ATOMIC SCALE ENGINEERING OF TOPOLOGICAL MATERIALS

By

MATTHEW J. BRAHLEK

A dissertation submitted to the

Graduate School–New Brunswick of

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Physics & Astronomy

Written under the direction of

Seongshik Oh

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE DISSERTATION

Atomic Scale Engineering of Topological Materials

By MATTHEW J. BRAHLEK

Dissertation Director:

Seongshik Oh

This dissertation seeks to understand the novel physical properties of a new class of materials called topological insulators (TIs) from an experimental standpoint, and, in particular, is concerned with electrical transport properties of atomic scale thin films structures of the TI Bi_2Se_3 grown by the molecular beam epitaxy technique. We first begin by reviewing the nascent field of TIs, which will act to motivate the work. This will be followed by a thorough introduction to the many aspects concerning the synthesis of thin film TIs by the atomically precise technique of molecular beam epitaxy. This is then followed by a brief introduction to, and the details of electrical transport, which is the main experimental technique used in this work. In the remaining chapters we apply these techniques to understand the fundamental properties of this new class of materials.

As a point of necessity, we first discuss the effects that exposure to atmospheric gases have on the transport properties of Bi_2Se_3. Following this, we examine the properties of $(Bi_{1-x}In_x)_2Se_3$, which transitions from a topological insulator ($x = 0$) into a topologically trivial band insulator ($x = 1$) through three quantum critical points. We then detail two studies that probe the intrinsic properties of the topological surface states. The first study deals with the emergence of

decoupled topological surface states when the metallic bulk state is suppressed. The second study presents a new phase called an inverse TI, whose electronic properties can be engineered chemically, which is found to allow control over the coupling between the topological surface states. This will lead into the final work, which is concerned with understanding the global properties of transport measurements on TIs.

Acknowledgements

From the easy times, to the frustrating points, the people around me are what have made my time in graduate school a wonderful experience. This includes my family, my fellow students, with whom, the inevitable struggles never seems too bad, and the many post-docs and professors who were always willing to take the time to share their knowledge of physics, or offer some valuable career or life advice.

I would like to first thank my advisor, Professor Seongshik Oh, for his limitless patience, friendship, and guidance. I am incredibly lucky to have had the chance to learn the art of science from someone with his passion and deep experimental instinct. I cannot imagine an advisor who has more respect for, and trust in his students. Seeing me through as his first graduate student will remain as one of the high points in my life.

I would like to thank my committee members Professors Vitaly Podzorov, Sevil Salur, and David Vanderbilt for their time, and insight, both of which, I have greatly benefited from.

I would like to specifically thank Professor Peter Armitage whose generosity with his time and advice has had a profound impact on my career. I would also like to thank Prof. Armitage and his student Liang Wu for their prolific scientific collaboration, through which, we have together gained deep insight into the fundamental physics of topological insulators.

I would like to thank my lab mate and friend Eliav Edrey for always keeping a sense of normalcy for me, and all of our fellow students, in the tumultuous early years. I am indebted to Yong-Seung Kim who patiently mentored me in my first year in the lab and was always happy to answer even the most naive question. I owe a tremendous amount to my lab mate and friend Nikesh Koirala, who has spent the last couple of years sharing half of the growth load and many long, tedious hours in the lab. Without his help, I could not have accomplished half that I have. I also want to thank my lab mates Namrata Bansal and Jisoo Moon. Finally, I would also like to

thank my lab mate and friend Maryam Salehi, with whom it has been such a joy working and learning together in the lab.

I would like to thank my classmates and friends Brett Manning, Sebastian Reyes-Lillo, and Tahir Yusufaly whose generosity and friendship has always been a joy.

I would like to thank some of my colleagues from neighboring labs, particularly Matthew Bell, Jixia Dai, Jianpeng Liu and Joshua Paramanandam, whose discussions and advice were always rewarding. I would also like to thank the following professors who were always generous in scientific collaboration, advice and often much needed equipment and parts: Robert Bartynski, Girsh Blumberg, Misha Gershensen, Vitaly Podzorov, David Vanderbilt, and Weida Wu. I would also like to thank Bill Schneider, Eric Paduch, and Ernie Erskine from the machine shop for their guidance, knowledge, and effort, which ultimately made this work possible.

Lastly, I am deeply indebted to my family. I want to especially thank my beautiful wife Alexandrea, who has so graciously tolerated my pursuit of physics, and never complained about my early mornings, late nights, or long weekends in the lab. To my Dad and Mom, whose encouragement early on set me on the path that I am today. To my Dad and Grandfathers, whose mastery of the world left me awestruck from an early age. To my brother Steven and sister Amanda whose exuberance in the shenanigans of our childhood shaped my desire to build and understand the physical world. Lastly, I want thank my little angel Amelia, who makes all this hard work worth it.

Dedication

Without my Dad and Mom, Brother and Sister, and especially my wife Alex and my wonderful daughter Amelia, I would not have had the focus and vigor to complete this work. Therefore this work is entirely dedicated to my family.

Table of contents

Abstract ii

Acknowledgements iv

Dedication vi

List of Tables xi

List of Figures xii

CHAPTER 1. PROLOGUE 1

1.1 Classifying insulators: band-insulators and the quantum hall state 3
- 1.1.1 Band theory of crystals 4
- 1.1.2 Topology: doughnuts, soccer balls and the quantum hall state 6

1.2 Topological insulators 9

1.3 Topological surface states compared to non-topological surface states 14

1.4 Experimental realization 18
- 1.4.1 2D TIs 19
- 1.4.2 Angle resolved photo-emission spectroscopy 19
- 1.4.3 3D TIs 21

1.5 New topological materials 23
- 1.5.1 Topological crystalline insulators 23
- 1.5.2 Topological Kondo insulators 24
- 1.5.3 3D Dirac materials 25

1.6 Scope of this work 26

CHAPTER 2. MOLECULAR BEAM EPITAXY: GROWTH OF TOPOLOGICAL MATERIALS 28

2.1 An abbreviated history of thin film deposition 30

2.2 Growth processes 36
- 2.2.1 Thermodynamics and kinetics 36
- 2.2.2 Physical and chemical adsorption 38
- 2.2.3 Accumulation and growth modes 40
- 2.2.4 Defects structures 43

2.3 Thin film growth techniques 45
- 2.3.1 Sputtering 46
- 2.3.2 Pulsed laser deposition 49
- 2.3.3 Molecular beam epitaxy 51

2.4 Molecular beam epitaxy at Rutgers University 53
2.4.1 The MBE system 53
2.4.2 Substrate handling 56
2.4.3 Thermal cell operation 59

2.5 Ultra-high vacuum 61
2.5.1 Necessity for ultra-high vacuum conditions 61
2.5.2 Achieving ultra-high vacuum 63

2.6 Atomic flux measurement 66
2.6.1 In situ quartz crystal microbalance 67
2.6.2 Rutherford backscattering spectroscopy 70

2.7 Structural characterization 77
2.7.1 Reflection high energy electron diffraction 77
2.7.2 X-ray diffraction 89
2.7.3 Atomic force microscopy 90

2.8 Van der Waals epitaxy of topological materials: A_2Se_3 chalcogenide family 93
2.8.1 Substrate consideration 94
2.8.2 Bi_2Se_3 on Si(111): the basic Bi_2Se_3 recipe 97
2.8.3 Bi_2Se_3 on Al_2O_3 100
2.8.4 Bi_2Se_3 on amorphous and non-symmetry matched substrates 103
2.8.5 In_2Se_3 on Al_2O_3 104
2.8.6 $(Bi_{1-x}In_x)_2Se_3$: film growth and flux control 104
2.8.7 Bi_2Se_3 on In_2Se_3 buffer layers 107

2.9 Conclusion 111

CHAPTER 3. ELECTRICAL TRANSPORT: THEORY AND PRACTICE 112

3.1 Semiclassical transport theory 113

3.2 Hall effect: Determination of the charge carrier density 118

3.3 Transport measurement: Practice 119
3.3.1 Van der Pauw and Hall bar geometries 119
3.3.2 Reverse field reciprocity 123
3.3.3 Symmetrization and data analysis 124
3.3.4 Sample preparation and instrumentation 127
3.3.5 Cryogenics 132
3.3.6 Magnet systems 135

CHAPTER 4. ATMOSPHERIC EFFECTS ON THE TRANSPORT PROPERTIES OF BI_2SE_3 138

4.1 Surface versus bulk state in topological insulator Bi_2Se_3 under environmental disorder 139

CHAPTER 5. TOPOLOGICAL PHASE TRANSITION IN $(BI_{1-X}IN_X)_2SE_3$ 147

5.1 Topological-metal to band-insulator transition in $(Bi_{1-x}In_x)_2Se_3$ thin film: DC transport and angle resolved photo-emission spectroscopy 148

5.2 A sudden collapse in the transport lifetime across the topological phase transition in $(Bi_{1-x}In_x)_2Se_3$ as measured by time domain terahertz spectroscopy 159

5.3 Conclusion and future prospects 163

CHAPTER 6. EMERGENCE OF DECOUPLED TOPOLOGICAL SURFACE STATES IN BULK-INSULATING TOPOLOGICAL INSULATOR THIN FILMS AND INVERSE TOPOLOGICAL INSULATOR HETEROSTRUCTURES 165

6.1 Weak anti-localization 166

6.2 Emergence of decoupled surface transport channels in bulk insulating Bi_2Se_3 thin films 170

6.2.1 Cu-doped Bi_2Se_3 thin films 170
6.2.2 Compensation doping of Bi_2Se_3 174
6.2.3 Shubnikov de Hass oscillations and Lifshitz-Kosevish analysis 178
6.2.4 Shubkinov de Hass oscillations in Cu-doped Bi_2Se_3 180
6.2.5 Angle resolved photo-emission spectra compared with transport data 182
6.2.6 Weak anti-location and decoupled topological surface states 183
6.2.7 Origin of suppressed bulk conduction: thickness dependant band bending 187
6.2.8 Origin of suppressed bulk conduction: Impurity band 192
6.2.9 Concluding remarks 194

6.3 Inverse topological insulator with tunable intersurface coupling 195

6.3.1 Experimental methods 197
6.3.2 Bi_2Se_3-In_2Se_3-Bi_2Se_3 inverse TI heterostructure 198
6.3.3 Computational methods 201
6.3.4 Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 heterostructure and tunable quantum mechanical tunneling 204
6.3.5 Concluding remarks 208

6.4 Conclusion 209

CHAPTER 7. TRANSPORT PROPERTIES OF TOPOLOGICAL INSULATOR: BAND BENDING, METAL-TO-INSULATOR TRANSITIONS, AND WEAK ANTI-LOCALIZATION 210

7.1 Bulk metal-to-insulator transition 210
7.2 Band bending effects 215
7.3 Weak anti-localization 224
7.4 Conclusion 229

CHAPTER 8. EPILOGUE 230

8.1 Review of results 230
8.2 Future research directions 232

CHAPTER 9. 235

9.1 List of abbreviations 235
9.2 References 238

List of Tables

Table 4-1 Taken from two 16 QL samples grown on Al_2O_3, this table shows the carrier density, n_{SS} (surface state) and n_{BS} (bulk state) and the corresponding mobilites obtained by the non-linear fitting process for G_{xy}. 144

Table 6-1 Comparison of data obtained by the Hall effect, SdH oscillations, and values calculated via an empirical curve fitted to an ARPES spectrum. 183

List of Figures

Figure 1-1 A comparison of the fundamental physics at high energy scales, and increasing complexity. 2

Figure 1-2 Two examples of how the ground state of a system breaks the symmetry. 3

Figure 1-3 An illustration of the band theory of crystals. 5

Figure 1-4 A comparison between a band insulator and a quantum hall insulator. 6

Figure 1-5 The real space chiral edge mode for the quantum Hall state , and the corresponding band diagram 7

Figure 1-6 A comparison of the quantum hall state and the quantum spin hall state 9

Figure 1-7 An illustration of the spin-orbit coupling effect. 10

Figure 1-8 Surface state dispersion that exists inside the gap formed by the bulk conduction and valence bands. 11

Figure 1-9 A schematic showing how the electrons in TI states are protected from 180° backscattering. 13

Figure 1-10 A schematic of the electronic potential near an interface. 14

Figure 1-11 A plot of E vs. κ^2 gives insight into the physics of the wave function in the bulk and close to the surface. 16

Figure 1-12 A comparison of surface states, bulk states and surface-bulk resonance states. 18

Figure 1-13 Experimental evidence for the 2D TI state. 20

Figure 1-14 A schematic setup of a typical ARPES experiment. 21

Figure 1-15 The topological crystal insulator state. 23

Figure 1-16 The topological Dirac semimetal phase. 25

Figure 2-1 A comparison of the different crystal geometries that are synthesised for scientific studies, where examples are the topological material Bi_2Se_3. 29

Figure 2-2 A naturally occurring example of epitaxial growth of rutile (TiO_2) on hematite (Fe_2O_3). 31

Figure 2-3 The increases in the electron mobility of GaAs over time. 34

Figure 2-4 This schematic depicts the physical and chemical processes of binding an element *A* to a substrate *S*. 38

Figure 2-5 A scanning tunnelling microscopy image of the growth of Cr oxide on a Pt(111) surface..... 39

Figure 2-6 The possible structures a material can take when accumulated on a substrate. 40

Figure 2-7 A schematic of the possible growth modes when depositing a film on a substrate. 42

Figure 2-8 A schematic showing examples of the three classes of crystallographic defects. 44

Figure 2-9 A schematic describing how a magnetron sputter deposition system operates. 48

Figure 2-10 A schematic of a pulsed laser deposition system. 50

Figure 2-11 A schematic of a molecular beam epitaxy system. 52

Figure 2-12 The MBE system at Rutgers University. 55

Figure 2-13 Substrate mounting..... 57

Figure 2-14 The process of transferring a substrate from atmosphere through to the growth chamber. 58

Figure 2-15 Various images of an MBE thermal cell during routine maintenance. 59

Figure 2-16 Cut-away diagrams of the various types of pumps used on the MBE chamber, with the corresponding pictures of the pumps on the MBE chamber directly below. 62

Figure 2-17 Images of the QCM system on the MBE. 68

Figure 2-18 A typical Rutherford backscattering spectroscopy experiment. 71

Figure 2-19 Simulations of RBS spectra for various film structures. 73

Figure 2-20 An RBS spectrum used in a calibration experiment on $(Bi_{1-x}In_x)_2Se_3$ where x was targeted to be ~ 30%, and a thickness of ~ 20 nm (1 QL ≈ 1 nm), which was used to grow films within ~ 1% of the targeted composition. 75

Figure 2-21 Images and schematics for in situ reflection high-energy electron diffraction. 78

Figure 2-22 The universal behavior of the mean free path of electrons in solids as a function of energy for various elements. 79

Figure 2-23 The Ewald-construction for diffraction of an incoming electron beam by a crystal lattice. 81

Figure 2-24 A schematic of a RHEED pattern formed from the diffraction of an electron beam off of a 2D surface. 83

Figure 2-25 Various crystal parameters obtained from RHEED images. 85

Figure 2-26 Various RHEED images to illustrate typical crystal structures of varying quality and various surface morphologies. 86

Figure 2-27 A schematic of the surface during deposition for the layer-by-layer and step-flow growth modes, respectively. 88

Figure 2-28 X-ray diffraction showing the QL peaks $(Bi_{1-x}In_x)_2Se_3$. 89

Figure 2-29 Atomic force microscopy. 91

Figure 2-30 Several AFM images of Bi_2Se_3 grown on different substrates that show the different surface morphologies. 93

Figure 2-31 Various structural measurements of Bi_2Se_3 grown on Si (111) by the two step growth method. 97

Figure 2-32 Various RHEED images which were used to determine the optimal growth conditions. 99

Figure 2-33 Structural data for Bi_2Se_3 grown on Al_2O_3 substrates. 101

Figure 2-34 Structural measurements of Bi_2Se_3 grown on amorphous SiO_2. 104

Figure 2-35 Structural data for $(Bi_{1-x}In_x)_2Se_3$ films. 106

Figure 2-36 Transport properties versus thickness that originates from defects in the materials. 108

Figure 2-37 A graphical depiction of the growth recipe for Bi_2Se_3 on the In_2Se_3 buffer layer. 110

Figure 3-1 A schematic of the acceleration of free carriers and the scattering processes in a material under an applied electric field. 115

Figure 3-2 A schematic of a Hall effect measurements on a sample of width *w* and thickness *t*. 119

Figure 3-3 A schematic of the different sample geometries used for electric transport measurement. 122

Figure 3-4 A demonstration of the need for, and the result of, data symmetrisation. 125

Figure 3-5 Lead and sample preparation..... 127

Figure 3-6 A schematic of the source meter switch setup. 128

Figure 3-7 The front interface of the custom program designed and written to perform transport measurements. 130

Figure 3-8 The flow diagram for the progression of the measurement process..... 131

Figure 3-9 Images of the American Magnetic, Inc liquid helium cryostat used to perform transport measurements down to 1.5 K and up to 9 T. 133

Figure 3-10 Images of the Advance Research Systems cryostat and GMW magnet used to perform transport measurements down to 4 K and up to 0.6 T. 134

Figure 3-11 The variable ramp rate magnetic field control program. 136

Figure 4-1 A band bending diagram showing the change from depletion to accumulation. 139

Figure 4-2 The response of Bi_2Se_3 thin films to exposure to atmospheric gases. 141

Figure 4-3 The response of Bi_2Se_3 thin films grown on Al_2O_3 to exposure to atmosphere. 142

Figure 4-4 The response of Cu-Bi_2Se_3 thin films grown on Al_2O_3 to exposure to atmosphere. .. 145

Figure 5-1 Structural data for $(Bi_{1-x}In_x)_2Se_3$ thin films for various values of x. 149

Figure 5-2 Scanning tunnelling microscopy images taken from Bi_2Se_3 and $(Bi_{1-x}In_x)_2Se_3$ with $x \approx 5\%$. 150

Figure 5-3 Resistance versus temperature for $(Bi_{1-x}In_x)_2Se_3$ thin films from the TI regime through to the band insulating regime. 152

Figure 5-4 Magnetotransport data and ARPES data for $(Bi_{1-x}In_x)_2Se_3$ thin films from the TI regime through to the band insulating regime. 155

Figure 5-5 A schematic showing how the conduction band and valence band evolve spatially across the topological phase transition. 157

Figure 5-6 The real conductance versus frequency for Bi_2Se_3 and $(Bi_{1-x}In_x)_2Se_3$. 160

Figure 5-7 The scattering rate for $(Bi_{1-x}In_x)_2Se_3$ thin films showing a thickness-dependant jump across the TI phase transition. 161

Figure 5-8 The phase diagram summarizes the behavior of the topological properties as a function of thickness and concentration..... 163

Figure 6-1 Cartoons showing how weak localizatoin and weak anti-localization occur due to coherenet backscattering along time reversable paths in the absence and presence of the spin-momentum locking, respectively. 167

Figure 6-2 Structural measurements showing that the Cu-doped Bi_2Se_3 remain single phase. ... 172

Figure 6-3 Hall effect measurements as a function of Cu-doping (fixed thickness ≈ 20 QL) and thickness (fixed Cu concentration, $x \approx 2.5$-4.0%)..... 173

Figure 6-4 Schematic diagrams showing how doping Bi_2Se_3 by Cu occurs. 175

Figure 6-5 XPS data for Cu-2p3/2 and 2p1/2 for bulk Cu and Cu-Bi_2Se_3 with $x \approx 0 – 22\%$..... 177

Figure 6-6 SdH oscillations and comparison with Hall effect for an optimally-doped 20 QL sample ($x \approx 2.5–4.0\%$). 179

Figure 6-7 An ARPES spectrum overlaid by the empirical functions, $E_{ave}(k) = 1.9k + 12.6k^2 + 2300k^6$ for the TSS and $E_{CB}(k) \approx \hbar^2k^2/(2m^*) + E_{DP}$ for the conduction band with $m^*/m \approx 0.15$ and $E_{DP} \approx 230$ meV. 182

Figure 6-8 Weak anti-localization effect showing the thickness and doping dependent coupling of the TSS. 184

Figure 6-9 Resistance vs. temperature for t = 2, 10, 20, 75, and 150 QL (1 QL ≈ 0.95 nm). 186

Figure 6-10 Band-bending dependence on the TSS carrier density: n_{ss} represents the total carrier density from top and bottom TSS. 189

Figure 6-11 This illustrates how band bending behaves as the films are made thinner. 190

Figure 6-12 As the films are made thicker, the conduction band minimum as a function of distance from the surface dips down closer to the E_F. 191

Figure 6-13 A schematic and TEM image showing a conventional TI vs an inverse TI. 196

Figure 6-14 Weak anti-localization and the first principles calculations for Bi_2Se_3–In_2Se_3–Bi_2Se_3 films. 199

Figure 6-15 Weak anti-localization and a schematic of the quantum tunnelling model for Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$–Bi_2Se_3 films. 205

Figure 6-16 Weak anti-localization for Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 and a phase diagram for interlayer coupling. 207

Figure 7-1 The calculated Fermi energy, E_F, relative to the conduction band minimum, as a function of buk carrier density. 212

Figure 7-2 The Mott and Ioffe-Regel criteria showing when a metal-to-insulator transition will occur based on the mobility and the bulk carrier density. 214

Figure 7-3 Band bending direction depending on the surface carrier density. 215

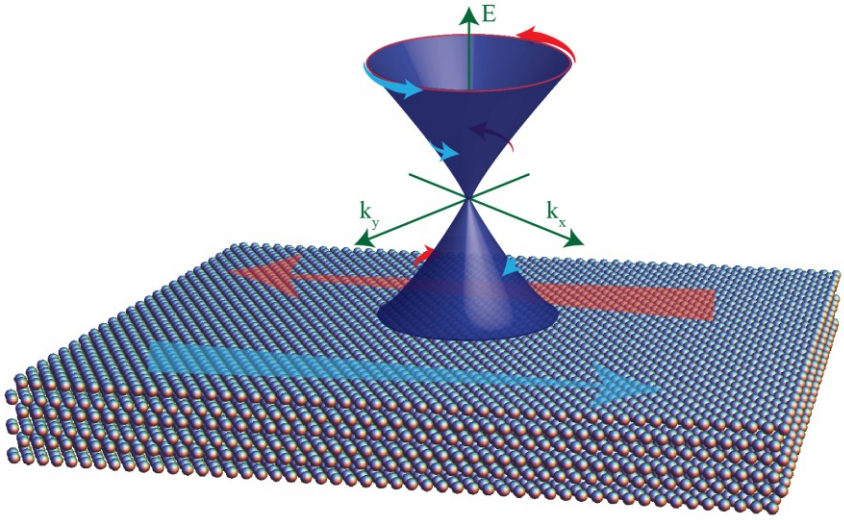
Figure 7-4 Band bending parameters for upward band bending. 218

Figure 7-5 Upward band bending vs thickness. 223

Figure 7-6 The $\tilde{A}$ parameter in the HLN fomula vs Bi_2Se_3 film thickness for various samples with different qualities. 226

Chapter 1.

Prologue[1]

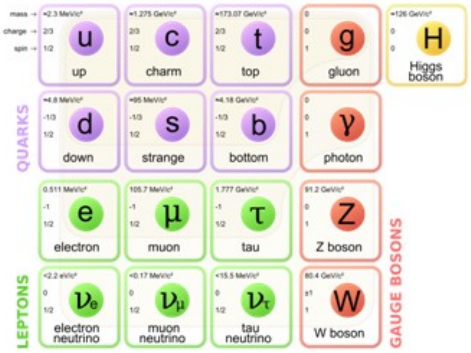


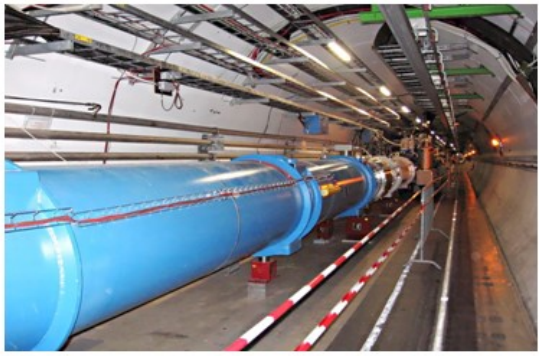
The basic goal pursued by physics is to understand the natural world, and find the most fundamental laws. Only a short time after the discovery of the atom, more than 100 years ago, the efforts of the physics community began to split roughly into two directions: one group sought to find the most basic building blocks by probing fundamental particles and interactions at ever decreasing length scales and increasing energy scales ([49]Close, 2004). Utilizing giant particle colliders, such efforts brought to the forefront the interworking of the fundamental forces in nature and an exotic zoo of particles that make up the standard model, as shown in Figure 1-1a. While as Figure 1-1b shows, the second group took a direction seeking fundamental explanations to the phenomena that emerged when large groups of simple building blocks are assembled. This resulted in discoveries of complex phases such as superconductivity ([5]Anderson, 1972), and the basic understanding of semiconductors, which have enabled the development of the entire electronics industry that we rely on today. The work presented here takes the direction of the latter path, and is concerned with understanding the basic properties of peculiar electronic states that emerge only on the surfaces of a very special class of materials.

[1] A illustration showing the linear dispersion and spin-momentum locking of the electrons on the surface of a topological insulator

a

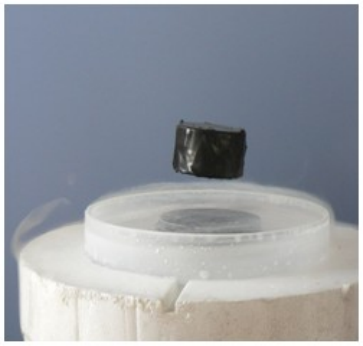
Subatomic physics: length scale $\lesssim 10^{-18}$ m, energy scale $\gtrsim 10^{13}$eV





b

Condensed matter physics: particle number ~ 10^{23} atoms



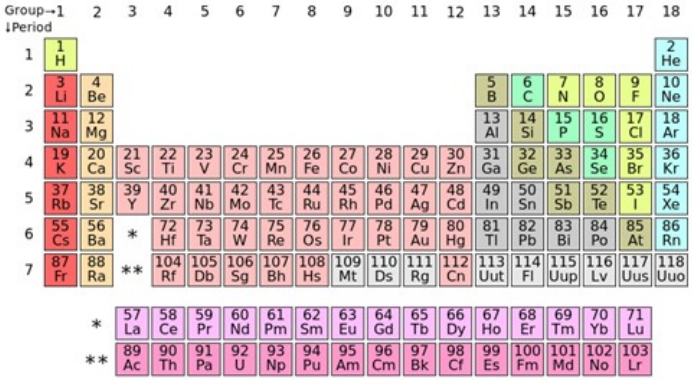


Figure 1-1 A comparison of the fundamental physics at high energy scales (a), and increasing complexity (b). a Subatomic physics is composed of a zoo of particles summarized by the standard model (left), which are probed at the Large Hadron Collider (right) and other particle accelerators. **b** In the world of atomic physics, novel electronic phases such as high temperature superconductivity emerge (left, showing magnetic levitation by the Meissner effect) when complex structures are formed from the multitude of elements in the periodic table (right). (Images adapted from Wikipedia.org)

This chapter will review and give a basic introduction to the rapidly developing field that has been coined *topological insulators* ([137]Moore and Balents, 2007). First, we will sketch a basic picture of what is topological about topological insulators, which will be followed with a brief review of the initial breakthroughs that ignited the field, and will carry through to incorporate the state of research today and introduce some of the new topological materials. Finally, we will give an overview of the contents and the goals of this work. (See [85]Hasan and Kane, 2010;[136]Moore,

2010;[157]Qi and Zhang, 2010;[56]Culcer, 2012 for more in depth discussions and a full set of references therein).

1.1 Classifying insulators: band-insulators and the quantum hall state

The idea of symmetry is intimately related to the fundamental laws of physics. Physical systems can be classified by the symmetries that they spontaneously break. As shown in Figure 1-2a-b, the magnetic interaction in solids obeys rotational invariance, but the ground state is chosen along a particular axis, which breaks the global rotational symmetry (for the case of ferromagnetism

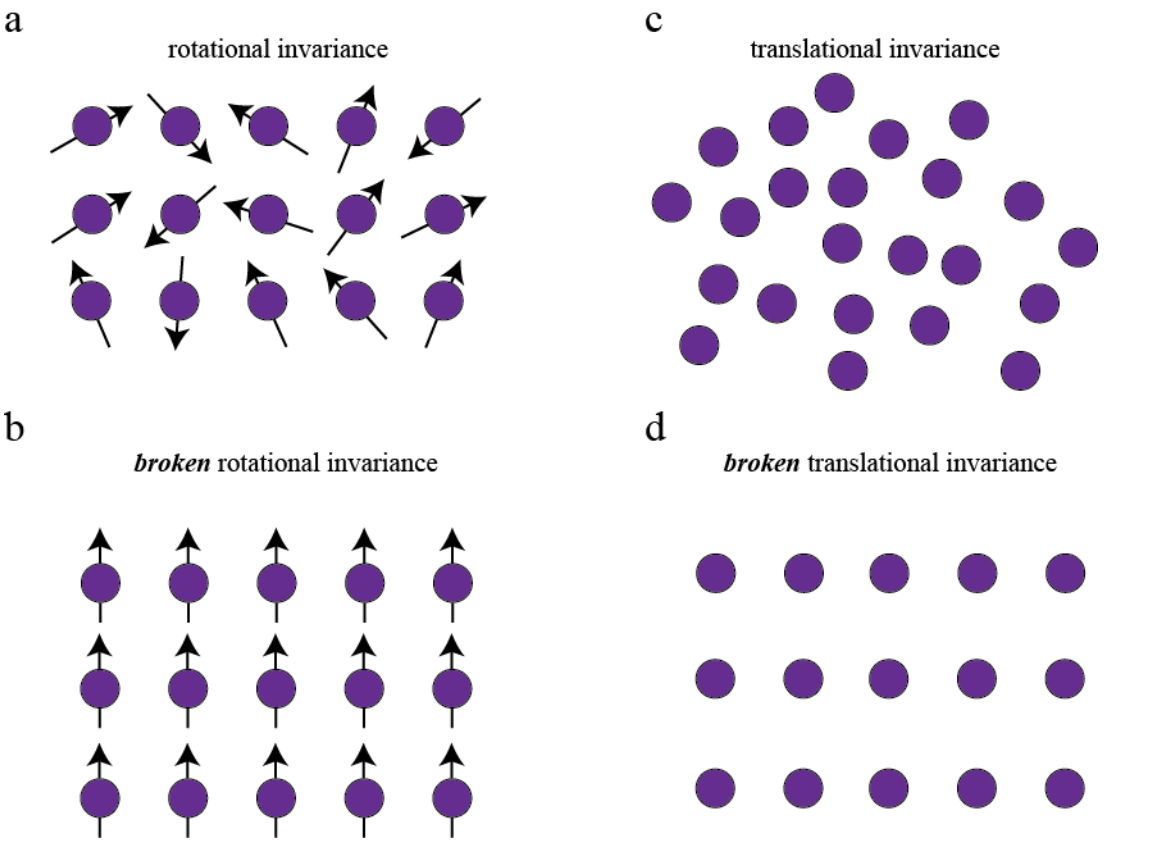


Figure 1-2 Two examples of how the ground state of a system breaks the symmetry. a-b The magnetic interaction does not break rotation invariance (**a**), while the ground state of the ferromagnetic interaction is one which chooses a particular direction thereby breaking rotational invariance (**b**). **c-d** The translation invariance of atoms placed at random locations (**c**) is broken when it is energetically favorable to crystallize into a lattice (**d**).

this has been a contentious issue, see [6]Anderson *et al.*, 1990;[151]Peierls *et al.*, 1991;[150]Peierls, 1992). Similarly, the translational invariance of the electrostatic interaction that drives the random mixture of atoms shown in Figure 1-2c into the crystalline solid in Figure 1-2d is spontaneously broken because the full translational invariance is replaced by translation invariance by a lattice vector. This idea of spontaneously broken symmetry, envisaged first by Lev Landau, was thought to be an encompassing means for classifying phases of matter. However, with the discovery of the quantum hall state (QHS) in the 1980s, it has been found that this had to be augmented by the notion topological order. Before moving on to the topological classification scheme, it will be instructive to briefly introduce the band theory of crystals.

1.1.1 Band theory of crystals

Much like the Newtonian soccer ball shown in Figure 1-3, in free space electrons can take a continuous range of energy with $E = p^2/(2m)$. However, if electrons are placed in a crystal, they see a periodic background potential due to the atomic cores, which modifies the energy relation. Electrons moving through the crystal can diffract off the atomic sites, which causes destructive and constructive interference along certain crystal directions when the electrons have certain wavelengths. This is qualitatively similar to the interference pattern formed by light on a diffraction grating. The electrons that scatter destructively effectively become localized, which, as shown in Figure 1-3, gives rise to the continuous energy band being split into a spectrum of energy bands that are separated by gaps.

The position of these bands determines the electronic properties of the material. For example, electrons, which distribute themselves energetically according to Fermi-Dirac statistics, will, like water filling a container, fill the lowest energy bands first up to a point called the Fermi level E_F, which is analogous the surface of the water. If E_F falls inside the band gap (E_B, which is typical of order ~1 eV), it will leave a completely filled band below (the valence band, VB), and a completely unfilled band above (the conduction band, CB); the material will be an insulator and

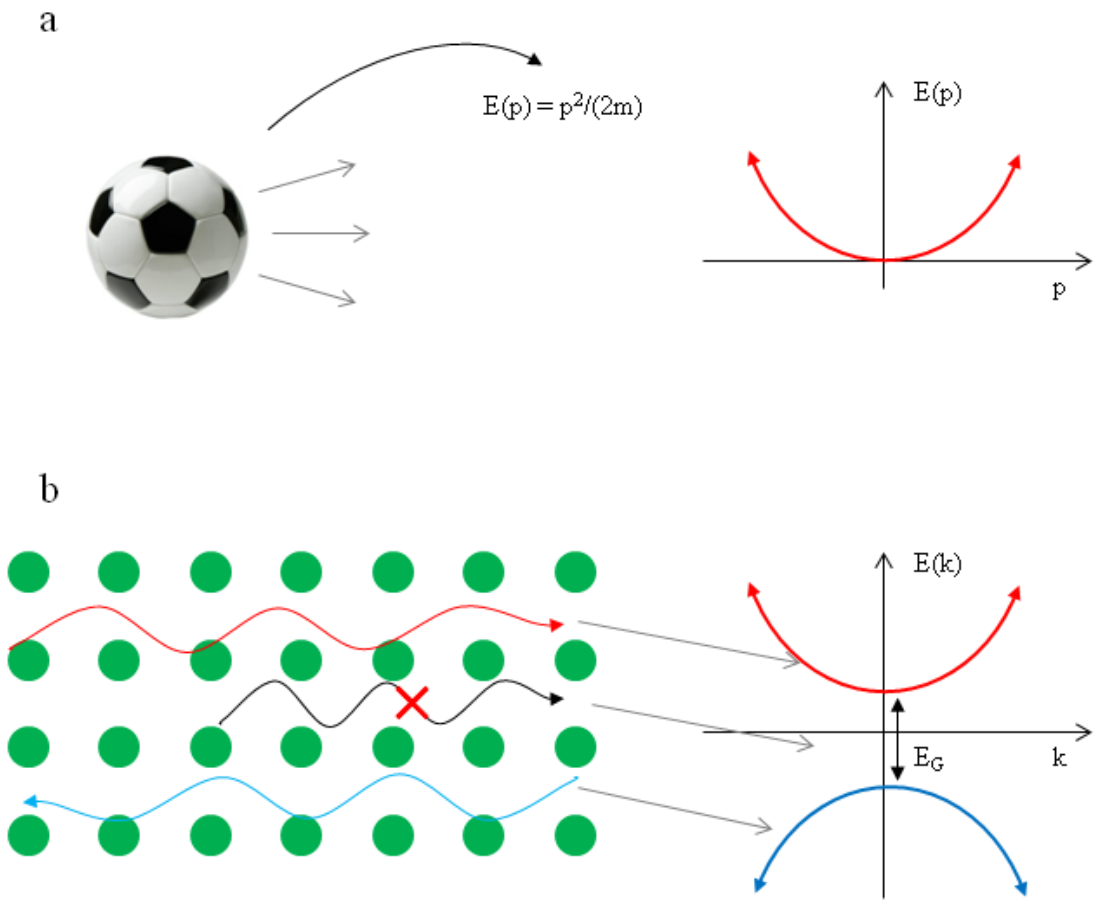


Figure 1-3 An illustration of the band theory of crystals. a For classical objects, and quantum mechanical objects in free space, there are continuous energy states, while in **b** the lattice of crystal causes destructive interference which induces gaps (E_G) in the energy diagram for certain energy ranges.

will only allow charge to flow if the applied electric field is sufficient to excite an electron from the VB to the CB. In contrast, a crystal will be metallic when a band is partially filled, which allows charges to be excited and conduct with an arbitrarily small electric field. In this sense, the vacuum of free space is an insulator with band gap of $\sim 1 \times 10^6$ eV, which is the excitation energy for pair-production of electron-position pairs (however, if there are free electrons floating around, then the vacuum can become metallic because the free electrons can be accelerated by an arbitrary small electric field, such is the case for a plasma). As we will see next, invoking the notion of topological order can be used to classify these various insulating states.

1.1.2 Topology: doughnuts, soccer balls and the quantum hall state

The common notion of topology is typified in the separation of soccer balls and doughnuts into different classes based on the number of holes each item has. In this sense, and as we will see, in the sense of electronic materials, topology is the ability to describe objects based on the broad details that are not sensitive to minor changes. Gauss and Bonnet made this precise by defining a single number called the genus, which is sufficient to distinguish geometric objects ([140]Nakahara, 1990). For the example of a 3D object, this number is constructed by integrating the local curvature of the object over the entire 2D surface. This number is found to be quantized to integer values that correspond to the number of holes each object has, and it is insensitive to smooth deformations. This can be illustrated by the pedagogical example of classifying the English alphabet into three classes. Class 1: [*C, E, F, G, H, I, J, K, L, M, N, S, T, U, V, W, X, Y, Z*], Class 2:[*A, R, D, O, P, Q*], and finally Class 3:[*B*]. All the objects within each class can be smoothly deformed to another element within the same class (e.g. *P* can be deformed to make *A*), but can only be deformed to an object in another class by, putting it in mathematically rigorous terms, filling in a hole or poking a new hole (e.g. *T* can be deformed to *P* only if a hole is made), which in mathematical terms is considered to be *a dramatic change*. By extending this idea of geometric topology, we can move on and discuss the first example of a topological material called the

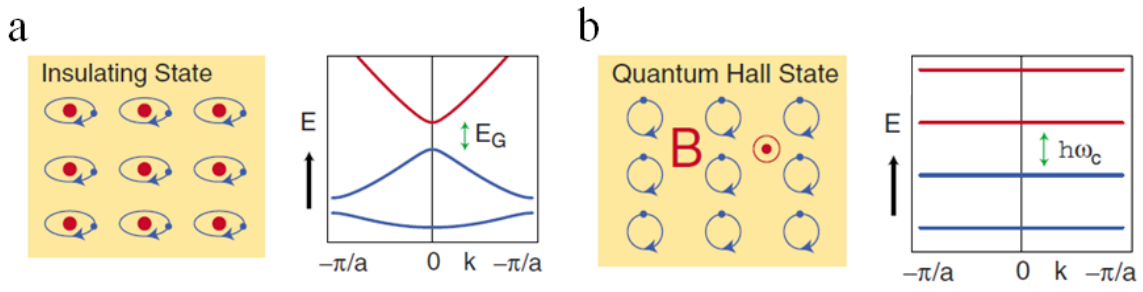


Figure 1-4 A comparison between a band insulator and a quantum hall insulator. a (left) A diagram of electrons localized around atoms and the corresponding band diagram (right) where the Fermi level sits in the bulk band gap. **b** The situation is nominally identical in the QHS with electrons localized around magnetic flux lines, which act analogously as the atoms, and the corresponding band diagram where the energy bands are separated by the cyclotron energy $\hbar\omega$. (Adapted from [85]Hasan and Kane, 2010)

quantum hall state (QHS).

When a magnetic field is applied perpendicularly to a 2D electron gas (2DEG), the free electrons within feel the Lorentz force, and begin to make cyclotron orbits. When the size of the orbits become sufficiently small, the 2D electrons become localized and this system becomes insulating (this final localization is due to the presence of a sufficiently small amount of disorder that prevents the drift of the orbits in the applied electric field, but this is beyond the scope of the discussion here (see [154]Prange and Girvin, 1990)). This semi-classical notion can be replaced by a more precise quantum mechanical model that predicts that these cyclotron orbits become quantized, and the continuous energy bands become split into distinct energy levels, called Landau levels (LL). Just like the band insulator discussed above, if the Fermi falls in between LL the system will be insulating, and at partial filling then the system is metallic. As shown in Figure 1-4, looking at the energy diagram alone, one cannot distinguish this state from that of a normal band insulator because these states are nominally identical from the point of view of symmetry.

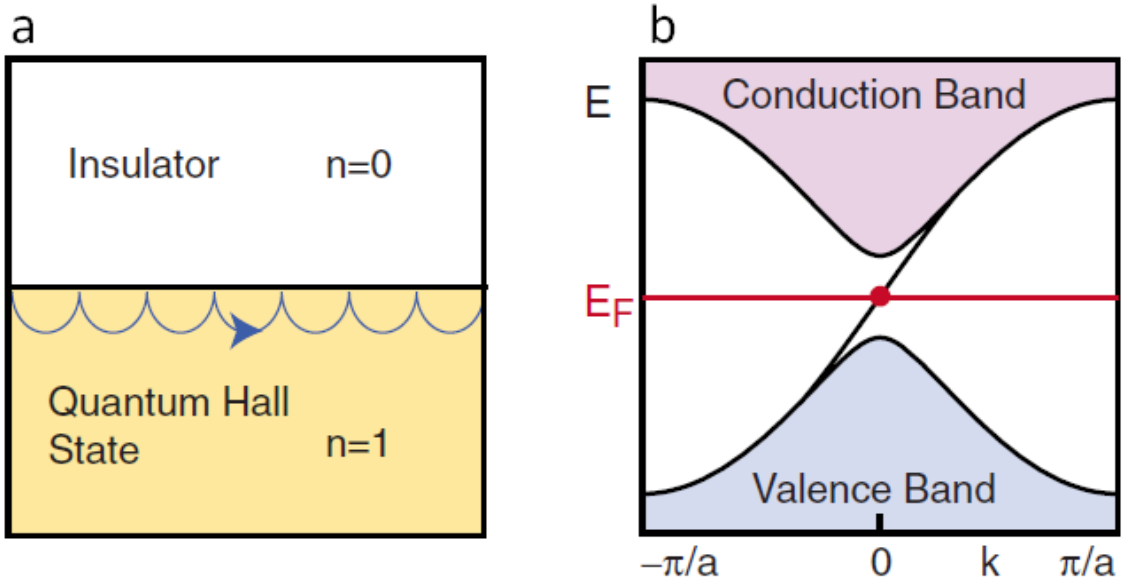


Figure 1-5 The real space chiral edge mode for the quantum Hall state (a), and the corresponding band diagram (b). The bulk boundary correspondence theorem states that at the interface between two insulators with different topological invariants there exist edge modes with the number of channels given by the difference in the invariants across the interface. (Adapted from [85]Hasan and Kane, 2010)

The most telling difference between the insulating behavior of the QHS and that of normal insulator is that the QHS is not entirely insulating. Rather, a very special metallic state exists at the edge of the 2D material with nearly zero resistivity (this is the system with the lowest known resistance other than superconductors) ([111]Klitzing *et al.*, 1980). The development of this metallic channel on the boundary is a direct consequence of the topology of the materials ([183]Thouless *et al.*, 1982).

The topological description of matter comes about in a similar way as with geometric objects. The role of the geometric shape is replaced by characteristics of the wave-function of the electrons and the ordering of the energy bands ([21]Berry, 1984). In the case of the QHS, a topological invariant emerges and is known as the Chern number, which is quantized to integer values that correspond to the number of LL below the Fermi level. This topological invariant separates these states from vacuum and normal band insulators, which always have a Chern number equal to zero. Physically, the QHS is easily distinguishable from the normal insulating states by the quantized conduction channels that exist on the edge, which can be ascribed to the topological nature through the bulk-boundary correspondence theorem (BBC) ([85]Hasan and Kane, 2010). The BBC theorem states that at the interface between materials described by different topological invariants, there necessarily exists gapless metallic states with the number of channels given by the difference between the topological invariant across the interface. For the case of the QHS shown in Figure 1-5, the number of edge modes, with quantized conductance, is given by the difference in Chern number between the 2DEG and the surrounding materials (typically a normal insulator or vacuum with Chern number equal to zero). The topologically non-trivial QHS requires a large magnetic field to push it into an insulating state, and the first system theorized to show a non-trivial topological phase without a net magnetization was imagined in the late-1980s by F. D. M. Haldane ([83]Haldane, 1988), and more recently in work led by C. L. Kane ([102]Kane and Mele, 2005;[103]Kane and Mele, 2005;[75]Fu and Kane, 2007;[76]Fu *et al.*, 2007), S.-C. Zhang ([19]Bernevig *et al.*, 2006;[20]Bernevig and Zhang, 2006;[156]Qi *et al.*, 2006;[155]Qi *et al.*, 2008), and J. E.

Moore and L. Balents ([137]Moore and Balents, 2007), which brought about the nascent field of topological insulators.

1.2 Topological insulators

Topological insulators (TI) are like a version of the QHS that obey time-reversal symmetry (TRS), so much that, the 2D TI state is also known as the *quantum spin hall state* (QSHS). This is because the QSHS can be formed by two QHS moving in opposite directions around the system with anti-parallel spin polarizations (see Figure 1-6). The magnetic field in QHS explicitly breaks TRS, which is reflected in the edge modes that are spin degenerate and propagate around the system in one direction; under time reversal, the edge modes change direction, which breaks the forward-backward time symmetry. In the QSHS, the edge modes emerge because there is an effective magnetic field that is intrinsic to the material, which obeys TRS. This effective magnetic field is provided by the spin-orbit coupling (SOC) effect, which can be thought of as the magnetic field that is created in the rest frame of the electrons; as the electrons are moving along, they see the charged nuclei moving backwards, which, as shown in Figure 1-7, in their reference frame, creates an effective magnetic field. In materials with heavy nuclear masses, the SOC field can be much stronger than external applied fields.

TIs are characterized by a topological invariant that belongs to the Z_2 family of numbers,

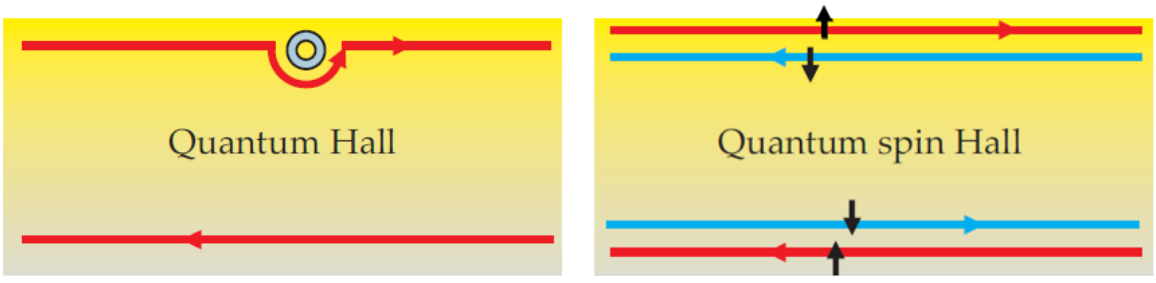


Figure 1-6 A comparison of the quantum hall state (left) and the quantum spin hall state (right). The QHS has a single chiral edge mode that moves in one direction, which violates time-reversal symmetry. While, the QSHS has edge modes that move in both directions with opposite spin polarizations; this state obeys time-reversal symmetry. (Adapted from [157]Qi and Zhang, 2010)

which gave rise to the original name Z_2 *topological insulators* that has since been shorten to just *topological insulators* ([137]Moore and Balents, 2007). The Z_2 class can take only one of two values, 1 or 0, and, therefore, based on the BBC theorem interfacial states should emerge at the interface between materials that are described by different Z_2 invariants. For 2D TIs this interface state is an edge state composed of an odd number of metallic modes, and for 3D TIs, the focus of this work, this state is a 2DEG, and these topologically induced states display many novel properties, which are of great interest to the scientific community with many projected electronic applications ranging from quantum computation ([177]Stern and Lindner, 2013) to spin-based electronics ([212]Yokoyama and Murakami, 2014).

As mentioned previously, the origin for the non-trivial topology is rooted in the strong SOC. In TIs, SOC is large enough to alter the band structure by reversing the ordering of the energy bands nearest to E_F, and thereby changing the topological invariant from 0 to 1. This band

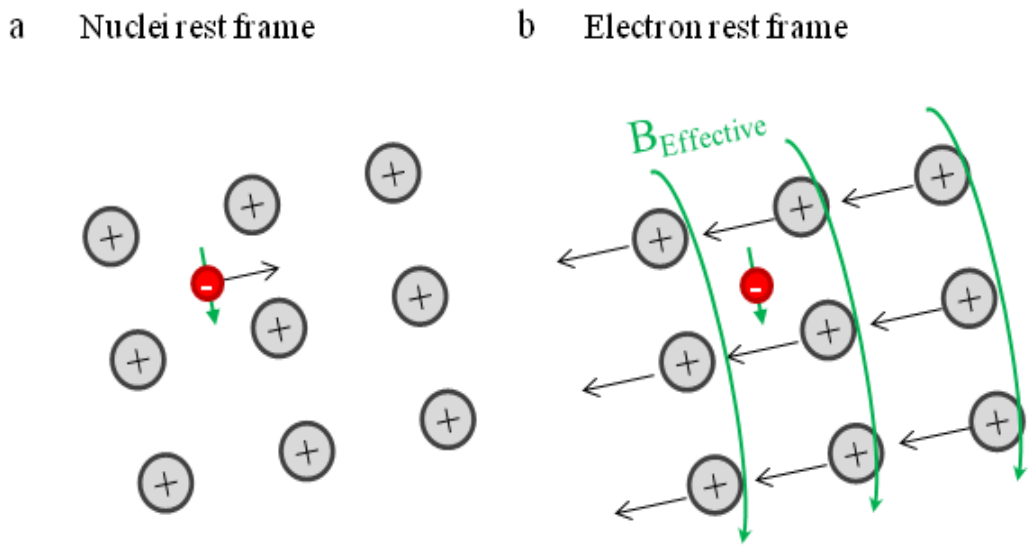


Figure 1-7 An illustration of the spin-orbit coupling effect. a An electron moving in the reference frame of the nuclei. **b** If the reference frame is boosted to that of the electron, the positively charged nuclei appear be moving with opposite velocity. This, then, generates an effective magnetic field, which the spin of the electron orientates parallel to. This interaction is due to spin-orbital coupling, and in topological insulators this is the origin for the spin-momentum locking.

inversion can be physically seen to give rise to the emergence of the surface bands by the following. Imagine two materials, one with inverted band, and a second with non-inverted bands: at the interface between these two materials the bands should be connected in a continuous fashion, and, therefore, they must cross at the interface. This gives rise to metallic interfacial states because the Fermi level must cross these intersecting bands, which results in partial filling (i.e. metallic) of the bands at the interface. In fact, this idea can be traced back to the early 1980s ([192]Volkov and Pankratov, 1985;[70]Fradkin *et al.*, 1986), but it was only with the language of topology that these ideas took full shape.

These topological states are immune to many forms of disorder, which is a reflection of their topological origin. To understand why this is indeed true, we can make use of Kramers' theorem ([85]Hasan and Kane, 2010), which states that for a system that is invariant under time-reversal (i.e. $t \rightarrow -t$), then the energy band must be two-fold degenerate at all the time-reversible points in the Brillouin zone. This means that at the points $k = 0$ and $\pm\pi/a$ (for 2D, but is general

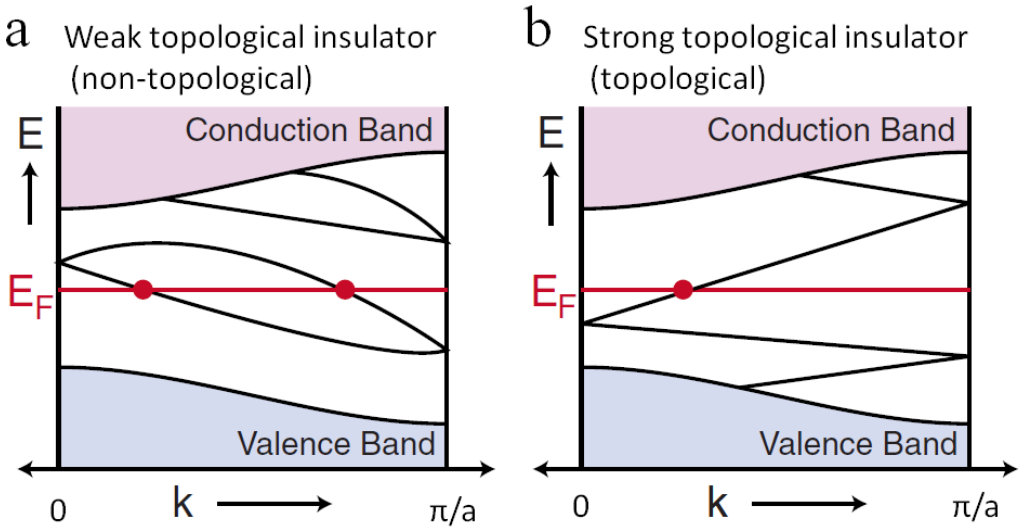


Figure 1-8 Surface state dispersion that exists inside the gap formed by the bulk conduction and valence bands. There are two different ways to connect Kramers pairs at the time reversible points in the Brillioun zone ($k = 0$ and $\pm\pi/a$, but $k = -k$ by time reversal symmetry, so only $k > 0$ is shown). **a** The situation where there are an even number of Fermi level (E_F) crossings, which is adiabatically connectable to trivial insulator states. While in **b** there are an odd number of crossings, which guarantee the surface states exist, since there will always be at least one crossing of the Fermi level. (Adapted from [85]Hasan and Kane, 2010 and [76]Fu *et al.*, 2007)

for 3D) that the surface bands cross, and since SOC acts like a magnetic field, it splits the bands based on the electrons' spins away from the time-reversible points. To see how this protects the existence of the topological states, Figure 1-8 shows two possible ways to connect the time-reversible-points, with an even number of Fermi level crossings (Figure 1-8a) or with an odd number of Fermi level crossings (Figure 1-8b) ([76]Fu *et al.*, 2007). For the case of an even number of crossings, one can imagine shifting the bands around by the application of disorder to the system; this can result in zero crossings of the Fermi level, which is the manifestation of adiabatically deforming this system into a trivial, or weak topological state (geometrically this is analogous to morphing the letter *A* into the letter *P*). The other situation is having an odd number of crossings of the Fermi level; one quickly realizes that no matter how these energy bands are deformed that there will be at least one, but more generally an odd number of Fermi level crossings regardless of how the band structure is shifted, which shows that the surface will always remain metallic.

From Figure 1-8, one can see another interesting physical property of these states, which is their energy momentum relations, or dispersion relation; near the time-reversible points $k = 0$ and $\pm\pi/a$, they disperse linearly with momentum. Compared to normal electrons and holes, which disperse like $E \sim k^2$, these surface states share a similarity with massless photons whose dispersion is linear in momentum. For a 2D surface, this linear dispersion is promoted to a cone, often referred to as a *Dirac cone*, and shares traits with the most famous material with Dirac electrons, graphene ([78]Geim and Novoselov, 2007). The main electronic difference between the surface states of TIs and graphene is the electron degeneracy. In graphene there are two Dirac cones per unit cell with each cone containing both up and down spin electrons. However, TI surface states have an odd number of Dirac cones (in the simplest case one Dirac cone per unit cell), and each being spin-polarized, where only one spin species occupies each state. This is due to the strong SOC field that spin polarizes the carriers, and has many possible applications due to the spin-polarized currents in the field of spintronics.

Lastly, backscattering of electrons in TIs is suppressed. Electrical resistance is caused when electrons are deflected from forward motion. In the edge states of the QHS, electrons cannot backscatter since the states are chiral and move in only one direction, and, therefore, there are no states for the electrons to scatter into; the consequence of which, is the extremely low resistance in the QHS. In TIs, backscattering by 180° is also suppressed, which, as shown in Figure 1-9, is due to the large SOC that simultaneously winds the electron spin by π on the clockwise loop and -π on the counterclockwise loop. This accumulated phase causes the probability that an electron will backscatter by exactly 180° to be zero. As will be seen later in this work, for the case of 3D TIs, this does not result in ultra-high mobility conductance through the topological surface states (TSS), which is often a source of confusion. This is because scattering to angles less than 180° is a source of resistance and is not forbidden, and eliminating the 180° channel only decreases the resistance in the range of a few percent. However, this suppressed backscattering can be seen in transport measurements as the weak anti-localization effect, and, as we will see in Chapter 6, this effect will prove to be an important phenomenon to help understand the physics of the TSS through transport measurements.

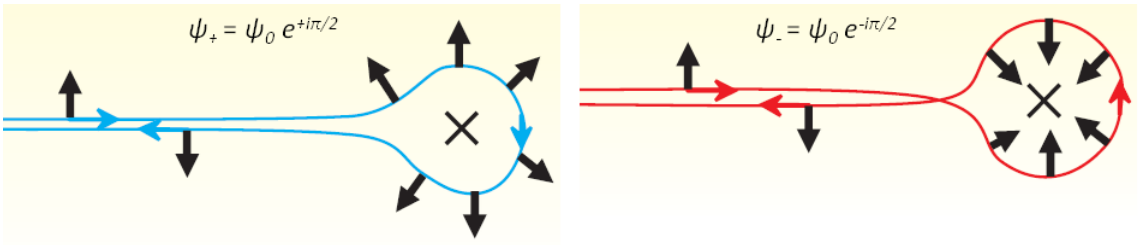


Figure 1-9 A schematic showing how the electrons in TI states are protected from 180° backscattering. When an electron backscatters off an impurity the amplitude to backscatter is composed of clockwise (left) and counterclockwise (right) components. Due to the strong SOC the spin of the electron is locked perpendicular to the momentum, and the electron accumulates a net phase of $\pm\pi$, with the factor of 1/2 in the exponent is due to the fermionic nature of electrons. Adding the contributions shows that the amplitude vanishes for exact 180° backscattering. (Adapted from [157]Qi and Zhang, 2010)

1.3 Topological surface states compared to non-topological surface states

The existence of surface states have been long studied, and dates back to the initial work by Tamm and Shockley in the early 1930s ([178]Tamm, 1932;[171]Shockley, 1939). The surface states in semiconductors and metals have been widely studied with broad ranging implications from semiconductors used in transistors, to the use as surface catalysis. From the viewpoint of topological surface states these classical, or non-topological surface states, come in even numbers and therefore lack the topological protection that TI surface states, which come in odd numbers, have. From the context of this classical notion of surface states, TSS are remarkable, and, therefore, in this section we will introduce surface states in a more classical way following A. Zangwill ([214]Zangwill, 1988), which allows one to see the unique nature of the TSS.

Central to the existence of an electronic state bound to the interface of a material is the modification of the electric potential at a surface. In an infinite crystal the potential is of the form

$$V(z) = V(z + a)$$

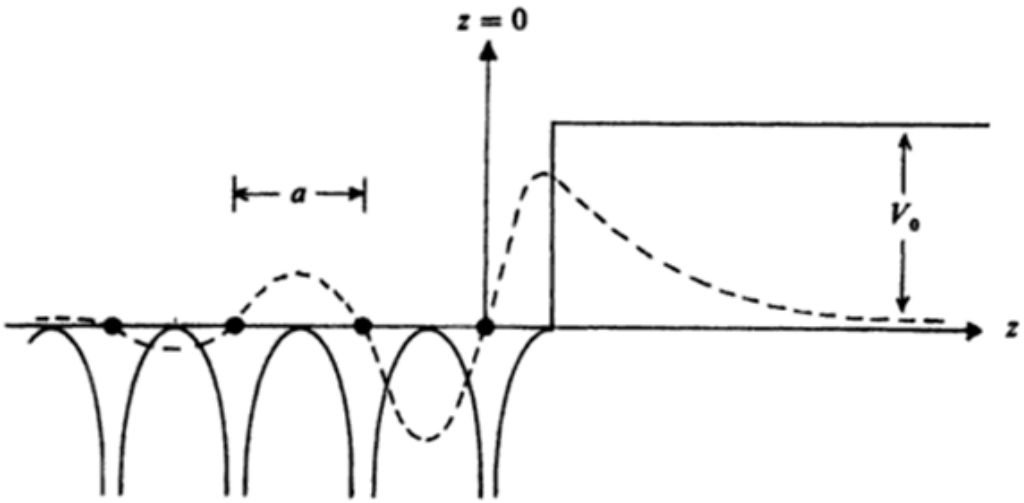


Figure 1-10 A schematic of the electronic potential near an interface. The solid line shows how the potential is modified near the boundary of the crystal at $z = 0$, and the dotted line shows the surface state wave function which decays into the bulk ($z < 0$) and into the vacuum ($z > 0$). (Adapted from [214]Zangwill, 1988)

where a is the lattice vector. With this type of potential, all the notions of bulk crystals come out naturally. For example, it gives rise to the entirety of band theory, which allows one to understand an immense amount about the physics of the bulk crystals. This periodic potential breaks down close to the surface of a material and needs to be modified to capture the new physics. As shown in Figure 1-10, in the simple nearly-free 1d electron approximation the modification due to the surface can be made as the following

$$V(z) = \begin{cases} -V_0 + 2V_g \cos(gz) & z < 0 \text{ Bulk} \\ 0 & z > 0 \text{ Vacuum} \end{cases}$$

where the surface is placed at $z = 0$, $z < 0$ is inside the crystal, as we will see V_g will give the bulk energy gap, and $g = 2\pi/a$, where a is, again, the lattice vector. This model captures the periodic nature of the screened ion cores, and the binding energy of the surface. Our goal now is to solve the 1d Schrodinger equation with the above potential. Since deep in the bulk the electrons are in nearly free Bloch states, and just outside of the crystal the electron wave function will decay exponentially, we can take a trial wave function of the form

$$\psi_k(z) = \alpha e^{ikz} + \beta e^{i(k+g)z}.$$

Plugging this trial wave function into the Schrodinger equation yields

$$(k^2 - V_0 - E)\alpha e^{ikz} + V_g\alpha(e^{i(k+g)z} + e^{i(k-g)z}) + ((k-g)^2 - V_0 - E)\beta e^{i(k+g)z}$$

$$+ V_g\beta\left(e^{i(k)z} + e^{i(k-2g)z}\right) = 0.$$

To satisfy this equation the coefficients for α and β must vanish independent of z. Therefore, we get the following secular equation

$$\begin{bmatrix} k^2 - V_0 - E & V_g \\ V_g & (k-g)^2 - V_0 - E \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

For this equation to be satisfied, the determinant of the 2 × 2 matrix on the left needs to vanish, which yields the following equation relating k and E

$$E = -V_0 + \kappa^2 + \frac{g^2}{4} \pm \sqrt{g^2\kappa^2 + V_g^2}$$

where $\kappa^2 = g^2/4 - gk + k^2$. This equation gives the various energy eigenvalues, which needs to be understood with the corresponding wave function given by

$$\psi_k(z) \propto e^{i\kappa z} \cos\left(\frac{gz}{2} + \delta\right)$$

where δ is a phase factor that is determined by the boundary conditions at $z = 0$. Since this eigenfunction can be physical for complex κ, it is instructive to plot E versus κ^2 to determine the energy spectrum. As shown in Figure 1-11, for $V_g > 0$, κ^2 can take both positive and negative values. For positive κ^2, κ is real, and the solution corresponds to bulk states which satisfy Bloch's theorem. As κ^2 is decreased to zero, we can see that the top and bottom sheets of the $E(\kappa^2)$ function are separated by a gap of $2V_g$, which corresponds to the bulk energy gap. As κ^2 begins to

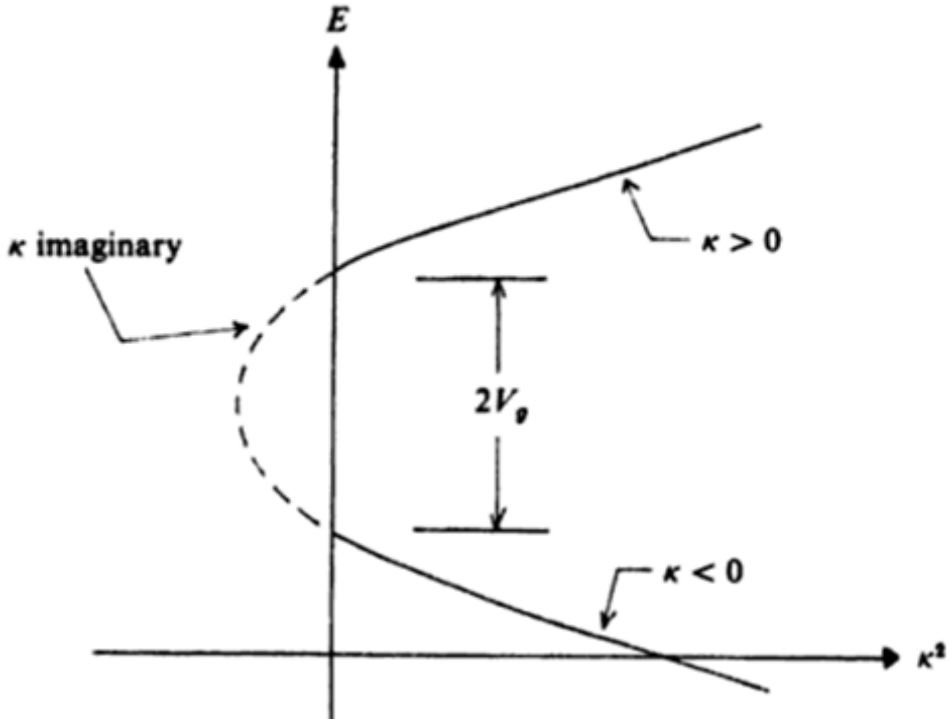


Figure 1-11 A plot of E vs. κ^2 gives insight into the physics of the wave function in the bulk and close to the surface. For $\kappa^2 > 0$, κ is real, and, therefore, the wave function is periodic, this is a bulk state. For $\kappa^2 < 0$, κ is imaginary, and the wave function decays into the bulk, and is therefore a surface states. (Adapted from [214]Zangwill, 1988)

take negative values, $E(\kappa^2)$ smoothly transitions into the gap; the most drastic physical change comes from exponential in the wave function. Since κ is now complex, these solutions then do not satisfy Bloch's theorem, and, therefore, they cannot be bulk states. This is directly seen because the now real exponential term forces the wave function to decay into the bulk (see Figure 1-10 dotted line). This means that the electrons that occupy these states are bound to the surface at $z = 0$, but are free to move parallel to the surface in the x-y direction. The simple model given above elegantly captures how the surface states emerge when the periodic bulk potential is truncated. However, it does not capture the variety of surface states that occur in many different materials ranging from metals and insulators to liquids, such as liquid helium ([51]Cole, 1974).

Surface states commonly occur in many different metals ranging from Cu and Au to W, and typically have a band width of the order of eV ([214]Zangwill, 1988). In metals, these surface states always occur in parallel with bulk states that are projected onto the surface. The main difference is that the bulk states disperse with k_z (often referred to as $k_{\perp)}$ and a true surface state will not. As shown in the projected band structure diagram in Figure 1-12, a true surface state will not be degenerate with a bulk state. When the surface band does overlap a bulk band, the surface and bulk states hybridize and the electrons disperse in k_z. This state is typically referred to as a surface-bulk resonance.

In insulators and semiconductors such as diamond, GaAs, and Si, surface states may also be present (in even numbers, as opposed to the odd numbers of surface states that occur in TIs, see 1.2). However, their character is quite different from those in metals. Since there will be no bulk state, insulators cannot host surface-bulk resonances; if a state exists on the surface it will be a true surface state. Typically, when a surface state does exist it usually resides at the Fermi level near the center of the band gap, and will often be associated with dangling bonds. The overlap of these states is very low, and, therefore, the surface band will be flat, which gives rise to a very large effective mass, and, therefore, a large density of states ($D_{2d}(E) \sim m^*$). This large density of states often causes surface pinning of the Fermi level because there is a large disparity in the

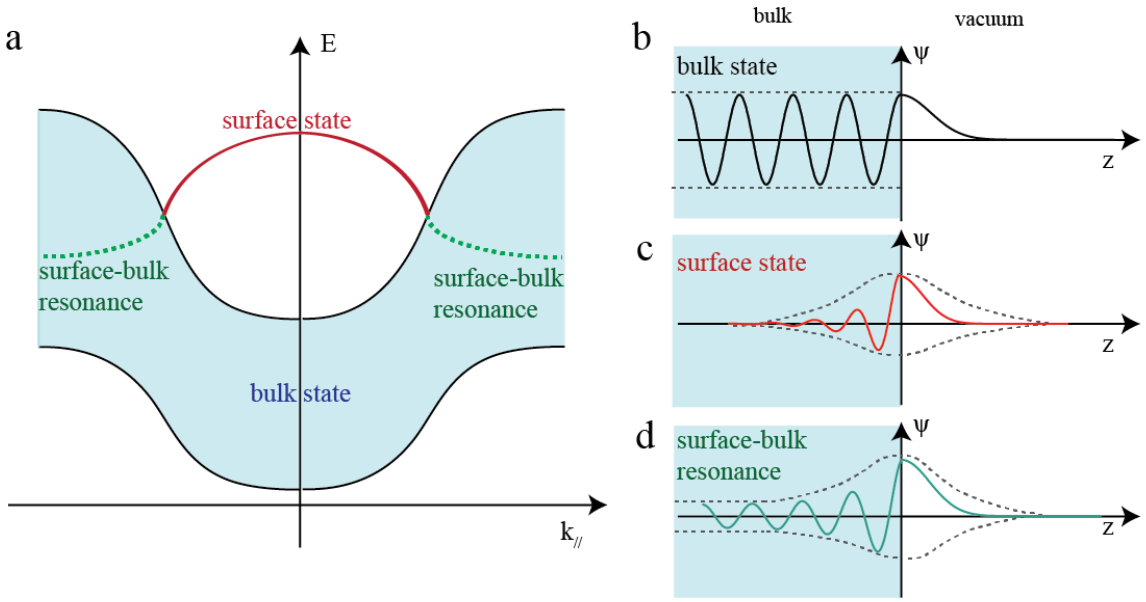


Figure 1-12 A comparison of surface states, bulk states and surface-bulk resonance states. a An idealized projected band diagram of E plotted versus $k_{//}$ for all values of k_z ($k_\perp$). The shaded region shows states that change with k_z, which are bulk states (**b**). The solid red line is a state that does not disperse with k_z, and is therefore a surface state (**c**). Lastly, the dotted overlapping region is a surface-bulk resonance state which is a hybrid of a bulk and a surface state (**d**). **b-d** shows the wave functions for each case in **a**. (Recreated from [190]Vanderbilt, 2013)

amount of doping that it takes to shift the Fermi level on surface versus the bulk, and, therefore, the Fermi level is stuck to this energy level on the surface; this often gives rise to large band-bending effects. Lastly, the robustness of these surface states to remain metallic is very low (as described in Figure 1-8a). Often these dangling bond states are only stable at high temperature, and are easily destroyed by surface contamination. Additionally, since the effective mass is very large, these states are also highly susceptible to localization effects that occur with a small amount of disorder (e.g. molecular adsorption, defects, etc).

1.4 Experimental realization

Although experimental evidence for novel surface states existed in materials that are known to be topological insulators, today ([185]Urazhdin *et al.*, 2004), it was not until 2007 that 2D TI theory ([19]Bernevig *et al.*, 2006;[20]Bernevig and Zhang, 2006;[156]Qi *et al.*, 2006) was confirmed ([117]König *et al.*, 2007). Later, measurements on $Bi_{1-x}Sb_x$ followed in 2008 that confirmed the theory ([75]Fu and

Kane, 2007;[76]Fu *et al.*, 2007) that showed 3D TIs existed with an odd number of spin-polarized Dirac surface states within the bulk energy gap by angle-resolved photo-emission spectroscopy (ARPES) ([94]Hsieh *et al.*, 2008), which was followed by the discovery of the canonical chalcogenide family of TIs by ARPES ([44]Chen *et al.*, 2009;[204]Xia *et al.*, 2009) and scanning tunneling spectroscopy (STM) ([165]Roushan *et al.*, 2009). Since the initial discoveries, many materials have been found to belong the TI class ([7]Ando, 2013), and it is also interesting to note the naturally occurring minerals like Kawazulite (approximate chemical composition $Bi_2(Te,Se)_2(Se,S)$) have been found to be a 3D topological insulator ([77]Gehring *et al.*, 2013).

1.4.1 2D TIs

2D TIs were first theoretically predicted in HgTe/CdTe heterostructures ([19]Bernevig *et al.*, 2006;[20]Bernevig and Zhang, 2006;[156]Qi *et al.*, 2006), and later they were found to exist when the thickness of the HgTe exceeded the critical value of 6.3 nm ([117]König *et al.*, 2007), which is the thickness where the bulk bands invert due to quantum confinement effects. As shown in Figure 1-13, for thicknesses below this critical value, the films are found to be insulating; when the thickness is increased just beyond 6.3 nm, then the conductance is found to be nearly quantized at $\sim 2e^2/h$ when the E_F is moved into the bulk gap. The quantized conductance, which is only expected in 2D TIs and not in 3D TIs, is expected to be $2e^2/h$ based on applying the Landauer Büttiker formalism with the expected properties for 1D TIs edge channels. Before we discuss the experimental realization of 3D TIs, it will be instructive to take a detour to discuss the ARPES technique before discussing further measurements.

1.4.2 Angle resolved photo-emission spectroscopy

ARPES is a surface sensitive technique that typically uses X-rays to photo-excite electrons out of the material. As shown in Figure 1-14, the ejected electrons are then sorted by energy and angle, which can be used to experimentally construct the energy-momentum relation. For an incoming photon with energy $\hbar\omega$ conservation of energy requires $E_{Kin} = \hbar\omega - E_B - \phi$, where E_{Kin} is the

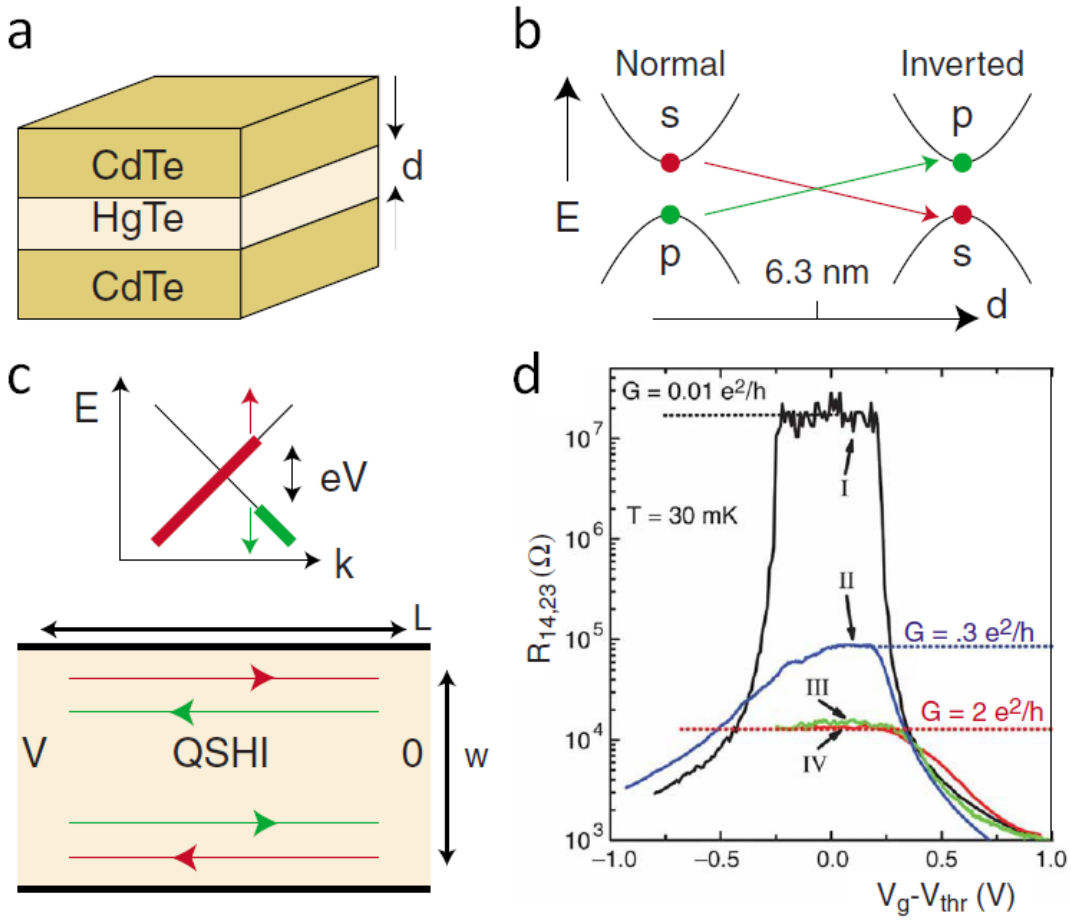


Figure 1-13 Experimental evidence for the 2D TI state. a A schematic of a 2D TI that is formed in a heterostructure of HgTe of thickness *d* grown on a CdTe substrate with a CdTe capping layer. **b** When *d* exceeds 6.3 nm, the conduction and valence bands invert, which pushes the HgTe into a topological state and spin polarized edge modes appear as shown in **c**. This is consistent with the experimental data shown in **d**, where the conductance becomes quantized when the thickness is increased 5.5 nm (I), to 7.3 nm (II-IV). (Adapted from [85]Hasan and Kane, 2010 with experimental data taken from [117]König *et al.*, 2007)

kinetic energy that the electron has once it has left the materials, ϕ is the electron work function that is unique to each material, and E_B is the binding energy. Conservation of momentum then requires $\hbar k_{initial,//} = \hbar k_{final,//} = \sqrt{2mE_{Kin}}\sin\theta$, and in the perpendicular direction $\hbar k_{\perp} = \sqrt{2m(E_{Kin}\cos^2\theta + V_0)}$, where V_0 is related to the electron work function, and is found by looking at the $\theta = 0$ electrons. By measuring the angle θ that the electrons are emitted at, and their final kinetic energy, one can then trace back to get their initial energies and momentum within the

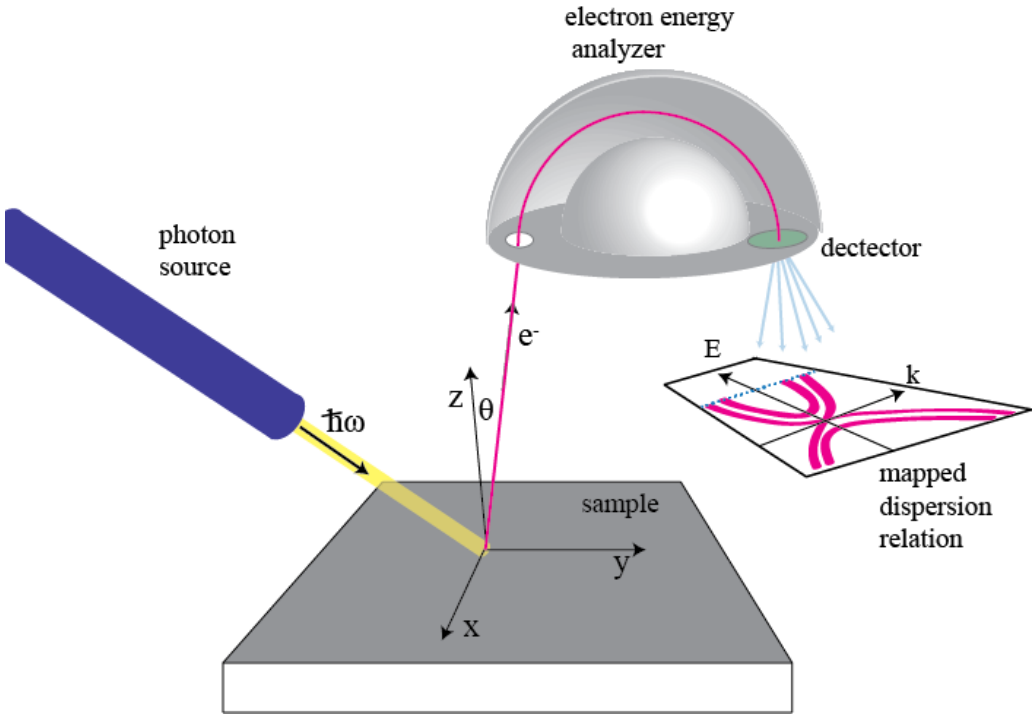


Figure 1-14 A schematic setup of a typical ARPES experiment. The photon source is typically a synchrotron X-ray source, which can continuously tune the wavelength, *hv*. The incoming X-ray ejects an electron, which is then measured, and from this, information about the initial state of the electron in the material can be gained.

crystal prior to excitation, and from this information the dispersion relation can be constructed. Further, by looking at $k_{\perp}$, which is often referred to as k_z, one can tell the dimensionality of the state. For a 3D electron the state should disperse with k_z, but for a 2D state, since it is in theory perfectly confined in the z direction, it should not change with k_z because, to first-order, $k_z \propto 1/z$, where z is the confinement thickness. ARPES provides a very straightforward means for determining the surface electronic properties of TIs, but also many other materials.

1.4.3 3D TIs

3D TIs were first discovered in $Bi_{1-x}Sb_x$ alloys that were being studied for applications in thermoelectrics ([80]Goldsmid, 2010). In fact, anecdotally, the researchers have said that the surface states were obscuring their measurements on the bulk bands, and it was the development of the TI theory that prompted them to looked more closely at the surface bands and realize the novel

properties. $Bi_{1-x}Sb_x$ showed that for $x \approx 10\%$ the bulk bands invert and five Dirac cones emerge at the surface, and disperse linearly with energy ([94]Hsieh *et al.*, 2008). Bi and Sb are both metals, and theory predicts the bulk is also metallic; this, in conjunction with the five Dirac cones on the surface, make this compound complicated to study, and quickly after the confirmation of the TI nature, new materials with better properties were discovered.

Using the theme of elements with heavy nuclei (those at the bottom of the periodic table), the chalcogenide family of TIs were experimentally ([44]Chen *et al.*, 2009;[204]Xia *et al.*, 2009) and theoretically ([216]Zhang *et al.*, 2009) discovered roughly simultaneously in 2009. Theoretically it was predicted that the SOC was large enough in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 to invert the energy bands, thus making them topologically non-trivial, while in Sb_2Se_3 the SOC was not large enough, and, therefore, it remained topologically trivial. The band structure was experimentally verified by ARPES measurements, which showed that unlike $Bi_{1-x}Sb_{x,}$ the surface band structure was much simpler, exhibiting only one Dirac cone per unit cell, and the bulk band gap for these materials was in the range of several hundred meV, which promised the possibility of room temperature TI effects without any bulk conduction.

Initially, Bi_2Se_3 and Bi_2Te_3 received the bulk of the interest due to their large gaps of ~ 300 and 165 meV ([44]Chen *et al.*, 2009;[204]Xia *et al.*, 2009). The position of the Dirac point in Bi_2Se_3 is more accessible than in Bi_2Te_3, which is buried in the bulk valence band, while the Dirac point in Bi_2Se_3 is nearly in the center of the bulk band gaps. However, these materials quickly showed that despite the large band gap, that the bulk states were populated by unavoidable defects, which contributed bulk carriers in parallel to the surface states. This caused problems with direct transport measurements, because the bulk state provides additional electrical pathways that are in parallel with the surface channels, which causes complications in analyzing transport data on these materials. To this end, understanding and solving this problem is a major theme of this work.

1.5 New topological materials

Since the initial demonstration of the topological nature of materials, much work has been done on searching for new topological phases that are driven by novel mechanisms. Unlike most of condensed matter physics (and more generally most of physics), which is driven mainly by experimental work that is followed by theory, most of the work in TIs is generally theoretically driven, where experiments follow. Theory work has spurred the experimental work on several burgeoning TIs materials including topological crystalline insulators, topological Kondo insulators, and most recently 3D Dirac materials.

1.5.1 Topological crystalline insulators

Topological crystalline insulators (TCI) are a new type of topological materials, and unlike the previously mentioned TI materials where an odd number of Dirac cones were protected by time-reversal symmetry, TCIs are protected by the mirror crystal symmetry. Another departure is that these materials are not predicted to have the required odd number of Dirac cones, but rather require an even number of Dirac states on the faces that are perpendicular to the mirror planes (Figure 1-15). The initial predictions by L. Fu ([74]Fu, 2011) and later by T. H. Hsieh *et al* ([96]Hsieh

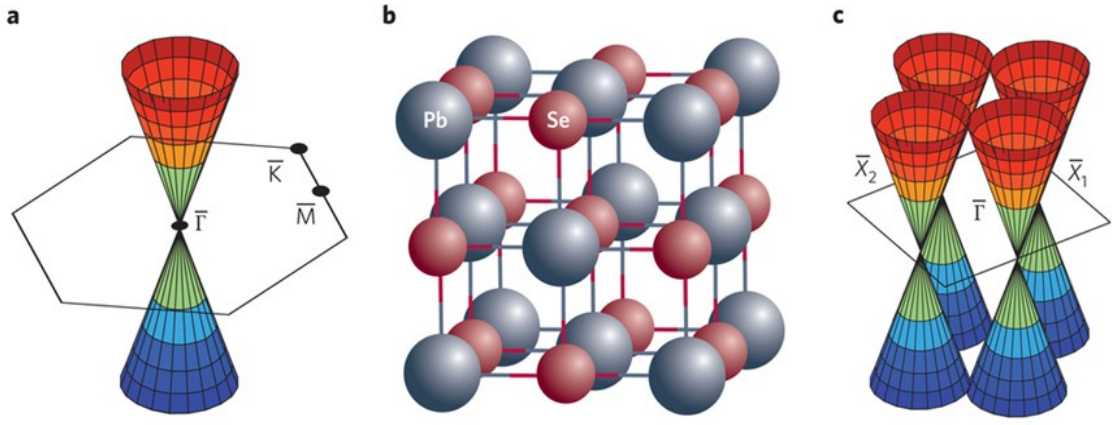


Figure 1-15 The topological crystal insulator state. a The surface dispersion relation for Bi_2Se_3, which contains one Dirac cone. For topological crystalline insulating such as PdSe shown in **b**, there exists four Dirac cones as shown in **c** by [63]Dziawa *et al.*, 2012. (Adapted from [69]Fiete, 2012)

et al., 2012) predicted that the SnTe, PbTe, and PbSe class of materials should be TCI, and experimental confirmation verification came in on $Pb_{1-x}Sn_xTe$ and SnTe in late 2012 ([63]Dziawa *et al.*, 2012;[179]Tanaka *et al.*, 2012;[207]Xu *et al.*, 2012;[145]Okada *et al.*, 2013). These experiments showed that with decreasing temperature, the lattice undergoes a change, and with this, the electronic states evolve into a topological phase that is in good agreement with theory.

