Copyright
by
Yang Gao
2016

The Dissertation Committee for Yang Gao
certifies that this is the approved version of the following dissertation:

Semiclassical Dynamics up to Second Order in Uniform Electromagnetic Fields: Theory and Applications

Committee:

Qian Niu, Supervisor

Gregory A.Fiete

Chih-Kang Shih

Allan H.MacDonald

Lorenzo Sadun

Semiclassical Dynamics up to Second Order in Uniform Electromagnetic Fields: Theory and Applications

by

Yang Gao, B.S.,B.S.,M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2016

Dedicated to my parents, Yujun Gao and Sumei Pan.

Acknowledgments

Seven years is a significant period of time in life. Seven years ago, I graduated from Peking University in China, imagining my next stage of life in US. Now I am preparing my dissertation for doctorate degree. At this point, I want to thank this wonderful opportunity granted by physics department at University of Texas at Austin. Austin is such a great place to stay. I met a lot of friendly people who were willing to help me to fit in. I also enjoy the warm winter and lovely spring and fall. The barbecue is so good that after eating some, even summer seems to be good.

More importantly, I am glad to stay in Professor Niu's group for five years and a half. Professor Niu is a wonderful advisor – knowledgeable, kindly, and responsible. Doing research is always challenging. Well-formulated theories I learned from textbook are like Plato's imaginary Utopia, – they are so perfect and beautiful that I sometimes doubted the meaning of my own research and whether or not I can make any progress. At that time, Professor Niu was very helpful and he would use a famous Chinese philosopher Yang-ming Wang's belief of knowledge as action to encourage me. I enjoy numerous hours of discussion with him, which greatly helps me to not only make progress in my research, but also establish real interest in physics. Moreover, I want to especially thank his belief in me.

I want to thank delightful discussions with Xiang Hu, Shengyuan A.Yang, Hua Chen, Ran Cheng, Xiao Li, and Zhenhua Qiao. They helped me to make fewer mistakes in my research. I also want to thank Prof. Allan H.MacDonald and Prof. Shengyuan A.Yang, who kindly agreed to write recommendation letter for my postdoc application.

I want to thank my fiancee Wei, who helped me a lot during my difficult time. I also thank my parents for their support and belief in me. Finally, I want to thank all friends who helped me in these years.

Semiclassical Dynamics up to Second Order in Uniform Electromagnetic Fields: Theory and Applications

Publication No. _____

Yang Gao, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Qian Niu

We derive a complete semiclassical theory up to second order with respect to electromagnetic fields, by establishing the equations of motion for the velocity and the force up to second order and modifying the band energy up to second order. With this theory, we are able to calculate magnetoelectric polarizability, nonlinear anomalous Hall conductivity, and the magnetic susceptibility. Finally, we derive a density quantization scheme which is the generalization of Onsager's rule to any polynomial order, and show its theoretical and experimental implications.

Table of Contents

List of Figures

Chapter 1

Introduction

Ever since their discovery, geometrical concepts such as Berry phase and Berry curvature are more and more important in the study of solid-state physics, since they are related to a vast variety of response functions of Bloch electrons. For example, the Chern number as obtained from the integration of the Berry curvature, yields the quantum Hall conductivity, and opens a new field of defining the states of matter. The Berry curvature is also essential in understanding the spin/valley Hall effect, the anomalous Hall effect, the magnetization, and so on, while the Berry phase finds its application in the adiabatic pumping, Landau level quantization, modern theory of polarization, and so on.

Among all those theories that use the Berry curvature and Berry phase, the semiclassical theory is a particularly simple but powerful one. It interprets the carrier in solids as a wave-packet of certain Bloch states, sharply localized in k-space, and uses the center of mass position and gauge-invariant crystal momentum of the wave-packet as the only dynamical variables [1] required to describe the evolution of the wave-packet. The dynamics of this position and

[1] A new variable representing the internal degree of freedom may also appear if the wave-packet consists of Bloch states from multiple bands.

momentum is then obtained through the Euler-Lagrange equation from the Lagrangian of the wave-packet. Compared with the textbook equations of motion, an anomalous velocity is identified, manifesting as the cross product of the force and the Berry curvature. Together with a correction to the electronic density of states, it suggests a non-canonical structure in the electronic dynamics. Meanwhile, the electronic band energy is also modified by a Zeeman-type energy due to the orbital magnetic moment, which completes the semiclassical description of Bloch electrons up to first order in the electromagnetic field. This framework is well exemplified in various scenarios, as summarized in [84].

However, one will need a semiclassical theory up to second order to derive response functions such as the magnetoelectric polarizability, the magnetic susceptibility, the non-linear anomalous Hall conductivity, and so on. The central work in this dissertation is thus how to accurately construct a wave-packet for this semiclassical theory up to second order, and how to use this theory to calculate the above mentioned response functions.

Our work is organized as follows. In Chapter 2, the semiclassical theory up to second order is derived. One can see that the form of the equations of motion stays the same, while the Berry curvature and the band energy are corrected up to first and second order, respectively. In Chapter 3, the magnetoelectric coupling is derived, as an example to demonstrate the validity of our theory. Moreover, the non-linear anomalous Hall conductivity is derived, which can compete with the ordinary Hall effect in relatively dirty materials.

In Chapter 4, the semiclassical theory is used to study static magnetic susceptibility from electronic orbital motion, and it offers a fresh understanding to various mechanisms behind this complicated response function. In Chapter 5, the Onsager's rule is generalized, to incorporate not only the Berry phase and magnetic moment correction, but also higher order magnetic responses, such as the magnetic susceptibility. This generalized quantization scheme is due to the quantization of semiclassical electron density, and have promising utility both theoretically and experimentally.

Chapter 2

Semiclassical Theory up to Second Order

In this chapter, I will first review the background for the response functions of Bloch electrons and the semiclassical theory to first order to calculate it. Then I will present the semiclassical theory up to second order. [1]

2.1 General Background

We are interested in the response functions of Bloch electrons to small uniform external electromagnetic fields. The Hamiltonian for this problem generally has the following form (we chose e and $\hbar$ to be unity):

$$\hat{H} = \hat{H}_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{q}) + \boldsymbol{E} \cdot \boldsymbol{q}\,. \tag{2.1}$$

The difficulty to solve the above problem lies in the fact that direct perturbation to the above Hamiltonian using the scalar and vector potential is invalid, since those potentials are essentially unbounded.

There are many ways to avoid this difficulty, such as the standard techniques to calculate response functions in the linear response theory [39,

[1] This semiclassical theory up to second order part (section 3,4,5) is based on the formalism part in *Y.Gao, S. A. Yang, and Q. Niu, Phys. Rev. Lett.112, 166601 (2014)* and *Y. Gao, S. A. Yang, and Q. Niu, Phys. Rev. B 91, 214405 (2015)* . This general formalism of the second order semiclassical theory is derived by me.

71, 25, 64, 65, 24, 78], the spatial perturbation technique [1, 29, 45, 68, 80, 5, 80, 48], and the semiclassical theory to be mentioned below. Among them, we want to discuss a framework based on phase space quantum mechanics at first, since it has a particularly simple physical structure (similar to Foldy-Wouthuysen transformation in relativistic physics [21]), and is closely related to the semiclassical theory we will discuss later.

This phase space quantum mechanics is first formulated by Moyal [50], and later translated in the solid-state context by Blount [5, 6]. Its central idea is to formulate the quantum mechanics using continuous phase space variables $\boldsymbol{p}$ (momentum) and $\boldsymbol{q}$ (position) by constructing a distribution function from the quantum mechanical state. The complete theory consists of three parts: (1) it gives the distribution function in terms of $\boldsymbol{p}$ and $\boldsymbol{q}$ from the quantum mechanical state; (2) it offers a systematic way to find a dynamics variable $O(\boldsymbol{p}, \boldsymbol{q})$ in the phase space as the exact counterpart of the corresponding operator $\hat{O}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}})$ in the Hilbert space; (3) for any two dynamical variable $O(\boldsymbol{p}, \boldsymbol{q})$ and $T(\boldsymbol{p}, \boldsymbol{q})$, one can define a Moyal star operation $O(\boldsymbol{p}, \boldsymbol{q}) \star T(\boldsymbol{p}, \boldsymbol{q})$ as an asymptotic expansion over the commutator $i[\hat{p}_i, \hat{q}_i] = \hbar$, to exactly recover the operator product $\hat{O}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}})\hat{T}(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}})$ in the Hilbert space.

In the solid-state context, under an external magnetic field, the magnetic field will modify the commutator in the following way: $i[\hat{p}_1, \hat{p}_2] = \hbar B$. Therefore, the asymptotic expansion in the Moyal product is now with respect to $\hbar B$, which is certainly valid at small B. With this understanding in mind, the method of solving the dynamics from Eq.(2.1) can be easily constructed

based on the Foldy-Wouthuysen transformation: (1) find the Hamiltonian $H_{nm}(\boldsymbol{p},\boldsymbol{q})$ in the phase space, and the key to understand the dynamics is to find a unitary transformation S to diagonalize $H_{nm}(\boldsymbol{p},\boldsymbol{q})$ order by order (we comment that the product operation in diagonalization is the Moyal product); (2) based on the Bloch states $|u_n(\boldsymbol{p})\rangle$ at $B=0$, one can construct the following transformation matrix $(S^{(0)})_{nm}(\boldsymbol{p},\boldsymbol{p}')=\langle u_n(\boldsymbol{p})|u_n(\boldsymbol{p}')\rangle$, where $|u_n(\boldsymbol{p})\rangle$ is the periodic part of the Bloch state; (3) then $S=(S^{(0)})_{nm}(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{q},0)$ can diagonalize $H_{nm}(\boldsymbol{p},\boldsymbol{q})$ at the leading order; (4) one can then find a series of matrices $S^{(i)}$, which is of i-th order with respect to B, and the Moyal product $\Pi_{\ell=0}^{i}S^{(\ell)}$ is unitary up to i-th order; (5) it can be proved that the above constraint only determines the Hermitian part of $S^{(i)}$, and its anti-Hermitian part is determined by requiring that $\Pi_{\ell=0}^{i}S^{(\ell)}$ can diagonalize $H_{nm}(\boldsymbol{p},\boldsymbol{q})$ up to i-th order; (6) all other operators should transform accordingly based on $\Pi_{\ell=0}^{i}S^{(\ell)}$.

With this complete understanding of the Hilbert space and operators, one can in principle calculate any physical observable. In fact, Blount has employed this formalism to analyze the orbital part of the magnetic susceptibility [5]. However, the disadvantage of this formalism is that the phase space variables $\boldsymbol{p}$ and $\boldsymbol{q}$ are canonical variables but not physical variables, so they have the gauge-fixing issue (here the gauge refers to the $U(1)$ transformation of the Bloch states).

2.2 First Order Semiclassical Theory

The above phase space quantum mechanics formalism is closely related to the semiclassical theory. In fact, it corresponds to the re-quantization of the semiclassical theory. To demonstrate this, we will first introduce the semiclassical theory up to first order [75, 84].

We start from constructing the wave-packet. For simplicity, we assume that the chemical potential falls in one band labelled by 0, which is well separated from other bands by a band gap. We note that this band 0 is from the local Hamiltonian obtained from the full quantum Hamiltonian by evaluating the gauge potentials at the center of mass position $\boldsymbol{r}_c$ of the wave-packet, $\hat{H}_c(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}}) = \hat{H}_0(\hat{\boldsymbol{p}} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{r}_c, \hat{\boldsymbol{q}}) + \boldsymbol{E} \cdot \boldsymbol{r}_c$, where $\hat{H}_0(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}})$ is the unperturbed Hamiltonian, $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$ are momentum and position operators, and we set $e = \hbar = 1$ to simplify notations and use symmetric gauge in the vector potential. Then the wave-packet for this problem only consists of Bloch states from this band 0:

$$|W\rangle = \int d\boldsymbol{p} C_0 e^{i\boldsymbol{p}\cdot\boldsymbol{q}} |u_0\rangle \,, \tag{2.2}$$

where $|u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{r}_c)\rangle$ is the periodic part of the Bloch states. This wave-packet is assumed to be sharply centered around some point $\boldsymbol{p}_c$ in the Brillouin zone, and subject to the constraint that the center of mass position of the wave-packet coincides with the presumed value $\boldsymbol{r}_c$.

From this wave-packet, we can derive the Lagrangian up to first order

as follows:

$$L = \langle W|i\partial_t - \hat{H}_c - \hat{H}'|W\rangle\,, \tag{2.3}$$

where $\hat{H}' = \frac{1}{4m}\boldsymbol{B}\cdot[(\hat{\boldsymbol{q}}-\boldsymbol{r}_c)\times\hat{\boldsymbol{V}} - \hat{\boldsymbol{V}}\times(\hat{\boldsymbol{q}}-\boldsymbol{r}_c)] + \boldsymbol{E}\cdot(\hat{\boldsymbol{q}}-\boldsymbol{r}_c)$ is the gradient correction to the local Hamiltonian $\hat{H}_c$. Obviously, this Lagrangian is only a function of $\boldsymbol{r}_c$ and $\boldsymbol{p}_c$. Since $\boldsymbol{p}_c$ is not gauge-invariant under a magnetic field, we can re-organize terms in L so that it is a function of $\boldsymbol{r}_c$ and $\boldsymbol{k}_c = \boldsymbol{p}_c + \frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c$, both of which have clear physical meanings. We comment that using $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ does not introduce any gauge issue, but it has a price that $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ are generally non-canonical.

This non-canonical geometry of the semiclassical dynamics emerges in the Euler-Lagrange equations of motion [75] based on L:

$$\dot{\boldsymbol{r}}_c = \frac{\partial\tilde{\varepsilon}}{\partial\boldsymbol{k}_c} - \dot{\boldsymbol{k}}_c\times\boldsymbol{\Omega}, \tag{2.4}$$

$$\dot{\boldsymbol{k}}_c = -\boldsymbol{E} - \dot{\boldsymbol{r}}_c\times\boldsymbol{B}\,. \tag{2.5}$$

where $\tilde{\varepsilon} = \varepsilon_0 - \boldsymbol{B}\cdot\boldsymbol{m}$ with ε_0 being the band dispersion for band 0, $\boldsymbol{m} = -\frac{1}{2}\mathrm{Im}\langle\boldsymbol{\partial}u_0|\times(\varepsilon_0-\hat{H}_c)|\boldsymbol{\partial}u_0\rangle$ being the orbital magnetic moment in band 0, and $\boldsymbol{\Omega} = i\langle\boldsymbol{\partial}u_0|\times|\boldsymbol{\partial}u_0\rangle$ being the Berry curvature in band 0. The second term in the velocity is usually referred to as the anomalous velocity, and it will introduce a factor $1+\boldsymbol{B}\cdot\boldsymbol{\Omega}$ in front of the phase space volume element $d\boldsymbol{r}_c d\boldsymbol{k}_c$, which represents the violation of the Liouville's theorem, and shows the non-canonical nature between $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$.

The non-canonical structure in the dynamics can also be seen from the

connection between canonical and physical variables:

$$\boldsymbol{r}_c = \boldsymbol{q} + \boldsymbol{a}_0 \,, \tag{2.6}$$

$$\boldsymbol{k}_c = \boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{q} + \boldsymbol{B} \times (\boldsymbol{r}_c - \boldsymbol{q}) \,, \tag{2.7}$$

where $\boldsymbol{a}_0 = \langle u_0 | i\boldsymbol{\partial}_{\boldsymbol{p}} | u_0 \rangle$ is the Berry connection in band 0, $\boldsymbol{p}$ and $\boldsymbol{q}$ are canonical variables. The appearance of $\boldsymbol{a}_0$ directly changes the Lie bracket between different components of $\boldsymbol{r}_c$, and introduces the non-canonical structure.

Another way to understand the above semiclassical theory is by re-quantizing it. By promoting the canonical variables in $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ and hence in $\tilde{\varepsilon}$ into the corresponding operators $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$, we essentially obtain an effective quantum mechanical theory for this single band problem, as discussed in [13]. It is interesting to examine the wave function associated with this re-quantization. Note that based on the wave-packet in Eq.(2.2), the Langrangian in Eq.(2.22) can be put in the following form:

$$L = \int d\boldsymbol{p} C_0^\star (i\partial_t - \mathcal{H}) C_0 \,, \tag{2.8}$$

where $\mathcal{H}$ is exactly $\tilde{\varepsilon}$ after the above re-quantization. This coincidence implies that C_0 is the wave function acting by the Hamiltonian after re-quantizing the semiclassical energy. Therefore, the semiclassical theory at first order is actually a transformation from the local Bloch basis to this new basis spanned by different C_0 and this transformation is obviously unitary up to first order. It is in this sense that the semiclassical theory is actually closely related to the phase space quantum mechanics formalism discussed previously.

To make this relation more complete, we can examine the way the position $\hat{\boldsymbol{q}}$ and gauge-invariant crystal momentum $\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \hat{\boldsymbol{q}}$ transform. Based on the definition of $\boldsymbol{r}_c$, it is easy to prove that

$$\langle W|\hat{q}|W\rangle = \int d\boldsymbol{p} C_0^\star (i\partial_p + \boldsymbol{a}_0) C_0 \,. \tag{2.9}$$

Now we will focus on the crystal momentum $\boldsymbol{p}$. $\boldsymbol{p}$ is related to the translational operator $T_{\boldsymbol{R}}$ as follows:

$$\boldsymbol{p} = T_{-\boldsymbol{R}}(-i\partial_{\boldsymbol{R}}) T_{\boldsymbol{R}} \,. \tag{2.10}$$

Therefore, we have

$$\begin{aligned}
\langle W|\boldsymbol{p}|W\rangle &= \langle W|T_{-\boldsymbol{R}}(-i\partial_{\boldsymbol{R}})T_{\boldsymbol{R}}|W\rangle \\
&= \int d\boldsymbol{p} C_0^\star(\boldsymbol{p}) C_0 \langle u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)| \\
&\qquad\qquad e^{-i\boldsymbol{p}\cdot\boldsymbol{q}} T_{-\boldsymbol{R}}(-i\partial_{\boldsymbol{R}}) e^{i\boldsymbol{p}\cdot\boldsymbol{R}} e^{i\boldsymbol{p}\cdot\boldsymbol{q}} |u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times(\boldsymbol{r}_c + \boldsymbol{R}))\rangle \\
&= \int d\boldsymbol{p} C_0^\star(\boldsymbol{p}) C_0 \langle u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)| \\
&\qquad\qquad e^{-i\boldsymbol{p}\cdot\boldsymbol{q}} T_{-\boldsymbol{R}} \boldsymbol{p} e^{i\boldsymbol{p}\cdot\boldsymbol{R}} e^{i\boldsymbol{p}\cdot\boldsymbol{q}} |u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times(\boldsymbol{r}_c + \boldsymbol{R}))\rangle \\
&+ \int d\boldsymbol{p} C_0^\star(\boldsymbol{p}) C_0 \langle u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)| \\
&\qquad\qquad e^{-i\boldsymbol{p}\cdot\boldsymbol{q}} T_{-\boldsymbol{R}} e^{i\boldsymbol{p}\cdot\boldsymbol{R}} e^{i\boldsymbol{p}\cdot\boldsymbol{q}} (-i\partial_{\boldsymbol{R}}) |u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times(\boldsymbol{r}_c + \boldsymbol{R}))\rangle \\
&= \int d\boldsymbol{p} C_0^\star(\boldsymbol{p}) \boldsymbol{p} C_0(\boldsymbol{p}) + \int d\boldsymbol{p} C_0^\star(\boldsymbol{p}) \frac{1}{2}\boldsymbol{B}\times\boldsymbol{a}_0 C_0(\boldsymbol{p}) \,.
\end{aligned} \tag{2.11}$$

From the transformation $\boldsymbol{p} \to \boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\boldsymbol{a}_0$, we know that $\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\hat{\boldsymbol{q}} \to \boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times i\boldsymbol{\partial}_{\boldsymbol{p}} + \boldsymbol{B}\times\boldsymbol{a}_0$. Therefore, the transformation of $\hat{\boldsymbol{q}}$ and $\boldsymbol{p} + \frac{1}{2}\boldsymbol{B}\times\hat{\boldsymbol{q}}$ exactly coincides with the relation of $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ to canonical variables as shown

in Eq.(2.6). This is consistent with obtaining new operators from $\Pi_{\ell=0}^{i} S^{(\ell)}$ in the phase space quantum mechanics formalism.

2.3 Horizontal and Vertical Mixing

Despite of the success of the semiclassical theory up to first order, there are many response functions which require a semiclassical theory up to second order. For this purpose, the wave-packet in Eq.(2.2) is no longer appropriate. Notice that from the connection between the semiclassical theory and the phase space quantum mechanics formalism, the semiclassical theory up to second order corresponds to applying a unitary transformation to remove the first order interband matrix element in the Hamiltonian. This will inevitably introduce the coupling between different bands. Therefore, the wave-packet should be most appropriately constructed as follows:

$$|W\rangle = \int d\boldsymbol{p} e^{i\boldsymbol{p}\cdot\boldsymbol{q}} \left(C_0|u_0\rangle + \sum_{n\neq 0} C_n|u_n\rangle \right) , \tag{2.12}$$

where the interband mixing is directly added through C_n and $|u_n\rangle$. We comment that C_n is of first order in electromagnetic field, since we only want to remove the first order interband matrix element of the Hamiltonian.

C_n is connected to C_0 by requiring that $|W\rangle$ is a correct quantum mechanical state:

$$i\partial_t|W\rangle = (\hat{H}_0 + \hat{H}')|W\rangle \, . \tag{2.13}$$

This constraint yields (see Appendix A for details):

$$C_n = -i\frac{1}{2}[\boldsymbol{B} \times (i\boldsymbol{\partial} + \boldsymbol{a}_0 - \boldsymbol{r}_c)C_0] \cdot \boldsymbol{A}_{n0} - \frac{G_{n0}}{\varepsilon_0 - \varepsilon_n} C_0 \, , \tag{2.14}$$

where $G_{n0} = -\boldsymbol{B}\cdot\mathcal{M}_{n0} + \boldsymbol{E}\cdot\boldsymbol{A}_{n0}$, $\mathcal{M}_{n0} = \frac{1}{2}(\sum_{m\neq 0}\boldsymbol{V}_{nm}\times\boldsymbol{A}_{m0} + \boldsymbol{v}_0\times\boldsymbol{A}_{n0})$, $\boldsymbol{A}_{n0} = \langle u_n|i\boldsymbol{\partial}|u_0\rangle$ is the interband Berry connection, $\boldsymbol{V}_{nm} = \langle u_n|\hat{\boldsymbol{V}}|u_m\rangle$ is the velocity matrix element, and $\boldsymbol{v}_0 \equiv \boldsymbol{V}_{00}$. The subscripts n, m, and 0 are band indices. ε represents the band energy. The partial derivative $\boldsymbol{\partial}$ is with respect to the crystal momentum.

There is another way to view this interband mixing from C_n. We can renormalize the $C_n|u_n\rangle$ in Eq.(2.12) to $|u_0\rangle$, and obtain the following modified periodic part of the local Bloch state:

$$|\tilde{u}\rangle = \lambda|u_0(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\hat{\boldsymbol{q}})\rangle + \sum_{n\neq 0}\frac{G_{n0}}{\varepsilon_0-\varepsilon_n}|u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle\,, \tag{2.15}$$

where λ is a normalization factor to ensure $\langle\tilde{u}|\tilde{u}\rangle = 1$ to first order. It is interesting to notice that the interband part in Eq.(2.12) actually has two effects in this modified Bloch states in Eq.(2.15), and the one involving λ is an intraband effect.

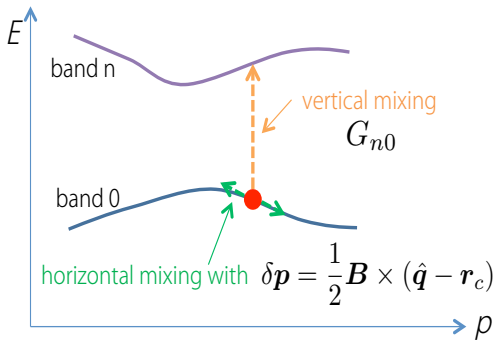

Figure 2.1: (color online) Schematic figure showing the horizontal mixing and the vertical mixing of Bloch states.

We will first discuss the second term in Eq.(2.15). It consists of local Bloch states at the same momentum from all the other bands ($n \neq 0$) and respects the lattice translational symmetry. The essential quantity G_{n0} is invariant under the $U(1)$ gauge transformation to the Bloch states. Meanwhile, its form simply represents the coupling between electromagnetic fields and interband matrix elements of electric and magnetic dipole moment. We comment that $\mathcal{M}_{n0}$ includes the time dependence of $\boldsymbol{r}_c$ in the Bloch states, which is in fact the nonadiabatic effect. The remaining part in $\boldsymbol{B} \cdot \mathcal{M}_{n0}$ is from the interband part ($\boldsymbol{A}_{n0}$) of the position operator $\hat{\boldsymbol{q}}$ in $\hat{H}'$ in the Bloch representation. We call this correction with G_{n0} the vertical mixing since it mixes Bloch states from different bands at the same k-point in the Brillouin zone (as illustrated in Fig.(2.1)).

