Twisted Bilayer Graphene

Amir Shapour Mohammadi

Junior Paper
Department of Physics
Princeton University
Fall 2022

Adviser: F. D. M. Haldane
Second reader: Shinsei Ryu

This paper represents my work in accordance with University regulations.

/s/ Amir Shapour Mohammadi

Abstract

The purpose of this paper is to provide a self-contained review of modern literature on twisted bilayer graphene (TBG). An overview of the necessary tools for modelling electrons in a lattice will be provided culminating in the tight-binding model. We first understand why graphene is special in its own right by computing its band structure and understanding its close connection with the Dirac equation. We will then closely follow the work of Bitstrizer and MacDonald to derive their eponymous continuum model of TBG which immensely reduces the computation cost of band structure calculations. This model will then be used numerically compute the band structure and other electronic properties. The key feature of TBG is the existence of flat bands at so-called magic angles of twist at which interactions become dominant leading to rich phenomena including ferromagnetism, superconductivity, and fractional Chern insulator phases. An overview of band topology and geometry will be provided which will allow us to understand the Haldane model for Chern insulator phases in graphene. We then build on this idea and review recent literature exploring fractional Chern insulator phases in magic-angle TBG.

Contents

Table 1: Description of most used variables.

Graphene	
a_c	Carbon-carbon spacing (≈ 1.42Å)
$\mathbf{a}_i$	Primitive lattice vectors
$\boldsymbol{\tau}_i$	Nearest-neighbor hopping vectors
$\mathbf{K}, \mathbf{K}'$	Dirac points
α	Sublattice index
η	Valley index
m	Sublattice potential
t	Hopping amplitude
v_F	Fermi velocity
Twisted Bilayer Graphene	
θ	Twist angle
Ω	Area of moire unit cell
N	Number of graphene unit cells in each layer
$\mathbf{a}_i^m$	Moire primitive lattice vectors
$\mathbf{b}_i^m$	Moire primitive reciprocal vectors
$\mathbf{q}_i^m$	Moire reciprocal nearest-neighbor hopping vectors
k_θ	Length of moire hopping vector
$\mathbf{Q}$	Moire reciprocal lattice vectors
$t_{\alpha\beta}$	Hopping amplitude
Band Topology and Geometry	
Ω	Berry curvature
$\mathcal{C}$	Chern number

Introduction

For the past few decades, graphene has been at the center of condensed matter physics. The material was first experimentally realized in 2004 by isolating layers of graphite using household scotch tape. Graphene is comprised of carbon atoms which bond together to form a honeycomb lattice. Its band structure reveals that electrons near so-called Dirac points remarkably behave as if they were massless and relativistic, paving the way for powerful applications and novel theoretical effects.

Meanwhile, the theory of topological insulators and quantum Hall effects were flourishing in the 1980s. A topological insulator is a material whose bulk acts as an insulator but with edges capable of conduction. The first realization of a topological insulator was the integer quantum Hall effect (IQHE) which had the notable feature of quantized Hall (transverse) conductivity. It was discovered experimentally by von Klitzing in 1980 followed by its relatively straightforward theoretical explanation. In 1987, Haldane then used graphene as a platform for the Chern insulator, a lattice analogue of the IQHE with broken time-reversal symmetry. The Chern insulator exhibits the quantum anomalous Hall effect (QAHE) with origins in the geometric properties of the material's band structure. In 1982, the fractional quantum Hall effect (FQHE) was discovered by experimentalists who noticed that the Hall conductivity was also quantized at fractional values of the quantum conductance. The explanation for the phenomena was much harder than that of the IQHE due to the fundamental role of electron interactions.

More recently, the search for materials with novel topological effects has been growing. One of the main contenders is twisted bilayer graphene, a material composed of two parallel layers of graphene with one layer rotated with respect to the other. The dynamics are highly sensitive to twist angle, and commensurate angles of rotation, the bilayer system exhibits a moire periodic structure. At so-called magic angles, the interaction between the layers forms bands with nearly zero band width known as flat bands. Electrons in these flat bands are highly sensitive to interactions which give rise to rich phenomena including the fractional Chern insulator, a lattice analogue of the FQHE. The objective of this paper is to provide an overview of the theory of twisted bilayer graphene and the existence fractional Chern insulator phases in a simplified limit.

1 Condensed Matter Physics

We begin with a review of condensed matter physics which provides a quantum description of materials. As the name suggests, condensed matter deals primarily with solids but is typically used as an umbrella term which may include liquids and other exotic phases. We follow the theory outlined in [?, ?]. We develop the theory in two dimensions, but the arguments below readily generalize to arbitrary dimension.

1.1 Lattices

Most solids are crystals (periodic lattices) where the lattice sites correspond to atoms. A Bravais lattice is the structure formed by a set of vectors

$$\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2, \qquad n_1, n_2 \in \mathbb{Z} \tag{1}$$

where $\mathbf{a}_j$ are linearly independent vectors called the primitive lattice vectors. The choice of such vectors is not unique. Standard examples of Bravais lattices include the triangular and rectangular

lattices.[1] However, not all crystals fall into this definition. For instance, the honeycomb lattice of graphene is not a Bravais lattice but is instead a so-called lattice with basis where for each lattice vector $\mathbf{R}$, we attribute the relative atom positions $\boldsymbol{\tau}_i$. If there is more than one atom per lattice vector, then the lattice cannot be described as a Bravais lattice. For graphene, there are two atoms per lattice vector and we may think of each set of atoms $\mathbf{R} + \boldsymbol{\tau}_i$ as a sublattice forming a triangular Bravais lattice.

Since the lattice is periodic, we may choose a suitable region of space, called a unit cell, to tile the plane. We define the primitive unit cell as the smallest such region and refer to it as the unit cell. Another notion of a unit cell is the Wigner-Seitz unit cell which is constructed by choosing a lattice vector and forming the shape enclosed by the intersections of the perpendicular bisectors to the nearest lattice vectors. This construction reveals the symmetries of the lattice and will be key to the discussion of the reciprocal lattice.
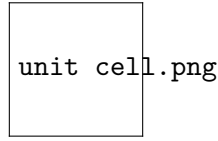

Figure 1: Non-primitive unit cell, primitive unit cell and Wigner-Seitz unit cell of a triangular lattice [**?**].

1.2 Reciprocal lattice

The periodicity of the lattice naturally leads us to techniques of Fourier analysis. We consider a function $F(\mathbf{r})$ periodic in $\mathbf{a}_j$[2] and take its Fourier transform

$$\tilde{F}(\mathbf{g}) = \int d^2\mathbf{r} F(\mathbf{r}) e^{i\mathbf{g}\cdot\mathbf{r}}. \tag{2}$$

For $\tilde{F}(\mathbf{g})$ to be nonzero, the integrand must be invariant under the translations $F(\mathbf{r}) \to F(\mathbf{r} + \mathbf{a}_j)$ and consequently $e^{i\mathbf{g}\cdot\mathbf{a}_j} = 1$ for all j. This fixes the form of $\mathbf{g}$

$$\mathbf{g} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2, \qquad n_1, n_2 \in \mathbb{Z} \tag{3}$$

where $\mathbf{b}_j$ are called primitive reciprocal vectors and are defined by the equation $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$.[3] As the name suggests, the primitive reciprocal vectors define the reciprocal lattice, a momentum space analog of the real space lattice. We define the Brillouin Zone (BZ) as the Wigner-Seitz unit cell of the reciprocal lattice and explore its properties in the next section.

1.3 Electrons in a lattice

Suppose we have a lattice with primitive lattice vectors $\mathbf{a}_j$ and length $N_j|\mathbf{a}_j|$ along the direction of $\mathbf{a}_j$. The total number of unit cells is then simply $N = N_1 N_2$. We impose periodic boundary conditions on the lattice which although does not correspond to the open boundary conditions of lattices in reality, we are mainly interested in the dynamics of electrons in the bulk and may neglect

---

[1]There are only five distinct types of Bravais lattices in 2 dimensions. [**?**]

[2]We use this terminology to state that $F(\mathbf{r} + \mathbf{a}_j) = F(\mathbf{r})$ for all j.

[3]δ_{ij} is the Kronecker delta defined as $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$.

boundary pathologies. We consider an electrostatic potential $U(\mathbf{r})$, periodic in $\mathbf{a_j}$, which is due to the ions in the lattice. The Hamiltonian describing an electron in this lattice is

$$H(\mathbf{r}) = \frac{-\hbar^2 \nabla^2}{2m} + U(\mathbf{r}). \tag{4}$$

We wish to understand the eigenstates of H, and we do so by utilizing its discrete translational symmetries. The proof is short but the result places strong restrictions on the form of its eigenstates. We first note that $\left[H(\mathbf{r}), T_{\mathbf{a_j}}\right] = 0$ where $T_{\mathbf{a_j}} = e^{-\mathbf{a_j} \cdot \nabla}$ is the operator corresponding to translation by $\mathbf{a}_j$. This result is basic by noting $\left[f(\nabla), T_{\mathbf{a}_j}\right] = 0$ for any function $f(\nabla)$, and $T_{\mathbf{a}_j}^\dagger U(\mathbf{r}) T_{\mathbf{a}_j} = U(\mathbf{r} + \mathbf{a}_j) = U(\mathbf{r})$ and therefore $\left[U(\mathbf{r}), T_{\mathbf{a}_j}\right] = 0$. We also note that the translation operators trivially commute with each other $\left[T_{\mathbf{a}_i}, T_{\mathbf{a}_j}\right] = 0$. Since H and the translation operators commute, we may construct a basis of functions that are simultaneously eigenstates of H and $T_{\mathbf{a}_j}$. We let $f(\mathbf{r})$ be an element of such a basis

$$T_{\mathbf{a}_j} f(\mathbf{r}) = \lambda_j f(\mathbf{r}), \qquad H(\mathbf{r}) f(\mathbf{r}) = E f(\mathbf{r}) \tag{5}$$

where $\lambda_j^{N_j} = 1$ by noting that the periodic boundary condition imposed on the lattice requires that $T_{\mathbf{a}_j}^{N_j} = \sigma_0$.[4] Therefore, λ_j must be a root of unity $\lambda_j = e^{2\pi i n_j / N_j}$, $n_j \in \mathbb{Z}$. We define the quasi-momentum $\mathbf{k} = \sum_{j=1}^2 n_j \mathbf{b}_j / N_j$ where $\mathbf{b}_j$ are the primitive reciprocal vectors satisfying $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij}$ and may then write $T_{\mathbf{a}_j} f(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{a}_j} f(\mathbf{r})$. Note that the eigenvalues $e^{i\mathbf{k} \cdot \mathbf{a}_j}$ of $T_{\mathbf{a}_j}$ are invariant under $\mathbf{k} \to \mathbf{k} + \mathbf{b}_j$. Therefore, equivalence in momentum space is equality modulo a reciprocal lattice vector and all unique momenta lie in the BZ, a set of size N. This equivalence also places boundary conditions on the BZ which makes its topology that of the torus $\mathbb{T}$.[5] This notion becomes useful when dealing with topological aspects of the BZ such as integration.

Returning to our problem, we can label the eigenstates of the translation operator by $\mathbf{k}$ and label that of the Hamiltonian by a band index n. In a lattice, the energy levels are discrete and referred to as bands, similar to those of a harmonic oscillator potential.[6] In general they are $\mathbf{k}$-dependent and define the band structure of the material, or the range of energies an electron at a given momentum is allowed to have. We have the simultaneous eigenstates

$$T_{\mathbf{a}_j} f_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{a}_j} f_{n,\mathbf{k}}(\mathbf{r}), \qquad H(\mathbf{r}) f_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}} f_{n,\mathbf{k}}(\mathbf{r}). \tag{6}$$

By definition of the translation operators, we have $T_{\mathbf{a}_j} f_{n,\mathbf{k}}(\mathbf{r}) = f_{n,\mathbf{k}}(\mathbf{r} + \mathbf{a}_j) = e^{i\mathbf{k} \cdot \mathbf{a}_j} f_{n,\mathbf{k}}(\mathbf{r})$, and therefore we may write

$$f_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k} \cdot \mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \qquad u_{n,\mathbf{k}}(\mathbf{r} + \mathbf{a}_j) = u_{n,\mathbf{k}}(\mathbf{r}), \qquad \int_\Omega d^2\mathbf{r} |u_{n,\mathbf{k}}(\mathbf{r})|^2 = 1 \tag{7}$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ is called the Bloch function and Ω denotes the unit cell. We see that the form of $f_{n,\mathbf{k}}(\mathbf{r})$ is that of a plane-wave modulated by a function periodic in the unit cell.

---

[4]We use the standard notation for Pauli matrices throughout this paper. We also include $\sigma_0 = \sigma_0$.

[5]In other words, there is a homeomorphism between the BZ and the torus which preserves all topological properties such as connectedness.

[6]Recall that in a harmonic oscillator potential, the energy levels are given by $E_n = \hbar\omega \left(n + \frac{1}{2}\right)$ where $n \in \mathbb{Z}_{\geq 0}$ and ω is the frequency of the oscillator.

1.4 Wannier orbitals

Since the eigenstates $f_{n,\mathbf{k}}(\mathbf{r})$ take the form of a plane wave, they are not localized, contradicting our intuition for localized atomic orbitals. To remedy this issue, we define the Wannier function of band n and lattice site $\mathbf{R}$ as the Fourier transform

$$W_{n,\mathbf{R}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} f_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} u_{n,\mathbf{k}}(\mathbf{r}) \tag{8}$$

where the summation is over the BZ. Note that $u_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r} - \mathbf{R})$ and therefore $W_{n,\mathbf{R}}(\mathbf{r})$ only depends on the difference $\mathbf{r} - \mathbf{R}$. The Wannier functions are orthogonal in both indices

$$\int d^2\mathbf{r} W^*_{n,\mathbf{R}}(\mathbf{r}) W_{n',\mathbf{R}'}(\mathbf{r}) = \delta_{n,n'}\delta_{\mathbf{R},\mathbf{R}'} \tag{9}$$

and correspond to localized atomic orbitals. They generally overlap but are modulated by a complex phase so that the overlap between them is integrated out on large scales. The localization of Wannier functions is an open problem which we briefly discuss in the appendix after developing the formalism of Chern numbers.[7]

1.5 Second quantization

We will now introduce (fermionic) second quantization which will help simplify notation when dealing with many-electron states. Recall that electrons are fermions which are defined by the anti-symmetric transformation of their many-particle composite state under permutation of particle labels,

$$\psi(\alpha_1,\ldots,\alpha_i,\ldots,\alpha_j,\ldots,\alpha_N) = -\psi(\alpha_1,\ldots,\alpha_j,\ldots,\alpha_i,\ldots,\alpha_N), \tag{10}$$

where ψ is the composite N-particle state and α_i is the collection of all properties of particle i. We neglect spin in this paper and focus on position $\alpha_i = \mathbf{r}_i$. We assume that $f_m(\mathbf{r})$ form an orthonormal basis of single-particle states. We construct many-particle states by first defining Fock states which serve as fundamental building blocks. We define the N-particle Fock state as a sum over permutations

$$\psi_{m_1,\ldots,m_N}(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_{\mathcal{P}} \text{sign}(\mathcal{P}) f_{m_1}(\mathbf{r}_{\mathcal{P}_1}) \cdots f_{m_N}(\mathbf{r}_{\mathcal{P}_N}) \tag{11}$$

where $\text{sign}(\mathcal{P})$ is the sign of the permutation and $\sqrt{N!}$ is the normalization factor. It is simple to see that this definition satisfies the requirement of anti-symmetry (10). However, this presentation of the N-particle state is clearly inconvenient and we are motivated to find a more simple language. We define the fermion creation (annihilation) operator $c^\dagger_m$ (c_m) of the single-particle state $f_m(\mathbf{r})$, $\langle \mathbf{r}| c^\dagger_m |0\rangle = f_m(\mathbf{r})$ where $|0\rangle$ is the vacuum or empty state. These fermion operators satisfy the anti-commutation relations

$$\left\{ c_m, c^\dagger_n \right\} = c_m c^\dagger_n + c^\dagger_n c_m = \delta_{mn}, \qquad \{c_m, c_n\} = \left\{ c^\dagger_m, c^\dagger_n \right\} = 0, \tag{12}$$

---

[7]Note that the Wannier function is not unique since it is defined in terms of the Bloch functions whose phase (gauge) can be transformed non-trivially by a continuous $\mathbf{k}$-dependent function. However, one can choose an optimal gauge of the Bloch functions to maximize the locally of the Wannier function and obtain the so-called maximally-localized symmetric Wannier function.

as required by anti-symmetry.[8] Note that the ordering of operators matters since they do not commute, and therefore we impose the ordering that $c_m^\dagger$ (c_m) with smaller (larger) m acts earlier on the state.[9] We can now write the N-particle Fock state (11) in this language as

$$|m_1, ..., m_N\rangle = \prod_{j=1}^{N} c_{m_j}^\dagger |0\rangle .$$

(13)

