2

Review of electronic structure theory

The basic concepts of the electronic band structure of crystalline materials, notably Bloch's theorem and its consequences, are developed in any standard solid state physics text. These concepts are briefly reviewed in the first part of this chapter, in part to establish the notation to be used later. For the reader interested in modern computational methods for solving for the electronic structure, the book by Martin (2004) is recommended. The second half of this chapter provides an introduction to one particularly simple approach, that of empirical tight-binding theory, that is often capable of giving a useful qualitative picture in terms of bonding between localized orbitals. The concepts to be introduced in the later chapters of this book will often be illustrated via examples that are based on the tight-binding description.

2.1 Electronic Hamiltonian and Bloch functions

The Hamiltonian for a system of electrons in a crystalline solid is known. All the difficulties are associated with the need to simplify the Hamiltonian to a tractable form.

2.1.1 Reduction to a single-particle Hamiltonian

We begin with the Hamiltonian for a finite system such as a molecule or cluster, which can be written as

$$H = T_{\mathrm{e}} + T_{\mathrm{n}} + V_{\mathrm{e\text{-}e}} + V_{\mathrm{n\text{-}e}} + V_{\mathrm{n\text{-}n}} + \ldots \tag{2.1}$$

where T_{e} and T_{n} are electron and nuclear kinetic energies, and $V_{\mathrm{e\text{-}e}}$, $V_{\mathrm{n\text{-}e}}$, and $V_{\mathrm{n\text{-}n}}$ are the electron-electron, electron-nuclear, and nuclear-nuclear Coulomb

interactions. Explicitly,

$$T_{\mathrm{e}} = \sum_i \frac{p_i^2}{2m}\,, \tag{2.2}$$

$$T_{\mathrm{n}} = \sum_I \frac{P_I^2}{2M_I}\,, \tag{2.3}$$

$$V_{\text{e-e}} = \frac{1}{2}\sum_{ij} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \tag{2.4}$$

$$V_{\text{n-e}} = -\sum_{iI} \frac{e^2 Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}\,, \tag{2.5}$$

$$V_{\text{n-n}} = \frac{1}{2}\sum_{IJ} \frac{e^2 Z_I Z_J}{|\mathbf{R}_i - \mathbf{R}_j|}\,, \tag{2.6}$$

where i runs over the electrons in the system with coordinates $\mathbf{r}_i$, momenta $\mathbf{p}_i$, charge $-e$, and mass m, while I runs over the nuclei with coordinates $\mathbf{R}_i$, momenta $\mathbf{P}_i$, charge eZ_I, and mass M_I. Gaussian units are used throughout this chapter. The spin of the electrons has been suppressed for the moment, and the ellipsis '...' in Eq. (2.1) indicates other terms that have been neglected, including relativistic corrections, interactions with external electric and magnetic fields, and interactions involving nuclear magnetic moments.

We first apply the *adiabatic approximation*, which states that the nuclear coordinates can be treated as slow variables because the nuclei are so much heavier than the electrons. Adopting this framework, we now treat the nuclear coordinates $\mathbf{R}_I$ as classical variables that are fixed by hand, or else are evolving slowly according to some classical equations of motion, and concentrate on solving the electronic Hamiltonian

$$H_{\mathrm{elec}} = T_{\mathrm{e}} + V_{\text{n-e}} + V_{\text{e-e}}\,. \tag{2.7}$$

Defining

$$V_{\mathrm{ext}}(\mathbf{r}) = \sum_I \frac{e^2 Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{2.8}$$

to be the bare potential felt by the electrons in the Coulomb field of the nuclei, which are regarded as "external" to the electron system, this can be written as

$$H_{\mathrm{elec}} = \sum_i \left[\frac{p_i^2}{2m} + V_{\mathrm{ext}}(\mathbf{r}_i)\right] + \frac{1}{2}\sum_{ij} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \tag{2.9}$$

which already appeared as Eq. (1.19) aside from some small changes of notation.

Equation (2.9) is famously impossible to solve when the number of electrons N becomes large. The problem does not reside in the kinetic-energy or external-potential terms, which are single-particle operators, but in the Coulomb interaction term, which is a two-particle operator. This means that the many-body electron wavefunction $\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$ cannot be written as a single Slater determinant or otherwise factored into any elementary form, so that the complexity of the wavefunction, and the time to an approximate solution, grow exponentially with N. In the quantum chemistry community, this problem is typically tackled by expanding Ψ as a linear combination of some subset of Slater determinants that are deemed most relevant, with an artful truncation of small terms in the expansion. While successful for atoms and molecules of modest size, these approaches still tend to become intractable for larger systems such as biomolecules and solids.

For those systems, a very widely used approach is that of *density-functional theory* (DFT), which reduces Eq. (2.9) to an approximate one-particle mean-field Hamiltonian. DFT was introduced...

In the DFT approach, one constructs an effective single-particle Kohn-Sham Hamiltonian

$$H_{\mathrm{KS}} = \frac{p^2}{2m} + V_{\mathrm{KS}}^{[n]}(\mathbf{r}) \tag{2.10}$$

where the superscript on the Kohn-Sham potential V_{KS} indicates that it is a functional of the electron density $n(\mathbf{r})$. Specifically,

$$V_{\mathrm{KS}}^{[n]}(\mathbf{r}) = V_{\mathrm{ext}}(\mathbf{r}) + V_{\mathrm{H}}^{[n]}(\mathbf{r}) + V_{\mathrm{XC}}^{[n]}(\mathbf{r}) \tag{2.11}$$

where

$$V_{\mathrm{H}}^{[n]}(\mathbf{r}) = e^2 \int d^3r' \, \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2.12}$$

is the Hartree potential, reflecting the Coulomb repulsion from the electron cloud, and

$$V_{\mathrm{XC}}^{[n]}(\mathbf{r}) = \frac{\delta E_{\mathrm{XC}}[n]}{\delta n(\mathbf{r})} \tag{2.13}$$

is the exchange-correlation potential, which encodes non-classical many-body effects. Here V_{XC} is written formally as a functional derivative of the exchange-correlation energy $E_{\mathrm{XC}}[n]$, which plays a central role in the DFT formalism, but approximations are applied in practice. The eigensolution of the Kohn-Sham Hamiltonian of Eq. (2.10) is then

$$H_{\mathrm{KS}} |\psi_i\rangle = E_i |\psi_i\rangle \tag{2.14}$$

where E_i and $|\psi_i\rangle$ are the single-particle Kohn-Sham eigenvalues and eigenvectors.[1] The ground-state solution is given by identifying the lowest N eigenvalues and constructing the charge density

