The Dissertation Committee for Christopher William Mann Certifies that this is the approved version of the following dissertation:

Scanning tunneling microscopy of Bi_2Se_3 and $Cu_xBi_2Se_3$

Committee:

Chih-Kang Shih, Supervisor

Alex de Lozanne

Gregory Fiete

Li Shi

Emanuel Tutuc

Scanning tunneling microscopy of Bi_2Se_3 and $Cu_xBi_2Se_3$

by

Christopher William Mann, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2013

Acknowledgements

I'm very grateful to many, many people for helping me throughout my degree process. I want to thank everyone in the Shih Lab: whether it was with a trick for solving a problem or a deep insight that alters my understanding of a system, I wouldn't have been able to complete my degree without the help of the whole lab.

I'm grateful to everyone in my family for their support, not just in my PhD process, but in their support of my interest in science from a young age. I'm extremely grateful to Jennifer Comstock, my fiancée, for her support, patience, and feedback. I also would not have made it this far into research without an early introduction to nanoscience from Tom Reynolds, whose insights helped pave an extremely valuable foundation for launching a career in research.

Finally, I want to thank Professor Shih for encouraging me through trajectory changes. The initial opportunity to work on plasmonics and nanofabrication techniques has provided a valuable skill set that is already paying dividends. The opportunity to pursue my own research on the graphene project prior to landing on Bi_2Se_3 was invaluable. The continued support throughout our multiple iterations on Bi_2Se_3 has helped transition my doctoral project from an interesting data set to a scientifically rigorous investigation of topological insulators, a field I initially found daunting and unapproachable.

Without the support of everyone mentioned above, and many others, this thesis would never have happened. To everyone who helped me reach this point, thank you.

Scanning tunneling microscopy of Bi_2Se_3 and $Cu_xBi_2Se_3$

Christopher William Mann, PhD

University of Texas of Austin, 2013

Supervisor: Chih-Kang Shih

Recently, Bi_2Se_3 was added to a new class of materials known as topological insulators. While several studies have provided tantalizing hints towards novel physical properties, such as backscatter suppression and spin-polarized transport, several concerns remain in actual materials. In particular, high defect densities, strong surface band bending, and potential fluctuations have been observed.

Here, scanning tunneling microscopy and spectroscopy are used to reveal surface effects in Bi_2Se_3 and $Cu_xBi_2Se_3$. First, a detailed examination of defects in bulk-grown samples is described. Then, I provide an analysis of molecular beam epitaxy results, done in collaboration with colleague Yuxuan Chen. Following this, I provide a detailed study of individual point defects in Cu-doped Bi_2Se_3 and examine how Cu is incorporated into the Bi_2Se_3 lattice. Finally, through spectroscopic analysis, a novel depth-sensitive measurement of the local band bending field is developed. Furthermore, for the first time, fluctuations of the Dirac point can be correlated to specific near-surface defects, namely Se vacancies.

These analyses provide valuable insights into the preparation of future samples for the investigation of topological insulators.

Table of Contents

List of Figures

Chapter 1: Introduction to scanning tunneling microscopy

OVERVIEW

Key to the operation of scanning tunneling microscopy is the ability to maintain a tunneling current between a probe and a sample; generally speaking, the operating parameters of interest are angstrom-scale distances and picoamp-scale currents. This chapter describes the technical aspects of scanning probe microscopy in order to lay a foundation for the later scientific discussions. An introduction to quantum tunneling is provided, followed by a technical discussion of the information that can be obtained using this technique.

BACKGROUND AND THEORY

Quantum theory of tunneling

Tunneling is a quantum process where particles in one system can travel through classically-forbidden barriers. Classically, when a particle is confined by a potential, it can only escape if it has enough energy—for example, for a Mars-bound rocket to escape Earth's pull, it's necessary for the rocket to achieve escape velocity. In a quantum system, it's possible for a particle to leave a confining potential without ever achieving 'escape velocity.'

To illustrate this process, consider the 1D time-independent Schrödinger equation in atomic units (this section follows from [1]):

$$H\psi = \mathrm{E}\psi \qquad \text{Eq. 1}$$

$$H = -\frac{1}{2}\frac{\partial^2}{\partial x^2} + V(x) \qquad \text{Eq. 2}$$

The Hamiltonian, H, contains a kinetic energy term, $-\frac{1}{2}\frac{\partial^2}{\partial x^2}$, and a potential energy term, $V(x)$. To demonstrate quantum tunneling through a barrier, consider an electron scattering off a potential $V(x) = \alpha\ \delta(x)$. On either side of the barrier, the electron is free and its wavefunction can be written:

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx}, x < 0 \\ Fe^{ikx} + Ge^{-ikx}, x > 0 \end{cases} \qquad \text{Eq. 3}$$

where A and G are the amplitudes of incident waves and B and F are the amplitudes of waves propagating away from the barrier, shown in Figure 1, and k is the wave vector ($E = k^2/2$). Because ψ must be a continuous function, at $x = 0$, $A + B = F + G$. Typically, the derivative of ψ must also be continuous, but the delta function is infinite in height, so a different boundary condition is used: $\frac{\partial\psi}{\partial x}\Big|_{+\epsilon} - \frac{\partial\psi}{\partial x}\Big|_{-\epsilon} = 2\int_{-\epsilon}^{+\epsilon} V(x)dx = 2\alpha\psi(0)$ for arbitrarily small ϵ. With these two boundary conditions, the following relation is found: $F - G = A(1 - 2i\alpha/k) - B(1 + 2i\alpha/k)$.

Assuming that a wave is inbound from the left and has a chance to scatter through the barrier or be reflected, then $G = 0$, $B = \frac{-i\alpha/k}{1+i\alpha/k}A$, and $F = \frac{1}{1+i\alpha/k}A$. The reflection coefficient is the relative amplitudes of inbound and reflected waves, or $R = \frac{|B|^2}{|A|^2} = \frac{\alpha^2}{k^2+\alpha^2}$ and the transmission coefficient is $T = \frac{|F|^2}{|A|^2} = \frac{k^2}{k^2+\alpha^2} \neq 0$.

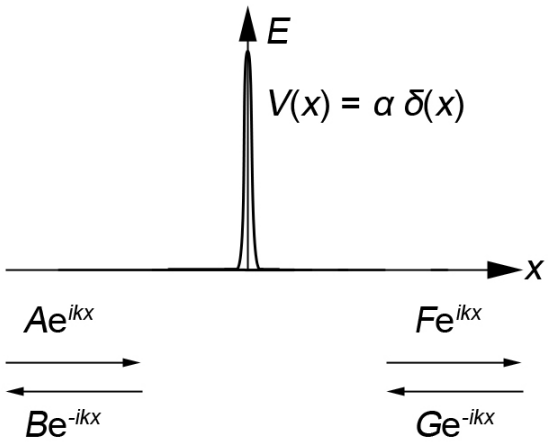


Figure 1. Scattering from a 1D delta function.

Adapted from [1], the time-independent approach to tunneling through a delta function barrier uses counter-propagating waves. The delta function is technically an infinite-height barrier; despite an infinite height, the particle has a probability to tunnel through. Classically, no particles can transmit through the barrier.

In other words, despite having an infinite potential, there is a nonzero transmission coefficient—this means that the particle can tunnel through the barrier. It's possible to take advantage of this property to study the properties of materials. While the barrier in this illustration was an infinite delta function, the junctions used in tunneling spectroscopy are often trapezoids and can allow considerable current through the junction.

Band theory of solids

Excellent introductions to band theory are available in a variety of textbooks—for example, [1–4]. However, as the concept of a topological insulator is relatively

new, this section is intended to lay a foundation for the later discussions about topological insulators and impurity states in subsequent chapters.

In solids, the density of states is determined by allowed energy 'bands.' Energy bands are the result of discrete atomic states trending towards a continuum of states. To develop the principle more completely, consider a single hydrogen atom:

$$H\psi_{nlm} = E_{nlm}\psi_{nlm}$$
$$E_{nlm} = -\frac{1}{n^2} \text{ x } 13.6\text{eV}$$

The allowed states strictly follow a $1/n^2$ trend and can only exist at these discrete states. However, if the hydrogen atom is placed near a second hydrogen nucleus, the electron will interact with both nuclei and the system will be perturbed. This perturbation showcases the origin of bonding in molecular systems—in particular, the separation between the two nuclei will determine which energy levels are available and what, if any, separation leads to energy minimization, as shown in Figure 2. If 2 electrons are in this system, they would fill the lowest energy state. This is the bonding state and would constitute the highest occupied molecular orbital, or HOMO, state. The anti-bonding state is then the lowest unoccupied molecular orbital, or LUMO. The relative positions of the HOMO and LUMO are critical in determining the behavior of a quantum system and, as described in Chapter 3, can describe the topology of the system.

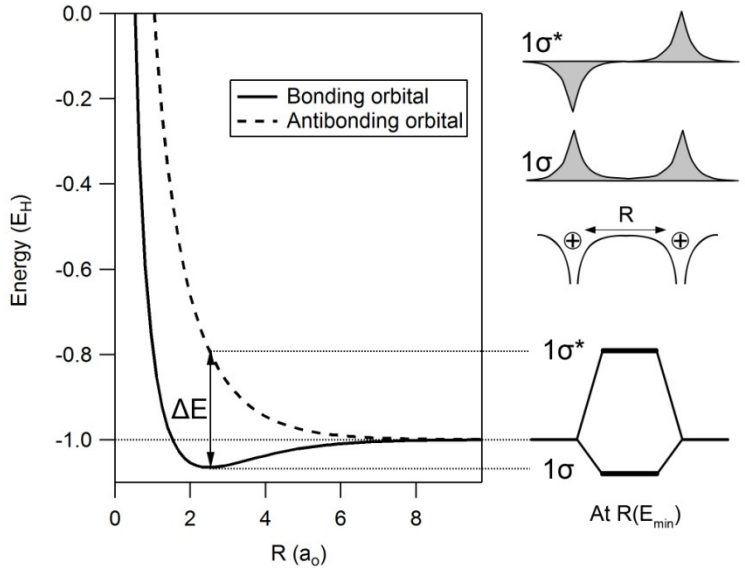


Figure 2. Hydrogen molecule ion: energy vs. nucleus separation.

Energy splitting as a function of separation showing the bonding and anti-bonding states. Near $2.5a_o$, the energy is minimized, suggesting a stable state. Calculation performed using the LCAO method with 1s hydrogen orbitals.

As more particles are added to the system, the energy levels will split again. Consequently, when enough atoms are added, instead of discrete energy levels, a band is formed. As an illustration, Figure 3 shows the levels splitting for 1s states of a 1D hydrogen-like solid. In the limit of infinite particles, the electronic states approach a continuum, or a band. The band can accept particles at any energy within the band, but the density of states is not necessarily uniform; some energy levels may allow a higher density of states.

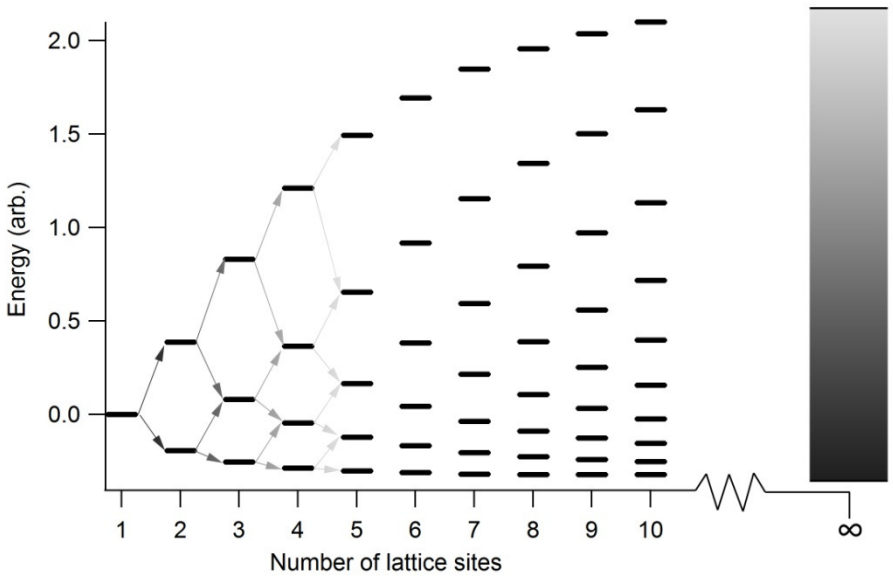


Figure 3. 1D hydrogen-like solid energy levels determined from 1s states.

The hydrogen ion states, similar to Figure 2, split when more atoms are added. By adding more atoms in a 1D chain, the levels continue to split until, eventually, a band is formed. To calculate this illustration, the LCAO method was used with 1s orbitals using the following charge-neutral Hamiltonian: $H = -\frac{\nabla^2}{2} - \frac{1}{N}\sum\frac{1}{|\vec{r}-\vec{R}_i|} + \frac{1}{N^2}\sum\frac{1}{|\vec{R}_j-\vec{R}_i|}$. These energies are extracted from $R(E_{min})$. This cartoon Hamiltonian remains charge-neutral with multiple ion cores and a single electron hopping between them, but does not conserve electrostatic energy. Therefore, the computed $E(R = \infty)$ values are subtracted from each simulation to account for the energy offset.

When multiple bands are present in a system, the highest full band takes on the role of the HOMO and the lowest empty band takes on the role of the LUMO: these are termed the valence band and conduction band, respectively. The conduction band allows electron conduction, as its name implies. Simply put, in order for an electron to move through a material, additional sites must be available to hop into. If a single electron is added to an empty band, then all positions within the band are available for hopping, and the electron can conduct through the material. Conversely, if there are no available positions, conductivity will go to zero and the material will behave like an insulator.

Insulators and semiconductors

The energy gap between the valence band and conduction band is the band gap and the energy scale of the band gap determines whether the material is insulating or semiconducting. For a small gap, it's possible for thermal fluctuations or impurities to promote electrons from the valence band to the conduction band—while the material starts as an insulator, these perturbations turn it into a partially conducting system, or a semiconductor. In general, semiconductors have small band gaps, allowing them to be easily tuned (by impurities or electrical gating, for instance) to create the desired conductivity. Insulators have larger gaps than semiconductors, so thermal fluctuations or electrical gating is typically insufficient to create carriers.

Conductors

Conductors are zero-gap materials, so impurities or perturbations are not required to promote an electron to the conduction band. As discussed in Chapter 3, this means that conductors have a different topology from insulators. Generally speaking, scanning tunneling microscopy is best suited for studying semiconductors and conductors.

TOPOGRAPHY

STM can provide rudimentary topographic information about samples, though its quantitative interpretation can be tricky. Because the tunneling current falls off exponentially with distance, creating a map of constant current is a rough approximation of the actual topography. However, the interpretation is complicated by the fact that non-physical electronic features can also show up in such a map.

However, for a great many studies, this approximation provides sufficient information to draw meaningful conclusions.

In order to maintain a constant tunneling current, precise control of relative position is required. A class of materials known as piezoelectrics is well-suited for this task: by applying a voltage, the material moves, though only slightly. In the case of STM piezo tubes, it's not uncommon to apply 150V in order to achieve a deflection of 1μm. The most common piezoelectric material used in scanning probe microscopy is PZT, or lead zirconate titanate.

PZT experiences a few key problems: hysteresis, vibration, and creep. [5] With hysteresis: sweeping from 0V to 10V and back to 0V will not return the piezo to the original location. At lower temperatures, however, the hysteresis is reduced, incentivizing lower temperature studies. If the piezo is scanned close to its resonant frequency, significant vibrations can result, typically leading to much slower scan speeds than are fundamentally possible. Finally, the piezo will experience creep that often will not stabilize for long periods of time. One commonly-observed consequence of creep is the smearing that occurs at the beginning of a scan frame. Scanning at lower temperature will greatly reduce creep.

Spatial resolution is dependent on the tunneling conditions, though atomic resolution is readily achieved for some samples. In general, the shape of the STM tip is convoluted into the data, as illustrated in Figure 4. In some cases, features appear to be repeated, like Figure 4c & d, but it's actually a multiple tip. Complicating the matter, crystals exhibit long-range order, so repeated features are innately part of many samples. To separate actual long-range order from multi-tip conditions, it's often necessary to find defects or step edges. If defects always show up in pairs, or

triplets, it's likely that the tip is duplicating the features artificially. At step edges, it's also possible to find duplicates or otherwise examine the sharpness of the tip; isolated steps are more useful, but this can be done on vicinal substrates, as well.

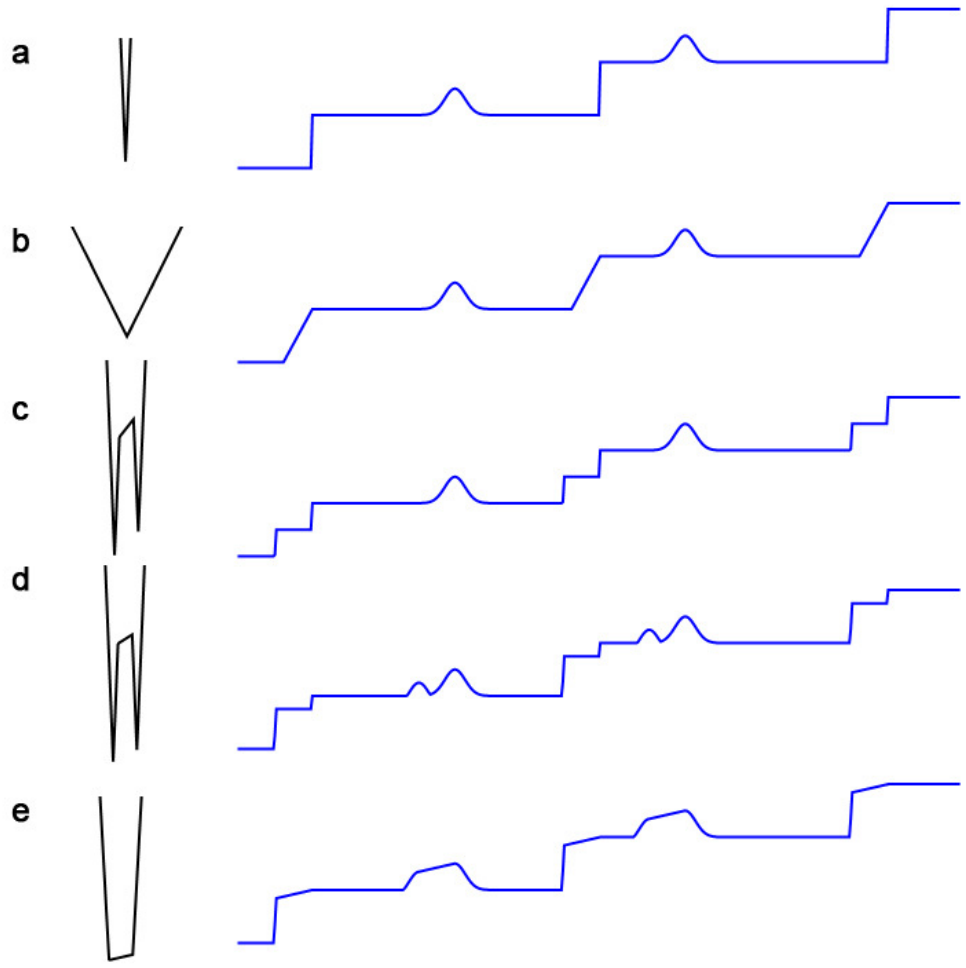


Figure 4. Tip shape convolution with topography.

The shape of the STM tip, left, determines the observed topography, right. **a** A delta-function tip exactly reproduces the topography. **b** A blunt tip smears out sharp features like atomic steps, but has less impact on broad features. **c** A double tip can create the illusion of multiple steps when there is actually only one, and can generate different topographs depending on the height of the features and the double tip, like in **d**. **e** Other tip shapes will also be convoluted into the geometry, leading to misleading shapes.

SPECTROSCOPY

Introduction

STM is not limited to topographic information. Because the tunneling mechanism itself depends on the available tunneling states, information about the electronic structure is accessible. In particular, tunneling is a measure of the coupling of filled electronic states in one system and empty electronic states in another system. By sweeping the relative voltage and observing how much more current flows at different biases, or taking a dI/dV sweep, it's possible to map out the tip-sample joint density of states, or the joint spectral functions. The density of states is generally proportional to the spectral function, however, so in most cases the difference is moot. While there is no such thing as a calibrated STM tip, metals are generally a good material for spectroscopy.

The strength of tunneling spectroscopy is its ability to extract information about the electronic structure of materials. While the example above shows that tunneling is energy and *k*-dependent, the actual measured tunneling current is directly representative of the number of electronic states available to tunnel to and from. Following from [3], the Hamiltonian representing such a process can be written:

$$H_T = \sum_{\nu\mu} \left(T_{\nu\mu} c_{1,\nu}^{\dagger} c_{2,\mu} + T_{\nu\mu}^{*} c_{2,\mu}^{\dagger} c_{1,\nu} \right), \qquad \text{Eq. 4}$$

where H_T is the tunneling Hamiltonian, c represent the electron operators for systems 1 and 2, the single-particle states $|\nu\rangle$ and $|\mu\rangle$ are complete sets for systems 1 and 2, respectively, and

$$T_{\nu\mu} = \int d\vec{r}\, \psi_{\nu}^{*}(\vec{r}) H(\vec{r}) \psi_{\mu}(\vec{r}) \qquad \text{Eq. 5}$$

is the tunneling matrix element between the states with $H(\vec{r})$ the single-particle Hamiltonian. When the two systems are at different biases, V, it's possible to achieve a net tunneling current, given by

$$I = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sum_{\nu\mu} \left|T_{\nu\mu}\right|^2 A_{sample}(\nu, \omega) A_{tip}(\mu, \omega + eV)[n_F(\omega + eV) - n_F(\omega)], \qquad \text{Eq. 6}$$

where I is the tunneling current, n_F is the Fermi distribution function, and A represents the spectral function for each system. This equation provides a number of insights into the tunneling process. For instance, the tunneling current is weighted by a tunneling matrix element, $T_{\nu\mu}$, which will change depending on the actual symmetry of the states. For instance, if the STM tip has a strong p-wave orbital extending from the tip, it will couple differently to the sample than if the tip is more s-like, and thereby change the measured data. Next, the tunneling current depends on the available states in both systems—the spectral function, $A(\nu, \omega)$, is a measure of how many states are available at energy ω and is proportional to the density of states. Sweeping the bias, V, allows the states in both systems to be measured. Finally, the tunneling current also depends on the temperature—which states contribute are broadened at finite temperature by $n_F(\omega)$.

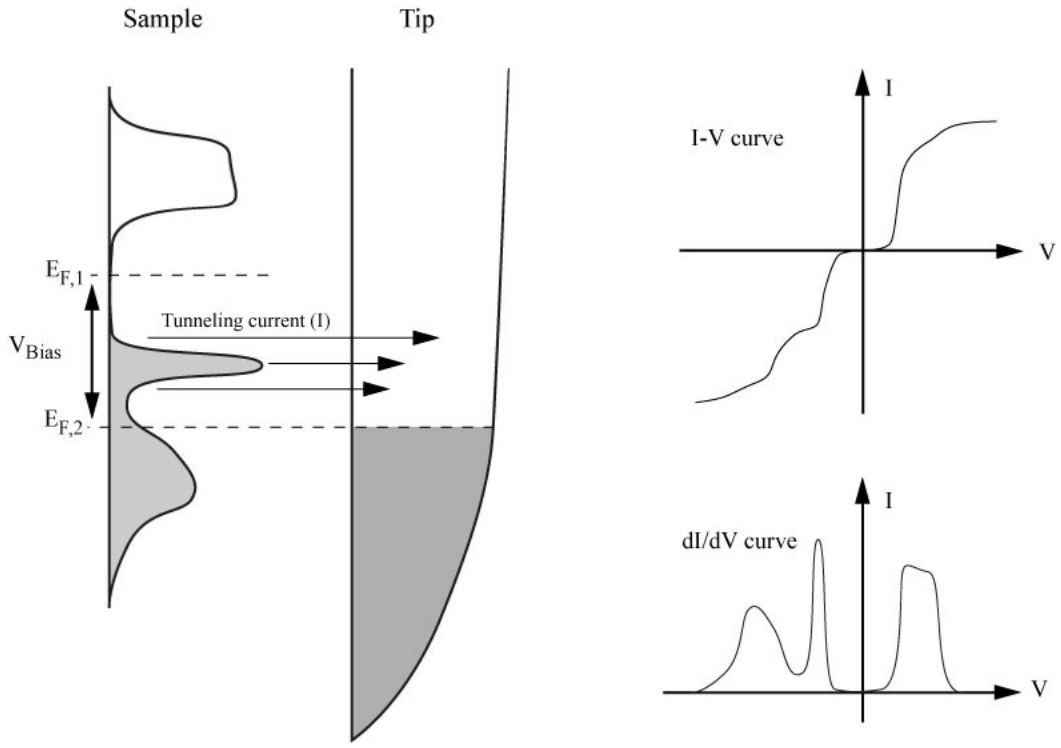


Figure 5. Principle of tunneling spectroscopy.