We also define the fermion real space (or field) creation (annihilation) operators at position $\mathbf{r}$, $\psi^\dagger(\mathbf{r})$ ($\psi(\mathbf{r})$) which satisfy the anti-commutation relations

$$\left\{\psi(\mathbf{r}), \psi^\dagger(\mathbf{r}')\right\} = \delta(\mathbf{r} - \mathbf{r}'), \qquad \left\{\psi(\mathbf{r}), \psi(\mathbf{r}')\right\} = \left\{\psi^\dagger(\mathbf{r}), \psi^\dagger(\mathbf{r}')\right\} = 0$$

(14)

and are related to the single-particle creation operators

$$c_m^\dagger = \int d^2\mathbf{r} f_m(\mathbf{r})\psi^\dagger(\mathbf{r}), \qquad \psi^\dagger(\mathbf{r}) = \sum_m f_m^*(\mathbf{r})c_m^\dagger$$

(15)

where N is the volume of the space. The real space operators act on the vacuum state $\psi^\dagger(\mathbf{r}) |0\rangle = |\mathbf{r}\rangle$, $\psi(\mathbf{r}) |0\rangle = 0$ and are used to formulate the real space basis

$$|\mathbf{r}_1, ..., \mathbf{r}_N\rangle = \prod_{j=1}^{N} \psi^\dagger(\mathbf{r}_j) |0\rangle .$$

(16)

In general, we can write any first-quantized operator $\hat{O}$ in second-quantized language

$$O = \int d^2\mathbf{r} \psi^\dagger(\mathbf{r})\hat{O}\psi(\mathbf{r}).$$

(17)

As a concrete example, consider the free-particle Hamiltonian $H = \hat{\mathbf{p}}^2/2m$ with single-particle basis $f_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}/\sqrt{N}$ where N denotes the volume of the space. The single-particle operators and field operators are simply

$$c_\mathbf{k}^\dagger = \frac{1}{\sqrt{N}} \int d^2\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}}\psi^\dagger(\mathbf{r}), \qquad \psi^\dagger(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}} a_\mathbf{k}^\dagger,$$

(18)

and the total momentum operator $\hat{\mathbf{p}} = -i\hbar\nabla$ can be written in second-quantized language

$$p = -i\hbar \int d^2\mathbf{r}\psi^\dagger(\mathbf{r})\nabla\psi(\mathbf{r}) = \hbar\sum_\mathbf{k} \mathbf{k} a_\mathbf{k}^\dagger a_\mathbf{k},$$

(19)

where we used (18) for the expanding the field operators. Of particular importance, we can rewrite Bloch functions in second quantization

$$|u_{n,\mathbf{k}}\rangle = c_{n,\mathbf{k}}^\dagger |0\rangle = \frac{1}{\sqrt{N}} \int d^2\mathbf{r} u_{n,\mathbf{k}}(\mathbf{r})\psi^\dagger(\mathbf{r}) |0\rangle$$

(20)

---

[8]The second equation is the Pauli-exclusion principle which states that a state can only be occupied by at most one electron.

[9]For instance, $-c_1^\dagger c_2^\dagger |0\rangle = c_2^\dagger c_1^\dagger |0\rangle = |1, 2\rangle$.

which can be thought of as the Bloch function of the second-quantized Hamiltonian.

1.6 Tight-binding model

Now that we can describe many-particle states in a simple language, we move on to describing the movement of electrons in a lattice. We assume the existence of localized Wannier orbitals and model electron movements as hopping between the orbitals. This model is known as the tight-binding model and can provide very powerful insight. We denote the first and second-quantized Hamiltonian as $\mathcal{H}$ and H respectively. We define the single-particle Wannier creation operators and corresponding states

$$|n,\mathbf{R}\rangle = c^\dagger_{n,\mathbf{R}}|0\rangle, \qquad c^\dagger_{n,\mathbf{R}} = \int d^2\mathbf{r}\, W_{n,\mathbf{R}}(\mathbf{r})\psi^\dagger(\mathbf{r}), \qquad \psi^\dagger(\mathbf{r}) = \sum_{n,\mathbf{R}} W^*_{n,\mathbf{R}}(\mathbf{r})c^\dagger_{n,\mathbf{R}}. \tag{21}$$

We can mathematically represent the action of an electron hopping from $|n',\mathbf{R}'\rangle$ to $|n,\mathbf{R}\rangle$ using the operator $tc^\dagger_{n,\mathbf{R}}c_{n',\mathbf{R}'}$ where t is called the hopping amplitude and assigns a strength to such a hopping. Quite literally, we are annihilating the former state and creating the latter. Assuming that the Hamiltonian can be expanded in this basis of operators, we define the tight-binding Hamiltonian

$$H = \int d^2\mathbf{r}\,\psi^\dagger(\mathbf{r})\mathcal{H}(\mathbf{r})\psi(\mathbf{r}) = \sum_{n,\mathbf{R}}\sum_{n',\mathbf{R}'} c^\dagger_{n,\mathbf{R}}H_{n,\mathbf{R};n',\mathbf{R}'}c_{n',\mathbf{R}'} \tag{22}$$

where

$$H_{n,\mathbf{R};n',\mathbf{R}'} = \langle n,\mathbf{R}|\,H\,|n',\mathbf{R}'\rangle = \int d^2\mathbf{r}\, W^*_{n,\mathbf{R}}(\mathbf{r})\mathcal{H}(\mathbf{r})W_{n',\mathbf{R}'}(\mathbf{r}) \tag{23}$$

is the hopping amplitude from orbital $|n',\mathbf{R}'\rangle$ to $|n,\mathbf{R}\rangle$. Since the Wannier function and Hamiltonian are invariant under translation by a lattice vector, the hopping amplitude is only a function of the relative distance between the orbitals $H_{n,\mathbf{R};n',\mathbf{R}'} = t_{nn';\mathbf{R}-\mathbf{R}'}$. The hopping amplitude usually decays exponentially with respect to relative distance since the Wannier orbitals are localized. Therefore, it is typical to only include nearest-neighbor hoppings, or set $t_{nn';\mathbf{R}-\mathbf{R}'} = 0$ if $\mathbf{R},\mathbf{R}'$ are not nearest-neighbors. Under this nearest-neighbor approximation, the tight-binding Hamiltonian can be written

$$H = \sum_{n,n'}\sum_{\langle\mathbf{R},\mathbf{R}'\rangle} t_{nn';\mathbf{R}-\mathbf{R}'}c^\dagger_{n,\mathbf{R}}c_{n',\mathbf{R}'} \tag{24}$$

where $\langle\mathbf{R},\mathbf{R}'\rangle$ denotes nearest-neighbors. Note that the tight-binding model completely neglects interactions between the electrons. One may also include $n_{n,\mathbf{R}}n_{n',\mathbf{R}}$ terms where $n_{n,\mathbf{R}} = c^\dagger_{n,\mathbf{R}}c_{n,\mathbf{R}}$ is called the number operator and counts the number of electrons in the corresponding Wannier orbital. This model which only couples number operators of the same Wannier orbital is called the Hubbard model. The extended Hubbard model couples number operators of different Wannier orbitals and can be written

$$H = \sum_{n,n'}\sum_{\langle\mathbf{R},\mathbf{R}'\rangle} (t_{nn';\mathbf{R}-\mathbf{R}'}c^\dagger_{n,\mathbf{R}}c_{n',\mathbf{R}'} + U_{n,\mathbf{R};n',\mathbf{R}'}n_{n,\mathbf{R}}n_{n',\mathbf{R}}) \tag{25}$$

where $U_{n,\mathbf{R};n',\mathbf{R}'}$ denotes some interaction amplitude between the corresponding Wannier orbitals. The band structure of the tight-binding model can be solved via matrix diagonalization while the Hubbard model requires more intricate computations. We will now apply a tight-binding model to graphene and derive its remarkable band structure.

6

2 Graphene

We have covered the basic theory of condensed matter physics which will allow us to derive the band structure of graphene. The objective of this section is to demonstrate that electrons near so-called Dirac points in the BZ behave as massless Dirac fermions. We follow [?, ?] for the tight-binding and continuum model derivations.

2.1 Dirac equation

We first introduce the Dirac equation, a relativistic reformulation of the Schrodinger equation. One is motivated to find a quantum description which agrees with the notion of Lorentz-invariance postulated in the theory of special relativity. It is clear that the time-dependent Schrodinger equation (in the absence of a potential)

$$-\frac{\hbar^2 \nabla^2}{2m} \psi(\mathbf{r}) = i\hbar \partial_t \psi(\mathbf{r}) \tag{26}$$

is not Lorentz-invariant since it contains a first-order time derivative but a second-order space derivative. In this derivation, we work in standard units ($\hbar = c = 1$) and we use Einstein summation notation[10]. One may try to build a relativistic Schrodinger equation by solving for solutions to the energy-momentum relation $p^\mu p_\mu - m^2 = 0$ where p^μ is the four-momentum $(E, \mathbf{p})$.[11] A clever technique is to factor this equation is by writing it as a sum of (generally non-abelian[12]) operators γ^μ

$$p^\mu p_\mu - m^2 = (\gamma^\mu p_\mu + m)(\gamma^\mu p_\mu - m) \tag{27}$$

and solving for the constraints. Having done so, we arrive at the Dirac equation $\gamma^\mu p_\mu - m = 0$ where γ^μ satisfy the anti-commutation relations $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ and $\eta^{\mu\nu} = \text{diag}(1, -1, -1, -1)$ is the Minkowski metric tensor [?]. Rewriting this in the position representation and tracking down the units, we have

$$(i\hbar\gamma^\mu \partial_\mu - mc)\psi(\mathbf{r}) = 0 \tag{28}$$

where ∂ is the four-gradient $(\frac{1}{c}\partial_t, -\nabla)$. The Dirac equation describes Dirac fermions[13] with mass m. One canonical representation of the operators γ^μ is the chiral representation

$$\gamma^0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \tag{29}$$

where σ_i are the Pauli matrices. In the case of a massless particle ($m = 0$), the Dirac equation reduces to $\gamma^\mu \partial_\mu \psi(\mathbf{r}) = 0$ which written in the chiral representation and having decoupled the off-diagonal terms becomes

$$\pm c(\boldsymbol{\sigma} \cdot \hat{\mathbf{p}})\psi_\pm(\mathbf{r}) = i\hbar\partial_t \psi_\pm(\mathbf{r}) \tag{30}$$

where $\psi(\mathbf{r}) = (\psi_+(\mathbf{r}), \psi_-(\mathbf{r}))$. It is clear that the Dirac equation admits both positive and negative energy solutions $E = \pm\sqrt{(pc)^2 + (mc^2)^2}$. We will soon see that graphene can be described by the

---

[10] Double indices, one raised and one lowered, denote summation $a^\mu b_\mu = \sum_\mu a^\mu b_\mu$. The index corresponds to a component of the vector. For instance, if $\mathbf{x} = (a, b, c)$ then $x^1 = a$, $x^2 = b$, $x^3 = c$.

[11] This is just the standard energy-momentum relation in special relativity $E^2 = (pc)^2 + (mc^2)^2$ written in terms of the four-momentum.

[12] We use the mathematical term non-abelian to refer to operators A, B which do not commute $[A, B] \neq 0$.

[13] A Dirac fermion is a spin-1/2 fermion which is its own anti-particle. This includes electrons but more generally all fermions in the standard model [?].

massless Dirac equation and that these positive and negative energy solutions correspond to the conduction and valence bands respectively.

2.2 Atomic structure of graphene

Before we derive the tight-binding model, we will first discuss the atomic structure of graphene. As mentioned before, graphene is a honeycomb lattice of carbon atoms each of which have six electrons. The two natural allotropes of carbon are graphite and diamond, and its synthetic allotropes include fullerene (1985), carbon nanotubes (1991), and graphene (2004). The six electrons in carbon are distributed as $1s^2 2s^2 2p^2$.[14] The 2s, $2p_x$, and $2p_y$ orbitals hybridize[15] to form three in-plane $2sp^2$ orbitals each of which form a σ bond with a neighboring carbon atom [?]. The σ bonds form angles of $2\pi/3$ radians with each other and are responsible for the honeycomb structure of graphene and its immense structural stability.[16] This leaves the p_z orbital which is transverse to the graphene plane and forms π bonds with neighboring p_z orbitals. Each p_z orbital contributes one electron which can hop between neighboring atoms along the π bonds, enabling electrical conduction. In the theory of band structure, bands near the Fermi level are divided into valence and conduction bands. Electrons in the valence bands can be excited into the conduction bands where they are free to hop between lattice sites, conducting electricity.
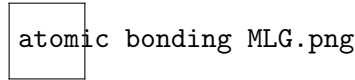

Figure 2: Transverse π bonds and in-plane σ bonds in graphene which form the honeycomb lattice structure [?].

2.3 Tight-binding model

We are ready to compute the tight-binding model of graphene. Graphene is a honeycomb lattice with two atoms in the unit cell. The honeycomb lattice can be considered as two triangular Bravais lattices translated with respect to one another. Therefore, we may attribute a sublattice index α to each of the two atoms in the unit cell. We first describe the primitive lattice vectors

$$\mathbf{a}_1 = \boldsymbol{\tau}_3 - \boldsymbol{\tau}_2 = \sqrt{3}a_c(1,0), \qquad \mathbf{a}_2 = \boldsymbol{\tau}_1 - \boldsymbol{\tau}_2 = \frac{\sqrt{3}}{2}a_c(1,\sqrt{3}) \tag{31}$$

where

$$\boldsymbol{\tau}_1 = a_c(0,1), \qquad \boldsymbol{\tau}_2 = -\frac{a_c}{2}(\sqrt{3},1), \qquad \boldsymbol{\tau}_3 = \frac{a_c}{2}(\sqrt{3},-1) \tag{32}$$

are the nearest-neighbor hopping vectors and $a_c \approx 1.42\text{Å}$ is the carbon-carbon spacing. We may then describe the lattice vectors (corresponding to the positions of atoms)

$$\mathbf{R}_\alpha = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + \boldsymbol{\tau}_\alpha, \qquad n_1, n_2 \in \mathbb{Z} \tag{33}$$

---

[14]We are using standard orbital notation ab^c where a corresponds to the energy index, b is the orbital angular momentum, and c is the number of electrons occupying the orbital. The p orbitals are made up of p_x,p_y, and p_z orbitals where the Cartesian index corresponds to the direction the orbital points. In this discussion, we take the graphene sheet to lay in the xy plane which we refer to as in-plane.

[15]sp^2 hybridization involves the mixing of one s and two p orbitals in which one electron in the s orbital is promoted to one of the 2p orbitals.

[16]Diamond uses all four of carbon's valence electrons to form σ bonds although surprisingly it is weaker than graphene. It is also a poor electrical conductor since it does not form p_z orbitals.

where we choose $\boldsymbol{\tau}_A = \mathbf{0}, \boldsymbol{\tau}_B = -\boldsymbol{\tau}_3$. We then compute the primitive reciprocal vectors

$$\mathbf{b}_1 = \frac{2\pi}{3a_c}(\sqrt{3}, -1), \qquad \mathbf{b}_2 = \frac{4\pi}{3a_c}(0, 1). \tag{34}$$

We can write the nearest-neighbor tight-binding Hamiltonian

$$H = -t \sum_{\langle \mathbf{R}_A^i, \mathbf{R}_B^j \rangle} (c_{\mathbf{R}_A^i}^\dagger c_{\mathbf{R}_B^j} + h.c.) + m \sum_\alpha (-1)^\alpha \sum_{\mathbf{R}_\alpha^i} c_{\mathbf{R}_\alpha^i}^\dagger c_{\mathbf{R}_\alpha^i} \tag{35}$$

where $h.c.$ denotes the Hermitian-conjugate of the prior term $(a + h.c. = a + a^\dagger)$. m is called the sublattice potential and assigns an energy $\pm m$ to sublattice A and B respectively, and t is the hopping amplitude. We now transform this Hamiltonian into momentum space using the relation

$$c_{\mathbf{R}_\alpha^i}^\dagger = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_\alpha^i} c_{\mathbf{k},\alpha}^\dagger \tag{36}$$

where N denotes the the number of unit cells. The $\alpha\alpha$ hopping term can be computed

$$m \sum_{\mathbf{R}} c_{\mathbf{R}_\alpha}^\dagger c_{\mathbf{R}_\alpha} = \frac{m}{N} \sum_{\mathbf{k},\mathbf{k}'} \sum_{\mathbf{R}_\alpha} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}_\alpha} c_{\mathbf{k},\alpha}^\dagger c_{\mathbf{k},\alpha} = \pm m \sum_{\mathbf{k}} c_{\mathbf{k},\alpha}^\dagger c_{\mathbf{k},\alpha} \tag{37}$$

and a similar computation shows that the AB term becomes

$$-t \sum_{\langle \mathbf{R}_A, \mathbf{R}_B \rangle} c_{\mathbf{R}_A}^\dagger c_{\mathbf{R}_B} = -t \sum_{\mathbf{k}} \sum_{i=1}^3 e^{-i\mathbf{k}\cdot\boldsymbol{\tau}_i} c_{\mathbf{k},A}^\dagger c_{\mathbf{k},B} \tag{38}$$

and the BA term is simply its Hermitian-conjugate. The momentum space tight-binding Hamiltonian therefore can be written as

$$H(\mathbf{k}) = \begin{pmatrix} m & -tf(\mathbf{k}) \\ -tf^*(\mathbf{k}) & -m \end{pmatrix}, \qquad f(\mathbf{k}) = \sum_{j=1}^3 e^{-i\mathbf{k}\cdot\boldsymbol{\tau}_j} \tag{39}$$

where $*$ denotes complex conjugation and we are using the basis $(c_{\mathbf{k},A}^\dagger, c_{\mathbf{k},B}^\dagger)^T$. We can easily solve for the eigenvalues of $H(\mathbf{k})$ to obtain the band energies

$$E_{\pm 1, \mathbf{k}} = \pm t\sqrt{|f(\mathbf{k})|^2 + m^2} \tag{40}$$

where ± 1 corresponds to the conduction and valence bands respectively. We see that increasing the sublattice potential widens the band gap and for $m \neq 0$, graphene is an insulator while for $m = 0$, it is a semimetal.[17]