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \,, \tag{2.15}$$

which is then plugged back into Eqs. (2.12) and (2.13) to obtain a new Kohn-Sham potential H_{KS}. The process is iterated until a self-consistent solution for $n(\mathbf{r})$ and $V_{\mathrm{KS}}(\mathbf{r})$ is obtained. The most time-consuming step is typically the eigensolution of Eq. (2.14), but since this is now only a single-particle problem, it can be carried out efficiently even for large molecules and solids.

This gain in efficiency comes at the cost of introducing an approximation for the exchange-correlation functional $E_{\mathrm{XC}}[n]$ appearing in Eq. (2.13). The simplest is the local-density approximation (LDA), which leads to an exchange-correlation potential of the form

$$V_{\mathrm{XC}}^{[n]}(\mathbf{r}) = \mu_{\mathrm{XC}}(n(\mathbf{r})) \tag{2.16}$$

where $\mu_{\mathrm{XC}}(n)$ is a function encoding the exchange-correlation potential felt by an electron in a uniform electron gas of density n, and this function is evaluated at the local density $n(\mathbf{r})$. This function is extracted once and for all from high-level many-body calculations on the uniform electron gas. Many more sophisticated functionals have since been invented, among which variants of the generalized-gradient approximation (GGA) are especially popular. For an overview...

For our purposes, all we really have to know is that the many-electron problem can be reduced to an approximate one-particle form, as given by Eq. (2.10), with reasonable accuracy for materials that are not-too-strongly correlated. We shall work under this approximation henceforth. We shall also assume that the expectation value of any one-particle observable $\mathcal{O}$ (e.g., dipole operator, momentum or current operator, etc.) is given by

$$\langle \mathcal{O} \rangle = \sum_{i=1}^{N} \langle \psi_i | \mathcal{O} | \psi_i \rangle \tag{2.17}$$

where the $|\psi_i\rangle$ are the occupied Kohn-Sham orbitals.

One other approximation that is commonly used in practice is the *pseudopotential approximation*, in which the attraction of an electron to nucleus I in Eq. (2.8) is replaced by an effective attraction $V_I^{\mathrm{PS}}(|\mathbf{r} - \mathbf{r}_I|)$. The pseudopotential V^{PS} is constructed, once and for all for each atomic species

[1] Note that i now runs over one-particle states, not over electrons as in Eq. (2.9).

needed for the calculation, in such a way as to reproduce the properties of the valence electrons as closely as possible while avoiding the need to include core electrons in the calculation. Again, the text by Martin (2004) is recommended for an overview of this topic.

2.1.2 Spin, spin-orbit coupling, and external fields

The development in Sec. 2.1.1 was written for the case of spinless electrons. In reality, of course, electrons have to be treated as spinors. For materials that do not contain heavy atoms from the bottom of the periodic table, spin-orbit coupling can usually be safely neglected; in this case the spin-up and spin-down systems of electrons can be treated independently.[2] The Kohn-Sham equations are then easily generalized by attaching spin labels to the wavefunctions, potentials, and densities. That is, spin-up electrons obey

$$\left(\frac{p^2}{2m} + V_{\text{ext}} + V_{\text{H}} + V_{\text{XC},\uparrow}\right) |\psi_{i\uparrow}\rangle = E_{i\uparrow}\, |\psi_{i\uparrow}\rangle \tag{2.18}$$

and similarly for down electrons. The Hartree potential remains as written in Eq. (2.12), where $n(\mathbf{r}) = n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r})$ is now the total electron density. The exchange correlation potential of Eq. (2.16) has to be generalized to the pair of potentials $\mu_\uparrow^{\text{XC}}(n_\uparrow, n_\downarrow)$ and $\mu_\downarrow^{\text{XC}}(n_\uparrow, n_\downarrow)$, which are adapted from results on the uniform spin-polarized electron gas. This theory goes under the name of the local spin-density approximation (LSDA). For non-magnetic systems the spin-up and spin-down systems are identical; one can then treat the electrons as though they were spinless, only inserting a factor of two in the right-hand side of Eq. (2.15) – and, more generally, Eq. (2.17) – to account for the spin degeneracy.

