Contents

List of Figures

List of Tables

Chapter 4

Electronic Band Structure of Crystals

4.1 Energy Bands in Solids

4.1.1 Bloch's theorem

The Hamiltonian for an electron in a crystal is

$$H = -\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2 + V(\boldsymbol{r}) \quad , \tag{4.1}$$

where $V(\boldsymbol{r}) = V(\boldsymbol{r}+\boldsymbol{R})$ for all $\boldsymbol{R} \in \mathcal{L}$, where $\mathcal{L}$ is the direct Bravais lattice underlying the crystal structure. The potential $V(\boldsymbol{r})$ describes the crystalline potential due to the ions, plus the average (Hartree) potential of the other electrons. The lattice translation operator is $t(\boldsymbol{R}) = \exp(i\boldsymbol{R}\cdot\boldsymbol{p}/\hbar) = \exp(\boldsymbol{R}\cdot\boldsymbol{\nabla})$. Acting on any function of $\boldsymbol{r}$, we have

$$t(\boldsymbol{R})\,f(\boldsymbol{r}) = f(\boldsymbol{r}+\boldsymbol{R}) \quad . \tag{4.2}$$

Note that lattice translations are unitary, *i.e.* $t^\dagger(\boldsymbol{R}) = t^{-1}(\boldsymbol{R}) = t(-\boldsymbol{R})$, and they satisfy the composition rule $t(\boldsymbol{R}_1)\,t(\boldsymbol{R}_2) = t(\boldsymbol{R}_1+\boldsymbol{R}_2)$. Since $\big[t(\boldsymbol{R}), H\big] = 0$ for all Bravais lattice vectors $\boldsymbol{R}$, the Hamiltonian H and all lattice translations $t(\boldsymbol{R})$ may be simultaneously diagonalized. Let $\psi(\boldsymbol{r})$ be such a common eigenfunction. Since $t(\boldsymbol{R})$ is unitary, its eigenvalue must be a phase $\exp(i\theta_{\boldsymbol{R}})$, and as a consequence of the composition rule, we must have $\theta_{\boldsymbol{R}_1+\boldsymbol{R}_2} = \theta_{\boldsymbol{R}_1} + \theta_{\boldsymbol{R}_2}$. This requires that $\theta_{\boldsymbol{R}}$ be linear in each of the components of $\boldsymbol{R}$, *i.e.* $\theta_{\boldsymbol{R}} = \boldsymbol{k}\cdot\boldsymbol{R}$, where $\boldsymbol{k}$ is called the *wavevector*. Since $\exp(i\,\boldsymbol{G}\cdot\boldsymbol{R}) = 1$ for any $\boldsymbol{G} \in \hat{\mathcal{L}}$, *i.e.* for any reciprocal lattice vector (RLV) $\boldsymbol{G}$ in the reciprocal lattice $\widehat{\mathcal{L}}$, the wavevector is only defined modulo $\boldsymbol{G}$, which means that $\boldsymbol{k}$ may be restricted to the first Brillouin zone of the reciprocal lattice. The quantity $\hbar\boldsymbol{k}$ is called the *crystal momentum*. Unlike ordinary momentum $\boldsymbol{p}$, crystal momentum is only conserved modulo $\hbar\boldsymbol{G}$.

The energy E will in general depend on $\boldsymbol{k}$, but there may be several distinct energy eigenstates with the same value of $\boldsymbol{k}$. We label these different energy states by a discrete index n, called the

band index. Thus, eigenstates of H are labeled by the pair $(n, \boldsymbol{k})$, with

$$H\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = E_n(\boldsymbol{k})\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \qquad , \qquad t(\boldsymbol{R})\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{4.3}$$

This is the content of *Bloch's theorem*. Note that the *cell function* $u_{n\boldsymbol{k}}(\boldsymbol{r}) \equiv \psi_{n\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$ is periodic in the direct lattice, with $u_{n\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}) = u_{n\boldsymbol{k}}(\boldsymbol{r})$. Thus, each Bloch function $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ may be written as the product of a plane wave and a cell function, *viz.*

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \equiv \langle\,\boldsymbol{r}\,|\,n\boldsymbol{k}\,\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad . \tag{4.4}$$

We may always choose the Bloch functions to be periodic in the reciprocal lattice, *i.e.* $\psi_{n,\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}}(\boldsymbol{r})$. This choice entails the condition $u_{n,\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{r}) = u_{n\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}$. However, there is no guarantee that $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ is continuous as a function of $\boldsymbol{k} \in \mathbb{T}^d$. As an example, consider the one-dimensional Bloch function $\psi_{nk}(x) = L^{-1/2}\,e^{i(G_n+k)x}$, where n labels the reciprocal lattice vector $G_n = 2\pi n/a$. If $\psi_{nk}(x)$ is taken to be continuous as a function of k, then clearly $\psi_{n,k+G_m}(x) = \psi_{n,k}(x)\,e^{iG_m x} = \psi_{n+m,k}(x) \neq \psi_{nk}(x)$.

The Bloch states, being eigenstates of a Hermitian operator, satisfy the conditions of completeness,

$$\sum_{n,\boldsymbol{k}} |\,n\boldsymbol{k}\,\rangle\langle\,n\boldsymbol{k}\,| = 1 \quad , \tag{4.5}$$

and orthonormality,

$$\langle\,n\boldsymbol{k}\,|\,n'\boldsymbol{k}'\,\rangle = \delta_{nn'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \tag{4.6}$$

Here we have assumed quantization of $\boldsymbol{k}$ in a large box of dimensions $L_1 \times L_2 \times L_3$. Each allowed wavevector then takes the form $\boldsymbol{k} = \left(\frac{2\pi n_1}{L_1}, \frac{2\pi n_2}{L_2}, \frac{2\pi n_3}{L_3}\right)$, where $n_{1,2,3} \in \mathbb{Z}$. In the thermodynamic limit, where $L_{1,2,3} \to \infty$, we have

$$\sum_{\boldsymbol{k}} \longrightarrow N v_0 \int \frac{d^d\boldsymbol{k}}{(2\pi)^d} \quad , \tag{4.7}$$

where d is the dimension of space ($d = 3$ unless otherwise noted), v_0 is the unit cell volume in real space, *i.e.* the volume of the Wigner-Seitz (WS) cell[1], and $N = L_1 \cdots L_d / v_0$ is the number of unit cells in the system, and is assumed to be thermodynamically large. Thus, we have from Eqn. 4.5,

$$\delta(\boldsymbol{r}-\boldsymbol{r}') = N v_0 \sum_n \int \frac{d^d\boldsymbol{k}}{(2\pi)^d}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r})\,\psi^*_{n\boldsymbol{k}}(\boldsymbol{r}') \quad . \tag{4.8}$$

One can see how the above equation is true in the simple case where $u_{n\boldsymbol{k}}(\boldsymbol{r}) = 1$ and $\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = V^{-1/2}\,e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, with $V = N v_0$.

[1]Elsewhere in these notes we denote the WS cell volume by Ω and the Brillouin zone volume by $\hat{\Omega}$.

4.1.2 Schrödinger equation

The potential $V(\boldsymbol{r})$ has a discrete Fourier representation as

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \quad , \tag{4.9}$$

where the sum is over all reciprocal lattice vectors $\boldsymbol{G} \in \widehat{\mathcal{L}}$. Since $V(\boldsymbol{r})$ is a real function, we must have $V_{-\boldsymbol{G}} = V_{\boldsymbol{G}}^*$. Any Bloch function $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ may also be written as a Fourier sum, *viz.*

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}}(\boldsymbol{k})\, e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}} \quad . \tag{4.10}$$

If we choose the Bloch functions to be periodic in the reciprocal lattice, then $C_{\boldsymbol{G}}(\boldsymbol{k}) = C(\boldsymbol{G}+\boldsymbol{k})$ is a function of $\boldsymbol{G}+\boldsymbol{k}$. Here, we have suppressed the band index n, and the wavevector $\boldsymbol{k} \in \hat{\Omega}$ must lie within the first Brillouin zone. The Schrödinger equation $H\psi_{\boldsymbol{k}}(\boldsymbol{r}) = E(\boldsymbol{k})\,\psi_{\boldsymbol{k}}(\boldsymbol{r})$ then takes the form

$$\frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m}\, C_{\boldsymbol{G}}(\boldsymbol{k}) + \sum_{\boldsymbol{G}'} V_{\boldsymbol{G}-\boldsymbol{G}'}\, C_{\boldsymbol{G}'}(\boldsymbol{k}) = E(\boldsymbol{k})\, C_{\boldsymbol{G}}(\boldsymbol{k}) \quad . \tag{4.11}$$

Note that we have one such equation for each wavevector $\boldsymbol{k} \in \hat{\Omega}$. This equation can be written in matrix form, as

$$\sum_{\boldsymbol{G}'} H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})\, C_{\boldsymbol{G}'}(\boldsymbol{k}) = E(\boldsymbol{k})\, C_{\boldsymbol{G}}(\boldsymbol{k}) \quad , \tag{4.12}$$

where, for each $\boldsymbol{k}$, $H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})$ is an infinite rank matrix,

$$H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}) = \frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m}\,\delta_{\boldsymbol{G}\boldsymbol{G}'} + V_{\boldsymbol{G}-\boldsymbol{G}'} \quad , \tag{4.13}$$

whose rows and columns are indexed by reciprocal lattice vectors $\boldsymbol{G}$ and $\boldsymbol{G}'$, respectively. The solutions, for any fixed value of $\boldsymbol{k}$, are then labeled by a band index n, hence

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \overbrace{\left(\sum_{\boldsymbol{G}} C_{\boldsymbol{G}}^{(n)}(\boldsymbol{k})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}}\right)}^{\text{cell function } u_{n\boldsymbol{k}}(\boldsymbol{r})} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \quad . \tag{4.14}$$

Note how the cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is explicitly periodic under direct lattice translations $\boldsymbol{r} \to \boldsymbol{r}+\boldsymbol{R}$. Note also that $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is an eigenfunction of the unitarily transformed Hamiltonian

$$\begin{aligned} H(\boldsymbol{k}) \equiv e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\, H\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} &= \frac{(\boldsymbol{p}+\hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r}) \\ H(\boldsymbol{k})\, u_{n\boldsymbol{k}}(\boldsymbol{r}) &= E_n(\boldsymbol{k})\, u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad . \end{aligned} \tag{4.15}$$

4.1.3 $V = 0$: empty lattice

Consider the case of $d = 1$ with $V = 0$, *i.e.* an empty lattice. We can read off the eigenvalues of $H_{GG'}$ from Eqn. 4.13: $E_{nk} = \hbar^2(G+k)^2/2m$, where the band index n identifies the reciprocal lattice vector $G = 2\pi n/a$, where a is the lattice spacing[2]. The first Brillouin zone $\hat{\Omega}$ is the region $k \in \left[-\frac{\pi}{a}\,,\,\frac{\pi}{a}\right]$. Fig. 4.1 shows how the usual ballistic dispersion $E(q) = \hbar^2 q^2/2m$ is "folded" into the first Brillouin zone by translating sections by integer multiples of the primary reciprocal lattice vector $b \equiv 2\pi/a$.

4.1.4 Perturbation theory

Let's consider the case where the potential $V(\boldsymbol{r})$ is weak. This is known as the *nearly free electron* (NFE) model. The matrix form of the Hamiltonian $H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})$ is given by

$$H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}) = \begin{pmatrix} \frac{\hbar^2(\boldsymbol{G}_1+\boldsymbol{k})^2}{2m} + V_0 & V_{\boldsymbol{G}_1-\boldsymbol{G}_2} & V_{\boldsymbol{G}_1-\boldsymbol{G}_3} & \cdots \\ V^*_{\boldsymbol{G}_1-\boldsymbol{G}_2} & \frac{\hbar^2(\boldsymbol{G}_2+\boldsymbol{k})^2}{2m} + V_0 & V_{\boldsymbol{G}_2-\boldsymbol{G}_3} & \cdots \\ V^*_{\boldsymbol{G}_1-\boldsymbol{G}_3} & V^*_{\boldsymbol{G}_2-\boldsymbol{G}_3} & \frac{\hbar^2(\boldsymbol{G}_3+\boldsymbol{k})^2}{2m} + V_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad . \tag{4.16}$$

Suppose we perturb in the off-diagonal elements, going to second order in $V_{\boldsymbol{G}-\boldsymbol{G}'}$. We then obtain

$$E_{\boldsymbol{G}}(\boldsymbol{k}) = E^0_{\boldsymbol{G}}(\boldsymbol{k}) + V_0 + \sum_{\boldsymbol{G}'(\neq\boldsymbol{G})} \frac{\left|V_{\boldsymbol{G}-\boldsymbol{G}'}\right|^2}{E^0_{\boldsymbol{G}}(\boldsymbol{k}) - E^0_{\boldsymbol{G}'}(\boldsymbol{k})} + \mathcal{O}(V^3) \quad , \tag{4.17}$$

where the unperturbed eigenvalues are $E^0_{\boldsymbol{G}}(\boldsymbol{k}) = \hbar^2(\boldsymbol{G}+\boldsymbol{k})^2/2m$. Note that the term with $\boldsymbol{G}' = \boldsymbol{G}$ is excluded from the sum. Here and henceforth, we shall set $V_0 \equiv 0$. The denominator in the above sum can vanish if a $\boldsymbol{G}'$ can be found such that $E^0_{\boldsymbol{G}}(\boldsymbol{k}) = E^0_{\boldsymbol{G}'}(\boldsymbol{k})$. In this case, the calculation fails, and we must use degenerate perturbation theory.

[2]Since the lattice is empty, we can use any value for a we please. The eigenspectrum will be identical, although the labeling of the eigenstates will depend on a since this defines the size of the Brillouin zone.

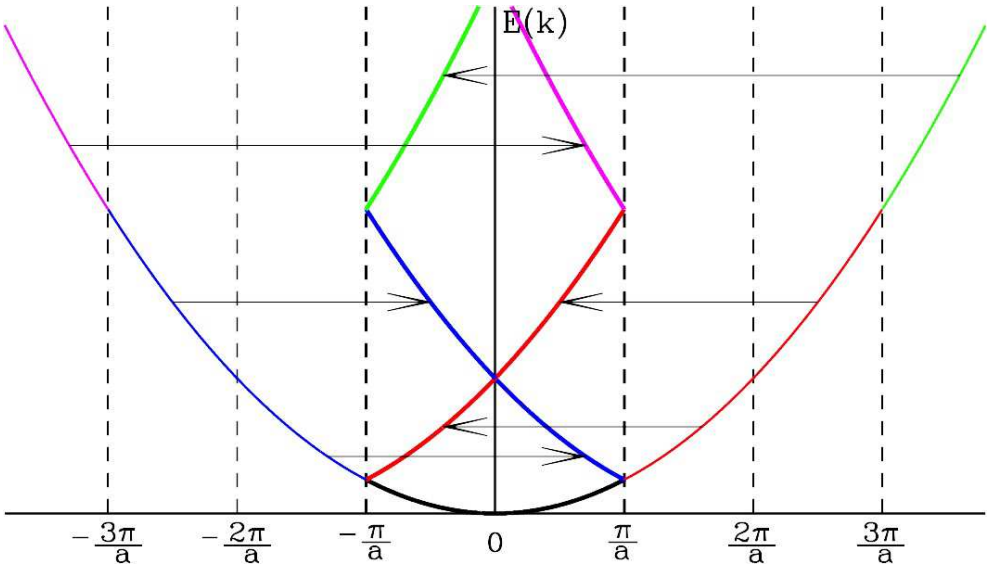

Figure 4.1: Band structure for an empty one-dimensional lattice, showing how the quadratic dispersion is "folded" from the extended zone picture into the first Brillouin zone.

Zone center (Γ), $\boldsymbol{G} = 0$

Let's first look in the vicinity of the zone center, labeled Γ, *i.e.* $\boldsymbol{k} \approx 0$. For the band associated with $\boldsymbol{G} = 0$, we have

$$\begin{aligned} E_0(\boldsymbol{k}) &= \frac{\hbar^2 \boldsymbol{k}^2}{2m} - \frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{\boldsymbol{G}^2 + 2\boldsymbol{G}\cdot\boldsymbol{k}} + \mathcal{O}(V^3) \\ &= \frac{\hbar^2 \boldsymbol{k}^2}{2m} - \frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^2} \left\{ 1 - \frac{2\,\boldsymbol{G}\cdot\boldsymbol{k}}{|\boldsymbol{G}|^2} + \frac{4(\boldsymbol{G}\cdot\boldsymbol{k})^2}{|\boldsymbol{G}|^4} + \ldots \right\} + \mathcal{O}(V^3) \quad . \end{aligned} \tag{4.18}$$

Since $V_{-\boldsymbol{G}} = V_{\boldsymbol{G}}^*$, the second term inside the curly bracket vanishes upon summation, so we have

$$E_0(\boldsymbol{k}) = \Delta + \tfrac{1}{2}\hbar^2 (m^*)^{-1}_{\mu\nu}\, k^\mu\, k^\nu + \ldots \quad , \tag{4.19}$$

where

$$\begin{aligned} \Delta &= -\frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^2} \\ (m^*)^{-1}_{\mu\nu} &= \frac{1}{m}\,\delta_{\mu\nu} - \left(\frac{4\sqrt{m}}{\hbar^2}\right)^2 \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^6}\, G^\mu\, G^\nu + \ldots \quad . \end{aligned} \tag{4.20}$$

Here Δ is the *band offset* relative to the unperturbed case, and $(m^*)^{-1}_{\mu\nu}$ are the components of the inverse *effective mass tensor*. Note that the dispersion in general is no longer isotropic. Rather,

the effective mass tensor $m^*_{\mu\nu}$ transforms according to a tensor representation of the crystallographic point group. For three-dimensional systems with cubic symmetry, m^* is a multiple of the identity, for the same reason that the inertia tensor of a cube is $I = \frac{1}{6}Ma^2\,\mathsf{diag}(1,1,1)$. But for a crystal with a tetragonal symmetry, in which one of the cubic axes is shortened or lengthened, the effective mass tensor along principal axes takes the general form $m^* = \mathsf{diag}(m_x, m_x, m_z)$, with $m_x \neq m_z$ in general.

Zone center (Γ), $\boldsymbol{G} = \pm\boldsymbol{b}_{1,2,3}$

Consider the cubic lattice with primitive direct lattice vectors $\boldsymbol{a}_j = a\,\hat{\mathbf{e}}_j$ and primitive reciprocal lattice vectors $\boldsymbol{b}_j = \frac{2\pi}{a}\,\hat{\mathbf{e}}_j$. For $\boldsymbol{k} = 0$, the six bands corresponding to $\boldsymbol{G} = \pm\boldsymbol{b}_j$ with $j \in \{1,2,3\}$ are degenerate, with $E^0_{\boldsymbol{G}} = 2\pi^2\hbar^2/ma^2$. Focusing only on these rows and columns, we obtain a 6×6 effective Hamiltonian,

$$H_{6\times6} = \begin{pmatrix} \frac{\hbar^2}{2m}(\boldsymbol{b}_1+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_1} & V_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V_{\boldsymbol{b}_1+\boldsymbol{b}_3} \\ V^*_{2\boldsymbol{b}_1} & \frac{\hbar^2}{2m}(\boldsymbol{b}_1-\boldsymbol{k})^2 & V^*_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V^*_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V^*_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_1-\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V_{\boldsymbol{b}_1+\boldsymbol{b}_2} & \frac{\hbar^2}{2m}(\boldsymbol{b}_2+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_2} & V_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V_{\boldsymbol{b}_2+\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V^*_{2\boldsymbol{b}_2} & \frac{\hbar^2}{2m}(\boldsymbol{b}_2-\boldsymbol{k})^2 & V^*_{\boldsymbol{b}_2+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2-\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V_{\boldsymbol{b}_2+\boldsymbol{b}_3} & \frac{\hbar^2}{2m}(\boldsymbol{b}_3+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2+\boldsymbol{b}_3} & V_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V^*_{2\boldsymbol{b}_3} & \frac{\hbar^2}{2m}(\boldsymbol{b}_3-\boldsymbol{k})^2 \end{pmatrix} . \tag{4.21}$$

To simplify matters, suppose that the only significant Fourier components $V_{\boldsymbol{G}}$ are those with $\boldsymbol{G} = \pm 2\boldsymbol{b}_j$. In this case, the above 6×6 matrix becomes block diagonal, *i.e.* a direct sum of 2×2 blocks, each of which resembles

$$H_{2\times2}(\Gamma) = \begin{pmatrix} \frac{\hbar^2}{2m}(\boldsymbol{b}_j+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_j} \\ V^*_{2\boldsymbol{b}_j} & \frac{\hbar^2}{2m}(\boldsymbol{b}_j-\boldsymbol{k})^2 \end{pmatrix} . \tag{4.22}$$

Diagonalizing, we obtain

$$E_{j,\pm}(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{b}_j^2}{2m} + \frac{\hbar^2\boldsymbol{k}^2}{2m} \pm \sqrt{\left(\frac{\hbar^2}{m}\,\boldsymbol{b}_j\cdot\boldsymbol{k}\right)^2 + |V_{2\boldsymbol{b}_j}|^2} \quad . \tag{4.23}$$

Assuming cubic symmetry with $V_{\boldsymbol{b}_1} = V_{\boldsymbol{b}_2} = V_{\boldsymbol{b}_3} = V$, we obtain six bands,

$$E_{j,\pm}(\boldsymbol{k}) = \frac{2\pi^2\hbar^2}{ma^2} + \frac{\hbar^2\boldsymbol{k}^2}{2m} \pm \sqrt{\left(\frac{2\pi\hbar^2}{ma}\,k_j\right)^2 + |V|^2} \quad . \tag{4.24}$$

The band gap at $\boldsymbol{k} = 0$ is then $2\,|V|$.

Zone edge (X), $\boldsymbol{G} = 0$

Consider now the case $\boldsymbol{k} = \frac{1}{2}\boldsymbol{b} + \boldsymbol{q}$ with $|\boldsymbol{q}a| \ll 1$, and the band $\boldsymbol{G} = 0$. This state is nearly degenerate with one in the band with $\boldsymbol{G} = -\boldsymbol{b}$. Isolating these contributions to $H_{\boldsymbol{G}\boldsymbol{G}'}$, we obtain the 2×2 matrix

$$H_{2\times 2}(\mathrm{X}) = \begin{pmatrix} \frac{\hbar^2}{2m}\big(\frac{1}{2}\boldsymbol{b} + \boldsymbol{q}\big)^2 & V_{-\boldsymbol{b}} \\ V_{\boldsymbol{b}} & \frac{\hbar^2}{2m}\big(-\frac{1}{2}\boldsymbol{b} + \boldsymbol{q}\big)^2 \end{pmatrix} \quad , \tag{4.25}$$

with dispersion

$$E_\pm(\boldsymbol{k}) = \frac{\hbar^2 \boldsymbol{b}^2}{8m} + \frac{\hbar^2 \boldsymbol{q}^2}{2m} \pm \sqrt{\left(\frac{\hbar^2}{2m}\,\boldsymbol{b}\cdot\boldsymbol{q}\right)^2 + |V_{\boldsymbol{b}}|^2} \quad . \tag{4.26}$$

The band gap is again $2\,|V_{\boldsymbol{b}}|$.

4.1.5 Solvable model : one-dimensional Dirac comb

Consider the one-dimensional periodic potential,

$$V(x) = -W_0 \sum_{n=-\infty}^{\infty} \delta(x - na) \tag{4.27}$$

with $W_0 > 0$. Define $W_0 \equiv \hbar^2/2m\sigma$, where σ is the *scattering length*. The Hamiltonian is then

$$H = -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m\sigma}\sum_n \delta(x - na) \quad . \tag{4.28}$$

The eigenstates of H must satisfy Bloch's theorem: $\psi_{nk}(x+a) = e^{ika}\,\psi_{nk}(x)$. Thus, we may write

$$\begin{aligned} x \in [-a\,,\,0] \quad &: \quad \psi_{nk}(x) = A\,e^{iqx} + B\,e^{-iqx} \\ x \in [0\,,\,+a] \quad &: \quad \psi_{nk}(x) = e^{ika}\,\psi_{nk}(x-a) \\ &\qquad\quad = A\,e^{i(k-q)a}\,e^{iqx} + B\,e^{i(k+q)a}\,e^{-iqx} \quad . \end{aligned} \tag{4.29}$$

Continuity at $x = 0$ requires $\psi_{nk}(0^-) = \psi_{nk}(0^+)$, or

$$A + B = A\,e^{i(k-q)a} + B\,e^{i(k+q)a} \quad . \tag{4.30}$$

A second equation follows from integrating the Schrödinger equation from $x = 0^-$ to $x = 0^+$:

$$\begin{aligned} \int_{0^-}^{0^+} dx\; H\,\psi_{nk}(x) &= \int_{0^-}^{0^+} dx\,\left\{-\frac{\hbar^2}{2m}\,\frac{d^2\psi_{nk}}{dx^2} - \frac{\hbar^2}{2m\sigma}\,\psi_{nk}(x)\,\delta(x)\right\} \\ &= \frac{\hbar^2}{2m}\Big[\psi'_{nk}(0^-) - \psi'_{nk}(0^+)\Big] - \frac{\hbar^2}{2m\sigma}\,\psi_{nk}(0) \quad . \end{aligned} \tag{4.31}$$

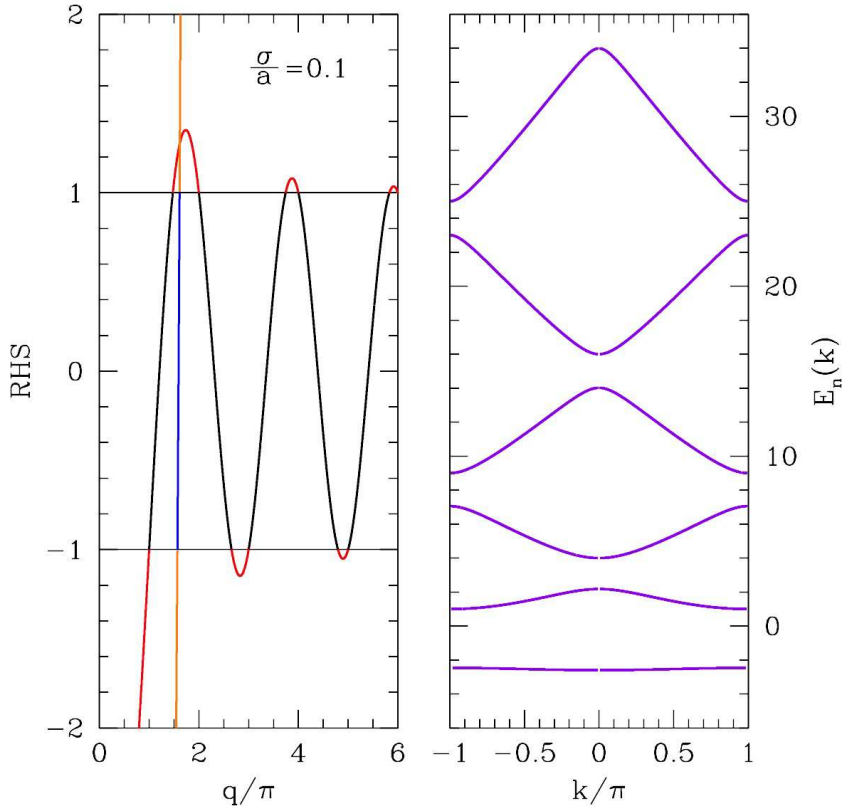

Figure 4.2: Left: Plots of the RHS of Eqns. 4.35 (black and red) and 4.36 (blue and orange) for the Dirac comb potential with scattering length $\sigma = 0.1\,a$. Allowed solutions (black and blue portions) must satisfy $\mathrm{RHS} \in [-1, 1]$. Right: Corresponding energy band structure.

Since $H\psi_{nk}(x) = E_n(k)\psi_{nk}(x)$, the LHS of the above equation is infinitesimal. Thus,

$$\psi'_{nk}(0^-) - \psi'_{nk}(0^+) = \frac{1}{\sigma}\,\psi_{nk}(0) \quad , \tag{4.32}$$

or

$$A - B - A\,e^{i(k-q)a} + Be^{i(k+q)a} = \frac{A+B}{iq\sigma} \quad . \tag{4.33}$$

The two independent equations we have derived can be combined in the form

$$\begin{pmatrix} e^{i(k-q)a} - 1 & e^{i(k+q)a} - 1 \\ e^{i(k-q)a} - 1 - \frac{i}{q\sigma} & 1 - e^{i(k+q)a} - \frac{i}{q\sigma} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = 0 \quad . \tag{4.34}$$

In order that the solution be nontrivial, we set the determinant to zero, which yields the condition

$$\cos(ka) = \cos(qa) - \frac{a}{2\sigma} \cdot \frac{\sin(qa)}{qa} \quad . \tag{4.35}$$

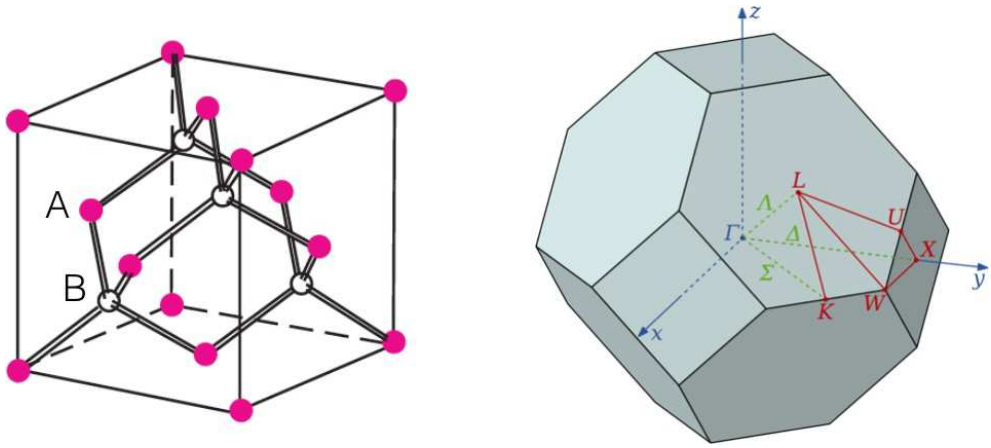

Figure 4.3: Left: The zincblende structure consists of two interpenetrating fcc lattices. Right: First Brillouin zone for the fcc lattice, with high symmetry points identified. From Wikipedia.

This is to be regarded as an equation for $q(k)$, parameterized by the dimensionless quantity σ/a. The energy eigenvalue is $E_n(k) = \hbar^2 q^2/2m$. If we set $q \equiv iQ$, the above equation becomes

$$\cos(ka) = \cosh(Qa) - \frac{a}{2\sigma} \cdot \frac{\sinh(Qa)}{Qa} \quad . \tag{4.36}$$

Here we solve for $Q(k)$, and the energy eigenvalue is $E_n(k) = -\hbar^2 Q^2/2m$. In each case, there is a discrete infinity of solutions indexed by the band index n. Results for the case $\sigma = 0.1\,a$ are shown in Fig. 4.2. In the limit $\sigma \to 0$, the solutions to Eqn. 4.35 are $q = k + \frac{2\pi n}{a}$, and we recover the free electron bands in the reduced zone scheme. When $\sigma = 0$, the only solution to Eqn. 4.36 is $Q = 0$ for the case $k = 0$.

As $q \to \infty$, the second term on the RHS of Eqn. 4.35 becomes small, and the solution for the $n^{\rm th}$ band ($n \in \mathbb{Z}_+$) tends to $q = k + 2\pi n/a$. The band gaps at $k = 0$ and $k = \pi$ become smaller and smaller with increasing band index. Clearly Eqn. 4.36 has no solutions for sufficiently large Q, since the RHS increases exponentially. Note that there is one band in the right panel of Fig. 4.2 with negative energy. This is because we have taken the potential $V(x)$ to be attractive. Recall that the potential $V(x) = -W_0\,\delta(x)$ has a single bound state $\psi_0(x) = \frac{1}{2\sqrt{\sigma}}\,e^{-|x|/2\sigma}$, again with $\sigma \equiv \hbar^2/2mW_0$. For the Dirac comb, the bound states in different unit cells overlap, which leads to dispersion. If $W_0 < 0$, the potential is purely repulsive, and all energy eigenvalues are positive. (There is no solution to Eqn. 4.36 when $\sigma < 0$.)

4.1.6 Diamond lattice bands

In dimensions $d > 1$, the essential physics is similar to what was discussed in the case of the NFE model, but the labeling of the bands and the wavevectors is more complicated than in the $d = 1$ case. Consider the zincblende structure depicted in the left panel of Fig. 4.3. Zincblende

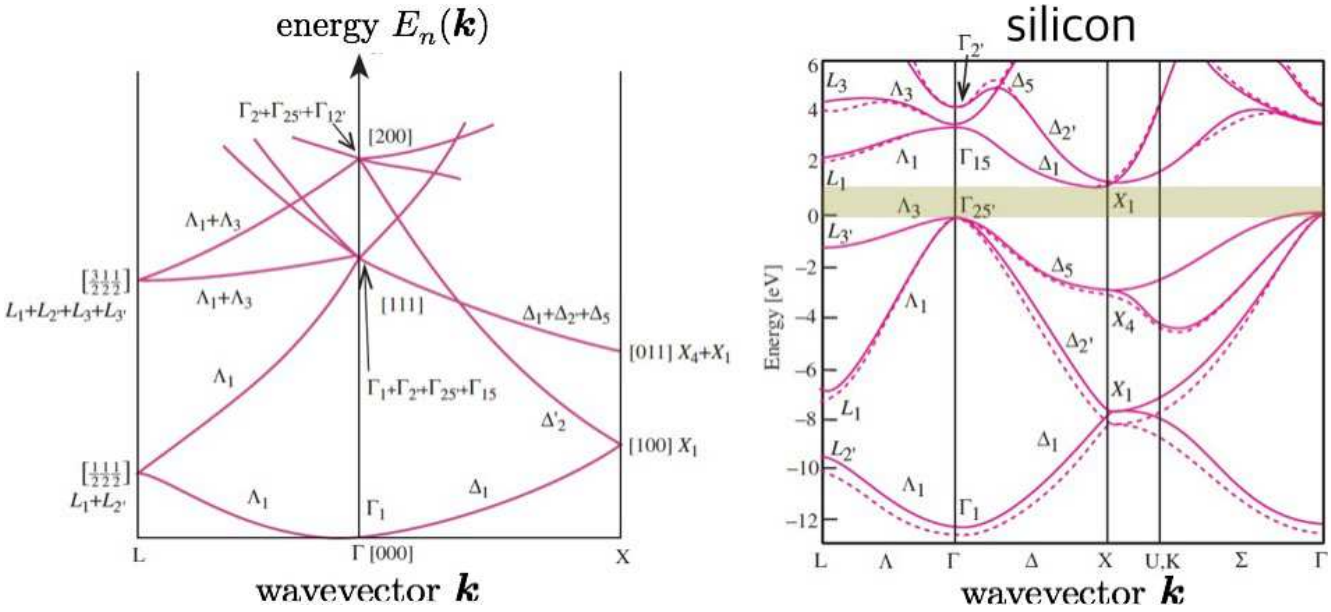

Figure 4.4: Left: Empty lattice s-orbital bands for the diamond structure. Right: Electronic band structure of Si based on local (dashed) and nonlocal (solid) pseudopotential calculations. The shaded region contains no electronic eigenstates, and reflects a (indirect) *band gap*. From ch. 2 of P. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, 1996).

consists of two interpenetrating face centered cubic lattices (labeled A and B in the figure), and is commonly found in nature (*e.g.*, in GaAs, InP, ZnTe, ZnS, HgTe, CdTe, *etc.*). Diamond is a homonuclear form of zincblende in which the ions on the two fcc sublattices are identical; the most familiar examples are C (carbon diamond) and Si. For both zincblende and diamond, the underlying Bravais lattice is fcc; the first Brillouin zone of the fcc lattice is depicted in the right panel of Fig. 4.3.

The left panel of Fig. 4.4 depicts the empty lattice (free electron) energy bands for the diamond structure along linear segments $\mathrm{L}\Gamma$ and $\Gamma\mathrm{X}$ (see Fig. 4.3 for the letter labels of high symmetry points). Energy levels at high symmetry points are labeled by reciprocal lattice vectors (square brackets) in the extended zone scheme, all in units of $2\pi/a$, where a is the size of the unit cube in Fig. 4.3. The other labels denote group representations under which the electronic eigenstates transform. In the extended zone scheme, the dispersion is $E(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$, so all the branches of the dispersion in the reduced zone scheme correspond to displacements of sections of this paraboloid by RLVs.