The tunneling current is proportional to the overlap of states between the tip and the sample. Offsetting the sample (or tip) bias allows additional states to tunnel, thereby increasing the current. By sweeping the bias and measuring the current (taking an IV curve), the states are summed. The derivative, then, is ideally a representation of the density of states. Adapted from [3].

Because the tunneling matrix element is actually fairly complicated and depends on the shape of the STM tip itself, which may not be reproducible, it's common to instead examine a derivative of this expression. If a material with a fairly featureless tunneling spectrum is used for the STM tip, such as a simple metal, an approximation can be made:

$$\sum_{\mu} \left|T_{\nu\mu}\right|^2 A_{tip}(\mu, \omega + eV) \approx constant. \qquad \text{Eq. 7}$$

Consequently, by taking the derivative of the current with respect to bias,

$$\frac{dI}{dV} \propto \int_{-\infty}^{\infty} d\omega \left(\frac{-\partial n_F(\omega + eV)}{\partial \omega}\right) \sum_{\nu} A_{sample}(\nu, \omega). \quad \text{Eq. 8}$$

A final assumption is typically used, as well: at low temperatures, the Fermi function becomes a delta function, giving:

$$\frac{dI}{dV} \propto \sum_{\nu} A_{sample}(\nu, -eV). \quad \text{Eq. 9}$$

Therefore, by finding the derivative of an I-V tunneling curve, it's possible to extract the approximate spectral function, as shown schematically in Figure 5. By taking such data at numerous locations, it's possible to generate a map of the local states as a function of position. This is, in fact, one of the most significant capabilities of STM: accessing the density of states with subatomic precision allows direct insights into the roles of individual atomic defects and, on occasion, can even provide momentum-resolved information, as well.

dI/dV techniques

Another technique combines spectroscopy and topography and can produce a map of the local dI/dV structure as a function of position: lock-in dI/dV imaging. Lock-in amplifiers work by producing an oscillatory function, $sin(\omega t)$, then measuring the sample response and multiplying the results together. The orthogonality of sine functions results in the local derivative. More generally, higher harmonics effectively measure higher derivatives.

In the course of my studies, I found that spectroscopic data at liquid nitrogen temperatures were severely influenced by drift. While the scan may begin on a feature of interest, by the time enough dI/dV sweeps have been averaged, the STM tip is no longer on the feature. One solution to this problem was to use multipass scans at

constant current—each pass has a different bias and the corresponding dI/dV map is saved. Then, the drift-corrected topographic maps at each bias are overlayed and the dI/dV data at each pass can be aligned. A dI/dV data set can be extracted from this, though it is no longer taken at constant height—instead, it is taken at constant current.

Unfortunately, the simple density-of-states picture breaks down for real systems, partly due to a nonlinear tunneling transmission factor that depends on bias. [6] The observed density-of-states will be different for different forms of spectroscopy. In Eq. 9, the assumption is made that the tunneling current depends only on how many states are available to tunnel into. However, a more accurate form is: [7]

$$\frac{dI}{dV} \propto \rho_{tip}(0)\rho_{sample}(eV)T(eV, eV, z) \qquad \text{Eq. 10}$$

with ρ the density of states and

$$T(E_1, E_2, z) = \exp\left[-z\sqrt{\frac{4m}{\hbar^2}(\phi_{tip} + \phi_{sample} + E_2 - 2E_1)}\right]. \qquad \text{Eq. 11}$$

where ϕ represents the work function. The T factor is the transmission coefficient and it depends nonlinearly on the tip-sample separation and the sample bias. This term is nonlinear in z and bias. Some of the data included in this thesis is constant-current spectroscopy, so understanding this nonlinearity is important. The current can be approximated as:

$$I(z) \propto \int_0^{eV} \rho_{sample}(E)\rho_{tip}T(E, eV, z)dE \qquad \text{Eq. 12}$$

Clearly, the height, z, changes to accommodate the current—for a given bias, the STM tip height is adjusted until it acquires the current set point, and this relation is not always linear or simple. To simulate the consequences of this for a known (and not simple) ρ_{sample}, it's necessary to iteratively solve for the height at which the current

will match the setpoint. In Chapter 7, this process is used to simulate the differences between the dI/dV curves at constant height and constant current. Therefore, I have included the qualitative process for simulating such data here:

For V in Range(V_{min}, V_{max}, step = dV):

Let $\rho_{sample}(V) = f(V)$.

Find the tip-sample separation for current setpoint I_p:

$z(V)$ = FindRoot($I_p - \int_0^V f(E)\, T(E,V,z)dE = 0, initial\ guess: z = 1nm$)

$dI/dV(V) = f(V)T(V,V,z(V))$

FindRoot can be a Newton-Cotes solver, for instance. From there, the dI/dV curve is simulated at constant current. To simulate constant height spectroscopy from a known $f(V)$, it's only necessary to use the final term in the pseudo-code.

Momentum from dI/dV

Mapping the local spectral density can provide additional information, such as scattering states. In the presence of defects, surface scattering states can produce higher densities of states locally, from which it is often possible to extract momentum information from the surface states.

The connection between wavelength and energy is not trivial. First, consider a simple standing wave pattern of ripples with no asymmetry. The standing waves are functions only of radial position. Wavelength is well-defined in Fourier space, so the Fourier transform of a scattered wave pattern produces a map of the scattering states in k-space: the 2D FFT provides angle-dependent momentum measurements. This

technique, known as FT-STM, is capable of providing insights about which states are available in a sample.

To illustrate this, consider a circular Fermi surface, $E(\vec{k}) = k_F\sqrt{{k_x}^2 + {k_y}^2}$. An electron at one point of the circle, $\vec{k}_1$, is allowed to scatter to any other point on the circle, $\vec{k}_2$. This scattering vector, $\vec{q} = \vec{k}_2 - \vec{k}_1$, is the result of the mixing of the initial and final states. Consequently, at a local point defect, a scattering state with wave vector magnitude q forms. [8] The question, then, is which wave vectors will dominate the scattering? In general, several factors are required to answer this question [9]:

1. What is the potential, *V*, that initiates the scattering? Is the scatterer isotropic and sharp?
2. What is the shape of the local Fermi surface? Where are the stationary points?
3. Are there symmetry requirements, such as spin transitions, that must be considered?
4. Is the local Fermi surface flat?

The answer to (1) determines whether a simple scattering theory can be used. If the scatterers are sharp and s-like, then the analysis is greatly simplified. For (2), in the case of the circular Fermi surface, for any given incoming wave vector $\vec{k}_1$, there is exactly 1 stationary point: $-\vec{k}_1$. The stationary points are positions where, for a slight change in the outgoing wave vector, $\vec{k}_1+\vec{\delta}$, the total scattering vector length, q, is unchanged to first order. More precisely, assuming that the scattering surfaces can be locally parameterized by a parameter t, where does $\frac{\partial E(\vec{q}(t))}{\partial t} = 0$? These states are reinforced and will have the strongest signal. For (3), in the case of topological

insulators, time-reversal symmetry prevents backscattering, so the only stationary point in the above model system, with $q = 2k_F$, is actually suppressed. Finally, in some systems like *d*-wave superconductors, there are flat regions of the Fermi surface that can dominate the scattering vectors in a way that is analogous to (2).

Chapter 2: Experimental apparatus

OVERVIEW

The STM used in this study is lovingly referred to as LT-STM3, or Low Temperature STM #3, as it's the 3rd generation design in the Shih Lab. Hyoungdo Nam was responsible for the primary chamber design and I was responsible for the supporting components, maintenance, and operation. The scan head was designed by RHK and uses stick-slip piezo walkers. The ability to move in X, Y, and Z allows position-dependent studies to be conducted. The STM controller was manufactured by Nanonis and uses LabView. The experimental chamber, shown schematically in Figure 6, is stainless steel and is maintains a vacuum of ~5×10^{-11} torr.

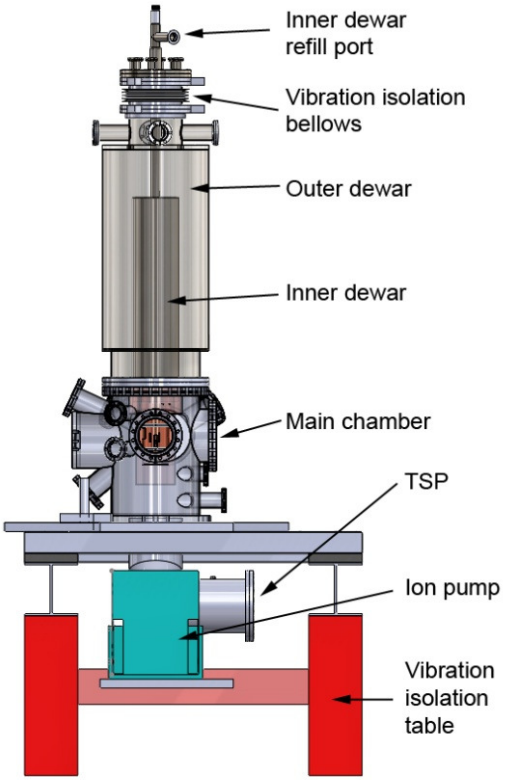


Figure 6. UHV STM chamber diagram.

The instrument consists of several parts: a vacuum chamber pumped by an ion pump, a vibration isolation table and a vibration isolation bellows, an inner and outer dewar, and the main chamber itself.

VACUUM SYSTEMS

To maintain sample cleanliness, it's highly desirable to keep the system clean. While argon and nitrogen environments are often acceptable, the system used here was maintained at ultra-high vacuum (UHV). UHV systems require additional considerations over standard high vacuum systems: vacuum greases cannot be used to seal gaskets, all surfaces must be completely oil-free, and an extensive bake-out is typically required.

Vacuum compatible materials

Many traditional materials used in science and engineering are not permitted in UHV environments: plastics can outgas and cause contamination, brass and some steels contain volatile substances like zinc and phosphorous, which can contaminate a UHV system, traditional solder is not vacuum compatible, and all materials must be able to survive a bake-out up to at least 150^{o}C.

Preparing a part for vacuum

Prior to introducing anything into a UHV system, a proper cleaning routine is required. Oils and other organics on the surface of materials, including grease from machining or oils from hands, can contaminate the vacuum system. The recommended process for preparing most materials is:

1. Rinse and scrub with alconox soap and water to remove gross debris.
2. Sonicate 15-30 minutes in alconox soap and water. Do not include too much soap, there shouldn't be any physical soap grains left at the end of the sonication. Repeat if the part was recently returned from the machine shop and has extra oils on it.

3. Sonicate 15-30 minutes in DI water. Repeat if the water is still soapy.
4. Sonicate 15-30 minutes in acetone to start dehydrating the material.
5. Sonicate 15-30 minutes in isopropanol (IPA) to remove residual water or acetone.
6. Remove the material from the IPA and immediately blow dry with canned air or N_2 gas. Allowing the IPA to dry on the surface will allow any residual dissolved contaminants to redeposit onto the material; blowing the IPA off prevents this.

This process should be modified if special materials or contaminants are expected. For instance, this process will not remove waxy contaminants. Because wax is used as a binder for some processes, extra steps must be taken, such as a methylene chloride rinse or an acid wash. Any new materials should be fully vetted prior to adding them to a vacuum system in order to prevent contamination.

Plastics

In practice, there are a few plastics which can be used in limited quantities: Teflon and Kapton have exceptional outgassing characteristics [10] and can be used as coatings for low-current wires in some cases. However, to minimize the amount of plastic, it's recommended to use very thin wires. For high-current wires where heating occurs, it's recommended that ceramic beads be used to separate uncoated copper wires.

Gate valves also require plastics to be effective. However, these parts are generally purchased and will behave nominally if the manuals are followed—they often have specific bake-out requirements.

Stainless steel

Stainless steel is a common choice for vacuum systems. Unfortunately, water and hydrogen gas are relatively soluble in stainless, so an extensive bake-out process is required to achieve UHV. If placed directly into a vacuum, water and hydrogen will slowly outgas from stainless steel, causing a reduced vacuum. The solution to this problem is to heat it up and evolve as many gases as possible while still pumping on the system, thereby removing them from the vacuum. Then, upon cooling down, the outgassing rate will be substantially lower. Stainless steel can be baked to 450^{o}C, but the bake-out temperatures are typically limited by the scientific apparatuses in the vacuum system—the piezos used in STM systems limit bake-out temperatures to ~150^{o}C.

Aluminum

Non-anodized aluminum is an excellent material for vacuum systems. The native oxide on aluminum creates a barrier that prevents hydrogen and water from penetrating into the metal, so it has very low outgassing rates. A low-temperature (120^{o}C) bake-out is typically recommended to desorb anything on the surface, but, along with de-greasing steps, this is the extent of the preparation. Anodized aluminum is not recommended: the anodization process creates a porous surface that can trap residual gasses.

Copper

Oxygen-free copper can be used in vacuum systems and is commonly required for making thermal contact between the inside and outside of a vacuum chamber. Additionally, because copper is a soft metal, it can be used as a gasket. In LT-STM3, copper gaskets are used for all gaskets with the exception of gate valves.

Epoxies

Several epoxies are available for UHV conditions. For electrically conductive epoxy, we used EpoTek H-20E with a full 2-hour cure to reduce outgassing. For thermally conductive, electrically insulating epoxy, we used EpoTek H70, also with a 2-hour curing schedule.

Pumping, baking, and leak-checking procedure

Upon sealing the chamber, it is then pumped down with a roughing pump to ~10^{-4} torr. Because the chamber is quite large, this can take some time (15+ minutes). If, after a few minutes, the roughing pump still makes gargling sounds, there may be a significant leak somewhere. Check the lines and system. To be safe, I recommend waiting 30 minutes before moving on. At this point, it's safe to turn on the ion gauge—make sure the current multiplier is set for 0.1x sensitivity.

The turbo pump can be started once the chamber is in the 10^{-4} torr range. If it does not reach the 'Normal' state and stay there, it's quite likely that there is a leak in the chamber. At this pressure range, the best way to check for leaks is with an IPA spray bottle—go over each newly-sealed flange and add a small amount of IPA. If the pressure goes up, the IPA is leaking into the vacuum system. Tighten the bolts on the flange—if the pressure does not immediately recover, it may be necessary to vent and change the gasket. If, upon venting and changing the gasket, the flange still leaks, inspect the knife-edge of the flanges for dings or chips.

Once the chamber reaches the low 10^{-6} range, start outgassing the titanium sublimation pump (TSP). The TSP works by running a large current through a titanium rod, causing it to heat up and begin sublimating. The sublimated titanium

redeposits on the sides of the vacuum chamber, where it acts as a getter pump: when a contaminant, like water, lands on a surface with freshly-deposited titanium, it reacts with the titanium and is thereby removed from the vacuum. However, the TSP must be fully outgassed prior to being functional, and outgassing it can take quite some time. Flash the TSP with increasing time and current, never letting the pressure in the chamber reach above $5x10^{-4}$ torr. The pump is fully outgassed once it can stay on at 45A for ~1 minute without the pressure rising—often, the pressure will begin falling rapidly, instead of rising. In this case, allow it to fall until it reaches a minimum, then turn off the TSP.

At this point the pressure is often near 10^{-8} torr. Prior to wrapping the chamber with aluminum foil in order to bake it, a second leak check is recommended. Go through all of the newly-sealed flanges and repeat the IPA test. If the pressure goes up when IPA is sprayed on a flange, tighten the flange and try it again. If the flange continues to leak, it won't get better by baking! Vent the chamber and replace the gasket.

If no leaks are apparent, it's time to begin the bake-out. Choose several major flanges in the system and attach thermocouples. The chamber should be wrapped in a single layer of aluminum foil in order to distribute the heat evenly. All viewports or ceramic feedthroughs should be wrapped in two layers of aluminum foil. The aluminum foil should then be wrapped with heating tapes. Do not allow the heating tapes to overlap each other, they can cause intense localized heating that may cause the heating tapes to overheat and fail (this can lead to a rather complicated wrapping scheme, a background in topology is recommended!). Wrap the heating tapes with 2 layers of aluminum foil—the purpose of this layer is to keep the heat in the system.

Plug the heating tapes into the variacs and raise the variac to ~30%. With the clamp multimeter, ensure that equal currents are flowing through both wires of the heating tapes. If they're uneven, this may indicate a short somewhere, which could lead to a fire.

Any heating tapes attached to flanges with viewports or ceramics should be ramped slower than bare-metal regions. Ramp the variacs up until the chamber is at the desired bake-out temperature, typically 150°C for most regions, but 120°C for flanges attached to sensitive equipment. Bake the chamber for a minimum of 48 hours. Once the pressure stabilizes in the low 10^{-7} torr range, start turning on the ion pump. Flip the switch to Start mode, turn on the pump—the pressure will spike. Turn it off if the pressure reaches the 10^{-5} torr range. Allow the pressure to recover, then turn on the pump again. Repeat this until the pump will stay on in Protected mode.

Every several hours, flash the TSP to help pump the chamber down. Once the pressure reaches the low 10^{-8} torr regime, change the sensitivity on the ion gauge to 1x. Close the right angle valve to the turbo pump—the chamber is now completely sealed. Once the desired pressure is reached, ramp down the variacs to 75% of their bake-out values, then 50% of that, then turn them off and allow the chamber to cool. Flash the TSP periodically during the cooldown.

The chamber pressure should reach the low 10^{-10} torr range—change the ion gauge sensitivity to 10x. If the pressure stabilizes above 5×10^{-10} torr, there is likely a leak. Using IPA, re-check the flanges. If the pressure fluctuates, either up or down, tighten the flange. In this vacuum regime, the IPA will sometimes clog a leak rather than leak through it, and the pressure will temporarily drop instead of go up, which is why any pressure fluctuations are suspect.

If the IPA test does not reveal a leak, attach the residual gas analyzer (RGA) to the side port and pump and bake it overnight. The RGA is sensitive to the mass of contaminants, so a spectrum can be obtained. Water appears at 18 m/z, and is a good indicator of the quality of the bake-out. Leaks appear at 28 m/z and 44 m/z, which is nitrogen or CO_2. If there is a strong water peak, but a weak N_2 peak, it's likely that the chamber needs to be baked longer. If both peaks are strong, there may be a leak, so a helium leak check is the next step.

Place a plastic bag of a large portion of the chamber, starting at the top. Track the helium concentration on the RGA, and slowly add helium gas to the bag until it begins to inflate. If the helium level on the RGA is unchanged, repeat the bag test over the next part of the chamber. If a leak is found, the helium peak will shoot up on the RGA. Allow the helium peak to recover prior to attempting to identify the specific problematic flange—the residual helium at the site of the leak can take some time to effuse into the chamber and may be temporarily saturated due to the bag test. Using the needle point attachment for the helium line, dip the end of it into a glass of water. Slowly increase the helium flow rate until the water bubbles at ~ 1Hz. Using the needle point, check around each flange, listening to the RGA whine out the helium level. If a leak is found, tighten the flange and check the pressure. If the pressure does not recover, it will likely be necessary to vent the chamber, replace the gasket, and perform the bake-out again.

For the first week at UHV, continue flashing the TSP daily. After that, it can be flashed weekly (depending on how often it's used).

CRYOGENICS

Once the pressure reaches $2x10^{-10}$ torr or lower, add liquid nitrogen (LN_2) to the outer dewar, then the inner dewar. The cold parts of the chamber will adsorb residual gases and the pressure will drop substantially. To prevent the majority of these species from adsorbing on the STM scanhead, the outer dewar should be filled first. The outer dewar should be refilled every 2 days, the inner dewar should be refilled once a week. A photograph of the inner dewar, when removed for maintenance, is shown in Figure 7.

To add liquid helium to the inner dewar, it's extremely important that there be no thermal leaks from the inner to the outer dewar. Any stray wires bridging the two dewars could cause significant losses of liquid helium gas. Center the inner dewar as well as possible—with the wobble stick, you can rock the inner dewar and see if it continues to sway back and forth, or if it is quickly damped. Adjust the centering until it sways as freely as possible.

Before the transfer, blow out the liquid nitrogen in the inner dewar with room temperature N_2 gas. When the inner dewar is completely empty, continue flowing N_2 gas until the temperature reaches ~90K—if there is any residual liquid nitrogen at the bottom, it will freeze upon adding liquid helium, and could cause problems. At 90K, remove the N_2 gas line and immediately begin the transfer of liquid helium. Time it for 3 minutes—if the helium filling isn't complete, it's likely that there is a thermal leak. Stop filling with helium and attempt to re-center the inner dewar. If this doesn't help, it's necessary to vent the STM and inspect for thermal leaks.

Track the cool down using the diode thermometer—once the inner dewar temperature has stabilized, the thermal leakage should be minimized. To do this, run a gas line from the inner dewar vent port to a syringe needle, then into a glass of water.

As the liquid helium boils off, it will cause bubbles in the water. Slowly adjust the centering screws on the STM, pausing between each adjustment. The goal is to minimize the rate of bubbling, as this indicates that the boil-off rate is minimized. The response time can be surprisingly quick, as liquid helium has an extremely low thermal capacity. The syringe needle is useful as it has a small orifice size—if the rate of bubbling is too large, remove the needle.

The rate of bubbling is actually an excellent indicator of whether there is still helium left in the inner dewar. Upon refilling, the bubbling rate will be relatively high. It will stabilize to a very low rate when it's ready for taking data. Once dewar is nearly empty, the boil-off rate will increase rapidly and the pressure in the chamber will begin to rise. These indicators can be used to decide when to refill the dewar again.

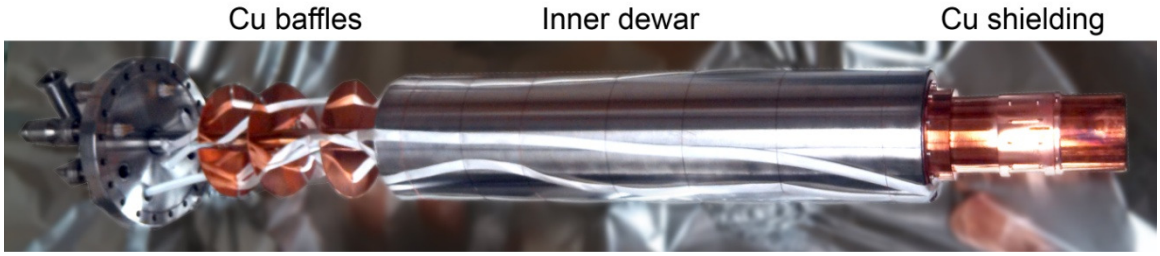


Figure 7. Inner dewar.

The inner dewar consists of a UHV vacuum flange, Cu baffles to radiatively isolate the inner dewar from the outside, the inner dewar itself, and Cu shielding to provide a thermally-equilibrated shielding around the STM scanhead, which is inside the shielding.

TIP PREPARATION

Preparing tungsten STM tips is a straightforward procedure that is well established in the literature; see [11] for an early example. The following is the technique I used to prepare reproducible STM tips, as illustrated in Figure 8:

1. Prepare a batch of 2M KOH.
2. Create a suspended droplet of KOH on a platinum ring.
3. Lower a tungsten wire, diameter 0.010", into the droplet.
4. Apply a 15V bias pulse to the wire. Bubbles will start to form. Turn off the bias before the bubbles become too intense.
5. Repeat the bias pulse until bubbles no longer form. At this point, the majority of the ions have been depleted and etching will proceed much slower.
6. Apply a large bias, 15V to 18V, until the tip decapitates. Recover the tip.
7. Promptly rinse the tip in DI H_2O.
8. Examine under optical microscope. If it's sharp enough that you cannot see curvature at 40x, the tip is likely good.
9. Transfer into the vacuum chamber as soon as possible.

This procedure is based on the principle of electrochemical polishing. At low bias, ~2-5V, electrochemical etching takes place, leaving a pitted surface behind. The pitted surface is fairly diffuse and is easily observable during tip approach, which is a

good thing, but it's also a less stable STM tip. Above 8V, the field lines become extremely sharp and cause micro-roughness on the tungsten tip to etch away, leaving behind a smooth surface.

