IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **20** (2008) 193202 (17pp) doi:10.1088/0953-8984/20/19/193202

TOPICAL REVIEW

Berry curvature, orbital moment, and effective quantum theory of electrons in electromagnetic fields

Ming-Che Chang[1] **and Qian Niu**[2]

[1] Department of Physics, National Taiwan Normal University, Taipei, Taiwan
[2] Department of Physics, University of Texas at Austin, Austin, TX, USA

E-mail: changmc@phy.ntnu.edu.tw and niu@physics.utexas.edu

Received 9 January 2008, in final form 20 March 2008
Published 11 April 2008
Online at stacks.iop.org/JPhysCM/20/193202

Abstract

Berry curvature and orbital moment of the Bloch state are two basic ingredients, in addition to the band energy, that must be included in the formulation of semiclassical dynamics of electrons in crystals, in order to give proper account of thermodynamic and transport properties to first order in the electromagnetic field. These quantities are gauge invariant and have direct physical significance as demonstrated by numerous applications in recent years. Generalization to the case of degenerate bands has also been achieved recently, with important applications in spin-dependent transport. The reader is assured that a knowledge of these ingredients of the semiclassical dynamics is also sufficient for the construction of an effective quantum theory, valid to the same order of the field, using a new quantization procedure that generalizes the venerable Peierls substitution rule. We cite the relativistic Dirac electron and the carrier in semiconductors as two prime examples to demonstrate our theory and compare with previous work on such systems. We also establish general relations between different levels of effective theories in a hierarchy.

(Some figures in this article are in colour only in the electronic version)

Contents

1. Introduction 2
2. Semiclassical theory 3
   2.1. Bloch band and Berry curvature 3
   2.2. Self-rotating angular momentum of a wavepacket 3
   2.3. Wavepacket energy in electromagnetic fields 3
   2.4. Wavepacket dynamics: Lagrangian and equations of motion 4
3. Semiclassical theory for a relativistic electron 5
   3.1. Basics of the Dirac theory 5
   3.2. Basics of the Dirac wavepacket 5
   3.3. Dynamics of the Dirac wavepacket 6
4. Canonical quantization 6
   4.1. Deriving canonical variables 7
   4.2. Effective quantum Hamiltonian 8
   4.3. Relativistic Pauli Hamiltonian 8
   4.4. Discussion 8
5. Hierarchy of effective theories 9
   5.1. Gauge-invariant ingredients of an effective theory 9
   5.2. Position and velocity 10
   5.3. Dirac theory as an effective theory 11
6. Applications to semiconductor physics 12
   6.1. Berry connection in the Kane model 12
   6.2. Berry curvature, orbital moment, and equations of motion 13
   6.3. Effective Hamiltonians 14
7. Conclusions 14
Acknowledgments 15
Appendix A. The Pauli wavepacket 15
Appendix B. Spin precession for a particle with AMM 15
Appendix C. Calculation of the Berry connection 16
Appendix D. Berry connection in the Luttinger model 16
References 16

1. Introduction

Berry's phase is a unifying concept in physics [1, 2] and the semiclassical dynamics with Berry curvature included has found wide applications in solid state physics. The Berry curvature appeared quite early in history in the one-band effective dynamics of a Bloch electron [3, 4]. It generates an anomalous velocity that is transverse to the applied electric field. When applied to the quantum Hall system, the semiclassical theory can explain the Hall current and the quantization of the Hall conductivity [5–7]. In recent years, it has helped in solving the mystery of the anomalous Hall effect in ferromagnetic materials [8–10]. It is also a primary notion motivating the proposal of the intrinsic spin Hall effect [11, 12] and the optical Hall effect [13, 14].

Despite various attempts, the effective dynamics remains largely semiclassical, and a systematic quantum version is still lacking. Such a quantum theory would be needed if one wants to know, for example, the full intricate structure of a band under a magnetic field (the Hofstadter spectrum) [15]. Here we report our recent finding of a quantization procedure by generalizing the Peierls substitution rule [16].

For an effective quantum theory without the Berry phase, it is a common practice to use the Peierls substitution [17, 18]. This is clearly inadequate when the band has a Berry curvature. When this fails, one typically blames the neglect of interband couplings, and expands the range of the Hilbert space by including more bands. One message that we would like to convey is that there is sufficient information in the semiclassical formalism for constructing the correct quantum theory. The Berry curvature and orbital moment in the one-band semiclassical formulation contain all the information of interband coupling that is needed for an effective quantum theory for the band (or a subset of bands) under consideration. This finding may have an enormous impact on research based on first-principle band structure calculations. One is now assured that, by calculating the three essential physical quantities, the Bloch energy $E(\mathbf{q})$, the Berry curvature $\mathbf{F}(\mathbf{q})$, and the self-rotating angular momentum of the wavepacket $\mathbf{L}(\mathbf{q})$, all electronic properties to first order in external fields can be obtained.

On the other hand, in the 1950s and 1960s, systematic studies were taken to develop effective Hamiltonians for a Bloch electron in an electromagnetic field [19–21]. They are exact up to a certain order in the external fields. It is not immediately obvious how the Berry curvature emerges from such an effective quantum theory. The effective Hamiltonian involves quantities that are gauge dependent (depending on the phase choice of the basis Bloch functions), and the position operator conjugate to the momentum is not really the physical position operator. However, miraculously, when one considers the Ehrenfest equations of motion for the physical position and crystal momentum, one obtains the semiclassical theory with the gauge-invariant Berry curvature and the orbital moment properly included.

In this review, we start with the gauge-invariant semiclassical theory and quantize it using the generalized Peierls substitution [16]. Naturally, our effective Hamiltonians are in agreement with earlier results obtained by canonical transformation [22, 23]. Although this route toward an effective quantum theory is less direct compared with earlier works, our approach is physically more transparent. The wavepacket point of view is certainly intuitive, and the gauge-invariant Berry curvature and orbital moment in the semiclassical theory are of direct physical significance. Furthermore, our method of quantization can also be applied to more general situations, such as the crystal deformation described in [24].

We do not attempt to give a comprehensive review on the topics related to the Berry phase or semiclassical dynamics in solids. They can be found in the following reviews in [25–28]. Instead, we focus on the semiclassical transport in solid that involves spin (or quasi-spin) degrees of freedom. The development is mainly based on the multi-band generalization of the semiclassical dynamics [29], which will be reviewed in section 2. This formalism applies to carriers with spins and has a richer dynamics than the one-band theory.

Throughout this review, we will use the Dirac theory of relativistic electron to illustrate the formalism. Dirac theory is one of the simplest that admits a non-Abelian gauge structure. In section 3, the three basic quantities mentioned above are calculated exactly for the Dirac electron. They are used to obtain the equations of motion and the equation for spin precession.

Section 4 is devoted to the quantization of the semiclassical theory. The Landau levels in a semiconductor are not usually obtained by solving the Schrödinger equation treating the lattice potential and the vector potential on an equal footing. Such an approach is difficult and one ends up with a complex Hofstadter spectrum [15]. Instead, one first solves the band energy, then uses the Peierls substitution to obtain the effective quantum Hamiltonian. The Landau levels are the eigenenergies of this effective Hamiltonian [17, 18]. In section 4, we derive the generalized Peierls substitution and the effective Hamiltonians for systems with Berry curvature [16]. The Dirac theory is again used as an example.

Compared to the original quantum theory, the re-quantized version is an effective one applying to a smaller Hilbert space (of only one energy band, for example). The re-quantized theory can have its own semiclassical approximation. One can in turn re-quantize this semiclassical theory, and such a process can go on and on until the resulting theory is trivial. This way, one can generate a whole series of hierarchy theories with different levels of approximation. In the extreme case, in a system with self-similarity in energy scales, such as the Hofstadter spectrum, such a series is essentially infinite. The general relations between the theories with different levels of approximation will be explored in section 5.

In section 6, equipped with the tools developed and confirmed in preceding sections, we turn our attention to the carrier dynamics in semiconductors within the framework of the eight-band Kane model. This investigation offers fresh perspectives on familiar concept in semiconductors, such as the origin of the anomalous g-factor and the relation between Berry connection and spin–orbit interaction. The review is summarized and concluded in section 7.

2. Semiclassical theory

In this section, we introduce the basic physical quantities essential to the semiclassical theory. These include the Bloch energy, the Berry curvature, and the orbital moment. We will then briefly summarize the multi-band wavepacket formulation of Culcer *et al* [29, 30]. Such a formulation serves as a basic tool in this review. This summary is intended to be self-contained. However, readers not familiar with the wavepacket formulation are advised to consult earlier literature on the simpler one-band theory first [7, 24]. In the one-band version, the wavefunction has only one component and the gauge structure is Abelian; in the multi-band version, the wavefunction is a multi-component spinor and the gauge structure becomes non-Abelian.

2.1. Bloch band and Berry curvature

Consider an electron in a periodic crystal with the following Schrödinger equation:

$$H_0(\mathbf{r},\mathbf{p})\psi_{n\mathbf{q}} = E_{0,n}(\mathbf{q})\psi_{n\mathbf{q}}, \tag{1}$$

where n and $\mathbf{q}$ are band index and crystal momentum, and $E_{0,n}(\mathbf{q})$ and $\psi_{n\mathbf{q}}$ are the Bloch energies and Bloch states. The Bloch states can always be written in the form

$$\psi_{n\mathbf{q}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}u_{n\mathbf{q}}(\mathbf{r}), \tag{2}$$

where $u_{n\mathbf{q}}(\mathbf{r})$ is a cell-periodic function. It is useful to define $\tilde{H} = \mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}H\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}$ and write the Schrödinger equation in an alternative form,

$$\tilde{H}_0(\mathbf{r},\mathbf{p};\mathbf{q})u_{n\mathbf{q}} = E_{0,n}(\mathbf{q})u_{n\mathbf{q}}. \tag{3}$$

In the semiclassical formulation, one considers a wavepacket moving within a subset $\{E_{0,n}(\mathbf{q}), n = 1\cdots D\}$ of the full spectrum. Such a subset is degenerate in energy and is gapped from the rest of the energy bands. Within such a subspace, one can define two geometric quantities: a metric that defines the quantum distance between two states [31], which is not a concern of this review, and a *connection* defined as

$$\mathbf{R}_{mn}(\mathbf{q}) = \mathrm{i}\left\langle u_{m\mathbf{q}}\left|\frac{\partial u_{n\mathbf{q}}}{\partial\mathbf{q}}\right\rangle\right. . \tag{4}$$

This is the non-Abelian Berry connection, which is a vector-valued $D\times D$ matrix [32].

Following the rule of parallel transport, a state circling a closed loop C in the Brillouin zone may acquire a unitary (Berry) rotation. The Berry rotation for an infinitesimal loop surrounding a point $\mathbf{q}$ is proportional to the curvature $\mathcal{F}$ at this point,

$$\mathcal{F}(\mathbf{q}) = \nabla_{\mathbf{q}}\times\mathcal{R} - \mathrm{i}\mathcal{R}\times\mathcal{R}. \tag{5}$$

Throughout this review, symbols in calligraphic fonts represent matrices (boldfaced calligraphic fonts are vector-valued matrices). The Berry curvature matrix is gauge covariant under a gauge rotation.

If the subset comprises only of one band with $D = 1$, then $\mathcal{R}$ and $\mathcal{F}$ become vectors $\mathbf{R}$ and $\mathbf{F}$, and the Berry rotation is simply a $U(1)$ Berry phase, which is invariant under a gauge transformation. One can prove that, if the system has time-reversal symmetry, then $\mathbf{F}(-\mathbf{q}) = -\mathbf{F}(\mathbf{q})$. Also, if there is space-inversion symmetry, then $\mathbf{F}(-\mathbf{q}) = \mathbf{F}(\mathbf{q})$. Therefore, when both symmetries exist, there can be no Berry curvature. However, in a multi-band theory, the Berry curvature can exist even if both symmetries exist.

2.2. Self-rotating angular momentum of a wavepacket

One can construct a wavepacket by linear superposition of the Bloch states. In order for the wavepacket to remain intact with a meaningful center of mass, we require it to reside on a set of bands closely packed in energy so that it will not split into separate packets. Before turning on an external field, the unperturbed Hamiltonian $H_0(\mathbf{r},\mathbf{p})$ is assumed to be solvable with degenerate Bloch states $|\psi_{n,\mathbf{q}}\rangle$, $n = 1\cdots D$. If necessary, a small degeneracy-breaking interaction can also be included. But it has to be treated as a perturbation[3].

The wavepacket is expanded as

$$|W\rangle = \sum_{n=1}^{D}\int \mathrm{d}^3\, q a(\mathbf{q},t)\eta_n(\mathbf{q},t)|\psi_{n,\mathbf{q}}\rangle, \tag{6}$$

where $\eta_n(\mathbf{q},t)$ is normalized as a unit vector at each $\mathbf{q}$, $a(\mathbf{q},t)$ is a narrow distribution that centers at $\mathbf{q}_{\mathrm{c}}(t)$ in momentum space and is normalized as $\int \mathrm{d}^3\, q|a|^2 = 1$. The wavepacket is also centered at $\mathbf{r}_{\mathrm{c}}$ in the real space. Therefore, one requires $\langle W|\mathbf{r}|W\rangle = \mathbf{r}_{\mathrm{c}}$. Obviously, one also has $\langle W|H_0|W\rangle = E_0(\mathbf{q}_{\mathrm{c}})$.

The wavepacket is often found to be spinning with respect to its own center of mass. The self-rotating angular momentum of a wavepacket is $\mathbf{L} = \langle W|(\mathbf{r}-\mathbf{r}_{\mathrm{c}})\times\mathbf{p}|W\rangle$. It is difficult to calculate $\mathbf{L}$ under the present form since the exact shape of the wavepacket is not specified. Fortunately, for a wavepacket narrowly distributed around $\mathbf{q}_{\mathrm{c}}$, it is possible to rewrite it in another form using the Bloch states only [7],

$$\mathbf{L}(\mathbf{q}_{\mathrm{c}}) = \mathrm{i}\frac{m}{\hbar}\left\langle\frac{\partial u}{\partial\mathbf{q}_{\mathrm{c}}}\right|\times\left[\tilde{H}_0 - E_0(\mathbf{q}_{\mathrm{c}})\right]\left|\frac{\partial u}{\partial\mathbf{q}_{\mathrm{c}}}\right\rangle, \tag{7}$$

where the cell-periodic function without a subscript is defined as $|u\rangle = \sum_{n=1}^{D}\eta_n|u_n\rangle$. In the discussions below, we will also encounter an angular momentum matrix with the following matrix elements:

$$\mathbf{L}_{nl}(\mathbf{q}_{\mathrm{c}}) = \mathrm{i}\frac{m}{\hbar}\left\langle\frac{\partial u_n}{\partial\mathbf{q}_{\mathrm{c}}}\right|\times\left[\tilde{H}_0 - E_0(\mathbf{q}_{\mathrm{c}})\right]\left|\frac{\partial u_l}{\partial\mathbf{q}_{\mathrm{c}}}\right\rangle. \tag{8}$$

Obviously, after spinor average, one obtains the angular momentum in equation (7): $\mathbf{L} = \boldsymbol{\eta}^{\dagger}\boldsymbol{\mathcal{L}}\boldsymbol{\eta} = \sum_{nl}\eta_n^*\mathbf{L}_{nl}\eta_l$.

