Carnegie Mellon University
MELLON COLLEGE OF SCIENCE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN THE FIELD OF PHYSICS

TITLE: "Topological Effects in Two-Dimensional Systems."

PRESENTED BY: Xiaoou Zhang

ACCEPTED BY THE DEPARTMENT OF PHYSICS

Di Xiao	5/20/19
DI XIAO, CHAIR PROFESSOR	DATE
Scott Dodelson	5/23/19
SCOTT DODELSON, DEPT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

Rebecca Doerge	5/23/19
REBECCA DOERGE, DEAN	DATE

Topological Effects in Two-Dimensional Systems

by

Xiaoou Zhang

Doctoral Thesis for the Degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

Advised by Professor Di Xiao

May 20, 2019

Abstract

Two-dimensional (2D) systems are the hatcheries of novel phenomena in condensed matter physics. For example, in graphene and transition metal dichalcogenides, the electronic behavior near the Fermi level is described by a 2D Dirac fermion model, which is the origin of many interesting Berry phase effects; recent studies for 2D ferromagnetic insulators such as CrI_3 revealed a series of novel optical and transport phenomena. The excitations in 2D systems manifest completely different properties compared to unconstrained free excitations in three-dimensional systems.

The excitations of interest in this thesis include electrons, excitons, phonons and magnons. We explain the nontrivial properties of these excitations in 2D systems from a topological point of view. We first show that we can construct electronic Chern insulators using graphene-hexagonal Boron Nitride superlattices. An effective mass theory for the conduction band electrons is derived using the Foldy-Wouthuysen (FW) transformation, the band projection method and the wave packet theory. This effective mass theory demonstrates how the Berry curvature distinguishes 2D Bloch band systems from free electron systems. Secondly, we show that the interaction between magnons and phonons can generate Berry curvatures, which can lead to the thermal Hall effect for magnon-phonon hybrid excitations even when the isolated magnon and phonon systems do not show thermal Hall effect separately. We also provide an analytical expression for the thermal Hall conductance as a function of the Berry curvature using the wave packet theory which is confirmed by the linear response theory. Finally, we unveil the importance of another topological number, the winding number. We found that the angular momenta of the bright exciton states in chiral fermion systems are determined by the winding number and the crystal symmetry. Based on our theory, we propose two chiral fermion systems capable of hosting dark s-like excitons: gapped surface states of a topological crystalline insulator with C_4 rotational symmetry and biased $3R$-stacked MoS_2 bilayers.

Acknowledgments

First and foremost, I want to thank my advisor Professor Di Xiao for guiding me through the five-year journey of scientific research. He is always motivating and supportive when I need help. He taught me to analyze problems from a high level rather than get submerged by details, while he is willing to patiently go though every details when I need knowledge to start a new topic or pick up a new technique. I enjoy every discussion with Di, where I learn lessons not only about physics, but also about life. I feel really fortunate to have Di as my advisor. Without him, this thesis would not exist.

I am grateful for the helpful advice and ideas from Professor Satoshi Okamoto and Professor Junren Shi. Satoshi provided important insight about Dzyaloshinskii-Moriya interactions in ferromagnetic systems and produced beautiful figures in the paper. Junren provided crucial advise on solving the magnon-phonon hybrid systems which saved me from a six-month-struggle. Many thanks to my committee members Professor Noa Marom, Professor Michael Widom and Professor Riccardo Penco for their valuable feedbacks during my rehearsal talks for the APS March Meeting and my talk for annual review.

I also want to thank all of my friends that are bonded together by physics. Postdoc Wenyu Shan spared no effort in helping me with the first paper in my career. I learnt a lot from him about the exciton physics and academic writing. Postdoc Yinhan Zhang helped me with his deep understanding in the linear response calculation for the thermal Hall conductance. Without his help, it will take several more months before the paper can be submitted. I am also thankful for all the informative discussions with postdoc Yang Gao, former postdocs Ran Cheng, Jianhui Zhou, and my former officemates Nikhil Sivadas, Matthew Daniels, Andrew Smith, Caitlin Carnahan and Nishchay Suri.

Last but not least, I want to thank all the support and love from my family. I am blessed to meet my wife Li He in Pittsburgh and she has been an inseparable part of my life since then. No matter I am in happiness or sadness, she is always with me. I also want to thank my parents for their unconditional and selfless support since the day I was born.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since the discovery of monolayer graphene in 2004 [5], 2D systems have attracted tremendous attention in recent studies. Graphene is a 2D metallic system with an ultrahigh electron mobility [6]. The discovery of graphene is followed by the fabrication of the first 2D semiconductor, monolayer MoS_2 in 2005 [7]. Monolayer MoS_2 has a direct band gap within the visible frequency range [8, 9], which makes it a good candidate to conduct optoelectronic studies. In 2017, the first 2D ferromagnetic insulator, monolayer CrI_3 is discovered [10]. It has a ferromagnetic ground state with spins aligned perpendicular to the 2D plane. The eruption of discoveries in recent years have unveiled a large landscape of 2D systems.

2D systems demonstrate unique properties due to the reduced dimensionality. As a consequence of the reduced dimensionality, the screening effects in 2D systems are weakened, thus many-body interactions are more prominent. For example, strong exciton effects have been observed in MoS_2 [11–13]. Moreover, compared to three dimensional systems, 2D systems are easier to be tuned by a gate voltage. One example is the tunable energy gap in the gated bilayer graphene [14]. In addition, stacking multiple layers of 2D materials on top of each other gives rise to van der Waals heterostructures [15]. These heterostructures have demonstrated emergent properties that are unseen in single layer systems, such as long-lifetime interband excitons where Bose-Einstein condensation can be realized [16, 17]. Furthermore, different twisting angles between layers with similar crystal structures lead to Moiré patterns with diverse properties. Recent studies have realized the superconducting state [18, 19] and the Mott insulator state [20] in the Moiré patterns formed by graphene layers.

Among the garden variety of novel phenomena in 2D systems, many of them are governed by the topological properties of the materials. The Berry phase of π around the Dirac point in graphene is responsible for the Landau level with zero energy [21]; Berry curvature leads to an anomalous velocity perpendicular to the driving force and generates the Hall current [1, 22]; the nonzero Chern number in 2D electronic systems gives rise to robust edge states crossing the band gap [23] and quantized Hall conductance [24–26]. Topology is a powerful tool to study condensed matter

problems and is expected to solve more mysteries in 2D systems.

In this chapter, I will first provide an introduction to the basic concepts of Berry phase and Berry curvature. Then I will introduce the wave packet theory, which will be used to derive the effective single band theory for electrons and the thermal Hall conductance in the following chapters. Finally, I will introduce three interesting 2D systems: graphene, transition metal dichalcogenides, and ferromagnetic insulators. These systems set the stage for the interesting phenomena in the following chapters.

1.1 Basics of Berry Phase

1.1.1 The Origin of Berry Phase

The Berry Phase originates from adiabatic evolutions of quantum systems [27]. Consider a system whose Hamiltonian $H(\boldsymbol{R})$ depends on a time dependent vector parameter represented by $\boldsymbol{R}(t) = (R_1(t), R_2(t), ...)$. The eigenvalue problem of this system is given by

$$H(\boldsymbol{R})|n(\boldsymbol{R})\rangle = E_n(\boldsymbol{R})|n(\boldsymbol{R})\rangle \ , \tag{1.1}$$

where $|n(\boldsymbol{R})\rangle$ is the n-th eigenstate with energy $E_n(\boldsymbol{R})$. If the parameter $\boldsymbol{R}(t)$ changes slowly enough, ie., the adiabatic condition

$$\hbar|\langle m(\boldsymbol{R})|\partial_t H(\boldsymbol{R})|n(\boldsymbol{R})\rangle| \ll (E_n - E_m)^2 \tag{1.2}$$

is satisfied, then a system initialized in an eigenstate $|n(\boldsymbol{R}(0))\rangle$ will always remain in the same instantaneous eigenstate $|n(\boldsymbol{R}(t))\rangle$ at any time t [28, 29]. Since the phase degree of freedom of the eigenstates is still unfixed, the wave function of a system starting from $|n(\boldsymbol{R}(0))\rangle$ has the following form:

$$|\psi_n(t)\rangle = e^{i\gamma_n(t)} e^{-i\int_0^t dt' E_n(\boldsymbol{R}(t'))/\hbar} |n(\boldsymbol{R}(t))\rangle \ . \tag{1.3}$$

Here the phase factor has two contributions. $e^{-i\int_0^t dt' E_n(\boldsymbol{R}(t'))/\hbar}$ is the dynamical phase, which also exists in systems described by time-independent Hamiltonians. The extra phase factor $e^{i\gamma_n(t)}$ is the geometric phase (the reason of the name will be explained later). This phase can be calculated by substituting the wave function (1.3) into the time-dependent Schrodinger equation $i\partial_t|\psi_n(t)\rangle = H(\boldsymbol{R}(t))|\psi_n(t)\rangle$, which gives

$$\gamma_n = \int_{\mathcal{C}} d\boldsymbol{R} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) \ , \tag{1.4}$$

where $\mathcal{C}$ is the contour describing the time evolution of $\boldsymbol{R}(t)$ in the parameter space, and $\boldsymbol{\mathcal{A}}_n(\boldsymbol{R})$ is the Berry connection defined as

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) = i\langle n(\boldsymbol{R})|\partial_{\boldsymbol{R}}|n(\boldsymbol{R})\rangle \ . \tag{1.5}$$

Note that the geometric phase γ_n in Eq. (1.4) is only explicitly related to $\boldsymbol{R}$ while the time evolution of $\boldsymbol{R}(t)$ is not important. That is why this phase factor is called the geometric phase.

Earlier belief [30] regards the geometric phase γ_n as a trivial quantity which can always be eliminated by a smooth gauge transformation. However, Berry [31] shows that this is not always possible if $\boldsymbol{R}(t)$ travels in a loop such that $|n(\boldsymbol{R}(0))\rangle = |n(\boldsymbol{R}(T))\rangle$. Under a gauge transformation $|n(\boldsymbol{R})\rangle \to e^{i\alpha(\boldsymbol{R})}|n(\boldsymbol{R})\rangle$, the Berry connection transforms accordingly:

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) \to \boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) - \partial_{\boldsymbol{R}}\alpha(\boldsymbol{R}) \; . \tag{1.6}$$

Therefore, the geometric phase transforms into $\gamma_n \to \gamma_n + \alpha(\boldsymbol{R}(t)) - \alpha(\boldsymbol{R}(0))$. It appears that we can always make $\gamma = 0$ by choosing a gauge transformation which satisfies $\alpha(\boldsymbol{R}(t)) - \alpha(\boldsymbol{R}(0)) = \gamma_n$. However, this is not true if the contour $\mathcal{C}$ is a closed loop. Under this circumstance, $|n(\boldsymbol{R}(0))\rangle = |n(\boldsymbol{R}(T))\rangle$ requires that the factor $e^{i\alpha(\boldsymbol{R})}$ should be a single valued function, meaning $\alpha(\boldsymbol{R}(t)) - \alpha(\boldsymbol{R}(0)) = 2m\pi$ where m is an integer. Under this constraint, γ_n is defined up to an integer times 2π and cannot always be made to 0 by a continuous gauge transformation. The geometric phase defined on a closed loop given by

$$\gamma_n = \oint_{\mathcal{C}} d\boldsymbol{R} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) \tag{1.7}$$

is known as the Berry phase. A continuous gauge transformation can only change the Berry phase by $2m\pi$.

Later in this thesis, we will frequently use a gauge independent quantity related to the Berry connection which is know as the Berry curvature. In three dimensional space, the Berry curvature is defined by

$$\boldsymbol{\Omega}_n(\boldsymbol{R}) = \nabla_{\boldsymbol{R}} \times \boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) \; ; \tag{1.8}$$

thus the Berry phase can be written as a surface integral over the Berry curvature

$$\gamma_n = \int_{\mathcal{S}} d\boldsymbol{S} \cdot \boldsymbol{\Omega}_n(\boldsymbol{R}) \; , \tag{1.9}$$

where $\mathcal{S}$ is the surface enclosed by the contour $\mathcal{C}$.

The Berry curvature is gauge independent. The gauge independence can be proved by applying the gauge transformation in Eq. (1.6) to the definition of the Berry curvature, while it can also be seen in the following equivalent expression for the z-component of Berry curvature

$$\Omega_n^z = i \sum_{m \neq n} \frac{\langle n|\partial_{R_x} H|m\rangle\langle m|\partial_{R_y} H|n\rangle - m \leftrightarrow n}{(E_n - E_m)^2} \; . \tag{1.10}$$

The right hand side of Eq. (1.10) is independent of the gauge choice of the wave functions $|n\rangle$.

In the later discussion for the wave packet theory in Section 1.2, we will show that the Berry curvature and the Berry connection play a similar role as the magnetic field and the vector potential in electrodynamics.

Note that the adiabatic condition in Eq. (1.2) must be satisfied in order to have a well defined Berry phase. One example of the violation of the adiabatic condition is a system with degeneracies, where the Berry curvature in Eq. (1.10) diverges. In this case, we need to consider the non-Abelian Berry curvature [32–34], which is beyond the scope of this thesis.

1.1.2 Berry Phase in Bloch Systems

The Berry phase is a successful tool in understanding a large range of interesting phenomena in Bloch band systems, such as the quantized adiabatic transport [35] and the intrinsic Hall effect [24–26]. Later in this thesis, I will show two other applications of the Berry phase in Bloch systems, including engineering topological Chern bands (Chapter 2) and the thermal Hall effect of magnon-phonon hybrid systems (Chapter 3).

We start from an electron in a periodic potential which satisfies $V(\boldsymbol{r}) = V(\boldsymbol{r}+\boldsymbol{a})$. According to Bloch theorem [36], the eigenstates for the periodic Hamiltonian have the following form:

$$\psi_{n\boldsymbol{q}}(\boldsymbol{r}) = e^{i\boldsymbol{q}\boldsymbol{r}} u_{n\boldsymbol{q}}(\boldsymbol{r}) \ , \tag{1.11}$$

where $u_{n\boldsymbol{q}}(\boldsymbol{r})$ is a periodic function with the same periodicity of the potential $V(\boldsymbol{r})$ such that $u_{n\boldsymbol{q}}(\boldsymbol{r}+\boldsymbol{a}) = u_{n\boldsymbol{q}}(\boldsymbol{r})$ [1]. n is the band index, and $\boldsymbol{q}$ is the wave vector. For simplicity, we assume that there is no degeneracy in the band structure such that the band index n is well defined. Note that $\psi_{n\boldsymbol{q}}(\boldsymbol{r}+\boldsymbol{a}) = e^{i\boldsymbol{q}\cdot\boldsymbol{a}}\psi_{n\boldsymbol{q}}(\boldsymbol{r})$, meaning the boundary conditions for different $\boldsymbol{q}$ are different, thus the eigenstates $\psi_{n\boldsymbol{q}}(\boldsymbol{r})$ with different $\boldsymbol{q}$ are not in the same Hilbert space [27]. In order to solve this problem, we do a unitary transformation to define a $\boldsymbol{q}$ dependent Hamiltonian.

$$H_{\boldsymbol{q}} = e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} H e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \ . \tag{1.12}$$

The eigenstates of $H_{\boldsymbol{q}}$ are just $u_{n\boldsymbol{q}}(\boldsymbol{r})$, whose boundary conditions for different $\boldsymbol{q}$ are the same.

The Berry connection and Berry curvature are defined using $u_{n\boldsymbol{q}}(\boldsymbol{r})$, and are written as

$$\boldsymbol{A}_n(\boldsymbol{q}) = i\langle u_{n\boldsymbol{q}}(\boldsymbol{r})|\partial_{\boldsymbol{q}}|u_{n\boldsymbol{q}}(\boldsymbol{r})\rangle \ , \tag{1.13}$$

and

$$\boldsymbol{\Omega}_n(\boldsymbol{q}) = i\nabla_{\boldsymbol{q}} \times \langle u_{n\boldsymbol{q}}(\boldsymbol{r})|\partial_{\boldsymbol{q}}|u_{n\boldsymbol{q}}(\boldsymbol{r})\rangle \ . \tag{1.14}$$

[1] Note that the form of $u_{n\boldsymbol{q}}(\boldsymbol{r})$ implies a gauge choice.

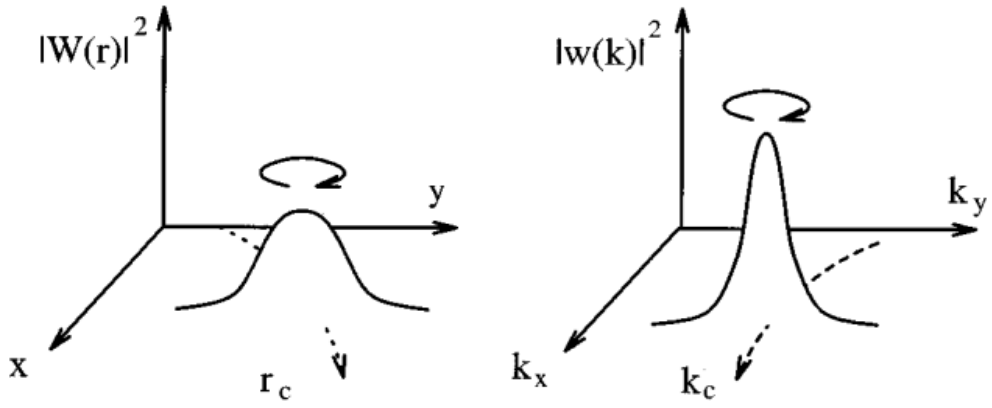

Figure 1.1: Schematic plot for the wave packet in the real space (left) and the momentum space (right). $|W(\boldsymbol{r})|$ and $|w(\boldsymbol{k})|$ are the envelope functions in the real space and momentum space, respectively. Figure is from Ref. [1].

In 2D Bloch systems, we are especially interested in the Berry phase around a loop enclosing the entire Brillouin zone (divided by 2π):

$$C_n = \frac{1}{2\pi}\int_{BZ} \Omega_n^z(\boldsymbol{q}) \ . \tag{1.15}$$

Since the Brillouin zone is a closed manifold, C_n is an integer known as the Chern number. The Chern number is a robust topological invariant. It cannot be changed without closing the band gap [37]. A nonzero Chern number is related to a series of interesting phenomena such as edge states crossing the band gap [23] and quantized Hall conductance [24–26].

1.2 Wave Packet Theory

The wave packet theory provides a semi-classical single electron picture to describe the motion of electrons in a Bloch band systems with smoothly varying external potentials. In the wave packet theory, electrons are described by a wave packet centered around $\boldsymbol{r}_c$ in the real space and $\boldsymbol{q}_c$ in the momentum space (see Fig. 1.1). The task is deriving the equation of motion for $\boldsymbol{r}_c$ and $\boldsymbol{q}_c$ under a slowly varying external potential. Note that we will set $\hbar = 1$ in future discussions.

We first construct the wave function of the wave packet. Consider a general form of the Hamiltonian

$$H[\boldsymbol{r}, \boldsymbol{p}; \beta_1(\boldsymbol{r}, t), ..., \beta_r(\boldsymbol{r}, t)] \ , \tag{1.16}$$

where $\beta_i(\boldsymbol{r}, t)$ are parameters that describe the external potential. Note that $\boldsymbol{r}$ and $\boldsymbol{p}$ are operators. Suppose the external potential changes slowly in the space, such that

it barely changes within the spread of the wave packet. Under this approximation, the Hamiltonian felt by a wave packet centered at $\boldsymbol{r}_c$ is given by

$$H = H_c + \sum_i \partial_{\beta_i} H_c \partial_{\boldsymbol{r}_c} \beta_i \cdot (\boldsymbol{r} - \boldsymbol{r}_c) + \mathcal{O}(\partial^2_{\boldsymbol{r}_c} \beta_i (\boldsymbol{r} - \boldsymbol{r}_c)^2) \ , \tag{1.17}$$

where $H_c = H[\boldsymbol{r}, \boldsymbol{p}; \beta_1(\boldsymbol{r}_c, t), ..., \beta_r(\boldsymbol{r}_c, t)]$ is the local Hamiltonian where the external potential is represented by its value at the center of the wave packet. For slowly varying potentials, only the first order expansions of $\beta_i(\boldsymbol{r}_c, t)$ are kept in the first order wave packet theory. Since $\beta_i(\boldsymbol{r}_c, t)$ are numbers rather than operators, the eigenstates of the local Hamiltonian are still Bloch wave functions. The wave function of the n-th band with momentum $\boldsymbol{q}$ and energy $E_{c,n}(\boldsymbol{r}_c, \boldsymbol{q}, t)$ is given by

$$H_c |\psi_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle = E_{c,n}(\boldsymbol{r}_c, \boldsymbol{q}, t) |\psi_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle \ . \tag{1.18}$$

The wave packet of the n-th band [2] is formed by a linear combination of the Bloch states $|\psi_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle$, which is given by

$$|\Psi_n\rangle = \int dq^3 w(\boldsymbol{q}, t) |\psi_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle = \int dq^3 w(\boldsymbol{q}, t) e^{i\boldsymbol{q}\cdot\boldsymbol{r}} |u_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle \ , \tag{1.19}$$

where $w(\boldsymbol{q}, t)$ is the envelope function and $|u_{n\boldsymbol{q}}(\boldsymbol{r}_c, t)\rangle$ is the periodical part of the Bloch function.

The wave packet center $\boldsymbol{r}_c$ and $\boldsymbol{q}_c$ are given by the envelope function. $\boldsymbol{q}_c$ is simply the center of the envelope function

$$\boldsymbol{q}_c = \int dq^3 \boldsymbol{q} |w(\boldsymbol{q}, t)|^2 \ . \tag{1.20}$$

The wave packet center in real space is given by $\boldsymbol{r}_c = \langle \Psi_n | \boldsymbol{r} | \Psi_n \rangle$. Simple derivation shows that [1, 22]

$$\boldsymbol{r}_c = \partial_{\boldsymbol{q}_c} \gamma_c + i \langle u_{n\boldsymbol{q}_c} | \partial_{\boldsymbol{q}_c} | u_{n\boldsymbol{q}_c} \rangle \ , \tag{1.21}$$

where $\gamma_c(t) = \gamma(\boldsymbol{q}_c, t)$ is the phase of the envelope function and is defined by $w(\boldsymbol{q}, t) = |w(\boldsymbol{q}, t)| e^{-i\gamma(\boldsymbol{q}, t)}$.

The equation of motion for $\boldsymbol{r}_c$ and $\boldsymbol{q}_c$ can be derived from the Lagrangian

$$\mathcal{L} = \langle \Psi_n | i \frac{d}{dt} - H | \Psi_n \rangle \ . \tag{1.22}$$

After some coarse-graining approximations which can be found in Ref. [1, 22], the Lagrangian takes the form of

$$\mathcal{L} \approx -\mathcal{E} + \boldsymbol{q}_c \cdot \dot{\boldsymbol{r}}_c + \dot{\boldsymbol{q}}_c \cdot i \langle u_n | \partial_{\boldsymbol{q}_c} | u_n \rangle + \dot{\boldsymbol{r}}_c \cdot i \langle u_n | \partial_{\boldsymbol{r}_c} | u_n \rangle + i \langle u_n | \partial_t | u_n \rangle \ , \tag{1.23}$$

[2] Here we assume that the bands have no degeneracy such that band indices are well defined.

where

$$\mathcal{E} = E_{c,n}(\boldsymbol{r}_c, \boldsymbol{q}_c, t) - \mathrm{Im}\langle \partial_{\boldsymbol{r}_c} u_n | \cdot (E_{c,n} - H_c) | \partial_{\boldsymbol{q}_c} u_n \rangle \ . \tag{1.24}$$

With the wave packet Lagrangian in Eq. (1.23), one can derive the equation of motion using $\dot{\boldsymbol{r}}_c = \partial_{\boldsymbol{q}_c} \mathcal{L}$ and $\dot{\boldsymbol{q}}_c = -\partial_{\boldsymbol{r}_c} \mathcal{L}$ which gives

$$\dot{\boldsymbol{r}}_c = \partial_{\boldsymbol{q}_c} \mathcal{E} - \boldsymbol{\Omega}_{\boldsymbol{qr}} \cdot \dot{\boldsymbol{r}}_c - \boldsymbol{\Omega}_{\boldsymbol{qq}} \cdot \dot{\boldsymbol{q}}_c + \boldsymbol{\Omega}_{t\boldsymbol{q}} \ , \tag{1.25}$$

$$\dot{\boldsymbol{q}}_c = -\partial_{\boldsymbol{r}_c} \mathcal{E} + \boldsymbol{\Omega}_{\boldsymbol{rr}} \cdot \dot{\boldsymbol{r}}_c + \boldsymbol{\Omega}_{\boldsymbol{rq}} \cdot \dot{\boldsymbol{q}}_c + \boldsymbol{\Omega}_{t\boldsymbol{r}} \ , \tag{1.26}$$

where

$$\Omega_{q_\alpha q_\beta} = \langle \partial_{q_\alpha} u | \partial_{q_\beta} u \rangle - \langle \partial_{q_\beta} u | \partial_{q_\alpha} u \rangle \tag{1.27}$$

is simply the Berry curvature, and $\boldsymbol{\Omega}_{\boldsymbol{rr}}$, $\boldsymbol{\Omega}_{\boldsymbol{qr}}$, $\boldsymbol{\Omega}_{t\boldsymbol{r}}$, $\boldsymbol{\Omega}_{t\boldsymbol{q}}$ are defined in the same way.

As an application, consider a wave packet in the electromagnetic field. The Hamiltonian takes the form of

$$H = H_0(\boldsymbol{q} + e\boldsymbol{A}(\boldsymbol{r}, t)) - e\phi(\boldsymbol{r}, t) \ , \tag{1.28}$$

where $\boldsymbol{A}$ is the vector potential, ϕ is the scalar potential, and $H_0(\boldsymbol{q})$ is the momentum dependent Bloch Hamiltonian defined in Eq. (1.12). Define the mechanical momentum

$$\boldsymbol{k} = \boldsymbol{q} + e\boldsymbol{A}(\boldsymbol{r}_c, t) \ . \tag{1.29}$$

Then the energy term in Eq. (1.24) becomes the Bloch band energy modified by the orbital magnetization energy and the electrostatic energy:

$$\mathcal{E} = E_{c,n}(\boldsymbol{k}_c, t) - e\phi(\boldsymbol{r}, t) - \boldsymbol{M} \cdot \boldsymbol{B} \ , \tag{1.30}$$

where B is the magnetic field, and the orbital magnetization is given by

$$\boldsymbol{M} = e\mathrm{Im}\langle \partial_{\boldsymbol{k}_c} u_n | \times (E_{c,n} - H_c) | \partial_{\boldsymbol{k}_c} u_n \rangle \ . \tag{1.31}$$

Here we have used $\partial_{\boldsymbol{q}_c} = \partial_{\boldsymbol{k}_c}$, $\partial_{\boldsymbol{r}_c} = e\partial_{\boldsymbol{r}_c} \boldsymbol{A} \cdot \partial_{\boldsymbol{k}}$ and $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. Using similar calculations, the Lagrangian can be simplified as

$$\begin{aligned} \mathcal{L} &= -\mathcal{E} + \boldsymbol{k}_c \cdot \dot{\boldsymbol{r}}_c - e\boldsymbol{A} \cdot \dot{\boldsymbol{r}}_c + \dot{\boldsymbol{k}}_c \cdot i\langle u_n | \partial_{\boldsymbol{k}_c} | u_n \rangle \\ &= -\mathcal{E} + \boldsymbol{k}_c \cdot \dot{\boldsymbol{r}}_c - e\boldsymbol{A} \cdot \dot{\boldsymbol{r}}_c + \dot{\boldsymbol{k}}_c \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{k}_c) \ . \end{aligned} \tag{1.32}$$

Notice that the last term is rewritten using the Berry connection $\boldsymbol{\mathcal{A}}_n$. Since the Berry connection and the vector potential enters the Lagrangian in the similar way, sometimes $\boldsymbol{\mathcal{A}}_n$ is regarded as the vector potential in the momentum space, and the Berry connection which is given by $\boldsymbol{\Omega}_n = \nabla \times \boldsymbol{\mathcal{A}}_n$ is the magnetic field in the momentum space [38].

Using the Euler-Lagrangian equation, the equation of motion for the wave packet in the magnetic field is given by

$$\dot{\boldsymbol{r}}_c = \partial_{\boldsymbol{k}_c} \mathcal{E} - \dot{\boldsymbol{k}}_c \times \boldsymbol{\Omega}_n(\boldsymbol{k}_c) \ , \tag{1.33}$$

$$\dot{\boldsymbol{k}}_c = -e\boldsymbol{E} - e\dot{\boldsymbol{r}}_c \times \boldsymbol{B} \ , \tag{1.34}$$

where $\boldsymbol{E} = -\nabla\phi$ is the electric field. The equation of motion also demonstrates the similarity between the magnetic field and Berry curvature. Notice that the second term in Eq. (1.33) can contribute to a velocity term that is perpendicular to the direction of $\boldsymbol{E}$. This term is called the anomalous velocity and can contribute to the thermal Hall effect. Later in this thesis, I will generalize the wave packet theory to the magnon-phonon coupled system to calculate the thermal Hall conductance.

We point out that the mechanical variables $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ are not canonical variables because of the last two terms in Eq. (1.32). We define canonical position and momentum variables as $\boldsymbol{k} = \boldsymbol{k}_c - e\boldsymbol{A}$ and $\boldsymbol{r} = \boldsymbol{r}_c - \boldsymbol{\mathcal{A}}$. It is easy to show that the Poisson bracket of these canonical variables is given by $\{r_i, k_j\}_{\text{Poission}} = \delta_{ij}$. In the canonical quantization procedure, the commutation relations of canonical operators satisfy $[r_i, k_j] = i\delta_{ij}$ and $[p_{cx}, p_{cy}] = [r_{cx}, r_{cy}] = 0$. On the other hand, the components of the mechanical variables are non-commutative. In a two dimensional system, the commutators are given by

$$[p_{cx}, p_{cy}] = -ieB_z \ , \tag{1.35}$$

and

$$[r_{cx}, r_{cy}] = i\Omega_n^z \ . \tag{1.36}$$

1.3 Background for 2D Systems

1.3.1 Graphene

Graphene is a single layer of carbon atoms with a 2D honeycomb structure, as shown in Fig. 1.2(a). The electronic ground state for an isolated carbon atom is s^2p^2. In graphene, the sp^2 hybridization leads to three hybridized orbitals and one p^z orbital. The symmetry axises of the three hybridized orbitals lie in the graphene plane form 120° angles between each other. Carbon atoms are connected by the σ-bonds formed by these hybridized orbitals, thus the honeycomb configuration is the direct consequence of the sp^2 hybridization. These orbitals are filled and become the valence bands of the graphene spectrum. On the other hand, the p^z orbitals are perpendicular to the graphene plane and forms π bonds. Since the p^z orbitals are half filled, the electronic properties of graphene near the Fermi level is dominated by the carbon p^z orbitals.

We construct a tight binding model of the p^z electrons. Using the nearest neighbor approximation, the tight binding Hamiltonian is given by [39]

$$H_{\text{g}} = -t\sum_{\langle i,j\rangle} a_i^\dagger b_j + H.c. \ , \tag{1.37}$$

where a_i and b_j are the annihilation operators of electrons on the A and B sublattices which are denoted in Fig. 1.2(a), and $t \approx 2.8$eV is the hopping energy. Since the

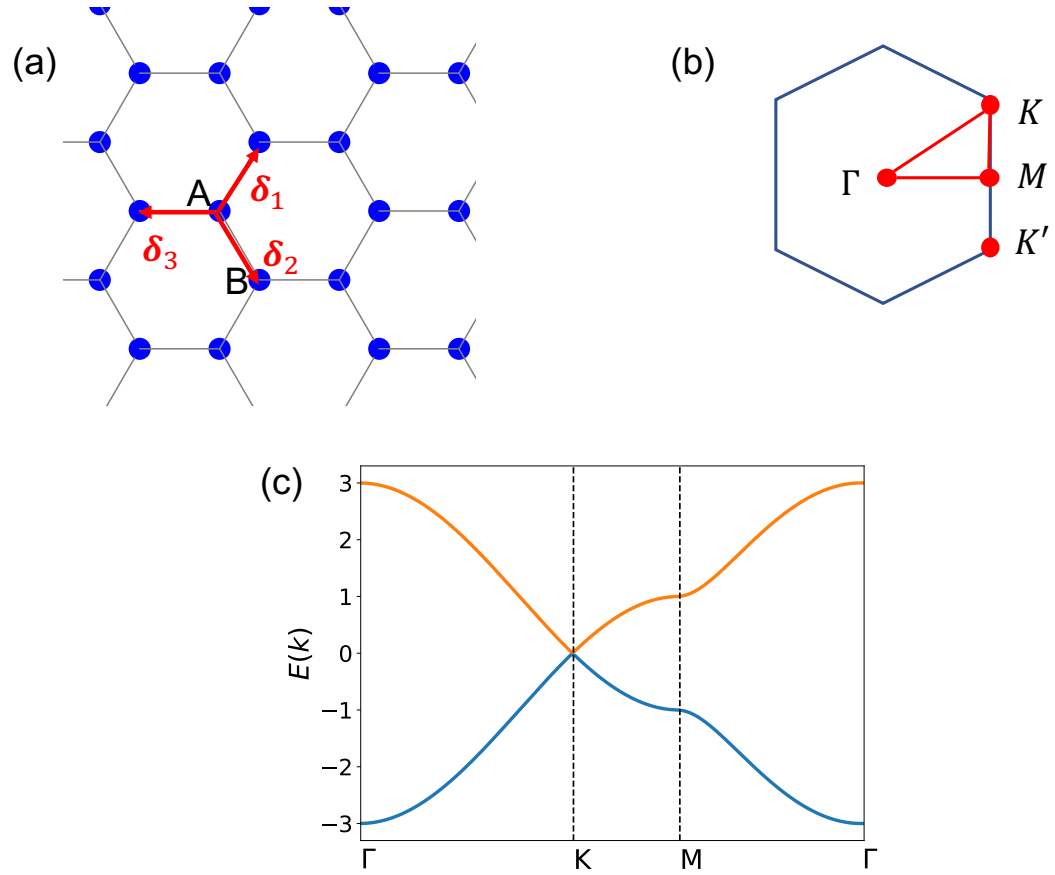

Figure 1.2: (a): The lattice structure of monolayer graphene, (b): the Brillouin zone showing the high symmetry points. (c): The band dispersion along the symmetry line $\Gamma - K - M - \Gamma$. The spectrum has the dispersion of massless Dirac fermions near the degeneracy at the K point.

hopping does not flip spins, spin is a conserved quantity and the spin index is ignored in Eq. (1.37). The three nearest neighbors vectors are given by $\boldsymbol{\delta}_1 = a(1/2, \sqrt{3}/2)$, $\boldsymbol{\delta}_2 = a(1/2, -\sqrt{3}/2)$, and $\boldsymbol{\delta}_3 = a(-1, 0)$, where a is the lattice constant.

In order to derive the band structure of the monolayer graphene, we do a Fourier transformation to Eq. (1.37), which gives rise to

$$H_{\mathrm{g}} = \sum_{\boldsymbol{q}} (a^{\dagger}_{\boldsymbol{q}}, b^{\dagger}_{\boldsymbol{q}}) H_{\mathrm{g}}(\boldsymbol{q}) \begin{pmatrix} a_{\boldsymbol{q}} \\ b_{\boldsymbol{q}} \end{pmatrix} , \tag{1.38}$$

where $a_{\boldsymbol{q}} = \sum_i a_i e^{i\boldsymbol{q}\cdot\boldsymbol{R}_{Ai}}$, $b_{\boldsymbol{q}} = \sum_i b_i e^{i\boldsymbol{q}\cdot\boldsymbol{R}_{Bi}}$, and

$$H_g(\boldsymbol{q}) = t \begin{pmatrix} 0 & \sum_{i=1}^{3} e^{i\boldsymbol{q}\cdot\boldsymbol{\delta}_i} \\ \sum_{i=1}^{3} e^{-i\boldsymbol{q}\cdot\boldsymbol{\delta}_i} & 0 \end{pmatrix} . \tag{1.39}$$

We are interested in the form of $H_g(\boldsymbol{q})$ at high symmetry points, as shown in Fig. 1.2(b). At the corner of the Brillouin zone, $K = (2\pi/(3a), 2\pi/(3\sqrt{3}a))$, $K' = (2\pi/(3a), -2\pi/(3\sqrt{3}a))$. The momentum dependent Hamiltonian matrix near the K point is given by

$$H_g^K(\boldsymbol{q}) = \frac{3ta}{2}\begin{pmatrix} 0 & k_+ \\ k_- & 0 \end{pmatrix} , \tag{1.40}$$

and the Hamiltonian matrix at the K' is simply given by the complex conjugate of Eq. (1.40). Notice that Eq. (1.40) takes the form of a Hamiltonian for a massless Dirac fermion, and the Dirac cone spectrum is shown in Fig. 1.2(a). This band dispersion is completely different from that of a free electron.

1.3.2 Transition Metal Dichalcogenides

Transition metal dichalcogenides are described by the chemical formula MX_2, where M=Mo or W is the transition metal atom and X=S or Se is the chalcogen atom. Transition metal dichalcogenides consist of strongly bonded X-M-X layers. Different X-M-X layers are weakly bonded by Van der Waals interactions, such that bulk MX_2 materials can be easily exfoliated into 2D samples. Taking MoS_2 as an example, the bulk MoS_2 has 2H stacking as shown in Fig. 1.3(a), and the structure of monolayer MoS_2 is shown in Fig 1.3(b).

