Magnetization of the metallic surface states in topological insulators

C J Tabert$^{1,2}$ and J P Carbotte$^{3,4}$

$^1$ Department of Physics, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada
$^2$ Guelph-Waterloo Physics Institute, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada
$^3$ Department of Physics, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4M1, Canada
$^4$ Canadian Institute for Advanced Research, 180 Dundas St W, Toronto, Ontario M5G 1Z8, Canada

E-mail: ctabert@uoguelph.ca

Received 16 September 2014, revised 30 October 2014
Accepted for publication 10 November 2014
Published 27 November 2014

Abstract
We calculate the magnetization of the helical metallic surface states of a topological insulator. We account for the presence of a small sub-dominant Schrödinger piece in the Hamiltonian in addition to the dominant Dirac contribution. This breaks particle-hole symmetry. The cross-section of the upper Dirac cone narrows while that of the lower cone broadens. The sawtooth pattern seen in the magnetization of the pure Dirac limit as a function of chemical potential ($\mu$) is shifted; but, the quantization of the Hall plateaus remains half integral. This is verified by taking the derivative of the magnetization with respect to $\mu$. We compare our results with those when the non-relativistic piece dominates over the relativistic contribution and the quantization is integral. Analytic results for the magnetic oscillations are obtained where we include a first order correction in the ratio of non-relativistic to relativistic magnetic energy scales. Our fully quantum mechanical derivations confirm the expectation of semiclassical theory except for a small correction to the expected phase. There is a change in the overall amplitude of the magnetic oscillations. The Dingle and temperature factors are modified.

Keywords: magnetization, topological insulators, Hall conductivity

(Some figures may appear in colour only in the online journal)

1. Introduction
A topological insulator is a bulk insulator with a metallic spectrum of topologically protected helical surface states [1–3]. The helical fermions which exist at the surface [4–7] display an odd number of Dirac points. As examples, Bi$_2$Se$_3$ [5] has a single point while samarium hexaboride has three [8]. In such systems, the in-plane spin of the electron is perpendicular to its momentum. Recently, angle-resolved photoemission spectroscopy (ARPES) experiments have mapped out the surface bands [4, 5] and confirmed the predicted Dirac-like spectrum and spin arrangement. In contrast to graphene, whose low-energy Dirac cones are particle-hole symmetric, the surface states of a topological insulator exhibit band bending and take an hourglass shape with the valence band below the Dirac point displaying significantly more outward bending than the corresponding inward-bending of the conduction band [4, 5, 7, 9–12]. This behaviour can be captured by adding a Schrödinger mass term to the ideal linear Dirac Hamiltonian which remains dominant in topological insulators. The resulting particle-hole asymmetry has important ramifications on the physics of such systems. As examples, Wright and Mackenzie [10] have discussed its effect on the Berry phase while Wright [13] describes its role on measurements of the Chern number and the phase transition from the spin Hall to quantum anomalous Hall phase. Shubnikov–de-Haas (SdH) oscillations [14] are also expected to be altered [15–19]. Li et al. [20] have considered the Hall conductivity and optical absorption [12]. Particle-hole asymmetry splits the interband magneto-optical absorption lines of the pure Dirac case [21, 22] into two. This is due to the broken degeneracy of the energy difference between
the valence band Landau levels $-m$ and $-(m+1)$ to conduction levels $m+1$ and $m$, respectively. Schafgans et al. [23] have given results of magneto-optical measurements in the topological insulator Bi$_2$Se$_3$.

Here we consider the magnetization ($M$) and, in particular, describe how the sub-dominant Schrödinger term changes its dependence on chemical potential ($\mu$). The derivative of $M$ with respect to $\mu$ is of particular interest as it is related to the underlying quantization of the Hall plateaus. We compare our results with those in the opposite limit when the Schrödinger term dominates. This yields a nearly parabolic electronic dispersion which is slightly modified by a small spin–orbit interaction [24, 25]. This limit is relevant to the field of spintronic semiconductors which has been extensively studied in the past [24–33]. In addition to the Hall plateaus, the oscillations in the magnetization are also affected by the Schrödinger term.

Our manuscript is organized as follows: in section 2, we introduce the appropriate low-energy Hamiltonian and Landau level spectrum resulting from a finite magnetic field. Section 3 contains a discussion of the grand thermodynamic potential on which our calculations are based. Both relativistic and non-relativistic limits are considered. Numerical results are presented for the evolution of the magnetization as a function of chemical potential for several ratios of the relevant energy scales. The results are compared to the pure Dirac limit and the differences are emphasized. We calculate the derivative $\partial M(\mu)/\partial \mu_{\mid_B}$, which is related to the quantization of the Hall conductivity through the Streda formula [34]. For comparison, similar results are obtained in the spintronic regime where the Schrödinger scale dominates and the Dirac term is a small perturbation. In section 4, we give a fully quantum mechanical derivation of the effect a sub-dominant Schrödinger term has on the quantum oscillations. Finite temperature effects are discussed as is the effect of impurity scattering in the constant scattering rate approximation. Our conclusions follow in section 5.

2. Low-energy hamiltonian

In the absence of a magnetic field, the low-energy helical surface fermions of a topological insulator are well described by the Bychkov–Rashba Hamiltonian [24, 25]

$$
H = \frac{\hbar^2 k^2}{2m} + \hbar v_F (k_x \sigma_y - k_y \sigma_x),
$$

(1)

where $\sigma_x$ and $\sigma_y$ are the usual Pauli matrices associated with spin and $k$ is the momentum measured relative to the $\Gamma$ point of the surface Brillouin zone. The first term is the familiar parabolic Schrödinger piece for describing an electron with effective mass $m$. The second term describes massless Dirac fermions which move with a Fermi velocity $v_F$. Equation (1) can be solved to give the energy dispersion

$$
\epsilon_{\pm}(k) = \frac{\hbar^2 k^2}{2m} \pm \hbar v_F k.
$$

(2)

