Title
Engineering Low Dimensional Materials with van der Waals Interaction

Author
Jin, Chenhao

Publication Date
2017

Peer reviewed|Thesis/dissertation

Engineering Low Dimensional Materials with van der Waals Interaction

By

Chenhao Jin

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Physics
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Feng Wang, Chair
Professor Michael F. Crommie
Professor Junqiao Wu

Spring 2017

Abstract

Engineering Low Dimensional Materials with van der Waals Interaction

By

Chenhao Jin

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Feng Wang, Chair

Two-dimensional van der Waals materials grow into a hot and big field in condensed matter physics in the past decade. One particularly intriguing thing is the possibility to stack different layers together as one wish, like playing a Lego game, which can create artificial structures that do not exist in nature.

These new structures can enable rich new physics from interlayer interaction: The interaction is strong, because in low-dimension materials electrons are exposed to the interface and are susceptible to other layers; and the screening of interaction is less prominent. The consequence is rich, not only from the extensive list of two-dimensional materials available nowadays, but also from the freedom of interlayer configuration, such as displacement and twist angle, which creates a gigantic parameter space to play with.

On the other hand, however, the huge parameter space sometimes can make it challenging to describe consistently with a single picture. For example, the large periodicity or even incommensurability in van der Waals systems creates difficulty in using periodic boundary condition. Worse still, the huge superlattice unit cell and overwhelming computational efforts involved to some extent prevent the establishment of a simple physical picture to understand the evolution of system properties in the parameter space of interlayer configuration.

In the first part of the dissertation, I will focus on classification of the huge parameter space into subspaces, and introduce suitable theoretical approaches for each subspace. For each approach, I will discuss its validity, limitation, general solution, as well as a specific example of application demonstrating how one can obtain the most important effects of interlayer interaction with little computation efforts. Combining all the approaches introduced will provide an analytic solution to cover majority of the parameter space, which will be very helpful in understanding

the intuitive physical picture behind the consequence of interlayer interaction, as well as its systematic evolution in the parameter space.

Experimentally, optical spectroscopy is a powerful tool to investigate properties of materials, owing to its insusceptibility to extrinsic effects like defects, capability of obtaining information in large spectral range, and the sensitivity to not only density of states but also wavefunction through transition matrix element. Following the classification of interlayer interaction, I will present optical spectroscopy studies of three van der Waals systems: Two-dimensional few layer phosphorene, one-dimensional double-walled nanotubes, and two-dimensional graphene/hexagonal Boron Nitride heterostructure. Experimental results exhibit rich and distinctively different effects of interlayer interaction in these systems, as a demonstration of the colorful physics from the large parameter space. On the other hand, all these cases can be well-described by the methods developed in the theory part, which explains experimental results quantitatively through only a few parameters each with clear physical meaning. Therefore, the formalism given here, both from theoretical and experimental aspects, offers a generally useful methodology to study, understand and design van der Waals materials for both fascinating physics and novel applications.

To My Parents

Donghua Chen & Guohua Jin

Table of contents

Acknowledgements

I am grateful to my advisor, Prof. Feng Wang, for his continuous instructions and support over the five years. I have learnt a lot from him; besides numerous techniques and knowledge, what I feel particularly precious is the rigor in doing science and the importance of bigger physics picture. It is really a great and unique experience of doing research here.

I would like to also thank Prof. Y.R. Shen, who is always glad to offer guidance on all aspects. It is with both great pleasure and pressure to talk to him, because his questions are so critical from his sharp vision. I feel so lucky to have the chance to closely observe the high standards from a great scientist like him, which always inspires me to improve myself.

It is also extremely fortunate to have all the colleagues in the Wang group. In junior years I was in close collaboration with Dr. Kaihui Liu and Dr. Zhiwen Shi, two fellow postdoctoral researchers, who introduced me to the field of low dimensional materials such as carbon nanotubes and graphene. I really appreciate the guidance they gave me when I first entered the world of graduate research, and the nice atmosphere of collaboration. Dr. Xiaoping Hong and Dr. Long Ju shared a lot of tips and tricks in building optical setups with me, and I was deeply impressed when watching them assembling optical components in a thoughtful and flexible way, into a powerful tool to realize fascinating physics. I really want to thank Dr. Sufei Shi. He is like a big brother to me, helping and encouraging me in all aspects. In senior years I was working closely together with Dr. Jonghwan Kim, who is one of the kindest people I ever met. He is always ready to help everyone, and takes lots of efforts to make the group more organized. His encouragements helped me to get through the difficult time when experiments did not go smoothly. I would also like to thank Iqbal Utama for the help on device fabrication. And lastly I want to thank all the other fellow students and postdocs in Wang group for their support and help. They are Dr. Huilin Han, Dr. Yaqing Bie, Dr. Yinghui Sun, Dr. Tsung-Ta Tang, Dr. JiHun Kang, Dr. Rai Takahashi Kou, Dr. Chaw Keong Yong, Dr. Hans Kleemann, Dr. Sihan Zhao, Dr. Lili Jiang, Dr. Chan Shan Yang, Dr. Jason Horng, Dr. Bo Zeng, Steve Drapcho, Sheng Wang, Halleh B. Balch, Tairu Lyu, Emma Regan, Matthew Kam, Tao Zhao, Puiyee Lee.

It was great pleasure to work with different group in Berkeley. I want to thank Prof. Alex Zettl and Dr. Qin Zhou from Zettl group for the close and long-term collaboration. I also want to thank Prof. Junqiao Wu and Dr. Joonki Suh from Wu group for the help on device fabrication.

Chapter 1 – Introduction

1.1 Background and Motivation

The emergence of two-dimensional (2D) materials is one of the major developments in condensed matter physics for the last decade, as it marked the capture of the last new land in spatial dimensional, decades after the development on zero-dimensional quantum dot and one-dimensional nanowire/tubes. It is impressive to see that this field is still growing after more than 10 years since the first discovery of monolayer graphene. Improving material quality and searching for new 2D materials are of course always among the objectives; the Holy Grail to fabricate 2D integrated circuit with ultimate scalability and potentially higher power efficiency is also a persistent goal in pursuit. Another unique aspect of 2D materials that greatly contribute to their prosperity, however, is the virtually infinite new possibilities from combining layers into artificial van der Waals systems.

The interlayer interaction is not only strong, leading to significant modification of material properties; but also rich, forming a gigantic parameter space through choices of materials and their alignment. Fascinating physics has indeed been observed from the huge parameter space, such as the versatile engineering of material properties through different layer configuration. Even more interestingly, parts of the parameter space can be difficult to reach in natural system, enabling realization of long-sought models such as the Hofstadter butterfly pattern. The dynamics of excited states can also be efficiently tuned through interlayer charge transfer or electron-phonon interaction.

These exciting new possibilities emphasize the necessity of a proper language to describe van der Waals systems. Given the huge parameter space, it would be helpful if the language can provide an intuitive picture of how system properties are evolving in the parameter space through, for example, pinpointing and isolating the most important part in the interaction. Various candidates for this goal have been around in literatures, which, nevertheless, can often create confusions without specifying the range of application. In the theory part, I aim to provide a set of approaches that are consistent with each other, because they are essentially coming from the same parent methodology; however each particularly suitable for describing specific region in the parameters space. In this way, confusion between different approaches can be removed and one would be able to find a physically-intuitive approach almost anywhere in the parameter space.

In the experiment part, I want to demonstrate the great potential of van der Waals systems from actual observation of fascinating physics with optical spectroscopy studies. In addition, the comparison to theory part confirms the accuracy and usefulness of formalism developed.

1.2 Outline of Thesis

The dissertation starts with a series of theory-oriented discussions in chapter 2, in which I will classify the large parameter space of interlayer interaction in van der Waals materials, and introduce suitable approaches in each case.

Section 2.2 introduced the most general approach, direct diagonalization, based on tight binding model. Section 2.3 classifies the parameter space into commensurate and incommensurate systems, from which we find deficiency of direct diagonalization in the latter case. Therefore, section 2.4 introduces the momentum space diagonalization method, which can further be expanded to different orders, with the two leading orders corresponding to the "direct mixing" and "effective potential" approaches discussed in section 2.5. With the expansion, the criterion of "weak coupling" and "strong coupling" can also be naturally derived. While the momentum space diagonalization method based on the "global picture" is generally suitable for the "weak coupling" case, a "local picture" is introduced in section 2.6 as an appropriate tool in the "strong coupling" regime. The equivalence between the two pictures is also discussed.

Following chapters will focus on experimental studies with optical spectroscopy to demonstrate the great potential of van der Waals systems in engineering material properties in different regimes of the parameter space.

Chapter 3 reveals the conveniently tunable bandgap and electronic structure in few layer phosphorene. Theoretically, this corresponds to the simplest case in "direct mixing", i.e. "direct mixing" at zero momentum, allowing the description of optical resonances in arbitrary layer phosphorene with single master equation. Experimentally, optical absorption spectroscopy is performed to directly demonstrate a tunable bandgap from 0.34-1.73eV in phosphorene, as well as extra optical resonances above the bandgap from higher subband transitions. The nice match between theoretical and experimental results confirms the reliability of both approaches and enables the determination of interlayer coupling strength. These observations suggest phosphorene as versatile platform for applications in both electronic and optoelectronics from the infrared to visible spectral range, and provide guidance to further engineering its properties.

Reference: (*1*). Likai Li*, Jonghwan Kim*, Chenhao Jin*, ***et al.*** *Nat Nano* **12**, 21 (2016)

Chapter 4 shows another example of "direct mixing", though much more complicated than in chapter 3, that electronic structure of 1D incommensurate double-walled nanotubes (DWNTs) can be strongly modified by intertube electronic interaction. Experimentally, through both TEM and optical absorption measurement on the same individual DWNT, we simultaneously determine the atomic and electronic structures of the DWNT, and observe a large shift of optical transition energies in DWNT compared to constituent single-walled nanotube. The transition energy shifts exhibit an intricate and interesting family pattern, which can be well described by theory developed here. These observations for the first time demonstrate strong intertube electronic hybridization in incommensurate 1D system, as opposed to previous belief that

intertube electronic coupling between incommensurate DWNT is negligible; therefore providing key information to understand and engineer the electronic and optical properties of 1D materials.

Reference: (*2*). Kaihui Liu*, Chenhao Jin*, *et al. Nat Phys* **10**, 737 (2014)

Chapter 5 studies graphene/hBN as one representative van der Waals heterostructure, where the "effective potential" approach applies. Interestingly, discussions in chapter 2 suggest this effective potential to have a spinor, rather than scalar nature, which directly accesses graphene's pseudospin degree of freedom. Experimentally, micro-infrared optical absorption spectroscopy was performed on the graphene/hBN heterostructure. With the capability of probing not only density of states but also wavefunction, the peculiar response from optical spectroscopy directly confirms the spinor nature of the effective potential and the dominant role of the "pseudospin-mixing" component, which opens an "inverse gap" in the Moiré Brillouin Zone. The "pseudospin-mixing" potential and its nontrivial effect on the pseudospin structure of graphene are virtually impossible to achieve with applied external field. These results therefore demonstrate the unique capabilities of van der Waals systems to generate unusual type of effective potential allowing versatile engineering on the properties of low-dimensional materials.

Reference: (*3*). Zhiwen Shi*, Chenhao Jin*, *et al. Nat Phys* **10**, 737 (2014)

Chapter 6 briefly discusses some more recent results not included in the main body of the dissertation. These contents consider a broader scope of interlayer interaction, such as the interlayer electron-phonon interaction and interlayer charge transfer process, and are also closely related to discussions in previous chapters; therefore can be of great interest in future studies.

References: (*4, 5*). Chenhao Jin*, Jonghwan Kim*, *et al. Nat Phys* **13**, 127 (2017), Jonghwan Kim*, Chenhao Jin*, *Arxiv*, 1612.05359 (2017)

Chapter 2 – Tuning Single Particle Eigen-states in van der Waals structures

2.1 Introduction

Quantum mechanics is one of the most significant progresses in the physics history, which allows accurate description of motion of particles, e.g. electrons, with simple and elegant forms: By writing down the Hamiltonian and solving Schrödinger or Dirac equations, the eigenstates of the system, both energy and wavefunction, can in principle be fully obtained. Because the behavior of a system, such as electrical, optical, thermal, magnetic properties etc., are often determined by the response of the system to external perturbations, and ultimately by the motion of particles constituting the system; Obtaining the eigenstates of a system is usually the first step to understand its behavior theoretically, and tuning the eigenstates of a system is one of the most important methods to engineer its properties.

For a simple system like non-interacting Fermi gas, its eigenstates can be readily calculated, so are its physical properties. However, when the interaction between particles is turned on, the difficulty starts increasing exponentially with particle number. While motion of two interacting particles can still be solved with pencil and paper within an hour, three particles will already need numerical assistance. For a macroscopic system which contains billons of trillions of particles, the time it takes to solve their motion with brute force would be much longer than the age of the universe.

Fortunately, a few approximations can be made to significantly simply the problem. Born-Oppenheimer approximation allows for separation of interactions at different energy scale, and isolating the equation of motion for electronic part alone. Fermi liquid theory states that the ground state of an interacting Fermi system can be adiabatically connected to a non-interacting system; and that the interacting electrons can be considered as non-interacting quasi-particles with renormalized dynamical properties. These approximations work quite well in many real systems, e.g. crystals, which make them quite similar to the simple free electron gas, with one difference: Periodicity. While free electron gas has translational symmetry with arbitrary displacement, the Hamiltonian of a crystal is only symmetric to a displacement of multiple lattice vectors.

It is therefore clear that the periodicity of a crystal plays an important role in determining its eigenstates. From this perspective, low dimensional van der Waals materials are really a unique platform since their periodicity can be conveniently tuned through artificially stacking layers into specific alignment; and the effects of such periodicity will be strong enough as all the electrons are at the interface and therefore sensitive to the interaction between layers. By tuning the strength and pattern of interlayer coupling, not only properties of materials can be significantly

modified, but also rich physical phenomena can emerge, among which many are difficult to be realized in traditional systems. On the other hand, such overly large parameter space makes it sometimes confusing to correctly describe the effects of interlayer interaction, since there are multiple ways and each may be only valid within certain range.

In this chapter, I will try to provide a comprehensive description of the interlayer interaction in low dimensional van der Waals systems by classifying the interaction into several parameter subspaces, and providing qualitative picture and quantitative analyzing methods suitable for each region.

To achieve this, we first need the tool to obtain eigenstates of systems at arbitrary parameter configuration (of interlayer coupling strength and pattern). The tool I will use here is tight binding model. The following Section 2.2 provides a brief review of important results of tight binding model, which will be used throughout the chapter.

2.2 Review of Tight Binding Model

2.2.1 General solution to periodic system: Direct Diagonalization

Discussion of tight binding model can be found in almost every solid state physics textbook. Therefore here I will not provide a detailed and rigorous derivation. Instead, a practical approach will be used.

Given H to be the Hamiltonian of a crystal, Schrödinger equation reads $H\psi = E\psi$ (2-1), where E and ψ are the eigen-energy and wavefunction, respectively. The crystal is composed of atoms tightly bound to ions. Therefore it is convenient to use the wavefunction of isolated atoms as the basis. i.e. $\psi(\boldsymbol{r}) = \sum_{\boldsymbol{R}_n} b_n \varphi(\boldsymbol{r} - \boldsymbol{R}_n) = b_n \varphi_n$ (2-2). Here $\varphi(\boldsymbol{r} - \boldsymbol{R}_n)$ is the local wavefunction (or local orbital) of the atom at $\boldsymbol{R}_n$, b_n is its coefficient. Einstein convention is used to simplify notation of summation. Multiply Eq. (2-1) by φ_m on left and use Eq. 2-2, we have:

$$\varphi_m H\psi = b_n \varphi_m H\varphi_n = b_n E\varphi_m\varphi_n \quad (2\text{-}3)$$

Define $t_{mn} = \varphi_m H\varphi_n$, $s_{mn} = \varphi_m\varphi_n$, Eq. 2-3 can be rewritten as:

$$(t_{mn} - Es_{mn})b_n = 0 \quad (2\text{-}4)$$

Eq. 2-4 naturally correspond to a matrix equation $(T - ES)X = 0$ (2-5), where the elements of matrix T and S are given by t_{mn} and s_{mn}, respectively. $X = (b_1, b_2, \dots)^{\mathrm{T}}$. Solving Eq. 2-5 is standard, and the eigen energy can be obtained directly from $\det|T - ES| = 0$.

Note that s_{mn} physically represents the wavefunction overlap between two local orbitals at different sites. Assuming the local orbitals are local enough that the above overlap is very small, we have $s_{mn} = \delta_{mn}$ and $S = I$. With such simplification, Eq. 2-5 becomes $TX = EX$ (2-6): Solving eigen-energy and wavefunction of the system becomes equivalent to the eigenvalue problem of matrix T. Therefore, we may call this method **Direct Diagonalization**.

Eq. 2-5 is generally true for system with arbitrary (or no) periodicity. However, if the system consists of N atoms, an N by N matrix is necessary to be solved. For a macroscopic system, this is still not possible. However, by taking advantage of translational symmetry, we can further simplify the calculation: for system with certain periodicity, the momentum $\boldsymbol{k}$ is a good quantum number. According to Bloch theorem, Eq. 2-2 can be reduced to

$$\psi^{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_l} \sum_n b_n^{\boldsymbol{k}} \varphi_n(\boldsymbol{r} - \boldsymbol{R}_l) e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} \quad (2\text{-}7)$$

Here $\boldsymbol{R}_l$ is arbitrary lattice vector, n labels all different local orbitals within a unit cell (either multiple orbitals from the same atom or multiple atoms from the same unit cell). The above equation have clear physical meaning: In a periodic system, the summation over all local orbitals can be done into two steps; first summing within a unit cell (over n), then summing over all unit cells (over $\boldsymbol{R}_l$). Following the same procedure as in Eq. 2-3, we now obtain:

$$H\psi^{\boldsymbol{k}} = \sum_{\boldsymbol{R}_l}\sum_{n} b_n^{\boldsymbol{k}}\varphi_m(\boldsymbol{r}-\boldsymbol{R}_s)H\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} = E\sum_{\boldsymbol{R}_l}\sum_{n} b_n^{\boldsymbol{k}}\varphi_m(\boldsymbol{r}-\boldsymbol{R}_s)\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l}$$

Again assuming $\varphi_m(\boldsymbol{r}-\boldsymbol{R}_s)\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l) = \delta_{mn}\delta_{ls}$, the above equation becomes:

$$\sum_{n} b_n^{\boldsymbol{k}}\sum_{\boldsymbol{R}_l}\varphi_m(\boldsymbol{r}-\boldsymbol{R}_s)H\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} = b_n E e^{i\boldsymbol{k}\cdot\boldsymbol{R}_s}$$

$$\sum_{n} b_n^{\boldsymbol{k}}\left[\sum_{\boldsymbol{R}_l}\varphi_m(\boldsymbol{r}-\boldsymbol{R}_s)H\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_l-\boldsymbol{R}_s)} - E\delta_{mn}\right] = 0$$

With translational symmetry, $\boldsymbol{R}_s$ can be arbitrary lattice vector. Choosing $\boldsymbol{R}_s$ to be 0, we have

$$\sum_{n} b_n^{\boldsymbol{k}}\left[\sum_{\boldsymbol{R}_l}\varphi_m(\boldsymbol{r})H\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} - E\delta_{mn}\right] = 0$$

which again corresponds to solving the eigen value and vector of matrix T, where

$$T_{mn}^{\boldsymbol{k}} = \textstyle\sum_{\boldsymbol{R}_l}\varphi_m(\boldsymbol{r})H\varphi_n(\boldsymbol{r}-\boldsymbol{R}_l)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} \quad (2\text{-}8)$$

2.2.2 Few remarks

Eq. 2-6 takes the same form as the Schrödinger equation, because it is simply the matrix form of Schrödinger equation in the basis of local orbitals. Comparing Eq. 2-6 and Eq. 2-1 directly gives the second quantization form of the Hamiltonian: $H = \sum_{m.n} t_{mn}C_m^+C_n$, where C_m^+ and C_m are creation and annihilation operators of local orbital m.

Comparing Eq. 2-8 to 2-6, the problem of solving a matrix containing all orbitals within the system, which is impractical, has been reduced to solving a matrix containing only orbitals within a unit cell, which is often doable. As an extreme example, regardless of dimension, if a simple lattice has only one atom in each unit cell, and only one relevant orbital from each atom, then only a 1x1 matrix needs to be solved, which actually does not need to be solved at all.

Since multiple orbitals from the same atom and multiple atoms in the same unit cell are on an equal footing in Eq. 2-8, from now on we assume that only one orbital is involved from each atom. To calculate matrix element $T_{mn}^{\boldsymbol{k}}$, Eq. 2-8 in principle requires summing over the coupling between atom m in a given unit cell and atom n in each unit cell. However, since both atom m and n have localized wavefunction, it is natural that their coupling will decay very fast over their spatial separation. In fact, only a few unit cells need to be considered for atom n where it is close enough to the given atom m. Depending on how fast the coupling decay over space and how accurate the result need to be, different approximations can be made. For example, the nearest-

neighbor (NN) approximation assumes that only two atoms next to each other have finite hopping, and will contribute to $T_{mn}^{\boldsymbol{k}}$.

In reality, if N atoms exist in a unit cell, we can follow the below steps to write down the matrix element T_{mn}:

1. Choose any unit cell (which cannot changed afterwards), label each atom with 1,2, …, N.

2. Starting from atom 1 in the chosen unit cell, find all atoms close enough to it so that they will have finite coupling.

3. For each of the atom found with finite coupling, determine its label j, and add the coupling to matrix element T_{1j}. If this atom belongs to a different unit cell from the chosen one, and the lattice vector connecting them is $\boldsymbol{R}_l$, then an additional phase factor $e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l}$ needs to be added.

4. In step 3, note that the "coupling " between atom j with itself (i.e. with atom j in the exactly same unit cell), $E_j = \varphi_j H \varphi_j$, is roughly the energy of isolated atom j, which will be added to the diagonal term T_{jj}.

5. Keep doing step 3 for all atoms from the chosen unit cell.

2.3 Classification of Interlayer Interaction, I: Commensurate and Incommensurate system

2.3.1 Example of Direct Diagonalization: Mono and Bilayer Graphene

The formalism developed in last section is generally applicable to any system, certainly including van der Waals structures stacked from multiple layers: we can write out the matrix T for the overall, multi-layer system and solve it as a whole. This method is valid for all pattern and strength of interlayer interaction, and can in principle produce all the information we need.

Indeed, for many layered systems, the above description is good enough. Taking graphene as an example: For monolayer graphene, each unit cell has 2 atoms (usually called A and B sublattices). Following the steps from last section, the matrix $T^{\boldsymbol{k}}$ (or $H^{\boldsymbol{k}}$) should be 2-by-2, which reads:

$$H_{ML}^{\boldsymbol{k}} = \begin{pmatrix} E_A & -\gamma_0 f(\boldsymbol{k}) \\ -\gamma_0 f^*(\boldsymbol{k}) & E_B \end{pmatrix}, f(\boldsymbol{k}) = 1 + e^{i\boldsymbol{k}\cdot\boldsymbol{a_1}} + e^{i\boldsymbol{k}\cdot\boldsymbol{a_2}}$$

Here only nearest neighbor interaction is considered, $\boldsymbol{a_1}, \boldsymbol{a_2}$ are real space primitive lattice vectors of graphene; $-\gamma_0$ is the hopping parameter between two neighboring carbon atoms, where γ_0~3eV.

For bilayer graphene, the matrix H can be written in exactly the same way, except for the fact that now there are 4 atoms (A1, B1 from 1st layer and A2, B2 from 2nd layer) in each unit cell; and therefore H should be 4-by-4 matrix, which takes the form(*6, 7*):

$$H_{BL}^{\boldsymbol{k}} = \begin{pmatrix} E_{A1} & -\gamma_0 f(\boldsymbol{k}) & \gamma_4 f(\boldsymbol{k}) & \gamma_3 f^*(\boldsymbol{k}) \\ -\gamma_0 f^*(\boldsymbol{k}) & E_{B1} & \gamma_1 & \gamma_4 f(\boldsymbol{k}) \\ \gamma_4 f^*(\boldsymbol{k}) & \gamma_1 & E_{A2} & -\gamma_0 f(\boldsymbol{k}) \\ \gamma_3 f(\boldsymbol{k}) & \gamma_4 f^*(\boldsymbol{k}) & -\gamma_0 f^*(\boldsymbol{k}) & E_{B2} \end{pmatrix}$$

While more complicated than monolayer case, the above matrix is still obtained following the exact process in section 2.2.2: we label the four atoms from 1 to 4, and write done the coupling between each two of them as the matrix element. All the γ symbols are just parameters describing interaction between different pair of atoms. e.g. $H_{23} = \gamma_1$ describes the coupling between B1 and A2 atoms, which corresponds to an interlayer coupling term.

Both $H_{ML}^{\boldsymbol{k}}$ and $H_{BL}^{\boldsymbol{k}}$ can be solved analytically to obtain the band structure and eigen-wavefunction of mono- and bi-layer graphene, which has proven amazingly successfully in understanding physical properties of graphene. By expanding $H_{ML}^{\boldsymbol{k}}$ around $\boldsymbol{k} = \boldsymbol{K}$, so that $f(\boldsymbol{k}) = p_x - ip_y$ ($\boldsymbol{p} = \boldsymbol{k} - \boldsymbol{K}$), $H_{ML}^{\boldsymbol{k}}$ can be rewritten into the form of Weyl equation for ultra-

relativistic particles, which essentially gives the linear Dirac cone dispersion and all the spinor physics.

Similar expansion can be done for $H_{BL}^{\boldsymbol{k}}$ that clearly demonstrates the effect of interlayer interaction. For example, at $\boldsymbol{k} = \boldsymbol{K}$ so that $f(\boldsymbol{k}) = 0$, for monolayer this requires that both bands are degenerate at energy 0 (Assuming that two sublattices are the same and thus $E_A = E_B$). In bilayer, however, the interlayer interaction γ_1 leads to a partial lift of degeneracy: two of the four bands are raised/lowered to the energy of $\pm\gamma_1$due to antibonding/bonding between sublattice B1 and A2, while the other two bands from A1 and B2 remain at energy 0. When $\boldsymbol{k}$ is slightly away from $\boldsymbol{K}$, with $\boldsymbol{p} = (\boldsymbol{k} - \boldsymbol{K})$ expansion we can readily obtain the analytical form of the band dispersion, which turns out to be a parabolic dispersion corresponding to an effective mass $\sim\gamma_1/\gamma_0^2$. Detailed discussion can be found in various references.

2.3.2 Commensurate and Incommensurate System

Given the success of direct diagonalization, it is naturally to ask when we need additional methods, or if we need them at all. The answer is: YES.

Considering a twisted bilayer graphene with twist angle θ. Apparently, the unit cell size of the bilayer system (which is the common unit cell of both layers) depends on angle θ. When $\theta = 0$, the unit cell size is the same as monolayer, such as the Bernal-stacked case in last section, for which direct diagonalization is a good approach. However, the bilayer unit cell (BUC) size can be much larger than that of monolayer (MUC), which can contain thousands of atoms. This means that a matrix of the order of thousands needs to be diagonalized. Worse still, the bilayer unit cell size can be infinite, for example when θ is infinitely small. Since we cannot diagonalize a matrix with infinitely large dimension, the direct diagonalization method automatically fails.

The last case is often considered the "**strict**" definition of incommensurate system. Just like two incommensurate numbers do not have a finite common multiple, two periodic system are incommensurate if no finite common periodicity can be found. Therefore, direct diagonalization is apparently not suitable for "Strict" incommensurate system.

Even for the second case where the bilayer unit cell has medium size, the direct diagonalization method also has problems. Although a matrix with dimension of thousands is solvable with the help of modern computer, such size is large enough to prevent any analytical solution even with expansion around certain momentum. This is not always a problem, but does make it harder to build an intuitive physics picture that relates the system configuration to its properties, since now solving the eigen-states of the system becomes a process in black box with pure computational efforts. Especially often when the properties of each isolated layer is elegantly simple and well-studied, it is counter-intuitive to completely neglect those beautiful results and do everything by brute force.

Another apparent problem of direct diagonalization method is its discontinuity. For example, if certain twist angle θ gives a bilayer unit cell size with one thousand atoms, by increasing the angle to $(\theta + \delta)$, where δ is small, the new bilayer unit cell can have one million, one billion, or even infinite number of atoms as long as δ is small enough. On the other hand, if we smoothly change θ in the parameter space, the properties of the system should also change smoothly, unless encountering some phase transitions or quantum critical points. Therefore, we would expect that small δ will not affect much the properties of the system, and that diagonalizing the above matrices with dimension of either one thousand, one million, one billion or infinite should give almost the same results.

The above case shows an intrinsic problem of direct diagonalization method, and actually it is also a problem for the definition of incommensurate system, since incommensurate and commensurate systems can be infinitely close in parameter space and in their properties. Therefore, here I prefer to introduce a "**practical**" version of the definition: the system can be considered "incommensurate" if the relative position between atoms from one layer and the other layer are almost random. This definition is compatible with the strict one. Taking one-dimensional case as an example, where the periodicity is simply the length of unit cell. If two layers are strictly incommensurate, the local offset between interlayer atoms will never be the same (or they will be commensurate). Therefore when we consider a long enough total length, the local offset will be approximately random. On the other hand, Bernal-stacked graphene is definitely not incommensurate as relative configuration between interlayer atoms are constant everywhere.

All the notion of "incommensurate" in the following sections are referring to the practical definition if without specification. At this point, such definition is still very vague, e.g. how random is "almost random". In section 2.4.2 we will see further, and more quantitative discussion.

Now we can conclude that, direct diagonalization is a good method for commensurate system, but not necessarily for incommensurate system. We should bear in mind that, however, the problems of direct diagonalization **do not mean that it is wrong**. In the next sections, we will try to develop alternative languages specifically for incommensurate system, with certain approximations. These methods can have advantages such as that they can require less computation efforts, better represent the nature of interlayer interaction, and provide more intuitive picture into the system properties. Nevertheless, they should still be all compatible with the direct diagonalization method. Therefore, it is often helpful to combine multiple methods together. For example, we can use other methods to obtain a simple and intuitive picture of some most important properties of the system; and then use direct diagonalization to confirm that all these results are valid.

2.4 Momentum Space Diagonalization

2.4.1 Isolation of Interlayer Interaction

As we discussed in the last section, direct diagonalization, the most universal method, might not be the best one in all situations. One fundamental reason is that it treats intralayer and interlayer interaction on an equal footing, while the parameter space we want to (and we can) explore is only for interlayer interaction, since we cannot modify much to intralayer interaction. Also, it will be desirable if we can start with the well-known monolayer and investigate how it is modified by interlayer interaction, instead of starting over from the beginning.

These considerations make it advantageous to isolate the interlayer interaction to discuss its effects at different parameters. Taking a system with two layers as example, the isolation of interlayer interaction is quite straightforward.

Following Eq. 2-6, assuming the combined bilayers have $(N_1 + N_2)$ atoms, where N_1 and N_2 are atoms from each layer. The general eigen state ψ of the system can be written as $\psi = (\psi^{1T}, \psi^{2T})^T$, which has $(N_1 + N_2)$ components. The first N_1 components, ψ^1, and the rest N_2 components, ψ^2, represents the contribution from atoms from 1st and 2nd layer, respectively.

With the above notation, the Hamiltonian of the whole system is a matrix with order $(N_1 + N_2)$:

$$H = \begin{pmatrix} H_1 & H_{int} \\ H_{int}^+ & H_2 \end{pmatrix} = H_0 + H_I \quad (2\text{-}9)$$

$$H_0 = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}, H_I = \begin{pmatrix} 0 & H_{int} \\ H_{int}^+ & 0 \end{pmatrix}$$

Where H_1 (N_1 by N_1) and H_2 (N_2 by N_2) are the Hamiltonian for isolated 1st and 2nd layer, in which all elements describe interaction between atoms within the same layer; while each element in H_{int} describes the coupling between two atoms from different layers. e.g. interlayer interaction.

If we already know each layer very well, which means that we have already solved all the eigen states of layer 1 $\psi^{1\boldsymbol{k}}(\boldsymbol{r}) = C_{\boldsymbol{k}}^{1+}|0\rangle$ with energy $E_{\boldsymbol{k}}^1$. Then we have $H_1 = \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}^1 C_{\boldsymbol{k}}^{1+} C_{\boldsymbol{k}}^1$, or $\psi^{1\boldsymbol{k}} H_1 \psi^{1\boldsymbol{k'}} = E_{\boldsymbol{k}}^1 \delta_{\boldsymbol{kk'}}$. Similarly, we have $H_2 = \sum_{\boldsymbol{q}} E_{\boldsymbol{q}}^2 C_{\boldsymbol{q}}^{2+} C_{\boldsymbol{q}}^2$, or $\psi^{2\boldsymbol{q}} H_1 \psi^{1\boldsymbol{2}} = E_{\boldsymbol{q}}^2 \delta_{\boldsymbol{qq'}}$. In addition, by assuming that wavefunction overlap between two local orbitals at different layers is negligible, i.e. $\psi^{1\boldsymbol{k}} \psi^{2\boldsymbol{q}} = 0$, all the Bloch states $\psi^{1\boldsymbol{k}}$ and $\psi^{2\boldsymbol{q}}$ will form an orthogonal basis. We can therefore try to write and solve the Hamiltonian in this new basis of Bloch states, which can be called **momentum space diagonalization**. Similar to the case in section 2.2, we now need calculate coupling matrix element between two basis states $\psi^{p\boldsymbol{k}}$ and $\psi^{r\boldsymbol{k'}}$, where n, m labels layer and can either be 1 or 2.

The **first advantage** with such basis is that, we have $\psi^{p\boldsymbol{k}}H_0\psi^{r\boldsymbol{k}'} = E^p_{\boldsymbol{k}}\delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{pr}$ and $\psi^{p\boldsymbol{k}}H_I\psi^{p\boldsymbol{k}'} = 0$: the intralayer interaction part, H_0, is purely diagonal and extremely simple; while the interlayer interaction part, H_I only contributes to off-diagonal elements. We can therefore separate the two parts and focus on the interlayer interaction part only.

2.4.2 Gift from Incommensurability

The final task is to calculate $M^{\boldsymbol{k}\boldsymbol{q}} = \psi^{1\boldsymbol{k}}H_I\psi^{2\boldsymbol{q}}$, the off-diagonal matrix element from interlayer interaction. To do this, we first expand the Bloch states back into local orbitals:

$$\psi^{1\boldsymbol{k}} = \sum_{\boldsymbol{R}_1}\sum_n b_n^{\boldsymbol{k}}\varphi_n^1(\boldsymbol{r}-\boldsymbol{R}_1)e^{i\boldsymbol{k}\cdot\boldsymbol{R}_1}$$

$$\psi^{2\boldsymbol{q}} = \sum_{\boldsymbol{R}_2}\sum_m a_m^{\boldsymbol{q}}\varphi_m^2(\boldsymbol{r}-\boldsymbol{R}_2)e^{i\boldsymbol{q}\cdot\boldsymbol{R}_2} \quad (2\text{-}10)$$

Where $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ are summing over unit cells of layer 1 and layer 2 respectively.

$$M^{\boldsymbol{k}\boldsymbol{q}} = \psi^{1\boldsymbol{k}}H_I\psi^{2\boldsymbol{q}} = \sum_{\boldsymbol{R}_1}\sum_n b_n^{\boldsymbol{k}*}\varphi_n^{1*}(\boldsymbol{r}-\boldsymbol{R}_1)e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_1}H_I\sum_{\boldsymbol{R}_2}\sum_m a_m^{\boldsymbol{q}}\varphi_m^2(\boldsymbol{r}-\boldsymbol{R}_2)e^{i\boldsymbol{q}\cdot\boldsymbol{R}_2}$$

$$= \sum_n\sum_m b_n^{\boldsymbol{k}*}a_m^{\boldsymbol{q}}\sum_{\boldsymbol{R}_1}\sum_{\boldsymbol{R}_2} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_1}e^{i\boldsymbol{q}\cdot\boldsymbol{R}_2}\langle\varphi_n^1(\boldsymbol{r}-\boldsymbol{R}_1)|H_I|\varphi_m^2(\boldsymbol{r}-\boldsymbol{R}_2)\rangle$$

$$= \sum_n\sum_m b_n^{\boldsymbol{k}*}a_m^{\boldsymbol{q}}\sum_{\boldsymbol{R}_1} e^{-i(\boldsymbol{k}-\boldsymbol{q})\cdot\boldsymbol{R}_1}\sum_{\boldsymbol{R}_2} e^{i\boldsymbol{q}\cdot(\boldsymbol{R}_2-\boldsymbol{R}_1)}\langle\varphi_n^1(\boldsymbol{r}-\boldsymbol{R}_1)|H_I|\varphi_m^2(\boldsymbol{r}-\boldsymbol{R}_2)\rangle$$

$$= \sum_n\sum_m b_n^{\boldsymbol{k}*}a_m^{\boldsymbol{q}}\sum_{\boldsymbol{R}_1} e^{-i(\boldsymbol{k}-\boldsymbol{q})\cdot\boldsymbol{R}_1}\sum_{\Delta\boldsymbol{R}} e^{i\boldsymbol{q}\cdot\Delta\boldsymbol{R}}\langle\varphi_n^1(\boldsymbol{r})|H_I|\varphi_m^2(\boldsymbol{r}-\Delta\boldsymbol{R})\rangle \quad (2-11)$$

Here $\Delta\boldsymbol{R} = \boldsymbol{R}_2 - \boldsymbol{R}_1$. Eq. 2-11 looks rather complicated, however has a very simple physical meaning: To calculate the coupling matrix element between two interlayer Bloch states, we need sum over the contribution from all interlayer atom pairs (each with a proper phase). This can be done in two steps. First for a given atom in layer 1, we sum its interaction with all atoms in layer 2 (summing over $\Delta\boldsymbol{R}$). Second we sum all atoms within layer 1 (summing over $\boldsymbol{R}_1$). The interaction between each pair of interlayer atoms, $t_{nm}(\Delta\boldsymbol{R}) = \langle\varphi_n^1(\boldsymbol{r})|H_I|\varphi_m^2(\boldsymbol{r}-\Delta\boldsymbol{R})\rangle$, is often called "**pair interaction**". Intuitively, the pair interaction will depend on atom species and the relative displacement between the two atoms.

