

This is to certify that the

dissertation entitled

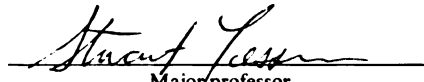
Low Temperature Scanning Tunneling Microscopy
Studies of the Surface and Impurity States
in Narrow Gap Semiconductors

presented by

SERGEI URAZHDIN

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Physics



Major professor

S. Tesser

Date 03/09/02



PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

LOW TEMPERATURE SCANNING TUNNELING MICROSCOPY STUDIES OF
THE SURFACE AND IMPURITY STATES IN NARROW GAP
SEMICONDUCTORS

By

Sergei Urazhdin

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics and Astronomy

2002

ABSTRACT

LOW TEMPERATURE SCANNING TUNNELING MICROSCOPY STUDIES OF THE SURFACE AND IMPURITY STATES IN NARROW GAP SEMICONDUCTORS

by

Sergei Urazhdin

In the techniques part of the thesis, the scanning tunneling microscopy (STM) technique and experimental methods, as well as methods of interpretation and theoretical analysis of STM data are described. The issues of design and development of STM scanning head, low temperature apparatus, and a highly sensitive low-temperature current amplifier are discussed. In the experimental part, STM measurements on the layered narrow gap semiconductors Bi_2Se_3 and Bi_2Te_3 are described. These semiconductors belong to a family of materials important for thermoelectric applications. We discuss the surface states observed with the STM probe, the presence of which is quite surprising in the strongly layered materials. Both *ab initio* and tight-binding model calculations are presented. The calculations elucidate the nature of the surface states and link them to the specific properties of the near-gap electronic structure, which are quite general to Bi_2Se_3 , Bi_2Te_3 and other materials with similar structure. We make a connection between the near-gap electronic structure and the thermoelectric performance of the narrow gap chalcogenide semiconductors. We also discuss the STM observation of defect states in Bi_2Se_3 doped with excess Bi. These states form peculiar clover-leaf shapes at the sample surface. As follows from the tight-binding analysis, the impurity states result from a subtle interplay of the surface and defect properties. We discuss possible implications for the STM observations of impurity states in other semiconducting compounds.

Acknowledgments

I would like to thank my wife, Melissa, and my parents for immense support they have given me over the many years of scientific struggle. It is their constant encouragement that made this work possible. I would also like to thank the people who collaborated in the project or gave invaluable advice: my adviser Prof. S. H. Tessmer, Prof. Norman O. Birge, Prof. M. I. Dykman, Prof. S. D. Mahanti, D. Bilc, Prof. W. Pratt Jr., Theodora Kyratsi, Prof. M. G. Kanatzidis, S. Lol, and Prof. T. Hogan. Besides sharing their insight in physics, they gave me a lot of encouragement and moral support, and made my graduate studies a very personal experience. I would like to thank my friends for sharing their life and research experiences. Finally, I would like to acknowledge the Physics and Astronomy department at Michigan State University for inviting me to the United States to do the things I enjoy. This work was in part supported by the Center for Fundamental Materials Research at Michigan State University and a grant from the National Science Foundation (DMR-0075230).

Contents

Acknowledgments	iii
1 Introduction	1
2 Principles, Methods, and Implementation of STM	4
2.1 A Very Simple Picture of STM Operation	6
2.2 STM Theory Based on Second Quantization	7
2.2.1 Metallic Electrodes	9
2.2.2 Electron Tunneling at Finite Temperature	10
2.2.3 Transmission Amplitude	11
2.3 STM Methods	12
2.3.1 Topographic Imaging	13
2.3.2 Multiple Bias Voltage Topographic Imaging	14
2.3.3 Point Spectroscopy	14
2.3.4 Current Imaging Tunneling Spectroscopy	15
2.3.5 Effective Work Function Measurement	16
2.4 STM Operation and Major Design Issues	17
2.4.1 Mechanical Stability of STM	17
2.4.2 A Low Temperature Tunnel Current Amplifier	21
3 STM of Surface and Defect states in Bi_2Se_3	29
3.1 Bi_2Se_3/Bi_2Te_3 Structure	29

3.2	Surface Preparation	32
3.3	Atomic Resolution STM Imaging of Bi_2Se_3	33
3.4	STM of Topographic Defects in Bi_2Se_3	36
3.5	STS Studies of Bi_2Se_3	39
3.6	Defect States in Semiconductors	41
3.7	STM Observations of Impurity States in Semiconductors	45
3.8	Impurity States in Bi-Doped Bi_2Se_3	47
4	Theoretical Modeling and Analysis of STM on Bi_2Se_3	56
4.1	Modeling STM with Band Structure Calculations	56
4.2	<i>ab initio</i> Modeling of STM on Bi_2Se_3	57
4.3	A Tight-Binding Model of Surface and Impurity States in Bi_2Se_3 . .	61
4.3.1	The Nature of the VBM State in Bi_2Se_3	64
4.3.2	Electronic Topological Transitions	65
4.3.3	Bi_{Se} Defect States in Bi_2Se_3	66
4.4	Defect States in Medium-Gap Semiconductors	67
5	Surface Effects in Bi_2Te_3	75
5.1	STM Studies of Bi_2Te_3	75
5.2	Surface Band Bending	77
5.3	Interpretation of the Electron Photoemission Experiments	80
5.4	The Thermoelectric Performance of Bi_2Te_3	81
6	Summary	84
A	STM Extensions Employing Capacitive Coupling	87
B	The Layered Nature of Bi_2Se_3 and Related Compounds	90
C	Semiclassical Theory of Thermoelectricity	93
C.1	Phenomenological Treatment of Transport	93

C.2	Semiclassical Transport Equations	94
C.3	Figure of Merit	96
C.4	A Single Parabolic Band Approximation	98
C.5	Beyond a Single Parabolic Band Approximation	100

List of Tables

3.1	Lattice parameters for Bi_2Se_3 and Bi_2Te_3 from Ref. [27].	31
4.1	Parameters used for the LCAO model calculation	62
B.1	Electronic density in the three filled levels of a 5-atom unit representing a single layer.	91

List of Figures

1	Schematic of a typical STM setup.	5
2	A schematic of electron tunneling between a tip and a sample through an insulating barrier. The electron wave function is periodic (Bloch) in the sample and the tip, and exponentially decaying in the vacuum gap, as the energy of the electron is below the vacuum barrier.	6
3	(a) Schematic of our Besocke-style STM design after Ref. [13]. (b) A photograph of our implementation of the design.	18
4	Schematic of the combined setup of an STM and a sample processing UHV chamber with direct sample transfer between them. The hatched bars represent the stages of vibration isolation loaded with lead bricks, as explained in the text.	20
5	Schematic of the circuit.	22
6	Equivalent circuit diagram.	23
7	Measured gain as a function of frequency vs. calculation.	24
8	Measured noise level (crosses). The Dashed line represents the Johnson noise amplitude of the measurement resistor at 4.2 K.	26
9	dI/dV spectrum of a 200nm thick superconducting Pb film at T=4.2 K. Squares: measurement, solid line: a fit as explained in the text. . . .	27

10	Bi_2Se_3 (Bi_2Te_3) structure. Top: View parallel to the layers. The rhombohedral unit cell is shown. In a dashed frame the chains of the $pp\sigma$ -bonds are plotted as discussed later. Bottom: A top view of the rhombohedral (1,1,1) plane. Open circles represent Se1(Te1) atomic positions, solid circles—subsurface Bi positions.	30
11	Rhombohedral Brillouin zone of Bi_2Se_3 and Bi_2Te_3 . Γ is the center of the zone, Z is [1,1,1] zone edge, M is a [1,1,-2] zone-edge point. . .	32
12	6.2x6.2 nm anomalous STM topographical images of Bi_2Se_3 surface. All the images are acquired in a single sequence under identical conditions over the same area of the surface, with bias voltage $V=-300$ mV, tunneling current $I = 0.75$ nA. Bottom insets show enhanced auto-correlated images. Top insets show schematically the tip geometry deduced from each image.	34
13	1.4x1.2 nm STM topographical images of Bi_2Se_3 surface, acquired at -0.7 V bias voltage, and 200 pA tunneling current. The black to white scale corresponds to 30 pm. Autocorellation was used to enhance the image.	35
14	(a), (b), (d) Topographic images of defects on the surface of nominally stoichiometric Bi_2Se_3 . The parameters are: (a) area 71x63 nm, $V=-50$ mV, $I=20$ pA, (b) area 8.5x7.5 nm, $V=-400$ mV, $I=100$ pA, (d) area 7.2x6.3 nm, $V=-600$ mV, $I=100$ pA. (c) Cross-section of image (b).	37
15	(a) $pp\sigma$ bonding chains in Bi_2Se_3 , with $pp\pi$ -type interaction between chains. (b) Origin of the topographic protrusions shown in Figure 14 from $pp\sigma$ chains.	39
16	Typical low-temperature dI/dV spectra of the stoichiometric Bi_2Se_3 surface. Left: large energy scale spectrum. A logarithmic vertical scale is used. Right: near-gap conductance acquired at a smaller tip-sample separation.	40

17	Hydrogenic impurity levels created by an ionized shallow donor impurity.	41
18	Semiclassical picture of the ionized donor effect on the local electronic properties.	44
19	(a) A 30×30 nm topographic map of $\text{Bi}_2\text{Se}_{2.85}$ acquired at bias $V = -0.2$ V, $I = 50$ pA. The total scale is 1.6 \AA . (b)-(f) CITS maps acquired simultaneously with (a) by fixing the tip at each point of the map (a) and measuring the differential conductance at different bias voltages, using a standard lock-in technique with ac bias voltage modulation of 4 mV. Biases used for CITS images are (b) -200 mV (the same as topographic mode bias), (c) -350 mV, (d) -400 mV, (e) -450 mV, and (f) -550 mV.	52
20	Differential conductance spectra acquired in $\text{Bi}_2\text{Se}_{2.95}$ in the vicinity of a clover-shaped impurity. The distance from the center is marked on the right. Curves are offset for clarity. Spectra (a) acquired along one of the clover lobes, and (b) along a line between the lobes, as shown in insets.	53
21	(a) A 5.1×5.1 nm topographic map encompassing one of the clover-like features. Sample bias voltage is -0.3 V, current setpoint 0.4 nA. (b) CITS current image map acquired at -0.7 V bias simultaneously with (a).	54

22	(a) A 3.5×3.5 nm topographic image of the same feature as Figure 21. The acquisition parameters are $V = -0.6$ V bias voltage and 1.0 nA current set-point. The inferred positions of the atoms terminating the $pp\sigma$ -bonded chains going through the defect atom are shown in small circles. The corresponding position of the subsurface defect is shown with a larger circle. (b) Crossection through (a). The defect corrugation is more than 1\AA , while the atomic corrugations away from the defect are only 20 pm. (c) A schematic of the interpretation of (a) in terms of $pp\sigma$ -chains.	55
23	Left: band structure calculation for the bulk Bi_2Se_3 . Right: band structure calculation for Bi_2Se_3 in slab geometry, with distance between 3-layer slabs increased by 7\AA . For notations see Figure 11. The position of the Fermi level is fixed arbitrarily.	70
24	(a) STM differential conductance spectra calculated as described in the text <i>vs.</i> the measurement, (b) Total calculated bulk <i>vs.</i> slab DOS.	71
25	Topographic map of Bi_2Se_3 calculated as described in the text.	71
26	Left: Schematic of a tight binding one-dimensional model used. Open orbitals represent Se atoms, solid orbitals represent Bi atoms. Right: Energy levels of a 5-atom unit representing a single layer. The non-bonding level is shown further split by the interlayer interaction V_3	72
27	(a) Evolution of the energy levels of a chain consisting of 5-atom Se1-Bi-Se2-Bi-Se1 units as the number of units is increased. (b) Amplitudes u_i (Equation (4.2)) in the VBM state for a 16-unit chain plotted as a function of atomic position along the chain. Only the first 30 amplitudes are plotted. Solid circles — Se1 positions, open circles — Se2 positions.	73

28	(a) Energy levels of a 16-unit chain (a total of 80 atoms) as a function of the position of a Bi_{Se} defect. The levels are calculated using a tight-binding model. (b) Same as Figure 27(b), with Bi_{Se3} defect introduced, shown with concentric circles. (c) A semiclassical schematic of the formation of the impurity resonance.	74
29	(a) A typical differential conductance spectrum of Bi_2Te_3 , (b) bulk vs. slab calculated DOS.	75
30	(a) Topographic image of a $0.5 \times 0.5 \mu\text{m}$ area of a Bi_2Te_3 sample doped with excess Te. The area has a high density of step edges and is not representative of the typical cleaved surface. Bias voltage $V = -80 \text{ mV}$. (b) A dI/dV map acquired simultaneously with (a). (c) Simultaneous cross-sections through (a) and (b). (d) Differential conductance spectra representative of high and low intensity areas in (b).	78
31	Left: Kelvin probe technique. The tip is vibrated above the surface, resulting in capacitively induced ac current through the tip dependant on the applied dc bias. Right: Scanning Capacitance microscopy technique. The small ac modulation of the bias voltage results in a capacitively induced current in the tip.	87
32	Contact potential effects.	89
33	Schematic of a setup for the analysis of thermoelectric phenomena. The direction of the heat flow is shown for n-type material, dominated by the Peltier contribution.	93
34	Schematic of a thermoelectric refrigerator setup.	96

Chapter 1

Introduction

The 20th century has seen a revolution in the understanding of electronic properties of solid state matter and subsequent applications of this knowledge. As an example it would suffice to mention the invention of the solid state transistor following the progress made in semiconductor physics. This invention shaped the technological development of the society in the last part of the century. The constant increase of the speed and efficiency of the semiconductor-based electronic devices has been achieved by minituarization. The recent progress made in nanofabrication techniques brings the size of semiconductor devices near the limits where quantum-mechanical and individual properties of a particular device become important [1]. This brings the need for a better understanding of the local properties of semiconductors, how they are affected by inhomogeneities: impurities, surface effects, etc. Moreover, with the recent emergence of the new field of quantum computing, it has been proposed to use an impurity state in semiconductor medium as an elementary quantum information unit, a qubit [2]. The possible implementations of such devices rely on the detailed knowledge of the near-surface impurity properties.

Scanning tunneling microscopy is a tool uniquely suited for the experimental investigation of local electronic properties with atomic spatial resolution, and sub-meV spectroscopic resolution of the electronic density of states at cryogenic temper-

atures. A number of STM studies of impurity states in semiconductors have been published [35]–[43], but a complete theoretical understanding of the experimental observations has not yet been achieved [33, 34, 45, 46]. The main difficulties are the identification of the impurity type and position with respect to the surface, factors which are important for the electronic structure of the impurity state. We present STM studies of the impurity states in a narrow gap semiconductor Bi_2Se_3 , which belongs to a class of semiconducting chalcogenides important for thermoelectric applications. Due to the bonding pattern in Bi_2Se_3 and similar compounds, the position of the impurity atom with respect to the surface can be determined precisely. We show that indeed the surface electronic properties in the case of Bi_2Se_3 are a dominant factor in the formation of the impurity state. These results have an important implication for understanding of the experimental data on impurities in other semiconducting compounds widely used in electronic devices, including GaAs and Si.

Our STM study of Bi_2Te_3 and Bi_2Se_3 is also motivated by their importance for the thermoelectric devices: Bi_2Te_3 alloys have been the most efficient room temperature thermoelectric materials for the past 40 years. Surprisingly enough, the exact origin of the electronic properties leading to high thermoelectric performance still remains elusive. It precludes researchers from significant gains in further optimization of thermoelectric materials. A dimensionless figure of merit used to characterize the thermoelectric performance of a material is defined as [64]

$$ZT = \frac{T\alpha^2\sigma}{\kappa}, \quad (1.1)$$

where α is Seebeck coefficient, σ is electrical conductivity, and κ is thermal conductivity (which includes both lattice and electronic contributions). It is shown in Appendix C, that the figure of merit can be related to the basic electronic and lattice properties of the material. In the semiclassical approximation, the figure of merit at the optimized doping value can be expressed through a single parameter B_i , combin-

ing all the material-specific properties [66],

$$B_i = \frac{1}{3\pi^2} \left(\frac{2kT}{\hbar^2} \right)^{3/2} \frac{(m_1 m_2 m_3)^{1/2}}{m_i} \frac{k^2 T N \tau}{\kappa_{ph}}, \quad (1.2)$$

where N is the degeneracy of the band extremum, m_i is the effective carrier mass in the direction of transport, τ is the electronic relaxation time, κ_{ph} is the lattice contribution to thermal conductivity. The figure of merit increases together with B_i , and Equation (1.2) shows which properties need to be optimized to improve thermoelectric efficiency. In the case of Bi_2Te_3 , the especially beneficial factors are low lattice thermal conductivity and 6-fold degeneracy of band extrema [73]. The thermoelectric devices based on Bi_2Te_3 alloys with Bi_2Se_3 and Sb_2Te_3 achieve $ZT \approx 1$, about 3 times less than the thermomechanical freon-based refrigerators, deterring their wider industrial use. Significant effort in designing materials with improved ZT factors has been directed towards further reduction of the thermal conductivity and increasing the degeneracy of the band extrema [69, 71, 73].

Our scanning tunneling microscopy studies and the following analyses provide a new perspective on the electronic structure of Bi_2Te_3 , Bi_2Se_3 , and related compounds, giving a rather general explanation for a number of existing experimental observations. The STM observation of prominent surface states in Bi_2Te_3 and Bi_2Se_3 is interpreted in terms of the strong dependence of the near-gap electronic structure on the weak interlayer coupling. This underscores the importance of considering the electron-lattice coupling, an effect beyond the simple semiclassical model typically employed in the analysis of the thermoelectric properties. This hypothesis gives a new possible direction for the ongoing search for better thermoelectric materials.

Chapter 2

Principles, Methods, and Implementation of STM

The STM operation, schematically shown in Figure 1, is based on the quantum-mechanical effect of electron tunneling through the space between a metallic tip and the surface of the material under study, when the tip is positioned within a few angstroms from the surface. STM of inert materials can be performed in ambient air, although atomic resolution studies of many materials require ultra-high vacuum of better than 10^{-10} torr. The tip is positioned with sub-angstrom precision by a piezoelectric tube. The voltages applied to the piezotube quadrants are controlled by electronics and can be correlated with the current detected by the tunneling current amplifier. For example, in a feedback mode the tip position is adjusted by a negative feedback circuit to keep the tunneling current to a set value. This is a typical mode of STM operation. The tip voltage is usually the virtual ground of the input of the current amplifier, while an adjustable bias voltage is applied to the sample to induce the tunneling current. The STM technique was originally employed [3] to resolve the surface properties of silicon on the atomic level. The observation of the surface reconstruction demonstrated that the new microscopic technique provided local information with resolution beyond that of the traditional indirect macroscopic

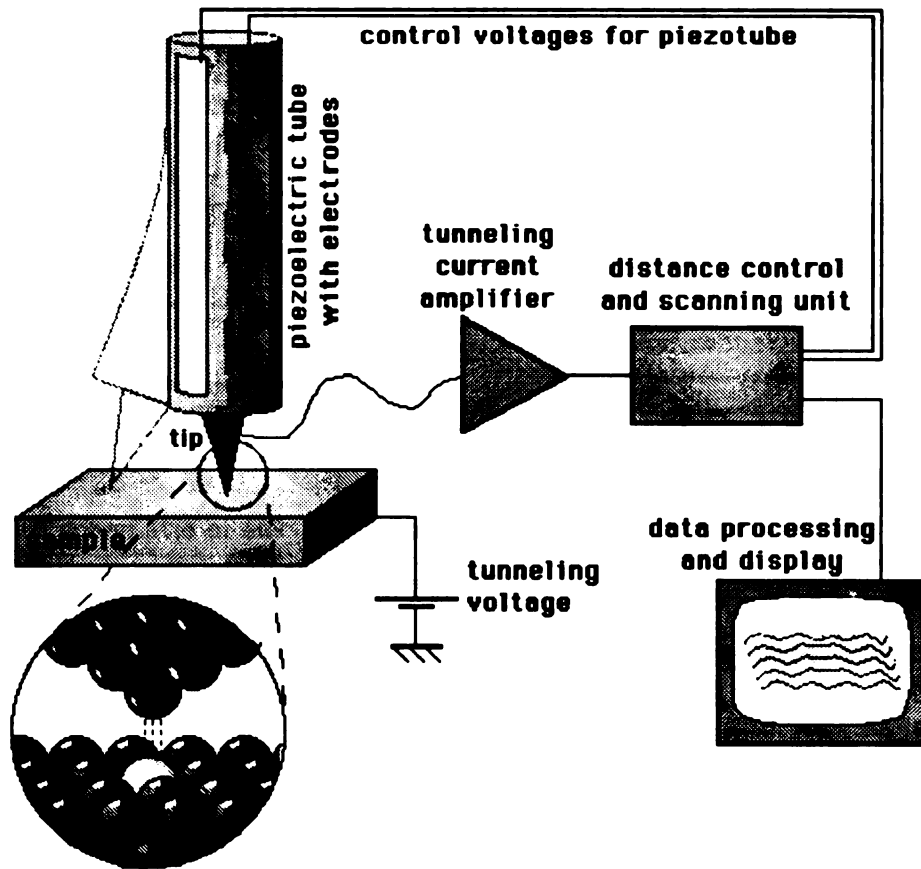


Figure 1: Schematic of a typical STM setup.

methods. Soon after, it was shown that, in addition to measuring surface topography, it is also a powerful probe of local electronic structure. The parameters that can be varied in STM operation are: tip position in space (3 coordinates), voltage applied between the tip and the sample (we will refer to it as sample bias, or simply bias, as usually the tip is at the virtual ground of the tunneling current amplifier), and the tunneling current. Of course, additional external conditions can be varied, for example magnetic field and temperature. The versatility of the STM probe arises from the ability of vary several parameters while performing local measurements. After deriving the basic theory of STM, we will describe a number of modes in which STM can operate, and the information that can be obtained from each mode.

2.1 A Very Simple Picture of STM Operation

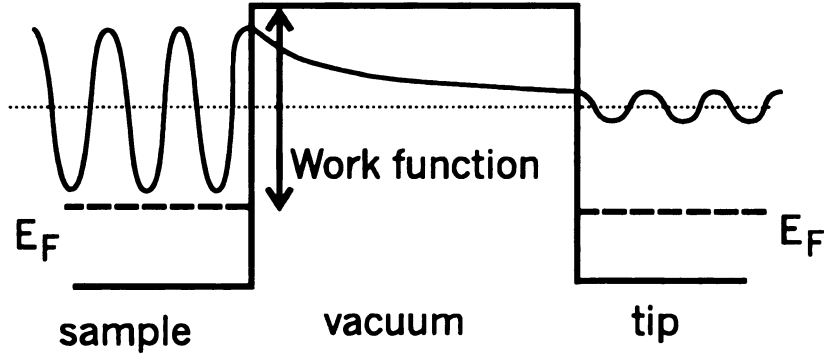


Figure 2: A schematic of electron tunneling between a tip and a sample through an insulating barrier. The electron wave function is periodic (Bloch) in the sample and the tip, and exponentially decaying in the vacuum gap, as the energy of the electron is below the vacuum barrier.

A convenient and simple model (shown in Figure 2) for the analysis of the electron tunneling is a narrow uniform gap between two electrodes with flat interface, representing the tip and sample surface. Consider two electrodes s (sample) and t (tip) biased with a voltage V with respect to each other. An intuitive picture of electron tunneling at zero temperature is that due to the applied bias voltage the electrons in the tip occupy energy levels that are empty in the sample. Due to the coupling of the electronic states through the vacuum gap these electrons tunnel into the sample at a rate determined by the overlap of the wave functions in the gap and the number of states of the same energy that are occupied in the tip and unoccupied in the sample

$$I = e \int_{E=E_F}^{E=E_F+eV} dE |M|^2 g_s(E) g_t(E - eV). \quad (2.1)$$

Here M is the tunneling matrix element determined by the overlap of the electronic wave functions, and g_s , g_t are electronic densities of states in the electrodes. It is not immediately obvious that when the quasiparticle nature of single particle excitations in electrodes is taken into account, g_s , g_t in Equation (2.1) should represent

single particle density of states, since electrons become "undressed" in the process of tunneling . Equation 2.1 also fails to describe superconducting electrodes. Josephson current can exist between them with no voltage bias applied. A more rigorous approach is also needed to describe the local and specifically surface effects observed with STM, not present in Equation (2.1).