A schematic of the energy dispersion is shown in figure 1. Figure 1(a) shows the surface-state band structure in the topological insulator regime (i.e. the Dirac term dominates). The blue cones correspond to the pure Dirac limit ($m \rightarrow \infty$) while the red dispersion results from a non-infinite mass. In figure 1(b), the spintronic limit is shown (i.e. the Schrödinger term dominates). The blue parabola corresponds to the pure Schrödinger case ($v_F \rightarrow 0$) while the two offset red parabolas result from a finite $v_F$. It should be clear from this figure that, while the same Hamiltonian describes both sets of dispersion curves, in the confines of a Brillouin zone, these are very different and, as will be seen, lead to different physics. In fact, the low-energy Hamiltonian does not allow these two regimes to be connected continuously; but, it does describe each separately given appropriate values of the two characterizing parameters $v_F$ and $m$. For the specific case of the topological insulator Bi$_2$Te$_3$, detailed band structure calculations [35, 36] give: $v_F = 4.3 \times 10^5$ m s$^{-1}$ and $m = 0.09m_e$, where $m_e$ is the bare mass of an electron [20]. For the typical spintronic materials, $v_F$ is reduced to $O(10^4)$ m s$^{-1}$ at most [32, 33] and $m$ remains at approximately 0.09$m_e$. The continuum Hamiltonian of equation (1) allows the valence band to bend back over and eventually cross the zero energy axis. This is not physical in the context of topological insulators. Therefore, when using this model, it is important to apply an appropriate momentum cutoff to prevent this spurious behaviour.

To discuss the theory of magnetic oscillations, we need to consider the effect of an external magnetic field $B$ which we
orient perpendicular to the surface of the insulator ($\hat{z}$). To do this, we work in the Landau gauge so that the magnetic vector potential, given by $B = \nabla \times \mathbf{A}$, is written as $\mathbf{A} = (0, Bx, 0)$. We then make the Peierls substitution on the momentum $\hat{p}_i \rightarrow \hat{p}_i - q \hat{A}_i$, where $q = -e$ is the charge of the electron. Thus, equation (1) becomes

$$H = \frac{\hbar^2}{2m} \left[ (-i \partial_y)^2 + (-i \partial_x + eBx/\hbar)^2 \right] + \hbar v_F \left[ (-i \partial_y)\sigma_y - (-i \partial_x + eBx/\hbar)\sigma_x \right]. \quad (3)$$

Here, we are not including a Zeeman splitting term. Its possible effects on the magnetization have already been considered in the work of Wang et al. [34] who found it to have a negligible effect on all the Landau level energies except for that at $N = 0$. Therefore, it will not appreciably affect the phenomena of interest in this paper. Equation (3) can be solved to give the Landau level dispersion

$$E_{N,s} = \frac{\hbar^2 N}{2ml_B^2} + s \sqrt{\frac{\hbar^2}{2ml_B^2} + \frac{2N\hbar^2v_F^2}{l_B^2}}, \quad (4)$$

where $l_B = \sqrt{\hbar/|eB|}$ is the magnetic coherence length, $s = \pm$ for the conduction and valence bands, respectively, and $N > 0$ is an integer and gives the Landau level index. The $N = 0$ Landau level must be treated carefully and is given by

$$E_{N=0} = \frac{\hbar^2}{2ml_B^2}. \quad (5)$$

For convenience, we define both Schrödinger and Dirac energy scales which are given by $E_0 = \hbar e|B|/m$ and $E_1 = \hbar v_F \sqrt{|eB|/\hbar}$, respectively. It is also useful to introduce a dimensionless parameter $P = E_1^2 / E_0^2$. The limit $P \rightarrow \infty$ corresponds to the pure Dirac system while $P \rightarrow 0$ in the pure Schrödinger case.

Using these definitions, the Landau level spectrum for $N \neq 0$ can be expressed as

$$E_{N,s} = E_0 N + s \sqrt{\frac{E_0}{2} + 2NE_1^2}, \quad (6)$$

or, equivalently,

$$E_{N,s} = E_0 N + sE_0 \sqrt{\frac{1}{4} + 2NP}. \quad (7)$$

For $N = 0$,

$$E_{N=0} = \frac{E_0}{2}. \quad (8)$$

One must be careful when dealing with the $s = -$ levels for $E_1 > E_0$. As previously mentioned, we are working with a continuum model in which the valence band can bend back toward the zero energy axis. As a result, that unphysical portion of the band structure can also condense into Landau levels. Indeed, one finds that for large $N$, the $E_{N,-}$ levels begin to increase in energy. While some bending may be characteristic of a topological insulator, one must not allow the valence band to cross the energy axis. This is done by applying a momentum cutoff when $B = 0$. For a finite $B$, we must apply an appropriate cutoff on $N$ to ensure none of the $s = -$ levels become positive.

3. Grand thermodynamic potential

3.1. Dirac limit

Our discussion of the magnetic response of the surface charge carriers begins with the grand thermodynamic potential $\Omega(T, \mu)$. For the relativistic Dirac case with particle-hole symmetry, Sharapov et al. [37] start from

$$\Omega(T, \mu) = -T \int_{-\infty}^{\infty} N(\omega) \ln \left( 2 \cosh \frac{\omega - \mu}{2T} \right) d\omega, \quad (9)$$

where $T$ is the temperature, $\mu$ is the chemical potential, and $N(\omega)$ is the density of states. In the absence of impurities, $N(\omega)$ is a series of Dirac delta functions located at the various Landau level energies. Equation (9) can be rewritten as

$$\Omega(T, \mu) = -T \int_{-\infty}^{\infty} N(\omega) \ln \left( 1 + e^{(\omega - \mu)/T} \right) d\omega + \frac{1}{2} \int_{-\infty}^{\infty} \mu N(\omega) d\omega - \frac{1}{2} \int_{-\infty}^{\infty} \omega N(\omega) d\omega. \quad (10)$$

