far higher epitaxial intensities.

4.3.2: XPS Equipment and Measurement Description

XPS and ARXPS data were collected using a Thermo Scientific Theta Probe™ system using monochromatic Al Kα (1486.6 eV) x-rays of 400 μm spot-size. All samples were loaded simultaneously into the UHV chamber, and the measurements were made at a base pressure of 8 × 10⁻⁹ Torr. The measurements were taken with an energy step of 0.05 eV and at pass energy settings of 100 and 50eV for Bi 5d and Se 3d peak regions, respectively. The energy resolution is estimated to be 1.0 eV from the full width at half maximum of the Au 4f⁷/₂ core level. In the absence of a well-defined C 1s peak after exfoliation, the energy scale of the XPS spectra was calibrated with respect to the Se 3d⁵/₂ core level (53.7 eV) according to the value in the National Institute of Standards and Technology (NIST) database. We estimate the uncertainty with all peak binding energies to be 0.05 eV from peak fitting. ARXPS data were measured by collecting photoelectrons in 16 channels over an angular range of 24.875 – 81.125 using a spherical sector analyzer fitted with an electrostatic lens having 60 angular acceptance and an angular dispersive multichannel detector.

4.4: Transmission Electron Microscopy

Transmission Electron Microscopy (TEM) is a multifunctional technique that, while time intensive, can provide a wealth of information from a localized area regarding crystallinity, structural defects, chemistry, and electrical behavior. Many of these can be done simultaneously. As an imaging technique, its combination of fast acquisitions and sub-angstrom resolution are
unique, although some scanning tunneling microscopes may image a surface with similar resolution. Recently, advanced ‘intelligent’ software has enabled automated sample preparation and data acquisition, drastically decreasing the human time and effort requirements to take these data. While these experiments are still time intensive and destructive, and cannot be used in-line to assess samples at a manufacturing scale, this new automation is facilitating the expansion of TEM systems into production-level quality assurance and research and development, making them ever more relevant and in demand.

4.4.1: Theory of TEM

For an electron to transmit through a sample, it must be traveling with extremely high kinetic energy, and the sample must be very thin. In order to project an electron at such rates, a TEM uses accelerating voltage in the range of 100 kV – 1 MV. This results in a relativistic velocity and wavelength

\[ v = c \left[ 1 \left( \frac{eV}{m_0c^2} + 1 \right) \right]^{\frac{1}{2}} = h \left[ 2m_0eV \left( 1 + \frac{eV}{2m_0c^2} \right) \right]^{\frac{1}{2}} \quad \text{Eq. 4-32} \]

For an accelerating voltage of 200 keV, the de Broglie wavelength is 2.5 pm, which gives a Rayleigh resolution limit of \( d_R = \frac{0.61}{\sin b} \frac{0.61}{2.2} = 2.2 \text{ pm} \). However, no electron source or lens systems behave ideally, and the aberrations in real TEM systems each effectively increase the probe size. These may originate from the source, \( d_s = \frac{2}{\sqrt[4]{I}} \frac{1}{C_s} \) (where \( I \) is the current, \( b \) is the brightness, and \( a \) is the convergence angle), from spherical aberrations, \( d_{sph} = C_s a^3 \), and from chromatic aberrations, \( d_{c} = C_c \frac{E}{E} \). The final resolution of a TEM can thus be approximated as

\[ d = \sqrt{d_s^2 + d_{sph}^2 + d_c^2 + d_R^2} \]  

While picometer resolution may not yet be attainable in modern TEM, sub-
angstrom features may be resolved with aberration correction. This process involves the use of extra lenses and monochromators which each must be tuned individually.

A TEM can be operated in two modes. Conventional TEM is performed with a beam that is parallel upon incidence with the sample. The exiting beam is scattered (at varying intensities) into a diffraction pattern, which may be measured, or some diffraction spots may be selected with apertures to produce images. Propagated diffractions create images that may contain varying information depending on the choice of spot(s), and are generated by different forms of contrast. Many TEMs are also capable of operating in scanning TEM (STEM) mode, where the beam converges to a small point in the sample in a conical shape (Fig. 4-13). In some ways, this mode is analogous to an ultra high-energy scanning electron microscope. Like TEM, this mode produces both a direct beam and scattered electrons that are still able to transmit through the sample, but apertures are replaced by detectors to generate different types of images.

Amplitude contrast imaging in TEM and STEM can be taken in bright field (BF) or dark field (DF) mode. The former is a measurement of electrons that pass directly through the sample without being scattered, and the latter involves selecting one or more scattered electron beams to propagate toward the detector to create an image. BF and DF mode are each capable of displaying mass-thickness contrast. This is a mode in which elastic Rutherford scattering events decrease the current transmitted through the sample as \( \frac{I}{I_0} = 1 - e^{-Qt} \approx Qt \), where \( Q \) is the total elastic scattering cross section (which takes into account atomic weight, number density, and the single atom cross section) and \( t \) is the sample thickness.\(^{108}\) Additionally, DF STEM imaging itself can be performed in two different ways, depending on the exit angle of the scattered transmitted beam. In addition to elastic scattering events that deflect the beam at relatively low angles, inelastic scattering events generate highly deflected beams that may be detected to generate images. These are often
called ‘Z-contrast’ images, as the intensity of inelastically scattered electrons is proportional to $Z^2$, making this technique highly sensitive to heavier elements. This may be performed in annular dark field (ADF) or high angle annular dark field (HAADF) mode. Both of these modes utilize disc-shape detectors around the direct beam to collect electrons scattered above ~10 mrad. ADF was developed earlier, and functions by collecting electrons scattered at an intermediate angle ($10 < \theta < 50$ mrad), and may be susceptible to signals from diffracted electrons that can deflect at lower angles. HAADF collects electrons inelastically scattered at a high angle ($> 50$ mrad), and is a commonly used tool in high resolution STEM imaging. To produce these images, the STEM beam rasters a (aberration corrected, in the high resolution case) sub-angstrom sized probe across the sample, and electrons scattered from individual atomic columns are detected, thus producing an atomic-resolution direct image of the crystal lattice, wherein brightness correlates with heavier elements.

4.4.2: TEM Experiments

The TEM used in this work FEI Titan³ G2 microscope, operating at 300 keV in HAADF STEM mode with spherical aberration correction. The source is a field emission gun (FEG) equipped with a monochromator, which in this case, is used to decrease the beam intensity at all energies rather than to filter for a narrow energy band. This lowered current level facilitates high resolution measurements. Emitted electrons are accelerated into the column, where 3 condenser lenses generate a highly focused probe. Before the beam impacts the sample, a spherical aberration corrector further decreases the probe size to sub-Angstrom levels needed for high resolution STEM imaging. The beam then interacts with the sample, deflecting off of atomic columns, and
inelastically scattered electrons are collected by the HAADF detector to produce atomic-resolution images.

4.5: Atomic Force Microscopy

Atomic force microscopy (AFM) is a high-resolution, potentially non-destructive topographical measurement technique. It can produce maps of surface features that contain sub-nm resolution height information over > 1000 μm² surface areas, generate 3D profiles of nanoscale features such as FinFETs, identify biological formations with unique structures, etc. It is an essential tool for any topographical studies, and can be used to measure surface feature dimensions with higher throughput (although lower resolution) than TEM, and higher resolution (although lower throughput) than optical methods. Its ability to measure surface textures makes it highly valuable for fabrication processes that critically depend on line edge roughness or line width roughness.

4.5.1: Theory of AFM

In topographic scanning modes, AFM uses a vibrating, tipped cantilever to detect a sample surface level with high accuracy. The sharp, pyramidal tip is located under one end of the cantilever, and the other end is mounted to the microscope. The most common type of measurement is called ‘tapping mode’ AFM, in which the cantilever is vibrated at its resonant frequency and placed near the surface. Meanwhile, a laser is reflected off the top surface of the vibrating cantilever, and a photodetector detects the frequency and amplitude of these vibrations. (Fig. 4-14) Highly sensitive piezo motors in the z-direction move the cantilever down until its vibrations cause they tip to lightly tap the sample surface. The force that the surface generates on
the tip causes the cantilever to deviate slightly from its maximum amplitude and its resonant frequency, \( f_0 = \frac{1}{4\pi^2} \sqrt{\frac{k}{m_0}} \), where \( m_0 \) is the effective mass of the cantilever and \( k \) is its spring constant,

\[
k = \frac{6Ec^3}{4l^3}
\]

\((E \text{ is the cantilever’s Young’s modulus, } w \text{ is its width, } t \text{ is its thickness, and } l \text{ is its length}).

These deviations are corrected by a feedback loop that uses a PID error minimization algorithm to adjust the tip height and keep the amplitude of oscillation at its set point. This set point is not directly at the center of the resonant frequency peak during scanning, but (typically) just below it, which allows the microscope to discriminate between amplitude changes due to higher tip force levels (which lower the amplitude from the resonant peak) and lower tip force levels (which raise the amplitude toward the resonant peak). The force that the surface generates on the tip is caused by changes in \( z \)-height at topographical features. The high sensitivity of this cantilever amplitude modulation detection mode and the PID correction algorithm help generate highly sensitive topographical surface maps.

**Figure 4-14:** AFM operation. (a) The tip interacts with the sample while a laser reflects off its back and into a detector. (b) The amplitude and frequency of the tip’s vibrations are measured during the scan. As the height of the surface increases, the force on the tip increases, and its vibration amplitude decreases. As the height of the surface decreases, the force on the tip decreases, and its amplitude increases toward the resonant frequency peak.
4.5.2: AFM Measurements

AFM measurements in this thesis were performed with a Dimension Icon AFM in tapping mode. The probes used in topographical measurements were OTESPA aluminum back-coated, etched silicon tips from Bruker. These have resonant frequencies near 300 kHz, and the aluminum back coating helps create stronger reflections. Scans were performed at 512 or 1024 px$^2$ resolutions at rastering frequencies below 1 Hz. No damage was measured on any of the samples during acquisitions. Some images contained highly regular corrugations that were assumed to be artifacts due to their coexistence with more defined features of expected sizes and depths, such as QL steps. These may have been caused by the slight surface tilt of exfoliated bulk samples. These were removed with Fourier filters in analysis.

4.6: Temperature-Programmed Desorption

Temperature-programmed desorption (TPD) is a method of tracking surface desorption rates as a function of temperature. Because rates can be tracked at UHV pressures, this technique is capable of measuring the vapor pressure of thin epitaxial layers. Analysis with TPD can be informative about the kinetics of a desorption event, the coverage of adsorbates on a surface, and with a sufficiently accurate system, it can provide the activation energies of individual desorptions. This may be employed to find the energy of binding sites at the surface of samples.

4.6.1: Theory of TPD

Desorption is a thermodynamic process in which surface molecules evaporate into the surrounding gaseous environment. The rate at which molecules desorb is dependent on the energy
of their surface bond and the temperature of the system, and is given by an Arrhenius relationship,

\[ \dot{\text{des}} = 0 \exp \frac{E_a}{kT} \]. For homogeneous materials simply evaporating into the surrounding environment, the surface evaporation rate (in molecules cm\(^{-2}\) sec\(^{-1}\)) is given by the surface molecular density and their desorption frequency,

\[ E = \dot{\text{des}} \] \hspace{1cm} \text{Eq. 4-33}

For systems in which a surface has a specified fractional coverage of evaporating adsorbates, a model may created in which evaporation rate is dependent on fractional surface coverage. The basis for this is given by the Polanyi-Wigner (P-W) equation,

\[ r(\theta) = \frac{dn}{d\theta} = \dot{\theta} = n \exp \frac{E_a}{kT} \] \hspace{1cm} \text{Eq. 4-34}

Here, \(\theta\) is the fractional surface coverage and \(n\) is the kinetic order of the desorption event. Different \(n\) values correspond to different kinetics of desorption e.g. \(n=0\) gives the equation for bulk or multilayer desorption in which the desorption rate is independent of coverage, \(n=1\) describes direct desorption from a surface in which the desorption rate is linearly dependent on coverage, and \(n=2\) gives the equation for a multi-step surface desorption process (for example, recombination of two H atoms to desorb as the gaseous H\(_2\) molecule). The \(n=0\) case results in evaporation rates that are purely dependent on the atomic surface density and the Arrhenius desorption frequency, and the evaporation rate reverts to Eq. 4-33. The P-W equation (Eq. 4-34) describes the rate of surface desorption as a function of time.

TPD is generally performed by linearly increasing the temperature of the sample \(i.e.\)

\[ T(t) = T_0 + t \]. This allows substitution of the time variable, \(\frac{dn}{dt} = \frac{-\dot{\theta}}{n} = \frac{-\dot{\theta}}{\theta} \), and the P-W may be rewritten
In this study, the desorbing material is a bulk Se cap, which experiences a zero order kinetics, and simplifies the equation to \( \frac{\partial q}{\partial T} = \tau_0 \exp \left( \frac{E_a}{k_B T} \right) \). This describes what fraction of the surface will desorb per second per increase in degree Kelvin. Desorption rates are directly proportional to the partial pressure they create, \( \frac{\partial P}{\partial T} \), which is an easily measurable thermodynamic variable. The kinetics of desorption can then be derived with pressure measurements,

\[
\frac{\partial P}{\partial T} = \tau_0 \exp \left( \frac{E_a}{k_B T} \right)
\]

**Eq. 4-36**

In ideal conditions, TPD can be used to measure this pressure change during a linear temperature ramp, and the activation energy can be extracted from an Arrhenius plot \( \ln \frac{P}{T} = \ln \frac{P}{T_0} + C \) by taking fitting the slope of the curve as \( 1/T \), \( k \frac{\partial P}{\partial T} \ln \frac{P}{T} = E_a \).

Another method to predict desorption rates in UHV conditions makes use of the vapor pressure at the surface. This is the pressure required to create an equilibrium such that the rate at which molecules are emitted is the same as the rate at which they are adsorbed. The rate of adsorption is calculated with the flux, \( g = \frac{1}{3} n \bar{v} \), where \( n \) is the volumetric molecular density of the gas, and \( \bar{v} \) is the average thermal velocity. This is the rate at which gas molecules impinge on an area of the surface. The average thermal velocity is computed with Boltzmann statistics,

\[
\bar{v} = \int \int \int f(v)dv = \int_{\frac{m}{2kT}}^{\infty} v^2 \int_{\frac{m}{2kT}}^{\infty} v^2 e^{-\frac{m v^2}{2kT}} dv = \sqrt{\frac{8kT}{m}}
\]

**Eq. 4-37**

Using the ideal gas law allows substitution of \( n \), and the flux is

\[
g = \frac{P}{\sqrt{\frac{2m}{kT}}} \]

**Eq. 4-38**

The units for this are molecules cm\(^2\) sec\(^{-1}\). The equilibrium is described by the case when the net
flux is zero, as given by the difference between the vapor pressure caused by desorption from a surface and the ambient partial pressure of the surrounding gas,

\[ \frac{d}{dt} = \frac{P_n - P_{v}}{\sqrt{2\pi m k T}} \]

Eq. 4-39

In a UHV environment, the partial pressure is very low, but a material’s vapor pressure may be high. The surface and surrounding gas systems are not in equilibrium when the vapor pressure dominates the pressure of the ambient gas. In this case, the net flux (desorption) is given by the vapor pressure. Further, if material density is known, Eq. 4-2 can be modified to calculate changes in thickness due to desorption. This can be expressed as

\[ \frac{d}{dt} = \frac{P_n}{\sqrt{2\pi m k T}} \]

Eq. 4-40

where \( n \) is the molecular density of the material.