2.2 STM Theory Based on Second Quantization

A more rigorous description of electron tunneling can be constructed using a second quantization formalism. Electron tunneling in this approach naturally arises as a result of the perturbation introduced by the proximity of the electrodes. The Josephson effect can also be incorporated by inclusion of higher-order terms in the Hamiltonian. This aspect, however, will not be addressed here. Tunneling between metallic electrodes with an independent of energy electronic density of states results in Ohmic behaviour of the tunneling junction, i.e. a linear relationship between the applied voltage and the tunneling current. Deviations from the Ohmic behaviour, i.e. nonlinearities in current-voltage dependence, result from energy dependence of the electronic density of states of the electrodes, as in the simplistic Equation (2.1), and energy dependence of the tunneling matrix element. The latter effect can be usually neglected for sufficiently small bias voltage. However, we show in this section that the combination of the tunneling matrix element and electronic density of states appearing in Equation (2.1) should be understood as the local surface electronic density of states, not necessarily reflecting the bulk electronic properties of the electrodes. The surface effects turn out to be an important factor for the correct interpretation of the STM experiments on semiconductors.

Let's first assume that the electrodes are separated by a significant distance, so

the system is described by a Hamiltonian

$$H_0 = \sum_{s,\sigma} \epsilon_s a_{s\sigma}^\dagger a_{s\sigma} + \sum_{t,\sigma} \epsilon_t b_{t\sigma}^\dagger b_{t\sigma}, \quad (2.2)$$

where $a_{s\sigma}^\dagger$ is a creation operator for a quasiparticle in the sample with quantum number s and spin σ , $b_{t\sigma}^\dagger$ is a creation operator for a quasiparticle in the tip etc. As the electrodes are brought in close proximity to each other, electrons can tunnel between them, with tunneling process described by $a^\dagger b$ and $b^\dagger a$ operators. The single-quasiparticle perturbation to the Hamiltonian (2.2) due to the overlap of electronic states in electrodes is

$$H_1 = \sum_{s,t\sigma} (t_{st} a_{s\sigma}^\dagger b_{t\sigma} + t_{ts} b_{t\sigma}^\dagger a_{s\sigma}), \quad (2.3)$$

where

$$\begin{aligned} t_{st} &= \langle a_{s\sigma}^\dagger 0 | H | b_{t\sigma}^\dagger 0 \rangle \\ t_{ts} &= \langle b_{t\sigma}^\dagger 0 | H | a_{s\sigma}^\dagger 0 \rangle = t_{st}^* \end{aligned} \quad (2.4)$$

are integrals characterizing the overlap of the electronic states in the electrodes. Suppose that a voltage V is applied between the electrodes, so that energies of the states in the tip (counted from the fixed energies of the sample states) are modified

$$\epsilon_t \rightarrow \epsilon_t + eV, \quad (2.5)$$

where e is the absolute value of the electron charge. According to the Fermi golden rule, if the state i is filled, and state f is empty, the transition rate between these states is

$$\Gamma_{if} = \frac{2\pi}{\hbar} |\langle f | H | i \rangle|^2 \delta(E_f - E_i). \quad (2.6)$$

In a statistical ensemble described by a finite temperature T , the total transition rate between the electrodes is described by the sum of partial rates (2.6), weighted by the Fermi distribution function $f(\epsilon) = \frac{1}{e^{(\epsilon-\mu)/k_B T} + 1}$

$$\begin{aligned} \Gamma_{st} &= \frac{2\pi}{\hbar} \sum |t_{st}|^2 f(\epsilon_s) (1 - f(\epsilon_t)) \delta(\epsilon_t - \epsilon_s - eV) \\ \Gamma_{ts} &= \frac{2\pi}{\hbar} \sum |t_{ts}|^2 f(\epsilon_t) (1 - f(\epsilon_s)) \delta(\epsilon_t - \epsilon_s - eV) \end{aligned} \quad (2.7)$$

where Γ_{st} and Γ_{ts} are the tunneling rates from the sample to the tip and *vice versa*. Here we assume that the tunneling current does not significantly violate the equilibrium distribution in the electrodes. Since $|t_{st}|^2 = |t_{ts}|^2$, we will omit the subscript from here on. The notation t is now used both as a subscript numbering the tip states and for the transmission amplitude, but confusion can be hopefully avoided due to the different meanings of the two. The average tunneling current is

$$I = -e(\Gamma_{st} - \Gamma_{ts}) = -\frac{2e\pi}{\hbar} \sum_{s,t,\sigma} |t|^2 (f(\epsilon_s) - f(\epsilon_t)) \delta(\epsilon_t - \epsilon_s - eV). \quad (2.8)$$

It is convenient to replace summation with integration over the energies, multiplying the expression by the density of quasiparticle states per unit energy. Assuming also no spin dependence,

$$\sum_{s,t,\sigma} \rightarrow 2 \int d\epsilon_s g_s(\epsilon_s) d\epsilon_t g_t(\epsilon_t) \quad (2.9)$$

where $g_s(E), g_t(E)$ are densities of states in the sample and the tip respectively. Therefore

$$I = -\frac{4e\pi}{\hbar} \int d\epsilon |t|^2 g_s(\epsilon) g_t(\epsilon + eV) (f(\epsilon) - f(\epsilon + eV)). \quad (2.10)$$

Transmission amplitude t and density of states $g(\epsilon)$ give complementary contributions to (2.10). The product tg_s can be related to the local density of the sample electronic states in the vacuum gap. This statement will be quantified in the analysis that follows.

2.2.1 Metallic Electrodes

To further simplify expression (2.10) the transmission amplitude t is usually approximated by a constant independent of energy. If the density of states is also assumed to be independent of energy (large spherical Fermi surface), then the integral in (2.10) can be calculated analytically

$$\int d\epsilon (f(\epsilon) - f(\epsilon + eV)) = eV, \quad (2.11)$$

and the result is independent of temperature. Formula (2.10) then can be written as Ohm's law

$$I = \frac{V}{R} \quad (2.12)$$

where

$$R = \frac{\hbar}{\pi 4e^2 |t|^2 g_s g_t} \quad (2.13)$$

The quantum-mechanical origin of tunneling resistance does not show up in formula (2.12). However, in (2.13) $\frac{\hbar}{\pi 4e^2}$ is, to the precision of a factor of order unity, a quantum resistance. In derivation of (2.12) we assumed weak coupling of the electronic states in the electrodes so that it could be treated perturbatively. To satisfy this condition, the tunneling resistance (2.13) should be much larger than the quantum resistance $R_Q = 26 \text{ k}\Omega$. Typical tunneling resistance in STM experiments is at least $1 \text{ M}\Omega$. Experiments with point contacts indeed show strong deviations from Ohmic behavior starting at $R \approx 500 \text{ k}\Omega$. In the simplified picture, illustrated previously in Figure 2, at small electrode separations the vacuum barrier height is lowered by the image potential [7]. This increases the effect of the electrostatic barrier bending (due to the applied bias voltage) on the amplitude of the tunneling current, distorting the linear relation (2.12).

2.2.2 Electron Tunneling at Finite Temperature

Assume that the density of states in the tip is independent on energy. From (2.10) the differential tunneling conductance defined as $G(V) = \frac{dI(V)}{dV}$ is

$$G = \frac{4e^2 \pi g_t}{\hbar} \int d\epsilon |t|^2 g_s(\epsilon) \frac{\partial f(\epsilon + eV)}{\partial \epsilon}. \quad (2.14)$$

At $T=0$ the Fermi distribution becomes a step function, its derivative is a δ -function, so the differential conductance is a measurement of density of states

$$G = \frac{4e^2 \pi g_t}{\hbar} |t|^2 g_s(E_F - eV). \quad (2.15)$$

At finite temperatures the differential conductance represents the density of states convoluted with a curve $\frac{\partial f(\epsilon+eV)}{\partial \epsilon}$, effectively smearing the measurement on the scale of about $3.5kT$, where k is Boltzmann constant. At room temperature the scale of smearing is about 100 mV, while at liquid He-4 temperature of 4.2K it is roughly 1 mV.

2.2.3 Transmission Amplitude

Transmission amplitude t and sample density of states g_s enter the expression (2.10) for the tunneling current only as a product $|t|^2 g_s$. This combination can be related to the local sample density of states at the position of the tip. Although the influence of the specific properties of the electronic states of the tip on the STM performance has been investigated [4], the general irreproducibility of the tip properties limits the possibility for a complete analysis of particular experimental data. Typically an approximation of spherically symmetric tip wave functions is made to simplify the analysis.

The wave function of the sample electrons in the vacuum gap can be expanded in the form

$$\psi_s = \Omega_s^{-1/2} \sum_{\mathbf{G}} a_{\mathbf{G}} \exp[-(k^2 + |\mathbf{k}_{\mathbf{G}}|^2)^{1/2} r_{\perp}] \exp(i\mathbf{k}_{\mathbf{G}} \mathbf{r}_{\parallel}), \quad (2.16)$$

which is a completely general expansion in the region of negligible crystal potential. Here Ω_s is sample volume, $k = \hbar^{-1}(2m\phi)^{1/2}$ is the minimum inverse decay length in vacuum for states close to Fermi energy, ϕ is the work function, \mathbf{G} is a two-dimensional reciprocal lattice vector along the surface, \mathbf{r}_{\parallel} , r_{\perp} are coordinate components respectively parallel and perpendicular to the surface, and $\mathbf{k}_{\mathbf{G}} = \mathbf{k}_{\parallel} + \mathbf{G}$. k_{\parallel} is the wave vector along the surface, which is a good quantum number. In most studies, especially those concerned with small energy scale phenomena like superconductivity, the dependence of the wave function decay rate on the momentum $\mathbf{k}_{\mathbf{G}}$ in (2.16) is neglected. However, in discussing the anomalous atomic corrugation in graphite, this

dependence turns out to be an important factor due to divergences appearing in this case for quasi-momentum at the edge of the Brillouin zone [5].

It was shown in Ref. [6] that the transmission amplitude can be approximated by

$$t_{st} = \frac{\hbar^2}{2m} 4\pi\Omega_t^{-1/2} R e^{kR} \psi_s(\mathbf{r}_0), \quad (2.17)$$

where \mathbf{r}_0 is the position of the center of curvature the tip, R is its radius, and Ω_t is the tip volume $\Omega_t = \frac{4\pi}{3} R^3$. The transmission amplitude (2.17) can be expressed through the value of surface wave function ψ_s at a single position \mathbf{r}_0 due to the specific assumptions made about the wave function (expression (2.16)), and a similar assumption for the tip wave function behavior. However, it does not imply that only $\psi_s(\mathbf{r}_0)$ is the only physically relevant quantity in the tunneling ocess. Equation (2.17) facilitates an easy way to model the STM tunneling current. Electronic structure calculations provide the charge density as a function of position and energy, in particular at the position \mathbf{r}_0 above the sample surface. The lattice calculations can be used as a model of the surface by separating slabs of material (thick enough to recover the electronic properties of the bulk) by sufficiently large gaps, in a supercell geometry. Formula (2.10) together with (2.17) can be used to calculate the tunneling current. As the absolute distance between the STM tip and the surface is usually not known, only the relative variations of the current with energy or lateral position above the surface are of interest, and the current (or differential conductance) is typically normalized to some arbitrary scale.

2.3 STM Methods

The standard methods of STM include topographic imaging, multiple bias voltage topographic imaging, point spectroscopy, current imaging tunneling spectroscopy, and effective work function measurement. Besides that, in our setup with a highly sensitive current amplifier (described in Section 2.4) we were able to perform Kelvin probe

and scanning capacitance measurements, which are extensions of the STM technique in which capacitively induced current in the tip is measured. These extensions of STM are described in Appendix 31.

2.3.1 Topographic Imaging

The very first STM technique invented, constant current surface topographic imaging is probably the most widely used. The atomic-scale resolution is achieved due to the extreme sensitivity of the tunneling current on the tip-sample separation: It has been derived in Sec. (2.2), that the tunneling current falls off as a function of tip-sample separation roughly as

$$I \propto |\psi_s|^2 = \exp(-2kr_{\perp}), \quad (2.18)$$

where $k_{\perp} = \hbar^{-1}(2m\phi)^{1/2}$, ϕ is the surface work function, and m is the actual electron mass (which is the parameter relevant in the vacuum gap). If k_{\perp} is measured in \AA^{-1} , then numerically

$$k_{\perp}(\text{\AA}^{-1}) \approx 0.51\sqrt{\phi(\text{eV})}. \quad (2.19)$$

The typical value of the surface work function is 3–6 eV, resulting in $k \approx 1\text{\AA}^{-1}$. Therefore, tunneling current (2.18) varies roughly by an order of magnitude as the tip-sample separation is changed by 1\AA . When the tip is scanned above the sample surface, atomic scale surface topography variations produce significant tunneling current variations. In practice, the desired value of the tunneling current is fixed, and the tip vertical position above the surface is adjusted by the feedback circuit to keep the set current value. As the tip is scanned laterally above the sample surface, its vertical position is recorded. Such a measurement has been shown [3] to be sensitive enough to detect atomic scale variations on the sample surface.

It should be pointed out here that the tunneling current is proportional to the densities of states in both the STM tip and the sample and the square of the trans-

mission amplitude between the two. The transmission amplitude between a periodic atomically flat surface and a tip at a fixed distance from the surface plane also varies with atomic periodicity. This, in turn, produces the atomic corrugations in the topographic image acquired with a fixed current. However, this does not mean that the protrusions in the topographic image correspond to the positions of actual atoms [9, 5]. If there are several inequivalent atomic positions at the surface (like on (1,1,0) surface of GaAs), then, depending on the bias voltage, different atoms can give strong contributions to atomic corrugations, resulting in bias-dependant shift in the STM atomic images.

Another important aspect of the topographic imaging follows from Equation (2.16). The variation of the wave function at the surface can be expanded in the reciprocal surface lattice vectors. For sufficiently large tip-sample separation all high order components are exponentially small and can be neglected. Therefore, the topographic image of an atomically flat surface typically represents only the zero-order terms of the surface wave function variation, i.e. just a product of two sinusoidal plane waves.

2.3.2 Multiple Bias Voltage Topographic Imaging

The tunneling current is proportional to the density of states integrated between the Fermi-level and the applied bias voltage. Therefore, the corrugations in topographic STM images reflect the local variation in the density of states. By comparing the topographic images acquired at various bias voltages important spectroscopic information can be gained, i.e. information about the local relative variation of the density of states.

2.3.3 Point Spectroscopy

This technique is a direct analog of the much earlier developed planar junction tunneling technique [12]. The position of the STM tip can be fixed with respect to the

surface. The tunneling current is subsequently measured as a function of the applied bias voltage. The differential conductance spectrum obtained by differentiation of the $I(V)$ curve is a direct measure of the surface density of states, as has been shown in Section 2.2. A variation of this technique is based on lock-in detection. The bias voltage is modulated with a small ac signal:

$$V = V_0 + \delta V \cos(\omega t) \quad (2.20)$$

The resulting tunneling current is to the first order in δV

$$I(V) = I(V_0) + \left(\frac{dI}{dV} \right)_{V_0} \delta V \cos(\omega t). \quad (2.21)$$

Therefore, the ac component of the tunneling current, which can be measured with a lock-in amplifier, is proportional to the differential conductance at the applied dc bias voltage.

2.3.4 Current Imaging Tunneling Spectroscopy

With increased capabilities of STM controllers, current imaging tunneling spectroscopy (CITS) technique has become increasingly popular. The tip is scanned over the sample surface in the feedback (topographic imaging) mode. At each point of the image the feedback is disabled, the bias voltage is changed to a different value, and the tunneling current is measured. The acquired map of tunneling current values constitutes a current image. If the surface is electronically homogeneous, i.e. the electronic properties are identical at each point, then the CITS map is flat. The features in the CITS map appear due to the relative variations of density of states at CITS imaging bias as compared to the topography imaging bias voltage.

An alternative method of CITS mapping is to measure $I(V)$ curves at each point of the image and then create a map of current at a given bias normalized by the current value at some other fixed bias voltage. This method requires a significant amount of memory and acquisition time, but has the capability to create a CITS

map representing arbitrary bias conditions within the limits set by the acquired $I(V)$ curves.

2.3.5 Effective Work Function Measurement

The origin of the versatility of the STM technique lies in the significant number of parameters which can be varied in the process of acquisition: three coordinates representing the tip position, the bias voltage and the tunneling current. One of these parameters is determined as a function of the rest. In the topographic imaging method described above, the tip height is measured as a function of its lateral position above the surface; in point spectroscopy the current is measured as a function of bias; in CITS the current is measured as a function of its lateral position. The only feasible combinations left are to measure the current as a function tip height or tip height as a function of the applied bias. While the latter technique gives only convoluted information about the electronic properties, the former is known as a measurement of the surface effective work function. From Equation (2.16) it follows that the tunneling current exponentially depends on the tip-sample separation

$$I(r_{\perp}) = \sum_G I_G \exp[-(k^2 + |k_G|^2)^{1/2} r_{\perp}], \quad (2.22)$$

where I_G is a factor depending only on the lateral tip position. By measuring the tunneling current dependence on the tip-sample separation at various bias voltages, information about the wave vectors of the electronic states near Fermi-level can be obtained [9]. The application of this method has been limited by the strong dependence of the experimental results on the poorly reproducible tip properties and the quality of the sample surface. If the vacuum tunneling conditions are violated, e.g. due to an insulating layer of hydrocarbons or water on the surface, the exponential dependence (2.22) is drastically modified to a much weaker decay. In this mode, the effective work function measurement is widely used to characterize the quality of

the sample surface and the vacuum tunneling conditions. If tunneling occurs through an insulating layer significantly reducing the tunnel barrier, this compromises the spectroscopic measurements due to the electrostatic barrier distortion caused by the applied bias voltage. The typical effective work function values for metallic surfaces are about 1 eV, much lower than the actual work function values of 3–6 eV. This discrepancy has been a subject of theoretical research [7, 8]. In all the STM experiments presented in this thesis the effective work function measurement was performed prior to the other measurements to ensure the quality of the sample and tip surfaces.

2.4 STM Operation and Major Design Issues

STM technique is often referred to as art, rather than science, because of the multitude of issues an experimentalist has to deal with to achieve successful STM operation. Experiments are usually repeated a number of times to verify the meaningfulness and reproducibility of the results. The important issues one has to deal with include:

- Mechanical Stability of the STM
- Sample surface preparation and preservation (UHV vs. non-UHV environment)
- STM tip preparation, reproducibility and stability
- Tunneling current amplification and noise filtering

Below we will discuss some of the important considerations and developments in the design of our STM system.

2.4.1 Mechanical Stability of STM

As discussed in Section 2.3.5, the tunneling current changes roughly by an order of magnitude as the tip-sample separation is varied by 1 Å. In spectroscopic measurements, the tip position is usually fixed and then the current is measured as a function

of applied bias voltage. The mechanical vibrations should not produce variations in tunneling current of more than a few percent; therefore the necessary tip-sample mechanical stability has to be better than 0.1\AA . Such an incredible mechanical stability is achieved by use of extremely stiff STM scanner head designs and at least one or two stages of vibration isolation. High frequencies of the normal vibrational modes of a stiff scanning head couple very weakly to the predominantly low frequency mechanical vibrations present in a laboratory.

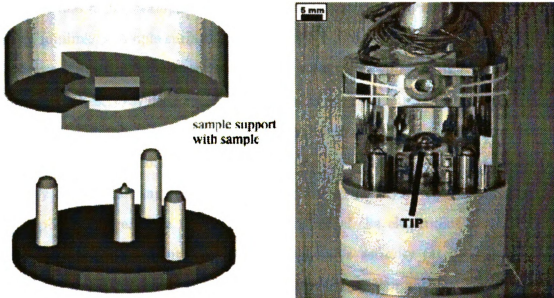


Figure 3: (a) Schematic of our Besocke-style STM design after Ref. [13]. (b) A photograph of our implementation of the design.

Our design of the STM scanning head is shown in Figure 3(a). The sample holder is machined in the shape of three spiral ramps (shown on top). Three outer piezoelements (tubes) are terminated with smooth balls and support the sample holder. The motion of the sample holder with respect to the tip, which is attached to the central piezotube, is achieved by sawtooth-like motion of the supporting piezoelements. During the fast part of the pulse the sample holder slides with respect to the supporting balls due to inertia. During the slow part of the sawtooth pulse the sample holder moves along with the supporting elements due to the friction between the ramps

and the supporting balls. Depending on the signs of the pulses applied to individual piezoelements, positioning in all three directions is achieved.

Although not the most mechanically stable of the existing designs [14], it has a number of advantages. The design is simple and robust, and incorporates very few parts. It also allows probing large areas of the sample through the lateral coarse positioning, a feature not available in many other STM designs, including Ref. [14]. Another feature incorporated in our design is *in situ* transfer of the sample between the STM and a UHV sample processing chamber, shown in Figure 4. Such a coupled design minimizes sample surface contamination between the sample preparation and STM experiment. The first microscope we built (see Figure 3(b)) incorporated 1" long piezotubes, providing mechanical stability of tip-sample separation of about 0.1 Å with a double-stage external vibration isolation. A later version incorporated shorter 0.75" long piezoelements with thicker walls, further increasing the mechanical stability of the scanning head and reducing the sensitivity of the piezoelements to the applied voltage. The typical electrical noise of the electronics controlling the piezoelements [15] is about 2 mV RMS. With the sensitivity (at T=300 K) of the 1" long scanning piezotube of 1500 Å/V for lateral motion and 50 Å/V for motion perpendicular to the surface, the electrical noise in the control electronics results in the mechanical vibration of the tip with respect to the surface of about 3 Å RMS laterally and 0.1 Å RMS perpendicular to the surface. These numbers give the lower limit for the STM lateral resolution and vertical stability at room temperature. Achieving atomic resolution with such a high level of vibrations at room temperature has proven to be extremely difficult. By reducing the sensitivity in the later version of the STM scanning head we were able to decrease the effect of the electrical noise on the mechanical stability. The sensitivity was decreased 5 times for the piezo motion along the surface, and 2 times for the motion perpendicular to the surface.

To reduce the coupling of external vibrations to the STM setup, external and/or internal vibration isolation systems are usually implemented. Internal vibration iso-

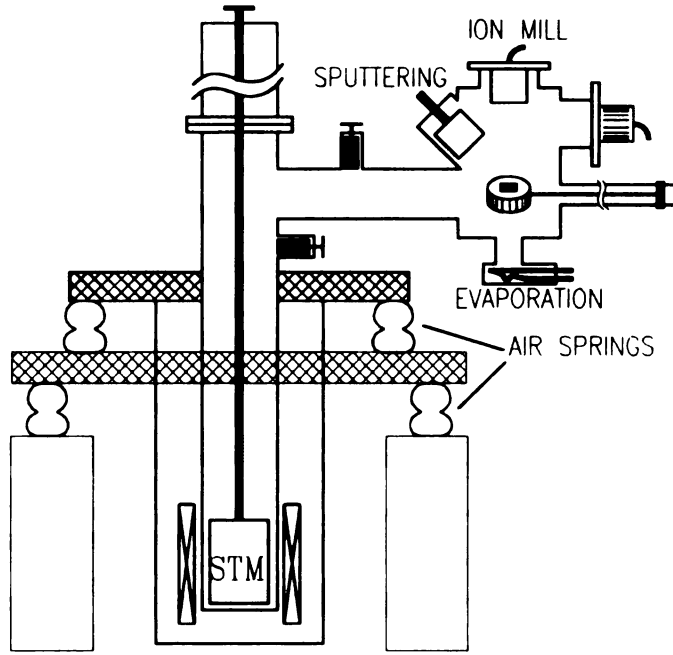


Figure 4: Schematic of the combined setup of an STM and a sample processing UHV chamber with direct sample transfer between them. The hatched bars represent the stages of vibration isolation loaded with lead bricks, as explained in the text.

lation, typically soft springs or bellows separating the STM head from the rest of the system, prove to be very efficient due to the very low associated resonance frequency of the suspended scanning head. In our design the necessity for the sample transfer between the sample preparation UHV chamber and STM has limited the available vibration isolation to the external type. The whole low temperature STM setup together with the rigidly attached preparation chamber were suspended on air springs. Initially we incorporated only a single stage, later extending it to a double stage. Model calculations [16] have shown that inclusion of the second stage with proper loading of both stages increases the damping factor by at least two orders of magnitude at frequencies above 50 Hz. The optimal calculated weight of both stages has been achieved with lead bricks totaling 200 kg for the intermediate stage and about 100 kg for the final stage. This brings the total weight of the final stage, including the UHV sample processing chamber and a cryostat, to about 300 kg. Subsequent

tip-sample vibration amplitude measurements showed, however, that no significant gain in stability had been achieved by incorporation of the second vibration isolation stage. It is likely that other noise sources dominated the test results, in particular electrical noise in the piezotubes, STM tip instability, and acoustic to mechanical coupling. The latter has been verified by producing sound, and observing the effect on the stability of the STM tunneling current.