The first term on the right hand side has the form of the usual non-relativistic grand potential [$\Omega_{NR}(T, \mu)$]; except, now there are negative energy states. The second term is half the chemical potential times the total number of states in our bands. Therefore, it does not contribute to the magnetization [$M(T, \mu)$] which is given by the first derivative of $\Omega(T, \mu)$ with respect to $B$ at fixed chemical potential, i.e. $M(T, \mu) = -[\partial \Omega(T, \mu)/\partial B]_B$. The final term is zero in graphene because of particle-hole symmetry. Thus, the magnetization calculated from the relativistic grand potential reduces correctly to that of $\Omega_{NR}$. While we could proceed directly from $\Omega_{NR}$, it is convenient to keep the second term of equation (10). At zero temperature, the first two terms of equation (10) reduce to

$$\Omega(T = 0, \mu) = \frac{1}{2} \int_{-\infty}^{0} (2\omega - \mu) N(\omega) d\omega + \int_{0}^{\mu} (\omega - \mu) N(\omega) d\omega + \frac{1}{2} \int_{-\infty}^{\infty} \mu N(\omega) d\omega, \quad (11)$$

which can be used to derive the results for graphene as well as a topological insulator. The density of states for our topological insulator has the form

$$N(\omega) = \frac{eB}{\hbar} \left[ \delta(\omega - E_0/2) + \sum_{N=1,\pm}^{\infty} \delta(\omega - E_{N,s}) \right], \quad (12)$$

where $s = \pm$ gives the conduction and valence band, respectively. For graphene, $E_0 = 0$ and hence the first term in equation (12) is a Dirac delta function at $\omega = 0$ which must be duly noted. For graphene, $N(\omega) d\omega = \int_{-\infty}^{0} N(\omega) d\omega$ while for a topological insulator with all $E_{N,-} < 0$, $\int_{-\infty}^{0} N(\omega) d\omega = \int_{-\infty}^{0} N(\omega) d\omega$.
\[
\int_{\infty}^{0} N(\omega) d\omega = \int_{-\infty}^{0} N(\omega) d\omega + eB/\hbar \quad \text{and thus}
\]

\[
\Omega(0, \mu) = \int_{0}^{\mu} (\omega - \mu) N(\omega) d\omega + \int_{-\infty}^{0} \omega N(\omega) d\omega + \frac{eB\mu}{2\hbar},
\]

for a topological insulator, and

\[
\Omega(0, \mu) = \int_{0}^{\mu} (\omega - \mu) N(\omega) d\omega + \int_{-\infty}^{0} \omega N(\omega) d\omega,
\]

for graphene. In both cases, the vacuum contribution \(\Omega_0(0)\) is \(\int_{-\infty}^{0} \omega N(\omega) d\omega\) which can depend on \(B\) but not on \(\mu\). For graphene, this has been worked out in detail by Sharapov et al. [37] and found to go like \(B^{3/2}\) (see their equation (A5)).

When interested in the changes in magnetization for fixed \(B\) as a function of \(\mu\), this term can be dropped as it simply adds a constant background. We also note that

\[
\left( \frac{\partial \Omega(0, \mu)}{\partial \mu} \right)_B = -\int_{0}^{\mu} N(\omega) d\omega + \frac{eB}{2\hbar},
\]

for a topological insulator,

\[
\left( \frac{\partial \Omega(0, \mu)}{\partial \mu} \right)_B = -\int_{0}^{\mu} N(\omega) d\omega,
\]

for graphene, and that the vacuum does not appear in this quantity. For a topological insulator excluding the vacuum contribution, we obtain,

\[
\tilde{\Omega}(0, \mu) = \frac{eB}{\hbar} \left[ \frac{\mu}{2} + (E_0/2 - \mu) \Theta(\mu - E_0/2) \right.
\]

\[
+ \sum_{N=1}^{\infty} \left( E_{N,+} - \mu \right) \Theta(\mu - E_{N,+}) \left. \right],
\]

where we have used equation (12) for the density of states, and have assumed that all \(E_{N,+}\) energies remain negative. The magnetization \(M(\mu)\) as a function of \(\mu\) derived from equation (17) is, by arrangement, zero for zero chemical potential. In figure 2, we display results for \(M(\mu)\) as a function of \(\mu\) for two fixed values of magnetic field (\(B = 2\) and 4 Tesla). Including a vacuum contribution would simply shift the curves up by a constant. In all cases, the Dirac energy \(E_1/\sqrt{B}\) is set at 10.4 meV T\(^{-1}\) which is a characteristic value of topological insulators. Two values of the Schrödinger scale are considered. \(E_0/B = 0\) (figure 2(a)) corresponds to the pure Dirac case (e.g. graphene) and is included only for comparison. Figure 2(b) shows \(E_0/B = 1.1\) meV T\(^{-1}\). In all cases, the magnetization displays a saw-tooth oscillation pattern where the location of the vertical jumps is best understood by examining a plot of the Landau level dispersion shown in figure 3 for the specific case of \(E_0/B = 1.1\) meV T\(^{-1}\). The black lines correspond to the Landau level energies as a function of \(B\). The vertical green and orange lines correspond to the two values of \(B\) which yield the curves in figure 2(b). The colored circles mark the intersection of the Landau levels with the lines of constant \(B\).
It is clear that the intersection of the Landau levels with these constant magnetic field lines, correspond to the vertical jumps in $M(\mu)$ seen in figure 2(b).

The slope of $M(\mu)$ is of particular importance as it is related to the quantized Hall conductivity ($\sigma_H$) through the Streda formula [34] $\frac{\partial M(\mu)}{\partial \mu} = (1/e)\sigma_H$. Indeed, by examining the inset of figure 2(a) (pure Dirac), we see a quantization in the slope at values of $v = \pm 1/2, \pm 3/2, \ldots$. Note: for graphene, two-fold valley and spin degeneracies would be included which give rise to the well known filling factors $\pm 2, \pm 6, \pm 10, \ldots$. In the inset of figure 2(b) (small $E_0$), we again see a half-integer quantization of the slope [20, 34]; however, now it begins at $v = \pm 1/2$. This is due to the Schrödinger term moving the location of the $N = 0$ Landau level to positive energy. Therefore, at energies below $E_0/2$, the highest filled level is $N = 1, s = -$. If we were to look at negative values of $\mu$, we would see the full negative half-integer quantization. We note that the addition of a small Schrödinger correction to the Hamiltonian does not break the half-integer quantization of the pure Dirac limit. It does, however, result in a negative Hall conductivity for positive $\mu < E_0/2$.