The **key step to further simplify** Eq. 2-11 is to take advantage of incommensurability. As discussed in section 2.3.2, incommensurability means that the relative displacement $\Delta\boldsymbol{R}$ is random. Therefore, we can replace the first summation over $\Delta\boldsymbol{R}$ in Eq. 2-11 with an integral:

$$\sum_{\Delta\boldsymbol{R}} e^{i\boldsymbol{q}\cdot\Delta\boldsymbol{R}}\langle\varphi_n^1(\boldsymbol{r})|H_I|\varphi_m^2(\boldsymbol{r}-\Delta\boldsymbol{R})\rangle \to \int d\boldsymbol{R}\, e^{i\boldsymbol{q}\cdot\boldsymbol{R}}\langle\varphi_n^1(\boldsymbol{r})|H_I|\varphi_m^2(\boldsymbol{r}-\boldsymbol{R})\rangle$$

$$M^{\boldsymbol{kq}} = \sum_n \sum_m b_n^{\boldsymbol{k}*} a_m^{\boldsymbol{q}} \sum_{\boldsymbol{R}_1} e^{-i(\boldsymbol{k}-\boldsymbol{q})\cdot\boldsymbol{R}_1} \int d\boldsymbol{R}\, e^{i\boldsymbol{q}\cdot\boldsymbol{R}} t_{nm}(\boldsymbol{R}) \quad (2-12)$$

The integral over $\boldsymbol{R}$ goes over the whole two-dimensional surface of layer 2. Such approximation may look weird at first: for a given atoms in layer 1, its displacement from layer 2 atoms are discrete and regular, and therefore definitely not random. However, we need take into account the second summation over $\boldsymbol{R}_1$, that we ultimately will consider all atoms from layer 1 instead of a single given atom. As a result, the above approximation will be valid as long as the interlayer displacement is random after considering all atoms pairs, which is exactly the practical definition of incommensurate system.

Another way to think is, we can spatially shift all the atoms from layer 1 to a given atom, each through a corresponding lattice vector. Then we can also shift the lattice of layer 2 by the same amount (i.e. by all lattice vectors of layer 1), which will result in many lattices with different offsets. Now layer 1 becomes a single atom, and we just need consider the coupling between this single atom in layer 1and all the shifted lattices in layer 2. For strictly incommensurate system, all the shifted lattices of layer 2 will not overlap, and they will ultimately form a continuous and homogeneous distribution of atoms covering everywhere, justifying the integral in Eq. 2-12. This is also approximately true for the "practically" incommensurate system, where all the shifted lattices form a dense but not necessarily continuous lattice; however already good enough to support the approximation in Eq. 2-12.

Eq. 2-12 can now be conveniently calculated, because the summation over $\boldsymbol{R}_1$ and integral over $\boldsymbol{R}$ are independent and can be obtained separately:

$$M^{\boldsymbol{kq}} = \sum_n \sum_m b_n^{\boldsymbol{k}*} a_m^{\boldsymbol{q}}\, t_{nm}(\boldsymbol{q}) \delta_{\boldsymbol{k}+\boldsymbol{G}_1,\boldsymbol{q}} \quad (2\text{-}13)$$

The summation over $\boldsymbol{R}_1$ gives the momentum selection rule, $\delta_{\boldsymbol{k}+\boldsymbol{G}_1,\boldsymbol{q}}$, which restrict states that can couple. Once the selection rule is satisfied, the integral determines the amplitude of the coupling matrix element: simply the Fourier component of the pair interaction, $t_{nm}(\boldsymbol{q})$.

2.4.3 Discussion on Interlayer Coupling Matrix Element

It is quite surprising to see that the complicated expression in Eq. 2-11 ends up with the simple and elegant form in Eq. 2-13. Therefore momentum space diagonalization is intuively a more suitable method for incommensurate system, as it takes full advantage of the incommensurability.

To fully appreciate the power of this method, we need take a closer look of Eq. 2-13. The selection rule requires:

$$\boldsymbol{k}_1 + \boldsymbol{G}_1 = \boldsymbol{q} = \boldsymbol{k}_2 + \boldsymbol{G}_2 \quad (2\text{-}14)$$

where $\boldsymbol{k_1}, \boldsymbol{k_2}$ are two bloch states from layer 1 and layer 2 (restricted in respective first Brillouin zone); and $\boldsymbol{G}_1$, $\boldsymbol{G}_2$ are reciprocal lattice vectors of layer 1 and layer 2, respectively. This selection rule has very simple physical meaning, that the momentum has to conserve during interlayer coupling, with the help of additional momentum provided by lattice of both layers.

The amplitude of matrix element, $t_{nm}(\boldsymbol{q})$, can be directly obtained if the pair interaction $t_{nm}(\boldsymbol{R})$ is given, which can often be done by introducing an empirical potential through comparing to experimental or first-principle calculation results. Regardless of the exact form of $t_{nm}(\boldsymbol{R})$, we know that it should decrease with the increase of R: when the two atoms are more separated, their interaction should be weaker. Therefore, the pair interaction can often be modeled with an exponential decay over separation,

$$t_{nm}(\Delta r) = t_0 e^{-\Delta r/\lambda} \qquad (2\text{-}15).$$

Here for simplicity we assume that the interaction is isotropic, which may not be true for orbitals with in-plane anisotropy such as p_x, p_y. In those cases, an additional in-plane angle needs to be introduced. The generalization, however, will be straightforward.

We note that, Δr in Eq. 2-15 is the **total distance** between two atoms, while in the expression of Eq. 2-12 all the $\boldsymbol{R}$ are referring to **in-plane displacement**. Therefore we have

$$\Delta r = \sqrt{d^2 + R^2}$$

$$t_{nm}(\boldsymbol{R}) = t_0 e^{-\sqrt{d^2+R^2}/\lambda} = t_0 e^{-\sqrt{d^2+x^2+y^2}/\lambda} \approx t_0 e^{-\left(\frac{d}{\lambda}+\frac{x^2+y^2}{2\lambda d}\right)} \qquad (2\text{-}16)$$

Where d is the out-of-plane displacement, i.e. separation between the two layers, which can often be considered a constant and obtained experimentally. λ is the characteristic decay length of interaction, which is often in the order of atomic radii. Usually $d \gg \lambda$, which allows us to further expand the pair interaction as in the last step of Eq. 2-16.

With Eq. 2-16, we can directly obtain the analytical form of $t_{nm}(\boldsymbol{q})$:

$$t_{nm}(\boldsymbol{q}) = \int d\boldsymbol{R}\, e^{i\boldsymbol{q}\cdot\boldsymbol{R}} t_{nm}(\boldsymbol{R}) = t_0 e^{-\frac{d}{\lambda}} \int dx e^{iqx} t_0 e^{-\frac{x^2}{2\lambda d}} \int dy\, e^{-\frac{y^2}{2\lambda d}}$$

$$\sim t_0 e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2}q^2} \qquad (2\text{-}17)$$

Without loss of generality, we have define the direction of $\boldsymbol{q}$ to be x, and the perpendicular in-plane direction to be y.

The important implication from Eq. 2-17 is that, the amplitude of coupling decays very fast with increasing q. This conclusion is generally true for all types of reasonable pair interaction, and is not limited to the specific form in Eq. 2-15.

Combining the above observation and the selection rule in Eq. 2-14, the **second advantage** of momentum space diagonalization now becomes apparent: It can provide a very simple and intuitive, often analytical picture to understand the most important properties of the system.

These properties are often related to electronic states at specific momentum $\boldsymbol{k}$, such as at band extrema, at Fermi level, or in a certain energy range. Now we can selectively isolate states with one or few momentum and investigate their behavior with interlayer interaction.

For example, if we are interested in state $\boldsymbol{k}_1$ from layer 1**,** according to selection rule in Eq. 2-14, we can determine all states $\boldsymbol{k}_2$ in layer 2 that satisfy the selection rule, by drawing a reciprocal lattice of layer 1 centered at $\boldsymbol{k}_1$ in momentum space. All states $\boldsymbol{k}_2$ that can couple to $\boldsymbol{k}_1$ are simply states on this lattice. Furthermore, Eq. 2-17 indicates that, only states that are close enough to **Γ** point (with zero momentum) will have non-negligible coupling amplitude. With a proper cutoff, usually only a few $\boldsymbol{k}_2$ need to be considered. Afterwards, the same procedure can be done iteratively for all $\boldsymbol{k}_2$ involved, until a second cutoff is reached. Such process implies the possibility to classify the interaction based on number of iteration, or equivalently order of expansion, which we will discuss in section 2.5.

The **third advantage** of momentum space diagonalization can also be directly seen from Eq. 2-14 and Eq. 2-17, that the inherent discontinuity problem of direct diagonalization is completely gone. Continuous change of interlayer configuration necessarily leads to smooth change in their lattice vector, and also reciprocal lattice vector $\boldsymbol{G}_1$, $\boldsymbol{G}_2$. Therefore all quantities in the equations will be continuous to interlayer configuration, even after doing derivative up to arbitrary order.

One may argue, however, that being always continuous is not always correct: at some critical points in parameter space, the system may indeed become discontinuous. For example, in twisted bilayer graphene, people both predict theoretically and observe experimentally that the Fermi velocity get strongly renormalized at small twist angle(*8-11*), which does not smoothly dependent on twist angle anymore (after certain order of derivative). How does such discontinuity come out of all the continuous quantities? The answer is simple: In the above situation, although the selection rule and coupling matrix element are still continuous, the iteration process will converge very slowly, and may ultimately not converge. Just like the fact that the sum of an infinite series may not be continuous if it is not convergent, even if all the elements are continuous; the system properties can be discontinuous even if each iteration is continuous. Such situation can be called "**strong coupling regime**", and will be discussed more detailedly in section 2.6.

2.4.4 Example: Indirect Bandgap in Twisted Bilayer MoS2

Combining Eq. 2-14 and 2-17 can in principle give information of Bloch states at any momentum $\boldsymbol{k}$. A especially simple case is when the state of interest is at $\boldsymbol{k} = 0$, i.e. Γ point of the Brillouin Zone. From Eq. 2-14, $\boldsymbol{k}_1 = 0$ state form the 1st layer will always satisfy the selection rule to couple with $\boldsymbol{k}_2 = 0$ state form the 2nd layer. Furthermore, such coupling will be particularly strong according to Eq. 2-17. The strong coupling here can be intuitively understood

as when both $\boldsymbol{k}_1, \boldsymbol{k}_2$ are equal to zero, all the atoms are in-phase; and their pair interaction directly adds up without any cancellation. Therefore, considering only the coupling between $\boldsymbol{k}_1 = 0$ and $\boldsymbol{k}_2 = 0$ states is usually sufficient, and coupling to all other states can be largely neglected.

The situation is actually quite often where **Γ** point property is important, since **Γ** point is often a band extrema as is required by various symmetries. One example is phosphorene that has a direct bandgap at **Γ** point, which will be discussed in detail in chapter 3. Here we take bilayer MoS2 as another example.

It is now well-known that monolayer MoS2 is semiconductor with direct bandgap at K point, and multilayer, including bilayer MoS2 has indirect bandgap(*12, 13*). While the direct bandgap nature makes monolayer transition metal dichalcogenides (often shorted as MX2) an attractive candidate for optoelectronics application, the transition between indirect to direct bandgap depending on layer number is often invoked as an excellent example of how interlayer interaction can efficiently tune the properties of the system.

Such transition can be readily understood with the knowledge we have developed in this section: In bilayer MX2 the **Γ** point energy of valence band is shifted up a lot and becomes the new valence band maximum. The conduction band minimum may also change to away from K point depending on specific materials, but the energy difference is not as large. The energy shift of **Γ** point can be directly obtained by considering the coupling between $\boldsymbol{k}_1 = 0$ and $\boldsymbol{k}_2 = 0$ states, which are simply the same **Γ** point state from each layer, $\psi^{1\boldsymbol{\Gamma}}$and $\psi^{2\boldsymbol{\Gamma}}$. These two states are degenerate to start with, and therefore will split into two energy levels with energy and wavefunction:

$$E_{\pm} = \pm t_0 e^{-\frac{d}{\lambda}}, \psi_{\pm} = \psi^{1\boldsymbol{\Gamma}} \pm \psi^{2\boldsymbol{\Gamma}} \quad (2\text{-}18)$$

The splitting of **Γ** point into two bands can be clearly seen in any DFT band structure calculations of bilayer MX2(*12, 13*). Furthermore, Eq. 2-18 should be generally true for bilayer MX2 with any twist angle. In principle, the energy shift $E_{\pm}$ may slightly deviate from the above value in commensurate case, such as AB stacked bilayer. However, it turns out that even AB and AA stacked bilayer MoS2 still follow the same scaling law. As a result, the shift of **Γ** point state at valence band, which mostly determines the indirect bandgap size in bilayer MX2, is only determined by interlayer separation d, regardless of the twist angle. This poses a sharp contrast to the cases of graphene where states near **K** point are important, because for **Γ** point the twist angle will not affect the momentum selection rule between $\boldsymbol{k}_1 = 0$ and $\boldsymbol{k}_2 = 0$ states.

Another difference between **Γ** point and **K** point states is, as we discussed at the beginning of this section, **Γ** point can have particularly strong interlayer coupling. i.e. Eq. 2-17 becomes maximum when $\boldsymbol{q} = 0$. In addition, the dominant orbitals may also vary with momentum, which will affect t_0, and therefore the coupling amplitude. For example, **K** point states in MX2 are more dominated by d orbitals from M atoms(*14*), which is quite localized and also spatially

separated from the other layer by two layers of X atoms. The above two reasons make **K** point coupling much weaker than **Γ** point, resulting in much less energy shift. Therefore, the **Γ** point at valence band ultimately becomes higher than **K** point in bilayer MX2.

The scaling law for energy shift and indirect bandgap size in twisted bilayer MX2, as well as the much smaller shift of **K** point states, has been confirmed both by experiment and by DFT calculation(*15*).This is a good demonstration of the momentum space diagonalization method we introduce here, that we can obtain all the results in a physically intuitive and computationally effortless way.

2.5 Classification of Interlayer Interaction, II: Direct Mixing and Effective Potential

2.5.1 Relation Between Momentum Space Diagonalization and Perturbation Theory

As we discussed in the section 2.4.3, the momentum space diagonalization method provides a possibility to obtain the modified properties of specific Bloch states at $\boldsymbol{k}$ with expansion. Naturally, we then classify interlayer interaction depending on the leading order in the expansion. In fact, different orders of expansion are related to different orders of perturbation theory. To see such relation, we can first write down the general perturbation theory for a quantum system:

$$\begin{aligned}|p\boldsymbol{k}\rangle' - |p\boldsymbol{k}\rangle = &+\sum_{r_1\boldsymbol{q}_1}|r_1\boldsymbol{q}_1\rangle\frac{\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1}}\\ &+\sum_{r_1\boldsymbol{q}_1}\sum_{r_2\boldsymbol{q}_2}|r_2\boldsymbol{q}_2\rangle\frac{\langle r_2\boldsymbol{q}_2|H_I|r_1\boldsymbol{q}_1\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1})(E_{p\boldsymbol{k}}-E_{r_2\boldsymbol{q}_2})}\\ &-\sum_{r_1\boldsymbol{q}_1}|r_1\boldsymbol{q}_1\rangle\frac{\langle p\boldsymbol{k}|H_I|p\boldsymbol{k}\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{\left(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1}\right)^2}\\ &-\frac{1}{2}|p\boldsymbol{k}\rangle\sum_{r_1\boldsymbol{q}_1}\frac{\langle p\boldsymbol{k}|H_I|r_1\boldsymbol{q}_1\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{\left(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1}\right)^2}\end{aligned}\quad(2-19)$$

Here $|p\boldsymbol{k}\rangle$ is a state of interest in the original system, and $|p\boldsymbol{k}\rangle'$ is the modified state in the perturbed system. All states on the right hand side of the equation are unperturbed states. The first line corresponds to 1st order perturbation in wavefunction, while the other three lines correspond to terms in 2nd order perturbation in wavefunction. Orders equal and higher than 3 are neglected here, but can be written in similar way if necessary.

Eq. 2-19 can be further simplified from the discussion in section 2.4.1: We have $\psi^{p\boldsymbol{k}}H_I\psi^{p\boldsymbol{k}'}=0$, i.e. the interlayer interaction only directly couples states from one layer to the other layer, but not two states of the same layer. As a result , the third line of Eq. 2-19 is simply 0, and Eq. 2-19 can be rewritten as:

$$|p\boldsymbol{k}\rangle' - |p\boldsymbol{k}\rangle = +\sum_{r_1\boldsymbol{q}_1}|r_1\boldsymbol{q}_1\rangle\frac{\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1}}-\frac{1}{2}|p\boldsymbol{k}\rangle\sum_{r_1\boldsymbol{q}_1}\frac{\langle p\boldsymbol{k}|H_I|r_1\boldsymbol{q}_1\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{\left(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1}\right)^2}+\cdots$$

$$+\sum_{r_1\boldsymbol{q}_1}\sum_{r_2\boldsymbol{q}_2}|r_2\boldsymbol{q}_2\rangle\frac{\langle r_2\boldsymbol{q}_2|H_I|r_1\boldsymbol{q}_1\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1})(E_{p\boldsymbol{k}}-E_{r_2\boldsymbol{q}_2})}+\cdots \quad (2-20)$$

Eq. 2-20 has re-arranged the perturbation into two lines. The first line only involves matrix element $\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle$. Comparing to the iteration process in section 2.4.3., this matrix element appears in the first step of the iteration, when we consider all states $|p_1\boldsymbol{k}_1\rangle$ that can directly couple to the given state $|p\boldsymbol{k}\rangle$. And apparently we have $|r_1\boldsymbol{q}_1\rangle = |p_1\boldsymbol{k}_1\rangle$, as they both represents states that directly couple to $|p\boldsymbol{k}\rangle$. The omitted terms at right hand side are those of higher orders, however still only involving $|r_1\boldsymbol{q}_1\rangle$ states.

On the other hand, the second line in Eq. 2-20 involves a new set of states $|r_2\boldsymbol{q}_2\rangle$, which cannot directly couple to $|p\boldsymbol{k}\rangle$. Instead, they only directly couple to states $|r_1\boldsymbol{q}_1\rangle$; and then, by using $r_1\boldsymbol{q}_1$ as an intermediate bridge, couple to $|p\boldsymbol{k}\rangle$. This is closely related to the second step of the iteration process, that we include all states $|p_2\boldsymbol{k}_2\rangle$ that can directly couple to states considered in the first step. Again we have $|r_2\boldsymbol{q}_2\rangle = |p_2\boldsymbol{k}_2\rangle$, as they both represents states that directly couple to $|r_1\boldsymbol{q}_1\rangle$, or $|p_1\boldsymbol{k}_1\rangle$.

The above correspondence can be extended to any order. For example, if we keep writing Eq. 2-20 up to nth line, the last line will take the form:

$$\sum_{r_1\boldsymbol{q}_1}\sum_{r_2\boldsymbol{q}_2}\cdots\sum_{r_n\boldsymbol{q}_n}|r_n\boldsymbol{q}_n\rangle\frac{\langle r_n\boldsymbol{q}_n|H_I|r_{n-1}\boldsymbol{q}_{n-1}\rangle\ldots\langle r_2\boldsymbol{q}_2|H_I|r_1\boldsymbol{q}_1\rangle\langle r_1\boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle}{(E_{p\boldsymbol{k}}-E_{r_n\boldsymbol{q}_n})\ldots(E_{p\boldsymbol{k}}-E_{r_2\boldsymbol{q}_2})(E_{p\boldsymbol{k}}-E_{r_1\boldsymbol{q}_1})} \quad (2-21)$$

Which describes a process that the state $|r_n\boldsymbol{q}_n\rangle$ couples to $|p\boldsymbol{k}\rangle$ through n steps, with (n-1) intermediate states. Apparently, only states able to couple to $|r_{n-1}\boldsymbol{q}_{n-1}\rangle$ need, and should be considered in the above equation. Comparing to the nth step in the iteration, which adds all states $|p_n\boldsymbol{k}_n\rangle$ that can directly couple to states in the (n-1)th step, i.e. $|p_{n-1}\boldsymbol{k}_{n-1}\rangle$, we can reach a conclusion by recursion that $|r_n\boldsymbol{q}_n\rangle \equiv |p_n\boldsymbol{k}_n\rangle$, that the nth step in the iteration and nth order perturbation always introduce exactly the same states.

However, we note that, the close relation above does not mean a one-to-one correspondence between the nth step in the iteration and nth order perturbation. For example, the first line in Eq. 2-20 includes both 1st and 2nd order perturbation term, and we can add even higher order terms afterwards. Nevertheless, as long as only states $|r_1\boldsymbol{q}_1\rangle$ are involved, this type of perturbation is fully captured in the first step of interaction. For example, in section 2.4.4 where $|p\boldsymbol{k}\rangle$ and $|r_1\boldsymbol{q}_1\rangle$ are **Γ** point states from each layer, respectively. The fact that $E_{p\boldsymbol{k}} = E_{r_1\boldsymbol{q}_1}$ makes it impossible to use perturbation theory, which can be clearly seen in the first line of Eq. 2-20. In contrast, in 2.4.4, we already see that everything can be solved by considering two states from just one-step iteration.

2.5.2 Weak, Strong, Perturbative and Non-perturbative

The relation between the iteration process and the perturbation theory now becomes clear: Each step of the iteration corresponds to one line, instead of one term, in Eq. 2-20. Therefore, the case of $|\langle r_1 \boldsymbol{q}_1|H_I|p\boldsymbol{k}\rangle| > \left|E_{p\boldsymbol{k}} - E_{r_1\boldsymbol{q}_1}\right|$, which is often considered as the strong coupling condition in perturbation theory, is **NOT a strong coupling condition** in the iteration method, since the required steps (or lines in Eq. 2-20) may still be only a few.

To avoid confusion, from now on we will introduce a few terms to describe different situations discussed above. We call it **"weak"** if only a few steps in the iteration is required or equivalently, **a few lines** in Eq. 2-20 need to be considered. We call it **"perturbative"** only **a few terms** in Eq. 2-20 need to be considered. We therefore have "**weak perturbative coupling**" condition when a few lines, and a few terms from each line, are important; "**weak non-perturbative coupling**" if at least one line have near-degenerate coupling and cannot be treated perturbative; "**strong coupling**" if many, or even infinite lines are required to fully describe the system. Because infinite lines necessarily mean infinite terms, "**strong coupling**" is always "**non-perturbative**".

Now we can easily see that the example of 2.4.4 belongs to "**weak non-perturbative coupling**". When using momentum space diagonalization method, we will ultimately solve the eigen-values of a matrix involving all relevant Bloch states, which is valid for both **weak perturbative coupling** and **weak non-perturbative coupling** cases. On the other hand, perturbation theory only works for **weak perturbative coupling**. The reason we still want to introduce it here is that when applicable, it can give a particularly simple and potentially intuitive picture to understand the properties of the system. e.g. If only a few terms are important, Eq. 2-20 guarantees an analytical solution to the modified eigen-states in the coupled system when we combine the matrix element obtained in Eq. 2-17. In addition, it is straightforward to separate the contribution from each term, therefore one can conveniently divide the parameter space into partitions where one term may dominates, and pinpoint the most relevant interaction in a specific sub-space.

In this section, we will focus on the **weak coupling** regime, and the **strong coupling** case will be discussed in section 2.6.

2.5.3 Direct Mixing and Effective Potential

Direct Mixing

Similar to perturbation theory that is classified according to the order of perturbation, the iteration process of momentum space diagonalization can also be classified by the number of steps to consider.

In many cases, the first line in Eq. 2-20 is good enough to describe the major change in the system's properties from interlayer interaction. We can call this situation "**Direct Mixing**" as the first line in Eq. 2-20 describes a direct coupling between **two states from different layers**. The modified eigen-wavefunction and the corresponding change in eigen-energy can be particularly interesting for states at specific momentum $\boldsymbol{k}$, such as those at the band extrema; since these states can determine important properties of the system like bandgap, as we have seen in section 2.4.4.

In chapter 3 and 4, I will give two examples that direct mixing can provide a versatile and efficient way to engineer the bandgap in systems such as few layer phosphorene and double-walled nanotubes (DWNTs), which are directly probed and confirmed by optical spectroscopy.

Another situation when specific states become especially important is when they have **non-perturbative coupling**, and in the extreme case, **degenerate coupling**. Generally, while non-degenerate coupling tends to shift or slightly modify the profile of the system's spectrum, degenerate coupling can induce a much more dramatic and fundamental effect, such as emergence of new resonance or bandgap, which can qualitatively change the behavior of the system. One simple and famous example is Peierls instability, in which one-dimensional crystal develops a bandgap from perturbed lattice potential. The bandgap forms at the edge of the Brillouin zone, where degenerate coupling happens.

It is therefore desirable to find all states $\boldsymbol{k}$ with **degenerate coupling** since these states are likely responsible for any qualitatively new behavior of the system. This can be simply done by combining the selection rule from Eq. 2-14 and the degenerate coupling condition: We can plot the energy-momentum dispersions of both layers together in periodic Brillouin Zone scheme, all the states with degenerate coupling will be at crossing points between the two dispersions. This is because, according to Eq. 2-14, only states with the same momentum in the periodic Brillouin Zone can satisfy the momentum selection rule; and they should also have the same energy to satisfy degenerate coupling conditions. This method also allows explicit selection of spectrum range by plotting only the relevant energy range in band dispersion. Similarly, the momentum range, i.e. how many period of Brillouin Zone needs to be plotted, can be determined by comparing Eq. 2-17 to a certain cutoff of interaction strength (for example, one can use the energy resolution of relevant experimental techniques as the cutoff).

One example of how degenerate coupling in direct mixing can give rise to new interesting spectral features in twisted bilayer graphene will be given in section 2.5.6.

Effective Potential

From the above criterion, we can see that in some systems the degenerate coupling may not happen at all from direct mixing. One simple example is that, if layer 2 is a wide bandgap insulator that does not have any electronic state within the energy range of interest, then certainly

there will no interlayer band crossing or degenerate coupling from direct mixing. Especially if $(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}_1}) \gg \langle 2\boldsymbol{q}_1|H_I|1\boldsymbol{k}\rangle$ holds for all states of interest (in layer 1), i.e. the highly perturbative regime, the 1st line in Eq. 2-20 will only lead to a shift or slight change in the spectrum of the system, as we have discussed before. While this shift can still be important for some states such as those at the band extrema; for other states it may not be very interesting, and can often be neglected. On the other hand, the 2nd line in Eq. 2-20 can still have non-trivial effects as the denominator involves $(E_{1\boldsymbol{k}} - E_{1\boldsymbol{q}_1})$, which can be 0 and induces degenerate coupling. In such situation, the major and qualitative change of the system may completely originate from the 2nd line, which describes the mixing between **two states from the same layer**, in sharp contrast with the "Direct Mixing" case.

Once effects of direct mixing can be neglected, the role of layer 2 becomes quite subtle: It only serves as a bridge to mix two states both from layer 1, but not directly involved in mixing. On the other hand, to mix two states from layer 1 can also be done by simply applying an external potential to layer 1. Therefore, we can in principle replace the effect of interlayer interaction with a potential solely acting on layer 1, and completely get rid of layer 2 afterwards. From Eq. 2-20, this can be done by defining the effective potential below:

$$\sum_{\boldsymbol{q}_1}\sum_{\boldsymbol{q}_2}|1\boldsymbol{q}_2\rangle\frac{\langle 1\boldsymbol{q}_2|H_I|2\boldsymbol{q}_1\rangle\langle 2\boldsymbol{q}_1|H_I|1\boldsymbol{k}\rangle}{(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}_1})(E_{1\boldsymbol{k}} - E_{1\boldsymbol{q}_2})} = \sum_{\boldsymbol{q}_2}|1\boldsymbol{q}_2\rangle\frac{\langle 1\boldsymbol{q}_2|V_{eff}|1\boldsymbol{k}\rangle}{(E_{1\boldsymbol{k}} - E_{1\boldsymbol{q}_2})}$$

$$\rightarrow \sum_{\boldsymbol{q}_1}\frac{\langle 1\boldsymbol{q}_2|H_I|2\boldsymbol{q}_1\rangle\langle 2\boldsymbol{q}_1|H_I|1\boldsymbol{k}\rangle}{(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}_1})(E_{1\boldsymbol{k}} - E_{1\boldsymbol{q}_2})} = \frac{\langle 1\boldsymbol{q}_2|V_{eff}|1\boldsymbol{k}\rangle}{(E_{1\boldsymbol{k}} - E_{1\boldsymbol{q}_2})}$$

$$\rightarrow \langle 1\boldsymbol{q}_2|V_{eff}|1\boldsymbol{k}\rangle = \sum_{\boldsymbol{q}_1}\frac{\langle 1\boldsymbol{q}_2|H_I|2\boldsymbol{q}_1\rangle\langle 2\boldsymbol{q}_1|H_I|1\boldsymbol{k}\rangle}{(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}_1})} \quad (2-22)$$

The right hand side of 1st line in Eq. 2-22 is simply the first order perturbation wavefunction of state $|p\boldsymbol{k}\rangle$ upon an external potential V_{eff}. In other words, if we are able to find an effective potential V_{eff} satisfying Eq. 2-22, it will be indeed possible to forget about all the details of layer 2 and use this effective potential instead. We therefore call such situation "**Effective Potential**".

The advantage of effective potential description is apparent: it can significantly simplify the system into a single layer, and can reduce interlayer interaction into an external potential on the single layer, which has been intensively-studied for decades. This allows for directly borrowing concepts and conclusions from similar systems and studies; and further approximation such as continuum model.

2.5.4 Determining the Effective Potential

Momentum of the Effective Potential and Periodicity of the Moiré Superlattice

With the significant advantages of using effective potential description, now the problem is how to determine V_{eff}. Naturally, one can do a Fourier transform and expand V_{eff} into Fourier components:

$$V_{eff} = \sum_{\boldsymbol{k_j}} V_j \, e^{i\boldsymbol{k_j}\cdot\boldsymbol{r}} \quad (2-23)$$

where each term corresponds to one Fourier component with momentum $\boldsymbol{k_j}$. Now the question become how to determine V_j and $\boldsymbol{k}_j$. From the definition in Eq. 2-22, we have:

$$\langle 1\boldsymbol{q}_2 | V_{eff} | 1\boldsymbol{k}\rangle = \left\langle 1\boldsymbol{q}_2 \left| \sum_{\boldsymbol{k_j}} V_j \, e^{i\boldsymbol{k_j}\cdot\boldsymbol{r}} \right| 1\boldsymbol{k} \right\rangle = \sum_{\boldsymbol{k_j}} \langle 1\boldsymbol{q}_2 | V_j | 1\boldsymbol{k}\rangle \, \delta(\boldsymbol{k} + \boldsymbol{k}_j + \boldsymbol{G}_1 - \boldsymbol{q}_2) \quad (2-24)$$

The δ function again comes from the fact that both $|1\boldsymbol{q}_2\rangle$ and $|1\boldsymbol{k}\rangle$ are Bloch states from the 1st layer, and therefore the momentum conservation is only good up to a reciprocal lattice vector $\boldsymbol{G}_1$.

We can now compare Eq. 2-24 to the right hand side of Eq. 2-22 to obtain the momentum $\boldsymbol{k_j}$ of the effective potential. From the momentum selection rule Eq. 2-14, the matrix element $\langle 1\boldsymbol{q}_2 | H_I | 2\boldsymbol{q}_1\rangle\langle 2\boldsymbol{q}_1 | H_I | 1\boldsymbol{k}\rangle$ is non-zero only when:

$$\boldsymbol{q}_2 + \boldsymbol{G''}_1 = \boldsymbol{q}_1 + \boldsymbol{G''}_2 = \boldsymbol{K}_1, \qquad \boldsymbol{q}_1 + \boldsymbol{G'}_2 = \boldsymbol{k} + \boldsymbol{G'}_1 = \boldsymbol{K}_2$$
$$\rightarrow \boldsymbol{k} + \boldsymbol{G'}_1 - \boldsymbol{G'}_2 + \boldsymbol{G''}_2 - \boldsymbol{G''}_1 = \boldsymbol{q}_2$$
$$\rightarrow \boldsymbol{k} + \boldsymbol{G}_1 - \boldsymbol{G}_2 = \boldsymbol{q}_2 \quad (2-25)$$

We have used the fact that $(\boldsymbol{G'}_1 - \boldsymbol{G''}_1)$ is still a lattice vector of layer 1, and similarly to $(\boldsymbol{G'}_2 - \boldsymbol{G''}_2)$. Comparing Eq. 2-24 and 2-25 directly gives the expression of $\boldsymbol{k_j}$:

$$\boldsymbol{k}_j = \boldsymbol{G}_1 - \boldsymbol{G}_2 \quad (2-26)$$

The value of $\boldsymbol{k_j}$ here is **limited to the 1st Brillouin Zone of layer 1**.

The amplitude of V_j will depend on both matrix element $\langle 1\boldsymbol{q}_2 | H_I | 2\boldsymbol{q}_1\rangle$ and $\langle 2\boldsymbol{q}_1 | H_I | 1\boldsymbol{k}\rangle$. According to Eq. 2-17, they should decrease quite fast with $|\boldsymbol{K}_1|$ and $|\boldsymbol{K}_2|$. If $\boldsymbol{G}_1 = (\boldsymbol{G'}_1 - \boldsymbol{G''}_1)$ has large amplitude, then at least one of $|\boldsymbol{G'}_1|$ and $|\boldsymbol{G''}_1|$ should be large, so as one of $|\boldsymbol{K}_1|$ and $|\boldsymbol{K}_2|$. As a result, we can conclude that the amplitude V_j should also decrease quite fast with both $|\boldsymbol{G}_1|$ and $|\boldsymbol{G}_2|$, though in a more complicated manner.

Things can become particularly simple if the reciprocal lattice vectors of the two layers are very close, e.g. when they have the same type of lattice but slightly different lattice constant and/or a small twist angle. Assuming the primitive reciprocal lattice vectors of the 1st layer are $\boldsymbol{b}_{11}$ and

$\boldsymbol{b}_{12}$; and $\boldsymbol{b}_{21}$ and $\boldsymbol{b}_{22}$ for the 2nd layer. Now $\boldsymbol{k}_j = n_1\boldsymbol{b}_{11} + m_1\boldsymbol{b}_{12} - (n_2\boldsymbol{b}_{21} + m_2\boldsymbol{b}_{22})$. Since $\boldsymbol{b}_{11}$ and $\boldsymbol{b}_{12}$ are very close to $\boldsymbol{b}_{21}$ and $\boldsymbol{b}_{22}$, we must have $n_1 = n_2 = n$ and $m_1 = m_2 = m$ to keep $\boldsymbol{k}_j$ in 1st Brillouin Zone of layer 1:

$$\begin{aligned} \boldsymbol{k}_j &= n(\boldsymbol{b}_{11} - \boldsymbol{b}_{21}) + m(\boldsymbol{b}_{12} - \boldsymbol{b}_{22}) \\ &= n\boldsymbol{b}_1^M + m\boldsymbol{b}_2^M \qquad (2-27) \end{aligned}$$

Here $\boldsymbol{b}_1^M = \boldsymbol{b}_{11} - \boldsymbol{b}_{21}$ and $\boldsymbol{b}_2^M = \boldsymbol{b}_{12} - \boldsymbol{b}_{22}$ are the primitive reciprocal lattice vectors of the so called "**Mini Brillouin Zone**", or "**Moiré Brillouin Zone**". The corresponding real-space lattice vectors, $\boldsymbol{a}_1^M$ and $\boldsymbol{a}_2^M$, define the Moiré superlattice from by the two layers. This can be a general way to **determine the Moiré superlattice periodicity** in the system. Eq. 2-27 then dictates that all the momentum in the effective potential will be a reciprocal lattice vector of the Moiré superlattice. In this sense, we may also call the effective potential a "**Moiré potential**" in this case.

In addition, the conclusion that V_j should decrease quite fast with both $|\boldsymbol{G}_1|$ and $|\boldsymbol{G}_2|$ now becomes that V_j should decrease quite fast with n and m. Combining with Eq. 2-27, the effective potential can have a very simple and intuitive real-space picture: The effective potential should follow the same periodicity as the moiré superlattice in real space, therefore the only existing Fourier components are given by the reciprocal lattice vectors of the Moiré Brillouin Zone. Furthermore, the effective potential is not likely to change much faster than the length scale of a unit cell, thus the amplitude of the Fourier components should decrease quite fast with increasing momentum, and only the first few need to be considered.

Given the elegance of such explanation, I want to note its limitation at the same time. The Moiré periodicity obtained above is only a "**quasi-periodicity**" that is approximately correct when the two layers have similar lattices; however will fail when the two lattices are very different. For example, when $b_{11} = 0.7b_{21}$. In such situation, important components of $\boldsymbol{k}_j$ are not fully described by Eq. 2-27, and their amplitudes will not necessarily decrease with $|\boldsymbol{k}_j|$. Therefore, the Moiré periodicity defined from Eq. 2-27 is not a good, or real periodicity of the system anymore. One should instead refer to Eq. 2-26 to fully determine the effective potential.

Now we have obtained all the momentum involved in the effective potential, which already gives important implications even without knowing the amplitude of each component. For example, using the criterion $E_{1\boldsymbol{k}} = E_{1(\boldsymbol{k}+\boldsymbol{k}_j)}$ (2-28), we can already determine states with degenerate coupling, and in turn positions on the energy spectrum that will be strongly modified.

Amplitude of the Effective Potential

The last piece of information is the amplitude V_j. While the general calculation of matrix element is already given in Eq. 2-13 and 2-17, here again I want to simplify the problem a bit more by looking at some special but useful cases.

Let us stay with the case in Eq. 2-27, and calculate V_1 for one of the first Fourier components $\boldsymbol{k}_1 = \boldsymbol{b}_{11} - \boldsymbol{b}_{21}$. First we need expand the three involved Bloch states into local orbitals:

$$|1\boldsymbol{k}\rangle = \sum_{\boldsymbol{R}_1} \sum_n b_n^{\boldsymbol{k}} \varphi_n^1(\boldsymbol{r} - \boldsymbol{R}_1) e^{i\boldsymbol{k}\cdot\boldsymbol{R}_1}$$

$$|1\boldsymbol{k}'\rangle = \sum_{\boldsymbol{R}_1'} \sum_{n'} b_{n'}^{\boldsymbol{k}'} \varphi_{n'}^1(\boldsymbol{r} - \boldsymbol{R}_1') e^{i\boldsymbol{k}'\cdot\boldsymbol{R}_1'}$$

$$|2\boldsymbol{q}\rangle = \sum_{\boldsymbol{R}_2} \sum_m a_m^{\boldsymbol{q}} \varphi_m^2(\boldsymbol{r} - \boldsymbol{R}_2) e^{i\boldsymbol{q}\cdot\boldsymbol{R}_2} \quad (2\text{-}29)$$

From the left hand of Eq. 2-22 we have:

$$\langle 1\boldsymbol{k}' | V_{eff} | 1\boldsymbol{k} \rangle = \sum_{\boldsymbol{R}_1'} \sum_{n'} \sum_{\boldsymbol{R}_1} \sum_n b_n^{\boldsymbol{k}} \varphi_n^1(\boldsymbol{r} - \boldsymbol{R}_1) e^{i\boldsymbol{k}\cdot\boldsymbol{R}_1} b_{n'}^{\boldsymbol{k}'*} \varphi_{n'}^{1*}(\boldsymbol{r} - \boldsymbol{R}_1') e^{-i\boldsymbol{k}'\cdot\boldsymbol{R}_1'} V_1 e^{i\boldsymbol{k}_1\cdot\boldsymbol{r}}$$

$$= \sum_{n'} \sum_n b_{n'}^{\boldsymbol{k}'*} b_n^{\boldsymbol{k}} V_1 \delta_{nn'} \quad (2\text{-}30)$$

In which we have used the fact that $\varphi_n^1(\boldsymbol{r} - \boldsymbol{R}_1)$ and $\varphi_{n'}^{1*}(\boldsymbol{r} - \boldsymbol{R}_1')$ are both localized in space, so that $\langle \varphi_{n'}^1(\boldsymbol{r} - \boldsymbol{R}_1') | V_1 e^{i\boldsymbol{k}_1\cdot\boldsymbol{r}} | \varphi_n^1(\boldsymbol{r} - \boldsymbol{R}_1) \rangle \approx V_1 \delta_{nn'} \delta_{\boldsymbol{R}_1\boldsymbol{R}_1'}$.