When present, the spin-orbit interaction mixes spin-up and spin-down electron states in such a way that a true spinor treatment is unavoidable. The spin-orbit coupling (SOC) is a relativistic effect, and becomes important for heavy elements, roughly those with atomic numbers $Z > 50$, because the electrons travel at relativistic speeds in the core region of the atom. The Kohn-Sham equation then takes the form

$$\left[\frac{p^2}{2m} + V_{\text{ext}} + V_{\text{H}} + V_{\text{XC}} + \mathbf{W}_{\text{XC}} \cdot \boldsymbol{\sigma} + h_{\text{SOC}}\right] \begin{pmatrix} \psi_{i\uparrow} \\ \psi_{i\downarrow} \end{pmatrix} = E_i \begin{pmatrix} \psi_{i\uparrow} \\ \psi_{i\downarrow} \end{pmatrix} \tag{2.19}$$

where V_{XC} and $\mathbf{W}_{\text{XC}}$ are the scalar and spin-dependent portions of the exchange-correlation potential respectively, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector

[2] As an exception, the presence of spin frustration can sometimes lead to noncollinear structures even when spin-orbit coupling is negligible.

of Pauli spin matrices. The spin-orbit Hamiltonian $h_{\rm SOC}$ is derived from the nonrelativistic limit of the Dirac equation and can be written as

$$h_{\rm SOC} = \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\nabla V(\mathbf{r})\times\mathbf{p} \tag{2.20}$$

where $V(\mathbf{r}) = -e\phi(\mathbf{r})$ is the electrostatic potential felt by the electron. (Recall that Gaussian units are being used.) Since $\nabla\phi$ is strongest in the core regions, $h_{\rm SOC}$ is often approximated by a sum of on-site "atomic" contributions of the form $\xi(r)\mathbf{L}\cdot\mathbf{S}$, where $\mathbf{L}$ and $\mathbf{S}$ are orbital and spin angular momenta on the site in question and $\xi(r)$ is a radial function related to $d\phi/dr$ in the core via

$$\xi(r) = \frac{-e}{2m^2c^2}\,\frac{1}{r}\,\frac{d\phi}{dr} \tag{2.21}$$

The spin-orbit coupling will play an important role in the discussion of topological insulators in Ch. 6.

For the case of a crystal under the influence of external electric and magnetic fields, expressed as $\boldsymbol{\mathcal{E}} = -\nabla\phi - c^{-1}d\mathbf{A}/dt$ and $\mathbf{B} = \nabla\times\mathbf{A}$ in terms of the scalar and vector potentials $\phi(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$, Eq. (2.19) needs to be extended in three ways: (i) the addition of a term $-e\phi(\mathbf{r})$ to describe the electrostatic potential; (ii) the canonical replacement $\mathbf{p}\to\mathbf{p}+e\mathbf{A}(\mathbf{r})/c$; and (iii) the addition of a Zeeman interaction $-\gamma\mathbf{B}(\mathbf{r})\cdot\mathbf{S}$, where γ is the gyromagnetic ratio ($\gamma = -e/mc$ in Gaussian units when taking the g-factor of the electron to be exactly $g_e = 2$, or $\gamma = -g_e e/2mc$ more generally). In the simplest case of a free electron in the presence of external fields the electronic Hamiltonian then takes the form

$$H = \frac{1}{2m}\big(\mathbf{p}+\frac{e}{c}\mathbf{A}(\mathbf{r})\big)^2 - e\phi(\mathbf{r}) + \frac{\hbar e}{2mc}\mathbf{B}(\mathbf{r})\cdot\boldsymbol{\sigma}\,. \tag{2.22}$$

In the following, we will typically suppress the spin indices and neglect external fields and spin-orbit coupling for simplicity.

2.1.3 Crystal potential, Bloch's theorem, and reciprocal space

Taking the crystal Hamiltonian to be given by Eq. (2.10), we simplify the notation further to

$$H = \frac{p^2}{2m} + V(\mathbf{r})\,. \tag{2.23}$$

For a 3D crystal the potential obeys the periodicity conditions

$$V(\mathbf{r}+\mathbf{R}_n) = V(\mathbf{r}) \tag{2.24}$$

for any lattice vector

$$\mathbf{R_n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3\,, \tag{2.25}$$

where $\mathbf{n} = (n_1, n_2, n_3)$ is a triplet of integers and $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are three primitive lattice vectors. Then the system is invariant under any of the three translation operators $T_{\mathbf{a}_j}$, where $T_\mathbf{u}$ is defined as the translation operator that shifts the entire system by displacement $\mathbf{u}$, i.e., $T_\mathbf{u}\, f(\mathbf{r}) = f(\mathbf{r}-\mathbf{u})$. Thus, the commutators $[H, T_{\mathbf{a}_j}]$ all vanish, and moreover the three translation operators commute with one another, so it follows that the eigenfunctions of H can be chosen to be simultaneous eigenfunctions of all three of the $T_{\mathbf{a}_j}$. Because $T_{\mathbf{a}_j}$ is unitary, its eigenvalues must have unit norm and can thus be written as $e^{-i\kappa_j}$ for some phase angle κ_j.

This leads directly to Bloch's theorem. The above conditions imply that the eigenvectors of H can be labeled as $\psi_{n\boldsymbol{\kappa}}$, where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)$ is a triplet of phase angles and n is an additional label (the "band index"[3]) that counts states of the same $\boldsymbol{\kappa}$ in order of increasing energy, such that

$$H|\psi_{n\boldsymbol{\kappa}}\rangle = E_{n\boldsymbol{\kappa}}\,|\psi_{n\boldsymbol{\kappa}}\rangle\,. \tag{2.26}$$

The label $\boldsymbol{\kappa}$ indicates that the state transforms under translations as $T_{\mathbf{a}_j}|\psi_{n\boldsymbol{\kappa}}\rangle = e^{-i\kappa_j}\,|\psi_{n\boldsymbol{\kappa}}\rangle$, or equivalently,

$$\psi_{n\boldsymbol{\kappa}}(\mathbf{r} + \mathbf{R_n}) = e^{i\boldsymbol{\kappa}\cdot\mathbf{n}}\psi_{n\boldsymbol{\kappa}}(\mathbf{r})\,. \tag{2.27}$$

Note that κ_j and $\kappa_j + 2\pi$ correspond to the same eigenvalue $e^{-i\kappa_j}$ of $T_{\mathbf{a}_j}$, and thus label the same state, so that κ_j is best regarded as living on the circumference of a unit circle. By the same token, the set of labels (κ_1, κ_2) in 2D has the topology of a torus, and the triplet of labels $(\kappa_1, \kappa_2, \kappa_3)$ in 3D has the topology of a 3-torus. The fact that these are closed spaces has strong implications for the existence of topological properties, as will be discussed in later chapters.