2.4.2 A Low Temperature Tunnel Current Amplifier

In a typical STM measurement, the electronic properties on the energy scale of 10 mV to 1 V are investigated in a weak coupling regime (tunneling resistance of at least 1 M Ω). Therefore, the tunneling current to be detected is about 10 nA, requiring a low noise tunnel current amplifier. Typical STM current amplifiers have noise level referred to the input current of about $10^{-14} - 10^{-13} \text{ A}/\sqrt{\text{Hz}}$. The noise level is significantly increased in the low-temperature STM setup, where long cables are used to connect the tunnel junction to the room temperature amplifier. The additional noise comes from the coupling to electromagnetic radiation, as well as the triboelectric effect (current induced in the cables by their mechanical vibration).

The current amplifier performance is also compromised by the dual role it has to play in the STM setup. On the one hand, it is used to detect dc tunnel current enabling the feedback loop to keep a constant tip-sample separation. On the other, in spectroscopic measurements, it is used for ac current lock-in detection resulting from a small ac variation of the bias voltage. The small sustainable dissipation level in sorption pumped He3-based systems or dilution refrigerators (typically in the μW) range, and the lack of available space in proximity to the STM tip also contribute to the difficulties in the development of suitable low-temperature (LT) amplifiers. Another major issue in the LT electronics design is a limited choice of LT compatible active devices. Bipolar transistors, a natural choice for current to voltage converters, cannot be used at temperatures below about 100 K due to carrier freeze-out.

A number of amplifier designs have been proposed based on field effect transistors (FET) [17]–[19]. These devices show improved performance down to liquid nitrogen temperatures due to decreased carrier thermal scattering, but at lower temperatures the scattering becomes dominated by the impurities and the depletion mode performance is decreased due to carrier freeze-out [21]. Moreover, FET operation in saturation mode results in power dissipation in the mW range, making the heat load unacceptable for He3 based systems.

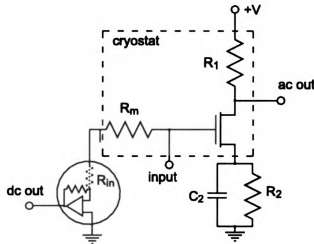


Figure 5: Schematic of the circuit.

We chose to use a cryogenic high electron mobility transistor [22] (HEMT) as an active device in our circuit, due to its inherently superior low-temperature performance, as well as low input capacitance. Since the operation of the device does not rely on thermally activated charge carriers, it does not suffer from carrier freeze-out. Moreover, carrier mobility increases monotonically down to liquid helium temperatures, as scattering of the charge carriers off the originating dopant atoms is greatly reduced due to their spatial separation in the heterostructure geometry (modulation doping). With the transistor operated at source-drain voltage well below saturation, the power dissipation is reduced by 2–3 orders of magnitude compared to satura-

tion mode operation. As our measurements show (see below), the output noise of the circuit is dominated by the Johnson noise of the measurement resistor, thus justifying the choice of HEMT operation mode. One of the important advantages of our amplifier design is a high bandwidth allowing the use of this amplifier not only for the tunneling current detection, but also for Kelvin probe and local capacitance measurements [10].

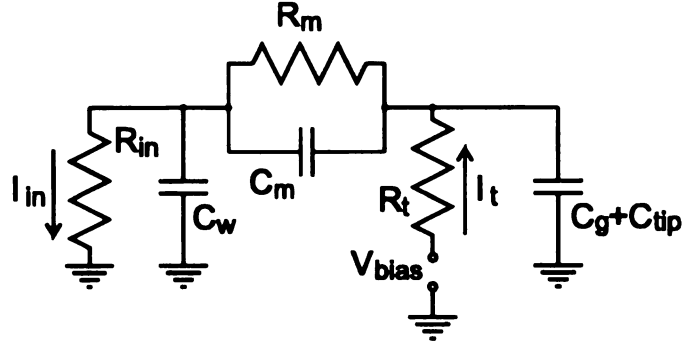


Figure 6: Equivalent circuit diagram.

We use dedicated circuits for separate amplification of the dc tunnel current used for feedback loop operation (dc part) and ac spectroscopic lock-in detection. This allows us to optimize the parameters of each circuit specifically for its application. Fig. 5 shows the schematic of the circuit, with the cryogenic part boxed by the dashed line. The typical operating parameters of the circuit are $V=5$ V, $R_m=18$ M Ω , $R_1=100$ K Ω , $R_2=10$ K Ω , and $C_2=1$ pF. In the ac preamplifier circuit, the current is converted into a voltage across the measuring resistor R_m . The resistor is connected to the virtual ground input of the room temperature current amplifier (see, e.g., Ref. [18, 19]) with input impedance R_{in} much less than R_m . R_1 in series with a voltage source serves as a current source for the transistor, resulting in the output voltage proportional to the HEMT drain-source resistance variation. R_2 shunted by C_2 is used to bias the gate of the HEMT with respect to the drain-source channel.

A major source of noise in the LT circuit is Johnson noise in the measurement

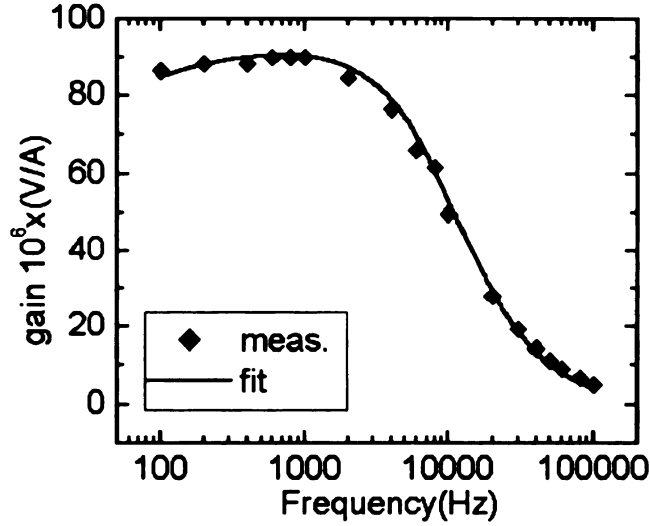


Figure 7: Measured gain as a function of frequency vs. calculation.

resistor. Referred to the voltage on the measurement resistor, it is given by [20]

$$V_J = \sqrt{4kT\Delta f ReZ} \quad (2.23)$$

where k is the Boltzmann constant, T is the absolute temperature of the resistor, Δf is the bandwidth of the lock-in measurement and ReZ is the real part of the circuit input impedance. Figure (6) shows the equivalent schematic of the circuit used to calculate the circuit input impedance and other electrical parameters. The schematic shows stray capacitances of the cable C_w , measurement resistor C_m and the tip C_{tip} , as well as HEMT gate to source-drain channel capacitance C_g . R_{in} represents the input impedance of the room temperature current amplifier. I_{in} represents the part of the tunnel current I_t flowing through the room temperature current amplifier.

The input impedance of the circuit can be represented as

$$Z = \frac{R_m}{1 + i\omega C_{shunt} R_m} \quad (2.24)$$

where C_{shunt} is the capacitive contribution of the circuit elements to the input impedance and is discussed below. To characterize the sensitivity of the current amplifier, we relate the voltage noise (2.23) to the input current using $I_J = V_J/|Z|$. Plugging (2.24)

into (2.23) we get

$$I_J = \sqrt{\frac{4kT\Delta f}{R_m}} \quad (2.25)$$

Formula (2.25) shows that to minimize the Johnson noise amplitude referred to the input current, it is beneficial to use the largest possible value of the measurement resistor. However, the input impedance of the circuit has to be much smaller than the signal source impedance (tunnel resistance). As the measurement resistor value approaches the tunnel resistance, the voltage drop across the measurement resistor becomes significant, distorting the voltage scale of the spectroscopic measurements. In an alternative approach, the measurement resistor can be made significantly larger and placed in the feedback loop of an inverting amplifier [18] resulting in a small input impedance. However this approach leads to significant design complications not compatible with low temperature operation. In STM measurements, the tunnel resistance is typically in the $1\text{M}\Omega$ - $1\text{G}\Omega$ range, limiting the possible value of the measurement resistor to the $\text{M}\Omega$ range. For measurement resistor values comparable to the tunnel resistor, however, the bias voltage distortion has to be taken into explicit account in analyzing the spectroscopic data. The measurement resistor value also determines the bandwidth of the amplifier. An RC low-pass filter is formed by the measurement resistor in parallel with the shunt capacitance, which includes the HEMT gate to source-drain channel capacitance C_g , stray capacitance of the tip and its connecting leads C_{tip} and measurement resistor self-capacitance C_m (see Fig. 6).

We used an amorphous granular film resistor [23] with $R_m=18\text{ M}\Omega$ at cryogenic temperatures and $C_m=0.3\text{ pF}$, and an uncased HEMT [22] with $C_g=0.3\text{ pF}$. By placing the circuit in close proximity to the tip we were able to keep the shunt capacitance to about 1 pF . This value has been estimated by fitting the calculated frequency dependence of the circuit gain, using a circuit schematic Figure 6 to the measured at $T=4.2\text{ K}$ values, as shown in Figure 7. A good overall fit between the calculation and the measurement supports the model of the circuit suggested

by Figure 6. Due to the low input capacitance of the circuit, the resulting roll-off frequency (3 dB point) is quite high at $f_{3dB} = 1/2\pi C_{shunt} R_m \approx 8 kHz$.

The optimal frequency of the ac excitation in lock-in based spectroscopic measurements is determined mostly by the balance between the $1/f$ amplifier noise increasing at low frequencies and the reduced gain at high frequencies due to the signal low-pass filtering by C_{shunt} . Our measurements show that the $1/f$ noise edge of the circuit is below 1 kHz, resulting in a flat noise level in a frequency range of 1–100 kHz as plotted in Fig. 8. The HEMT is operated in the linear regime at a drain-source resistance that matches the impedance of the high capacitance cable (C_w 500 pF) that connects its output to the room temperature electronics, resulting in transistor voltage gain close to unity. Therefore it can be thought of as an impedance matcher from the high output impedance of the signal source (measurement resistor) to the low impedance of the load. The noise amplitude referred to the input current measured at 4.2 K is about $4 fA/\sqrt{Hz}$ in the frequency range of 3 kHz–40 kHz. This level is consistent with the level of Johnson noise on the measurement resistor of $3.7 fA/\sqrt{Hz}$ and it is expected that at 0.3 K operating temperature the noise level will reach a value of about $1 fA/\sqrt{Hz}$ [24].

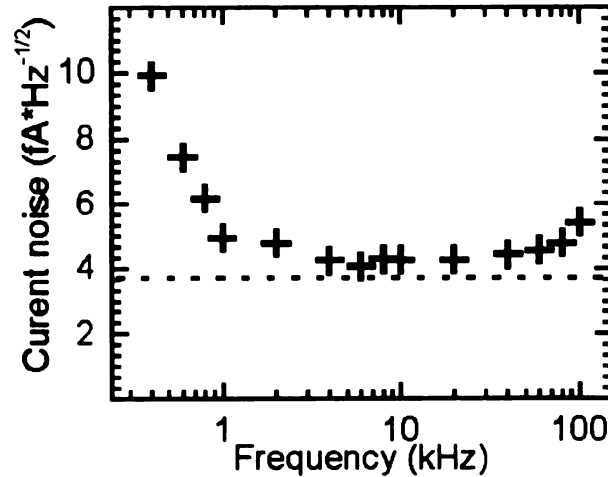


Figure 8: Measured noise level (crosses). The Dashed line represents the Johnson noise amplitude of the measurement resistor at 4.2 K.

To provide for effective STM feedback loop operation, the ac amplifier circuit elements should not result in low-pass filtering of the current going to the dc amplifier. As can be seen from Fig. 6, such a low pass filter is formed due to the capacitance $C_{tip} + C_g$ drawing part of the signal current. Therefore, we should require that the impedance due to the shunt capacitance in the feedback loop bandwidth should be greater than the input impedance of the dc current amplifier circuit including bias resistor

$$\frac{1}{C_{shunt}} \gg \frac{1}{C_w + \frac{1}{2\pi f_{FB} R_{in}}} + \frac{1}{C_m + \frac{1}{2\pi f_{FB} R_m}} \quad (2.26)$$

which is well satisfied for the circuit parameters given in Fig. 5 and typical feedback loop bandwidth f_{FB} of 2 kHz. Eq. 2.26 shows the importance of minimizing the input capacitance of the amplifier. If the circuit is placed just several centimeters away from the tip, this results in an effective bandwidth of the feedback loop of less than 1 kHz, compromising the STM topographic measurements.

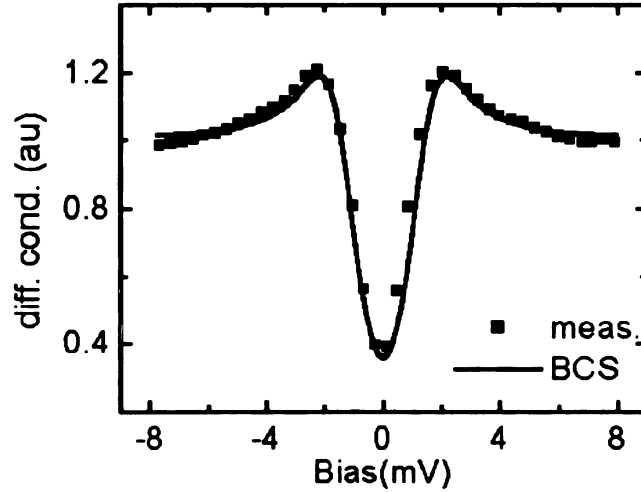


Figure 9: dI/dV spectrum of a 200nm thick superconducting Pb film at $T=4.2$ K. Squares: measurement, solid line: a fit as explained in the text.

An example of the operation of the circuit is demonstrated in Fig. 9. A differential conductance spectrum of a superconducting Pb film was measured at $T=4.2$ K. The

lock-in detection technique was used with ac bias modulation amplitude of 0.4 mV rms. The acquisition rate was 100 ms/point (lock-in time constant of 30 ms), and the tunnel resistance R_t of 0.4 G Ω . The fit was performed based on BCS theory for the nominal Pb gap value Δ of 1.35 meV. Besides the temperature smearing, an additional energy smearing due to the ac excitation voltage is accounted for by convolution with a Lorentzian. The spectrum shows good agreement with the BCS curve and a low noise level consistent with the results of the circuit noise measurements. The parameters of the circuit are optimized for compatibility with most LT STM systems. The analysis shows that the value of the measurement resistor determines the performance of the circuit. Therefore, if a narrow current feedback loop bandwidth is acceptable, and high tunnel resistances are used, the noise level referred to the input current can be easily improved by increasing the measurement resistor value. For example, increasing R_m to 100 M Ω will result in noise level of about 2 fA/ $\sqrt{\text{Hz}}$ at T=4.2 K.

Chapter 3

STM of Surface and Defect states in Bi_2Se_3

Bi_2Te_3 and Bi_2Se_3 are narrow gap semiconductors with gap values of about 0.15 eV and 0.3 eV respectively [25, 26]. These compounds are structurally identical and are similar in many properties. However, the electronic structure of Bi_2Se_3 is simpler than Bi_2Te_3 (as discussed in detail below), making it a convenient model system for the studies of the family of narrow gap chalcogenide semiconductors. On the other hand, Bi_2Te_3 and its alloys with Sb_2Te_3 are the best room temperature thermoelectric materials known for over 40 years. Surprisingly, the origin of the excellent thermoelectric properties is not clear to date. In the following chapters it will be demonstrated how the STM technique applied to Bi_2Se_3 and Bi_2Te_3 provides new insight in the electronic properties of these materials. We will also link the observations to the thermoelectric performance.

3.1 Bi_2Se_3/Bi_2Te_3 Structure

The crystal structure of Bi_2Se_3 (and Bi_2Te_3) is rhombohedral with space group $D_{3d}^5(R\bar{3}m)$. It can be represented as a stack of hexagonally arranged (see bottom of

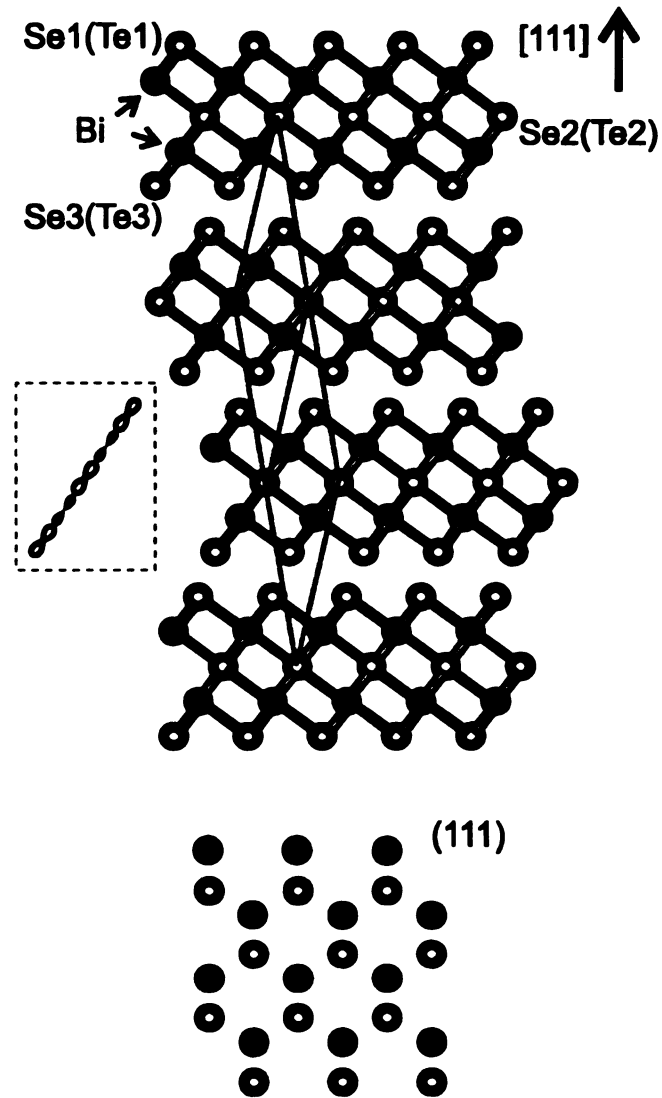


Figure 10: Bi_2Se_3 (Bi_2Te_3) structure. Top: View parallel to the layers. The rhombohedral unit cell is shown. In a dashed frame the chains of the $pp\sigma$ -bonds are plotted as discussed later. Bottom: A top view of the rhombohedral $(1,1,1)$ plane. Open circles represent $Se1(Te1)$ atomic positions, solid circles—subsurface Bi positions.

Figure 10) atomic planes, each consisting of only one atomic type, as shown in Figure 10. Five atomic planes are stacked in a close-packed *fcc* fashion in order Se1(Te1)-Bi-Se2(Te2)-Bi-Se1(Te1), in a five-plane layer sometimes called quintuple-layer leaf. Proximity of the surface breaks the equivalence of the Se1(Te1) and Se3(Te3) positions. We use Se1(Te1) for the surface Se(Te) positions, and Se3(Te3) for the 5th from the surface atomic plane positions. The rhombohedral unit cell spans three layers and contains 5 atoms. Since the unit cell is needle-like, pointing across the layers in [1,1,1] direction, the Brillouin zone, shown in Figure 11 is pancake-like, squeezed in this direction. The lattice parameters are given in Table 3.1. All the atoms in the rhombohedral cell lie along the [1,1,1] direction. The positions of the atoms are given in units of the rhombohedral lattice constant a . $a = 9.84\text{\AA}$ for Bi_2Se_3 and 10.45\AA for Bi_2Te_3 . The rhombohedral angle α is $24^\circ 8'$ for both compounds. It is interesting to note that Bi_2Se_3 and Bi_2Te_3 have identical (within crystallographic precision) lattice parameters scaled by the unit cell size. The likely explanation for this surprising fact lies in the nature of the bonding in the structure, with the ionicity of the atoms largely governing the difference in the interatomic distances in the two compounds. The ionicity of atoms in Bi_2Se_3 is higher than in Bi_2Te_3 , leading to a more closely packed structure. However, it is not clear why the *interlayer* distances also scale, since (as will be discussed later) they are governed by a more complicated bonding mechanism.

Atom	Site index	Position
Bi	2c	0.399
Se1(Te1)	2c	0.792
Se2(Te2)	1a	0.000

Table 3.1: Lattice parameters for Bi_2Se_3 and Bi_2Te_3 from Ref. [27].

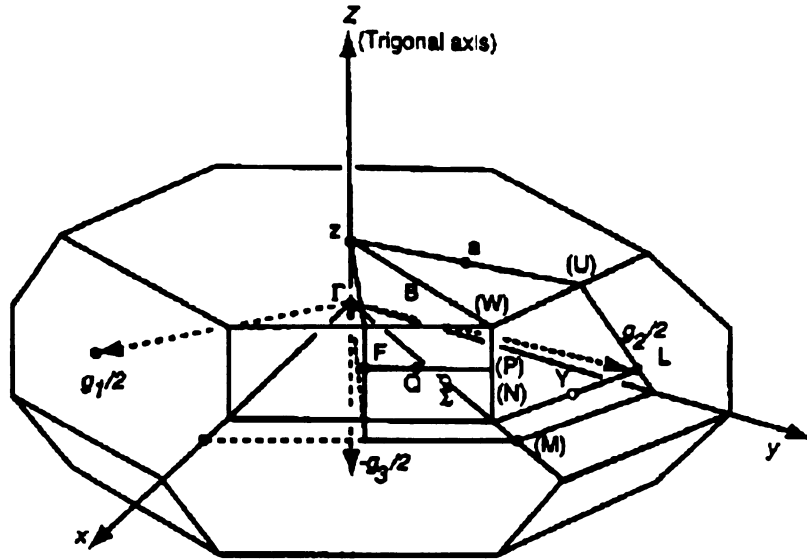


Figure 11: Rhombohedral Brillouin zone of Bi_2Se_3 and Bi_2Te_3 . Γ is the center of the zone, Z is $[1,1,1]$ zone edge, M is a $[1,1,-2]$ zone-edge point.

3.2 Surface Preparation

The stoichiometric and substitutionally doped Bi_2Se_3 and Bi_2Te_3 single crystal samples were grown by a directional solidification technique [28]. The layered compounds Bi_2Se_3 and Bi_2Te_3 can be easily cleaved with a scotch tape along the $(1,1,1)$ surface, similarly to graphite. The surface is sufficiently inert to obtain atomic resolution in air immediately after cleaving. However, atomic resolution does *not* imply that there is not a layer of adsorbates on the surface. We were able to obtain more reproducible results by cleaving the samples in a glove box directly attached to the STM setup or *in situ* in the UHV chamber. Cleaving inside the chamber was achieved by gluing a piece of thick aluminum wire perpendicular to the surface, and then knocking it off in vacuum using a magnetically coupled manipulator, as shown in Figure 4. Using a long manipulator stick the sample was subsequently transferred under vacuum through a gate-valve to the STM. After the sample transfer, the gate between the STM and the UHV chamber was closed, and the gate to the He-4 storage dewar with liquid He-4 was opened, letting the He-4 gas into the chamber with the STM.

He-4 gas produced by liquid He-4 boil-off in the dewar is sufficiently clean and inert, and in the direct immersion setup contaminants are efficiently cryo-pumped by the large cold surface of the dewar and He-4 liquid. The vacuum achievable in the STM setup with the gate to the He-4 dewar closed is about 10^{-4} mbar. It proved to be sufficiently good for studying relatively inert Bi_2Se_3 and Bi_2Te_3 surfaces. A similar setup was also used for a study of gold nanoparticles on the surface of superconductor NbSe₂ [29], and atomic resolution on NbSe₂ was also attained. After the transfer, the microscope together with the sample were gradually cooled over a period of 3-5 hours to the temperature of 4.2 K. Such slow cooling minimizes the thermal stress on the piezoelements to avoid their physical damage.

3.3 Atomic Resolution STM Imaging of Bi_2Se_3

Both chemically etched and mechanically cut tips have been used for the experiments, with similar success. Typically less than half of the tips are metallic (as seen from the tunneling measurements) and sufficiently stable for STM studies. The best results have been obtained by softly crashing the tips into a clean gold surface prior to using them for STM analysis. An alternative tip preparation technique, not accessible in our setup, is field emission. By biasing the sample surface with a sufficiently high positive voltage of about 100 V, emission of atoms from the end of the tip is initiated. This procedure is repeated until a stable tip is produced.