2.4 Continuum model

We now turn off the sublattice potential ($m = 0$) and note that the valence and conduction bands intersect at reciprocal lattice sites, forming so-called Dirac cones. Recall that momenta are equivalent modulo reciprocal lattice vectors, and therefore we restrict our attention to the BZ. There

---

[17]If the bands were gapped then we would have a (bulk) insulator, and if they overlapped, we would have a metal. Their intersection corresponds to a semimetal.

are six corners of the BZ, with half corresponding to each "reciprocal sublattice". To distinguish between the sublattice of the real space and that of the reciprocal space, we use the term valley, denoted $\eta = \pm 1$, to refer to the latter. There are two valleys in the BZ and each can be represented by a lattice vector, which as is convention, we choose the points

$$\mathbf{K} = \frac{4\pi}{3\sqrt{3}a_c}(1, 0), \qquad \mathbf{K}' = -\mathbf{K} \tag{41}$$

as representatives of valley $\eta = \pm 1$ respectively. These sites are called Dirac points and are precisely where the valence and conduction bands intersect. We now formulate a low-energy continuum model of graphene by expanding the Hamiltonian about these Dirac points. We take the momenta $\mathbf{k}$ to be a continuous set in the BZ which eliminates the notion of a lattice. For each valley, we Taylor expand the off-diagonal entries of the tight-binding Hamiltonian (104) to first-order

$$f(\mathbf{k} + \mathbf{K}) \approx -\frac{3a_c}{2}(k_x + ik_y), \qquad f(\mathbf{k} + \mathbf{K}') \approx -\frac{3a_c}{2}(k_x - ik_y) \tag{42}$$

where $|\mathbf{k}| \ll |\mathbf{K}| = |\mathbf{K}'|$ and $\mathbf{k}$ is measured from the Dirac points. The continuum model Hamiltonian is therefore

$$h(\mathbf{k}) = v_F(\tau_z \sigma_x k_x + \sigma_y k_y) \tag{43}$$

where τ_z is the σ_z Pauli matrix acting on the valley subspace ($\tau_z = \pm 1$ for $\eta = \pm 1$ respectively), and $v_F = \frac{3}{2}ta_c$ is the Fermi velocity. The conduction and valence bands are given by $E_\pm = \pm v_F |\mathbf{k}|$ respectively and form so-called Dirac cones when energy is plotted against momentum. The linearity of this band dispersion reveals that electrons near the Dirac points of graphene behave as though they were relativistic and massless, but there is an even more direct relation. Recall the massless Dirac equation (with velocity v_F instead of c and denoting momentum as $\mathbf{k}$ instead of $\mathbf{p}$)

$$H(\mathbf{k}) = \begin{pmatrix} 0 & v_F \mathbf{k} \cdot \boldsymbol{\sigma} \\ v_F \mathbf{k} \cdot \boldsymbol{\sigma} & 0 \end{pmatrix} \tag{44}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y)$. If we perform an appropriate unitary transformation[18] on this Hamiltonian, we have

$$U^\dagger H(\mathbf{k}) U = \begin{pmatrix} v_F \mathbf{k} \cdot \boldsymbol{\sigma}^* & 0 \\ 0 & v_F \mathbf{k} \cdot \boldsymbol{\sigma} \end{pmatrix} \tag{45}$$

where $\boldsymbol{\sigma}^* = (-\sigma_x, \sigma_y)$. We have precisely the continuum model Hamiltonian of graphene (43). Note that in the case of the Dirac equation, the two-component spinor basis corresponds to the spin degree of freedom of the fermion, but we have neglected spin in our derivation of (43). At first, it seems to be a mystery how the graphene basis fits into the picture of spin. Originally, the tight-binding Hamiltonian (116) was written in the sublattice basis but we see that it is treated on equal footing with spin. Therefore the sublattice degree of freedom may be interpreted as giving rise to a so-called pseudospin. In this context, the pseudospin is governed by Dirac equations of opposite helicity.[19]

---

[18]We omit the exact expression of U for brevity.

[19]Helicity is concerned with the relationship between spin and momentum. The two Dirac equations involving $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^*$ are said to belong to right-handed and left-handed helicity respectively.
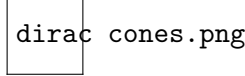
Figure 3: Band structure of graphene illustrating the intersection of the valence and conduction band forming Dirac cones. [**?**].

3 Twisted Bilayer Graphene

Now that we have developed a theory of graphene, we switch our attention to twisted bilayer graphene. The aim of this section is to derive the Bistritzer-MacDonald continuum model and study its consequences.

3.1 Moire lattice

We begin by providing an introduction to the lattice structure of twisted bilayer graphene (TBG). TBG is comprised of two parallel layers of graphene separated by a distance $d_\perp \approx 3.34$Å with one layer rotated relative to the other by an angle θ. For general angle, TBG does not form a periodic structure, and therefore we cannot apply Bloch's Theorem to understand the eigenstates of the Hamiltonian. However, for an infinite, discrete set commensurate angles, a periodic structure known as the moire pattern is formed. The moire lattice of TBG is a honeycomb lattice with unit cell usually much larger than that of graphene.

There are two typical stacking arrangements for untwisted bilayer graphene ($\theta = 0$) that we must keep in mind. AA stacking corresponds with aligned sublattice sites: A(B) points of one layer on top of A(B) points of the other. AB or Bernal stacking corresponds with the lattice sites of one layer on top of the centers of the honeycombs of the other layer and can be achieved by shifting one layer by a hopping vector $\boldsymbol{\tau}$. The lattice sites of the two layers are related

$$\mathbf{R}' = \mathbf{M}_\theta(\mathbf{R} - \boldsymbol{\tau}) + \mathbf{d}, \qquad \mathbf{R}' = \mathbf{M}_\theta \mathbf{R} + \mathbf{d} \tag{46}$$

for AB and AA stacking respectively. We denote $\mathbf{R}'$ as lattice vectors of layer $l = +1$ and $\mathbf{R}$ as that of layer $l = -1$. $\mathbf{M}_\theta$ is the rotation matrix corresponding to a counterclockwise rotation by θ, and $\mathbf{d}$ represents a translation of the $l = +1$ layer after rotating. It can be shown that smoothly varying $\mathbf{d}$ only translates the moire pattern and therefore, we set $\mathbf{d} = 0$ as it does not affect the calculations. However, note that if $\theta = 0$ then the lattice structure varies non-trivially with $\mathbf{d}$. We
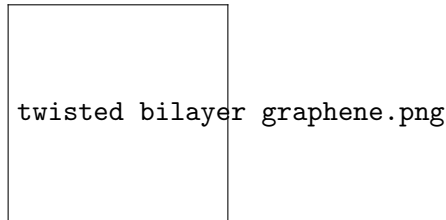

Figure 4: AA stacking, AB stacking with a small relative twist, AB stacking, and the moire pattern formed by twisted bilayer graphene. The light regions of the moire pattern correspond to AA stacking [**?**].

can solve for the conditions of periodicity and compute the set of commensurate angles

$$\theta = 2\tan^{-1}\left(\frac{\sqrt{3}}{1+2n}\right), \qquad n \in \mathbb{Z}_{\geq 1} \tag{47}$$

where $n(n+1)+1$ is the number of graphene unit cells of a single layer in the moire unit cell. Unless stated otherwise, we now assume a commensurate angle θ. We also assume the coordinate convention of rotating one layer by $\theta/2$ ($l = +1$) and the other by $-\theta/2$ ($l = -1$). The moire lattice has primitive lattice vectors and primitive reciprocal vectors [?]

$$\mathbf{a}_{1(2)} = \frac{\sqrt{3}a_0}{4\sin(\theta/2)}(\sqrt{3}, \mp 1), \qquad \mathbf{b}_{1(2)}^m = \frac{4\pi\sin(\theta/2)}{3a_0}\left(1, \mp\sqrt{3}\right) \tag{48}$$

respectively and we may describe the moire reciprocal lattice vectors

$$\mathbf{Q}_\pm = n_1\mathbf{b}_1^m + n_2\mathbf{b}_2^m + \mathbf{q}_{1(2)}^m, \qquad n_1, n_2 \in \mathbb{Z} \tag{49}$$

where

$$\mathbf{q}_{1(2)}^m = k_\theta\left(\pm\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \quad \mathbf{q}_3^m = -(\mathbf{q}_1^m + \mathbf{q}_2^m), \quad k_\theta = \frac{8\pi}{3a_c}\sin\left(\frac{\theta}{2}\right) = |\mathbf{q}_i^m| \tag{50}$$

are the nearest-neighbor hopping vectors in the moire reciprocal lattice. The $\pm$ in $\mathbf{Q}_\pm$ refers to the "sublattice" degree of freedom in the honeycomb moire reciprocal lattice. We will expand on this notion later.

3.2 Hopping

Now that we have an explicit construction of the moire lattice, we turn our attention to developing a continuum model for TBG. If we consider only intralayer coupling, then the bilayer Hamiltonian is simply

$$\mathcal{H}(\mathbf{k}) = |-1\rangle\, h_{-\theta/2}(\mathbf{k})\,\langle -1| + |+1\rangle\, h_{\theta/2}(\mathbf{k})\,\langle +1| \tag{51}$$

where we denote the projection onto layer i as $|i\rangle\langle i|$ and $h_\Phi(\mathbf{k})$ is the continuum model Hamiltonian of graphene (43) rotated by an angle Φ

$$h_\Phi(\mathbf{k}) = v_F k(\tau_z \cos(\Phi_k - \Phi)\sigma_x + \sin(\Phi_k - \Phi)\sigma_y) \tag{52}$$

where Φ_k is the angle that $\mathbf{k}$ makes with the k_x axis. This system is not particularly interesting since we just have two copies of graphene with no interactions. Realistically, electrons may hop between the layers, and we now focus on computing this interlayer hopping amplitude. We assume that the graphene valleys $\eta = \pm 1$ are sufficiently far from each other so that we can neglect inter-valley hoppings and can treat the valleys independently. We also assume an AA stacking arrangement with the origin at a honeycomb lattice point. The momentum Bloch basis for each layer corresponding to sublattice α and momentum $\mathbf{p}$ are

$$\left|\Psi_{\mathbf{p}\alpha}^{(-1)}\right\rangle = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}_\alpha}e^{i\mathbf{p}\cdot\mathbf{R}_\alpha}|\mathbf{R}_\alpha\rangle, \qquad \left|\Psi_{\mathbf{p}\alpha}^{(+1)}\right\rangle = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}_\alpha}e^{i\mathbf{p}\cdot\mathbf{R}_\alpha}|\mathbf{R}_\alpha'\rangle \tag{53}$$

where N is the number of graphene unit cells in each layer, which we take to be equal. These expressions are the inverse of (36) written in first-quantization. We now expand the interlayer

hopping Hamiltonian $\mathcal{H}$ in this basis

$$\mathcal{H}_{\mathbf{kp'}}^{\alpha\beta} = \left\langle \Psi_{\mathbf{k}\alpha}^{(-1)} \right| \mathcal{H} \left| \Psi_{\mathbf{p'}\beta}^{(+1)} \right\rangle \tag{54}$$

which describes the process of hopping from sublattice β and momentum $\mathbf{p'} = \mathbf{M}_\theta \mathbf{p}$ in layer $l = +1$ to sublattice α and momentum $\mathbf{k}$ in layer $l = -1$. The momenta are measured relative to the center of the graphene BZ. We can compute hoppings from layer -1 to layer $+1$ using the complex conjugate of this expression. Throughout this derivation, primed coordinates are rotated by $\mathbf{M}_\theta$ and unprimed coordinates are not rotated. We invoke the two-center approximation for the interlayer hopping amplitude

$$\langle \mathbf{R}_\alpha | \mathcal{H} | \mathbf{R}'_\beta \rangle = t_{\alpha\beta}(\mathbf{R}_\alpha - \mathbf{R}'_\beta) \tag{55}$$

where $\mathbf{R}_\alpha, \mathbf{R}'_\beta$ are graphene lattice vectors, and we note that the hopping amplitude may be sublattice-dependent. Following the derivation in [?], we now rewrite the hopping amplitude in momentum space with the intention of later expanding it about the Dirac points

$$\begin{aligned}
\mathcal{H}_{\mathbf{kp'}}^{\alpha\beta} &= \frac{1}{N} \sum_{\mathbf{R}_\alpha, \mathbf{R}'_\beta} t_{\alpha\beta}(\mathbf{R}_\alpha - \mathbf{R}'_\beta) e^{i(\mathbf{p}\cdot\mathbf{R}'_\beta - \mathbf{k}\cdot\mathbf{R}_\alpha)} \\
&= \frac{1}{N} \sum_{\mathbf{R}_\alpha, \mathbf{R}'_\beta} \left(\frac{1}{N} \sum_{\mathbf{q}} \frac{t_{\mathbf{q}}^{\alpha\beta}}{\Omega} e^{i\mathbf{q}\cdot(\mathbf{R}'_\beta - \mathbf{R}_\alpha)} \right) e^{i(\mathbf{p}\cdot\mathbf{R}'_\beta - \mathbf{k}\cdot\mathbf{R}_\alpha)}
\end{aligned} \tag{56}$$

where we rewrote the hopping amplitude using its Fourier transformation

$$t_{\alpha\beta}(\mathbf{R}_\alpha - \mathbf{R}'_\beta) = \frac{1}{N} \sum_{\mathbf{q}} t_{\alpha\beta}(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}'_\beta - \mathbf{R}_\alpha)}, \qquad t_{\alpha\beta}(\mathbf{q}) = \frac{t_{\mathbf{q}}^{\alpha\beta}}{\Omega}. \tag{57}$$

The above summation is performed over the moire BZ and $\Omega = \sqrt{3}a_c^2/2$ is the area of the moire unit cell. Summations involving $\mathbf{R}$ are done over all graphene lattice vectors. Numerical computations of π-band tight-binding models suggest that

$$\frac{t_{\mathbf{q}}^{\alpha\beta}}{\Omega} = t_0 \exp(-\gamma_{\alpha\beta}(|\mathbf{q}|d_\perp)^\mu), \tag{58}$$

where $\gamma_{AB} = 0.13, \mu = 1.25$, and $t_0 = 2$ eV. We assume these parameters in our band structure computations. We later define the intra-sublattice hopping amplitudes $t_{\mathbf{q}}^{\alpha\alpha}$ in terms of $t_{\mathbf{q}}^{AB}$. The expression (56) can be rearranged

$$\frac{1}{N^2} \sum_{\mathbf{R}_\alpha, \mathbf{R}'_\beta} \sum_{\mathbf{q}} \frac{t_{\mathbf{q}}^{\alpha\beta}}{\Omega} e^{i(\mathbf{p}+\mathbf{q})\cdot\mathbf{R}'_\beta} e^{-i(\mathbf{k}+\mathbf{q})\cdot\mathbf{R}_\alpha} = \frac{1}{N} \sum_{\mathbf{G}'_2} \frac{t_{\mathbf{G}'_2 - \mathbf{p}}^{\alpha\beta}}{\Omega} e^{i\mathbf{G}'_2 \cdot \boldsymbol{\tau}'_\beta} \sum_{\mathbf{R}_\alpha} e^{-i(\mathbf{k}-\mathbf{p}+\mathbf{G}'_2)\cdot\mathbf{R}_\alpha} \tag{59}$$

where on the right-hand-side we used the Poisson summation formula for the sum over $\mathbf{R}'_\beta$

$$\sum_{\mathbf{R}'_\beta} e^{i\mathbf{R}'_\beta \cdot(\mathbf{p}+\mathbf{q})} = e^{i\boldsymbol{\tau}'_\beta \cdot(\mathbf{p}+\mathbf{q})} \sum_{\mathbf{R}'} e^{i\mathbf{R}'\cdot(\mathbf{p}+\mathbf{q})} = N e^{i\boldsymbol{\tau}_\beta \cdot(\mathbf{p}+\mathbf{q})} \sum_{\mathbf{G}'_2} \delta_{\mathbf{p}+\mathbf{q}, \mathbf{G}'_2} \tag{60}$$

and enforced the delta function by setting $\mathbf{q} = \mathbf{G}'_2 - \mathbf{p}$. We denote graphene reciprocal lattice vectors by $\mathbf{G}$. Summations involving $\mathbf{G}$ are done over all reciprocal lattice vectors. Repeating for

$\mathbf{R}_\alpha$, we have

$$\sum_{\mathbf{G}_1,\mathbf{G}_2'} \frac{t_{\mathbf{k}+\mathbf{G}_1}^{\alpha\beta}}{\Omega} e^{i\mathbf{G}_2'\cdot\boldsymbol{\tau}_\beta'} e^{-i(\mathbf{k}-\mathbf{p}+\mathbf{G}_2')\cdot\boldsymbol{\tau}_\alpha} \delta_{\mathbf{p}+\mathbf{G}_2',\mathbf{k}+\mathbf{G}_1} \tag{61}$$

where we used the Poisson summation formula

$$\sum_{\mathbf{R}_\alpha} e^{i(\mathbf{p}-\mathbf{k}-\mathbf{G}_2')\cdot\mathbf{R}_\alpha} = N e^{i\boldsymbol{\tau}_\alpha\cdot(\mathbf{p}-\mathbf{k}-\mathbf{G}_2')} \sum_{\mathbf{G}_1} \delta_{\mathbf{p}-\mathbf{k}-\mathbf{G}_2',\mathbf{G}_1} \tag{62}$$

and set $\mathbf{G}_2' = -\mathbf{G}_2'$ for symmetry. We finally arrive at the interlayer hopping matrix element

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} = \sum_{\mathbf{G}_1,\mathbf{G}_2'} \frac{t_{\mathbf{k}+\mathbf{G}_1}^{\alpha\beta}}{\Omega} e^{i(\mathbf{G}_1\cdot\boldsymbol{\tau}_\alpha-\mathbf{G}_2'\cdot\boldsymbol{\tau}_\beta')} \delta_{\mathbf{k}+\mathbf{G}_1,\mathbf{p}+\mathbf{G}_2'}. \tag{63}$$

Another convention that we will now use to derive the BM Hamiltonian uses AB Bernal stacking for which a similar derivation shows

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} = \sum_{\mathbf{G}_1,\mathbf{G}_2'} \frac{t_{\mathbf{k}+\mathbf{G}_1}^{\alpha\beta}}{\Omega} e^{i(\mathbf{G}_1\cdot\boldsymbol{\tau}_\alpha-\mathbf{G}_2'\cdot(\boldsymbol{\tau}_\beta'-\boldsymbol{\tau}))} \delta_{\mathbf{k}+\mathbf{G}_1,\mathbf{p}'+\mathbf{G}_2'} \tag{64}$$

where $\boldsymbol{\tau}_A = \mathbf{0}, \boldsymbol{\tau}_B = \boldsymbol{\tau}$, and $\boldsymbol{\tau}$ is the vector describing the AB stacking arrangement (any of the graphene hopping vectors).