On the contrary, the first term in Eq.(2.15) only contains the Bloch state inside the same band (band 0). Compared with the local Bloch state $|u_0(\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{r}_c)\rangle$, its crystal momentum is shifted to $\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \hat{\boldsymbol{q}}$, suggesting that the lattice translational symmetry is broken. We comment that this term can be also obtained if we first take the position operator in the vector potential in the exact Hamiltonian $\hat{H}$ as a c-number and later recover it in $|u_0\rangle$ as an operator. This shift of momentum is due to the intraband part of the position operator $\hat{\boldsymbol{q}}$ in $\hat{H}'$ in the Bloch representation. We comment that this intraband correction to $|u_0\rangle$ can be rewritten simply in terms of the shift of momentum $\delta\boldsymbol{p} = \frac{1}{2}\boldsymbol{B} \times (\hat{\boldsymbol{q}} - \boldsymbol{r}_c)$, and reads $\delta\boldsymbol{p} \cdot \hat{\boldsymbol{D}}|u_0\rangle$, where $\hat{\boldsymbol{D}} = \boldsymbol{\partial} + i\boldsymbol{a}_0$ is the gauge-covariant derivative acting on the Bloch states and $\boldsymbol{a}_0 = \langle u_0|i\boldsymbol{\partial}|u_0\rangle$ is the

intraband Berry connection. Since the correction from $\delta \boldsymbol{p}$ mixes Bloch states at neighboring k-points in the same band, we call it the horizontal mixing.

In conclusion, the magnetic field affects the local Bloch states $|u_0\rangle$ in two ways: (i) it vertically mixes $|u_0\rangle$ with the Bloch states $|u_n\rangle$ from other bands; (ii) it also horizontally shifts the Bloch states along the path $\delta \boldsymbol{p}$ according to the affine connection (Berry connection) $\boldsymbol{a}_0$ in the Brillouin zone. We illustrate the two types of mixing of Bloch states schematically in Fig.2.1.

2.4 Positional Shift

As explained in the last section, in the second order theory, we need to use $|\tilde{u}\rangle$ to construct the wave-packet. The difference between $|\tilde{u}\rangle$ and $|u_0\rangle$ is that $|\tilde{u}\rangle$ contains the first order correction $|u_0'\rangle$ from the gradient perturbation $\hat{H}'$. Therefore, the wave-packet now acquires a shift in its center of mass position given by $\boldsymbol{a}_0' = \langle u_0|i\partial_{\boldsymbol{p}}|u_0'\rangle + c.c.$. It corresponds to a first order correction to the Berry connection $\boldsymbol{a}_0 = \langle u_0|i\partial_{\boldsymbol{p}}|u_0\rangle$ of the unperturbed band, but is invariant under the $U(1)$ gauge transformation to the local Bloch states, as can be easily checked using the orthogonality between $|u_0'\rangle$ and $|u_0\rangle$. Therefore it is a physical quantity and represents the shift of the wave packet center due to external fields. Indeed, $\boldsymbol{a}_0'$ transforms as a spatial vector under symmetry operations, e.g. it is odd under spatial inversion and even under time reversal. This positional shift is the key concept in understanding the semiclassical equations of motion up to second order.

From the wave-packet in Eq.(2.12), we can explicitly express this posi-

tional shift:

$$\boldsymbol{a}_0' = \sum_{n\neq 0} \frac{G_{0n}\boldsymbol{A}_{n0} + c.c.}{(\varepsilon_0 - \varepsilon_n)} + \frac{1}{4}\partial_i[(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i \boldsymbol{A}_{n0} + c.c.]\,. \tag{2.16}$$

The first term is due to the horizontal mixing, and the second term is due to the vertical mixing. The above expression has another form:

$$a_i' = F_{ij}B_j + G_{ij}E_j\,, \tag{2.17}$$

where

$$F_{ij} = \mathfrak{Im}\sum_{n\neq 0}\frac{(V_i)_{0n}(\omega_j)_{n0}}{(\varepsilon_0-\varepsilon_n)^2},\ G_{ij} = 2\mathfrak{Re}\sum_{n\neq 0}\frac{(V_i)_{0n}(V_j)_{n0}}{(\varepsilon_0-\varepsilon_n)^3}\,, \tag{2.18}$$

with $\boldsymbol{\omega}_{n0}$ defined as

$$(\omega_j)_{n0} = -i\epsilon_{jk\ell}\sum_{m\neq 0}\frac{[(V_k)_{nm} + (v_k)\delta_{nm}](V_\ell)_{m0}}{\varepsilon_m - \varepsilon_0}\,. \tag{2.19}$$

Here and hereafter, summation is implied over repeated spatial indices. Since $\boldsymbol{a}_0'$ contains the interband velocity, it is not a single band property. All the quantities in Eq.(2.18) can be readily evaluated in first principle calculations.

As a concrete example of the positional shift, we consider a generic two-band model Hamiltonian with

$$\hat{H}_0 = h_0 + \boldsymbol{h}\cdot\boldsymbol{\sigma} \tag{2.20}$$

where $\boldsymbol{\sigma}$ is the vector of Pauli matrices and h's have arbitrary dependence on the crystal momentum. The energy band dispersion is $\varepsilon_\pm = h_0 \pm h$. Assume

the two bands are fully gapped with $h \neq 0$. The positional shift for the lower band can be calculated from Eq.(2.17) and (2.18) with

$$F_{ij} = -\frac{g_{ik}\epsilon_{k\ell j}\partial_{p_\ell} h_0}{4h} - \frac{1}{8}\epsilon_{jk\ell}\Gamma_{\ell ki}, \quad G_{ij} = -\frac{1}{4h}g_{ij}, \tag{2.21}$$

where $g_{ik} = \partial_{p_i}\boldsymbol{n} \cdot \partial_{p_k}\boldsymbol{n}$ (with $\boldsymbol{n} = \boldsymbol{h}/h$) is the quantum metric of the band, and $\Gamma_{\ell ki} = \frac{1}{2}(\partial_{p_i} g_{\ell k} + \partial_{p_k} g_{\ell i} - \partial_{p_\ell} g_{ki})$ is the corresponding Christoffel symbol [3]. Like the Berry curvature, the quantum metric is also a geometric physical quantity, which defines the infinitesimal distance in the Hilbert space on the Brillouin zone. Meanwhile the Christoffel symbol defines the affine geometry of the Brillouin zone [3]. They together make the Brillouin zone a Riemannian manifold. It has been proposed that the quantum metric could be probed by measuring the current noise spectrum [53]. Our result shows that g and Γ are also closely connected with the positional shift, hence might be probed in second order effects.

2.5 Equations of Motion up to Second Order

In this section, we will derive the correct equations of motion up to second order and discuss its connection with the first order semiclassical theory. From the wave-packet in Eq.(2.12) and using similar methods as in [75], we find the following effective Lagrangian:

$$L = -(\boldsymbol{r}_c - \boldsymbol{a}_0 - \boldsymbol{a}_0') \cdot \dot{\boldsymbol{k}}_c - \frac{1}{2}\boldsymbol{B} \times \boldsymbol{r}_c \cdot \dot{\boldsymbol{r}}_c - \tilde{\varepsilon}\,, \tag{2.22}$$

where $\boldsymbol{k}_c = \boldsymbol{p}_c + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{r}_c$ is the gauge invariant crystal momentum and $\tilde{\varepsilon}$ is the semiclassical energy accurate to second order. Its calculation needs the

second order gradient correction $\hat{H}'' = \frac{1}{8}\Gamma_{ij}[\boldsymbol{B}\times(\hat{\boldsymbol{q}}-\boldsymbol{r}_c)]_i[\boldsymbol{B}\times(\hat{\boldsymbol{q}}-\boldsymbol{r}_c)]_j$ to the local Hamiltonian $\hat{H}_c$, where $\Gamma_{ij}=\partial_{p_ip_j}\hat{H}_0$ is the Hessian matrix. Its form is given as follows (see Appendix B for details):

$$\begin{aligned}\tilde{\varepsilon} =& \varepsilon_0 - \boldsymbol{B}\cdot\boldsymbol{m} \\ &+ \frac{1}{4}(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m}) - \frac{1}{8}\epsilon_{sik}\epsilon_{tj\ell}B_sB_tg_{ij}\alpha_{k\ell} \\ &- \boldsymbol{B}\cdot(\boldsymbol{a}_0'\times\boldsymbol{v}_0) + \nabla\cdot\boldsymbol{P}_{\mathrm{E}} \\ &+ \sum_{n\neq 0}\frac{G_{0n}G_{n0}}{\varepsilon_0-\varepsilon_n} + \frac{1}{8m}(B^2g_{ii} - B_ig_{ij}B_j). \end{aligned} \tag{2.23}$$

Here $\boldsymbol{\Omega} = -\mathrm{Im}\langle\boldsymbol{\partial}u_0|\times|\boldsymbol{\partial}u_0\rangle$ is the Berry curvature, $\boldsymbol{m} = -\frac{1}{2}\mathrm{Im}\langle\boldsymbol{\partial}u_0|\times(\varepsilon_0-\hat{H}_c)|\boldsymbol{\partial}u_0\rangle$ is the orbital magnetic moment, $g_{ij} = \mathrm{Re}\langle\partial_iu_0|\partial_ju_0\rangle - (a_0)_i(a_0)_j$ is the quantum metric of k-space[3, 53], $\alpha_{k\ell}=\partial_{k\ell}\varepsilon_0$ is the inverse of effective mass tensor, $\boldsymbol{a}_0'$ is the field-induced positional shift of the wave-packet center, and $\boldsymbol{P}_{\mathrm{E}} = (1/4)[\langle(\boldsymbol{B}\times\hat{\boldsymbol{D}})u_0|(\hat{\boldsymbol{V}}+\boldsymbol{v}_0)\cdot|(\boldsymbol{B}\times\hat{\boldsymbol{D}})u_0\rangle + c.c.]$ is a single band quantity representing the energy polarization density in k-space. Indices i, j, k, ℓ, s and t refer to Cartesian coordinates and repeated indices are summed over. ϵ_{sik} is the totally antisymmetric tensor in three dimension. For the last term in Eq.(2.23), we choose $\Gamma_{ij}=\delta_{ij}/m$ (for Pauli and Schrödinger Hamiltonians) for simplicity and a more general formula is given in Appendix B. All physical quantities in Eq.(2.23) should be understood as functions of the gauge-invariant crystal momentum $\boldsymbol{k}_c$, and the partial derivatives are with respect to $\boldsymbol{k}_c$.

In Eq.(2.23), it is easy to check that each term is gauge-invariant. In fact, they can be classified by their geometrical and physical meaning. The

first two terms in Eq.(2.23) have simple structures, and they are the band energy plus the magnetic dipolar energy, which is expected based on the first order semiclassical theory. We call the two terms in the second line the geometrical energies, since they consist of single band geometrical quantities, i.e. the Berry curvature and the quantum metric. In the Brillouin zone, the Hilbert space with single band Bloch states $|u_0\rangle$ forms a fiber bundle, whose curvature is characterized by the Berry curvature $\boldsymbol{\Omega}$[84], and the distance in which is captured by the quantum metric[53]. For the remaining two quantities, α_{kl} depends only on the band dispersion and the magnetic moment $\boldsymbol{m}$ is a single band quantity. It is interesting to note that α_{ij} and $\boldsymbol{m}$ actually form a conjugate pair: they are proportional to the real and the imaginary part of $\delta_{ij}/m + 2\langle\partial_i u_0|(\varepsilon_0 - \hat{H}_c)|\partial_j u_0\rangle$, respectively. So are the quantum metric and the Berry curvature, with respect to the quantity $\langle\partial_i u_0|\partial_j u_0\rangle - (a_0)_i(a_0)_j$. Thus the less obvious meaning of g_{ij} and $\boldsymbol{m}$ can be understood from their well studied conjugate partners $\boldsymbol{\Omega}$ and α_{ij}.

The first term in the third line of Eq.(2.23) is a real space polarization energy. It is due to the fact that the magnetic field shifts the wave-packet center by $\boldsymbol{a}_0'$, and hence modifying the magnetic dipole moment. This modified magnetic dipole couples to the magnetic field to change the wave-packet energy. The next term is a k-space polarization energy. This can be understood by noticing that the the momentum shift $\delta\boldsymbol{p}$ gives rise to a second order energy polarization in k-space, $(1/2)(\hat{H}'\delta\boldsymbol{p} + c.c.)$. Similar to the relation between electric polarization and charge, the divergence of such energy polar-

ization yields a local energy correction. We find that this term is a single band quantity, and is related to the quadrupole moments of the velocity operator.

In the fourth line of Eq.(2.23), the first term is a standard second order perturbation energy through virtual interband transitions. The last term in Eq.(2.23) is from the perturbation of $\hat{H}''$. Note that this term vanishes for the Dirac Hamiltonian, and for the nonrelativistic Pauli and Schrödinger Hamiltonians, it comes with the bare electron mass m.

The above geometrical and physical meanings of the second order wave-packet energy can also be implied from the vertical and horizontal mixings as shown in Eq.(2.15). Such two types of corrections in $|\Psi\rangle$ originates from the Bloch represention of $\hat{H}'$. Therefore, they enter the wave packet energy in Eq.(2.23) through both $\hat{H}'$ and $|\Psi\rangle$. If two horizontal mixings are combined to yield a second order energy, only the neighborhood in the Brillouin zone is involved, and we should obtain a purely geometrical contribution as in the second line in Eq.(2.23). On the contrary, if two vertical mixings are combined, then virtual interband transition is involved, and we obtain an interband effect as the first term in the fourth line of Eq.(2.23). If the horizontal and vertical mixing are combined together, we obtain the k-space polarization energy, which is a single band but not necessarily geometrical quantity.

By applying the Euler-Lagrangian equations of motion, we can obtain

the following dynamics from Eq.(2.22):

$$\dot{\boldsymbol{r}}_c = \frac{\partial \tilde{\varepsilon}}{\partial \boldsymbol{k}_c} - \dot{\boldsymbol{k}}_c \times \tilde{\boldsymbol{\Omega}}, \tag{2.24}$$

$$\dot{\boldsymbol{k}}_c = -\boldsymbol{E} - \dot{\boldsymbol{r}}_c \times \boldsymbol{B}\,, \tag{2.25}$$

where $\tilde{\Omega} = \nabla \times (\boldsymbol{a}_0 + \boldsymbol{a}_0')$. To our surprise, compared with the first order semiclassical dynamics in Eq.(2.4) and (2.5), these second order dynamics possess exactly the same structure. The only differences are that the band energy is now corrected up to second order as shown in Eq.(2.23) and that the Berry curvature now has a first order correction as given by the curl of $\boldsymbol{a}_0'$. We comment that this change of the Berry curvature does not affect the Chern number. The reason is that $\boldsymbol{a}_0'$ is a physical vector and periodic in the Brillouin zone, so the integration of its curl over the whole Brillouin zone vanishes. With this equations of motion, one can easily derive the modification to the phase space density of states, and it reads as $1 + \boldsymbol{B} \cdot \tilde{\boldsymbol{\Omega}}$, i.e. we can simply change $\boldsymbol{\Omega}$ in the first order result by our new field-dependent Berry curvature $\tilde{\Omega}$.

From the Lagrangian in Eq.(2.22), we can also derive the connection between physical variables $\boldsymbol{r}_c$, $\boldsymbol{k}_c$ and canonical variables $\boldsymbol{q}$, $\boldsymbol{p}$. The method is that this connection can change L to a standard form: $L = \boldsymbol{p} \cdot \dot{\boldsymbol{q}} - \varepsilon(\boldsymbol{p}, \boldsymbol{q})$, where ε is an appropriate energy. The result reads:

$$\boldsymbol{r}_c = \boldsymbol{q} + \boldsymbol{a}_0 + \frac{1}{2}(\boldsymbol{B} \times \boldsymbol{a}_0 \cdot \boldsymbol{\partial}_{\boldsymbol{p}})\boldsymbol{a}_0 + \frac{1}{2}\boldsymbol{\Omega} \times (\boldsymbol{B} \times \boldsymbol{a}_0) + \boldsymbol{a}_0'\,, \tag{2.26}$$

$$\boldsymbol{k}_c = \boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{q} + \boldsymbol{B} \times (\boldsymbol{r}_c - \boldsymbol{q})\,. \tag{2.27}$$

Note that in Eq.(2.26) and (2.27), the argument for $\boldsymbol{a}_0$ and $\boldsymbol{a}_0'$ is $\boldsymbol{p} + \frac{1}{2}\boldsymbol{B} \times \boldsymbol{q}$. Previously people thought that the first four terms on the right hand side

of Eq.(2.26) would be sufficient to first order in the fields [13], but now one observes that this is incomplete without the positional shift $\boldsymbol{a}_0'$. We comment that this relation can also be obtained by similar methods discussed in section 2.2.

Chapter 3

Magnetoelectric Coupling and Non-linear Anomalous Hall Effect

With the equations of motion in Eq.(2.24) and (2.25), and the band energy in Eq.(2.23), we obtain a complete semiclassical theory up to second order. We can now use the free energy to obtain equilibrium response functions and use the Boltzmann transport theory to obtain transport response functions. As examples, in this chapter, we will derive the magnetoelectric polarizability and the non-linear anomalous Hall effect from the semiclassical theory up to second order. [1]

3.1 Magnetoelectric Coupling

Usually when one applies a magnetic field, one expects a magnetization based on the magnetic susceptibility. Similarly, electric field can induce a polarization due to polarizability. However, the cross effect, namely the magnetization due to the electric field and the polarization due to the magnetic

[1] This chapter is based on *Y.Gao, S. A. Yang, and Q. Niu, Phys. Rev. Lett.112, 166601 (2014)*. My contribution is that I derive the general formalism and apply it to these two applications in the two-band model, and that I partially help to establish the conceptual understanding.

field is also possible. The corresponding response function is defined as the magnetoelectric polarizability as follows [17, 18]:

$$\alpha_{ij} = \frac{\partial P_i}{\partial B_j} = \frac{\partial M_j}{\partial E_i}\,, \tag{3.1}$$

where P and M are polarization and magnetization, respectively. In multiferroics materials, this type of response function is usually mediated by phonons. The purely electronic orbital part is derived in [18]. It consists of two contributions, one is topological since it involves the integration of Chern-Simons 3-form, and the other one is cross-gap, since it involves virtual transition between occupied and unoccupied states. As a validity check of the semiclassical theory, here we will show that it can be used to derive this electronic orbital part of magnetoelectric effect.

The polarization in solids is obtained as follows [36]:

$$\boldsymbol{P} = \int \frac{d\boldsymbol{k}}{8\pi^3}\boldsymbol{a}_0\,, \tag{3.2}$$

where $\boldsymbol{a}_0$ is the Berry connection in one band. This can be understood based on Eq.(2.6). The canonical position actually yields the lattice vector, which should be discarded since the polarization is only well-defined up to lattice vectors.

An external magnetic field will change this expression in two ways: (1) the density of states is changed to $1 + \boldsymbol{B}\cdot\tilde{\boldsymbol{\Omega}}$; (2) the Berry connection is also changed according to Eq.(2.26). We take account these two changes, and obtain the following formula from Eq.(3.2):

$$\boldsymbol{P}' = -\int_{BZ} \frac{d^3k}{(2\pi)^3}\left(\frac{1}{2}(\Omega_0\cdot\boldsymbol{a}_0)\,\boldsymbol{B} + \boldsymbol{a}_0'\right)\,. \tag{3.3}$$

The first term in Eq.(3.3) is the Chern-Simons 3-form for the Abelian case, which is the essential quantity in the classification of three dimensional topological insulators [59]. It only involves the Berry connection and Berry curvature of the unperturbed band, and can be derived within the framework of the first order semiclassical theory [17, 86]. An additional term from the field-induced band mixing was envisioned in [86], but its calculation had to wait for a full quantum perturbation formulation [18]. We now see that this additional term has a very compact form of the positional shift integrated over the Brillouin zone. Our results in Eq.(3.3) agrees exactly with the quantum calculation, confirming the reliability of our semiclassical theory. Moreover, if we apply our formula for the positional shift induced by an electric field, we also get the correct result for the electric polarizability.

As an example, we will calculate the magnetoelectric polarization for a generic two-band model in Eq.(2.20). The topological part (the first term in Eq.(3.3)) is quantized and well understood [59, 60, 28, 31, 32, 22, 19, 42], so we focus on the magnetoelectric polarization due to the positional shift, which requires broken time reversal and spatial inversion symmetry [18]. For the case of lower band being occupied, we have:

$$P'_\mu = \int \frac{d^3k}{8\pi^3} \frac{g_{\mu\nu} f_\nu}{4h} = \int \frac{d^3k}{8\pi^3} G_\mu B \,. \tag{3.4}$$

In the second expression, $\boldsymbol{G} = (4h)^{-1}(\hat{z}\cdot\boldsymbol{\partial}h_0 \times \boldsymbol{\partial}n_i)\boldsymbol{\partial}n_i$, with $\hat{z}$ being the direction of the magnetic field. We note that, if h_0 is a constant $\boldsymbol{G}$ vanishes. This is consistent with an earlier finding that a non-zero orbital magnetoelectric polarization must break particle-hole symmetry [18].

From the symmetry analysis, one can build a minimal lattice model that realizes this effect in 2D. Notice that for model Eq.(2.20) in 2D the topological part of magnetoelectric polarization, i.e. the first term in Eq.(3.3) vanishes [86] hence only the contribution from $\boldsymbol{a}_0'$ exists. Moreover, since $\boldsymbol{a}_0'$ transforms as a spatial vector and it must lie in the plane, in general it must vanish if the system has in-plane rotational symmetry. And if in-plane mirror symmetry exists, $\boldsymbol{P}'$ would be restricted to be along the normal direction of the mirror line (see Fig.3.1).

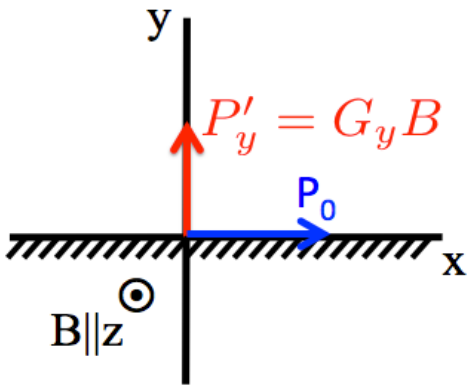

Figure 3.1: (color online) Magnetoelectric Polarization in a 2D system with a mirror line along x axis. The mirror symmetry requires the zeroth order polarization $\boldsymbol{P}_0$ to be along the mirror line, and requires the first order $\boldsymbol{P}'$ to lie in the perpendicular direction.

These symmetry constraints provide guidance for the construction of the lattice model. The Haldane model with opposite onsite potentials on two sublattices breaks the time reversal and spatial inversion symmetry, so it may be a good candidate to realize such a system [27]. However, the original

Haldane model has three-fold rotational symmetry in the material plane, which will suppress the orbital magnetoelectric polarization. So we change the local magnetic field to break the rotational symmetry. The details of the model are shown in Fig.3.2. Here the inter-sublattice hopping amplitudes (between A and B sites) are equal to t_1. The intra-sublattice hopping (between two nearest A sites or B sites) has the same strength t_2. However the phases of the hopping depend on the local field and are subject to symmetry constraints. Since under the in-plane mirror reflection, the magnetic flux behaves as the magnetic field and will flip the sign. Without loss of generality, we adjust the local field so that the upper two intra-sublattice hopping phases equal to ϕ for the counterclockwise direction. Then the in-plane mirror symmetry requires that the corresponding two lower hopping phases are $-\phi$, and that the two vertical intra-sublattice hopping phases vanish.

3.2 Non-linear Anomalous Hall Effect

The semiclassical dynamics can also be used to calculate the transport current. Here we consider the intrinsic current which does not depend on the transport relaxation time. The semiclassical current is defined by

$$\boldsymbol{j} = \int \frac{d\boldsymbol{k}}{(2\pi)^3} \mathcal{D} \dot{\boldsymbol{r}} f_0 \,, \tag{3.5}$$

where $\mathcal{D} = 1 + \boldsymbol{B} \cdot \tilde{\boldsymbol{\Omega}}$ is the density of states, $\dot{\boldsymbol{r}}$ is given in Eq.(2.24), and f_0 is the equilibrium Fermi distribution function. It is important to note that the argument of f must be the modified band energy $\tilde{\varepsilon}$ as given in Eq.(2.23).

