The Zeeman, spin-orbit, and quantum spin Hall interactions in anisotropic and low-dimensional conductors

Aiying Zhao¹,², Qiang Gu¹,*, Timothy J Haugan³ and Richard A Klemm²,³,*

¹ Department of Physics, University of Science and Technology Beijing, Beijing 100083, People’s Republic of China
² Department of Physics, University of Central Florida, Orlando, FL 32816-2385, United States of America
³ U. S. Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433-7251, United States of America

E-mail: qgu@ustb.edu.cn and richard.klemm@ucf.edu

Received 3 July 2020, revised 21 September 2020
Accepted for publication 10 November 2020
Published 3 December 2020

Abstract
To construct a microscopic theory of electrons and holes in anisotropic conductors that self-consistently treats their band effective mass anisotropy with their interactions with applied electric and magnetic fields, the Dirac equation is extended for an electron or hole in an orthorhombically-anisotropic conduction band. Its covariance is established both by a modified version of the Klemm–Clem transformations to a space in which it is isotropic, and also in its fully anisotropic form by making the most general proper and improper Lorentz transformations, proving its validity in both the relativistic and non-relativistic limits. The appropriate Foldy–Wouthuysen transformations are extended to expand about the non-relativistic Hamiltonian limit to fourth order in the inverse of the particle’s Einstein rest energy. The results have important consequences for magnetic measurements of many classes of clean anisotropic semiconductors, metals, and superconductors. In all of these cases, the Zeeman interaction is found to depend strongly upon the effective mass anisotropy. When an electron or hole is traveling in an atomically thin one-dimensional conduction band, its Zeeman, spin–orbit, and quantum spin Hall interactions are vanishingly small. Accurate expressions for the Zeeman, spin–orbit and quantum spin Hall interactions for two-dimensional conductors are provided.

Keywords: lower-dimensional conductors, relativistic effects, anisotropy effects, Zeeman interactions, quantum spin Hall interactions, spin-orbit interactions

Supplementary material for this article is available online
(Some figures may appear in colour only in the online journal)

1. Introduction

Unlike a trapped proton or chargeless neutron that have magnetic moments [1, 2], Dirac showed that the magnetic moment of an electron or positron arises only from its charge and mass

* Authors to whom any correspondence should be addressed.

while in motion [3]. In free space in three dimensions (3D), the magnetic moment due to the motion of an electron or hole is \( g \mu_B = \frac{e |e|}{2m} \), where \( |e| \) is its charge, \( \pm m \) is its mass, \( \hbar = \frac{h}{2\pi} \), \( h \) is Planck’s constant, and \( g = 2 \). But what is the \( g \)-factor when an electron propagates in a conduction band of an anisotropic or low-dimensional metal or superconductor? What is it if the
crystal were a monolayer of FeSe or NbSe2 [4, 5], either metallic or superconducting twisted bilayer graphene [6], a metallic single-walled carbon nanotube [7, 8] or a chain of single atoms of Au or Al [9–11]? What is it for a hole in its crystalline conduction band? These are fundamental questions that have been completely ignored in essentially all papers written on semiconductors, metals, and superconductors. Here we present a rigorous answer to all of these questions. The results imply that strong changes in the interpretations of many existing experiments are necessary, and many theoretical treatments of lower-dimensional conductors need to be greatly modified. In particular, the Zeeman, spin–orbit, and quantum spin Hall interactions of an electron or hole depend strongly upon the anisotropy and/or dimensionality of the conduction band.

In semiconductors such as Si and Ge, and in metallic Bi, the lowest energy conduction bands have ellipsoidal symmetry due to their crystal electric potentials, with ε0(p) = \[ε_0(\mathbf{p}) = \sum_{j=0}^{3} \mathbf{p}_j^2 / (2m_j)\] about some point in the first Brillouin zone [12–15]. The normal state of the high-temperature superconductor YBa2Cu3O7–δ (YBCO) is metallic in all three orthorhombic directions [16, 17], so that ε0(p) reasonably has that form, with m3 ≫ m1, m2. Much greater anisotropy of the normal state conduction band of the layered organic superconductor κ-(ET)2Cu(SCN)2, where ET is bis(ethylenedithio)tetrathiafulvalene, was found by de Haas van Alphen measurements [18]. In addition, many materials of current interest contain elements with 5d electronic orbitals, for which the conduction bands exhibit strong relativistic effects [19].

Although standard temperature T adiabatic Knight shift K(T) measurements on the planar 63Cu nuclei of YBCO showed the standard K(T) behavior through the superconducting transition temperature Tc, when the strong time-independent magnetic induction \(B_0 \parallel \mathbf{c}\) (perpendicular to the direction of mass m3) and the component \(B_1(t)\) oscillatory in time t was perpendicular to that direction, measurements when \(B_0 \parallel \mathbf{c}\) and \(B_1(t) \perp B_0\) showed anomalous, T-independent K behavior through the superconducting state [20]. Those K(T) measurements were inconsistent with the standard model that neglects any anisotropy features of the Fermi surface and of the Zeeman interaction [21]. In fact, Slichter remarked that those results were ‘foolish’, as he did not have an explanation other than it might have something to do with planar near-neighbor Cu–Cu couplings [22]. Further questions about Knight shift measurements were raised by the apparent incompatibility of pulsed, non-adiabatic 17O K(T) measurements when \(B_0 \parallel \mathbf{c}\) with scanning tunneling measurements of the superconducting gap in the highly layered Sr2RuO4 [23–26]. However, the Knight shift dimensionality issue still remains unresolved in the quasi-one-dimensional organic superconductor (TMTSF)2PF6, where TMTSF is tetramethyl–tetrathiafulvalene [27], for which strong Zeeman dimensionality issues could arise for both the strong constant \(B_0\) and the weaker perpendicular \(B_1(t)\).