4.6.2: TPD Experiments

TPD was performed in a chamber custom built by Carl Ventrice and his research group (4-15). It contains a UHV chamber for measurements and a high vacuum chamber for pre-pumping samples, thus allowing successive measurements without exposing the UVH chamber to air, and obviating the need for bake-outs between samples. The TPD measurements rely on temperature regulation and mass spectrometry systems that run simultaneously.

Temperature is constantly measured by a type-K thermocouple in physical contact with the sample surface. This thermocouple is thermally isolated from the rest of the Cu column by ceramic mounting washers. As the voltages produced by thermocouples are low, this signal is increased by a preamplifier before being fed into a data acquisition board which is able to communicate with a nearby computer. Behind the sample, there is a tungsten filament that radiates when at a sufficiently high current. An annular aperture on the sample mounting plate allows this to directly
reach the back of the sample. On the computer, a program written in LabVIEW monitors the temperature of the sample surface, and sends an outgoing signal to a power source to supply current to the W filament, and is therefore able to adjust the heating rate based on the recent temperature trend. This program relies on a PID error minimization model in order to keep a steady, linear ramp rate.

The mass spectrometry is conducted with a Hiden Analytical HAL 201 RC quadrupole mass spectrometer (QMS), which is a type of a residual gas analyzer (RGA). Prior to entering the QMS, molecules are ionized with 70 eV electrons. The QMS is composed of a four parallel metal rods, which when observed along their long axis appear as corners of a square. Rods in opposite corners

Figure 4-15: Schematic of TPD system. (a) After insertion through a load-lock, the sample is placed in a large UHV chamber. Signals from the thermocouple are fed into a nearby computer, where a PID program regulates heating to a linear ramp by signaling a power source to increase (or decrease) the current running through a heating filament behind the sample. Meanwhile, a mass spectrometer tracks the partial pressure of desorbing gases. (b) The sample is mounted on a plate with an annular hole that lets the filament radiation through to directly impact the back of the sample. A thermocouple in physical contact with the surface sends temperature readings to the computer.
are paired, and given an equal bias. The two pairs have opposite biases, which are fluctuated at radio frequencies. These biases are able to filter out singly ionized molecules according to their masses, and the QMS is able to quickly sweep through relevant mass ranges in order to quickly detect desorbing molecules. This model is has a mass range of 200 amu, which disallows the direct measurement of Bi desorption. However, all isotopes of Se atoms are measurable in this range.

Desorption rates, as estimated from vapor pressure, can be calculated in two ways. One is to use a lookup table, such as from Honig and Kramer\textsuperscript{109}, in order to find the vapor pressure. Another way is to use perform a direct calculation with the semi-empirical simplified Antoine equation, \( \log_{10} P_v = A + \frac{B}{T} \), and use reported literature values for the \( A \) and \( B \) parameters. For known solids, the vapor pressure can be used to calculate temperature. In the system used for these experiments, the temperature could not be accurately calibrated for \( \text{Al}_2\text{O}_3/\text{Bi}_2\text{Se}_3/\text{Se} \) samples. The thermocouple read accurately near room temperature, but high temperature tests involving electron beam heating of a Si coupon sample showed significant (consistently >200 °C) discrepancies between the thermocouple and a visible wavelength pyrometer. As the pyrometer could not read below 600 °C, these trials were conducted between 600 and 1100 °C. As this region is far outside the range of interest (Se expected to desorb near 130 °C and \( \text{Bi}_2\text{Se}_3 \) expected to desorb near 325 °C), and the measured values between the pyrometer and thermocouple were significantly different, the initial Se desorption event was used to calibrate temperature. The thickness change of the Se cap due to evaporation is given by \textbf{Eq. 4-40}. The vapor pressure of Se reaches 5×10\textsuperscript{-5} torr near 130 °C\textsuperscript{109,110}, Se has three atoms per unit cell, and the volume of a unit cell is \( V = a^2c\sin 60 = 81.78 \text{ Å}^3 \), giving a thickness evaporation rate of 2.77 nm sec\textsuperscript{-1}. This is approximately the average rate needed to
evaporate the Se cap in 2 minutes, as shown in Fig. 4-16(a), and thus is assigned to the middle of the Se desorption increase ramp. The desorption, as indicated by high partial pressures of residual gases, is then plotted as a function of temperature in the region where temperature is a linear function of time.

Unfortunately, this method of temperature calibration is not accurate beyond a short range, and not reliable between samples. Se cap desorption events were measured between 60 and 110 °C, so no single TPD spectrum could be used as a general calibration. Rather, TPD measurements were performed at a nominally linear temperature ramp, and the temperature data were recalculated based on the Se peak position in each measurement. TPD measurements that were calibrated this way indicated higher temperatures than possible beyond the Se desorption peak. Figure 4-16(b) shows all of the gas species measured in TPD, and shows that a temperature of 375 °C is not sufficient to produce a second Se peak due to Bi$_2$Se$_3$ desorption. The partial pressure of Bi$_2$Se$_3$ is

![Figure 4-16](image)

**Figure 4-16:** Typical temperature calibration and full array of TPD-measured molecules. (a) Temperature calibrations were performed near the midpoint of Se desorption rate, and taken to indicate a vapor pressure of $5 \times 10^{-5}$ Torr. (b) Even this direct calibration is not valid, as it results in ‘calibrated’ temperatures of up to 375 °C in the linear ramp range that do not show Bi$_2$Se$_3$ desorption. TPD is therefore only used as an indicator of desorbed layers, but not as a valid measurement for desorption temperatures.
known to reach $10^4$ Torr at 375 °C, and we therefore conclude that no attempted method of TPD temperature calibration is sufficiently reliable to produce the correct temperature readings. TPD is therefore used as an indicator of desorption events, and is taken to be a reasonable qualitative guideline to temperatures near Se desorption peak, but is not sufficiently accurate for calculations regarding the kinetics of Se or Bi$_2$Se$_3$ desorption.

4.7: Low Energy Electron Diffraction

Low energy electron diffraction (LEED) is a method of probing the crystalline structure of samples very close to the surface. It relies on the wave nature of the electrons to diffract with the periodicity of the lattice and the low kinetic energy of the electrons to only produce signal within the top ~1 nm of the surface. It was one of the early confirmations of the de Broglie wavelength of electrons, as it created lattice patterns that be compared to reference x-ray diffraction experiments (a 100 eV electron has $\lambda = 1.2264 \, \text{Å}$, similar to the Cu $K_{\alpha}$ wavelength of 1.5406 Å) where the wavelength dependence of light diffraction was known. It is a standard surface science tool, operating in UHV conditions to indicate the 2D crystallinity of the growths at surfaces. Further, LEED can be used in a quantitative mode where the intensity of diffraction points as a function of the electron accelerating voltage can be informative about specific atomic positions (as opposed to lattice parameters) by comparison to a dynamical scattering model, or used to calculate the work function at the surface by detection of emitted secondary electrons as a function of incident electron accelerating voltage and sample bias. As only the verification of intact crystal structure was needed in this thesis, LEED was only used for lattice assessment. It should be noted that observation of a LEED pattern does not guarantee that the whole surface is ordered, as LEED
spot sizes are often on the order of 1 mm$^2$, and a detailed assessment of surface crystallinity requires a supplemental microscopic characterization technique.

4.7.1: Theory of LEED

LEED is a highly surface sensitive measurement. This is due to the low kinetic energy of the electrons traveling through the solid sample, and the low mean free path that such a low kinetic energy generates. Unlike in particle travel through gas or vacuum, which may be modeled with a Beer-Lambert function, electrons traveling through a solid are subject to many different scattering events, such as interactions with phonons, plasmons, individual valence electrons, etc. These interaction events happen at wildly varying energy levels, and thus a complex model is required to calculate the mean free path of an electron as a function of its kinetic energy. In lieu

![Figure 4-17: Universal mean free path curve of electrons traveling through solids. For kinetic energies between 20 and 200 eV, this is less than 1 nm, which indicates that electrons traveling at these energies scatter rapidly. This produces a high degree of surface sensitivity in LEED, as few electrons are backscattered after the top 1 nm of the surface. Adapted from [5].](image)
of such a calculation, a reference to a “universal curve” for the mean free path of electrons will suffice (Fig. 4-17). Because surface sensitivity is desired in LEED, these experiments are usually carried out with accelerating voltages between 20 and 200 V. The mean free path in this energy region is below 1 nm, which ensures that electrons that impinge on the sample surface are rapidly scattered, and that any information gathered from scattered electrons originates near the surface. LEED detects electrons that are elastically and coherently backscattered. The coherence maintained during this process leads to a collective phase acquisition during scattering, which

\[ k - k_0 = g_{hkl} \]

**Figure 4-18:** Schematic of crystal diffractions in LEED. The incoming beam, labeled \( k_0 \), is diffracted with reciprocal lattice vector \( g_{hkl} \), and leaves the crystal with momentum \( k \). Because the interaction volume is so short, the reciprocal lattice vectors are blurred in the surface direction, thus allowing many reciprocal lattice vectors with \( g \parallel g_{surf} \) to produced the same diffraction. Even though this \( k-k_0 \) does not exactly coincide with a reciprocal lattice vector, the overlap with the truncation rods allows diffraction.
creates a diffraction pattern. This can be described by the change in momentum (similar to XRD) that an electron receives during the scattering process,

\[ \vec{k} - \vec{k}_0 = \vec{g}_{hkl} \quad \text{Eq. 4-41} \]

\( k \) (\( k_0 \)) is the outgoing (incoming) electron momentum (from the de Broglie equation \( \hbar k = p = \sqrt{2mE} \)), and \( g_{hkl} \) is a reciprocal lattice vector. Because the interaction depth is so low, the effective reciprocal lattice points tend to spread out into truncation rods in the surface direction, thus allowing some diffractions that coincide with changes in \( k \) along the surface plane, but do not exactly coincide with the surface direction of the \( g_{hkl} \) vector (Fig. 4-18). This allows diffractions in many Laue zones to be measured simultaneously.

4.7.2: LEED Experiments

LEED is conducted with an electron gun pointed normal to the sample surface (Fig. 4-19) and a hemispherical detector grid surrounding it. The gun filament is negatively biased to extract electrons, which are then accelerated at voltages below 100 eV toward the sample. Upon reflection, these electrons leave the surface with angles acquired by their diffraction vectors. The detector system is composed of four concentric hemispherical grids and a phosphor screen. The middle grids are negatively biased, and serve to filter out inelastically scattered electrons from the surface that would otherwise generate noisy signals. The outer two grids are attached to ground and prevent stray fields. Diffracted electrons have sufficient energy to surpass the filters, and impact the biased (2-5 keV) phosphor screen, which creates a visible LEED pattern.
Scanning tunneling microscopy (STM) is a method of imaging a surface by rastering a conductive tip close enough to it that electrons are able to tunnel through the vacuum that separates the tip from the surface (Fig. 4-20). Its resolution (sub-angstrom) and surface sensitivity make it a uniquely useful tool (Fig. 4-21). Its applications to basic science earned its inventors a Nobel Prize in 1986. STM has been used to observe surface diffusion and the resultant growth and expansion of terraces, to image crystalline surface reconstructions, and to manipulate the positions of atoms by attracting them with a bias ramp, dragging them with the STM tip, and applying another bias ramp to release them\textsuperscript{112}. Scanning tunneling spectroscopy is an extension of STM, in which the tip bias is varied and the current generated between the tip and sample is measured. This technique
gives electrical information at the surface with high spatial resolution, thus allowing some electronic properties to be mapped as a function of position.

4.8.1: Theory of STM/STS

STM measurements are conducted such that the probe tip never touches the sample surface, although it approaches very closely. There is no direct line by which electrons may classically conduct in these experiments. Instead, the fundamental physics of STM/STS involves quantum tunneling. Electrons on the surface of the sample do not have sufficient energy to make it over the energetic barrier of vacuum, but with sufficient bias and minimal distance from tip to sample, the tunneling current rises to measurable levels. Calculations of tunneling current in a typical STM environment require a brief derivation.

An electron traveling with energy $E$ can be described by the Schrodinger equation

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi + V(x)\psi = E\psi.$$  

The wave function has the solution

**Figure 4-20:** Schematic of an STM. An atomically sharp tip is positioned within 1 nm of a surface and biased. The resulting tunneling current is monitored, and usually kept constant with a feedback loop that is sensitive to the sample-tip distance.
For regions where $U(z) < E$, $k$ is real, and electrons oscillate at with wave number $k$. This is a classically viable energy regime. However, for regions where $U(z) > E$, classical particles are forbidden. In the quantum regime, though, this creates an imaginary $k$ value, and the wave function has exponentially decreasing amplitude. For an abrupt transition into the high potential, this amplitude is dependent on the distance traveled into the classically forbidden region, $x - x_U$, and the probability that an electron may have a given position in this potential is $P(x) = |\psi_0|^2 \exp(-2k(x - x_U))$. In the case of STM, the distance between the tip and surface is on the order of 1 nm or less, and will be given here by $d$, and the difference between the high vacuum potential between the surface and tip, and the average electron energy ($E_F$), is given by the work function, $\phi$. The probability of an electron tunneling to the tip in such a system is

$$P_{tunnel} = |\psi_0|^2 \exp\left(-\frac{\sqrt{8mE_F}}{\hbar^2} \phi\right)$$

This probability is extremely small for $d > 1$ nm, as in most systems, the work function is
significantly greater than $\frac{e^2}{md}$ (the product $\frac{me^2}{\hbar} \phi = 62$ for the electron rest mass, $d = 1$ nm, and the work function of Si, 4.75 eV, resulting in a negligible probability). However, the tip distance can be modified with sub Angstrom-level precision in an STM, allowing the surface-tip distance to decrease to a few angstroms or less, greatly increasing tunneling probability and resulting in measurable current levels. Further, a bias between the sample and tip increases the tunneling probability by changing the Fermi level of the tip or sample. Instead of a constant high vacuum potential, this results in a ‘slide’ down the potential barrier from sample to tip (or from tip to sample given the opposite bias).

In order to calculate real current, a model that takes into account temperature and density of states DOS is required. The rate at which an electron can transfer between states of different energies, e.g. surface and tip states, is given by Fermi’s golden rule, $R_{\alpha \beta} = \frac{2\pi}{\hbar} |M_{\beta\alpha}|^2 \delta(E_i - E_f)$, where $M$. For varying surface and tip DOS, many such transfer events may occur, thus requiring an integration

$$I_{\alpha \beta} = \frac{2\pi}{\hbar} \int_{-\infty}^{\infty} |M_{\beta\alpha}|^2 N_t(E - eV) [1 - f(E - eV)] N_s(E) f(eV) dE$$

Eq. 4-44

Here, $N_{t,s}$ are the DOS of the tip and sample, and $f$ is the Fermi-Dirac distribution. The multiple uses of $f$ indicate that there must be both a filled state to tunnel from (in the sample, in this case), and an empty state to tunnel into (in the tip). In real systems, although bias pushes current in one direction, there’s a real probability of back current, and thus the net current requires a sum of current directionalities,

$$I_{\alpha} = \frac{2\pi}{\hbar} \int_{-\infty}^{\infty} |M_{\beta\alpha}|^2 N_t(E - eV) N_s(E) [f(eV) - f(E - eV)] dE$$

Eq. 4-45

Further simplifications can be made for low temperatures and biases. At $T = 0$, $f$ acts as a step function, which allows the integration to be made between $E_F$ and $E_F + eV$. For tips that have near
constant densities of states in these energy ranges, that term can be treated as a constant. Finally, a consideration by Bardeen showed that for most realistic quantum tunneling situations, the matrix element is virtually independent of the difference in energies between the two sides. The tunneling current may finally be written

\[ I_{\text{ne}} = \frac{2\pi e}{h} |M_{ji}|^2 N_f \int_{E_f}^{E_f+eV} N_i(E) dE \quad \text{Eq. 4-46} \]

and the sample’s local density of states (LDOS) at the tunneling site can be represented by the derivative

\[ \frac{dI}{dV} = \frac{2\pi e}{h} |M_{ji}|^2 N_f N_i(E_f-eV) \propto N_i(E_f-eV) \quad \text{Eq. 4-47} \]

This is the essential equation that governs STS. It allows a surface LDOS to be investigated to high precision.