To see this robust quantization, return to equation (17). Taking the derivative with respect to $\mu$, we obtain

$$\frac{\partial \tilde{\Omega}}{\partial \mu} = \frac{eB}{h} \left[ \frac{1}{2} - \Theta(\mu - E_0/2) - \sum_{N=1}^{\infty} \Theta(\mu - E_{N,s}) \right],$$

(18)

for $\mu > 0$, where we ignore the $\delta$-functions resulting from the derivatives of the $\Theta$-functions. Differentiating with respect to $B$, we obtain

$$-\frac{\partial}{\partial B} \frac{\partial \tilde{\Omega}}{\partial \mu} \equiv \frac{\partial M}{\partial \mu} = \frac{e}{\hbar} \left[ -\frac{1}{2} + \Theta(\mu - E_0/2) \right]$$

$$+ \sum_{N=1}^{\infty} \Theta(\mu - E_{N,s})].$$

(19)

It is clear, that for $\mu < E_0/2$, the Hall plateau occurs at $-1/2$, for $E_0/2 < \mu < E_1$, the quantization is $1/2$, etcetera.

3.2. Comparison with the Schrödinger Limit

For comparison, it is useful to consider the dominant-Schrödinger regime ($E_0 \gg E_1$). This limit is well understood, and will, therefore, not be discussed in detail. In this system, the grand thermodynamic potential is [37]

$$\Omega_{NR}(T, \mu) = -T \int_{-\infty}^{\infty} N(\omega) \ln \left(1 + e^{(\mu - \omega)/T} \right) d\omega, \quad (20)$$

where, again, $N(\omega)$ is given by equation (12). At $T = 0$, this gives

$$\Omega_{NR}(0, \mu) = -\frac{eB}{\hbar} \left[ (\mu - E_0/2) \Theta(\mu - E_0/2) \right]$$

$$+ \sum_{N=1}^{\infty} (\mu - E_{N,s}) \Theta(\mu - E_{N,s})].$$

(21)

Again, the magnetization is given by $M = -\partial \Omega_{NR}/\partial B$.

In the pure Schrödinger limit ($E_1 = 0$), the Landau levels evolve linearly with $B$ unlike the $\sqrt{B}$ dependence observed for Dirac fermions (see figure 3(a)). The $N = 0, 1, 2, \ldots$ levels for $s = -$ are degenerate in energy with the $N = 1, 2, 3, \ldots$ Landau levels for $s = +$. There is no $N = 0, s = -$ level. The Hall conductivity is given by an even-integer quantization (i.e. $\sigma_H = (e^2/\hbar)v$ where $v = 0, 2, 4, \ldots$).

If a small relativistic contribution is included, the results of figures 4(a) and (b) are obtained for the Landau level dispersion as a function of $B$ and $M(\mu)$ as a function of $\mu$, respectively. The Landau levels are no longer linear in $B$ but have an additional $\sqrt{B}$ dependence. As a result, the degeneracy of the levels is broken. The impact of the degeneracy breaking of the Landau levels on the magnetization is seen in figure 4(b). For small $E_1$, the Landau level spectrum is close to the sum of two Schrödinger systems which are separated by a small energy difference. As a result, the peaks in the magnetization split in two but remain close in energy. This affects the quantization of the Hall conductivity (see the inset of figure 4(b)). With the breaking of the two-fold degeneracy, $\sigma_H$ is no longer quantized in even integers but can take any integer value [20]. For small $E_1$, the odd integer plateaus exist over a much smaller range of $\mu$ than the even-integer values. As $E_1$ is increased, the odd integer steps become more prominent. Unlike, the robust half-integer quantization seen in the relativistic limit (see the previous subsection), the addition of a relativistic term to the pure Schrödinger regime changes the Hall conductivity from an even-integer to an integer quantization. This quantization can...
be seen by returning to equation (21) and taking the derivative with respect to \( \mu \), giving

\[
\frac{\Theta(M/\mu)}{\mu} = \frac{e \hbar}{e} \left[ \Theta(\mu - E_0/2) + \sum_{N=1,\sigma=\pm} \Theta(\mu - E_N,\sigma) \right].
\]

Taking the derivative with respect to \( B \), we obtain

\[
\frac{\partial}{\partial B} \frac{\partial\Omega_{NR}}{\partial\mu} = \frac{e \hbar}{e} \left[ \Theta(\mu - E_0/2) + \sum_{N=1,\sigma=\pm} \Theta(\mu - E_N,\sigma) \right].
\]

For the pure Schrödinger limit, \( E_{N,-} \) is degenerate with \( E_{N-1,+} \). For \( \mu < E_0/2 \), \( \sigma_H = 0 \). The first plateau occurs for \( E_0/2 < \mu < E_1, \sigma_H = 0 \), and has a value of 2 since \( E_1 = E_0/2 \). The next step occurs for \( \mu > E_1, \sigma_H = 0 \), which is degenerate with \( E_{2,-} \), so \( \sigma_H \) is incremented by \( 2e^2/h \). All further steps are also by two. The inclusion of a small spin–orbit contribution breaks the degeneracy of the \( s = \pm 1 \) bands and now the steps occur at all integer values. As the energy difference between \( E_{N,-} \) and \( E_{N-1,+} \) is small for \( E_0 \gg E_1 \), the even integer steps are only visible over a small range of \( \mu \). For a more detailed discussion of the Hall effect in this system, the reader is referred to [20].