1.5.2 Topological Kondo insulators

A long standing problem has been the unusual behavior of the electrical resistance versus temperature in the strongly correlated compound SmB_6 ([142]Nickerson *et al.*, 1971). Since then, SmB_6 has become a prototypical material associated with being Kondo insulators ([53]Coleman, 2007). Kondo insulators occur when the itinerate electrons hybridize with the local electrons, which opens a small gap at the Fermi level. This gap is usually of order of tens of meV, and, therefore, at low temperatures these materials show a weak divergence in the resistance versus temperature, which indicates the insulating behavior setting in. The resistivity of SmB_6 shows the requisite upturn in the resistance with decreasing temperature, but generally levels off, and is flat as T goes towards zero, which indicates a metallic state. This none-diverging behavior has long puzzled scientists, and in 2010 M. K. Dzero and coworkers pointed out that the strong correlations can drive this phase into a topological phase indicated by a non-trivial Z_2 index ([62]Dzero *et al.*, 2010). Then, according to the BBC theorem metallic surface states would necessarily emerge, which would electrically short the Kondo insulating bulk state at low temperatures, and would appear as the leveling off of the resistance versus temperature. With this new motivation, experimentalists have begun to look at the low energy physics in SmB_6 for signatures of the topological surface states ([107]Kim *et al.*, 2013;[122]Li *et al.*, 2013;[141]Neupane *et al.*, 2013;[182]Thomas *et al.*, 2013;[200]Wolgast *et al.*, 2013;[218]Zhang *et al.*, 2013). Unlike previously mentioned TIs, the in-gap states of SmB_6 are very flat and, therefore, are more difficult to resolve.

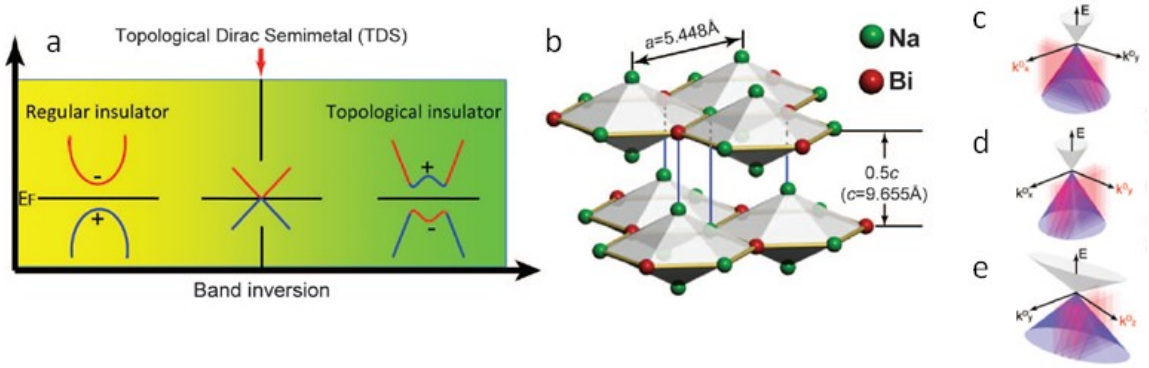


Figure 1-16 The topological Dirac semimetal phase. a A schematic of how a normal insulator transitions into a topological insulator. At the crossover point, there is a band touching that result in a topological Dirac semimetal state. Such a state exists in Na_3Bi whose crystal structure is shown in **b**. Z. K. Liu *et al* confirmed that the band structure of Na_3Bi disperses linearly in k for k_x, k_y, and k_z (as shown in **c**). For the 2D surface of a topological insulator, the Dirac state will disperses linearly in k for k_x, k_y, and will not change with k_z. (Adapted from [128]Liu *et al.*, 2014)

1.5.3 3D Dirac materials

Lastly, we will discuss a new class of materials where the bulk state disperses linearly like the Dirac surface states of a TI. As shown in Figure 1-16a, these so-called topological Dirac semimetals (TDS) or 3D Dirac metals, can be thought to emerge right at the crossover between normal insulators and TIs where the bulk valence and conduction bands touch at the point of band inversion ([197]Wang *et al.*, 2012;[213]Young *et al.*, 2012), and in this way TDS can be thought of as a 3D analogue of the 2D Dirac material graphene. Although it has been shown that one can carefully interpolate between a normal insulator and a TI (see Chapter 5, and [170]Sato *et al.*, 2011;[209]Xu *et al.*, 2011;[28]Brahlek *et al.*, 2012), chemically tuning to the exact point of band inversion would be difficult if not impossible. However, for certain stoichiometric compounds it has been shown that a 3D Dirac point can be protected by crystal symmetry, and, therefore, should be stable ([197]Wang *et al.*, 2012; Young, 2012 #238). As Figure 1-16 shows, the first observations of these materials came in Na_3Bi by Z. K. Liu *et al* ([128]Liu *et al.*, 2014), where ARPES measurements showed that the electrons near the surface disperse linearly in E for k_x and

k_y and most importantly k_z, which was quickly followed by experiments that confirmed that the compound Cd_3As_2 is also a TDS ([27]Borisenko *et al.*, 2014;[101]Jeon *et al.*, 2014;[127]Liu *et al.*, 2014).

In these TDS, an open question, which still remains to be experimentally realized in these materials, is what happens when inversion symmetry is broken ([82]Halász and Balents, 2012), or time-reversal symmetry is broken ([34]Burkov and Balents, 2011). If either of these symmetries is broken, then the 3D Dirac points of the parent materials are split into Weyl points, and a new topological phase emerges called a Weyl semimetal. This Weyl semimetal phase, whose ancestors first emerged in particle physics to describe neutrinos, has been predicted to host novel surface states called Fermi arcs ([194]Wan *et al.*, 2011), and could be a new platform to realize spin based devices ([221]Zhu and Hoffman, 2014).

1.6 Scope of this work

This work attempts to address many of the experiment questions regarding 3D topological insulators. On the materials side:

1. What are the key properties to synthesize high quality topological insulator thin films by molecular beam epitaxy?
2. Current topological materials are not true bulk-insulators; rather, due to intrinsic defects the Fermi level is pinned in the bulk bands, and results in a metallic bulk channel that allows electrical current to flow in parallel to the topological surface channels. Therefore, can a true bulk-insulating topological insulator be realized by suppressing these intrinsic defects during synthesis?

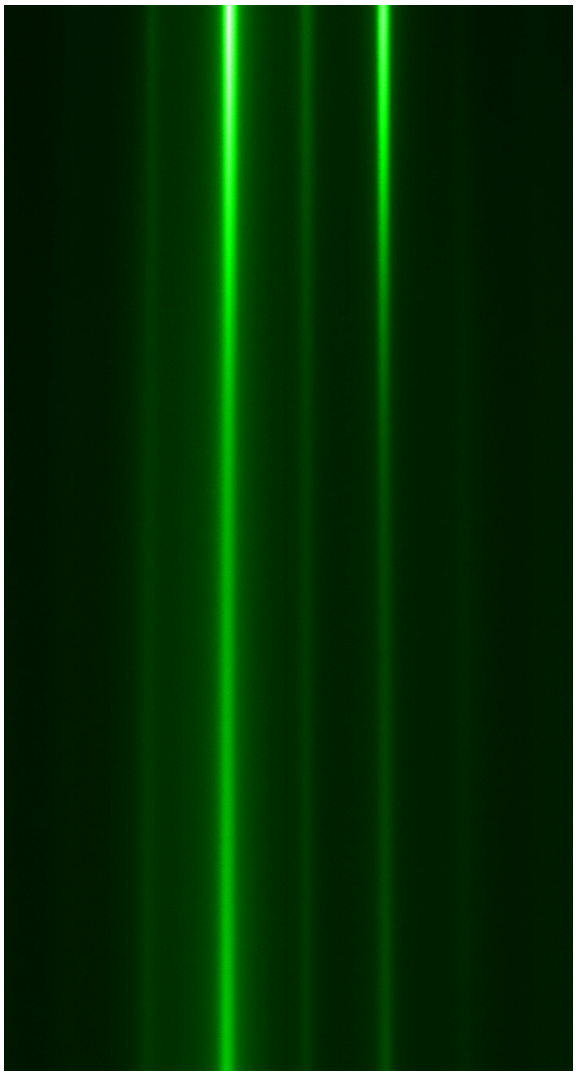
On the physics side:

3. What are the intrinsic transport signatures of the topological surface states?
4. How do materials become topological?

This work will be divided into two parts. The first part (Chapter 2-3) will deal mainly with experimental details regarding the growth by molecular beam epitaxy (Chapter 2), and transport measurements (Chapter 3). The remaining part will be dedicated to experimental results (Chapter 4-7). Finally, in Chapter 8 we will close with an overview and an outlook for future research in TIs.

Chapter 2.

Molecular beam epitaxy: growth of topological materials[2]



From the beginning of our lives, we develop an instinct for our physical world through passive experimentation. As physicists, we continue to refine this understanding of the world just as children do, but our play is carried out in a more controlled and methodical manner. Unlike the seemingly chaotic experimentation of children, our work hinges on the ability to produce an isolated physical system where the experimental parameters can be changed in a controllable way. This can range from giant accelerators for probing the fundamental interactions of particles, to the realm of solid state physics where complex electronic states are probed at the nanoscale. The latter typically relies on the synthesis of high quality crystalline samples composed of chemical elements that can span a near infinite number of possibilities.

[2] A RHEED image showing the expansion of the inplane lattice constant during the deposition of In_2Se_3 on Bi_2Se_3

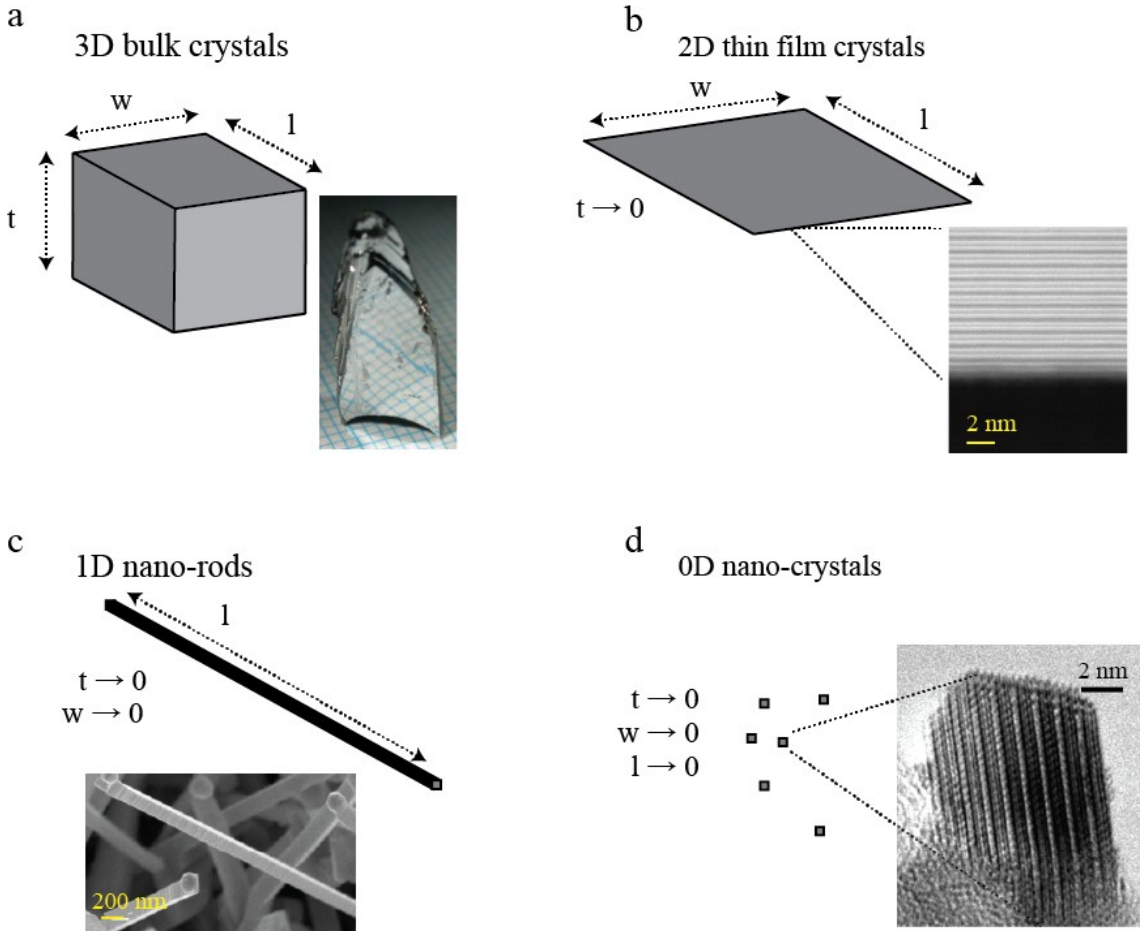


Figure 2-1 A comparison of the different crystal geometries that are synthesised for scientific studies, where examples are the topological material Bi_2Se_3. a Bulk crystals have a significant extent in all three dimensions (Image adapted from [97]Huang *et al.*, 2012), **b** Hypothetically thinning a bulk crystal to zero thickness results in a pseudo 2D thin film (Courtesy of M.-G. Han, and Y. Zhu). **c** Further thinning of the width results in a pseudo 1D nanorod (Image adapted from [220]Zhu *et al.*, 2013) **d** By reducing the length down, pseudo 0D nanocrystals are obtained (Image adapted from [159]Qiu *et al.*, 2004).

As shown in Figure 2-1, crystalline samples that are used in practical solid state experiments come in roughly four varieties: bulk crystals, thin films crystals, nanorods, and nanocrystals. Bulk crystals are large 3D objects that typically exceed millimeters in size; the commonplace example is a shimmering diamond that sits in a wedding ring, but, in contrast to common perceptions, even the gold that the diamond is set in has a well-defined crystal structure and should be thought of as a bulk crystal. Thin films can be thought of as bulk crystal with one dimension thinned down to the nanometer scale, thereby creating a planer, pseudo 2D crystal.

Extending this further, one can imagine reducing one of the dimensions of a thin film down to the nanometer scale, which results in pseudo 1D nanorods. Lastly, reducing the length of a nanorod down to the nanometer scale will result in a pseudo-0D nanocrystal. All of these crystal-types can display a huge variety of physical phenomena, and, therefore, represent a huge area of research in their own right. In this chapter, our main focus is on thin film crystals and their synthesis by molecular beam epitaxy (MBE).

We will first give a brief history of the various methods that have been developed to synthesize thin films, and continue with an introduction to, and a comparison of, the most common techniques used to produce atomically precise thin film samples. Following this, we will discuss the basic experimental techniques used to create and characterize thin films. Lastly, we will discuss the particular example of growing the chalcogenide family of topological insulators, which will be the focus of this thesis.

2.1 An abbreviated history of thin film deposition

In the first section, we will sketch out the brief history of thin film deposition, which will give a sense of how and why these techniques were developed. A full historical account would constitute a book in its own right, and, therefore, we will primarily cover the milestones that brought us to the current point; for a more extensive historical account see [148]Pashley, 1975 and [88]Henini, 2013.

Deposition of very thin layers of materials likely dates back thousands of years to the beating of gold to form an ultra-thin (sub-micron in thickness) foil, which was then physically pressed onto smooth glass to form mirrors (although gold foil can be used to coat or gild many different surface besides glass), and it is plausible that the mysterious Bagdad Battery may have been used for a rudimentary form of electroplating dating back to between 250 BC – 224 AD ([73]Frood, 2003). More modern deposition techniques have their beginning in the early 1800s with the advent of Alessandro Volta's voltaic pile, which was used, soon after its development, for

electro-plating. Later, vacuum-based physical vapor deposition techniques arose in the mid-1800s with Faraday's initial use of arced electricity to evaporate metal films (for complete timeline of physical vapor deposition see [133]Mattox, 2003). This was later built upon for the coating of glass for the use as mirrors, and in 1884, Edison filed for a patent for the arc deposition process. In 1887, the vacuum sublimation of Pt films was reported, which is the predecessor of MBE, and was followed by the first successful electron beam evaporator that was patented in 1907 by Pirani. Despite many technical advances over the next 60 years, the fundamentals of the deposition processes remained the same until the advent of pulsed laser deposition (PLD) in 1965, which was followed in the late 1960s by J. R. Author and Y. Cho, who led the initial work on developing the first modern realization of MBE ([9]Arthur, 1968;[46]Cho, 1969;[47]Cho, 1970;[48]Cho, 1971).

The early development of thin film deposition was focused on depositing thin layers of metals on various underlying materials, called a substrate, which did not play a direct role in the

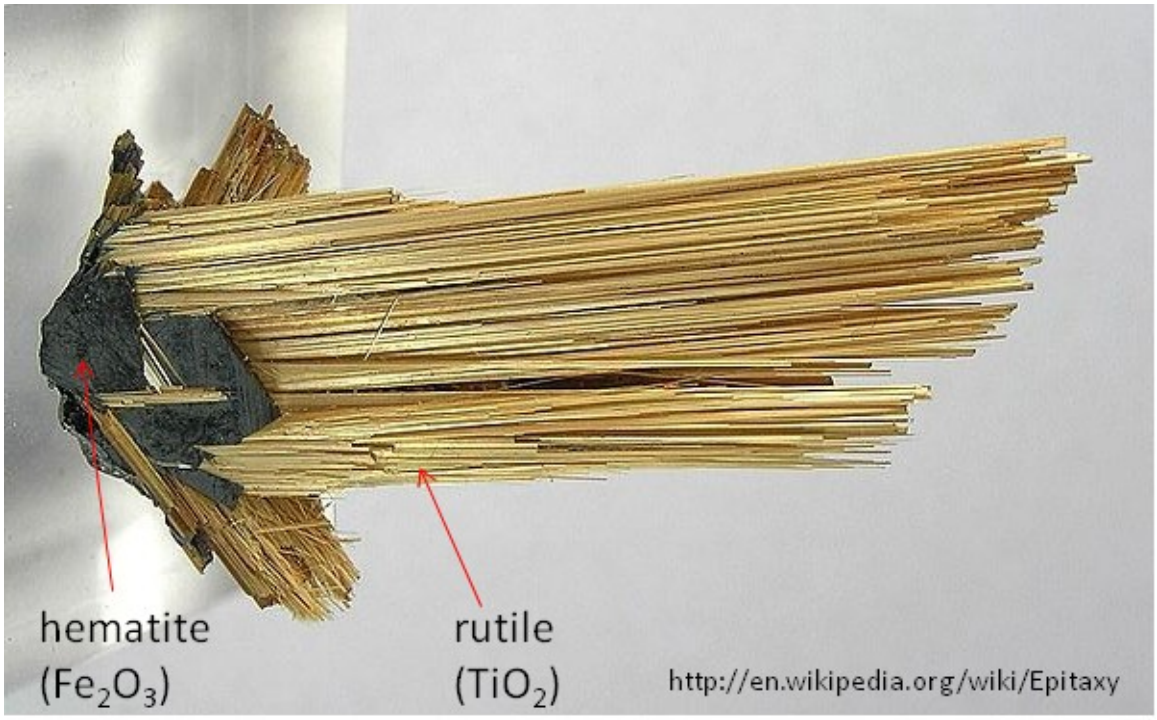


Figure 2-2 A naturally occurring example of epitaxial growth of rutile (TiO_2) on hematite (Fe_2O_3). (The entire crystal is over six inches in length.)

properties of the metallic over-layer; however, attempts to deposit thin layers that followed the underlying crystallinity of the substrate material were initially motivated by naturally occurring mineral deposits which contained dissimilar crystals grown together (see Figure 2-2 for naturally occurring epitaxial crystal of TiO_2 grown on Fe_2O_3). The first attempts to reproduce this effect synthetically were by M. L. Frankenhiem, who grew sodium nitrate on calcite by a solid solution process ([71]Frankenheim, 1836). His work was later built upon by T. V. Barker in 1906-8 who performed the first systematic studies of liquid phase growth by depositing various alkali halides on each other ([11]Baker, 1906;[12]Baker, 1907;[13]Baker, 1908); however, this effort was hampered due to the rudimentary liquid phase growth, and the optical microscopy, which was the only observation tool at the time. Remarkably, this effort predated X-ray diffraction, which, when developed, proved to be a better probe. The epitaxial growth of one crystal type on another was then pushed forward by the work of Royer in the 1920s, and it was Royer who introduced the term epitaxy, which is Latin, meaning "the arrangement on" referring to the arrangement of a new crystal on top of an existing crystal ([166]Royer, 1928;[167]Royer, 1935); Royer initially intended this to mean the arrangement of a dissimilar crystal, which would be referred to today as homoepitaxy, but epitaxy often refers to the growth of the same material on a crystal of identical species, which is called heteroepitaxy.

What ultimately allowed substantial progress in the growth of epitaxial thin films were the development of reliable high vacuum technology and the use of electron diffraction techniques. A reliable high vacuum is critical to the synthesis of epitaxial films because it minimizes the reaction of the film with the residual gases in the vacuum chamber. It was not until the 1960s when bakeable stainless steel vacuum chambers, with high performance pumping systems, particular ion and turbo-molecular pumps (often called turbo-pump), were widely available, which allowed reliable operation at the ultra-high vacuum with pressures less than 10^{-8} Torr. The high vacuum environment is also critical to the creation and stability of the electron beams that are used in monitoring the surface structure during deposition.

Electrons have a distinct advantage over X-rays in probing the structural properties because the energy of the electrons can be tuned to form a diffraction pattern solely from the first few layers of atoms on the surface, and thereby, probing the structural properties of the deposited over-layer without being affected by the underlying substrate (see 2.7.1). Electron diffraction was first demonstrated by E. W. Davisson and L. H. Germer ([58]Davisson and Germer, 1927) using electrons with ~100 eV of energy incident on a Nickel target. They showed that a diffraction pattern is formed due to the combination of the wave nature of the electrons on the periodic atomic structure of the Ni crystal; for this work, they were awarded the 1934 Nobel Prize in Physics. During 1930-50s, electron diffraction gave much needed experimental insight into how the growth parameters, such as rate, substrate temperature and the residual gases affect the growth; all of which demonstrated the possibilities and power of the epitaxial growth technique. High performance transmission electron microscopy (TEM) systems began to come online in the mid-1950s. Through both realspace imagining, as well as diffraction images, TEM revealed the role of kinetics in the growth process, as well as, providing critical experimental data concerning defect structures and their origins.

Prior to late 1950s and the early 1960s, studies of epitaxially deposited films were primarily focused on metals. However, driven by the technological applicability of semiconductors, people began exploring epitaxial growth of various semiconductor materials. This motivated the work of J. R. Author and Y. Cho, who successfully grew GaAs films in the first modern molecular beam epitaxy system operating at ultra-high vacuum and equipped with reflection high energy electron diffraction (RHEED) that was capable of monitoring the growth in real time ([9]Arthur, 1968;[46]Cho, 1969;[47]Cho, 1970;[48]Cho, 1971). During the following decades, many advances in deposition technology were driven to improve transport measurements on epitaxial grown GaAs.

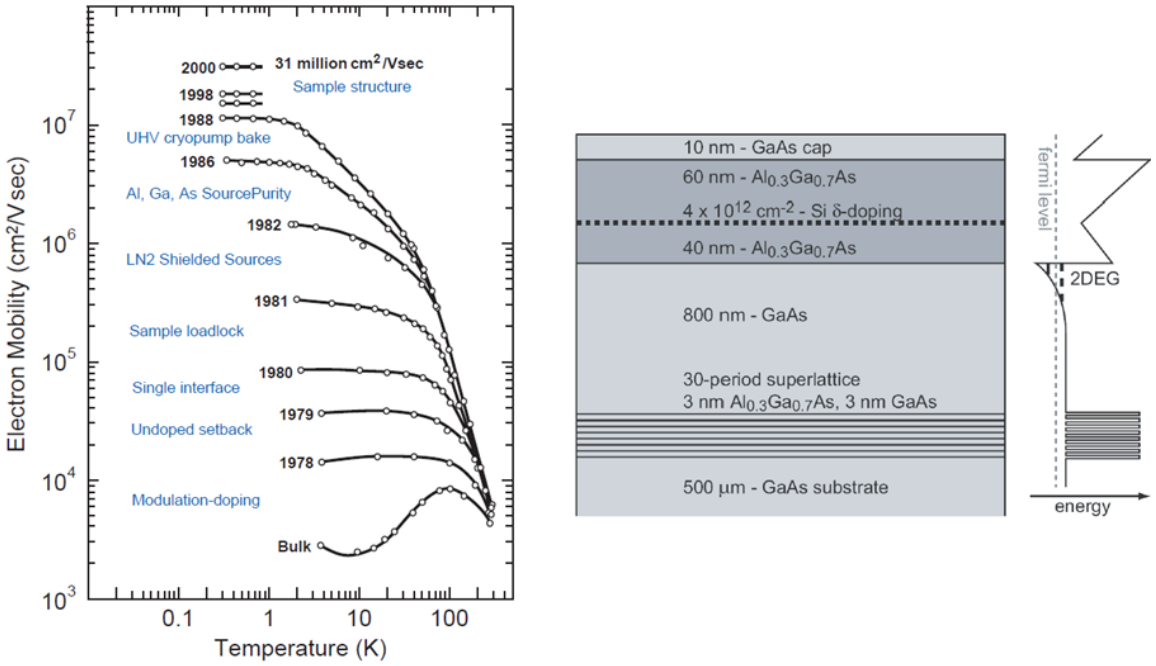


Figure 2-3 The increases in the electron mobility of GaAs over time. The main improvements were due to the use of modulation doping, schematically shown on the right, in conjunction with better vacuum techniques, and improved material qualities of the GaAs substrates and the source materials. (Adapted from [153]Pfeiffer and West, 2003)

As shown in Figure 2-3, the electron mobility for bulk GaAs crystals is limited to approximately ~10^4 cm^2/Vs, and through the use of MBE, the mobilities have steadily risen to exceed ~30×10^6 cm^2/Vs, which is the highest known electron mobility for any system. Macroscopically, the electron mobility is a measure of how much conductance each electron carriers, or in a microscopic picture, as exemplified by the Drude model, the electron mobility is a measure of how the electrons' motion is disrupted, or scattered, by crystal defects (see Chapter 3). The electron mobility does not provide detailed information as to defect type, but it is the most sensitive means to measure the relative concentration of defects.

The increase in the electron mobility, shown in Figure 2-3, is the result of the control of defects through the epitaxial growth technique. In semiconductors, the origin of carriers is chemical doping by adding foreign atoms with valence states that are different from the host atoms. These foreign atoms will act as defects, which degrade the overall mobility. The key

aspect to increasing the mobility of GaAs thin films was found by creating a heteroepitaxial structure, called a heterostructure, formed by GaAs and GaAlAs. As shown in Figure 2-3, at the interface between these heterogeneous materials the energy bands align in such a way to create a depression in the potential energy landscape, which is called a quantum well. This, then, acts to attract electrons away from the region of chemical doping, towards the depression where they are trapped, and no longer encounter these defects. This is the key to increasing the mobility in these quantum well structures, which is referred to as modulation doping. With the electrons no longer seeing the chemical defects, they can flow extraordinarily long distances, in excess of ~10 μm or ~50,000 atomic sites, without scattering (see [33]Brill and Heiblum, 1996).

The initial development of modulation doping increased the mobility by less than an order of magnitude, and was mainly limited by the intrinsic defects that occurred in the quantum well. This was found to be due to chemical defects resulting from residual gas in the vacuum chamber, outgassing of vacuum components, outgassing of chemical impurities from the substrate, and chemical contaminates in the individual source elements. Over the next several decades, these were improved by better vacuum techniques including pumping and baking schemes that pushed the vacuum to below 10^{-11}-10^{-12} Torr, and better quality substrates and elemental materials (see [68]Farrow, 1995). The modulation doping technique in GaAs showed that thin film materials can achieve results that are not achievable in bulk crystals. The ultra-high mobility electrons that resulted from these quantum wells structures led to the discovery of the quantum hall state ([111]Klitzing *et al.*, 1980) and later the fractional quantum hall state ([184]Tsui *et al.*, 1982), which, as mentioned in Chapter 1, are the first known topological state of matter, and these discoveries went on to win the 1985, and 1998 Nobel Prizes in Physics.

In closing, thin films technology impacts our daily lives in many ways from overlooked technologies like anti-reflective coating on our glasses and windows, to wear resistance surface coatings on mechanical parts that vastly extend the usable lifespans in extreme and corrosive environments, thus enabling reliable technologies like the modern gas turbine engines for jet

travel, and many critical components in nuclear reactors among many others ([72]Freund and Suresh, 2003). Thin film coatings have a great impact in medical science and are key components in many surgical devices and advanced imagining technologies. Probably one of the most ubiquitous and key places thin films appear are in electronic devices. Simply put, without thin films technologies the modern microprocessor would not be possible, and consequently, any electronic device used in our modern lives. To put the technological importance of thin film technology in perspective, the United States National Academy of Engineering has compiled a list of the 20 most significant achievements of the 20th century, and thin film technology is a key to the modern realization of virtually all the achievements on that list (see [54]Constable and Somerville, 2003).

2.2 Growth processes

As shown in the previous section, the process of depositing one material on another material has been extensively studied for hundreds of years. This involves many physical and chemical processes, the study of which constitutes an entire field of research in their own right. In this section we will introduce the basic physics and chemistry involved in the deposition of materials on a substrate, and how these materials behave once deposited (together, these processes are loosely referred to as the growth process). Understanding the growth process opens a huge number of routes to designing functionalities into materials and makes possible the ultimate goal of this work, which is the ability to tailor electronic properties at the atomic level, as well as, having a broad potential for future scientific research, and technological applications.

2.2.1 Thermodynamics and kinetics

Extending back to the early work of Langmuir, the interaction between the substrate *S*, and the adsorbing atoms *A*, should be thought of as a class of chemical reaction exemplified by $A + S \rightarrow AS$. Therefore, as in the general process of any chemical reaction, the process of thin film deposition involves both thermodynamic and kinetin processes. Thermodynamics tells what

equilibrium state is favored, while the kinetic conditions affect how the reaction proceeds and what equilibrium configurations are accessible. Controlling these aspects is crucial to being able to produce thin films with specific desired qualities. For a preferred reaction to take place (the preferred reaction is always dependant on the desired outcome—success for one situation may be failure for another) the Gibbs free energy must decrease. The change in Gibbs free energy is given by $\Delta G = \Delta H - T\Delta S$, where H is the enthalpy, T is the temperature and S is the entropy. If $\Delta G > 0$ the reaction is prohibited under ambient conditions, requiring additional energy to proceed, and for $\Delta G < 0$ the reaction is permissible; lastly, if $\Delta G = 0$, then the system is in equilibrium.

To illustrate this, we can take the example of depositing a metal layer on a SiO_2 substrate. Depositing Cu or Al on SiO_2 produces different results, which are dictated by the free energy of the following reactions: $4/3Al + O_2 \rightarrow 2/3Al_2O_3$, $Si + O_2 \rightarrow SiO_2$, and $2Cu + O_2 \rightarrow 2CuO$, which have changes in free energy ΔG_{AlO}, ΔG_{SiO}, and ΔG_{CuO} where $\Delta G_{AlO} < \Delta G_{SiO} < \Delta G_{CuO}$. Then, if depositing Cu or Al on a SiO_2 substrate we must therefore consider the following reduction reactions $4/3Al + SiO_2 \rightarrow 2/3Al_2O_3 + Si$, and $2Cu + SiO_2 \rightarrow 2CuO + Si$, which would indicate that the Al and Cu have reacted with the substrate and reduced the O from the Si. The change in free energy of the first reaction is negative, while the second is positive; this indicates that Al will tend to pull O out of the substrate and form insulating Al_2O_3, while Cu will form metallic Cu on the surface leaving the SiO_2 unchanged. How fast these reactions occur depends on the kinetic conditions; at high temperatures the reaction will likely proceed, but it is possible that at a sufficiently low temperature, there will be insufficient thermal interaction between the Al and SiO_2 for the reaction to go forward. This relatively simple example illustrates the complex interplay between theromodynamics and kinetics during the growth of thin films, and as we will see later, this will be important for the growth of topological insulator materials.

2.2.2 Physical and chemical adsorption

The growth of thin films involves the creation of molecular beams that condense onto a substrate (see the following section for more information on the various methods for achieving elemental or molecular beams). How the elements adsorb onto the substrate involves weak physical adsorption (simply called physisorption), prior to the crystallization provided by chemical bonding or chemical adsorption (shortened to chemisorption). At the surface of a substrate, there is a weak potential-well due to attractive Van der Waals forces, often modeled by the 6-12 Lennard-Jones potential $U\ (r) = 4\epsilon((\sigma/r)^{12} - (\sigma/r)^{6})$, where r is the distance from the surface, ϵ is the potential well depth, and σ is the distance where the potential is zero (see Figure 2-4a), and the binding energies of physisorption are typically in the range of ~10-100 meV. During deposition, atoms are sent toward the substrate and are initially trapped at the surface by this weak attractive force. For this to occur the incoming atom needs to have a low enough energy that

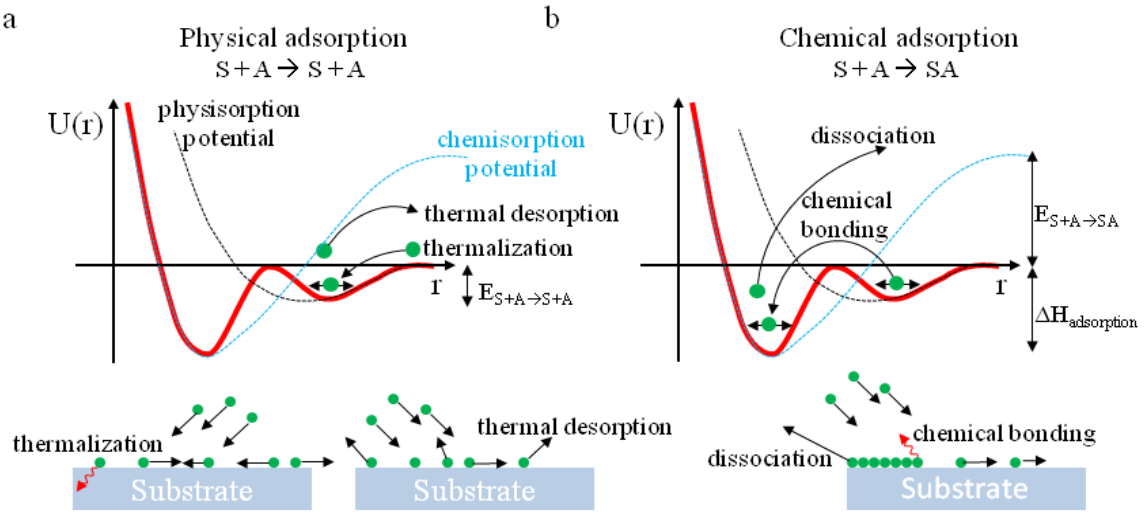


Figure 2-4 This schematic depicts the physical and chemical processes of binding an element *A* to a substrate *S*. a The process of physical adsorption of atoms on the surface of a crystalline substrate. Atoms that are adsorbed onto the surface can move in the plane of the substrate due to thermal energy, as shown on the left. While, if the temperature is too high, the atoms can be ejected off the surface, in a process of thermal desorption, as shown on the right. **b** The process of physical adsorption gives way to chemical adsorption as the atoms are fixed more permanently in their crystal locations.

can be transferred inelastically to the substrate in a process of thermal accommodation referred to as thermalization; else, the atom will recoil back into the vacuum. Once physisorption does occur, if the substrate temperature is too high, then the atoms will be desorbed off the substrate, and sent back into vacuum, and, obliviously, growth cannot occur. In contrast, the substrate temperature needs to be sufficiently high so that the atoms that are adsorbed onto the substrate have enough energy to remain mobile, so they can diffuse around and find the optimal place to chemically adsorb. The physisorption to chemisorption process usually occurs at a nucleation point, typically a point of stronger confinement, such as a crystalline step edge. This is a much stronger adsorption process with a typical binding energy in range of 1-10 eV (see Figure 2-4b and Figure 2-5).

It should be noted that the actual form of this potential energy curve depends on many factors, and is more complex than the simple Lennard-Jones potential and the schematic version shown in Figure 2-4; these factors can range from local geometry, for example a defect or step

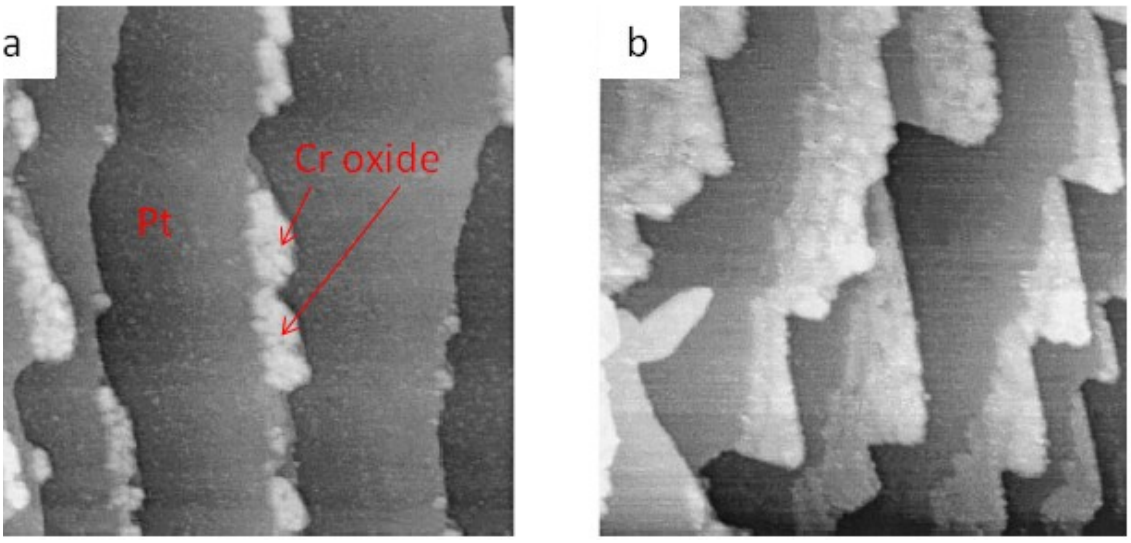


Figure 2-5 A scanning tunnelling microscopy image of the growth of Cr oxide on a Pt(111) surface. Images taken at 8% (a) and 40% (b) of a monolayer**.** This is a typical example of how the growth of epitaxial thin films proceed from the step edges. (Adapted from [38]Chambers, 2000)

edge, to a different chemical structure, or even the presence of other mobile atomic species. An example of the latter that is critical to TIs is the formation of Bi_2Se_3 by co-deposition of Bi and Se, both of which tend to physisorb, but quickly desorb (no potential minimum in the chemisorption curve) if alone. However, in the presence of the other species a minimum in the chemisorption potential develops corresponding to the formation of Bi_2Se_3. For a more in depth discussion of the physisorption and chemisorption processes see chapter 8-9 of [214]Zangwill, 1988.

2.2.3 Accumulation and growth modes

Once a material adsorbs onto a substrate and begins to accumulate, the atoms can arrange in one of the several ways shown in Figure 2-6. Under optimized conditions the adsorbed atoms can form a single crystalline phase where the crystal structure is the same across the entire substrate (see Figure 2-6a); this generally requires the deposition rate and temperature of the substrate to be optimized based on the specific type of material being deposited, and typically requires the temperature to be high enough to maximize surface mobility, while low enough to prevent thermal desorption. Additionally, the growth rate needs to be slow enough to give the adsorbed

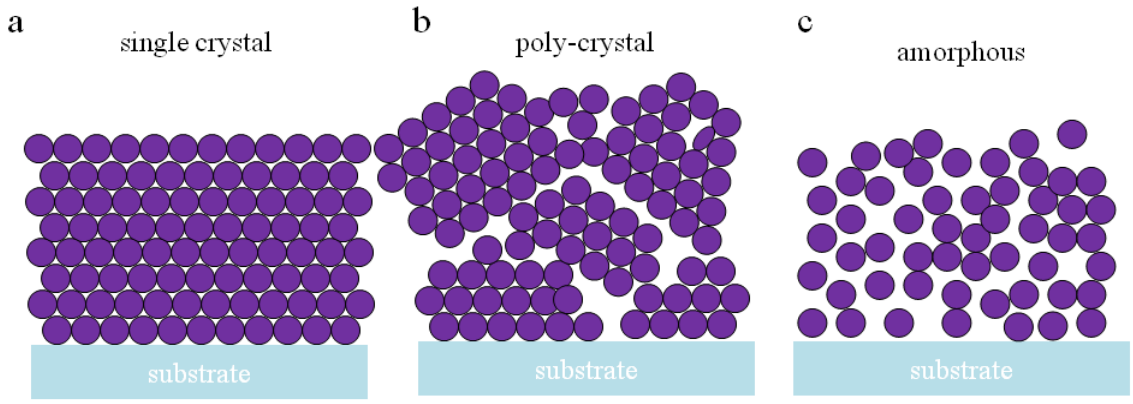


Figure 2-6 The possible structures a material can take when accumulated on a substrate. If the atoms arrange coherently in a lattice that extends across the entire crystal, then it is referred to as single crystal (**a**). If the crystal structure forms small randomly oriented domains, then it is called poly-crystalline (**b**). If the atoms arrange with no order then it is called amorphous (**c**).

atoms sufficient time to find their crystal location prior to being enveloped by the next layer of atoms; depending on the material and temperature, this is typically at least several seconds per mono-layer. The terminating surface of the substrate is critical because it acts as the seed crystal for the thin film to orient itself upon. If the quality of the substrate is low, or if the substrate is dirty, this is transferred into the film, and it will also be of low quality. The main goal in this work is to optimize the growth condition and achieve single crystal films of topological insulator materials (see 2.8).

Polycrystalline arrangement occurs when the film is composed of many small crystalline domains with random orientation. One of the many reasons this may occur is that the crystal structure of the substrate is not of the same symmetry as the film. This is the case when the substrate has 4-fold cubic symmetry and the film to be grown has 6-fold hexagonal symmetry; for example, when Bi_2Se_3, which has 6-fold symmetry, is grown on a (001) orientated $SrTiO_3$ substrate which has 4-fold symmetry (see 2.8.4). Another common reason for polycrystalline growth is if the substrate conditions are not optimized, and the film is polymorphic (meaning multiple crystal structures are available under a given set of conditions); this is case for growing In_2Se_3 on Al_2O_3, where the crystal symmetry is the same, but with a large lattice mismatch (see 2.8.5), which results in the formation of In_2Se_3 with differing crystal structures, different phases, or many randomly orientated domains.

Lastly, if the randomness extends down to the atomic level, and there is no order, then the structure is called amorphous. This type of structure often occurs when the growth temperature is too low, and the adsorbed atoms cannot diffuse around to find the optimal crystal positions prior to the next layer of atoms being adsorbed; roughly speaking, to achieve an amorphous film the diffusion time needs to be significantly longer than the deposition rate (inverse rate to be more precise). In the case of Se, an amorphous structure is relatively easy to achieve when the substrate is at room temperature, while for many metals, achieving an amorphous structure requires extremely fast deposition rates at cryogenic temperatures ([144]Ohring, 2002). In contrast, many

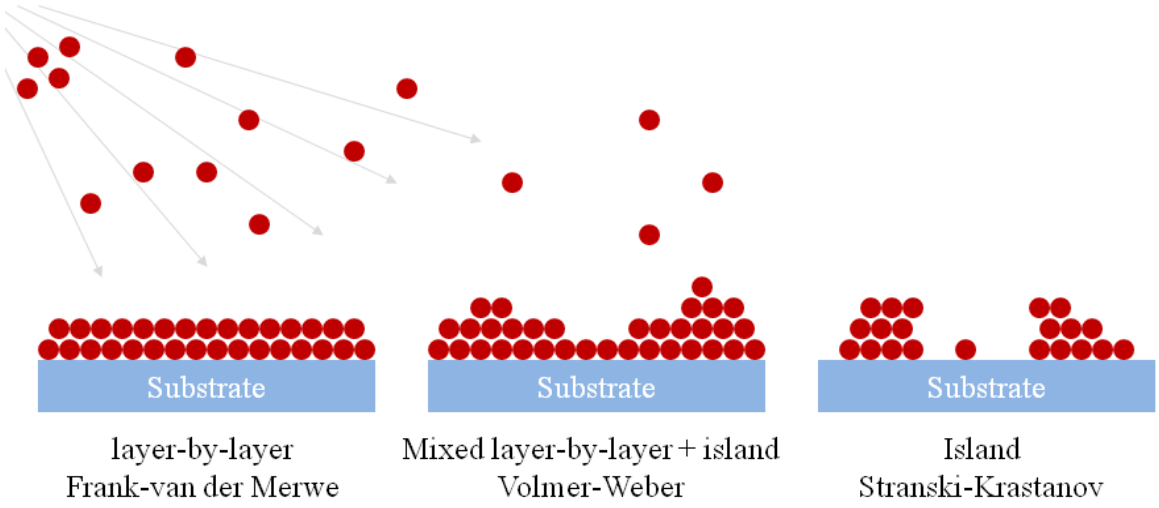


Figure 2-7 A schematic of the possible growth modes when depositing a film on a substrate. Left: shows the layer-by-layer growth mode, where each subsequent atomic layer begins once the previous layer is completed. Right: island growth occurs when the atoms clump up on the surface. In reality, growth typically is somewhere in between layer-by-layer and island growth mode as shown in the middle.

materials spontaneously form an amorphous structure, as is the case with Si and Al in an oxygen environment.

The growth process can proceed in a number of topographical ways. As shown in Figure 2-7, for the example of single crystals, on one end is the 2D atomic-layer by atomic-layer (shortened to layer-by-layer) growth mode, where the adsorbed atoms complete a single uniform layer before starting the subsequent layer; this growth mode is often referred to as the Frank-van der Merwe growth mode. On the opposite end, the adsorbed atoms coalesce around nucleation points and form 3D multi layered islands, which is often referred to as the Volmer-Weber growth mode. In this mode, full coverage is achieved only when neighboring islands grow large enough and merge together. Generally, the growth progresses as some mixture of layer-by-layer and island growth, where the coverage is uniform, but with some amount of roughness to the surface; this is referred to as the Stranski-Krastanov mode. The growth mode that is favored is determined by a large number of thermodynamic and kinetic factors including the relative substrate and film chemistry, the substrate's surface morphology and crystal structure, deposition rate, deposition method and substrate temperature.

The structural compatibility of the substrate crystal and the film to be grown plays a large role in the growth of thin films. The key properties include the crystal symmetry of the surface of the substrate and if it is compatible with the surface symmetry of the film; also, to a lesser extent the bulk symmetry can also play a role in the growth, and the subsequent electronic properties, because a film's properties can change depending on if the bulk structure is continued into the film ([164]Rondinelli *et al.*, 2012). Another factor is the lattice spacing of the substrate, as compared to the lattice spacing of the film. If the lattice spacing of the substrate is larger, then the film will need to be stretched, or tensile strained, whereas if it is too small, then the film will need to be compressively strained to grow epitaxially. The level of strain is quantifiable by the lattice mismatch $f = (a_{film} - a_{s)} / a_s$, where a_{film} is the in-plane bulk lattice constant of the film to be grown, and a_S is the substrate lattice constant. If this strain is small, then it is possible for the film to grow in a layer-by-layer mode because the energetic cost of staining the film is small enough for the surface tension energetics to dominate. In contrast, if the mismatch is too large, then the strain energy will dominate over the surface tension energy required to form a layer-by-layer film, and may result in island-like growth where the film grows off of these nucleation points and takes a more energetically favorable structure rather than straining to that of the substrate. Lastly, if the film is initially strained and layer-by-layer growth is achieved, the strain can only be maintained below a critical thickness. This is because there is a competition between the energy required to maintain the strain, and the energy cost of alleviating the strain through the formation of defects (generally vertical cracks or edge dislocations).

2.2.4 Defects structures

In ubiquitous existence, crystallographic defects play a major role in the properties of thin films, and by controlling the growth parameters, defects can be minimized, or in certain situations, if desired, defects can be induced. Defects broadly fall into three classes: point defects, line defects, and planar defects. Point defects occur on the atomic level, mainly due to the inclusion or the

absences of a single atom, and as shown in Figure 2-8a, which can be a missing atom in the crystal structure, a foreign atom that sits on an atomic site, or an atom, foreign or native, that sits interstitially in between the atomic sites. Point defects are localized, and, therefore, can be thought of as 0D.

Line defects, however, occur as the name suggests, along a 1D line. As shown in Figure 2-8b, an example of a line defect is an edge dislocation, which, as shown, extends perpendicular to the page, where an extra row of atoms is inserted into the crystal lattice. These types of defects can occur at the heterointerface during growth to alleviate strain at the interface. As shown in

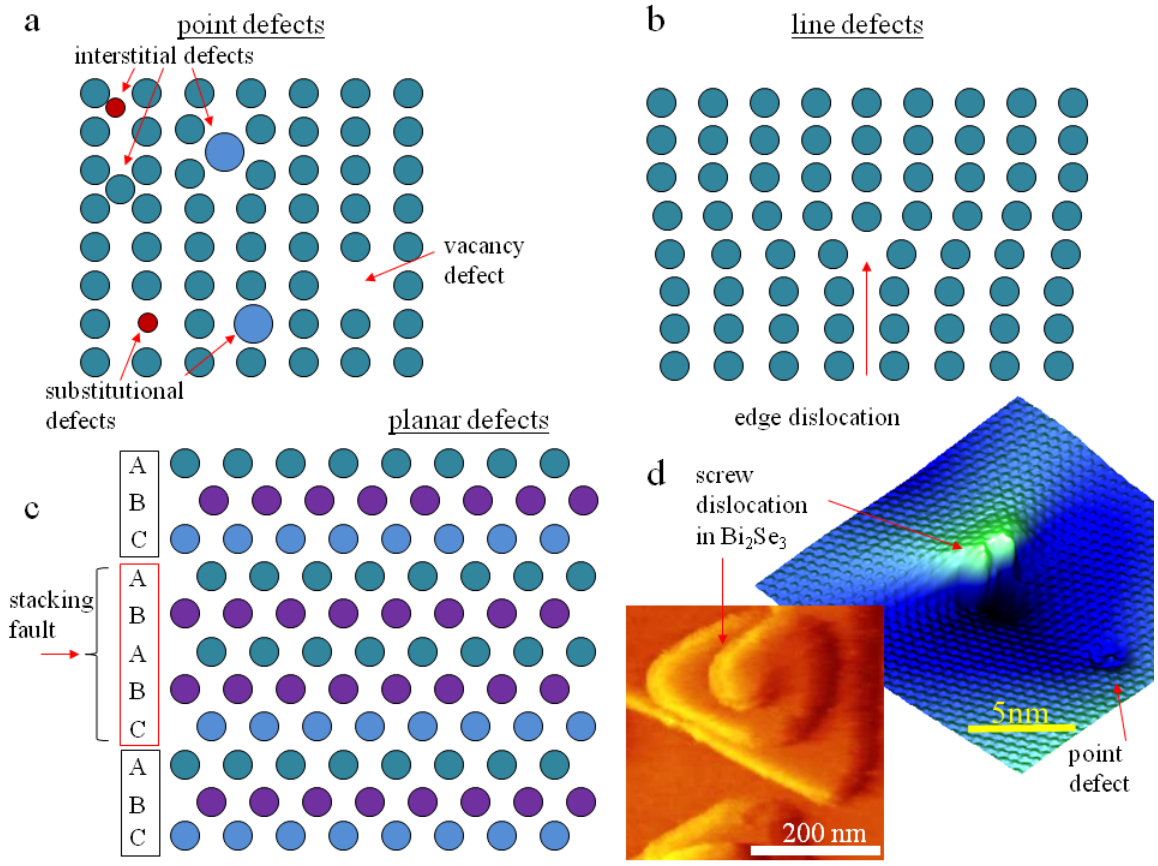


Figure 2-8 A schematic showing examples of the three classes of crystallographic defects. Several examples of point defects are shown in **a**, while **b** and **c** show 2D cuts of line and planar defects, respectively (the crystal structure in **b** and **c** extend perpendicular into the page). **b** An edge dislocation where an extra column of atoms is grown into the crystal structure. **c** A stacking fault where the stacking order of the atomic planes, ABC-ABC-ABC …, is interrupted with the inclusion of an extra A layer in the form of ABC-ABABC-ABC… **d** An example of a point and line defect in Bi_2Se_3 (Image courtesy of Jixia Dai and Weida Wu).

Figure 2-8d and Figure 2-30, a second example that occurs in Bi_2Se_3 is a screw dislocation, which forms when the surface layer grows back around and onto itself forming a spiral-like structure. Away from the axis the crystal structure is normal, but is interrupted along the 1D axis.

Lastly, a planar defect, which is 2D, and can occur at a grain boundary in a poly-crystalline sample (Figure 2-6b) or where the stacking order of a crystal is interrupted; this is shown in Figure 2-8c by a schematic of a hexagonal close-packed structure whose lowest energy stacking order is ABC-ABC-ABC… where A, B, and C represent identical layers that are rotated 120° around the vertical to form the close-packed structure. A stacking fault occurs when this ABC order is interrupted and a layer is formed with ABA, BAB, BCB, CBC, ACA or CAC (Figure 2-8c is shown with ABA). Such defects occur in Bi_2Se_3, and can be seen in the transmission electron microscopy image shown in Figure 2-34b. While this is not an exhaustive list of the defect types, these particular examples have been chosen because they are of particular importance in the growth of topological materials, and these defects, and others, will be discussed in more detail, later.

2.3 Thin film growth techniques

Most techniques used for thin film deposition fall somewhere within the two broad classes of either physical vapor deposition (PVD) or chemical vapor deposition (CVD), but there does exist PVD-CVD hybrid techniques that have been used to improve film quality (see chapter 17 of [88]Henini, 2013). In PVD, the material to be deposited is ejected from a target by physical means, whether thermal evaporation or expulsion by some external bombardment. In contrast, CVD generally employees a carrier gas that contains the elements required for the desired material; through a chemical reaction in the vicinity of the substrate, the desired elements are freed up and deposed as a film, with the leftover components remaining in the gas phase, which are then pumped away. This work involves the use of PVD processes, so in this section we will first give a

basic description of dominate PVD processes of sputtering, pulsed laser deposition, and molecular beam epitaxy (MBE), and contrast the strengths of each technique.

2.3.1 Sputtering

Sputtering is a PVD process where material is ejected by means of bombardment of the target material by ionized atoms in the form of a plasma (see [144]Ohring, 2002 for a more thorough introduction). With a negative external voltage bias applied to the target, the positively charged ions are then accelerated towards the surface of the target, and upon impact, particulates (from individual atoms to larger clusters of atoms) are ejected from the target in the direction of the substrate. The gas that is used to form the plasma is typically highly purified argon, due to its inert nature, abundance, and large mass; in reactive sputtering the plasma can be composed of argon mixed with a reactive gas such as oxygen, nitrogen or sulfur, which is present to react with, and alter, the chemical composition of the film. Sputtering is capable of ablating virtually any material that can be formed into a suitable target because it is possible to maintain the plasma under a direct current (DC) bias for metallic targets, and under an alternating current (AC, typically in the radio frequency (RF) range) for insulating targets.

The dominate form of sputter deposition used is magnetron sputtering, which employs a magnetic field to confine the charged plasma near the target (see Figure 2-9). The confining field is created by strong neodymium magnets, and allows the minimum background pressure of argon to be much lower; typically in the range of 10^{-3} Torr compared to 10^{-2} Torr for non-magnetron sputtering. The lower working pressure increases the efficiency because the ejected particles en route to the substrate are not deflected due to scattering with the background gas. As shown in section 2.5, the mean free path for a given pressure P, is given by $\lambda \approx 5\times10^{-3}/P$, where the units of λ are cm, and P is in Torr. From this we can see that for magnetron sputtering at ~ 10^{-3} Torr, the mean free path is of order ~5 cm, which is similar to the substrate-target separation, and for the non-magnetron case at ~ 10^{-2} Torr, the mean free path drops to ~0.5 cm, so a sputtered particle en

route to the substrate will be scattered many times, and the deposit rate will be substantially decreased.

The sputtering process begins with the initial ionization of a few argon ions, which are accelerated away from the cathode. As they are accelerated, they collide with other argon atoms, which, in turn, are ionized, and also begin to accelerate; in this way an ionization cascade effect occurs, and if the conditions are sufficiently habitable, the plasma can be sustained. The critical property of the plasma is that it acts like a metal, which electrically couples the anode and cathode. In DC sputtering, the discharge current continuously flows between the anode and cathode, which remains at constant potential; in the plasma, the current is composed of positively charged argon ions that are transported to the target, and negatively charged electrons flow from ground into the target to compensate built up charge. If the target is insulating, this positive charge quickly builds up on the target, thereby deflecting incoming argon ions, which prevents continuous sputtering. To avoid this, the voltage bias can be varied at radio frequencies (which has an inverse period that is similar to the time it takes to build change on the target). In forward bias, the argon ions are driven towards and into the target, which sputters the surface of the target, but this simultaneously causes charge build up. Then, in reverse bias, electrons flow from the cathode through the plasma, and arrive at the surface of the target neutralizing the positive charge there. Then, upon repeating this cycle, continuous sputtering is achieved. For both DC and RF sputtering, the deposition rate can be controlled by varying the power to the target. This allows deposition rates that are low enough to achieve epitaxial growth.

Through both DC and RF sputtering nearly any material can be deposited, which includes materials whose vapor pressures are extremely low, as well as multi-elemental targets. The latter avoids the arduous task of trying to match the rates of different targets to achieve a given stoichiometry (this is the case for MBE, see below). However, in many cases the sputtering processes will not result in a stoichiometric film even if the target itself is stoichiometeric, which is common in multielement targets where one of the elements is more volatile than the other.

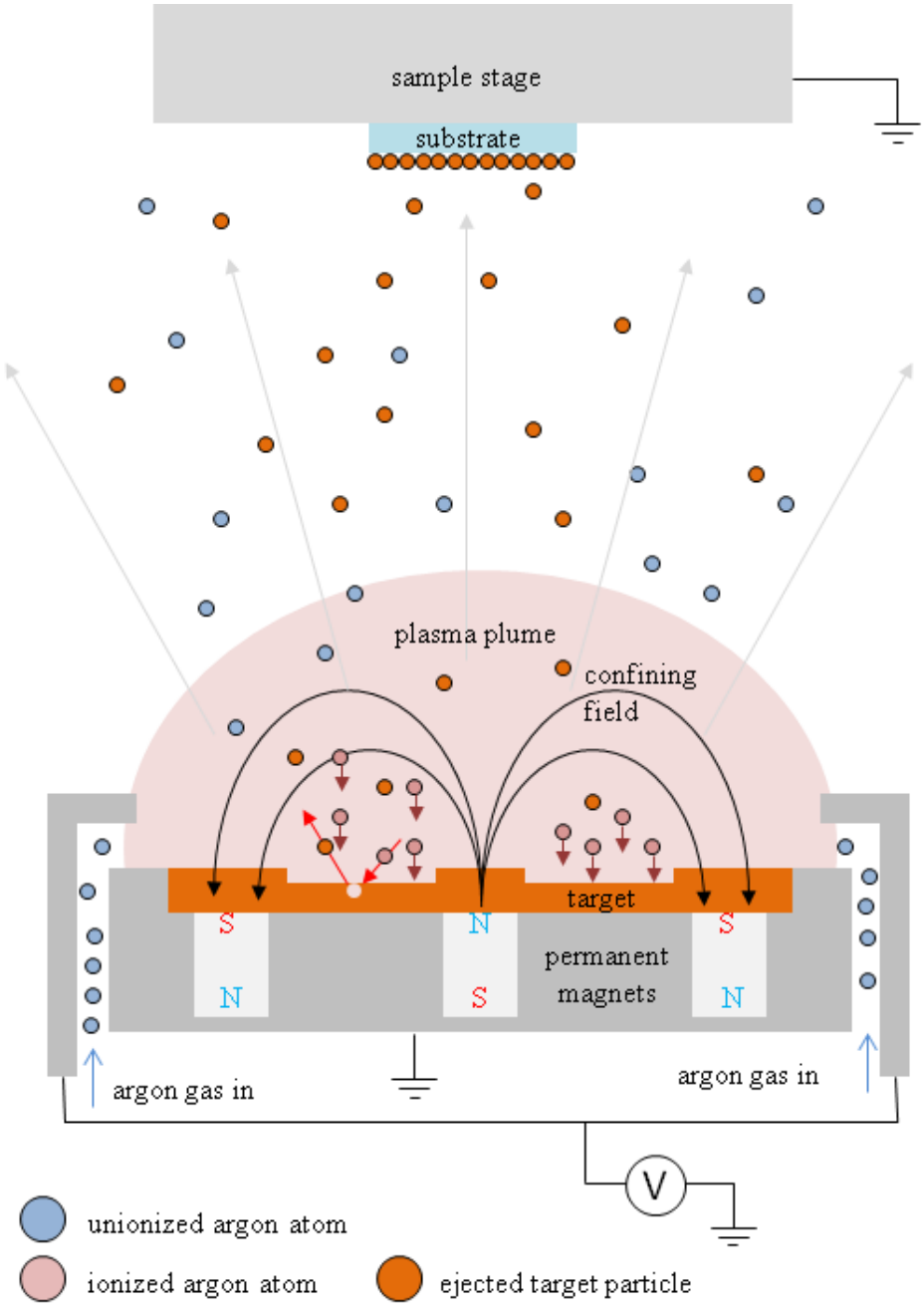


Figure 2-9 A schematic describing how a magnetron sputter deposition system operates. The argon gas is ionized and confined by the magnetic field produced by the permanent magnets, and then accelerated downward by the applied voltage. Upon impact, the target material is ejected from the target and flies towards the substrate.

The energy of the ejected particulates can be relatively large in sputtering, which can be in the range of 1-10 eV. This can have both positive and negative effects. On the positive side,

this high energy can impart more surface energy, and result in better surface diffusion, which allows epitaxial growth to be accessible at lower substrate temperatures. This high incident energy can also be a determent because the surface energy can be so high, that it desorbs the atoms off the surface in a process called resputtering. In many cases resputtering prefers one element type, and can make controlling stoichiometry difficult. This process is not detrimental in the cases where the desorption of the elements in multielement films is similar, and can be minimized by off-axis sputtering, where the targets are placed at nearly 90° relative to the surface of the substrate.

2.3.2 Pulsed laser deposition

Like sputtering, pulsed laser deposition (PLD, Figure 2-10) is a PVD process that bombards the surface of a target by an intense laser that usually operates in the ultra-violet to the visible frequency range (this may be chosen to be a wavelength that the target adsorbs well). The laser causes intense local heating and vaporization of the target, which ejects material in a plume of plasma, sending particulates of the target material towards the substrate. The repetition rate of the laser is typically in the range of 10-100 Hz, and the pulse length is much shorter in the range of nano-seconds. This long delay in repeated pulses is a key to tailoring the ablation process to achieve the optimal growth conditions ([64]Eason, 2007). PLD systems are typically equipped with rotatable target carousals, which can hold typically 5-10 separate targets; this gives a large flexibility to the types of materials that can be deposited in succession, and makes possible creating epitaxial heterostructures on the atomic scale.

Like sputtering, PLD is also capable of ablating nearly any material due to the extremely high temperatures achieved during the intense laser heating. Again, this has both positive and negative aspects. On the positive side, when applicable, this allows the use of stoichiometeric targets that ensures stoichiometeric films, which simplifies the difficult task of manually adjusting the stoichiometery, as in the case of MBE. However, since the ejected particulates in

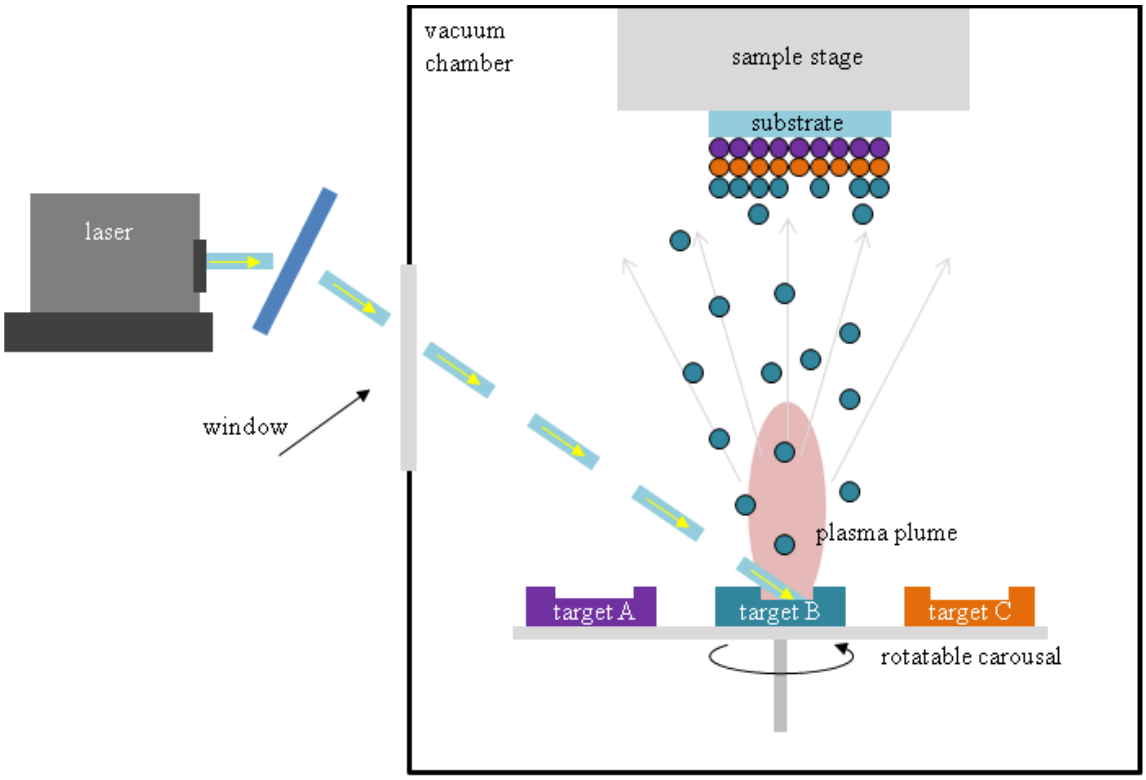


Figure 2-10 A schematic of a pulsed laser deposition system. This shows the laser pulses, which are generated outside of the vacuum space, and are impinged on one of several targets that are mounted on the rotatable carousal used to select the target. The resulting ablation process sends material towards the substrate where the growth takes place.

PLD have energies in excessive of 1-10 eV, there also exists the same problem of ejecting more volatile elements off the surface of the film, which can yield non-stoichiometeric films, and in some cases induce defects on the surface ([65]Eckstein, 2007). Despite these issues, PLD has been shown to produce very high quality films.

Due to the pulsed nature of PLD, the growth process is different compared to sputtering and MBE. In PLD the laser heating happens very quickly, on the order of nanoseconds, and the subsequent ablation of the target and deposition on the substrate happens within microseconds. This is followed by a 10-100 millisecond gap, where the adsorbed material diffuses and finds the lowest energy atomic position for chemisorbtion to occur (see chapter 5 of [64]Eason, 2007).

Unlike sputtering, PLD can be performed with or without a background gas. This flexibility has many advantages and can offer further control over the growth. The residual

background gas can be used for reaction with the film, which is commonly used in the growth of oxide materials; also, the background pressure can be adjusted to increase scattering with the background gas, which lowers the incident energy of the incoming particulates and lessen the probability to eject the adsorbed material. Lastly, the lower working pressure of PLD allows in situ and real time monitoring by RHEED, which the high pressures of sputtering makes difficult. This in situ monitoring is critical to optimizing the growth parameters, and monitoring film quality.

2.3.3 Molecular beam epitaxy

Molecular beam epitaxy is a PVD technique that uses single element thermal sources that can be used to evaporate individually or simultaneously onto a substrate (see Figure 2-11). The thermal evaporation process is very controllable, gentle, and reproducible, which gives the ability to control the ratio and rates of each element by simply adjusting the cell temperatures; these together make MBE the most controllable PVD growth process. Typically, MBE systems are equipped with ~5-15 individual sources, each containing an individual element. Nearly all the elements in the periodic table can be used in MBE by either gaseous infusion, or thermal evaporation. For the refractory elements that cannot be thermally evaporated, like Zr, Nb, Mo, Ta, W, Re, Os, and Ir, with several other elements which are on the border line like B, Rh, and Pt, electron beam evaporators can be used. However, this type of source imparts higher energies to the ejected particulates, and is not as stable as thermal cells. Like PLD, MBE is typically performed in ultra-high vacuum, and can make use of real time in situ monitoring by RHEED.

Unlike PLD and sputtering where the stoichiometery of the film is largely set by the target, in MBE the composition is determined through a more manual process. For many binary materials such as ZnO, GaAs, and in the present case Bi_2Se_3, the stoichiometery is set naturally because these materials are self-limiting based on the cation flux, and, therefore, are always grown in an excess flux of the anion. For example, in the growth of Bi_2Se_3 the Bi and Se flux

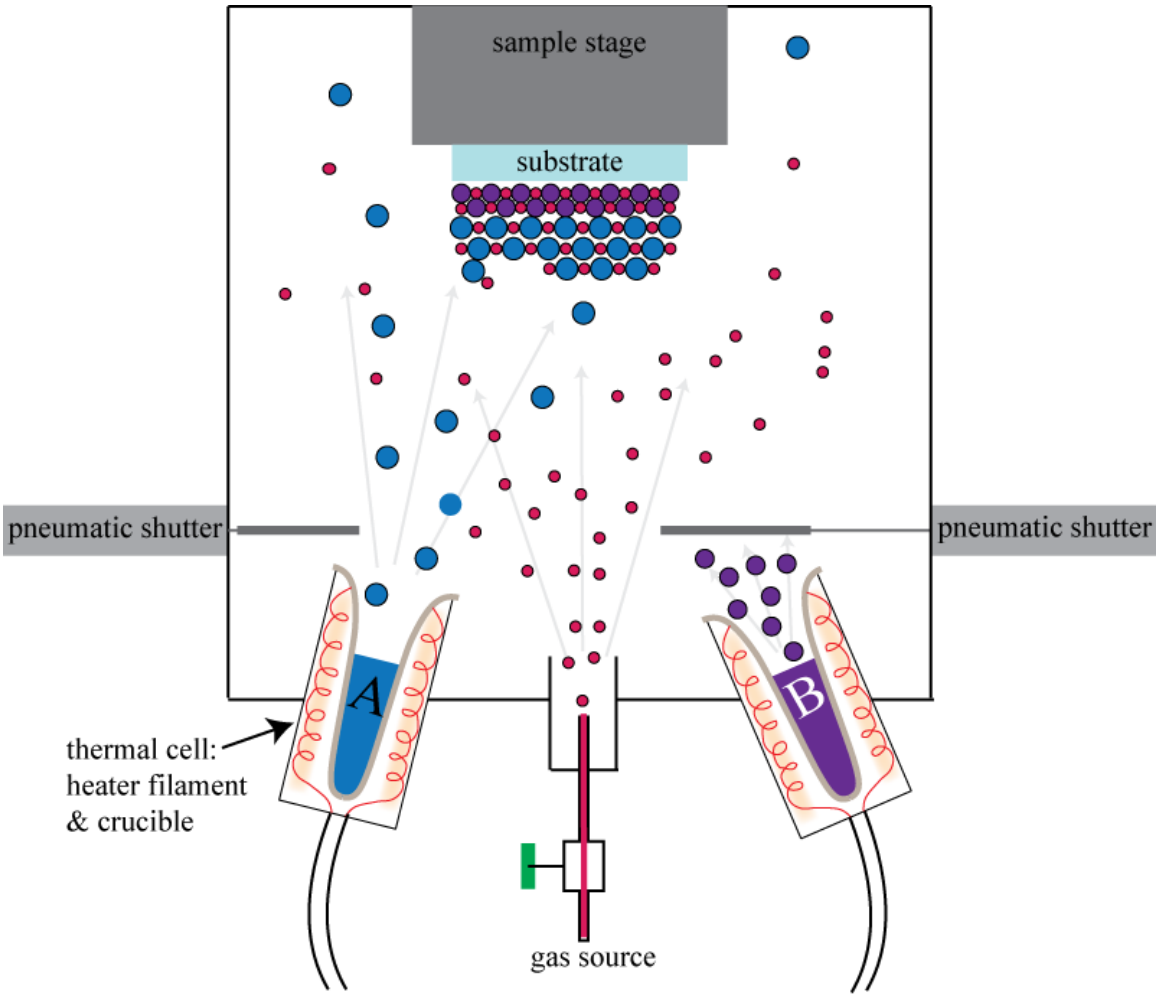


Figure 2-11 A schematic of a molecular beam epitaxy system. As shown, there are two thermal cells with the elemental source inside a crucible that are surrounded by heater filaments, and one gas cell. The cells are blocked from the substrate by shutters which are opened to initiate growth.

ratio is typically one part Bi to ten parts Se, however, most of the Se atoms desorb, which results in a stoichiometeric film with a ratio of two to three; for these types of materials the error in the stoichiometery can be below the 0.0001% level, which simply cannot be achieved with other PVD processes due to the relatively large energies used. MBE is also the best technique to create alloys and films with a controlled doping dependence. PLD and sputtering require separate targets for each composition, whereas MBE can simply change the relative cell temperatures to control composition; this is the case for a material system like $(Bi_{1-x}In_x)_2Se_3$, which can be grown for any x from 0 to 1 by MBE, but would be very difficult to achieve by PLD or sputtering.

For more complex materials like the perovskites with structure ABO_3, the A-to-B ratio must be done manually (although some perovskite oxides can be grown in a self-limiting fashion, for example $PbTiO_3$ ([181]Theis *et al.*, 1998), and more recently $SrTiO_3$ and others by using metal-organic gas source(s) in conjunction with thermal sources (see [88]Henini, 2013 chapter 17)). This can be done either by adjusting the thermal cell temperature, so the flux of A is equal to the flux of B, and then co-depositing A and B in an oxygen environment, or this same structure can be achieved by depositing repeated monolayers of A-O and then B-O_2, where the fluxes of A and B must be known, but can be different; these methods will invariably result in some level of none stoichiometery, typically in the range of ~1% error, attributable to errors in the measurement of the elemental fluxes, and the control of the shutters, which is in a similar regime as both sputtering and PLD. Compared to sputtering and PLD the energetics of thermal evaporation is much lower, in the rage 10-100 meV, so the ejection of adsorbed elements is not problematic, since the energy is of similar scale to the physisorption energy. Moreover, the lower kinetic energy makes possible the growth of materials which are closer to thermodynamically unstable or metastable, and would relax to the more stable configuration with the available energy.

2.4 Molecular beam epitaxy at Rutgers University

2.4.1 The MBE system

The MBE reactor used at Rutgers University during this work was designed and build by SVT Associates, to the specifications of Professor Seongshik Oh in 2007-08. This system was originally purposed to grow oxide materials, and included components that can withstand the corrosive environment that is required to grow these materials. However, due to initial problems with vacuum leaks in 2008-2010, projects involving oxides materials were postponed; despite this, the vacuum conditions were sufficient for the growth of topological insulator materials. In 2010 these problem were fix, but TI materials have remained the main material system grown in this chamber.

Figure 2-12 shows images of the MBE system at Rutgers University that was used to grow all the samples in this dissertation. The entire MBE system is composed of three individual vacuum chambers to maintain UHV conditions in the growth chamber. The load-lock chamber (Figure 2-12b) is where the substrates are loaded and removed from the MBE system, and is the only chamber that is regularly opened to the atmosphere. This chamber is pumped by a turbo pump that is backed by either an oil-free scroll or diaphragm pump, which allows the chamber pressures to range from atmosphere (10^3 Torr ≈ 1 atmosphere) down to 10^{-6} Torr. Vertically below the load-lock chamber is the preparation (prep) /sputtering chamber (Figure 2-12c), where the substrates are staged for transfer into the into the growth chamber (Figure 2-12d) or into sputtering stage (within the prep chamber). Both the prep and growth chambers are pumped by cryo-pumps, with typical pressures ranging from 10^{-6} -10^{-10} Torr. The base pressures for the growth and prep chambers are ~ 10^{-10} Torr and ~ 10^{-9} Torr, respectively; the higher pressure of the prep chamber is due to its exposure to the lower vacuum of the load-lock during transfer.

As with all modern ultra-high vacuum (UHV) systems, this MBE chamber is constructed out of 300 series stainless steel, which is rolled out of sheets ~0.1 inches thick, and are welded together or onto universal ConFlat (CF) flanges for attaching vacuum components (all vacuum welding is always performed on the vacuum side to minimize surface area and prevent any crevices). As shown in Figure 2-12a-g, one can see many standard UHV parts that enable external operation of such a chamber without breaking vacuum: for sight into the chamber, sealed glass windows are used. For mechanical manipulation, a feed through system that either uses an accordion-like baffle (on all gate valves, the load-lock to prep transfer elevator, the quartz crystal microbalance, the prep and growth stages, and the cell shutters) or a magnetic feed through, which employs two permanent magnets, one on the outside which is used to manipulate another magnet that is mounted inside the chamber (the prep-to-growth transfer arm, rotatable growth stage, all window shutters). Electrical components are fed into vacuum using stainless steel flanges fused to ceramic inserts, which are used to electrically isolate and make a vacuum seal

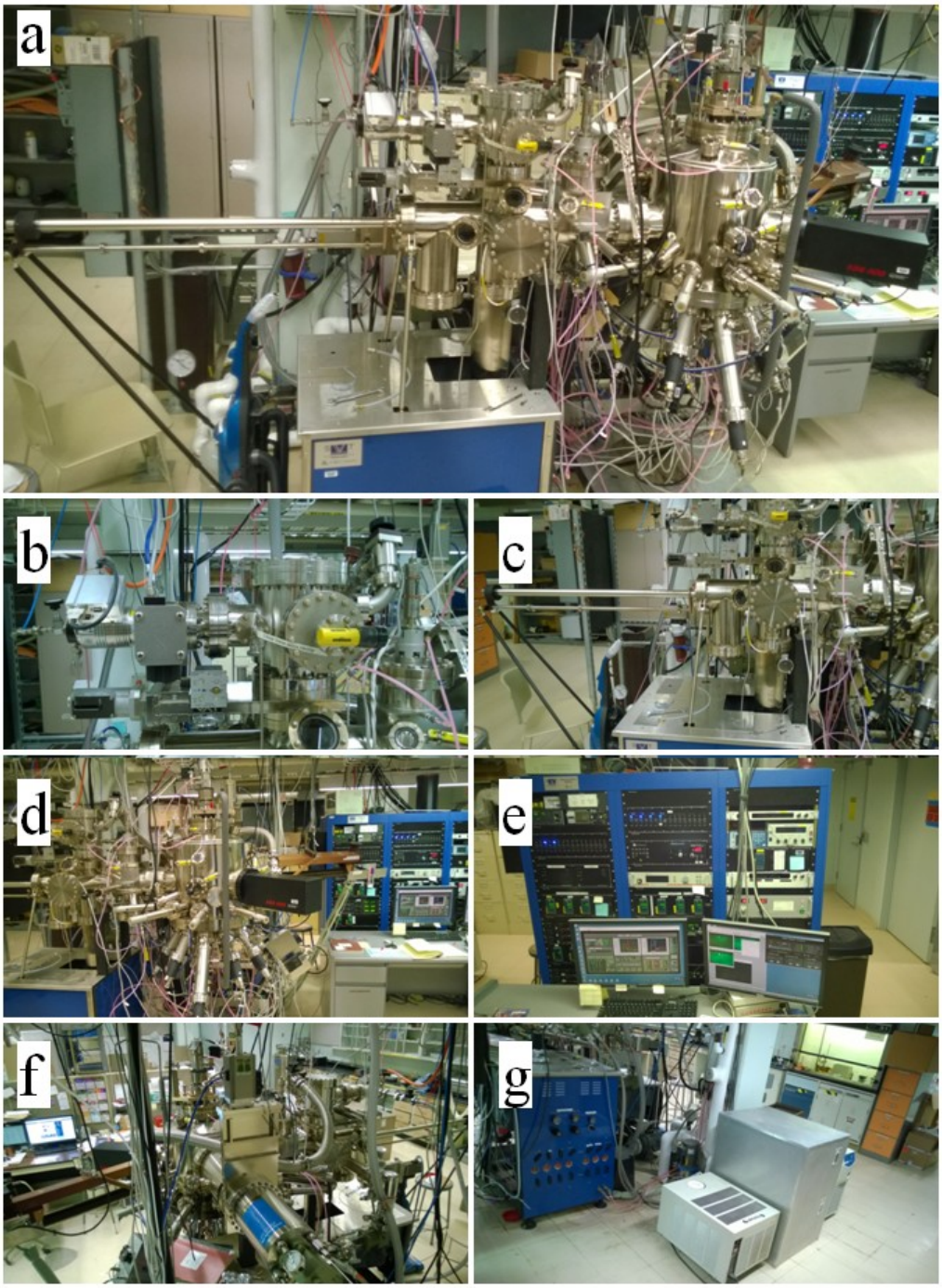


Figure 2-12 The MBE system at Rutgers University. a A wide scale view from the front of the MBE. **b** The load lock chamber, where substrates are loaded from the atmosphere and transferred vertically downward into the prep chamber. **c** The prep/sputtering chamber, where substrates are transferred from the storage cassette into the growth chamber or into the sputtering stage. **d** The growth chamber where thermal deposition takes place. **e** The rack for the electronics that control all devices on the MBE system, and the PC which is the primary means for controlling the deposition. **f** An alternative view from the back of the MBE. **g** A view of the support equipment. The blue panel contains the compressed air and cooling manifolds. The boxes from left to right are the QCM water chiller, the cryopump compressors (within the aluminium sound baffling boxes), and on the right the home-built heat exchanger that is responsible for all the cooling to the entire system.

around the electrical wires (used for thermal cell and sputter gun power, the RHEED electron gun, thermocouples, and all pressure gauges). Lastly, gas is fed into the vacuum system by numerous gas valves, by either leak valves, which provide a slow and controlled flow of gas, or shut-off valves, which are used for venting.

The seemingly tangled mess of wires and tubes leading into the MBE system are electrically wiring, cooling lines, purified gas sources, compressed air (pneumatic) lines, and vacuum exhaust lines. All of the controls for the pneumatics and electronics are stationed in the electronic rack shown in Figure 2-12e, which are all interfaced in real time with the PC in the foreground. Most growth processes are controlled remotely through the PC, which facilitates ease of operation, but more importantly, it ensures consistent operation during the multi-stepped growth process. Lastly, Figure 2-12g shows the supporting equipment used to operate the MBE system; in this panel, one can see all the cooling lines, which are pink in color due to the glycol that is used as a coolant, which is less corrosive and more resistant to algae growth than deionized water. The coolant is circulated through a home-built 150,000 BTU heat exchanger that thermally links the coolant lines to the building-wide coolant system that operates at ~10° C. Inside the metallic box shown in Figure 2-12g are the helium compressors that are used to compress and cool gaseous helium, which is then sent to the MBE, where it is used as a refrigerant to cool the cryo-pumps to cryogenic temperatures.

2.4.2 Substrate handling

The MBE system is capable of handling up to 10 substrates, which can range from as small as a few millimeters to as large as three inches in diameter, with the most commonly used size being 10×10 mm^2. As shown in Figure 2-13, any substrate less than 3 inches is attached to a 3 inch diameter, 0.030 inch thick metal plate, constructed from either molybdenum or Inconel (an austenitic alloy based on nickel and chromium), which are used due to their ability to withstand high temperatures. The metal plates have a window milled through, which allows the substrate to

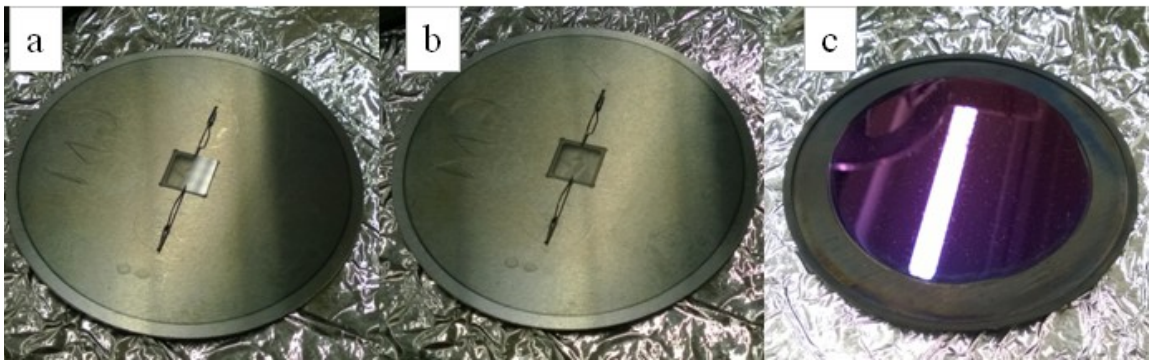


Figure 2-13 Substrate mounting. a-b Two views of a molybdenum plate holding a 10 × 10mm^2 Al_2O_3 substrate, which is more visible in **a**. In **b** the through hole is visible, which is to enable the substrate to be heated directly by radiation for the back, and the shoulder when the substrate sits down into, where it can be held in place by the tungsten-rhenium wires. **c** A 3 inch Si wafer that is mounted in a holding ring (the plate shown in **a-b** is ultimately placed in one of these rings).

be heated directly by radiation from the heating element on the growth stage. There is a shoulder milled on the perimeter of the window, which prevents lateral movement of the substrate on the plate. To fix the substrate in place vertically, 0.016 inch tungsten-rhenium alloyed (97% tungsten-3% rhenium) wires are threaded through holes in the plate, and are bent into place in order to apply slight force perpendicular to the surface of the substrate. The tungsten-rhenium alloy is used because it can withstand high temperatures while retaining its elasticity. Once attached to the plate, the substrates are cleaned in a UV ozone oven for 5 minutes, before being mounted into a holding ring and placed in the substrate cassette.