The anisotropic Ginzburg–Landau model of superconductors, in which the effective mass of the paired quasiparticles is assumed to have orthorhombic anisotropy [17, 28], has been widely used to fit the anisotropy of the upper critical magnetic induction \(B_{c2}\) in the low \(B\) direction not too far below the superconducting transition temperature \(T_c\), and can also be accurately applied to the anisotropy of the lower critical induction \(B_{c1}\) [17, 29]. But such a phenomenological model cannot be relevant at high \(B\) values exceeding the Pauli limit \(B_p = 1.867 T_c (T/K)\), since it does not contain any Pauli pair-breaking due to the Zeeman interaction. In the case of many layered superconductors, the Klemm–Luther–Beasley (KLB) model has been used to provide a possible explanation of the Pauli limit being exceeded for the field parallel to the layers based upon strong spin–orbit impurity scattering of the paired conduction particles in the dirty limit [30].

Now excellent quality lower-dimensional conductors and superconductors can be prepared in the clean limit [6], so it is important to reassess the effects of the crystal anisotropy upon the Zeeman energy of the conduction particle. We note in particular that clean bi- and trilayers of superconducting NbSe2 exhibited \(B_{c2}(T)\) that violated the Pauli limit by about a factor of 5, and near to \(T_c\), a clean monolayer of NbSe2 had an even larger \(B_{c2}(T)\) [5], but the samples are too clean to apply the KLB model. We emphasize that the conventional Pauli limit due to the rest-mass Zeeman interaction with \(g = 2\) is a relativistic effect arising from the linear Dirac equation for the isotropic vacuum motion of the charge and mass of an electron or hole when expanded to second order in the particle’s inverse Einstein rest energy. Hence it is inconsistent to treat the orbital motion of the paired particles with effective mass anisotropy, but to completely ignore such anisotropy with regards to their spins. It is crucial to construct a single theory of an electron or hole in an anisotropic or lower-dimensional conduction band that is valid in both the relativistic and non-relativistic limits, in order to treat the non-relativistic orbital and relativistically-induced spin aspects of effective mass anisotropy self consistently.

Here we present a theory of an electron or hole in an orthorhombically anisotropic conduction band that is valid in the entire region from the relativistic to the non-relativistic limits. In the relativistic limit, we prove its covariance in two ways. We first employ a modified form of the Klemm–Clem transformations (KCT) to transform it to gauge-invariant isotropic form [17, 28], in which its invariance under each of the improper transformations of charge conjugation, parity, and time reversal (CPT) is easily established. Second, we demonstrate in the supplementary material that in its fully relativistic form, it is invariant under all of those transformations [31]. Before those transformations, the appropriate Foldy–Wouthuysen (FW) transformations of the anisotropic Hamiltonian are greatly extended to evaluate its expansion about the non-relativistic limit to order \((mc^2)^{-4}\) [32], where \(mc^2\) is the particle’s Einstein rest energy, in order to investigate the dimensionality of the Zeeman, spin–orbit, and quantum spin Hall interactions with great precision.

2. The anisotropic Dirac equation

The wavefunction \(\psi(r, t)\) for the particle in an orthorhombically anisotropic conduction band satisfies the Schrödinger
equation $\hat{H}\psi(r, t) = i\hbar \frac{\partial \psi(r, t)}{\partial t}$, based upon the modified Hamiltonian

$$\hat{H} = \hat{T} + \varepsilon,$$

(1)

where $\varepsilon(r, t) = q\Phi(r, t)$ and

$$\hat{T} = \sqrt{m^2 c^4} \sum_{\mu=1}^{3} \tilde{\Pi}_\mu^\alpha / m_\mu + m^2 c^4,$$

(2)

where $\Pi_\mu = p_\mu - qA_\mu(r, t)$, and $m_\mu$, $p_\mu$ and $A_\mu$ are the components of the particle’s effective mass, its quantum-mechanical momentum, and the magnetic vector potential, respectively, and we assume the $t$ dependencies of the scalar and vector potentials $\Phi(r, t)$ and $A(r, t)$ are slow with respect to differences in the particle’s energies divided by $\hbar$, so that they can be treated adiabatically. Gauge invariance of $\hat{T}$ is easiest to establish in the transformed representation, as was done for the anisotropic Ginzburg–Landau model [17, 28], and as shown in the following.