Due to the difficulty of controlling the shape, stability, and electronic properties of the STM tip, reproducibility and common sense are used to establish that the STM results are representative of the sample properties and not dominated by the tip. In particular, the STM topographic images can be thought of as a convolution of the surface and tip geometry. Figure 12 illustrates such an effect of the STM tip geometry on the topographic images. It shows a series of STM images of Bi_2Se_3 surface, acquired in a single sequence under identical experimental conditions over the same

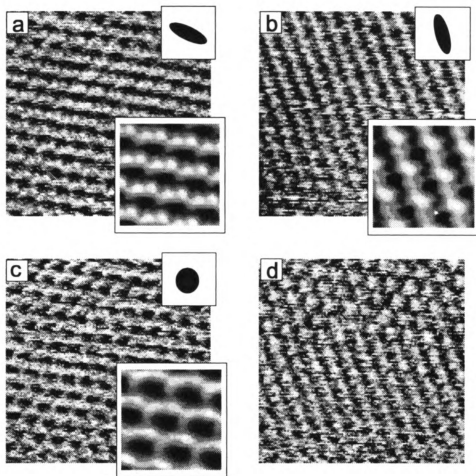


Figure 12: 6.2x6.2 nm anomalous STM topographical images of Bi_2Se_3 surface. All the images are acquired in a single sequence under identical conditions over the same area of the surface, with bias voltage $V = -300$ mV, tunneling current $I = 0.75$ nA. Bottom insets show enhanced autocorrelated images. Top insets show schematically the tip geometry deduced from each image.

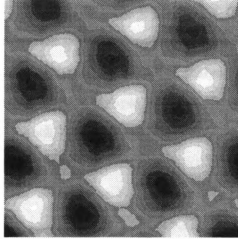


Figure 13: 1.4x1.2 nm STM topographical images of Bi_2Se_3 surface, acquired at -0.7 V bias voltage, and 200 pA tunneling current. The black to white scale corresponds to 30 pm. Autocorrelation was used to enhance the image.

sample area. The images do not possess the hexagonal symmetry of the Bi_2Se_3 (1,1,1) surface, and they significantly vary as the consecutive scans are performed. In image (a), atomic corrugations appear as horizontal stripes; in (b) corrugations appear as vertical stripes; (c) nearly reproduces the hexagonal symmetry of the surface. In image (d) the perceived symmetry of the atomic corrugations changes in the process of image acquisition, so the top of the image has hexagonal symmetry, while the bottom part is similar to (b). We explain all these effects by the marginal stability of the tip used for the acquisition of images 12, resulting in the variation of the tip shape between images (a)-(c) and during the acquisition of image (d). Such a variation of the tip shape possibly resulted from the interaction of the sample surface with the atoms weakly attached to the end of the tip. The various shapes of the the tip end are shown in Figure 12 schematically in the top insets. These shapes were deduced from the shapes of the atomic corrugations of the topographic images.

The *ab initio* modeling of STM topographic images [49] of Bi_2Se_3 indicates that both Se atoms from the surface atomic plane and Bi atoms from subsurface atomic plane can be seen under appropriate bias conditions. In particular, the empty elec-

tronic states are Bi-dominated, so the STM topographic image acquired with a positive sample bias exhibits atomic corrugations shifted away from the surface Se positions, closer to the subsurface Bi positions. Filled states are Se-dominated, and the atomic corrugations seen at negative sample bias reflect Se positions, so the images are shifted with respect to each other. It will be shown below that at bias voltages of about -0.5 to -0.3 V, tunneling occurs from the Bi-dominated surface states of Bi_2Se_3 . Therefore, at moderate negative bias voltage the atomic corrugations due to both surface Se and subsurface Bi can be seen, as shown in Figure 13.

3.4 STM of Topographic Defects in Bi_2Se_3

The defect types to consider include [30] substitutional defects (one atom type occupying the site of another type), interstitial (Frenkel) defects (atom placed between the sites of a perfect crystal), and vacancies. The electronic properties of defects depend sensitively on their density and particular configuration [32]. As illustrated in Figure 14, the (1,1,1) surface topography of a nominally stoichiometric Bi_2Se_3 sample exhibits variations due to local defects. The scale in the images corresponds to about 100 pm topographic variation, with atomic corrugations only about 20 pm high. The electronic properties in the vicinity of the topographic defects (as tested with STS techniques described in Section 2.3) do not show any significant deviation from the average properties. The shapes of the protrusions (Figure 14) are triangular, correlated with the symmetry of the lattice, and all the protrusions have similar size. The likely origin of these protrusions is interlayer interstitial Bi defects between the surface and second from the surface layer. Below the arguments in favor of this hypothesis are listed. These arguments are highly suggestive and need to be supported by model calculations.

- The low solubility of Bi in Bi_2Se_3 [25] is likely due to the high energy of the Bi_{Se} defect (here we use A_B notation for a substitutional A-type defect on site

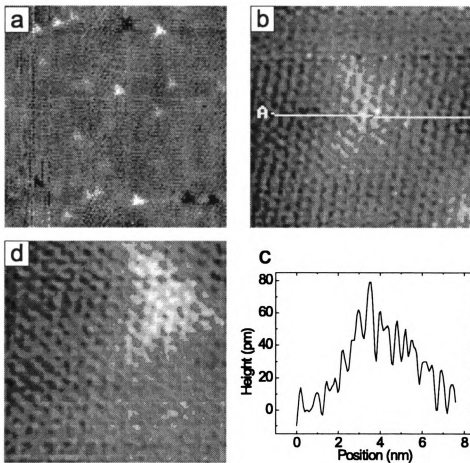


Figure 14: (a), (b), (d) Topographic images of defects on the surface of nominally stoichiometric Bi_2Se_3 . The parameters are: (a) area 71x63 nm, $V=-50$ mV, $I=20$ pA, (b) area 8.5x7.5 nm, $V=-400$ mV, $I=100$ pA, (d) area 7.2x6.3 nm, $V=-600$ mV, $I=100$ pA. (c) Cross-section of image (b).

B). On the other hand, interstitial Bi defects between the layers should have lower energy by a simple electrostatic argument.

- The nominally stoichiometric Bi_2Se_3 samples are n-type with carrier concentration of about 10^{19} cm^{-3} due to the statistical balance of defects in as grown samples. If Bi atoms form high density interstitial defects between the layers, then the excess Se form Se_{Bi} defects. Both Bi and Se have a partly filled p-band, $6p$ with 3 electrons and $4p$ with 4 electrons respectively. Therefore, Se_{Bi} can be represented as a defect with a single extra nuclear charge and an electron bound to it, i.e. a donor.
- It is well established [27] that both valence and conduction bands are formed almost exclusively by the $4p$ and $6p$ orbitals of Se and Bi respectively. For each atom, except for Se1 (see Figure 10), the nearest neighbors from the adjacent atomic planes form (to the precision of a few degrees) almost a regular octahedron, so the bonding can be roughly approximated by strongly $pp\sigma$ interacting chains of atomic p -orbitals with a weaker $pp\pi$ -type interaction between adjacent chains (Figure 15(a)). These bonds create the directions of the highest stiffness of the material, so an interstitial defect between the layers creates a protrusion that propagates to the surface in a triangular shape, as shown in Figure 15(b). This explains why all the protrusions in Figure 14 are similar in size. Here we assume that the topographic variations originating at interlayer defects between the deeper layers are sufficiently relaxed at the surface to be unobservable with STM.
- By simple counting of defects in Figure 14(a) we can confirm that the defect density is consistent with the natural doping level of the crystal, if one accepts that the protrusions come only from the interstitial defects between the first and the second surface layer.

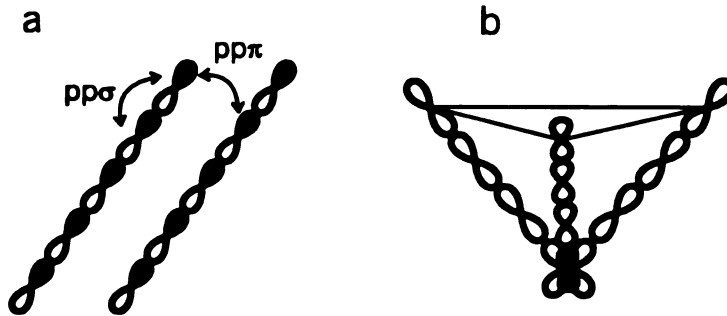


Figure 15: (a) $pp\sigma$ bonding chains in Bi_2Se_3 , with $pp\pi$ -type interaction between chains. (b) Origin of the topographic protrusions shown in Figure 14 from $pp\sigma$ chains.

- The topographic image of Figure 13 was acquired simultaneously with the images of Figure 14. While the surface Se atomic plane has 6-fold symmetry, the bulk structure has only 3-fold symmetry, reflected both in the shape of the protrusions (Figure 14) and the high-resolution topographic images, Figure 13. The highest-strength $pp\sigma$ bonding picture we invoke to explain the triangular shapes of the protrusions is consistent with the crystal symmetry inferred from Figure 13.

3.5 STS Studies of Bi_2Se_3

The established bulk semiconducting gap value in Bi_2Se_3 is about 0.3 eV [27, 25], both from *ab initio* calculations and photoabsorption edge measurements. We expect to see negligible differential conductance at energies between the conduction band minimum (close to Fermi level in the n-type samples, corresponding to bias voltage $V=0.0$ V) and -0.3 eV (or -0.3 V sample bias).

In Figure 16 the typical low-temperature dI/dV spectra of the stoichiometric Bi_2Se_3 surface are presented. These spectra were acquired directly using a lock-in detection technique with a low-temperature current amplifier described in Section 2.4.

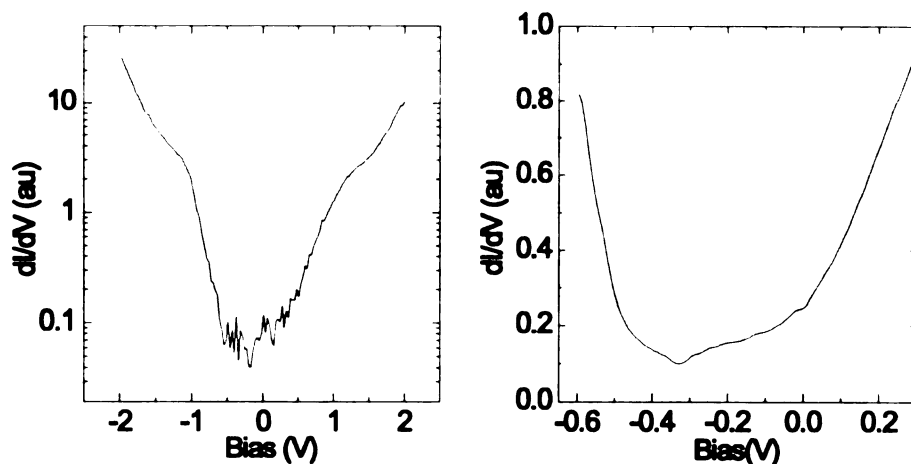


Figure 16: Typical low-temperature dI/dV spectra of the stoichiometric Bi_2Se_3 surface. Left: large energy scale spectrum. A logarithmic vertical scale is used. Right: near-gap conductance acquired at a smaller tip-sample separation.

A highly sensitive measurement over the full bias voltage sweep range could not be achieved. Therefore, two measurements were performed: large energy scale spectrum (shown on the left), and a near-gap spectrum acquired at a smaller tip-sample separation and over a smaller bias range (right). The conductance at bias voltages used in the small bias scale plot is dominated by noise in the large scale plot.

It can be clearly seen from plots Figure 16 that the differential conductance spectra are indeed suppressed between -0.3 V and 0.0 V. However, the conductance remains finite at these energies. These characteristics were highly reproducible and independent of the position. This lack of spatial variation excludes the explanation of in-gap conductance by the impurity states. At $T=4.2$ K the temperature smearing is about 1 mV, negligible on the energy scale of spectra Figure 16. Therefore, in-gap finite conductance also cannot be attributed to the temperature effects. We have put forward a hypothesis that the effect is likely due to the presence of surface states [31]. It has been confirmed by model *ab initio* calculations which will be described below.

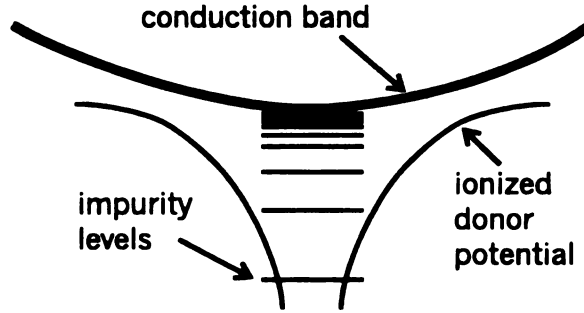


Figure 17: Hydrogenic impurity levels created by an ionized shallow donor impurity.

The surface effects are also responsible for the suppressed conductance just below the gap, roughly between -0.3 V and -0.5 V . Similar surface effects are also observed in the differential conductance spectra of Bi_2Te_3 which will be discussed below. For the analysis of the experimental data we will use a simple one-dimensional tight bonding model to show that such surface effects are indeed quite general.

3.6 Defect States in Semiconductors

In elemental semiconductors (for example, group IV: diamond, Si or Ge) substitutional defects from the neighboring groups (III or V) tend to form shallow impurity levels (acceptor and donor respectively). A simple qualitative picture of a P donor in Si is that of four covalent sp^3 -bonds connecting each atom with its neighbors and delocalizing the valence electrons. A P atom in Si has an extra nuclear charge and an extra electron, which can be easily broken off the atom and delocalized into the conduction band. The remaining positive nuclear charge is screened over a distance of about 100\AA by the dielectric response of the crystal. The ionized donor charge creates a system of hydrogen-like levels (Figure 17), modified by the screening effect. Consider an electron moving in a screened donor potential. The Hamiltonian describing the electron dynamics is [32]

$$H = H_0 + U(\mathbf{r}) = -\frac{\hbar^2}{2m_0}\nabla^2 + V(\mathbf{r}) - \frac{e^2}{\kappa r} \quad (3.1)$$

where m_0 is a free electron mass, $V(\mathbf{r})$ is the periodic crystal potential, e is electron charge, r is the distance from the impurity center, κ is the lattice dielectric constant. Bloch functions are solutions of the Shrödinger equation without the impurity potential $U(\mathbf{r})$

$$\phi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{V_0^{1/2}} u_{n,\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k}\mathbf{r}), \quad (3.2)$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ is a periodic function with lattice period, V_0 is the crystal volume, \mathbf{k} is a wave vector, n numbers the bands and spins. Here we assume that the band extremum is non-degenerate and (for simplicity) is at the center of the Brillouin zone, so that the effective mass approximation can be used for the band dispersion around $k = 0$

$$E_n(k) = \frac{\hbar^2 k^2}{2m}, \quad (3.3)$$

where m is effective mass generally not equal to the free electron mass m_0 . It is convenient to look for the eigenstates of Hamiltonian (3.1) in terms of eigenstates (3.2) of the unperturbed Hamiltonian

$$\psi = \sum_{n',\mathbf{k}'} B_{n'}(\mathbf{k}') \phi_{n',\mathbf{k}'}(\mathbf{r}). \quad (3.4)$$

Substitute (3.4) into Shrödinger equation $(H_0 + U - E)\psi = 0$ and find a projection on $\phi_{n,\mathbf{k}}(\mathbf{r})$

$$(E_n(\mathbf{k}) - E) B_n(\mathbf{k}) + \sum_{n',\mathbf{k}'} U_{n',\mathbf{k}'}^{n,\mathbf{k}} B_{n'}(\mathbf{k}') \quad (3.5)$$

where

$$U_{n',\mathbf{k}'}^{n,\mathbf{k}} = (\phi_{n,\mathbf{k}}, U \phi_{n',\mathbf{k}'}) = \frac{1}{V_0} \int u_{n,\mathbf{k}}^* u_{n',\mathbf{k}'} \exp[i(\mathbf{k}' - \mathbf{k})\mathbf{r}] U(\mathbf{r}) d\mathbf{r} \quad (3.6)$$

The characteristic size of the electron localization around the impurity is typically much larger than the lattice spacing, therefore coefficients $B_{n'}(\mathbf{k}')$ in the expansion (3.4) are large only for small \mathbf{k}' . In (3.6) we make an approximation $u_{n,\mathbf{k}} \approx u_{n,0}$. Then in the integral (3.6) we can separate the product of two functions dependant on \mathbf{r} : the first is u^*u , rapidly changing in space with periodicity of the lattice, the

second is $U(r)\exp[(\mathbf{k} - \mathbf{k}')\mathbf{r}]$, relatively slowly changing in space. We can write 3.6 approximately as

$$U_{n',\mathbf{k}}^{n,\mathbf{k}} \approx \int_{\text{cell}} \frac{1}{V_0} u_{n,0}^* u_{n',0} d\mathbf{r} \int \exp[i(\mathbf{k}' - \mathbf{k})\mathbf{r}] U(r) d\mathbf{r} \quad (3.7)$$

where in the first integral we integrate over one unit cell. Due to orthonormality of the Bloch functions

$$\int_{\text{cell}} \frac{1}{V_0} u_{n,0}^* u_{n',0} d\mathbf{r} = \delta_{nn'}. \quad (3.8)$$

We get a closed system of equations for unknown $B_n(\mathbf{k})$

$$(E_n(\mathbf{k}) - E)B_n(\mathbf{k}) + \sum_{\mathbf{k}'} U(\mathbf{k}, \mathbf{k}') B_{n'}(\mathbf{k}') = 0 \quad (3.9)$$

where

$$U(\mathbf{k}, \mathbf{k}') = \frac{1}{V_0} \int \exp[i(\mathbf{k}' - \mathbf{k})\mathbf{r}] U(r) d\mathbf{r} = -\frac{4\pi e^2}{\kappa V_0 (\mathbf{k} - \mathbf{k}')^2} \quad (3.10)$$

Plugging expression (3.10) and (3.3) into (3.9), we obtain

$$\left(\frac{\hbar^2 k^2}{2m} - E \right) B_n(\mathbf{k}) - \frac{4\pi e^2}{\kappa V_0} \sum_{\mathbf{k}'} \frac{1}{(k - k')^2} B_n(\mathbf{k}') = 0. \quad (3.11)$$

Using the fact that $B(\mathbf{k})$ decays very rapidly away from the center of the Brillouin zone, we can extend the summation over \mathbf{k}' to infinity and change to the spatial representation, introducing

$$F(\mathbf{r}) = \frac{1}{V_0^{1/2}} \sum_{\mathbf{k}} B_n(\mathbf{k}) \exp(i\mathbf{k}\mathbf{r}). \quad (3.12)$$

Multiplying (3.11) by $\exp(i\mathbf{k}\mathbf{r})$ and summing over \mathbf{k} , we get

$$\left(-\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{\kappa r} \right) F(\mathbf{r}) = E F(\mathbf{r}). \quad (3.13)$$

This equation is identical to the Shrödinger equation for the hydrogen atom, except that the electron mass is replaced by the effective mass of the conduction band, and the dielectric constant is different from its vacuum value. The usual expression for the eigenenergies of equation (3.13) is

$$E(t) = -\frac{1}{t^2} \frac{e^4 m}{2\epsilon^2 \hbar^2}, \quad t = 1, 2, \dots \quad (3.14)$$

and the ground state ($t = 1$) wave function is

$$\psi_0 = \frac{1}{(\pi a_{eff}^3)^{1/2}} u_{c,0}(\mathbf{r}) \exp\left(-\frac{r}{a_{eff}}\right), \quad (3.15)$$

where $a_{eff} = \hbar^2 \epsilon / m e^2$ is the effective Bohr radius.

The effective radius of the wave function is typically much larger than the lattice constant due to the large value of dielectric constant and small value of effective mass. In GaAs, for example, $\kappa = 12.6$, $m = 0.066m_0$, $a_{eff} \approx 100\text{\AA}$, so the inverse decay length a_{eff}^{-1} is only a small fraction of the reciprocal lattice vector. Solving (3.12) for B , we get

$$B_n(k) = \frac{8\pi^{1/2}}{V_0^{1/2} a^{5/2}} \frac{1}{(k^2 + a^{-2})^2} \quad (3.16)$$

Consistent with the assumptions we made in deriving (3.14) and (3.15), $B(\mathbf{k})$ decays rapidly away from the center of the Brillouin zone.

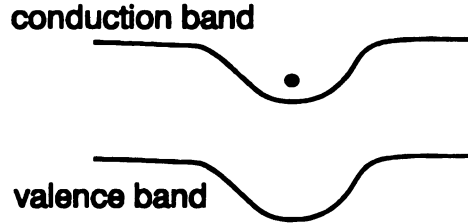


Figure 18: Semiclassical picture of the ionized donor effect on the local electronic properties.

A connection can be made to the semiclassical (Thomas-Fermi) picture of an ionized impurity, for example a donor state. The charge locally bends the conduction and valence bands, as shown in Figure 18, so the local conduction-like states are created below the bulk conduction band minimum. This is a direct semiclassical analogy of the quantum picture of impurity states described above. The bending of

the valence band semiclassically approximates the impurity screening by the electronic system [30].

3.7 STM Observations of Impurity States in Semiconductors

In spite of a long-established theory of impurity states in semiconductors [32], the physics of the STM observation of impurity states in semiconductors is not well established yet. Here the current state of the field is briefly reviewed. Modeling of the tunneling between the STM tip and sample surface shows that in a typical STM experiment the small tunneling current of about 1 nA is carried through a small area of the sample surface of about 0.2 nm^2 , i.e. the current density is extremely high, more than $5 \times 10^5 \text{ A/cm}^2$. If a single impurity state is to be observed, it has to be able to sustain such a high tunneling current. In particular, in derivation of expression 2.10 equilibrium charge distributions in the electrodes were assumed, which holds only if the carrier relaxation rates in the electrodes are sufficiently higher than the rate of electron tunneling between the electrodes. For example, although a localized state might have a large amplitude at the tip position, yielding a high tunneling current (according to Equation 2.10), in reality such a localized state does not contribute to the STM tunneling current because a single tunneling event saturates this state, and after that the charge decays to its equilibrium value over the long relaxation time, after which another tunneling event can occur. For the tunneling current $I = 1 \text{ nA}$ the electron tunneling rate is $\tau = \frac{I}{e_0} \approx 10^{10} \text{ sec}^{-1}$. Therefore, the charge relaxation time should be at most in the picosecond range in order not to inhibit the tunneling current through the defect state. Special relaxation mechanisms have to be invoked to provide such short relaxation times: hopping conductivity through the defect band [33], tunneling into the bulk bands mediated by the surface states, and relaxation through recombination processes — phonon emission, Auger recombination

and radiative processes have been suggested [34]. The particular mechanisms and their applicability to the specific experimental conditions will not be discussed here, since the theoretical background for such considerations does not seem to be well established.

Experimental STM data on defect states in semiconductors are available predominantly for GaAs [35]–[43], with most of the studies concentrating on the topographic aspects of impurities, and only a few [35, 36, 43, 44] are concerned with the spectral properties. The defect electronic states appear as very broad (about 0.5 eV wide) resonances deep inside the semiconducting gap. The width of the defect states observed in STM experiments is significantly larger than the room temperature smearing. It can not be explained by the broadening effects of the high impurity concentrations, resulting in the formation of the impurity band. Such an explanation would imply unreasonably high defect densities, comparable to the densities of the host material. It has been well established that the dominant naturally occurring defect in GaAs is As_{Ga} , in the bulk forming a deep defect state in the gap [45], which has been identified with a so called *EL2* defect state. However, the positions of the defect states deep inside the semiconducting gap in all the experiments on shallow dopants are not consistent with the established bulk shallow impurity energy levels [32]. It is possible that regardless of doping only the *EL2* deep defect states were observed.

There is another explanation for the experimental data. Surface effects in GaAs have been shown [46, 47] to be important for the formation of the impurity states originating from the surface defects. It is possible that the electronic states observed in Ref. [35]–[43] were in fact dominated by the surface effects. This hypothesis will be discussed in detail below in conjunction with the analysis of the impurity states in Bi_2Se_3 . Because of the high sensitivity of the defect states to the position with respect to the cleavage plane it is important to know the *exact* position of the impurity atom. Unfortunately, in the experiments Ref. [35]–[43] the atomic impurity types under study could be inferred only from indirect observations, and the exact position with

respect to the surface could not be determined. This imposes considerable limitations on the theoretical understanding of the experiments.

From this point of view Bi_2Se_3 and similar compounds can be considered as model systems for understanding of the properties of the near-surface impurity states: As described in Section 3.4, the bonding in Bi_2Se_3 is dominated by the strongly $pp\sigma$ -bonded chains, with three almost orthogonal chains passing through each atom [27]. The $pp\pi$ -interaction between the chains is about 4-5 times [52] weaker (per bond), still non-negligible for even a rough quantitative analysis. Qualitatively, however, such a bonding scheme indicates that it might be possible to directly determine the position of a substitutional impurity with respect to the surface: a substitutional defect is likely to produce a perturbation in the electronic local density of states (DOS) predominantly along the three $pp\sigma$ chains passing through the defect atom. If the defect atom is near the surface, it would result in three spots of modified DOS around the atoms terminating these chains at the surface, directly relating the relative positions of the spots to the position of the impurity atom under the surface. This idea provides a nice interpretation for the defect state observations in Bi_2Se_3 presented below.