2.3. Wavepacket energy in electromagnetic fields

An external electromagnetic field can be added to H_0 using a scalar potential $\phi(\mathbf{r})$ and a vector potential $\mathbf{A}(\mathbf{r})$, giving the full Hamiltonian H. The wavepacket energy $E(\mathbf{r}_{\mathrm{c}},\mathbf{q}_{\mathrm{c}}) = \langle W|H|W\rangle$ differs from $E_0(\mathbf{q}_{\mathrm{c}})$ as a result of these external fields. Such an energy shift can be calculated using a gradient

[3] Such an addition is referred to as H_n in Culcer's paper [29].

expansion [24, 29]: since the wavepacket is localized, we only need to know the values of $\phi(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ near $\mathbf{r}_c$. Therefore, they are expanded as

$$\phi(\mathbf{r}) \simeq \phi(\mathbf{r}_c) + (\mathbf{r}-\mathbf{r}_c)\cdot\frac{\partial\phi}{\partial\mathbf{r}_c}, \tag{9}$$

$$\mathbf{A}(\mathbf{r}) \simeq \mathbf{A}(\mathbf{r}_c) + (\mathbf{r}-\mathbf{r}_c)\cdot\frac{\partial\mathbf{A}}{\partial\mathbf{r}_c}. \tag{10}$$

Such an approximation is valid only if the potentials vary slowly across the wavepacket.

The values of $\phi_c = \phi(\mathbf{r}_c)$ and $\mathbf{A}_c = \mathbf{A}(\mathbf{r}_c)$ do not need to be small compared to the unperturbed part since they only contribute to a shift in energy by $-e\phi_c$ and a shift in crystal momentum by $e\mathbf{A}_c$ (the charge of the carrier is $-e$). However, the field strengths are limited by the adiabatic condition: the total probability in these D bands should remain conserved. There can be no tunneling out of this subspace.

Assuming the electric and magnetic fields are uniform, choosing the symmetric gauge for the vector potential, and neglecting nonlinear terms in $\mathbf{A}$, one finds

$$\begin{aligned} E(\mathbf{r}_c,\mathbf{q}_c) &= \langle W|H_c|W\rangle + e\langle W|(\mathbf{r}-\mathbf{r}_c)|W\rangle\cdot\mathbf{E} \\ &\quad + \frac{e}{2m}\langle W|(\mathbf{r}-\mathbf{r}_c)\times(\mathbf{p}+e\mathbf{A}_c)|W\rangle\cdot\mathbf{B} \\ &= E_0(\mathbf{k}_c) - e\phi(\mathbf{r}_c) + \frac{e}{2m}\mathbf{L}(\mathbf{k}_c)\cdot\mathbf{B}, \end{aligned} \tag{11}$$

where

$$H_c \equiv H_0(\mathbf{r},\mathbf{p}+e\mathbf{A}_c) - e\phi_c. \tag{12}$$

Notice that $\langle W|(\mathbf{r}-\mathbf{r}_c)|W\rangle$ is zero and $\mathbf{k}_c \equiv \mathbf{q}_c + (e/\hbar)\mathbf{A}_c$. If a charged wavepacket is self-rotating, then it has an orbital magnetic moment $\mathbf{m} = -e/(2m)\mathbf{L}$ coupling with the magnetic field and gives the Zeeman energy.

Since nonlinear terms have been neglected in the calculation above, the following result is accurate only to linear order of the fields. The nonlinear electromagnetic effect is thus outside the scope of this review.

2.4. Wavepacket dynamics: Lagrangian and equations of motion

Instead of solving the Schrödinger equation, we use the time-dependent variational method to study the wavepacket evolution in a crystal. Recall that in the usual time-independent variational method one proposes a sensible wavefunction with a few adjustable parameters. These parameters are then determined by minimizing the energy $\langle H\rangle$ of the proposed state. In the time-dependent variational method, the wavepacket is parametrized by the center of mass, $\mathbf{r}_c(t)$ and $\mathbf{q}_c(t)$, and the spinor amplitude $\boldsymbol{\eta} = (\eta_1,\eta_2,\ldots,\eta_D)^T$. Any possible change of shape of the wavepacket will be ignored. One first calculates the effective Lagrangian,

$$L_{\rm eff}(\mathbf{r}_c,\mathbf{q}_c,\boldsymbol{\eta};\dot{\mathbf{r}}_c,\dot{\mathbf{q}}_c,\dot{\boldsymbol{\eta}}) = \langle W|\left(\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t} - H\right)|W\rangle, \tag{13}$$

then obtains the effective action $S_{\rm eff} = \int \mathrm{d}t\, L_{\rm eff}$ for a given trajectory. Unlike the energy $\langle H\rangle$ in the time-independent variational method that depends on only a few parameters, this effective action is a *functional* of the trajectory and the time evolution of the spinor amplitude: $(\mathbf{r}_c(t),\mathbf{q}_c(t),\boldsymbol{\eta}(t))$. Therefore, one needs to use the method of variation to determine the one that extremizes this action. The Euler–Lagrange equations for this effective Lagrangian give us equations of motion of the wavepacket.

By carefully analyzing the time-derivative term in equation (13), one obtains

$$\mathrm{i}\hbar\langle W|\frac{\mathrm{d}}{\mathrm{d}t}|W\rangle = \hbar\left\langle u\middle|\mathrm{i}\frac{\mathrm{d}u}{\mathrm{d}t}\right\rangle + \hbar\mathbf{q}_c\cdot\dot{\mathbf{r}}_c. \tag{14}$$

The first term can further be decomposed as

$$\hbar\left\langle u\middle|\mathrm{i}\frac{\mathrm{d}u}{\mathrm{d}t}\right\rangle = \mathrm{i}\hbar\boldsymbol{\eta}^\dagger\frac{\partial\boldsymbol{\eta}}{\partial t} + \hbar\dot{\mathbf{k}}_c\cdot\boldsymbol{\mathcal{R}}(\mathbf{k}_c), \tag{15}$$

where we have used the gauge-invariant crystal momentum $\mathbf{k}_c = \mathbf{q}_c + (e/\hbar)\mathbf{A}_c$ in place of $\mathbf{q}_c$ as a basic variable. Therefore, the effective Lagrangian is

$$\begin{aligned} L_{\rm eff}(\mathbf{r}_c,\mathbf{k}_c,\boldsymbol{\eta};\dot{\mathbf{r}}_c,\dot{\mathbf{k}}_c,\dot{\boldsymbol{\eta}}) &= \mathrm{i}\hbar\boldsymbol{\eta}^\dagger\frac{\partial\boldsymbol{\eta}}{\partial t} \\ &\quad + \hbar\dot{\mathbf{k}}_c\cdot\boldsymbol{\mathcal{R}} + (\hbar\mathbf{k}_c - e\mathbf{A}_c)\cdot\dot{\mathbf{r}}_c - E(\mathbf{r}_c,\mathbf{k}_c), \end{aligned} \tag{16}$$

in which $E = \langle W|H|W\rangle$ is the wavepacket energy calculated in equation (11) using the gradient expansion. Notice that there are three pairs of independent variables, as indicated in the argument of $L_{\rm eff}$.

After lengthy derivations, the Euler–Lagrange equations of this effective Lagrangian give the following (coupled) dynamical equations of the wavepacket [29]:

$$\hbar\dot{\mathbf{k}}_c = -e\mathbf{E} - e\dot{\mathbf{r}}_c\times\mathbf{B}, \tag{17}$$

$$\hbar\dot{\mathbf{r}}_c = \frac{1}{\mathrm{i}}\boldsymbol{\eta}^\dagger\left[\mathrm{i}\frac{\partial}{\partial\mathbf{k}_c} + \boldsymbol{\mathcal{R}}, \mathcal{H}\right]\boldsymbol{\eta} - \hbar\dot{\mathbf{k}}_c\times\boldsymbol{\eta}^\dagger\boldsymbol{\mathcal{F}}\boldsymbol{\eta}, \tag{18}$$

$$\mathrm{i}\hbar\dot{\boldsymbol{\eta}} = \left(\frac{e}{2m}\boldsymbol{\mathcal{L}}\cdot\mathbf{B} - \hbar\dot{\mathbf{k}}_c\cdot\boldsymbol{\mathcal{R}}\right)\boldsymbol{\eta}, \tag{19}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the external fields, and $\boldsymbol{\mathcal{R}}$, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\mathcal{L}}$ are all evaluated at the center of the wavepacket. The Hamiltonian matrix within the commutator in equation (18) has the form

$$\mathcal{H}(\mathbf{r}_c,\mathbf{k}_c) = E_0(\mathbf{k}_c) - e\phi(\mathbf{r}_c) + \frac{e}{2m}\boldsymbol{\mathcal{L}}(\mathbf{k}_c)\cdot\mathbf{B}, \tag{20}$$

in which $E_0 - e\phi$ is multiplied by an identity matrix. Obviously, the spinor average of $\mathcal{H}$ is the wavepacket energy in equation (11), $E = \boldsymbol{\eta}^\dagger\mathcal{H}\boldsymbol{\eta} = \sum_{nl}\eta_n^* H_{nl}\eta_l$.

From the spinor equation (19), one can immediately obtain the equation for the spin vector, $\mathbf{J} = \boldsymbol{\eta}^\dagger\boldsymbol{\mathcal{J}}\boldsymbol{\eta}$, where $\boldsymbol{\mathcal{J}}$ is the usual angular momentum matrix in quantum mechanics,

$$\mathrm{i}\hbar\dot{\mathbf{J}} = \boldsymbol{\eta}^\dagger\left[\boldsymbol{\mathcal{J}}, \mathcal{H} - \hbar\dot{\mathbf{k}}_c\cdot\boldsymbol{\mathcal{R}}\right]\boldsymbol{\eta}. \tag{21}$$

The only term in $\mathcal{H}$ that would contribute to the commutator above is the Zeeman coupling. The correction to the Hamiltonian that is proportional to the Berry connection $\boldsymbol{\mathcal{R}}$ is in fact a spin–orbit coupling. This will become clear in later discussions.

3. Semiclassical theory for a relativistic electron

To have a better grasp of the theoretical formulation in the previous section, let us take the relativistic particle as an illustrative example [33, 34]. The wavepacket is living in the electron subspace. Since the electron states are separated from the positron states by a huge energy gap mc^2, the adiabatic condition is hardly violated. It fails only if electron–positron creation is not negligible. Furthermore, unlike the Bloch states in a solid, the energy eigenstates of a free Dirac particle are analytically known. Therefore, the calculations of **R**, **F**, and **L** are clean and exact. These quantities are essential to the dynamics of a Dirac wavepacket in an electromagnetic field.

3.1. Basics of the Dirac theory

The Dirac Hamiltonian in the presence of an electromagnetic field is

$$H = c\boldsymbol{\alpha}\cdot(\mathbf{p}+e\mathbf{A}) + \beta mc^2 - e\phi(\mathbf{r}), \tag{22}$$

where $\boldsymbol{\alpha}$ and β are the Dirac matrices,

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \qquad \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{23}$$

in which $\boldsymbol{\sigma}$ are the Pauli matrices and I is the 2×2 identity matrix. The off-diagonal matrix $c\boldsymbol{\alpha}$ is the velocity operator for the Dirac electron.

In the absence of an external field, one can perform the following unitary transformation to diagonalize the Hamiltonian, $H_0 = \mathrm{e}^{\mathrm{i}S}H_{0d}\mathrm{e}^{-\mathrm{i}S}$, where $H_{0d} = \sqrt{c^2|\mathbf{p}|^2 + m^2c^4}\beta$ has positive-energy and negative-energy branches (each branch has two degenerate levels). The unitary operator $\mathrm{e}^{\mathrm{i}S} = e^{\frac{\omega}{2}\beta\boldsymbol{\alpha}\cdot\mathbf{p}/|\mathbf{p}|}$, where $\tan\omega = p/mc$ [35]. The four independent plane-wave solutions of H_0 are of the form $\psi_l(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}u_l(\mathbf{r})$, where the spinor u_l can be obtained from the four-component *unit spinor* e_l via the unitary rotation, $u_l = \mathrm{e}^{\mathrm{i}S}e_l$.

We emphasize that these solutions remain normalized at any velocity. The equally valid solutions obtained from the unit spinors via a Lorentz boost are not normalized, since the particle density is not invariant under the Lorentz transformation. These two sets of solutions have slightly different normalization constants, but otherwise are of the same form. However, the Berry connection $\boldsymbol{\mathcal{R}}$ in equation (4) can be Hermitian *only if* we use the normalized eigenstates, which are obtained from the unitary rotation.

3.2. Basics of the Dirac wavepacket

The Dirac wavepacket is constructed from the free-particle states with positive energy $E_0 = \sqrt{c^2\hbar^2q^2 + m^2c^4}$ (see figure 1),

$$|W\rangle = \sum_{n=1}^{2}\int \mathrm{d}^3q\, a(\mathbf{q},t)\eta_n(\mathbf{q},t)|\psi_n\rangle. \tag{24}$$

In such a subspace with twofold spin degeneracy, the Berry connection and curvature are 2×2 matrices. Using the definitions in equations (4) and (5), after some calculations, one can obtain

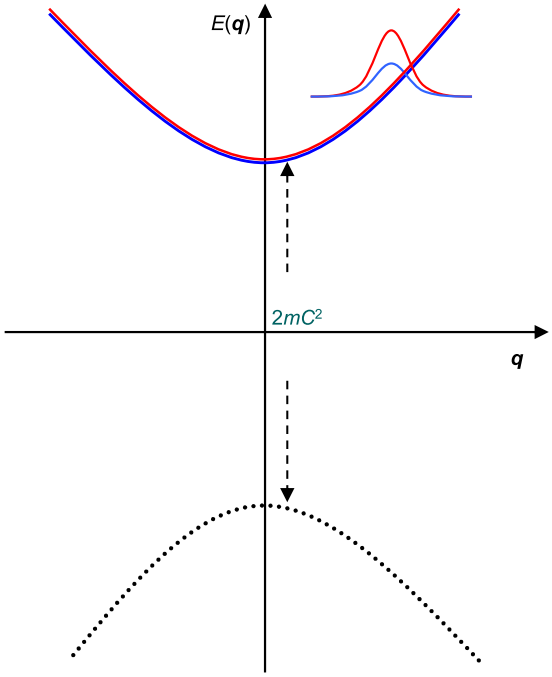

Figure 1. The Dirac spectrum has two branches separated by a huge mass gap. Each branch is twofold degenerate. Therefore, the wavepacket on the positive-energy branch has two localized probability amplitudes.

$$\boldsymbol{\mathcal{R}} = \frac{\lambda_c^2}{2\gamma(\gamma+1)}\mathbf{q}\times\boldsymbol{\sigma}, \tag{25}$$

$$\boldsymbol{\mathcal{F}} = -\frac{\lambda_c^2}{2\gamma^3}\left(\boldsymbol{\sigma} + \lambda_c^2\frac{\mathbf{q}\cdot\boldsymbol{\sigma}}{\gamma+1}\mathbf{q}\right), \tag{26}$$

where $\lambda_c = \hbar/mc$ is the Compton wavelength, and $\gamma(q) \equiv \sqrt{1+(\hbar q/mc)^2}$. The dilation factor γ reduces to one when the electron is at rest ($q = 0$).