Dirac first showed that one can linearize the isotropic version of this Hamiltonian by the use of matrices based upon the Pauli matrices [3]. Here we generalize it to orthorhombic anisotropy, obtaining

$$\hat{H}\psi = \left[\hat{O} + \beta mc^2 + \varepsilon\right] \psi = i\hbar \frac{\partial \psi}{\partial t},$$

(3)

$$\hat{O} = c\alpha \cdot \Pi,$$

(4)

where $\Pi = p - qA$, $p \rightarrow -ih\nabla$, and

$$\tilde{\alpha}_\mu = \begin{bmatrix} 0 & \tilde{\sigma}_\mu \\ \tilde{\sigma}_\mu & 0 \end{bmatrix}, \tilde{\sigma}_\mu = \frac{\sigma_\mu}{\sqrt{m_\mu/m}},$$

(5)

for $\mu = 1, 2, 3$, the $\sigma_\mu$ are the Pauli matrices, and both the $\tilde{\alpha}_\mu$ and $\beta$ are rank-4 matrices, where 1 represents the rank-2 identity matrix [3, 33].

These four traceless matrices satisfy $\{\tilde{\alpha}_\mu, \tilde{\alpha}_\nu\} = 2\delta_{\mu\nu} m_\mu/m_\nu$, $\{\tilde{\alpha}_\mu, \beta\} = 0$, and the eigenvalues of the $\tilde{\alpha}_\mu$ and $\beta$ are respectively $\pm \sqrt{m_\mu/m_\nu}$ and $\pm 1$.

From equations (3)–(5), the $\mu$th component of the probability current is $j_\mu = \bar{\psi}\tilde{\alpha}_\mu \psi$, and since $\rho = \bar{\psi}\psi$, the continuity equation $\frac{\partial \rho}{\partial t} + \frac{\partial j_\mu}{\partial x_\mu} = 0$, is still satisfied with effective mass anisotropy.

2.1. Proof of covariance

Details of the proof of covariance of this anisotropic Dirac equation, equation (3), that it satisfies the most general proper Lorentz transformations (a rotation about all three orthogonal axes and a boost to a general special relativistic reference frame moving at a constant velocity in an arbitrary direction) and improper Lorentz transformations (reflection, time reversal, and charge conjugation), are posted in the supplementary material [31]. However, a much simpler proof is given here. We use a version of the KCT that were used to transform an orthorhombically anisotropic Ginzburg–Landau model of an anisotropic superconductor into isotropic form [17], which is exact at $B_2 = \mu_0 H$, where $H$ is the applied magnetic field. We make an anisotropic scale transformation of the spatial parts of equation (3),

$$\frac{\partial}{\partial x_\mu} = \sqrt{m_\mu/m} \frac{\partial}{\partial x_\mu'},$$

(6)

$$A_\mu = \sqrt{m_\mu/m} A_\mu',$$

(7)

which transforms it to

$$H'\psi = [c\alpha \cdot \Pi' + \beta mc^2 + q\Phi] \psi = i\hbar \frac{\partial \psi}{\partial t},$$

(8)

where $\Pi' = p'/qA'$, and

$$\alpha_\mu = \begin{bmatrix} 0 & \sigma_\mu \\ \sigma_\mu & 0 \end{bmatrix},$$

(9)

which has precisely the same form as the isotropic Dirac equation. Obviously, equation (8) is gauge-invariant [17, 28], covariant, and further details are provided elsewhere [31].

Clearly, this transformation preserves the Maxwell equation of no monopoles, $\nabla \times B' = 0$, provided that

$$B_\mu' = C \sqrt{m/m'} B_\mu',$$

(10)

for $\mu = 1, 2, 3$, where

$$C = (m/m')^{3/2},$$

(11)

where $m = (m_1 m_2 m_3)^{1/3}$ is the geometric mean effective mass. This preserves the required relation

$$B' = \nabla' \times A'.$$

(12)

When electric fields are present, we also require

$$E_\mu(r, t) = \sqrt{m_\mu/m} E_\mu'(r', t),$$

(13)

and

$$x_\mu' = \sqrt{m/m'} x_\mu.$$
They also dropped the third order Zeeman term.

We note that the fourth order term containing \[\hat{\Gamma}, \varepsilon\] vanishes, since \(\Phi(r, t)\) commutes with each component of \(E(r, t)\).

Although \(\beta\) is a rank-4 matrix, it has only two diagonal elements, so that the Hamiltonian for an electron or a hole in an anisotropic conduction band is obtained respectively \([32, 33]\) with \(\beta = 1, -1\) and \(q = [-\varepsilon], [+\varepsilon]\).

We evaluated \(H_{3D}\) both by using the above procedure with the effective masses and in the primed notation by first making the KCT. We note that \(\hat{\mathcal{O}} = \hat{\alpha} \cdot \hat{\Pi} = \alpha \cdot \hat{\Pi}'\), and that \(\hat{\Gamma} = i\hbar q \hat{\alpha} \cdot \hat{E}(r, t) = i\hbar q \hat{\alpha} \cdot \hat{E}'(r', t)\), where

We then may write

\[
H_{3D} = \beta \left( m^2 \frac{\Pi^2}{2m} - \mu_B \sigma \cdot \mathbf{B} + \frac{\mu_B^2}{2mc^2} \mathbf{E}^2 \right)
- \frac{1}{2m^2} \left( \frac{\Pi^2}{2m} - \mu_B \sigma \cdot \mathbf{B} \right)^2 - \frac{i\hbar^2 \hat{\alpha} \cdot \hat{\Pi} \partial \hat{\Pi}}{8mc^2 \partial t} - \frac{i\hbar^2 \hat{\alpha} \cdot \hat{\Pi} \partial \hat{\Pi} \left( \mathbf{E}^2 \right)}{8mc^2 \partial t} - \frac{\mu_B}{4mc^2} \left( \hbar \nabla' \cdot \mathbf{E} + \left( 2 \mathbf{E} \times \hat{\Pi}' + i\hbar \nabla' \times \mathbf{E}' \right) \cdot \sigma \right)
- \frac{\mu_B}{192mc^2} \nabla' \cdot \mathbf{E} + q \Phi(r, t) \right),
\]

where \(X_{3D}'\) is given elsewhere \([31]\).