Previous study shows that when MoS_2 is reduced from a bulk material to a monolayer flake, it evolves from a indirect gap insulator to a direct gap insulator [8]. The band edges of the conduction and valence bands of monolayer MoS_2 appear at the K and K' point in the momentum space [40], which is shown in Fig. 1.3(c). The atomic orbitals that contribute to the band edges states are mainly half filled d-orbitals of Mo atoms, which are further split by the discrete crystalline symmetry. Using the bases $|\phi_c\rangle = |d_{z^2}\rangle$ and $|\phi_v\rangle = |d_{x^2-y^2}\rangle + i\tau|d_{xy}\rangle$, where $\tau = \pm 1$ represents the K and K' valley, the effective Hamiltonian near the band edge is given by [2]

$$H_{TMD} = at(\tau k_x\sigma_x + k_y\sigma_y) + \frac{\Delta}{2}\sigma_z - \lambda\tau\frac{\sigma_z - 1}{2}\hat{s}_z . \tag{1.41}$$

Here σ_i and $\hat{s}_i$ are the Pauli matrices in the orbital space and the spin space, respectively. a and t are the lattice constant and the hopping energy in MoS_2. Note that Eq. (1.41) is very similar to the Hamiltonian of graphene (see Eq. (1.40)), while there are two significant differences. First, since the inversion symmetry is broken in monolayer MoS_2, the Δ term in Eq. (1.41) opens a large gap with an order of magnitude of eV at the band edge. Second, unlike graphene, the d-electrons in MoS_2 have a strong spin orbital coupling effect, which is given by the λ term in Eq. (1.41). This term splits the valence band while has tiny effect on the conduction band (see Fig. 1.3(c)). Notice that the spin is a good quantum number at the band edge.

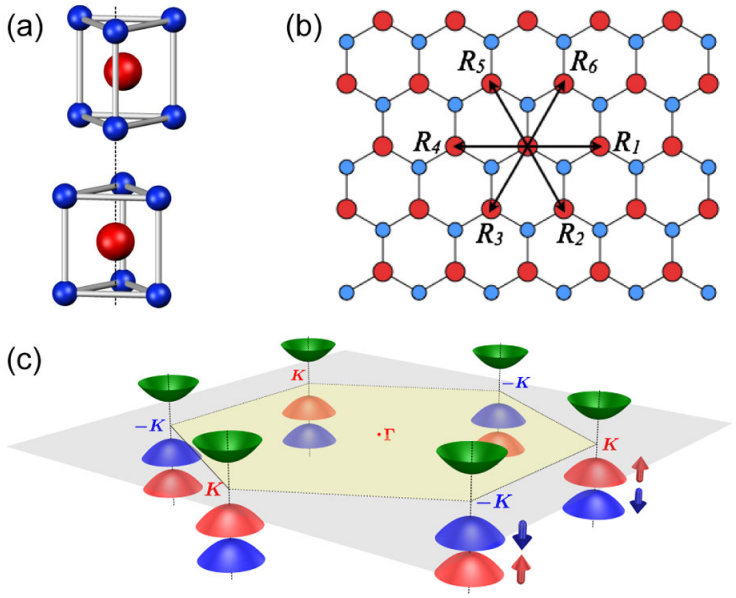

Figure 1.3: (a): The 2H stacking in bulk MoS_2. The red dots are Mo atoms, and the blue dots are S atoms, (b): The lattice structure of the monolayer MoS_2, (c): The band dispersion of MoS_2. This figure is from Ref. [2]

.

1.3.3 Ferromagnetic Insulators

In ferromagnetic insulators, the electrons are strongly bonded at their corresponding lattice sites, thus the charge degree of freedom is frozen. However, from a magnetically ordered ground state, spin excitations such as magnons can be generated by the spins that deviate from their ground state alignments. Spin excitations can generate spin and energy flux without causing Joule heating.

Previous studies have unveiled a series of novel properties of ferromagnetic insulators. The thermal Hall effect for magnons and spinons has been observed in a pyrochlore lattice [41, 42] and a frustrated quantum magnet system [43], respectively. Recently, giant magnon thermal Hall effects have been detected in high temperature superconductors [44, 45]. In all these systems, the thermal Hall effect is attributed to the 2D planes in the 3D systems, such as the Kagome plane in the pyrochlore lattice and the CuO_2 plane in the high temperature superconductor. Therefore, it is important to study the 2D ferromagnetic insulators. Although the Mermin-Wagner theorem prohibits magnetic orders in the isotropic 2D Heisenberg model [46], ferromagnetic states in 2D systems are realized in CrI_3 [10], where the out-of-plane anisotropy breaks the continuous rotation symmetry. Experimental study shows that the 2D CrI_3 systems demonstrate layer dependent magnetic phases [10], helical photo luminescence [47] and giant tunneling magnetoresistance [48].

1.4 Summary

In this chapter, we have introduced the Berry curvature, wave packet theory and some interesting 2D systems including graphene, transition metal dichalcogenides and ferromagnetic insulators. In the following chapters, we will show different ways to control the topological properties in 2D systems. In Chapter 2, we show that the twisting of graphene-hBN superlattice can tune the Chern number between zero and nonzero; in Chapter 3, we control the thermal Hall conductance in a magnon-phonon hybrid system using a gate voltage; in Chapter 4, we use a gate voltage to tune the winding number in a chiral fermion system, such that the lowest exciton state can be tuned between dark and bright.

Chapter 2

Engineering Topological Bands with Superlattices

2.1 Introduction

Nonzero Berry curvatures are ubiquitous, and can exist in common materials such as fcc Fe [49]. However, the nonzero Chern number, as an integral of the Berry curvature in the Brillouin zone, is much less common because of the cancellation of Berry curvatures from different parts of the Brillouin zone.

One way of engineering nonzero Chern numbers from nonzero Berry curvatures is using the superlattice structure. The superlattice produces large crystal cells in the real space. Thus the original Brillouin zone is folded into much smaller sublattice Brillouin zones and the energy bands are split into sub-bands. Since each sub-band only inherits the Berry curvature from the original system within a small area in the momentum space, the cancellation of Berry curvatures can be avoided. The superlattice structure provides a possibility to give rise to sub-bands with nonzero Chern numbers.

Song et al [3] predicted the appearance of topological nontrivial Chern bands using the graphene-hexagonal Boron Nitride (hBN) superlattice. In this chapter, we provide an explanation for the nonzero Chern number from a single band effective mass picture using different methods including the Foldy-Wouthuysen (FW) transformation, the band projection method and the wave packet theory. Using the single band effective mass model, the superlattice problem is reduced to solving the Bloch band structure of a free electron in a periodic potential. We show that the multi-band effect in the single band model is included in the correction terms from the Berry curvature and from the linear expansion term of the Hamiltonian around the wave packet center. The latter correction term has been ignored in previous studies, but we show that it must be included when the superlattice potential couples to pseudo-spin degree of freedoms.

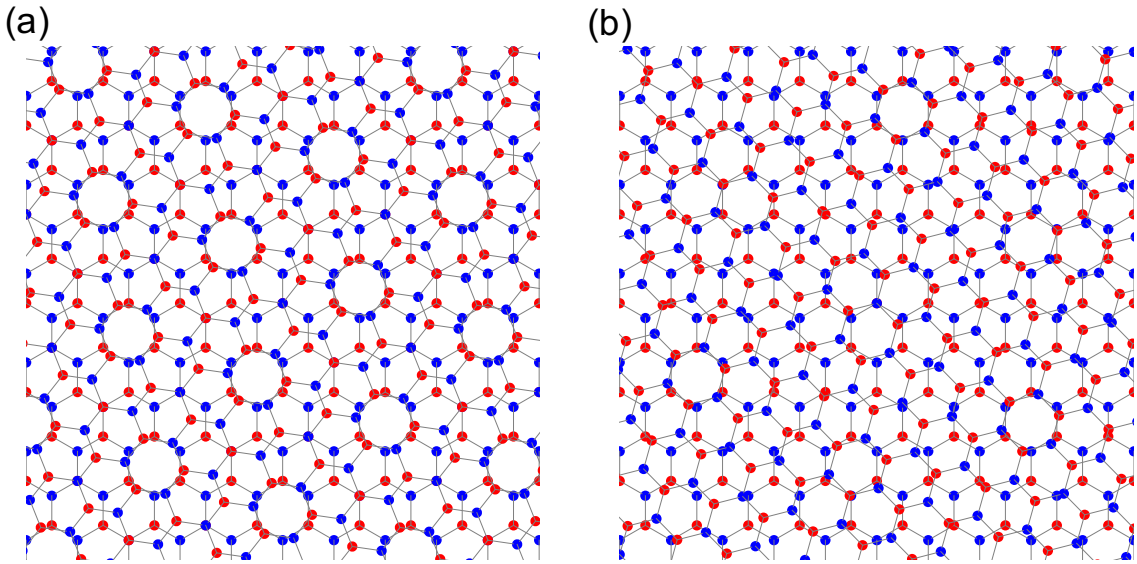

Figure 2.1: The Moiré pattern for the graphene-hBN superlattice with (a) commensurate, and (b) non-commensurate alignments. Different color represents different layers.

2.2 Graphene-hBN Superlattice

Since graphene and hBN have similar lattice constants, the Graphene-hBN heterostructure shows very long spacial periodicity with a small twisting angle. The size of the superlattice unit cell is in the order of 100 atomic distances [3]. When the rotation angle satisfies [50]

$$\cos\theta = \frac{3i^2 + 3i + 1/2}{3i^2 + 3i + 1} \,, \tag{2.1}$$

where i is a nonnegative integer, the Moiré pattern shows perfect periodicity. Fig. 2.1(a) shows the commensurate stacking where the two layers are rotated by $\arccos(13/14)$. More generally, the superlattice only has pseudo-periodicity which correspond to the non-commensurate stacking (see Fig. 2.1(b)). The graphene-hBN superlattice have attracted a lot of attentions [51, 52]. Recently, a gate tunable Mott insulator state was discovered in trilayer graphene-hBN superlattices [53].

The long periodicity simplifies the construction of the superlattice Hamiltonian. The hBN layer can be considered as an insulator with the valence band well below the Fermi level and the conduction band well above the Fermi level [54]. The electronic properties are dominated by the graphene layer, and the hBN layer provides a periodic potential with the same periodicity of the superlattice. A simple Hamiltonian has the following form

$$H_{\text{supper}} = H_0 + V(\boldsymbol{r}) \,, \tag{2.2}$$

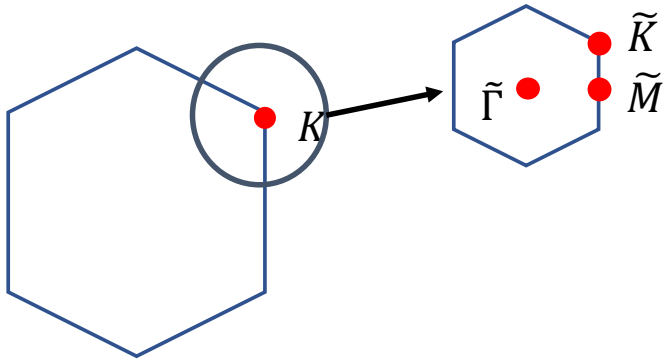

Figure 2.2: The superlattice Brillouin zone which comes from folding the original Brillouin zone near the K point.

where $H_0 = \alpha \boldsymbol{p} \cdot \boldsymbol{\sigma} + \Delta\sigma_z$ is the effective model for graphene near the K point with a global gap opened by the broken inversion symmetry of the superlattice, and the superlattice potential is given by

$$\begin{aligned} V(\boldsymbol{r}) &= V_0(\boldsymbol{r}) + V_z(\boldsymbol{r})\sigma_z \\ &= V_0 \sum_{i=1}^{3} \exp(i\boldsymbol{G_i} \cdot \boldsymbol{r}) + V_z\sigma_z \sum_{i=1}^{3} \exp(i\boldsymbol{G_i} \cdot \boldsymbol{r}) \ . \end{aligned} \tag{2.3}$$

$V(\boldsymbol{r})$ originates from the coupling of the two layers. Here $\boldsymbol{G}_i$ is the basis vector in the reciprocal space of the superlattice. Note that it is also possible to include σ_x and σ_y terms in $V(\boldsymbol{r})$, which originate from stains. Here we ignore these terms for simplicity.

The superlattice potential $V(\boldsymbol{r})$ folds the graphene Brillouin zone around the K point and forms a sub-Brillouin zone, as shown in Fig 2.2. At the boundary of the sub-Brillouin zone, gaps are opened by $V(\boldsymbol{r})$. The high symmetry points are denoted by $\tilde{\Gamma}$, $\tilde{K}$, etc.

2.2.1 Scalar Coupling

We first consider the simple case when the σ_z term is absent in $V(\boldsymbol{r})$. Instead of studying the motion of the Dirac particle described by the 2×2 Hamiltonian H_0 in the superlattice potential, we first reduce H_0 to a single band effective mass model, thus the superlattice problem becomes studying an electron with an effective mass in a Bloch system. The multi-band effect in H_0 is manifested by the correction of Berry curvature.

The wave packet equation of motion Eq. (1.33) and (1.34) show the similar role played by the Berry curvature and the magnetic field. Since the magnetic vector potential enters the Hamiltonian by the Peierls substitution $\boldsymbol{p} \to \boldsymbol{p} - e\boldsymbol{A}$, the Berry

connection of H_0, which is the "magnetic field" in the momentum space, can be introduced by the substitution $\boldsymbol{r} \to \boldsymbol{r} + \boldsymbol{\mathcal{A}}$ [38, 55]. In addition, since the superlattice Brillouin zone only includes a small region around the K point, we replace the Berry curvature $\Omega_z(\boldsymbol{p})$ by its value at the band edge, which is given by

$$\Omega_z = -sign(\Delta)\frac{\alpha^2}{2\Delta^2} \,. \tag{2.4}$$

Using the symmetric gauge, the substitution for the position operator is written as

$$\boldsymbol{r} \to \boldsymbol{r} + \boldsymbol{\Omega} \times \boldsymbol{p}/2 \,. \tag{2.5}$$

Now we use this substitution to derive an effective model for the graphene-hBN superlattice. Under a large band gap assumption where $\alpha|\boldsymbol{G}_i| \ll \Delta$, the Berry curvature is small according to Eq. (1.10). Therefore, we can apply the substitution Eq. (2.5) in $V_0(r)$ and apply a Taylor expansion to the first order of $\boldsymbol{\Omega}$:

$$\begin{aligned} V_0 &\to V_0(\boldsymbol{r} + \frac{1}{2}\boldsymbol{\Omega} \times \boldsymbol{p}) \\ &= V_0(r) - \frac{1}{2}\boldsymbol{\Omega} \cdot (\nabla V_0(r) \times \boldsymbol{p}) \,. \end{aligned} \tag{2.6}$$

Now we construct the effective model of the conduction band electron in H_0. The effective Hamiltonian is given by the effective mass approximation of the conduction band electron and the substitution in Eq. (2.5), which is written as

$$H = \frac{p^2}{2m} + V(\boldsymbol{r}) - \frac{1}{2}\boldsymbol{\Omega} \cdot (\nabla V_0(\boldsymbol{r}) \times \boldsymbol{p}) \,. \tag{2.7}$$

Here $m = |\Delta|/\alpha^2$ is the effective mass of the conduction band electron of H_0. The last term in the effective model shows that the Berry curvature describes the price paid when reducing the multiple band theory to a single band theory.

We solve the eigenvalue and the Berry curvature of the effective model using the central equation. The idea of central equation is using the plane wave basis to expand the Bloch Hamiltonian [56]. For a general Bloch system, the Hamiltonian is given by

$$H_{\text{Bloch}} = \frac{p^2}{2m} + \tilde{V}(\boldsymbol{r}) \,, \tag{2.8}$$

where the periodic potential can be expanded by a Fourier series using the reciprocal lattice points $\boldsymbol{G}_i$ as $\tilde{V}(\boldsymbol{r}) = \sum_i \tilde{V}_{\boldsymbol{G}_i} \exp(i\boldsymbol{G}_i \cdot \boldsymbol{r})$. We also expand the eigenfunctions using the Fourier series $\psi(\boldsymbol{r}) = \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \exp(i\boldsymbol{k}_i \cdot \boldsymbol{r})$. Therefore, we achieve an eigenvalue problem known as the central equation

$$\frac{k^2}{2m}c_{\boldsymbol{k}} + \sum_i \tilde{V}_{\boldsymbol{G}_i} c_{\boldsymbol{k}-\boldsymbol{G}_i} = E c_{\boldsymbol{k}} \,. \tag{2.9}$$

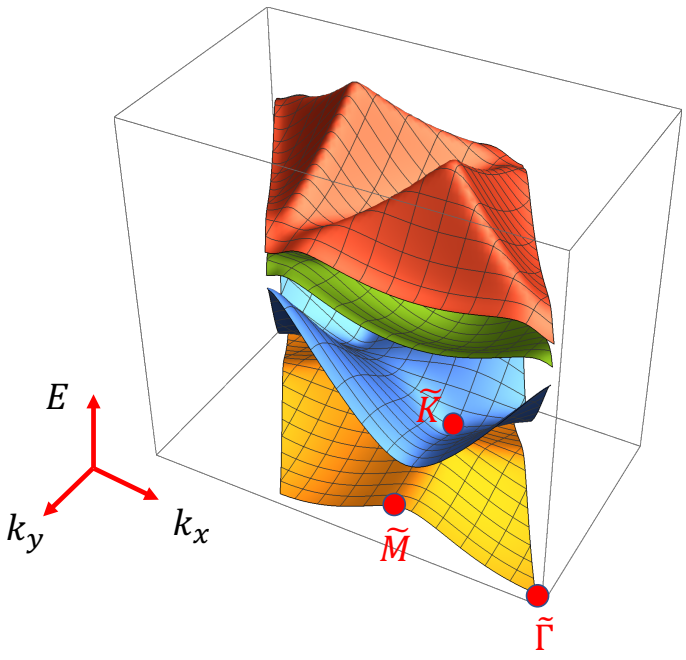

Figure 2.3: The energy spectrum of the superlattice Hamiltonian using the single band approximation. We used the parameters form Ref [3].

Note that different wave vectors $\boldsymbol{k}$ in the first Brillouin zone are isolated in the central equation. For each $\boldsymbol{k}$ value, the central equation provides a matrix with basis $c_{\boldsymbol{k}+\boldsymbol{G}_i}$, and the spectrum is given by the eigenvalue of this matrix. The dimension of this matrix dependents on the number of reciprocal vectors $\boldsymbol{G}_i$, which can be restricted to a finite number using a cut-off in numerical calculations. Using 100 reciprocal lattice points, we get the superlattice spectrum shown in Fig 2.3.

The phase diagram in Fig. 2.4(b) shows Chern numbers with different signs of parameters. In order to show the difference between the trivial and nontrivial phase, Berry curvatures are plotted in Fig 2.5. For the topological nontrivial case, the majority contribution of Berry curvatures comes from the $\tilde{K}$ point, where the band gap is split by the superlattice potential, while for the trivial case, the Berry curvatures from different momentums cancel each other.

When $V(\boldsymbol{r})$ is coupled with σ_z, the method show above should be modified. The following sections will derive the effective single band model using the FW transformation, projection onto conduction band states, and wave packet theory.

2.2.2 FW Transformation

Under the large band gap limit where $\alpha|\boldsymbol{G}_i| \ll \Delta$ and $V(\boldsymbol{r}) \ll \Delta$, the FW transformation [57] uses a series of unitary transformations on Hamiltonian (2.2) and gives rise to an approximately diagonal Hamiltonian. The off-diagonal terms after performing

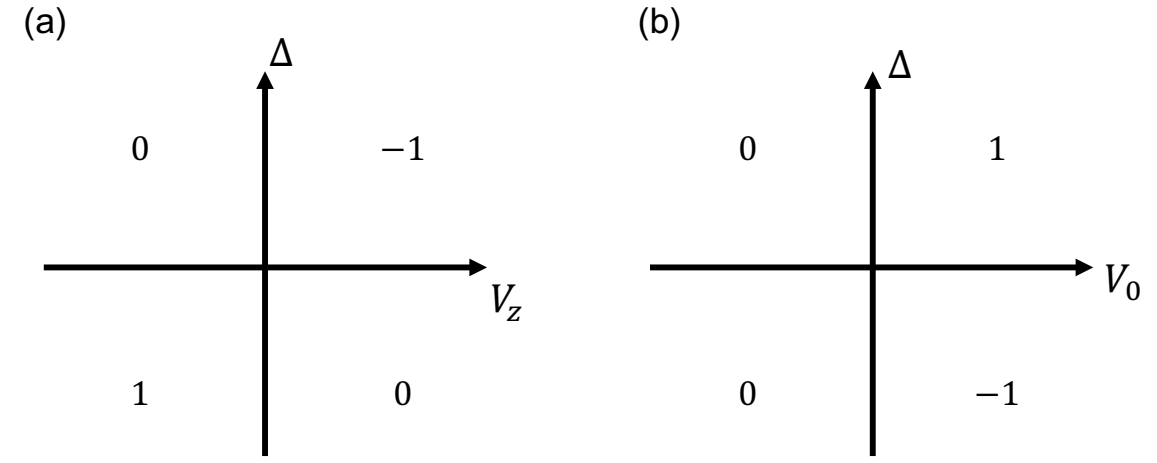

Figure 2.4: Phase diagram when V(r) is coupled with: (a) σ_3 and with (b) I_2. The Chern numbers are given for different parameters.

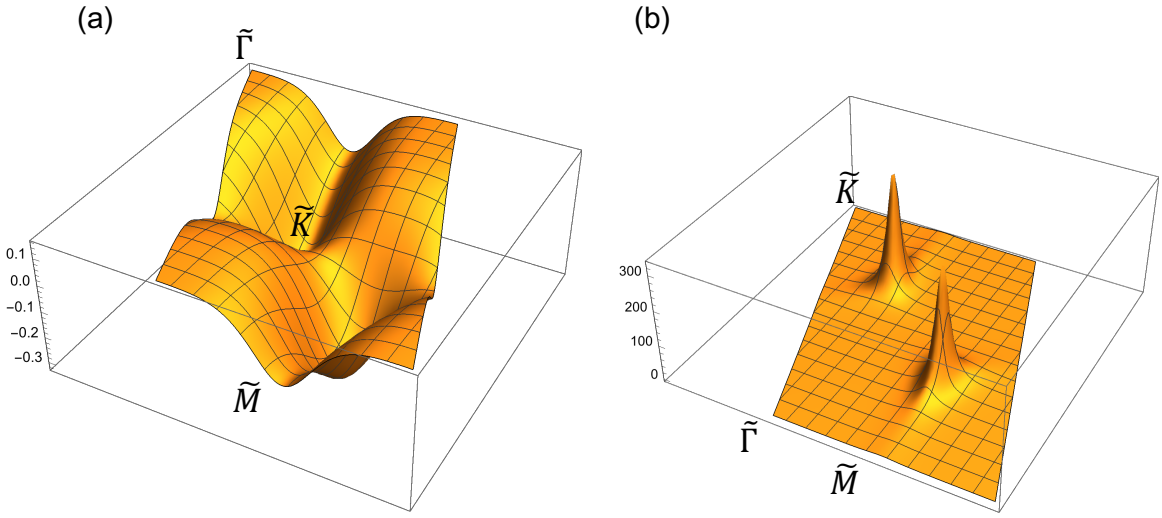

Figure 2.5: Berry curvature distribution for topological trivial (a) and non-trivial case (b) when $V(\boldsymbol{r})$ is a scaler potential.

the FW transformation are in high order of $\alpha|p|/\Delta$ and $V(\boldsymbol{r})/\Delta$.

We start from the superlattice Hamiltonian Eq. (2.2) and set $V_0 = 0$ for simplicity. The Hamiltonian becomes:

$$H = \alpha \boldsymbol{p} \cdot \boldsymbol{\sigma} + \Delta\sigma_z + V_z(\boldsymbol{r})\sigma_z \ . \tag{2.10}$$

Under a unitary transformation defined by a Hermitian matrix S, the Hamiltonian is

transformed into:

$$\begin{aligned} H' &= e^{iS} H e^{-iS} \\ &= H + i[S,H] + \frac{i^2}{2!}[S,[S,H]] + \frac{i^3}{3!}[S,S[S,H]]] + \dots . \end{aligned} \tag{2.11}$$

The off-diagonal perturbations in the Hamiltonian are called odd terms, and the diagonal perturbations are called even terms. We represent the odd terms by $\hat{P}$ and even terms by $\hat{V}$. For now, $\hat{P} = \alpha \boldsymbol{p} \cdot \boldsymbol{\sigma}$ and $\hat{V} = V_z(\boldsymbol{r})\sigma_z$. We find the unitary transformations defined by S to eliminate the odd terms $\hat{P}$ up to the order of $1/\Delta^3$.

First, we set $S = -i\sigma_z \hat{P}/2\Delta$. The commutators in the expansion Eq. (2.11) are written as

$$i[S,H] = -\hat{P} + \frac{1}{\Delta}\sigma_z \hat{P}^2 + \frac{\sigma_z}{2\Delta}[\hat{P}, \hat{V}(r)] , \tag{2.12}$$

$$\frac{i^2}{2!}[S[S,H]] = -\frac{1}{2\Delta}\sigma_z \hat{P}^2 - \frac{1}{2\Delta^2}\hat{P}^3 - \frac{1}{8\Delta^2}[\hat{P},[\hat{P},\hat{V}(r)]] , \tag{2.13}$$

$$\frac{i^3}{3!}[S,S[S,H]]] = \frac{1}{6\Delta^2}\hat{P}^3 + o(\frac{1}{\Delta^3}) . \tag{2.14}$$

Therefore, the new Hamiltonian after the transformation is

$$\begin{aligned} H' &= \Delta\sigma_z + (\frac{\sigma_z}{2\Delta}[\hat{P},\hat{V}] - \frac{1}{3\Delta^2}\hat{P}^3) + (\hat{V} + \frac{1}{2\Delta}\sigma_z\hat{P}^2 - \frac{1}{8\Delta^2}[\hat{P},[\hat{P},\hat{V}]]) \\ &= \Delta\sigma_z + \hat{P}' + \hat{V}' . \end{aligned} \tag{2.15}$$

Now new odd terms become $\hat{P}'$ and new even terms become $\hat{V}'$, where $\hat{P}'$ now has an order of $1/\Delta$. Similarly, we can do two consecutive unitary transformations with the form of $S' = -i\sigma_z\hat{P}'/2\Delta$ and $S'' = -i\sigma_z\hat{P}''/2\Delta$, where $\hat{P}''$ is the odd term after the S' transformation and is in the order of $1/\Delta^2$. After these two transformations, the odd terms are finally in the order of $1/\Delta^3$. Ignoring the $\mathcal{O}(1/\Delta^3)$ terms, the approximately diagonal Hamiltonian is given by

$$\begin{aligned} H''' =& \Delta\sigma_z + \frac{\alpha^2}{2\Delta}\sigma_z p^2 + V_z(\boldsymbol{r})\sigma_z \\ &- \frac{\alpha^2}{8\Delta^2}(-\nabla^2 V_z(\boldsymbol{r})\sigma_z - 4i\nabla V_z(\boldsymbol{r})\cdot\boldsymbol{p}\sigma_z + 4V_z(\boldsymbol{r})p^2\sigma_z + 2(\nabla V_z(\boldsymbol{r})\times\boldsymbol{p})_z) . \end{aligned} \tag{2.16}$$

The two diagonal elements represent the energies of conduction band and valence band electrons, respectively. In order to study the electronic excitations, we focus on the conduction band which is described by the effective Hamiltonian

$$\begin{aligned} H_{cond} =& \frac{\alpha^2}{2|\Delta|}p^2 + V_z(\boldsymbol{r}) \\ &- \frac{\alpha^2}{8\Delta^2}(-\nabla^2 V_z(\boldsymbol{r}) - 4i\nabla V_z(\boldsymbol{r})\cdot\boldsymbol{p} + 4V_z(\boldsymbol{r})p^2 + 2(\nabla V_z(\boldsymbol{r})\times\boldsymbol{p})_z) . \end{aligned} \tag{2.17}$$

The first line describs a free electron in a periodic potential, where the effective mass is given by $m = |\Delta|/\alpha^2$. The second line originates from the down-folding process of the FW transformation. It represents the residual effect of the other band when we build the single band effective mass model that only includes the conduction band.

We can also write the effective Hamiltonian Eq. (2.17) in terms of the effective mass m and the Berry curvature Ω_z in Eq. (2.4). The form of the effective Hamiltonian depends on the sign of gap Δ. For $\Delta > 0$

$$H_{cond} = \frac{p^2}{2m} - i\Omega_z(\nabla V_z(\boldsymbol{r}) \cdot \boldsymbol{p}) + \Omega_z V_z(\boldsymbol{r})p^2 + \frac{1}{2}\boldsymbol{\Omega} \cdot (\nabla V_z(\boldsymbol{r}) \times \boldsymbol{p}) + V_z(\boldsymbol{r}) \,. \quad (2.18)$$

Here the $\nabla^2 V$ term has been thrown away by assuming that $V(\boldsymbol{r})$ changes smoothly, which is true for a superlattice potential [3]. For $\Delta < 0$

$$H_{cond} = \frac{p^2}{2m} - i\Omega_z(\nabla V_z(\boldsymbol{r}) \cdot \boldsymbol{p}) + \Omega_z V_z(\boldsymbol{r})p^2 - \frac{1}{2}\boldsymbol{\Omega} \cdot (\nabla V_z(\boldsymbol{r}) \times \boldsymbol{p}) - V_z(\boldsymbol{r}) \,. \quad (2.19)$$

Note that the format of the effective Hamiltonian for conduction band electrons when $V(\boldsymbol{r})$ is a scalar does not depend on the sign of Δ (see Eq. (2.7)).

We calculate the Berry curvature and Chern number using the effective Hamiltonian Eq. (2.18) and (2.19). The phase diagram is shown in Fig 2.4(b). Using the central equation method (see Eq. (2.9)), the Chern number is nonzero when V_z and Δ share the same sign, and the sign of the Chern number is controlled by the sign of V_z and Δ, which agrees with the result from Song's paper [3] using the full Hamiltonian. Therefore, the effective single band model is good enough to study the topological phase transition of the superlattice system. Note that V_z and Δ have the same sign for commensurate stackings and have different signs for incommensurate stackings [3].

Now we can see the difference between $V_0(\boldsymbol{r})$ and $V_z(\boldsymbol{r})\sigma_z$. If the superlattice potential only contains $V_0(\boldsymbol{r})$, the effective Hamiltonians (written in terms of the effective mass m and the Berry curvature Ω_z) in Eq. (2.7) for both $\Delta > 0$ and $\Delta < 0$ have the same form. Therefore, changing the sign of Δ will flip the sign of Ω_z, which gives rise to the time reversal counterpart, thus the sign of Chern number is changed. In addition, different signs of V_0 can be proved to change the Chern number back and forth between zero and nonzero by changing the sign of the gap opened at the $\tilde{M}$ point. However, for the case of $V(r)\sigma_z$ coupling, changing the sign of Δ will change the form of the conduction band Hamiltonian in Eq. (2.18) and (2.19), by effectively changing the sign of V_z and Ω_z simutanously. According to the result for the $V_0(\boldsymbol{r})$ case, changing the sign of $V(r)\sigma_z$ will change Chern number from nonzero (zero) to zero (nonzero).

2.2.3 Projection on the Conduction Band

We can also derive a single band theory from Hamiltonian with σ_z coupling in Eq. (2.10) by spanning it with the states only on the conduction band of H_0. The

conduction band states of H_0 have the following form:

$$|\Psi\rangle = \sum_{\boldsymbol{k}} f(\boldsymbol{k})|\psi_c(\boldsymbol{k})\rangle \ , \tag{2.20}$$

where $f(\boldsymbol{k})$ is the envelope function and $|\psi_c(\boldsymbol{k})\rangle$ is the eigenstate of the conduction band. Suppose $\Delta > 0$, the conduction band eigenstate is given by

$$|\psi_c(\boldsymbol{k})\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \begin{pmatrix} \cos(\theta_{\boldsymbol{k}}/2) \\ \sin(\theta_{\boldsymbol{k}}/2)e^{i\phi_{\boldsymbol{k}}} \end{pmatrix} \ , \tag{2.21}$$

and the spherical angles are given by

$$\cos(\theta_{\boldsymbol{k}}) = \frac{\Delta}{\sqrt{\alpha^2 k^2 + \Delta^2}} \tag{2.22}$$

and

$$\tan\phi_{\boldsymbol{k}} = \frac{k_y}{k_x} \ . \tag{2.23}$$

Note that the gauge choice in Eq. (2.21) has no singularity at $\boldsymbol{k} = 0$ when $\Delta > 0$. Using the Schrodinger equation $H|\Psi\rangle = E|\Psi\rangle$, and multiplying $\langle\psi_{c\boldsymbol{k}}|$ on the left, we can get its matrix form:

$$\sqrt{\alpha^2 k^2 + \Delta^2} f_{\boldsymbol{k}} + V_{\boldsymbol{k}-\boldsymbol{k}'}\langle u_{c\boldsymbol{k}}|\sigma_z|u_{c\boldsymbol{k}'}\rangle f_{\boldsymbol{k}'} = E f_{\boldsymbol{k}} \ . \tag{2.24}$$

Where $V_{\boldsymbol{k}-\boldsymbol{k}'}$ is the Fourier component of $V_z(\boldsymbol{r})$ given by $V_{\boldsymbol{k}} = \int dr^2 V_z(\boldsymbol{r}) \exp(i\boldsymbol{k}\cdot\boldsymbol{r})$. The additional σ_z in $\langle u_{c\boldsymbol{k}}|\sigma_z|u_{c\boldsymbol{k}'}\rangle$ makes this model different from the scaler coupling case where the corresponding inner product $\langle u_{c\boldsymbol{k}}|u_{c\boldsymbol{k}'}\rangle \approx 1 + \langle u_{c\boldsymbol{k}}|\partial_{\boldsymbol{k}}|u_{c\boldsymbol{k}}\rangle \cdot (\boldsymbol{k}' - \boldsymbol{k})$ simply gives rise to a term related to the Berry connection. With the form of the wave function (2.21), and using the large band gap limit $\alpha^2 k^2 \ll \Delta^2$, for $\Delta > 0$, we have:

$$\begin{aligned}
\langle u_{c\boldsymbol{k}}|\sigma_z|u_{c\boldsymbol{k}'}\rangle &= \cos(\theta_{\boldsymbol{k}}/2)\cos(\theta_{\boldsymbol{k}'}/2) - \sin(\theta_{\boldsymbol{k}}/2)\sin(\theta_{\boldsymbol{k}'}/2)e^{i(\phi_{\boldsymbol{k}'}-\phi_{\boldsymbol{k}})} \\
&= \cos(\theta_{\boldsymbol{k}}/2)\cos(\theta_{\boldsymbol{k}}/2 + \Delta\theta/2) - \sin(\theta_{\boldsymbol{k}}/2)\sin(\theta_{\boldsymbol{k}}/2 + \Delta\theta/2)e^{i\Delta\phi} \\
&\approx \cos\theta_{\boldsymbol{k}} - \sin\theta_{\boldsymbol{k}}\frac{\Delta\theta}{2} - i\Delta\phi\frac{1-\cos\theta_{\boldsymbol{k}}}{2} \\
&= 1 - \frac{\alpha^2 k^2}{2\Delta^2} - \alpha^2\frac{k_x\Delta k_x + k_y\Delta k_y}{2\Delta^2} - \alpha^2\frac{i}{4\Delta^2}(-k_y\Delta k_x + k_x\Delta k_y) \ .
\end{aligned} \tag{2.25}$$

Here $\Delta\theta = \theta_{\boldsymbol{k}'} - \theta_{\boldsymbol{k}}$ and $\Delta\phi = \phi_{\boldsymbol{k}'} - \phi_{\boldsymbol{k}}$. In the last line, we used the differential relation

$$\Delta\theta = \frac{k_x\Delta k_x + k_y\Delta k_y}{\alpha^2 k^2} \ , \tag{2.26}$$

and

$$\Delta\phi = \frac{-k_y\Delta k_x + k_x\Delta k_y}{\alpha^2 k^2} \ . \tag{2.27}$$

where $\Delta \boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k}$.

The current gauge choice of the wave function in Eq. (2.21) has a singularity when $\Delta < 0$. Therefore, when Δ is negative, we need to choose a new gauge by multiplying Eq. (2.21) by $e^{-i\theta_{\boldsymbol{k}}}$. This new gauge choice will lead to the following result:

$$\begin{aligned}
\langle u_{c\boldsymbol{k}}|\sigma_z|u_{c\boldsymbol{k}'}\rangle &= \cos(\theta_{\boldsymbol{k}}/2)\cos(\theta_{\boldsymbol{k}'}/2)e^{-i(\phi_{\boldsymbol{k}'}-\phi_{\boldsymbol{k}})} - \sin(\theta_{\boldsymbol{k}}/2)\sin(\theta_{\boldsymbol{k}'}/2) \\
&= \cos(\theta_{\boldsymbol{k}}/2)\cos(\theta_{\boldsymbol{k}}/2+\Delta\theta/2)e^{-i\Delta\phi} - \sin(\theta_{\boldsymbol{k}}/2)\sin(\theta_{\boldsymbol{k}}/2+\Delta\theta/2) \\
&\approx \cos\theta_{\boldsymbol{k}} - \sin\theta_{\boldsymbol{k}}\frac{\Delta\theta}{2} - i\Delta\phi\frac{1+\cos\theta_{\boldsymbol{k}}}{2} \\
&= -(1-\frac{\alpha^2k^2}{2\Delta^2} - \alpha^2\frac{k_x\Delta k_x + k_y\Delta k_y}{2\Delta^2}) - \alpha^2\frac{i}{4\Delta^2}(-k_y\Delta k_x + k_x\Delta k_y)\ .
\end{aligned} \tag{2.28}$$

Substituting Eq. (2.25) and (2.28) back to equation (2.24) and using the Fourier transformation, the same result as equation (2.18) and (2.19) can be reached.