Once transferred into the STM, it's important to e-beam clean it, as tungsten forms a non-conducting native oxide. To e-beam clean the STM tip, a tungsten filament is heated by passing a current through it. The tip is held at high voltage (typically ramped from 10V to 300V over several minutes), allowing hot electrons from the tungsten filament to bombard the tip. Once the tip is clean, it is transferred to the STM.

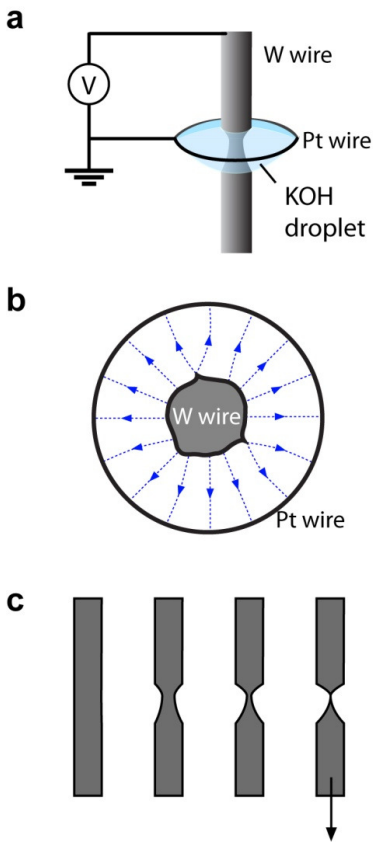


Figure 8. STM tip etching process.

a Tungsten (W) wire is threaded through a platinum (Pt) loop with a KOH droplet suspended in it. A bias is applied to the W wire, the Pt loop is held at ground. **b** The approximate E-field lines around a W wire. The surface roughness has been exaggerated for the illustration. At high bias, the field lines concentrate at bumps, causing them to etch faster. **c** The tip etching process is self-limiting: when the wire has etched through, it falls off, preventing over etching.

SAMPLE TRANSFER AND MANIPULATION

Cleaving samples *in situ* is fairly straightforward. In fact, preventing the sample from cleaving before reaching high vacuum is often a larger problem than achieving a good cleave. To prepare samples for cleaving, a small dab of epoxy is placed on a sample holder followed by the sample. Another small dab of epoxy is

placed on the sample followed by a stainless steel handle which is later used to cleave the sample. Epotek H20E has low outgassing and good electrical conductivity and was used for all of the cleavage samples used in this thesis. Typically, the epoxy is over-cured to reduce the outgassing in vacuum and ensure a strong bond. It's worth mentioning that, to prevent substantial thermal contraction, lower baking temperatures are recommended.

One difficulty in epoxying small parts together is holding them in place during the curing process. To solve this, I designed a mini-breadboard to support the samples during preparation, as shown in Figure 9.

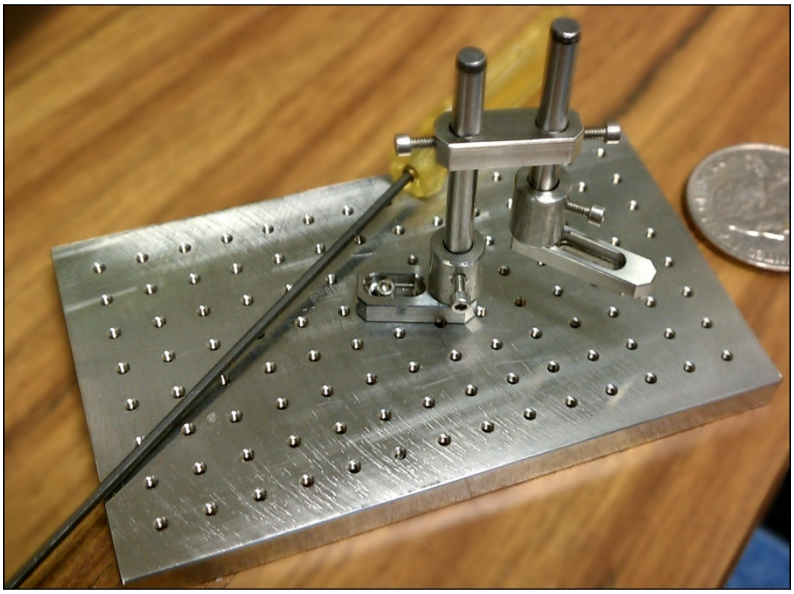


Figure 9. Miniature breadboard.

To hold samples in place while allowing epoxy to dry, as well as manipulation of small parts, I designed a miniature breadboard with supporting components. The holes are 0-80. The rods are standard 1/8" dowels which are commercially available. Many of the supporting components are miniaturized versions of parts available for optical tables.

PIEZO MOTORS

In order to achieve precise motion and maintain a rigid structure, piezo motors are used. By applying a bias to a piezo stack, the stack deforms. To move the STM tip or the sample, piezo stacks are used to grip the work piece. If a bias is applied gently, such as a slow ramp up, the work piece moves with the piezo stack deformation. Generally speaking, this is considered fine motion. If the bias is applied quickly, the piezo stack can move abruptly and, with the right conditions, slip—this is the same effect as pulling a tablecloth off from under a dining table set. Repeating this process leads to arbitrarily large displacements, and is referred to as 'coarse motion' for piezos motors. The process is shown schematically in Figure 10.

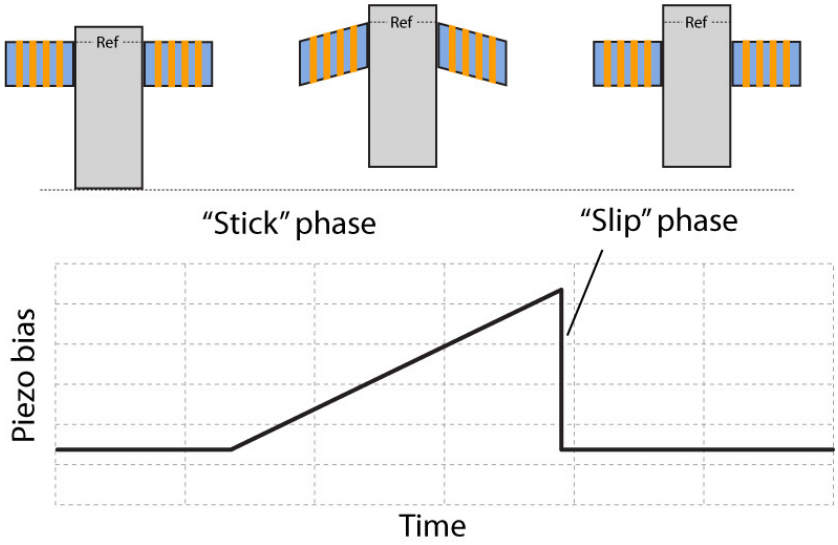


Figure 10. Principle behind stick-slip motion.

As the bias is ramped, the piezo stack deforms and the work piece moves with it. When the bias changes abruptly, the stack can move faster than the work piece, generating a net motion. Repeating this produces coarse motion.

Because a single deformation phase can be as small as a few nanometers, and the coarse-motion phase can be run virtually indefinitely, this process allows extremely fine motion over large length scales. Therefore, this process is used for bringing the STM tip into tunneling range (~1nm, sometimes less) of the sample. During the fine motion deflection, the STM tip is constantly measuring for a current. If the fine motion reaches the end of its stroke without achieving the predefined set point, it takes a coarse step and repeats. This process can be extremely slow if approaching from, say, 1cm away (on the order of 12 hours), so it's advisable that a coarse approach without searching for a feedback current is performed prior to the fine approach. It can be difficult to prevent the STM tip from coming into contact with the sample, so some tricks are described in the 'Lighting and optical access' section.

When the STM tip comes into physical contact with the sample, this is referred to as a 'crash,' and is highly undesirable—it blunts the STM tip, damages the sample, and typically means that it will be 24 hours until the next tip is ready for sample measurement. A variety of things can lead to this:

- Performing too aggressive of a coarse approach
- Having a non-conducting sample, or a sample that does not conduct in the bias range used for fine approach (make sure the voltage is larger than the material's band gap)
- Setting the current feedback set point too high
- Taking coarse approach steps that are larger than the fine-approach steps—the coarse step should be <75% of the fine-approach step, just to be cautious

- Not checking all wires for continuity after installing the insert (sometimes, the process of placing the inner dewar in its seat can cause wires to be tugged or jostled)
- Having a loose piezo motor or tip

Other problems can arise, as well, but these cover the majority of failures that I experienced during operation of the STM. In general, it's a good practice to test the STM prior to performing the bake-out: if a wire is loose or broken, this is a good time to discover it!

In order to achieve precise X-Y motion for the sample holder, a shear piezo stack (such as that described in Figure 10) is stacked on top of a second piezo stack at 90 degrees: this allows Stack 1 to perform X motion, Stack 2 to perform Y motion. Unfortunately, the X and Y motion can have cross-talk and lead to non-orthogonal motion. However, for the majority of STM experiments, this is not an issue.

LIGHTING AND OPTICAL ACCESS

For STMs with optical access, such as the one used in this study, it's possible to perform a coarse approach to within a few microns. Because the fine-approach process is time consuming, a good coarse approach can save many hours. However, this is not trivial and requires some mastery of fairly subtle details, particularly for sample stages with multiple degrees of freedom where misalignment is easy. Here, I provide a summary of some tricks of the trade:

- Look for the STM tip's reflection in the sample (Figure 11a & f). If the distance from the tip to the reflection of the tip is 2*d*, the distance of the tip to the sample is *d*, so this provides a fairly successful way to avoid crashing the tip.
- A good STM tip is actually quite reflective. Instead of lighting the STM tip, light the background (Figure 11c, e, & f).
- When possible, design in multiple viewports for orthogonal viewing angles.

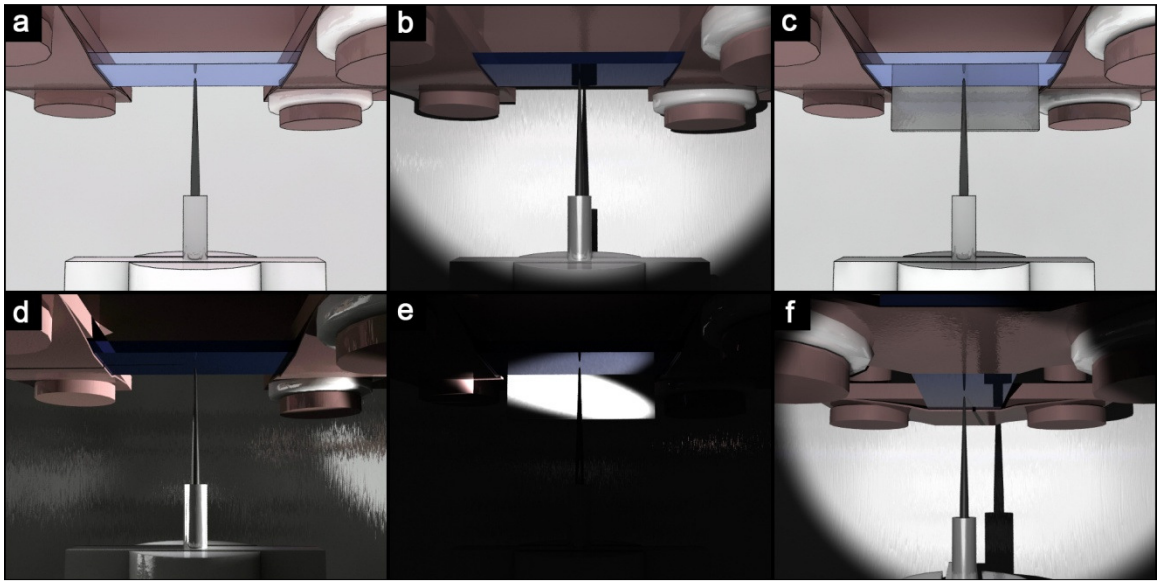


Figure 11. STM lighting configurations.

Several configurations are available for lighting the sample and STM tip, each has advantages and disadvantages. **a** Geometry of STM tip approaching a sample (blue), as well as its reflection in the sample. **b** Rendering of geometry in **a** with shadows, showing how the position of the STM tip can be difficult to determine. **c** Geometry of STM tip approaching a sample with a diffuse white background behind it. **d** Rendering of **a** with lighting from the side. While the shadows are now less of a problem than in **b**, it's difficult to see the reflection of the tip. **e** Rendering of geometry in **c** with the diffuse white background lit from the side. This setup is excellent for achieving extremely high signal, as a fiber optic can be focused onto the white background at any required intensity. **f** View of **b** from the side. This is the standard way to view the sample, as the length of the sample provides enough surface to actually see the reflection in, while the background is sufficiently bright, as well.

STM SCANHEAD

The STM topography mechanism is based on a feedback loop between the Z-axis and the tunneling current, as illustrated in Figure 12. For the STM used, the Z-axis is summed into the X and Y signals. The piezo tube principle is similar to the piezo stacks, where a bias induces a deflection. By holding the center at ground, and bias difference between the tube center and the outer wall is converted into a deflection. To achieve X-Y scanning, the scantube is divided into 4 quadrants: +X, -X, +Y, and -Y. Applying a bias to all 4 quadrants results in deflection in the Z direction. Therefore, in this piezo configuration, it's necessary to control Z-deflection by controlling an offset bias on all of the X and Y signals. Other piezo configurations are possible, but this configuration has the advantage that the interior of the tube is always at ground, thereby providing additional shielding for the tunneling current signal.

If the STM tip does not sense enough current, it will move closer to the sample until it achieves the desired set point. Similarly, if the STM tip senses too much current, it will retract until it finds the desired set point. Like many systems with active feedback, it's necessary to tune the integration time and gain—too much gain, or too short of an integration, and it's possible to set up an oscillatory conditions that can destroy the tip or sample.

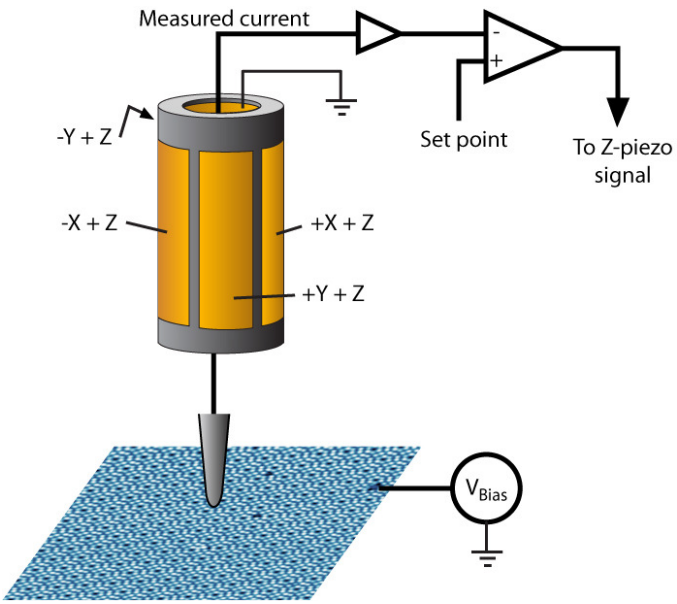


Figure 12. Illustration of tunneling feedback mechanism.

The sample is held at the requisite bias while the tip is held at ground. Any current flowing between the sample and the tip goes through a FEMTO DLPCA-200 preamplifier and is converted to a voltage. The voltage signal is then read by the Nanonis controller, which compares it with the desired set point. If there is not enough current, the controller extends the STM tip further; if there is too much current, the controller retracts the STM tip.

The scanhead used for LT-STM3 is an RHK Pan-type STM, shown in Figure 13. The piezo tube center is held at ground, so the Z-axis is summed into the X and Y quadrants. The sample stage has X and Y coarse motion, while the piezo tube is attached to a motor with Z coarse motion. In practice, the piezo tube approaches the sample from below. This way, if there is a sudden physical jolt on the system, the tip is more likely to fall *away* from the sample, rather than into it.

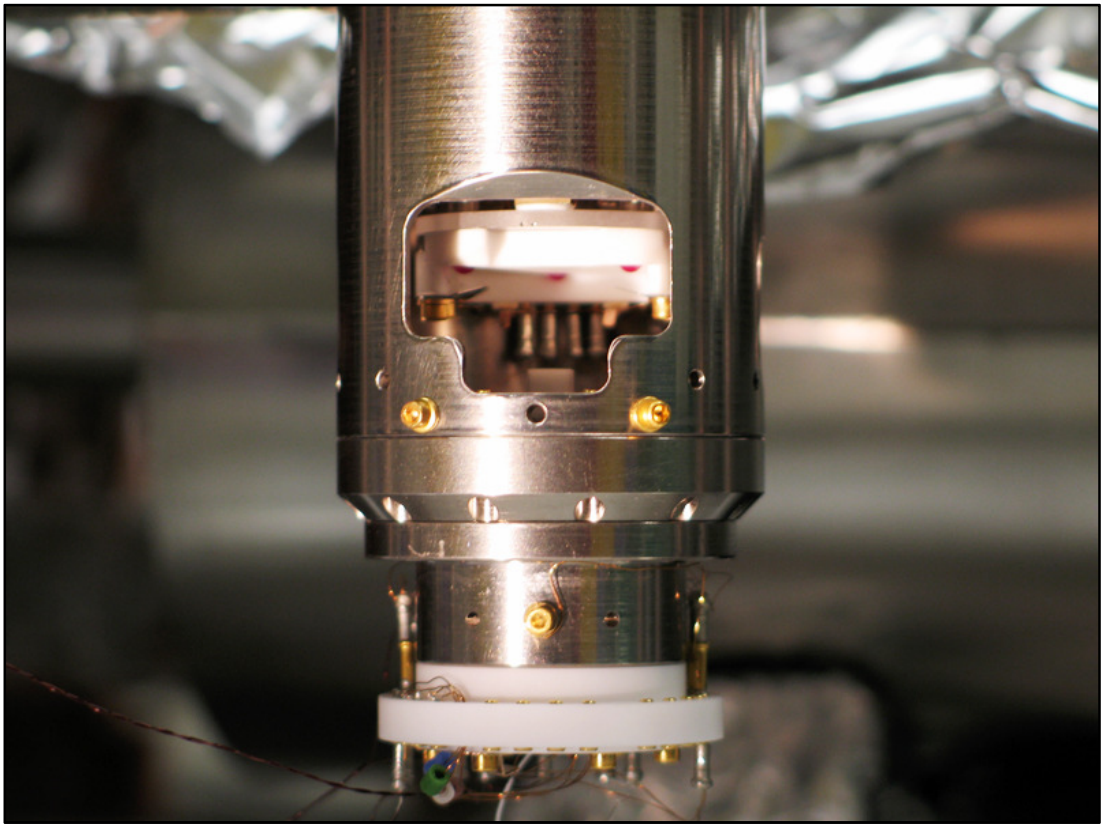


Figure 13. RHK pan-type STM

The STM used in our studies was manufactured by RHK. The body is titanium and the sample stage is macor (white interior part).

STM CONTROLLER

The STM scan controller used in these studies is a Nanonis design. The data generated can be viewed by most freely-available scanning probe analysis software, such as Gwyddion [12] and WSXM. [13] The Nanonis box uses digital control for all features, allowing active control of every channel. However, the following problems are known to be caused by digital controllers:

- +/- 1 LSB issues—DAC boards can only guaranteed accuracy to 1 least-significant bit. Consequently, for a 16 bit DAC, the 16^{th} bit may be wrong, leading to nonlinearities.
- Transient voltages on the DAC—DAC boards store numbers as 1's and 0's. Each 1 is a positive voltage, so switching from a 1 to a 0 necessarily has a lag. Consequently, for a switch between a number with many 1's to a number with many 0's, a significant transient voltage can occur.

We observed that the lock-in signal was able to detect digital artifacts which we have associated with transient voltages due to their proximity to known biases with many 1-0 transitions. A consistent example is the measurement of a 'gap' in the dI/dV signal at 0V for just about any sample (or for no sample at all). To circumvent this issue, we found that either the numeric derivative of the signal (for dI/dV curves) or external, analog lock-ins, could remove this issue in some data sets. However, some issues remained in the Z-piezo signal that was, somehow, transferred to the current signal. To illustrate the problem, a bias sweep with simultaneous lock-in for the Z-piezo channel was recorded, as shown in Figure 14. For a perfect system, this would be a flat response curve instead of the reproducible structure shown below. Additional discussion of these artifacts is provided in Chapter 7.

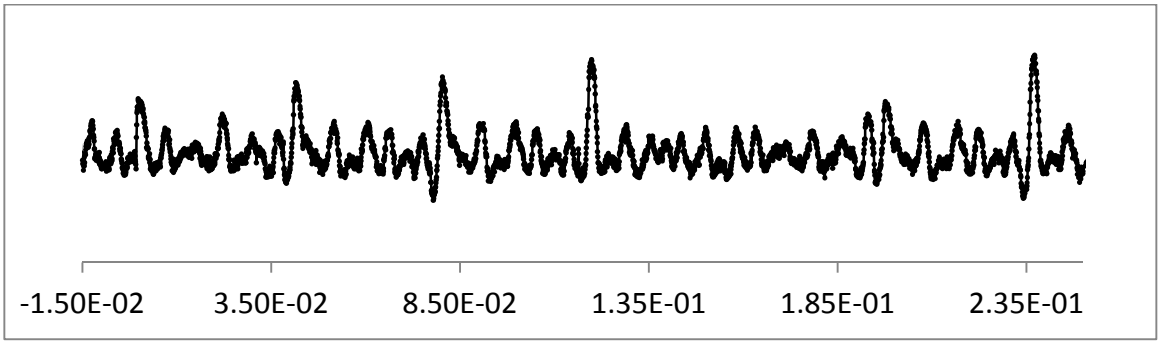


Figure 14. Nonlinearities in the lock-in signal from a Nanonis controller.

Sweeping a channel on the Nanonis controller and simultaneously measuring the lock-in signal shows reproducible structures. These are likely the result of transient voltages in the DAC, but could be related to other digital artifacts.

Chapter 3: Topological insulators

INTRODUCTION

Topological insulators (TIs) are a member of a new class of materials where the *boundary* of a material hosts a topologically protected metallic surface state, due to the topology of the *bulk* of the material. While the bulk of a TI has a conduction band and valence band, like a semiconductor or insulator, the boundary state is gapless and protected by time-reversal symmetry. [14] Therefore, to understand the topological insulator, and why it's considered a new electronic state of matter, it's necessary to discuss topological properties in condensed matter physics. For a more complete introduction to topological insulators, see [14–16].

Topological properties of insulators

Semiconductors and insulators both share similar bulk electronic structure: simply, they consist of a filled band and an empty band separated by a gap. Therefore, they are topologically equivalent: by tuning the gap, you can continuously tune between an insulator and a semiconductor. [14] More fundamentally, if we think of these systems simply as energy states with a gap, then the HOMO and LUMO of the hydrogen molecule can be included in the same topology. However, not all gapped electronic systems are topologically equivalent.

Fundamentally, an insulator does not conduct electricity because all valence band states are filled and all conduction band states are empty—there are no free carriers. There is a gapped system, where all 'valance band' states are filled, and all 'conduction band' states are empty, yet a net current can result: the quantized Hall

system. [17] When a 2D electron gas (2DEG) is placed in a strong magnetic field, the electron states become quantized Landau levels; in the presence of an electric field, the cyclotron orbits drift and produce a net current. [14] However, the 'band structure' looks equivalent to an insulator, illustrating that other considerations can be necessary to describe electronic properties of some systems. In particular, the difference between a conventional insulator and one exhibiting a quantized Hall response is topological in nature.

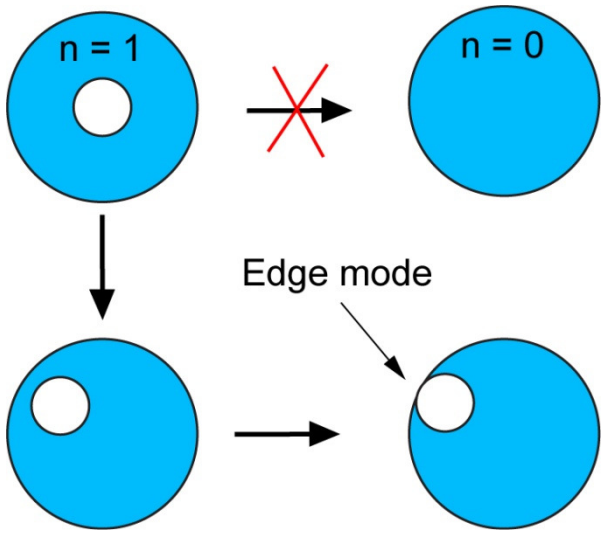


Figure 15. Topological sets and formation of edge modes.