Previously, Bjorken and Drell followed the procedure of Foldy and Wouthuysen to obtain \([32, 33]\),

\[
H_{8D} = \beta \left( m^2 \frac{\Pi^2}{2m} - \mu_B \sigma \cdot \mathbf{B} + \frac{\mu_B^2}{2mc^2} \mathbf{E}^2 \right)
- \frac{\mu_B}{4mc^2} \left( \hbar \nabla' \cdot \mathbf{E} + \left( 2 \mathbf{E} \times \hat{\Pi}' + i\hbar \nabla' \times \mathbf{E}' \right) \cdot \sigma \right)
+ q \Phi(r, t).
\]

They noted that Foldy and Wouthuysen omitted the term proportional to \(\nabla' \cdot \mathbf{E}\), and we note that Bjorken and Drell replaced the gauge-invariant operator \(\hat{\Pi}\) by the momentum \(p\), except in its leading (non-relativistic) term in equation \((17)\).

They also dropped the third order Zeeman term.

The spin-dependent terms in \(H_{3D}\) are

\[
\begin{align*}
H_{3D}' &= -\beta \mu_B \left( \sigma \cdot \mathbf{B} + \frac{1}{4m^2c^2} \right) + \frac{\mu_B}{192mc^2} X_{3D,(0,0)}', \\
H_{3D}'' &= \beta \frac{\mu_B}{4mc^2} \left( \frac{\Pi^2}{2m} - \mu_B \sigma \cdot \mathbf{B} \right),
\end{align*}
\]
where $\Pi_x = \Pi_1(x,t) = -i\hbar \alpha - qa_1(x,t)$ and the last line of order $(mc^2)^{-4}$ arises from the first term in $Z_{3D}$ shown elsewhere in this 1D limit [31]. Since the four-dimensional matrix $\beta$ in the 3D Dirac equation has eigenvalues $\pm 1$, which respectively correspond to electrons or holes, one could replace it with the eigenvalues of a Pauli matrix such as $\sigma_z$. Explicitly, none of the spin–orbit, quantum spin Hall, or Zeeman interactions exist in 1D, as sketched in figure 1(a).

This was checked to order $\beta$ using the orthorhombically anisotropic form of the FW transformations and taking the limits $m_x, m_y \to \infty$. In the case of the tight-binding model treated in many books and papers [36–39], $m_1$ is obtained from the inverse curvature of the 1D conduction band at the Fermi energy $E_F$, as sketched in figure 1(b). Of course, in a quasi-1D metal, $g_z$ is given by equation (22).

We could have linearized $H_{1D}$ using just two Pauli matrices $\sigma_1$ and $\sigma_3$,

$$H_{1D} \to c \sqrt{\frac{m}{m_1} \sigma_1 \Pi_1 + \sigma_3 mc^2 + q\Phi(x,t)}.$$

(24)

When squared for $\Phi = 0$, equation (24) leads to the 1D version of the Klein–Gordon equation, and the 1D FW transformations of it also leads to equation (23). Such a model has been extended to include particle–particle interactions, and studied by many workers [36–39].

For an electron or hole in a 2D metal with $m_1 \to \infty$, the anisotropic 2D Hamiltonian $H_{2D}$ with $m_1 \neq m_2$ to order $(mc^2)^{-4}$ is given elsewhere [31]. For the isotropic 2D case $m_1 = m_2 = m_1$, it reduces to

$$H_{2D} = \beta \left( mc^2 + \frac{\Pi_1^2}{2m_1} - \frac{\beta_B}{\sigma_1 B_\perp} + \frac{\beta_B}{8m_1 c^2} E^2 \right) - \frac{1}{2mc^2} \left( \frac{\Pi_1^2}{2m_1} - \frac{\beta_B}{\sigma_1 B_\perp} \right)^2 - \frac{i\hbar \mu_B^2}{8m_1 c^2} \frac{\partial}{\partial t} E^2 + \frac{\beta_B X_{2D}}{192m_1 c^2} + q\Phi(x,t) + O(mc^{-2}).$$

(25)

where $\Pi_1 = \Pi_{11}(x,y,t) \hat{x} + \Pi_{12}(x,y,t) \hat{y}$ is defined following equation (4), $E_1 = E_1(x,y,t) \hat{x} + E_2(x,y,t) \hat{y}$, $\mu_B = \mu_0/(2m)$ is the planar Bohr magneton, and $X_{2D}$ is given elsewhere [31]. We note that the 2D Zeeman, spin–orbit, quantum spin Hall, and mixed terms may be written as

$$H_{2D}^Z = -\beta \mu_B \frac{E}{mc} \left( \sigma_1 B_\perp - \frac{\{\Pi_1^2, \sigma B_\perp\}}{4m_1 c^2} \right) + \mu_B \sum_{i=1}^3 X_{2D(i)} \frac{1}{192m_1 c^2},$$