Figure 2-14a-e shows the process of loading a substrate from atmosphere to the growth chamber. Once the substrates are prepared, they are loaded into the substrate cassette, and then the load-lock chamber is vented. In the load-lock the substrate cassette sits in a retractable holder. The load-lock chamber is pumped down first by rough pump, then by turbo down to a pressure of around 10^{-6} Torr, which takes about 10 minutes (alternatively without using the turbo pump, the load lock can be pumped by the rough pump alone to $< 10^{-2}$ Torr, then the rough pump can be valved off, and the gate valve can be opened, which in total takes about 20-30 minutes and simply raises the pressure in the prep slightly). At this point, the gate valve is opened, and the elevator is

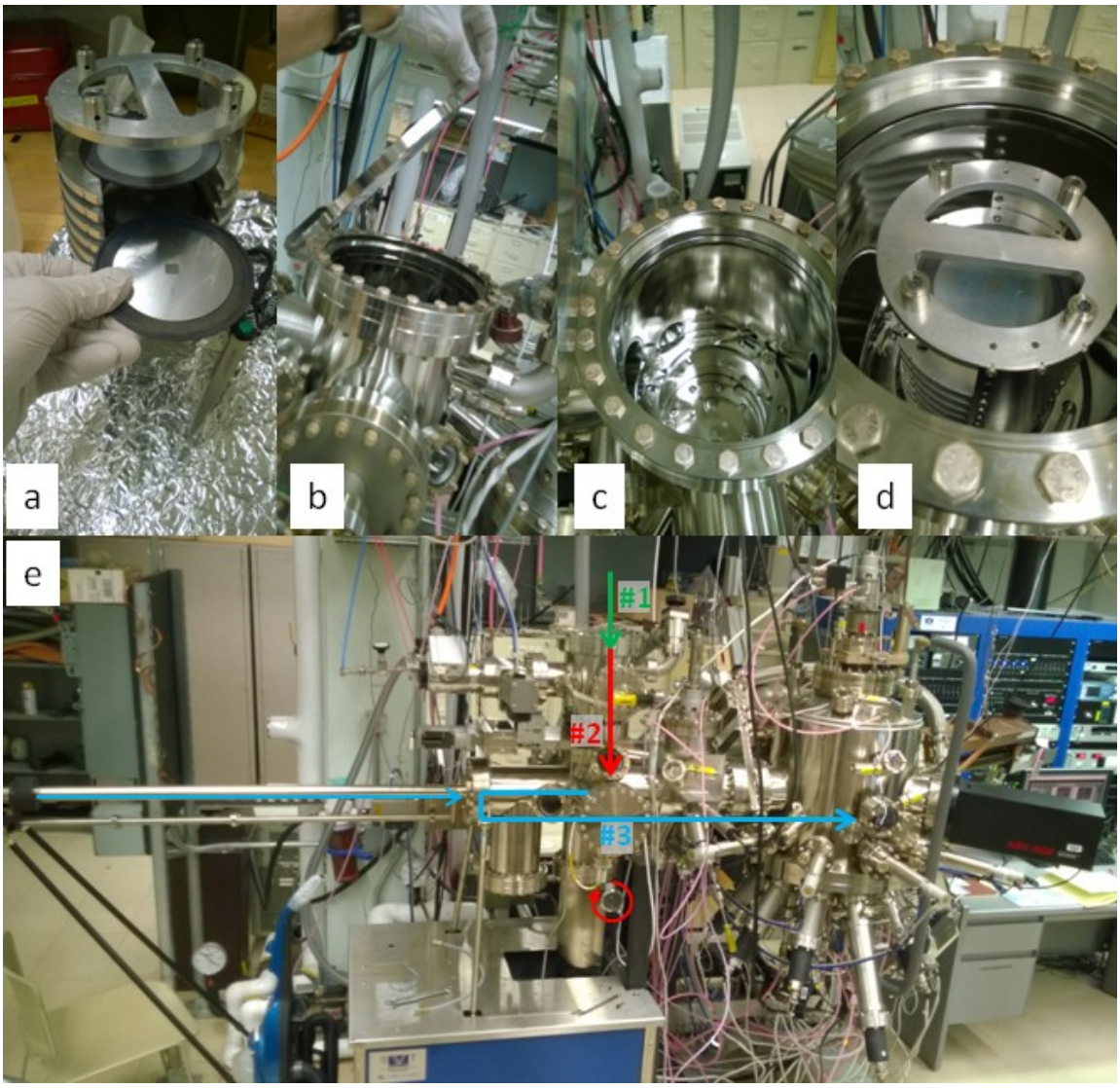


Figure 2-14 The process of transferring a substrate from atmosphere through to the growth chamber. a Loading the substrate holder and ring into the sample cassette. **b** The load-lock chamber vented. **c-d** The retractable holder that the substrate cassette sits on. **e** The process of transferring the cassette into the load lock (green, #1), then down into the prep chamber by rotating the knob which raises and lowers the elevator (red, #2). Finally, the transfer arm is inserted into the cassette, the substrate is retracted towards the left, the cassette is lowered, and finally the transfer arm is pushed towards the right into the growth chamber (blue, #3).

raised until the cassette is pushed up above the retractable holder, which is then retracted, and the cassette can be lowered down into the prep chamber, and the gate valve is then closed. Once in the prep chamber, the substrate to be used can be removed by raising the cassette until it is aligned with the horizontal transfer arm. The desired substrate and ring are then removed from the

cassette, and the cassette is lowered below the transfer arm. Then, the gate valve that separates the growth chamber and the prep chamber is opened, and the growth stage is lowered to accept the new substrate. Then the substrate is transferred into the growth chamber, where the stage is raised, and the arm is retracted. Once in the chamber, the substrate and ring are checked visually, and by RHEED to ensure correct alignment, then the growth can proceed.

2.4.3 Thermal cell operation

The operation of thermal cells starts with the characteristics of the specific elements to be used. The main properties to consider are the reactivity, and how the vapor pressure of the individual element depends on temperature. Highly purified elements are available commercially with purities better than 99.999%. Before loading the elements into the thermal cell, the crucible is cleaned by out gassing to high temperatures in vacuum to lessen the chance of contamination of the source material. The material is then loaded into the crucible, with the amount depending on the size of the crucible, and the application. For Se, which is heavily used, up to ~30 cm^3 are used, which typically lasts for ~ 6 months, while for elements that are not used as frequently ~3-5 cm^3

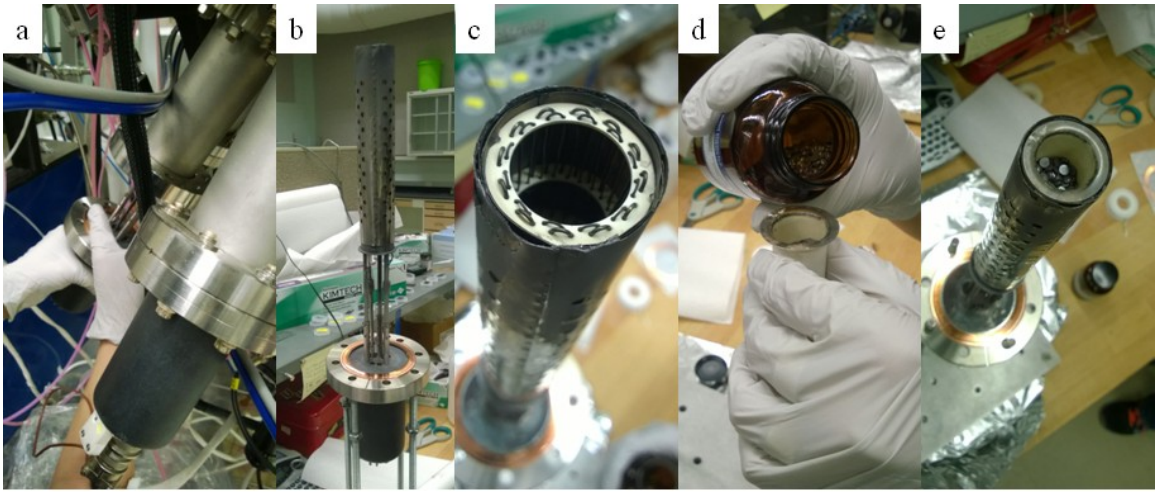


Figure 2-15 Various images of an MBE thermal cell during routine maintenance. a A low temperature thermal cell that is being removed from the MBE chamber to recharge Se. In **b-c** the Se thermal cell is shown while mounted on a maintenance stand; in **c** the heating filament is visible with the crucible removed. Shown in **d**, the crucible as it is being refilled with Se shot, and in **e** the thermal cell with refilled crucible in place ready for reinstallation.

is more typical, which might last for years of use (see Figure 2-15). The materials that the crucibles are made of are critical to avoid decomposition during use, due to chemical reaction with the element. Some common crucible materials used are Al_2O_3, pyrolytic boron nitride, vitreous carbon, graphite, and tantalum. The compatibility of the element and the crucible is typically looked up (for example see the Kurt J. Lesker Material Deposition table available at https://www.lesker.com/newweb/deposition_materials/MaterialDeposition.cfm?pgid=0). Currently (2014), in our chamber Sr, In, Se, and Sb are in pyrolytic boron nitride crucibles, Sn, Bi, Ni, Fe, and Cu are in Al_2O_3 crucibles, and lastly La is in a Ta crucible (prior to this La was in a Al_2O_3 crucible which reacted and formed a crack). Since many of the materials used for crucibles are brittle, it is especially important to avoid mechanic stress. This occurs most frequently during melting and solidification of the source material, and to minimize stress, a very slow ramp rate is used near the melting point, typically 1-2 °C/min.

The vapor pressure of an element determines the temperature where there will be a sufficient flux in order to achieve significant deposition. To first-order, the vapor pressure varies exponentially with temperature as $P(T) \sim Exp(-E_A/k_BT)$, where the activation energy is E_A. To achieve the desired flux, typically in the range of $\sim 10^{13}$ /cm^{2}s (which yields approximately a monolayer per minute since a typical atomic density is around $\sim 10^{15}$ /cm^2), the vapor pressure needs to be in the range of $\sim 10^{-4}$-10^{-3} Torr, which will give approximately the cell temperature required for deposition. This gives only an estimation and the exact temperature required to achieve the desired flux should be measured. The working cell temperature can have huge variation for different elements: Se, with its very high vapor pressure, requires only around 150-170 °C, while elements like La, and Ti, both of which sit at the upper limit for what is possible to thermally evaporate, require temperatures in the range of 1700 °C, and elements like Bi, and In are right in the middle, requiring temperatures around 600-700°C. In the MBE system, the thermal cells are segregated by temperature range because the higher temperature cell required more specialize design. There are 5 low temperature cells (below 1000°C) containing Bi, Sr, Sb,

In and Se, 2 medium temperature (below 1400°C) cells containing Cu and Sn, and 3 high temperature cells (below 1800°C) containing Fe, Ni, and La.

2.5 Ultra-high vacuum

2.5.1 Necessity for ultra-high vacuum conditions

In the MBE chamber, the pressure is in the range of ~10^{-10} Torr, which is equivalent to the pressure of the residual gas of interstellar space. There are two main reasons why such a high level of vacuum is required for the growth of thin films by MBE. One is that at this low pressure atom-to-atom scattering is reduced, and the mean free path is increased, which enables line-of-sight evaporation, and less diffusion for the elemental and electron beams; in this regard, if diffusion of elemental sources occurred, then this would likely lead to inter-source contamination, which would then cause film contamination. To see how the mean free path roughly depends on pressure we can calculate it as follows. Imagine a background gas of atoms with an average separation l, and, therefore, a density of $n = 1/l^3$. The mean free path is given by the distance traveled vt (v is velocity and t is the time between scattering events), divided by the number of gas particles n, within the volume sweep out by the traveling particle $V = Area \times length = (\pi d^2)\times(vt)$, where d is the atomic diameter. The mean free path is then $\lambda = vt/(Vn) = vt/(\pi d^2 vtn) = 1/(\pi d^2 n) = k_B T/\pi d^2 P$, where $n = N/V = k_B T/P$ with P being the pressure. For the case of nitrogen, this simplifies to $\lambda \approx 5\times10^{-3}/P$, where the units are given in cmTorr. From this we can see that the mean free path of the background gas drops to roughly the size of the chamber (~1 m^3) at around 10^{-5} Torr, while for the typical operating pressure of 10^{-10} Torr, the mean free path is about ~1000 km, and, therefore, we can safely assume no diffusion occurs.

The second and probably more important reason for the extremely low residual background pressure is the need to lower the impurity defects in the thin films. If the background pressure is high, then as the film grows, the residual gas molecules may adsorb onto the surface, and become incorporated in the film. Such an effect may be disastrous depending on the native

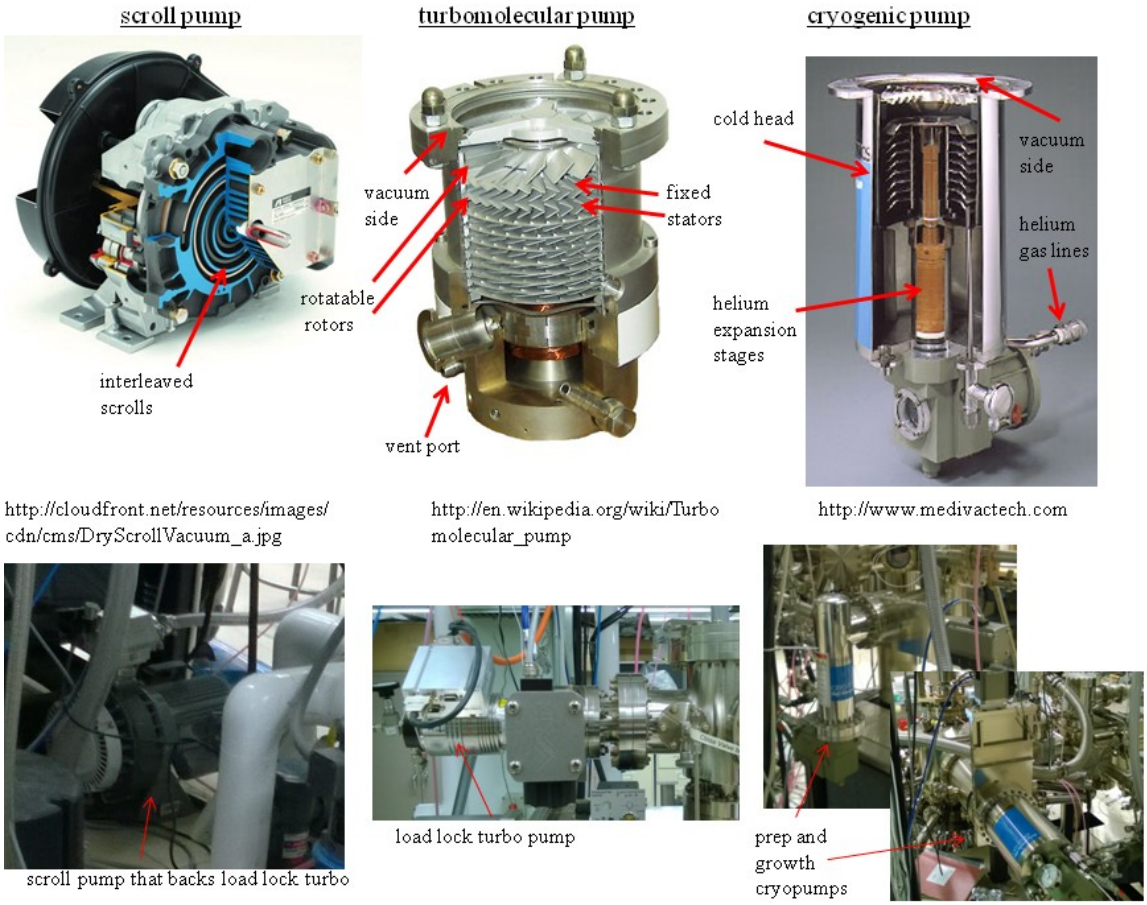


Figure 2-16 Cut-away diagrams of the various types of pumps used on the MBE chamber, with the corresponding pictures of the pumps on the MBE chamber directly below.

defect level inside the film. As an upper limit we can calculate how long it will take to cover the surface of the film by a single monolayer of residual gas assuming a sticking coefficient of one (i.e. every atom that hits the surface will stick). On the walls of a vacuum chamber the gas molecules exert a force $N = PA = \Delta p/\Delta t$, where A is the area, and Δp is the transferred momentum of the gas molecules in a time interval Δt. If we take $\Delta p = N_A \Delta p_A$ where N_A is the number of atoms that impact with individual momentum $\Delta p_A = mv$, with v being the average velocity, which is related to the temperature by the equipartition theorem to be $1/2mv^2 = 1/2k_BT$. Substituting these yields $PA = mvN_A/\Delta t$. Now, $N_A/(\Delta tA)$ should be recognized as a molecular flux F, which counts how many atoms per unit area are impacting in a given time interval. Rearranging this yields the

final result of $F = \frac{N_A}{\Delta t A} \approx P/\sqrt{mk_BT}$. With this we can calculate the time it will take for one monolayer to accumulate for a given pressure. Typically one monolayer is ~10^{15} atoms/cm^2, and for nitrogen, the time to accumulate one monolayer on the surface of a thin film is found to be about ~1 second if the vacuum is ~10^{-6} Torr. Therefore at 10^{-6} Torr if the sticking coefficient is unity, then a film will be mostly contamination (however, the sticking coefficient is likely much less than one). Now, at 10^{-10} Torr, the time required to accumulate one monolayer increases to roughly 2 hours, which, assuming a deposition rate of 1 monolayer per minute yields a defect density of $\lesssim$ 1%, which is acceptable for many systems, such as Bi_2Se_3, whose native defect density of Se vacancies is in this range. However, in other systems like GaAs, is it critical to have a defect density $\lesssim$ 0.0001%, which requires a vacuum below 10^{-11} Torr, and much faster deposition rate to achieve this extremely low defect density ([68]Farrow, 1995).

2.5.2 Achieving ultra-high vacuum

There are many types of pumps that can be used to achieve different levels of vacuum (see Figure 2-16, and [144]Ohring, 2002 chapter 2 for a thorough review). Vacuums pumps fall into three categories. Positive displacement pumps are probably the typical image that is conjured when one thinks of a pump. These pumps used a cyclic motion to compress the residual gas on the vacuum side, which is then vented to the atmosphere. Piston, scroll, rotary vane and diaphragm pumps all fall into this category. These pumps only achieve low vacuum in the range of $\gtrsim 10^{-3}$ Torr; below this pressure, the backflow on the vent cycle eventually equals the amount of gas that is being exhausted, and, therefore, are eventually not effective. Typically in UHV systems, these pumps, often referred to as rough pumps, are used to back the second category of pumps called momentum transfer pumps. It should be noted that in any UHV system it is critical to avoid the use of any rough pump that uses oil, which may leak into the vacuum chamber with the back streaming gas and cause severe contamination.

Momentum transfer pumps operate by transferring momentum to the incoming residual gas particles, which are directed out of the vacuum space towards the vent port. At the vent port the pressure builds to a sufficiently high value, which allows it to be effectively pumped by a rough pump. The momentum transfer can be achieved by either diffusion pumps or, most commonly, turbomolecular pumps (simply called turbo pumps). Diffusion pumps use super-heated oil that spays in a jet directed away from the vacuum side and towards the vent port. This jet pushes and confines the residual gas at the vent port. Typically, the oil's finite vapor pressure leads to oil contamination of the chamber, and the high operating temperatures make this type of pump uncommon, due to availability of the much cleaner turbo pumps.

Turbo pumps use a series of turbine rotors that are interleaved with stators (stator—stationary rotor) with opposite pitches. The rotor spins at speeds in excess of 100,000 rpm. This high speed efficiently directs residual gas molecules out of the vacuum and compresses them so they can be effectively pumped out of the chamber by a rough pump attached to the vent port. These pumps can be used to achieve pressures as low as 10^{-9} Torr, but like displacement pumps, the rate the gas is exhausted will eventually be equal to the rate of the back streaming gas as the pump reaches its base pressure. On the MBE system, the load lock chamber is pumped by a turbo pump that is backed by an oil free scroll pump.

To achieve pressures below the capability of a turbo pump, the back streaming in the exhaust step can be avoided by capturing the residual gas, which is the basis of operation for entrapment pumps. The most common forms of entrapment pumps are ion and cryogenic pumps. Ion pumps operate by ionizing the residual gas molecules, which are then accelerated into a cathode where they are driven in, and chemically bonded to, the cathode material. Typically, titanium is used as the cathode materials, which, during operation will often sputter onto the chamber walls, which will chemically absorb more residual gas by acting as a getter pump.

On the MBE chamber the prep and the growth chambers are pumped by cryogenic pumps (cryopumps). Cryopumps utilize compressed He as a refrigerant gas which is expanded within the

cold head and cools it down to ~10-20 K. The cold head then adsorbs and traps residual gas because at this low temperature there is insufficient thermal energy for the gas molecules to desorb off the cryohead; this process of trapping gas on a cold surface is analogous to water molecules being pumped out of the atmosphere by the surfaces of a cold bottle of soda. To increase the pumping ability, cryopump heads are often coated with materials that have naturally high surface areas such as charcoal or zeolite. The increased surface areas increase both the pumping efficiency and duration between regeneration steps (see below). Entrapment pumps work well for nearly all gases, but have trouble pumping gases that do not condense easily, with particular problems pumping He gas. Since cryopumps do not continuously exhaust the pumped gas, they eventually have to be shut down, and warmed up to exhaust the trapped gas. In the MBE system this regenerative (regen) step is performed once every year, either when the system is shut down for major maintenance, or when the background pressure is found to begin to slowly rise or anomalous spikes in the pressure are observed.

Critical to achieving UHV conditions is the ability to eliminate all sources of gas, which is carried out during the initial pump down from atmosphere. The two main sources of gases are vacuum leaks and trapped gases, particularly H_2O, that are adsorbed onto the chamber walls during exposure to atmosphere. Leaks are located by using the in situ residual gas analyzer (mass spectrometer) that is set to detect trace amounts of He gas. To detect a leak, He gas is sprayed from a small tube that scanned around the outside of the chamber; when the tube passes over a leak, the analyzer detects an increase in the He inside the chamber, which signals the location of the leak. The main source of leaks during regular maintenance usually involves improperly tightening CF flanges; CF flanges should be tightened in a star-type pattern to slowly and evenly compress the Cu gasket into the knife-edges on the flange surfaces. However, welded components can develop microcracks, which arise from mechanical or thermal stress, and can be serious enough to require disassembly of the chamber for repair. Even small leaks in the MBE chamber will typically prevent the vacuum level from dropping below ~10^{-8} Torr.

Once the initial volume of gas has been pumped out, the adsorbed gases on the chamber wall limit the chamber from dropping below ~10^{-7}-10^{-8} Torr. Over time, these gases will eventually be pumped out of the chamber, but depending on the amount of the adsorbed gas, and the total surface area this process can take several months to as long as a year. To expedite the process, the chamber is enclosed in a thermal blanket and baked at a temperature of ~150-200°C. This high temperature increases the desorption rate, and the adsorbed gas can be driven off in as little as a few days to a week. For minor maintenance, such as recharging an element in a thermal cell, the MBE chamber will be vented with dry N_2 gas, and will only be open for 5-20 minutes under continuous N_2 flow. In these brief exposures, only a minimal amount of gas is adsorbed, and the system will not require baking with the vacuum level dropping down to the range of 10^{-9} Torr within a few days. This is sufficiently low for growth to resume, and over the course of a week or two, will slowly achieve the typical base pressure of 10^{-10} Torr.

2.6 Atomic flux measurement

The knowledge of how many atoms are arriving at the substrate in a given amount of time, called the atomic flux, is the key to achieving precise atomic scale thin films. This requires the ability to measure accurately and precisely the atomic flux of each individual element. In situ, the dominate forms for measuring the flux are a quartz crystal microbalance (QCM), ion gauges, and atomic adsorption. QCM is the only method to directly measure the flux, but requires the substrate to be moved, and, therefore, cannot be measured in real time during the growth. Ion gauges, which measure the vapor pressure at the substrate location, need to be calibrated against another method to give a flux, and also requires the substrate to be moved out of the growth location. Lastly, atomic absorption measures the change in intensity of a light beam as it passes through the molecular beam, which also needs to be calibrated to give the flux at the sample surface, but this does not require the substrate to be moved, and, therefore, the flux can be measured in real time. Ex situ, the most accurate measure of composition, and, therefore, flux, is done by Rutherford

backscattering spectroscopy (RBS), which operates by measuring the energy dependence of the backscattered ions from an impinged ion beam; RBS is able to identify the total concentration of individual chemical species with depth resolution within a film, which is typically used in conjunction with the in situ methods.

Primarily in this work, we make use of the combination of QCM and RBS to complimentary determine the flux of each element, which together provide an accuracy of $\lesssim 1\%$. When combined with the high stability of the thermal cells, flux measurements for Bi_2Se_3 based materials can be done on a weekly basis.

2.6.1 In situ quartz crystal microbalance

The quartz crystal microbalance, or QCM, gets its name from its capability to measure tiny masses, which can be as small as ~10^{-8} grams. In the MBE chamber the QCM is installed on a retractable bellow system, which allows it be inserted into the vacuum chamber at the substrate location when the growth stage is raised (see Figure 2-17). The QCM operates by measuring the change in resonance frequency of a thin quartz crystal as material is deposited on its surface. The quartz crystal can be thought of qualitatively like a spring with one fixed end, and a mass attached to the opposite end. The spring-mass system has a natural resonance frequency give by $\omega_0 = \sqrt{k/m}$, where k is the stiffness of the spring, and m is the mass on the end of the spring. If one imagines the spring is oscillating, and the mass changes by a small factor Δm, then the resonances frequency decreases by $\Delta\omega$, which can easily be measured, and since k is constant, the change in mass is given to first-order by $\Delta\omega \approx -\omega_0\Delta m/(2m)$. In the analogy with the spring and mass system, the spring constant is provided by the elastic modulus of the quartz crystal which is usually cut at ~35° with respect to the *c*-axis (so-called AT type) with a resonance frequency constant of 1656 kHz-mm. The mass on the end of the spring corresponds to the total mass of the crystal, and the added mass is that of the film deposited on the surface.

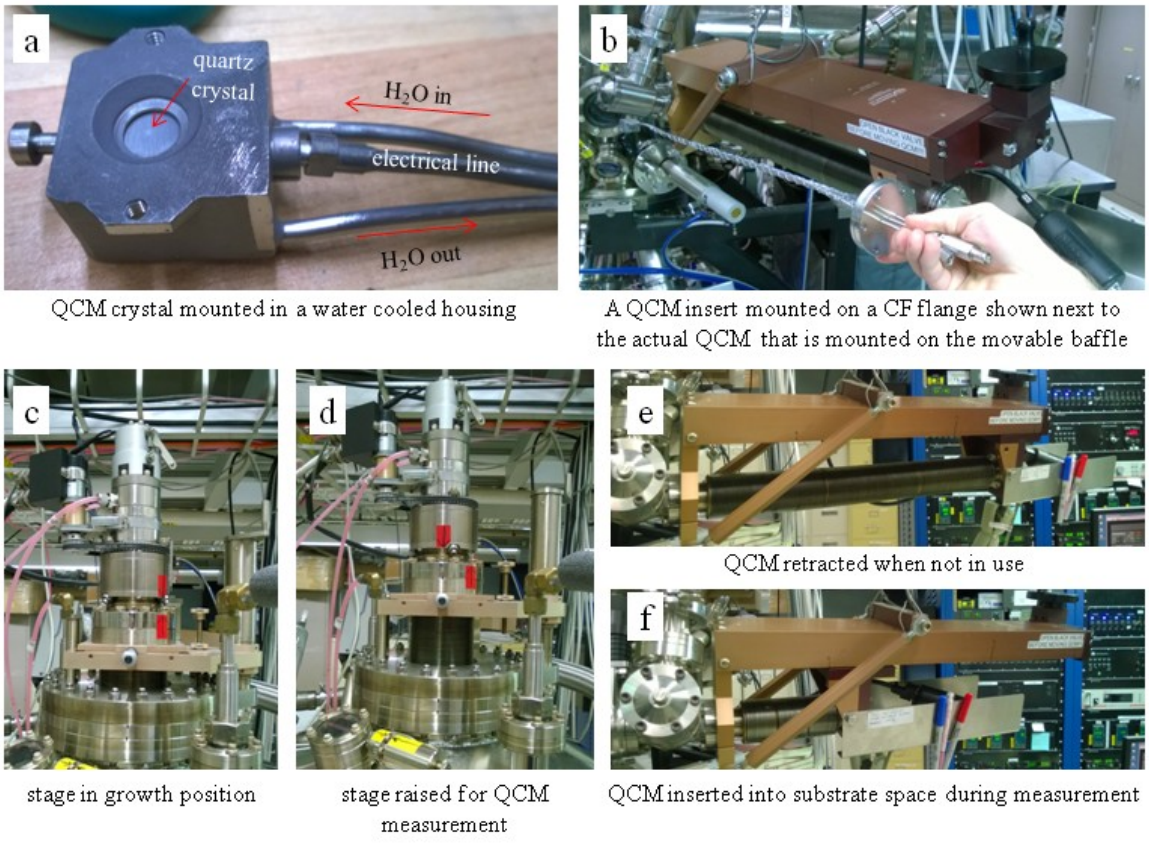


Figure 2-17 Images of the QCM system on the MBE. a The measurement head where the quartz crystal is mounted, attached to long electrical and cooling lines, which are mounted on a CF feedthrough flange that is mounted on a retractable baffle (**b, e, f**). To insert the QCM, the growth stage is raised (**c-d**) and the QCM is inserted (**e-f**)

For a quartz crystal of thickness d_q, and elastic wave velocity v_q, the resonance frequency is given by $f = v_q/(2d_q)$. If a film is deposited on one of the surfaces of the quartz crystal, then the thickness expands to $d_q + \Delta d$, where $\Delta d = \Delta m/(A\rho)$, Δm is the added mass, A is the film area, and ρ is the mass density of the film. Now, in the small thickness limit, a film with thickness Δd, will produce the same shift in resonance frequency as adding the equivalent thickness of quartz Δd_q, which yields a change in frequency given by $\Delta f = v_q/(2(d_q + \Delta d)) - v_q/(2d_q) \approx v_q/(2(d_q + \Delta d_q)) - v_q/(2d_q)$; this can be expanded in the limit of a small change in thickness to give $\Delta f \approx -\Delta d_q v_q/(2d^2) = -f^2\delta m/(C\rho A)$, where the resonance frequency constant $C = v_q/2 \approx 1656$ kHz-mm. With the knowledge of the geometry and material properties of the quartz crystal (C, and A) and the mass density of the deposited element (ρ), one can measure the change in resonance frequency, which

together yields the accumulated mass over a given period of time. This can then be transformed to a flux by converting the mass density to atomic number density, which gives information as to how many atoms are arriving at the substrate per unit of time. This method can be quite sensitive because changes of ~ 1 Hz at a frequency of 10 MHz are relatively easy to measure, which can easily give thickness in the range of $\lesssim 1$Å (adapted from [144]Ohring, 2002).

The previous derivation was in the limit of small thickness of the accumulated layer, but more realistic models are typically used, which are programmed directly into the commercially available control units. In a typical measurement, one simply inputs the element, its mass density, and *Z*-ratio, which is a measure of the films' density and shear modulus relative to that of the quartz (usually close to unity), and measures the flux for a given element at a given temperature. However, the measurement procedure is more complex because spurious effects are always present.

The two main effects which throw off QCM measurements are geometrical effects, and the heating of the quartz crystal due to radiation from hot thermal cell. The former is usually corrected for by determining a calibration factor, which empirically accounts for any tilting of the quartz crystal relative to the substrate, and any difference in position between the quartz crystal and the substrate. Radiative heating can introduce inaccuracies to the flux measurement because the changing temperature naturally shifts the resonance frequency, which would appear as false accumulation of mass on the crystal. This is difficult to quantify because the heating process occurs in a non-controllable way. To minimize this effect, the quartz crystal is mounted inside a cooled housing, with only minimal surface exposed. The QCM system is cooled by its own dedicated chiller, which is less susceptible to spurious changes in coolant temperature due to changes in the heating load (as with a shared chiller).

To account for this thermal effect and any geometric effect quantitatively, a calibration factor is used for each element, which is determined by crosscalibration with RBS. To measure this factor, one simply grows a test film, after which, flux measurements are made consecutively

on each element by QCM. To account for the thermal effect, the QCM measurement for a particular element is started, and after a given period of time the flux is recorded. Then this value is compared to the true flux value obtained by RBS, and the ratio of the two gives the calibration factor, which accounts for both the thermal effect and any geometric effect. In the case of Bi and In, which are the two elements whose flux values are most relevant for this work, the calibration factor between QCM and RBS are ~84% and ~86% respectively, which have been measured at 5 minutes after opening the shutters. This means that the true values of the Bi flux is ~84% that of the value measured by QCM taken at 5 minutes. This method is quite reproducible, with deviation in the QCM reading at the 5 minute point being in the range of ~1%, and, therefore, values obtained from averaging many measurements should yield a flux within about $\lesssim$ 1%.

2.6.2 Rutherford backscattering spectroscopy

The basic principles of Rutherford backscattering spectroscopy or RBS extends back to the early days of modern physics with Ernest Rutherford's famous experiment (conducted by Hans Geiger and Ernest Marsden) where alpha particles (He^{2+}) were impinged on a thin Au foil and the backscattered ions indicated the existence of a dense nuclear core. Although similar in principle, modern RBS systems require much more complex instrumentation. The Laboratory for Surface Modification at Rutgers University hosts a 1.7 megavolt Tandetron accelerator that is used for RBS measurements (see Figure 2-18). This system is a tandem accelerator, which accelerates the ions, usually alpha particles, in a two-step process that shares the same high voltage terminal. Initially the source of alpha particles (helium gas) and the target are near ground potentials, which are separated by the positively charged high voltage terminal. The helium gas is sent through an ionizer that negatively ionizes them by adding an electron. With a net charge they can then be accelerated towards the positive terminal; near the positive terminal the helium ions pass through an electron stripper, which removes electrons, thereby changing the net charge on the ions from 1^- to either 1^+ or 2^+. This change in ionization allows them to be further accelerated as they move

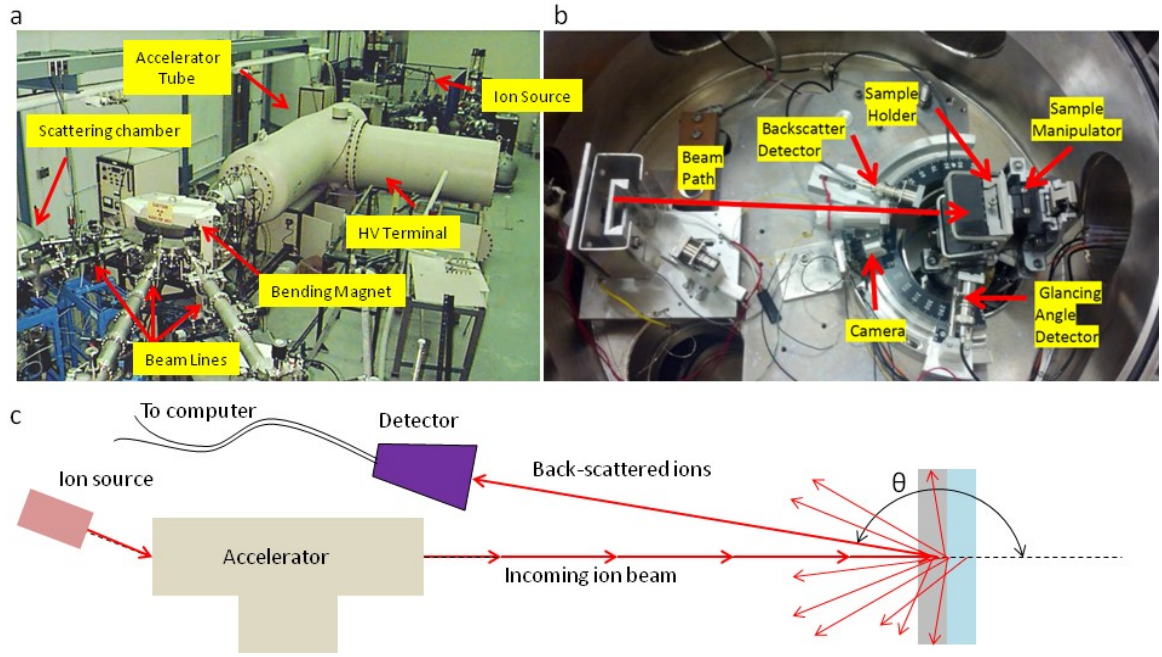


Figure 2-18 A typical Rutherford backscattering spectroscopy experiment. a-b The Tandetron accelerator housed at the Laboratory for Surface Modification at Rutgers University. **a** A wide angle view of the entire system, which is about ~50 feet in length. **b** An image inside the scattering chamber (**a-b** adapted from [17]Bartynski, 2013). **c** A schematic of a typical RBS experiment, where the ions are generated, accelerated, then impinged on the sample. After backscattering they are finally measured by a detector at a fixed angle θ, as a function of energy.

away from the positive terminal toward the target that is at ground potential. The ions are then impinged onto the sample surface at normal incidence, and the backscattered ions are measured at 165° relative to the incoming beam by a silicon detector.

Upon reaching the target, the ions interact with (scatter off of) the atoms in the sample by simple Coulomb repulsion, which conserves the total energy and momentum (this is, of course, true only if there are no induced nuclear reactions near this energy, which can occur in this MeV energy range for certain nuclei; for example O^{16} has a nuclear resonance at around ~2.5 MeV). Assuming no nuclear reactions take place, we can apply the hard-sphere approximation to the kinetics, which yields $E_1 = kE_0$ where E_0 is the incoming energy prior to scattering and E_1 is the measured outgoing energy after scattering. k is called the kinetic factor, and is given by

$$k = \frac{\left(\sqrt{m_1^2 - m_2^2 \sin^2\theta} + m_1 \cos\theta\right)^2}{m_1 + m_2}$$

where θ is the angle of the backscattered ion, with mass m_1, relative to the incoming ion beam (see Figure 2-18c), and m_2 is the mass of the nuclei in the film. The Coulomb potential between the nuclei gives the standard Rutherford cross-section

$$\sigma = \left(\frac{Z_1 Z_2 e^2}{4E_0}\right)^2 \frac{1}{sin^4\,\theta/2},$$

for a target atom with nuclear charge Z_2, and incoming ion with nuclear charge Z_1. The cross-section and kinetic factor are sensitive to the elemental species in the films that scatter the incoming ion; this allows RBS to probe the chemical composition of films, and, as we will show next, these together give a clear signature for many additional properties of the films.

From the Rutherford cross-section together with the kinetic factor one can see the power of the RBS technique. If the incoming ions scatter off the surface atoms, then the scattering parameters will be determined by k and σ, as given above, with the incoming energy taken to be the beam energy. However, the penetration depth of the ions can be in excess of 1 μm, and as the ions penetrate into the sample, they continuously lose energy due to the diffuse interaction with the electrons surrounding the nuclei; upon backscattering with a nucleus, they continuously lose energy as they emerge out of the sample. This continuous loss of energy enables RBS to probe the composition of the film with depth resolution, which is manifest in the line shape of the energy versus intensity plot.

To understand how RBS provides depth resolution, we can take a specific example for a bilayer film composed of one layer of Au and another layer of Cu on an Al_2O_3 substrate. If RBS spectra are taken from Al_2O_3 – Au – Cu, and Al_2O_3 – Cu – Au the peaks will appear at different energies. Figure 2-19a shows the simulated spectra from ~2 nm Au on ~10nm Cu (red upper curve), and ~10 nm Cu on ~2 nm Au (blue lower curve). In comparing these curves, one can notice several distinct features which allow one to determine the composition of the film with

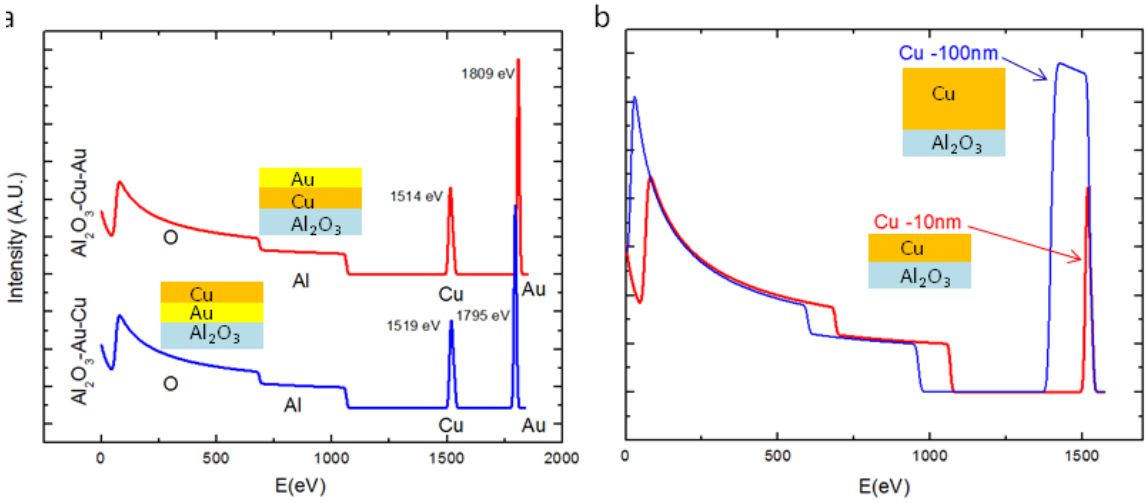


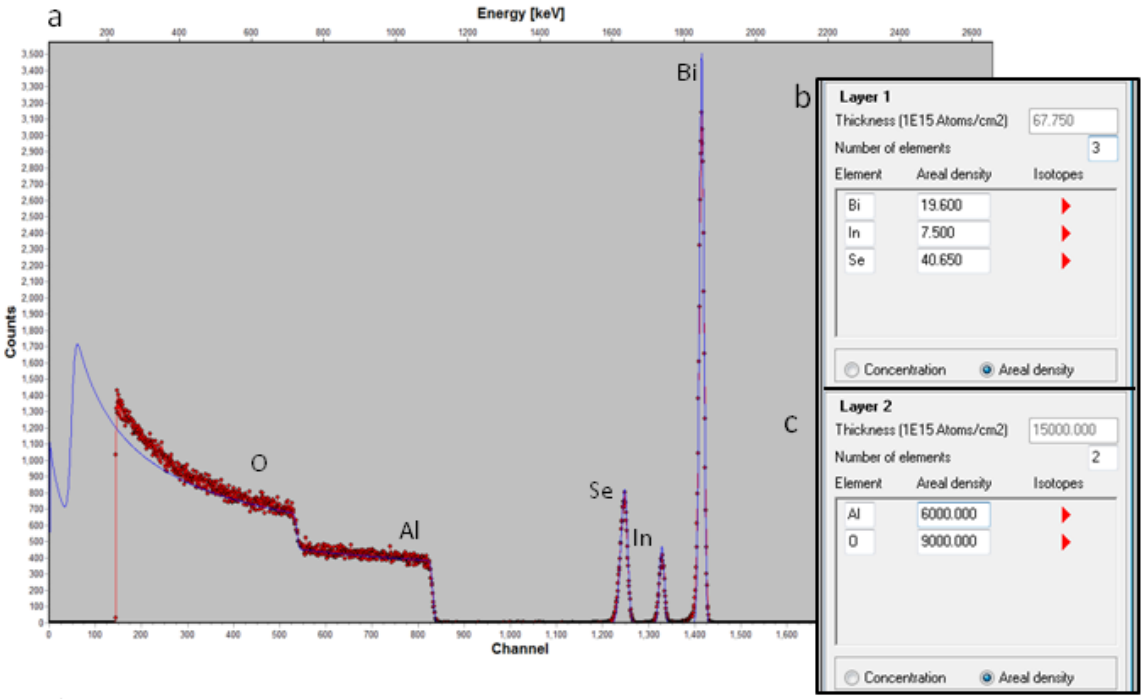
Figure 2-19 Simulations of RBS spectra for various film structures. a Two RBS spectra simulated using SIMNRA (see text) for Al_2O_3 substrates with Au – Cu (lower blue curve), and a Cu—Au (upper red curve) films. This shows the shift in energy for the Cu and Au peaks based on the layering sequence of the films. **b** The spectra (normalized to the Al edge) for a thick and a thin Cu film (blue and red curves respectively). The thin film shows a single peak, while the thick film shows a peak that is elongated and plateau-like due to the continuous energy loss of the ion as they travel deeper into the film, which is also seen in the shift of the substrate peaks to lower energy.

depth resolution. The first feature to notice is that when the Cu layer is on top of the Au, then the gold peak is shifted to a lower energy (similarly for the Cu peaks when the Au is on top). This is because the ions that are able to scatter off of the gold have already lost energy as they penetrate through the Cu layer. The second thing, which is difficult to see from the spectra, is that the intensity of Au peak is increased by about ~10% when underneath the Cu layer. This can be attributed to Rutherford cross-section, which is inversely proportional to the square of the incident energy, and, consequently, an ion with smaller energy (i.e. an ion that has passed through the Cu layer) is more likely to backscatter than if the same ion had a higher amount of energy (this enhanced scattering at lower energy is technologically very important in the ion implantation step in the manufacturing of Si transistors). Furthermore, this enhancement at lower energies is the reason why the Al and the O in the substrate appear as long plateaus that rise with decreasing energy rather than peaks like the Cu and the Au. Similarly, this elongation of the backscattering peak can be seen in the peaks that originate from a single Cu film on a Al_2O_3 substrate. As shown

in Figure 2-19b, as thickness of the Cu layer is increased from ~10 nm to ~100 nm, the peak becomes an elongated plateau. Similar to the substrate peaks, this is due to ions that are scattered from increasing depths of the film, which have a significantly higher probability to scatter due to the loss of energy.

The major use of RBS in this work is as an independent calibration of elemental flux by determining stoichiometry and film composition. For simplicity, typical film thicknesses are kept under $\lesssim$ 30 nm where the surface-layer approximation is said to be valid, and prevents overlapping of the elemental peaks. In this approximation all atoms are assumed to be near the surface layer, which prevents the need for more complex models for determining the films' parameters. The number of ions that are expected to be backscattered off a particular element (i = element) in the film is given by $A_i = \sigma_i \Omega Q N_i$ where σ_i is the cross-section for a given element at a given set of beam parameters (which is typically found in a table), Ω is the solid angle of the detector, Q is the total number of incoming ions (the beam current multiplied by the exposure time), and N_i is the areal density of the particular element in the film given in units of atoms per area. From this, one can determine the stoichiometry and areal density of the film simply by integrating the spectrum for a particular element (which yields the total number of counts A) and using the known cross-section and the beam and detector parameters. However, in practice numerical software packages are available which simplify this analysis.

SIMNRA is the standard software used to analyze backscattering spectra, and generates simulated backscattering curves for direct comparison to data. This allows one to calculate film parameters, and gives a qualitative comparison to the experimental data. To use SIMNRA to determine a film's parameters requires some guesses as to the composition of the film; SIMNRA calculates RBS spectra by inputting the elemental composition in terms of layers, which requires the user to select the elements and their areal density, then, based on this information, SIMNRA simulates the spectra. Comparing the simulated spectrum to an experimental spectrum is subjective, so the following method has been developed to determine the makeup of the film: (1)



d

	areal density	
Bi	1.96E+16	1/cm^2
In	7.50E+15	1/cm^2
Se	4.07E+16	1/cm^2
totaltime	1340	sec
indium%	27.7	%

flux	
1.46E+13	1/cm^2s
5.60E+12	1/cm^2s

atomic density	
1.35E+15	1/cm^2
1.44E+15	1/cm^2

	effective thickness	
	14.5	QL
	5.2	QL
total:	19.8	QL

Figure 2-20 An RBS spectrum used in a calibration experiment on $(Bi_{1-x}In_x)_2Se_3$ where *x* was targeted to be ~ 30%, and a thickness of ~ 20 nm (1 QL ≈ 1 nm), which was used to grow films within ~ 1% of the targeted composition. a The spectrum and numerical fit. **b-c** The input windows in the SIMNRA program for the film layer (**b**) and the substrate layer (**c**) used in generating the simulated spectrum for this samples. **d** The actual calculation to determine flux (given by atomic areal density divided by total deposition time).

Using the numerical integration function find the area under the experimental curve, and subtract off a background value, which is typically ~1-2 counts per channel times the number of channels. This gives the experimental number of counts that is used to match the simulated number of counts. (2) Adjust the areal density of each element until the total counts under the simulated curve matches the experimental number of counts minus the background. (3) If these parameters do indeed match the film parameters, then the leading edges of the experimental and simulated

spectra should exactly line up for all elements in the film, and for the elements in substrate the leading edges should also line up, as well as the top of the curves, which is a measure of the accuracy of the beam current. With this method the RBS spectra are accurate within about ~ 1%, and are free from any subjective errors. A sample of this method can be seen in Figure 2-20 for $(Bi_{1-x}In_x)_2Se_3$ where the composition x and total thickness were determined by RBS.

The previous method assumes the surface-layer approximation, which neglects any variation in the composition. To account for compositional variation in the thickness, one can extend this method by fixing the total composition and redistribution of the elements in various layers. This can be done until the shape of the experimental curve coincides with the shape of the simulated curve. Lastly, RBS can be sensitive to the structural by the so-called channeling effect. This occurs when the surface of the film backscatters ions and effectively creates ion shadows because the ions that are not backscattered by the surface layer and are spatially located in-between the atomic position and can be moved into the film unimpeded (no backscattering) by moving down these channels in the lattice. This effect decreases the total backscattering, and by comparing the total counts for a spectrum measured with the sample that was continuously rotated in a random manner (to eliminate the channeling effect), to one where the channeling effect is at its maximum (usually along one of the principle crystal axes), one can determine the quality of the crystal structure (e.g. no channeling effect occurs for a highly disordered film). Although, both of these effects are not typically looked at in an RBS experiment used for calibration purposes, these techniques were used successfully in the case medium energy ion scattering (a specialization of RBS which is more sensitive to interfaces) on thin Bi_2Se_3—In_2Se_3 heterostructures, which were used to quantify the level of diffusion between In and Bi at the interface ([119]Lee *et al.*, 2014).

2.7 Structural characterization

Achieving high quality thin films is closely tied to the ability to probe the structural properties by many different techniques, which when used together, give a global view of how the growth conditions affect the films structure. Structural measurements fall into two categories: the first are measurements that probe the real space morphology or crystal structure with the most common techniques being atomic force microscopy (AFM), which gives a broad view of surface morphology, and scanning tunneling microscopy (STM) and transmission electronic microscopy (TEM), which give atomic scale information. All of these techniques image the local crystal structure from the nanometer to the atomic scale. The second type of structural measurements are diffraction techniques, which rely on the wave nature of particles or light that produces an interference pattern when the waves pass through a crystal. These patterns can be used to reveal many details about the underlying crystal structure. The most common technique that is used in nearly every branch of science is X-ray diffraction (XRD), which can be used to probe as few as tens of nanometers in depth at lower energies, to thicknesses of meters at the highest energy ([149]Pecharsky and Zavalij, 2003). In regard to thin films, electron diffraction techniques are more useful because electrons penetrate only a few nanometers, which is better suited for resolving surface features ([50]Cohen and Ichimiya, 2004). In this section we will introduce the techniques used to analyze the structural properties of the films used in this work, and how they are feedback, and subsequently used, to optimize the growth conditions.

2.7.1 Reflection high energy electron diffraction

In regards to MBE, the single most important technique that enables the growth of high quality atomic scale films is in situ reflection high energy electron diffraction (RHEED). With RHEED one can monitor in real time the crystal structure of the terminating layer of a film; this allows many growth trials to be performed quickly, in sequence (usually on the same sample), and yields unambiguous details of morphology and film structure. With this capability, one can pinpoint the

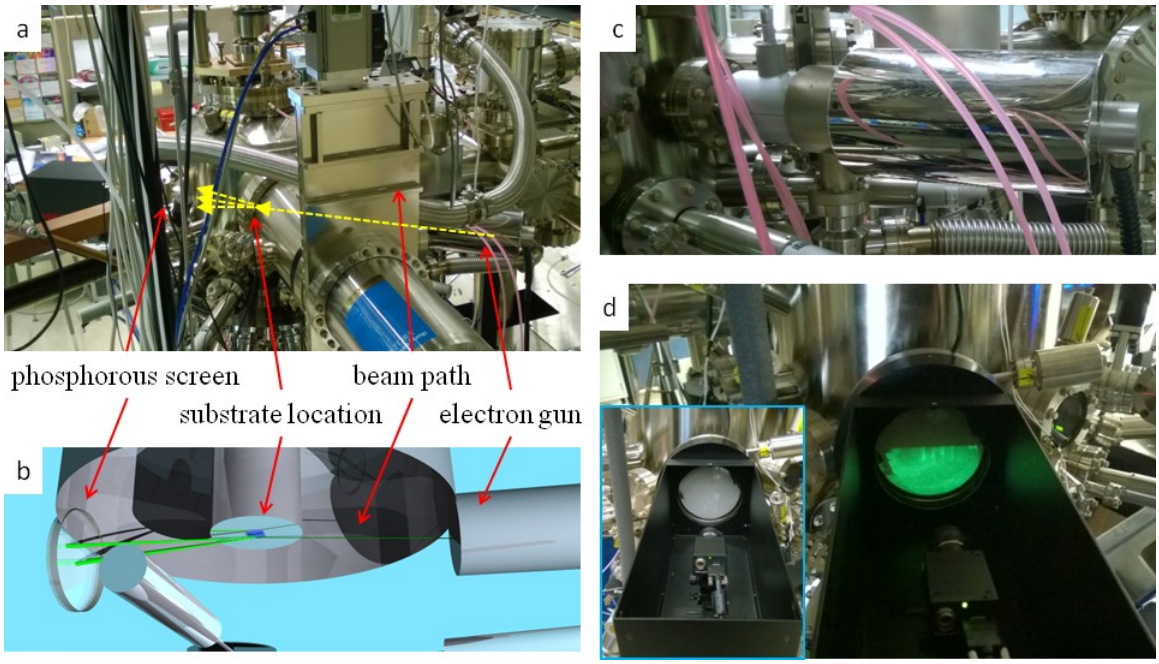


Figure 2-21 Images and schematics for in situ reflection high-energy electron diffraction. a-b The experimental geometry for RHEED. **c** The electron gun that is used to generate the electron beam. Visible in this image is the CF flange used to mount the gun on the MBE chamber, the power cable, and the differential pumping line which is connected to the prep chamber used to keep the pressure at the filament low when the chamber is filled with a reactive gas such as oxygen. **d** The light proof box (with top removed), the CCD camera, and the phosphorus coated screen that is used to image the diffraction pattern with (inset) and without the lights on.

optimal growth conditions usually within a few days to a week. In comparison to ex situ structural measurements, to gain similar information as obtained with RHEED, one must use a combination of structural probes at the end of each growth trial, which may take more than an order of magnitude as long, and cannot be performed as easily as RHEED.

As shown in Figure 2-21, RHEED is performed in a glancing angle geometry, where a 5-50 keV electron beam is impinged at a shallow angle (~1°); in the experiments presented here the beam energy used during film growth is fixed, and is always kept at 10 keV for uniform comparison between samples over time. The beam width is ~ 1 mm, and, therefore, the diffraction image that is produced is an average over a large portion of the sample. After the beam diffracts off of the sample surface, it is impinged and fluoresces on a phosphorous screen that is mounted on a CF flange with a window (Figure 2-21d). The screen is then imaged by a CCD

camera that is housed in a light proof box that is mounted directly on the outside of the CF flange, and the image is then read and analyzed by the computer using commercially available RHEED software (k-Space Associates, Inc.).

One of the most important features of RHEED that makes it especially useful is its surface sensitivity, which is enabled by the short mean free path of electrons in materials. The mean free path of electrons in solids follows a nearly universal behavior, which is roughly material independent. This behavior is summarized in the so-called universal curve, shown in Figure 2-22, which shows a minimum mean free path of ~5-10Å at an energy ~50 eV; this short mean free path is what gives RHEED its surface sensitivity. This is because the electrons that form the diffraction pattern are those with the same wavelength; thus, the electrons that are inelastically scattered cannot contribute. Therefore, the diffraction image is formed by electrons that are within the mean free path of the surface, and if the energy is targeted at the minimum of the universal curve, then the image formed is due only to the surface electrons. In the case of low energy electron diffraction, where the electron beam is at normal incidence, the energy is directly

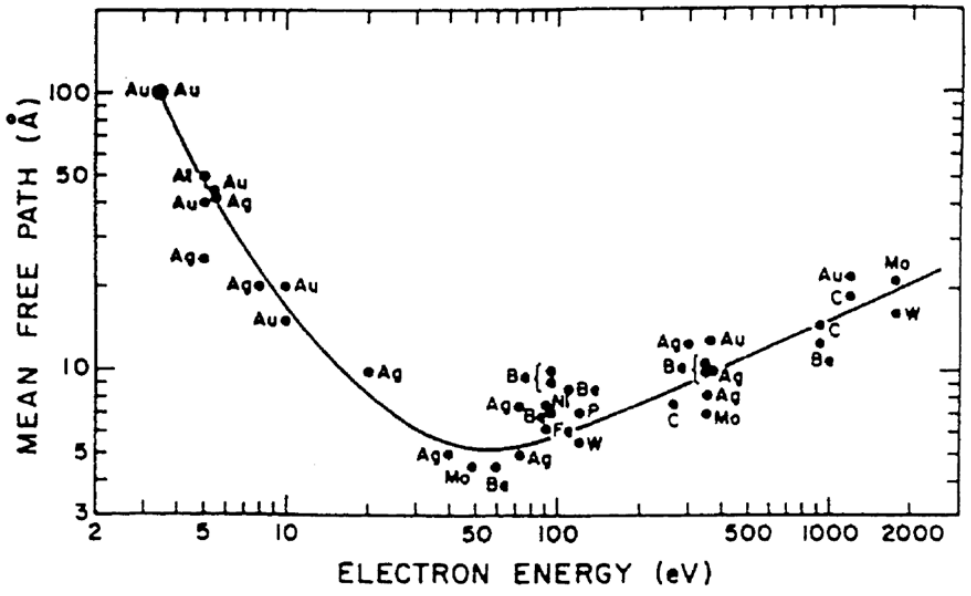


Figure 2-22 The universal behavior of the mean free path of electrons in solids as a function of energy for various elements. (Adapted from [214]Zangwill, 1988)

targeted at ~ 50 eV, and yields a surface diffraction pattern; in contrast RHEED, with its higher electron energy and low scattering angle is more complex, and satisfies the universal curve by its kinetic energy in the perpendicular direction ($E \sim E_0 sin^2(\theta) \sim$ 10 keV × $sin^2(1°)$ ~ 100 eV → ~ 10 Å), while in the direction parallel to the surface, the mean free path can be much longer ($E \sim E_0 cos^2(\theta) \sim$ 10keV$cos^2(89°)$ ~ 10keV → > 100 Å), which, as we will see next, aids in distinguishing the surface morphology.

One of the most important features of RHEED is its sensitivity to surface morphology, and particularly roughness. Based on the diffraction pattern, one can easily distinguish a surface that is 2D and flat, versus one that is rough with 3D features. The most elegant way to see these is through the geometric Ewald-construction for the diffraction condition shown in Figure 2-23. An incoming beam of either photons or electrons (we will focus on electrons here) can be described as a plane wave with wave vector give by $\boldsymbol{k_0}$, which gives the beam direction and wavelength. When the beam is impinged upon a crystal lattice, the electrons diffract off the lattice positions along certain directions. The interactions that result in diffraction are elastic, in that, they preserve energy and momentum. The magnitude of the momentum is conserved ($|\boldsymbol{k}_0| = |\boldsymbol{k}_{Final}|$); however, the direction of the momentum is changed, which can result in $\boldsymbol{k_{Final}}$ taking any angle. Along most directions the amplitudes of the scattered electrons add incoherently, and do not contribute to the formation of the diffraction pattern. However, along certain crystal directions the difference in the path lengths traveled by the electrons are an integer multiple of the wavelength, and, therefore, they add coherently; this is known as the Bragg condition. Ewald gave a geometric method to easily determine the possible directions a diffraction peak will occur along ([67]Ewald, 1969), which can be seen by the following.

A crystal lattice in real space can be transformed into *k*-space to another crystal lattice with *k*-space lattice vectors given to be $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, and the incoming beam of electrons has wave vector $\boldsymbol{k}_0$. As shown in Figure 2-23, Ewald prescribed that if the incoming wave vector is graphically terminated on a lattice position, and a circle is traced out around this point with radius

given by $|\boldsymbol{k}_0|$ (called the Ewald circle or sphere for the case of a 3D lattice, as we will see), then where this circle intercepts a lattice point will correspond to a diffraction peak. This is true because for elastic collisions the magnitude of the momentum is conserved ($|\boldsymbol{k}_{Final}| = |\boldsymbol{k}_{R,i}|$) and, therefore, the outgoing wave vector will also lie on the Ewald circle. Now when the outgoing wave vector terminates on another lattice point, then the Bragg condition is automatically satisfied simply because the difference between the incoming and the outgoing wave vectors are a integer multiple of the lattice vectors, $\Delta\boldsymbol{k} = \boldsymbol{k}_0 - \boldsymbol{k}_{R,i} = \boldsymbol{G}_i$, and a diffraction peak will occur along this direction.

When this is generalized to 2D and 3D, we see that there will be distinct differences in the RHEED pattern formed. For the case of a 3D crystal lattice, transformation to *k*-space yields

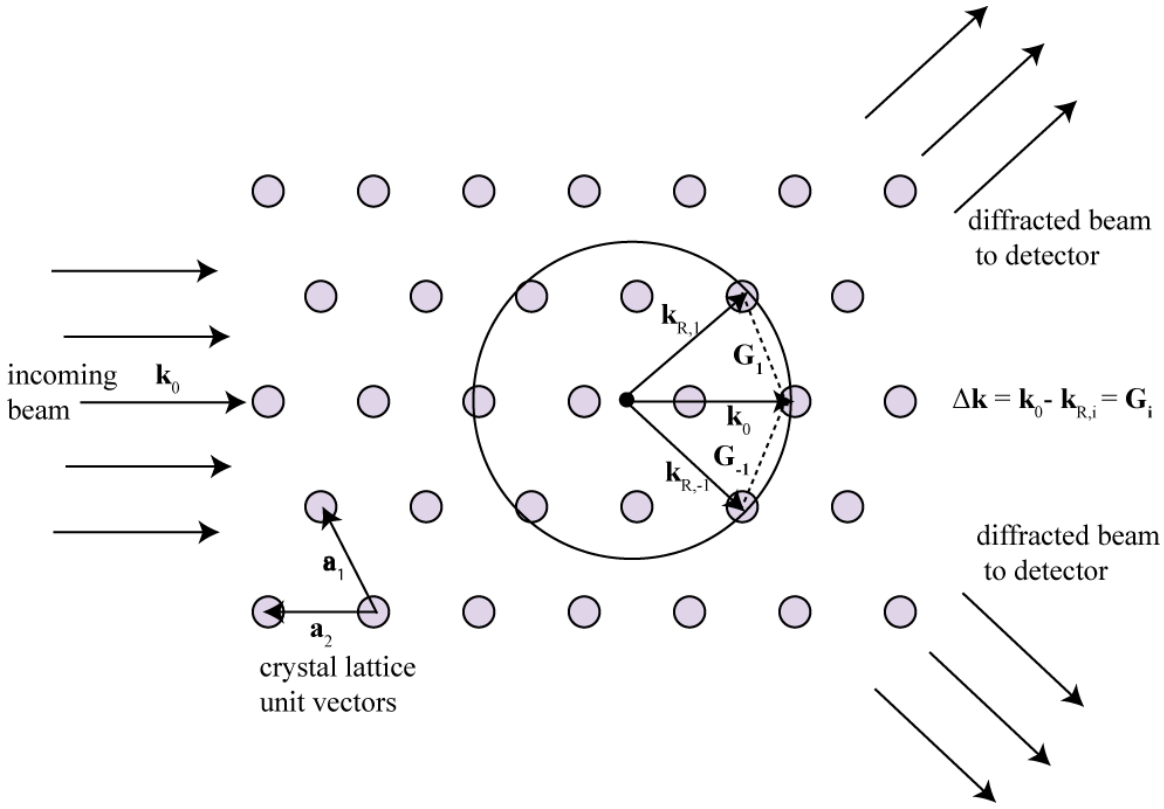


Figure 2-23 The Ewald-construction for diffraction of an incoming electron beam by a crystal lattice.

another 3D lattice, whose lattice points are localized in 3D. However, for the case of a 2D lattice whose real space points are localized in 3D, the *k*-space analog is a 2D lattice whose points are localized in 2D, but delocalized in the third dimension: a 2D lattice that is localized in the *z* direction can be approximated by a delta-function $\delta(z)$; transitioning to *k*-space involves a Fourier transformation, which, for a 2D lattice yields $F[A\delta(z)] \rightarrow A$, a constant. In this approximation, a real space 2D lattice in *k*-space should be thought of as a series of rods, called reciprocal rods, whose axes are along the *z*-direction. Now, if we imagine extending the Ewald construction to both 2D and 3D crystals, a distinct difference will appear in the RHEED images that can be seen as follows.

In a RHEED experiment the beam comes in at a very shallow angle, and if the surface of the crystal is atomically flat, the electrons only enter the top layer of atoms. The subsequent diffraction patterned is formed along the directions where the reciprocal rods penetrate the Ewald sphere with the resulting pattern having peaks that are located on a ring (see Figure 2-24a- b). Upon rotation of the sample, the peaks track vertically on the screen in a continuous way, which is due to the fact that the 2D reciprocal space lattice is composed of rods, rather than points; this can be seen by comparing how, upon rotation, Figure 2-24b changes to d.

When the surface is not atomically flat and has 3D structures, the electrons penetrate these structures, and form a transmission pattern. The corresponding Ewald-construction is a 3D lattice in *k*-space, and the diffraction is satisfied when the Ewald sphere intersects this lattice, also forming peaks that resemble the underlying crystal structure. When the sample is rotated, the behavior of this transmission pattern is distinctly different from that of a 2D lattice; as the Ewald sphere is rotated, the peaks do not change in a continuous fashion; rather, they blink-on and blink-off as the lattice is rotated through the Ewald sphere. From this character, one can immediately determine whether the surface of a thin film is atomically flat or composed of 3D structures.

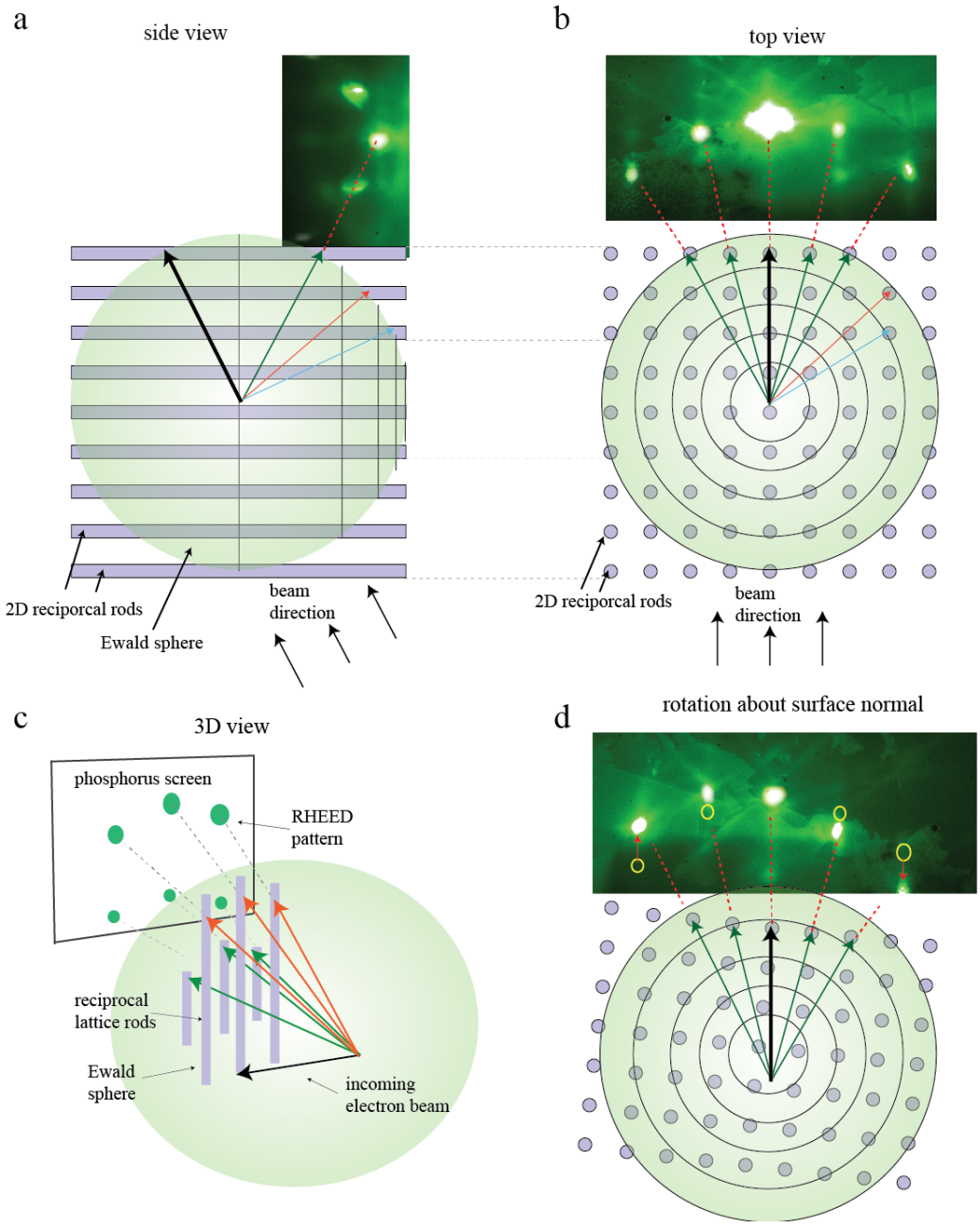


Figure 2-24 A schematic of a RHEED pattern formed from the diffraction of an electron beam off of a 2D surface. Perspective from the side (**a**), top (**b**) and a 3D view (**c**). If the surface is 2D, then the diffraction peaks track vertically when the sample is rotated (**d**), while for a 3D surface, the peaks will blink on and off. In **d** the open circles are the positions of the peaks in **b**.

Qualitative information such as surface morphology, crystal structure, and its overall quality, can be determined simply by examining the RHEED patterns. Figure 2-26 shows several examples of RHEED images for various surface morphologies and crystal structures with

different crystal qualities. The highest quality crystals are typically the single crystal substrates which are used as a temple for growing the epitaxial films. Figure 2-26a and b show Al_2O_3 and Si(111) substrates, which are the dominate substrates used for the films grown in this work. In addition to the main peaks that are visible, the curved diagonal lines, call Kikuchi lines, are due to a combination of an inelastic scattering event with subsequent elastic Bragg scattering. These lines are formed when electrons enter the crystal and suffer an inelastic scattering event where the electron loses energy and changes direction. Along certain crystal directions these inelastically scattered electrons can satisfy the Bragg condition while scattering out of the crystal, which results in the continuous streaks emanating out of the diffraction peaks. This type of scattering requires long mean free paths to fulfill the Bragg scattering out of the crystal, and, therefore, requires high crystal quality; the presence or absence of Kikuchi lines, therefore, gives information about the crystalline quality. This can be seen in the comparison of the Kikuchi lines in Figure 2-26a-b to those in d and e: the lines in a and b are sharp and easily visible, while in d no Kikuchi lines are visible, and in d the lines are more broad; this indicates that the films in d and e are more disordered than the single crystal substrates in a and b.

Another indication of structural quality is the shape of the main diffraction peaks. Ideal 2D crystals will have diffraction peaks that are sharp and localized, and if the structural quality is worse, and the surface is rough, then the main diffraction peaks will elongate in the vertical direction and become more diffuse overall (more precisely, this occurs when the mean free path of the electrons becomes shorter than the film's mean terrace size). This can be seen again by comparing the RHEED images in Figure 2-26a, b, d, and e. In the single crystal substrates of a and b, the surfaces are highly 2D with large flat terrace that extend hundreds of nanometers between unit cell step edges, which is indicated by the sharp and highly localized diffraction peaks. In d, the RHEED image is formed by the low temperature growth step of Bi_2Se_3 (the rational for this will be discussed in 2.8), which acts as a seed layer, but is structurally and morphologically poor. Upon annealing the Bi_2Se_3 film to a higher temperature, and subsequently

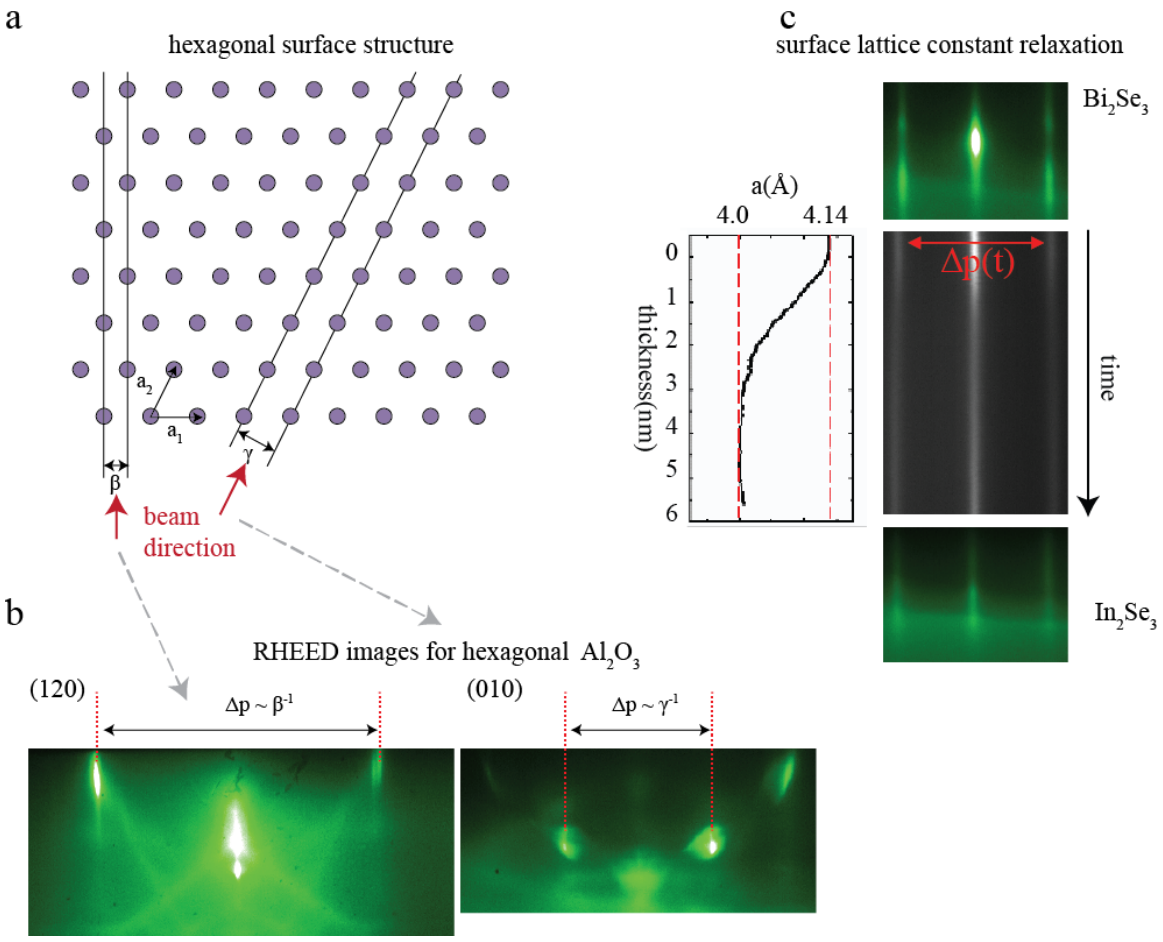


Figure 2-25 Various crystal parameters obtained from RHEED images. a A schematic hexagonal terminated crystal where there are two periodic structures. **b** RHEED images along with the electron beam parallel to these direction (these show a common problem— the crackly texture is Se accumulation on the screen). **c** The change in lattice constant can be measured during the deposition; this data was taken as In_2Se_3 (bottom) was deposited on Bi_2Se_3 (top).

depositing more Bi_2Se_3, the structural quality improves and the surface becomes more 2D, which is indicated by the sharper, more localized diffraction peaks, and, as previously mentioned, the appearance of Kikuchi lines.

In addition to qualitative information, RHEED can give quantitative information about the surface structure, as well as, the growth rate of the film. By rotating the sample, and noting the angular separation between repeating patterns, yields the rotational symmetry of the crystal's surface structure; the most common symmetries being cubic symmetry (4-fold, 90° between repetitions) and hexagonal symmetry (6-fold, 60° between repetitions); in this work, TI materials

naturally grow along an axis which terminates with a hexagonal surface, and, therefore, the substrates used are almost all terminated along a crystal direction with a hexagonal surface.

The real space distance between the main peaks in the RHEED image gives information as to the lattice spacing. The diffraction pattern formed in a RHEED image is in *k*-space, which measures the inverse of the period of the real space repetition; from Figure 2-25a, we can see that for a hexagonal surface there are two unique periodicities, which we refer to as β and γ. In real space distances $\beta < \gamma$, but in *k*-space, which measures inverse distances, the peak separation

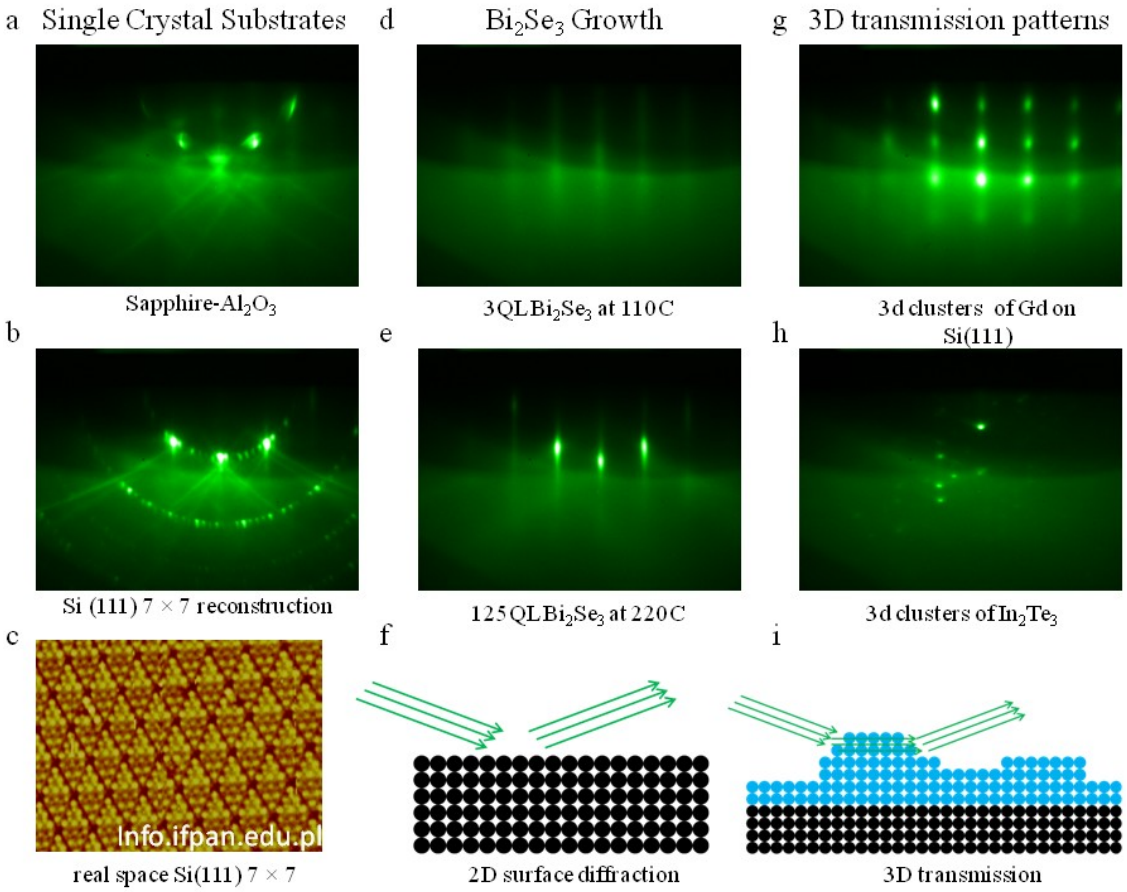


Figure 2-26 Various RHEED images to illustrate typical crystal structures of varying quality and various surface morphologies. a-b The image formed by single crystal substrates, with the real space image in **c** corresponding to the RHEED pattern in **b** showing a 7 × 7 surface reconstruction. **d-e** Two Bi_2Se_3 films with the film in **d** having lower quality than the film in **e**. The RHEED images in **a-f** are typical of a 2D surface, whereas, **h-i** are typical of 3D transmission patterns formed when the electrons transmit through 3D surface structures.

formed by the β periodicity is larger than that of the γ periodicity (see Figure 2-25b). This inverse relationship between real space and *k*-space can also be seen by comparing the RHEED images in Figure 2-26a and b, in which the Si (111) has a surface periodicity that is roughly seven times larger than that of Al_2O_3, and, therefore, has a 6 additional peaks between main peaks (seven times smaller periodicity in *k*-space).

To obtain a numerical value for lattice spacing one can use the wavelength of the electrons, and the geometry of the sample and the detector to find the angle between the central and first peak, then the lattice spacing can be calculated; however, in practice to find the lattice spacing for an unknown material, it is usually easier to simply take the ratio of the peak spacing and lattice constant for the unknown material with a known material (the substrate). This can be used to measure the change in lattice constant in real time during a deposition. Figure 2-25c shows the change in peak separation during the deposition of In_2Se_3 on Bi_2Se_3; from this it can be seen that the structure attains the lattice constant of In_2Se_3 within ~ 2-3 nm (2-3 unit cells).

Finally, in certain situations, information from RHEED images can be used to measure the deposition rate and, therefore, the flux. This can be done by monitoring the intensity of a RHEED peak during the growth of a film. If the growth proceeds in a layer-by-layer fashion, then there will be a change in intensity of the peak that oscillates with a frequency that is the same as the inverse growth rate (see Figure 2-27a). This is due to the increase in surface roughness that causes more diffuse scattering, which decreases the intensity of the diffraction peak. As the layer approaches full coverage, then the diffuse scattering decreases, which gives rise to the intensity of the peak recovering. The time for this recovery can then be used, with the knowledge of the surface crystal structure, to determine the elemental flux. However, this might not always be possible because RHEED oscillations may not occur. For example, in the step-flow growth mode, which occurs when nucleation proceeds in a continuous fashion starting from a step edge, and as shown in Figure 2-27b, the intensity does not oscillate in time because the surface morphology of the top layer does not significantly change. Layer-by-layer growth typically takes place at a lower

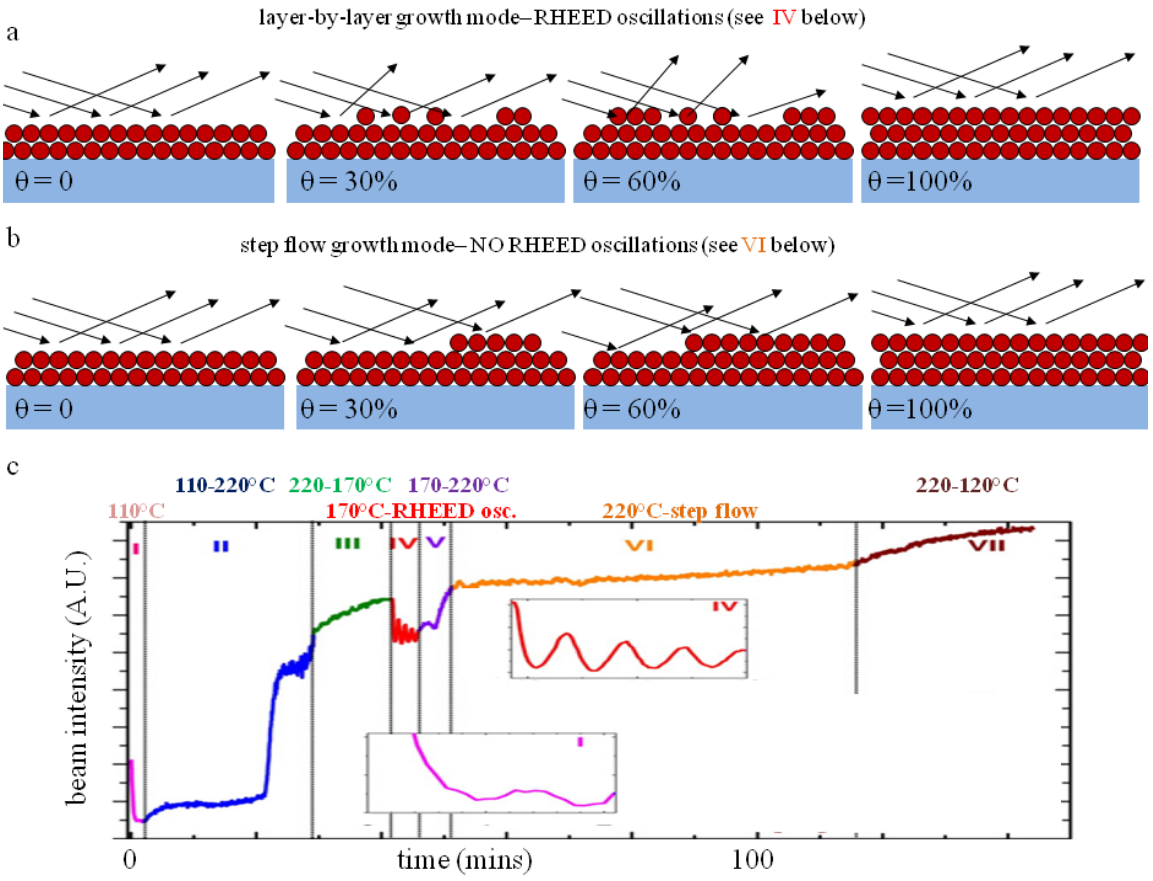


Figure 2-27 A schematic of the surface during deposition for the layer-by-layer and step-flow growth modes, respectively. In **a**, the intensity of the diffraction spot oscillates with coverage, while in **b**, the intensity does not oscillate. Both behaviors can be seen in the intensity verse time plot in **c** for Bi_2Se_3 deposited at various temperatures (Data from [16]Bansal *et al.*, 2011).

temperature than step-flow, because layer-by-layer growth proceeds from lower energy nucleation points, which, with increasing temperature, becomes unfavorable and nucleation begins to favor the step edges; since step-flow requires higher temperatures, the crystalline quality is typically higher than that of the layer-by-layer mode. In Bi_2Se_3 films, both modes can be seen: as shown in Figure 2-27c, layer-by-layer growth occurs at 170°C, while at 220°C, no oscillations are present indicating step-flow. It should be noted that one must be cautious of RHEED oscillations because they can give non-accurate values. RHEED oscillations may yield spurious data when the substrate temperature is inhomogeneous over the surface, which can result in the growth mode also being inhomogeneous across the substrate; this may cause the oscillation

frequency to be different from that of the deposition rate. Therefore, as with all techniques that can be used to measure the flux, RHEED oscillations must also be checked against QCM or RBS for confirmation.