On the other hand, using Eq. 2-13, we have:

$$\langle 1\boldsymbol{k}' | H_I | 2\boldsymbol{q} \rangle \langle 2\boldsymbol{q} | H_I | 1\boldsymbol{k} \rangle = \sum_{n'} \sum_{m'} b_{n'}^{\boldsymbol{k}'*} a_{m'}^{\boldsymbol{q}} t_{n'm'}(\boldsymbol{K}_2) \sum_n \sum_m a_m^{\boldsymbol{q}*} b_n^{\boldsymbol{k}} t_{nm}^*(\boldsymbol{K}_1)$$

$$= \sum_{n'} \sum_n b_{n'}^{\boldsymbol{k}'*} b_n^{\boldsymbol{k}} \left[\sum_{m'} \sum_m a_{m'}^{\boldsymbol{q}} a_m^{\boldsymbol{q}*} t_{nm}^*(\boldsymbol{K}_1) t_{n'm'}(\boldsymbol{K}_2) \right] \quad (2\text{-}31)$$

Comparing Eq. 2-30 and 2-31, it seems that we need let

$$V_1 \delta_{nn'} = \frac{1}{(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}})} \sum_{m'} \sum_m a_{m'}^{\boldsymbol{q}} a_m^{\boldsymbol{q}*} t_{nm}^*(\boldsymbol{K}_1) t_{n'm'}(\boldsymbol{K}_2) \quad (2\text{-}32)$$

If $N_1 = 1$, i.e. there is only 1 atom (or orbital) in the unit cell of layer 1, we can conclude straightforwardly that $V_1 = \sum_{m'} \sum_m a_{m'}^{\boldsymbol{q}} a_m^{\boldsymbol{q}*} t_m(\boldsymbol{K}_1) t_{m'}(\boldsymbol{K}_2)$. However, when $N_1 > 1$, the above equality seems impossible since the LHS is a simple number, while the RHS have multiple components in the space expanded by n and n'; and there is no reason that the sum $\sum_{m'} \sum_m a_{m'}^{\boldsymbol{q}} a_m^{\boldsymbol{q}*} t_{nm}(\boldsymbol{K}_1) t_{n'm'}(\boldsymbol{K}_2)$ should give anything like $\delta_{nn'}$.

In fact, this does not mean that we cannot use the effective potential description when $N_1 > 1$. On the contrary, this is **one of the most interesting aspects of the effective potential** created by interlayer interaction: The mismatch in Eq. 2-32 origins from the fact that we assume V_1 to be a simple number. As a result, $V_1 e^{i\boldsymbol{k}_1\cdot\boldsymbol{r}}$ is a simple spatially-varying scalar potential that can be fully described by an electrostatic potential with periodicity of the Moiré superlattice.

However, given the multiple periodicities involved in the intricate interplay between layers, such as $\boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G^M}$, it is hard to imagine that the effect can be solely represented by a simple scalar potential. Indeed, Eq. 2-32 just means that this would be impossible when $N_1 > 1$.

On the other hand, if we let V_1 to be a matrix instead of a number. Eq. 2-30 will become:

$$\langle 1\boldsymbol{k}'|V_{eff}|1\boldsymbol{k}\rangle = \sum_{n'}\sum_{n} b_{n'}^{\boldsymbol{k}'*} b_{n}^{\boldsymbol{k}} V_{1,n'n} \qquad (2\text{-}33)$$

Where V_1 is a N_1 by N_1 matrix, whose elements are $V_{1,n'n}$. Comparing to Eq. 2-31 becomes now natural, which gives:

$$V_{1,nn'} = \frac{1}{(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}})} \sum_{m'}\sum_{m} a_{m'}^{\boldsymbol{q}} a_{m}^{\boldsymbol{q}*} t_{nm}^{*}(\boldsymbol{K}_1) t_{n'm'}(\boldsymbol{K}_2) \qquad (2\text{-}34)$$

i.e. All the matrix elements of V_1 can be directly obtained. We note that, the above calculation only considers a single state $|2\boldsymbol{q}\rangle$ as the intermediate state. If multiple states from layer 2 can contribute to the same component in V_{eff}, we will need sum their contribution.

Few Remarks

The ability to provide a matrix potential instead of simple scalar one is a quite unique consequence of interlayer interaction. Components in the matrix V_1 can have fundamentally different physical meaning from each other, some are otherwise almost impossible to achieve by applying any artificial external field. As a result, they will generate distinctively different effects on the electronic properties of the system, which offers unique and exciting opportunities. One example of the matrix potential in graphene/hBN heterostructure will be calculated in detail in section 2.5.7; and the induced fascinating physics is demonstrated by experimental studies in Chapter 5.

From Eq. 2-34, apparently the effective potential V_1 will depend on momentum $\boldsymbol{k}$. While it is still fine to calculate V_1 at each specific $\boldsymbol{k}$, it turns out that in many cases V_1 can be approximately considered a constant if we are focusing on electronics states around a specific $\boldsymbol{k}$. In Eq. 2-34, $(E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}})$ can be quite insensitive to $\boldsymbol{k}$, especially when $|E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}}|$ is much larger than the energy range we are interested in. Similarly, $t_{nm}(\boldsymbol{K}_1)$ and $t_{n'm'}(\boldsymbol{K}_2)$ can also be treated as constant if we focus on a small enough momentum range. Therefore, the main problem will be coming from $a_{m'}^{\boldsymbol{q}}$ and $a_{m}^{\boldsymbol{q}*}$, which may change sensitively regardless of how small a range of $\boldsymbol{k}$ we are looking at. For example, $a_{m}^{\boldsymbol{q}}$ of two states can change from $\frac{\sqrt{2}}{2}$ to $-\frac{\sqrt{2}}{2}$ when near the Dirac point of graphene even if they are infinitely close.

There is one case, however, in which we have no worry about $a_{m}^{\boldsymbol{q}}$: if $N_2 = 1$, and therefore $a_{m}^{\boldsymbol{q}} \equiv 1$. In fact, in many situations this will be approximately true. For example, if layer 2 is a wide bandgap insulator, the wide bandgap usually means that the valence band and conduction band are coming from different orbitals that almost do not mix between each other. In this

situation, although there can be more than one orbitals in each unit cell of layer 2, only one is contributing to the effective potential at a time, and we still have $a_m^{\boldsymbol{q}} \approx 1$. It is then possible to calculate the effective potential only once for all electronic states within certain range.

2.5.5 Summary

In present section 2.5, we have classified the momentum space diagonalization method for incommensurate system into weak, strong, perturbative and non-perturbative regime; and then focused on the weak coupling regime by further classifying it into direct coupling and effective potential cases. Here I want to briefly summarize the difference and similarities of these two cases.

Fundamentally, both direct mixing and effective potential origin from interlayer interaction, but correspond to different order: The former describes a direct mixing between two states from different layers; while the latter is a mixing between two states from the same layer, mediated by a third state from the other layer.

To determine states with degenerate coupling, one need take into account band dispersion from both layers in the direct mixing case, by overlaying them and searching for crossing points; however only band dispersion from single layer in the effective potential case, by shifting it by certain amount and find crossing points with the original dispersion. If specific electronic state satisfy one criterion, it is likely that the corresponding process plays a critical role in understanding its behavior upon interlayer interaction. One specific state cannot satisfy two criterion at the same time, otherwise the system will be in strong coupling regime.

Practically, if the two layers both have electronic states in the energy range of interest, especially if they are the same type of materials, the direct mixing process is likely to be the one giving major effects. On the other hand, if one layer does not have states in the energy range of interest, especially when it is a wide bandgap insulator for substrate or encapsulation purpose, the effective potential description can be a better choice. We note, however, that the two processes are not mutually-exclusive and can be considered simultaneously.

2.5.6 Example of Direct Mixing: Twisted Bilayer Graphene

This section will give one representative example of how direct mixing can modify the properties of the materials, and the model materials here will be twisted bilayer graphene.

Twisted bilayer graphene is one of the first and most intensively studied bilayer van der Waals structure. The richly varied behaviors depending on twist angle is a good example that the large

parameter space of interlayer interaction can provide new rich possibilities: Near the Dirac point, the bilayer with large twist angle will recover the linear band dispersion as in monolayer(*16-18*); when twist angle becomes smaller, the Fermi velocity will get renormalized and decrease(*19-21*); even smaller twisted angle can result in more bizarre behaviors such as flat bands and zero energy nodes(*8-11*). Away from the Dirac point, new Van Hove singularities have been observed in scanning tunneling spectroscopy from interlayer interaction(*22*), which can further affect the system's properties on various aspects such as Raman response(*23, 24*); and can be used to determine the interlayer twist angle.

Here I want to show that, given such richly varied behaviors depending on twist angle, we can qualitatively understand all of them in a quite simple and straightforward way, with the knowledge we have built in previous sections.

Firstly, we have mentioned in section 2.5.3 that two types of electronic states are of particular interest, one type is those near important momentum $\boldsymbol{k}$ in monolayer; and the other type is those have degenerate coupling. For graphene, the most worth-investigating electronics states in monolayer are of course those near the Dirac point. We can therefore take a look at them first.

Given a state in layer 1 at momentum $\boldsymbol{k} = (\boldsymbol{K}_1 + \boldsymbol{p})$, where $\boldsymbol{K}_1$ is the Dirac point in monolayer 1 and $\boldsymbol{p}$ is very small. Because of the linear dispersion near Dirac cone, its energy is simply $E_{1\boldsymbol{k}} = v_F p$, where v_F is the Fermi velocity in graphene.

States from layer 2 that can have direct mixing with $|1\boldsymbol{k}\rangle$ are given by the selection rule Eq. 2-14. For a qualitative understanding, let us just take one of these states, $|2\boldsymbol{k}\rangle$, and see how the mixing between them will modify the system.

We first need determine the energy of $|2\boldsymbol{k}\rangle$. For this, we use the Brillouin zone of layer 2 as the reference frame:

$$E_{2\boldsymbol{k}} = v_F|\boldsymbol{k} - \boldsymbol{K}_2| = v_F|\boldsymbol{p} + (\boldsymbol{K}_1 - \boldsymbol{K}_2)| = v_F|\boldsymbol{p} + \Delta\boldsymbol{K}| \qquad (2\text{-}35)$$

Here $\Delta\boldsymbol{K}$ is the difference between Dirac points in two layers due to interlayer twist. $|\Delta\boldsymbol{K}| = 2K\sin(\frac{\theta}{2}) \approx K\theta$ when θ is small. Therefore

$$E_{2\boldsymbol{k}} = v_F(K\theta + p\cos\alpha) \qquad (2\text{-}36)$$

Where α is the angle between $\boldsymbol{p}$ and $\Delta\boldsymbol{K}$, and we have assumed $p \ll \Delta K$.

In non-perturbative regime, we can directly obtain the effect of direct mixing between $|1\boldsymbol{k}\rangle$ and $|2\boldsymbol{k}\rangle$ from the 1st term in Eq. 2-20, and the corresponding energy shift is simply a second-order perturbation in energy:

$$\Delta E_{1\boldsymbol{k}} = \frac{|M_{1\boldsymbol{k},2\boldsymbol{k}}|^2}{E_{1\boldsymbol{k}} - E_{2\boldsymbol{k}}} = -\frac{|M|^2}{v_F[K\theta + p(\cos\alpha - 1)]} \approx -\frac{|M|^2}{v_F K\theta} - \frac{|M|^2}{v_F(K\theta)^2}p(1 - \cos\alpha) \quad (2-37)$$

The last term in Eq. 2-37 is proportional to p, meaning that the Fermi velocity near Dirac point will be changed by the direct mixing between $|1\boldsymbol{k}\rangle$ and $|2\boldsymbol{k}\rangle$. To quantitatively obtain the

amount of Fermi velocity renormalization will require considering direct mixing between $|1\boldsymbol{k}\rangle$ and at least six states from layer 2, three from conduction band and three from valence band, each can have different M and α. However, we can already reach important conclusions from Eq. 2-37: first, the linear dispersion will be recovered; and second, the relative change in Fermi velocity $\Delta v_F/v_F$ scales like $\frac{|M|^2}{(v_F K\theta)^2}$, which will be negligible at large twist angle θ, but can be significant when θ is small. At even smaller θ, the system will enter strong coupling regime, which can give rise to more bizarre behaviors.

Now we have understood the behavior of states near Dirac point, and can turn to the second type of interesting states: those have degenerate coupling. Determination of these states can be done straightforwardly following the discussion in section 2.5.3, and can often benefit from symmetry. Let us again consider the example of mixing between $|1\boldsymbol{k}\rangle$ and $|2\boldsymbol{k}\rangle$. Degenerate coupling condition requires $\boldsymbol{k}$ to satisfy:

$$E_{1\boldsymbol{k}} = v_F|\boldsymbol{k} - \boldsymbol{K}_1| = E_{2\boldsymbol{k}} = v_F|\boldsymbol{k} - \boldsymbol{K}_2| \qquad (2\text{-}38)$$

Apparently, the lowest energy $\boldsymbol{k}$ satisfying the above equation is at the middle point between the line connecting $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ in momentum space, with the energy $E_{1\boldsymbol{k}} = E_{2\boldsymbol{k}} = v_F K\sin(\frac{\theta}{2})$. The degenerate coupling between these two states lead to avoided crossing and Van Hove singularities, which strongly modify the system's density of states (DOS) near this energy, as is observed in ref. (*22*). Since this will happen both in conduction band and valence band, the transition between states near the two Van Hove singularities create a large joint density of states (JDOS) near $2E_{\boldsymbol{k}} = 2v_F K\sin(\frac{\theta}{2})$, which leads to enhancement of Raman signal when optical excitation are in resonance with this energy(*23*, *24*).

The above example of twisted bilayer graphene clearly shows the power of the framework here in qualitatively understanding the consequence of seemingly complicated interlayer interaction. Furthermore, quantitative information are also readily available by doing one more step of considering all relevant states. Two examples on black phosphorus and double-walled nanotubes (DWNTs) will be given in chapter 3 and 4, which provide quantitative explanation to experimental results from optical spectroscopy studies of these systems.

2.5.7 Example of Effective Potential: Graphene/hBN Heterostructure

Graphene and hexagonal boron nitride (hBN) have close lattice structure: both with hexagonal lattice, and the lattice constant is only different by ~2%. The heterostructure formed by stacking graphene and hBN together therefore are quite similar to bilayer graphene in atomic configuration. However, the physics induced by interlayer interaction is distinctively different in these two systems. In graphene/hBN heterostructure, experiments have observed emergence of

replica of Dirac points away from the original Dirac point in monolayer, often called "mini Dirac points"(*25, 26*); with an external out-of-plane magnetic field, the interplay between original Dirac points and "mini Dirac points" give rise to a beautiful fractal pattern in magneto-transport studies(*27-29*), which is the first experimental demonstration of the famous Hofstadter's butterfly pattern that was predicted 40 years ago(*30*). We can now readily understand the drastically different consequences of interlayer interaction between the two systems, as the majority of effects will be coming from direct mixing in bilayer graphene, but effective potential for graphene/hBN heterostructure.

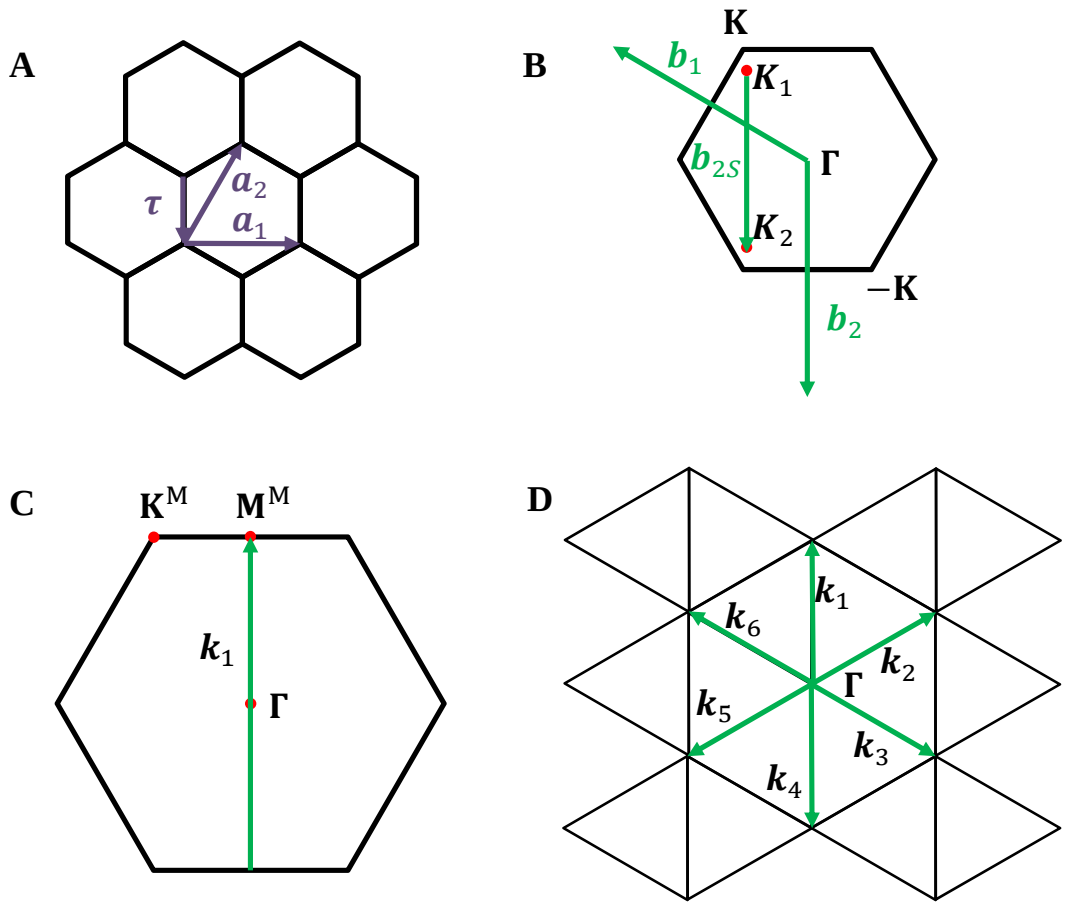

Figure 2.1 (**A**) Real space lattice structure of graphene, $\boldsymbol{a}_1, \boldsymbol{a}_2$ are the two primitive lattice vectors. $\boldsymbol{\tau}$ is the real space vector connecting two sublattices in the same unit cell. (**B**) First Brillouin Zone of isolated graphene monolayer in momentum space, $\boldsymbol{b}_1, \boldsymbol{b}_2$ are the two primitive reciprocal lattice vectors. Two inequivalent Dirac points are at **K** and **−K**. $\boldsymbol{K}_2$ and $\boldsymbol{K}_1$ are differed by a reciprocal lattice vector of the substrate, $\boldsymbol{b}_{2S}$, and therefore represent the same substrate state. This substrate state acts as a bridge to couple two graphene states at $\boldsymbol{K}_2$ and $\boldsymbol{K}_1$. The difference between $\boldsymbol{b}_2$ and $\boldsymbol{b}_{2S}$ is exaggerated for clarity. (**C**) Mini-Brillouin Zone (MBZ) of graphene/hBN heterostructure. High symmetry points are labeled as $\mathbf{M}^M$ and $\mathbf{K}^M$. $\boldsymbol{k}_1$ is one of the first Fourier components in the effective potential, which is simply $-\boldsymbol{b}_2^M$. (**D**)

Available momentum in the effective potential are all on the reciprocal lattice formed by $\boldsymbol{b}_1^M$ and $\boldsymbol{b}_2^M$.The six components with smallest momentum k are labeled by $\boldsymbol{k}_1$ to $\boldsymbol{k}_6$. All of them are related by symmetry.

The general formalism to obtain the effective potential has been given in section 2.5.4. Here I will follow the process and quantitatively calculate the effective potential for a representative interlayer configuration, where the graphene and hBN have zero twist angle.

First of all, let us apply our conclusion developed in section 2.5.4 and determine how V_{eff} should look like in this case:

Given the small lattice mismatch between graphene and hBN lattices, results from Eq. 2-27 can be safely used. We can therefore define the **Mini Brillouin Zone** (MBZ) in momentum space with two primitive reciprocal lattice vectors $\boldsymbol{b}_1^M = \boldsymbol{b}_{11} - \boldsymbol{b}_{21}$ and $\boldsymbol{b}_2^M = \boldsymbol{b}_{12} - \boldsymbol{b}_{22}$. The MBZ has the same hexagonal shape as the Brillouine Zone (BZ) of graphene and hBN, however with a size ~2% as large. Correspondingly, the Moiré superlattice in real space will be ~50 times larger than the unit cell in graphene and hBN.

An illustration of the MBZ is shown in Figure 2.1C. High symmetry points of the MBZ are labeled as M^M, K^M and Γ, where the superscript "M" indicates that they are referring to MBZ instead of BZ of isolated graphene or hBN.

Eq. 2-27 also states that the amplitude V_j decreases quite fast with k_j, and that we might only consider the components with smallest k_j. To do this, we can draw the reciprocal lattice formed by $\boldsymbol{b}_1^M$ and $\boldsymbol{b}_2^M$, and select points closest to the center. Since the reciprocal lattice is a triangular lattice, there are six points closest to the center, given by $\boldsymbol{k}_1$ to $\boldsymbol{k}_6$ in Figure 2.1D. A representative component, $\boldsymbol{k}_1 = -\boldsymbol{b}_2^M$, is drawn in the MBZ in Figure 2.1C, from which we can clearly see the relation between the MBZ and the momentum in the Moiré potential.

The Moiré potential now takes the form:

$$V_{eff} = \sum_{j=1}^{6} V_j e^{i\boldsymbol{k}_j \cdot \boldsymbol{r}} \quad (2-39)$$

Before really going into the derivation of V_j, it is always good to first further simply the problem with more general consideration like symmetry. Both hBN and graphene have three-fold rotational symmetry, so is their Moiré superlattice. V_1, V_3, V_5 are therefore related by the rotational symmetry and not independent; same argument applies to V_2, V_4, V_6.

Another restriction to V_{eff} is that it must be Hermitian. i.e. $V_{eff} = V_{eff}^+$, which relates V_1 and V_4. Therefore all the six components in Figure 2.1D are related and only one is independent, which we can choose to be V_1. From the discussion in section 2.5.4, in general V_1 is a matrix potential with dimension $N_1 = 2$. A general form of 2-by-2 matrix is:

$$V_1 = \begin{pmatrix} u_0 + iu_3 & u_1 + iu_2 \\ u_1 - iu_2 & u_0 - iu_3 \end{pmatrix} \quad (2-40)$$

where all the 4 parameters u_j are in principle complex numbers. Comparing to the Hamiltonian of electronic states in graphene near $\boldsymbol{K}$ point in Eq. 2-9, we can see that the term u_1, u_2 actually correspond to vector potential in x and y direction, respectively. This can be seen by replacing $\boldsymbol{p}$ with $(\boldsymbol{p} - e\boldsymbol{A})$ in Eq. 2-9. Therefore, u_2 term is effectively a vector potential $\boldsymbol{A} = \frac{u_2}{e}\hat{\boldsymbol{y}}e^{i\boldsymbol{k_1}\cdot\boldsymbol{r}}$, with $\hat{\boldsymbol{y}}$ the unit vector along y axis. Because that both $\boldsymbol{A}$ and $\boldsymbol{k_1}$ are along y direction, the corresponding magnetic field $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \equiv 0$. In other words, the u_2 term in Eq. 2-40 will not have any real effect and can be gauged away. The expression of V_1 now becomes:

$$V_1 = \begin{pmatrix} u_0 + iu_3 & u_1 \\ u_1 & u_0 - iu_3 \end{pmatrix} \quad (2-41)$$

We note that, to be able to assume V_1 as a constant, all discussion here are necessarily limited to electronic states in a small momentum range. i.e. near one Dirac point ($\boldsymbol{K}$ point) of BZ of isolated graphene; effective potential for the other inequivalent Dirac point ($-\boldsymbol{K}$ point) can be directly obtained from time reversal symmetry:

$$\widetilde{V_1} = \begin{pmatrix} u_0^* - iu_3^* & u_1^* \\ u_1^* & u_0^* + iu_3^* \end{pmatrix} \quad (2-42)$$

Another symmetry we can take advantage of is inversion symmetry. While it is not apparent why V_{eff} should be inversion symmetric, we can infer it from several approximations:

First, the important BN states involved as intermediate states, i.e. $|2\boldsymbol{q}\rangle$ in Eq. 2-30, are also near band edge in hBN's energy dispersion. As we have mentioned in section 2.5.4, the large bandgap of hBN indicates that the conduction and valence band edges are likely belong to different orbitals that do not mix much. Indeed, the conduction band minimum and valence band maximum are mainly from p orbitals of Boron and Nitrogen atom, respectively. Therefore, when considering the conduction band state from hBN as intermediate state $|2\boldsymbol{q}\rangle$, only Boron atoms are important; and the hBN layer is approximately equivalent to a triangular lattice of Boron atoms only, which is inversion-symmetric.

Second, it is reasonable to assume that both graphene and hBN layers are sufficiently flat. If the interlayer separation along z (out-of-plane) direction is a constant, then the pair-interaction between two interlayer atoms will only depend on their displacement in xy plane. As a result, the symmetry of effective potential should not depend on the interlayer separation between the two layers, and we can shift them vertically to be in the same plane.

Third, now we effectively have a graphene layer and a Boron layer in the same xy plane, and both are inversion symmetric. Due to the lattice mismatch, the relative displacement between unit cells of the two layers will change over space, and we can always find a point where the two inversion centers overlap. Therefore the whole system should be approximately inversion-symmetric, so is the effective potential.

The above argument is generally true for two incommensurate layers that both are inversion-symmetric or effectively inversion-symmetric. However, given the approximations involved, the inversion-asymmetric part of the effective potential also exists, but should be much smaller.

Let us neglect the inversion-asymmetric part for now. The inversion symmetry put an another restriction on the parameters u_j by requiring:

$$\begin{pmatrix} 1 \\ e^{-i\theta} \end{pmatrix}^{+} V_1 \begin{pmatrix} 1 \\ e^{i\theta} \end{pmatrix} = \begin{pmatrix} e^{-i\theta} \\ 1 \end{pmatrix}^{+} \widetilde{V_1} \begin{pmatrix} e^{i\theta} \\ 1 \end{pmatrix} \quad (2-43)$$

Here $\begin{pmatrix} 1 \\ e^{-i\theta} \end{pmatrix}$ and $\begin{pmatrix} 1 \\ e^{i\theta} \end{pmatrix}$ describes two electronic states of graphene near $\boldsymbol{K}$ point that can be coupled by the component V_1 of the effective potential, $\theta \in (0, \pi)$ is the phase difference between the two sublattices. Inversion symmetry operation will bring a state near $\boldsymbol{K}$ point to near $-\boldsymbol{K}$, and also exchange the two sublattices, therefore giving the expression on the right hand side. By plug in V_1 and $\widetilde{V_1}$ into Eq. 2-43, one can readily see that all the parameters u_j in Eq. 2-41 **must be real**.

It is quite impressive that the general formalism developed in section 2.5.4, combined with symmetry consideration for specific systems, can reduce the interaction in a complex incommensurate system to such simple expression as in Eq. 2-41. It then becomes very convenient to discuss the physical nature of each component of the potential, as well as their effects on the system, even without knowing their exact values. These discussions will be provided in chapter 5, combined with experimental studies. Here, I will go back to the original task of present section, and finally calculate the values of these components.

According to Eq. 2-34, we have:

$$V_{1,nn'} \approx \frac{1}{(E_{1\boldsymbol{k}} - E_S)} t_n^*(\boldsymbol{K}_1) t_{n'}(\boldsymbol{K}_2) \quad (2-44)$$

In which E_S is the energy of substrate state from hBN, and $S = B, N$; $a_m^{\boldsymbol{q}} \approx 1$ due to little mixing between Boron and Nitrogen orbitals at band edge; $\boldsymbol{K}_1 \approx \frac{1}{3}(\boldsymbol{b}_1 - \boldsymbol{b}_2), \boldsymbol{K}_2 \approx \frac{1}{3}(\boldsymbol{b}_1 + 2\boldsymbol{b}_2)$, with $\boldsymbol{b}_1, \boldsymbol{b}_2$ the primitive reciprocal lattice vectors of graphene, see Figure 2.1B. We can then calculate $t_n(\boldsymbol{K}_1)$ and $t_{n'}(\boldsymbol{K}_2)$ following similar steps as in Eq. 2-17:

$$t_1(\boldsymbol{K}_1) = \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_1 \cdot \boldsymbol{R}} \left\langle \varphi_g^1\left(\boldsymbol{r} + \frac{\boldsymbol{\tau}}{2}\right) \middle| H_I \middle| \varphi_S^2(\boldsymbol{r} - \boldsymbol{R}) \right\rangle$$

$$= e^{-i\boldsymbol{K}_1 \cdot \frac{\boldsymbol{\tau}}{2}} \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_1 \cdot \boldsymbol{R}} \langle \varphi_g^1(\boldsymbol{r}) | H_I | \varphi_S^2(\boldsymbol{r} - \boldsymbol{R}) \rangle \sim t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2} K^2} e^{-i\frac{\pi}{3}}$$

$$t_2(\boldsymbol{K}_1) = \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_1 \cdot \boldsymbol{R}} \left\langle \varphi_g^1\left(\boldsymbol{r} - \frac{\boldsymbol{\tau}}{2}\right) \middle| H_I \middle| \varphi_S^2(\boldsymbol{r} - \boldsymbol{R}) \right\rangle$$

$$= e^{i\boldsymbol{K}_1 \cdot \frac{\boldsymbol{\tau}}{2}} \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_1 \cdot \boldsymbol{R}} \langle \varphi_g^1(\boldsymbol{r}) | H_I | \varphi_S^2(\boldsymbol{r} - \boldsymbol{R}) \rangle \sim t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2} K^2} e^{i\frac{\pi}{3}}$$

$$t_1(\boldsymbol{K}_2) = \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_2\cdot\boldsymbol{R}} \left\langle \varphi_g^1\left(\boldsymbol{r}+\frac{\boldsymbol{\tau}}{2}\right)\middle| H_I \middle| \varphi_S^2(\boldsymbol{r}-\boldsymbol{R})\right\rangle$$

$$= e^{-i\boldsymbol{K}_2\cdot\frac{\boldsymbol{\tau}}{2}} \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_2\cdot\boldsymbol{R}} \langle \varphi_g^1(\boldsymbol{r})|H_I|\varphi_S^2(\boldsymbol{r}-\boldsymbol{R})\rangle \sim t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2}K^2} e^{i\frac{\pi}{3}}$$

$$t_2(\boldsymbol{K}_2) = \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_2\cdot\boldsymbol{R}} \left\langle \varphi_g^1\left(\boldsymbol{r}-\frac{\boldsymbol{\tau}}{2}\right)\middle| H_I \middle| \varphi_S^2(\boldsymbol{r}-\boldsymbol{R})\right\rangle$$

$$= e^{i\boldsymbol{K}_2\cdot\frac{\boldsymbol{\tau}}{2}} \int d\boldsymbol{R}\, e^{i\boldsymbol{K}_2\cdot\boldsymbol{R}} \langle \varphi_g^1(\boldsymbol{r})|H_I|\varphi_S^2(\boldsymbol{r}-\boldsymbol{R})\rangle \sim t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2}K^2} e^{-i\frac{\pi}{3}} \quad (2-45)$$

Here $\boldsymbol{\tau} = \frac{1}{3}(\boldsymbol{a}_1 - 2\boldsymbol{a}_2)$ is the real space vector connecting two sublattices in graphene unit cell, see Figure 2.1A. Plugging Eq. 2-45 into Eq. 2-44, we now have:

$$V_1 = \frac{|M_S|^2}{(E_{1\boldsymbol{k}} - E_S)} \begin{pmatrix} e^{i\frac{2\pi}{3}} & 1 \\ 1 & e^{-i\frac{2\pi}{3}} \end{pmatrix} \approx \frac{|M_S|^2}{E_S} \begin{pmatrix} 1-\frac{\sqrt{3}}{2}i & -1 \\ -1 & 1+\frac{\sqrt{3}}{2}i \end{pmatrix} \quad (2-46)$$

in which $M_S \sim t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2}K^2}$, and we have used $E_{1\boldsymbol{k}} \approx 0$. By comparing Eq. 2-46 and Eq. 2-41, the value of parameters can be obtained as:

$$(u_0, u_3, u_1) = \frac{|M_S|^2}{E_S}\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}, -1\right) \quad (2-47)$$

Indeed, all the parameters u_j are real, since the inversion symmetry is automatically included in our calculation of $t(\boldsymbol{K})$.

Similar results are also given in several references, with potentially different notations on symbols(*31, 32*). The above calculation can also be readily extended to the case of finite twist angle between graphene and hBN substrate.

2.6 Classification of Interlayer Interaction, III: Strong and Weak Coupling

2.6.1 Introduction

I have already given a brief discussion on the condition of "**strong coupling**" and "**weak coupling**" in section 2.5.2, that "**strong coupling**" happens if many, or even infinite lines in Eq. 2-20 are of similar amplitude and require consideration. From such conclusion, it is quite straightforward to mathematically derive a criterion as definition of strong coupling regime. However, following this way, the whole process will be like obtaining an abstract equation from another abstract equation, which may prevents one from seeing a deeper physics picture behind. Therefore, here I want to take an alternative route: First discuss some physical consequences of strong coupling, then the implications of these consequences, which naturally sets requirements on the strong coupling regime.

2.6.2 Two (Conflicting) Pictures of Describing Incommensurate Systems

Reviewing all literature discussing properties of incommensurate systems (or systems with large common unit cell size), we will find two main methodologies adopted to solve the problem:

One methodology is by considering **all atoms in the system**, like what we have been doing from section 2.1 to 2.5. The direct diagonalization, momentum space diagonalization, and the following derivative methods are all essentially belong to this methodology, just with different perspectives and approximations. Using this methodology, we necessarily obtain the **global properties** of the system: all the obtained eigen-states are Bloch states extended to infinitely far away in real space; and the corresponding energy dispersion is also for the whole system. For example, a gap in the energy spectrum will be a "**global bandgap**" that does not allow particles to travel through (unless considering the tunneling process).

Another methodology, however, is by considering the "**local properties**" of the system (*33, 34*). Since the interlayer displacement is changing over real space, we may focus on the local configuration between two layers, and obtain a local band structure. For example, in twisted bilayer graphene, if we only look at the range of one or a few unit cells, then locally it may resemble AB-stacking, AA-stacking, etc., depending on the center position. One may say that the band structure should be like AB-stacking at point 1, but like AA-stacking at point 2. For many systems, since the bandgap can certainly depend on stacking order, following the same argument

one can obtain a "**local bandgap**" at point 1, and a different "**local bandgap**" at point 2, and even map out the distribution of "**local bandgap**" over space.

Now there is an apparent problem: These two pictures seem to be not compatible, at least not intuitively. We can easily raise tons of questions. For example, how to understand the concept of "**local bandgap**" in the global picture? What is the eigen-wavefunction of states at "**local bandgap**", and what are their relation to the extended Bloch states in the global picture?

Worse still, the two seemingly conflicting pictures both have reasons to be valid, from different perspectives. The global picture is a direct result of symmetry and Bloch theorem that should be generally correct. It then seems necessary to include all atoms in a periodic system to account for the behavior of a Bloch state, because the boundary condition always play a critical rule in Bloch theorem. The local picture, however, also has its own reasoning as one can argue that when the change of local configuration over space is very slow, particles at one point should only feel the local environment and behave accordingly.

We now find that, all the above conflict is originating from one fundamental question: To determine the behavior of particles at a specific point (in real space), how large a range of the system is necessary to consider? An intuitive answer is: the range of the wavefunction of that particle, since that defines the region where the particle "lives". This answer, however, does not really solve the conflict; and we may still ask whether its wavefunction should extended to infinity as described by a Bloch state, or should be localized.

The answer to the this question is actually very simple: The wavefunction is always extended formally, so Bloch theorem is always correct; however, in some situations, it can be also localized at the same time.

How can a wavefunction be both extended and localized? Let us consider a simple example. For an isolated quantum well, it is well known that if the potential well is deep enough, there can be bound states in the well. i.e. The wavefunction ψ_n will only penetrate a short distance into the potential wall if the kinetic energy is small compared to the amplitude of the wall. Naturally, we call a particle at such state a "localized" particle in the quantum well.

Now let us move one step forward, and consider a periodic array of the above quantum wells, with lattice vector $\boldsymbol{R}_1$. Assuming the separation between neighboring wells is large enough so that the "localized" particle in different quantum wells does not see each other at all. As a result, the particle will not feel the change of environment and will stay in the original state ψ_n. In this sense, this "localized" state ψ_n should be still an eigen-state of the system, even now in a periodic system.

On the other hand, one can also use Bloch theorem and band theory to obtain energy dispersion and corresponding Bloch states, which is simply,

$$\psi_n^{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_1} e^{i\boldsymbol{k}\cdot\boldsymbol{R}_1}\, \psi_n(\boldsymbol{r} - \boldsymbol{R}_1)$$

$$E_n^{\boldsymbol{k}} \equiv E_n \quad (2\text{-}48)$$

The Bloch states are still extended over the whole space from superposition of every quantum well in the array, with the phase difference described by momentum $\boldsymbol{k}$. However, due to the lack of interaction between any two quantum wells, the band dispersion is completely flat, and that the Bloch states are all degenerate regardless of momentum. One can therefore re-construct the basis through an arbitrary linear combination of Bloch states. As a result, both the "localized" states $\psi_n(\boldsymbol{r}-\boldsymbol{R}_1)$ and Bloch states $\psi_n^{\boldsymbol{k}}(\boldsymbol{r})$ are eigen states of the system with energy E_n, just from different basis.

The example above is an extreme case where the exact eigen-states can be both "extended" and "localized"; and therefore both the global and local pictures are equivalent and will give the same results. In reality, there could always be finite hopping between different quantum wells. While the Bloch states are still exact eigen-states, the localized states are not anymore. However, if the interaction is small, then the induced modification should be also small; and the localized picture should still give an approximately correct results.