(26)

$$H_{2D}^{SO} = -\frac{\beta \mu_B}{2mc^2} \frac{h(\nabla \times E) \sigma_1}{\sigma_1} + \frac{\mu_B}{192m_1 c^2} \sum_{i=1}^5 X_{2D(i)} \frac{1}{192m_1 c^2},$$

(27)

$$H_{2D}^{QSH} = -\frac{\beta \mu_B}{192m_1 c^2} \frac{E \cdot \sigma_1}{\sigma_1} + \mu_B \frac{X_{2D(i)}}{192m_1 c^2},$$

(28)

and $X_{2D(i)}$ are given elsewhere [31].

Some examples of the highly anisotropic nature of the 2D Zeeman interaction are sketched in figures 2 and 3. In figure 2(a), $B || z$ is normal to the conduction plane, and an electron or hole experiences a normal Zeeman interaction only modified by its planar effective mass $m_1$. This splits the conduction and hole energy bands, the former of which is sketched in figure 2(b). However, for $B \perp z$, as sketched in figure 2(c), the 2D motion of an electron or hole does not allow for it to have a Zeeman interaction with that $B$, so its conduction and hole energy bands are unsplit, the former of which is sketched in figure 2(d).

In figure 3(a), the Fermi energy $E_F$ is added to the sketch shown in figure 2(b) for $B || z$. The first Brillouin zone in reciprocal space of the spin–split bands for a square 2D lattice is sketched for electrons in figure 3(b) and for holes in figure 3(c). In figure 3(d), spin-split bands with $B || z$ are shown for a square 2D conductor with both electron and hole bands.

The leading contribution to the 2D Zeeman interaction, $-\beta \mu_B \sigma_1 B_\perp$, is equivalent for electrons and holes, and vanishes for $B$ parallel to the infinitesimally thin 2D metallic film. However, finite thicknesses of real 2D materials could result in $0 < g_1 < 2$, as from equation (22). Twisted-bilayer graphene and $\kappa(ET)_2Cu(SCN)_2$ are in the clean limit and $B_{1,2,3}(T)$ is
in those materials is consistent with the clean limit \( g_s < 2 \) [6, 40], as sketched by the solid blue curve in figure 4. This may also be the case in a large number of other clean 2D superconductors, including monolayer FeSe, monolayer NbSe_2, gated MoS_2, and possibly few-layer stanene [4, 5, 41–44]. The leading terms for spin–orbit coupling and the quantum spin Hall interactions have coefficients that differ only by a factor of 2 and when one replaces \( i\hbar \nabla \) by \(-\mathbf{p}\), the sign. These coefficients are not arbitrary, as \( m_1 \) can be measured by cyclotron resonance experiments.

As for \( H_{1D} \), we could rewrite \( H_{2D} \) as

\[
H_{2D} \rightarrow e \sum_{\mu \nu} \sqrt{\frac{m}{m_\|}} \sigma_{\mu \nu} \Pi_{\mu \nu} + \alpha_3 mc^2 + q\Phi(x, y, t),
\]

(30)

which can also be completely described in terms of rank-2 Pauli matrices, which satisfy the requirements that \( \{\sigma_i, \sigma_j\} = 2\delta_{ij} \) and that they all be traceless and have eigenvalues \( \lambda = \pm 1 \) [33], and use the 2D FW transformations to obtain equation (25). Thus, this simplified Hamiltonian is sufficient to generate the Zeeman, spin–orbit, and quantum spin Hall interactions in 2D, as described in detail in equations (26)–(29).

An electron or hole in an isotropic 3D conduction band with \( m_1 = m_2 = m_3 = m_e \) satisfies the Schrödinger equation \( H_{3D}\mathbf{\psi} = \hat{H}(\mathbf{\psi})/\hbar \), given by equation (16) with \( E' = \sqrt{m/m_e} E, \mathbf{I}' = \sqrt{m/m_e} \mathbf{I}, \) and \( \mathbf{B}' \rightarrow (m/m_e)\mathbf{B} \).

4. Discussion

Microscopically, the nucleus of an atom moves slowly inside a 3D electronic shell, and as for the Dirac equation of a free electron, the 3D relativistic motion of each of its neutrons and protons leads to it having an overall spin I and a nuclear Zeeman energy that can be probed by a time t-dependent external \( \mathbf{B}(t) \) in nuclear magnetic resonance (NMR) and in Knight shift measurements when in a metal [45, 46]. The orbital electrons bound to that nucleus also move in a nearly isotropic 3D environment, and have a much larger Zeeman interaction with \( \mathbf{B}(t) \), modified only by the \( V(r) = -|e|\Phi(r) \) of nearby atoms.