The right panel of Fig. 4.4 depicts the energy bands of crystalline silicon (Si), which has the diamond lattice structure. Notice how the lowest L-point energy levels are no longer degenerate. A gap has opened, as we saw in our analysis of the NFE model. Indeed, between $E = 0$ and $E = 1.12\,\mathrm{eV}$, there are *no* electronic energy eigenstates. This is the 'band gap' of silicon. Note that the $1.12\,\mathrm{eV}$ gap is *indirect* – it is between states at the Γ and X points. If the minimum en-

material	gap (eV)	type	material	gap	type	material	gap	type
C	5.47	indirect	Si	1.14	indirect	h-BN	5.96	direct
Ge	0.67	indirect	Sn	$\lesssim 0.08$	indirect	AlN	6.28	direct
GaN	3.44	direct	InN	0.7	direct	ZnO	3.37	direct
GaAs	1.43	direct	InP	1.35	direct	ZnSe	2.7	direct
GaP	2.26	indirect	InAs	0.36	direct	ZnS	3.54	direct
GaSb	0.726	direct	InSb	0.17	direct	ZnTe	2.25	direct
CdS	2.42	direct	CdTe	1.49	direct	Cu_2S	1.2	indirect

Table 4.1: Common semiconductors and their band gaps. From Wikipedia.

ergy gap occurs between levels at the same wavevector, the gap is said to be *direct*. In intrinsic semiconductors and insulators, *transport* measurements typically can provide information on indirect gaps. *Optical* measurements, however, reveal direct gaps. The reason is that the speed of light is very large, and momentum conservation requires optical transitions to be essentially vertical in $(\boldsymbol{k}, E)$ space.

4.2 Metals and Insulators

4.2.1 Density of states

In addition to energy eigenstates being labeled by band index ν and (crystal) wavevector $\boldsymbol{k}$, they are also labeled by spin polarization $\sigma = \pm 1$ relative to some fixed axis in internal space (typically $\hat{\boldsymbol{z}}$)[3]. The component of the spin angular momentum along $\hat{\boldsymbol{z}}$ is then $S^z = \frac{1}{2}\hbar\sigma = \pm\frac{1}{2}\hbar$. Typically, $E_\nu(\boldsymbol{k}, \sigma)$ is independent of the spin polarization σ, but there are many examples where this is not the case[4].

The *density of states* (DOS) per unit energy per unit volume, $g(\varepsilon)$, is given by

$$g(\varepsilon) = \frac{1}{V}{\sum_{\nu,\boldsymbol{k},\sigma}}' \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \overset{V\to\infty}{=} \sum_\nu \sum_\sigma \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d}\, \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \quad . \tag{4.37}$$

Here we assume box quantization with $\boldsymbol{k} = \left(\frac{2\pi j_1}{L_1}, \ldots, \frac{2\pi j_d}{L_d}\right)$, where j_1 *etc.* are all integers. The volume associated with each point in $\boldsymbol{k}$ space is then $\Delta V = (2\pi/L_1)\cdots(2\pi/L_d) = (2\pi)^d/V$,

[3]Here we denote the band index as ν, to obviate confusion with the occupancy n below.
[4]If there is an external magnetic field H, for example, the energy levels will be spin polarization dependent.

which establishes the above equality in the thermodynamic limit. We can also restrict our attention to a particular band ν and spin polarization σ, and define

$$g_{\nu\sigma}(\varepsilon) = \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d}\, \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \quad . \tag{4.38}$$

Finally, we may multiply by the real space unit cell volume v_0 to obtain $\overline{g}(\varepsilon) \equiv v_0\, g(\varepsilon)$, which has dimensions of inverse energy, and gives the number of levels per unit energy per unit cell.

Examples

Consider the case of a one-dimensional band with dispersion $E(k) = -2t\cos(ka)$. The density of states per unit cell is

$$\overline{g}(\varepsilon) = a\int\limits_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{dk}{2\pi}\, \delta(\varepsilon + 2t\cos ka) = \frac{1}{\pi}\left(B^2 - \varepsilon^2\right)^{-1/2} \Theta(B^2 - \varepsilon^2) \quad , \tag{4.39}$$

where $B = 2t$ is half the bandwidth. *I.e.* the allowed energies are $\varepsilon \in [-B, +B]$. Note the square root singularity in $\overline{g}_{d=1}(\varepsilon)$ at the band edges.

Now let's jump to d space dimensions, and the dispersion $E(\boldsymbol{k}) = -2t\sum_{i=1}^d \cos(k_i a)$. The DOS per unit cell is then

$$\overline{g}_d(\varepsilon) = \int\limits_{-\pi}^{\pi} \frac{d\theta_1}{2\pi} \cdots \int\limits_{-\pi}^{\pi} \frac{d\theta_d}{2\pi}\, \delta\big(\varepsilon + 2t\cos\theta_1 + \cdots + 2t\cos\theta_d\big) = \frac{1}{\pi}\int\limits_0^\infty du\, \cos(\varepsilon u)\, \big[J_0(2tu)\big]^d \quad , \tag{4.40}$$

where each $\theta_j = k_j a$, and we have invoked an integral representation of the Dirac δ-function. Here $J_0(x)$ is the ordinary Bessel function of the first kind. Since $\int_{-\infty}^{\infty} d\varepsilon\, \cos(\varepsilon u) = 2\pi\,\delta(u)$, it is easy to see that $\int_{-\infty}^{\infty} d\varepsilon\, \overline{g}(\varepsilon) = 1$, *i.e.* that the DOS is correctly normalized. For $d = 2$, the integral may be performed to yield

$$\overline{g}_2(\varepsilon) = \frac{2}{\pi^2 B}\, \mathsf{K}\Big(\sqrt{1 - (\varepsilon/B)^2}\Big)\, \Theta(B^2 - \varepsilon^2) \quad , \tag{4.41}$$

where

$$\mathsf{K}(k) = \int\limits_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2\sin^2\theta}} \tag{4.42}$$

is the complete elliptic integral of the first kind[5], and $B = 4t$ is the half bandwidth. The function $\overline{g}_2(\varepsilon)$ has a logarithmic singularity at the band center $\varepsilon = 0$, called a *van Hove singularity*.

[5]There is an unfortunate notational variation in some sources, which write $K(m)$ in place of $\mathsf{K}(k)$, where $m = k^2$.

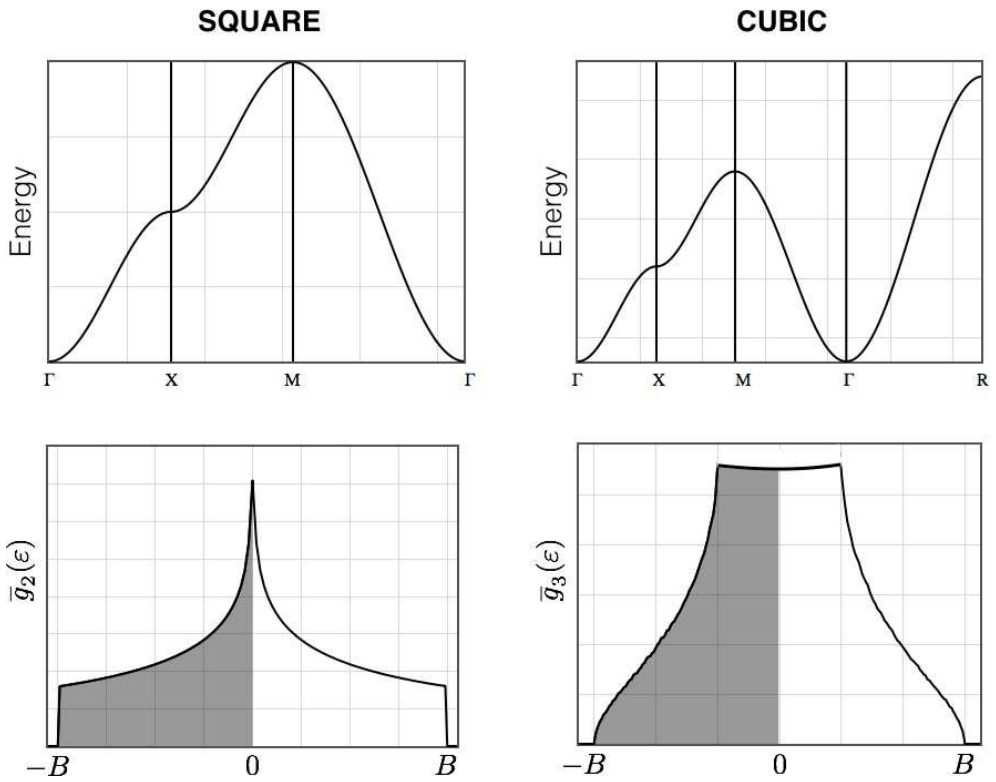

Figure 4.5: Upper left: Two-dimensional dispersion $E(k_x, k_y)$ along high symmetry lines in the 2D square lattice first Brillouin zone. Lower left: Corresponding density of states $\overline{g}_2(\varepsilon)$. Upper right: Three-dimensional dispersion $E(k_x, k_y, k_z)$ along high symmetry lines in the 2D cubic lattice first Brillouin zone. Lower right: Corresponding density of states $\overline{g}_3(\varepsilon)$. Shaded regions show occupied states for a lattice of s-orbitals with one electron per site. Figures from http://lampx.tugraz.at/∼hadley/ss1/bands/tbtable/tbtable.html .

The results for $d = 2$ and $d = 3$ are plotted in Fig. 4.5. The logarithmic van Hove singularity at $\varepsilon = 0$ is apparent in $\overline{g}_2(\varepsilon)$. The function $\overline{g}_3(\varepsilon)$ has van Hove singularities at $\varepsilon = \pm\frac{1}{3}B$, where the derivative $g_3'(\varepsilon)$ is discontinuous. In the limit $d \to \infty$, we can use the fact that $J_0(x) = 1 - \frac{1}{4}x^2 + \ldots$ to extract

$$\overline{g}_{d\gg1}(\varepsilon) = \sqrt{\frac{d}{\pi B^2}}\, e^{-d\varepsilon^2/B^2} = (4\pi dt^2)^{-1/2}\, \exp\bigl(-\varepsilon^2/4dt^2\bigr) \quad . \tag{4.43}$$

We recognize this result as the Central Limit Theorem in action. With $E(\boldsymbol{k}) = -2t\sum_{i=1}^{d}\cos\theta_i$ and θ_i uniformly distributed along $[-\pi, \pi]$, the standard deviation σ is given by

$$\sigma^2 = (2t)^2 \times d \times \langle\cos^2\theta\rangle = 2dt^2 \quad , \tag{4.44}$$

exactly as in Eqn. 4.43

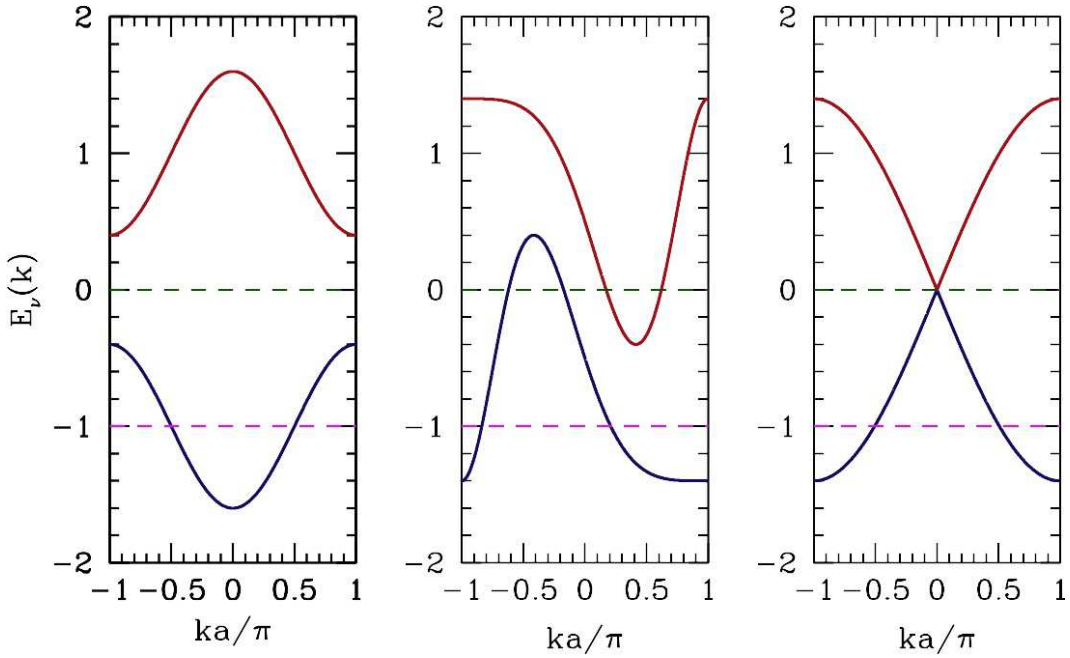

Figure 4.6: Three one-dimensional band structures. Valence bands are shown in dark blue, and conduction bands in dark red. Left: Non-overlapping bands with forbidden region $E \in [-0.4, 0.4]$. Center: At each point there is a direct gap, but the indirect gap is negative and there is no forbidden region. Right: Linear crossing leading to cusp-like band touching. In all cases, when $\varepsilon_{\rm F} = -1$ (dashed magenta line), the Fermi energy cuts through the bottom band and the system is a metal. When $\varepsilon_{\rm F} = 0$ (dashed green line), the system at the left is an insulator, with $g(\varepsilon_{\rm F}) = 0$. The system in the middle is a metal, with $g(\varepsilon_{\rm F}) > 0$ The system on the right also has a finite density of states at $\varepsilon_{\rm F}$, but in two space dimensions such a diabolical point in the dispersion, where $\varepsilon(\boldsymbol{k}) = \pm\hbar v_{\rm F}|\boldsymbol{k}|$ results in a continuously vanishing density of states $g(E)$ as $E \to \varepsilon_{\rm F} = 0$. Such a system is called a *semimetal*.

Band edge behavior

In the vicinity of a quadratic band minimum, along principal axes of the effective mass tensor, we have

$$E(\boldsymbol{k}) = \Delta + \sum_{i=1}^{d} \frac{\hbar^2 k_i^2}{2m_i^*} \quad , \tag{4.45}$$

and the density of states is

$$\begin{aligned} \overline{g}(\varepsilon) &= v_0 \int \frac{dk_1}{2\pi} \cdots \int \frac{dk_d}{2\pi}\, \delta\left(\varepsilon - \Delta - \sum_i \frac{\hbar^2 k_i^2}{2m_i^*}\right) \\ &= \frac{v_0}{2} \cdot \frac{\sqrt{2m_1^*}}{h} \cdots \frac{\sqrt{2m_d^*}}{h}\, \Omega_d \, (\varepsilon - \Delta)^{\frac{d}{2}-1} \quad , \end{aligned} \tag{4.46}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in d space dimensions. Consistent with Fig. 4.5, $g_{d=2}(\varepsilon)$ tends to a constant at the band edges, and then discontinuously drops to zero as one exits the band. In $d = 3$, $\overline{g}_{d=3}(\varepsilon)$ vanishes as $(\varepsilon - \Delta)^{1/2}$ at a band edge.

4.2.2 Fermi statistics

If we assume the electrons are noninteracting[6], the energy of the state for which the occupancy of state $|\nu \boldsymbol{k}\sigma\rangle$ is $n_{\nu\boldsymbol{k}\sigma}$ is

$$E\big[\{n_{\nu\boldsymbol{k}\sigma}\}\big] = \sum_\nu \sum_\sigma {\sum_{\boldsymbol{k}}}' E_\nu(\boldsymbol{k},\sigma)\, n_{\nu\boldsymbol{k}\sigma} \quad . \tag{4.47}$$

The Pauli exclusion principle tells us that a given electronic energy level can accommodate at zero or one fermion, which means each $n_{\nu\boldsymbol{k}\sigma}$ is either 0 or 1. At zero temperature, the N electron ground state is obtained by filling up all the energy levels starting from the bottom of the spectrum, with one electron per level, until the lowest N such levels have been filled. In $E_\nu(\boldsymbol{k},\sigma)$ is independent of σ, then there will be a twofold *Kramers degeneracy* whenever N is odd, as the last level filled can either have $\sigma = +1$ or $\sigma = -1$[7].

At finite temperature $T > 0$, the thermodynamic average of $n_{\nu\boldsymbol{k}\sigma}$ is given, within the grand canonical ensemble, by

$$\big\langle n_{\nu\boldsymbol{k}\sigma}\big\rangle = \frac{1}{\exp\Big(\frac{E_\nu(\boldsymbol{k},\sigma)-\mu}{k_{\rm B}T}\Big)+1} \equiv f\big(E_\nu(\boldsymbol{k},\sigma)-\mu\big) \quad , \tag{4.48}$$

where $f(x)$ is the *Fermi function*,

$$f(x) = \frac{1}{e^{x/k_{\rm B}T}+1} \quad . \tag{4.49}$$

The total electron number density is then

$$n(T,\mu) = \frac{N}{V} = \int\limits_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, f(\varepsilon-\mu) = \sum_\nu \sum_\sigma \int \frac{d^dk}{(2\pi)^d} \frac{1}{\exp\Big(\frac{E_\nu(\boldsymbol{k},\sigma)-\mu}{k_{\rm B}T}\Big)+1} \quad . \tag{4.50}$$

This is a Gibbs-Duhem relation, involving the three intensive quantities (n, T, μ). In principle it can be inverted to yield the chemical potential $\mu(n,T)$ as a function of number density and temperature. When $T = 0$, we write $\mu(n, T=0) \equiv \varepsilon_{\rm F}$, which is the *Fermi energy*. Since the Fermi function becomes $f(x) = \Theta(-x)$ at zero temperature, we have

$$n(\varepsilon_{\rm F}) = \int\limits_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\, g(\varepsilon) \quad . \tag{4.51}$$

[6]Other, that is, than the mean "Hartree" contribution to the potential $V(\boldsymbol{r})$.

[7]There can be additional degeneracies. For example, in $d = 1$ if, suppressing the band index, $E(k) = E(-k)$, then each level with $k \neq 0$ and $k \neq \pi/a$ is fourfold degenerate: $(k\uparrow, k\downarrow, -k\uparrow, -k\uparrow)$.

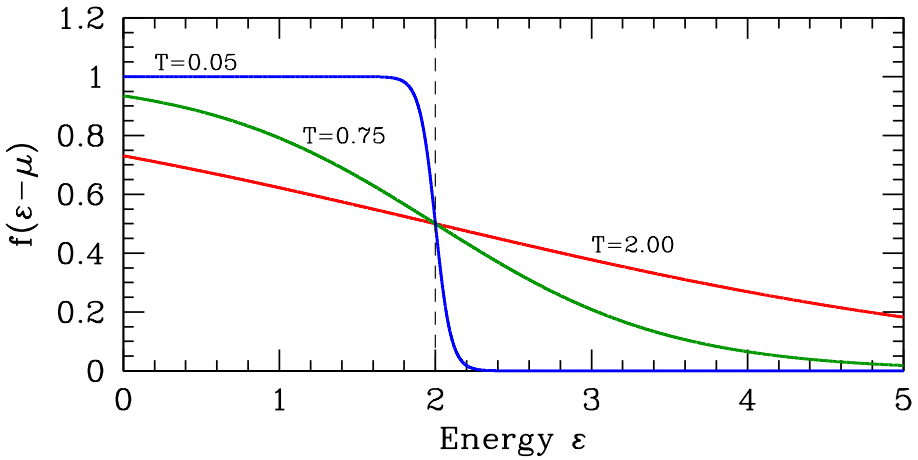

Figure 4.7: The Fermi distribution, $f(x) = \big[\exp(x/k_{\rm B}T) + 1\big]^{-1}$. Here we have set $k_{\rm B} = 1$ and taken $\mu = 2$, with $T = \frac{1}{20}$ (blue), $T = \frac{3}{4}$ (green), and $T = 2$ (red). In the $T \to 0$ limit, $f(x)$ approaches a step function $\Theta(-x)$.

This is to be inverted to obtain $\mu(n, T = 0) = \varepsilon_{\rm F}(n)$.

4.2.3 Metals and insulators at $T = 0$

At $T = 0$, the ground state is formed by filling up each single particle state $|\, n\boldsymbol{k}\sigma \,\rangle$ until the source of electrons (*i.e.* the atoms) is exhausted. Suppose there are $N_{\rm e}$ electrons in total. If there is a finite gap Δ between the $N_{\rm e}^{\rm th}$ and $(N_{\rm e}+1)^{\rm th}$ energy states, the material is an *insulator*. If the gap is zero, the material is a *metal* or possibly a *semimetal*. For a metal, $g(\varepsilon_{\rm F}) > 0$, whereas for a semimetal, $g(\varepsilon_{\rm F}) = 0$ but $g(\varepsilon) \sim |\varepsilon - \varepsilon_{\rm F}|^\alpha$ as $\varepsilon \to \varepsilon_{\rm F}$, where $\alpha > 0$.

Under periodic boundary conditions, there are N quantized wavevectors $\boldsymbol{k}$ in each Brillouin zone, where N is the number of unit cells in the crystal. Since, for a given band index n and wavevector $\boldsymbol{k}$ we can accommodate a maximum of two electrons, one with spin $\uparrow$ and the second with spin $\downarrow$, each band can accommodate a total of $2N$ electrons. Thus, if the number of electrons per cell $N_{\rm e}/N$ is not a precise multiple of two, then *necessarily* at least one of the bands will be partially filled, which means the material is a metal. Typically we only speak of valence and conduction electrons, since the core bands are all fully occupied and the high energy bands are all completely empty. Then we can define the *electron filling factor* $\nu = \widetilde{N}_{\rm e}/N$, where $\widetilde{N}_{\rm e}$ is the total number of valence plus conduction electrons. As we have just noted, if $\nu \neq 2k$ for some $k \in \mathbb{Z}$, the material is a metal.

Is the converse also the case, *i.e.* if $\nu = 2k$ is the material always an insulator? It ain't necessarily so! As the middle panel of Fig. 4.6 shows, it is at least in principle possible to have an arrangement of several partially filled bands such that the total number of electrons per site is an even integer. This is certainly a nongeneric state of affairs, but it is not completely ruled out.

4.3 Tight Binding Model

4.3.1 Bands from atomic orbitals

A crystal is a regular assembly of atoms, which are bound in the crystalline state due to the physics of electrostatics and quantum mechanics. Consider for the sake of simplicity a homonuclear Bravais lattice, *i.e.* a crystalline lattice in which there is the same type of atom at every lattice site, and in which all lattice sites are equivalent under translation. As the lattice constant a tends to infinity, the electronic energy spectrum of the crystal is the same as that of each atom, with an extensive degeneracy of N, the number of unit cells in the lattice. For finite a, the atomic orbitals on different lattice sites will overlap. Initially we will assume a Bravais lattice, but further below we shall generalize this to include the possibility of a basis.

Let $|\, n\boldsymbol{R}\,)$ denote an atomic orbital at Bravais lattice site $\boldsymbol{R}$, where $n \in \{1\mathrm{s}, 2\mathrm{s}, 2\mathrm{p}, \ldots\}$. The atomic wavefunctions[8],

$$\varphi_{n\boldsymbol{R}}(\boldsymbol{r}) = (\, \boldsymbol{r} \,|\, n\boldsymbol{R}\,) = \varphi_n(\boldsymbol{r} - \boldsymbol{R}) \quad , \tag{4.52}$$

Atomic orbitals on the same site form an orthonormal basis: $(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}\,) = \delta_{nn'}$. However, orbitals on different lattice sites are not orthogonal, and satisfy

$$(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}'\,) = \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R})\, \varphi_{n'}(\boldsymbol{r} - \boldsymbol{R}') \equiv S_{nn'}(\boldsymbol{R} - \boldsymbol{R}') \quad , \tag{4.53}$$

where $S_{nn'}(\boldsymbol{R}-\boldsymbol{R}')$ is the *overlap matrix*. Note that $S_{nn'}(0) = \delta_{nn'}$. If we expand the wavefunction $|\,\psi\,\rangle = \sum_{n,\boldsymbol{R}} C_{n\boldsymbol{R}}\, |\, n\boldsymbol{R}\,)$ in atomic orbitals, the Schrödinger equation takes the form

$$\sum_{n',\boldsymbol{R}'} \Big\{ \overbrace{(\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}'\,)}^{H_{nn'}(\boldsymbol{R}-\boldsymbol{R}')} - E\, \overbrace{(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}'\,)}^{S_{nn'}(\boldsymbol{R}-\boldsymbol{R}')} \Big\}\, C_{n'\boldsymbol{R}'} = 0 \quad , \tag{4.54}$$

where

$$\begin{aligned} H_{nn'}(\boldsymbol{R} - \boldsymbol{R}') &= (\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}'\,) \\ &= \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R}) \left\{ -\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r}) \right\} \varphi_{n'}(\boldsymbol{r} - \boldsymbol{R}') \quad , \end{aligned} \tag{4.55}$$

where $V(\boldsymbol{r}) = \sum_{\boldsymbol{R}} v(\boldsymbol{r} - \boldsymbol{R})$ is the lattice potential. Note that

$$\begin{aligned} H_{nn'}(\boldsymbol{R}) = \tfrac{1}{2}\Big(E_n^{\rm at} + E_{n'}^{\rm at}\Big)\, S_{nn'}(\boldsymbol{R}) + \tfrac{1}{2}\int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R}) \Big\{ v(\boldsymbol{r}) + v(\boldsymbol{r} - \boldsymbol{R}) \Big\}\, \varphi_{n'}(\boldsymbol{r}) \\ + \sum_{\substack{\boldsymbol{R}' \\ (\neq 0, \boldsymbol{R})}} \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R})\, v(\boldsymbol{r} - \boldsymbol{R}')\, \varphi_{n'}(\boldsymbol{r}) \quad . \end{aligned} \tag{4.56}$$

[8] In our notation, $|\, \boldsymbol{r}\,) = |\, \boldsymbol{r}\,\rangle$, so $(\, \boldsymbol{r} \,|\, \boldsymbol{r}'\,) = \delta(\boldsymbol{r} - \boldsymbol{r}')$.

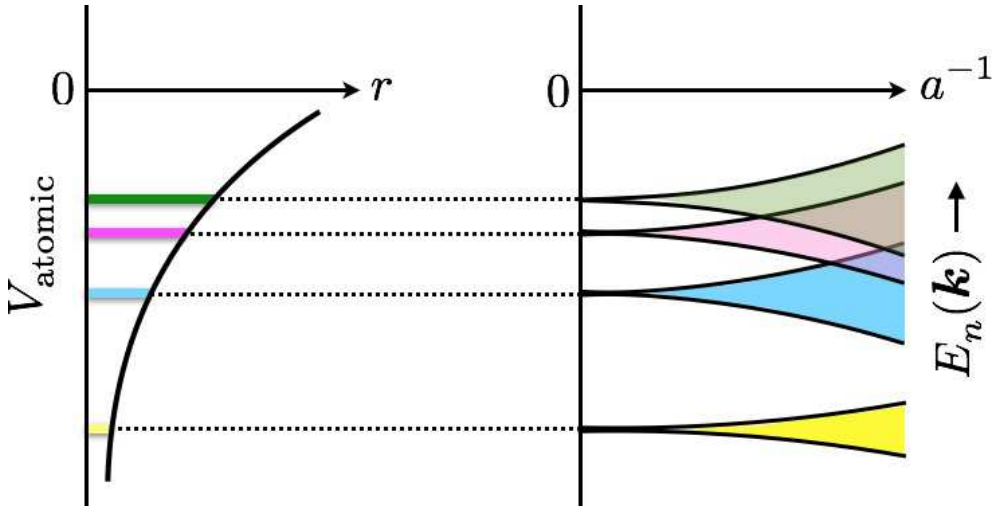

Figure 4.8: Left: Atomic energy levels. Right: Dispersion of crystalline energy bands as a function of interatomic separation.

We can simplify Eqn. 4.54 a bit by utilizing the translational invariance of the Hamiltonian and overlap matrices. We write $C_{n\boldsymbol{R}} = C_n(\boldsymbol{k})\,\exp(i\boldsymbol{k}\cdot\boldsymbol{R})$, as well as

$$\widehat{S}_{nn'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\,S_{nn'}(\boldsymbol{R}) \quad , \quad \widehat{H}_{nn'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\,H_{nn'}(\boldsymbol{R}) \quad . \tag{4.57}$$

The Schrödinger equation then separates for each $\boldsymbol{k}$ value, *viz.*

$$\sum_{n'}\Big\{\widehat{H}_{nn'}(\boldsymbol{k}) - E(\boldsymbol{k})\,\widehat{S}_{nn'}(\boldsymbol{k})\Big\}\,C_{n'}(\boldsymbol{k}) \quad . \tag{4.58}$$

Note that

$$\widehat{H}^*_{n'n}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,H^*_{n'n}(\boldsymbol{R}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,H_{nn'}(-\boldsymbol{R}) = \widehat{H}_{nn'}(\boldsymbol{k}) \quad , \tag{4.59}$$

hence for each $\boldsymbol{k}$, the matrices $\widehat{H}_{nn'}(\boldsymbol{k})$ and $\widehat{S}_{nn'}(\boldsymbol{k})$ are Hermitian.

Suppose we ignore the overlap between different bands. We can then suppress the band index, and write[9]

$$\begin{aligned} S(\boldsymbol{R}) &= \int d^dr\;\varphi^*(\boldsymbol{r}-\boldsymbol{R})\,\varphi(\boldsymbol{r}) \\ H(\boldsymbol{R}) &= E^{\rm at}\,S(\boldsymbol{R}) + \sum_{\boldsymbol{R}'\neq 0}\int d^dr\;\varphi^*(\boldsymbol{r}-\boldsymbol{R})\,v(\boldsymbol{r}-\boldsymbol{R}')\,\varphi(\boldsymbol{r}) \quad ; \end{aligned} \tag{4.60}$$

[9]The student should derive the formulae in Eqn. 4.60. In so doing, is it necessary to presume that the atomic wavefunctions are each of a definite parity?

Note that $S(0) = 1$. The tight binding dispersion is then

$$E(\boldsymbol{k}) = \frac{\widehat{H}(\boldsymbol{k})}{\widehat{S}(\boldsymbol{k})} = E^{\rm at} - \frac{\sum_{\boldsymbol{R}} t(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}}{\sum_{\boldsymbol{R}} S(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}} \quad , \tag{4.61}$$

where

$$t(\boldsymbol{R}) = E^{\rm at} S(\boldsymbol{R}) - H(\boldsymbol{R}) = -\int d^d r\, \varphi^*(\boldsymbol{r}-\boldsymbol{R}) \left\{ \sum_{\boldsymbol{R}'\neq 0} v(\boldsymbol{r}-\boldsymbol{R}') \right\} \varphi(\boldsymbol{r}) \quad . \tag{4.62}$$

Let's examine this result for a d-dimensional cubic lattice. To simplify matters, we assume that $t(\boldsymbol{R})$ and $S(\boldsymbol{R})$ are negligible beyond the nearest neighbor separation $|\boldsymbol{R}| = a$. Then

$$\begin{aligned} E(\boldsymbol{k}) &= E^{\rm at} - \frac{t(0) + 2\,t(a)\sum_{j=1}^d \cos(k_j a)}{1 + 2\,S(a)\sum_{j=1}^d \cos(k_j a)} \\ &\approx E^{\rm at} - t(0) - 2\big[t(a) - t(0)S(a)\big] \sum_{j=1}^d \cos(k_j a) + \ldots \quad , \end{aligned} \tag{4.63}$$

where we expand in the small quantity $S(a)$.

Remarks

Eqn. 4.58 says that the eigenspectrum of the crystalline Hamiltonian at each crystal momentum $\boldsymbol{k}$ is obtained by simultaneously diagonalizing the matrices $\widehat{H}(\boldsymbol{k})$ and $\widehat{S}(\boldsymbol{k})$. The simultaneous diagonalization of two real symmetric matrices is familiar from the classical mechanics of coupled oscillations. The procedure for complex Hermitian matrices follows along the same lines:

(i) To simultaneously diagonalize $\widehat{H}_{nn'}(\boldsymbol{k})$ and $\widehat{S}_{nn'}(\boldsymbol{k})$, we begin by finding a unitary matrix $U_{na}(\boldsymbol{k})$ such that

$$\sum_{n,n'} U^\dagger_{an}(\boldsymbol{k})\, \widehat{S}_{nn'}(\boldsymbol{k})\, U_{n'a'}(\boldsymbol{k}) = s_a(\boldsymbol{k})\, \delta_{aa'} \quad . \tag{4.64}$$

The eigenvalues $s_a(\boldsymbol{k})$ are all real and are presumed to be positive[10].

(ii) Next construct the Hermitian matrix

$$\widehat{L}_{bb'}(\boldsymbol{k}) = \sum_{n,n'} s_b^{-1/2}(\boldsymbol{k})\, U^\dagger_{bn}(\boldsymbol{k})\, \widehat{H}_{nn'}(\boldsymbol{k})\, U_{n'b'}(\boldsymbol{k})\, s_{b'}^{-1/2}(\boldsymbol{k}) \quad . \tag{4.65}$$

[10] In fact there are many zero eigenvalues, as we shall discuss below. Still this won't prove fatal to our development so long as we operate in some truncated Hilbert space.

This may be diagonalized by a unitary matrix $V_{ab}(\boldsymbol{k})$, *viz.*

$$\sum_{b,b'} V^\dagger_{ab}(\boldsymbol{k})\,\widehat{L}_{bb'}(\boldsymbol{k})\,V_{a'b'}(\boldsymbol{k}) = E_a(\boldsymbol{k})\,\delta_{aa'} \quad . \tag{4.66}$$

The eigenvalues are then the set $\{E_a(\boldsymbol{k})\}$.

(iii) Define the matrix

$$\Lambda_{nb}(\boldsymbol{k}) = \sum_a U_{na}(\boldsymbol{k})\,s_a^{-1/2}(\boldsymbol{k})\,V_{ab}(\boldsymbol{k})\,s_b^{1/2}(\boldsymbol{k}) \quad , \tag{4.67}$$

or, in abbreviated notation, $\Lambda = U s^{-1/2}\,V s^{1/2}$, suppressing the $\boldsymbol{k}$ label and defining the square matrix $s = \mathsf{diag}(s_1, s_2, \ldots\}$. Note that $\Lambda^\dagger = s^{1/2}\,V^\dagger s^{-1/2}\,U^\dagger$ but $\Lambda^{-1} = s^{-1/2}\,V^\dagger s^{1/2}\,U^\dagger$ and thus $\Lambda^\dagger \neq \Lambda^{-1}$. Nevertheless, Λ simultaneously diagonalizes $\widehat{H}$ and $\widehat{S}$:

$$\begin{aligned} \Lambda^\dagger_{an}(\boldsymbol{k})\,\widehat{S}_{nn'}(\boldsymbol{k})\,\Lambda_{n'a'}(\boldsymbol{k}) &= s_a(\boldsymbol{k})\,\delta_{aa'} \\ \Lambda^\dagger_{an}(\boldsymbol{k})\,\widehat{H}_{nn'}(\boldsymbol{k})\,\Lambda_{n'a'}(\boldsymbol{k}) &= h_a(\boldsymbol{k})\,\delta_{aa'} \quad , \end{aligned} \tag{4.68}$$

where $h_a(\boldsymbol{k}) = s_a(\boldsymbol{k})\,E_a(\boldsymbol{k})$.

Thus, the band energies are then given by

$$E_a(\boldsymbol{k}) = \frac{h_a(\boldsymbol{k})}{s_a(\boldsymbol{k})} \quad . \tag{4.69}$$

This all seems straightforward enough. However, implicit in this procedure is the assumption that the overlap matrix is nonsingular, which is clearly wrong! We know that the atomic eigenstates at any *single* lattice site must form a complete set, therefore we must be able to write

$$\varphi_n(\boldsymbol{r} - \boldsymbol{R}) = \sum_{n'} A_{nn'}(\boldsymbol{R})\,\varphi_{n'}(\boldsymbol{r}) \quad . \tag{4.70}$$

Therefore the set $|\,n\boldsymbol{R}\,)$ is massively degenerate. Fortunately, this problem is not nearly so severe as it might first appear. Recall that the atomic eigenstates consist of *bound states* of negative energy, and *scattering states* of positive energy. If we restrict our attention to a finite set of atomic bound states, the overlap matrix remains nonsingular.