Such situation is quite similar to the case of comparison between direct diagonalization and momentum space diagonalization at the end of section 2.3.2, that direct diagonalization is always mathematically valid, however not always physically intuitive; while momentum space diagonalization provides better physics picture, but are limited to certain conditions. Here the global picture **always gives the correct results**, but may prevent one to capture what is really happening. For example, the Bloch state in Eq. 2-48 does take form of a periodic state with momentum $\boldsymbol{k}$. However the momentum here is somewhat artificial, snice it does not change anything. On the contrary, the local picture directly obtain ψ_n, which contains the really important information. Furthermore, adopting the local picture can significantly reduce the computation efforts since only one or a few unit cells require consideration, as opposed to the potentially thousands of unit cells in a superlattice.

We have now solved the "conflict" between the global and local pictures by finding out that there is actually no real conflict. Instead, the latter can be an approximately correct and potentially more intuitive picture under some conditions. The remaining question is, what these conditions are. While the condition to use local picture is quite straightforward for an array of quantum wells, it is far less clear in the context of van der Waals systems.

One can make an educated guess, however, that the answer should be "**strong coupling**" regime. The reason is simple: the eigen-states in an isolated monolayer are described by Bloch states with equal amplitude in each unit cell. However, the local picture requires that the eigen-states after interlayer interaction become localized in one or few unit cells out of thousands, which is definitely a "strong" modification to the system. Moreover, from simple uncertainty principle, a localized state in real space must be a superposition of Bloch states of a large range in momentum space. To involve momentum of a large range, many steps are required in the

iteration process in section 2.4.3, which is exactly the criterion for "**strong coupling**" regime that we have discussed in section 2.5.2.

2.6.3 Conditions of Strong Coupling Regime

Having established the relation between the local picture and "**strong coupling**" regime, we now have an alternative way to understand the quantitative condition of "**strong coupling**" regime, as it should also be the condition for using the local picture; and we can thereby get some inspiration from the simple quantum well case.

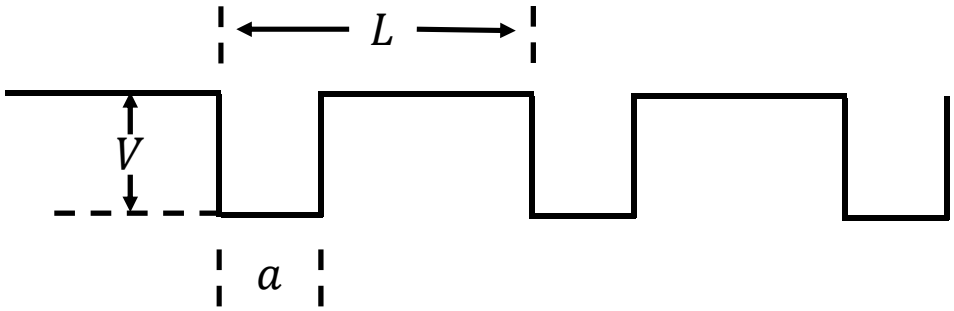

Figure 2.2 Illustration of an array of quantum well with periodicity L. Each quantum well has width a and height V.

Assuming we have an array of quantum well with periodicity L; and each quantum well has width a and height V, as illustrated in Figure 2.2. The eigen-states and energy for an isolated quantum well is well-known:

$$E \sim \frac{\hbar^2}{2ma^2}, \kappa \sim \frac{\sqrt{2m(E-V)}}{\hbar} \quad (2-49)$$

Where E is the kinetic energy of the particle, and κ is its momentum outside the quantum well. For a bound state, we have $E < V$, and therefore the amplitude of wavefunction will decay outside the well, with characteristic length

$$\lambda \sim \kappa^{-1} \sim \frac{\hbar}{\sqrt{2m(E-V)}} \quad (2-50)$$

As is discussed in the last section, using the local picture requires that the two neighboring quantum wells have little hopping with each other, therefore:

$$\lambda \ll (L-a) \to V \gg \left(\frac{\hbar^2}{2ma^2}+\frac{\hbar^2}{2m(L-a)^2}\right) \sim \frac{\hbar^2}{2mL} \sim E_k \quad (2-51)$$

We have assumed in the last step that a and $(L-a)$ are comparable, which should be true in a van der Waals system where local configuration is continuously changing over space. In reality, there are cases where the change is not continuous, e.g. when the strain between layers are concentrated at some narrow domain walls. One can adjust the ratio between a and L in the above toy model accordingly.

The condition given in Eq. 2-51 can be intuitively understood as a competition between kinetic energy and the potential energy, since it is natural for a particle to be localized if the confining potential energy is much greater than its kinetic energy. The interesting part here, however, is that the relevant kinetic energy is determined by the periodicity of the superlattice, which can be conveniently tuned in the parameter space of interlayer configuration. Such tunability could give rise to rich possibilities, which will be further discussed in section 2.7.

We can now go back to Eq. 2-20 to see if the condition for "**strong coupling**" obtained here from an intuitive picture does match the one we expect from pure mathematical derivation.

To have a more direct comparison, let us consider the effect potential picture, which focus on the mix between two states within the same layer. These type of terms correspond to even number of lines in Eq. 2-20, with the first term:

$$\sum_{r_1\boldsymbol{q}_1}\sum_{r_2\boldsymbol{q}_2}|r_2\boldsymbol{q}_2\rangle\frac{\langle r_2\boldsymbol{q}_2|V_{eff}|pk\rangle}{(E_{p\boldsymbol{k}}-E_{r_2\boldsymbol{q}_2})} \sim \sum_{r_1\boldsymbol{q}_1}\sum_{r_2\boldsymbol{q}_2}|r_2\boldsymbol{q}_2\rangle\frac{V_{eff}}{\Delta E} \quad (2-52)$$

The amplitude of wavefunction mixing is in the order of $\frac{V_{eff}}{\Delta E}$, where ΔE is the energy difference between two states. In the approximation of large Moiré superlattice (Eq. 2-27), the momentum difference between the two states should be $(\boldsymbol{q}_2-\boldsymbol{k})=\Delta\boldsymbol{k}\sim\boldsymbol{b}^M$. For states near the band edge, this correspond to an energy difference of $\Delta E\sim\frac{\hbar^2(b^M)^2}{2m}$.

Similarly, the second term (4th line in Eq. 2-20) is in the order of:

$$\sim|r_4\boldsymbol{q}_4\rangle\frac{V_{eff}^2}{(E_{p\boldsymbol{k}}-E_{r_2\boldsymbol{q}_2})(E_{p\boldsymbol{k}}-E_{r_4\boldsymbol{q}_4})}\sim|r_4\boldsymbol{q}_4\rangle\left(\frac{V_{eff}}{\Delta E}\right)^2 \quad (2-53)$$

We have used the fact that the since $(\boldsymbol{q}_4-\boldsymbol{q}_2)$ and $(\boldsymbol{q}_2-\boldsymbol{k})$ both $\sim\boldsymbol{b}^M$, so does $(\boldsymbol{q}_4-\boldsymbol{k})$. We can further write down higher order terms, which, following the same argument, should scale like $\left(\frac{V_{eff}}{\Delta E}\right)^n$. Therefore, how fast the series converge, and how many steps are required in the iteration process, will be ultimately determined by the comparison between V_{eff} and ΔE; and the strong coupling condition should be $V_{eff}\gg\Delta E$. Note that ΔE here is exactly the kinetic energy corresponding to the momentum of Moiré superlattice, we therefore have reached the same conclusion as in Eq. 2-51.

2.6.4 Equivalence between the Two Pictures: 1D Toy Model Example

We have now resolved the fundamental "conflict" between the two pictures: the local picture can be an approximately correct, and physically more intuitive language in the "**strong coupling**" regime. However, we have yet answered all questions raised at the beginning of section 2.6.2, such the meaning of "local bandgap" in the global picture.

To further confirm the condition of using the local picture, as well as to establish a correspondence to the global picture for various concepts, here I will take a simple 1D toy model as an example.

Bandgap Change Obtained from Both Pictures

Considering the simplest possible periodic system, a 1D atom chain with simple unit cell, with lattice constant a and nearest-neighbor hopping $(-t)$. The eigen-energy and eigen-wavefunction of the system is well known to be (if not, one can refer to section 2.2):

$$E^k = -2t\cos(ka) \quad (2-54)$$

Now let us simulate the superlattice effect by adding a Moiré potential on top of the above 1D-lattice, with the form:

$$V(x) = -V_0 \cos\left(\frac{2\pi}{Na}x\right) = -V_0 \cos(k_0 x) \quad (2-55)$$

Which corresponds to a periodic potential (in real space) with periodicity Na. We can then try to solve the modified system with both the global and local pictures, and see at what condition they will give consistent results.

One of the most important properties of a system is its bandgap. Assuming that there are some other bands below the energy spectrum in Eq. 2-54 that are not affected by $V(x)$ (for example, from other orbitals not considered in Eq. 2-54 that do not have interlayer interaction), the bandgap of the system will be solely determined by the minimum energy in Eq. 2-54. For unperturbed system, the band minimum will be at $k = 0$, with $E = -2t$.

Let us first use the local picture to calculate the bandgap change with the periodic potential described in Eq. 2-55. The local picture suggests that we can consider the "local environment" of each unit cell, and calculate a "local bandgap" for each of them. Note that **this does not mean** that we consider a system of an isolated single unit cell, since it is apparently unreasonable to remove any intralayer bonding. Instead, the "local environment" only refers to the interlayer interaction part, or the effective potential in the above example. Therefore, the correct way of adopting the local picture is to determine the local environment (i.e. interlayer configuration,

Moiré potential, etc.) at a given point, and assume that this local environment applies to everywhere. For example, a twisted bilayer graphene may locally resemble AB or AA stacking. To calculate the local properties, we should in turn consider an AB or AA stacked bilayer graphene, i.e. a periodic bilayer graphene that is AB- or AA- stacked everywhere.

Back to the 1D model, for a unit cell at position $x = na$, the "local environment" is described by a potential $V_n = -V_0\cos(\frac{2\pi}{N}n)$. As discussed above, obtaining the local properties requires considering a system with $V(x) \equiv V_n$ everywhere. i.e. a constant external potential. The "local energy dispersion" then simply reads:

$$E_n^k = -2t\cos(ka) + V_n$$
$$E_n'^g = E^g + V_n \quad (2-56)$$

Here E^g is the bandgap of the unperturbed system, and $E_n'^g$ is the "local bandgap" at n^{th} unit cell with the external potential.

To have a direct comparison with the global picture, it will be helpful to obtain a "global bandgap". Intuitively, the "global bandgap" should be the smallest one among all "local bandgap", because within which no state should exist. From Eq. 2-56, the "global bandgap" should be at $n = 0$ with

$$E'^g = E^g - V_0 \quad (2-57)$$

Interestingly, the "global bandgap" obtained from the local picture is not dependent on the periodicity of $V(x)$. This result is definitely not unconditional: from the global picture, the energy shift of $k = 0$ state (in second perturbation regime) can be directly obtained as:

$$\Delta E^{k=0} \sim -\frac{V_0^2}{2(E^{k=k_0} - E^{k=0})} \quad (2-58)$$

which not only depends on k_0, but also scales with V_0^2. The discrepancy here is expected since the local picture **should not work** in weak perturbative regime.

To quantitatively evaluate the deviation between the global bandgap change obtained from the two pictures, we define the following error function of global bandgap change:

$$\varepsilon(V_0, N) = \frac{\Delta E^g_{local\ picture} - \Delta E^g_{global\ picture}}{\Delta E^g_{local\ picture}} = \frac{-V_0 - \Delta E^g_{global\ picture}}{-V_0} \quad (2-59)$$

in which we have used $\Delta E^g_{local\ picture} \equiv -V_0$. $\Delta E^g_{global\ picture}$ can be obtained numerically following methods in section 2.2, which should generally depend on both V_0 and N.

Figure 2.3A shows numerical results of the error function ε depending on the potential amplitude V_0 and periodicity N. Apparently, the deviation between the two pictures decreases with increase of both the potential amplitude V_0 and periodicity N. This result is consistent with Eq. 2-51, which states that the local picture can give approximately correct results when potential energy

V_0 is much larger than the kinetic energy E_k associated with superlattice periodicity. Here the kinetic energy E_k is:

$$E_k = \frac{\hbar^2 k_0^2}{2m} = \left(\frac{2\pi}{N}\right)^2 t \quad (2-60)$$

in which we have used the effective mass $m = \frac{\hbar^2}{2a^2 t}$ obtained from Eq. 2-54. Increasing N will decrease the kinetic energy E_k, therefore making the local picture more accurate.

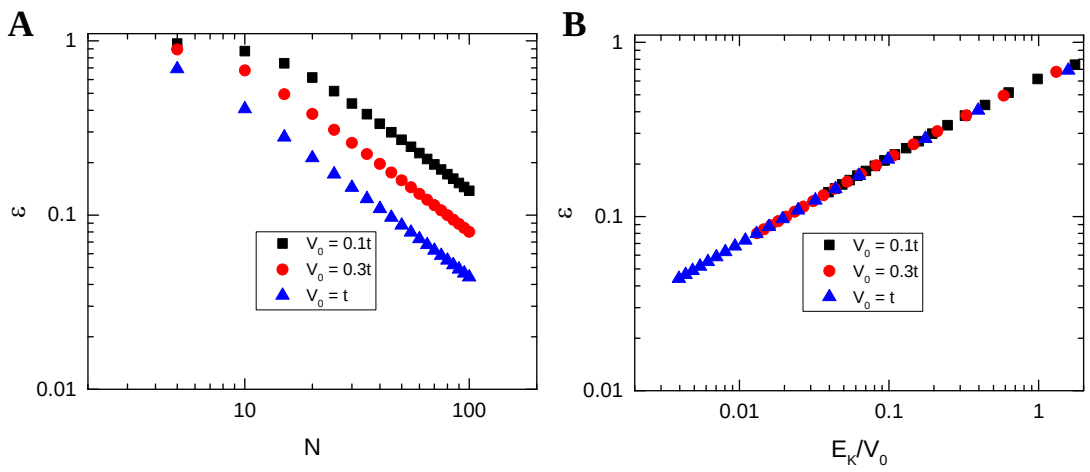

Figure 2.3 (**A**) Error of global bandgap change as a function of potential amplitude V_0 and periodicity N. (**B**) The error between two pictures only depends on the ratio between kinetic potential energy.

We can further plot the error ε directly as a function of the ratio between kinetic and potential energy $\eta = E_k/V_0$, as shown in Figure 2.3B. Interestingly, all data points now collapse onto one straight line, indicating that the ε should be function of only one parameter η. Such observation further confirms our conclusion in Eq. 2.51, that the only important criterion to determine the validity of the local picture, as well as the strong coupling regime, is the ratio between kinetic and potential energy.

Now we can now take a closer and more quantitative look at the relation between ε and η. First, it would be good to know that how small a ratio η is necessary for the local picture to give a reasonably accurate bandgap change. From Figure 2.3B, when the ratio $\eta = 0.1$, the error $\varepsilon \sim 20\%$. If we consider this amount of error as "barely acceptable", then it means that the potential energy needs be at least one order larger than the kinetic energy. In reality, how small a ε is acceptable usually depend on actual systems, such as other relevant energy scales. A few remarks and thoughts on this will be provided in section 2.6.5.

In addition, the $\eta - \varepsilon$ dependence behaves like a straight line in Figure 2.3B, implying a power law scaling. One can directly obtain the power index to be $1/2$ from the slope of the straight line. We therefore have:

$$\varepsilon \sim \eta^{1/2} \sim N^{-1} V_0^{-\frac{1}{2}} \quad (2-61)$$

Localization of Wavefunction at the Global Bandgap

Now let us take a look at the wavefunction of the $k = 0$ state at band minimum. According to discussion in section 2.6.2, its wavefunction should be localized when the local picture becomes approximately correct.

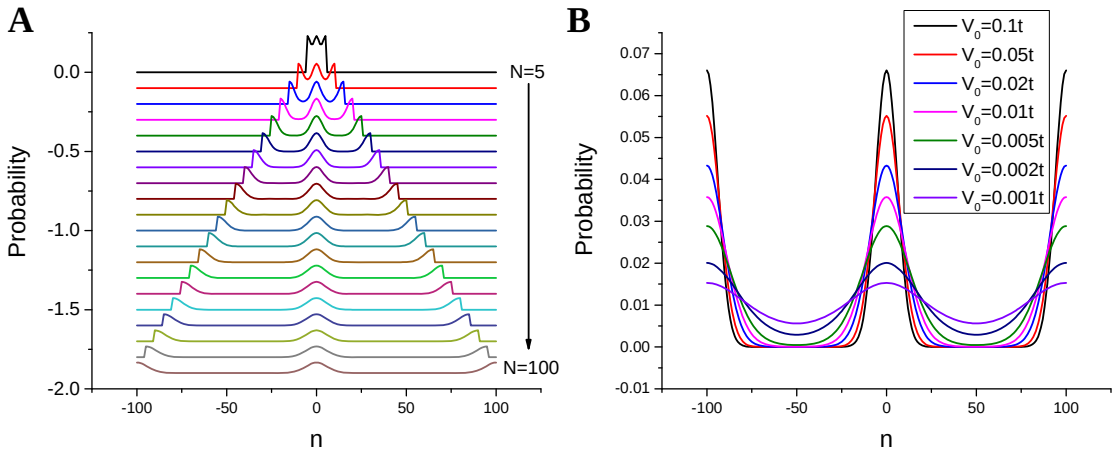

Figure 2.4 (**A**) Probability distribution of the $k = 0$ state with external potential of different periodicity N. Only two superlattice unit cells are shown (each corresponding to N original unit cells). $V_0 = 0.1t$. Each curve is shifted downward for visual clarity. (**B**) Probability distribution of the $k = 0$ state with external potential of different amplitude V_0. $N = 100$.

Figure 2.4 shows the probability distribution of $k = 0$ state in real space as a function of the periodicity and amplitude of the external potential. Only two superlattice unit cells are shown (i.e. new unit cell with the external potential), which can be repeated on both sides to obtain the probability distribution over the entire 1D space. Clearly, we find that the probability becomes more "localized" (compared to superlattice unti cell) with increasing N and V_0, consistent with expectation from the simple picture we introduce in section 2.6.3.

Wavefunction of States away from the Global Bandgap

Up to here, the results in this section from the toy model are merely confirmation of conclusions in previous sections. Now we can try to answer the meaning of "local bandgap" in the global picture. In Eq. 2-56, we have already obtained the size of "local bandgap" at each unit cell with index n: Given a specific n, the "local bandgap" takes a value of $(E^g + V_n)$, which is between $(E^g - V_0)$ to $(E^g + V_0)$ over the whole atom chain. One may guess that the state at the "local bandgap" should again be localized at unit cell n, similar to the case for the global bandgap state at $n = 0$.

However, we may then ask another question: considering the local band dispersion 2-56 at any unit cell $n' < n$, there should also be a state of the same energy $(E^g + V_n)$. What is the wavefunction distribution of these states? If they are also localized in each unit cell, we will end up with n degenerate states each localized at a unit cell. On the other hand, if we take every two unit cells as the smallest unit in the local picture, following the same argument, we would expect $n/2$ denegerate states at this energy. Nevertheless, the local picture necessarily requires that the external potential change very slowly over space; therefore considering one or two unit cells as the smallest unit should not make any difference.

The only way to resolve the above conflict is to admit that there is only one state at this energy, which spans over all the n unit cells. Apparently, such conclusion always holds regardless of the smallest unit chosen in the local picture. Furthermore, we can handwavingly provide an explanation: Because unit cells with $n' \leq n$ all have a state at this energy, the state at this energy should be at all unit cells with $n' \leq n$. This argument may seem a little weird, but one can always refer to the quantum well picture in section 2.6.3 to help understand.

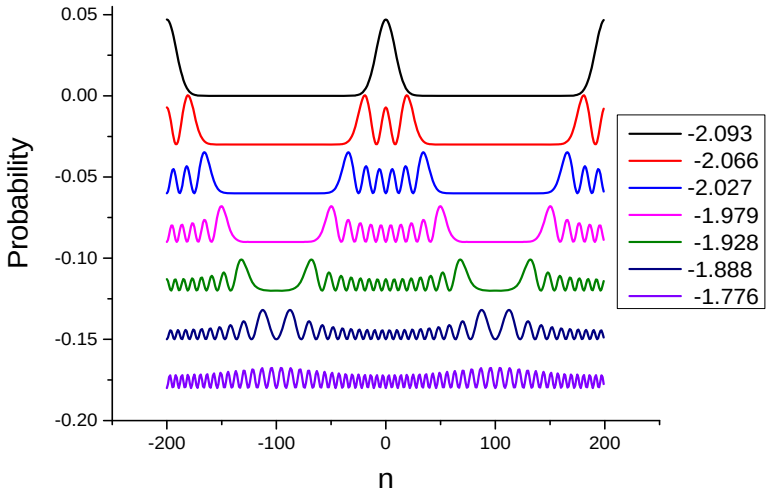

Figure 2.5 Probability distribution of eigen-states at different energy. All energy in the legend is in unit of t. $V_0 = 0.1t, N = 200$. State at the global bandgap (black) is localized at the center in real space; while

states at "local bandgap" spans over a larger range. Once the energy is above the maximum of all local minimum, states become completely delocalized.

Figure 2.5 shows numerical results of wavefunction probability distribution for states at different energy. Indeed, we find that states away from the global bandgap are not as localized; instead, they span a larger range up to a point beyond which its energy will be lower than the "local band minimum". Such observation is exactly what we expect from the above discussions. If the energy further increases and becomes higher than all the "local band minimum", then the corresponding state should be everywhere. Here the maximum of all the "local band minimum" is roughly at $(-2t + V_0) = -1.9t$. Therefore the two states in navy and purple are completely delocalized.

2.6.5 Practical Considerations When Using the Local Picture

We have finally answered all questions about the relation between the local and global picture. Here I want to give a brief summary of system properties that can be conveniently obtained from the local picture (especially form an experimental perspective) when it is valid, and things one need keep in mind to ensure that it is indeed valid.

The spatial distribution of "local bandgap" is readily available information one can obtain with the local picture. As we have discussed in last section, the "local bandgap" at a certain point indicates that no state within it can span to this point. i.e. the local density of states (LDOS) at this point should be zero for the whole energy range within the "local bandgap". Experimentally, LDOS can be directly measured from scanning tunneling spectroscopy (STS) technique; therefore the spatial distribution of "local bandgap" directly gives the STS mapping results in experiment, and vice versa.

One can further obtain the wavefunction distribution of states at any given energy by adopting the criterion that "if this energy exists in the local energy spectrum of one point, the corresponding state should span to this point". We have already seen a simple example in section 2.6.4, and one can easily extend it to more complicated cases, such as a monolayer graphene with a mass term $m(\boldsymbol{r}) = m_0 \cos(\boldsymbol{k} \cdot \boldsymbol{r})$. It is well-known that the mass term opens a gap in energy spectrum, with the gap size equal to the amplitude of the mass term. As a result, any state at energy $E < m_0$ should be localized to the range $m(\boldsymbol{r}) \leq E$; and the most localized state will have energy $E = 0$, which is completely localized at $m(\boldsymbol{r}) = 0$. An alternative way to reach the same conclusion is by using the language in topological insulator, that the Hamiltonian for two regions with $m(\boldsymbol{r}) > 0$ and $m(\boldsymbol{r}) < 0$ cannot continuously connect to each other without closing the gap, therefore a gapless state must emerge at the edge of $m(\boldsymbol{r}) = 0$.

Given the usefulness of the local picture, one has to be particularly careful regarding its validity. The first thing to check is of course if the strong coupling condition 2-51 is reached. It is often

difficult to predict if the condition will be satisfied before doing calculation. One can therefore adopt the concept of reductio ad absurdum by assuming that the local picture is valid to start with, which greatly reduces the computation effort and gives important information such as local band dispersion. With all this information, one can go back and see whether the strong coupling condition is satisfied, which should now be easy.

How accurate the local picture need to be often depends on actual system. Still using the example of bandgap change: if the error of the local picture is much less than any relevant energy scale, then it is acceptable. One relevant energy scale is of course the bandgap itself, and it is apparent that the result will not be acceptable if error becomes a large portion of it. There can also be other relevant energy. For example, if there are multiple band extrema in the system and some are very close in energy, the energy difference between them will then become an important energy scale; since error on this order may lead to incorrect conclusion on whether the system will have a direct or indirect bandgap. Thermal energy is another energy scale often involved in various processes. Therefore, one needs carefully evaluate the error and its effect before drawing any conclusion from the local picture.

Even if the strong coupling condition is reached, there are still potential problems one will encounter when using the local picture. For example, how to calculate the local properties from local configuration? It is quite straightforward for bilayers with the same unit cell size, such as twisted bilayer graphene that we have discussed in section 2.6.4. What if the unit cells are different, such as WSe2 on WS2? To make computation possible, we still need a periodic boundary condition. Given the few percent lattice mismatch between them, we have three choices: compress the WSe2 unit cell by ~4% percent to match with WS2; or stretch the WS2 unit cell by ~4% percent to match with WSe2; or just let both lattices relax (which most likely leads to both stretch of WS2 and compress of WSe2 to match with each other). Which one is correct?

To answer this question, we can again refer to the criterion we introduced in section 2.6.4, that we want to maintain the intralayer coupling part, and consider the local interlayer configuration. Therefore, we should fix the intralayer hopping of each layer (i.e. assume that it does not change with bond length). If this is not practical, a compromised way is that when we are calculating properties of WSe2, we can fix its unit cell and stretch WS2 to match with it; and do the opposite when calculating properties of WS2. In this way, we at least have not changed the intralayer part of the target layer. On the other hand, doing it in an incorrect way will introduce few percent of artificial strain, which can easily give error of hundreds of meV.

As an ultimate criterion, one can always compare to the global picture, if possible, to ensure that the results from the local picture are correct. By combining with the global picture, one can also use the local picture to only provide an intuitive physical picture for important results, which just need to be qualitatively correct; and numerical accuracy can become less important.

2.7 Summary of the Theory Section

Summary of the Classification and Methods

In the entire section 2, I have been classifying van der Waals systems into different categories in light of interlayer interaction, and presenting corresponding theoretical methodologies to investigate the properties in each case. Although discussion here is focusing on single particle picture, one can always further turning on many body interaction later based on the modified eigen-states. See chapter 6 for a general discussion on many body effects.

Below I briefly summarize all the categories and suitable methodologies:

In all cases, **direct diagonalization** is mathematically correct, however not necessarily practical or intuitive.

Depending on interlayer configuration, the system can be classified into "**commensurate**" and "**incommensurate**" cases.

In **commensurate** case, **direct diagonalization** is practical and can be a good approach. (Section 2.2 and 2.3)

The **Incommensurate** case, depending on the amplitude and periodicity of interlayer interaction, can be further classified into "**strong coupling**" and "**weak coupling**" regime.

In case of "**strong coupling**", the "**local picture**" can be a suitable method. (Section 2.6).

In case of "**weak coupling**", **momentum space diagonalization** is a generally applicable tool. (Section 2.4)

Interlayer interaction in the "**weak coupling**" regime can be further expanded into different orders, and classified depending on the important order. The two lowest-order effects can be described by "**direct mixing**" and "**effective potential**" approaches. (Section 2.5)

One will find that the above classification and methodologies almost cover all van der Waals systems in reality. Therefore, one can always readily choose the corresponding approaches for an actual problem.

New Rich Possibilities from van der Waals Systems

Providing a thorough classification and investigation methods for van der Waals systems is only one major goal of this section. In addition, I hope that the all these discussions have also provided a general picture that the van der Waals systems are indeed interesting, and can offer fascinating physics. While at the beginning of this section I have already given a brief introduction that interlayer interaction in low dimensional systems can create new rich

possibilities; now with all the knowledge and examples given in the section, we can look back to gain a deeper understanding of why this is the case.

The richness of physics from the large parameter space of interlayer interaction can already be seen from the many different regimes classified in the section. The commensurate coupling case, which only accounts for a tiny little bit of the parameter space, is relatively easy to study and usually investigated first. On the other hand, the incommensurate case occupies a much richer parameter space and is becoming center topic of research in many systems. The fundamental reason to all these new possibilities can be traced back to the general rule governing the behavior of a physical system: A real physical system always involve multiple relevant energy, length, time etc. Their competition largely determine the ultimate phase of the system. For example, whenever we see a "critical temperature", most likely some other energy scale is competing with thermal energy; and type II superconductor is an interplay between superconducting coherent length and London magnetic field penetration depth.

The **first unique advantage** of van der Waals system is that, one can artificially create a length scale and/or energy scale to compete with other existing ones, which might not be achievable in traditional systems. For example, the observation of Hofstadter butterfly pattern requires the periodicity of the system to be comparable to the magnetic length, which demands an impractically high magnetic field in traditional solid state systems. However, van der Waals structure can easily create such periodicity through formation of Moiré superlattice.

Things can become even more interesting if we consider the strong coupling regime. As discussed in section 2.6, the relevant kinetic energy now is determined by the periodicity of the superlattice. While kinetic energy is ubiquitous for every particle, it is also one of the most uninteresting objects to study: When the kinetic energy is too high, the particle becomes effectively a free particle, which is simple but trivial. On the other hand, when other interactions are able to compete, we start to see variety of fascinating behavior, such as localization, paring, quantization, Mott transition etc. However, practically it can often be challenging to lower the kinetic energy to a point that other specific energy start to play a critical role. This problem can be again solved in van der Waals systems, as we can readily achieve a very low kinetic energy through a large superlattice periodicity; which enables exciting opportunities when other interaction, such as Coulomb interaction, starts to dominate.

Naturally, the **second advantage** of van der Waals system will be tunability. Closely following the above discussion, if it is possible to bring the system into a new phase not achievable previously, there must be also a transition region during the process. Since the energy scale we "create" is readily tunable through design of the interlayer configuration, it would provide a convenient and unique opportunity to observe the complex interplay of different scales near the transition, such as quantum criticality.

The **third advantage** might look more subtle, as we discussed the section 2.5, that the effect the interlayer interaction cannot be described by a simple electrostatic potential in most cases, due to the involvement of several energy and length scales. In contrast, it is generally a matrix that can

directly access multiple degree of freedoms (DOFs) of the system, e.g. the pseudospin DOF in graphene. Many components of the interlayer interaction can be difficult to achieve with any other existing field in a solid state system. Therefore again new possibilities become available. Furthermore, all these components are tunable through careful design of the system.

Given all these exciting new possibilities, the field of van der Waals systems is only at a starting point. For example, while many interesting results have already been reported from "weak coupling" systems, the "strong coupling" regime is still relatively unexplored, which could be even more interesting. Part of the reason is the potentially higher requirement on sample quality, which is becoming less and less a problem nowadays thanks to the rapid progress on sample preparation and device fabrication technique, such as polymer transfer, BN encapsulation, graphite electrode, etc. Therefore, it is reasonable to expect that van der Waals systems will keep acting as a seedbed from which fascinating physics will grow.

Chapter 3 – Bandgap Engineering in Phosphorene

3.1 Introduction and Background

Since the discovery of 2D crystals, one major pursuit is always towards their application in electronics and photonics, given their intrinsic advantages over bulk materials (e.g. Silicon and GaAs) on aspects such as scalability and power consumption (*12, 28, 35-40*). The electronic and photonic applications of these materials will fundamentally depend on whether they are metal, semiconductor, or insulator; and what type of light they can have strong interaction with, which are ultimately determined by their energy gap. In this sense, one of the most important task is always to find a 2D material with the correct bandgap for specific device. This can certainly be done by searching from the rich bank of layered materials and trying thinning down each to see if any will match with the goal. Indeed, intensively studied 2D materials now include semi-metallic graphene with zero bandgap, semiconducting transition metal dichalcogenides (TMDs) with 1.5~2eV bandgap, and insulating hexagonal boron nitride (hBN) with giant bandgap ~5.8eV. However, achieving desirable bandgap by searching new materials often involves lots of trial and error, and may not match exactly with the desired value.

Alternatively, one can try to modify the bandgap of an existing material. The advantage is apparent: one can often achieve greater tunability, which allows for fine tuning and optimization for specific purposes. Depending on the method used for bandgap engineering, such modification may not be necessarily static, enabling further possibilities such as high speed modulator.

As discussed in chapter 2, one unique tool for bandgap engineering in 2D materials is through interlayer interaction. To maximize the tuning range, we want the interlayer interaction to be as strong as possible. According to section 2.4, this is usually achieved through "direct mixing" at **Γ** point in the Brillouin Zone. One example is illustrated in section 2.4.4 on bilayer MoS2, in which the strong interlayer interaction at **Γ** point shift the valence band maximum to **Γ** point in bilayer. Following the same argument, if one material has a direct bandgap already at **Γ** point in monolayer, we would expect it to maintain a direct bandgap when getting thicker, meanwhile with changing bandgap size. Monolayer phosphorene (or monolayer black phosphorous) is one such material. In this chapter, we will investigate the layer-tunable electronic structure in phosphorene from both theoretical and experimental perspectives. Part of the discussion is also presented in a published paper (*1*).

3.2 Tunable Electronic Structure in Few Layer Phosphorene

Monolayer phosphorene has a predicted direct bandgap of ~1.5eV at **Γ** point. Among the five valence electrons of phosphorus atom, three will form near sp^3 bonds with the three adjacent atoms (since it is "near" sp^3, the angle will be different from exact sp^3 bonds); and the other two form a lone-pair, giving a puckered honeycomb structure (*41*), as illustrated in Figure 3.1.

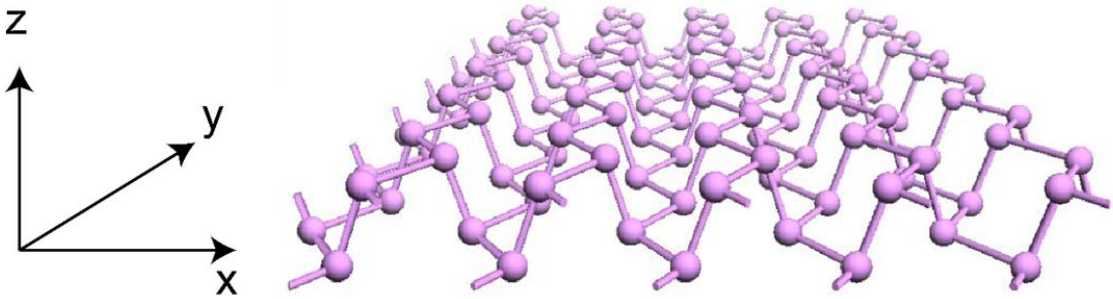

Figure 3.1 Atomic structure of monolayer phosphorene. Each phosphorus atom is covalently bonded with three adjacent phosphorus atoms, forming a puckered honeycomb structure

Following the discussion in section 2.4.4, the situation for **Γ** point coupling is particularly simple, as it is usually sufficient to only consider coupling between **Γ** point states from different layers. Here the situation is even simpler in that few layer phosphorene is regularly stacked without any twist angle; and therefore the coupling between **Γ** point states become a strict selection rule. One can expect, however, that even at finite twist angle, this conclusion can still approximately hold in many situations, similar to the case of bilayer MoS2. What is different here, is that phosphorene has strong in-plane anisotropy, so does the band extrema states at **Γ** point. As a result, the coupling amplitude between states at **Γ** point from different layers will generally depend on their relative rotation angle, in contrast to the MoS2 case. One can still, however, use the same formalism in section 2.4 to obtain the coupling matrix element at arbitrary configuration.

Here I will focus on the simplest case, i.e. naturally stacked few layer phosphorene, as it is experimentally ready. The bandgap change in bilayer phosphorene is already given in section 2.4.4, which can be viewed as bonding and anti-bonding states between two atoms. The extension to arbitrary layer number N is straightforward: We can treat the conduction band minimum state (which is at **Γ** point) in each layer of phosphorene as an atom with energy E^c, the interlayer coupling will be effectively the hopping between these atoms. By assuming that only coupling between two neighboring layers are important, with matrix element amplitude γ^c. The system of N layer phosphorene can be mapped into a simple model of finite 1D atom chain:

$$H_N^s = \sum_{j=1}^{N} E^s C_j^+ C_j + (\sum_{j=1}^{N-1} \gamma^s C_{j+1}^+ C_j + \text{h. c.}) \quad (3-1)$$

where $s = c, v$ stands for conduction and valence band, respectively. C_j^+ / C_j $(j = 1 \dots N)$ is the creation/annihilation operator on j^{th} atom (i.e. j^{th} layer), $C_j^+|0\rangle = |\phi_j\rangle$ is the state from j^{th} atom (i.e. **Γ** point state from j^{th} layer), γ^s is real since $\langle\phi_1|\gamma^s|\phi_2\rangle = \langle\phi_2|\gamma^s|\phi_1\rangle$. The eigen wavefunction of the Hamiltonian can be generally written as $|\psi\rangle = \sum_{j=1}^{N} a^j |\phi_j\rangle$, which has a clear physical meaning that the eigen wavefucntion $|\psi\rangle$ of N layer BP have contribution from each layer; and that the coefficient a^j contains full information of how different layers are mixed. We have assumed that conduction band states do not interact with valence band states, which is reasonable given their large energy separation.

The eigen-energy and wavefunction of Eq. 3-1 can be solved by directly diagonalizing the N-by-N tridiagonal matrix. Alternatively, here we will use a simpler and more intuitive approach, owing to analogy to the 1D lattice. An infinite 1D lattice supports a travelling wave, with eigen wavefunction $a^j = e^{ikj}$ and eigenenergy $E_k = E^s + 2\gamma^s cosk$. For a finite 1D lattice, the two atoms at the end have only one nearest neighbor (to the inner side of the chain), and the boundary condition is equivalent to having two additional fixed atoms on each end (i.e 0^{th} and $(N+1)^{th}$ atoms with $a^0 = a^{N+1} \equiv 0$). This finite 1D lattice of N+2 atoms with fixed ends can be easily solved, giving N eigen-states and corresponding eigen-energy:

$$a_{N,n}^j = \sin(k_{N,n} j), k_{N,n} = \frac{N+1-n}{N+1}\pi, n = 1 \dots N$$

$$E_{N,n}^{(s)} = E^s + 2\gamma^s cosk_{N,n} = E^s - 2\gamma^s cos\left(\frac{n}{N+1}\pi\right) \quad (3-2)$$

Here n indexes the N eigen-states and eigen-energy, which is in reality the n conduction and valence subbands at **Γ** point in N layer phosphorene.

These eigen-states have well-known physical meaning for a real 1D lattice: When the travelling wave gets reflected by the two fixed ends, the eigen-states will become standing waves, with different number of nodes. From the wavefunction given in Eq. 3-2, the number of nodes of each mode is simply given by $(N-n)$. Note that each atom represents one phosphorene layer; therefore the standing wave actually describes the distribution of the eigen-states over each layer, i.e. along z direction. For example, a standing wave with more nodes corresponds to a faster oscillation of the wavefunction along z direction.