3.8 Impurity States in Bi-Doped Bi_2Se_3

The topographic defects in nominally stoichiometric Bi_2Se_3 have been discussed in Section 3.4. STS measurements do not show any significant variation of the local electronic properties in the vicinity of such defects. On the other hand, a number of facts about the Bi_2Se_3 enable one to infer the nature of these defects. Another more controllable way to study defects is by intentional doping. By simply doping Bi_2Se_3 with excess Bi or Se, antisite defects (acceptor and donor type, respectively) may be introduced. Because of the low solubility of Bi in Bi_2Se_3 [25], doping the samples with excess Bi does not result in significant carrier concentration changes, probably

due to the formation of interstitial (interlayer) Bi defects and/or Se vacancies. This effect is governed by the statistical distribution of the defects in semiconductors, driven by the differences in the defect formation energies [30]. Our initial interest in studying Bi_2Se_3 samples doped with excess Bi was stimulated by such an unusual doping behavior. As it follows from the discussion below, the separate issues of the interplay of the surface properties and defect states can be analyzed based on the STM studies of Bi antisite defects in Bi_2Se_3 . There are 3 electrons in the Bi valence p -shell, and 4 electrons in the Se p -shell. In the bulk, Bi_{Se} antisite defects likely form shallow acceptor levels inside the semiconducting gap, according to the theory developed in Section 3.6. Our study shows that near-surface Bi_{Se} defects can be seen with STM as resonant levels *inside* the valence band due to the effect of the surface.

We have studied heavily doped $\text{Bi}_2\text{Se}_{2.85}$ and $\text{Bi}_2\text{Se}_{2.95}$ samples. The differential CITS maps of the $\text{Bi}_2\text{Se}_{2.85}$ surface acquired at a number of bias voltages are presented in Figure 19. The surface (topographic image 19(a)) is atomically flat, with variations in the topography mostly due to the adsorbates on the surface. However, there seems to be a certain degree of correlation between the enhanced conductance in CITS images Figure 19 (b)–(f) and the topographic imperfections. These are likely due to correlations between the local lattice/surface defects and minor variations in the electronic structure. Striking clover-shaped regular features of similar size (about 2.5 nm across) appear in the images (d), (e), at CITS acquisition bias around -400 to -450 mV. The features correspond to locally enhanced conductance at these bias voltages, i.e. local resonances. A more precise spectroscopic analysis is given below. The orientation of the clover-shaped features follows the three-fold symmetry of the (1,1,1) surface. These features are not observed in nominally stoichiometric Bi_2Se_3 samples, therefore they can be attributed to Bi_{Se} antisite defect states.

The data discussed below were obtained on a more weakly doped $\text{Bi}_2\text{Se}_{2.95}$ sample, which exhibits similar features with a reduced density. The data were acquired with a redesigned more stable microscope. As is clear from Figure 19, the impurity

states are spread at energies around -0.4 to -0.45 eV. Figure 20 shows a series of differential conductance spectra in the vicinity of a clover-shaped defect. The spectra were obtained by numerical differentiation of 60 I-V curves, with setpoint parameters: sample bias voltage $V=-0.3$ V, and current $I=0.8$ nA. At the measurement temperature of 4.2K, thermal broadening is negligible on the displayed bias voltage scale, so the spectra represent the local surface density of states. Several important features of these spectra can be contrasted to the simple hydrogenic picture of the impurity states discussed in Section 3.6:

- The impurity states are highly spatially anisotropic.
- The impurity state appears in the differential conductance spectra as a broad resonance, as opposed to a sharp almost localized state only slightly widened by the thermal smearing and relaxation processes.
- The resonance energy lies inside the valence band, in the energy range where differential conductance is suppressed away from the impurity. In the conventional picture given in Section 3.6 the impurity states appear in the gap.
- The resonance amplitude is largest 1.2 nm away from the impurity state center along the clover lobes.
- It is clear that depending on the position of the impurity atom with respect to the surface the three $pp\sigma$ chains creating clover-like impurity shapes will terminate at different spacing, creating different sizes of the "clovers". However, all the observed clover-shaped defect states have the same size.

The listed features suggest that the impurity state is, as expected, dominated by the contribution of $pp\sigma$ -bonded chains. The three spots with enhanced conductivity around the atoms terminating the chains at the surface create the clover-like appearance of the defect state at the surface. All the clovers have the same size, therefore the resonance originates only from a certain atomic plane below the surface. This

fact together with the energies at which the resonant states appear are analysed in detail below.

To clarify the positions of the originating defect atoms, in Figure 21 we zoom in on one of the clover-shaped features. While the topography image Fig. 21(b) acquired at a bias voltage of -0.3 V shows a periodic atomic structure, the CITS image Fig. 21(b), acquired simultaneously at -0.7 V, has strong local conductance variations due to the impurity state, also with discernible atomic corrugations. It is clear from the cross-section plot Fig. 21(c), that the atomic corrugations in the two images are offset with respect to each other. This offset persists away from the defects, and reflects the strong contribution of subsurface Bi atoms to the in-gap states (between around -0.3 V and 0 V), while the valence band states (below -0.3 V) are dominated by the surface Se. As a result, the topographic image acquired at -0.3 V does not reflect the surface Se atomic positions, but is shifted towards the positions of subsurface Bi atoms. On the other hand, the current image acquired at -0.7 V is dominated by the contributions of the surface Se atoms, resulting in the observed offset between the two images. This picture is supported by the first principles modeling of the experiments [49].

As the topographic STM images merely reflect the local DOS integrated between zero (Fermi energy) and the bias voltage, the topographic map Fig. 22 of the same area acquired with a bias voltage below the energy of the impurity state indeed has a significant local variation, consistent with that in Fig. 21(b). From the topographic image Fig. 22(a) we can identify the impurity position on the atomic scale. The largest enhancement correlates with positions of three surface Se atoms (marked with small black circles) forming a regular triangle. These atoms terminate three *ppσ*-bonded chains passing through the Se1 site five atomic layers below the surface, for which we also use notation Se3. The origin of the clover-shaped electronic features exclusively from the Se3 atomic plane is also supported by the low surface density of these features, considering the high doping level of the sample. It is surprising that

the electronic properties associated with impurities in the Se3 position are so special. This issue will be addressed in the following analysis.

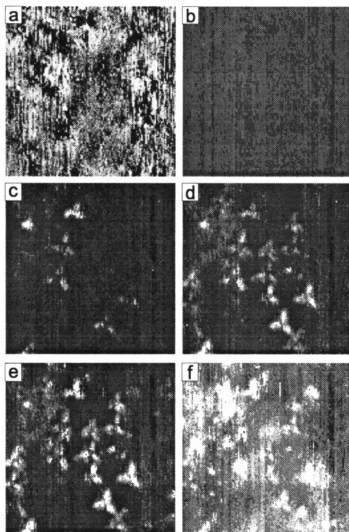


Figure 19: (a) A 30×30 nm topographic map of $\text{Bi}_2\text{Se}_{2.85}$ acquired at bias $V = -0.2$ V, $I = 50$ pA. The total scale is 1.6 \AA . (b)-(f) CITS maps acquired simultaneously with (a) by fixing the tip at each point of the map (a) and measuring the differential conductance at different bias voltages, using a standard lock-in technique with ac bias voltage modulation of 4 mV. Biases used for CITS images are (b) -200 mV (the same as topographic mode bias), (c) -350 mV, (d) -400 mV, (e) -450 mV, and (f) -550 mV.

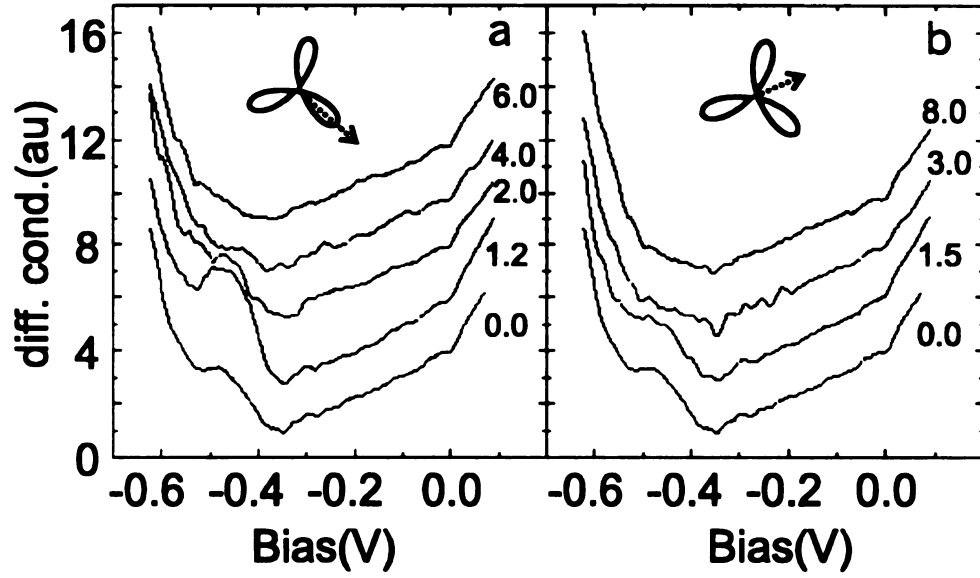


Figure 20: Differential conductance spectra acquired in $\text{Bi}_2\text{Se}_{2.95}$ in the vicinity of a clover-shaped impurity. The distance from the center is marked on the right. Curves are offset for clarity. Spectra (a) acquired along one of the clover lobes, and (b) along a line between the lobes, as shown in insets.

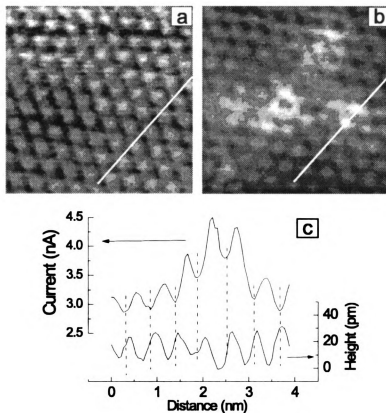


Figure 21: (a) A 5.1×5.1 nm topographic map encompassing one of the clover-like features. Sample bias voltage is -0.3 V, current setpoint 0.4 nA. (b) CITS current image map acquired at -0.7 V bias simultaneously with (a).

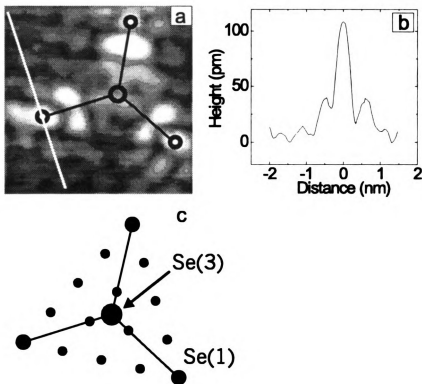


Figure 22: (a) A 3.5×3.5 nm topographic image of the same feature as Figure 21. The acquisition parameters are $V = -0.6$ V bias voltage and 1.0 nA current set-point. The inferred positions of the atoms terminating the $p-p-\sigma$ -bonded chains going through the defect atom are shown in small circles. The corresponding position of the sub-surface defect is shown with a larger circle. (b) Crossection through (a). The defect corrugation is more than 1\AA , while the atomic corrugations away from the defect are only 20 pm. (c) A schematic of the interpretation of (a) in terms of $p-p-\sigma$ -chains.

Chapter 4

Theoretical Modeling and Analysis of STM on Bi_2Se_3

4.1 Modeling STM with Band Structure Calculations

In many cases the surface effects are neglected in the analysis of the STM data. It is assumed that STM spectroscopic information directly reflects the bulk electronic properties of the sample, a simplistic picture briefly described in the beginning of Section 2.2. This indeed holds in most cases for metals, for example, proving STM to be an invaluable probe of superconductivity [48]. The electronic structure of semiconductors is not as robust as that of metals, and the surface can have a profound effect [9]. Model calculations of the STM, therefore, must incorporate the surface in some way. The most straightforward way is to perform a complete band structure calculation for a crystal terminated with a surface. Equation 2.10 with transmission amplitude expressed from Equation 2.17 relates the local electronic density of states several angstroms away from the surface to the STM tunneling current. The surface usually can not be included explicitly in band structure calculations, except

for relatively small cluster calculations. Such cluster calculations face the opposite problem of not incorporating a sufficient number of atoms to imitate the bulk properties. In band structure calculations performed for periodic systems, the surface can be modeled by a slab geometry [6]; sufficiently thick slabs of the crystal are separated by a significant distance from each other (of about 10\AA) and are used as supercells (i.e. with slabs repeated periodically). The thickness of the slabs has to be sufficient enough for the bulk properties to evolve. In three-dimensional materials the determination of the minimum slab thickness sufficient to reproduce both surface and bulk properties is not straightforward. Starting from some small value, the thickness of the slab has to be increased until the bulk properties evolve and the results of the model calculation stabilize.

4.2 *ab initio* Modeling of STM on Bi_2Se_3

In case of Bi_2Se_3 the choice of the proper slab thickness is simplified by the strongly layered nature of the material. Due to the weak coupling between the layers, electronic structure of even a single layer is close to that of the bulk (terminated by two surfaces). The analysis presented here, in particular the tight-binding model discussed below, shows the importance of the weak interlayer coupling for *certain* electronic features, related to the the clover-shaped impurity states described in Section 3.8. To incorporate this effect, in band structure supercell calculations we¹ include three quintuple layers. This comprises the actual bulk unit cell, i.e. in this case the supercell contains the same number of atoms as a unit cell, with the distance between the 3-layer slabs increased. The band structure calculations were performed [49] in the local density approximation (LDA) of the density functional theory [50]. The calculations were based on a full relativistic linear augmented plane wave method

¹The theoretical analysis described in this Section was performed in collaboration with D. Bilc and Prof. S. D. Mahanti at the Department of Physics and Astronomy, MSU.

(LAPW). The relativistic corrections (spin-orbital interaction effects) are important in Bi_2Se_3 because of the large atomic number of Bi.

In Figure 23 the bulk and slab band structure calculations are compared. The bulk calculation correctly reproduces the established semiconducting gap value of 0.3 eV. The high dispersion directions of the bands can be simplistically thought of as the directions of the dominant bonding underlying the band formation. In spite of the weak interlayer coupling, the band structure calculation for a slab shows some significant deviations from the bulk calculation:

- The number of bands in the higher valence band energy range (between about -0.2 eV and 0.0 eV) is reduced.
- There is finite density of states within the bulk gap.
- There is no dispersion along ΓZ direction, perpendicular to the layers.

The last effect is simply due to the negligible bonding between the slabs. The other two are surface effects, consequently having a profound impact on the STM differential conductance spectra. In particular, from the analysis of the partial DOS, the bulk high valence band (HVB) states are mostly Se1-type. Appreciable dispersion along the ΓZ direction indicates that coupling perpendicular to the layers is important for the formation of these states. In the slab, the HVB states are predominantly Se3- and Se4-type. Here the numbers correspond to the Se atomic plane positions with respect to the surface, e.g. Se1 position is at the surface. The HVB is recovered away from the surface, but the surface Se1 contribution to the HVB states is suppressed by lowering the Se1-type states in energy. Since mostly the interlayer Se1-Se1 bonds are broken at the surface, this leads to the conclusion that the valence band maximum (VBM) states result from antibonding interlayer Se1-Se1 coupling. Such an unusual VBM formation mechanism is thus responsible for the significant difference between the band structures in Figure 23.

The Fermi level in the calculations Figure 23 is arbitrarily assigned. To relate to the STM data on n-doped samples, we need to adjust it correspondingly, moving up by 0.3 eV. Then the reduced differential conductance in Figure 16 between the VBM at -0.3 V and around -0.5 V directly corresponds to the suppressed HVB states in Figure 23.

The finite DOS within the bulk gap observed in STM conductance is also clearly reproduced by the slab DOS calculation (figure 24(b)). The in-gap band is predominantly of Bi character and has very high dispersion along the surface (e.g. ΓM or ΓK directions). Given this information the likely origin of these states is rehybridization of Bi conduction states along the surface, bringing the bonding-type states down below the bulk conduction band minimum. This effect is probably caused by the charge transfer at the surface due to Se1-Se1 broken bonds. Further analysis is necessary to clarify the physics of the phenomenon. These in-gap states are not reproduced by the non-self consistent tight bonding calculations we present in the next Section.

As described in Chapter 2.2, the STM differential conductance spectra are well reproduced by the local DOS calculated at the position of the tip. Such calculations are currently in progress at MSU. They require a significant amount of resources due to the high energy resolution required and the low electronic densities at energies close to the semiconducting gap. The calculation of the partial density of states at the positions of the surface Se1 atoms is relatively simpler because of the higher electronic density at the atomic positions. Such a calculation is also a good approximation for the STM spectra because it is representative of the surface electronic properties. The calculation shown in Figure 24(a) reproduces the finite surface DOS in the semiconducting gap, and the suppressed differential conductance just below the gap, important surface effects discussed above. The low (as compared to the measured) calculated density of in-gap states could have several explanations. First, the partial Se1 density of states is not expected to quantitatively reproduce in detail the STM spectroscopic measurements which reflect the partial DOS *above* the sam-

ple surface. Second, the surface in-gap states are likely very sensitive to the surface structural relaxation, which has not been addressed in the calculation. Third, the calculations shown here are performed at the limit of the capabilities of the LDA approximation in terms of spectroscopic precision, and it is possible that due to the limitations of the method such a calculation *in principle* cannot give a quantitative account of the experiment. In particular, it is well known that LDA calculations typically underestimate semiconducting gap values by about 50%. As we have shown, a much better agreement between LDA calculations and experiment is found in case of semiconductors under study.

Comparison of the bulk and slab total DOS in Figure 24(b) leads to conclusions consistent with the analysis given for the the band structure calculations. The major difference between the spectra is in the HVB energy range. The variations in the conduction band part of DOS in the slab geometry are due to van Hove singularities introduced by making the system effectively two-dimensional.

The topographic images can also be modeled with band structure calculations. The simplest approximation for the topographic image is a map of the local charge density integrated between Fermi level (0 V) and the specified bias voltage, calculated at a set of points lying in a plane above the surface (or in a supercell calculation between the slabs, but closer to one of the two Se1 atomic planes) . Strictly speaking, such a map reflects the tunneling current map acquired by fixing the tip at a certain distance from the surface plane and scanning laterally above the surface. However, it is also a reasonable approximation for the topographic image acquired at a fixed current, scaled by the surface effective work function. The relation can be expressed by

$$\Delta z = \left(\frac{dI}{dz} \right)^{-1} \Delta I, \quad (4.1)$$

where Δz is the tip height variation in the fixed current topogaphic imaging mode, ΔI is the calculated current variation, and $\frac{dI}{dz}$ is related to the effective work function through Equation (2.22). In Figure 25 a calculated charge density map of Bi_2Se_3

surface is presented. The local partial charge density is integrated between 0 V and -0.3 V at each point of a plane positioned between the slabs 3 Å above the Se1 plane (about 6 Å from the Se1 plane of the next slab). Since the tunneling current decreases by roughly an order of magnitude per Å of tip-sample separation (see Section 2.3.5), the charge density map reflects the contribution of only the closest slab. Both the surface Se1 and subsurface Bi atomic positions give contributions to the map at the -0.3 V bias. This is consistent with the experimental observations described in Chapter 3 and the expected predominant contribution of Bi to the in-gap surface states.

It is difficult to model the impurity states with the *ab initio* calculations used for the analysis of the surface effects. In a supercell geometry the supercell size has to be sufficiently large to eliminate the interaction between different impurity states. This results in supercell sizes of at least dozens of angstroms, containing more than a hundred of atoms [46, 47]. While not impossible in principle, such a calculation requires a significant amount of resources, both in terms of computer memory and computation time. Below we will present a simple, very rough, tight-binding model which nevertheless conveys the important physics, and enables one to draw important conclusions from the analysis of the defect states. A big advantage of the model as compared to the *ab initio* calculations is a very clear qualitative picture it suggests.

4.3 A Tight-Binding Model of Surface and Impurity States in Bi_2Se_3

Below we will develop a simple one-dimensional qualitative tight binding model in the approximation of linear combination of atomic orbitals (LCAO). It lacks a number of important elements necessary to reproduce the quantitative features of the electronic structure. Instead it offers simplicity and a clear qualitative picture of the physical effects it describes.

The wave function in LCAO approximation is

$$\psi = \sum_{\mathbf{i}} u_{\mathbf{i}} \phi_{\mathbf{i}}, \quad (4.2)$$

where $\phi_{\mathbf{i}}$ are atomic wave functions. The Hamiltonian is a sum of the atomic terms and overlap integrals

$$(\phi_j, H \phi_{\mathbf{i}}) = E_{\mathbf{i}} \delta_{ij} + V_{ij}, \quad (4.3)$$

which is a completely general form in the approximation (4.2). We use the simplest non-self consistent approximation for $E_{\mathbf{i}}$ with empirical atomic terms [52]. This approximation is usually made for covalent bonding, where the effects due to charge transfer are not important, but it is sufficient for the qualitative analysis presented here. We take into account only the $pp\sigma$ nearest neighbor interaction between the nearest neighbors, bonding each atom with six other atoms in the neighboring atomic planes. This approximation splits the LCAO Shrödinger equation into a set of independent systems describing $pp\sigma$ -bonded chains, as shown in Figure 10. The chains cross each layer in three mutually nearly perpendicular directions, with a 5-atom unit Se1-Bi-Se2-Bi-Se1 representing a layer. The three remaining nonzero overlap integrals in (4.3) are V_1 , V_2 , and V_3 , corresponding to Se1-Bi, Se2-Bi, and Se1-Se1 $pp\sigma$ bonds (Figure 26). These overlap integrals can be approximately calculated from the interatomic distances, using a formula [52]

$$V_{pp\sigma}(\text{eV}) \approx \frac{28.2}{d^2(\text{\AA})}, \quad (4.4)$$

where d is the distance between atoms in angstroms.

Parameter	E_{Se}	E_{Bi}	V_1	V_2	V_3
Value (eV)	-9.5	-7.0	3.26	3.05	2.49

Table 4.1: Parameters used for the LCAO model calculation

Consider first a 5-atom unit representing a single layer. The symmetry of such a unit is $D_{\infty h}$, which includes a continuous rotation around the axis of the chain and an inversion through the center Se2 atom. The eigenfunctions of the Hamiltonian can be selected to transform under one of two irreducible representations: a_{1g} , symmetric under reflection, or the antisymmetric a_{2u} . The Hamiltonian of a 5-atom unit is

$$H = \begin{bmatrix} E_{Se} & V_1 & 0 & 0 & 0 \\ V_1 & E_{Bi} & V_2 & 0 & 0 \\ 0 & V_2 & E_{Se} & V_2 & 0 \\ 0 & 0 & V_2 & E_{Bi} & V_1 \\ 0 & 0 & 0 & V_1 & E_{Se} \end{bmatrix} \quad (4.5)$$

In Figure 26 the level scheme of the 5-atom unit is shown. Two lowest energy states are bonding, two highest energy states are antibonding, and one state is nonbonding, i.e. it has nonzero contributions only from Se orbitals

$$\psi_0 = \frac{1}{\sqrt{2 + (V_1/V_2)^2}} \left[\phi_{Se1} + \phi_{Se3} - \left(\frac{V_1}{V_2} \right) \phi_{Se2} \right], \quad (4.6)$$

and its energy is $E_0 = E_{Se} = -9.0 \text{ eV}$. Here the Se3 position is equivalent to Se1. The energies of the other levels are -13.8 eV , -11.7 eV , -4.8 eV , and -2.7 eV . The energy span of levels for a single unit is, therefore, 11.1 eV . Asserting that the model captures the dominant contribution to the hybridization in the system, the energy span of the p-type bands obtained with a much more precise calculation [27] is about 10.8 eV . The s-type bands are positioned 10 eV below the Fermi level and practically do not hybridize with the p-states. To figure out the filling of the levels, we consider 3 p-electrons on each Bi atom, and 4 p-electrons on each Se atom, a total of 18 p-electrons for a 5-atom unit. This number has to be divided by 3 since each atom in fact belongs to 3 $pp\sigma$ -bonded chains. Taking also into account the double spin degeneracy of each level shown in Figure 26, the three bottom levels are filled, and the two highest (antibonding) levels are empty.