4.3.2 Wannier functions

Suppose a very nice person gives us a complete set of Bloch functions $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$. We can then form the linear combinations

$$W_n(\boldsymbol{r} - \boldsymbol{R}) \equiv \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\,e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})}\,u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{4.71}$$

were N is the number of unit cells and $\chi_n(\boldsymbol{k})$ is a smooth function of the wavevector $\boldsymbol{k}$ which satisfies $\chi_n(\boldsymbol{k}+\boldsymbol{G}) = \chi_n(\boldsymbol{k})$. The $\boldsymbol{k}$ sum is over all wavevectors lying within the first Brillouin zone. Writing $W_n(\boldsymbol{r}) = \langle\, \boldsymbol{r} \,|\, n\boldsymbol{R} \,\rangle$, we have

$$|\, n\boldsymbol{R} \,\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, |\, n\boldsymbol{k} \,\rangle \quad , \tag{4.72}$$

and the overlap matrix is

$$\langle\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}' \,\rangle = \int d^d r\; W_n^*(\boldsymbol{r}-\boldsymbol{R})\, W_{n'}(\boldsymbol{r}-\boldsymbol{R}') = \delta_{nn'}\, \delta_{\boldsymbol{R}\boldsymbol{R}'} \quad . \tag{4.73}$$

The functions $W_n(\boldsymbol{r}-\boldsymbol{R})$ are called *Wannier functions*. They are linear combinations of Bloch states within a single energy band which are localized about a single Bravais lattice site or unit cell. Since the Wannier states are normalized, we have $\int d^d r\, \big|W_n(\boldsymbol{r}-\boldsymbol{R})\big|^2 = 1$, which means, if the falloff is the same in all symmetry-related directions of the crystal, that the envelope of $W_n(\boldsymbol{r}-\boldsymbol{R})$ must decay faster than $|\boldsymbol{r}-\boldsymbol{R}|^{-d/2}$ in d dimensions. For core ionic orbitals such as the 1s states, the atomic wavefunctions themselves are good approximations to Wannier states. Note that our freedom to choose the phase functions $\chi_n(\boldsymbol{k})$ results in many different possible definitions of the Wannier states. One desideratum we may choose to impose is to constrain the phase functions so as to minimize the expectation of $(\boldsymbol{r}-\boldsymbol{R})^2$ in each band.

Closed form expressions for Wannier functions are hard to come by, but we can obtain results for the case where the cell functions are constant, *i.e.* $u_{n\boldsymbol{k}}(\boldsymbol{r}) = (Nv_0)^{-1/d}$. Consider the cubic lattice case in $d=3$ dimensions, where $v_0 = a^3$. We then have

$$\begin{aligned} W(\boldsymbol{r}-\boldsymbol{R}) &= v_0^{1/2} \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})} \\ &= \left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_x\, e^{ik_x(x-X)}\right]\left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_y\, e^{ik_y(y-Y)}\right]\left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_z\, e^{ik_z(z-Z)}\right] \\ &= \left[\frac{\sqrt{a}\,\sin\big[\frac{\pi}{a}(x-X)\big]}{\pi(x-X)}\right]\left[\frac{\sqrt{a}\,\sin\big[\frac{\pi}{a}(y-Y)\big]}{\pi(y-Y)}\right]\left[\frac{\sqrt{a}\,\sin\big[\frac{\pi}{a}(z-Z)\big]}{\pi(z-Z)}\right] \quad , \end{aligned} \tag{4.74}$$

which falls off as $|\Delta x\,\Delta y\,\Delta z|^{-1}$ along a general direction in space[11].

The Wannier states are not eigenstates of the crystal Hamiltonian. Indeed, we have

$$\begin{aligned} \langle\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}' \,\rangle &= \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'} e^{-i\chi_n(\boldsymbol{k})}\, e^{i\chi_{n'}(\boldsymbol{k}')}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{-i\boldsymbol{k}'\cdot\boldsymbol{R}'} \langle\, n\boldsymbol{k} \,|\, H \,|\, n'\boldsymbol{k}' \,\rangle \\ &= \delta_{nn'}\, v_0 \int \frac{d^dk}{(2\pi)^d}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')}\, E_n(\boldsymbol{k}) \quad , \end{aligned} \tag{4.75}$$

[11] Note that $W(x,0,0)$ falls off only as $1/|x|$. Still, due to the more rapid decay along a general real space direction, $W(\boldsymbol{r})$ is square integrable.

which is diagonal in the band indices, but not in the unit cell labels.

4.3.3 Tight binding redux

Suppose we have an orthonormal set of orbitals $|\, a\boldsymbol{R}\,\rangle$, where a labels the orbital and $\boldsymbol{R}$ denotes a Bravais lattice site. The label a may refer to different orbitals associated with the atom at $\boldsymbol{R}$, or it may label orbitals on other atoms in the unit cell defined by $\boldsymbol{R}$. We presume that $a \in \{1, \ldots, N_{\rm orb}\}$ with $N_{\rm orb}$ finite[12]. The most general tight binding Hamiltonian we can write is

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, |\, a\boldsymbol{R}\,\rangle\langle\, a'\boldsymbol{R}'\,| \quad , \tag{4.76}$$

where $H_{aa'}(\boldsymbol{R} - \boldsymbol{R}') = H^*_{a'a}(\boldsymbol{R}' - \boldsymbol{R}) = \langle\, a\boldsymbol{R}\,|\, H\,|\, a'\boldsymbol{R}'\,\rangle$ is the Hamiltonian matrix, whose rows and columns are indexed by a composite index combining both the unit cell label $\boldsymbol{R}$ and the orbital label a. When $\boldsymbol{R} = \boldsymbol{R}'$ and $a = a'$, the term $H_{aa}(0) = \varepsilon_a$ is the energy of a single electron in an isolated a orbital. For all other cases, $H_{aa'}(\boldsymbol{R} - \boldsymbol{R}') = -t_{aa'}(\boldsymbol{R} - \boldsymbol{R}')$ is the hopping integral between the a orbital in unit cell $\boldsymbol{R}$ and the a' orbital in unit cell $\boldsymbol{R}'$. Let's write an eigenstate $|\,\psi\,\rangle$ as

$$|\,\psi\,\rangle = \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{4.77}$$

Applying the Hamiltonian to $|\,\psi\,\rangle$, we obtain the coupled equations

$$\sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, \psi_{a'\boldsymbol{R}'}\, |\, a\boldsymbol{R}\,\rangle = E \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{4.78}$$

Since the $|\, a\boldsymbol{R}\,\rangle$ basis is complete, we must have that the coefficients of $|\, a\boldsymbol{R}\,\rangle$ on each side agree. Therefore,

$$\sum_{\boldsymbol{R}'} \sum_{a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, \psi_{a'\boldsymbol{R}'} = E\, \psi_{a\boldsymbol{R}} \quad . \tag{4.79}$$

We now use Bloch's theorem, which says that each eigenstate may be labeled by a wavevector $\boldsymbol{k}$, with $\psi_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}}\, u_a(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$. The $N^{-1/2}$ prefactor is a normalization term. Multiplying each side by by $e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$, we have

$$\sum_{a'} \left(\sum_{\boldsymbol{R}'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \right) u_{a'}(\boldsymbol{k}') = E(\boldsymbol{k})\, u_a(\boldsymbol{k}) \quad , \tag{4.80}$$

which may be written as

$$\sum_{a'} \widehat{H}_{aa'}(\boldsymbol{k})\, u_{a'}(\boldsymbol{k}) = E(\boldsymbol{k})\, u_a(\boldsymbol{k}) \quad , \tag{4.81}$$

[12] If our unit cell contained one s-orbital for each of the r basis sites, then $N_{\rm orb} = r$. But we may have multiple orbitals for each atom/ion within the unit cell.

where

$$\widehat{H}_{aa'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} H_{aa'}(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.82}$$

Thus, for each crystal wavevector $\boldsymbol{k}$, the $u_{a\boldsymbol{k}}$ are the eigenfunctions of the $N_{\rm orb} \times N_{\rm orb}$ Hermitian matrix $\widehat{H}_{aa'}(\boldsymbol{k})$. The energy eigenvalues at wavevector $\boldsymbol{k}$ are given by $\mathsf{spec}\big\{\widehat{H}(\boldsymbol{k})\big\}$, *i.e.* by the set of eigenvalues of the matrix $\widehat{H}(\boldsymbol{k})$. There are $N_{\rm orb}$ such solutions (some of which may be degenerate), which we distinguish with a band index n, and we denote $u_{n\boldsymbol{a}}(\boldsymbol{k})$ and $E_n(\boldsymbol{k})$ as the corresponding eigenvectors and eigenvalues. We sometimes will use the definition $\hat{t}_{aa'}(\boldsymbol{k}) \equiv -\widehat{H}_{aa'}(\boldsymbol{k})$ for the matrix of hopping integrals.

4.3.4 Interlude on Fourier transforms

It is convenient to use second quantized notation and write the Hamiltonian as

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, c^\dagger_{a\boldsymbol{R}}\, c_{a'\boldsymbol{R}'} \quad , \tag{4.83}$$

where $c^\dagger_{a\boldsymbol{R}}$ creates an electron in orbital a at unit cell $\boldsymbol{R}$. The *second quantized* fermion creation and annihilation operators satisfy the anticommutation relations

$$\big\{c_{a\boldsymbol{R}}\,,\, c^\dagger_{a'\boldsymbol{R}'}\big\} = \delta_{\boldsymbol{R}\boldsymbol{R}'}\,\delta_{aa'} \quad . \tag{4.84}$$

To quantize the wavevectors, we place our system on a d-dimensional torus with N_j unit cells along principal Bravais lattice vector $\boldsymbol{a}_j$ for all $j \in \{1, \ldots, d\}$. The total number of unit cells is then $N = N_1 N_2 \cdots N_d$. Consider now the Fourier transforms,

$$c_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} c_{a\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad c^\dagger_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} c^\dagger_{a\boldsymbol{k}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.85}$$

and their inverses

$$c_{a\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} c_{a\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad c^\dagger_{a\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} c^\dagger_{a\boldsymbol{R}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.86}$$

One then has $\big\{c_{a\boldsymbol{k}}\,,\, c_{a'\boldsymbol{k}'}\big\} = \delta_{aa'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'}$, which says that the Fourier space operators satisfy the same anticommutation relations as in real space, *i.e.* the individual $\boldsymbol{k}$ modes are orthonormal. This is equivalent to the result $\langle\, a\boldsymbol{k}\,|\,a'\boldsymbol{k}'\,\rangle = \delta_{aa'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'}$, where $|\,a\boldsymbol{k}\,\rangle = N^{-1/2}\sum_{\boldsymbol{R}} |\,a\boldsymbol{R}\,\rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$. The Hamiltonian may now be expressed as

$$H = \sum_{\boldsymbol{k}} \sum_{a,a'} \widehat{H}_{aa'}(\boldsymbol{k})\, c^\dagger_{a\boldsymbol{k}}\, c_{a'\boldsymbol{k}} \quad , \tag{4.87}$$

where $\widehat{H}_{aa'}(\boldsymbol{k})$ was defined in Eqn. 4.82 above.

You must, at the very deepest level of your soul, internalize Eqn. 4.85. Equivalently, using bra and ket vectors,

$$\langle\, a\boldsymbol{R}\,| = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \langle\, a\boldsymbol{k}\,| \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad |\, a\boldsymbol{R}\,\rangle = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} |\, a\boldsymbol{k}\,\rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.88}$$

and

$$\langle\, a\boldsymbol{k}\,| = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} \langle\, a\boldsymbol{R}\,| \, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad |\, a\boldsymbol{k}\,\rangle = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} |\, a\boldsymbol{R}\,\rangle\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.89}$$

To establish the inverse relations, we evaluate

$$\begin{aligned} |\, a\boldsymbol{R}\,\rangle &= \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} |\, a\boldsymbol{k}\,\rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ &= \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \left(\frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}'} |\, a\boldsymbol{R}'\,\rangle\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}'} \right) e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ &= \sum_{\boldsymbol{R}'} \left(\frac{1}{N}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \right) |\, a\boldsymbol{R}'\,\rangle \quad . \end{aligned} \tag{4.90}$$

Similarly, we find

$$|\, a\boldsymbol{k}\,\rangle = \sum_{\boldsymbol{k}'} \left(\frac{1}{N}\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} \right) |\, a\boldsymbol{k}'\,\rangle \quad . \tag{4.91}$$

In order for the inverse relations to be true, then, the quantities in round brackets in the previous two equations must satisfy

$$\frac{1}{N}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} = \delta_{\boldsymbol{R}\boldsymbol{R}'} \quad , \quad \frac{1}{N}\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} = \delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \tag{4.92}$$

Let's see how this works in the $d = 1$ case. Let the lattice constant be a and place our system on a ring of N sites (*i.e.* a one-dimensional torus). The k values are then quantized according to $k_j = 2\pi j/a$, where $j \in \{0, \ldots, N-1\}$. The first equation in Eqn. 4.92 is then

$$\frac{1}{N}\sum_{j=0}^{N-1} e^{-2\pi i j(n-n')/N} = \delta_{nn'} \quad , \tag{4.93}$$

where we have replaced $\boldsymbol{R}$ by na and $\boldsymbol{R}'$ by $n'a$, with $n, n' \in \{1, \ldots, N\}$. Clearly the above equality holds true when $n = n'$. For $n \neq n'$, let $z = e^{-2\pi i(n-n')/N}$. The sum is $1 + z + \ldots + z^{N-1} = (1 - z^N)/(1 - z)$. But $z^N = 1$ and $z \neq 1$, so the identity is again verified.

If we do not restrict n and n' to be among $\{1, \ldots, N\}$ and instead let their values range freely over the integers, then the formula is still correct, provided we write the RHS as $\delta_{n,n' \bmod N}$. Similarly, we must understand $\delta_{\boldsymbol{R}\boldsymbol{R}'}$ in Eqn. 4.92 to be unity whenever $\boldsymbol{R}' = \boldsymbol{R} + l_1\, N_1\, \boldsymbol{a}_1 + \ldots + l_d\, N_d\, \boldsymbol{a}_d$, where each $l_j \in \mathbb{Z}$, and zero otherwise. Similarly, $\delta_{\boldsymbol{k}\boldsymbol{k}'}$ is unity whenever $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{G}$, where $\boldsymbol{G} \in \widehat{\mathcal{L}}$ is any reciprocal lattice vector, and zero otherwise.

4.3.5 Examples of tight binding dispersions

One-dimensional lattice

Consider the case of a one-dimensional lattice. The lattice sites lie at positions $X_n = na$ for $n \in \mathbb{Z}$. The hopping matrix elements are $t(j) = t\,\delta_{j,1} + t\,\delta_{j,-1}$, where j is the separation between sites in units of the lattice constant a. Then $\hat{t}(k) = 2t\cos(ka)$ and the dispersion is $E(k) = -2t\cos(ka)$. Equivalently, and quite explicitly,

$$\begin{aligned} H &= -t\sum_n \Big(|\,n+1\,\rangle\langle\,n\,| + |\,n\,\rangle\langle\,n+1\,| \Big) = -\frac{t}{N}\sum_k \sum_{k'} \sum_n e^{-ik'(n+1)a}\, e^{ikna}\, |\,k\,\rangle\langle\,k'\,| + \text{H.c.} \\ &= -t\sum_k \sum_{k'} \left(\frac{1}{N}\sum_n e^{i(k-k')na} \right) e^{-ik'a}\, |\,k\,\rangle\langle\,k'\,| + \text{H.c.} = -2t\sum_k \cos(ka)\, |\,k\,\rangle\langle\,k\,| \quad , \end{aligned} \tag{4.94}$$

since the term in round brackets is $\delta_{kk'}$, as per Eqn. 4.92.

s-orbitals on cubic lattices

On a Bravais lattice with one species of orbital, there is only one band. Consider the case of s orbitals on a d-dimensional cubic lattice. The hopping matrix elements are

$$t(\boldsymbol{R}) = t\sum_{j=1}^{d} \Big(\delta_{\boldsymbol{R},\boldsymbol{a}_j} + \delta_{\boldsymbol{R},-\boldsymbol{a}_j} \Big) \quad , \tag{4.95}$$

where $\boldsymbol{a}_j = a\,\hat{\mathbf{e}}_j$ is the j^{th} elementary direct lattice vector. Taking the discrete Fourier transform (DFT) as specified in Eqn. 4.82,

$$\hat{t}(\boldsymbol{k}) = 2t\sum_{j=1}^{d} \cos(k_j a) \quad . \tag{4.96}$$

The dispersion is then $E(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k})$. The model exhibits a *particle-hole symmetry*,

$$\widetilde{c}_{\boldsymbol{k}} \equiv c^\dagger_{\boldsymbol{k}+\boldsymbol{Q}} \quad , \tag{4.97}$$

where $\boldsymbol{Q} = \frac{\pi}{a}\big(\hat{\mathbf{e}}_1 + \ldots + \hat{\mathbf{e}}_d\big)$. Note $\hat{t}(\boldsymbol{k}+\boldsymbol{Q}) = -\hat{t}(\boldsymbol{k})$.

s-orbitals on the triangular lattice

The triangular lattice is depicted as the lattice of black dots in the left panel of Fig. 4.9. The elementary direct lattice vectors are

$$\boldsymbol{a}_1 = a\left(\tfrac{1}{2}\hat{\mathbf{e}}_1 - \tfrac{\sqrt{3}}{2}\hat{\mathbf{e}}_2\right) \quad , \qquad \boldsymbol{a}_2 = a\left(\tfrac{1}{2}\hat{\mathbf{e}}_1 + \tfrac{\sqrt{3}}{2}\hat{\mathbf{e}}_2\right) \quad , \tag{4.98}$$

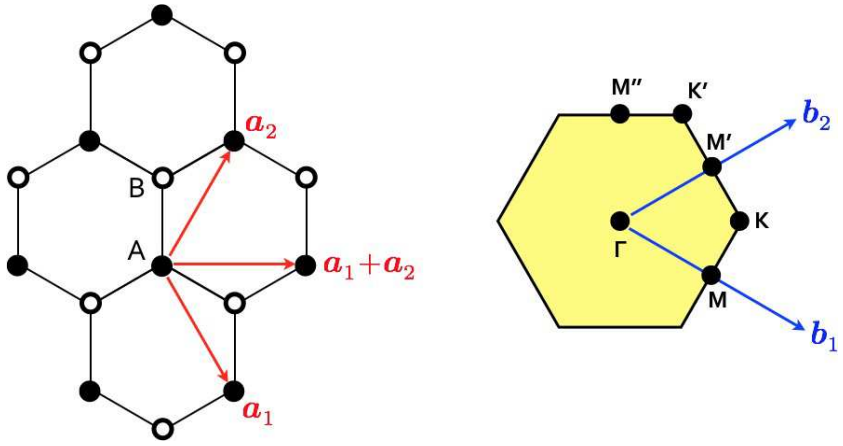

Figure 4.9: Left: The honeycomb lattice is a triangular lattice (black sites) with a two element basis (add white sites). $\boldsymbol{a}_{1,2}$ are elementary direct lattice vectors. Right: First Brillouin zone for the triangular lattice. $\boldsymbol{b}_{1,2}$ are elementary reciprocal lattice vectors. Points of high symmetry Γ, K, K′, M, M′, and M″ are shown.

and the elementary reciprocal lattice vectors are

$$\boldsymbol{b}_1 = \frac{4\pi}{\sqrt{3}\,a}\left(\tfrac{\sqrt{3}}{2}\hat{\mathrm{e}}_1 - \tfrac{1}{2}\hat{\mathrm{e}}_2\right) \quad , \quad \boldsymbol{b}_2 = \frac{4\pi}{\sqrt{3}\,a}\left(\tfrac{\sqrt{3}}{2}\hat{\mathrm{e}}_1 + \tfrac{1}{2}\hat{\mathrm{e}}_2\right) \quad . \tag{4.99}$$

Note that $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\,\delta_{ij}$. The hopping matrix elements are

$$t(\boldsymbol{R}) = t\,\delta_{\boldsymbol{R},\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_2} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_2} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_3} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_3} \quad , \tag{4.100}$$

where $\boldsymbol{a}_3 \equiv \boldsymbol{a}_1 + \boldsymbol{a}_2$. Thus,

$$\begin{aligned}\hat{t}(\boldsymbol{k}) &= 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1) + 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_2) + 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_3) \\ &= 2t\cos(\theta_1) + 2t\cos(\theta_2) + 2t\cos(\theta_1+\theta_2) \quad .\end{aligned} \tag{4.101}$$

Here we have written

$$\boldsymbol{k} = \frac{\theta_1}{2\pi}\,\boldsymbol{b}_1 + \frac{\theta_2}{2\pi}\,\boldsymbol{b}_2 + \frac{\theta_3}{2\pi}\,\boldsymbol{b}_3 \quad , \tag{4.102}$$

and therefore for a general $\boldsymbol{R} = l_1\,\boldsymbol{a}_1 + l_2\,\boldsymbol{a}_2 + l_3\,\boldsymbol{a}_3$, we have

$$\boldsymbol{k}\cdot(l_1\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + l_3\boldsymbol{a}_3) = l_1\theta_1 + l_2\theta_2 + l_3\theta_3 \quad . \tag{4.103}$$

Again there is only one band, because the triangular lattice is a Bravais lattice. The dispersion relation is $E(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k})$. Unlike the case of the d-dimensional cubic lattice, the triangular lattice energy band does not exhibit particle-hole symmetry. The extrema are at $E_{\rm min} = E(\Gamma) = -6t$, and $E_{\rm max} = E(\mathrm{K}) = +3t$, where $\Gamma = 0$ is the zone center and $\mathrm{K} = \frac{1}{3}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$ is the zone corner, corresponding to $\theta_1 = \theta_2 = \frac{2\pi}{3}$.

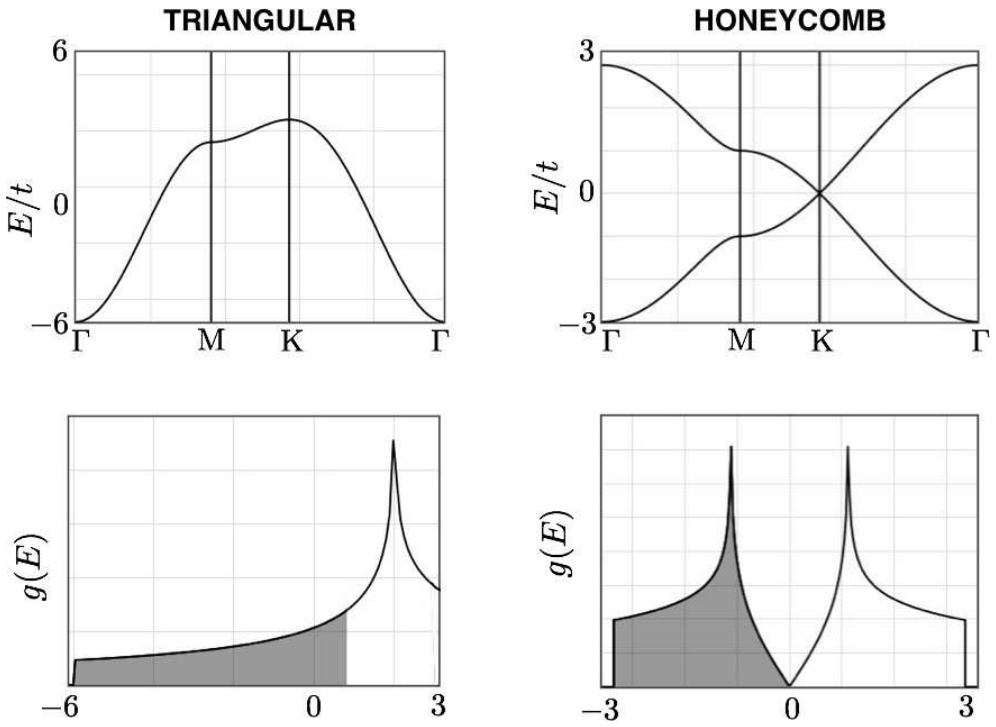

Figure 4.10: Energy bands and density of states for nearest neighbor s-orbital tight binding model on triangular (left) and honeycomb (right) lattices.

Graphene: π-orbitals on the honeycomb lattice

Graphene is a two-dimensional form of carbon arrayed in a honeycomb lattice. The electronic structure of carbon is $1\mathrm{s}^2\,2\mathrm{s}^2\,2\mathrm{p}^2$. The $1s$ electrons are tightly bound and have small overlaps from site to site, hence little dispersion. The $2s$ and $2p_{x,y}$ orbitals engage in sp^2 hybridization. For each carbon atom, three electrons in each atom's sp^2 orbitals are distributed along bonds connecting to its neighbors[13]. Thus each bond gets two electrons (of opposite spin), one from each of its carbon atoms. This is what chemists call a σ-bond. The remaining p_z orbital (the π orbital, to our chemist friends) is then free to hop from site to site. For our purposes it is equivalent to an s-orbital, so long as we don't ask about its properties under reflection in the x-y plane. The underlying Bravais lattice is triangular, with a two element basis (labelled A and B in Fig. 4.9). According to the left panel of Fig. 4.9, the A sublattice site in unit cell $\boldsymbol{R}$ is connected to the B sublattice sites in unit cells $\boldsymbol{R}$, $\boldsymbol{R}+\boldsymbol{a}_1$, and $\boldsymbol{R}-\boldsymbol{a}_2$. Thus, the hopping matrix element between the A sublattice sites in unit sell 0 and the B sublattice site in unit cell $\boldsymbol{R}$ is given by

$$t_{\mathrm{AB}}(\boldsymbol{R}) = t\,\delta_{\boldsymbol{R},0} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_2} \quad , \tag{4.104}$$

[13] In diamond, the carbon atoms are fourfold coordinated, and th orbitals are sp^3 hybridized.

and therefore $\hat{t}_{\rm AB}(\boldsymbol{k}) = t\big(1 + e^{-i\theta_1} + e^{i\theta_2}\big)$. We also have $\hat{t}_{\rm BA}(\boldsymbol{k}) = \hat{t}^*_{\rm AB}(\boldsymbol{k})$ and $\hat{t}_{\rm AA}(\boldsymbol{k}) = \hat{t}_{\rm BB}(\boldsymbol{k}) = 0$. Thus, the Hamiltonian matrix is

$$H(\boldsymbol{k}) = -\begin{pmatrix} 0 & \hat{t}_{\rm AB}(\boldsymbol{k}) \\ \hat{t}^*_{\rm AB}(\boldsymbol{k}) & 0 \end{pmatrix} = -t\begin{pmatrix} 0 & 1 + e^{-i\theta_1} + e^{i\theta_2} \\ 1 + e^{i\theta_1} + e^{-i\theta_2} & 0 \end{pmatrix} \quad , \tag{4.105}$$

and the energy eigenvalues are

$$E_\pm(\boldsymbol{k}) = \pm\big|\hat{t}_{\rm AB}(\boldsymbol{k})\big| = \pm t\sqrt{3 + 2\cos\theta_1 + 2\cos\theta_2 + 2\cos(\theta_1 + \theta_2)} \quad . \tag{4.106}$$

These bands are depicted in the right panels of Fig. 4.10. Note the band touching at K (and K′), which are known as *Dirac points*. In the vicinity of either Dirac point, writing $\boldsymbol{k} = \boldsymbol{K} + \boldsymbol{q}$ or $\boldsymbol{k} = \boldsymbol{K}' + \boldsymbol{q}$, one has $E_\pm(\boldsymbol{k}) = \pm\hbar v_{\rm F}|\boldsymbol{q}|$, where $v_{\rm F} = \frac{\sqrt{3}}{2}\,ta/\hbar$ is the *Fermi velocity*. At the electroneutrality point (*i.e.* one π electron per site), the Fermi levels lies precisely at $\varepsilon_{\rm F} = 0$. The density of states vanishes continuously as one approaches either Dirac point[14].

p-orbitals on the cubic lattice

Finally, consider the case where each site hosts a trio $(p_x\,,\,p_y\,,\,p_z)$ of p-orbitals. Let the separation between two sites be $\boldsymbol{R}$. Then the 3×3 hopping matrix between these sites depends on two tensors, $\delta_{\mu\nu}$ and $\widehat{\boldsymbol{R}}_\mu\widehat{\boldsymbol{R}}_\nu$. When $\hat{\eta}$ lies along one of the principal cubic axes, the situation is as depicted in Fig. 4.11. The hopping matrix is

$$t_{\mu\nu}(\boldsymbol{R}) = t_{\rm w}(R)\,\delta_{\mu\nu} - \big(t_{\rm w}(R) + t_{\rm s}(R)\big)\,\hat{R}_\mu\,\hat{R}_\nu \quad , \tag{4.107}$$

where the weak and strong hoppings $t_{\rm w,s}$ are depicted in Fig. 4.11. We can now write

$$t_{xx}(\boldsymbol{R}) = -t_{\rm s}\left(\delta_{\boldsymbol{R},\boldsymbol{a}_1} + \delta_{\boldsymbol{R},-\boldsymbol{a}_1}\right) + t_{\rm w}\left(\delta_{\boldsymbol{R},\boldsymbol{a}_2} + \delta_{\boldsymbol{R},-\boldsymbol{a}_2} + \delta_{\boldsymbol{R},\boldsymbol{a}_3} + \delta_{\boldsymbol{R},-\boldsymbol{a}_3}\right) \quad , \tag{4.108}$$

and therefore

$$\hat{t}_{\mu\nu}(\boldsymbol{k}) = 2\begin{pmatrix} -t_{\rm s}\cos\theta_1 + t_{\rm w}(\cos\theta_2 + \cos\theta_3) & 0 & 0 \\ 0 & -t_{\rm s}\cos\theta_2 + t_{\rm w}(\cos\theta_1 + \cos\theta_3) & 0 \\ 0 & 0 & -t_{\rm s}\cos\theta_3 + t_{\rm w}(\cos\theta_1 + \cos\theta_2) \end{pmatrix} \tag{4.109}$$

which is diagonal. The three p-band dispersions are given by the diagonal entries.

4.3.6 Bloch's theorem, again

In Eqn. 4.76,

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\,|\,a\boldsymbol{R}\,\rangle\langle\,a'\boldsymbol{R}'\,| \quad ,$$

[14] *I.e.*, the DOS in either the K or K′ valley.

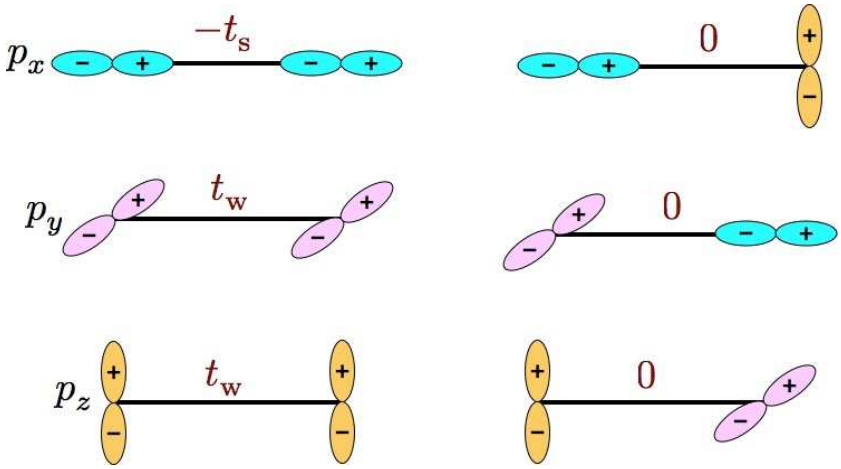

Figure 4.11: Matrix elements for neighboring tight binding orbitals of p symmetry.

$\boldsymbol{R}$ and $\boldsymbol{R}'$ labeled Bravais lattice sites, while a and a' labeled orbitals. We stress that these orbitals don't necessarily have to be located on the same ion. We should think of $\boldsymbol{R}$ and $\boldsymbol{R}'$ labeling *unit cells*, each of which is indeed associated with a Bravais lattice site. For example, in the case of graphene, $|\, a\boldsymbol{R}\,\rangle$ represents an orbital on the a sublattice in unit cell $\boldsymbol{R}$. The eigenvalue equation may be written

$$\widehat{H}_{aa'}(\boldsymbol{k})\, u_{na'}(\boldsymbol{k}) = E_n(\boldsymbol{k})\, u_{na}(\boldsymbol{k}) \quad , \tag{4.110}$$

where n is the band index. The function $u_{na}(\boldsymbol{k})$ is the *internal wavefunction* within a given cell, and corresponds to the cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ in the continuum, with $a \leftrightarrow (\boldsymbol{r} - \boldsymbol{R})$ labeling a position within each unit cell. The full Bloch state may then be written

$$|\, \psi_{n\boldsymbol{k}}\,\rangle = |\, \boldsymbol{k}\,\rangle \otimes |\, u_{n\boldsymbol{k}}\,\rangle \quad , \tag{4.111}$$

so that

$$\begin{aligned} \psi_{n\boldsymbol{k}}(\boldsymbol{R}, a) &= \Big(\langle\, \boldsymbol{R}\,| \otimes \langle\, a\,|\Big)\Big(|\, \boldsymbol{k}\,\rangle \otimes |\, u_{n\boldsymbol{k}}\,\rangle\Big) \\ &= \langle\, \boldsymbol{R}\,|\, \boldsymbol{k}\,\rangle\, \langle\, a\,|\, u_{n\boldsymbol{k}}\,\rangle = \frac{1}{\sqrt{N}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, u_{na}(\boldsymbol{k}) \quad . \end{aligned} \tag{4.112}$$

Here we have chosen a normalization $\sum_a \big|u_{na}(\boldsymbol{k})\big|^2 = 1$ within each unit cell, which entails the overall normalization $\sum_{\boldsymbol{R},a} \big|\psi_{n\boldsymbol{k}}(\boldsymbol{R}, a)\big|^2 = 1$.

4.3.7 Go flux yourself : how to add magnetic fields

To simplify matters, we consider only s-orbitals on two-dimensional lattices. The general tight-binding Hamiltonian is written

$$H = -\sum_{\boldsymbol{r}<\boldsymbol{r}'} \Big(t_{\boldsymbol{r}\boldsymbol{r}'}\, c^\dagger_{\boldsymbol{r}}\, c_{\boldsymbol{r}'} + t^*_{\boldsymbol{r}\boldsymbol{r}'}\, c^\dagger_{\boldsymbol{r}'}\, c_{\boldsymbol{r}}\Big) \quad , \tag{4.113}$$

where the notation $\boldsymbol{r} < \boldsymbol{r}'$ means that each pair $(\boldsymbol{r}, \boldsymbol{r}')$ summed only once. We may write $t_{\boldsymbol{r}\boldsymbol{r}'} = t^*_{\boldsymbol{r}'\boldsymbol{r}} = |t_{\boldsymbol{r}\boldsymbol{r}'}| \exp(iA_{\boldsymbol{r}\boldsymbol{r}'})$, where $A_{\boldsymbol{r}\boldsymbol{r}'}$ is a gauge field living on the links of the lattice. Let p denote a plaquette on the lattice. Then the dimensionless flux ϕ_p (in units of $\hbar c/e$) through plaquette p is

$$\phi_p = \sum_{\langle \boldsymbol{r}\boldsymbol{r}' \rangle \in \partial p} A_{\boldsymbol{r}\boldsymbol{r}'} \quad , \tag{4.114}$$

where the sum is taken in a counterclockwise fashion along the links on the boundary of p. The tight-binding Hamiltonian exhibits a *gauge invariance* under the combined operations

$$\begin{aligned} c_{\boldsymbol{r}} &\to e^{i\alpha_{\boldsymbol{r}}}\, c_{\boldsymbol{r}} \\ A_{\boldsymbol{r}\boldsymbol{r}'} &\to A_{\boldsymbol{r}\boldsymbol{r}'} + \alpha_{\boldsymbol{r}} - \alpha_{\boldsymbol{r}'} \quad . \end{aligned} \tag{4.115}$$

Consider now the case of the square lattice. It is clear that any configuration of the $A_{\boldsymbol{r}\boldsymbol{r}'}$ which is periodic in the structural unit cell, *i.e.* under translations by elementary direct lattice vectors, must correspond to $\phi_p = 0$ for every plaquette p[15]. This is because the phase $A_{\boldsymbol{r}\boldsymbol{r}'}$ is associated with the *directed link* from $\boldsymbol{r}$ to $\boldsymbol{r}'$, and parallel links on opposite sides of the elementary square plaquette will yield equal and opposite values of $A_{\boldsymbol{r}\boldsymbol{r}'}$ because they are traversed in opposite directions. *In order to describe nonzero flux per plaquette, the configuration of the lattice gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ must break lattice translational symmetry*[16]. Consider the case where $\phi = \pi$ in each plaquette. A configuration for the gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ yielding this flux distribution is shown in the left panel of Fig. 4.12. All links have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$, hence $t_{\boldsymbol{r}\boldsymbol{r}'} = t \exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) = t$, except for the links depicted with slashes, for which $A_{\boldsymbol{r}\boldsymbol{r}'} = \pi$ and $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. The *magnetic unit cell* is now a 2×1 block consisting of one cell from each sublattice (blue and red). We call this a magnetic unit cell to distinguish it from the *structural unit cell* of the underlying square lattice. The *structural* Bravais lattice is square, with elementary direct lattice vectors are $\boldsymbol{a}_1 = a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. But the *magnetic* Bravais lattice is rectangular, with elementary RLVs $\boldsymbol{a}_1 = 2a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. From Bloch's theorem, the phase of the wavefunction varies by $\exp(i\boldsymbol{k} \cdot \boldsymbol{a}_1) \equiv \exp(i\theta_1)$ across the unit cell in the x-direction, and by $\exp(i\boldsymbol{k} \cdot \boldsymbol{a}_2) \equiv \exp(i\theta_2)$ in the y-direction. The Hamiltonian is

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 + e^{i\theta_1} \\ 1 + e^{-i\theta_1} & -2\cos\theta_2 \end{pmatrix} \tag{4.116}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\cos^2(\frac{1}{2}\theta_1) + \cos^2\theta_2}$. The band gap collapses at two points: $(\theta_1, \theta_2) = (\pi, \pm\frac{1}{2}\pi)$. Writing $(\theta_1, \theta_2) = (\pi + \delta_1\,,\, \pm\frac{1}{2}\pi + \delta_2)$, we find

$$E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\sin^2(\tfrac{1}{2}\delta_1) + \sin^2\delta_2} = \pm 2ta\,\sqrt{q_1^2 + q_2^2} + \mathcal{O}(|\boldsymbol{q}|^3) \quad , \tag{4.117}$$

[15] More precisely, if $A_{\boldsymbol{r}\boldsymbol{r}'}$ is periodic in the structural unit cell, then each structural unit cell is congruent to a zero flux state. However, it may be that a structural cell is comprised of more than one elementary plaquette, as is the case with the triangular lattice (each structural cell consists of two triangles), or that there are closed loops which don't correspond to a structural unit cell due to further neighbor hoppings. In such cases, there may be closed loops on the lattice whose flux is not congruent to zero. See §4.3.7 for some examples.