2.2.4 From Wave Packet Theory

Consider a wave packet localized near $\boldsymbol{k}_c$ in the momentum space and $\boldsymbol{r}_c$ in the real space. The wave packet wave function is written as

$$|W\rangle = \int dk^2 w(\boldsymbol{k})|\psi_{c\boldsymbol{k}}\rangle\ , \tag{2.29}$$

and the center of mass position and momentum is give by

$$\boldsymbol{r_c} = \langle W|\boldsymbol{r}|W\rangle\ , \tag{2.30}$$

$$\boldsymbol{k_c} = \langle W|\boldsymbol{k}|W\rangle \tag{2.31}$$

Where $|\psi_{c\boldsymbol{k}}\rangle = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})|u_{c\boldsymbol{k}}\rangle$ is the Bloch wave function of the conduction band of H_0. If the superlattice potential $V_z(\boldsymbol{r})$ changes smoothly within the length scale of the wave packet, the local Hamiltonian felt by the wave packet is approximated by

$$H_c(\boldsymbol{r}) = \alpha\boldsymbol{p}\cdot\boldsymbol{\sigma} + (\Delta + V_z(\boldsymbol{r}_c) + \nabla V_z(\boldsymbol{r}_c)\cdot(\boldsymbol{r}-\boldsymbol{r}_c))\sigma_z\ . \tag{2.32}$$

From the wave packet theory (see Eq. (1.24)), the effective Hamiltonian of this conduction band wave packet is written as

$$\langle W|H_c|W\rangle = E_c - Im[\langle\frac{\partial u_{c\boldsymbol{k}_c}}{\partial\boldsymbol{r}_c}|\cdot(E_c - H_c(\boldsymbol{r}_c))|\frac{\partial u_{c\boldsymbol{k}_c}}{\partial\boldsymbol{k}_c}\rangle]\ , \tag{2.33}$$

where $E_c = \langle W|H_c(\boldsymbol{r}_c)|W\rangle = \sqrt{\alpha^2k_c^2 + (\Delta + V(\boldsymbol{r}_c))^2}$ is the conduction band energy of the local Hamiltonian $H_c(\boldsymbol{r}_c)$. By assuming $\Delta \gg \alpha|\boldsymbol{G}_i|$ and $\Delta \gg V(\boldsymbol{r}_c)$, which

also implies $sign(\Delta) = sign(\Delta + V(\boldsymbol{r}_c))$, we get the follow approximate expression for E_c

$$\begin{aligned} E_c &= |\Delta + V(\boldsymbol{r}_c)|\sqrt{1 + \frac{k_c^2}{(\Delta + V(\boldsymbol{r}_c))^2}} \\ &\approx |\Delta + V(\boldsymbol{r}_c)|(1 + \frac{k_c^2}{2(\Delta + V(\boldsymbol{r}_c))^2}) \\ &= |\Delta| + sign(\Delta)V(\boldsymbol{r}_c) + sign(\Delta)\frac{k_c^2}{2(\Delta + V(\boldsymbol{r}_c))} \\ &\approx |\Delta| + sign(\Delta)V(\boldsymbol{r}_c) + \frac{k_c^2}{2|\Delta|} - sign(\Delta)\frac{k_c^2 V(\boldsymbol{r}_c)}{2\Delta^2} \end{aligned} \tag{2.34}$$

We calculate the second term in Eq. (2.33) which comes from expanding the Hamiltonian around the wave packet center. Here $|u_{c\boldsymbol{k}}\rangle$ has the same form as the spinor part of expression (2.21), while we need to include $V_z(\boldsymbol{r})\sigma_z$ as a spatially varying part of Δ, thus there is $\boldsymbol{r}_c$ dependence of $\theta_{\boldsymbol{k}}$, which is given by

$$\cos\theta_{\boldsymbol{k}} = \frac{\Delta + V(\boldsymbol{r}_c)}{\sqrt{\alpha^2 k^2 + (\Delta + V(\boldsymbol{r}_c))^2}} .$$

After simple calculations, the energy correction term of Eq. (2.33) can be simplified to:

$$-Im[\langle\frac{\partial u_{c\boldsymbol{k}_c}}{\partial \boldsymbol{r}_c}| \cdot (E_c - H_c(\boldsymbol{r}_c))|\frac{\partial u_{c\boldsymbol{k}_c}}{\partial \boldsymbol{k}_c}\rangle] = -\frac{\alpha^2}{2\Delta^2}(\nabla V \times \boldsymbol{k}_c)_z . \tag{2.35}$$

Note that the second term in Eq. (2.33) has no contribution in the $V_0(\boldsymbol{r})$ case while it is nonzero when $V(\boldsymbol{r})$ involves the σ_z coupling. We point out that the finite size of the wave packet is important in the σ_z coupling case.

So far, everything is written in terms of the position and momentum of the wave packet center which are numbers. In order to convert $\boldsymbol{r}_c$ and $\boldsymbol{k}_c$ to operators $\boldsymbol{r}$ and $\boldsymbol{p}$, two steps are required. First, we need to implement the Weyl ordering [58], which requires that the $\boldsymbol{r}$ and $\boldsymbol{p}$ operators must be symmetric. Second, we need to convert the mechanical position $\boldsymbol{r}_c$ to the canonical position, which is given by the substitution $\boldsymbol{r}_c = \boldsymbol{r} - \frac{1}{2}\boldsymbol{\Omega} \times \boldsymbol{k_c}$. Therefore, the effective Hamiltonian from $\langle W|H|W\rangle$ in Eq. (2.33) is given by:

$$\begin{aligned} &H_{cond} \\ &= sign(\Delta)V(\boldsymbol{r}_c) + \frac{p^2}{2|\Delta|} - \frac{sign(\Delta)}{8\Delta^2}(-\nabla^2 V(\boldsymbol{r}_c) - 4i\nabla V(\boldsymbol{r}_c)\cdot\boldsymbol{p} + 4V(\boldsymbol{r}_c)p^2) - \frac{\alpha^2}{2\Delta^2}(\nabla V \times \boldsymbol{p})_z \\ &\overset{\mathbf{r_c}\to\mathbf{r}-\frac{1}{2}\mathbf{\Omega}\times\mathbf{k_c}}{=\!=} sign(\Delta)V(\boldsymbol{r}) - \frac{\alpha^2}{4\Delta^2}(\nabla V \times \boldsymbol{p})_z + \frac{p^2}{2|\Delta|} - \frac{sign(\Delta)}{2\Delta^2}(-i\nabla V(\boldsymbol{r})\cdot p + V(\boldsymbol{r})p^2) . \end{aligned} \tag{2.36}$$

Note that the second derivative terms of $V(\boldsymbol{r})$ in the last line are ignored. Therefore, the Hamiltonian from the wave packet theory has the same form of the effective Hamiltonian in Eq. (2.18) and (2.19) from FW transformation.

2.3 Summary

In this chapter, we derived a single band effective mass theory to study the band topology of a Dirac electron from graphene in a superlattice potential. We show that the effective model for the superlattice potential with inner degrees of freedom (such as σ_z coupling) is different from that of a scalar external potential. The effective model is derived using the FW transformation, the band projection method and the wave packet theory. We show that in addition to the substitution $\boldsymbol{r} \to \boldsymbol{r} - \frac{1}{2}\boldsymbol{\Omega} \times \boldsymbol{k_c}$ to the local Hamiltonian of the wave packet center, we also need to consider the linear expansion of the Hamiltonian around the wave packet center which cannot be ignored when the coupling involves σ_z terms. The effective theory leads to the correct Chern numbers for different phases calculated by the complete theory.

Chapter 3

Thermal Hall Effect Induced by Magnon-Phonon Interactions

3.1 Introduction

The spin-lattice interaction in solids is responsible for a wide spectrum of cross-correlated phenomena. A well-known example is the coupling between dielectric and magnetic order in multiferroics [59–61]. It can also manifest in the dynamics of elementary excitations such as magnons and phonons, in the form of magnon-phonon interaction. For example, it has been demonstrated that magnons that couple to optical phonons can be launched by an electric field [62, 63], paving the way to the electric generation of magnon spin current [64]. On the other hand, the dynamics of phonons can be modified by the magnon-phonon interaction as well, as in the case of nonreciprocal sound propagations observed in Cu_2OSeO_3 with an applied magnetic field [65].

Another important scenario in which the magnon-phonon interaction is expected to play a significant role is the thermal Hall effect. In a magnetic insulator, the heat current can be carried by either magnons or phonons. Thus the thermal Hall effect can be used as an effective probe of these charge-neutral excitations. Indeed, thermal Hall effects attributed to magnons [41–43] and phonons [66–69] have been reported. Theoretical explanations have so far assumed that the low-energy excitations can be described by independent magnons or phonons [70–76]. However, if the magnon-phonon interaction is strong, considering the magnon-phonon hybrid as a whole is more appropriate. Recently, Takahashi and Nagaosa have studied the magnon-phonon interaction arising from long-range dipolar couplings [77]; Park and Young [78] have studied the magnon-phonon interaction from the antiferromagnetic exchange interaction in non-collinear magnetic systens. However, the consequence of the magnon-phonon interaction in collinear magnetic systems with short-range couplings (such as symmetric or antisymmetric exchange) on the thermal Hall effect is yet to be explored.

In this chapter, we investigate the effect of the magnon-phonon interaction on the thermal Hall effect. Using symmetry arguments, we show that the magnon-phonon interaction can induce a thermal Hall effect whenever the mirror symmetry in the direction of the magnetization is broken. In the limit of strong magnetic anisotropy this effect can be understood as a phonon Hall effect, driven by an effective magnetic field in the phonon sector introduced by the magnon-phonon interaction. In the more general case where the magnons and phonons are close in energy, we have developed a theory to treat both excitations on an equal footing. We demonstrate our theory in a collinear ferromagnet on a square lattice, with perpendicular easy-axis anisotropy and Dzyaloshinskii-Moriya (DM) interaction due to mirror symmetry breaking (Fig. 3.1). In this model, the thermal Hall effect is entirely due to the magnon-phonon interaction. The thermal Hall conductivity is estimated using realistic material parameters. Our result sheds new light on the dynamical aspect of the spin-lattice interaction, and may find applications in the emerging field of spin caloritronics [79].

3.2 Symmetry Consideration

We begin our discussion by analyzing the symmetry of a magnon-phonon coupled system. Consider a two-dimensional (2D) spin system described by the Hamiltonian

$$H_{\rm s} = -J\sum_{\langle i,j\rangle} \boldsymbol{s}_i \cdot \boldsymbol{s}_j - \frac{K}{2}\sum_i s_{iz}^2 + \sum_{\langle i,j\rangle} \boldsymbol{D}_{ij}\cdot(\boldsymbol{s}_i \times \boldsymbol{s}_j)\ , \tag{3.1}$$

where $J > 0$ represents the nearest-neighbor ferromagnetic exchange, and $K > 0$ is the perpendicular easy-axis anisotropy. The third term describes the DM interaction due to the out-of-plane mirror-symmetry breaking. Here $\boldsymbol{D}_{ij} = D\hat{\boldsymbol{R}}_{ij}\times\hat{z}$ with D being the strength of the DM interaction, and $\hat{\boldsymbol{R}}_{ij} \equiv (\boldsymbol{R}_i - \boldsymbol{R}_j)/(|\boldsymbol{R}_i - \boldsymbol{R}_j|)$ is the bond direction from site j to site i. The direction of $\boldsymbol{D}_{ij}$ is in-plane and perpendicular to the bond direction, as shown in Fig. 3.1(b). We restrict our discussion to $D < \sqrt{JK}/2$ such that the ground state remains a collinear ferromagnet.

The spin-wave Hamiltonian can be obtained by expanding the spin operator in Eq. (3.1) around its ground state expectation value, i.e., $\delta\boldsymbol{s}_i = \boldsymbol{s}_i - S\hat{z}$. To the lowest order, the spin-wave Hamiltonian reads

$$H_{\rm sw} = -J\sum_{\langle i,j\rangle} \delta\boldsymbol{s}_{i\perp}\cdot\delta\boldsymbol{s}_{j\perp} - (J\zeta + K)S\sum_i \delta s_{iz}\ , \tag{3.2}$$

where ζ is the coordination number. Note that the DM interaction is absent. This can be seen by expanding the DM interaction,

$$H_{\rm DMI} = DS\sum_{\langle i,j\rangle} \hat{\boldsymbol{R}}_{ij}\cdot(\delta\boldsymbol{s}_i - \delta\boldsymbol{s}_j) + \mathcal{O}(\delta s^3)\ . \tag{3.3}$$

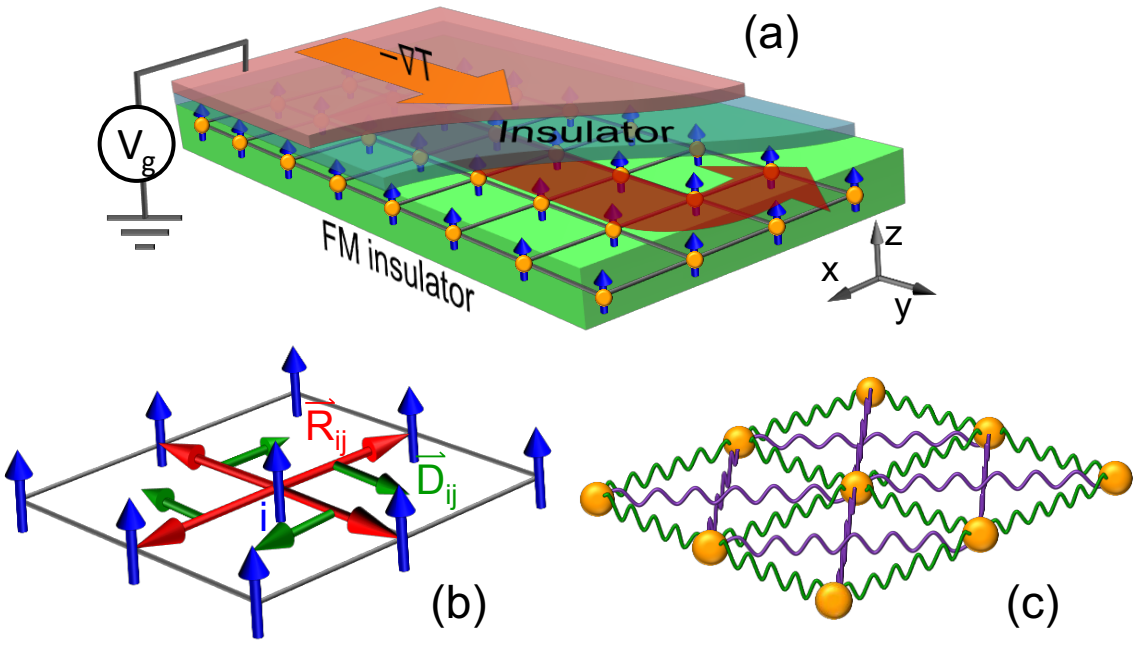

Figure 3.1: (a) The setup illustrates the thermal hall effect of the hybrid magnon-phonon system. In this example, the out-of-plane mirror symmetry is broken by an applied gate voltage. (b) For the spin system, the ferromagnetic Heisenberg exchange interaction and the anisotropy develop a collinear ferromagnetic state with an out-of-plane magnetization (blue arrow), and the out-of-plane mirror symmetry breaking produces an in-plane DM interaction (green arrow), perpendicular to the nearest-neighbor bond direction (red arrow); (c) For the phonon system, an idealized lattice vibration model with the first (green wavy line) and second nearest neighbor interaction (purple wavy line) are considered.

After summing over all lattice sites, the total DM interaction vanishes within the linear spin-wave theory. This is a general consequence of the DM vector $\boldsymbol{D}_{ij}$ being perpendicular to the magnetization. If $\boldsymbol{D}_{ij}$ is parallel to the magnetization, then the DM interaction explicitly enters into the spin-wave Hamiltonian and, as shown in previous work, gives rise to a thermal Hall effect carried by magnons [41, 42, 70–72].

Since the DM interaction is absent in the spin-wave Hamiltonian, the magnon subsystem alone does not exhibit the thermal Hall effect. This can also be understood by the following symmetry consideration. The thermal Hall effect is described by $\boldsymbol{j}^Q = \alpha_{xy}\hat{\boldsymbol{z}} \times \boldsymbol{\nabla} T$, where $\boldsymbol{j}^Q$ is the heat current, $\boldsymbol{\nabla} T$ is the temperature gradient, and α_{xy} is the thermal Hall conductivity. Even though the spin-wave Hamiltonian in Eq. (3.2) breaks the time-reversal symmetry, it remains invariant under the combined time-reversal ($\mathcal{T}$) and spin rotation ($\mathcal{C}_x$) by 180° around the x-axis (or any in-plane axis). Since $\boldsymbol{j}^Q$ is odd and $\boldsymbol{\nabla} T$ is even under $\mathcal{T}\mathcal{C}_x$, the existence of the $\mathcal{T}\mathcal{C}_x$ symmetry forbids the thermal Hall effect. This is reminiscent of a well-known fact about the

anomalous Hall effect: it vanishes in a uniform ferromagnet in the absence of the spin-orbit interaction [80].

For the phonon part, we consider a simple coupled-oscillator model described by the Hamiltonian

$$H_{\rm ph} = \sum_i \frac{\boldsymbol{p}_i^2}{2M} + \frac{1}{2}\sum_{i,j,\alpha,\beta} u_i^\alpha \Phi_{ij}^{\alpha\beta} u_j^\beta \ , \tag{3.4}$$

where M is the ion mass, $\boldsymbol{u}_i \equiv \boldsymbol{R}_i - \boldsymbol{R}_i^0$ is the displacement of the ith ion from its equilibrium position $\boldsymbol{R}_i^0$, $\boldsymbol{p}_i = \dot{\boldsymbol{u}}_i$ is the canonical momentum conjugate to $\boldsymbol{u}_i$, and $\Phi_{ij}^{\alpha\beta}$ is the dynamical matrix describing inter-ion interactions. Obviously, due to the presence of time-reversal symmetry, the phonon subsystem alone does not exhibit the thermal Hall effect either.

The magnon-phonon interaction enters through the dependence of the exchange interaction on the ion displacement $\boldsymbol{u}_i$, i.e., phonons. For the isotropic Heisenberg exchange, we find that expanding $J(R_{ij})$ in terms of $\boldsymbol{u}_i$ only normalizes the magnon energy, and cannot lead to the thermal Hall effect since it preserves the rotational symmetry in the spin space [77]. On the other hand, the in-plane DM interaction will have a nontrivial contribution to the magnon-phonon hybrid. Expanding the DM interaction in Eq. (3.3) to the first order in $\boldsymbol{u}_i$, we find

$$H_{\rm int} = \sum_{\langle i,j\rangle}\sum_{\alpha,\beta}(u_i^\alpha - u_j^\alpha)T^{\alpha\beta}(\boldsymbol{R}_{ij}^0)(\delta s_i^\beta - \delta s_j^\beta) \ , \tag{3.5}$$

where $T^{\alpha\beta}(\boldsymbol{R})$ is the magnon-phonon coupling matrix,

$$T^{\alpha\beta}(\boldsymbol{R}) = \frac{D}{|\boldsymbol{R}|}S[\delta^{\alpha\beta} - (1+\gamma)\hat{R}^\alpha \hat{R}^\beta] \ , \tag{3.6}$$

with $\gamma = -(dD/dR)/(D/R)$. In obtaining Eq. (3.6), we note that the DM interaction depends on both the bond length R_{ij} and the bond direction $\hat{R}_{ij}$. It is clear that at the lowest order of the expansion, only the in-plane phonon modes are involved in the magnon-phonon interaction, and we shall only consider these modes from now on.

Since the magnon-phonon interaction in Eq. (3.5) couples the spin $\delta\boldsymbol{s}$ to the displacement field $\boldsymbol{u}$, it can be regarded as an effective spin-orbit interaction for the magnon-phonon hybrid. In particular, it breaks the $\mathcal{TC}_x$ symmetry, making the thermal Hall effect possible. We have therefore found an interesting example in which neither the magnons nor the phonons alone exhibit the thermal Hall effect, but the magnon-phonon hybrid could via the magnon-phonon interaction.

We can also deduce the dependence of the thermal Hall conductivity α_{xy} on the DM interaction $\boldsymbol{D}_{ij}$ and the magnetization $\boldsymbol{M}$. Since α_{xy} is invariant under the out-of-plane mirror reflection, flipping the sign of D, which is determined by the mirror-symmetry breaking, does not change the sign of α_{xy}, i.e., α_{xy} must be an even function of D. However, if we flip the direction of the magnetization $\boldsymbol{M}$, the whole system turns into its time-reversal counterpart. Therefore, reversing the ground state magnetization changes the sign of α_{xy}.

3.3 Large Magnetic Anisotropy Limit

Having established the symmetry requirement for the magnon-phonon interaction induced thermal Hall effect, we now develop a quantitative theory. Let us first consider the limit of large magnetic anisotropy, $K \gg k_B T$. In this limit, the magnons are pushed well above phonons in energy, and the thermal transport is mainly contributed by phonons. We can thus integrate out the magnon degree of freedom to obtain an effective Hamiltonian for phonons [81]. Leaving the details in Appendix C, we find that the effective Hamiltonian for phonons is given by

$$H_{\rm ph}^{\rm eff} = \sum_{\boldsymbol{q}} \frac{(\boldsymbol{p}_{-\boldsymbol{q}} - \boldsymbol{\mathcal{A}}_{-\boldsymbol{q}}\boldsymbol{u}_{-\boldsymbol{q}})^T(\boldsymbol{p}_{\boldsymbol{q}} - \boldsymbol{\mathcal{A}}_{\boldsymbol{q}}\boldsymbol{u}_{\boldsymbol{q}})}{2M} + \frac{1}{2}\boldsymbol{u}_{-\boldsymbol{q}}^T \tilde{\boldsymbol{\Phi}}_{\boldsymbol{q}} \boldsymbol{u}_{\boldsymbol{q}} \ , \tag{3.7}$$

where $\tilde{\Phi}_{\boldsymbol{q}}^{\alpha\beta} \equiv \Phi_{\boldsymbol{q}}^{\alpha\beta} + \delta\Phi_{\boldsymbol{q}}^{\alpha\beta}$ is the renormalized dynamical matrix, and $\mathcal{A}_{\boldsymbol{q}}^{\alpha\beta}$ is the emergent gauge field experienced by phonons. Detailed calculation shows that $\delta\Phi_{\boldsymbol{q}}^{\alpha\beta}$ and $\mathcal{A}_{\boldsymbol{q}}^{\alpha\beta}$ are proportional to the real part and the imaginary part of the spin-spin response function of the ferromagnetic state, respectively (see Appendix C). Eq. (3.7) describes a phonon system in a perpendicular magnetic field, as mentioned in Ref. 82, and can lead to the thermal Hall effect of phonons.

We note that the mechanism of this phonon Hall effect is different from that originated from the Raman type spin-lattice interaction [73, 83]. In the Raman type interaction, the phonon modes couple to the static spin ground state, while in our model, phonons couple to magnons which describe the dynamic of the spin system.

3.4 Magnon-Phonon Hybrid

Generally, if the magnon and phonon bands are close in energy, we need to treat them on an equal footing and consider the complete Hamiltonian that includes both magnons and phonons, such that $H = H_{sw} + H_{\rm ph} + H_{\rm int}$. As a simple example, we consider a magnon-phonon interacting system on a 2D square lattice. The linear spin wave model in Eq. (3.2) can be solved by applying the Holstein-Primakoff transformation [84], $\delta s_{ix} = \sqrt{S/2}(a_i + a_i^\dagger)$, $\delta s_{iy} = -i\sqrt{S/2}(a_i - a_i^\dagger)$, and $\delta s_{iz} = -a_i^\dagger a_i$, where a_i and $a_i^\dagger$ are the creation and annihilation operators for magnons at the i-site. Note that we have set $\hbar = 1$. This transformation gives rise to the magnon band dispersion $E_{m\boldsymbol{q}} = 2SJ[2 - \cos(q_x a) - \cos(q_y a)] + K(2S-1)/2$. For the phonon part, we consider the first and the second nearest neighbor interactions without loss of generality (see Fig. 3.1(c)). The dynamic matrix in this case is given in Appendix B. The complete

Hamiltonian can be written in the momentum space as

$$\begin{aligned} H =& \sum_{\boldsymbol{q}} E_{m\boldsymbol{q}} a^\dagger_{\boldsymbol{q}} a_{\boldsymbol{q}} + \frac{\boldsymbol{p}_{-\boldsymbol{q}}\boldsymbol{p}_{\boldsymbol{q}}}{2M} + \frac{1}{2}\boldsymbol{u}^T_{-\boldsymbol{q}}\boldsymbol{\Phi}(\boldsymbol{q})\boldsymbol{u}_{\boldsymbol{q}} \\ &+ \sqrt{\frac{S}{2}} \sum_{\boldsymbol{\delta},\alpha} u^\alpha_{-\boldsymbol{q}}(1-e^{i\boldsymbol{q}\cdot\boldsymbol{\delta}})[T^{\alpha x}(\boldsymbol{\delta})(a_{\boldsymbol{q}} + a^\dagger_{-\boldsymbol{q}}) - iT^{\alpha y}(\boldsymbol{\delta})(a_{\boldsymbol{q}} - a^\dagger_{-\boldsymbol{q}})] \ , \end{aligned} \tag{3.8}$$

where $\boldsymbol{\delta}$ is the nearest neighbor vector, and $E_{m\boldsymbol{q}} = SJ(\zeta - \sum_{\boldsymbol{\delta}} e^{i\boldsymbol{q}\cdot\boldsymbol{\delta}}) + K(2S-1)/2$ is magnon dispersion.

The dynamics of the magnon-phonon hybrid excitation can be determined by the generalized Bogoliubov-de Gennes (BdG) equation. To do this, we transform into the Fourier space and work in the basis of $\hat{\boldsymbol{\psi}}_{\boldsymbol{q}} = [a_{\boldsymbol{q}}, a^\dagger_{-\boldsymbol{q}}, \tilde{u}^x_{\boldsymbol{q}}, \tilde{u}^y_{\boldsymbol{q}}, \tilde{p}^x_{-\boldsymbol{q}}, \tilde{p}^y_{-\boldsymbol{q}}]^T$, where the dimensionless operators are given by $\tilde{u}^\alpha_{\boldsymbol{q}} = \sqrt{M\Omega}u^\alpha_{\boldsymbol{q}}$ and $\tilde{p}^\alpha_{\boldsymbol{q}} = \sqrt{1/M\Omega}p^\alpha_{\boldsymbol{q}}$, and Ω is the vibration frequency of nearest neighbor ions. The quadratic Hamiltonian in Eq. (3.8) can always be written as

$$H = \frac{1}{2}\sum_{\boldsymbol{q}} \hat{\boldsymbol{\psi}}^\dagger_{\boldsymbol{q}} \mathcal{H}_{\boldsymbol{q}} \hat{\boldsymbol{\psi}}_{\boldsymbol{q}} \ . \tag{3.9}$$

Here the Hamiltonian matrix of the hybrid system $\mathcal{H}_{\boldsymbol{q}}$ has the form

$$\mathcal{H}_{\boldsymbol{q}} = \begin{pmatrix} E_{m\boldsymbol{q}} I_{2\times 2} & \boldsymbol{M}_1^\dagger & 0 \\ \boldsymbol{M}_1 & \tilde{\boldsymbol{\Phi}}(\boldsymbol{q}) & 0 \\ 0 & 0 & \Omega I_{2\times 2} \end{pmatrix} \ , \tag{3.10}$$

where $\boldsymbol{M}_1$ is proportional to the DM strength D, which is given by

$$\boldsymbol{M}_1 = \begin{pmatrix} H^{xx}_{\boldsymbol{q}} + iH^{xy}_{\boldsymbol{q}} & H^{yx}_{\boldsymbol{q}} + iH^{yy}_{\boldsymbol{q}} \\ H^{xx}_{\boldsymbol{q}} - iH^{xy}_{\boldsymbol{q}} & H^{yx}_{\boldsymbol{q}} - iH^{yy}_{\boldsymbol{q}} \end{pmatrix} \ , \tag{3.11}$$

where $H^{\alpha\beta}_{\boldsymbol{q}} = \sqrt{S/2M\Omega}\sum_{\boldsymbol{\delta}} T^{\alpha\beta}_{\boldsymbol{\delta}}(1-e^{i\boldsymbol{q}\cdot\boldsymbol{\delta}})$, and $\tilde{\boldsymbol{\Phi}}(\boldsymbol{q}) = \boldsymbol{\Phi}(\boldsymbol{q})/(M\Omega)$.

Now we solve the Heisenberg equation of motion $i\partial_t\hat{\boldsymbol{\psi}}_{\boldsymbol{q}} = [\hat{\boldsymbol{\psi}}_{\boldsymbol{q}}, H]$. Using the property $\hat{\boldsymbol{\psi}}^\dagger_{\boldsymbol{q}} = \hat{\boldsymbol{\psi}}_{-\boldsymbol{q}}$ and $\mathcal{H}^T_{-\boldsymbol{q}} = \mathcal{H}_{\boldsymbol{q}}$, we have

$$i\eta\partial_t\hat{\boldsymbol{\psi}}_{\boldsymbol{q}} = \mathcal{H}_{\boldsymbol{q}}\hat{\boldsymbol{\psi}}_{\boldsymbol{q}} \ , \tag{3.12}$$

where the matrix η is given by

$$\eta = [\hat{\boldsymbol{\psi}}_{\boldsymbol{q}}, \hat{\boldsymbol{\psi}}^\dagger_{\boldsymbol{q}}] = \begin{pmatrix} \sigma_z & 0 & 0 \\ 0 & 0 & iI_{2\times 2} \\ 0 & -iI_{2\times 2} & 0 \end{pmatrix} \ . \tag{3.13}$$

Here we have used $\eta^2 = I_{6\times 6}$.

The frequency of the magnon-phonon hybrid excitation can be derived by solving the eigenvalue problem of the generalized BdG equation $E_{n\boldsymbol{q}}\eta\psi_{n\boldsymbol{q}} = \mathcal{H}_{\boldsymbol{q}}\psi_{n\boldsymbol{q}}$ with $\psi_{n\boldsymbol{q}}$ being the coefficient matrix of basis $\hat{\psi}_{\boldsymbol{q}}$.

Here we discuss the orthonormal relation of eigenvectors $\psi_{n\boldsymbol{q}}$. The eigenvalue problems for $\psi_{n\boldsymbol{q}}$ is equivalent to

$$\tilde{H}_{\boldsymbol{q}}\psi_{n,\boldsymbol{q}} = E_{n\boldsymbol{q}}\psi_{n,\boldsymbol{q}} \ , \tag{3.14}$$

where $\tilde{H}_{\boldsymbol{q}} = \eta\mathcal{H}_{\boldsymbol{q}}$. Since $\eta\tilde{H}_{\boldsymbol{q}}$ is Hermitian, it is easy to prove the orthogonal relation $\psi^{\dagger}_{m\boldsymbol{q}}\eta\psi_{n\boldsymbol{q}} = 0$ if $n \neq m$. The eigenvectors are normalized by $E_{n\boldsymbol{q}}\psi^{\dagger}_{n\boldsymbol{q}}\eta\psi_{n\boldsymbol{q}} = 1$ [1]. Define $\bar{\psi}_{m\boldsymbol{q}} = \psi^{\dagger}_{m\boldsymbol{q}}\eta\tilde{H}_{\boldsymbol{q}} = E_{n\boldsymbol{q}}\psi^{\dagger}_{m\boldsymbol{q}}\eta$, and we get a clean orthonormal relation which can be written as

$$\bar{\psi}_{m,\boldsymbol{q}}\psi_{n,\boldsymbol{q}} = \delta_{mn} \ . \tag{3.15}$$

We note that the spectrum of magnon-phonon hybrid excitations has a particle-hole symmetry. Define a unitary transformation

$$U = \begin{pmatrix} \sigma_x & 0 \\ 0 & I_{4\times 4} \end{pmatrix} \ , \tag{3.16}$$

we show that $U^{\dagger}\tilde{H}_{\boldsymbol{q}}U = -\tilde{H}^{*}_{-\boldsymbol{q}}$. Using Eq.(3.14), the relation between the states with opposite energies is given by

$$\begin{aligned} U^{\dagger}\tilde{H}_{\boldsymbol{q}}UU^{\dagger}\psi_{n,\boldsymbol{q}} &= E_{n\boldsymbol{q}}U^{\dagger}\psi_{n,\boldsymbol{q}} \\ -\tilde{H}^{*}_{-\boldsymbol{q}}U^{\dagger}\psi_{n,\boldsymbol{q}} &= E_{n\boldsymbol{q}}U^{\dagger}\psi_{n,\boldsymbol{q}} \\ \tilde{H}_{\boldsymbol{q}}U^{T}\psi^{*}_{n,-\boldsymbol{q}} &= -E_{n-\boldsymbol{q}}U^{T}\psi^{*}_{n,-\boldsymbol{q}} \ . \end{aligned} \tag{3.17}$$

Therefore, if we denote the positive and negative branches with the upper script $(+)$ and $(-)$ respectively, then $\psi^{(-)}_{n,\boldsymbol{q}} = U^{T}\psi^{(+)*}_{n,-\boldsymbol{q}}$.

3.5 Thermal Hall Conductivity From Wave Packet Theory

We derive the thermal Hall conductance of the magnon-phonon hybrid excitation using the wave packet theory, and the result is confirmed by linear response calculations in Section 3.6. The wave packet of the hybrid excitation is written as $|W\rangle = \int dq^2 w(\boldsymbol{q},t)e^{i\boldsymbol{q}\cdot\boldsymbol{r}}|\psi_{n\boldsymbol{q}}\rangle$, where $w(\boldsymbol{q},t)$ is the envelop function centered around the center-of-mass momentum $\boldsymbol{q}_c = \int dq^2|w(\boldsymbol{q},t)|^2\boldsymbol{q}$. Accordingly, the center of the wave packet in real space is given by $\boldsymbol{r}_c = \langle W|\eta\boldsymbol{r}|W\rangle/\langle W|\eta|W\rangle$. We derive the equation of motion for the wave packet from the Lagrangian $\mathcal{L} = \langle W|i\eta d/dt -$

[1] Including the $E_{n\boldsymbol{q}}$ in the normalization relation is important, since the normalization for the positive and negative spectrum must be different by a minus sign.

$\mathcal{H}|W\rangle/\langle W|\eta|W\rangle$, and $\langle W|\mathcal{H}|W\rangle/\langle W|\eta|W\rangle = E_{n\boldsymbol{q}_c}$. Using the Euler-Lagrangian equation, we can obtain the equation of motion for the wave packet center $\boldsymbol{r}_c$ [75, 85–87]

$$\dot{\boldsymbol{r}_c} = \frac{\partial \mathcal{E}_{n\boldsymbol{q}_c}}{\partial \boldsymbol{q}_c} + \frac{1}{-}\nabla U(\boldsymbol{r}_c) \times \boldsymbol{\Omega}_n(\boldsymbol{q}_c) \ , \tag{3.18}$$

where $U(\boldsymbol{r})$ is the potential felt by the wave packet, and the Berry curvature is defined by $\Omega_n^z = \partial_{q_x} A_{ny} - \partial_{q_y} A_{nx}$, with $\boldsymbol{A}_n = i\langle\psi_{n\boldsymbol{q}}|\eta\partial_{\boldsymbol{q}}|\psi_{n\boldsymbol{q}}\rangle/\langle\psi_{n\boldsymbol{q}}|\eta|\psi_{n\boldsymbol{q}}\rangle$ being the Berry connection.

Using the equation of motion of the wave packet, the thermal Hall current for Bosonic excitations is given by [22, 71, 72, 88]:

$$\begin{aligned}\boldsymbol{j} = &\frac{k_B^2 T}{}\hat{\boldsymbol{z}} \times \nabla T \sum_n \int \frac{d^2q}{(2\pi)^2}\Omega_n^z(\boldsymbol{q}) \\ &\left[(1+\rho_{n\boldsymbol{q}})\ln^2\left(\frac{1+\rho_{n\boldsymbol{q}}}{\rho_{n\boldsymbol{q}}}\right) - \ln^2\rho_{n\boldsymbol{q}} - 2\mathrm{Li}_2(-\rho_{n\boldsymbol{q}})\right] .\end{aligned} \tag{3.19}$$

Here $\rho_{n\boldsymbol{q}} = 1/(e^{E_n(\boldsymbol{q})/k_BT} - 1)$ is the Bose-Einstein distribution function with a zero chemical potential, and the index n in the Berry curvature Ω_n^z is summed over all positive bands.