States with different topologies cannot be continuously transformed from one to the other. At the interface of two bands with different topologies, it's impossible to rectify the different topologies continuously: instead, a boundary mode is formed.

Two systems are considered topologically equivalent if there is a continuous, 1-1 transformation from one to the other without introducing any jumps or discontinuities in the bulk electronic structure. At the interface of a quantum Hall state with topological invariant (Chern number) n = 1, and an insulator with n = 0, the energy gap must vanish, otherwise the two states would be topologically

equivalent. [14] Consequently, an edge mode forms, as shown schematically in Figure 15. In the Quantum Hall Effect (QHE), there is a simple physical interpretation of this edge mode, as shown in Figure 16.

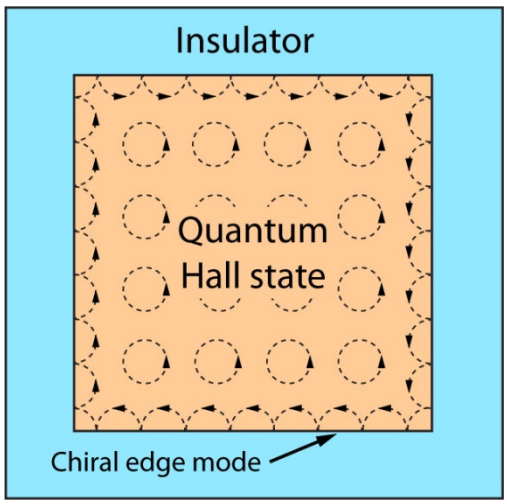


Figure 16. Edge modes in a 2D Quantum Hall system.

The cyclotron orbits in the Quantum Hall state turn into half-cyclotron orbits at the boundary and can support a direction current. Due to the chirality of the edge mode (it only goes in one direction), a scattering mode cannot backscatter the carrier.

In the integer QHE, the coupling is between the charge of the electron and the external magnetic field. If the external magnetic field in the Quantum Hall state is replaced with strong spin-orbit coupling, it's possible to achieve a similar system with a protected edge mode; this is a 2D topological insulator. [15]

Unlike the external magnetic field in the QHE, spin-orbit coupling is not a 2D effect and can be generalized to three dimensions. In a 3D TI, instead of generating a topologically-protected state on the edges, the entire surface is now the interface between the insulator and 3D TI—the surface supports the topologically protected modes.

While the edge modes of 2D TIs have 1D dispersion, the surfaces of 3D TIs have 2D dispersion. A useful model for this surface state was described by Fu [18], explaining the Dirac-like dispersion observed by [19,20] in Bi_2Te_3, a model 3D TI. In [18], the model Hamiltonian for the topological surface state is given by:

$$H(\vec{k}) = E_0(k) + v_k(k_x\sigma_y - k_y\sigma_x) + \frac{\lambda}{2}(k_+^3 + k_-^3)\sigma_z \qquad \text{Eq. 13}$$

where v_k is the Dirac velocity, σ_j is the Pauli matrix for spin in the j direction, λ is the warping parameter, and $k_\pm = k_x \pm ik_y$. To interpret this, first consider $\lambda = 0$. Without warping, the Hamiltonian describes a system where the wavevector, k_j, is always perpendicular to the spin—this is an example of spin-polarized transport and, similar to the case of the edge modes in the QHE, is a chiral state. There is an interesting consequence of this: backscatter suppression. In the case of backscattering, time-reversal symmetry requires that the incoming and outgoing spins be identical (barring a symmetry-breaking scattering center, like a magnetic defect). However, according to the above Hamiltonian, the available spin states at $+k_j$ and $-k_j$ are antiparallel. Therefore, backscattering is not an allowed process in topological insulators.

For $\lambda \neq 0$, backscattering is allowed through a 3rd order process: λ allows warping along the z-axis, so while the x and y components of spin at $+k_j$ and $-k_j$ are still antiparallel, the z components may not be. Therefore, understanding whether the Dirac cone is warped is important to transport properties, as it determines the degree of backscattering.

The dispersion relation from the Hamiltonian in Eq. 13 is:

$$E_{\pm}(\vec{k}) = E_0(k) \pm \sqrt{{v_k}^2 k^2 + \lambda^2 k^6 \cos^2(3\theta)}. \qquad \text{Eq. 14}$$

In the limiting case of no warping (λ), this reduces to the equation for a cone. The dispersion for E_+ is plotted in Figure 1Figure 17. From a materials perspective, one consequence of this is that topological insulators, despite having a bulk gap, will always have a conductive surface state.

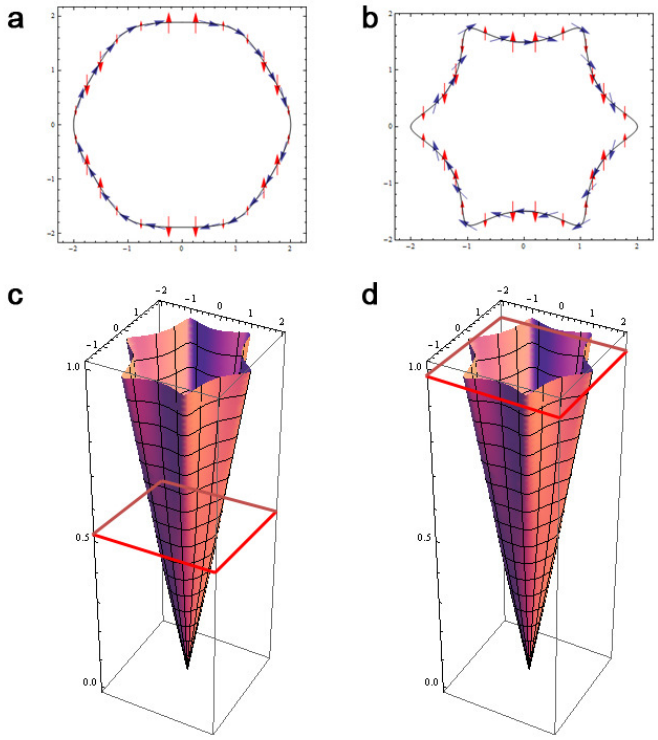


Figure 17. Dirac cone and spin components with warping.

Contours computed from Eq. 14. **a**,**b** Contours from the Dirac cone show the *x*- and *y*-components of the spin as blue arrows, *z*-component is shown as red arrows. For slight warping (**a**), the spins are essentially tangent to the Dirac cone—the *z*-component is exaggerated. For strong warping (**b**), the spins are no longer tangent to the Dirac cone and the *z*-component warping is quite strong. **c** and **d** show cross sections through the cone that correspond to the contours in **a** and **b**, respectively.

RELATIONSHIP TO THIS THESIS

The model topological insulator Bi_2Se_3 has both a band gap and a topological surface state which appears to follow Eq. 14. Unfortunately, it has two major problems in actual materials: a strong surface band bending field pulls the conduction band below the Fermi level, making the 'insulator' a conductor, and the surface is non-uniform, exhibiting potential fluctuations which would change the carrier helicity. Furthermore, in Cu-doped Bi_2Se_3, the incorporation of the Cu into the lattice isn't well-understood and has consequences for the surface states.

In this thesis, I have attempted to address these issues through STM investigation. In Chapter 4, I describe the basic impurities and structure of Bi_2Se_3 and describe results from cleaved, bulk Bi_2Se_3. Due to the large number of impurities in bulk Bi_2Se_3, a number of groups have considered preparing samples through molecular beam epitaxy (MBE). Chapter 5 describes STM results on MBE Bi_2Se_3 grown by colleague Yuxuan Chen. Chapter 6 examines defect states in bulk Cu-doped Bi_2Se_3, where we found that intercalants physically repel each other, possibly paving the way towards more uniform doping schemes. Finally, in Chapter 7, I describe a trick we discovered for examining the band bending field with nanometer resolution. Furthermore, we found that we can correlate potential fluctuations to specific point defects, namely Se vacancies. With these insights, Bi_2Se_3 is on a better track toward being the model system that it's expected to be.

Chapter 4: STM of Bi_2Se_3

INTRODUCTION & BACKGROUND

Bi_2Se_3 is a rhombohedral crystal with space group $D_{3d}^5(R\bar{3}m)$ that has a lamellar structure of quintuple layers. [16,21] Each quintuple layer has two equivalent Bi atom sites and three Se sites, two of which are equivalent, [16] as shown in Figure 18. The quintuple layers are separated by a van der Waals gap—the Se1-Se1 bonds are much weaker than the other bonds in the system. [22] Therefore, to prepare the sample for study by STM, the sample is simply cleaved in ultra-high vacuum—the cleavage plane is at the van der Waals gap. The surface that's revealed is the (111) plane and consists of a triangular lattice of Se atoms, as seen in Figure 19.

Known native defects include Se vacancies [23] and Bi substituted for Se, or Bi_{Se} antisites defects [24]. The lattice is surprisingly susceptible to distortion due to defects, so individual atomic defects often manifest as large lattice distortions, such as the Se vacancies observed in Figure 19. The system was originally of interest as a semiconductor and thermoelectric material, but has gained additional interest due to its topological insulator status. [16,21] In particular, the spin-orbit coupling in Bi and Se is strong enough that it cannot be neglected in the Hamiltonian: Bi_2Se_3 is a prime example of a topological insulator, so our attention turns to its surface state.

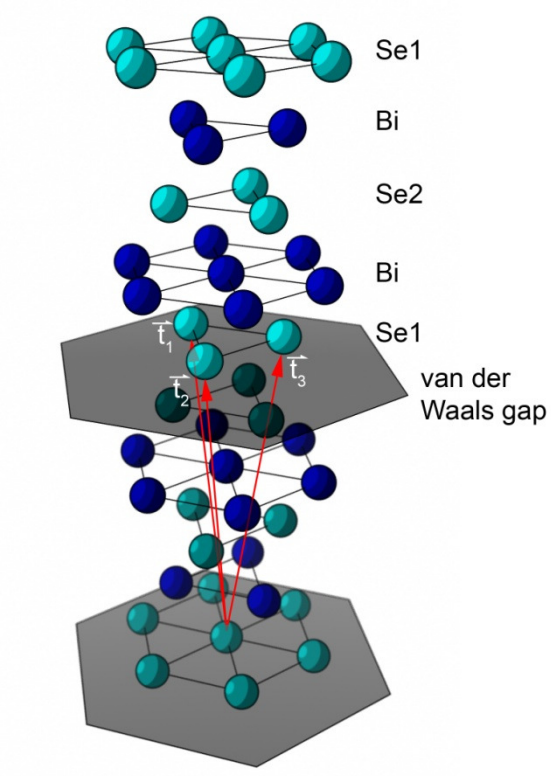


Figure 18. The structure of Bi_2Se_3.

The atomic structure of Bi_2Se_3 is shown above. The two equivalent Se positions, labeled Se1 and Se2, have different bonding configurations. The Se1 position is bonded to 6 other in-plane Se atoms and 3 neighboring Bi atoms, while the Se2 position is bonded to 6 other in-plane Se atoms and 6 neighboring Bi atoms—a vacancy at this position would break many more bonds than in the Se1 position. Both Bi positions are chemically equivalent. The Se1-Se1 bonds are van der Waals in nature, and are much weaker than the other bonds, making this a favorable cleavage plane. The lattice vectors, $\{\vec{t_1}, \vec{t_2}, \vec{t_3}\}$, are shown in red.

STM is especially well suited to study the surface states of conducting materials, so the topological surface state in Bi_2Se_3 should be readily accessible by STM. Unfortunately, the bulk conduction band also has a presence on the surface near the Fermi level, complicating the analysis. Topographic data near the Fermi level will reveal both the non-trivial surface state, but also the conduction band states (Figure 21). However, it is possible to investigate the surface state itself: dI/dV data in the bulk gap will reveal the gap states without bulk contributions (to first order).

Interestingly, for STM, the surface state will not reveal all of its secrets on a featureless surface: defects are a good thing. By examining how the surface state interacts with defects, it's possible to draw conclusions about the electronic structure. Therefore, the starting point for any investigation of surface states should be at the defects that the surface state interacts with.

POINT DEFECTS

Due to the high vapor pressure of selenium, preparing a stoichiometric combination of 2Bi + 3Se does not result in a stoichiometric Bi_2Se_3 crystal—much of the selenium ends up leaving the crystal. Consequently, it's common to end up with a large number of selenium vacancies (V_{Se}).

Early scanning tunneling microscopy (STM) reports of the surface of Bi_2Se_3 revealed two primary forms of native point defects: triangular topographic features and purely electronic defect states. [25] The triangular defects have been identified as Se vacancies which n-dope the sample, [26] and are prominently visible in Figure 19 and Figure 20a. The second defect type is the Bi_{Se} antisite (where a Bi atom replaces a Se lattice site) and is best observed near or below the Dirac point by *dI*/*dV* imaging (Figure 20b), which provides a map proportional to the local density of states at a given bias. The peculiar features observed for Bi_{Se}, like those shown in the atomic resolution topograph in Figure 20b, has led some recent studies to attribute them to surface-bulk scattering near the Dirac point [27]—however, angle resolved photoemission spectroscopy (ARPES) results have not observed a side lobe to the valence band, necessary for this interpretation, and the feature geometry does not change at different biases (Figure 22), which is inconsistent with the proposed

scattering states. We therefore attribute the defect features to impurity states unassociated with bulk-surface scattering.

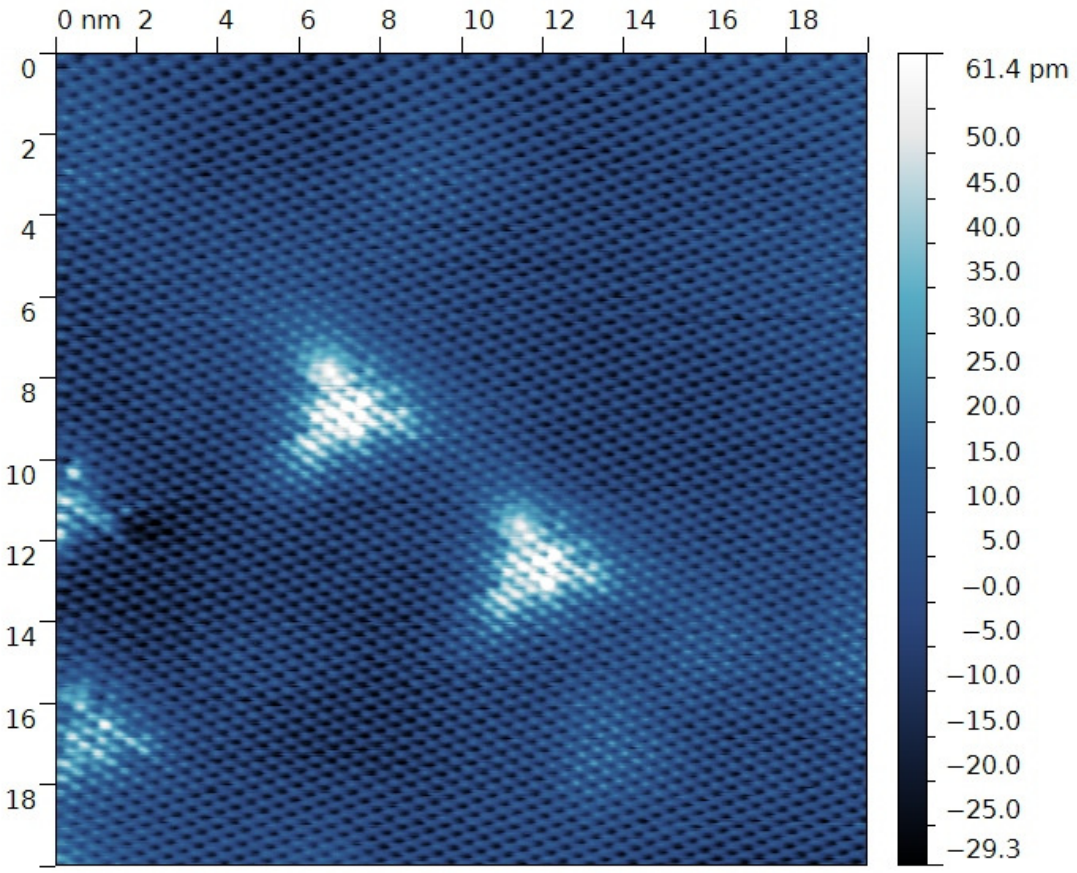


Figure 19. STM topograph of Se vacancies on Bi_2Se_3.

Upon cleaving, the Se1 (111) plane of Bi_2Se_3 is exposed—these atoms make up the lattice observed in the above figure. Sub-surface defects, such as the Se vacancies above, distort the atomic lattice and appear as topographic protrusions. These defects are likely on the bottom of the 1st quintuple layer or the top of the 2nd quintuple layer.

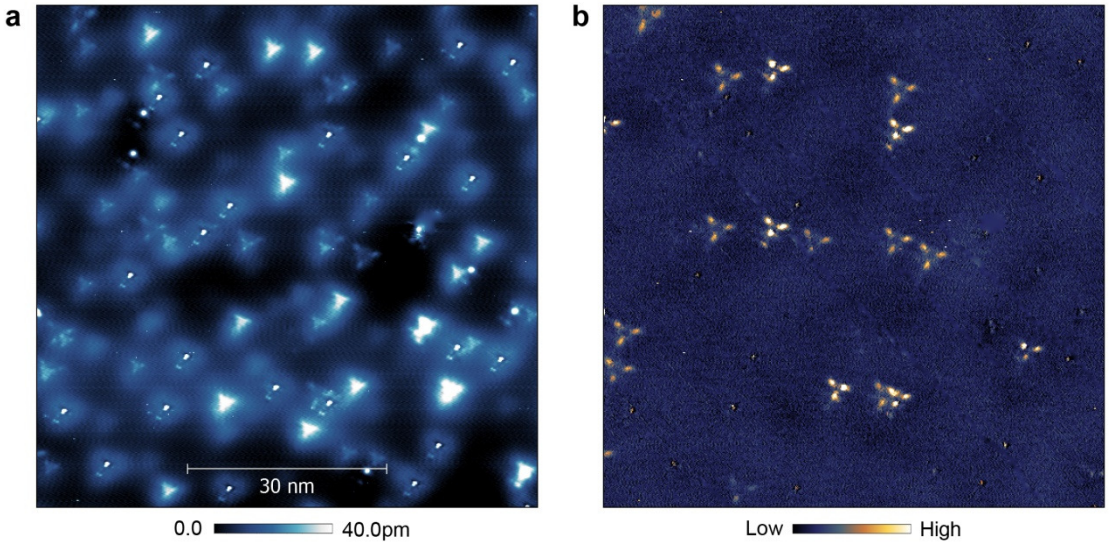


Figure 20. Bi_2Se_3 topography and dI/dV images.

The STM topography, **a**, and corresponding dI/dV image, **b**, show the variety of native defects in nominally stoichiometric Bi_2Se_3 at -380meV and 500pA.

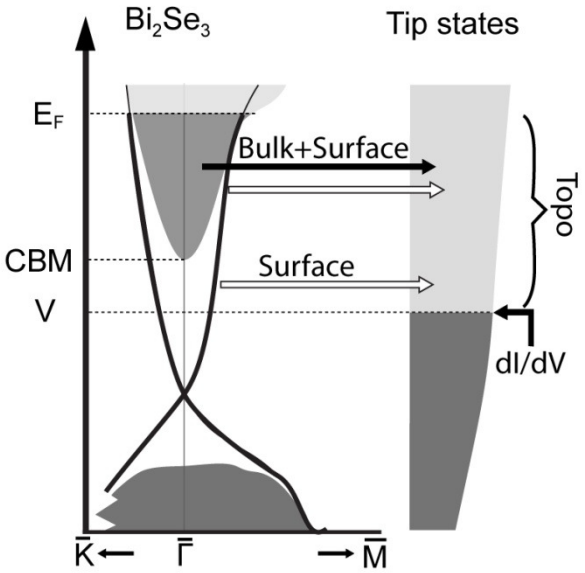


Figure 21. Tunneling cartoon for Bi_2Se_3.

By shifting the sample bias, the tip and sample states are offset from each other, allowing a net current between the STM tip and the Bi2Se3 states. Both bulk and surface states are sampled by the STM.

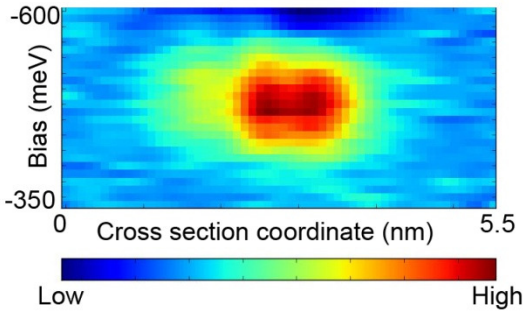


Figure 22. Cross section spectroscopy down the lobe of a BiSe defect.

The feature geometry is not strongly energy dependent. A dispersive feature, such as those associated with quasiparticle interference, would appear as a diagonal across this kind of plot.

ELECTRONIC PROPERTIES

Bi_2Se_3 is a semiconductor with a bulk gap, but it's also a topological insulator with a non-trivial surface state. As shown in Figure 23, the dI/dV curve indicates that the bulk states are separated by ~250meV. Between the bulk bands the surface state has a minimum near -375meV—this is regarded as the Dirac point, but due to tunneling nonlinearities, may not correspond quantitatively with the actual Dirac point position. The position of the Dirac point is important to technological applications: in order to create carriers with a helicity, the Fermi level must be tuned to the right part of the gap. Because the helicity is of opposite handedness above and below the gap, fluctuations of the Fermi level would result in different helicity in different regions on the sample. Chapter 6 is dedicated to examining one form of dopant that can tune the Fermi level. If interference between carriers of a well-defined helicity is desired, understanding how the potential fluctuates is essential—Chapter 7 discusses the influence of defects on the Dirac point position.

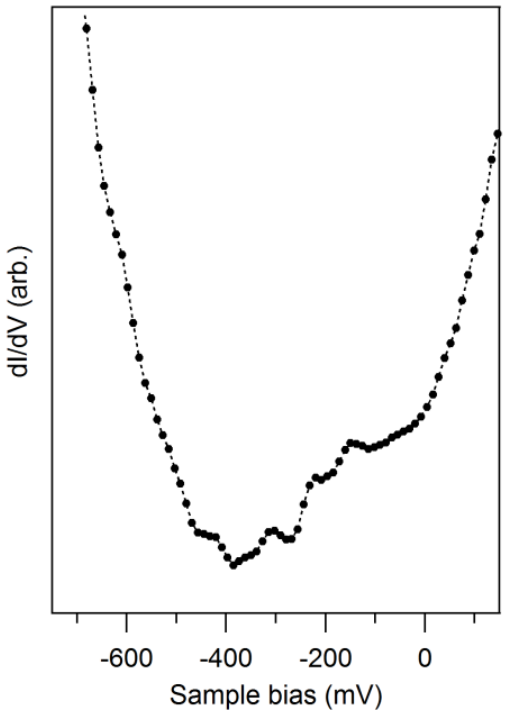


Figure 23. dI/dV from undoped Bi_2Se_3 showing a Dirac point near -375 meV.

The dI/dV structure shows a Dirac point near -375 meV. The conduction band minimum is near -250meV, while the valence band onset is near -500meV. This data was determined through the numeric derivative of the current data to avoid digital artifacts in the lock-in amplifier. A Savitzky-Golay filter was used to take the derivative and smooth the data, though noise remains in this dI/dV curve.

Selenium vacancies are sufficiently common that researchers are investigating additional ways to prepare the samples. One example of this is molecular beam epitaxy, where atomic fluxes of the species to be deposited are impinged on a crystal substrate. Under the right conditions, it's possible to achieve epitaxial growth. The following chapter discusses the epitaxial growth of Bi_2Se_3.

Chapter 5: STM & MBE of Bi_2Se_3

Author Contributions

This section describes work done in collaboration with Yuxuan Chen, a colleague in the Shih Lab. Yuxuan was responsible for the MBE growth of Bi_2Se_3, while Chris Mann was responsible for the STM analysis and conceiving and demonstrating combinatorial substrates.