To cast Bloch's theorem into a more familiar form, we introduce the reciprocal lattice. Its primitive reciprocal lattice vectors $\mathbf{b}_j$ are defined so as to be dual to the $\mathbf{a}_j$ in the sense that $\mathbf{a}_i\cdot\mathbf{b}_j = 2\pi\delta_{ij}$. Explicitly, $\mathbf{b}_1 = 2\pi(\mathbf{a}_2\times\mathbf{a}_3)/V_{\text{cell}}$ (and similarly for $\mathbf{b}_2$ and $\mathbf{b}_3$ by permutation of indices) since the cell volume is $V_{\text{cell}} = \mathbf{a}_1\cdot\mathbf{a}_2\times\mathbf{a}_3$.[4] An arbitrary reciprocal lattice vector is of the form

$$\mathbf{G_m} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3\,, \tag{2.28}$$

and the real and reciprocal lattices are easily seen to obey the important

[3] The band index n is not to be confused with the lattice vector index $\mathbf{n} = (n_1, n_2, n_3)$.

[4] In 2D the corresponding construction is $\mathbf{b}_1 = 2\pi(\mathbf{a}_2\times\hat{\mathbf{n}})/A_{\text{cell}}$ and $\mathbf{b}_2 = 2\pi(\hat{\mathbf{n}}\times\mathbf{a}_1)/A_{\text{cell}}$, where $\hat{\mathbf{n}}$ is the unit normal in the direction of $\mathbf{a}_1\times\mathbf{a}_2$.

condition

$$e^{i\mathbf{G_m}\cdot\mathbf{R_n}} = 1 \tag{2.29}$$

for any $\mathbf{G_m}$ and any $\mathbf{R_n}$. Finally, we define the wavevector $\mathbf{k}$ to be

$$\mathbf{k} = \frac{\kappa_1}{2\pi}\mathbf{b}_1 + \frac{\kappa_2}{2\pi}\mathbf{b}_2 + \frac{\kappa_3}{2\pi}\mathbf{b}_3\,, \tag{2.30}$$

and adopt $\mathbf{k}$ in place of $\boldsymbol{\kappa}$ as the label of the eigenvector. Bloch's theorem then takes the familiar form

$$H|\psi_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}\,|\psi_{n\mathbf{k}}\rangle \tag{2.31}$$

where $\psi_{n\mathbf{k}}$ transforms under translations as

$$\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\,\psi_{n\mathbf{k}}(\mathbf{r}) \tag{2.32}$$

for any lattice vector $\mathbf{R}$.

We again emphasize that $\mathbf{k}$, being just a rescaled version of $\boldsymbol{\kappa}$, lives in a closed space having the topology of a 3-torus. This space is called the Brillouin zone (BZ). Alternatively, we can identify the BZ with a "unit cell" in reciprocal space, such that tiling it by all reciprocal lattice vectors $\mathbf{G}$ covers the reciprocal space exactly once. Indeed, notice that the wavevectors $\mathbf{k}$ and $\mathbf{k}+\mathbf{G}$ are duplicate labels for the same state, for any reciprocal lattice vector $\mathbf{G}$, so a single unit cell in reciprocal space is enough to label each state once and only once. Different conventions may be used for specifying this cell, including a parallelogram centered at the origin ($\kappa_j \in [-\pi, \pi]$), a parallelogram with a corner at the origin ($\kappa_j \in [0, 2\pi]$), or a Wigner-Seitz cell (the locus of reciprocal-space points closer to the origin than to any other $\mathbf{G}$-vector).[5] The groups of states with the same n index are referred to as "bands," and bands that are nowhere degenerate with the next lower or higher band are denoted as "isolated bands."

Equations (2.26-2.27) and Eqs. (2.31-2.32) are equivalent statements of Bloch's theorem, with the first pair being written in "internal coordinates" while the second pair is in the usual Cartesian coordinates. The former is often more convenient in practical calculations, but in any case we shall go back and forth freely between the representations as needed below.

The presentation above has mainly been framed in terms of ordinary 3D crystals; some hints were given about the 2D case, and the generalization to 1D is straightforward. It is worth adding a word here, however, about the terminology to be used in this book. Systems are classified here by the dimensionality of the lattice (or equivalently, of the reciprocal lattice).

[5] Some authors reserve the term "Brillouin zone" for the Wigner-Seitz reciprocal-space cell, but I adopt a looser terminology here.

Thus, "1D systems" will be taken to include periodic polymers or nanotubes embedded in 3D space, or nanoribbons embedded in 2D space, in addition to "true" 1D systems described Hamiltonians of the form $p^2/2m + V(x)$. Similarly, graphene and MoS_2 sheets will be regarded as "2D systems" even though they have some extent in 3D, and molecules and clusters will be classified as 0D. In this terminology, therefore, the dimensionality of the system is the same as the dimension of the BZ.