3.6 Thermal Hall Conductivity From Linear Response Theory

3.6.1 The Energy Current Operator

The energy current operator can be derived from the continuity equation

$$\dot{H}(\boldsymbol{x}) + \nabla \cdot \boldsymbol{j}_E(\boldsymbol{x}) = 0 \ , \tag{3.20}$$

where $H(\boldsymbol{x})$ is the local energy operator defined as [89, 90]:

$$H(\boldsymbol{x}) = \sum_i \Delta(\boldsymbol{x} - \boldsymbol{R}_i) H_i \ . \tag{3.21}$$

Here $\Delta(\boldsymbol{x})$ is a smooth function that is localized in a small region near $\boldsymbol{x} = 0$ and satisfies $\int \Delta(\boldsymbol{x})d\boldsymbol{x} = 1$. H_i is given by

$$H_i = H_{i,spin} + H_{i,ph} + H_{i,int} \ , \tag{3.22}$$

where

$$H_{is} = -\frac{J}{2}\sum_{j\in i} \boldsymbol{s}_i \cdot \boldsymbol{s}_j - \frac{K}{2}s_{iz}^2 + \sum_{j\in i}\frac{\boldsymbol{D}_{ij}}{2}\cdot(\boldsymbol{s}_i \times \boldsymbol{s}_j) \ . \tag{3.23}$$

$$H_{i,int} = \sum_{j\in i}\sum_{\alpha,\beta}\frac{1}{2}(u_i^\alpha - u_j^\alpha)T_{ij}^{\alpha\beta}(\delta s_i^\beta - \delta s_j^\beta)\ , \tag{3.24}$$

$$H_{i,ph} = \frac{\boldsymbol{p}_i^2}{2M} + \frac{1}{2}\sum_j \Phi_{ij}^{\alpha\beta}u_i^\alpha u_j^\beta\ . \tag{3.25}$$

The sub-index $j \in i$ in the summation means site j is a nearest neighbor of site i. Note that $\sum_i H_i = H$.

The energy current operator derived from the continuity equation is defined up to a curl of a vector field. We show that (see Appendix A.1 for details)

$$\boldsymbol{J}^E(\boldsymbol{q}) = \frac{\Delta(\boldsymbol{q})}{8}\sum_{\boldsymbol{q}_1}\hat{\boldsymbol{\psi}}_{\boldsymbol{q}_1}^\dagger\eta(\boldsymbol{v}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1} + \tilde{H}_{\boldsymbol{q}_1}\boldsymbol{v}_{\boldsymbol{q}_1} + \boldsymbol{v}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}} + \tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\boldsymbol{v}_{\boldsymbol{q}_1+\boldsymbol{q}})\hat{\boldsymbol{\psi}}_{\boldsymbol{q}_1+\boldsymbol{q}}\ , \tag{3.26}$$

where $v_{\boldsymbol{q}} = \partial_{\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}}$ is the velocity matrix. For a small $\boldsymbol{q}$, we simply set $\Delta(\boldsymbol{q}) = 1$. Equivalently, we can write the energy current operator in a more compact format:

$$\boldsymbol{J}^E(\boldsymbol{q}) = \frac{1}{8}\sum_{\boldsymbol{q}_1}\bar{\hat{\psi}}_{\boldsymbol{q}_1}(\boldsymbol{v}_{\boldsymbol{q}_1} + \boldsymbol{v}_{\boldsymbol{q}_1})\hat{\psi}_{\boldsymbol{q}_1+\boldsymbol{q}} + \hat{\psi}_{\boldsymbol{q}_1}^\dagger(\boldsymbol{v}_{\boldsymbol{q}_1} + \boldsymbol{v}_{\boldsymbol{q}_1})\bar{\hat{\psi}}_{\boldsymbol{q}_1+\boldsymbol{q}}^\dagger\ . \tag{3.27}$$

We write the energy current operator using the creation ($\hat{\Psi}^\dagger$) and annihilation ($\hat{\Psi}$) operators of the hybrid excitations, which are defined by

$$\hat{\psi}_{\boldsymbol{q}} = \sum_{n>0}\sqrt{E_{n,\boldsymbol{q}}}\psi_{n,\boldsymbol{q}}\hat{\Psi}_{n,\boldsymbol{q}} + \sqrt{E_{n,-\boldsymbol{q}}}U^T\psi_{n,-\boldsymbol{q}}^*\hat{\Psi}_{n,-\boldsymbol{q}}^\dagger\ . \tag{3.28}$$

The creation and annihilation operators satisfies the commutation relation for Bosons

$$[\hat{\Psi}_{n,\boldsymbol{q}}, \hat{\Psi}_{m,\boldsymbol{q}}^\dagger] = \delta_{mn}\ . \tag{3.29}$$

In the next section about the linear response theory, we will need the time dependent energy current operator. Since the time dependent operators for hybrid excitations at time t are written as $\hat{\Psi}_{n,\boldsymbol{q}}(t) = \hat{\Psi}_{n,\boldsymbol{q}}e^{-iE_{n\boldsymbol{q}}t/}$ and $\hat{\Psi}_{n,\boldsymbol{q}}(t)^\dagger = \hat{\Psi}_{n,\boldsymbol{q}}^\dagger e^{iE_{n\boldsymbol{q}}t/}$, the time dependent energy current operator is given by

$$\begin{aligned}\boldsymbol{J}^E(\boldsymbol{q},t) =& \frac{1}{8}\sum_{n,m>0,\boldsymbol{q}_1}[\hat{\Psi}_{n,\boldsymbol{q}_1}^\dagger\boldsymbol{\Gamma}_{mn}^{(1)}(\boldsymbol{q}_1,\boldsymbol{q})\hat{\Psi}_{m,\boldsymbol{q}_1+\boldsymbol{q}} + \hat{\Psi}_{n,\boldsymbol{q}_1}^\dagger\boldsymbol{\Gamma}_{mn}^{(2)}(\boldsymbol{q}_1,\boldsymbol{q})\hat{\Psi}_{m,-\boldsymbol{q}_1-\boldsymbol{q}}^\dagger \\ &+ \Psi_{n,-\boldsymbol{q}_1}\boldsymbol{\Gamma}_{mn}^{(3)}(\boldsymbol{q}_1,\boldsymbol{q})\hat{\Psi}_{m,\boldsymbol{q}_1+\boldsymbol{q}} + \Psi_{n,-\boldsymbol{q}_1}\boldsymbol{\Gamma}_{mn}^{(4)}(\boldsymbol{q}_1,\boldsymbol{q})\hat{\Psi}_{m,-\boldsymbol{q}_1-\boldsymbol{q}}^\dagger]\ ,\end{aligned} \tag{3.30}$$

where the vertices are given by

$$\boldsymbol{\Gamma}_{mn}^{(1)}(\boldsymbol{q}_1,\boldsymbol{q},t) = \sqrt{E_{n\boldsymbol{q}_1}E_{m\boldsymbol{q}_1+\boldsymbol{q}}}\psi_{n,\boldsymbol{q}_1}^\dagger\eta\partial_{\boldsymbol{q}_1}(\tilde{H}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1} + \tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}})\psi_{m,\boldsymbol{q}_1+\boldsymbol{q}}e^{i(E_{n\boldsymbol{q}_1}-E_{m\boldsymbol{q}_1+\boldsymbol{q}})t/}\ , \tag{3.31}$$

$$\boldsymbol{\Gamma}_{mn}^{(2)}(\boldsymbol{q}_1,\boldsymbol{q},t) = \sqrt{E_{n\boldsymbol{q}_1}E_{m-\boldsymbol{q}_1-\boldsymbol{q}}}\psi_{n,\boldsymbol{q}_1}^\dagger\eta\partial_{\boldsymbol{q}_1}(\tilde{H}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1} + \tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}})U^T\psi_{m,-\boldsymbol{q}_1-\boldsymbol{q}}^*e^{i(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1-\boldsymbol{q}})t/}\ , \tag{3.32}$$

$$\mathbf{\Gamma}^{(3)}_{mn}(\boldsymbol{q}_1,\boldsymbol{q},t)=\sqrt{E_{n-\boldsymbol{q}_1}E_{m\boldsymbol{q}_1+\boldsymbol{q}}}\psi^T_{n,-\boldsymbol{q}_1}U^*\eta\partial_{\boldsymbol{q}_1}(\tilde{H}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1}+\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}})\psi_{m,\boldsymbol{q}_1+\boldsymbol{q}}e^{i(-E_{n-\boldsymbol{q}_1}-E_{m\boldsymbol{q}_1+\boldsymbol{q}})t/}\ , \tag{3.33}$$

$$\mathbf{\Gamma}^{(4)}_{mn}(\boldsymbol{q}_1,\boldsymbol{q},t)=\sqrt{E_{n-\boldsymbol{q}_1}E_{m-\boldsymbol{q}_1-\boldsymbol{q}}}\psi^T_{n,-\boldsymbol{q}_1}U^*\eta\partial_{\boldsymbol{q}_1}(\tilde{H}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1}+\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}})U^T\psi^*_{m,-\boldsymbol{q}_1-\boldsymbol{q}}e^{-i(E_{n-\boldsymbol{q}_1}-E_{m-\boldsymbol{q}_1-\boldsymbol{q}})^t} \tag{3.34}$$

Similarly, we can write the total Hamiltonian using the operators of the hybrid excitations. According to Eq. (3.9), the total Hamiltonian can be written as

$$H=\frac{1}{2}\sum_{\boldsymbol{q}}\bar{\hat{\psi}}_{\boldsymbol{q}}\hat{\psi}_{\boldsymbol{q}}\ , \tag{3.35}$$

or equivalently,

$$H=\sum_{n>0,\boldsymbol{q}}E_{n\boldsymbol{q}}(\hat{\Psi}^\dagger_{n,\boldsymbol{q}}\hat{\Psi}_{n,\boldsymbol{q}}+\frac{1}{2})\ . \tag{3.36}$$

3.6.2 Thermal Conductivity from Linear Response

The linear response theory calculates the response function of mechanical forces. However, since the temperature gradient is a statistical force, we cannot calculate its response directly using the linear response theory. Instead, we introduce a gravitational field [91] $\psi_G(\boldsymbol{r})$. In the existence of the gravitational field, the Hamiltonian becomes

$$H^\psi=\int d\boldsymbol{r}(1+\psi_G(\boldsymbol{r}))H(\boldsymbol{r})=H_0+\int d\boldsymbol{r}\psi_G(\boldsymbol{r})H(\boldsymbol{r})\ . \tag{3.37}$$

The expectation value of the total energy current under the gravitational field is given by $\langle J^\psi_E\rangle=Tr(\rho J^\psi_E)$, where ρ is the density matrix and J^ψ_E is the energy current operator in the existence of the gravitational field. According to the Einstein's relation [75], the transport coefficient of $T\nabla(1/T)$ is the same as that of $\nabla\psi_G(\boldsymbol{r})$. Therefore, in order to calculate the thermal Hall conductance, we only need to calculate the response function of the gravitational field.

The gravitational field modifies both the density matrix and the energy current. With the gravitational field, the density matrix becomes

$$\rho=\frac{1}{Z}\exp[-\int\beta_0(1+\psi_G(\boldsymbol{r}))H(\boldsymbol{r})d\boldsymbol{r}]=\rho_0+\rho_1\ , \tag{3.38}$$

where $\rho_0=\frac{1}{Z}\exp[-\int\beta_0H(\boldsymbol{r})d\boldsymbol{r}]$ is the density matrix in the equilibrium case. For the energy current $\boldsymbol{J}^\psi_E(\boldsymbol{r})$, we can do the same calculation as the $\psi_G=0$ case shown in the last section. It is easy to show that the energy current operator defined in Eq. (3.26) follows the the scaling law [75] $\boldsymbol{J}^\psi_E(\boldsymbol{r})\to(1+\psi_G(\boldsymbol{r}))^2\boldsymbol{J}_E$. We expand the energy current operator up to the first order of the gravitational field

$$\langle\boldsymbol{J}^\psi_E\rangle=Tr[(\rho_0+\rho_1)(1+\psi_G(\boldsymbol{r}))^2\boldsymbol{J}_E]\approx Tr(\rho_1\boldsymbol{J}_E)+2Tr(\rho_0\psi_G(\boldsymbol{r})\boldsymbol{J}_E)\ . \tag{3.39}$$

The first term is the Kubo term, which can be calculated using Kubo formula. The second term is the energy magnetization term.

Kubo Term

Generally, in the existence of a general time independent uniform external field described by a coupling strength F and an operator $\hat{Q}$, the total Hamiltonian is written as

$$H = H_0 + H_1 = H_0 - F\hat{Q} \ . \tag{3.40}$$

The standard Kubo formula calculates the expectation value of the response function to the external field. If the expectation value of operator $\hat{P}$ is the response function, Kubo formula gives

$$\langle \hat{P} \rangle = \frac{1}{-} \int_0^\infty dt e^{-\epsilon t} \int_0^\beta d\lambda \langle \dot{Q}(-i\lambda - t) P(0) \rangle F \ . \tag{3.41}$$

We put our case into the general linear response framework. The perturbation term from the gravitational field is written as

$$\begin{aligned} H_1 &= \int d\boldsymbol{r} \psi_G(\boldsymbol{r}) H(\boldsymbol{r}, t) \\ &= \int d\boldsymbol{r} \int_{-\infty}^t dt' \psi_G(\boldsymbol{r}) \dot{H}(\boldsymbol{r}, t') \\ &= -\int d\boldsymbol{r} \int_{-\infty}^t dt' \psi_G(\boldsymbol{r}) \nabla \cdot \boldsymbol{J}_E(\boldsymbol{r}, t') \\ &= \int d\boldsymbol{r} \int_{-\infty}^t dt' \nabla \psi_G(\boldsymbol{r}) \cdot \boldsymbol{J}_E(\boldsymbol{r}, t') \end{aligned} \tag{3.42}$$

Note that we have used the continuity equation (3.20) in the third line. Suppose the gradient of the gravitational field is a constant. Compared with the standard Kubo formula, we have $F = -\nabla\psi_G$, $\dot{Q} = \int \boldsymbol{J}_E(\boldsymbol{r}, t) = \boldsymbol{J}_E(\boldsymbol{q}, t)|_{\boldsymbol{q}=0}$ and $P = \boldsymbol{J}_E(\boldsymbol{q}, t)|_{\boldsymbol{q}=0}/V$. In the following discussions, I will simply write $\boldsymbol{J}_E(\boldsymbol{q}, t)|_{\boldsymbol{q}=0}$ as $\boldsymbol{J}_E(0, t)$. Accordingly, the response function for $\nabla\psi_G$ is given by

$$\sigma_\psi^{xy,kubo} = -\frac{1}{V} \int_0^\infty dt e^{-\epsilon t} \int_0^\beta d\lambda \langle J_E^y(0, -i\lambda - t) J_E^x(0, 0) \rangle \ . \tag{3.43}$$

Using the Einstein's relation, the response function of a constant $\nabla\psi_G$ correspond to that of a constant temperature gradient $T\nabla(1/T)$. Converting $\sigma_\psi^{xy,kubo}$ to the response function of ∇T, we get

$$\sigma^{xy,kubo} = \frac{1}{VT_0} \int_0^\infty dt e^{-\epsilon t} \int_0^\beta d\lambda \langle J_E^y(0, -i\lambda - t) J_E^x(0, 0) \rangle \ . \tag{3.44}$$

This is the Kubo term for the thermal Hall conductance.

The calculation for the correlation function $\langle J_E^y(0,-i\lambda-t)J_E^x(0,0)\rangle$ is given in Appendix A.2. We show that the Kubo term can be written as

$$\sigma^{xy,kubo}=\sum_{m,n,\boldsymbol{q}_1}\frac{(\mathcal{V}^x_{\boldsymbol{q}_1,n,m}\mathcal{V}^y_{\boldsymbol{q}_1,m,n}-\mathcal{V}^y_{\boldsymbol{q}_1,n,m}\mathcal{V}^x_{\boldsymbol{q}_1,m,n})E_{n\boldsymbol{q}_1}E_{m\boldsymbol{q}_1}n_{n\boldsymbol{q}_1}}{32iVT_0(E_{n\boldsymbol{q}_1}-E_{m\boldsymbol{q}_1})^2}\ , \tag{3.45}$$

where

$$\mathcal{V}_{\boldsymbol{q}_1,m,n}=2\psi^\dagger_{n,\boldsymbol{q}_1}\eta\partial_{\boldsymbol{q}_1}\tilde{H}^2_{\boldsymbol{q}_1}\psi_{m,\boldsymbol{q}_1}=2\psi^\dagger_{n,\boldsymbol{q}_1}\eta(v_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1}+\tilde{H}_{\boldsymbol{q}_1}v_{\boldsymbol{q}_1})\psi_{m,\boldsymbol{q}_1}\ . \tag{3.46}$$

Notice that the indices n,m are summed over both the positive and negative spectrums. Since $\eta\tilde{H}$ is Hermitian, we have $(\eta\tilde{H})^\dagger=\tilde{H}^\dagger\eta=\eta\tilde{H}$. Therefore, we can write $\mathcal{V}_{\boldsymbol{q}_1,m,n}$ using the normalized wave functions

$$\mathcal{V}_{\boldsymbol{q}_1,m,n}=2(\psi^\dagger_{n,\boldsymbol{q}_1}\partial_{\boldsymbol{q}_1}\tilde{H}^\dagger_{\boldsymbol{q}_1}\bar{\psi}^\dagger_{m,\boldsymbol{q}_1}+\bar{\psi}_{n,\boldsymbol{q}_1}\partial_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1}\psi_{m,\boldsymbol{q}_1})\ . \tag{3.47}$$

Magnetization Term

The energy current density contributed by the magnetization term is given by

$$\begin{aligned}\langle\boldsymbol{J}_E^{mag}\rangle=&\frac{2}{V}\int d\boldsymbol{r}Tr(\rho_0\psi_G(\boldsymbol{r})\boldsymbol{J}_E(\boldsymbol{r}))\\=&\frac{2}{V}\int d\boldsymbol{r}Tr(\rho_0\psi_G(\boldsymbol{r})\nabla\times\hat{\boldsymbol{M}}_E(\boldsymbol{r}))\\=&-\frac{2}{V}\int d\boldsymbol{r}Tr(\rho_0\nabla\psi_G(\boldsymbol{r})\times\hat{\boldsymbol{M}}_E(\boldsymbol{r}))\\=&-\frac{2}{V}\int d\boldsymbol{r}\nabla\psi_G(\boldsymbol{r})\times\boldsymbol{M}_E(\boldsymbol{r})\ ,\end{aligned} \tag{3.48}$$

where $\hat{\boldsymbol{M}}_E(\boldsymbol{r})$ is the energy magnetization operator, and $\boldsymbol{M}_E(\boldsymbol{r})=Tr[\rho_o\hat{\boldsymbol{M}}_E(\boldsymbol{r})]$ is the expectation value of $\hat{\boldsymbol{M}}_E(\boldsymbol{r})$. If the gradient of the gravitational field is a constant, the energy current from the magnetization is given by

$$\langle\boldsymbol{J}_E^{mag}\rangle=-\frac{2}{V}\nabla\psi_G\times\boldsymbol{M}_E\ , \tag{3.49}$$

where $\boldsymbol{M}_E=\int d\boldsymbol{r}\boldsymbol{M}_E(\boldsymbol{r})$. Using the Einstein's rule, the magnetization term of the energy current under a temperature gradient can be written as

$$\langle\boldsymbol{J}_E^{mag}\rangle=\frac{2}{T_0V}\nabla T\times\boldsymbol{M}_E\ . \tag{3.50}$$

Qin's article [75] shows that if the energy current operator satisfies the scaling law, which is true for our case, the energy magnetization has a simple expression:

$$2\boldsymbol{M}_E-T\frac{\partial\boldsymbol{M}_E}{\partial T}=\frac{\beta_0}{2i}\nabla_{\boldsymbol{q}}\times\langle h_{-\boldsymbol{q}};\boldsymbol{J}_E(\boldsymbol{q})\rangle_0|_{\boldsymbol{q}\to0}=\tilde{\boldsymbol{M}}_E\ , \tag{3.51}$$

where

$$h_{\boldsymbol{q}} = \frac{1}{2}\sum_{\boldsymbol{k}} \bar{\hat{\psi}}_{\boldsymbol{k}} \hat{\psi}_{\boldsymbol{k}+\boldsymbol{q}} \ , \tag{3.52}$$

and $\langle h_{-\boldsymbol{q}}; \boldsymbol{J}_E(\boldsymbol{q})\rangle_0$ is the canonical correlation function is defined by

$$\langle h_{-\boldsymbol{q}}; \boldsymbol{J}_E(\boldsymbol{q})\rangle_0 = \frac{1}{\beta_0}\int_0^{\beta} d\lambda \langle h_{-\boldsymbol{q}}(-i\lambda)\boldsymbol{J}_E(\boldsymbol{q})\rangle \ . \tag{3.53}$$

Note that $\tilde{\boldsymbol{M}}_E$ only has the z component in a two dimensional system. Calculating the correlation function gives

$$\tilde{M}_E^z = \frac{i\beta}{16}\lim_{\boldsymbol{q}\to 0}\partial_{q_y}\sum_{\boldsymbol{q}_1,m,n} \bar{\psi}_{n,\boldsymbol{q}_1+\boldsymbol{q}}\psi_{m,\boldsymbol{q}_1}(\mathcal{V}^x_{mn}(\boldsymbol{q}_1,\boldsymbol{q})+\mathcal{V}^{x*}_{nm}(\boldsymbol{q}_1,\boldsymbol{q}))E_{m\boldsymbol{q}_1}E_{n\boldsymbol{q}_1+\boldsymbol{q}}\frac{n_{n,\boldsymbol{q}_1+\boldsymbol{q}}-n_{m,\boldsymbol{q}_1}}{E_{m\boldsymbol{q}_1}-E_{n\boldsymbol{q}_1+\boldsymbol{q}}} \ , \tag{3.54}$$

where $\boldsymbol{\mathcal{V}}_{nm}(\boldsymbol{q}_1,\boldsymbol{q}) = \bar{\psi}_{n,\boldsymbol{q}_1}\partial_{\boldsymbol{q}_1}(\tilde{H}_{\boldsymbol{q}_1}+\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}})\psi_{m,\boldsymbol{q}_1+\boldsymbol{q}} + \psi^{\dagger}_{n,\boldsymbol{q}_1}\partial_{\boldsymbol{q}_1}(\tilde{H}^{\dagger}_{\boldsymbol{q}_1}+\tilde{H}^{\dagger}_{\boldsymbol{q}_1+\boldsymbol{q}})\psi_{m,\boldsymbol{q}_1+\boldsymbol{q}}$.

Total Thermal Hall Conductance

Combining the magnetization term and the Kubo term, the total thermal Hall conductance is given by

$$\sigma^{xy} = \sigma^{xy,kubo} + \frac{2M_E^z}{T_0 V} \ . \tag{3.55}$$

Using exactly the same derivation as shown in [75], we get

$$\sigma^{xy} = -\frac{\pi}{6\beta}C_{ph} - \frac{1}{T_0 V}\sum_{n>0,\boldsymbol{q}}\int_{E_{n\boldsymbol{q}}}^{\infty}\epsilon^2\frac{dn_\epsilon}{d\epsilon}\Omega_{n,\boldsymbol{q}}d\epsilon \ . \tag{3.56}$$

In the first term, $C_{ph} = \frac{2\pi}{V}\sum_{n>0,\boldsymbol{q}}\Omega_{n\boldsymbol{q}}$, which is 0 in our case. Simple derivation shows that the second term is equivalent to the result from the wave packet theory in Eq. (3.19).

3.7 Materials Consideration

We now estimate the size of the magnon-phonon interaction induced thermal Hall effect using realistic parameters. Let us consider the recently discovered monolayer ferromagnet CrI_3 [10], in which the Cr^{3+} ions carry $S = 3/2$, the lattice constant is $a = 5$ Å and the mass M is about 50 times of the proton mass which is $Mc^2 = 5\times 10^{10}$ eV [2]. The exchange coupling $J = 2.2$ meV and anisotropy $K = 1.36$ meV [92, 93]. We will also set the lattice vibration frequency for the nearest neighbor interactions

[2] The Cr^{3+} ions in CrI_3 forms a honeycomb lattice. However, since the thermal Hall effect is dominated by low-energy Γ-point physics, our square lattice mode should be sufficient for an order of magnitude estimation.

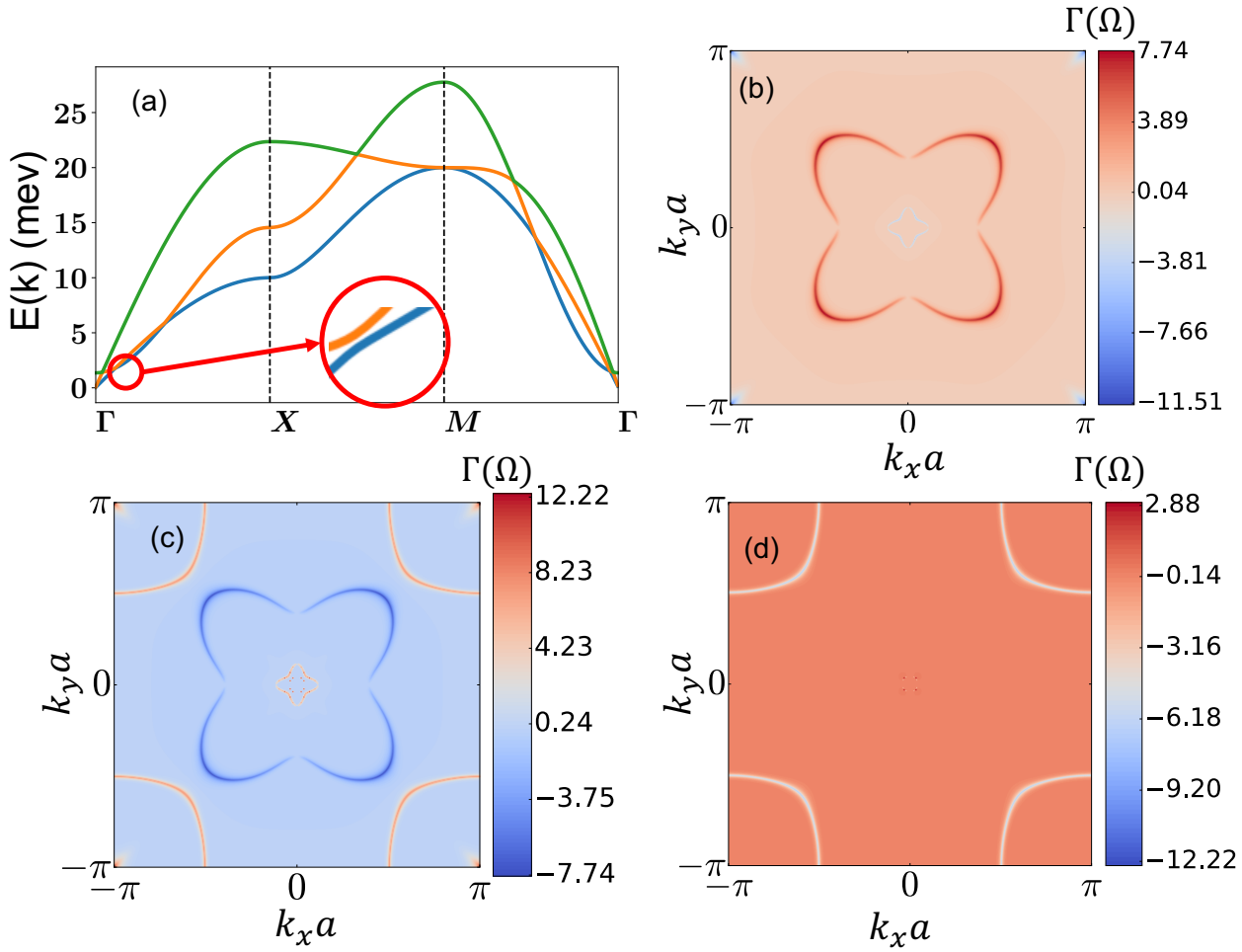

Figure 3.2: The band structure and Berry curvature using the parameters in the main text. (a) The band structure of the magnon-phonon hybrid system along the high symmetry line $\Gamma - X - M - \Gamma$. The bands from the highest to the lowest are denoted by green, orange, and blue, respectively; (b,c,d) The distribution of Berry curvatures in log-scale $\Gamma(\Omega^z) \equiv \text{sign}(\Omega^z)\ln(1 + |\Omega^z|)$: (b) for the blue band; (c) for the orange band; (d) for the green band.

at 10 meV, and for the second nearest neighbor interactions at 5 meV [92]. For simplicity, we have set $\gamma = 0$ in Eq. (3.6) [3]. Fig. 3.2(a) shows the band structure of the positive spectrum with $D = 0.93$ meV. The magnon and phonon bands have

[3]In the superexchange model, the DM strength $D \sim t_0 t'/U$, where t_0 and t' are the spin-independent and spin-dependent hopping integrals and U is the onsite Coulomb interaction [94]. In general, the hopping integrals have a power law dependence on the interatomic distance R [95], and the dependence of U on R can be neglected. If $D \sim R^{-n}$ then $dD/dR = -nD/R$. Therefore $\gamma = -(dD/dR)/(D/R) = n$ should have the same order as unity, and our numerical calculation based on $\gamma = 0$ should give an order of magnitude estimation of the thermal Hall conductivity.

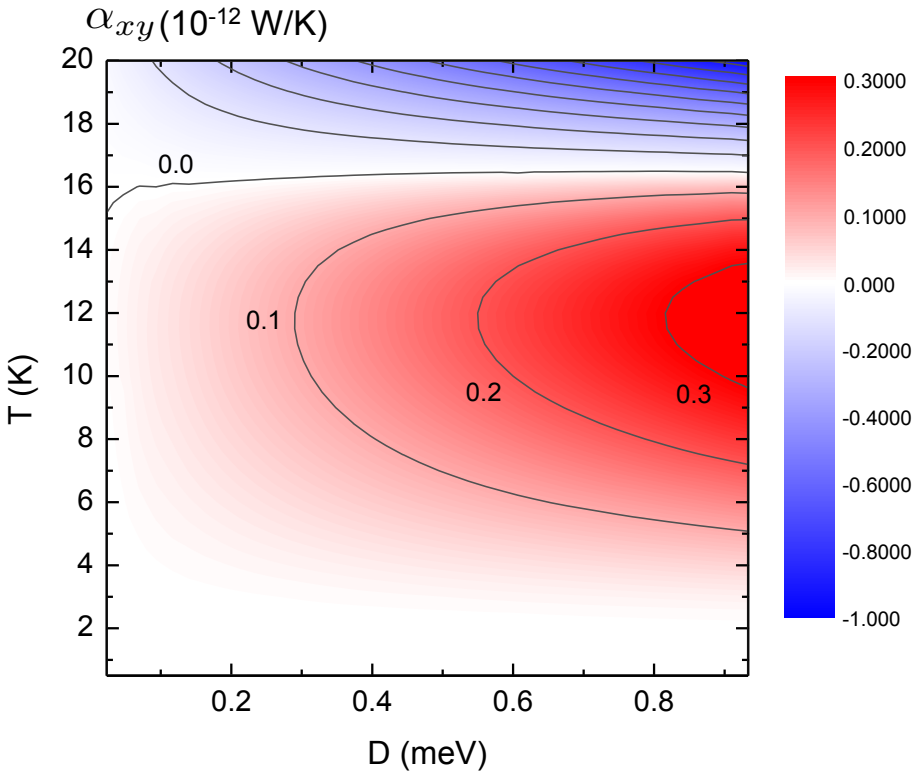

Figure 3.3: The thermal Hall conductivity α_{xy} for different temperatures and different strengths of DM interaction D. Other parameters are defined in the main text.

many crossing points, but they can be removed by the magnon-phonon interaction, although the gaps are too small to be seen on the figure. These avoided band crossing points have a large contribution to the Berry curvature. Fig. 3.2(b,c,d) show the distribution of the Berry curvatures of the three bands in log scale, which is given by $\Gamma(\Omega^z) = \text{sign}(\Omega^z)\ln(1 + |\Omega^z|)$.

The dependence of the thermal Hall conductivity α_{xy} on both temperature and the strength of the DM interaction is shown in Fig 3.3. For $T = 20$ K, $D = 0.93$ meV, we get $\alpha_{xy} \sim 9 \times 10^{-13}$ W/K. In Ref. [41], the magnon thermal Hall conductance of the 3D sample is around 10^{-3} WK^{-1}m^{-1}. If we assume that the thickness of a monolayer sample is 5 Å, then the thermal Hall conductance for one monolayer is about 5×10^{-13} W/K. Therefore, the order of magnitude of the thermal Hall conductance from our model is comparable to that from the magnon Hall effect. Note that the thermal Hall conductance flips sign when temperature is around 17 K. This is due to the contribution from the Berry curvatures of the higher bands near the Γ point, which have the opposite sign and larger magnitude compared with that from the lowest band.

3.8 Summary

In summary, we have proposed a new mechanism for the thermal Hall effect in an exchange spin-wave system by magnon-phonon interactions. The key ingredient is an out-of-plane magnetization and an in-plane DM vector due to mirror symmetry breaking. This can be realized in bulk crystals with broken mirror symmetry, or in low-dimensional systems where the mirror symmetry is broken by either a heterointerface or an applied gate voltage. Our result revealed the crucial role of the magnon-phonon interaction in the thermal Hall effect, and may find applications in the emerging field of spin caloritronics [79].

Chapter 4

Optical Selection Rule of Excitons in Gapped Chiral Fermion Systems

4.1 Introduction

Our understanding of optical absorption in semiconductors relies on two essential approximations [96]. The first is the effective mass approximation [97], in which the electron and the hole are considered as two particles moving with the effective masses of the conduction and valence bands, respectively. In the presence of the Coulomb interaction, the electron-hole pair will form a hydrogen-like bound state known as the exciton [98], which plays a crucial role in semiconductor optics. The second approximation is the electric dipole approximation. Within this approximation, the inter-band optical transition is usually understood in terms of the transition between atomic orbitals that make up the Bloch functions. Together, these two approximations yield the optical selection rule for excitons, as derived in a classic paper by Elliott [99]: if the band edge transition is dipole-allowed, then only the s-like excitons are bright and the rest are dark. Despite its simplicity, this theory is quite versatile and can be further generalized to include complications such as band degeneracy, anisotropy, and spin-orbit interaction.

However, the validity of the above theory has been recently challenged in a new class of materials called gapped chiral fermion (CF) systems. Examples include gapped topological surface states [100], biased bilayer graphene [101, 102], and monolayers of group-VI transition metal dichalcogenides such as MoS_2 [2, 103, 104]. It has been shown that in these systems the effective mass approximation must be modified to include the Berry phase [27] carried by the CFs to give a proper account of the exciton energy spectrum [105, 106]. At the same time, anomalous exciton optical selection rule has also been found in these systems. For example, it has been shown that both the s-like and d-like excitons are bright in monolayer MoS_2, and their optical transitions have opposite circular polarization [107], while in biased bilayer graphene it is the p-like excitons that are bright [101]. These results suggest that a new exciton

optical selection rule must be established in gapped CF systems.

In this chapter, we show that the exciton optical selection rule in gapped CF systems is governed by their winding number w [see Eq. (4.23) below], a topological property of the Bloch bands [108, 109]. Specifically, we find that the bright excitons in an isotropic CF system have angular momentum $m = w \pm 1$. When the full rotational symmetry is reduced to discrete C_N symmetry by crystal field effect, the allowed angular momentum of bright excitons expands to $m = w \pm 1 + nN$, where n is an integer. Our theory thus gives a unified view of the optical selection rule previously found in various gapped CF systems [100, 101, 107]. To further demonstrate our theory, we propose two gapped CF systems capable of hosting dark s-like excitons. The first is gapped surface states of a topological crystalline insulator with C_4 symmetry. The second is $3R$-stacked MoS_2 bilayers. In the latter case, we show that gating can be used to tune the s-like exciton from bright to dark by changing the winding number. The value of the gate voltage to realize such a dark-bright transition is within experimental reach. Our study, together with the previous result on the Berry phase effect on the exciton spectrum [105, 106], provides a basic description of the electronic structure of excitons in gapped CF systems.