3.3 Bistritzer-MacDonald continuum model

We wish to expand (64) for $\mathbf{k}, \mathbf{p}'$ near the Dirac points to understand how the Dirac cones are altered with the introduction of interlayer hopping. We now measure $\mathbf{k}, \mathbf{p}'$ from the Dirac points instead of the center of the BZ. The infinite double sum may seem unwieldy at first glance, but since the hopping amplitude $t_{\mathbf{k}}^{\alpha\beta}$ decays exponentially with respect to $|\mathbf{k}|$, a reasonable approximation would be to truncate the sum with $|\mathbf{G}_1 + \mathbf{k} - \mathbf{K}|$ sufficiently small. Therefore, we set $\mathbf{G}_1 = \mathbf{K}$ which then requires $\mathbf{G}_2 = \mathbf{G}_1$ since the Kronecker delta requires $\mathbf{k} - \mathbf{p}' = \mathbf{G}_2' - \mathbf{G}_1$ and we are considering $\mathbf{k}, \mathbf{p}'$ small. This implies that $\mathbf{k} - \mathbf{p}' = \mathbf{q}_j^m$ where $\mathbf{q}_j^m$ are the moire hopping vectors defined at the beginning of this chapter. By considering only these terms, we have

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} \approx \sum_{j=1}^{3} T_j^{\alpha\beta} \delta_{\mathbf{k},\mathbf{p}'+\mathbf{q}_j^m}, \qquad T_j^{\alpha\beta} = \frac{t_{\mathbf{K}}^{\alpha\beta}}{\Omega} e^{i\mathbf{K}\cdot(\boldsymbol{\tau}_\alpha-\boldsymbol{\tau}_\beta+\boldsymbol{\tau})}. \tag{65}$$

where we approximated $t_{\mathbf{k}+\mathbf{K}}^{\alpha\beta} \approx t_{\mathbf{K}}^{\alpha\beta}$. The T_j matrices have explicit form

$$T_1 = \begin{pmatrix} w_0 & w_1 \\ w_1 & w_0 \end{pmatrix}, \qquad T_2 = \begin{pmatrix} w_0 e^{-i\phi} & w_1 \\ w_1 e^{i\phi} & w_0 e^{-i\phi} \end{pmatrix}, \qquad T_3 = \begin{pmatrix} w_0 e^{i\phi} & w_1 \\ w_1 e^{-i\phi} & w_0 e^{i\phi} \end{pmatrix}, \tag{66}$$

where $\phi = 2\pi/3$ and we define the hopping strengths $w_0 = w_{AA} = w_{BB} = t_{\mathbf{K}}^{AA}/\Omega$, $w_1 = w_{AB} = w_{BA} = t_{\mathbf{K}}^{AB}/\Omega$.[20] In experiment, interactions between the layers corrugates twisted bilayer graphene unevenly for AA and AB stacking regions. The atoms are displaced transverse to the graphene plane and this displacement is greater at the AA regions than the AB regions [?]. Therefore,

---

[20]We assume there is no sublattice potential which distinguishes between the two sublattices.

interlayer hopping is weaker at the AA regions since the atoms are further apart, and experiments show typical values of $w_0/w_1 \approx 0.7$ [?, ?]. Note that $w_0/w_1 = 1$ corresponds to zero corrugation, and therefore the regions have equal interlayer hopping amplitude, while $w_0/w_1 = 0$ corresponds to corrugation that completely inhibits hopping in the AA regions. We will soon study these two limits with the last chapter dedicated to the so-called chiral limit $w_0/w_1 = 0$.

Combining interlayer hopping (65) with intralayer hopping (52), we have the BM continuum model

$$\mathcal{H}_{BM}(\mathbf{k}) = \begin{pmatrix} h_{\theta/2}(\mathbf{k}) & T_1 & T_2 & T_3 \\ T_1^\dagger & h_{-\theta/2}(\mathbf{k}+\mathbf{q_1}) & 0 & 0 \\ T_2^\dagger & 0 & h_{-\theta/2}(\mathbf{k}+\mathbf{q_2}) & 0 \\ T_3^\dagger & 0 & 0 & h_{-\theta/2}(\mathbf{k}+\mathbf{q_3}) \end{pmatrix} \tag{67}$$

where $\mathbf{k}$ takes values in the moire BZ and acts on the four two-component spinors $\Psi = (\psi_0, \psi_1, \psi_2, \psi_3)$ [?, ?]. ψ_0 is at momentum $\mathbf{k}$ in layer $l = +1$ and ψ_j are at momentum $\mathbf{k}+\mathbf{q}_j^m$ in $l = -1$. The diagonal terms correspond to intralayer hopping, and the off-diagonal terms to interlayer hopping. This expression (67) can be used as an expansion about either valley by simply using the appropriate form of the graphene Hamiltonian h^η. Note that although we assumed a commensurate angle in the last section, the continuum model above holds for non-commensurate angles as well since there is no long a notion of the moire lattice. We may write $\mathcal{H}_{BM}$ in real space for $\eta = 1$ by taking the Fourier transform

$$\mathcal{H}_{BM}(\mathbf{r}) = \begin{pmatrix} -i\hbar v_F \boldsymbol{\sigma}_{\theta/2} \cdot \nabla & T(\mathbf{r}) \\ T^\dagger(\mathbf{r}) & -i\hbar v_F \boldsymbol{\sigma}_{-\theta/2} \cdot \nabla \end{pmatrix}, \qquad T^{\alpha\beta}(\mathbf{r}) = \sum_{j=1}^{3} e^{-i\mathbf{q}_j^m \cdot \mathbf{r}} T_j^{\alpha\beta} \tag{68}$$

where $\boldsymbol{\sigma}_\phi = e^{-\phi\sigma_z/4}(\sigma_x, \sigma_y)e^{\phi\sigma_z/4}$ [?]. $\mathcal{H}_{BM}(\mathbf{r})$ acts on the spinor $\Phi = (\psi_1, \chi_1, \psi_2, \chi_2)$ where the index denotes the layer and ψ, χ correspond to sublattice A, B respectively. We will come back to (68) in the last chapter when we solve for the wave functions of the flat bands in the chiral limit $(w_0/w_1 = 0)$.

3.4   Derivation of first magic angle

We now assume $w_0 = w_1 = w$ and prove that the Fermi velocity vanishes, and therefore we have exactly flat (dispersionless) bands in the low-energy limit. We perform the derivation below with $\eta = 1$ although it is simple to also work out for the other valley $\eta = -1$. In the limit of small twist angle θ, we may approximate $h_{\pm\theta/2}(\mathbf{k}+\mathbf{q}_i^m) \approx h(\mathbf{k}+\mathbf{q}_i^m) = v_F\boldsymbol{\sigma} \cdot (\mathbf{k}+\mathbf{q}_i^m)$ which we denote h_i. We first neglect interlayer coupling $w = 0$, yielding

$$\mathcal{H}_{BM}^0(\mathbf{k}) = \text{diag}(h_0, h_1, h_2, h_3) \tag{69}$$

where we include $h_0 = v_F\boldsymbol{\sigma} \cdot \mathbf{k}$. Since $\mathcal{H}_{BM}^0$ is in block-diagonal form, its eigenvalues are those of h_i which include $\pm v_F|\mathbf{k}|$ due to h_0. Consequently, there is a two-fold degeneracy at zero energy when $\mathbf{k} = \mathbf{0}$. We will now see that this degenerate subspace is robust under turning on w by constructing the zero-energy eigenstates explicitly. We are investigating solutions of the equation $\mathcal{H}_{BM}\Phi = \mathbf{0}$ which is a system of four equations involving two-component spinors ψ_j. Three of the equations have solutions

$$\psi_j = -wh_j^{-1}T_j^\dagger\psi_0 = \mathbf{0}, \qquad j = 1, 2, 3. \tag{70}$$

We apply T_j to the left-hand side and sum over j to obtain

$$\sum_{j=1}^{3} T_j \psi_j = \sum_{j=1}^{3} T_j(-h_j^{-1} T_j^\dagger \psi_0) = -\sum_{j=1}^{3} (T_j h_j^{-1} T_j^\dagger) \psi_0 = \mathbf{0} \tag{71}$$

where we used

$$\sum_{j=1}^{3} T_j h_j^{-1} T_j^\dagger = 0 \tag{72}$$

which can be shown by direct computation. Therefore, the two-fold degenerate zero-energy eigenvalues of $\mathcal{H}_{BM}$ are sustained for $w \neq 0$. We will now consider $\mathbf{k} \neq \mathbf{0}$ and investigate the resulting perturbation of the zero-energy spectrum. We define the normalized eight-component basis of this two-dimensional zero-energy subspace

$$\tilde{\psi}_i = \frac{1}{1+6\alpha^2}\left(e_i, \quad wh_1^{-1} T_1^\dagger e_i, \quad wh_2^{-1} T_2^\dagger e_i, \quad wh_3^{-1} T_3^\dagger e_i\right), \qquad i = 1, 2 \tag{73}$$

where e_i are the canonical unit vectors in $\mathbb{R}^2$ and $\alpha^2 = w/(v_F k_\theta)$. We finally project $\mathcal{H}_{BM}$ onto this subspace resulting in the low-energy Hamiltonian

$$\begin{aligned}
\mathcal{H}_{BM}^L(\mathbf{k}) &= \left(\left\langle \tilde{\psi}_m \middle| \mathcal{H}_{BM}(\mathbf{k}) \middle| \tilde{\psi}_n \right\rangle\right) \\
&= \frac{v_F}{1+6\alpha^2}\left(\boldsymbol{\sigma} \cdot \mathbf{k} + w^2 \sum_{j=1}^{3} T_j h_j^{-1}(\boldsymbol{\sigma} \cdot \mathbf{k})h_j^{-1} T_j^\dagger\right) \\
&= \tilde{v}_F \boldsymbol{\sigma} \cdot \mathbf{k}
\end{aligned} \tag{74}$$

where in the last line we defined the Fermi velocity $\tilde{v}_F = v_F(1 - 3\alpha^2)/(1 + 6\alpha^2)$. We can see that there are solutions for which $\tilde{v}_F = 0$ including $\theta \approx 1.05°$, and the Fermi velocity vanishes. We have proven the existence of flat bands in the low-energy limit. The twist angles for which $\tilde{v}_F$ vanishes are called magic angles, and the resulting material is called magic-angle twisted bilayer graphene (MATBG).

3.5 Relaxed truncation

The BM continuum model Hamiltonian (67) is the most truncated case which captures the flat bands of MATBG. We will obtain a more precise Hamiltonian by relaxing the requirement that $\mathbf{G}_1 = \mathbf{K}$ which we used to only consider the $t_\mathbf{K}^{\alpha\beta}$ term. In preparation, we write the Hamiltonian

$$\mathcal{H}(\mathbf{k}) = \begin{pmatrix} \mathcal{H}^{++}(\mathbf{k}) & \mathcal{H}^{+-} \\ \mathcal{H}^{-+} & \mathcal{H}^{++}(\mathbf{k}) \end{pmatrix} \tag{75}$$

where the $\pm$ indices represent hopping between layers (e.g. $\mathcal{H}^{+-}$ describes hopping from layer $l = +1$ to $l = -1$). For $\mathcal{H}(\mathbf{k})$ to be Hermitian, it is required that $\mathcal{H}^{-+} = (\mathcal{H}^{+-})^\dagger$. Recall the interlayer tunnelling matrix element (64)

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} = \sum_{\mathbf{G}_1, \mathbf{G}_2'} \frac{t_{\mathbf{k}+\mathbf{G}_1}^{\alpha\beta}}{\Omega} e^{i(\mathbf{G}_1 \cdot \boldsymbol{\tau}_\alpha - \mathbf{G}_2 \cdot (\boldsymbol{\tau}_\beta - \boldsymbol{\tau}))} \delta_{\mathbf{k}+\mathbf{G}_1, \mathbf{p}'+\mathbf{G}_2'}. \tag{76}$$

As before, we require $\mathbf{k}, \mathbf{p}'$ to be near the center of the moire BZ. If $\mathbf{G}_1 \neq \mathbf{G}_2$, then the Kronecker delta would be null for all $\mathbf{k}, \mathbf{p}'$ sufficiently small since then $\mathbf{G}_2' - \mathbf{G}_1$ is much larger than $\mathbf{k} - \mathbf{p}'$. We therefore require $\mathbf{G}_1 = \mathbf{G}_2$ and simplify the interlayer hopping

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} = \sum_{\mathbf{G}_1} \frac{t_{\mathbf{k}+\mathbf{G}_1}^{\alpha\beta}}{\Omega} e^{i\mathbf{G}_1 \cdot (\boldsymbol{\tau}_\alpha - \boldsymbol{\tau}_\beta + \boldsymbol{\tau})} \delta_{\mathbf{k}-\mathbf{p}', \mathbf{G}_1' - \mathbf{G}_1} \tag{77}$$

which is nonzero if and only if $\mathbf{k} - \mathbf{p}' = \mathbf{G}_1' - \mathbf{G}_1$ for some $\mathbf{G}_1$. This requirement is satisfied for any $\mathbf{k}, \mathbf{p}'$ that are moire reciprocal vectors which we denote $\mathbf{Q}$ and defined at the beginning of this chapter. Since $t_{\mathbf{k}+\mathbf{G}_1}$ decays exponentially with respect to $|\mathbf{k} + \mathbf{G}_1| \approx |\mathbf{G}_1|$, we only consider $\mathbf{k} - \mathbf{p}' = \mathbf{G}_1' - \mathbf{G}_1$ with magnitude less than some sufficiently large radius N in the moire reciprocal lattice. This is because as the magnitude of $\mathbf{G}_1$ increases, that of the difference $\mathbf{G}_1' - \mathbf{G}_1$ also increases. We therefore consider only a truncated set of points $|\mathbf{G}_1| < M$ which defines a truncated lattice of reciprocal vectors with radius less than or equal to N, $\mathbf{k} - \mathbf{p}' \in \mathbf{Q}^N$ in the moire reciprocal lattice. Note that the BM Hamiltonian utilized the same idea but only kept the four smallest magnitude elements of $\mathbf{Q}$, namely the null vector and the three nearest-neighbor hopping vectors. We now consider this truncated moire reciprocal lattice $\mathbf{Q}^N$.

Recall that the moire BZ is a honeycomb lattice with a "sublattice" degree of freedom which having truncated the space, we denote as $\mathbf{Q}_\pm^N$. It can be shown that this sublattice degree of freedom corresponds to a combination of layer and valley. We can write $\mathbf{Q}_\pm = \mathbf{Q}_{l\eta=\pm1}$ where as before, $l = \pm1$ denotes the graphene layer and $\eta = \pm1$ denotes the valley.