It is remarkable that even a free electron would have a non-trivial gauge structure. The Berry connection vanishes at zero velocity but the Berry curvature reduces to a non-zero quantity proportional to the Pauli matrix, $-(\lambda_c^2/2)\boldsymbol{\sigma}$. The Berry curvature is not originated from level crossing and therefore is nothing like a monopole field. A similar result has been reported in [36, 37].

The other truly remarkable result is that such a free-electron wavepacket is in fact self-rotating with an orbital angular momentum [34]. A straightforward application of equation (8) gives[4]

$$\boldsymbol{\mathcal{L}} = \frac{\hbar}{\gamma^2}\left(\boldsymbol{\sigma} + \lambda_c^2\frac{\mathbf{q}\cdot\boldsymbol{\sigma}}{\gamma+1}\mathbf{q}\right). \tag{27}$$

[4] It has a functional form similar to $\boldsymbol{\mathcal{F}}$. This is related to the fact that, when $\boldsymbol{\mathcal{L}}$ is written in the form of equation (51) (shown later), it differs from an alternative form of the Berry curvature in equation (50) only by a factor of energy (apart from some constants). In the two-branch model we used, this only brings about a difference of a factor of $2E_0$.

After spinor average, this self-rotating wavepacket generates a magnetic moment $\mathbf{m} = -e/(2m)\mathbf{L}$ that couples with the magnetic field and gives rise to the Zeeman energy in equation (11). The coefficient $-e/(2m)$ finds its origin in the fact that $\mathbf{L} = \int \mathrm{d}^3 r(\mathbf{r}-\mathbf{r}_\mathrm{c}) \times m\mathbf{j}(\mathbf{r})$, where $\mathbf{j}$ is the current density, while $\mathbf{m} = \frac{1}{2}\int \mathrm{d}^3 r(\mathbf{r}-\mathbf{r}_\mathrm{c}) \times (-e)\mathbf{j}(\mathbf{r})$.

When the momentum q is zero, the orbital angular momentum $\mathcal{L}$ reduces to $\hbar\boldsymbol{\sigma}$, twice the spin angular momentum. It produces a Zeeman energy identical to the usual one in which the spin is $(\hbar/2)\boldsymbol{\sigma}$ and the g-factor is 2. However, we emphasize that the g-factor in our formulation equals one, not two! That is, in the wavepacket description, there is no need to assign a special g-factor to the electron spin.

Neither in the Dirac equation nor in its four-component wave functions can one find explicit trace of the spin. The established mathematical procedures, such as the Foldy and Wouthuysen transformation [22], for extracting the Zeeman interaction are usually complicated, and the appearance of the intrinsic magnetic moment for the electron has always been mysterious. The wavepacket approach offers an alternative and very intuitive picture explaining the origin of the electron spin: it is indeed a self-rotating motion that produces the angular momentum in equation (27). Such a delightful result is a natural consequence of the wavepacket theory and requires no forced assumption. For more discussions on the spinning Dirac wavepacket, one can also see [38].

3.3. Dynamics of the Dirac wavepacket

Once the quantities $\mathcal{H}$, $\boldsymbol{\mathcal{R}}$, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\mathcal{L}}$ are known, one can proceed to derive the equations of motion in equations (17)–(19). Correct to linear order in fields, the equations of motion for the center of mass are

$$\hbar\dot{\mathbf{k}}_\mathrm{c} = -e\mathbf{E} - e\frac{\hbar\mathbf{k}_\mathrm{c}}{\gamma m} \times \mathbf{B}, \tag{28}$$

$$\dot{\mathbf{r}}_\mathrm{c} = \frac{\hbar\mathbf{k}_\mathrm{c}}{\gamma m} + \frac{e}{\hbar}\left(\mathbf{E}\times\mathbf{F} + \mathbf{B}\cdot\mathbf{F}\frac{\hbar\mathbf{k}_\mathrm{c}}{\gamma m}\right), \tag{29}$$

where $\gamma = \gamma(\mathbf{k}_\mathrm{c})$ and $\mathbf{F} = \boldsymbol{\eta}^\dagger\boldsymbol{\mathcal{F}}\boldsymbol{\eta}$ is the spinor-averaged Berry curvature. The equation for spin precession in equation (21) for the Dirac electron has the following form:

$$\langle\dot{\boldsymbol{\sigma}}\rangle = \frac{e}{\gamma m}\left[\mathbf{B} + \mathbf{E}\times\frac{\hbar\mathbf{k}_\mathrm{c}}{(\gamma+1)mc^2}\right]\times\langle\boldsymbol{\sigma}\rangle. \tag{30}$$

This is the Bargmann–Michel–Telegdi (BMT) equation for a relativistic electron [39]. The second term proportional to the electric field couples the orbital motion with the spin. It is originated from the $-\hbar\dot{\mathbf{k}}\cdot\boldsymbol{\mathcal{R}}$ term in equation (21) and thus is closely related to the Berry connection.

If we turn off the magnetic field in equations (28) and (29), then the trajectory is much simpler but still non-trivial. There is a linear acceleration along the direction of the electric field, $\hbar\dot{\mathbf{k}}_\mathrm{c} = -e\mathbf{E}$, which is reasonable. However, there is a transverse velocity as a result of the Berry curvature in equation (29), even though there is no external force along this direction! The magnitude of the transverse particle momentum is $m(e/\hbar)\mathbf{E}\times\mathbf{F}$, which reduces to $\mathbf{E}\times\mathbf{m}/c^2$ at low velocity, where $\mathbf{m} = -e/(2m)\mathbf{L}$ (see figure 2).[5]

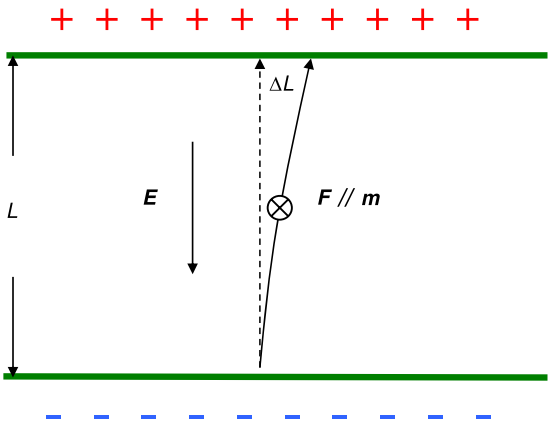

Figure 2. An electron is accelerated by a static electric field in the capacitor. The Berry curvature and the magnetic moment point into the page, giving rise to an anomalous velocity pointing to the right.

From a historical perspective, one can also understand such a curved trajectory in the following way: the theory of relativity tells us that a moving magnetic dipole $\mathbf{m}$ generates an effective electric dipole $\mathbf{p}_\mathrm{eff} = \mathbf{v}\times\mathbf{m}/c^2$ [41]. Being in a uniform field, the electron feels no gradient dipole force. However, it acquires a dipole energy $-\mathbf{p}_\mathrm{eff}\cdot\mathbf{E}$. Such a term contributes to the classical Hamiltonian. Therefore, from the Hamiltonian equation $\dot{\mathbf{r}} = \partial H/\partial\mathbf{p}$, one would get an extra velocity $\mathbf{v} = \mathbf{E}\times\mathbf{m}/(mc^2)$, the same as the particle velocity above[6].

It would be very impressive to confirm such a trajectory shift in free space. The electron can be accelerated by a parallel-plate capacitor with a potential difference V_0. Aligning the electron spin perpendicular to the electric field maximizes the effect. When moving from one plate to another, the electron trajectory will be shifted sideways by a distance $\lambda_\mathrm{c}\sqrt{eV_0/2mc^2}$ (assuming the electron is initially at rest). That is, for a potential difference of 1 MeV, the shift is of the order of the Compton wavelength, which is two orders of magnitude smaller than the size of a hydrogen atom. One can alternate the polarity of the electrodes to synchronize with the electron motion and amplify the shift. Nevertheless, a huge potential difference is required and such an observation will not be easy.

4. Canonical quantization

The wavepacket approach for electron transport, even though approximate in general, can be accurate at a certain limit. One

[5] Whenever a magnetic dipole $\mathbf{m}$ produced by a current loop (such as the wavepacket here) is put in an electric field $\mathbf{E}$, it is accompanied by a hidden momentum $\mathbf{p}_\mathrm{hidden} = \mathbf{m}\times\mathbf{E}/c^2$ [40, 41]. When this is taken into account, one finds that the part of $\hbar\mathbf{k}_\mathrm{c}$ that is perpendicular to $\mathbf{E}$ remains conserved, consistent with its equation of motion.

[6] In addition to the velocity mentioned above, the effective electric dipole is closely related to the electric polarization generated by a spin current; see [42]. In the free space, a spin current can be modeled by counter-propagating beams of bare spin-up and spin-down electrons. It has been known for a long time that the charge center of the beams does not coincide with its mass center [43]. This is a direct result of the fact that the effective polarizations of both beams point in the same direction.

can take advantage of the simplicity of the semiclassical theory and re-quantize it to study various quantum phenomena. Under this certain limit, the prediction of the simplified (effective) quantum theory could be as accurate as the original quantum theory.

In an earlier work, we used the Bohr–Sommerfeld condition to re-quantize the cyclotron orbit and successfully predicted the positions of Landau levels in magnetic Bloch bands [7]. This approach, however, is limited to integrable systems with closed trajectories. Here we use the method of canonical quantization that does not have these restrictions and, moreover, can be applied to the multi-band cases. We demonstrate that the knowledge of $\mathcal{H}$, $\mathcal{R}$, $\mathcal{F}$, and $\mathcal{L}$ in the semiclassical formulation is sufficient for us to obtain the correct quantum theory.

In this section, we first find out the classical canonical variables, then promote the Poisson bracket to a commutator. As a result, the dynamical variables become operators acting on a Hilbert space and the theory is thus 'quantized'. That is, we can turn the wavepacket energy into an effective quantum Hamiltonian. The Dirac wavepacket will again be used as an example and its effective quantum Hamiltonian turns out to be the well known Pauli Hamiltonian.

4.1. Deriving canonical variables

The complementary pair $(\mathbf{r}, \mathbf{p})$ are canonical variables if their equations of motion have the following canonical form:

$$\dot{\mathbf{r}} = \frac{\partial E}{\partial \mathbf{p}}; \qquad \dot{\mathbf{p}} = -\frac{\partial E}{\partial \mathbf{r}}. \tag{31}$$

This also means that their Poisson brackets have the canonical form

$$\begin{aligned} \{r_\alpha, r_\beta\} &= 0, \\ \{p_\alpha, p_\beta\} &= 0, \\ \{r_\alpha, p_\beta\} &= \delta_{\alpha\beta}. \end{aligned} \tag{32}$$

The variables $\mathbf{r}_c$ and $\mathbf{k}_c$ that depict the trajectory of the wavepacket are not canonical variables because their equations of motion are not of the form above. This is due to the gauge potentials, $\mathbf{A}(\mathbf{r}_c)$ and $\mathbf{R}(\mathbf{k}_c)$, that generate rich dynamics but are otherwise harmless. However, in order to re-quantize the theory, one needs to find canonical variables with a simple Poisson bracket. In principle, this can always be done, as assured by the Darboux theorem [44]. Investigation along this path offers a powerful tool for the quantization.

In general, there is no easy way to exactly construct the canonical variables for systems with non-Abelian gauge potentials. However, for the wavepacket theory in this review, which is accurate only to linear fields, there is a simple way of finding them even for the multi-band version. It is crucial to observe that the source of non-canonicity is the presence of the gauge potentials, $\mathbf{A}(\mathbf{r}_c)$ and $\mathbf{R}(\mathbf{k}_c)$, in the Lagrangian. They need to be removed when the Lagrangian is written in the canonical variables $\mathbf{r}$ and $\mathbf{p}$. That is, once the Lagrangian in equation (16) can be written in the standard form (up to a total time derivative),

$$L_{\rm eff} = \mathrm{i}\hbar\boldsymbol{\eta}^\dagger \frac{\partial \boldsymbol{\eta}}{\partial t} + \mathbf{p}\cdot\dot{\mathbf{r}} - E(\mathbf{r}, \mathbf{p}), \tag{33}$$

then the new variables $\mathbf{r}$ and $\mathbf{p}$ will automatically be canonical[7].

It is reasonable to make the initial guess

$$\hbar\mathbf{k}_c = \hbar\mathbf{k} + e\mathbf{A}(\mathbf{r}_c); \qquad \mathbf{r}_c = \mathbf{x} + \mathbf{R}(\mathbf{k}_c), \tag{35}$$

and rewrite the effective Lagrangian in equation (16) in the new variables, $\mathbf{x}$ and $\mathbf{k}$, as

$$L_{\rm eff} = \mathrm{i}\hbar\boldsymbol{\eta}^\dagger \frac{\partial \boldsymbol{\eta}}{\partial t} + e\dot{\mathbf{A}}\cdot\mathbf{R} + \hbar\mathbf{k}\cdot\dot{\mathbf{x}} - E(\mathbf{x}, \mathbf{k}) + \frac{\mathrm{d}}{\mathrm{d}t}(\hbar\mathbf{k}\cdot\mathbf{R}), \tag{36}$$

in which the total time derivative will be ignored. This is still not of the standard form. To fix it, we choose a gauge $\mathbf{A} = \frac{1}{2}\mathbf{B}\times\mathbf{r}_c$ and write $e\dot{\mathbf{A}}\cdot\mathbf{R} = \frac{e}{2}\mathbf{R}\times\mathbf{B}\cdot(\dot{\mathbf{x}}+\dot{\mathbf{R}})$. After some rearrangement, one will get

$$e\dot{\mathbf{A}}\cdot\mathbf{R} + \hbar\mathbf{k}\cdot\dot{\mathbf{x}} = \left[\hbar\mathbf{k} + \frac{e}{2}\mathbf{R}(\mathbf{k}_c)\times\mathbf{B}\right]\cdot\dot{\mathbf{x}} + \mathbf{G}(\mathbf{k}_c)\cdot\hbar\dot{\mathbf{k}}_c, \tag{37}$$

where $G_\alpha(\mathbf{k}_c) \equiv (e/\hbar)(\mathbf{R}\times\mathbf{B})\cdot\partial\mathbf{R}/\partial k_{c\alpha}$.