Further STM physics considerations can be made for the often-spherical geometry of the tip, the transfer matrix value, the real DOS for both the tip and sample, and the effect of orbital shape on current levels, but those are left to the reader.

STM experiments are usually conducted such that the tip bias is held constant during scans, and the tip distance is modified such that the current is maintained at a specified level. As shown in Eq. 4-32, higher currents require the tip to be closer to the surface, and vice versa. This relationship makes the STM sensitive to surface topography. As the tip moves to a feature protruding out of the surface, the current is briefly increased as the tip distance decreases, and the system raises the tip in order to keep the current at the specified level. Likewise, dips in topography briefly decrease the current, and the system responds by pushing the tip downward toward the surface, maintaining the current level.
4.8.2: STM/STS Experiments

Before inserting each $\text{Bi}_2\text{Se}_3$ sample into the STM, the sample is placed on a molybdenum plate and loaded into a UHV chamber for annealing, in order to degas the surface. A radiative heater is utilized to remove the Se cap in samples that had not been previously annealed. For samples that were thermally decapped in a previous chamber (TPD), the STM anneals had maximum temperatures that were 20 °C below the initial anneal, in order to minimize further modifications to the $\text{Bi}_2\text{Se}_3$. This heating chamber is located in situ with the STM, connected by UHV transfer tubes, allowing surface measurements to be performed without breaking vacuum, thus minimizing surface contamination. Samples were photographed in order to ensure that the anneals caused complete Se cap removal.

All STM and STS measurements are performed with an Omicron LT STM. Measurements are done at room temperature (300K) and under UHV. The probe is a Pt-Ir wire that has been mechanically cut to form an atomically sharp tip. Each STM measurement is done with a tip bias of 0.487V to ensure stable tunneling above the band gap of $\text{Bi}_2\text{Se}_3$. A feedback loop is monitored with a tunneling current of 2.1 nA in order to maintain a constant current across the 1000×1000 nm image area. Each STM image is taken with 2.5 nm step resolution.

In STS measurements, a 10×10 grid area is used over the 1 um² imaging area, producing spectral maps with 100 nm node spacing. The tip bias in each spectrum is swept from +0.487 V to -0.487 V with a resolution of 2.5 meV, and data is acquired at a rate of 320 μs per point. The feedback loop is disabled during STS acquisitions. Forward and reverse sweeps are performed at each grid location for a total of 200 spectra taken per image. Because we assume minimal hysteresis, these are then averaged to increase the quality of the data. A 50us delay is used between each sweep of STS to account for any change in the tip location due to drift.
Chapter 5: Surface Oxidation in an Ambient Environment

Aging and surface oxidation are highly disruptive processes at the surface of Bi$_2$Se$_3$. The alterations they cause in the chemistry, structure, and electronic behavior near the surface can have detrimental impacts on attempts to utilize TSSs$^{40,70}$, but are not well understood mechanistically. It is critical to the TI behavior of Bi$_2$Se$_3$ that the Fermi level remains in the gap, even when the crystal is exposed to air, as is a likely scenario during fabrication. Exposure to various air gases (CO, H$_2$O, O$_2$…) effects have been correlated with band bending and the emergence of a Rashba-split 2DEG$^{92}$, which exhibits metallic conduction with a sufficiently high DOS to drown out conductivity measurements of the TSSs.$^{40}$ Various properties of the oxide itself are also not well documented. Bi and Se oxidation states have been shown to grow on the surface,$^{70,113}$ but overall composition of the oxide, the rate of its growth, the surface morphology it creates, and its electrical and optical properties must be studied in order to assess the impact that oxidation has on the Bi$_2$Se$_3$ crystal and its surface states. Therefore, attempts to characterize these processes are critical to a comprehensive physical picture of the overall electrical behavior of Bi$_2$Se$_3$ and its TSSs. This is a necessary preliminary step to implementing Bi$_2$Se$_3$ in devices, as fabrication processes involving Bi$_2$Se$_3$ must be sensitive to its defect vulnerabilities to circumvent them.

5.1: Sample Preparation

Ex-situ cleaved Bi$_2$Se$_3$ crystals from Sigma Aldrich (CAS 12068-69-8, SKU 401080) were used for all of the measurements. Cleavage was performed with Scotch tape. This method has been shown to produce high quality flake samples,$^9$ and can be used in a similar fashion to produce large, flat, shiny measurement areas on the bulk surface. Samples were exposed to air at 21 ± 2 °C
with humidity levels at 45 ± 5 percent. These conditions were held constant during the aging periods of each sample, which ranged from 0 minutes to 10.5 days.

5.2: Measurement Results

This study makes use of a hybrid approach to characterizing the aging process. The goals are to understand the rate at which the surface oxidizes in air, to observe the surface morphology that it creates, and to develop optical models for both the bulk Bi$_2$Se$_3$ and its surface oxide. Multiple experimental methods are necessary to accomplish this. The methodology of our study involves characterizing the bulk of the samples with XRD and HAADF STEM to assess their crystallinity, exposing them to air in a controlled environment for specified time periods, and characterizing the time-dependent oxidation process with a variety of instruments. These include XPS for near-surface chemical analysis, AFM for topography, and both spectroscopic ellipsometry and second harmonic generation for optical properties. This generates a complete characterization of the oxidation process and its physical properties, and is informative about the time scale of oxidation caused by air exposure as well as the electronic properties of the oxide, both of which impact the device development process. As we are looking to create a reference for this aging process, it is important to ensure that our samples are of high quality before they are modified.

5.2.1: Verification of Single Crystal Sample Quality

Before conducting a variety of measurements on Bi$_2$Se$_3$ to test its oxidation dynamics, the crystalline quality of the samples was assessed. Previous microscopy studies of Bi$_2$Se$_3$ and Bi$_2$Te$_3$ show a variety of lattice defects including Bi-Bi layers in the bulk of Bi$_2$Te$_3$\cite{114} and deviations from the ideal lattice (Fig. 3-1) could have significant effects on the oxidation process. The QL
Figure 5-1: Aberration-corrected HAADF STEM images of bulk Bi$_2$Se$_3$. (a) A top-down image of the crystal along the [0001] zone axis, showing its hexagonal structure, along with a simulated atomic model in the upper left, and the power spectrum of the image in the upper right. Though each QL has a honeycomb structure (two lateral sites occupied by both Bi and Se, and one singly occupied only by Se, generating one relatively weak site per hexagonal cell, and thus a honeycomb structure when viewed in the [0001] axis), their rhombohedral stacking leads to a 2D hexagonal close-packed image. A hexagonal unit cell is outlined in red. Differences between the hexagonal shape and the observed image are attributed to sample drift. (b) A cross-sectional image of the crystal. In hexagonal notation, the zone axis is normal to the plane (0 1 0). The QL structure of the crystal is apparent from the vdw gaps separating sections of five atomic layers. One can also see the Se-Bi-Se-Bi-Se order of QLs from the relatively high brightness of the Bi atoms. A QL is outlined in green, with the red dotted extension outlining a unit cell. Each figure has an ideal atomic schematic and a power spectrum in the insets. The lattice parameters measured from these images are $a = 0.429$ nm and $c = 2.958$ nm. Adapted from [115].
structure is integral to understanding the rates of oxidation and the presence of major structural defects may change this time scale, or disrupt the expected equilibrium, drastically. To assess the crystallinity of these samples, local measurements are made with TEM to show pristine QL structure by direct imaging methods, and bulk measurements are made with XRD to show that the expected QL structure found in TEM is present across the sample. Further, XRD pole figures show the lack of twinning or major rotational stacking defects between QLs in the crystals we studied.

HAADF STEM was conducted in cross-sectional and top-down geometries in order to obtain images of the Bi$_2$Se$_3$ crystal structure, as shown in Fig. 5-1. These images were taken with the assistance of Brendan O’Brien, formally of CNSE. The cross-sectional intensity seen in each QL indicates the expected Se-Bi-Se-Bi-Se structure, and each QL is vertically stacked with smooth, vdW-gapped interfaces. Because the $\{01\bar{T}0\}$ normal zone axis is used in this cross section, it does not indicate the presence or absence of distinct twinned grains, but a lack of visible grain boundaries (and Bi bilayers) indicate a high quality single crystal. The STEM image and electron diffraction pattern of the samples used in this study show that these crystals have the expected lattice structure and lattice constant values, as is readily seen by comparison of atoms in the bulk STEM images with a schematic of the ideal bulk structure in the inset. Hexagonal lattice parameters were measured at $a \sim 0.429$ nm and $c \sim 2.958$ nm respectively, which are both within 3% of values found in literature. Aberrations from regular hexagonal shape in the top down images are attributed to drift-induced stretching effects. However, because these are highly localized measurements, bulk crystallinity measurements are also conducted.

XRD characterization also shows a high quality QL structure (Fig. 5-2). XRD $\omega$-2θ measurements yielded strong diffraction peaks at the $(0\ 0\ 0\ 3n)$ indices, and produced a $c$ value of 2.833 nm, which is within 1% of literature values. The diffraction peak intensities follow those
previously reported\textsuperscript{117}. Pole figures were measured in order to search for grains in the samples over a larger volume than possible with STEM. The (0 0 0 3n) peaks were chosen for their high intensity (highest predicted intensity in powder diffraction experiments\textsuperscript{118}) and their three-fold symmetry. They are highly visible and indicative of the presence or absence of twinned grains. This can be seen in the azimuthal symmetry of the pole figure, which becomes 6-fold for samples containing roughly equal amounts of grains with the two predominant orientations being rotated by 60°. This was observed in MBE-grown samples, as discussed in Ch. 6. For flat samples, the plane normals that give the position of diffraction spots have expected tilt angles of \(\theta = \cos^{-1} \left( \frac{\mid hk \mid}{\mid h,k \mid} \right) = \cos^{-1} \frac{s/c}{\left( \frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} \right)^{1/2}} = 58°\). These spots may stray from the expected value for tilted samples such as the bulk growths involved in this study. The acquired pole figure has no 60°-separated twin diffraction, and thus indicates that only one grain orientation is present in these samples (Fig. 5-2).

**Figure 5-2:** Omega-2Theta XRD scan of Bi\textsubscript{2}Se\textsubscript{3}. This measurement was performed on a large granule, of approximately 2x2 cm in area and 0.5 mm in thickness. The (0 0 0 3n) indices are the only c-axis planes that are not forbidden. Thus, we have labeled each peak with its l index. These scans give values of 28.427 Å per unit cell, or 9.476 Å per QL. Adapted from \textsuperscript{115}.
We thus conclude that since local measurements definitively show a vdW-gapped structure, and XRD shows a well-layered structure in the bulk as well as a single azimuthal orientation, these may be considered to represent high quality Bi$_2$Se$_3$ systems, devoid of twinned grains. Since the measurements of the lattice constants of bulk Bi$_2$Se$_3$ produced values close to the literature, and top down and cross-sectional lattice images show no signs of defects, the origin of time dependent changes to this crystal sample of Bi$_2$Se$_3$ presented in the next section are considered to arise from the basic properties of the exfoliated surface of crystalline Bi$_2$Se$_3$. The quintuple layer (QL) structure and weak vdW inter-QL bonds of the Bi$_2$Se$_3$ crystal allow for samples to be prepared via micro-mechanical exfoliation, as has been done with graphene since 2004. Scotch tape exposes a fresh QL as the sample surface, enabling measurements on pristine surfaces, and further measurements as the samples age in air.
5.2.2: AFM Characterization of Time-Dependent Surface Morphology

The time dependence of the surface morphology, structure, and chemical state were studied by AFM, ARXPS, XPS, and RA-SHG. First, the observation of islands on the surface by AFM is presented. AFM analysis was done with the open-source scanning probe microscopy software.

Figure 5-4: Time-dependent surface topography of Bi$_2$Se$_3$. All scans were taken with Bruker TESPA AFM tips. Until 33 minutes after exfoliation, the crystal is atomically smooth. After 33 minutes, surface patches arise, but stop growing at ~100 minutes. They are very uniform in height (3.6 ± 0.2 Å). We hypothesize that these patches are Bi bilayers. The background roughens through the full time scale, increasing from 0.4 Å RMS at 11 minutes to 1.1 Å RMS at 1251 minutes. Adapted from [115].
Gwyddion\textsuperscript{119}. AFM scans completed within 25 minutes after exfoliation show atomically flat surface topographies. Initial surface roughness in the sample shown in Fig. 5-4 is 0.43 Å RMS. Subsequent scans show the nucleation and growth of patches that within measurement error are uniform in height (3.6 ± 0.2 Å). These patches and their growth process are also noticeably different from samples that experience buildup of contaminating particles (Fig. 5-5). Their height above background is consistent with the height of Bi bilayers in the bulk of Bi\textsubscript{2}Te\textsubscript{3}\textsuperscript{114} which is structurally similar to Bi\textsubscript{2}Se\textsubscript{3}. Based on this consideration, we hypothesize that the patches are bilayers of Bi. The existence of Bi bilayers on the surface of Bi\textsubscript{2}Se\textsubscript{3} samples is well documented in literature.\textsuperscript{117,120} This patchy growth provides an explanation for why surface Bi bilayers are observed by some methods and not by others.\textsuperscript{71,120} Some attempts to identify these patches with

\textbf{Figure 5-5:} Reference topography scan of a contaminated surface. Time-resolved AFM images of a clean surface (a), followed by patch growth (b-c), and then dirt adhesion (c-d). These patches, as in other measurements, have a uniform thickness of 3.6 ± 0.2 Å. The bright spots in (d) have thicknesses roughly equivalent to their diameters (>100 nm). Whereas the patches are quite large and flat, the dirt is roughly spherical, and contributes significantly to image aberrations. Adapted from [115].
Auger electron spectroscopy were conducted, but returned inconclusive based on the spatial resolution of the tool. If they are Bi bilayers, these scans are also indicative of a vertical diffusion process, in which Bi rises to the surface. We expect that this diffusion may alter the band structure of Bi$_2$Se$_3$ at the surface, changes the conductance of the topologically protected surface states, and could contribute to 2D electron gas formation. The majority of patch growth occurs before 2 hours. After 2 hours, the sample appears to roughen ubiquitously, and we attribute this change to surface oxidation in agreement with XPS measurements to follow. This roughening process slows over time, leaving the background with RMS roughness of 1.1 Å after 22 hours. It is possible that oxide growth up to a certain length of time after the exfoliation provides a barrier for further surface diffusion, thus distinguishing the growth of Bi bilayers on top of the oxidized surface from the initial oxide growth.