[16] By "nonzero" flux, we mean $\phi \bmod 2\pi \neq 0$.

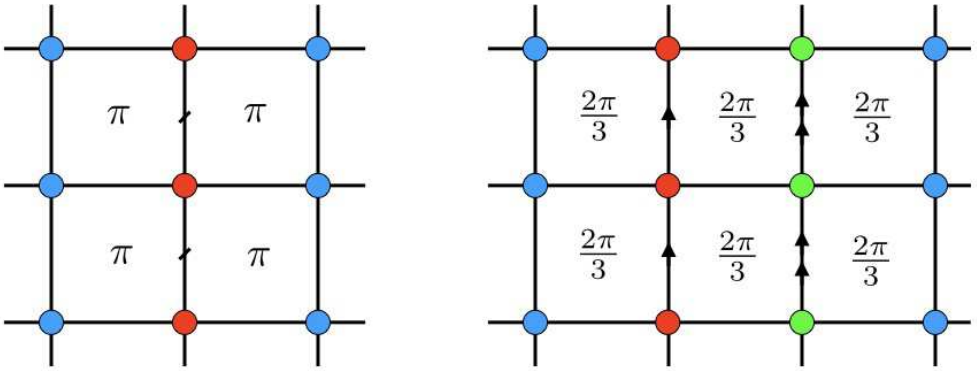

Figure 4.12: Gauges for the square lattice Hofstadter model. Left: $\phi = \pi$ case. $t_{\boldsymbol{r}\boldsymbol{r}'} = t$ on all links except those with slashes, where $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. Right: $\phi = \frac{2}{3}\,\pi$. Each arrow corresponds to a factor of $\exp(2\pi i/3)$.

which is a Dirac cone! Thus, the dispersion for the square lattice π flux model has two Dirac points. Here, $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_{\rm D}$ is the wavevector measured from either Dirac point[17].

The π flux state is time-reversal symmetric, since under time reversal we have $\exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) \to \exp(-iA_{\boldsymbol{r}\boldsymbol{r}'})$, hence $\phi_p \to -\phi_p$. But flux is only defined modulo 2π, hence $\pi \to -\pi \cong \pi$ yields the same flux configuration.

A more interesting state of affairs pertains for the case $\phi = \frac{2}{3}\pi$, for which a valid gauge configuration $A_{\boldsymbol{r}\boldsymbol{r}'}$ is shown in the right panel of Fig. 4.12. Now there are three sites per unit cell: A (blue), B (red), and C (green). The Bloch phase accrued across the magnetic unit cell in the $\pm\hat{\boldsymbol{x}}$ direction is $e^{\pm i\theta_1}$, and in the $\pm\hat{\boldsymbol{y}}$ direction is $e^{\pm i\theta_2}$. Thus

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & e^{i\theta_1} \\ 1 & 2\cos(\theta_2 + \frac{2\pi}{3}) & 1 \\ e^{-i\theta_1} & 1 & 2\cos(\theta_2 + \frac{4\pi}{3}) \end{pmatrix} \quad . \tag{4.118}$$

The general case where the flux per structural unit cell is $\phi = 2\pi p/q$ is known as the Hofstadter model[18]. In this case, the magnetic unit cell is a $q \times 1$ block, and the resulting $q \times q$ Hamiltonian is given by

$$\begin{aligned} H_{jj'}(\boldsymbol{\theta}) = -2t\cos\Big(\theta_2 + \tfrac{2\pi(j-1)p}{q}\Big)\,\delta_{jj'} - t\,\delta_{j',j+1}\big(1-\delta_{j,q}\big) - t\,e^{-i\theta_1}\,\delta_{j,q}\,\delta_{j',1} \\ - t\,\delta_{j',j-1}\big(1-\delta_{j,1}\big) - t\,e^{i\theta_1}\,\delta_{j,1}\,\delta_{j',q} \quad . \end{aligned} \tag{4.119}$$

[17] Note $\theta_1 = 2k_x a = \pi + 2q_x a$ and $\theta_2 = k_y a = \frac{1}{2}\pi + q_y a$, hence $\delta_1 = 2q_x a$ and $\delta_2 = q_y a$.
[18] See D. R. Hofstadter, *Phys. Rev. B* **14**, 2239 (1976).

In other words,

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & 0 & \cdots & 0 & e^{i\theta_1} \\ 1 & 2\cos\left(\theta_2 + \frac{2\pi p}{q}\right) & 1 & & & 0 \\ 0 & 1 & 2\cos\left(\theta_2 + \frac{4\pi p}{q}\right) & 1 & & \vdots \\ \vdots & 0 & 1 & \ddots & & \vdots \\ 0 & & & & 1 & \\ e^{-i\theta_1} & 0 & & \cdots & 1 & 2\cos\left(\theta_2 + \frac{2\pi(q-1)p}{q}\right) \end{pmatrix} . \tag{4.120}$$

There are thus q magnetic subbands. Note that

$$H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = XUH(\theta_1, \theta_2)\, U^\dagger X^\dagger \quad , \tag{4.121}$$

where $X_{ij} = \delta_{i,j+1 \bmod q}$ and $U = \mathsf{diag}(1, e^{-i\theta_1}, \cdots, e^{-i\theta_1})$. Thus,

$$\mathsf{spec}\, H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = \mathsf{spec}\, H(\theta_1, \theta_2) \quad , \tag{4.122}$$

as we saw explicitly in the $q = 2$ case above. A plot of the magnetic subbands in (E, ϕ) space, known as *Hofstadter's butterfly*, is shown in Fig. 4.13.

In the limit where the denominator q of the flux $\phi = 2\pi p/q$ is large (for fixed p), the flux per cell is very small. We then expect to recover the continuum Landau level spectrum $E_n = (n+\frac{1}{2})\hbar\omega_{\rm c}$. To express this in terms of the flux ϕ, note that the $B = 0$ dispersion is

$$E(\boldsymbol{k}) = -2t\cos(k_x a) - 2t\cos(k_y a) = -4t + t\boldsymbol{k}^2 a^2 + \ldots \quad , \tag{4.123}$$

which allows us to identify the effective mass m from the coefficient of the $\boldsymbol{k}^2$ term, with the result $m = \hbar^2/2ta^2$. The magnetic field is the flux per unit area, hence $B = \phi\hbar c/ea^2$. Thus,

$$\hbar\omega_{\rm c} = \frac{\hbar eB}{mc} = \frac{\hbar e}{c} \times \frac{\phi\,\hbar c}{ea^2} \times \frac{2ta^2}{\hbar^2} = 2\phi\, t \quad . \tag{4.124}$$

This describes the corners of the Hofstadter butterfly in Fig. 4.13, where continuum Landau levels radiate outward from the energies $\pm 4t$ according to

$$E_n(\phi) = \pm\Big(4t - (2n+1)\,\phi\, t\Big) \qquad \text{and} \qquad E_n(\phi) = \pm\Big(4t - (2n+1)\,(2\pi - \phi)\, t\Big) \quad , \tag{4.125}$$

for $\phi \ll 1$.

Unit cells with zero net flux

As mentioned in a footnote above, it is not quite true that a lattice gauge field $A_{rr'}$ which is periodic in the underlying Bravais lattice unit cell leads to zero net flux in every plaquette or

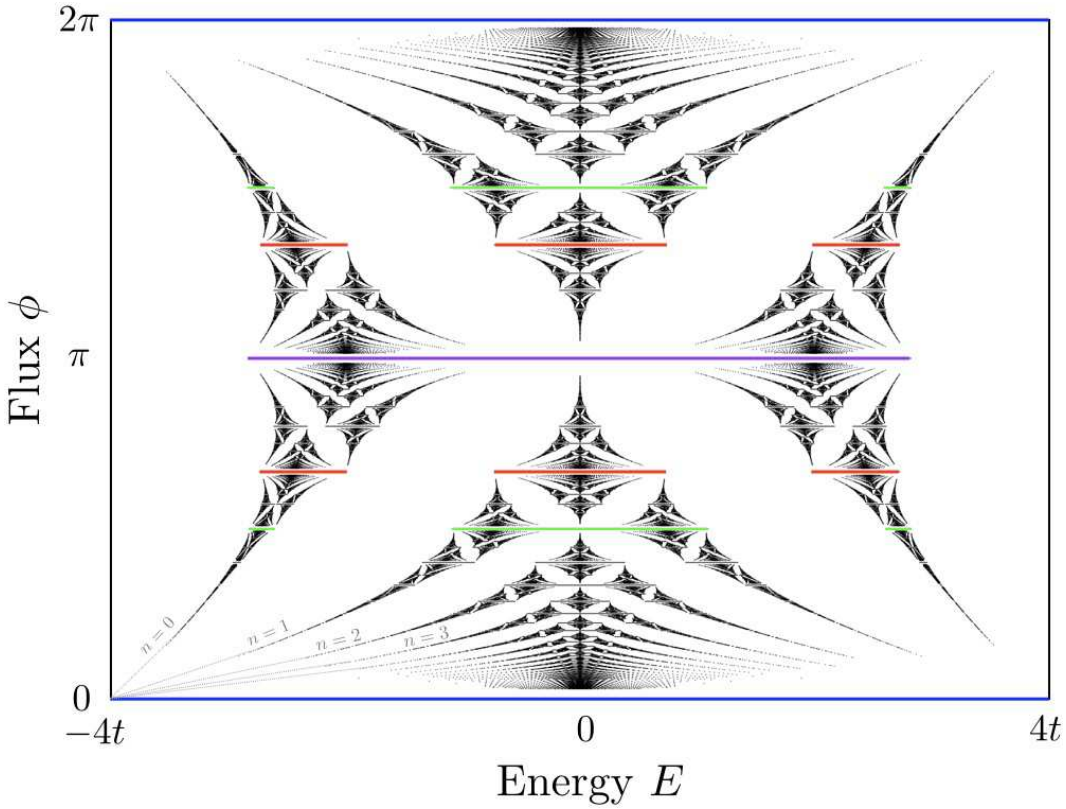

Figure 4.13: Magnetic subbands for the square lattice Hofstadter model for flux per plaquette $\phi \in [0, 2\pi]$. Blue bands at $\phi = 0$ and $\phi = 2\pi$ have the full bandwidth $W = 8t$. At $\phi = \pi$ (purple), there are two subbands with $E_- \in [-2\sqrt{2}t, 0]$ and $E_+ \in [0, 2\sqrt{2}t]$ which touch at $E = 0$. Similarly, at $\phi = \pm\frac{1}{2}\pi$ (green), there are four subbands, with the central two bands touching at $E = 0$. At $\phi = \pm\frac{2}{3}\pi$ (red), there are three subbands. Continuum Landau levels are shown radiating from the lower left corner.

closed loop of links on the lattice. Two counterexamples are shown in Fig. 4.14. The first example is that of the triangular lattice, where each structural unit cell is a rhombus consisting of two elementary triangular plaquettes. Consider now the situation where each horizontal link carries a U(1) phase α, *i.e.* $A_{\boldsymbol{r}\boldsymbol{r}'} = t\, e^{i\alpha}$, while the remaining links all have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$. Computing the U(1) flux by taking the directed sum counterclockwise over each triangle, we see that all the up triangles carry flux $\phi_\triangle = \alpha$, while all the down triangles carry flux $\phi_\triangledown = -\alpha \cong 2\pi - \alpha$. If, as before, we take the elementary direct lattice vectors to be $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ and write $\boldsymbol{k} = \sum_{j=1}^{2} \theta_j\, \boldsymbol{b}_j/2\pi$, with $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\, \delta_{ij}$, then the tight binding Hamiltonian for this triangular lattice model is given by $H = \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}\, a^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}}$, where

$$\begin{aligned} E_{\boldsymbol{k}} &= -2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1 + \boldsymbol{k}\cdot\boldsymbol{a}_2 + \alpha) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_2) \\ &= -2t\cos(\theta_1 + \theta_2 + \alpha) - 2t\cos\theta_1 - 2t\cos\theta_2 \quad . \end{aligned} \tag{4.126}$$

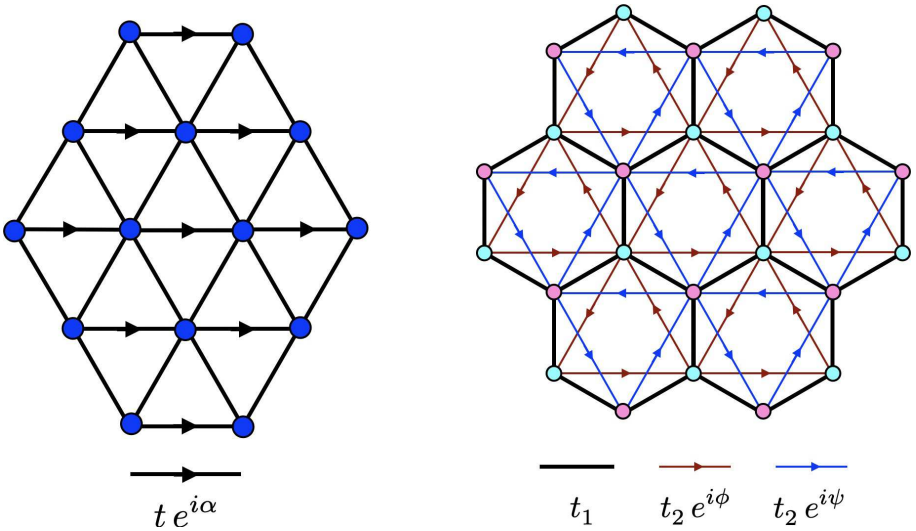

Figure 4.14: Two models with zero net flux per unit cell which still break time reversal symmetry. Left: The unit cell of the triangular lattice consists of two triangles.

A more interesting state of affairs is depicted in the right panel of Fig. 4.14, which is graphene augmented by nearest neighbor same-sublattice hopping terms, which is the celebrated Haldane honeycomb lattice model[19]. Inscribed in each hexagonal cell are one up-triangle of A sites, depicted by blue dots in the figure, and one down-triangle of B sites, depicted as pink dots in the figure. Again we take $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ for the underlying A Bravais lattice, with the basis vectors $\mathbf{0}$ and $\boldsymbol{\delta}_1 = a\hat{\boldsymbol{y}}$. The nearest neighbor hoppings between A and B sites all are taken to have (real) amplitude t_1, while the inscribed A and B same-sublattice hoppings are taken to be $t_2\,e^{i\phi}$ and $t_2\,e^{i\psi}$, respectively, and taken in the counterclockwise direction around the inscribed $\triangle$ and $\triangledown$ paths. An on-site energy term $\pm m$, called the *Semenoff mass*, is added to the hopping terms. Taking the product of the real t_1 amplitudes around any hexagonal unit cell along the thick black lines, we see that the flux per unit cell is congruent to zero. However, *within* each unit cell there is nonzero flux, distributed nonuniformly such that the total flux is zero. The corresponding tight binding Hamiltonian is given by

$$\begin{aligned} H = m\sum_i \big(a_i^\dagger\,a_i - b_i^\dagger\,b_i\big) - t_1\sum_{\langle ij\rangle}\big(a_i^\dagger\,b_j + b_j^\dagger\,a_i\big) \\ - t_2\sum_{\triangle}\big(e^{i\phi}\,a_i^\dagger\,a_j + e^{-i\phi}\,a_j^\dagger\,a_i\big) - t_2\sum_{\triangledown}\big(e^{i\psi}\,b_i^\dagger\,b_j + e^{-i\psi}\,b_j^\dagger\,b_i\big) \quad , \end{aligned} \tag{4.127}$$

where the same-sublattice hopping along the triangles $\triangle$ and $\triangledown$ inscribed within each hexagon

[19] F. D. M. Haldane, *Phys. Rev. Lett.* **61**, 2015 (1988).

are to be traversed in a counterclockwise direction. Fourier transforming to $\boldsymbol{k}$-space, one has

$$\begin{aligned} H = \sum_{\boldsymbol{k}} \Big\{ &- t_1 \left(1 + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) a_{\boldsymbol{k}}^\dagger \, b_{\boldsymbol{k}} - t_1 \left(1 + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) b_{\boldsymbol{k}}^\dagger \, a_{\boldsymbol{k}} \\ &+ \Big[m - 2t_2 \, \mathsf{Re}\left(e^{i\phi} \, e^{i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\phi} \, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\phi} \, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big] \, a_{\boldsymbol{k}}^\dagger \, a_{\boldsymbol{k}} \\ &+ \Big[-m - 2t_2 \, \mathsf{Re}\left(e^{i\psi} \, e^{-i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\psi} \, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\psi} \, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big] \, b_{\boldsymbol{k}}^\dagger \, b_{\boldsymbol{k}} \Big\} \quad , \end{aligned} \tag{4.128}$$

and thus we may write

$$H = \sum_{\boldsymbol{k}} \begin{pmatrix} a_{\boldsymbol{k}}^\dagger & b_{\boldsymbol{k}}^\dagger \end{pmatrix} \begin{pmatrix} H_{\mathsf{AA}}(\boldsymbol{k}) & H_{\mathsf{AB}}(\boldsymbol{k}) \\ H_{\mathsf{BA}}(\boldsymbol{k}) & H_{\mathsf{BB}}(\boldsymbol{k}) \end{pmatrix} \begin{pmatrix} a_{\boldsymbol{k}} \\ b_{\boldsymbol{k}} \end{pmatrix} \quad , \tag{4.129}$$

where

$$\begin{aligned} H_{\mathsf{AA}}(\boldsymbol{k}) &= m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi \\ H_{\mathsf{BB}}(\boldsymbol{k}) &= -m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\psi + 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\psi \\ H_{\mathsf{AB}}(\boldsymbol{k}) &= H_{\mathsf{BA}}^*(\boldsymbol{k}) = -t_1\big(1 + e^{i\theta_1} + e^{-i\theta_2}\big) \quad . \end{aligned} \tag{4.130}$$

In the Haldane model one has $\psi = \phi$, in which case we may write $H(\boldsymbol{k})$ in terms of Pauli matrices, as

$$\begin{aligned} H(\boldsymbol{k}) = &-2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - t_1\big(1 + \cos\theta_1 + \cos\theta_2\big)\,\sigma^x \\ &+ t_1\big(\sin\theta_1 - \sin\theta_2\big)\,\sigma^y + \Big(m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi\Big)\,\sigma^z \quad . \end{aligned} \tag{4.131}$$

What makes the Haldane model so interesting is that its band structure is *topological* over a range of the dimensionless parameters m/t_2 and ϕ.[20] More on this below!

4.3.8 General flux configuration on the square lattice

More generally, consider a magnetic unit cell formed by an $M \times N$ block of structural unit cells, as depicted in Fig. 4.15. Each structural cell p is labeled by the indices (m, n), where the Bravais lattice site its lower left corner is $ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$. To assign the lattice gauge fields, do the following. For $\boldsymbol{r} = ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = ma\hat{\boldsymbol{x}} + (n+1)a\hat{\boldsymbol{y}}$ with $n < N$, let $A_{\boldsymbol{r}\boldsymbol{r}'} = \sum_{i=1}^{m-1} \phi_{i,n}$. For the $n = N$, we include the Bloch phase θ_2, so that $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_2 + \sum_{i=1}^{m-1} \phi_{i,n}$, also noting that $(m, N+1) \cong (m, 1)$. This sets $A_{\boldsymbol{r}\boldsymbol{r}'}$ for all vertical (y-directed) links. The only horizontal links for which $A_{\boldsymbol{r}\boldsymbol{r}'}$ are nonzero are those with $\boldsymbol{r} = Ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = a\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$; note $(M+1, n) \cong (1, n)$. Then $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_1 - \sum_{i=1}^{M} \sum_{j=1}^{n} \phi_{i,j}$. One can check that this prescription yields the desired flux configuration, as well as the two Bloch phases.

[20] Without loss of generality, we may set $t_1 \equiv 1$.

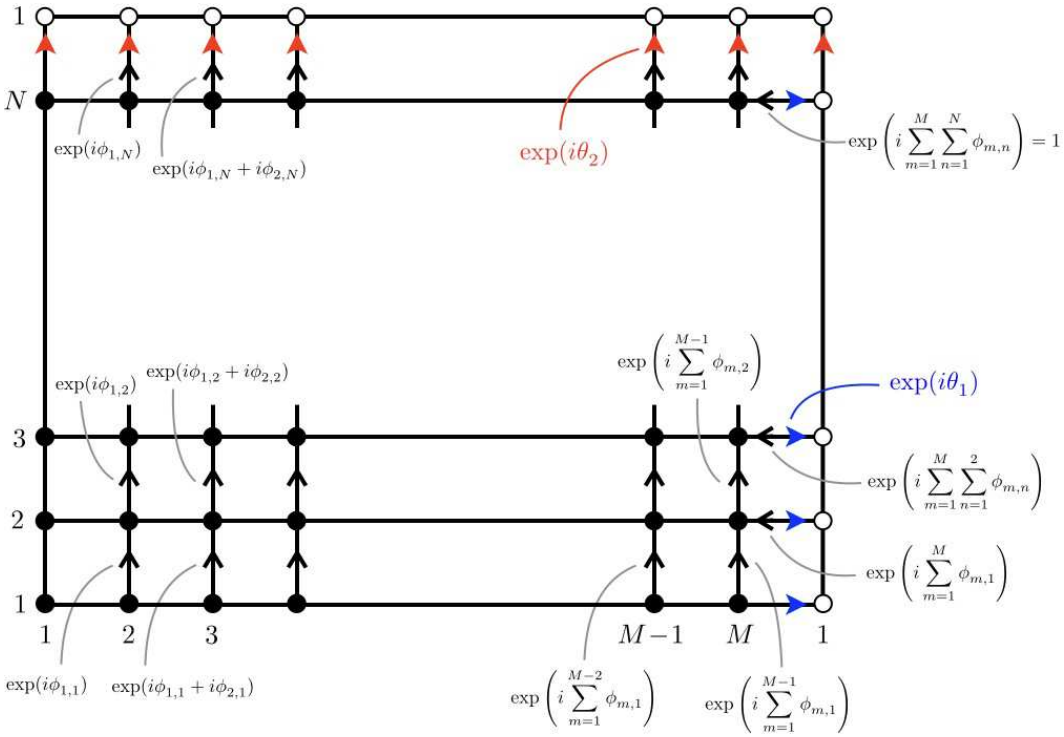

Figure 4.15: Lattice gauge field configuration for a general $M \times N$ rectangular lattice with flux $\phi_{m,n}$ in unit cell (m, n) and Bloch phases (θ_1, θ_2).

4.4 Topological Band Structures

4.4.1 SSH model

Consider the long chain polymer polyacetylene, with chemical formula $(\mathrm{CH})_x$, where x can be fairly large ($x \sim 10^4$ for example). In the *trans* form of $(\mathrm{CH})_x$, each C atom has two C neighbors and one H neighbor (see Fig. 4.16). The electronic structure of the carbon atom is $1\mathrm{s}^2\,2\mathrm{s}^2\,2\mathrm{p}^2$. The two $1\mathrm{s}$ electrons are tightly bound and are out of the picture. Three of the remaining $n = 2$ electrons form sp^2 planar hybrid bonding orbitals. This leaves one electron per carbon, which is denoted by π and has the symmetry of p_z, where backbone of the molecule is taken to lie in the (x, y) plane. If we were to model $(\mathrm{CH})_x$ by a one-dimensional tight binding model for these π-electrons, we'd write

$$H = -t \sum_{n=1}^{N} \sum_{\sigma=\pm} \left(c^\dagger_{n,\sigma}\, c_{n+1,\sigma} + c^\dagger_{n+1,\sigma}\, c_{n,\sigma} \right) = -2t \sum_{k,\sigma} \cos(ka)\, c^\dagger_{k,\sigma}\, c_{k,\sigma} \quad , \tag{4.132}$$

where σ is the spin polarization and where $c_{k,\sigma} = N^{-1/2} \sum_n e^{-ikna}\, c_{n,\sigma}$. We assume periodic boundary conditions $c_{n+N,\sigma} = c_{n,\sigma}$, which entails the mode quantization $k = 2\pi j/Na$ with

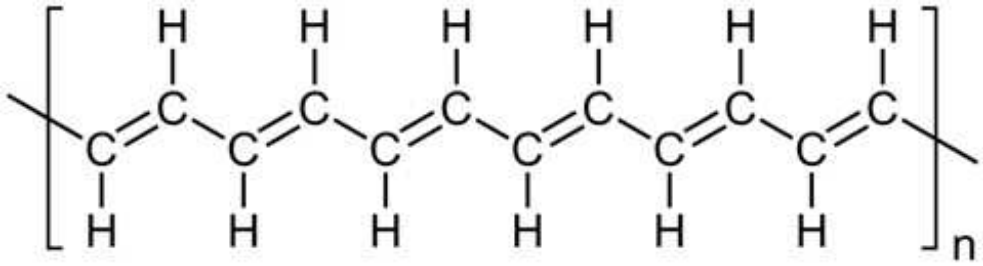

Figure 4.16: *trans*-polyacetylene, CH_x, consists of a backbone of C atoms with alternating single and double bonds.

$j \in \{1, \ldots, N\}$. In the thermodynamic limit $N \to \infty$ we may restrict $|k| \leqslant \frac{\pi}{a}$ to the first Brillouin zone.

The ground state is obtained by filling each of the $\frac{1}{2}N$ lowest energy states $\varepsilon_k = -2t\cos(ka)$ with a pair of $\uparrow$ and $\downarrow$ states, thus accounting for the N π-electrons present. Thus, with $\theta \equiv ka$, we have

$$E_0 = -4Nt \int_{-\pi/2}^{\pi/2} \frac{d\theta}{2\pi} \cos\theta = -\frac{4Nt}{\pi} \quad . \tag{4.133}$$

But there is a way for the system to lower its energy through *spontaneous dimerization*. Let's see how this works. We allow the nuclear centers of the carbon atoms to move a bit, and we write

$$H = -t \sum_{n=1}^{N} \Big(1 - \alpha(u_{n+1} - u_n)\Big)\Big(c^\dagger_{n,\sigma}\, c_{n+1,\sigma} + c^\dagger_{n+1,\sigma}\, c_{n,\sigma}\Big) + \sum_{n=1}^{N} \left(\frac{p_n^2}{2M} + \tfrac{1}{2}K(u_{n+1} - u_n)^2\right) \quad , \tag{4.134}$$

where u_n is the displacement of the $n^{\rm th}$ carbon atom and p_n is its momentum in the direction along the $(\mathrm{CH})_x$ backbone, and M is the atomic mass of carbon. The parameter α accounts for the exponential falloff of the π-electron wavefunctions from each nuclear center. We expect then that the hopping integral between atoms n and $n+1$ will be given by $t_{n,n+1} = t\, e^{-\alpha(u_{n+1} - u_n)}$, so $t_{n,n+1}$ increases if the distance $x_{n+1} - x_n = a + u_{n+1} - u_n$ decreases, *i.e.* if $u_{n+1} - u_n < 0$, and decreases if $u_{n+1} - u_n > 0$. The parameter α has dimensions of inverse length, and we presume that the lattice distortions are all weak, which licenses us to expand the exponential and write $t_{n,n+1} \approx t\,\big(1 - \alpha(u_{n+1} - u_n)\big)$. This is known as the Su-Schrieffer-Heeger (SSH) model[21].

For spontaneous dimerization, which breaks the lattice translation symmetry from $\mathbb{Z}_N$ to $\mathbb{Z}_{N/2}$, we write

$$u_n = (-1)^n\, \zeta + \delta u_n \quad . \tag{4.135}$$

[21] W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979).

The phonon part of the Hamiltonian now becomes

$$H_{\rm ph} = \overbrace{\sum_{n=1}^{N}\left(\frac{p_n^2}{2M} + \tfrac{1}{2}K\,(\delta u_{n+1} - \delta u_n)^2\right)}^{H^0_{\rm ph}} + 2NK\zeta^2 + 4K\zeta\sum_n (-1)^n\,\delta u_n \quad . \tag{4.136}$$

We assume N to be even. We take as a trial state for the phonons the ground state of $H^0_{\rm ph}$, which is a product of harmonic oscillator states for each of the phonon modes with

$$H^0_{\rm ph} = \sum_k \hbar\omega_k\left(A_k^\dagger\,A_k + \tfrac{1}{2}\right) \quad , \tag{4.137}$$

where $\omega_k = 2(K/M)^{1/2}\,|\sin(\frac{1}{2}ka)|$ is the phonon dispersion and

$$A_k = \left(\frac{1}{2\hbar M\omega_k}\right)^{1/2}\hat{p}_k + \left(\frac{M\omega_k}{2\hbar}\right)^{1/2}\delta\hat{u}_k \tag{4.138}$$

is the (first quantized) phonon annihilation operator, with $\begin{pmatrix}\hat{p}_k\\ \delta\hat{u}_k\end{pmatrix} = N^{-1/2}\sum_n e^{-ikna}\begin{pmatrix}p_n\\ \delta u_n\end{pmatrix}$. At $T=0$, we have $E_0^{\rm ph} = \frac{1}{2}\sum_k\hbar\omega_k = 4N\hbar\sqrt{K/M}$. Note also that in the phonon ground state $|\,\Psi_0^{\rm ph}\,\rangle$ we have $\langle\,\Psi_0^{\rm ph}\,|\,\delta u_n\,|\,\Psi_0^{\rm ph}\,\rangle = 0$ for all n. If we take the expectation value of H from Eqn. 4.134 in the state $|\,\Psi_0^{\rm ph}\,\rangle$, we obtain the effective electronic Hamiltonian

$$\begin{aligned} H_{\rm eff} &= E_0^{\rm ph} + 2NK\zeta^2 - \sum_{n=1}^{N_{\rm c}}\sum_\sigma\Big(t_1\,a^\dagger_{n,\sigma}\,b_{n,\sigma} + t_2\,b^\dagger_{n,\sigma}\,a_{n+1,\sigma} + {\rm H.c.}\Big) \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \sum_{k,\sigma}\begin{pmatrix}a^\dagger_{k,\sigma} & b^\dagger_{k,\sigma}\end{pmatrix}\begin{pmatrix}0 & t_1 + t_2\,e^{-ikb}\\ t_1 + t_2\,e^{ikb} & 0\end{pmatrix}\begin{pmatrix}a_{k,\sigma}\\ b_{k,\sigma}\end{pmatrix} \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 + \sum_{k,\sigma}\left|t_1 + t_2\,e^{ikb}\right|\left(\gamma^\dagger_{+,k,\sigma}\,\gamma_{+,k,\sigma} - \gamma^\dagger_{-,k,\sigma}\,\gamma_{-,k,\sigma}\right) \quad , \end{aligned} \tag{4.139}$$

where $N_{\rm c} = \frac{1}{2}N$, $b = 2a$ is the unit cell size, $a_{n,\sigma} \equiv c_{2n-1,\sigma}$, $b_{n,\sigma} \equiv c_{2n,\sigma}$, and

$$t_{1,2} = \big(1 \mp 2\alpha\zeta\big)\,t \quad . \tag{4.140}$$

We see that there are two states for each wavevector k and spin polarization σ, labeled by indices $\pm$, with energies $\varepsilon_{\pm,k,\sigma} = \pm\big|t_1 + t_2\,e^{ikb}\big|$. At $T=0$, we fill the bottom $(-)$ band with electrons of both spin polarizations, and the total energy is then (substituting $\theta \equiv kb \in [-\pi,\pi]$)

$$\begin{aligned} E_0^{\rm tot} &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \frac{N}{2\pi}\int_{-\pi}^{\pi}d\theta\,\sqrt{t_1^2 + t_2^2 + 2t_1t_2\cos\theta} \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \frac{4Nt}{\pi}\,\mathsf{E}(\kappa) \quad , \end{aligned} \tag{4.141}$$

where

$$\kappa = \frac{2\sqrt{t_1 t_2}}{|t_1 + t_2|} = \sqrt{1 - 4\alpha^2\zeta^2} \quad . \tag{4.142}$$

and where $\mathsf{E}(\kappa)$ is the complete elliptic integral of the second kind,

$$\mathsf{E}(\kappa) = \int_0^{\pi/2} d\vartheta \, \sqrt{1 - \kappa^2 \sin^2\vartheta} \quad . \tag{4.143}$$

We are interested in the case where $\alpha|\zeta| \ll 1$, *i.e.* small dimerization amplitude. Thus $\kappa \approx 1$, and in this limit we have[22]

$$\mathsf{E}(\kappa) = 1 + \tfrac{1}{2}\kappa'^2\Big(\ln(4/\kappa') - \tfrac{1}{2}\Big) + \ldots \quad , \tag{4.144}$$

where $\kappa' = \sqrt{1-\kappa^2} = 2\alpha\zeta$. We now have that the energy per site is given by

$$\frac{E_0^{\rm tot}}{N}(\zeta) = 4\hbar\sqrt{\frac{K}{M}} + 2K\zeta^2 - \frac{4t}{\pi} - \frac{8t}{\pi}\,\alpha^2\zeta^2 \ln\left(\frac{2}{\sqrt{e}\,\alpha\,\zeta}\right) + \ldots \quad . \tag{4.145}$$

Differentiating with respect to the dimerization parameter ζ, we obtain a minimum at

$$\zeta^* = \frac{2}{\sqrt{e}\,\alpha}\,\exp\left(-\frac{\pi K}{4\alpha^2 t}\right) \quad , \tag{4.146}$$

in the limit where $\alpha^2 t \ll K$. Thus, the system prefers to break the discrete translational symmetry of the one-dimensional lattice and to be spontaneously dimerized. The tendency for one-dimensional electronic systems coupled to phonons to spontaneously dimerize is known as the *Peierls instability*.

Topological aspects of the SSH model

We now suppress spin and investigate the following Hamiltonian,

$$\begin{aligned} H_0 &= -\sum_{n=1}^{N_{\rm c}} \Big(t_1\, a_n^\dagger\, b_n + t_2\, b_n^\dagger\, a_{n+1} + \text{H.c.} \Big) \\ &= -\sum_k \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} \begin{pmatrix} 0 & t_1 + t_2\, e^{-ikb} \\ t_1 + t_2\, e^{ikb} & 0 \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \quad , \end{aligned} \tag{4.147}$$

where the second and third lines are written in terms of second-quantized fermionic creation $(a^\dagger, b^\dagger)$ and annihilation (a, b) operators. This corresponds to a model with $N_{\rm c}$ unit cells where

[22] See Gradshteyn and Ryzhik, formula 8.114.3.

there are alternating strong (t_1) and weak (t_2) bonds, with periodic boundary conditions. The number of sites is $N = 2N_{\rm c}$. An arbitrary single particle eigenstate may be written

$$|\,\psi\,\rangle = \sum_{n=1}^{N_{\rm c}} \Big(A_n\, a_n^\dagger + B_n\, b_n^\dagger \Big)\,|\,0\,\rangle \quad . \tag{4.148}$$

In terms of the c-number coefficients $\{A_n, B_n\}$, the Schödinger equation may be written as the system

$$\begin{aligned} E A_n &= -t_2 B_{n-1} - t_1 B_n \\ E B_n &= -t_1 A_n - t_2 A_{n+1} \end{aligned} \quad , \tag{4.149}$$

with periodic boundary conditions $A_{n+N_{\rm c}} = A_n$ and $B_{n+N_{\rm c}} = B_n$. We can write this system in matrix form as

$$\begin{pmatrix} 0 & t_1 \\ t_2 & E \end{pmatrix} \begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = - \begin{pmatrix} E & t_2 \\ t_1 & 0 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} E^2 - t_1^2 & E\,t_2 \\ -E\,t_2 & -t_2^2 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad . \tag{4.150}$$

With translational invariance, we can demand

$$\begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = z \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad , \tag{4.151}$$

where $z = \exp(ikb)$. With this substitution in the above equation, we obtain the system

$$\begin{pmatrix} E^2 - t_1^2 - z\,t_1\,t_2 & E\,t_2 \\ -E\,t_2 & -t_2^2 - z\,t_1\,t_2 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} = 0 \quad , \tag{4.152}$$

and setting the determinant to zero recovers the dispersion $E = \pm|t_1 + z\,t_2|$.

