THERMOELECTRIC AND THERMOSPINTRONIC TRANSPORT IN
DIRAC MATERIAL-BASED NANOSTRUCTURES

by

Po-Hao Chang

A dissertation submitted to the Faculty of the University of Delaware in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in Physics

Winter 2016

© 2016 Po-Hao Chang
All Rights Reserved
THERMOELECTRIC AND THERMOSPINTRONIC TRANSPORT IN
DIRAC MATERIAL-BASED NANOSTRUCTURES

by

Po-Hao Chang

Approved:

Edmond R. Nowak, Ph.D.
Chair of the Department of Physics

Approved:

George H. Watson, Ph.D.
Dean of the College of Science

Approved:

Ann L. Ardis, Ph.D.
Interim Vice Provost for Graduate and Professional Education
I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: 
Branislav K. Nikolić, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: 
John Clem, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: 
John Q. Xiao, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: 
Predrag Lazic, Ph.D.
Member of dissertation committee
ACKNOWLEDGEMENTS

I would like to express my special gratitude to my advisor Prof. Branislav. K. Nikolić, for all of his invaluable advices on both research as well as being supportive on my career. I would also like to thank my committee members, Prof. Predrag Lazic, Prof. John Clem, Prof. John Q. Xiao for serving as my committee members.

Thanks to Ada (Lin Yuan) for being there for me, to Hao Shen, Wei-Fan Kuan and Shu-Ching Ou who came in the same year as me. We helped each other get through nasty paperworks during the first year. I literally could not do it without you. I would also like to thank all of my friends in UD for their supports and helps. In particular I would like to thank, Dr. Hector Mera for all the helps and encouragements in both research and job hunting. Also I am grateful to Kamal Saha, Farzad Mahfouzi, Taha Salavati-fard, Brian Kelly, Kevin Haughey, Matthew Jaffee, Tim McClory, Matt Civiletti, Tarique Aziz, Son-Hsien Chen, Derya Vural, Heng-Yu Chen, Curtis Walkons, Prachanda Subedi, Seth Meiselman, Prayash Sharma Pyakurel, Yue Pan, Hang Chen and Yunsong Xie and many others for their helps and friendship I would also like to thank the secretaries in Department of Physics and Astronomy specially Mrs. Maura Perkins and Mrs. Debra Morris, for coping with all of my botherings patiently and always being more than helpful. Thanks to the Prof. Edmond R. Nowak who treats us like family.

Softball seasons played a big part of my life in UD. Not only it helped me overcome frustrations from work but enriched my life and expanded my social circle. My sincere thanks go to Ryan Stearrett and John C. Meyer who introduced me to join the softball team and taught me how to play. I would also like to thank James Dalessio, Rob Vander Decker, Travis Lutman and all the other team members in Honey Badgers
who accepted me as as part of the team and helped me improve. I really had a great
time.

    Last but not the least, a special thanks to my lovely family. No words can
express my gratitude enough to my mother, father and brother. For all of the sacrifices
that you have made mean everything to me. Your encouragements and endless spiritual
support throughout my Ph.D. study and all my life are what support me and keep me
going thus far.
TABLE OF CONTENTS

LIST OF FIGURES ....................................................... ix
ABSTRACT ................................................................. xvi
LIST OF ABBREVIATIONS AND SYMBOLS ........................... xix

Chapter

1 INTRODUCTION ............................................................ 1
  1.0.1 Overview ......................................................... 1
  1.0.2 Dirac Materials .................................................. 2
  1.0.3 Lattice and Electron Graphene Structure of Monolayer
       Graphene ......................................................... 4
       1.0.3.1 Graphene Nanoribbons ................................. 7
  1.0.4 Topological Insulators .......................................... 8
       1.0.4.1 GNR-based 2D TI ....................................... 9
       1.0.4.2 3D TI .................................................... 10
  1.1 Thermoelectric Energy Conversion ............................... 10
       1.1.1 Efficiency of Thermoelectric Conversion ............... 13

2 INTRODUCTION TO NON-EQUILIBRIUM GREEN’S
   FUNCTION ............................................................... 16
  2.1 Retarded Green’s Function in Single-particle Quantum Mechanics 16
       2.1.1 Retarded Green’s function as a propagator ............ 18
       2.1.2 Space discretization in 1D case: ....................... 19
       2.1.3 Open quantum system ................................... 20
       2.1.4 Physical observables from one-particle Green’s functions 22
LIST OF FIGURES

1.1 Bulk graphene sheet and GNRs with different edge types highlighted by different colors. ................................................. 4

1.2 (a) Graphene is made of carbon atoms arranged in honeycomb lattice formed by two triangular structure sublattices highlighted with blue and red colors. (b) Reciprocal lattice of graphene where K’ and K are two inequivalent points and both correspond to Dirac points. ........................................ 5

1.3 Energy-momentum dispersion for infinite graphene. The valence and conduction bands of graphene only touch at six discrete points within 1st BZ. These discrete points are often referred to as Dirac points. In low-energy excitations where the energy–momentum dispersion relation close to Dirac points shows linear behavior and described by Dirac-like as Eq. (1.1) and points where bands touch are called Dirac points. ...................................................... 6

1.4 (a) The electronic band structure computed for a supercell of Bi$_2$Se$_3$ film of thickness 5 QLs using ncDFT with pseudo-atomic localized basis functions [159] implemented in ATK [158] package. This is contrasted with the electronic band structure obtained using ncDFT with PW basis set implemented in VASP package. [169] (b) Zero-bias transmission function of Bi$_2$Se$_3$ thin film in the two-terminal geometry of Fig. 6.1, for electrons injected along the Γ–M direction ($k_y = 0$) in the inset of Fig. 6.2(a), computed using NEGF+ncDFT formalism implemented in ATK package. Adapted from Ref. [16]. ............................. 11

1.5 (a) Schematic view of a device attached to hot and cold macroscopic reservoirs that induces the temperature gradient the electrons around hot area diffuse faster than cold ones therefore resulting in charge accumulation on both sides. (b) The difference of Fermi-Dirac distribution functions is anti-symmetric around $E_F$, therefore the optimization of $I_T$ relies on the asymmetry of transmission function. ..................................................... 12
1.6 Consider an N-type cooling system. The heat is pumped from cold side $T_C$ to hot side $T_H$. While the heat is carried to the hot end, part of the flow is canceled out by the heat dissipation and the temperature gradient.

2.1 Two-terminal open system, the leads are semi-infinite and connected smoothly at $\pm \infty$ to reservoirs biased by the chemical potential difference $\mu_L - \mu_R$.

2.2 (a) Exact contour. (b) Keldsh contour.

2.3 In 3D vibration, there are three independent parameters to describe force tensor between each pair of carbons.

3.1 Bandstructures for (a) 20 ZGNR 3NN TB (b) 20 ZGNR (1NN TB) (c) (8,1)-CGNR 3NNTB (d) (8,1)-CGNR 1NN TB. As the results indicate, the subbands around DP show distortion by considering 3NN and therefore gain nonzero velocities to contribute to the transport.

3.2 The phonon thermal conductance as a function of temperature for 8-ZGNR with and without a single nanopore (of diameter $D = 0.59$ nm) computed by coupling NEGF formalism to empirical 4NFC, [74] empirical 5NFC, [75] Brenner EIP, [76] and DFT (using the basis of local DZP orbitals and PBE exchange-correlation functional). The FC matrix in the case of Brenner EIP and DFT methodology was computed using the GPAW package. [77]

3.3 Schematic view of: (a) 20-ZGNR (composed of 20 zigzag chains); and (b) (8,1)-CGNR with chiral angle $\theta = 5.8^\circ$. The size of the nanopores, assumed to be drilled in the GNR interior away from its zigzag or chiral edges, and the distance between them is illustrated by plotting two repeated supercells of each GNR. The length of these GNRs in the actual calculations is set to $L \simeq 1.2 \mu m$ which supports 300 nanopores. Adapted from Ref. [13].
3.4 (Color online) Spatial profiles of local charge currents in (a) 20-ZGNR and (c) (8,1)-CGNR with nanopores for electronic transport close ($E_F = -0.43$ eV) to the DP. The corresponding current profiles over the transverse cross section of nanoribbons are shown in panels (b) and (d) for both infinite homogeneous GNRs and GNRs with nanopores. Note that the sum of bond currents $J_{nm}/V$, which describe charge flow from site $n$ to site $m$ of the honeycomb lattice if hopping $t_{nm} \neq 0$ is non-zero between the two sites, gives the conductance $G = I/V$ ($I$ is the total current in the leads and $V \to 0$ is small bias voltage driving the linear-response transport). Adapted from Ref. [13].

3.5 (Color online) (a) The zero-bias electronic transmission $T_{el}(E)$ for an infinite homogeneous 20-ZGNR or 20-ZGNR with periodic array of identical nanopores shown in Fig. 3.3(a). (b) The thermopower at two different temperatures for finite length 20-ZGNR with nanopores. (c) Zero-bias electronic transmission for an infinite homogeneous (8,1)-CGNR or (8,1)-CGNR with periodic array of identical nanopores shown in Fig. 3.3(b). (d) The thermopower at two different temperatures for finite length (8,1)-CGNR with nanopores. Adapted from Ref. [13].

3.6 (Color online) (a) The phonon transmission function $T_{ph}(\omega)$ and (b) the corresponding phonon thermal conductance $\kappa_{ph}$ for an infinite homogeneous 20-ZGNR or 20-ZGNR with periodic array of identical nanopores shown in Fig. 3.3(a). (c) The phonon transmission function and (d) the corresponding phonon thermal conductance $\kappa_{ph}$ for an infinite homogeneous (8,1)-CGNR or (8,1)-ZGNR with a periodic array of identical nanopores shown in Fig. 3.3(b). Adapted from Ref. [13].

3.7 (Color online) The thermoelectric figure of merit $ZT$ vs. energy at two different temperatures for (a),(b) ZGNR+nanopores and (c),(d) CGNR+nanopores. In panels (a) and (c) we assume a periodic array of identical nanopores, as illustrated in Fig. 3.3, while in panels (b) and (d) the nanopore diameter and the distance between neighboring pores is a uniform random variable. The curves plotted in panels (b) and (d) are computed for a single sample.
4.1 (a) The central region, consisting of GNR (of width $W = 6.16$ nm and length $L \approx 1.35 \, \mu m$ used in our calculations) with indium adatoms and (253 along chosen $L$) nanopores, is attached to hot and cold macroscopic reservoirs. The adatoms shown in the center of randomly chosen $n_{ad} \approx 19\%$ hexagons locally enhance SOC within those hexagons which creates an energy gap $E_G$ and a pair of helical quantum states on each edge channeling flow of electrons of opposite spin in opposite directions. We consider GNRs with: (b) perfect edge of zigzag type; or (c) disordered edge created by removing one or two edge carbon atoms in each GNR supercell. (d) The total density of states, where the gap $E_G \approx 17.3$ meV around the Dirac point at zero energy is filled by contribution from the helical edge states, is insensitive to the randomness of adatom configuration or spatial inhomogeneities. Adapted from Ref. [14]...

4.2 (a) The local DOS within GNR + heavy-adatoms with a single nanopore in its interior at $E_F = 0.001$ eV. (b) The spatial profile of local currents over the transverse cross section of system in (a). The sum of all bond currents $J_{mn} / V_b$, describing [98] charge flow from site $m$ to site $n$ of the honeycomb lattice at bias voltage $V_b$, gives the electrical conductance $G(E_F) = e^2 T_{el}(E_F)/h$ at zero temperature. Due to helical edge quantum states determining the local DOS and local currents, both quantities have non-zero value only around the edges. (c) and (d) The zero-bias electronic transmission function $T_{el}(E)$ for setups in Figs. 4.1(b) and 4.1(c), respectively. The blue and red lines in panels (b)–(d) are obtained in the presence of both heavy adatoms and nanopores, while black dotted line in panels (c) and (d) is computed after nanopores in the two setups from Fig. 4.1 are removed. (e) and (f) The Seebeck coefficient at $T = 40$ K computed by plugging $T_{el}(E)$ from panels (c) and (d) into Eq. (2.76), respectively. Panels (e) and (f) also show the power factor $S^2 G$. Adapted from Ref. [14]...
4.3 **The phononic band structure and thermal conductance.** Solid line in both panels (a) and (b) plots the phonon band structure for $4 \times 4$ supercell of graphene, with one indium adatom per supercell, where SOC is not included. Dotted line in panel (a) plots phonon dispersion for the same supercell of pristine graphene without any adatoms, while dashed line in panel (b) includes both adatoms and SOC. All curves are obtained from first-principles calculations using VASP [170]. (c) and (d) The phononic transmission function $T_{\text{ph}}(\omega)$ for setups in Fig. 4.1(b) and 4.1(c), respectively, where solid lines are based on the dashed line dispersion in panel (b). Dotted lines in panels (c) and (d) plot quantized $T_{\text{ph}}(\omega)$ for an infinite homogeneous ZGNR of the same width and edge shape as in Fig. 4.1, but without any adatoms, nanopores or edge disorder. (e) and (f) Solid and dotted lines plot the phononic thermal conductance, computed by plugging the values of $T_{\text{ph}}(\omega)$ from solid and dotted lines in panels (c) and (d), respectively, into Eq. (2.70). Additional dashed line in panels (e) and (f) plots $\kappa_{\text{ph}}$ for the same setups as in Figs. 4.1(b) and 4.1(c), respectively, but where the presence of heavy adatoms and local SOC induced by them is neglected. Adapted from Ref. [14].

4.4 **The thermoelectric figure of merit.** Dashed line plots $ZT$ at $T = 36$ K for GNR + heavy-adatoms + nanopores setup in Fig. 4.1(b), while solid line takes into account additional edge disorder for setup in Fig. 4.1(c). The inset shows dependence of the peak values of $ZT$ on temperature. Adapted from Ref. [14].

5.1 (Color online) Schematic view of the three-terminal junction where a thin film of Bi$_2$Se$_3$ is attached to two nonmagnetic metallic leads kept at different temperatures $T_1 > T_2$. The SSE voltage signal $V_{\text{ISHE}}$ is measured across the edges of the third nonmagnetic metallic lead attached to the top TI surface a function of its position along the x axis. This lead acts as voltage probe attached to a macroscopic reservoir at potential $V_3$ to ensure zero net charge current through it. We orient quintuple layers of Bi$_2$Se$_3$ at an angle of 45° with respect to the direction (x axis) of electron propagation on the TI surface that is orthogonal to these layers. Adapted from Ref. [15].
5.2 (Color online) The equilibrium expectation values of the spin operator for surface-state electrons on the Bi and Se sublattices as a function of the in-plane momentum \( \mathbf{p} = (p_x, p_y) \). The Bi\(_2\)Se\(_3\) crystal is assumed to fill the half-space \( z < 0 \), so that its infinite surface in the \( xy \)-plane is orthogonal to its QLs that are also oriented perpendicularly to the unit vector \( \mathbf{n} = (e_x + e_y)/\sqrt{2} \), as illustrated in Fig. 5.1. Adapted from Ref. [15].

5.3 (Color online) The components \((P_x, P_y, P_z)\) and the magnitude \(|\mathbf{P}|\) of the spin-polarization vector of charge current outflowing into lead 2, after unpolarized charge current is injected from lead 1 into the two-terminal version of junction in Fig. 5.1. These are plotted vs. (a) the Fermi energy \( E_F \), or (b) the length \( L_x \) of the TI film. The linear response charge conductance of the same two-terminal junction vs. (c) the Fermi energy, or (d) the inverse length \( 1/L_x \). Adapted from Ref. [15].

5.4 (Color online) (a) The SSE signal \( V_{ISHE}/(L_y \Delta T) \) as a function of the position of N lead 3 in Fig. 5.1 displaced between the contacts with N leads 1 and 2 sandwiching TI samples of length \( L_x = 100a \) or \( L_x = 200a \). (b) Dependence of \( |V_{ISHE}|_{\text{max}}/(L_y \Delta T) \) on the Fermi energy \( E_F \). (c) Voltage \( V_3 L_y/\Delta T \) applied to N lead 3 in order to bring the net charge current \( I_3 \equiv 0 \). (d) Conventional charge and spin-dependent Seebeck coefficients vs. \( E_F \) for the two-terminal version of junction in Fig. 5.1. Adapted from Ref. [15].

6.1 Schematic view of a two-terminal setup where a thin film of Bi\(_2\)Se\(_3\) is attached to two macroscopic reservoirs biased by the electrochemical potential difference \( eV_b = \mu_L - \mu_R \). We assume that clean Bi\(_2\)Se\(_3\) film is infinite along the \( x \)-axis (i.e., the direction of transport) and the \( y \)-axis, while its thickness along the \( z \)-axis is chosen as 5 QLs. The shaded cell defines thickness \( d_x \approx 5 \, \text{Å} \) over which \( S(\mathbf{r}) \) from Fig. 6.2 is averaged and then plotted in Fig. 6.3, together with LDOS, over the corresponding section within the \( yz \)-plane defined by the cell. Adapted from Ref. [16].
6.2 (a) The arrangement of Bi and Se atoms within a supercell of Bi$_2$Se$_3$ thin film of thickness 5 QLs. The inset in panel (a) shows BZ in the $k_x$-$k_y$ plane at $k_z = 0$ with special $k$-points $\Gamma$, $M$, and $K$ marked. (b) The vector field of nonequilibrium $S(\mathbf{r})$ within selected planes shown in (a), generated by injection of unpolarized charge current along the $x$-axis (see also Fig. 6.1). The planes 1 and 3 correspond to the top and bottom metallic surfaces of Bi$_2$Se$_3$ thin film, while plane 2 resides in the bulk at a distance $d \approx 0.164$ nm away from plane 1. (c) The vector fields in (b) projected onto each of the selected planes in (a). The real space grid of $\mathbf{r}$-points in panels (b) and (c) has spacing $\approx 0.4$ Å. Adapted from Ref. [16].

6.3 (a),(b) The spatial profile of LDOS at $E_F$ over the cross section of the cell denoted in Fig. 6.1. (c),(d) The spatial profile of the components of $S(\mathbf{r})$, obtained by averaging its texture plotted in Fig. 6.2, over the cell of Bi$_2$Se$_3$ thin film marked in Fig. 6.1. The direction of injected charge current for the results in panels (a) and (c), or the results in panels (b) and (d), is denoted in panel (e) relative to the orientation of the lattice of Bi and Se atoms. The bottom surface of Bi$_2$Se$_3$ is located at $z = 0$ nm, and the top TI surface is located at $z \approx 4.56$ nm. Adapted from Ref. [16].
ABSTRACT

The growing need for power due to the rapid developments of the technologies has urged both engineers and scientists to study more sustainable types of energy. On the other hand, the improvement of our abilities although enable us, for example, to double the number of transistors in a dense integrated circuit approximately every two years (Moore’s law), comes with side effect due to overheating. Taking advantage of thermoelectric effect has thus become one of the obvious solutions for the problems. But due to the poor efficiency of electricity-heat conversion, there are still challenges to be overcome in order to fully utilize the idea.

In the past few years, the realization of graphene along with the discoveries of topological insulators (TI) which are both considered as Dirac material (DM) have offer alternative routs for improving the energy conversion efficiency through different approaches as well as novel quantum effects of materials themselves for investigation.

The aim of this thesis is to present contributions to improving the efficiency of thermoelectric conversion as well as analyzing spin transport phenomena that occur in nano-devices. This thesis spans the areas of thermoelectric (TE) effect, spin-Seebeck effect (SSE) and the spin transport on the 3D topological insulator (TI). The different methods have been applied ranging from tight-binding (TB) approximation to density function theory (DFT) combined with non-equilibrium function (NEGF) techniques.

Outline of the Thesis

The general background of the Dirac materials (DMs) used to improve the efficiency and the basic concept of TE effect will be introduced in Chapter 1. The basic tools based on NEGF required to calculate both thermoelectric (TE) and spin transport will be presented in Chapter 2. The applications and the results are covered in the subsequent chapters as follows:
1. **EDGE CURRENTS AND NANOPORE ARRAYS IN ZIGZAG AND CHIRAL GRAPHENE NANORIBBONS AS A ROUTE TOWARD HIGH-ZT THERMOELECTRICS:**

We analyze electronic and phononic quantum transport in zigzag or chiral graphene nanoribbons (GNRs) perforated with an array of nanopores. Since local charge current profiles in these GNRs are peaked around their edges, drilling nanopores in their interior does not affect edge charge currents while drastically reducing phonon heat current in sufficiently long wires. The combination of these two effects can yield highly efficient thermoelectric devices with maximum $ZT \simeq 5$, at both liquid nitrogen and room temperature achieved, in $\sim 1\mu m$ long zigzag GNRs with nanopores of variable diameter and spacing between them. Our analysis is based on the nonequilibrium Green function formalism combined with the $\pi$-orbital tight-binding Hamiltonian with up to third nearest-neighbor hopping for electronic subsystem, or with empirical fifth-nearest-neighbor force-constant (5NNFC) model for phononic subsystem. Additionally, we demonstrate that different empirical FC models typically overestimate the phonon conductance when compared to first-principles results.

2. **GIANT THERMOELECTRIC EFFECT IN GRAPHENE-BASED TOPOLOGICAL INSULATORS WITH NANOPORES:**

Designing thermoelectric materials with high figure of merit $ZT = S^2GT/\kappa$ requires fulfilling three often irreconcilable conditions, i.e., the high electrical conductance $G$, small thermal conductance $\kappa$ and high Seebeck coefficient $S$. Nanostructuring is one of the promising ways to achieve this goal as it can substantially suppress lattice contribution to $\kappa$. However, it may also unfavorably influence the electronic transport in an uncontrollable way. Here we theoretically demonstrate that this issue can be ideally solved by fabricating graphene nanoribbons with heavy adatoms and nanopores. These systems, acting as a two-dimensional topological insulator with robust helical edge states carrying electrical current, yield a highly optimized power factor $S^2G$ per helical conducting channel. Concurrently, their array of nanopores impedes the lattice thermal conduction through the bulk. Using quantum transport simulations coupled with first-principles electronic and phononic band structure calculations, the thermoelectric figure of merit is found to reach its maximum $ZT \simeq 11$ at $T = 40$ K. This paves a way to design high-$ZT$ materials by exploiting the nontrivial topology of electronic states through nanostructuring.

3. **Spin-Seebeck effect on the surface of topological insulator due to nonequilibrium spin-polarization parallel to the direction of thermally driven electronic transport:**
We study the transverse spin-Seebeck effect (SSE) on the surface of a three-dimensional topological insulator (TI) thin film, such as Bi$_2$Se$_3$, which is sandwiched between two normal metal leads. The temperature bias $\Delta T$ applied between the leads generates surface charge current which becomes spin-polarized due to strong spin-orbit coupling on the TI surface, with polarization vector acquiring a component $P_x \simeq 60\%$ parallel to the direction of transport. When the third nonmagnetic voltage probe is attached to the portion of the TI surface across its width $L_y$, pure spin current will be injected into the probe where the inverse spin Hall effect (ISHE) converts it into a voltage signal $|V_{ISHE}|^{max}/\Delta T \simeq 2.5 \mu V/K$ (assuming the SH angle of Pt voltage probe and $L_y = 1$ mm). The existence of predicted nonequilibrium spin-polarization parallel to the direction of electronic transport and the corresponding electron-driven SSE crucially relies on orienting quintuple layers (QLs) of Bi$_2$Se$_3$ orthogonal to the TI surface and tilted by 45$^\circ$ with respect to the direction of transport. Our analysis is based on the Landauer-Büttiker-type formula for spin currents in the leads of a multi-terminal quantum-coherent junction, which is constructed using nonequilibrium Green function formalism within which we show how to take into account arbitrary orientation of QLs via the self-energy describing coupling between semi-infinite normal metal leads and TI.