We also assume the convention of considering valley as a separate parameter and therefore the layer l fully describes the lattice $\mathbf{Q}_l \equiv \mathbf{Q}_\pm$. Interlayer hopping requires $\mathbf{k}, \mathbf{p}'$ to be on opposite moire sublattices $\mathbf{k} \in \mathbf{Q}_+$ and $\mathbf{p}' \in \mathbf{Q}_-$, or the reverse. Having chosen vectors $\mathbf{k}, \mathbf{p}'$ which are in $\mathbf{Q}^N$, the summation over $\mathbf{G}_1$ reduces to one term

$$\mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} = \frac{t_{\mathbf{k}+\mathbf{G}_{\mathbf{k}-\mathbf{p}'}}^{\alpha\beta}}{\Omega} e^{i\mathbf{G}_{\mathbf{k}-\mathbf{p}'} \cdot (\boldsymbol{\tau}_\alpha - \boldsymbol{\tau}_\beta + \boldsymbol{\tau})} \tag{78}$$

where $\mathbf{G}_{\mathbf{k}-\mathbf{p}'} = (\mathbf{M}_\theta - \sigma_0)^{-1}(\mathbf{k} - \mathbf{p}')$ and σ_0 is the identity matrix.

We define the single-particle states $|\mathbf{q}_l, \alpha, \eta, s\rangle$ and the corresponding creation (annihilation) operator $c_{\mathbf{q}_l,\alpha,\eta,s}^\dagger$ ($c_{\mathbf{q}_l,\alpha,\eta,s}$) where s denotes the spin degree of freedom and $\mathbf{q}_l \in \mathbf{Q}_l$. Since we are not considering spin as part of the dynamics, we cease writing its index but note its two-fold degeneracy. We write the interlayer Hamiltonian in second-quantization

$$\mathcal{H}_\eta^{+-} = \sum_{\mathbf{k}_-, \mathbf{p}'_+ \in \mathbf{Q}} \sum_{\alpha,\beta} \mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} c_{\mathbf{k}_-,\alpha,\eta}^\dagger c_{\mathbf{p}'_+,\beta,\eta} \approx \sum_{\mathbf{k}_-, \mathbf{p}'_+ \in \mathbf{Q}^N} \sum_{\alpha,\beta} \mathcal{H}_{\mathbf{k}\mathbf{p}'}^{\alpha\beta} c_{\mathbf{k}_-,\alpha,\eta}^\dagger c_{\mathbf{p}'_+,\beta,\eta}, \quad \mathcal{H}_\eta^{-+} = (\mathcal{H}_\eta^{+-})^\dagger. \tag{79}$$

where we denote $\mathbf{k}_\pm, \mathbf{p}'_\pm$ as meaning that they belong to the appropriate $\mathbf{Q}_\pm$. The intralayer Hamiltonian can also be written

$$\mathcal{H}_\eta^{ll}(\mathbf{k}) = \sum_{\alpha,\beta} \sum_{q_l \in \mathbf{Q}^N} h_l^{\eta,\alpha,\beta}(\mathbf{k} + \mathbf{q}_l) c_{\mathbf{q}_l,\alpha,\eta}^\dagger c_{\mathbf{q}_l,\beta,\eta} \tag{80}$$

where $\mathbf{k}$ here is unrelated to the interlayer hopping momenta $\mathbf{k}, \mathbf{p}'$ we were describing above. We are simply evaluating the intralayer Hamiltonian H_η^{ll} at $\mathbf{k}$. In addition, we denote $h_{l=\pm1} = h_{\pm\theta/2}$.

3.6 Band structure

We have constructed the relaxed BM Hamiltonian which we will now use to compute the band structure of MATBG. We first construct a nice basis for the moire reciprocal lattice

$$\mathbf{q} = n_1\mathbf{q}_1^m + n_2\mathbf{q}_2^m + n_3\mathbf{q}_3^m, \qquad n_i \in \mathbb{Z}. \tag{81}$$

where $\mathbf{q}_i^m$ are the moire reciprocal nearest-neighbor hopping vectors and are defined at the beginning of this chapter. This is clearly an over-complete basis and to remedy this we (arbitrarily) set $n_3 = 0$ and have

$$\mathbf{q} = n_1\mathbf{q}_1^m + n_2\mathbf{q}_2^m \tag{82}$$

where the site $\mathbf{q}$ corresponds to a moire honeycomb center, $\mathbf{Q}_+$, or $\mathbf{Q}_-$ lattice vectors for $(n_1 + n_2)$ mod $3 = 0, 1, 2$ respectively. We solve for the eigenvalues of the Hamiltonian matrix to recover the band structure. We first generate the truncated moire reciprocal lattice $\mathbf{Q}^N$ by generating a sufficiently large list of these vectors and only keeping those with magnitude less than N. We perform this computation for each valley separately. Interlayer hopping can be computed by creating a ranked list of neighbors for each lattice vector and computing (78) for each hopping. We then include intralayer hopping by computing the diagonal elements which corresponds to (80). To calculate the band structure, we must diagonalize this matrix for a grid of momenta $\mathbf{k}$ in the moire BZ. The lattice is only generated once in this process and the construction of the Hamiltonian is practically instant. The diagonalization however is computationally intensive.[21] Nonetheless, we note that the BM model matrix is significantly smaller than that of previous models, namely a standard tight-binding model on the moire lattice [?].
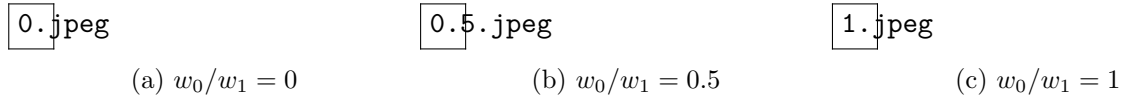

(a) $w_0/w_1 = 0$ (b) $w_0/w_1 = 0.5$ (c) $w_0/w_1 = 1$

Figure 5: Numerical computations of the band structure for $\eta = 1$, $\theta = 1.05°$. We see that the flat bands are increasingly isolated for decreasing w_0/w_1 and intersect with neighboring bands for $w_0/w_1 = 1$. See the appendix for a link to the code.

3.7 Projecting onto flat bands

We now project the real space wave functions onto the flat bands in an effort to understand where the flat band electrons are located in the lattice. Throughout this section, any summations over $\mathbf{k}$ are implicit over the moire BZ. We can write the Hamiltonian in diagonal form

$$\mathcal{H}_\eta(\mathbf{k}) = \sum_{n,\eta} \epsilon_{\mathbf{k},n,\eta} c_{\mathbf{k},n,\eta}^\dagger c_{\mathbf{k},n,\eta} \tag{83}$$

where we have written the eigenstates

$$c_{\mathbf{k},n,\eta}^\dagger = \sum_{l,\alpha} \sum_{\mathbf{q}\in\mathbf{Q}_{l\eta}} u_{\mathbf{k},\mathbf{q}_l,\alpha,n,\eta} c_{\mathbf{k},\mathbf{q}_l,\alpha,\eta}^\dagger \tag{84}$$

---

[21]The matrix is sparse especially when only including nearest-neighbor hoppings. Specialized numerical software can be used to efficiently compute the eigenvalues given that the matrix is Hermitian, sparse, and we are only interested in the lowest magnitude bands. In particular, for a sparse matrix one can use Lanczos sparse-matrix diagonalization method.

corresponding to band n and energy $\epsilon_{\mathbf{k},n,\eta}$. We included the $\mathbf{k}$ index in the creation operator $c^\dagger_{\mathbf{k},\mathbf{q}_l,\alpha,\eta}$ to denote that it is an eigenstate of the Hamiltonian evaluated at $\mathbf{k}$. Rewriting in first-quantization, we have the eigenstates

$$|\psi_{\mathbf{k},n,\eta}\rangle = \sum_{l,\alpha} \sum_{\mathbf{q}\in\mathbf{Q}_{l\eta}} u_{\mathbf{k},\alpha,n,\eta} |\mathbf{k}, \mathbf{q}_l, \alpha, n, \eta\rangle \tag{85}$$

We define the projection operator which projects states onto the flat bands [?]

$$P_{\mathbf{k},\eta} = \sum_{n\in\{1,2\}} \frac{1}{|\psi_{\mathbf{k},n,\eta}|^2} |\psi_{\mathbf{k},n,\eta}\rangle \langle \psi_{\mathbf{k},n,\eta}| \tag{86}$$

where we denote the band indices of the flat bands by $n = 1, 2$. Note that P_η has trivial action on the eigenstates

$$P_{\mathbf{k},\eta} |\psi_{\mathbf{k},n,\eta}\rangle = \begin{cases} |\psi_{\mathbf{k},n,\eta}\rangle & n \in \{1,2\}, \\ 0 & n \notin \{1,2\} \end{cases}. \tag{87}$$

Instead we will project the real space wave functions onto the flat bands which will reveal where the electrons are located. We introduce the continuous real space basis [?]

$$c^\dagger_{l,\alpha,\eta}(\mathbf{r}) = \frac{1}{\sqrt{N\Omega}} \sum_{\mathbf{k}} \sum_{\mathbf{q}\in\mathbf{Q}_{l\eta}} e^{-i(\mathbf{k}+\mathbf{Q})\cdot\mathbf{r}} c^\dagger_{\mathbf{k},\mathbf{q}_l,\alpha,\eta}, \qquad |\mathbf{r}, l, \alpha, \eta\rangle = c^\dagger_{l,\alpha,\eta}(\mathbf{r}) |0\rangle \tag{88}$$

where $\mathbf{r}$ is a continuous variable in the moire unit cell [?]. We now sum over valley since they both contribute to the real space density

$$|\mathbf{r}, \alpha, l\rangle = \sum_\eta |\mathbf{r}, \alpha, l, \eta\rangle. \tag{89}$$

We then project the real space basis onto the flat bands

$$|\mathbf{r}, \alpha, l\rangle \rightarrow \sum_{\mathbf{k}} \sum_\eta P_{\mathbf{k},\eta} |\mathbf{r}, \alpha, l, \eta\rangle_{\mathbf{k}} = \lambda_{l,\alpha} |\psi_{l,\alpha}(\mathbf{r})\rangle \tag{90}$$

yielding the real space density $|\psi_{\alpha,l}(\mathbf{r})\rangle$. $\lambda_{l,\alpha}$ is chosen so that $|\psi_{l,\alpha}(\mathbf{r})\rangle$ is normalized, and $|\mathbf{r}, \alpha, l, \eta\rangle_{\mathbf{k}}$ denotes the $\mathbf{k}$ component of $|\mathbf{r}, \alpha, l, \eta\rangle$. The magnitude-squared of $|\psi_{l,\alpha}(\mathbf{r})\rangle$ corresponds to the real space density of electrons on sublattice α and layer l. Hence, we have built the formalism for computing where the electrons in these bands are located.

4 Band Topology and Geometry

We now develop the theory of band topology and geometry to better understand the global and local properties of bands. These tools will be first applied to graphene as well as the Haldane model for Chern insulators. In the last chapter, we build on the material presented here and use it to understand the band geometry of chiral MATBG.

4.1 Berry phase

We present the theory of Berry phase in a general framework and then apply it to lattices, closely following the derivations in [?, ?, ?]. Consider a non-degenerate Hamiltonian which depends on a

d-dimensional parameter $\boldsymbol{\lambda}$ and satisfies the time-independent Schrodinger equation

$$H(\boldsymbol{\lambda}) \,|\psi_n(\boldsymbol{\lambda})\rangle = E_n \,|\psi_n(\boldsymbol{\lambda})\rangle \tag{91}$$

with band index n. Suppose that a quantum state starts as an eigenstate $|\psi_n(t=0)\rangle = |\psi_n(\boldsymbol{\lambda})\rangle$ of $H(\boldsymbol{\lambda}(t=0))$. If the parameter $\boldsymbol{\lambda}$ changes sufficiently slowly, then the Adiabatic Theorem implies that $|\psi_n(t)\rangle$ will be an instantaneous eigenstate of $H(\boldsymbol{\lambda}(t))$ up to a phase factor

$$|\psi_n(t)\rangle = e^{i\Phi_n(t)} \,|\psi_n(\boldsymbol{\lambda})\rangle \,. \tag{92}$$

We are interested in computing the phase $\Psi_n(t)$. We know from the time-dependent Schrodinger equation that

$$i\hbar \,\langle\psi_n(t)| \,\frac{\mathrm{d}}{\mathrm{d}t} \,|\psi_n(t)\rangle = \langle\psi_n(t)| \,H(\boldsymbol{\lambda}(t)) \,|\psi_n(t)\rangle = E_n(\boldsymbol{\lambda}). \tag{93}$$

Expanding the left-hand-side using (92), we have

$$\frac{1}{\hbar} E_n(\boldsymbol{\lambda}) = i \,\langle\psi_n(t)| \,\frac{\mathrm{d}}{\mathrm{d}t} \,|\psi_n(t)\rangle = -\frac{\mathrm{d}\Phi_n(t)}{\mathrm{d}t} + i\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} \cdot \langle\psi_n(\boldsymbol{\lambda})| \,\nabla_{\boldsymbol{\lambda}} \,|\psi_n(\boldsymbol{\lambda})\rangle \,. \tag{94}$$

We now assume that $\boldsymbol{\lambda}$ is varied in a closed loop C and solve for the phase

$$\Phi_n(t) = -\frac{1}{\hbar} \int_0^t dt' E_n(\boldsymbol{\lambda}(t')) + \int_C d\lambda^\mu \mathcal{A}_\mu \tag{95}$$

where we have defined the (abelian[22]) Berry connection

$$\mathcal{A}_\mu^n = i \,\langle\psi_n(\boldsymbol{\lambda})|\partial_\mu \psi_n(\boldsymbol{\lambda})\rangle \tag{96}$$

which is always real and we denoted $\partial_\mu = \partial_{\lambda_\mu}$. The first term of (95) is the usual dynamical phase associated with time-evolution[23] and the second term is known as the Berry phase and has purely geometric origins.[24] The Berry connection is not an observable since it is not gauge-invariant. Transforming the gauge of the wave function $|\psi_n(\boldsymbol{\lambda})\rangle \to e^{i\alpha(\boldsymbol{\lambda})} |\psi_n(\boldsymbol{\lambda})\rangle$ transforms the Berry connection non-trivially $\mathcal{A}_\mu^n \to \mathcal{A}_\mu^n - \partial_\mu\alpha$. We can define the gauge-invariant Berry curvature anti-symmetric[25] tensor

$$\Omega_{\mu\nu}^n = \partial_\mu\mathcal{A}_\nu^n - \partial_\nu\mathcal{A}_\mu^n = -2\mathrm{Im} \,\langle\partial_\mu\psi_n(\boldsymbol{\lambda})|\partial_\nu\psi_n(\boldsymbol{\lambda})\rangle \tag{97}$$

which is simply the exterior derivative of the 1-form Berry connection $\Omega = d\mathcal{A}$. We can use Stokes' Theorem to rewrite the Berry phase in terms of the Berry curvature

$$\Phi_C^n = \oint_C d\lambda^\mu \mathcal{A}_\mu^n = \int_S d\lambda^\mu d\lambda^\nu \Omega_{\mu\nu}^n \tag{98}$$

---

[22] We treated the non-degenerate case, but for a degenerate Hamiltonian (i.e. bands cross) we partition the Hilbert space and we define the non-abelian Berry connection ($N \times N$ matrix) $\mathcal{A}_\mu^{mn}(\mathbf{k}) = i \,\langle\psi_m(\boldsymbol{\lambda})|\partial_\mu\psi_n(\boldsymbol{\lambda})\rangle$ for each N-dimensional degenerate subspace. The evolution of the state (92) generalizes to $|\psi(t)\rangle = e^{i\phi(t)}\boldsymbol{\Gamma}\,|\psi\rangle$ where $\phi(t)$ is the dynamical phase, $\boldsymbol{\Gamma} = \mathcal{P}\exp\int_C d\lambda^\mu \mathcal{A}_\mu$, and $\mathcal{P}$ denotes path ordering. In the non-degenerate case, $\boldsymbol{\Gamma}$ becomes diagonal and we have our usual Berry phase. For the remainder of this paper we solely use the abelian variant.

[23] We assume the Hamiltonian commutes with itself so we have the simple integral expression for the dynamical phase.

[24] In mathematics, the notion of Berry phase can be explained by the general theory of holonomy.

[25] Anti-symmetry is the requirement that $\Omega_{\mu\nu} = -\Omega_{\nu\mu}$ which immediately implies that the diagonal terms vanish $\Omega_{\mu\mu} = 0$.

where C is the boundary of the surface S. It is simple to show that Φ_C is only gauge-invariant if C is a closed loop.

We consider electromagnetism as an analogy with the formalism above. We have a magnetic field $\mathbf{B}$ which can be written as the curl of the vector potential $\mathbf{B} = \nabla \times \mathbf{A}$. We can then write the magnetic flux enclosed by a closed loop C as

$$\Phi_C = \oint_C d\mathbf{r} \cdot \mathbf{A} = \int_S d\mathbf{S} \cdot \mathbf{B}. \tag{99}$$

where S is the surface enclosed by C. This is nothing but the three-dimensional form of the Berry phase equation (98). In two and three dimensions, we can write the Berry curvature as the curl of the pseudo-vector Berry connection $\Omega = \nabla \times \mathcal{A}$. This analogy motivates one to call the Berry connection and Berry phase as the Berry potential and Berry flux respectively, as some sources often do [?], although we use the former vocabulary in this paper. Therefore, just as the magnetic field can be thought of as a point source of magnetic flux, the Berry curvature can also be thought of as a point source of Berry flux.