2.7.2 X-ray diffraction

In addition to RHEED, X-ray diffraction (XRD) provides an additional structural measurement, which furthers confirms crystal properties and phase. As discussed in the previous section, RHEED provides detailed information for the in-plane structure, but provides little data about the out of plane structure. As a complement to RHEED, XRD provides detailed data for the out of plane structure, but it is harder to obtain in-plane structural data. In this work XRD was only performed on the samples early in the development of the growth recipe, and in cases where confirmation was need to insure the absences of any second-phase growth.

The canonical introduction to XRD is the development of Bragg's law that most people learn in their first year physics courses. Along this line, for an X-ray with wavelength λ one obtains the condition for coherent diffraction to be $n\lambda = 2d\ sin(\theta_n)$ where d is one of the

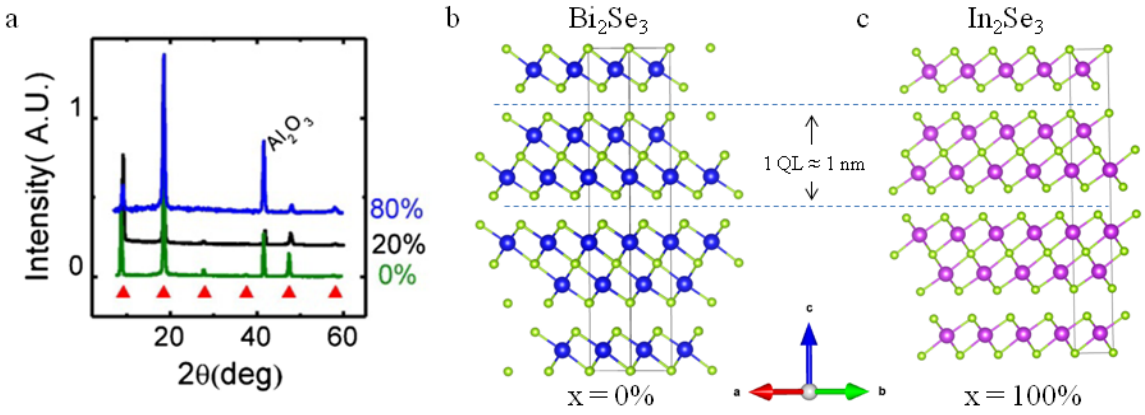


Figure 2-28 X-ray diffraction showing the QL peaks $(Bi_{1-x}In_x)_2Se_3$. a The XRD data for $(Bi_{1-x}In_x)_2Se_3$ films grown on Al_2O_3 substrates for $x \approx 0$, 20, and 80%. The red triangles mark the *c*-axis peaks for bulk crystals. (Data taken from [28]Brahlek *et al.*, 2012) **b-c** Schematics of the crystal structure for Bi_2Se_3 and In_2Se_3 with quintuple layers (QL) corresponding to the peaks in **a**.

periodicity of the crystal and n is an integer with a corresponding diffraction peak at angle θ_n. For the XRD measurements in this work, X-rays were generated by impinging electrons onto a copper anode, which generates continuous a bremsstrahlung radiation background with sharp *k*-alpha radiation (decay of an excited 2*p* electron to a vacant 1*s* state) with wavelength $\lambda_{k\alpha} \approx 1.5418$ Å. The *k*-alpha radiation was isolated by diffracting the beam off of a graphite monochromator. The diffraction data were measured by a Bruker HiStar multi-wire area detector mounted on a goniometer for measuring at different angles.

Typically the strongest peaks are those corresponding to the unit cell periodicities, and in thin films, this is usually the periodicity along the growth direction, which for Bi_2Se_3 is along the *c*-axis. As an aside, for thin films where there are heterostructure layers, there may be additional XRD peaks due to this additional periodicity; since the distances between these peaks are usually larger than the atomic spacing, they tend to have a narrower spacing, and measuring this spacing can give a measurement of the thickness, which can be used as an additional calibration of flux. Figure 2-28, shows XRD data taken from several $(Bi_{1-x}In_x)_2Se_3$ films with $x \approx 0$, 20, and 80%. From this, the main periodicity is from the (0,0,3*n*), due to the *c*-axis lattice periodicity; for Bi_2Se_3 and In_2Se_3 this corresponds to the unit cell layer called a quintuple layers (QL). Measuring the angular spacing yields a *c*-axis lattice constant of ~ 0.95 Å. The narrow widths of the peaks indicate the high quality; using the Scherrer equation it can be seen that the width of the peaks are limited by the thickness of the film, which indicates any domains or disorder are on a scale that is much larger than the film thickness of ~ 50 nm.

2.7.3 Atomic force microscopy

Atomic force microscopy (AFM) is a scanning probe technique that is most typically used to measure surface topography, but can be extended to study many other properties such as imaging magnetic or electric domains. In this work AFM is used to provide specific surface morphology

that is used in conjunction with RHEED, which provides non-specific details of the surface structures (*e.g.* 2D versus 3D).

AFM operates on the simple principle that a small force exists between two objects (the film's surface and a small, very sharp tip that is formed on the end of a cantilever) when they are brought in close proximity; this force is very short range, and composed of many forces such as Van der Waals, and dipole-dipole, but the actual form is not important. When the AFM tip is brought close to the film's surface, it feels this effective force, which deflects the cantilever that the tip is mounted on (Figure 2-29). The main operational principle is then to keep this force at a constant magnitude by monitoring how much the cantilever bends, which results in keeping the tip at a constant height above the surface; this then gives information about the surface topography (that is assuming the film-tip force does not change over the surface of the sample). In practice this is achieved by monitoring the position of the cantilever and tip by reflecting a

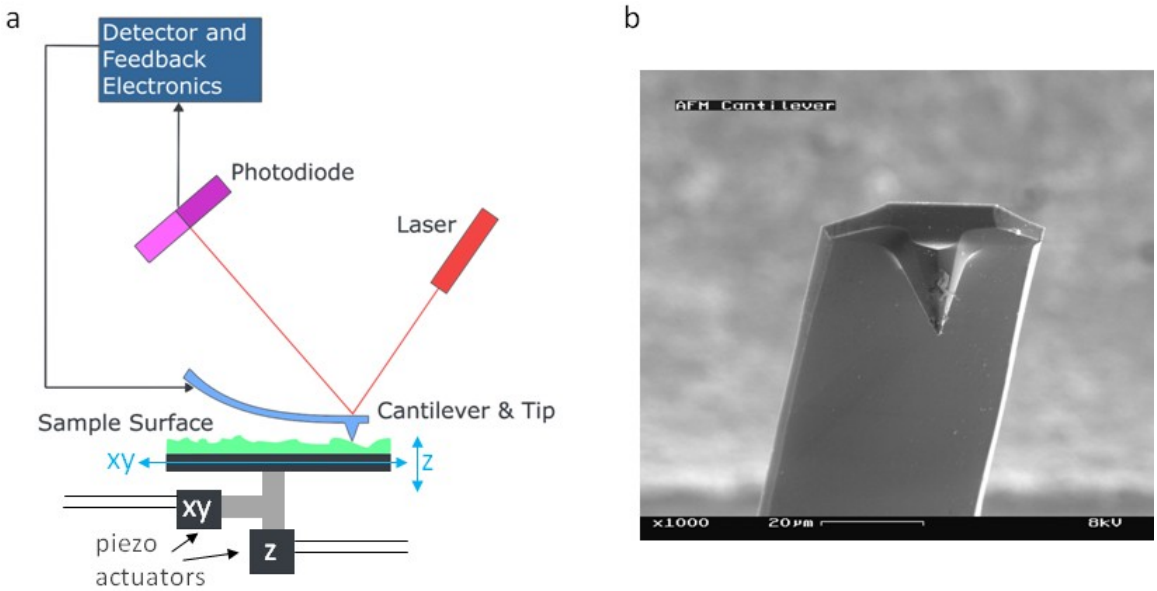


Figure 2-29 Atomic force microscopy. a A schematic setup for an atomic force microscope (AFM). The sample is mounted on piezo-electric actuators, which scan the sample in *x* and *y* directions (horizontal), and maintain a constant distance between tip and the sample in the *z* direction (vertical); this is controlled through a feedback loop to maintain the reflected laser spot in a constant position on the photodiode. **b** A scanning electron micrograph of the tip end on an AFM cantilever. (Adapted from wikipedia.org/Atomic_force_microscopy)

laser beam off the end of the cantilever, which is then read by a photodiode array that tracks the location of the reflected spot. This signal is then fed back to control the vertical position of the sample, which is raised or lowered to keep the laser spot at a constant position on the photodiode array. This is done with the position of the tip fixed in *xyz*, while the sample, which is mounted on *x*, *y* and *z* piezo-electric actuators, is scanned in the *xy* plane (horizontal), and in the *z* axis (vertical) the sample-tip position is held constant by maintaining a fixed location of the laser spot on the photodiode. The signal from the *z* piezo-electric actuator is then converted to a height, and recorded as a function of *x* and *y*, which then produces a topographical image. In this work, the films were scanned in the so-called tapping mode, which avoids dragging the tip across the surface, and thus avoids damaging the soft Bi_2Se_3 surface.

Figure 2-30 shows a sample of AFM images for Bi_2Se_3 films grown on different substrates. By RHEED, all these images show nominal 2D behavior, but the AFM images reveal different surface morphologies. The steps in the AFM images correspond to 1 QL, which is about ~ 1 nm in height. From this, one can see that the ratio of the height over the width of the features is around 1%, and, therefore, despite the appearance, the surfaces are quite flat, which confirms the RHEED data. Figure 2-30a and b show Bi_2Se_3 grown on Si (111) and Al_2O_3 both of which have hexagonal terminated structure that match Bi_2Se_3, which allows Bi_2Se_3 to orientate itself crystallographically. This can be seen by the coherently oriented triangular features, which are the lowest energy surface structure. In Figure 2-30c and d, Bi_2Se_3 was grown on amorphous SiO_2, which is void of surface structure and $SrTiO_3$ (001), which has cubic surface symmetry, respectively. The Bi_2Se_3 in c-d is morphologically different from the Bi_2Se_3 grown on a hexagonal surface (a-b), because there is no preferred orientation, and at different nucleation points the Bi_2Se_3 orientates itself in random directions; despite this, the Bi_2Se_3 does correctly orientate along the *c*-axis, which, as we will see next, is quite distinct for Van der Waals materials.

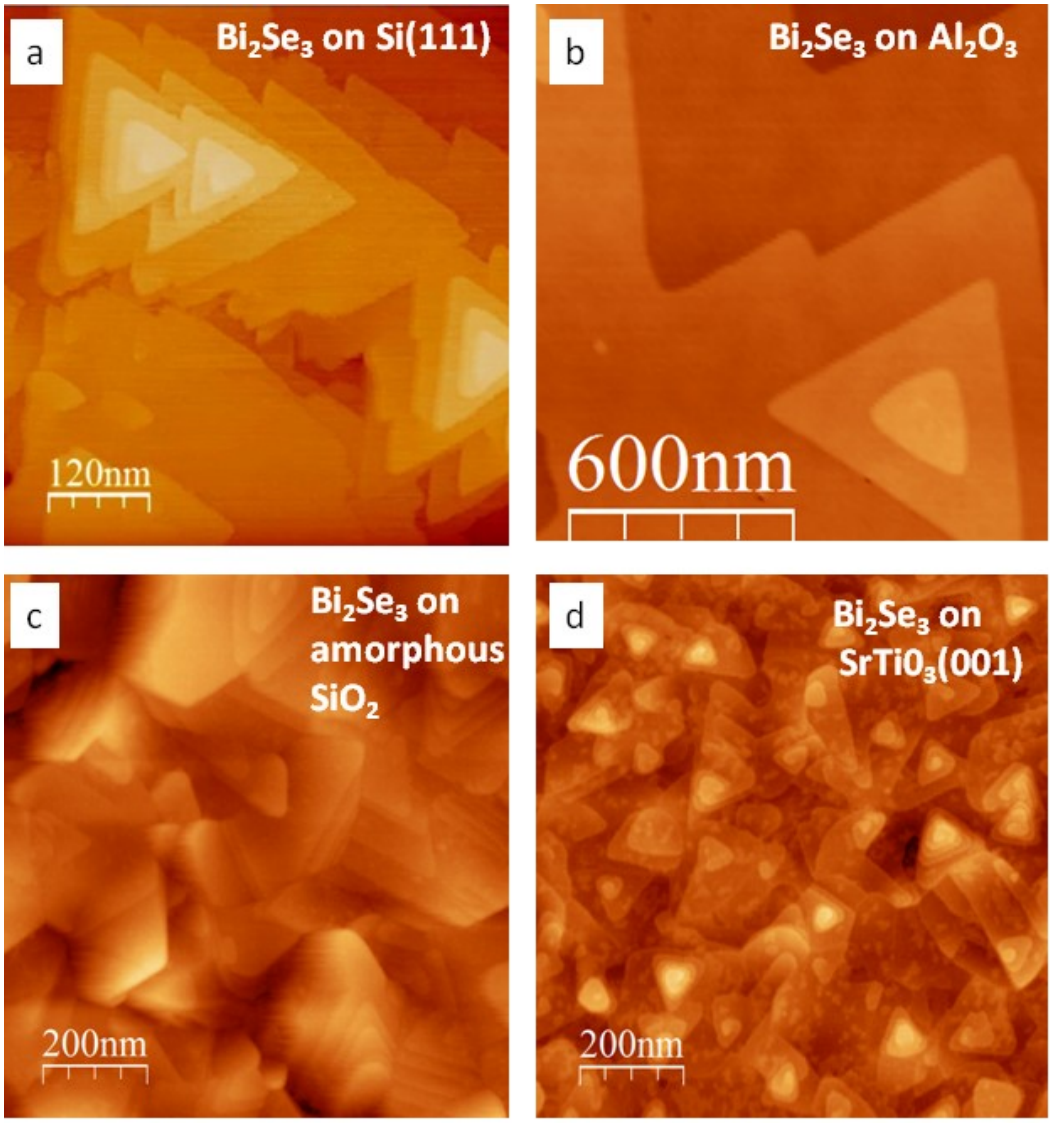


Figure 2-30 Several AFM images of Bi_2Se_3 grown on different substrates that show the different surface morphologies. a-b Bi_2Se_3 on Si(111) (~100 QL) and Al_2O_3 (0001) (~32 QL), both of which have hexagonal terminated surfaces that match Bi_2Se_3. **c-d** Bi_2Se_3 on amorphous SiO_2 (~32 QL) and cubic terminated $SrTiO_3$ (001) (~32 QL). In **a-d** all step-edge heights correspond to 1 QL ≈ 1 nm, showing that the surfaces are flat and 2D. Bi_2Se_3 on Si(111) ([16]Bansal *et al.*, 2011), and on Al_2O_3 (courtesy of Eliav Edrey).

2.8 Van der Waals epitaxy of topological materials: A_2Se_3 chalcogenide family

Van der Waals materials are present in our daily lives with the most recognizable material being graphite, which is most well known as pencil lead (which has nothing to do with the element lead). Graphite is composed of sheets that are held together by strong chemical bonds, and these

sheets are then held together by much weaker Van der Waals bonds. In fact, this much weaker Van der Waals bonding is what enables the sheets of graphite to slough off as the tip of a pencil is dragged against a piece of paper, thereby leaving semi-permanent marks. The strong in-plane bond and weaker out-of-plane bonding give these materials very different growth characteristics than normal chemically bonded materials (for example GaAs), which were realized as far back as the early 1990's in the growth of transition metal dichalcogenides such as $NbSe_2$, NbS_2, $MoSe_2$, and MoS_2 to name a few, and it was the growth of these materials that inspired the name *Van der Waals epitaxy* ([114]Koma, 1992).

In this section we will give an overview of the growth characteristics and procedures for growing thin films of these Van der Waals bonded topological materials. Although many different topological materials have been grown on various substrates, here we will take a chronological approach to how our growth recipe for the Bi_2Se_3 based TI materials developed to the point it is at today, and how the film properties became progressively better as our understanding of the growth improved.

2.8.1 Substrate consideration

As with the epitaxial growth of any material, the choice of substrate plays a critical role. With most chemically bonded materials, the choice of substrate basically centers on the surface structure and lattice mismatch. Chemical compatibility is only considered after structure and lattice mismatch are satisfied because without lattice matching the films cannot grow correctly. However, for these Van der Waals bonded topological materials, both structure and lattice mismatch play a less critical role and, therefore, chemical matching is promoted to a more significant position.

The substrates used in this work are all single crystal and available commercially. They were grown according to the method that produces the highest quality crystals of a sufficiently large size for each material; typically Czochralski or the Bridgman-Stockbarger processes are

used, and produce large, high-quality boules, which, for the case of single crystal silicon boules, can be as large as 45 cm in diameter and many meters long. To form substrates out of the a raw boule, the crystals were cleaved, or cut along a certain crystallographic direction in thin sheets called wafers; this cleaving can be done within an accuracy of around ~ 0.1°. Once the wafers are cut from the boule, they are then polished mechanically and electro-chemically, and finally the wafers are diced into the final size, which in this work is typically 1×1 cm^2. Altogether, this produces crystalline planes that are hundreds of nanometers in length for each unit cell step. All these steps are carried out in the factory, and then they are packed in ultra-clean bags for sale. This allows the processing steps to be minimal before using these substrates for epitaxial growth.

Prior to growth, all substrates are checked and cleaned by the following procedure. Before the substrates were unpacked, any tool that comes into contact with them was thoroughly cleaned with acetone then isopropyl alcohol, and then blown clean with dry N_2 gas to remove any dust particles; this includes the tweezers used for substrate manipulation, the scissors used to open the substrate bags, and the Teflon microscope slide. With these cleaned, the substrate bag was cut open, and the substrate was removed by grasping the center of the edge, by about < 1-2 mm. Then the substrate was blown off to remove any dust that has been attracted electro-statically to the surface. After this, the surfaces of the substrates were checked for cleanliness under an Olympus BH2 reflected light microscopy equipped with differential interference contrast. A small number of dust particles do not affect the growth because they are easily removed during the cleaning steps; however, any polishing fluid that is left over from the manufacturing process will spread over the surface during the cleaning, which causes the films to grow poorly; if any smears of polishing fluid were visible, then the substrate was discarded. After this the substrates were carefully mounted in the substrate holder (see Figure 2-13).

The cleaning was done in two steps. Once the substrates were found to be clean by optical microscope, then they were attached to the molybdenum plate as described in 2.4.2, and the surfaces were thoroughly blow clean with dry N_2 gas. The substrates were then cleaned ex

situ for 5 minutes in UV-generated ozone gas, which removes organic adsorbents. Once cleaned, the substrate plates were placed in the rings and then the substrate cassette, and placed in vacuum for loading into the growth chamber where in situ cleaning was performed as follows.

In the growth chamber the substrates were initially heated to a high temperature to further remove any surface contamination; this process varies for the different types of substrates. In the case of Al_2O_3 substrates, the temperature used for this outgassing step was 750°C, ramped to at 35°C/min; this is done in conjunction with filling the chamber with molecular oxygen to a pressure of 10^{-6} Torr, which acts to reduce the carbon in any organic contaminates to CO_2, which readily desorbs from the surface thereby cleaning the surface of the substrate. The Al_2O_3 substrates were left at 750°C for about 10 minutes then cooled to the growth temperature. For Si substrates with a thermally formed SiO_2 layer, the same outgassing step as Al_2O_3 was used, since the thermally formed SiO_2 is more stable than the native oxide layer on the Si(111) discussed next. For $SrTiO_3$, one must ensure that the substrates are not overheated, which creates oxygen vacancies and causes the substrate to become metallic. This may occur if the substrate is heated to 750°C, and, therefore, in this work to avoid this, the $SrTiO_3$ substrates were only heated to 600°C, and always in an oxygen environment.

For Si(111) substrates, the surface cannot be exposed to O_2 during the high temperature cleaning step, due to the formation of amorphous SiO_2 on the surface. The outgassing temperature used for Si(111) was around 830°C at 35°C/min; the actual temperature used varies from substrate to substrate, and was determined by the temperature where the native SiO_2 surface outgasses, leaving a clean Si(111) surface. The surface of Si(111) is very reactive, and, therefore, as soon as the cloudy amorphous SiO_2 RHEED pattern gives way to the crystalline Si(111) RHEED pattern, one must start cooling to lower the probability of surface contamination. In regard to the surface contamination, prior to growth on a Si(111) substrate, the growth stage must be outgassed to 900°C for ~30 minutes to desorb any Se or other contaminants preferentially to avoid adsorption and subsequent reaction on the clean substrates. For Si(111) the transition with

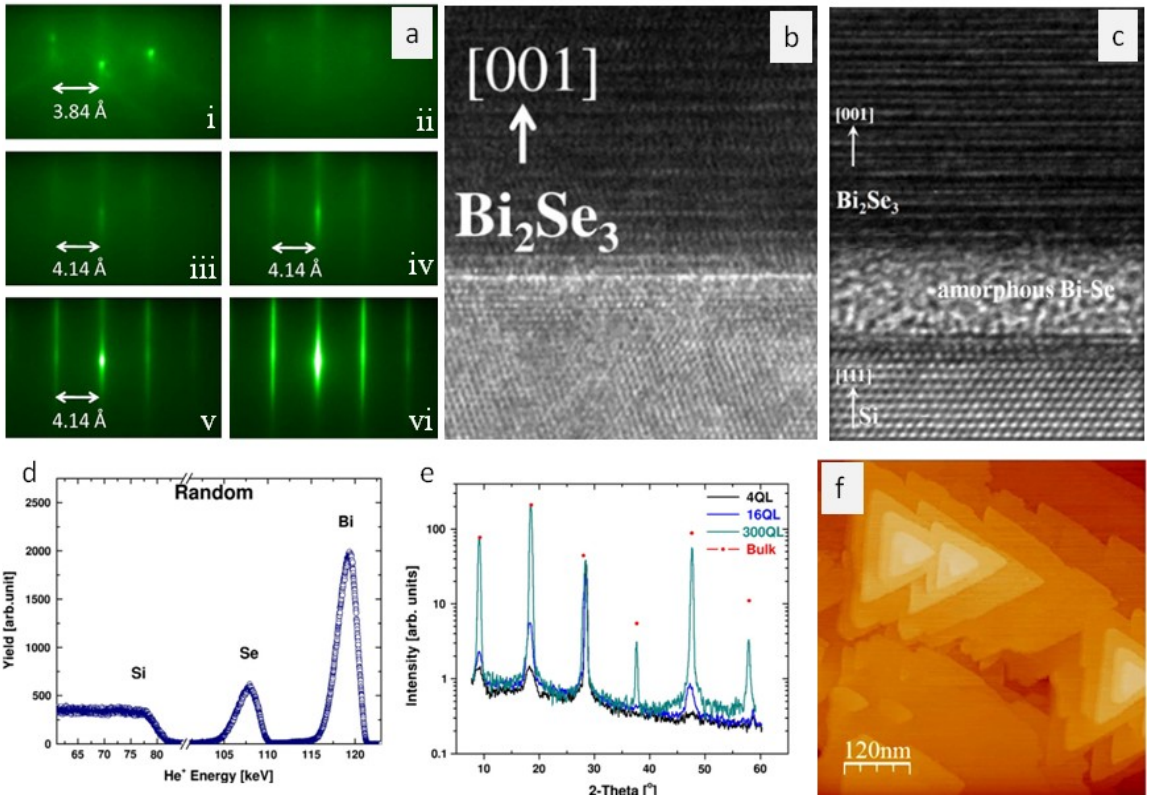


Figure 2-31 Various structural measurements of Bi_2Se_3 grown on Si (111) by the two step growth method. a RHEED images during the deposition. **i** Si(111) with the passivation layer of Se on top, which shows the lattice spacing of Si. **ii-iv** (0.5, 1 and 2 QL) the peak spacing shows that the lattice constant relaxes to that of Bi_2Se_3 within 1 QL. **v-vi** 2 QL Bi_2Se_3 annealed to 220°C and a final thickness of 16 QL. **b-c** TEM images of Bi_2Se_3 grown on Si by Se passivation (**b**) and by Bi passivation (c, [123]Li *et al.*, 2010), which shows a clean interface and an amorphous interface. **d** Ion scattering shows that the Bi:Se ratio is 2:3. **e** XRD shows that the film grows in the correct 1 nm QL which can be seen in **b**. **f** The surface is flat with the lowest energy surface being triangular which reflects the underlying crystal structure. (Adapted from [16]Bansal *et al.*, 2011)

decreasing temperature to the 7 × 7 surface reconstruction (see Figure 2-26b) was used as an indicator of the cleanliness of the surface. If even a monolayer of contamination is on the Si, then the surface will fail to reconstruct, and the substrate must be discarded.

2.8.2 Bi_2Se_3 on Si(111): the basic Bi_2Se_3 recipe

Bi_2Se_3 was first grown on Si(111) because there is a relatively good lattice matching (~7%), the proper terminating crystal structure, and known chemical compatibility; the latter has been known for over 20 years, when a single monolayer of Bi was shown to grow well on the Si (111) surface ([193]Wan *et al.*, 1991), which yields either 1/3 (α-phase) or a full monolayer (β-phase) coverage

with $\sqrt{3} \times \sqrt{3}\, R30°$ surface structure (relative to Si(111) in standard Wood's notation, see [214]Zangwill, 1988). This inspired the initial work on growing Bi_2Se_3 on Si(111), which used the β-phase Bi monolayer as a buffer layer on the Si(111) surface to terminate the dangling Si bonds. This step was often performed at, or below, room temperature, which results in an amorphous interfacial layer (see Figure 2-31c, [215]Zhang *et al.*, 2009;[123]Li *et al.*, 2010).

The more simple approach used in this work is to use Se to terminate the Si dangling bonds. The temperature of the Se termination is critical because if the temperature is too low, then the Se will collect and form an amorphous structure, and if the temperature is too high, then the Se reacts and forms 3D $SiSe_2$ clusters. At the optimum temperature of ~ 110°C, Se will form one complete monolayer by bonding to the Si (as seen by the transition from 7×7 to 1×1), and any subsequent Se will not bond and desorbs off the surface. As shown in Figure 2-32, at this temperature crystalline Bi_2Se_3 can be immediately deposited, and not react with the underlying Si due to the strong passivating bonds between the Se and the Si atoms on the surface. However, this is lower than the optimum growth temperature for Bi_2Se_3, and, therefore, only 3 QL Bi_2Se_3 is deposited at 110°C. Following this, the film is annealed to a higher temperature and the film quality of this initial 3 QL seed-layer improves, and at this elevated temperature all the subsequent Bi_2Se_3 layers were deposited to attain the desired thickness.

During the initial low temperature growth step the RHEED pattern progresses as follows: as shown in Figure 2-31a, the Si(111) 7 × 7 reconstruction pattern gives way to a 1 × 1 pattern during the initial Se deposition, where the lattice constant remains that of Si ~3.84 Å; this typically takes around ~3-4 seconds with the Se flux set to be around ~25× 10^{13} /cm^{2}s (this is roughly 10 times larger than the Bi flux). Following this, the Bi shutter is open, and the RHEED pattern progressive changes to that of Bi_2Se_3, with the lattice constant relaxing to ~ 4.138 Å within the first QL. The RHEED pattern continually improves until about 3 QL, and if the film is allowed to continue growing beyond 3 QL, the pattern eventually gets worse and more 3D

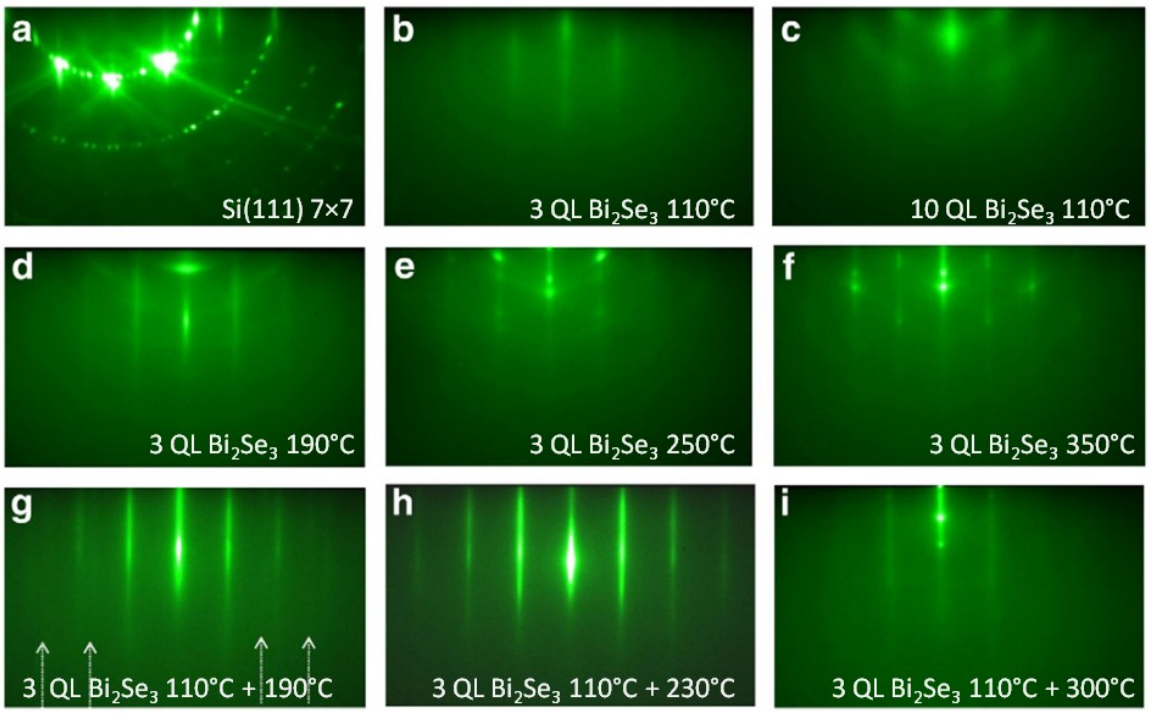


Figure 2-32 Various RHEED images which were used to determine the optimal growth conditions. a Bare Si(111) with no Bi_2Se_3. **b-c** 3 and 10 QL Bi_2Se_3 deposited at 110°C. **d-e** 3 QL grown at 190, 250, and 350°C. **g-h** 3 QL grown at 110°C followed by further deposition at 190, 230, and 300°C (the arrow in **g** are point to 30° rotated twin structure). From these images it can be seen that panel **b** followed by **h** yield the best results. (Adapted from [16]Bansal *et al.*, 2011)

(Figure 2-31a); this is due to the insufficient surface mobility of the adsorbed atoms, which is due to the low temperature, and, therefore, a higher temperature is required.

Beyond this low temperature passivation/seed-layer, RHEED indicated the optimum growth temperature to be ~220°C. This is seen by the appearance of RHEED oscillations at around ~170°C (Figure 2-27), which indicates a layer-by-layer growth mode. Beyond this, the RHEED oscillations diminished, and gave way to step-flow growth due to the larger surface mobility, which indicates the optimum growth temperature is beyond 170 °C. On the higher side, growing subsequent Bi_2Se_3 layers at ~250°C yields second phase features in the RHEED pattern, which points out that this higher temperature possibly induces some reaction with the Si substrate (see Figure 2-32).

Using the 110 °C and 220 °C growth procedure yields high quality Bi_2Se_3 films, which can be seen by the combination of structural measurements shown in Figure 2-31. Medium energy ion scattering (a process similar to RBS), shows that the film composition is stoichiometric Bi_2Se_3; additionally, this also shows the interface is sharp, and there is minimal atomic diffusion at the interface. This interfacial sharpness can also be seen in the cross-section transmission electron microscopy (TEM) image in Figure 2-31b; it can be seen that the transition from Si to Bi_2Se_3 is within about 1 nm, with negligible interfacial mixing and, in this image, the Van der Waals layered QL structure is clearly visible. The QL structure is also consistent with the XRD scans in Figure 2-31e, which shows that the *c*-axis is periodic with ~1 nm layers, and with no second-phase formed. Lastly, AFM measurements of the surface morphology confirm that the surface is indeed flat. Moreover, the AFM images in Figure 2-32f show that the lowest energy surface structure are triangular features; this shape is the one with the lowest surface free energy, and it can be shown that this shape always reflects the underlying 6-fold symmetry of the lattice (e.g. hexagonal or triangular). Applying Wulff's theorem (see [214]Zangwill, 1988) to this surface shows that the surface energy per unit area must differ by a factor of two along the hexagonal (100) compared to that of the (010) for the lowest energy surface to be triangular rather than hexagonal.

2.8.3 Bi_2Se_3 on Al_2O_3

The importance of chemical compatibility and lattice matching can be seen by the differences in the Bi_2Se_3 grown on Si(111) versus growth on Al_2O_3 (sapphire). Al_2O_3 has a much larger lattice constant than Si(111) at ~4.785Å, and has a very large lattice mismatch of ~ -14% with Bi_2Se_3. Despite this, Al_2O_3 is far more chemically inert than Si(111), which is due to the strong bonding between Al and O. In regard to the growth procedure, growing Bi_2Se_3 on Al_2O_3 uses nominally the same steps as growing Bi_2Se_3 on Si(111), but the rationale for such a method is completely different. As discussed previously, the two step growth procedure was used on Si(111) to avoid

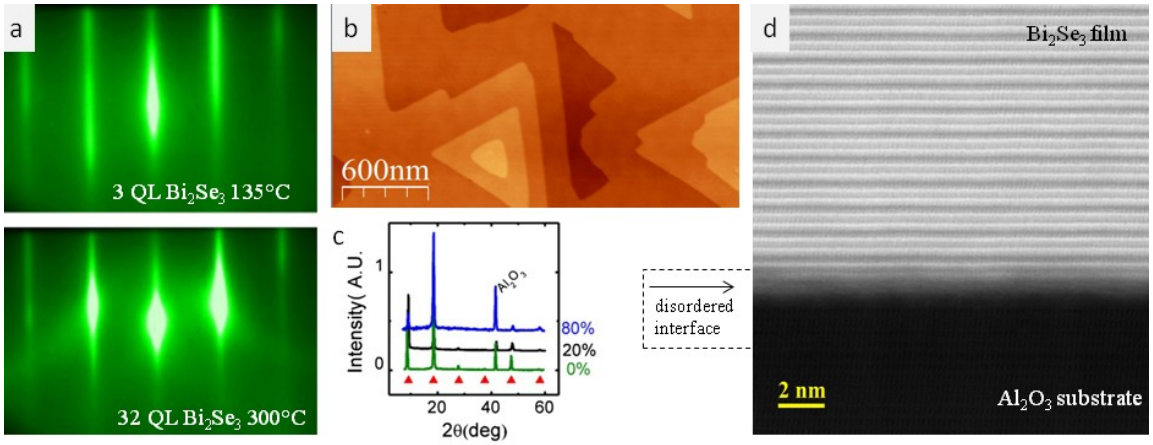


Figure 2-33 Structural data for Bi_2Se_3 grown on Al_2O_3 substrates. a RHEED images of the Bi_2Se_3 during the first 3 QL at 135°C, and the subsequent 32 QL at 300°C (adapted from [15]Bansal *et al.*, 2012). **b** AFM image of a 32 QL Bi_2Se_3 film on Al_2O_3, which show the same low energy surface structure as Bi_2Se_3 on Si(111), but with a larger feature size (courtesy of Eliav Edrey). **c** XRD measurement for $(Bi_{1-x}In_x)_2Se_3$ films grown on Al_2O_3 substrates for $x \approx 0$, 20, and 80%. The red triangles mark the *c*-axis peaks for bulk crystals. (Data taken from [28]Brahlek *et al.*, 2012.) **d** Transmission electron microscopy image of Bi_2Se_3 grown on Al_2O_3, where the interface is disordered (Courtesy of M.-G. Han and Y. Zhu, [113]Koirala *et al.*, 2014).

surface reaction and subsequent second-phase formation; for Bi_2Se_3 grown on Al_2O_3 there is no issue with surface reaction and second phase formation due to the chemical inertness of Al_2O_3, and, therefore, in theory one might be able to use a single step growth process and eliminate the low temperature growth step. However, achieving the initial heteroepitaxial layer on Al_2O_3 is difficult because the sticking coefficient for Bi alone, and Se alone, are nearly zero above ~190°C; it is only when both Se and Bi are present together that triggers chemisorption. Due to this effect, if one attempts to co-deposit Bi and Se to form Bi_2Se_3 on Al_2O_3 at, say, ~220°C, the coverage will be spotty and the thickness will not easily be controlled during the initial layers (i.e. the growth mode is closer to the Stranski-Krastanov type, see 2.2.3). Since it is easier to achieve heteroepitaxy at a lower temperature, one can use a similar two temperature grow scheme to achieve good coverage at low temperature, and annealing to the higher growth temperature to improve crystal quality as the subsequent layers are deposited. This works because once the initial seed layer is uniformly deposited, the effective sticking coefficients for Bi and Se become larger. The main advantage with growing Bi_2Se_3 on Al_2O_3 is that Al_2O_3 chemical inertness allows

the two step growth process to be done at the higher temperatures of ~135°C and ~300°C without causing a chemical reaction, which yields higher quality films.

Figure 2-33 shows structural data for Bi_2Se_3 grown on Al_2O_3. From the structural data it can be seen that Bi_2Se_3 is slightly different than that grown on Si(111), with the largest difference coming from transport data which will be presented later. RHEED shows that, like on Si(111), Bi_2Se_3 on Al_2O_3 is 2D and high quality, which is seen by the sharp diffraction peaks and the presence of Kikuchi lines; similarly, XRD scans confirm the presence of highly ordered QL along the *c*-axis. AFM images show one difference between Bi_2Se_3 grown on Al_2O_3 and Si(111). On Si(111) and Al_2O_3 the lowest energy surface are the same triangular shapes; however, on Al_2O_3, the feature size is much larger. This is due to the lower reactivity of the Al_2O_3 substrates. On Si(111), the surface is more reactive, and, therefore, the Bi_2Se_3 has more places to nucleate during the growth of the initial QL, which results in a rougher surface. On Al_2O_3, the only place for the Bi_2Se_3 to nucleate is the step edges, and since the growth extends from there, the surface feature size ultimately reflects the feature size of the step edges.

The cross-sectional transmission electron microscopy image shown in Figure 2-33d, shows that the interface between the Bi_2Se_3 and Al_2O_3 is disordered. Although, from this image, it may appear that the interfacial layer may be amorphous, RHEED shows that the first QL is coherently crystalline. The likely origin of this is disorder is the grain boundaries within the first QL. One can imagine that during the first QL, the film nucleates simultaneously from many locations on the surface. As the film grows away from these nucleation points, they will eventually meet. Since the substrate and film's lattice constants are 14% different, the grain boundary meets with spacing that is not close to a multiple of the lattice constant of Bi_2Se_3. Therefore, as the grain boundaries merge, there will be a high probably to introduce defects in the form of Se vacancies. The interesting thing about Bi_2Se_3 is beyond this first disordered QL, the film will grow without continuing these defects, which can be attributed to the layered Van der Waals structure. Although, at this moment this is unconfirmed, it is likely the origin for the

"thickness-independent" carrier density observed in Bi_2Se_3 grown directly on Al_2O_3 ([15]Bansal *et al.*, 2012), and is consistent with measurements of Cu-doped Bi_2Se_3 grown directly on Al_2O_3 (see Chapter 6, and [32]Brahlek *et al.*, 2014), and Bi_2Se_3 grown on an In_2Se_3 buffer layer (see 2.8.7, and [113]Koirala *et al.*, 2014).

2.8.4 Bi_2Se_3 on amorphous and non-symmetry matched substrates

Bi_2Se_3 can be grown on virtually any flat substrate that is chemically compatible with Bi and Se. This is possible due to the energetics of the Bi_2Se_3, which is virtually the only energetically stable phase along with the layered Van der Waals bonded nature. Following the same two step growth procedure, Bi_2Se_3 was grown on amorphous SiO_2 and cubic $SrTiO_3$ (001). With both of these substrates, RHEED showed that the films grew in a 2D layered mode, with random rotations in-plane. This is due to the lack of the proper crystalline substrate to coherently orient the film, and, therefore, during the deposition of the initial layers, each nucleation point will seed a Bi_2Se_3 crystallite with random in-plane orientation, but the energetics always favor the *c*-axis of the film to be aligned to the surface normal. In RHEED, this appears as long streaks that do not change when the film is rotated; to rationalize this pattern, one can imagine super-imposing RHEED images of a 2D single crystal at all angles simultaneously. Therefore, as shown in Figure 2-25, the diffraction peaks for the hexagonal (120) and the (010) rotations would appear simultaneously. This is confirmed in both AFM and XRD; in the AFM images shown in Figure 2-30, the lack of the coherently orientated triangular surface features seen in Bi_2Se_3 grown on Si(111) and Al_2O_3, indicate that the underlying crystal-symmetry is at least random in-plane, and possibly along the c-axis. Transmission electron microscopy images then confirm that the random nature is only in-plane and indicate that along the *c*-axis, the films are single crystal with the proper QL structure. However, TEM images show that these films have significant planar defects in the form of stacking faults and polycrystalline grains.

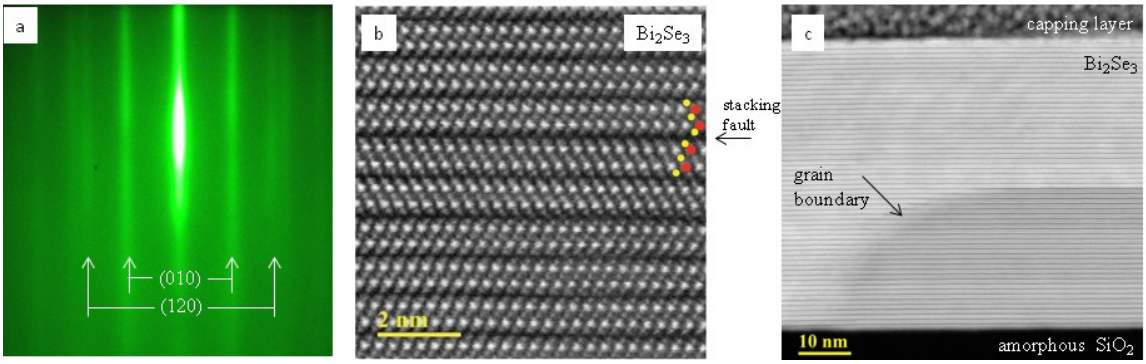


Figure 2-34 Structural measurements of Bi_2Se_3 grown on amorphous SiO_2. a RHEED shows no change with rotation, and the simultaneous appearance of the (010) and the (120) peaks indicate the film is poly-crystalline in-plane. **c-d** TEM confirm that the film is poly-crystalline in-plane but orientate along the *c*-axis. (TEM images courtesy of M.-G. Han and Y. Zhu)

2.8.5 In_2Se_3 on Al_2O_3

Unlike Bi_2Se_3, In_2Se_3 has as many as five different crystal structures, with only the α-phase being the layered Van der Waals bonded structure ([203]Wyckoff, 1964;[199]Watanabe *et al.*, 1989). Therefore growing α-In_2Se_3 is not as simple as Bi_2Se_3 because the energetics can easily favor one of the other phases. This can been seen if one tries to directly grow In_2Se_3 on Al_2O_3, which results in a 3D film structure with possibly many different phases. This is due to the large lattice mismatch of ~ -16%, which is too energetically costly to overcome. In the case of Bi_2Se_3, the lattice mismatch was accommodated by creating defects in the first QL, but in the case of In_2Se_3, it is energetically more favorable to grow in a possible combination of other phases, and 3D structures. As we will see next, there are two possible routes to obtain the targeted In_2Se_3 α-phase. In the case of $(Bi_{1-x}In_x)_2Se_3$, for $x < 1$, where the small amount of Bi stabilizes the α-phase, and the second route is growing In_2Se_3 directly on a Bi_2Se_3 buffer layer, in which, the α-phase is also favored.

2.8.6 $(Bi_{1-x}In_x)_2Se_3$: film growth and flux control

For the $(Bi_{1-x}In_x)_2Se_3$ films used in this work the growth recipe followed that of Bi_2Se_3 grown on Al_2O_3. The growth mode is the same for low x; however, as x was increased the film quality

decreased; this is mainly due to the large lattice mismatch, along with In_2Se_3 being more reluctant to relax than Bi_2Se_3. For the initial 3 QL growth, with increasing x the film became closer to amorphous, and completely amorphous beyond ~ 20-30%. Upon annealing to the final growth temperature of 300°C, the films crystallized to the Van der Waal α-phase for all $x < 1$, and further deposition followed this phase; this yielded films of reasonably good quality for all x.

The key to studying the intrinsic electronic properties of $(Bi_{1-x}In_x)_2Se_3$ is being able to reliably control x within ±1% of the targeted value. There are two routes to obtain $(Bi_{1-x}In_x)_2Se_3$ with the correct stoichiometry. The technique that was primarily used in this work was to adjust the individual fluxes (by adjusting the cell temperatures) for In and Bi such that their ratio gives the prescribed x. The second route was to use fixed In and Bi fluxes (fixed cell temperatures), and adjust how long each element's shutter was open within each QL, which is referred to as the shuttering technique; this technique relies on interdiffusion of In and Bi in the $(Bi_{1-x}In_x)_2Se_3$ structure to create a homogeneous mixture.

In practice, one needs to determine the cell temperature to achieve a desired value of x. To find the required cell temperatures, one must consider how the lattice constant changes with x, and how the individual fluxes change with temperature. The knowledge of how the lattice constant changes with composition is critical to the thickness control because this tells how many atoms are required to complete a QL unit. Vegard's law predicts that the lattice constant of an alloy should linearly interpolate between the end members ([59]Denton and Ashcroft, 1991), and RHEED measurements on $(Bi_{1-x}In_x)_2Se_3$ films show that Vegard's law holds (see Figure 2-35, [28]Brahlek *et al.*, 2012). This allows the lattice constant as a function of x to be approximated simply as $a_{BIS} = a_{BS}(x\text{-}1) + a_{IS}(x)$ where a_{BS} = 4.138 Å, and a_{IS} = 4.00Å ([203]Wyckoff, 1964;[28]Brahlek *et al.*, 2012). With the lattice parameter calculated, together with the growth rate (typically somewhere around ~1 min/QL), gives the total required flux (given by atomic density divided by the desired growth rate), which is composed of 1-x parts Bi and x parts In. This gives

the required flux for both Bi and In, which is obtained by adjusting the cell temperatures as follows.

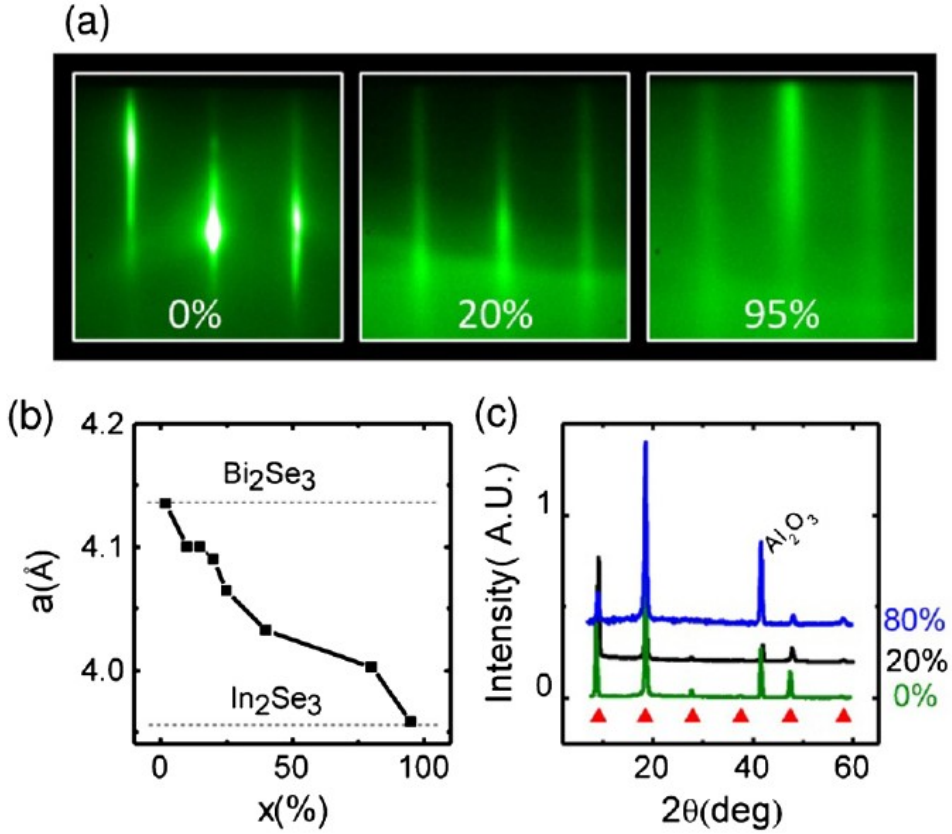


Figure 2-35 Structural data for $(Bi_{1-x}In_x)_2Se_3$ films. a Typical RHEED images for x=0%, 20%, and 95%. The 2nd-phase free vertical streaks indicate atomically flat films for the full range of x. **b** Lattice constant as a function of x obtained from the horizontal separation of the maxima in **a**. **c** X-ray diffraction data for x=0%, 10%, and 80%. The (003n) peaks, marked by triangles for n=1–6, provide a quintuple layer thickness of ~9.5 $\AA$, almost independent of x. (Adapted from [28]Brahlek *et al.*, 2012)

The most practical way to determine the temperature required to give a specified flux is to precisely measure the flux at a single temperature then make use of a semiempirical function to extrapolate the flux at other temperatures. For a specific element, the temperature dependence of the flux is given by

$$F(T) = F_0 \frac{P(T)}{P_0} \sqrt{\frac{T_0}{T}}$$

where F_0, and P_0 are flux and the vapor pressure of the given element at a temperature T_0. $P(T)$ is the temperature dependence of the vapor pressure, which, to first order, is exponential in the temperature, but in practice the most accurate way to determine the vapor pressure is to make use of empirical values that are available as charts ([131]Margrave, 1967). To do this, we simply fit a parabolic function to $\log(P(T))$ vs T between vapor pressures of $\sim 10^{-7} - 10^{-3}$, which provides an analytic expression $P(T)$. Then, by simply measuring F_0 at a single temperature T_0, and $P_0 = P(T_0)$, yields an analytic approximation to how the flux changes with temperature. In practice, this function was programmed into an Excel file with the growth rate and desired composition x as the inputs, with the output being the cell temperatures. When checked against RBS this method has been found to be accurate to less than ~1%.

2.8.7 Bi_2Se_3 on In_2Se_3 buffer layers

As discussed in section 2.8.3, Bi_2Se_3 grown on Al_2O_3 is the highest quality from a structural point of view; this is barring the high concentration of defects confined within the first QL, as shown in Figure 2-36. In transport measurements, these interfacial defects are seen as a thickness-independent carrier density ([15]Bansal *et al.*, 2012, [87]Hellerstedt *et al.*, 2014). As shown in Figure 2-36a, by extrapolating the carrier density down to zero thickness it can be seen that this interfacial defect layer is given to be $n_{Interface} \approx 3 \times 10^{13}$ /cm^2. For a surface defect layer plus bulk defects, the areal carrier density should increase with thickness like $n_{2D} = n_{Interface} + N_{Bulk} \times t$, with t being the thickness of the film. In the limit of $N_{Bulk} \times t \ll n_{Interface}$, the expression for the areal carrier density reduces to $n_{2D} \approx n_{Interface}$, which, in this limit, is independent of thickness. As shown in Figure 2-36a, the carrier density versus thickness for Bi_2Se_3 on Al_2O_3 shows this behavior. It should be noted that this is not necessarily evidence for transport through the TSS because a material without TSS would exhibit the same behavior. From this data, one can estimate an upper bound for the bulk carrier density N_{Bulk}, by analyzing the variation in the n_{2D} vs thickness curve; from this, the upper bound on bulk defect density is $\sim 5 \times 10^{17}$ /cm^3, which would appear as an

areal carrier density of ~ (5 × 10^{10} /cm^2) × *t*(nm), where *t* is the thickness in nanometers. This shows that the bulk state would only appear when the thickness is in the range of several hundred

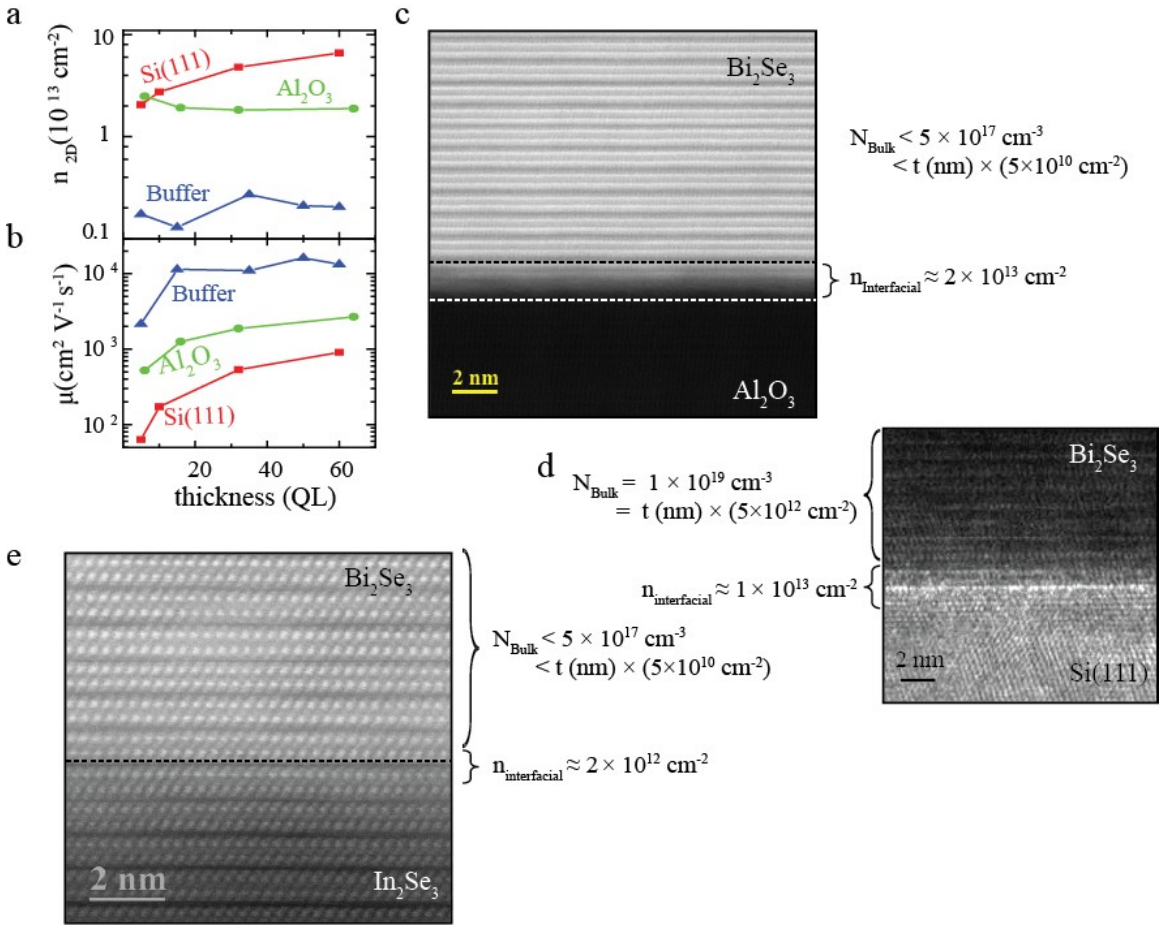


Figure 2-36 Transport properties versus thickness that originates from defects in the materials. a-b The carrier density (**a**) and mobility (**b**) vs thickness for Bi_2Se_3 grown on Si(111), Al_2O_3, and an In_2Se_3 buffer layer. The origin of the thickness dependence can be traced back to structural properties as shown in the transmission electron microscopy images in **c-e** (Bi_2Se_3 on Al_2O_3, Si (111), and In_2Se_3 buffer layer, respectively). (Adapted from [113]Koirala *et al.*, 2014)

nanometers. As can be seen in Figure 2-36b, this defect layer pins the mobility to values less than ~2000 cm^2/Vs, which is due to the large level of disorder associated with these defects.

In contrast to Bi_2Se_3 grown on Al_2O_3, Bi_2Se_3 grown on Si(111) has a lower interface defect level ~1 × 10^{13} /cm^2, which is likely due to the better lattice matching, and less defects associated the relaxation in the first QL. The carrier density vs thickness for Bi_2Se_3 grown on Si(111) does not show the surface defect dominance as with Bi_2Se_3 grown on Al_2O_3; this is

because there is a significantly higher bulk defect density at ~1×10^{19} /cm^3, which is likely due to the more reactive Si(111) substrate that induces more bulk defects.

The observation of Bi_2Se_3 grown on Si(111) having fewer interfacial defects than Bi_2Se_3 grown on Al_2O_3, which has a significant lower bulk defect density, motivated us to find a substrate that was chemically compatible with Bi_2Se_3, like Al_2O_3, but had a closer lattice match, like Si(111). It was quickly realized that In_2Se_3 was an optimal candidate, but In_2Se_3 substrates are not commercially available, so for practical purposes In_2Se_3 must be grown on a substrate as a buffer layer prior to the deposition of Bi_2Se_3. Due to its chemical inertness Al_2O_3 would be the optimum substrate to grow the In_2Se_3 buffer layer on, but as discussed in 2.8.5, high quality epitaxial In_2Se_3 cannot be grown directly on Al_2O_3. In spite of this, the following scheme has successfully been developed.

It has been found that high quality In_2Se_3 can be grown on Bi_2Se_3. So, as summarized in Figure 2-37, the following scheme is the basis for the In_2Se_3 buffer layer: On an Al_2O_3 substrate grow an initial layer of Bi_2Se_3, then a layer of In_2Se_3, and finally a second layer of Bi_2Se_3, which will not contain the defect laden interfacial layer as with the first layer. The problem with this scheme is that in a transport experiment, electric current will flow through both the upper and lower layers of Bi_2Se_3, thereby confounding the results. With the knowledge that $(Bi_{1-x}In_x)_2Se_3$ becomes insulating around $x \gtrsim 25\%$ ([28]Brahlek *et al.*, 2012), a thin layer of Bi_2Se_3, ~3 QL, can be deposited on Al_2O_3 at 135 °C, and annealed to 300 °C, where 20 QL of In_2Se_3 is deposited. This structure has a resistance in the range of ~kΩ, but upon annealing to 600 °C, the Bi and In intermix across the interface, making a non-homogeneous mixture of $(Bi_{1-x}In_x)_2Se_3$; after this annealing step, the resistance of the film becomes immeasurable large, which indicates $x > 25\%$ throughout the film. Therefore, the defects that are created at the Bi_2Se_3-Al_2O_3 interface are rendered inert by burying them within this now insulating layer. With this achieved, the film temperature can be lowered to the optimum temperature of 275 °C and the final Bi_2Se_3 can be grown. An unexpected problem arose with the previously mentioned scheme, in that, In was

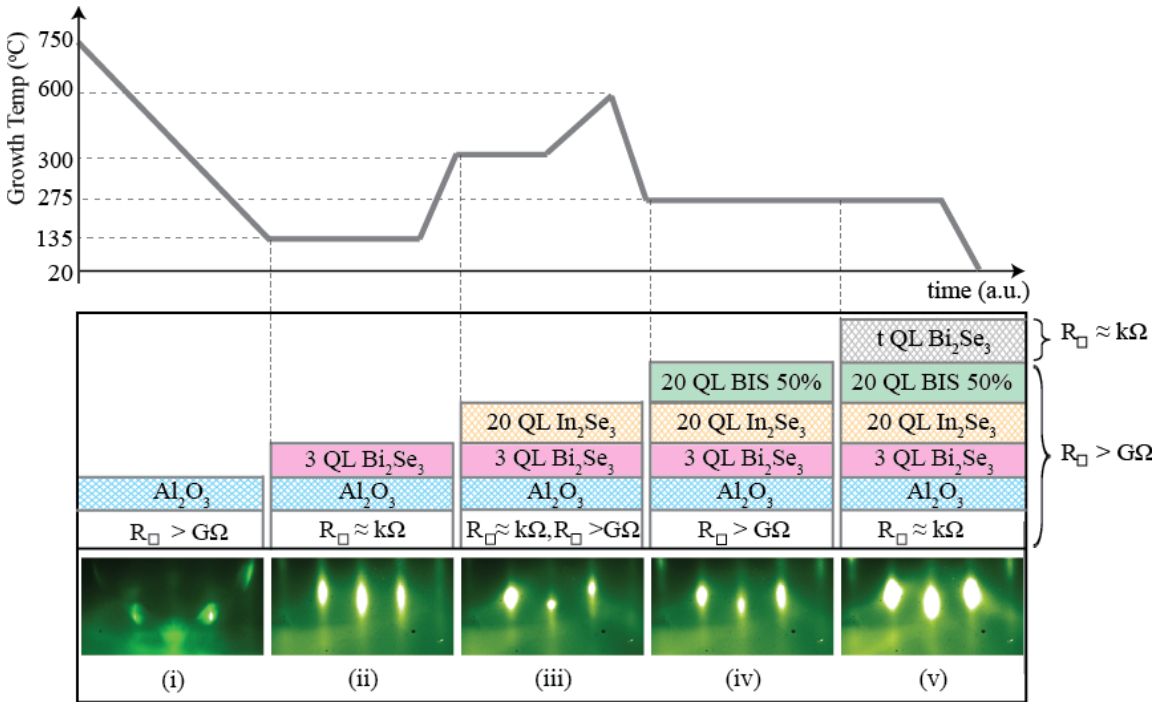


Figure 2-37 A graphical depiction of the growth recipe for Bi_2Se_3 on the In_2Se_3 buffer layer. After the Al_2O_3 substrate is outgassed to 750 °C (**i**), it was cooled to 135 °C where 3 QL of Bi_2Se_3 is deposited (**ii**). Following this, the film is annealed to 300 °C where 20 QL In_2Se_3 is deposited, and subsequently annealed to 600 °C to suppress the conduction in the initial 3 QL of Bi_2Se_3 (**iii**). After cooling to 275 °C, 20 QL of $(Bi_{0.5}In_{0.5})_2Se_3$ (labeled BIS 50%) is deposited, which remains immeasurably insulating (**iv**). This is followed by the final deposition of Bi_2Se_3 to achieve the desired thickness. (Adapted from [113]Koirala *et al.*, 2014)

found to diffuse up from the In_2Se_3 layer when Bi_2Se_3 was grown directly on top of In_2Se_3. To remedy this, a $(Bi_{0.5}In_{0.5})_2Se_3$ spacer layer was grown between the In_2Se_3 layer and final Bi_2Se_3. With this additional layer, the resistance of the buffer layer remained immeasurably insulating, and suppressed In diffusion to below an acceptable level of < 1% on the surface as measured by scanning tunneling microscopy.

When grown on this buffer layer scheme, Bi_2Se_3 showed remarkably improved transport properties. As shown in Figure 2-36a-b, total carrier density decreases by over an order of magnitude to a minimum of around ~2×10^{12} /cm^2, while the mobility increased to a maximum value of around ~15,000 cm^2/Vs, which is an improvement by nearly a factor of 10 compared to Bi_2Se_3 grown on Al_2O_3, and is highest reported mobility of thin film Bi_2Se_3. The thickness dependence shows that interfacial defects still dominate in the thin regime with a density of

around ~2 × 10^{12} /cm^2. However, with this lower interface defect density, the bulk state is more evident than Bi_2Se_3 grown on Al_2O_3, with a similar upper bound around ~5 × 10^{17} /cm^2. In closing, it should be noted that the improvements at the material level were, as with GaAs, motivated by transport, which is the most sensitive technique to the presence of defects. Each step of the growth scheme was optimized by transport measurements because the defects are so dilute that they are beyond the sensitivity of techniques like RHEED.

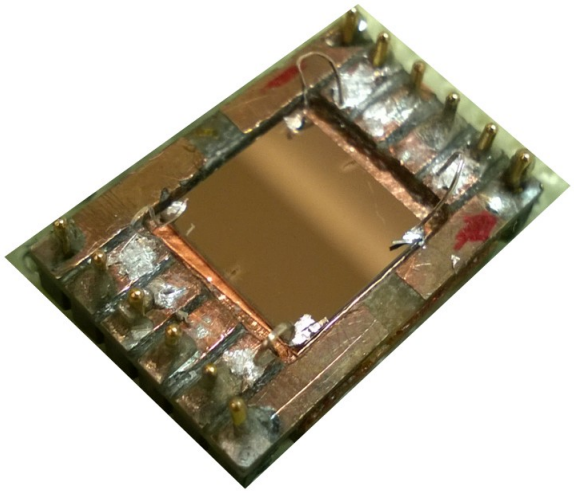
2.9 Conclusion

In closing, we have thoroughly discussed the many aspects required to grow thin film crystals. Navigating the large parameter space to determining the optimum growth parameters can seem to be a daunting task. However, with a well honed understanding of the fundamental chemical and physical processes, the optimum conditions can be found. As we have shown, this takes a combination of many probes and techniques, all of which give an individual perspective of the qualities of thin films.

Going forward in the growth of topological materials, MBE has proven to be a vital tool to discover new materials as well as improve the current set of materials. One of the major problems with TIs is the metallic bulk channel that results from defects. This is a problem at the materials processing level, and as we will show in Chapter 6, this can be solved through improvements in the growth. Beyond this there are still many points where the material properties can be further improved. A particular tough problem is the integration of Bi_2Se_3 with other materials. Currently, there are very few materials that can be grown underneath, or on top of Bi_2Se_3 without chemically reacting. The ability to integrate topological materials with dielectrics, ferromagnetic, and superconductors will allow the novel physics of topological surface states to be illuminated. In the coming years, as the growth parameters are further refined and new materials are found, many novel states will emerge out of TI systems that are grown by molecular beam epitaxy.

Chapter 3.

Electrical transport: Theory and practice[3]



In the previous chapter we went into great detail describing the processes used to synthesis high quality topological insulator thin films by molecular beam epitaxy. However, the ultimate goal of this work is to understand the novel physical states that reside on the surface of topological materials – a task that inevitably requires the collection of experiment data. Thus, in this chapter, we will introduce transport theory through a simple and pedagogical route where the basic physics can be understood without broaching the deeper complexities of transport theory, which can be very difficult and represents an entire field in its own right. This simple knowledge will allow us to understand the basic physics that emerges from transport measurements on topological insulator films in the later chapters. We will follow this with the details regarding the experimental processes and procedures used in this work to obtain the transport data.

Transport measurements belong to a broad class of probes that study the response of a material to an applied gradient. These phenomena range across all of science: from chemistry and biology, every engineering discipline, to materials science and physics. One can see this by imagining the multitude of examples that fall under the following categories: fluid flowing in response to a difference in pressure (momentum transport), diffusion of a solution across a

[3] A topological insulator thin film, wired and ready for electrical transport measurements

membrane (mass transport), heat flowing from a hot source towards a cooler sink (energy transport), and the main focus of this dissertation is how electrons flow through a material in response to an applied electric field (electronic transport, see [26]Bird *et al.*, 2002 for more details on general transport phenomena). In general, the functionality of all practical electronic devices ultimately rely on how electrons flow through a material. For example, the fundamental operational principle for transistors, and, therefore, the functionality of all computers, is the ability of an applied voltage to change how electricity flows through a material: the transistor is in an *on* state (binary 0) if electricity flows unimpeded, and an *off* state (binary 1) if electricity does not flow at all. In this way, to realize some of the practical properties of the novel surface states in topological insulators, one must understand their transport properties.

3.1 Semiclassical transport theory

The analysis of electronic transport pre-dates the quantum mechanical model of atomic physics; despite this, the basic description and understanding of the simplest transport model has remained relatively unchanged and relevant to understand the fundamental processes involved. We will start with an introduction to transport with the simple Drude model, which provides a qualitative understanding of, and the most basic quantitative framework for, analyzing and understanding transport processes (the following loosely follows [104]Kasap, 2002). More rigorous theories arrive at the same basic results, which would take us too far afield, and therefore we will simply provide references to more advanced texts (see [10]Ashcroft and Mermin, 1976 and [81]Grosso and Pastori Parravicini, 2000).

In the early 1900s, Paul Drude sought to understand the behavior of the electrons in metals; motivated by the classical theory of gases, and pre-dating the theory of quantum mechanics, he applied classical mechanics to the valence electrons in a metal by treating them as if they behaved like molecules in the gas phase ([61]Drude, 1900;[60]Drude, 1900). Despite the simple

approximations, Drude's model, with very little modification for quantum mechanical behavior, captured the physics strikingly well.

Electricity is the flow of charge, and in a material is caused when electrons move in response to an applied field, which is most commonly in response to an applied electrical field (although charge flow can be induced by an applied temperature gradient, thermoelectricity, or a stress field, piezoelectricity, to name a few). The current density J, or charge per unit area per unit time, is given in terms of how much charge Δq, passes through a cross sectional area A, in a given time interval Δt, by

$$J = \frac{\Delta q}{A\Delta t}.$$

In an experiment, the current density is typically a macroscopic quantity, usually measured in a millimeter scale device. The physics that determines the current density is, however, a microscopic quantity governed by processes that takes place in the range of nanometers. This is of ultimate interest to us, and, therefore, we must connect these microscale processes to the macroscale measureable quantities. To do this we can recast the current density in terms of the average velocity of charges as

$$J = qNv_d$$

where N is the number of charges per unit volume, q is the charge of individual carriers ($-e$ for electrons, and $+e$ for holes) and v_d is the drift velocity given by

$$v_d = \frac{1}{M}\sum_j^M v_j$$

with M being the total number of free charges in the material, each having a velocity v_j. The drift velocity provides a connection between the macroscale average velocity that determines the current, and the velocities of the individual charges that occurs on the microscale.

With an applied electric field E, each charge in the material feels a constant electrostatic force given by $F = qE$, and, therefore, should feel a constant acceleration, and the velocity should increase with time as $v_j = at + v_{0,j}$, where $v_{0,j}$ is the initial velocity prior to the application of the

electric field. With this, the drift velocity would be proportional to *t*, and, therefore, one would expect the current density should increase with time; this predication is in contrast to the well-known behavior of how electricity flows through a material. What is experimentally seen is that the current density, and, therefore, the drift velocity, initially increases with time; however, it quickly reaches a steady state as the material begins to dissipate heat. Therefore, the electrons must be dumping the energy they accumulated during their acceleration through an inelastic scattering process.

For moving charges in the material, Bloch's theorem implies that inelastic scattering must come from non-periodic potentials (see [10]Ashcroft and Mermin, 1976;[81]Grosso and Pastori Parravicini, 2000), and, therefore, as shown in Figure 3-1, the dissipative processes must be associated with scattering from non-periodic defects in the crystal lattice or other mechanisms such as phonons. Since these scattering events dissipate energy, the incoming and outgoing velocities are neither of the same magnitude nor direction. This allows us to treat each charge

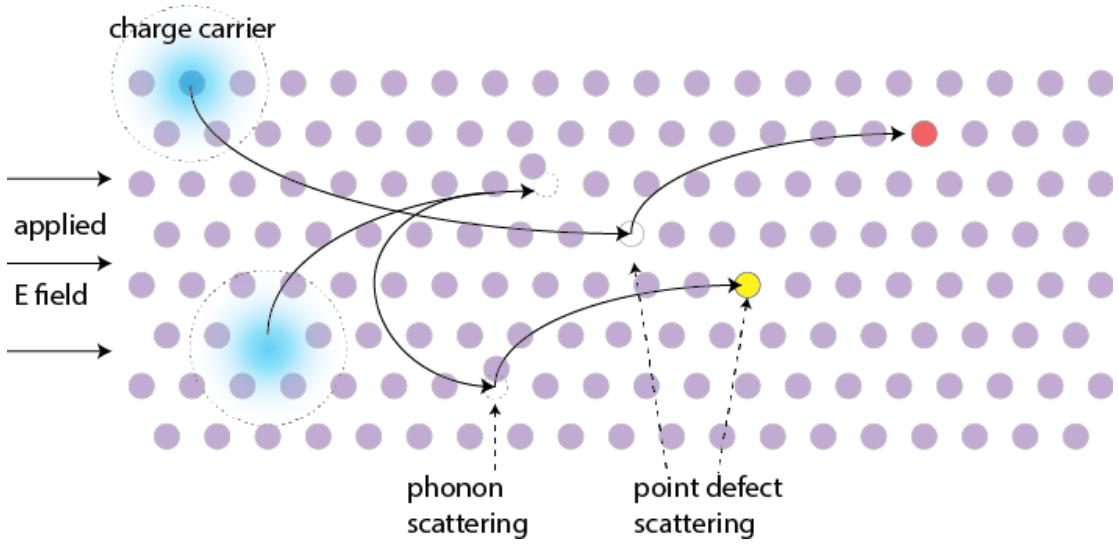


Figure 3-1 A schematic of the acceleration of free carriers and the scattering processes in a material under an applied electric field. Scattering is caused by non-periodic potentials within the crystal lattice. Shown here is scattering by phonons (displacements of the atoms in the crystal), and point defects, which can be foreign atoms or vacancies.

carrier as starting from a random velocity $v_{0,j}$, following the last collision at time $t_{0,j}$; the final velocity at the time of the subsequent collision t_j, is then given by

$$v_j = v_{0,j} + a(t_j - t_{0,j}) = v_{0,j} + \frac{qE}{m_q}(t_j - t_{0,j}).$$

Since it is impossible to know the precise initial velocity nor the time between collisions for each charge, the only way to proceed is to average the previous equation and use some sort of mean time between collisions. When averaged over all particles, the average velocity becomes the drift velocity

$$v_d = \frac{1}{M}\sum_j^M \left(v_{0,j} + \frac{qE}{m_q}(t_j - t_{0,j}) \right) = \frac{1}{M}\sum_j^M \frac{qE}{m_q}(t_j - t_{0,j}) = \frac{qE}{m_q}\tau$$

where the average of $v_{0,j}$ is zero since the charges start in random directions with random speeds, and $\tau = M^{-1}\sum(t_j - t_{0,j})$ is the average time between collisions, which is often called the relaxation time. With this, we can substitute back into the microscopic equation for the current density to obtain

$$J = qNv_d = \frac{q^2 N\tau}{m_q} E = \sigma E$$

where Ohm's law is used, which states that the current density is proportional to the applied electric field with the proportion constant being the conductivity $\sigma = \frac{q^2 N\tau}{m_q} = qN\mu$, where $\mu = \frac{q\tau}{m_q}$, is called the mobility.

Measurements of conductivity and carrier density (see 3.2) yield the mobility. Macroscopically, the mobility can be thought of simply as how much conductivity each charge carries, i.e. $\mu = \sigma/(Nq)$. This, however, is ultimately determined by microscopic qualities, which are fundamental properties of the material such as charge carriers' mass (see below), and the dominant scattering processes, which we will be of particular interest throughout this work.

The simple Drude theory, which is correct in form, does require some modification for quantum mechanical effects that are not captured by the semiclassical theory. Particularly, it was

found early on that the charge mass m_q, varies from metal-to-metal, and while close to the free electron mass in some materials, in other materials, m_q could be several orders of magnitude larger (aptly called heavy Fermion materials), or in semiconductors, m_q can be drastically smaller than m_e. We mentioned previously that Bloch's theorem states that electrons cannot be scattering by the periodic potential of the lattice, but despite this, the ionic potentials of stationary atoms in the lattice do have an effect on the motion of the electrons. To intuitively see this we can, again, imagine a charge that has scattered, and begins to accelerate, then is scattered again. The average distance between scattering events, called the mean free path, is generally much larger than the lattice spacing, and, therefore, as the charge moves between scattering events, it passes by many lattice sites. When moving across these lattice sites, the charge does interact with the atoms in the lattice, which in-turn affects the motion of the charge; we can model this semiclassically as an effective force $F_{Lattice}$. With this, the charge accelerates according to $m_q a = F_E + F_{Lattice}$. Since it is complicated to know the exact form of $F_{Lattice}$, we can divide through by m_q which yields $a = (F_E + F_{Lattice})/m_q$, because the acceleration is well defined. This enables us to redefine the right-hand-side to be $(F_E + F_{Lattice})/m_q = F_E/m^*$, where m^* is called the effective mass, and takes into account the effects of the lattice on the mass of the charge. This effective mass is a fundamental property of the bonding of the underlying atomic structure (see [10]Ashcroft and Mermin, 1976;[57]Davies, 1998;[81]Grosso and Pastori Parravicini, 2000). This renormalized mass can range from several orders of magnitude smaller than the free electron mass in semiconductors where the *p* and *s* orbital overlap is large and therefore the effects of the lattice are small, to materials with *d* and *f* orbital's where the atomic overlap is small and the effective mass can rival that of the proton mass (~$1000m_e$). Describing the charges that move through a crystal as nearly free electrons that have been dressed up with a new mass, is a remarkable emergence of the collective behavior. This ability to recast strongly interacting electrons as weakly interacting *quasiparticles* originated with the work by Lev Landau, and has been one of the major triumphs of many-body physics of the twentieth century (see [52]Coleman, 2003).

3.2 Hall effect: Determination of the charge carrier density

The number of free carriers in a material can give deep insight into the physics that is at play. It determines whether the carriers move along, basically not seeing each other (uncorrelated), or if the position of each electron is determined by its neighbor's position (strongly correlated). This also determines whether a material is a good metal or on the verge of transforming into an insulating state. The method for determining the number of charge carriers dates back to the work of Edwin Hall (hence the name Hall effect, see [84]Hall, 1879), and is simple enough to be part of most undergraduate physics lab projects.

The Hall effect derives from the Lorentz response of a charged particle to an applied magnetic field. Imagine a sheet of a homogeneous material of width w (y direction) and thickness t (z direction), and, as shown in Figure 3-2, if we apply a current I along the x axis and a uniform magnetic field along the z axis with magnitude B, then the carriers in the materials will respond by deflecting towards the positive y direction according to the right-hand rule. With this deflection, there will be more carriers at the top of the material than the bottom as shown in Figure 3-2, which will create an electric field in the y direction given by E_y. The Lorentz force that the charge carriers feel is given by $F_L = qv_dB$, and in steady state conditions the Lorentz force will eventually be equilibrated by the electric field associated with the buildup of charge thereby providing the following relation $qE_y = qv_dB$. Experimentally, E_y can be indirectly found by measuring the voltage transverse to the applied current (Figure 3-2). This voltage is given by $V_{Hall} = \int_0^w E_y dy = E_y w = wv_dB$, with w being the width of the material. From this, the Hall resistance is found by dividing the Hall voltage by the total applied current, which can be recast in terms of the current density times the cross sectional area, $A = wt$. This then yields $I = JA = J(wt) = qNv_d(wt)$, and therefore $R_{xy} = V_{Hall}/I = (wv_dB)/(qNv_dwt) = B/(qNt) = B/(qn)$, where $n = Nt$ is the number of charges per unit volume times the thickness (interchangeably referred to as n_{2d}), which gives the total number of carriers per unit area of the sample, simply called the areal carrier

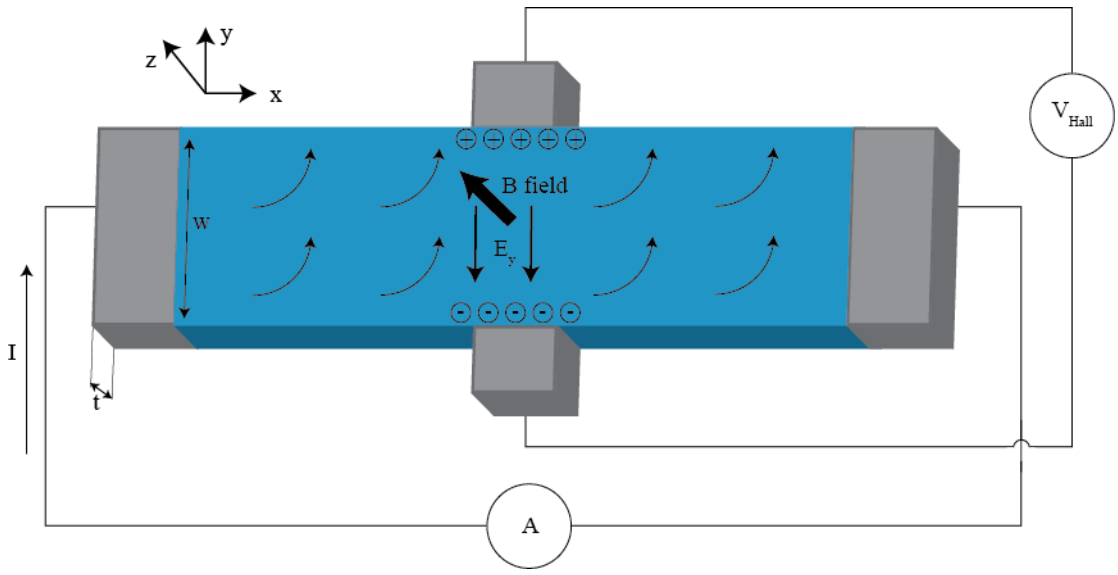


Figure 3-2 A schematic of a Hall effect measurements on a sample of width *w* and thickness *t*. The current flows along the *x* direction, and the magnetic field points out of the page (along the *z*). The Hall voltage V_{Hall} is measured perpendicular to this along the *y* direction.

density. In addition to measuring the carrier density, the sign of the slope of R_{xy} allows one to distinguish electrons (negative charge carriers/negative slope) from holes (positive charge carriers/positive slope).

3.3 Transport measurement: Practice

3.3.1 Van der Pauw and Hall bar geometries

For low frequency (below ~gigahertz) transport measurements, one must attach metallic leads in order to direct the electric field into the sample, and to extract the proper resistivity values the material must be of a certain, well-defined shape. As shown in Figure 3-3, the two dominate geometries used in direct current (DC) transport measurements are Hall bar and Van der Pauw (see [57]Davies, 1998). Due to its simplicity, the main geometry used in this work was the square Van der Pauw because the films were grown on square substrates and this does not require an extra processing step as Hall bar would.

Both Hall bar and Van der Pauw are examples of four-probe measurement techniques, which, as the name implies, requires at least four leads to be attached to the sample. The most common notion of resistance measurements are two-probe techniques (picture measuring the resistance of a resistor with a handheld multimeter), which sources current and measures the corresponding voltage drop along the same leads. A complication that arises from using the two probe technique is this measures the resistance of the leads and the resistance of the contacts in series with the resistance of the sample. This is problematic because it is impossible to determine if a measured quantity originates from the material of interest, or is due to the lead or contact resistance. Making direct contact to the sample with both the voltage and current probes (i.e. two probes for current and two more for voltage, hence four probes) eliminates both contact and lead resistance. This relies on the input impedance of the source meter being very high, which limits the amount of current flowing through the voltage leads, which would thereby create additional voltage drops within the leads that would introduce lead and contact resistances. Standard laboratory source meters can have input impedances in excess of 10^{10} Ω, thus making any lead or contact resistance completely negligible.

The Van der Pauw configuration assumes the following: (1) the sample is of uniform thickness. (2) The sample is simply connected (no holes). (3) The sample is relatively isotropic. (4) The leads are located at the edges of the sample. (5) The contact area of the leads is much smaller than the total area of the film. With these assumptions L. J. Van der Pauw showed, in the late 1950s, that one can obtain the resistivity (with thin films the sheet resistance R_S, is a more typical quantity, which is related to the resistivity by $\rho = R_S t$) through an averaging procedure with the leads in various locations ([188]Van der Pauw, 1958;[189]Van der Pauw, 1958).