4. Nonequilibrium spin density around surfaces of current-carrying topological insulator thin film: A first-principles quantum transport study: We study microscopic details, over $\lesssim 1$ Å length scale, of the nonequilibrium spin density $S(\mathbf{r})$ driven by unpolarized charge current injection into a ballistic thin film of Bi$_2$Se$_3$ as a prototypical topological insulator (TI) material. We find a complex nonequilibrium spin texture on the surfaces of TI, as well as within $\simeq 2$ nm thick layer near the surfaces driven by penetration of evanescent wavefunctions from the metallic surfaces into the bulk. Averaging the spin texture over few Å yields large nonzero component of $S(\mathbf{r})$ in the direction transverse to the current flow. In addition, we find an order of magnitude smaller out-of-plane component when the direction of injected current with respect to Bi and Se atoms probes the largest hexagonal warping of the Dirac-cone dispersion on its surface. Our analysis is based on an extension of the nonequilibrium Green functions combined with density functional theory to situations involving noncollinear spins and spin-orbit coupling. We also demonstrate how DFT calculations with properly optimized local orbital basis set can precisely match putatively more accurate calculations with plane-wave basis set for the supercell of Bi$_2$Se$_3$. 

xviii
LIST OF ABBREVIATIONS AND SYMBOLS

2DEG  two-dimensional electron gas
2D    two-dimensional
3D    three-dimensional
AGNR  armchair graphene nanoribbon
ARPES angle resolved photoemission spectroscopy
BZ    Brillouin zone
CGNR  chiral graphene nanoribbon
DFT   density functional theory
DM    Dirac material
DP    Dirac point
FDD   Fermi-Dirac distribution
Gr    graphene
GF    Green’s function
GNR   graphene nanoribbon
TE    thermoelectric
TB    tight-binding
ncDFT noncollinear density functional theory
NEGF  non-equilibrium Green’s function
QL    quintuple layer
RHS   right hand side
SHE   spin Hall effect
SOC   spin-orbital coupling
SSE   spin Seebeck effect
TI    topological insulator
TRS   time-reversal symmetry
WF    wavefunction
ZGNR  zigzag graphene nanoribbon
Tr(A) trace of matrix A
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>operator (matrix) defined in a single particle Hilbert space.</td>
</tr>
<tr>
<td>$G^r$</td>
<td>retarded Green function.</td>
</tr>
<tr>
<td>$G^a$</td>
<td>advanced Green function.</td>
</tr>
<tr>
<td>$G^&lt;$</td>
<td>lesser Green function.</td>
</tr>
<tr>
<td>$G^&gt;$</td>
<td>greater Green function.</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION

1.0.1 Overview

In order to meet the growing demand for energy, the need for finding more sustainable type of energy urged a great deal of attention in studying and improving the performance of thermoelectrics [1]. It has been well known that large portion of energy produced turns into waste heat. Therefore, finding out a way to utilizing thermoelectric effect to recycle waste heat effectively is crucial. But due to the poor energy conversion efficiency, the usage of thermoelectrics is not widespread. Two decades ago, Hicks and Dresselhaus [2] proposed the idea of improving thermoelectrics by looking into the electronic properties of low-dimensional materials. It has ever since become the major inspiration for the direction in the recent developments. Since in the low-dimension structures, the quantum effects become significant such as quantum confinement, different quantities that control the efficiency can be improved independently. The idea is further extended by the recent theoretical study by Humphrey and Linke [3] which predicted the possibility of achieving the efficiency of energy conversion comparable to the Carnot limit at a single energy level channel between hot and cold reservoirs. Despite the prediction is based on ideal condition, it points out a clear direction for further studies. One of the biggest advantage of thermoelectrics is that, besides its sustainability, it contains no moving part unlike traditional type of engine or refrigerator. It could be particularly useful for local cooling systems or power generators that recycle waste heat from small part of devices like CPU chips. More recently, the spin Seebeck effect (SSE) where spin current or spin accumulation induced by a temperature gradient applied across the device [12], has been observed in a wide range of materials.
The recent discoveries of graphene (Gr) and topological insulators (TIs), which are both considered as Dirac material (DM) [8], due to their low-energy excitations behaving as massless Dirac fermions, offer wide range of possibilities for improving the efficiency through nanostructuring. Graphene as monolayer of graphite, is a pure two-dimensional material, composed of carbon atoms. On the other hand, in the low energy excitation (around Fermi energy $E_F$), both three-dimensional (3D) and two-dimensional (2D) TI only conduct with spin-polarized electrons at the boundaries (i.e., surfaces and edges receptively). With their low dimensionality and special electronic structures, separated conduction channels through edge (surface) states, improving power factors while suppressing thermal transport becomes possible [113, 10, 11].

The first part of the thesis is focused on the optimizations of the thermoelectrics based on graphene nanoribbons (GNRs) [13]. However, manufacturing of GNRs with perfectly sharp edges has posed a great challenge from experimental perspective. In the following chapter, we extend the interested materials from normal GNRs to graphene-based topological insulator whose edge states are protected against non-magnetic edge disorder and backscattering is suppressed. Therefore perfect edges for GNRs are no longer necessary [14]. In the second part we move our focus to the studying of phenomena of spin currents on the surfaces of 3D TI [16] and its application for spin-Seebeck (SSE) [15].

1.0.2 Dirac Materials

The DMs refer to crystalline lattices whose low-energy excitations are described by the relativistic Dirac or Weyl Hamiltonians. These materials are usually semiconductors with strongly coupled bands near a level-crossing. Dirac equation was originally used in particle physics to describe the particle behavior that satisfies the relativistic linear energy-momentum dispersion, until recently it has also been adapted in condensed matter fields despite the underlying physics are different. Unlike particle physics, the energy-momentum dispersion $E(k)$ of DM described by bloch-waves can satisfy the Dirac linear dispersion around certain points.
Started from the most celebrated material graphene, there are also other types of DMs—such as artificial graphene, topological insulator, d-wave superconductors and Weyl semimetal [20, 21, 22, 23, 24, 25, 26]—which may look completely different at first glance but in fact share a fundamental similarity. Unlike conventional metal or semiconductor, these materials show linear energy-momentum dispersions of low-energy excitation which are governed by Dirac-like equation with speed of light replaced by $v_F$

$$H = v_F \sigma \cdot p$$

(1.1)

rather than the Schrödinger equation. Quasi-particles described by such type of Dirac-like Hamiltonians are often called Dirac fermions and the low-energy dispersions characterized by the Dirac-like Hamiltonians reveal cone-shape linear feature are often referred to as Dirac cone shown in Fig. (1.3)

Based on the observed similarities, these are collectively called DMs. The emergence of Dirac point can be derived from symmetry where its type varies from material to material. The possible origins of symmetry include sublattice (pseudo-spin) symmetry or time-reversal symmetry (TRS). The similarities among the mathematical descriptions for relativistic quantum mechanics and for DMs’ low-energy excitations, also make DMs possibly inherit plentiful novel phenomena directly from particle physics. Besides massless-like behavior from linear dispersion, according to Eq. (1.1), even if there is a non-zero mass term (or gap), the positive and negative eigenstates correspond to electron and hole respectively are originated from the same spinor wave function hence linked via charge conjugation. This interconnection between conduction and valence band results in the same effective mass for both states. It also implies the analogy between DMs and particle/antiparticle pair in quantum electrodynamics (QED) that such interconnection between quasi-particles leads to chiral symmetry.

Chiral symmetry is the invariance under parity by a Dirac fermion. For any given particle, the chirality can is defined as $\Lambda = \sigma p/|p|$ and used to described the connection between the orientation of spin and direction of particle momentum. Unlike in relativistic quantum mechanics, here spin does not necessarily need to be real
Figure 1.1: Bulk graphene sheet and GNRs with different edge types highlighted by different colors.

electron spin. It could be other degrees of freedom such as in the case of graphane, \( \sigma \) represents pseudo-spin describing sublattice symmetry where spin-up and spin-down correspond to site-A and site-B in the unitcell or vise versa. The concept of chirality directly leads to one of the most striking phenomenon, Klein paradox, in both theoretical prediction [27] and experimental observation [28]. The nearly perfect electron-to-hole conversion at a potential barrier has been observed [28].

In the rest of the dissertation, we will focus on graphene, 3D and 2D TI-based devices.

1.0.3 Lattice and Electron Graphene Structure of Monolayer Graphene

Graphene-based materials, including 2D graphene and GNRs, although share common underlying electronic structure, their exact details depend on the quantum confinement effects. The differences among their electronic states could be significant in the presence of disorder. The earliest theoretical study of graphene can be traced back to 1947, Wallace used it as a starting point for understanding the electronic
Figure 1.2: (a) Graphene is made of carbon atoms arranged in honeycomb lattice formed by two triangular structure sublattices highlighted with blue and red colors. (b) Reciprocal lattice of graphene where $K'$ and $K$ are two inequivalent points and both correspond to Dirac points.

properties of 3D graphite. But the emergent massless Dirac equation was pointed out by Semenoff, DiVincenzo and Mele 40 years later [17].

Graphene is a one-layer sheet of graphite that consists of only carbon atoms arranged in honeycomb lattices. Its unit cell is made of two inequivalent sub-lattices. It is now one of the most famous DMs due to its merely one atom thickness and many extraordinary electronic properties. Each of the carbon atoms in graphene forms three strong covalent $\sigma$ bonds with three neighboring atoms and left one unpaired out-of-plane $p_z$ dislocalized electrons. The TB Hamiltonian for graphene can be easily expressed in a second quantization form as

$$H = \sum_n \varepsilon_n \hat{c}_n^\dagger \hat{c}_n + t_0 \sum_{\langle n,m \rangle} \hat{c}_n^\dagger \hat{c}_m \quad (1.2)$$

where $t_0$ is hopping constant describes the probability one electron jump from one site to its neighbors through overlap of their orbitals, $\hat{c}_n^\dagger$ ($\hat{c}_n$) creates (annihilates) $p_z$ electron at site $n$ and $\varepsilon_n$ is on-site energy. If the electric potential through out the
Figure 1.3: Energy-momentum dispersion for infinite graphene. The valence and conduction bands of graphene only touch at six discrete points within 1st BZ. These discrete points are often referred to as Dirac points. In low-energy excitations where the energy–momentum dispersion relation close to Dirac points shows linear behavior and described by Dirac-like as Eq. (1.1) and points where bands touch are called Dirac points.

whole graphene sheet is uniform, then the on-site energy is a constant \( \varepsilon_n = \varepsilon_0 \) and only shifts the bands. When graphene is clean we can simply drop out the constant and focus on the second term that gives the electron dynamics.

To calculate energy spectrum for a periodic system, according to Bloch’s theorem, the Eq. (1.2) can be Fourier transformed into the representation in terms of Bloch states of momentum \( \mathbf{k} \):

\[
H = \begin{bmatrix}
0 & h_0^* \\
h_0 & 0
\end{bmatrix},
\]

(1.3)

where \( h_0 = 1 + \exp(i \mathbf{k} \cdot \mathbf{a}_1) + \exp(i \mathbf{k} \cdot \mathbf{a}_2) \). By solving the eigenvalue problem, one immediately obtains one positive and one negative eigenenergy which correspond to conduction and valence bands respectively

\[
E(\mathbf{k}) = \pm t_0 |1 + \exp(i \mathbf{k} \cdot \mathbf{a}_1) + \exp(i \mathbf{k} \cdot \mathbf{a}_2)|. 
\]

(1.4)
There are two inequivalent points $K$ and $K'$, within the irreducible first Brillouin zone (BZ), that each can not be obtained from the other by simply adding or subtracting reciprocal lattice vectors. To analyze the low-energy excitation, we further expand Eq. (1.4) around $K$ through Taylor expansion by considering $\mathbf{k} = \mathbf{K} + \triangle \mathbf{k}$

$$E(\triangle \mathbf{k}) \cong \pm \left[ \left( \frac{\partial h_0}{\partial k_x} \right)_{k_x = K_x} (k_x - K_x) + \left( \frac{\partial h_0}{\partial k_y} \right)_{k_y = K_y} (k_y - K_y) \right]$$

$$= \pm v_F \hbar \sqrt{\triangle k_x^2 + \triangle k_y^2}$$

$$= \pm v_F \hbar \triangle k$$

where $v_F = \sqrt{3} t_0 a / 2$. Eq. (1.0.3) shows a linear dependence of $\triangle \mathbf{k}$ around $K$. A similar equation can be obtained for $K'$ by expanding Eq. (1.4) around $K'$. Both points are often called Dirac points as the energy-momentum dispersions of quasi-particles around them are linear and can be described by Dirac-like Hamiltonian that is formally equivalent to the 2D massless Dirac equation where the effective mass tensor $1/m^*(\text{Eq. 1.5})$

$$\frac{1}{m^*} = \hbar^{-2} \frac{\partial^2 E}{\partial k_i \partial k_j}$$

(1.5)

diverges at both points in which conventional effective mass fail to explain.

### 1.0.3.1 Graphene Nanoribbons

A pure graphene is a zero-gap material, so its application as a replacement for Si in semiconductors would require extra tuning as well as taking advantage of quantum confinement. It has been shown that the reduction of the GNRs width could enhance the performance [29, 31, 30] and therefore provide an alternative route for the usages of graphene.

GNRs are graphene in belt shapes that can be obtained by cutting along certain directions. The electronic properties respond to disorders differently depending on the type of the edges. They have been classified into three most common categories: GNRs with zigzag shape edges (ZGNRs) in Fig. (1.1); armchair shape edges (AGNRs); and
chiral edges which mix portions of both types of edges (CGNRs). Earlier experimental works have already introduced couple of different methods to fabricate GNRs [32, 35, 36], as well as both top-down [37] and bottom-up approach [38] to control the formation of sharp GNR edges.

One special transport feature is the emergence of edge currents around DPs both in ZGNRs and CGNRs without external field or SOC. With the fact that in the two types of GNRs mentioned above, the majority of currents flow through edges, the idea has been adopted in proposals of applications ranging from thermoelectrics to DNA sequencing [39, 40].

1.0.4 Topological Insulators

Topological insulator is characterized by its insulating bulk and perfectly conducting boundaries. The earliest idea that is similar to properties of TI, known as quantum hall effect (QHE), was proposed in 1975 [45] by considering two-dimensional electron gas (2DEG) subjected to a strong external magnetic field to break TRS. The idea can be understood by the semiclassical picture that a 2DEG gets localized by magnetic field due to Lorentz force

$$\mathbf{F} = q \left( \mathbf{E} + (\mathbf{v} \times \mathbf{B}) \right)$$

and for those electrons around edges do not have enough space to complete a full circular motion bounce off from the edges and keep leaping forward along the directions defined by edges. Such semiclassical skipping orbits around the edges become, after quantization, chiral edge states as wavefunctions whose probability amplitude is nonzero only around sample edge.

In TIs, the intrinsic SOC preserves TRS. The SOC acting on electrons can be thought of as effective magnetic fields that opposite spin feels the opposite direction. Bulk-boundary correspondence is a fundamental consequence of topological classification and unifying feature for all the TIs. For gapped bandstructures, the gap has to close at any interfaces where the topological invariant changes. For instance, both
vacuum and TIs belong to different topological classifications as insulators and with
different topological invariants, so the interfaces of TI must have gapless states. This
type of gapless states can not be removed by boundary passivation and the backscat-
tering is suppressed since spin-flip is forbidden by TRS. The simplest way to explain
whether band structures are equivalent is through continuous change. If one can be de-
formed into another without closing the energy gap while still conserving symmetries,
these bandstructures are considered topologically equivalent.

1.0.4.1 GNR-based 2D TI

A parallel topic in GNRs research is aiming at GNR-based topological insulator
also known as quantum spin Hall insulator. The original model was introduced by
Haldane [46] in 1988. The work predicted that a two-dimensional lattice with “2D
graphite” structure exhibits quantization of the Hall conductance without the presence
of external magnetic field. It was later extended to describe the quantum spin Hall
effect in graphene [47], by introducing second-nearest neighbor hopping term which
 corresponds to SOC into TB model in Eq. (1.2). This is the so called Kane-Mele
Hamiltonian

\[
\hat{H} = -t \sum_{\langle mn \rangle, \sigma} \hat{c}_{m \sigma}^\dagger \hat{c}_{n \sigma} + i \lambda_{\text{SO}} \sum_{\langle\langle mn \rangle\rangle, \sigma, \sigma'} \nu_{mn} \hat{c}_{m \sigma}^\dagger \hat{s}_{\sigma \sigma'} z \hat{c}_{n \sigma'}. \tag{1.6}
\]

The spin-dependent hopping in the second term, where \( \nu_{mn} = 1 \) for moving coun-
terclockwise around the hexagon and \( \nu_{mn} = -1 \) otherwise, acts between next-nearest
neighbor sites. The strength of such SOC is parameterized by \( \lambda_{\text{SO}} \). The introduced
SOC term induces QSH phase that turns GNR into 2D TI whit insulating bulk and
conducting edges protected by TSR. The intrinsic TI gap induced by QSH phase that
corresponds to \( \lambda_{\text{SO}} \) in pristine graphene has been theoretically predicted to 24 \( \mu \text{eV} \)
[96]. The phase can only be observed as low temperature as 0.28 K.

Even though the GNR with intrinsic \( \lambda_{\text{SO}} \) sets up a perfect example as 2D TIs,
minuscule size of gap makes impractical for application. Nevertheless, Ref. [PRX]
has predicted that by doping with a small concentration of heavy atoms onto graphene
sheet, the intrinsic $\lambda_{SO}$ can be greatly enhanced by orders of magnitude. This prediction brings back the research interests in its applications as well as motivates the studies targeting on the further enhancement of SOC in graphene.

### 1.0.4.2 3D TI

3D TIs are characterized by its insulating bulk surrounded by conducting surface states that are described as two-dimensional Dirac fermions. The first strong 3D TI discovered is $\text{Bi}_{1-x}\text{Sb}_x$ alloys which have been experimentally studied for their thermoelectric properties. Later on, $\text{Bi}_2\text{Se}_3$, $\text{Bi}_2\text{Te}_3$, and $\text{Sb}_2\text{Te}_3$ were soon identified as 3D TI [42, 43, 44]. $\text{Bi}_2\text{Se}_3$ is prototypical to its large gap (0.3 eV) and it has DP in the gap, unlike $\text{Bi}_2\text{Te}_3$ where DP is below top of valence band, located right at $\Gamma$ point that has been observed through APRES [41]. Figure. 1.4(a) shows first-principles calculations with two different packages based on Density-functional theory (DFT).

Theoretically, in 2006, Fu, Kane and Mele generalized 2D QSH to 3D using diamond lattices and similar Hamiltonian to Eq. (1.6). Soon after, a simple effective four-band model Hamiltonian was proposed by Zhang and the collaborators that can be applied to a large family of 3D TIs with relatively small computational cost and has led to a series of predictions for applications [15]. The most recent study predicts several nontrivial features of $\text{Bi}_2\text{Se}_3$ through first-principle calculations [16]. As opposed to simplified 2D surface state model, the results reveal complex noncollinear spin textures as well as finite width of surface states in fact spread over almost two QLs. The results imply the necessity of accurate first-principles calculation in order to capture the details of the surface states.

### 1.1 Thermoelectric Energy Conversion

The Seebeck effect (SSE) is the direct conversion between temperature difference $\Delta T$ and electric voltage $\Delta V$. The effect can be used for power generator by recycling waste heat or as a local cooling device with a given electric voltage. Its physical origin can be understood using a simple picture. For a given $\Delta T$ applied across the device,
Figure 1.4: (a) The electronic band structure computed for a supercell of Bi$_2$Se$_3$ film of thickness 5 QLs using ncDFT with pseudo-atomic localized basis functions [159] implemented in ATK [158] package. This is contrasted with the electronic band structure obtained using ncDFT with PW basis set implemented in VASP package. [169] (b) Zero-bias transmission function of Bi$_2$Se$_3$ thin film in the two-terminal geometry of Fig. 6.1, for electrons injected along the $\Gamma$–$M$ direction ($k_y = 0$) in the inset of Fig. 6.2(a), computed using NEGF+ncDFT formalism implemented in ATK package. Adapted from Ref. [16].
Figure 1.5: (a) Schematic view of a device attached to hot and cold macroscopic reservoirs that induces the temperature gradient the electrons around hot area diffuse faster than cold ones therefore resulting in charge accumulation on both sides. (b) The difference of Fermi-Dirac distribution functions is anti-symmetric around $E_F$, therefore the optimization of $I_T$ relies on the asymmetry of transmission function.
as described in Fig. 1.5 (a), the electrons on both hot and cold ends diffuse differently at different temperatures and hence build up charge accumulation that induce electric field in the opposite direction to the direction of $\Delta T$. The ratio between the given $\Delta T$ and $\Delta V$ gives Seebeck coefficient:

$$S = -\frac{\Delta V}{\Delta T}.$$ 

A more accurate description relies on the concepts of density of states (DOS) and transmission function $\mathcal{T}(E)$ of nanostructures. The central quantity of the thermoelectric effect is the temperature-driven current $I_T$ which in the Landauer-Büttiker formula can be described as integration of transmission function $\mathcal{T}(E)$ over energy weighted by the difference of Fermi-Dirac distribution (FDD) functions at two different temperatures

$$I_T = \frac{e}{h} \int \mathcal{T}(E) [f_H(E) - f_L(E)] dE. \quad (1.7)$$

As Fig. 1.5 (b) indicates, the difference of FDD functions is roughly anti-symmetric around $E_F$. Therefore, the integration of flat transmission yields zero net current. So the key idea to maximize $I_T$ requires the asymmetry of transmission function to isolate the contribution above (below) $E_F$ from the contribution below (above) $E_F$. The step-like transmission function in Fig. 1.5 (b) is the advantageous situation for the optimization of $I_T$ where the only charge carriers with energy above $E_F$ contribute to net $I_T$.

### 1.1.1 Efficiency of Thermoelectric Conversion

The whole effect can be summarized, for simplicity, by two 1D coupled equations

$$E_x = \frac{1}{\sigma} J_x + S \frac{dT}{dx}, \quad (1.8)$$

$$J_{Qx} = \Pi J_x - \kappa \frac{dT}{dx}. \quad (1.9)$$
Figure 1.6: Consider an N-type cooling system. The heat is pumped from cold side $T_C$ to hot side $T_H$. While the heat is carried to the hot end, part of the flow is canceled out by the heat dissipation and the temperature gradient.

Equation (1.8) is modified from the well-known current density equation with additional correction term coming from Seebeck effect and Eq. (1.9) states the heat flow contains contributions from both electric current $J_x$ and temperature gradient ($dT/dx$), where $\Pi$ is Peltier coefficient represent the ratio between electric current and the heat flow in the absence of $dT/dx$.

To estimate the performance of thermoelectric effect, we start with an example of cooling system in 1D cause in (Fig. (1.6)). If we want to pump heat current from cold end to hot end, following Eq. (1.9), assuming heat dissipation is random and roughly half amount of heat going in $+x$ direction and half in $-x$ direction, we then end up with additional third term

$$J_{Qx} = \Pi \frac{I_x}{A} - \kappa \frac{dT}{dx} - \frac{I_x^2 R_x}{2A} \quad (1.10)$$

where $R_x = \rho L_x / A$. The optimized situation happens under two conditions: (i) $dJ_{Qx}/dI = 0$ gives maximum $J_{Qx}$. (ii) the biggest temperature difference has been reached so that $J_{Qx}$ can not be increased further. By satisfying both conditions, we
obtain

\[ Z = \frac{S^2 \sigma}{\kappa}, \]  

is so called TE figure of merit. \( Z \) is a preliminary estimation for the efficiency of thermoelectrics, and it is often treated as a dimensionless parameter by multiplying it with temperature, \( ZT = TS^2 \sigma / \kappa \),

15
Chapter 2

INTRODUCTION TO NON-EQUILIBRIUM GREEN’S FUNCTION

The traditional tools for electronic and spintronic device simulations, based on classical drift-diffusion or semiclassical Boltzmann equation, cannot be used for quasiballistic nanometer-size active region attached to much larger reservoirs. The proper description of such open quantum systems can be achieved using quantum master equations (QMEs) \[48, 49, 50\] for the reduced density matrix of the active region or the NEGF formalism \[91\]. The former is typically used when the active region is weakly coupled to the reservoirs and can handle only a small number of atomic orbitals \[49\], while the latter is employed in the opposite limit and can handle large number of atomic orbitals in the active region \[51, 52\] in open quantum systems we have encountered can be divided into two different types of regions. The first type is finite size and often called central or active region which we are interested in and all the interactions take place while the second type is semi-infinite and describes leads. The effects of leads can be viewed as interactions between central region and leads. It appear in the equation of motions for the central region as complex energy term known as embedded self-energy (SE). The embedded SE has different origin from the SE that is due to many-body interactions, but its emergence in the equation of motions (EOM) of the system shares the same mathematical expression and can therefore be treated perturbatively through Dyson equations.