The term proportional to $\mathbf{G}$ can be removed by further shifting the position variable, $\mathbf{x} = \mathbf{r} + \mathbf{G}$. After throwing away a total time derivative $\mathrm{d}(\mathbf{G}\cdot\hbar\mathbf{k}_c)/\mathrm{d}t$ and a term explicitly nonlinear in B, we finally have

$$L_{\rm eff} = \mathrm{i}\hbar\boldsymbol{\eta}^\dagger \frac{\partial \boldsymbol{\eta}}{\partial t} + \left[\hbar\mathbf{k} + \frac{e}{2}\mathbf{R}(\mathbf{k}_c)\times\mathbf{B}\right]\cdot\dot{\mathbf{r}} - E. \tag{38}$$

Obviously, the quantity in the square bracket should be identified as the canonical momentum $\mathbf{p}$ conjugate to the variable $\mathbf{r}$ (see equation (33)). Therefore, the connections between new and old variables are [16]

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_c - \mathbf{R}(\mathbf{k}_c) - \mathbf{G}(\mathbf{k}_c), \\ \mathbf{p} &= \hbar\mathbf{k}_c - e\mathbf{A}(\mathbf{r}_c) - \frac{e}{2}\mathbf{B}\times\mathbf{R}(\mathbf{k}_c). \end{aligned} \tag{39}$$

The inverses of the relations are (again accurate to linear fields)

$$\begin{aligned} \mathbf{r}_c &= \mathbf{r} + \mathbf{R}(\boldsymbol{\pi}) + \mathbf{G}(\boldsymbol{\pi}), \\ \hbar\mathbf{k}_c &= \boldsymbol{\pi} + e\mathbf{B}\times\mathbf{R}(\boldsymbol{\pi}), \end{aligned} \tag{40}$$

where $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}(\mathbf{r})$. This is the correct generalization of the Peierls substitution for systems with Berry connection $\mathbf{R}$. It reduces to the usual Peierls substitution when $\mathbf{R} = 0$. Similar equations can be found via a route slightly more complicated than the one reported here [46].

Two comments are in order. First, the center-of-mass variables are gauge invariant, but the canonical variables are not. Therefore, only the center-of-mass variables can be considered as physical variables. Second, the G-term would further shift the position operator, but does not influence the velocity to linear order of fields. Similarly, the $\mathbf{B}\times\mathbf{R}$-term would shift the momentum, but does not alter the force to the same order.

[7] It is possible to rewrite the first term in $L_{\rm eff}$ as the Wess–Zumino term,

$$\mathrm{i}\boldsymbol{\eta}^\dagger\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = -S\int_0^1 \mathrm{d}\lambda\hat{\mathbf{S}}_\lambda\cdot\frac{\partial\hat{\mathbf{S}}_\lambda}{\partial\lambda}\times\frac{\partial\hat{\mathbf{S}}_\lambda}{\partial t}, \tag{34}$$

where $\mathbf{S}$ is the spin vector. See [45] for more details. However, we use the spinor amplitude $\boldsymbol{\eta}$ as a basic variable in the Lagrangian, instead of the spin vector $\mathbf{S}$.

4.2. Effective quantum Hamiltonian

When written in the canonical variables, the wavepacket energy in equation (11) becomes

$$E(\mathbf{r},\mathbf{p}) = E_0(\boldsymbol{\pi}) - e\phi(\mathbf{r}) + e\mathbf{E}\cdot\mathbf{R}(\boldsymbol{\pi}) + \frac{e}{2m}\mathbf{B}\cdot\left[\mathbf{L}(\boldsymbol{\pi}) + 2\mathbf{R}\times m\frac{\partial E_0}{\partial\boldsymbol{\pi}}\right]. \tag{41}$$

The third term on the RHS comes from expanding the electrostatic potential $\phi(\mathbf{r}_c)$ about the canonical position $\mathbf{r}$ to first order, with nonlinear terms neglected. It has the form of a dipole energy, with $\mathbf{R}$ being the displacement of the charge center. Later we will see that this term represents the usual spin–orbit interaction. The last term has the form of the Zeeman energy, but in addition to the usual magnetic moment generated by $\mathbf{L}$ there is an extra term from expanding the band energy $E_0(\mathbf{k}_c)$ about $\boldsymbol{\pi}$ to first order. We will call it a Yafet term [21, 47, 48].

The effective quantum Hamiltonian can be obtained from the wavepacket energy by the standard procedure of promoting the variables to operators,

$$[r_\alpha, r_\beta] = 0, \qquad [p_\alpha, p_\beta] = 0, \qquad [r_\alpha, p_\beta] = \mathrm{i}\hbar\delta_{\alpha\beta}.$$

In addition, the basic quantities, $\mathbf{R}$, $\mathbf{F}$, and $\mathbf{L}$, also need to be promoted to matrices, $\mathcal{R}$, $\mathcal{F}$, and $\mathcal{L}$.

The effective Hamiltonian, being a function of $\boldsymbol{\pi}$, is invariant under the electromagnetic gauge rotation. However, the terms explicit in the Berry connection $\mathbf{R}$, such as the dipole energy and the Yafet term, are not invariant under the gauge rotation that re-shuffles the spinor basis. Therefore, it would be tricky to give a clear physical interpretation to each individual term. The original semiclassical theory based on the physical center-of-mass variables, on the other hand, is free of any gauge ambiguity.

4.3. Relativistic Pauli Hamiltonian

In this subsection, we will again take the Dirac electron as an example to illustrate the power of this re-quantization procedure. Combining equations (11) and (27), one has the energy of a Dirac wavepacket ($E_0 = \gamma mc^2$),

$$\begin{aligned} E(\mathbf{r}_c,\mathbf{k}_c) &= E_0(\mathbf{r}_c) - e\phi_c + \frac{e}{2m}\mathbf{L}(\mathbf{k}_c)\cdot\mathbf{B} \\ &= \gamma mc^2 - e\phi(\mathbf{r}_c) + \frac{\mu_B}{\gamma^2}\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{B} \\ &\quad + \frac{\mu_B}{\gamma^2(\gamma+1)}\frac{\hbar\mathbf{k}_c}{mc}\cdot\langle\boldsymbol{\sigma}\rangle\frac{\hbar\mathbf{k}_c}{mc}\cdot\mathbf{B}, \end{aligned} \tag{42}$$

in which all the γ are functions of $\mathbf{k}_c$ and $\mu_B = e\hbar/2m$. In the next step, the generalized Peierls substitution in equation (40) is used to write the energy in canonical variables. One needs, for example,

$$\begin{aligned} \gamma(\mathbf{k}_c)mc^2 &= mc^2\sqrt{1+\left(\frac{\boldsymbol{\pi}+e\mathbf{B}\times\mathbf{R}}{mc}\right)^2} \\ &\simeq \gamma(\boldsymbol{\pi})mc^2 + \mu_B\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{B}\left(\frac{\gamma-1}{\gamma^2}\right) \\ &\quad - \frac{\mu_B}{m^2c^2}\frac{\boldsymbol{\pi}\cdot\langle\boldsymbol{\sigma}\rangle\boldsymbol{\pi}\cdot\mathbf{B}}{\gamma^2(\gamma+1)}, \end{aligned} \tag{43}$$

in which all the γ are functions of $\boldsymbol{\pi}$. After some more calculations, one obtains an energy slightly simpler in form,

$$E(\mathbf{r},\mathbf{p}) = \gamma(\boldsymbol{\pi})mc^2 - e\phi(\mathbf{r}) + \frac{\mu_B}{\gamma}\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{B} + \frac{\mu_B}{\gamma(\gamma+1)}\frac{\boldsymbol{\pi}}{mc^2}\times\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{E}. \tag{44}$$

In the final step, one promotes the classical canonical variables $\mathbf{r}$ and $\mathbf{p}$ to quantum canonical variables, and $\langle\boldsymbol{\sigma}\rangle$ to Pauli matrices. The result is the effective quantum Hamiltonian for an electron living on the positive-energy branch. It allows no inter-branch transition (i.e. pair creation), but otherwise is valid for all velocities of the electron. At low velocity, it reduces to the familiar Pauli Hamiltonian.

This is by no means a trivial result since the gauge-invariant momentum operator $\boldsymbol{\pi}$ appears inside the square root in the denominator. In the usual textbook derivation, one performs successive Foldy–Wouthuysen (FW) transformations [49], treating the electron velocity as an expansion parameter. Such a unitary transformation quickly becomes very complicated at high velocity. One has to ingeniously come up with the correct unitary operator directly, assisted by some guessing, in order to reach the closed form in equation (44) [50]. By comparison, the approach presented here offers a simple and systematic way to derive an effective Hamiltonian when one has difficulty implementing an FW-like transformation.

4.4. Discussion

In addition to the re-quantization approach reported above, we would like to mention two other methods for obtaining an effective Hamiltonian. One method uses the unitary transformation to block-diagonalize a Hamiltonian. This is the method used by Foldy and Wouthuysen to obtain the Pauli Hamiltonian from the Dirac Hamiltonian [22]. Another method uses projection operators to restrict the extent of a Hilbert space [51]. This method has been implemented in single-particle as well as many-body quantum systems. More discussion of the effective Hamiltonian can also be found in Nenciu's review [27].

The zero-field Dirac Hamiltonian can be diagonalized by the unitary rotation given below equation (23). However, an external electric or magnetic field generates off-diagonal terms in the Hamiltonian matrix. One can perform the second unitary rotation U_{BD} to eliminate such couplings between positive- and negative-energy branches. This is the block-diagonalization process used in the FW transformation [22]. The Dirac Hamiltonian then consists of two disjoined pieces. Each piece is the effective 2×2 Hamiltonian for the corresponding energy branch. That is, $H_{\mathrm{eff},\pm} = P_\pm U_{BD}^\dagger H U_{BD} P_\pm$, where $P_\pm$ is the projection operator onto the positive-energy branch or the negative-energy branch.

This diagonalization method was used extensively in the 1950s and 1960s to obtain the effective Hamiltonian of a Bloch electron [4, 20, 21, 52]. In principle, one can systematically go beyond the linear order and obtain effective interactions

nonlinear in field, but the derivation could be (in Kohn's words) 'shockingly complicated' [53].

One thing to keep in mind is that, as a result of the unitary transformation, the basis states are changed. Therefore, the explicit representation of an observable, such as the position operator, would also change. This fact is stated most clearly in the original paper by Foldy and Wouthuysen [22, 54]. For example, in the one-band effective theory, $\mathbf{r}_{\mathrm{eff},n} = P_n U_{\mathrm{BD}}^{\dagger} \mathbf{r} U_{\mathrm{BD}} P_n$, where P_n projects to the nth band. In the simplest case when $U_{\mathrm{BD}} = 1$, the effective position operator has a simple correction from the Berry connection [4],

$$\mathbf{r}_{\mathrm{eff},n} = P_n \mathbf{r} P_n = \left\langle \psi_{n\mathbf{q}'} \left| \mathrm{i}\hbar \frac{\partial}{\partial \mathbf{p}} \right| \psi_{n\mathbf{q}} \right\rangle = \left[\mathrm{i}\frac{\partial}{\partial \mathbf{q}} + \mathbf{R}_n(\mathbf{q}) \right] \delta(\mathbf{q} - \mathbf{q}'), \tag{45}$$

where $\mathbf{R}_n(\mathbf{q})$ is the Berry connection of the nth band. It is not unusual to apply only the projection and ignore the unitary transformation: the off-diagonal matrix elements in the Hamiltonian are often linear in field. After the unitary rotation, they would usually translate to quadratic terms in the block-diagonalized Hamiltonian. Therefore, if one is only interested in an effective Hamiltonian linear in field, then it may not be necessary to perform the rotation U_{BD}. In such cases, the effective position operator is obtained simply by the one-band projection.

It is also possible to obtain an effective Hamiltonian using nothing but the projection operator. One divides the complete Hilbert space into two orthogonal parts, one relevant and one irrelevant. Consider a time-independent Schrödinger equation,

$$(H_0 + V)|\Psi\rangle = E|\Psi\rangle, \tag{46}$$

where V is a perturbation, and $H_0|\Psi^{(0)}\rangle = E_0|\Psi^{(0)}\rangle$. It can be shown that the projected (relevant) state $P|\Psi\rangle$ satisfies the following equation [51, 55]:

$$H_{\mathrm{eff}}(P|\Psi\rangle) = E(P|\Psi\rangle). \tag{47}$$

The effective Hamiltonian has the following form:

$$H_{\mathrm{eff}} = H_0 + PVP + PV\frac{1}{E_0 - QHQ}VP + PV\frac{1}{E_0 - QHQ}V\frac{1}{E_0 - QHQ}VP + \cdots, \tag{48}$$

where $Q = 1 - P$ projects to irrelevant subspace. This projection method has been used, for example, in deriving the Fano–Feschbach resonance [56], the 'poor man's method' of renormalization [57], and the one-band effective Hamiltonians of quantum Hall systems [58, 59].

5. Hierarchy of effective theories

The process of constructing a semiclassical theory out of a given quantum theory, then using the re-quantization method to turn it back to an effective quantum theory, can be executed iteratively to generate a hierarchy of effective theories.

The Dirac theory in equation (22) for lattice electrons can be considered as the theory at the top of the solid-state-theory hierarchy. Instead of solving the Dirac–Bloch spectrum directly, which is more difficult, one can first treat the periodic potential $\phi(\mathbf{r})$ as a *smooth* external potential. This is justified since the Dirac wavepacket can be as small as the Compton wavelength, two orders of magnitude smaller than the size of an atom. Therefore, earlier discussion on the Dirac wavepacket in section 3 is fully applicable in such a periodic environment. One first has the semiclassical Dirac theory, then the re-quantization leads to the two-component Pauli theory. We emphasize again that, from the semiclassical point of view, the electron spin is a direct result of the self-rotation of the Dirac wavepacket.

Starting from the quantum Pauli theory, one can solve for the Bloch energy bands with the relativistic and the spin–orbit effects. By focusing on closely packed multiple bands of interest, one can climb down the hierarchy a notch further. For example, in semiconductor physics, one is often interested in only the conduction and the valence bands above and below the chemical potential. Therefore, depending on the desired accuracy, one can construct an effective theory having only four bands (Luttinger model), eight bands (Kane model), or 12 bands (extended Kane model) near the fundamental gap. This will be the subject of the next section. In this section, the hierarchy structure is studied from a more general perspective.

5.1. Gauge-invariant ingredients of an effective theory

In order for the re-quantization to work correctly, it is essential that the three gauge-invariant quantities, E_0, $\mathbf{F}$, and $\mathbf{L}$ (or their matrix version for multiple bands, which are gauge covariant), are encoded in the semiclassical description. Since the Bloch energy E_0 is a rather well known concept, we focus only on the latter two quantities here.