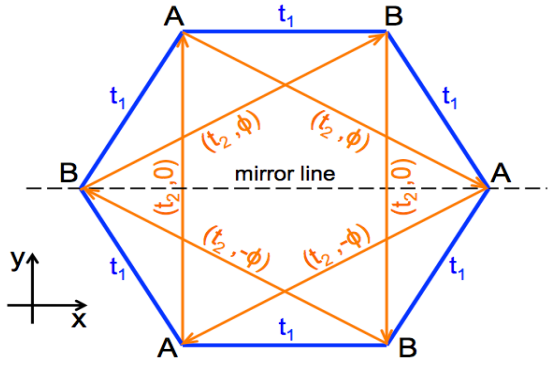

Figure 3.2: (color online) Modified Haldane model with a mirror line symmetry to realize magnetoelectric coupling in 2D. The first argument in the bracket for the second nearest neighbour hopping is the hopping strength, and the second one is the phase due to the local field. Phases in the upper and lower part of the cell have opposite signs because they behave as the magnetic field under reflection.

The reason is that under this Fermi function, the current that only depends on B vanishes, as expected from the fact that the magnetic field alone does not drive a current.

We want to derive the intrinsic contribution to the current. Therefore, the distribution function f_0 can be expanded as follows:

$$f_0(\tilde{\varepsilon}) = f_0(\varepsilon_0) + \frac{\partial f_0}{\partial \varepsilon_0}\varepsilon' + \frac{\partial f_0}{\partial \varepsilon}\varepsilon'' + \frac{1}{2}\frac{\partial^2 f_0}{\partial \varepsilon^2}(\varepsilon')^2 + \cdots , \tag{3.6}$$

where $\varepsilon' = -\boldsymbol{B}\cdot\boldsymbol{m}$ is the first order correction to the semiclassical energy, and ε'' is the second order correction.

To derive the nonlinear anomalous Hall effect, we plug (3.6) into Eq.(3.5)

and keep the current at second order. The last two terms in Eq.(3.6) cancel the corresponding terms in the center of mass velocity, i.e. the second order correction to the wave packet energy does not enter into the final result. The final result is clean and simple:

$$\boldsymbol{j}' = \boldsymbol{E} \times \int [\boldsymbol{v}_0 \times \boldsymbol{a}_0' + \boldsymbol{\Omega}(\boldsymbol{B} \cdot \boldsymbol{m})] \frac{\partial f_0}{\partial \varepsilon_0} \frac{d^3k}{(2\pi)^3} \,. \tag{3.7}$$

Or more explicitly we can write

$$\frac{\partial^2 j_i'}{\partial E_j \partial E_\ell} = \int (v_i G_{j\ell} - v_j G_{i\ell}) \frac{\partial f_0}{\partial \varepsilon_0} \frac{d^3k}{(2\pi)^3} \,, \tag{3.8}$$

$$\frac{\partial^2 j_i'}{\partial E_j \partial B_\ell} = \int (v_i F_{j\ell} - v_j F_{i\ell} + \epsilon_{ijk} \Omega_k m_\ell) \frac{\partial f_0}{\partial \varepsilon_0} \frac{d^3k}{(2\pi)^3}, \tag{3.9}$$

We call the two response functions in Eq.(3.8) and (3.9) the electric nonlinear anomalous Hall effect and the magneto nonlinear anomalous Hall effect respectively. We call such currents as anomalous because they are not caused by the Lorentz force as in the case of the ordinary Hall effect. These response functions can be directly evaluated by first principle methods.

From Eq.(3.5) we see that the intrinsic nonlinear current is normal to the electric field, so it is purely of Hall type. Also, the appearance of $\partial f_0/\partial \varepsilon_0$ shows that it is a Fermi surface property. The second term in Eq.(3.7) contains only the Berry curvature and magnetic moment, so it can be derived from an naive extension of the first order theory and has been discussed in the study of anomalous Hall transport in multi-valley systems [9]. The first term comes from the correction of the Berry curvature due to the positional shift found in this work, so it can only be derived within the semiclassical theory up to second

order. We will show that this term is quite nontrivial. Moreover, we note that for in-phase oscillating $\boldsymbol{E}$ and $\boldsymbol{B}$ fields, the first order intrinsic anomalous Hall response vanishes upon time average, hence the DC intrinsic anomalous Hall current would be dominated by the nonlinear response $\boldsymbol{j}'$.

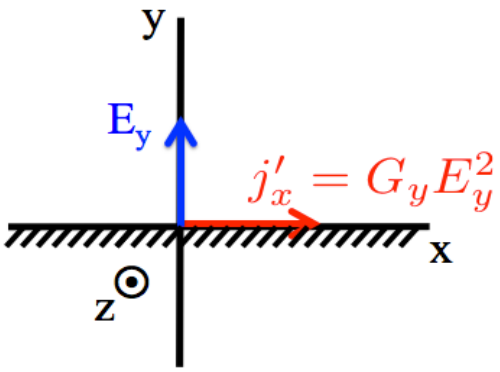

Figure 3.3: (color online) The electric nonlinear anomalous Hall effect in a 2D system with a mirror line along x axis. The linear anomalous Hall current vanishes due to the mirror symmetry, but the nonlinear anomalous Hall current can exist along the mirror line if the electric field is applied along the perpendicular direction.

First we will discuss the electric nonlinear anomalous Hall effect with $\boldsymbol{B} = 0$. Then the intrinsic nonlinear Hall conductivity $\sigma'_{xy} = \partial j_x/\partial E_y$ is proportional to the electric field and only the term with positional shift in Eq.(3.5) contributes. For the generic two band model (2.20), the result is $\sigma'_{xy} = -\int(\partial f_0/\partial\varepsilon_0)\boldsymbol{G}\cdot\boldsymbol{E}d^3k/(2\pi)^3$. Interestingly, since it involves the $\boldsymbol{G}$ vector, it is actually connected to the orbital magnetoelectric polarizability. In fact, the two effects have the same symmetry properties, requiring both time

reversal and spatial inversion symmetries to be broken in the system. This nonlinear anomalous Hall current will dominate if the corresponding linear current vanishes due to symmetry constraint. For example, if a 2D system has a mirror line perpendicular to the electric field, then the linear intrinsic current vanishes because the (unperturbed) Berry curvature has a sign change under mirror operation, while the nonlinear current could be finite (see Fig.3.3).

Now we discuss the magneto nonlinear anomalous Hall effect. It actually does not have such a stringent symmetry constraint, since the current transforms in the same way as the product of electric and magnetic fields under both time reversal and spatial inversion. So it is much easier to realize in real systems. Furthermore, if the system itself has time reversal symmetry (neglect the small Zeeman splitting due to the external magnetic field), both the linear anomalous Hall effect and the electric nonlinear anomalous Hall effect vanish, and the magneto nonlinear anomalous Hall effect dominates. For the two band model, we find that

$$\sigma'_{xy} = \int \frac{d^3k}{(2\pi)^3} \frac{\partial f_0}{\partial \varepsilon_0} \Big[\frac{g_{ij}}{4h} (\hat{\boldsymbol{z}} \times \boldsymbol{v})_i \, (\boldsymbol{B} \times \boldsymbol{\partial} h_0)_j \\ - \frac{1}{8} (\hat{\boldsymbol{z}} \times \boldsymbol{v})_i \epsilon_{k\ell j} B_k \Gamma_{j\ell i} + h(\boldsymbol{\Omega} \cdot \hat{\boldsymbol{z}})(\boldsymbol{\Omega} \cdot \boldsymbol{B}) \Big] , \tag{3.10}$$

where $\boldsymbol{v} = \boldsymbol{\partial}(h_0 - h)$ and $\boldsymbol{\Omega} = \frac{1}{2}\epsilon_{ijk} n_i \boldsymbol{\partial} n_j \times \boldsymbol{\partial} n_k$ is the unperturbed Berry curvature. As a concrete example, let's consider a 2D gapped Dirac model with $h_0 = 0$ and $\boldsymbol{h} = (vk_x, vk_y, \Delta)$, which is widely used to study systems such as symmetry-breaking graphene, MoS_2, topological insulator surfaces with time reversal symmetry breaking, and topological insulator thin films [52, 85, 60,

28]. Here v is the fermi velocity and Δ is the gap parameter. Then the first term in Eq.(3.10) vanishes immediately due to the particle-hole symmetry. Consider an in-plane electric field and an out-of-plane magnetic field, we obtain that (at zero temperature)

$$\sigma'_{xy} = -e^3 \frac{v^2(v^2 p_F^2 + 2\Delta^2)}{16\pi(v^2 p_F^2 + \Delta^2)^2} B \,, \tag{3.11}$$

where $p_F = \hbar k_F$ being the Fermi momentum and we assume the Fermi level is in the upper band. We have recovered factors e and $\hbar$ in Eq.(3.11) to generate the correct unit. In realistic materials, there may be valley degeneracy, such as the inequivalent K and K' connected by time reversal symmetry in MoS_2 or graphene (with inversion symmetry breaking), the contributions to this magneto nonlinear anomalous Hall effect from the two valleys in fact add together rather than cancel each other as in the first order response [88].

We comment that for magneto nonlinear anomalous Hall effect, since both the electric field and magnetic field are applied (in a normal configuration), there is also the ordinary Hall response due to Lorentz force. The ordinary Hall conductivity also has a linear B dependence. However there is an important difference. The ordinary Hall conductivity is proportional to the square of relaxation time (or longitudinal conductivity), while our intrinsic nonlinear conductivity does not have such extrinsic dependence. On the other hand, in terms of the Hall resistivity, the ordinary effect has an intrinsic looking form $\rho_{xy}^{\rm ord} = -B/ne$, where n is the carrier density, while intrinsic nonlinear Hall resistivity acquires a dependence on the square of the longitudinal

resistivity. This is well understood as a result of matrix inversion between the conductivity and resistivity tensors with the usual condition that longitudinal coefficients are much bigger than the transverse ones. Therefore from this discussion, like the linear anomalous Hall effect, our nonlinear effect would become important for more resistive samples.

As a concrete example of the difference between these two effects, we consider the gapped Dirac model. From Eq.(3.11), we obtain the following ratios between resistivities from these two effects

$$\frac{\rho'_{xy}}{\rho^{\rm ord}_{xy}} = \left(\rho_{xx}\frac{e^2}{4h}\right)^2 \left[1-\left(\frac{\Delta}{\varepsilon_F}\right)^4\right], \tag{3.12}$$

where $\varepsilon_F = \sqrt{v^2 p_F^2 + \Delta^2}$ is the Fermi energy, $\rho'_{xy} \simeq \sigma'_{xy}\rho^2_{xx}$ is obtained by inverting σ'_{xy}, and ρ_{xx} is the longitudinal resistivity. In Eq.(3.12), the first factor is proportional to ρ^2_{xx}, which is universal (model independent). The second factor shows a simple dependence on Fermi energy for the gapped Dirac model: it vanishes at the band bottom and quickly saturates with increasing Fermi energies. Our predictions can be tested by the standard Hall bar measurement on single layer MoS_2 or Bi_2Se_3 thin films with longitudinal resistivity tuned by temperature, film thickness, or doping. The difference in the scaling in terms of the longitudinal resistivity as shown in Eq.(3.12) can be used to disentangle the ordinary Hall effect and the nonlinear anomalous Hall effect.

Chapter 4

Static Magnetic Susceptibility at Zero field

In this chapter, I will formulate the general theory of static magnetic susceptibility from my semiclassical theory. [1]

4.1 General Theory

The magnetic susceptibility is a second order response function, defined from the free energy density G by:

$$\chi_{ij} = -\frac{\partial^2 G}{\partial B_i \partial B_j}\,. \tag{4.1}$$

In atomic physics, various mechanisms for magnetic susceptibility are identified. They include the Pauli paramagnetic susceptibility, which is due to the spin angular momentum of electrons, the Van-Vleck paramagnetic susceptibility, which is due to the interband transition induced by the magnetic vector potential, the Langevin or Larmor diamagnetic susceptibility, which is due to the second order perturbative Hamiltonian from the vector potential, and the Lan-

[1] This chapter is based on *Y. Gao, S. A. Yang, and Q. Niu, Phys. Rev. B 91, 214405 (2015)*. My contribution is that I derive the general formalism of the susceptibility and calculate it to the honeycomb model and the dice lattice. I also derive the physical meaning of various contributions to the susceptibility.

dau diamagnetic susceptibility, which is purely due to the non-commutative relation between different components of the mechanical momentum.

In the solid-state context, the Landau magnetism has been re-discovered as the Peierls-Landau magnetism. It was later realized that unlike in the atomic physics, the Peierls-Landau magnetism can be paramagnetic around the saddle points inside the band [78]. The Pauli paramagnetism for spins is easy to understand. However, other mechanisms are rather difficult to understand.

The standard technique to calculate the magnetic susceptibility includes the linear response theory [39, 71, 25, 64, 65, 24, 78], and the spatial perturbation technique [1, 29, 45, 68, 80, 5, 80, 48]. However, they have their own disadvantages. The linear response theory can calculate the magnetic susceptibility response as a whole, but it is rather difficult to identify different mechanisms that contribute to this response function. So for each individual materials, one has to calculate the susceptibility independently, and a universal understanding beyond each specific model is hard to establish. The spatial perturbation technique is rather complicated in mathematics. Although it is possible to envision mechanisms of susceptibility, as shown in [5], usually canonical variables are involved in the derivation, which are not independent of the choice of the phase of the Bloch state. This gauge issue comprises the understanding of various mechanisms.

Our semiclassical theory up to second order as established in Chapter 2 is a good candidate to understand the complicated orbital part of the magnetic

susceptibility. On one hand, we have the complete dynamical theory up to second order, and our theory only involves the physical variables, so it does not have any gauge issue; on the other hand, the form of the semiclassical version of the free energy up to second order has been established in [5]. The thermodynamic free energy $G = \mathrm{Tr}[g(\hat{H})]$, where $g(\varepsilon) = -k_B T \ln(1+\exp[(\mu-\varepsilon)/k_B T])$, T is the temperature and μ is the chemical potential. Under external magnetic field, the semiclassical limit of the free energy density written in terms of physical variables reads as [5]:

$$G = \int_{\mathrm{BZ}} [\mathcal{D} g(\tilde{\varepsilon}) + g_{\mathrm{L}}] \frac{d^3 k_c}{8\pi^3}. \tag{4.2}$$

Here V is the system volume and $\mathcal{D} = 1 + \boldsymbol{B}\cdot(\boldsymbol{\Omega} + \nabla \times \boldsymbol{a}_0')$ is the modified density of states. The first term in the bracket of Eq.(4.2) is from the semiclassical free energy density with second order energy correction. The second term g_{L} is the Peierls-Landau magnetic energy: $g_{\mathrm{L}} = -(f'/48) B_s B_t \epsilon_{sik} \epsilon_{tj\ell} \alpha_{ij} \alpha_{k\ell}$, where f' is the energy derivative of the Fermi distribution function f. For isotropic bands, the effective mass tensor α takes a diagonal form, and g_{L} reduces to its familiar form. This term originates from the discreteness of the Landau levels, and appears when we transform the free energy from the quantum version to its semiclassical limit [5].

With the help of Eq.(2.23), we can expand the free energy in Eq.(4.2) to second order: $G = V \int_{\mathrm{BZ}} (g_0 + g' + g'') d^3 k_c/(8\pi^3)$. At zeroth order, $g_0 = g(\varepsilon_0)$. At first order, $g' = -\boldsymbol{B}\cdot\boldsymbol{m} f + \boldsymbol{B}\cdot\boldsymbol{\Omega} g_0$, which yields the same magnetization as in Ref. [72]. The second order g'' is required for the magnetic susceptibility

$\chi_{ij} = -(1/V)(\partial^2 G/\partial B_i \partial B_j)_{\mu,T,V}$, and reads (for simplicity we take $\Gamma_{ij} = \delta_{ij}/m$)

$$\begin{aligned} g'' = & g_{\rm L} + \frac{f'}{2}(\boldsymbol{B}\cdot\boldsymbol{m})^2 - f'\boldsymbol{v}_0\cdot\boldsymbol{P}_{\rm E} \\ & + f\frac{G_{0n}G_{n0}}{\varepsilon_0-\varepsilon_n} + \frac{f}{8m}(B^2 g_{ii} - B_i g_{ij} B_j) \\ & - \frac{3f}{4}(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m}) - \frac{f}{8}\epsilon_{sik}\epsilon_{tj\ell}B_s B_t g_{ij}\alpha_{k\ell}. \end{aligned} \tag{4.3}$$

This second order free energy yields a susceptibility tensor as follows:

$$\begin{aligned} \chi_{st} = & \frac{f'}{24}\epsilon_{sik}\epsilon_{tj\ell}\alpha_{ij}\alpha_{k\ell} - f' m_s m_t \\ & + \frac{f'}{4}(\epsilon_{sik}\epsilon_{tj\ell} + \epsilon_{tik}\epsilon_{sj\ell})(v_0)_k \langle \hat{D}_i u_0 | (\hat{V}+v_0)_\ell | \hat{D}_j u_0\rangle \\ & - \frac{(\mathcal{M}_{n0})_s(\mathcal{M}^\star_{n0})_t + (\mathcal{M}_{n0})_t(\mathcal{M}^\star_{n0})_s}{\varepsilon_0-\varepsilon_n} f - \frac{f}{4m}(g_{ii}\delta_{st} - g_{st}) \\ & + \frac{3f}{4}(\Omega_s m_t + \Omega_t m_s) + \frac{f}{4}\epsilon_{sik}\epsilon_{tj\ell} g_{ij}\alpha_{k\ell}\,. \end{aligned} \tag{4.4}$$

From Eq.(5.52), we find that the susceptibility only contains two types of contributions, one from Fermi surface and the other one from Fermi sea. Magnetisms in the first line of Eq.(5.52) are contributions from the Fermi surface. The first two contributions are the Peierls-Landau magnetism, and the Pauli paramagnetism for the orbital moment $\boldsymbol{m}$. In solids, the Peierls-Landau magnetism can be paramagnetic, especially near the band saddle points [78]. These two contributions are prominent around singular points where the density of states diverges. For example, around the saddle point, Peierls-Landau magnetism dominates in general [78]. The third term is due to the k-space energy polarization in Eq.(2.23), and is first identified here. Similar to the Pauli and

Peierls-Landau magnetism, it also involves only single band quantities. This term generally compete with the Pauli orbital and Peierls-Landau magnetisms as illustrated in Fig.(4.1), except at the band maxima or minima where $|\boldsymbol{v}_0|$ vanishes.

The other terms in Eq.(5.52) are Fermi sea contributions. The first term in the second line yields the Van Vleck susceptibility originated from the vertical mixing energy in Eq.(2.23). It is due to the interband transition from the vector potential, and is the only interband contribution to the orbital magnetic susceptibility. It is always paramagnetic after summing over all the occupied bands, similar to the Van Vleck paramagnetic susceptibility in atomic systems. The second one yields a Langevin-like magnetic susceptibility from the last term in Eq.(2.23). It can be expressed in a compact form using the quantum metric g_{ij}, which describes the intrinsic fluctuation of position-position operators $(\hat{q}_i\hat{q}_j)$ in the Bloch representation: $g_{ij} = \mathrm{Re}[(A_i)_{0n}(A_j)_{n0}]$. For Pauli or Schrödinger Hamiltonian with constant mass, this term yields diamagnetic response along directions that diagonalize g_{ij}. Its expression will change for effective Hamiltonians with a general Hessian matrix Γ_{ij} as given in Appendix B. For Dirac Hamiltonian, the Langevin-like magnetic susceptibility vanishes.

Magnetic free energies in the third line in Eq.(5.52) have no analogs as in atomic physics or for free particles, and are first identified here. We call the susceptibility from these two terms the geometrical magnetic susceptibility, because they are due to the geometrical energies in Eq.(2.23) from

the horizontal mixing of Bloch states and the geometrical correction to the density of states in Eq.(4.2). Notice that the first term consists of the Berry curvature, it is important when the band structure contains monopole or other nontrivial topological structures. For example, for two band systems with the particle-hole symmetry, geometrical magnetic susceptibility always yields a diamagnetic susceptibility and is a prominent or even dominant contribution in the band gap.

4.2 In Atomic Insulator Limit

A good way to understand various mechanisms of the magnetic susceptibility in Eq.(5.52) is to examine various terms in the atomic insulator limit, where the hopping between lattice sites is suppressed, and the lattice is simply an array of atomic insulators. Then the susceptibilities in the second line in Eq.(5.52) will reduce to the familiar Van Vleck paramagnetic susceptibility and Langevin diamagnetic susceptibility in atomic physics. To demonstrate this, we use the Wannier function representation. Notice that the periodic part of the Bloch function can be expressed in terms of the Wannier function:

$$|u_n(\boldsymbol{k}_c)\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}_c\cdot(\boldsymbol{r}-\boldsymbol{R})} |W_n(\boldsymbol{r},\boldsymbol{R})\rangle \,, \tag{4.5}$$

where n is the band index and $|W_n(\boldsymbol{r},\boldsymbol{R})\rangle$ is the Wannier function localized at $\boldsymbol{R}$ for band n. With the help of Eq.(4.5), we can express the Van Vleck paramagnetic susceptibility in terms of Wannier functions. To simplify the discussion, we consider an insulator (with each band either completely filled

or empty) with a single atom at each lattice site and then suppress the inter-lattice-site hopping. The Van Vleck term reads:

$$
\begin{aligned}
&\int \frac{d^3k_c}{8\pi^3}\sum_n f(\varepsilon_0(\boldsymbol{k}_c))\frac{G_{0n}G_{n0}}{\varepsilon_0-\varepsilon_n}\\
&=\int \frac{d^3k_c}{8\pi^3}\sum_n \frac{f(\varepsilon_0(\boldsymbol{k}_c))}{\varepsilon_0-\varepsilon_n}\frac{1}{N}\Bigg|\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2}\frac{1}{2}e^{i\boldsymbol{k}_c\cdot(\boldsymbol{R}_1-\boldsymbol{R}_2)}\boldsymbol{B}\cdot\\
&\Big[\langle W_n(\boldsymbol{R}_2)|\hat{\boldsymbol{V}}\times(\boldsymbol{r}-\boldsymbol{R}_1)|W_0(\boldsymbol{R}_1)\rangle\\
&-\frac{1}{N}\sum_{\boldsymbol{R}_3,\boldsymbol{R}_4}e^{i\boldsymbol{k}_c\cdot(\boldsymbol{R}_3-\boldsymbol{R}_4)}\langle W_n(\boldsymbol{R}_2)|\hat{\boldsymbol{V}}|W_0(\boldsymbol{R}_1)\rangle\times\langle W_0(\boldsymbol{R}_4)|(\boldsymbol{r}-\boldsymbol{R}_3)|W_0(\boldsymbol{R}_3)\rangle\Big]\Bigg|^2\\
&=\int \frac{d^3k_c}{8\pi^3}\sum_n \frac{f(\varepsilon_0(\boldsymbol{k}_c))}{\varepsilon_0-\varepsilon_n}\frac{1}{N}\Bigg|\sum_{\boldsymbol{R}_1}\frac{1}{2}\boldsymbol{B}\cdot\\
&[\langle W_n(\boldsymbol{R}_1)|\hat{\boldsymbol{V}}\times(\boldsymbol{r}-\boldsymbol{R}_1)|W_0(\boldsymbol{R}_1)\rangle\\
&-\frac{1}{N}\sum_{\boldsymbol{R}_3}\langle W_n(\boldsymbol{R}_1)|\hat{\boldsymbol{V}}|W_0(\boldsymbol{R}_1)\rangle\times\langle W_0(\boldsymbol{R}_3)|(\boldsymbol{r}-\boldsymbol{R}_3)|W_0(\boldsymbol{R}_3)\rangle]\Bigg|^2\\
&=\frac{N}{V}\sum_n\frac{f(\varepsilon_0)}{\varepsilon_0-\varepsilon_n}\big|\langle W_n(\boldsymbol{R})|\frac{1}{2}\boldsymbol{B}\cdot[\hat{\boldsymbol{V}}\times(\boldsymbol{r}-\boldsymbol{R})]|W_0(\boldsymbol{R})\rangle\big|^2\,, \qquad (4.6)
\end{aligned}
$$

where N is the total number of atoms and V is the sample volume. In Eq.(4.6), the first equality is simply the expansion of the Van Vleck paramangetism for a single band ε_0 in terms of the Wannier function. The second equality is derived by taking the atomic insulator limit and setting all inter-lattice-site hopping to zero: $\langle W_n(\boldsymbol{R}_2)|\hat{\boldsymbol{V}}\times(\boldsymbol{r}-\boldsymbol{R}_1)|W_0(\boldsymbol{R}_1)\rangle\propto\delta_{\boldsymbol{R}_1,\boldsymbol{R}_2}$, $\langle W_n(\boldsymbol{R}_2)|\hat{\boldsymbol{V}}|W_0(\boldsymbol{R}_1)\rangle\propto\delta_{\boldsymbol{R}_1,\boldsymbol{R}_2}$ and $\langle W_0(\boldsymbol{R}_4)|(\boldsymbol{r}-\boldsymbol{R}_3)|W_0(\boldsymbol{R}_3)\rangle\propto\delta_{\boldsymbol{R}_3,\boldsymbol{R}_4}$. The last equality accounts for the fact that for atomic insulators, we only have flat band ε_0 with no $\boldsymbol{k}_c$ dependence, which coincides with the atomic energy level. Meanwhile,

$\langle W_0(\boldsymbol{R}_3)|(\boldsymbol{r}-\boldsymbol{R}_3)|W_0(\boldsymbol{R}_3)\rangle = 0$, since the electron is bound with each atom and its position expectation should coincide with the atom position. We also use the fact that $\langle W_n(\boldsymbol{R})|\frac{1}{2}\boldsymbol{B}\cdot[\hat{\boldsymbol{V}}\times(\boldsymbol{r}-\boldsymbol{R})]|W_0(\boldsymbol{R})\rangle$ are identical for all lattice sites. It is easy to check that the final result in Eq.(4.6) is exactly the Van-Vleck susceptibility in atomic physics.