2.1 Retarded Green’s Function in Single-particle Quantum Mechanics

In static case, considering single particle Schrödinger equation

\[ [H_0(r) + V(r)]\Psi = E\Psi, \] (2.1)
where stationary
\[ H_0(r)\Psi_0 = E\Psi_0, \]
and \( \Psi_0 \) is eigenstates of \( H_0(r) \) and can be solved exactly while \( V(r) \) is treated as a perturbation. The Green’s function is defined as a general solution to the differential operator \( [E - H_0] \). It represents the response to an unit excitation at a given energy

\[ [E - H_0(r)]G_0(r, r'; E) = \delta(r - r'). \]  (2.2)

We can also write down similar expression for Eq. (2.1)

\[ [E - H_0(r) + V(r)]G(r, r'; E) = \delta(r - r'). \]

It is natural to define
\[ G_0^{-1}(r; E) = E - H_0(r), \]  (2.3)
and
\[ G_0^{-1}(r; E)G_0^{-1}(r, r'; E) = \delta(r - r'). \]  (2.4)

the Schrödinger equation can then be rewritten as

\[ [G_0^{-1}(r) + V(r)]\Psi = 0, \]  (2.5)

and the solution has an integral form
\[ \Psi(r) = \Psi_o(r) + \int dr' G_0(r, r'; E)V(r')\Psi(r'). \]  (2.6)

The expression of total wavefunction (WF) has an iterative form which contains itself on the RHS of Eq. (2.6). We can iterate the solution to re-express the solution in terms of free particle quantities \( \Psi_0 \) and \( G_0 \), to obtain
\[ \Psi(r) = \Psi_o(r) + G_0V\Psi_o + G_0V G_0V\Psi_0 + G_0V G_0 V G_0V\Psi_0, \]
\[ = \Psi_o(r) + (G_0 + G_0 V G_0 + G_0 V G_0 V G_0)V\Psi_0. \]

Compare with Eq. (2.6), we observe that \( G \) also has an iterative form
\[ G = G_0 + G_0 V G. \]  (2.7)

Equation 2.7 is called Dyson equation and it is in a particularly useful in perturbation theory which re-express \( G \) in terms of on-interacting Green’s function \( G_0 \).
2.1.1 Retarded Green’s function as a propagator

According to the definition of the GF, the total WF function due to arbitrary source can also be recovered from

$$\Psi(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}'; E)s(\mathbf{r}')d\mathbf{r}',$$

or its discretized form in matrix representation

$$\Psi = [E - H]^{-1}s = Gs.$$ (2.8)

To ensure the WFs converge to zero at infinite, we impose boundary conditions by adding an infinitesimal imaginary term $\pm i\eta$. It leads to the most common solutions retarded GF $G^r$ for equation

$$[E - H(\mathbf{r}) + i\eta]G^r(E)(\mathbf{r} - \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'),$$

that corresponds to the response at $\mathbf{r}$ due to unit excitation applied at $\mathbf{r}'$ and its Hermitian conjugate, advanced GF $G^a(E)$

$$[E - H(\mathbf{r}) - i\eta]G^a(E)(\mathbf{r} - \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}').$$

The $G$ has been introduced here is the non-interacting version of the retarded single particle GF in energy representation. It behaves as a propagator and propagates wave function. The idea will become more clear if we consider the it in time representation by Fourier transformation. Equation. (2.1) becomes

$$[-i\frac{\partial}{\partial t} - H_0(\mathbf{r}) - V(\mathbf{r})]\Psi = 0.$$

The solution $\Psi$ has a time-dependent form

$$\Psi(\mathbf{r}, t) = \int d\mathbf{r}'G_0(\mathbf{r}, \mathbf{r}', t'; E)\Psi(\mathbf{r}', t').$$

For time-dependent problem, the solution of time-evolved WF $\Psi(t)$ with propagator reads

$$\langle r|\Psi(t)\rangle = \int dr' \langle r|e^{-iH(t-t')}|r'\rangle\langle r'|\Psi(t')\rangle.$$
By observation, we can define retarded GF by
\[ G^r(r, t, r', t') = -i\theta(t - t') \langle r | e^{-iH(t-t')} | r' \rangle, \]
and its hermitian conjugate advanced GF
\[ G^a(r, t, r', t') = -i\theta(t' - t) \langle r | e^{-iH(t-t')} | r' \rangle, \]
which are both solutions of time-dependent Schrödinger equation.

2.1.2 Space discretization in 1D case:

If the Hamiltonian is discretized in a way that the interaction only involves orbitals on neighboring sites, it is also known as TB method. In the following, we start with simple 1D case for space discretization. With space coordinate \( r \) and WF \( \Psi(r) \) be described in vector form \( [a, 2a, ..., ja, ...] \) and \( [\Psi_1, \Psi_2, ..., \Psi_j, ...] \) with spacing \( a \). The Hamiltonian operator \( \hat{H} \) acting on wave function \( \Psi \) can be rewritten as
\[
[\hat{H}\Psi]_{x=ja} = -\left[ \frac{\hbar^2}{2m} \frac{d^2}{dx^2} \right]_{x=ja} + U_j \Psi_j, \tag{2.11}
\]
Applying the finite difference method, we obtain expressions for first derivative
\[
\left( \frac{\partial \Psi}{\partial x} \right)_{j+1/2} = \frac{\Psi_{j+1} - \Psi_j}{a}, \quad \left( \frac{\partial \Psi}{\partial x} \right)_{j-1/2} = \frac{\Psi_j - \Psi_{j-1}}{a},
\]
and the second derivative
\[
\left( \frac{\partial^2 \Psi}{\partial x^2} \right)_j = \frac{\left( \frac{\partial \Psi}{\partial x} \right)_{j+1/2} - \left( \frac{\partial \Psi}{\partial x} \right)_{j-1/2}}{a} = \left( \frac{1}{a^2} \right) \left[ \Psi_{j+1} - 2\Psi_j + \Psi_{j-1} \right],
\]
with this approximation we can rewrite the Eq. (2.11)
\[
[\hat{H}\Psi]_{x=ja} = (U_j + 2t)\Psi_j - t\Psi_{j-1} - t\Psi_{j+1},
\]
where \( t \equiv \hbar^2/2ma^2 \).

\( \mathbf{H} \) in this space grid representation has a matrix form,
\[
\mathbf{H} = \begin{pmatrix}
... & -t & 0 & 0 & 0 \\
-t & U_{-1} + 2t & -t & 0 & 0 \\
0 & -t & U_0 + 2t & -t & 0 \\
0 & 0 & -t & U_1 + 2t & -t \\
0 & 0 & 0 & -t & ...
\end{pmatrix}.
\]
and from Eq. (2.10) we obtain

$$G'(E) = [(E + i\eta)1 - H]^{-1}.$$ 

2.1.3 Open quantum system

To continue the analysis in previous 1D case, since it is periodic and infinite in transport direction, we normally need to divide the system into three regions (Fig. 2.1), left lead, central region and right lead. The division allows us to define a region of interest and reduce the Hamiltonian to solvable dimension. Both leads are assumed semi-infinite and connected smoothly at ±∞ to reservoirs and the central region is what we are interested in and, in many cases, where interactions take place. According to this idea, Hamiltonian can be block-divided into a general form

$$H = \begin{bmatrix} H_L & H_{LC} & 0 \\ H_{CL} & H_C & H_{RC} \\ 0 & H_{CR} & H_R \end{bmatrix}.$$ 

Assuming two leads are decoupled, which can always be true by expanding the central region. The coupling at the Lead 1/Central interface and Central region/Lead
2 interface are expressed in a matrix form $\tau_{R/L}$. One of the advantage of GF is that the central region can be arbitrary shape

$$\begin{bmatrix}
G_L & G_{LC} & 0 \\
G_{CL} & G_C & G_{RC} \\
0 & G_{CR} & G_R
\end{bmatrix} = \begin{bmatrix}
E - H_L + i\eta & \tau_L & 0 \\
\tau_L^\dagger & E - H_C & \tau_R \\
0 & \tau_R^\dagger & E - H_R + i\eta
\end{bmatrix}^{-1}\quad (2.12)
$$

Based on the previous assumption central region $G_C$ is what we are interested, we are looking for the expression for $G_C$. To simplify the problem, only one (left) interface is considered at the moment. Eq. (2.12) reduces to

$$\begin{bmatrix}
G_L & G_{LC} \\
G_{CL} & G_C
\end{bmatrix} = \begin{bmatrix}
E - H_L + i\eta & \tau_L \\
\tau_L^\dagger & E - H_C
\end{bmatrix}^{-1} \quad (2.13)
$$

From two linear equations both involving $G_C$ and $G_{LC}$

$$[E - H_L + i\eta] G_{LC} + \tau_L G_C = 0,$$

$$\tau_L^\dagger G_{LC} + [E - H_C] G_C = 1,$$

by eliminating $G_{LC}$, one obtains equation

$$G_C = [E - H_C - \Sigma]^{-1}, \quad (2.14)$$

where $\Sigma_L = \tau_L^\dagger g_L \tau_L$ and $g_L = [E - H_L + i\eta]$.

The term $\Sigma_L$ appearing in Eq. (2.14) is a complex term that describes the effect how electron can enter or escape from the central region and can be viewed as the presence of source and drain. The size of is then reduced to finite by replacing semi-infinite lead with an effective “source” or “drain” adding to the interface.

According to the previous assumption that leads are decoupled, we can recover the full system by simply adding the effect from the right lead to the Hamiltonian. The retarded GF for this open system becomes

$$G^r = G_C = [E - H_C - \Sigma_L - \Sigma_R]. \quad (2.15)$$

More importantly, the ideas of source and drain are considered as interactions of finite central region interacting with semi-infinite leads.
2.1.4 Physical observables from one-particle Green’s functions

With the Knowledge of $G^r$ and $\Sigma^r$, there are several physical quantities can be directly calculated. The density of states $n(E)$ can be expressed as

$$n(E) = \frac{1}{2\pi} \text{Tr}[A(E)],$$

where $A(E)$ is the spectral function defined as

$$A(E) = i[G^r(E) - G^a(E)],$$

whose diagonal elements give local density of states (LDOS) $\rho(r; E) = \rho(r, r; E)$.

$$\rho(r; E) = \frac{1}{2\pi} A(r, r, E) = -\frac{1}{\pi} \text{Im}[G^r(r, r; E)],$$

or in discretized matrix form of basis representation

$$\rho_i(E) = \frac{1}{2\pi} A_{ii}(E) = -\frac{1}{\pi} \text{Im}[G^r_{ii}(E)].$$

The level broadening function is defined as

$$\Gamma_\alpha = i[\Sigma^r_\alpha - \Sigma^a_\alpha], \quad (2.16)$$

that describes the rate of electron escaping from the central region.

The transmission function between the lead $p$ and lead $q$, can be expressed as

$$T_{pq} = \text{Tr}[\Gamma_p G^r \Gamma_q G^a], \quad (2.17)$$

and the current formula

$$I_p = \frac{e}{\hbar} \int dE T_p(E) (f_p - f_q) = \frac{e}{\hbar} \int dE \text{Tr} [\Gamma_p G^r \Gamma_q G^a] (f_p - f_q).$$

This is known as Landauer-Büttiker formula. A more rigorous derivation will be introduced in Sec. 2.4.
2.1.5 A simple picture for NEGF and density matrix $G^<$

Unlike semi-classical picture, FDD function $f$ is not enough in phase coherent transport problem where phase-relationships among different states are not negligible. Therefore a more general quantity, lesser GF $G^<$ which takes diagonal terms into account is introduced. In fact lesser GF is closely related to density matrix. In steady-state one particle less GF for many-body (MB) non-interacting system is defined as

$$
\langle \hat{c}_\beta ^\dagger \hat{c}_\alpha \rangle = \frac{\hbar}{i} G^<_{\alpha\beta}(t, t) = \frac{\hbar}{i} G^<_{\alpha\beta}(\tau = 0) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} dE G^<_{\alpha\beta}(E),
$$

(2.18)

and the density matrix in matrix form states

$$
\rho = \frac{1}{2\pi i} \int_{-\infty}^{\infty} dE \ G^<(E),
$$

(2.19)

where $\hat{c}_\beta ^\dagger$ and $\hat{c}_\alpha$ are field operators. This can be understood with a simple picture [138]. In one-particle description, creation and annihilation operators reduce to one particle WF and its complex conjugate. The lesser GF then reduces to

$$
G^<_{\alpha\beta} = (2\pi i) \Psi^\dagger_{\beta} \Psi_{\alpha},
$$

and the matrix form

$$
G^< = (2\pi i) \Psi \Psi^\dagger.
$$

The lesser GF was originally introduced by Keldysh [33] and in principle determines most important kinetic properties, such as distribution function, charge, and current. From the definition of density matrix defined in Eq. (2.18), the diagonal elements are directly related to electron density at the the particular state $\alpha$, depends on which basis representation is chosen

$$
n_\alpha = i \langle \hat{c}_\alpha ^\dagger \hat{c}_\alpha \rangle = -i G^<_{\alpha\alpha}(E).
$$

(2.20)

In the absence of the many-body interactions, the lesser GF can be obtained exactly through the kinetic equation

$$
G^<(E) = G^r(E) \Sigma^>(E) G^a(E),
$$

(2.21)
where $\Sigma^<_\alpha = f_\alpha(E)\Gamma_\alpha(E)$, $\Sigma^>_\alpha = (1 - f_\alpha(E))\Gamma_\alpha(E)$ and $f_\alpha$ are FDs of given conditions in the leads.

2.2 Time-dependent Problems

To be able to derive a more general expression for single particle GFs of MB system, we begin with a time dependent problem and the Shrödinger equation reads

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle.$$ 

The solution for time-independent Hamiltonian ($H=H(t_0)$) has a simple form

$$|\Psi(t)\rangle = e^{-i\hat{H}(t_0)(t-t_0)} |\Psi(t_0)\rangle,$$

whereas for time-dependent Hamiltonian $H=H(t)$, WF is governed by a more complicated evolution operator $U(t,t_0)$, time-evolved ket is

$$|\Psi(t)\rangle = U(t,t_0) |\Psi(t_0)\rangle,$$  \hspace{1cm} (2.22)

where general operator for Eq. (2.22)

$$U(t,t_0) = T \left\{ e^{-i \int_{t_0}^{t} dt' H(t')} \right\},$$  \hspace{1cm} (2.23)

and $T$ is the time ordering operator.

The expectation value of the observable can be expanded as

$$\langle \Psi(t)| O(t) |\Psi(t)\rangle = \langle \Psi(t_0)| U(t_0,t)O(t)U(t,t_0) |\Psi(t_0)\rangle,$$

and its ensemble average in Heisenberg picture reads

$$O(t) = \langle O_H(t) \rangle \equiv Tr\{\rho_0 O_H(t)\} = Tr\{\rho_0 U(t_0,t)O(t)U(t,t_0)\}.$$  \hspace{1cm} (2.24)

For $t > t_0$ the evolution operator $U(t,t_0)$ follows chronological order while $U(t_0,t)$ follows anti-chronological order. Applying Eq. (2.23) properly we find

$$O(t) = \langle \Psi(t_0)| \bar{T} \left\{ e^{-i \int_{t_0}^{t} dt' H(t')} \right\} O(t) T \left\{ e^{-i \int_{t_0}^{t} dt' H(t')} \right\} |\Psi(t_0)\rangle,$$  \hspace{1cm} (2.25)
where $\bar{T}$ is revers-time-ordering operator. We may say that the system starts from ket state $|\Psi_0\rangle$ evolves from $t_0$ to $t$, after $O(t)$ acts, and then back from $t$ to $t_0$. The evolution consists of forward and backward branches on the time axis which forms a round trip contour. With this contour idea, we can introduce an oriented “contour” that combines both branches [91]

$$C = C_- \oplus C_+,$$

so that Eq. (2.25) can be re-expressed in a more elegant and general form

$$O(z) = \langle \Psi_0 | T_C \left\{ e^{-i \int_C dz' H(z')} \right\} O(z) | \Psi_0 \rangle.$$  

(2.26)

Taking statistical average into account does not change the form Eq. (2.26), but extend the contour to imaginary time in vertical axis.

Following the idea above, we can define physical operators on the contour for time-dependent problems. The definition of the expectation value in Eq. (2.24) can also be applied to calculate the expectation value of two-operator object, often called two-time correlator operator. The one-particle NEGF of MB systems is then defined as

$$G(z, z') = \langle \mathbf{r} | G(z, z') | \mathbf{r}' \rangle \equiv -\frac{i}{\hbar} \langle T_C \left\{ \psi_H(r, z) \psi_H^\dagger(r', z') \right\} \rangle.$$

The definition is not physically intuitive at the moment, but it gives some hints as it is a statistical average of creation of particle at position $r'$ and time $z'$ correlates with its annihilation at position $r$ and time $z$. This field operator form of expression is particularly useful since all the many-body operators can be built from field operators. In fact for equal $t$ and $r$ case, the result is directly related to electron density.

### 2.3 Equation of Motion of $G^<$

The EOM for the NEGF can be introduced through two different but equivalent formulations: the Kadanoff-Baym and the Keldysh formulation. The relation between the the Kadanoff-Baym (KB) equation and Keldysh equation is in fact that the one takes the form of ordinary differential equation plus a boundary condition and the other one takes the form of corresponding integral equation.
The Keldysh contour simplifies the problems by utilizing the property of time evolution operator as unitary transformation, see Fig. 2.2. In stationary case as the initial condition is irrelevant, one can extend the contour $C$ without changing the expectation value

$$
T \left\{ e^{-i \int_{C} dz' H(z')} O(t_-) \right\} = U(t_0, \infty) U(\infty, t) O(t) U(t, t_0) = U(t_0, t) O(t) U(t, t_0),
$$

and similarly for $T \left\{ e^{-i \int_{C} dz' H(z')} O(t_+) \right\}$. The contour idea is particularly useful when combining with adiabatic assumption in treating non-equilibrium problem. In the case of open quantum system, all the regions are assumed to be decoupled at $t = -\infty$ and the couplings are gradually turned on until $t_0$ the interaction is fully established. The advantage of this approach is that a well-defined state in which the system was prepared at $t = -\infty$ would define all the expectation values. The vertical part of contour that comes from statistical average, which has been proven only relevant for initial condition and the effects decay exponentially, can be neglected.

The contour representation itself is rather impractical for calculations. One prefers to work with real time integrals. The common procedure of converting the contour into real-time integrals is called analytic continuation. We here follow the formulation by Langreth. The contour $C_K$ depicted in Fig. 2.2 consists of two branches:
Each time arguments of the Green’s function can reside either on the first or second part of the contour. In summary, contour-ordered GFs can be classified into four different Green’s functions based on the relationships between time arguments and each of components carries different types of information. The most common GFs that correspond to different components on contour are lesser GF

\[ G^\leq(t, t') = -i \langle \psi_H(t') \psi^\dagger_H(t) \rangle. \]

that describing electron density, greater GF that describing holes occupation

\[ G^\geq(t, t') = i \langle \psi^\dagger_H(t) \psi_H(t') \rangle. \]

retarded GF

\[ G^r = +\theta(t - t') [G^>(t, t') - G^<(t, t')], \tag{2.27} \]

acting as evolution operator and it’s hermitian conjugate advanced GF \( G^a = [G^r]^\dagger \).

### 2.3.1 Kadanoff-Baym formulation

It is sometimes more straightforward to derive current expression starting with KB formulation. Similar to Eq. (2.7), Dyson’s equations can be written as

\[ G(r, t; r', t') = G_0(r, t; r', t') + \int d1d2 G(r, t; 1) \Sigma(1; 2) G_0(2; r', t'), \tag{2.28} \]

and the other equivalent equation, known as adjoint Dyson’s equation,

\[ G(r, t; r', t') = G_0(r, t; r', t') + \int d1d2 G_0(r, t; 1) \Sigma(1; 2) G(2; r', t'). \tag{2.29} \]

where 1 and 2 in the arguments of GFs are given space-time coordinate \((r_1, t_1)\) and \((r_2, t_2)\) respectively. By definition, equations of motion (EOMs) of free particle GFs are

\[ \left[i\hbar \frac{d}{dt_1} - H_0(1)\right] G_0(1; 2) = \delta_{1,2}, \]

and it’s adjoint equation

\[ \left[i\hbar \frac{d}{dt_2} - H_0(2)\right] G_0(1; 2) = \delta_{1,2}. \]
The EOMs of $G(r,t;r',t')$ that contains the effect of leads can be obtained by acting $[i\hbar d/dt_1 - H_0(1)]$ and $[i\hbar d/dt_2 - H_0(2)]$ on Dyson’s equations Eq. (2.29) respectively,

$$\left[i\hbar \frac{d}{dt_1} - H_0(1)\right] G(1;2) = \delta_{1,2} + \int d3\Sigma(1;3)G(3;2),$$  \hspace{1cm} (2.30)

and its adjoint equation

$$\left[-i\hbar \frac{d}{dt_2} - H_0(2)\right] G(1,2) = \delta_{1,2} + \int d3G(1;3)\Sigma(3;2)$$  \hspace{1cm} (2.31)

Converting both Eq. (2.30) and (2.31) to real time representation by applying Langreth rules $D = \int_C AB$, $D^\approx = \int_C [A^r B^\approx + A^\approx B^a]$ leads to

$$\left[i\hbar \frac{d}{dt_1} - H_0(1)\right] G^\approx(1;2) = \int d3[\Sigma^r(1;3)G^\approx(3;2) + \Sigma^\approx(1;3)G^a(3;2)],$$  \hspace{1cm} (2.32)

and its adjoint equation

$$\left[-i\hbar \frac{d}{dt_2} - H_0(2)\right] G^\approx(1;2) = \int d3[G^r(1;3)\Sigma^\approx(3;2) + G^\approx(1;3)\Sigma^a(3;2)]$$  \hspace{1cm} (2.33)

Applying Eq. (2.27) by subtracting $G^<$ of Eq. (2.32) from $G^>$ of Eq. (2.33), one obtains EOM for retarded GF. [34]

$$\left[i\hbar \frac{d}{dt_1} - H_0(1)\right] G^r(1;2) - \int d3[\Sigma^r(1;3)G^r(3;2)] = \delta_{1,2}$$  \hspace{1cm} (2.34)

and

$$\left[i\hbar \frac{d}{dt_2} - H_0(2)\right] G^r(1;2) - \int d3[\Sigma^r(1;3)G^r(3;2)] = \delta_{1,2}.$$  \hspace{1cm} (2.35)

2.3.2 Keldysh formulation

Instead of working with KB relations, in integro-differential form, it is sometimes more convenient to consider integral equations. One can apply Langreth’s rules directly to Dyson’s equation Eq. (2.29) and obtain

$$G^< = G^<_0 + G^r_0\Sigma^r G^< + G^r_0\Sigma^a G^a + G^<_0\Sigma^a G^a$$  \hspace{1cm} (2.36)
To make the equations more easily readable, the notation where a product of two terms is interpreted as a matrix product in the internal variables (space, time, etc.) has been used. An expression for $G^<$ can be derived by iterating and re-arranging Eq. (2.36)

$$G^< = (1 + G^r \Sigma^r)G^<_0 (1 + G^a \Sigma^a) + G^r \Sigma^r G^a$$  \hspace{1cm} (2.37)

It is sufficient to consider only the second term throughout our studies since the first term on the RHS of Eq. (2.37) corresponds to initial conditions which is irrelevant non-interacting steady-state systems. Equation. (2.37) can then be expressed in integral form

$$G^<(12) = \int d^3 d^4 G^r(13) \Sigma^<(34) G^a(42), \hspace{1cm} (2.38)$$

also known as Keldysh equations. $G^>$ can be obtained with the similar procedure.

Under steady-state condition, where Hamiltonian does not depend on time, two-time dependence reduce to the dependence of $\tau = t_1 - t_2$, it is advantageous to work on energy space through Fourier transformation (FT). The matrix form in a given basis representation reads

$$G(E) = \frac{1}{\hbar} \int d\tau \exp(iE\tau/\hbar) G(\tau).$$

Applying to Eq. (2.34) and (2.38) directly recover Eq. (2.15) and (2.21) respectively.