2.1.4 Electron counting

In order to count electron states, it is convenient to impose periodic boundary conditions on a supercell built from $N_1 \times N_2 \times N_3 = N$ primitive cells, with the understanding that the limit $N_j \to \infty$ will be taken later. Then **k**-space becomes discretized in the sense that only $\psi_{n\mathbf{k}}$ with $\mathbf{k} = (n_1/N_1)\mathbf{b}_1 + (n_2/N_2)\mathbf{b}_2 + (n_3/N_3)\mathbf{b}_3$ are consistent with these boundary conditions. Then for any given band, there are N states in the supercell of volume NV_{cell}, or one per V_{cell}. This establishes the principle that a filled band contributes a density of exactly one electron per unit cell.[6]

It also implies that the density of states per unit volume of **k**-space is $V_{\text{cell}}/(2\pi)^3$. A useful heuristic relation embodying this principle is

$$\sum_{\mathbf{k}} \quad \longleftrightarrow \quad \frac{V}{(2\pi)^3} \int d^3k. \tag{2.33}$$

where $V = NV_{\text{cell}}$ is the supercell volume. Then, for example, the average of some quantity $g(\mathbf{k})$ over the BZ can be written equivalently as

$$\overline{g} = \frac{1}{N} \sum_{\mathbf{k}}^{\text{BZ}} g(\mathbf{k}) = \frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} g(\mathbf{k})\, d^3k\,. \tag{2.34}$$

As another example, the density of states $\rho(E)$, defined as the number of states per unit energy interval per unit volume, is given by

$$\rho(E) = \frac{1}{(2\pi)^3} \sum_n \int_{\text{BZ}} \delta(E - E_{n\mathbf{k}})\, d^3k \tag{2.35}$$

where δ is a Dirac delta-function. So, for example, the electron density $n(\mathbf{r})$ (number of electrons per unit volume) can be written in terms of the Bloch functions as

$$n(\mathbf{r}) = \frac{1}{N} \sum_{n\mathbf{k}} f_{n\mathbf{k}} \, |\psi_{n\mathbf{k}}(\mathbf{r})|^2 = \frac{V_{\text{cell}}}{(2\pi)^3} \sum_n \int_{\text{BZ}} f_{n\mathbf{k}} \, |\psi_{n\mathbf{k}}(\mathbf{r})|^2 \, d^3k\,, \tag{2.36}$$

[6] When spin degeneracy is present, a "band" is often taken to comprise both spins; in this case it contributes two electrons per cell.

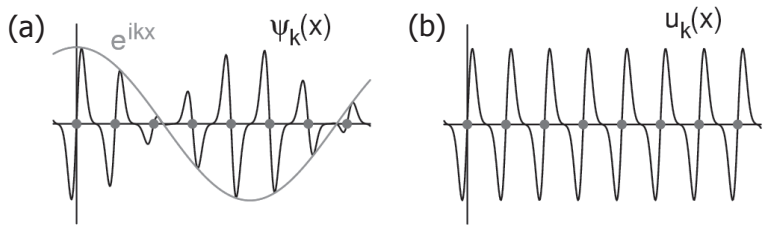


Figure 2.1 (a) Bloch wavefunction $\psi_k(x)$ showing how it is modulated by an envelope function e^{ikx}. (b) Definition of cell-periodic Bloch function $u_k(x)$, defined in such a way that the envelope function has been factored out.

where $f_{n\mathbf{k}}$ is the Fermi function taking values 1 or 0 for occupied and unoccupied states respectively. That is, the electronic contribution to the charge density is just $\rho_{\text{elec}}(\mathbf{r}) = -en(\mathbf{r})$.

2.1.5 Cell-periodic Bloch functions

The Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$ obey twisted periodic boundary conditions on the primitive cell as specified in Eq. (2.32). This means that Bloch functions with different $\mathbf{k}$ obey different boundary conditions, a feature that would be awkward for many purposes, and it is often convenient to work instead with the *cell-periodic Bloch functions*

$$u_{n\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{n\mathbf{k}}(\mathbf{r}) \tag{2.37}$$

which instead obey ordinary boundary conditions

$$u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r})\,. \tag{2.38}$$

We can then think of the Bloch function $\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$ as a product of a plane-wave envelope function $e^{i\mathbf{k}\cdot\mathbf{r}}$ times an underlying cell-periodic function $u_{n\mathbf{k}}(\mathbf{r})$, as illustrated in Fig. 2.1. Viewed as functions living in one primitive cell, the $u_{n\mathbf{k}}$ all belong to the same Hilbert space (periodic functions defined on this cell), whereas the ordinary Bloch functions do not. This has many desirable properties; for example, derivatives such as $\nabla_{\mathbf{k}}u_{n\mathbf{k}}$ are well-defined periodic functions belonging to the same Hilbert space, whereas $\nabla_{\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r}) = e^{\imath\mathbf{k}\cdot\mathbf{r}}[\nabla_{\mathbf{k}}u_{n\mathbf{k}}(\mathbf{r}) + i\mathbf{r}u_{n\mathbf{k}}(\mathbf{r})]$ has a piece that blows up linearly with distance from the origin. For these reasons it will turn out to be essential to use the $u_{n\mathbf{k}}$ and not the $\psi_{n\mathbf{k}}$ when defining Berry phases and related quantities for the Bloch functions in Sec. 3.4.

It is also conventional to define the **k**-dependent effective Hamiltonian

$$H_{\mathbf{k}} = e^{-i\mathbf{k}\cdot\mathbf{r}}\, H\, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.39}$$

such that

$$H_{\mathbf{k}}\,|u_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}\,|u_{n\mathbf{k}}\rangle\,. \tag{2.40}$$

In the same spirit, any one-particle operator $\mathcal{O}$ can be converted into a **k**-dependent operator in this representation via $\mathcal{O}_{\mathbf{k}} = e^{-i\mathbf{k}\cdot\mathbf{r}}\mathcal{O}e^{i\mathbf{k}\cdot\mathbf{r}}$, such that expectation values take the form

$$\langle\mathcal{O}\rangle = \frac{1}{N}\sum_{n\mathbf{k}}\langle\psi_{n\mathbf{k}}|O|\psi_{n\mathbf{k}}\rangle = \frac{1}{N}\sum_{n\mathbf{k}}\langle u_{n\mathbf{k}}|\mathcal{O}_{\mathbf{k}}|u_{n\mathbf{k}}\rangle \tag{2.41}$$

where N is the number of **k**-points in the BZ mesh. Of particular interest is the velocity operator

$$\mathbf{v} = \frac{-i}{\hbar}\,[\mathbf{r}, H]\,, \tag{2.42}$$

which is the Heisenberg-picture time derivative of the position operator **r**; in the transformed representation $\mathbf{v}_{\mathbf{k}}$ obeys the useful formula

$$\mathbf{v}_{\mathbf{k}} = \frac{1}{\hbar}\nabla_{\mathbf{k}} H_{\mathbf{k}} \tag{2.43}$$

(see Ex. 2.1.1).