We now apply the theory of Berry phase to lattices. As a special case of the above, we let the parameter be momentum. As is convention, we ignore the plane-wave part of the eigenstates and define the Berry curvature

$$\Omega^n_{\mu\nu} = -2\mathrm{Im} \langle \partial_\mu u_{n,\mathbf{k}} | \partial_\nu u_{n,\mathbf{k}} \rangle \tag{100}$$

where we use the ket form of the Bloch function (20). The expression for the Berry phase is the same (98). For the remainder of the paper, we will only compute the Berry curvature using the above formula and will not use the Berry connections. Also note that when we consider two-dimensional lattices, the Berry curvature only has one unique non-zero entry Ω^n_{xy}.

4.2 Chern insulator

The definitions above may seem rather abstract and disconnected from reality, but we will now see how they can give rise to macroscopic phenomena. We consider an electron in a lattice and an electric field which drives current. The semiclassical equation of motion of an electron is derived in the appendix and we cite the result

$$\frac{d\mathbf{r}}{dt} = \frac{1}{\hbar}\nabla_\mathbf{k} E_{n,\mathbf{k}} + \mathbf{v}_a, \qquad \frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar}\mathbf{E}, \qquad \mathbf{v}_a = -\frac{d\mathbf{k}}{dt} \times \Omega^n_{xy}(\mathbf{k}) \tag{101}$$

where $\mathbf{v}_a$ is called the anomalous velocity and is responsible for the anomalous Hall (transverse) conductivity σ_{xy}. The term "anomalous" is used to differentiate the present case with that of the integer quantum Hall effect (IQHE) in which it is not the geometry of the material causing nonzero Hall conductivity but rather a magnetic field, disorder, and time-reversal symmetry. Nonetheless, many of the same effects are present. Using these equations, we can calculate the current density

$$\mathbf{j} = \sigma\mathbf{E} + \sigma_{xy}\hat{\mathbf{z}} \times \mathbf{E} \tag{102}$$

where σ is the usual longitudinal conductivity[26]. For an insulator with Fermi energy in the band gap, we can derive the TKNN formula [?]

$$\sigma_{xy}^n = \frac{e^2}{2\pi\hbar}\mathcal{C}_n, \qquad \mathcal{C}_n = \frac{1}{2\pi}\int_{BZ} d^2k\,\Omega_{xy}^n(\mathbf{k}) \tag{103}$$

where $\mathcal{C}_n$ is called the Chern number and is an integer [?, ?].[27] Remarkably, the Hall conductivity is quantized in units of the quantum conductance $e^2/2\pi\hbar$ and is purely a consequence of the Berry curvature. The Chern number of a band tells us the quantized value of the Hall conductivity. A material with net non-zero Chern number is called a Chern insulator and a band with non-zero Chern number is called a Chern band. The total Hall conductivity is simply the sum over the Hall conductivities of the filled bands. This phenomena, known as the quantum anomalous Hall effect (QAHE) was first experimentally realized in 2013 [?]. We will not derive the IQHE in this paper, but we note that one of its key features is quantized Hall conductivity, just as we derived here. In addition, while time-reversal symmetry is fundamental to the IQHE, we will see that it in fact must be broken for the QAHE.

4.3 Space-inversion and time-reversal symmetry

We will now investigate the effects of space-inversion and time-reversal symmetry on the Chern number. The time-reversal operator Θ flips the sign of momentum $\mathbf{k} \to -\mathbf{k}$ and leaves position unchanged $\mathbf{r} \to \mathbf{r}$. The band structure must also be even in momentum $\Theta^\dagger E_n(\mathbf{k})\Theta = E_n(-\mathbf{k})$. It is straightforward to verify that $\Theta^\dagger\Omega(\mathbf{k})\Theta = -\Omega(-\mathbf{k})$, and therefore, assuming time-reversal symmetry, $\Omega(\mathbf{k})$ is odd in $\mathbf{k}$ [?]. We then find that the Chern number vanishes since it is the integral of an odd function over a closed space. We see that for a band to have a non-trivial Chern number, time-reversal symmetry must be broken. On the other hand, the space-inversion operator $\mathcal{X}$ inverts both position and momentum $\mathbf{r} \to -\mathbf{r}$, $\mathbf{k} \to -\mathbf{k}$. It can be shown that assuming space-inversion symmetry, the Berry curvature is even in $\mathbf{k}$ [?]. Combining the space-inversion and time-reversal symmetries, we have that the Berry curvature must be identically zero. This trivially implies that the Chern number vanishes for all bands but is clearly a much stronger statement.

We now apply these results to graphene. Recall the tight-binding Hamiltonian for graphene (104) with a (real) momentum-dependent sublattice potential

$$H(\mathbf{k}) = \begin{pmatrix} m(\mathbf{k}) & -tf(\mathbf{k}) \\ -tf^*(\mathbf{k}) & -m(\mathbf{k}) \end{pmatrix}, \qquad f(\mathbf{k}) = \sum_{j=1}^{3} e^{-i\mathbf{k}\cdot\boldsymbol{\tau}_j}. \tag{104}$$

The time-reversal operator can be decomposed as $\Theta = UK$ where U is unitary and K is the complex conjugation operator. Therefore, the Hamiltonian transforms as $\Theta^\dagger H(\mathbf{k})\Theta = H^*(-\mathbf{k})$. We see that $H(\mathbf{k})$ obeys time-reversal symmetry given that $m(\mathbf{k}) = m(-\mathbf{k})$ since $f^*(-\mathbf{k}) = f(\mathbf{k})$. On the other hand, space-inversion transforms $\mathbf{k} \to -\mathbf{k}$ and $\boldsymbol{\tau}_j \to -\boldsymbol{\tau}_j$, leaving the Hamiltonian invariant given that $m(\mathbf{k}) = -m(-\mathbf{k})$.[28] As we expect, if m is a constant then space-inversion symmetry is broken if and only if $m \neq 0$ since the sublattices are (not) distinguished if m is non-zero (zero). We see

---

[26]The expression for σ differs from that of the Drude model in that it takes into consideration a relaxation time for scattering effects.

[27]An important property of tight-binding models is that the sum of its Chern numbers vanish. It is a mathematical statement but can also be understood physically since if all the bands are completely filled, then the Hall conductivity vanishes as an electric field can not impart current.

[28]The diagonal entries of the Hamiltonian switch since sublattices are inverted which results in the minus sign.

that graphene obeys both space-inversion and time-reversal symmetry if and only if $m = 0$, and obeys time-reversal symmetry for any constant m. [29] From the arguments above, we have that graphene is not a Chern insulator for constant m since it obeys time-reversal symmetry.

4.4 Graphene

We now turn our attention to applying the theory of Berry phase to graphene. A sublattice potential m is included to gap the bands. Recall the continuum model Hamiltonian

$$h(\mathbf{k}) = v_F(\tau_z \sigma_x k_x + \sigma_y k_y) + m\sigma_z \tag{105}$$

where τ_z and σ_i act on valley and sublattice subspace respectively. We denote the Bloch functions of band $n = \pm 1$ (conductance and valence bands respectively) and valley $\eta = \pm 1$ as $u^\eta_{\pm 1, \mathbf{k}}(\mathbf{r})$. Since we are working in a two-dimensional parameter space (k_x, k_y), there is only one unique nonzero entry of the Berry curvature tensor

$$\Omega^{n\eta}_{xy} = -2\mathrm{Im}\left\langle \partial_{k_x} u^\eta_{n,\mathbf{k}} \middle| \partial_{k_y} u^\eta_{n,\mathbf{k}} \right\rangle = \frac{\mathrm{sign}(\eta n)v_F^2 m}{2(m^2 + (v_F k)^2)^{3/2}}. \tag{106}$$

where $k = |\mathbf{k}|$. We can clearly see that if the system satisfies space-inversion and time-reversal symmetry ($m = 0$), then the Berry curvature vanishes everywhere except the origin where it is undefined. We compute the Berry phase of an electron orbiting the closed loop C containing a Dirac point corresponding to valley η

$$\Phi^{n\eta}_C = \int_S d^2k\ \Omega^{n\eta}_{xy} = \mathrm{sign}(\eta n)\pi \left(1 - \frac{m}{\sqrt{m^2 + (v_F r)^2}}\right) \tag{107}$$

where in the last equality, r denotes the radius of C which we take to be a circle. Setting $m = 0$, we have $\Phi^{n\eta}_C = \mathrm{sign}(\eta n)\pi$ and note that the Berry curvature converges to $\mathrm{sign}(\eta n)\pi\delta(\mathbf{k})$. Consequently, for any closed loop C enclosing a Dirac point, the Berry phase is constant irrespective of the geometry of the loop. The quantization of Berry phase at the Dirac points is a simple result of space-inversion and time-reversal symmetries. Acting on the Berry phase with these symmetries, we have

$$\mathcal{X}^\dagger \Theta^\dagger \Phi_C \Theta \mathcal{X} = \int_S dk^\mu dk^\nu (\mathcal{X}^\dagger \Theta^\dagger \Omega_{\mu\nu}(\mathbf{k}) \Theta \mathcal{X}) = -\int_S dk^\mu dk^\nu \Omega_{\mu\nu}(\mathbf{k}) = -\Phi_C \tag{108}$$

and therefore under space-inversion and time-reversal symmetries, $\Phi_C = -\Phi_C$. Since Φ_C is only well-defined modulo 2π, we have $\Phi_C = -\Phi_C + 2n\pi$ ($n \in \mathbb{Z}$), and finally, $\Phi_C = n\pi$. Therefore, Φ_C can only attain the values 0 and π modulo 2π. This immediately fixes the form of the Berry curvature

$$\Omega^n_{\mu\nu}(\mathbf{k}) = \sum_i a_i \pi \delta(\mathbf{k} - \mathbf{k}'_i), \qquad a_i \in \mathbb{Z}_{\geq 0} \tag{109}$$

where the delta functions serve as point sources of Berry curvature [?]. Assuming space-inversion and time-reversal symmetry, a smooth change of parameters can only result in a smooth variation of the positions of these point sources. In particular, they cannot be created (annihilated) without

---

[29]Hexagonal boron nitride (hBN) also forms a honeycomb-like lattice but with unequal bond lengths. In this case, there is a clear distinction between the sublattices (corresponding to boron and nitrogen atoms) which breaks space-inversion symmetry and contributes an sublattice potential.

combining Dirac points of same (opposite) Berry curvature sign which is a non-smooth process.

4.5 Haldane model for Chern insulator

We saw above that the Chern number is trivial for graphene due to time-reversal symmetry. However, it turns out, as Haldane discovered in 1987 that a modified version of graphene serves as an excellent model for constructing a Chern insulator. Once again, recall the graphene continuum model Hamiltonian

$$H = v_F(\tau_z \sigma_x k_x + \sigma_y k_y) + m\sigma_z. \tag{110}$$

With broken sublattice symmetry (resulting in broken space-inversion symmetry), a band gap opens and therefore the Berry curvature is well-defined and may be nonzero. However, the Chern number is still trivial due to time-reversal symmetry. With the above model, the Berry curvature has opposite sign at each valley. Therefore, to yield a non-trivial Chern number, these signs must be aligned so that they sum constructively. A simple solution would be to have a valley-dependent sublattice potential

$$H = v_F(\tau_z \sigma_x k_x + \sigma_y k_y) + m\tau_z \sigma_z. \tag{111}$$

While this seems like a quick fix, it is not clear what physical system this corresponds to. Therefore, we are motivated to explore a tight-binding model on graphene called the Haldane model which will reduce to (111) in the continuum limit. Recall the nearest-neighbor tight-binding model of graphene (35)

$$H = -t_1 \sum_{\langle \mathbf{R}_A^i, \mathbf{R}_B^j \rangle} (c_{\mathbf{R}_A^i}^\dagger c_{\mathbf{R}_B^j} + h.c.) + m \sum_\alpha (-1)^\alpha \sum_{\mathbf{R}_\alpha^i} c_{\mathbf{R}_\alpha^i}^\dagger c_{\mathbf{R}_\alpha^i} \tag{112}$$

We now include imaginary next-to-nearest-neighbor (NNN) hoppings induced by a magnetic field coupling electrons of the same sublattice. We first briefly recall the Aharonov-Bohm Effect. Consider an electron which moves along the closed loop C containing a magnetic field $\mathbf{B}$. We cite the result which states that the electron gains a geometric phase $|\psi\rangle \to e^{i\Phi_C}|\psi\rangle$ after traversing the loop C where

$$\Phi_C = \frac{e}{\hbar} \int_S d\mathbf{S} \cdot \mathbf{B} \tag{113}$$

and S is the surface enclosed by C [?]. If the orientation of the loop C changes (the electron changes direction) then that of the surface also changes and we have $\Phi_{-C} = -\Phi_C$. We note that this phase is gauge-invariant and is also a physical observable as long as C is closed. We let Φ_α denote the phase gained by an electron on a single NNN hop on the α sublattice ($3\Phi_\alpha$ corresponds to a closed loop). For simplicity, we let $\Phi_A = \Phi_B = \Phi$. The NNN tight-binding Hamiltonian then takes the
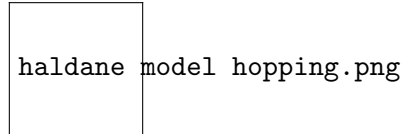

Figure 6: Illustration of counterclockwise NNN hopping in the Haldane model [?].

form

$$H_{NNN} = -t_2 \sum_\alpha \sum_{\langle\langle \mathbf{R}_\alpha^i, \mathbf{R}_\alpha^j \rangle\rangle} (e^{i\Phi_\alpha} c_{\mathbf{R}_\alpha^i}^\dagger c_{\mathbf{R}_\alpha^j} + h.c.) \tag{114}$$

where the sum is only over counter-clockwise hoppings and the double angle brackets denotes NNN. We consider both NN and the imaginary NNN hoppings to form the Hamiltonian

$$H = H_{NN} + H_{NNN} \tag{115}$$

which we can write in the momentum space basis $(c_{\mathbf{k},A}^\dagger, c_{\mathbf{k},B}^\dagger)^T$ as

$$H(\mathbf{k}) = \begin{pmatrix} -2t_2\gamma_\Phi(\mathbf{k}) + m & -t_1 f(\mathbf{k}) \\ -t_1 f^*(\mathbf{k}) & -2t_2\gamma_{-\Phi}(\mathbf{k}) - m \end{pmatrix} \tag{116}$$

where $\gamma_\Phi(\mathbf{k}) = e^{i\Phi}(e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot(\mathbf{a}_2-\mathbf{a}_1)} + e^{i\mathbf{k}\cdot\mathbf{a}_2})$ [?, ?] and $\mathbf{a}_i$ are the primitive lattice vectors of graphene. This Hamiltonian can be written as

$$H(\mathbf{k}) = H_0(\mathbf{k})\sigma_0 + H_x(\mathbf{k})\sigma_x + H_y(\mathbf{k})\sigma_y + H_z(\mathbf{k})\sigma_z \tag{117}$$

where $H_z = \mp 3\sqrt{3}t_2\sin\phi + m$ at the $\eta = \pm 1$ valley respectively [?]. In addition, H_x, H_y vanish at the valleys and therefore we have a band gap for $H_z \neq 0$. For $|m| < \left|3\sqrt{3}t_2\sin\Phi\right|$, the sign of H_z changes at the valleys (assuming $\Phi \neq n\pi$), just as we had intended for with the $m\tau_z\sigma_z$ term in (111), and therefore we have a pair of Chern bands. For $|m| > \left|3\sqrt{3}t_2\sin\Phi\right|$, the sign of H_z is constant across the two valleys yielding a trivial band.[30] Therefore, if we fill the valence band in the former case, we have a Chern insulator, the first example of a topological insulator outside the realm of the integer quantum Hall effect.

5 Chiral MATBG and Fractional Chern Insulators

We have now explored models for both graphene and magic-angle twisted bilayer graphene. We derived the Haldane model in the previous section which was historically important for proving that the effects of the integer quantum Hall effect are general and can be found in other physical scenarios. We will now show that chiral magic-angle twisted bilayer graphene (chiral MATBG) also admits novel topological behavior, namely a phase with origins in the fractional quantum Hall Effect but in a lattice setting. Although we will not derive the fractional quantum Hall effect, we will cite basic facts which the interested reader can follow up on. In 1982, it was observed that the Hall resistivity was also quantized at fractional filling factors ν, first for $\nu < 1$, or fractionally filling the lowest Landau level. It was understood that such an effect had origins in electron interactions since the bands are only partially filled. As we mentioned, similar phases have been suggested in chiral MATBG. This novel behavior has both experimental and theoretical underpinnings but we dedicate this chapter to deriving the latter. In particular, we will find that the flat bands of chiral MATBG are remarkably similar to the bands of the lowest Landau level (LLL), corresponding to the lowest energy state of a particle confined in two dimensions and subject to a transverse magnetic field. Per definition, all electrons in a flat band have the same energy regardless of their momentum. Consequently, electrons are highly sensitive to interactions since there is no other dominant interaction between them and the environment or even their kinetic energy. This notion is fundamental to understanding why interactions play such a key role in both chiral MATBG as well as the LLL.