### 2.4 Current Operator

In this section expression of $I_{op}$ will be introduced and link to several well-known formulas. Based on the conservation of charge (particle), the continuity equation states

$$e \frac{\partial \rho}{\partial t} + \nabla \cdot j = 0., \hspace{1cm} (2.39)$$

Compare with the EOM of $G^<$ which can be obtained by setting $t_1 = t_-$, $t_2 = t_+$ and subtracting Eq. (2.33) from Eq. (2.32)

$$i\hbar \frac{d}{dt} G^<(t,t) = [H_0, G^<(t,t)] + [\Sigma^r G^< + \Sigma^< G^a - G^r \Sigma^r - G^< \Sigma^a](t,t). \hspace{1cm} (2.40)$$
Equation (2.40) is a rather general form that we directly derive from EOM of Dyson equations without making further assumption. It is somewhat redundant, at least in non-interacting case, so the way to simplify it depends on the problems at hand. The first term on the RHS of Eq. (2.40) corresponds the current flowing through the central interacting area which is useful for local current/bond current calculations, while the second term accounts for the interaction with leads at the boundaries which is convenient for calculating the total current flow through any given terminal. In the non-interacting case, the SEs are only from leads. The sum of second term then yield zero net current to preserve current and particle conservation.  

2.4.1 Current in the terminals

As we explained previously, it is often more convenient to focus on the second term for the terminal currents. Here we can define a current operator based on the second term of Eq. (2.40)

\[ I_{\text{op}} = \frac{e}{\hbar} \left[ \Sigma^r G^< + \Sigma^< G^a - G^< \Sigma^a - G^r \Sigma^< \right]. \tag{2.41} \]

And to calculate current, we only need to take trace of Eq. (2.41) and integrate over energy

\[ I = \int dE \text{Tr}[I_{\text{op}}(E)] \]

\[ I = \frac{e}{\hbar} \int dE \text{Tr} \left[ \Sigma^r(E) G^<(E) + \Sigma^<(E) G^a(E) - G^<(E) \Sigma^a(E) - G^r(E) \Sigma^<(E) \right]. \tag{2.42} \]

In non-interacting two-terminal system, the net current given by Eq. (2.42) is zero. This means in steady-state transport, what enter the bulk area through left lead

\[ 1 \text{ The statement does not hold in the presence of interaction, since the SEs due to many-particle interaction are obtained approximately. For instance, in the case of electron-phonon interaction, at least up to self-consistent Born approximation should be considered to ensure the current conservation.} \]
should be equal to what flows out through right lead. To extend this idea, we can find the current at any terminals from Eq. (2.42) at specific lead $\alpha$, by separating the contributions with respect to relevant SE $\Sigma^{i}_\alpha (i = \gtrless, r/a)$

$$I = \sum_{\alpha} I_{\alpha},$$  \hspace{1cm} (2.43)

where

$$I_{\alpha} = \frac{e}{\hbar} \int dE \text{Tr} \left[ \Sigma^{r}_\alpha (E) G^{\gtrless}(E) + \Sigma^{<}_\alpha (E) G^{a}(E) - G^{<}(E) \Sigma^{a}(E) - G^{r}(E) \Sigma^{<}_\alpha (E) \right].$$  \hspace{1cm} (2.44)

Many commonly used expressions can be derived from Eq. (2.44) and automatically work for multi-terminal system.

**Meir-Wingreen formula**

Using cyclic permutation property of the trace, from Eq. (2.43) and regrouping the terms on the RHS, we can write down

$$I = \frac{e}{\hbar} \int dE \text{Tr} \left\{ [\Sigma^{r}(E) - \Sigma^{a}(E)] G^{\gtrless}(E) + [G^{a}(E) - G^{r}(E)] \Sigma^{<}(E) \right\},$$

with the help of the identities

$$G^{r} - G^{a} = G^{>} - G^{<},$$  \hspace{1cm} (2.45)

and

$$\Sigma^{r} - \Sigma^{a} = \Sigma^{>} - \Sigma^{<},$$  \hspace{1cm} (2.46)

We obtain Meir-Wingreen’s current expression

$$I_{p} = \frac{e}{\hbar} \int dE \text{Tr} \left[ \Sigma^{\gtrless}_{p}(E) G^{>}(E) - \Sigma^{>}_{p}(E) G^{<}(E) \right].$$  \hspace{1cm} (2.47)
Landauer-Büttiker formula

Using Meir-Wingreen expression and combined with Eq. (2.21), Eq. (2.47) becomes

\[ I_p = \frac{e}{\hbar} \int dE \text{Tr} \left[ \Sigma_p^< G^r \Sigma_p^> G^a - \Sigma_p^> G^r \Sigma_p^< G^a \right]. \]  

(2.48)

where

\[ \Sigma^\approx = \sum_p \Sigma^\approx_p, \]  

(2.49)

\[ \Sigma^> = \sum_p (1 - f_p(E)) \Gamma_p(E), \]  

(2.50)

and

\[ \Gamma = \sum_p \Gamma_p. \]  

(2.51)

Combining Eq. (2.49), (2.50), (2.51) and (2.48), Eq. (2.48) becomes

\[ I_p = \frac{e}{\hbar} \int dE \text{Tr} \left[ \Sigma_p^< G^r \left( \sum_p (1 - f_p) \Gamma_p \right) G^a - \Sigma_p^> G^r \left( \sum_p f_p \Gamma_p \right) G^a \right], \]

and between any two terminals

\[ I_{pq} = \frac{e}{\hbar} \int dE \text{Tr} \left[ f_p \Gamma_p G^r \left( \sum_m (1 - f_m) \Gamma_m \right) G^a - (1 - f_p) \Gamma_p G^r \left( \sum_n f_n \Gamma_n \right) G^a \right]. \]

(2.52)

The Eq. (2.52) can be applied to multi-terminal case,

\[ I_p = \sum_q \frac{e}{\hbar} \int dE \text{Tr} \left[ \Gamma_p G^r \Gamma_q G^a \right] (f_p - f_q), \]

which sums up all the contributions to terminal \( p \) from all the other terminals.
2.4.2 Bond current

The meaning of the first term on RHS of Eq. (2.40) will be more clear with the following example. With information carried by $G^<$, the current from site to site can be calculated directly as long as the basis set representing Hamiltonian is orthogonal. Since we limit our discussion to steady-state noninteracting case, the sum of second term in Eq. (2.40) is zero obeying the particle conservation, and the remaining part of the equation becomes

$$
\left[ i\hbar \left( \frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2} \right) + \frac{\hbar^2}{2m} (\nabla_{r_1}^2 - \nabla_{r_2}^2) - (V(1) - V(2)) \right] G^<(12) = 0 \quad (2.53)
$$

Take the same limit $1 \rightarrow 2$

$$
i\hbar \lim_{t_2 \rightarrow t_1} \left[ \frac{\partial}{\partial t_1} G^<(12) + \frac{\partial}{\partial t_2} G^<(12) \right] + \nabla \left( \frac{\hbar^2}{2m} \lim_{r_2 \rightarrow r_1} [\nabla_{r_1} - \nabla_{r_2}] G^<(12) \right) = 0 \quad (2.54)
$$

This recovers the continuity equation

$$
e \frac{\partial \rho(r_1, t_1)}{\partial t_1} + \nabla \cdot J(r_1, t_1) = 0
$$

where

$$
J(r_1, t_1) = \frac{q\hbar^2}{2m} \lim_{r_2 \rightarrow r_1} [\nabla_{r_1} - \nabla_{r_2}] G^<(12)
$$

and steady state condition, current density does not depend on time

$$
J(r_1) = \frac{q\hbar^2}{2m} \lim_{r_2 \rightarrow r_1} [\nabla_{r_1} - \nabla_{r_2}] G^<(12).
$$

The particle density can be viewed as the sum of the densities over all sites inside the central region,

$$
\rho = \sum_m \rho_m = \frac{N}{V} = \frac{1}{V} \sum_m N_m
$$

so what happens at a given site $m$ can be described by

$$
\frac{d\rho_m(t)}{dt} = [\nabla \cdot J]_m = \sum_{i=1}^3 \frac{J_i(r_m + \Delta x_i/2) - J_i(r_m - \Delta x_i/2)}{\Delta x_i} \quad (2.55)
$$

in discretized real space representation, where $\Delta x_i, i = 1, 2, 3$ correspond to the small displacement in direction $x, y$ and $z$ respectively.
A cross section needs to be specified to measure the current pass through it. This sometimes introduces additional numerical complexity, if the space discretization is not in square lattice shape, i.e. graphene. To better visualize the situation, we consider one-dimensional transport case along $x$-axis, Eq. (2.55) then read
\[
\frac{d\rho_m(t)}{dt} = -[\nabla \cdot \mathbf{J}]_m = - \frac{J_{m+\frac{1}{2}}(E) - J_{m-\frac{1}{2}}(E)}{\Delta x}
\]
and
\[
\frac{d\rho_m(t)}{dt} = [H, G^<(E)]_{mm} = \frac{q}{A[\Delta x]} \sum_{\sigma\sigma'} \sum_{m'} \left[ G^<_{mm',\sigma\sigma'} H_{mm} - H_{mm'}^{\sigma\sigma'} G^<_{m'm,\sigma'\sigma} \right]
\]
where $A$ is the area of the cross section.

A more general derivation is to work with second quantization operators. The second term of Eq. (2.54) in discretized real space representation at site $m$ reads
\[
i\hbar \frac{d}{dt} G^<_{mm}(E) = [H, G^<_{mm}(E)]
\]
which follows Heisenberg EOM. The charge continuity equation on the lattice site $m$.
\[
\frac{d\hat{N}_m}{dt} = \frac{1}{i\hbar} \left[ \hat{N}_m, H \right]
\]
where $N_m = c^\dagger_{m',\sigma'} c_{m\sigma}$ is the number operator at site $m$ and it leads to
\[
\frac{d\hat{N}_m}{dt} = \sum_{k=x,y,z} \left( \hat{j}_{m,m+e_k} - \hat{j}_{m-e_k,m} \right) = 0.
\]
This equation introduces the bond charge-current operator.

By simply carrying out the commutator calculation in Eq. (2.56), the current which connects site $m$ to site $m'$ can be calculated through equation
\[
J_{mm'} = \sum_{\sigma\sigma'} J_{mm'}^{\sigma\sigma'}
\]
where
\[
J_{mm'}^{\sigma\sigma'} = \frac{e}{i\hbar} \left[ c^\dagger_{m',\sigma'} t_{mm'}^{\sigma\sigma'} c_{m\sigma} - c^\dagger_{m\sigma} t_{m'm}^{\sigma'\sigma} c_{m'} \right]
\]
Finally the quantum-statistical average of the bond charge-current operator in non-equilibrium state can be expressed in terms of $G^<$
\[
\langle J_{mm'}^{\sigma\sigma'} \rangle = \frac{e}{i\hbar} \left[ G^<_{mm',\sigma\sigma'} t_{mm'}^{\sigma\sigma'} - t_{mm'}^{\sigma'\sigma} G^<_{m'm,\sigma'\sigma} \right]
\]
The detail can be found in Appendix 34.
2.4.3 Spin current

In steady-state situations the density matrix can be expressed in terms of lesser GF as in Eq. (2.19) and expectation value of any physical observable can be obtained by

\[ \langle O \rangle = \text{Tr}\left[ \frac{1}{2\pi i} \int OG^\prec(E)dE \right] \]  \hspace{1cm} (2.57)

Inserting Eq. (2.57) into Eq. (2.40), we have a new EOM for physical observable \( O \)

\[ \frac{d\langle O \rangle}{dt} = -i\text{Tr}[\{OH - HO\}G^\prec] - i\text{Tr}[O(\Sigma^r G^\prec + \Sigma^\prec G^a - G^\prec \Sigma^a - G^a \Sigma^\prec)]. \]  \hspace{1cm} (2.58)

If the observable is spin \( \sigma \), the Eq. (2.58) becomes the expression for spin current.

\[ \frac{d\langle S \rangle}{dt} = -i\text{Tr}[\{\sigma H - H\sigma\}G^\prec] - i\text{Tr}[\sigma(\Sigma^r G^\prec + \Sigma^\prec G^a - G^\prec \Sigma^a - G^a \Sigma^\prec)]. \]  \hspace{1cm} (2.59)

In fact, the change of magnetic moment corresponds to torque and similar to the case of charge current, the first term and second term on the RHS correspond to spin-flip in the central region and spin currents flow across boundaries respectively. We can derive a general expression to compute the terminal spin currents at the boundaries by rewriting the second term of Eq. (2.59) with the operator defined in Eq. (2.41)

\[ I_{op} = \text{Tr}[\sigma I_{op}]. \]

One thing worth noticing is that, in the series of derivations for different current formulas, the simplification mainly depends on the invariance under cyclic permutations which are not allowed in any arbitrary permutations. So despite the similarity between charge and spin currents, the similar derivations to WG and LB do not hold.

In a special case that considers only normal metal leads i.e. without magnetization and spin-orbit coupling, where \( \sigma \) commutes with \( \Sigma_i \) \( (i = \tilde{z}, r/a) \). The similar derivation as we did for MW and LB can then be applied along with invariant cyclic permutation property of trace, very similar forms to MW and LB can be obtained

\[ I_{p}^{S_{\alpha}} = \frac{e}{h} \int dE \text{Tr}\{\hat{\sigma}_{\alpha}[\Sigma_{p}^\prec(E)G^\succ(E) - \Sigma_{p}^\succ(E)G^\prec(E)]\}. \]  \hspace{1cm} (2.60)
\[ I_{p}^{S_{\alpha}} = \frac{e}{\hbar} \sum_{q} \int dE \text{Tr} [\hat{\sigma}_{\alpha} \Gamma_{q}(E)G(E)\Gamma_{p}(E)G^\dagger(E)] \{ f_{p}(E) - f_{q}(E) \}. \quad (2.61) \]

In our study Ch 5, we apply Eq. (2.61) to calculate spin current at the probe. This is totally appropriate, since we intentionally choose metal leads to avoid having SOC in the leads, such that spin axes are well defined.

### 2.4.3.1 Non-equilibrium lesser GF

In non-interacting case, lesser GF in Eq. (2.38) can be described in terms of retarded components of GF and embedded SE

\[ G^{\less}(E) = G^r [i f_{L} \Gamma_{L} + i f_{R} \Gamma_{R}] G^{a}, \quad (2.62) \]

and it can be separated into equilibrium and non-equilibrium terms

\[ G^{\less} = \text{Im}[G^r](f_{L} + f_{R}) + i G^r \frac{\Gamma_{L} - \Gamma_{R}}{2} G^{a}(f_{L} - f_{R}). \quad (2.63) \]

So far, we have not made any additional assumption besides non-interacting system in the presence of external bias \( V \), thus Eq. (2.63) is still general. One major advantage of using Eq. (2.63) is that, while performing integration non-equilibrium part only needs to be integrated within bias window and yields zero otherwise.

### 2.5 NEGF Modeling of Phonon Transport

Phonon transport can be understood as heat transfer due to lattice vibration and is treated as spring and mass particle system. In the following example we consider only monoatomic case follows Ref. [113]. It is reasonable to assume for an arbitrary ion, it only moves around its equilibrium points \( R_{i} \), so the position of the ion can be divided into a constant part \( R_{i} \) and the time dependent variation part \( x_{i}(t) \)

\[ \tilde{R}_{i}(t) = R_{i} + x_{i}(t), \]
Combining Newton’s second law and Hook’s law the force equation of the many particle system can be described as

\[ M \ddot{x} = \sum_j K_{ij} \Delta x_{ij} = \sum_j K_{ij} (x_j - x_i) = \sum_j \Phi_{ij} (x_j - x_i) \quad (2.64) \]

By taking advantage of the periodic nature of the crystal, we can choose Bloch-wave type of solution

\[ x^k_i(t) = A e^{i(kR_i - \omega t)}, \]

substitute into Eq. (2.64), the equation becomes

\[ \omega^2 MA = \sum_j \Phi_{ij} e^{ik(R_i - R_j)} A. \]

Since the system is periodic, similar to electronic dispersion, we can consider only one supercell in reciprocal space by setting \( i = 0 \), we then have linear equation in matrix form

\[ \mathbf{D}(k) \mathbf{A} = \omega^2 \mathbf{A}, \quad (2.65) \]

where

\[ \mathbf{D}(k) = \frac{1}{M} \sum \Phi_{0j} e^{iqkR}. \]

The phonon dispersion or vibrational mode can be obtained, by solve the eigenvalue problem of Eq. (2.65). Despite the analogy between electronic and phononic cases, Eq. (2.65) is derived from force instead of energy. The phonon dispersion should be square root of eigenvalue \( \omega^2 \) versus \( k \). It’s straightforward to generalize beyond monoatomic case. One only needs to expand the unit cell.

### 2.5.1 Application to graphene

For the case of graphene as in Fig. (2.3), we can use the symmetry to deduce the number of parameters. By assuming the force constants between every pairs of
**Figure 2.3:** In 3D vibration, there are three independent parameters to describe force tensor between each pair of carbons.

carbons with the same distance and orientation are the same, the force tensor between site A and site B$_1$ in Fig. (2.3) is a 3 by 3 tensor

$$
\Phi^0 = \begin{pmatrix}
\phi_r^{(i)} & 0 & 0 \\
0 & \phi_{tt}^{(i)} & 0 \\
0 & 0 & \phi_{to}^{(i)}
\end{pmatrix}.
$$

The force constants for the pairs with the same distance but different orientations can be obtained by unitary transformation

$$
\Phi^{(i,j)} = U_z^{-1}(\theta_{ij})\Phi^0 U_z(\theta_{ij}),
$$

where $\theta_{ij}$ is the relative angle to A-B$_1$ pair.

To calculate open systems with two semi-infinite leads in stead of bulk system, we do not need to consider the periodicity by FT. As long as we know the “force constants” between each pair of sites, we can write down the force tensors.

$$
\Phi_{ij} = \frac{1}{\sqrt{M_jM_i}} \begin{cases} 
\Phi_{ij} , & i \neq j \\
-\Sigma \Phi_{ii} , & i = j
\end{cases}
$$

(2.66)

The diagonal term $\Phi_{ii}$ is the sum of all the forces acting on the site $i$. 

38
From the observation of the analogy between electronic, the phonon retarded GF is obtained in the same fashion as the electronic one in Eq. (2.15) but with substitutions $H \rightarrow K$ and $E \rightarrow \omega^2 M$ and $\Sigma_{L,R} \rightarrow \Pi_{L,R}$. The dynamical matrix is

\[
D(\omega) = [\omega^2 M - K - \Pi_L(\omega) - \Pi_R(\omega)]^{-1} \tag{2.67}
\]

Here $K$ is the force constant matrix with matrix elements in Eq. (2.66), $M$ is a diagonal matrix with the atomic masses, $\Pi_{L,R}$ are the SEs, and

\[
\Lambda_{L,R}(\omega) = i[\Pi_{L,R}(\omega) - \Pi_{L,R}^\dagger(\omega)]. \tag{2.68}
\]

Phonon transmission then follows a similar form

\[
\mathcal{T}_{ph}(\omega) = \text{Tr}\{\Lambda_R(E)D(\omega)\Lambda_L(E)D^\dagger(\omega)\}, \tag{2.69}
\]

which are mutually connected with electronic case through the Landauer-type formula [141, 107]

\[
\kappa_{ph} = \frac{\hbar^2}{2\pi k_B T^2} \int_0^\infty d\omega \omega^2 T_{ph}(\omega) \frac{e^{\hbar\omega/k_BT}}{(e^{\hbar\omega/k_BT} - 1)^2}, \tag{2.70}
\]

and

### 2.6 ZT Figure of Merit from NEGF for Nanostructure

In nanoscale, instead of using the intrinsic properties of materials such as conductivities and thermoconductivity in Eq. (1.11), the ZT is expressed in terms of conductance $G_e$ and thermal conductance $\kappa$ which both depends on geometry

\[
ZT = \frac{TS^2 G_e}{(\kappa_{el} + \kappa_{ph})}. \tag{2.71}
\]

This new expression implies additional degrees of freedom for us to optimize the ZT.

As new definition of ZT introduced in Eq. (2.71) indicates, all the quantities are related to electronic transport can be calculated through standard NEGF except
for the phonon contribution to thermal conductance should be obtained by Eq. (2.70).

It is known that both charge current and heat current can be driven by external bias ($\Delta \mu$) and temperature gradient ($\Delta T$). They form two coupled linear of equations

$$\begin{bmatrix} \frac{i_e}{e} \\ i_Q \end{bmatrix} = \begin{bmatrix} K_0 & K_1 \\ K_1 & K_2 \end{bmatrix} \begin{bmatrix} \Delta \mu \\ \Delta T/T \end{bmatrix},$$

(2.72)

where

$$K_n = \frac{2}{\hbar} \int \frac{dE}{2\pi} (E - \mu)^n T_e(E) \left( \frac{\partial f}{\partial E} \right).$$

(2.73)

The thermal conductance ($\kappa_{el}$) is measured at the condition $I_e = 0$, therefore

$$\Delta \mu = -\frac{K_1 \Delta T}{K_0}.$$ 

(2.74)

Plug Eq. (2.74) into Eq. (2.72), the second equation becomes

$$I_Q = \frac{1}{T} \left( K_2 - \frac{K_1^2}{K_0} \right) \Delta T.$$ 

By definition, one obtains the equation for $\kappa_{el}$

$$\kappa_{el} = -\frac{I_Q}{\Delta T} \bigg|_{I=0} = (K_2 - K_1^2/K_0)/T,$$

(2.75)

Seebeck coefficient $S$

$$S = -\frac{\Delta \mu}{\Delta T} \bigg|_{I=0} = K_1/eTK_0 = K_1k_B/eK_0,$$

(2.76)

and for electronic conductance $G_e$

$$G_e = e^2K_0.$$ 

(2.77)
3.1 Introduction

The recent explosion of research on graphene—one-atom-thick allotrope of carbon—has been largely focused on its unique electronic structure and transport properties governed by the two-dimensional honeycomb lattice of carbon atoms. [53] Very recently, the exploration of its thermal and thermoelectric properties has been initiated by measuring the thermopower [146] $S$ and phonon thermal conductivity [55] $\kappa_{\text{ph}}$ of large-area graphene. The measured values [146] of $S \simeq 100 \, \mu\text{V/K}$ near the Dirac point (DP), as well as the room-temperature $\kappa_{\text{ph}} \simeq 4000 \, \text{W/mK}$ (averaged over values obtained using different samples and experimental techniques [55]) which outperforms virtually all other known materials, point out that large-area graphene is not suitable for thermoelectric applications.

Thermoelectrics transform temperature gradients into electric voltage and vice versa. Although a plethora of thermoelectric energy harvesting and cooling applications has been envisioned, their usage is presently limited by their small efficiency. [56] Thus, careful tradeoffs are required to optimize the dimensionless figure of merit defined in Eq. 2.71 which quantifies the maximum efficiency of a thermoelectric cycle conversion in the linear-response regime where a small voltage $V = -S\Delta T$ exactly cancels the current induced by the small thermal bias $\Delta T$. This is due to the fact that $ZT$ contains unfavorable combination of $S$, average temperature $T$, electronic conductance $G$ and thermal conductance $\kappa_{\text{el}} + \kappa_{\text{ph}}$. The total thermal conductance has contributions from both electrons $\kappa_{\text{el}}$ and phonons $\kappa_{\text{ph}}$. The devices with $ZT > 1$ are regarded as
good thermoelectrics, but values of $ZT > 3$ are required for thermoelectric devices to compete in efficiency with conventional power generators and refrigerators. [56]. with $ZT \simeq 2 – 3$ that are stable over a broad temperature range and with low parasitic losses is presently considered as a realistic goal. [56].

Thus, a number of proposals have been put forth to evade the problem of high lattice thermal conductivity of large-area graphene that could open a pathway for its thermoelectric applications. For example, large-area graphene could reach $ZT \approx 0.3$ if perforated by the so-called antidot lattice tailored to impede phonon propagation. [109] Switching to quasi-one-dimensional graphene nanoribbons (GNRs) makes possible further enhancement of $ZT$ where it has been predicted that long ($\sim 1 \mu m$) GNRs with zigzag edges and disorder introduced around edges by removing carbon atoms could reach $ZT \simeq 4$ at room temperature. [113] Another route is to engineer structural defects in GNRs that can block phonons while retaining quasiballistic electronic transport. [10]

However, it is more advantageous to search for high-$ZT$ devices among nanowires [11] with well-defined edges since edge or surface disorder can affect electronic conductance significantly. For example, the experiments on etched GNRs with rough edges find Coulomb blockade effects and transport gap much larger than the band gap. [58]

### 3.2 Methodology

In this section we explain details of our models for electronic and phononic subsystems, which are coupled to nonequilibrium Green function formalism (NEGF) as explained in Ch. 2 to treat their elastic quantum transport. [141]

#### 3.2.1 Model Hamiltonian for GNRs

The early theoretical studies of ZGNR-based devices have utilized [62] a simplistic tight-binding model (TBM) with single $\pi$-orbital per site and the nearest-neighbor hopping only, or its long-wavelength (continuum) approximation—the Dirac-Weyl Hamiltonian [63]—valid close to the DP. However, both of these models predict [98, 62] that the transmission function of an infinite homogeneous ZGNR is $T_{el} = 1$
around the DP and that current density profile is peaked [98] in the middle of ZGNR (even though local density of states reaches maximum around the edges [98]). This contradicts first-principles calculations, [65] or TBM with up to third nearest-neighbor [66] hopping parameters fitted to such first-principles calculations, which predict $T_{cl} = 3$ around the DP, as well as that the local current density is mostly confined to flow around the zigzag edges. [67] It is worth mentioning that the majority of recent studies focused on the thermoelectric properties of ZGNRs with edge disorder [113] or finite length graphene antidot lattice [109] have utilized the TBM with nearest-neighbor hopping, so that a possibility to exploit “edge currents” around zigzag or chiral edges for thermoelectric device applications has been overlooked.