By similar derivations, we find that the Langevin term in the atomic insulator limit reads:

$$\begin{aligned}&\int\frac{d^3k_c}{8\pi^3}\sum_n f(\varepsilon_0(\boldsymbol{k}_c))\frac{f}{8m}(B^2g_{ii}-B_ig_{ij}B_j)\\ =&\frac{N}{V}\frac{f(\varepsilon_0)}{8m}\langle W_0(\boldsymbol{R})||\boldsymbol{B}\times(\boldsymbol{r}-\boldsymbol{R})|^2|W_0(\boldsymbol{R})\rangle\,,\end{aligned}\tag{4.7}$$

which is the familiar Langevin diamagnetic susceptibility in atomic systems.

Now we show that the geometrical magnetic susceptibility vanishes in the atomic insulator limit. Therefore, our formula indeed reduces to the correct result in the atomic insulator limit. The first term of the geometrical

susceptibility reads as:

$$
\begin{aligned}
&-\frac{3}{4}\int\frac{d^3k_c}{8\pi^3}f(\varepsilon_0(\boldsymbol{k}_c))(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m})\\
&=\frac{3}{8}\int\frac{d^3k_c}{8\pi^3}f(\varepsilon_0(\boldsymbol{k}_c))\frac{1}{N^2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2,\boldsymbol{R}_3,\boldsymbol{R}_4}e^{i\boldsymbol{k}_c\cdot(\boldsymbol{R}_1+\boldsymbol{R}_3-\boldsymbol{R}_2-\boldsymbol{R}_4)}\\
&(\langle W_0(\boldsymbol{R}_2)|[(\boldsymbol{r}-\boldsymbol{R}_2)\times(\boldsymbol{r}-\boldsymbol{R}_1)]\cdot\boldsymbol{B}|W_0(\boldsymbol{R}_1)\rangle\\
&\times\langle W_0(\boldsymbol{R}_4)|[\boldsymbol{B}\times(\boldsymbol{r}-\boldsymbol{R}_4)]\cdot[(\varepsilon_0-\hat{H}_c)(\boldsymbol{r}-\boldsymbol{R}_3)]|W_0(\boldsymbol{R}_3)\rangle\Big)\\
&=\frac{3}{8}\frac{f(\varepsilon_0)}{N^2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2,\boldsymbol{R}_3,\boldsymbol{R}_4}\delta_{(\boldsymbol{R}_1+\boldsymbol{R}_3-\boldsymbol{R}_2-\boldsymbol{R}_4),0}\,(\langle W_0(\boldsymbol{R}_2)|[(\boldsymbol{r}-\boldsymbol{R}_2)\times(\boldsymbol{r}-\boldsymbol{R}_1)]\cdot\boldsymbol{B}|W_0(\boldsymbol{R}_1)\rangle\\
&\times\langle W_0(\boldsymbol{R}_4)|[\boldsymbol{B}\times(\boldsymbol{r}-\boldsymbol{R}_4)]\cdot[(\varepsilon_0-\hat{H}_c)(\boldsymbol{r}-\boldsymbol{R}_3)]|W_0(\boldsymbol{R}_3)\rangle\Big)
\end{aligned}
\tag{4.8}
$$

In the above equation, the second equality uses the fact that in the atomic insulator limit, the band dispersion becomes flat. Notice that in the last line in Eq.(4.8), $(\boldsymbol{r}-\boldsymbol{R}_2)\times(\boldsymbol{r}-\boldsymbol{R}_1)=0$ if $\boldsymbol{R}_1=\boldsymbol{R}_2$. Therefore, to obtain a nonzero contribution, we must have $\boldsymbol{R}_1\neq\boldsymbol{R}_2$ and $\boldsymbol{R}_3\neq\boldsymbol{R}_4$ due to the Kronecker delta function. This part of geometrical contribution contains only inter-lattice-hopping effect, so it must vanish in the limit of atomic insulators.

Now we discuss the second term in the geometrical magnetic free energy.

$$
\begin{aligned}
&-\frac{1}{8}\int\frac{d^3k_c}{8\pi^3}f(\varepsilon_0(\boldsymbol{k}_c))\epsilon_{sik}\epsilon_{tj\ell}B_sB_tg_{ij}\alpha_{k\ell}\\
&=\frac{1}{8}\int\frac{d^3k_c}{8\pi^3}f(\varepsilon_0(\boldsymbol{k}_c))\epsilon_{sik}\epsilon_{tj\ell}B_sB_t\frac{1}{N^2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2,\boldsymbol{R}_3,\boldsymbol{R}_4}e^{i\boldsymbol{k}_c\cdot(\boldsymbol{R}_1+\boldsymbol{R}_3-\boldsymbol{R}_2-\boldsymbol{R}_4)}(\boldsymbol{R}_3-\boldsymbol{R}_4)_k(\boldsymbol{R}_3-\boldsymbol{R}_4)_\ell\\
&\langle W_0(\boldsymbol{R}_4)|\hat{H}_c|W_0(\boldsymbol{R}_3)\rangle\Big(\langle W_0(\boldsymbol{R}_2)|(\boldsymbol{r}-\boldsymbol{R}_2)_i(\boldsymbol{r}-\boldsymbol{R}_1)_j|W_0(\boldsymbol{R}_1)\rangle\\
&-\frac{1}{N}\sum_{\boldsymbol{R}_3,\boldsymbol{R}_4}e^{i\boldsymbol{k}_c\cdot(\boldsymbol{R}_5-\boldsymbol{R}_6)}\langle W_0(\boldsymbol{R}_2)|(\boldsymbol{r}-\boldsymbol{R}_1)_j|W_0(\boldsymbol{R}_1)\rangle\langle W_0(\boldsymbol{R}_6)|(\boldsymbol{r}-\boldsymbol{R}_5)_i|W_0(\boldsymbol{R}_5)\rangle\Big)\\
&=\frac{1}{8}\frac{f(\varepsilon_0)}{N^2}\epsilon_{sik}\epsilon_{tj\ell}B_sB_t\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2,\boldsymbol{R}_3,\boldsymbol{R}_4}\delta_{(\boldsymbol{R}_1+\boldsymbol{R}_3-\boldsymbol{R}_2-\boldsymbol{R}_4),0}\\
&(\boldsymbol{R}_3-\boldsymbol{R}_4)_k(\boldsymbol{R}_3-\boldsymbol{R}_4)_\ell\langle W_0(\boldsymbol{R}_4)|\hat{H}_c|W_0(\boldsymbol{R}_3)\rangle\langle W_0(\boldsymbol{R}_2)|(\boldsymbol{r}-\boldsymbol{R}_2)_i(\boldsymbol{r}-\boldsymbol{R}_1)_j|W_0(\boldsymbol{R}_1)\rangle\,.
\end{aligned}
\tag{4.9}
$$

Due to the prefactor $(\boldsymbol{R}_3-\boldsymbol{R}_4)$, it is obvious that this remaining contribution to the geometrical magnetic free energy also contains only inter-lattice-hopping effect, so it also vanishes in the limit of atomic insulators.

Based on the above Wannier function representation of the magnetic susceptibility, it is interesting to notice that the Fermi sea contribution to the orbital susceptibility is due to two types of effects, i.e. intra-lattice-site transition similar to that in the atomic physics and inter-lattice-site hopping which is unique in crystalline solids. Our classification in Eq.(5.52) thus provides a reasonable extrapolation of the orbital susceptibility from atomic systems to crystalline solids: on one hand, the Van Vleck paramagnetic susceptibility and Langevin magnetic susceptibility reduce to their counterparts in atomic systems and the geometrical magnetic susceptibility vanishes in the limit of atomic insulators; on the other hand, even though they all contain inter-lattice-

site hopping contribution, the Van Vleck paramagnetic susceptibility and the Langevin magnetic susceptibility in solids still preserve their essential properties established in the atomic systems, i.e. the Van Vleck paramagnetic susceptibility is always paramagnetic and depends on the energy interval between two electronic states and the Langevin magnetic susceptibility is diamagnetic along directions that diagonalize the quantum metric. Furthermore, we emphasize that the geometrical magnetic susceptibility is indeed unique in crystalline solids and is a novel mechanism of orbital susceptibility that only depends on the geometrical quantities in k-space.

In the above we have identified six different mechanisms to the orbital magnetic susceptibility. Here we will argue that they can dominate in different scenarios. In atomic systems, the Van Vleck paramagnetic susceptibility is generally small due to the large separation between electronic levels, although it has a notable exception in the atomic Lanthanide series, where the energy interval between the ground state and the first excited states is small. In solids, different energy levels can be very close, e.g. near the topological transition points. In this case, the Van Vleck paramagnetic susceptibility become large. However, we will show that in the circumstance where the linear band crossing occurs, it is another contribution, i.e. the geometrical contribution, that dominates over other orbital susceptibilities. The Langevin magnetic susceptibility in atomic systems is also small and only important for close shell atoms. Likewise, in solids the Langevin magnetic susceptibility is usually discussed for insulators where electrons are localized. However, our theory suggests that

it is connected to the intrinsic expansion of the localized wave-packet, i.e. the quantum metric. Therefore, for both insulators and metals, the Langevin magnetism is well defined and can be sizable as discussed later. It can even dominate over other susceptibilities as in the continuum model of the double layer graphene.

In conclusion, within the wave-packet semiclassical approach, since the Bloch electron energy is derived to second order in the magnetic field and classified into gauge-invariant terms with clear physical meaning as in Eq.(2.23), it yields a fresh understanding of the complex behavior of orbital magnetic susceptibility. We are able to answer the following questions: how the intrinsic geometry of Bloch bands affects the second order response to electromagnetic fields, and whether or not there are additional geometrical quantities emerging in the orbital magnetic susceptibility. We identify a geometrical contribution from the Fermi sea, which we call the geometrical susceptibility, in the sense that it involves geometrical quantities including the Berry curvature and the quantum metric. The geometrical susceptibility is a novel mechanism for orbital magnetic susceptibility, which provides the dominant diamagnetic response around the band gaps, and is especially important in strongly spin/pseudospin-orbit coupled systems such as topological insulators and 2D semimetals. We will present examples later.

Moreover, we derive a novel Fermi surface contribution, arising from the energy polarization in the Brillouin zone, and competing with Pauli and Peierls-Landau magnetism. To our delight, these Fermi surface contributions,

together with a Langevin-like magnetic susceptibility and the geometrical susceptibility, can be calculated based on Bloch states inside a single Bloch band, and the only interband contribution is the Van Vleck paramagnetic susceptibility. The various terms can dominate over different energy range of the Bloch band, as illustrated later. We emphasize that the above understanding of the orbital magnetic susceptibility is under the assumption of the minimal coupling, in which case the magnetic field modifies the Hamiltonian only through the magnetic vector potential. The generalization beyond the minimal coupling assumption will be discussed later.

4.3 Example I: Gapped Dirac Model

The first example we choose is the gapped Dirac model. The model Hamiltonian is given by:

$$H = v_f(k_1\sigma_1 + k_2\sigma_2) + \Delta\sigma_3\,, \tag{4.10}$$

where v_f is the Fermi velocity, and σ_1, σ_2, and σ_3 are Pauli matrices. Using Eq.(5.52) to calculate the susceptibility, we find that the energy polarization, the Van-Vleck, and the Langevin susceptibility all vanish. So the only three non-trivial contributions are:

$$\begin{aligned}
\mu_{PL} &= \frac{\Delta^2 v_f^2}{8\pi|\mu|^3} \\
\mu_{LD} &= -\frac{\Delta^2 v_f^2}{24\pi|\mu|^3} \\
\mu_{CP} &= -\frac{\Delta^2 v_f^2}{12\pi|\mu|^3}\,,
\end{aligned} \tag{4.11}$$

Here we assume the chemical potential μ falls inside the band (either valence or conduction band). μ_{PL}, μ_{LD}, and μ_{CP} are the Pauli, Peierls-Landau, and the geometrical susceptibility, respectively. One can see that these three contributions exactly cancel each other. The total susceptibility is 0 when μ falls inside the band. However, when the Fermi energy falls in the gap, there will be a diamagnetic susceptibility due to the geometrical susceptibility, and the Pauli and Peierls-Landau susceptibility vanish since there are no Fermi surfaces.

An interesting fact is that the susceptibility satisfies the following sum rule: $\int \chi d\mu = -\frac{v_f^2}{6\pi}$, which does not depend on the value of the band gap. At the limit of $\Delta \to 0$, the susceptibility when μ falls in the gap becomes larger and larger, but it keeps zero when μ is in the band. Therefore, it gives a delta-function type behavior for the susceptibility at $\Delta = 0$.

4.4 Example II: Double Layer Graphene

As the second example of our theory, we calculate the magnetic susceptibility in the low energy model of the double layer graphene. The model Hamiltonian is given by

$$h = \begin{pmatrix} \Delta & -\frac{(k_1 - ik_2)^2}{2m} \\ -\frac{(k_1 + ik_2)^2}{2m} & -\Delta \end{pmatrix}, \tag{4.12}$$

where Δ gives the band gap, and m is the effective mass.

We will first calculate the magnetic susceptibility from the Landau level spectrum. The Landau level in the conduction band for this low energy model

is

$$\varepsilon_{LL} = \sqrt{\Delta^2 + \frac{B^2}{m^2}(n-1/2)^2 - \frac{B^2}{4m^2}}\,, \tag{4.13}$$

where $n = 0, 1, 2, \cdots$.

We set $x = (n-1/2)B$, then the free energy is

$$\begin{aligned} F &= -\frac{1}{2\pi\beta}\sum_n \ln(1 + e^{2\beta\mu} + e^{\beta\mu}(e^{\beta\varepsilon_{LL}} + e^{-\beta\varepsilon_{LL}}) \\ &= -\frac{1}{2\pi\beta}\int_{-B}^{\infty} \ln(1 + e^{2\beta\mu} + e^{\beta\mu}(e^{\beta\varepsilon_{LL}} + e^{-\beta\varepsilon_{LL}})dx + F_{LD} \\ F_{LD} &= \frac{B^2}{48\pi\beta}\frac{\beta e^{\beta\mu}(e^{\beta\varepsilon_{LL}} - e^{-\beta\varepsilon_{LL}})}{1 + e^{2\beta\mu} + e^{\beta\mu}(e^{\beta\varepsilon_{LL}} + e^{-\beta\varepsilon_{LL}})}\frac{x/m^2}{\sqrt{\Delta^2 + x^2/m^2}}\Big|_{x=0}^{x=\infty} = \frac{B^2}{48\pi m}\,. \end{aligned} \tag{4.14}$$

In the above equation, we have used the Euler-Maclaurin formula to transform the summation over the Landau level index to the integration over the continuous variable x. We only keep terms up to second order with respect to B, since the susceptibility is a second order response function, and all higher order terms do not contribute.

Therefore, the magnetic susceptibility is

$$\begin{aligned}
\chi &= -\frac{\partial^2 F}{\partial B^2}\Big|_{B=0} \\
&= \frac{1}{2\pi}\int_0^\infty \frac{e^{\beta\mu}(e^{\beta\varepsilon_{LL}} - e^{-\beta\varepsilon_{LL}})}{1 + e^{2\beta\mu} + e^{\beta\mu}(e^{\beta\varepsilon_{LL}} + e^{-\beta\varepsilon_{LL}})}\frac{\partial^2\varepsilon}{\partial B^2}\Big|_{B=0} dx - \frac{1}{24\pi m} \\
&= -\frac{1}{8\pi m^2}\int_0^\infty \frac{1}{\varepsilon}\frac{e^{\beta\mu}(e^{\beta\varepsilon_{LL}} - e^{-\beta\varepsilon_{LL}})}{1 + e^{2\beta\mu} + e^{\beta\mu}(e^{\beta\varepsilon_{LL}} + e^{-\beta\varepsilon_{LL}})} dx - \frac{1}{24\pi m} \\
&= -\frac{1}{8\pi m^2}\int_{x_0}^\infty \frac{1}{\varepsilon} dx - \frac{1}{24\pi m} \\
&= -\frac{1}{8\pi m}\int_{\theta_0}^{\pi/2} \frac{1}{\Delta/\cos\theta}\frac{\Delta}{\cos^2\theta} d\theta - \frac{1}{24\pi m} \\
&= -\frac{1}{16\pi m}\ln\frac{1+\sin\theta}{1-\sin\theta}\Big|_{\theta_0}^{\pi/2} - \frac{1}{24\pi m} \\
&= -\frac{1}{16\pi m}\ln\frac{1+\sin\theta_\Lambda}{1-\sin\theta_\Lambda} + \frac{1}{16\pi m}\ln\frac{1+\sin\theta_0}{1-\sin\theta_0} - \frac{1}{24\pi m}\,, \qquad (4.15)
\end{aligned}$$

where $x = \Delta m \tan\theta$, θ_Λ corresponds to high energy cut-off, and $\Delta = |\mu|\cos\theta_0$.

Now I will calculate the susceptibility for the conduction band from the semiclassical theory. Note that in the generic 2-band model $\hat{H} = h_1\sigma_1 + h_2\sigma_2 + h_3\sigma_3$, we have

$$\begin{aligned}
h_1 &= -\frac{k_1^2 - k_2^2}{2m} \\
h_2 &= -\frac{k_1 k_2}{m} \\
h_3 &= \Delta\,. \qquad (4.16)
\end{aligned}$$

Therefore, the wave function for the valence band is

$$\psi = \frac{1}{\sqrt{2\varepsilon_0(\varepsilon_0 - \Delta)}}\begin{pmatrix} h_3 - \varepsilon_0 \\ h_1 + ih_2 \end{pmatrix}\,. \qquad (4.17)$$

The Hessian matrix of the Hamiltonian are:

$$\Gamma_{11} = -\frac{1}{m}\sigma_1\,, \quad \Gamma_{22} = -\Gamma_{11}\,, \quad \Gamma_{12} = -\frac{1}{m}\sigma_2\,. \tag{4.18}$$

Then the expectation of the Hessian matrix in the valence band is

$$\langle\psi|\Gamma_{11}|\psi\rangle = \frac{h_1}{m\varepsilon_0} = -\langle\psi|\Gamma_{22}|\psi\rangle\,, \quad \langle\psi|\Gamma_{12}|\psi\rangle = \frac{h_2}{m\varepsilon_0}\,. \tag{4.19}$$

Now we calculate a few important quantities used in Eq.(5.52) for the magnetic susceptibility, including the Berry curvature Ω, the magnetic moment m_o, the density of states $g(\varepsilon_0)$ at the energy ε_0, the inverse effective mass tensor α_{ij}, the quantum metric g_{ij}. The results read:

$$\begin{aligned}
\Omega &= -\frac{\Delta k^2}{2m^2\varepsilon_0^3} \\
m_0 &= \frac{\Delta k^2}{2m^2\varepsilon_0^2} \\
g(\varepsilon_0) &= \frac{1}{2\pi}\frac{m\varepsilon_0}{\sqrt{\varepsilon_0^2-\Delta^2}} \\
\alpha_{11} &= \frac{k^2}{2m^2\varepsilon_0} + \frac{\Delta^2 k_1^2}{m^2\varepsilon_0^3} \\
\alpha_{22} &= \frac{k^2}{2m^2\varepsilon_0} + \frac{\Delta^2 k_2^2}{m^2\varepsilon_0^3} \\
\alpha_{12} &= \frac{\Delta^2 k_1 k_2}{m^2\varepsilon_0^3} \\
g_{11} &= \frac{k^2}{4m^2\varepsilon_0^2}\left(1 - \frac{k^2 k_1^2}{4m^2\varepsilon_0^2}\right) \\
g_{22} &= \frac{k^2}{4m^2\varepsilon_0^2}\left(1 - \frac{k^2 k_2^2}{4m^2\varepsilon_0^2}\right) \\
g_{12} &= \frac{-k^4 k_1 k_2}{16m^4\varepsilon_0^4}\,.
\end{aligned} \tag{4.20}$$

From the above quantities, various susceptibilities can be easily calculated. The Pauli paramagnetic free energy is

$$F_{PL} = -\frac{1}{2}g(\mu)B^2 m_0^2 \Rightarrow \chi_{PL} = \frac{\Delta^2\sqrt{\mu^2-\Delta^2}}{2\pi m|\mu|^3}\,. \tag{4.21}$$

The Landau diamagnetic free energy is

$$\begin{aligned}
F_{LD} &= \frac{1}{24}B^2(\alpha_{11}\alpha_{22}-\alpha_{12}^2)g(\mu) \\
\Rightarrow \chi_{LD} &= -\frac{1}{12}\frac{m|\mu|}{2\pi\sqrt{\mu^2-\Delta^2}}\left[\left(\frac{k^2}{2m^2\varepsilon_0}+\frac{k_1^2\Delta^2}{m^2\varepsilon_0^3}\right)\left(\frac{k^2}{2m^2\varepsilon_0}+\frac{k_2^2\Delta^2}{m^2\varepsilon_0^3}\right)-\frac{\Delta^4k_1^2k_2^2}{m^4\varepsilon_0^6}\right] \\
&= -\frac{1}{24\pi}\frac{m|\mu|}{\sqrt{\mu^2-\Delta^2}}\frac{\mu^2-\Delta^2}{m^2\mu^2}\left(1+\frac{2\Delta^2}{\mu^2}\right) \\
&= -\frac{1}{24\pi m}\frac{\sqrt{\mu^2-\Delta^2}}{|\mu|}\left(1+\frac{2\Delta^2}{\mu^2}\right)\,.
\end{aligned} \tag{4.22}$$

The geometrical susceptibility due to the coupling of the density of states and the magnetic dipole moment is:

$$\begin{aligned}
\chi_{CP} &= 2\int \Omega m\frac{d^2k}{4\pi^2} \\
&= -\int\frac{\Delta^2k^4}{2m^4\varepsilon_0^5}\frac{dk^2}{4\pi} \\
&= -\int\frac{\Delta^2}{\pi m\varepsilon_0^5}\frac{k^4}{4m^2}d\frac{k^2}{2m} \\
&= -\int_{\theta_0}^{\pi/2}\frac{\Delta^2\Delta^2\tan^2\theta}{\pi m\Delta^5/\cos^5\theta}\frac{\Delta}{\cos^2\theta}d\theta \\
&= -\frac{1}{\pi m}\int_{\theta_0}^{\pi/2}\sin^2\theta\cos\theta d\theta \\
&= -\frac{1}{3\pi m}(1-\sin^3\theta_0) \\
&= -\frac{1}{3\pi m}\left(1-\frac{(\mu^2-\Delta^2)\sqrt{\mu^2-\Delta^2}}{|\mu|^3}\right)\,.
\end{aligned} \tag{4.23}$$

Now we calculate the second order energy. The geometrical energy from the Berry curvature is:

$$\frac{1}{4}(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m}) = -\frac{1}{16}\frac{B^2\Delta^2k^4}{m^4\varepsilon_0^5} = -\frac{1}{4}\frac{B^2\Delta^2(\varepsilon_0^2-\Delta^2)}{m^2\varepsilon_0^5}\,. \tag{4.24}$$

The geometrical term from the quantum metric is (notice that the effective mass would pick up a minus sign for the valence band):

$$\begin{aligned}
-\frac{1}{8}\varepsilon_{\theta i\alpha}\varepsilon_{\phi j\beta}B_\theta B_\phi\alpha_{ij}g_{\alpha\beta} &= -\frac{1}{8}(\alpha_{11}g_{22}+\alpha_{22}g_{11}-2\alpha_{12}g_{12})B^2 \\
&= \frac{1}{8}\frac{B^2k^2}{4m^2\varepsilon_0^2}\left[\left(1-\frac{k^2k_1^2}{4m^2\varepsilon_0^2}\right)\left(\frac{k^2}{2m^2\varepsilon_0}+\frac{\Delta^2k_2^2}{m^2\varepsilon_0^3}\right)\right. \\
&\qquad \left.+\left(1-\frac{k^2k_2^2}{4m^2\varepsilon_0^2}\right)\left(\frac{k^2}{2m^2\varepsilon_0}+\frac{\Delta^2k_1^2}{m^2\varepsilon_0^3}\right)+\frac{k^2k_1k_2}{2m^2\varepsilon_0^2}\frac{\Delta^2k_1k_2}{m^2\varepsilon_0^3}\right] \\
&= \frac{1}{8}\frac{B^2k^2}{4m^2\varepsilon_0^2}\frac{k^2}{2m^2\varepsilon_0}\left(1+\frac{3\Delta^2}{\varepsilon_0^2}\right) \\
&= \frac{B^2}{16m^2}\frac{\varepsilon_0^2-\Delta^2}{\varepsilon_0^3}\left(1+\frac{3\Delta^2}{\varepsilon_0^2}\right)\,.
\end{aligned} \tag{4.25}$$