Now suppose we remove the link between the b orbital in unit cell $n = N_{\rm c}$ and the a orbital in unit cell $n = 1$. We now have

$$\begin{pmatrix} A_2 \\ B_1 \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} E^2 - t_1^2 & 0 \\ -E\,t_2 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ B_{N_{\rm c}} \end{pmatrix} \quad , \quad \begin{pmatrix} A_{N_{\rm c}} \\ B_{N_{\rm c}-1} \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} 0 & -E\,t_2 \\ 0 & E^2 - t_1^2 \end{pmatrix} \begin{pmatrix} A_1 \\ B_{N_{\rm c}} \end{pmatrix} \quad , \tag{4.153}$$

as well as Eqn. 4.150 for $n \in \{2, \ldots, N_{\rm c} - 1\}$. We now show that there exist two eigenstates in the thermodynamic limit which are degenerate with $E = 0$ and which are exponentially localized in the vicinity of one of the ends of the chain provided $|r| < 1$, where $r = t_1/t_2$. With $E = 0$ we have from the first of the above equations that $A_2 = -rA_1$ and $B_1 = 0$. Iterating Eqn. 4.150, we then find $A_n = (-r)^{n-1} A_1$ and $B_n = 0$. In the limit $N_{\rm c} \to \infty$, so long as $|r| < 1$, we obtain a normalized wavefunction with $A_1 = \sqrt{1 - r^2}\, e^{i\alpha}$, where α is an arbitrary phase. A second zero mode is elicited by starting with the second of the above equations, which yields $B_{N_{\rm c}-n} = (-r)^n B_{N_{\rm c}}$ and $A_{N_{\rm c}-n} = 0$, which is normalized by setting $B_{N_{\rm c}} = \sqrt{1 - r^2}\, e^{i\beta}$. For $N_{\rm c}$ finite, these modes will mix and undergo level repulsion, resulting in two levels of opposite energy exponentially close to $E = 0$. Thus, for $|r| < 1$ there are two $E = 0$ edge states. For $|r| > 1$, there are no normalizable mid-gap (*i.e.* $E = 0$) modes. Of course, the midgap states *will* appear for $|r| > 1$ if we instead cut the bond between the a and b orbitals within the $n = 1$ cell.

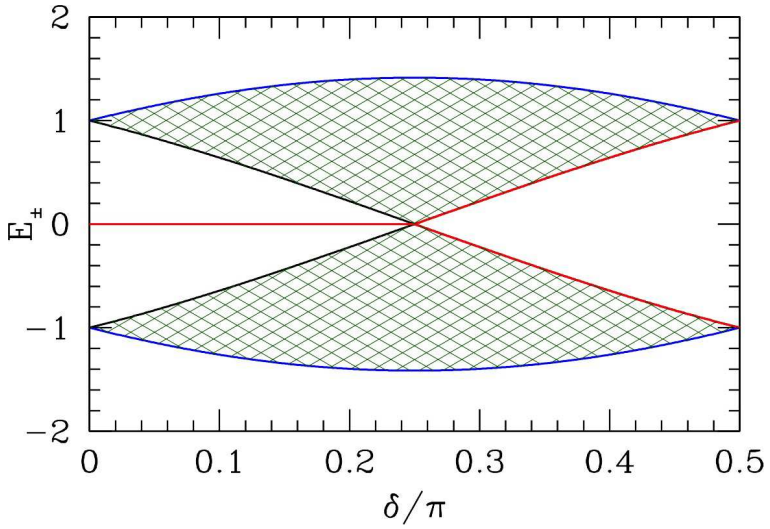

Figure 4.17: Spectrum of the SSH Hamiltonian on a finite chain with $N_{\rm c} \to \infty$ with $t_1 = t \sin\delta$ and $t_2 = t\cos\delta$. The two $E = 0$ edge states exist provided $|t_1| < |t_2|$, *i.e.* $|\delta| < \frac{\pi}{4}$.

4.4.2 Polarization and geometric phase

The individual cell functions for the Bloch states of the SSH model are given by the spinors

$$\vec{u}_\pm(k) = \begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix} \quad , \tag{4.154}$$

which are eigenstates of the Hamiltonian $H(k)$, *viz.*

$$\overbrace{-\begin{pmatrix} 0 & t(k) \\ t^*(k) & 0 \end{pmatrix}}^{H(k)} \overbrace{\begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix}}^{\vec{u}_\pm(k)} = E_\pm(k) \overbrace{\begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix}}^{\vec{u}_\pm(k)} \quad , \tag{4.155}$$

where $t(k) \equiv t_1 + t_2\, e^{-ikb}$ and $E_\pm(k) = \pm|t(k)|$. We define the *polarization* $P_\pm$ of each band as

$$P_\pm = i \int\limits_{-\pi/b}^{\pi/b} \frac{dk}{2\pi} \, \langle\, \vec{u}_\pm(k) \,|\, \partial_k \,|\, \vec{u}_\pm(k) \,\rangle = \int\limits_{-\pi/b}^{\pi/b} \frac{dk}{2\pi} \, A_\pm(k) \quad , \tag{4.156}$$

where

$$A_\pm(k) = \left\langle\, \vec{u}_\pm(k) \,\middle|\, i\,\frac{\partial}{\partial k} \,\middle|\, \vec{u}_\pm(k) \,\right\rangle \tag{4.157}$$

is the *Berry connection*, which plays a role similar to an electromagnetic vector potential. However, the polarization is defined only up to an integer multiple, because if we make an allowed gauge transformation $|\,\vec{u}_\pm(k)\,\rangle \to e^{-i\varphi(k)}\,|\,\vec{u}_\pm(k)\,\rangle$, where $e^{-i\varphi(k)}$ is *single-valued* on the Brillouin

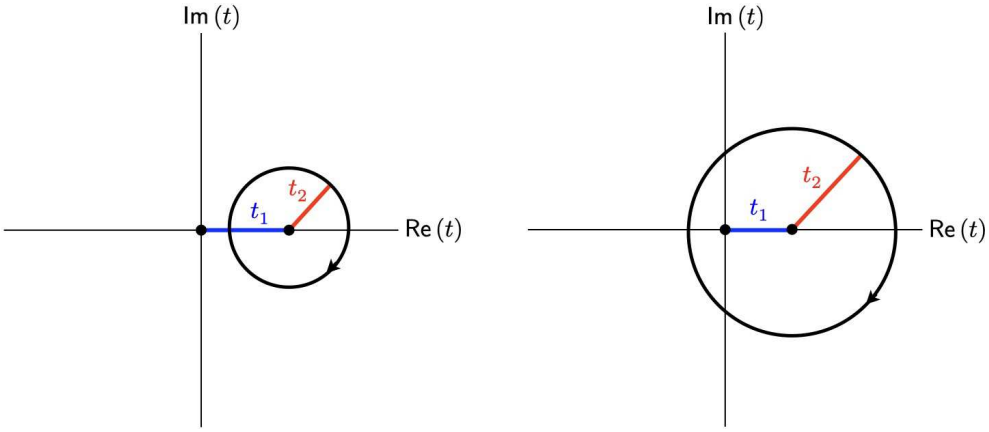

Figure 4.18: Winding of $t(k) = t_1 + t_2\, e^{-ikb}$ in the SSH model. Left: $|t_1| > |t_2|$, $W = 0$. Right: $|t_1| < |t_2|$, $W = -1$.

zone, then $P_\pm \to P_\pm + \big[\varphi(\pi/b) - \varphi(-\pi/b)\big]/2\pi = P_\pm + n$ where $n \in \mathbb{Z}$. Thus, it is the *phase* $\exp(2\pi i P_\pm)$ which is well-defined and gauge-invariant. We now compute this phase and show that it characterizes the two sectors $|r| > 1$ and $|r| < 1$, where $r \equiv t_1/t_2$.

Solving the Schrödinger equation for the cell functions, we obtain the solutions

$$u_{\mathsf{A}\pm}(k) = \frac{1}{\sqrt{2}} \quad , \qquad u_{\mathsf{B}\pm}(k) = \mp\frac{1}{\sqrt{2}}\,\frac{t^*(k)}{|t(k)|} = \mp\frac{1}{\sqrt{2}}\,e^{-i\Theta(k)} \quad , \tag{4.158}$$

where $\Theta(k) \equiv \arg t(k)$. Thus, $A_\pm(k) = \frac{1}{2}\frac{\partial\Theta}{\partial k}$ and

$$2\pi P_\pm = \frac{1}{2}\oint d\Theta = \pi W \quad , \tag{4.159}$$

where W is the integer *winding number* of the angle $\Theta(k)$ around the Brillouin zone $kb \in [-\pi, \pi]$. Examining the function $t(k) = t_1 + t_2\, e^{-ikb}$, we see that $\Theta(k)$ has winding $W = 0$ if $|t_1| > |t_2|$ and $W = -1$ if $|t_1| < |t_2|$. Thus, $\exp(2\pi i P_\pm) = +1$ for both $\pm$ bands in the nontopological sector where $|t_1| > |t_2|$, and $\exp(2\pi i P_\pm) = -1$ for both $\pm$ bands in the topological sector where $|t_1| < |t_2|$. Thus, the polarization phase neatly distinguishes the two sectors, which, as we saw in the previous section, also differ in their eigenspectra on open chains, where the topological sector exhibits two $E = 0$ edge states which are not present in the nontopological sector. The different winding sectors are depicted in Fig. 4.18.

Explicit breaking of translational symmetry

Suppose the A and B site ions are not both carbon, but are chemically different species. Then there will in general be a difference in the on-site energies for the individual π orbitals, and we

should write

$$H(k) = -\begin{pmatrix} m & t(k) \\ t^*(k) & -m \end{pmatrix} = -E_k \begin{pmatrix} \cos\theta_k & \sin\theta_k\, e^{i\phi_k} \\ \sin\theta_k\, e^{-i\phi_k} & -\cos\theta_k \end{pmatrix} \tag{4.160}$$

where $E_k = \sqrt{|t(k)|^2 + m^2}$ and

$$\cos\theta_k = \frac{m}{E_k} \quad , \quad \sin\theta_k\, e^{i\phi_k} = \frac{t(k)}{E_k} \quad . \tag{4.161}$$

The eigenvectors of $H(k)$ are

$$|\,\vec{u}_-(k)\,\rangle = \begin{pmatrix} \cos\frac{1}{2}\theta_k \\ \sin\frac{1}{2}\theta_k\, e^{i\phi_k} \end{pmatrix} \quad , \quad |\,\vec{u}_-(k)\,\rangle = \begin{pmatrix} \sin\frac{1}{2}\theta_k\, e^{i\phi_k} \\ -\cos\frac{1}{2}\theta_k \end{pmatrix} \quad , \tag{4.162}$$

with corresponding eigenvalues $\mp E_k$. We then have

$$\begin{aligned} A_\pm(k) &= \left\langle\, \vec{u}_\pm(k) \,\middle|\, i\,\frac{\partial}{\partial k} \,\middle|\, \vec{u}_\pm(k) \,\right\rangle = \pm\tfrac{1}{2}(1-\cos\theta_k)\,\frac{\partial\phi_k}{\partial k} \\ &= \frac{1}{2}\left(1 - \frac{m}{E_k}\right)\frac{\partial\,\mathsf{arg}\,t_k}{\partial k} \end{aligned} \tag{4.163}$$

Adopting again the definition $\Theta(k) = \phi_k = \mathsf{arg}\,t(k)$, note the following limits:

$$A_\pm(k\,;\,m=0) = \frac{1}{2}\,\frac{\partial\Theta}{\partial k} \quad , \quad A_\pm(k\,;\,m=-\infty) = \frac{\partial\Theta}{\partial k} \quad , \quad A_\pm(k\,;\,m=\infty) = 0 \quad . \tag{4.164}$$

For general m, we have that $\exp(2\pi i P_\pm)$ are no longer quantized at ± 1 and can take any value on the unit circle. The reason is that the Hamiltonian $H(k)$ in Eqn. 4.160 is intrinsically complex, which is a consequence of the breaking of inversion symmetry $\mathcal{I}$. You might think that the $m = 0$ case also breaks $\mathcal{I}$ as it is also complex, but in that case $H(k) = -\mathsf{Re}\,t(k)\,X + \mathsf{Im}\,t(k)\,Y$ is unitarily equivalent to a real Hamiltonian by rotating by $\frac{1}{2}\pi$ about the $\hat{\boldsymbol{x}}$ axis in internal space, *i.e.* by writing $H'(k) = U^\dagger H(k)\,U$ with $U = 2^{-1/2}(1+iX)$, whence $H'(k) = -\mathsf{Re}\,t(k)\,X + \mathsf{Im}\,t(k)\,Z$. When $m \to \pm\infty$, we obtain a trivial topology, because $\exp(2\pi i P_\pm) = +1$ in both limits[23]. Note also that for $m = 0$ the center of each link is a valid center of inversion symmetry, which is not the case when $m \neq 0$.

What happens to the edge states? Clearly they must split in energy because the sites on either end are members of different sublattices. Since the edge states we obtained previously live entirely on the A and B sublattices, respectively, their energies when $m \neq 0$ are, simply, $E_{\rm edge} = \pm m$. The bulk dispersion takes the form $E_{\rm bulk} = \pm E_k = \pm\sqrt{|t(k)|^2 + m^2}$. The edge states once again join the bulk spectrum when $|t_1| = |t_2|$, where $E_{k=\pi/b} = \pm m$.

[23] When $m = -\infty$ and $A_\pm(k) = \partial\Theta/\partial k$, we have $P_\pm = W$ rather than $P_\pm = \frac{1}{2}W$ as in the $m = 0$ case we studied previously, hence $\exp(2\pi i P_\pm) = +1$.

4.4.3 Domain wall states in the Dirac equation

Dirac Hamiltonians and Clifford algebras

A famous exchange (though perhaps somewhat apocryphal)[24]:

> Bohr to Dirac (∼1928): "What are you working on, Mr. Dirac?"
>
> Dirac to Bohr: "I am trying to take the square root of something."

What was Dirac trying to take the square root of? The Klein-Gordon equation, apparently:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 + \frac{m^2c^2}{\hbar^2}\right)\psi = 0 \quad . \tag{4.165}$$

Dirac didn't like the fact that the KG equation is second order in space and time. The famous Dirac Hamiltonian is given by

$$H = c\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta mc^2 \quad , \tag{4.166}$$

where $\{\alpha_1, \alpha_2, \alpha_3, \beta\}$ are mutually anticommuting Hermitian *matrices* that are elements of a *Clifford algebra*, which is an associative algebra consisting of the identity (1) and d elements Γ^μ with $\mu \in \{1, \ldots, d\}$, each with $\mathrm{Tr}\,\Gamma^\mu = 0$, and satisfying the relations

$$\left\{\Gamma^\mu, \Gamma^\nu\right\} = 2\delta^{\mu\nu} \quad , \tag{4.167}$$

where $\{\bullet, \bullet\}$ is the anticommutator[25]. Note that $(\Gamma^\mu)^2 = 1$ for each element of the algebra. This can be accomplished with matrices of rank $2^{[d/2]}$, where $[d/2]$ is the greatest integer less than or equal to $\frac{1}{2}d$, in the following manner. Starting with $d = 2$, we may take $\Gamma^1 = X$ and $\Gamma^2 = Y$, where X and Y are the Pauli matrices σ^x and σ^y. For $d = 3$, we add $\Gamma^3 = -i\Gamma^1\Gamma^2 = Z$. For $d = 4$, take

$$\Gamma^1 = X \otimes \mathbb{I} = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \\ +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \end{pmatrix} \quad , \quad \Gamma^2 = Y \otimes \mathbb{I} = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ +i & 0 & 0 & 0 \\ 0 & +i & 0 & 0 \end{pmatrix}$$
$$\Gamma^3 = Z \otimes X = \begin{pmatrix} 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad , \quad \Gamma^4 = Z \otimes Y = \begin{pmatrix} 0 & -i & 0 & 0 \\ +i & 0 & 0 & 0 \\ 0 & 0 & 0 & +i \\ 0 & 0 & -i & 0 \end{pmatrix} \quad , \tag{4.168}$$

[24] See P. A. M. Dirac and the Discovery of Quantum Mechanics by Kurt Gottfried (2010).

[25] More generally, we can take $d = t+s$ and $\{\Gamma^\mu, \Gamma^\nu\} = 2\eta^{\mu\nu}$, where $\eta^{\mu\nu} = 0$ if $\mu \neq \nu$, $\eta^{\mu\nu} = +1$ if $\mu = \nu \in \{1, \ldots, t\}$, and $\eta^{\mu\nu} = -1$ if $\mu = \nu \in \{t+1, \ldots, d\}$, which define the Clifford algebra $\mathcal{C}_{t,s}$. We restrict our attention to the Clifford algebras $\mathcal{C}_{d,0}$.

where $\mathbb{I}$ is the 2×2 unit matrix. For $d = 5$, we add

$$\Gamma^5 = -\Gamma^1\Gamma^2\Gamma^3\Gamma^4 = Z \otimes Z = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} \quad . \tag{4.169}$$

In general, for $d = 2k$ or $d = 2k + 1$, we may write

$$\begin{aligned} \Gamma^1 &= \overbrace{X \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I}}^{k \text{ terms}} & \Gamma^{2k-1} &= Z \otimes Z \otimes \cdots \otimes Z \otimes X \\ \Gamma^2 &= Y \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I} & \Gamma^{2k} &= Z \otimes Z \otimes \cdots \otimes Z \otimes Y \\ \Gamma^3 &= Z \otimes X \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I} & \Gamma^{2k+1} &= Z \otimes Z \otimes \cdots \otimes Z \otimes Z \, . \end{aligned} \tag{4.170}$$

The last of these pertains when $d = 2k + 1$ and is equivalent to

$$\Gamma^{2k+1} = (-i)^k \, \Gamma^1 \, \Gamma^2 \cdots \Gamma^{2k} \, . \tag{4.171}$$

Now you know something about Clifford algebras!

A very nice thing about Clifford algebras is that if we write

$$H = \boldsymbol{a} \cdot \boldsymbol{\Gamma} = a_\mu \, \Gamma^\mu \quad , \tag{4.172}$$

then

$$H^2 = a_\mu \, a_\nu \, \Gamma^\mu \, \Gamma^\nu = \tfrac{1}{2} \, a_\mu \, a_\nu \left\{ \Gamma^\mu, \Gamma^\nu \right\} = \boldsymbol{a}^2 \quad . \tag{4.173}$$

Since both H and H^2 are of rank 2^k, where $d = 2k$ or $d = 2k+1$, we have that the spectrum of H must consist of a 2^{k-1}-fold degenerate set of eigenvalues $E_- = -|\boldsymbol{a}|$, and a 2^{k-1}-fold degenerate set of eigenvalues $E_+ = +|\boldsymbol{a}|$.

Note that the dimension of the space of rank-n Hermitian matrices is n^2 (n independent real elements along the diagonal and $\frac{1}{2}n(n-1)$ independent complex elements above the diagonal). Thus for $n = 2^k$ we require 2^{2k} parameters to describe the most general Hamiltonian, whereas $H = \boldsymbol{a} \cdot \boldsymbol{\Gamma}$ is specified only $2k$ or $2k + 1$ parameters (or $2k + 2$ if we include real multiples of the identity). You may have noticed that the matrix $Z \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I}$, for example, is missing among the list Eqn. 4.170 for all $k > 1$. We can use the matrices in the Clifford algebra to build a full set of 2^{2k} independent Hermitian matrices, which will no longer be mutually anticommuting, by taking various products of matrices in the Clifford algebra. For example, with $k = 2$ and $d = 2k+1 = 5$ we have only five matrices in the CA but 15 independent traceless 4×4 Hermitian matrices in all. The missing 10 matrices are given by $\Gamma^{\mu\nu} = -i\Gamma^\mu\Gamma^\nu$ for $\mu < \nu$.

But we digress! Let us get back to the Dirac equation! For the Dirac Hamiltonian $H = \boldsymbol{a}(\boldsymbol{p}) \cdot \boldsymbol{\Gamma}$ with $\boldsymbol{a}(\boldsymbol{p}) = (p_x, p_y, p_z, m, 0)$, we have two massive Dirac cones with

$$E_\pm(\boldsymbol{p}) = \pm\sqrt{c^2\boldsymbol{p}^2 + m^2c^4} \quad . \tag{4.174}$$

Each energy level is doubly degenerate, corresponding to two electron spin polarizations. The gamma matrices are taken from the $k = 2$ Clifford algebra and are of rank four. In one space dimension, $H(p) = cpX + mc^2Z$ where X and Z are Pauli matrices (*i.e.* $k = 1$ Clifford algebra).

Domain wall bound states in $d = 1$

The bulk dispersion and ground state energy of the massive Dirac Hamiltonian are even functions of m. What happens if we continuously interpolate between negative and positive values of m? Consider the $d = 1$ Hamiltonian

$$H = -i\hbar c\, X\, \frac{\partial}{\partial x} + m(x)\, c^2\, Z = \begin{pmatrix} m(x)\, c^2 & -i\hbar c\, \partial_x \\ -i\hbar c\, \partial_x & -m(x)\, c^2 \end{pmatrix} \quad . \tag{4.175}$$

We have in mind a $m(x)$ smoothly interpolating between $m(-\infty) = -m$ and $m(+\infty) = +m$, but we can easily generalize to the case where $m(-\infty) = m_1$ and $m(+\infty) = m_2$. Over most of $x \in \mathbb{R}$, the mass $m(x)$ is roughly constant with either $m(x) = m_1$ or $m(x) = m_2$. In these regions, the bulk dispersion $E_{\pm,k} = \pm\sqrt{(\hbar ck)^2 + m_{1,2}^2\, c^4}$ holds. We now show that when the product $m_1 m_2 < 0$ that there is a precise zero energy eigenstate localized along the interface between the two regions[26]. Setting $E = 0$, the Schrödinger equation $H\vec{\psi}(x) = E\vec{\psi}(x)$ can be written as

$$\frac{\partial\vec{\psi}(x)}{\partial x} = -\frac{c}{\hbar}\, m(x)\, Y\, \vec{\psi}(x) \quad . \tag{4.176}$$

This requires that the spinor component of $\vec{\psi}(x)$ must be an eigenstate of Y, so we write

$$\vec{\psi}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i\eta \end{pmatrix} f(x) \quad , \tag{4.177}$$

where $\eta = \pm 1$, hence $Y\vec{\psi}(x) = \eta\,\vec{\psi}(x)$. Thus,

$$\frac{\partial \ln f(x)}{\partial x} = -\eta\, \frac{c}{\hbar}\, m(x) \qquad \Rightarrow \qquad f(x) = A\, \exp\left\{ -\,\eta\, \frac{c}{\hbar} \int_0^x dx'\, m(x') \right\} \quad , \tag{4.178}$$

where A is a normalization constant. In order to have a normalizable solution, though, we must have that $m(-\infty)\, m(+\infty) < 0$, in which case we choose

$$\eta = \mathsf{sgn}\big[m(+\infty) - m(-\infty)\big] \quad . \tag{4.179}$$

Helical edge states in $d = 2$

Let's now proceed to $d = 2$ and write

$$H = -i\hbar c\, \Gamma^1\, \frac{\partial}{\partial x} - i\hbar c\, \Gamma^2\, \frac{\partial}{\partial y} + m(x)\, c^2\, \Gamma^4 \quad . \tag{4.180}$$

[26] Two efficiently brief and readable texts on the general subject of topological bands in condensed matter, see S.-Q. Shen, *Topological Insulators* (Springer, 2012) or J. K. Asbóth, L. Oroszlány, and A. Pályi, *A Short Course on Topological Insulators* (Springer, 2016). More extensive treatments may be found in the now standard texts by Bernevig and Hughes (Princeton, 2013) and by Vanderbilt (Cambridge, 2018).

The mass $m(x)$ is presumed to be dependent only on the x coordinate, again interpolating between positive and negative values at $x = \pm\infty$. The wavefunction $\vec{\psi}(x,y)$ is now a four component spinor. We assume that $\vec{\psi}(x,y)$ corresponds to a plane wave in the y-direction, *i.e.* parallel to the domain wall. Note that $\Gamma^2 = Y \otimes \mathbb{I}$ has the following normalized eigenvectors:

$$\vec{\xi}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\i\\0\end{pmatrix} \quad , \quad \vec{\xi}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\i\end{pmatrix} \quad , \quad \vec{\xi}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-i\\0\end{pmatrix} \quad , \quad \vec{\xi}_4 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\-i\end{pmatrix} \quad , \tag{4.181}$$

where $\Gamma^2\,\vec{\xi}_{1,2} = +\vec{\xi}_{1,2}$ and $\Gamma^2\,\vec{\xi}_{3,4} = -\vec{\xi}_{3,4}$. We will also need the following, which you can easily check with $\Gamma^1 = X \otimes \mathbb{I}$ and $\Gamma^4 = Z \otimes Y$:

$$\begin{aligned}
&\Gamma^1\,\vec{\xi}_1 = +i\,\vec{\xi}_3 \quad , \quad \Gamma^1\,\vec{\xi}_2 = +i\,\vec{\xi}_4 \quad , \quad \Gamma^1\,\vec{\xi}_3 = -i\,\vec{\xi}_1 \quad , \quad \Gamma^1\,\vec{\xi}_4 = -i\,\vec{\xi}_2 \\
&\Gamma^4\,\vec{\xi}_1 = +i\,\vec{\xi}_4 \quad , \quad \Gamma^4\,\vec{\xi}_2 = -i\,\vec{\xi}_3 \quad , \quad \Gamma^4\,\vec{\xi}_3 = +i\,\vec{\xi}_2 \quad , \quad \Gamma^4\,\vec{\xi}_4 = -i\,\vec{\xi}_1 \quad .
\end{aligned} \tag{4.182}$$

The domain wall bound states are of two types, which we call I and II:

$$\begin{aligned}
\vec{\psi}_{\rm I}(x,y) &= A\,f(x)\,e^{ik_y y}\left(\alpha\,\vec{\xi}_1 + \beta\,\vec{\xi}_2\right) \\
\vec{\psi}_{\rm II}(x,y) &= B\,g(x)\,e^{ik_y y}\left(\gamma\,\vec{\xi}_3 + \delta\,\vec{\xi}_4\right) \quad .
\end{aligned} \tag{4.183}$$

For type I solutions, we write $E = \hbar c k_y$, in which case insertion into the Schrödinger equation $H\vec{\psi} = E\vec{\psi}$ yields the equation

$$\hbar c\left(\alpha\,\vec{\xi}_3 + \beta\,\vec{\xi}_4\right)\partial_x f + m(x)\,c^2\left(-i\beta\,\vec{\xi}_3 + i\alpha\,\vec{\xi}_4\right) f = 0 \quad . \tag{4.184}$$

The solvability condition here requires that the ratios of the $\vec{\xi}_4$ and $\vec{\xi}_3$ coefficients inside each of the round brackets must be the same, *i.e.* $\beta/\alpha = (-i\alpha)/(i\beta) = -\alpha/\beta$. Thus $\beta^2 = -\alpha^2$ and we may take

$$\begin{aligned}
(\alpha,\beta) = (1,+i) \;:\; \partial_x \ln f &= -\frac{c}{\hbar}\,m(x) \\
(\alpha,\beta) = (1,-i) \;:\; \partial_x \ln f &= +\frac{c}{\hbar}\,m(x) \quad .
\end{aligned} \tag{4.185}$$

For type II solutions, we write $E = -\hbar c k_y$, whence

$$-\hbar c\left(\gamma\,\vec{\xi}_1 + \delta\,\vec{\xi}_2\right)\partial_x g + m(x)\,c^2\left(-i\delta\,\vec{\xi}_1 + i\gamma\,\vec{\xi}_2\right) g = 0 \quad , \tag{4.186}$$

and we have the following possibilities:

$$\begin{aligned}
(\gamma,\delta) = (1,+i) \;:\; \partial_x \ln g &= +\frac{c}{\hbar}\,m(x) \\
(\gamma,\delta) = (1,-i) \;:\; \partial_x \ln g &= -\frac{c}{\hbar}\,m(x) \quad .
\end{aligned} \tag{4.187}$$

We therefore have the following solutions:

- Type I : $E = +\hbar c k_y$, $\Gamma^2\,\vec{\psi} = +\vec{\psi}$:

$$\begin{aligned}\vec{\psi}_{\rm Ia}(x,y) &= A\,e^{ik_y y}\,\exp\left\{-\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_1 + i\vec{\xi}_2\big)\\ \vec{\psi}_{\rm Ib}(x,y) &= A\,e^{ik_y y}\,\exp\left\{+\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_1 - i\vec{\xi}_2\big)\end{aligned} \tag{4.188}$$

  and

- Type II: $E = -\hbar c k_y$, $\Gamma^2\,\vec{\psi} = -\vec{\psi}$:

$$\begin{aligned}\vec{\psi}_{\rm IIa}(x,y) &= B\,e^{ik_y y}\,\exp\left\{+\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_3 + i\vec{\xi}_4\big)\\ \vec{\psi}_{\rm IIb}(x,y) &= B\,e^{ik_y y}\,\exp\left\{-\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_3 - i\vec{\xi}_4\big)\quad .\end{aligned} \tag{4.189}$$

Thus, if $m(-\infty) < 0 < m(+\infty)$, the normalizable solutions are $\vec{\psi}_{\rm Ia}$ and $\vec{\psi}_{\rm IIb}$, while for the case $m(-\infty) > 0 > m(+\infty)$, the normalizable solutions are $\vec{\psi}_{\rm Ib}$ and $\vec{\psi}_{\rm IIa}$. Note that the time-dependences are as follows:

$$\begin{aligned}\text{type I}\ :\ \vec{\psi}_{\rm I}(x,y) \propto f(x)\,e^{ik_y(y-ct)} \quad &\Rightarrow \quad \text{up-mover, } v_y = +c\\ \text{type II}\ :\ \vec{\psi}_{\rm II}(x,y) \propto g(x)\,e^{ik_y(y+ct)} \quad &\Rightarrow \quad \text{down-mover, } v_y = -c\quad .\end{aligned} \tag{4.190}$$

Note that the product of $\mathsf{sgn}(v_y)$ and the eigenvalue η of Γ^2 is $\chi = \eta\ \mathsf{sgn}(v_y) = +1$ in both cases. We call χ the *chirality*.

It should now come as no surprise that for the $d = 3$ Dirac Hamiltonian

$$H = -i\hbar c\,\Gamma^1\frac{\partial}{\partial x} - i\hbar c\,\Gamma^2\frac{\partial}{\partial y} - i\hbar c\,\Gamma^3\frac{\partial}{\partial z} + m(x)\,c^2\,\Gamma^4 \tag{4.191}$$

that we obtain two chiral domain-wall surface states with $E = \pm\hbar c\sqrt{k_y^2 + k_z^2}$.

4.4.4 The adiabatic theorem and Berry's phase

Consider a Hamiltonian $H(\boldsymbol{\lambda})$ dependent on a set of parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_K\}$, and let $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ satisfy the time-independent Schrödinger equation,

$$H(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = E_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \quad . \tag{4.192}$$

Now let $\boldsymbol{\lambda}(t)$ be continuously time-dependent, and consider the time-*dependent* Schrödinger equation

$$i\hbar\frac{d}{dt}\,|\,\Psi(t)\,\rangle = H\big(\boldsymbol{\lambda}(t)\big)\,|\,\Psi(t)\,\rangle \quad . \tag{4.193}$$

The adiabatic theorem states that if $\boldsymbol{\lambda}(t)$ evolves extremely slowly, then each solution $|\,\Psi_n(t)\,\rangle$ is proportional to $|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$, with

$$\big|\,\Psi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle = \exp\big(i\gamma_n(t)\big)\,\exp\!\left(-\,\frac{i}{\hbar}\int^t\!dt'\;E_n\big(\boldsymbol{\lambda}(t')\big)\right)\big|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle \quad , \tag{4.194}$$

with corrections which vanish in the limit $|\dot{\boldsymbol{\lambda}}|/|\boldsymbol{\lambda}| \to 0$. Taking the time derivative and then the overlap with the bra vector $\langle\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,|$, one obtains the result

$$\frac{d\gamma_n(t)}{dt} = i\,\big\langle\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big|\,\frac{d}{dt}\,\big|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle = \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})\cdot\frac{d\boldsymbol{\lambda}}{dt} \equiv \mathcal{A}_n(t) \quad , \tag{4.195}$$

where

$$\mathcal{A}_n^\mu(\boldsymbol{\lambda}) = i\,\big\langle\,\varphi_n(\boldsymbol{\lambda})\,\big|\,\frac{\partial}{\partial\lambda_\mu}\,\big|\,\varphi_n(\boldsymbol{\lambda})\,\big\rangle \tag{4.196}$$

is the *Berry connection*. We have already met with such a quantity in Eqn. 4.157 above. Note that $\mathcal{A}_n^\mu(\boldsymbol{\lambda})$ is real. In particular, if $\boldsymbol{\lambda}(t)$ traverses a closed loop $\mathcal{C}$ with infinitesimal speed, then the wavefunction $|\,\Psi_n(t)\,\rangle$ will accrue a *geometric phase* $\gamma_n(\mathcal{C})$, given by

$$\gamma_n(\mathcal{C}) = \oint\limits_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{4.197}$$

also called *Berry's phase*[27].

In the adiabatic limit, the dynamical phase $\hbar^{-1}\int^t dt'\;E_n\big(\boldsymbol{\lambda}(t')\big)$ becomes very large whenever $\langle E_n\rangle \neq 0$, because the path $\boldsymbol{\lambda}(t)$ is traversed very slowly. We may remove this dynamical phase by defining the Hamiltonian

$$\widetilde{H}_n(\boldsymbol{\lambda}) \equiv H(\boldsymbol{\lambda}) - E_n(\boldsymbol{\lambda}) \quad . \tag{4.198}$$

We define $|\,\widetilde{\Psi}_n(t)\,\rangle$ as the solution to the Schrödinger equation

$$i\hbar\,\frac{d}{dt}\,|\,\widetilde{\Psi}_n(t)\,\rangle = \widetilde{H}_n\big(\boldsymbol{\lambda}(t)\big)\,|\,\widetilde{\Psi}_n(t)\,\rangle \tag{4.199}$$

in the adiabatic limit. The adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ are the same as before, but now satisfy the zero energy condition $\widetilde{H}_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = 0$. Clearly $|\,\widetilde{\Psi}_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle = \exp\big(i\gamma_n(t)\big)\,|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$

[27] See M. V. Berry, *Proc. Roy. Soc. A* **392**, 45 (1984).

and the dynamical phase has been removed. However, note that the geometrical phase γ_n does not depend on the elapsed time, but only on the path traversed, *viz.*

$$\gamma_n = \gamma_n(\boldsymbol{\lambda}) = \int_{\boldsymbol{\lambda}_0}^{\boldsymbol{\lambda}} d\boldsymbol{\lambda}' \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}') \quad , \tag{4.200}$$

where $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(0)$, and where the integral is taken along the path in traversed by $\boldsymbol{\lambda}$.

4.4.5 Connection, curvature, and Chern numbers

The mathematical structure underlying this discussion is that of the *Hermitian line bundle*, the ingredients of which are (i) a *base space* $\mathcal{M}$ which is a topological manifold; this is the parameter space for $\boldsymbol{\lambda}$, and (ii) to each point $\boldsymbol{\lambda} \in \mathcal{M}$ is associated a *fiber* which is the adiabatic wavefunction $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathcal{H}$, which is a complex one-dimensional subspace of some Hilbert space $\mathcal{H}$. As $\boldsymbol{\lambda}$ moves around the base space $\mathcal{M}$, the fiber twists around. The adiabatic theorem furnishes us with a way of defining *parallel transport* of $|\,\widetilde{\Psi}_n(\boldsymbol{\lambda})\,\rangle$ along the curve $\mathcal{C}$[28]. The object $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ is the *connection* and the geometric phase $\gamma(\mathcal{C})$ is the *holonomy* of the connection[29]. As a holonomy, $\gamma(\mathcal{C})$ depends only on the curve $\mathcal{C}$ and not on where along the curve one starts.