However, when an atomic electron or hole is excited into a crystalline conduction or valence band, respectively, it leaves that atomic site and moves with wave vector \( \mathbf{k} \) across the crystal. Its motion depends upon the crystal structure, and can be highly anisotropic. In an isotropic, 3D metal, \( E(k) = \hbar^2 k^2/(2m) \) for free electrons. These states are filled at \( T = 0 \) up to the Fermi energy \( E_F \) and \( H_{3D}^{\text{I}} = -\mu_3 \mathbf{g} \cdot \mathbf{B} \). But in Si and Ge [14], the lowest energy conduction bands can be expressed as \( E(k) = \hbar^2 \sum_{\mathbf{k}} (|k_\perp - k_0\parallel|^2/2m) \) about some minimal point \( \mathbf{k}_0 \), and the \( m_\parallel \) differ significantly from \( m \).

In a purely 1D metal, the conduction electrons move rapidly along the chain of atomic sites, as sketched in figure 1(a), usually with a tight-binding 1D band \( E(k) \) as sketched in figure 1(b), and \( H_{1D}^{\text{I}} = 0 \). When an electron is in a quasi-1D superconductor such as (TMTCF)_2PF_6 [27], \( E(k) \) is highly anisotropic, the transport normal to the conducting chains is by weak hopping, so the effective masses in those directions greatly exceed \( m \), and \( H_{1D}^{\text{I}} \) should decrease very substantially in magnitude from the isotropic 3D value.

Similarly, in 2D metals, such as monolayer or gated NbSe_2, MoS_2, WTe_2 [5, 41, 47, 48], and twisted bilayer graphene [6], the effective mass normal to the conducting plane is effectively infinite. As sketched in figure 2, the direction of \( \mathbf{B} \) is very important. When \( \mathbf{B} \) is normal to that plane, as in figure 2(a), the spins of the conduction electrons eventually align either parallel or anti-parallel to \( \mathbf{B} \), giving rise to a Zeeman interaction that can differ from that of a free electron only by the effective mass \( m_\parallel \). There are two energy dispersions \( E(k) \) for up and down spin conduction electrons, as sketched in figure 2(b).
In an infinitesimally thin conducting layer, the Zeeman Hamiltonian to \( O(m c^2)^{-1} \) is given by equation (26). When \( B \) lies within the 2D conduction plane, as sketched in figure 2(c), the Zeeman interactions vanish, so their spin states are effectively random, and there is only one conduction band, as sketched in figure 2(d).

Furthermore, the dimensionality of the particle’s motion on the Zeeman interaction can greatly affect the upper critical field in highly anisotropic superconductors. Figure 4 is a sketch of the generic behavior expected for the upper critical induction \( B_{c2}(T) \) for \( B = \mu_0 H \) applied parallel and perpendicular to a 2D film. The red dashed horizontal line is the conventional Pauli limiting induction \( B_p = 1.86 T_c \) T/K for an isotropic superconductor with \( m_\delta = m \) and \( g = 2 \), a situation that is often not realized in real materials, especially unconventional ones such as heavy fermion materials. However, for the clean layered superconductors twisted bilayer graphene and \( \kappa-(ET)_2\text{Cu(SCN)}_2 \), while there are strong Pauli limiting effects, \( B_{c2}(T) \) generically follows the Tinkham thin film formula \( B_{c2}(T) = \mu_0 \sqrt{3} \phi_0 / (\pi s \xi(T)) \) [6, 17, 40, 49], plotted as the solid blue curve in figure 4, where \( s \) is the film thickness, \( \phi_0 = h/(2e) \) is the superconducting flux quantum, \( \mu_0 \) is the magnetic permeability of vacuum, and \( \xi(T) \) is the Ginzburg–Landau coherence length parallel to the film. However, in both twisted bilayer graphene and \( \kappa-(ET)_2\text{Cu(SCN)}_2 \), the value of \( s \) required to fit the data is about a factor of 30–35 larger than the actual layer thickness [6, 40], so a new theory of clean 2D and layered superconductors that does not depend upon the layer thickness \( s \) is greatly needed.

In the case of the dirty limit with \( g < 2 \), sketched in figure 4 by the dashed blue curve, \( B_{c2}(T) \) exceeds the conventional Pauli limit \( B_p \) at low \( T \), usually by spin–orbit scattering, but flattens out at low \( T \) values, with zero slope at \( T = 0 \). This is typical of many dirty 2D superconductors and highly anisotropic layered superconductors [17, 30]. In addition, the solid black curve for \( B_{c2}(T) \) with \( g > 2 \), is also predicted by our theory, but has not been appreciated at all in the literature: for \( m_\delta < m \), the effective Pauli limiting is stronger than in the conventional theory with \( g = 2 \). We emphasize that there can be a huge difference between monolayer and few-layer superconductors: the recent \( B_{c2}(T) \) results for monolayer NbSe\(_2\) greatly exceed those results for bilayer and trilayer NbSe\(_2\) [5]. This is clearly a dimensionality effect, and \( g \) for the monolayer is most likely substantially less than for the bilayer and trilayer with this \( B \) direction.