Combinatorial Substrates and Molecular Beam Epitaxy

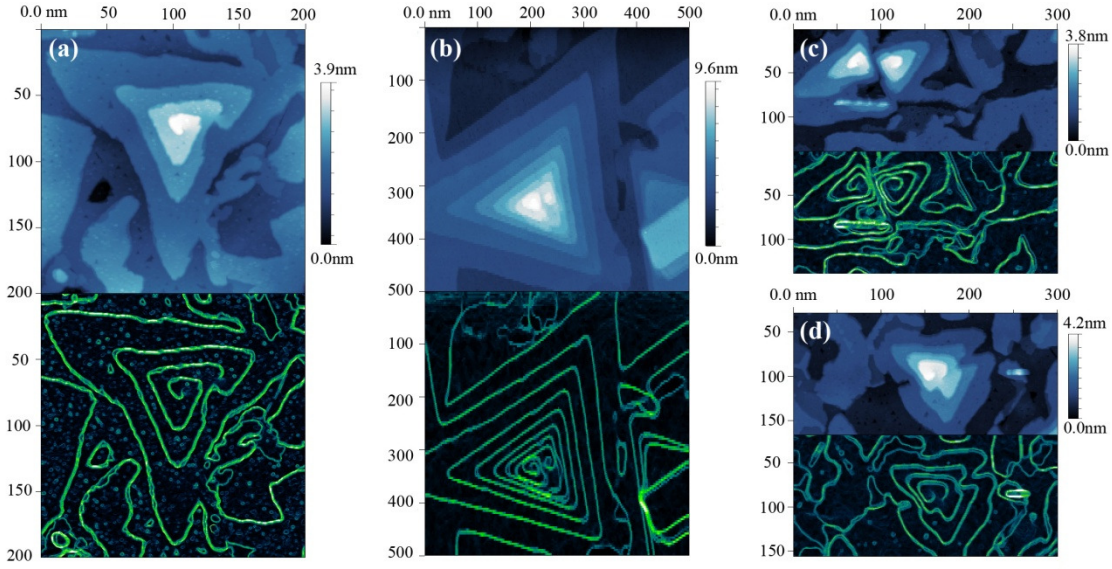


Figure 24. Spiral growth of Bi_2Se_3 on Si(111).

STM topographs, blue, and edge-detection, green, of spiral growth modes of Bi_2Se_3 on Si(111). Spiral formation often leads to pairing which can form extended pyramid structures.

Molecular beam epitaxy (MBE) is the process of growing a crystal by vacuum deposition onto a crystalline substrate. The orientation and atomic step structure of

the substrate alters the growth, leading to different growth morphology—for instance, in the case of GaN, a spiral growth mode can be suppressed by using vicinal substrates, allowing the step-growth mode to dominate over the spiral growth mode. [28] When a new material system is first investigated for MBE growth, the parameter space that must be studied is enormous: substrate temperature, substrate orientation, substrate step structure, deposition rates, precursor processing, and so forth. Here, we report a combinatorial technique to greatly reduce the parameter space: a textured substrate that allows the simultaneous study of a variety of atomic step structures. By fabricating microdomes into an epitaxial substrate, we show that it's possible to reproducibly generate vicinal regions, evenly-spaced atomic steps, and step-free regions. Because these studies are performed simultaneously, the use of a combinatorial substrate removes uncertainty due to unintentional variations from one growth to the next and is well suited for research MBE.

The principle form of a combinatorial substrate is a microtextured crystal. By preparing a structure in the geometry of a convex dome, multiple miscuts are simultaneously produced around the dome. The as-prepared substrate can be relatively rough, so an annealing process is necessary. In our case, we studied Si(111), where the annealing process allows each face of the microdome to relax into a thermodynamically-preferred orientation. The result is a variety of miscuts being simultaneously formed around the dome. Cylinder-type structures allow better isolation of specific step directions, but we have found their application to be more limited.

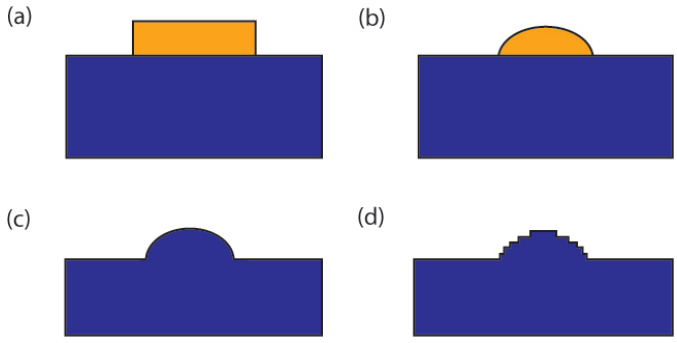


Figure 25. Process flow for fabricating combinatorial substrates on Si.

(a) Lithography determines the geometry of photoresist pillars, (b) which are exposed to acetone vapor and subsequently reflow into dome-like shapes. (c) The photoresist pattern is then transferred into the Si substrate by dry etching in SF_6. After cleaning the substrate, it is annealed to generate step structures (d).

As a case study, we prepared Si microdomes through reflow of photoresist pillars, followed by transfer into the Si substrate by plasma etching, illustrated in Figure 25. [29] The substrates are cleaned by sonicating in acetone and isopropyl alcohol, followed by 30 minutes of ozone exposure. The substrates are then transferred into a vacuum chamber and resistively heated until the surface oxide has been removed, leaving a clean 7x7 surface. The surface is then annealed at ~1200^{o}C, resulting in the structures seen in Figure 26(a). When the ozone cleaning step is skipped, residual carbon on the surface results in pinned steps and large nodules, as shown in Figure 27.

The step structure depends on the initial substrate miscut, but for a given miscut, the results are reproducible. For high-miscut Si(111), which typically prefers step bunching, we can generate uniform atomic steps along [1 1 -2]. Along [-1 -1 2], vicinal steps appear with. At the top of the microdome, despite the substrate being vicinal, large step-free regions are created. The size of this flat structure depends on

the annealing conditions, but can reach upwards of 20µm x 20µm for larger microstructures.

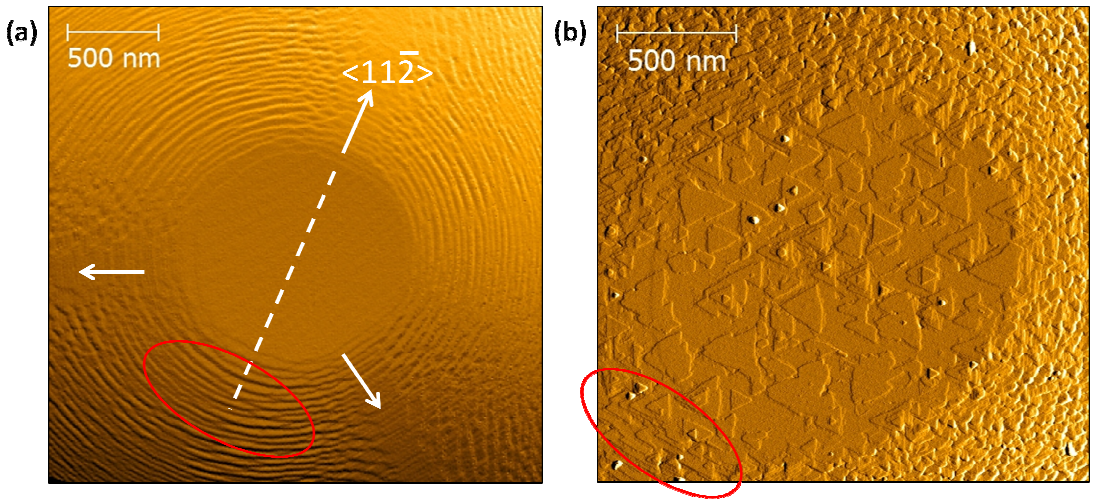


Figure 26. AFM of the top of microdome before and after growth.

(a) AFM amplitude-error image of the top of a Si dome after flashing to reconstruct the surface, showing the step orientation dependence and a large flat region at the top. (b) AFM amplitude-error image of 6QL of Bi_2Se_3 grown on a similar Si dome. In the central region, nucleation and some spiral growth is observed. On the vicinal regions, the [1 1 -2] directions do not provide particularly smooth growth despite the uniform steps, while the [-1 -1 2] directions appears to provide relatively linear growth fronts. AFM data compliments of Yuxuan Chen.

As observed in Figure 26(b), the most uniform growth fronts are not observed along the uniform step direction; rather, the regions with bunched steps have the most linear growth fronts. The step-free region at the top demonstrates nucleation and growth of triangular structures. Significantly, not all of the triangles point in the same direction. This indicates that, for flat growth, it's possible to end up with domain rotation at the Se-Se van der Waals interface between quintuple layers. The ability to demonstrate growth mode dependence on multiple substrate morphologies within a single AFM scan frame is remarkable.

One drawback of this technique is that strain can be introduced by the curvature of the substrate for thicker samples. In particular, any differences in the coefficient of thermal expansion between the substrate and the deposited material could lead to morphological changes at the characterization temperature, which would be amplified by the curvature of the microdome. However, for nucleation and early-growth studies where local forces dominate the morphology, as well as *in situ* studies, the use of a convex (or concave) substrate should play an insignificant role.

We have demonstrated a combinatorial substrate for epitaxial growth. The ability to consolidate multiple substrate miscuts into a single region allows a significant reduction in the parameter space required to investigate the epitaxial growth of new compounds. Techniques already exist for preparing similar microstructures on a variety of substrates, so we expect that this technique will find broad application in research MBE.

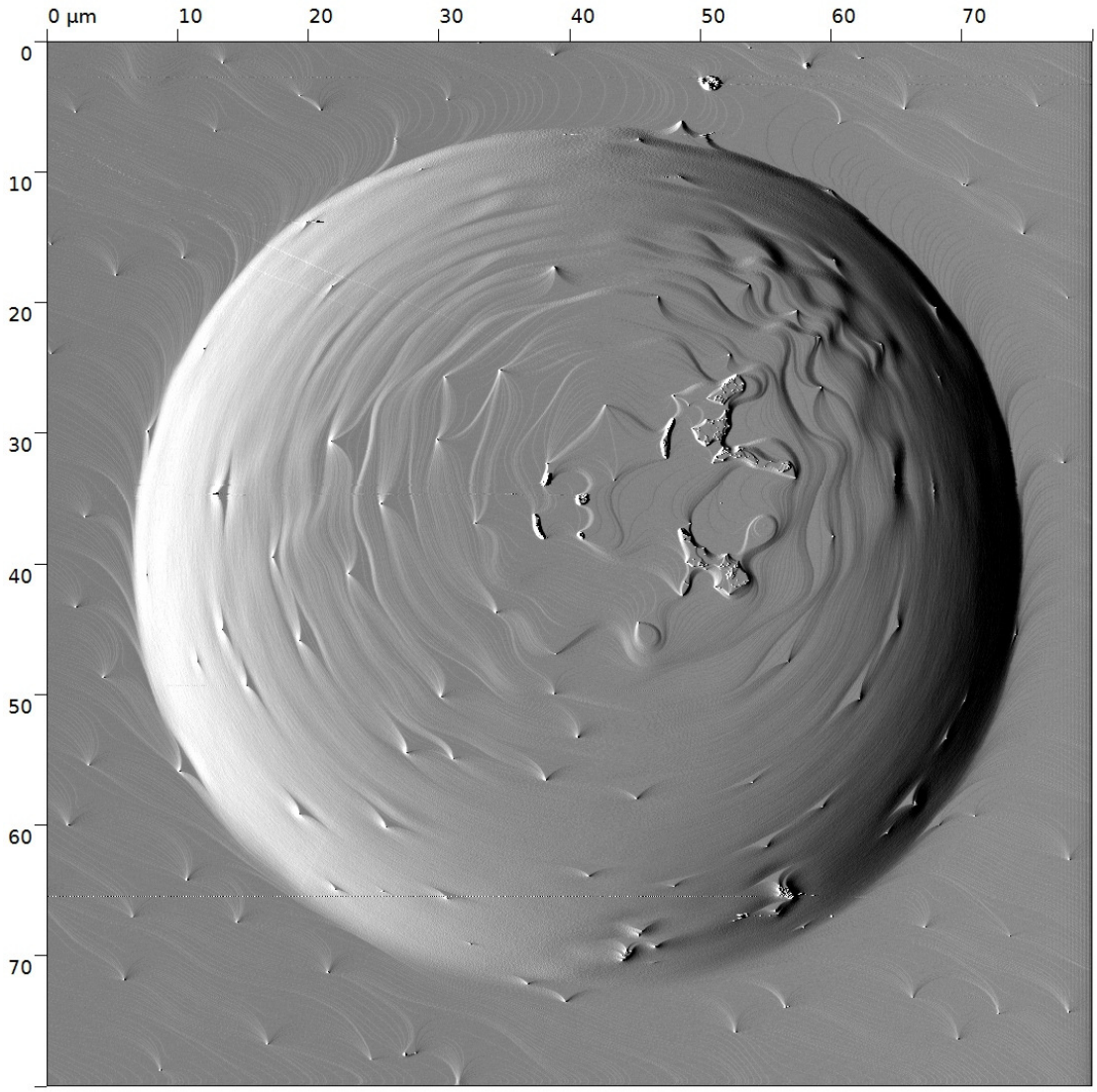


Figure 27. Amplitude error image of an unclean microdome.

The ozone cleaning step is necessary to remove the residual photoresist. The residual photoresist can pin the steps during the anneal and result in the structures seen above.

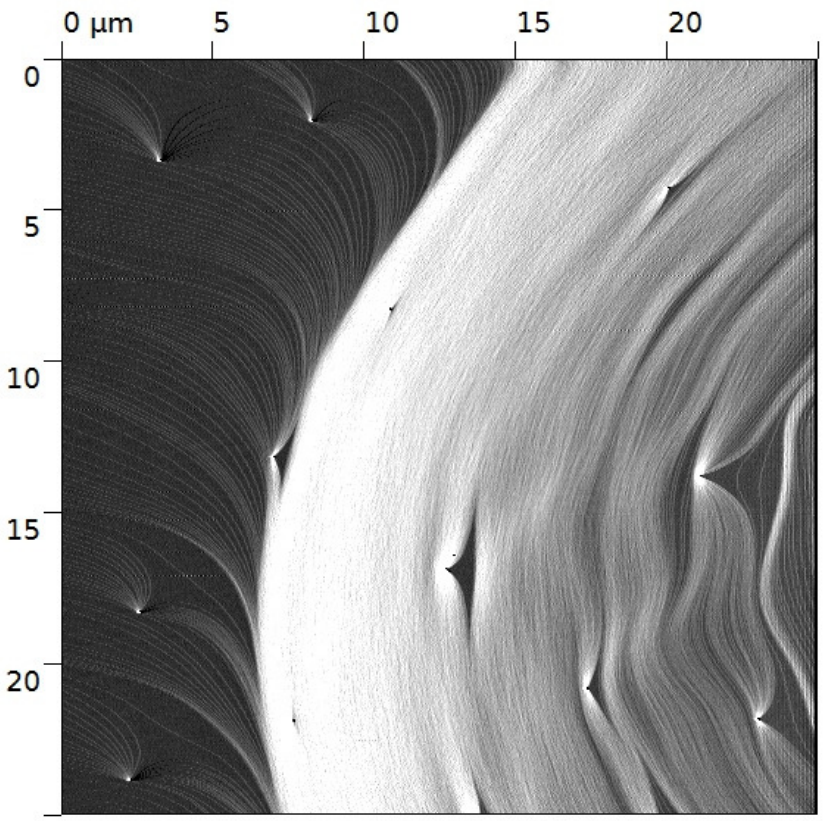


Figure 28. Close up of Figure 27.

Individual atomic steps are visible in amplitude error images, as shown in this subsection of Figure 27. The pinning sites cause ripples and nonuniformities in the step structure.

Chapter 6: Pair correlation analysis of $Cu_xBi_2Se_3$

AUTHOR CONTRIBUTIONS

The following chapter was performed in collaboration with Dr. Damien West and Professor Shengbai Zhang of Rensselaer Polytechnical Institute and Dr. Ireneusz Miotkowski and Professor Yong P. Chen of Purdue University.

All authors contributed to the intellectual content of this section. Ireneusz Miotkowski and Yong Chen provided the samples. Damien West performed and analyzed density functional theory simulations under the guidance of Shengbai Zhang. Chris Mann collected and analyzed the STM data and prepared the text, Chih-Kang Shih advised on the experiment and provided input on the data analysis and the manuscript.

INTRODUCTION & MOTIVATION

To produce bulk-insulating topological insulators, Bi_2Se_3 has emerged as a particularly strong candidate due to its bulk gap. [21] However, strong band bending at the surface has limited its viability, as the bulk conduction band is pulled down below the Fermi level and allows non-topological carriers. [30] Recent fabrication efforts have focused on techniques to tune the band bending; compensating dopants have proven to be good candidates. [31] One compensating dopant is Cu: while many studies find that Cu is an n-type dopant in Bi_2Se_3, other studies have found that it can be used to prevent n-type doping—Cu is an amphoteric defect, where intercalated Cu is expected to n-dope the sample and Cu_{Bi} substitutions are expected to p-dope the sample. [32] Here, we investigate the role of Cu with an atomic-scale resolution using

scanning tunneling microscopy and *ab initio* simulations. We find that Cu_{Bi} and interstitial defects show minimal interaction with the surface state, while Cu intercalants can give rise to quasiparticle scattering states. Additionally, through radial pair correlation analysis, the Cu intercalants exhibit repulsive potentials. Furthermore, the physical size of these defects is sufficiently large that traditional quasiparticle scattering analysis fails.

Samples were grown using a Bridgman-type method similar to [33] and were prepared for analysis by cleaving in ultrahigh vacuum (UHV) at room temperature, then immediately transferred to the STM at 77K. The room temperature cleave allows surface Cu, previously intercalated, to rearrange. We observed ~500pm nodules on the surface that we attribute to Cu clusters, shown in Figure 29. All high resolution data was acquired in regions away from these large defects.

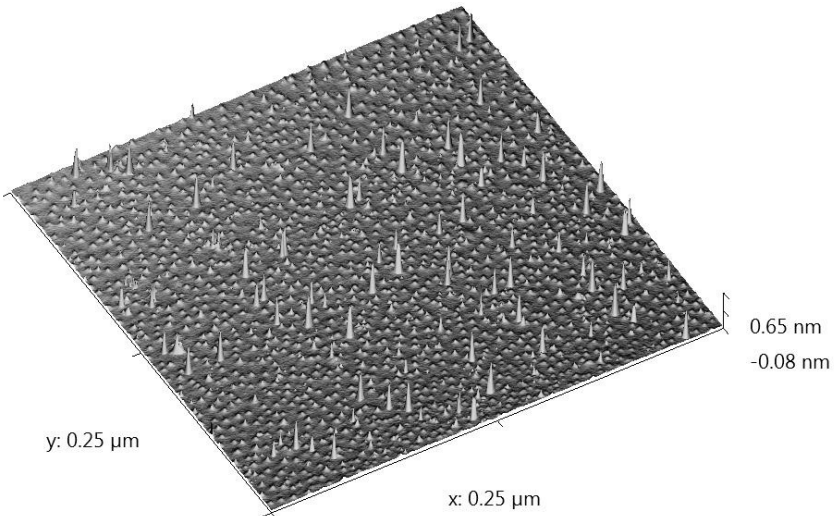


Figure 29. Topography of $Cu_xBi_2Se_3$ after cleaving at room temperature.

$Cu_{0.05}Bi_2Se_3$ after a room temperature cleave shows nodules on the surface. To prevent the tip from changing, we take data between these larger features.

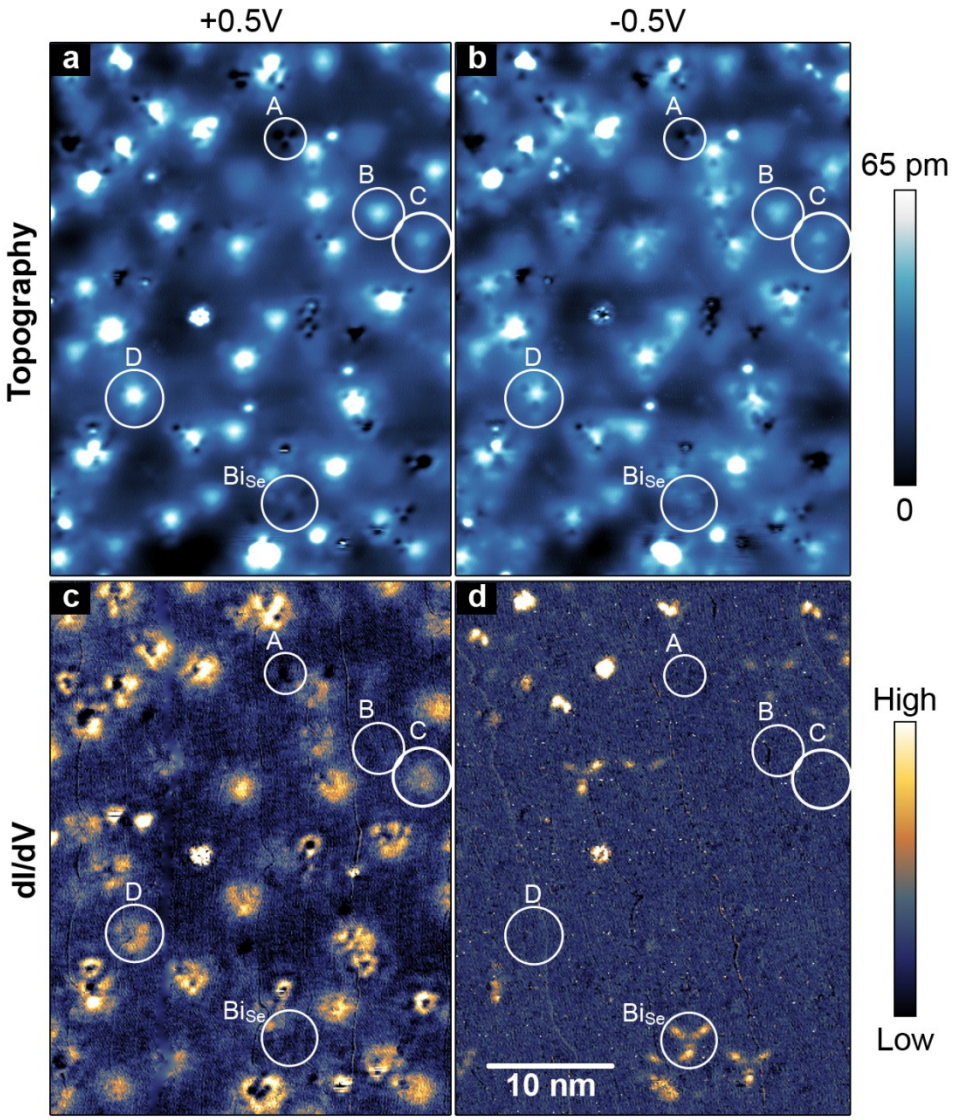


Figure 30. Topography and dI/dV images of $Cu_xBi_2Se_3$.

Topography (top) and dI/dV (bottom) at +0.5V (left) and -0.5V (right) of the same region of $Cu_{0.2}Bi_2Se_3$ at 500pA. At -0.5V, the Bi_{Se} antisite defects are weak features in the topography (**c**), though are prominent features in the dI/dV image (**d**).

Smaller defects with identifiable structure are visible in Figure 30a and b—in addition to the nodules, we can identify 5 additional defects that are not observed in

nominally undoped Bi_2Se_3, labeled Defects A-E (Only defects A-D are visible in Figure 30). Because the features in the topographs at +500meV and -500meV, respectively, are nearly identical, this indicates that the STM-observed defect structures are primarily due to lattice distortion, rather than electronic states.

Though Cu intercalants have long been expected to be the origin of n-doping in $Cu_xBi_2Se_3$, [32,34] a quantitative picture of the electronic interaction has not been developed. We observed several Cu defects in $Cu_xBi_2Se_3$ with different topographic and tunneling signatures Figure 30a-d, several of which have been observed by STM before (Defects A-C); [35,36] however, because STM lacks chemical specificity, we cannot definitively identify these defects without resorting to *ab initio* simulations of STM topographs. To simulate the Cu defects, we performed density functional theory calculations with single (7x7) quintuple-layer (QL) slabs using 98 Bi and 147 Se atoms; the technique is otherwise identical to that used in [33].