4.3.1 The Nature of the VBM State in Bi_2Se_3

Consider the highest filled nonbonding state of a 5-atom chain (4.6). If the interlayer interaction V_3 is included as a perturbation, the correction to the energy of this state is to the first order

$$\Delta E_0 = \pm (\psi_0, H_1 \psi_0) = \pm \frac{V_3}{2 + (V_1/V_2)^2}, \quad (4.7)$$

where H_1 is the contribution to the hamiltonian from the interlayer coupling, and only bonding or antibonding in the sense of the interlayer coupling states are considered (taking "–" for bonding and "+" for antibonding). In a crystal the bonding and antibonding states represent the extrema of the band formed through the interlayer coupling. The correction due to the interlayer coupling is proportional to the square of the wave function amplitude on the interlayer Se1 atoms. (4.7) is large because the wavefunction (4.6) has a significant amplitude on Se1. In the model the surface effects amount to the absence of the contribution (4.7) at the end of the $pp\sigma$ chain, i.e. at the surface.

In Fig. 27(a) the evolution of the energy levels of a chain consisting of 5-atom units is plotted as a function of the number of units, modeling the evolution of bulk states from the states of a single layer. The contribution of the interlayer coupling increases with the number of units. As discussed above, this also increases the HVB bonding-antibonding splitting, raising the VBM energy and reducing the gap. The gap value saturates at about 10 units incorporated. The saturated gap value does not reproduce the experimental bulk value of the semiconducting gap mainly because of the ignored $pp\pi$ interactions between the chains, and neglected spin-orbit interactions, which have a significant contribution due to the high Bi atomic number [53], as well as non-self consistent approach to the calculation of the energy levels. In Figure 27(b) the amplitudes u_i of the VBM state are plotted. In agreement with the general considerations, the VBM state is antibonding, as all the Se1 orbitals

have contributions with the same sign. Although in contrast to the non-bonding state of a 5-atom unit, the amplitudes on Bi are not zero, the Bi amplitudes are lower than Se. An important conclusion drawn from Figure 27(b) is the decay of the HVB states at the surface. This result is qualitatively consistent with much more precise *ab initio* calculations presented in Section 4.1. In semiclassical terms, this effect is expressed as lowering of the VBM energy at the surface, with the value of semiconducting gap larger at the surface than in the bulk. Such an effect of the surface proximity on the semiconducting gap is similar in GaAs and is opposite in Si [9]. This similarity leads to a number of analogies between certain aspects of the near-surface impurity states in these semiconductors, which we will discuss below. The tight-binding model does not reproduce the finite surface density of states in the gap. It was shown in Section 4.1 that these states have a high dispersion direction parallel to the surface. This implies that the states are formed by the hybridization along the surface. On the other hand, in the tight-bonding model the system is approximated by a one-dimensional chain with only one atom representing the surface, which eliminates the surface hybridization effects.

4.3.2 Electronic Topological Transitions

Dramatic increase in the carrier concentration is observed in p-doped semiconductor Bi_2Te_3 upon cooling. The structure of Bi_2Te_3 is identical to Bi_2Se_3 , and the physics of the gap formation is similar. A large carrier density increase has been observed in $BaBiTe_3$ upon application of external pressure [55]. The structure of $BaBiTe_3$ is in some respects similar to Bi_2Te_3 . It contains blocks of Bi_2Te_3 structure, and like in Bi_2Te_3 the HVB states are formed predominantly from the Te_1 -type states [53]. A similar, albeit much weaker, temperature effect is also observed in Bi_2Se_3 [59]. This is in contrast to the behavior of typical semiconductors, in which the thermally activated charge carriers freeze out at low temperatures [32]. The situation in narrow gap semiconductors is complicated by the high level of natural doping, such that the

material is usually near the degenerate limit.

Here we propose a hypothesis connecting the temperature/pressure dependence of the carrier concentration with the other electronic properties discussed in this chapter. As both *ab initio* and tight binding calculations have shown, the HVB in Bi_2Se_3 is very sensitive to the interlayer coupling. A similar effect is present in Bi_2Te_3 , although in contrast to Bi_2Se_3 the gap is believed to form as a result of anticrossing of the HVB and lower conduction band states [53]. The elastic constant corresponding the interlayer spacing is small, as follows from the small interlayer surface energy. This implies that the interlayer spacing varies significantly with external conditions, e.g. temperature and pressure. Such a strong interlayer spacing variation consequently results in significant variations of the properties of the semiconducting gap, affecting the carrier concentration. It is also likely that coupling of charge carriers to the acoustic phonons associated with the interlayer distance variations is anomalously strong, resulting in strong inelastic carrier scattering [26]. Polaronic effects are also likely to be prominent in these materials: higher local electronic density increases the distance between layers due to the electrostatic repulsion. This results in the formation of Bi-dominated in-gap states. We would like emphasize that this hypothesis has not been proposed or explored before, and it requires a more detailed theoretical analysis.

4.3.3 Bi_{Se} Defect States in Bi_2Se_3

The simplicity of the developed tight-binding model provides for an easy qualitative analysis of impurity states. Assuming that the clover-shaped impurity states described in Section 14 originate from Bi_{Se} antisite defects, we employ the model to explain why only Bi_{Se3} defect states are observed with STM, and are seen as resonances in the valence band. In Figure 28(a) the near-gap energy levels of a 16-unit chain are plotted as the Bi_{Se} impurity is placed on different Se sites along the chain. A dramatic effect of the surface proximity on the energy of the defect state is clear.

The second layer (positions 4–6) is only weakly affected by the proximity of the surface through the interlayer coupling to the surface layer, so the impurity level is split from the VBM and its energy is almost independent of the position. As the impurity position approaches the surface (positions 1–3), the impurity level energy is reduced as the surface gap opens up, so that impurity state merges with the bulk valence states, forming a resonance. In Figure 28(b) the VBM state with a Bi_{Se3} defect is plotted. In the model it is localized at the surface. By including the $pp\pi$ -bonding between the 1-d chains the HVB states are widened. As a result the Bi_{Se3} defect state likely becomes a resonance inside the valence band. Only the Bi_{Se3} defect state is observed in the experiment because Bi_{Se1} and Bi_{Se2} defect states are so much lowered in energy by the proximity of the surface that they form very small amplitude broad resonances in the valence band, while the Bi_{Se} impurities in the layers further from the surface are not seen with STM since they produce bound states in the gap, weakly coupling to the surface in-gap conducting states. Such states cannot sustain STM tunneling current.

4.4 Defect States in Medium-Gap Semiconductors

A brief review of STM experiments on GaAs and similar materials has been given in Section 3.7. In semiclassical terms, GaAs (and InGaAs) (1,1,0) surface is similar in electronic properties to Bi_2Se_3 , i.e. the valence band maximum energy decreases at the surface [56]. In spite of extensive experimental and theoretical research, the conduction band behavior at the surface is not well established, and a clear connection to the experimental measurements has not been made. We attribute this lack of understanding to the limited capabilities of the *ab initio* methods for the calculations of unoccupied states, as well as ambiguous interpretation of the inverse photoemission spectroscopy results [56].

The lack of understanding of the surface electronic properties in III-V semicon-

ductors complicates the interpretation of the surface STM probe in terms of the bulk bands. Here we suggest a possible interpretation of the results presented in Ref. [35, 36], consistent with our analysis of impurity states in Bi_2Se_3 .

At small impurity concentrations, the spatial scale of band bending is large enough to prevent direct electron tunneling from the STM tip into the bulk bands. The gap observed with STM is therefore dominated by the surface band bending. The Fermi level is typically pinned either at the top of the bulk valence band or at the bottom of the bulk conduction band, depending on doping (in low temperature-grown GaAs, the Fermi level can be pinned to the middle of the gap by the midgap EL2 donor states, produced by the As_{Ga} antisite defects). While in Si the in-gap surface states pin the Fermi level to the middle of the gap, it is not the case for GaAs (1,1,0) surface. In Ref. [57] various positions of the Fermi level with respect to the band edges were reported depending on doping. In interpretation of the impurity states presented in Ref. [35], it is important to note that in all the differential conductance spectra of GaAs with various dopings, the Fermi level was close to the center of the gap, and not pinned to one of the band edges. It is possible that the STM spectroscopy was dominated by the surface bending/gap opening effect, determining the positions of the band edges at the surface. According to the surface electronic properties calculations [56], the surface valence band maximum is 0.7 eV lower than in the bulk, and in LDA calculations the conduction band minimum is also lowered at the surface by the same amount. The Green function calculations, however, place the surface conduction band minimum at approximately the same energy as in the bulk. Generally, this technique has proven to give more realistic gap values [58]. Assuming the bulk p-doping of the samples, the position of the surface valence band maximum at about -0.7 eV below the Fermi level is consistent with the experimental spectra of Ref. [35]. It has been shown [45], that for a defect atom positioned in the surface layer of the (1,1,0) surface of GaAs, the defect state energy is strongly affected by the surface effects. For example, the As_{Ga} antisite defect forms a resonance *below* the

bulk valence band maximum. This naturally explains the wide impurity resonances between the Fermi level and the valence band maximum seen in the experiment. A similar explanation can be given for the impurity resonances observed on the (1,1,0) surface of InGaAs [36], and other III-V compound semiconductors. However, a more detailed analysis is needed to quantify the suggested picture, and clarify the surface band structure as seen with STM.

The qualitative picture described here also applies to Si, where semiclassically the semiconducting gap decreases at the surface, an effect which is also dependant on the surface reconstruction. In this case, the defect states formed by the near-surface impurities should appear deep inside the bulk semiconducting gap. These states are only confined to the surface and cannot be seen spectroscopically with STM. Consistent with this expectation, we know of no reports of the electronic structure STM studies of local defect states in Si.

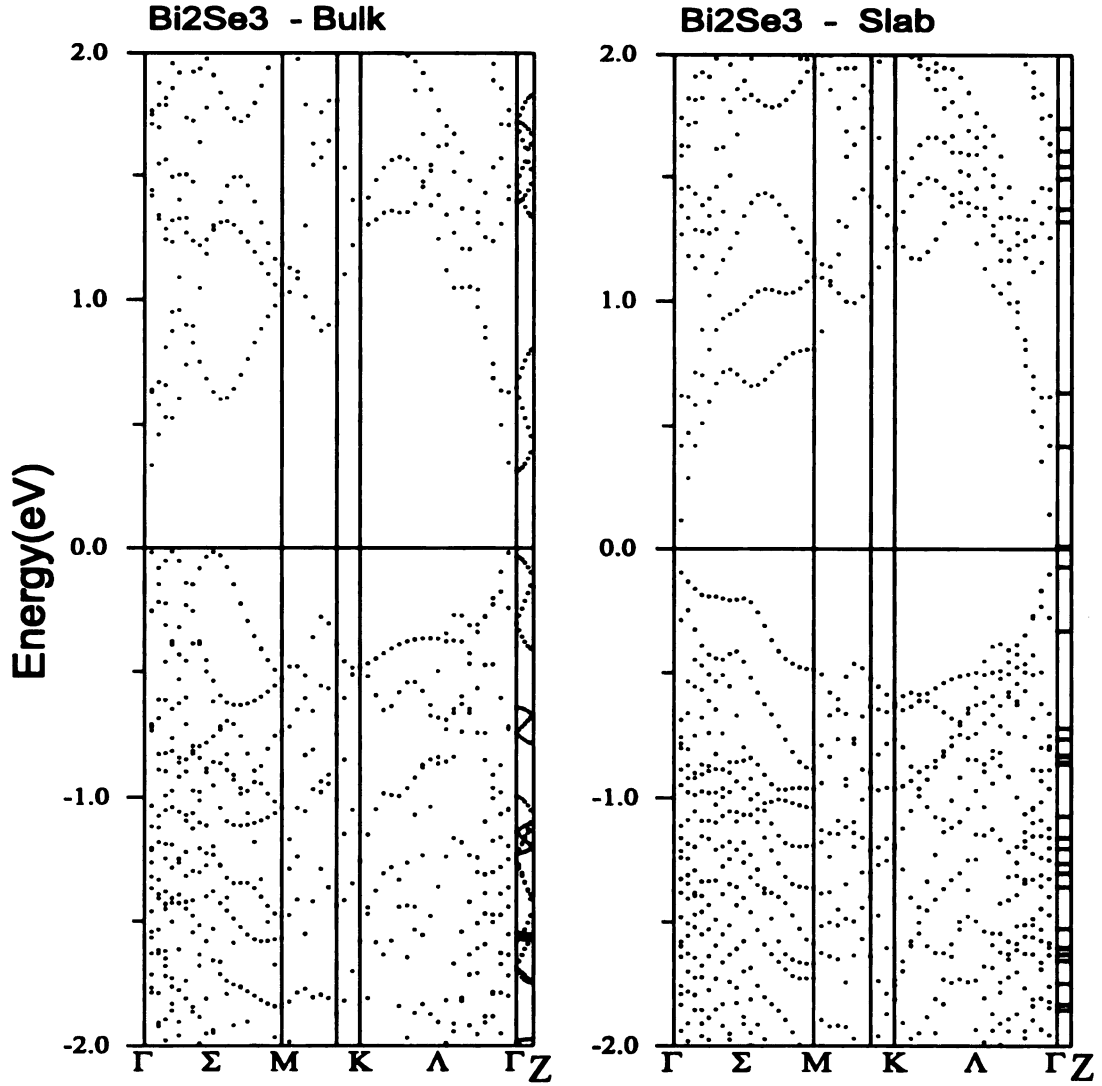


Figure 23: Left: band structure calculation for the bulk Bi_2Se_3 . Right: band structure calculation for Bi_2Se_3 in slab geometry, with distance between 3-layer slabs increased by 7\AA . For notations see Figure 11. The position of the Fermi level is fixed arbitrarily.

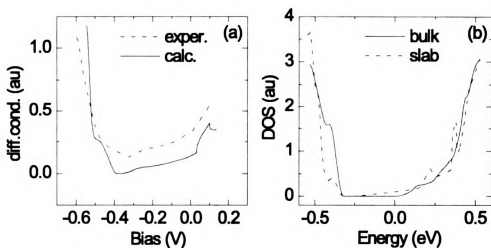


Figure 24: (a) STM differential conductance spectra calculated as described in the text *vs.* the measurement, (b) Total calculated bulk *vs.* slab DOS.

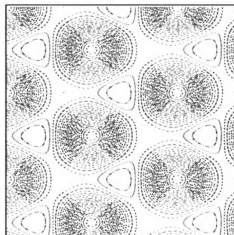


Figure 25: Topographic map of Bi_2Se_3 calculated as described in the text.

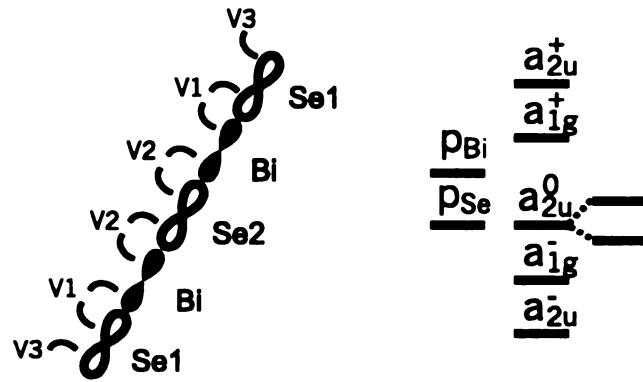


Figure 26: Left: Schematic of a tight binding one-dimensional model used. Open orbitals represent Se atoms, solid orbitals represent Bi atoms. Right: Energy levels of a 5-atom unit representing a single layer. The nonbonding level is shown further split by the interlayer interaction V_3 .

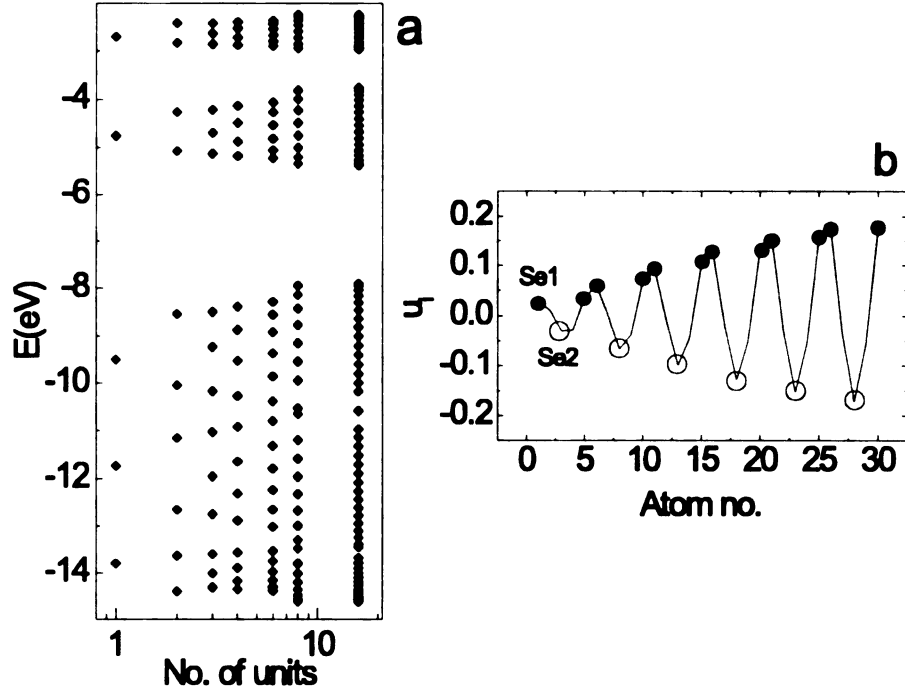


Figure 27: (a) Evolution of the energy levels of a chain consisting of 5-atom Se1-Bi-Se2-Bi-Se1 units as the number of units is increased. (b) Amplitudes u_i (Equation (4.2)) in the VBM state for a 16-unit chain plotted as a function of atomic position along the chain. Only the first 30 amplitudes are plotted. Solid circles — Se1 positions, open circles — Se2 positions.

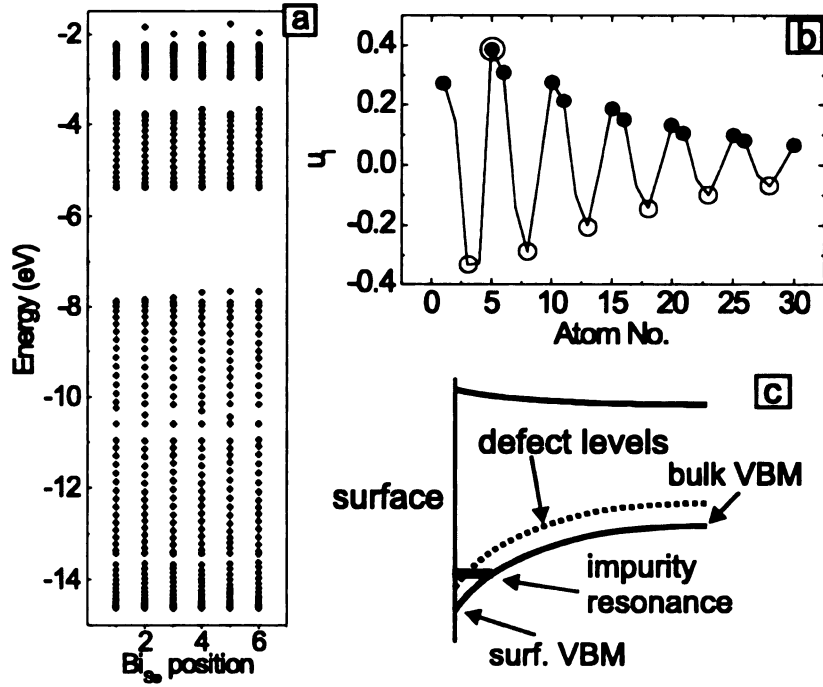


Figure 28: (a) Energy levels of a 16-unit chain (a total of 80 atoms) as a function of the position of a Bi_{Se} defect. The levels are calculated using a tight-binding model. (b) Same as Figure 27(b), with Bi_{Se3} defect introduced, shown with concentric circles. (c) A semiclassical schematic of the formation of the impurity resonance.

Chapter 5

Surface Effects in Bi_2Te_3

5.1 STM Studies of Bi_2Te_3

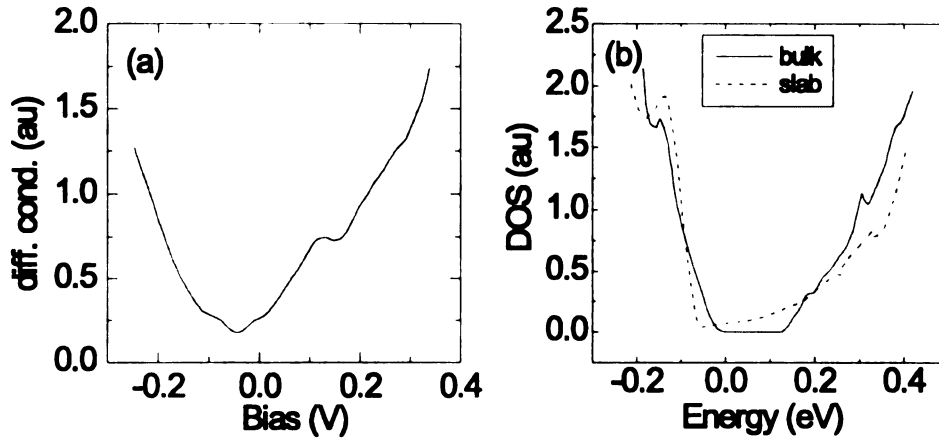


Figure 29: (a) A typical differential conductance spectrum of Bi_2Te_3 , (b) bulk vs. slab calculated DOS.

Bi_2Te_3 is a narrow gap semiconductor with the same structure as Bi_2Se_3 . It is slightly less ionic, as Te is heavier than Se and is less electronegative. As a result, the lattice parameters are larger than in Bi_2Se_3 . It was demonstrated in Section 4.3, that

in Bi_2Se_3 the semiconducting gap is formed as a result of the bonding-antibonding splitting of the atomic terms. The physics of the gap formation in Bi_2Te_3 is similar, although due to weaker bonding the gap value is smaller, about 0.15 eV. We should mention that the experimental and theoretical results for the gap value [25, 26, 27, 53, 54] vary from 0.05 eV to 0.2 eV due to the complicated structure of the band extrema and the indirect gap. The conduction band minimum and valence band maximum in Bi_2Te_3 are off the center of the Brillouin zone, and 6-fold degenerate [25]. In Bi_2Se_3 , on the other hand, the band extrema are both at the center of the Brillouin zone. This difference is believed to account for the difference in the thermoelectric performance of the two compounds. The high degeneracy of the conduction and valence band extrema in Bi_2Te_3 has been suggested to originate from the anticrossing of the conduction and valence bands [53]. Spin-orbit interaction effects have also been shown to be important for the formation of the semiconducting gap.

In this section we will discuss some electronic properties of Bi_2Te_3 observed with STM. In Figure 29(a) a typical differential conductance spectrum of naturally doped Bi_2Te_3 is shown. Nominally stoichiometric Bi_2Te_3 is naturally p-doped, with carrier concentration of about 10^{19} cm^{-3} . From the features of the spectrum it can be assumed that the gap energies lie between 0.0 V and about 0.18 V. The in-gap conductance is finite and the conductance just below the VBM is suppressed. These features are clearly similar to those in Bi_2Se_3 discussed in the previous chapters. Figure 29(b) shows the total calculated DOS of the bulk Bi_2Te_3 *vs.* a similar calculation for the slab. The calculation method and parameters were identical to those used in the analysis of Bi_2Se_3 described in Section 4.1. The bulk DOS reproduces the generally accepted semiconducting gap value of about 0.15 eV. In-gap surface states appear in the slab geometry calculation. It is clear from both the measured differential conductance and the calculation that the surface effects in Bi_2Te_3 are likely similar to the effects in Bi_2Se_3 which were discussed in the previous chapters. More detailed theoretical studies of the surface effects in Bi_2Te_3 are underway [49].

We have not studied Bi_2Te_3 doped with excess Bi, but the general similarity of the spectroscopic features to those of Bi_2Se_3 suggests that similar resonant states are likely to be observed. It is not clear if, as in Bi_2Se_3 , these states would necessarily originate from Bi_{Te} defects in the Te3 atomic plane. It is possible that the resonant states could originate from the Te1 or Te2 plane or from more than one atomic plane, resulting in clover-shaped electronic features with different sizes.