The energy order between different modes will depend on the sign of γ^s. To obtain this information, one can refer to the bulk black phosphorous situation, where people already know that both the conduction band minimum and valence maximum are at **Z** point (*41*). Note that such observation does not conflict with our expectation that the direct bandgap should maintain at **Γ** point: Eq. 3-2 is from the perspective of 2D Brillouin Zone in xy plane, and **Z** point is

indeed projected to **Γ** point in xy plane, along with a fast oscillation on z direction. If we take $N \to \infty$ limit in Eq. 3-2, such fast oscillation on z direction is corresponding to the $n = N$ mode. Therefore, the fact that the direct bandgap in bulk black phosphorous is at **Z** point of the 3D Brillouin Zone indicates that γ^c is positive, and γ^v is negative in Eq. 3-2; and therefore n also indexes the conduction and valence subbands starting from the bandgap.

Having obtained the eigen-states and eigen-energy of band extrema states in N layer phosphorene, we can now readily predict the electronic and optical responses of the system. First, as expected, the direct bandgap has a sensitive dependence on layer number, which could be very useful in electronics and optics. Furthermore, the other subbands in Eq. 3-2 should give a series of interband transitions at higher energy, which can potentially access even larger energy range for application in optics. Given the standing wave-like nature of each subband, the optical transition can only happen between two subbands with the same node number, as they are otherwise orthogonal to each other. The transition between n^{th} valence to n^{th} conduction band will have the energy:

$$E_{N,n}^{R} = E_{\mathrm{g0}} - 2(\gamma^c - \gamma^v)\cos\left(\frac{n}{N+1}\pi\right) \qquad (3-3)$$

e.g. $n = 1$ corresponds to bandgap transitions, $n = 2$ and $n = 3$ describes the first and second above-bandgap resonances, respectively. $E_{\mathrm{g0}} = (E^c - E^v)$ is the bandgap in isolated monolayer phosphorene. There will be altogether N resonances.

The tunability of electronic structure can be further expanded with an external field, such as an electric field along z direction. This problem can again be mapped to a 1D atom chain with an electric field along the chain, and one can immediately write down the effective Hamiltonian:

$$H_N^S = \sum_{j=1}^{N}(E^s - jV)C_j^+ C_j + (\sum_{j=1}^{N-1}\gamma^s C_{j+1}^+ C_j + \mathrm{h.c.}) \qquad (3-4)$$

where V is the electrostatic potential drop over neighboring layers. Solving above eigen-value problem is straightforward, and will be omitted here. One can already see without real calculation, however, that the conduction band minimum and valence band maximum state will shift to the two ends of the chain (i.e. the two outermost layer); and the bandgap energy should decrease.

3.3 Experimental Observation of Layer Dependent Electronic Structure

3.3.1 Reliable Determination of Electronic Structure with Absorption

In the last section, we have theoretically investigated the highly-tunable electronic structure in few layer phosphorene, which could bridge the much-need bandgap range from 0.3eV to 1.5eV in 2D materials. Given the conciseness of Eq. 3-2 and 3-3 that describe subbands and optical transitions through only two parameters with well-defined physical meaning; it seems straightforward to link theory and experiment to further confirm both the theoretical picture and experimental results.

However, this seemingly simple task has not been done until recently. Actually, previously measured layer-dependent optical bandgap in phosphorene is often compared to a power-law model $E_N^R = aN + bN^2 + \cdots$, where the coefficients are fitting parameters without clear physical meaning.

Part of the reason is that, almost all previous experimental measurements are using photoluminescence (PL) spectroscopy to probe the optical band gap (*37, 42-44*). While PL is a powerful and convenient technique to investigate the properties of materials, it has certain shortcoming, among which one of the most critical is its sensitivity to sample quality. Actually, the ratio between photon number in excitation light and PL emission, the quantum yield, is often used as a figure of merit to evaluate the quality of a certain sample. Such sensitivity is ultimately coming from the fact that PL emission involves relaxation process: the spontaneous emission rate of a given type of excited state can often be considered constant, therefore how much light it emits in its life largely depend on how long it lives, i.e. its total lifetime. The existence of defect can greatly decrease the total lifetime of excited states by providing additional non-radiative decay channel, which often dominate the overall decay process.

The defect states can not only significantly eliminate the PL signal we want to study (e.g. from the direct bandgap). Even worse, since they are usually quite localized and are more energetically favorable than bandgap states, once excited electron and holes fall into a defect state, it may get trapped for a longtime. As a result, although the spontaneous emission rate of the defect states is often quite low, their integrated emission can still be significant given the long lifetime. In a defect-rich sample, majority of the experimentally measured PL emission can come from the in-gap defect states, which may not contain any direct information of the bandgap. Therefore, it requires careful control experimental and data analysis to probe bandgap energy with PL measurements alone. Such problem is further amplified in phosphorene system, which is long known to be air-sensitive, and can easily become “defect-rich” if without careful sample preparation process.

To avoid inaccurate information, it would be helpful to find an alternative technique that is not sensitive to defect states. From the above discussion, the key to avoid defect effects is to exclude the relaxation in the process. One such technique is optical absorption spectroscopy. As opposed to PL spectroscopy, no relaxation process in involved in absorption spectroscopy; and the absorption cross section is only determined by the oscillator strength of a state (transition dipole moment and density of states), while not affected by lifetime (*45*). Therefore, if there is 1% defect states in the system, generally the absorption signal from defects will be at 1% level of the overall signal. Therefore the defect signal will not dominate unless the defect density become as high as the material itself. One well-known example to illustrate the difference between absorption and PL spectroscopy is the comparison between monolayer and bilayer MoS2. While the PL in bilayer is orders of magnitude weaker than in monolayer due to the direct-to-indirect bandgap transition and significantly shorter lifetime direct bandgap excitons; absorption of bilayer is roughly twice as of monolayer because lifetime is not relevant here.

An another advantage of absorption spectroscopy is that, again owing to the lack of relaxation process, we can also probe properties of states away from bandgap, which are very difficult to measure from PL because they will relax down to band edge (and further to in-gap defect states) very quickly. The information of higher energy transitions can provide direct comparison to the subbands transition with $n > 1$ in Eq. 3-3.

3.3.2 Experimental Configuration

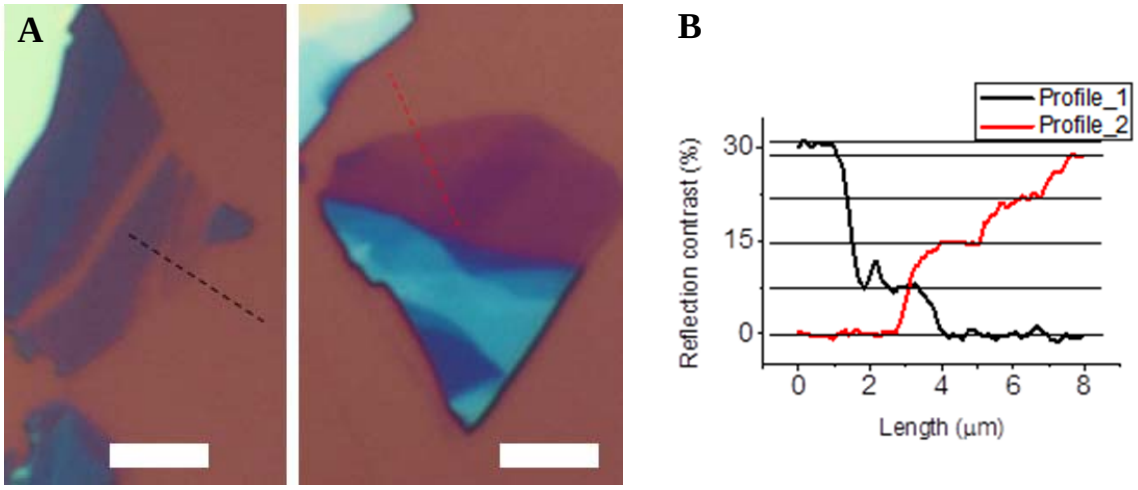

Figure 3.2 (**A**) Optical microscope image of few layer phosphorene on silicon substrate with 300nm silicon oxide. (**B**) Preliminary determination of layer number through red channel contrast.

Even with a defect-insensitive technique, we certainly still want the sample to have high quality by itself. To avoid sample degradation in air, all samples are fabricated in an inert gas glove box with oxygen and moisture levels lower than 1 ppm. In addition, we adopt BN-encapsulation to further prevent contamination: Phosphorene flakes are first exfoliated onto silicon substrates with a 300 nm oxide layer (Fig. 3.2A). The optical contrast of phosphorene flakes under optical microscope shows step like feature, which can be used to preliminarily identify layer number (Fig. 3.2B). Later this layer number information will be confirmed with absorption spectroscopy studies. Phosphorene fakes are then transferred to sapphire substrate with a top hBN encapsulation using a dry transfer technique in the glove box (*46*). During optical spectroscopy studies, the samples are kept in high vacuum.

The phosphorene samples are encapsulated between sapphire substrate and hBN, as illustrated in Figure 3.3. The optical absorption spectrum of few layer phosphorene is obtained through measuring its reflectance contrast $\Delta R/R$ at 77 K.

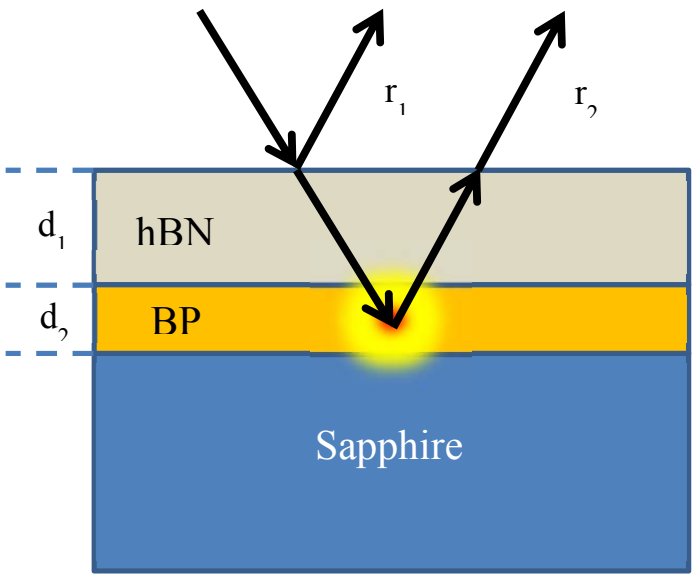

Figure 3.3 Light reflection from few layer black phosphorous (BP) under experimental configuration. The incident light is normal to sample surface, however drawn as oblique for visual clarity.

The reflectance contrast spectrum $\Delta R/R$ from phosphorene is related to its complex dielectric constant through Fresnel equation. In the visible and near-infrared spectral region, the refractive index of hBN and sapphire can be approximated as a constant with $n_{BN} = n_{Sapphire} = 1.75$ (denoted as n_1). The hBN thickness (d_1) is approximately 15 nm in our devices. We denote the phosphorene dielectric constant and effective thickness as $\varepsilon_2(\omega)$ and d_2, respectively.

Optical reflection changes from atomically thin layers such as phosphorene can be treated as a small perturbation, and it has the form of (*47*, *48*).

$$\frac{\Delta R}{R} = -Re[\sigma \cdot c_l^2 / c_r] \quad (3-5)$$

where c_l is the local field factor at the sample, c_r is the reflection coefficient, and $\sigma = \frac{\omega d_2}{i}(\varepsilon_2 - 1)$ is the two-dimensional optical conductivity of black phosphorous. The reflection is dominated by the hBN-air interface r_1 as shown in Figure 3.3, therefore $c_r = r_1 = \frac{n_0 - n_1}{n_0 + n_1}$. The local electric field at sample has the form $c_l = t_1 e^{i\varphi}$. Here $t_1 = \frac{2n_0}{n_0 + n_1}$ is the transmission from air to hBN, and $\varphi = 2\pi \frac{n_1 \omega d_1}{c_0}$ is the phase change of light propagating through the hBN layer. The final result is.

$$\frac{\Delta R}{R} = \frac{4\omega d_2}{n_1^2 - 1} \mathrm{Im}\left[(\varepsilon_2 - 1)e^{i2\varphi}\right] = \frac{4\omega d_2}{n_1^2 - 1} \mathrm{Im}\left[(\varepsilon_2 - 1)e^{i\frac{4\pi\omega d_1}{c_0}}\right] \quad (3-6)$$

If the hBN thickness is negligible (*i.e.* $e^{i\varphi} = 1$), the reflectance contrast is proportional to the imaginary part of ε_2 (*i.e.* optical absorption). However, the finite hBN thickness makes $e^{i\varphi}$ complex; thus both real and imaginary part of ε_2 will contribute to the reflectance contrast. Eq. 3-6 shows that the contribution from real part of ε_2 increases with light frequency. In reality, the thickness of hBN is quite small compared to wavelength range of measurement; and we can determine the optical absorption resonance energies from the reflection peaks with an uncertainty less than 25 meV.

We combine the supercontinuum laser and tungsten lamp to cover an energy range from near-infrared to visible. The incident light is focused onto the phosphorene flakes in by an 40X objective with 0.60NA, and the reflected light was collected by the same objective to analyze in a spectrometer equipped with both silicon and InGaAs arrayed detectors. The polarization of incident light was controlled using a broadband calcite half-wave plate placed in between the objective and beamsplitter to avoid polarization distortion from the beamsplitter. For comparison, we also measured the absorption spectrum of a thick phosphorene flake (~ 100 nm) using Fourier Transform Infrared Spectroscopy (FTIR) at room temperature, which yields the electronic structure of black phosphorus in the bulk limit.

3.3.3 Experimental Results

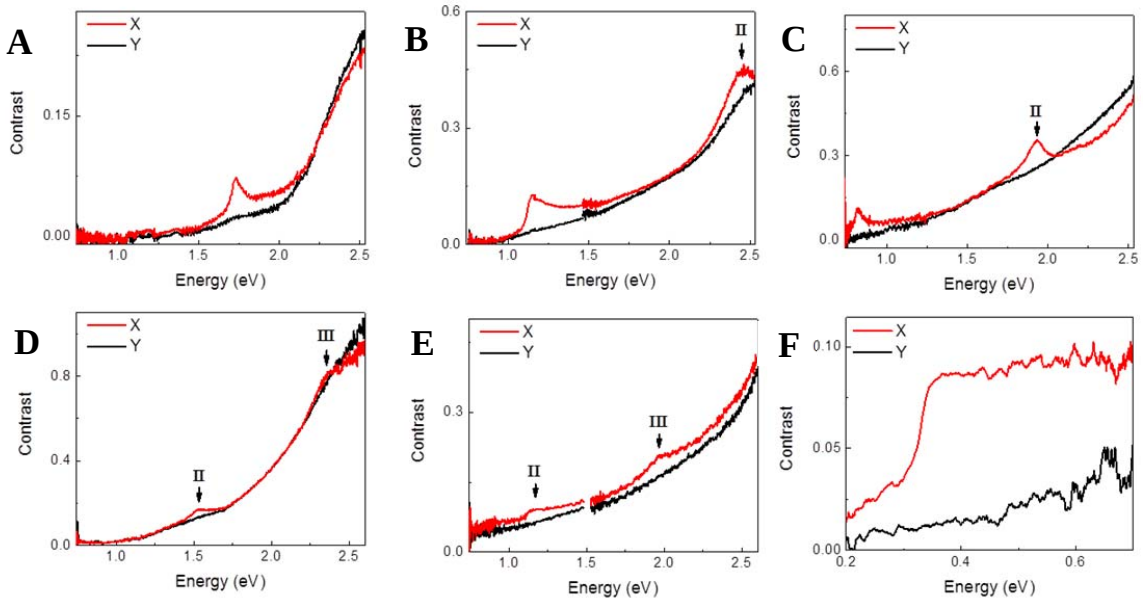

Figure 3.4 (**A-E**) Reflectance contrast spectrum of monolayer (**A**), bilayer (**B**), trilayer (**C**), tetralayer (**D**), and pentalayer (**E**) phosphorene at 77K. Red and black curves are spectrum taken with incident light polarized along x and y direction, respectively. Optical resonances only show up in x-polarization configuration, which are coming from subband transitions at center of Brillouin Zone. II and III labels transitions away from the direct bandgap with $n = 2$ and $n = 3$. (**F**) FTIR reflectance spectra of bulk black phosphorous at room temperature, the step like feature in x-polarization corresponds to the direct bandgap of ~0.35eV in bulk black phosphorous.

Figure 3.4 shows the measured reflectance contrast spectrum of mono- to penta-layer phosphorene (Figure 3.4A-E), as well as of bulk black phosphorous (Figure 3.4F). Each sample is measured with two incident light polarization along x and y direction, respectively. All spectra have prominent polarization dependence, with resonance features only showing up in x-polarization configuration. In monolayer phosphorene, a prominent absorption peak shows up at 1.73 eV for *x*-polarized incident light (Fig. 3.4A, red curve), which can be attributed to lowest energy exciton arising from direct bandgap transitions, as no optical absorption is observed for photon energies below this resonance. This optical bandgap decreases systematically with layer number, to 1.15 eV for bilayer, 0.83 eV for trilayer, below 0.75 eV (the lower detection limit of our spectral range) for tetralayer and pentalayer, and ultimately 0.35eV in bulk black phosphorous (Fig. 3.4B-F).

In addition to optical bandgap transitions, additional resonances show up at higher energy from subband transitions with $n > 1$, as expected from our discussion in section 3.2. We label the

resonances from $n = 2$ and $n = 3$ subband transitions with "II" and "III" in Figure 3.4, which also shows a systematic evolution with layer number: transition "II" does not exist in monolayer, emerges at 2.44 eV in bilayer, and then decreases to 1.93 eV, 1.52 eV, and 1.16 eV in trilayer, tetralayer, and pentalayer, respectively. Transition "III" should in principle show up in trilayer, however is still out of our experimental spectral range; therefore is only observed in tetralayer and tetralayer at 2.31 eV and 1.94 eV, respectively.

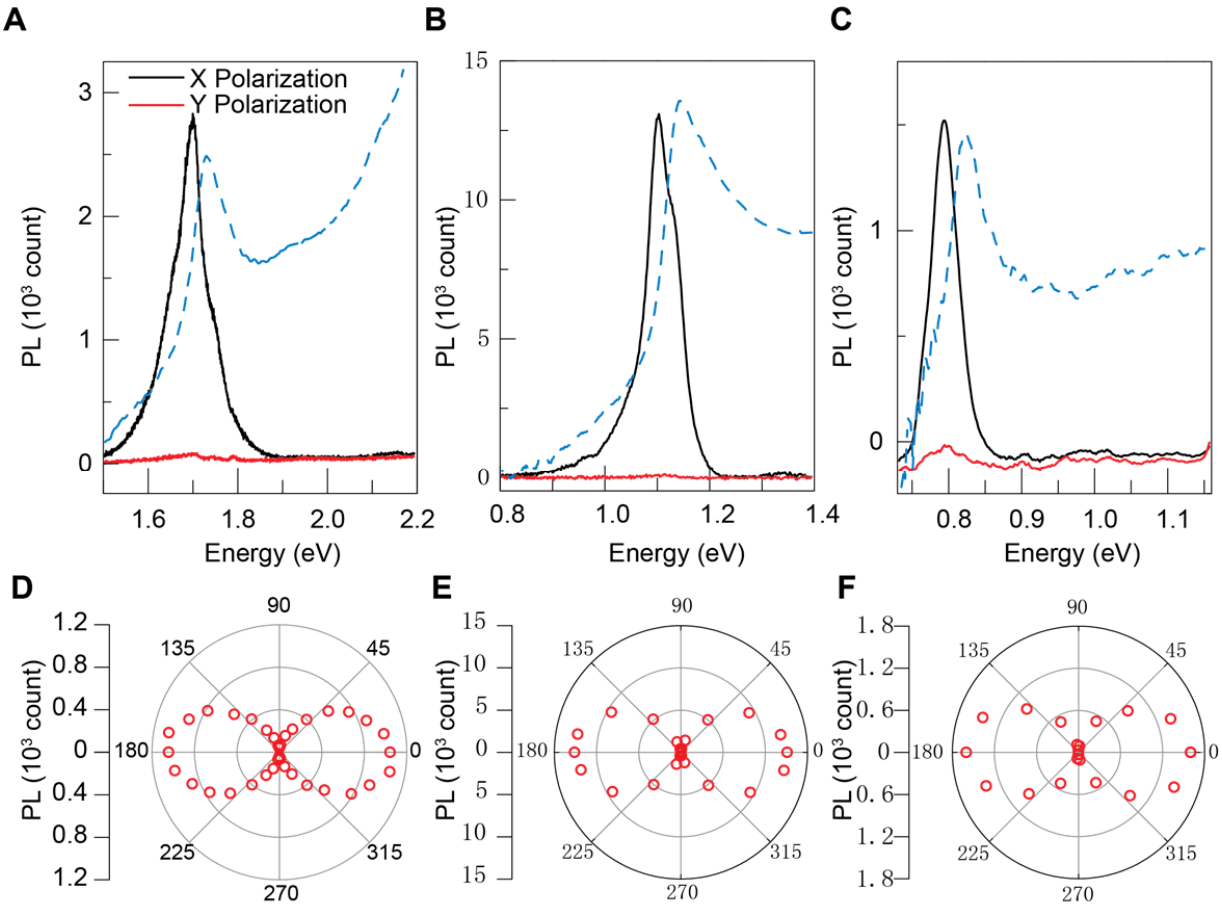

Figure 3.5 (**A-C**) Photoluminescence spectrum monolayer (**A**), bilayer (**B**) and trilayer (**C**) measured at 77K. Black and red curves are spectrum taken with incident light polarized along x and y direction, respectively. Blue dashed lines show reflectance contrast spectra of each layer as reference. The good match between resonances in absorption and PL confirms the direct bandgap nature of few layer phosphorene. (**D-F**) Polarization dependent PL intensity corresponding to (**A-C)** as a function of angle between detection direction and x direction. The near-perfect $\cos^2\theta$ dependence indicates the linear-polarized nature of PL, as expected from theory.

We can also examine the PL emission from few layer phosphorene. Figure 3.5 shows the measured PL spectrum for monolayer to trilayer phosphorene at 77K, with comparison to corresponding reflectance spectrum. The PL peaks are very close to the optical bandgap transitions in reflectance contrast, confirming the direct bandgap nature of the system. We also observe similar polarization dependence, that the emission only shows up at x-polarization. Note that here the polarization refers to detection direction, and excitation light is always unpolarized with energy of 2.33eV. Figure 3.5D-F further shows the PL emission with respect to θ, the angle between detection direction and x direction. The near-perfect $\cos^2\theta$ dependence indicates that the PL emission is almost completely linearly-polarized, though excited with unpolarized light.

3.3.4 Comparison to Theoretical Model

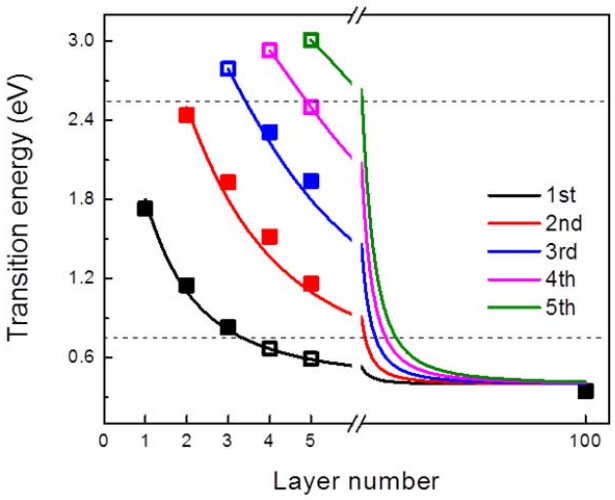

Figure 3.6 Comparing experimental and theoretical evolution of subband transition energy. Solid line is from the theoretical prediction in Eq. 3-3, with different color corresponding to subband transition with different n from 1 to 5. Measured resonance energies are shown as solid symbols, with dashed lines mark the experimentally available spectral range. Open symbols represent theoretically predicted resonances that are outside of the experimental spectral range.

With the experimentally measured subband transition energy available, we can now have a direct comparison with theoretical predictions in Eq. 3-3. Figure 3.6 summarizes all the measured resonance energy dependence on layer number as solid symbols, with different color representing different subband index n. Solid line of the same color illustrates the predicted transition energy at corresponding n, as given by Eq. 3-3. Two fitting parameters $E_{g0} = 1.8\ eV$ and $\gamma^c - \gamma^v = 0.73\ eV$ are used for all curves. The two horizontal dashed lines mark the

available spectral range in our measurement, outside which symbols are shown in open because they are theoretically predicted but not observable due to experimental limitation.

Results from the simple 1D model match quite well with experimental observation for all resonances, including the direct bandgap in bulk. Such good match further confirms the validity of both our theoretical model and experimental technique. Furthermore, we are able to quantitatively obtain important parameters of interlayer interaction, which is indeed very strong.

Another important experimental observation from Figure 3.4 and 3.5 is the apparent polarization dependence that all the absorption and emissions only show up in x-polarization configuration. For monolayer, such behavior can be understood from its symmetry: Monolayer phosphorene has mirror symmetry with respect to the xz plane. As a result, Electron wavefunctions at **Γ** point must be eigen-states to the $y \to -y$ operation, with eigen-value of either 1 or -1. Indeed, both the conduction band minimum (CBM) and valence band maximum (VBM) of phosphorene are composed of s, p_x and p_z orbitals and are even with respect to the xz mirror plane. As a result, excitation from VCM to CBM (or emission from CBM to VCM) is only allowed through an interaction that is also even under this symmetry. While an electric filed along x is even to $y \to -y$ operation, electric field along y is odd; therefore the latter is strictly forbidden at **Γ** point, giving the well-defined polarization selection rule in monolayer phosphorene.

In few layer phosphorene, states from each layer will mix in a way described by Eq. 3-2. However, the mix mainly changes the distribution of wavefunction along z direction, leaving the in-plane part largely unaffected. Therefore again from the simple 1D model, we expect the in-plane polarization selection rules for all subband transitions to be the same as in monolayer, which is exactly what we observe experimentally.

.

3.4 Discussion and Outlook

The study in this chapter is straightforward in concept, however already presents many useful and important results: The combination of physical theoretical model and reliable experimental technique provide an accurate determination of layer-dependent direct bandgap as well as higher subbands transitions in few-layer phosphorene, which serves as a fundamental guidance to its application in electronics and optoelectronics. The tunability of electronic states, high carrier mobility and two-dimensional nature, together make few layer phosphorene a desirable platform for novel devices. In addition, the perfect match between results from the simple model and reflectance spectrum unambiguously identifies the strong interlayer interaction as the origin to the systematic evolution of electronic structure in few-layer phosphorene; and also allows us to quantitatively determine the interlayer coupling strength.

On theoretical aspect, this study is an example application of our discussion in chapter 2, though the "simplest of simplest" type. However, we can already see the power of the formalism introduced in chapter 2, that we can understand the physical picture behind the interlayer interaction-induced electronic structure change in a simple and elegant way. e.g. with only one parameter in the present case. This parameter, the interlayer coupling strength (or matrix element amplitude), is in the order of 1eV, which is indeed quite strong. Such strong coupling can also be intuitively understood as a result from direct mixing at **Γ** point, as has been discussed at the beginning of this section (which is not the only factor determining the coupling matrix element though, see section 2.4.3).

We note that, although the 1D model does a good job in explaining all the experimental observation, it also has limitations. First, it is still a single-particle model which does not take into account many-body effect. However, one can justify its use from the fact that the exciton binding energy and quasi-particle bandgap renormalization (i.e. Coulomb interaction of electron-electron and electron-hole) will large cancel each other. One extreme case is that, if Coulomb interaction is infinitely long range with no length dependence at all, then the electron-electron and electron-hole interaction will completely cancel as long as the total charge is conserved. As a result, the optical bandgap can often be represented reasonably well by single-particle bandgap, which is demonstrated in the bandgap calculation of many systems using density functional theory (DFT) methods.

Another deficiency of the model is that we only consider one valence and conduction band from each layer. As layer number increases, Eq. 3-2 tells us that the energy span of all the subbands will significantly increase, which can also be clearly seen in Figure 3.6. Therefore, when the layer number is too large, high energy subbands in Eq. 3-2 will inevitably overlap with the low energy subbands from other bands in monolayer that are not considered in the model. In such case, their interaction will be important; and the 1D model can become significantly inaccurate.

Thus the 1D model will ultimately fail if both n and N are large. Note that if only N is large but n is small then the model will still work.

In principle, however, we can always improve the model by considering more bands at **Γ** point in monolayer, which can be mapped into multiple 1D atom chains interacting with each other following the same concept as in section 3.2. In this way, the model can in principle work in an arbitrarily larger energy range, as long as all important orbitals are considered.

Chapter 4 – Strongly Coupled Electronic States in One-dimensional van der Waals Systems

4.1 Introduction and Background

In the last section, we have seen one example of how interlayer interaction can act as a convenient and powerful tool to engineer the properties of 2D materials. The explanation of interlayer coupling effects is particularly simple because of the nature of **Γ** point interaction. Given the rich and useful effects of interlayer interaction in 2D van der Waals systems, it is naturally to ask: can we extend to other dimension?

To answer this question, we can look back to see why the concept of material engineering through interlayer van der Waals coupling is only intensively studied recently, accompanying the emergence of 2D materials. If we put two bulk graphite together, it is hard to imagine that their bulk properties will change significantly, since most of the atoms are deeply buried below the interface and can hardly see the other material unless there are some macroscopic field produced. As a result, combining two graphite will lead to nothing but bigger graphite, which is quite trivial. In two-dimension, however, the "bulk" of the material is really just the surface. As a result, all the electrons are exposed at the interface and are close to the other layer; which can result in strong modification of their behaviors even with non-covalent, van der Waals interaction.

Another factor further facilitating the strength of interlayer interaction is the reduction of dielectric screening in low dimension. It is well known that a perfect metal sphere can completely screen external electric field from its interior spaces, because the free carriers inside metal will response to the external electric field and rearrange to cancel it upon equilibrium. In actual case the screening effect will not be always perfect. However as long as particles, such as electrons, phonons, etc. in a system are responding to external potential or interaction, their redistribution will generate additional field that partially cancels the original one, as described by their dielectric function or susceptibility. In 3D system, such screening effect can often reduce the potential/interaction to be orders magnitude smaller than the original one (or bare one). However, in 2D system the material itself is limited to a plane; therefore the screening effect coming from all particles inside the material is also much less significant. Intuitively, one can think of the penetration depth of metal; and the field will be finite everywhere inside the metal if its thickness is smaller than the penetration depth.

The reduced dielectric screening effectively increases the amplitude of any potential or interaction. One representative consequence of such strong interaction is the orders of magnitude stronger binding energy for many-body sates like excitons, trions, etc., in atomically-thin semiconductors than in the case of quantum wells (*49-51*). Such strong many-body interaction enables various new perspectives of 2D materials that I will not discuss in detail here. The point

is, owing to the same reason, interlayer interaction will also be particularly strong in low dimension, which, combined with its large parameter space, creates a whole field of rich new possibilities.

With the above discussion, now we can answer the initial question: It is difficult to extend to 3D; however 1D may have similar effect. The "similarity" comes from the fact that in 1D the electrons are also exposed to the interface; and the dielectric screening is largely reduce, actually even more reduced than 2D case, from the outstandingly large exciton binding energy of carbon nanotubes (*52*, *53*). On the other hand, 1D is never identical to 2D, since almost all the important theoretical models will give different results in different dimensions, in many cases even qualitatively different. In this way, we would expect that the interlayer interaction in 1D van der Waals systems should, like in 2D, gives rise to strong and rich modification to system properties; however in a way that might be fundamentally different from 2D case. In this chapter, I will focus on these new exciting possibilities in 1D. Part of the discussion is also presented in a published paper (*2*).

4.2 New Rich Possibilities from Interlayer Interaction in One-Dimensional Double Walled Carbon Nanotubes (DWNTs)

4.2.1 Carbon Nanotubes: Ideal System to Study One-Dimensional Physics

In last section, we have seen the potential new rich possibilities in one-dimensional van der Waals systems from interlayer interaction. The first step to realize all these possibilities is to find suitable materials as platform and playground. In 2D, the "suitable materials" compose a long list including graphene, transition metal dichalcogenide, phosphorene, boron nitride and many more. It is therefore natural to look into their 1D counterpart, among which carbon nanotube stands out as an almost ideal system to study.

Carbon nanotubes have been intensively studied as a model 1D system owing to its attractive properties on both fundamental and application aspects. Studies on quantum conductance and intriguing Luttinger liquid physics has always chosen nanotube as the subject. On the other hand, field effect transistor based on carbon nanotubes has also been a central topic in new-generation circuits owing to its great potential in scalability and reducing the power consumption.

More fundamentally, what benefits carbon nanotube on both aspects is its well-defined but richly-varied structures: A single-walled carbon nanotubes (SWNTs) has an atomic structure uniquely defined by its **chiral indices**, i.e. two integer numbers (n, m). What is impressive, a SWNT can have near-perfect lattice ordering without single defect over a large length, since its structure dos not have any edge except for at the two ends. In addition, its electronics properties sensitively depend on the chirality (or chiral indices), from metallic to semiconducting with different bandgap size. Given the infinite combination of integer pairs (n, m), SWNT will also have infinite number of unique structures.

These advantages of carbon nanotube make it very preferable for the study of interlayer interaction. Similar to the 2D case, the necessary preparation before isolating and understanding interlayer interaction is to understand individual layer first. Therefore, here I will first give a brief introduction to SWNTs.

4.2.2 Real- and Momentum-Space Structure of Single-Walled Nanotubes

To understand the electronic structure of SWNT, a commonly adopted method is from the perspective of graphene, as nanotube can be considered a rolled-up version of graphene. If we neglect the curvature effect, which is reasonable if the radius is much larger than the carbon-carbon bond length, then nanotube's electronics properties will have direct relation to graphene's.

Certainly, the electronic properties should also depend on how we "roll-up" the graphene sheet, which is given by the chiral indices (n, m): Assuming the two primitive lattice vector of graphene to be $\boldsymbol{a}_1, \boldsymbol{a}_2$, we can choose an arbitrary lattice vector $\boldsymbol{C}_h = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$ as the chiral vector of the nanotube, and roll the graphene up by connecting both ends of $\boldsymbol{C}_h$ to form a circle. Figure 4.1A illustrates one example with $\boldsymbol{C}_h = 4\boldsymbol{a}_1 + 2\boldsymbol{a}_2$, corresponding to a SWNT with chiral indices $(n, m) = (4,2)$. Depending on the ratio between n and m, there will be an angle α between $\boldsymbol{C}_h$ and $\boldsymbol{a}_1$, given by $\alpha = \tan^{-1}\frac{\sqrt{3}m}{2n+m}$. α is often called "**chiral angle**" since it represents the orientation of the tube.

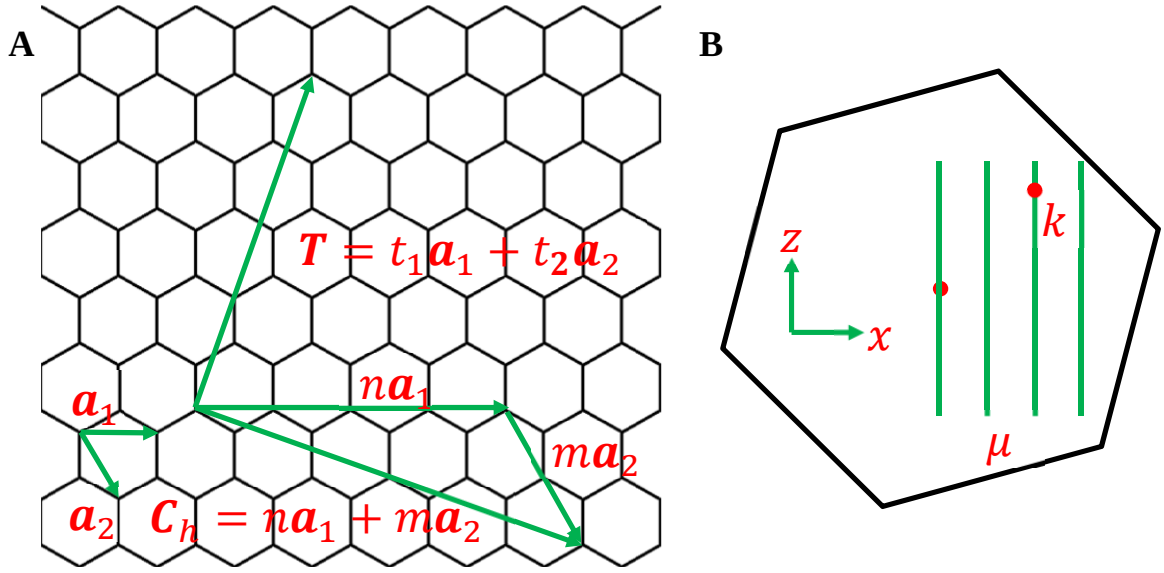

Figure 4.1 (**A**) Relation between nanotube and graphene through the chiral vector $\boldsymbol{C}_h$. A nanotube with chiral indices (n, m) corresponds to a chiral vector $\boldsymbol{C}_h = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$, and is formed by rolling up the chiral vector into a circle. In the above example we have $(n, m) = (4,2)$. The 1D lattice vector $\boldsymbol{T}$ is always perpendicular to $\boldsymbol{C}_h$. (**B**) Electronic states of SWNT in 2D Brillouin Zone of graphene. The boundary condition along the periphery of tube leads to quantized cutting lines on which eigen-states reside, labeled by discrete integer μ. On the other hand, along the tube the periodic distribution of atoms gives a continuous momentum k.

Because the chiral vector $\boldsymbol{C}_h$ forms the circumference of the cylinder, the tube axis should always be perpendicular to $\boldsymbol{C}_h$. We can therefore determining the 1D periodicity of a given nanotube along the axis by searching for the shortest lattice vector $\boldsymbol{T}$ of graphene that is perpendicular to $\boldsymbol{C}_h$. One can show that a general form of periodicity $\boldsymbol{T}$ can be expressed as

$$\boldsymbol{T} = (t_1\boldsymbol{a}_1 + t_2\boldsymbol{a}_2) = \frac{(n+2m)}{d_R}\boldsymbol{a}_1 - \frac{(2n+m)}{d_R}\boldsymbol{a}_2 \quad (4-1)$$

Where $d_R = \gcd(n+2m, 2n+m)$ is the greatest common divisor of $(n+2m)$ and $(2n+m)$.

The close relation between nanotube and graphene in real space necessarily indicates their correspondence in momentum space, i.e. band dispersion. To see this, one can unfold the nanotube into a piece of graphene with finite width $\boldsymbol{C}_h$ (which is defined as x direction) and infinite length on the perpendicular direction (z direction). In real nanotube, the two ends of $\boldsymbol{C}_h$ are connected, so does any eigen-state of the system. Therefore, to make the "unfolded" graphene piece equivalent to nanotube, we need require its eigen-state to satisfy a boundary condition:

$$\psi(\boldsymbol{r}) = \psi(\boldsymbol{r} + \boldsymbol{C}_h) \quad (4-2)$$

As a result, compare to a freestanding graphene sheet, the eigen-states in unfolded nanotube cannot freely choose a momentum in the 2D Brillouin Zone of graphene. Instead, they can only reside on discrete cutting lines separated by $\Delta\boldsymbol{k}_\mu = 2\pi/C_h$. These cutting lines are labeled by an integer μ, often called the **cutting line number**.