Most importantly, the recent experiments have confirmed the existence of “edge currents” in metallic ZGNRs by actually utilizing them to increase the heat dissipation around edge defects and, thereby, rearrange atomic structure locally until sharply defined zigzag edge is achieved. [112] Also, the very recent chemical synthesis [69] of (8,1)-CGNRs via carbon nanotube unzipping method have exhibited properties in sub-nanometer-resolved scanning tunneling microscopy and spectroscopy that can only be explained by the existence of smooth edges supporting edge quantum states (i.e., wave-functions whose probability density is large around the edges). Although ZGNRs [70] or CGNRs [69] are insulating at very low temperatures due to one-dimensional spin-polarized edge states coupled across the width of the nanoribbon, such unusual magnetic ordering and the corresponding band gap is easily destroyed [70] $T > 10$ K.

In our study we adopt the TBM with single $\pi$-orbital per site:

$$\hat{H} = \sum_n \varepsilon_n \hat{c}_n^\dagger \hat{c}_n - \sum_{n,m} t_{n,m} \hat{c}_n^\dagger \hat{c}_m,$$

(3.1)

to describe the electronic subsystem of 20-ZGNR and (8,1)-CGNR in Fig. 3.3. The operators $\hat{c}_n^\dagger$ ($\hat{c}_n$) create (annihilate) electron in the $\pi$-orbital located on site $n$ of the honeycomb lattice whose lattice constant is $a \approx 0.246$ nm. For impurity-free GNRs assumed here, the on-site potential is set to zero $\varepsilon_n = 0$. We consider up to third nearest-neighbor [66] hopping parameters—he $t_{n,dAB}^{n+dAB} = 2.7$ eV, $t_{n,dAA}^{n+dAA} = t_{n,dBB}^{n+dBB} = 0.2$.
Figure 3.1: Bandstructures for (a) 20 ZGNR 3NN TB (b) 20 ZGNR (1NN TB) (c) (8,1)-CGNR 3NNTB (d) (8,1)-CGNR 1NN TB. As the results indicate, the subbands around DP show distortion by considering 3NN and therefore gain nonzero velocities to contribute to the transport.

\[ eV, \quad t_{n+0.5}^{d_{AB}} = 0.18 \text{ eV} \]—which describe the nearest-, next-nearest- and next-next-nearest neighbor hopping, respectively. Since the honeycomb lattice of graphene is composed of two triangular sublattices \( A \) and \( B \), the parameters \( t_{n+0.5}^{d_{AB}} \) and \( t_{n+0.5}^{d_{AB}'} \) describe intersublattice hopping, while \( t_{n+0.5}^{d_{AA}} = t_{n+0.5}^{d_{BB}} \) describes the intrasublattice hopping.

The Fig. 3.1 shows how taking higher order TB affects subband structure around DP. Unlike 3.1 (b) and (d) both are the results obtained by the less accurate 1NN TB Hamiltonians have exact flat bands that correspond to edge state, 3.1 (a) and (c) consider up to 3NN TB Hamiltonian have tilted subbands that indicate nonzero velocities of edge states and can therefore contribute to the transport.

In realistic devices, active region consisting of ZGNR or CGNR of finite length with nanopores will eventually need to be connected to metallic electrodes. However,
since GNR+nanopores devices we analyze are rather long $\sim 1\ \mu m$, and screening takes place over a distance much shorter than the active region, it is justified to use semi-infinite homogeneous ZGNRs or CGNRs as leads for simplicity.

### 3.2.2 Phonon Transmission from Empirical Method

In the elastic transport regime, where electron-phonon and phonon-phonon scattering can be neglected, independent electron and phonon transport quantities can be obtained from NEGF-based formulas whose technical details can be found in Chapter 2. This methodology does not take into account the resistive umklapp phonon-phonon scattering which plays an important role in interpretation of experiments on room-temperature lattice thermal conductivity of large-area graphene. [55] However, this effect, which is easy to describe using the Boltzmann equation but is very expensive computationally within the NEGF formalism, [71] does not play an important role in GNRs depicted in Fig. 3.3 because their width is much smaller than the mean-free path $\ell \simeq 677$ nm due to phonon-phonon scattering in large-area graphene at room temperature. [72]

The widely used methodology to compute $\kappa_{ph}$ of GNRs, as introduced in Chapter 2, couples [113, 73] NEGF to empirical fourth-nearest-neighbor force-constant [74] (4NNFC) or 5NNFC [75] models. The parameters of 4NNFC and 5NNFC models have been refined over the years to reproduce the newly acquired experimental data on the phonon dispersion of graphite, [75] fit DFT calculations [74] for infinite graphene sheets, and satisfy the symmetry imposed conditions (such as rotational invariance [74]). However, direct application of these models to GNRs is not warranted since force constants on the edge carbon atoms will be modified when compared to those in their interior. While the resulting shift in the phonon density of states and the corresponding reduction of $\kappa_{ph}$ are typically assumed to lead only to a minor improvement, [113] here we explicitly compare $\kappa_{ph}$ obtained from these models to first-principles calculations. We also include the Brenner empirical interatomic potential [76] (EIP) which offers much
Figure 3.2: The phonon thermal conductance as a function of temperature for 8-ZGNR with and without a single nanopore (of diameter $D = 0.59$ nm) computed by coupling NEGF formalism to empirical 4NNFC, [74] empirical 5NNFC, [75] Brenner EIP, [76] and DFT (using the basis of local DZP orbitals and PBE exchange-correlation functional). The FC matrix in the case of Brenner EIP and DFT methodology was computed using the GPAW package. [77]

faster numerics than full DFT methodology while being able to match the DFT results for some device geometries. [141, 109]

The first-principles extraction of the FC matrix $K$ is performed via the GPAW package, [77] which is a real space electronic structure code based on the projector augmented wave method. [78] The electronic wavefunctions are expanded in atomic orbitals with a double-zeta polarized (DZP) basis set, and Perdew-Burke-Ernzerhof (PBE) parametrization of the generalized gradient approximation for the exchange-correlation functional is used. The whole active region, composed of a segment of 8-ZGNR with or without a nanopore and few layers of the semi-infinite 8-ZGNR leads, is first relaxed to a maximum force of 0.01 eV/Å per atom. Subsequently, we displace each atom $I$ by $Q_{I\alpha}$ in the direction $\alpha = \{x, y, z\}$ to get the forces $F_{J\beta}(Q_{I\alpha})$ on
atom $J \neq I$ in direction $\beta$. The elements of $K$-matrix are then computed from finite differences

$$K_{I\alpha,J\beta} = \frac{[F_{J\beta}(Q_{I\alpha}) - F_{J\beta}(-Q_{I\alpha})]}{2Q_{I\alpha}}.$$ 

The intra-atomic elements are calculated by imposing momentum conservation, such that $K_{I\alpha,J\beta} = -\Sigma_{J\neq I}K_{I\alpha,J\beta}$. In the case of Brenner EIP-based calculation, we initially relax the active region and then compute the force constant between atom $I$ in direction $\alpha$ and atom $J$ in direction $\beta$ using analytical derivatives, $K_{I\alpha,J\beta} = \partial U/\partial R_{I\alpha} \partial R_{J\beta}$, where $U$ is the total energy. These calculations are also performed using the GPAW package. \cite{77}

Since calculations based on 4NNFC and 5NNFC models do not include passivation of edge carbon atoms with (usually assumed) hydrogen, we do not include hydrogen atoms in DFT- or Brenner EIP-based analysis. This approximation is further justified by the first-principles results of Ref. Tan\textsuperscript{2010} where ZGNRs with and without hydrogen-passivation exhibit virtually the same $\kappa_{ph}$ due to the fact that edge C-C bonds are only slightly perturbed in the presence of hydrogen.

The comparison of different $\kappa_{ph}$ values, computed by coupling NEGF to four different FC matrices in Fig. 3.2, shows that all three empirical models overestimate the phonon thermal conductance of ZGNR+nanopore when compared to first-principles result. In the case of an infinite homogeneous ZGNR, Brenner EIP and DFT calculations yield virtually the same $\kappa_{ph}$ in Fig. 3.2, while 4NNFC and 5NNFC models lead to an overestimate of this quantity. While the first-principles calculations of $\kappa_{ph}$ are too expensive to be applied to our $L \simeq 1.2 \mu$m GNRs (that are also wider than the 8-ZGNR example used in Fig. 3.2), the comparison of different methods applied to a testbed in Fig. 3.2 demonstrates that actual $ZT$ or the proposed ZGNR and CGNR devices will be even higher than the one computed in Fig. 3.7 using 5NNFC model.

### 3.3 Electronic and Phonoic Transport of GNRs with Nanopores

In this section, we exploit the peculiar electronic transport properties of GNRs with zigzag (ZGNR) or chiral (CGNR) edges, illustrated in Fig. 3.3, where the local
Figure 3.3: Schematic view of: (a) 20-ZGNR (composed of 20 zigzag chains); and (b) (8,1)-CGNR with chiral angle $\theta = 5.8^\circ$. The size of the nanopores, assumed to be drilled in the GNR interior away from its zigzag or chiral edges, and the distance between them is illustrated by plotting two repeated supercells of each GNR. The length of these GNRs in the actual calculations is set to $L \simeq 1.2\,\mu\text{m}$ which supports 300 nanopores. Adapted from Ref. [13] charge current density carried by quasiparticles sufficiently close to the DP is peaked around nanoribbons edges as demonstrated in Fig. 3.4(b) and (c). Thus, drilling nanopores [111] in the ZGNR or CGNR interior will not substantially modify such “edge currents.” This is confirmed in Fig. 3.4 using spatial profiles of bond currents, as well as by the transmission function in Figs. 3.5(a) and 3.5(c) which is reduced from $T_{el}(E) = 3$ in infinite homogeneous GNRs to $T_{el}(E) \simeq 2$ around the DP for both ZGNR and CGNR with an array of nanopores. Furthermore, $T_{el}(E)$ around the DP does not change as one increases the length of GNRs because “edge currents” propagate quasiballistically.

The nanopore arrays have been explored before [60] in bulk thermoelectric materials. Furthermore, their fabrication in graphene has been pursued recently by a variety of experimental techniques. [111] Since they break homogeneity of the nanowire, they can substantially impede the propagation of phonons in sufficiently long GNRs. This is corroborated by our results for the phonon transmission function in Figs. 3.6(a) and
Figure 3.4: (Color online) Spatial profiles of local charge currents in (a) 20-ZGNR and (c) (8,1)-CGNR with nanopores for electronic transport close \( E_F = -0.43 \) eV) to the DP. The corresponding current profiles over the transverse cross section of nanoribbons are shown in panels (b) and (d) for both infinite homogeneous GNRs and GNRs with nanopores. Note that the sum of bond currents \[ J_{nm}/V, \] which describe charge flow from site \( n \) to site \( m \) of the honeycomb lattice if hopping \( t_{nm} \neq 0 \) is non-zero between the two sites, gives the conductance \( G = I/V \) (\( I \) is the total current in the leads and \( V \to 0 \) is small bias voltage driving the linear-response transport). Adapted from Ref. [13].

3.6(b) and the corresponding lattice thermal conductance in Figs. 3.6(c) and 3.6(d). The ZGNR and CGNR length is chosen as \( L \approx 1.2 \) µm, which is close to the limit beyond which further increase of \( L \) does not reduce \( \kappa_{ph} \) significantly. The number of nanopores hosted by GNRs of these length is 300.
Figure 3.5: (Color online) (a) The zero-bias electronic transmission $T_{el}(E)$ for an infinite homogeneous 20-ZGNR or 20-ZGNR with periodic array of identical nanopores shown in Fig. 3.3(a). (b) The thermopower at two different temperatures for finite length 20-ZGNR with nanopores. (c) Zero-bias electronic transmission for an infinite homogeneous (8,1)-CGNR or (8,1)-CGNR with periodic array of identical nanopores shown in Fig. 3.3(b). (d) The thermopower at two different temperatures for finite length (8,1)-CGNR with nanopores. Adapted from Ref. [13].

3.4 Figure of Merit

Combining these features of the edge states, we are able to obtain maximum $ZT \simeq 4$ at $T = 77$ K and $ZT \simeq 2$ at $T = 300$ K in Fig. 3.7(a) for the case of 20-ZGNR whose identical nanopores are arranged in a periodic array. The values of $ZT$ for (8,1)-CGNR with periodic array of nanopores are lower, as shown in Fig. 3.7(c). In realistic GNR-based devices, it may be challenging [111] to control the pore arrangement to a high precision assumed in Fig. 3.4. Therefore, in Figs. 3.7(b) and 3.7(d) we assume...
that the pore diameter $D$ is a uniform random variable, such as $D \in [9d_{AB}, 15d_{AB}]$ for pores in 20-ZGNR or $D \in [3d_{AB}, 7d_{AB}]$ for pores in (8,1)-CGNR, as well as that position of nanopores is shifted randomly by $\Delta x \in [-2d_{AB}, 2d_{AB}]$ in 20-ZGNR or $\Delta x \in [-0.5d_{AB}, 0.5d_{AB}]$ in (8,1)-CGNR where $d_{AB} \approx 0.142$ nm is the C-C bond length. This yields maximum value of $ZT \simeq 5$ in our study, at both $T = 77$ K and $T = 300$ K, as shown in Fig. 3.7(b).
3.5 Chapter Summary

In this chapter, based on Ref. [13], we predicted that ZGNRs and CGNRs perforated by an array of nanopores in their interior could serve as the building blocks of highly efficient thermoelectric devices. This is due to the fact that local charge current density is peaked around their edges, as demonstrated explicitly by Fig. 3.4 and confirmed experimentally, [112] so that nanopores do not impede such “edge currents” while drastically reducing phonon conduction in sufficiently long ZGNRs or CGNRs.

In the case of periodic array of identical nanopores, we find that largest $ZT \simeq 4$ at...
$T = 77$ K and $ZT \simeq 2$ at $T = 300$ K can be reached using ZGNR-based devices. On the other hand, if the pore diameter takes a random value within some interval and the distance between the pores is varied, then we find a possibility of even higher figure of merit which can reach $ZT \simeq 5$ at both $T = 77$ K and $T = 300$ K in the case of ZGNR+nanopores devices.
Chapter 4

GIANT THERMOELECTRIC EFFECT IN GRAPHENE-BASED TOPOLOGICAL INSULATORS WITH NANOPORES

4.1 Introduction

The major directions to increase $ZT$ have been focused on either materials with high power factor $S^2G$, such as doped narrow-gap semiconductors; or on minimizing $\kappa_{ph}$ by enhanced phonon scattering in different frequency ranges, such as through nanostructuring [80, 81]. Although nanostructuring has progressed rapidly over the past decade [80, 81], it typically affects bulk electronic states of conventional materials in unfavorable way for thermoelectricity. Thus, the recently discovered topological insulator (TI) materials [83, 84] are of particular interest. The key ingredient in this new class of materials is strong spin-orbit coupling (SOC) which opens an energy gap $E_G$ in the bulk and generates conducting edge (in two dimensions) or surface (in three-dimensions) electron states robust against backscattering off nonmagnetic disorder [84]. Interestingly, $\text{Bi}_2\text{Te}_3$ as one of the prime examples of 3D TIs [83] is well-known to be one of the best bulk thermoelectrics with $ZT \simeq 1$ [82]. Recent efforts have demonstrated how using nanocomposites of bulk and thin film $\text{Bi}_2\text{Te}_3$ can lead to $ZT \simeq 2.5$ [80, 81, 82]. However, none of these findings relies on the topological surface states whose contribution to $S$ and $G$ would be insensitive to disorder introduced to suppress $\kappa_{ph}$.

The very recent attempts to design thermoelectrics based on three-dimensional [85, 86, 87, 88] (3D) or two-dimensional (2D) [89, 90] TIs are mostly qualitative and lack information on their phonons. In addition, many candidate 3D TI materials are unintentionally doped in the bulk which makes difficult to reach the topological transport
Figure 4.1: (a) The central region, consisting of GNR (of width $W = 6.16$ nm and length $L \approx 1.35$ µm used in our calculations) with indium adatoms and (253 along chosen $L$) nanopores, is attached to hot and cold macroscopic reservoirs. The adatoms shown in the center of randomly chosen $n_{ad} \approx 19\%$ hexagons locally enhance SOC within those hexagons which creates an energy gap $E_G$ and a pair of helical quantum states on each edge channeling flow of electrons of opposite spin in opposite directions. We consider GNRs with: (b) perfect edge of zigzag type; or (c) disordered edge created by removing one or two edge carbon atoms in each GNR supercell. (d) The total density of states, where the gap $E_G \approx 17.3$ meV around the Dirac point at zero energy is filled by contribution from the helical edge states, is insensitive to the randomness of adatom configuration or spatial inhomogeneities. Adapted from Ref. [14].

regime where electrons, behaving as massless Dirac fermions, propagate exclusively on their surfaces [88, 83].

In this chapter we present a theoretical design of a high-$ZT$ system, using graphene-based 2D TI nanowire depicted in Fig. 4.1. The device idea can be consider as an extension of Ch. 3, as the edge-sates in TI have different origin and are protected by TRS. For the systems we accurately obtain all quantities in Eq. (2.71) via the (NEGF) methodology as introduced in chapter 2, combined with first-principles calculations of both electronic and phononic band structure.
4.2 Methodology

The electronic subsystem of GNR + heavy-adatoms is described by the tight-binding Hamiltonian of Kane-Mele type [84] with a single $p_z$ orbital per site of the honeycomb lattice

$$\hat{H} = -t \sum_{\langle mn \rangle, \sigma} \hat{c}_m^{\dagger} \sigma \hat{c}_n \sigma + i \lambda_{\text{SO}} \sum_{P} \sum_{\langle \langle mn \rangle \rangle \in P, \sigma, \sigma'} \nu_{mn} \hat{c}_m^{\dagger} \sigma \hat{s}_z^{\sigma} \hat{c}_n \sigma'. \quad (4.1)$$

Here the operator $\hat{c}_m^{\dagger} \sigma$ ($\hat{c}_m \sigma$) creates (annihilates) electron on site $m$ of the lattice in spin state $\sigma$ and $\hat{s}_z$ is the Pauli matrix. The nearest-neighbor hopping $t = 2.7$ eV in the first term in Eq. (4.1) sets the unit of energy scale. The spin-dependent hopping in the second term, where $\nu_{mn} = 1$ for moving counterclockwise around the hexagon and $\nu_{mn} = -1$ otherwise, acts between next-nearest neighbor sites of only those hexagons $P$ of the honeycomb lattice which host indium adatoms. The strength of such SOC, which can be viewed as locally enhanced version of the tiny intrinsic SOC in pristine graphene [96], is parameterized by $\lambda_{\text{SO}}$.

This minimal effective model in Eq. (4.1) is sufficient [94] to fit—using $\lambda_{\text{SO}} = 0.0037t$—the low-energy spectrum obtained from first-principles calculations for $4 \times 4$ graphene supercell with two indium adatoms using VASP package [170], which gives $E_G \approx 11.5$ meV. The electron-core interactions are described by the projector augmented wave (PAW) method [171, 118], and we use Perdew-Burke-Ernzerhof (PBE) [119] parametrization of the generalized gradient approximation (GGA) for the exchange-correlation functional. The cutoff energies for the plane wave basis set used to expand the Kohn-Sham orbitals are 500 eV for all calculations. A $11 \times 11 \times 1 k$-point mesh within Monkhorst-Pack scheme is used for the Brillouin zone (BZ) integration. Structural relaxations and total energy calculations are performed ensuring that the Hellmann-Feynman forces acting on ions are less than 0.005 eV/Å.

Starting from the matrix representation $\mathbf{H}$ of the Hamiltonian in Eq. (4.1) and $\Sigma_{L,R}$, the self-energies introduced by the semi-infinite ideal (without disorder, adatoms or nanopores) ZGNR leads assumed to be attached to 2D TI wire in Figs. 4.1(b) or
4.1(c), the transport calculations follow the standard procedures explained in Chapter
2. We compute the electronic retarded GF, and the level broadening matrices \( \Gamma_{L,R}(E) \)
defined in Eq. (2.15) and Eq. (2.16) allow us to obtain the electronic zero-bias transmission function, Eq. (2.17), which determines all the electronic transport related quantities as introduced in 4.2.

The phononic band structure plotted in Figs. 4.3(a) and 4.3(b) was computed via first-principles methodology using combined VASP [170] and Phonopy packages [116]. The details of VASP calculations are the same as delineated above (except that we use \( 3 \times 3 \times 1 \) \( k \)-point mesh), but here we start from \( 4 \times 4 \) graphene supercell hosting one indium adatom and then enlarged it to \( 8 \times 8 \) supercell in order to capture accurately force constants between a range of neighboring carbon atoms or carbon atoms and indium adatoms. Figures 4.3(a) and 4.3(b) demonstrate appearance of new low energy bands due to the presence of indium adatoms. Although the effect of SOC on phononic band structures can be profound for materials containing heavy elements, especially on surfaces and in thin films (as exemplified by the recent calculations on \( \text{Bi}_2\text{Te}_3 \) and \( \text{Bi}_2\text{Te}_3 \) [120]), the inclusion of SOC in Fig. 4.3(b) generates only a small difference.

For phonon transport, we also follow the standard procedure explained in Ch . To construct the empirical up to fourth-nearest neighbors force constant model, we varied and optimized the FCs to fit as closely as possible the phononic dispersions plotted in Figs. 4.3(a) and 4.3(b). Using the FC matrix \( K \), the diagonal matrix \( M \) containing atomic masses, and self-energies \( \Pi_{L,R} \) of the semi-infinite ideal ZGNR leads [obtained using FCs extracted from the dotted line in Fig. 4.3(a)], we compute the phononic version of the retarded GF, through Eq. (2.67), this together with the level broadening matrices Eq. (2.68), gives the phononic transmission function defined by Eq. (2.69), which determines \( \kappa_{\text{ph}} \) through Eq. (2.70).

The significant difference between \( \kappa_{\text{ph}} \) for GNRs with nanopores but neglecting heavy adatoms [dashed line in Figs. 4.3(e) and 4.3(f)] and \( \kappa_{\text{ph}} \) when heavy adatoms and the corresponding SOC are included [solid line in Figs. 4.3(e) and 4.3(f)] confirms the
necessity for the procedure delineated above. We note that the values of \( \kappa_{\text{ph}} \) [solid lines in Figs. 4.3(e) and 4.3(f)] based on FCs extracted from the phononic band structure of bulk graphene with heavy adatoms in Fig. 4.3(b) are most likely overestimated—more precise FCs would require computationally very expensive procedure which considers a large number of atoms confined within the nanoribbon geometry and in the presence of nanopores [141, 114].

4.3 Electronic Transport

Thus far, 2D TIs have been realized experimentally using cumbersome-to-grow HgTe/CdTe [92], or somewhat more accessible InAs/GaSb [93], quantum wells. A much simpler system—graphene with randomly distributed heavy adatoms—has been conjectured recently via first-principles studies [94, 95]. For example, among many possible heavy adatoms indium and thallium favor high-symmetry position in the center of the hexagons of honeycomb lattice of carbon atoms, while being nonmagnetic and without inducing Rashba SOC that would compete with the emergence of the 2D TI phase [94]. We consider two types of graphene nanoribbons (GNRs) + nanopores as the central region of the two-terminal setup in Fig. 4.1(a), where the edge of GNR is either perfectly ordered and chosen to be of zigzag type in Fig. 4.1(b), or lightly disordered in Fig. 4.1(c). We assume that \( n_{\text{ad}} \approx 19\% \) of GNR hexagons are randomly selected and covered by heavy adatoms of indium, which locally enhance [94] tiny (due to lightness of carbon atoms) intrinsic SOC coupling [96] already present in graphene. This helps to increase the bulk band gap from \( E_G/k_B \approx 0.28 \text{ K} \) [96] in pristine graphene to \( E_G/k_B \approx 200 \text{ K} \) for the chosen adatom type and \( n_{\text{ad}} \).