Exercises

Exercise 2.1.1

(a) Using only the very general definition of the velocity operator given in Eq. (2.42), show that the transformed velocity operator $\mathbf{v}_{\mathbf{k}} = e^{-i\mathbf{k}\cdot\mathbf{r}}\mathbf{v}e^{i\mathbf{k}\cdot\mathbf{r}}$ is given by Eq. (2.43).

(b) For an electron (mass m, charge $-e$) in the presence of an external scalar potential $\phi(\mathbf{r})$ and vector potential $\mathbf{A}(\mathbf{r})$, the Hamiltonian is

$$H = \frac{1}{2m}\big(\mathbf{p} + \frac{e}{c}\mathbf{A}(\mathbf{r})\big)^2 - e\phi(\mathbf{r})$$

Find $\mathbf{v}$, convert to $\mathbf{v}_{\mathbf{k}}$, and check consistency with the expression in part (a).

2.2 Tight-binding model Hamiltonians

Using modern methods of computational electronic-structure theory, the Kohn-Sham equations introduced in Sec. 2.1.1 can be solved to very good

precision; the book by Martin (2004) gives a good overview of the technical methods employed to do so. However, one is often interested not so much in the full results of an accurate calculation, but in some simplified representation that is more likely to provide physical and chemical understanding. For this, localized-orbital and tight-binding (TB) methods are frequently adopted.

2.2.1 Finite systems

We consider first a localized system such as a molecule or cluster consisting of atoms indexed by μ located at positions $\boldsymbol{\tau}_\mu$. We choose some atomic-like orbitals $\varphi_{\mu\alpha}$ on site μ, where α runs over the orbitals on a given atom. These will typically be s-, p-, or d-like orbitals taking the form of a radial function times the appropriate $Y_{lm}(\theta,\phi)$ spherical harmonics. Introducing a compound index $j = \{\mu\alpha\}$ that runs over all M localized orbitals in the basis, we can write the trial Hamiltonian eigenstates as

$$\psi_n(\mathbf{r}) = \sum_j C_{nj}\,\varphi_j(\mathbf{r} - \boldsymbol{\tau}_j) \tag{2.44}$$

where the C_{nj}, the expansion coefficients of the n'th eigenstate and the j'th basis orbital, are to be determined. Plugging this into the Schrödinger equation $H|\psi_n\rangle = E_n|\psi_n\rangle$, the solutions are given by solving the matrix equation

$$(H - E_n\,S)\,C_n = 0 \tag{2.45}$$

where H and S are the $M \times M$ Hamiltonian and overlap matrices

$$\begin{aligned} H_{ij} &= \langle\varphi_i|H|\varphi_j\rangle\,, \\ S_{ij} &= \langle\varphi_i|\varphi_j\rangle\,, \end{aligned} \tag{2.46}$$

and C_n is the M-component column vector of coefficients C_{nj}. Equation (2.45) is a generalized eigenvalue problem that can be solved using standard computational methods, and the Hamiltonian eigenstates can be constructed using Eq. (2.44).

There are different philosophies for the implementation of this approach, depending on the desired accuracy. At the accurate end of the spectrum are the methods of quantum chemistry, where one chooses a fairly extensive set of basis functions, typically including two or more radial functions per angular-momentum channel, and the calculations are carried out in the context of Hartree-Fock or DFT at a minimum, and more often using correlated wave-function approaches. The GAUSSIAN code package is a well-known implementation of this kind. At the intermediate level are methods

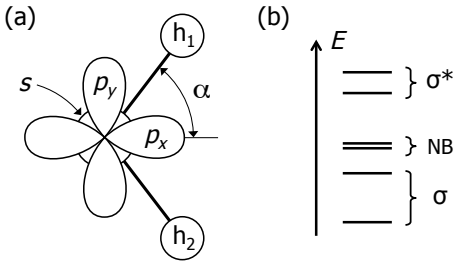


Figure 2.2 (a) Empirical tight-binding model for the water molecule. The oxygen p_z orbital, directed out of the plane, is not shown. (b) Eigenspectrum showing unoccupied antibonding (σ^*) orbitals and occupied nonbonding (NB) and bonding (σ) orbitals.

that sometimes go under the name of "linear combination of atomic orbitals" (LCAO), in which one adopts a minimal basis, typically just one radial function per angular-momentum channel, but still computes the matrix elements and overlaps of Eq. (2.46) explicitly, usually at the DFT level. Here we shall be interested in an implementation that lies at the low-accuracy end of this spectrum, known as the *empirical tight-binding* approach, where the idea is to construct a simple model containing the essential ingredients so as to enhance physical understanding.

The philosophy of empirical TB is, first, to focus just on those basis orbitals φ_j needed to describe the valence and low-lying conduction states of the system of interest. But second, these orbitals are never explicitly constructed; instead, the Hamiltonian and overlap matrix elements between them, given by Eq. (2.46), are parametrized in a model-building sense. Typically the hopping matrix elements between orbitals are truncated to connect only nearest-neighbor, or perhaps second-neighbor, sites, and the overlap matrix is often taken to be the unit matrix ("orthogonal tight binding"). Equation (2.45) is again solved using standard computational methods to get the energies E_n, but the corresponding eigenstates $\psi_n(\mathbf{r})$ are never actually constructed; the vector of coefficients $\mathcal{C}_n$ plays the role of eigenstate instead.