4.2 Light-Electron Interactions in Semiconductors

4.2.1 Stimulated Absorption

Isolated Atoms

The interactions between light and electrons in semiconductors involve multiple processes. Initially, the semiconductor is in the electronic ground state, with valence band filled and conduction band empty. The photon field, described by a classical plane wave $\boldsymbol{A} = \boldsymbol{A}_0 \cos(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t)$, provides a perturbation that induces stimulated absorption. An electron in the valence band can absorb a photon and be excited into the conduction band, leaving a hole behind in the valence band. This process is described by the following Hamiltonian:

$$H = \frac{(\boldsymbol{p} - e\boldsymbol{A}(\boldsymbol{r},t))^2}{2m} + V(\boldsymbol{r}) \approx H_0 + H_1 \,, \tag{4.1}$$

where $H_0 = p^2/2m + V(\boldsymbol{r})$ gives rise to the Bloch band with energy $E_{n\boldsymbol{k}}$ and $V(\boldsymbol{r})$ is a periodic potential. The perturbation from the photon field is given by

$$H_1 = -\frac{e}{m}\boldsymbol{A}(\boldsymbol{r},t)\cdot\boldsymbol{p} = -e\boldsymbol{A}(\boldsymbol{r},t)\cdot\boldsymbol{v} \,, \tag{4.2}$$

where the velocity operator is defined as $\boldsymbol{v} = \boldsymbol{p}/m$. We derive the transition probability using the time dependent perturbation theory. Consider an electron initially

in a valence band state $\psi_{v\boldsymbol{k}_0}$. In the presence of the photon field, the wave function after a time interval t is given by

$$\Psi(t) = \sum_{n\boldsymbol{k}} c_{n\boldsymbol{k}}(t) e^{-iE_{n\boldsymbol{k}}t} \psi_{n\boldsymbol{k}} \ . \tag{4.3}$$

Using the Schrodinger equation, the time evolution of the coefficients $c_{n\boldsymbol{k}}(t)$ is given by

$$i\dot{c}_{n\boldsymbol{k}}(t) = \sum_{n'\boldsymbol{k}'} e^{i(E_{n\boldsymbol{k}} - E_{n'\boldsymbol{k}'})t} \langle \psi_{n\boldsymbol{k}} | H_1 | \psi_{n'\boldsymbol{k}'} \rangle c_{n'\boldsymbol{k}'}(t) \ . \tag{4.4}$$

According to the time dependent perturbation theory, we replace $c_{n'k'}(t)$ on the right hand side with its initial value $c_{n\boldsymbol{k}}(0) = \delta_{n,v}\delta_{\boldsymbol{k}_0,\boldsymbol{k}}$. For $n, k \neq v, k_0$, the coefficients are given by:

$$\begin{aligned} c_{n\boldsymbol{k}}(t) =& -i \int_0^t dt \sum_{n'\boldsymbol{k}'} e^{i(E_{n\boldsymbol{k}} - E_{n'\boldsymbol{k}'})t} \langle \psi_{n\boldsymbol{k}} | H_1 | \psi_{n'\boldsymbol{k}'} \rangle \delta_{n',v}\delta_{\boldsymbol{k}_0,k'} \\ =& i \int_0^t dt e^{i(E_{n\boldsymbol{k}} - E_{v\boldsymbol{k}_0})t} \boldsymbol{A}_0 \cdot \int dr^3 (e^{i(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t)} + H.c.) \psi^*_{n\boldsymbol{k}}(r) e\boldsymbol{v} \psi_{v\boldsymbol{k}_0}(r) \ . \end{aligned} \tag{4.5}$$

When the perturbation is turned on after a long time, we take the limit $t \to \infty$. Therefore, the integral over t gives rise to a δ-function that conserves the total energy of electrons and photons. Then we define the periodic part of the Bloch function $u_{n\boldsymbol{k}}$ by $\psi_{n\boldsymbol{k}} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}}$, and separate the integral over $\boldsymbol{r}$ into a sum over all lattice points $\boldsymbol{R}_i$ and an integral in a unit cell. The sum over $\boldsymbol{R}_i$ gives rise to a δ-function that leads to momentum conservation, and the integral in a unit cell gives the velocity matrix $\langle u^*_{n\boldsymbol{k}} | \boldsymbol{v} | u_{v\boldsymbol{k}_0} \rangle$. Therefore, the coefficients $c_{n\boldsymbol{k}}$ are written as

$$c_{n\boldsymbol{k}} = -\delta(E_{n\boldsymbol{k}} - E_{v\boldsymbol{k}_0} - \omega)\delta(\boldsymbol{q} + \boldsymbol{k}_0 - \boldsymbol{k}) e\boldsymbol{A}_0 \cdot \langle u^*_{n\boldsymbol{k}} | \boldsymbol{v} | u_{v\boldsymbol{k}_0} \rangle \ . \tag{4.6}$$

The first two δ-functions represent the conservation of momentum and energy, and the last term is the off-diagonal matrix element of the velocity operator. Since the momentum of the photon is usually small, we can assume $\boldsymbol{q} \approx 0$, thus the momentum of the electron before and after the transition is approximately the same. Therefore, when the conservation of momentum and energy is satisfied, the possibility of the stimulated absorption is

$$|c_{nk_0}|^2 = |e\boldsymbol{A}_0 \cdot \boldsymbol{V}_{nv}|^2 \ , \tag{4.7}$$

where $\boldsymbol{V}_{nv} = \langle nk_0 | \boldsymbol{v} | vk_0 \rangle$ is the velocity matrix.

Bloch Band Systems

In Bloch band system, the definition of velocity operator needs to be modified. According to the Heisenberg equation of motion, the time derivative of the position

operator is given by $\dot{\boldsymbol{r}} = -i[\boldsymbol{r}, H]$. For Bloch band systems, we work under a $\boldsymbol{k}$-dependent representation, where the operators are given by $O(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} O e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$. Thus $\boldsymbol{v}(\boldsymbol{k}) = -ie^{-i\boldsymbol{k}\cdot\boldsymbol{r}}[\boldsymbol{r}, H]e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$.

Now we need to know the velocity matrix under the tight binding basis. Define $|u_n(\boldsymbol{k})\rangle = \sum_i e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_i - \boldsymbol{r})}\phi_n(\boldsymbol{R}_i)$, where $\phi_n(\boldsymbol{R}_i)$ is atomic state n on the atom site R_i. In the absence of the magnetic field, the tight bonding Hamiltonian is given by $H = \sum_{m,n,i,j} t_{ij}^{mn} |\phi_m(\boldsymbol{R}_i)\rangle\langle\phi_n(\boldsymbol{R}_j)|$. Under the basis of $|u_n(\boldsymbol{k})\rangle$, the velocity matrix is written as

$$
\begin{aligned}
v_{mn}(\boldsymbol{k}) =& \langle u_m(\boldsymbol{k})|\boldsymbol{v}(\boldsymbol{k})|u_n(\boldsymbol{k})\rangle \\
=& \sum_{i,j} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}\langle\phi_m(\boldsymbol{R}_i)| - i[\boldsymbol{r}, H]|\phi_n(\boldsymbol{R}_j)\rangle \\
=& -i\sum_{i,j}\sum_{i',j',m',n'} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}[\langle\phi_m(\boldsymbol{R}_i)|\boldsymbol{r}|\phi_{m'}(\boldsymbol{R}_{i'})\rangle t_{i'j'}^{m'n'}\langle\phi_{n'}(\boldsymbol{R}_{j'})|\phi_n(\boldsymbol{R}_j)\rangle \\
& - \langle\phi_m(\boldsymbol{R}_i)|\phi_{m'}(\boldsymbol{R}_{i'})\rangle t_{i'j'}^{m'n'}\langle\phi_{n'}(\boldsymbol{R}_{j'})|\boldsymbol{r}|\phi_n(\boldsymbol{R}_j)\rangle] \\
=& -i\sum_{i.j}\sum_{i',j',m',n'} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}(\langle\phi_m(\boldsymbol{R}_i)|\boldsymbol{r}|\phi_{m'}(\boldsymbol{R}_{i'})\rangle t_{i'j}^{m'n} - t_{ij'}^{mn'}\langle\phi_{n'}(\boldsymbol{R}_{j'})|\boldsymbol{r}|\phi_n(\boldsymbol{R}_j)\rangle) \,.
\end{aligned}
\tag{4.8}
$$

In the last line, we suppose that the atomic wave functions on different sites do not overlap, thus $\langle\phi_m(\boldsymbol{R}_i)|\phi_n(\boldsymbol{R}_j)\rangle = \delta_{mn}\delta_{ij}$.

The position $\boldsymbol{r}$ can be separated into the equilibrium lattice position $\boldsymbol{R}_i$ and the relative position to the lattice point given by $\tilde{\boldsymbol{r}} = \boldsymbol{r} - \boldsymbol{R}_i$. Define the polarization matrix as $\langle\phi_m(\boldsymbol{R}_i)|\tilde{\boldsymbol{r}}|\phi_n(\boldsymbol{R}_j)\rangle = \delta_{ij}\langle\phi_m(\boldsymbol{R}_i)|\tilde{\boldsymbol{r}}|\phi_n(\boldsymbol{R}_i)\rangle = \delta_{ij}\boldsymbol{r}^{mn}$. Substituting into Eq. (4.8), the velocity matrix can be written as

$$
\begin{aligned}
v_{mn}(\boldsymbol{k}) =& -i\sum_{i,j,m',m'} [e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}(\boldsymbol{r}^{mm'}t_{ij}^{m'n} - t_{ij}^{mn'}\boldsymbol{r}^{nn'}) + e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}(\boldsymbol{R}_i - \boldsymbol{R}_j)t_{ij}^{mn}] \\
=& -i\sum_{m',m'}(\boldsymbol{r}^{mm'}t_k^{m'n} - t_k^{mn'}\boldsymbol{r}^{nn'}) + \frac{\partial t^{mn}(\boldsymbol{k})}{\partial \boldsymbol{k}} \,,
\end{aligned}
\tag{4.9}
$$

where $t^{mn}(\boldsymbol{k}) = \sum_{ij} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}t_{ij}^{mn}$. Note that the matrix $\boldsymbol{t}(\boldsymbol{k})$ is sometimes written as $H_0(\boldsymbol{k})$ since it is the $\boldsymbol{k}$-dependent Hamiltonian matrix under the orbital basis. The first term comes from single atomic transition, whose magnitude (and thus the selection rule) is controlled by the polarization matrix $\boldsymbol{r}^{mm'}$. This term is not included in Ref [110]. The second term is related to the hopping of electrons among lattices.

4.2.2 Relaxation

The electron-hole pairs generated from stimulated absorptions initially have high energies (so called hot excitons). They can be cooled down by relaxation processes,

assisted by phonon and defects. Ultra fast relaxation processes have been observed for double layer graphene [111] and transition metal dichalcogenides [112]. The relaxation time is in the order of a few picoseconds. The relaxation process moves the electrons and holes near their corresponding band edge, then they form bond states called excitons due to the coulomb interaction. Details about excitons will be discussed in Section 4.3

4.2.3 Spontaneous Emission

Exciton states are not stable and can spontaneously emit photons without the perturbation from external electric magnetic field. The possibility of such emission is described by the oscillator strength which is defined as [113]:

$$O_n = \frac{\text{Possibility of spontaneous emission from state n in unit time}}{\text{Energy lost in unit time for classical oscillator/Total energy of classical oscillator}} . \tag{4.10}$$

It describes which portion of the emission predicted by the classical charged oscillator model actually happens. For a classical charged oscillator that oscillates in the $\hat{z}$ direction, it carries a dipole $\boldsymbol{p}_{os}(t) = p_{os}(0)\cos(\omega t)\hat{\boldsymbol{z}}$. The energy loss per unit time can be calculated by the projection of Poynting vector $\boldsymbol{S}$ on a sphere with a large radius R. Classical electrodynamic calculation gives

$$\begin{aligned} -\frac{dW}{dt} =& \int d\Omega \int_0^R r^2 d\boldsymbol{r} \cdot \boldsymbol{S} \\ =& \frac{p_{os}(0)^2\omega^4}{12\pi\epsilon_0}, \end{aligned} \tag{4.11}$$

where $\int d\Omega$ is an integral over the solid angle. Note that we have set the velocity of light equals 1. Furthermore, the energy of the oscillator is $W = \frac{1}{2}\mu\omega^2 z_0^2$, where μ is the mass of the oscillator and $z_0 = p_{os}(0)/e$ is the amplitude of the oscillation. Therefore, the denominator in equation (4.10) is written as:

$$\gamma = -\frac{dW}{dt}/W = \frac{e^2\omega^2}{6\pi\epsilon_0\mu} . \tag{4.12}$$

The probability of spontaneous emissions is described by the Einstein coefficient, which is defined by

$$A_n = \frac{e^2\omega}{3\pi\epsilon_0}|\langle GS|\boldsymbol{v}|ex_n\rangle|^2 , \tag{4.13}$$

where $|GS\rangle$ is the ground state and $|ex_n\rangle$ is the n-th excited state, with their corresponding energy E_{ex} and E_g. Therefore, combined with Eq. (4.11), the oscillator strength is given by

$$\begin{aligned} O_n =& \frac{A^n}{3\gamma} \\ =& \frac{2}{3\mu(E_{ex} - E_g)}|\langle GS|\boldsymbol{v}|ex_n\rangle|^2 . \end{aligned} \tag{4.14}$$

The factor 3 in equation (4.14) is included in the denominator since the oscillator can be polarized in all 3 directions. Similarly, the oscillator strength for photons with a specific polarization $\hat{\boldsymbol{e}}$ is given by

$$O_n|_{\hat{\boldsymbol{e}}} = \frac{2}{3\mu(E_{ex} - E_g)} |\langle GS|\hat{\boldsymbol{e}} \cdot \boldsymbol{v}|ex_n\rangle|^2 \ . \tag{4.15}$$

4.2.4 Summary

In sum, the optical processes in semiconductors are included in Fig. 4.1. An electron-hole pair is generated by incoming photons, then this pair is relaxed to excitonic bond states. Each bond state can emit photons via spontaneous emissions, which is also known as the recombination process for excitons. Note that there might be more complicated processes that have not been shown in Fig 4.1. For example, some non-radiative recombination processes such as the Auger effect [114, 115] have not been taken into consideration.

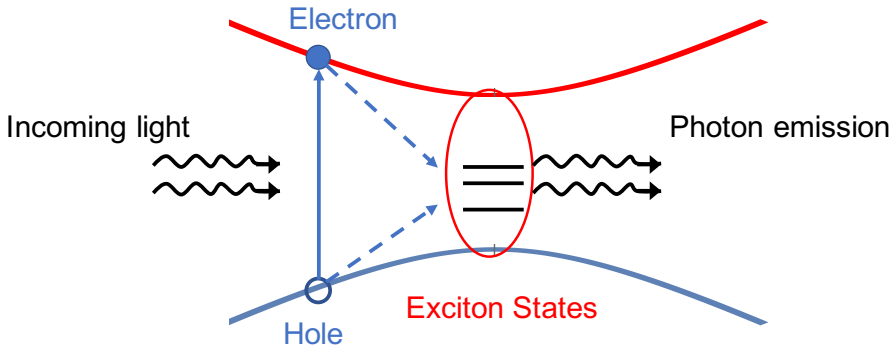

Figure 4.1: Optical process in semiconductors

4.3 Exciton States and Exciton Optical Selection Rule

4.3.1 Angular Momentums of Excitons

An exciton in a general two-band model can be written as a linear combination of electron-hole pairs,

$$|\Psi^{ex}(\boldsymbol{q})\rangle = \sum_{\boldsymbol{k}} f_{\boldsymbol{q}}(\boldsymbol{k}) a^\dagger_{c\boldsymbol{k}+\boldsymbol{q}} a_{v\boldsymbol{k}} |\Omega\rangle \ . \tag{4.16}$$

Here $|\Omega\rangle$ is the semiconductor ground state with the valence band filled and the conduction band empty, and $a^\dagger_{c\boldsymbol{k}+\boldsymbol{q}}$ ($a_{v\boldsymbol{k}}$) creates an electron (hole) in the conduction

(valence) band. The coefficient $f_{\boldsymbol{q}}(\boldsymbol{k})$ is the exciton envelope function, where $\boldsymbol{q}$ and $\boldsymbol{k}$ are the center-of-mass and relative momentum of the electron-hole pair, respectively. For photo-excited excitons, the center-of-mass momentum $\boldsymbol{q}$ is negligible, which will be set to zero and omitted hereafter. In the isotropic model, the angular momentum m is a good quantum number, thus the envelope functions have the following form

$$f_m(\boldsymbol{k}) = \tilde{f}_m(|\boldsymbol{k}|)e^{im\phi_{\boldsymbol{k}}} \ . \tag{4.17}$$

We note that the angular momentum of the envelope function is gauge dependent. In general, the envelope function is calculated using the Bethe-Salpeter equation [100, 116]:

$$(2\epsilon_{\boldsymbol{k}} + \Sigma_{\boldsymbol{k}})f(\boldsymbol{k}) - \sum_{k'} U(\boldsymbol{k}, \boldsymbol{k}')f(\boldsymbol{k}') = Ef(\boldsymbol{k}), \tag{4.18}$$

where $U(\boldsymbol{k}, \boldsymbol{k}') = V(|\boldsymbol{k} - \boldsymbol{k}'|)\langle c\boldsymbol{k}|c\boldsymbol{k}'\rangle\langle v\boldsymbol{k}'|v\boldsymbol{k}\rangle$, and $V(|\boldsymbol{k} - \boldsymbol{k}'|)$ is Coulomb interaction in momentum space, and the self energy correction $\Sigma_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V(|\boldsymbol{k}-\boldsymbol{k}'|)(|\langle v\boldsymbol{k}'|v\boldsymbol{k}\rangle|^2 - |\langle c\boldsymbol{k}|v\boldsymbol{k}'\rangle|^2)$. Using the Fourier transformation of the polar angle of $\boldsymbol{k}$, which is define by $\phi_{\boldsymbol{k}} = \arctan k_y/k_x$, we get the Bethe-Salpeter equation in angular momentum representation

$$(2\epsilon_k + \Sigma_k)\tilde{f}_m(|\boldsymbol{k}|) - \sum_{n,k'} U_{mn}(|\boldsymbol{k}|, |\boldsymbol{k}'|)\tilde{f}_n(|\boldsymbol{k}'|) = E\tilde{f}_m(|\boldsymbol{k}|) \ , \tag{4.19}$$

where

$$U_{mn}(|\boldsymbol{k}|, |\boldsymbol{k}'|) = \int d\phi_{\boldsymbol{k}} d\phi_{\boldsymbol{k}'} U(\boldsymbol{k}, \boldsymbol{k}')e^{-im\phi_{\boldsymbol{k}}}e^{in\phi_{\boldsymbol{k}'}} \ . \tag{4.20}$$

We show that the gauge choice of wave functions may influence the definition of angular momentum. Suppose we give a phase factor to the conduction band, which leads to the new wave function: $|c\boldsymbol{k}\rangle' = e^{i\phi_{\boldsymbol{k}}}|c\boldsymbol{k}\rangle$. This gauge transformation increases the angular momentum of the Bloch wave function m_{band} by 1. In addition, $U'(\boldsymbol{k}, \boldsymbol{k}') = U(\boldsymbol{k}, \boldsymbol{k}')e^{i\phi_{\boldsymbol{k}'}}e^{-i\phi_{\boldsymbol{k}}}$ while Σ_k is unchanged, thus the Bethe-Salpeter equation in the angular momentum space becomes

$$(2\epsilon_k + \Sigma_k)\tilde{f}_m(|\boldsymbol{k}|) - \sum_{n,k'} U'_{m-1,n-1}(|\boldsymbol{k}|, |\boldsymbol{k}'|)\tilde{f}_n(|\boldsymbol{k}'|) = E\tilde{f}_m(|\boldsymbol{k}|) \ . \tag{4.21}$$

Therefore, $\tilde{f}_m(|\boldsymbol{k}|)$ becomes $\tilde{f}_{m-1}(|\boldsymbol{k}|)$ after the gauge transformation. We note that the total angular mementum of the exciton state is unchanged under the gauge transformation. Although the angular momentum for the envelope function m_{env} is decreased by 1, m_{band} increases by 1, thus the entire angular momentum $m_{total} = m_{env} + m_{band}$ is unchanged. We will come back to the gauge freedom in the discussion of the exciton states in Chiral Fermion systems.

4.3.2 Optical Selection Rule of Excitons

The selection rule for the band edge absorption of electrons can be derived from the velocity matrix $V_{vc} = \langle v\boldsymbol{k}|\boldsymbol{v}|c\boldsymbol{k}\rangle$ (see equation (4.6)). The transition is not allowed only when the velocity matrix is zero at the band edge. It should be pointed out that there are generally two contributions to the velocity matrix element (see Section 4.2.1 for details): one is from the electron hopping between lattice sites, and the other from the dipole transition between localized orbitals [117]. Here we only consider the former contribution while neglecting the latter. This is justified for the systems considered in this chapter. In MoS_2 the conduction and valence band edges are mainly formed by the Mo d-orbitals, with slight mixing from the S p-orbitals [118]. There is no dipole transition between the even-parity d-orbitals, and transitions between d- and p-orbitals are negligible. Similarly, in gapped graphene systems the atomic orbitals involved are carbon p_z-orbitals, and optical transitions among them are dipole-forbidden.

On the other hand, for the spontaneous emission of excitons, we need to consider the additional contribution from the envelope function. Substituting the exciton state in Eq. (4.16) into the oscillator strength in Eq. (4.15), the oscillator strength of an exciton with angular momentum m under circular polarization is given by

$$O_m = \frac{1}{\mu E_m^{\text{ex}}} \sum_{\eta=\pm} \left| \int d\boldsymbol{k}\, \tilde{f}_m(|\boldsymbol{k}|) e^{im\phi_{\boldsymbol{k}}} v_\eta(\boldsymbol{k}) \right|^2 , \tag{4.22}$$

where $v_\eta(\boldsymbol{k}) = \langle v\boldsymbol{k}|\hat{v}_\eta|c\boldsymbol{k}\rangle$ is the interband matrix element of the velocity operator $\hat{v}_\eta = \hat{v}_x + i\eta\hat{v}_y$ with $\hat{v}_{x,y} = \partial H_0(\boldsymbol{k})/\partial k_{x,y}$. Here $H_0(\boldsymbol{k})$ is the $\boldsymbol{k}$ dependent Bloch Hamiltonian (see the text under Eq. (4.9)), E_m^{ex} is the exciton energy and μ is the reduced mass. The exciton state is dark if $O_m = 0$.

4.4 General Gapped Chiral Fermion Systems

We begin with the $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian for an isotropic two-dimensional CF model with an integer winding number w,

$$H_0 = \begin{pmatrix} \Delta & \alpha(|\boldsymbol{k}|)e^{iw\phi_{\boldsymbol{k}}} \\ \alpha(|\boldsymbol{k}|)e^{-iw\phi_{\boldsymbol{k}}} & -\Delta \end{pmatrix} , \tag{4.23}$$

where 2Δ is the energy gap and $\phi_{\boldsymbol{k}} = \tan^{-1}(k_y/k_x)$. This Hamiltonian describes a wide range of material systems. For example, both gapped topological surface states [100] and monolayer MoS_2 [2] have $\alpha(|\boldsymbol{k}|) \propto |\boldsymbol{k}|$ with the winding number $w = 1$, and biased bilayer graphene has $\alpha(|\boldsymbol{k}|) \propto |\boldsymbol{k}|^2$ with $w = 2$ [119]. In fact, in graphene multilayers, w can be made arbitrary integral values [120]. We note that this model also includes the special case of zero winding number even though it cannot

be called a chiral fermion anymore. The energy dispersion of this model is given by $\varepsilon_{c,v} = \pm\varepsilon_{\boldsymbol{k}} = \pm\sqrt{\Delta^2 + \alpha^2(|\boldsymbol{k}|)}$ with the corresponding eigenstates

$$|c\boldsymbol{k}\rangle = \begin{pmatrix} \cos\frac{\theta_k}{2} \\ \sin\frac{\theta_k}{2} e^{-iw\phi_{\boldsymbol{k}}} \end{pmatrix}, \quad |v\boldsymbol{k}\rangle = \begin{pmatrix} \sin\frac{\theta_k}{2} e^{iw\phi_{\boldsymbol{k}}} \\ -\cos\frac{\theta_k}{2} \end{pmatrix}, \tag{4.24}$$

where $\theta_k = \cos^{-1}(\Delta/\epsilon_{\boldsymbol{k}})$. The wave functions have a U(1) gauge freedom. Here we fix the gauge by demanding that both $|c\boldsymbol{k}\rangle$ and $|v\boldsymbol{k}\rangle$ have no singularity at the band edge ($\boldsymbol{k} = 0$). Under this gauge choice, the labeling of excitons by their angular momenta returns to that of the hydrogenic model in the large gap limit [105].

We calculate the velocity operator near the band edge. The velocity operator is given by $\boldsymbol{v} = \frac{\partial H_0}{\partial \boldsymbol{k}}$, such that

$$v_+ = \begin{pmatrix} 0 & \alpha_1(|\boldsymbol{k}|)e^{i(w+1)\phi_{\boldsymbol{k}}} \\ \alpha_2(|\boldsymbol{k}|)e^{-i(w-1)\phi_{\boldsymbol{k}}} & 0 \end{pmatrix}, \tag{4.25}$$

and

$$v_- = \begin{pmatrix} 0 & \alpha_2(|\boldsymbol{k}|)e^{i(w-1)\phi_{\boldsymbol{k}}} \\ \alpha_1(|\boldsymbol{k}|)e^{-i(w+1)\phi_{\boldsymbol{k}}} & 0 \end{pmatrix}, \tag{4.26}$$

where

$$\alpha_1(|\boldsymbol{k}|) = \frac{d\alpha(|\boldsymbol{k}|)}{d|k|}|\boldsymbol{k}|^w, \tag{4.27}$$

$$\alpha_2(|\boldsymbol{k}|) = \frac{d\alpha(|\boldsymbol{k}|)}{d|k|}|\boldsymbol{k}|^w + 2w\alpha(|\boldsymbol{k}|)|\boldsymbol{k}|^{w-1}. \tag{4.28}$$

In the large band gap limit, we can write the wave functions of the conduction (c) and valence band (v) to the lowest order of $|\boldsymbol{k}|$,

$$|c\boldsymbol{k}\rangle = \begin{pmatrix} 1 \\ \frac{1}{2}\frac{\alpha(|\boldsymbol{k}|)}{\Delta}k_-^w \end{pmatrix}, \quad |v\boldsymbol{k}\rangle = \begin{pmatrix} \frac{1}{2}\frac{\alpha(|\boldsymbol{k}|)}{\Delta}k_+^w \\ -1 \end{pmatrix}. \tag{4.29}$$

It is straightforward to show that

$$\langle v\boldsymbol{k}|v_+|c\boldsymbol{k}\rangle = [\frac{\alpha^2|\boldsymbol{k}|^{2w}}{4\Delta^2}\alpha_1(|\boldsymbol{k}|) - \alpha_2(|\boldsymbol{k}|)]e^{i(-w+1)\phi_{\boldsymbol{k}}} \propto e^{i(-w+1)\phi_{\boldsymbol{k}}}, \tag{4.30}$$

$$\langle v\boldsymbol{k}|v_-|c\boldsymbol{k}\rangle = [\frac{\alpha^2|\boldsymbol{k}|^{2w}}{4\Delta^2}\alpha_2(|\boldsymbol{k}|) - \alpha_1(|\boldsymbol{k}|)]e^{i(-w-1)\phi_{\boldsymbol{k}}} \propto e^{i(-w-1)\phi_{\boldsymbol{k}}}. \tag{4.31}$$

It then follows from Eq. (4.22) that after angular average only exciton states with $m = w \pm 1$ have nonzero oscillator strength. In addition, optical transitions to these two angular momentum states always have opposite circular polarization. We emphasize that it is the $\boldsymbol{k}$-space phase winding of the velocity matrix element, a feature not available in the atomic transition picture, that determines the exciton optical selection rule of gapped CF systems.

Although both $w+1$ and $w-1$ states are bright, their oscillator strength can be quite different. For simplicity, we assume $\alpha(|\boldsymbol{k}|) = \alpha|\boldsymbol{k}|^w$. The velocity matrix elements take the form

$$\begin{aligned}\langle v\boldsymbol{k}|\hat{v}_+|c\boldsymbol{k}\rangle &= -2\alpha w\cos^2\frac{\theta_k}{2}k^{w-1}e^{-i(w-1)\phi_{\boldsymbol{k}}}\ ,\\ \langle v\boldsymbol{k}|\hat{v}_-|c\boldsymbol{k}\rangle &= 2\alpha w\sin^2\frac{\theta_k}{2}k^{w-1}e^{-i(w+1)\phi_{\boldsymbol{k}}}\ .\end{aligned} \tag{4.32}$$

In the large band gap limit, i.e., $\Delta \gg \alpha k_B^w$ where k_B is the inverse of the exciton Bohr radius, we have $\cos\frac{\theta_{\boldsymbol{k}_B}}{2} \gg \sin\frac{\theta_{\boldsymbol{k}_B}}{2}$. In this case, the $m = w-1$ exciton states are much brighter than the $m = w+1$ states.

So far, we have only considered the isotropic case. However, in a crystalline environment the C_∞ symmetry is reduced to C_N by the crystal field effect, which will modify the optical selection rule. The modifications come from two places. First, the exciton state with angular momentum m is mixed with those with angular momentum $m+nN$,

$$f_m(\boldsymbol{k}) \to \tilde{f}_m(|\boldsymbol{k}|)e^{im\phi_{\boldsymbol{k}}} + \sum_{n\neq 0} c_n\tilde{f}_{m+nN}(|\boldsymbol{k}|)e^{i(m+nN)\phi_{\boldsymbol{k}}}, \tag{4.33}$$

where n is an integer and c_n is the coefficient for each angular momentum channel, whose form has been derived in Appendix D. Secondly, the velocity matrix element is also expanded into a series of angular momentum channels (see Appendix D)

$$\langle v\boldsymbol{k}|\hat{v}_\pm|c\boldsymbol{k}\rangle = \sum_n v_n e^{-i(w\mp 1+nN)\phi_{\boldsymbol{k}}}. \tag{4.34}$$

According to Eq. (4.22), the exciton selection rule now reads

$$m = w \pm 1 + nN\ . \tag{4.35}$$

This is a reflection of the fact that in a C_N invariant system the angular momentum is only defined modulo N [4]. Finally, we note that the optical transitions to the m- and $(m+nN)$-states have the same circular polarization.

4.5 Examples

4.5.1 Monolayer MoS_2

Now we examine our theory in the two previously studied systems. The first one is monolayer MoS_2 with winding number $w = 1$ (see section 1.3.2). According to our theory, the s- and d-like excitons should be bright with opposite circular polarizations when the crystal field effect is ignored, and the s-state should be much brighter than the d-state due to the relatively large band gap in MoS_2 ($\alpha k_B/\Delta \sim 0.1$) [2, 105]. If we turn on the crystal field, the symmetry is reduced from C_∞ to C_3. In this case,

the p-like state with $m = -1$, which is dark in the isotropic model, becomes bright and has the same polarization as the d-like excitons with $m = 2$. This result agrees with the direct calculation in a recent study [107].

4.5.2 Bilayer Graphene

The second example is the biased bilayer graphene [101], which is described by the following effective Hamiltonian [119]

$$H_{\mathrm{BLG}} = \begin{pmatrix} \Delta & \alpha k_+^2 \\ \alpha k_-^2 & -\Delta \end{pmatrix} + 3\gamma_3 \begin{pmatrix} 0 & k_- \\ k_+ & 0 \end{pmatrix} , \tag{4.36}$$

where $k_\pm = k_x \pm i k_y$, and γ_3 is the interlayer hopping amplitude. The first term in H_{BLG} describes an isotropic CF model with winding number $w = 2$. This term alone would give rise to dark s-states, since only the $m = 1$ and $m = 3$ states are bright. However, in the presence of the γ_3 term, which reduces the C_∞ symmetry to C_3, the optical transition to s-like states are turned on, and have opposite circular polarization compared to the p-like states. Similarly, the $m = -2$ states also become bright (see Fig. 4.2).

To estimate the crystal field effect, we have carried out a perturbative calculation by treating γ_3 as a small quantity in the large band gap limit (see Appendix D). We find that the modification to the exciton envelope function is a higher order contribution, and the main effect of the crystal field comes from its modification to the velocity matrix element, which is proportional to γ_3. Accordingly, the ratio of the oscillator strength between the s- and p-states should be proportional to $9\gamma_3^2/(2\alpha|k_B|)^2$. According to Ref. [101], the k-space radius of the exciton envelope function is $k_B \sim 0.02$ Å^{-1}, which gives $9\gamma_3^2/(2\alpha|k_B|)^2 \sim 0.02$. Note that from a pure group theory point of view, we can also come to the conclusion that the s-like excitons are bright. In contrast, our theory provides a quantitative estimation of the brightness of the s-state.

We can check the validity of this argument with numerical calculations using a more complete four-band model for the biased bilayer graphene. The band structure for bilayer graphene near the Fermi surface also has degenerates at the K and K' valleys. With a gap Δ opened by the gate voltage, the $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian near the K valley for the biased bilayer graphene is given by [39]

$$H'_{BLG} = \begin{pmatrix} \Delta & \gamma_0 k_+ & 0 & 3a\gamma_3 k_- \\ \gamma_0 k_- & \Delta & \gamma_1 & 0 \\ 0 & \gamma_1 & -\Delta & \gamma k_+ \\ 3a\gamma_3 k_+ & 0 & \gamma k_- & -\Delta \end{pmatrix} . \tag{4.37}$$

Here γ_0, γ_1, γ_3 correspond to intralayer NN, interlayer NN and interlayer NNN respectively, and a is the lattice constant. The basis for H'_{BLG} consists of A and B sublattices in the up (layer 1) and down (layer 2) layer, which are denoted as A_1, B_1, A_2, B_2 in Fig. 4.3. The band structure near the K valley is shown in Fig. 4.4.

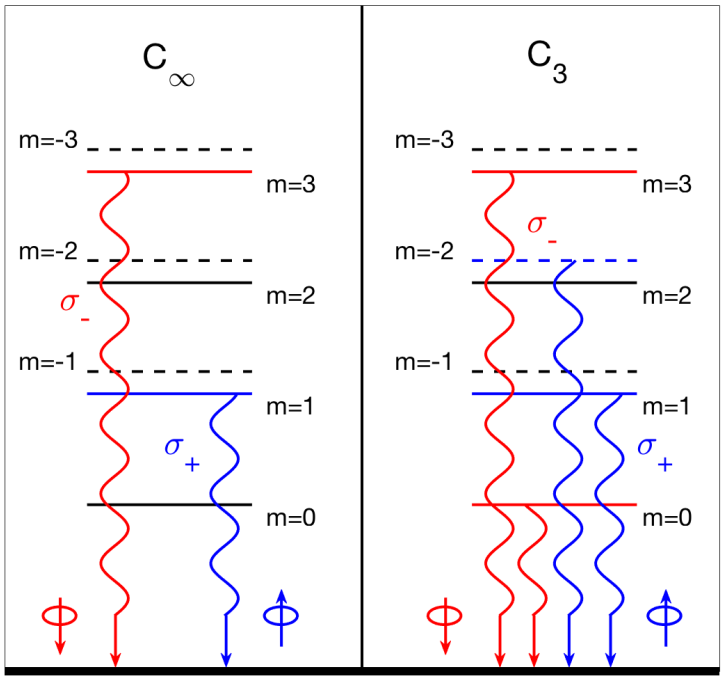

Figure 4.2: The exciton optical selection rule of the $w = 2$ chiral fermion model when the symmetry is reduced from (a) C_∞ to (b) C_3. The black lines indicate dark states, and the red (blue) lines are bright states with σ_- (σ_+) polarization. The solid lines represent positive angular momenta, and the dashed lines represent negative angular momenta.

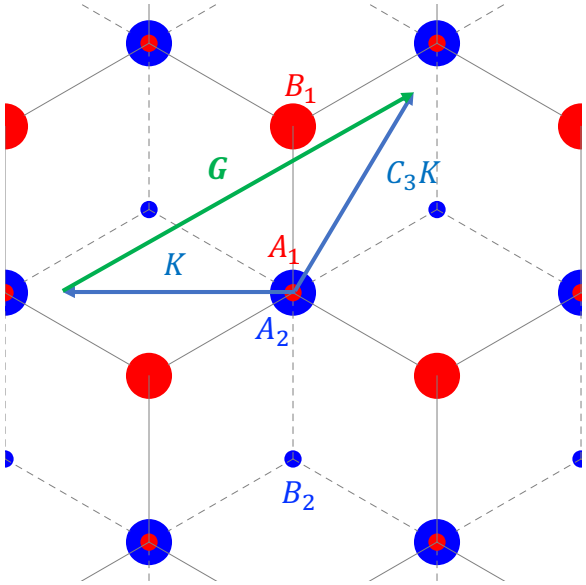

Figure 4.3: A-B stacking (Bernal) bilayer graphene. Different colors correspond to different layers. The blue arrows shows the wave vector K, and the green arrow shows the reciprocal lattice vector $\boldsymbol{G}$.

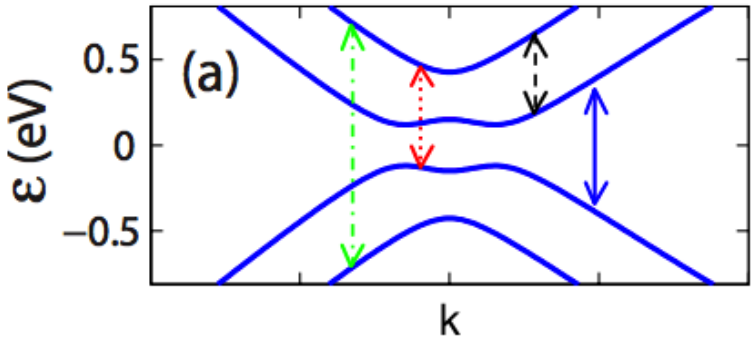

Figure 4.4: Band structure of bilayer graphene near the K valley. Figure is from Ref [4].

At the band edge, the wave functions for the lowest conduction and highest valence band are given by $\begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}^T$ and $\begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}^T$ [1]. The velocity matrix involving these two bands is denoted as $\boldsymbol{V}_{vc}(\boldsymbol{k})$.