4. Magnetic oscillations

4.1. Dirac limit

We now turn our attention to the purely oscillating part of the magnetization. To begin, note that equation (12) can be written as

\[
N(\omega) = \frac{eB}{\hbar} \frac{d}{d\omega} \left[ \Theta(\omega - E_0/2) + \sum_{N=1,\sigma=\pm} \Theta(\omega - E_N,\sigma) \right].
\]

The oscillatory part of equation (24) is obtained by applying the Poisson formula

\[
\sum_{N=1}^{\infty} F(N) = - \frac{1}{2} F(0) + \int_0^\infty F(x) \, dx + 2 \sum_{k=1}^{\infty} \int_0^\infty F(x) \cos(2\pi kx) \, dx.
\]

This gives (see equation (19) in [38]),

\[
N(\omega) = \frac{eB}{\hbar} \frac{d}{d\omega} \left\{ \frac{1}{2} \Theta(\omega - E_0/2) - \frac{1}{2} \Theta(\omega + E_0/2) + \left[ \Theta(\omega + E_0/2) + \Theta(\omega - E_0/2) - \Theta(\omega - \omega_{\text{min}}) \right] \times \left[ x_1 + \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin(2\pi kx_1) \right] + \Theta(\omega - \omega_{\text{min}}) \times \left[ x_2 + \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin(2\pi kx_2) \right] \right\},
\]

where

\[
\omega_{\text{min}} = - \frac{E_0 P}{2} - \frac{E_0}{8P},
\]

and

\[
x_i = P + \frac{\omega}{E_0} + \left( \frac{\omega}{E_0} \right)^2 \sqrt{p_0^2 + \frac{2P\omega}{E_0} + \frac{1}{4}}.
\]
All the necessary information about the de-Haas–van-Alphen (dHvA) oscillations can be extracted from the first term of equation (13). Thus, we examine

\[ \tilde{\Omega}(0, \mu) = \int_0^\mu (\omega - \mu)N(\omega)d\omega, \]  

(29)

where the vacuum, and \( eB\mu/(2h) \) terms are excluded as we are only interested in the oscillating part of \( M(\mu) \). In addition, only the trigonometric terms of \( N(\omega) \) [equation (26)] will contribute to the oscillations. As is argued by Suprunenko et al [38], since we are working at low energy, we should only consider small values of \( N \) and thus, in the relativistic regime, we should only include the \( x_1 \) term.

Focusing only on the terms which contribute to the oscillations, the magnetization is given by

\[ \tilde{M}(0, \mu) = -\frac{d}{dB} \int_0^\mu (\omega - \mu)N_{osc}(\omega)d\omega, \]  

(30)

where

\[ N_{osc}(\omega) = \frac{eB}{\hbar} \sum_{k=1}^\infty \frac{d}{d\omega} N_k^k(\omega), \]  

(31)

and

\[ N_k^k(\omega) = \frac{\Theta(\omega + E_0/2) + \Theta(\omega - E_0/2) - \Theta(\omega - \omega_{\text{min}})}{\pi k} \]  

(32)

As we take only positive \( \mu \) and \( E_0 \), this may be written as

\[ N_k^k(\omega) = \frac{\Theta(\omega - E_0/2) \sin(2\pi k x_1)}{\pi k}. \]  

(33)

In the relativistic limit, we use the approximation

\[ x_1 \approx \frac{\omega^2}{2E_1^2} \left(1 - \frac{\omega}{mv_F^2}\right) = \frac{E_0}{8mv_F^2}, \]  

(34)

which can be found by Taylor expanding equation (28). In what follows, we will only keep amplitudes which are first order in \( 1/m \). Substituting the \( k \)-th component of equation (31) into equation (30), and integrating by parts, we obtain

\[ \tilde{M}_k(0, \mu) \approx \frac{d}{dB} \left\{ \frac{eB}{\hbar} \int_0^\mu \frac{\partial}{\partial\omega} N_k^k(\omega)d\omega \right\}. \]  

(35)

Therefore, assuming \( \mu > E_0/2 \), we have

\[ \tilde{M}_k(0, \mu) \approx \frac{d}{dB} \left\{ \frac{eB}{\pi k h} \int_0^\mu \sin(2\pi k x_1)d\omega \right\} \]  

(36)

We ignore the second term as, for the limit of interest (\( B \to 0 \), \( E_0 \to 0 \)); hence, expanding the sine to lowest order and performing the integral, gives a magnetization that goes like \( B^2 \) and is thus negligible.

We are left with evaluating

\[ M_{osc}(0, \mu) = \frac{e}{\pi k h} \int_0^\mu \left[ \sin(2\pi k x_1) - 2\pi k x_1 \cos(2\pi k x_1) \right] d\omega. \]  

(37)

To solve this integral, define

\[ y \equiv \omega^2 \left(1 - \frac{\omega}{mv_F^2}\right) = \frac{E_0^2}{4}. \]  

(38)

For \( m \to \infty, y \approx \omega^2 \) so \( \sqrt{y} \approx \omega \). Therefore, we can write

\[ y \approx \omega^2 \left(1 - \frac{\sqrt{y}}{mv_F^2}\right) = \frac{E_0^2}{4}. \]  

(39)

Our simplified integral is now

\[ M_{osc}(0, \mu) \approx \frac{e}{2\pi k h} \int_0^\alpha \left[ \sin \left( \frac{\pi k y^{1/2}}{E_1^2} \right) - \frac{\pi k y^{1/2}}{E_1^2} \cos \left( \frac{\pi k y^{1/2}}{E_1^2} \right) \right] \frac{dy}{\sqrt{y}} \times \frac{1 + \sqrt{\pi} \mu}{mv_F^2}, \]  

(40)

where

\[ \alpha = \mu^2 \left(1 - \frac{\mu}{mv_F^2}\right) = \frac{E_0^2}{4}. \]  

(41)

We then make the substitution \( y = x\alpha \), to obtain

\[ M_{osc}(0, \mu) \approx \frac{e}{2\pi k h} \int_0^1 \left[ \sqrt{\frac{\mu^2[1 - \mu/(mv_F^2)]}{x}} \left( \sin(ax) - ax \cos(ax) \right) \right] dx, \]  

(42)

where \( a \equiv \pi k\alpha/E_1^2 \).