5.2 Surface Band Bending

Bi_2Te_3 can be substitutionally doped with excess Te to produce n-type samples. The data presented in this section were acquired on an n-doped with about 1.7% of excess Te sample. An interesting spectroscopic feature observed on this sample is a strong variation of the differential conductance spectra correlated with the surface termination. The crystal was cleaved with a scotch tape, exposing an atomically flat surface with some terracing. A topographic image of one of the areas with a high density of step edges is shown in Figure 30(a). Because of the large difference between the inter-layer and intralayer surface energies, it is expected that the step edge height should be a multiple of the quintuple layer thickness, which is about 10\AA . A cross-section plot Figure 30(c), however, shows that the terrace step heights are smaller than the thickness of the quintuple-layer. Another interesting feature of image 30(a) is the regularity of the topographic pattern. It is correlated with the lattice symmetry and is likely due to the lower bonding strength in the high symmetry directions along the layer. A differential conductance map Figure 30(b), acquired simultaneously with the topographic image, exhibits a correlation with Figure 30(a). To clarify the origin of the contrast in image 30(b), in Figure 30(d) we plot the differential conductance spectra acquired at points of the surface corresponding to different intensities. As can be seen from images 30(b),(d), the differential conductance spectra of areas with high differential conductance at the bias voltage of -80 mV exhibit spectroscopic features

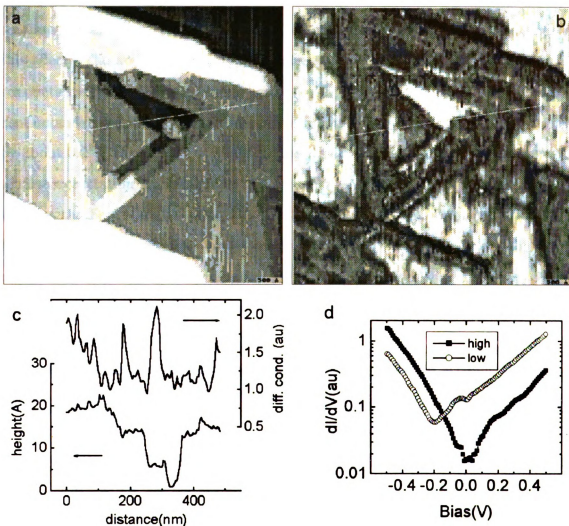


Figure 30: (a) Topographic image of a $0.5 \times 0.5 \mu\text{m}$ area of a Bi_2Te_3 sample doped with excess Te. The area has a high density of step edges and is not representative of the typical cleaved surface. Bias voltage $V = -80 \text{ mV}$. (b) A dI/dV map acquired simultaneously with (a). (c) Simultaneous cross-sections through (a) and (b). (d) Differential conductance spectra representative of high and low intensity areas in (b).

similar to those of p-doped samples (compare to Figure 29). They are typical of most of the sample surface. STM current through these areas, however, gradually changes their properties to n-type-like. We attribute this effect to the charging of adsorbates on the surface. Measurement under UHV conditions is likely to clarify the origin of this effect. The areas with low (before extensive tunneling) at -80 mV differential conductance constitute only a small percentage of the sample surface. The differential conductance spectra are roughly similar to the spectra of p-doped samples, shifted by about 0.2 V, close to the semiconducting gap value. This effect is consistent with a naive expectation for the difference between n-type and p-type samples: In n-type samples the Fermi level is pinned at the bottom of the conduction band, while in the p-type sample it is close to the top of the valence band. Therefore the difference between the spectra of p-type and n-type spectra is a rigid shift by the value of the semiconducting gap. Such a simple picture, however, breaks down when surface effects are taken into account.

To explain the observed phenomenon, we need to take into account the surface electronic structure. However small the interlayer coupling is, the surface electronic properties are different from the bulk. As was shown in previous Section, in Bi_2Te_3 the surface density of states in the semiconducting gap is finite due to the in-gap surface states which appear at the Te1-terminated surface, an effect similar to that observed in Bi_2Se_3 . The in-gap surface states originate from the empty conduction band states. In undoped samples the presence of such states would pin the surface Fermi level position to the valence band top. From Figure 30 we conclude that the bulk doping in the studied n-type sample is not sufficient to pin the Fermi level position at the surface to the bottom of the conduction band, as it is the case in the bulk. Therefore, the STM differential conductance spectra of n-type samples are in most areas similar to those of p-type samples because of the surface band bending. Due to the statistical variations in the density of defects in Bi_2Te_3 doped with excess Te, the density of the edge dislocations is increased, with higher probability of cleavage

along other than Te1 atomic planes. The resulting surface states are *different* than in the case of Te1 surface termination. It is likely that for certain terminations no conduction-like empty surface states appear in the gap. As a result the Fermi level is pinned at the bottom of the conduction band, resulting in spectra consistent with n-doping at the surface. The dependence of the surface electronic properties on the surface termination requires a careful theoretical analysis. Detailed surface band structure calculations are likely to clarify the observed effects.

5.3 Interpretation of the Electron Photoemission Experiments

Electron photoemission [61] experiments have become an important technique for studying the electronic band structure. In these experiments, a monochromatic photon beam is incident on the crystal surface, and the distribution of the emitted photoelectrons is analyzed. Due to the short mean free path of an excited electron in the solid [62], the technique typically probes the electronic properties of the surface layer about the 5 to 10 Å deep. We have shown that the electronic properties of Bi_2Te_3 (similarly to Bi_2Se_3), are significantly affected by the proximity at the surface. In particular, in photoemission studies of Bi_2Te_3 , in-gap states have been observed [63], which were originally attributed to donor impurity bands. Moreover, the difference between the spectra of p- and n-doped crystals could not be reduced to a simple rigid shift of the band structure due to the difference in the Fermi level position. It is clear from the STM studies, that the surface band bending and surface states play an important role for the surface electronic properties, and cannot be ignored in the analysis of the photoemission experiments.

5.4 The Thermoelectric Performance of Bi_2Te_3

Bi_2Te_3 combines a number of established features making it the best currently known room temperature thermoelectric material [69]. It is shown in Appendix C that in the semiclassical approximation the thermoelectric performance of a material can be described in terms of its basic electronic and lattice properties: effective electron and hole masses, degeneracy of conduction and valence band extrema, semiconducting bandgap, doping level, and lattice thermal conductivity. According to the semiclassical formulae derived in Appendix C, a large ZT-factor in Bi_2Te_3 is achieved through [66]

- 6-fold degeneracy of the active band extrema,
- highly anisotropic effective mass tensor, $m_x = 0.02m_0$, $m_y = 0.08m_0$, $m_z = 0.32m_0$,
- low phonon thermal conductivity $\kappa_{ph} \approx 1.5 W m^{-1} K^{-1}$,
- a gap size of about $6kT$ at room temperature, making the electron-hole recombination effect relatively insignificant.

Combining all the factors, an optimized ZT value of 0.5 is calculated in a single-band approximation [66]. This is less than the experimentally established value of 0.67 [67]. The value calculated in Ref. [66] is in fact overestimated: The transport is assumed in the direction of the highest carrier mobility yielding the highest thermoelectric performance. This value is multiplied by the degeneracy factor of 6. It is clear that the 6 effective mass ellipsoids have different orientations, resulting in averaging of the anisotropy of the transport properties and reduction of the calculated ZT value by roughly a factor of 2. Furthermore, the calculated value is decreased by the bipolar contribution and interband scattering. From where does the discrepancy come?

The STM study and analysis of surface and defect states presented in this dissertation gives a new perspective of some of the unusual electronic properties of narrow gap

chalcogenide semiconductors, including Bi_2Te_3 . It has been shown that the higher valence band states in Bi_2Se_3 are very sensitive to the interlayer coupling, which can be easily varied with external conditions. The near-gap structure of Bi_2Te_3 is more subtle due to a smaller gap value and larger spin-orbit coupling effects [53]. Detailed band structure calculations remain to be done. It is clear, however, that the near-gap electronic structure of Bi_2Se_3 and all related compounds is highly sensitive to the interlayer coupling. It seems to be a general consequence of the bonding scheme in these materials. Several important implications for the transport and thermoelectric properties follow from the connection between the weak interlayer coupling and low energy electronic structure:

- Defects producing local variation in the interlayer coupling, e.g. Se1/Te1 vacancies, interlayer Bi interstitial defects, and $Bi_{Se/Te}$ antisite defects, should result in strong scattering of charge carriers, reducing the electrical conductivity. Such defects are likely to have a relatively low formation energy. In particular, Bi_2Se_3 is more ionic than Bi_2Te_3 , and its natural n-doping indicates that the density of interstitial Bi is higher than in Bi_2Te_3 , resulting in strong impurity scattering. It is also established that doping Bi_2Te_3 with excess Bi results in drastic decrease of the hole mobility [25]. This probably results from the strong hole scattering on the interlayer Bi defects.
- Temperature and pressure, including chemical pressure in alloys and mixed compounds, should have a strong effect on the near-gap electronic structure. Such an effect has indeed been experimentally observed [26, 55], although theoretical aspects have not been explored yet. It is also known [25] that alloying of Bi_2Te_3 with other compounds significantly changes the gap value and drastically affects the thermoelectric performance.
- The details of the near-gap electronic structure of Bi_2Te_3 remain elusive even from the highly accurate electronic structure calculations [53, 54], because

they are strongly dependent on the approximations made. In particular, LDA method usually applied for the band structure calculations fails at low electronic density and high density gradient. This situation is encountered for the interlayer coupling, responsible for details of the formation of the semiconducting gap.

- Experimentally, multiple cracks in the samples should result in appearance of in-gap states or reduction of the semiconducting gap observed with macroscopic probes.
- The phonon modes corresponding to the interlayer distance variations are heavy, and are likely important for the low lattice conductivity through Umklapp scattering. An interesting consequence of the electronic structure analysis is the anomalous electron coupling to these modes. The enhanced phonon-drag effect could be responsible for the experimental ZT value being higher than that calculated in the elementary free-electron approach. Such a strong phonon-drag effect on the thermoelectric properties has been observed in metals and semiconductors at low temperatures [76].
- One of the attractive possibilities for improving the ZT factors of thermoelectric devices has been proposed by reducing the dimensionality of the active device [66, 72]. It has been conjectured [66] that, due to the increased anisotropy of transport, the optimal thermoelectric performance would be achieved, for example, in a single layer of Bi_2Te_3 patterned in a narrow strip. Our STM study shows that the advantageous bulk electronic properties of Bi_2Te_3 can not be projected to nanostructures, in which surface effects would play a dominant role. In fact, a good thermoelectric performance for nanostructures has been achieved in other semiconducting systems [72, 71].

Chapter 6

Summary

STM and its extensions represent an invaluable tool providing complimentary information to other imaging and spectroscopic methods. In this dissertation, various experimental STM techniques and the theoretical methods used for their interpretation have been discussed. Some technological innovations in the low-temperature STM which we have developed have also been described.

We have performed and analyzed STM low temperature measurements on the narrow gap semiconductors Bi_2Se_3 and Bi_2Te_3 . Two important experimental results presented in this dissertation are:

- The surface electronic structure of Bi_2Se_3 and Bi_2Te_3 gives an important contribution to the near-gap surface electronic properties observed with STM. This observation leads to a number of important consequences for our understanding of the bulk electronic structure as well.
- The defect states observed with STM in Bi_2Se_3 doped with excess Bi are resonances inside the bulk valence band, in the energy range where the valence band is suppressed at the surface. Thus surface effects provide a mechanism for the charge relaxation of the impurity states through their hybridization to the bulk valence band.

We have employed LDA as well as tight binding calculations for the theoretical analysis of the STM experimental data. From our calculations it follows that the weak interlayer coupling, frequently described as Van der Waals-type, is in fact of covalent type. The near-gap electronic states are especially sensitive to the interlayer coupling strength. At the surface, this results in a strong variation of the electronic structure and appearance of in-gap surface states. This effect is observed with STM. Together with the band bending effect, this effect also accounts for the discrepancy between the experimental photoemission spectroscopy results and bulk band structure calculations. Besides the surface states, the strong dependence of the near-gap states on the interlayer coupling gives a clue to the previously unexplained physical effects, such as increase of carrier concentration and conductivity with pressure or as the temperature is lowered. We also predict several unusual properties which follow from the nature of the near-gap electronic structure, such as anomalous electron-phonon coupling, importance of the polaronic effects, and strong electron scattering on the interstitial defects. A deeper understanding and observation of the predicted effects will likely have an important impact on the future design and improvements of thermoelectric materials.

Observation and interpretation of the surface states in Bi_2Se_3 leads to an intuitively transparent explanation for the origin of the impurity states observed in the material doped with excess Bi. Bi_{Se} antisite defects are acceptors, so in the bulk they produce shallow electronic states above the VBM. These states are bound, and cannot support current, making them invisible to STM. Near the surface, the semiconducting gap in Bi_2Se_3 is increased, and as a result the impurity states are lowered in energy by the proximity of the surface. For certain defect positions, the impurity states are high intensity resonances below the bulk VBM, easily observable with STS. The direct observation of the correlation between the defect position with respect to the surface and its electronic properties in Bi_2Se_3 suggests an interpretation for the STM experiments on a number of other semiconducting systems, in particular GaAs and

InGaAs. We believe that in these semiconductors, surface properties are extremely important for the STM observations of the impurity states.

Appendix A

STM Extensions Employing Capacitive Coupling

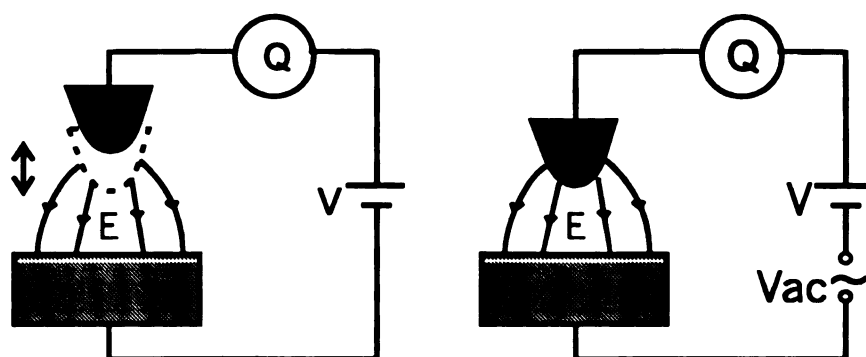


Figure 31: Left: Kelvin probe technique. The tip is vibrated above the surface, resulting in capacitively induced ac current through the tip dependant on the applied dc bias. Right: Scanning Capacitance microscopy technique. The small ac modulation of the bias voltage results in a capacitively induced current in the tip.

The Kelvin probe and scanning capacitance microscopy techniques are outside the scope of STM methods since they do not employ tunneling current. Instead, the capacitively induced ac current in the tip is detected. However, these techniques do not require any modifications to the standard STM setup, thus representing a natural extension of STM. The Kelvin probe is a method to measure the relative value of the

work function, i.e. the work function of one metal as compared to the other used as a probe. Consider a parallel plate capacitor with a dc voltage V applied between the plates. The capacitance is

$$C = \frac{\epsilon_0 A}{d}, \quad (\text{A.1})$$

where A is the area of each plate, d is the distance between the plates. If the plates are vibrated with respect to each other with a small amplitude

$$d(t) = d_0 + \delta d \cos \omega t. \quad (\text{A.2})$$

then the current through the external circuit changes according to

$$I = \frac{dQ}{dt} = \frac{d}{dt}(CV) \approx \frac{V\epsilon_0 A \omega \delta d}{d_0^2} \left(\sin \omega t - \frac{\delta d}{d_0} \sin 2\omega t + \dots \right) \quad (\text{A.3})$$

There is an ac current induced in the circuit unless the voltage between the plates is zero. This simple conclusion has to be modified when extended to the plates made of dissimilar materials. The effect of the finite contact potential (different surface work functions of the two capacitor plates) is demonstrated in Figure 32. The top schematic shows 2 neutral dissimilar metals with different Fermi levels as measured from the vacuum level. The difference between the Fermi levels is called a contact potential. When contact is made between the metals, the charge redistributes between them so that the Fermi levels are equilibrated, as shown in the middle schematic. However, this results in electrical field between the metals originating on the charges accumulated at the interface. If the distance between the plates is varied, electrical current will flow in the circuit. When external potential is applied equal to the contact potential, it eliminates the electrical field between the metals, as shown in the bottom part of Figure 32. Thus, when two dissimilar metals are used to form plates of the capacitor, the potential should be understood as electrochemical potential, so the ac current in the circuit is reduced to zero when the external potential is equal to the work function difference (contact potential) between the two metals. In Kelvin probe setup based on an STM the role of one of the plates is played by

the STM tip, the other is the sample, and the bias voltage is adjusted to minimize the ac current induced by the vibration of the tip. The typical STM tip is about 100 nm in diameter, and it can be positioned about 20 nm away from the surface and vibrated with an amplitude of about 10 nm at a frequency of about 1 kHz. The electrochemical potential difference of 1 mV then results in ac current with amplitude of 10^{-16} A. This rough estimate leads to the conclusion that a special highly sensitive current amplifier is needed for an effective Kelvin probe local measurement. A highly sensitive low temperature amplifier described in Section 2.4.2 has been successfully applied for this measurement [10, 77].

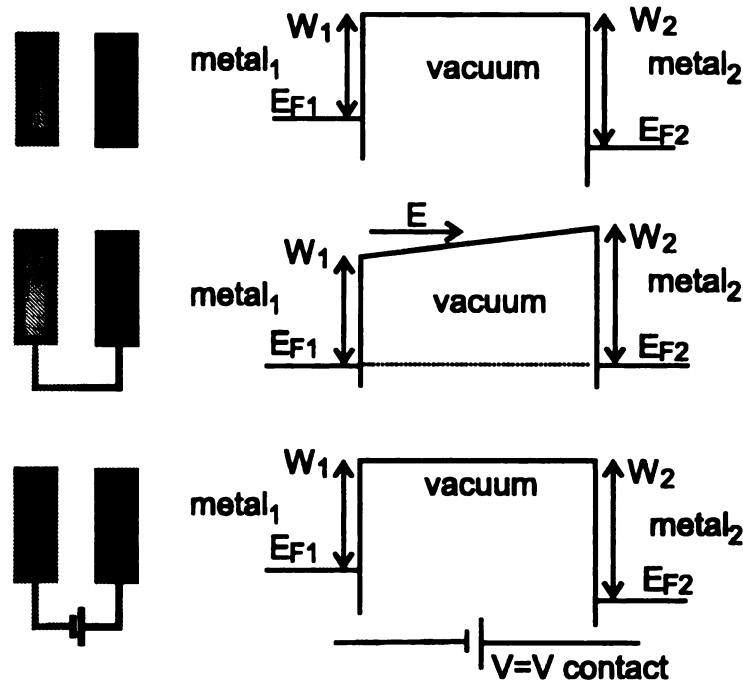


Figure 32: Contact potential effects.

Here we only briefly mention the scanning capacitance technique which, like Kelvin probe, is based on the capacitive coupling between the tip and the sample. The ac current in the tip is induced by a small modulation of the bias voltage. The field originating locally from the sample depends on certain electronic properties of the sample at the Fermi energy. The method proved effective in imaging of local inhomogeneities in the Quantum Hall regime of the two-dimensional electron gas [10].

Appendix B

The Layered Nature of Bi_2Se_3 and Related Compounds

The surface energy is defined as the amount of external work necessary to separate two faces of a crystal by a sufficiently large fixed finite distance. This definition is similar to that of the work function [30]. The seeming arbitrariness is due to the crystal shape-dependant macroscopic electric field between different faces of crystal, resulting in size- and shape-dependant macroscopic contribution to the work necessary for the separation of two surfaces by an infinite distance. The interlayer surface energy in Bi_2Se_3 is orders of magnitude lower than the intralayer surface energy [70]. The mechanical weakness of crystals against cleaving between the layers results in multiple cracks in macroscopic samples [25], complicating transport measurements. The origin of such a strongly layered structure of Bi_2Se_3 and related materials has been attributed to the weak Van der Waals-type coupling between the layers [25, 26, 27, 53, 54]. Strictly speaking, Van der Waals bonding happens only in molecular crystals [30], where atomic levels are complete, and the 0th-order electrostatic interactions between atoms are absent. In Bi_2Se_3 and similar compounds this is far from being the case. *ab initio* calculations can be utilized for the analysis of the atomic charge distributions. In order to make a qualitative estimate, however,

the tight binding model developed above is sufficient. For a 5-atom unit, the charge distribution in each of the 3 filled levels is shown in Table B.1.

	Se1	Bi	Se2
-13.8	0.11	0.19	0.38
-11.74	0.34	0.16	0
-9.5	0.32	0	0.36
total	0.77	0.35	0.74
ionicity	0.62	-0.9	0.44

Table B.1: Electronic density in the three filled levels of a 5-atom unit representing a single layer.

The atomic charge (in electrons) is calculated as a square of the atomic contribution to the wave function. To calculate the ionicity, we multiply the total charge given in Table B.1 by 6 (to take into account the double spin degeneracy, and that three identical chains pass through each atom), and subtract the number of electrons in a neutral atom. The calculated ionicity is rather weak as compared to the valency of each atom. Due to such weak ionicity, the bonding in Bi_2Se_3 and related compounds is usually referred to as mixed ionic-covalent. For example, while a neutral Se atom has four 3p electrons, in the tight-binding model (Table B.1, last row) Se1 has only 0.6 additional electrons. This small extra charge on the interlayer Se1 atoms should result in *repulsive* Madelung contribution to the surface energy, opposite to the induced dipole-dipole attraction (Van der Waals forces), which has been suggested as the dominant bonding mechanism between the layers [25, 26, 27, 53, 54]. On the other hand, *ab initio* calculations presented in Section 4.1 have demonstrated the importance of the interlayer hybridization and surface states for the interpretation of the experimental STM spectroscopic data. Therefore, it is more appropriate to interpret the low interlayer surface energy in terms of the competition between the negative Madelung contribution to the surface energy and positive quantum-mechanical con-

tribution (covalent bonding) [54]. Since the attractive forces associated with covalent bonding are local, they have a shorter range than the electrostatic contribution. By measuring the interaction between layers as a function of their separation the hypothesis proposed here can be verified. This can be done experimentally or theoretically, by calculating the crystal energy dependence on the separation between the layers. When the interlayer separation is increased by a small value, the attraction between layers should change over to electrostatic repulsion. At larger distances, the electrostatic repulsion between the layers is suppressed as all the layers are electroneutral, and at sufficient separation they can be treated as consisting of alternating uniformly charged planes. In such a geometry, the electrostatic forces are exactly compensated.

Appendix C

Semiclassical Theory of Thermoelectricity

C.1 Phenomenological Treatment of Transport

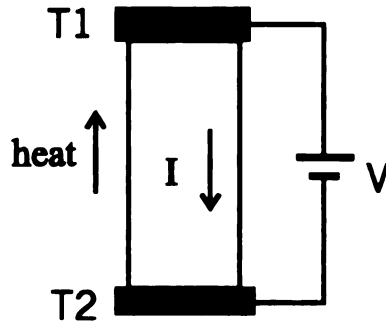


Figure 33: Schematic of a setup for the analysis of thermoelectric phenomena. The direction of the heat flow is shown for n-type material, dominated by the Peltier contribution.

Consider a setup shown in Figure (33). A temperature difference $\Delta T = T_1 - T_2$ is maintained across a slab of a material under study, and a voltage V is also applied between the ends. The temperature and potential gradients induce carrier flow in the system that we would like to express in terms of the material parameters. Phenomenologically, the transport in the setup shown in Figure 33 can be described

by a system of equations

$$\begin{cases} \mathbf{j} = \sigma[\mathbf{E} - \Delta T] \\ \mathbf{j}^Q = \pi\mathbf{j} - \kappa\Delta T \end{cases} \quad (\text{C.1})$$

where \mathbf{j} is electrical current density, \mathbf{j}^Q is thermal current density, σ is electrical conductivity, \mathbf{E} is electrical field, π is Peltier coefficient, κ is thermal conductivity. The thermodynamic relationship $\pi = T\alpha$ was first established by William Thomson (Lord Kelvin) in mid-1800s. Assuming that the spatial variation of the variables is slow compared to the Fermi wave-length, we will apply a standard semi-classical approach to express the phenomenological transport coefficients in terms of the electronic properties of the system. This approximation is not always justified for semiconductors, where the carrier density is low. The typical carrier concentration in narrow gap semiconductors used in thermoelectric applications is $n \approx 10^{19} \text{cm}^{-3}$. In a free electron approximation this yields

$$\lambda_F \approx \frac{2r_s}{a_0} \text{\AA} \approx 100 \text{\AA}, \quad (\text{C.2})$$

where a_0 is the Bohr radius, $r_s \approx n^{-1/3}$. The Fermi wavelength is almost two orders of magnitude larger than in typical metals, setting the scale for the allowed by the semiclassical model variations of the external parameters.

C.2 Semiclassical Transport Equations

The description of transport in semiclassical formalism employs nonequilibrium distribution function $f(\mathbf{r}, \mathbf{k}, t)$ defined as the density of electrons in the semiclassical phase space. In equilibrium

$$f_0(\mathbf{r}, \mathbf{k}, t) \equiv f(E(\mathbf{k})) = \frac{1}{e^{(E-\mu)/kT} + 1}. \quad (\text{C.3})$$

Here and below μ is understood as "internal" chemical potential, i.e. counted, for example, from the bottom of the conduction band. By this definition, chemical potential variation in metals is negligible due to the condition of electroneutrality.