On the other hand, along axial direction the unfolded tube is infinitely long with a periodicity of T. This indicates that eigen-states can sit at wherever they like on the cutting line, described by a continuous momentum k along the z direction; however only those within $[-\frac{\pi}{T}, \frac{\pi}{T})$ are independent. In this way, any eigen-state of given nanotube takes the form:

$$\psi^{\sigma,\mu,k}(x,z) = \psi^{\sigma,\boldsymbol{k}_{2D}}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_j} \varphi^{\sigma,\boldsymbol{k}_{2D}}\left(\boldsymbol{r} - \boldsymbol{R}_j\right) e^{i\boldsymbol{k}_{2D}\cdot\boldsymbol{R}_j}, \boldsymbol{k}_{2D} = \mu\frac{2\pi}{C_h}\hat{\boldsymbol{x}} + k\hat{\boldsymbol{z}}, k \in \left[-\frac{\pi}{T}, \frac{\pi}{T}\right)$$

$$E^{\sigma,\mu,k} = E^{\sigma,\boldsymbol{k}_{2D}} \approx \sigma v_F k_{2D} \quad (4-3)$$

Here $\psi^{\boldsymbol{k}_{2D}}(\boldsymbol{r})$ is simply the wavefunction of graphene at 2D momentum $\boldsymbol{k}_{2D}$ (Note that the graphene is in the xz plane due to our definition), and $\boldsymbol{R}_j$ is summing over unit cells of graphene; $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}}$ are unit vectors along x and z direction. In calculating the energy we have assumed a linear band dispersion, which is only valid when k_{2D} is reasonably small. $\sigma = +1/-1$ stands for conduction/valence band states in graphene, respectively.

With the wavefunction obtained in the reference frame of unfolded nanotube, we can now roll it back, and see how the wavefunction looks like in reality. Naturally, we will adopt cylindrical coordinate (r, θ, z), corresponding to radial distance, azimuth angle, and axial height, respectively. For nanotube rolled-up from an atomically thin graphene sheet, we can neglect its thickness and assume that $r \equiv C_h/2\pi$. Then the coordinate (θ, z) in cylindrical frame will have one-to-one correspondence to coordinate (x, z) in Eq. 4-3 through $\theta = \frac{x}{r}, z = z$. We can therefore directly write down the wavefunction in cylindrical coordinate as:

$$\psi^{\sigma,\mu,k}(\theta,z)=\psi^{\sigma,\mu,k}\left(\frac{x}{C_h},z\right)=\sum_{\boldsymbol{R}_j}\varphi^{\sigma,\boldsymbol{k}_{2D}}\left(\boldsymbol{r}-\boldsymbol{R}_j\right)e^{i(\mu\theta_j+kz_j)} \quad (4-4)$$

where (θ_j, z_j) is the cylindrical coordinate of graphene unit cell at $\boldsymbol{R}_j$. From Eq. 4-4, we can clearly see the physical meaning of both (μ, k): μ is an angular momentum and describes the angular distribution of wavefunction, which has to be quantized because of the 2π periodicity requirement of any angular distribution. On the contrary, k is a one-dimensional momentum coming from nanotube's translational symmetry along axial direction, and therefore can be continuous.

4.2.3 Existence of Interlayer Electronic Coupling in 1D

With the knowledge of individual "layer", here SWNT, now we are finally at a point to try to investigate interlayer, or intertube interaction in DWNTs. A DWNT is formed by two concentric SWNTs, just like rolling up a piece of bilayer graphene. Given the intimate relation between SWNT and single layer graphene, one would naturally expect DWNT to resemble the behavior of bilayer graphene, which we have intensively discussed in Chapter 2. If so, supposedly a variety of new physics should show up depending on specific regime of interlayer configuration, such as commensurate and incommensurate, strong and weak coupling etc..

However, as we mentioned in the beginning, there can be intrinsic difference between 1D and 2D systems. One example is that, the parameter space of interlayer interaction in 1D is even larger. While in 2D bilayer graphene the only relevant parameter is the interlayer twist angle; in 1D this is apparently not the case: even if the difference between their chiral angle, $(\alpha_1-\alpha_2)$, is fixed; both inner- and outer-tubes can still vary significantly, e.g. from metallic to semiconducting tubes, depending on α_1 and α_2 individually. Therefore it is necessary to at least consider both chiral angles, instead of only their difference. Furthermore, even if both chiral angle are fixed, the diameter will still play critical role in determining the properties of each tube, as well as their intertube interaction, which adds additional knobs to tune.

Such huge parameter space should create a large playground for potentially new physics. Several previous studies have already attempted to investigate the consequence from intertube interaction. However, the results suggest that there is virtually no intertube electronic coupling in incommensurate DWNTs (*54*, *55*). Here the definition of "incommensurate" DWNT is quite similar to incommensurate bilayer in 2D: The 1D periodicity of a SWNT, $T=\sqrt{3(n^2+m^2+nm)}/d_R$, is an irrational number; and therefore T_1 and T_2 of inner- and outer-tubes are in general strictly incommensurate unless $\frac{n_1}{m_1}=\frac{n_2}{m_2}$, or equivalently $\alpha_1=\alpha_2$. i.e. zero twist angle. While in 2D the zero twist angle case, such as AB stacked bilayer graphene is naturally available. In 1D it is extremely rare because the probability is too low for two SWNTs

to have the same chiral angle and proper radius difference simultaneously; and almost all experimentally observed DWNTs are incommensurate ones (*56-65*). As a result, theoretical predictions seem to suggest the non-existence of intertube electronic coupling in almost all realistic DWNTs.

Such conclusion, though disappointing, seems to have some support both from theoretical and experimental aspects. Theoretically, the negligible intertube coupling was explained as a consequence of the random local intertube configuration from incommensurability, which gives a random phase when summing up contribution from different atom sites and ultimately cancel with each other. Experimentally, transport measurement on multi-walled nanotubes (MWNTs) surprisingly resembled behavior of a single-walled tube (*54*), which was believed to originate from the lack of intertube interaction and therefore tunneling; and only the outermost tube is conducting current.

With our discussion in chapter 2, however, we can immediately see that both arguments are not complete: As we have derived in section 2.4.2, incommensurability indeed leads to cancellation of interlayer coupling in most cases, however not when the two interlayer electronic states satisfy specific momentum selection rules (see Eq. 2-14). Similarly, we also already know that multi-layer graphene will show an effective interlayer decoupling and single-layer behavior near Dirac point at large twist angle (section 2.5.6), which is apparently very similar to the case of multiple-walled nanotubes. Therefore, effective monolayer (or single-walled) behavior at specific energy does not necessarily suggest complete non-existence of interlayer interaction. For example, specific states in bilayer graphene away from the Dirac point can always be strongly modified, even at large twist angle. Thus we would expect interlayer electronic coupling to persist in incommensurate DWNTs.

What is different in 1D DWNTs, however, is that the above literature also investigated situation of small intertube twist angle; and it seems that states near Dirac point are still largely unaffected by interlayer interaction. This poses a sharp contrast to 2D bilayer graphene case, where states near Dirac point should be strongly modified at small twist angle, even forming flat bands and zero energy modes if entering the strong coupling regime.

Such difference indicates that 1D and 2D cases are not completely identical, which is hardly surprising at all. Below I will follow similar approaches as in chapter 2, section 2.4.2, to obtain the interlayer (or intertube) coupling selection rules and matrix element in 1D DWNTs. Afterwards, we can have a direct comparison between 1D and 2D cases, which will help us understand the similar and different aspects in their behavior.

4.2.4 Interlayer Electronic Coupling in Incommensurate DWNTs

Selection Rule and Matrix Element

Similar to the derivation in section 2.4.2, to consider the intertube electronic coupling between two states $\psi_1^{\mu k\sigma}$ and $\psi_2^{\mu' k'\sigma'}$, we can start with expanding them into local orbitals of each carbon atom:

$$\psi_1^{\mu k\sigma} = \frac{1}{\sqrt{2N}} \sum_{\boldsymbol{R}_1} e^{i(\mu\theta_1 + kz_1)} \left[\varphi(\boldsymbol{r} - \boldsymbol{R}_1) + \sigma e^{i\phi_1}(\boldsymbol{r} - \boldsymbol{R}_1 - \boldsymbol{\tau})\right] = \psi_A^{\mu k} + \sigma\psi_B^{\mu k} \quad (4-5)$$

$$\psi_2^{\mu' k'\sigma'} = \frac{1}{\sqrt{2N}} \sum_{\boldsymbol{R}_2} e^{i(\mu'\theta_2 + k'z_2)} \left[\varphi(\boldsymbol{r} - \boldsymbol{R}_2) + \sigma' e^{i\phi_2}(\boldsymbol{r} - \boldsymbol{R}_2 - \boldsymbol{\tau}')\right] = \psi_A^{\mu' k'} + \sigma'\psi_B^{\mu' k'}$$

in which we have separated the contribution from A and B sublattices of atoms by splitting the total wavefunction into two parts. $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ sum over all graphene unit cells in the inner- and outer-tube, respectively, with their cylindrical coordinate given by (θ_1, z_1) and (θ_2, z_2). ϕ_1, ϕ_2 represent the phase difference between the two sublattices, which depend on $\boldsymbol{k}_{2D}$. The coupling matrix element between $\psi_1^{\mu k\sigma}$ and $\psi_2^{\mu' k'\sigma'}$ can therefore be expanded as:

$$M_{12} = \left\langle \psi_1^{\mu k\sigma} \middle| H_I \middle| \psi_2^{\mu' k'\sigma'} \right\rangle = M_{12}^{AA} + \sigma' M_{12}^{AB} + \sigma M_{12}^{BA} + \sigma\sigma' M_{12}^{BB}$$

$$M_{12}^{AA} = \frac{1}{2\sqrt{NN'}} \sum_{\boldsymbol{R}_1} e^{i[(\mu'-\mu)\theta_1 + (k'-k)z_1]} \sum_{\Delta\boldsymbol{R}} e^{i(\mu'\Delta\theta + k'\Delta z)}\, t(\Delta\boldsymbol{R})$$

$$M_{12}^{AB} = \frac{1}{2\sqrt{NN'}} e^{i\phi_2} \sum_{\boldsymbol{R}_1} e^{i[(\mu'-\mu)\theta_1 + (k'-k)z_1]} \sum_{\Delta\boldsymbol{R}} e^{i(\mu'\Delta\theta + k'\Delta z)}\, t(\Delta\boldsymbol{R} + \boldsymbol{\tau}')$$

$$M_{12}^{BA} = \frac{1}{2\sqrt{NN'}} e^{-i\phi_1} \sum_{\boldsymbol{R}_1} e^{i[(\mu'-\mu)\theta_1 + (k'-k)z_1]} \sum_{\Delta\boldsymbol{R}} e^{i(\mu'\Delta\theta + k'\Delta z)}\, t(\Delta\boldsymbol{R} - \boldsymbol{\tau})$$

$$M_{12}^{BB} = \frac{1}{2\sqrt{NN'}} e^{i(\phi_2 - \phi_1)} \sum_{\boldsymbol{R}_1} e^{i[(\mu'-\mu)\theta_1 + (k'-k)z_1]} \sum_{\Delta\boldsymbol{R}} e^{i(\mu'\Delta\theta + k'\Delta z)}\, t(\Delta\boldsymbol{R} + \boldsymbol{\tau}' - \boldsymbol{\tau}) \quad (4-6)$$

The expression of above matrix element is exactly the 1D counterpart of Eq. 2-11. Again we can take advantage of the random interlayer configuration and replace the second summation over $\Delta\boldsymbol{R}$ with an integration:

$$\sum_{\Delta\boldsymbol{R}} e^{i(\mu'\Delta\theta + k'\Delta z)}\, t(\Delta\theta, \Delta z) \to \frac{1}{S} \iint r_2 d\theta dz e^{i(\mu'\theta + k'z)} t(\theta, z) \quad (4-7)$$

Here the integration goes over the whole outer-tube, and S is the graphene unit cell area. Eq. 4-7 is simply the Fourier component of the pair interaction in the cylindrical coordinate. With such

simplification, we can further obtain the selection rule and amplitude of matrix element, similar to the process from Eq. 2-13 to Eq. 2-17. Using an pair interaction in exponential decay form

$$t(\Delta\theta, \Delta z) \approx \gamma e^{-\sqrt{\Delta r^2 + (\Delta r \Delta\theta)^2 + \Delta z^2}/\lambda} \quad (4-8)$$

In which λ is the characteristic decay length of the pair interaction, and $\Delta r \equiv (r_2 - r_1)$ is the radius difference between the two tubes. i.e. intertube distance. We then have:

$$M_{12}^{AA} = \delta_{\boldsymbol{q}+\boldsymbol{G}_1,\boldsymbol{q}'} \frac{2\pi\lambda\Delta r}{S} \gamma e^{-\frac{\Delta r}{\lambda}} e^{-\frac{\lambda\Delta r}{2}(k'^2 + \frac{\mu'^2}{r_1 r_2})} \quad (4-9)$$

where $\boldsymbol{q} = \boldsymbol{k}_{2D} = \frac{\mu}{r_1}\hat{\boldsymbol{x}} + k\hat{\boldsymbol{z}}$, $\boldsymbol{q}' = \frac{\mu'}{r_2}\frac{r_2}{r_1}\hat{\boldsymbol{x}} + k'\hat{\boldsymbol{z}} = \mathrm{P}\boldsymbol{k}'_{2D}$. P is defined as an operatior on vector that stretches its x component by $\eta = \frac{r_2}{r_1}$, while keeps its z component unchanged. With such definition, we can write the selection rule as:

$$\boldsymbol{k}_{2D} + \boldsymbol{G}_1 = \mathrm{P}(\boldsymbol{k}'_{2D} + \boldsymbol{G}_2) \quad (4-10)$$

Eq. 4-10 looks very similar to Eq. 2-14, the selection rule in 2D system. However they are definitely not identical, from the additional operatior P. Such discrepancy originates from the intrinsic difference between bilayer graphene in 2D and DWNTs in 1D: Although the latter is indeed a rolled-up of the former, it is rolled up from two pieces of graphene with different width, as can be seen from the different circumference of inner- and outer-tubes. During the rolling up process, we make the two pieces of graphene with different width into the same angle 2π, therefore introducing an intrinsic strain in the outer tube in the real space (by taking the inner tube as reference). In other words, to construct a bilayer graphene with the same interlayer atom configuration as in a DWNT, we need compress the second layer (corresponding to the outer-tube) by $\eta = r_2/r_1$ to make its width the same as the first layer. Such intrinsic compress in real space also leads to a stretch in momentum space described by the operation P, as shown in Figure 4.2. Pink and grey hexagons represent Brillouin Zone of 1st and 2nd layer graphene, they are twisted against each other due to the different chiral angle of the inner- and outer-tubes. In addition, the Brillouin Zone of the 2nd layer is stretched by η along the x direction. Because this "twisted and stretched" 2D bilayer graphene is completely equivalent to 1D DWNT, we can directly apply the 2D selection rule Eq. 2-14, which states that the momentum need be conserved. Immediately, we will reach the same conclusion as Eq. 4-10.

Another way to see the difference between 2D and 1D is directly from the perspective of 1D. Similar to 2D case where the 2D momentum need be conserved; in 1D case we would naturally expect both the angular momentum μ and 1D linear momentum k to be conserved. While it is not so apparent how lattices of both tubes can provide additional momentum, at least two interlayer states should satisfy the selection rule if $\mu = \mu'$ and $k = k'$. According to Eq. 4-3, however, they will have different $\boldsymbol{k}_{2D}$ in the 2D Brillouin Zone of graphene. Such difference comes from the different cutting line spacing between inner- and outer-tubes; and ultimately originates from the intrinsic strain involved.

Practically, it would be convenient to adopt the picture in Figure 4.2 to determine states satisfying the selection rule. Since this picture considers a 2D bilayer graphene system, the process is essentially the same as in chapter 2: For a given electronic state at $\boldsymbol{k}_{2D}$ in layer 1, we can draw a reciprocal lattice ($\boldsymbol{k}_{2D} + \boldsymbol{G}_1$), illustrated as an array of dots in Figure 4.2. All states from layer 2 that satisfy the selection rule will be on this lattice. Note that, although all states on the lattice correspond to the same momentum in layer 1, they represent different states in layer 2 since is Brillouin Zone is not only twisted but also intrinsically stretched compared to layer 1.

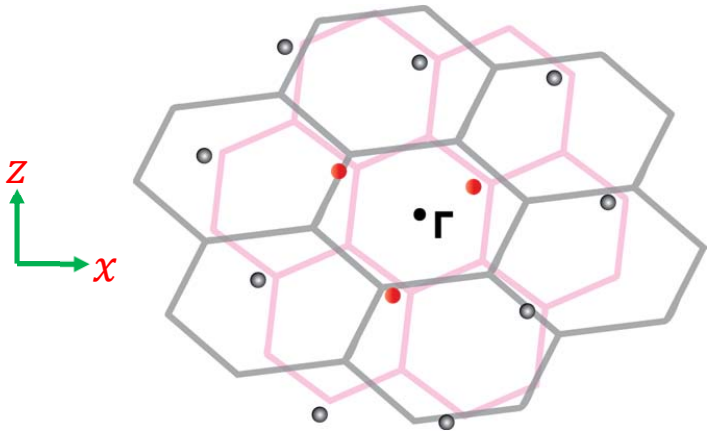

Figure 4.2 Interlayer coupling selection rule in 1D DWNT. The 1D DWNT system is equivalent to a 2D system of bilayer graphene, with the 2nd layer intrinsically compressed by $\eta = r_2/r_1$ along x. As a result, the Brillouin Zone of 2nd layer graphene (grey hexagons) is stretched by η along x direction compared to 1st layer (pink hexagons). In such picture, the selection rule will be the same as 2D case: We can draw the reciprocal lattice of 1st layer centering at the given momentum, states from the 2nd layer satisfy the selection rule only when they are on this lattice. The three red dots on the lattice labels states closest to the center of Brillouin Zone, therefore having the strongest coupling.

Modification of Optical Transition Energy at Van Hove Singularities

Following similar procedure as in Eq. 4-7 to 4-9, we can obtain $M_{12}^{AA}, M_{12}^{AB}, M_{12}^{BA}, M_{12}^{BB}$, respectively, which gives M_{12}:

$$M_{12} = C_{12} M_{12}^{AA}$$

$$C_{12} = \left(1 + \sigma e^{-i\beta_1}\right)\left(1 + \sigma e^{i\beta_2}\right) M_{12}^{AA}, \qquad \beta_1 = \phi_1 - \boldsymbol{q} \cdot \boldsymbol{\tau},\ \beta_2 = \phi_2 - \boldsymbol{q}' \cdot \boldsymbol{\tau}' \quad (4-11)$$

Similar to Eq. 2-17, the exponential decay in Eq. 4-9 indicates a fast decrease of coupling amplitude with increase of q'; and therefore only those close to center of Brillouin Zone will have strong coupling, shown as the three red dots in Figure 4.2.

With both selection rule and amplitude of matrix element obtained, we can start investigating modification of electronic structure in DWNT from interlayer interaction. One unique consequence of 1D system is the emergence of Van Hove singularities in the density of states at band extrema. The extremely large density of states near Van Hove singularities can easily dominate system behavior, such as giving prominent optical transitions between pairs of van Hove singularities. It is therefore natural to look at how these Van Hove singularities and induced optical transitions are modified by interlayer electronic interaction.

Assuming we are in the simplest, weak perturbative regime. Since both inner- and outer-tubes have electronic states in the same energy range, most likely the dominant effect will be from direct mixing. The induced energy change of an arbitrary state (μ, k, σ) from layer 1 is approximately:

$$\delta E_1 = \sum_{j=1}^{3} \frac{\left|M_{12}^{j}\right|^2}{E_1 - E_2^{j}} \quad (4-12)$$

In which we only consider the coupling to states from layer 2 at three important momentum closest to **Γ** point, and neglect the coupling between conduction band and valence band states because they will have large energy difference.

It is then straightforward to calculate the energy shift for a given optical transition from $(\mu, k, -)$ to $(\mu, k, +)$. We can choose (μ, k) to be at the i^{th} Van Hove singularity of layer 1 starting from the valence band maximum/lowest conduction minimum. This type of optical transition is often called S_{ii} or M_{ii} transition depending on whether the tube is semiconducting or metallic. Its energy change, ΔE_{ii} is:

$$\Delta E_{ii} = \Delta E^{\mu k+} - \Delta E^{\mu k-} = \sum_{j=1}^{3} \frac{\left|M_{12}^{\mu k+,\mu_j' k_j'+}\right|^2}{E_1^{\mu k+} - E_2^{\mu_j' k_j'+}} - \sum_{j=1}^{3} \frac{\left|M_{12}^{\mu k-,\mu_j' k_j'-}\right|^2}{E_1^{\mu k-} - E_2^{\mu_j' k_j'-}}$$

$$\approx \sum_{j=1}^{3} \frac{\left|M_{12}^{\mu k+,\mu_j' k_j'+}\right|^2 + \left|M_{12}^{\mu k-,\mu_j' k_j'-}\right|^2}{E_1^{\mu k+} - E_2^{\mu_j' k_j'+}} = \sum_{j=1}^{3} \frac{\left|A_j\right|^2}{\Delta E_j} \quad (4-13)$$

where $\left|A_j\right|^2 = \left|M_{12}^{\mu k+,\mu_j' k_j'+}\right|^2 + \left|M_{12}^{\mu k-,\mu_j' k_j'-}\right|^2 = (\left|C_{12,j}^{+}\right|^2 + \left|C_{12,j}^{-}\right|^2)\left|M_{12,j}^{AA}\right|^2$ is the effective coupling matrix element combining the contribution from both conduction band and valence band energy shift; $\Delta E_j = E_1^{\mu k+} - E_2^{\mu_j' k_j'+}$. We have used $(E_1^{\mu k+} - E_2^{\mu_j' k_j'+}) \approx -(E_1^{\mu k-} - E_2^{\mu_j' k_j'-})$ due to the approximate electron-hole symmetry for states close to **K** and **K**' points in 2D Brillouin Zone of graphene.

Rich Parameter Space of Interlayer Interaction

Eq. 4-13 indicates that the energy shift will generally depends on the amplitude of matrix element A_j, as well as the energy difference ΔE_j. Similar to the case in 2D, if we confine our discussion to states within a reasonably small range in momentum space (here it means optical transitions of reasonably low energy), then $M_{12,j}^{AA}$ in Eq. 4-9 can be considered approximately as constant. Though there is still a factor ($\left|C_{12,j}^{+}\right|^2 + \left|C_{12,j}^{-}\right|^2$) between A_j and $M_{12,j}^{AA}$, one can show that such factor has a large chance to take value around 8. As a result, the variation of A_j can be largely described by a simple scaling law $e^{-\Delta r/\lambda}$. i.e. the variation of intertube spacing. Figure 4.3A shows the maximum A_j ($j = 1-3$) calculated for each experimentally observed optical transitions in many DWNTs. Indeed, A_{max} is roughly following a simple scaling law of $e^{-\Delta r/\lambda}$, as given by the red solid line; and their non-perfect match is understandable as a result of all the approximations we make. In calculating A_j we used parameters $\gamma = 405\gamma_0 \approx 1200\text{eV}$ and the characteristic length $\lambda = 0.045\text{nm}$ (*66, 67*) in Eq. 4-9, where $\gamma_0 \approx 3\text{eV}$ is the (intralayer) nearest-neighbor hopping constant in graphene.

The simple behavior of A_j in Eq. 4-13 suggests that the large parameter space of interlayer interaction in 1D should be mainly coming from the denominator, ΔE_j. This conclusion is similar to the situation in 2D, as we have seen in various examples in chapter 2, that the selection rule instead of the amplitude of matrix element gives all the richly varied effects, such as weak and strong, perturbative and non-perturbative coupling. From our discussion at the beginning of section 4.2.3, we would expect the coupling in 1D DWNT to depend on the chiral angles of both inner- and outer-tubes, as well as tube diameter and specific transition energy. Figure 4.3B shows the calculated maximum $1/\Delta E_j$ ($j = 1-3$) depending on inner-tube chiral angle θ_i and intertube twist angle θ_{twist} for inner-tube transition S_{22}, at fixed tube diameter $d_i = 1.0\text{nm}$ and $d_o = 1.7\text{nm}$. The value of $(1/\Delta E)_{max}$ shows systematic evolution with respect to chiral angles of both tube; as well as distinct family pattern between the $\text{mod}(n-m,3) = 1$ family (left half) and $\text{mod}(n-m,3) = 2$ family (right half). The $\text{mod}(n-m,3) = 0$ family corresponds to metallic nanotube and is not shown here. $(1/\Delta E)_{max}$ varies dramatically over the parameter space, and can be very large in specific region (white color contour). These regions can have particularly strong modification on electronics structure of DWNTs, even entering non-perturbative or strong coupling regime.

We can therefore focus on these "strong-coupling contours" by investigating its evolution in the parameter space. In Fig. 4.3C we examine its dependence on different optical transitions for a fixed nanotube diameter ($d_i = 1\text{nm}$, $d_o = 1.7\text{nm}$). The strong-coupling contour line shrinks with decreasing transition index (or equivalently lower transition energy). It suggests that the probability to have strong inter-tube coupling becomes lower for low energy transitions.

Figure 4.3D shows the diameter dependence of the strong-coupling contour line for the S_{22} transition with $d_i = 0.5, 1.0, 2.0\ nm$. The strong-coupling contour line pattern shrinks with increasing diameter, largely due to the lower energy of S_{22} transition in a larger diameter tube. Besides, we notice that at the large diameter limit, this contour line approaches $\theta_{twist} = 0$, the requirement of strong coupling in 2D. But for small diameter nanotubes, the contour pattern line rotates away from the line of constant twist angle.

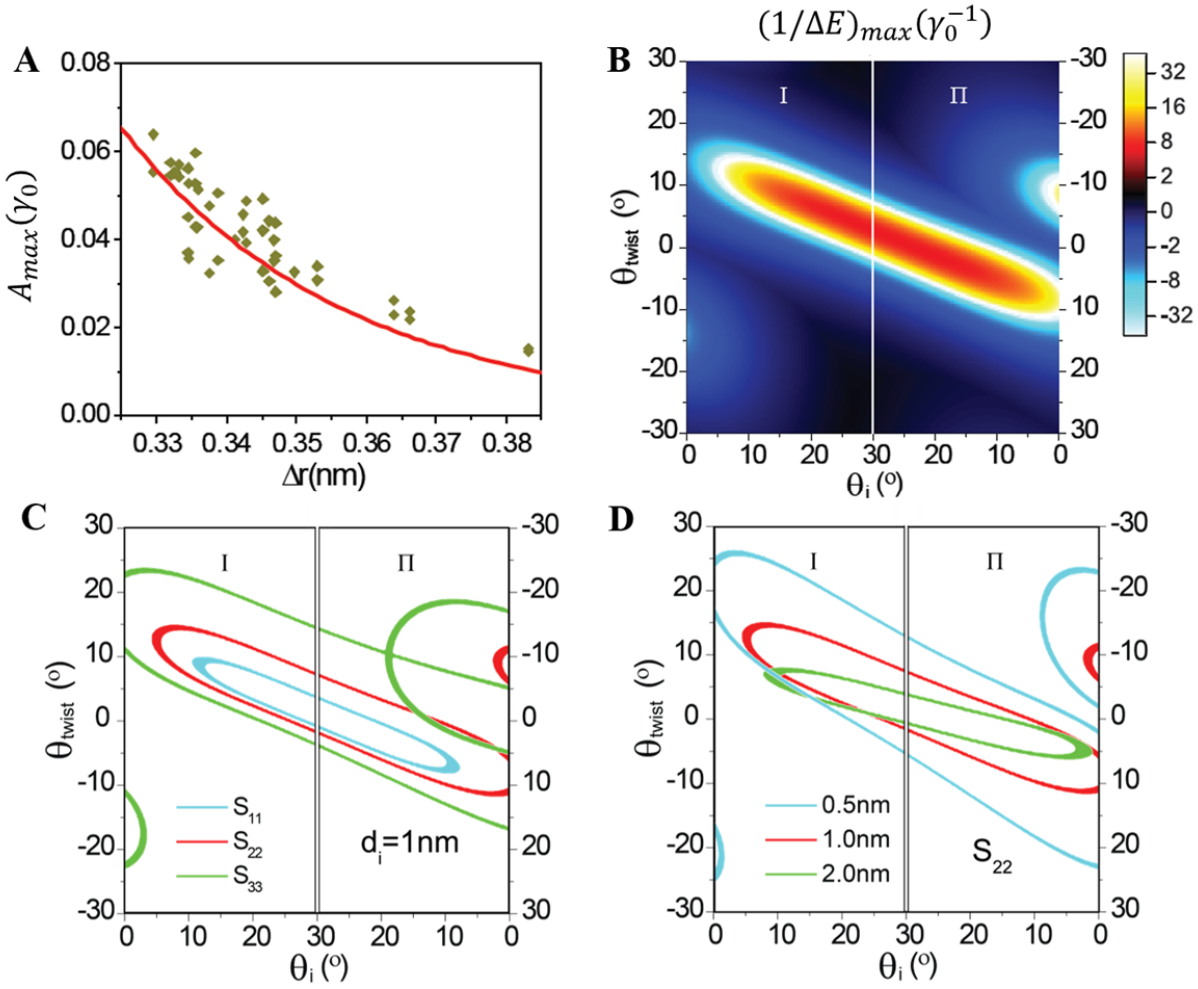

Figure 4.3 (**A**) Calculated maximum matrix element A_{max} for experimentally observed optical transitions, which roughly follows a simple exponential scaling law with intertube distance (red solid line). (**B**) Calculated dependence of $(1/\Delta E)_{max}$ on inner-tube chiral angle θ_i and intertube twist angle θ_{twist} for the inner-tube S_{22} transition with tube diameter $d_i = 1.0$nm and $d_o = 1.7$nm. The value of $(1/\Delta E)_{max}$ shows an interesting evolution with respect to chiral angles of both tube; as well as distinct pattern between the $\mathrm{mod}(n-m,3) = 1$ family and $\mathrm{mod}(n-m,3) = 2$ family. White contour correspond to small ΔE, and strong intertube coupling. (**C**) Transition energy dependence of the strong-coupling contour with $d_i = 1.0$nm. The contour shrinks to the center as transition energy decreases. (**D**) Inner-tube diameter dependence of the strong-coupling contour for S_{22} transition, with intertube spacing is fixed to 0.35nm. The contour rotates away from $\theta_{twist} = 0$ at small diameter, due to the intrinsic strain in 1D.

4.2.5 Distinctive Pattern of Interlayer Interaction in 1D

Now we can go back to answer questions raised at the end of section 4.2.3: Why states near Dirac point cannot have strong interlayer coupling even if the twist angle is small? Figure 4.3C indicates that if the transition energy gets further closer to zero (the Dirac point), then the strong coupling contour will ultimately shrink to a single point at the center. i.e. It will not have strong intertube coupling unless it is an armchair/armchair tube, which is however extremely rare experimentally. The fundamental origin of such distinctive behavior from 2D is the intrinsic "strain" η in 1D DWNTs. Such "strain" leads to a stretch of the Brillion Zone of the outer-tube. As a result, if we focus on an electronic state at Dirac point in the inner-tube; the outer-tube states that satisfy the selection rule will generally be away from the Dirac point depending on the amount of strain, which creates a large energy difference, and therefore preventing strong coupling to happen.

Such intrinsic "strain" also makes the parameter space for interlayer interaction in 1D even larger than 2D. As shown in Figure 4.3D, the larger "strain" in smaller diameter tube will rotate the strong coupling contour further away from the horizontal line of zero twist angle, which from another aspect demonstrate the necessity to consider chiral angles of each tube as well as their diameters; instead of just the twist angle.

4.3 Experimental Observation of Strong Interlayer Electronic Coupling in Incommensurate DWNTs

4.3.1 Experimental Configuration

The new rich possibilities from interlayer interaction in incommensurate DWNTs are still almost a virgin land; and a good first step is always starting with the simplest case: weak perturbative coupling regime. The induced energy shift of optical transitions is already given in Eq. 4-13, which can be directly probed with optical spectroscopy.

Before any comparison between experiment and theory, however, one need first determine the chiral indices (n_1, m_1) and (n_2, m_2) for inner and outer tubes, just like determining the twist angle in twisted bilayer graphene. This requires an independent measurement to provide structural information of constituent nanotubes in a DWNTs, which is achieved here by measuring the electron diffraction pattern of given nanotube (*60, 61*). With the obtained structural information, we can further obtain its optical transitions from single-tube absorption spectroscopy (*68, 69*) to experimentally investigate the energy shift induced by interlayer electronic coupling, as illustrated in Figure 4.4.

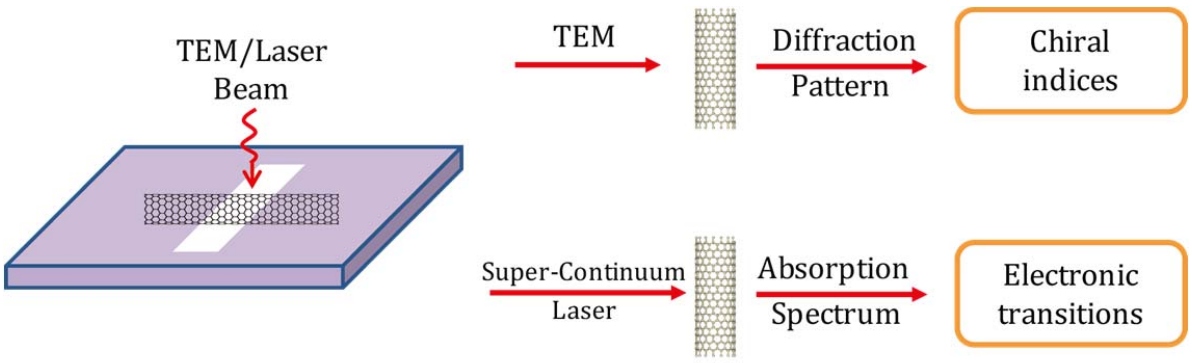

Figure 4.4 Experimental Configuration. Chiral indices and optical transitions of the same individual DWNT is obtained from TEM diffraction pattern and absorption spectrum, respectively.

4.3.2 Experimental Results

Figure 4.5A and 4.5B are the electron diffraction pattern and absorption spectrum of a representative DWNT, respectively. The electron diffraction pattern (4.5A) unambiguously

determines the DWNT chiral indices $(n_o, m_o)/(n_i, m_i)$ to be (22,9)/(11,11), which corresponds to a semiconducting outer-wall nanotube with diameter of 2.16 nm and an armchair metallic inner-wall nanotube with diameter of 1.49 nm. Its absorption spectrum (Fig. 4.5B) shows four prominent optical resonances at 1.58, 1.66, 2.08, and 2.41 eV. In comparison, the isolated (22,9) outer SWNT have three optical transitions S_{33}^o, S_{44}^o, S_{55}^o at 1.66, 2.17, and 2.58 eV, respectively, and the isolated (11,11) inner SWNT has one M_{11}^i transition at 1.77 eV in the experimental spectral range *(70)*. These transition energies of the isolated SWNTs are indicated by vertical dashed lines in Fig. 4.5B. There is a one-to-one correspondence between the DWNT optical resonances and those from the constitutive SWNTs, but the resonance energies are shifted by -80, -90, -170 and -110 meV for the S_{33}^o, S_{44}^o, S_{55}^o and M_{11}^i transitions, respectively.

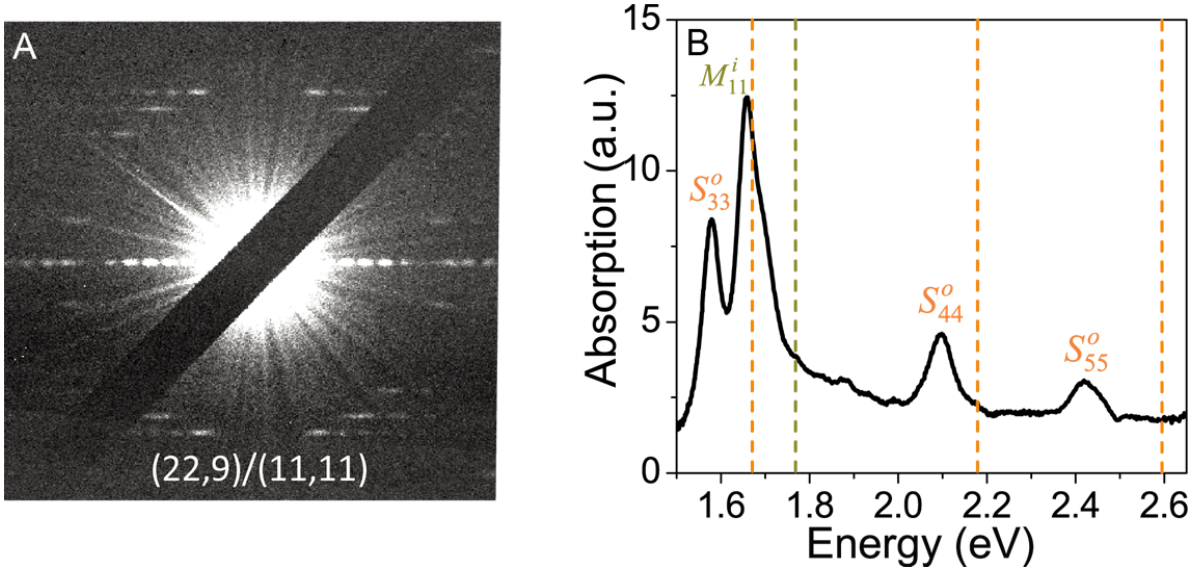

Figure 4.5 (**A**) TEM Diffraction pattern and (**B**) optical absorption spectrum of a representative DWNT. The chiral indices can be unambiguously determine as (22,9)/(11,11) for the outer- and inner- tube, which allows for the prediction of absorption resonances in isolated tubes without intertube interaction, shown as the dashed vertical lines. By comparing experimentally measured DWNT absorption spectrum to that of constituent SWNTs, we find a one-to-one correspondence of transitions, and therefore can assign each optical resonance in DWNT spectrum to a specific transition from either inner- or outer-tube. Interestingly, there is a large energy shift between the same transitions in SWNT and DWNT, which varies between each transition.

We performed the combined electron diffraction and single-tube absorption measurements on 28 individual suspended DWNTs with total 99 optical transitions. In all studied DWNTs, we can map each observed optical transition to that from corresponding isolated constituent SWNT, but the resonance energy is always shifted.