The numerical results below have used the parameters in [101] and [39]. We first use the tight binding model in Ref. [39], and calculate the distribution of the

[1] That means electron and hole are in different layers. One property of interlayer exciton is that they have long lifetime.

magnitude of the velocity matrix with σ_+ polarization (V_{vc}^+) in an extended Brillouin Zone. The result is shown in Fig. 4.5. The dominant contribution to the velocity matrix $|V_{vc}^+|$ comes from the Dirac points, and the distribution of $|V_{vc}^-|$ is simply a time reversal counterpart. Now we focus on the K point and use the $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian in Eq. (4.37). Fig. 4.7 shows the velocity matrix distribution near the K point with $\sigma_\pm$ polarized light ($|V_{vc}^\eta|$ with $\eta = \pm$). Fig. 4.8 separates $|V_{vc}^\eta|$ into different angular momentum channels for both σ_- and σ_+ polarizations. As shown in the left figure of Fig. 4.8, the $m = 0$ channel of $|V_{vc}^+|$ is nonzero, thus the s-exciton is bright for the σ_+ polarization. On the other hand, the p-exciton ($m = -1$) is dark with σ_+ polarization. For the σ_- polarization, the p-state is bright while the s-state is dark according to the right figure of Fig. 4.7. This numerical result agrees with the optical selection rule discussed earlier.

Fig. 4.6 shows the distribution of the degree of polarization (DOP) in an extended Brillouin Zone using the tight binding model in Ref. [39]. Here DOP is defined by

$$\mathrm{DOP} = \frac{|V_{vc}^+|^2 - |V_{vc}^-|^2}{|V_{vc}^+|^2 + |V_{vc}^-|^2}. \tag{4.38}$$

We note that the DOP changes sign at the K point. Fig. 4.9 shows DOP along the k_y direction near the K point using the $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian in Eq. (4.37), and the band structure is also shown as a reference. A sharp peak that changes the sign of DOP appears at the K point which is absent when γ_3 is ignored [4].

For the completeness of the discussion, we provide a group theory argument for the optical selection rule of the s-exciton. We first consider the point group of the bilayer graphene. Under the Bernal stacking, the sublattice A_1 and A_2 in the two layers overlap in the $x-y$ plane (see Fig. 4.3). The bilayer graphene without a gate voltage has a $D_{3\mathrm{d}}$ symmetry [121], and the group of wave vector (GWV) at the K point is reduced to C_3 with or without gate voltage.

In group theory, the optical selection rule is derived from the irreducible representation (IR) of the band edge states. When $k = 0$, the wave functions of the conduction (valence) band are linear combinations of the B_1 (B_2) (labeled in Fig 4.3) atomic wave functions, respectively, and are given by

$$\psi_{c(v)K}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_{B1(B2)}} e^{i\boldsymbol{K}\cdot\boldsymbol{R}_{B1(B2)}}\phi(\boldsymbol{r} - \boldsymbol{R}_{B1(B2)}) \,, \tag{4.39}$$

where $\phi(\boldsymbol{r})$ is the wave function of the p_z orbitals on carbon atoms. Note that $\phi(\boldsymbol{r})$ is invariant under any rotations in the $x-y$ plane, and changes sign under the mirror reflection $z \to -z$. $\boldsymbol{R}_{B1(B2)} = \boldsymbol{R}_i + \tilde{\boldsymbol{R}}$, where $\boldsymbol{R}_i$ is the position of the unit cell, $\tilde{\boldsymbol{R}}$ is the relative position of the lattice point in the unit cell. We set $\tilde{\boldsymbol{R}} = 0$ at the A_1 point. To find out the IR of the wave functions, we can simply apply the symmetry operations in GWV. For a C_3 rotation around the O point in Fig. 4.3, The wave

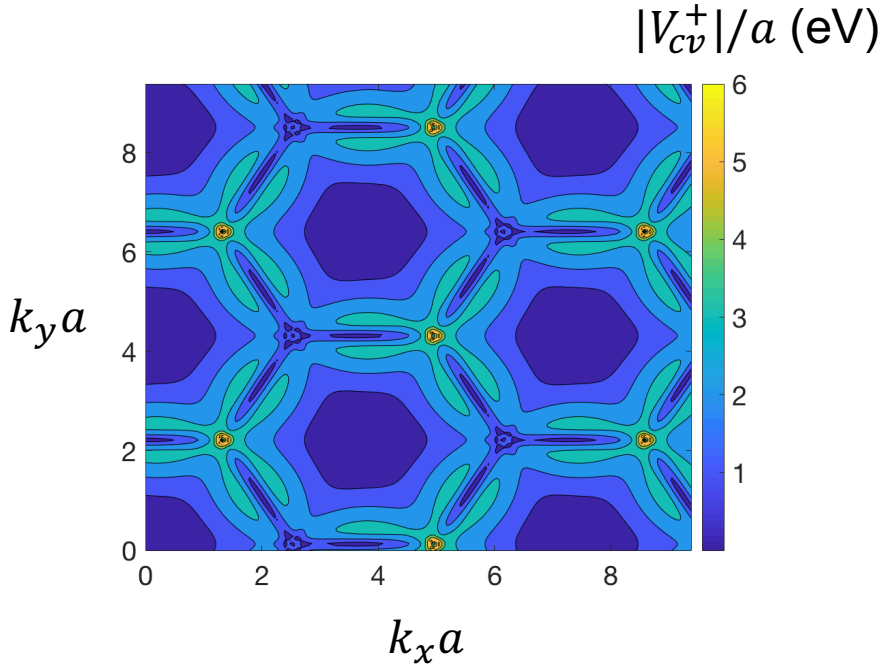

Figure 4.5: The distribution of $|V_{vc}^{+}|/a$ of biased graphene bilayers in an extended Brillouin zone with σ_+ polarization using the tight binding model. The distribution of $|V_{vc}^{-}|$ is simply a time reversal counterpart. The major contribution to $|V_{vc}^{\pm}|$ comes from the K and K' point.

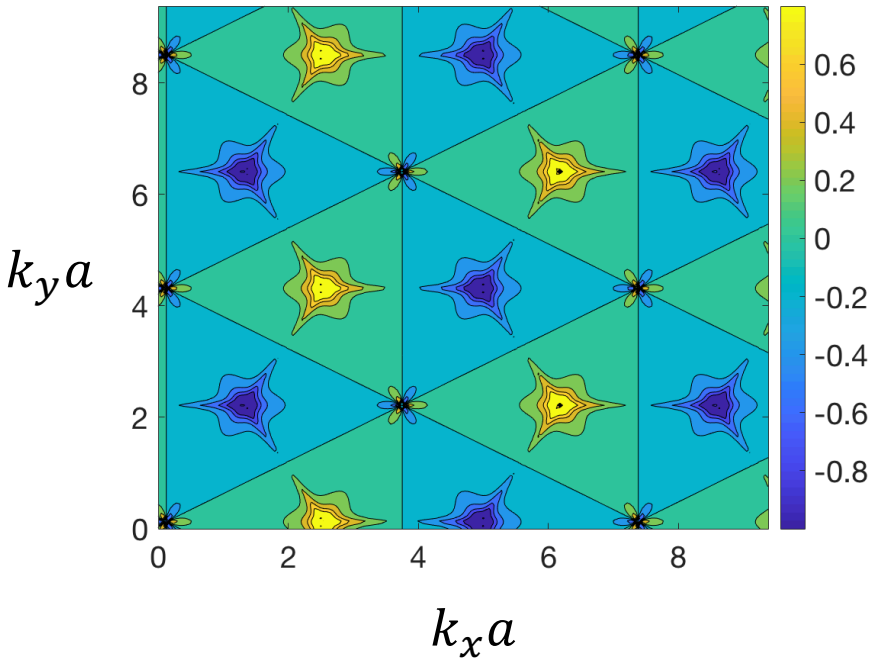

Figure 4.6: The distribution of degree of polarization DOP of biased graphene bilayers in an extended Brillouin zone using the tight binding model.

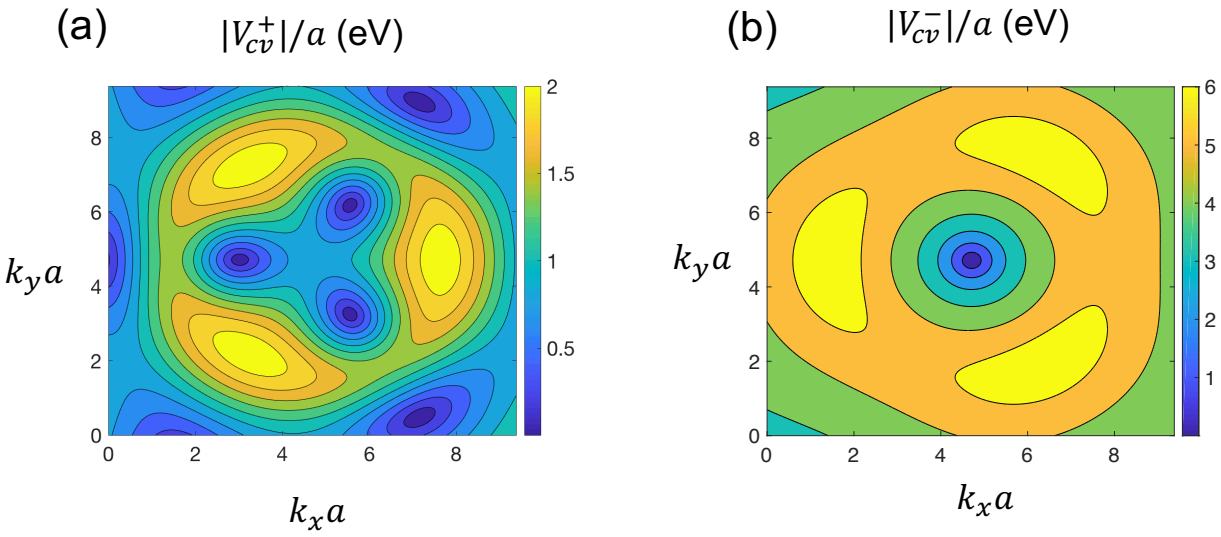

Figure 4.7: $|V_{vc}^{\eta}|/a$ of biased graphene bilayers near the K point for σ_+ (a) and σ^- (b) polarization.

function transforms as:

$$\begin{aligned} C_3\psi_{cK}(r) =& \sum_{\boldsymbol{R}_i} e^{i\boldsymbol{K}\cdot(\boldsymbol{R}_i+\tilde{\boldsymbol{R}})} C_3\phi(\boldsymbol{r}-(\boldsymbol{R}_i+\tilde{\boldsymbol{R}})) \\ =& \sum_{\boldsymbol{R}_i} e^{i\boldsymbol{K}\cdot(\boldsymbol{R}_i+\tilde{\boldsymbol{R}})} \phi(C_3^{-1}\boldsymbol{r}-(\boldsymbol{R}_i+\tilde{\boldsymbol{R}})) \\ =& \sum_{\boldsymbol{R}_i} e^{i\boldsymbol{K}\cdot(\boldsymbol{R}_i+\tilde{\boldsymbol{R}})} \phi(\boldsymbol{r}-C_3(\boldsymbol{R}_i+\tilde{\boldsymbol{R}}))e^{i\alpha} \\ =& e^{i(C_3\boldsymbol{K}-\boldsymbol{K})\cdot\tilde{\boldsymbol{R}}} e^{i\alpha}\psi_{cK}(r). \end{aligned} \tag{4.40}$$

Here α is the phase contributed by the atomic wave function under a three-fold rotation. We have $\alpha = 0$ for the p_z orbital. Therefore, the wave function gets an additional phase factor $e^{i(C_3\boldsymbol{K}-\boldsymbol{K})\cdot\tilde{\boldsymbol{R}}} = e^{i\boldsymbol{G}\cdot\tilde{\boldsymbol{R}}}$, where $\boldsymbol{G}$ is the reciprocal lattice vector shown in Fig. 4.3. Using $R_{B1(2)} = (0, \pm a)$, $K(K') = (\mp\frac{4\pi}{3\sqrt{3}a}, 0)$ and $\boldsymbol{G} = (\frac{2\sqrt{3}\pi}{3a}, \frac{2\pi}{3a})$, the phase change for ψ_{B_1} and ψ_{B_2} is $\pm\frac{2\pi}{3}$ at the K point and $\mp\frac{2\pi}{3}$ at K'. If we concentrate on the K valley, the B_1 and B_2 bands are in the E and E^* representation.

The optical selection rule of the s-exciton is determined by the velocity matrix $\langle B_2|\boldsymbol{v}|B_1\rangle$. Therefore, the s-state is bright if the direct product $(E^*)^* \bigotimes \Gamma^v \bigotimes E = E \bigotimes \Gamma^v \bigotimes E$ contains the identity IR, where Γ^v is the IR of the velocity operator. Since $\boldsymbol{v}$ should transform as a vector, it is in the E (E^*) representation, if the circular polarization is σ_+ (σ_-). It is straightforward to show that $\Gamma^v = E$ gives the identity IR, thus the s-exciton is bright at the K valley for σ_+ polarization, and vanishes for σ_- polarization.

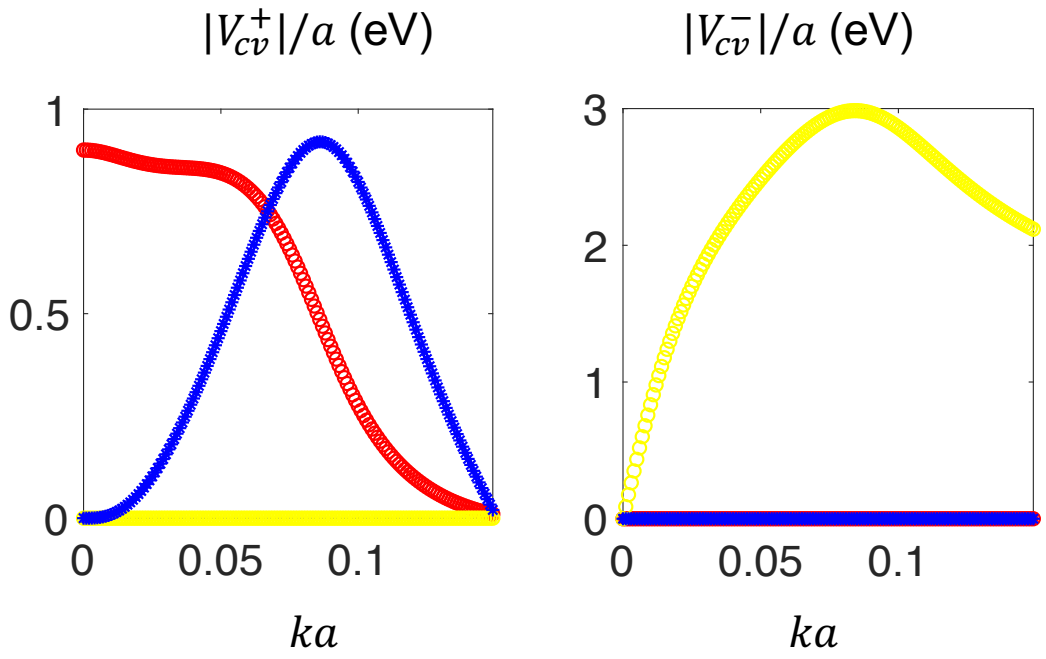

Figure 4.8: $|V_{vc}^{\eta}|/a$ of biased graphene bilayers for different angular momentum channels with σ_+ (left figure) and σ_- (right figure) polarization. The red line represents the $m = 0$ channel, the yellow line represents $m = -1$ and the blue line represents $m = -3$. The s-state and the $m = -3$ state are bright, with the opposite circular polarization to the bright p-state.

4.5.3 Topological Crystalline Insulators

The fact that it is the C_3 symmetry that turns the s-like excitons bright in a $w = 2$ CF system suggests that by switching to a different rotational symmetry, the s-states can remain dark. One such system is the gapped surface states of a topological crystalline insulator with a possible C_4 rotational symmetry [122]. The tight bonding Hamiltonian originates from the linear combination of p_x and p_y orbitals on two sublattices. Two orbitals and two sublattices leads to a tedious 4×4 bulk Hamiltonian, which will not be listed here. On $(0, 0, 1)$ surface where the C_4 symmetry is preserved, there are two edge states crossing the band gap, which are degenerate at $\bar{M}$ point. Projecting the bulk Hamiltonian to the p_x and p_y surface states, the Hamiltonian for the surface states is given by

$$H_s = \frac{k^2}{2m_0} I_{2\times2} + \frac{k_x^2 - k_y^2}{2m_1}\sigma_z + \frac{k_x k_y}{2m_2}\sigma_x \,. \tag{4.41}$$

We can transform the basis from $p_{x,y}$ into $p_\pm$ by applying a unitary transformation

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}. \tag{4.42}$$

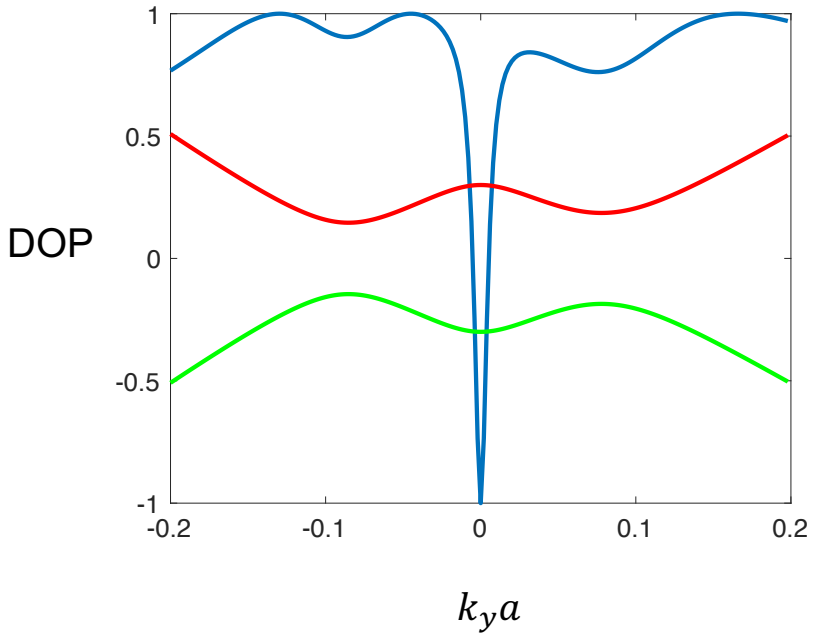

Figure 4.9: Degree of circular polarization (blue line) of biased graphene bilayers near the K valley along k_y direction. Conduction band (red) and valence band (green) are shown as a reference. The sudden sign change of DOP near the K valley is unseen when γ_3 is absent.

Ignoring the $I_{2\times 2}$ term, the surface state Hamiltonian Eq. (4.41) can be conveniently written in the following form

$$H_s' = a_1 \begin{pmatrix} V_z & k_+^2 \\ k_-^2 & -V_z \end{pmatrix} + a_2 \begin{pmatrix} V_z & k_-^2 \\ k_+^2 & -V_z \end{pmatrix}. \tag{4.43}$$

where V_z is the gap opened by a time-reversal breaking perturbation [122]. This model is a mixture of CFs with $w = \pm 2$. The simultaneous existence of both winding numbers reduces the rotational symmetry to C_4, and the s-states remain dark.

Now we check the optical selection rule by numerical calculations. First consider the case $a_1 = 1$ and $a_2 = 0$, where the Hamiltonian retreats to that for the CF with winding number $w = 2$. As shown in Fig. 4.10, the velocity matrix is nonzero for $m = 1, 3$ while vanishes for $m = 0$, which agrees with the general optical selection rule for CF models. Note that the values of velocity matrices in the right figures are very small ($\sim 10^{-17}$) which lead to the messy data points.

For the completely mixed case where $a_1 = a_2 = 0.5$, the velocity matrices in different momentum channels for both $\sigma_\pm$ polarizations are shown in Fig. 4.11. The red lines for the s-channel always vanish for both polarizations thus the s-state is dark. This result also agrees with our earlier discussion.

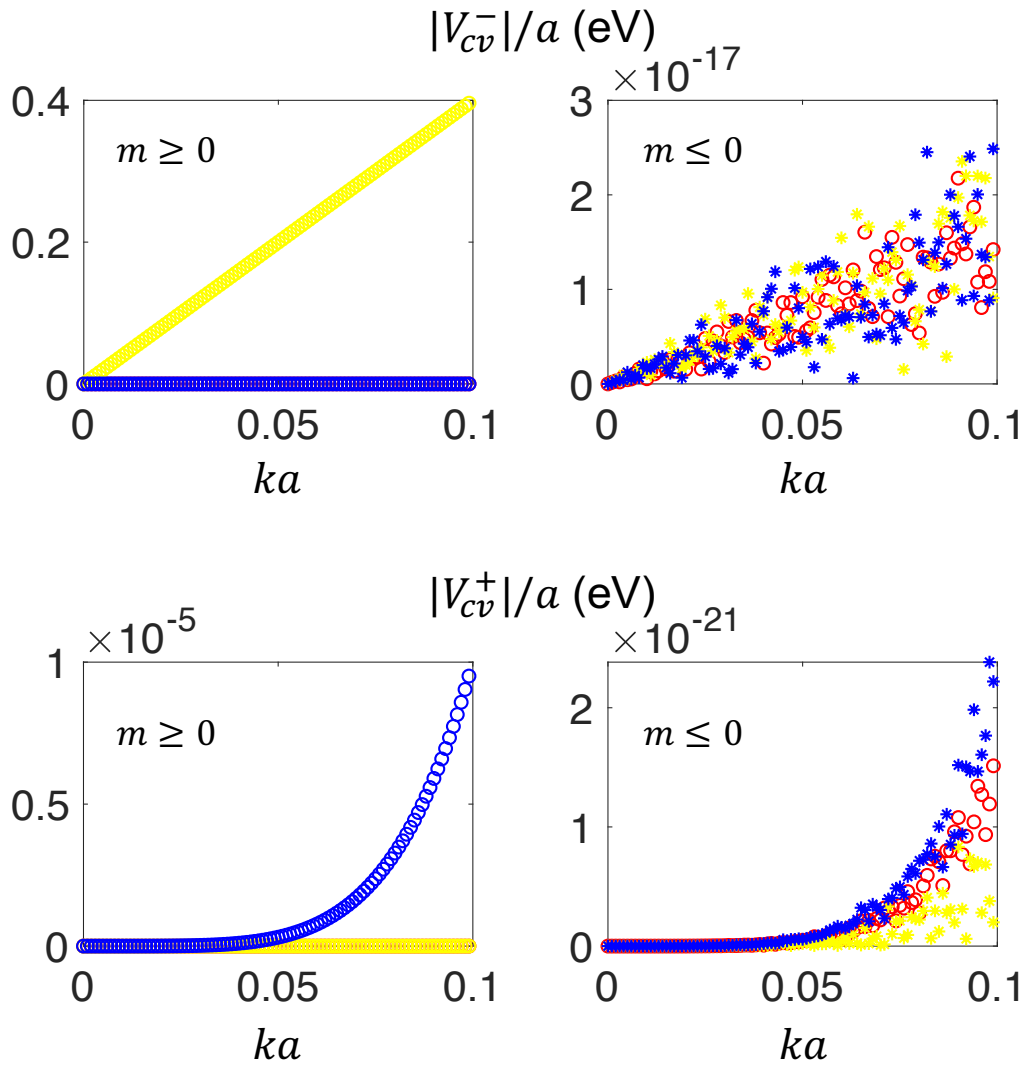

Figure 4.10: Velocity matrix $|V_{vc}^{\eta}|/a$ for TCI when $a_1 = 1$ and $a_2 = 0$. Up (down) figures corresponds to $\sigma_-(\sigma_+)$ polarization, while left (right) figures are for positive (negative) angular momentum channel. Red lines are m=0, and yellow lines are $m = \pm 1$ (positive sign for left figure, negative for right figure), bule lines are $m = \pm 3$ (sign choice is the same as yellow line).

4.5.4 3R-Stacked MoS_2 Bilayer

Apart from varying the symmetry group, we can also obtain dark s-states by switching to a different winding number while keeping the C_3 symmetry. For this purpose, let us consider $3R$-stacked MoS_2 bilayers. In the $3R$-stacked bilayer structure, the top layer is shifted relative to the bottom layer along the honeycomb armchair edge, as shown in Fig. 4.12. Neglecting the spin degree of freedom, the effective Hamiltonian

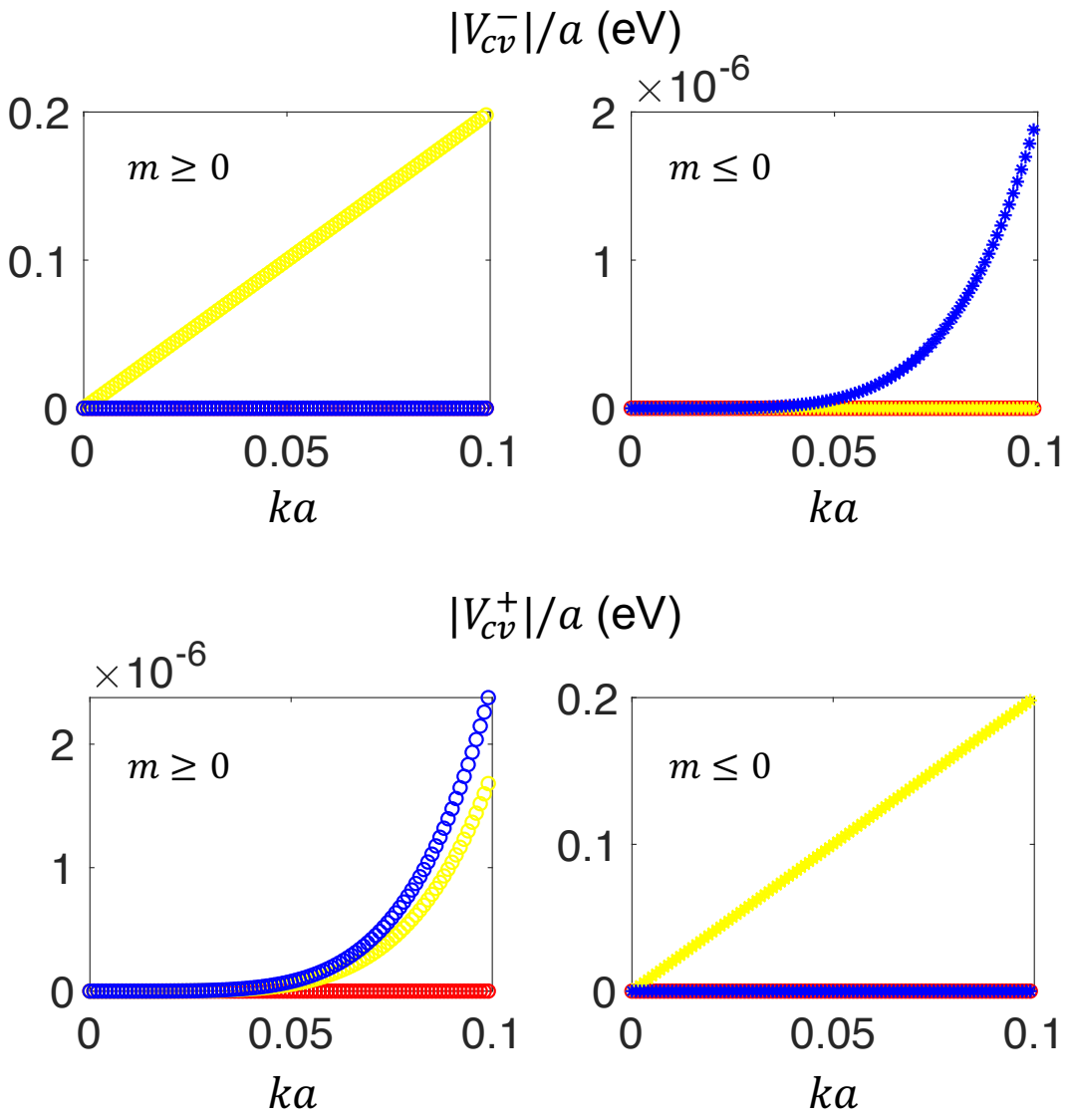

Figure 4.11: Velocity matrix $|V_{vc}^{\eta}|/a$ for TCI when $a_1 = 0.5$ and $a_2 = 0.5$. See the captions for figure 4.10 for the meaning of these lines.

at one of the corners of the hexagonal Brillouin zone is given by (see Appendix E)

$$H_{3\mathrm{R}} = \begin{pmatrix} \Delta_I + V_g & v_0 k_- & 0 & 0 \\ v_0 k_+ & -\Delta_I + V_g & \gamma_1 & 0 \\ 0 & \gamma_1 & \Delta_I - V_g & v_0 k_- \\ 0 & 0 & v_0 k_+ & -\Delta_I - V_g \end{pmatrix}, \tag{4.44}$$

where Δ_I is the gap opened by the broken inversion symmetry in each monolayer, V_g is the out of plane gate voltage, and v_0 and γ_1 are the intralayer and interlayer hopping

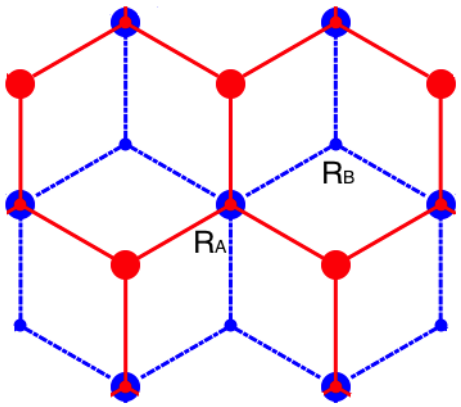

Figure 4.12: Top view of $3R$-stacked MoS_2 bilayers. The large dots are Mo atoms and the small ones are S atoms. Red (blue) dots refer to the atoms in layer 1 (2).

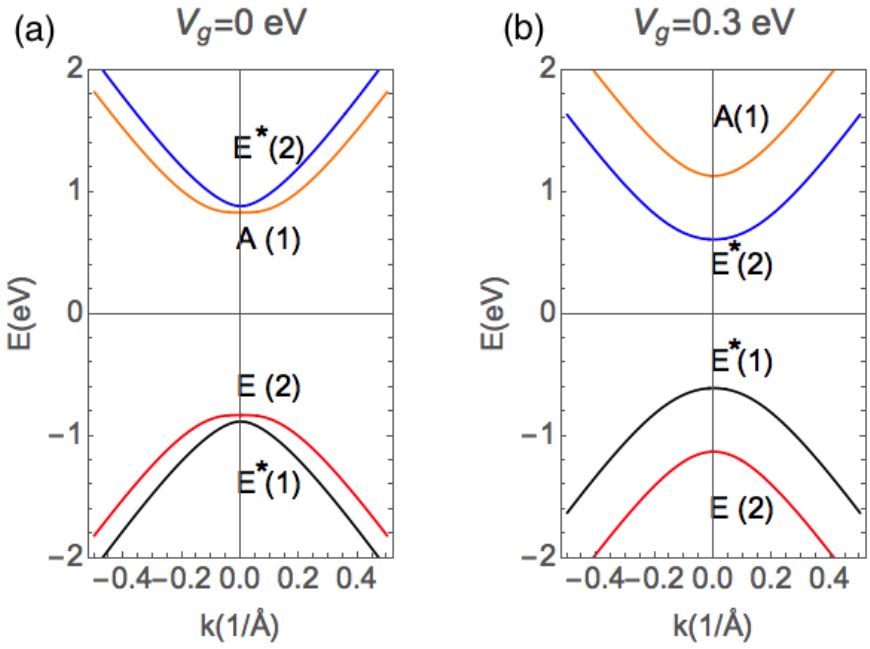

Figure 4.13: Band structure of a biased $3R$-MoS_2 bilayer at (a) $V_g = 0$ eV and (b) $V_g = 0.3$ eV. Bands with different colors belong to different irreducible representations of C_3 group and layer number $(1, 2)$. The parameters used are $\Delta_I = 0.83$ eV and $v_0 = 3.5$ eV Å. We used a large interlayer hopping term, $\gamma_1 = 0.3$ eV, to make the band separation visible.

coefficients, respectively. We have only kept the isotropic part of the Hamiltonian, which is sufficient to demonstrate the essential physics. An interesting feature of this system is that by varying V_g one can switch the band order within the conduction and valence bands (see Fig. 4.13). If we assume that Δ_I is large compared to the interlayer hopping constant γ_1, the critical value of the gate voltage V_{gc} at the band crossing point is approximately $\gamma_1^2/2\Delta_I$. For intralayer band gap $\Delta_I \approx 0.8$ eV, and interlayer hopping $\gamma_1 \approx 0.05$ eV [2], the required V_{gc} is about 1.5 meV which is not difficult to achieve in experiment [123].

The ability to switch the bands is important because the winding number is a topological quantity, it can only be changed via band crossing if the rotational symmetry is kept invariant. To find the winding number before and after the band crossing we downfold the Hamiltonian (4.44) to project out the higher conduction band and the lower valence band [124] (see Appendix F for details). Before the band crossing, i.e.,

[2] These data are the hopping constants for 2H-stacked MoS_2 bilayers [123], as an estimate for 3R stacking.

$V_g < V_{gc}$, the downfolded Hamiltonian reads

$$H_{\text{before}} = \begin{pmatrix} \Delta_I + V_g & \frac{v_0^2\gamma_1}{4\Delta_I V_g}k_-^2 \\ \frac{v_0^2\gamma_1}{4\Delta_I V_g}k_+^2 & -\Delta_I - V_g \end{pmatrix} . \tag{4.45}$$

We can see that the winding number $w = 2$, similar to the biased bilayer graphene. This is not surprising because each monolayer MoS_2 carries winding number $w = 1$, and in the $3R$-stacking, one can simply add the winding numbers together [16]. In this case, the s-like exciton is bright in the presence of the C_3 symmetry. After the band crossing, i.e., $V_g > V_{gc}$, the 2×2 Hamiltonian is

$$H_{\text{after}} = \begin{pmatrix} -\sqrt{\gamma_1^2 + (\Delta_I - V_g)^2} & -\frac{v_0^2\gamma_1(\Delta_I+V_g)}{4(\Delta_I-V_g)\Delta_I V_g}k^2 \\ -\frac{v_0^2\gamma_1(\Delta_I+V_g)}{4(\Delta_I-V_g)\Delta_I V_g}k^2 & \sqrt{\gamma_1^2 + (\Delta_I - V_g)^2} \end{pmatrix} . \tag{4.46}$$

Clearly the winding number is changed to $w = 0$. Hence $m = \pm 1$ states become bright. Turning on C_3 symmetry makes $m = \pm 1, \pm 4, \ldots$ states bright, but the s-states remain dark.

Up to now, we have omitted the valley degree of freedom, which exists in most chiral fermion systems such as graphene and MoS_2 monolayers. Different valleys carry opposite winding number as a result of the time-reversal symmetry. The corresponding optical transitions therefore have opposite circular polarization. However, intervalley coupling of exciton states via the same circularly polarized light is unlikely since the bright exciton states in the two valleys usually have different energies (for the same circular polarization).

4.6 Summary

In conclusion, we have established a new optical selection rule of excitons in gapped CF system. We found that the angular momentum of bright excitons are $w \pm 1$ in the isotropic cases, and the circular polarizations of these two states are opposite. When the crystal field effect is taken into account, the optically bright excitons have angular momentum $(w \pm 1) + nN$ if the system has N-fold rotational symmetry. We showed that by proper combinations of the winding number and rotational symmetry, one can engineer dark s-states in CF systems. The occurrence of dark excitons has already been under intense experimental investigation [125–128]. Such dark exciton has prolonged lifetime [129], and can be utilized to realize exciton condensation [130–132] and implement quantum information protocols [133, 134].

Note added—Upon the completion of this work, we have become aware of a recent paper, Ref. [135], which also studied the exciton optical selection rule in graphene systems.

Appendices

Appendix A

Linear Response Theory for the Thermal Hall Conductance

A.1 Energy Current Operator

The energy current operator can be derived from the continuity equation

$$\dot{H}(\boldsymbol{x}) + \nabla \cdot \boldsymbol{j}_E(\boldsymbol{x}) = 0 \ , \tag{A.1}$$

where

$$H(\boldsymbol{x}) = \sum_i \Delta(\boldsymbol{x} - \boldsymbol{R}_i) H_i \ . \tag{A.2}$$

H_i is defined by

$$H_i = H_{i,spin} + H_{i,ph} + H_{i,int} \ , \tag{A.3}$$

with each term defined in the main text. $\dot{H}(\boldsymbol{x})$ is given by the Heisenberg equation

$$\dot{H}(\boldsymbol{x}) = \frac{1}{i\hbar}[H(\boldsymbol{x}), H] \ , \tag{A.4}$$

which contains the contribution from the commutators between H_{is}, $H_{i,int}$ and $H_{i,ph}$. I will use the $\dot{H}_{spin}(\boldsymbol{x}) \equiv \sum_{i,j} \Delta(\boldsymbol{x} - \boldsymbol{R}_i)[H_{i,spin}, H_{j,spin}]/(i\hbar)$ term as an example, and other terms just follow similar calculations.