Let us consider the first integral of equation (42). We have

\[ I_1 = \int_0^1 [\sin(ax) - ax \cos(ax)] \frac{dx}{\sqrt{x}} \]  

= \int_0^1 \sin(ax) \frac{dx}{\sqrt{x}} - \int_0^1 \sqrt{x} \frac{d}{dx} \sin(ax) \, dx. \]  

(43)

This can be integrated by parts to give

\[ I_1 = 3 \sqrt{\frac{\pi}{2a}} S \left( \sqrt{\frac{2a}{\pi}} \right) - \sin a, \]  

(44)

where we have used the definition of the Fresnel sine integral

\[ S(z) = \int_0^z \sin \left( \frac{\pi t^2}{2} \right) \, dt. \]  

(45)
We are interested in the oscillations for small $B$. In the limit $B \to 0$, $a \to \infty$. Expanding $S(\sqrt{2a/\pi})$ to first order in $1/a$, we obtain
\[
S(\sqrt{2a/\pi}) \approx \frac{1}{2} \frac{1}{\sqrt{2\pi a}} \cos a;
\]
and thus, in the limit of interest,
\[
I_1 \approx -\sin a.
\]

Next, consider the second term of equation (42). This is a standard integral and gives
\[
I_2 = \int_0^1 [\sin(ax) - ax \cos(ax)]dx \approx -\sin a
\]
in the limit of $B \to 0$. Combining $I_1$ and $I_2$, we obtain
\[
M_{osc}(0, \mu) \approx -\frac{e}{\hbar} \sum_{k=1}^\infty \left[ 1 + \frac{\mu}{2mv_F^2} \right] \sin(2\pi k x_1) \frac{\pi k}{\pi k},
\]
where
\[
x_1 = \frac{\mu^2}{2\hbar v_F e B} \left( 1 - \frac{\mu}{mv_F^2} \right) - \frac{\hbar e B}{8mv_F^2}.
\]
If we compare this to the customary [39]
\[
x_1 = \frac{\hbar A(\mu)}{2\pi e B} - \gamma,
\]
the coefficient of the $1/B$ dependence in equation (50) is indeed the area of the cyclotron orbit:
\[
A(\mu) \approx \frac{\pi \mu^2}{\hbar^2 v_F^2} \left( 1 - \frac{\mu}{mv_F^2} \right).
\]
The remainder is a phase shift which has a linear dependence on $B$. It is new and is not part of a standard semiclassical quantization scheme [14]. It is
\[
\gamma = \frac{\hbar e B}{8mv_F^2}.
\]
In the pure Dirac limit ($m \to \infty$), $A(\mu)$ reduces to the correct value [38, 39] of $\pi \mu^2/(\hbar^2 v_F^2)$. The inclusion of a small Schrödinger contribution gives a correction of $-\{\mu^2/(\hbar^2 v_F^2)[\mu/(mv_F^2)]\}$ to $A(\mu)$ and the cyclotron orbit area is reduced due to the narrowing of the conduction band (see figure 1(a)). The pure Dirac limit should also have a phase shift of $0$ associated with a Berry’s phase of $\pi$ [38]. Here, we find the inclusion of an additional $E_0$ term results in a finite phase shift of $\hbar e B/(m^2 v_F^2)$ which is linear in $B$ but very small for $m \to \infty$. There is also a correction to the overall amplitude of the quantum oscillations. While there can be a significant Schrödinger contribution to the low-energy Hamiltonian of a topological insulator, we find here [equation (53)] that the phase offset $\gamma$ of the magnetic oscillations is essentially zero; this is in agreement with previous semiclassical considerations [10, 13, 14, 40]. This is the same result as for pure relativistic particles for which the Berry phase is $\pi$. The result is in agreement with experimental findings. For example, the high-field SdH data in Bi$_2$Te$_2$Se by Xiong et al [41] shows zero offset. The same holds for many other topological insulators as documented in the extensive review by Ando [17]. This signature of zero offset has often been used to distinguish between oscillations coming from the surface states and those originating from the bulk [17]. In SdH (oscillations in the conductivity) and dHvA (magnetization oscillations), experiment can also extract the cyclotron orbit area [42]. Equation (52) shows that this quantity contains a correction from the pure Dirac result of order $\mu/(mv_F^2)$ which may allow one to extract the Schrödinger mass $m$ from such data.

4.2. Dingle factor

The Landau level broadening due to impurity scattering can be included by convolving the grand thermodynamic potential with a scattering function $P_\Gamma(\epsilon)$. That is [37],
\[
\Omega(\mu) = \int_{-\infty}^{\infty} d\omega P_\Gamma(\omega - \mu) \Omega(\omega),
\]
where
\[
P_\Gamma(\epsilon) = \frac{\Gamma}{\pi (\epsilon^2 + \Gamma^2)},
\]
and $\Gamma$ is a small broadening parameter.