To measure the sheet resistance in Van der Pauw geometry, one must measure the raw resistance along two sides of the film. As shown in Figure 3-3a, in the longitudinal Van der Pauw geometry, the first measurement is done by sourcing current from lead 1 to 2, and measuring the corresponding voltage drop across 3 and 4; we define this to be $R_{12,34} = V_{34}/I_{12}$. The second

measurement is done by rotating the leads by 90° and gives $R_{14,23} = V_{23}/I_{14}$. Technically for full Van der Pauw, one should average the reversed resistance values with the leads rotated 180°, $R_{34,12}$ and $R_{23,14}$, but the reciprocity theorem in 3.3.2 states that $R_{12,34}(B) = R_{34,12}(-B)$, so when $B = 0$ both $R_{12,34}$ and $R_{34,12}$ should be the same, and for resistance measurements in magnetic field this additional averaging is accounted for through a symmetrization procedure (see 3.3.3). With $R_{12,34}$ and $R_{14,23}$ obtained, L. J. Van der Pauw showed that

$$e^{-\pi R_{12,34}/\mathrm{Rs}} + e^{-\pi R_{14,23}/\mathrm{Rs}} = 1.$$

For a relatively isotropic material $R_{12,34}$ can be approximated to be same as $R_{14,23}$, and, therefore, the Van der Pauw equation simple reduces to

$$R_S \approx \frac{\pi}{ln\,(2)} R_{xx}$$

where R_{xx} is experimentally taken to be the average of $R_{12,34}$ and $R_{14,23}$, and $\pi/ln(2)$ is known as the geometric factor. Lastly, in the Van der Pauw geometry, the Hall resistance is obtained by applying the current across opposite corners as shown in Figure 3-3b. Since Hall resistance does not depend on sample geometry, as with sheet resistance, there needs to be no geometric factor applied to obtain the carrier density.

The same transport data can be obtained for the Hall bar geometry in Figure 3-3c. In the Hall bar configuration, the lead configuration is typically better defined than in Van der Pauw, which is susceptible to errors based on the uncertainty in the placement of the leads. Also, in Hall bar the leads are placed away from the active material, and therefore the finite size of the leads plays little role. However, as mentioned before, creating a Hall bar pattern requires extra processing steps to define the shape, and as we will see in Chapter 4, the transport properties of TI materials are susceptible to changes when exposed to atmospheric gases and other contaminants, and therefore the uncertainty introduced by the Van der Pauw geometry is typically much smaller than the changes introduced during these extra processing steps.

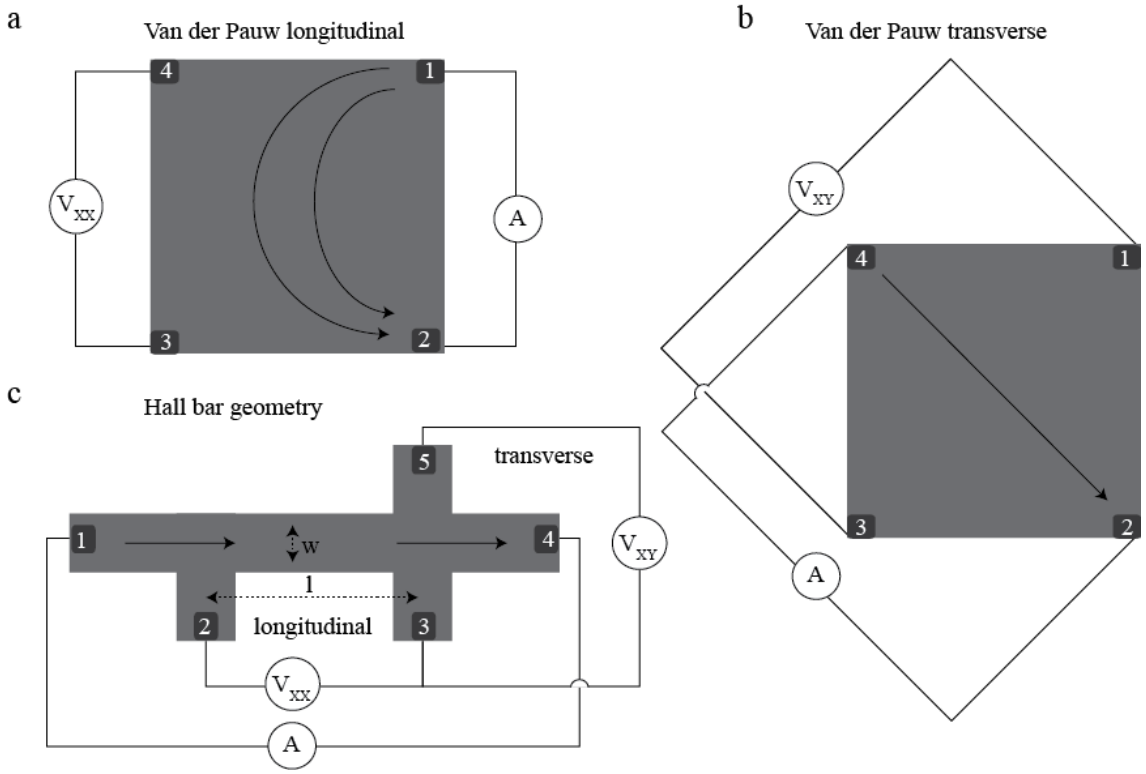


Figure 3-3 A schematic of the different sample geometries used for electric transport measurement. a-b The sheet and Hall resistance can be obtained with square Van der Pauw geometry by measuring in the longitudinal (parallel to the current) and transverse (perpendicular to the current) configuration as shown in **a** and **b**, respectively. **c** The Hall bar geometry where the current is sourced from left to right, and the longitudinal and transverse voltage drops are measured.

To obtain the sheet resistance in the Hall bar geometry, one measures $R_{14,23}$; the geometric factor is simply obtained by calculating the aspect ratio of the rectangle that separates leads number 2 and 3. This yields $R_S = R_{xx}(l/w)$, where l and w are defined in Figure 3-3c. Since R_S, is independent of the overall size, and the lead size does not play a role, effects from sample inhomogeneity can be reduced by using microprocessing techniques and defining a Hall bar in the range of μm, which, when applied to topological materials, has the future possibility of revealing coherent quantum mechanical effects such as Shubnikov de Hass oscillation and the quantized Hall effect, which may be unknowingly canceled due to small inhomogeneities within the sample, when measured using the cm scale Van der Pauw geometry, but this is a digression, and is an issue for future work.

3.3.2 Reverse field reciprocity

A very useful method used in transport measurements that dates back to the work of L. Onsager in the 1930s, is the reverse-field reciprocity relation ([146]Onsager, 1931;[147]Onsager, 1931), which was later expanded on by H. H. Sample *et al* ([169]Sample *et al.*, 1987) and by M. Cornils and O. Paul ([55]Cornils and Paul, 2008). It states that for resistance measurements in a magnetic field (+B), one can obtain the resistance measurement for the reversed field (-B) simply by switching the voltage and current leads without actually changing the magnetic field: for an arbitrary shaped, Ohmic ($J = \sigma E$) sample that is simply connected, if leads are placed at arbitrary positions 1, 2, 3, and 4, if initially, the current flows into lead 1 and out of lead 2 (I_{12}), with voltage probes connected to leads 3 and 4 ($V_{3,4}$), then the measured resistance is given by $R_{12,34} = V_{34}/I_{12}$. Now if a magnetic field B is applied, the resistance becomes $R_{12,34}(+B)$, and if the magnetic field is reversed, the resistance is given by $R_{12,34}(-B)$. The reverse-field reciprocity relation states that $R_{34,12}(+B) = R_{12,34}(-B)$.

With commonly available automatic electronic switch systems (see 3.3.4), the practical application of the reverse-field reciprocity relation comes about because lead reversal is generally much easier than field reversal, due to the large currents and high inductance of typical electromagnets (see 3.3.6). Moreover, with single pole magnetic facilities or permanent magnet systems, the reverse field is not available; one could, in theory, obtain the reverse field values by physically flipping the sample by 180° and performing the measurement again, but with the reverse-field reciprocity relation, one can easily obtain reverse field values in real time simply by switching the current and voltage leads.

The second scenario, where the reverse-field reciprocity relation is useful, is when the carrier density needs to be measured as a function of temperature. One method is to lower the temperature and fix it at discrete points, then scan B while measuring R_{xy} versus field. This is a time consuming procedure that can be complicated by temperature drift during the field scan. The reverse-field reciprocity relation allows this process to be simplified, because one can fix the

magnetic field at $+B$ while measuring $R_{12,34}(+B) = R_{xy}(+B)$, and by the reverse-field reciprocity relation $R_{34,12}(+B) = R_{12,34}(-B) = R_{xy}(-B)$, which enables the calculation of $n_{2d} = 1/(q(R_{xy}(+B) - R_{xy}(-B))/(2B))$ as a function of temperature, without ever actually reversing the magnetic field.

3.3.3 Symmetrization and data analysis

In Van der Pauw geometry, and less so in Hall bar geometry, there is always unavoidable mixing of the longitudinal, R_{xx} and R_{yy} and transverse, R_{xy}, components of the resistance. This is due to the inevitable misplacement of the leads on the sample, as can be seen in the drawing in Figure 3-4a. The experimentally measured voltage-drops will, therefore, always contain some component of the transverse and longitudinal components, which can be written as

$$R_{xy,measured} = aR_{xy} + bR_{xx}$$

$$R_{xx,measured} = cR_{xx} + dR_{xy}$$

where, for well-placed leads, the coefficients b and d are small and a and c are near unity. For resistivity measurements at zero magnetic fields, the value of $R_{xy} = B/(qn) \approx 0$, and, therefore, mixing of R_{xy} into R_{xx} is negligible yielding $R_{xx,measured} \approx R_{xx}$ (R_{xy} at zero field has no meaning, so there is no worry if R_{xx} mixes into R_{xy}). However, in an applied magnetic field R_{xy} is no longer negligible, so the mixing of components is problematic. This issue can be solved, not through better placement of the leads, but numerically, post-measurement, by using the different fundamental behavior of R_{xx} and R_{xy}.

If we go back to the semiclassical equation of motion for a charge under the influence of an applied electric and magnetic field (see [81]Grosso and Pastori Parravicini, 2000) we have

$$m^*\boldsymbol{a} = q\boldsymbol{E} + q\boldsymbol{v} \times \boldsymbol{B} - \frac{m^*}{\tau}\boldsymbol{v}$$

where the last term represents an effective dissipative force. Since in the steady state $\boldsymbol{a}$ is effectively zero, this equation can be rearranged to yield

$$v_x = \frac{q\tau}{m^*}E_x - \frac{eB}{m^*}\tau v_y$$

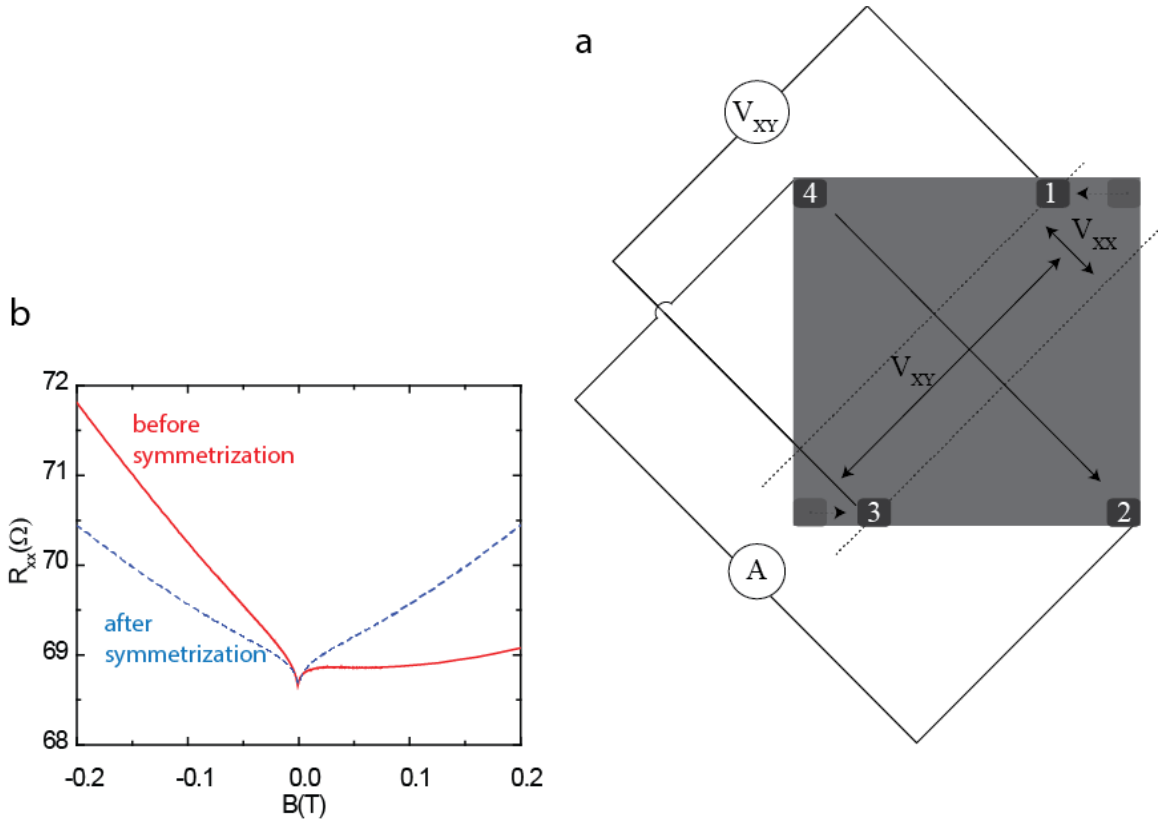


Figure 3-4 A demonstration of the need for, and the result of, data symmetrisation. a A schematic of how the unintentional mispositioning of leads 1 and 3 introduces mixing of into the measurement of R_{xy}. **b** The raw measurement of R_{xx} (solid red curve) is clearly non-symmetric, whereas after the symmetrisation (dashed blue curve), is clearly symmetric.

$$v_y = \frac{q\tau}{m^*}E_y + \frac{eB}{m^*}\tau v_x.$$

These equations can be solved to yield

$$v_x = \frac{q\tau}{m^*}\frac{1}{1+\left(\frac{eB\tau}{m^*}\right)^2}\left(E_x - \frac{eB\tau}{m^*}E_y\right)$$

for the velocity in the x, and for the velocity in the y

$$v_y = \frac{q\tau}{m^*}\frac{1}{1+\left(\frac{eB\tau}{m^*}\right)^2}\left(\frac{eB\tau}{m^*}E_x + E_y\right).$$

Recalling the $\boldsymbol{J} = qn\boldsymbol{v} = \sigma\boldsymbol{E}$, where σ, now explicitly a matrix, can be assembled out of the previous expressions to be

$$\sigma = \frac{ne^2\tau}{m^*}\frac{1}{1+\left(\frac{eB}{m^*}\tau\right)^2}\begin{pmatrix} 1 & -\frac{e\tau B}{m^*} \\ \frac{e\tau B}{m^*} & 1 \end{pmatrix},$$

which can be inverted to yield the resistivity matrix

$$\sigma^{-1} = \rho = \frac{m^*}{ne^2\tau}\begin{pmatrix} 1 & \frac{e\tau B}{m^*} \\ -\frac{e\tau B}{m^*} & 1 \end{pmatrix}.$$

The diagonal components of the resistivity matrix are related to the longitudinal resistance, as shown previously, and are *even functions of magnetic field*, while the off diagonal elements are the transverse Hall resistance, which is an *odd function of the magnetic field*. This shows that under the reversal of magnetic field, the transverse and the longitudinal parts of the resistance have opposite behaviors, which can be used to separate them mathematically.

To eliminate the odd components from $R_{xx,measured}$, we can use its evenness under magnetic field. For $+B$

$$R_{xx,measured}(+B) = cR_{xy}(+B) + dR_{xx}(+B)$$

and for $-B$

$$R_{xx,measured}(-B) = -cR_{xy}(+B) + dR_{xx}(+B).$$

Therefore assuming c ≈ 1, and adding the two previous equations, we get

$$R_{xx}(B) = (R_{xx,measured}(+B) + R_{xx,measured}(-B))/2.$$

Similarly, to remove the even component from R_{xy} we get

$$R_{xy,measured}(+B) = aR_{xy}(+B) + bR_{xx}(+B)$$

and

$$R_{xy,measured}(-B) = -aR_{xy}(+B) + bR_{xx}(+B),$$

which, assuming $a \approx 1$, and subtracting the two previous equations yields

$$R_{xy}(B) = (R_{xy,measured}(+B) - R_{xy,measured}(-B))/2.$$

In an actual experiment, R_{xx}, R_{yy}, and R_{xy} are measured at discrete B values (i.e. R_{xx} might be measured for positive values at B_n = 1.25, 1.28, 1.31 etc, while the corresponding negative values could be B_n = -1.24, 1.27, 1.30 etc). This, then, requires one to change the sign of B (i.e. $-B_n = -1 \times B_n$), and then using a numerical interpolation function to calculate the corresponding resistance

values at the discrete negative values of the field. In practice this is all done with a script written in the data analysis program Origin, and takes only a few seconds to complete, and yields very good results, as can be seen by comparing the data in Figure 3-4b before and after symmetrization.

3.3.4 Sample preparation and instrumentation

With a sample that is ready to be measured, the first thing to do is to securely attach it to the sample holder. This needs to be done in a way that maintains good thermal contact between the sample holder, and the substrate, and thus the film. This is achieved using double-sided copper tape. The tape used was 3M 1182 tape that is made from 1 cm wide soft copper foil of thickness ~ 0.002", and the adhesive is a special formulation with high electrical and thermal conductance attained by mixing in small metal particles. Prior to attaching the sample, the tape is sized to fit on the sample holder, slightly less than 1×1 cm^2, and the back liner is removed to expose a single adhesive side. The tape is then pressed into the sample holder, with special attention paid to not forming any air bubbles or folds in the tape. Once the tape is pressed onto the sample holder, the top liner is removed, the sample is place in position and firmly pushed into the tape, and is ready for the leads to be attached.

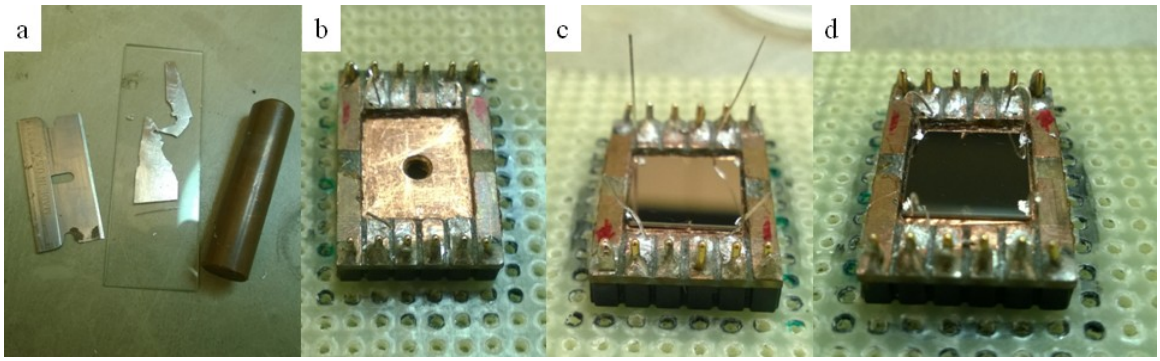


Figure 3-5 Lead and sample preparation. a The indium foil on a glass slide used to create the wires that contact the sample. Shown is the copper rod used to flatten the indium foil and the razor that is used to cut the wire that sits below the foil on the slide. **b** The sample holder used. **c** The sample holder with a sample attached by copper tape, with indium wires prior to contact being made. **d** The same sample, now with the indium leads attached by gently pressing them with toothpicks.

There are many methods to make proper electrical contact, which often depend on the specific types of materials being used. Substantial resistance can arise from the contacts between the metal leads and the semiconductor material. This resistance can be non-Ohmic, in that $J \neq \sigma E \rightarrow V \neq IR$ or may be Ohmic depending on the relative material properties of the sample and the metal that is being used to make contact. When using any material, one should always check this by measuring voltage versus current, which, if the contact is Ohmic, should be linear (i.e. $V = IR$). In theory, the four probe geometry should eliminate any contact resistance, but it is always best to use materials that make Ohmic contacts. In the case of topological material, the lead material

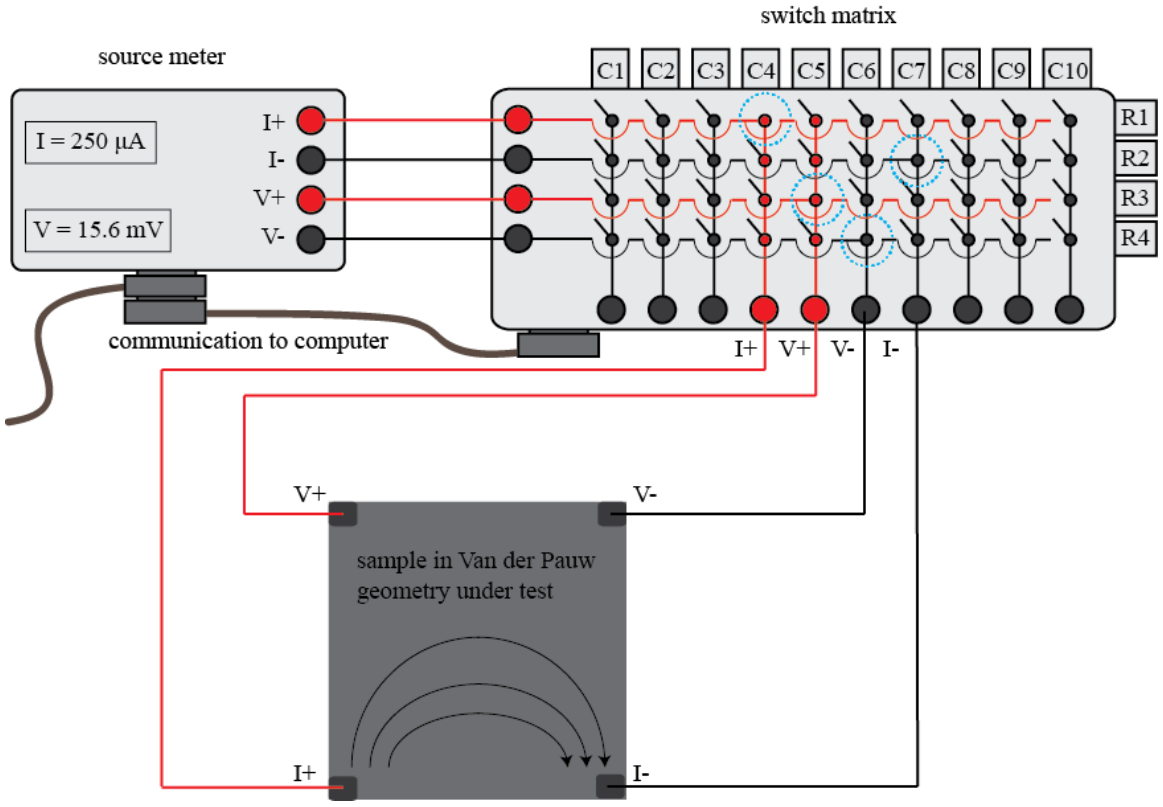


Figure 3-6 A schematic of the source meter switch setup. In the configuration shown, the current and voltage leads go from the source meter to the rows of the switch, and the columns of the switch connect directly to the sample. As shown, the switch is closed (highlighted by blue circles) at R1-C4, R2-C7, R3-C5, R4-C6, thereby performing a longitudinal measurement (R_{xx}). If, however, the switch would be closed at R1-C4, R2-C6, R3-C5, R4-C7, then, the measurement would be the transverse Hall voltage (R_{xy}).

used is indium, which is found to make Ohmic contacts.

Indium is a very soft and ductile metal, which makes forming it easy. To make the In wires, In shot is used. It is first roughly flattened in a clamp, or by squashing it flat with pliers. Then, the flattened piece is placed on a glass slide, and further flattened using a Cu rolling-pin, a metal cylinder (Cu was used simply because there was a small Cu rod on hand, likely any smooth cylinder would do), and gently rolling the rod back and forth over the In, thereby successfully pressing it into a foil with thickness in the range < 0.005", which can be used to make ~ 50-100 individual wires. Once pressed into a thin foil, leads can be made by using a sharp razorblade and cutting thin wires out of the foil, roughly 0.010-0.015" in width and roughly 1 cm long (see Figure 3-5). The wires can be gently picked up with fine tweezers, and one end is placed on the sample holder's contact pad, and pressed by a toothpick to mechanically and electrically attach it. Using toothpicks, the other end is gently directed to the sample surface, and the gently pressed in to make mechanical and electrical contact. This is repeated for all leads. The leads are tested mechanically by nudging with a toothpick to ensure they are attached, and electrically tested by measuring two probe resistance between different leads, and ensuring the resistance values between all leads are relatively close, for typical films ~ 1-2 kΩ. The sample holder is then attached to the cryostat (see 3.3.5), and transport measurements can begin.

The electrical measurements are performed using a source meter in combination with an electrical switch system. The source meter used was typically a high impedance single channel Keithley 2400, but a Keithley 2636 dual channel source meter was also used. The electrical switch used was a Keithley 7001, which is used to automatically change the lead configuration in about ~0.1 seconds. The current and the voltage leads were attached to the rows of the switch, and the leads that were attached to the sample were connected to the columns of the switch as shown in Figure 3-6. Inside the switch matrix all rows are electrically attached to all columns through electro-mechanical relays (electro-mechanical on-off switches). When the switch receives a specific signal from the computer to, say, close R1 to C7, then the relay that connects

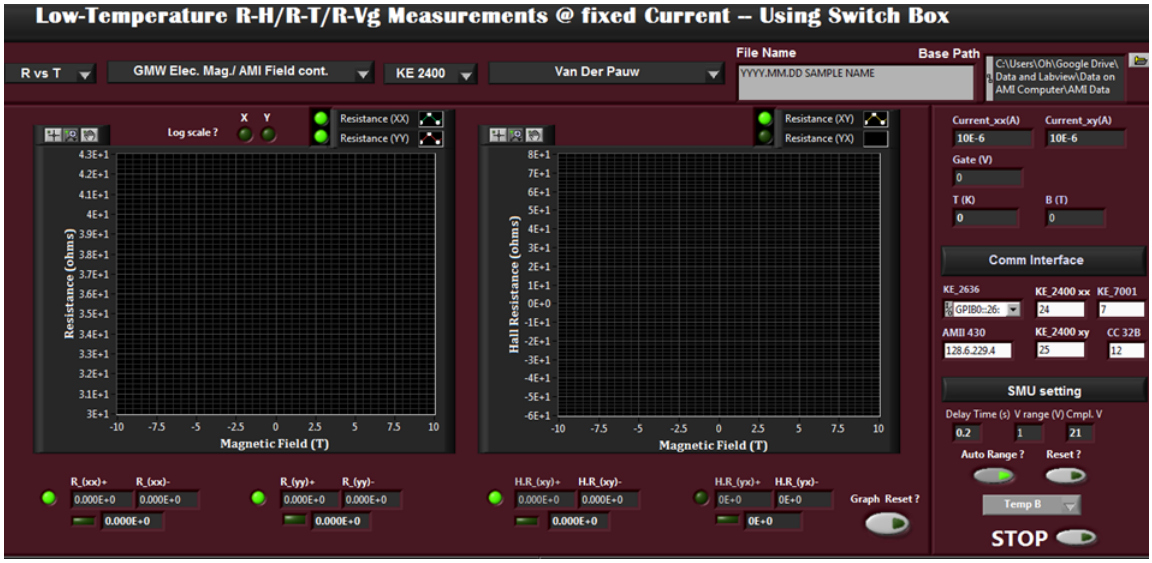


Figure 3-7 The front interface of the custom program designed and written to perform transport measurements. At the top are options, from left to right are measurement resistance versus temperature, magnetic field, or gate voltage, magnetic field system, source meter, and final measurement geometry. On the right side are various parameters that are adjusted for each particular measurement. Within this one program, a host of measurements can be performed all within the same interface.

R1 to C7 shorts those two lines together. In this way the various connections can be quickly changed between the different Van der Pauw configurations, thereby performing measurements of R_{xx}, R_{yy}, and R_{xy} together with only a short delay between each.

The measurement process is controlled automatically through communication with the computer via GPIB lines. The control program was home-built using the LabView programming language, and was designed for flexibility in the measurement process. As shown in Figure 3-7, the user-interface has many options for measurement of resistance versus temperature, magnetic field, or gate voltage, with options for source meter, magnet system, and sample geometry. The program automatically switches and measures R_{xx}, R_{yy}, and R_{xy}, with the specific switch configuration being changed inside the LabView code.

Figure 3-8, shows a flow chart of the logic used with the measurement program. First, for measurement of R_{xx}, the program closes the switch at the specified locations so the current and

voltage probes are attached to the correct leads to measure R_{xx}. The program measures the temperature, magnetic field, and gate voltage. Then, the specified current is sourced, and after a short delay, which allows the current to stabilize, the voltage is measured. This delay is very important because spurious effects can be introduced into the measurement; the time constant should be longer than the resistance-capacitance time constant for the circuit, where the capacitance and resistance are for the entire circuit. Typically for a metallic sample, a value of around ~0.1 second is sufficient, but for more insulating samples, values upwards of 1 second may be needed. Following the measurement of R_{xx} with $+I$, the current is then reversed to $-I$, and a second voltage measurement is made. Lastly, the two results are averaged, stored and plotted in real time. The averaging of both positive and negative currents is critical because there can be spurious effects, typically thermoelectric effects (thermoelectricity is the creation of an electric field in response to a thermal gradient), within the sample, and from the wires that connect the

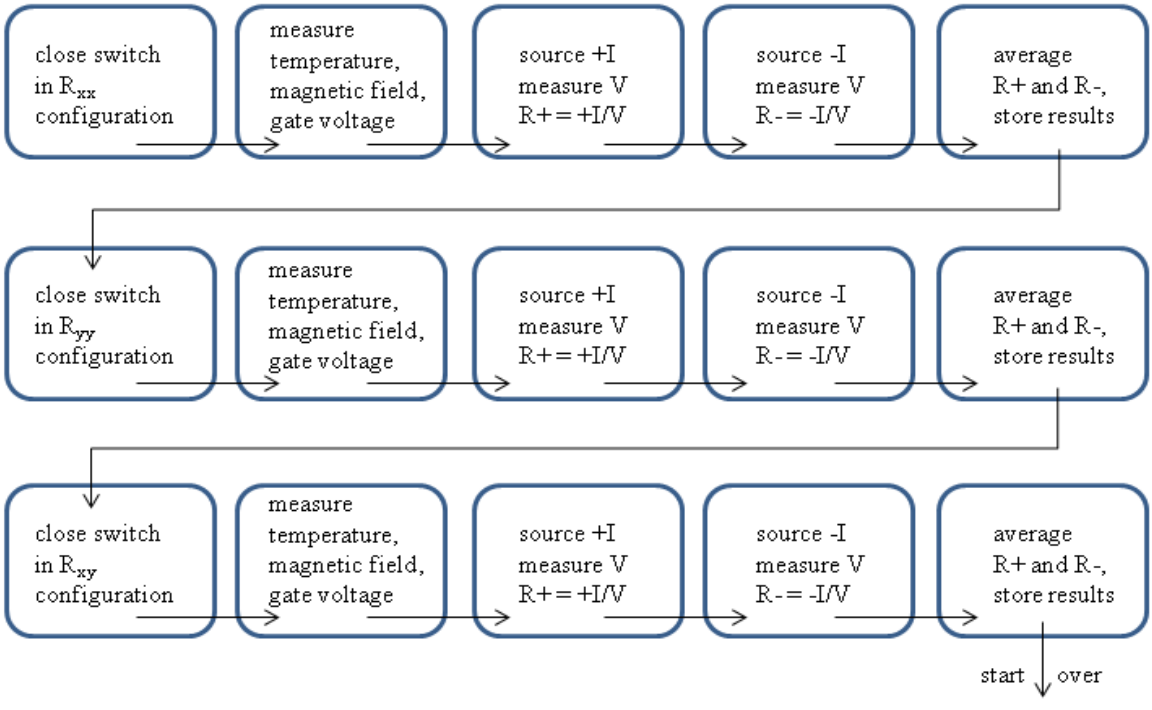


Figure 3-8 The flow diagram for the progression of the measurement process.

electronics, at room temperature, to the sample inside the cryostat, at low temperature, thereby introducing inaccuracies into the measurement. The forward and reverse current averaging eliminates these effects because the polarity of thermoelectric voltage does not depend on the direction of the sources current.

3.3.5 Cryogenics

Most quantum mechanical effects in transport measurements can only be seen at low temperatures. This is due to the large thermal fluctuations that destroy coherent quantum phenomena. In addition to this, the behavior with changing temperature gives insight into the physics at play in a particular material, which, all together, motivates us to study the transport properties of these topological materials at low temperatures. The typical base temperatures used in this work are between 1.5— 6 K (-272 — -267C, or -457— -449F), which is achieved separately in two different systems. Here, we will briefly discuss the operational principles of the systems that make it possible to probe these materials at low temperature (see [14]Balshaw, 1996 for a good introduction to cryogenics).

The first system we will discuss uses liquefied helium and nitrogen to achieve low temperatures and was designed and manufactured by American Magnetic Incorporated. As shown in the image and cut-away diagram in Figure 3-9, the outer container is simply a vacuum Dewer used to minimize thermal leakage from the outside walls at room temperature to the cryogenic liquids within. The outer most chamber is a liquid nitrogen (LN, 77 K) belly that is used to thermally insulate the liquid helium (LHe, 4.2 K), which is stored in the chamber inside and below the LN belly. Both LN and LHe are bought (at about $1000/ 100 liter Dewer for LHe) preliquefied in supply Dewers and manually transferred into the system. Situated inside the LHe belly at the bottom is a 9 T superconducting magnet (see 3.3.6), and at the center of the bore of the magnet is the sample space. The sample space consists of a vacuum space that isolates the sample tube. The space that separates the LHe belly from the sample tube must be in vacuum to

minimize heat leakage from the sample tube into the LHe belly while the sample is heated to room temperature. Lastly, the sample is attached to a variable temperature insert that is removable for sample attachment, then inserted into the sample tube for measurement.

To cool the sample from room temperature to 4.2 K, the sample is inserted at room temperature and a small valve is opened that links the LHe belly to the sample tube. This valve, called a needle valve, allows a small controlled flow of LHe into the sample space, which brings the sample down to 4.2K. Further cooling can be achieved by valving off the sample tube, and fully opening the needle valve, which allows a significant amount of LHe to flow into the sample space, at which point, the needle valve is closed and the vacuum line to the sample tube is reopened to begin the evaporation of the LHe. Pumping this LHe in the sample tube allows

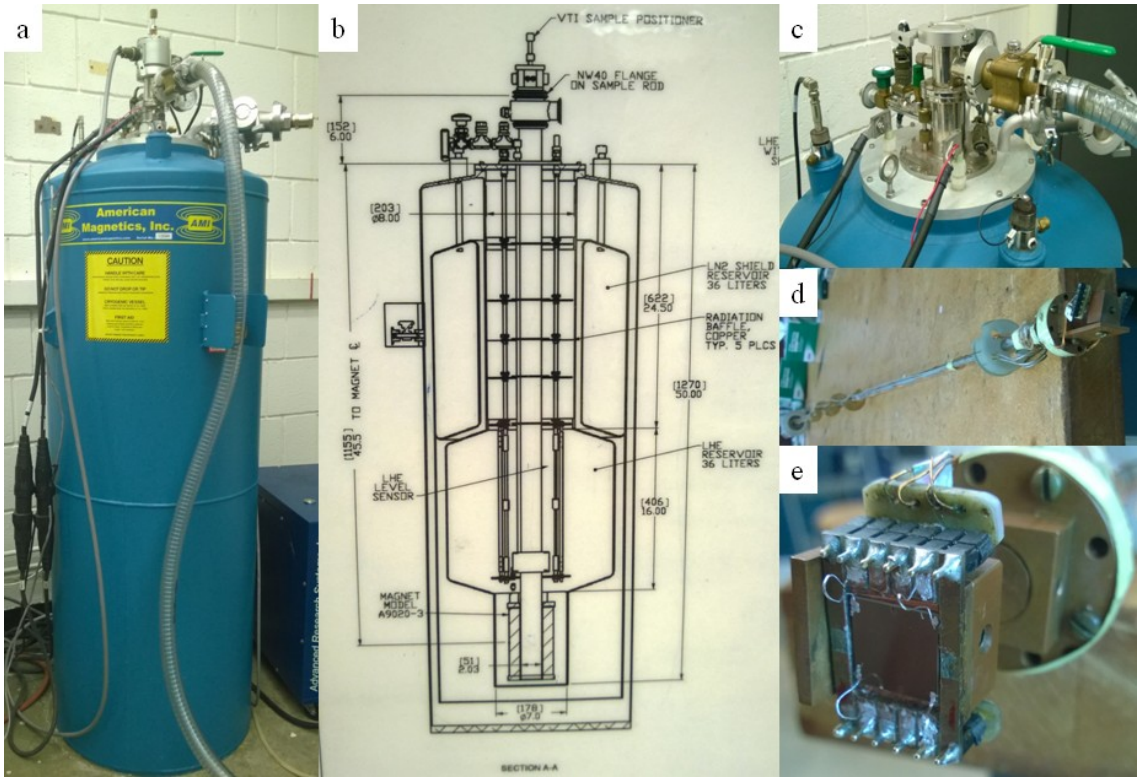


Figure 3-9 Images of the American Magnetic, Inc liquid helium cryostat used to perform transport measurements down to 1.5 K and up to 9 T. a-c Images of the Dewer, along with a cut-away showing the inner construction (courtesy of American Magnetics, Inc.). **d** The variable temperature insert, which has been removed (compare upper most component in **a** to **c**). **e** The end if the variable temperature insert with the sample holder and a sample attached prior to being inserted into the cryostat for measurement.

cooling down to 1.5K, which the sample will remain at for ~3-5 hours while the measurement is performed.

The second cryogenic system used was designed and manufactured by Advanced Research Systems (see Figure 3-10), and is a closed-system that does not use any preliquified cryogenic liquid (at significant cost saving), but rather, operates by compressing and cooling, and

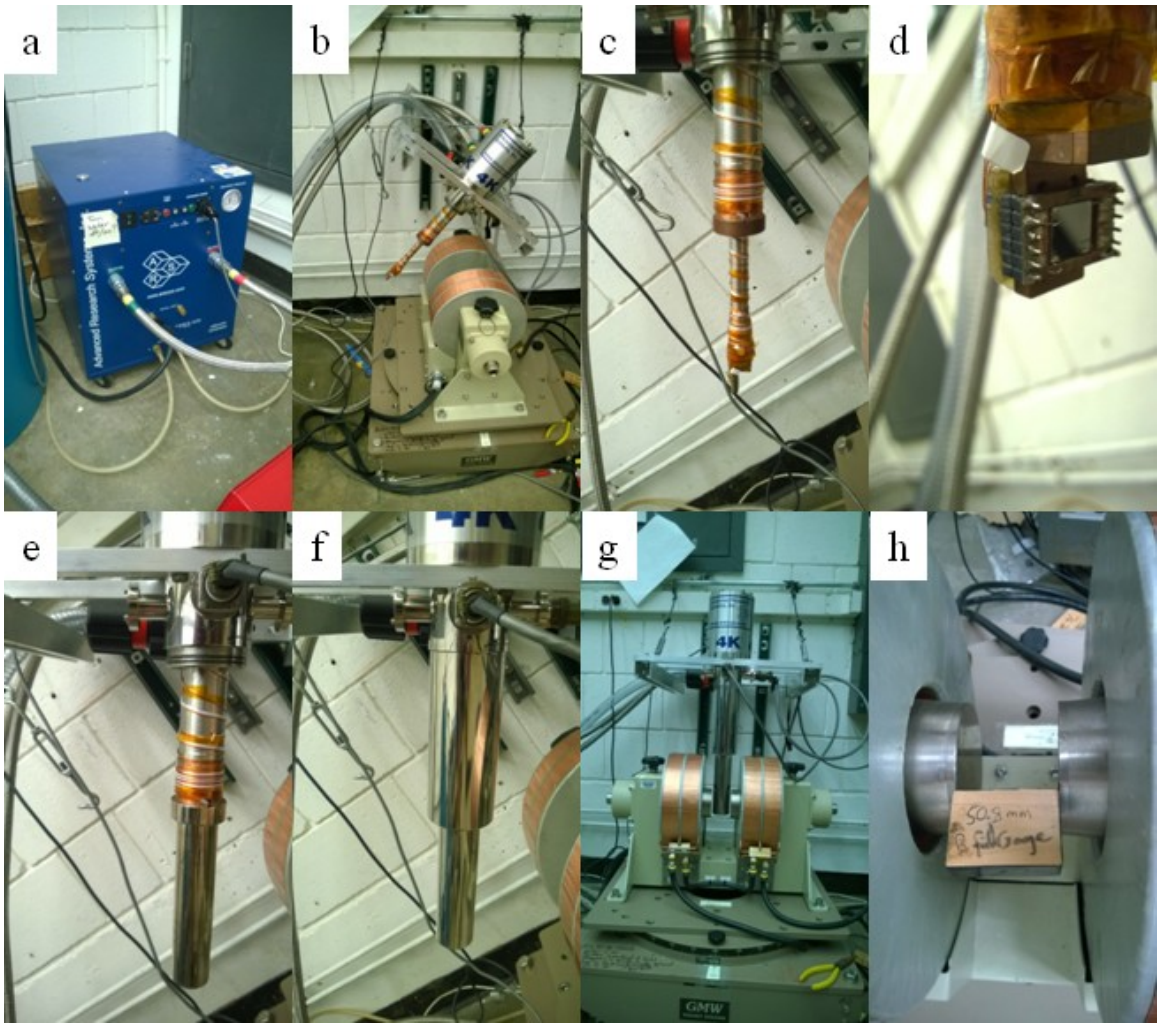


Figure 3-10 Images of the Advance Research Systems cryostat and GMW magnet used to perform transport measurements down to 4 K and up to 0.6 T. a The helium compressor. The green line is the supply line containing compressed helium cooled to room temperature, while the red line is the expanded helium return line. **b-c** The cold head rotated out of the magnet, prior to a sample being loaded. **d** A sample mounted on the cold head. **e** The cold head with the nickel coated copper radiation shield attached to the 20 K stage. **f** The vacuum shroud attached (note the double O-ring vacuum seals in **e** that the shroud slides over). **g** The cold head rotated down, and positioned within the bore of the GMW electromagnet with experiment started. **h** The bores of the magnet separated by 50.8 mm.

then re-expanding the helium as a refrigerant gas. The main operation principle of this system is as follows: Externally compressed (~ 300 psi) and cooled (room temperature) helium gas is transferred down a high-pressure line into the chiller head. In the chiller head the He gas is expanded using the four-step Gifford-McMahon refrigeration cycle (see [8]ARS, 2014). The warm, expanded helium gas is then sent back to the compressor where it is compressed and cooled back to room temperature.

This system consists of a cold head where the sample is mounted, which reaches 4-6 K, and is shielded from 300 K radiation by a nickel coated copper can that is thermally locked to the 20 K first stage heat station. The sample space is enclosed by a vacuum shroud, and pumped down to below mTorr level prior to cooling. The entire system is mounted on a rotatable gimbal system that is used to rotate the sample inside the bore of an external electromagnet (see 3.3.6).

3.3.6 Magnet systems

Each cryogenic system is outfitted with its own electromagnet, which enables Hall effect measurements to be performed at low temperature. The AMI system uses a niobium-titanium alloy superconducting magnet, while the ARS system using a GMW Associates electromagnet. Both magnets are controlled by bi-polar four quadrant power supplies used to supply current to the magnets, which are used to achieve 9 T (AMI) and ~0.6 T (GMW) maximum fields in their current configurations. The AMI magnet has an inductance of around 9.45 H, and a field to current ratio of 0.119 T/A. The GMW system has an inductance of around 60 mH, and a field to current ratio of around 0.013 T/A.

The AMI system uses factory installed firmware to control current, which directly controls the ramping of the magnetic field, with maximum ramp rate of ~ 0.024 T/s; using a ramp rate higher than this may cause the magnet to quench (transition into a nonsuperconducting resistive state). The GMW system uses a power supply that only controls current, and the current to magnetic field conversion is controlled through a homebuilt LabView program. This was done,

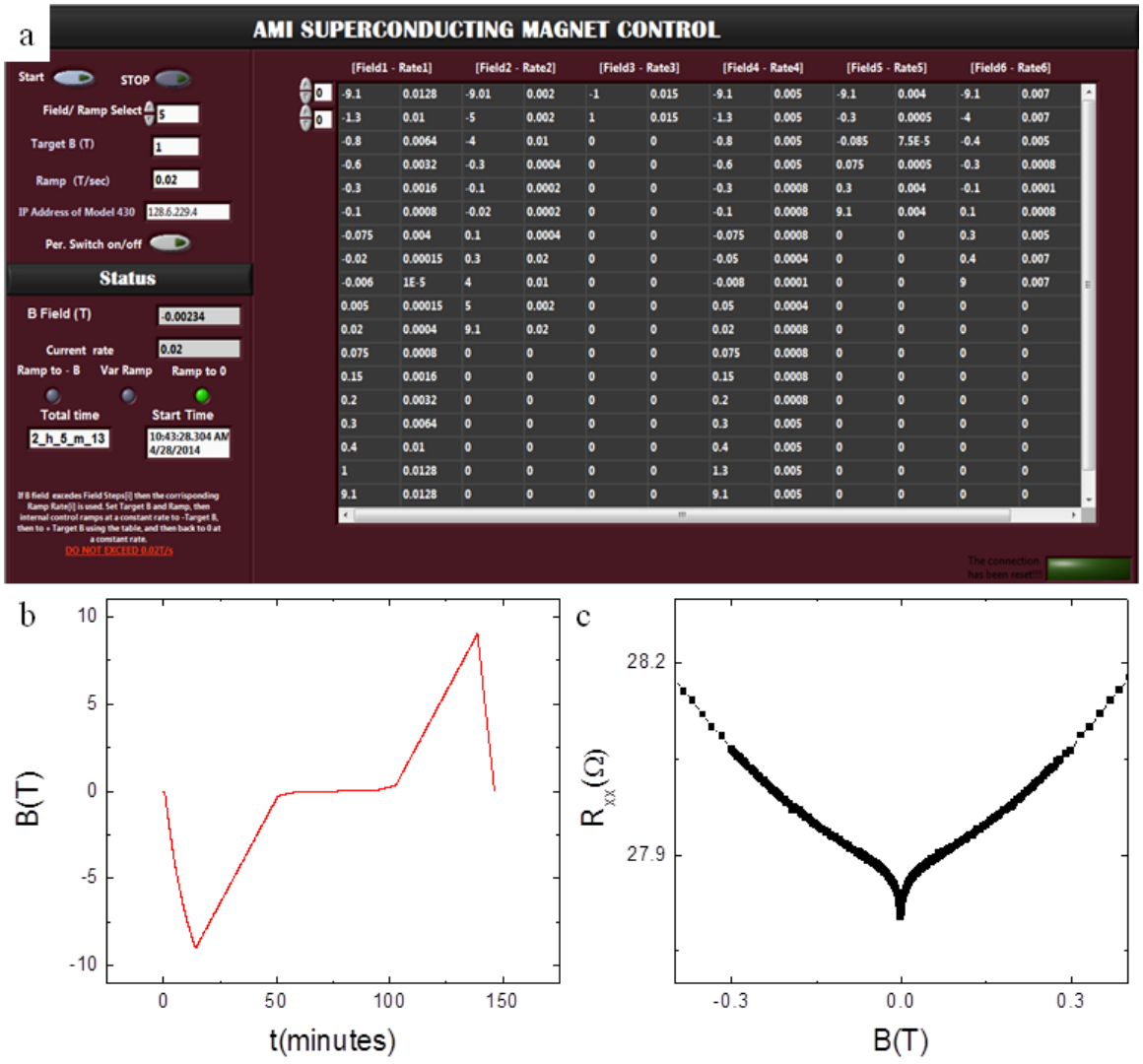


Figure 3-11 The variable ramp rate magnetic field control program. a The front panel of the magnetic control program for the AMI (there is a similar program for the GMW electromagnet). The spreadsheet contains fields and ramp rates, which is selected before the sweep begins. **b** A plot of the magnetic field versus time. The first leg is to ramp from 0 T – -9 T using a constant ramp rate. Then, from -9 T to 9 T using the ramp rates and field change values specified in the select column of the spread sheet. The last leg uses a constant ramp rate to scan to 0 T, where the measurement is automatically stopped. **c** The result is high resolution of the resistance values at low field.

using a calibration curve (current versus magnetic field) provided by GMW at a bore spacing of 55.5 mm (see Figure 3-10h). When checked against the AMI system (simply by measuring the Hall effect on the same sample), the two systems agree within ~3%.

Of particular interest to topological insulators is the weak anti-localization effect (WAL), which is a feature in R_{xx} versus magnetic field. The WAL effect appears as a sharp increase in

resistance with the application of a small magnetic field, and therefore requires a very slow ramp rate around $B = 0$ T; the WAL effect in the highest quality samples require a ramp rate as small as 0.000075 T/s. Using this ramp rate with the AMI's 9 T maximum field would require 18T/(0.000075 T/s) = 70 hours for the sweep to complete, which is impractical. This motivated the development of a LabView program that controls the ramp rate of the magnet based on the magnetic field. As shown in Figure 3-11a, the program using a column of field values and corresponding ramp rates, which it uses to adjust the ramp rate based on the magnetic field as compared to the field values in the column. In this way, the program can slow the ramp rate down near zero field (or any field of interest), and then increase the ramp rate again once the field is beyond $B = 0$T (see Figure 3-11b). The result is a high resolution of resistance versus magnetic field near zero field (see Figure 3-11c), and total sweep time in the range of ~ 2 - 3 hours.

Chapter 4.

Atmospheric effects on the transport properties of Bi_2Se_3

One of the necessities of the measurement process is moving the samples between the ultra-high vacuum that growth by molecular beam epitaxy takes place under, and the measurement system. In this work, primarily transport measurement were performed in the cryogentic magnet systems discussed in Chapter 3, but, other measurements were carried out in collaboration with other groups that range from terahertz spectroscopy, to scanning tunneling microscopy and angle resolved photoemission spectroscopy. Therefore, before one can begin to understand the fundamental properties of the topological materials, one must understand the chemical role that environmental gasses play. In the case of topological materials this is of special importance because the topological surface states (TSS), which are the main interest in these materials, sit on the interface that is exposed. In this chapter we discuss what effects atmospheric gases have on the transport properties of Bi_2Se_3 and their implication on the bulk state as well as the TSS ([29]Brahlek *et al.*, 2011).

4.1 Surface versus bulk state in topological insulator Bi_2Se_3 under environmental disorder[4]

It is known that band bending occurs at the surface of Bi_2Se_3 ([22]Bianchi *et al.*, 2010). This is seen by plotting the bottom of the conduction band (CB) versus depth from the surface (Figure 4-1). After cleaving bulk samples, the bottom of the CB near the surface is above that of the bulk, forming a depletion layer (Figure 4-1a). The material then relaxes to the thermodynamic ground state by natural *n*-type doping from either exposure to atmosphere and/or as Se atoms diffuse out, and over time the depletion layer becomes an *n*-type accumulation layer near the surface ([92]Hor *et al.*, 2009;[3]Analytis *et al.*, 2010;[4]Analytis *et al.*, 2010;[22]Bianchi *et al.*, 2010;[115]Kong *et al.*, 2011) (Figure 4-1b). The exact time is highly dependent on the environment, and is estimated to be as little as a few seconds in atmosphere to as long as several hours in vacuum ([3]Analytis *et al.*, 2010;[22]Bianchi *et al.*, 2010). Despite the sensitivity of the Fermi level, ARPES done on bulk

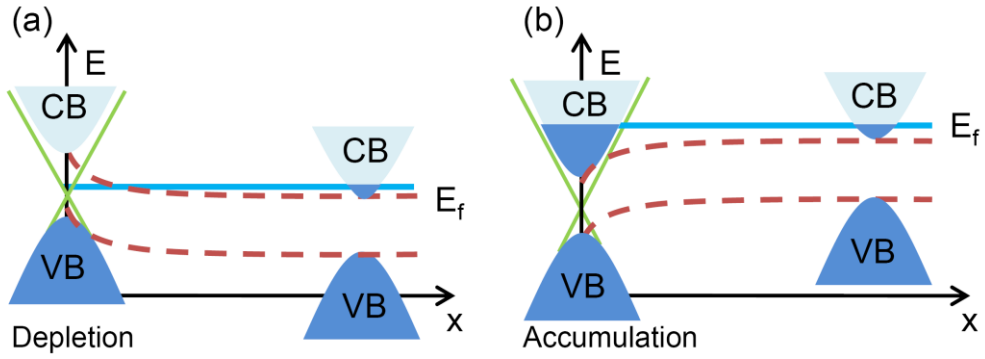


Figure 4-1 A band bending diagram showing the change from depletion to accumulation. The dotted red curves show band bending by tracking the bottom of the conduction band and the top of the valence band (VB) as a function of depth, *x*, from the surface of the material. The blue line, labeled as E_f, represents the Fermi energy. **a** Cleaved bulk samples exhibit an initial depletion region near the surface. **b** Unavoidable *n*-type doping occurs by exposure to atmosphere or vacuum, and the depletion region becomes an accumulations region. Thin films that are grown in vacuum and exposed to atmosphere always result in an accumulation region as in **b**. (Adapted from [29]Brahlek *et al.*, 2011)

[4] This section is based on work that is published in M. Brahlek *et al,* Applied Physics Letters **99**, 012109 (2011).

samples cleaved in atmosphere have shown that the SS still remains resolvable ([3]Analytis *et al.*, 2010), suggesting that the SS survives the atmospheric disorder. However, in transport studies, the robustness of the SS predicted by theory and confirmed in ARPES measurements has never been confirmed. Instead, transport studies of bulk samples showed that the quantum oscillations of the SS, which were measured with the surface in the depletion state, disappears when the Bi_2Se_3 samples are exposed to air for an hour or so ([4]Analytis *et al.*, 2010). Although the latter experiment seems to be in contradiction to the robustness of the SS as observed in the ARPES measurement, in fact, it is not, because as the surface is exposed to air, the depletion layer transforms into an accumulation layer, which washes away the quantum oscillation signal even if the SS is still there. In other words, this previous transport study does not imply that the SS disappears with exposure to air, and whether the surface transport properties of Bi_2Se_3 survive atmospheric disorders is still an open question. Our recent achievement ([15]Bansal *et al.*, 2012) of isolating the surface transport properties even with the surface in the accumulation mode has made it possible to address this question here.

Atmospheric gas is mainly composed of N_2, O_2, and water vapor. N_2 is basically chemically inert (a result of the strong triple bond between the N atoms), while O_2 readily ionizes. Therefore O and water are the most probable components to affect the electronic properties of Bi_2Se_3. The effect of O and water has to be taken in conjunction with the effects of the Se vacancies which *n*-dope the material (Figure 4-2a). The rate at which each of these components diffuses in and out of the structure is a complex process, which depends both on the relative concentrations and on the intrinsic diffusion coefficient of each species ([10]Ashcroft and Mermin, 1976); due to the imprecise mixture of atmosphere, it is difficult to label what the ultimate effects on the electronic properties are.

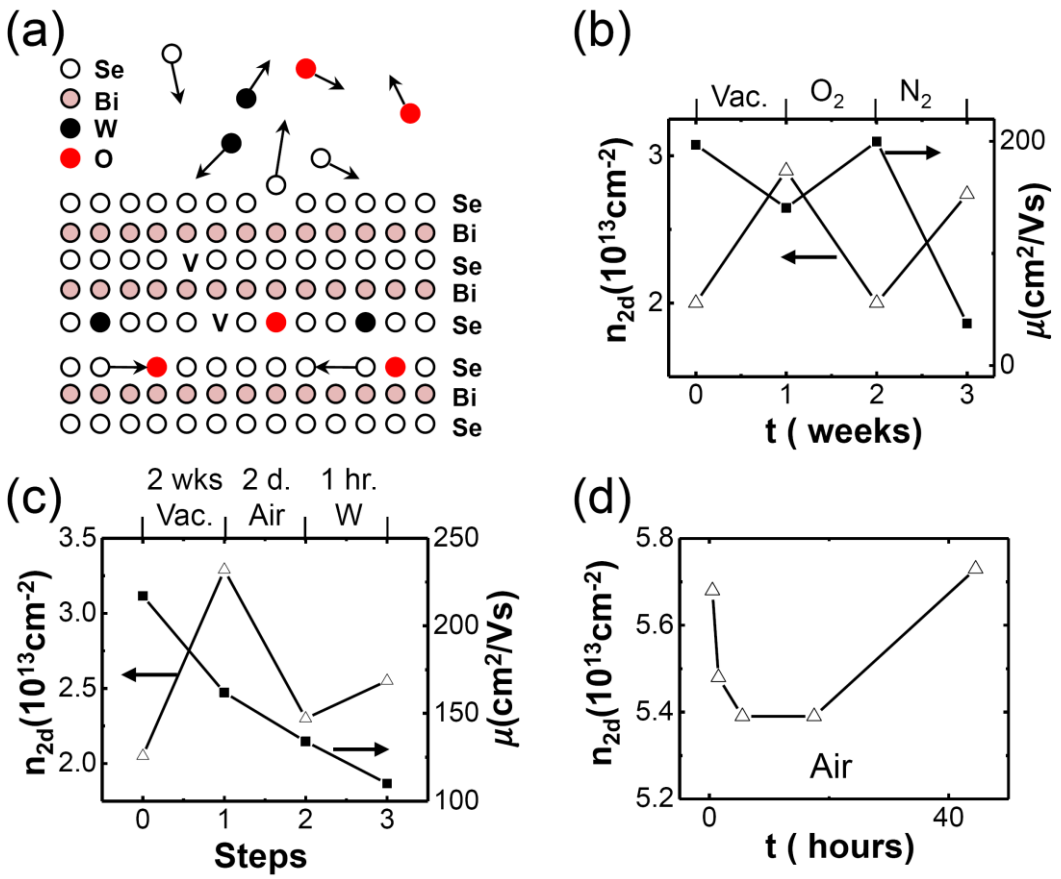


Figure 4-2 The response of Bi_2Se_3 thin films to exposure to atmospheric gases. a A cartoon showing the bulk contamination process. Se atoms diffuse out leaving Se vacancies (V), and O atoms and water molecules (W) naturally diffuse into the structure. **b** The carrier density (left axis open triangles), and the mobility (right axis solid squares); immediately after growth, 1 week in vacuum (Vac), 1 week in O_2 gas, and 1 week in N_2 gas. **c** A similar plot showing the response of the carrier density and mobility to exposure to vacuum for 2 weeks, air for 2 days, and lastly liquid water for 1 hour. **d** The response of the carrier density to air exposure. (Adapted from [29]Brahlek *et al.*, 2011)

To systemically test the effects of each component we exposed an 8 quintuple layer (QL—1 QL ≈ 1 nm) Bi_2Se_3 sample grown on Si to vacuum, and 1 atmosphere of O_2 and N_2 gas for periods of 1 week, and we used the standard Hall effect to determine the transport properties. Figure 4-2b shows the reversible tuning of the carrier density; when exposed to vacuum directly after growth the carrier density increased by 50%, and decreased by 50% when exposed to O_2. Finally, when the sample was exposed to pure N_2 gas the carrier density increased by a similar amount; this implies that N_2 environment is almost equivalent to vacuum with regard to Se

vacancy formation. The corresponding mobilities are also plotted in Figure 4-2b on the right axis; the mobility measurements showed the same reversible effect confirming the origin to be the diffusion of Se and O in and out of the sample. The Se vacancies boost the carrier density and also act as point defects which decrease the mobility. O, which is chemically similar to Se at the impurity level, fills vacant Se slots, and thus pushes the mobility up and the carrier density down. The general trend in mobility is expected to be downward, because various contaminants will generally degrade the sample quality, and thus the fact that the mobility reversibly increased during oxygen exposure indicates that the sample quality remained nearly constant during this time.

Figure 4-2c shows data for another 8 QL sample on Si, which was exposed to vacuum for 2 weeks, followed by atmosphere for 2 days, and lastly liquid water for 1 hour. The carrier

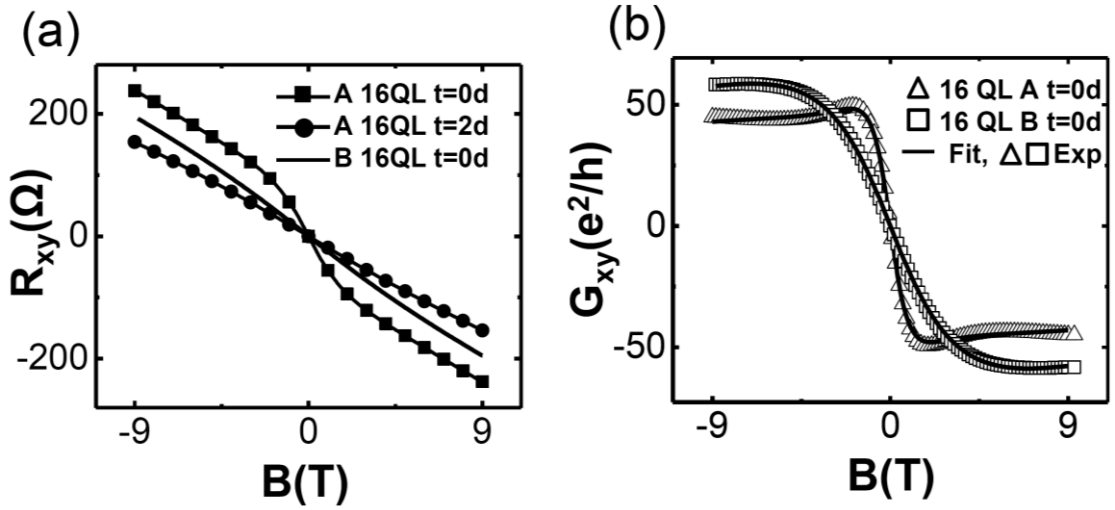


Figure 4-3 The response of Bi_2Se_3 thin films grown on Al_2O_3 to exposure to atmosphere. a Non-linear Hall effect shows that two types of carriers contribute to the conduction. Sample A (solid squares) at $t = 0$ days, shows the lowest bulk carrier density; at $t = 2$ days (solid circles) the bulk carrier density of sample *A* has increased due to *n*-type doping, there-by overwhelming the surface signal and thus reducing the nonlinearity. The solid curve shows data from sample *B* whose bulk carrier density is initially larger. **b** The fitting parameters were extracted using G_{xy} rather than R_{xy} for functional simplicity. The fitted curves are represented by solid curves, while the experimental data are depicted by open triangles (sample *A*, $t = 0$), and open squares (sample *B*, $t = 0$). (Adapted from [29]Brahlek *et al.*, 2011)

density in this sample similarly increased while in vacuum, dropped in atmosphere, but increased again in water. The data shown in Figure 4-2b-c suggest that N_2 (vacuum) and water provide *n*-type carriers through Se out-diffusion or water in-diffusion and only oxygen behaves as a *p*-type dopant; this observation is consistent with previous reports by other groups ([4]Analytis *et al.*, 2010;[45]Chen *et al.*, 2010;[115]Kong *et al.*, 2011). Accordingly, atmosphere, which is composed of N_2, O_2 and water, can function as either a *p*- or *n*-type dopant depending on which of these components dominates. In Figure 4-2c, two days of air exposure shows *p*-type doping, i.e. decrease in *n*-type carrier density, for the 8 QL sample, suggesting dominance of the oxygen effect. On the other hand, another sample (32 QL on Si) in Figure 4-2d exhibited *p*-type doping initially, but after 20 hours, the trend changed to *n*-type doping, suggesting that initial dominance of the oxygen effect was later taken over by the selenium and water diffusion effects. These observations show that atmosphere can behave as both *p*- and *n*-type dopant depending on the detailed competition between oxygen and the rest.

Although the above study of films grown on Si substrates show that environment disorder has significant effects on the transport properties of Bi_2Se_3, they do not resolve the question of how the surface states are affected by these factors, because the carrier densities and mobilities in these samples were dominated by the bulk states ([108]Kim *et al.*, 2011). As shown previously, samples grown on Al_2O_3 substrates enabled separation of the surface and bulk transport properties ([15]Bansal *et al.*, 2012). In the standard Hall effect, R_{xy} scales linearly with the applied magnetic field. However, if two types of carriers contribute, each with a different mobility, say, for surface and bulk respectively, R_{xy} scales nonlinearly with the applied magnetic field (see Figure 4-3). For two types of carriers n_1 and n_2 with mobilities μ_1 and μ_2 respectively, the Hall resistance can be obtained by inverting the conductivity tensor with diagonal elements $G_{xx} = G_{yy}$, and off-diagonal elements $G_{xy} = -G_{yx}$, where

$$G_{xx}(B) = e\left(\frac{n_1\mu_1}{1+\mu_1^2B^2} + \frac{n_2\mu_2}{1+\mu_2^2B^2}\right); G_{xy}(B) = eB\left(\frac{n_1\mu_1^2}{1+\mu_1^2B^2} + \frac{n_2\mu_2^2}{1+\mu_2^2B^2}\right).$$

Upon inversion, the off-diagonal element of the resistance tensor, the Hall resistance, is found to be ([174]Smith, 1978;[175]Steinberg *et al.*, 2010;[15]Bansal *et al.*, 2012)

$$R_{xy}(B) = -\left(\frac{B}{e}\right)\frac{(n_1\mu_1^2 + n_2\mu_2^2) + B^2\mu_1^2\mu_2^2(n_1+n_2)}{(n_1\mu_1 + n_2\mu_2)^2 + B^2\mu_1^2\mu_2^2(n_1+n_2)^2}.$$

Using this function (in addition to $R_{xx}(B = 0)$, and $R_{xy}(B \to 0)/B$ which reduces the 4 parameters down to 2) we could fit the measured data and thereby determine the transport properties of the surface and bulk states ([15]Bansal *et al.*, 2012). In our previous work, we showed that over a range of 2-2750 nm in thickness one carrier density was independent of thickness while the other varied with the thickness ([15]Bansal *et al.*, 2012). From this observation we identified the first type of carriers with the surface layer, confined to 1 QL, and the second with the bulk. This allowed us to fix the thickness of our samples to determine how each of the surface and bulk transport properties changes under atmosphere.

Figure 4-3a and Table 4-1 show the measured data for $R_{xy}(B)$ for two 16 QL samples grown on Al_2O_3. The transport properties for sample *A* were measured directly after growth and measured again 2 days later. The dramatic change in character of the curves can be associated with an increase in bulk carriers. The fitting revealed that the bulk carrier density of sample *A* increased by 300% from 0.30×10^{13}cm^{-2} to 1.1×10^{13}cm^{-2}. However, the surface carriers remained nearly constant during the exposure to atmosphere, changing only 10% from 3.3×10^{13}cm^{-2} to

Table 4-1 Taken from two 16 QL samples grown on Al_2O_3, this table shows the carrier density, n_{SS} (surface state) and n_{BS} (bulk state) and the corresponding mobilites obtained by the non-linear fitting process for G_{xy}. (Adapted from [29]Brahlek *et al.*, 2011)

	n_{SS} (10^{13}cm^{-2})	n_{BS} (10^{13}cm^{-2})	μ_{SS} (cm^2/Vs)	μ_{BS} (cm^2/Vs)
A (*t* = 0.5 hrs)	3.3	0.30	560	6400
A (*t* = 48 hrs)	3.0	1.1	440	1500
B (*t* = 0.5 hrs)	3.0	0.95	510	2100

3.0×10^{13}cm^{-2}. The data obtained from sample *B*, which was grown under identical conditions and tested directly after growth, exhibited almost the same surface carrier density of 3.0×10^{13}cm^{-2}. However the bulk carrier density was 0.95×10^{13}cm^{-2}, which is over three times larger than that of sample *A* after growth. Regarding the mobilities, the bulk mobility of sample *A* showed a sharp decrease from 6400 to 1500 cm^2/Vs over two days. This drop is due to Se vacancies being produced along with contamination by other atmospheric impurities, both acting as scattering centers. In contrast, the surface mobility changed much less from 560 to 440 cm^2/Vs. These observations suggest that most of the changes observed in the samples grown on Si substrates are indeed from the bulk transport properties and in contrast, the surface states remain robust against air exposure. The robustness of the surface states in air is also supported by a recent optical measurement that observed strong surface transport on Bi_2Se_3 films left in air for several weeks ([186]Valdés Aguilar *et al.*, 2012).

Lastly, in Cu-Bi_2Se_3 thin films, the bulk channel is suppressed and the SS become decoupled when the film is $\gtrsim$ 20 QL and the Cu concentration is $x \approx 2.5 - 4\%$ (x = Cu/Bi × 100%)

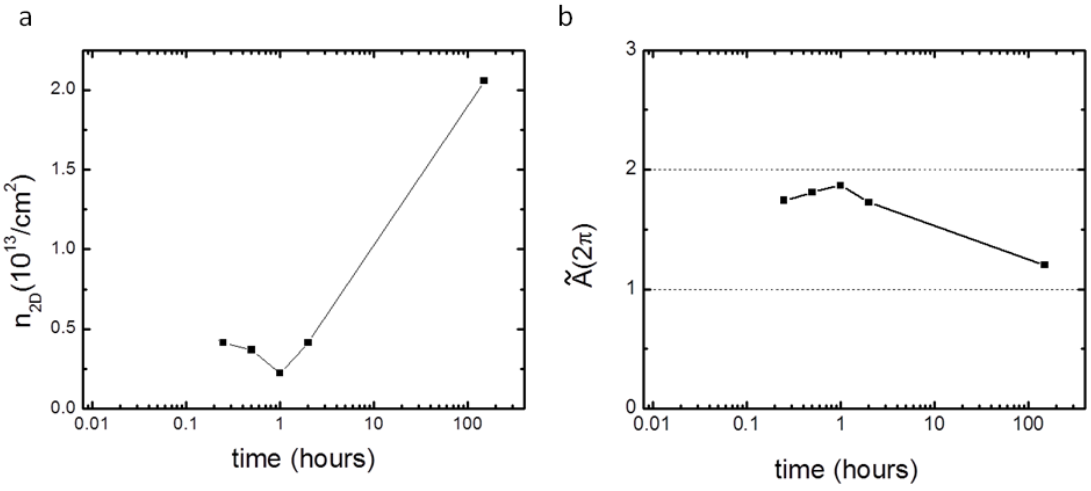


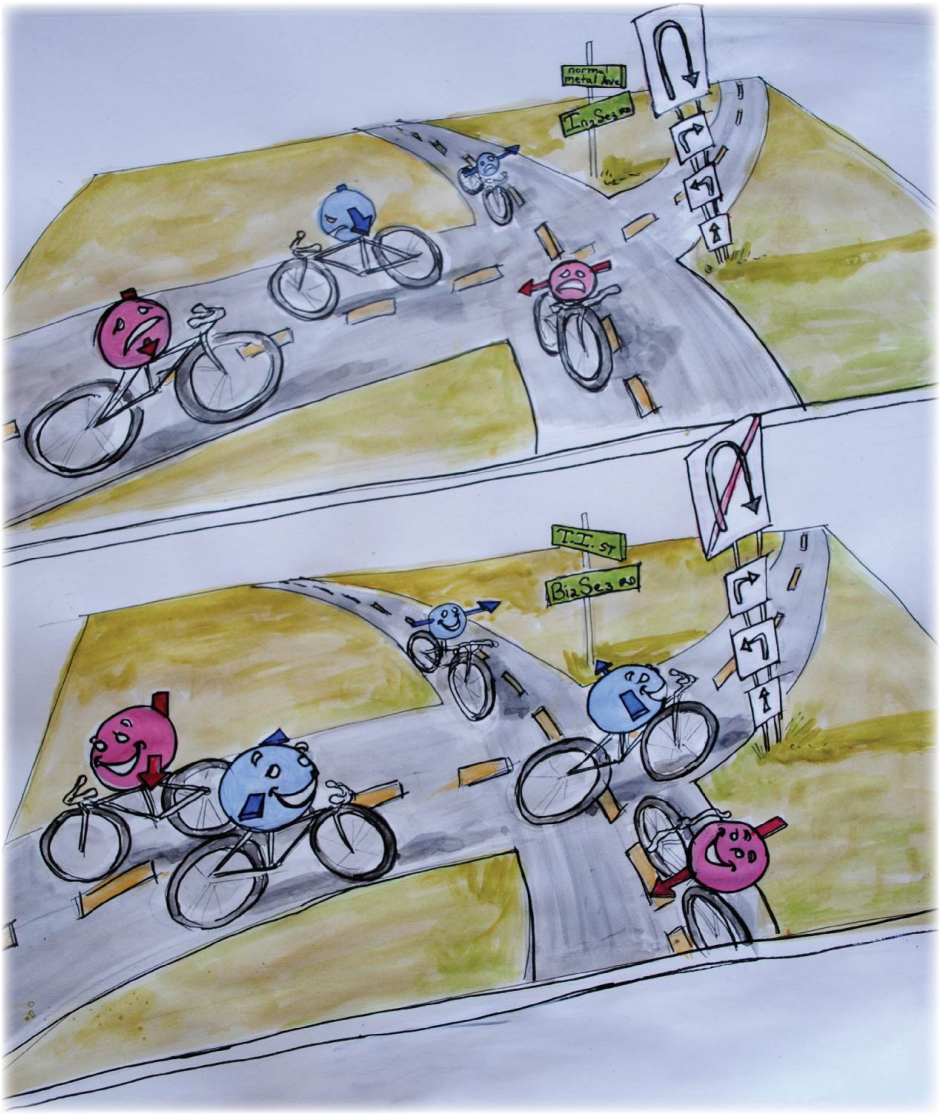
Figure 4-4 The response of Cu-Bi_2Se_3 thin films grown on Al_2O_3 to exposure to atmosphere. **a** The carrier density versus exposure time shows that the carrier density rises. **b** As the carrier density rises, the bulk channel becomes repopulated, which is reflected as the $\tilde{A}$ parameter in the weak anti-localization effect decreases from the decoupled ($\tilde{A}$ = 2), to coupled ($\tilde{A}$ = 1).

([32]Brahlek *et al.*, 2014, and Chapter 6). As shown in Figure 4-4, the carrier concentration remains in the optimum range of below 1×10^{13} /cm^2, for roughly 2 hours. After ~ 100 hours of exposure, the carrier concentration rises to ~ 2×10^{13} /cm^2, which, as mentioned before, is likely due to water diffusion. This high carrier density is sufficient to repopulate the bulk state, which is reflected in the effective channel number $\tilde{A}$, extracted from the weak anti-localization effect, decreasing from 2 (decoupled SS), down to 1 (SS coupled by a conduction bulk channel) ([32]Brahlek *et al.*, 2014). As previously stated, due to the ill-defined nature of atmospheric gas, the exact time that the bulk state remains depleted is unique for this particular measurement, but the overall trend is general in that at long times Cu-Bi_2Se_3 goes towards a bulk-metallic state.

To conclude, we investigated how the electronic properties of the topological insulator Bi_2Se_3 change under a variety of conditions. While the bulk transport properties were significantly affected by environmental factors such as vacuum, oxygen and moisture, the surface states remained stable. These observations resolve the previous discrepancy between transport and ARPES studies regarding the robustness of the surface states, and support the theoretical prediction of the surface state immunity to non-magnetic disorders. However, they raise another issue of the fragility of the bulk properties. Finding a way to stabilize the TI bulk properties seems to be the key to the future of TI applications.

Chapter 5.

Topological phase transition in $(Bi_{1-x}In_x)_2Se_3$ [5]



One challenge of this work is to understand the most basic electronic phenomena of topological insulators (TI). A fortuitous property of the TI Bi_2Se_3 is that it shares the same structure as the non-topological band-insulator In_2Se_3, which enables the solid mixture $(Bi_{1-x}In_x)_2Se_3$ to be stable

[5] A cartoon depicting the electrons in In_2Se_3, a non-topological insulator, suffering backscattering (top), while the electrons in Bi_2Se_3, a topological insulator, are free from such backscattering (bottom). Courtesy of Alexandrea Seiler-Brahlek.

for all x. With this mixed structure one can use x, in a sense, as a knob to tune the topological properties from a normal insulator through into the TI state. This ability to turn the topological properties *on* and *off* systematically has proven to be extremely useful, and we are only at the beginning of much wider studies to explore the fundamentals of how the topological phase emerges out of a non-topological normal insulator. Furthermore, as we will see in Chapter 6, there are many additional applications to other experimental structures in TIs that make use to the combination of Bi_2Se_3 with $(Bi_{1-x}In_x)_2Se_3$ as a solid mixture, or as a heterostructure.

In this chapter, we will focus on several measurements on isolated $(Bi_{1-x}In_x)_2Se_3$ films, with varying composition and film thickness. To probe the crossover from TI (Bi_2Se_3 x = 0%) to full band insulating non-TI (In_2Se_3 x = 100%) several techniques are employed to gain a thorough understanding of the physics of the transitions. We will first discuss the earliest work, which used direct current (DC) transport measurements combined with angle resolved photo-emission spectroscopy (ARPES) to provide details as to how the TI phase transitioned into a normal metal and then into a band insulating phase ([28]Brahlek *et al.*, 2012). We will then discuss a later study that employed terahertz spectroscopy to explore the regime of the TI phase transition in more depth ([202]Wu *et al.*, 2013). Lastly, we will provide some of the future prospects for experiments that are currently underway to explore the topological to non-topological insulator transition.

5.1 Topological-metal to band-insulator transition in $(Bi_{1-x}In_x)_2Se_3$ thin film: DC transport and angle resolved photo-emission spectroscopy[6]

This section will focus on transport and ARPES measurements on thin films of $(Bi_{1-x}In_x)_2Se_3$, where x varies from 0 to 1. Bi_2Se_3 (x = 0) is a proto-typical topological insulator. However, its bulk Fermi energy (E_F) always stays in the conduction band (CB), and thus it is better to call Bi_2Se_3 a topological metal (TM), as we will continue throughout this chapter. On the other hand,

[6] This section is based on work that is published in M. Brahlek *et al,* Physical Review Letters **109**, 186403 (2012).

In_2Se_3 (x = 1) is a topologically trivial band insulator (BI) with energy gap around 1.3 eV ([199]Watanabe *et al.*, 1989). Although metal-insulator transitions occur in many other systems (see [98]Imada *et al.*, 1998 for a thorough review), and a topological metal to non-topological metal transition has been observed in $TlBi(Se_{1-x}S_x)_2$ ([170]Sato *et al.*, 2011;[209]Xu *et al.*, 2011), until now no material system has exhibited tunability from a topological metal state to a truly insulating state. Here, we show that $(Bi_{1-x}In_x)_2Se_3$ is the first system to show such a broad tunability.

For a TM to transform into a BI, there have to be at least two quantum critical points. As Bi atoms are replaced by lighter In atoms and the SOC weakens, a quantum critical point will

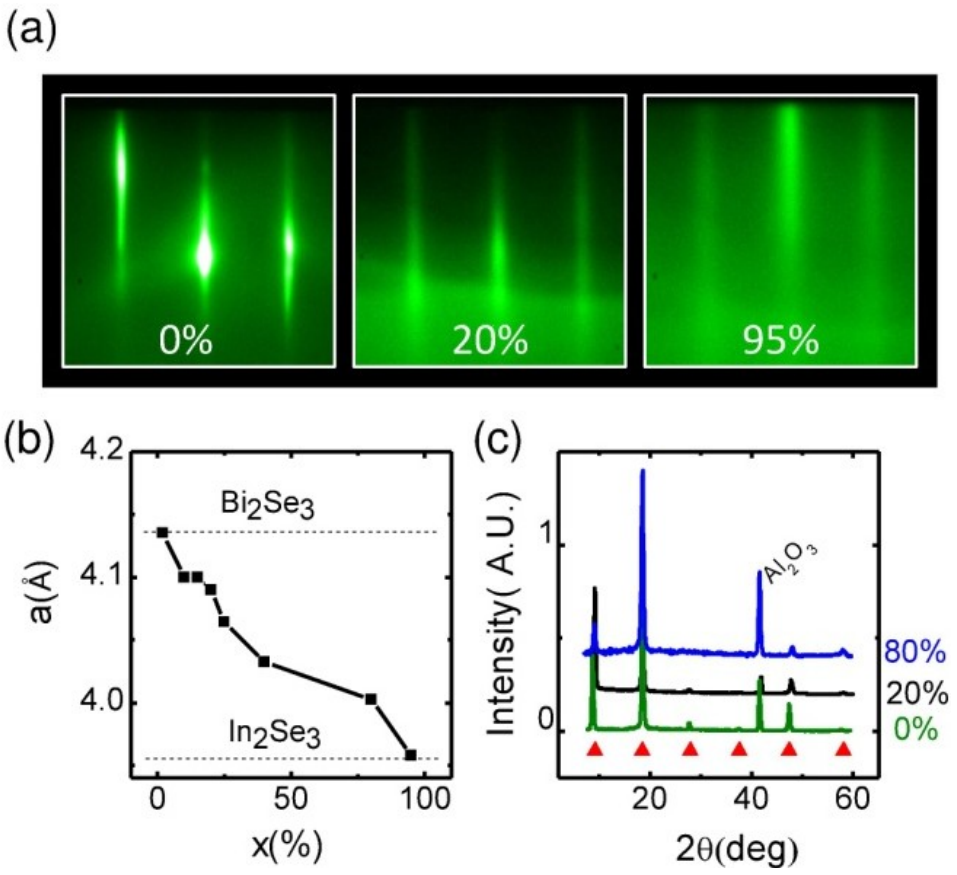


Figure 5-1 Structural data for $(Bi_{1-x}In_x)_2Se_3$ thin films for various values of x. **a** Typical RHEED images for x = 0, 20 and 95%. The 2nd-phase free vertical streaks indicate atomically flat films for the full range of x. **b** Lattice constant as a function of x obtained from the horizontal separation of the maxima in a. **c** X-ray diffraction data for x = 0, 10, 80%. The (003n) peaks, marked by triangles for n = 1 – 6, provide a quintuple layer thickness of ~9.5 Å, almost independent of x. (Adapted from [28]Brahlek *et al.*, 2012)

emerge where the energy bands invert, and the system will go from a topologically non-trivial state to a topologically trivial state. At another quantum critical point, E_F will drop below the CB minimum and the system will become an insulator. Depending on which one occurs first, a TM could become a non-topological normal metal (NM) and then a BI, or it could become a TI first and then a BI. We find that the former is the case in this material. In addition to these two quantum phase transitions, we also find that the material goes through a variable range hopping insulator (VI) phase before becoming a BI.

As detailed in Chapter 2, we prepared thin films of $(Bi_{1-x}In_x)_2Se_3$ by molecular beam epitaxy (MBE). The Bi and In source fluxes were calibrated in situ with a quartz crystal micro-balance, and ex situ with Rutherford backscattering spectroscopy (RBS) measurements; together these provide accuracy to within about $\Delta x \approx \pm 1\%$ (see 2.8.6 for more details). The thickness of the films was fixed at 60 QL (QL, where 1 QL ≈ 1 nm), and were grown on 10 × 10 mm^2 Al_2O_3 (0001) substrates ([16]Bansal *et al.*, 2011). Transport measurements were carried out down to 1.5 K

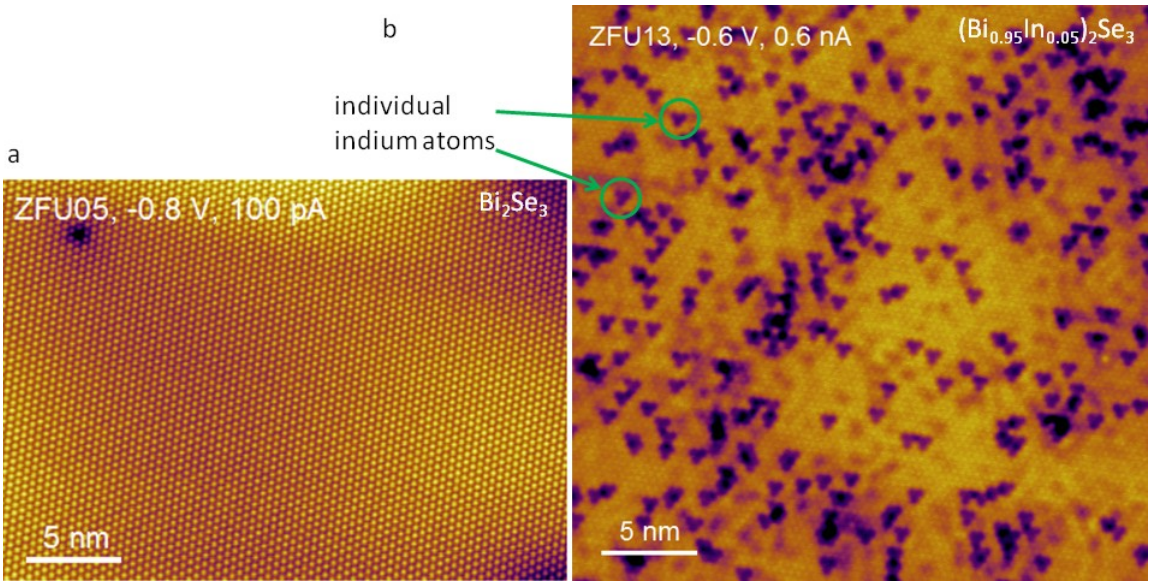


Figure 5-2 Scanning tunnelling microscopy images taken from Bi_2Se_3 and $(Bi_{1-x}In_x)_2Se_3$ with $x \approx 5\%$. These images show that in **a,** Bi_2Se_3 shows no defects within the scan area, while in **b,** individual (no In_2Se_3 clusters) In atoms are visible as triangular shaped disturbances to the local electron density. (Image courtesy of Jixia Dai and Weida Wu)

in a magnetic field up to 9 T using the standard Van der Pauw geometry (see section 3.3). The ARPES measurements were performed at the Advance Light Source in the Lawrence Berkeley National Laboratory. Both ARPES and transport samples were grown in identical conditions except the ARPES samples were capped in situ by ~100 nm-thick Se. The Se capping was removed in the ARPES chambers by heating the samples up to ~250 °C while keeping the chamber pressure below 1×10^{-9} Torr. ARPES measurements were performed at 20 K with 29-64 eV photon energy, with the chamber pressure less than 5×10^{-11} Torr.