The first Heissian free energy is (there is a sign difference in the expecations of the Heissian for the conduction and valence band):

$$\begin{aligned}
\frac{1}{8}\sum_{(m,n)\neq 0}(\boldsymbol{B}\times\boldsymbol{A}_{0m})_i(\Gamma_{ij})_{mn}(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j &= \frac{1}{8}\varepsilon_{\theta i\alpha}\varepsilon_{\phi j\beta}B_\theta B_\phi(\Gamma_{ij})_n g_{\alpha\beta} \\
&= \frac{1}{8}((\Gamma_{11})_n g_{22}+(\Gamma_{22})_n g_{11}-2(\Gamma_{12})_n g_{12})B^2 \\
&= \frac{1}{8}\frac{B^2k^2}{4m^2\varepsilon_0^2}\left[\left(1-\frac{k^2k_1^2}{4m^2\varepsilon_0^2}\right)\frac{h_1}{m\varepsilon_0}\right. \\
&\qquad \left.-\left(1-\frac{k^2k_2^2}{4m^2\varepsilon_0^2}\right)\frac{h_1}{m\varepsilon_0}-2\frac{k^2k_1k_2}{4m^2\varepsilon_0^2}\frac{h_2}{m\varepsilon_0}\right] \\
&= \frac{1}{8}\frac{B^2k^2}{4m^2\varepsilon_0^2}\left(\frac{k^2h_1^2}{2m^2\varepsilon_0^3}+\frac{k^2h_2^2}{2m^2\varepsilon_0^3}\right) \\
&= \frac{B^2}{16}\frac{(\varepsilon_0^2-\Delta^2)^2}{m^2\varepsilon_0^5}\,.
\end{aligned} \tag{4.26}$$

Now we will calculate the second Hessian free energy:

$$
\begin{aligned}
-\frac{1}{16}(\boldsymbol{B}\times\boldsymbol{\partial})_i(\boldsymbol{B}\times\boldsymbol{\partial})_j\langle 0|\Gamma_{ij}|0\rangle &= -\frac{B^2}{16}(\partial_1\partial_1(\Gamma_{22})_0+\partial_2\partial_2(\Gamma_{11})_)-2\partial_1\partial_2(\Gamma_{12})_0)\\
&= -\frac{B^2}{16}\partial_2\left(\frac{k_2}{m^2\varepsilon_0}-\frac{h_1}{m\varepsilon_0^2}\frac{k^2k_2}{2m^2\varepsilon_0}\right)\\
&\quad -\frac{B^2}{16}\partial_1\left(\frac{k_1}{m^2\varepsilon_0}+\frac{h_1}{m\varepsilon_0^2}\frac{k^2k_1}{2m^2\varepsilon_0}\right)\\
&\quad +\frac{B^2}{16}2\partial_1\left(-\frac{k_1}{m^2\varepsilon_0}-\frac{h_2}{m\varepsilon_0^2}\frac{k^2k_2}{2m^2\varepsilon_0}\right)\\
&= -\frac{B^2}{4m^2\varepsilon_0}\left(1-\frac{k^4}{4m^2\varepsilon_0^2}-\frac{3k^4}{8m^2\varepsilon_0^2}+\frac{3k^4}{8m^2\varepsilon_0^4}\frac{k^4}{4m^2}\right)\\
&= -\frac{\Delta^2B^2}{4m^2\varepsilon_0^3}\left(1-\frac{3(\varepsilon_0^2-\Delta^2)}{2\varepsilon_0^2}\right)\\
&= -\frac{\Delta^2B^2}{4m^2\varepsilon_0^3}\frac{3\Delta^2-\varepsilon_0^2}{2\varepsilon_0^2}\,. \qquad (4.27)
\end{aligned}
$$

Now we will sum up all the second order energy:

$$
\begin{aligned}
I &= -\frac{\Delta^2B^2}{4m^2\varepsilon_0^3}\frac{3\Delta^2-\varepsilon_0^2}{2\varepsilon_0^2}+\frac{B^2}{16}\frac{(\varepsilon_0^2-\Delta^2)^2}{m^2\varepsilon_0^5}\\
&\quad +\frac{B^2}{16m^2}\frac{\varepsilon_0^2-\Delta^2}{\varepsilon_0^3}\left(1+\frac{3\Delta^2}{\varepsilon_0^2}\right)-\frac{1}{4}\frac{B^2\Delta^2(\varepsilon_0^2-\Delta^2)}{m^2\varepsilon_0^5}\\
&= \frac{B^2}{8m^2\varepsilon_0}-\frac{\Delta^2B^2(\varepsilon_0^2+2\Delta^2)}{8m^2\varepsilon_0^5}\,. \qquad (4.28)
\end{aligned}
$$

The corresponding susceptibility is

$$
\begin{aligned}
\chi_I &= -\int\frac{1}{4m^2\varepsilon_0}\frac{dk^2}{4\pi}+\int\frac{\Delta^2(\varepsilon_0^2+2\Delta^2)}{4m^2\varepsilon_0^5}\frac{dk^2}{4\pi}\\
&= -\frac{1}{8m\pi}\int\frac{1}{\Delta/\cos\theta}\frac{\Delta}{\cos^2\theta}d\theta+\frac{1}{8m\pi}\int\frac{\Delta^2(\Delta^2\tan^2\theta+3\Delta^2)}{\Delta^5/\cos^5\theta}\frac{\Delta}{\cos^2\theta}d\theta\\
&= -\frac{1}{16\pi m}\ln\frac{1+\sin\theta}{1-\sin\theta}\Big|_{\theta_0}^{\theta_\Lambda}+\frac{1}{24\pi m}(1-\sin^3\theta_0)+\frac{3}{8\pi m}\left(1-\sin\theta_0-\frac{1}{3}(1-\sin^3\theta_0)\right)\,. \qquad (4.29)
\end{aligned}
$$

Combining Eq.(4.21), (4.22), (4.23) and (4.29), we obtain the total susceptibility:

$$\begin{aligned}
\chi &= \frac{1}{2\pi m}\sin\theta_0\cos\theta_0^2 - \frac{1}{24\pi m}\sin\theta_0(1+2\cos^2\theta_0) - \frac{1}{3\pi m}(1-\sin^3\theta_0) \\
&\quad - \frac{1}{16\pi m}\ln\frac{1+\sin\theta_\Lambda}{1-\sin\theta_\Lambda} + \frac{1}{16\pi m}\ln\frac{1+\sin\theta_0}{1-\sin\theta_0} \\
&\quad + \frac{1}{24\pi m}(1-\sin^3\theta_0) + \frac{3}{8\pi m}\left(1-\sin\theta_0 - \frac{1}{3}(1-\sin^3\theta_0)\right) \\
&= -\frac{1}{16\pi m}\ln\frac{1+\sin\theta_\Lambda}{1-\sin\theta_\Lambda} + \frac{1}{16\pi m}\ln\frac{1+\sin\theta_0}{1-\sin\theta_0} \\
&\quad - \frac{1}{3\pi m} + \frac{1}{24\pi m} + \frac{3}{8\pi m} - \frac{1}{8\pi m} \\
&\quad + \frac{1}{2\pi m}\sin\theta_0 - \frac{1}{24\pi m}\sin\theta_0 - \frac{1}{12\pi m}\sin\theta_0 - \frac{3}{8\pi m}\sin\theta_0 \\
&\quad - \frac{1}{2\pi m}\sin^3\theta_0 + \frac{1}{12\pi m}\sin^3\theta_0 + \frac{1}{3\pi m}\sin^3\theta_0 - \frac{1}{24\pi m}\sin^3\theta_0 + \frac{1}{8\pi m}\sin^3\theta_0 \\
&= -\frac{1}{16\pi m}\ln\frac{1+\sin\theta_\Lambda}{1-\sin\theta_\Lambda} + \frac{1}{16\pi m}\ln\frac{1+\sin\theta_0}{1-\sin\theta_0} - \frac{1}{24\pi m}\,. \qquad (4.30)
\end{aligned}$$

This is the same result as obtained directly from the quantum mechanics.

4.5 Example III: Tight-binding Graphene

Our theory of magnetic susceptibility is not only valid for low energy models, but also applicable in various realistic models, such as from first principle calculations or the tight-binding approximations. Various terms in Eq.(5.52) can readily be evaluated in first principle calculations. As a concrete example to show how various terms contribute to the total magnetic susceptibility and the importance of the geometrical magnetic susceptibility, we now

consider the following tight-binding model defined on a honeycomb lattice [52]:

$$\hat{H} = -t \sum_{\langle i,j \rangle} c_i^\dagger c_j - t' \sum_{\langle\langle i,j \rangle\rangle} c_i^\dagger c_j + \Delta \sum_i \xi_i c_i^\dagger c_i \,, \tag{4.31}$$

where $c_i(c_i^\dagger)$ is the electron annihilation (creation) operator on site i, the first and the second terms are the nearest-neighbor and second-neighbor hopping terms, the third term is a staggered potential with $\xi_i = \pm 1$ for the two sublattices, and t, t' and Δ are the strengths of the terms. The staggered potential breaks the inversion symmetry and generates a gap of 2Δ in the spectrum. The second-neighbor hopping is introduced for breaking the particle-hole symmetry.

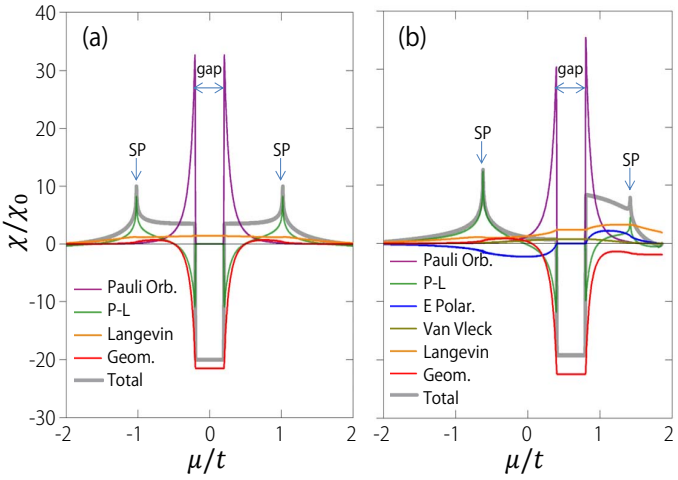

Figure 4.1: (color online) Orbital magnetic susceptibility for the lattice model (4.31) as a function of μ. χ is in units of $\chi_0 = e^2a^2t/(4\pi^2\hbar^2)$, a is the bond length. Here $\Delta = 0.2t$, (a) $t' = 0$ and (b) $t' = 0.1t$. Here P-L, E Polar, and SP stand for the Peierls-Landau, energy-polarization, and saddle point, respectively.

The various contributions to the orbital magnetic susceptibility are plotted in Fig.4.1 with (a) and without (b) particle-hole symmetry. We first consider the simple case with $t' = 0$, such that the particle-hole symmetry is preserved. Then the energy polarization and the Van Vleck contributions vanish identically, since G_{n0} vanishes in any two band system with particle-hole symmetry. From Fig.4.1(a), one observes that in the gap the Fermi surface terms vanish and the geometrical magnetic susceptibility dominates, which leads to a large diamagnetic response. The magnitude of the geometrical magnetic susceptibility decreases rapidly away from the gap and it (along with Peierls-Landau term) is compensated largely by the Pauli orbital paramagnetic susceptibility which is peaked at the band edges where $\boldsymbol{m}$ takes its largest value [9]. Two noticeable paramagnetic peaks are observed around the band saddle points due to the Peierls-Landau contribution, which is a general feature as discussed before [78]. Further away from the gap region, the susceptibility decreases gradually to zero. Our result of χ agrees with that from the exact quantum treatment [64].

The physics around gap can be described by the gapped Dirac model $\hat{h} = vk_1\sigma_1 + vk_2\sigma_2 + \Delta\sigma_3$ with $v = 3at/2$ where a is the nearest-neighbor bond length [52]. This model is widely used in the study of graphene, MoS_2, topological insulator surfaces and thin films [52, 85, 60, 28]. Here σ's are the Pauli matrices. Near the band edge ($|\mu| > \Delta$), the three competing magnetic susceptibilities have already been given in section 3, and the results can well describe the susceptibility near the band edge. Moreover, we emphasize

that for systems with two valleys connected by time reversal operation, such as graphene or MoS_2, the geometrical magnetic susceptibilities from the two valleys have the same sign. Note that in this low energy model the total susceptibility vanishes identically outside gap, which seems contradicting to the result in Fig.4.1(a) where one sees a finite paramagnetic plateau. The difference is due to two Fermi sea contributions including the Langevin and a term in geometrical magnetic susceptibility resulting from the nonzero Hessian Γ_{ij} (whereas Γ_{ij} vanishes in the low energy model) which produce an overall shift of χ. This was known as the "lattice contribution" in previous studies [25], since it vanishes at the limit of $a \to 0$ and $t \to \infty$ with $v_f = ta$ a constant. This agreement demonstrates the validity of our theory. We comment that it is easy to show that the magnitude of this paramagnetic plateau does not depend on the band gap. The value is same when the gap closes.

We point out that if one starts from the Pauli or Schrödinger Hamiltonian, the Langevin magnetic susceptibility is diamagnetic as discussed previously. But for an effective Hamiltonian as given in Eq.(4.31), the Langevin magnetic susceptibility can be paramagnetic as shown in Fig.4.1. Therefore, this paramagnetic plateau is sizable as in Fig.4.1 only when these two bands are well separated from other bands such that the tight binding model is an appropriate approximation.

Now we consider the case with a finite t', such that the particle-hole symmetry is broken. The results are shown in Fig.4.1(b), the geometrical magnetic susceptibility still dominates in the gap and one notes that the para-

magnetic plateau near gap is suppressed in the valence band but enhanced in the conduction band. Now the energy polarization and the Van Vleck contributions are finite due to a finite G_{n0}. With current parameters, the Van Vleck paramagnetic susceptibility takes a small value around the gap region, while the energy polarization contribution takes opposite signs between the two bands. The energy polarization term, along with the enhanced Langevin and Pauli terms are the main contributions to a large paramagnetic response between the conduction band edge and the saddle point. This is different from the usual orbital paramagnetic susceptibility resulting from the Peierls-Landau contribution [78]. In fact, the contribution from Peierls-Landau is less important even in the region near and above the saddle point in conduction band. The total susceptibility there is more affected by the competition between geometrical, Langevin, and energy polarization terms. This is in contrast to valence band where the Peierls-Landau dominates while other contributions are suppressed.

This example illustrates that: (1) the geometrical magnetic susceptibility is an important contribution, especially around the band gap; (2) different terms in our classification dominate over different energy ranges; (3) it is possible to enhance the paramagnetic susceptibility by breaking the particle-hole symmetry. In addition, we note that for a generic two band model $\hat{h} = h_0 + \boldsymbol{h} \cdot \boldsymbol{\sigma}$, $G_{n0} = -\boldsymbol{B} \cdot (\boldsymbol{\partial} h_0 \times \boldsymbol{A}_{n0})$ is finite when particle-hole symmetry is broken. From Eq.(5.52), we find that only the Van Vleck paramagnetic susceptibility depends quadratically on $\boldsymbol{\partial} h_0$, while all other terms are either

independent or only have linear dependence. This implies that Van Vleck susceptibility can in principle dominate for large $|\boldsymbol{\partial} h_0|$, which also leads to a strong paramagnetic response.

4.6 Discussion

Our theory could also be generalized beyond the minimal coupling. Based on the Foldy-Wouthuysen transfromation in solid state context, the correction beyond minimal coupling to the Hamiltonian is [21, 6]: $\Delta\hat{H} = \boldsymbol{B}\cdot\hat{\boldsymbol{\mu}} + B_i\hat{h}_{ij}B_j$, where $\hat{\boldsymbol{\mu}}$ and $\hat{h}_{ij}$ are appropriate matrix operators as functions of $(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)$. Then the expectation value of the second term directly adds to the second order wave-packet energy, while the first term modifies the orbital magnetic moment. For example, $\hat{\boldsymbol{\mu}}$ could stand for the electron spin magnetic moment, whose diagonal and off-diagonal parts add to the corresponding orbital moment $\boldsymbol{m}$ and G_{n0}, respectively.

In summary, based on the compact gauge-invariant expression of the Bloch wave-packet energy as in Eq.(2.23) which is correct to second order in external magnetic field, we obtain a complete and compact formula for the orbital magnetic susceptibility, with important advantage that each term is gauge invariant and has clear physical meanings. We find that other than the familiar Pauli and Peierls-Landau magnetic susceptibilities, the orbital susceptibility in solids also consists of the Van Vleck paramagnetic susceptibility and Langevin magnetic susceptibility, which reduce to their counterparts in atomic physics in the limit of atomic insulators. More importantly, we identify

two new contributions: the geometrical magnetic susceptibility derived from the Berry curvature and the quantum metric which can dominate in a small energy gap, and the k-space energy polarization magnetic susceptibility, which competes with Peierls-Landau and Pauli magnetic susceptibility on a Fermi surface. These two susceptibilities will vanish in the atomic insulator limit, so they are unique in the solid-state context. We illustrate that the Pauli, Peierls-Landau and geometrical magnetic susceptibility depend solely on band geometrical quantities and affect the orbital susceptibility greatly.

Chapter 5

Landau Level Quantization: Generalization of Onsager's Rule

5.1 Onsager's Rule and Berry Phase Correction

Ever since the discovery of the quantum behavior from the black body radiation, a fascinating subject is the connection between the quantum mechanics and the classical mechanics. Bohr and Sommerfeld established the famous Bohr-Sommerfeld rule to generate the discrete energy levels in hydrogen atoms from the classical orbits with continuously varying energies. Then Einstein envisioned a generalized version of the quantization rule for an integrable dynamical system. He argued that the dynamics is actually described by a series of invariant tories, which are determined by various action variables. In the solid-state context, quantization of electronic states into Landau levels by a magnetic field has particular importance, since it can cause oscillations in magnetization (de Haas-van Alphen effect) and conductivity (Shubnikov-de Haas effect), which contains a wealth of information about the band structure and geometric properties of Bloch states. Onsager translated the Bohr-Sommerfeld rule into the Onsager's rule [57] in the solid-state context, which

reads:

$$S = 2\pi \left(n + \frac{1}{2} \right) \frac{eB}{\hbar} \,, \tag{5.1}$$

where S is the area enclosed by the semiclassical orbits in k-space. The shape of Fermi surfaces of metals and semiconductors was obtained in this way.

However, later it was realized that this quantization rule has the following corrections:

$$S = 2\pi \left(n + \frac{1}{2} + \gamma \right) \frac{eB}{\hbar} \,, \tag{5.2}$$

where γ is some additional shift of the level index that depends on the magnetic field and the constant of motion. The first order correction in γ has been found to relate to the Berry phase. A good example is the Landau level near the Dirac point in 2D monolayer graphene. The conduction band carries a uniform Berry phase π, corresponding to $\gamma = -1/2$. Therefore, the Landau level starts right from the Dirac point, where the energy is zero, different from the Landau level of free electrons, which has a nontrivial zero-point energy. This shift of Landau level index can be used to determine the Berry phase in such materials, and the Berry phase in graphene has already been measured in this way.

However, the Berry phase correction can only yield the correct Landau levels up to first order, not to arbitrary order. One can naively generalize this Berry phase correction by correcting both the Berry phase and the band energy to higher order, but this cannot generate Landau levels correct to higher orders. As an example, we analyze the gapped Dirac model. The model Hamiltonian is

$$\hat{H} = vk_1\sigma_1 + vk_2\sigma_2 + \Delta\sigma_3 \,. \tag{5.3}$$

If an magnetic field is applied along the z-direction, the hamiltonian is

$$\hat{H} = \begin{pmatrix} \Delta & v(p_1 - ip_2 - \frac{1}{2}By - \frac{i}{2}Bx) \\ v(p_1 + ip_2 - \frac{1}{2}By + \frac{i}{2}Bx) & -\Delta \end{pmatrix} . \tag{5.4}$$

To exactly diagonalize the above Hamiltonian, we define

$$a^\dagger = \frac{1}{\sqrt{2B}}(p_1 - ip_2 - \frac{1}{2}By - \frac{i}{2}Bx) \tag{5.5}$$

$$a = \frac{1}{\sqrt{2B}}(p_1 + ip_2 - \frac{1}{2}By + \frac{i}{2}Bx) . \tag{5.6}$$

Then it is easy to check that $[a^\dagger, a] = 1$. The eigenvalue satisfies:

$$\begin{pmatrix} \Delta - \varepsilon & v(p_1 - ip_2 - \frac{1}{2}By - \frac{i}{2}Bx) \\ v(p_1 + ip_2 - \frac{1}{2}By + \frac{i}{2}Bx) & -\Delta - \varepsilon \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0 , \tag{5.7}$$

which gives

$$(\Delta - \varepsilon)a_1 = v\sqrt{2B}a^\dagger a_2 \tag{5.8}$$

$$(-\Delta - \varepsilon)a_2 = v\sqrt{2B}aa_1 . \tag{5.9}$$

so the exact spectrum for the conduction band reads

$$\varepsilon_n = \sqrt{\Delta^2 + 2v^2Bn} . \tag{5.10}$$

Now we use the Berry phase correction to calculate Landau levels. The semiclassical energy for the graphene system:

$$\varepsilon = \varepsilon_0 - \boldsymbol{B}\cdot\boldsymbol{m} + \boldsymbol{B}\cdot\boldsymbol{v}_0 \times \boldsymbol{a}_0' + \frac{1}{8}\alpha_{ij}(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j , \tag{5.11}$$

where

$$\varepsilon_0 = \sqrt{\Delta^2 + v^2 k^2} \tag{5.12}$$

$$m_z = -\frac{\Delta v^2}{2\varepsilon^2} \tag{5.13}$$

$$\alpha_{ij} = \partial_{ij}\varepsilon_0 = \begin{pmatrix} \frac{v^2}{\varepsilon}\left(1 - \frac{v^2 k_1^2}{\varepsilon^2}\right) & -\frac{v^4 k_1 k_2}{\varepsilon^3} \\ -\frac{v^4 k_1 k_2}{\varepsilon^3} & \frac{v^2}{\varepsilon}\left(1 - \frac{v^2 k_2^2}{\varepsilon^2}\right) \end{pmatrix} \tag{5.14}$$

The quantization rule with the Berry phase correction reads:

$$S = 2\pi \frac{eB}{\hbar}\left(n + \frac{1}{2} - \frac{1}{2\pi}\oint (\boldsymbol{a}_0 + \boldsymbol{a}_0') \cdot d\boldsymbol{k}\right). \tag{5.15}$$

Here $\boldsymbol{a}_0'$ is the field induced positional shift, and it modifies the Berry connection and hence the Berry phase up to first order. For the case of graphene, since the semiclassical energy up to second order (Eq.(9)) only depends on k^2, which means that the equal-energy surface is a circle. This fact will greatly simplify the Landau level calculation. We get from the quantization rule that

$$k^2 = 2\pi B\left(n + \frac{1}{2} + \frac{1}{2} - \frac{\Delta}{2\varepsilon_0} - \frac{1}{2\pi}\oint \boldsymbol{a}_0' d\boldsymbol{k}\right). \tag{5.16}$$

Inserting this quantization for k^2 into the semiclassical energy Eq.(5.11). The last term will vanish. The third term gets cancelled. The remaining part is

$$\varepsilon = \varepsilon_0 + \frac{\Delta B v^2}{2\varepsilon_0^2}. \tag{5.17}$$

Then

$$\varepsilon_0 = \sqrt{\Delta^2 + v^2 k^2} \tag{5.18}$$

$$= \sqrt{\Delta^2 + 2v^2 B(n+1) - \frac{v^2 \Delta B}{\varepsilon}} \tag{5.19}$$

$$= \varepsilon_n - \frac{v^2 \Delta B}{2\varepsilon_n \varepsilon} - \frac{1}{8}\frac{v^4 \Delta^2 B^2}{\varepsilon_n^3 \varepsilon^2} \tag{5.20}$$

$$= \varepsilon_n - \frac{v^2 \Delta B}{2\varepsilon_n^2} - \frac{v^2 \Delta^2 B^2}{4\varepsilon_n^5} - \frac{1}{8}\frac{v^4 \Delta^2 B^2}{\varepsilon_n^3 \varepsilon^2}\,. \tag{5.21}$$

The total energy would be

$$\varepsilon = \varepsilon_0 + \frac{\Delta B v^2}{2\varepsilon_0^2} \tag{5.22}$$

$$= \varepsilon_n - \frac{v^2 \Delta B}{2\varepsilon_n^2} - \frac{v^4 \Delta^2 B^2}{4\varepsilon_n^5} - \frac{1}{8}\frac{v^4 \Delta^2 B^2}{\varepsilon_n^3 \varepsilon^2} + \frac{\Delta B v^2}{2\varepsilon_n^2} + \frac{v^4 \Delta^2 B^2}{2\varepsilon_n^5} \tag{5.23}$$

$$= \varepsilon_n + \frac{\Delta^2 B^2 v^4}{8\varepsilon_n^5}\,. \tag{5.24}$$

Therefore, the semiclassical theory would miss a second order energy when calculating the Landau level. Based on the definition of the magnetic susceptibility, if the quantization rule is correct, this missing second order energy should cause the semiclassical theory to fail to give the correct susceptibility. However, as discussed in the previous chapter, the susceptibility from the semiclassical theory is exact. So the quantization rule must be modified.