In the multi-band version, the Berry curvature and self-rotating angular momentum defined in equations (5), (8) can be rewritten in alternative forms more convenient for calculation [16]. First we write equation (5) in the component form,

$$\mathbf{F}_{mn} = \frac{\partial}{\partial \mathbf{q}} \times \left\langle u_m \left| \mathrm{i}\frac{\partial}{\partial \mathbf{q}} u_n \right\rangle - \mathrm{i}\sum_{l\in \mathrm{in}} \left\langle u_m \left| \mathrm{i}\frac{\partial}{\partial \mathbf{q}} u_l \right\rangle \times \left\langle u_l \left| \mathrm{i}\frac{\partial}{\partial \mathbf{q}} u_n \right\rangle \right. \right. \right. \right.$$
$$= \mathrm{i}\left\langle \frac{\partial}{\partial \mathbf{q}} u_m \right| \times \left| \frac{\partial}{\partial \mathbf{q}} u_n \right\rangle - \mathrm{i}\sum_{l\in \mathrm{in}} \left\langle \frac{\partial}{\partial \mathbf{q}} u_m \middle| u_l \right\rangle \times \left\langle u_l \left| \frac{\partial}{\partial \mathbf{q}} u_n \right\rangle \right., \tag{49}$$

where l sums over the states *inside* the wavepacket space. One can insert a complete set $\sum_l$ to the first cross-product term, where l sums over the states in the whole Hilbert space. After canceling with the $\sum_{l\in \mathrm{in}}$ in the second term, one has

$$\mathbf{F}_{mn} = \mathrm{i}\sum_{l\in \mathrm{out}} \mathbf{R}_{ml} \times \mathbf{R}_{ln}, \tag{50}$$

where $\mathbf{R}_{ml}$ is the Berry connection, and l sums over the states *outside* of the wavepacket space. Similarly, by inserting a complete set to the bracket in equation (8), and using $\tilde{H}_0|u_{l(\in \mathrm{in})}\rangle = E_0|u_{l(\in \mathrm{in})}\rangle$, one has

$$\mathbf{L}_{mn} = \frac{m_0}{\mathrm{i}\hbar}\sum_{l\in \mathrm{out}} (E_{0,m} - E_{0,l})\mathbf{R}_{ml} \times \mathbf{R}_{ln}. \tag{51}$$

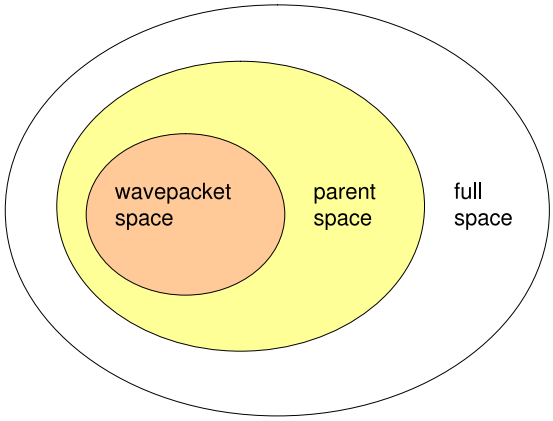

Figure 3. The relations between wavepacket space, parent space, and full space.

In the following, these two formulas are used to calculate the Berry curvature and the orbital angular momentum of a subspace.

In principle, equations (50) and (51) are exact if the summation is over the entire Hilbert space outside the wavepacket space. In practice, one often starts with a limited parent space, thereby neglecting terms from remote states outside this parent space (see figure 3). Let us call the approximate Berry curvature and orbital angular momentum obtained this way $\mathcal{F}'$ and $\mathcal{L}'$. In addition, let us assume that the neglected remote states generate a Berry curvature $\mathcal{F}_{\rm p}$ and an orbital angular momentum $\mathcal{L}_{\rm p}$ in the parent space, which is a subspace of the full space (see figure 3). Then from equations (50) and (51) they give corrections to the approximate $\mathcal{F}'$ and $\mathcal{L}'$ as follows:

$$\begin{aligned} \mathcal{F} &= \mathcal{F}' + P\mathcal{F}_{\rm p}P, \\ \mathcal{L} &= \mathcal{L}' + P\mathcal{L}_{\rm p}P, \end{aligned} \tag{52}$$

where P is the projection from the parent space to the wavepacket subspace. These relations describe precisely how the Berry curvature and orbital angular momentum should be renormalized. There is no need to refer back to the remote bands, once the Berry curvature and orbital angular momentum of the parent space are known.

Two comments are in order: first, the Berry curvature of the full Hilbert space for a given quantum model is always zero, since there is no 'outside' state in equation (50) [60]. The same is true for the self-rotating angular momentum. However, whether a Hilbert space is full or not is a relative concept. From the perspective of a more detailed quantum model with more degrees of freedom, the original Hilbert space is no longer full, but is only a subspace of a larger Hilbert space. Therefore, this subspace now can have its own Berry curvature. For example, from the viewpoint of the Dirac theory in equation (22), the positive-energy branch has a non-zero Berry curvature, whereas the four-component state itself is 'flat'. However, this latter fact is no longer true if the Dirac theory is considered as an effective theory of an even more complete theory (see section 5.3 for more discussions).

Second, the Berry curvature often can be attributed to a topological object in the parameter space. For example, it could be due to a $U(1)$ monopole at the location of level crossing [1]. The monopole charge is related to the first Chern number in the mathematical theory of fiber bundles [61]. The non-Abelian Berry curvature in this review is a result of projection, not level crossing. It has been shown that the Berry curvature in the Dirac theory is generated from a meron [37], an instanton with a topological charge 1/2. In the four-band Luttinger model (see next section), the Berry curvature is generated from an instanton with integer topological charge [62, 63].

5.2. Position and velocity

As we mentioned earlier in section 4.4, rotating the basis states of a Hilbert space amounts to a rotation of all operators, not only the Hamiltonian operator. Therefore, the effective Hamiltonian is only part of the story. From the 'parent theory' point of view, one also needs to know how the physical variables (e.g. $\mathbf{r}_{\rm c}$) are related to the canonical variables (e.g. $\mathbf{r}$, which is unphysical due to its gauge dependence) in the Hamiltonian to avoid mistakes.

In the generalized Peierls substitution (neglecting the G-term), $\mathbf{r}_{\rm c} = \mathbf{r} + \mathbf{R}$, $\mathbf{r}$ is the position operator of the re-quantized theory. The re-quantized theory can have its own semiclassical theory, with its own wavepacket centered at $\mathbf{r}_a \equiv \langle \mathbf{r} \rangle$. On the other hand, there is the original wavepacket centered at $\mathbf{r}_{\rm c}$. How would the motions of these two wavepackets relate to each other if they are living in the *same* Hilbert space?

To illustrate the difference, we consider only the simpler case with $\mathbf{B} = 0$. A more detailed discussion using the Dirac electron as an example can be found in appendix A. For the original wavepacket, with the help of equations (17), (18), and (20), one has

$$\hbar \dot{\mathbf{r}}_{\rm c} = \left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{k}_{\rm c}} \right\rangle + e\mathbf{E} \times \langle \mathcal{F} \rangle \,, \tag{53}$$

where the angular bracket performs the spinor average. On the other hand, the other wavepacket is assumed to live in the *full* Hilbert space of the *effective* theory with $H_{\rm eff}(\mathbf{r}) = H_0 - e\phi(\mathbf{r}) + e\mathbf{E}\cdot\boldsymbol{\mathcal{R}}$. As we have just explained in the previous subsection, the Berry curvature of a full Hilbert space is always zero. Therefore, one can choose a trivial gauge for the Berry connection inside the bracket in equation (18) and obtain

$$\hbar \dot{\mathbf{r}}_a = \left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{k}_a} \right\rangle + e \left\langle \frac{\partial}{\partial \mathbf{k}_a} \mathbf{E} \cdot \boldsymbol{\mathcal{R}} \right\rangle, \tag{54}$$

in which $\mathbf{k}_a = \mathbf{k}_{\rm c}$ when $\mathbf{B} = 0$ (so the subscript will be dropped). The transverse velocities in equation (53) and equation (54) are obviously different.

One can improve the result in equation (54) by including the missing piece, $\langle \dot{\boldsymbol{\mathcal{R}}} \rangle = \langle [\boldsymbol{\mathcal{R}}, \mathcal{H}]/\mathrm{i}\hbar \rangle + \langle \partial \boldsymbol{\mathcal{R}}/\partial t \rangle$. It gives the following correction:

$$\hbar \langle \dot{\boldsymbol{\mathcal{R}}} \rangle = -\mathrm{i}e \, \langle [\boldsymbol{\mathcal{R}}, \mathbf{E} \cdot \boldsymbol{\mathcal{R}}] \rangle - e \left\langle \mathbf{E} \cdot \frac{\partial}{\partial \mathbf{k}} \boldsymbol{\mathcal{R}} \right\rangle. \tag{55}$$

When combined with $\hbar\dot{\mathbf{r}}_a$, it is not difficult to check that one would indeed recover the correct result in equation (53).

For example, the Pauli theory is a valid effective theory of the Dirac theory. They both produce the same positive-energy spectrum. However, starting from the Pauli theory alone, *in the non-relativistic limit* and with $\mathbf{B} = 0$ for simplicity, the velocity of a Pauli wavepacket reads

$$\dot{\mathbf{r}}_a = \frac{\hbar\mathbf{k}}{m} + \frac{e\lambda_c^2}{4\hbar}\langle\boldsymbol{\sigma}\rangle \times \mathbf{E}. \tag{56}$$

On the other hand, for a Dirac wavepacket, one has

$$\dot{\mathbf{r}}_c = \frac{\hbar\mathbf{k}}{m} + \frac{e\lambda_c^2}{2\hbar}\langle\boldsymbol{\sigma}\rangle \times \mathbf{E}. \tag{57}$$

The transverse velocity of the Pauli wavepacket is smaller by a factor of two. That is, the Pauli theory does not always give a correct dynamical prediction *even in the low-velocity limit*. The lesson is that the effective Hamiltonian alone does not specify the effective theory completely. To avoid mistakes, one also needs to know how the physical variables are related to the canonical variables of the theory.

5.3. Dirac theory as an effective theory

At finer length scale or higher energy scale, even the Dirac theory itself is not complete. It needs to be embedded in a larger Hilbert space with more degrees of freedom. One example is the theory of electrons interacting with a quantized electromagnetic field. Another example is the theory for nucleons with inner structures. In these cases, the Dirac theory itself is an effective theory projected from a larger Hilbert space including photons or quarks.

By now, one should be familiar with the fact that such a projection can often generate a gauge structure (due to $\mathcal{R}$ and $\mathcal{F}$) and a magnetic moment (due to $\mathcal{L}$) in the effective theory. Indeed, the anomalous magnetic moment (AMM) of a Dirac particle originates either from the coupling with a quantized electromagnetic field, or from the internal motion of the constituent quarks in a nucleon. It contributes an energy in addition to those in equation (22) [49],

$$H' = a\mu_B\beta\left(\boldsymbol{\Sigma}\cdot\mathbf{B} - \mathrm{i}\boldsymbol{\alpha}\cdot\frac{\mathbf{E}}{c}\right), \tag{58}$$

where $a = g/2 - 1$ and $\boldsymbol{\Sigma}$ is the 4×4 spin matrix. Such an interaction will be treated as a perturbation to H_0.

The perturbation energy is calculated as $E' = \langle W|H'|W\rangle = \sum_{mn}\eta_m^*\langle u_m|H'|u_n\rangle\eta_n$ [29], where $|u_m\rangle$ ($m = 1, 2$) are the positive-energy eigenstates of H_0 (being evaluated at $\mathbf{k}_c$). With the help of the two identities,

$$\langle u_m|\beta\boldsymbol{\Sigma}|u_n\rangle = \left[\boldsymbol{\sigma} - \frac{\lambda_c^2}{\gamma(\gamma+1)}\mathbf{k}_c(\mathbf{k}_c\cdot\boldsymbol{\sigma})\right]_{mn}, \tag{59}$$

$$\mathrm{i}\langle u_m|\beta\boldsymbol{\alpha}|u_n\rangle = \left[\frac{\lambda_c}{\gamma}\boldsymbol{\sigma}\times\mathbf{k}_c\right]_{mn}, \tag{60}$$

one has

$$E' = \frac{a\mu_B}{\gamma mc^2}\hbar\mathbf{k}_c\times\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{E} + a\mu_B\left[\langle\boldsymbol{\sigma}\rangle - \frac{\hbar\mathbf{k}_c\cdot\langle\boldsymbol{\sigma}\rangle\hbar\mathbf{k}_c}{\gamma(\gamma+1)m^2c^2}\right]\cdot\mathbf{B}. \tag{61}$$

Unlike the wavepacket energy in equation (11), the spin–orbit coupling is now explicit from the very beginning (before the Peierls substitution). It indicates that, when the particle has an AMM, the true center of charge $\mathbf{r}_c'$ has been displaced further from the $\mathbf{r}_c$ of a Dirac wavepacket. Such a displacement $\mathbf{R}'$ generates a spin–orbit interaction $e\mathbf{E}\cdot\mathbf{R}'$. By comparing with the spin–orbit energy in equation (61), we obtain $\mathbf{R}' = a(\lambda_c^2/2\gamma)\mathbf{k}_c\times\langle\boldsymbol{\sigma}\rangle$. Indeed, if one simply replaces the $\mathbf{r}_c$ and $e/2m$ in equation (11) with $\mathbf{r}_c' = \mathbf{r}_c + \mathbf{R}'$ and $eg/4m$ respectively, then the energy E' above can be exactly reproduced.

The equations of motion for a particle with AMM will not be reported in detail. A short discussion on its spin precession can be found in appendix B.

Following the same re-quantization scheme using the generalized Peierls substitution, we can again obtain the relativistic Pauli Hamiltonian with the AMM correction. First, by combining the energies in equations (42) and (61), the total energy becomes

$$\begin{aligned}E(\mathbf{r}_c,\mathbf{k}_c) = {}& \gamma mc^2 - e\phi(\mathbf{r}_c) + \mu_B\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{B}\left(\frac{1}{\gamma^2}+a\right)\\ &+ \frac{a\mu_B}{\gamma mc^2}\hbar\mathbf{k}_c\times\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{E}\\ &+ \frac{\mu_B}{m^2c^2}\hbar\mathbf{k}_c\cdot\langle\boldsymbol{\sigma}\rangle\hbar\mathbf{k}_c\cdot\mathbf{B}\left[\frac{1}{\gamma^2(\gamma+1)} - \frac{a}{\gamma(\gamma+1)}\right],\end{aligned} \tag{62}$$

in which $\gamma = \gamma(\mathbf{k}_c)$. Then one can use the generalized Peierls substitution to write the energy in canonical variables. Such a result is a generalization of equation (44). Promoting the canonical variables to quantum conjugate variables, we finally obtain the following relativistic Pauli Hamiltonian:

$$\begin{aligned}\mathcal{H}(\mathbf{r},\mathbf{p}) = {}& \gamma mc^2 - e\phi(\mathbf{r}) + \mu_B\boldsymbol{\sigma}\cdot\mathbf{B}\left(\frac{1}{\gamma}+a\right)\\ &+ \frac{\mu_B}{\gamma mc^2}\boldsymbol{\pi}\times\boldsymbol{\sigma}\cdot\mathbf{E}\left(\frac{1}{1+\gamma}+a\right)\\ &- \frac{a\mu_B}{m^2c^2}\frac{\boldsymbol{\pi}\cdot\boldsymbol{\sigma}\boldsymbol{\pi}\cdot\mathbf{B}}{\gamma(\gamma+1)},\end{aligned} \tag{63}$$

in which $\gamma = \gamma(\boldsymbol{\pi})$. This result is an extension of the low-velocity one reported in [64].