Figure 5-1 shows how the structure of $(Bi_{1-x}In_x)_2Se_3$ changes with x. The reflection high energy electron diffraction (RHEED) images in Figure 5-1a, and the X-ray diffraction data in Figure 5-1c show that all our samples are of a single phase within the instrument limit. This is confirmed by the scanning tunneling microscopy image in Figure 5-2, which shows that In atoms occur individually, and not in local In_2Se_3 clusters. Lastly, in Figure 5-1b the in-plane lattice constants estimated from the RHEED images gradually decreased from 4.14 Å for Bi_2Se_3 to 3.95 Å for In_2Se_3.

$R(T)$ curves in Figure 5-3a show how the system changes from a metallic state to an insulating state as x increases. Samples with $x < 15\%$ show metallic temperature dependence, whereas for $x \gtrsim 15\%$, the $R(T)$ curves turn upward as temperature decreases. To see this, we plotted $R(1.5\ \mathrm{K})/R_{min}$ vs. x in Figure 5-3b; for $x > 18\%$, $R_{min} = R(270\mathrm{K})$ because $R(T)$ monotonically decreased with increasing temperature below 270 K, the highest temperature used in this experiment. Figure 5-3b shows that $R(1.5\ \mathrm{K})/R_{min}$ remains unity for $x \lesssim 15\%$ and it sharply increases for larger x, which indicates that a metal-to-insulator transition occurs at $x = 15\%$: we define this composition as x_I. Although this analysis suggests that the system becomes an insulator for $x > 15\%$, ARPES in Figure 5-3f shows that the Fermi level remains in the conduction band up to $x > 20\%$.

Such an insulating state is best described by the variable range hopping (VRH) mechanism ([139]Mott and Davis, 1979). With disorder, even if a band is partially filled, electrons

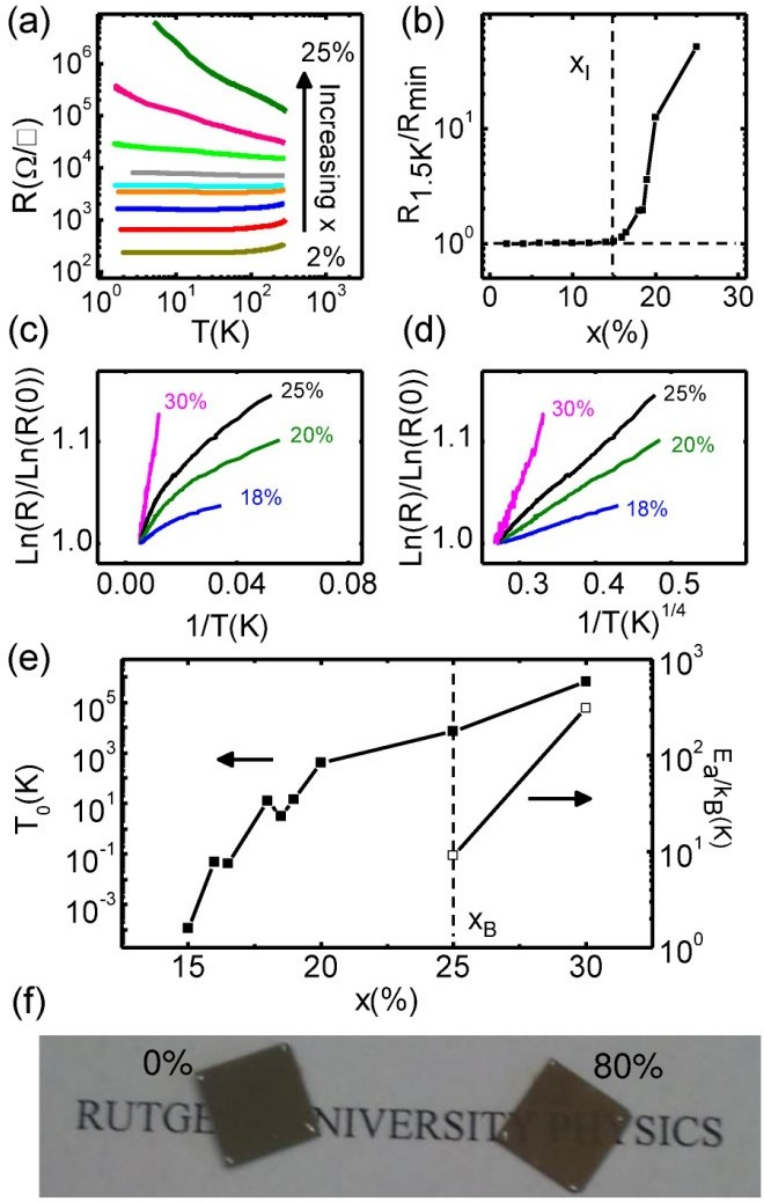


Figure 5-3 Resistance versus temperature for $(Bi_{1-x}In_x)_2Se_3$ thin films from the TI regime through to the band insulating regime. a Resistance vs. temperature for 60 QL films with various values of x. From bottom up x = 2.0, 8.0, 10, 12, 15, 16, 18, 20, and 25% **b** R(1.5 K)/R_{min} vs. x, exhibiting a metal-insulator transition at $x \approx 15\%$, defined as x_I. **c** and **d** show how resistance scales with temperature. Linearity in **c** and **d** implies, respectively, activation and 3d-VRH temperature dependence. **e** T_0 for the 3d-VRH (left axis) and E_a for the activation (right axis) formula as a function of x. The vertical dotted line is placed at x = 25% as an approximate transition point (x_B) between VI and BI. **f** A photograph of 60 QL films with x = 0 and 80%. (Adapted from [28]Brahlek *et al.*, 2012)

can be localized as temperature decreases. According to the VRH model, $R(T)$ is described by $R(T) = R_0 exp[(T_0/T)^{1/(D+1)}]$ where T_0 is a characteristic temperature scale for hopping, D is the

effective dimensionality of the conducting channel, and R_0 is the characteristic resistance of the sample. For $x > 15\%$, the mean free path (Figure 5-4b) is less than ~1 nm, much smaller than the film thickness (60 nm). In this case, the resistance is dominated by the bulk scattering effect with negligible surface contribution, and the film behaves effectively as a 3D system. Figure 5-3d indeed shows that the temperature dependence of the resistance follows this 3D-VRH formula quite nicely for $x > 15\%$.

From the fitting parameter T_0, plotted in Figure 5-4e, we can further check the validity of the 3D-VRH model as a function of x. Because $T_0 \propto 1/N(E_F)$, where $N(E_F)$ is the density of states at the Fermi level ([139]Mott and Davis, 1979), as the Fermi level approaches the bottom of the CB and $N(E_F)$ reduces to zero with increasing x, T_0 diverges and eventually the VRH mechanism breaks down. Once the Fermi level drops below the CB minimum, the system becomes a BI and the temperature dependence of the resistance should switch to the activation formula, $R(T) = R_0 exp(E_a/k_BT)$, where k_B is Boltzmann's constant, and E_a is the energy difference between the CB minimum and E_F. Considering that the band gap of Bi_2Se_3 (In_2Se_3) is ~0.3 eV (~1.3 eV), any energy scale governing their transport properties cannot be larger than ~1 eV (or ~10,000 K). The 3D-VRH fitting provides T_0 =500, 5000, and 6.5×10^5 K for x = 20, 25 and 30%. This implies that the VRH mechanism is valid for x = 20%, but it is on the verge of breaking down for x = 25% and clearly invalid for x = 30%. On the other hand, E_a/k_B obtained from the activation fitting for x = 30% is 300 K (≈ 30 meV), and this is physically valid considering the band gap shown in Figure 5-4f. Fitting the low temperature data of x = 25% (Figure 5-3c) to the activation formula yields $E_a/k_B \approx 10$ K (≈ 1 meV). Considering that this activation energy is larger than the minimum measured temperature (1.5 K), the activation fitting does not violate self-consistency for the 25% data. In other words, x = 25% can barely belong to either side of the VI or the BI within our measurement limit. Summarizing these transport and ARPES analyses, we conclude that x = 20% is a VI, x = 25% is on the boundary between a VI and a BI and x = 30% is a BI.

For x > 30% the samples are immeasurably insulating even at room temperature. The metal-insulator transition in these samples is also observable in their optical transparencies: in Figure 5-3f, 60 QL Bi_2Se_3 is completely opaque and mirror-like, whereas $(Bi_{0.2}In_{0.8})_2Se_3$ of the same thickness is semi-transparent even if the band gap of In_2Se_3 is smaller than the energy of visible light.

Figure 5-4a presents the results obtained from the Hall effect measurements. Below x = 6%, the Hall resistance, R_{xy}, versus B was nonlinear, which indicates the presence of multiple channels contributing to the Hall effect ([15]Bansal *et al.*, 2012). This non-linearity disappears above 6% as shown in the inset of Figure 5-4a, where dR^N_{xy}/dB ($R^N_{xy} = R_{xy}(B)/R(9\text{ T})$) is used as a measure of the non-linearity for x = 2, 6, and 8%. The disappearance of the non-linearity occurs near a point where the surface band disappears in the ARPES spectra (Figure 5-4f), and is likely a signature of the topological phase transition. Extracting individual carrier densities from non-linear Hall effect data requires a multi-carrier fitting process. To avoid such a complication, we took the slope of the R_{xy} vs. B curves at high field, which is inversely proportional to the sum of the sheet carrier densities, to obtain the total sheet carrier density, n_{2d}, (right axis) and the mobility, μ, (left axis) at 1.5 K. In Figure 5-4a, the carrier density monotonically decreased with increasing x, and became too small to be properly measured for x > 20%; thus n_{2d} for 25 and 30% were simply taken to be zero. One notable feature in Figure 5-4a is that the mobility drops fast right above x = 6% as the non-linearity disappears in $R_{xy}(B)$. This accelerated mobility drop is possibly due to enhanced backscattering as the topological surface bands disappear (see 5.2).

Several key factors must be considered to understand this topological phase transition. In the case of $TlBi(Se_{1-x}S_x)_2$ ([170]Sato *et al.*, 2011;[209]Xu *et al.*, 2011), the replacement of Se by S weakened the SOC and contracted the lattice constant, which resulted in the band inversion and the topological transition at $x \approx 50\%$, while the bonding orbital (p) characters remain unchanged with x. In contrast, in the $(Bi_{1-x}In_x)_2Se_3$ system has, in addition to the SOC and the lattice constant effects, the bonding orbitals change with x: Bi ($6p^3$) has only p orbitals involved in the bonding,

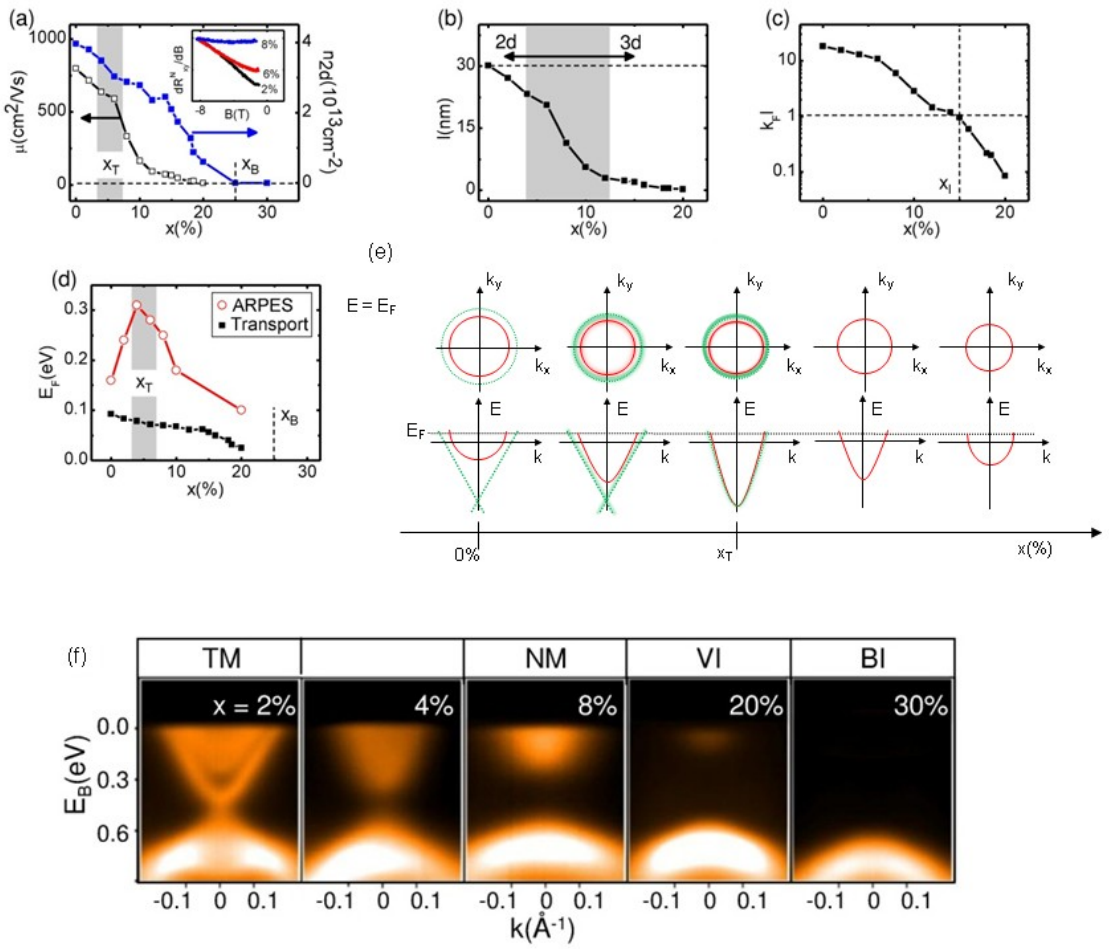


Figure 5-4 Magnetotransport data and ARPES data for $(Bi_{1-x}In_x)_2Se_3$ thin films from the TI regime through to the band insulating regime. a Mobility (open squares/left axis) and carrier density (solid squares/right axis) vs. x at 1.5 K. Inset: dR^N_{xy}/dB vs. B showing that the non-linearity disappears between x = 6 and 8%. **b** Mean free path, l, vs. x. As the mean free path becomes much smaller than half of the film thickness (dotted line), the effective dimensionality transforms to 3d. **c** k_Fl vs. x. This value drops below one above $x \approx 15\%$. **d** The Fermi energy, E_F, relative to the CB minimum, vs. x, estimated from ARPES and transport. **e** A schematic of how the band structure changes near x_T. Top panel: Cross sections of the Fermi surface at E_F. Bottom panel: E vs. k band dispersion. The red solid (green dotted) lines represent the bulk conduction (surface) bands: for simplicity, the valence bands are not shown here. As x increases toward x_T, the bulk gap closes, and once x goes beyond x_T, the gap reopens and grows with increasing x; during this process, the conduction and valence bands invert and the surface band disappears. x_T may not be a single point if the gap-opening at the Dirac point and disappearance of the surface band do not occur simultaneously. **f** ARPES spectra for x = 2, 4, 8, 20, and 30%. TM, NM, VI, and BI represent topological metal, non-topological metal, variable range hopping insulator, and band insulator, respectively. (Adapted from [28]Brahlek *et al.*, 2012)

whereas In ($5s^25p^1$) has both *s* and *p* orbitals. This orbital mixing may have a significant effect on the topological transition of $(Bi_{1-x}In_x)_2Se_3$, considering that the transition occurs at a much smaller value of x than expected.

In $(Bi_{1-x}In_x)_2Se_3$ this observation of a band inversion transition at an anomalously small value of x inspired a very thorough study by first principles calculation. J. Liu and D. Vanderbilt ([125]Liu and Vanderbilt, 2013) extensive studied how the TI to non-TI transition in $(Bi_{1-x}In_x)_2Se_3$ and $(Bi_{1-x}Sb_x)_2Se_3$ (the latter is only a hypothetical structure because there is a structural transition around $x \approx 25\%$, which drives the system into a non-topological state ([216]Zhang *et al.*, 2009;[118]Lee *et al.*, 2013;[125]Liu and Vanderbilt, 2013)), and confirmed that there are distinct differences between the two systems. $(Bi_{1-x}Sb_x)_2Se_3$, like $TlBi(Se_{1-x}S_x)_2$, only has p orbitals near the Fermi energy, and, therefore, is found to follow a linear gap closure picture and has a topological critical point near ~80%. Similarly, for $(Bi_{1-x}In_x)_2Se_3$ without the In 5*s* orbital, band inversion takes place at around $x \approx 50\%$ following a similar linear dependence. In contrast, as mentioned above, in $(Bi_{1-x}In_x)_2Se_3$ the In 5*s* orbitals plays a more prominent role in the band inversion than SOC; when the 5*s* orbitals were restored the transition was pushed to a lower value < 12%, which closely matches the experimental values shown here.

As graphically illustrated in Figure 5-5, the effect of finite sample thickness has to be considered. In an ideal TI, the penetration depth of the SS into the bulk is given, to first-order, by $\lambda_{ss} = \hbar v_F / E_g$, where v_F is the Fermi velocity of the Dirac band and E_g is the bulk band gap ([124]Linder *et al.*, 2009). As x approaches the topological transition point (x_T), E_g becomes smaller and λ_{ss} increases. If λ_{ss} approaches the film thickness, the wavefunctions of the top and bottom surface states should start overlapping with each other, leading to a gap opening at the Dirac point even before the bulk band gap closes and surface band completely disappears ([219]Zhang *et al.*, 2010). This expectation is indeed consistent with our observation: a gap opened in the ARPES spectrum at $x \approx 3\%$ (Figure 5-4f), whereas the signature of the surface band at the Fermi level survived up to $x \approx 7\%$ in the transport properties (Figure 5-4a). In this case, x_T is not a single point, but rather should be defined broadly over ~3-7%, whose sharpness may also depend

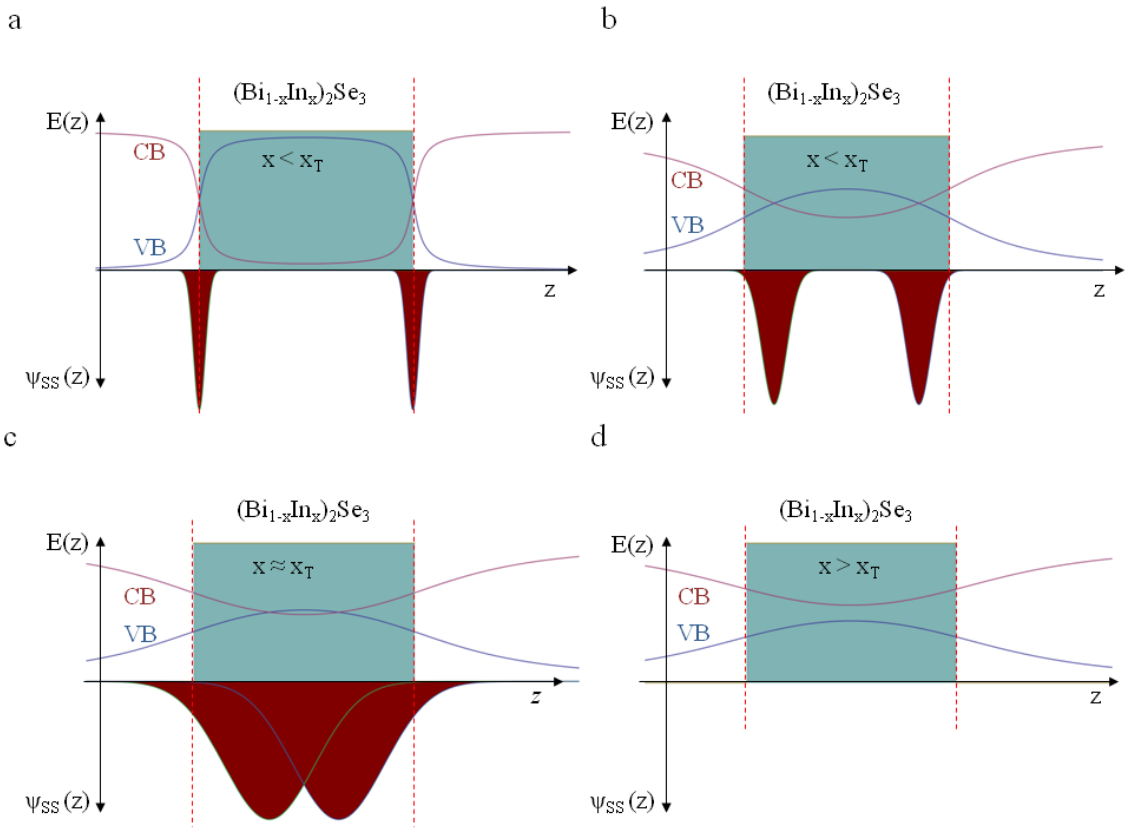


Figure 5-5 A schematic showing how the conduction band (CB) and valence band (VB) evolve spatially across the topological phase transition. a-d The CB and the VB cross near the interface (dashed red lines) between vacuum and $(Bi_{1-x}In_x)_2Se_3$ (upper part of each panel) and the surface state wavefunction (lower part of each panel) expand as a function of x near the quantum critical point x_T where the horizontal axes being the thickness.

on the sample thickness. This scenario is strongly supported by an experiment ([202]Wu *et al.*, 2013) that were performed after the publication of this work, and is detailed in 5.2.

The mobility (μ) and carrier density (n_{2d}) obtained from the Hall effect allows calculation of the mean free path, $l = (\hbar\mu/e)(3\pi^2 n_{2d}/t)^{1/3}$, where t is the film thickness; l decreases monotonically with x as shown in Figure 5-4b. According to the Ioffe-Regel criterion ([1]A. F. Ioffe, 1960), a material remains metallic only if $k_F l \geq 1$, where $k_F = (3\pi^2 n_{2d}/t)^{1/3}$ is the 3D Fermi wave vector. Figure 5-4b shows that $k_F l$ continuously diminishes as x increases, and crosses the Ioffe-Regal criterion of $k_F l = 1$ exactly at the composition ($x = 15\%$) where the metal-insulator transition occurs in Figure 5-3b.

Figure 5-4d shows two separately estimated Fermi levels (E_F) relative to the bottom of the CB from transport and ARPES measurements. The transport E_F is calculated from n_{2d} assuming a 3D Fermi surface, using $E_F = \hbar^2(3\pi^2 n_{2d}/t)^{2/3}/(2m^*)$ and $m^* \approx 0.15m_e$ ([3]Analytis *et al.*, 2010), and the ARPES E_F is simply read off from the spectra. In Figure 5-4d, the first notable feature is the presence of a peak in the ARPES data around x_T and its absence in the transport data. This can be understood by considering how the band structure changes near the critical point. Throughout the topological transition near x_T, as illustrated in Figure 5-4e, a dramatic change occurs around the Dirac point and the CB minimum, while at E_F the change is only marginal with the overall size of the Fermi circle getting gradually reduced and the surface band smearing out. This explains why the transport E_F, which simply represents the size of the Fermi circle, does not show the kind of peak observed in the ARPES data. Another observation is that the ARPES E_F is generally higher than the transport E_F. This is mainly due to the downward band bending on Bi_2Se_3 surfaces ([3]Analytis *et al.*, 2010;[22]Bianchi *et al.*, 2010). Therefore, the ARPES E_F, which measures the surface Fermi level, should always be higher than the transport E_F, which is an average value of the surface and bulk Fermi levels.

Finally, Figure 5-4f shows how the ARPES spectrum evolves with increasing x. For $x <$ 3%, the surface band is clearly visible, implying that the system is a topological metal. Although a gap is formed at the Dirac point in the x = 4% spectrum, the surface bands seem to disappear completely only for $x >$ 7% according to the transport analysis. The system becomes a 3D-VRH insulator for $x >$ 15%, and a full band insulator for $x \gtrsim$ 25%. $(Bi_{1-x}In_x)_2Se_3$ is unique in that it is so far the only system that provides not only topological but also electronic tunabilities with a structural match to Bi_2Se_3 ([198]Wang *et al.*, 2011). Therefore, it offers an ideal venue to fabricate tunable topological devices with seamless epitaxial tunnel barriers and gate dielectrics.

5.2 A sudden collapse in the transport lifetime across the topological phase transition in $(Bi_{1-x}In_x)_2Se_3$ as measured by time domain terahertz spectroscopy[7]

Time-domain terahertz spectroscopy (TDTS) is measurement technique that is performed in transmission geometry, and the complex transmission function (real plus imaginary components) is directly measured. This can be then used to measure the real and complex conductivity of a material (see [186]Valdés Aguilar *et al.*, 2012;[187]Valdés Aguilar *et al.*, 2013;[202]Wu *et al.*, 2013 for more details on TDTS measurements on TI thin films). Therefore, in regards to these measurements, the Drude model developed in Chapter 3 can be extended (see [81]Grosso and Pastori Parravicini, 2000 for example) to account for the finite frequency response of the conductivity by

$$\sigma = \frac{\sigma_0}{1-i\omega\tau},$$

where $\sigma_0 = e^2 n\tau/m^*$ is the Drude conductivity, n is the carrier density, ω is the frequency, and τ is the scattering time ($\tau^{-1} = \Gamma$ is the scattering rate). The conductivity can be separated into the real component

$$Re(\sigma) = \frac{\sigma_0}{1+(\omega\tau)^2},$$

and the complex component

$$Im(\sigma) = \frac{\sigma_0(\omega\tau)}{1+(\omega\tau)^2}.$$

In this way, one can obtain data similar to DC transport, but in a complementary manner. Terahertz spectroscopy sits in the regime between electronics ($\omega_{THz} \sim 10^{12}$ Hz > $\omega_{Eletronics}$) and photonics ($\omega_{THz} < \omega_{Photonics}$), and is a frequency scale that matches much of the novel low energy physics; for example, in TIs the scattering rate, Γ, is in the THz regime.

[7] This section is based on work that is published in L. Wu, M. Brahlek *et al* Nature Physics **9**, 410 (2013).

Figure 5-6 shows the real conductance versus frequency for Bi_2Se_3 as a function of thickness (a and b) and for $(Bi_{1-x}In_x)_2Se_3$ fixed at 64 QL with increasing In concentration (c). Two features are prominent: the peak at zero frequency is the Drude response (marked by a green arrow), and the second peak, near 2 THz, is a phonon (marked by a red arrow). The conductance can be modeled as a low frequency Drude term describing the response of the free carriers, and a Drude-Lorentz terms which accounts for the phonon. As shown in Figure 5-6d, there is good agreement between the model and the data; from this, the transport parameters can be extracted,

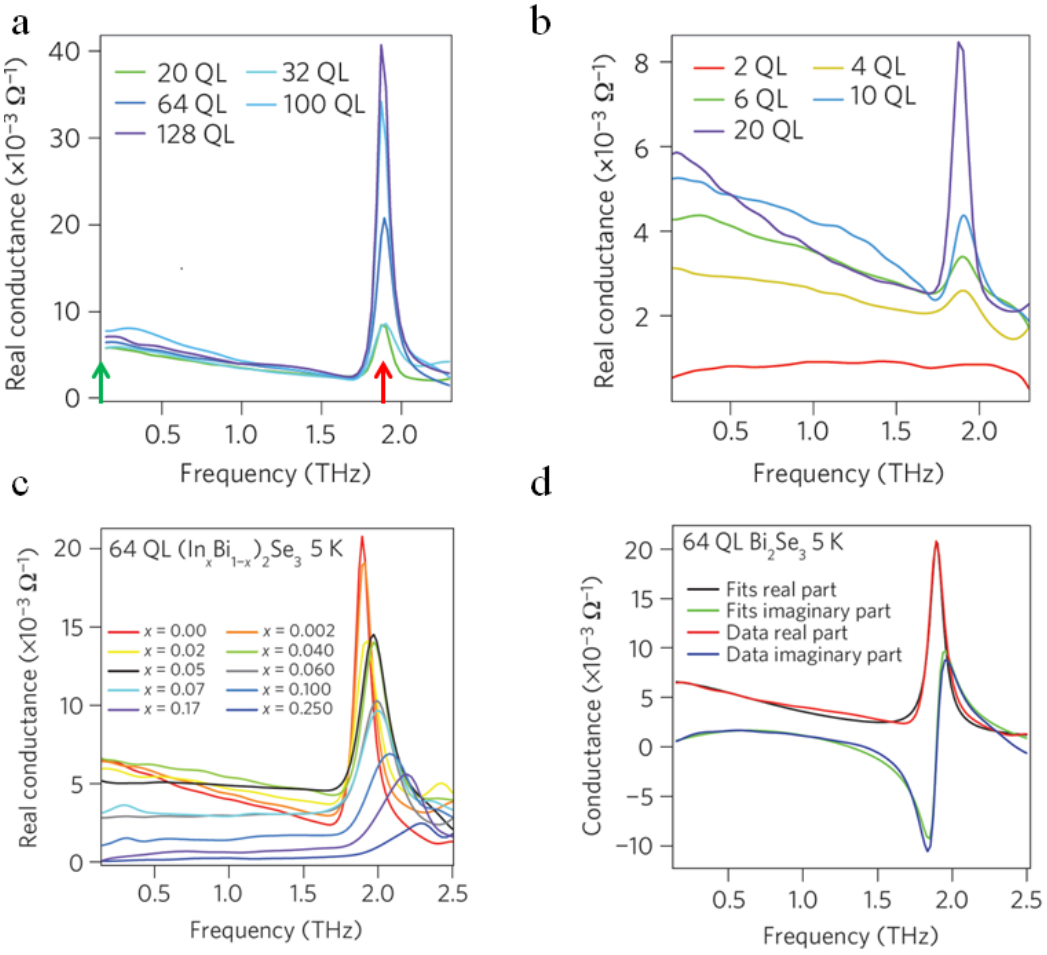


Figure 5-6 The real conductance versus frequency for Bi_2Se_3 and $(Bi_{1-x}In_x)_2Se_3$. For various thickness (**a-b**), and $(Bi_{1-x}In_x)_2Se_3$ with thickness fixed at 64 QL for various values of *x* (**c**). The red arrow shown in **a** marks the phonon near 2 THz, and the green arrow marks the zero frequency peak of the Drude term. **d** The comparison between the model and data for the real and imaginary parts of the conductance. (Adapted from [202]Wu *et al.*, 2013)

which have been shown be a highly sensitive probe of the TSS ([186]Valdés Aguilar *et al.*, 2012, [202]Wu *et al.*, 2013 and [187]Valdés Aguilar *et al.*, 2013)

Of particular interest in this study is the dependence of the scattering rate $\Gamma = \tau^{-1}$ across the critical composition x_T. As shown in Figure 5-4, the mobility (recall, $\mu = e\tau/m^* = e/(m^*\Gamma)$) monotonically decreases with increasing In concentration. For $x \lesssim 5$ - 6%, the mobility decreases nearly linearly with x, and right around $x \approx 5$ - 6%, the rate the mobility decreases with increasing x sharply increases, resulting in a telling kink in the μ vs x curve. With knowledge of x_T, Brahlek *et al* posited that this kink is consistent with the TI to non-TI phase transition ([28]Brahlek *et al.*, 2012). Such an affect is likely since the surface electrons are topologically protected from 180° backscattering, and, therefore, as x_T is crossed, one should expect to see an increase (decrease) in scattering rate (mobility) as the 180° backscattering is regained.

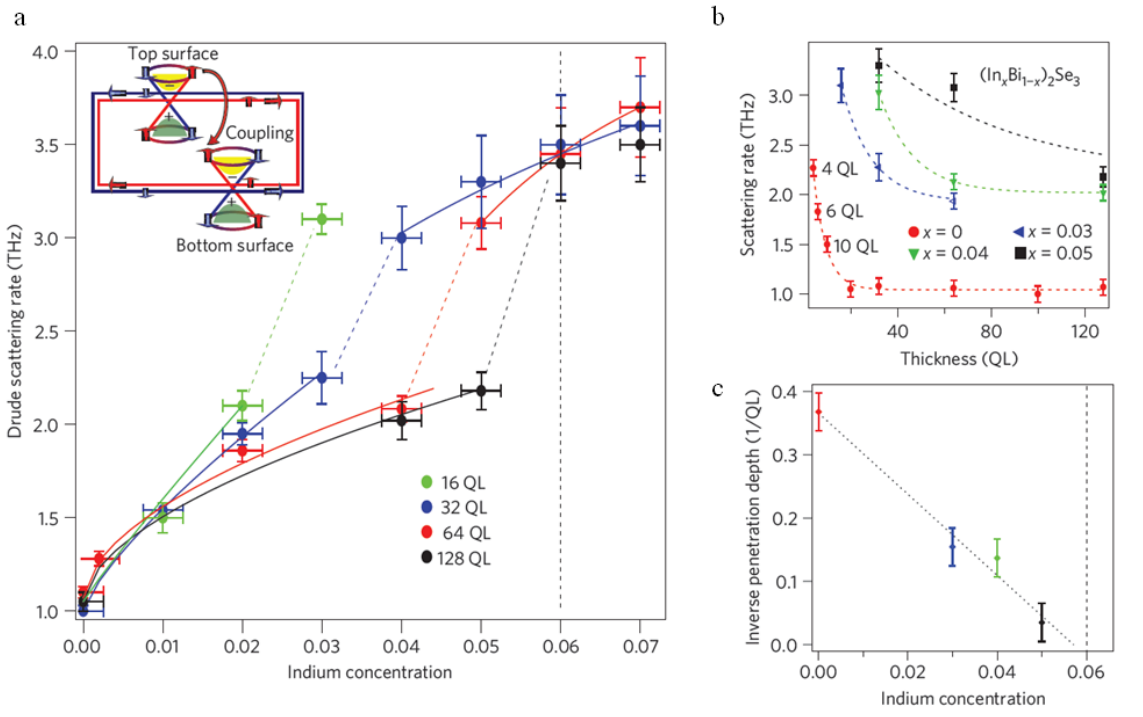


Figure 5-7 The scattering rate for $(Bi_{1-x}In_x)_2Se_3$ thin films showing a thickness-dependant jump across the TI phase transition. a The scattering rate versus In concentration plotted for various thicknesses. **b** The scattering rate versus thickness plotted for various concentrations. **c** The inverse penetration depth obtained by fitting the data in **b** to an exponential decay model (see dashed lines), and extrapolating this data to zero yields a bulk critical concentration of ~6%. (Adapted from [202]Wu *et al.*, 2013)

[202]Wu *et al.*, 2013 sought to systematically study this enhancement in backscattering by using the finite thickness effect ([202]Wu *et al.*, 2013). As mentioned previously the penetration depth of the TSS into the bulk is given by $\lambda_{ss} = \hbar v_F/E_g$, and near x_T, $E_g \rightarrow 0$, and therefore λ_{SS} should eventually increase to be comparable to the film thickness. When this happens, the wavefuctions on opposite sides of the film begin to overlap, and due to their opposite spin momentum locking, backscattering will be allowed. When this happens, there should be a measureable increase in the scattering rate, which should be a function of both the thickness and the concentration of In. As shown in Figure 5-7, with fixed film thickness, and increasing In concentration, the scattering rate increases with x, and then jumps near a critical value; then, with decreasing thickness, this jump in scattering rate is systematically pushed to smaller values of x.

This behavior can be quantified by plotting the scattering rate versus thickness for fixed In concentration (Figure 5-7b). The decay with increasing thickness can be fit well to the exponential decay model $\Gamma = 1/(A + Be^{-t/(2\gamma_{SS}(x))})$, where $\gamma_{SS}(x)$ is the concentration dependant penetration depth. Figure 5-7c shows that $\gamma_{SS}(x)$ decreases linearly with increasing x, and vanishes very close to 6%, which represents the bulk limit where the TI to non-TI crossover resides. This behavior is consistent with the critical thicknesses measured by ARPES in Bi_2Se_3 films ([219]Zhang *et al.*, 2010), and the mass acquisition in bulk crystals of $TlBi(Se_{1-x}S_x)_2$ ([170]Sato *et al.*, 2011), which was attributed to the SS penetration on the macroscale of the bulk crystals used.

To conclude, Figure 5-8 shows a phase diagram summarizing how a topological material decays into a non-topological state. This map shows four distinct regimes: The upper left hand portion is the 3D TI phase, in which there exists well defined TSS on opposite surfaces of the film that are independent and decoupled due to the large spatial separation. In the infinite thickness limit, as the indium concentration increases (near top moving from left to right) to a critical value the bands invert, and the phases changes from a topological to a non-topological phase; such a transition happens near a single critical value of x with finite thickness effect

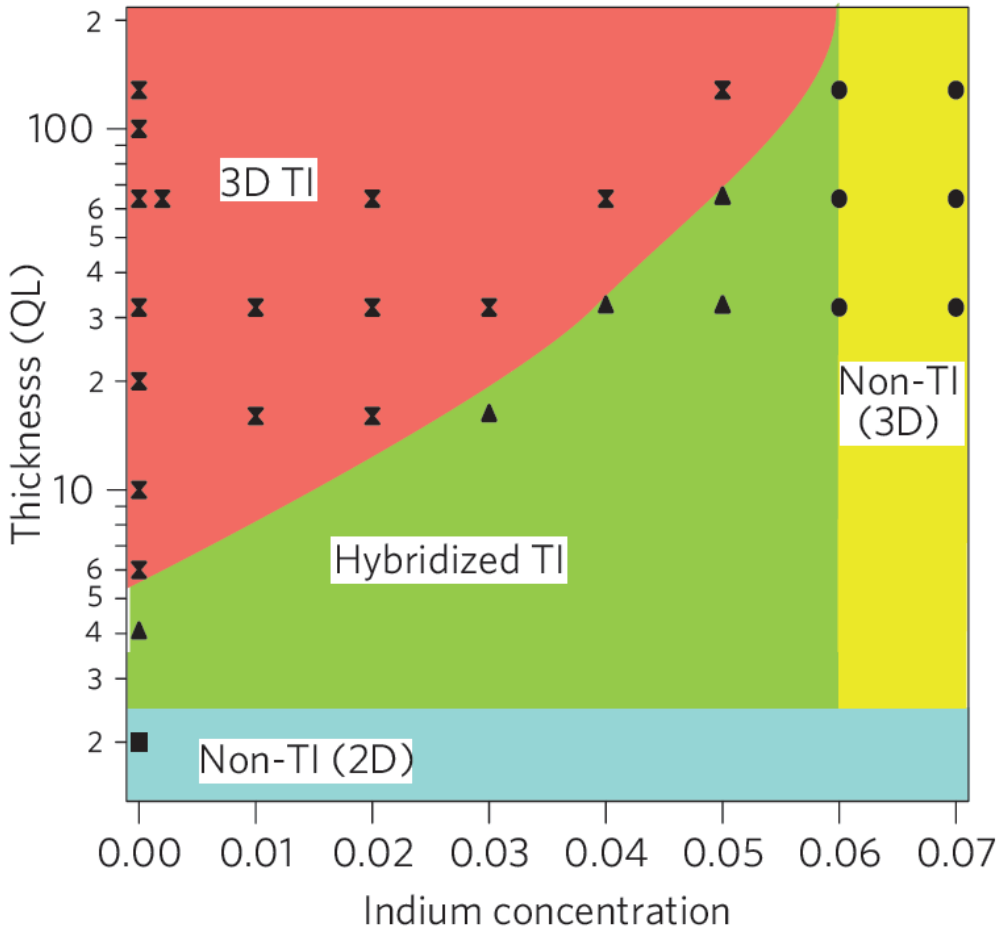


Figure 5-8 The phase diagram summarizes the behavior of the topological properties as a function of thickness and concentration. (Adapted from [202]Wu *et al.*, 2013)

playing a minor role. As the thickness is decreased (moving vertically downward on the right side in Figure 5-8), this transition from topological to non-topological is smeared out as the finite thickness effect starts to take a more prominent role. In this regime the TSS begin to expand into the bulk, and spin dependant coupling is allowed from opposite surfaces; this is denoted as hybridized TI regime. Eventually, with decreasing thickness, the topological phase eventually fully decays into a pseudo 2D material regardless of indium concentration.

5.3 Conclusion and future prospects

The two studies given in this chapter are only the beginning of many studies underway that are aimed to understand how materials attain their topological properties. Further, such a system is a

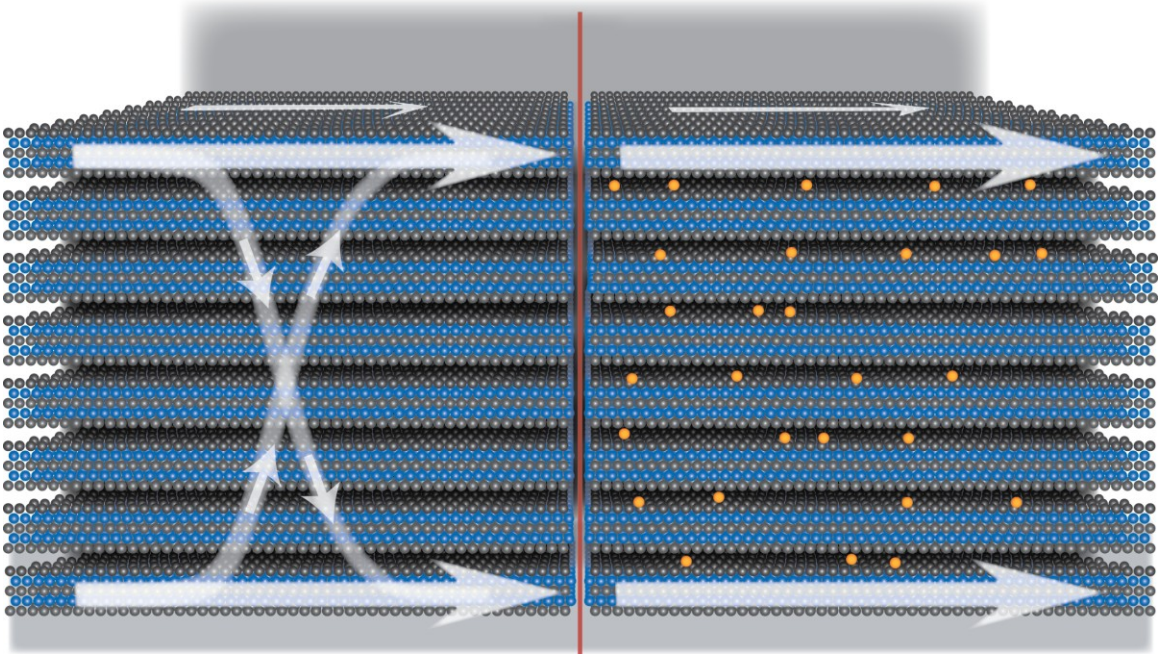
great test bed for determining properties that are associated with the topological nature because in this system one can controllably tune the topology from non-trivial to trivial. Lastly, the combination of Bi_2Se_3 and the tunable material $(Bi_{1-x}In_x)_2Se_3$ have already proven to be of great use as a probe of the emergence the TSS, which is one of the main topics in Chapter 6.

At this point there are several promising directions directly related to studying the TI to non-TI phase in Bi_2Se_3 thin films by DC transport. As discussed in Chapter 2, the Fermi energy is pinned into the conduction band mainly due to the large number of interfacial defects at the Bi_2Se_3-Al_2O_3 interface. This is also true in the $(Bi_{1-x}In_x)_2Se_3$ system, and, as discussed here, this high E_F may obscure much of the physics that happens near the Dirac point. One future direction is to study $(Bi_{1-x}In_x)_2Se_3$ films that are grown on the In_2Se_3 buffer layer. By alleviating the initial interfacial defects and lowering E_F, the physics of the TI phase transition may be further illuminated.

Another direction that offers a different prospective on the TI to non-TI phase transition is amorphous Bi_2Se_3. It is found that if Bi_2Se_3 is deposited near room temperature, then the surface mobility of the Bi and Se atoms is so low that no crystal structure can form, and the film accumulates in an amorphous structure. Electrically this material is found to be completely insulating and therefore non-topological since to be topological, there must be metallic states at the interface. The interesting feature comes when annealing to higher temperature: the amorphous film crystallizes to the typical layered Bi_2Se_3 structure, and therefore becomes topological. This disorder-insulator to topological-insulator is a promising system to probe the susceptibility of the TI phase to increasing disorder. Such a system invokes very fundamental questions about the TI phase; it is often stated that TI phases are immune to disorder, i.e. guaranteed to exist. The simple fact that this amorphous phase shows no metallic states, either surface or bulk states, shows that there is a critical level of disorder that kills the TI phase. A full study of this system is currently underway, and many of these questions will soon be examined.

Chapter 6.

Emergence of decoupled topological surface states in bulk-insulating topological insulator thin films and inverse topological insulators heterostructures[8]



One of the major problems that topological insulators suffer from is the metallic bulk channel that allows electric current to flow in parallel to the metallic surface states. As mentioned in previous chapters this is due to the natural defects that occur in these materials. These defects free up charge carriers that are otherwise bound by chemical bonds, and eventually these carriers begin to occupy the bulk bands, which results in the development of the metallic bulk state. This parallel conduction channel is not a very serious problem for certain surface sensitive techniques like angle resolved photo-emission spectroscopy, which can easily distinguish between bulk and surface states. However, like measuring resistors in parallel, the surface state transport becomes

[8] The metallic bulk state in Bi_2Se_3 couples the topological surface state, and the addition of Cu depletes the bulk state and decouples the topological surfaces states. (Adapted from Editor's Suggested articles Physical Review Letters).

tangled up with transport through the metallic bulk state, and it is not a simple, nor an obvious task to distinguish between effects that originate from the surface states, the bulk states, or a combination between the two.

In this chapter we present two studies that illuminate some of novel transport effects that original from the topological surface states (TSS). There are two routes that we use to isolate the transport properties from the TSS. The first is on isolated Bi_2Se_3 films where Cu-doping is used to suppress the parasitic bulk channel. The second is by creating an entirely new artificial topological material known as an *inverse topological insulator*. By analyzing the weak anti-localization effect, both system show that the metallic bulk channel ties the TSS together, and as the bulk state is depleted, the surface state coupling gradually dies out when the thickness of the bulk insulating layer exceeds a critical value.

6.1 Weak anti-localization[9]

Weak anti-localization (WAL) and weak localization (WL) are effects in transport measurements where the conductance is enhanced or suppressed due to coherent back scattering, and are usually seen as a sharp change in the conductance with the application of a small magnetic field. The origin of this effect is due to coherent back scattering of the electrons along time reversible paths (see Ref. [130]MacKinnon, 2001 for more details). As shown in Figure 6-1, an electron can trace out many closed paths, which return the electron to its original path with opposite direction. The paths that contribute to this effect are those where no inelastic scattering events occur; such events cause the electrons to lose their phase information, which suppresses the interference effect. The amplitude to scatter backwards is then the sum of the amplitudes to go around the path clockwise (CW) and counter clockwise (CCW), during which the electron accumulates a phase that is given by the spin operator for an electron, $\boldsymbol{S}$, dotted into the angle operator, $\boldsymbol{\theta} = \phi\boldsymbol{n}$, where

[9] This section is based on work that is to be published in M. Brahlek *et al,* ArXiv **1408.1614 ,** (2014).

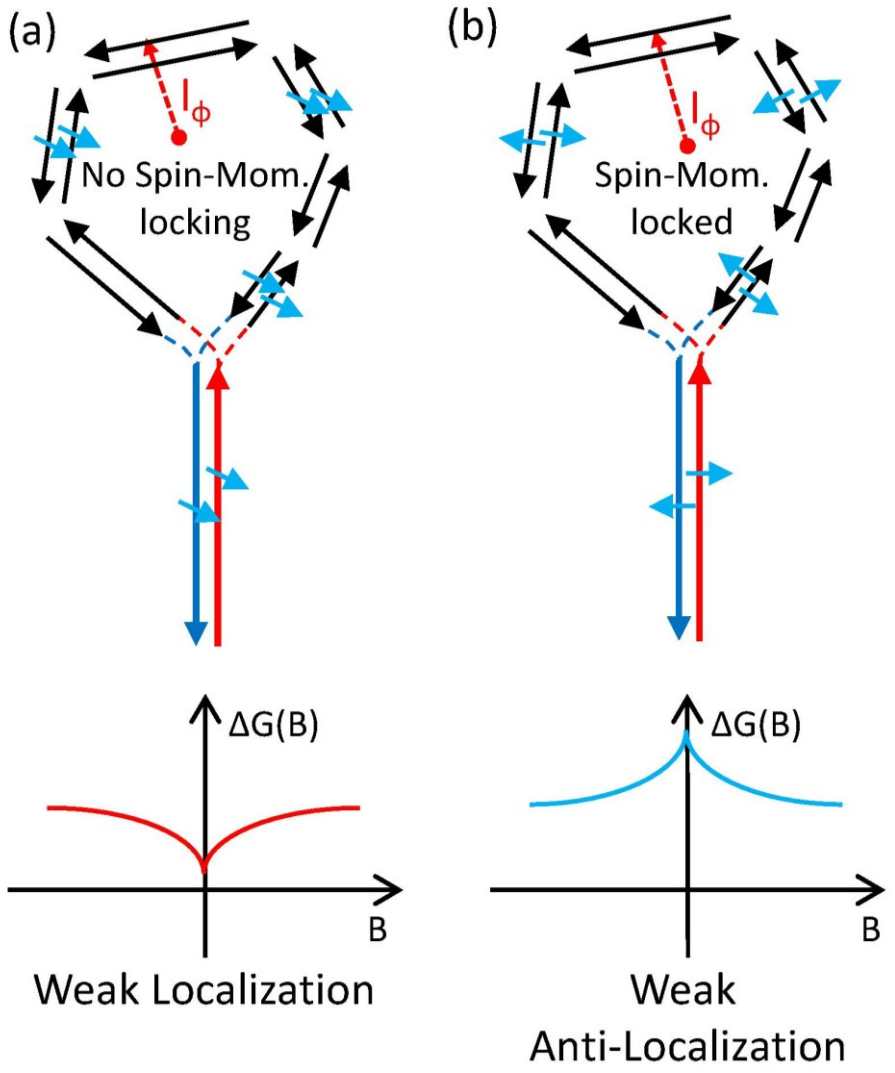


Figure 6-1 Cartoons showing how weak localizatoin (a) and weak anti-localization (b) occur due to coherenet backscattering along time reversable paths in the absence and presence of the spin-momentum locking, respectively. (Adapted from [30]Brahlek *et al.*, 2014)

ϕ is the angle accumulated in the spin part of the wave function and $\boldsymbol{n}$ is the unit vector along the rotation direction. This then yields

$$\psi_{CW} = \psi_0 e^{-i\frac{\boldsymbol{S}}{\hbar}\cdot\boldsymbol{\theta}_{CW}} = \psi_0 e^{\frac{i\phi}{2}}$$

$$\psi_{CCW} = \psi_0 e^{-i\frac{\boldsymbol{S}}{\hbar}\cdot\boldsymbol{\theta}_{CCW}} = \psi_0 e^{\frac{-i\phi}{2}}$$

and the amplitude ψ_0 contains any other phase information accumulated in the spatial part of the wave function; in the absence of magnetic field, magnetic impurities or other dephasing scattering

events, these two time reversal paths share the same spatial amplitude of ψ_0. The amplitude to return to the original position is then given by the sum of the CW and CCW components

$$\psi = \psi_0 e^{\frac{i\phi}{2}} + \psi_0 e^{\frac{-i\phi}{2}} = 2\psi_0 \cos\left(\frac{\phi}{2}\right).$$

From this we can see that there is a big difference between electrons without spin-momentum locking and those with strong spin momentum locking. For the former case, the angle the electron's spin sweeps out as it traverses the paths is negligibly small so $\psi_{WL} \approx 2\psi_0$. Compared with the case of no interference (i.e. classical case), which yields a probability of $2|\psi_0|^2$ for backscattering, this constructive interference doubles the probability of backscattering to $4|\psi_0|^2$. This enhanced backscattering effect is the root of the weak localization effect. For the latter case, where there is significant spin-momentum locking, the angle the spin sweeps out as it traverses the time reversible paths is π so $\psi_{WAL} \approx 2\psi_0 \cos(\pi/2) \approx 0$. This destructive interference effect suppresses the backscattering leading to the weak anti-localization effect.

Now, an applied magnetic field breaks the time reversal symmetry and the two opposite paths accrue opposite phases of equal magnitude, $\beta = e\Phi/\hbar$, which is proportional to the magnetic flux, Φ, through the time reversible loops. With this additional phase factor the wave function becomes

$$\psi = \psi_0 e^{i\beta} e^{\frac{i\phi}{2}} + \psi_0 e^{-i\beta} e^{\frac{-i\phi}{2}} = 2\psi_0 \cos\left(\frac{\phi}{2} + \beta\right)$$

and the exact destructive (WAL) or constructive (WL) interference becomes dephased as the magnetic field (hence β) increases. This phase factor strongly depends on the flux through the loops, and therefore with a small increase in the magnetic field, the localization effect quickly dies out: however, the periodic dependence on the magnetic field as implied by this formula does not actually occur because the area that determines the flux varies with different electron paths, and the correct treatment requires more rigorous approaches such as the HLN formalism to be discussed below. This gives rise to the sharp cusp at low magnetic field, which is a signature of WAL and WL as shown in Figure 6-1. In WAL (WL) the destructive (constructive) interference

effect suppresses (enhances) the backscattering and leads to enhanced (reduced) conductance in the absence of magnetic field in Figure 6-1b (Figure 6-1a). The magnitude of this enhancement (reduction) of the conductance due to WAL (WL) is typically of the order of 10% of the overall conductance in TIs, which is significant, yet implying that the other scattering channels account for the majority of the scattering.

The WAL and WL effects are typically analyzed using the HLN formalism, which was developed in the 1980s for 2D systems ([89]Hikami *et al.*, 1980). The following formula was derived to quantify the correction to the conductance: $\Delta G(B) = \tilde{A}e^2/(2\pi h)[ln(B_\phi/B) - \Psi(1/2 + B_\phi/B)]$, where h is Planck's constant, $\tilde{A}$ is a parameter related to the number of 2D conductive channels (1 for each channel), B_ϕ is the de-phasing field, and $\Psi(x)$ is the digamma function. The two independent parameters are $\tilde{A}$ and B_ϕ (in the original HLN formalism, the prefactor was in units of $1/(2\pi)$ per channel, but for simplicity, we incorporate this 2π term into the prefactor so that $\tilde{A} = 1$, instead of $1/(2\pi)$, implies a single channel).

B_ϕ is a magnetic field that roughly gives the width of the WAL cusp in the ΔG vs B plot. To get a physical understanding, B_ϕ can be related to a length scale by $B_\phi = \hbar/(4el_\phi^2)$, where l_ϕ is the de-phasing length. Roughly speaking all time reversible paths that are within the de-phasing length will contribute to the WAL effect. Such a parameter can then be thought of like a mean-free path, which is dependent on the microscopic details such as disorder. In contrast, the other parameter, $\tilde{A}$, is less sensitive to such details; $\tilde{A}$ is related to the number of channels in the system. For an ideal TI with surface states on the top and bottom, each should carry a channel number of 1, and if B_ϕ is similar, then the conductance should add, which would give an effective $\tilde{A} = 1 + 1 = 2$. However, in all known TIs, there is a metallic bulk channel, which complicates the analysis of this parameter. Despite this, $\tilde{A}$ provides vital clues to help unravel transport measurements.

6.2 Emergence of decoupled surface transport channels in bulk insulating Bi_2Se_3 thin films[10]

Existing TIs suffer from parasitic bulk conduction due to unavoidable self-doping effects. Progress has been made to reduce the bulk conductance both in bulk crystals utilizing anion mixing ([160]Ren *et al.*, 2010;[206]Xiong *et al.*, 2012), and in thin films, where the large surface to volume ratio has made it relatively easy to achieve dominant surface transport properties ([15]Bansal *et al.*, 2012). However, when TI samples are made thin, the conducting channel provides an inter-surface conduction path, and this causes the top and bottom TSS to be coupled into a single channel in weak anti-localization (WAL) effect ([43]Chen *et al.*, 2010;[108]Kim *et al.*, 2011;[126]Liu *et al.*, 2011;[176]Steinberg *et al.*, 2011;[15]Bansal *et al.*, 2012;[180]Taskin *et al.*, 2012); it should be noted that this indirect coupling through the bulk state in the WAL effect occurs over a length scale far greater than the critical thickness (~6 quintuple layers, QL ≈ 1 nm) of direct coupling that has been observed in ARPES ([219]Zhang *et al.*, 2010). Further effort to suppress the bulk conduction in thin TIs led to a significant reduction in the mobilities, suppression of the surface quantum oscillations and strong inter-surface coupling ([91]Hong *et al.*, 2012;[105]Kim *et al.*, 2012). Here, we report that by fully suppressing the parasitic channels in Bi_2Se_3 thin films through compensation doping Bi_2Se_3 by Cu, the TSS on opposite surfaces become fully decoupled from the bulk and each other with enhanced transport properties.

6.2.1 Cu-doped Bi_2Se_3 thin films

The Bi, Se, and Cu fluxes were calibrated in situ by a quartz crystal micro-balance, and ex situ using Rutherford backscattering spectroscopy. To track the doping level, the same Se and Bi cell temperatures were used for all samples, and by varying the temperature of the Cu cell , we were able to control the doping level to within ~ ±1.0% of the target concentration. The films produced

[10] This section is based on work that is published in M. Brahlek *et al,* Physical Review Letters **113**, 026801 (2014).

were single crystals of Bi_2Se_3 free of any Cu-based intergrowth; Figure 6-2 shows structural measurements confirming this. Figure 6-2a shows X-ray diffraction (XRD) scans for Cu doped Bi_2Se_3 at various doping levels. The only peaks detected were the standard (003n) peaks of Bi_2Se_3, and the peak of the underlying Al_2O_3 substrate (at ~41°). Similarly for reflection high energy electron diffraction (RHEED) images shown in Figure 6-2b-c, no extra peaks were observed in between the main peaks indicating no second phase existed. Both XRD and RHEED showed sharp peaks, indicating the films were of high quality single crystals that were atomically flat. This was confirmed by topographic measurements by atomic force microscopy (AFM) shown in Figure 6-2d-f.

Although a number of valence 2+ elements have been shown to suppress the naturally occurring *n*-type carriers in bulk Bi_2Se_3 crystals, most of them were ineffective in thin films. Interestingly, copper, which has been used to induce superconductivity in Bi_2Se_3 ([93]Hor *et al.*, 2010), turns out to effectively suppress the *n*-type carriers in Bi_2Se_3 ([191]Vaško *et al.*, 1974) thin films without degrading their mobilities. In high mobility undoped Bi_2Se_3 films, the Hall resistance, R_{xy}, is typically non-linear due to multiple contributions from the TSS and other non-topological channels ([15]Bansal *et al.*, 2012). However, as shown in Figure 6-3a-b, as the Cu concentration (x = Cu/Bi × 100%) was increased, R_{xy} went from non-linear to linear at $x \approx 2.5$ - 4.0%, and then became non-linear again beyond $x \approx 4.0\%$. This implies that at the optimal doping ($x \approx 2.5$ - 4.0%), the transport is dominated by one type of carriers and thus n_{Hall} can be extracted simply from $R_{xy} = B/(en_{Hall})$ (these, and all measurements herein, were performed at 1.5K).

To see the doping effect of Cu on Bi_2Se_3 films, Figure 6-3c shows that the Hall carrier density, which exhibits a clear minimum at the doping where R_{xy} became linear; the minimum Hall carrier density for these Cu-doped Bi_2Se_3 films is $n_{Hall} \approx 5.0 \times 10^{12}$ /cm^2, which is an order of magnitude lower than that of undoped Bi_2Se_3 ($n_{Hall} \approx 40 \times 10^{12}$ /cm^2) grown in otherwise identical conditions ([15]Bansal *et al.*, 2012). According to ARPES measurements on Cu-doped Bi_2Se_3 ([201]Wray *et al.*, 2010), the band structure of TSS remained almost unchanged beyond these doping

levels. As shown in section 6.2.5, comparing the measured sheet carrier density with ARPES spectra ([204]Xia *et al.*, 2009;[22]Bianchi *et al.*, 2010 ;[100]Jenkins *et al.*, 2012) estimates a surface E_F of ~146 meV above the Dirac point, which corresponds to ~70 meV below the bottom of the conduction band.

As can be seen in Figure 6-3d, the overall trend in the Hall mobility versus x is monotonically downward (black solid line) with a more unusual spike in mobility near the optimal doping (red solid line). The downward trend is naturally expected with the increase in

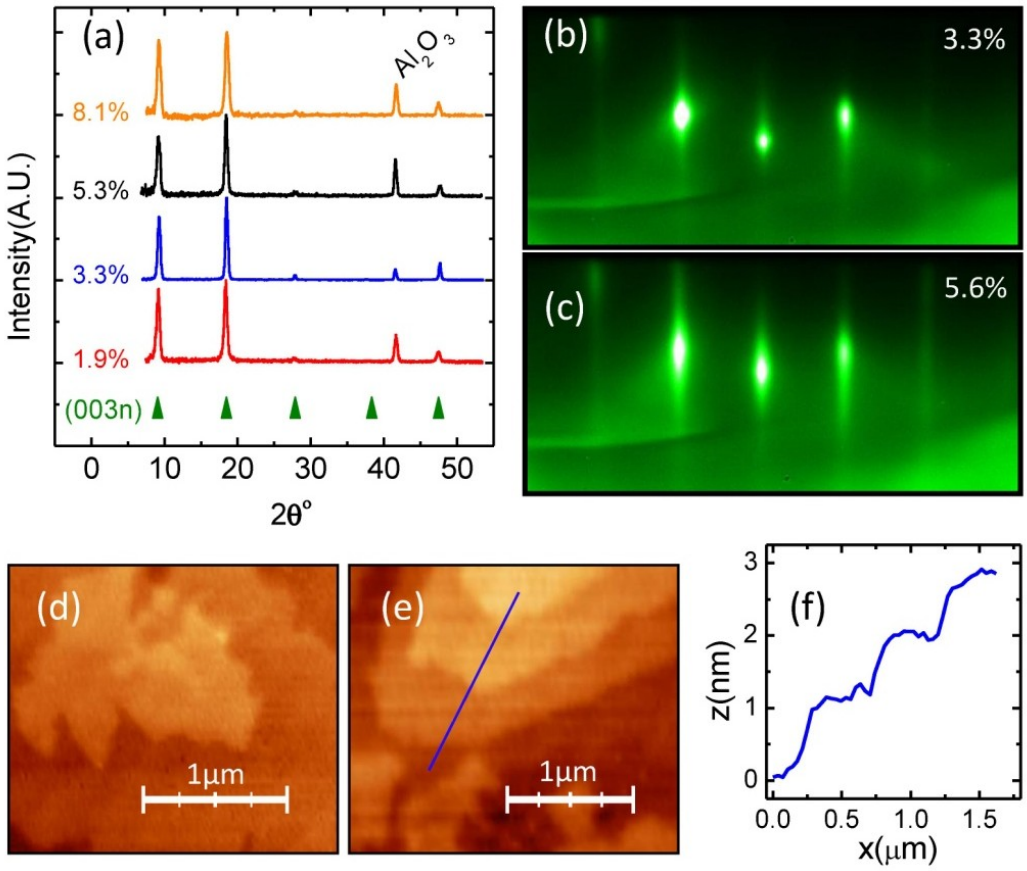


Figure 6-2 Structural measurements showing that the Cu-doped Bi_2Se_3 remain single phase. Cu-intergrowth would appear as extra peaks in the diffraction plots (**a-c**) or as clusters in the topographical images (**d-e**). **a** X-ray diffraction data at various values of x with the thickness fixed at 20 QL. The triangles mark the (003n) peaks of Bi_2Se_3, and the peak at ~41° marks the Al_2O_3 substrate peak. **b** Reflection high energy electron diffraction images also only show peaks for Bi_2Se_3 with no second phase present. **d-f** Atomic force microscopy images for $x \approx 3.3\%$ show large flat terraces with feature size in excess of 1μm. The line cut in (**f**) corresponds to the blue line in (**e**), which shows the step height to be ~ 1 nm -- the height for a single quintuple layer. (Adapted from [32]Brahlek *et al.*, 2014)

impurity scatting due to Cu dopants. In this regard, the upturn emphasized in red in Figure 6-3d is surprising; one likely explanation is that Cu atoms do not act solely as chargeless defects, but rather Cu facilitates Bi_2Se_3 to grow with fewer Se vacancies near the critical regime, and thus Cu acts effectively as p-type dopant while enhancing the mobility.

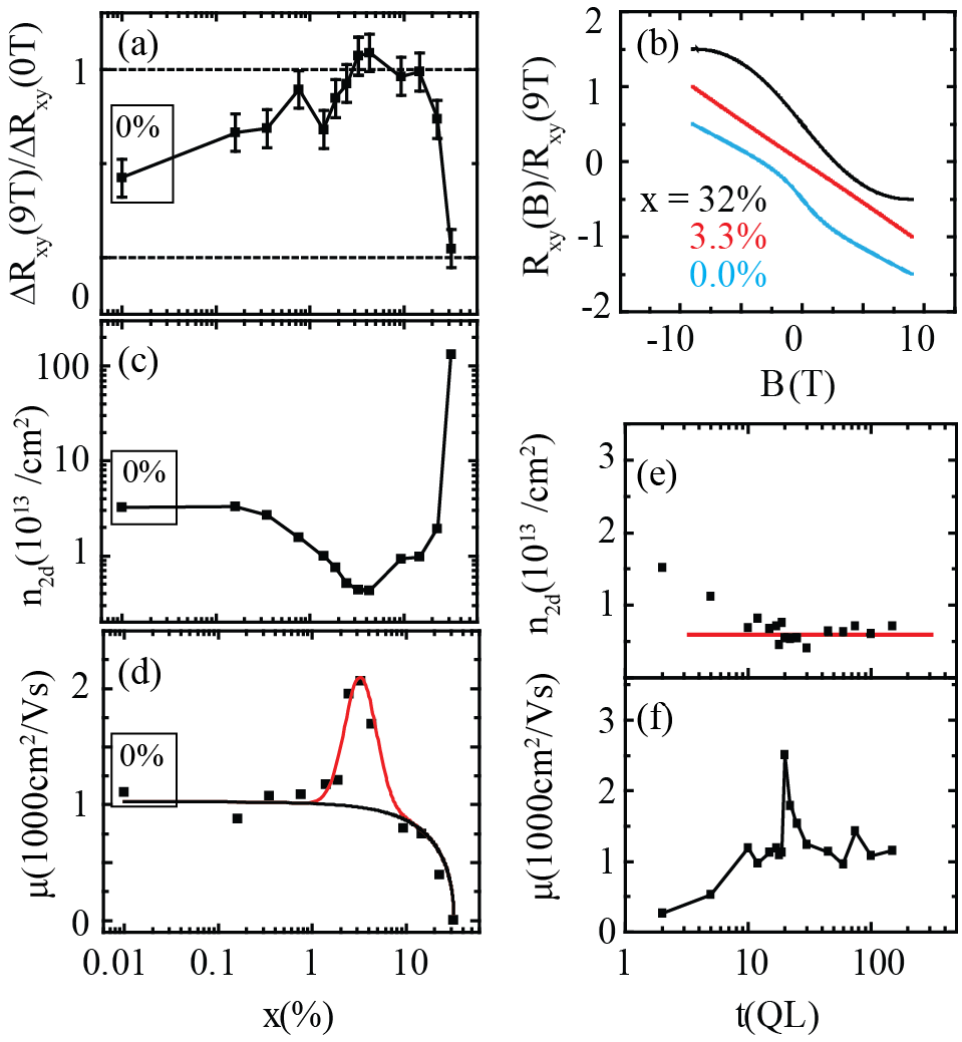


Figure 6-3 Hall effect measurements as a function of Cu-doping (fixed thickness ≈ 20 QL) and thickness (fixed Cu concentration, x ≈ 2.5-4.0%). a Ratio of the high field to the low field slope of R_{xy}. Where the ratio becomes unity, only a single carrier type contributes to the Hall effect. **b** R_{xy}/R_{xy}(9 T) vs. B showing R_{xy} is linear only in the optimum doping range. **c** Hall carrier density vs. x showing a minimum at x ≈ 2.5-4.0%. **d** Mobility vs. x showing the overall monotonic decrease of the mobility (black guide line) with increasing Cu doping and mobility peak (red guide line) near the optimal doping. Carrier density (**e**) and mobility (**f**) vs. thickness exhibiting nearly constant carrier density of ~6.0 ± 2.0 × 10^{12} /cm^2 and an anomalous mobility peak near 20 QL. (Adapted from [32]Brahlek *et al.*, 2014)

Regarding n_{Hall} vs. thickness (Figure 6-3e), between ~10 – 150 QL, the Hall carrier density was nearly constant at $n_{Hall} \approx 6.0 \pm 2.0 \times 10^{12}$ /cm^2 (the slight increase in carrier density below 10 QL implies that Cu doping is less effective at small thicknesses.). This thickness-independent signature can originate either from a non-topological 2D electron gas (2DEG) due to quantum confinement or from the TSS. In order for a 2DEG to exist, downward band-bending is necessary at the surfaces ([15]Bansal *et al.*, 2012), but with the estimated surface E_F, which is significantly below the conduction band minimum, if any band-bending is present it must be upward, giving rise to a depletion layer instead of an accumulation layer on the surfaces (see 6.2.7); such upward band-bending near the surface cannot harbor a 2DEG. This naturally leads to the conclusion that the data is most consistent with the TSS as the only surface channel in the optimally-Cu-doped samples. Lastly, unlike the sheet carrier densities, the mobility versus thickness (Figure 6-3f) does not show full thickness independence. Rather, at around 20 QL, there is an anomalous spike from ~1100 to 2500 cm^2/Vs. The exact origins of the mobility peaks observed in Figure 6-3d and f are currently unclear and need further studies.

6.2.2 Compensation doping of Bi_2Se_3

Unlike traditional semiconductor systems, compensation-doping Bi_2Se_3 has been found to be especially tricky. The energetics of the crystal structure make it possible for the dopants to naturally sit in many different locations within the lattice. This gives rise to different electrical behavior ranging from electron donor to electron acceptor, and even neutral behavior where there is no change in the carrier density. The exact behavior depends on the specific dopant and the conditions used to grow the crystal, and therefore molecular beam epitaxy (MBE) grown films may have a different behavior compared to single crystals.

This is best illustrated by comparing the successful use of Ca to compensate-dope bulk crystals of Bi_2Se_3 compared to our attempts to use Ca and Zn (both ionize to 2+, and therefore should be electron acceptors) to dope MBE thin films. In the bulk crystal case, Ca was

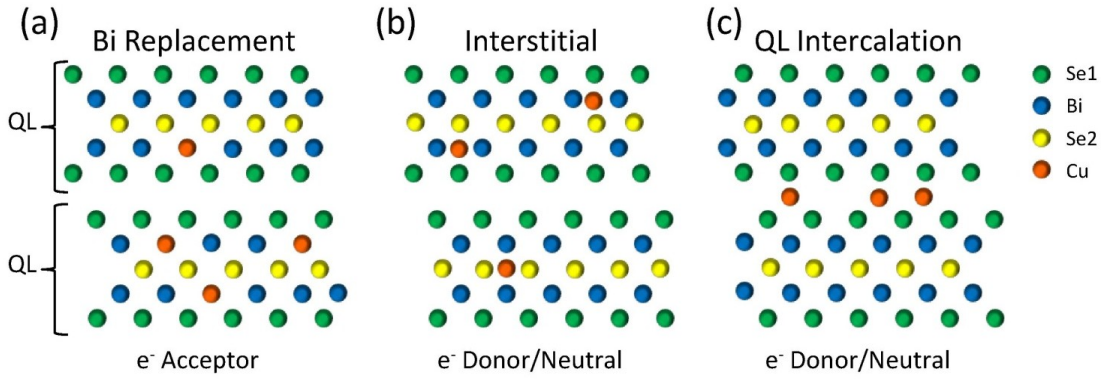


Figure 6-4 Schematic diagrams showing how doping Bi_2Se_3 by Cu occurs. For charge doping to occur, Cu atoms can replace Bi atoms as in **a**, they can sit interstitially in **b**, or they can intercalate in between the quintuple layers (QL) as in **c**. In **a** Cu atoms act as electron acceptors, while in **b** and **c** they can act as electron donors. However, types **b** and **c** can act as neutral dopants depending on the microscopic details, as discussed in the text. Overall, transport and XPS measurements indicate that Cu-doping works as both *p*- and *n*-type dopants, and therefore all of the above mechanisms should be present in our MBE grown thin films. (Adapted from [32]Brahlek *et al.*, 2014)

successfully used to make Bi_2Se_3 *p*-type at levels as low as 0.25% (x = Ca/Bi×100%) ([92]Hor *et al.*, 2009). This value is slightly larger than the value where exact charge compensation should occur; these undoped bulk crystals had a bulk electron density of ~10^{18}/cm^3, therefore the doping level for exact charge compensation is around ~0.1%, which results in a doping efficiency greater than ~ 50% (the remaining Ca atoms boost the hole carrier density). For Ca and Zn (a similar result was reported for Zn by [208]Xu *et al.*, 2012) doping Bi_2Se_3 thin films grown by MBE no change in the carrier density was observed to a doping level beyond ~10% which results in a doping efficiency of 0%. The difference between the behavior of Ca in MBE crystals and bulk crystals likely stems from the growth temperature. For MBE the films were grown at a temperature of around ~ 300°C for roughly an hour, while the bulk crystals were grown at a much hotter temperature of ~ 800°C for several days. The higher temperature can change the growth mode and provides more thermal energy for the dopant atoms to overcome any barrier to the chemical reaction with the Bi_2Se_3.

As mentioned before the difference in the electrical behavior of the dopants stem from the energetic of the Bi_2Se_3 crystal lattice. In Bi_2Se_3 there are three possible positions for the

dopants to reside. As illustrated in Figure 6-4a-c the dopant atoms can replace Bi atoms and chemically bond in the lattice, they can sit interstitially in between the chemically bonded Bi-Se sites, or lastly they can intercalate in between the van der Waals bonded QL. In the first case, if a valence 2+ atom, such as Ca, Zn, or Cu, replaces Bi then it will contribute 1 hole since it replaces Bi^{3+}. In the remaining two cases the dopants will contribute electrons, if ionized, or may remain inert and not contribute any electrons; this depends on the exact details of the crystal growth. Our transport measurements definitively show that in the case of MBE Bi_2Se_3 doped by Ca, Zn, or Cu, the majority of the atoms do not contribute any electrons, and therefore most of the atoms are either interstitial or intercalated in the case of Cu doping. This is confirmed by the X-ray photoemission spectroscopy (XPS) measurements shown in Figure 6-5a-b, which shows that the majority of Cu atoms in Bi_2Se_3 are in the Cu^{0+} state and remain effectively inert (if all Cu atoms contributed charge carriers, the Hall carrier density would be over two orders of magnitude larger than observed here). Moreover, these XPS spectra show that Cu-Bi_2Se_3 and pure Cu show nearly identical Cu 2*p*1/2 and 3/2 peaks, in both binding energy and line shape. This shows that the majority of the Cu sits either interstitially or intercalated in the QL gaps, and remains in an inert 0+ state working as either neutral (for low Cu content) or n-type (for high Cu content) dopants.

To see the main difference between Ca and Zn vs Cu, we can look at the difference in the carrier density versus doping level. Roughly speaking, Cu only shows a small change in the carrier density compared to the amount of Cu incorporated, whereas Ca and Zn showed no change whatsoever. This shows that most of the Cu remains inactive. There are two likely mechanisms: the first, and most obvious, is Bi replacement as shown in Figure 6-4a. For this case the reason for the better success of Cu, may stem from the electronegativity: Ca and Zn (~1.00, and ~1.65) are too different from Bi (~2.02), whereas Cu, whose electronegativity is ~1.90, is a much closer match to Bi. The second scenario may be that during the growth, Cu acts mostly as a flux, facilitating the growth of Bi_2Se_3 with fewer defects but not playing an active role electrically; this situation accounts for the increase in mobility near the optimum doping shown in Figure 6-3d.

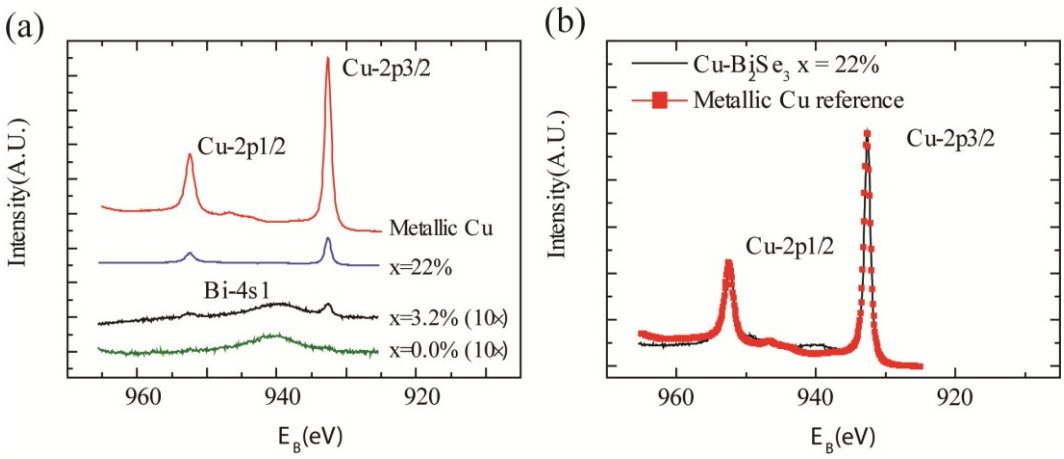


Figure 6-5 XPS data for Cu-2*p*3/2 and 2*p*1/2 for bulk Cu and Cu-Bi_2Se_3 with $x \approx 0 - 22\%$. a The Cu-peaks occur at the same binding energies (~932.6 eV and 952.5 eV respectively) for pure Cu and Cu-doped Bi_2Se_3. The Bi_2Se_3 also show a small peak in between the Cu peaks, which is the Bi-4*f* state (~940 eV). Binding energies are relative to C-1s, and $x \approx 0.0\%$ and 3.2% are multiplied by 10 for clarity. **b** In addition to the peaks occurring at the same binding energies, the curve shapes are nominally identical. This can be seen by superimposing the bulk Cu spectra on top of the Cu-Bi_2Se_3 with $x \approx 22\%$. Altogether the XPS data in **a-b** confirm the results from transport measurements (Figure 6-3) that show the majority of the Cu remains in an unionized state. (Adapted from [32]Brahlek *et al.*, 2014)

Overall, based on the current data we cannot say definitively the mechanism for the success of Cu doping in Bi_2Se_3, and this will remain a topic for further study.

In addition to the complications of the multiple locations of the dopant atoms and their different electrical behaviors, there is another complication. Bi_2Se_3 has both bulk and surface states, and depending on the alignment of their Fermi levels can frequently lead to band bending effects. These effects are apparent in the transport data shown in this chapter. Our undoped Bi_2Se_3 thin-films have a bulk doping level that sets the bulk carrier density to be around $\sim 5 \times 10^{17}$ /cm^2 or less, which corresponds to $E_F \approx 15$ meV above the conduction band minimum. On the other hand, the surface carrier density of these undoped Bi_2Se_3 thin-films is around 1.5×10^{13} /cm^2, which corresponds to a surface $E_F \approx 500$ meV above the Dirac point. Considering that the Dirac point is ~220 meV below the conduction band minimum, this implies that the undoped Bi_2Se_3 thin-films have downward band bending and accumulation layers on the surfaces ([15]Bansal *et al.*, 2012). For Cu-doped Bi_2Se_3, the surface Fermi level has dropped to around ~70 meV below the

bottom of the conduction band. The bulk Fermi level may have also dropped a little, but in thick films it can never drop below the conduction band minimum because of the Mott criterion as discussed in section 6.2.7; in other words, the bulk Fermi level is almost pinned at the bottom of the conduction band. This results in upward band bending and depletion layers on the surfaces of optimally-Cu-doped Bi_2Se_3 thin-films. This indicates that the main doping effect takes place at the surface, which is associated with the change in band-bending from downward (accumulation) to upward (depletion); more evidence and details on band-bending can be found in section 6.2.7.

6.2.3 Shubnikov de Hass oscillations and Lifshitz-Kosevish analysis

Shubnikov-de Hass (SdH) oscillations appear when a metal is subjected to a sufficiently strong magnetic field (such that $\mu \times B > 1$, where μ is the electron mobility, and B is the magnetic field. $\mu \times B \approx 2.5$ for our films.). The magnetic field collapses the continuous energy bands into a spectrum of individual energy levels. As the magnetic field increases, the energy level spacing increases. Therefore, increasing B pushes these energy levels across E_F, and thus gives rise to the resistance oscillating with increasing magnetic field ([172]Shoenberg, 1984).

Near the optimal doping, SdH oscillations were prominent enough to allow a standard Lifshitz-Kosevich (LK) analysis to be carried out on several samples as follows. The SdH oscillations of the resistance follows the LK formula $R_{SdH} = R_T R_D cos[2\pi(\Gamma/B+1/2+\beta)]$, where Γ is the period of oscillation, $2\pi\beta$ is the Berry phase. The first prefactor is given by $R_T = (\alpha T/B)/sinh(\alpha T/B)$, with $\alpha = 2\pi^2 m^* k_B/(\hbar e)$, where m^* is the cyclotron mass, k_B is Boltzmann's constant, T is the temperature. The second prefactor is the Dingle factor $R_D = exp[-\pi m^*/(eB\tau_D)]$, where τ_D is the Dingle scattering time, which is related to the quantum mobility by $\mu = e\tau_D/m^*$.

By measuring ΔR_{xx} vs. $1/B$ at different temperatures as shown in Figure 6-6a, we can extract the cyclotron mass. As shown in Figure 6-6d, at a fixed magnetic field the amplitude of the oscillations decays with increasing temperature. This isolates the R_T term in the LK formula: by fitting R_T vs T at fixed magnetic fields, we can extract the cyclotron mass. Once m^* is found,

R_T is determined, and then the Dingle factor, R_D, can be extracted by fitting the oscillations at fixed temperature to the LK formula, and then we can calculate τ_D.

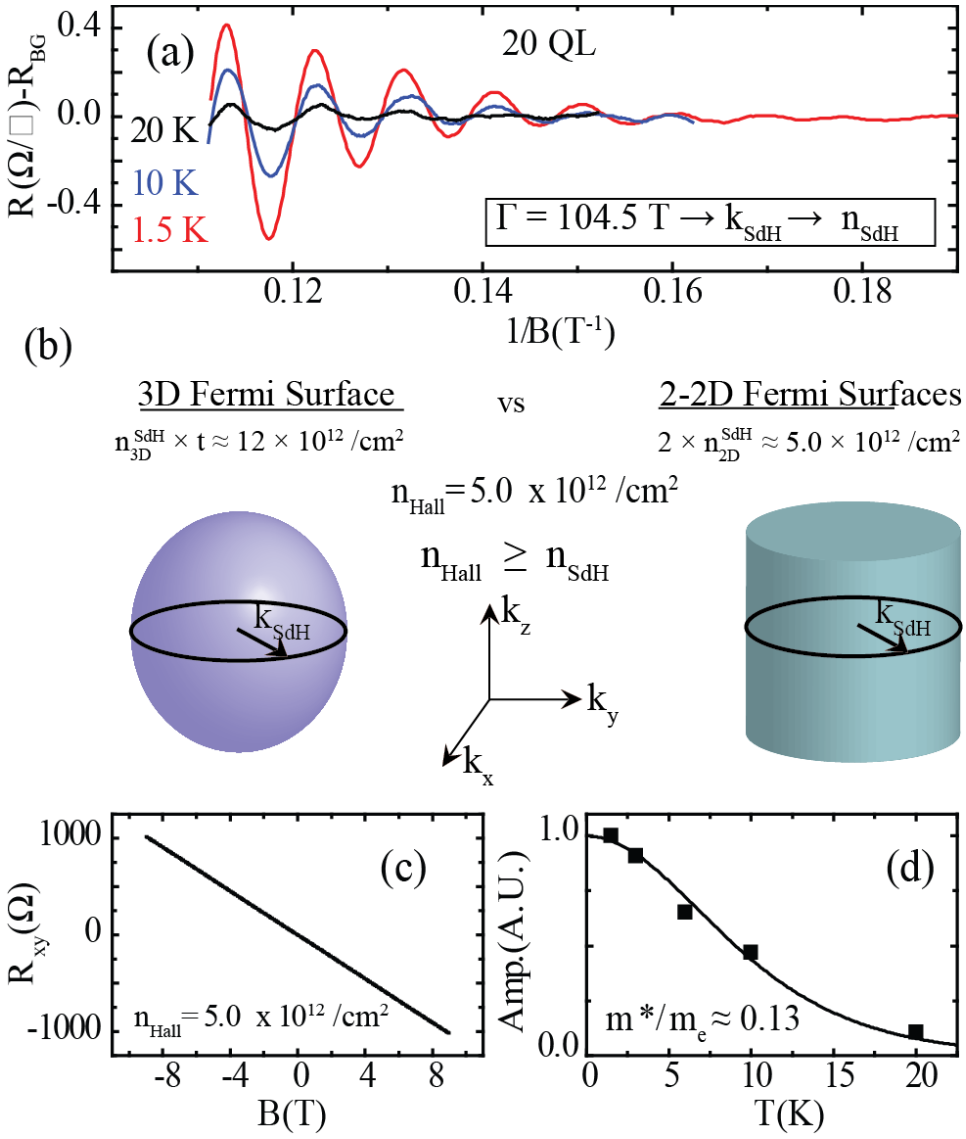


Figure 6-6 SdH oscillations and comparison with Hall effect for an optimally-doped 20 QL sample ($x \approx 2.5$–4.0%). a R_{xx} with a smooth polynomial background subtracted vs. $1/B$ taken at different temperatures. The frequency of oscillations, Γ, gives k_{SdH}, and therefore the carrier density n_{SdH}. **b** To calculate n_{SdH}, one must consider a 3D Fermi surface (left) versus two 2D TSS Fermi surfaces (right, one Fermi surface per TSS). The 3D Fermi surface would give an areal carrier density of at least ~12×10^{12} /cm^2, while the two 2D surfaces yield 5.0×10^{12} /cm^2. SdH oscillations can never produce sheet carrier densities higher than the Hall effect, ruling out the bulk origin of the SdH oscillation. **c** The Hall resistance, R_{xy}, yields a carrier density of $n_{Hall} \approx 5.0 \times 10^{12}$ /cm^2, which matches the 2D SdH estimate. **d** Amplitude of the SdH oscillation vs. temperature; the fit yields a cyclotron mass of 0.13 ± 0.01 times that of the electron mass. This agrees with the effective mass of the TSS as estimated from the ARPES spectra shown in Figure 6-7. (Adapted from [32]Brahlek *et al.*, 2014)

One standard use of SdH oscillations is to separate a 3D Fermi surface from a 2D Fermi surface. This is done by measuring the resistance versus magnetic field that is applied at an angle relative to the film's surface: if the oscillation is from a 3D Fermi surface, it should be roughly angle-independent, whereas if it is from a 2D Fermi surface, it should depend only on the surface normal component of the applied magnetic field. In the past, identification of the 2D Fermi surfaces by SdH oscillations has been used as direct evidence that the transport originates from the TSS. However, this is not necessarily true. In the case of a topologically trivial high mobility thin-films, the mean free path can be many times the film thickness. The transport signature of such a film would appear as if it was 2D from SdH oscillations, even if the transport does not originate from any surface states. An even more serious complication has been pointed out recently in undoped Bi_2Se_3 bulk crystals. Because of the intrinsic 2D nature of the Bi_2Se_3 crystal structure, originating from the weak van der Waals bonding between quintuple layers, even the bulk Bi_2Se_3 state can exhibit 2D behavior in SdH oscillations ([36]Cao *et al.*, 2012). What this shows is that the 2D signature observed by angle dependent SdH oscillations is less stringent than showing the transport is thickness independent as shown here, which is possible only if the conduction occurs through the surface states with negligible bulk contribution.