5.2 Beyond the Berry Phase Correction

The above discrepancy finds its root in the very formulation of the Bohr-Sommerfeld rule. The process of the Bohr-Sommerfeld rule, or its general version the Einstein-Brillouin-Keller quantization rule, starts from the

observation that for a completely integrable system in d-dimension, there is d constants of motion, which are the action variables in the action-angle formulation of the classical mechanics. For simplicity, we consider the system with one action variable J. Here in the Landau level quantization process, J is proportional to the area enclosed by the semiclassical orbit in k-space. Then we can write the modified energy $\tilde{\varepsilon}$ as a function of J: $\tilde{\varepsilon} = \tilde{\varepsilon}(J)$. The right hand side of Eq.(5.2) chooses appropriate values of J for the Landau level formation, which then yields the Landau level energy according to $\tilde{\varepsilon}(J)$.

However, the above process is based on the assumption that operators $\hat{\tilde{\varepsilon}}$ and $\hat{\tilde{\varepsilon}}(\hat{J})$ are the same, which cannot be taken a priori. Under the Landau gauge for the magnetic vector potential, we can write the energy as $\tilde{\varepsilon}(k_x, k_y + Bx)$ and $\tilde{\varepsilon}[J(k_x, k_y + Bx)]$. If we directly promote the dynamical variable k_x, k_y, x and y into operators $\hat{k}_x$,$\hat{k}_y$, $\hat{x}$ and $\hat{y}$ according to $[\hat{x}, \hat{k}_x] = [\hat{y}, \hat{k}_y] = i\hbar$ and $[\hat{x}, \hat{k}_y] = [\hat{y}, \hat{k}_x] = 0$, it is easy to check that the operator $\hat{\tilde{\varepsilon}}(\hat{k}_x, \hat{k}_y + B\hat{x})$ is different from $\hat{\tilde{\varepsilon}}[J(\hat{k_x}, \hat{k}_y + B\hat{x})]$, because they have different arrangments of the two arguments $\hat{k}_x$ and $\hat{k}_y + B\hat{x}$, which do not commute with each other in the operator form. Actually, it has been long established in the phase space quantum mechanics [50, 81], that the dynamical function $\tilde{\varepsilon}(k_x, k_y + Bx)$ corresponds to a quantum mechanical operator by re-arranging the two arguments k_x and $k_y + Bx$ into totally symmetric form and promoting them into operators, which does not agree with the operator $\hat{\tilde{\varepsilon}}[J(\hat{k_x}, \hat{k}_y + B\hat{x})]$ from the Bohr-Sommerfeld rule in general.

As a concrete example, we illustrate how to correctly get the Landau

level up to second order for the gapped Dirac model, by adding the difference between the operator $\hat{\tilde{\varepsilon}}(\hat{k}_x, \hat{k}_y + B\hat{x})$ and $\hat{\tilde{\varepsilon}}[J(\hat{k_x}, \hat{k}_y + B\hat{x})]$. Notice that in this model we can choose $J = k^2$, and we label $\boldsymbol{p} = \boldsymbol{k} + \boldsymbol{A}$ where $\boldsymbol{A}$ is the vector potential. Then we have $[\hat{p}_i, \hat{p}_j] = \epsilon_{ijk} i B_k$. Quantum mechanically, $\hat{p}^2\hat{p}^2$ is different from $\hat{p}_i\hat{p}_j\hat{p}_i\hat{p}_j$. To get the right operator form we need to use the total symmetric form for all the $\hat{p}_i$'s. The difference between two operators $\hat{A}(\hat{p}_i)\hat{B}(\hat{p}_i)$ and its total symmetric form is

$$\begin{aligned}
<< AB >> =& \frac{1}{2} i \left(\frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \right) \\
& - \frac{1}{8} \left(\frac{\partial^2 A}{\partial p_i \partial p_j} \frac{\partial^2 B}{\partial q_i \partial q_j} + \frac{\partial^2 A}{\partial q_i \partial q_j} \frac{\partial^2 B}{\partial p_i \partial p_j} - 2 \frac{\partial^2 A}{\partial p_i \partial q_j} \frac{\partial^2 B}{\partial q_i \partial p_j} \right) \\
=& i \epsilon_{ij\theta} B_\theta \partial_j A \partial_i B - \frac{1}{8} B_\theta B_\phi \epsilon_{m\theta i} \epsilon_{n\phi j} \frac{\partial^2 A}{\partial k_i \partial k_j} \frac{\partial^2 B}{\partial k_m \partial k_n} \,. \qquad (5.25)
\end{aligned}$$

For simplicity, we consider a rotational symmetric energy spectrum $\varepsilon_0(k)$. To evaluate the exact difference in ε_0, notice that $\varepsilon_0 = f(k^2)$. We take the Laplace transformation of ε_0:

$$\phi(s) = \int_0^\infty f(x) e^{-xs} dx \,. \qquad (5.26)$$

Then

$$f(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \phi(s) e^{xs} ds \,. \qquad (5.27)$$

We can evaluate the difference in e^{xs} due to the noncommutability of $x = k^2$:

$$e^{s\hat{x}} = e^{xs} + \sum_n \frac{s^n}{n!} C_n^2 x^{n-2} << x, x >> \qquad (5.28)$$

$$+ \sum_n \frac{s^n}{n!} C_n^3 x^{n-3} << x, x, x >> + \cdots \qquad (5.29)$$

Finally, we can get

$$f(\hat{x}) = f(x) - \frac{1}{8}B_\lambda B_\mu \epsilon_{\lambda\ell k}\epsilon_{\mu\nu\rho}\left(\frac{1}{2}f''\partial_{\ell\nu}x\partial_{k\rho}x + \frac{1}{3}f'''\partial_{\ell\nu}x\partial_k x\partial_\rho x\right). \quad (5.30)$$

In general, we can change x into the action variable J, we have

$$\tilde{\varepsilon}(\hat{J}) = \hat{\tilde{\varepsilon}} - \frac{1}{8}B_s B_t \epsilon_{sij}\epsilon_{tk\ell}\left(\frac{1}{2}\tilde{\varepsilon}''\partial_{ik}\hat{J}(k_1,k_2)\partial_{j\ell}\hat{J}(k_1,k_2)\right.$$
$$\left. + \frac{1}{3}\tilde{\varepsilon}'''\partial_{ik}\hat{J}(k_1,k_2)\partial_j\hat{J}(k_1,k_2)\partial_\ell\hat{J}(k_1,k_2)\right), \quad (5.31)$$

where ϵ_{sij} is the totally antisymmetric tensor in three dimension. i, j, k, ℓ, s and t refer to Cartesian coordinates, and repeated index is summed. $\tilde{\varepsilon}''$ and $\tilde{\varepsilon}'''$ is the second order and third order derivatives of $\tilde{\varepsilon}$ with respect to $\hat{J}$. Note that the second term on the right hand side of Eq.(5.31) only involves $J(k_1,k_2)$ instead of $J(k_1,k_2+Bx)$.

The semiclassical theory has the energy corresponds to $f(k_1,k_2)$, or the operator from the total symmetrization of $f(x(k_1,k_2))$. However, the quantization we have done above corresponds to $f(\hat{x})$. So we must take account the difference between these two, which means that the exact Landau level from the semiclassical theory should be

$$\varepsilon = \varepsilon_n + \frac{\Delta^2 B^2 v^4}{8\varepsilon_n^5} \quad (5.32)$$
$$+ \frac{1}{8}B_\lambda B_\mu \epsilon_{\lambda\ell k}\epsilon_{\mu\nu\rho}\left(\frac{1}{2}f''\partial_{\ell\nu}x\partial_{k\rho}x + \frac{1}{3}f'''\partial_{\ell\nu}x\partial_k x\partial_\rho x\right). \quad (5.33)$$

Since $f(x) = \sqrt{\Delta^2 + v^2 x}$ and $x = k^2$, we can calculate the energy. The result is $\varepsilon = \varepsilon_n + O(B^3)$. Therefore, we have corrected the Landau level energy from the semiclassical theory up to second order.

As a second example, we consider the low energy model in the double layer graphene. The model Hamiltonian is given by $\hat{\mathrm{H}} = -\frac{k_1^2-k_2^2}{2}\sigma_1 - k_1k_2\sigma_2 + \Delta\sigma_3$, where k_1 and k_2 are momentum in x and y direction, 2Δ is the band gap, and σ_1, σ_2 and σ_3 are Pauli matrices. Under a B field along z direction, Landau levels in the conduction band can be solved exactly: $\varepsilon_{\mathrm{quan}} = \sqrt{\Delta^2 + n(n-1)B^2}$, with $n = 0, 1, 2, \cdots$.

To apply the Onsager's rule, we start with the band energy modified by the magnetic field. We will focus on the conduction band. According to the discussion in Chapter 4 and given that this model has the particle-hole symmetry, the band energy up to second order in B for the conduction band reads:

$$\begin{aligned}\tilde{\varepsilon} = \varepsilon_0 - Bm + \frac{1}{4}(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m}) - \frac{1}{8}\epsilon_{sik}\epsilon_{tj\ell}B_sB_tg_{ij}\alpha_{k\ell} - B\cdot(\boldsymbol{a}_0'\times\boldsymbol{v}_0) \\ + \frac{1}{8}\sum_{(m,n)\neq 0}(\boldsymbol{B}\times\boldsymbol{A}_{0m})_i(\hat{\Gamma}_{ij})_{mn}(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j - \frac{1}{16}(\boldsymbol{B}\times\boldsymbol{\partial})_i(\boldsymbol{B}\times\boldsymbol{\partial})_j(\hat{\Gamma}_{ij})_{00}\,.\end{aligned} \tag{5.34}$$

Here ε_0 is the conduction band dispersion without any magnetic field, $m = \frac{\Delta k^2}{2\varepsilon_0^2}$ is the orbital magnetic moment; $\Omega = \frac{\Delta k^2}{2\varepsilon_0^3}$ is the Berry curvature; g_{ij} is the quantum metric tensor:

$$g = \frac{k^2}{4\varepsilon_0^2}\begin{pmatrix} 1-\frac{k^2k_1^2}{4\varepsilon_0^2} & -\frac{k^2k_1k_2}{4\varepsilon_0^2} \\ -\frac{k^2k_1k_2}{4\varepsilon_0^2} & 1-\frac{k^2k_2^2}{4\varepsilon_0^2}\end{pmatrix}; \tag{5.35}$$

$\alpha_{k\ell}$ is the effective mass tensor:

$$\alpha = \begin{pmatrix} \frac{k^2}{2\varepsilon_0}+\frac{\Delta^2k_1^2}{\varepsilon_0^3} & \frac{\Delta^2k_1k_2}{\varepsilon_0^3} \\ \frac{\Delta^2k_1k_2}{\varepsilon_0^3} & \frac{k^2}{2\varepsilon_0}+\frac{\Delta^2k_2^2}{\varepsilon_0^3}\end{pmatrix}; \tag{5.36}$$

$\boldsymbol{v}_0 = \frac{k^2}{2\varepsilon_0}(k_1, k_2)$ is the band velocity; $\boldsymbol{a}_0'$ is the field induced positional shift:

$$\boldsymbol{a}_0' = \frac{(k_2, -k_1)}{2\varepsilon_0^2}\left(1 - \frac{3k^4}{8\varepsilon_0^2}\right) B\,; \tag{5.37}$$

$\boldsymbol{A}_{0m}$ is the interband Berry connection; $\hat{\Gamma}_{ij} = \partial_{ij}\hat{H}$ is the Hessian matrix.

Therefore, we can obtain the modified band energy up to second order in B:

$$\tilde{\varepsilon} = \varepsilon_0 - B\frac{\Delta k^2}{2\varepsilon_0^2} - \frac{B^2}{8\varepsilon_0}\left(1 - \frac{\Delta^2(\varepsilon_0^2 + 2\Delta^2)}{\varepsilon_0^4}\right) - \frac{k^4 B^2}{4\varepsilon_0^3}\left(1 - \frac{3k^4}{8\varepsilon_0^2}\right). \tag{5.38}$$

Obviously, the equal-energy contour of $\tilde{\varepsilon}$ is a circle, and its area is $S = \pi k^2$. From the Onsager's rule, we obtain that

$$S = 2\pi\left(n + \frac{1}{2} - \frac{\Gamma(\mu)}{2\pi}\right) B\,, \tag{5.39}$$

where $\Gamma = \oint(\boldsymbol{a}_0 + \boldsymbol{a}_0') \cdot d\boldsymbol{k}$ is the Berry phase up to first order in B, and $n = 0, 1, 2, \cdots$. The Berry phase contribution from $\boldsymbol{a}_0$ is

$$\oint \boldsymbol{a}_0 \cdot d\boldsymbol{k} = \int \Omega d^2k = 2\pi\left(1 - \frac{\Delta}{\varepsilon_0}\right). \tag{5.40}$$

The field induced positional shift $\boldsymbol{a}_0'$ yields:

$$\oint \boldsymbol{a}_0' \cdot d\boldsymbol{k} = -\frac{\pi B k^2}{\varepsilon_0^2}\left(1 - \frac{3k^4}{8\varepsilon_0^2}\right). \tag{5.41}$$

Based on Eq.(5.39-5.41), we have

$$\pi k^2 = 2\pi\left[n - \frac{1}{2} + \frac{\Delta}{\varepsilon_0} + \frac{k^2 B}{4\varepsilon_0^2}\left(1 - \frac{3k^4}{8\varepsilon_0}\right)\right] B. \tag{5.42}$$

We then plug Eq.(5.42) into Eq.(5.38) to obtain the Landau level energy $\tilde{\varepsilon}$ from the Onsager's rule. The exact Landau level energy $\varepsilon_{\text{quan}}$ has been given

in the main text. We will expand the difference between $\tilde{\varepsilon}$ and $\varepsilon_{\rm quan}$ up to second order in B:

$$\begin{aligned}\tilde{\varepsilon} &= \sqrt{\varepsilon_{\rm quan}^2 + \frac{B^2}{4} + \frac{\Delta^2 B^2}{\varepsilon_0^2} + \frac{2\Delta(n-1/2)B^2}{\varepsilon_0}} - \frac{\Delta B k^2}{2\varepsilon_0^2} - \frac{B^2}{8\varepsilon_0}\left(1 - \frac{\Delta^2(\varepsilon_0^2+2\Delta^2)}{\varepsilon_0^4}\right) \\ &= \varepsilon_{\rm quan} + \frac{\Delta^2 B^2}{2\varepsilon_{\rm quan}^3} + \frac{\Delta B^2(n-1/2)}{\varepsilon_{\rm quan}^2[1+\Delta B^2(n-1/2)/(\varepsilon_{\rm quan}^3)]} - \frac{\Delta^2 B^4 (n-1/2)^2}{2\varepsilon_{\rm quan}^5} \\ &- \frac{\Delta B}{2}\frac{2(n-1/2+\Delta/\varepsilon_{\rm quan})B}{\varepsilon_{\rm quan}^2[1+2\Delta B^2(n-1/2)/(\varepsilon_{\rm quan}^3)]} + \frac{B^2\Delta^2(\varepsilon_{\rm quan}^2+2\Delta^2)}{8\varepsilon_{\rm quan}^5} + O(B^3) \\ &= \varepsilon_{\rm quan} + \frac{\Delta^2 B^2}{8\varepsilon_{\rm quan}^5}(\varepsilon_{\rm quan}^2 - 2\Delta^2) + O(B^3)\,. \end{aligned} \tag{5.43}$$

Therefore, the Onsager's rule is not guaranteed to produce correct Landau levels even with the higher order Berry phase and band structure corrections. By adding the correction in Eq.(5.31), the second term in the last line of Eq.(5.43) will be cancelled, and we have the correct Landau level up to second order.

5.3 Density Quantization Scheme

There is a systematic way to generalize the Onsager's rule to any order of magnetic field. It is given by the quantization of the semiclassical electron density at zero temperature. In this section, we will derive this density quantization scheme.

Without loss of generality we derive the quantization rule in two dimension. We start from the observation that each Landau level carries the same density of states in two dimensional electron gas. It is well known that

when the Fermi energy falls between neighboring Landau levels in two dimensional electron gas, the transverse conductivity is quantized (the quantum Hall effect): $\sigma_{\text{trans}} = ne^2/h$, where n is an integer called Landau level filling factor. The transverse current from this conductivity can also be derived by multiplying the density of states ρ with the drift velocity: $\boldsymbol{j} = e\rho(\boldsymbol{E} \times \boldsymbol{B})/B^2$, where $\boldsymbol{E}$ is in-plane electric field and $\boldsymbol{B}$ is out-of-plane magnetic field. Therefore, the density of states is also quantized $\rho = neB/h$, and each Landau level has the same density of states eB/h.

Similar observation also holds for crystalline solids, where the Hofstadter spectrum emerges instead of Landau levels. When the Fermi energy falls in the spectral gap, the carrier density ρ reads as [79, 46, 14]: $\rho = nB/\phi_0 + t/A$, where $\phi_0 = h/e$ is the magnetic flux quantum, A is the unit cell area, n is an integer connected to the quantum Hall effect, and t is an integer connected to the filling index in the mini-band. We will focus on the case with $t = 0$ due to the following recursive structure in the Hofstadter butterfly[11, 30, 79, 46, 14]: for the 'pure case' with a flux $\phi/\phi_0 = 1/q$ ($\phi = BA$ and q is an integer), $t = 0$ and the spectrum generally follows a set of Landau levels arising from the original band structure at $B = 0$; on the other hand, for the situation with $\phi/\phi_0 = p/q$ (p and q are co-prime integers) with $1 < p < q/2$, the spectrum can be renormalized to a pure case arising from the magnetic Bloch band at a certain rational flux ϕ', and the flux is renormalized to $\phi - \phi'$ with $(\phi - \phi')/\phi_0 = 1/q'$ (q' is an integer). Therefore, the carrier density is $\rho = nB/\phi_0$ and each mini-band in the Hofstadter spectrum carries the same

density of states eB/h.

The Landau level spectrum $\varepsilon_n(B)$ (level index $n = 0, 1, \cdots$) from a band minimum in two dimensions can be expressed as $\rho(B, T, \mu) = (B/\phi_0) \sum_n f[\varepsilon_n(B) - \mu]$ with $\phi_0 = h/e$ the flux quantum, f the Fermi function and μ the chemical potential. At $T = 0$, according to above analysis, the electron density is a staircase function with constant risers B/ϕ_0 located at $\mu = \varepsilon_n(B)$ (see Fig.(5.1)).

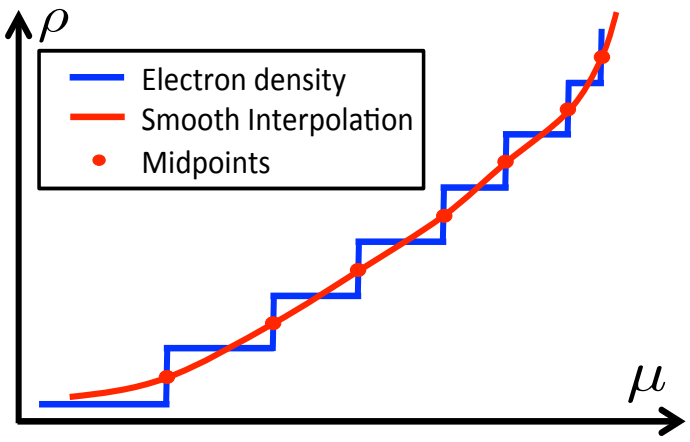

Figure 5.1: Zero-temperature electron density and its smooth interpolation.

The spectrum $\varepsilon_n(B)$ can be interpolated as follows. Without loss of generality, we write $\varepsilon_n(B) = g(x_n, B)$ with $x_n = (n + 1/2)B/\phi_0$ being the zero-temperature electron density when the n-th Landau level is half-filled. If we allow x change continuously, the function $\varepsilon = g(x, B)$ smoothly interpolate points $\varepsilon_n(B)$ at $x = x_n$ for each value of B. $g(x, B)$ is a monotonic function of x for each B, inherited from the fact that $\varepsilon_n(B)$ increases with n. Therefore,

we can invert $g(x, B)$ and obtain $x = x(\mu, B)$, represented as the smooth curve in Fig.(5.1). By construction, the interpolation condition implies that this smooth curve goes through the midpoints of the staircase risers.

In the following, we will show that the smooth interpolation curve $x(\mu, B)$ is actually the semiclassical electron density $\rho_{\rm semi}(B, T \to 0, \mu)$. Notice that there are three energy scales in the exact electron density $\rho(B, T, \mu)$: the Landau level spacing determined by B, the variation scale of the Fermi function determined by $k_B T$, and the chemical potential μ. We assume μ falls inside the band, and thus has a finite distance from the band minimum. To calculate finite temperature semiclassical electron density $\rho_{\rm semi}(B, T, \mu)$, we set the level spacing to be much smaller than $k_B T$. Then, using the Euler-Maclaurin formula, we can transform the sum over n in $\rho(B, T, \mu)$ to an integration over a continuous variable x:

$$\rho_{\rm semi}(B, T, \mu) = \int_{B/2\phi_0}^{\infty} f dx + \frac{B f|_{x=B/2\phi_0}}{2\phi_0} + \mathcal{R}\,, \tag{5.44}$$

where $f = f[g(x, B) - \mu]$. $\mathcal{R}$ contains all the remainder terms.

Eq.(5.44) can be greatly simplified. Terms in $\mathcal{R}$ are evaluated at $x = B/2\phi_0$, i.e. at the 0-th Landau level near the band minimum. Here we have ignored the contribution at $x = \infty$ since they correspond to very high energy, and this is still valid for realistic models in solid state physics which contain singular points. Moreover, terms in $\mathcal{R}$ are proportional to successively higher power of B and contain $\partial f/\partial\mu$ and its higher order derivatives with respect to μ. The limit $T \to 0$ in $\rho_{\rm semi}(\rho, T \to 0, \mu)$ is understood as $k_B T$ being

much smaller than the distance between μ and band minimum, but still much bigger than the level spacing. Therefore, each term in $\Re$ is exponentially small. At this limit of $k_BT \to 0$, Eq.(5.44) may not converge, in which case it is understood as an asymptotic expansion. Therefore, we can keep finite number of terms in Eq.(5.44) and then set $k_BT \to 0$. The result is simply $\rho_{\rm semi}(B, T \to 0, \mu) = x(\mu, B)$.

The coincidence between the interpolation curve $x(\mu, B)$ and the semiclassical electron density at the limit of zero temperature yields a density quantization scheme:

$$\rho(B, T \to 0, \varepsilon_n) = x_n \,. \tag{5.45}$$

At finite temperature, we have the following power series expansion: for any $N \geq 0$, $\rho_{\rm semi}(B, T, \mu) = \sum_{m=0}^{N} R_m(\mu, T) B^m/m! + O(B^{N+1})$, where $R_m(\mu, T)$ is the zero-field response function defined from the exact electron density $\rho(B, T, \mu)$ by $R_m(\mu, T) = \lim_{B\to 0} \partial^m \rho/\partial B^m$. At $T \to 0$, we treat the above expansion in the asymptotic sense, and obtain $\rho_{\rm semi}(B, T \to 0, \mu) = \sum_{m=0}^{N} R_m(\mu, T \to 0) B^m/m! + O(B^{N+1})$. Application of the interpolation condition $x_n = x(\varepsilon_n, B)$ to this expression then yields

$$\left(n + \frac{1}{2}\right) \frac{B}{\phi_0} = \sum_{m=0}^{N} R_m(\varepsilon_n, T \to 0) \frac{B^m}{m!} + O(B^{N+1}) \,. \tag{5.46}$$

Eq.(5.46) is our density quantization scheme: we first calculate $R_m(\mu, T \to 0)$ from the semiclassical theory, and then solve the spectrum ε_n from Eq.(5.46). Obviously, this spectrum can be accurate to arbitrary order of B as long as corresponding $R_m(\mu, T \to 0)$ are given.

We comment that these $R_m(\mu, T \to 0)$ are connected to the response of the smooth semiclassical free energy $G(B, \mu, T \to 0)$ with respect to B. Here and hereafter, we use R_m to represent $R_m(\mu, T \to 0)$. Based on the Maxwell relation, we have $R_0 = \partial G|_{B=0}/\partial\mu$, $R_1 = \partial M/\partial\mu$, $R_2 = \partial\chi/\partial\mu$, $\cdots$, where $M = -(\partial G/\partial B)|_{B=0}$ is the magnetization, $\chi = -(\partial^2 G/\partial B^2)|_{B=0}$ is the susceptibility, and so on.