Figure 4.6 summarizes the observed resonance energy shifts for all optical transitions in DWNTs sorted by the shift value. All the DWNTs consist of incommensurate inner and outer walls, but the optical transition energy shift can be quite large, varying from a red shift of 190 meV to a blue shift of 50 meV in different DWNT species. Even for optical transitions from the same DWNT, the energy shifts can vary significantly (See Figure 4.5). Indeed, our observation signifies a very strong effect from inter-tube coupling in DWNTs, which depends sensitively on the chiral indices of the DWNT as well as the exact electronic state of specific transition.

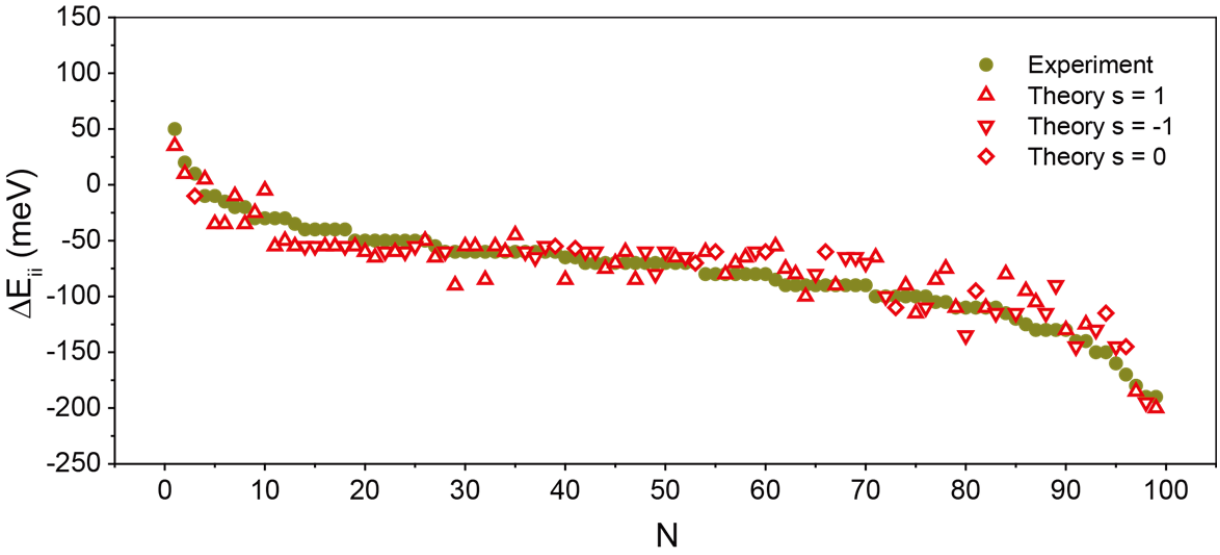

Figure 4.6 Summary of experimentally measured transition energy shift of 99 optical transitions from 28 DWNTs (yellow circles). The average redshift can be explained by the dielectric screening between the inner- and outer-tubes. However the large variation of shift from -200meV redshift to even 50meV blueshift requires the existence of another type of interlayer interaction. i.e. interlayer electronic coupling. Red symbols are theoretical prediction of transition energy shift. $s = +1/-1$ depending on whether the inner- and outer-tube have the same handedness or not.

Before quantitatively comparing the experimental results to theory developed in section 4.2, we should first take a look at other effects that can induce transition energy shift in DWNTs. In previous literature the only well-established interaction between adjacent SWNTs is dielectric screening. It has been shown in previous studies of nanotube in different environments (*71, 72*) that dielectric screening can lead to a redshift in optical transition energies, and the redshift value is similar for all optical transitions. This dielectric screening effect can account for an average redshift of optical transition energies observed in DWNTs (Figure 4.6), but it cannot explain the very large and strongly transition-dependent variations in optical transition shifts, especially the significant blue shifts for certain optical transitions. We can therefore treat it as a constant Δ, and obtain the overall shift of optical transition energy ΔE_{ii}^{tot} from

$$\Delta E_{ii}^{tot} = \Delta E_{ii}^{el} + \Delta \quad (4-14)$$

Red symbols in Figure 4.6 show the theoretical energy shift for the 99 optical transitions with $\Delta \equiv -55\text{meV}$. We note that, the handedness ($n > m$ or $n < m$) does not affect the optical spectra of SWNT; however in DWNT the relative handedness between constituent tubes can affect the intertube electronic coupling strength due to different intertube twist angle (unless the DWNT contains at least one armchair or zigzag tube). Here we introduce an "s" parameter indicating whether the inner and outer tubes have the same ($s = 1$) or opposite ($s = -1$) handedness. The electron diffraction pattern cannot determine the relative handedness (i.e. the s parameter) experimentally. Instead we assign the s parameter based on a better match between the experimental and theoretical transition energy shifts. The up triangle, down triangle and diamond in Figure 4.6 represent transitions in DWNT with $s = 1$, -1 and 0 (contains armchair or zigzag tube), respectively. The agreement between experimental and theoretical results is quite nice given that all parameters in the model in section 4.2 are obtained from literature values, and the only fitting parameter is Δ, the energy shifts from the dielectric screening.

4.4 Summary and Outlook

In this chapter, we have unambiguously revealed the strong intertube electronic coupling in DWNTs from both theoretical and experimental perspectives. Such observation is certainly critical to our understanding of electronic properties of double-walled and multi-walled carbon nanotubes. More generally, however, the study here demonstrate the possibility to strongly modify material properties in 1D incommensurate systems through interlayer electronic coupling, which opens up new exciting opportunities to engineer 1D materials for electronic and optical application.

From the theoretical aspect, the framework we adopt in this chapter is essentially an extension of chapter 2, which further demonstrates the universality of methodologies developed in chapter 2, that it should be applicable to general incommensurate systems regardless of dimension.

Compared the the 2D case, 1D incommensurate system will have an even larger parameter space from interlayer configuration and even stronger interlayer coupling from reduced screening. Such huge playground should be able to give rise to rich and fascinating physics. First of all, most of the discussions in chapter 2 should be able to find their counterpart in 1D, such as weak and strong coupling, perturbative and non-perturbative coupling, direct mixing and effective potential etc. The present study is only focusing on the simplest case, i.e. direct mixing in weak perturbative coupling regime. While being an important first step, it is only a starting point followed by a long road filled with excitement.

Furthermore, we would expect many more possibilities once starting considering the difference between 1D and 2D. The greater quantum confinement in 1D can lead to distinctive behavior on almost every aspect. The larger parameter space in 1D is only one consequence; emergence of van Hove singularities and Peierls instability are among other famous examples, both coming from the peculiar Fermi surface structure in 1D. Things can become even more interesting when we include Coulomb interaction: the Coulomb force can have distinctive effect in 1D, such as the Luttinger liquid behavior. It would be certainly interesting to see the interplay between different interactions, especially when one of them, the interlayer interaction, is highly tunable.

Chapter 5 – Pseudospin Mixing Potential in Graphene/hBN Heterostructure

5.1 Introduction and Background

From Direct Mixing to Effective Potential

In chapter 3 and 4, we have seen two representative examples of how direct mixing in low-dimensional materials can strongly modify their properties; ranging from the simplest case of **Γ** point coupling in well-aligned phosphorene layers; to much more complicated one-dimensional "bilayers" that are not only twisted but also intrinsically strained.

There are two main reasons why we have only considered the effect of direct mixing: First, we have been only looking on the first type of interesting states, i.e. those at band extrema. Because the direct mixing process is of the lowest order perturbation, it will generally generate the largest energy shift, which is our main focus in previous two chapters. In addition, the involved layers are from the same material, and therefore they all have electronic states in the energy range of interest, enabling strong interaction and also formation of interesting pattern depending on interlayer configuration.

On the other hand, as we have discussed in chapter 2, section 2.5.3, if we are interested in the second type of phenomena, i.e. dramatic change of spectral response instead of a simple shift; then in some systems the 2nd order term in Eq. 2-22, the effective potential, may play the dominant role. A representative example is a hetero-bilayer where only one layer has electronic states in the energy range of interest. Graphene/hBN is one such system, as the Dirac point of graphene is deeply sitting within the large bandgap of hBN. Therefore, we would expect the effective potential to take a dominant role here when considering states with degenerate coupling. i.e. in the weak non-perturbative regime.

Exciting Possibilities from the Effective Poetential

Following discussion in chapter 2, section 2.7, one unique and exciting aspect of the effective potential is its capability of providing some components that can be otherwise extremely difficult to generate. What is better, the spectral change of a system is usually coming from electronic states that are near-degenerate to start with, whose properties are therefore susceptible to even a small effective potential. For example, a gap will be opened between two degenerate states given arbitrarily small coupling. As a result, the spectral change of the system will generally be very sensitive to different types of effective potential generated by interlayer interaction, which provides both a convenient method to investigate components in the effective potential; as well

as great opportunities to engineer materials properties in a designed way with specific component in the potential.

In this chapter, I will present one such example in graphene/hBN heterostructure: Electrons in graphene are described by relativistic Dirac-Weyl spinors with a two-component pseudospin(*38, 40, 73-82*). The control of pseudospin, such as opening a pseudospin gap at the Dirac point (*83-89*), has been among the center issues in graphene research. Practically, however, it is rather difficult to directly access the pseudospin degree of freedom in graphene with traiditional external field. On the other hand, the interaction in van der Waals heterostructure creates a "spinor potential" that can readily manipulate the pseudospin in distinctively different ways with different components.

As an overview of the chapter, I will first elucidate the respective role of direct mixing and effective potential in graphene/hBN system, which helps further clarify situations where the latter one can be dominant. Then I will combine results from chapter 2, section 2.5.7 to discuss the distinctive physical nature of each component in the effective potential, as well as how they can directly manipulate graphene pseudospin in different ways. Section 5.3 will present experimental studies that reveal peculiar spectral change in graphene/hBN heterostructure, which suggests the dominance of "pseudospin-mixing" component in the effective potential, and is consistent with theoretical result. Part of the discussion is also presented in a published paper (*3*).

5.2 Effective Potential in Graphene/hBN Heterostructure

5.2.1 Role of Direct Mixing

In chapter 2, section 2.5.3, I have given a general argument on conditions that the direct mixing may be neglected. Now we can see how the argument works in the specific example of graphene/hBN heterostructure.

Let us first calculate the effect of direct mixing. The wavefunction mixing given by 1st line of Eq. 2-22 should be very small, because of the large energy difference in denominator. Therefore let us consider the corresponding energy shift of states and the change in energy spectrum of the system.

The energy shift from direct mixing is given by (in perturbative regime):

$$\Delta E_{1\boldsymbol{k}} = \sum_{\boldsymbol{q}} \frac{|\langle 2\boldsymbol{q}|H_I|1\boldsymbol{k}\rangle|^2}{E_{1\boldsymbol{k}} - E_{2\boldsymbol{q}}} \quad (5-1)$$

which is simply the 2nd order perturbation in energy. Layer 1, 2 represent the graphene and hBN layer, respectively. Similar to derivation in chapter 2, section 2.5.7, again we can divide states from hBN to boron state (conduction band) and nitrogen state (valence band), and consider them separately. Therefore, assuming $\boldsymbol{k}$ is close to $\boldsymbol{K}$ so that $\boldsymbol{k} \approx \frac{1}{3}(\boldsymbol{b}_1 - \boldsymbol{b}_2)$, there are three substrate states with largest coupling matrix element:

$$\boldsymbol{q}_1 = \boldsymbol{k} \approx \frac{1}{3}(\boldsymbol{b}_1 - \boldsymbol{b}_2), \boldsymbol{q}_2 = \boldsymbol{k} + \boldsymbol{b}_2 \approx \frac{1}{3}(\boldsymbol{b}_1 + 2\boldsymbol{b}_2), \boldsymbol{q}_3 = \boldsymbol{k} - \boldsymbol{b}_1 \approx \frac{1}{3}(-2\boldsymbol{b}_1 - \boldsymbol{b}_2) \quad (5-2)$$

The wavefunction of graphene state $\boldsymbol{k}$ can be expanded in to contribution from two sublattices:

$$|1\boldsymbol{k}\rangle = \psi^{1\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{2N}} \sum_{\boldsymbol{R}_1} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \left[\varphi(\boldsymbol{r} - \boldsymbol{R}_1) + e^{i\phi(\boldsymbol{k})}(\boldsymbol{r} - \boldsymbol{R}_1 - \boldsymbol{\tau})\right] = \psi_A^{1\boldsymbol{k}} + \psi_B^{1\boldsymbol{k}} \quad (5-3)$$

$\phi(\boldsymbol{k})$ represent the phase difference between the two sublattices. Following the derivation in Eq. 2-45, we have:

$$\langle 1\boldsymbol{k}|H_I|2\boldsymbol{q}\rangle = \frac{1}{\sqrt{2}}\left[t_1(\boldsymbol{q}) + e^{i\phi(\boldsymbol{k})}t_2(\boldsymbol{q})\right]$$

$$\approx \frac{1}{\sqrt{2}} t_S e^{-\frac{d}{\lambda}} e^{-\frac{\lambda d}{2}K^2} \left(1 + e^{i\phi(\boldsymbol{k})} e^{i\boldsymbol{q}\cdot\boldsymbol{\tau}}\right) = \frac{M_S}{\sqrt{2}} \left(1 + e^{i\phi(\boldsymbol{k})} e^{i\boldsymbol{q}\cdot\boldsymbol{\tau}}\right) \quad (5-4)$$

where $S = B, N$ to represent the contribution from boron and nitrogen states, respectively. We therefore have:

$$|\langle 2\boldsymbol{q}|H_I|1\boldsymbol{k}\rangle|^2 = M_S^2[1 + \cos(\phi(\boldsymbol{k}) + \boldsymbol{q}\cdot\boldsymbol{\tau})] \quad (5-5)$$

Plugging Eq. 5-2 into Eq. 5-5 and using $\boldsymbol{\tau} = \frac{1}{3}(\boldsymbol{a}_1 - 2\boldsymbol{a}_2)$, Eq. 5-1 becomes:

$$\Delta E_{1\boldsymbol{k}} \approx -\frac{|M_S|^2}{E_S}\left[3 + \cos(\phi) + \cos\left(\phi + \frac{2\pi}{3}\right) + \cos\left(\phi + \frac{4\pi}{3}\right)\right] \equiv -\frac{3|M_S|^2}{E_S} \quad (5-6)$$

We find that although $\phi(\boldsymbol{k})$ is $\boldsymbol{k}$-dependent, the energy shift is not. As a result, the effect of direct mixing is a constant shift in energy for all electronic states reasonably close to the Dirac point. i.e. equivalent to a constant potential with amplitude $3|M_S|^2/E_S$. Such amplitude is actually comparable or even larger than the effective potential given by Eq. 2-47; however, it only induces a rigid shift in the energy spectrum of graphene, instead of any new spectral feature. These results are consistent with the general argument in section 2.5.3. Therefore, while direct mixing is still critical in some situations such as when the energy shift or the overall interlayer interaction is important, it can be safely neglected if we are focusing on the spectral profile change of the system.

5.2.2 Physical Nature of Different Components in Spinor Potential

In chapter 2, section 2.5.7, we have already calculated the effective potential:

$$V_{eff} = \sum_{j=1}^{6} V_j e^{i\boldsymbol{k}_j \cdot \boldsymbol{r}}, V_1 = \begin{pmatrix} u_0 + iu_3 & u_1 \\ u_1 & u_0 - iu_3 \end{pmatrix} \quad (5-7)$$

Overall, it is a two-by-two matrix acting on the two sublattices of graphene. These two sublattices lead to wavefunction that mimics the two-component spinor wavefunction for spin 1/2 particle in Pauli formalism; and is often called "pseudo-spin" degree of freedom. The effective potential is actually a "spinor" potential in this sense.

Apparently, the three components, (u_0, u_3, u_1), will have distinctive physical meaning. Figure 5.1 uses a 1D lattice to illustrate physical nature of each component in the effective potential. To capture the two component physics, each unit cell necessarily has two atoms, labeled by the blue rectangle. The simplest component, diagonal symmetric term u_0, is the same at the two sublattices of graphene, shown as a constant potential over a unit cell in Figure 5.1. It corresponds to "pseudospin-blind" potential because it will not differentiate between the two sublattices.

Similarly, the u_3 term is diagonal but anti-symmetric; and therefore should be the opposite at the two sublattices, see Figure 5.1. We may in turn call it "pseudospin-dependent" potential.

The last term, u_1, is quite special in that it is off-diagonal. While superficially it modifies the interaction between two sublattices and can be called a "pseudospin-mixing" potential; its physical nature is a little bit subtle: Instead of simply changing the nearest-neighbor hopping amplitude γ_0, it modifies bonding along different directions in different way, as illustrated in the

bottom panel of Figure 5.1 (Not exactly the same as 2D graphene case, however conceptually similar).

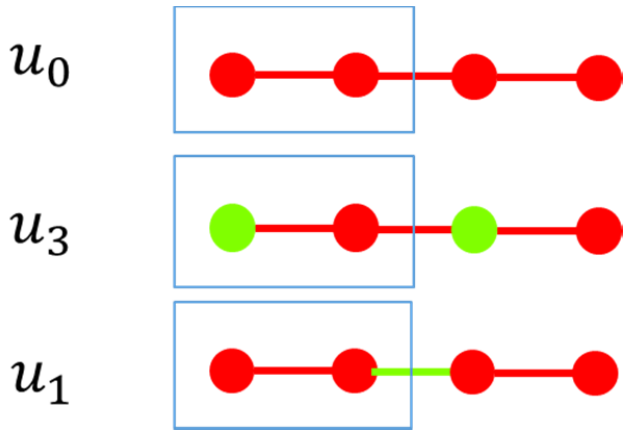

Figure 5.1 Illustration of different components from the effective potential in a 1D lattice. Each unit cell has two atoms, as indicated by the blue rectangle.

Now we can take a look how difficult it is to generate these components if without van der Waals system. The u_0 term is reasonably easy to generate since it is equivalent to an electrostatic potential that varies slowly over space. However, such slowly-varying potential only have minor effect on the properties of graphene. e.g. It cannot open any gap the energy spectrum, therefore may not be very interesting (*74, 75*).

The u_3 term, or the "mass term", has been the pursuit by many studies, since it can open a bandgap, which is critical for using graphene as field effect transistor. However, u_3 term oscillates extremely fast spatially in the order of graphene unit cell (~0.25nm); and therefore is virtually impossible to achieve with external electric field.

The u_1 term is even more peculiar. The 1D case in Figure 5.1 corresponds to the famous polyacetylene structure, which can host fascinating physics such as half-integer charge and domain wall transport. In 2D honeycomb system the u_1 term is described by Kekulé distortion, a lattice model predicted to show various interesting behaviors (*90, 91*). Nevertheless, it is not even possible to generate with commonly used external fields.

On the contrary, viewing from the perspective of interlayer interaction, we can naturally understand how these components are generated in van der Waals systems: The u_0 term has a large spatial periodicity equal to that of the Moiré superlattice, and therefore is coming from the average effect of Boron and Nitrogen atoms. In contrast, the "pseudospin-dependent" potential u_3 originates from the fast spatial oscillation between Boron and Nitrogen atoms. In addition, a single Boron or Nitrogen atom can simultaneously couple to two neighboring graphene atoms,

acting as a bridge that effectively modify the bonding between the two graphene atoms. Such modification will generally depend on the direction of the bonding, giving rise to the u_1 term.

5.2.3 Gap Opening from the Spinor Potential

With above discussion, now we can confidently say that the interlayer interaction in van der Waals systems can indeed generate effective potential that is otherwise is difficult to achieve. While have already given a brief preview of effects from each component, in this section I will discuss in more detail.

According to chapter 2, the states degenerately coupled by effective potential will be at the boundary of the Mini Brillouin Zone (MBZ), and the lowest energy one is at the center of zone boundary ($\mathbf{M}^{\mathrm{M}}$ point in Figure 2.1C) in the specific case of graphene. Hereafter we will focus on the behavior of $\mathbf{M}^{\mathrm{M}}$ point states with the modification of effective potential. For simplicity, we call $\mathbf{M}^{\mathrm{M}}$ point **m** point in short from now on.

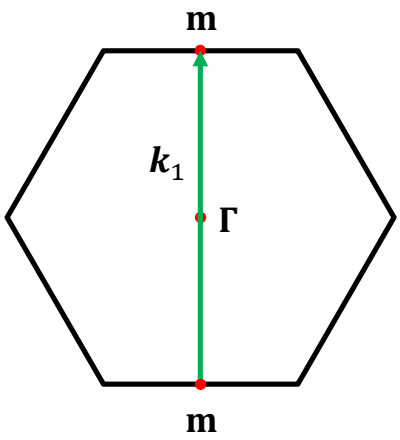

Figure 5.2 Mini Brillouin Zone (MBZ) of graphene/hBN system. Two **m** points on each end of MBZ can be degenerately coupled by $\boldsymbol{k}_1$ component of the effective potential. Original Dirac point of graphene is folded to **Γ** point in MBZ.

Two **m** points on each end of MBZ can be degenerately coupled by $\boldsymbol{k}_1$ component of the effective potential, see Figure 5.2. Their coupling matrix element is given by:

$$\langle \psi_1^\sigma | V_{eff} | \psi_2^\sigma \rangle = \frac{1}{2} \begin{pmatrix} 1 \\ -i\sigma \end{pmatrix} \begin{pmatrix} u_0 + iu_3 & u_1 \\ u_1 & u_0 - iu_3 \end{pmatrix} \begin{pmatrix} 1 \\ -i\sigma \end{pmatrix} = i(u_3 - \sigma u_1) \quad (5-8)$$

where $\sigma = +, -$ labels **m** point states from conduction band and valence band, respectively.

First thing we notice is that, there is finite coupling between originally degenerate states, which should generally lead to gap-opening. i.e. A gap should open at **m** point of MBZ, with gap size $|2(u_3 - \sigma u_1)|$. While both u_3 and u_1 component can open a gap at **m** point, u_0 cannot, which is expected because a scalar potential like u_0 cannot directly access pseudospin degree of freedom.

An even more interesting part is the sign of gap: the coupling matrix elements are the same at conduction and valence band with pseudospin-dependent potential u_3; however the opposite with pseudospin-mixing potential u_1. As a result, u_3 and u_1 will create gaps with opposite sign in either conduction or valence band. Such concept of "negative gap" is always fascinating, because it implies band inversion, and is closely related to the topological insulator. For example, topologically protected edge states will generally emerge at the boundary between regions with positive gap and negative gap.

The different sign of gaps opened by u_3 and u_1 is definitely an intriguing demonstration of the distinctive consequence of different components from the effective potential, as well as the rich physics involved. However, it can be difficult to determine whether an insulator is topological or trivial by merely measuring the bulk properties. Similarly, if we only look at the density of states or energy dispersion of the graphene/hBN system, both u_3 and u_1 potential can generate a spectrum that is very close to each other. e.g. both can open gaps of the same size in conduction and valence band. Actually, even if only the u_0 term is present and no gap is opened, the density of states might still look reasonably similar if its amplitude is large enough.

As a result, direct observation of the spinor potential has been challenging. For example, the density of states change in graphene/BN heterostructures revealed by previous scanning tunneling spectroscopy (STS) and transport measurements can be largely accounted for by a scalar periodic potential (*27-29, 92-94*), since they cannot directly probe the pseudospin degree of freedom in graphene.

On the other hand, optical transitions is not only sensitive to joint density of states (JDOS), but also the wavefunction of initial and final states from the optical selection rule. Therefore optical spectroscopy can be one potential tool to directly prove the pseudospin structure of graphene and investigate different components of the spinor potential. In the following section, we will calculate the consequence of each component on the optical spectrum of the heterostructure.

5.2.4 Optical Selection Rules with the Spinor Potential

Optical Selection Rules Near m Point of Mini Brillouin Zone

An illustration of band dispersion in MBZ is shown in Figure 5.3B-D, where a gap will be opened at **m** point in both conduction and valence band with either u_3 or u_1 term. We can label the above/below gap state within conduction/valence band with 2e, 1e, 1h, 2h, respectively. These four states are results of the degenerate coupling between **m** point states at each end of

MBZ, shown as state "1" and "2" in Figure 5.3A. Without effective potential, the unperturbed conduction/valence band states at "1" and "2" are defined as $\psi_1^c, \psi_1^v, \psi_2^c, \psi_2^v$, respectively.

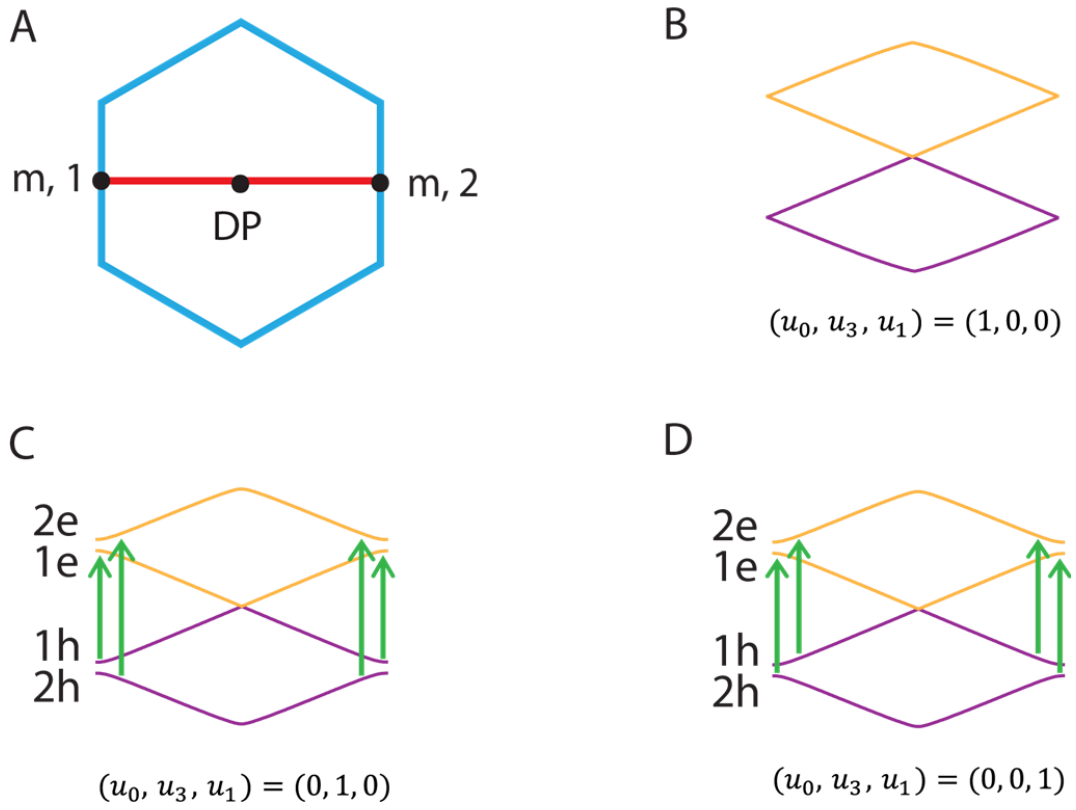

Figure 5.3 (**A**) Illustration of degenerate coupling between two states "1" and "2" both at **m** point of MBZ. (**B-D**) Optical selection rules with each component of the spinor potential. (**B**) u_0 component cannot open a gap at **m** point of MBZ. Both u_3 (**C**) are u_1 (**D**) components can open a gap at **m** point of MBZ, giving very similar energy dispersion. However, the optical transitions are only allowed between symmetric bands for u_3 component (**C**), and between largely parallel bands for u_1 component (**D**).

Considering the case of only pseudospin-dependent potential u_3, the wavefunctions of states 2e, 1e, 1h, 2h at **m** point are:

$$\psi_{2e} = (\psi_1^c + i\psi_2^c)/\sqrt{2}, \qquad \psi_{1e} = (\psi_1^c - i\psi_2^c)/\sqrt{2},$$
$$\psi_{1h} = (\psi_1^v + i\psi_2^v)/\sqrt{2}, \qquad \psi_{2h} = (\psi_1^v - i\psi_2^v)/\sqrt{2} \quad (5-9)$$

For both conduction and valence sides, the higher energy bands (2e and 1h) have the same "+" sign due to the same coupling matrix elements. Based on optical transition matrix element in bare graphene (*95*), for unperturbed graphene electronic states the optical matrix element is:

$$\langle\psi_1^v|H_{opt}|\psi_2^v\rangle = \langle\psi_1^v|H_{opt}|\psi_2^c\rangle = \langle\psi_1^c|H_{opt}|\psi_2^v\rangle = -\langle\psi_1^c|H_{opt}|\psi_2^c\rangle = 0$$

$$\langle\psi_1^v|H_{opt}|\psi_1^c\rangle = \langle\psi_2^c|H_{opt}|\psi_2^v\rangle = -\langle\psi_1^c|H_{opt}|\psi_1^v\rangle = -\langle\psi_2^v|H_{opt}|\psi_2^c\rangle$$
$$= iv_F\hat{\boldsymbol{z}}\cdot(\boldsymbol{E}\times\hat{\boldsymbol{k}}_1) = T_2 \quad (5-10)$$

Here $H_{opt} = \boldsymbol{E}\cdot\boldsymbol{\nabla}$ is the optical transition Hamiltonian, $\boldsymbol{E}$ is the light electric field, $\hat{\boldsymbol{k}}_1$ is the unit vector along momentum of state "1", $\hat{\boldsymbol{z}}$ is the unit vector along z direction perpendicular to graphene plane.

The optical transition matrix element for perturbed states

$$\langle\psi_{1h}|H_{opt}|\psi_{2e}\rangle = \langle\psi_{2h}|H_{opt}|\psi_{1e}\rangle = (\langle\psi_1^v|H_{opt}|\psi_1^c\rangle + \langle\psi_2^v|H_{opt}|\psi_2^c\rangle)/2 = 0$$
$$\langle\psi_{1h}|H_{opt}|\psi_{1e}\rangle = \langle\psi_{2h}|H_{opt}|\psi_{2e}\rangle = (\langle\psi_1^v|H_{opt}|\psi_1^c\rangle - \langle\psi_2^v|H_{opt}|\psi_2^c\rangle)/2 = T_2 \quad (5-11)$$

Therefore only transitions between 1h-1e and 2h-2e states are allowed (Figure 5.3C).

In the case of pseudospin-mixing potential u_1, we have:

$$\psi_{2e} = (\psi_1^c - i\psi_2^c)/\sqrt{2}, \quad \psi_{1e} = (\psi_1^c + i\psi_2^c)/\sqrt{2},$$
$$\psi_{1h} = (\psi_1^v + i\psi_2^v)/\sqrt{2}, \quad \psi_{2h} = (\psi_1^v - i\psi_2^v)/\sqrt{2} \quad (5-12)$$

The optical transition matrix element

$$\langle\psi_{1h}|H_{opt}|\psi_{1e}\rangle = \langle\psi_{2h}|H_{opt}|\psi_{2e}\rangle = (\langle\psi_1^v|H_{opt}|\psi_1^c\rangle + \langle\psi_2^v|H_{opt}|\psi_2^c\rangle)/2 = 0$$
$$\langle\psi_{1h}|H_{opt}|\psi_{2e}\rangle = \langle\psi_{2h}|H_{opt}|\psi_{1e}\rangle = (\langle\psi_1^v|H_{opt}|\psi_1^c\rangle - \langle\psi_2^v|H_{opt}|\psi_2^c\rangle)/2 = T_2 \quad (5-13)$$

The pseudospin-mixing potential u_1 therefore limits the transitions to between 1e-2h and 2e-1h states.

The last component u_0, does not open a gap at **m** point of MBZ, and therefore will not strongly affect optical transitions near **m** point of MBZ, see Figure 5.3B.

Optical Selection Rules away from m Point of Mini Brillouin Zone

Having obtained the optical selection rules near **m** point of MBZ, we need also examine states away from **m** point to see if they have any contribution to absorption spectrum. These states can have very different perturbation from the spinor potential: while states near MBZ boundary (region II in Fig. 5.4A) are described using the degenerate perturbation theory; states away from MBZ boundary (region I in Fig. 5.4A) only couple to other states with very different energies and the coupling can be described by the non-degenerate perturbation theory. Therefore they may have distinctive selection rules.

Consider one typical transition of this type from two states a, b at $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$. Figure 5.4B shows the two states in graphene's original Brillouin zone. The two wavevectors need satisfy the momentum selection rule $|\boldsymbol{k}_a - \boldsymbol{k}_b| = |\boldsymbol{k}_1| = q_M$. As a result, they will be folded back to the

same point in MBZ, as shown in Figure 5.4A. The involved four unperturbed graphene states are $\psi_a^c, \psi_a^v, \psi_b^c, \psi_b^v$.

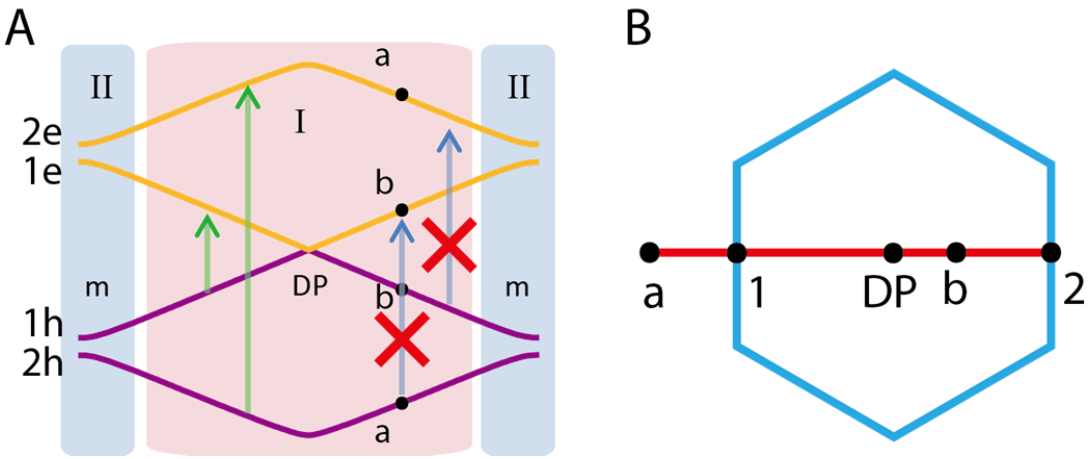

Figure 5.4 (**A**) Electronic coupling between states away from the MBZ boundary (region I, the red shaded area) are described by non-degenerate perturbation theory. Optical transitions in this region are only allowed between symmetric bands, and the resulting absorption is almost identical to that from corresponding transitions in bare graphene. In region II (blue shaded area) the Moiré superlattice potential mixes almost degenerate states, and leads to strong modification in optical absorption. (**B**) Representative coupled states from region I (a and b) and from region II (1 and 2) in the original graphene Brillouin zone. The blue hexagon is the MBZ.

Using the second order perturbation theory, we can obtain the perturbed electronic state wavefunction as

$$\psi_a^{v\prime} = \psi_a^v + \frac{u_0}{-E_M}\psi_b^c = \psi_a^v - C_1\psi_b^c$$

$$\psi_a^{c\prime} = \psi_a^c + \frac{u_0}{E_M}\psi_b^v = \psi_a^v + C_1\psi_b^v$$

$$\psi_b^{v\prime} = \psi_b^v + \frac{u_0}{-E_M}\psi_a^v = \psi_b^v - C_1\psi_a^c$$

$$\psi_b^{c\prime} = \psi_b^c + \frac{u_0}{-E_M}\psi_a^v = \psi_b^c + C_1\psi_a^v \quad (5-14)$$

Similar to Eq. 5-10, optical transition matrix element between unperturbed states are given by:

$$\langle\psi_a^v|H_{opt}|\psi_b^v\rangle = \langle\psi_a^v|H_{opt}|\psi_b^c\rangle = \langle\psi_a^c|H_{opt}|\psi_b^v\rangle = -\langle\psi_a^c|H_{opt}|\psi_b^c\rangle = 0 \quad (5-15)$$

$$\langle\psi_a^v|H_{opt}|\psi_a^v\rangle = \langle\psi_b^c|H_{opt}|\psi_b^c\rangle = -\langle\psi_a^c|H_{opt}|\psi_a^c\rangle = -\langle\psi_b^v|H_{opt}|\psi_b^v\rangle = -v_F\boldsymbol{E}\cdot\hat{\boldsymbol{k}}_a = T_1$$

$$\langle\psi_a^v|H_{opt}|\psi_a^c\rangle = \langle\psi_b^c|H_{opt}|\psi_b^v\rangle = -\langle\psi_a^c|H_{opt}|\psi_a^v\rangle = -\langle\psi_b^v|H_{opt}|\psi_b^c\rangle = iv_F\hat{\boldsymbol{z}}\cdot(\boldsymbol{E}\times\hat{\boldsymbol{k}}_a) = T_2$$

Eq. 5-15 shows that, in unperturbed bare graphene optical transitions between ψ_a^v and ψ_b^c are not allowed because $\mathbf{k}_a \neq \mathbf{k}_b$, as expected. In general, this constraint can be relaxed by the Moiré potential because it mixes electronic states with different momentum. However, we find that optical transition between $\psi_a^{v\prime}$ and $\psi_b^{c\prime}$ is still forbidden because the transition matrix element is:

$$\langle\psi_a^{v\prime}|H_{opt}|\psi_b^{c\prime}\rangle = C_1\left(\langle\psi_a^v|H_{opt}|\psi_a^v\rangle - \langle\psi_b^c|H_{opt}|\psi_b^c\rangle\right) = C_1(T_1 - T_1) = 0 \quad (5-16)$$

Meanwhile for transitions between symmetric bands we have (to first order)

$$\langle\psi_a^{v\prime}|H_{opt}|\psi_a^{c\prime}\rangle = \langle\psi_a^v|H_{opt}|\psi_a^c\rangle = T_2$$

$$\langle\psi_b^{v\prime}|H_{opt}|\psi_b^{c\prime}\rangle = \langle\psi_b^v|H_{opt}|\psi_b^c\rangle = -T_2 \quad (5-17)$$

Therefore we obtain the selection rule in region I that optical transitions only happen between symmetric bands, which give negligible optical conductivity change compared to bare graphene. We have not assumed any specific value of Moiré potential, therefore this conclusion is valid for arbitrary effective potential in Eq. 5-7.

5.2.5 Optical Conductivity Change from the Spinor Potential

Combining the discussion for both states near and away from **m** point of MBZ, we can predict the optical conductivity change of graphene with each component in the spinor potential.

With zero gap at the m point, the effect of a pseudospin-blind potential on the optical absorption is rather small (Figure 5.5B).

As for the pseudospin-dependent potential u_3, since optical transitions can only happen between symmetric bands, the optical conductivity will look quite similar to electron density of states (*92, 94*) except that the energy scale is multiplied by 2. As a result, an absorption dip will emerge at energy of $E_M = v_F q_M \approx 380$meV, see Figure 5.5C. We can therefore call it a "normal" gap.