DEFECT IDENTIFICATION

Based on our simulations (Figure 31a-e), we have found that Defect A (Figure 31a) is consistent with an interstitial Cu below a surface Se atom. Defects B and C (Figure 31b & c) are consistent with intercalants, though the simulations for the two defect sites (H and T) are qualitatively similar. We concluded that Defect B is at the H-site because the lobe intensities in the simulation are relatively stronger than the simulation for the T-site. Defect D (Figure 31d) is consistent with a Cu_{Se} substitutional defect and can be identified by its protruding central feature and depressed lobes. Defect E (Figure 31e) is consistent with a Cu_{Bi} defect, and is only visible under ideal tunneling conditions due to its weak signature. Interestingly, this

feature is expected to be responsible for p-type doping effects in $Cu_xBi_2Se_3$, though we do not observe any localized spectroscopic features to support this—apparently, its doping effects are delocalized. Defect D has a weak spectroscopic state ~500meV above E_F. Additionally, Defect D is rarely observed in $Cu_{0.05}Bi_2Se_3$, while it is fairly abundant in $Cu_{0.12}Bi_2Se_3$ (Figure 32), suggesting that it has a higher energy of formation than the other defects.

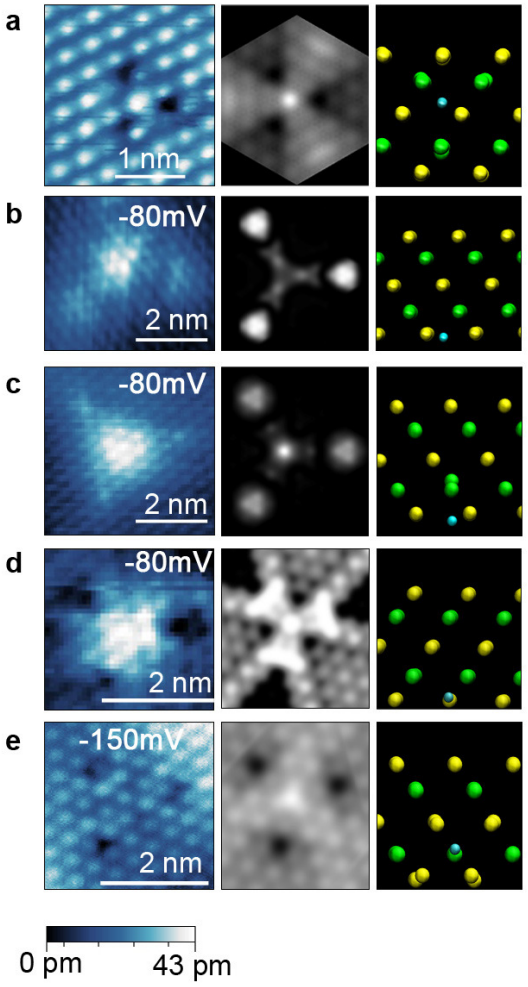


Figure 31. Detailed defect analysis and comparison with simulation.

a Defect A atomic resolution topograph at 1V, 480pA, left, simulated STM topograph, center, and the location of the interstitial Cu used for the simulation (the Cu is positioned underneath a Se atom).

b Defect B topographs, left, and the simulated topograph of an H-site intercalant, right.

c Defect C topographs, left, and the simulated topograph of a T4-site intercalant, right.

d Newly identified defect D, left, and the simulated topograph for a Cu_{Se} defect in the bottom Se layer of the sample.

e Newly identified defect E, left, and the simulated topograph for a Cu_{Bi} defect in the bottom Bi layer, right.

Color range for **e** is 18pm, all others is 43pm.

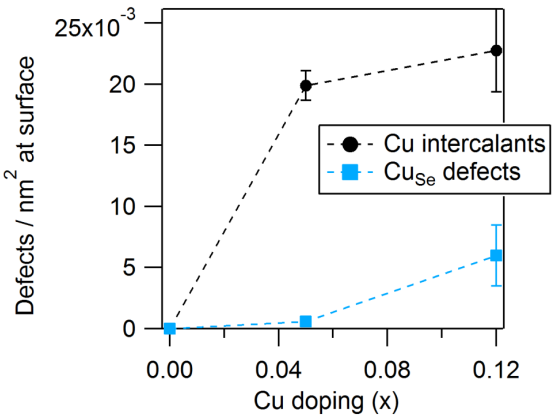


Figure 32. Defect counts taken from various cleaves.

At x=0.05, the Cu_{Se} count is significantly smaller than the Cu intercalant count. However, at x=0.12, the CuSe count is a significant portion of the defects, indicating that it has a higher energy of formation.

PAIR CORRELATION STUDY

By examining the 200nm x 200nm topograph of $Cu_{0.05}Bi_2Se_3$ in Figure 33a, we found that along the [1 0 -1] direction, the radial distribution function for Cu intercalants becomes structured. In this nearly-dilute regime, the existence of any structure in the pair correlation function, $C(\mathbf{r}_1,\mathbf{r}_2)$, is significant as it indicates a preference towards self-ordering. Therefore, to ensure that the radial distribution is not influenced by finite size effects that could create similar features, we normalized the pair correlation function by the average result of 100,000 simulated random ensembles, denoted $R^*(\mathbf{r}_1,\mathbf{r}_2)$, using identical defect counts and box dimensions as the STM data, similar to Ref. [37]. More information can be found in the Methods section.

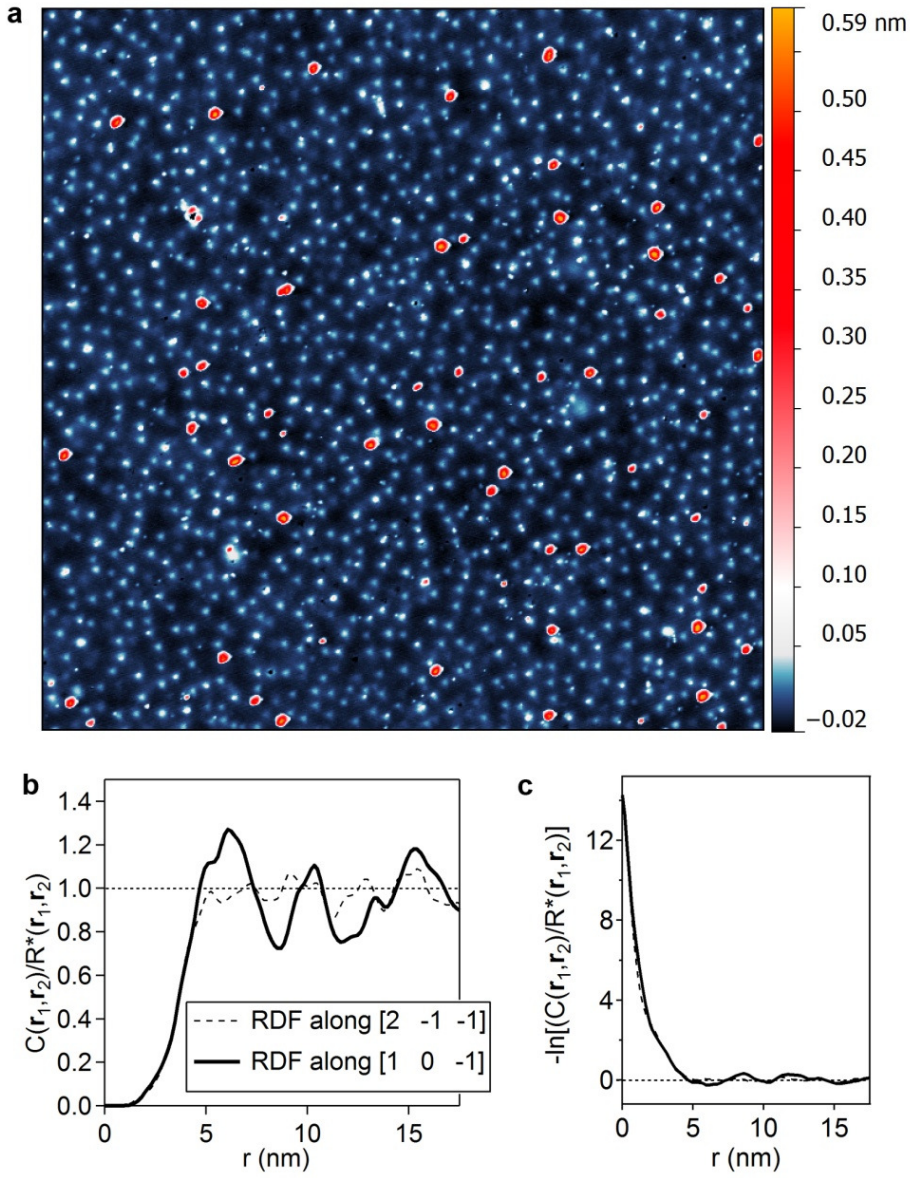


Figure 33. Radial distribution function and potential of mean force.

a 200nm x 200nm topography of nominally $Cu_{0.05}Bi_2Se_3$ showing Cu intercalants, white spots, and Cu clusters, red dots. **b** Radial distribution function of Cu intercalants extracted from **a** along directions of major symmetry. **c** A repulsive interaction is shown by the potential of mean force, determined by taking the log of the radial distribution function.

Though STM provides a static picture of the intercalation scheme, any observed ordering is the result of energy minimization of mobile defects. Therefore, to approximate the interactions between intercalants, we turn to the potential of mean force, $-kT \cdot \ln(C(\mathbf{r}_1,\mathbf{r}_2))$, which is a representation of the mean pair interaction. The sharp peak for small radii (Figure 33c) indicates that the Cu intercalants physically repel each other. Interestingly, this is in contrast to the surface adatoms which prefer to arrange into clusters. Most likely, the intercalants are fully ionized and exhibit screened Coulombic interactions, while the adatoms are not. Using STM, we are unable to determine if the Cu intercalant distribution formed during the sample anneal or after cleaving—however, assuming that the defects rearranged at room temperature, the interaction potential peaks near ~300meV, while higher temperatures would mean even stronger repulsive interactions.

CONSEQUENCES FOR SURFACE STATE INVESTIGATIONS

The observed anisotropic geometries of these defects has consequences for quasiparticle interference studies: the common assumption that scattering centers are both small and isotropic relative to the quasiparticle wavelength [9] is not necessarily supported. It's possible that an *isotropic* Fermi surface, such as the 2DEG formed on the bulk conduction band of Bi_2Se_3, can exhibit *anisotropic* interference effects due to the geometry of the scatterers. For instance, in Figure 34b and Figure 37, we observe a triangular scattering geometry from surface defects, but we cannot verify its origin—is this due to a warped Fermi surface, or a triangular scattering potential, or both?

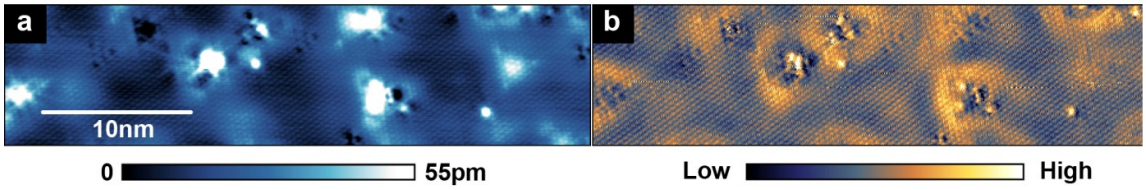


Figure 34. Topography and quasiparticle states on $Cu_xBi_2Se_3$.

a Atomic resolution topography at -80meV shows a variety of defects, while the corresponding dI/dV data (**b**) shows localized states appear near larger defect complexes. More subtle defects do not appear to play a significant role on the surface state.

To better motivate the concern, consider the following: an isotropic point source generates an isotropic wave pattern, as shown in Figure 35. An isotropic point source with an anisotropic arrangement, however, gives rise to an anisotropic wave pattern, like those shown in Figure 36. An additional complication impacts the interpretation of the results: the crystallographic deformations created by the defects cause significant nonlinearities in spectroscopy which require a detailed understanding of the transmission barrier as a function of energy and height. [7]

Prior to arriving at this understanding, we naively approached the scattering data like similar studies on more traditional materials (including superconductors and metal surface states): the process is to acquire a dI/dV map, then Fourier transform it. Performing this analysis on $Cu_{0.05}Bi_2Se_3$, including the data in Figure 37, provided a convincing dispersion curve. However, we cannot quantitatively analyze the data: we have insufficient information to account for the nonlinear transmission effects in the data, nor can we decouple the scatterer geometry from the patterns without knowing a functional form for the scattering potential.

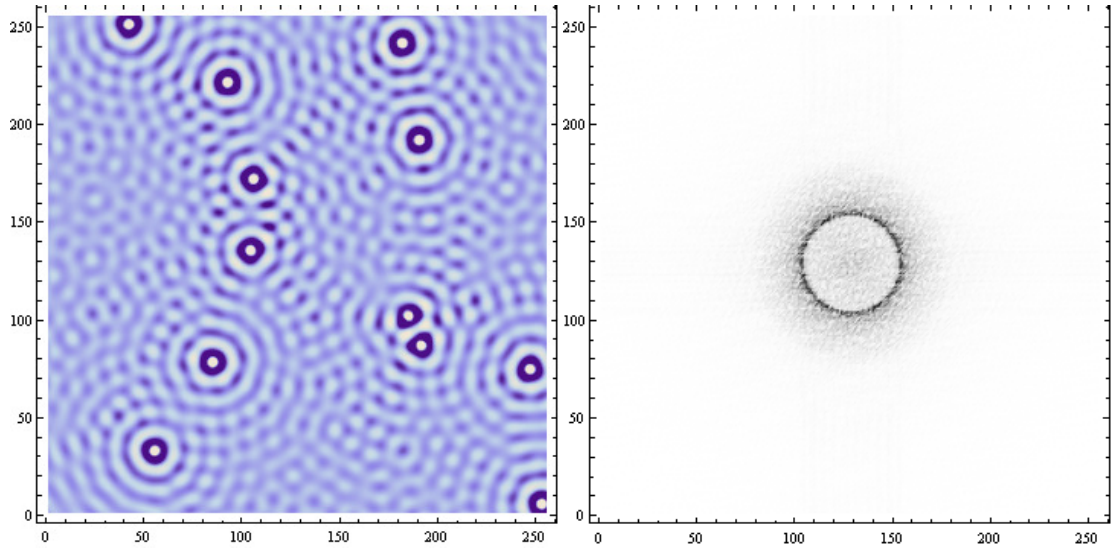


Figure 35. Isotropic point sources and their FFT.

Left, isotropic point sources are placed randomly within a domain and simulated with a 1/r falloff. Right, the Fourier transform has a sharp peak at the wavelength of the point sources, as expected.

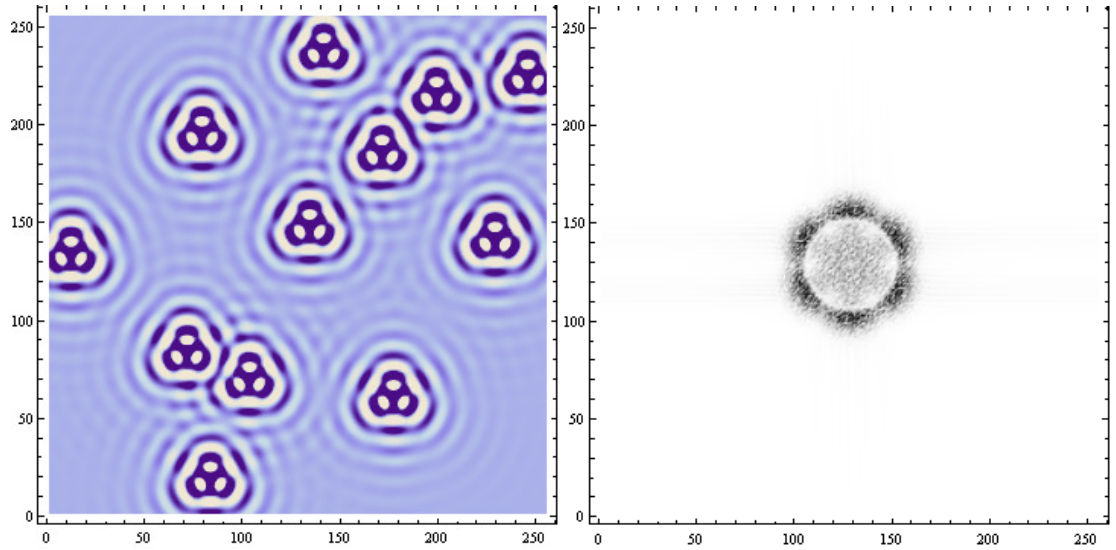


Figure 36. Triangular point sources and their FFT.

Left, sets of 3 isotropic point sources are placed randomly in a domain. Consequently, they generate a triangular wave pattern, despite the fact that each point source is actually isotropic. Right, the FFT has 6 lobes rather than a single ring. Such a pattern can arise either from anisotropic waves or from anisotropic wave sources, making it difficult to distinguish the origin without detailed analysis of the wave patterns themselves.

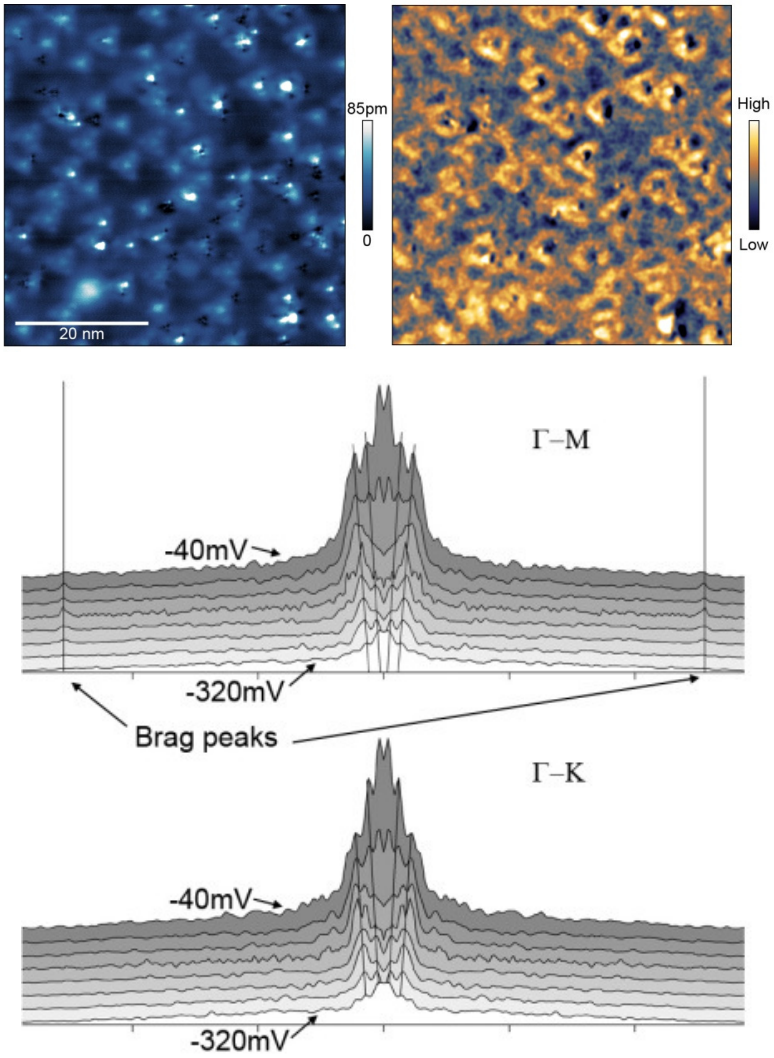


Figure 37. Quasiparticle analysis of $Cu_xBi_2Se_3$.

Topograph (top left) and dI/dV image (top right) at -40meV on $Cu_{0.05}Bi_2Se_3$. The triangular features in the dI/dV image could be due to a triangular scattering potential, instead of a warped Fermi surface. These features are dispersive, as seen in the cross sections of the FFTs (bottom) at different biases, but we cannot assign their origin due to the non-trivial scattering potentials.

Future quasiparticle interference studies on Bi_2Se_3, and other similar materials with large anisotropic potentials, will require additional analysis to ensure the scattering geometry does not leave its signature in the interference patterns. In

particular, the soft potentials created by these triangular protrusions may prove impossible to model with absolute certainty.

In summary, by identifying multiple Cu defects with STM and *ab initio* simulations of $Cu_xBi_2Se_3$, we have found disparate influences due to different defects. Through analysis of the radial distribution function, we have observed that Cu intercalants exhibit repulsive forces that, in combination with the enhanced mobility within the van der Waals gap, mediate self-ordering. While interstitial and Cu_{Bi} defects appear to have a delocalized influence on the surface state, Cu intercalants exhibit strong long-range anisotropy, not only in their structural effects, but also on the electronic structure near the surface.

METHODS

The potential of mean force should, at zero radius, go to infinity. However, to reduce the noise, we used a Gaussian kernel for the data with h = 1 nm. This creates a small, but nonzero, value at 0 radius.

To simulate the defect ensembles, we used a total of ~10^9 defect pairs. Each simulation has the same dimensions and number of particles as the real data, which is 200nm x 200nm and 1021 defects. Therefore, each ensemble has (1021 * 1020)/2 pairs; of course, inside the cutoff window (17.5nm, in this case), there are far fewer pairs. The simulation was run until there were 10^9 pairs within the cutoff window, which ended up being ~100,000 ensembles.

The number of ensembles computed is fairly large and is best performed using a parallel multi-core algorithm in a compiled language. Here, I include pseudo-code for generating simulated RDF data:

```
MainLoop():
        Number of ensembles = 100,000;
        Initialize thread manager();
        For each ensemble:
                Create EnsembleThread();

EnsembleThread():
        Dim = [200,200];
        Number of particles = 1021;
        For each particle:
                Particle.position = [Dim[0]*Random(), Dim[1]*Random()];
        Calculate RDF();
        Return RDF();
End;
```

The resulting averaged function is extremely smooth—however, each simulation is noisy due to the random distribution of particles. To determine whether our data is within 1σ of the standard deviation, we kept track of the pointwise standard deviation. Figure 38 shows the average RDF with the pointwise standard deviation for the non-interacting system. Therefore, any real system that falls inside these bounds is within the bounds of a truly random system. Normalizing a measured system by this data would provide a function that is ≈ 1 for all values if the distribution is random. Therefore, to determine whether the structure in our measured radial distribution function data is within the standard deviation, we normalized the pair correlation

function, $C(\mathbf{r}_1,\mathbf{r}_2)$, by $R^*(\mathbf{r}_1,\mathbf{r}_2)+\sigma$ and $R^*(\mathbf{r}_1,\mathbf{r}_2)-\sigma$. The results are shown in Figure 39. Along [2 -1 -1], the fluctuations beyond 5nm are within $\pm\sigma$ of 1, suggesting that the structure is noise. Along [1 0 -1], however, the fluctuations beyond 5nm vary more than the 1σ value, suggesting that the features are real. Both data sets, however, confirm the repulsive potential.

The lack of structure along [2 -1 -1] may be due to the semi-dilute regime for this data. In particular, we expect a highly dilute system would show no structure along any direction, while a higher density system would show complicated structure along all major symmetries.

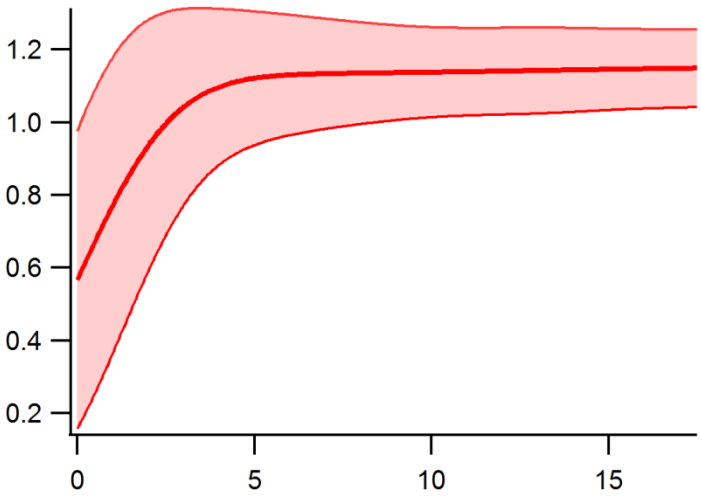
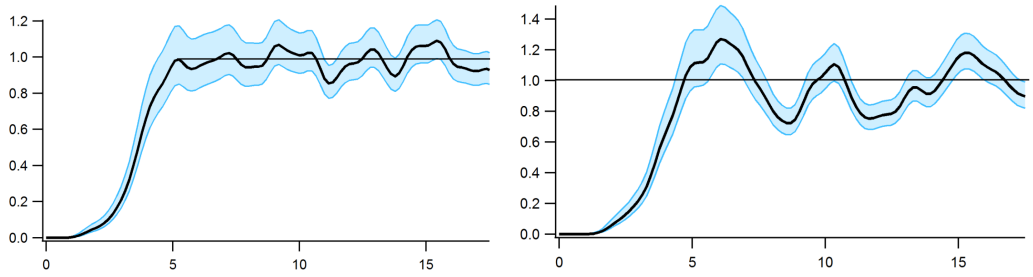


Figure 38. Pair correlation function of a random distribution with no interactions.