Remarkably, despite completely random position of heavy adatoms, such disordered system has extremely stable 2D TI phase (which is actually stabilized by the randomness of adatom distribution [97]) that does not require disorder averaging [90] because each sample has the same bulk energy gap \( E_G \approx 17.3 \text{ meV} \). The gap is visible in the DOS in Fig. 4.1(d) which, together with the local DOS in Fig. 4.2(a) and related
Figure 4.2: (a) The local DOS within GNR + heavy-adatoms with a single nanopore in its interior at $E_F = 0.001$ eV. (b) The spatial profile of local currents over the transverse cross section of system in (a). The sum of all bond currents $J_{mn}/V_b$, describing charge flow from site $m$ to site $n$ of the honeycomb lattice at bias voltage $V_b$, gives the electrical conductance $G(E_F) = e^2 T_{el}(E_F)/h$ at zero temperature. Due to helical edge quantum states determining the local DOS and local currents, both quantities have non-zero value only around the edges. (c) and (d) The zero-bias electronic transmission function $T_{el}(E)$ for setups in Figs. 4.1(b) and 4.1(c), respectively. The blue and red lines in panels (b)–(d) are obtained in the presence of both heavy adatoms and nanopores, while black dotted line in panels (c) and (d) is computed after nanopores in the two setups from Fig. 4.1 are removed. (e) and (f) The Seebeck coefficient at $T = 40$ K computed by plugging $T_{el}(E)$ from panels (c) and (d) into Eq. (2.76), respectively. Panels (e) and (f) also show the power factor $S^2G$. Adapted from Ref. [14].
spatial profile of local charge currents in Fig. 4.2(b) (both of which are confined around
the GNR edges), do not contain any signatures of spatial inhomogeneities.

Due to time-reversal symmetry (TRS), the edge currents cannot be reduced
by interior nonmagnetic impurities and vacancies, or by edge disorder like the one
introduced in Fig. 4.1(c). This leads to quantized zero-bias electronic transmission
function $T_{el}(E)$ in the Landauer-Büttiker approach to quantum transport [91], shown
in Figs. 4.2(c) and 4.2(d) for systems illustrated in Figs. 4.1(b) and 4.1(c), respectively. The quantized value $T_{el}(E) = 2$ reflects ballistic transport of electrons through
one-dimensional counter-propagating and spin-polarized edge states, often labeled as
“helical” [84], where TRS forces electrons of opposite spin to flow in opposite directions. The total number of spin-polarized conducting channels on both edges is four, as
illustrated in Fig. 4.1(a), where electrons in two of these channels moving in the same
direction give $T_{el}(E) = 2$ (which was also employed as the experimental signature of
2D TI phase in the early experiments [92]).

We note that dotted horizontal line in Figs. 4.2(c) and 4.2(d) plots $T_{el}(E) = 2$
within a wider range of energies $E_F \in [-0.4 \text{ eV}, 0.4 \text{ eV}]$ for a uniform ZGNR + heavy-
adatoms, which is inherited from the underlying subband structure of GNRs with zigzag
edges (ZGNRs) [99]. However, once nanopores and/or edge disorder are introduced the
quantized $T_{el}(E) = 2$ in Figs. 4.2(c) and 4.2(d) persists only when the Fermi energy
(whose position can be controlled by the gate voltage) is within the bulk gap $E_G$. In
fact, the quantization of $T_{el}(E)$ due to protection of helical edge states by TRS occurs
in a range of energies smaller than the naïvely expected $E_F \in [-E_G/2, E_G/2]$. This is
due to the fact that the width of non-zero LDOS around edges in Fig. 4.2(a) increases
as one moves away from the Dirac point (DP) at $E_F = 0$, so that when states from
opposite edges start to overlap a minigap is created thereby removing the crossing
point in the inset of Fig. 4.1(a) and protection by TRS [99]. Although no symmetry
prevents inelastic backscattering off phonons, quantized transmission is expected to be
insensitive to such inelastic mechanisms to leading order [101].

The Seebeck coefficient is obtained from the electronic transmission function as
defined in Eq. (2.76) and Eq. (2.73). The integrals in Eq. (2.73) also determine [141, 140] electronic charge conductance, \( G_e(E_F) = e^2 K_0(E_F) \) and thermal conductance Eq. (2.75). At low temperatures \( (k_B T \ll E_F) \), Eq. (2.76) can be approximated [140] by the so-called Mott formula

\[
S(E_F) \approx \frac{\pi^2 k_B^2 T}{3e^2 T_{el}(E_F)} \left[ \frac{dT_{el}(E_F)}{dE} \right]
\]

which explicitly shows that large values of \( S \), plotted in Figs. 4.2(e) and 4.2(f), require steep variation of \( T_{el}(E) \). This occurs around \( E \simeq \pm 4.5 \) meV in Figs. 4.2(c) or \( E \simeq \pm 3.8 \) meV in Fig. 4.2(d), where the details of the steep change from \( T_{el}(E) = 2 \) to a vanishing value are also controlled by the nanopores and/or edge disorder. The electron-like (for \( E > 0 \)) and hole-like (for \( E < 0 \)) transport give contributions to \( S \) of opposite sign, so that \( S \equiv 0 \) exactly at the DP (as observed in the experiments on large-area graphene [146]).

While low \( T_{el}(E) \) also helps to increase \( S \), its large values too far away from the DP are irrelevant because thermoelectric performance depends [103] on the power factor \( S^2 G_e \). In fact, \( S^2 G_e \) shown in Figs. 4.2(e) and 4.2(f) has exactly the same shape as the one obtained from the Mahan-Sofo (MS) mechanism [104] based on \( T_{el}(E) = M \delta(E) \), which ensures \( \kappa_{el} \to 0 \) and also sets putative [103, 105] upper limit on \( S^2 G_e \). However, the peak value \( (S^2 G_e)_{max} \approx 8 k_B^2 / h \) per spin-polarized conducting channel in Figs. 4.2(e) and 4.2(f) is larger than the corresponding \( (S^2 G_e)_{max}^{MS} \approx 5.76 k_B^2 / h \) obtained [105] in the MS mechanism. Thus, 2D TI nanowires utilize their two spin-polarized edge conducting channels most efficiently for thermoelectricity, which can be traced back to the (approximate) boxcar functional shape [103] of \( T_{el}(E) \) in Figs. 4.2(c) and 4.2(d). Although only two available conducting channels make the total power factor of a single graphene-based 2D TI nanowire minuscule compared to bulk 3D materials [105], quantities shown in Fig. 4.2 do not scale with the wire width so that one can substantially increase the total \( S^2 G_e \) by patterning very large number of very narrow GNRs connected in parallel.
4.4 Phonon Transport

The phononic band structure in Fig. 4.3(b) for graphene supercell with indium adatoms and SOC switched on is computed from first-principles using VASP simulation package [170]. This serves as in input for quantum transport calculation of the phononic transmission function $T_{ph}(\omega)$, plotted in Figs. 4.3(c) and 4.3(d), and $\kappa_{ph}$ in Figs. 4.3(e) and 4.3(f). They are mutually connected through the Landauer-type formula Eq. (2.70). The very high Debye temperature ($\simeq 2100$ K) of graphene necessitates quantum transport treatment of phonon propagation captured by Eq. (2.70). However, this formula does not take into account the resistive umklapp phonon-phonon scattering. Nevertheless, such effect is expected to be irrelevant for GNRs depicted in Fig. 4.1 because their width is much smaller than the mean-free path due to phonon-phonon scattering (e.g., $\ell \simeq 600$ nm or $\ell \simeq 100$ nm at room temperature for graphene freely suspended or lying on SiO$_2$ substrate, respectively [108]).

Although the thermal conductivity of freely suspended large-area graphene at room temperature is among the highest of any known material [108], it decreases significantly when this 2D material is in contact with a substrate or confined into nanoribbons. We further reduce the ballistic value of $\kappa_{ph}$ for GNRs, plotted as dotted line in Figs. 4.3(e) and 4.3(f), by two orders of magnitude via introduction of nanopores, as demonstrated by solid line in Figs. 4.3(e) and 4.3(f). This reduction occurs in sufficiently long GNRs, $L \approx 1.35$ $\mu$m, where we check that decrease of $\kappa_{ph}$ with increasing $L$ and the number of nanopores saturates around this length [109]. Although nanopore arrays have been considered theoretically as a way to reduce lattice thermal conductivity of bulk materials [87, 110], graphene with its high mechanical stability makes it actually possible to fabricate nanopore arrays by a variety of recently developed techniques [111].

4.5 Optimized ZT

The combination of electronic and phononic transport quantities shown in Figs. 4.2 and 4.3, respectively, generates maximum $ZT_{max} \approx 3$ for both setups Figs. 4.1(b) and
Figure 4.3: The phononic band structure and thermal conductance. Solid line in both panels (a) and (b) plots the phonon band structure for $4 \times 4$ supercell of graphene, with one indium adatom per supercell, where SOC is not included. Dotted line in panel (a) plots phonon dispersion for the same supercell of pristine graphene without any adatoms, while dashed line in panel (b) includes both adatoms and SOC. All curves are obtained from first-principles calculations using VASP [170]. (c) and (d) The phononic transmission function $\mathcal{T}_{ph}(\omega)$ for setups in Fig. 4.1(b) and 4.1(c), respectively, where solid lines are based on the dashed line dispersion in panel (b). Dotted lines in panels (c) and (d) plot quantized $\mathcal{T}_{ph}(\omega)$ for an infinite homogeneous ZGNR of the same width and edge shape as in Fig. 4.1, but without any adatoms, nanopores or edge disorder. (e) and (f) Solid and dotted lines plot the phononic thermal conductance, computed by plugging the values of $\mathcal{T}_{ph}(\omega)$ from solid and dotted lines in panels (c) and (d), respectively, into Eq. (2.70). Additional dashed line in panels (e) and (f) plots $\kappa_{ph}$ for the same setups as in Figs. 4.1(b) and 4.1(c), respectively, but where the presence of heavy adatoms and local SOC induced by them is neglected. Adapted from Ref. [14].
Figure 4.4: The thermoelectric figure of merit. Dashed line plots $ZT$ at $T = 36$ K for GNR + heavy-adatoms + nanopores setup in Fig. 4.1(b), while solid line takes into account additional edge disorder for setup in Fig. 4.1(c). The inset shows dependence of the peak values of $ZT$ on temperature. Adapted from Ref. [14].

Note that the main purpose of introducing light edge disorder in Fig. 4.1(c) is to highlight how GNR-based 2D TIs do not require large effort [112] to control the position of edge carbon atoms, rather than to introduce additional phonon scattering off edge roughness [113, 114]. The value of $ZT$ can be tuned significantly by changing the charge density (i.e., the corresponding $E_F$) via the gate voltage, where the advantage over other recent proposals [141, 113, 114] for thermoelectrics based on topologically trivial GNRs is insensitivity of the position of the peaks of $ZT(E_F)$ to microscopic details of the system. That is, the energy $E_F$ at which $ZT$ reaches maximum in Fig. 4.4 depends only on $T$ and $E_G$ (governed by the adatom coverage and adatom type).

4.6 Chapter Summary

In this chapter, based on Ref. [14], we designed a high-$ZT$ thermoelectric where graphene nanoribbons with arbitrary shaped edges and nanopores in their interior are covered with heavy adatoms of indium using quantum transport modeling combined...
with first-principles electronic and phononic band structure calculations. The adatoms provide sufficiently strong local SOC in some fraction of randomly chosen hexagons, so that such inhomogeneous SOC opens both the bulk band gap $E_G \approx 17.3$ meV (for $n_{ad} \approx 19\%$ of hexagons covered) and generates topologically protected helical edge states. The electronic transmission through helical edge states in the form of the (approximate) boxcar function of width $\lesssim E_G$ generates power factor $S^2 G$ per conducting channel which is actually larger than the one obtained from the celebrated Mahan-Sofo mechanism using delta function transmission. This feature combined with two orders of magnitude reduction of phononic thermal conductance by the nanopores leads to the thermoelectric figure of merit for this system $ZT \simeq 3$ at low temperatures $T \simeq 40$ K. Since the existing bulk thermoelectric materials are very inefficient at low operating temperatures $T \sim 10$ K where they give $ZT \lesssim 0.01$ [81, 82], TI-based high-$ZT$ thermoelectrics at low temperatures designed by our study are very attractive for applications like radioisotope thermoelectric generators on spacecrafts or cooling of electronic satellite components. Although bulk materials are deemed necessary for large-scale power generation [80], GNRs underlying our proposal are single-atom-thick and with electronic transport properties which do not scale with their width, so that very high packing density [105] of GNRs connected in parallel is possible within a 3D volume. Other choices [95] for heavy adatoms include Os, Ir and Cu-Os or Cu-Ir dimers, which could generate larger gap $E_G \gtrsim 0.2$ eV using smaller adatom coverage $n_{ad} \simeq 2\%$ (via different hybridization mechanism between carbon and adatom orbitals than in the case of indium or thallium), thereby making it possible to also tune the optimal operating temperature. Interestingly, if the boxcar-shaped transmission function is preserved at finite bias voltage, the same system could optimize efficiency at finite power output in the nascent field of nonlinear thermoelectrics [115].
Chapter 5

SPIN-SEEBECK EFFECT ON THE SURFACE OF A TOPOLOGICAL INSULATOR DUE TO NONEQUILIBRIUM SPIN-POLARIZATION PARALLEL TO THE DIRECTION OF THERMALLY DRIVEN ELECTRONIC TRANSPORT

5.1 Introduction

The spin-Seebeck effect (SSE) is a recently observed phenomenon where spin current or spin accumulation is induced by a temperature gradient applied across a ferromagnetic material. [121, 122]. At first sight, the SSE appears to be a counterpart of the traditional charge-Seebeck effect (CSE) where the temperature gradient across a conductor induces electrical current (in closed circuits) or voltage (in open circuits). However, the SSE has been observed in a surprisingly wide range of materials, including ferromagnetic insulators where the CSE does not exist. [121, 122]

In the so-called transverse SSE measurement geometry, illustrated in Fig. 5.1, a temperature gradient applied longitudinally over a strip of magnetic material induces a spin signal detected by measuring voltage $V_{\text{ISHE}}$ generated via the inverse spin Hall effect (ISHE) in the nonmagnetic metallic probe (such as Pt) attached on top of the strip across its width. The voltage signal $V_{\text{ISHE}}$ is found to be approximately linear (in ferromagnetic metals and insulators [122]) or a hyperbolic sine (in ferromagnetic semiconductors [123]) function of the probe position in the longitudinal direction over the length of several millimeters. Since this is surprisingly long when compared to the usual electronic spin-dependent length scales, recent theories of the SSE have focused on the interplay of magnons and phonons out of equilibrium. [124].

On the other hand, the role of spin-polarized conduction electrons in SSE generation has been much less explored. The need for this has been prompted by the very
Figure 5.1: (Color online) Schematic view of the three-terminal junction where a thin film of Bi$_2$Se$_3$ is attached to two nonmagnetic metallic leads kept at different temperatures $T_1 > T_2$. The SSE voltage signal $V_{\text{ISHE}}$ is measured across the edges of the third nonmagnetic metallic lead attached to the top TI surface a function of its position along the x axis. This lead acts as voltage probe attached to a macroscopic reservoir at potential $V_3$ to ensure zero net charge current through it. We orient quintuple layers of Bi$_2$Se$_3$ at an angle of 45° with respect to the direction (x axis) of electron propagation on the TI surface that is orthogonal to these layers. Adapted from Ref. [15].

Recent experimental [125] unveiling of “giant” $V_{\text{ISHE}}$ (up to a thousand times larger than observed in measurements on magnetic materials [122]) in the transverse SSE setup where a nonmagnetic semiconductor InSb was placed in a large longitudinal (parallel to the temperature gradient) external magnetic field. Aside from the magnitude of $V_{\text{ISHE}}$, which is speculated to arise in the interplay of spin-orbit coupling (SOC) and enhancement of the phonon drag contribution to both the spin- and charge-Seebeck coefficients for electrons pushed into the ultraquantum limit by the applied magnetic field, another puzzle for SSE theories posed by Ref. [125] is that $V_{\text{ISHE}}$ did not change sign under the reversal of the magnetic field direction.

Here we show that basic phenomenology of the experiment in Ref. [125] can
be recreated without applying any external magnetic field. The role of the magnetic field was to spin polarize electrons in the direction of transport (by means of Zeeman splitting, further amplified by SOC in InSb), as well as to confine their spatial motion so that electrons spiral in the $yz$ plane (with cyclotron orbits that are quantized into Landau levels) as they translate along the $x$ axis. In order to generate the same spin polarization along the electron transport direction, we employ a thin film of a recently discovered three-dimensional topological insulator (3D TI) material, such as $\text{Bi}_2\text{Se}_3$ assumed here, which is attached to three normal (i.e., nonmagnetic) metal (N) leads, as illustrated in Fig. 5.1.

The 3D TI materials [84] possess a usual band gap in the bulk, while hosting metallic surfaces whose low-energy quasiparticles are massless Dirac fermions with spins locked to their momenta due to strong Rashba-type SOC. [155] In particular, $\text{Bi}_2\text{Se}_3$ realization of TI is a strongly anisotropic material composed of quintuple layers (QLs) of Bi and Se atoms, where one QL consists of three Se layers strongly bonded to two Bi layers in between. [84] While $\text{Bi}_2\text{Se}_3$ is always unintentionally $n$-type doped by Se vacancies, charge carriers in the bulk of films of thickness $\lesssim 10$ nm can be completely removed by a gate electrode. [127]

One of the key effects [129, 130, 131, 132] that 3D TIs bring into spintronics is nonequilibrium spin density in the direction transverse to injected unpolarized charge current, which is much larger [131] than in the case [154, 134, 135] of two-dimensional electron gases (2DEGs) with the Rashba SOC. Our first principal result [see Figs. 5.3(a) and 5.3(b)] demonstrates that additional component of nonequilibrium spin density and polarization can be induced in the direction parallel to injected charge current, on the proviso that QLs are oriented as shown in Fig. 5.1. Our second principal result [see Figs. 5.4(a)–(c)] shows that this indeed makes possible non-zero SSE signal in three-terminal geometry of the junction depicted in Fig. 5.1.

The chapter is organized as follows: In Sec. 5.2 we employ the nonequilibrium Green function (NEGF) formalism [91] to obtain the Landauer-Büttiker (LB-type) formula for spin currents in the leads of a multi-terminal quantum-coherent junction.
driven by both voltage bias and temperature bias in the linear-response regime. Section 5.3 explains our Hamiltonian model for the TI film, as well as the construction of the retarded GF for an open system TI + semi-infinite N leads where the orientation of QLs shown in Fig. 5.1 is taken into account through the self-energy entering the retarded GF. In Sec. 5.4 we analyze the spin-polarization vector of the charge current, as well as charge conductance, for a TI film attached to two N leads. In Sec. 5.5 we predict the magnitude of the voltage signal generated across the third N lead in the three-terminal junction depicted in Fig. 5.1, while also contrasting its features with those of conventional charge and spin-dependent Seebeck coefficients that would be measured between terminals 1 and 2 in Fig. 5.1. We conclude in Sec. 5.6.

5.2 Spin Currents in Multi-terminal Quantum-coherent Conductors Driven by Voltage and Temperature Biases

We use the same units for the total charge \( I_p = I_p^\uparrow + I_p^\downarrow \) and total spin \( I_p^{S\alpha} = I_p^\uparrow - I_p^\downarrow \) currents flowing through lead \( p \), which are constructed from spin-resolved charge currents \( I_p^\sigma \) with the spin quantization axis for \( \sigma = \uparrow, \downarrow \) chosen along \( e_\alpha \). There has been a lively debate [136, 137] in the literature on the proper derivation of the multi-terminal LB-type formula [138] which connects spin current \( I_p^{S\alpha} \) flowing through the semi-infinite ideal (i.e., charge- and spin-interaction-free) metallic lead \( p \) attached to a quantum-coherent conductor due to voltages \( V_p \) applied at the external macroscopic reservoirs into which the leads terminate at infinity. The debate was spurred by one of the early derivations, [139] using the traditional scattering matrix framework [19], which predicted unphysical \( I_p^{S\alpha} \neq 0 \) in equilibrium \( V_p = \text{const} \).

Here we derive LB-type formula for spin currents driven by both voltage and temperature biases. We bypass the issue of unphysical equilibrium total spin currents [136, 137] by starting from the outset from a general NEGF-based expression Eq. (2.61) for spin current in lead \( p \).
\[ I_{p}^{S_{\alpha}} = \frac{e}{\hbar} \sum_{q} \int dE \, \text{Tr} [\hat{\sigma}_{\alpha} \Gamma_{q}(E) \mathbf{G}(E) \Gamma_{p}(E) \mathbf{G}^{\dagger}(E)] \{ f_{p}(E) - f_{q}(E) \} \]

This is actually the difference of spin-resolved charge currents given by the well-known Meir-Wingreen formula Eq.(2.47), where \((\hat{\sigma}_{x}, \hat{\sigma}_{y}, \hat{\sigma}_{z})\) is the vector of the Pauli matrices.

The two fundamental objects of the NEGF formalism are the retarded \(\mathbf{G}(E)\) and the lesser GFs defined in Eq. (2.15) Eq. (2.21)—describe the density of available quantum states and how electrons occupy those states, respectively. [91]

In the elastic transport regime the total spin current in lead \(p\) can be obtained by Eq. (2.61) expanding \(f_{p}(E) - f_{q}(E)\) to linear order in \(T_{p} - T_{q}\) and \(V_{p} - V_{q}\), we finally get the desired multiterminal LB-type formula for spin current driven by both temperature and voltage biases in the linear-response regime:

\[
I_{p}^{S_{\alpha}} = \frac{e^{2}}{\hbar} \sum_{q} \int dE \, \text{Tr} [\hat{\sigma}_{\alpha} \Gamma_{q}(E) \mathbf{G}(E) \Gamma_{p}(E) \mathbf{G}^{\dagger}(E)]
\times \left\{ \frac{\partial f}{\partial E} \left[ \frac{E - E_{F}}{eT} (T_{p} - T_{q}) - (V_{p} - V_{q}) \right] \right\}.
\] (5.1)

Note that the usual expression [138] for the total charge current \(I_{p}\) in lead \(p\) is the same as Eq. (5.1), except that \(\hat{\sigma}_{\alpha} \mapsto \hat{\sigma}_{0}\) where \(\hat{\sigma}_{0}\) is the unit \(2 \times 2\) matrix.

Applying Eq. (3) to the three-terminal junction in Fig. 5.1 gives

\[
I_{p}^{S_{\alpha}} = \frac{e^{2} L_{y}}{2\pi \hbar} \sum_{q=1}^{3} \int \int dE \, dk_{y} \, \text{Tr} [\hat{\sigma}_{\alpha} \Gamma_{q}(E) \mathbf{G}(E) \Gamma_{p}(E) \mathbf{G}^{\dagger}(E)]
\times \left\{ \frac{\partial f}{\partial E} \left[ \frac{E - E_{F}}{eT} (T_{p} - T_{q}) - (V_{p} - V_{q}) \right] \right\},
\] (5.2)

and \(V_{1} = V_{2} \neq V_{3}\). By imposing the condition \(I_{p} = 0\) in one of the leads, such as lead \(p = 3\) in Fig. 5.1, the linear system of equations in Eq. (4) can be solved to find voltage \(V_{3}\) that has to be applied to convert this lead into a voltage probe employed in SSE experiments.