The *curvature tensor* for the bundle is given by

$$\begin{aligned} \Omega_n^{\mu\nu}(\boldsymbol{\lambda}) &= \frac{\partial \mathcal{A}_n^\nu}{\partial \lambda_\mu} - \frac{\partial \mathcal{A}_n^\mu}{\partial \lambda_\nu} \\ &= i \left\langle \frac{\partial \varphi_n}{\partial \lambda_\mu} \middle| \frac{\partial \varphi_n}{\partial \lambda_\nu} \right\rangle - i \left\langle \frac{\partial \varphi_n}{\partial \lambda_\nu} \middle| \frac{\partial \varphi_n}{\partial \lambda_\mu} \right\rangle \quad . \end{aligned} \tag{4.201}$$

Using completeness of the $|\,\varphi_n\,\rangle$ basis, we may write the curvature tensor as

$$\Omega_n^{\mu\nu}(\boldsymbol{\lambda}) = i \sum_l{}' \frac{\left\langle \varphi_n \middle| \frac{\partial H}{\partial \lambda_\mu} \middle| \varphi_l \right\rangle \left\langle \varphi_l \middle| \frac{\partial H}{\partial \lambda_\nu} \middle| \varphi_n \right\rangle - (\mu \leftrightarrow \nu)}{(E_n - E_l)^2} \quad , \tag{4.202}$$

where the prime on the sum indicates that the term $l = n$ is to be excluded. We see that in this formulation the curvature tensor is actually independent of any phase convention for the adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$. So long as the adiabatic eigenstate $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ remains nondegenerate, the denominator in Eqn. 4.202 remains nonzero, hence the curvature tensor $\Omega(\boldsymbol{\lambda})$ is nonsingular. The same cannot be said about the connection $\boldsymbol{\mathcal{A}}(\boldsymbol{\lambda})$, however, because it is

[28] Note that $|\,\Psi_n(t)\,\rangle$, which depends explicitly on elapsed time and not solely on the position $\boldsymbol{\lambda}$ along its trajectory, can not be said to be parallel transported along any curve.

[29] See B. Simon, *Phys. Rev. Lett.* **51**, 2167 (1983).

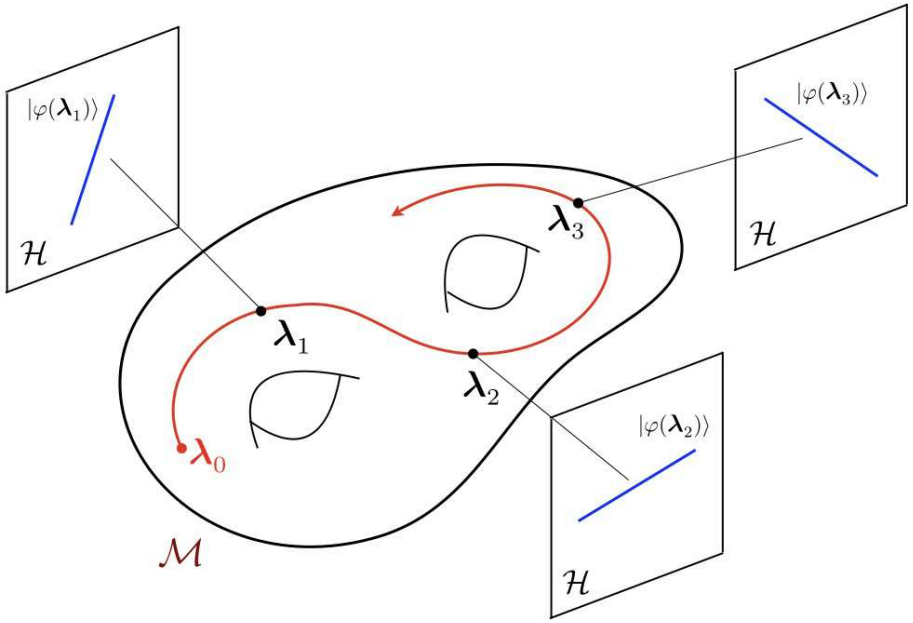

Figure 4.19: A Hermitian line bundle consists of a base space $\mathcal{M}$ and a fiber $|\,\varphi(\boldsymbol{\lambda})\,\rangle$ which twists as the point $\boldsymbol{\lambda}$ moves around the base space.

gauge-covariant. This means that if we multiply the adiabatic wavefunctions by phases, with $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \to \exp(if_n(\boldsymbol{\lambda}))\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$, the connection changes accordingly, with

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \to \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) - \frac{\partial f_n(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \quad . \tag{4.203}$$

How can we fix a gauge in order to give unambiguous meaning to $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$? One way might be to demand that the adiabatic wavefunction amplitude be real and positive at some particular point in space $\boldsymbol{r}_0$, *i.e.* $\langle\, \boldsymbol{r}_0\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathbb{R}_+$ for all $\boldsymbol{\lambda} \in \mathcal{M}$. For lattice-based models, where the adiabatic wavefunction is a vector of amplitudes for each orbital and each site within the appropriate unit cell, we could similarly demand that one of these amplitudes be real and positive. *This prescription fails if there exists a value of* $\boldsymbol{\lambda}$ *for which this wavefunction amplitude vanishes.*

As we are about to discover, the integral of the curvature over a two-dimensional base space is a topological invariant, meaning that it remains fixed (and indeed quantized) under continuous deformations of the Hamiltonian $H(\boldsymbol{\lambda})$. Using Stokes' theorem, we can turn an area integral of the curvature into line integrals of the connection. However, having chosen a particular gauge for the adiabatic wavefunctions, it may be that the connection is singular at certain points. Therefore the line integrals cannot be completely collapsed, and we obtain the result

$$\int\limits_{\mathcal{M}} d^2\lambda\; \Omega_n^{12}(\boldsymbol{\lambda}) = -\sum_i \oint\limits_{\mathcal{C}_i} d\boldsymbol{\lambda} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{4.204}$$

where the loop $\mathcal{C}_i$ encloses the $i^{\rm th}$ singularity $\boldsymbol{\lambda}_i$ of the connection in a counterclockwise man-

ner[30]. This is the generalization to Hermitian line bundles of the index formula in Eqn. 4.299 for the Gauss-Bonnet theorem. Quantization follows by writing $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = e^{iq_i\zeta(\boldsymbol{\lambda}-\boldsymbol{\lambda}_i)}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ in the vicinity of $\boldsymbol{\lambda} = \boldsymbol{\lambda}_i$, where q_i is an integer and

$$\zeta(\boldsymbol{\lambda}-\boldsymbol{\lambda}_i) = \tan^{-1}\left(\frac{\lambda_2-\lambda_{i,2}}{\lambda_1-\lambda_{i,1}}\right) \quad . \tag{4.205}$$

The integers q_i are chosen to 'unwind' the singularity at each $\boldsymbol{\lambda}_i$, so as to make the gauge transformed connection $\widetilde{\boldsymbol{\mathcal{A}}}{}^{\mu}_n(\boldsymbol{\lambda}) \equiv i\,\langle\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,|\,\boldsymbol{\nabla}_{\boldsymbol{\lambda}}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ nonsingular[31]. We then obtain

$$C_n \equiv \frac{1}{2\pi}\int\limits_{\mathcal{M}} d^2\lambda\;\Omega^{12}_n(\boldsymbol{\lambda}) = \sum_i q_i \quad . \tag{4.206}$$

Thus $C_n \in \mathbb{Z}$ is the *Chern number* of the Hermitian line bundle corresponding to the adiabatic wavefunction $|\,\varphi_n\,\rangle$.

The simplest nontrivial example is that of a spin-$\frac{1}{2}$ object in a magnetic field $\boldsymbol{B}(t)$, with

$$H(t) = g\mu_{\rm B}\boldsymbol{B}\cdot\boldsymbol{\sigma} = g\mu_{\rm B}B\begin{pmatrix}\cos\theta & \sin\theta\exp(-i\phi)\\ \sin\theta\exp(i\phi) & \cos\theta\end{pmatrix} \quad , \tag{4.207}$$

where $\boldsymbol{B} = B\,\hat{\boldsymbol{n}}$ is the adiabatic parameter which varies extremely slowly in time. The adiabatic wavefunctions are

$$|\,\varphi_+(\theta,\phi)\,\rangle = \begin{pmatrix}u\\ v\end{pmatrix} \qquad , \qquad |\,\varphi_-(\theta,\phi)\,\rangle = \begin{pmatrix}-\bar{v}\\ \bar{u}\end{pmatrix} \quad , \tag{4.208}$$

where $\hat{\boldsymbol{n}} = (\sin\theta\cos\phi\,,\,\sin\theta\sin\phi\,,\,\cos\theta)$, $u = \cos(\frac{1}{2}\theta)$, and $v = \sin(\frac{1}{2}\theta)\exp(i\phi)$. The energy eigenvalues are $E_\pm = \pm g\mu_{\rm B}B$. We now compute the connections,

$$\begin{aligned}\mathcal{A}_+ &= i\,\langle\,\varphi_+\,|\,\frac{d}{dt}\,|\,\varphi_+\,\rangle = i\big(\bar{u}\dot{u}+\bar{v}\dot{v}\big) = -\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = -\tfrac{1}{2}\,\dot{\omega}\\ \mathcal{A}_- &= i\,\langle\,\varphi_-\,|\,\frac{d}{dt}\,|\,\varphi_-\,\rangle = i\big(u\dot{\bar{u}}+v\dot{\bar{v}}\big) = +\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = +\tfrac{1}{2}\,\dot{\omega} \quad ,\end{aligned} \tag{4.209}$$

where $d\omega = \dot{\omega}\,dt$ is the differential solid angle subtended by the path $\hat{\boldsymbol{n}}(t)$. Thus, $\gamma_\pm(\mathcal{C}) = \mp\frac{1}{2}\,\omega_{\mathcal{C}}$ is $\mp$ half the solid angle subtended by the path $\hat{\boldsymbol{n}}_{\mathcal{C}}(t)$ on the Bloch sphere. We may now read off the components $\mathcal{A}^\theta_\pm = 0$ and $\mathcal{A}^\phi_\pm = \mp\frac{1}{2}(1-\cos\theta)$ and invoke Eqn. 4.201 to compute the curvature,

$$\Omega^{\theta\phi}_\pm(\theta,\phi) = \mp\tfrac{1}{2}\sin\theta \quad . \tag{4.210}$$

[30] We must assume that the base space $\mathcal{M}$ is orientable.

[31] Note that we have employed a singular gauge transformation, which is necessary to do the desired unwinding. Also note that the integers q_i should also carry a band index n, which we have suppressed here for notational simplicity.

Note then that the integral of the curvature over the entire sphere is given by

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \; \Omega_\pm^{\theta\phi}(\theta,\phi) = 2\pi\, C_\pm \quad , \tag{4.211}$$

where $C_\pm = \mp 1$ is the Chern number. Equivalently, note that both connections are singular at $\theta = \pi$, where the azimuthal angle is ill-defined. This singularity can be gentled through an appropriate singular gauge transformation $|\,\varphi_\pm\,\rangle = e^{\pm i\phi}\,|\,\widetilde{\varphi}_\pm\,\rangle = e^{\mp i\zeta}\,|\,\widetilde{\varphi}_\pm\,\rangle$, where ζ is defined to be the angle which increases as one winds counterclockwise around the *south pole*, hence $\zeta = -\phi$. This corresponds to $q_\pm = \mp 1$ in our earlier notation, hence again $C_\pm = \mp 1$.

As we saw above, this is a general result: when the base space $\mathcal{M}$ is two dimensional: the integral of the curvature over $\mathcal{M}$ is 2π times an integer. This result calls to memory the famous Gauss-Bonnet theorem (see §4.7 below for more), which says that the integral of the Gaussian curvature K over a two-dimensional manifold $\mathcal{M}$ is

$$\int_{\mathcal{M}} dS\; K = 4\pi(1-g) \quad , \tag{4.212}$$

where g is the *genus* (number of holes) in the manifold $\mathcal{M}$. In the Gauss-Bonnet case, the bundle construction is known as the *tangent bundle* of $\mathcal{M}$, and the corresponding connection is determined by the Riemannian metric one places on $\mathcal{M}$. However, *independent of the metric*, the integral of K is determined solely by the global topology of $\mathcal{M}$, *i.e.* by its genus. Thus, in three-dimensional space, a sphere inherits a metric from its embedding in $\mathbb{R}^3$. If you distort the sphere by denting it, locally its curvature K will change, being the product of the principal radii of curvature at any given point. But the integral of K over the surface will remain fixed at 4π. Just as the genus g of a Riemann surface is unaffected by simple deformations but can change if one does violence to it, such as puncturing and resewing it[32], so is the Chern number invariant under deformations of the underlying Hamiltonian, provided one does not induce a level crossing of the adiabatic eigenstate $|\,\varphi_n\,\rangle$ with a neighboring one. Also, note that if the connection $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ can be defined globally on $\mathcal{M}$, *i.e.* with no singularities, then $C_n = 0$.

4.4.6 Two-band models

For the two band ($S = \frac{1}{2}$) system with Hamiltonian $H = g\mu_{\rm B} B\,\hat{\boldsymbol{n}}(\boldsymbol{\lambda})\cdot\boldsymbol{\sigma}$, one can verify that we may also write the Chern numbers as[33]

$$C_\pm = \pm\frac{1}{4\pi}\int_{\mathcal{M}} d^2\lambda\; \hat{\boldsymbol{n}}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_1}\times\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_2} \quad . \tag{4.213}$$

[32] M. Gilbert's two commandments of topology: (I) Thou shalt not cut. (II) Thou shalt not glue.

[33] The dependence of the magnitude $B = |\boldsymbol{B}|$ on $\boldsymbol{\lambda}$ is irrelevant to the calculation of the Chern numbers. The equivalence $\hat{\boldsymbol{n}} = z^\dagger\boldsymbol{\sigma}z$ for $z = \begin{pmatrix} u \\ v \end{pmatrix}$ is known as the *first Hopf map* from CP^1 to $\mathbb{S}^2$.

In this formulation, the Chern number has the interpretation of a *Pontrjagin number*, which is a topological index classifying real vector bundles (more in §4.7.2 below). Thus, for a tight binding model on a bipartite lattice, the most general Hamiltonian may be written

$$H(\boldsymbol{k}) = d_0(\boldsymbol{k}) + \boldsymbol{d}(\boldsymbol{k}) \cdot \boldsymbol{\sigma} \quad , \tag{4.214}$$

where $\boldsymbol{k}$ is the wavevector and where each $d^\mu(\boldsymbol{k})$ is periodic under translations of $\boldsymbol{k}$ by any reciprocal lattice vector $\boldsymbol{G}$. In this case $\lambda_{1,2} = k_{x,y}$ are the components of $\boldsymbol{k}$, and $\mathcal{M} = \mathbb{T}^2$ is the Brillouin zone torus. Note that the sum of the Chern numbers for each of the $+$ and $-$ bands is zero. As we shall see below with the TKNN problem, for a larger spin generalization, *i.e.* when the magnetic unit cell contains more than two basis elements, the sum $\sum_a C_a$ of the Chern numbers over all bands also vanishes. Consider the two band model with

$$H(\boldsymbol{\theta}) = \begin{pmatrix} m - 2t\cos\theta_1 - 2t\cos\theta_2 & \Delta\,(\sin\theta_1 - i\sin\theta_2) \\ \Delta\,(\sin\theta_1 + i\sin\theta_2) & -m + 2t\cos\theta_1 + 2t\cos\theta_2 \end{pmatrix} \quad . \tag{4.215}$$

As before, $\theta_\mu = \boldsymbol{k}\cdot\boldsymbol{a}_\mu$. Note $H(\boldsymbol{\theta}) = \boldsymbol{d}(\boldsymbol{\theta})\cdot\boldsymbol{\sigma}$ with

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{\theta}) &= \big(\Delta\sin\theta_1\,,\,\Delta\sin\theta_2\,,\,m - 2t\cos\theta_1 - 2t\cos\theta_2\big) \\ &\equiv |\boldsymbol{d}|\,\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big) \quad . \end{aligned} \tag{4.216}$$

Note the adiabatic parameters here are θ_1 and θ_2, upon which ϑ and χ are parametrically dependent. Does $\boldsymbol{d}(\boldsymbol{\theta})$ wind around the Brillouin zone torus, yielding a nonzero Chern number?

First, you might be wondering, where does this model come from? Actually, it is the Hamiltonian for a $p_x + ip_y$ superconductor, but we can back out of $H(\boldsymbol{\theta})$ a square lattice insulator model involving two orbitals a and b which live on top of each other at each site, and are not spatially separated[34]. The parameter m reflects the difference in the local energies of the two orbitals. The nearest neighbor hopping integrals between like orbitals are $t_{aa} = t$ and $t_{bb} = -t$, but $t_{ab}(\pm\boldsymbol{a}_1) = \pm\frac{i}{2}\,\Delta$ and $t_{ab}(\pm\boldsymbol{a}_2) = \pm\frac{1}{2}\Delta$, with $t_{ba}(-\boldsymbol{a}_{1,2}) = t^*_{ab}(+\boldsymbol{a}_{1,2})$ due to hermiticity.

The energy eigenvalues are

$$E_\pm(\boldsymbol{\theta}) = \pm\sqrt{\Delta^2\sin^2\theta_1 + \Delta^2\sin^2\theta_2 + (m - 2t\cos\theta_1 - 2t\cos\theta_2)^2} \quad . \tag{4.217}$$

The Wigner-von Neumann theorem says that degeneracy for complex Hamiltonians like ours has codimension three, meaning one must fine tune three parameters in order to get a degeneracy. The reason is that for $H = \boldsymbol{d}\cdot\sigma$ describing two nearby levels, the gap is $2|\boldsymbol{d}|$, thus in order for the gap to vanish we must require three conditions: $d_x = d_y = d_z = 0$. For the real case where $d_y = 0$ is fixed, we only require two conditions, *i.e.* $d_x = d_z = 0$. For our model, the gap collapse requires

$$\begin{aligned} \Delta\,\sin\theta_1 &= 0 \\ \Delta\,\sin\theta_2 &= 0 \\ m - 2t\cos\theta_1 - 2t\cos\theta_2 &= 0 \quad . \end{aligned} \tag{4.218}$$

[34] In this model they are both s-orbitals, which is unphysical.

Thus, degeneracies occur at $(\theta_1, \theta_2) = (0, 0)$ when $m = 4t$, at (π, π) when $m = -4t$, and at $(0, \pi)$ and $(\pi, 0)$ when $m = 0$. It is clear that for $|m| > 4t$ both Chern numbers must be zero. This is because for $m > 4t$ we have $d_z(\theta_1, \theta_2) > 0$ for all values of the Bloch phases, while for $m < -4t$ we have $d_z(\theta_1, \theta_2) < 0$. Thus in neither case can the $\boldsymbol{d}$ vector wind around the Bloch sphere, and the Pontrjagin/Chern indices accordingly vanish for both bands.

Now consider the case $m \in [0, 4t]$. Recall that the eigenfunctions are given by

$$|\,\varphi_+\,\rangle = \begin{pmatrix} \cos(\frac{1}{2}\vartheta) \\ \sin(\frac{1}{2}\vartheta)\, e^{i\chi} \end{pmatrix} \quad , \quad |\,\varphi_-\,\rangle = \begin{pmatrix} -\sin(\frac{1}{2}\vartheta)\, e^{-i\chi} \\ \cos(\frac{1}{2}\vartheta) \end{pmatrix} \quad , \tag{4.219}$$

with eigenvalues $\pm|\boldsymbol{d}|$. The singularity in both $|\,\varphi_\pm(\theta_1, \theta_2)\,\rangle$ occurs at $\vartheta = \pi$. Recall that $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, which entails $\boldsymbol{d} = (0\,,\,0\,,\,-|\boldsymbol{d}|)$, *i.e.* $d_x = d_y = 0$ and $d_z < 0$. This only occurs for $(\theta_1, \theta_2) = (0, 0)$. All we need to do to compute the Chern numbers is to identify the singularity in $\zeta(\theta_1, \theta_2)$ about this point, *i.e.* does $\zeta = -\chi$ wind clockwise or counterclockwise, in which case $C_+ = -1$ or $C_+ = +1$, respectively. Treating $\theta_{1,2}$ as very small, one easily obtains $\zeta = -\tan^{-1}(\theta_2/\theta_1)$, which is to say clockwise winding, hence $C_\pm = \mp 1$. *Exercise* : Find $C_\pm$ for $m \in [-4t, 0]$.

Haldane honeycomb model

In §4.3.7, we met Haldane's famous honeycomb lattice model, $H(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) + \boldsymbol{d}(\boldsymbol{\theta}) \cdot \boldsymbol{\sigma}$, with

$$\begin{aligned} d_0(\boldsymbol{\theta}) &= -2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1 + \theta_2)\big]\cos\phi \\ d_x(\boldsymbol{\theta}) &= -t_1\big(1 + \cos\theta_1 + \cos\theta_2\big) \\ d_y(\boldsymbol{\theta}) &= t_1\big(\sin\theta_1 - \sin\theta_2\big) \\ d_z(\boldsymbol{\theta}) &= m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1 + \theta_2)\big]\sin\phi \quad . \end{aligned} \tag{4.220}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) \pm |\boldsymbol{d}(\boldsymbol{\theta})|$. Now is quite easy to demonstrate that $\big|\sin\theta_1 + \sin\theta_2 - \sin(\theta_1 + \theta_2)\big| \leqslant \frac{3}{2}\sqrt{3}$, and therefore that the $\boldsymbol{d}(\boldsymbol{\theta})$ cannot wind if $|m| > 3\sqrt{3}\,t_2\,|\sin\phi|$ and $C_\pm = 0$. As above, we set $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, and the singularity in both wavefunctions occurs at $\vartheta = \pi$, which requires $d_x(\boldsymbol{\theta}) = d_y(\boldsymbol{\theta}) = 0$ and $d_z(\boldsymbol{\theta}) < 0$. This in turn requires $\theta_1 = \theta_2 = \frac{2}{3}\pi s$ where $s = \mathsf{sgn}(\sin\phi)$. We now write $\theta_j = \frac{2}{3}\,\pi s + \delta_j$ and find

$$\tan\chi = \frac{d_y(\boldsymbol{\theta})}{d_x(\boldsymbol{\theta})} = \left(\frac{\delta_2 - \delta_1}{\delta_1 + \delta_2}\right)\mathsf{sgn}(\sin\phi) = s\,\tan(\alpha - \tfrac{\pi}{4}) \quad , \tag{4.221}$$

where $\boldsymbol{\delta} \equiv |\boldsymbol{\delta}|\,(\cos\alpha, \sin\alpha)$. Thus, $\zeta = -\chi$ winds in the same sense as α if $s < 0$ and in the opposite sense if $s > 0$. Thus we conclude $C_\pm = \mp\mathsf{sgn}(\sin\phi)$. The topological phase diagram for the Haldane honeycomb lattice model is shown in Fig. 4.20. The phase space is a cylinder in the dimensionless parameters $\phi \in [-\pi, \pi]$ and $m/t_2 \in \mathbb{R}$. Regions are labeled by the Chern numbers $C_\pm$ of the two energy bands.

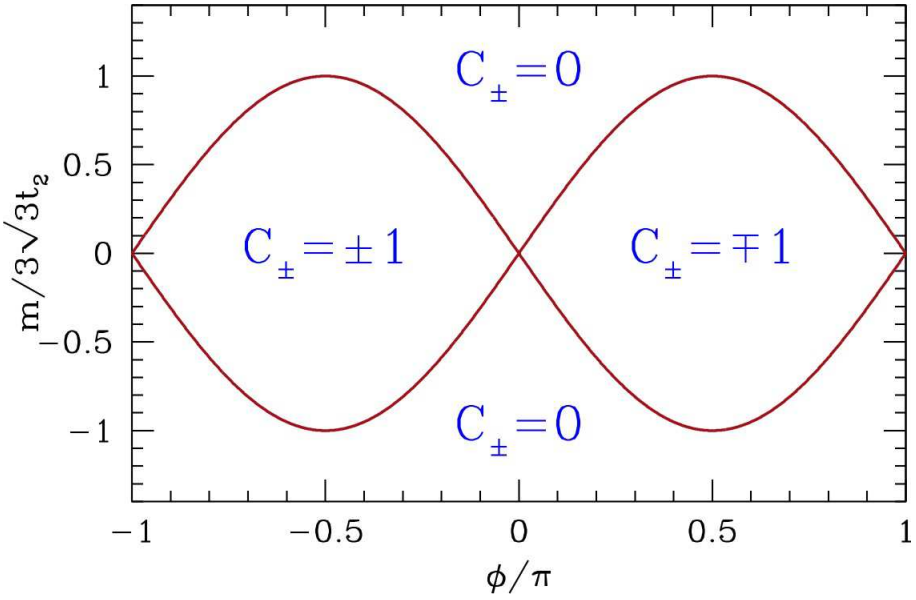

Figure 4.20: Topological phase diagram for the Haldane honeycomb lattice model, in which regions in the $(\sin\phi, m/t_2)$ cylinder are labeled by the Chern numbers $C_\pm$ of the energy bands.

Note on broken symmetries

We saw above how the band structure of spinless π-orbitals on graphene results in two Dirac points at wavevectors $\boldsymbol{K}$ and $\boldsymbol{K}' = -\boldsymbol{K}$. If we adopt a pseudospin convention for the valleys, with Pauli matrices $\boldsymbol{\tau}$, and where $\tau^z = \pm 1$ corresponds to the $\pm\boldsymbol{K}$ valley, it turns out that we may in one stroke write the long wavelength graphene Hamiltonian as

$$H_0 = \tfrac{\sqrt{3}}{2}\, ta\big(q_x\, \sigma^x\, \tau^z + q_y\, \sigma^y\big) \quad , \tag{4.222}$$

where σ^y again operates in the A/B sublattice space. This Hamiltonian is symmetric under the operations of parity ($\mathcal{P}$) and time-reversal ($\mathcal{T}$). Under $\mathcal{P}$, we switch valleys, switch sublattices, and send $q_x \to -q_x$. Under $\mathcal{T}$, we switch valleys and send $\boldsymbol{q} \to -\boldsymbol{q}$. It is also important to remember that $\mathcal{T}$ is antiunitary, and includes the complex conjugation operator $\check{K}$. The matrix parts of these operators, *i.e.* other than their actions on the components of $\boldsymbol{q}$, are given by

$$\mathcal{P} = \sigma^y\, \tau^z \qquad , \qquad \mathcal{T} = i\tau^y \check{K} \quad . \tag{4.223}$$

Note that $\mathcal{T}^2 = -1$ and $\mathcal{T}^{-1} = -\mathcal{T} = \check{K}\tau^y\,(-i)$. Of course $\mathcal{P}^2 = 1$ and thus $\mathcal{P}^{-1} = \mathcal{P}$. One can now check explicitly that $\mathcal{P}H_0\mathcal{P}^{-1} = \mathcal{T}H_0\mathcal{T}^{-1} = H_0$.

There are three ways to introduce a gap into the model, *i.e.* to gap out the dispersion at the $\boldsymbol{K}$ and $\boldsymbol{K}'$ points at the two inequivalent Brillouin zone corners:

1. The first way is by introducing a Semenoff mass term, which is of the form $V_{\rm S} = \Delta_{\rm S}\, \sigma^z$. This turns graphene into boron nitride (BN), distinguishing the local π-orbital energies

on the B and N sites. One can check that

$$\mathcal{P}\,\sigma^z\,\mathcal{P}^{-1} = -\sigma^z \quad , \quad \mathcal{T}\,\sigma^z\,\mathcal{T}^{-1} = +\sigma^z \quad , \tag{4.224}$$

and therefore the Semenoff mass breaks parity and preserves time-reversal.

2. The second way comes from the Haldane honeycomb lattice model at $m = 0$, where $V_{\rm H} = \Delta_{\rm H}\,\sigma^z\tau^z$.

$$\mathcal{P}\,\sigma^z\tau^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z \quad , \quad \mathcal{T}\,\sigma^z\tau^z\,\mathcal{T}^{-1} = -\sigma^z\tau^z \quad . \tag{4.225}$$

This term, the Haldane mass, preserves parity but breaks time-reversal. It leads to a topological band structure in which the bands are classified by nonzero Chern numbers.

3. The third way involves introducing the physical electron spin, and arises from spin-orbit effects. It essentially is described by two copies of the Haldane model, in which the up and down spin electrons couple oppositely to magnetic flux. This was first discussed by Kane and Mele[35], and is described by the perturbation $V_{\rm KM} = \Delta_{\rm KM}\,\sigma^z\tau^z s^z$. where $\boldsymbol{s}$ is the electron spin operator . The Kane-Mele mass term preserves $\mathcal{P}$ and $\mathcal{T}$ symmetries:

$$\mathcal{P}\,\sigma^z\tau^z s^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z s^z \quad , \quad \mathcal{T}\,\sigma^z\tau^z s^z\,\mathcal{T}^{-1} = +\sigma^z\tau^z s^z \quad . \tag{4.226}$$

Therefore, following the tried and true rule in physics that "everything which is not forbidden is compulsory", there *must* be a KM mass term in real graphene. The catch is that it is extremely small because graphene is a low-Z atom, and first principles calculations[36] conclude that the spin-orbit gap is on the order of $10\,$mK – too small to be observed due to finite temperature and disorder effects. However, there are many materials (Bi bilayers, HgTe/CdTe heterostructures, various three-dimensional materials such as α-Sn, $\rm Bi_{\it x}Sb_{1-\it x}$ and others) where the effect is predicted to be sizable and where it is indeed observed. This is the essence of topological insulator behavior.

4.4.7 The TKNN formula

Recall the Hamiltonian of Eqn. 4.120 for the isotropic square lattice Hofstadter model with flux $\phi = 2\pi p/q$ per unit cell. A more general version, incorporating anisotropy which breaks 90°

[35] C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
[36] See Y. Yao *et al.*, *Phys. Rev. B* **75**, 041401(R) (2007).

rotational symmetry, is given by[37]

$$H(\theta_1,\theta_2) = -\begin{pmatrix} 2t_2\cos\theta_2 & t_1 & 0 & \cdots & 0 & t_1\,e^{-i\theta_1} \\ t_1 & 2t_2\cos\left(\theta_2+\frac{2\pi p}{q}\right) & t_1 & & & 0 \\ 0 & t_1 & 2t_2\cos\left(\theta_2+\frac{4\pi p}{q}\right) & t_1 & & \vdots \\ \vdots & 0 & t_1 & \ddots & & \vdots \\ & & & & & t_1 \\ 0 & & & & t_1 & 2t_2\cos\left(\theta_2+\frac{2\pi(q-1)p}{q}\right) \\ t_1\,e^{i\theta_1} & 0 & & \cdots & t_1 & \end{pmatrix} . \tag{4.227}$$

This is a $q\times q$ matrix, and the q eigenvectors $|\,\varphi_n(\boldsymbol{\theta})\,\rangle$ are labeled by a band index $n\in\{1,\ldots,q\}$, with component amplitudes $\varphi_{a,n}(\boldsymbol{\theta})$ satisfying

$$H_{aa'}(\boldsymbol{\theta})\,\varphi_{a',n}(\boldsymbol{\theta}) = E_n(\boldsymbol{\theta})\,\varphi_{a,n}(\boldsymbol{\theta}) \quad . \tag{4.228}$$

From Wigner-von Neumann, we expect generically that neighboring bands will not cross as a function of the two parameters (θ_1,θ_2), because degeneracy has codimension three. Thus, associated with each band n is a Chern number C_n. By color coding each spectral gap according to the Chern number of all bands below it, J. Avron produced a beautiful and illustrative image of Hofstadter's butterfly, shown in Fig. 4.21 for the isotropic square lattice and in Fig. 4.22 for the isotropic honeycomb lattice.

It turns out that the Chern number is not just an abstract topological index. It is in fact the dimensionless Hall conductivity σ_{xy} itself, provided the Fermi level lies in a gap between magnetic subbands. This was first discovered by Thouless, Kohmoto, Nightingale, and den Nijs, in a seminal paper known by its authors' initials, TKNN[38]. In fact, we've developed the theory here in reverse chronological order. First came TKNN, who found that the contribution $\sigma_{xy}^{(n)}$ to the total Hall conductivity from a band lying entirely below the Fermi level is given by $\sigma_{xy}^{(n)} = \frac{e^2}{h}\,C_n$, where

$$C_n = \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_1 \int\limits_0^{2\pi} d\theta_2 \left(\left\langle \frac{\partial\varphi_n}{\partial\theta_1}\,\middle|\,\frac{\partial\varphi_n}{\partial\theta_2}\right\rangle - \left\langle \frac{\partial\varphi_n}{\partial\theta_2}\,\middle|\,\frac{\partial\varphi_n}{\partial\theta_1}\right\rangle \right) \tag{4.229}$$

is an integral over the Brillouin zone. They proved that this expression is an integer, because invoking Stokes' theorem,

$$C_n = \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_2 \left\langle \varphi_n\,\middle|\,\frac{\partial\varphi_n}{\partial\theta_2}\right\rangle\Bigg|_{\theta_1=0}^{\theta_1=2\pi} - \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_1 \left\langle \varphi_n\,\middle|\,\frac{\partial\varphi_n}{\partial\theta_1}\right\rangle\Bigg|_{\theta_2=0}^{\theta_2=2\pi} \quad . \tag{4.230}$$

[37] We drop the hat on $\hat H(\boldsymbol{\theta})$ but fondly recall that $\hat H_{aa'}(\boldsymbol{\theta})$ is the lattice Fourier transform of $H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')$.
[38] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).

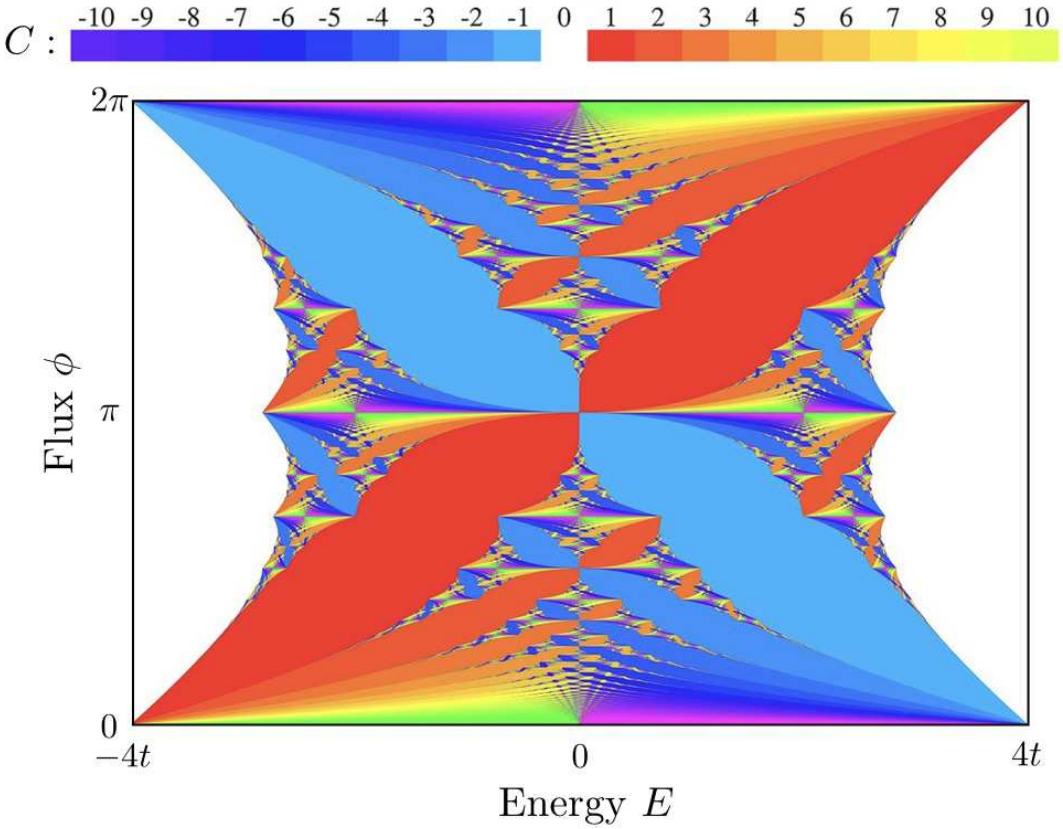

Figure 4.21: Avron's depiction of the Hofstadter butterfly for the isotropic square lattice system. The band gap regions are color coded by Chern number, C, which is the sum of the Chern numbers of all bands below a given gap. White regions correspond to $C = 0$. See J. E. Avron, *Colored Hofstadter butterflies*, in *Multiscale Methods in Quantum Mechanics*, P. Blanchard and G. Dell'Antonio, eds. (Birkhäuser, 2004).