We will first show that the flat band wave function of chiral MATBG is holomorphic, or

---

[30]For $|m| = \left|3\sqrt{3}t_2\sin\Phi\right|$, the bands intersect at a point and the Berry curvature is trivial. This point marks a topological phase transition since above this point we have a Chern insulator while below it we have a trivial insulator.

complex-differentiable. We will then introduce Landau levels and the wave function of the lowest Landau level in the symmetric gauge. We will show that since both wave functions are holomorphic, they admit a simplified geometry which fundamentally connects them and has been established in the literature to imply the existence of fractional Chern insulator (FCI) phases. At the end, we will briefly review current experimental work in uncovering FCI phases in chiral MATBG and other materials. We note that this section is not entirely rigorous and the arguments are oftentimes heuristic rather than mathematical. However, we encourage the reader to refer to the citations if they are interested in the exact proofs. The goal of this chapter is to introduce the key ingredients for FCI phases in MATBG rather than provide a rigorous derivation.

5.1   Chiral MATBG flat band wave function

We begin by investigating the chiral limit of MATBG ($w_0 = 0$) known as chiral MATBG. This regime is called chiral because, as we will see, the Hamiltonian obeys chiral (or sublattice) symmetry. Recall the real space BM Hamiltonian for valley $\eta = 1$ (68)

$$H = \begin{pmatrix} -i\hbar v_F \boldsymbol{\sigma}_{\theta/2} \cdot \nabla & T(\mathbf{r}) \\ T^\dagger(\mathbf{r}) & -i\hbar v_F \boldsymbol{\sigma}_{-\theta/2} \cdot \nabla \end{pmatrix} \tag{118}$$

where we eliminated the rotation in the Pauli matrices by rotating the spinors. H acts on the spinor $\Phi_{\mathbf{k}} = (\psi_1, \chi_1, \psi_2, \chi_2)$ where the index denotes the layer and ψ, χ correspond to sublattice A, B respectively. Note the chiral symmetry $\{H, \sigma_z^1\} = \mathbf{0}$ where $\sigma_z^1 = \sigma_0 \otimes \sigma_z$. We are motivated to recast the Hamiltonian in a basis that takes advantage of this symmetry by interchanging χ_1 and ψ_2 to achieve

$$H = \begin{pmatrix} 0 & D^*(-\mathbf{r}) \\ D(\mathbf{r}) & 0 \end{pmatrix}, \qquad D(\mathbf{r}) = \begin{pmatrix} -2i\bar{\partial} & \alpha U(\mathbf{r}) \\ \alpha U(-\mathbf{r}) & -2i\bar{\partial} \end{pmatrix} \tag{119}$$

where $U(\mathbf{r}) = e^{-i\mathbf{q}_1 \cdot \mathbf{r}} + e^{i\phi} e^{-i\mathbf{q}_2 \cdot \mathbf{r}} + e^{-i\phi} e^{-i\mathbf{q}_3 \cdot \mathbf{r}}$, and $\alpha = w_{AB}/(\hbar v_F k_\theta) > 0$ is the only (dimensionless) parameter and therefore fully controls the dynamics of the system [?,?,?]. $\bar{\partial} = (\partial_x + i\partial_y)/2$ is called the anti-holomorphic derivative in complex analysis and commutes with holomorphic functions.[31]

We are in search of the wave functions of the perfectly flat bands of this Hamiltonian. It is clear that the zero modes will be sublattice-polarized in the sense that they are eigenstates of the chiral operator $\sigma_z^2 = \sigma_z \otimes \sigma_0$.[32] This is evident by noting that the solution to the positive chirality zero mode $D(\mathbf{r})\psi_{\mathbf{k}}(\mathbf{r}) = 0$ implies that $\chi_{\mathbf{k}}(\mathbf{r}) = \psi_{\mathbf{k}}^*(-\mathbf{r})$ solves the negative chirality zero mode equation $D^*(-\mathbf{r})\chi_{\mathbf{k}}(\mathbf{r}) = 0$. We assume this expression for $\chi_{\mathbf{k}}(\mathbf{r})$ when appropriate and now focus on solving $D(\mathbf{r})\psi_{\mathbf{k}}(\mathbf{r}) = 0$. The existence of perfectly flat bands, which we earlier derived in the low-energy limit, implies that $D(\mathbf{r})\psi_{\mathbf{k}}(\mathbf{r}) = 0$ has solutions for all $\mathbf{k}$ in the moire BZ.

Due to symmetry considerations, it can be shown that $\psi_{\mathbf{K}}$ is a zero mode for all α [?]. We can compute this solution as a perturbation in α [?]

$$\psi_{\mathbf{K}}(\mathbf{r}) \approx \begin{pmatrix} 1 + \alpha^2 u_2(\mathbf{r}) + \alpha^4 u_4(\mathbf{r}) \\ \alpha u_1(\mathbf{r}) + \alpha^3 u_3(\mathbf{r}) \end{pmatrix}. \tag{120}$$

---

[31]A holomorphic function $f(z)$ means that it is complex-differentiable, $\partial f(z) = (\partial_x - i\partial_y)f(z) = 0$ where $z = x + iy$.

[32]Reshuffling the basis also changed the presentation of the chirality operator $\sigma_z^1 \to \sigma_z^2$.

where the $u_n(\mathbf{r})$ can be computed step-by-step numerically by solving the differential equation

$$-2i\overline{\partial}u_n(\mathbf{r}) + U((-1)^n\mathbf{r})u_{n-1}(\mathbf{r}) = 0. \tag{121}$$

We now solve for zero modes at $\mathbf{k} \neq \mathbf{K}$ and begin by noticing that $D(\mathbf{r})(f(z)\psi_{\mathbf{K}}) = 0$ for $f(z)$ holomorphic and $z = x + iy$ since $D(\mathbf{r})$ only has anti-holomorphic derivatives $\overline{\partial}$ which commute with holomorphic functions. Therefore we assume an ansantz

$$\psi_{\mathbf{k}}(\mathbf{r}) = f_{\mathbf{k}}(z)\psi_{\mathbf{K}}(\mathbf{r}). \tag{122}$$

Note that if $f(z)$ is holomorphic then Louville's theorem from complex analysis implies that $f(z)$ either diverges at ∞ or is constant. Both cases violates Bloch periodicity, and therefore we resort to meromorphic functions.[33] The poles of $f(z)$ must exactly interfere with the zeros of $\psi_{\mathbf{K}}(\mathbf{r})$ so that $f(z)$ is well-behaved. It turns out that it is exactly at the magic angles that $\psi_{\mathbf{K}}(\mathbf{r})$ has zeros for some $\mathbf{r}$ and we have this flat band solution. The Bloch periodicity of $f(z)$ and certain other requirements fix its form in terms of Jacobi theta functions which we omit since it will not be used explicitly. However, we note the important result that the Bloch functions $u_{\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{\mathbf{k}}(\mathbf{r})$ can be chosen to be holomorphic in $k = k_x + ik_y$. This will be fundamental to the connection between these wave functions and those of the LLL.

Therefore, we have found two holomorphic zero modes of H, the positive chirality $\psi_{\mathbf{k}}(\mathbf{r})$ and the negative chirality $\chi_{\mathbf{k}}(\mathbf{r})$ which can be shown to have Chern numbers ± 1 respectively [?]. If we then include the same analysis for the Hamiltonian of valley $\eta = -1$ and include spin-degeneracy, we have a total of 4 (Chern) bands with Chern number 1, and 4 with Chern number -1 [?]. We note for later reference that the Berry curvatures of these Chern bands is spread out in the moire BZ with little variation. Note that Berry curvature is focused around sections of the band structure with high curvature or kinks. Since MATBG has flat bands, and assuming they are sufficiently gapped, it is not surprising that the Berry curvature is nearly uniform. Recall in the computation of the Berry curvature for graphene that as the gap between the conduction and valence bands closed, the Berry curvature was more focused and intense at the Dirac points. We direct the reader to [?] for a numerical calculation of this variation and comparison with other models.

5.2   Magic angles

We take a short detour to compute the magic angles, or the angles for which the Fermi velocity vanishes, now that we can compute the exact form of the flat band wave function using (120, 121) and the aforementioned form of $\chi_{\mathbf{K}}(\mathbf{r})$. We compute the normalized Fermi velocity[34]

$$v_F(\alpha) \approx |\nabla_{\mathbf{k}}E_{\mathbf{k}}|_{\mathbf{k}=0} = \left|\frac{\langle\Phi_{\mathbf{K}}|\nabla_{\mathbf{k}}H_{\mathbf{k}}|\Phi_{\mathbf{K}}\rangle}{\langle\Phi_{\mathbf{K}}|\Phi_{\mathbf{K}}\rangle}\right|_{\mathbf{k}=0}, \qquad H_{\mathbf{k}} \approx \begin{pmatrix} 0 & \overline{k}\sigma_0 \\ k\sigma_0 & 0 \end{pmatrix} \tag{123}$$

where we approximated

$$D(\mathbf{r}) = \begin{pmatrix} -2i\overline{\partial} & \alpha U(\mathbf{r}) \\ \alpha U(-\mathbf{r}) & -2i\overline{\partial} \end{pmatrix} \approx k\sigma_0 \tag{124}$$

---

[33]A complex function is meromorphic if it is holomorphic in all of $\mathbb{C}$ except for some isolated sets of points at which it diverges nicely (pole). A subset of meromorphic functions are holomorphic functions, namely they do not have any poles.

[34]It is normalized in the sense that it is written in units of the Fermi velocity of graphene.

by neglecting the off-diagonal terms (assuming small α) and transforming the derivative operators into momentum space [?]. Substituting for $\Phi_{\mathbf{K}}$, we have [?]

$$
\begin{aligned}
v_F(\alpha) &\approx \frac{|\langle \psi_{\mathbf{K}}^*(-\mathbf{r})|\psi_{\mathbf{K}}(\mathbf{r})\rangle|}{\langle \psi_{\mathbf{K}}|\psi_{\mathbf{K}}\rangle} \\
&= \frac{1 - 3\alpha^2 + \alpha^4 - \frac{111}{49}\alpha^6 + \frac{143}{294}\alpha^8 + \dots}{1 + 3\alpha^2 + 2\alpha^4 + \frac{6}{7}\alpha^6 + \frac{107}{98}\alpha^8 + \dots} \approx \frac{1 - 3\alpha^2}{1 + 6\alpha^2}.
\end{aligned}
\tag{125}
$$

Note that earlier, we had derived the approximate form in the regime of unit hopping ratio $w_0/w_1 = 1$.[35] The two regimes agree due to the robustness of the first magic angle and it is not suggested that the trend continues. We also note that the spacing between solutions of $v_F(\alpha) = 0$ is approximately constant, $\Delta\alpha \approx 3/2$. Hence, we can compute all magic angles in the chiral regime, a great advancement from our earlier tricky derivation.

5.3   Landau levels

We now derive the form of the wave function of the lowest Landau level and compare their geometry in the next few sections. We begin with a discussion of Landau levels following the standard derivation presented in [?, ?]. In the presence of a transverse magnetic field, classically we expect an electron confined to two dimensions to undergo cyclotron motion. When we solve this problem using quantum mechanics, we see that the energy levels are discretized similar to a harmonic oscillator potential except that each energy level is wildly degenerate. These energy levels are called Landau levels and the lowest level is called the lowest Landau level (LLL). We will see that the LLL wave function has a very similar geometry to that of chiral MATBG.

The Hamiltonian of an electron with charge $-e$ in the presence of a magnetic field $\mathbf{B}$ (and in the absence of an electric field) can be written

$$
H = \frac{1}{2m}(\mathbf{p} + e\mathbf{A})^2
\tag{126}
$$

where we assume $\mathbf{B} = \nabla \times \mathbf{A} = B\hat{\mathbf{z}}$. We define the mechanical momentum operator

$$
\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}
\tag{127}
$$

which is Hermitian and satisfies the commutation relation $[\pi_x, \pi_y] = -ie\hbar B$.[36] We further define the raising and lowering operators

$$
a = \frac{1}{\sqrt{2e\hbar B}}(\pi_x - i\pi_y), \qquad \left[a, a^\dagger\right] = 1
\tag{128}
$$

and rewrite the Hamiltonian

$$
H = \frac{\boldsymbol{\pi}^2}{2m} = \hbar\omega_B\left(a^\dagger a + \frac{1}{2}\right)
\tag{129}
$$

where $\omega_B = eB/m$ is the cyclotron frequency. We may then construct the ground state $|0\rangle$ (and higher states) by acting on it with the lowering operator $a|0\rangle = 0$ and solving the differential

---

[35]An even earlier paper than that of Bistritzer and MacDonald reported a normalized Fermi velocity of $1 - 9\alpha^2$, which agrees with (125) in the small α regime [?].

[36]The mechanical momentum corresponds to the usual momentum of the electron is the sense that in the Heisenberg picture, $\boldsymbol{\pi} = m d\mathbf{x}/dt$.

equation. We then find that the state $|n\rangle = (a^\dagger)^n/\sqrt{n!}\,|0\rangle$ has energy

$$E_n = \hbar\omega_B\left(n + \frac{1}{2}\right), \qquad n \in \mathbb{Z}_{\geq 0}. \tag{130}$$

These equally-spaced energy levels are called Landau levels. It turns out that these energy levels are wildly degenerate, and this degeneracy can be revealed by using the symmetric gauge

$$\mathbf{A} = -\frac{1}{2}\mathbf{r} \times \mathbf{B} = \frac{1}{2}B(-y, x) \tag{131}$$

which breaks translation symmetry in both $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. We are searching for new raising and lowering operators which commute with $a, a^\dagger$ to reveal the degeneracy. We define the Hermitian operators

$$\tilde{\boldsymbol{\pi}} = \mathbf{p} - e\mathbf{A}, \qquad [\tilde{\pi}_x, \tilde{\pi}_y] = ie\hbar B \tag{132}$$

which satisfy the trivial commutation relations $[\pi_i, \tilde{\pi}_i] = 0, [\pi_x, \tilde{\pi}_y] = [\pi_y, \tilde{\pi}_x] = 0$ in the symmetric gauge. Consequently, we are motivated to define the raising and lowering operators

$$b = \frac{1}{2e\hbar B}(\tilde{\pi}_x + i\tilde{\pi}_y), \qquad \left[b, b^\dagger\right] = 1 \tag{133}$$

which commute with the $a, a^\dagger$ raising and lowering operators. We may solve for the ground state by solving $a\,|0,0\rangle = 0$, and define the general state

$$|n, m\rangle = \frac{(a^\dagger)^n (b^\dagger)^m}{\sqrt{n!m!}}\,|0,0\rangle, \qquad E_{n,m} = E_n = \hbar\omega_B\left(n + \frac{1}{2}\right), \qquad n, m \in \mathbb{Z}_{\geq 0}. \tag{134}$$

Note that as before, the energy level of the $|n, m\rangle$ state only depends on n.

5.4 Lowest Landau level

We will now compute the wave function of the lowest Landau level (LLL) ($n = 0$) by solving the differential equation

$$a\,|0, m\rangle = 0, \qquad a = -\sqrt{2}i\left(l_B\bar{\partial} + \frac{z}{4l_B}\right) \tag{135}$$

for general m where we defined the magnetic length $l_B = \sqrt{\hbar/eB}$ for convenience. We find that the LLL wave function takes the form [?]

$$\psi_{LLL}(z) = f(z)e^{-|z|^2/4l_B^2} \tag{136}$$

where $z = x + iy$ and $f(z)$ is any holomorphic function.[37] Note that in both chiral MATBG and the present case, we have an anti-holomorphic derivative which presents itself as a holomorphic degree of freedom in the wave functions.

We will now heuristically discuss the geometric properties of this wave function for the sake of brevity. Although the Landau levels were derived in the absence of a lattice, it turns out that one can use translation symmetries of the magnetic field to place restrictions on the form of the wave function. We can define so-called magnetic Bloch bands which are analogous to the bands of an

---

[37]We may further restrict the form of the wave function by requiring it to be annihilated by the b lowering operator, but it is not important to the discussion here.

electron in a lattice. Namely, they have a band index and are generally momentum-dependent. It turns out that the eigenstates of the Hamiltonian are familiarly of the form $\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r})$. We can then quantify the geometry of the Landau levels by computing the Berry curvature using the (magnetic) Bloch functions above. For all Landau levels, the Berry curvature is a constant $-l_B^2$ and the Chern number is simply $-\mathrm{sign}(1/l_B^2)$.[38] We direct the reader to [?] for a derivation of these results. There are a few important results that we emphasize. Namely, the Landau levels are flat bands (energy is momentum-independent), all Landau levels have constant, nonzero Berry curvature, and they are also Chern bands (non-zero Chern number). On the other hand, chiral MATBG also has flat Chern bands and nearly constant Berry curvature. It appears that their geometries match up very well. However, we are missing a key ingredient which we will now discuss.