5.2.3: Characterization of Chemical Changes during Oxidation with XPS and ARXPS

XPS and ARXPS provide insight into the growth of the oxide overlayer over a period of ~ 1.5 weeks after exfoliating the samples. Fig. 5 contains angle-integrated, time dependent XPS

<table>
<thead>
<tr>
<th>Sample ID (time since exfoliation)</th>
<th>Binding energy (eV) ± 0.05 eV</th>
<th>Thickness (nm) ± 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bi 5d$_{5/2}$ in</td>
<td>Se 3d$_{5/2}$ in</td>
</tr>
<tr>
<td>t ~ 0 min</td>
<td>25.2</td>
<td>--</td>
</tr>
<tr>
<td>t ~ 119 min</td>
<td>25.2</td>
<td>--</td>
</tr>
<tr>
<td>t ~ 189 min</td>
<td>25.2</td>
<td>26.4</td>
</tr>
<tr>
<td>t ~ 309 min</td>
<td>25.2</td>
<td>26.4</td>
</tr>
<tr>
<td>t ~ 479 min</td>
<td>25.2</td>
<td>26.4</td>
</tr>
<tr>
<td>t ~ 1.5 week</td>
<td>25.2</td>
<td>26.4</td>
</tr>
</tbody>
</table>

Table 5-1: Results of time dependent XPS and ARXPS investigation on Bi$_2$Se$_3$ after exfoliation along (0001). Adapted from [115].
spectra showing aging of the Bi$_2$Se$_3$ surface, and provides information about the constituent elements by means of the Se 3d, Bi 5d, and O 1s level peaks. Table 5-1 lists the values of binding energy of the samples after exfoliation. Spin orbital splitting leads to a pair of peaks for the Se 3d (3d$_{5/2}$ and 3d$_{3/2}$; separated by $\sim$ 0.9 eV) and Bi 5d (5d$_{5/2}$ and 5d$_{3/2}$; separated by $\sim$ 3.0 eV) levels. The freshly exfoliated sample (t = 0 minutes) shows no signs of oxidation as evident by a flat O 1s peak and the absence of higher binding energy peaks in the Se 3d and Bi 5d regions. We estimate the binding energy reference for Bi metal 5d$_{5/2}$ peak as 24.0 ± 0.3 eV and Se chalcogenide 3d$_{5/2}$ as 55.4 ± 0.6 eV after calculating mean and average of all the reported values present at the NIST XPS database.$^{121}$ Thus, for the freshly exfoliated sample (t = 0 minutes), the Bi 5d$_{5/2}$ and Se 3d$_{5/2}$ in Bi$_2$Se$_3$ show a charge shift of +1.2 eV and −1.7 eV respectively which agrees well with the previously reported values for Bi$_2$Se$_3$.$^{122,123}$ The more electronegative Se (2.55 in Pauling scale) pulls a greater portion of the electron cloud towards itself as compared to the less electronegative

![Figure 5-6: XPS and ARXPS results. Angle summed XPS spectra of (a) Se 3d, (b) Bi 5d, (c) O 1s region for various times after exfoliation. The oxide peaks corresponding to SeO$_2$ (marked by a dotted box in (a)) and Bi$_2$O$_3$ (marked by vertical arrows in (b)) are observed after $\sim$119 minutes from exfoliation along with the observation of a strong O 1s peak. The results also indicate that oxygen adhesion and subsurface oxidation proceed simultaneously. Adapted from $^{115}$.](image)

127
Bi (2.02 in Pauling scale), thus giving some ionic nature to the bonds in the Bi$_2$Se$_3$ crystal.\textsuperscript{124} No evidence of oxidation is visible in the XPS spectra of samples up to $t \sim 119$ minutes from exfoliation (Fig. 5-6). AFM measurements show the growth of small patches (3.6 ± 0.2 Å in height, less than 3 μm in diameter) for $t < \sim 119$ minutes, and subsequent roughening of the whole surface. The relative concentration of those patches with respect to the analyzing volume is below the detection limit of our XPS measurement. For samples exposed to ambient air longer than $t \sim 189$ minutes, a sharp O 1s peak appears implying adsorption of oxygen on the (0001) surface of the crystal with time. At the same time, new features show up near the Se 3d (broad peak at \sim 59 eV) and Bi 5d (shoulders towards higher binding energy side) regions at higher binding energy values consistent with oxidized Se and Bi confirming the oxidation of both Bi and Se at the surface as discussed below. Top layer adsorption would produce O 1s peaks with no corresponding oxide peaks in the case that oxygen did not chemically bond to the surface, or O 1s and Se oxide peaks but no Bi oxide peaks in the case that oxygen only bonded to the topmost layer, as indicated by low patch coverage in AFM. The observed values at the higher binding energy region of the Se 3d and Bi 5d (Table 5-1) are close to the values for Bi$_2$O$_3$ and SeO$_2$, the most stable oxides of Bi and Se under ambient conditions. As the O 1s peak show only one binding energy value (531.5 eV), we conclude that the average bonding environment around the O atom stays similar over time and the O atom remains surrounded by both Bi and Se atoms, \textit{i.e.}, the chemical environment in the oxide overlayer is similar to a mixture of Bi$_2$O$_3$ and SeO$_2$. At the beginning of the oxidation process, it is possible that some O-vacancy related defects will remain in the oxide overlayer. Adsorption of more oxygen atoms over time will heal some of these defects and also increase the thickness of the oxide overlayer as more Bi and Se from the QL layer reacts with the atmospheric O. The O 1s, Se 3d oxide, and Bi 5d oxide peaks all continue to grow as evidenced by spectra
collected at subsequent times (Fig. 5-6) although the rate of growth of the oxide overlayer appears to decrease after ~8 hours. This may be due to the vdW gapped structure of Bi$_2$Se$_3$, which could act as a temporary barrier to oxidation beyond the first QL.$^{39}$ We expect that over longer time scales (months), oxidation may continue through further QLs via defect-mediated movement of O in the intercalation layer. The time dependence of the oxidation process is quite unusual, in that it displays a significant delay between surface exposure and oxygen incorporation. It has been reported that a delay to oxidation of up to 20 hours is normal for the (0001) surface of certain

![Figure 5-7: XPS scans of the Bi 4f region taken over various times. Before $t \sim 119$ minutes, no oxidation is observed in the XPS data (see Fig. 5). As such, the presence of the shoulders to the left of the main Bi 4f peaks before $t \sim 119$ minutes are attributed to overlapping Se 3p peaks.$^{19,20}$ For samples exposed to air for longer times, the Bi 4f oxide peaks also show up in the same region as that of the shoulders and grow in intensity over time. Adapted from $^{[115]}$.](image-url)
Bi$_2$Te$_3$ crystals. Further study of the factors relating the surface behavior of Bi$_2$Se$_3$ and Bi$_2$Te$_3$ (e.g., Bi presence, topological phase, QL structure, etc.) is required.

Another research group performed an analysis of the oxidation of Bi$_2$Se$_3$ and found from the presence of a shoulder in the Bi 4f peak that an oxide is present immediately after cleavage in all samples (Fig 2-24). Fig. 5-7 shows the time evolution of the Bi 4f XPS peaks. At $t = 0$ minutes, the main Bi 4f$_{7/2}$ and 4f$_{5/2}$ XPS peaks are accompanied by shoulders to the higher binding energy side. Although the Se 3p 3/2, 1/2 doublet alter the higher binding energy spectrum of the Bi 4f peaks, the effect of oxidation can be observed over time in the continued growth of higher binding energy shoulders. However, since there is no evidence of oxygen adsorption at the (0001) surface for $0 < t < \sim 119$ minutes (see the O 1s spectra in Fig. 5-6(c)), the presence of these Bi 4f shoulders may not be used for interpreting the early presence oxidation of the (0001) surface of Bi$_2$Se$_3$. These shoulder peaks are attributed to potential shake-up peaks and overlap with peaks from Se, as shown in the literature. For samples exposed to the ambient for longer time periods, the shoulder peaks grow in intensity as the contribution from the Bi 4f oxide peaks also appear in the same region. The Bi 5d peak region (Fig. 5-6(b)) is devoid of any such complications arising from peak overlap. Thus, we used the Bi 5d XPS peak region, along with the Se 3d and O 1s regions, to understand the growth mechanism of surface oxide on the Bi$_2$Se$_3$ (0001) surface.

Analysis of angle-resolved XPS data provides a profile of the growth of the oxygen overlayer with time. Among other factors, photoelectron counts are a function of collection angle from normal incidence ($\theta$), and electron origin depth. This is largely due to inelastic scattering events in the sample, which can be approximately modeled with a Beer-Lambert function: $I = I_0 \exp[-\alpha d/\cos(\theta)]$, where $I_0$ is the intensity of emitted electrons, $d$ is the vertical depth from the sample surface to the origin of the emitted electron, and $\alpha$ is a material-dependent density...
parameter. Fig. 5-8 shows the representative plots for Se 3d and Bi 5d regions’ ARXPS data collected for the sample ~1.5 week old. The relative intensities of the oxide signals increase in going from a smaller angle (more sensitive to signals coming from the bulk) to angles close to the surface implying the depth of the oxide overlayer is limited to few QLs during our measurement time. Applying a maximum entropy approach\(^{126,127}\) to the normalized ARXPS data as implemented in the ARProcess software in Thermo Avantage, growth of the oxide overlayer was quantified as shown in Fig. 5-9. Oxide depth appears to increase from 0 nm before 119 minutes to 1 nm at 189 minutes, to ~1.9 nm at the end of 1.5 weeks. This matches previously reported values very closely, which describe the oxidation process of the top-most QL in Bi\(_2\)Se\(_3\).\(^{70}\) It also indicates that the
expansion of a QL due to oxidation increases its volume by ~2x from 0.96 nm in an oxidized QL. This expansion rate is similar to that seen in Si\textsuperscript{128}.

5.2.4: Time-Sensitive Optical Characterization with Ellipsometry

Spectroscopic ellipsometry was used to characterize the changes in freshly exfoliated Bi\textsubscript{2}Se\textsubscript{3} continuously over a 27-hour period. Three repetitions of the experiment, each at different angles, gave consistent results. Standard ellipsometry data Ψ, the relative polarization amplitude change,
and Δ, the relative phase shift of polarizations, are shown in Fig. 5-10(a,b). During the first three hours after exfoliation, Δ initially decreases rapidly followed by a period of slow decrease, while
\( \Psi \) shows little change over time (Fig. 5-10(e)). Changes in \( \Psi \) are small at most energies (wavelengths). The well-known Drude approximation indicates that at an appropriately selected wavelength \( \Delta \) is a linear function of film thickness for thin dielectric films on metals or semiconductors. Sub monolayer changes in \( \Delta \) can easily be measured.\(^{40}\) That is only one possible explanation for the observed changes in \( \Delta \) during the first 3 hours. \( \Delta \) shows significant change at higher energies (smaller wavelengths) over the time frame when surface patch growth is observed by AFM. Because XPS does not observe an oxide layer prior to 3 hours, the changes in \( \Delta \) may be due to the formation and growth of surface patches. Bi\(_2\)Se\(_3\) thicknesses of > 1 mm lead to complete absorption of light above the 0.3 eV bandgap. Thus, bulk Bi\(_2\)Se\(_3\) behaves as an infinitely thick substrate for our entire measurement range of 1.24 to 4.0 eV, and is treated accordingly in our models. Initial measurements after exfoliation are used to extract a dielectric function of bulk Bi\(_2\)Se\(_3\). As can be seen in Fig. 5-10(e), the main peak in the imaginary portion of the dielectric function, \( \varepsilon_2 \), is located at 2.00 eV, and the real portion of the dielectric function, \( \varepsilon_1 \), crosses zero at 2.27 eV. This matches well with FP-LAPW band structure calculations\(^{129}\), for which transitions \( F_{iv} \rightarrow F_{vc}, \quad \bar{v} \rightarrow v, \quad L_{iv} \rightarrow L_{vc}, \) and \( Z_{iv} \rightarrow Z_{vc} \) near 1.96 eV contribute most strongly to the optical response of Bi\(_2\)Se\(_3\), and with previous experiments, in which pseudodielectric functions were calculated from ellipsometric data\(^{130}\). Utilizing these substrate properties and the observed changes in \( \Psi \) and \( \Delta \) over time, we are able to model the changes to the (0001) surface that we observe with our other methods. Because of the variety of growing species and topographical changes to the surface, e.g. sparse unidentified patches, possible 2D electron gas formation, bismuth and selenium oxide growth, surface roughening, as well as the thinness of these new layers (sub-nanometer in most cases), monitoring the individual contributions of these changes is currently beyond the analytical capabilities of our system. Thus, for the first 3 hours the surface ‘oxide’ layer used in
this analysis models the total optical response of a combination of changes in the surface. After 3 hours, the model tracts the observed growth in the oxide layer. It may be possible to track each of source of change individually in the future, but without access to reference data on the exact Se or Bi oxides that grow on Bi₂Se₃’s (0001) surface, or any more information about the identity of surface patches, we have not yet been able to extract meaningful analysis from each affect simultaneously. Thus, we have generated a single dielectric function that most strongly resembles that of previously reported bismuth and selenium oxides¹³¹–¹³³, which we consider to be the major

**Figure 5-11:** Contrast maps for Bi₂Se₃ on a plain Si substrate (a) and on a Si substrate with 300 nm thermal SiO₂ (b). Light regions indicate higher reflectivity of a flake at a given wavelength and flake thickness relative to the reflectivity of the substrate. Likewise, dark regions indicate lower reflectivity of a flake relative to the reflectivity of the substrate. For regions at 0 contrast, there is no difference between the intensity of light reflected from the flake and from the substrate. It is clear that having a 300 nm thermal SiO₂ layer between the substrate and the flake causes higher levels of contrast, especially in the 375 nm and 630 nm regions. Adapted from [¹¹⁵].
contributor to ellipsometric signal generated by our observed surface alterations, after the first 3 hours of ambient exposure. The ‘oxide’ dielectric model is composed of Tauc-Lorenz oscillators, and has a band gap of 2.7 eV (Fig. 8f), which matches previously reported values for Bi oxide.\textsuperscript{131,132} Se oxide may also contribute to $\varepsilon_2$ at energies above its band gap of 2.8 eV\textsuperscript{133}, but we cannot currently parse the contributions of each oxide species.

In addition to providing information about electronic behavior, these dielectric functions (refractive indices) can be used to model reflectivity and the contrast of Bi$_2$Se$_3$ flakes on variety of substrates as seen in an optical microscope with the Fresnel equations at normal incidence:

$$r_{0,1} = \frac{n_0 - n_1}{n_0 + n_1} \quad k = \frac{2}{\lambda} n_k d$$

$$r_{0, k+1} = \frac{n_k r_{k+1} \exp(i 2 n_k d)}{1 + n_k r_{k+1} \exp(-i 2 n_k d)}$$

Eq. 5-1

Here, $r_{0,1}$ represents the electric field reflectance at the interface between layers 0 (taken as ambient) and 1, $n_k$ is layer k’s complex refractive index, and $\beta_k$ gives the absorption of and phase acquired by light of wavelength $\lambda$ propagating through layer $k$ of thickness $d$, and $r_{0, k+1}$ is the field reflectance produced by layers 0 through $k+1$ The reflection intensity is given by $|r|^2$. Fig. 5-11 shows maps of the contrast created by unoxidized flakes on a Si wafer, as well as on a Si wafer with 300 nm of thermal oxide, as a function of flake thickness and incident wavelength. The equation for contrast is $C(l, d) = \frac{|r_{\text{flake}}|^2}{|r_{\text{sub}}|^2}$. For flakes between 0 and 20 nm thick, an increase in thickness causes an increase in contrast with the background. However, since the trend across wavelengths changes minimally with changes in thickness, the change in color of a flake may not be a reliable metric for flake thickness.