As an example, let's set up and solve an empirical TB model of the water molecule. We take the relevant orbitals to be $|s\rangle$, $|p_x\rangle$, $|p_y\rangle$, $|p_z\rangle$, $|h_1\rangle$, and $|h_2\rangle$, as illustrated in Fig. 2.2(a), and assume they are orthonormal. The first four are located on the oxygen and are assigned site energies (E_s, E_p, E_p, E_p), while the last two are s orbitals on the hydrogens and have site energies $(E_\mathrm{h}, E_\mathrm{h})$. We also let there be non-zero hoppings $t_s = \langle s|H|h_1\rangle$ between the oxygen s orbital and a hydrogen, and t_p between an oxygen p-orbital and a

hydrogen toward which it is directed. Then the 6×6 Hamiltonian matrix is

$$H_{\mathrm{H_2O}} = \begin{pmatrix} E_s & 0 & 0 & 0 & t_s & t_s \\ 0 & E_p & 0 & 0 & t_p \cos\alpha & t_p \cos\alpha \\ 0 & 0 & E_p & 0 & t_p \sin\alpha & -t_p \sin\alpha \\ 0 & 0 & 0 & E_p & 0 & 0 \\ t_s & t_p \cos\alpha & t_p \sin\alpha & 0 & E_h & 0 \\ t_s & t_p \cos\alpha & -t_p \sin\alpha & 0 & 0 & E_h \end{pmatrix} \tag{2.47}$$

where α is the half-bond-angle sketched in the figure. We thus arrive at simple but instructive six-parameter model. The eigenenergies E_j are the six solutions of the secular equation $\det[H - E\mathbb{1}] = 0$, where $\mathbb{1}$ is the 6×6 unit matrix.

While this looks difficult to solve by hand, the problem can be simplified by noting that the matrix can be block-diagonalized on the basis of symmetry. The orbital $|p_z\rangle$ is the only one that is odd in z, and thus forms an eigenvector by itself. The remaining five that are even in z can be grouped into three that are even and two that are odd in y, after reorganizing the H orbitals into linear combinations $[|h_1\rangle \pm |h_2\rangle]/\sqrt{2}$. Still, a closed-form solution requires the roots of a cubic equation.

On the other hand, a direct solution on the computer is straightforward. Based on chemical experience and/or higher-level calculations, the choice of parameter values $E_s = -1.5$, $E_p = -1.2$, $E_h = -1.0$, $t_s = -0.4$, $t_p = -0.3$ and $\alpha = 52^\circ$ (energies in Ry) is roughly representative of H_2O. This parameter set yields the spectrum shown in Fig. 2.2(b) with corresponding eigenvalues and eigenvectors

```
n   eigval   eigvec
1   -1.896   (  0.80,   0.20,   0.00,   0.00,   0.40,   0.40 )
2   -1.458   ( -0.00,   0.00,   0.80,   0.00,   0.42,  -0.42 )
3   -1.242   ( -0.34,   0.93,   0.00,   0.00,   0.11,   0.11 )
4   -1.200   (  0.00,  -0.00,  -0.00,   1.00,  -0.00,   0.00 )
5   -0.742   ( -0.00,   0.00,   0.60,   0.00,  -0.57,   0.57 )
6   -0.562   (  0.49,   0.32,  -0.00,  -0.00,  -0.57,  -0.57 )
```

The four lowest-energy states are doubly occupied (counting spin) and the last two are empty.

In the spirit of empirical TB, we do not expect these results to be highly accurate, but they provide some useful qualitative information and insight. For example, the two lowest eigenstates can be seen to be bonding combinations of O *sp* hybrids and H *s* orbitals, while the unoccupied ones are the corresponding antibonding orbitals, and the middle states have an O non-bonding character. The symmetries are also evident from an inspection

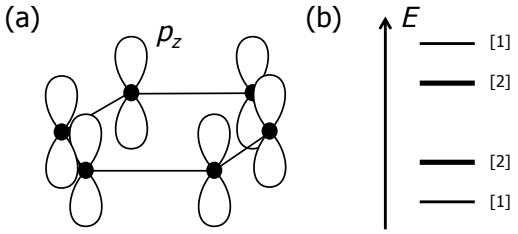


Figure 2.3 (a) Empirical tight-binding model for the π manifold on the benzene molecule (hydrogen atoms not shown). (b) Eigenspectrum of the model; numbers in brackets indicate the degeneracy.

of the eigenvectors; for example, the state at E_1 is fully symmetric in both y and z.

As a second example, suppose we are interested in the manifold of π states on the benzene molecule, made from linear combinations of p_z orbitals on the C atoms as shown in Fig. 2.3(a). These are distinguished from the C–C and C–H σ-bonding and antibonding states by being odd under the mirror M_z and being closer to the Fermi energy, so we discard all other orbitals and just model the behavior of the π manifold. This time we write the Hamiltonian in the notation of raising and lowering operators as

$$H = E_p \sum_j c_j^\dagger c_j + t\Big(\sum_j c_{j+1}^\dagger c_j + \text{h.c.}\Big) \tag{2.48}$$

where j runs over the six orbitals around the ring ($j+1$ is interpreted mod 6) and E and t are site energies and nearest-neighbor hoppings respectively. This model is easily solved; using the six-fold symmetry the eigenvectors are of the form $C_{nj} = e^{2\pi inj/6}$. The spectrum of states is sketched in Fig. 2.3(b); the three lowest states are occupied.