5.5 Comparison of geometries

We have now derived both the wave functions of chiral MATBG and the LLL. We wish to compare the geometry of these wave functions but it is not completely clear how to do so. In the previous chapter, we introduced the Berry curvature as a tool to quantify the change in phase due to adiabatic transport through the Hilbert space. However, this completely ignores the change in magnitude which holds valuable information. We will now introduce the quantum geometric tensor, a single object which will capture both quantities. We follow the derivation in [?].

As in the Berry phase section, we consider a parameter-dependent non-degenerate Hamiltonian $H(\boldsymbol{\lambda})$ and consider a state $|\psi(\boldsymbol{\lambda})\rangle$ in the Hilbert space. We are in search of a metric on the Hilbert space that will allow us to define a notion of distance, capturing both differences in phase and magnitude. We expand the infinitesimal distance with respect to the parameter space

$$
\begin{aligned}
ds^2 &= |\psi(\boldsymbol{\lambda} + d\boldsymbol{\lambda}) - \psi(\boldsymbol{\lambda})|^2 \\
&= \langle \partial_\mu \psi | \partial_\nu \psi \rangle \, d\lambda^\mu d\lambda^\nu = (\gamma_{\mu\nu} + i\sigma_{\mu\nu})d\lambda^\mu d\lambda^\nu
\end{aligned}
\tag{137}
$$

where in the last step we simply defined $\gamma_{\mu\nu}$ and $\sigma_{\mu\nu}$ as the real and imaginary parts of $\langle \partial_\mu \psi | \partial_\nu \psi \rangle$ respectively. We often omit writing the dependence on the parameter $\boldsymbol{\lambda}$ when it is clear from context. Note that since $\langle \partial_\mu \psi | \partial_\nu \psi \rangle = \langle \partial_\nu \psi | \partial_\mu \psi \rangle^*$, we must have $\gamma_{\mu\nu} = \gamma_{\nu\mu}, \sigma_{\mu\nu} = -\sigma_{\nu\mu}$. Therefore, the distance reduces to

$$
ds^2 = \gamma_{\mu\nu}d\lambda^\mu d\lambda^\nu
\tag{138}
$$

as the anti-symmetry of $\sigma_{\mu\nu}$ eliminates it under summation. We might be quick to then claim that $\gamma_{\mu\nu}$ can serve as a metric, but it is not gauge-invariant. Indeed, the gauge transformation $|\psi'(\boldsymbol{\lambda})\rangle = e^{i\alpha(\boldsymbol{\lambda})}|\psi(\boldsymbol{\lambda})\rangle$ transforms the above quantities

$$
\gamma'_{\mu\nu} = \gamma_{\mu\nu} - \mathcal{A}_\mu \partial_\nu \alpha - \mathcal{A}_\nu \partial_\mu \alpha + \partial_\mu \alpha \partial_\nu \alpha, \qquad \sigma'_{\mu\nu} = \sigma_{\mu\nu}
\tag{139}
$$

where $\mathcal{A}_\mu = i\langle \psi | \partial_\mu \psi \rangle$ is the Berry connection and transforms as $\mathcal{A}'_\mu = \mathcal{A}_\mu + \partial_\mu \alpha$. By noting how the Berry connection transforms under gauge transformation, we define the symmetric gauge-invariant quantum metric

$$
g_{\mu\nu} = \gamma_{\mu\nu} - \mathcal{A}_\mu \mathcal{A}_\nu.
\tag{140}
$$

This quantum metric defines a notion of distance between states on the projected Hilbert space in which we compare states modulo their phase, so-called quantum rays. That is, we are quantifying

---

[38]The integration of the Berry curvature is carried out over the so-called magnetic Brillouin zone, a direct analog for the (lattice) Brillouin zone but defined using magnetic translation symmetry.

the distance between states in terms of changes in their magnitude rather than their phase. It is clear that if we are only considering the magnitudes of the states, then this quantity is manifestly gauge-invariant. On the other hand, $\gamma_{\mu\nu}$ measures distances between states in the usual sense including phase and magnitude. It is not too hard to verify that we can express

$$|\langle\psi(\boldsymbol{\lambda})|\psi(\boldsymbol{\lambda}+d\boldsymbol{\lambda})\rangle| = 1 - \frac{1}{2}g_{\mu\nu}d\lambda^\mu d\lambda^\nu. \tag{141}$$

To unify all the quantities, we define the quantum geometric tensor (QGT)

$$\eta_{\mu\nu} = \langle\partial_\mu\psi|\,(\sigma_0 - |\psi\rangle\,\langle\psi|)\,|\partial_\nu\psi\rangle \tag{142}$$

which captures all geometric information of the normalized band state $|\psi\rangle$. In particular, the previous quantities as well as the Berry curvature can be recovered

$$g_{\mu\nu} = \mathrm{Re}\eta_{\mu\nu}, \qquad \sigma_{\mu\nu} = \mathrm{Im}\eta_{\mu\nu}, \qquad \Omega_{\mu\nu} = -2\mathrm{Im}\eta_{\mu\nu} \tag{143}$$

and we see that $\sigma_{\mu\nu}$ is essentially the Berry curvature [?, ?].

We now apply the above formalism to capture the geometry of a band whose Bloch function u_k is holomorphic in $k = k_x + ik_y$. Due to holomorphicity, we have the relations $\partial_{k_x}u_k = (\partial_k + \bar{\partial}_k)u_k = \partial_k u_k$ ($\bar{\partial}u_k = 0$) and similarly $\partial_{k_y}u_k = i\partial_k u_k$ from which it follows that [?]

$$\eta(\mathbf{k}) \propto \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \tag{144}$$

Consequently, the holomorphicity of the wave functions of chiral MATBG and the LLL simplify the geometry of the Hilbert space to the above metric, namely they are proportional.[39] The requirement of proportionality of the QGT is actually much stronger than that which is required to realize FCI phases [?]. The true, weaker requirement is out of the scope of this paper but can be found in [?]. They are associated with matching the band geometry of the LLL, which as we mentioned, is known to exhibit fractional quantum Hall phases due to electron interactions. The requirements include flat Chern bands, sufficiently uniform Berry curvature, and the so-called determinant condition which relates the Berry curvature tensor to the quantum metric g and is implied by (144). Hence, all the conditions for the realization of FCI phases are met in chiral MATBG.

5.6 Flat band wave functions as generalized LLL

The similar band geometries of chiral MATBG and the LLL begs the question of whether there is a more direct relation between their physical origins. It seems to be a mathematical coincidence due to the presence of an anti-holomorphic operator in both cases. In this section, we show that the chiral MATBG wave function can be thought of as a generalized Landau level, that is a Landau level with non-constant magnetic field.

We consider an electron confined in two dimensions and subject to a transverse, inhomogenous magnetic field. This system typically forms dispersive so-called generalized Landau levels and has non-holomorphic eigenstates. However, both dispersivity (non-flat bands) and non-holomorphicity contradict our analysis of chiral MATBG. It turns out that there are two special physical models that do have these desirable properties: a Dirac particle and a particle in curved (Kähler) space.

---

[39]We direct the reader to [?] for a direct computation of the QGT of the LLL.

We consider the former since it allows for a more general form of the magnetic field [**?**]. The Dirac Hamiltonian with an inhomogenous magnetic field can be written as

$$H = \boldsymbol{\sigma} \cdot \boldsymbol{\pi} = 2 \begin{pmatrix} 0 & \pi \\ \overline{\pi} & 0 \end{pmatrix} \tag{145}$$

where $\pi = (\pi_x - i\pi_y)/2$, $\overline{\pi} = (\pi_x + i\pi_y)/2$. As usual $\boldsymbol{\pi} = \mathbf{p} - e\mathbf{A}$, $B = \nabla \times \mathbf{A}$, and we choose $c = 1$. $\boldsymbol{\sigma}$ represents a spin degree of freedom, but we may as well interpret it as a sublattice degree of freedom as we mentioned before when we discussed the pseudospin of graphene. We note that this Hamiltonian looks vaguely similar to that of chiral MATBG (119), and we will see that they also have similar (exact) eigenstates for an appropriate form of the magnetic field.

Expressed in position space, it is clear that $\overline{\pi} = -i\overline{\partial} - e\overline{A} = -i\overline{\partial} - e(A_x + iA_y)/2$ is an anti-holomorphic operator. As in the case of chiral MATBG, the eigenstates of H are sublattice-polarized since the zero mode $\pi\psi(\mathbf{r}) = 0$ immediately gives us $\overline{\pi}\psi^*(\mathbf{r}) = 0$. Therefore, we can find the LLL by computing the zero mode of $\overline{\pi}$, which turns out to be

$$\psi_{LLL}(\mathbf{r}) = f(z)e^{-K(\mathbf{r})/4l_{B_0}^2} \tag{146}$$

where $B(\mathbf{r}) = \hbar/e\nabla^2 K(\mathbf{r})$ and $f(z)$ is any holomorphic function [**?**]. One can match this equation to the chiral MATBG wave function we derived earlier to yield $K(\mathbf{r})$ for which the expressions become equal.[40] We have found that that the flat bands in chiral MATBG can be exactly replicated by a Dirac particle in an inhomogenous magnetic field. The magnetic field turns out to be periodic in the moire unit cell and immensely inhomogenous. There is debate about whether the formulation of the flat bands as a generalized LLL is a special feature of chiral MATBG or is a general property of materials which host FCI phases and satisfy the geometric requirement (144).[41]

5.7 Experimental evidence for FCI phases in MATBG

As of 2018, FCI phases have only been observed in AB Bernal stacked (untwisted) bilayer graphene aligned with hexagonal boron nitride (hBN) and an intense magnetic field which is responsible for flattening and creating Chern bands [**?**, **?**, **?**].[42] The discovery of flat bands in MATBG has made it a desirable material to host FCI phases without the need for an intense magnetic field. However, the analysis we carried out above is ideal in the sense that the chiral limit of MATBG is an analytic simplification and in experiment the ratio of hopping strengths w_0/w_0 is around 0.7 [**?**, **?**]. Nonetheless, the chiral limit provides us with a foundational understanding in the simple limit that we use to motivate the behavior at more mathematically complicated regimes. Most notably, the Berry curvature in laboratory MATBG is not as evenly spread out as in the chiral limit which may violate the requirements for a FCI phase [**?**]. A relatively weak magnetic field is used to spread out the Berry curvature.[43] We emphasize that the role of the magnetic field is not, as in the case of (untwisted) bilayer graphene, to create the Chern bands. In 2021, this setup for MATBG was carried out and eight FCI phases were observed at weak magnetic field [**?**]. Consequently, the study also motivated the realization of FCI phases at zero magnetic field, which have not yet been realized.

---

[40]We omit the expression of $K(\mathbf{r})$ for brevity.

[41]This formulation of the flat bands also allows us to understand how they behave in the presence of a physical magnetic field [**?**].

[42]hBN is responsible for gapping the bands since it violates space-inversion symmetry and introduces a sublattice potential on one of the layers.

[43]The magnetic field flattens the bands to spread out the Berry curvature.

6   Conclusions

In this paper, we provided an overview of graphene and explored recent literature on twisted bilayer graphene. The flat bands of magic-angle twisted bilayer graphene were central to our discussion and give rise to fractional Chern insulator phases. Nonetheless, there are many other interesting phases due to the strong interactions present in flat bands, and it would be a nice continuation of the material we presented here to study them as well. While we pursued numerical computations of the band structure, Berry curvature, and density of states of twisted bilayer graphene, we did not consider interactions in these calculations due to time considerations. Many of the articles cited in this paper directly considered numerical simulations of interactions, and so it appears essential to carry out such work. It seems to be a promising avenue to continue experimental work in realizing fractional Chern insulator phases in (non-chiral) magic-angle twisted bilayer graphene without the need for flatting out the Berry curvature.

7   Appendix

We dedicate this appendix to material which either did not fit in the footnotes or were not entirely necessary to understand the section. The code for the numerical computations including the band structure, density of states, and Berry curvature of TBG can be found at `https://github.com/amirsm02/mohammadi_haldane_JPI`.

7.1   Electron semiclassical equation of motion

We will heuristically derive the electron equation of motion used in (101). Recall the action from from classical mechanics

$$dS = \mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}, t)dt = \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}} \cdot \dot{\mathbf{r}} - H \right) dt = d\mathbf{r} \cdot \mathbf{p} - dtE \tag{147}$$

where we made the distinction $\mathbf{p} = \partial \mathcal{L}/\partial \dot{\mathbf{r}}$ and $H = E$. We use the ansantz for a (semiclassical) wave packet

$$\psi(\mathbf{r}, t) = \gamma(\mathbf{r}, t)e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} = \gamma(\mathbf{r}, t)e^{\frac{i}{\hbar}(\mathbf{p} \cdot \mathbf{r} - Et)} = \gamma(\mathbf{r}, t)e^{i\Phi(t)} \tag{148}$$

where $\gamma(\mathbf{r}, t)$ can be treated as a constant over distances much smaller than the de Broglie wavelength $2\pi/|\mathbf{k}|$ and over times much shorter than the period of oscillation $2\pi/\omega$. We notice the elegant connection between the classical action and the phase of this wave packet

$$d\Phi(t) = \frac{1}{\hbar}(d\mathbf{r} \cdot \mathbf{p} - dtE) = \frac{1}{\hbar}dS = \frac{1}{\hbar}dt\mathcal{L}. \tag{149}$$

A wave packet centered at position $\mathbf{r}$ and momentum $\mathbf{k}$ located in a lattice has form

$$|\mathbf{r}, \mathbf{k}\rangle \approx e^{i\Phi_n(t)} |u_n(\mathbf{k})\rangle \tag{150}$$

We will now independently expand the phase of this wave packet in terms of quantities we are already familiar with. We assume the wave packet is in the presence of an electromagnetic field with four-potential $(\mathbf{A}, A_0)$ with magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ and electric field $\mathbf{E} = -\nabla A_0 - \partial_t \mathbf{A}$. There are four contributions to the phase: the usual dynamical phase due to the time-dependent Schrodinger equation, the Bloch phase, the Berry phase, and the phase due to the electromagnetic

gauge field, written in order

$$\Phi_n(t) = -\frac{1}{\hbar} \int dt E_{n,\mathbf{k}} + \int d\mathbf{r} \cdot \mathbf{k} + \int d\mathbf{k} \cdot \mathcal{A}_n(\mathbf{k}) + \frac{e}{\hbar} \int (-d\mathbf{r} \cdot \mathbf{A}(\mathbf{r}) + dt A_0(\mathbf{r})) \qquad (151)$$

which gives us the Lagrangian $\mathcal{L}(\mathbf{k}, \frac{d\mathbf{k}}{dt}, \mathbf{r}, \frac{d\mathbf{r}}{dt}, t) = \hbar d\Phi_n/dt$. We now apply the Euler-Lagrange equation for $\mathbf{k}$

$$0 = \frac{\partial \mathcal{L}}{\partial \mathbf{k}} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{k}}} = \hbar\frac{d\mathbf{r}}{dt} - \frac{dE_{n,\mathbf{k}}}{d\mathbf{k}} + \hbar\frac{d\mathbf{k}}{dt} \times \Omega^n(\mathbf{k}) \qquad (152)$$

and we arrive and the intended result [**?**]

$$\frac{d\mathbf{r}}{dt} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E_{\mathbf{k}} - \frac{d\mathbf{k}}{dt} \times \Omega^n(\mathbf{k}), \qquad \hbar\frac{d\mathbf{k}}{dt} = -e\mathbf{E} - e\frac{d\mathbf{r}}{dt} \times \mathbf{B} \qquad (153)$$

where the second equation comes from the Euler-Lagrange equation for $\mathbf{r}$ and is simply the Lorentz force equation of motion. Notice that this correspondence remarkably implies that the Berry curvature and the Fermi velocity $v_F = \frac{1}{\hbar}\nabla_{\mathbf{k}} E_{\mathbf{k}}$ take the place of the magnetic field and electric field in momentum space.

7.2   Localization of Wannier orbitals

Localization of Wannier orbitals are a sought-after trait of systems. It makes numerical calculations converge exponentially and in the absence of an analytic proof of its existence, it is typically numerically suggested or assumed nonetheless ([**?**]). Wannier orbitals are exponentially localized if and only if the Bloch functions $u_{n,\mathbf{k}}(\mathbf{r})$ are analytic with respect to $\mathbf{k}$ [**?**]. In 1D systems, it has been proven that $u_{n,\mathbf{k}}(\mathbf{r})$ can always be constructed to be analytic, and therefore Wannier orbitals can also be constructed to be exponentially localized. In general, the localization of the Wannier orbitals is directly related to the regularity, or smoothness, of the Bloch functions (i.e. more regularity implies more localized). In 2D and 3D, it can be shown that the Bloch functions are analytic, and therefore the Wannier orbitals are exponentially localized if and only if the Chern numbers of all bands are trivial. This immediately suggests that time-reversal symmetry implies exponential localization since all Chern numbers are required to be trivial. On the other hand, Chern insulators have non-trivial Chern numbers by definition and therefore have non-localized Wannier orbitals, or in other words, it is impossible to construct exponentially localized Wannier orbitals. This captures the essence of why bands with non-trivial Chern number are called topological. Their behavior cannot be captured using local orbitals and can only be explained using their global properties.