5.2.5: Non-Linear Optical Characterization with Second Harmonic Generation

The nonlinear optical technique of second harmonic generation (SHG) is particularly suited for the study of Bi$_2$Se$_3$, because SHG arises from symmetry breaking at the surface, and thus should
detect changes at the surface preferentially over the bulk. The generated second harmonic light of frequency $2\omega$ using the incident fundamental light at frequency $\omega$ is forbidden in centrosymmetric systems such as Bi$_2$Se$_3$. However, at the surface, there is a break in the symmetry due to the transition to ambient, allowing SHG from the crystal surface. Therefore, SHG is a powerful probe for surface-specific properties, such as surface electronic, spin, and lattice structures.\textsuperscript{134,135} In our rotational-anisotropy SHG (RA-SHG) experiment, the sample was rotated azimuthally about the surface normal (0001), and azimuthal angle dependent SHG intensity was measured for p-in/p-out polarization configuration. In Fig. 5-12, the 3-fold symmetric RA-SHG signal appears at the beginning stage after sample cleaving, corresponding to the 3-fold symmetry of the Bi$_2$Se$_3$ crystal, and it develops over time. Initially, there are three major lobes at 60°, 180°, and 300° and three minor lobes at 0°, 120°, and 240°, of sample azimuthal angles. The minor lobes appear in a fresh sample when the outgoing beam is nearly parallel to the (01T5) plane normal, and the major lobes appear in a fresh sample when the incident laser is nearly antiparallel to the (01T5) plane normal, as verified with XRD (Fig. 5-13). As the (01T5) plane normal is parallel to the Bi-Se$_{(2)}$ bond at the top QL surface, alignment between the top Bi-Se$_{(2)}$ bond and the outgoing SHG beam produces minor lobes, and alignment between the top Bi-Se$_{(2)}$ bond and the incident laser produces major lobes. The major lobes decrease in amplitude while the minor lobes increase, eventually becoming equal in amplitude for a 6-fold symmetric RA-SHG scan at ~3.5 hours. Subsequently, the lobes continue to grow at 0° (and other equivalent angles) and shrink at 180°, thus switching their major and minor characters. When the sample has not experienced any prior laser exposure history, this major-to-minor lobe switching process is complete at 12 hours with a majority of the change
occurring by 6 hours. Analysis was done with the following RA-SHG equation for light that was p-polarized before and after reflection:

$$I^2 (\phi) = |a_0 + a_3 \cos(3\phi)|^2$$  \hspace{1cm} \text{Eq. 5-2}

Here $a_0$ and $a_3$ are respectively the isotropic and anisotropic complex coefficients, both determined by Fresnel factors of the beam configuration and electric susceptibility of Bi$_2$Se$_3$, and $\phi$ is the sample azimuthal angle between the incident plane and the mirror symmetry plane, with the directionality defined as in Fig. 5-14. Both coefficients are time-dependent when the surface

---

**Figure 5-12:** Measured rotational-anisotropy SHG (RA-SHG) scans in a dry air environment immediately, and at 1, 2, 3, 4, 12, and 24 hours after Bi$_2$Se$_3$ sample cleaving. Each scan was taken at a different sample spot so that the measurement spot had not experienced any prior laser exposure before each measurement. Thin lines are experimental data, and thick smooth curves are fits to Eq. (2). The empirical fit parameters $a_0$ and $a_3$ are plotted as a function of time. While $a_3$ only undergoes minor changes in the early stages of measurement, $a_0$ increases significantly and monotonically throughout the full oxidation process. Adapted from [115].
is modified, and can be described as linear functions of the second-order susceptibility of Bi$_2$Se$_3$ as follows\textsuperscript{136}:

\[ a_0 = a_{00} + ia_{01} = c_{00} zzz + c_{02} xxx + c_{03} xzx \]

\[ a_5 = c_3 xxx \]

Eq. 5-3

Here $\chi^{zzz}$, $\chi^{zzx}$, $\chi^{xxz}$, $\chi^{xxx}$ are the four non-zero components of the second-order susceptibility tensor $\chi^{(2)}$ of the 3-fold symmetric Bi$_2$Se$_3$ surface,\textsuperscript{137} and $c_{01}$, $c_{02}$, $c_{03}$, $c_3$ are constants that are
determined by Fresnel factors of the beam configuration and dielectric functions of the Bi$_2$Se$_3$. Although both $a_0$ and $a_3$ are complex in general, in our experiment, $a_3$ can be chosen to be real, and $a_0$ requires only a minor imaginary contribution. The significant changes in $a_0$ can be attributed to surface charge accumulation$^{137,138}$ and the trapping of photoinjected electrons at the Bi$_2$Se$_3$/oxide interface, which produces an electric field-induced SHG effect. Figs. 5-12(a-c) show the measured RA-SHG scans in a dry air environment immediately, and at 1, 2, 3, 4, 12, and 24 hours after Bi$_2$Se$_3$ sample cleaving. To avoid the effects of prior laser exposure, the sample was shifted to a different measurement spot before each RA-SHG measurement. The thin lines are experimental data of SHG, and the thick smooth curves are fits to Eq. (2). The $a_0$ term contains a contribution from the surface dipole induced SHG and electric field induced SHG, as described by Eq. (3). The impact of the $a_0$ term changes as the surface oxidizes. Fig. 5-12(d) shows empirical fits to $a_0$ and $a_3$ over time. The major change to $a_3$ to occurs within 1 hour. As other measurement methods detect no oxide in that time, we attribute this change in $a_3$ to be either caused by 2D electron gas formation from band bending$^{137,138}$ or Se vacancy formation. After the first hour, the value of $a_3$ remains near 1.44. The effects of charge accumulation, however, change the isotropic parameter $a_0$ from -0.19 to +0.19 while the first QL is being oxidized. The $a_0$ trend closely resembles the trend of oxide thickness observed in the previous sections of this work. We thus consider the surface dipole SHG contribution, as represented by $a_0$, to be a valid metric for measuring oxide thickness via RA-SHG.

We can reasonably assume that a 6-fold RA-SHG pattern corresponds to the time when the top two layers and part of the middle layer of the top QL is oxidized, and a full major-to-minor lobe switching corresponds to the time when a full QL is oxidized. In comparison to the XPS results
with regard to the time scales of surface oxidation, RA-SHG shows a gradual evolution of oxidation within a QL during the 12 hours period after sample cleaving.

5.3: Discussion

The surface aging process occurs in sequential stages (Fig. 12). First, surface islands of height $3.6 \pm 0.2$ Å form over the first 25 minutes as measured by AFM. The ellipsometry parameter $\Delta$ also changes during the same time period. Since no O 1s peak is observed by XPS, the observations are interpreted as being due to segregation of Bi to the surface over the first 25 minutes. The AFM data shows that patch growth continues for the first two hours, followed by surface roughening. The presence of a surface oxide layer composed of both Bi oxide and Se oxide is clearly present after 189 minutes (~3 hours) in the XPS data. Calculated values of the oxide thickness from both XPS and SE find that the oxide film present at ~3 hours is ~1 nm thick. Using the estimated expansion factor of 2 (as discussed in section 3), the first $\frac{1}{2}$ QL is oxidized at ~3 hours. Here we refer to the oxidation of the top two layers plus part of the middle layer as the top $\frac{1}{2}$ QL. The RA-SHG data show that half of the change in the top QL has occurred by ~3 hours, based on the six-fold symmetric RA-SHG pattern and the associated data from other methods. Thus the oxidation starts soon after 2 hours and continue until the first QL is oxidized at ~1.5 weeks. The RA-SHG for the p-in/p-out polarization combination shows that most of the

![Figure 5-15: Schematic of 2-step surface aging process. (a) Initially, a freshly cleaved sample (0001) surface is atomically smooth. (b) Between 25 and 120 minutes after exfoliation, thin (~3.6 Å) patches nucleate and grow on the crystal surface. (c) After 120 minutes, the surface begins to oxidize and roughen (not shown), obscuring the patches and expanding the surface QL, eventually reaching an oxide thickness of ~1.9 nm. Adapted from [115].](image-url)
oxidation induced change is done after ~12 hours. There is only a minimal change in the RA-SHG data between 12 and 24 hours. The single oxide film model for the SE data also shows a small change in oxide thickness between 12 and 24 hours. The dielectric function generated shows that this oxide layer has a sufficient band gap to electrically insulate TSSs from conduction to layers ostensibly in contact with the Bi$_2$Se$_3$ surface. This may act as a unwanted barrier to transmission in device applications of Bi$_2$Se$_3$ if not accounted for.
Chapter 6: Characterization of Bi$_2$Se$_3$ with a Selenium Protective Layer and after Selenium Layer Removal

The surface of Bi$_2$Se$_3$ is fragile, and may easily be altered by any of the various defects mentioned in Ch. 3. As previously described, defects such as surface oxidation and Se or Bi segregation or desorption alter the surface band structure. In order to utilize a Bi$_2$Se$_3$ in an electronic device, the material must be kept in pristine condition. This may be a particularly difficult task during the fabrication process, where the variety of chemical environments that a material may be exposed to (including ion etching, chemical and physical depositions, polishing, etc.) are capable of inducing defects at the Bi$_2$Se$_3$ surface. Although some evidence supports the viability of surface oxide removal process as well as an electrostatic doping correction, the capacity to repair a Bi$_2$Se$_3$ surface once it is integrated into a device is miniscule. Some promising means of producing these devices involve protecting the Bi$_2$Se$_3$ surface during fabrication with an inert surface layer, or cap. This may be formed in a variety of ways. Some oxides that are familiar to the semiconductor industry, such as ZnO, are inert. These types of dielectric caps may both protect and maintain a pristine crystal, as well as isolate the conduction at the Bi$_2$Se$_3$ surface. This is in contrast to an inert metal deposition, such as gold, which may not interact with the Bi$_2$Se$_3$, but provides conductive states for TSSs to transfer into, thereby losing their helical character. In either of these cases, the cap may not be applied in the same chamber that the Bi$_2$Se$_3$ is grown (given that a dedicated chamber is a likely requirement for high quality Bi$_2$Se$_3$ growths), and there may be no way to selectively remove the cap without damaging the Bi$_2$Se$_3$ surface, as may be required in order to make contact to a device. One solution to these problems is to use bulk Se as a cap. It is trivial to introduce this deposition into a MBE Bi$_2$Se$_3$ growth process, as it is only necessary to curtail the Bi flux and allow Se to accumulate at the surface once the desired Bi$_2$Se$_3$ layer has been
grown. A sufficiently thick Se cap serves as a barrier to surface damage and contamination. Moreover, it is conceivable that this cap may improve the chemistry of Bi$_2$Se$_3$ films that contain Se vacancies, as the cap may act as a quasi-infinite Se source from which Se may diffuse toward vacancies present in the adjacent Bi$_2$Se$_3$ layer. However, the process of removing the cap for the deposition of contacts to the Bi$_2$Se$_3$ has not yet been sufficiently examined. One way that this may be accomplished makes use of the high vapor pressure of Se. At relatively low temperatures (< 150 °C), Se desorbs in a vacuum environment at the rate of > 1 nm/s. Thus, a wide range of annealing temperatures exist at which Se can be removed in bulk (over the course of minutes for sub-micron caps), while the underlying Bi$_2$Se$_3$ phase remains. An investigation of this process, and its impact on the crystallinity, chemistry, and electronic properties of the Bi$_2$Se$_3$ follows.

6.1: Sample Preparation

Se-capped Bi$_2$Se$_3$ samples were grown by Dr. Lee Walsh and Prof. Chris Hinkle’s research group at the University of Texas at Dallas. Bi$_2$Se$_3$ growth was performed in a V80H MBE system that is part of a three-chamber MBE cluster system with each of the growth chambers interconnected with UHV transfer tubes (base pressure = 10-11 mbar). The TMD growth chamber is equipped with Knudsen effusion cells for the evaporation of Bi and Se and in-situ reflection high-energy electron diffraction (RHEED) for characterization of the grown films. The sapphire substrates were sequentially cleaned in acetone, methanol and isopropyl alcohol for 10 minutes each. The sapphire samples were then loaded into the system and annealed at 600 °C for 90 minutes and 750 °C for 10 minutes to degas and complete the surface clean. Before each growth, the Bi and Se sources were outgassed for two hours. The Se:Bi flux ratio was maintained at 20:1 in order to minimize n-doping. The Se and Bi shutters were opened and closed simultaneously to initiate
growth. Growth was performed using a two-step process where a nucleation step was performed at 110 °C and then the Bi shutter was closed while the sample was annealed to 320 °C under a Se flux to improve the crystallinity, the second step was then performed at 320 °C until the desired cap thickness of 200 nm was achieved. The growth rate was determined to be ~0.4 nm/min.

Three rounds of growths were conducted and shipped to Albany for experiments. Samples from the first round are named SBn, where n indicates the sample in the set. This round also includes one sample that was never capped, Bi1, which is used as a reference as no annealing was necessary to reveal its surface. The second set of samples contained severe abnormalities, and is generally not included in analysis. Samples from round three are labeled Rd3_n, where n again indicates the sample in the set. All three rounds of samples were sealed in protective packaging before transportation and arrived in nominally ideal condition.

6.2: Results

The purpose of these experiments is to test whether an optimized annealing process may remove the Se cap, while maintaining an intact, high quality Bi₂Se₃ film. The criteria for assessing the quality of these films are their crystallinity, surface chemistry, and key electronic properties such as the Fermi level. The films must retain a highly crystalline structure, and they must maintain the expected, appropriate properties throughout the annealing process. Therefore, these criteria are measured before and after anneals. The preliminary reference measurements are performed with XPS and SEM to assess the Bi₂Se₃ crystal structure and cap coverage. It is necessary for the usage of Bi₂Se₃ TSSs in devices that the Fermi level resides in the band gap. Thus, the chemical analysis of the film surface is paramount in diagnosing the validity of the annealing decap technique, and the utility of the Se cap in general. To this end, particular care was taken in examining the surface
chemistry and the $E_F$ after anneals. STS electrical analysis is performed in addition to XPS chemical analysis, so that direct measurements of the Fermi levels of these samples could be made.

### 6.2.1: Crystallinity and Chemistry of a Capped Sample

Because the goal of this study is to measure the crystalline and chemical changes in the MBE Bi$_2$Se$_3$ film that occur with Se-decapping anneals, it is necessary to make preliminary structural and chemical measurements that serve as a baseline. A survey consisting of SEM imaging, XRD, and XPS spectra is performed in order to generate a complete picture of the initial samples. Initially, AFM measurements were attempted, as the Se cap was expected to be amorphous, and only rough on the order or $< 10$ nm. However, it was immediately apparent that the surface had much larger topographical features. This motivated SEM imaging, which provided more complete morphological information.

SEM imaging is performed with a Zeiss LEO 1550 at 5 kV accelerating voltage and with the system’s InLens detector. A standard aperture of 30 μm is used, which provides enough signal with minimum sample damage to clearly see surface features. Images are taken normal to the surface and up to 75° from normal incidence in order to allow vertical measurements of the features. All images of the cap clearly show crystallites laying on the surface (Fig. 6-1). In many areas, these are rods, lying flat and bundled together. In areas with well-aligned rods, patches of rods are rotated in 30° intervals, making right angles and triangular structures. These are measured at up to 5 μm in length. There are also areas where crystallites have a less defined
Figure 6-1: SEM images of Se-capped Bi$_2$Se$_3$ films, showing the crystallites on the cap surface. (a) Large domains exist on the surface of Se cap (bright and dark). In some, there are large crystallite formation areas that are oriented in symmetric directions. In others, the orientation of crystallites appears more random. (b) An image taken at 75° from normal shows the heights of crystallites. Some are up to 200 nm.
patches of crystallites with well-defined angular separations. However, they still form bordered structures on the order of 0.5 μm that appear crystalline. In both areas, crystallites of up to 200 nm in height are measured. There are no large areas that are smooth, which indicates that while the cap structure is not as expected, it appears to completely cover and protect the film.