In many models [30, 50, 51], the Zeeman interaction was assumed to be that of a free electron moving isotropically in three spatial dimensions (3D). The Fulde–Ferrell–Larkin–Ovchinnikov (FFLO) state was predicted to occur in clean superconductors with a spatially varying order parameter at low \( T \) and high \( B \) values exceeding the ‘conventional’ Pauli limit \( B_p \), but has never been observed in an isotropic superconductor [17]. Although it was claimed to have been observed in heavy fermion superconductors such as CeCoIn\(_5\) [52], that second phase has since been attributed to spin–density wave formation [40, 53]. Since then, it has only been claimed to have been observed in clean layered superconductors for \( B \) parallel to the layers [40, 54].

The standard KLB theory of the upper critical field in layered superconductors was designed to explain the violation of that conventional Pauli limit of \( B_{c2}(T) \) in dirty layered superconductors in the limit of strong spin–orbit scattering, while assuming that \( g \approx 2 \) for all \( B \) directions [17, 30]. In that theory, \( B_{c2}(T) \) was obtained in the standard pair-breaking theory,

\[
\ln(T/T_c) + \frac{1}{2} + \frac{\alpha_{ph}}{2\pi k_B T} - \psi(1/2) = 0, \tag{31}
\]

where \( \psi(z) \) is the digamma function, \( I = g\mu_B B_{c2}(T)/2 \) and the energy eigenvalue of the relevant Matthieu equation, \( \varepsilon \rightarrow I^2/2(2h) \), in the high \( B_{c2} \) limit, where \( J \) is the interlayer electron hopping energy, \( \tau \) and \( \tau_{so} \) are the total and spin–orbit intralayer scattering times, and \( g \approx 2 \) was assumed without having derived it for that field direction [30]. However, for monolayer superconductors, or layered superconductors consisting of superconducting monolayers or bilayers well separated by insulating layers, \( g \) depends strongly upon the direction of \( B \), and vanishes when \( B \) is parallel to infinitely thin layers! Setting \( g \rightarrow 0 \) in equation (31) leads to an unbounded \( B_{c2}(T) \) below the dimensional crossover temperature \( T^* \) [17, 30]! Of course, in the KLB theory, the dirty layers with strong spin–orbit scattering were assumed to be infinitely thin. In the Tinkham thin film formula described above and used to sketch the top curve in figure 4, the thickness \( s \) of the single layer is the relevant parameter that limits the value of \( B_{c2}(T) \). Hence, the KLB theory should be reformulated to treat the clean limit of each infinitely thin superconducting layer, and to set \( g < 2 \) as a ‘fitting’ parameter in calculating \( B_{c2}(T) \). No present software routines have been developed to calculate the anisotropy of \( g \) in films a single atom thick, but it is imperative that this be done. Such a calculation could provide a quantitative fit to \( B_{c2}(T) \) in superconducting twisted bilayer graphene and in \( \kappa-(ET)_2\text{Cu(SCN)}_2 \) [6, 40], and most likely in gated MoS\(_2\), WTe\(_2\), few-layer stanene, and mono- and few-layer NbSe\(_2\) [5, 43].

Although there is as yet no experimental evidence of superconductivity in 1D systems, there are a number of quasi-1D superconductors, such as (SN)\(_4\) and (TMTSF)\(_2\)PF\(_6\) [27, 37–39]. But it is interesting to consider the possibility of producing a very narrow superconducting filament, such as a carbon nanotube covered by a monolayer of Nb. Since somewhat larger versions of this have been routinely made as nano-SQUIDs for magnetic field detectors at the nm scale [55], it might be possible to make a single-walled carbon nanotube coated with a monolayer of Pb or Nb. Then, it would be interesting to measure \( B_{c2}(T) \) for such a quasi-1D superconductor. Superconductivity in zero applied magnetic field was also reported in carbon nanotubes coated with a Mo/Ge alloy [56].

In addition to the upper critical field, the anisotropic Zeeman interaction can influence the interpretations of Knight shift measurements. The Knight shift is the relative change in the NMR resonance frequency for a nuclear spin when it is in a metal (or superconductor) from when it is in an insulator or vacuum. In both cases, the nuclear spin of an atom
interacts with that of one of its orbital electrons via the hyperfine interaction. But when that atom is in a metal, the orbital electron can sometimes be excited into the conduction band, traveling throughout the crystal, and then returning to the same nuclear site, producing the leading order contribution to the Knight shift \cite{45,46}. The dimensionality of the motion of the electron in the conduction band is therefore crucial in interpreting Knight shift measurements of anisotropic materials, as first noticed in the anisotropic three-dimensional superconductor, YBa$_2$Cu$_3$O$_{7-\delta}$ \cite{20}, for which anomalous, $T$-independent behavior of the planar $^63$Cu $K(T)$ was observed for $B_0 \parallel \hat{e}_c$, and $B_1(t)$ lying in the most conducting $ab$ plane. However, for $B_0 \perp \hat{e}_c$, a conventional change in $K(T)$ through $T_c$ was seen, both for the planar and chain $^63$Cu spins. It is evident that the Yosida theory of an isotropic, single band superconductor cannot easily explain such data \cite{21}. Similarly, equation \eqref{22} implies that $K(T)$ for the quasi-one-dimensional superconductor (TMTSF)$_2$PF$_6$ should be nearly constant through the superconducting transition, as observed \cite{27}. A recent $K(T)$ measurement on Sr$_2$RuO$_4$ under uniaxial planar pressure did show a substantial $K(T)$ variation below $T_c$ for $B_0 \parallel \hat{e}_c$ \cite{24}, as for YBCO \cite{20}, in agreement with scanning tunneling microscopy results of a nodeless, nearly isotropic superconducting gap in Sr$_2$RuO$_4$, which has two electron-like Fermi surfaces and one hole Fermi surface \cite{26}.