Because a truly random distribution will have some fluctuations, we need to know how significant they are to determine the validity of our data. Furthermore, by sampling a finite size, it's possible to end up with finite-size effects that can resemble structure. Therefore, we prepared this simulated data: the thick red line is the average of 10^9 ensembles, while the thin red lines are the +/- 1σ values.



[2 -1 -1] **[1 0 -1]**

Figure 39. Radial distribution functions along two major-symmetry directions.

While there is structure along both directions, the fluctuations along the [2 -1 -1] direction are within 1σ of the mean, while along [1 0 -1], the fluctuations deviate more significantly. We suspect that for higher doping densities, the signal along both directions would be more structured.

Chapter 7: Band bending and potential fluctuations in Bi_2Se_3 and $Cu_xBi_2Se_3$

AUTHOR CONTRIBUTIONS

The following section is based on the publication *Mapping the 3D surface potential in Bi_2Se_3*. [33] The work was performed in collaboration with Dr. Damien West and Professor Shengbai Zhang of Rensselaer Polytechnical Institute and Dr. Ireneusz Miotkowski and Professor Yong P. Chen of Purdue University.

All authors contributed to the intellectual content of this section. Ireneusz Miotkowski and Yong Chen provided the samples. Damien West performed and analyzed the density functional theory simulations under the guidance of Shengbai Zhang. Chris Mann collected and analyzed the STM data, performed the nonlinear transmission coefficient simulations, and prepared the text, Chih-Kang Shih advised on the experiment and provided input on the data analysis and the manuscript.

INTRODUCTION

Bi_2Se_3 has emerged as a particularly interesting host of topological physics due to its non-trivial surface state. [14,15,38] Initial studies indicated that theoretical predictions were largely true—topological protection reduces scattering and carriers are helical Dirac-like fermions. [39–42] However, further investigation raised several issues with its physical and structural properties in actual materials: Bi_{Se} antisites may allow scattering from the surface to the bulk, [27,43] strong band bending creates an additional, topologically trivial, surface state, [30,44–46] and deep, unobservable charges may cause fluctuations in the Dirac point. [39]

To address these concerns, we conducted scanning tunneling microscopy studies on nominally stoichiometric Bi_2Se_3. We have identified two distinct Bi_{Se} antisite distributions using spectroscopic imaging and *ab initio* simulations; their energy states are sensitive to the local valence band position and operate as depth-dependent potential sensors. Using these sensors, we generate nanometer-scale measurements of a strong surface band bending field that is tunable with bulk dopants, which we demonstrate with measurements on $Cu_xBi_2Se_3$. Additionally, by mapping the lateral fluctuations of the Dirac point with *dI*/*dV* imaging, we show spatial correlation of the potential fluctuations with specific near-surface point defects. Disparate influences are observed for different defects: Se vacancies cause Dirac point shifts, while Bi_{Se} defects follow the Dirac point shifts.

RESULTS

Band Bending

We observed that Bi_{Se} defects show tunneling amplitude maxima at vastly different voltages—for example, Figure 40f shows *dI*/*dV* spectroscopy for two defects whose tunneling maxima are separated by ~150 meV. By analyzing the *dI*/*dV* spectra from numerous Bi_{Se} defects, a histogram analysis shows that there is a bimodal distribution (Figure 45a). Therefore, contrary to previous reports, [25] these nearly-identical features are due to Bi_{Se} defects at *two* separate Se layers. From *ab initio* calculations, we can rule out the middle layer (Bi_{Se2}) due to its high energy of formation, [47] which means that the defects are in structurally equivalent locations (i.e., the top or bottom of a quintuple layer—these positions are symmetric). Absent an external field, two equivalent defects at two different depths should be energetically

degenerate. However, in this case, the degeneracy is broken by a vertical electric field: *mapping the defect energies at different depths maps the local band bending*. Based on our *ab initio* simulations of STM topographs for these defects (Figure 40b,c) we attribute the higher amplitude defect to Bi_{Se} on the 5^{th} layer and the lower amplitude defect to Bi_{Se} on the 6^{th} atomic layer, which correspond to the bottom of the 1^{st} quintuple layer and the top of the 2^{nd} quintuple layer, respectively (see Figure 40e and Figure 42). The n-shifted defect has a higher tunneling amplitude, meaning that it is closer to the surface/STM tip, demonstrating downward surface band bending (SBB) of the bulk conduction and valence bands. Downward SBB is consistent with ARPES observations, [30] and is possibly due to more Se vacancies forming at the sample-vacuum interface following the ultra high vacuum sample cleave.

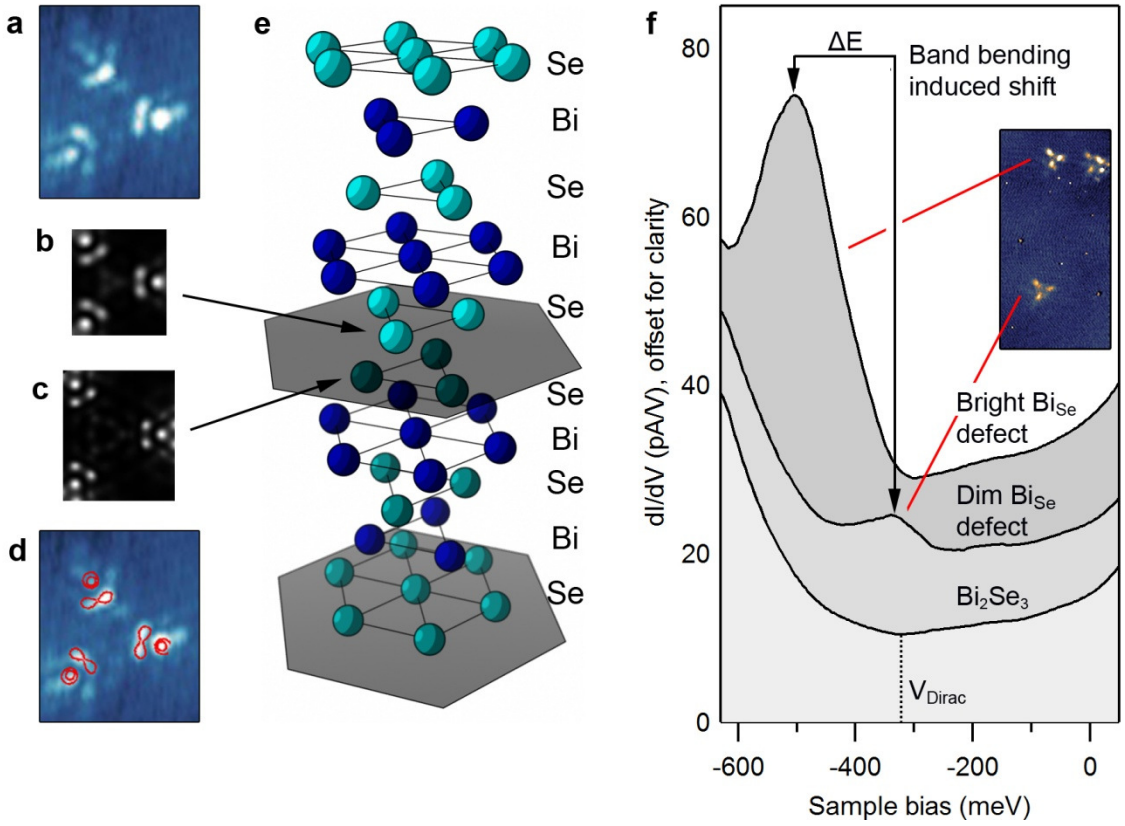


Figure 40. Topographic and spectroscopic identification of Bi_{Se} defects.

a, 2.9 nm x 3.5 nm atomic resolution imaging of a Bi_{Se} defect at -500 meV and *ab initio* STM simulations for a Bi_{Se} defect in the 5th (**b**) and 6th (**c**) layer at the same scale. **d**, Contours from **b** overlaid on **a**. **e**, Atomic structure of Bi_2Se_3, where hexagons separate the quintuple layers. **f**, dI/dV spectroscopy on two defect types, taken on their lobes: the dim defect type and a bright defect type as well as spectroscopy taken away from any defects. The bright and dim defects show a significant difference in their energy position on undoped Bi_2Se_3. To physically distinguish the defects, the lobes on the dim defects are slightly further from the center, more likely corresponding to a deeper defect as per the simulation in **c**, whereas the bright defect is more compact, better corresponding to **b**.

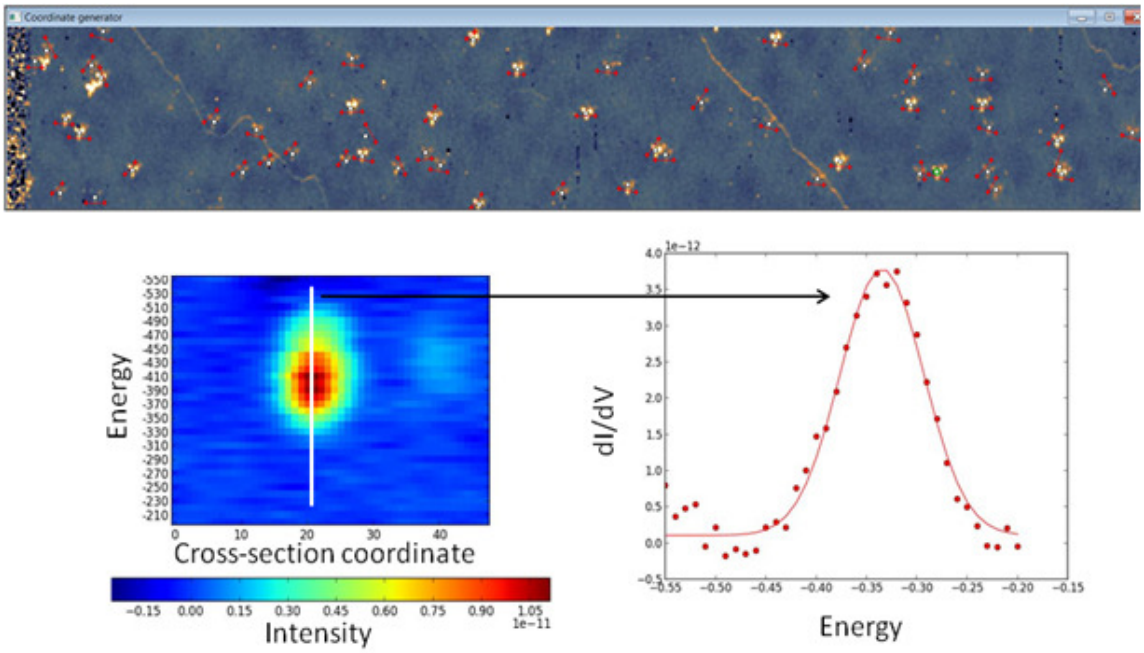


Figure 41. The dI/dV data processing technique

dI/dV spectroscopy is extracted from multipass data scans, one such scan is picked as an illustration above. We select spectroscopy across the lobes of each identifiable defect for multiple scans from Bi_2Se_3 (above), $Cu_{0.05}Bi_2Se_3$, and $Cu_{0.12}Bi_2Se_3$. Nearby spectra not on a defect state is then subtracted from the spectroscopy in order to fit the defect peak. A Gaussian peak was found to fit quite well. The following number of Bi_{Se} spectroscopic scans were analyzed for each sample:

- Bi_2Se_3: N=66
- $Cu_{0.05}Bi_2Se_3$: N=20
- $Cu_{0.12}Bi_2Se_3$: N=32

Note, the lines running diagonally through the dI/dV image at the top are due to a rare digital artifact from the Nanonis controller; these lines were removed from the data used in the Letter. They can be identified as artifacts because they show up precisely at piezo voltages associated with strong bit flipping. As an illustration, if the DAC is changing the bias from (say) 31 mV to 32 mV, this is 011111 to 100000 on the digital controller; each 1 puts extra voltage on the DAC that must be dissapated to achieve the next bias. If enough bits are flipping at once, the transient voltages are transferred to the piezo scanner and can be observed by the lock-in amplifier. The effect is too subtle to observe on topographic scans.

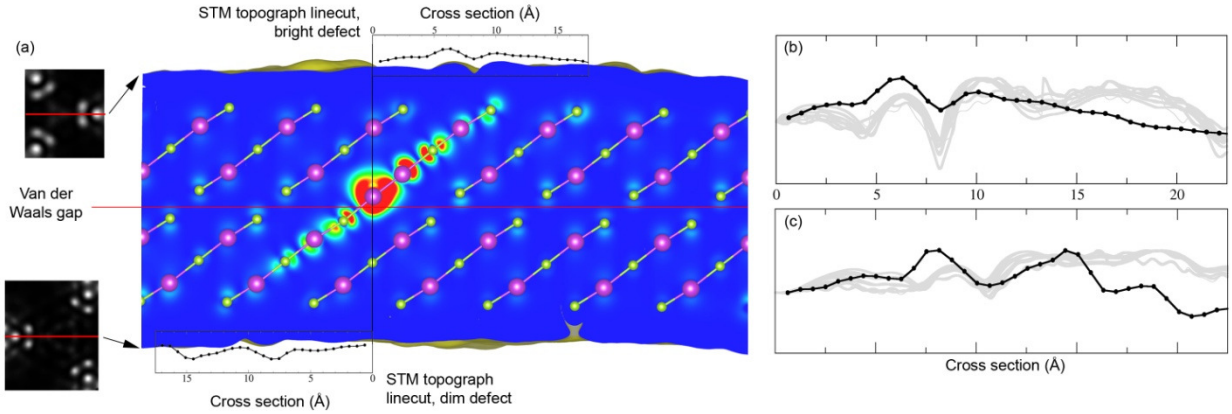


Figure 42. Defect simulation and comparisons.

(a) A 2-layer slab is solved by DFT, performed by Damien West. 2 surfaces are accessible in this simulation, allowing 5th and 6th layer Bi_{Se} defects to be simulated simultaneously—STM cross sections were extracted from topographic data and are shown with the computed isosurface. Experimental line-cuts from the defect center through one of the bright lobes of the defect topographic image for (b) the 'bright' defect and (c) the 'dim' defect are shown as the dark black lines. These are compared to 10 different scaled simulated linecuts with isosurface values which vary by 5orders of magnitude and Fermi energies which vary by 0.5eV, shown in grey. This demonstrates that the qualitative features of the simulation are insensitive to the particular choice of parameters and that the 'dim' defect peak positions correlate strongest with the 6th layer simulation, while the 'bright' defect peak positions correlate best with the 5th layer simulation. The relative overestimation of charge density at distances far from the defect center can be attributed to overlap of neighboring cell images.

These defects should retain their position relative to the valence band, so adding an n-type dopant, like incorporating Cu into the bulk, should n-shift the entire electronic structure including the Bi_{Se} distributions, without altering the nature of the Bi_{Se} defect itself (see Figure 43 and Figure 29 for more information). Indeed, we observed this effect—we also observed the two Bi_{Se} distributions shift closer together (Figure 45a). From the band bending picture, this means that the SBB is reduced (Figure 45b). Because more carriers are introduced with higher n-type doping (more of the bulk conduction band is below the Fermi level), we believe screening effects are causing the reduction in the band bending potential field.

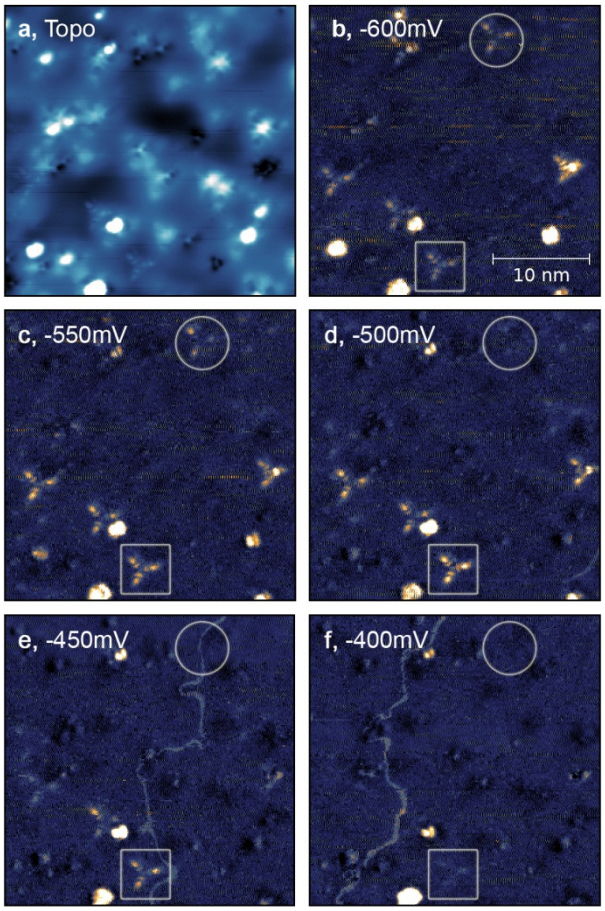


Figure 43. Impact of Cu doping on Bi_{Se} defects.

Cu doping introduces a number of new defects, as shown in the topograph in **a**. However, as shown in the dI/dV data in **b**-**f**, the distinct Bi_{Se} defect is still present. Indeed, it still shows up at different biases depending on its layer, as highlighted by the circled and squared defects.

We have considered the possibility that the band bending was induced by the electric field from the STM tip, an effect known as tip-induced band bending (TIBB). We performed spatial spectroscopic imaging at a variety of tunneling set points (150pA to 500pA) and observed less than a 5meV shift in the impurity peak position (Figure 44), indicating that TIBB is not the dominant source of the energy shift.

Presumably, as the surface is already n-doped even in the case of zero Cu doping, the available free carrier concentration is high enough to screen out the tip-induced field.

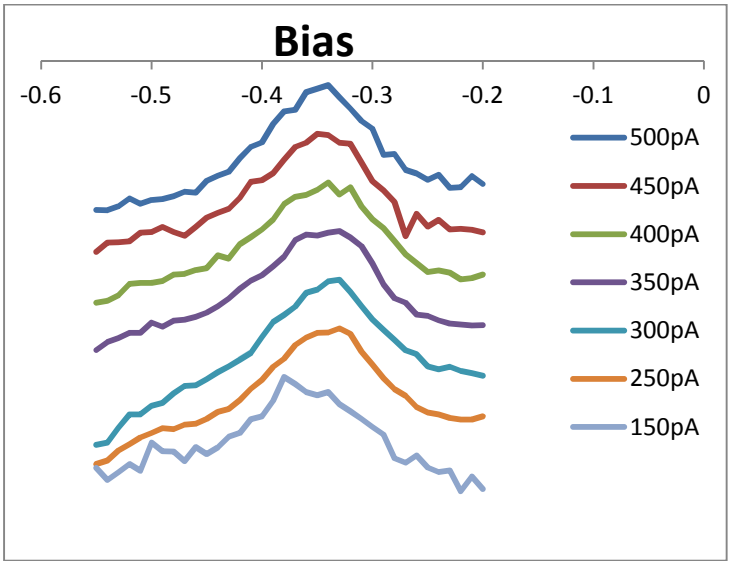


Figure 44. Tip-induced band bending study at several setpoints.

While there may be a small shift in the peak position, it is only a few millivolts—substantially less than the observed band bending. The data shown above is dI/dV data at constant current, followed by background subtraction.

To quantitatively approximate the magnitude of SBB as a function of doping (Figure 46a), the energy separation of the defect distributions is divided by the approximate separation between the Se layers terminating adjacent quintuple layers, as determined by local-density approximation (LDA) energy relaxation calculations. We note that intercalation of Cu increases the separation between layers, [36] and any TIBB effects would further amplify the observed bending. Therefore, the values we have measured are upper bounds.

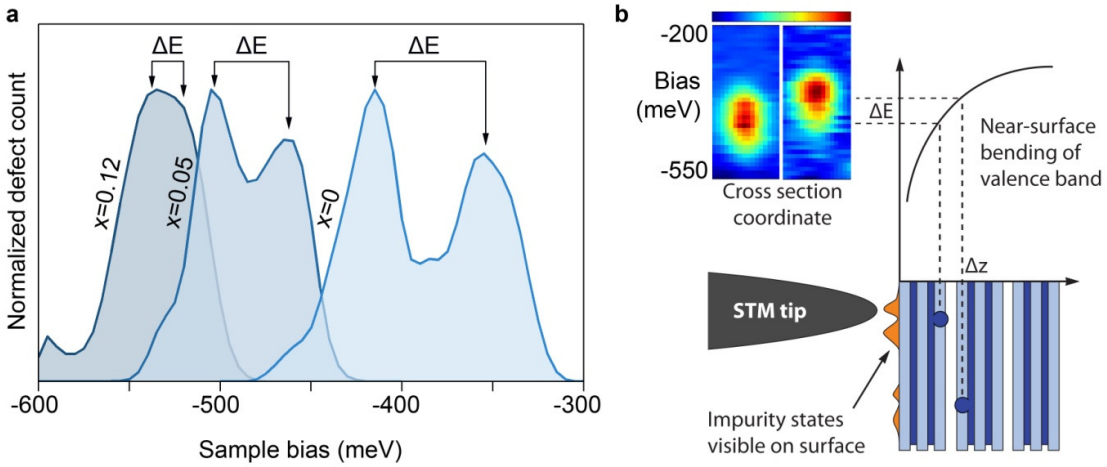


Figure 45. Defect distributions and detecting near-surface band bending.

a, Histograms of the Bi_{Se} defect energies in $Cu_xBi_2Se_3$, extracted from multipass dI/dV images, show a decreasing gap between the peaks as the Cu content is increased. **b**, Schematic of the proposed band-bending mechanism responsible for the energy difference between the defects; the orange profile represents the Bi_{Se} defect's observable states on the Bi_2Se_3 surface. Cross sectional dI/dV linescans from two different defects on Bi_2Se_3 are shown for direct comparison, as well. Color scale ranges: left, $1.4x10^{-11}$ pA/meV, right, $5.0x10^{-12}$ pA/meV.

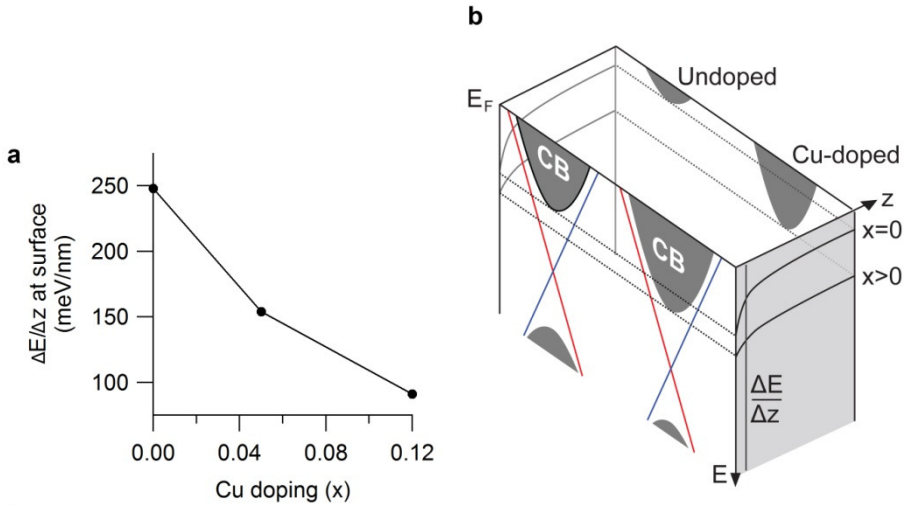


Figure 46. Influence of Cu doping on the surface potential**.**

a, Band bending as a function of doping level. To numerically determine the band bending, we used the distance determined by *ab initio* simulations of 0.254 nm. **b**, Schematic showing the surface potential dependence on doping. In Cu-doped samples, more of the bulk conduction band is below the Fermi energy, providing enough carriers to screen the band bending.