For the $\dot{H}_{spin}(\boldsymbol{x})$ term, we can ignore the contribution from the DM term, which does not appear in the linear spin wave Hamiltonian. Using the commutation relation between spin operators, we have

$$\sum_{j \in i} \sum_{\langle i', j' \rangle} [\boldsymbol{s}_i \cdot \boldsymbol{s}_j, \boldsymbol{s}_{i'} \cdot \boldsymbol{s}_{j'}] = -2i \sum_{j \in i} \sum_{k \in j, k \neq i} \boldsymbol{s}_i \cdot (\boldsymbol{s}_j \times \boldsymbol{s}_k) \tag{A.5}$$

and

$$[\boldsymbol{s}_i \cdot \boldsymbol{s}_j, s_i'^{z2}] = i\{\delta_{ji'}[-s_i^x(s_j^z s_j^y + s_j^y s_j^z) + s_i^y(s_j^z s_j^x + s_j^x s_j^z)] + \delta_{ii'}[-i(s_i^z s_i^y + s_i^y s_i^z)s_j^x + i(s_i^z s_i^x + s_i^x s_i^z)s_j^y]\} \tag{A.6}$$

Using the commutation relation derived above and the Holstein-Primakoff transformation, the Heisenberg equation becomes

$$\begin{aligned}\dot{H}_{spin}(\boldsymbol{x}) =&\frac{1}{i\hbar}\sum_i \Delta(\boldsymbol{x}-\boldsymbol{R}_i)\{-i\frac{J^2}{2}\sum_{j\in i}\sum_{k\in j,k\neq i}\boldsymbol{s}_i\cdot(\boldsymbol{s}_j\times\boldsymbol{s}_k)\\ &+\frac{JK}{4}\sum_{j\in i}\sum_{i'}[\boldsymbol{s}_i\cdot\boldsymbol{s}_j,s_{i'}^2]+\frac{JK}{4}\sum_{\langle i',j'\rangle}[s_i^2,\boldsymbol{s}_{i'}\cdot\boldsymbol{s}_{j'}]\}\\ =&\frac{1}{i\hbar}\sum_i \Delta(\boldsymbol{x}-\boldsymbol{R}_i)\{-\frac{J^2S_0^2}{2}\sum_{j\in i}\sum_{k\in j,k\neq i}(a_j^\dagger a_k-a_k^\dagger a_j+a_k^\dagger a_i-a_i^\dagger a_k+a_i^\dagger a_j-a_j^\dagger a_i)\\ &+\frac{JK}{4}\sum_{j\in i}\sum_{i'}[\boldsymbol{s}_i\cdot\boldsymbol{s}_j,s_{i'}^2]+\frac{JK}{4}\sum_{\langle i',j'\rangle}[s_i^2,\boldsymbol{s}_{i'}\cdot\boldsymbol{s}_{j'}]\}\\ =&A+B\ ,\end{aligned} \tag{A.7}$$

where A is contributed by only the exchange term, and B is proportional to K. We consider these two terms separately.

We use the anti-symmetric format of the operators in the A term:

$$\begin{aligned}A=&\frac{iJ^2S_0^2}{4\hbar}\sum_i\sum_{j\in i}\sum_{k\in j,k\neq i}\{[\Delta(\boldsymbol{x}-\boldsymbol{R}_i)-\Delta(\boldsymbol{x}-\boldsymbol{R}_k)](a_k^\dagger a_i-a_i^\dagger a_k)\\ &+[\Delta(\boldsymbol{x}-\boldsymbol{R}_i)-\Delta(\boldsymbol{x}-\boldsymbol{R}_j)](a_i^\dagger a_j-a_j^\dagger a_i)\\ &+2[\Delta(\boldsymbol{x}-\boldsymbol{R}_j)+(\boldsymbol{R}_i-\boldsymbol{R}_j)\cdot\nabla\Delta(\boldsymbol{x}-\boldsymbol{R}_j)](a_j^\dagger a_k-a_k^\dagger a_j)\}\ .\end{aligned} \tag{A.8}$$

In the last line, suppose there is no sublattice degree of freedom, the $(\boldsymbol{R}_i-\boldsymbol{R}_j)\cdot\nabla\Delta(\boldsymbol{x}-\boldsymbol{R}_j)$ term vanishes after summing over the i index. In addition, suppose the envelope function $\Delta(x)$ is smooth enough, thus we can do a Taylor expansion $\Delta(\boldsymbol{x}-\boldsymbol{R}_i)-\Delta(\boldsymbol{x}-\boldsymbol{R}_k)=(\boldsymbol{R}_i-\boldsymbol{R}_k)\cdot\nabla(\Delta(\boldsymbol{x}-\boldsymbol{R}_i)+\Delta(\boldsymbol{x}-\boldsymbol{R}_k))/2$, and the current contributed by the A term is given by

$$\begin{aligned}\boldsymbol{j}_A(\boldsymbol{x})=&-\frac{iJ^2S_0^2}{8\hbar}\sum_i\sum_{j\in i}\sum_{k\in j,k\neq i}\{(\boldsymbol{R}_i-\boldsymbol{R}_k)[\Delta(\boldsymbol{x}-\boldsymbol{R}_i)+\Delta(\boldsymbol{x}-\boldsymbol{R}_k)](a_k^\dagger a_i-a_i^\dagger a_k)\\ &+(\boldsymbol{R}_i-\boldsymbol{R}_j)[\Delta(\boldsymbol{x}-\boldsymbol{R}_i)+\Delta(\boldsymbol{x}-\boldsymbol{R}_j)](a_i^\dagger a_j-a_j^\dagger a_i)\\ &+(\boldsymbol{R}_j-\boldsymbol{R}_k)[\Delta(\boldsymbol{x}-\boldsymbol{R}_j)+\Delta(\boldsymbol{x}-\boldsymbol{R}_k)](a_j^\dagger a_k-a_k^\dagger a_j)\}\ .\end{aligned} \tag{A.9}$$

Now we apply a Fourier transformation $\boldsymbol{j}_A(\boldsymbol{q})=\int dx e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}\boldsymbol{j}(\boldsymbol{x})$. It is easy to show

that

$$\begin{aligned}\boldsymbol{j}_A(\boldsymbol{q}) =& \frac{J^2 S_0^2}{8\hbar}\Delta(\boldsymbol{q})[-\sum_{\boldsymbol{q}_1}\sum_{\delta_1,\delta_2}\partial_{\boldsymbol{q}_1}(e^{i\boldsymbol{q}_1\cdot(\boldsymbol{\delta}_1+\boldsymbol{\delta}_2)}+e^{-i\boldsymbol{q}_1\cdot(\boldsymbol{\delta}_1+\boldsymbol{\delta}_2)})(a^\dagger_{\boldsymbol{q}_1-\boldsymbol{q}}a_{\boldsymbol{q}_1}+a^\dagger_{\boldsymbol{q}_1}a_{\boldsymbol{q}_1+\boldsymbol{q}}) \\ &+2N\sum_{\boldsymbol{q}_1}\sum_{\delta}\partial_{\boldsymbol{q}_1}(e^{i\boldsymbol{q}_1\cdot\delta}+e^{-i\boldsymbol{q}_1\cdot\boldsymbol{\delta}})(a^\dagger_{\boldsymbol{q}_1-\boldsymbol{q}}a_{\boldsymbol{q}_1}+a^\dagger_{\boldsymbol{q}_1}a_{\boldsymbol{q}_1+\boldsymbol{q}})]\ ,\end{aligned} \tag{A.10}$$

where δ, δ_1, δ_2 are nearest neighbor vectors, and N is the number of the nearest neighbors.

Now consider the B term. Using the commutation relation, we get

$$\begin{aligned}B =& -\frac{JKS_0(2S_0-1)}{2i\hbar}\sum_i\Delta(\boldsymbol{x}-\boldsymbol{R}_i)\sum_{j\in i}(a_i^\dagger a_j-a_j^\dagger a_i) \\ =& -\frac{JKS_0(2S_0-1)}{4i\hbar}\sum_i\sum_{j\in i}[\Delta(\boldsymbol{x}-\boldsymbol{R}_i)-\Delta(\boldsymbol{x}-\boldsymbol{R}_j)](a_i^\dagger a_j-a_j^\dagger a_i) \\ =& -\frac{JKS_0(2S_0-1)}{8i\hbar}\sum_i\sum_{j\in i}(\boldsymbol{R}_i-\boldsymbol{R}_j)\cdot\nabla(\Delta(\boldsymbol{x}-\boldsymbol{R}_i)+\Delta(\boldsymbol{x}-\boldsymbol{R}_j))(a_i^\dagger a_j-a_j^\dagger a_i)\ .\end{aligned} \tag{A.11}$$

Therefore, the corresponding current from the B term is given by

$$\boldsymbol{j}_B(\boldsymbol{x}) = \frac{JKS_0(2S_0-1)}{8i\hbar}\sum_i\sum_{j\in i}(\boldsymbol{R}_i-\boldsymbol{R}_j)(\Delta(\boldsymbol{x}-\boldsymbol{R}_i)+\Delta(\boldsymbol{x}-\boldsymbol{R}_j))(a_i^\dagger a_j-a_j^\dagger a_i)\ . \tag{A.12}$$

Fourier transformation gives

$$\boldsymbol{j}_B(\boldsymbol{q}) = \frac{JKS_0(2S_0-1)}{8\hbar}\Delta(\boldsymbol{q})\sum_{\boldsymbol{q}_1}\sum_{\delta}\partial_{\boldsymbol{q}_1}(e^{i\boldsymbol{q}_1\cdot\delta}+e^{-i\boldsymbol{q}_1\cdot\boldsymbol{\delta}})(a^\dagger_{\boldsymbol{q}_1-\boldsymbol{q}}a_{\boldsymbol{q}_1}+a^\dagger_{\boldsymbol{q}_1}a_{\boldsymbol{q}_1+\boldsymbol{q}}) \tag{A.13}$$

Combining the contributions from A and B terms together, the energy current that contributes to $\dot{H}_{spin}$ is given by

$$\boldsymbol{j}_{E,spin}(\boldsymbol{q}) = \frac{\Delta(\boldsymbol{q})}{4}\sum_{\boldsymbol{q}_1}(\tilde{\boldsymbol{V}}^m_{\boldsymbol{q}}\tilde{H}^m_{\boldsymbol{q}}+H.c.)(a^\dagger_{\boldsymbol{q}_1-\boldsymbol{q}}a_{\boldsymbol{q}_1}+a^\dagger_{\boldsymbol{q}_1}a_{\boldsymbol{q}_1+\boldsymbol{q}})\ , \tag{A.14}$$

where $\tilde{H}^m_{\boldsymbol{q}} = S_0J(N-\sum_\delta e^{i\boldsymbol{q}_1\cdot\delta})+K(2S_0-1)/2$ is the magnon energy, and $\tilde{\boldsymbol{V}}^m_{\boldsymbol{q}} = \partial\tilde{H}^m_{\boldsymbol{q}}/\partial\boldsymbol{q}$ is the velocity of magnons.

Other commutators can be calculated using similar procedures, while we only need to notice that the interaction part need to take the symmetric format given in Eq. (3.24), rather than a more compact format $H_{i,int} = \sum_{j\in i}\sum_{\alpha,\beta}u_i^\alpha T_{ij}^{\alpha\beta}(\delta s_i^\beta - \delta s_j^\beta)$. Although both definitions of $H_{i,int}$ leads to the same total Hamiltonian, their corresponding energy current operators are slightly different. Combining all the commutators together, we get the total energy current operator for the magnon-phonon hybrid

excitation

$$\boldsymbol{J}^E(\boldsymbol{q}) = \frac{\Delta(\boldsymbol{q})}{8\hbar}\sum_{\boldsymbol{q}_1}\hat{\psi}^\dagger_{\boldsymbol{q}_1}\eta(\boldsymbol{v}_{\boldsymbol{q}_1}\tilde{H}_{\boldsymbol{q}_1}+\tilde{H}_{\boldsymbol{q}_1}\boldsymbol{v}_{\boldsymbol{q}_1}+\boldsymbol{v}_{\boldsymbol{q}_1+\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}+\tilde{H}_{\boldsymbol{q}_1+\boldsymbol{q}}\boldsymbol{v}_{\boldsymbol{q}_1+\boldsymbol{q}})\hat{\psi}_{\boldsymbol{q}_1+\boldsymbol{q}}\ , \quad \text{(A.15)}$$

where $v_{\boldsymbol{q}} = \partial_{\boldsymbol{q}}\tilde{H}_{\boldsymbol{q}}$ is the velocity matrix, $\tilde{H}_{\boldsymbol{q}}$ and η are the Hamiltonian matrix and the commutator matrix, respectively, and are given in the main text.

A.2 Kubo Term

The Kubo term in the thermal Hall conductivity is given by

$$\sigma^{xy,kubo} = \frac{1}{\hbar V T_0}\int_0^\infty dt e^{-\epsilon t}\int_0^{\beta\hbar} d\lambda\langle J^y_E(0,-i\lambda-t)J^x_E(0,0)\rangle\ . \quad \text{(A.16)}$$

Now we calculate the correlation functions

$$\begin{aligned}
&\langle J^E_y(0,-i\lambda-t)J^E_x(0,0)\rangle\\
=&\frac{1}{64\hbar^2}\sum_{m,n,m',n',\boldsymbol{q}_1,\boldsymbol{q}_1'}\{\Gamma^{(1)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(4)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}^\dagger_{n,\boldsymbol{q}_1}\hat{\Psi}_{m,\boldsymbol{q}_1}\hat{\Psi}_{n',-\boldsymbol{q}_1'}\hat{\Psi}^\dagger_{m',-\boldsymbol{q}_1'}\rangle\\
&+\Gamma^{(4)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(1)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}_{n,-\boldsymbol{q}_1}\hat{\Psi}^\dagger_{m,-\boldsymbol{q}_1}\hat{\Psi}^\dagger_{n',\boldsymbol{q}_1'}\hat{\Psi}_{m',\boldsymbol{q}_1'}\rangle\\
&+\Gamma^{(2)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(3)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}^\dagger_{n,\boldsymbol{q}_1}\hat{\Psi}^\dagger_{m,-\boldsymbol{q}_1}\hat{\Psi}_{n',-\boldsymbol{q}_1'}\hat{\Psi}_{m',\boldsymbol{q}_1'}\rangle\\
&+\Gamma^{(3)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(2)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}_{n,-\boldsymbol{q}_1}\hat{\Psi}_{m,\boldsymbol{q}_1}\hat{\Psi}^\dagger_{n',\boldsymbol{q}_1'}\hat{\Psi}^\dagger_{m',-\boldsymbol{q}_1'}\rangle\\
&+\Gamma^{(1)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(1)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}^\dagger_{n,\boldsymbol{q}_1}\hat{\Psi}_{m,\boldsymbol{q}_1}\hat{\Psi}^\dagger_{n',\boldsymbol{q}_1'}\hat{\Psi}_{m',\boldsymbol{q}_1'}\rangle\\
&+\Gamma^{(4)y}_{mn}(\boldsymbol{q}_1,0,-i\lambda-t)\Gamma^{(4)x}_{m'n'}(\boldsymbol{q}_1',0,0)\langle\hat{\Psi}_{n,-\boldsymbol{q}_1}\hat{\Psi}^\dagger_{m,-\boldsymbol{q}_1}\hat{\Psi}_{n',-\boldsymbol{q}_1'}\hat{\Psi}^\dagger_{m',-\boldsymbol{q}_1'}\rangle\}
\end{aligned} \quad \text{(A.17)}$$

We take the third line as an example. The correlation function gives

$$\langle\hat{\Psi}^\dagger_{n,\boldsymbol{q}_1}\hat{\Psi}^\dagger_{m,-\boldsymbol{q}_1}\hat{\Psi}_{n',-\boldsymbol{q}_1'}\hat{\Psi}_{m',\boldsymbol{q}_1'}\rangle = (\delta_{nn'}\delta_{mm'}\delta_{\boldsymbol{q}_1,-\boldsymbol{q}_1'}+\delta_{mn'}\delta_{nm'}\delta_{\boldsymbol{q}_1,\boldsymbol{q}_1'})n_{n,\boldsymbol{q}_1}n_{m,-\boldsymbol{q}_1}\ , \quad \text{(A.18)}$$

where $n_{n,\boldsymbol{q}_1}$ is the Bose-Einstein distribution function with zero chemical potential. The integral of the real and imaginary time gives

$$\begin{aligned}
&\int_0^\infty dt e^{-\epsilon t}\int_0^{\beta\hbar} d\lambda e^{i(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})(-i\lambda-t)/\hbar}n_{n,\boldsymbol{q}_1}n_{m,-\boldsymbol{q}_1}\\
=&\frac{e^{\beta(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})}-1}{i(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})^2/\hbar^2}\frac{1}{e^{\beta E_{n,\boldsymbol{q}_1}}-1}\frac{1}{e^{\beta E_{m,-\boldsymbol{q}_1}}-1}\\
=&\frac{1+n_{n,\boldsymbol{q}_1}+n_{m,-\boldsymbol{q}_1}}{i(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})^2/\hbar^2}\ .
\end{aligned} \quad \text{(A.19)}$$

Therefore, the contribution from the third line of Eq. (A.17) is

$$\sigma_3^{xy,kubo} = \sum_{m,n>0,\boldsymbol{q}_1} \frac{1+n_{n,\boldsymbol{q}_1}+n_{m,-\boldsymbol{q}_1}}{16iVT_0\hbar(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})^2} E_{n\boldsymbol{q}_1}E_{m-\boldsymbol{q}_1}\psi^\dagger_{n,\boldsymbol{q}_1}\eta\partial_{q_1^y}\tilde{H}^2_{\boldsymbol{q}_1}U^T\psi^*_{m,-\boldsymbol{q}_1} \\ [\psi^T_{n,\boldsymbol{q}_1}U^*\eta\partial_{-q_1^x}\tilde{H}^2_{-\boldsymbol{q}_1}\psi_{m,-\boldsymbol{q}_1}+\psi^T_{m,-\boldsymbol{q}_1}U^*\eta\partial_{q_1^x}\tilde{H}^2_{\boldsymbol{q}_1}\psi_{n,\boldsymbol{q}_1}] \ . \tag{A.20}$$

Since $U^\dagger\tilde{H}_{\boldsymbol{q}}U=-\tilde{H}^*_{-\boldsymbol{q}}$, we have $U^\dagger\tilde{H}^2_{\boldsymbol{q}}U=\tilde{H}^{*2}_{-\boldsymbol{q}}$. Combined with $U\eta^*U=-\eta$, we can do the following transformation

$$\begin{aligned}\psi^T_{n,\boldsymbol{q}_1}U^*\eta\partial_{-q_1^x}\tilde{H}^2_{-\boldsymbol{q}_1}\psi_{m,-\boldsymbol{q}_1} =&(\psi^\dagger_{n,\boldsymbol{q}_1}U\eta^*\partial_{-q_1^x}\tilde{H}^{*2}_{-\boldsymbol{q}_1}\psi^*_{m,-\boldsymbol{q}_1})^* \\ =&(\psi^\dagger_{n,\boldsymbol{q}_1}U\eta^*U\partial_{-q_1^x}\tilde{H}^2_{\boldsymbol{q}_1}U\psi^*_{m,-\boldsymbol{q}_1})^* \\ =&(\psi^\dagger_{n,\boldsymbol{q}_1}\eta\partial_{q_1^x}\tilde{H}^2_{\boldsymbol{q}_1}U\psi^*_{m,-\boldsymbol{q}_1})^* \\ =&\psi^T_{m,-\boldsymbol{q}_1}U\eta\partial_{q_1^x}\tilde{H}^{*2}_{\boldsymbol{q}_1}\psi_{n,\boldsymbol{q}_1} \ .\end{aligned} \tag{A.21}$$

Note that we have used the Hermitian property of $\eta\partial_{q_1^y}\tilde{H}^2_{\boldsymbol{q}_1}$. Therefore, $\sigma_3^{xy,kubo}$ can be written as

$$\sigma_3^{xy,kubo} = \\ \sum_{m,n>0,\boldsymbol{q}_1} \frac{1+n_{n,\boldsymbol{q}_1}+n_{m,-\boldsymbol{q}_1}}{8iVT_0\hbar(E_{n\boldsymbol{q}_1}+E_{m-\boldsymbol{q}_1})^2} E_{n\boldsymbol{q}_1}E_{m-\boldsymbol{q}_1}\psi^\dagger_{n,\boldsymbol{q}_1}\eta\partial_{q_1^y}\tilde{H}^2_{\boldsymbol{q}_1}U^T\psi^*_{m,-\boldsymbol{q}_1}\psi^T_{m,-\boldsymbol{q}_1}U\eta\partial_{q_1^x}\tilde{H}^2_{\boldsymbol{q}_1}\psi_{n,\boldsymbol{q}_1} \ . \tag{A.22}$$

Other lines in Eq.(A.17) can be calculated in the similar way. Combining all the contributions in Eq.(A.17), we get the thermal conductivity from the Kubo term:

$$\sigma^{xy,kubo} = \sum_{m,n,\boldsymbol{q}_1} \frac{(\mathcal{V}^x_{\boldsymbol{q}_1,n,m}\mathcal{V}^y_{\boldsymbol{q}_1,m,n}-\mathcal{V}^y_{\boldsymbol{q}_1,n,m}\mathcal{V}^x_{\boldsymbol{q}_1,m,n})E_{n\boldsymbol{q}_1}E_{m\boldsymbol{q}_1}n_{n\boldsymbol{q}_1}}{32iVT_0\hbar(E_{n\boldsymbol{q}_1}-E_{m\boldsymbol{q}_1})^2} \ , \tag{A.23}$$

where the vertices $\mathcal{V}^\sigma_{\boldsymbol{q}_1,n,m}$ are given in the main text.

Appendix B

The Dynamic Matrix for Phonons

Phonon modes originate from classical normal modes of vibrating ions. The Hamiltonian for ions is

$$H_{\rm ph} = \sum_i \frac{\boldsymbol{p}_i^2}{2M} + V(\{\boldsymbol{R}_i\}) \ , \tag{B.1}$$

where M is the ion's mass, and $V(\{\boldsymbol{R}_i\})$ is the potential energy between ions under some ion configuration $\{\boldsymbol{R}_i\}$. We expand the potential around the equilibrium configuration $\{\boldsymbol{R}_i^0\}$ under the harmonic approximation

$$V(\{\boldsymbol{R}_i\}) \approx V(\{\boldsymbol{R}_i^0\}) + \frac{1}{2}\sum_{i,j} \boldsymbol{u}_i \cdot \frac{\partial^2 V}{\partial \boldsymbol{u}_i \partial \boldsymbol{u}_j}\Big|_{\{\boldsymbol{R}_i=\boldsymbol{R}_i^0\}} \cdot \boldsymbol{u}_j. \tag{B.2}$$

where the deviation from its equilibrium position of the i-th ion $\boldsymbol{u}_i \equiv \boldsymbol{R}_i - \boldsymbol{R}_i^0$. Here, we assume that the potential $V(\{\boldsymbol{R}_i\})$ contains only the first and second nearest neighbor interactions

$$\begin{aligned}\frac{1}{2}\sum_{i,j} \boldsymbol{u}_i \cdot \frac{\partial^2 V}{\partial \boldsymbol{u}_i \partial \boldsymbol{u}_j}\Big|_{\{\boldsymbol{R}_i=\boldsymbol{R}_i^0\}} \cdot \boldsymbol{u}_j &\approx \sum_{\langle i,j\rangle} \frac{M\Omega^2}{2}[(\boldsymbol{u}_i - \boldsymbol{u}_j)\cdot \hat{\boldsymbol{R}}_{ij}^0]^2 + \sum_{\langle\langle i,j\rangle\rangle} \frac{M\Omega'^2}{2}[(\boldsymbol{u}_i - \boldsymbol{u}_j)\cdot \hat{\boldsymbol{R}}_{ij}^0]^2 \\ &\equiv \frac{1}{2}\sum_{\langle i,j\rangle} \boldsymbol{u}_i^T \boldsymbol{\Phi}_{ij} \boldsymbol{u}_j,\end{aligned} \tag{B.3}$$

with two vibrations frequencies Ω and Ω' corresponding to two nearest neighbor atoms and two second nearest neighbor atoms, and $\boldsymbol{\Phi}_{ij}$ is the dynamic matrix in the real space. Here we only consider the vibrations along the bonds, since the vibrations perpendicular to the bonds are higher order effects. Accordingly, the phonon Hamiltonian can be written in the momentum $\boldsymbol{q}$ space

$$H_{\rm ph} = \sum_{\boldsymbol{q}} \frac{\boldsymbol{p}_{-\boldsymbol{q}}\boldsymbol{p}_{\boldsymbol{q}}}{2M} + \frac{1}{2}\boldsymbol{u}_{-\boldsymbol{q}}^T \boldsymbol{\Phi}(\boldsymbol{q}) \boldsymbol{u}_{\boldsymbol{q}} \ , \tag{B.4}$$

where the dynamic matrix in the momentum space is given by

$$\begin{aligned}\frac{\boldsymbol{\Phi}(\boldsymbol{q})}{M} =& 2\Omega'^2(1-\cos q_y a\cos q_x a+\sigma_x \sin q_x a \sin q_y a)\\ &+2\Omega^2\left[\sigma_z\left(\sin^2\frac{q_x a}{2}-\sin^2\frac{q_y a}{2}\right)+\sin^2\frac{q_x a}{2}+\sin^2\frac{q_y a}{2}\right],\end{aligned} \tag{B.5}$$

where the 2×2 matrices $\sigma_{x,y,z}$ are the Pauli matrices.

Appendix C

Effective Phonon Model in the Square Lattice

We build an effective theory just for phonons by integrating out the magnon degree of freedom. Our starting point is the equation of motion for ions [81],

$$M\ddot{u}_i^\alpha = -\sum_{j,\beta} \Phi_{ij}^{\alpha\beta} u_j^\beta + \langle F_i^\alpha \rangle \ , \tag{C.1}$$

where $F_i^\alpha \equiv -\partial H_{\rm int}/\partial u_i^\alpha$ is the effective force operator acting on phonons from the magnon-phonon interaction. The expectation value $\langle F_i^\alpha \rangle$ should be evaluated in the magnon subsystem subjected to a time-dependent perturbation from the lattice vibration $\boldsymbol{u}_i$. The bracket $\langle ... \rangle$ denotes the statistical quantum average of the spin states. Following Eq. (5) in the main text, and transforming to the momentum space, we can write the effective force operator as

$$F_{\boldsymbol{q}}^\alpha(t) = -\sum_\beta T_{\boldsymbol{q}}^{\alpha\beta} \delta s_{\boldsymbol{q}}^\beta(t), \tag{C.2}$$

where the coupling matrix $T_{\boldsymbol{q}}^{\alpha\beta} = \sum_{\boldsymbol{\delta}_j} (1 - e^{-i\boldsymbol{\delta}_j \cdot \boldsymbol{q}}) T^{\alpha\beta}(\boldsymbol{\delta}_j)$ with $\boldsymbol{\delta}_j$ the nearest neighbor vector is a symmetric matrix. Standard linear response theory in the frequency representation can be explicitly written as [81]

$$\langle \boldsymbol{F}_{\boldsymbol{q}}(\omega) \rangle = \boldsymbol{T}(\boldsymbol{q}) \boldsymbol{\chi}(\boldsymbol{q};\omega) \boldsymbol{T}(-\boldsymbol{q}) \boldsymbol{u}_{\boldsymbol{q}}(\omega) \ , \tag{C.3}$$

where $\chi^{\alpha\beta}(\boldsymbol{q};\omega) = -\frac{i}{\hbar} \int dt e^{i\omega t} \Theta(t) \langle [\delta s_{\boldsymbol{q}}^\alpha(t), \delta s_{-\boldsymbol{q}}^\beta(0)] \rangle$ is the spin-spin response function of the ferromagnetic state. Here we use the convention that a bold form such as $\boldsymbol{\chi}$, $\boldsymbol{T}$, $\boldsymbol{F}$ and $\boldsymbol{u}$ denotes a tensor, and the plain form such as $\chi^{\alpha\beta}$ denotes a tensor component.

To proceed further, let us consider the low temperature regime where only the modes with low frequencies ω are important. Therefore, we can expand $\boldsymbol{\chi}(\boldsymbol{q};\omega)$ to the first order of ω, $\boldsymbol{\chi}(\boldsymbol{q};\omega) \approx \boldsymbol{\chi}_0(\boldsymbol{q}) + i\omega \boldsymbol{\chi}_1(\boldsymbol{q})$. Inserting Eq. (C.3) back to Eq. (C.1), we obtain

$$\left[\left(i\omega \boldsymbol{I} + \frac{\boldsymbol{g}(\boldsymbol{q})}{2M} \right)^2 + \frac{\boldsymbol{\mathcal{K}}(\boldsymbol{q})}{M} \right] \boldsymbol{u}_{\boldsymbol{q}} = 0 \ , \tag{C.4}$$

where $\boldsymbol{\mathcal{K}}(\boldsymbol{q}) = \boldsymbol{\Phi}(\boldsymbol{q}) + \delta\boldsymbol{\Phi}(\boldsymbol{q}) - \boldsymbol{\mathcal{A}}^2(\boldsymbol{q})/M$, with

$$\delta\boldsymbol{\Phi}(\boldsymbol{q}) = \boldsymbol{T}(\boldsymbol{q})\boldsymbol{\chi_0}(\boldsymbol{q})\boldsymbol{T}(-\boldsymbol{q}) \tag{C.5}$$

and

$$\boldsymbol{\mathcal{A}}(\boldsymbol{q}) = 2\boldsymbol{T}(\boldsymbol{q})\boldsymbol{\chi}_1(\boldsymbol{q})\boldsymbol{T}(-\boldsymbol{q})\ . \tag{C.6}$$

Note that these two corrections $\delta\boldsymbol{\Phi}$ and $\boldsymbol{\mathcal{A}}$ are proportional to the real and imaginary part of the spin-spin response function, respectively. It is straightforward to show that Eq. (C.4) can be simply derived from the effective Hamiltonian (Eq. (10)) in the main text. Therefore, Re$\boldsymbol{\chi}(\boldsymbol{q};\omega)$ provides a screening effect to the inter-atomic interaction, and Im$\boldsymbol{\chi}(\boldsymbol{q};\omega)$ provides an effective magnetic field for phonons.

To be specific, we calculate the $\delta\boldsymbol{\Phi}(\boldsymbol{q})$ and $\boldsymbol{\mathcal{A}}(\boldsymbol{q})$ terms in the square lattice. We first calculate $\boldsymbol{\chi}(\boldsymbol{q};\omega)$ using the analytical continuation of its corresponding Matsubara Green's function in the frequency representation. We get

$$\boldsymbol{\chi}(\boldsymbol{q};\omega) = \frac{2}{\hbar(\omega^2 - \omega_{m\boldsymbol{q}}^2)}\begin{pmatrix} -\omega_{m\boldsymbol{q}} & i\omega \\ -i\omega & -\omega_{m\boldsymbol{q}} \end{pmatrix}\ , \tag{C.7}$$

where $\omega_{m\boldsymbol{q}} = E_m(\boldsymbol{q})/\hbar$ is the magnon frequency. In the square lattice, we have $T^{xx(yy)}(\boldsymbol{q}) = -DS/a\{2 - 2\cos(q_{y(x)}a) - \gamma[2 - 2\cos(q_{x(y)}a)]\}$ and $T^{xy(yx)}(\boldsymbol{q}) = 0$.

Substituting $\boldsymbol{T}(\boldsymbol{q})$ and $\boldsymbol{\chi}(\boldsymbol{q};\omega)$ into Eq. (C.6) and (C.5) and gives the effective magnetic field

$$\mathcal{A}^{\alpha\beta}(\boldsymbol{q}) = -4\frac{c_0(\boldsymbol{q})}{\hbar\omega_{m\boldsymbol{q}}^2}\epsilon^{\alpha\beta}\ , \tag{C.8}$$

where ϵ is the antisymmetric unit tensor and $c_0(\boldsymbol{q}) = 2S^3D^2[1 - \cos(q_xa) - \gamma(1 - \cos(q_ya))][1 - \cos(q_ya) - \gamma(1 - \cos(q_xa))]/a^2$, and the screening term

$$\delta\Phi^{\alpha\beta}(\boldsymbol{q}) = -2\delta^{\alpha\beta}\frac{c_\alpha(\boldsymbol{q})}{\omega_{m\boldsymbol{q}}}\ , \tag{C.9}$$

where $c_{x(y)}(\boldsymbol{q}) = 2S^3D^2[1 - \cos(q_{y(x)}a) - \gamma(1 - \cos(q_{x(y)}a))]^2/a^2$. Accordingly, the leading order of the screening effect $\delta\boldsymbol{\Phi}$ and the effective magnetic field $\boldsymbol{\mathcal{A}}$ from the magnon-phonon coupling is independent of the temperature. All temperature dependent terms come from higher order expansions.

Appendix D

Optical Selection Rule with Crystal Field Effect

D.1 Envelope function modified by crystal field effect

In an isotropic system, the envelope function $f_m(\boldsymbol{k})$ for an exciton state with angular momentum m is given by the Bethe-Salpeter equation [105]

$$\sum_{\boldsymbol{k}'} H_{BS}(\boldsymbol{k},\boldsymbol{k}')f_m(\boldsymbol{k}') = E_m^{ex} f_m(\boldsymbol{k}), \tag{D.1}$$

where E_m^{ex} is the exciton energy,

$$H_{BS}(\boldsymbol{k},\boldsymbol{k}') = [(\epsilon_{c\boldsymbol{k}} - \epsilon_{v\boldsymbol{k}}) + \Sigma_{\boldsymbol{k}}]\delta_{\boldsymbol{k},\boldsymbol{k}'} - U(\boldsymbol{k},\boldsymbol{k}') \tag{D.2}$$

is the exciton Hamiltonian, and

$$U(\boldsymbol{k},\boldsymbol{k}') = V(|\boldsymbol{k}-\boldsymbol{k}'|)\langle c\boldsymbol{k}|c\boldsymbol{k}'\rangle\langle v\boldsymbol{k}'|v\boldsymbol{k}\rangle. \tag{D.3}$$

Here $V(|\boldsymbol{k}-\boldsymbol{k}'|)$ is the Coulomb interaction in momentum space, and $\epsilon_{c\boldsymbol{k}} - \epsilon_{v\boldsymbol{k}}$ is the band gap. The self energy $\Sigma_{\boldsymbol{k}}$ can be absorbed into the gap and will be ignored for a qualitative discussion.

The crystal field effect modifies $H_{BS}(\boldsymbol{k},\boldsymbol{k}')$ by changing the dispersion $\epsilon_{\boldsymbol{k}}$ and the interaction term $U(\boldsymbol{k},\boldsymbol{k}')$. We denote the perturbation term from the crystal field effect in the exciton Hamiltonian as $H'_{BS}(\boldsymbol{k},\boldsymbol{k}')$, and the envelope function of the s-state ($m=0$) after perturbation is

$$\begin{aligned} f'_0(\boldsymbol{k}) =& f_0(\boldsymbol{k}) + \sum_{m\neq 0} c_m f_m(\boldsymbol{k}) \\ =& f_0(\boldsymbol{k}) \\ &+ \sum_{m\neq 0} \frac{\int d^2\boldsymbol{k}'' d^2\boldsymbol{k}' f_m^*(\boldsymbol{k}'') H'_{BS}(\boldsymbol{k}'',\boldsymbol{k}') f_0(\boldsymbol{k}')}{E_0^{ex} - E_m^{ex}} f_m(\boldsymbol{k}). \end{aligned} \tag{D.4}$$

As an application, consider the 2×2 Hamiltonian of the biased bilayer graphene [119]:

$$H = H_0 + H_{warp} = \begin{pmatrix} V_g(1 - 2\frac{\gamma_0^2}{\gamma_1^2}k^2) & \frac{\gamma_0^2}{\gamma_1}k_+^2 \\ \frac{\gamma_0^2}{\gamma_1}k_-^2 & -V_g(1 - 2\frac{\gamma_0^2}{\gamma_1^2}k^2) \end{pmatrix} + \begin{pmatrix} 0 & 3\gamma_3 k_- \\ 3\gamma_3 k_+ & 0 \end{pmatrix} , \quad \text{(D.5)}$$

For simplicity, define $\Delta(\boldsymbol{k}) = V_g(1 - 2\gamma_0^2 k^2/\gamma_1^2)$ and $\alpha = \gamma_0^2/\gamma_1$. Then we write out the isotropic Hamiltonian H_0 in terms of the creation and annihilation operators of the eigenstates of H_0

$$H_0 = \sum_{n=c,v} \epsilon_{n\boldsymbol{k}} a_{n\boldsymbol{k}}^\dagger a_{n\boldsymbol{k}}, \quad \text{(D.6)}$$

where $\epsilon_{n\boldsymbol{k}} = \pm\sqrt{\Delta(k)^2 + \alpha^2 k^4}$ is the energy of H_0. Using the new basis $a_{c,v\boldsymbol{k}}$, the γ_3 term becomes

$$H_{warp} = \sum_{k,n,n'} H'_{n,n'} a_{nk}^\dagger a_{n'k} \ . \quad \text{(D.7)}$$

where

$$H'_{cc(vv)} = \pm\frac{3}{2}\gamma_3 k \sin\tilde{\theta}_{\boldsymbol{k}}(e^{-3i\phi_{\boldsymbol{k}}} + e^{3i\phi_{\boldsymbol{k}}}) \quad \text{(D.8)}$$

and

$$H'_{vc} = 3k\gamma_3(\cos^2\frac{\tilde{\theta}_{\boldsymbol{k}}}{2}e^{-i\phi_{\boldsymbol{k}}} + \sin^2\frac{\tilde{\theta}_{\boldsymbol{k}}}{2}e^{5i\phi_{\boldsymbol{k}}}) \ , \quad \text{(D.9)}$$

where the polar angle $\tilde{\theta}_{\boldsymbol{k}}$ is defined as

$$\sin\tilde{\theta}_{\boldsymbol{k}} = \frac{\alpha k^2}{\sqrt{\Delta(\boldsymbol{k})^2 + \alpha^2 k^4}} \ . \quad \text{(D.10)}$$

The crystal field effect modifies $H_{BS}(\boldsymbol{k}, \boldsymbol{k}')$ by changing the band gap $\epsilon_{c\boldsymbol{k}} - \epsilon_{v\boldsymbol{k}}$ and the interaction term $U(\boldsymbol{k}, \boldsymbol{k}')$. The influence on the band gap is attributed to $H'_{cc(vv)}$, which is proportional to $1/\Delta$. On the other hand, the crystal field effect in the $U(\boldsymbol{k}, \boldsymbol{k}')$ term is introduced by modifying the wave functions. Under the first order perturbation, the wave functions are given by

$$|c\boldsymbol{k}\rangle \to |c\boldsymbol{k}\rangle + \frac{H'_{vc}}{2\epsilon_{c\boldsymbol{k}}}|v\boldsymbol{k}\rangle \quad \text{(D.11)}$$

and

$$|v\boldsymbol{k}\rangle \to |v\boldsymbol{k}\rangle + \frac{H'_{cv}}{2\epsilon_{c\boldsymbol{k}}}|c\boldsymbol{k}\rangle \ . \quad \text{(D.12)}$$

Therefore, the perturbation to the Coulomb interaction term is given by

$$\begin{aligned} &U(\boldsymbol{k}, \boldsymbol{k}') \to U(\boldsymbol{k}, \boldsymbol{k}') \\ &+V(|\boldsymbol{k} - \boldsymbol{k}'|)(\frac{H'_{cv}}{2\epsilon_{c\boldsymbol{k}}}\langle v\boldsymbol{k}|c\boldsymbol{k}'\rangle + \frac{H'_{vc}}{2\epsilon_{c\boldsymbol{k}}}\langle c\boldsymbol{k}|v\boldsymbol{k}'\rangle - \frac{H'_{cv}}{2\epsilon_{c\boldsymbol{k}}}\langle v\boldsymbol{k}'|c\boldsymbol{k}\rangle - \frac{H'_{vc}}{2\epsilon_{c\boldsymbol{k}}}\langle c\boldsymbol{k}'|v\boldsymbol{k}\rangle), \end{aligned} \quad \text{(D.13)}$$

where $\langle v\boldsymbol{k}|c\boldsymbol{k}'\rangle = \sin\frac{\tilde{\theta}_{\boldsymbol{k}}}{2}\cos\frac{\tilde{\theta}'_{\boldsymbol{k}}}{2}e^{-2i\phi_{\boldsymbol{k}}} - \cos\frac{\tilde{\theta}_{\boldsymbol{k}}}{2}\sin\frac{\tilde{\theta}'_{\boldsymbol{k}}}{2}e^{-2i\phi_{\boldsymbol{k}'}}$. Checking term by term, we find that the corrections are proportional to either $\gamma_3 e^{\pm 3i\phi_{\boldsymbol{k}'}}$ or $\gamma_3 e^{\pm 3i\phi_{\boldsymbol{k}}}$.