To solve for the Dingle factor, consider the magnetization in the presence of impurities:
\[
M_{osc}(\mu, \Gamma) = \int_{-\infty}^{\infty} d\omega \frac{\Gamma}{\pi (\omega - \mu)^2 + \Gamma^2} M_{osc}(\omega),
\]
where $M_{osc}(\omega)$ is given by letting $\mu \to \omega$ in equation (49) for the relativistic regime. To obtain a first order approximation to the Dingle factor, we note that the Lorentzian is peaked around $\omega = \mu$ and that $\omega/(2mv_F^2)$ is small. We, therefore, fix $\omega$ at $\mu$ in these terms. Here, we also assume $\Gamma \ll \mu$. The $k^{th}$ component of equation (56) becomes
\[
M_{osc}^{(k)}(\mu, \Gamma) \approx \int_{-\infty}^{\infty} d\omega \frac{\Gamma}{\pi (\omega - \mu)^2 + \Gamma^2} \left[ -\frac{e}{\hbar} \frac{\omega}{2\pi k} \sin \left( \frac{k\pi}{E_1} \right) \frac{\omega (1 - \frac{\mu}{mv_F^2})}{mv_F^2} \right].
\]
If we ignore the phase factor for the purpose of integrating, and define $1/E_1^{(k)} = (1/E_1)[1 - \mu/(mv_F^2)]$, we have mapped our results onto the pure Dirac limit and obtain the renormalized results of [37] (see their equation (8.10)):
\[
M(\mu, \Gamma) \approx -\frac{e}{\hbar} \frac{\mu^2}{2\mu} - \frac{\Gamma^2}{2\mu} \sum_{k=1}^{\infty} \sin(2\pi k \bar{x}_1) \frac{\pi k}{R_D},
\]
where
\[
\bar{x}_1 \approx \frac{\mu^2 - \Gamma^2}{2\hbar v_F e B} \left( 1 - \frac{\mu}{mv_F^2} \right) - \frac{\hbar e B}{8mv_F^2}.
\]
and

$$R_D = \exp\left[-2\pi \Gamma k \frac{\mu}{\hbar v_F eB} \left(1 - \frac{\mu}{m v_F^2}\right)\right]. \quad (60)$$

This reduces to the expected [37] $R_D = \exp[-2\pi \Gamma \mu/(\hbar v_F^2 eB)]$ in the pure-Dirac limit. We note that in this regime, the damping is dependent on the chemical potential [37]; thus, for large $\mu$, magnetic oscillations would be difficult to observe [37]. When particle-hole symmetry is broken by a subdominant Schrödinger piece, the damping in the Dingle factor $R_D$ for a given $\mu$ is reduced by a factor of $1 - \mu/(m v_F^2)$ which makes it more favourable for observation of magnetic oscillations as compared to the pure relativistic case.

4.3. Finite T

Finite temperature effects are included by convolving $M_{\text{osc}}(\omega)$ with $P_T(\omega - \mu)$, where [37]

$$P_T(\omega) = \frac{1}{4T \cosh^2\left(\frac{\epsilon}{2T}\right)}, \quad (61)$$

and, $M_{\text{osc}}(\omega)$ is the result of letting $\mu \to \omega$ in equation (49). Again, the damping factor is peaked around $\omega = \mu$. By applying the same procedure as in section 4.2, we map our results onto those of [37] (see their equations (8.16) and (8.17)) and obtain

$$M(\mu, T) \approx -\frac{e}{2} \left(1 + \frac{\mu}{2m v_F^2}\right) \sum_{k=1}^{\infty} \frac{\sin(2\pi k x)}{\pi k} R_T, \quad (62)$$

where

$$R_T = \frac{k \lambda}{\sinh(k \lambda)}, \quad (63)$$

with

$$\lambda = \frac{2\pi^2 T \mu}{\hbar v_F eB} \left(1 - \frac{\mu}{m v_F^2}\right). \quad (64)$$

We note that $R_T$ is 1 in the limit $T \to 0$ as it must be. Like impurity scattering, the damping of oscillations due to temperature is dependent on $\mu$. For large $\mu$, magnetic oscillations are difficult to observe. The introduction of a small Schrödinger term reduces the damping by $1 - \mu/(m v_F^2)$ similar to what we found in the previous section for impurity scattering.

5. Conclusions

We have computed the magnetization of the topologically protected helical electronic states that exist at the surface of a three dimensional topological insulator. We focus on adding a small Schrödinger quadratic-in-momentum term to a dominant linear-in-momentum Dirac term. Typical for a topological insulator, is a Schrödinger mass ($m$) on the order of 1/10 the bare electron mass and a Dirac Fermi velocity ($v_F$) of order $4 \times 10^5$ m s$^{-1}$. Thus, the relevant magnetic energy scales for $B = 1$ T are $E_0 \sim 1.0$ meV and $E_1$ of order 10.0 meV. The ratio of Schrödinger to Dirac energy scales $E_0/E_1$ is much less than one. This is the opposite limit to that found for semiconductors used in spintronic applications where the spin–orbit coupling is 100 times smaller and $E_0/E_1$ is much greater than one.

In the topological insulator limit, the magnetization as a function of chemical potential $\mu$ shows the usual jagged sawtooth oscillations. The teeth occur at values of $\mu$ which reflect the underlying Landau level energies. While the distance between the teeth for fixed $B$ as a function of $\mu$ is altered for a small Schrödinger contribution, there is no qualitative difference to the pure Dirac limit. For comparison, and in sharp contrast to this finding, adding a small spin–orbit coupling term to a quadratic band introduces a new distinct set of teeth to those that arise when $v_F = 0$. These new teeth are much less prominent and appear ghosted on the side of and slightly displaced from the main set. They disappear when spin–orbit coupling is neglected and their distinctness increases with $v_F$.

The slope of the magnetization as a function of $\mu$ reflects the quantization of the Hall conductivity and is found to remain unchanged for a topological insulator as the Schrödinger contribution is increased and the half integer quantization of the pure Dirac regime is retained. Of course, the amount by which the chemical potential needs to be incremented to go from one plateau to the next is changed. These results are in striking contrast to the case when the spin–orbit term is small and the Schrödinger mass term dominates. While the spin–orbit coupling splits the Landau level degeneracy of the pure non-relativistic case, the well known integer quantization of the plateaus remains even though two distinct sets are observed. The second set merges with the dominant set for $v_F = 0$, and increases in prominence with increasing spin–orbit coupling.