But since $H(\theta_1, \theta_2)$ is doubly periodic with period 2π in each of its arguments, we must have

$$\begin{aligned} |\,\varphi_n(\theta_1, 2\pi)\,\rangle &= e^{if_n(\theta_1)}\,|\,\varphi_n(\theta_1, 0)\,\rangle \\ |\,\varphi_n(2\pi, \theta_2)\,\rangle &= e^{ig_n(\theta_2)}\,|\,\varphi_n(0, \theta_2)\,\rangle \quad . \end{aligned} \tag{4.231}$$

Thus, one finds

$$C_n = \frac{1}{2\pi}\Big(f_n(2\pi) - f_n(0) + g_n(0) - g_n(2\pi)\Big) \quad . \tag{4.232}$$

But we also have

$$\begin{aligned} |\,\varphi_n(0,0)\,\rangle &= e^{-if_n(0)}\,|\,\varphi_n(0,2\pi)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,|\,\varphi_n(2\pi,2\pi)\,\rangle \\ &= e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,|\,\varphi_n(2\pi,0)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,e^{ig_n(0)}\,|\,\varphi_n(0,0)\,\rangle \quad , \end{aligned} \tag{4.233}$$

and therefore $\exp(2\pi i\, C_n) = 1$ and $C_n \in \mathbb{Z}$. But just as Berry didn't know he had found a holonomy, TKNN didn't know they had found a Chern number. That mathematical feature

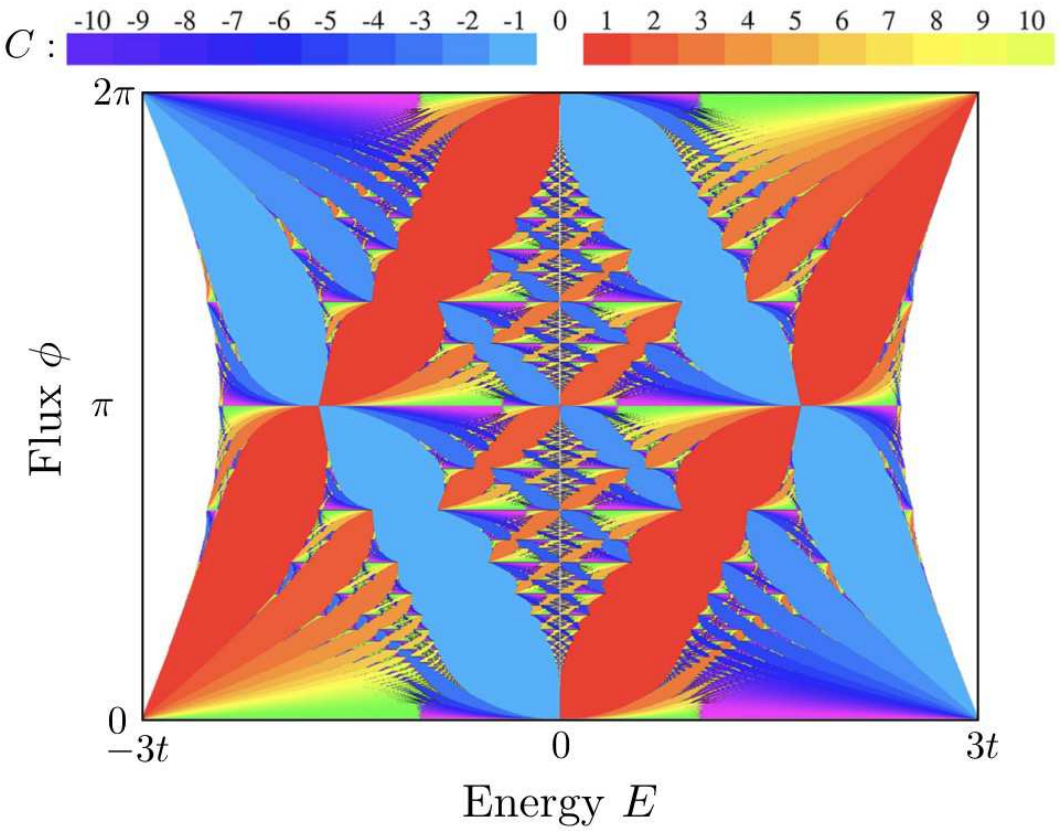

Figure 4.22: Colored Hofstadter butterfly for the honeycomb lattice system, from A. Agazzi, J.-P. Eckman, and G. M. Graf, *J. Stat. Phys.* **156**, 417 (2014).

was first elucidated by Avron, Seiler, and Simon[39], in a paper which is widely appreciated but which, understandably, is not known by its authors' initials.

To see why Hall conductivity is related to Berry curvature, consider an electric field $\boldsymbol{E} = E_y\,\hat{\boldsymbol{y}}$, and the single electron Hamiltonian $H(E_y) = H(0) - eE_y y$, where $H(0) = \frac{\pi^2}{2m} + V(\boldsymbol{r})$ has eigenstates $|\,\alpha\,\rangle$ and eigenvalues ε_α. First order perturbation theory in the electric field term says

$$|\,\alpha'\,\rangle = |\,\alpha\,\rangle - eE_y \sum_\beta{}' \frac{|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{4.234}$$

where the prime on the sum means the term with $\beta = \alpha$ is excluded. Let's now compute the expectation of the velocity operator v_x in the perturbed state $|\,\alpha'\,\rangle$. We have, to lowest order,

$$\langle\,\alpha'\,|\,v_x\,|\,\alpha'\,\rangle = -eE_y \sum_\beta{}' \frac{\langle\,\alpha\,|\,v_x\,|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle + \langle\,\alpha\,|\,y\,|\,\beta\,\rangle\langle\,\beta\,|\,v_x\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \tag{4.235}$$

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

We now invoke the Feynman-Hellman theorem, which says

$$\langle\,\alpha\,|\,y\,|\,\beta\,\rangle = \frac{\hbar}{i}\,\frac{\langle\,\alpha\,|\,v_y\,|\,\beta\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{4.236}$$

multiply by the electron charge $-e$, divide by the area of the system Ω, and sum using the Fermi distribution over the levels $|\,\alpha\,\rangle$, to obtain the current density j_x :

$$j_x = E_y \cdot \frac{e^2}{h} \cdot \frac{2\pi i\hbar^2}{\Omega} \sum_\alpha \sum_\beta f_\alpha\,(1-f_\beta)\,\epsilon_{ij}\,\frac{\langle\,\alpha\,|\,v_i\,|\,\beta\,\rangle\langle\,\beta\,|\,v_j\,|\,\alpha\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} \quad , \tag{4.237}$$

where f_α is the Fermi function at temperature T, chemical potential μ, and energy ε_α. The above expression for $\sigma_{xy} = j_x/E_y$ is known as the *Kubo formula* for the Hall conductivity. At $T = 0$, the Fermi distribution becomes the step function $f_\alpha = \Theta(E_{\mathsf{F}} - E_\alpha)$.

Suppose our system lies on a torus defined by the spatial periods $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$. Define the gauge transformed Hamiltonian

$$\widetilde{H}(\boldsymbol{\theta}) \equiv e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,H\,e^{+i\boldsymbol{q}\cdot\boldsymbol{r}} \quad , \tag{4.238}$$

where

$$\boldsymbol{q} = \theta_2\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_1}{\Omega} - \theta_1\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_2}{\Omega} \quad , \tag{4.239}$$

with $\Omega = \hat{\boldsymbol{z}}\cdot\boldsymbol{L}_1\times\boldsymbol{L}_2 = 2\pi\ell^2 p$ with $p\in\mathbb{Z}$, *i.e.* the total magnetic flux through the system is an integer multiple of the Dirac quantum. Then

$$\frac{\partial\widetilde{H}}{\partial\theta_i} = \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,\hbar\boldsymbol{v}\,e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \equiv \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot\hbar\widetilde{\boldsymbol{v}} \quad , \tag{4.240}$$

because $[H,\boldsymbol{r}] = (\hbar/i)\,\boldsymbol{v}$. Thus, defining $|\,\widetilde{\alpha}\,\rangle \equiv \exp(-i\boldsymbol{q}\cdot\boldsymbol{r})\,|\,\alpha\,\rangle$, and recalling the definition of the wavevector $\boldsymbol{q} = \epsilon_{ab}\,\theta_a\,\boldsymbol{L}_b\times\hat{\boldsymbol{z}}/\Omega$, we find

$$\frac{\partial\widetilde{H}}{\partial\theta_a} = \hbar\,\epsilon_{ab}\,\epsilon_{ij}\,\frac{\widetilde{v}_i\,L_b^j}{\Omega} \quad . \tag{4.241}$$

We then find

$$\sigma_{xy} = \frac{j_x}{E_y} = \sum_{\alpha\ \mathrm{occ}}\sigma_{xy}^{(\alpha)} \quad , \tag{4.242}$$

where the sum is over occupied states below the Fermi level, and where

$$\sigma_{xy}^{(\alpha)} = \frac{e^2}{h}\cdot 2\pi i\,{\sum_\beta}'\epsilon_{ij}\,\frac{\langle\,\widetilde{\alpha}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_i}\,|\,\widetilde{\beta}\,\rangle\langle\,\widetilde{\beta}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_j}\,|\,\widetilde{\alpha}\,\rangle}{(\varepsilon_\alpha-\varepsilon_\beta)^2} \quad , \tag{4.243}$$

which is precisely of the form of Eqn. 4.202. Thus, if we now *uniformly average over the boundary phases* θ_1 *and* θ_2 , we obtain

$$\left\langle \sigma_{xy}^{(\alpha)} \right\rangle = \frac{e^2}{h} \cdot \frac{i}{2\pi} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \sum_\beta{}' \epsilon_{ij} \frac{\left\langle \widetilde{\alpha} \,\middle|\, \frac{\partial \widetilde{H}}{\partial \theta_i} \,\middle|\, \widetilde{\beta} \right\rangle \left\langle \widetilde{\beta} \,\middle|\, \frac{\partial \widetilde{H}}{\partial \theta_j} \,\middle|\, \widetilde{\alpha} \right\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} = \frac{e^2}{h} C_\alpha \quad , \tag{4.244}$$

i.e. each filled band α contributes $\frac{e^2}{h} C_\alpha$ to the total Hall conductivity whenever the Fermi level at $T = 0$ lies in a gap between energy bands. For a crystalline (periodic) system, averaging over $\theta_{1,2}$ is tantamount to integrating over the Brillouin zone.

4.5 Semiclassical Dynamics of Bloch Electrons

Consider a time-evolving quantum mechanical state $|\psi(t)\rangle$. The time dependence of the expectation value $\mathcal{O}(t) = \langle \psi(t) | \mathcal{O} | \psi(t) \rangle$ satisfies

$$\frac{d\mathcal{O}}{dt} = \frac{i}{\hbar} \left\langle \, \psi(t) \,\middle|\, [H, \mathcal{O}] \,\middle|\, \psi(t) \, \right\rangle \quad . \tag{4.245}$$

Thus for $H = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r})$ we have $\frac{d}{dt} \langle \boldsymbol{r} \rangle = \langle \frac{\boldsymbol{p}}{m} \rangle$ and $\frac{d}{dt} \langle \boldsymbol{p} \rangle = -\langle \boldsymbol{\nabla} V \rangle$, a result known as Ehrenfest's theorem. There are a couple of problems in applying this to electrons in crystals, though. One is that the momentum $\boldsymbol{p}$ in a Bloch state is defined only modulo $\hbar \boldsymbol{G}$, where $\boldsymbol{G}$ is any reciprocal lattice vector. Another is that the potential $\Delta V(\boldsymbol{r}) = e\boldsymbol{E} \cdot \boldsymbol{r}$ breaks the lattice periodicity present in the crystal.

4.5.1 Adiabatic evolution

Here we assume $d = 3$ dimensions. Recall $\boldsymbol{E} = -\boldsymbol{\nabla}\phi - c^{-1}\partial_t \boldsymbol{A}$, so rather that taking $\boldsymbol{A} = 0$ and $\phi = -\boldsymbol{E} \cdot \boldsymbol{r}$ in the case of a uniform electric field, we can instead take $\boldsymbol{A}(t) = -c\boldsymbol{E}t$ and $\phi = 0$ and write

$$H(t) = \frac{\left(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(t)\right)^2}{2m} + V(\boldsymbol{r}) \quad , \tag{4.246}$$

with $\partial_t \boldsymbol{A} = -c\boldsymbol{E}$, and $\boldsymbol{\nabla} \times \boldsymbol{A} = \boldsymbol{B}$ if there is a magnetic field as well. We assume that the electric field is very weak, which means that we can treat the time dependent Hamiltonian $H(t)$ in the adiabatic limit[40].

We begin by reiterating some key results from §4.4.4. Consider a setting in which a Hamiltonian $H(\boldsymbol{\lambda})$ depends on a set of parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_K\}$. The adiabatic eigenstates of $H(\boldsymbol{\lambda})$ are denoted as $|n(\boldsymbol{\lambda})\rangle$, where $H(\boldsymbol{\lambda}) \, |n(\boldsymbol{\lambda})\rangle = E_n(\boldsymbol{\lambda}) \, |n(\boldsymbol{\lambda})\rangle$. Now suppose that $\lambda(t)$ varies with time.

[40] Technically, we should require there be a finite energy gap in order to justify adiabatic evolution.

Since the set $\big\{|n(\boldsymbol{\lambda})\rangle\big\}$ is complete, we may expand the wavefunction $|\psi(t)\rangle$ in the adiabatic basis, *viz.*

$$|\,\psi(t)\,\rangle = \sum_n a_n(t)\,|\,n\big(\boldsymbol{\lambda}(t)\big)\,\rangle \quad . \tag{4.247}$$

Now we impose the condition $i\hbar\,\partial_t\,|\psi(t)\rangle = H\big(\boldsymbol{\lambda}(t)\big)|\psi(t)\rangle$. We first define the phases $\phi_n(t)$ and $\gamma_n(t)$, where

$$\phi_n(t) = -\frac{1}{\hbar}\int^t\!dt'\;E_n\big(\boldsymbol{\lambda}(t')\big) \tag{4.248}$$

and where $\gamma_n(t)$ satisfies

$$\frac{d\gamma_n}{dt} = i\,\langle\,n\big(\boldsymbol{\lambda}(t)\big)\,|\,\frac{d}{dt}\,|\,n\big(\boldsymbol{\lambda}(t)\big)\,\rangle \quad . \tag{4.249}$$

Then, writing $a_n(t) \equiv\; e^{i\phi_n(t)}\,e^{i\gamma_n(t)}\,\alpha_n(t)$, we find

$$\frac{d\alpha_n}{dt} = -{\sum_l}'\,e^{i(\gamma_l-\gamma_n)}\,e^{i(\phi_l-\phi_n)}\,\alpha_l \quad , \tag{4.250}$$

where the prime on the sum indicates that the term $l = n$ is to be excluded. Now consider initial conditions where $a_l(0) = \delta_{ln}$. Since the evolution is adiabatic, the phases $\phi_l(t)$ are the fastest evolving quantities, with $\partial_t\,\phi_l = -E_l/\hbar = \mathcal{O}(1)$, as opposed to γ_l and α_l, which vary on the slow time scale associated with the evolution of $\boldsymbol{\lambda}(t)$. This allows us to approximately integrate the above equations to obtain

$$\alpha_n(t) \approx \alpha_n(0) = 1 \qquad , \qquad \alpha_l(t) \approx -i\hbar\,\frac{\langle\,l\,|\,\partial_t\,|\,n\,\rangle}{E_l - E_n} \quad . \tag{4.251}$$

Thus,

$$|\,\psi(t)\,\rangle \approx e^{i\phi_n(t)}\,e^{i\gamma_n(t)}\left\{|\,n(t)\,\rangle - i\hbar\,{\sum_l}'\frac{|\,l(t)\,\rangle\langle\,l(t)\,|\,d_t\,|\,n(t)\,\rangle}{E_l(t) - E_n(t)}\right\} \quad , \tag{4.252}$$

where each $|\,l(t)\,\rangle = |\,l\big(\boldsymbol{\lambda}(t)\big)\,\rangle$, and where $d_t = d/dt$ is the total time derivative. Note that we can write

$$\frac{d\,|n\rangle}{dt} = \frac{\partial\,|n\rangle}{\partial\lambda_\mu}\cdot\frac{d\lambda_\mu}{dt} \quad , \tag{4.253}$$

with an implied sum on μ, and with $i\,\langle n|\,d_t\,|n\rangle = \mathcal{A}_n^\mu\,\dot\lambda_\mu$, where

$$\mathcal{A}_n^\mu(\boldsymbol{\lambda}) \equiv i\,\big\langle\,n(\boldsymbol{\lambda})\,\big|\,\frac{\partial}{\partial\lambda_\mu}\,\big|\,n(\boldsymbol{\lambda})\,\big\rangle \tag{4.254}$$

is the *geometric connection* (or *Berry connection*) for the state $|\,n(\boldsymbol{\lambda})\,\rangle$. Note that the Berry connection is gauge-dependent, in that redefining $|\,\widetilde{n}(\boldsymbol{\lambda})\,\rangle \equiv e^{if_n(\boldsymbol{\lambda})}\,|\,n(\boldsymbol{\lambda})\,\rangle$ results in

$$\widetilde{\mathcal{A}}_n^\mu(\boldsymbol{\lambda}) = i\,\big\langle\,\widetilde{n}(\boldsymbol{\lambda})\,\big|\,\frac{\partial}{\partial\lambda_\mu}\,\big|\,\widetilde{n}(\boldsymbol{\lambda})\,\big\rangle = \mathcal{A}_n^\mu(\boldsymbol{\lambda}) - \frac{\partial f_n}{\partial\lambda_\mu} \quad . \tag{4.255}$$

If we require that the adiabatic wavefunctions be single-valued as a function of $\boldsymbol{\lambda}$, then *the integral of the Berry connection around a closed path is a gauge-invariant quantity*,

$$\gamma_n(\mathcal{C}) \equiv \oint\limits_{\mathcal{C}} d\lambda_\mu \, \mathcal{A}_n^\mu(\boldsymbol{\lambda}) \quad , \tag{4.256}$$

since $f_n(\boldsymbol{\lambda})$ can wind only by $2\pi k$, with $k \in \mathbb{Z}$, around a closed loop $\mathcal{C}$ in parameter space. If $\mathcal{C}$ is contractable to a point, then $k = 0$.

Now consider the cell function $|u_{n\boldsymbol{k}}\rangle$ as a function of the Bloch wavevector $\boldsymbol{k}$ for each band index n. The velocity operator is $\boldsymbol{v}(\boldsymbol{k}) = \hbar^{-1}\,\partial H(\boldsymbol{k})/\partial \boldsymbol{k}$, where $H(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} H e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$. We then have[41]

$$\frac{d\boldsymbol{r}}{dt} = \langle\, \psi(t) \,|\, \boldsymbol{v}(\boldsymbol{k}) \,|\, \psi(t) \,\rangle = \frac{1}{\hbar}\frac{\partial E_n(\boldsymbol{k})}{\partial \boldsymbol{k}} - \frac{d\boldsymbol{k}}{dt} \times \boldsymbol{\Omega}_n(\boldsymbol{k}) \tag{4.257}$$

where

$$\mathcal{A}_n^\mu(\boldsymbol{k}) = i \left\langle\, u_{n\boldsymbol{k}} \,\middle|\, \frac{\partial}{\partial k^\mu} \,\middle|\, u_{n\boldsymbol{k}} \,\right\rangle \quad , \qquad \Omega_n^\mu(\boldsymbol{k}) = \epsilon_{\mu\nu\lambda} \frac{\partial \mathcal{A}_n^\lambda(\boldsymbol{k})}{\partial k^\nu} \quad . \tag{4.258}$$

In vector notation, $\boldsymbol{\mathcal{A}}_n(\boldsymbol{k}) = i \langle\, u_{n\boldsymbol{k}} \,|\, \boldsymbol{\nabla}_{\boldsymbol{k}} \,|\, u_{n\boldsymbol{k}} \,\rangle$ and $\boldsymbol{\Omega}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}} \times \boldsymbol{\mathcal{A}}_n(\boldsymbol{k})$, where $\boldsymbol{\nabla}_{\boldsymbol{k}} = \frac{\partial}{\partial \boldsymbol{k}}$. Eqn. 4.257 is the first of our semiclassical equations of motion for an electron wavepacket in a crystal. The quantity $\boldsymbol{\Omega}_n(\boldsymbol{k})$, which has dimensions of area, is called the *Berry curvature* of the band $|u_{n\boldsymbol{k}}\rangle$. The second term in Eqn. 4.257 is incorrectly omitted in many standard solid state physics texts! When the orbital moment of the Bloch electrons is included, we must substitute[42]

$$E_n(\boldsymbol{k}) \to E_n(\boldsymbol{k}) - \boldsymbol{M}_n(\boldsymbol{k}) \cdot \boldsymbol{B}(\boldsymbol{r}, t) \quad , \tag{4.259}$$

where

$$M_n^\mu(\boldsymbol{k}) = e\,\epsilon_{\mu\nu\lambda}\,\mathrm{Im} \left\langle \frac{\partial u_{n\boldsymbol{k}}}{\partial k^\nu} \,\middle|\, \Big(E_n(\boldsymbol{k}) - H(\boldsymbol{k}) \Big) \,\middle|\, \frac{\partial u_{n\boldsymbol{k}}}{\partial k^\lambda} \right\rangle \quad , \tag{4.260}$$

and $H(\boldsymbol{k}) = \frac{(\boldsymbol{p}+\hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r})$ as before.

The second equation of semiclassical motion is for $d\boldsymbol{k}/dt$. This is the familiar equation derived from Newton's second law[43],

$$\frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar}\boldsymbol{E} - \frac{e}{\hbar c}\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B} - \frac{e}{2\hbar mc}\boldsymbol{\nabla}(\boldsymbol{\sigma}\cdot\boldsymbol{B}) \quad , \tag{4.261}$$

where we include the contribution from the Zeeman Hamiltonian $H_{\rm Z} = (e\hbar/2mc)\,\boldsymbol{\sigma}\cdot\boldsymbol{B}$. If we choose $\hat{\boldsymbol{z}}$ as the spin quantization axis, so $H_{\rm Z} = (e\hbar/2mc)\,\sigma B^z$, then we can combine the spin orbit force with that of the electric field and write

$$\frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar}\boldsymbol{E}_\sigma - \frac{e}{\hbar c}\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B} \quad , \tag{4.262}$$

where

$$\boldsymbol{E}_\sigma = \boldsymbol{E} + \frac{\sigma}{2mc}\boldsymbol{\nabla}B^z \quad . \tag{4.263}$$

[41] See §4.8.

[42] See G. Sundaram and Q. Niu, *Phys. Rev. B* **59**, 14915 (1999); also D. Xiao, M. Chang, and Q. Niu, *Rev. Mod. Phys.* **82**, 1959 (2010).

[43] Some subtleties in the derivation are discussed in A. Manohar, *Phys. Rev. B* **34**, 1287 (1986).

4.5.2 Violation of Liouville's theorem and its resolution

Our equations of motion for a wavepacket are thus

$$\begin{aligned}\dot{x}^\alpha + \epsilon_{\alpha\beta\gamma}\,\dot{k}^\beta\,\Omega_n^\gamma &= v_n^\alpha \\ \dot{k}^\alpha + \frac{e}{\hbar c}\,\epsilon_{\alpha\beta\gamma}\,\dot{x}^\beta B^\gamma &= -\frac{e}{\hbar}\,E^\alpha \quad ,\end{aligned} \tag{4.264}$$

where $\boldsymbol{v}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}}\,E_n(\boldsymbol{k})/\hbar$. These equations may be recast as

$$\begin{pmatrix} \delta_{\alpha\beta} & \epsilon_{\alpha\beta\gamma}\,\Omega_n^\gamma \\ \frac{e}{\hbar c}\,\epsilon_{\alpha\beta\gamma}\,B^\gamma & \delta_{\alpha\beta} \end{pmatrix} \begin{pmatrix} \dot{x}^\alpha \\ \dot{k}^\alpha \end{pmatrix} = \begin{pmatrix} v_n^\alpha \\ -\frac{e}{\hbar}\,E^\alpha \end{pmatrix} \quad . \tag{4.265}$$

Inverting, we find

$$\begin{aligned}\dot{x}^\alpha &= \big(1 + \tfrac{e}{\hbar c}\boldsymbol{B}\cdot\boldsymbol{\Omega}_n\big)^{-1}\Big\{v_n^\alpha + \tfrac{e}{\hbar c}\,(\boldsymbol{v}_n\cdot\boldsymbol{\Omega}_n)\,B^\alpha + \tfrac{e}{\hbar}\,\epsilon_{\alpha\beta\gamma}\,E^\beta\,\Omega_n^\gamma\Big\} \\ \dot{k}^\alpha &= -\tfrac{e}{\hbar}\,\big(1 + \tfrac{e}{\hbar c}\boldsymbol{B}\cdot\boldsymbol{\Omega}_n\big)^{-1}\Big\{E^\alpha + \tfrac{e}{\hbar c}\,(\boldsymbol{E}\cdot\boldsymbol{B})\,\Omega_n^\alpha + \tfrac{1}{c}\,\epsilon_{\alpha\beta\gamma}\,v_n^\beta\,B^\gamma\Big\} \quad .\end{aligned} \tag{4.266}$$

It is straightforward to derive the result

$$\frac{\partial\dot{x}^\alpha}{\partial x^\alpha} + \frac{\partial\dot{k}^\alpha}{\partial k^\alpha} = -\frac{\partial\ln D_n}{\partial x^\alpha}\,\frac{dx^\alpha}{dt} - \frac{\partial\ln D_n}{\partial k^\alpha}\,\frac{dk^\alpha}{dt} - \frac{\partial\ln D_n}{\partial t} = -\frac{d\ln D_n}{dt} \quad , \tag{4.267}$$

where

$$D_n(\boldsymbol{r},\boldsymbol{k},t) = 1 + \frac{e}{\hbar c}\,\boldsymbol{B}(\boldsymbol{r},t)\cdot\boldsymbol{\Omega}_n(\boldsymbol{k}) \tag{4.268}$$

is dimensionless. As discussed by Xiao, Shi, and Niu[44], this implies a violation of Liouville's theorem, as phase space volumes will then expand according to

$$\frac{d\ln\Delta V}{dt} = \boldsymbol{\nabla}_{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}} + \boldsymbol{\nabla}_{\boldsymbol{k}}\cdot\dot{\boldsymbol{k}} = -d\ln D_n(\boldsymbol{r},\boldsymbol{k},t)/dt \quad , \tag{4.269}$$

where $\Delta V = \Delta\boldsymbol{r}\,\Delta\boldsymbol{k}$ is a phase space volume element. Thus, $\Delta V(t) = \Delta V(0)/D_n(\boldsymbol{r},\boldsymbol{k},t)$, and this inconvenience can be eliminated by redefining the phase space metric as

$$d\mu = \frac{d^3r\,d^3k}{(2\pi)^3} \quad\longrightarrow\quad d\tilde{\mu} \equiv D_n(\boldsymbol{r},\boldsymbol{k},t)\,\frac{d^3r\,d^3k}{(2\pi)^3} \quad . \tag{4.270}$$

This means that the expectation of any local observable $\mathcal{O}$ is given by

$$\begin{aligned}\langle\mathcal{O}\rangle(\boldsymbol{r}',t) &= \sum_n \int\limits_{\hat{\Omega}} d\tilde{\mu}\,f_n(\boldsymbol{r},\boldsymbol{k},t)\,\langle\,u_{n\boldsymbol{k}}\,|\,\mathcal{O}\,|\,u_{n\boldsymbol{k}}\,\rangle\,\delta(\boldsymbol{r}-\boldsymbol{r}') \\ &= \sum_n \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\,D_n(\boldsymbol{r},'\boldsymbol{k},t)\,f_n(\boldsymbol{r}',\boldsymbol{k},t)\,\langle\,u_{n\boldsymbol{k}}\,|\,\mathcal{O}\,|\,u_{n\boldsymbol{k}}\,\rangle \quad ,\end{aligned} \tag{4.271}$$

[44] See D. Xiao, J. Shi, and Q. Niu, *Phys. Rev. Lett.*. **95**, 137204 (2005).

where $f_n(\boldsymbol{r},\boldsymbol{k},t)$ is the mean occupation number of the state $|\,n\boldsymbol{k}\,\rangle$ in the region spatially centered at $\boldsymbol{r}$, and where we have absorbed the spin polarization label σ into the band label n.[45] In equilibrium, $f_n(\boldsymbol{k})$ is the Fermi function $f\big(E_n(\boldsymbol{k})-\mu\big)$. The electrical current density carried by a given band n is then

$$\begin{aligned} j_n^\alpha(\boldsymbol{r},t) &= \int\limits_{\hat\Omega}\frac{d^dk}{(2\pi)^d}\,D_n(\boldsymbol{r},\boldsymbol{k},t)\,f_n(\boldsymbol{k})\,(-e\dot x) \\ &= -e\int\limits_{\hat\Omega}\frac{d^3k}{(2\pi)^3}\left\{v_n^\alpha+\frac{e}{\hbar c}\,(\boldsymbol{v}_n\cdot\boldsymbol{\Omega}_n)\,B^\alpha+\frac{e}{\hbar}\,\epsilon_{\alpha\beta\gamma}\,E^\beta\,\Omega_n^\gamma\right\}f_n(\boldsymbol{k}) \quad . \end{aligned} \tag{4.272}$$

Note the cancellation of the $D_n(\boldsymbol{k})$ factors in $d\mu$ and $\dot x$. Consider the case of a filled band, with $\boldsymbol{B}=0$. The total current density is then

$$\boldsymbol{j}_n = -\frac{e^2}{\hbar}\,\boldsymbol{E}\times\int\limits_{\hat\Omega}\frac{d^3k}{(2\pi)^3}\,\boldsymbol{\Omega}_n(\boldsymbol{k}) \quad . \tag{4.273}$$

Thus, *when the geometric curvature $\boldsymbol{\Omega}_n(\boldsymbol{k})$ is nonzero, a filled band may carry current*.

4.5.3 Bloch oscillations

Let's consider the simplest context for our semiclassical equations of motion: $d=1$ dimension, which means $\boldsymbol{B}=0$. We'll take a nearest neighbor s orbital hopping Hamiltonian, whose sole tight binding band has the dispersion[46] $E(k)=-2\beta\cos(ka)$. The semiclassical equations of motion are

$$\dot x=\frac{1}{\hbar}\frac{\partial E}{\partial k}=\frac{2\beta a}{\hbar}\,\sin(ka) \qquad , \qquad \dot k=-\frac{e}{\hbar}\,E \quad . \tag{4.274}$$

The second of these equations is easily integrated for constant E:

$$k(t)=k(0)-\frac{e}{\hbar}Et \quad , \tag{4.275}$$

in which case

$$\dot x=\frac{2\beta a}{\hbar}\,\sin\Big(k(0)\,a-\frac{eaEt}{\hbar}\Big) \qquad\Rightarrow\qquad x(t)=x(0)+\frac{2\beta}{eE}\left[\cos\Big(k(0)\,a-\frac{eaEt}{\hbar}\Big)-\cos\big(k(0)\,a\big)\right] \quad . \tag{4.276}$$

Note that $x(t)$ oscillates in time! This is quite unlike the free electron case, where we have $m\ddot x=-eE$, yielding ballistic motion $x(t)=x(0)+\dot x(0)\,t-\frac{eE}{2m}\,t^2$, *i.e.* uniform acceleration $(-eE/m)$. This remarkable behavior is called a *Bloch oscillation*.

[45] This allows for intrinsic spin structure in the cell wavefunctions, which is the case when spin-orbit terms are present.

[46] We write the hopping integral as β so as to avoid any confusion with the time variable, t.

The period of the Bloch oscillations is $\tau_{\rm B} = \hbar/eaE$. Let's estimate $\tau_{\rm B}$, taking $E = 1\,{\rm V/cm}$ and $a = 3\,$Å. We find $\tau_{\rm B} = 1.4 \times 10^{-7}\,{\rm s}$, which is much larger than typical scattering times due to phonons or lattice impurities. For example, the thermal de Broglie lifetime is $\hbar/k_{\rm B}T = 2.5 \times 10^{-14}\,{\rm s}$ at $T = 300\,{\rm K}$. Thus, the wavepacket never makes it across the Brillouin zone - not even close. In the next chapter, we will see how to model charge transport in metals.

4.6 *Ab initio* Calculations of Electronic Structure

4.6.1 Orthogonalized plane waves

The plane wave expansion of Bloch states in Eqn. 4.10,

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})\, e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}} \tag{4.277}$$

is formally correct, but in practice difficult to implement. The main reason is that one must keep a large number of coefficients $C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})$ in order to get satisfactory results, because the interesting valence or conduction band Bloch functions must be orthogonal to the core Bloch states derived from the atomic 1s, 2s, *etc.* levels. If the core electrons are localized within a volume $v_{\rm c}$, Heisenberg tells us that the spread in wavevector needed to describe such states is given by $\Delta k_x\,\Delta k_y\,\Delta k_z \gtrsim v_{\rm c}^{-1}$. In $d = 3$, the number of plane waves we need to describe a Bloch state of crystal momentum $\hbar\boldsymbol{k}$ is then

$$N_{\rm pw} \approx \frac{\Delta k_x\,\Delta k_y\,\Delta k_z}{\hat{v}_0} = \frac{1}{8\pi^3}\cdot\frac{v_0}{v_{\rm c}} \quad , \tag{4.278}$$

where $\hat{v}_0 = {\rm vol}(\hat{\Omega})$ is the volume of the first Brillouin zone (with dimensions $[\hat{v}_0] = L^{-d}$). If the core volume $v_{\rm c}$ is much smaller than the Wigner-Seitz cell volume v_0, this means we must retain a large number of coefficients in the expansion of Eqn. 4.277.

Suppose, however, an eccentric theorist gives you a good approximation to these core Bloch states. Indeed, according to Eqn. 4.58, if we define $C_n(\boldsymbol{k}) \equiv S^{1/2}_{nn'}(\boldsymbol{k})\, D_{n'}(\boldsymbol{k})$, then the coefficients $D_n(\boldsymbol{k})$ satisfy the eigenvalue equation

$$S^{-1/2}(\boldsymbol{k})\, H(\boldsymbol{k})\, S^{-1/2}(\boldsymbol{k}) \boldsymbol{D}_n(\boldsymbol{k}) = E_n(\boldsymbol{k})\, \boldsymbol{D}_n(\boldsymbol{k}) \quad , \tag{4.279}$$

Thus, for each $\boldsymbol{k}$ the eigenvectors $\big\{\boldsymbol{D}_a(\boldsymbol{k})\big\}$ of the Hermitian matrix $\widetilde{H}(\boldsymbol{k}) \equiv S^{-1/2}(\boldsymbol{k})\, H(\boldsymbol{k})\, S^{-1/2}(\boldsymbol{k})$ yield, upon multiplication by $S^{1/2}(\boldsymbol{k})$, the coefficients $C_n(\boldsymbol{k})$. Here we imagine that the indices n and n' are restricted to the core levels alone. This obviates the subtle problem of overcompleteness of the atomic levels arising from the existence of scattering states. Furthermore, we

may write

$$S_{nn'}(\boldsymbol{k}) = \delta_{nn'} + \Sigma_{nn'}(\boldsymbol{k})$$
$$\Sigma_{nn'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}\neq 0} S_{nn'}(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \tag{4.280}$$

Now since we are talking about core levels, the contribution $\Sigma_{nn'}(\boldsymbol{k})$, which involves overlaps on different sites, is very small. This means the inverse square root, $S^{-1/2} \approx 1 - \frac{1}{2}\Sigma + \mathcal{O}(\Sigma^2)$, can be well-approximated by the first two terms in the expansion in powers of Σ.