Since the perturbation term from $U(\boldsymbol{k},\boldsymbol{k}')$ is in the order of $1/\Delta^2$, H'_{BS} is dominated by the correction of the band gap under the large band gap limit, which gives

$$H'_{BS} = \frac{3\gamma_3\alpha}{\Delta}(k_+^3 + k_-^3)\ . \tag{D.14}$$

In the large band gap limit $\Delta \gg \tilde{\alpha}k_B^w$, we can estimate the coefficients c_m according to Eq. (D.4) as

$$c_{-3} \propto \frac{\alpha|k_B|^2}{E_0^{ex} - E_{-3}^{ex}}\frac{3\gamma_3|k_B|}{\Delta}\ , \tag{D.15}$$

$$c_3 \propto \frac{\alpha|k_B|^2}{E_0^{ex} - E_3^{ex}}\frac{3\gamma_3|k_B|}{\Delta}\ . \tag{D.16}$$

All wave vectors are estimated by $k_B = 2\pi/a_B$ and a_B is the exciton Bohr radius. The first terms in Eq. (D.15) and (D.16) have an order of magnitude of 1. The second terms are proportional to the warping term γ_3 which is regarded as a small quantity compared with the band gap Δ. Note that in the first-order perturbation theory, c_3 and c_{-3} are the only nonzero coefficients. With higher-order perturbation theory, c_{3n} will become nonzero, where n is an integer.

D.2 Velocity operator modified by crystal field effect

For the Hamiltonian Eq. (D.5), the wave functions of the conduction and valence band are

$$|c\boldsymbol{k}\rangle = \begin{pmatrix} 1 \\ \frac{\alpha}{2\Delta}k_-^2 + \frac{3\gamma_3}{2\Delta}k_+ \end{pmatrix} \tag{D.17}$$

$$|v\boldsymbol{k}\rangle = \begin{pmatrix} \frac{\alpha}{2\Delta}k_+^2 + \frac{3\gamma_3}{2\Delta}k_- \\ -1 \end{pmatrix} \tag{D.18}$$

in the large band gap limit. To the leading order of γ_3, we can show that

$$\begin{aligned}\langle v\boldsymbol{k}|v_+|c\boldsymbol{k}\rangle =& -4\alpha|k|e^{-i\phi_{\boldsymbol{k}}} \\ &+ \frac{\alpha^2|k|^4}{4\Delta^2}6\gamma_3 e^{-4i\phi_{\boldsymbol{k}}}\ ,\end{aligned} \tag{D.19}$$

$$\begin{aligned}\langle v\boldsymbol{k}|v_-|c\boldsymbol{k}\rangle =& \frac{\alpha^2|k|^4}{\Delta^2}\alpha|k|e^{-3i\phi_{\boldsymbol{k}}} \\ &+ (\frac{\alpha^2|k|^4}{\Delta^2}6\gamma_3 - 6\gamma_3)\ .\end{aligned} \tag{D.20}$$

D.3 Selection rule modified by crystal field effect

Combining the results from section D.1 and D.2, we can evaluate the oscillator strength modified by the crystal field effect.

As mentioned in the main text, the s-state is dark in the isotropic $w = 2$ chiral fermion model, and is bright when the C_3 warping term is included. For a quantitative discussion, we calculate the oscillator strength for the s-state in gapped graphene bilayers,

$$\begin{aligned} O_0 =& \frac{1}{\mu E_m^{ex}} \sum_{\eta=\pm} |\sum_{\boldsymbol{k}} f_m(\boldsymbol{k}) v_\eta(\boldsymbol{k})|^2 \\ =& \sum_{\eta=\pm} |\sum_{\boldsymbol{k}} (\tilde{f}_0 + \sum_{n=\pm 1} c_{3n} \tilde{f}_{3n} e^{3ni\phi_{\boldsymbol{k}}}) (\sum_{n'=\pm 1} v_\eta^{3n'} e^{3n'i\phi_{\boldsymbol{k}}})|^2 \\ \approx& |\sum_{|k|} 6\tilde{f}_0 \gamma_3|^2, \end{aligned} \tag{D.21}$$

where $v_\eta^m(|\boldsymbol{k}|) = \int d\phi_{\boldsymbol{k}} e^{-im\phi_{\boldsymbol{k}}} v_\eta(\boldsymbol{k})$ and $f_m(\boldsymbol{k}) = \tilde{f}_m(|\boldsymbol{k}|) e^{im\phi_{\boldsymbol{k}}}$. In the last line of Eq. (D.21), only the leading order terms are preserved, which come from the modification of the velocity matrix. Note that the perturbation terms from the envelope function are high order corrections.

Now we compare the oscillator strength of the s-state with that of the bright p-state ($m = 1$) and $m = 3$ state. The oscillator strengths of p-state and $m = 3$ state are given by

$$O_1 = |\sum_{|k|} \tilde{f}_1 4\alpha |k||^2 \,, \tag{D.22}$$

$$O_3 = |\sum_{|k|} \tilde{f}_3 \frac{\alpha^2 |k|^4}{\Delta^2} \alpha |k||^2 \,. \tag{D.23}$$

The relative oscillator strength can be written as

$$O_0/O_1 \sim (\frac{3\gamma_3}{2\alpha |k_B|})^2 \tag{D.24}$$

and

$$O_0/O_3 \sim (\frac{4\Delta^2}{\alpha^2 |k_B|^4})^2 O_0/O_1. \tag{D.25}$$

Accordingly, the relative oscillator strength between the s-state and the p-state is given by the ratio between the off-diagonal terms in H_{warp} and H_0, which are $\gamma_3|k_B|$ and $\alpha|k_B|^2$, respectively. Moreover, compared with O_0/O_1, the relative oscillator strength O_0/O_3 is enhanced by a factor of $\frac{4\Delta^2}{\alpha^2|k_B|^4}$, which is large in the large band gap limit.

Appendix E

Effective $k \cdot p$ Hamiltonian

Generally, the tight binding theory involves multiple atomic orbitals. In some cases, part of these orbitals form a set of energy bands that have similar energies, and this set of bands is called a nearly degenerate set (NDS). It is a good approximation to only study the subspace spanned by NDS. In this appendix, we use the $\boldsymbol{k}\cdot\boldsymbol{p}$ approximation to construct effective Hamiltonians spanned by NDS.

We consider a simple case where the NDS consists of two bands denoted by m and n, such that the energy of band m and n are very closed to each other at the $\boldsymbol{k}_0$ point. The Schrodinger equation for the periodic part of Bloch function is given by

$$(\frac{p^2}{2M}+V(\boldsymbol{r})+\frac{\hbar\boldsymbol{k}_0\cdot\boldsymbol{p}}{M}+\frac{\hbar\boldsymbol{k}\cdot\boldsymbol{p}}{M})u_{n,\boldsymbol{k}_0+\boldsymbol{k}}=(E_n(\boldsymbol{k}_0+\boldsymbol{k})-\frac{\hbar^2(\boldsymbol{k}+\boldsymbol{k}_0)^2}{2M})u_{n,\boldsymbol{k}_0+\boldsymbol{k}}, \quad \text{(E.1)}$$

where $\boldsymbol{k}$ is the deviation from $\boldsymbol{k}_0$. The $\boldsymbol{k}\cdot\boldsymbol{p}$ approximation regards $u_{n,\boldsymbol{k}_0}$ as the zeroth order wave function, and $H'=\hbar\boldsymbol{k}\cdot\boldsymbol{p}/M$ is the perturbation Hamiltonian. In the second order degenerate perturbation theory, the effective Hamiltonian is written as (see Ref. [136]):

$$H=\begin{pmatrix} E_m+\frac{\hbar\boldsymbol{k}}{M}\cdot\langle m|\boldsymbol{p}|m\rangle & \frac{\hbar\boldsymbol{k}}{M}\cdot\langle m|\boldsymbol{p}|n\rangle+\sum\limits_{\alpha\notin NDS}\frac{\langle m|\boldsymbol{p}|\alpha\rangle\cdot\boldsymbol{k}\boldsymbol{k}\cdot\langle\alpha|\boldsymbol{p}|n\rangle}{E_m-E_\alpha} \\ \frac{\hbar\boldsymbol{k}}{M}\cdot\langle n|\boldsymbol{p}|m\rangle+\sum\limits_{\alpha\notin NDS}\frac{\langle n|\boldsymbol{p}|\alpha\rangle\cdot\boldsymbol{k}\boldsymbol{k}\cdot\langle\alpha|\boldsymbol{p}|m\rangle}{E_n-E_\alpha} & E_n+\frac{\hbar\boldsymbol{k}}{M}\cdot\langle n|\boldsymbol{p}|n\rangle \end{pmatrix}. \quad \text{(E.2)}$$

Note that the k and k^2 terms in the off-diagonal elements originate from the first and second order perturbation, respectively.

We use the biased bilayer graphene as an example. Four bands are relevant near the Fermi level, as discussed in the main text. At the K point, the p_z orbitals from B_1 and B_2 are close to each other and form a NDS (see Fig. 4.3 for the position of lattice points). The group of wave vector of the K point is C_3, and the B_1 and B_2 states are in E and E^* IR, respectively. We also consider the second order contributions from A_1 and A_2 orbitals. At the band edge, these two orbitals mix and form a conduction band above NDS and valence band below NDS (see Eq. (4.37) and Fig. 4.4). Since

both A_1 and A_2 orbitals at the K points are in A representation, their mixtures are also in A representation. In addition, H' can be decomposed using the basis functions of C_3 group as

$$H' \sim \boldsymbol{k}\cdot\boldsymbol{p} = \frac{1}{2}((k_x+ik_y)(p_x-ip_y)+(k_x-ik_y)(p_x+ip_y)) \ , \tag{E.3}$$

where $p_x \pm p_y$ belongs to E and E^* IR. A nonzero matrix elements $\langle m|H'|n\rangle$ requires that $\Gamma^{m*}\bigotimes\Gamma^{H'}\bigotimes\Gamma^{n}$ contains an identity representation. After checking the multiplication table for the C_3 group, we get $\langle B_1|H'|A\rangle \sim k_x+ik_y$, $\langle B_2|H'|A\rangle \sim k_x-ik_y$ and $\langle B_1|H'|B_2\rangle \sim k_x-ik_y$. Therefore, in the off-diagonal term H_{B_1,B_2}, the first order perturbation gives k_x-ik_y, and the second order term is $(k_x+ik_y)^2$. This result agrees with the effective two-band model in Eq. (4.36).

This analysis can go beyond two band models. Here we build a four-band model for $3R$-stacked MoS_2. The first step is figuring out the irreducible representations of the four basis, which can be seen by deriving the transformation law of the atomic wave functions and the plane wave part of Bloch wave functions under the symmetry operations. On one hand, the band edge states consist of d_{z^2} and $d_{x^2-y^2}+id_{xy}$ orbitals on Mo atoms from both layers at valley $K=(-\frac{4\pi}{3\sqrt{3}a},0)$, where a is the lattice constant of MoS_2. The atomic orbital d_{z^2} is invariant under a clockwise three-fold rotation $\hat{C}_3$ while $\phi(r)=d_{x^2-y^2}+id_{xy}$ orbital gets a phase factor $e^{-i\frac{2\pi}{3}}$. On the other hand, the plane wave part of the Bloch function also gives rise to a phase factor under a three-fold rotation. Together, the transformation rule follows Eq. (4.40) in the main text, which is re-written here

$$C_3\psi_K(r) = e^{i(C_3\boldsymbol{K}-\boldsymbol{K})\cdot\tilde{\boldsymbol{R}}}e^{i\alpha}\psi_K(r) \ . \tag{E.4}$$

Here $\boldsymbol{R}_i$ is the position of the unit cell, $\tilde{\boldsymbol{R}}$ is the relative position of the lattice point in the unit cell, and α is the phase contributed by the atomic orbital under a three-fold rotation. As mentioned earlier, we have $\alpha=0$ for d_{z^2} and $\alpha=-\frac{2\pi}{3}$ for $d_{x^2-y^2}+id_{xy}$. According to the top view of the 3R-stacked MoS_2 bilayer in Fig. 4.12, the R_A point is the rotation center of the three-fold rotation, so $\boldsymbol{R}_A=0$, and $\boldsymbol{R}_B=(\frac{\sqrt{3}}{2}a,\frac{1}{2}a)$. Combined with $C_3\boldsymbol{K}-\boldsymbol{K}=(\frac{2\sqrt{3}\pi}{3a},\frac{2\pi}{3a})$, and $(C_3\boldsymbol{K}-\boldsymbol{K})\cdot\boldsymbol{R}_B=-\frac{2\pi}{3}$, the d_{z^2} and $d_{x^2-y^2}+id_{xy}$ bands on layer 1 (red in Fig. 4.12) are in $A(1)$ and $E^*(1)$ irreducible representations, while on layer 2 (blue in Fig. 4.12) they are in $E^*(2)$ and $E(2)$ irreducible representations. The numbers in the brackets label the layers where the states are located in.

Using the same group theory analysis as the two-band case, the effective four-band Hamiltonian written in the basis of $A(1)$, $E^*(1)$, $E^*(2)$, $E(2)$ is given by

$$H_{3R} = \begin{pmatrix} \Delta_I & v_0(k_x-ik_y) & v_1(k_x-ik_y) & v_3(k_x+ik_y) \\ v_0(k_x+ik_y) & -\Delta_I & \gamma_1 & v_2(k_x-ik_y) \\ v_1(k_x+ik_y) & \gamma_1 & \Delta_I & v_0(k_x-ik_y) \\ v_3(k_x-ik_y) & v_2(k_x+ik_y) & v_0(k_x+ik_y) & -\Delta_I \end{pmatrix} . \tag{E.5}$$

The diagonal 2×2 blocks are the Hamiltonians for the two MoS_2 monolayers, and the off-diagonal blocks represent the interlayer coupling. Assuming that the interlayer coupling is weak, we can ignore the v_1, v_2 and v_3 terms which are zero at $\boldsymbol{k} = 0$.

Appendix F

Downfolding into the two-band model

We start from the effective $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian in Eq. (E.5). In the simplified case $v_1 = v_2 = v_3 = 0$ and by applying an out-of-plane gate voltage V_g, we have the four-band model of the MoS_2 bilayer

$$H_{3\mathrm{R}} = \begin{pmatrix} \Delta_I + V_g & v_0 k_- & 0 & 0 \\ v_0 k_+ & -\Delta_I + V_g & \gamma_1 & 0 \\ 0 & \gamma_1 & \Delta_I - V_g & v_0 k_- \\ 0 & 0 & v_0 k_+ & -\Delta_I - V_g \end{pmatrix} . \tag{F.1}$$

This produces Eq. (13) in the main text. Note that the C_∞ symmetry is retained in this simplified model.

We first diagonalize the four-band model (F.1) at $\boldsymbol{k} = 0$ by a unitary transformation U. To the first order of γ_1/Δ_I,

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{\gamma_1}{2E_4} \\ 0 & 0 & -\frac{\gamma_1}{2E_4} & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} , \tag{F.2}$$

where $E_4 = \sqrt{(\Delta_I - V_g)^2 + \gamma_1^2}$. Applying the unitary transformation U on $H_{3\mathrm{R}}$, we have

$$\begin{aligned} \tilde{H}_{3\mathrm{R}} &= U^{-1} H_{3\mathrm{R}} U \\ &= \begin{pmatrix} \Delta_I + V_g & 0 & v_0 k_- & \frac{\gamma_1 v_0}{2E_3} k_- \\ 0 & -\Delta_I - V_g & -\frac{\gamma_1 v_0}{2E_3} k_+ & v_0 k_+ \\ v_0 k_+ & -\frac{\gamma_1 v_0}{2E_3} k_- & -E_4 & 0 \\ \frac{\gamma_1 v_0}{2E_3} k_+ & v_0 k_- & 0 & E_4 \end{pmatrix} . \end{aligned} \tag{F.3}$$

Now the basis are in $A(1)$, $E(2)$, $E^*(1)$, $E^*(2)$ irreducible representations. When V_g is small, the lowest conduction band and the highest valence band are made up of

the upper two basis $A(1)$, $E(2)$. In this case, we downfold the four-band model into a two-band model that only includes the $A(1)$ and $E(2)$ basis. To the second order of $\boldsymbol{k}$, the downfolded Hamiltonian is given by [124]

$$\begin{aligned}(H_{\text{before}})_{mm'} =&(\tilde{H}^0_{3\text{R}})_{mm'} \\ &+\sum_l \frac{(\tilde{H}^1_{3\text{R}})_{ml}(\tilde{H}^1_{3\text{R}})_{lm'}}{2}(\frac{1}{E_m - E_l} + \frac{1}{E_{m'} - E_l}),\end{aligned} \tag{F.4}$$

where $m, m' = 1, 2$, $l = 3, 4$, $E_n = (\tilde{H}_{3\text{R}})_{nn}$ and $\tilde{H}^0_{3\text{R}}$, $\tilde{H}^1_{3\text{R}}$ contain the diagonal and off-diagonal terms in $\tilde{H}_{3\text{R}}$, respectively. It is straightforward to get the downfolded Hamiltonian

$$H_{\text{before}} = \begin{pmatrix} \Delta_I + V_g & \frac{v_0^2\gamma_1}{4\Delta_I V_g}k_-^2 \\ \frac{v_0^2\gamma_1}{4\Delta_I V_g}k_+^2 & -\Delta_I - V_g \end{pmatrix} . \tag{F.5}$$

Here we get a chiral fermion model with $w = 2$, as given by Eq. 14 in the main text. This model is similar to biased graphene bilayers where the s-state is bright in a C_3 invariant system.

When V_g is large enough to switch the band order, the lowest conduction band and the highest valence band are given by $E^*(1)$ and $E^*(2)$ basis. Similar to H_{before}, we get the downfolded two-band model after the band crossing

$$H_{\text{after}} = \begin{pmatrix} -\sqrt{\gamma_1^2 + (\Delta_I - V_g)^2} & -\frac{v_0^2\gamma_1(\Delta_I + V_g)}{4(\Delta_I - V_g)\Delta_I V_g}k^2 \\ -\frac{v_0^2\gamma_1(\Delta_I + V_g)}{4(\Delta_I - V_g)\Delta_I V_g}k^2 & \sqrt{\gamma_1^2 + (\Delta_I - V_g)^2} \end{pmatrix} . \tag{F.6}$$

We see that the winding number is changed from 2 to 0 after the band crossing. In this case, the s-state is always dark even when the discrete C_3 symmetry is considered.

Bibliography

[1] M.-C. Chang and Q. Niu, Phys. Rev. B **53**, 7010 (1996).

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

[3] J. C. W. Song, P. Samutpraphoot, and L. S. Levitov, Proceedings of the National Academy of Sciences **112**, 10879 (2015), https://www.pnas.org/content/112/35/10879.full.pdf .

[4] W. Yao, D. Xiao, and Q. Niu, Phys. Rev. B **77**, 235406 (2008).

[5] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004), https://science.sciencemag.org/content/306/5696/666.full.pdf .

[6] K. I. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, and H. Stormer, Solid State Communications **146**, 351 (2008).

[7] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proceedings of the National Academy of Sciences **102**, 10451 (2005).

[8] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).

[9] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).

[10] B. Huang, G. Clark, E. Navarro-Moratalla, D. R. Klein, R. Cheng, K. L. Seyler, D. Zhong, E. Schmidgall, M. A. McGuire, D. H. Cobden, *et al.*, Nature **546**, 270 (2017).

[11] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).

[12] A. Ramasubramaniam, Phys. Rev. B **86**, 115409 (2012).

[13] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. Lett. **113**, 076802 (2014).

[14] Y. Zhang, T.-T. Tang, C. Girit, Z. Hao, M. C. Martin, A. Zettl, M. F. Crommie, Y. R. Shen, and F. Wang, Nature **459**, 820 (2009).

[15] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 EP (2013).

[16] H. Min, R. Bistritzer, J.-J. Su, and A. H. MacDonald, Phys. Rev. B **78**, 121401 (2008).

[17] F.-C. Wu, F. Xue, and A. H. MacDonald, Phys. Rev. B **92**, 165121 (2015).

[18] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, and P. Jarillo-Herrero, Nature **556**, 80 EP (2018).

[19] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Nature **556**, 43 EP (2018).

[20] G. Chen, L. Jiang, S. Wu, B. Lyu, H. Li, B. L. Chittari, K. Watanabe, T. Taniguchi, Z. Shi, J. Jung, Y. Zhang, and F. Wang, Nature Physics **15**, 237 (2019).

[21] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, Nberrbeature **438**, 201 (2005).

[22] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).

[23] B. I. Halperin, Phys. Rev. B **25**, 2185 (1982).

[24] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).

[25] J. E. Avron, R. Seiler, and B. Simon, Phys. Rev. Lett. **51**, 51 (1983).

[26] Q. Niu, D. J. Thouless, and Y.-S. Wu, Phys. Rev. B **31**, 3372 (1985).

[27] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).

[28] T. Kato, Journal of the Physical Society of Japan **5**, 435 (1950).

[29] A. Messiah, *Quantum Mechanics*, Vol. II (North Holland, Amsterdam, 1999).

[30] V. Fock, Zeitschrift für Physik **49**, 323 (1928).

[31] M. V. Berry, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **392**, 45 (1984).

[32] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).

[33] D. Culcer, Y. Yao, and Q. Niu, Phys. Rev. B **72**, 085110 (2005).

[34] R. Shindou and K.-I. Imura, Nuclear Physics B **720**, 399 (2005).

[35] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[36] N. W. Ashcroft and N. D. Mermin, (Saunders, Philadelphia, 2010).

[37] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[38] H. M. Price, T. Ozawa, and I. Carusotto, Phys. Rev. Lett. **113**, 190403 (2014).

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

[40] L. F. Mattheiss, Phys. Rev. B **8**, 3719 (1973).

[41] Y. Onose, T. Ideue, H. Katsura, Y. Shiomi, N. Nagaosa, and Y. Tokura, Science **329**, 297 (2010).

[42] M. Hirschberger, R. Chisnell, Y. S. Lee, and N. P. Ong, Phys. Rev. Lett. **115**, 106603 (2015).

[43] M. Hirschberger, J. W. Krizan, R. J. Cava, and N. P. Ong, Science **348**, 106 (2015).

[44] G. Grissonnanche, A. Legros, S. Badoux, E. Lefrançois, V. Zatko, M. Lizaire, F. Laliberté, A. Gourgout, J. Zhou, S. Pyon, *et al.*, arXiv:1901.03104 (2019).

[45] J. H. Han, J.-H. Park, and P. A. Lee, arXiv:1903.01125 (2019).

[46] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).

[47] K. L. Seyler, D. Zhong, D. R. Klein, S. Gao, X. Zhang, B. Huang, E. Navarro-Moratalla, L. Yang, D. H. Cobden, M. A. McGuire, *et al.*, Nature Physics **14**, 277 (2018).

[48] T. Song, X. Cai, M. W.-Y. Tu, X. Zhang, B. Huang, N. P. Wilson, K. L. Seyler, L. Zhu, T. Taniguchi, K. Watanabe, *et al.*, Science **360**, 1214 (2018).

[49] Y. Yao, L. Kleinman, A. H. MacDonald, J. Sinova, T. Jungwirth, D.-s. Wang, E. Wang, and Q. Niu, Phys. Rev. Lett. **92**, 037204 (2004).

[50] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. Lett. **99**, 256802 (2007).

[51] Z.-G. Chen, Z. Shi, W. Yang, X. Lu, Y. Lai, H. Yan, F. Wang, G. Zhang, and Z. Li, Nature Communications **5**, 4461 EP (2014).

[52] Z. Shi, C. Jin, W. Yang, L. Ju, J. Horng, X. Lu, H. A. Bechtel, M. C. Martin, D. Fu, J. Wu, K. Watanabe, T. Taniguchi, Y. Zhang, X. Bai, E. Wang, G. Zhang, and F. Wang, Nature Physics **10**, 743 EP (2014).

[53] G. Chen, L. Jiang, S. Wu, B. Lyu, H. Li, B. L. Chittari, K. Watanabe, T. Taniguchi, Z. Shi, J. Jung, *et al.*, arXiv:1803.01985 (2018).

[54] H. Henck, D. Pierucci, G. Fugallo, J. Avila, G. Cassabois, Y. J. Dappe, M. G. Silly, C. Chen, B. Gil, M. Gatti, F. Sottile, F. Sirotti, M. C. Asensio, and A. Ouerghi, Phys. Rev. B **95**, 085410 (2017).

[55] M.-C. Chang and Q. Niu, Journal of Physics: Condensed Matter **20**, 193202 (2008).

[56] C. Kittel, P. McEuen, and P. McEuen, *Introduction to solid state physics*, Vol. 8 (Wiley New York, 1996).

[57] W. Greiner *et al.*, *Relativistic quantum mechanics*, Vol. 3 (Springer, 1990).

[58] T. Lee, Contemp. Concepts Phys. **1**, 447 (1981).

[59] W. Eerenstein, N. D. Mathur, and J. F. Scott, Nature **442**, 759 (2006).

[60] S.-W. Cheong and M. Mostovoy, Nat. Mater. **6**, 13 (2007).

[61] Y. Tokura, S. Seki, and N. Nagaosa, Rep. Prog. Phys. **77**, 076501 (2014).

[62] A. Pimenov, A. Mukhin, V. Y. Ivanov, V. Travkin, A. Balbashov, and A. Loidl, Nat. Phys. **2**, 97 (2006).

[63] Y. Takahashi, R. Shimano, Y. Kaneko, H. Murakawa, and Y. Tokura, Nat. Phys. **8**, 121 (2012).

[64] W. Chen and M. Sigrist, Phys. Rev. Lett. **114**, 157203 (2015).

[65] T. Nomura, X.-X. Zhang, S. Zherlitsyn, J. Wosnitza, Y. Tokura, N. Nagaosa, and S. Seki, arXiv:1809.08775 .

[66] C. Strohm, G. L. J. A. Rikken, and P. Wyder, Phys. Rev. Lett. **95**, 155901 (2005).

[67] T. Ideue, T. Kurumaji, S. Ishiwata, and Y. Tokura, Nat. Mater. **16**, 797 (2017).

[68] K. Sugii, M. Shimozawa, D. Watanabe, Y. Suzuki, M. Halim, M. Kimata, Y. Matsumoto, S. Nakatsuji, and M. Yamashita, Phys. Rev. Lett. **118**, 145902 (2017).

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

[70] H. Katsura, N. Nagaosa, and P. A. Lee, Phys. Rev. Lett. **104**, 066403 (2010).

[71] R. Matsumoto and S. Murakami, Phys. Rev. Lett. **106**, 197202 (2011).

[72] R. Matsumoto and S. Murakami, Phys. Rev. B **84**, 184406 (2011).

[73] L. Sheng, D. N. Sheng, and C. S. Ting, Phys. Rev. Lett. **96**, 155901 (2006).

[74] Y. Kagan and L. A. Maksimov, Phys. Rev. Lett. **100**, 145902 (2008).

[75] T. Qin, J. Zhou, and J. Shi, Phys. Rev. B **86**, 104305 (2012).

[76] M. Mori, A. Spencer-Smith, O. P. Sushkov, and S. Maekawa, Phys. Rev. Lett. **113**, 2 (2014).

[77] R. Takahashi and N. Nagaosa, Phys. Rev. Lett. **117**, 1 (2016).

[78] S. Park and B.-J. Yang, arXiv preprint arXiv:1901.06500 (2019).

[79] G. E. W. Bauer, E. Saitoh, and B. J. van Wees, Nature Materials **11**, 391 (2012).

[80] D. Gosálbez-Martínez, I. Souza, and D. Vanderbilt, Phys. Rev. B **92**, 085138 (2015).

[81] D. Liu and J. Shi, Phys. Rev. Lett. **119**, 075301 (2017).

[82] J.-S. Wang and L. Zhang, Phys. Rev. B **80**, 012301 (2009).

[83] H. Capellmann and S. Lipinski, Zeitschrift für Physik B Condensed Matter **83**, 199 (1991).

[84] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).

[85] R. Cheng, S. Okamoto, and D. Xiao, Phys. Rev. Lett. **117**, 217202 (2016).

[86] Y. Zhang, S. Okamoto, and D. Xiao, Phys. Rev. B **98**, 035424 (2018).

[87] C. Zhang and Q. Niu, Phys. Rev. A **81**, 053803 (2010).

[88] D. Xiao, Y. Yao, Z. Fang, and Q. Niu, Phys. Rev. Lett. **97**, 026603 (2006).

[89] R. J. Hardy, Phys. Rev. **132**, 168 (1963).

[90] R. Matsumoto, R. Shindou, and S. Murakami, Phys. Rev. B **89**, 054420 (2014).

[91] J. M. Luttinger, Phys. Rev. **135**, A1505 (1964).

[92] W.-B. Zhang, Q. Qu, P. Zhu, and C.-H. Lam, J. Mater. Chem. C **3**, 12457 (2015).

[93] J. L. Lado and J. Fernández-Rossier, 2D Materials **4**, 035002 (2017).

[94] T. Moriya, Phys. Rev. **120**, 91 (1960).

[95] W. A. Harrison, *Elementary Electronic Structure* (World Scientific, Singapore, 1999).

[96] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors*, 4th ed. (Springer-Verlag Berlin Heidelberg, 2010).

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

[98] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[99] R. J. Elliott, Phys. Rev. **108**, 1384 (1957).

[100] I. Garate and M. Franz, Phys. Rev. B **84**, 045403 (2011).

[101] C.-H. Park and S. G. Louie, Nano letters **10**, 426 (2010).

[102] L. Ju, L. Wang, T. Cao, T. Taniguchi, K. Watanabe, S. G. Louie, F. Rana, J. Park, J. Hone, F. Wang, and P. L. McEuen, Science **358**, 907 (2017).

[103] M. M. Glazov, L. E. Golub, G. Wang, X. Marie, T. Amand, and B. Urbaszek, Phys. Rev. B **95**, 035311 (2017).

[104] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, and B. Urbaszek, arXiv:1707.05863 (2017).

[105] J. Zhou, W.-Y. Shan, W. Yao, and D. Xiao, Phys. Rev. Lett. **115**, 166803 (2015).

[106] A. Srivastava and A. Imamoğlu, Phys. Rev. Lett. **115**, 166802 (2015).

[107] P. Gong, H. Yu, Y. Wang, and W. Yao, Phys. Rev. B **95**, 125420 (2017).

[108] J. N. Fuchs, F. Piéchon, M. O. Goerbig, and G. Montambaux, Euro. Phys. J. B **77**, 351 (2010).

[109] Y. Gao and Q. Niu, Proc. Natl. Acad. Sci. **114**, 7295 (2017).

[110] B. A. Bernevig and T. L. Hughes, *Topological insulators and topological superconductors* (Princeton University Press, Princeton, NJ, 2013).

[111] R. W. Newson, J. Dean, B. Schmidt, and H. M. van Driel, Opt. Express **17**, 2326 (2009).

[112] M. Danovich, I. Aleiner, N. D. Drummond, and V. Fal'ko, IEEE J. Sel. Top. Quantum Electron **23** (2017).

[113] A. Corney, *Atomic and laser spectroscopy* (Clarendon Press Oxford, 1978).

[114] F. Wang, Y. Wu, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. B **73**, 245424 (2006).

[115] M. Danovich, V. Zólyomi, V. I. Fal'ko, and I. L. Aleiner, 2D Materials **3**, 035011 (2016).

[116] F. Wu, F. Qu, and A. H. MacDonald, Phys. Rev. B **91**, 075310 (2015).

[117] T. G. Pedersen, K. Pedersen, and T. B. Kriestensen, Phys. Rev. B **63**, 201101 (2001).

[118] G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, and D. Xiao, Phys. Rev. B **88**, 085433 (2013).

[119] E. McCann and V. I. Fal'ko, Phys. Rev. Lett. **96**, 086805 (2006).

[120] H. Min and A. H. MacDonald, Phys. Rev. B **77**, 155416 (2008).

[121] L. M. Malard, M. H. D. Guimarães, D. L. Mafra, M. S. C. Mazzoni, and A. Jorio, Phys. Rev. B **79**, 125426 (2009).

[122] L. Fu, Phys. Rev. Lett. **106**, 106802 (2011).

[123] S. Wu, J. S. Ross, G.-B. Liu, G. Aivazian, A. Jones, Z. Fei, W. Zhu, D. Xiao, W. Yao, D. Cobden, and X. Xu, Nat. Phys. **9**, 149 (2013).

[124] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems* (Springer, 2003).

[125] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (2014).

[126] X.-X. Zhang, Y. You, S. Y. F. Zhao, and T. F. Heinz, Phys. Rev. Lett. **115**, 257403 (2015).

[127] X.-X. Zhang, T. Cao, Z. Lu, Y.-C. Lin, F. Zhang, Y. Wang, Z. Li, J. C. Hone, J. A. Robinson, D. Smirnov, S. G. Louie, and T. F. Heinz, Nat. Nanotechnol. **12**, 883 (2017).

[128] B. Scharf, G. Xu, A. Matos-Abiague, and I. Žutić, Phys. Rev. Lett. **119**, 127403 (2017).

[129] C. Robert, T. Amand, F. Cadiz, D. Lagarde, E. Courtade, M. Manca, T. Taniguchi, K. Watanabe, B. Urbaszek, and X. Marie, Phys. Rev. B **96**, 155423 (2017).

[130] M. Combescot, O. Betbeder-Matibet, and R. Combescot, Phys. Rev. Lett. **99**, 176403 (2007).

[131] R. Anankine, M. Beian, S. Dang, M. Alloing, E. Cambril, K. Merghem, C. G. Carbonell, A. Lemaitre, and F. Dubin, Phys. Rev. Lett. **118**, 127402 (2017).

[132] K. H. Lee, C. Lee, H. Min, and S. B. Chung, arXiv:1710.00841 (2017) .

[133] E. Poem, Y. Kodriano, C. Tradonsky, N. H. Lindner, B. D. Gerardot, P. M. Petroff, and D. Gershoni, Nature Physics **6**, 993 (2010).

[134] I. Schwartz, D. Cogan, E. R. Schmidgall, Y. Don, L. Gantz, O. Kenneth, N. H. Lindner, and D. Gershoni, Science **354**, 434 (2016).

[135] T. Cao, M. Wu, and S. G. Louie, Phys. Rev. Lett. **120**, 087402 (2018).

[136] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio, *Group theory: application to the physics of condensed matter* (Springer Science & Business Media, 2007).