Eq.(5.46) is our new quantization rule and we demonstrate its significance through its relation with the Onsager's rule. First, we truncate its right hand side at the zeroth order term. Since $R_0 = S_0/(4\pi^2)$ where S_0 is the k-space area enclosed by the equal-energy contour in the band structure ε_0 with the energy μ, we obtain the Onsager's rule: $S_0 = 2\pi\left(n + \frac{1}{2}\right)\frac{eB}{\hbar}$.

Then we truncate Eq.(5.46) at the first order term $BR_1 = B\partial M/\partial\mu$. We consider the spinless case for simplicity. M contains contributions from the orbital magnetic moment m and the Berry curvature Ω [87, 72]: $M = \int(mf - \Omega g)d^2k/(4\pi^2)$. Here $g = \int f d\varepsilon$. Therefore,

$$\frac{\partial M}{\partial\mu} = -\int(mf' - \Omega f)\frac{d^2k}{4\pi^2}\,. \tag{5.47}$$

If we combine the first term with R_0 and move the second term to the left hand side of Eq.(5.46), we obtain the following quantization condition:

$$S' = 2\pi\left(n + \frac{1}{2} - \frac{\Gamma(\mu)}{2\pi}\right)\frac{eB}{\hbar}\,, \tag{5.48}$$

where $S' = \int f(\varepsilon_0 - Bm - \mu)d^2k$ is the area enclosed by the equal-energy contour in the modified band structure $\varepsilon_0 - Bm$ with the energy μ, and $\Gamma(\mu)$

is the Berry phase associated with the semiclassical orbit. This is exactly the modified Onsager's rule due to the Berry phase and magnetic moment effect [11, 84, 82, 47].

Even though the equivalence at first order, our quantization scheme still has a few advantages over the modified Onsager's rule in previous work. First, it offers a concrete and general form where the orbital magnetization M is the essential ingredient as a combination of the Berry phase and orbital magnetic moment effect. Moreover, since response functions in Eq.(5.46) only contain the Fermi function in the original band ε_0, the equal-energy contour in ε_0 is required. This is a computational advantage over directly finding the equal-energy contour of $\varepsilon_0 - Bm$ as in Eq.(5.48)[11], since m can be dramatically enhanced around the point where two bands are near each other, creating significant numerical error in the equal-energy-contour finding process.

We comment that naively generalizing the Berry phase and magnetic moment to higher order in magnetic field does not guarantee a correct Landau level spectrum at higher order. Actually, if we truncate Eq.(5.46) at the second order term, since $R_2 = \partial\chi/\partial\mu$, it is the magnetic susceptibility that contributes, which is different from this naive generalization.

Eq.(5.46) can be easily generalized for Landau levels of holes. In this case, we have a quantization of empty states. Therefore, the magnetic responses on the right hand side of Eq.(5.46) must be calculated from the band maximum, i.e. we can replace the Fermi distribution function f by $1 - f$.

In summary, Eq.(5.46) interprets the Onsager's rule as from the zeroth order contribution to the semiclassical electron density and modification from the Berry phase and magnetic moment as from the first order magnetic response $R_1 = \partial M/\partial\mu$. It also suggests that further corrections to Onsager's rule corresponds to successively higher order magnetic response functions.

5.4 Application in continuum and lattice models

As a concrete example, we consider the continuum model in double layer graphene, with the Hamiltonian given in the last section.

Now we apply our quantization scheme up to second order. The conduction band dispersion is $\varepsilon_{\rm c} = \sqrt{\Delta^2 + k^4/4}$. Its equal-energy contour is a circle with the area $S_0 = \pi k^2$. The magnetization and susceptibility can be easily calculated from the semiclassical theory, with $\partial M/\partial\mu = 1/(2\pi)$ and $\partial\chi/\partial\mu = 1/(8\pi\sqrt{\mu^2 - \Delta^2})$. Therefore, Eq.(5.46) yieds:

$$\left(n + \frac{1}{2}\right) B = \sqrt{\varepsilon_2^2 - \Delta^2} + \frac{B}{2} + \frac{B^2}{8} \frac{1}{\sqrt{\varepsilon_2^2 - \Delta^2}} \,. \tag{5.49}$$

Solving the above equation by including the zeroth, first and second order terms on its right hand side and expanding the results at a small magnetic field, we obtain $\varepsilon_0 = \varepsilon_{\rm quan} + B\sqrt{1 - \Delta^2/\varepsilon_{\rm quan}^2} + O(B^2)$, $\varepsilon_1 = \varepsilon_{\rm quan} + B^2/8 + O(B^3)$, and $\varepsilon_2 = \varepsilon_{\rm quan} + O(B^3)$. These results clearly demonstrates that our quantization scheme can correct the Onsager's rule order by order if corresponding magnetic response functions are known. In comparison, the direct generalization of the Berry phase and magnetic moment to higher order cannot yield

correct Landau levels up to second order, as discussed in the previous section.

Moreover, Eq.(5.46) sometimes yields exact Landau levels after truncated at a finite order. As an example, we consider the Dirac model in graphene: $\hat{H} = vk_1\sigma_1 + vk_2\sigma_2 + \Delta\sigma_3$, where v is the Fermi velocity. Now we apply our quantization scheme up to first order. The conduction band dispersion is $\varepsilon_c = \sqrt{\Delta^2 + v^2k^2}$. Since $R_1 = \partial M/\partial\mu = 1/(4\pi)$, Eq.(5.46) yields the exact Landau levels $\varepsilon_{\text{quan}} = \sqrt{\Delta^2 + 2nv^2B}$. This coincidence implies that all R_m at $m \geq 2$ for this model must vanish when μ falls inside the band.

An interesting comment is that in the above two models, $\partial M/\partial\mu$ are both constants, causing constant shifts of the level index: $n + 1/2 \to n - 1/2$ and $n + 1/2 \to n$, respectively. Our theory suggests that such constant shifts are dominated by the Berry phase effect for higher levels away from the band edge and by the orbital magnetic moment for lower levels near the band edge. The Berry phase effect contributes alone only when the band gap vanishes.

Our quantization scheme applies to the lattice model as well. As an example, we consider the following tight-binding graphene model [52]: $\hat{H} = -t\sum_{\langle i,j\rangle} c_i^\dagger c_j$, where t is the strength of the nearest neighbor hopping. We will focus on the conduction band. To obtain Landau levels, we analyze the topology of the equal-energy contour first. The conduction band has two degenerate valleys around two energy minimum with $\varepsilon = 0$ at K and K' point, one saddle point with $\varepsilon = t$, and one maximum with $\varepsilon = 3t$ at the Γ point. According to the Morse index theorem [7], there are two sets of degenerate electron-like semiclassical orbits at the range $0 < \varepsilon < t$ which enclose K and

K' point separately, and one set of hole-like semiclassical orbits at the range $t < \varepsilon < 3t$ which enclose the Γ point. Different sets of semiclassical orbits yield Landau levels with different sets of level index.

Here we comment that the valley/spin degeneracy in 2D materials as shown in our tight-binding model are common. Then a magnetic field may lift these degeneracies and create multiple sets of Landau levels, as illustrated in [41, 15]. Eq. (5.45) and (5.46) is then valid independently for each species of carriers. In other words, to calculate the Landau levels for each species, the electron density in Eq.(5.45) and magnetic response functions in Eq.(5.46) are for the same species.

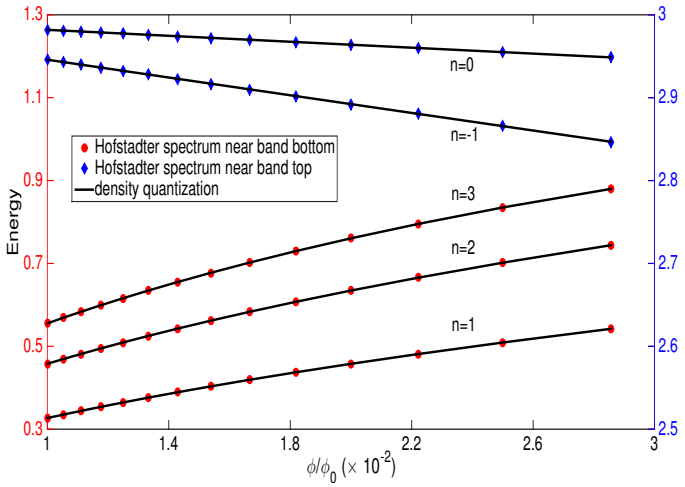

Figure 5.2: (color online) The Hofstadter spectrum and the Landau level energy based on Eq.(5.46). We calculate three levels from the band bottom and two levels from the band top.

With the above analysis in mind, we calculate magnetic responses only

from one valley from the semiclassical theory, and use Eq.(5.46) up to second order to obtain the zeroth Landau level. In Fig.(5.2), we compare the results with the Hofstadter spectrum at small magnetic field range, and they agree very well.

We comment that mini-bands in Hofstadter spectrum have finite bandwidths. This corresponds to the fact that Landau levels are highly degenerate and tunneling between degenerate orbits broadens Landau levels [82], which is beyond the scope of our theory. However, the resulting bandwidth for small B has the form $e^{-B_0/B}$ where B_0 is some constant [37, 82, 26]. So it is exponentially small at small B except near the saddle point where the semiclassical theory breaks down. Therefore, this broadening effect does not introduce any contradiction to our asymptotic power series expansion in Eq.(5.46).

There is another delightful generalization of our theory to match more parts of the Hofstadter spectrum. Near each of the following rational flux $\phi/\phi_0 = p/q$ where p and q are co-prime integers with $p \neq 1$ and $p < q$, there are independent sets of Landau levels, which represent the self-similar property of the Hofstadter spectrum [30, 46]. Our quantization scheme can be applied to obtain these Landau levels, if we replace the original band structure by the magnetic Bloch band at $\phi/\phi_0 = p/q$. This type of calculations with first order accuracy has been performed in [11].

5.5 Experimental Implication

To illustrate the implication of our quantization scheme in experiments, we truncate Eq.(3) at second order yields:

$$n = \frac{\phi_0 S_0}{4\pi^2 B} + \left(\frac{\Gamma(\mu)}{2\pi} - \frac{1}{2} + \phi_0 \langle m \rangle D \right) + \frac{B\phi_0}{2} \frac{\partial \chi}{\partial \mu} , \tag{5.50}$$

where $D = \int f' d^2k/(4\pi^2)$ is the density of states at Fermi surface and $\langle m \rangle = \int m f' d^2k/(4\pi^2 D)$ is the averaged magnetic moment over the Fermi surface. The terms in the bracket comes from $\partial M/\partial \mu$.

The Berry phase in graphene has already been measured through the Shubnikov-de Haas oscillations [54, 92], by linearly fitting the level index n to the inverse of the magnetic field B and reading the interception to the n axis. However, in similar experiments on the gapless surface mode of three-dimensional topological insulators, the measured Berry phase usually deviates from the expected value π [76, 62, 77, 2, 66, 89, 16, 8, 90, 83]. This puzzle is closely related to the above mentioned higher order corrections to the level index, since they will lead to a nonlinear relation between n and $1/B$.

Eq.(5.50) offers a universal explanation to those experimental puzzles in the Berry phase measurements. It shows that if $\partial \chi/\partial \mu$ vanishes, n depends linearly on $1/B$, with the constant term determined by Berry phase and averaged value of magnetic moment. Therefore, a non-vanishing $\partial \chi/\partial \mu$ certainly causes deviation from this linear relation. Meanwhile, Eq.(5.50) explains the shift of Berry phase in previous measurements. If we combine the last two

terms in Eq.(5.50), even though $\langle m \rangle = 0$, we can find that $B\partial\chi/\partial\mu$ effectively shifts the Berry phase after linearly fitting n to $1/B$.

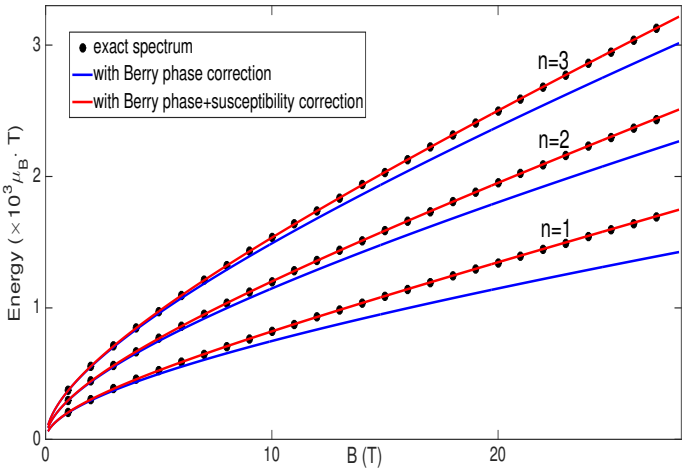

Figure 5.3: Comparing exact spectrum with the spectrum calculated from the quantization rule in Eq.(5.50) with Berry phase alone ($\langle m \rangle = 0$ for Eq.(5.51)) and Berry phase plus magnetic susceptibility correction. We calculate the spectrum for $n = 1$, $n = 2$, and $n = 3$ level. We use experimentally determined parameters $m/m_0 = 0.13$, $g_s = 76$, and $v_f = 3 \times 10^5$m/s [76].

As a concrete example, we consider the following model for surface states of three dimensional topological insulators [76] (for simplicity, we choose e, $\hbar$ to be unity):

$$\hat{\mathrm{H}} = v_f(k_1\sigma_2 - k_2\sigma_1) + \frac{k^2}{2m} - \frac{1}{2}g_s\mu_B B\sigma_3\,, \tag{5.51}$$

where v_f is the Fermi velocity, m is the effective mass, g_s is the surface g-factor, μ_B is the Bohr magneton, k_1 and k_2 are momentum in x and y direction, and σ_1, σ_2 and σ_3 are Pauli matrices. The three terms in Eq.(5.51) represents spin-orbit coupling, kinetic energy, and spin Zeeman energy, respectively. Under a

B field along z direction, Landau levels in the conduction band can be solved exactly: $\varepsilon_{\rm quan} = nB/m + \sqrt{2v_f^2 nB + (m_0/m - g_s/2)^2 \mu_B^2 B^2}$, where m_0 is the free electron mass.

For a finite m and g_s, the susceptibility does not vanish in general. According to the semiclassical theory, if μ falls inside the conduction band, we have

$$\frac{\partial \chi}{\partial \mu} = -\frac{m\mu_B^2}{2\pi v_f} \frac{\left(\frac{1}{m/m_0} - \frac{1}{2}g_s\right)^2}{k_f + mv_f}\,, \tag{5.52}$$

where $k_f = -mv_f + \sqrt{m^2 v_f^2 + 2m\mu}$ is the Fermi wave vector. The Berry phase is still π and $\langle m \rangle$ still vanishes. Therefore, Eq.(5.50) yields

$$n = k_f^2/2B - \frac{mB\mu_B^2}{2v_f} \frac{\left(\frac{1}{m/m_0} - \frac{1}{2}g_s\right)^2}{k_f + mv_f}\,. \tag{5.53}$$

The second term in Eq.(5.53) clearly shows the deviation from linear relation between n and $1/B$, as long as $m_0/m - g_s/2$ does not vanish.

The importance of susceptibility in Eq.(5.53) is clearly seen in Fig.(5.3). We see that the Berry phase correction alone cannot reproduce the exact spectrum very well. But with the additional correction from susceptibility one can significantly reduce the error.

The Berry phase and magnetic susceptibility can be obtained by quadratically fitting nB to B. Then according to Eq.(5.46) and (5.50), if $\langle m \rangle$ vanishes, the Berry phase solely appears in the coefficient of the B-linear term. As an example, by using the exact spectrum as shown in Fig.(**??**), the second order fitting gives a Berry phase 0.99π. To compare, if we fit n linearly to $1/B$, a

Berry phase 0.68π is obtained. Our method is clearly more accurate than the previous one. Moreover, a $\partial\chi/\partial\mu$ with the value 0.022 (in units of $\phi_0^{-1}\mathrm{T}^{-1}$) is given by the coefficient of B^2 term, which is very close to its exact value 0.023.

We comment that for two dimensional materials with spin or valley degeneracy, if one can isolate the Landau level spectrum for each spin or valley species, then by employing Eq.(5.50), the magnetic response functions for that species can be obtained. The total magnetic response functions are simply the summation of contributions from all species.

Appendices

Appendix A

First Order Correction in the Wave-Packet

In this appendix, we will derive the first order correction in the wave-packet. The wave-packet can be put in the following form:

$$|\Psi\rangle = \int d\boldsymbol{p}\, e^{i\boldsymbol{p}\cdot\boldsymbol{q}} \left(C_0(\boldsymbol{p})|u_0(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle + \sum_{n\neq 0} C_n(\boldsymbol{p})|u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \right). \tag{A.1}$$

From now on, $u_0(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times r_c)$ and $u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)$ are denoted by u_0 and u_n in shorthand. In Eq.(A.1), similarly as in the first order semiclassical theory, $C_0(\boldsymbol{p})$ is required to be peaked around some point $\boldsymbol{p}_c$: $|C_0|^2 \approx \delta(\boldsymbol{p}-\boldsymbol{p}_c)$. $C_n(\boldsymbol{p})$ incorporates the interband mixing corresponding to the components due to $|u_0'\rangle$ and is hence at least of first order in external fields. C_n and C_0 are not independent: their relation should be determined by requiring that the wave packet state satisfying the Schrödinger equation $\hat{H}|\Psi\rangle = i\partial_t|\Psi\rangle$.

Since $|u_0\rangle$ and $|u_n\rangle$ all depends on t implicitly through $\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c(t)$, the dynamical part of the Schrödinger equation is

$$\begin{aligned} i\partial_t|\Psi\rangle = &\int d\boldsymbol{p}\,\varepsilon C_0 e^{i\boldsymbol{p}\cdot\boldsymbol{q}}|u_0\rangle + \sum_{n\neq 0}\int d\boldsymbol{p}\,\varepsilon C_n e^{i\boldsymbol{p}\cdot\boldsymbol{q}}|u_n\rangle \\ &+ \int d\boldsymbol{p}\, C_0 e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\frac{1}{2}\boldsymbol{B}\times\dot{\boldsymbol{r}}_c\cdot i\partial_{\boldsymbol{p}}|u_0\rangle + \sum_{n\neq 0}\int d\boldsymbol{p}\, C_n e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\frac{1}{2}\boldsymbol{B}\times\dot{\boldsymbol{r}}_c\cdot i\partial_{\boldsymbol{p}}|u_n\rangle\,. \end{aligned} \tag{A.2}$$

In Eq.(A.2), the first two terms describe the effect of the dynamical phase. The remaining two terms are due to the change of the Bloch states in the parameter space spanned by $\boldsymbol{r}_c$, which give rise to the Berry phase effect on electron dynamics.

The energetic part of the Schrödinger equation is

$$\begin{aligned}\hat{H}|\Psi\rangle = &\int d\boldsymbol{p}\,\varepsilon_0 e^{i\boldsymbol{p}\cdot\boldsymbol{q}} C_0(\boldsymbol{p})|u_0\rangle + \int d\boldsymbol{p}\, C_0 e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\hat{H}'|u_0\rangle \\ &+ \sum_{n\neq 0}\int d\boldsymbol{p}\,\varepsilon_n e^{i\boldsymbol{p}\cdot\boldsymbol{q}} C_n(\boldsymbol{p})|u_n\rangle + \sum_{n\neq 0}\int d\boldsymbol{p}\, C_n e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\hat{H}'|u_n\rangle\,, \qquad \text{(A.3)}\end{aligned}$$

where $\hat{H}''$ is discarded because we only focus on the leading order contribution to C_n.

The Schrödinger equation sets up a group of linear equations for $C_0(\boldsymbol{p})$ and $C_n(\boldsymbol{p})$. All terms in this equation array can be determined by changing the integration variable from $\boldsymbol{p}$ to $\boldsymbol{p}'$ in Eq.(A.2) and (A.3) and take the inner product $\langle u_n|e^{-i\boldsymbol{p}\cdot\boldsymbol{q}}$ to both sides of the Schrödinger equation. For the dynamical part, we have

$$\begin{aligned}\langle u_n|e^{-i\boldsymbol{p}\cdot\boldsymbol{q}} i\partial_t|\Psi\rangle &= C_0(\boldsymbol{p})\frac{1}{2}\boldsymbol{B}\times\dot{\boldsymbol{r}}_c\cdot\langle u_n|i\partial_{\boldsymbol{p}}|u_0\rangle + \varepsilon C_n(\boldsymbol{p}) \\ &\quad + \sum_{m\neq 0} C_m(\boldsymbol{p})\frac{1}{2}\boldsymbol{B}\times\dot{\boldsymbol{r}}_c\cdot\langle u_n|i\partial_{\boldsymbol{p}}|u_m\rangle \\ &= C_0(\boldsymbol{p})\frac{1}{2}\boldsymbol{B}\times\boldsymbol{v}_0\cdot\boldsymbol{A}_{n0} + \varepsilon C_n(\boldsymbol{p})\,. \qquad \text{(A.4)}\end{aligned}$$

In the last step, the term with C_m is discarded since C_m is already of first order in fields, which makes the term of second order in total. We also substitute $\dot{\boldsymbol{r}}_c$ by the band group velocity $\boldsymbol{v}_0$, which is valid since the correction to $\boldsymbol{v}_0$ is at

least of first order in fields. $\boldsymbol{A}_{n0}$ is the interband Berry connection. We find that the Berry phase effect contributes to the interband mixing, which is half the gauge independent work done by the Lorentz force.

For the energetic part Eq.(A.3), we have:

$$\begin{aligned}
\langle u_n|e^{-i\boldsymbol{p}\cdot\boldsymbol{q}}\hat{H}|\Psi\rangle &= \int d\boldsymbol{p}'\varepsilon_0(\boldsymbol{p}')C_0(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
&+ \sum_{m\neq 0}\int d\boldsymbol{p}'\varepsilon_n(\boldsymbol{p}')C_m(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}|u_m(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
&+ \int d\boldsymbol{p}'C_0(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{H}'|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
&+ \sum_{m\neq 0}\int d\boldsymbol{p}'C_m(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{H}'|u_m(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
&= \varepsilon_n(\boldsymbol{p})C_n(\boldsymbol{p}) \\
&+ \sum_{m\neq 0}\int d\boldsymbol{p}'C_m(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{H}'|u_m(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle\,.
\end{aligned} \tag{A.5}$$

We equate Eq.(A.4) with Eq.(A.5) to solve $C_n(\boldsymbol{p})$:

$$\begin{aligned}
(\varepsilon_0-\varepsilon_n)C_n &= -\frac{1}{2}\boldsymbol{B}\times\boldsymbol{v}_0\cdot\boldsymbol{A}_{n0}C_0 \\
&+ \sum_{m\neq 0}\int d\boldsymbol{p}'C_m(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{H}'|u_m(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle\,.
\end{aligned} \tag{A.6}$$

Since $\hat{H}' = \frac{1}{4}\boldsymbol{B}\cdot((\boldsymbol{q}-\boldsymbol{r}_c)\times\hat{\boldsymbol{V}}-\hat{\boldsymbol{V}}\times(\boldsymbol{q}-\boldsymbol{r}_c))$, the last term of Eq.(A.6)

can be evaluated as follows:

$$
\begin{aligned}
I = & \frac{1}{4}\boldsymbol{B}\times i\partial_{\boldsymbol{p}}\cdot\int d\boldsymbol{p}' C_0(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{\boldsymbol{V}}|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
& -\frac{1}{4}\boldsymbol{B}\cdot\int d\boldsymbol{p}' C_0(\boldsymbol{p}') i\langle\partial_{\boldsymbol{p}}u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|\times e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{\boldsymbol{V}}|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
& +\frac{1}{4}\boldsymbol{B}\cdot\int d\boldsymbol{p}' i\partial_{\boldsymbol{p}'}C_0(\boldsymbol{p}')\times\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{\boldsymbol{V}}|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
& -\frac{1}{4}\boldsymbol{B}\cdot\int d\boldsymbol{p}' C_0(\boldsymbol{p}')\langle u_n(\boldsymbol{p}+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)|e^{i(\boldsymbol{p}'-\boldsymbol{p})\cdot\boldsymbol{q}}\hat{\boldsymbol{V}}\times i\partial_{\boldsymbol{p}'}|u_0(\boldsymbol{p}'+\frac{1}{2}\boldsymbol{B}\times\boldsymbol{r}_c)\rangle \\
& -\frac{1}{2}C_0(\boldsymbol{p})\boldsymbol{B}\times\boldsymbol{r}_c\cdot\boldsymbol{V}_{n0} \\
= & \frac{1}{4}\boldsymbol{B}\times i\partial_{\boldsymbol{p}}\cdot(C_0(\boldsymbol{p})\boldsymbol{V}_{n0})-\frac{1}{4}\boldsymbol{B}\cdot i\langle\partial_{\boldsymbol{p}}u_n|\times\hat{\boldsymbol{V}}|u_0\rangle C_0(\boldsymbol{p}) \\
& +\frac{1}{4}\boldsymbol{B}\cdot i\partial_{\boldsymbol{p}}C_0(\boldsymbol{p})\times\boldsymbol{V}_{n0}-\frac{1}{4}\boldsymbol{B}\cdot\langle u_n|\hat{\boldsymbol{V}}\times i\partial_{\boldsymbol{p}}|u_0\rangle C_0(\boldsymbol{p})-\frac{1}{2}C_0(\boldsymbol{p})\boldsymbol{B}\times\boldsymbol{r}_c\cdot\boldsymbol{V}_{n0}\,.
\end{aligned}
\tag{A.7}
$$

We can group the terms appearing in Eq.(9) as

$$
\begin{aligned}
I = & \frac{1}{2}\boldsymbol{B}\cdot(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)C_0\times\boldsymbol{V}_{n0}+\frac{1}{4}\boldsymbol{B}\cdot[i\partial_{\boldsymbol{p}}\times\boldsymbol{V}_{n0}+(\boldsymbol{a}_n-\boldsymbol{a}_0)\times\boldsymbol{V}_{n0}]-\boldsymbol{B}\cdot\boldsymbol{M}_{n0}C_0 \\
& -\frac{1}{4}\boldsymbol{B}\cdot(\boldsymbol{v}_0\times\boldsymbol{A}_{n0}+\boldsymbol{v}_n\times\boldsymbol{A}_{n0})C_0,
\end{aligned}
\tag{A.8}
$$

where $\hat{\boldsymbol{D}}=i\partial_{\boldsymbol{p}}+\boldsymbol{a}$ is the covariant derivative, and $\boldsymbol{M}_{n0}=-\frac{1}{4}\sum_{m\neq 0,n}(\boldsymbol{A}_{nm}\times\boldsymbol{V}_{m0}-\boldsymbol{V}_{nm}\times\boldsymbol{A}_{m0})$. The last three terms in Eq.(A.8) can be combined to give the interband element of the magnetic moment operator, which renders a clean result:

$$
I=\frac{1}{2}\boldsymbol{B}\cdot(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)C_0\times\boldsymbol{V}_{n0}-\boldsymbol{B}\cdot\mathcal{M}^r_{n0}\,,
\tag{A.9}
$$

where $\mathcal{M}^r_{n0}$ is the interband magnetic moment of the wave packet in the center of mass reference frame:

$$
\mathcal{M}^r_{n0}=-\frac{1}{4}[i\partial_{\boldsymbol{p}}\times\boldsymbol{V}_{n0}+(\boldsymbol{a}_n-\boldsymbol{a}_0)\times\boldsymbol{V}_{n0}]+\boldsymbol{M}_{n0}+\frac{1}{4}(\boldsymbol{v}_0\times\boldsymbol{A}_{n0}+\boldsymbol{v}_n\times\boldsymbol{A}_{n0})\,.
\tag{A.10}
$$

The first term in Eq.(A.10) is anti-Hermitian, while the other two terms are Hermitian.