XRD is again performed with a Bruker Bede Metrix-L Diffractometer, with Cu Kα x-rays at 1.5406 Å. Because the cap is less than 1 μm thick, and the Bi$_2$Se$_3$ film is less than 50 nm thick, these measurements are able to acquire diffractions from the Se cap, the Bi$_2$Se$_3$ film, and the Al$_2$O$_3$ substrate. An ω-2θ scan along the surface normal direction shows the (004) peak of the Si carrier wafer, the (006n) peaks of the Al$_2$O$_3$ substrate, the (0 0 0 3n) peaks of the Bi$_2$Se$_3$ film, and some peaks from the crystalline Se cap (Fig. 6-2). Again, the Bi$_2$Se$_3$ intensities are as

![XRD spectrum](image)
expected, showing strong diffractions from the (0006) and (0 0 0 15) peaks. This scan also contains strong evidence of good epitaxial MBE growth, as the \( l \) planes of the hexagonal Al\(_2\)O\(_3\) substrate and Bi\(_2\)Se\(_3\) films are in alignment. Pole figures along the \((01\bar{1} 5)\) planes are also performed in order to assess the polycrystallinity of the samples. As in the case of the oxidation study, these are chosen for their high intensity, as well as for the grain information contained in their 3-fold symmetry, or deviance from it. Three sets of film growths were used in this study, and each produces a different quality pole figure, indicating differing crystallinity between the growth rounds (Fig. 6-3). In all sample sets, the 4-fold symmetric Si (111) peaks from the carrier wafer are also present. In the first round, pole figures are generally 6-fold symmetric, barring tilt effects. In the second round, pole figures were still six-fold symmetric, but the peaks produced less intensity, and the images
were noisier. In the third round, the pole figures are three fold symmetric, with one 3-fold symmetric set of points producing significantly higher intensity than the other. This indicates that there was a dominant grain direction in the third round of samples, but not in the first two. The poor diffraction quality of the second set of samples also indicates that the crystalline structures in that round were not well formed. This data in all measurements of the second round is similarly poor, and complete crystallographic and chemical analyses were not performed on these samples. However, from these XRD measurements, it is concluded that crystallinity seen in rounds 1 and 3 were suitable for study of the effects of thermal decap anneals.

XPS measurements are also conducted prior to annealing. These provide a chemical identity of the capped surface, and may indicate gaps in the cap that could expose the Bi$_2$Se$_3$ film to the surrounding environment. Fig. 6-4 shows the resultant Se 3p, Bi 5d, and O 1s spectra. There is a significant Se 3d peak in these capped samples, as expected. The Se 3d$_{5/2}$ peak is centered at 54.73 eV. This is a 1 eV shift from the peak center previously seen in cleaved Bi$_2$Se$_3$ samples (53.7). This shift is due to the change in Se chemical bonding states. Whereas the Se cap is unary, the Se atoms in Ch. 5 were influenced by surrounding donor Bi atoms, which had the effect of driving electrons toward the Se and lowering its core binding energies. The Bi 5d and O 1s spectra are not present. This indicates that no Bi was near enough to the surface to eject photoelectrons, *i.e.* the film is completely capped. It also indicates that only trace amounts of oxygen may be on the surfaces of these capped samples. This is interesting to note, as it indicates that the Se caps do not oxidize. The Se oxide seen in Bi$_2$Se$_3$ is likely facilitated (or perhaps
enabled) by the larger chemical or structural environment of the Bi$_2$Se$_3$ surface. From these data, we gather that the thermodynamic process of breaking Se-Se bonds to form Se-O in these crystallites is less energetically favorable than the process of breaking Se-Bi bonds to form Se-O at the Bi$_2$Se$_3$ surface, or that the kinetics pose a large, inhibitive barrier to the reaction.

### 6.2.2: Cap Desorption and Thermal Effects on Surface Crystallinity

The first experiment performed on these MBE films was aimed at tracking the surface desorption caused by various annealing temperatures, and subsequently measuring the crystallinity of the surface of the Bi$_2$Se$_3$ film. Annealing and desorption measurements were conducted with TPD, as described in Ch. 4.7. Samples are loaded into the high vacuum pre-pump chamber and left to outgas for a minimum of 12 hours before anneals were performed. Because the system was not capable of measuring atomic masses above 200 amu, the cap desorption and Bi$_2$Se$_3$ desorption events were measured with Se spectra. Both of the major isotopes, $^{78}$Se and $^{80}$Se (which have
expected ratios of 23.77%:49.61%, or 1:2.09) are measured during these processes. Anneals from room temperature ranging between 90 °C and 500 °C showed multiple Se desorption events. Those measured near 130 °C are ascribed to the Se cap desorption, and used for temperature calibration. As discussed in Ch 4.7, while this was the best method found to calibrate TPD anneals, it is not accurate at higher temperatures, and thus only used qualitatively, as an indicator of desorption events. Measurements that contained a second Se desorption peak, resulted in a visibly damaged Bi$_2$Se$_3$ film. These were usually close to 400 °C, which is known to be due to bad temperature measurement based on reports of Bi$_2$Se$_3$ decomposition in literature. This indicates that the second Se peak is due to Bi$_2$Se$_3$ decomposition and desorption (Fig. 6-5). All peaks showed $^{78}$Se:$^{80}$Se ratios that were in agreement with the expected values.

**Figure 6-5:** Decomposition and desorption of the Bi$_2$Se$_3$ film. (a) TPD measurements show a second, minor peak of both Se isotopes at high temperature. This is measured at a much higher temperature than the expected Bi$_2$Se$_3$ desorption temperature of 325 °C, as discussed earlier. The inaccuracy in this early measurement is exacerbated by the thermocouple’s placement near the edge of the sample, a thicker thermocouple wire, and metallic mount washers, which were replaced for later measurements.
Most samples are subjected to TPD anneals between the Se and Bi$_2$Se$_3$ desorption temperatures, as this is the region of interest. Crystallinity is measured with three different surface-sensitive techniques: LEED, AFM, and STM. Not all samples were measured by all techniques, as LEED is the only measurement that could be performed \textit{in situ} with TPD, and time spent out of vacuum is known to affect the Bi$_2$Se$_3$ surface. The time of this exposure was minimized in order to keep surface contamination low. For exposure times on the order of hours, oxidation is likely to take place (see Ch. 4). As oxidation penetrates a full quintuple layer, expanding it to 2 nm thick, the surface crystallinity at the surface may be severely altered if insufficient care is taken during transfer between measurement stations. Transfers to AFM and STM were completed within 10 minutes, during which time the sample was enclosed in a dry plastic case. Many \textit{ex situ} measurements showed no signs of degradation due to transfer.

LEED is performed \textit{in situ} immediately after the decapping anneal. All samples from rounds 1 and 3 heated between the Se and Bi$_2$Se$_3$ temperatures produced LEED patterns that indicated a high degree of surface crystallinity. The acceleration energies of these experiments are set to 30, 72.7, and 96.3 eV in order to see the reciprocal lattice of the surface at different scales (Fig. 6-6). As LEED is only sensitive to the in-plane crystal structure, diffraction spots in these measurements are denoted by the first three indices of Miller hexagonal 4-notation, \textit{e.g.} $(10\bar{1}n) \rightarrow (10\bar{1})$. The 30 eV images show three $\{10\bar{1}\}$ diffraction spots and their $\{\bar{1}01\}$ twins. In the first set of samples, these are approximately equal in intensity. This indicates that there are twinned grains at the surface of near equal density. In the third set of samples (Fig. 6-6(b)), there is visible 3-fold symmetry in the $\{10\bar{1}\}$ diffractions, indicating that the twin domain dominance seen in the preliminary XRD pole figures appears both in the bulk and on the surface. 72.7 eV LEED images capture both the $\pm\{10\bar{1}\}$ and $\pm\{11\bar{2}\}$ families of diffractions, and the 96.3 eV images capture out to
3-fold symmetry can still be seen in these higher energy images, but the intensity difference between twins is smaller. This slight intensity change requires a further quantitative analysis, which is beyond the scope of this work. These LEED measurements show a similar level of crystallinity to that shown in the preliminary XRD measurements, albeit at the
surface level and not throughout the bulk of the film. This demonstrates that there are areas of surface where the crystal lattice remains intact during the decapping anneal, although as LEED is a macroscopic measurement, these patterns do not rule out the presence of defects in the surface structure. As we expect the crystallinity in the bulk of these films to be as good or better than that seen at the surface (which may be subject to atmospheric effects), these LEED experiments allow us to conclude that significant portions of the Bi$_2$Se$_3$ films maintain their crystallinity during the thermal decapping process at all attempted temperatures.

Topographic measurements were performed with a Bruker Dimension Icon AFM (Fig. 6-7). Scans were taken at 1, 5, and 10 $\mu$m$^2$ in order to see the changes in topography induced by annealing over a large area, as well as to see the details of grains. However, STM topographic measurements are limited to a maximum of 1 $\mu$m$^2$, and are better suited for observations of detailed grain structure changes. Samples from round 1 were subjected to AFM measurements after TPD decapping and LEED measurement. These films have pyramidal topographies, composed of QL layered triangular grains, which shrink in area with their vertical stacking (Fig. 6-7). These features are similar to the MBE-grown films seen in literature$^{37}$, and create surfaces with ~20 nm roughnesses. While these films are not flat enough to be suitable for most applications, their > 1 $\mu$m$^2$ grains make them useful for intragranular measurements of chemical alterations, and for observations of the changes in crystallinity of visible features. Sample Bi1 was unique in that it was never capped. This served as a reference for the effect of annealing on grain structures. All round 1 samples were topographically similar, and contained grains of up to ~1.5 $\mu$m$^2$. Decapping anneals did not appear to significantly alter the triangular structure of the grains, although no samples displayed the sharp 60° corners seen in sample Bi1. While LEED indicated little change of the lattice at the surface, AFM shows that the geometry of grains is likely slightly altered by
It should be noted that because all annealing temperatures produced similar topographies, there is some ambiguity in whether the capping or decapping process created these modifications to grain structure. However, as the film growth process was identical between Bi1 and SB5, and Se accumulation is not expected to damage the surface, we conclude that the changes in grain seen in AFM are likely due to thermal annealing.

STM imaging was performed with an Omicron LT STM. Samples Rd3_1, Rd3_5, and Rd3_6 were all annealed in situ prior to STM measurement. These samples were initially annealed at different temperatures, e.g. sample Rd3_1 was annealed at 165 °C, sample Rd3_6 was annealed at 200 °C, etc. Each was also subsequently ramped to higher temperatures at approximately 35 °C intervals until the Bi2Se3 film decomposed, and measurements were taken between each heating step. The Se and Bi2Se3 desorption events (as could be seen by visible changes in the surface) were used to calibrate the anneal temperatures. These measurements were difficult and inconsistent (in
that many images contained too many artifacts to be of use, and image quality was difficult to correlate with sample preparation conditions), but some scans provided topographic information at small scales. No attempts were made at atomic resolution imaging. Most of the scans in this study were 1 \text{um}^2, which is on the order of the size of grains seen in round 1 and 3 samples. Samples subjected to this annealing and measurement process displayed generally triangular pyramidal stacked QL grain structures. Some grains showed a spiral vertical growth pattern (Fig. 6-8(b)), and the corners and edges of these round 3 samples were more irregular than the annealed round 1 samples. Because the first set of samples (which were annealed to different temperatures in TPD) all had straighter edges, and the topography of the 3\textsuperscript{rd} set of samples were generally unaffected by successively hotter anneal temperatures, it is likely that the difference in edge shape seen in these images is due to slightly different growth conditions, and not due to the annealing process. Similar to AFM images, STM topographic scans show that annealing between the Se and Bi\textsubscript{2}Se\textsubscript{3} temperatures had minimal effect on the surface topography of these films, and that these films remain crystalline during the annealing process.

6.2.3: Thermal Effects on the Surface Chemistry and Fermi Level

The surface chemistry and Fermi level of these samples is of critical importance, as it is the determining factor to whether the TSS are overwhelmed by metallic bulk conduction. Some samples that were subjected to TPD and LEED measurements were then taken \textit{ex situ} to the XPS system in order to characterize the surface chemistry of the Bi\textsubscript{2}Se\textsubscript{3} film. Others were annealed \textit{in
situ in the STM and immediately measured with STS to find the Fermi level. The second set of samples produced linear current levels near 0 bias. This behavior is indicative of metallic states, and is not an expected quality in ideal TIs. The first and third set of samples each contained a region near 0 bias where there was very little current, due to the band gap.

XPS was performed under the same conditions as in previous measurements. Spectra were acquired at the Se 3d, Bi 5d, and O 1s peaks so that chemical changes between the capped and decapped samples could be measured (Fig. 6-9). The annealed films demonstrated the expected surface composition. The Bi 5d signals are strong. Because XPS extinction distances are low, this excludes the possibility of any Se cap greater than a few nm. Additionally the Se 3d$_{5/2}$ peak was measured at 53.63 eV, which is within normal error of the Se binding energy in Bi$_2$Se$_3$, and significantly different from the cap Se 3d$_{5/2}$ binding energy of 54.73 eV. The higher energy portion of this spectrum is entirely attributable to the Se 3d$_{3/2}$ peak. There was some oxygen measured in most post-decap XPS experiments. This was to be expected from ex situ measurements, as even

![XPS spectra taken on sample Rd3_2 after TPD decapping. This sample was annealed to 165 °C, and shows a complete decap, as well as some oxidation. (a) The lack of Se cap can be seen from the energy of the Se 3d peak, which has shifted 1 eV to its expected Bi$_2$Se$_3$ chemical state position of 52.63. (b) Because XPS is sensitive to within a few surface nm, the high Bi signal also indicates that there is no Se cap. However, there is some Bi oxidation, as can be seen by the left-skewed Bi 5d doublets. (c) Oxygen is more present at this surface than it was in the capped sample surface. This may originate from the oxidized film or from adsorbed hydrocarbons.](image)
rapid transfers between tools result in air exposure, and once adsorbed, may result in oxidation within the time period that XPS is performed.

The main results of this study are seen in STS measurements. The critical criterion for annealing is that it does not shift the $E_F$ into the bulk or conduction band. STS is used to directly calculate the $E_F$ by measuring shifts in the LDOS at the surface (see Ch. 3.9). This is possible because not only is the DOS function of Bi$_2$Se$_3$ known, the Dirac point provides a clear reference.

STM data is used to calculate how far the Fermi energy has shifted from the Dirac point, $E_F - E_D$. STS measurements are performed on samples from round 3 after in situ anneals. Each sample is heated to an initial temperature, measured, and heated again at a higher temperature for a subsequent STS measurement. This is repeated until the Bi$_2$Se$_3$ film decomposes. In all measurements, the samples are held at ground while the tip was biased. A lock-in amplifier was to perform $dI/dV$ measurements on some of the first set of samples, but equipment issues prevented its further use. Instead, IV measurements are performed and fit with the analytical function,

$$f(x) = e^{b_1(x-x_1)} - e^{b_2(x+x_2)} + Cx + D$$  \hspace{1cm} \text{(Eq. 6-1)}$$
in proximity to inflection point, which is taken to represent the Dirac point, as can be seen in Fig.
Many spectra were too noisy for evaluation due to either the sample quality after Se cap layer desorption or equipment issues, and the spatial distribution of Fermi levels was found to be inconsistent, and thus a statistical approach is taken to assess the doping levels of these films. Over 200,000 IV spectra were analyzed between samples Rd3_1 and Rd3_6. Spectra with a high noise level or were taken when the system was not functioning correctly were not used in the analysis. Selection of the spectra was done using the spectral shape as well as the error of the fit curve. Fermi shifts are extracted from the remaining spectra, as shown in Fig. 6-11.