Using the spin-dependent transmission function

\[
\mathcal{T}_{21}^{\alpha}(E) = \text{Tr} [\hat{\sigma}_{\alpha} \Gamma_{2}(E) \mathbf{G}(E) \Gamma_{1}(E) \mathbf{G}^{\dagger}(E)]
\]
of the two-terminal version of junction in Fig. 5.1, we can compute the following integrals [140, 141]

\[ K_n^\alpha(\mu) = \frac{1}{\hbar} \int_{-\infty}^{\infty} dE \mathcal{T}^\alpha_{21}(E)(E - E_F)^n \left( -\frac{\partial f}{\partial E} \right), \]  

(5.3)

which yield the three spin-dependent Seebeck coefficients \( S_{\text{spin}}^\alpha = K_1^\alpha/(eTK_0^\alpha) \) for the chosen Pauli matrix \( \hat{\sigma}_\alpha \), or the charge-Seebeck coefficient \( S_{\text{charge}} = K_1/(eTK_0) \) when \( \hat{\sigma}_\alpha \mapsto \hat{\sigma}_0 \).

### 5.3 Hamiltonian and Retarded Green’s Function for Multi-terminal TI-based Junctions

The junction in Fig. 5.1 is modeled on the simple cubic lattice with lattice spacing \( a \), which is assumed to be periodically repeated in the \( y \)-direction. The TI thin film has finite length \( L_x \), while it is sufficiently thick \( L_z = 30a \) to ensure no coupling between the top and the bottom metallic surfaces which penetrate as evanescent states into the bulk of the TI film and whose overlap would open a minigap at the Dirac point (DP) in ultrathin films. [166] The TI thin film is described using the minimal tight-binding Hamiltonian with four orbitals per site [143]

\[
H_{\text{TI}} = \sum_{n,k_y}(c_{n,k_y}^\dagger [M_0 + C\mathbf{1} + T_y e^{ik_ya} + T_y^1 e^{-ik_ya}]c_{n,k_y})
\]  

(5.4)

\[
+ \sum_{n,k_y,a=x,z}(c_{n,k_y}^\dagger T_\alpha c_{n+a,k_y} + \text{H.c.}),
\]  

(5.5)

where \( T_\alpha = B\hat{\sigma}_z \otimes \sigma_0 - iA\hat{\sigma}_x \otimes \hat{\sigma}_\alpha/2; \) \( M_0 = (M - 6B)\hat{\sigma}_z \otimes \hat{\sigma}_0; \) \( \hat{\sigma}_0 \) is the unit \( 2 \times 2 \) matrix; and \( \mathbf{1} = \hat{\sigma}_0 \otimes \hat{\sigma}_0 \). Here \( c_{n,k_y} = (\hat{c}_{+\uparrow}, \hat{\hat{c}}_{+\downarrow}, \hat{\hat{c}}_{-\uparrow}, \hat{\hat{c}}_{-\downarrow})^T \) annihilates electron in different orbitals on site \( n \) with the transverse momentum \( k_y \). The numerical values of the parameters are chosen as: \( M = 0.3 \text{ eV}; \) \( A = 0.5 \text{ eV}; \) and \( B = 0.25 \text{ eV}. \) The bottom of the band of TI is shifted by \( C = 3.0 \text{ eV}. \)

When applying the in-plane external magnetic field \( B_x \), the Zeeman term changes to \( M_0 \mapsto M_0 + \Delta \hat{\sigma}_0 \otimes \hat{\sigma}_x \) where \( \Delta = g_\parallel \mu_B B_x \). We also apply an additional Zeeman
term $\Delta \hat{\sigma}_0 \otimes \hat{\sigma}_z$ with $\Delta = 0.5$ eV to the bottom TI surface in order to split its Dirac cone and block current through it.

The semi-infinite N leads made of nonmagnetic metallic material are described by a tight-binding Hamiltonian with a single orbital per site

$$H_N = \sum_{n,\sigma,k_y} \varepsilon_{n,k_y} \hat{c}_{n\sigma,k_y} \hat{c}_{n\sigma,k_y}^\dagger - \gamma \sum_{n,\sigma,k_y,\alpha=x,z} (\hat{c}_{n\sigma,k_y} \hat{c}_{n+\epsilon\alpha,\sigma,k_y} + H.c.). \quad (5.6)$$

where the operators $\hat{c}_{n\sigma}^\dagger$ ($\hat{c}_{n\sigma}$) create (annihilate) electron with spin $\sigma$ on site $n$ with the transverse momentum $k_y$. The kinetic energy $\varepsilon_{n,k_y} = -2\gamma \cos k_y a$ is equivalent to an increase in the on-site energy, and the nearest neighbor hopping is set at $\gamma = 1.0$ eV.

The evaluation of Eq. (5.2) relies crucially on the construction of the proper coupling matrix $\tau$ between $\hat{H}_{TI}$ in Eq. (5.4) and $\hat{H}_N$ in Eq. (5.6) since $\tau$ enters into the retarded GF. For example, the Hamiltonian of the composite system semi-infinite-N-lead-1 + TI-thin-film is given by

$$H_{N+TI} = \begin{pmatrix} H_N & \tau_1 \\ \tau_1^\dagger & H_{TI} \end{pmatrix}. \quad (5.7)$$

The retarded GF of the TI film alone, viewed as an open quantum system, is defined by [138, 91]

$$G(E) = [E - H_{TI} - \Sigma_1(E)]^{-1}. \quad (5.8)$$

Here the retarded self-energy introduced by the semi-infinite N lead 1 is

$$\Sigma_1(E) = \tau_1^\dagger \cdot g(E) \cdot \tau_1, \quad (5.9)$$

and ($\eta$ is positive infinitesimal)

$$g(E) = [E + i\eta - H_N]^{-1}, \quad (5.10)$$

is the retarded GF [138, 91] of N lead 1. The same procedure would be repeated when more than one N lead is attached to the TI thin film to get

$$G(E) = [E - H_{TI} - \Sigma_1(E) - \Sigma_2(E) - \Sigma_3(E)]^{-1}, \quad (5.11)$$
for the three-terminal junction in Fig. 5.1.

**Figure 5.2:** (Color online) The equilibrium expectation values of the spin operator for surface-state electrons on the Bi and Se sublattices as a function of the in-plane momentum $p = (p_x, p_y)$. The Bi$_2$Se$_3$ crystal is assumed to fill the half-space $z < 0$, so that its infinite surface in the $xy$-plane is orthogonal to its QLs that are also oriented perpendicularly to the unit vector $\mathbf{n} = (e_x + e_y)/\sqrt{2}$, as illustrated in Fig. 5.1. Adapted from Ref. [15].

The conventionally assumed identical orientation of spin (i.e., expectation value of the spin operator) on Bi and Se sublattices, where spin on both sublattices follows “helical” texture shown in Fig. 5.1 for Bi sublattice, is valid only on the (111) surface of the TI crystal that coincides with the plane of the QL. For TI surface other than (111), the spin operators for electrons residing on the Bi and Se sublattices are inequivalent. [144] Nevertheless, for interpreting spin- and angular- resolved photoemission spectroscopy (spin-ARPES) experiments [84] or for attaching the TI sample to N leads, where photoelectrons or electrons injected or absorbed from N leads do not carry a sublattice index, it is advantageous to select the standard relation between the Pauli matrices and the spin operator, $\hat{S} = \hbar \hat{\sigma}/2$.

The spinors $u_{\text{Bi}}$ and $u_{\text{Se}}$, associated with each sublattice when inequivalent spin operators are used, have to be unitarily transformed [144] to $u_{\text{Bi}} \mapsto u_{\text{Bi}}$ and $u_{\text{Se}} \mapsto i(\hat{\sigma} \cdot \mathbf{n})u_{\text{Se}}$. Here $\mathbf{n}$ is the unit vector normal to the QL. This specifies $2 \times 4$ coupling matrices for each N lead $p = 1$–$3$ as

$$
\tau_p = \begin{pmatrix}
t_{\text{Bi}} & 0 & in_z t_{\text{Se}} & (in_x + n_y) t_{\text{Se}} \\
0 & t_{\text{Bi}} & (in_x - n_y) t_{\text{Se}} & -in_z t_{\text{Se}}
\end{pmatrix}.
$$

(5.12)
For the setup in Figs. 5.1 and 5.2 we use \( n = (110) \) in Eq. (5.12). In addition, the hopping parameters between orbitals in the N leads and those on the Bi or Se sublattice are chosen as: \( t_{	ext{Bi}} = 0.4 \) eV for leads 1 and 2; \( t_{	ext{Bi}} = 0.45 \) eV for lead 3; \( t_{	ext{Se}} = 0.8 \) eV for leads 1 and 2; and \( t_{	ext{Se}} = 0.9 \) eV for lead 3.

5.4 Nonequilibrium Spin Polarization in Two-terminal TI-based Junctions

![Figure 5.3:](image)

Figure 5.3: (Color online) The components \( P_x, P_y, P_z \) and the magnitude \( |P| \) of the spin-polarization vector of charge current outflowing into lead 2, after unpolarized charge current is injected from lead 1 into the two-terminal version of junction in Fig. 5.1. These are plotted vs. (a) the Fermi energy \( E_F \), or (b) the length \( L_x \) of the TI film. The linear response charge conductance of the same two-terminal junction vs. (c) the Fermi energy, or (d) the inverse length \( 1/L_x \). Adapted from Ref. [15].

When unpolarized charge current is injected from N lead 1, the ensemble of outgoing spins in N lead 2 of N1/TI/N2 two-terminal junction is characterized by the
spin density matrix

\[ \hat{\rho}_{\text{spin}}^{\text{out}} = \frac{1}{2} (1 + \mathbf{P} \cdot \hat{\sigma}) \]

whose polarization vector is given by \( P_\alpha = I_p^{S_\alpha} / I_p \) for such setup. We first demonstrate in Fig. 5.3(a) that an unpolarized charge current injected from N lead 1 into the top surface of a two-terminal junction (i.e., when the third N lead in Fig. 5.1 is removed) will exit into N lead 2 with non-zero spin-polarization vector \( \mathbf{P} = (P_x, P_y, P_z) \) which includes a component \( P_x \neq 0 \) in the direction of transport. Since Dirac fermions on the opposite surfaces of TI have opposite chiralities, which generates opposite spin-polarization for currents flowing through the top and bottom surface that would cancel in the total current in N lead 2, we block transport through the bottom surface by introducing an energy gap into its Dirac cone (e.g., due to coating by magnetic film [128]). Figure 5.3(b) reveals that current spin-polarization is established on a very short length scale of \( \approx 10 \) lattice spacings, so that this mechanism can operate near or under the contacts with N leads even in the presence of inevitable spin or charge dephasing mechanisms (the spin dephasing time for the in-plane spin components on the surface of TI is the same as the momentum relaxation time [129]).

Although the \( g \)-factor in the Zeeman term \(-g_\parallel \mu_B B_x \hat{\sigma}_x\) introduced by the external magnetic field applied parallel to the top and bottom surfaces of Bi\(_2\)Se\(_3\) is renormalized [155] \( g_\parallel = 23 \) due to strong SOC effects, [145] changing its sign \( B_x \rightarrow -B_x \) has virtually no effect on the spin-polarization vector governed by the strong surface SOC. For example, \( P_x = 0.605 \) (at \( E_F = 3.04 \) eV selected for illustration) in zero magnetic field \( B_x = 0 \) is virtually indistinguishable from \( P_x = 0.609 \) at large external magnetic field \( B_x = -10 \) T applied opposite to the direction of electron transport. Thus, the same mechanism—momentum-dependent effective magnetic field associated with SOC which is much stronger than any external one—can be invoked to explain why \( V_{\text{ISHE}} \) did not change sign upon reversing \( B_x \rightarrow -B_x \) in the experiment of Ref. [125]. This requires that 2D hole gas (see Fig. 1 in Ref. [125]) formed at the interface between Pt probe and InSb has SOC with a component of its effective magnetic field pointing in
the direction of transport (as it would be the case in the presence of the Dresselhaus SOC [155]).

At first sight, the surface of 3D TI is expected to spin-polarize charge current in the transverse direction only, [129, 130, 131, 132] \( P_y \neq 0 \) while \( P_x = P_z = 0 \). This is due to the fact that Dirac cone energy-momentum dispersion on the surface of TI and spin-orthogonal-to-momentum locking within it, as observed in spin-ARPES experiments, [84] is routinely described by an effective 2D Hamiltonian [84, 143] taking form of the massless Rashba model, [155]

\[
\hat{H} = v_F (\hat{\sigma} \times \hat{p}) \cdot \hat{e}_z. \tag{5.13}
\]

Here \( v_F \) is the Fermi velocity, \( \hat{p} = (\hat{p}_x, \hat{p}_y) \) is the momentum operator in 2D and \( \hat{\sigma} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z) \) is the vector of the Pauli matrices. Thus, when charge current flows on the TI surface longitudinally, this Hamiltonian predicts induction of nonequilibrium transverse spin density [129, 130, 131] \( S_y \) and the corresponding spin-polarization \( P_y \neq 0 \) of the current outflowing into the attached N leads. [132] This effect is the counterpart of the one predicted long ago, [154, 134] and observed in recent experiments, [135] for the Rashba spin-split 2DEGs, except that \( S_y \) on the TI surface is larger by a factor \( \hbar v_F / \alpha \gg 1 \) (the Rashba SOC term in 2DEGs is given by [155] \( \alpha (\hat{\sigma} \times \hat{p}) \cdot \hat{e}_z / \hbar \)).

However, such conclusion is an artifact of a naïve identification of \( \hat{\sigma} \) operator with the true electron spin, which becomes invalid when the TI surface does not co-incide with the QL plane. [144] Since low-energy Hamiltonian models [143] of Bi\(_2\)Se\(_3\) operate with the pseudospin degree of freedom describing states with support on the Bi and Se sublattices, the attachment of the TI thin film to semi-infinite N leads that inject or absorb electrons carrying real spin requires to properly interpret their mutual coupling when studying spin-dependent electron transport. Aligning QLs of Bi\(_2\)Se\(_3\) perpendicularly to the infinite TI surface in \( xy \)-plane, and at an angle of 45° with respect to the \( yz \)-plane, will generate difference in spin textures on the two sublattices shown in Fig. 5.2. This motivates our proposal for the junction setup in Fig. 5.1, where
longitudinal spin polarization \([P_x \neq 0 \text{ in Fig. 5.3(a) and (b)}]\) is driven by the surface-state electrons on the Se sublattice. This emerges in addition to the amply studied (in the diffusive \([129, 130, 131]\) or in the ballistic \([132]\) transport regimes) transverse nonequilibrium spin density and polarization \([P_y \neq 0 \text{ in Fig. 5.3(a) and (b)}]\) that is predominantly generated by the Bi sublattice. Note that \(P_z\) component in Fig. 5.3(a) and (b) remains non-zero even if both surfaces are open for transport, or if the cross section in the \(yz\)-plane becomes infinite, since it originates from electrons tunneling through the bulk of the TI.

Figure 5.3(c) plots the linear-response conductance

\[
G_e = \lim_{V_b \to 0} \frac{I_2}{V_b}
\]

of the TI thin film attached to two N leads as a function of the Fermi energy \(E_F\) when small bias voltage \(V_b = V_1 - V_2\) drives charge current \(I_2\). The \(G_e\) vs. \(E_F\) dependence exhibits a V-shape (slightly asymmetric due to the attached N leads) familiar from graphene, \([146, 147]\) with a minimum conductivity

\[
\sigma = GL_x/L_y
\]

reached at the DP located at \(E_F = 3.0\) eV. Even though the density of states vanishes at the DP, so that \(\sigma\) should apparently approach zero at the DP, for this ballistic junction it remains non-zero due to evanescent wavefunctions injected by the metallic N leads. While they are similar to the well-known metal induced gap states in metal-semiconductor junctions, such states typically penetrate only a few atomic lengths into the semiconductor where the depth of penetration decreases with increasing band gap. On the other hand, evanescent states in N/TI junctions penetrate a much longer distance due to zero energy gap at the DP, as observed also in N/graphene junctions. \([148]\) Figure 5.3(d) shows accidental (\(\beta \approx 1\)) Ohmic scaling \(G_e \propto L_y/L_x^\beta\) (for \(L_y/L_x \gg 1\)) at DP, so that evanescent mode quantum transport in N/TI or N/graphene junctions is termed \([147]\) “pseudo-diffusive.”
Figure 5.4: (Color online) (a) The SSE signal $V_{\text{SHE}}/(L_y \Delta T)$ as a function of the position of N lead 3 in Fig. 5.1 displaced between the contacts with N leads 1 and 2 sandwiching TI samples of length $L_x = 100a$ or $L_x = 200a$. (b) Dependence of $|V_{\text{SHE}}|_{\text{max}}/(L_y \Delta T)$ on the Fermi energy $E_F$. (c) Voltage $V_3 L_y/\Delta T$ on the Fermi energy $E_F$. (d) Conventional charge and spin-dependent Seebeck coefficients vs. $E_F$ for the two-terminal version of junction in Fig. 5.1. Adapted from Ref. [15].

5.5 Voltage signal of SSE in three-terminal TI-based junctions

When the third N lead, assumed to be made of a heavy metal with sufficiently large [149] SH angle $\theta_{\text{SH}}$, is attached to the top surface of the TI, as shown in Fig. 5.1, spin current $I^{S\alpha}_3$ will be injected into it. Besides using temperature bias $\Delta T = T_1 - T_2 = 2$ K at average temperature $T = (T_1 + T_2)/2 = 50$ K to drive SSE, we also apply voltage $V_3$ to the macroscopic reservoir (attached to N lead 3 at infinity) in order to ensure that net charge current through it remains zero $I_3 \equiv 0$ and $I^{S\alpha}_3 \neq 0$. 
is pure. The profile of $V_3$ across the TI thin film is plotted in Fig. 5.4(c). We assume that the reservoir temperature

$$T_3(x) = T_1 - x(T_1 - T_2)/L_x$$

decreases linearly as N lead 3 is displaced between the contacts of TI film with N lead 1 and 2.

The central quantity in the theories [124] of transverse SSE is $I_{3x}^S$, component of the pure spin current injected into N lead 3, which we compute per lattice spacing $a$ because of assumed periodicity of system in Fig. 5.1 in the $y$-direction. Since $I_{3x}^S$ carries spins pointing along the $x$-axis, the ISHE mechanism illustrated in Fig. 5.1 will accumulate charges on the opposite edges of N lead 3 in the transverse direction. These generate electric field $E_{ISHE}$ and the corresponding voltage signal [124]

$$V_{ISHE} = E_{ISHE}^y L_y = \theta_{SH}^{Pt} I_{3x}^S e \rho^{Pt} / W.$$ 

To facilitate comparison with experiments, [123, 125] Fig. 5.4 plots $V_{ISHE}/(L_y \Delta T)$ which has the same unit (after multiplying the results in Fig. 4 by $L_y$) as the conventional charge-Seebeck coefficient $S_{charge} = -(V_1 - V_2)/(T_1 - T_2)$ measured [146] on two-terminal junctions. For this purpose, we assume that $I_{3x}^S$ is converted into $V_{ISHE}$ via the ISHE operating within N lead 3 of width $W = 6a$ which is made of Pt with resistivity $\rho^{Pt} = 105 \text{n}\Omega\text{m}$ and with putative [149] SH angle $\theta_{SH}^{Pt} = 0.08$.

The dependence of $V_{ISHE}/(L_y \Delta T)$ on the position of N lead 3 shown in Fig. 5.4(a) is asymmetric, thereby exhibiting a fundamental property of the transverse SSE where its signal changes sign between cold and hot ends of the sample. [122, 123, 125] The maximum signal $|V_{ISHE}|_{max}/(L_y \Delta T)$ is reached around the sample edges and it is independent of length $L_x$, as shown in Fig. 5.4(a). Unlike the spin-dependent Seebeck coefficient $S_{spin}^x$ plotted in Fig. 5.4(d), which quantifies spin current injected into N lead 2 due to temperature bias applied to the two-terminal version of junction in Fig. 5.1, $|V_{ISHE}|_{max}/(L_y \Delta T)$ vs. $E_F$ within the bulk gap of TI plotted in Fig. 5.4(b) is unrelated to conventional CS coefficient $S_{charge}$. Note that both $S_{spin}^x$ and $S_{charge}$ in Fig. 5.4(d)
exhibit the same dependence on $E_F$ as $S_{\text{charge}}$ measured on graphene. [146] This is due to the fact that electron- and hole-like transport gives contributions to these coefficients of opposite sign, so that $S_{\text{spin}}^z = S_{\text{charge}} \equiv 0$ exactly at the DP while reaching maximum absolute value few $k_B T$ away from it.

5.6 Chapter Summary

In this chapter, based on Ref. [15], we predict that thermally driven charge current on the surface of 3D TI thin film, realized using Bi$_2$Se$_3$ whose QLs are oriented at an angle of 45° with respect to the direction of transport while being perpendicular to the TI surface (see Fig. 5.1), will become spin-polarized due to strong surface SOC. In addition to amply studied [129, 130, 131, 132] nonequilibrium transverse spin polarization on the TI surface parallel to QLs, for the orientation of QLs we propose in Fig. 5.1 the spin-polarization vector will acquire an additional component parallel to the direction of charge transport which is generated by the Se sublattice. This makes it possible to recreate the phenomenology of electron-driven SSE recently observed in InSb, [125] but in the absence of any external magnetic field. The predicted magnitude of SSE signal shown in Figs. 5.4(a) and 5.4(b) can be translated into experimentally measurable voltage by multiplying it with $L_y$ and $\Delta T$, e.g., $|V_{\text{ISHE}}|_{\text{max}} \simeq 5 \mu V$ if we assume $L_y = 1 \text{ mm}$, $\Delta T = 2 \text{ K}$ and $a \approx 3.4 \text{ nm}$ (which is the effective lattice constant of our simple cubic lattice for the distance $c \approx 2.9 \text{ nm}$ between QLs oriented as in Fig. 6.1). We believe that this value could be further enhanced by including phonon-electron drag effect, speculated to play a key role in achieving the “giant” magnitude of SSE measured in Ref. [125], which we relegate to future studies.
Chapter 6

NONEQUILIBRIUM SPIN DENSITY AROUND SURFACES OF CURRENT-CARRYING TOPOLOGICAL INSULATOR

The newly discovered three-dimensional topological insulator (3D TIs) materials possess a usual band gap in the bulk while also hosting metallic surfaces. The low-energy quasiparticles on these surfaces behave as massless Dirac fermions whose spins are locked to their momenta due to strong spin-orbit coupling (SOC). [84] Such spin-momentum locking is viewed as a resource which has swiftly ignited pursuit of their spintronic applications. [131] For example, very recent experiments [150] have demonstrated magnetization dynamics of a single ferromagnetic metallic (FM) overlayer deposited on the surface of 3D TI due to current-induced SO torques. Another recent experiment [151] has detected spin-to-charge conversion [152, 153] when precessing magnetization of the FM overlayer pumps pure spin current into the metallic surface of 3D TI.

The microscopic mechanism behind these phenomena can be traced to the so-called Edelstein effect (EE), originally predicted [154] for a diffusive two-dimensional electron gas (2DEG) with the Rashba SOC [155] and observed much later experimentally. [156] In the EE in Rashba-split 2DEGs, longitudinal unpolarized charge current flowing along the $x$-axis drives a nonequilibrium spin density $\mathbf{S}(\mathbf{r}) = (0, S_y, 0)$ pointing in the transverse direction. The two-terminal TI geometry spin density $\mathbf{S}(\mathbf{r})$ is expected to be substantially enhanced [131] (by a factor $\hbar v_F/\alpha_R \gg 1$, where $v_F$ is Fermi velocity in TI and $\alpha_R$ is the strength [155] of the Rashba SOC in 2DEG). This enhancement can be explained by the spin-momentum locking along the single Fermi circle, [84] formed in $k$-space at the intersection of the Dirac cone energy-momentum dispersion.
Figure 6.1: Schematic view of a two-terminal setup where a thin film of Bi$_2$Se$_3$ is attached to two macroscopic reservoirs biased by the electrochemical potential difference $eV_b = \mu_L - \mu_R$. We assume that clean Bi$_2$Se$_3$ film is infinite along the $x$-axis (i.e., the direction of transport) and the $y$-axis, while its thickness along the $z$-axis is chosen as 5 QLs. The shaded cell defines thickness $d_x \approx 5$ Å over which $S(r)$ from Fig. 6.2 is averaged and then plotted in Fig. 6.3, together with LDOS, over the corresponding section within the $yz$-plane defined by the cell. Adapted from Ref. [16].

and the Fermi energy plane, in contrast to spin-momentum locking along the two circles [155] in the case of Rashba 2DEG which counter the effect of each other. A class of recent experiments [157] has also probed $S(r)$ directly in three-terminal geometry where nonmagnetic electrodes inject unpolarized charge current into a TI, while a third FM contact deposited in the middle of the top surface of the TI film detects a voltage signal when a non-zero $S(r)$ is induced. This methodology allows one to quantify the projection of $S(r)$ onto the magnetization of the third FM contact.