Since the magnetic moments of the neutron and a trapped proton are now known to 8 and 10 significant figures \cite{1,2}, respectively, similar experiments should be performed on a trapped electron. Since its spin only arises from the relativistic motion of its mass and charge, and $g = 2$ only arises from that 3D motion in vacuum, we predict that if an electron can be trapped as for the double-Penning traps used to measure the magnetic moment of a nearly stationary proton \cite{1}, its $g$ value should be vanishingly small. We also strongly encourage magnetic moment measurements both normal and parallel to the conducting single atom chains of Au, Ag, Cu, or Al \cite{9–11}, for which $g$ is predicted to vanish for both magnetic field directions. Similarly, we strongly encourage magnetic moment measurements for fields parallel to monoatomic conducting films.

5. Summary

The Dirac equation is extended to treat a relativistic charge $q$ moving in an orthorhombically anisotropic conductor. The norm for this model with metric $\tilde{g}$ is invariant under the most general proper Lorentz transformation $\tilde{\alpha}$, the matrix representation of which exhibits $O(3,1)$ group symmetry, and this anisotropic Dirac equation is demonstrated to be covariant, precisely as for the isotropic Dirac equation. This model applies to large classes of anisotropic semiconductors, metals, and superconductors. Although overlooked by many workers, the $A$ in $\Pi$ plays an important role in the quantum spin Hall Hamiltonian, which is distinctly different from that of spin–orbit coupling in topological insulators, and a proposed experiment to test this result will be published elsewhere.

This model has profound consequences for Pauli limiting effects upon $B_{\perp z}$ for $B$ parallel to the low mass direction(s) of clean, highly anisotropic superconductors, and for the temperature dependence of Knight shift measurements in highly anisotropic superconductors. We encourage measurements at higher fields and lower $T$ values to confirm our prediction that $B_{\perp z}(0)$ could greatly exceed the standard Pauli limit in clean monolayer and bilayer superconductors, such as gated and pure transition metal dichalcogenides \cite{5,41}. In monolayer NbSe$_2$, the $g$ factor for $B$ parallel to the layers appears to be less than 0.3 \cite{5}. For bilayer and trilayer NbSe$_2$, the analogous $g$-factor appears to be about 0.5–0.7 \cite{5}. $g$ for $B$ parallel to the twisted bilayers of graphene and to the layers of $\kappa-(ET)_2$Cu(NCS)$_2$ appear to be on the order of 1.5–1.7 \cite{6,40,54}. Since the low-$T$ regime of $\kappa-(ET)_2$Cu(NCS)$_2$ was claimed to be in an FFLO state, but $B_{\perp z}(T)$ is very similar to that of twisted bilayer graphene, the data suggest that it may be the intrinsic $B_{\perp z}(T)$ behavior for a clean 2D superconductor.

Finally, the general quantum spin Hall Hamiltonian is given by equation \eqref{20} for this model, and it is equation \eqref{28} in a 2D metal. We emphasize that in many papers on the quantum spin Hall effect, the $A$ term was omitted, and therefore the full ramifications of the quantum spin Hall effect in conductors have not yet been observed \cite{35}. We further note that the coefficient $\mu_B/(2m_1c^2)$ is not a free parameter, as $\mu_B = q\hbar/(2m)$ is the Bohr magneton for a hole and minus the Bohr magneton for an electron, and $m_1$ is the particle’s planar effective mass that is measurable for any 2D metal by cyclotron resonance with $B$ normal to the conducting plane.

Acknowledgments

The authors acknowledge helpful discussions with Luca Argenti, Thomas Bullard, Kazuo Kadowaki, Anthony J Leggett, Qiang Li, and Jingchuan Zhang. This work was supported by the National Natural Science Foundation of China through Grant no. 11874083. AZ was also supported by the China Scholarship Council. RAK was partially supported by the U.S. Air Force Office of Scientific Research (AFOSR) LRIR #18RQCOR100, and the AFRL/SFFP Summer Faculty Program provided by AFRL/RQ at WPAFB.

ORCID IDs

Qiang Gu https://orcid.org/0000-0002-9325-0871
Richard A Klemm https://orcid.org/0000-0001-6827-0187

References

[1] Schneider G et al 2017 Science 358 1081–4
[2] Beringer J (Particle Data Group) 2012 Phys. Rev. D 86 010001
[3] Dirac P A M 1928 Proc. R. Soc. A 117 610–24
Dirac P A M 1928 Proc. R. Soc. A 118 351–61
[4] He S et al 2013 Nat. Mater. 12 605–10
[5] Xi X, Wang Z, Zhao W, Park J-H, Law K T, Berger H, Forró L, Shan J and Mak K F 2016 Nat. Phys. 12 139–43
[6] Cao Y, Fatemi V, Fang S, Watanabe K, Taniguchi T, Kaxiras E and Jarillo-Herrero P 2018 Nature 556 43–50