Figure 6-10: Typical STS results and analytical fitting used to derive $E_F - E_D$. These spectra were measured on sample Rd1_1 after it was annealed to 265 °C. (a) IV spectrum, with data in circles and the fit function in red. (b) Direct derivation of the data in (a), calculated with significant smoothing. As most spectra had more noise, which was amplified by derivation, fitting the IV curves was found to be more reliable. The dI/dV spectrum shows the Dirac point (minimum dI/dV) at higher energy than the Fermi level (0 bias), indicating that this sample is slightly p-doped.
Figure 6-11: Histograms of Fermi level shifts in sample Rd3_1 after annealing temperatures of 150 (a), 180 (b), 210 (c), 240 (d), 265 (e), and 295 °C (f). Each inset shows the number of fit spectra used to construct the histogram, \( n \), the median Fermi level shift, \( E^* \), and the standard deviation of the set of measurements. While no median Fermi level is shifted more than 35 meV from the Dirac point, there was significant variance in these measurements.
resolution of each of these measurements is determined by the Fermi-Dirac distribution. At room temperature, the FWHM of its derivative is $2kT \ln \left(3 + 2\sqrt{2} \right) = 0.092$ eV. Thus, each individual measurement may not be as precise as desired for this experiment, but the acquisition of many ensures the accuracy of the $E_F$ study. The histograms constructed from filtered $E_F - E_D$ show low median shifts (maximum $|E_F - E_D|$ median is 35 meV), but high standard deviations (maximum is 125 meV). The trend of these medians and standard deviations is more easily seen in Fig. 6-12. In sample Rd3_1, the standard deviation of Fermi level shifts is lowest below anneals of 200 °C. For anneals of 210 °C and above, the standard deviations rise above 100 meV. This indicates that lower temperature anneals produce higher spatial uniformity of Fermi levels in these films. Increasing anneal temperatures do not lead to uniform increases in the Fermi level. In both samples Rd3_1 and Rd3_6, the highest median Fermi level occurs between anneals of 210 and 255 °C. These median shifts were 35 and 37 meV, respectively. Hotter anneals appear to correlate with a return to the Dirac point. This effect is not well understood, and will be discussed in the analysis section. Gaps of time between measurements are also significant factors in the Fermi level shifts, as clearly shown in Fig. 6-12(b). After sample Rd3_6 was annealed to 225 °C, it was left in vacuum for more than three weeks while scans were intermittently performed. This dwell time appears to cause a rise in the Fermi level, which is reset by the next anneal. During STS measurements, at no time did the median Fermi level rise or fall to levels near the valence or conduction bands, as the band gap of 300 meV is an order of magnitude larger than measured $E_F$ shifts.
Figure 6-12: Statistical trends of Fermi level shifts over various anneals and time. (a) Sample Rd3_1 was annealed at 150 °C and higher until the Bi₂Se₃ film decomposed. The numbers above each line indicate how many days had passed since the film was decapped. (b) Data from sample Rd3_1 plotted as a function of time rather than temperature. (c) Sample Rd3_6 was annealed from 225 °C until it decomposed. The effects of dwell time are emphasized in this sample.
6.3: Discussion

A variety of experiments were performed on Se-capped Bi$_2$Se$_3$ films to test whether they may remain structurally and chemically intact during a Se decapping anneal process. Many temperatures were tested between the Se desorption temperature and the Bi$_2$Se$_3$ decomposition temperature. Comparison between XRD measurements conducted before decapping and LEED, AFM, and STM conducted after decapping shows that the anneals have little affect on the structure of the film. LEED shows that the lattice is intact at the surface, and AFM and STM topographical scans show that triangular grains and pyramidal structures, as often seen in literature, remain after anneals. Chemical changes at the surface are measured with XPS and STS. The former shows that the Se bonds detected at the surface contain the expected chemical shifts for Bi$_2$Se$_3$. The latter shows a variety of annealing effects on the Fermi levels of the films. Between annealing temperatures of 150 and 255 °C, the Fermi level tends to rise. However, subsequent hotter anneals correlate with a return to the Dirac point. The only proposed $p$-doping mechanism for these films is adsorption of O$_2$. Oxygen may have been adsorbed during measurement or between anneals, as some samples were left exposed to 10$^{-10}$ Torr for as much as three weeks. However, no $p$-doping trends are seen during in situ dwell times. Because there are multiple factors that may potentially affect the electrical environment at the surface of these films, including oxidation and Se evaporation, it is difficult to determine the cause of some changes. However, it may be concluded that thermal annealing is a suitable method to decap these Bi$_2$Se$_3$ films, as they maintain the structure of their growths, and exhibit no significant Fermi level shifting during the process.
Chapter 7: Discussion and Conclusions

This thesis illustrates the potential utility of Bi$_2$Se$_3$ as a TI material, some of the current obstacles barring applications including a detailed study of oxide growth, and shows that a Se cap may be used to protect the Bi$_2$Se$_3$ surface.

The physics that govern the TI class was discussed in detail. This provides the reader information about how the time reversal polarization of $Z_2$ topological insulators may be calculated, and how it acts as a topological invariant. This is central to understanding the emergence of TSS and their protection. After the lengthy explication of TI physics, the emergent phenomena of TSSs and their potential applications were discussed. Finally, a variety of defects in Bi$_2$Se$_3$ seen in the literature were introduced, and an assessment of their potential impacts on surface conduction was presented. The experimental work in this paper is aimed at creating a better understanding of the generation of oxide formation, a common defect, and how it and other defects may be avoided. A chapter regarding the experiments used in this thesis followed. It included information about vacuum and surface science, as well as the theory and practical applications of XRD, spectroscopic ellipsometry, XPS, TEM, AFM TPD, LEED, and STM/STS. These tools were all essential in the hybrid approach taken to studying Bi$_2$Se$_3$ defects.

There were two major experimental studies presented in this thesis. The first investigated the oxidation process at the surface of exfoliated Bi$_2$Se$_3$ samples. This has been the subject of some debate in the academic literature, with some research groups reporting immediate oxidation of all samples upon exposure to air, and others reporting no oxidation up to a time period of years. This study addresses the issue with a hybrid approach, using multiple techniques to assess the dynamics of oxidation at the Bi$_2$Se$_3$ surface and generating multiple optical methods to non-destructively track its formation. The samples used in this study were initially measured with XRD and TEM to
show that they were of high quality. TEM demonstrated a pristine lattice along the (0001) and (10T0) zone axes, showing the hexagonal lattice with expected lattice parameters in plane view and intact and vdW separated QLs in the cross section. As TEM may only be conducted over small areas that do not necessarily represent the macroscopic properties of a sample, XRD was performed to supplement measurements of the bulk crystals. Omega-2theta and pole figures showed the expected vertical separation of layers in each sample, and pole figures showed that all grains were oriented in the same direction. This showed that these samples while there may have been some structural imperfections in these crystals, they were nearly single crystalline and suitable for studies of morphological and chemical changes. AFM measurements showed that the exfoliation technique produced atomically flat surfaces, but when exposed to air, showed the growth of unidentified surface patches and began to roughen over the time scale of hours. XPS performed on samples exposed to air over the same time period showed the nucleation and growth of a surface oxide. Chemical shifts in the Se and Bi spectra show that this layer contains both Bi and Se oxides. Samples that were exposed for a variety of time periods before XPS measurement show that no oxide forms within 2 hours of exposure, but that a surface oxide layer is universally present when crystals are exposed to air for 3 or more hours. Subsequent spectroscopic ellipsometry measurements confirm this time scale. Optical models created from 245 to 1700 nm wavelengths facilitate the measurement of this oxide growth, and provide a means to monitor the surface quality in real time. The dielectric model of this ‘oxide layer’ shows a significant band gap of 2.7 eV. Additional SHG measurements were performed and calibrated with XRD. This experiment showed the impact of surface bond directions oxide formation on SHG. Second harmonic waves are generated away from the surface bond direction when the Bi$_2$Se$_3$ surface is fresh, and are generated toward the surface bond direction when it is oxidized.
The second experimental study in this thesis is aimed at solving the issue of defect state formation at the surface of MBE-grown Bi$_2$Se$_3$ films. The proposed technique is Se surface capping, which may protect the film from physical damage, environmental contamination, and Se evaporation. This may mitigate or eliminate the effects of 2DEG formation, n-doping, and surface oxidation. In order for this solution to be viable for most experiments or device fabrications, the Se cap must be removable with minimal damage or alterations to the underlying Bi$_2$Se$_3$ film. The high vapor pressure of Se ($5 \times 10^{-5}$ Torr at 130 °C) motivates a series of low temperature anneals (150 - 295 °C) to thermally desorb the cap. The criteria laid out to assess the quality of the film are that it maintains its initial crystallinity and that the Fermi level at the surface remains inside the band gap. This is a strict requirement for any implementations of Bi$_2$Se$_3$ that involve measurements or manipulations of the TSSs, as metallic behavior overwhelms the TSS signals when $E_F$ is in the bulk or conduction band. As in the oxidation study, initial measurements were performed to assess the crystallinity of MBE-grown Bi$_2$Se$_3$ films. Multiple growth rounds created differing crystallinities, as assessed with omega-2theta patterns and pole figures, and showed strong 6-fold symmetric diffractions in the first round of samples, poor though still 6-fold symmetric diffractions in the second, and strong 3-fold symmetric diffractions from the third round of samples. Anneals were performed with TPD measurements or in-situ in the STM chamber. Samples annealed with TPD were subjected to LEED measurements that showed surface crystallinity that was highly correlated with pre-anneal XRD measurements. Samples retained their 6-fold or 3-fold symmetries, and the intensities of LEED patterns after annealing matched the intensities of XRD pole figures. Subsequent AFM and STM measurements were performed on annealed samples and compared to a sample that had never been capped or annealed. These displayed similar surface QL pyramid features. Annealed samples maintained triangular surface
geometries, although corners and edges of these grains were more rounded, and in some cases, more irregular than seen in the never-annealed sample. XPS showed the effect of anneals on the surface chemistry of the films. All anneals resulted in ~1 eV binding energy shifts of the Se 3d peak. This indicates that no Se-Se bonds remain on the surface, and that the Se cap is removed to within detection limits. Shifts of the Fermi energy were measured with STS. IV spectra show the LDOS at the sample surface, which is indicative of the Fermi level relative to the Dirac point energy. Anneals of up to 250 °C showed increasing n-doping due to $E_F$ shifts above the Dirac point, with maximum $E_F - E_D$ reaching 38 meV. This was expected from Se vaporization trends. However, anneals beyond 250 °C resulted regressions of $E_F$ back to $E_D$. This behavior is still not well understood, and is left to future directions in the field. Higher anneal temperatures also appeared to cause a larger spatial distribution of $E_F$ shifts, as is visible in the statistical analysis of the STS data. Dwell time also appeared to be a factor in Fermi level shifts. A sample left in UHV for three weeks showed a steady increase in the Fermi level over time. This effect was negated by the next anneal. It is possible that this trend was due to surface contamination in the chamber that was removed by subsequent heating.

These studies show that Bi$_2$Se$_3$, while delicate, may be placed in optimized conditions that allow its reliable usage in devices and experiments. The surface oxidation study presented rectifies some disagreements in literature about the time scale of oxidation. A previous study that showed immediate oxidation incorrectly attributed this to the presence of a Se 3p peak in XPS measurements that occurs at the same energy as the Bi 4f oxide peak. While immediate oxidation is not the cause of degradation in TSS signals, exposure to air reliably causes oxidation to occur in Bi$_2$Se$_3$, even at atomically flat surfaces. The surface patches in AFM may also serve to bridge a gap in literature regarding the existence of Bi bilayers in these systems. While unconfirmed, we
propose that these may be caused by surface Bi segregation. They are of uniform thickness similar to previously reported Bi bilayers, they appear immobile during measurement, they display a nucleation, growth, and stagnation process that appears less random than adsorption of surface contaminants, and differ categorically from other scans showing significant surface contamination. Additionally, ellipsometric and SHG methods developed to monitor the oxidation process may be used in the future to help understand the mechanisms of oxidation, and could be used to help parse the different defects that emerge at the Bi$_2$Se$_3$ surface. While ‘oxide’ layer of the ellipsometric model may actually be a representation of multiple surface modifications at some time scales, the insulating properties it exhibits pose an additional challenge to fabricating a Bi$_2$Se$_3$ devices, as making electrical contact through this oxide layer may be a non-trivial task. SHG could also be used as a method of non-destructively determining the orientation of Bi$_2$Se$_3$ crystals during the growth process to supplement or supplant RHEED. The analysis of Se decapping anneals shows that this method can be used to protect the surface of Bi$_2$Se$_3$. This may be paramount to fabrication procedures, where the surrounding environment may differ significantly from MBE growth chamber conditions. As the oxidation study discusses and the literature reports, these surfaces are highly fragile in the face of oxidating species, and may be detrimentally doped by exposure to other elements in fabrication chambers such as metals. The Se capping procedure allows the Bi$_2$Se$_3$ film to remain in good condition until surface exposure is desired, and the thermal decapping procedure shows minimal effects on the quality of the Bi$_2$Se$_3$ films. Ideal decapping temperatures were found under 200 °C, where the Se cap was completely removed, high surface crystallinity of the films was confirmed, and $E_F$ was measured very near $E_D$. Excessive annealing temperatures result in some $n$-doping and higher degrees of spatial $E_F$ variance on the surface, although the median $E_F$ was never shown to shift to the valence or conduction band for any attempted anneal.
Prior to such alterations, the Bi$_2$Se$_3$ samples decomposed. We can therefore conclude that this is a promising method to help the implementation of Bi$_2$Se$_3$ in devices.

Further experiments are needed to understand the behavior of the Bi$_2$Se$_3$ surface. We were unable to definitively answer questions regarding Bi bilayer formations or what the microscopic behavior is that causes anneals above 250 °C to show $E_{\text{db}} < E_F$. Dynamic SIMS is a likely technique to help address these issues, as it is precise to orders of magnitude lower than the surface coverage of observed patches requires. As the only tool available during this thesis work utilizes a Bi beam, such experiments were not possible. Additionally, SIMS may help determine the Se concentration trends for Bi$_2$Se$_3$ samples under various thermodynamic conditions. For known Se concentration and $E_F$ levels, the ionization energy of Bi donors near Se vacancies may be assessed, and used to help determine what surface doping levels are purely attributed to Se evaporation. Many characterization steps are also necessary for the fabrication process. Interactions between various metal contacts have been assessed, and the basic properties of TSSs are well known, but few devices have been fabricated. Decoupling the surfaces is a necessary step toward spin filtering. This may involve multiple vertical contact layers to each terminal of the sample so that the net spin conduction of the top and bottom surfaces do not cancel. This may also be accomplishable by doping the surfaces to significantly different levels, as net spin conduction of any completely filled surface Dirac cone is zero, due to the opposite helicities of the top and bottom halves of each cone. It may also be useful to decap parts of these films while leaving other parts unaffected. This may be accomplishable with precise, targeted heating methods such as rapid laser annealing. Implementations of Bi$_2$Se$_3$ for more exotic purposes such as quantum computing may soon be possible, and outside the realm of TI, new topological systems such as Weyl and Dirac semimetals.
are recent subjects of academic interest. Each prospect will require rigorous scientific study and due diligence, but may result in a dynamic new regime of solid state physics and its applications.
Chapter 8: Appendix

8.1: Quantum Hall Calculations from Linear Response Theory

The Kubo formula used by Thouless et al. to calculate the quantum Hall effect can be derived from linear response theory. This derivation involves calculating the linear response of a system (given by the Hamiltonian) subject to a time-dependent perturbation. In order to do so, it is common to use the interaction picture, which is distinct from the Schrodinger and Heisenberg pictures in that both wavefunctions and operators are functions of space and time.