On the other hand, the total chemical potential variation (measured from the vacuum level) can be large due to the electric field.

In a stationary state the distribution function can be determined from the linearized semiclassical Boltzmann equation [30]

$$-\frac{\partial f_0}{\partial E} \mathbf{v} \left[\frac{E - \mu}{T} \nabla T + \nabla \mu + e \mathbf{E} \right] = -\mathbf{v} \frac{\partial g}{\partial \mathbf{r}} + \frac{e}{\hbar} \left(\frac{\mathbf{v}}{c} \times \mathbf{B} \right) \frac{\partial g}{\partial \mathbf{k}} + I_{col}, \quad (\text{C.4})$$

where $\mathbf{v} = \frac{1}{\hbar} \frac{\partial E(\mathbf{k})}{\partial \mathbf{k}}$, $g(\mathbf{r}, \mathbf{k}, t) = f(\mathbf{r}, \mathbf{k}) - f_0(\mathbf{r}, \mathbf{k})$, \mathbf{B} is magnetic field (enters the equation only in second order), I_{col} is a collision integral describing electron scattering. Usually a relaxation-time approximation is made

$$I_{col} = -\frac{f - f_0}{\tau} = -\frac{g}{\tau}, \quad (\text{C.5})$$

which is a mathematical expression for the assumption that the distribution of electrons emerging from collisions is independent of that just prior to collisions. τ is the relaxation time, which can be dependent on \mathbf{k} , \mathbf{r} . If the distribution $g(\mathbf{r}, \mathbf{k})$ is known, the currents are determined by

$$\begin{cases} \mathbf{j} = -e \int \frac{d^3 \mathbf{k}}{4\pi^3} \mathbf{v}(\mathbf{k}) g(\mathbf{k}) \\ \mathbf{j}^Q = \int \frac{d^3 \mathbf{k}}{4\pi^3} (E(\mathbf{k}) - \mu) \mathbf{v}(\mathbf{k}) g(\mathbf{k}). \end{cases} \quad (\text{C.6})$$

Equation (C.4) used to find $g(\mathbf{r}, \mathbf{k})$ leads to the following expressions for the transport coefficients

$$\sigma = e^2 K^{(0)} \quad (\text{C.7})$$

$$\kappa = \frac{K^{(2)}}{T} - \frac{K^{(1)}(K^{(0)})^{-1}K^{(1)}}{T} \quad (\text{C.8})$$

$$\pi = T\alpha = -\frac{1}{e}(K^{(0)})^{-1}K^{(1)} \quad (\text{C.9})$$

$$(\text{C.10})$$

where

$$K_{ij}^{(n)} = \int \frac{d^3 \mathbf{k}}{4\pi^3} \left(-\frac{\partial f_0}{\partial E} \right) \tau (E - \mu)^n v_i v_j. \quad (\text{C.11})$$

C.3 Figure of Merit

In setup shown in Figure 33 the heat absorbed or dissipated at each end of the slab has contributions from both the Peltier term in equation (C.1), and Joule heating in the slab, which can be expressed as

$$\frac{dj^Q}{dx} = \rho J^2 = \frac{d}{dx} [\pi j - \kappa \nabla T]. \quad (\text{C.12})$$

Integrating this equation, we get [64]

$$\begin{aligned} j^Q(0) &= -\frac{\kappa}{L}(T_2 - T_1) + \pi j - \rho j^2 L/2 \\ j^Q(L) &= -\frac{\kappa}{L}(T_2 - T_1) + \pi j + \rho j^2 L/2, \end{aligned} \quad (\text{C.13})$$

where "0" and "L" are coordinates of the ends of the slab in Figure 33. From Equations (C.13) it follows that the heat absorbed(dissipated) on either end of the slab is a sum three contributions: heat conductivity (first term in (C.13)), thermoelectric heat current (second term), and Joule heat. The Joule heat is evenly dissipated at both ends of the slab.

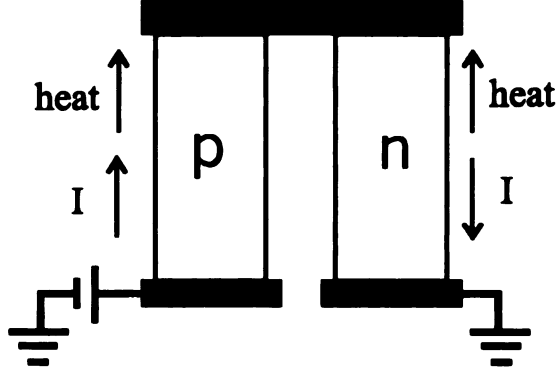


Figure 34: Schematic of a thermoelectric refrigerator setup.

Consider now a setup shown in Figure 34. Two oppositely doped semiconductors are joined on one end, and electric potential difference is applied across the free ends. This setup is used in thermoelectrical applications to minimize the thermal conduction through the connecting wires, which would diminish the thermoelectric

efficiency in a setup such as shown in Figure 33. From Equations (C.13))

$$j_{tot}^Q(0) = (\pi_n - \pi_p)j - (\kappa_n + \kappa_p)\frac{T_2 - T_1}{L} - j^2\left(\frac{1}{\sigma_n} + \frac{1}{\sigma_p}\right)/2. \quad (C.14)$$

The maximal temperature difference sustained by the refrigerator can be found by optimizing this expression with respect to the electrical current and setting it to zero [65]

$$(T_2 - T_1)_{max} = L \frac{(\pi_n - \pi_p)^2}{2\left(\frac{1}{\sigma_n} + \frac{1}{\sigma_p}\right)(\kappa_n + \kappa_p)}. \quad (C.15)$$

Under these conditions the thermoelectric heat absorption is compensated at the cold end by the Joule heat dissipation and thermal conductivity. It is usually assumed for simplicity, that $\pi_n = -\pi_p$, and the value (C.15) is treated as the property of either material in the junction. A dimensionless figure of merit is then defined for a material as

$$ZT = \frac{\pi(T)^2\sigma}{T\kappa} = \frac{T\alpha^2\sigma}{\kappa}. \quad (C.16)$$

Materials with high figure of merit can sustain higher temperature difference in a thermoelectric refrigerator. Formula C.16 applied to metals yields very small figure of merit for metals, where the thermal and electrical conductivities are connected by the Wiedemann-Franz law

$$\kappa = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 T\sigma, \quad (C.17)$$

and Seebeck coefficient, which is related to the electron-hole assymetry, is very small. On the other hand, in semiconductors the Seebeck coefficient can be large, resulting in high thermoelectric performance.

The dimensionless figure of merit (C.16) also enters the expression for the efficiency of the thermoelectric device, defined as the ratio of the heat removed from the cold end to the power produced by the external source. Using expressions (C.1) and (C.14),

$$\eta = \frac{j^Q(0)}{jV} = \frac{\pi j - j^2\rho L/2 - \kappa\Delta T}{\alpha j\Delta T + j^2\rho L}. \quad (C.18)$$

Maximizing (C.14) with respect to j , we get

$$\begin{aligned} j_\eta &= \frac{\alpha \Delta T}{\rho [(1+ZT_m)^{(1/2)}-1]} \\ \eta_{max} &= \frac{T_1}{\Delta T} \frac{[(1+ZT_m)^{1/2}-T_2/T_1]}{(1+ZT_m)^{1/2}+1}, \end{aligned} \quad (C.19)$$

where $T_m = (T_1 + T_2)/2$. From this formula it is clear that the thermoelectric material performance approaches the performance of an ideal Carnot machine $\eta_c = \frac{T_1}{\Delta T}$ as $ZT \rightarrow \infty$.

C.4 A Single Parabolic Band Approximation

It has been shown in Section C.3 that the thermoelectric performance of a material can be characterized by a single parameter (C.16). Usually the ZT value is expressed in terms of the basic parameters of the electronic structure within the framework of the effective mass approach. The transport currents are first assumed to be carried by a single band with effective masses m_1, m_2, m_3 characterizing the dispersion in the band. This assumption allows to simplify the integrals (C.11). It is, however, poorly justified for Bi_2Te_3 , due to the multivalley nature of the band extrema with highly anisotropic dispersion, and presence of several active conduction and valence bands [25, 54].

In the effective mass approximation, K -matrices are diagonal in the coordinates diagonalizing effective mass tensor. Introducing reduced chemical potential $\zeta = \mu/kT$, we can transform (C.11) to

$$K_i^{(n)} = \frac{\sqrt{m_1 m_2 m_3} 2^{3/2} \tau (kT)^{n+3/2}}{3m_i \hbar^3} \int_0^\infty dx \left(3/2 + \frac{nx}{x-\zeta}\right) x^{1/2} (x-\zeta)^n f_0 \quad (C.20)$$

with the integral dependent only on ζ . Introducing the Fermi-Dirac functions F_i

$$F_i = F_i(\zeta) = \int_0^\infty \frac{x^i dx}{e^{x-\zeta} + 1}, \quad (C.21)$$

we can rewrite all the transport coefficients

$$\sigma_i = e^2 K^{(0)} = 2^{3/2} (kT)^{7/2} \tau \frac{\sqrt{m_1 m_2 m_3}}{m_i} F_{1/2} \quad (C.22)$$

$$\alpha_i = -\frac{1}{eT}(K^{(0)})^{-1}K^{(1)} = -\frac{k}{e}\left(\frac{5F_{3/2}}{3F_{1/2}} - \zeta\right) \quad (\text{C.23})$$

$$\kappa_i = \frac{k^2T}{e^2}\sigma_i \left[\frac{7F_{5/2}}{3F_{1/2}} - \left(\frac{5F_{3/2}}{3F_{1/2}}\right)^2 \right]. \quad (\text{C.24})$$

Combing the factors together, we get for the figure of merit

$$ZT = \frac{\left(\frac{5F_{3/2}}{3F_{1/2}} - \zeta\right)^2}{\frac{7F_{5/2}}{3F_{1/2}} - \left(\frac{5F_{3/2}}{3F_{1/2}}\right)^2}. \quad (\text{C.25})$$

As the reduced chemical potential is decreased below the bottom of the conduction band (or increased above the valence band maximum, in case of hole conduction), the electronic system eventually becomes nondegenerate. In this limit, $F_\alpha \approx \alpha F_{\alpha-1}$, so expression C.25 can be approximated by

$$ZT = \frac{2}{5} \left(\frac{5}{2} - \zeta\right)^2. \quad (\text{C.26})$$

This factor diverges as $\zeta \rightarrow -\infty$. In the actual systems this does not happen mainly because of the lattice contribution to the heat conductivity and the presence of the second band limiting the variation of ζ before Seebeck coefficient α is significantly reduced due to the comparable compensating electron and hole transport currents. Replacing in formula (C.16) $\kappa \rightarrow \kappa_{el} + \kappa_{ph}$ to account for the lattice thermal conductivity, we get [66]

$$ZT_i = \frac{\frac{3}{2} \left(\frac{5F_{3/2}}{3F_{1/2}} - \zeta\right)^2}{\frac{1}{B_i} + \frac{7F_{5/2}}{2F_{1/2}} - \frac{25F_{3/2}^2}{6F_{1/2}^2}}, \quad (\text{C.27})$$

where

$$B_i = \frac{1}{3\pi^2} \left(\frac{2kT}{\hbar^2}\right)^{3/2} \frac{(m_1 m_2 m_3)^{1/2}}{m_i} \frac{k^2 T N \tau}{\kappa_{ph}}. \quad (\text{C.28})$$

Here the degeneracy of the band extremum N is included, which has been suggested to play an important role for the thermoelectric properties of Bi_2Te_3 [25], where $N = 6$. m_i is the effective mass in the direction of transport. In formula (C.28) we have separated the effect of material properties, which are present only in C , and the effect of doping or carrier concentration, contributing to the Fermi-Dirac integrals.

The doping is easily optimized for given material properties. It can be shown that for the optimized doping the figure of merit diverges with C [68]. Therefore, for optimizing (C.28) it is beneficial to use anisotropic materials with high average mass $\sqrt{m_1 m_2 m_3}$, low m_i , and high degeneracy of the band extremum N . In polycrystalline samples high anisotropy of the effective masses might still prove beneficial due to the percolation-type network formation. It is also clear that the small value of the lattice heat conductivity is needed. Typically the lattice conductivity is lower if phonons are heavy with low Debye temperature. The high rate of umklapp phonon-phonon scattering in this case reduces the thermal conductivity.

C.5 Beyond a Single Parabolic Band Approximation

The one-band approximation used in the previous discussion is not valid for the narrow gap semiconductors. The gap is of the same order of magnitude as the temperature, therefore both conduction and valence bands participate in transport. The presence of the second band suppresses the Seebeck coefficient. From (C.7)

$$\alpha_{tot} = -\frac{1}{eT} \frac{K_e^{(1)} + K_h^{(1)}}{K_e^{(0)} + K_h^{(0)}} = \frac{\alpha_e \sigma_e + \alpha_h \sigma_h}{\sigma_e + \sigma_h}. \quad (\text{C.29})$$

Since electron and hole contributions have opposite signs, the Seebeck coefficient is reduced by the proximity of the second band

$$\kappa = \kappa_e + \kappa_h + \frac{\kappa_e \kappa_h}{\kappa_e + \kappa_h} (\alpha_e - \alpha_h)^2 T. \quad (\text{C.30})$$

The physical picture of this effect is a drift of thermally excited at the hot end electron-hole pairs to the cold end with subsequent recombination. This process contributes to the thermal conductivity, but does not affect the electrical conductivity.

To explain the decrease of the conductivity in heavily doped Bi_2Te_3 , it has been

suggested [25, 54] that as doping is increased, a second band with a lower mobility is activated. As the electrons get scattered into the second band, the average mobility is reduced by the presence of the second band.

The phonon drag effect is yet another complication not accounted for in formula (C.27); Drift of charge carriers in the presence of potential gradient produces phonon current as a result of electron-phonon scattering. Similarly, phonon drift in the presence of temperature gradient induces charge carrier current due to electron-phonon scattering. Phonon-drag effect results in the experimental thermopower values in some semiconductors more than an order of magnitude larger than the predicted values based on the effective mass semiclassical theory [76]. Strong electron-phonon interaction likely enhances phonon-drag effect in Bi_2Te_3 . We do not discuss this effect in detail here. The phonon-drag effects in thermoelectric semiconductors need further theoretical studies.

Bibliography

- [1] M. H. Devoret and R. J. Schoelkopf, *Nature*, **406**, 1039 (2000).
- [2] B. E. Kane, *Nature*, **393**, 133 (1998).
- [3] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, *Appl. Phys. Lett.*, **40**, 178 (1982); *Phys. Rev. Lett.*, **49**, 57 (1982); *Phys. Rev. Lett.*, **50**, 120 (1983).
- [4] M. Tsukada, K. Kobayashi and S. Ohnishi, *J. Vac. Sci. Technol. A*, **8**(1), 160 (1990).
- [5] J. Tersoff, *Phys. Rev. Lett.*, **57**, 440 (1986).
- [6] J. Tersoff and D. R. Hamann, *Phys. Rev. Lett.*, **50**, 1998 (1983); J. Tersoff and D. R. Hamann, *Phys. Rev.*, **B 31**, 805 (1985).
- [7] G. Binnig, N. Garcia, H. Rohrer, J. M. Soler, and F. Flores, *Phys. Rev.*, **B 30**, 4816 (1984).
- [8] J. Tersoff and N. D. Lang, in *Scanning Tunneling Microscopy*, Methods in Experimental Physics Vol. 27 (Academic Press, New York, 1993).
- [9] J. A. Stroscio and R. M. Feenstra, in *Scanning Tunneling Microscopy*, Methods in Experimental Physics Vol. 27 (Academic Press, New York, 1993).
- [10] S. H. Tessmer, P. I. Glicofridis, L. S. Levitov, R. C. Ashoori and M. R. Melloch, *Nature*, **392**, 51 (1998).

- [11] S. Urazhdin et al., submitted to *Physica C*.
- [12] Ivar Giaever, *Phys. Rev. Lett.*, **5**, 147,464 (1960)
- [13] J. Frohn, J. F. Wolf, K. Besocke, and M. Teske, *Rev. Sci. Instrum.*, **60**, 1200 (1989).
- [14] See, for example, Chr. Wittneven *et al.*, *Rev. Sci. Instrum*, **68**, 3806 (1997).
- [15] Purchased from RHK Technology Inc., 1750 West Hamlin Rd., Rochester Hills, MI 48309.
- [16] D. Lanterman, Summer REU project at MSU, 2000.
- [17] A. Lee, *Rev. Sci. Instrum.*, **64(8)**, 2373 (1993).
- [18] B. Michel, L. Novotny and U. Durig, *Ultramicroscopy*, **42-44**, 1647 (1992).
- [19] H. Birk, K.Oostveen and C.Shonenberger, *Rev. Sci. Instrum.*, **67(8)**, 2977 (1996).
- [20] S.M. Rytov, *Introduction to Statistical Radiophysics*, 2nd ed., ("Nauka", Moscow, 1976) (in russian).
- [21] Proceedings of the Cold Electronics Workshop, October 4-5, 1983, Pasadena, California, ed. E.Tward, R.Kirschman.
- [22] Fujitsu Microelectronics, Inc., San Jose, CA.
- [23] Eltec Instruments Inc., Daytona Beach, FL.
- [24] R.C. Ashoori, H.L. Stormer, J.S. Weiner, L.N. Pfeiffer, S.J. Pearton, K.W. Baldwin and K.W. West, *Physica*, **B 189**, 117 (1993).
- [25] D. R. Lovett, *Semimetals and Narrow-Bandgap Semiconductors* (Pion Limited, London, UK, 1977).

- [26] G. A. Thomas, D. H. Rapke, R. B. Van Dover, L. F. Mattheis, W. A. Surden, L. F. Schneemaper, and J. V. Waszczak, *Phys. Rev.*, **B 46**, , (1)553 (1992).
- [27] S. K. Mishra, S. Satpathy, and O. Jepsen, *J. Phys.: Cond. Matt.*, **9**, 461 (1997).
- [28] Samples were grown by Theodora Kyratsi at M. G. Kanatzidis lab, Department of Chemistry, MSU.
- [29] S. H. Tessmer, D. J. Van Harlingen, and J. W. Lyding, *Rev. Sci. Instrum.*, **65**, 2855 (1994); S. H. Tessmer, D. J. Van Harlingen, and J. W. Lyding, *Phys. Rev. Lett.*, **70**, 3135 (1993).
- [30] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, (Sounders College Publishing, 1976).
- [31] S. Urazhdin, D. Bilc, S. H. Tessmer, S. D. Mahanti, Theodora Kyratsi and M. G. Kanatzidis, submitted to *Phys. Rev. Lett.*
- [32] B. I. Schklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors*, (Nauka, Moscow, USSR, 1979), *in russian*.
- [33] X. de la Broïse *et al.*, *Phys. Rev.*, **B 61**, 2138 (2000).
- [34] D. Lawunmi and M. C. Payne, *J. Phys.: Condens. Matter*, **7**, 5155 (1995)
- [35] R. M. Feenstra, J. M. Woodall, and G. D. Pettit, *Phys. Rev. Lett.*, **71**, 1176 (1993).
- [36] B. Grandidier, Huajie Chen, R. M. Feenstra, D. T. McInturf, P. W. Juodawlkis and S. E. Ralph, *Appl. Phys. Lett.*, **74**, 1439 (1999);
- [37] P. N. First, J. A. Stroschio, R. A. Dragoset, D. T. Pierce, and R. J. Celotta, *Phys. Rev. Lett.*, **63**, 1416 (1989).

- [38] G. Lengel, R. Wilkins, G. Brown, M. Weimer, J. Gryko and R. E. Allen, Phys. Rev. Lett., **72**, 836 (1994).
- [39] J. F. Zheng, X. Liu, N. Newman, E. R. Weber, D. F. Ogletree, and M. Salmeron, Phys. Rev. Lett., **72**, 1490 (1994).
- [40] M. B. Johnson, O. Albrechtsen, R. M. Feenstra, and H. W. M. Salemink, Appl. Phys. Lett., **63**, 2923 (1993).
- [41] Z. F. Zheng, M. B. Salmeron, E. R. Weber, Appl. Phys. Lett., **64**, 1836 (1994).
- [42] Kuo-Jen Chao, Arthur R. Smith, and Chih-Kang Shih, Phys. Rev., **B 53**, 6935 (1996).
- [43] B. Grandidier, D. Stievenard, J. P. Nys, and X. Wallart, Appl. Phys. Lett., **72**, 2454 (1998).
- [44] Chr. Wittneven, R. Dombrowski, M. Morgenstern, and R. Wiesendanger, Phys. Rev. Lett., **81**, 5616 (1998).
- [45] Jaroslaw Dabrowski and Matthias Scheffler, Phys. Rev., **B 40**, 10391 (1989).
- [46] R. B. Capaz, K. Cho, and J. D. Joannopoulos, Phys. Rev. Lett, **75**, 1811 (1995).
- [47] S. B. Zhang, Phys. Rev., **B 60**, 4462 (1999).
- [48] J. Bardeen, Phys. Rev. Lett., **6**, 57 (1961).
- [49] Performed by D. Bilc at MSU using commercial software wien97, to be published.
- [50] P. Hohenberg and W. Kohn, Phys. Rev., **136**, B864 (1964); W. Kohn and L. J. Sham, Phys. Rev., **140**, A1133 (1965).
- [51] N. D. Lang and W. Kohn, Phys. Rev., **B 1**, 4555 (1970).

- [52] W. A. Harrison, *Electronic Structure and the Properties of Solids* (W. H. Freeman and Co., San Francisco, 1980).
- [53] P. Larson, S. D. Mahanti, and M. G. Kanatzidis, Phys. Rev., **B 61**, 8162 (2000).
- [54] S. J. Youn and A. J. Freeman, Phys. Rev., **B 63**, 085112-1 (2001).
- [55] J. F. Meng, N. V. Chandra Shekar, J. V. Badding, D.-Y. Chung and M. G. Kanatzidis, J. Appl. Phys., **90**, 2836 (2001).
- [56] X. Zhu, S. B. Zhang, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett., **63**, 2112 (1989).
- [57] R. M. Feenstra, Phys. Rev., **B 50**, 4561 (1994).
- [58] W. G. Aulbur, L. Jonsson, and J. W. Wilkins in *Solid State Physics* vol. 54 (Academic Press, San Diego, 2000).
- [59] S. Urazhdin, T. Hogan, and S. Lol, unpublished.
- [60] as follows from LDA calculations performed by D. Bilc at MSU, unpublished.
- [61] S. Hufner, *Photoemission Spectroscopy*, (Springer, New York, 1995).
- [62] M. P. Seah, W. A. Dench, Surf. and Interface Anal., **1**, 2 (1979).
- [63] V. A. Greanya, W. C. Tonjes, Rong Liu, C. G. Olson, D.-Y. Chang, and M. G. Kanatzidis, Phys. Rev., **B 62**, 16425 (2000).
- [64] M. P. Marder, *Condensed matter physics*, (John Wiley, New York, 2000).
- [65] H. J. Goldsmid in *CRC Handbook of Thermoelectrics*, ed. D. M. Rowe, (CRC Press, Inc., 1995).
- [66] L. D. Hicks and M.S. Dresselhaus, Phys. Rev., **B 47**, 12727 (1993); L. D. Hicks and M.S. Dresselhaus, *ibid.*, 16631 (1993).

- [67] H. J. Goldsmid, *Thermoelectric Refrigeration* (Plenum, New York, 1964).
- [68] C. M. Bhandari in *CRC Handbook of Thermoelectrics*, ed. D. M. Rowe, (CRC Press, Inc., 1995).
- [69] M. G. Kanatzidis, *Semiconductors and Semimetals*, **69**, 51 (2001).
- [70] M. G. Kanatzidis, private communications.
- [71] G. Mahan, B. Sales, and J. Sharp, *Physics Today*, **50(3)**, 42 (1997).
- [72] F. J. DiSalvo, *Science*, **285**, 703 (1999).
- [73] G. D. Mahan, in *Solid State Physics* vol. 51 (Academic Press, Orlando, 1998).
- [74] R. G. Wyckoff, *Crystal Structures* Vol. 2 (John Wiley and Sons, New York, 1967).
- [75] David J. Singh, *Plane Waves, Pseudopotentials and the LAPW Method*, (Kluwer Academic Publishers, Boston, 1994).
- [76] Frank J. Blatt, *Physics of Electronic Conduction in Solids*, (McGraw-Hill, New York, 1968).
- [77] S. Urazhdin, S. H. Tessmer, Norman O. Birge, W. K. Neils, and D. J. Van Harlingen, to be submitted to *Physica C*.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02334 8034