6.2.4 Shubkinov de Hass oscillations in Cu-doped Bi_2Se_3

Undoped or non-optimally-doped Bi_2Se_3 samples were afflicted by multiple low-amplitude SdH channels, thus making LK analysis inapplicable ([15]Bansal *et al.*, 2012); as shown in Figure 6-6a, however, an optimally doped 20 QL sample yielded a well-defined frequency of oscillation, Γ = 104.5 T. From $\Gamma = \hbar S_F/(2\pi e)$, where $S_F = \pi k_F^2$, the Fermi wave vector, k_F, was extracted, and the sheet carrier density can be extracted from this and compared with the Hall effect as follows .

As shown on the right side of Figure 6-6b, if two 2D TSS (from top and bottom surfaces) contribute, then the two surfaces will individually contribute a spin-polarized 2D carrier density of $n_{2D}^{SdH} = k_F^2/(4\pi)$. This yields a total carrier density of $2 \times n_{2D}^{SdH} = 5.0 \times 10^{12}$ /cm^2, which fully

matches the Hall effect value of n_{Hall} = 5.0 × 10^{12} /cm^2 (Figure 6-6c) and is also consistent with the estimations based on ARPES data (see section 6.2.5). For consistency, if we assume an elliptical bulk 3D Fermi surface as the origin of the SdH oscillations (Figure 6-6b, left side), we obtain a bulk carrier density of $n_{3D}^{SdH} \approx 6 \times 10^{18}$ /cm^3, which leads to conflicts with many other quantities. First of all, such a bulk state should yield an areal carrier density of $n_{3D}^{SdH} \times t \approx 12 \times 10^{12}$ /cm^2, which cannot be reconciled with the independently measured Hall effect, n_{Hall} = 5.0 × 10^{12} /cm^2. This is *unphysical* because the SdH oscillations, which capture only the high mobility channel, *can never yield an areal carrier density larger than the Hall effect*, which captures all carriers regardless of their mobilities. Also, such a bulk carrier density will put the surface Fermi level about 290 meV above the Dirac point, and thus a parallel surface carrier density of ~2.0 × 10^{13} /cm^2 should also be present yielding a Hall carrier density of ~3.0 × 10^{13} /cm^2, which is about six times larger than the measured value of n_{Hall}. Another inconsistency is that a bulk state of 6 × 10^{18} /cm^3 should clearly show up in the n_{Hall} vs *thickness* plot in Figure 6-3e. For example, at 150 QL, n_{Hall} should be around 6 × 10^{18} /cm^3 × 150 nm ≈ 10^{14} /cm^2, which is more than an order of magnitude larger than the measured value. Altogether, we can conclude that the origin of the SdH oscillations is from the TSS, and there is negligible bulk contribution. It is worth noting that in all previous reports, the surface SdH oscillations accounted for only a small fraction of the total sheet carrier density as measured by the Hall effect.

Lastly, the decay of the oscillations with temperature and magnetic field yielded a cyclotron mass of $m^*/m_e \approx 0.13 \pm 0.01$, and quantum mobility of around ~1400 cm^2/Vs. The former is exactly consistent with the value that is estimated from ARPES (see 6.2.5), and the latter is slightly lower, but fully consistent with the Hall mobility. The Berry's phase can be obtained by extrapolating linearly the Landau index, n, as $1/B \rightarrow 0$, which gives $n \approx 0.25$ from the lowest resolvable Landau indices of n = 12-16. Since this is very far from the so-called quantum limit (Landau index of 1) the extrapolated error is large, which yields inconclusive results as to the Berry's phase of this channel ([4]Analytis *et al.*, 2010;[205]Xiong *et al.*, 2013). Altogether, the

analyses of the SdH oscillations are fully consistent with the surface band parameters as inferred by ARPES (see next), and for the first time, account for the entire Hall carrier density.

6.2.5 Angle resolved photo-emission spectra compared with transport data

The clearest studies of the TSS have been done using angle resolved photo-emission spectroscopy (ARPES). Therefore, it is best to use their direct measurements of the band-structure as a guide to understand our transport data. To minimize ambiguous estimations, we made use of an empirical curve fit to the surface dispersion relations to estimate the parameters that we compare with transport data. As detailed elsewhere ([100]Jenkins *et al.*, 2012), an isotropic dispersion for the TSS can be fit to ARPES spectra that yields $E_{ave}(k) = 1.9k + 12.6k^2 + 2300k^6$, where k is in units of 1/Å and E is in units of eV. As shown in Figure 6-7, this function is overlaid on an typical ARPES

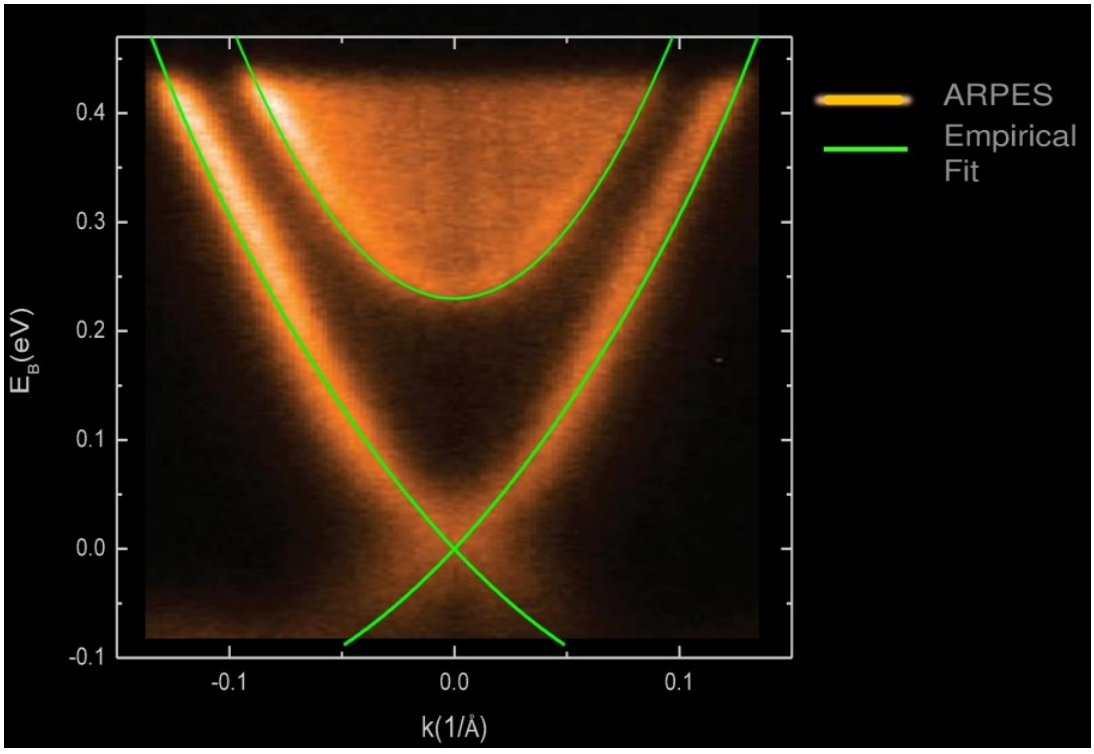


Figure 6-7 An ARPES spectrum taken from [22]Bianchi *et al.*, 2010 overlaid by the empirical functions, $E_{ave}(k) = 1.9k + 12.6k^2 + 2300k^6$ for the TSS and $E_{CB}(k) \approx \hbar^2k^2/(2m^*) + E_{DP}$ for the conduction band with $m^*/m \approx 0.15$ and $E_{DP} \approx 230$ meV. Parameters obtained from this function can be directly compared to transport data (see Table 6-2). (Adapted from [32]Brahlek *et al.*, 2014)

spectrum taken from [22]Bianchi *et al.*, 2010, which was chosen for its clarity. The empirical function fits equally well all ARPES spectra taken for bulk crystals, and thin films (both undoped and Cu-doped). Therefore, from this we calculate and then compare measured transport properties with what we expect from the empirical band structure.

One quantity that is typically measured by transport is the carrier density per area, n_{Hall}. This is done by measuring the transverse Hall resistance, R_{xy}, which is then related to the carrier density by $R_{xy} = B/(en_{Hall})$, where B is the magnetic field, and e is the electron charge. From this we can directly calculate the Fermi wave vector to be, k_F = 0.056 /Å, from $n_{Hall} = 2k_F^2/(4\pi)$, where the two in the numerator is from the fact that there are TSS on both the top and bottom surfaces. From here, we can use the empirical dispersion curve to calculate the Fermi energy to be $E_F \approx$ 146 meV above the Dirac point, and about ~69 ± 15 meV (depending on the distance from the conduction band minimum and the Dirac point, which we take to be 215 ± 15 meV) below the conduction band bottom. At this energy, the Fermi velocity is $v_F(k) = dE/d(\hbar k) \approx 5.0 \times 10^5$ m/s, and the cyclotron mass is $m^*(k)/m_e = \hbar k/(v_F\, m_e) \approx 0.13 \pm 0.01$. As shown in Table 6-2, these values can be compared to the data obtained by Hall effect measurements and SdH oscillations.

Table 6-2 Comparison of data obtained by the Hall effect, SdH oscillations, and values calculated via an empirical curve fitted to an ARPES spectrum. From this we see that all three techniques agree that the measured transport properties emanate from the TSS. (Adapted from [32]Brahlek *et al.*, 2014)

Measurement	k_F(1/Å)	v_F(10^5m/s)	n_{2D}(10^{12}/cm^2)	m^*/m	E_F(meV)
Hall Effect	0.056	-	5.0	-	-
SdH Oscillations	0.056	5.0	5.0	0.13±0.01	-
ARPES (@$k_{F,Hall}$)	0.056	5.0	-	0.13±0.01	146

6.2.6 Weak anti-location and decoupled topological surface states

With the general transport properties established above, we will now discuss the main finding in this study: the decoupling of the surface states with increasing thickness and Cu concentration. This was found by analyzing the WAL effect ([43]Chen *et al.*, 2010;[108]Kim *et al.*, 2011;[176]Steinberg

et al., 2011;[15]Bansal *et al.*, 2012;[56]Culcer, 2012), which appears as a cusp in magneto-conductance at low magnetic field (Figure 6-8a). This magneto-conductive effect is quantitatively described by the Hikami-Larkin-Nagaoka (HLN) formula: $\Delta G(B) = \tilde{A}e^2/(2\pi h)[ln(B_\varphi/B) - \Psi(1/2 + B_\varphi/B)]$, where h is Planck's constant, $\tilde{A}$ is a parameter related to the number of conducting 2D channels ($\tilde{A}$ = 1 for each channel), B_φ is the de-phasing field, and $\Psi(x)$ is the digamma function ([89]Hikami *et al.*, 1980). The channel number, $\tilde{A}$, in undoped Bi_2Se_3 thin films is always found to be around ~1, that is one channel, independent of sample thickness ([108]Kim *et al.*, 2011;[15]Bansal *et al.*, 2012). This implies that the two TSS in undoped Bi_2Se_3 films are coupled to the conducting bulk state

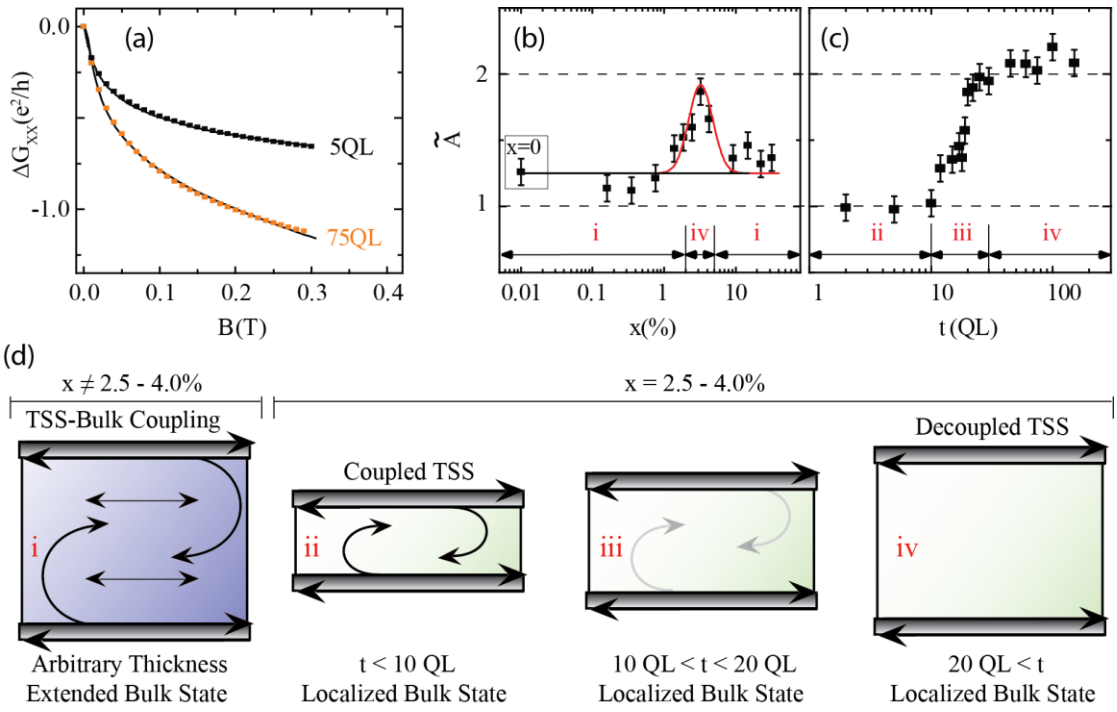


Figure 6-8 Weak anti-localization (WAL) effect showing the thickness and doping dependent coupling of the TSS. a The change in conductance taken at low field showing the WAL effect. The solid lines are the raw data, and the solid squares are the fit to the HLN formula. In (**b**) and (**c**), $\tilde{A}$ shows the effective number of 2D channels as a function of (b) composition (x = Cu/Bi × 100%) for fixed thickness (20 QL) and (c) thickness for fixed composition ($x \approx 2.5 - 4.0\%$). **d** A cartoon showing top and bottom surface coupling in the WAL effect. (i) Samples with extended bulk states behave as one channel over a wide thickness range, whereas the optimally Cu-doped samples with localized bulk states exhibit (ii-iv) decoupled surface states as the film becomes thicker than an inter-surface coupling length. (Adapted from [32]Brahlek *et al.*, 2014**)**

and together they effectively act as one channel. Here, it is important to note that this indirect coupling in the WAL effect occurs far beyond the thickness limit (~6 QL) of direct coupling caused by wavefunction overlap between top and bottom surfaces previously observed in ARPES ([219]Zhang *et al.*, 2010).

In Cu-doped Bi_2Se_3 films the behavior is quite different. Figure 6-8b shows how the effective number of the surface states changes with increasing Cu-doping. The first thing to notice is that at doping levels above and below the critical value of 2.5 – 4.0%, the channel number is nearly constant at ~1. Consistent with other transport properties, this indicates that the bulk is conducting and the entire film effectively acts as one channel. In the critical regime, the channel number sharply increases to ~2. This coincides with the minimum in the carrier density vs. x (see Figure 6-3c). Figure 6-8c shows $\tilde{A}$ at fixed x, while varying the thickness. As mentioned before, undoped Bi_2Se_3 yields $\tilde{A} \approx 1$ independent of thickness. However, optimally doped Cu-Bi_2Se_3 films exhibit qualitatively different thickness dependence. For thickness less than 10 QL, the fitting parameter $\tilde{A}$ quantizes at 1, corresponding to one channel; from 10 – 20 QL $\tilde{A}$ smoothly increased to 2, corresponding to two channels; and from 20 –150 QL $\tilde{A}$ stays constant at 2.

A simple explanation for the data shown in Figure 6-8b-c is as follows (see Figure 6-8d i-iv for corresponding schematic): below 10 QL, the TSS on the top and bottom surfaces can communicate (via tunneling or hopping) through the thin bulk, even if extended bulk states are suppressed, and remain strongly coupled, thus behaving as one channel. Between 10 – 20 QL, the TSS start to decouple as the communication dies out, and above 20 QL, which is too thick for inter-surface coupling to occur, the two TSS are completely decoupled from each other and behave as two isolated channels, as expected for bulk insulating systems ([42]Chen *et al.*, 2011;[176]Steinberg *et al.*, 2011;[106]Kim *et al.*, 2013).

This process can be better described by comparing the scattering times between available states at the Fermi level and the de-phasing time in the WAL effect ([176]Steinberg *et al.*, 2011). If the bulk has extended states at the Fermi level, surface electrons can coherently scatter in and out of the bulk states before the backscattered electrons de-phase; then, the system will behave effectively as one channel because the coherent time-reversible paths of all the transport channels are mixed up: this is the case for Bi_2Se_3 films with conducting bulks states as described in Figure 6-8d i. Even if extended bulks states are absent at the Fermi level, if the opposite surfaces are within tunneling (or hopping) regime such that the inter-surface tunneling (or hopping) time is shorter than the de-phasing time, then the system should still behave as one channel: Figure 6-8d ii. However, if available bulk states are localized at the Fermi level and the sample becomes thick enough such that the inter-surface scattering time becomes larger than the WAL de-phasing time, the two surfaces behave as two separate channels in the WAL effect because electrons on each

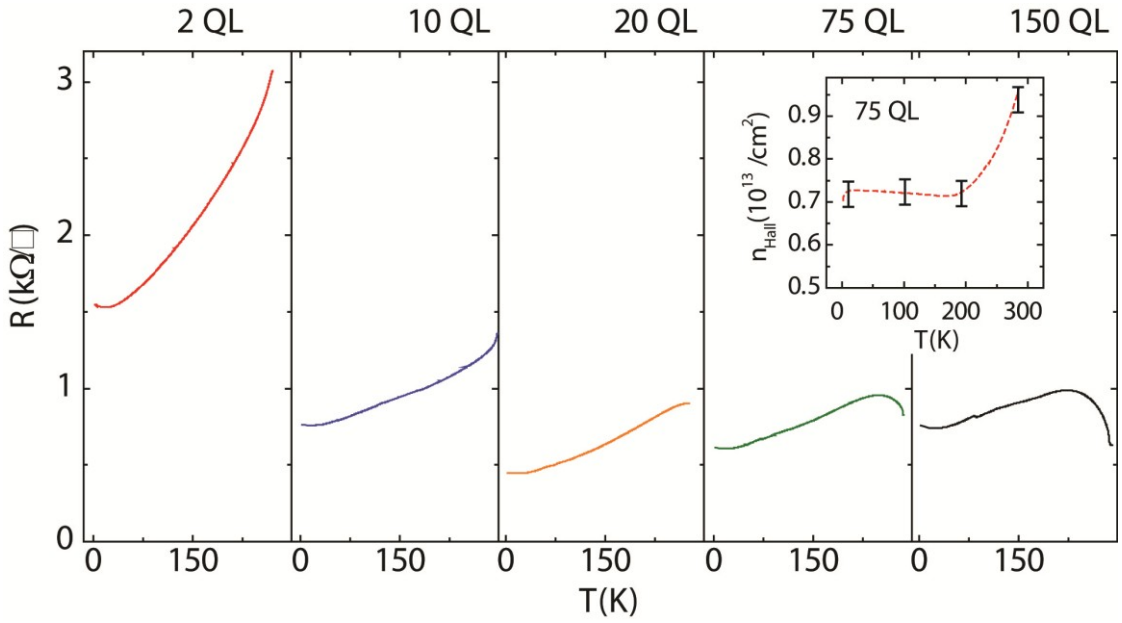


Figure 6-9 Resistance vs. temperature for t = 2, 10, 20, 75, and 150 QL (1 QL ≈ 0.95 nm). The hump at high temperature (250 - 300 K) is associated with the bulk-carriers freezing-out. This can be seen in the inset, which shows n_{Hall} vs. T for a 75 QL film. (Adapted from [32]Brahlek *et al.*, 2014)

surface complete their time-reversal paths before they scatter into opposite surfaces ([42]Chen *et al.*, 2011;[176]Steinberg *et al.*, 2011;[106]Kim *et al.*, 2013): Figure 6-8d iv.

6.2.7 Origin of suppressed bulk conduction: thickness dependant band bending

As discussed in the in the previous sections, all transport measurements coherently indicate that at low temperature only the topological surface states contribute to the conduction. Based on the resistance versus temperature data shown in Figure 6-9, we discuss two scenarios for the observed suppression of the bulk conduction.

The major problem in Bi_2Se_3 and all topological insulators is the existence of a bulk channel in parallel with the TSS. Much work has been done in bulk crystals and thin-films to eliminate the bulk channel (when we refer to the bulk channel, we mean this to incorporate bulk conduction plus any bulk related 2DEG—bulk is anything except the TSS) ([41]Checkelsky *et al.*, 2009;[4]Analytis *et al.*, 2010;[160]Ren *et al.*, 2010;[91]Hong *et al.*, 2012;[105]Kim *et al.*, 2012;[206]Xiong *et al.*, 2012). To date, however, the lowest volume carrier density reported is ~10^{16} /cm^3, and it was found that these crystals were dominated by bulk conductance ([35]Butch *et al.*, 2010). The fact that the bulk of Bi_2Se_3 remained conducting even at a bulk carrier density of ~10^{16} /cm^3 is not surprising; this can be understood by considering the Mott criterion, which gives the critical dopant density where a metal-to-insulator transition occurs ([66]Edwards and Sienko, 1982). This critical value occurs when the mean dopant spacing is of the same order as the effective Bohr radius of the defect states. In a crystal with a dielectric constant ε ($\varepsilon \approx 113$ for Bi_2Se_3 [163]Richter and Becker, 1977), the effective Bohr radius is given by $a_B = \varepsilon/(m^*/m) \times 0.05\ \text{nm} \approx 40\ \text{nm}$, where $m^* \approx 0.15m$ is the effective mass and m is the electron mass ([3]Analytis *et al.*, 2010). Then the critical dopant density, N_{CD}, is given by $a_B N_{CD}^{1/3} \approx 0.26$. From this we get $N_{CD} = (0.26/a_B)^3 \approx 3 \times 10^{14}$ /cm^3. Therefore, for bulk crystals unless the dopant density is below ~3×10^{14} /cm^3, their bulk states will be always metallic with E_F pinned to the conduction band minimum. A dopant density of order ~10^{14}/cm^3 is achievable in semiconductors such as Si and GaAs; however, since

the chemical bonding in Bi_2Se_3 is much weaker, it seems to be thermodynamically impossible for a defect density to reach such a low level in bulk crystals of Bi-based TI materials. However, in Bi_2Se_3 there is significant band-bending due to a large mismatch in the surface and bulk Fermi levels (before equilibrium), which we argue below can be employed to lower E_F into the bulk band gap in thin-films.

Band-bending can be qualitatively understood by knowing the surface carrier density, n_{ss}, along with the knowledge that the Mott-criterion fixes E_F in the bulk to be close to the conduction band minimum. In the previous sections we showed that if the top and bottom TSS carry a total density of $n_{ss} = 1.0 \times 10^{13}$ /cm^2, then E_F at the surface exactly touches the bottom of the conduction band. This then implies that if $n_{ss} = 1.0 \times 10^{13}$ /cm^2, band-bending is absent, and the conduction band minimum is at the same level from the surface into the bulk (see Figure 6-10a). If $n_{ss} > 1.0 \times 10^{13}$ /cm^2, which is the case for undoped Bi_2Se_3 where n_{ss} is typically ~3.0×10^{13} /cm^2 (see ref. [15]Bansal *et al.*, 2012), then the bands must be bent downward close to the surfaces (see Figure 6-10b). This gives rise to an accumulation region near the surface, which has been shown to harbor a quantum-well based 2DEG ([22]Bianchi *et al.*, 2010;[15]Bansal *et al.*, 2012). However, for optimally Cu-doped Bi_2Se_3 thin-films, $n_{ss} < 1.0 \times 10^{13}$ /cm^2 ($n_{ss} = 5.0 \times 10^{12}$ /cm^2, more specifically); this then implies that the bands bend upward close to the surfaces, which gives rise to a depletion region near the surface (see Figure 6-10c); such a state cannot have a non-topological 2DEG. Thus, in optimally Cu-doped Bi_2Se_3 the direction of the surface band-bending is upward with a depletion layer formed at the surface.

To describe the depletion region formed at the surface of the Cu-doped Bi_2Se_3 quantitatively, one needs to solve the Poisson equation self-consistently, but to capture a rough physical picture we can, to first approximation, solve it exactly by assuming the bulk dopants are distributed uniformly throughout the bulk. The Poisson equation is

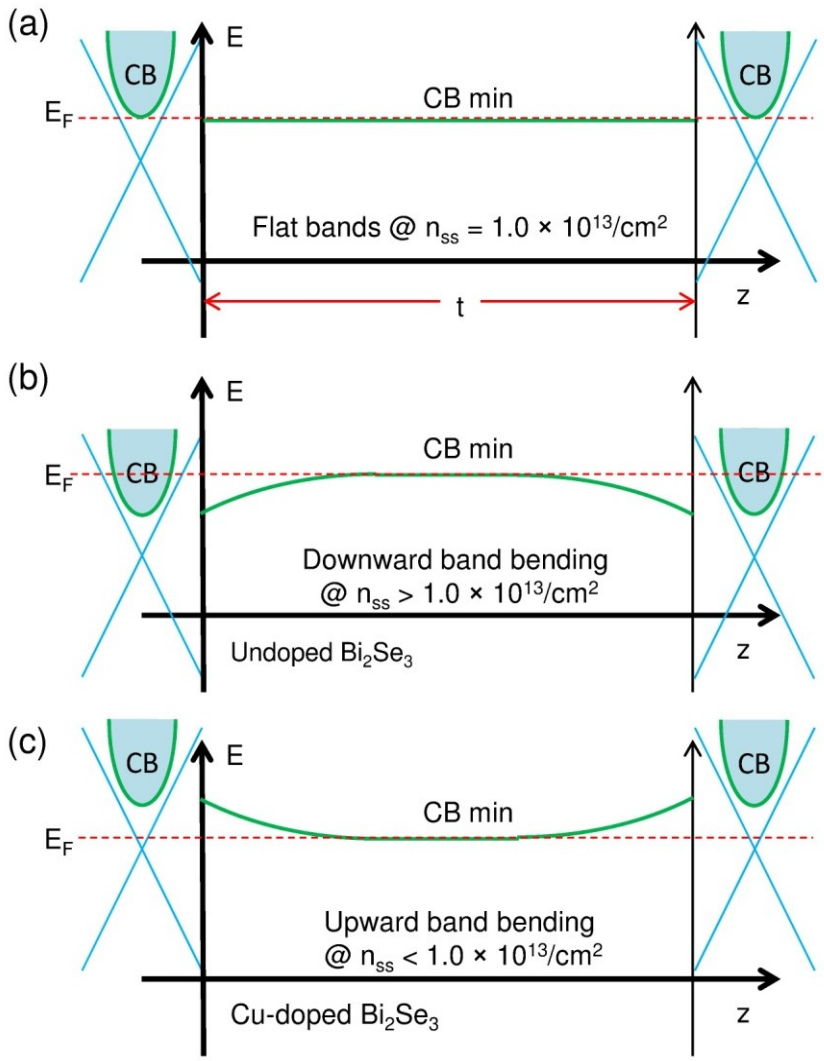


Figure 6-10 Band-bending dependence on the TSS carrier density: n_{ss} represents the total carrier density from top and bottom TSS. So long as the bulk donor density remains larger than ~3 × 10^{14} /cm^3, the Mott-criterion requires that E_F remains pinned to the conduction band (CB) minimum deep in the bulk. **a** A surface carrier density of n_{ss} = 1.0 × 10^{13} /cm^2 fixes E_F at the CB bottom on the surface, and therefore the bands must be flat. **b** If n_{ss} > 1.0 × 10^{13} /cm^2, then bands must bend downward. This is the case for undoped Bi_2Se_3. **c** If n_{ss} < 1.0 × 10^{13} /cm^2, then bands must bend upward. This is the case for optimally Cu-doped Bi_2Se_3. (Adapted from [32]Brahlek *et al.*, 2014)

$$\nabla^2 V(z) = -\frac{e^2 N_{BS}}{\varepsilon\varepsilon_0}$$

where V is the potential energy as a function of distance from the surface, z, e is the electron charge, and N_{BS} is the bulk dopant density. Subjecting this equation to the boundary conditions of $V(z = z_d) = 0$, and $V(z = 0) = \Delta V$, where ΔV is the energy difference between the bands deep in the

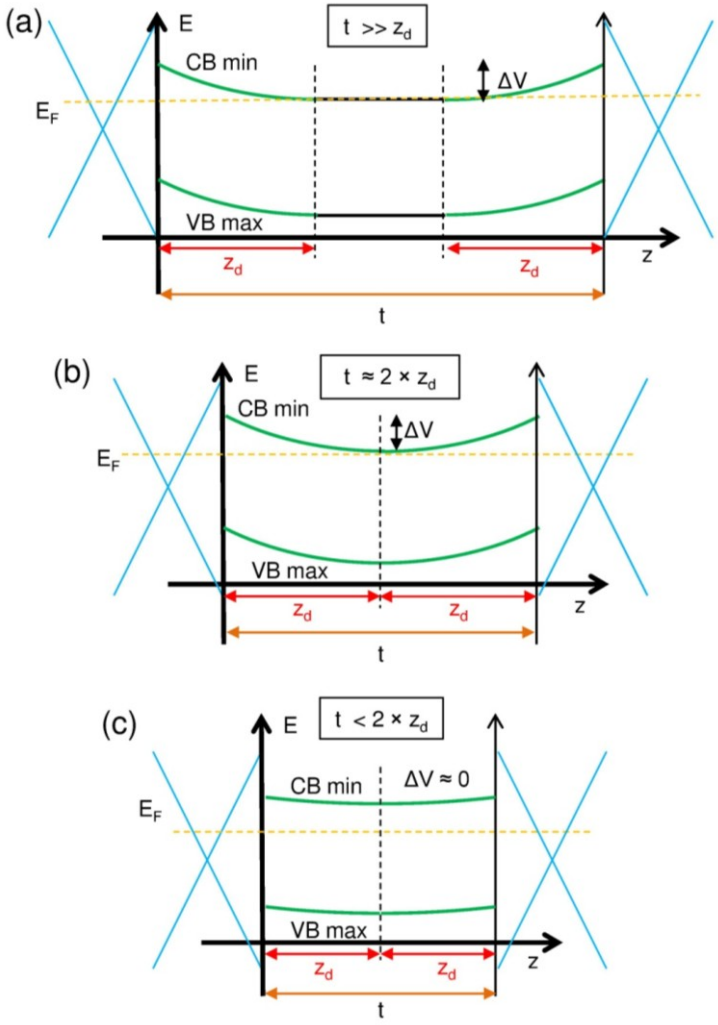


Figure 6-11 This illustrates how band bending behaves as the films are made thinner. In **a** $t \gg z_d$, so the film behaves as if it were a bulk sample. In **b** $t \approx 2 \times z_d$, and there still exists significant band-bending. In the limit of $t < 2 \times z_d$, as shown in **c**, z_d is limited by the film thickness, and hence the level of band-bending, ΔV, diminishes. (Adapted from [32]Brahlek *et al.*, 2014)

bulk and at the surface, and z_d is the distance over which the band-bending occurs in the bulk limit of $t \gg z_d$ (see Figure 6-11). The solution to Poisson's equation for one surface is then

$$V(z) = \frac{e^2 N_{BS}}{2\varepsilon\varepsilon_0}(z - z_d)^2, (z < z_d)$$

$$V(z) = 0, (z > z_d)$$

where the boundary conditions require that $\Delta V = e^2 z_d{}^2 N_{BS}/(2\varepsilon\varepsilon_0)$ and the electric field vanishes beyond $z = z_d$.

One thing we can note from this is that in the thin-film limit ($t < 2 \times z_d$), z_d is limited to half the film thickness. Therefore when $t < 2 \times z_d$, ΔV must decrease accordingly, and hence as the thickness decreases, $\Delta V \propto t^2 \rightarrow 0$, which implies that the band-bending is too energetically costly to be sustained and vanishes as the films are made thinner (see Figure 6-11a-c). To estimate a lower bound on z_d, we can use the bulk dopant density from our undoped Bi_2Se_3 films ([15]Bansal *et al.*, 2012), whose estimated upper bound is $N_{BS} \approx 5 \times 10^{17}$ /cm^3, and since the Cu atoms lower the bulk donor density, we can take N_{BS} in our Cu-doped films to be $< 5 \times 10^{17}$ /cm^3. Assuming that the surface Fermi level is ~ 70 meV below the conduction band minimum as

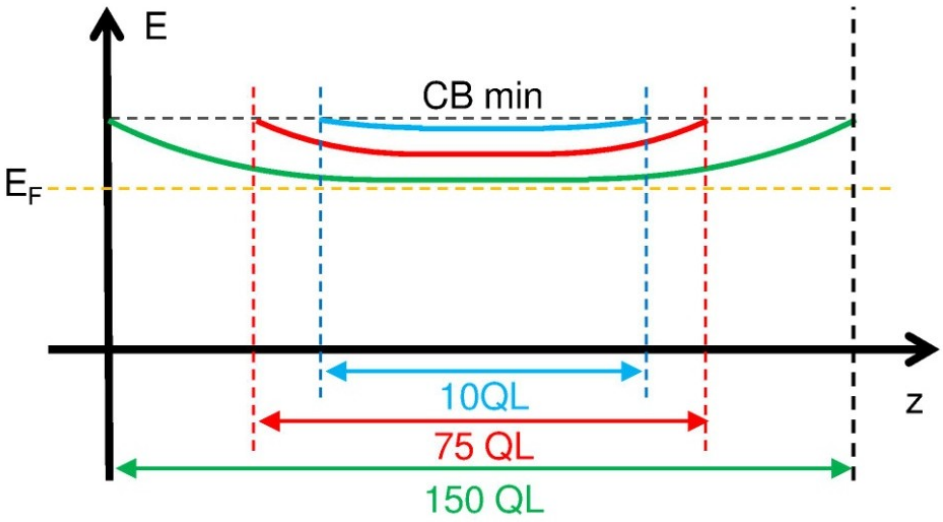


Figure 6-12 As the films are made thicker, the conduction band minimum as a function of distance from the surface dips down closer to the E_F. (Adapted from [32]Brahlek *et al.*, 2014)

estimated above in section B, and that E_F in the bulk at most touches the conduction band minimum (based on the Mott criteria), then we can take ΔV to be ~70 meV (see Figure 6-11a). Based on these values, the formula $\Delta V = e^2 z_d{}^2 N_{BS}/(2\varepsilon\varepsilon_0)$ then leads to z_d larger than 40 nm (see [3]Analytis *et al.*, 2010 where z_d is estimated to be of similar magnitude). Since at the surface E_F is pinned at ~70 meV below the conduction band minimum (which is set by the surface defect density), E_F must drop below the conduction band in the bulk as the film becomes thinner than 2 × z_d, which is an order of ~100 nm (see Figure 6-12). The effect is a true topological phase developing in the thin-film limit ($t \lesssim 100$ nm).

The band-bending as a function of thickness provides a natural explanation for the resistance versus temperature data shown in Figure 6-9. At high temperatures a hump develops for t > 30 QL; the size of the hump increased steadily with thickness. When the films are significantly thinner than twice the depletion length, the bands are almost flat, and the bulk E_F is as much as ~70 meV below the conduction band minimum; therefore, the bulk carriers are frozen all the way up to room temperature ($k_B T \approx 26$ meV). As the films become thicker, the band-bending develops, and the gap between the bulk E_F and the conduction band minimum ($E_{CB,min}$) gradually decreases. Then, activated bulk carriers can contribute to the conductance near room temperature and freeze out only at low temperatures when $k_B T \ll E_{CB,min} - E_{F,}$ which gives rise to a hump developing in R vs. T; this is qualitatively consistent with what we observed in Figure 6-9.

6.2.8 Origin of suppressed bulk conduction: Impurity band

The existence of impurity band in doped Bi_2Se_3 and other topological materials has been used to explain many R vs. T curves of low bulk conductance samples ([160]Ren *et al.*, 2010;[162]Ren *et al.*, 2012;[206]Xiong *et al.*, 2012). This is typically seen as an increase in the resistance with decreasing temperature. Therefore, despite the fact that the band-bending picture given above satisfactorily explains the data, it is appropriate to consider the impurity band effect as another scenario for the data shown in Figure 6-9.

In doped materials, as the doping density increases, the mean separation between dopants decreases and the wave functions of the electrons around the dopants will begin to overlap, and may eventually form a continuous impurity band in the bulk energy gap. If the impurity band is formed near or within the conduction band, then it goes back to the band bending scenario as discussed above, but if it is formed in the middle of the bulk band gap, the band bending picture is no longer valid. Due to the random location of dopant atoms, the mobility of impurity bands are, in general, very low, and at low temperatures the impurity band may be either metallic or insulating depending on the Ioffe-Regel criterion. According to the Ioffe-Regel criterion, a band will remain metallic only if the product of the Fermi wave vector, k_F, and the mean-free path, l, is larger than unity ([99]Ioffe and Regel, 1960). To estimate $k_F l$, we can take, as an upper bound, $\mu_{IB} <$ 50 cm^2/Vs ([206]Xiong *et al.*, 2012), and $n_{3D,IB} < 5 \times 10^{17}$ /cm^3, then $l = v_F\tau = (\hbar\mu_{IB}/e)(3\pi^2 n_{3D,IB})^{1/3}$, and $k_F = (3\pi^2\ n_{3D,IB})^{1/3}$, where the Fermi velocity is given by $v_F = \hbar k_F/m^*$, and the relaxation time is given by $\tau = m^*\mu/e$. Then, for the impurity band, if it exists, we finally get $k_F l < 0.2$. Since $k_F l$ is significantly smaller than unity, this impurity band should be insulating with the carriers freezing out at low temperature. This can also explain the carrier freeze-out and the hump feature in R vs. T as observed in Figure 6-9, even in the absence of the band bending effect.

If this low mobility bulk impurity band is the origin of the suppressed bulk conduction, then the inter-surface coupling will occur through hopping via the bulk states instead of direct inter-surface tunneling as would be the case for the strong insulating bulk state developed by the thickness-dependent band bending effect as discussed above. In the impurity band scenario, if the surface-bulk-surface hopping time becomes larger than the de-phasing time of the WAL effect on each surface, the WAL effect will exhibit two channels as observed here. In other words, both the impurity band picture and the thickness-dependent band bending picture can explain our data satisfactorily. In order to pinpoint the origin behind the suppression of the bulk conduction, more detailed studies will be needed.

6.2.9 Concluding remarks

In an ideal topological insulator films the bulk state, which is supposed to be insulating, should not provide any electric coupling between the two metallic surfaces. However, transport studies on existing TI films show that the topological states on opposite surfaces are electrically tied to each other at thicknesses far greater than the direct coupling limit where the surface wavefunctions overlap. Here, we have shown that as the conducting bulk channels were suppressed, the parasitic coupling effect diminished and the decoupled surface channels emerge as expected for ideal TIs. In Bi_2Se_3 thin films with fully suppressed bulk states, the two surfaces, which are directly coupled below ~10 QL, become gradually isolated with increasing thickness and were completely decoupled beyond ~20 QL.

This is the first demonstration of how the TSS on opposite surfaces can become decoupled with compensation doping and thickness control. This sheds light on the distinction between the direct coupling of surface states caused by wavefunction overlapping, and the indirect coupling that is mediated by the conducting bulk states. While the direct coupling effect exists only below ~6 QL according to ARPES measurements, the indirect coupling persists beyond hundreds of QLs in the standard bulk conducting TI samples according to transport measurements. Here we have demonstrated that as this indirect coupling is eliminated by suppressing the parasitic bulk channels, the thickness dependence of the ideal TSS emerges even in transport channels. Still, the critical decoupling thickness of 10~20 QL observed in the present work is significantly bigger than the critical thickness (6 QL) observed in ARPES, and whether this is due to the difference in sensitivity between transport and ARPES probes or due to an incomplete understanding of the underlying physics is an open question and needs further studies. Altogether, this study has brought us one step closer to the realization of pure topological devices.

6.3 Inverse topological insulator with tunable intersurface coupling[11]

The topological classification scheme is rooted in the ability to distill the global properties of an object down into a single number known as a topological invariant. This notion of topology can be extended from the archetypal example of geometric topology, where shapes are classified based solely on the number of holes, to electronic materials, where the main physical implication occurs on the boundary between materials that belong to different topological classes. In the 3-dimensional (3D) class of TIs, unusual topological surface states (TSS) form on the 2-dimensional (2D) boundary with non-TIs. These surface states have metallic gapless energy bands, which disperse linearly with momentum like massless photons, and the spins of the surface electrons are locked perpendicular to the direction of their momentum (see [85]Hasan and Kane, 2010;[158]Qi and Zhang, 2011). Since the initial discovery, experimental confirmation of these novel properties has largely been confined to surface-sensitive probes such as angle-resolved photo emission spectroscopy ([94]Hsieh *et al.*, 2008;[44]Chen *et al.*, 2009;[204]Xia *et al.*, 2009) and scanning tunneling microscopy ([165]Roushan *et al.*, 2009); direct transport measurements of the TSS have proven difficult due to the intrinsic defects that pin the Fermi level into either the conduction or valence bands, which effectively shorts the transport through the TSS ([4]Analytis *et al.*, 2010;[35]Butch *et al.*, 2010;[175]Steinberg *et al.*, 2010;[40]Checkelsky *et al.*, 2011;[108]Kim *et al.*, 2011;[126]Liu *et al.*, 2011;[15]Bansal *et al.*, 2012;[105]Kim *et al.*, 2012;[106]Kim *et al.*, 2013;[32]Brahlek *et al.*, 2014).

Figure 6-13a shows the typical experimental and theoretical framework that has been used to study TIs, which consists of a single slab of a TI (topological invariant of 1) surrounded by a non-TI (topological invariant of 0), and it is at this interface that the TSS are formed. Here we present a new platform to study the topological nature of materials by inverting the single-slab archetype and surrounding a non-TI by a TI. The bulk-boundary correspondence (BBC) theorem

[11] This section is based on work that is to be published in M. Brahlek *et al* (2014).

([85]Hasan and Kane, 2010) maintains that this inverse topological insulator (ITI) geometry is physically equivalent to the standard TI geometry because the TSS formed at such an interface are nominally identical, which can be seen by comparing Figure 6-13a and b. Through the use of molecular beam epitaxy (MBE), Figure 6-13c shows that this ITI geometry is feasible by inserting a layer of the tunable non-TI $(Bi_{1-x}In_x)_2Se_3$ (band gap ~1.3 eV at x = 100% [199]Watanabe *et al.*, 1989;[28]Brahlek *et al.*, 2012;[125]Liu and Vanderbilt, 2013;[202]Wu *et al.*, 2013) between two layers of the TI Bi_2Se_3 [198]Wang *et al.*, 2011. By varying the thickness of the non-TI barrier layer from the ultra-thin (1 quintuple layer, 1 QL ≈ 1 nm) to the thick regime (> 100 QL), we find that the Bi_2Se_3 layers become electrically isolated and the TSS emerge at the interfaces when the In_2Se_3 thickness is above ~3 QL, as shown by both transport measurements and first-principles calculations. Further, by decreasing x in the $(Bi_{1-x}In_x)_2Se_3$ barrier layer from x = 100% to 30%, its potential barrier height is lowered, and the coupling strength gradually increases for the same layer thickness. For a fixed composition x, the interfacial TSS undergoes a coupled-to-decoupled

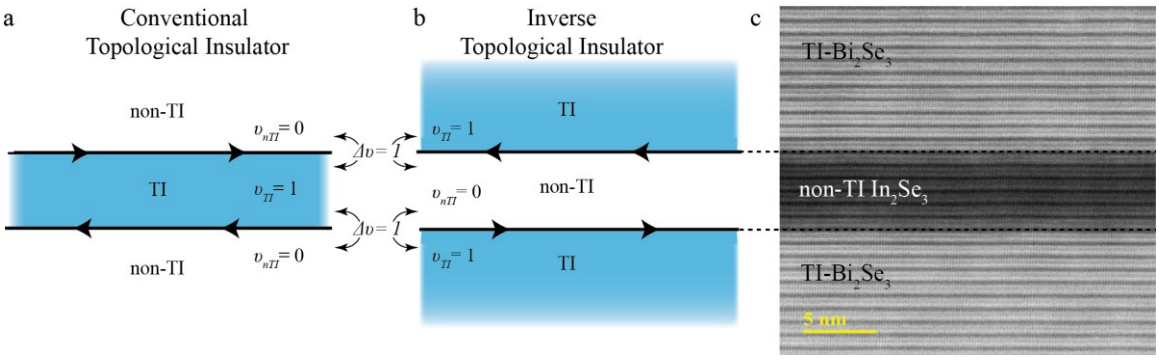


Figure 6-13 A schematic and TEM image showing a conventional TI vs an inverse TI. a-b Topological surface states form when there is a change in topological invariant (ν) across an interface (*i.e.* $\Delta\nu$ = 1). **a** Typical experimental paradigm: a topological material (characterized by a topological invariant ν_{TI} = 1) surrounded by non-topological insulator (ν_{nTI} = 0), which forms a metallic topological surface state at the interface (indicated by the arrows). **b** Inverse topological insulator (ITI) is formed by inverting the role of non-TI and the topological material, and a nominally identical surface state forms at this interface. **c** High-resolution high-angle annular dark-field scanning transmission electron microscopy image showing the physical realization of the ITI geometry (**b**) in a Bi_2Se_3-In_2Se_3-Bi_2Se_3 heterostructure. (Adapted from [31]Brahlek *et al.*, 2014)

transition as the thickness grows beyond a critical value. This transition from coupled to decoupled can be described by a simple analytic function of the thickness and composition, which suggests the presence of a fundamental mechanism behind the intersurface coupling through this true band-insulating barrier layer. However, when x is reduced below ~25%, the barrier layer becomes metallic and TSS remain fully coupled over the entire thickness range, with the analytical function breaking down.

6.3.1 Experimental methods

All samples were grown using 10 mm × 10 mm *c*-plane Al_2O_3 substrates. The first Bi_2Se_3 layer was grown according to the two-step growth method developed at Rutgers University where the first 3 QL was grown at 135°C, which was followed by slowly annealing the sample to 300°C, where the subsequent 27 QL of Bi_2Se_3 layers were grown. Once the first Bi_2Se_3 layer finished growth, the In_2Se_3 of the specified thickness was grown, followed by the remaining Bi_2Se_3 layer. All the samples were then capped by 50 QL of In_2Se_3 which stabilized the films during exposure to atmosphere. For the samples with $(Bi_{1-x}In_x)_2Se_3$ as the barrier layer, the same basic recipe was used. The Bi and In cell temperatures were adjusted such that when opened together the resulting film gave the concentration that was sought. All the concentrations were checked by a combination of ex situ Rutherford back scattering spectroscopy and in situ quartz crystal microbalance measurements, and the results were within ±1% of the target values.

All transport measurements were carried out at 1.5 K using the standard Van der Pauw lead geometry, and the magnetic field was applied perpendicular to the films' surface. The raw data was symmetrized to remove any odd component from R_{xx} and any even component from R_{xy}. The carrier density and mobility of the films ranged between 3-7 × 10^{13} /cm^2 and 500-1000 cm^2/Vs, and there was no correlation between the transport data and the value of $\tilde{A}$. From the WAL fitting, l_ϕ ranged between 50-100 nm and also showed little correlation with the other transport data or $\tilde{A}$. The temperature dependence of resistivity for all samples showed typical

monotonic decreasing behavior with decreasing temperature, which is typical of a metal. $\tilde{A}$ was independent of temperature below ~20 K, above which deviation occurred as thermal effect suppresses the WAL signal.

TEM sample preparation was carried out with focused-ion beam technique using 5 keV Ga^+ ions. A JEOL ARM 200CF equipped with a cold field-emission gun and double-spherical aberration correctors operated at 200 kV was used for high-angle annular dark-field scanning transmission electron microscopy with the collection angles ranging from 68 to 280 mrad (TEM performed at Brookhaven National Lab by Myung-Geun Han and Yimei Zhu).

6.3.2 Bi_2Se_3-In_2Se_3-Bi_2Se_3 inverse TI heterostructure

As discussed in 6.1, the weak anti-localization (WAL) effect is a common feature of the magneto-resistance in TI thin films (see Figure 6-14a for measurement geometry, and note that both layers of Bi_2Se_3 are equally contacted when the In wires are physically pressed into the film) ([42]Chen *et al.*, 2011;[108]Kim *et al.*, 2011;[126]Liu *et al.*, 2011;[176]Steinberg *et al.*, 2011;[15]Bansal *et al.*, 2012;[37]Cha *et al.*, 2012;[2]Aguilar *et al.*, 2013;[106]Kim *et al.*, 2013;[32]Brahlek *et al.*, 2014). The WAL effect is typically seen as a sharp cusp at small field in resistance vs magnetic field, and is quantifiable by the Hikami-Larkin-Nagaoka (HLN) formula $\Delta G(B) = \tilde{A}e^2/(2\pi h)(ln[\hbar/(4el_\phi^2 B)] - \Psi[1/2 + \hbar/(4el_\phi^2 B)])$, where h is Planck's constant, Ψ is the digamma function, $\tilde{A}$ is related to the number of 2D conducting channels ($\tilde{A} \approx 1$ for each channel), and l_ϕ is the de-phasing length ([89]Hikami *et al.*, 1980). In general l_ϕ is limited by the inelastic scattering length, which depends strongly on microscopic details such as disorder and phonons. In contrast, $\tilde{A}$ has been found to be much more robust, and monitoring the effective channel number gives information about the coupling between the various transport channels ([108]Kim *et al.*, 2011;[176]Steinberg *et al.*, 2011;[15]Bansal *et al.*, 2012;[106]Kim *et al.*, 2013;[32]Brahlek *et al.*, 2014). In TIs, $\tilde{A} \approx 1$ occurs when the bulk state is conducting and the surface-to-bulk scattering time is shorter than the electron dephasing time; in this limit electrons are allowed to scatter from the surface into the bulk before losing their phase

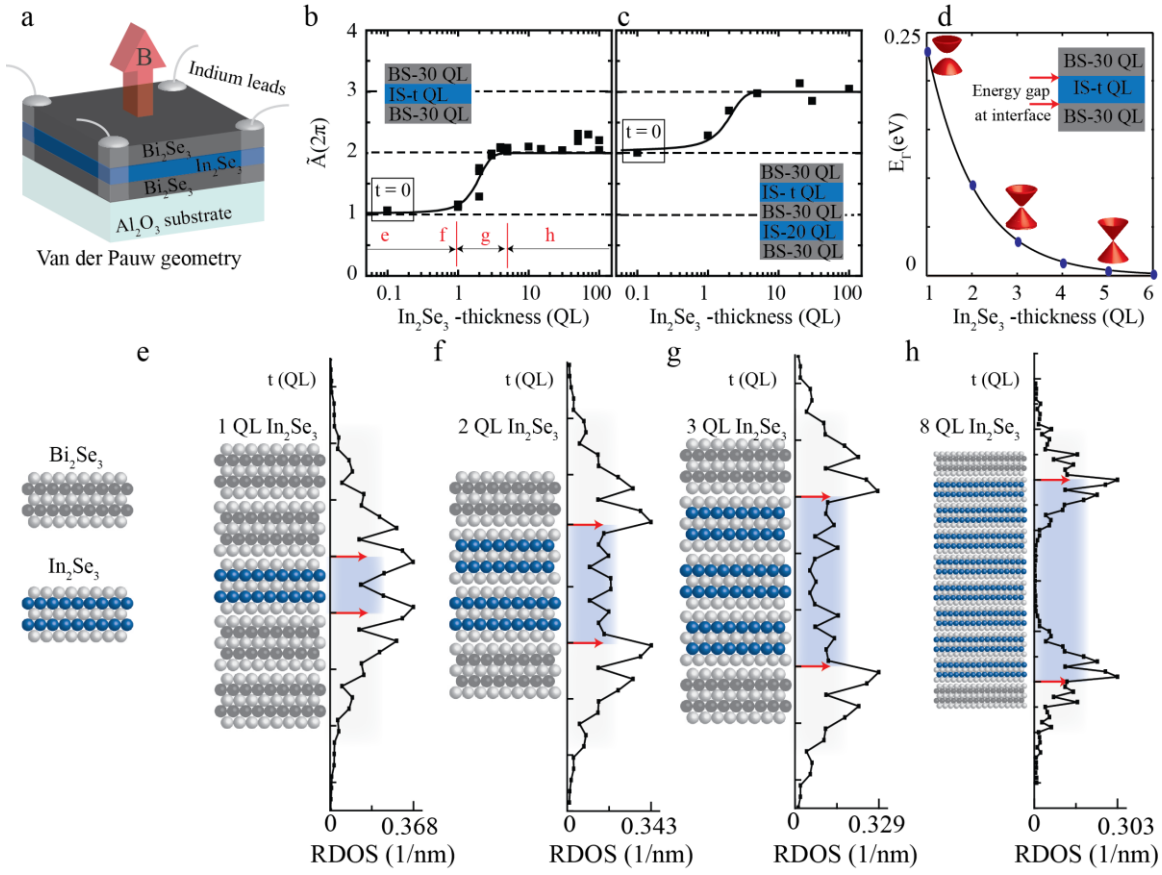


Figure 6-14 Weak anti-localization and the first principles calculations for Bi_2Se_3–In_2Se_3–Bi_2Se_3 films. a Schematic of the experimental setup where the films were grown on square substrates, and all layers of Bi_2Se_3 were contacted by pressed indium wires in the Van der Pauw geometry. **b-c** $\tilde{A}$ extracted from the HLN formula plotted versus thickness of the In_2Se_3 layer in Bi_2Se_3-In_2Se_3-Bi_2Se_3 (30-*t*-30 QL) (**b**), and Bi_2Se_3-In_2Se_3-Bi_2Se_3-In_2Se_3-Bi_2Se_3 (30-20-30-*t*-30 QL) (**c**). **d** Calculations of the energy gap at the interface of Bi_2Se_3-In_2Se_3 at the Γ point showing that the gap at the Dirac point closes as the In_2Se_3 thickness increases. **e-f** Schematic of the heterostructure formed for increasing In_2Se_3 thickness alongside the calculated real space density of the states (RDOS, see 6.3.3) around the Fermi level as a function of thickness, showing the emergence of the TSS at ~3 QL. (Adapted from [31]Brahlek *et al.*, 2014)

information ([129]Lu and Shen, 2011;[176]Steinberg *et al.*, 2011;[2]Aguilar *et al.*, 2013;[106]Kim *et al.*, 2013). Most TI films, including the Bi_2Se_3 films studied here, have conducting bulk states and thus exhibit $\tilde{A} \approx 1$ over a wide thickness range from 2 to above 100 QL ([108]Kim *et al.*, 2011;[15]Bansal *et al.*, 2012).

Introducing a variable-thickness In_2Se_3 layer in the center of a 60-QL Bi_2Se_3 slab induces an $\tilde{A} = 1 \rightarrow 2$ transition with increasing In_2Se_3 thickness, as shown in Figure 6-14b. Due to the large band gap of In_2Se_3, ~1.3 eV compared with ~0.3 eV for Bi_2Se_3, $\tilde{A}$ indicates the existence of

two channels even at a thickness of only ~3 QL; this indicates that the transport properties of the top and bottom Bi_2Se_3 layers are electrically isolated for barrier thicknesses of ~3 QL or more, indirectly implying the emergence of TSS at the interfaces. Figure 6-14c shows an extension of this experiment where another unit of Bi_2Se_3-In_2Se_3 has been added (Bi_2Se_3-In_2Se_3-Bi_2Se_3-In_2Se_3-Bi_2Se_3 with corresponding thicknesses of 30-20-30-*t*-30 QL); $\tilde{A}$ responds by transitioning from 2 → 3 with increasing In_2Se_3 thickness *t*, which confirms the counting nature of the channel interpretation of $\tilde{A}$ observed in Bi_2Se_3 with insulating bulk states ([2]Aguilar *et al.*, 2013;[106]Kim *et al.*, 2013;[32]Brahlek *et al.*, 2014).

To understand the microscopic origin of the emergence of the TSS near the transitional region, we carried out first-principles calculations based on density-functional theory (DFT). We first performed calculations on bulk Bi_2Se_3 and In_2Se_3, which were extended to Bi_2Se_3-In_2Se_3 supercells by allowing the construction of Wannierized effective Hamiltonians (see 6.3.3 for more details). Figure 6-14d shows the band gap at the Γ point near the interface of Bi_2Se_3-In_2Se_3, and confirms that the gapless nature of the TSS is attained with increasing In_2Se_3 thickness. The spatial electronic properties can be further seen by tracking the real space density of the states around the Fermi level (RDOS, see [217]Zhang *et al.*, 2012 and 6.3.3), as shown in Figure 6-14 e-h. This calculation shows that the RDOS increases near the Bi_2Se_3-In_2Se_3 interface even for a single QL of In_2Se_3, indicating a new state has begun to emerge. By 2-3 QL of In_2Se_3, the RDOS splits, peaking near the Bi_2Se_3-In_2Se_3 interfaces and diminishing near the center, which implies the formation of the TSS at the interfaces and development of an insulating bulk in the middle of In_2Se_3. The finite density of states in the In_2Se_3 region is due to the evanescent decay of the TSS wave functions into the In_2Se_3 layer, and extends around ~2 QL into the In_2Se_3 layer, which is better seen for the relatively thick 8QL In_2Se_3 in Figure 6-14h. This shows that the emergence of the TSS from the DFT calculation is consistent with the WAL $\tilde{A} = 1 \rightarrow 2$ transition, both of which suggest that the two Bi_2Se_3 layers are fully isolated beyond ~3 QL of In_2Se_3 in between.

6.3.3 Computational methods[12]

We study the tunneling between the topological surface states (TSS) in Bi_2Se_3-In_2Se_3-Bi_2Se_3 heterostructures based on density-functional theory (DFT) ([90]Hohenberg and Kohn, 1964;[112]Kohn and Sham, 1965). We first use the Quantum ESPRESSO package ([79]Giannozzi *et al.*, 2009) to perform calculations on bulk Bi_2Se_3 and In_2Se_3, with the generalized gradient approximation (GGA) ([152]Perdew *et al.*, 1996) to the exchange-correlation functional and fully relativistic norm-conserving pseudopotentials. The Brillouin zone (BZ) is sampled on an 8×8×8 Monkhorst-Pack [135]Monkhorst and Pack, 1976 *k* mesh, with an energy cutoff of 55 Ry (1 Ry ≈ 13.6 eV) for Bi_2Se_3 and 65 Ry for In_2Se_3. The first-principles output is fed into the Wannier90 package to produce Wannier functions (WFs) and to generate a realistic tight-binding (TB) model defined in the chosen Wannier basis ([138]Mostofi *et al.*, 2008;[132]Marzari *et al.*, 2012). 30 Wannier functions are constructed for Bi_2Se_3, including all the valence *p* orbitals, while four extra In 5*s* orbitals are included for In_2Se_3. Both models are constructed in such a way that they exactly reproduce the first-principles bandstructures within a certain energy range, spanning from 3 eV below to 3 eV above the Fermi level.

Once the first-principles TB model is obtained, we are ready to construct supercells including a Bi_2Se_3-In_2Se_3 interface. First, the Wannier-based model Hamiltonians for bulk Bi_2Se_3 and In_2Se_3, denoted as H_1 and H_2, are extrapolated to N_1 QL and N_2 QL slabs stacked in the [111] direction with open boundary conditions. These two isolated slabs are connected together in such a way that all the first-neighbor hoppings (by first-neighbor hopping, we actually refer to hopping terms between nearest-neighbor QLs) across the interface are taken as the average value of the corresponding hopping terms in the Bi_2Se_3 and In_2Se_3 bulk TB models. Then the periodic boundary condition is applied to the (N_1 + N_2)-QL slab to make it a periodic supercell. In our calculations, the total thickness of Bi_2Se_3 and In_2Se_3 is fixed to be N_1 + N_2 = 12 QLs, and the

[12] Calculations were carried out by Jainpeng Lui and David Vanderbilt, see Brahlek *et al*. 2014 for further details.

thickness of In_2Se_3 is varied from N_2 = 1 to 6 QLs (for the data shown in Figure 6-14h, $N_1 + N_2$ = 16 QLs with N_2 = 8). Working in the Wannier basis allows for the thickness of In_2Se_3 in the heterostructure to be highly tunable, and the computational cost is negligible compared with a fully self-consistent interface calculation.

In implementing this procedure, two issues need to be addressed. First, at the bulk level, we note that standard DFT tends to underestimate the energy of the In 5*s* level. Because the lowest conduction band and highest valence band of In_2Se_3 are dominated by In 5*s* and Se 4*p* orbitals respectively, DFT predicts a smaller band gap compared with experiment ([125]Liu and Vanderbilt, 2013). Here we adopt the corrective treatment described in [125]Liu and Vanderbilt, 2013 which involves applying a +0.79 eV rigid shift (taken from many-body GW calculations) to the four In 5*s* levels in the 34-band model for In_2Se_3, leaving all the other matrix elements unchanged.

Second, when constructing interface models, we have to take extra care of the band offset between the two bulk materials. Initially the zeroes of energy of the Wannierized tight-binding models for Bi_2Se_3 and In_2Se_3 are inherited from the respective bulk DFT bulk calculations, but as is well known, these are largely arbitrary, as they depend on irrelevant details such as the choice of pseudopotentials. We adopt the alignment method based on surface work functions ([173]Singh-Miller and Marzari, 2009) by carrying out self-consistent surface slab calculations on Bi_2Se_3 and In_2Se_3 slabs individually, from which we evaluate the difference between the average electrostatic potential energy deep in the bulk vs in the vacuum for each material. We do this by computing the macroscopic-averaged electrostatic potential $\overline{V}(z)$ from the microscopic potential $V(x,y,z)$ as

$$\overline{V}(z) = (cA)^{-1} \int_{z-c/2}^{z+c/2} dz \iint_A dxdy\, V(x,y,z),$$

where *c* and *A* are the cell height (size of a QL) and basal area respectively. For these calculations, a 3-QL slab is used, and slabs are separated from each other by a vacuum space of 2.9 nm. The

macroscopic averages of the electrostatic potentials are shown in the supplement of [31]Brahlek *et al.*, 2014. Note that due to the non-polar crystal structure and the homogeneous nature of the vacuum, $\overline{V}(z)$ remains constant both deep in the bulk and in vacuum. Aligning the vacuum levels, we conclude that the relative shift between the average electrostatic potential in bulk Bi_2Se_3 vs In_2Se_3 is $\Delta V = V_2 - V_1 = 1.776$ eV. Therefore, the arbitrariness in the energy zeroes can be removed by shifting all the Kohn-Sham eigenenergies of In_2Se_3 using $\tilde{E}_n(\mathbf{k}) = E_n(\mathbf{k}) + \Delta V$.

With the GW correction to In 5*s* levels and the shift ΔV on all the In_2Se_3 on-site energies, our interface model is ready to be used for the superlattice calculations. The eigenvalues are calculated in the (k_x, k_y) plane, setting k_z = 0. If the TSS do not interact, we expect to see a doubly degenerate gapless Dirac cone around Γ (k_x = 0, k_y = 0), but the energy spectrum should become gapped when a tunneling interaction is allowed. Therefore, the band gap at Γ, denoted as $\Delta(\Gamma)$, should provide a measure reflecting the tunneling amplitude between the TSSs. As shown in Figure 6-14d, $\Delta(\Gamma)$ is found to drop exponentially as the thickness of the In_2Se_3 layer increases. Setting 0.05 eV as a threshold below which the tunneling between the TSS is considered as negligible, the corresponding critical thickness t_c is about ~2.6 QLs, which agrees well with experimental data.

One may also be interested in the real-space distribution of the interface states, which can be easily calculated using the interface model described above. We define the following quantity as a weight of the real space density of the interface states around the Fermi level ([217]Zhang *et al.*, 2012): $\xi(z) = \sqrt{|\psi_\Gamma^v(z)|^2 + |\psi_\Gamma^c(z)|^2}$, where $\psi_\Gamma^v(z)$ and $\psi_\Gamma^c(z)$ are the components of the Bloch states at Γ projected onto the Wannier functions centered at z, and the superscripts v and c refer to the highest occupied and lowest unoccupied states respectively. In other words, if the Fermi level lies slightly above the conduction band minimum (CBM) at Γ, $\xi(z)^2$ measures the z-dependence of the charge density averaged over the x-y plane around the Fermi level. $\xi(z)$ is denoted as the real space density of the states (RDOS), as shown in Figure 6-14e-h.

6.3.4 Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 heterostructure and tunable quantum mechanical tunneling

It was previously shown that $(Bi_{1-x}In_x)_2Se_3$ undergoes compositional dependant topological and metal-to-insulator phase transitions: It first undergoes a TI to non-TI transition near $x \approx 3$ - 7% ([28]Brahlek *et al.*, 2012;[125]Liu and Vanderbilt, 2013;[202]Wu *et al.*, 2013), then transitions into a weakly insulating variable-range-hopping state near $x \approx 15\%$ and finally into a strong band insulator for $x > 25\%$ ([28]Brahlek *et al.*, 2012). Using this, we can explore the $\tilde{A} = 1 \rightarrow 2$ transition by systematically decreasing x in the $(Bi_{1-x}In_x)_2Se_3$ layer which lowers the potential height of the barrier. Plots of $\tilde{A}$ versus $(Bi_{1-x}In_x)_2Se_3$ thickness are presented in Figure 6-15a-c, which show that as x decreases from 40 to 30 and to 20%, the transition region is pushed to larger thickness.

These behaviors can be understood by applying an extremely simple quantum-mechanical tunneling model as follows. We think of the electrons in the individual Bi_2Se_3 layers as nearly free electrons due to the metallic states there, as illustrated in Figure 6-15d, and since the barrier layer of $(Bi_{1-x}In_x)_2Se_3$ is insulating, it can be thought of analogously as a potential barrier. The tunneling through this barrier will be described by a transmission coefficient T. If this tunneling is negligible, T will be zero, and the top and bottom metallic layers of Bi_2Se_3 will be decoupled from each other, leading to two WAL channels ($\tilde{A} = 2$). In the opposite limit, $T = 1$, the top and bottom Bi_2Se_3 layers will be strongly coupled and will act as a unit, leading to just one WAL channel ($\tilde{A} = 1$). The simplest way to connect these end points is through a linear approximation, $\tilde{A}(T) = 2 - T$ and using the standard quantum-mechanical form $T = 1/(1 + e^{2(t-t_0)/l_0})$ for free-electron tunneling through a square barrier, we obtain $\tilde{A}(t) = 2 - 1/(1 + e^{2(t-t_0)/l_0})$. This function, which is parameterized by the critical transition thickness t_0 and the transition width l_0, shows good agreement with the experimental data, which can be seen by comparing the experimental data to the fits in Figure 6-14b-c and Figure 6-15a-c.

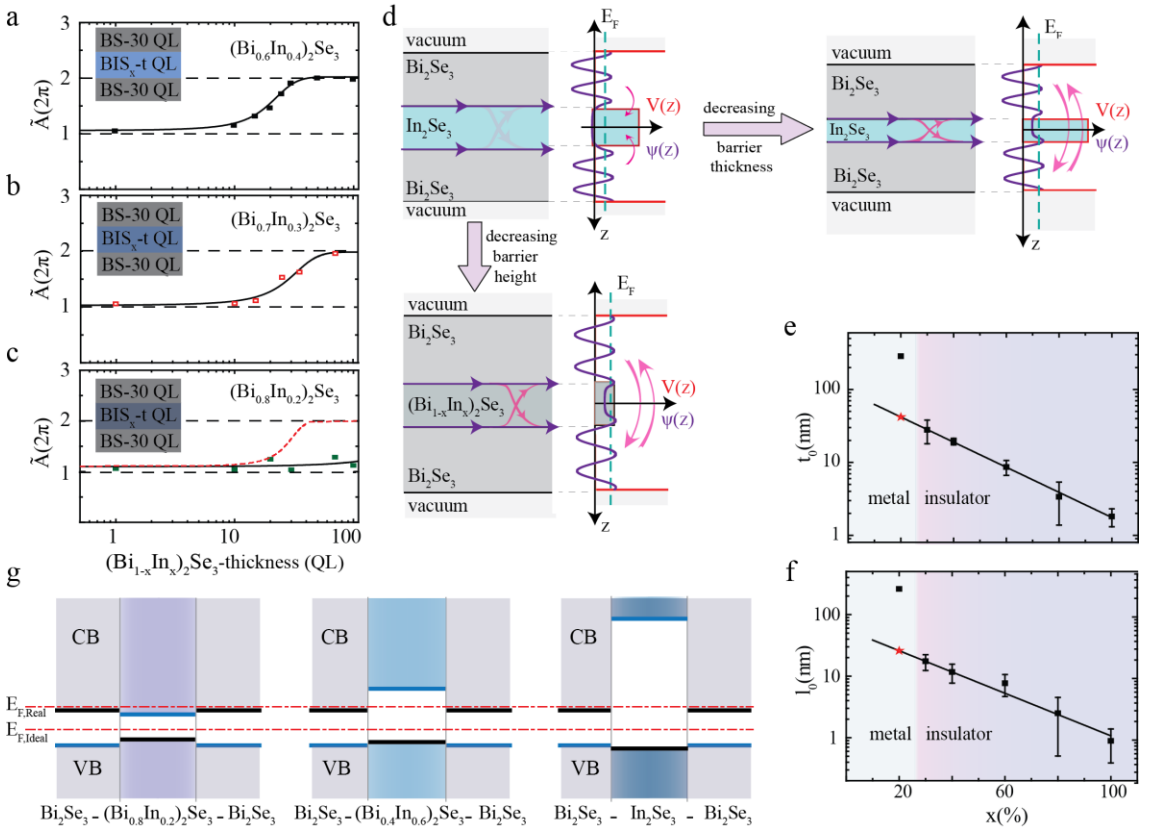


Figure 6-15 Weak anti-localization and a schematic of the quantum tunnelling model for Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$–Bi_2Se_3 films. a-c Thickness dependence of $\tilde{A}$ for Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 (30-*t*-30 QL) with x = 40, 30, and 20%, respectively. Symbols are experimental data, and the solid lines are the fits to the semi-empirical function (see text, and d). In **c** the solid line is the best fit to the data, while the red dotted line is the curve generated by extrapolating the exponential trend from the parameters shown as stars in **e-f**. **d** Schematic showing how the Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 structure can be thought of as a simple quantum-mechanical barrier separating the metallic states in Bi_2Se_3. Within this framework, it can be seen that tunneling increases as the barrier thickness is decreased or the height is lowered, thus accounting for the observed behavior of $\tilde{A}$. (**e-f**) Compositional dependence of the fitting parameters extracted from the curves in **a-c** and Figure 6-14b. Here t_0 (**e**) gives the critical thickness for the $\tilde{A}$ = 1→ 2 transition versus thickness, while l_0 (**f**) is the width of the transition region. The solid lines in **e** and **f** are exponential fits. The red stars are extrapolated parameters for x = 20%. Below $x \approx 25\%$ the data deviate from the exponential trend, consistent with an insulator-to-metal transition of the barrier layer. **g** Energy bands alignment of Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 for x = 20, 60 and 100% respectively. For $x \gtrsim 25\%$ the conduction band minimum of $(Bi_{1-x}In_x)_2Se_3$ is above the experimental (real) Fermi level $E_{F,Real}$, which makes the barrier layer insulating, while for $x \lesssim 25\%$ the conduction band minimum drops below $E_{F,Real}$, which makes the barrier metallic: with ideal Fermi levels ($E_{F,Ideal}$), the barrier should remain insulating even for x = 20%. (Adapted from [31]Brahlek *et al.*, 2014)

We have plotted the values of the fitting parameters t_0 and l_0 in Figure 6-15e-f, and it can be seen that for $x \gtrsim 30\%$, both t_0 and l_0 are almost exactly exponential functions of x. The

consistency of this trend implies that the barrier's electronic properties are described by the tunneling model in this composition range. However, the exponential dependence breaks down below x = 30%. To quantify this analytic behavior, we fit the experimental data to $t_0(x) = \tau e^{-x/x_0}$ and $l_0(x) = \lambda\, e^{-x/x_0}$, which results in $\tau \approx 90$ nm, $\lambda \approx 60$ nm, and both t_0 and l_0 consistently yield $x_0 \approx 25\%$, which coincides with the composition where the exponential trend breaks down. In order to see this more clearly, we generated the red dotted curve for x = 20% in Figure 6-15c by extrapolating the exponential behavior to x = 20% (see the red stars in Figure 6-15e-f); this curve clearly deviates from the experimental data for x = 20%. This indicates that below $x \approx 25\%$, the insulating behavior of $(Bi_{1-x}In_x)_2Se_3$ fully breaks down, giving way to a metallic regime. This composition coincides with the point where the band-insulating state dies out in homogeneous $(Bi_{1-x}In_x)_2Se_3$ films ([28]Brahlek *et al.*, 2012).

This transition can be understood from the band-alignment diagrams in Figure 6-15g generated by DFT calculations for Bi_2Se_3 and $(Bi_{1-x}In_x)_2Se_3$ with x = 20, 60 and 100%. What is noticeable is that the bottom of the conduction band for the barrier layer is clearly above that of Bi_2Se_3 for x = 60 and 100%, while for x = 20% the conduction band of barrier layer drops below that of Bi_2Se_3. If the Bi_2Se_3 crystals were perfect without any selenium vacancies, their Fermi levels ($E_{F,Ideal}$) would be at the Dirac point in the middle of the band gap. Then, the barrier would remain insulating all the way down and the band gap of the barrier would close at the topological phase transition near $x \approx$ 3-7%, so that x = 20% would behave as an insulating barrier like x = 60 and 100%. However, in practice, real Bi_2Se_3 layers have their Fermi levels ($E_{F,Real}$) slightly above the conduction band minimum, so that x = 20% behaves as a metal instead of an insulator. This analysis is fully consistent with the breakdown of the tunneling behavior for x below ~25% discussed above.

Figure 6-16a-c shows how $\tilde{A}$ changes with x for different barrier thicknesses (10, 20 and 30 QL). For each thickness, $\tilde{A}$ transitions from 1 to 2 with increasing x. If the empirical exponential relations hold between t_0 and x, and between l_0 and x, then there should be enough

information to generate a curve that fits these data points. Using $\tilde{A}(t) = 2 - \frac{1}{1+e^{2(t-t_0)/l_0}} \rightarrow \tilde{A}(t,x) = 2 - \frac{1}{1+e^{2(t-t_0(x))/l_0(x)}}$, $l_0(x) \approx 60e^{-x/25}$ and $t_0(x) \approx 90e^{-x/25}$ (the numerical values were obtained above) allows the generation of the curves for $\tilde{A}$ vs x (for greater than ~25%) with *no free parameters*. The solid red curves in Figure 6-16a-c agree well with the experimental data

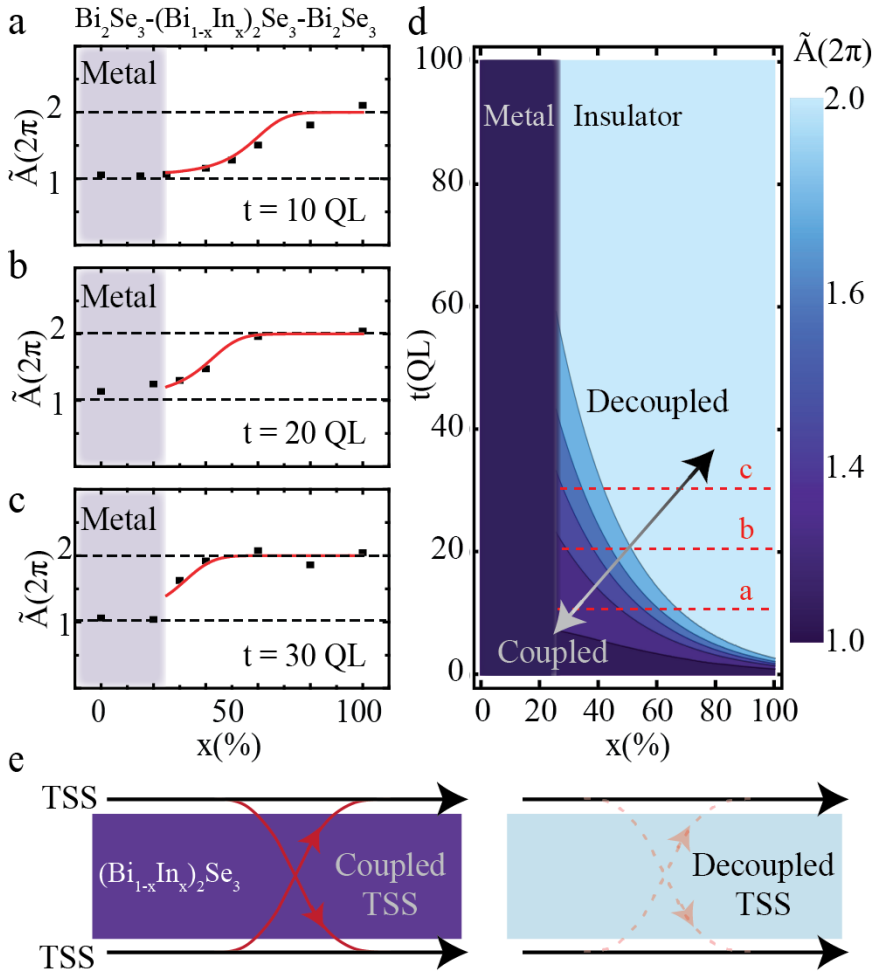


Figure 6-16 Weak anti-localization for Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 and a phase diagram for interlayer coupling. a-c Compositional dependence of $\tilde{A}$ with the $(Bi_{1-x}In_x)_2Se_3$ barrier thickness fixed at 10, 20, and 30 QL respectively. The solid red lines for each thickness were generated by using the semi-empirical function for $x > 25\%$, with the x dependence given by the analytic relations for l_0 and t_0 vs. x in Figure 6-15e-f, *with no free parameters*; this confirms that the exponential behavior holds for all $x > 25\%$. **d** Contour plot of the semi-empirical function for $\tilde{A}$ as a function of x and t for $x \gtrsim 25\%$; below $x \approx 25\%$, we took $\tilde{A}$ to be 1 due to the metallic nature of the barrier in this regime. The red dashed lines are cuts at constant thickness, which correspond to the solid red lines plotted in **a-c**. **e** Schematic showing the coupling ($\tilde{A} = 1$, left) and decoupling ($\tilde{A} = 2$, right) of the TSS across the $(Bi_{1-x}In_x)_2Se_3$ barrier layer. (Adapted from [31]Brahlek *et al.*, 2014)

for x > 25%, and this further confirms that the exponential relations for l_0 and t_0 hold for all x greater than ~25%.

Figure 6-16d summarizes the analytic behavior of $\tilde{A}$ as a function of thickness t and composition x, and Figure 6-16e sketches how the coupling between the TSS evolves with the barrier layer properties. The well-defined behavior of $\tilde{A}$ with both x and t suggests presence of a fundamental mechanism underlying here. However, the exponential dependence of t_0 and l_0 on x is anomalous because Vegard's law predicts that the lattice parameters and band gap of an alloy should, to first order, interpolate linearly between the end members ([59]Denton and Ashcroft, 1991), which would manifest itself as a power-law, not an exponential, change in both t_0 and l_0. Further studies will be required to resolve the origin behind this effect.

6.3.5 Concluding remarks

Massless, Diraclike topological surface states (TSS) form at the interface between materials that belong to different topological classes (topological insulator (TI), and non-TI), and have been realized in systems where a single slab of TI is surrounded by a non-TI (vacuum or any non-TI). We have extended the notion of a topological description of electronic materials and introduced the entirely new inverse topological insulator (ITI) system by sandwiching a layer of the tunable non-TI $(Bi_{1-x}In_x)_2Se_3$ (band gap 0 ~ 1.3 eV) between two layers of the prototypical TI Bi_2Se_3. In conventional single slab TIs, the bulk state is most often conducting, which allows deleterious coupling of the TSS. In contrast, the $(Bi_{1-x}In_x)_2Se_3$ barrier layer in the ITI system has been found to be a reliable and tunable insulator, which has allowed the bulk state between the TSS to be varied from metallic to a truly insulating state. On this new platform, we showed how the interfacial TSS emerge and become decoupled as the band gap and thickness of the non-TI layer was systematically changed..

By utilizing the ITI system, we have shown that the number of channels contributing to the WAL effect directly depends on the coupling strength through the barrier layer. When the

barrier layer is metallic, the two surfaces behave as one channel; this mimics the standard TI material, whose bulk states are metallic, thus causing only one WAL channel to appear. When the barrier layer is insulating, the two surfaces act either as one or two WAL channels depending on the barrier height and thickness. This exemplifies how the ideal TSS with fully insulating bulk states should behave and corroborates recent experiments on TI films showing the $\tilde{A} = 1 \rightarrow 2$ transition with film thickness ([106]Kim *et al.*, 2013;[32]Brahlek *et al.*, 2014). Beyond this, the ITI system provides a unique platform for probing buried topological surface states with tunable bulk properties. We anticipate that this ITI paradigm will provide new pathways to TI-based nanostructures, thereby initiating further discoveries and applications.

6.4 Conclusion

In this chapter we saw in two complementary routes that the TSS can be decoupled by tuning the electronic properties of the bulk state that connects them. In Cu-doped Bi_2Se_3, as the Cu concentration was increased, and the bulk channel became depleted, the TSS became decoupled. Similarly, in the inverse topological insulator system, as the In concentration was increased, and the $(Bi_{1-x}In_x)_2Se_3$ barrier layer became insulating, we saw a similar decoupling of the TSS, which was a function of both the barrier layer's concentration and thickness. An interested parallel can be seen in comparing the data shown in Figure 6-8c, and Figure 6-15a. The data in Figure 6-15a is for an ITI with $x = 40\%$; at this concentration the band gap is ~300 meV, which matches Bi_2Se_3, but is topological trivial. This $x = 40\%$ system can therefore be thought of as a non-topological version of Bi_2Se_3. It is remarkable that the data in Figure 6-8c, and Figure 6-15a are nearly identical, which is very strong evidence support the claim that Cu-doped Bi_2Se_3 is really a bulk insulating TI. The work with Cu-Bi_2Se_3 represents a proof-of-concept study that confirms that it is possible to make a true bulk insulating TI.

Chapter 7.

Transport properties of topological insulators: Band bending, metal-to-insulator transitions, and weak anti-localization

In this closing chapter we reanalyze some of the critical transport experiments and provide a coherent understanding of the current generation of topological insulators (TIs). Currently TI transport studies abound with widely varying claims of the surface and bulk states, often times contradicting each other, and a proper understanding of TI transport properties is lacking. According to the simple criteria given by Mott and Ioffe-Regel, even the best TIs are not true insulators in the Mott sense, and at best, are weakly-insulating bad metals. However, band bending effects contribute significantly to the TI transport properties including Shubnikov de-Haas oscillations, and we show that utilization of this band bending effect can lead to a Mott insulating bulk state in the thin regime. In addition, by reconsidering previous results on the weak anti-localization (WAL) effect with additional new data, we correct a misunderstanding in the literature and generate a coherent picture of the WAL effect in TIs.

7.1 Bulk metal-to-insulator transition[13]

A condition for a material to become topologically non-trivial is for certain energy bands to invert (other promising examples such as topological Kondo insulators have recently been predicted

[13] This chapter is based on work that is to be published in M. Brahlek *et al,* ArXiv **1408.1614,** (2014).

([62]Dzero *et al.*, 2010) with experiments in the initial phase ([107]Kim *et al.*, 2013;[141]Neupane *et al.*, 2013;[182]Thomas *et al.*, 2013;[200]Wolgast *et al.*, 2013;[218]Zhang *et al.*, 2013). Band inversion is most easily achieved in narrow gap semiconductors that are predicted to have non-inverted bands, and by incorporating realistic spin-orbit coupling the energy bands invert ([216]Zhang *et al.*, 2009). However, in real materials there are a finite number of thermodynamically induced atomic vacancies and anti-site defects that occur and act as charge defects, which shifts E_F. Since the band gap is narrow, E_F is easily moved into the bulk energy bands, and the material becomes metallic in the bulk. To get a feeling for how the bulk E_F changes as a function of bulk charge defect density, N_{BD}, we have plotted $E_F = \hbar^2/(2m^*)(3\pi^2 N_{BD})^{2/3}$ in Figure 7-1, where we assumed an isotropic Fermi surface with the effective mass given in terms of the electron mass to be $m^* \approx 0.15m_e$, a typical value for Bi_2Se_3 and Bi_2Te_3 ([35]Butch *et al.*, 2010). Here and for all the following discussions, unless explicitly specified, the temperature is assumed to be absolute zero, which is a valid assumption for the commonly measured transport properties of TIs. What can be seen from this is that E_F is pinned within a few meV of the conduction band minimum (or valence band maximum in *p*-type materials) when the carrier density drops below ~10^{17}-10^{18} /cm^3. E_F can only become unpinned from the conduction band minimum and fall into the band gap if a metal-to-insulator transition occurs with reduced carrier densities; these transitions can occur via two different processes, Mott transition and Ioffe-Regel criterion.