We emphasize that the presence of the electric dipole energy indicates that the theory is only an effective one. The position of the particle in the effective theory should differ from the true position by a Berry connection. Of course, in predicting the trajectory of the particle, one always has to refer to its true position. Therefore, in the Dirac theory with $H' = a\mu_B\beta(\boldsymbol{\Sigma}\cdot\mathbf{B} - \mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{E}/c)$, the true position is conjectured to be $\mathbf{r}_D + \mathbf{R}_D$, where $\mathbf{r}_D$ is the usual position operator in Dirac theory, and $\mathbf{R}_D = (a\mu_B/e)(-\mathrm{i}\beta\boldsymbol{\alpha})$ is the displacement required to generate the electric field term in H'. Thus, the true velocity operator of the particle should be

$$\dot{\mathbf{r}}_D + \dot{\mathbf{R}}_D = (1+a)c\boldsymbol{\alpha} - a\beta\frac{\boldsymbol{\pi}}{m} + a^2\frac{\mu_B}{mc}\left(\frac{\mathbf{E}}{c}\times\boldsymbol{\Sigma} + \gamma_5\mathbf{B}\right). \tag{64}$$

$$H = \begin{pmatrix}
E_g + \frac{p^2}{2m_0} & 0 & -\frac{1}{\sqrt{2}}Vp_+ & \sqrt{\frac{2}{3}}Vp_z & \frac{1}{\sqrt{6}}Vp_- & 0 & -\frac{1}{\sqrt{3}}Vp_z & -\frac{1}{\sqrt{3}}Vp_- \\
0 & E_g + \frac{p^2}{2m_0} & 0 & -\frac{1}{\sqrt{6}}Vp_+ & \sqrt{\frac{2}{3}}Vp_z & \frac{1}{\sqrt{2}}Vp_- & -\frac{1}{\sqrt{3}}Vp_+ & \frac{1}{\sqrt{3}}Vp_z \\
-\frac{1}{\sqrt{2}}Vp_- & 0 & -P-Q & -L & -M & 0 & \frac{1}{\sqrt{2}}L & \sqrt{2}M \\
\sqrt{\frac{2}{3}}Vp_z & -\frac{1}{\sqrt{6}}Vp_- & -L^* & -P+Q & 0 & -M & -\sqrt{2}Q & -\sqrt{\frac{3}{2}}L \\
\frac{1}{\sqrt{6}}Vp_+ & \sqrt{\frac{2}{3}}Vp_z & -M^* & 0 & -P+Q & L & -\sqrt{\frac{2}{3}}L^* & \sqrt{2}Q \\
0 & \frac{1}{\sqrt{2}}Vp_+ & 0 & -M^* & L^* & -P-Q & -\sqrt{2}M^* & \frac{1}{\sqrt{2}}L^* \\
-\frac{1}{\sqrt{3}}Vp_z & -\frac{1}{\sqrt{3}}Vp_- & \frac{1}{\sqrt{2}}L^* & -\sqrt{2}Q & -\sqrt{\frac{3}{2}}L & -\sqrt{2}M & -\Delta - \frac{\gamma_1 p^2}{2m_0} & 0 \\
-\frac{1}{\sqrt{3}}Vp_+ & \frac{1}{\sqrt{3}}Vp_z & \sqrt{2}M^* & -\sqrt{\frac{3}{2}}L^* & \sqrt{2}Q & \frac{1}{\sqrt{2}}L & 0 & -\Delta - \frac{\gamma_1 p^2}{2m_0}
\end{pmatrix}, \tag{65}$$

The magnitude of the anomaly a for an electron is of the order of 10^{-3} and might be too small for such a difference to be observed. However, the anomaly for a nucleon is of order one. Therefore, it is possible to observe and verify such a deviation in experiments using relativistic nucleon beams[8].

6. Applications to semiconductor physics

We now turn our attention to an important application of the semiclassical approach. Using the eight-band Kane model as the parent theory, different gauge structures in different subspaces of the Kane model will be studied [16]. It is shown that the Berry curvature in the conduction band gives the wavepacket a spin-dependent transverse velocity *even in bulk materials*. A similar situation exists in valence bands and produces the spin Hall effect predicted by Murakami *et al* [11, 12]. We have also calculated the self-rotating angular momenta of the wavepackets in the conduction band and the valence band, from which one can obtain the anomalous g-factors. Finally, the effective Hamiltonians for different subspaces will be studied.

6.1. Berry connection in the Kane model

Starting with the eight-band Kane model, we study the wavepacket dynamics in its subspaces, such as the conduction band, the heavy hole–light hole (HH–LH) complex, and the spin–orbit split-off (SO) band (see figure 4). Each band has a twofold spin degeneracy. Assuming the fundamental gap is located at $\mathbf{k} = 0$, then for a small k one can expand the states as $u(\mathbf{k}) = \sum_j c_j u_j(0)$. The Schrödinger equation, $Hu(\mathbf{k}) = Eu(\mathbf{k})$, has the following Hamiltonian [23]:

(see equation (65) given above)

[8] However, when the velocity is projected to the subspace of a particle with positive energy, one has $Pc\boldsymbol{\alpha}P = \boldsymbol{\pi}/(\gamma m)$. It would cancel with the term $P\beta\boldsymbol{\pi}/mP = \boldsymbol{\pi}/m$ to leading order of the velocity. Therefore, the leading (field-free) correction is of the order of $a(v/c)^2$ only.

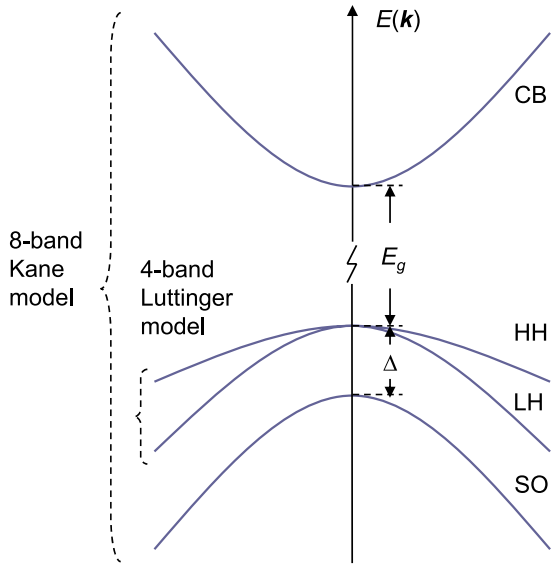

Figure 4. Schematic plot of the semiconductor band structure near the fundamental gap.

in which E_g is the fundamental gap, Δ is the spin–orbit split-off gap, $\mathbf{p} = \hbar\mathbf{k}$, and

$$P = \frac{\gamma_1}{2m_0}p^2, \qquad Q = \frac{\gamma}{2m_0}(p_\parallel^2 - 2p_z^2),$$
$$L = -\frac{\sqrt{3}\gamma}{m_0}p_z p_-, \qquad M = -\frac{\sqrt{3}\gamma}{2m_0}p_-^2, \tag{66}$$

$$V = \frac{\hbar}{m_0}\langle S|\hat{p}_x|X\rangle. \tag{67}$$

The matrix elements are ordered from higher energy to lower energy from the upper-left corner to the lower-right corner. That is, the upper-left 2×2 matrix is for the conduction electron. Simple truncation (projection) to the 4×4 matrix in the middle gives us the four-band Luttinger model for the HH–LH complex; the lower-right 6×6 matrix is for the six-band Luttinger model including the SO band. For demonstration, a simpler Hamiltonian with $\gamma_2 = \gamma_3 \equiv \gamma$ is chosen. Also, we do not distinguish between the usual band parameters and the reduced band parameters.

One first needs to obtain the basic quantities $\mathcal{R}$, $\mathcal{F}$, and $\mathcal{L}$. Because of the anomalous velocity (the second term on the RHS) in equation (18), a non-zero Berry curvature in the conduction band or the valence band arising from the projection would immediately imply a transverse motion of the carrier that resembles the Hall effect. In this subsection, we focus first on the Berry connection $\mathcal{R}$ defined in equation (4), in which $|u_n\rangle$ are the eigenstates of the above Hamiltonian. It is unlikely for one to obtain these eigenstates by analytic diagonalization. Therefore, an alternative approach given in appendix C is used. It allows one to obtain the small-k limit of the Berry connection.

Instead of going through the details, we just show the result for the Berry connection of conduction electron. The result correct to the order of k^1 is[9]

$$\mathcal{R} = \frac{V^2}{3}\left[\frac{1}{E_g^2} - \frac{1}{(E_g+\Delta)^2}\right]\boldsymbol{\sigma}\times\mathbf{k}. \tag{68}$$

Therefore, the dipole term $e\mathbf{E}\cdot\mathcal{R}$ becomes

$$H_{so} = e\mathbf{E}\cdot\mathcal{R} = \alpha\mathbf{E}\cdot\boldsymbol{\sigma}\times\mathbf{k}, \tag{69}$$

where $\alpha \equiv (eV^2/3)[1/E_g^2 - 1/(E_g+\Delta)^2]$. This accounts for the spin–orbit coupling of a conduction electron. It requires neither bulk-inversion asymmetry, as in Dresselhaus coupling [65], nor structure-inversion asymmetry, as in Rashba coupling [66]. However, unlike the Rashba coupling in asymmetric quantum wells, in which the interface electric field plays a crucial role, here one needs to supply an *external* electric field E. When there is no spin–orbit split-off ($\Delta = 0$), the coefficient α vanishes as expected.

In addition, the Berry connections for the HH–LH complex and the split-off band are shown in table 1. They have the same functional form as the one in equation (68), but differ in coefficients and the dimensions of spin matrices. Notice that the Berry connection for the HH–LH complex in table 1 looks very different from the one in [11] (see equation (D.1) in appendix D). The parent theory there is the four-band Luttinger model, instead of the eight-band Kane model. The Berry curvature generated from the $\mathcal{R}$ in equation (D.1) in fact is zero, indicating a 'flat' space (see the discussion near the end of section 5.1). On the other hand, the Berry connection for the HH–LH complex in table 1 yields a non-zero Berry curvature. Because of equation (52), this gives a correction to the result in [11] (see appendix D).

6.2. Berry curvature, orbital moment, and equations of motion

The Berry curvature can be calculated from equation (5) once the Berry connection is known, while the angular momentum is most easily calculated from equation (51). Notice that, in order to calculate $\mathcal{L}$, the knowledge of off-diagonal $\mathcal{R}$ is required (see equation (C.5) in appendix C). Here we simply show the result. For the conduction band, to the lowest non-trivial order, one has

$$\mathcal{F} = \frac{2V^2}{3}\left[\frac{1}{E_g^2} - \frac{1}{(E_g+\Delta)^2}\right]\boldsymbol{\sigma}, \tag{70}$$

[9] This result remains valid up to a $SU(2)$ gauge rotation.

Table 1. Berry connection, Berry curvature, and orbital angular momentum of the *electron* wavepacket in three disjoint subspaces of the eight-band Kane model. Only the leading order (in k) terms are shown. E_g and Δ are the energy gaps in figure 4, $V = \hbar\langle S|p_x|X\rangle/m_0$, and $\mathcal{J}$ is the spin-3/2 matrix operator.

	Conduction band	HH–LH complex	Split-off band
$\mathcal{R}$	$\frac{V^2}{3}[\frac{1}{E_g^2} - \frac{1}{(E_g+\Delta)^2}]\boldsymbol{\sigma}\times\mathbf{k}$	$-\frac{V^2}{3E_g^2}\mathcal{J}\times\mathbf{k}$	$-\frac{V^2}{3}\frac{1}{(E_g+\Delta)^2}\boldsymbol{\sigma}\times\mathbf{k}$
$\mathcal{F}$	$\frac{2V^2}{3}[\frac{1}{E_g^2} - \frac{1}{(E_g+\Delta)^2}]\boldsymbol{\sigma}$	$-\frac{2V^2}{3E_g^2}\mathcal{J}$	$-\frac{2V^2}{3}\frac{1}{(E_g+\Delta)^2}\boldsymbol{\sigma}$
$\mathcal{L}$	$-\frac{2m_0V^2}{3\hbar}(\frac{1}{E_g} - \frac{1}{E_g+\Delta})\boldsymbol{\sigma}$	$-\frac{2m_0}{3\hbar}\frac{V^2}{E_g}\mathcal{J}$	$-\frac{2m_0}{3\hbar}\frac{V^2}{E_g+\Delta}\boldsymbol{\sigma}$

and

$$\mathcal{L} = -\frac{2m_0V^2}{3\hbar}\left(\frac{1}{E_g} - \frac{1}{E_g+\Delta}\right)\boldsymbol{\sigma}. \tag{71}$$

For a semiconductor electron, the E_0 in equation (11) should have already included the Zeeman energy from the *spin* magnetic moment. Through the Zeeman term in equation (11), the orbital magnetic moment generated from equation (71) contributes an extra g-factor,

$$\delta g = -\frac{4}{3}\frac{m_0V^2}{\hbar^2}\left(\frac{1}{E_g} - \frac{1}{E_g+\Delta}\right). \tag{72}$$

This is of course the familiar anomalous g-factor of the conduction electron. This result gives the clearest possible interpretation of the textbook formula [67]: the anomalous g-factor in solid is indeed a result of the self-rotating motion of the electron wavepacket.

The Berry curvatures and orbital angular momenta for the other two wavepacket subspaces can be found in table 1.

Finally, we selectively comment on the *electron* dynamics in the HH–LH complex using the semiclassical equations of motion. For brevity, the subscripts c in $\mathbf{r}_c$ and $\mathbf{k}_c$ are dropped. In order to focus on the effect introduced by the applied field, we set $\gamma = 0$ to avoid the complication from internal spin–orbit coupling. Therefore, this is only a toy-model calculation. Whenever necessary, one can always numerically solve the coupled dynamical equations with realistic parameters without difficulty.

In the simplest case, one applies only a constant electric field ($\mathbf{B} = 0$). The anomalous velocity in equation (18) due to the Berry curvature is $-2eV^2/(3E_g^2)\mathbf{E}\times\mathbf{J}$, where $\mathbf{J} = \langle\mathcal{J}\rangle$. Being transverse to the electric field and spin dependent, it is in fact a spin Hall current.

From the equation of spin precession in equation (21), one obtains

$$\dot{\mathbf{J}} = \frac{\kappa\mu_B/\hbar}{E_g/m_0c^2}\mathbf{B}_{eff}\times\mathbf{J}, \tag{73}$$

where $\kappa\mu_B \equiv eV^2/(3\hbar E_g)$, and $\mathbf{B}_{eff} = \mathbf{E}\times\hbar\mathbf{k}/(m_0c^2)$. The spin simply precesses around the effective magnetic field in its own reference frame.

In the presence of a constant magnetic field ($\mathbf{E} = 0$), the dynamics of the HH–LH wavepacket is slightly more complicated [68]. To linear order of the field, the quasi-momentum decouples from the other degrees of freedom, $\hbar\dot{\mathbf{k}} = (-e/\hbar)\partial E_0/\partial\mathbf{k}\times\mathbf{B}$, where $E_0 = -\gamma_1\hbar^2k^2/2m_0$. The orbit

thus lies at the intersection of the Fermi surface and a plane perpendicular to the magnetic field. The component $\mathbf{k}_\parallel$ parallel to the magnetic field is a constant of motion. For a circular orbit, the frequency of rotation is the usual cyclotron frequency $\omega_c = eB\gamma_1/m_0$. The result so far is not different from the usual cyclotron motion without a Berry curvature [69].