Several groups have observed bulk 2D electron gas (2DEG) states appearing on samples left in vacuum, and the ARPES data suggests band bending is the culprit. [30,44–46] STM is surface sensitive, so this calls into question the results of all STM studies above the conduction band minimum on this material: is the measured surface state really of topological origin, or simply due to a traditional bulk 2DEG? If a 2DEG exists, it can be eliminated by reducing the SBB. To ensure that future STM studies are truly investigating the topological surface state in Bi_2Se_3, control of the surface potential will be essential—bulk n-type doping provides one such mechanism, as shown schematically in Figure 46b.

Correlating Potential Fluctuations

The Dirac point shifts rigidly with bulk band fluctuations [39] (Figure 47a) and the *dI*/*dV* data ~200meV below the Dirac point is effectively linear with bias, allowing us to map the *dI*/*dV* value below the Dirac point to the fluctuations of the Dirac point. This technique has been successfully applied to graphene, [48] and we confirmed the validity of this technique for Bi_2Se_3 from line spectroscopy data, as shown in Figure 47b and c. Figure 47d and e show the topography and corresponding map of Dirac point fluctuations, respectively, in nominally stoichiometric Bi_2Se_3. Se vacancies observed in Figure 47d, indicated with red circles in Figure 47e, can be correlated with n-shifts in the Dirac point. The energy of formation for Se vacancies increases as the sample is n-doped [47]—so they would preferentially form in more positively-shifted regions. Therefore, their correlation with negatively-shifted regions suggests they are causing the n-type fluctuations. It's also apparent that these defects are the origin of the fluctuations (responsible for the finite width of the peak energy distributions in

Figure 45a) in the Bi_{Se} distributions: negatively shifted Bi_{Se} defects are on negatively shifted regions of the sample and positively shifted Bi_{Se} defects are found on positively shifted regions of the sample (Figure 47e-g). Significantly, while these Bi_{Se} defects do have strong electronic interactions, they do not appear to have strong effects on the position of the Dirac point; rather, they appear to *follow* the bulk fluctuations.

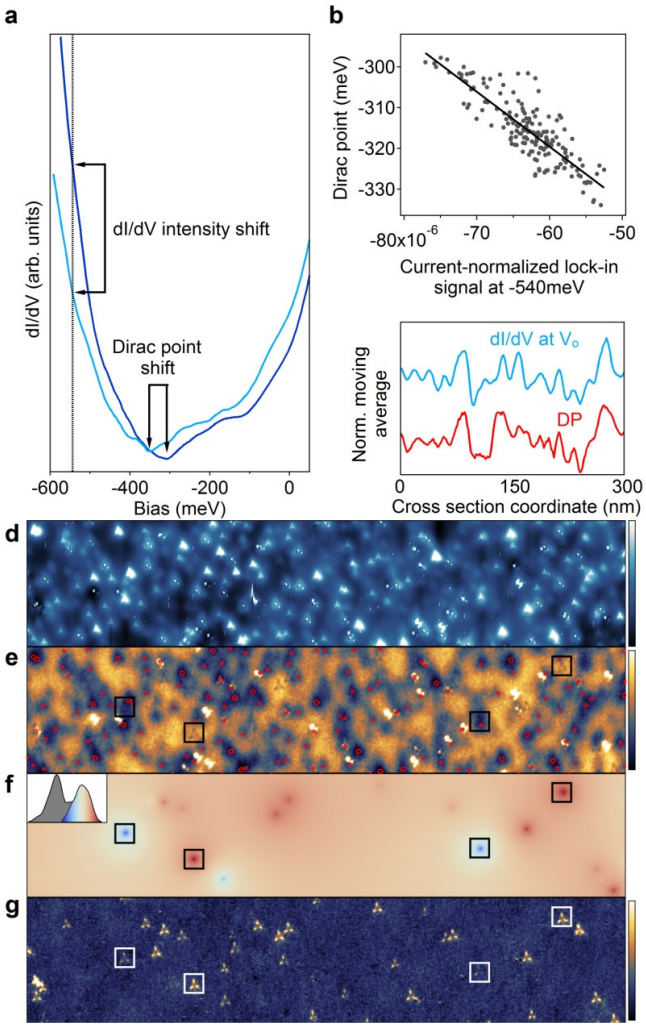


Figure 47. Correlating potential fluctuations with point defects.

a, Two $(dI/dV)/I_o$ curves taken from different regions on Bi_2Se_3 show how the Dirac point shift can be correlated, to first order, with the intensity shift in dI/dV intensity—this allows single-pass dI/dV scans to map the Dirac point fluctuations. V_o is the bias chosen in the approximately linear regime of the dI/dV data (-540 meV in this case) and I_o is the corresponding tunneling current. **b**, A scatter plot determined from line spectroscopy data across 300 nm of Bi_2Se_3 shows the first-order relation between the Dirac point and the bulk valence band; $R^2 = 0.72$. **c**, 7-point moving average from the line spectroscopy visually shows the correlation from **b**. A 350 nm x 50 nm region of Bi_2Se_3 is then investigated with this technique: topography (**d**) at -200 meV shows Se vacancies, and the dI/dV image of the same region taken at -540 meV shows fluctuations in the Dirac point (**e**). In **e**, the red circles represent bright triangular defects, determined from **d**. A visualization of the deviations from the distribution mean for dim-type Bi_{Se} defects is shown in **f**, where red points indicate a Bi_{Se} defect with an energy state that is more positive than the distribution mean, and blue points indicate a Bi_{Se} defect below the distribution mean (see inset). The dI/dV image at -350 meV (**g**) shows both dim and bright type defects. The fluctuations in the Bi_{Se} defect energy scale correlates directly with the fluctuations in

the Dirac point; a few extreme examples are highlighted in boxes in **e**-**g**. Color scales: **d**, 40pm, **e**, 30meV, **g**, 10pA/V.

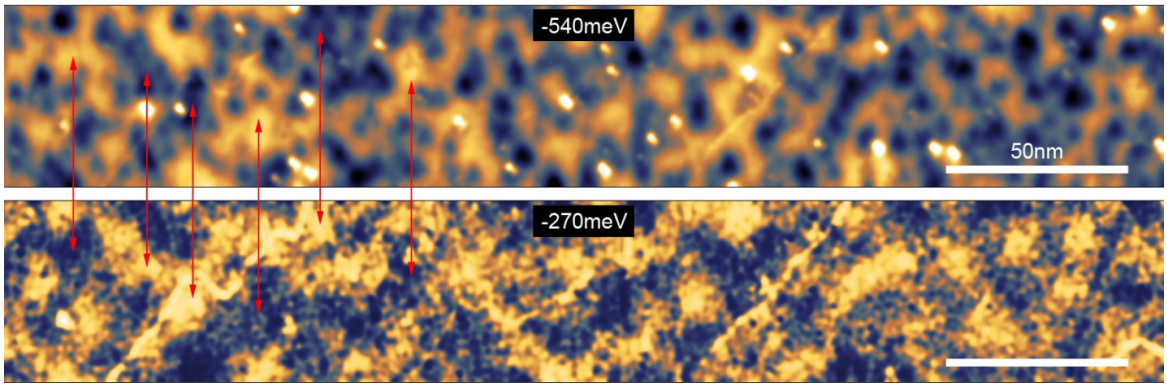


Figure 48. dI/dV mapping above and below the Dirac point.

The dI/dV mapping of the Dirac point also works above the Dirac point. Above are two scans of the same region on Bi_2Se_3 showing the Dirac point shifts using the same color scale as in Figure 47e. As expected, we observe a contrast reversal due to the rigid bulk shift. However, due to the reduced slope of the dI/dV curve closer to the Fermi level, this data is far noisier. We have included a few red arrows to aid the eye. The cross correlation coefficient is -0.50, quantitatively confirming the contrast reversal.

dI/dV images are often found to be inverse topographic images. [49] To verify that we are imaging bulk shifts, rather than an artifact associated with this effect, we examined the dI/dV data above the Dirac point, as shown in Figure 48. Because the slope of the dI/dV curve is positive above the Dirac point, instead of negative like at -540meV (see Figure 47a), we would expect a contrast reversal above the Dirac point. Indeed, we observed this effect. Both images were first subject to outlier point rejection, followed by Gaussian filters to remove high-frequency noise. The resulting cross-correlation coefficient between the dI/dV data at -540meV and -270meV is -0.50, which is strongly anti-correlated (perfect anticorrelation would be -1, no correlation would be 0). Therefore, for the case of Se vacancies, we have confirmed that they do, in fact, cause rigid shifts of the Dirac point.

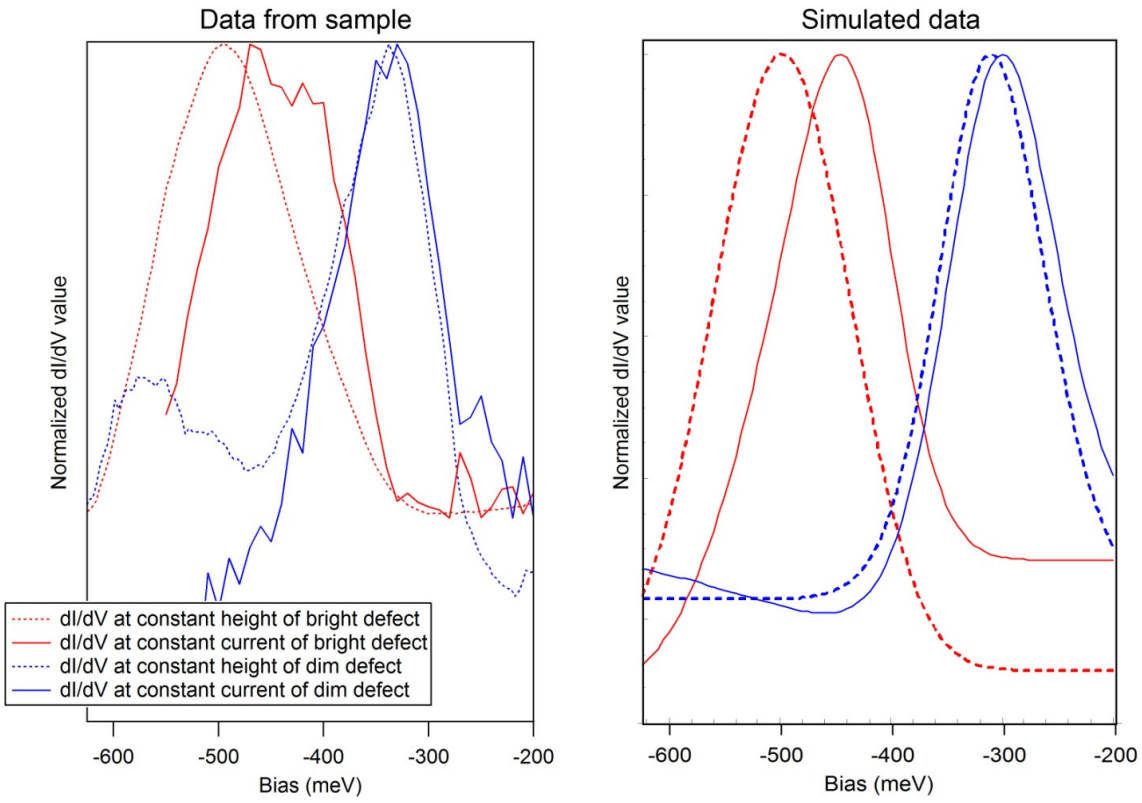


Figure 49. Differences between constant-current and constant-height spectroscopy.

The $\left.\frac{dI}{dV}\right|_z$, or more commonly dI/dV, peak positions are not the same as the peak positions determined from multi-pass dI/dV data, which is effectively $\left.\frac{dI}{dV}\right|_I$. The data presented in Figure 40f is $\left.\frac{dI}{dV}\right|_z$ data, and does not appear to fit within the histogram in Figure 45a, but it is included: the histograms were built with $\left.\frac{dI}{dV}\right|_I$ data. On the left is a plot of both forms of dI/dV data from both defect types. On the right is synthetic dI/dV data of both forms; we simulated the density of states as a Gaussian function on top of a polynomial background. To simulate the $\left.\frac{dI}{dV}\right|_I$ data, we used the method of [7]. The simulation shows that the differences in peak positions are actually to be expected—this is a consequence of the nonlinear tunneling transmission coefficient as a function of voltage. The taller the peak, the larger the expected difference in peak positions should be between the data types.

Grid dI/dV data is very sensitive to drift at liquid nitrogen temperatures—however, we found that by taking multi-pass dI/dV data, we were able to beat the drift velocity and generate reproducible spectroscopic data. During multi-pass dI/dV scans, the z-position of the STM tip changes during each pass to enforce a constant current,

causing a systematic shift in the peak positions that increases with larger biases, as seen in Figure 49. However, this does not create new spectroscopic peaks. In other words, the bimodal distributions we present are not artifacts of the multipass technique. In fact, constant-height dI/dV data show a *larger* degeneracy splitting between the two defect types.

DISCUSSION

Our reevaluation of the Bi_{Se} defect as an impurity state suggests that bulk-surface interactions are not directly evident. Furthermore, this classification allows the band bending to be directly visualized, highlighting the concerns that non-topological surface states may be inadvertently formed on Bi_2Se_3. To prevent the formation of such a state, it will be important to deliberately tailor the band bending using advanced synthesis techniques such as molecular beam epitaxy. Furthermore, by correlating observable defects with the fluctuations in the Dirac point through *dI/dV* imaging, we have shown that different defects influence the Dirac point disparately—as different preparation techniques produce different characteristic defects, an understanding of the interplay between specific defects and the local electronic structure is critical. We expect that such atomic-scale insights will enable application-optimal doping and growth schemes. Ultimately, Bi_2Se_3 remains a model system for exploring topological physics.

METHODS

Sample growth

Ternary $Cu_xBi_2Se_3$ bulk crystals were grown by the vertical Bridgman technique using a three-zone high-temperature furnace. The temperature profiles in the growth zone were designed to produce a 3–5ºC axial gradient. The starting materials used for crystal growth (6N Bi and 6N Se) were deoxidized and outgassed under vacuum directly before use. Stoichiometric amounts of Bi and Se along with 6N purity deoxidized Cu were enclosed in a graphitized fused quartz ampoule, evacuated for several hours, and sealed off under a dynamic vacuum of $\sim 10^{-8}$ torr. The pre-loaded charge was then placed in the middle zone of the furnace. After the preliminary reaction was completed at 400ºC, the charge was slowly heated up to 770ºC for approximately 6 hours and then the melt was allowed to homogenize for 24 hours under a high linear temperature gradient to promote mixing. The melt was solidified with a rate of about 1.5 mm/hour at a small linear gradient of 3-5ºC/cm.

Ab initio calculations

Our calculations are based on density functional theory with the Perdew-Burke-Ernzerhof approximation. [50] Interactions between ion cores and valence electrons are described by the projector augmented wave (PAW) method, [51,52] as implemented in the VASP package. [53,54] Plane-waves with a kinetic energy cutoff of 270eV were used as the basis set. We used a (7x7) double quintuple slab containing 196 Bi and 294 Se atoms with Gamma point sampling of the Brillouin zone. All atoms were allowed to relax with a convergence criterion for the structural relaxations of 0.04eV/Å. To compare with experiment, we simulate the STM images

based on the theory of Tersoff and Hamann. [55] When simulating STM images for comparison with experimental images, care has to be taken with regard to the location of the experimental Fermi energy. To correspond to the observed images at a -500meV bias, the Fermi energy window used to construct the charge density for the simulated images range from CBM+300meV to the valence band maximum (VBM)+100meV. This window is chosen relative to the band edges because the use of only two slabs increases the theoretical gap.

Scanning tunneling microscopy measurements

Samples were cleaved in an ultra high vacuum (UHV) with a base pressure $<10^{-10}$ torr, then immediately transferred to the STM at liquid nitrogen temperature. The STM is a Pan-type STM scan head from RHK in a home-built UHV chamber. Electrochemically-etched tungsten STM tips were clean by e-beam irradiation under UHV. Scanning tunneling spectroscopy (STS) data was taken using the built-in lock-in amplifier on a Nanonis STM controller at 780Hz and 10mV modulation amplitude. STM topographs were taken in constant current mode and the bias was applied to the sample. Bi_{Se} peak positions were determined from multipass scan data at constant current, unless otherwise noted, and were extracted from the lobes of the defects; the background spectra, averaged on either side of the given lobes, were subtracted from the impurity peak spectra to remove systematic offsets due to the shape of the nearby dI/dV spectra. More information is provided with Figure 49.

Bibliography

[1] D. J. Griffiths, *Introduction to Quantum Mechanics (2nd Edition)* (Pearson Prentice Hall, 2004), pp. 70–76.

[2] W. A. Harrison and Physics, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond (Dover Books on Physics)* (Dover Publications, 1989).

[3] H. Bruus and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction* (Oxford University Press, Oxford, 2004).

[4] J. M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, 1979), p. 452.

[5] D. Croft, G. Shedd, and S. Devasia, in *Proceedings of the American Control Conference* (2000), pp. 2123–2128.

[6] R. J. Hamers, Annu. Rev. Phys. Chem. **40**, 531 (1989).

[7] J. Li, W.-D. Schneider, and R. Berndt, Physical Review B **56**, 7656 (1997).

[8] L. Petersen, P. T. Sprunger, P. Hofmann, E. Laegsgaard, B. G. Briner, M. Doering, H.-P. Rust, A. M. Bradshaw, F. Besenbacher, and E. W. Plummer, Physical Review B **57**, 6858 (1998).

[9] G. A. Fiete and E. J. Heller, Reviews of Modern Physics **75**, 933 (2003).

[10] M. J. Viens, (n.d.).

[11] P. J. Bryant, H. S. Kim, Y. C. Zheng, and R. Yang, Review of Scientific Instruments **58**, 1115 (1987).

[12] D. Nečas and P. Klapetek, Central European Journal of Physics **10**, 181 (2011).

[13] I. Horcas, R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, and A. M. Baro, The Review of Scientific Instruments **78**, 013705 (2007).

[14] M. Hasan and C. Kane, Reviews of Modern Physics **82**, 3045 (2010).

[15] J. E. Moore, Nature **464**, 194 (2010).

[16] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nature Physics **5**, 438 (2009).

[17] K. v. Klitzing, G. Dorda, and M. Pepper, Physical Review Letters **45**, 494 (1980).

[18] L. Fu, Physical Review Letters **103**, 1 (2009).

[19] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, a V Fedorov, H. Lin, a Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **460**, 1101 (2009).

[20] Y. L. Chen, J. G. Analytis, J.-H. Chu, Z. K. Liu, S.-K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen, Science (New York, N.Y.) **325**, 178 (2009).

[21] Y. Xia, D. Qian, D. Hsieh, L. Wray, a. Pal, H. Lin, a. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature Physics **5**, 398 (2009).

[22] H. Lind, S. Lidin, and U. Häussermann, Physical Review B **72**, 184101 (2005).

[23] J. Navrátil, J. Horák, T. Plecháček, S. Kamba, P. Lošťák, J. S. Dyck, W. Chen, and C. Uher, Journal of Solid State Chemistry **177**, 1704 (2004).

[24] J. Horák, Z. Stary, P. Lošťák, and J. Pancíř, Journal of Physics and Chemistry of Solids **51**, 1353 (1990).

[25] S. Urazhdin, D. Bilc, S. Tessmer, S. Mahanti, T. Kyratsi, and M. Kanatzidis, Physical Review B **66**, 161306 (2002).

[26] Y. Hor, A. Richardella, P. Roushan, Y. Xia, J. Checkelsky, A. Yazdani, M. Hasan, N. Ong, and R. Cava, Physical Review B **79**, 195208 (2009).

[27] S. Kim, M. Ye, K. Kuroda, Y. Yamada, E. Krasovskii, E. Chulkov, K. Miyamoto, M. Nakatake, T. Okuda, Y. Ueda, K. Shimada, H. Namatame, M. Taniguchi, and A. Kimura, Physical Review Letters **107**, 056803 (2011).

[28] Y. Cui and L. Li, Physica Status Solidi (a) **188**, 583 (2001).

[29] D. A. Fletcher, K. B. Crozier, K. W. Guarini, S. C. Minne, G. S. Kino, L. Fellow, C. F. Quate, K. E. Goodson, and A. Member, **10**, 450 (2001).

[30] M. Bianchi, D. Guan, S. Bao, J. Mi, B. B. Iversen, P. D. C. King, and P. Hofmann, Nature Communications **1**, 128 (2010).

[31] J. G. Analytis, R. D. McDonald, S. C. Riggs, J.-H. Chu, G. S. Boebinger, and I. R. Fisher, Nature Physics **6**, 960 (2010).

[32] J. Weissenstein, Applied Physics **5**, 217 (1974).

[33] C. Mann, D. West, I. Miotkowski, Y. P. Chen, S. Zhang, and C.-K. Shih, Nature Communications **In press**, (2013).

[34] L. A. Wray, S.-Y. Xu, Y. Xia, Y. S. Hor, D. Qian, A. V. Fedorov, H. Lin, A. Bansil, R. J. Cava, and M. Z. Hasan, Nature Physics **6**, 855 (2010).

[35] Y.-L. Wang, Y. Xu, Y.-P. Jiang, J.-W. Liu, C.-Z. Chang, M. Chen, Z. Li, C.-L. Song, L.-L. Wang, K. He, X. Chen, W.-H. Duan, Q.-K. Xue, and X.-C. Ma, Physical Review B **84**, 075335 (2011).

[36] Y. S. Hor, a. J. Williams, J. G. Checkelsky, P. Roushan, J. Seo, Q. Xu, H. W. Zandbergen, A. Yazdani, N. P. Ong, and R. J. Cava, Physical Review Letters **104**, 057001 (2010).

[37] K.-J. Chao, C.-K. Shih, D. Gotthold, and B. Streetman, Physical Review Letters **79**, 4822 (1997).

[38] X.-L. Qi and S.-C. Zhang, Physics Today **63**, 33 (2010).

[39] H. Beidenkopf, P. Roushan, J. Seo, L. Gorman, I. Drozdov, Y. S. Hor, R. J. Cava, and A. Yazdani, Nature Physics **7**, 939 (2011).

[40] Y. Wang, D. Hsieh, D. Pilon, L. Fu, D. Gardner, Y. Lee, and N. Gedik, Physical Review Letters **107**, 207602 (2011).

[41] Y. Zhao, Y. Hu, L. Liu, Y. Zhu, and H. Guo, Nano Letters **11**, 2088 (2011).

[42] T. Hanaguri, K. Igarashi, M. Kawamura, H. Takagi, and T. Sasagawa, Physical Review B **82**, 081305 (2010).

[43] M. Romanowich, M.-S. Lee, D.-Y. Chung, S. D. Mahanti, M. G. Kanatzidis, and S. H. Tessmer, Physical Review B **87**, 085310 (2013).

[44] M. Bianchi, R. Hatch, J. Mi, B. Iversen, and P. Hofmann, Physical Review Letters **107**, 086802 (2011).

[45] P. King, R. Hatch, M. Bianchi, R. Ovsyannikov, C. Lupulescu, G. Landolt, B. Slomski, J. Dil, D. Guan, J. Mi, E. Rienks, J. Fink, A. Lindblad, S. Svensson, S. Bao, G. Balakrishnan, B. Iversen, J. Osterwalder, W. Eberhardt, F. Baumberger, and P. Hofmann, Physical Review Letters **107**, 096802 (2011).

[46] Z.-H. Zhu, G. Levy, B. Ludbrook, C. Veenstra, J. Rosen, R. Comin, D. Wong, P. Dosanjh, A. Ubaldini, P. Syers, N. Butch, J. Paglione, I. Elfimov, and A. Damascelli, Physical Review Letters **107**, 186405 (2011).

[47] D. West, Y. Y. Sun, H. Wang, J. Bang, and S. B. Zhang, Physical Review B **86**, 121201 (2012).

[48] Y. Zhang, V. W. Brar, C. Girit, A. Zettl, and M. F. Crommie, Nature Physics **5**, 722 (2009).

[49] R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, 1994), p. 637.

[50] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).

[51] P. E. Blöchl, Physical Review B **50**, 17953 (1994).

[52] G. Kresse and D. Joubert, Physical Review B **59**, 1758 (1999).

[53] T. Physik, T. U. Wien, and W. Hauptstrasse, Physical Review B **47**, 558 (1993).

[54] G. Kresse and J. Furthmuller, Physical Review B **54**, 11169 (1996).

[55] D. R. J. Tersoff, Hamann, Physical Review B **31**, 805 (1985).