When a dopant atom is embedded in an insulator, it forms a rescaled Coulomb potential, and this leads to hydrogenlike bound states, each with an effective Bohr radius given by $a_B = \varepsilon(m_e/m^*) \times 0.5$ Å, where ε is the dielectric constant, m_e is the electron mass, m^* is the effective mass, and 0.5 Å is the free space Bohr radius. In a crystal, as the dopant density is increased, the mean separation of the dopant atoms decreases and eventually becomes comparable to the effective Bohr radius. When this happens, electrons that were bound to the dopant sites can become delocalized and freely move to neighboring sites, which gives rise to the insulator becoming a metal. The Mott criterion predicts that this happens when the dopant density passes

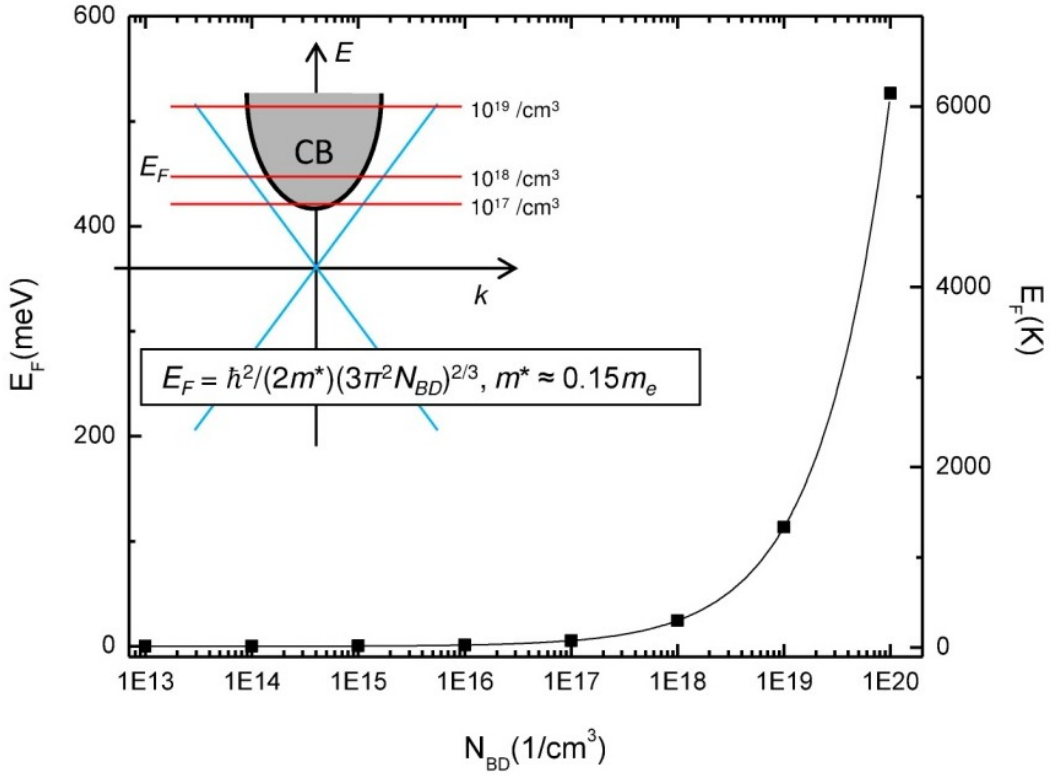


Figure 7-1 The calculated Fermi energy, E_F, relative to the conduction band minimum, as a function of buk carrier density. Below a bulk defect density of ~10^{17} /cm^3, E_F is within a few meV of the conduction minumum. However, based on the Mott criterion, E_F will stay pinned in the bottom of the conduction band until the bulk defect density drops below the critical value of ~3×10^{14} /cm^3. (Adapted from [30]Brahlek *et al.*, 2014)

the critical density given by $a_B N_M^{1/3} \approx 0.26$ (see [66]Edwards and Sienko, 1982). Rearranging this, we then get the critical value to be $N_M \approx (m^*/(\varepsilon m_e))^3 \times (1.4 \times 10^{23}$ /cm$^3) \approx 3 \times 10^{14}$ /cm^3 for both Bi_2Se_3 and Bi_2Te_3; this was found by using $\varepsilon \approx 110$ and $m^* \approx 0.15m_e$ ([35]Butch *et al.*, 2010). $N_M \approx 3 \times 10^{14}$ /cm^3 is extremely small compared to the corresponding numbers for common semiconductors such as Si where $N_M \approx 10^{18}$ /cm^3. In fact, for Bi_2Se_3 the lowest dopant density achieved thus far is ~10^{16} /cm^3 (see [35]Butch *et al.*, 2010); it was found that these crystals were dominated by bulk carriers, which can be rationalized by considering the Mott criterion. In standard semiconductors such as Si, the bonding is very strong, and therefore a defect density of ~10^{14} /cm^3 is achievable in the cleanest materials. However in current TI materials, the bonding is

much weaker, and therefore such a low defect density may be thermodynamically impossible to achieve either in intrinsic or in compensation-doped materials.

The Ioffe-Regel criterion explains how a metal transitions into an insulator with increased disorder and reduced carrier density. It states that the electrons in a metal become localized when the mean free path drops below the Fermi wave length ([99]Ioffe and Regel, 1960). In other words, the system remains metallic only if $k_F l > 1$, where k_F is the Fermi wave vector and l is the mean free path, which can easily be calculated based on typical values measured by transport. Assuming a 3D isotropic Fermi surface $k_F = (3\pi^2 N_{BD})^{1/3}$ and $l = (\hbar\mu/e)(3\pi^2 N_{BD})^{1/3}$ where μ is the electron mobility, e is the electron charge, and N_{BD} is the volume carrier density, which is related to the Hall carrier density by $N_{BD} = n_{Hall}/t$ with t being the crystal's thickness. By setting $k_F l = 1$, this can be rearranged into a function of critical mobility, μ_{IR}, vs N_{BD}; the result is $\mu_{IR} = (e/\hbar)(3\pi^2 N_{BD})^{-2/3}$. What this says is that if $\mu < \mu_{IR}$ then the system can transition into a weakly insulating phase even if the bulk defect density exceeds the critical value predicted by the Mott criterion. Such an insulating phase usually can be described by a variable range hopping mechanism, in which the conduction electrons typically hop a certain distance before becoming localized ([139]Mott and Davis, 1979). However, these types of insulating phases tend to remain quite conducting at any finite temperatures, only truly insulating in the limit of zero temperature, and are thus often called a bad metal.

The results for the Mott and the Ioffe-Regel criteria are summarized in Figure 7-2, where data have been compiled from single crystals and thin films of Bi_2Se_3, Bi_2Te_3, Bi_2Te_2Se as well as these compounds that have been intentionally charge-doped to compensate for the native carriers ([92]Hor *et al.*, 2009;[3]Analytis *et al.*, 2010;[4]Analytis *et al.*, 2010;[35]Butch *et al.*, 2010;[160]Ren *et al.*, 2010;[175]Steinberg *et al.*, 2010;[161]Ren *et al.*, 2011;[195]Wang *et al.*, 2011;[15]Bansal *et al.*, 2012;[91]Hong *et al.*, 2012;[162]Ren *et al.*, 2012;[206]Xiong *et al.*, 2012;[205]Xiong *et al.*, 2013). In Figure 7-2, the vertical line corresponds to the Mott criterion, where everything to the left is in a strong insulating state. The diagonal region roughly corresponds to the Ioffe-Regel criterion of $k_F l \sim 1$.

The region below this, but to the right of the Mott criterion, is in a bad metal phase. The region above the Ioffe-Regel criterion and to the right of the Mott criterion is a good metal. What is clear from this plot is that no samples come close to the critical density predicted by the Mott criterion. Currently, the samples that are closest to insulating are Bi_2Te_2Se bulk crystals, which show insulatorlike behavior in resistance vs temperature but have significant bulk carriers and are highly disordered with extremely small mobilities. These samples become more insulating at the cost of decreasing mobility. Optimally, one would want to move left horizontally, while maintaining the mobility.

Last of all, the Mott criterion gives critical insight in the position of the Fermi level deep

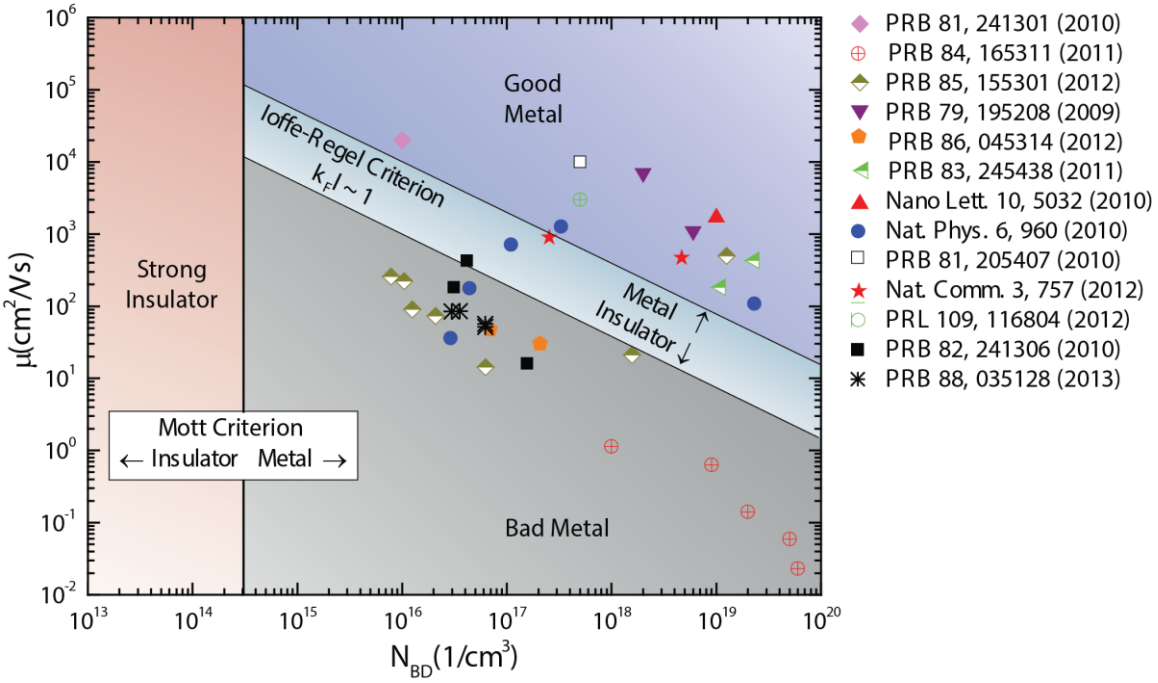


Figure 7-2 The Mott and Ioffe-Regel criteria showing when a metal-to-insulator transition will occur based on the mobility and the bulk carrier density. The vertical line corresponds to a critical bulk defect density of $N_M \approx 3 \times 10^{14}$ /cm^3, while the diagonal lines correspond to $k_F l \sim 0.3$-3: carrier densities and mobilities of representative Bi_2Se_3 and Bi_2Te_3 family of materials are shown together for comparison. What can be seen from this plot is that no known TI samples are truly an insulator in the Mott sense, and at best they are bad metals, exhibiting weakly insulating temperature dependence. Data are obtained from [92]Hor *et al.*, 2009;[3]Analytis *et al.*, 2010;[4]Analytis *et al.*, 2010;[35]Butch *et al.*, 2010;[160]Ren *et al.*, 2010;[175]Steinberg *et al.*, 2010;[161]Ren *et al.*, 2011;[195]Wang *et al.*, 2011;[15]Bansal *et al.*, 2012;[91]Hong *et al.*, 2012;[162]Ren *et al.*, 2012;[206]Xiong *et al.*, 2012;[205]Xiong *et al.*, 2013. (Adapted from [30]Brahlek *et al.*, 2014)

inside the bulk. Since all the transport data give a bulk carrier density that exceeds the Mott criterion by at least a few orders of magnitude, it is then implied that E_F deep in the bulk must be at the conduction band minimum. This does not contradict the Ioffe-Regel criterion which says the bulk electrons can be localized even if E_F is at the conduction band minimum, simply by making the system more disordered. We will see next that this is an important point when band bending effects are incorporated.

7.2 Band bending effects

Band bending occurs wherever there is a mismatch in Fermi levels, for example, at the interface between two semiconductors. In any case, for band bending to occur there needs to be charge transfer to equilibrate the Fermi level everywhere. This charge transfer then creates an electric field, and the associated potential shifts the bands in the vicinity of the charge transfer. In Bi_2Se_3 the band bending is due to a large mismatch between the surface and bulk Fermi levels, and then upon equilibrating, the bulk band shifts near the surfaces. This shift has been found in some samples to be upward giving rise to a depletion of carriers near the interface ([3]Analytis *et al.*, 2010;[4]Analytis *et al.*, 2010 [32]Brahlek *et al.*, 2014), while in other samples it has been found to

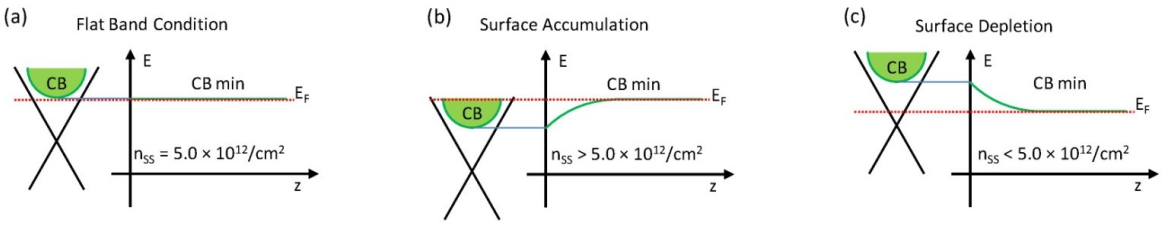


Figure 7-3 Band bending direction depending on the surface carrier density. a If the surface E_F is exactly at the conduction band minimum, then the surface carrier density for one TSS is equal to ~5.0 × 10^{12} /cm^2 and there is no band bending. **b** If the surface carrier density is above ~5.0 × 10^{12} /cm^2, then the bands must bend downward, and there is an accumulation of carriers at the surface. **c** If the surface carrier density is below ~5.0 × 10^{12} /cm^2, then the bands bend upward, and there is a depletion of bulk carriers near the surface. (Adapted from [30]Brahlek *et al.*, 2014)

shift downward, which gives rise to an accumulation region near the surface ([22]Bianchi *et al.*, 2010;[15]Bansal *et al.*, 2012). This downward bending also gives rise to a non-topological two dimensional electron gas (2DEG) because the downward bending creates a quantum well that quantizes the bulk bands near the surface ([110]King *et al.*, 2008;[22]Bianchi *et al.*, 2010;[18]Benia *et al.*, 2011;[25]Bianchi *et al.*, 2011;[109]King *et al.*, 2011;[23]Bianchi *et al.*, 2012;[24]Bianchi *et al.*, 2012).

The band bending direction is strongly dependent on both the surface and bulk defect levels. Given that the Mott criterion fixes E_F to be at the conduction band minimum and knowing the surface state carrier density, together these fix the band bending as follows ([32]Brahlek *et al.*, 2014). First, assume that at the surface E_F barely touches the CB minimum. Then assuming a Fermi velocity ($v_F \approx 4.0 \times 10^5$ m/s) and the energy of the CB minimum relative to the Dirac point ($E_{DP} \approx 210$ meV, see [95]Hsieh *et al.*, 2009;[22]Bianchi *et al.*, 2010), the surface carrier density can be calculated for one surface to be $n_{SS} \approx 5.0 \times 10^{12}$ /cm^2. Since E_F at the surface and in the bulk is at the CB minimum, the bands must be flat, and band bending is negligible (see Figure 7-3a). Therefore, $n_{SS} \approx 5.0 \times 10^{12}$ /cm^2 gives the flat band condition. Now, if n_{SS} increases above ~5.0×10^{12} /cm^2 then the conduction band bends downward near the surface, giving rise to an accumulation of carriers near the surface (see Figure 7-3b). Alternatively, if n_{SS} is less than ~5.0×10^{12} /cm^2 then the bands bend upward close to the surface, which implies that there should be a depletion of bulk carriers near the surface (see Figure 7-3c).

There is a large difference between upward and downward band bending. In the case of upward band bending with bulk carriers less than ~10^{18} /cm^3, the Poisson equation is sufficient to quantify the effect (see below). In the case of downward band bending, the Thomas-Fermi approximation is valid because the system near the surface is much more like a typical metal. For Thomas-Fermi screening, the length scale over which the screening occurs is given by $l_0 \approx ((\varepsilon_0 \pi^2 \hbar^2)/(k_F m^* e^2))^{1/2} \approx 6$ Å, where $k_F \approx 0.07$ /Å is the Fermi wave vector as typically measured by ARPES, $m^* \approx 0.15 m_e$ is the effective mass in terms of the electron mass m_e, e is the electron charge, and ε_0 is the free space dielectric constant. This shows that for downward band bending

the screening length is $\lesssim$ 1 nm. In contrast, for the case of upward band bending, the length scale is much larger, as will be shown below. Although downward band bending is more common with many TI materials, we will show below that if upward band bending can be achieved in thin film TIs, truly insulating bulk state could be implemented even in current generation TIs, whose bulk states are always metallic under the Mott criterion.

In general, band bending should be treated in a self consistent manner with a numerical solution to the coupled Poisson-Schrodinger equations. However, the case of upward band bending, shown in Figure 7-3c, can be described to first order by considering only the Poisson equation. This formalism is only valid in the limit of when E_F is close to the CB minimum and when the bulk defect density, N_{BD}, is uniform. From Figure 7-2, we can see that E_F begins to deviate from the CB bottom significantly when N_{BD} is larger than ~10^{18} /cm^3, and therefore, in the following we assume $N_{BD} < 10^{18}$ /cm^3; high quality TI samples (bulk crystals and thin films) with minimal defect densities satisfy this condition. The Poisson equation is

(1) $$\nabla^2 V(z) = -\frac{e^2 N_{BD}}{\varepsilon\varepsilon_0}$$

where $V(z)$ is the potential as a function of z, the depth from the surface of the TI. Then, by assuming N_{BD} is roughly constant and the potential is fully screened at a depth of z_D, the depletion depth, the solution for one surface is given by

(2) $$V(z) = \begin{cases} \frac{e^2 N_{BD}}{2\varepsilon\varepsilon_0}(z - z_D)^2 & z < z_D \\ 0 & z > z_D \end{cases}.$$

At $z = 0$, the CB minimum is at its largest value (see Figure 7-4a), $V(z = 0) = \Delta V$, which gives the following equation

(3) $$\Delta V = \frac{e^2 {z_D}^2 N_{BD}}{2\varepsilon\varepsilon_0}.$$

This is summarized in Figure 7-4a. The band bending can be related back to the various charge densities by the following. There are two intrinsic defect levels, the bulk defect density, N_{BD}, and the surface defect density, n_{SD}, which together set the initial surface and bulk Fermi levels, $E_{F,SD}$ and $E_{F,BD}$ (see Figure 7-4b) prior to equilibrium being established. For equilibrium to then be established, the Fermi level on the surface needs to match that of the bulk, and therefore a certain amount of charge, n_{CT}, needs to be transferred from the bulk onto the surface (see Figure 7-4c). Since we assumed a constant bulk defect density, n_{CT} is simply given by

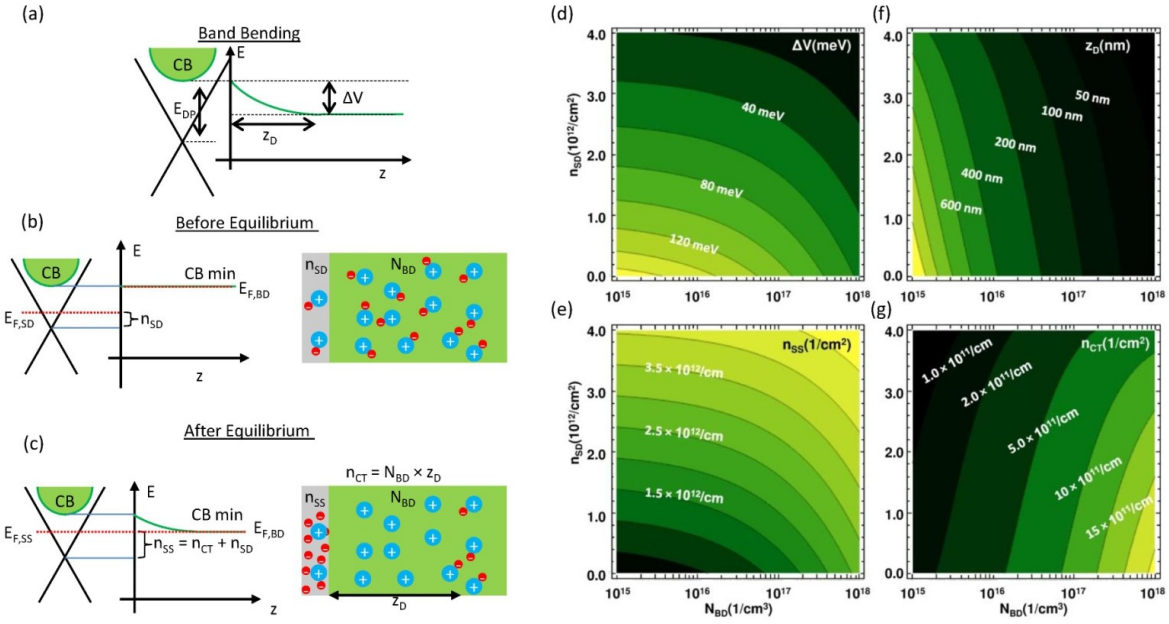


Figure 7-4 Band bending parameters for upward band bending. a A schematic of the band bending parameters for the case of upward band bending in the infinite thickness limit. ΔV gives the magnitude of the band bending, z_D is the depletion length, and E_{DP} is the energy seperation of conduction band minimum and the Dirac point. **b** Charge distribution and Fermi levels before equilibrium, determined by local native dopants. **c** Charge distribution and Fermi levels after equilibrium is reached. The charge transfer is given by $n_{CT} = N_{BD} \times z_D$, which causes a potential build-up, and therefore the bands bend. The Fermi levels are equilibrated everywhere due to the combined effect of the charge transfer, and the shift in potential. This can be made quantative in **d-g** by solving the Poisson equation which determines ΔV, z_D, n_{SS}, and n_{CT} as a function of intrinsic defect levels of the surface, n_{SD} (vertical axes), and bulk, N_{BD} (horizontal axes). (Adapted from [30]Brahlek *et al.*, 2014)

(4) $$n_{CT} = N_{BD} \times z_D.$$

So now, the total amount of charge on the surface is

(5) $$n_{SS} = n_{CT} + n_{SD}.$$

Lastly, we can get one more equation by looking at Figure 7-4a. By assuming E_F is at the conduction band bottom deep inside the bulk, we can see that the surface Fermi level, relative to the Dirac point, is related to the level of band bending, ΔV, by

(6) $$E_{F,SS} = \hbar v_F \sqrt{4\pi n_{SS}} = E_{DP} - \Delta V,$$

where we used the dispersion relation for a Dirac state $E_{F,SS} = \hbar v_F k_F = \hbar v_F (4\pi n_{SS})^{1/2}$, where $n_{SS} = k_F^2/(4\pi)$, and $E_{DP} \approx 210$ meV is the separation between the conduction band minimum and the Dirac point. Combining Eq. (3 - 6) gives a set of 4 equations with 6 unknowns, which allows us to solve the system by letting ΔV, z_D, n_{SS}, and n_{CT} be the dependent variables, which depend on n_{SD} and N_{BD}. We proceed this way because experimentally n_{SD} and N_{BD} are set by growth conditions and the subsequent exposure of the surface to the experimental environment (vacuum, atmosphere, etc). An analytic solution to this set of equations exists. However, it is too complicated to present explicitly, and the clearest way to understand the physics of the solution is the contour plots shown in Figure 7-4d-g as a function of n_{SD} and N_{BD}.

The contour plots in Figure 7-4d-g contain a large amount of information and can be understood by the following. The vertical axes are the intrinsic defect density (n_{SD}) on the surface, while the horizontal axes are the intrinsic bulk defect density (N_{BD}). In Figure 7-4d ΔV is plotted. For fixed N_{BD}, as n_{SD} increases ΔV decreases accordingly because the n_{SD} sets the initial E_F on the surface, and the closer the surface E_F is to the bulk conduction band minimum, the less the bands need to bend. In Figure 7-4f z_D is plotted. As N_{BD} decreases, z_D increases, which is expected because it takes a larger distance of the depletion zone to raise the surface E_F to that of the bulk

state; in a similar way z_D decreases for increasing n_{SD}. Figure 7-4e-g show the total surface states carrier density, n_{SS}, and the total charge transfer, n_{CT}. This shows that for low bulk carrier density ($\lesssim 5\times 10^{16}$ /cm^3) the carrier density on the surface is dominated by the surface defect level, and only when the bulk carrier density exceeds ~5×10^{16} /cm^3, is there significant charge transfer to the surface.

Figure 7-4d-g allows us to understand what happens when Bi_2Se_3 and Bi_2Te_3 are exposed to atmosphere or controlled surface contamination in ARPES experiments ([3]Analytis *et al.*, 2010;[25]Bianchi *et al.*, 2011). The long term effect of atmospheric exposure has been shown to increase the carrier density ([29]Brahlek *et al.*, 2011). What can be seen from Figure 7-4 is that if the crystals start with upward band bending and low carrier densities on the surface and in the bulk, then as time increases the band bending flattens out, and upward band bending will eventually give way to downward bending; such an effect has been linked to charge doping by water vapor in atmospheric gases ([4]Analytis *et al.*, 2010;[29]Brahlek *et al.*, 2011;[116]Kong *et al.*, 2011;[2]Aguilar *et al.*, 2013). This change of band bending character will be accompanied with the development of quantum well states on the surfaces, which are in parallel with the TSS.

The level and direction of band bending is crucial to observing Shubnikov de-Haas (SdH) oscillations from the TSS. There has been a puzzling difference between two sets of Bi_2Se_3 based samples reported in J. G Analytic *et al* ([4]Analytis *et al.*, 2010) and N. P. Butch *et al* ([35]Butch *et al.*, 2010). In the former study, SdH oscillations observed at high magnetic field were consistent with the TSS, while in the latter study no surface SdH oscillations were observed. The puzzle is that Butch *et al*'s samples were an order of magnitude lower in bulk defect density and still did not show surface SdH oscillations. The main difference between the samples in the two studies was the addition of a small amount of Sb in the Analytis *et al*'s. The addition of the Sb makes the initial state of the samples start with upward band bending ([3]Analytis *et al.*, 2010), while pure Bi_2Se_3 starts in an initial state with downward band bending. As shown in Figure 7-3b and c, if the band bending is downward, the size of the Fermi surface of the TSS is much larger than if the

bands bend upward. This affects the SdH oscillations from the TSS in two ways. First, the period of SdH oscillations depends inversely on the cross sectional area of the Fermi surface; consequently a large Fermi surface translates into a small period of oscillation, and thus it is harder to observe the SdH oscillations from the surfaces with downward band bending. The second complication is that with downward band bending, quantum well states, which are bulk states confined near the surface, develop in the conduction band ([25]Bianchi *et al.*, 2011). These states can have higher mobility than the topological surface states and since their Fermi surfaces are smaller than the TSS, the carrier density tends to be lower than that of the TSS. These together with the natural 2D nature of the quantum well states can make their SdH oscillations more prominent and easily confusable with the SdH oscillations that emanate from the TSS.

Since the amplitude of the SdH oscillations is exponentially sensitive to the mobility, the mobility dependence on band bending is another factor affecting observation of the SdH oscillations from the TSS. Band bending can affect the mobility from the TSS in two ways, both of which can be seen by considering the Drude mobility, $\mu = e\tau/m^*$, where τ is the carrier relaxation time and m^* is the effective mass (more specifically m^* is the cyclotron mass for a Dirac band). The first is that τ is related to the change in band bending direction. As mentioned above, for upward band bending to be present, the surface defect density needs to be sufficiently low, and if the surface defect density is too large then the band bending direction is downward. Because more defects imply shorter relaxation times, this directly implies that $\tau_{upward} > \tau_{downward}$. Secondly, for a Dirac band m^* is proportional to the Fermi level measured from the Dirac point, $m^* = E_F/v_F^2$, whereas for a parabolic bulk band m^* is constant. Because the Fermi level is lower for upward band bending than for downward bending while the Fermi velocity remains almost constant, this then implies $m^*_{upward} < m^*_{downward}$. Both effects lead to $\mu_{upward} > \mu_{downward}$. Altogether, this helps explain why Butch *et al*, whose samples had downward band bending, saw no sign of SdH oscillations from the TSS while those samples measured by Analytist *et al* that exhibited upward band bending revealed SdH oscillations consistent with the TSS.

Finally, from this analysis of upward band bending we can get some insight into the regime where the thickness is comparable with the depletion length. For the above analysis, it was assumed that the thickness of the crystal was much larger than the depletion length as shown in Figure 7-5a. On the other hand, when the thickness drops below twice the infinite limit depletion length, the solution to the Poisson equation (Eq. (2)) becomes

$$V(z) = \frac{e^2 N_{BD}}{2\varepsilon\varepsilon_0}(z - t/2)^2 \ (0 < z < t), \tag{7}$$

where t is the thickness of the crystal. Since in this finite thickness limit $t < 2 \times z_D$, the charge transferred from the bulk state to the surface is now limited to

$$n_{CT,F} = N_{BD} \times t/2 < N_{BD} \times z_D \, . \tag{8}$$

This implies then that the bulk is now completely depleted because all the free electrons are transferred out of the bulk onto the surface. As mentioned above, before equilibrium is established, the Fermi levels on the surface and in the bulk are set by the defect levels that reside in those regions, and the charge transfer raises the surface Fermi level up to that of the bulk. Now, if the available charge in the bulk, $N_{BD} \times t/2$, is insufficient to raise the Fermi level on the surface to the bottom of the conduction band, then all the free electrons are transferred out of the bulk and the bulk Fermi level simply drops below the conduction band minimum to establish equilibrium. In fact, this is not in contradiction with the Mott criterion, because the Mott criterion states that the Fermi level must be pinned in the conduction band only until the bulk carrier density remains above N_M (~3×10^{14} /cm^3) as shown in section II.

Following the analysis given above we can actually calculate the gap, Δ, between the conduction band minimum and the Fermi level as a function of film thickness, t. The main modification is to Eq. (6), which becomes

$$E_{F,SS} = \hbar v_F \sqrt{4\pi n_{ss}} = E_{DP} - \Delta V - \Delta(t), \tag{9}$$

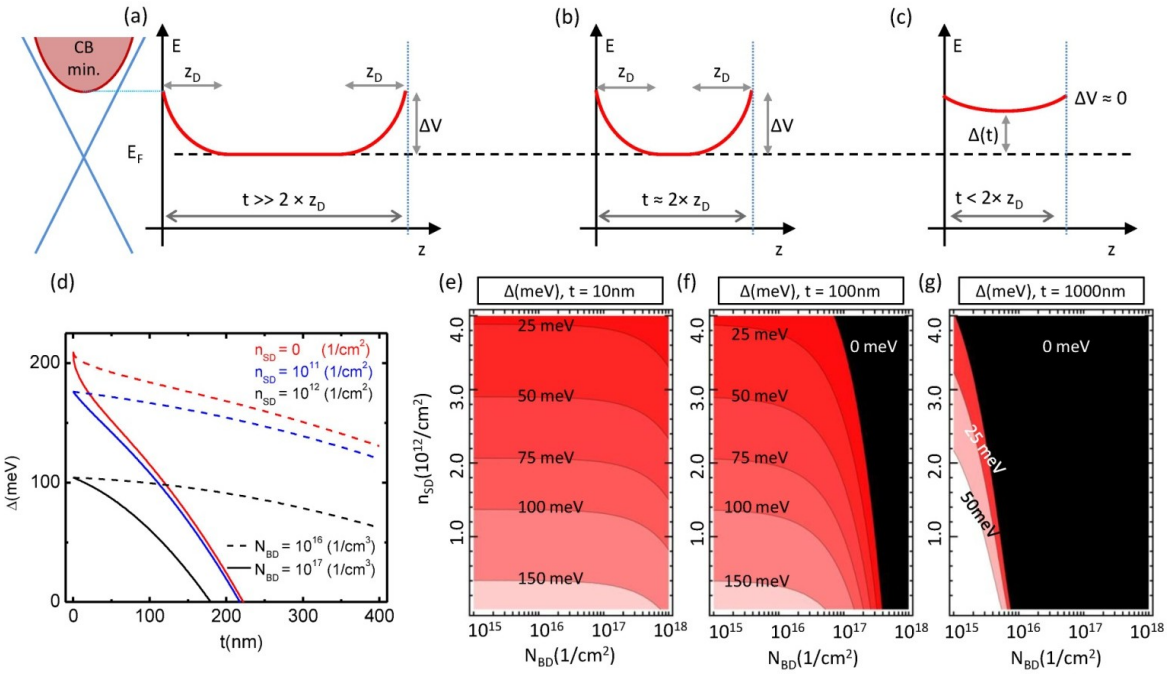


Figure 7-5 Upward band bending vs thickness. a In the bulk limit when the thickness of the TI, t, is much larger than the depletion length, z_D, the Fermi level sits at the bottom of the conduction band. **b** Band bending in the intermediate thickness (t ≈ $2z_D$). **c** Once the thickness is reduced below $2z_D$, the Fermi level drops below the bottom of the conduction band by an amount $\Delta(t)$. **d** This shows how Δ(t) changes for various values of the bulk and surface defect densities according to Eq. (10). The dotted lines are for $N_{BD} = 10^{16}$ /cm^3, and the solid lines are for $N_{BD} = 10^{17}$ /cm^3, and at t = 0 nm the upper two curves are for $n_{SD} = 0$ /cm^2, the next two are for $n_{SD} = 10^{11}$ /cm^2 and the lowest two are for $n_{SD} = 10^{12}$ /cm^2. **e-g** shows how $\Delta(t)$ changes for various N_{BD}, and n_{SD} while the thickness is fixed. (e) shows $\Delta(t)$ in the ultra thin regime (t = 10 nm), where it remains finite over the entire range of N_{BD} and n_{SD} considered. On the other hand, in the bulk limit shown in (g) at t = 1000 nm, $\Delta(t)$ is zero over most of the range of N_{BD} and n_{SD}. (Adapted from [30]Brahlek *et al.*, 2014)

where the Fermi level has settled at $\Delta(t)$ below the conduction band minimum (see Figure 7-5c). As was done with Eq. (6), Eq. (9) can be expanded in terms of the thickness and the bulk and surface defect densities, and then we can solve for $\Delta(t)$ only in terms of these parameters. This procedure yields

$$\Delta(t) = E_{DP} - \frac{e^2 N_{BD}}{2\varepsilon\varepsilon_0}\left(\frac{t}{2}\right)^2 - \hbar v_F\sqrt{4\pi(n_{SD} - N_{BD}t/2)}, \tag{10}$$

where, in the finite thickness limit, we used $\Delta V = (e^2 N_{BD}/(2\varepsilon\varepsilon_0))(t/2)^2$ and $n_{SS} = n_{SD} + n_{CT,F} = n_{SD} + N_{BD}(t/2)$. Eq. (10) shows that if the bulk and surface defect levels are low enough, there exists a

critical thickness for which the Fermi level drops below the conduction band minimum. This can be easily seen by plotting Eq. (10) for fixed defect densities against thickness in Figure 7-5d or for fixed thickness in the contour plots in Figure 7-5e-g. Figure 7-5e shows that in the ultra thin limit (t < 10 nm) the Fermi level should be below the conduction band minimum, so long as the surface carrier density is below the critical value where upward band bending gives way to downward band bending (as mentioned previously, this occurs when $n_{SS} > 5.0 \times 10^{12}$ /cm^2 for one surface). In contrast, in the bulk limit (t > 1000 nm) in Figure 7-5f, $\Delta(t)$ remains zero for experimentally achievable bulk defect density levels, which are ~10^{16} /cm^3 or higher, implying that Fermi levels deep inside the bulk still remain pinned to the bottom of the conduction band in bulk crystals. Altogether, this suggests that even if the bulk charge defect density is higher than the Mott criterion of ~3×10^{14} /cm^3, as far as the surface charge density is kept below the critical value (~5×10^{12} /cm^2) necessary for the upward band bending, making the samples thin will force the bulk Fermi level to drop below the conduction band minimum and lead to a truly insulating bulk state. In other words, through thin film engineering, it should be possible to achieve Mott insulating bulk state even in current generation TIs. The main experimental obstacle is that most TI materials such as Bi_2Se_3 have strong tendency toward downward, instead of upward, band bending. However, it was recently demonstrated that Cu doping in Bi_2Se_3 thin films can help stabilize the upward band bending and it does lead to the bulk insulating state as expected here ([32]Brahlek *et al.*, 2014).

7.3 Weak anti-localization

Weak anti-localization (WAL) and weak localization (WL) are effects in transport measurements where the conductance is enhanced or suppressed due to coherent back scattering, and are usually seen as a sharp change in the conductance with the application of a small magnetic field. The origin of this effect is due to coherent back scattering of the electrons along time reversible paths (see [130]MacKinnon, 2001 and Chapter 6.1 for more basic details regarding WAL and WL).

The WAL and WL effects are typically analyzed using the HLN formalism, which was developed in the 1980s for 2D systems ([89]Hikami *et al.*, 1980). The following formula was derived to quantify the correction to the conductance: $\Delta G(B) = \tilde{A}e^2/(2\pi h)[ln(B_\phi/B) - \Psi(1/2 + B_\phi/B)]$, where h is Planck's constant, $\tilde{A}$ is a parameter related to the number of 2D conductive channels (1 for each channel), B_ϕ is the de-phasing field, and $\Psi(x)$ is the digamma function. The two independent parameters are $\tilde{A}$ and B_ϕ (in the original HLN formalism, the prefactor was in units of $1/(2\pi)$ per channel, but for simplicity, we incorporate this 2π term into the prefactor so that $\tilde{A} = 1$, instead of $1/(2\pi)$, implies a single channel).

B_ϕ is a magnetic field that roughly gives the width of the WAL cusp in the ΔG vs B plot. To get a physical understanding, B_ϕ can be related to a length scale by $B_\phi = \hbar/(4el_\phi^2)$, where l_ϕ is the de-phasing length. Roughly speaking all time reversible paths that are within the de-phasing length will contribute to the WAL effect. Such a parameter can then be thought of like a mean-free path, which is dependent on the microscopic details such as disorder. In contrast, the other parameter, $\tilde{A}$, is less sensitive to such details; $\tilde{A}$ is related to the number of channels in the system. For an ideal TI with surface states on the top and bottom, each should carry a channel number of 1, and if B_ϕ is similar, then the conductance should add, which would give an effective $\tilde{A} = 1 + 1 = 2$. However, in all known TIs, there is a metallic bulk channel, which complicates the analysis of this parameter. Despite this, $\tilde{A}$ provides vital clues to help unravel transport measurements.

In thin-film samples of Bi_2Se_3 $\tilde{A}$ is not the ideal value of 2, but rather close to half of this at $\tilde{A} \approx 1$ [43]Chen *et al.*, 2010;[108]Kim *et al.*, 2011;[129]Lu and Shen, 2011;[176]Steinberg *et al.*, 2011;[15]Bansal *et al.*, 2012;[180]Taskin *et al.*, 2012 (see Figure 7-6). This suggests that there is a conducting bulk state in parallel with the two topological surface states. Electrons on the surface can couple to this bulk state, and together the entire film acts effectively as one channel, and therefore $\tilde{A} \approx 1$; this has been confirmed by electro-static gating experiments ([42]Chen *et al.*, 2011;[176]Steinberg *et al.*, 2011). As the gate-voltage depleted the bulk state near one surface, $\tilde{A}$

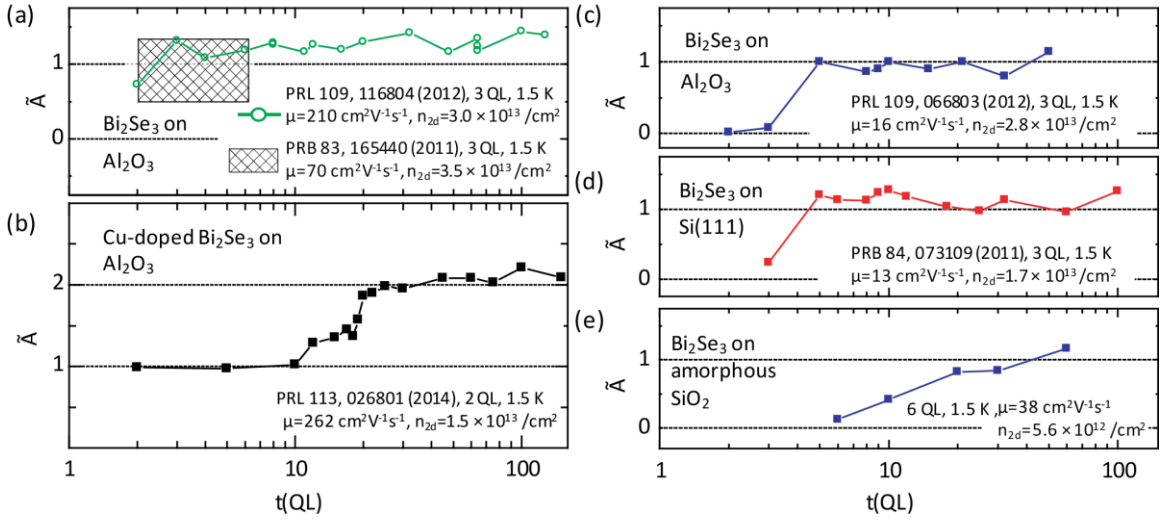


Figure 7-6 The $\tilde{A}$ parameter in the HLN fomula vs Bi_2Se_3 film thickness for various samples with different qualities. Mobilities and carrier densities of 2, 3 or 6 QL thick sample of each set are shown for quality comparison. **a** High mobility, bulk metallic Bi_2Se_3 films from [15]Bansal *et al.*, 2012 and [126]Liu *et al.*, 2011 showing that $\tilde{A}$ is ~1 over the entire thickness range without any 1-to-0 transition down to thickness of 2 QL. **b** High mobility, bulk insulating Cu-doped Bi_2Se_3 films from [32]Brahlek *et al.*, 2014 showing that $\tilde{A}$ is ~1 for 2-10 QL and transitions to ~2 for films thicker than ~20 QL: note absence of the 1-to-0 transition for samples down to 2 QL. (**c-e**) For lower mobility bulk metallic Bi_2Se_3 films grown on Al_2O_3 (**c**, [180]Taskin *et al.*, 2012), Si(111) (**d**, [108]Kim *et al.*, 2011), and amorphous SiO_2 **e**, $\tilde{A}$ is nearly constant at ~1 for thick films, but shows a transition from 1 to 0 with decreasing thickness. Note that the WAL 1-to-0 transition strongly depends on the mobilities of the samples. This proves that the metal-insulator transition and the associated WAL 1-to-0 transition observed when the surface Fermi level is far from the Dirac point **c-e** is due to enhanced disorder effect in the ultrathin regime and not due to a surface-hybridization gap opening at the Dirac point. (Adapted from [30]Brahlek *et al.*, 2014)

gradually increased from ~1 towards ~2, implying that two surface states mediated by the conducting bulk behave as one 2D channel, whereas two isolated surface states appear as two 2D channels in the WAL effect.

This $\tilde{A}$ parameter is therefore a sensitive indicator to distinguish between samples that have conducting bulk versus those with the bulk state depleted. Two experiments have made the best case for a truly depleted bulk state, one by surface-doping thin Bi_2Se_3 flakes in conjunction with electro-static gating ([106]Kim *et al.*, 2013), and the other achieved with compensation-doped Bi_2Se_3 thin-films ([32]Brahlek *et al.*, 2014). The likely explanation as to how these experiments got around the Mott criteria is the previously mentioned use of band bending in the ultra thin limit (*t*

$< 2 \times z_D$). The main result of both experiments is the observation that the $\tilde{A}$ parameter transitions from 1 to 2 as the bulk state transforms from metallic to insulating state. The former experiment by D. Kim *et al* ([106]Kim *et al.*, 2013) showed that while $\tilde{A} \approx 1$ regardless of thickness when the carrier density is high ($\gtrsim 5 \times 10^{12}$ /cm^2 per surface), as the carriers were depleted through gating $\tilde{A}$ transitioned from 1 to 2 for films thicker than ~10 QL but remained at ~1 for thinner samples. The latter experiment by M. Brahlek *et al* ([32]Brahlek *et al.*, 2014) provided a more extensive thickness dependence of the WAL parameter in Cu-doped ultralow carrier density (~2.5×10^{12} /cm^2 per surface) Bi_2Se_3 films. Like D. Kim *et al*, Brahlek *et al*, first showed that at fixed thickness (20 QL) as the bulk state was depleted by Cu doping, $\tilde{A}$ smoothly increased from 1 to 2. They further showed that when the bulk state is made insulating, for films below 10 QL, $\tilde{A}$ was exactly quantized at 1, but as the films were made thicker than 10 QL, the $\tilde{A}$ parameter increased from 1 towards 2, and from 20 QL to 150 QL $\tilde{A}$ remained quantized at 2: these data are reproduced in Figure 7-6b. Together, both studies suggest that when the coupling between the top and bottom surfaces is suppressed, the two surfaces contribute individually to the total channel number of two for $\tilde{A}$, but if coupling is allowed between the two surfaces either through bulk conduction or through tunneling in the ultrathin regime, they contribute as a whole one channel to $\tilde{A}$.

Lastly, WAL sometimes shows a transition from $\tilde{A} = 1$ to 0 in the ultrathin films ([108]Kim *et al.*, 2011;[180]Taskin *et al.*, 2012). Considering that a hybridization gap is formed at the Dirac point when Bi_2Se_3 films get thinner than ~6 QL according to an ARPES study ([219]Zhang *et al.*, 2010), such a transition is naturally expected when the Fermi level is tuned into this gap and was experimentally confirmed by D. Kim *et al* ([106]Kim *et al.*, 2013). However, similar transitions were observed even in films with their Fermi levels far above the Dirac point: first by us in Y.-S. Kim *et al* ([108]Kim *et al.*, 2011) and later by A. A. Taskin *et al* ([180]Taskin *et al.*, 2012). Although these studies report that the transition occurs at just below 5 QL, which happens to be close to the gap opening thickness as observed by the ARPES study, whether this is related to the gap opening at

the Dirac point is questionable considering that their Fermi levels are far (~300-500 meV) from the Dirac point; disorder and surface scattering in the ultrathin regime can easily drive otherwise metallic films into insulating state causing such a 1-to-0 WAL transition. Although Y.-S. Kim *et al* did not interpret it in terms of the gap opening process, A. A. Taskin *et al* interpreted a similar 1-to-0 transition as due to surface-hybridization gap opening at the Dirac point even though their surface Fermi levels were far from the gap. If this argument were correct, such a transition should have been observed at similar thicknesses in Bi_2Se_3 films regardless of their Fermi levels.

However, N. Bansal *et al* ([15]Bansal *et al.*, 2012), M. Liu *et al* ([126]Liu *et al.*, 2011), and M. Brahlek *et al* ([32]Brahlek *et al.*, 2014) showed that Bi_2Se_3 films with much higher mobilities did not exhibit such a 1-to-0 transition in $\tilde{A}$, which remained close to 1 all the way down to 2 QL. In Figure 7-6, we compare the published thickness-dependent WAL channel data from Bi_2Se_3 films and a new data set on amorphous SiO_2: the mobilities at ~1.5 K for their 2, 3 or 6 QL samples are also shown for quantifying the level of disorder in the ultrathin regime. The comparison shows that there is no universality in the 1-to-0 WAL transition and it strongly depends on, among other things, the mobility of the samples. While the higher mobility samples by Bansal *et al*, Liu *et al*, or Brahlek *et al* did not exhibit any transition all the way down to 2 QL, the lower mobility samples by Y. Kim *et al* and Taskin *et al* exhibited the 1-to-0 transition between 3-5 QL, and those on amorphous SiO_2, which had the lowest mobilities in the thin regime among the compared results, showed the transition at an even higher thickness. Furthermore, all the lower mobility samples exhibited clearly insulating temperature dependence at low temperatures, whereas the higher mobility samples remained pretty much metallic down to the lowest measurement temperatures. This mobility dependent metal-to-insulator transition can be partly understood with the above-discussed Ioffe-Regel criterion in the 3D context. It is also well known that 2D metallic films turn insulating once their sheet resistance becomes similar to or larger than the quantum resistance ($h/e^2 \approx 26$ kΩ) as the thickness is reduced, due to enhanced surface scattering and other disorder effects ([108]Kim *et al.*, 2011,[86]Haviland *et al.*, 1989). Altogether, this

comparison clearly shows that the WAL 1-to-0 transition observed in the low mobility samples whose surface Fermi levels are far from the surface-hybridization gap at the Dirac point is due to enhanced disorder in the ultrathin regime and not due to the gap opening at the Dirac point.

7.4 Conclusion

Transport measurements on TI materials are often more complex than non-TI materials due to the various topological and non-topological transport channels that are simultaneously present. To address these issues, we have presented a thorough analysis that takes into account the effects of TSS and non-topological bulk channels. By applying the simple Mott and Ioffe-Regel criteria that govern metal-to-insulator transitions, we have shown that it may be thermodynamically impossible to achieve a bulk insulating TI at finite temperatures, because the Mott criterion predicts that such crystals are required to have a bulk defect density below ~10^{14} /cm^3 (~1 part per billion). However, if upward band bending effects can be stabilized in these materials and the samples are made thinner than twice the depletion length, it may be possible to overcome the Mott criterion and achieve a true bulk insulating TI. The upward band bending seems to be also essential for clear observation of TSS SdH oscillations. Lastly, we have shown that the number of channels that appear in the WAL effect provides significant evidence pertaining to the electronic properties of the bulk as well as the surface states. It can be either one, two or zero depending on the thickness of the sample, metallicity of the bulk states, location of the surface Fermi level, and the level of disorder. Specifically, the WAL channel can transition from one to zero in ultrathin TI films not only when the Fermi level tunes into the hybridization gap at the Dirac point but also when disorder-driven metal-to-insulator transition occurs. Overall, with this comprehensive examination of the transport properties of the current TI materials, we have cleared up some misunderstandings in the literature and demonstrated that it is possible to achieve a true bulk insulating TI, which will hopefully stimulate and guide future studies in the field of TIs.

Chapter 8.

Epilogue

When this work started in 2010, the field of topological insulators (TIs) was in its infancy, and researchers around the world were scrambling to examine this new class of materials. Among this whirlwind, the first order of business was to establish a starting point by synthesizing and measuring the first generation of TI samples. From this initial data, the two conjoined directives mentioned in Chapter 1 were established: (1) Seeking to create a true topological insulator, which is free of any bulk conduction, make key material improvements, and (2) engineer topological material systems that reveal the novel physical properties of the topological surface states. Throughout this work we have shown it is possible to achieve success along both these directions; however, success is only in the sense that the door leading to further possibilities has begun to be opened, rather than a final closing word.

8.1 Review of results

In chapter 5 we showed that despite the existence of the metallic bulk channel, the $(Bi_{1-x}In_x)_2Se_3$ system clearly shows novel evidence of the decay of the topological phase into a trivial state as x was increased ([28]Brahlek *et al.*, 2012;[202]Wu *et al.*, 2013). For $x < 3 - 6\%$ the phase remained topological, while above this, the topological phase transitioned into a non-topological normal phase through a process of band inversion where the TSS extend into the bulk and finally merge

upon the point of the full transition. When x was tuned beyond the point of the topological phase transition, it remained in a normal metal phase until x exceeded ~20 – 30%, at which point, it transitions into a bulk-insulating state. A future research directive is to reperform this experiment, but on the films grown on the In_2Se_3 buffer layer, which reduces the bulk conduction and places the Fermi level closer to the Dirac point ([113]Koirala *et al.*, 2014). With this more ideal system, the physics of the TI phase transition, which is most dramatically apparent when the Fermi energy nears the Dirac point, will be more evident in transport experiments. Further, in this system it may be possible to control the Fermi level with an external gate voltage and push E_F through the Dirac point; a point that was not possible in the previous experiments due to the large number of carriers.

Chapter 6 and 7 were devoted to the emergence of the decoupled TSS in both the ideal TI system, and the entirely new inverse TI (ITI) phase. The former was realized in Cu-doped Bi_2Se_3, where it was shown the majority of the conductivity is through the topological surface states. Moreover, when the film's thickness was increased from 10 QL to beyond 20 QL, the surface states became decoupled as measured by the weak anti-localization effect ([32]Brahlek *et al.*, 2014). This work was built upon by creating an entirely new topological system (ITI) created out of two layers of the TI Bi_2Se_3, which were separated by a barrier layer of the tunable non-TI $(Bi_{1-x}In_x)_2Se_3$ (x > 20%). In this system it was seen that for x = 100%, the surface states emerged when the barrier layer was ~2-3 QL, and were already in the decoupled regime. Further, as x was decreased, the potential barrier decreased, and the decoupling was pushed to larger thicknesses, which was found to follow a very simple quantum tunneling model ([31]Brahlek *et al.*, 2014).

One of the biggest challenges in TI transport experiments has been to confirm that the transport signature emanates from the TSS. The fact that the two completely separate systems of Cu-doped Bi_2Se_3 and the Bi_2Se_3-$(Bi_{1-x}In_x)_2Se_3$-Bi_2Se_3 ITI system showed nominally identical behavior gives credence to the fact that the bulk state in Cu-doped Bi_2Se_3 has been substantially suppressed. This can be seen by comparing the reliably insulating $(Bi_{1-x}In_x)_2Se_3$ barrier layer with

$x \approx 40\%$, which has nearly the same band gap as Bi_2Se_3, to that of optimally Cu doped Bi_2Se_3. The decoupling behavior for this system exactly matches that seen in Cu-doped Bi_2Se_3.

8.2 Future research directions

Topological materials are teeming with future research possibilities. On the materials improvement side, these materials are in need of a reliable capping material to protect against reaction with atmospheric gases (see Chapter 4, and [29]Brahlek *et al.*, 2011). When deposited on top of Bi_2Se_3, most materials have been found to pull Se out of the surface layer, resulting in an increase of the carrier density. Therefore, a capping layer scheme must leave the Bi_2Se_3 unharmed. Such a capping layer should be completely insulating, and, therefore, has a secondary potential to act as a gate dielectric, so Fermi level tuning can be done simultaneously by an external gate voltage, which would further enable the exploration of the physics near the Dirac point.

On the physics side, there is a need to integrate TIs with ferromagnetic or superconducting materials. Recently, the quantum anomalous hall effect was observed in Cr-doped $(Bi_{1-x}Sb_x)_2Te_3$ ([39]Chang *et al.*, 2013;[143]Oh, 2013), which relied upon opening a gap at the Dirac point. The gap was a result of the time reversal symmetry breaking of the ferromagnetic phase that set in at a Curie temperature of ~15 K. This low Curie temperature resulted in a small gap, and therefore the quantum anomalous hall phase was only observed when the films were cooled to the miniscule temperature of ~0.030 K. By integrating Bi_2Se_3 with a strong ferromagnetic material, with Curie temperature in excess of several hundred Kelvin, direct exchanging coupling across the interface has the possibility to open up a substantially larger gap, thereby pushing the quantum anomalous phase to tens, if not hundreds of Kelvin. This would require a ferromagnetic material that was electrically insulating, with the ability to be grown on top, or underneath Bi_2Se_3, while maintaining a pristine interface.

A ferromagnetic-TI heterostructure would also be critical to exploit the strong spin-momentum locking of the TSS in practical devices. Electronic devices rely on the ability to

manipulate the charge degree of freedom to act as a binary *on-off* switch. Still in its infancy, spin based electronics, or spintronics, uses the spin of the electrons to perform such logic operation, or to act as data storage ([222]Žutić *et al.*, 2004). Relying on spin-momentum locking, TI materials offer a promising route to the realization of a practical spintronic devices, and could solve one of the major problems, which is spin-dephasing across the typical micron-scale devices; since electric charge can be robust to dephasing on length scales in excess of meters, the locking of spin to the motion of the charge can enable the spin direction to be preserved on length scales larger than the dephasing length. Secondly, the spin momentum locking offers a new route to create logic devices by coupling the resistance to an external ferromagnetic switch, which has the potential to operate as a spin-transistor ([212]Yokoyama and Murakami, 2014). All of these applications rely on the ability to seamlessly integrate a ferromagnetic material for use as either a spin-injector, or a spin-manipulator ([121]Li *et al.*, 2014;[134]Mellnik *et al.*, 2014). In this way, integration with ferromagnetic materials is a key to realizing many practical applications of TI materials.

Another promising direction is the search for new topological phases. Of particular interest has been the yet unsuccessful search for a topological superconductor. Theory predicts that when an *s*-wave superconductor is coupled with a strong spin-orbit coupled semiconductor (e.g. a TI such as Bi_2Se_3), then, due to the superconducting proximity gap, the TI takes on addition non-trivial properties ([158]Qi and Zhang, 2011;[120]Leijnse and Flensberg, 2012). As with the ferromagnetic material, this comes down to finding a superconductor that can be seamlessly integrated with Bi_2Se_3, so the superconducting electrons can penetrate through the interface without being stripped of their phase information. Of particular promise is the Van der Waals layered superconductor $NbSe_2$, which has already been shown to accept Bi_2Se_3 as a top layer ([196]Wang *et al.*, 2012). Achieving such a structure would be a first step to realizing novel excitations such as Majorana fermions, whose intrinsic, non-local nature give them distinct

advantages in applications such as quantum computing ([120]Leijnse and Flensberg, 2012, [177]Stern and Lindner, 2013).

Lastly, with the ever evolving understanding of topological order, one must look beyond the layered chalcogenide TIs, to new and more robust materials. Discussed at length throughout this dissertation was the need to find material improvements to get around the bulk metallic problem due to the Se vacancies in Bi_2Se_3. Therefore, a future goal is to find new topological materials that are naturally robust against the intrinsic defects that are ever-present in all materials. Of particular interest is to find a non-trivial phase in oxide based materials ([168]Rüegg and Fiete, 2011;[211]Yang *et al.*, 2011;[210]Yan *et al.*, 2013), which, when combined as heterostructures, could utilize the huge number of electronic phases known to exist in oxides, and this will open up a vast toolbox for engineering novel thin film phases in TI materials by molecular beam epitaxy.

Chapter 9.

9.1 List of abbreviations

(0D)	0-dimensional
(1D)	1-dimensional
(2D)	2-dimensional
(3D)	3-dimensional
(2DEG)	2-dimension electron gas
(AC)	Alternating current
(AFM)	Atomic force microscopy
(AMI)	American Magnetics, Inc
(ARPES)	Angle resolved photo-emission spectroscopy
(ARS)	Advanced Research Systems, Inc
(BBC)	Bulk boundary correspondence
(BI)	Band insulator
(BZ)	Brillouin zone
(CB)	Conduction band
(CBM)	Conduction band minimum
(CCW)	Counterclockwise
(CF)	ConFlat
(CVD)	Chemical vapor deposition
(CW)	Clockwise
(DC)	Direct current
(DFT)	Density-functional theory
(E_F)	Fermi energy (or Fermi level)

(GGA)	Generalized gradient approximation
(HLN)	Hikami-Larkin-Nagaoka
(ITI)	Inverse topological insulator
(LHe)	Liquid helium
(LK)	Lifshitz-Kosevich
(LL)	Landau level
(LN)	Liquid nitrogen
(MBE)	Molecular beam epitaxy
(NM)	Normal metal
(PLD)	Pulsed laser deposition
(PVD)	Physical vapor deposition
(QCM)	Quartz-crystal microbalance
(QHS)	Quantum hall state
(QL)	Quintuple layer (1 QL ≈ 1 nm)
(QSHS)	Quantum spin hall state
(RBS)	Rutherford backscattering spectroscopy
(RDOS)	Real space density of the states
(RF)	Radio frequency
(RHEED)	Reflection high energy electron diffraction
(SdH)	Shubnikov-de Hass
(SOC)	Spin orbit coupling
(SS)	Surface states
(STM)	Scanning tunneling microscopy
(TB)	Tight binding
(TCI)	Topological crystalline insulator
(TDS)	Topological Dirac semimetals
(TDTD)	Time-domain terahertz spectroscopy
(TEM)	Transmission electron microscopy
(THz)	Terahertz (10^{12} Hz)

(TI)	Topological insulator
(TM)	Topological metal
(TRS)	Time reversal symmetry
(TSS)	Topological surface states
(UHV)	Ultra-high vacuum
(VB)	Valence band
(VBM)	Valence band maximum
(VI)	Variable range hopping insulator
(VRH)	Variable range hopping
(WAL)	Weak anti-localization
(WF)	Wannier function
(WL)	Weak localization
(XPS)	X-ray photoemission spectroscopy
(XRD)	X-ray diffraction

9.2 References

(1) A. F. Ioffe, a. A. R. Regel., Prog. Semicond. **4**, (1960)

(2) Aguilar, R. V., L. Wu, A. V. Stier, et al., Journal of Applied Physics **113**, 153702 (2013)

(3) Analytis, J. G., J.-H. Chu, Y. Chen, et al., Physical Review B **81**, 205407 (2010)

(4) Analytis, J. G., R. D. McDonald, S. C. Riggs, et al., Nature Physics **6**, 960-964 (2010)

(5) Anderson, P. W., Science **177**, 393-396 (1972)

(6) Anderson, P. W., P. Langacker and A. K. Mann, Physics Today **43**, 117 (1990)

(7) Ando, Y., Journal of the Physical Society of Japan **82**, 102001 (2013)

(8) ARS, http://www.arscryo.com/TechNotes/CryocoolerPrincipleofOperation.html. (2014)

(9) Arthur, J. R., Journal of Applied Physics **39**, 4032-4034 (1968)

(10) Ashcroft, N. W. and N. D. Mermin, *Solid State Physics*, New York, Holt, Rinehart, and Winston.(1976)

(11) Baker, T. V., J. Chem. Soc. Trans. **89**, (1906)

(12) Baker, T. V., Mineral. Mag. **14**, (1907)

(13) Baker, T. V., Z. Kristallogr. **45**, (1908)

(14) Balshaw, N. H., *Practical Cryogenics*, Witney, Oxon, Oxford Instruments, Scientific Research Division.(1996)

(15) Bansal, N., Y. S. Kim, M. Brahlek, et al., Physical Review Letters **109**, 116804 (2012)

(16) Bansal, N., Y. S. Kim, E. Edrey, et al., Thin Solid Films **520**, 224-229 (2011)

(17) Bartynski, R. A., Surface Science Lecture Notes (unpublished) (2013)

(18) Benia, H. M., C. Lin, K. Kern and C. R. Ast, Physical Review Letters **107**, 177602 (2011)

(19) Bernevig, B. A., T. L. Hughes and S.-C. Zhang, Science **314**, 1757-1761 (2006)

(20) Bernevig, B. A. and S.-C. Zhang, Physical Review Letters **96**, 106802 (2006)

(21) Berry, M. V., Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **392**, 45-57 (1984)

(22) Bianchi, M., D. Guan, S. Bao, et al., Nature Commun. **1**, 128 (2010)

(23) Bianchi, M., R. C. Hatch, D. Guan, et al., Semiconductor Science and Technology **27**, 124001 (2012)

(24) Bianchi, M., R. C. Hatch, Z. Li, et al., ACS Nano **6**, 7009-7015 (2012)

(25) Bianchi, M., R. C. Hatch, J. Mi, et al., Physical Review Letters **107**, 086802 (2011)

(26) Bird, R. B., W. E. Stewart and E. N. Lightfoot, *Transport Phenomena*, New York, John Wiley & Sons, Inc.(2002)

(27) Borisenko, S., Q. Gibson, D. Evtushinsky, et al., Physical Review Letters **113**, 027603 (2014)

(28) Brahlek, M., N. Bansal, N. Koirala, et al., Physical Review Letters **109**, 186403 (2012)

(29) Brahlek, M., Y. S. Kim, N. Bansal, et al., Applied Physics Letters **99**, 012109 (2011)

(30) Brahlek, M., N. Koirala, N. Bansal and S. Oh, ArXiv **1408.1614**, (2014)

(31) Brahlek, M., N. Koirala, J. Liu, et al., Submitted (2014)

(32) Brahlek, M., N. Koirala, M. Salehi, et al., Physical Review Letters **113**, 026801 (2014)

(33) Brill, B. and M. Heiblum, Physical Review B **54**, R17280-R17283 (1996)

(34) Burkov, A. A. and L. Balents, Physical Review Letters **107**, 127205 (2011)

(35) Butch, N. P., K. Kirshenbaum, P. Syers, et al., Physical Review B **81**, 241301 (2010)

(36) Cao, H., J. Tian, I. Miotkowski, et al., Physical Review Letters **108**, 216803 (2012)

(37) Cha, J. J., D. Kong, S.-S. Hong, et al., Nano Letters **12**, 1107-1111 (2012)

(38) Chambers, S. A., Surface Science Reports **39**, 105-180 (2000)

(39) Chang, C.-Z., J. Zhang, X. Feng, et al., Science **340**, 167-170 (2013)

(40) Checkelsky, J. G., Y. S. Hor, R. J. Cava and N. P. Ong, Physical Review Letters **106**, 196801 (2011)

(41) Checkelsky, J. G., Y. S. Hor, M. H. Liu, et al., Physical Review Letters **103**, 246601 (2009)

(42) Chen, J., X. Y. He, K. H. Wu, et al., Physical Review B **83**, 241304 (2011)

(43) Chen, J., H. J. Qin, F. Yang, et al., Physical Review Letters **105**, 176602 (2010)

(44) Chen, Y. L., J. G. Analytis, J.-H. Chu, et al., Science **325**, 178-181 (2009)

(45) Chen, Y. L., J.-H. Chu, J. G. Analytis, et al., Science **329**, 659-662 (2010)

(46) Cho, A. Y., Surface Science **17**, 494-503 (1969)

(47) Cho, A. Y., Journal of Applied Physics **41**, 2780-2786 (1970)

(48) Cho, A. Y., Journal of Applied Physics **42**, 2074-2081 (1971)

(49) Close, F., *Partilce Physics: a Very Short Introduction*, Oxford, Oxford University Press.(2004)

(50) Cohen, P. I. and A. Ichimiya, *Reflection high-Energy Electron Diffraction* Cambridge, Cambridge University Press.(2004)

(51) Cole, M. W., Reviews of Modern Physics **46**, 451-464 (1974)

(52) Coleman, P., Annales Henri Poincaré **4**, 559-580 (2003)

(53) Coleman, P., *Handbook of Magnetism and Advanced Magnetic Materials*, New York, Wiley.(2007)

(54) Constable, G. and B. Somerville, *A Century of Innovation: Twenty Engineering Achievements that Transformed our Lives*, The National Academies Press.(2003)

(55) Cornils, M. and O. Paul, Journal of Applied Physics **104**, (2008)

(56) Culcer, D., Physica E: Low-dimensional Systems and Nanostructures **44**, 860-884 (2012)

(57) Davies, J. H., *The Physics of Low Dimensional Semiconductors*, New York, Cambridge University Press.(1998)

(58) Davisson, E. W. and L. H. Germer, Physical Review **30**, (1927)

(59) Denton, A. R. and N. W. Ashcroft, Physical Review A **43**, 3161-3164 (1991)

(60) Drude, P., Annalen der Physik **306**, 566-613 (1900)

(61) Drude, P., Annalen der Physik **308**, 369-402 (1900)

(62) Dzero, M., K. Sun, V. Galitski and P. Coleman, Physical Review Letters **104**, 106408 (2010)

(63) Dziawa, P., B. J. Kowalski, K. Dybko, et al., Nature Materials **11**, 1023-1027 (2012)

(64) Eason, R., *Pulsed Laser Deposition of Thin Films Applications-Led Growth of Functional Materials*, Hoboken, John Wiley & Sons.(2007)

(65) Eckstein, J. N., Nature Materials **6**, 473-474 (2007)

(66) Edwards, P. P. and M. J. Sienko, Accounts of Chemical Research **15**, 87-93 (1982)

(67) Ewald, P., Acta Crystallographica Section A **25**, 103-108 (1969)

(68) Farrow, R. F. C., Ed. *Molecular Beam Epitaxy Applications to Key Materials*. Park Ridge, Noyes Publications.(1995)

(69) Fiete, G. A., Nature Materials **11**, 1003-1004 (2012)

(70) Fradkin, E., E. Dagotto and D. Boyanovsky, Physical Review Letters **57**, 2967-2970 (1986)

(71) Frankenheim, M. L., Ann. Phys. **37**, (1836)

(72) Freund, L. B. and S. Suresh, *Thin Film Materials: Stress, Defects Formation and Surface Evolution*, Cambridge, Cambridge University Press.(2003)

(73) Frood, A. Riddle of 'Baghdad's batteries', BBC News(2003)

(74) Fu, L., Physical Review Letters **106**, 106802 (2011)

(75) Fu, L. and C. L. Kane, Physical Review B **76**, 045302 (2007)

(76) Fu, L., C. L. Kane and E. J. Mele, Physical Review Letters **98**, 106803 (2007)

(77) Gehring, P., H. M. Benia, Y. Weng, et al., Nano Letters **13**, 1179-1184 (2013)

(78) Geim, A. K. and K. S. Novoselov, Nature Materials **6**, 183-191 (2007)

(79) Giannozzi, P., S. Baroni, N. Bonini, et al., Journal of Physics: Condensed Matter **21**, 395502 (2009)

(80) Goldsmid, H. J., *Introduction to Thermoelectricity*, Berlin, Springer-Verlag.(2010)

(81) Grosso, G. and G. Pastori Parravicini, *Solid State Physics*, San Diego Academic Press.(2000)

(82) Halász, G. B. and L. Balents, Physical Review B **85**, 035103 (2012)

(83) Haldane, F. D. M., Physical Review Letters **61**, 2015-2018 (1988)

(84) Hall, E. H., The Johns Hopkins University Press **2**, 287 (1879)

(85) Hasan, M. Z. and C. L. Kane, Reviews of Modern Physics **82**, 3045-3067 (2010)

(86) Haviland, D. B., Y. Liu and A. M. Goldman, Physical Review Letters **62**, 2180-2183 (1989)

(87) Hellerstedt, J. T., M. T. Edmonds, J. H. Chen, et al., ArXiv **1405.5692**, (2014)

(88) Henini, M., Ed. *Molecular Beam Epitaxy From Research to Mass Production*. Wyman Street, Elsevier.(2013)

(89) Hikami, S., A. I. Larkin and N. Nagaoka, Progress of Theoretical Physics **63**, 707 (1980)

(90) Hohenberg, P. and W. Kohn, Physical Review **136**, B864-B871 (1964)

(91) Hong, S. S., J. J. Cha, D. Kong and Y. Cui, Nature Commun. **3**, 757 (2012)

(92) Hor, Y. S., A. Richardella, P. Roushan, et al., Physical Review B **79**, 195208 (2009)

(93) Hor, Y. S., A. J. Williams, J. G. Checkelsky, et al., Physical Review Letters **104**, 057001 (2010)

(94) Hsieh, D., D. Qian, L. Wray, et al., Nature **452**, 970-974 (2008)

(95) Hsieh, D., Y. Xia, L. Wray, et al., Science **323**, 919-922 (2009)

(96) Hsieh, T. H., H. Lin, J. Liu, et al., Nature Commun. **3**, 982 (2012)

(97) Huang, F. T., M. W. Chu, H. H. Kung, et al., Physical Review B **86**, 081104 (2012)

(98) Imada, M., A. Fujimori and Y. Tokura, Reviews of Modern Physics **70**, 1039-1263 (1998)

(99) Ioffe, A. F. and A. R. Regel, Prog. Semicond. **4**, 237 (1960)

(100) Jenkins, G. S., A. B. Sushkov, D. C. Schmadel, et al., Physical Review B **86**, 235133 (2012)

(101) Jeon, S., B. B. Zhou, A. Gyenis, et al., Nature Materials **13**, 851-856 (2014)

(102) Kane, C. L. and E. J. Mele, Physical Review Letters **95**, 226801 (2005)

(103) Kane, C. L. and E. J. Mele, Physical Review Letters **95**, 146802 (2005)

(104) Kasap, S. O., *Principles of Electronic Materials and Devices*, New York, McGraw-Hill Higher Education.(2002)

(105) Kim, D., S. Cho, N. P. Butch, et al., Nature Physics **8**, 459-463 (2012)

(106) Kim, D., P. Syers, N. P. Butch, et al., Nature Commun. **4**, 2040 (2013)

(107) Kim, D. J., S. Thomas, T. Grant, et al., Sci. Rep. **3**, (2013)

(108) Kim, Y. S., M. Brahlek, N. Bansal, et al., Physical Review B **84**, 073109 (2011)

(109) King, P. D. C., R. C. Hatch, M. Bianchi, et al., Physical Review Letters **107**, 096802 (2011)

(110) King, P. D. C., T. D. Veal and C. F. McConville, Physical Review B **77**, 125305 (2008)

(111) Klitzing, K. v., G. Dorda and M. Pepper, Physical Review Letters **45**, 494-497 (1980)

(112) Kohn, W. and L. J. Sham, Physical Review **140**, A1133-A1138 (1965)

(113) Koirala, N., M. Brahlek, M. Salehi and S. Oh, In Preparation (2014)

(114) Koma, A., Thin Solid Films **216**, 72-76 (1992)

(115) Kong, D., J. J. Cha, K. Lai, et al., ACS Nano **5**, 4698-4703 (2011)

(116) Kong, D., Y. Chen, J. J. Cha, et al., Nature Nanotechnology **6**, 705-709 (2011)

(117) König, M., S. Wiedmann, C. Brüne, et al., Science **318**, 766-770 (2007)

(118) Lee, C., R. He, Z. Wang, et al., Nanoscale (2013)

(119) Lee, H. D., C. Xu, S. M. Shubeita, et al., Thin Solid Films **556**, 322-324 (2014)

(120) Leijnse, M. and K. Flensberg, Semiconductor Science and Technology **27**, 124003 (2012)

(121) Li, C. H., O. M. J. van`t Erve, J. T. Robinson, et al., Nature Nanotechnology **9**, 218-224 (2014)

(122) Li, G., Z. Xiang, F. Yu, et al., ArXiv **1306.5221**, (2013)

(123) Li, H. D., Z. Y. Wang, X. Kan, et al., New Journal of Physics **12**, 103038 (2010)

(124) Linder, J., T. Yokoyama and A. Sudbø, Physical Review B **80**, 205401 (2009)

(125) Liu, J. and D. Vanderbilt, Physical Review B **88**, 224202 (2013)

(126) Liu, M., C.-Z. Chang, Z. Zhang, et al., Physical Review B **83**, 165440 (2011)

(127) Liu, Z. K., J. Jiang, B. Zhou, et al., Nature Materials **13**, 677-681 (2014)

(128) Liu, Z. K., B. Zhou, Y. Zhang, et al., Science **343**, 864-867 (2014)

(129) Lu, H.-Z. and S.-Q. Shen, Physical Review B **84**, 125138 (2011)

(130) MacKinnon, A., *Chapter 5. Low-Dimensional Semiconductor Structures Fundamentals and Device Applications*, Cambridge, Cambridge University Press.(2001)

(131) Margrave, J., *The Characterization of high-temperature vapors*, New York, John Wiley & Sons.(1967)

(132) Marzari, N., A. A. Mostofi, J. R. Yates, et al., Reviews of Modern Physics **84**, 1419-1475 (2012)

(133) Mattox, D. M., https://www.svc.org/HistoryofVacuumCoating/History-of-Vacuum-Coating.cfm#1600. (2003)

(134) Mellnik, A. R., J. S. Lee, A. Richardella, et al., Nature **511**, 449-451 (2014)

(135) Monkhorst, H. J. and J. D. Pack, Physical Review B **13**, 5188-5192 (1976)

(136) Moore, J. E., Nature **464**, 194-198 (2010)

(137) Moore, J. E. and L. Balents, Physical Review B **75**, 121306 (2007)

(138) Mostofi, A. A., J. R. Yates, Y.-S. Lee, et al., Computer Physics Communications **178**, 685-699 (2008)

(139) Mott, N. F. and E. A. Davis, *Electronic Processes in Non-Crystalline Materials 2nd Edition*, Oxford Clarendon Press.(1979)

(140) Nakahara, M., *Geometry, Topology and Physics* Bristol, Hilger.(1990)

(141) Neupane, M., N. Alidoust, S. Y. Xu, et al., Nature Commun. **4**, (2013)

(142) Nickerson, J. C., R. M. White, K. N. Lee, et al., Physical Review B **3**, 2030-2042 (1971)

(143) Oh, S., Science **340**, 153-154 (2013)

(144) Ohring, M., *The Materials Science of Thin Films Deposition and Structure*, San Diego, Academic Press.(2002)

(145) Okada, Y., M. Serbyn, H. Lin, et al., Science **341**, 1496-1499 (2013)

(146) Onsager, L., Physical Review **37**, 405-426 (1931)

(147) Onsager, L., Physical Review **38**, 2265-2279 (1931)

(148) Pashley, D. W., *Epitaxial Growth Ch. 1,* New York, Academic Press, INC. (1975)

(149) Pecharsky, V. K. and P. Y. Zavalij, *Fundamentals of Powder Diffraction and Structural Characterization of Materials*, New York, Springer.(2003)

(150) Peierls, R., Contemporary Physics **33**, 221-226 (1992)

(151) Peierls, R., T. A. Kaplan and P. W. Anderson, Physics Today **44**, 13 (1991)

(152) Perdew, J. P., K. Burke and M. Ernzerhof, Physical Review Letters **77**, 3865-3868 (1996)

(153) Pfeiffer, L. and K. W. West, Physica E: Low-dimensional Systems and Nanostructures **20**, 57-64 (2003)

(154) Prange, R. E. and S. M. Girvin, Eds. *The Quantum Hall Effect*. New York, Springer.(1990)

(155) Qi, X.-L., T. L. Hughes and S.-C. Zhang, Physical Review B **78**, 195424 (2008)

(156) Qi, X.-L., Y.-S. Wu and S.-C. Zhang, Physical Review B **74**, 085308 (2006)

(157) Qi, X.-L. and S.-C. Zhang, Physics Today **63**, 33-38 (2010)

(158) Qi, X.-L. and S.-C. Zhang, Reviews of Modern Physics **83**, 1057-1110 (2011)

(159) Qiu, X.-f., J.-J. Zhu, L. Pu, et al., Inorganic Chemistry Communications **7**, 319-321 (2004)

(160) Ren, Z., A. A. Taskin, S. Sasaki, et al., Physical Review B **82**, 241306 (2010)

(161) Ren, Z., A. A. Taskin, S. Sasaki, et al., Physical Review B **84**, 165311 (2011)

(162) Ren, Z., A. A. Taskin, S. Sasaki, et al., Physical Review B **85**, 155301 (2012)

(163) Richter, W. and C. R. Becker, physica status solidi (b) **84**, 619-628 (1977)

(164) Rondinelli, J. M., S. J. May and J. W. Freeland, MRS Bulletin **37**, 261-270 (2012)

(165) Roushan, P., J. Seo, C. V. Parker, et al., Nature **460**, 1106-1109 (2009)

(166) Royer, L., Bull. Soc. Fr. Mineral. Crist. **51**, (1928)

(167) Royer, L., Ann. Phys. **23**, (1935)

(168) Rüegg, A. and G. A. Fiete, Physical Review B **84**, 201103 (2011)

(169) Sample, H. H., W. J. Bruno, S. B. Sample and E. K. Sichel, Journal of Applied Physics **61**, (1987)

(170) Sato, T., K. Segawa, K. Kosaka, et al., Nature Physics **7**, 840-844 (2011)

(171) Shockley, W., Physical Review **56**, 317-323 (1939)

(172) Shoenberg, D., *Magnetic Oscillations in Metals*, Cambridge, Cambridge University Press.(1984)

(173) Singh-Miller, N. E. and N. Marzari, Physical Review B **80**, 235407 (2009)

(174) Smith, R. A., *Semiconductors*, Cambridge, Cambridge University Press.(1978)

(175) Steinberg, H., D. R. Gardner, Y. S. Lee and P. Jarillo-Herrero, Nano Letters **10**, 5032-5036 (2010)

(176) Steinberg, H., J. B. Laloë, V. Fatemi, et al., Physical Review B **84**, 233101 (2011)

(177) Stern, A. and N. H. Lindner, Science **339**, 1179-1184 (2013)

(178) Tamm, I., Phys. Z. Soviet Union **1**, 733 (1932)

(179) Tanaka, Y., Z. Ren, T. Sato, et al., Nature Physics **8**, 800-803 (2012)

(180) Taskin, A. A., S. Sasaki, K. Segawa and Y. Ando, Physical Review Letters **109**, 066803 (2012)

(181) Theis, C. D., J. Yeh, D. G. Schlom, et al., Thin Solid Films **325**, 107-114 (1998)

(182) Thomas, S., D. J. Kim, S. B. Chung, et al., ArXiv **1307.4133**, (2013)

(183) Thouless, D. J., M. Kohmoto, M. P. Nightingale and M. den Nijs, Physical Review Letters **49**, 405-408 (1982)

(184) Tsui, D. C., H. L. Stormer and A. C. Gossard, Physical Review Letters **48**, 1559-1562 (1982)

(185) Urazhdin, S., D. Bilc, S. D. Mahanti, et al., Physical Review B **69**, 085313 (2004)

(186) Valdés Aguilar, R., A. V. Stier, W. Liu, et al., Physical Review Letters **108**, 087403 (2012)

(187) Valdés Aguilar, R., L. Wu, A. V. Stier, et al., Journal of Applied Physics **113**, 153702 (2013)

(188) Van der Pauw, L. J., Philips Technical Review **20**, 220-224 (1958)

(189) Van der Pauw, L. J., Philips Research Reports **13**, 1-9 (1958)

(190) Vanderbilt, D., Surface Science Lecture Notes (unpublished) (2013)

(191) Vaško, A., L. Tichý, J. Horák and J. Weissenstein, Applied Physics A: Materials Science & Processing **5**, 217-221 (1974)

(192) Volkov, B. A. and O. A. Pankratov, Journal of Experimental and Theoritical Physics **42**, 178 (1985)

(193) Wan, K. J., T. Guo, W. K. Ford and J. C. Hermanson, Physical Review B **44**, 3471-3474 (1991)

(194) Wan, X., A. M. Turner, A. Vishwanath and S. Y. Savrasov, Physical Review B **83**, 205101 (2011)

(195) Wang, J., A. M. DaSilva, C.-Z. Chang, et al., Physical Review B **83**, 245438 (2011)

(196) Wang, M.-X., C. Liu, J.-P. Xu, et al., Science **336**, 52-55 (2012)

(197) Wang, Z., Y. Sun, X.-Q. Chen, et al., Physical Review B **85**, 195320 (2012)

(198) Wang, Z. Y., X. Guo, H. D. Li, et al., Applied Physics Letters **99**, 023112-023113 (2011)

(199) Watanabe, Y., S. Kaneko, H. Kawazoe and M. Yamane, Physical Review B **40**, 3133-3142 (1989)

(200) Wolgast, S., Ç. Kurdak, K. Sun, et al., Physical Review B **88**, 180405 (2013)

(201) Wray, L. A., S.-Y. Xu, Y. Xia, et al., Nature Physics **6**, 855-859 (2010)

(202) Wu, L., M. Brahlek, R. Valdes Aguilar, et al., Nature Physics **9**, 410-414 (2013)

(203) Wyckoff, R. Q. G., *Crystal Structure 2*, New York City, J. Wiley and Sons.(1964)

(204) Xia, Y., D. Qian, D. Hsieh, et al., Nature Physics **5**, 398-402 (2009)

(205) Xiong, J., Y. Khoo, S. Jia, et al., Physical Review B **88**, 035128 (2013)

(206) Xiong, J., Y. Luo, Y. Khoo, et al., Physical Review B **86**, 045314 (2012)

(207) Xu, S.-Y., C. Liu, N. Alidoust, et al., Nature Commun. **3**, 1192 (2012)

(208) Xu, S.-Y., M. Neupane, C. Liu, et al., Nature Physics **8**, 616-622 (2012)

(209) Xu, S.-Y., Y. Xia, L. A. Wray, et al., Science **332**, 560-564 (2011)

(210) Yan, B., M. Jansen and C. Felser, Nature Physics **9**, 709-711 (2013)

(211) Yang, K.-Y., W. Zhu, D. Xiao, et al., Physical Review B **84**, 201104 (2011)

(212) Yokoyama, T. and S. Murakami, Physica E: Low-dimensional Systems and Nanostructures **55**, 1-8 (2014)

(213) Young, S. M., S. Zaheer, J. C. Y. Teo, et al., Physical Review Letters **108**, 140405 (2012)

(214) Zangwill, A., *Physics at Surfaces*, Cambridge, Cambridge University Press.(1988)

(215) Zhang, G., H. Qin, J. Teng, et al., Applied Physics Letters **95**, - (2009)

(216) Zhang, H., C.-X. Liu, X.-L. Qi, et al., Nature Physics **5**, 438-442 (2009)

(217) Zhang, Q., Z. Zhang, Z. Zhu, et al., ACS Nano **6**, 2345-2352 (2012)

(218) Zhang, X., N. P. Butch, P. Syers, et al., Physical Review X **3**, 011011 (2013)

(219) Zhang, Y., K. He, C.-Z. Chang, et al., Nature Physics **6**, 584-588 (2010)

(220) Zhu, H., C. A. Richter, E. Zhao, et al., Sci. Rep. **3**, (2013)

(221) Zhu, Z. and J. E. Hoffman, Nature **513**, 319-320 (2014)

(222) Žutić, I., J. Fabian and S. Das Sarma, Reviews of Modern Physics **76**, 323-410 (2004)