Take a base 2D Hamiltonian and eigenstates \( H_0 |n\rangle = E_n |n\rangle \), where \( n = 0 \) denotes the ground state. To get the conductivity of this system, we perturb the system with an electric field, \( \vec{E}(t) = \vec{E} e^{-i\omega t} \), which gives the vector potential \( \vec{A}(t) = \frac{\vec{f}}{m_e c} e^{-i\omega t} \) (seen in Maxwell’s equation \( \vec{E} = -\partial_t \vec{A} - \vec{\nabla} \phi \) for \( \omega = 0 \)). Later, we’ll take the limit \( \omega \to 0 \), but it’s easier to have an oscillator for now. The addition of the electric field brings the Hamiltonian to \( H(t) = H_0 + H(t) \), where \( \Delta H(t) = -L_x L_y \dot{\vec{A}}(t) \), and \( \dot{\vec{J}} = \frac{\vec{p}}{m_e c} \) is the current density operator. For now, the goal is to calculate \( \langle J \rangle \). In the interaction picture, operators evolve via a unitary change of basis

\[
\hat{O}(t) = V(t - t_0) \hat{O}(t_0) V(t - t_0) \text{, where } V(t - t_0) = e^{-i\phi(t-t_0)} \text{, and states evolve as } (t \ t_0) = U(t \ t_0) \ (t_0) \text{, where } U(t) = T \exp\left(\frac{i}{\hbar} \int_{t_0}^{t} \Delta H(t') dt' \right) = T \exp\left(\frac{i}{\hbar} \int_{t_0}^{t} \Delta H(t') dt' \right) \text{ (where } T \text{ stands for time ordering). Thus, the current density at time } t \text{ of state } \hat{O} \text{ is:}
\]

\[
\langle \dot{J}(t) \rangle = \langle 0(t) | \hat{J}(t) | 0(t) \rangle
\]

\[
\langle J(t) \rangle = \langle 0(t_0) | U^* (t - t_0) \hat{J}(t) U(t - t_0) | 0(t_0) \rangle
\]

\[
\langle \dot{J}(t) \rangle = \langle 0(t_0) | T^* \exp\left(\frac{i}{\hbar} \int_{t_0}^{t} \Delta H(t') dt' \right) \hat{J}(t) T \exp\left(\frac{i}{\hbar} \int_{t_0}^{t} \Delta H(t') dt' \right) | 0(t_0) \rangle
\]
\begin{align*}
\langle \tilde{J}(t) \rangle &= \langle 0(t_0) \rangle \sum_{k=0}^{\infty} \left( \frac{1}{k!} \int_{t_0}^{t} \Delta H(t') \, dt' \right)^k \tilde{J}(0) \sum_{k=0}^{\infty} \left( \frac{1}{k!} \int_{t_0}^{t} \Delta H(t') \, dt' \right)^k \langle 0(t_0) \rangle \\
&= \langle 0(t_0) \rangle \left( 1 + \frac{1}{2} \int_{t_0}^{t} \Delta H(t') \, dt' \right) \tilde{J}(t) \left( 1 - \frac{1}{2} \int_{t_0}^{t} \Delta H(t') \, dt' \right) \langle 0(t_0) \rangle
\end{align*}

Taking only up to the linear terms of the matrix exponential leaves
\begin{align*}
\langle \tilde{J}(t) \rangle &= \langle 0(t_0) \rangle \left( 1 + \frac{1}{2} \int_{t_0}^{t} \Delta H(t') \, dt' \right) \tilde{J}(t) \left( 1 - \frac{1}{2} \int_{t_0}^{t} \Delta H(t') \, dt' \right) \langle 0(t_0) \rangle
\end{align*}

Taking only up to the linear terms of the perturbation leaves
\begin{align*}
\langle \tilde{J}(t) \rangle &= \langle 0(t_0) \rangle \tilde{J}(t) + \frac{1}{2} \int_{t_0}^{t} [\tilde{J}(t'), \tilde{A}(t')] \, dt' \langle 0(t_0) \rangle
\end{align*}

The first term is ignored as it represents the current in the absence of an electric field. Plugging in for the perturbation, we get:
\begin{align*}
\langle \tilde{J}(t) \rangle &= \langle 0(t_0) \rangle \tilde{J}(t) + \frac{1}{2} \int_{t_0}^{t} \left[ \tilde{J}(t'), \tilde{A}(t'), \tilde{J}(t) \right] \, dt' \langle 0(t_0) \rangle
\end{align*}

For clarity from here, we’ll take $L_x L_y$ to be the total sample area, and use Einstein summation notation over perturbation components $j$.
\begin{align*}
\langle \tilde{J}(t) \rangle &= \frac{A L_x L_y}{
\int_{t_0}^{t} \langle 0(t_0) \rangle \left[ \tilde{J}(t'), \tilde{A}(t), \tilde{J}(t) \right] \, dt' \langle 0(t_0) \rangle
\end{align*}

We may use $u$-substitution $t' = t - t'$. We also take $t_0 \to \infty$, and for simplicity, drop $t_0$ from $\langle 0 \rangle$.
\begin{align*}
\langle \tilde{J}(t) \rangle &= \frac{A L_x L_y}{\int_{t_0}^{t} \langle 0 \rangle \left[ \tilde{J}(t - t'), \tilde{A}(t), \tilde{J}(t) \right] \, dt'}
\end{align*}

To further simplify, we use $\hat{O}(t) = V^{-1}(t) \hat{O}(0) V(t)$:
\begin{align*}
\langle 0 \left[ \tilde{J}(t), \tilde{J}(t) \right] \rangle &= \langle 0 \left[ V^{-1}(t) \tilde{A}(0) V(t), V^{-1}(t) \tilde{J}(t'), \tilde{J}(t) V(t) \right] \rangle \\
&= \langle 0 \left[ V^{-1}(t) \tilde{J}(0) V(t), V^{-1}(t) \tilde{J}(0) V(t) \right] \rangle \\
&= \langle 0 \left[ V^{-1}(t) \tilde{J}(0) V(t), V^{-1}(t) \tilde{J}(0) V(t) \right] \rangle \\
&= \langle 0 \left[ e^{-\frac{A(t'{t_0})}{k}} \tilde{J}(0) V(t), V^{-1}(t) \tilde{J}(0) e^{-\frac{A(t'{t_0})}{k}} \right] \rangle \\
&= \langle 0 \left[ e^{-\frac{A(t'{t_0})}{k}} \tilde{J}(0), V(t) \right] \rangle e^{-\frac{A(t'{t_0})}{k}}
\end{align*}
Plugging back into get the current density:

\[
\langle \hat{J}_i(t) \rangle = \frac{N_{i}}{\hbar \omega} e^{-i\omega t} \int_{-\infty}^{0} \langle 0 | [\hat{J}_i(0), \hat{J}_i(t')] | 0 \rangle e^{i\omega t'} dt'
\]

Since the only \( t \)-dependence of the expected current density comes from \( e^{i\omega t} \), the system responds to an applied electric field at frequency \( \omega \) with current at the same frequency. This is the linear response that we’re interested in. To get the Hall conductivity from this point is trivial:

\[
\sigma_{xy}(\omega) = \frac{\Delta}{\hbar} \int_{0}^{\infty} e^{-\omega \tau} \sum_{n} \langle 0 | \hat{J}_y(0) | n \rangle \langle n | \hat{J}_x(t) | 0 \rangle - \langle 0 | \hat{J}_x(t) | n \rangle \langle n | \hat{J}_y(0) | 0 \rangle dt
\]

This is the Kubo formula applied to conductivity. There is a more general Kubo formula that expresses the reaction of any system in linear response to a time-dependent perturbation, but that is beyond the scope of this section. To see the topological implications of this formula, we further modify by expanding the commutator and multiplying by \( |n\rangle \langle n| \):

\[
\sigma_{xy}(\omega) = \frac{\Delta}{\hbar} \int_{0}^{\infty} e^{-\omega \tau} \sum_{n} \langle 0 | \hat{J}_y(0) | n \rangle \langle n | \hat{J}_x(t) | 0 \rangle e^{i\omega_0 \tau} - \langle 0 | \hat{J}_x(t) | n \rangle \langle n | \hat{J}_y(0) | 0 \rangle e^{i\omega_0 \tau} dt
\]

The next step is to plug in the evolution of \( \hat{J}_i(t) \). And again, \( \langle 0 | \hat{J}_i(t) | 0 \rangle = 0 \), so we may take the sum for \( n \neq 0 \), and integrate.

\[
\sigma_{xy}(\omega) = \frac{\Delta}{\hbar} \int_{0}^{\infty} e^{-\omega \tau} \sum_{n \neq 0} \langle 0 | \hat{J}_y(0) | n \rangle \langle n | \hat{J}_x(t) | 0 \rangle e^{i\omega_0 \tau} - \langle 0 | \hat{J}_x(t) | n \rangle \langle n | \hat{J}_y(0) | 0 \rangle e^{i\omega_0 \tau} dt
\]

\[
= \frac{\Delta}{\hbar} \int_{0}^{\infty} \sum_{n \neq 0} \langle 0 | \hat{J}_y(0) | n \rangle \langle n | \hat{J}_x(t) | 0 \rangle e^{i\omega_0 \tau} - \langle 0 | \hat{J}_x(t) | n \rangle \langle n | \hat{J}_y(0) | 0 \rangle e^{i\omega_0 \tau} dt
\]

\[
= \frac{\Delta}{\hbar} \sum_{n \neq 0} \frac{\langle 0 | \hat{J}_y(0) | n \rangle \langle n | \hat{J}_x(t) | 0 \rangle - \langle 0 | \hat{J}_x(t) | n \rangle \langle n | \hat{J}_y(0) | 0 \rangle}{\hbar \omega + (E_n - E_0) - \hbar \omega - (E_n - E_0)}
\]

From here, we expand the fractions in \( \hbar \omega \) as \( \frac{1}{\hbar \omega + (E_n - E_0)} = \frac{1}{(E_n - E_0)} - \frac{\hbar \omega}{(E_n - E_0)^2} + ... \). The first term vanishes, and we can plug in \( \omega = 0 \) to get the response for a DC field:
\[ \sigma_n = i\hbar \sum_{n \neq 0} \frac{\langle 0 | \hat{j}_n | n \rangle \langle n | \hat{j}_s | 0 \rangle - \langle 0 | \hat{j}_s | n \rangle \langle n | \hat{j}_s | 0 \rangle}{(E_n - E_0)^2} \]

8.2: Derivation of the Fresnel Equations

In optical experiments, the reflection plane is defined by the propagation vectors \((\mathbf{k}_i, \mathbf{k}_r)\) of the incident and reflected waves, which act as basis vectors for the plane, barring a normal reflection geometry. Light polarization is given by electric field components of the wave that are aligned in the reflection plane (\(p\) polarized, for parallel) and perpendicular to the reflection plane (\(s\) polarization, from the German ‘senkrecht’ for perpendicular). Given any sample, one can make a classical approximation wherein an incident polarized wave traveling in the \(\hat{x}\) \(\hat{z}\) reflection plane meets a locally flat surface (with surface normal \(\hat{z}\)), as given by Fig. 8-1. Because fields generated by the electrically polarized surface are oriented vertically, all EM fields in the plane are conserved:

\[
\begin{align*}
E_{ix} + E_{rx} &= E_{tx} \\
H_{ix} + H_{rx} &= H_{tx} \\
E_{iy} + E_{ry} &= E_{ty} \\
H_{iy} + H_{ry} &= H_{ty}
\end{align*}
\]  
(Eq. 8-1)

Figure 8-1: Schematic of an EM wave incident upon an interface between media. Amplitudes \(E_{ip}, E_{is}, E_{ip}, \) and \(E_{is}\), as well as their accompanying phase considerations, are all functions of the incident complex wave given by \(E_{ip}\) \(E_{is}\), as well as the angle of incidence \(\varphi_1\) and the refractive indices of both media, \(N_1\) and \(N_2\). These relations are given by the Fresnel equations, Eqs. 8-4 and 8-5. Figure adapted from [24].
Note that in the given geometry, $s$-polarization coincides with the $y$-axis, and $p$-polarization is in the $\hat{x}$ $\hat{z}$ plane. These equations constitute the basis from which Fresnel coefficients, which give the complex reflection and transmission of an interface, are derived. The Fresnel coefficients are functions of angle of incidence and the complex refractive indices of the interface materials. Thus, in order to derive the Fresnel coefficients, we must incorporate equations containing information about the angle of incidence and complex refractive indices of at the interface into the continuity equations.

Toward deriving reflection and transmission coefficients for $s$-polarized light, we make use of the electric field equation in the $\hat{y}$-direction, and create another from the magnetic continuity in the $\hat{x}$-direction. This is done through the relationship between the electric field, magnetic field, and propagation vector, $\vec{E} = \vec{k} \times \vec{H} / (\varepsilon_0 \mu_0)$ (where $\varepsilon = (\varepsilon_0 \mu_0)$ is the complex dielectric function of the media that the light is traveling through). Further, utilizing the definition of the complex refractive index, $k = -\varepsilon N$ (where $N = N(\varepsilon)$ is the complex refractive index), one can make a complete substitution of the desired variables to produce $N_1 \cos j_1 (E_{is} - E_{rs}) = N_2 \cos j_2 E_{ts}$. Combining this with the initial $\hat{y}$-direction electric field equation gives the Fresnel coefficients for $s$-polarized light:

$$r_s = \frac{N_1 \cos j_1 - N_2 \cos j_2}{N_1 \cos j_1 + N_2 \cos j_2} \quad \text{and} \quad t_s = \frac{2N_1 \cos j_1}{N_1 \cos j_1 + N_2 \cos j_2}. \quad \text{(Eq. 8-2)}$$

These can be further reduced with Snell’s law to remove the transmission angle:

$$r_s = \frac{N_1 \cos j_1 + \sqrt{N_1^2 - N_2^2 \sin^2 j_1}}{N_1 \cos j_1 + \sqrt{N_1^2 - N_2^2 \sin^2 j_1}} \quad \text{(Eq. 8-3)}$$

Fresnel coefficients for $p$-polarized light are calculated with the remaining continuity equations for electric field in the $\hat{x}$-direction and magnetic field in the $\hat{y}$-direction, as well as the previously
given electric/magnetic relation and definition of the refractive index. The $p$-polarized Fresnel coefficients, after Snell’s law substitution, are:

$$
R_p = \frac{N_1 \sqrt{N_2^2 - (N_1 \sin \theta_1)^2} \cos \theta_1}{N_1 \sqrt{N_2^2 + (N_1 \sin \theta_1)^2} \cos \theta_1} \quad \text{(Eq. 8-4)}
$$

$$
T_p = \frac{2N_1 \cos \theta_1}{N_1 \sqrt{N_2^2 - (N_1 \sin \theta_1)^2} \cos \theta_1} \quad \text{(Eq. 8-5)}
$$

Because the Fresnel coefficients are complex, they completely calculate the change of amplitude and phase of $s$- and $p$-polarized light at an interface. Thus, barring cross-polarization, which is not expected to occur in optically isotropic samples, they provide complete information for ellipsometric measurements.
References

21. (n.d.).
34 M.Z. Hasan and C.L. Kane, Rev. Mod. Phys. 82, 3045 (2010).
133 K. Tripathi, M. Husain, and M. Zulfiquar, Chalcogenide Lett. 6, 517 (2009).
136 Y.Q. An, A.J. Green, and Diebold, A.C., (n.d.).