Applying the Poisson formula to the thermodynamic potential, we extract an analytic expression for the magnetic oscillations $M_{\text{osc}}$ which includes a first correction in a small Schrödinger term. Our new expression properly reduces to the known result of [37] when we set $E_0$ to zero which is the relativistic generalization of the classical Lifshitz–Kosevich (LZ) result. In this case, the phase offset which appears in the semiclassical expression for $M_{\text{osc}}$ is equal to zero in contrast to the value of 1/2 of LK theory [39]. This difference can be traced back to a Berry phase which is $\pi$ in the Dirac limit and zero in the Schrödinger limit. When a sub-dominant mass term is added to the Dirac limit, the phase offset is no longer zero but has a correction of order $(E_0/E_1)^2$ which is equal to $\hbar e B/(8m^2 v_F^2)$. This new contribution to the phase is zero in the pure relativistic case $m \to \infty$ and is also negligible when the magnetic field tends to zero. At the same time, the Berry phase is totally unchanged by the $E_0/E_1$ correction term and retains its value of $\pi$. Nevertheless, in a topological insulator, there is a small phase offset in the quantum oscillations which has its origin in the bending of the electronic dispersion curves away from linearity which narrows slightly the cross-section of the conduction band Dirac cone.

We have considered the effect of impurity scattering on the magnetization in a constant scattering rate ($\Gamma$) approximation. We assume the same value of $\Gamma$ applies to all the Landau levels. This provides a Dingle factor in $M_{\text{osc}}$. In the pure-relativistic case, the exponential argument of the damping
factor depends linearly on chemical potential [37]. Here, we find a correction to this damping by a further multiplicative factor in the exponential of $1 - \mu/(mv_F^2)$. This reduces the effectiveness of damping over its pure relativistic value. A similar correction to the temperature factor is also noted.

Acknowledgments

We thank E J Nicol for discussions. This work has been supported by the Natural Science and Engineering Research Council of Canada and, in part, by the Canadian Institute for Advanced Research.

References

[1] Kane C L and Mele E J 2005 Phys. Rev. Lett. 95 226801
[2] Hasan M Z and Kane C L 2010 Rev. Mod. Phys. 82 3045–67
[3] Qi X L and Zhang S C 2011 Rev. Mod. Phys. 83 1057–110
[4] Chen Y L et al 2009 Science 325 178–81
[5] Hsieh D et al 2009 Nature 460 1101–5
[6] Chen Y L et al 2010 Science 329 659–62
[7] Xu S Y et al 2011 Science 332 560–4
[8] Dzero M, Roy B, Sau J D and Galitski V 2014 Phys. Rev. B 90 155314
[9] Hancock J N, van Mechelen J L M, Kuzmenko A B, van der Marel D, Brüne C, Novik E G, Astakhov G V, Buhmann H and Molenkamp L W 2011 Phys. Rev. Lett. 107 136803
[10] Wright A R and McKenzie R H 2013 Phys. Rev. B 87 085411
[11] Li Z and Carbotte J P 2013 Phys. Rev. B 87 155416
[12] Li Z and Carbotte J P 2013 Phys. Rev. B 88 045414
[13] Wright A R 2013 Phys. Rev. B 87 085426
[14] Fuchs J N, Piechon F, Goerbig M O and Montambaux G 2010 Eur. Phys. J. B 77 351
[15] Taskin A A and Ando Y 2011 Phys. Rev. B 84 035301
[16] Mikitik G P and Sharlai Y V 2012 Phys. Rev. B 85 033301
[17] Ando Y 2013 J. Phys. Soc. Japan 82 102001
[18] Raoux A, Morigi M, Fuchs J N, Piechon F and Montambaux G 2014 Phys. Rev. Lett. 112 026402
[19] Kishigi K and Hasegawa Y 2014 Phys. Rev. B 90 085427
[20] Li Z and Carbotte J P 2014 Phys. Rev. B 89 085413
[21] Gusynin V P, Sharapov S G and Carbotte J P 2007 J. Phys.: Condens. Matter 19 026222
[22] Pound A, Carbotte J P and Nicol E J 2012 Phys. Rev. B 85 125422
[23] Schafgans A A, Post K W, Taskin A A, Ando Y, Qi X L, Chapler B C and Basov D N 2012 Phys. Rev. B 85 195440
[24] Bychkov Y A and Rashba E I 1984 JETP Lett. 39 78
[25] Bychkov Y A and Rashba E I 1984 J. Phys. C: Solid State Phys. 17 6039
[26] Luo J, Munekata H, Fang F F and Stiles P J 1990 Phys. Rev. B 41 7685–93
[27] de Andrade e Silva E A, La Rocca G C and Bassani F 1994 Phys. Rev. B 50 8523–33
[28] Nitta J, Akazaki T, Takayanagi H and Enoki T 1997 Phys. Rev. Lett. 78 1335–8
[29] Grundler D 2000 Phys. Rev. Lett. 84 6074–7
[30] Wang Z F and Vasilopoulos P 2005 Phys. Rev. B 72 085344
[31] Zarea M and Ulloa S E 2005 Phys. Rev. B 72 085342
[32] Žutić I, Fabian J and Das Sarma S 2004 Rev. Mod. Phys. 76 323–410
[33] Fabian J, Matos-Abiague A, Ertler C, Stano P and Žutić I 2007 Acta Phys. Slov. 57 565–907
[34] Wang Z, Fu Z G, Wang S X and Zhang P 2010 Phys. Rev. B 82 085429
[35] Zhang H, Liu C X, Qi X L, Dai X, Fang Z and Zhang S C 2009 Nat. Phys. 5 438
[36] Liu C X, Qi X L, Zhang H, Dai X, Fang Z and Zhang S C 2010 Phys. Rev. B 82 045122
[37] Sharapov S G, Gusynin V P and Beck H 2004 Phys. Rev. B 69 075104
[38] Suprunenko Y, Gorbar E V, Sharapov S G and Loktev V M 2008 Fiz. Nizk. Temp. 34 1033–9
[39] Luk’yanchuk I A and Kopelevich Y 2004 Phys. Rev. Lett. 93 166402
[40] Taskin A A, Ren Z, Sasaki S, Segawa K and Ando Y 2011 Phys. Rev. Lett. 107 016801
[41] Xiong J, Luo Y, Kho Y, Jia S, Cava R J and Ong N P 2012 Phys. Rev. B 86 045314
[42] Lawson B J, Hor Y S and Li L 2012 Phys. Rev. Lett. 109 226406