In this chapter, we analyze manifestations of EE in the two-terminal TI thin film geometry illustrated in Fig. 6.1. We choose Bi$_2$Se$_3$ as the prototypical TI material—with its single Dirac cone in the surface band structure, relatively large bulk band gap, and Dirac point (DP) inside the gap [see Fig. 1.4(a)]—on which many recent experiments probing EE directly [157] or indirectly [150, 151] have been performed. The central region of the device in Fig. 6.1, which has length $L_x = 21.5$ Å along the $x$-axis and infinite width along the $y$-axis, is attached to two semi-infinite electrodes made of
the same material. The electrodes are assumed to terminate at infinity into macroscopic
Fermi liquid reservoirs where electrons are thermalized to acquire electrochemical po-
tential $\mu_L$ in the left reservoir and $\mu_R$ in the right one. The Hamiltonian of the central
region and the electrodes is obtained from the noncollinear density functional theory
(ncDFT) implemented by us in the ATK package [158] using optimized pseudo-atomic
localized basis functions [159] and SOC introduced via the total-angular-momentum-
dependent pseudopotentials. [160] The transport properties of the system in Fig. 6.1
are computed using the nonequilibrium Green function (NEGF) formalism [91], so that
our approach represents an extension of the widely used NEGF+DFT framework [161]
to quantum transport problems involving noncollinear spins and SOC.

In the simplest picture—based on the Dirac-type effective Hamiltonian in Eq. (5.13)
($\hat{\sigma}$ is the vector of the Pauli matrices; $\hat{p}$ is the momentum operator; and $e_z$ is the unit
vector along the $z$-axis in Fig. 6.1) describing electrons on the metallic surfaces of TIs—
the spin and momentum of electronic eigenstates are orthogonal to each other along
the single Fermi circle. This generates net homogeneous $S = (0, S_y, 0)$ after an applied
electric field $E_x$ shifts the Fermi circle [131, 150, 130] along the momentum parallel to
$E_x$. Such manifestation of EE persists in ballistic samples as well [132, 162] where there
is no electric field within the TI but instead one applies the electrochemical potential
difference $eV_b = \mu_L - \mu_R$ between macroscopic reservoirs to inject a current into the
TI, as illustrated in Fig 6.1. The relations $S_y \propto E_x$ or $S_y \propto V_b$, describing EE in the
diffusive or ballistic transport regimes, respectively, are allowed only in nonequilibrium
since in equilibrium $S$ changes sign under time reversal, and, therefore, has to vanish
(assuming absence of magnetic field).

On the other hand, DFT calculations [163, 164] reveal that equilibrium expecta-
tion value of spin in the eigenstates of Bi$_2$Se$_3$ surfaces tilts out of the 2D Brillouin zone
(BZ) due to hexagonal warping of the Dirac cone surface band structure. This was also
confirmed by spin- and angle-resolved photoemission spectroscopy (SARPES). [165]
Thus, in nonequilibrium one can expect to observe $S = (0, S_y, S_z)$ on the surface of
the TI, which calls to quantify [165] the ratios $|S_y|/S$ and $|S_z|/S$ and gain insight into
how the Bi and Se sublattices contribute [162] to $S$. Since SARPES experiments [165] show that stronger Fermi circle warping leads to a larger out-of-plane spin component, the texture of $S(r)$ should also depend on the direction of injected current relative to the orientation of the lattice of Bi and Se atoms.

6.1 Methodology

We now explain the technical details of our calculations. The extension of DFT to the case of spin-polarized systems is formally derived in terms of total electron density $n(r)$ and vector magnetization density $m(r)$. In the collinear DFT, $m(r)$ points in the same direction at all points in space, which is insufficient to study magnetic systems where the direction of the local magnetization is not constrained to a particular axis or systems governed by SOC. In ncDFT, [167] the exchange-correlation (XC) functional $E_{\text{XC}}[n, m(r)]$ depends on $m(r)$ pointing in arbitrary directions. The local density approximation (LDA) and most often employed version of generalized gradient approximation (GGA), implemented also by us in ATK, [158] make additional approximations [167] that lead to the XC magnetic field

$$B_{\text{XC}}(r) = \frac{\delta E_{\text{XC}}[n, m(r)]}{\delta m(r)},$$

being parallel everywhere to $m(r)$. The single-particle spin-dependent Kohn-Sham (KS) Hamiltonian in ncDFT takes the form

$$\hat{H}_{\text{KS}} = -\frac{\hbar^2 \nabla^2}{2m} + V_H(r) + V_{\text{XC}}(r) + V_{\text{ext}}(r) - \sigma \cdot B_{\text{XC}}(r),$$

(6.1)

where $V_H(r)$, $V_{\text{xc}}(r)$ and $V_{\text{ext}}(r)$ are the Hartree, XC and external potential, respectively. Diagonalization of $\hat{H}_{\text{KS}}$ proceeds by approximating the Hilbert space of all single-electron eigenfunctions with a finite set of basis functions. A popular basis set is plane-waves (PWs), where varying only one parameter (the energy cutoff) allows one to improve the basis systematically. Linear combination of atomic orbitals (LCAO) basis sets require more tuning, however, they simplify the NEGF calculations where one has to spatially separate system into the central region and semi-infinite electrodes, as illustrated in Fig. 6.1.
Since the pioneering predictions [168] of candidate materials realizing strong TIs via ncDFT calculations, the electronic band structure of TIs has almost exclusively been calculated using PWs. [163, 166] In Fig. 1.4(a) we demonstrate that ncDFT with LCAO basis implemented in ATK [158] can accurately reproduce the band structure of ncDFT calculations with PWs implemented in the VASP package. [169] The supercell considered in both calculations is shown in Fig. 6.2(a), which includes 5 QLs terminated by Se atomic layer on both the top and bottom surface, as well as 7.5 Å thick vacuum layer above and below these Se atomic layers.

In VASP calculations, [170] the electron-core interactions are described by the projector augmented wave method, [171] and we use Perdew-Burke-Ernzerhof (PBE) parametrization of GGA for the XC functional. The cutoff energy for the PW basis set is 350 eV. In both the ATK and the VASP calculations of Fig. 1.4(b) we employ $11 \times 11 \times 1$ $k$-point mesh within Monkhorst-Pack scheme for the Brillouin zone integration.

In the ATK calculations, the electron-core interactions are described by norm-conserving pseudopotentials. The pseudopotentials are obtained by mapping the solution of the Dirac equation, which naturally includes SOC, [160] to non-relativistic pseudopotential,

$$V_{PS} = V_L + V_{NL}^{1/2} + V_{NL}^{-1/2}$$

, with local contribution $V_L$ and non-local contributions $V_{NL}$ from the total angular momentum $j = l + 1/2$ and $j = l - 1/2$. The non-local terms are expanded in terms of SO projector functions,

$$V_{NL}^{\pm1/2} = \sum_{l,\xi,\alpha,\beta} \nu_{l,\xi,\alpha,\beta} P_{\alpha\beta}^{l+1/2,\xi}$$

, where $\nu_{l,\xi,\alpha,\beta}$ are normalization constants and the indices $\alpha, \beta$ denote the possible spin orientations ($\uparrow, \downarrow$). We use PBE GGA for the XC functional and a LCAO basis set $\{\phi_i\}$ generated by the OpenMX package, [172, 159] which consists of $s^2p^2d^1$ orbitals on Se atoms and $s^2p^2d^2$ on Bi atoms. These pseudoatomic orbitals were generated by a confinement scheme [159] with the cutoff radius 7.0 a.u. and 8.0 a.u. for Se and Bi.
atoms, respectively. The energy mesh cutoff for the real-space grid is chosen as 75.0 Hartree.

The eigenstates $|\Psi_n\rangle$ of the KS Hamiltonian in Eq. (6.1) make it possible to construct the equilibrium density matrix

$$\rho_{eq} = \sum_n |\Psi_n\rangle \langle \Psi_n| f(E)$$

for electrons at $\mu_L = \mu_R$ and temperature $T$ described by the Fermi distribution function $f(E)$. The local electron and magnetization density, as the central variables of ncDFT, are obtained from

$$n(r) = \langle r | \text{Tr}_{\text{spin}} [\rho_{eq}] | r \rangle$$
and

$$m(r) = \langle r | \text{Tr}_{\text{spin}} [\rho_{eq}\sigma] | r \rangle,$$

where the trace is taken over the spin Hilbert space. In steady-state nonequilibrium, with a dc current flowing between the left and right reservoirs in Fig. 6.1, we construct the nonequilibrium density matrix \cite{173} $\rho_{neq}$ in terms of NEGFs

$$\rho_{neq} = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} dE \, G_{<}(E) - \rho_{eq}. \quad (6.2)$$

This allows us to obtain $S(r)$ plotted in Figs. 6.2 and 6.3 from

$$S(r) = \frac{\hbar}{2} \text{Tr}_{\text{spin}} [\rho_{neq}\sigma]. \quad (6.3)$$

The NEGF formalism for steady-state transport operates with two central quantities— the retarded GF, $G(E)$, and the lesser GF, $G_{<}(E)$—which describe the density of available quantum states and how electrons occupy those states, respectively. In the absence of inelastic processes, these are given by $G = [E \mathbf{O} - \mathbf{H}_{KS} - \Sigma_L - \Sigma_R]$ and Eq. (2.21). Here the self-energies $\Sigma_{L,R}$ are due to semi-infinite electrodes and $\Gamma = i(\Sigma_{L,R} - \Sigma_{L,R}^\dagger)$. For the chosen LCAO basis set, the Hamiltonian matrix $\mathbf{H}_{KS}$ is composed of elements $\langle \phi_i | \hat{H}_{KS} | \phi_j \rangle$ and the overlap matrix $\mathbf{O}$ is composed of elements $\langle \phi_i | \phi_j \rangle$. 

86
In the linear-response transport regime considered here, Eq. (6.2) can be expanded \cite{173} to linear order in bias voltage $V_b$. Since $S(r)$ is zero in equilibrium (because of assumed absence of external magnetic field), the linear-response density matrix can be plugged into Eq. (6.3) in its simplified form, \cite{173}

$$
\rho_{\text{neq}} = \frac{eV_b}{2\pi} \int_{-\infty}^{+\infty} dE \, G\Gamma L G^\dagger \left(-\frac{\partial f}{\partial E}\right).
$$

Otherwise, the gauge-invariant form of $\rho_{\text{neq}}$ requires additional terms \cite{173} to properly remove the equilibrium expectation value of a considered physical quantity.

The retarded GF also allows us to obtain the transmission function of the device in Fig. 6.1, $T(E, k_y) = \text{Tr}[\Gamma_R G\Gamma L G^\dagger]$, which depends on energy and transverse momentum $k_y$ due to assumed periodicity in the $y$-direction. The total transmission function $T(E)$ is obtained by integrating over $k_y$, which determines the linear-response conductance via the Landauer formula, $G = \frac{e^2}{h} \int dE \, T(E) \left(-\frac{\partial f}{\partial E}\right)$. We confirm in Fig. 1.4(b) that $T(E, k_y = 0) = 2$ for $E$ within the bulk gap shown in Fig. 1.4(a) because only one doubly degenerate helical conducting channel is open for transport in that energy range \cite{174, 175} for injected electrons with momentum along the $\Gamma$–M direction ($k_y = 0$).

### 6.2 Spin Texture on the Surfaces of TI

Our principal results are shown in Figs. 6.2 and 6.3. When a small (ensuring linear-response transport regime) bias voltage $V_b$ is applied between the reservoirs in Fig. 6.1, the unpolarized charge current injected into Bi$_2$Se$_3$ thin film generates a nonequilibrium $S(r)$ whose complex inhomogeneous texture within three planes selected in Fig. 6.2(a) is plotted in Figs. 6.2(b) and 6.2(c). For the visualization we use real-space grid for $r$ whose spacing is $\simeq 0.4 \, \text{Å}$. On such a fine grid, all three components of $S(r)$ are non-zero. However, upon averaging $S(r)$ over a $d_x \simeq 5 \, \text{Å}$ thick cell denoted in Fig. 6.1, we obtain spatial profiles in Figs. 6.3(c) and 6.3(d) which show that $S_y$ is the largest component independently of the direction of incoming electrons. An order
Figure 6.2: (a) The arrangement of Bi and Se atoms within a supercell of Bi$_2$Se$_3$ thin film of thickness 5 QLs. The inset in panel (a) shows BZ in the $k_x$-$k_y$ plane at $k_z = 0$ with special $k$-points $\Gamma$, $M$, and $K$ marked. (b) The vector field of nonequilibrium $\mathbf{S}(\mathbf{r})$ within selected planes shown in (a), generated by injection of unpolarized charge current along the $x$-axis (see also Fig. 6.1). The planes 1 and 3 correspond to the top and bottom metallic surfaces of Bi$_2$Se$_3$ thin film, while plane 2 resides in the bulk at a distance $d \approx 0.164$ nm away from plane 1. (c) The vector fields in (b) projected onto each of the selected planes in (a). The real space grid of $\mathbf{r}$-points in panels (b) and (c) has spacing $\approx 0.4$ Å. Adapted from Ref. [16].
of magnitude smaller $S_z$ component shown in Fig. 6.3(d) appears for electrons incoming along current direction 2 marked in panel (e), which is in accord with SARPES experiments [165] finding largest warping and out-of-plane spin component along the corresponding direction in the BZ.

In Figs. 6.2 and 6.3 we also predict that a current-driven non-zero $\mathbf{S}(\mathbf{r})$ will appear not only on the TI surface, but also within $\simeq 2$ nm thick region of its bulk in the vicinity of the top or bottom surfaces. This feature is explained in Figs. 6.3(a) and 6.3(b) showing spatial profile of the local density of states (LDOS) at the Fermi energy $E_F$ over the cell depicted in Fig. 6.1. The non-zero LDOS and the corresponding $\mathbf{S}(\mathbf{r})$ in the bulk of the TI thin film stem from evanescent wavefunctions which originate from the top and bottom metallic surfaces and penetrate into the energy gap of the insulating bulk. The Bi$_2$Se$_3$ is a strongly anisotropic material composed of quintuple layers (QLs) of Bi and Se atoms, illustrated in Fig. 6.2(a), where one QL consists of three Se layers strongly bonded to two Bi layers in between. For Bi$_2$Se$_3$ film thinner than 5 QLs, the evanescent wavefunctions from the top and bottom metallic surface can overlap to create a minigap [163, 166] at the DP. We select the thickness of Bi$_2$Se$_3$ to be 5 QLs along the $z$-axis in Fig. 6.1, which ensures that the LDOS in Figs. 6.3 (a) and 6.3 (b) goes to zero on the plane half way between the top and bottom surfaces of the TI thin film.

6.3 Chapter Summary

In this chapter, based on Ref. [16], we obtained a nonequilibrium spin texture $\mathbf{S}(\mathbf{r})$ within a thin film of current-carrying Bi$_2$Se$_3$ TI material using a NEGF+ncDFT framework implemented recently by us in ATK package [158]. The non-zero texture appears on the TI metallic top and bottom surfaces, as well as within bulk layers of thickness $\simeq 2.0$ nm in the vicinity of the surfaces that effectively dope the bulk by evanescent wavefunctions. The spin texture is quite complex on length scales $\lesssim 1$ Å. However, upon averaging it over a few Å we find a simpler pattern—with either $\mathbf{S} \equiv (0, S_y, 0)$, or $\mathbf{S} \equiv (0, S'_y, S'_z)$ where $S'_y/S'_z \gg 1$—depending on the direction of injected current with
Figure 6.3: (a),(b) The spatial profile of LDOS at $E_F$ over the cross section of the cell denoted in Fig. 6.1. (c),(d) The spatial profile of the components of $S(\mathbf{r})$, obtained by averaging its texture plotted in Fig. 6.2, over the cell of Bi$_2$Se$_3$ thin film marked in Fig. 6.1. The direction of injected charge current for the results in panels (a) and (c), or the results in panels (b) and (d), is denoted in panel (e) relative to the orientation of the lattice of Bi and Se atoms. The bottom surface of Bi$_2$Se$_3$ is located at $z = 0$ nm, and the top TI surface is located at $z \approx 4.56$ nm. Adapted from Ref. [16].
respect to orientation of the lattice of Bi and Se atoms. Such dependency offers a novel probe, via transport electronic measurements, [157] of the hexagonal warping of the Dirac cone surface band structure. For the envisaged spintronic applications of TIs, it is essential to understand how $S(r)$ changes due to finite bias voltage or self-consistent coupling [176] to magnetization of a ferromagnetic (metal or insulator) overlayer, which we relegate to future studies.
BIBLIOGRAPHY


[77] https://wiki.fysik.dtu.dk/gpaw/.


generation in mesoscopic conductors with non-uniform magnetic fields,” New J.

[137] B. K. Nikolić, L. P. Zárbo, and S. Souma, “Mesoscopic spin Hall effect in multi-
075361.


[140] U. Sivan and Y. Imry, “Multichannel Landauer formula for thermoelectric trans-
port with application to thermopower near the mobility edge,” Phys. Rev. B 33
(1986) 551.

quantum transport modeling of thermoelectricity in single-molecule nanojunctions

186801 (2010); O. V. Yazyev, J. E. Moore, and S. G. Louie, Phys. Rev. Lett. 105


[145] H. K"ohler and E. W"ochner, “The g-factor of the conduction electrons in Bi2Se3,”


[148] R. Golizadeh-Mojirad and S. Datta, “Effect of contact induced states on mini-


Appendix A

DERIVATION OF BOND CURRENT

A.1 Derivation of Bond Current in second quantization form

\[ H = \sum_{m'n'} c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma}, \]

\[ \frac{dN_m}{dt} = \frac{1}{i\hbar} \{ N_m, H \}, \]

where \( N_m = c_{m's'}^\dagger c_{m\sigma} \)

\[ J_m = \frac{1}{i\hbar} \left[ N_m \left( \sum_{m'n'} c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma} \right) - \left( \sum_{m'n'} c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma} - \sum_{m'n'} c_{m's'}^\dagger N_{m'n'} c_{m\sigma} - c_{m's'}^\dagger N_{m'n'} c_{m\sigma} \right) \right], \]

where \( J_{mm'} = e \frac{1}{i\hbar} \left[ \sum_{\sigma\sigma'} \sum_{m'n'} \left( c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} - c_{m's'}^\dagger N_{m'n'} c_{m\sigma} - c_{m's'}^\dagger N_{m'n'} c_{m\sigma} \right) - \left( c_{m's'}^\dagger c_{m\sigma} \right) N_{m'n'} t_{mm'}^{\sigma\sigma'} \right], \]

and

\[ J_m = \sum_{\sigma\sigma'} \sum_{m'} c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma} - c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma}, \]

where

\[ J_{mm'}^{\sigma\sigma'} = \frac{e}{i\hbar} \left[ c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma} - c_{m's'}^\dagger t_{mm'}^{\sigma\sigma'} c_{m\sigma} \right]. \]
A.1.1 Todorov’s approach for bond currents

Start with wave function in orthogonal basis set representation, we then have wavefunction

\[ |\psi\rangle = \sum_{m,\gamma} \psi_{m\gamma} |\phi_{m\gamma}\rangle \]

and Hamiltonian,

\[ H = \sum_{m'\gamma m'\gamma'} |\phi_{m\gamma}\rangle H_{m\gamma m'\gamma'} \langle \phi_{m'\gamma'}| \]

In time-dependent case

\[ |\psi(t)\rangle = \sum_{m,\gamma} \psi_{m\gamma}(t) |\phi_{m,\gamma}\rangle, \]

\[ H |\psi(t)\rangle = i\hbar \dot{|\psi(t)}\rangle, \]

\[ \sum_{m',\gamma'} H_{m\gamma m'\gamma'} \psi_{m\gamma} = E \psi_{m\gamma}. \]

\( m \) is site index, \( \gamma \) is orbitals in each sites.

The projection operator, by definition, reads

\[ P_m = \sum_{\gamma} |\phi_{m\gamma}\rangle \langle \phi_{m\gamma}| \]

\[ \sum_{m} P_m = 1 \]

\( P_m \) is the electron probability at site \( m \). In Heisenberg picture

\[ P_m(t) = \langle \psi(t)| P_m |\psi(t)\rangle \]

and according to continuity equation

\[ J_m(t) = \langle \psi(t)| J_m |\psi(t)\rangle = \dot{P}(t) = \frac{1}{i\hbar} \langle \psi(t)| [P_m, H] |\psi(t)\rangle, \]
The current operator has expression
\[ J_m = \frac{1}{i\hbar} [P_m H - H P_m]. \]
Inserting projection operator again
\[ J_m = \frac{1}{i\hbar} \sum_{m' \neq m} [P_m H P_{m'} - P_{m'} H P_m], \]
and its basis representation
\[ J_m = \sum_{m', \gamma'} |\phi_{m\gamma}\rangle \langle \phi_{m\gamma}| H |\phi_{m'\gamma'}\rangle \langle \phi_{m'\gamma'}| - |\phi_{m'\gamma'}\rangle \langle \phi_{m'\gamma'}| H |\phi_{m\gamma}\rangle \langle \phi_{m\gamma}|, \]
which describe the sum of all the contributions of currents from neighboring sites. From observation, it is natural to define current flow that connects each pair of sites.
\[ J_m = \sum_{m'} J_{mm'} \]
\[ J_{m'm} = \sum_{\gamma'} |\phi_{m\gamma}\rangle \langle \phi_{m\gamma}| H |\phi_{m'\gamma'}\rangle \langle \phi_{m'\gamma'}| - |\phi_{m'\gamma'}\rangle \langle \phi_{m'\gamma'}| H |\phi_{m\gamma}\rangle \langle \phi_{m\gamma}|. \]
\[ J_{m'm'} = J_{m'm} = \sum_{\gamma'} |\phi_{m\gamma}\rangle H_{m\gamma m'\gamma'} \langle \phi_{m'\gamma'}| - |\phi_{m'\gamma'}\rangle H_{m'\gamma' m\gamma} \langle \phi_{m\gamma}|. \]
In second quantization form
\[ J_{m'm}^{\gamma\gamma'} = \sum_{\gamma'} [c_{m\gamma}^\dagger c_{m'\gamma'} H_{m\gamma m'\gamma'} - H.c] = \sum_{\gamma'} [c_{m\gamma}^\dagger c_{m'\gamma'} H_{m\gamma m'\gamma'} - H.c]. \]
A.2 Symmetrized form of $G^<$

Starting from the definition

$$G^<(E) = G^r[i f_L \Gamma_L + i f_R \Gamma_R]G^a.$$  

Adding and subtracting $G^r[i f_L \Gamma_R]G^a$, becomes

$$G^r[i f_L \Gamma_L + i f_L \Gamma_R]G^a + G^r[i f_R \Gamma_R - i f_L \Gamma_R]G^a = G^r[i f_L (\Gamma_L + \Gamma_R)]G^a + G^r \Gamma_R G^a (f_R - f_L)$$  

(A.4)

and doing the opposite by adding and subtracting $G^r[i f_R \Gamma_R]G^a$, becomes

$$G^r[i f_L \Gamma_L - i f_R \Gamma_L]G^a + G^r[i f_R \Gamma_R + i f_R \Gamma_L]G^a = G^r[i f_R (\Gamma_L + \Gamma_R)]G^a + G^r \Gamma_L G^a (f_L - f_R).$$  

(A.5)

Average the two equations Eq. (A.4) and Eq. (A.5)

$$G^< = \text{Im}[G^r](f_L + f_R) + i G^r(\frac{\Gamma_L - \Gamma_R}{2})G^a (f_L - f_R).$$
A.3 Prove in non-equilibrium situation, first and second term in the RHS of Eq. (2.40)

Start from the first term

\[
[\Sigma^r G^< + \Sigma^< G^a] + H.c.
\]

Using the Eq. (2.21) to replace \( G^< \)

\[
[\Sigma^r (G^r \Sigma^< G^a) + ([G^r]^{-1} G^r) \Sigma^< G^a] + H.c.
\]

factor out \( G^< \) and replace \( G^r \) with Eq. (2.14)

\[
[\Sigma^r + (E - H - \Sigma^r)] G^< + H.c.
\]

We then obtain

\[
[E - H] G^< - H.c
\]

and

\[
[H G^< - G^< H].
\]
Appendix B

LIST OF THESIS RELATED PUBLICATIONS