Eq.(A.6) and (A.9) together yield the correct coefficient C_n:

$$C_n = \frac{\frac{1}{2}\boldsymbol{B}\cdot(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)C_0\times\boldsymbol{V}_{n0}-\boldsymbol{B}\cdot\mathcal{M}_{n0}^{tot}C_0}{\varepsilon_0-\varepsilon_n}\,. \tag{A.11}$$

where $\mathcal{M}_{n0}^{tot}$ is the total interband magnetic moment:

$$\mathcal{M}_{n0}^{tot} = \mathcal{M}_{n0}^{r} + \frac{1}{2}\boldsymbol{v}_0\times\boldsymbol{A}_{n0}\,. \tag{A.12}$$

It is easy to check that C_n is gauge-indepedent. It contains both the contribution from the Berry phase effect and the perturbative Hamiltonian $\hat{H}'$.

Appendix B

Band Energy up to Second Order

The wave-packet energy has several parts. We first consider the contribution from $\hat{H}_c$ first.

$$\langle\Psi|\hat{H}_c|\Psi\rangle = \int d\boldsymbol{p} C_0^\star C_0 \varepsilon_0 + \sum_{n\neq 0}\int d\boldsymbol{p} C_n^\star C_n \varepsilon_n \,. \tag{B.1}$$

There is an important correction to Eq.(B.1): when the mixing of Bloch states C_n is taken into account, the normalization of the wave-packet is modified:

$$\langle\Psi|\Psi\rangle = \int \left(|C_0|^2 + \sum_{n\neq 0}|C_n|^2\right) d^3\boldsymbol{p} \,. \tag{B.2}$$

To make the Euler-Lagrangian method valid, the wave-packet must be normalized, at least up to second order for our purpose. Therefore, C_0 must have a second order correction: $C_0 \to (1+\delta)C_0$, with $\delta = -\frac{1}{2}\sum_{n\neq 0}\int |C_n|^2 d^3p$.

This correction δ contributes to the energy in Eq.(B.1):

$$\langle\Psi|\hat{H}_c|\Psi\rangle = \varepsilon_0 - \sum_{n\neq 0}\int d\boldsymbol{p} C_n^\star C_n(\varepsilon_0 - \varepsilon_n) + \delta\varepsilon_c \,. \tag{B.3}$$

Here, the term $\delta\varepsilon_c$ arises due to the horizontal mixing in the coefficient C_n. Compared with the renormalization condition in Eq.(B.2), the integration in Eq.(B.1) has an additional energy factor ε_0 in the integrand, which leads to

the additional term $\delta\varepsilon_c$ due to the derivative of C_0 involved in the horizontal mixing. However, this $\delta\varepsilon_c$ is cancelled by the contribution from the dynamical part of the Lagrangian:

$$\delta L_{dyn} = \int d\boldsymbol{p} C_0^\star \delta i\partial_t C_0 - \sum_{n\neq 0}\int d\boldsymbol{p} C_n^\star i\partial_t C_n \,. \tag{B.4}$$

We only need to keep the term up to the second order in $\boldsymbol{B}$:

$$\begin{aligned} i\partial_t C_n = {} & \frac{G_{n0}}{\varepsilon_0-\varepsilon_n} i\partial_t C_0 + \frac{i}{2}(\boldsymbol{B}\times\boldsymbol{A}_{n0})\cdot(i\boldsymbol{v}_0 - i\dot{\boldsymbol{r}}_c)C_0 \\ & + \varepsilon_0\frac{i}{2}(\boldsymbol{B}\times\boldsymbol{A}_{n0})\cdot(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)C_0 \,. \end{aligned} \tag{B.5}$$

Plug Eq.(B.5) into Eq.(B.4), and we have

$$\delta L_{dyn} = -\delta\varepsilon_c + \sum_{n\neq 0}\frac{1}{8}\alpha_{ij}(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j \,, \tag{B.6}$$

where $\alpha_{ij} = \partial_{ij}\varepsilon_0$ is the inverse effective mass tensor.

Then we calculate the contribution to the wave-packet energy from the gradient correction $\hat{H}'$:

$$\begin{aligned} \langle\Psi|\hat{H}'|\Psi\rangle = {} & -\boldsymbol{B}\cdot\boldsymbol{m} + \boldsymbol{B}\cdot(\boldsymbol{v}_0\times\boldsymbol{a}_0') \\ & + 2\sum_{n\neq 0}\frac{G_{0n}G_{n0}}{\varepsilon_0-\varepsilon_n} - \sum_{n\neq 0}\frac{1}{4}\alpha_{ij}(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j \\ & + \sum_{n\neq 0}\frac{1}{2}\partial_i[(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i G_{n0} + c.c.] \\ & + \sum_{n\neq 0}\frac{1}{4}\int d\boldsymbol{p}\{[\boldsymbol{B}\times(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)]_i^\star C_0^\star[\boldsymbol{B}\times(\hat{\boldsymbol{D}}-\boldsymbol{r}_c)]_j C_0 \\ & \qquad [-i(V_i)_{0n}(A_j)_{n0}] + c.c.\}\,, \end{aligned} \tag{B.7}$$

where $\boldsymbol{m}$ is the orbital magnetic moment: $\boldsymbol{m} = -\frac{1}{2}\mathrm{Im}\langle\boldsymbol{\partial}u_0| \times (\varepsilon_0 - \hat{H}_c)|\boldsymbol{\partial}u_0\rangle$, and $\boldsymbol{a}_0'$ is the positional shift:

$$\boldsymbol{a}_0' = \sum_{n\neq 0} \frac{G_{0n}\boldsymbol{A}_{n0}}{\varepsilon_0 - \varepsilon_n} + \frac{1}{4}\partial_i[(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i\boldsymbol{A}_{n0}] + c.c.\,. \tag{B.8}$$

The remaining contribution to the wave-packet energy is from the second order correction $\hat{H}''$:

$$\begin{aligned}
\langle\Psi|\hat{H}''|\Psi\rangle = &-\frac{1}{8}\langle 0|\Gamma_{ij}|0\rangle(\boldsymbol{B}\times\boldsymbol{r}_c)_i(\boldsymbol{B}\times\boldsymbol{r}_c)_j \\
&-\sum_{n\neq 0}\frac{1}{8}[(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i(\Gamma_{ij})_{n0}(\boldsymbol{B}\times\boldsymbol{r}_c)_j + c.c.] \\
&+\frac{1}{8}\int d\boldsymbol{p}(\boldsymbol{B}\times\boldsymbol{\partial}|C_0|)_i(\boldsymbol{B}\times\boldsymbol{\partial}|C_0|)_j\langle 0|\Gamma_{ij}|0\rangle \\
&+\frac{1}{8}\langle 0|\Gamma_{ij}|0\rangle(\boldsymbol{B}\times\boldsymbol{r}_c)_i(\boldsymbol{B}\times\boldsymbol{r}_c)_j \\
&+\sum_{n\neq 0}\frac{1}{8}[(\boldsymbol{B}\times\boldsymbol{A}_{0n})_i(\Gamma_{ij})_{n0}(\boldsymbol{B}\times\boldsymbol{r}_c)_j + c.c.] \\
&+\sum_{n\neq 0}\frac{1}{16}(\boldsymbol{B}\times i\boldsymbol{\partial})_i[(\Gamma_{ij})_{0n}(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j] - (i\leftrightarrow j) \\
&+\frac{1}{8}\sum_{(m,n)\neq 0}(\boldsymbol{B}\times\boldsymbol{A}_{0m})_i(\Gamma_{ij})_{mn}(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j\,.
\end{aligned} \tag{B.9}$$

After some cancellations, we have

$$\begin{aligned}
\langle\Psi|\hat{H}''|\Psi\rangle = &-\frac{1}{16}(\boldsymbol{B}\times\boldsymbol{\partial})_i(\boldsymbol{B}\times\boldsymbol{\partial})_j\langle 0|\mathit{\Gamma}_{ij}|0\rangle \\
&+\frac{1}{8}\sum_{(m,n)\neq 0}(\boldsymbol{B}\times\boldsymbol{A}_{0m})_i(\mathit{\Gamma}_{ij})_{mn}(\boldsymbol{B}\times\boldsymbol{A}_{n0})_j\,,
\end{aligned} \tag{B.10}$$

where $(\mathit{\Gamma}_{ij})_{mn} = \langle u_m|\mathit{\Gamma}_{ij}|u_n\rangle$. For relativistic Dirac Hamiltonian, L'' simply vanishes. For nonrelativistic Schrödinger and Pauli Hamiltonian, $\mathit{\Gamma}_{ij}$ is inverse

mass times the identity matrix, and Eq.(B.10) reduces to a compact form:

$$\langle\Psi|\hat{H}''|\Psi\rangle = \frac{1}{8m}(B^2 g_{ii} - B_i g_{ij} B_j)\,, \tag{B.11}$$

where $g_{ij} = \mathrm{Re}\langle\partial_i u_0|\partial_j u_0\rangle - a_i a_j$ is the quantum metric of k-space [3, 53].

By combining Eq.(B.3),(B.6),(B.7),(B.11) and the relation between effective mass tensor α_{ij} and the Hessian matrix Γ_{ij} of $\hat{H}_c$:

$$\langle 0|\Gamma_{ij}|0\rangle = \alpha_{ij} + \sum_{n\neq 0}[-i(A_i)_{0n}(V_j)_{n0} + c.c.]\,, \tag{B.12}$$

the wave-packet energy can be put in a compact form (assume $\Gamma_{ij} = \delta_{ij}/m$):

$$\begin{aligned}\tilde{\varepsilon} =& \varepsilon_0 - \boldsymbol{B}\cdot\boldsymbol{m} \\ &+ \frac{1}{4}(\boldsymbol{B}\cdot\boldsymbol{\Omega})(\boldsymbol{B}\cdot\boldsymbol{m}) - \frac{1}{8}\epsilon_{sik}\epsilon_{tj\ell}B_s B_t g_{ij}\alpha_{k\ell} \\ &- \boldsymbol{B}\cdot(\boldsymbol{a}_0'\times\boldsymbol{v}_0) + \nabla\cdot\boldsymbol{P}_{\mathrm{E}} \\ &+ \sum_{n\neq 0}\frac{G_{0n}G_{n0}}{\varepsilon_0 - \varepsilon_n} + \frac{1}{8m}(B^2 g_{ii} - B_i g_{ij} B_j).\end{aligned} \tag{B.13}$$

Various quantities in Eq.(B.13) have been explained in the main text.

Publication List

1. Y. Gao, S. A. Yang, and Q. Niu, *field induced positional shift of Bloch electrons and its dynamical implications*, Phys. Rev. Lett. 112, 166601 (2014).

2. Y. Gao, S. A. Yang, and Q. Niu, *geometrical effects in orbital magnetic susceptibility*, Phys. Rev. B 91, 214405 (2015).

3. H. Chen, Y. Gao, D. Xiao, A. H. MacDonald, and Q. Niu, *semiclassical theory of linear magnetoresistance in crystalline conductors with broken time-reversal symmetry*, arXiv: 1511.02557 (submitted to PRL).

4. Y. Gao and Q. Niu, *zero-field magnetic response functions in Landau levels*, arXiv: 1507.06342 (submitted to PRL).

Bibliography

[1] E. N. Adams. *Phys. Rev.*, 89:633, 1953.

[2] J. G. Analytis, R. D. McDonald, S. C. Riggs, J.-H. Chu, G. S. Boebinger, and I. R. Fisher. *Nature Phys.*, 10:960, 2010.

[3] J. Anandan and Y. Aharonov. *Phys. Rev. Lett.*, 65:1697, 1990.

[4] D. Bercious, D. F. Urban, H. Grabert, and W. Häusler. *Phys. Rev. A*, 80:063603, 2009.

[5] E. I. Blount. *Phys. Rev.*, 126:1636, 1962.

[6] E. I. Blount. *Phys. Rev.*, 128:2454, 1962.

[7] R. Bott. *Bull. Amer. Math. Soc. (N.S.)*, 7:331, 1982.

[8] C. Brüne, C. X. Liu, E. G. Novik, E. M. Hankiewicz, H. Buhmann, Y. L. Chen, X. L. Qi, Z. X. Shen, S. C. Zhang, and L. W. Molenkamp. *Phys. Rev. Lett.*, 106:126803, 2011.

[9] T. Cai, S. A. Yang, X. Li, F. Zhang, J. Shi, W. Yao, and Q. Niu. *Phys. Rev. B*, 88:115140, 2013.

[10] P. Carmier and D. Ullmo. *Phys. Rev. B*, 77:245413, 2008.

[11] M. C. Chang. *Rev. Mod. Phys.*, 82:1959, 2010.

[12] M. C. Chang and Q. Niu. *Phys. Rev. Lett.*, 75:1348, 1995.

[13] M.-C. Chang and Q. Niu. *J. Phys.: Condens. Matter*, 20:193202, 2008.

[14] C. R. Dean, L. Wang, P. Maher, C. Forsythe, F. Ghahari, Y. Gao, J. Katoch, M. Ishigami, P. Moon, M. Koshino, T. Taniguchi, K. Watanabe, K. L. Shepard, J. Hone, and P. Kim. *Nature*, 497:598, 2013.

[15] C. A. Duarte, G. M. Gusev, A. A. Quivy, T. E. Lamas, A. K. Bakarov, and J. C. Portal. *Phys. Rev. B*, 76:075346, 2007.

[16] B. Sacépé, J. B. Oostinga, J. Li, A. Ubaldini, N. J. G. Couto, E. Giannini, and A. F. Morpurgo. *Nature Commun.*, 2:575, 2011.

[17] A. M. Essin, J. E. Moore, and D. Vanderbilt. *Phys. Rev. Lett*, 102:146805, 2009.

[18] A. M. Essin, A. M. Turner, J. E. Moore, and D. Vanderbilt. *Phys. Rev. B*, 81:205104, 2010.

[19] C. Fang, M. J. Gilbert, and B. A. Bernevig. *Phys. Rev. B*, 86:115112, 2012.

[20] R. C. Fivaz. *Phys. Rev.*, 183:586, 1969.

[21] L. L. Foldy and S. A. Wouthuysen. *Phys. Rev.*, 78:29, 1950.

[22] L. Fu, C. L. Kane, and E. J. Mele. *Phys. Rev. Lett.*, 98:106803, 2007.

[23] J. N. Fuchs, F. Piéchon, M. O. Goerbig, and G. Montambaux. *Eur. Phys. J. B*, 77:351, 2010.

[24] H. Fukuyama. *Prog. Theor. Phys.*, 45:704, 1971.

[25] G. Gómez-Santos and T. Stauber. *Phys. Rev. Lett.*, 106:045504, 2011.

[26] J. P. Guillement, B. Helffer, and P. Treton. *J. Phys. France*, 50:2019, 1989.

[27] F. D. M. Haldane. *Phys. Rev. Lett.*, 61:2015, 1988.

[28] M. Z. Hasan and C. L. Kane. *Rev. Mod. Phys.*, 82:3045, 2010.

[29] J. E. Hebborn and E. H. Sondheimer. *J. Phys. Chem. Solids*, 13:105, 1960.

[30] D. R. Hofstadter. *Phys. Rev. B*, 14:2239, 1976.

[31] P. Hosur, S. Ryu, and A. Vishwanath. *Phys. Rev. B*, 81:045120, 2010.

[32] T. L. Hughes, E. Prodan, and B. A. Bernevig. *Phys. Rev. B*, 83:245132, 2011.

[33] T. Jungwirth, Q. Niu, and A. H. MacDonald. *Phys. Rev. Lett.*, 88:207208, 2002.

[34] Y. Kagan and L. A. Maksimov. *Phys. Rev. Lett.*, 100:145902, 2008.

[35] R. Karplus and J. M. Luttinger. *Phys. Rev.*, 95:1154, 1954.

[36] R. D. King-Smith and D. Vanderbilt. *Phys. Rev. B*, 47:1651, 1993.

[37] W. Kohn. *Proc. Phys. Soc.*, 72:1147, 1958.

[38] Th. Koschny and L. Schweitzer. *Physica B*, 298:52, 2001.

[39] M. Koshino and T. Ando. *Phys. Rev. B*, 76:085425, 2007.

[40] R. B. Laughlin. *Phys. Rev. Lett.*, 52:2304, 1984.

[41] X. Y. Lee, H. W. Jiang, and W. J. Schaff. *Phys. Rev. Lett.*, 83:3701, 1999.

[42] B. Leung and E. Prodan. *J. Phys. A: Math. Theor.*, 46:085205, 2013.

[43] R. G. Littlejohn and W. G. Flynn. *Phys. Rev. A*, 44:5239, 1991.

[44] D. Z. Liu, X. C. Xie, and Q. Niu. *Phys. Rev. Lett.*, 76:975, 1996.

[45] J. M. Luttinger and W. Kohn. *Phys. Rev.*, 97:869, 1955.

[46] A. H. MacDonald. *Phys. Rev. B*, 28:6713, 1983.

[47] G. P. Mikitik and Yu. V. Sharlai. *Phys. Rev. Lett.*, 82:2147, 1999.

[48] P. K. Misra and L. Kleinman. *Phys. Rev. B*, 5:4581, 1972.

[49] J. E. Moore and L. Balents. *Phys. Rev. B*, 75:121306, 2007.

[50] J. E. Moyal. *Proc. Camb. Phil. Soc.*, 45:99, 1949.

[51] N. Nagaosa, J. Sinova, and S.Onoda. *Rev. Mod. Phys.*, 82:1539, 2010.

[52] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. *Rev. Mod. Phys.*, 81:109, 2009.

[53] T. Neupert, C. Chamon, and C. Mudry. *Phys. Rev. B*, 87:245103, 2013.

[54] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov. *Nature (London)*, 438:197, 2005.

[55] M. Onoda and N. Nagaosa. *J. Phys. Soc. Jpn.*, 71:19, 2002.

[56] S. Onoda, N. Sugimoto, and N. Nagaosa. *Phys. Rev. Lett.*, 97:126602, 2006.

[57] L. Onsager. *Phil. Mag. 43*, 43:1006, 1952.

[58] H. Potempa and L. Schweitzer. *Physica B*, 298:52, 2001.

[59] X. L. Qi, T. L. Hughes, and S. C. Zhang. *Phys. Rev. B*, 78:195424, 2008.

[60] X. L. Qi and S. C. Zhang. *Rev. Mod. Phys.*, 83:1057, 2011.

[61] T. Qin, J. Zhou, and J. Shi. *Phys. Rev. B*, 86:104305, 2012.

[62] D.-X. Qu, Y. S. Hor, J. Xiong, R. J. Cava, and N. P. Ong. *Science*, 329:821, 2010.

[63] R. Rammal. *J. Physique*, 46:1345, 1985.

[64] A. Raoux, M. Morigi, J.-N. Fuchs, F. Piéchon, and G. Montambaux. *Phys. Rev. Lett.*, 112:026402, 2014.

[65] A. Raoux, F. Piéchon, J.-N. Fuchs, and G. Montambaux. *Phys. Rev. B*, 91:085120, 2015.

[66] Z. Ren, A. A. Taskin, S. Sasaki, K. Segawa, and Y. Ando. *Phys. Rev. B*, 82:241306, 2010.

[67] R. Resta. *Rev. Mod. Phys.*, 66:899, 1994.

[68] L. M. Roth. *J. Phys. Chem. Solids*, 23:433, 1962.

[69] L. M. Roth. *Phys. Rev.*, 145:434, 1966.

[70] R. Roy. *Phys. Rev. B*, 79:195322, 2009.

[71] G. A. H. Schober, H. Murakawa, M. S. Bahramy, R. Arita, Y. Kaneko, Y. Tokura, and N. Nagaosa. *Phys. Rev. Lett.*, 108:247208, 2012.

[72] J. Shi, G. Vignale, D. Xiao, and Q. Niu. *Phys. Rev. Lett.*, 99:197202, 2007.

[73] P. Středa. *Phys. Rev. B*, 82:045115, 2010.

[74] P. Středa and T. Jonckheere. *Phys. Rev. B*, 82:045115, 2010.

[75] G. Sundaram and Q. Niu. *Phys. Rev. B*, 59:14915, 1999.

[76] A. A. Taskin and Y. Ando. *Phys. Rev. B*, 84:035301, 2011.

[77] A. A. Taskin, K. Segawa, and Y. Ando. *Phys. Rev. B*, 82:121302, 2010.

[78] G. Vignale. *Phys. Rev. Lett.*, 57:358, 1991.

[79] G. H. Wannier. *Phys. Status Solidi B*, 88:757, 1978.

[80] G. H. Wannier and U. N. Upadhyaya. *Phys. Rev.*, 136:A803, 1964.

[81] H. Weyl. *Z. Physik*, 46:1, 1927.

[82] M. Wilkinson. *J. Phys. A*, 17:3459, 1984.

[83] F.-X. Xiang, X. L. Wang, M. Veldhorst, S.-X. Dou, and M. S. Fuhrer. *Phys. Rev. B*, 92:035123, 2015.

[84] D. Xiao, M. C. Chang, and Q. Niu. *Rev. Mod. Phys.*, 82:1348, 2010.

[85] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao. *Phys. Rev. Lett.*, 108:196802, 2012.

[86] D. Xiao, J. Shi, D. P. Clougherty, and Q. Niu. *Phys. Rev. Lett.*, 102:087602, 2009.

[87] D. Xiao, J. Shi, and Q. Niu. *Phys. Rev. Lett.*, 95:137204, 2005.

[88] D. Xiao, W. Yao, and Q. Niu. *Phys. Rev. Lett.*, 99:236809, 2007.

[89] J. Xiong, A. C. Petersen, D. Qu, R. J. Cava, and N. P. Ong. *Physica E*, 44:917, 2012.

[90] F. Xiu, L. He, Y. Wang, Z. Chen, J. Zou, A. Shailos, and K. L. Wang. *Nature Nano.*, 6:216, 2011.

[91] L. Zhang, J. Ren, J.-S. Wang, and B. Li. *Phys. Rev. Lett.*, 105:225901, 2010.

[92] Y. Zhang, Y.-W.Tan, H. L. Stormer, and P. Kim. *Nature (London)*, 438:201, 2005.