The equation of spin precession is

$$\dot{\mathbf{J}} = \omega_V \left[\hat{\mathbf{B}} + \delta\mathbf{B}_\parallel(\mathbf{k}) + \delta\mathbf{B}_\perp(\mathbf{k})\right] \times \mathbf{J}, \tag{74}$$

where $\omega_V \equiv 2\kappa\mu_B B/\hbar$, $\delta\mathbf{B}_\parallel = -(E_0/E_g)\hat{\mathbf{k}}_\perp \times (\hat{\mathbf{k}}_\perp \times \hat{\mathbf{B}})$, and $\delta\mathbf{B}_\perp = -(E_0/E_g)\hat{\mathbf{k}}_\parallel \times (\hat{\mathbf{k}}_\perp \times \hat{\mathbf{B}})$, in which $\hat{\mathbf{B}}$ and $\hat{\mathbf{k}}_{\parallel,\perp}$ are unit vectors. The correction terms are negligible when the carrier is near the band edge (where $E_0 \simeq 0$).

After some inspection, it can be found that the correction $\delta\mathbf{B}_\perp$ rotates at the cyclotron frequency ω_c. On the other hand, $\delta\mathbf{B}_\parallel$ is proportional to $\mathbf{B}\sin^2\theta(t)$, where $\theta(t)$ is the angle between $\mathbf{B}$ and $\mathbf{k}(t)$. It can be steady or oscillating, depending on whether the orbit is circular or non-circular. In the simplest case of a circular orbit, the effective magnetic field is rotating around $\mathbf{B}$ with a fixed magnitude, resembling that in the electron spin resonance. Therefore, the spin would flip back and forth with a frequency ω_c. This is a natural consequence of combining the circular motion with the spin–orbit coupling.

6.3. Effective Hamiltonians

We can use the generalized Peierls substitution in equation (40) to obtain the effective Hamiltonians. The canonical Hamiltonian in equation (41) then has the following form for the conduction band:

$$H(\mathbf{r},\mathbf{p}) = E_0(\boldsymbol{\pi}) - e\phi(\mathbf{r}) + \alpha\mathbf{E}\cdot\boldsymbol{\sigma}\times\boldsymbol{\pi} + \delta g\mu_B\mathbf{B}\cdot\frac{\hbar\boldsymbol{\sigma}}{2}, \tag{75}$$

where E_0 includes the Zeeman energy from the bare spin, α is given below equation (69), and δg is given in equation (72). The effective Hamiltonian for the spin–orbit split-off band has a similar form, but with $\alpha = -eV^2/[3(E_g+\Delta)^2]$ and $\delta g = -4m_0V^2/[3\hbar^2(E_g+\Delta)]$.

For a wavepacket in the HH–LH complex, one has

$$H(\mathbf{r},\mathbf{p}) = E_0(\boldsymbol{\pi},\boldsymbol{\mathcal{J}}) - e\phi(\mathbf{r}) + \alpha\mathbf{E}\cdot\boldsymbol{\mathcal{J}}\times\boldsymbol{\pi} + \delta g\mu_B\mathbf{B}\cdot\hbar\boldsymbol{\mathcal{J}}, \tag{76}$$

where $E_0(\boldsymbol{\pi},\boldsymbol{\mathcal{J}})$ is the four-band Luttinger Hamiltonian, $\alpha = -eV^2/(3E_g^2)$, and $\delta g = -2m_0V^2/(3\hbar^2E_g)$. These results are in agreement with those in [23] obtained using the Löwdin partition.

The semiclassical approach is more than an alternative derivation of the effective Hamiltonians. It tells an intelligible story behind the formalism. The various interaction terms in the effective Hamiltonian have clear interpretations because of the direct connection with $\mathcal{R}$ and $\mathcal{L}$. Furthermore, these quantities can immediately be integrated with the equations of motion to study transport properties. Comparing with other approaches, such as the theory of linear response, this formulation offers a particularly useful visual guidance when the dynamics of charge and spin are coupled.

7. Conclusions

The wavepacket dynamics in a solid can successfully explain various phenomena, such as the Bloch oscillation and the de Haas–van Alphen effect [69]. In the early primitive version, the particle velocity is simply the group velocity of the wavepacket. Over the years, people gradually realized that the Berry curvature of Bloch states can have important effect on the electron transport. It contributes to an anomalous velocity that is crucial to the anomalous Hall effect [3, 8]. Subsequently, the (one-band) semiclassical formalism, with the Berry curvature included, was developed and applied successfully to the quantum Hall effect [6, 7]. A more general framework, allowing, e.g., adiabatic crystal deformation was then advanced [24]. Generalization of this formalism to the multi-band case makes it capable of investigating spintronic transport involving spin or quasi-spin degrees of freedom [29, 30].

Based on the semiclassical theory, one can use either the canonical quantization or other methods of quantization to obtain an effective quantum theory. This review focuses on the recent progress regarding the canonical quantization of the multi-band semiclassical theory. Because of this progress, the two-way passage between semiclassical and quantum theories is now achieved. We explain in great detail how to use this new recipe of quantization with the generalized Peierls substitution. It is applied successfully to the theories of Dirac electrons and semiconductor carriers.

It is necessary to include the Berry curvature and orbital moment in order to account for physical effects to first order in external fields. Furthermore, these gauge-covariant quantities are also sufficient for building a correct quantum theory. Although this route for an effective quantum theory is less direct, it is physically more transparent than other approaches. Most importantly, an architecture that is rich in concepts is revealed along the way. For example, we now understand the deep connection between the Berry connection and the spin–orbit interaction and the origin of the anomalous g-factor, to name only two. The wavepacket point of view is certainly both intuitive and productive.

The re-quantized effective theory can have its own semiclassical theory, which in turn can be re-quantized. Therefore, a whole series of effective theories can be generated. We study the connections between these hierarchical theories, and explain how the gauge-covariant quantities, $\mathcal{F}$ and $\mathcal{L}$, can be renormalized from one level to another. We now have a clear overview of these hierarchical theories. Such a fundamental understanding certainly is not restricted to solid state physics. It can be valuable to physical theories in other disciplines.

Most of the topics covered in this review have close connections with the field of spintronics, which has been vigorously pursued in recent years. In the past, the Bloch energy has played a key role in *ab initio* calculations. The values are calculated and tabulated for all kinds of crystalline materials. These results are useful for many purposes, but are incomplete in light of this work. The other two essential quantities, $\mathbf{F}$ and $\mathbf{L}$, should also be calculated and tabulated

systematically. Such an endeavor would be very helpful for researchers to study more intricate electronic properties [70]. This subject is still at its dawning stage. We expect to see more exciting discoveries ahead of us.

Acknowledgments

The authors would like to thank D Culcer, C P Chuu, R Winkler, D Xiao, and W Yao for valuable discussions. They also thank the referee for pointing out the connection between this work and the electric polarization induced by spin current, which is added near the end of section 3.3. MCC was supported by the NSC of Taiwan (NSC 96-2112-M-003-010-MY3); QN was supported by DOE (DE-FG03-02ER45958), NSF (DMR-0404252/0606485), the Welch Foundation, and NSFC (No 10740420252).

Appendix A. The Pauli wavepacket

Starting from the non-relativistic Pauli Hamiltonian instead of the Dirac Hamiltonian, one can study the energy and dynamics of the Pauli wavepacket. The Hamiltonian can be expanded as

$$H = H_0 + \Delta H + H' = \frac{1}{2m}(\mathbf{p} + e\mathbf{A}_a)^2 - e\phi_a + \frac{e}{2m}\mathbf{L}\cdot\mathbf{B} + e\mathbf{E}\cdot(\mathbf{r}-\mathbf{r}_a) \tag{A.1}$$

$$+ \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B} + \frac{e\hbar}{4m^2c^2}\boldsymbol{\sigma}\cdot\mathbf{E}\times(\mathbf{p}+e\mathbf{A}_a)\,. \tag{A.2}$$

It consists of the unperturbed Hamiltonian H_0 (with two degenerate bands), the gradient correction ΔH, and the degeneracy-lifting perturbation H'. The third part H' has been mentioned in section 2 (footnote 3). Its influence on the equations of motion can be found in [29].

The Pauli Hamiltonian looks lengthier than the Dirac Hamiltonian. However, the eigenstates $|u_i\rangle$ of H_0 are simply the unit spinors, $(1,0)^{\mathrm{T}}$ and $(0,1)^{\mathrm{T}}$. Therefore, based on equations (4), (5), and (8), one has the trivial result with $\mathcal{R}=0$, $\mathcal{F}=0$, and $\mathcal{L}=0$. This simplifies the equations of motion significantly [29],

$$\hbar\dot{\mathbf{k}}_a = -e\mathbf{E} - e\dot{\mathbf{r}}_a\times\mathbf{B}, \tag{A.3}$$

$$\hbar\dot{\mathbf{r}}_a = \frac{\partial E_0}{\partial\mathbf{k}_a} + \boldsymbol{\eta}^\dagger\frac{\partial\mathcal{H}'}{\partial\mathbf{k}_a}\boldsymbol{\eta}, \tag{A.4}$$

$$\mathrm{i}\hbar\dot{\boldsymbol{\eta}} = \mathcal{H}'\boldsymbol{\eta}, \tag{A.5}$$

where $E_0 = \hbar^2k_a^2/2m$.

Substituting $\mathcal{H}' = \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B} + \frac{e\hbar^2}{4m^2c^2}\boldsymbol{\sigma}\cdot\mathbf{E}\times\mathbf{k}_a$ into the equations above, one immediately has the following equations of motion:

$$\hbar\dot{\mathbf{k}}_a = -e\mathbf{E} - e\dot{\mathbf{r}}_a\times\mathbf{B}, \tag{A.6}$$

$$\dot{\mathbf{r}}_a = \frac{\hbar\mathbf{k}_a}{m} + \frac{e\hbar}{4m^2c^2}\langle\boldsymbol{\sigma}\rangle\times\mathbf{E}, \tag{A.7}$$

$$\mathrm{i}\hbar\dot{\boldsymbol{\eta}} = \frac{e\hbar}{2m}\left(\boldsymbol{\sigma}\cdot\mathbf{B} + \frac{\hbar}{2mc^2}\boldsymbol{\sigma}\cdot\mathbf{E}\times\mathbf{k}_a\right)\boldsymbol{\eta}. \tag{A.8}$$

To compare these equations with those of the Dirac wavepacket in equations (28), (29), and (30), one needs the following two relations (see equation (40); $\mathbf{G}$ is neglected):

$$\begin{aligned}\mathbf{r}_{\mathrm{c}} &= \mathbf{r}_a + \mathbf{R},\\ \hbar\mathbf{k}_{\mathrm{c}} &= \hbar\mathbf{k}_a + e\mathbf{B}\times\mathbf{R}.\end{aligned} \tag{A.9}$$

One can check that the Lorentz force equation in equation (A.6) remains valid in either set of variables, $(\mathbf{r}_a,\mathbf{k}_a)$ or $(\mathbf{r}_{\mathrm{c}},\mathbf{k}_{\mathrm{c}})$. To linear order of fields, one can replace the $\mathbf{k}_a$ in equation (A.8) by $\mathbf{k}_{\mathrm{c}}$. Therefore, it also gives the same (non-relativistic) BMT equation as given in equation (30).

There are two main differences between equation (29) and equation (A.7). First, the velocity transverse to the electric field in equation (A.7) is smaller by a factor of two if one does not distinguish between $\mathbf{r}_{\mathrm{c}}$ and $\mathbf{r}_a$. Second, there is no magnetic-field-related velocity in equation (A.7). If the shift of position due to the Berry connection is taken into account, then in the non-relativistic limit one has

$$\dot{\mathbf{R}} = \frac{e\lambda_{\mathrm{c}}^2}{4\hbar}\langle\boldsymbol{\sigma}\rangle\times\mathbf{E} + \frac{e\lambda_{\mathrm{c}}^2}{4m}\mathbf{B}\times(\mathbf{k}_a\times\langle\boldsymbol{\sigma}\rangle). \tag{A.10}$$

The first term compensates the transverse velocity missing in equation (A.7) (see section 5.2). The second term would cancel with an extra term from the $e\mathbf{B}\times\mathbf{R}$ shift in momentum. Therefore, equation (A.7) gives

$$\dot{\mathbf{r}}_{\mathrm{c}} = \frac{\hbar\mathbf{k}_{\mathrm{c}}}{m} + \frac{e\hbar}{2m^2c^2}\langle\boldsymbol{\sigma}\rangle\times\mathbf{E}. \tag{A.11}$$

The $\mathbf{B}\cdot\mathbf{F}$-correction in equation (29) cannot be reproduced even if the R-shift has been considered. Such a correction amounts to a spin-dependent mass renormalization, which is not included in the Pauli equation from the beginning.

Appendix B. Spin precession for a particle with AMM

As a reference and without going into details, we show the spin precession equation for a particle with $a\neq 0$,

$$\langle\dot{\boldsymbol{\sigma}}\rangle = \frac{e}{\gamma mc}\left[(1+\gamma a)\mathbf{B} - a\frac{(\hbar\mathbf{k}_{\mathrm{c}}\cdot\mathbf{B})\hbar\mathbf{k}_{\mathrm{c}}}{(\gamma+1)m^2c^2} + \frac{1}{mc^2}\left(\frac{1}{\gamma+1}+a\right)\mathbf{E}\times\hbar\mathbf{k}_{\mathrm{c}}\right]\times\langle\boldsymbol{\sigma}\rangle. \tag{B.1}$$

If the particle is moving with the constraint $\boldsymbol{\beta}\cdot\mathbf{B} = \boldsymbol{\beta}\cdot\mathbf{E} = 0$, where $\boldsymbol{\beta} = \mathbf{v}/c = \hbar\mathbf{k}_{\mathrm{c}}/\gamma mc$, then the spin precession follows $\langle\dot{\boldsymbol{\sigma}}\rangle = \boldsymbol{\omega}_s\times\langle\boldsymbol{\sigma}\rangle$, in which

$$\boldsymbol{\omega}_s = \frac{e}{m}\left[\frac{\mathbf{B}}{\gamma} - \frac{1}{\gamma+1}\frac{\boldsymbol{\beta}}{c}\times\mathbf{E} + a\left(\mathbf{B} - \frac{\boldsymbol{\beta}}{c}\times\mathbf{E}\right)\right]. \tag{B.2}$$

On the other hand, the velocity vector orbits at a slightly different frequency, $\dot{\boldsymbol{\beta}} = \boldsymbol{\omega}_{\mathrm{c}}\times\boldsymbol{\beta}$, in which

$$\boldsymbol{\omega}_{\mathrm{c}} = \frac{e}{m}\left[\frac{\mathbf{B}}{\gamma} - \frac{\gamma}{\gamma^2-1}\frac{\boldsymbol{\beta}}{c}\times\mathbf{E}\right]. \tag{B.3}$$

Unlike the frequency of spin precession, such a frequency is independent of a. It is not difficult to find that their difference, $\boldsymbol{\omega}_a = \boldsymbol{\omega}_s - \boldsymbol{\omega}_{\mathrm{c}}$, is independent of the electric field if $\gamma_{\mathrm{magic}}^2 - 1 = 1/a$. This fact can be used on a beam line to eliminate the unwanted effect from the confining electric field [71].