We will write $|\,\mathsf{a}\boldsymbol{k}\,\rangle$ for a core Bloch state in band a, and $|\,\mathsf{a}\boldsymbol{R}\,\rangle$ for the corresponding core Wannier state[47]. Now define the projector,

$$\Pi = \sum_{\mathsf{a},\boldsymbol{k}} |\,\mathsf{a}\boldsymbol{k}\,\rangle\langle\,\mathsf{a}\boldsymbol{k}\,| = \sum_{\mathsf{a},\boldsymbol{R}} |\,\mathsf{a}\boldsymbol{R}\,\rangle\langle\,\mathsf{a}\boldsymbol{R}\,| \quad . \tag{4.281}$$

Note that $\Pi^2 = \Pi$, which is a property of projection operators. We now define

$$|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle \equiv \big(1-\Pi\big)|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle \quad , \tag{4.282}$$

where $|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle$ is the plane wave, for which $\langle\,\boldsymbol{r}\,|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle = V^{-1/2}\,e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}$. It is important to note that we continue to restrict $\boldsymbol{k} \in \widehat{\mathcal{L}}$ to the first Brillouin zone. Note that $\Pi\,|\phi_{\boldsymbol{G}+\boldsymbol{k}}\rangle = 0$, *i.e.* the state $|\phi_{\boldsymbol{G}+\boldsymbol{k}}\rangle$ has been *orthogonalized* to all the core orbitals. Accordingly, we call

$$\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r}) = \frac{e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}}{\sqrt{V}}\,\left\{1 - \sum_{\mathsf{a}} u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}\int d^d r'\,u^*_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r}')\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}'}\right\} \tag{4.283}$$

an *orthogonalized plane wave* (OPW).

As an example, consider the case of only a single core 1s orbital, whose atomic wavefunction is given by the hydrogenic form $\varphi(\boldsymbol{r}) = \frac{\alpha^{3/2}}{\sqrt{\pi}}\,e^{-\alpha r}$. The core cell function is then approximately

$$u_{\boldsymbol{k}}(\boldsymbol{r}) \approx \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} \varphi(\boldsymbol{r}-\boldsymbol{R})\,e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})} \quad . \tag{4.284}$$

We then have

$$\int d^3r'\,u^*_{\boldsymbol{k}}(\boldsymbol{r}')\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}'} \approx \sqrt{N}\,\big[\widehat{\varphi}(\boldsymbol{G}+\boldsymbol{k})\big]^* = \frac{8\pi^{1/2}\alpha^{5/2}N^{1/2}}{\big[\alpha^2+(\boldsymbol{G}+\boldsymbol{k})^2\big]^2} \tag{4.285}$$

and then

$$\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r}) = \frac{e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}}{\sqrt{V}}\,\left\{1 - \frac{8\,\alpha^4}{\big[\alpha^2+(\boldsymbol{G}+\boldsymbol{k})^2\big]^2}\sum_{\boldsymbol{R}} e^{-\alpha|\boldsymbol{r}-\boldsymbol{R}|}\,e^{-i(\boldsymbol{G}+\boldsymbol{k})\cdot(\boldsymbol{r}-\boldsymbol{R})}\right\} \quad . \tag{4.286}$$

[47] Recall that the Wannier states in a given band are somewhat arbitrary as they depend on a choice of phase.

For $\boldsymbol{G}+\boldsymbol{k}=0$, we have, in the $\boldsymbol{R}=0$ cell,

$$\phi_0(\boldsymbol{r}) \approx \frac{1}{\sqrt{V}}\left(1-8\,e^{-\alpha r}\right) \quad . \tag{4.287}$$

Note that the OPW states are not normalized. Indeed, we have[48]

$$\langle\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle = \int d^dr\,\left|\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r})\right|^2 = 1-\frac{1}{V}\sum_{\mathsf{a}}\left|\int d^dr\,u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{G}\cdot\boldsymbol{r}'}\right|^2 \quad . \tag{4.288}$$

The energy eigenvalues are then obtained by solving the equation $\det M_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k},E)=0$ for $E_n(\boldsymbol{k})$, where

$$\begin{aligned} M_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k},E) &= \langle\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,|\,H\,|\,\phi_{\boldsymbol{G}'+\boldsymbol{k}}\,\rangle \qquad (4.289)\\ &= \left[\frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m}-E\right]\delta_{\boldsymbol{G}\boldsymbol{G}'}+V_{\boldsymbol{G}-\boldsymbol{G}'}+\sum_{\mathsf{a}}\left[E-E_{\mathsf{a}}(\boldsymbol{k})\right]\langle\,\boldsymbol{G}+\boldsymbol{k}\,|\,\mathsf{a}\boldsymbol{k}\,\rangle\,\langle\,\mathsf{a}\boldsymbol{k}\,|\,\boldsymbol{G}'+\boldsymbol{k}\,\rangle \quad . \end{aligned}$$

The overlap of the plane wave state $|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle$ and the core Bloch state $|\,\mathsf{a}\boldsymbol{k}\,\rangle$ is given by

$$\langle\,\boldsymbol{G}+\boldsymbol{k}\,|\,\mathsf{a}\boldsymbol{k}\,\rangle = N^{1/2}\,v_0^{-1/2}\int\limits_{\Omega} d^dr\,u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \quad . \tag{4.290}$$

4.6.2 The pseudopotential

Equivalently, we may use the $|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle$ in linear combinations to build our Bloch states, *viz.*

$$|\,\psi_{n\boldsymbol{k}}\,\rangle = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}}^{(n)}(\boldsymbol{k})\,|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle \equiv (1-\Pi)\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle \quad , \tag{4.291}$$

where $|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}}^{(n)}(\boldsymbol{k})\,|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle$. Note that the *pseudo-wavefunction* $\widetilde{\psi}_{n\boldsymbol{k}}(\boldsymbol{r})$ is a sum over plane waves. What equation does it satisfy? From $H|\,\psi_{n\boldsymbol{k}}\,\rangle = E\,|\,\psi_{n\boldsymbol{k}}\,\rangle$, we derive

$$H|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle + \overbrace{(E-H)\,\Pi}^{V_{\rm R}}\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle = E\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle \quad . \tag{4.292}$$

Note that

$$V_{\rm R} = (E-H)\,\Pi = \sum_{\mathsf{a}}\left[E-E_{\mathsf{a}}(\boldsymbol{k})\right]\,|\,\mathsf{a}\boldsymbol{k}\,\rangle\langle\,\mathsf{a}\boldsymbol{k}\,| \tag{4.293}$$

[48] Note that $0 \leqslant \langle\,\phi_{\boldsymbol{k}}\,|\,\phi_{\boldsymbol{k}}\,\rangle \leqslant 1$.

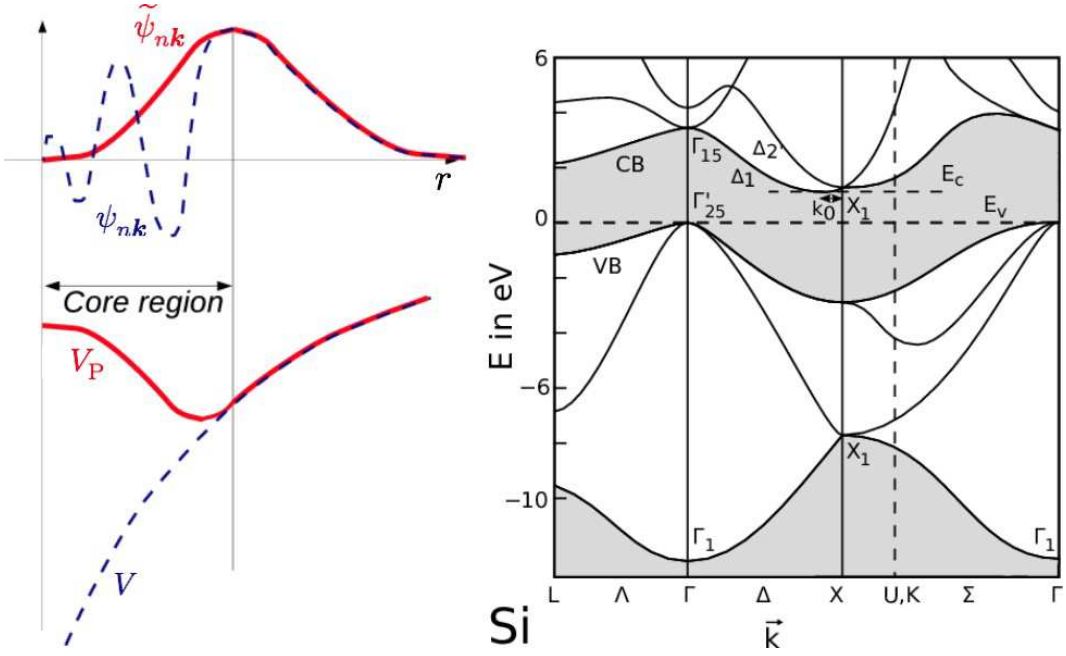

Figure 4.23: Left: Blue curves sketch the potential V in the vicinity of a nucleus, and corresponding valence or conduction band wavefunction $\psi_{n\boldsymbol{k}}$. Note that $\psi_{n\boldsymbol{k}}$ wiggles significantly in the vicinity of the nucleus because it must be orthogonal to the core atomic orbitals. Red curves sketch the pseudopotential $V_{\rm P}$ and corresponding pseudo-wavefunction $\widetilde{\psi}_{n\boldsymbol{k}}$. Right: pseudopotential calculation for the band structure of Si.

and

$$\langle\, \boldsymbol{r} \,|\, V_{\rm R} \,|\, \widetilde{\psi} \,\rangle = \int d^d r'\, V_{\rm R}(\boldsymbol{r},\boldsymbol{r}')\, \widetilde{\psi}(\boldsymbol{r}') \quad , \tag{4.294}$$

where

$$\begin{aligned} V_{\rm R}(\boldsymbol{r},\boldsymbol{r}') &= \sum_{{\rm a},\boldsymbol{k}} \big[E - E_{\rm a}(\boldsymbol{k})\big]\, \psi_{{\rm a}\boldsymbol{k}}(\boldsymbol{r})\, \psi^*_{{\rm a}\boldsymbol{k}}(\boldsymbol{r}') \\ &\approx \sum_{{\rm a},\boldsymbol{R}} (E - E_{\rm a})\, \varphi_{\rm a}(\boldsymbol{r}-\boldsymbol{R})\, \varphi^*_{\rm a}(\boldsymbol{r}'-\boldsymbol{R}) \quad , \end{aligned} \tag{4.295}$$

where $\varphi_{\rm a}(\boldsymbol{r})$ is the a$^{\rm th}$ atomic core wavefunction. The approximation in the second line above is valid in the limit where the core energy bands are dispersionless, and $E_{\rm a}(\boldsymbol{k})$ is replaced by the $\boldsymbol{k}$-independent atomic eigenvalue $E_{\rm a}$. We then use $\sum_{\boldsymbol{k}} |\,{\rm a}\boldsymbol{k}\,\rangle\langle\,{\rm a}\boldsymbol{k}\,| \;=\; \sum_{\boldsymbol{R}} |\,{\rm a}\boldsymbol{R}\,\rangle\langle\,{\rm a}\boldsymbol{R}\,|$, with $\langle\, \boldsymbol{r} \,|\, {\rm a}\boldsymbol{R}\,\rangle \approx \varphi_{\rm a}(\boldsymbol{r}-\boldsymbol{R})$. Because the atomic levels are highly localized, this means $V_{\boldsymbol{R}}(\boldsymbol{r},\boldsymbol{r}')$ is very small unless both $\boldsymbol{r}$ and $\boldsymbol{r}'$ lie within the same core region. Thus, $V_{\rm R}(\boldsymbol{r},\boldsymbol{r}')$ is "almost diagonal" in $\boldsymbol{r}$ and $\boldsymbol{r}'$. Since the energies of interest satisfy $E > E_{\rm a}$, the term $V_{\boldsymbol{R}}$ tends to add to what is a negative (attractive) potential V, and the combination $V_{\rm P} = V + V_{\rm R}$, known as the *pseudopotential*, is in general weaker the original potential. As depicted in the left panel of Fig.

4.23, whereas the actual valence or conduction band Bloch states $\psi_{\rm c/v,\boldsymbol{k}}(\boldsymbol{r})$ must wiggle rapidly in the vicinity of each nucleus, in order to be orthogonal to the atomic core states and thereby necessitating the contribution of a large number of high wavevector plane wave components, each pseudo-wavefunction $\widetilde{\psi}_{n\boldsymbol{k}}(\boldsymbol{r})$ is unremarkable in the core region, and can be described using far less information.

In fact, there is a great arbitrariness in defining the operator $V_{\rm R}$. Suppose we take $\widetilde{V}_{\rm R} = \Pi\, W$, where W is *any* operator. Note that

$$\widetilde{V}_{\rm R}(\boldsymbol{r},\boldsymbol{r}') = \langle\, \boldsymbol{r}\,|\,\Pi\, W\,|\,\boldsymbol{r}'\,\rangle = \sum_{\mathsf{a},\boldsymbol{k}} \psi_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,\langle\, \mathsf{a}\boldsymbol{k}\,|\,W\,|\,\boldsymbol{r}'\,\rangle \equiv \sum_{\mathsf{a},\boldsymbol{k}} \psi_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\, W^*_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r}') \quad . \tag{4.296}$$

This needn't even be Hermitian! The point is that $H = T + V$ and $\widetilde{H} = T + V + \widetilde{V}_{\rm R}$ have the same eigenvalues so long as they are acting outside the space of core wavefunctions. To see this, let us suppose

$$H\,|\,\psi\,\rangle = E\,|\,\psi\,\rangle \qquad , \qquad \widetilde{H}\,|\,\xi\,\rangle = \widetilde{E}\,|\,\xi\,\rangle \quad . \tag{4.297}$$

Then

$$\begin{aligned}\widetilde{E}\,\langle\,\psi\,|\,\xi\,\rangle = \langle\,\psi\,|\,\widetilde{H}\,|\,\xi\,\rangle &= \langle\,\psi\,|\,(H + \Pi\, W)\,|\,\xi\,\rangle \\ &= E\,\langle\,\psi\,|\,\xi\,\rangle + \langle\,\psi\,|\,\Pi\, W\,|\,\xi\,\rangle \quad .\end{aligned} \tag{4.298}$$

Now if $|\,\psi\,\rangle$ lies in the complement of that part of the Hilbert space spanned by the core states, then $\Pi\,|\,\psi\,\rangle = 0$, and it follows that $(E - \widetilde{E})\,\langle\,\psi\,|\,\xi\,\rangle = 0$, hence $E = \widetilde{E}$, so long as $\langle\,\psi\,|\,\xi\,\rangle \neq 0$. If we want $V_{\rm P}$ to be Hermitian, a natural choice might be $\widetilde{V}_{\rm R} = -\Pi\, V\, \Pi$, which gives $V_{\rm P} = V - \Pi\, V\, \Pi$. This effectively removes from the potential V any component which can be constructed from core states alone.

4.7 Appendix I : Gauss-Bonnet and Pontrjagin

4.7.1 Gauss-Bonnet theorem

There is a deep result in mathematics, the Gauss-Bonnet theorem, which connects the *local geometry* of a two-dimensional manifold to its *global topology*. The content of the theorem is as follows:

$$\int\limits_{\mathcal{M}} dS\; K = 2\pi\,\chi(\mathcal{M}) = 2\pi \sum_i \mathsf{ind}\,(\boldsymbol{V})\;, \tag{4.299}$$

where $\mathcal{M}$ is a 2-manifold (a topological space locally homeomorphic to $\mathbb{R}^2$), K is the local *Gaussian curvature* of $\mathcal{M}$, given by $K = (R_1\,R_2)^{-1}$, where $R_{1,2}$ are the principal radii of curvature at a given point, and dS is the differential area element. Here $\boldsymbol{V}(\boldsymbol{x})$ is a vector field on $\mathcal{M}$, and

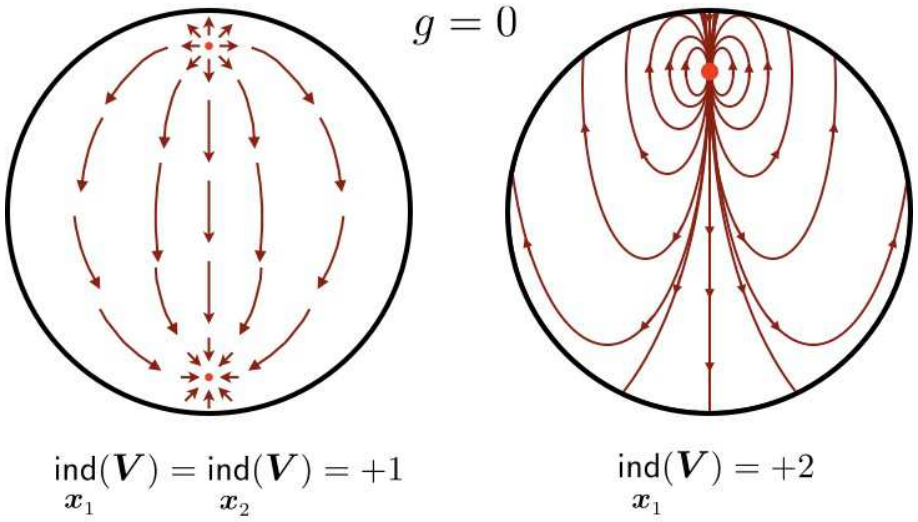

Figure 4.24: Two smooth vector fields on the sphere $\mathbb{S}^2$, which has genus $g = 0$. Left panel: two index $+1$ singularities. Right panel: one index $+2$ singularity.

$\mathsf{ind}_{\boldsymbol{x}_i}(\boldsymbol{V})$ refers to the *index* of $\boldsymbol{V}$ at its $i^{\rm th}$ singularity $\boldsymbol{x}_i$. The index is in general defined relative to any closed curve in $\mathcal{M}$, and is given by the winding number of $\boldsymbol{V}(\boldsymbol{x})$ around the curve, *viz.*

$$\mathop{\mathsf{ind}}_{\mathcal{C}}(\boldsymbol{V}) = \oint\limits_{\mathcal{C}} d\boldsymbol{x} \cdot \boldsymbol{\nabla} \tan^{-1}\left(\frac{V_2(\boldsymbol{x})}{V_1(\boldsymbol{x})}\right) \quad . \tag{4.300}$$

If $\mathcal{C}$ encloses no singularities, then the index necessarily vanishes, but if $\mathcal{C}$ encloses one or more singularities, the index is an integer, given by the winding number of $\boldsymbol{V}$ around the curve $\mathcal{C}$.

The quantity $\chi(\mathcal{M})$ is called the *Euler characteristic* of $\mathcal{M}$ and is given by $\chi(\mathcal{M}) = 2 - 2g$, where g is the *genus* of $\mathcal{M}$, which is the number of holes (or handles) of $\mathcal{M}$. Furthermore, $\boldsymbol{V}(\boldsymbol{x})$ can be *any* smooth vector field on $\mathcal{M}$, with $\boldsymbol{x}_i$ the singularity points of that vector field[49].

To apprehend the content of the Gauss-Bonnet theorem, it is helpful to consider an example. Let $\mathcal{M} = \mathbb{S}^2$ be the unit 2-sphere, as depicted in fig. 4.24. At any point on the unit 2-sphere, the radii of curvature are degenerate and both equal to $R = 1$, hence $K = 1$. If we integrate the Gaussian curvature over the sphere, we thus get $4\pi = 2\pi\,\chi\big(\mathbb{S}^2\big)$, which says $\chi(\mathbb{S}^2) = 2 - 2g = 2$, which agrees with $g = 0$ for the sphere. Furthermore, the Gauss-Bonnet theorem says that *any* smooth vector field on $\mathbb{S}^2$ *must* have a singularity or singularities, with the total index summed over the singularities equal to $+2$. The vector field sketched in the left panel of fig. 4.24 has two index $+1$ singularities, which could be taken at the north and south poles, but which could be anywhere. Another possibility, depicted in the right panel of fig. 4.24, is that there is a one singularity with index $+2$.

In fig. 4.25 we show examples of manifolds with genii $g = 1$ and $g = 2$. The case $g = 1$ is the

[49]The singularities $\boldsymbol{x}_i$ are fixed points of the dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$.

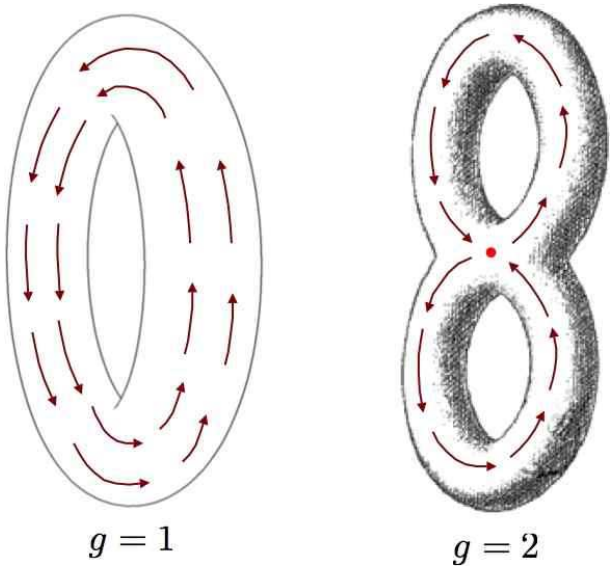

Figure 4.25: Smooth vector fields on the torus $\mathbb{T}^2$, and on a 2-manifold $\mathcal{M}$ of genus $g = 2$

familiar 2-torus, which is topologically equivalent to a product of circles: $\mathbb{T}^2 \cong \mathbb{S}^1 \times \mathbb{S}^1$, and is thus coordinatized by two angles θ_1 and θ_2. A smooth vector field pointing in the direction of increasing θ_1 never vanishes, and thus has no singularities, consistent with $g = 1$ and $\chi\big(\mathbb{T}^2\big) = 0$. Topologically, one can define a torus as the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, or as a square with opposite sides identified. This is what mathematicians call a 'flat torus' – one with curvature $K = 0$ everywhere. Of course, such a torus cannot be embedded in three-dimensional Euclidean space; a two-dimensional figure embedded in a three-dimensional Euclidean space inherits a metric due to the embedding, and for a physical torus, like the surface of a bagel, the Gaussian curvature is only zero *on average*.

The $g = 2$ surface $\mathcal{M}$ shown in the right panel of fig. 4.25 has Euler characteristic $\chi(\mathcal{M}) = -2$, which means that any smooth vector field on $\mathcal{M}$ must have singularities with indices totalling -2. One possibility, depicted in the figure, is to have two saddle points with index -1; one of these singularities is shown in the figure (the other would be on the opposite side).

4.7.2 The Pontrjagin index

Consider an N-dimensional vector field $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$, and let $\hat{\boldsymbol{n}}(\boldsymbol{x})$ be the unit vector field defined by $\hat{\boldsymbol{n}}(\boldsymbol{x}) = \boldsymbol{V}(\boldsymbol{x})/|\boldsymbol{V}(\boldsymbol{x})|$. Consider now a unit sphere in $\hat{\boldsymbol{n}}$ space, which is of dimension $(N-1)$. If we integrate over this surface, we obtain

$$\Omega_N = \oint d\sigma_a \, n^a = \frac{2\pi^{(N-1)/2}}{\Gamma\big(\frac{N-1}{2}\big)} \,, \tag{4.301}$$

which is the surface area of the unit sphere $\mathbb{S}^{N-1}$. Thus, $\Omega_2 = 2\pi$, $\Omega_3 = 4\pi$, $\Omega_4 = 2\pi^2$, *etc.*

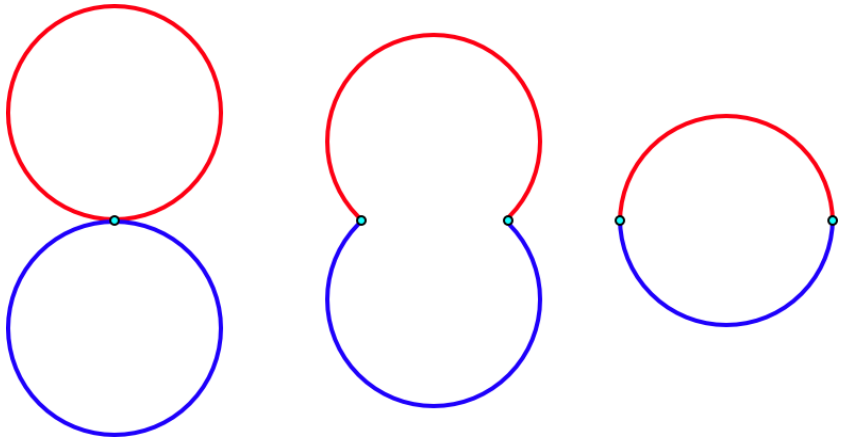

Figure 4.26: Composition of two circles. The same general construction applies to the merging of n-spheres $\mathbb{S}^n$, called the *wedge sum*.

Now consider a change of variables to those over the surface of the sphere, $(\xi_1, \ldots, \xi_{N-1})$. We then have

$$\Omega_N = \oint\limits_{\mathbb{S}^{N-1}} d\sigma_a \, n^a = \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{4.302}$$

The topological charge is then

$$Q = \frac{1}{\Omega_N} \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{4.303}$$

The quantity Q is an *integer topological invariant* which characterizes the map from the surface $(\xi_1, \ldots, \xi_{N-1})$ to the unit sphere $|\hat{\boldsymbol{n}}| = 1$. In mathematical parlance, Q is known as the *Pontrjagin index* of this map.

This analytical development recapitulates some basic topology. Let $\mathcal{M}$ be a topological space and consider a map from the circle $\mathbb{S}^1$ to $\mathcal{M}$. We can compose two such maps by merging the two circles, as shown in fig. 4.26. Two maps are said to be *homotopic* if they can be smoothly deformed into each other. Any two homotopic maps are said to belong to the same *equivalence class* or *homotopy class*. For general $\mathcal{M}$, the homotopy classes may be multiplied using the composition law, resulting in a group structure. The group is called the *fundamental group* of the manifold $\mathcal{M}$, and is abbreviated $\pi_1(\mathcal{M})$. If $\mathcal{M} = \mathbb{S}^2$, then any such map can be smoothly contracted to a point on the 2-sphere, which is to say a trivial map. We then have $\pi_1(\mathcal{M}) = 0$. If $\mathcal{M} = \mathbb{S}^1$, the maps can wind nontrivially, and the homotopy classes are labeled by a single integer winding number: $\pi_1(\mathbb{S}^1) = \mathbb{Z}$. The winding number of the composition of two such maps is the sum of their individual winding numbers. If $\mathcal{M} = \mathbb{T}^2$, the maps can wind nontrivially around either of the two cycles of the 2-torus. We then have $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$, and in general $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$. This makes good sense, since an n-torus is topologically equivalent to a product of n circles. In some cases, $\pi_1(\mathcal{M})$ can be nonabelian, as is the case when $\mathcal{M}$ is the genus $g = 2$ structure shown in the right hand panel of fig. 4.25.

In general we define the $n^{\rm th}$ *homotopy group* $\pi_n(\mathcal{M})$ as the group under composition of maps from $\mathbb{S}^n$ to $\mathcal{M}$. For $n \geqslant 2$, $\pi_n(\mathcal{M})$ is abelian. If $\dim(\mathcal{M}) < n$, then $\pi_n(\mathcal{M}) = 0$. In general, $\pi_n(\mathbb{S}^n) = \mathbb{Z}$. These $n^{\rm th}$ homotopy classes of the n-sphere are labeled by their associated Pontrjagin index Q.

4.8 Appendix II : Derivation of Eqn. 4.257

Consider the Hamiltonian $H(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} H e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, for which $H(\boldsymbol{k})\,|u_{n\boldsymbol{k}}\rangle = E_n(\boldsymbol{k})\,|u_{n\boldsymbol{k}}\rangle$, where $u_{n\boldsymbol{k}}(\boldsymbol{r}) = \langle\,\boldsymbol{r}\,|\,u_{n\boldsymbol{k}}\,\rangle = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ are the cell functions. We consider the wavevector $\boldsymbol{k} = \boldsymbol{k}(t)$ to be an adiabatic parameter which is slowly varying with time. In this setting, the *time-dependent* solutions to the Schrödinger equation,

$$i\hbar\,\frac{\partial}{\partial t}\,|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle = H(\boldsymbol{k})\,|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle \quad , \tag{4.304}$$

are, per Eqn. 4.251, given by

$$|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle = e^{i\phi_{n\boldsymbol{k}}(t)}\,e^{i\gamma_{n\boldsymbol{k}}}\left\{|\,u_{n\boldsymbol{k}}\,\rangle - i\hbar{\sum_{\ell}}'\frac{|\,u_{\ell\boldsymbol{k}}\,\rangle\,\langle\,u_{\ell\boldsymbol{k}}\,|\,d_t\,|\,u_{n\boldsymbol{k}}\,\rangle}{E_\ell(\boldsymbol{k}) - E_n(\boldsymbol{k})}\right\} + \ldots \quad , \tag{4.305}$$

where the prime on the sum indicates that the state $|u_{n\boldsymbol{k}}\rangle$ is excluded. Higher order terms are negligible in the adiabatic limit. Here

$$\phi_{n\boldsymbol{k}}(t) = -\frac{1}{\hbar}\int^t dt'\;E_n\big(\boldsymbol{k}(t')\big) \quad , \tag{4.306}$$

where$\boldsymbol{k} = \boldsymbol{k}(t)$, and

$$\frac{d\gamma_{n\boldsymbol{k}}}{dt} = i\,\langle\,u_{n\boldsymbol{k}}\,|\,\frac{d}{dt}\,|\,u_{n\boldsymbol{k}}\,\rangle = i\,\langle\,u_{n\boldsymbol{k}}\,|\,\frac{\partial}{\partial k^\alpha}\,|\,u_{n\boldsymbol{k}}\,\rangle\,\frac{dk^\alpha}{dt} \quad . \tag{4.307}$$

We now compute the expectation of the velocity operator, $v^\alpha = \hbar^{-1}\,\partial_\alpha H(\boldsymbol{k})$, where we abbreviate $\partial_\alpha \equiv \partial/\partial k^\alpha$:

$$\begin{aligned}\big\langle\,\tilde{u}_{n\boldsymbol{k}}(t)\,\big|\,v^\alpha\,\big|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\big\rangle &= \langle\,u_{n\boldsymbol{k}}\,|\,\hbar^{-1}\,\partial_\alpha H(\boldsymbol{k})\,|\,u_{n\boldsymbol{k}}\,\rangle \\ &\quad - i{\sum_{\ell}}'\frac{\langle\,u_{n\boldsymbol{k}}\,|\,\partial_\alpha H(\boldsymbol{k})\,|\,u_{\ell\boldsymbol{k}}\,\rangle\,\langle\,u_{\ell\boldsymbol{k}}\,|\,d_t\,|\,u_{n\boldsymbol{k}}\,\rangle - \langle\,u_{n\boldsymbol{k}}\,|\,d_t\,|\,u_{\ell\boldsymbol{k}}\,\rangle\,\langle\,u_{\ell\boldsymbol{k}}\,|\,\partial_\alpha H(\boldsymbol{k})\,|\,u_{n\boldsymbol{k}}\,\rangle}{E_\ell(\boldsymbol{k}) - E_n(\boldsymbol{k})} \quad .\end{aligned} \tag{4.308}$$

Now

$$\begin{aligned}\frac{\partial}{\partial k^\alpha}\,\langle\,u_{n\boldsymbol{k}}\,|\,H(\boldsymbol{k})\,|\,u_{\ell\boldsymbol{k}}\,\rangle &= \frac{\partial E_n(\boldsymbol{k})}{\partial k^\alpha}\,\delta_{n\ell} \\ &= E_\ell(\boldsymbol{k})\left\langle\,\frac{\partial u_{n\boldsymbol{k}}}{\partial k^\alpha}\,\middle|\,u_{\ell\boldsymbol{k}}\,\right\rangle + \left\langle\,u_{n\boldsymbol{k}}\,\middle|\,\frac{\partial H(\boldsymbol{k})}{\partial k^\alpha}\,\middle|\,u_{\ell\boldsymbol{k}}\,\right\rangle + E_n(\boldsymbol{k})\left\langle\,u_{n\boldsymbol{k}}\,\middle|\,\frac{\partial u_{\ell\boldsymbol{k}}}{\partial k^\alpha}\,\right\rangle \quad ,\end{aligned} \tag{4.309}$$

which says

$$\langle\, u_{n\boldsymbol{k}} \,|\, \partial_\alpha H(\boldsymbol{k}) \,|\, u_{\ell\boldsymbol{k}} \,\rangle = \partial_\alpha E_n(\boldsymbol{k})\, \delta_{n\ell} + \big(E_n(\boldsymbol{k}) - E_\ell(\boldsymbol{k})\big)\, \langle\, u_{n\boldsymbol{k}} \,|\, \partial_\alpha \,|\, u_{\ell\boldsymbol{k}} \,\rangle \quad . \tag{4.310}$$

Thus, we have

$$\begin{aligned}
\langle\, \tilde{u}_{n\boldsymbol{k}}(t) \,|\, v^\alpha \,|\, \tilde{u}_{n\boldsymbol{k}}(t) \,\rangle &= \hbar^{-1}\, \partial_\alpha E_n(\boldsymbol{k}) + i \sum_\ell{}' \Big(\langle\, u_{n\boldsymbol{k}} \,|\, \partial_\alpha \,|\, u_{\ell\boldsymbol{k}} \,\rangle\, \langle\, u_{\ell\boldsymbol{k}} \,|\, d_t \,|\, u_{n\boldsymbol{k}} \,\rangle - \text{c.c.} \Big) \\
&= \hbar^{-1}\, \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\langle\, \partial_\alpha u_{n\boldsymbol{k}} \,|\, d_t u_{n\boldsymbol{k}} \,\rangle - \langle\, \partial_\alpha u_{n\boldsymbol{k}} \,|\, u_{n\boldsymbol{k}} \,\rangle\, \langle\, u_{n\boldsymbol{k}} \,|\, d_t u_{n\boldsymbol{k}} \,\rangle - \text{c.c.} \Big) \\
&= \hbar^{-1}\, \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\langle\, \partial_\alpha n \,|\, \partial_\beta n \,\rangle\, \dot{k}^\beta - \langle\, \partial_\alpha n \,|\, n \,\rangle\, \langle\, n \,|\, \partial_\beta n \,\rangle\, \dot{k}^\beta - \text{c.c.} \Big) \\
&= \hbar^{-1}\, \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\partial_\alpha \langle\, n \,|\, \partial_\beta n \,\rangle - \partial_\beta \big\langle\, n \,\big|\, \partial_\alpha n \,\big\rangle \Big)\, \dot{k}^\beta \quad ,
\end{aligned} \tag{4.311}$$

where $|\, \partial_\alpha u_{n\boldsymbol{k}} \,\rangle \equiv |\, \partial_\alpha n \,\rangle$, where we ultimately suppress the wavevector label $\boldsymbol{k}$ for notational convenience. Recall now the definitions in Eqn. 4.258,

$$\mathcal{A}_n^\alpha(\boldsymbol{k}) = i \,\Big\langle\, u_{n\boldsymbol{k}} \,\Big|\, \frac{\partial}{\partial k^\alpha} \,\Big|\, u_{n\boldsymbol{k}} \,\Big\rangle \qquad , \qquad \Omega_n^\alpha(\boldsymbol{k}) = \epsilon_{\alpha\beta\gamma}\, \frac{\partial \mathcal{A}_n^\gamma(\boldsymbol{k})}{\partial k^\beta} \quad . \tag{4.312}$$

In our abbreviated notation, $\mathcal{A}_n^\alpha = i\,\langle\, n \,|\, \partial_\alpha\, n \,\rangle$, and we obtain

$$\frac{dx^\alpha}{dt} = \langle\, \psi_{n\boldsymbol{k}}(t) \,|\, v^\alpha \,|\, \psi_{n\boldsymbol{k}}(t) \,\rangle = \hbar^{-1}\, \partial_\alpha E_n(\boldsymbol{k}) - \epsilon_{\alpha\beta\gamma}\, \dot{k}^\beta\, \Omega_n^\gamma(\boldsymbol{k}) \quad , \tag{4.313}$$

since

$$\begin{aligned}
\epsilon_{\alpha\beta\gamma}\, \dot{k}^\beta\, \Omega_n^\gamma &= \epsilon_{\alpha\beta\gamma}\, \dot{k}^\beta\, \epsilon_{\gamma\alpha\beta}\, \partial_\alpha\, \mathcal{A}_n^\beta = \epsilon_{\alpha\beta\gamma}\, \epsilon_{\alpha\beta\beta}\, \dot{k}^\beta\, \partial_\alpha\, \mathcal{A}_n^\beta = \big(\delta_{\alpha\alpha}\, \delta_{\beta\beta} - \delta_{\alpha\beta}\, \delta_{\beta\alpha}\big)\, \dot{k}^\beta\, \partial_\alpha\, \mathcal{A}_n^\beta \\
&= \big(\partial_\alpha \mathcal{A}_n^\beta - \partial_\beta \mathcal{A}_n^\alpha\big)\, \dot{k}^\beta = i \Big(\partial_\alpha \,\langle\, n \,|\, \partial_\beta\, n \,\rangle - \partial_\beta \,\big\langle\, n \,\big|\, \partial_\alpha\, n \,\big\rangle \Big)\, \dot{k}^\beta \quad .
\end{aligned} \tag{4.314}$$