In contrast, the allowed transitions in the pseudospin-mixing potential u_1 case are between largely parallel bands with almost the same transition energy at $E_M = v_F q_M \approx 380$meV, leading to a large joint density of states. An absorption peak will therefore emerge at this energy, opposite to the case of u_3, as shown in Figure 5.5D. The gap generated by u_1 is therefore can be termed as "inverse" gap.

When both u_3 and u_1 exist, the nature of the mini-gap opened at **m** point is determined by their relative amplitude: if $|u_1| < |u_3|$, the mini-gap is more similar to a "normal" gap, and gives an optical absorption dip at E_M; on the other hand, $|u_1| > |u_3|$ leads to an "inverse" type of gap and an optical absorption peak at E_M; $|u_1| = |u_3|$ corresponds to mini-gap closing at either valence of conduction band.

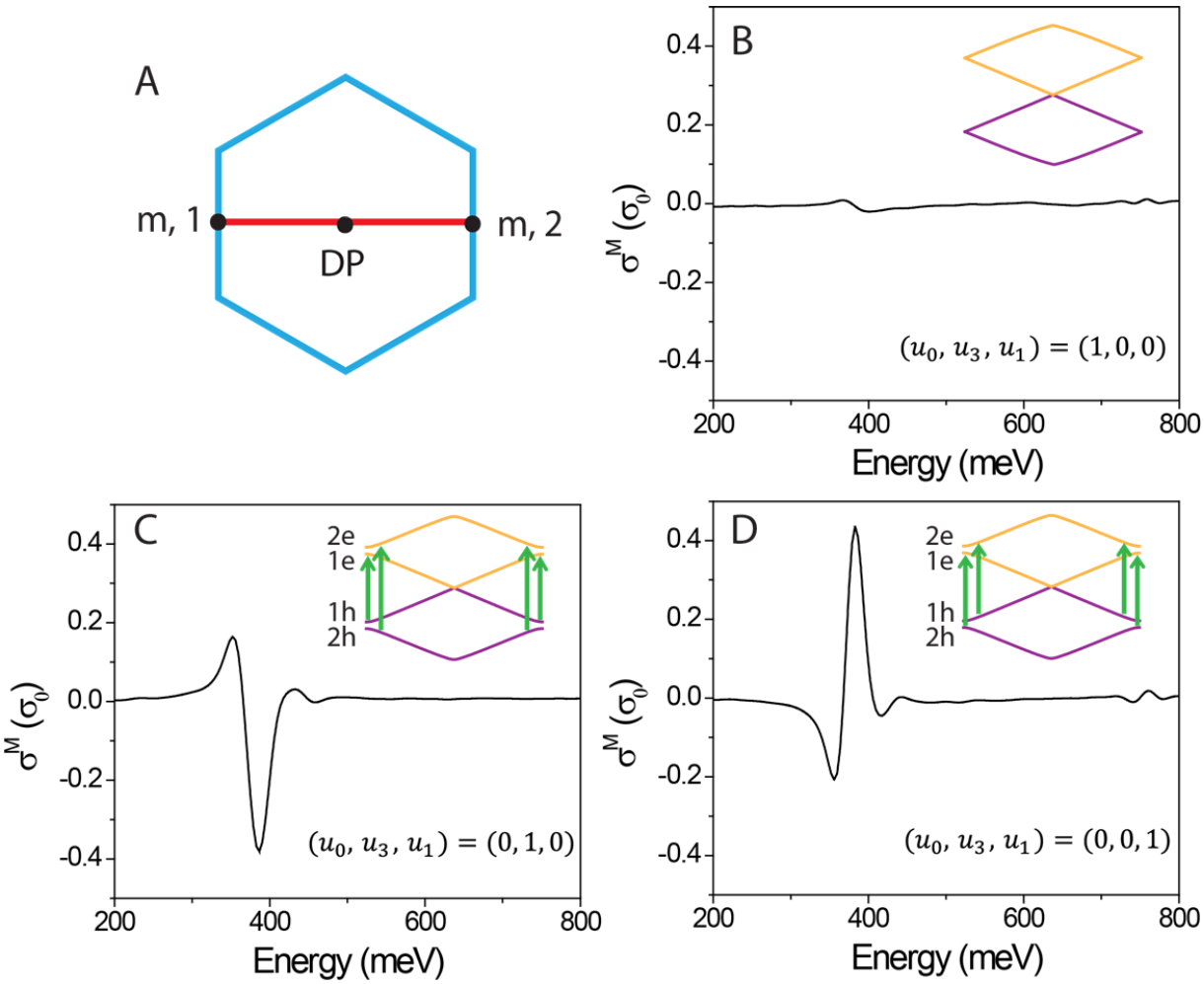

Figure 5.5 (**A**) Illustration of degenerate coupling between two states "1" and "2" both at **m** point of MBZ. (**B-D**) Optical conductivity change of graphene with each component of the spinor potential with amplitude of 10meV. Inset shows corresponding band structure in MBZ, as well as optical selection rules. (**B**) u_0 component cannot open a gap at **m** point of MBZ, therefore only slightly affects the optical spectrum of graphene. In contrast, both u_3 (**C**) are u_1 (**D**) components can open a gap at **m** point of MBZ, giving very similar energy dispersion. However, the optical selection rules is the opposite, leading to opposite change in optical conductivity.

As have been pointed out in section 5.2.3, the two types of gaps here, the "inverse" gap and "normal" gap, gives an interesting analogy to topological insulator and normal insulator. Our gap equation $u_3 \mp u_1$ is very similar to the gap equation $\Delta_3 \mp \Delta_1$ obtained in the four band topological insulator model by Kane and Mele (*96*). In that model, topological phase transition happens at $\Delta_3 = \Delta_1$, which is analogous to the transition between "inverse" and "normal" gaps in our case. And this transition corresponds to a change from absorption peak to dip at EM in

optical spectra. In this sense, optical spectroscopy is an ideal tool to investigate the distinctive consequences from different components in the spinor Moiré potential, owing to its transition matrix element dependence, and therefore sensitivity to wavefunction.

5.3 Probing the Spinor Potential in Graphene/hBN Heterostructure with Optical Spectroscopy

5.3.1 Absorption Spectra of Graphene/hBN Heterostructure

This section will presents experimental results from optical absorption measurement of graphene/hBN heterostructure. The graphene/hBN structure is obtained through directly growing graphene on hBN following a van der Waals epitaxy mode (*93*). Such growth method provides both high and uniform graphene coverage on hBN, see AFM image in Figure 5.6A; as well as preferential interlayer configuration of zero twist angle, as confirmed by the 15nm Moiré superlattice periodicity observed in high resolution AFM (Figure 5.6A inset). The transport behavior of a two terminal device (Figure 5.6B) is similar to results from previous studies (*27-29, 93*), which shows prominent resistance peak at $V_g = 0$ V and $V_g = -40$ V, see Figure 5.6C. These two peaks correspond to the original Dirac point and the **m** point in MBZ, as we have discussed in last section (see also Figure 5.6C inset). In addition, apparent asymmetry is observed between electron- and hole-doping side: the side resistance peak on the hole-doping side is much more obvious. This behavior will be explained later in section 5.3.2.

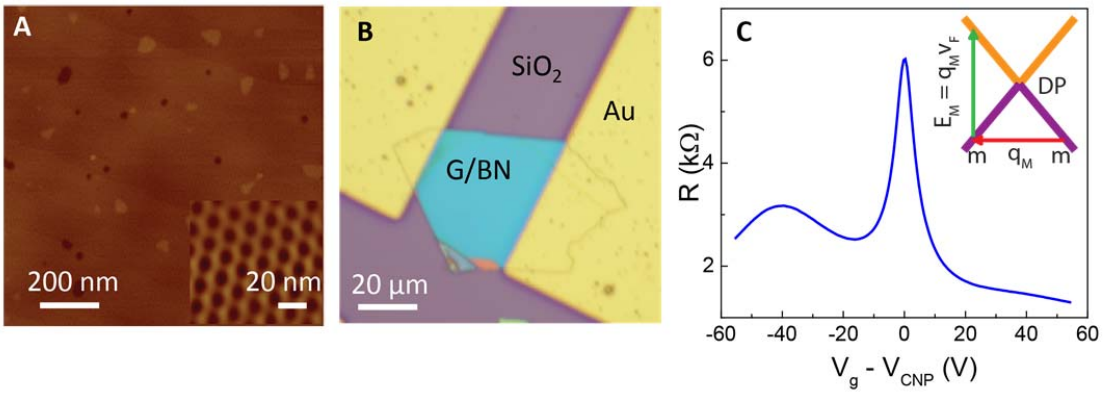

Figure 5.6 (**A**) AFM image of graphene epitaxially grown on hBN. Inset shows high resolution AFM image, where a Moiré pattern of ~15nm is observed, corresponding to zero twist angle between graphene and hBN. (**B**) Optical microscope image of graphene/hBN device on Si/SiO2 substrate. (**C**) Representative transport data from graphene/hBN device. The two resistance peaks at $V_g = 0$ V and $V_g = -40$ V correspond to the original Dirac point and mini-Dirac point on hole side. Inset illustrates the linear band dispersion of graphene, with Moiré wavevector q_M and corresponding energy E_M shown in red and green arrows.

To probe components in the spinor effective potential, we performed infrared micro-spectroscopy on graphene/BN heterostructures, as illustrated in Figure. 5.7A.

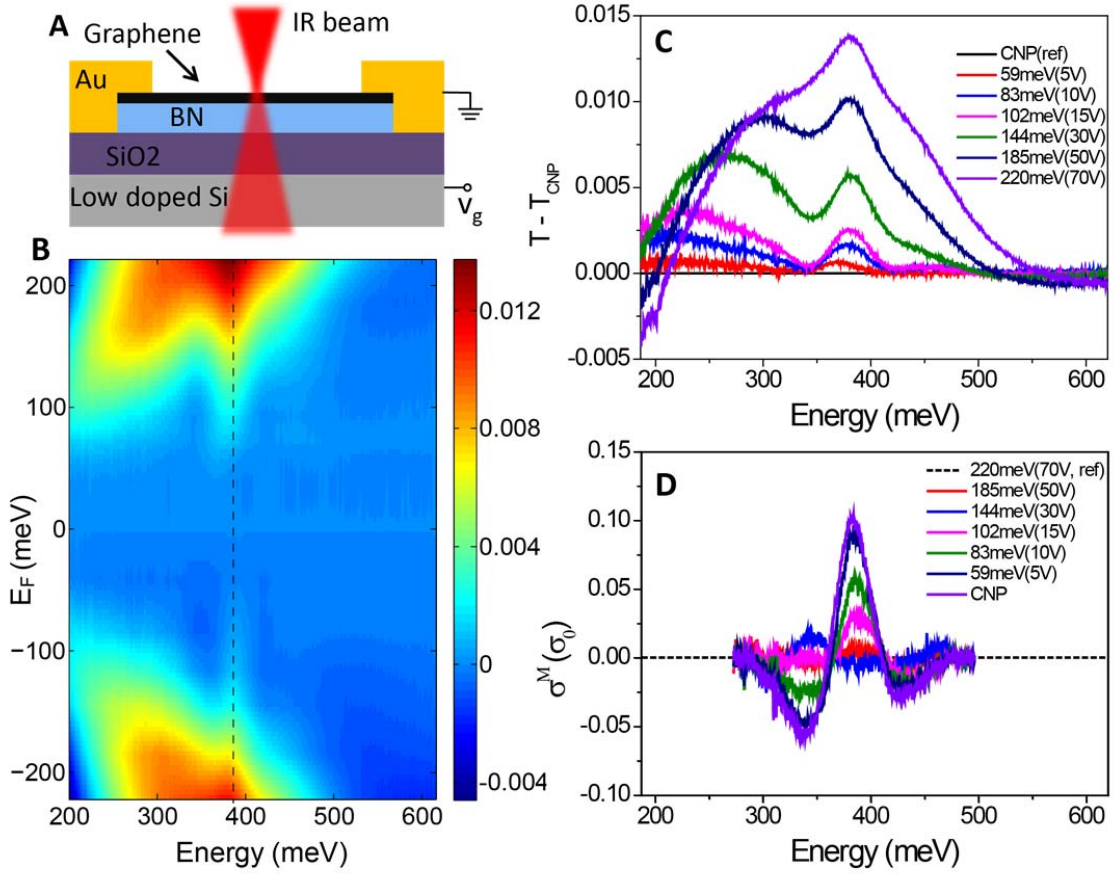

Figure 5.7 (**A**) Illustration of experimental setup measuring the optical conductivity of graphene/hBN heterostructure in infrared (IR) range. (**B**) Two dimensional mapping of the transmission difference $T - T_{\mathrm{CNP}}$ depending on excitation light energy (x axis) and doping level (y axis). T_{CNP} is the transmission spectra when graphene is at charge neutral. The broad feature shifting with E_F originates from the blocking of interband transitions; while the sharp feature at fixed energy $E_M \approx 380$meV (vertical dashed line) is a consequence of the Moiré potential. (**C**) $T - T_{\mathrm{CNP}}$ spectra at representative doping level. (**D**) Optical conductivity change σ^M induced by the Moiré potential at different gate voltages. $\sigma_0 = \pi e^2/2h$ is graphene universal conductivity.

Figure 5.7B shows the measured transmission difference ($T - T_{\mathrm{CNP}}$) of graphene/BN heterostructures depending on both excitation light energy (x axis) and doping level (y axis), with T_{CNP} the transmission spectra when graphene is at charge neutral. Two prominent features are observe in the 2D-plot: One broad feature shifting with E_F, which has been observed in bare graphene on SiO_2/Si substrate (*77-79*), and was assigned to the Pauli-blocking of interband transitions when E_F is larger than half of the transition energy. Another sharp feature, on the

contrary, always show up at $E_M \approx 380$meV regardless of E_F. From our discussion, this feature should be originating from the effect of spinor effective potential.

Figure 5.7C shows several $(T - T_{\mathrm{CNP}})$ spectra at representative E_F level (and corresponding gate voltage), from which we can clearly see the feature at $E_M \approx 380$meV. To isolate the response to the Moiré potential, we subtract the broad background and extract the interlayer interaction-induced optical conductivity change in Figure 5.7D. We have taken $\sigma^{\mathrm{M}}_{70\mathrm{V}}$ as the refenrence, because at $V_{\mathrm{g}} = 70$ V ($E_{\mathrm{F}} = 220$ meV) the optical absorption around E_M should be negligible due to Pauli blocking at $2E_F > E_M$.

5.3.2 Comparison to Theory

With the experimentally obtained optical conductivity change in Figure 5.7D , we can now have a direct comparison with theoretical prediction in section 5.2.5.

When graphene is charge neutral, a prominent absorption peak emerges at $E_M \approx 380$meV. According to our discussion in section 5.2.5, this suggest that the dominant component in the spinor potential is the "pseudospin-mixing" type u_1, and that an "inverse" gap is opened at **m** point in MBZ.

Theoretically, we have obtained expression of the effective potential at zero twist angle in Eq. 2-47: $(u_0, u_3, u_1) = V_S(\frac{1}{2}, -\frac{\sqrt{3}}{2}, -1)$. Indeed, we find that $|u_1| > |u_3|$, i.e. the pseudospin-mixing potential u_1 will dominate, and an absorption peak should emerge at E_M. We can further directly compare the experimental optical conductivity change to the one calculated with parameters given by Eq. 2-47, see Figure 5.8. $V_S = \frac{|M_S|^2}{E_S} \approx 10$meV is estimated from the pair-interaction between carbon-carbon atoms, with modification to consider the different p orbital radii of boron, carbon and nitrogen atoms (roughly 80, 60, and 50 pm, respectively). The theoretical and experimental results match amazingly well, given that there is no fitting parameters in the theoretical model. (The only two parameters, γ and λ in the pair interaction $t(r) = \gamma e^{-r/\lambda}$, are also obtained from generally accepted values in literature).

Another unanswered question is the prominent asymmetry in electron- and hole-doping side of gate-dependent transport data in Figure 5.6C. Such observation can also be naturally explained by the gap equation at **m** point in MBZ, see Eq. 5-8. The gap size on hole side, $|u_1 + u_3|$, is an order of magnitude larger than $|u_1 - u_3|$, the gap on electron side, with the above parameters. Figure 5.8B illustrates the band dispersion of graphene in MBZ with the Moiré Potential, from which we can clearly see a much larger gap opened on the hole side at **m** point in MBZ

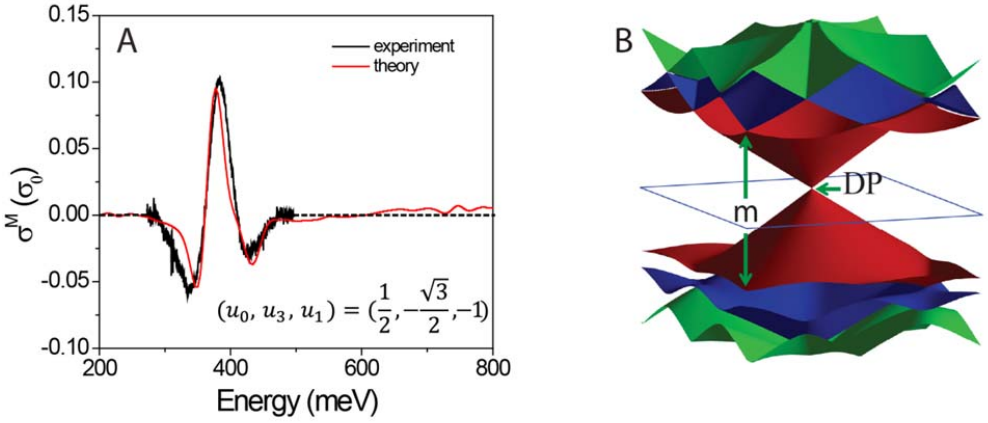

Figure 5.8 (**A**) Comparison between experimental and theoretical optical conductivity change with the Moiré potential. $V_S = 10\text{meV}$. (**B**) Band dispersion of graphene in MBZ with the Moiré potential. A much larger gap is opened at **m** point on the hold side (vertical green arrow), giving the prominent asymmetry in transport.

5.4 Summary and Outlook

In this chapter, we have examined the interlayer interaction in graphene/hBN, as a representative example of both van der Waals heterostructure and the effective potential description in the weak non-perturbative regime. The experimental results confirm our expectation from chapter 2, that indeed the effective potential can provide rich effects on material properties through directly accessing the sublattice degree of freedom of the system. For graphene, such degree of freedom corresponds to the two-component pseudo-spin, and different components in the effective potential therefore correspond to "pseudospin-blind", "pseudospin-dependent", and "pseudospin-mixing" potential. One can readily extend the conclusion to arbitrary system: generally the diagonal term in the effective potential corresponds to a modification of the on-site energy of sublattices (or orbitals), while the off-diagonal term modifies the hopping between them. These components can be difficult to realize unless with an optical lattice; however are also highly desirable, since many fascinating physics emerge exactly from manipulating these parameters in toys models.

Another important message from this chapter is the unique advantage of optical spectroscopy in investigating the distinctive phases of a system: Optical transitions are sensitive to the transition matrix element, and therefore naturally encode wavefunction information of initial and final states. For example, while the energy dispersion of two bands may look almost identical after band inversion, their wavefunction will be inverted. Therefore, if originally optical transitions are only allowed between the upper band and some other states; then after band inversion only the lower band can have optical transitions to those states, which provides a natural way to distinguish between the two cases without measuring the boundary states. Indeed, the use of optical technique to study phases with nontrivial topology, such as Weyl semimetal, has been growing rapidly, because the topology information is necessarily stored in the wavefunction.

From theoretical aspect, we again see the success of approaches developed in chapter 2, this time in the case of effective potential description. It is quite impressive that the theoretical predictions match quite well with experimental results even without fitting parameter. Combining great simplicity and reasonable accuracy, the effective potential description will be very helpful in providing a physically intuitive picture relating interlayer configuration to its effect on system properties; as well as guidance to directional engineering via design of van der Waals structures.

Chapter 6 – Things Not Included

I hope that present article can provide useful tools to investigate the properties of van der Waals heterostructures. Certainly, not all situations are covered; therefore here I will briefly discuss things not included to avoid potential confusion.

Firstly, all the discussions are based on tight binding model and apparently do not include many-body interaction, i.e. Coulomb interaction. As we have explained in chapter 2, the single particle picture can be good enough to determine certain properties in many cases, especially when we are focusing on the effects of interlayer interaction but not many-body interaction. Even in case that we do want to focus on many-body effects, such as excitons, obtaining the single particle eigen-states with interlayer interaction is still important, because single particle eigen-states are often the basis and starting point to solve equations considering higher order interactions. e.g. Exciton wavefunction can be expressed as an superposition of product of electron and hole eigenstates, as one can see from the Bethe-Salpeter equation. Furthermore, much information can already be acquired without even going into the complicated Feynman diagram. For example, some selection rules should still hold regardless of the strength of interaction. Another example is to obtain band parameters, such as the effective mass, from the single particle picture first, and then calculate exciton binding energy semi-classically with hydrogen model.

Secondly, we have been discussing "eigen-states", or steady states of the system, therefore all the properties we obtained can be essentially called "static properties". On the other hand, the "dynamic properties", e.g. dynamics of excited states, can also be dramatically modified through interlayer interaction. One representative example is the charge transfer process in van der Waals heterostructure with type II band alignment. The induced ultrafast charge separation into different layers, and thereby exciton-free-carrier conversion, can be of great potential in optoelectronics, spintronics and valleytronics, allowing us to realize both near-unity generation efficiency of spin/valley polarization and microseconds-long valley lifetime in a recent work (*4*). These dynamic properties are also closely related to the static properties, because the scattering cross section is ultimately another type of matrix element; therefore we would expect that the concepts and approaches we have developed should still work, in slightly different form.

All the results up to this point are still considering interactions between electrons from different layers. However, there are also other important quasi-particles and interactions between them, such as the electron-phonon interaction. Recently we demonstrate that strong interlayer electron-phonon interaction indeed exist in the van der Waals heterostructure of WSe2/hBN (*5*), leading to the emergence of hBN ZO mode and an "X" peak in the emission and absorption spectrum, respectively. The interlayer electron-phonon coupling should generally exist in van der Waals systems, enabling another set of new rich possibilities, such as interlayer-mediated charge density wave and superconductivity. Given the similarity between phonon modes and electron modes, e.g. their periodicity and energy dispersion, we would again expect a reasonably straightforward transfer of results here into the context of interlayer electron-phonon interaction.

References

1. L. K. Li *et al.*, Direct observation of the layer-dependent electronic structure in phosphorene. *Nat Nanotechnol* **12**, 21 (Jan, 2017).
2. K. H. Liu *et al.*, Van der Waals-coupled electronic states in incommensurate double-walled carbon nanotubes. *Nat Phys* **10**, 737 (Oct, 2014).
3. Z. W. Shi *et al.*, Gate-dependent pseudospin mixing in graphene/boron nitride moire superlattices. *Nat Phys* **10**, 743 (Oct, 2014).
4. J. Kim *et al.*, Observation of Ultralong Valley Lifetime in WSe2/MoS2 Heterostructures. *Arxiv*, 1612.05359 (2017).
5. C. Jin *et al.*, Interlayer electron-phonon coupling in WSe2/hBN heterostructures. *Nat Phys* **13**, 127 (02//print, 2017).
6. J. C. Slonczewski, P. R. Weiss, Band Structure of Graphite. *Phys Rev* **109**, 272 (1958).
7. E. McCann, M. Koshino, The electronic properties of bilayer graphene. *Rep Prog Phys* **76**, (May, 2013).
8. G. T. de Laissardiere, D. Mayou, L. Magaud, Localization of Dirac Electrons in Rotated Graphene Bilayers. *Nano Lett* **10**, 804 (Mar, 2010).
9. E. S. Morell, J. D. Correa, P. Vargas, M. Pacheco, Z. Barticevic, Flat bands in slightly twisted bilayer graphene: Tight-binding calculations. *Phys Rev B* **82**, (Sep 13, 2010).
10. R. Bistritzer, A. H. MacDonald, Moire bands in twisted double-layer graphene. *P Natl Acad Sci USA* **108**, 12233 (Jul 26, 2011).
11. L. J. Yin *et al.*, Landau quantization and Fermi velocity renormalization in twisted graphene bilayers. *Phys Rev B* **92**, (Nov 10, 2015).
12. K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, Atomically Thin MoS2: A New Direct-Gap Semiconductor. *Phys Rev Lett* **105**, (Sep 24, 2010).
13. A. Splendiani *et al.*, Emerging Photoluminescence in Monolayer MoS2. *Nano Lett* **10**, 1271 (Apr, 2010).
14. E. Cappelluti, R. Roldan, J. A. Silva-Guillen, P. Ordejon, F. Guinea, Tight-binding model and direct-gap/indirect-gap transition in single-layer and multilayer MoS2. *Phys Rev B* **88**, (Aug 8, 2013).
15. K. H. Liu *et al.*, Evolution of interlayer coupling in twisted molybdenum disulfide bilayers. *Nat Commun* **5**, (Sep, 2014).
16. C. Berger *et al.*, Electronic confinement and coherence in patterned epitaxial graphene. *Science* **312**, 1191 (May 26, 2006).
17. X. S. Wu, X. B. Li, Z. M. Song, C. Berger, W. A. de Heer, Weak antilocalization in epitaxial graphene: Evidence for chiral electrons. *Phys Rev Lett* **98**, (Mar 30, 2007).

18. J. Hass *et al.*, Why multilayer graphene on 4H-SiC(000(1)over-bar) behaves like a single sheet of graphene. *Phys Rev Lett* **100**, (Mar 28, 2008).

19. A. Luican *et al.*, Single-Layer Behavior and Its Breakdown in Twisted Graphene Layers. *Phys Rev Lett* **106**, (Mar 21, 2011).

20. J. M. B. L. dos Santos, N. M. R. Peres, A. H. Castro, Graphene bilayer with a twist: Electronic structure. *Phys Rev Lett* **99**, (Dec 21, 2007).

21. S. Shallcross, S. Sharma, E. Kandelaki, O. A. Pankratov, Electronic structure of turbostratic graphene. *Phys Rev B* **81**, (Apr 15, 2010).

22. G. H. Li *et al.*, Observation of Van Hove singularities in twisted graphene layers. *Nat Phys* **6**, 109 (Feb, 2010).

23. K. Kim *et al.*, Raman Spectroscopy Study of Rotated Double-Layer Graphene: Misorientation-Angle Dependence of Electronic Structure. *Phys Rev Lett* **108**, (Jun 14, 2012).

24. R. W. Havener, H. L. Zhuang, L. Brown, R. G. Hennig, J. Park, Angle-Resolved Raman Imaging of Inter layer Rotations and Interactions in Twisted Bilayer Graphene. *Nano Lett* **12**, 3162 (Jun, 2012).

25. C. H. Park, L. Yang, Y. W. Son, M. L. Cohen, S. G. Louie, New generation of massless Dirac fermions in graphene under external periodic potentials. *Phys Rev Lett* **101**, (Sep 19, 2008).

26. M. Yankowitz *et al.*, Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nat Phys* **8**, 382 (May, 2012).

27. L. A. Ponomarenko *et al.*, Cloning of Dirac fermions in graphene superlattices. *Nature* **497**, 594 (May 30, 2013).

28. C. R. Dean *et al.*, Hofstadter's butterfly and the fractal quantum Hall effect in moire superlattices. *Nature* **497**, 598 (May 30, 2013).

29. B. Hunt *et al.*, Massive Dirac Fermions and Hofstadter Butterfly in a van der Waals Heterostructure. *Science* **340**, 1427 (Jun 21, 2013).

30. D. R. Hofstadter, Energy-Levels And Wave-Functions Of Bloch Electrons In Rational And Irrational Magnetic-Fields. *Phys Rev B* **14**, 2239 (1976).

31. J. R. Wallbank, A. A. Patel, M. Mucha-Kruczynski, A. K. Geim, V. I. Falko, Generic miniband structure of graphene on a hexagonal substrate. *Phys Rev B* **87**, (Jun 4, 2013).

32. M. Kindermann, B. Uchoa, D. L. Miller, Zero-energy modes and gate-tunable gap in graphene on hexagonal boron nitride. *Phys Rev B* **86**, (Sep 13, 2012).

33. Q. Tong *et al.*, Topological mosaics in moire superlattices of van der Waals heterobilayers. *Nat Phys* **advance online publication**, (2016).

34. J. C. W. Song, P. Samutpraphoot, L. S. Levitov, Topological Bloch bands in graphene superlattices. *P Natl Acad Sci USA* **112**, 10879 (Sep 1, 2015).

35. K. Watanabe, T. Taniguchi, H. Kanda, Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. *Nat Mater* **3**, 404 (Jun, 2004).

36. L. K. Li *et al.*, Black phosphorus field-effect transistors. *Nat Nanotechnol* **9**, 372 (May, 2014).

37. H. Liu *et al.*, Phosphorene: An Unexplored 2D Semiconductor with a High Hole Mobility. *Acs Nano* **8**, 4033 (Apr, 2014).

38. K. S. Novoselov *et al.*, Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197 (Nov 10, 2005).

39. X. X. Xi *et al.*, Strongly enhanced charge-density-wave order in monolayer NbSe2. *Nat Nanotechnol* **10**, 765 (Sep, 2015).

40. Y. B. Zhang, Y. W. Tan, H. L. Stormer, P. Kim, Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201 (Nov 10, 2005).

41. R. Hultgren, N. S. Gingrich, B. E. Warren, The atomic distribution in red and black phosphorus and the crystal structure of black phosphorus. *J Chem Phys* **3**, 351 (Jun, 1935).

42. J. Yang *et al.*, Optical tuning of exciton and trion emissions in monolayer phosphorene. *Light-Sci Appl* **4**, (Jul, 2015).

43. R. J. Xu *et al.*, Extraordinarily Bound Quasi-One-Dimensional Trions in Two-Dimensional Phosphorene Atomic Semiconductors. *Acs Nano* **10**, 2046 (Feb, 2016).

44. S. Zhang *et al.*, Extraordinary Photoluminescence and Strong Temperature/Angle-Dependent Raman Responses in Few-Layer Phosphorene. *Acs Nano* **8**, 9590 (Sep, 2014).

45. C. Jin *et al.*, On Optical Dipole Moment and Radiative Recombination Lifetime of Excitons in WSe2. *Advanced Functional Materials*, n/a (2016).

46. L. Wang *et al.*, One-Dimensional Electrical Contact to a Two-Dimensional Material. *Science* **342**, 614 (Nov 1, 2013).

47. K. F. Mak *et al.*, Measurement of the Optical Conductivity of Graphene. *Phys Rev Lett* **101**, (Nov 7, 2008).

48. F. Wang *et al.*, Gate-variable optical transitions in graphene. *Science* **320**, 206 (Apr 11, 2008).

49. A. Chernikov *et al.*, Exciton Binding Energy and Nonhydrogenic Rydberg Series in Monolayer WS2. *Phys Rev Lett* **113**, (Aug 13, 2014).

50. M. M. Ugeda *et al.*, Giant bandgap renormalization and excitonic effects in a monolayer transition metal dichalcogenide semiconductor. *Nat Mater* **13**, 1091 (Dec, 2014).

51. Z. L. Ye *et al.*, Probing excitonic dark states in single-layer tungsten disulphide. *Nature* **513**, 214 (Sep 11, 2014).

52. J. Maultzsch *et al.*, Exciton binding energies in carbon nanotubes from two-photon photoluminescence (vol 72, art. no. 241402(R), 2005). *Phys Rev B* **74**, (Oct, 2006).

53. F. Wang, G. Dukovic, L. E. Brus, T. F. Heinz, The optical resonances in carbon nanotubes arise from excitons. *Science* **308**, 838 (May 6, 2005).

54. S. Frank, P. Poncharal, Z. L. Wang, W. A. de Heer, Carbon nanotube quantum resistors. *Science* **280**, 1744 (Jun 12, 1998).

55. S. Uryu, T. Ando, Electronic intertube transfer in double-wall carbon nanotubes. *Phys Rev B* **72**, 245403 (Dec, 2005).

56. M. Kociak *et al.*, Linking Chiral Indices and Transport Properties of Double-Walled Carbon Nanotubes. *Phys Rev Lett* **89**, 155501 (2002).

57. F. Villalpando-Paez *et al.*, Raman Spectroscopy Study of Isolated Double-Walled Carbon Nanotubes with Different Metallic and Semiconducting Configurations. *Nano Lett* **8**, 3879 (Nov, 2008).

58. J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, L. A. Nagahara, Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science* **300**, 1419 (May 30, 2003).

59. A. Hashimoto *et al.*, Atomic correlation between adjacent graphene layers in double-wall carbon nanotubes. *Phys Rev Lett* **94**, 045504 (Feb 4, 2005).

60. K. Hirahara *et al.*, Chirality correlation in double-wall carbon nanotubes as studied by electron diffraction. *Phys Rev B* **73**, 195420 (May, 2006).

61. K. H. Liu *et al.*, Direct determination of atomic structure of large-indexed carbon nanotubes by electron diffraction: application to double-walled nanotubes. *J Phys D Appl Phys* **42**, 125412 (Jun 21, 2009).

62. K. H. Liu *et al.*, Chirality-Dependent Transport Properties of Double-Walled Nanotubes Measured in Situ on Their Field-Effect Transistors. *J Am Chem Soc* **131**, 62 (Jan 14, 2009).

63. D. Levshov *et al.*, Experimental Evidence of a Mechanical Coupling between Layers in an Individual Double-Walled Carbon Nanotube. *Nano Lett* **11**, 4800 (Nov, 2011).

64. K. H. Liu *et al.*, Quantum-coupled radial-breathing oscillations in double-walled carbon nanotubes. *Nat Commun* **4**, 1375 (Jan, 2013).

65. R. Saito, M. S. Dresselhaus, G. Dresselhaus, *Physical Properties of Carbon Nanotubes*. (World Scientific Publishing Company, 1998).

66. R. Saito, G. Dresselhaus, M. S. Dresselhaus, Electronic-Structure of Double-Layer Graphene Tubules. *Journal of Applied Physics* **73**, 494 (Jan 15, 1993).

67. S. Roche, F. Triozon, A. Rubio, D. Mayou, Conduction mechanisms and magnetotransport in multiwalled carbon nanotubes. *Phys Rev B* **64**, (Sep 15, 2001).

68. J. Lefebvre, P. Finnie, Polarized light microscopy and spectroscopy of individual single-walled carbon nanotubes. *Nano Res* **4**, 788 (Aug, 2011).

69. K. H. Liu *et al.*, High-throughput optical imaging and spectroscopy of individual carbon nanotubes in devices. *Nat Nanotechnol* **8**, 917 (Dec, 2013).

70. K. Liu *et al.*, An atlas of carbon nanotube optical transitions. *Nat Nanotechnol* **7**, 325 (2012).

71. J. Lefebvre, J. M. Fraser, Y. Homma, P. Finnie, Photoluminescence from single-walled carbon nanotubes: a comparison between suspended and micelle-encapsulated nanotubes. *Appl Phys a-Mater* **78**, 1107 (May, 2004).

72. M. Steiner *et al.*, How does the substrate affect the Raman and excited state spectra of a carbon nanotube? *Appl Phys a-Mater* **96**, 271 (Aug, 2009).

73. K. S. Novoselov *et al.*, Electric field effect in atomically thin carbon films. *Science* **306**, 666 (Oct, 2004).

74. A. H. C. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, The electronic properties of graphene. *Reviews of Modern Physics* **81**, 109 (2009).

75. M. I. Katsnelson, K. S. Novoselov, A. K. Geim, Chiral tunnelling and the Klein paradox in graphene. *Nat Phys* **2**, 620 (2006).

76. A. K. Geim, K. S. Novoselov, The rise of graphene. *Nat Mater* **6**, 183 (2007).

77. F. Wang *et al.*, Gate-variable optical transitions in graphene. *Science* **320**, 206 (Apr 11, 2008).

78. Z. Q. Li *et al.*, Dirac charge dynamics in graphene by infrared spectroscopy. *Nat Phys* **4**, 532 (Jul, 2008).

79. J. Horng *et al.*, Drude conductivity of Dirac fermions in graphene. *Phys Rev B* **83**, 165113 (Apr 15, 2011).

80. H. Min, G. Borghi, M. Polini, A. H. MacDonald, Pseudospin magnetism in graphene. *Phys Rev B* **77**, 041407 (Jan, 2008).

81. J. Jung, F. Zhang, A. H. MacDonald, Lattice theory of pseudospin ferromagnetism in bilayer graphene: Competing interaction-induced quantum Hall states. *Phys Rev B* **83**, 115408 (Mar 7, 2011).

82. P. San-Jose, E. Prada, E. McCann, H. Schomerus, Pseudospin Valve in Bilayer Graphene: Towards Graphene-Based Pseudospintronics. *Phys Rev Lett* **102**, 247204 (Jun 19, 2009).

83. Y. B. Zhang *et al.*, Direct observation of a widely tunable bandgap in bilayer graphene. *Nature* **459**, 820 (Jun, 2009).

84. E. McCann, Asymmetry gap in the electronic band structure of bilayer graphene. *Phys Rev B* **74**, 161403 (2006).

85. T. Ohta, A. Bostwick, T. Seyller, K. Horn, E. Rotenberg, Controlling the electronic structure of bilayer graphene. *Science* **313**, 951 (Aug, 2006).

86. C. H. Lui, Z. Li, K. F. Mak, E. Cappelluti, T. F. Heinz, Observation of an electrically tunable band gap in trilayer graphene. *Nat Phys* **7**, 944 (Dec, 2011).

87. Y. W. Son, M. M. Cohen, S. G. Louie, Half-metallic graphene nanoribbons. *Nature* **444**, 347 (2006).

88. M. Y. Han, B. Ozyilmaz, Y. Zhang, P. Kim, Energy band gap engineering of graphene nanoribbons. *Phys. Rev. Lett.* **98**, 206805 (2007).

89. P. Avouris, Z. H. Chen, V. Perebeinos, Carbon-based electronics. *Nat Nanotechnol* **2**, 605 (Oct, 2007).

90. S. H. Lee *et al.*, Band Gap Opening by Two-Dimensional Manifestation of Peierls Instability in Graphene. *Acs Nano* **5**, 2964 (Apr, 2011).

91. B. Roy, I. F. Herbut, Unconventional superconductivity on honeycomb lattice: the theory of Kekule order parameter. *Phys Rev B* **82**, (Jul 20, 2010).

92. M. Yankowitz *et al.*, Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nat Phys* **8**, 382 (May, 2012).

93. W. Yang *et al.*, Epitaxial growth of single-domain graphene on hexagonal boron nitride. *Nat Mater* **12**, 792 (Sep, 2013).

94. C.-H. Park, L. Yang, Y.-W. Son, M. L. Cohen, S. G. Louie, New generation of massless Dirac fermions in graphene under external periodic potentials. *Phys Rev Lett* **101**, 126804 (Sep 19, 2008).

95. R. R. Nair *et al.*, Fine structure constant defines visual transparency of graphene. *Science* **320**, 1308 (Jun 6, 2008).

96. C. L. Kane, E. J. Mele, Quantum Spin Hall Effect in Graphene. *Phys Rev Lett* **95**, 226801 (2005).