Appendix C. Calculation of the Berry connection

We explain briefly the approach used in calculating the Berry connection of the conduction band. The connections for other subspaces can be calculated in a similar manner. For the Kane model in equation (65), we can separate the Hamiltonian into the following three parts:

$$H(\mathbf{k}) = H_0 + H_1^D(\mathbf{k}) + H_1^{\mathrm{OD}}(\mathbf{k}) = \begin{pmatrix} h_0 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} h_1 & 0 \\ 0 & h_1' \end{pmatrix} + \begin{pmatrix} 0 & M \\ M^\dagger & 0 \end{pmatrix}, \quad \text{(C.1)}$$

where h_0 is the part of the conduction-band Hamiltonian when $\mathbf{k} = 0$. The rest of the Hamiltonian is separated into H_1^D with only diagonal blocks, plus H_1^{OD} with only off-diagonal blocks.

The Hamiltonian can be diagonalized using a unitary transformation,

$$\begin{aligned} H^D &= \mathrm{e}^{-\mathrm{i}S} H \mathrm{e}^{\mathrm{i}S} \\ &= H + \mathrm{i}[H, S] - \tfrac{1}{2}[S, [S, H]] + \cdots \\ &= H_0 + H_1^D + H_1^{\mathrm{OD}} + \mathrm{i}[H_0, S] + \mathrm{i}[H_1^D, S] \\ &\quad + [H_1^{\mathrm{OD}}, S] - \tfrac{1}{2}[S, [S, H_0]] + \cdots . \end{aligned} \quad \text{(C.2)}$$

One divides the hermitian matrix S into the following form:

$$S = \begin{pmatrix} S_{00} & S_{01} \\ S_{01}^\dagger & S_{11} \end{pmatrix}, \quad \text{(C.3)}$$

where S_{00} has the same rank as h_0. We then demand the off-diagonal blocks in equation (C.2) to vanish to the leading power of k. This enables us to solve the off-diagonal part S_{01}, which has the same power of k as H_1^{OD}. To linear order of S, as far as the Berry curvature is concerned, one need not solve the block-diagonal part S_{00} since it merely re-shuffles the basis states of the wavepacket subspace. Such a gauge rotation would not change the Berry curvature. Once the unitary matrix $U = \mathrm{e}^{\mathrm{i}S}$ is known (to this order), the Berry connection can be calculated using $\boldsymbol{\mathcal{R}} = \mathrm{i}U^\dagger \partial U/\partial \mathbf{k}$.

For example, the upper block-diagonal part of the Berry connection is

$$P_{\mathrm{U}} \boldsymbol{\mathcal{R}} P_{\mathrm{U}} = \frac{\mathrm{i}}{2} S_{01}^\dagger \frac{\partial S_{10}}{\partial \mathbf{k}} + \mathrm{H.c.}, \quad \text{(C.4)}$$

where P_{U} projects to the upper diagonal block. The leading off-diagonal block is ($P_L + P_{\mathrm{U}} = 1$)

$$P_{\mathrm{U}} \boldsymbol{\mathcal{R}} P_L = -\frac{\partial S_{01}}{\partial \mathbf{k}}. \quad \text{(C.5)}$$

Off-diagonal Berry connection is required in the calculation of the orbital angular momentum in equation (51).

Appendix D. Berry connection in the Luttinger model

In an earlier study, the Berry connection in the valence band was investigated based on the four-band Luttinger model [11]. For reference, the result is briefly reviewed here. The Berry connection calculated using the Luttinger model is (θ and ϕ are the angles of $\mathbf{k}$ in the spherical coordinate)

$$\boldsymbol{\mathcal{R}} = \frac{1}{k} \mathcal{J}_y \hat{\theta} - \frac{1}{k} (\mathcal{J}_x - \operatorname{cosec}\theta \mathcal{J}_z) \hat{\phi}, \quad \text{(D.1)}$$

which generates zero Berry curvature, $\boldsymbol{\mathcal{F}} = 0$.

After being projected to the HH (LH) subspace, one has

$$\begin{aligned} P_{\mathrm{HH}} \boldsymbol{\mathcal{R}} P_{\mathrm{HH}} &= \frac{3}{2k} \operatorname{cosec}\theta \sigma_z \hat{\phi}, \\ P_{\mathrm{LH}} \boldsymbol{\mathcal{R}} P_{\mathrm{LH}} &= \frac{1}{2k} \operatorname{cosec}\theta \sigma_z \hat{\phi} - \frac{\sigma_x}{k} \hat{\phi} + \frac{\sigma_y}{k} \hat{\theta}. \end{aligned} \quad \text{(D.2)}$$

It has been emphasized [11] that the Berry connection for the HH states (with $m_j = \pm 3/2$) is diagonal (Abelian) because the vector operator $\mathrm{i}\partial/\partial\mathbf{k}$ cannot connect states differing in angular momentum by more than one. On the other hand, the connection of the LH states is non-Abelian. The *projected* Berry connection generates a non-zero Berry curvature in each of their own subspaces, $\boldsymbol{\mathcal{F}}'_{\mathrm{HH/LH}} = \mp 3/(2k^2)\sigma_z \hat{r}$. They have the same form as a magnetic monopole field in the k-space with a charge $\mp 3/2\sigma_z$.

According to the renormalization relations in equation (52), since the HH–LH complex inherited a Berry connection from the four remote bands in the Kane model, the Berry curvature of the HH (or LH) subspace would obtain a correction,

$$\boldsymbol{\mathcal{F}}_{\mathrm{HH/LH}} = \boldsymbol{\mathcal{F}}'_{\mathrm{HH/LH}} + P_{\mathrm{HH/LH}} \boldsymbol{\mathcal{F}}_{\mathrm{p}} P_{\mathrm{HH/LH}}, \quad \text{(D.3)}$$

where $\boldsymbol{\mathcal{F}}_{\mathrm{p}}$ is the curvature of the HH–LH complex listed in table 1 and $P_{\mathrm{HH/LH}}$ is the projection from four bands to two bands.

References

[1] Berry M V 1984 *Proc. R. Soc.* A **392** 45
[2] Shapere A and Wilczek F (ed) 1989 *Geometric Phases in Physics* (Singapore: World Scientific)
[3] Adams E N and Blount E I 1959 *J. Phys. Chem. Solids* **10** 286
[4] Blount E I 1962 *Solid State Physics* vol 13, ed F Seitz and D Turnbull (New York: Academic)
[5] Thouless D J, Kohmoto M, Nightingale P and den Nijs M 1982 *Phys. Rev. Lett.* **49** 405
Kohmoto M 1993 *J. Phys. Soc. Japan* **62** 659
[6] Chang M C and Niu Q 1995 *Phys. Rev. Lett.* **75** 1348
[7] Chang M C and Niu Q 1996 *Phys. Rev.* B **53** 7010
[8] Karplus R and Luttinger J M 1954 *Phys. Rev.* **95** 1154
[9] Taguchi Y, Oohara Y, Yoshizawa H, Nagaosa N and Tokura Y 2001 *Science* **291** 2573
Jungwirth T, Niu Q and MacDonald A H 2002 *Phys. Rev. Lett.* **88** 207208
Fang Z *et al* 2003 *Science* **302** 92
[10] Zeng C, Yao Y, Niu Q and Weitering H 2006 *Phys. Rev. Lett.* **96** 37204
[11] Murakami S, Nagaosa N and Zhang S C 2003 *Science* **301** 1348
[12] Murakami S, Nagaosa N and Zhang S C 2004 *Phys. Rev.* B **69** 235206
[13] Onoda M, Murakami S and Nagaosa N 2004 *Phys. Rev. Lett.* **93** 83901
Sawada K and Nagaosa N 2005 *Phys. Rev. Lett.* **95** 237402
Onoda M, Murakami S and Nagaosa N 2006 *Phys. Rev.* E **74** 66610
[14] Bliokh K Y *et al* 2006 *Phys. Rev. Lett.* **96** 073903
Bliokh K Y 2006 *Phys. Rev. Lett.* **97** 043901
Duval C *et al* 2006 *Phys. Rev.* D **74** 021701 (R)
[15] Hofstadter D 1976 *Phys. Rev.* B **14** 2239
[16] Chang M C and Niu Q 2008 submitted
[17] Peierls R 1933 *Z. Phys.* **80** 763
[18] Luttinger J M 1951 *Phys. Rev.* **84** 814

[19] Luttinger J M and Kohn W 1955 *Phys. Rev.* **97** 869
[20] Kohn W 1959 *Phys. Rev.* **115** 1460
[21] Roth L M 1962 *J. Phys. Chem. Solids* **23** 433
[22] Foldy L and Wouthuysen S 1950 *Phys. Rev.* **78** 29
[23] Winkler R 2003 *Spin–Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems* (Berlin: Springer)
[24] Sundaram G and Niu Q 1999 *Phys. Rev.* B **59** 14 915
[25] Resta R 1994 *Rev. Mod. Phys.* **66** 899
Resta R 2000 *J. Phys.: Condens. Matter* **12** R107
[26] Bohm A, Mostafazadeh A, Koizumi H, Niu Q and Zwanziger J 2003 *The Geometric Phase in Quantum Systems* (Heidelberg: Springer)
[27] Nenciu G 1991 *Rev. Mod. Phys.* **63** 91
[28] Teufel S 2003 *Adiabatic Perturbation Theory in Quantum Dynamics* (Berlin: Springer)
[29] Culcer D, Yao Y and Niu Q 2005 *Phys. Rev.* B **72** 85110
[30] Shindou R and Imura K-I 2005 *Nucl. Phys.* B **720** 399
[31] Marzari N and Vanderbilt D 1997 *Phys. Rev.* B **56** 12 847
[32] Wilczek F and Zee A 1984 *Phys. Rev. Lett.* **52** 2111
[33] Bliokh K Y 2005 *Europhys. Lett.* **72** 7
[34] Chuu C P, Chang M C and Niu Q 2008 submitted
[35] Baym G 1990 *Lectures on Quantum Mechanics* (Reading, MA: Addison-Wesley)
[36] Mathur H 1991 *Phys. Rev. Lett.* **67** 3325
[37] Shankar R and Mathur H 1994 *Phys. Rev. Lett.* **73** 1565
[38] Huang K 1952 *Am. J. Phys.* **20** 479
[39] Bargmann V, Michel L and Telegdi V L 1959 *Phys. Rev. Lett.* **2** 435
[40] Griffiths D J 1999 *Introduction to Electrodynamics* 3rd edn (Englewood Cliffs, NJ: Prentice-Hall)
Jackson J D 1999 *Classical Electrodynamics* 3rd edn (New York: Wiley)
[41] Penfield P Jr and Haus H A 1967 *Electrodynamics of Moving Media* (Cambridge, MA: MIT Press)
[42] Katsura H, Nagaosa N and Balatsky A V 2005 *Phys. Rev. Lett.* **95** 057205
[43] Frenkel J 1926 *Z. Phys.* **37** 243
also see Bruno P and Dugaev V K 2005 *Phys. Rev.* B **72** 241302(R)
[44] Arnold V I 1989 *Mathematical Methods of Classical Mechanics* 2nd edn (New York: Springer)
Xiao D, Shi J and Niu Q 2005 *Phys. Rev. Lett.* **95** 137204
[45] Sachdev S 1999 *Quantum Phase Transition* (Cambridge: Cambridge University Press) chapter 13
[46] Gosselin P, Bérard A and Mohrbach H 2006 *Preprint* hep-th/0603192
Gosselin P, Hanssen J and Mohrbach H 2006 *Preprint* cond-mat/0611628
[47] Blount E I 1962 *Phys. Rev.* **126** 1636
[48] Yafet Y 1962 *Solid State Physics* vol 13, ed F Seitz and D Turnbull (New York: Academic)
[49] Bjorken J D and Drell S D 1964 *Relativistic Quantum Mechanics* (New York: McGraw-Hill) chapter 4
[50] See equation (33) in Silenko A J 2003 *J. Math. Phys.* **44** 2952
[51] Feschbach H 1958 *Ann. Phys.* **5** 363
[52] Kane E O 1957 *J. Phys. Chem. Solids* **1** 249
[53] For a recent generalization to many-body systems, see Kita T and Arai M 2005 *J. Phys. Soc. Japan* **74** 2813
[54] Blount E I 1962 *Phys. Rev.* **128** 2454
[55] Löwdin P-O 1951 *J. Chem. Phys.* **19** 1396
[56] See, for example Pethick C J and Smith H 2002 *Bose–Einstein Condensation in Dilute Gases* (Cambridge: Cambridge University Press)
[57] Anderson P W 1984 *Basic Notions of Condensed Matter Physics* (Reading, MA: Addison-Wesley) p 194
[58] Rammal R and Bellissard J 1990 *J. Physique* **51** 1803
[59] Chang M C, Yang M F and Hong T M 1997 *Phys. Rev.* B **56** 3602
[60] Avron J E 1995 *Les Houches Proc.* ed E Akkermans, G Montambaux and J L Pichard (Amsterdam: North-Holland)
[61] Simon B 1983 *Phys. Rev. Lett.* **51** 2167
[62] Samuel J 1996 *Phys. Rev. Lett.* **76** 717
[63] Avron J E, Sadun L, Segert J and Simon B 1989 *Commun. Math. Phys.* **124** 595
[64] See p 288 in Brown L S and Gabrielse G 1986 *Rev. Mod. Phys.* **58** 233
[65] Dresselhaus G 1955 *Phys. Rev.* **100** 580
[66] Rashba E I 1960 *Sov. Phys.—Solid State* **2** 1224
Bychkov Y A and Rashba E I 1984 *JETP Lett.* **39** 78
also see Chazalviel J-N 1975 *Phys. Rev.* B **11** 3918
and the nice review by Zutic I, Fabian J and Das Sarma S 2004 *Rev. Mod. Phys.* **76** 323–410
[67] See problem (9.16) in Yu P Y and Cardona M 2003 *Fundamentals of Semiconductors* 3rd edn (Berlin: Springer)
[68] Winkler R 2004 *Phys. Rev.* B **70** 125301
[69] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders)
[70] Yao Y, Kleinman L, MacDonald A H, Sinova J, Jungwirth T, Wang D-s, Wang E and Niu Q 2004 *Phys. Rev. Lett* **92** 037204
Guo G Y, Yao Y and Niu Q 2005 *Phys. Rev. Lett.* **94** 226601
Wang X, Yates J R, Souza I and Vanderbilt D 2006 *Phys. Rev.* B **74** 195118
[71] Carey R M *et al* 1999 *Phys. Rev. Lett.* **82** 1632