Once the TB Hamiltonian matrix elements H_{ij} are specified, no other information is needed to compute the spectrum of energy eigenstates as above. For some other purposes, however, additional information may be needed. For example, in general one should specify the position matrix elements

$$X_{ij} = \langle\varphi_i|x|\varphi_j\rangle\,, \quad Y_{ij} = \langle\varphi_i|y|\varphi_j\rangle\,, \quad Z_{ij} = \langle\varphi_i|z|\varphi_j\rangle\,, \tag{2.49}$$

which may be needed, e.g., for computing electric dipoles. In the spirit of the empirical TB method, we typically make the simplest possible approximation, namely that these matrices are diagonal with $\langle\varphi_i|\mathbf{r}|\varphi_j\rangle = \delta_{ij}\,\boldsymbol{\tau}$, where $\boldsymbol{\tau}_j$ is the nuclear coordinate of the atom on which the orbital is centered.[7]

[7] The diagonal approximation misses some important physics in the case of orbitals of mixed parity on the same site. For example, a matrix element of the form $\langle s|x|p_x\rangle$ between s and p_x

2.2.2 The PythTB package

It turns out that the concepts at the heart of this book – Berry phases, Berry curvature, electric polarization, magnetoelectric couplings, topological insulators – can all be illustrated insightfully using TB models. This was the motivation for the development of the PythTB software package, developed mainly by students in the Department of Physics and Astronomy at Rutgers University several years ago. PythTB is a software package written in the Python programming language, and is designed to allow the user to construct and solve TB models of the electronic structure, not only for finite systems such as those introduced above, but also for crystals, slabs, ribbons, polymer chains, and other configurations that display periodicity in one or more dimensions. It is also rich with features for computing Berry phases and related properties.

The PythTB package is described in some detail in App. A, but we introduce it here by using it to solve the H_2O and benzene model systems discussed above. Aside from a few lines of header material that are not shown, the code needed to solve the water-molecule model looks like this:

```
#!/usr/bin/env python

# ------------------------------------------------------------
# tight-binding model for H2O molecule
# ------------------------------------------------------------

#import the pythtb module
from pythtb import *

#import pythtb extensions
import ptbe

# geometry: bond length and half bond-angle
b=1.0; angle=54.0*np.pi/180

# site energies [O(s), O(p), H(s)]
eos=-1.5; eop=-1.2; eh=-1.0

# hoppings [O(s)-H(s), O(p)-H(s)]
ts=-0.4; tp=-0.3

# define frame for defining vectors: 3D Cartesian
lat=[[1.0,0.0,0.0],[0.0,1.0,0.0],[0.0,0.0,1.0]]

# define coordinates of orbitals: O(s,px,py,pz) ; H(s) ; H(s)
orb=[ [0.,0.,0.], [0.,0.,0.], [0.,0.,0.], [0.,0.,0.],
      [b*np.cos(angle), b*np.sin(angle),0.],
      [b*np.cos(angle),-b*np.sin(angle),0.] ]

my_model=tbmodel(0,3,lat,orb)
```

orbitals on the same site is needed to describe the fact that the center of charge of an sp^3 hybrid is displaced from the atomic center.

```
my_model.set_onsite([eos,eop,eop,eop,eh,eh])
my_model.set_hop(ts,0,4)
my_model.set_hop(ts,0,5)
my_model.set_hop(tp*np.cos(angle),1,4)
my_model.set_hop(tp*np.cos(angle),1,5)
my_model.set_hop(tp*np.sin(angle),2,4)
my_model.set_hop(-tp*np.sin(angle),2,5)

my_model.display()

(eval,evec)=my_model.solve_all(eig_vectors=True)

# my_print(eval,evec)

# signs of evec's are regularized by the following rule
# (this is arbitrary and not very important)
for i in range(len(eval)):
  if sum(evec.real[i,1:4]) < 0:
    evec[i,:]=-evec[i,:]

ptbe.print_eig_real(eval,evec)
```

The first line imports the PythTB module. The next few lines define the various model parameters, then the `lat=` line specifies that coordinates will be given in a 3D Cartesian space, while the `orb` line specifies the locations of the atoms in this space. The `my_model=` line calls the `tbmodel` function of the PythTB package with four input variables. The last two are just the `lat` and `orb` variables just defined; the first two specify the number of periodic dimensions (none since this is a finite model) and the dimension of the atomic coordinate space (three here). The Python variable `my_model` is assigned a "value" embodying all this information. The model is then augmented further by defining the onsite energies and hoppings using the `set_onsite` and `set_hop` functions applied to `my_model`. (Note that Python indexes arrays starting from 0, not 1, so the six orbitals are labeled as {0,1,2,3,4,5}.) Finally, the `solve_all` function is applied to `my_model` and the resulting eigenvalues and eigenvectors are returned in arrays `eval` and `evec`. The last line prints these to the terminal. The `tbmodel`, `set_onsite`, `set_hop`, and `solve_all` functions (and many more not used here) are provided by the PythTB module that was imported in the first line. The output, produced by the `ptbe.print_eig_real` function produces output that ends as follows:

```
n   eigval   eigvec
0   -1.896   (  0.80,   0.20,   0.00,   0.00,   0.40,   0.40 )
1   -1.458   ( -0.00,   0.00,   0.80,   0.00,   0.42,  -0.42 )
2   -1.242   ( -0.34,   0.93,   0.00,   0.00,   0.11,   0.11 )
3   -1.200   (  0.00,  -0.00,  -0.00,   1.00,  -0.00,   0.00 )
4   -0.742   ( -0.00,   0.00,   0.60,   0.00,  -0.57,   0.57 )
5   -0.562   (  0.49,   0.32,  -0.00,  -0.00,  -0.57,  -0.57 )
```

This was the basis of the results presented earlier on p. 46.[8] A PythTB code for solving the benzene model of Eq. (2.48) can be found in App. A.

Obviously the reader is not expected to follow this example of PythTB programming in any detail at this point, but the above should serve to illustrate the main features. A TB model can be specified in a few lines of user-written code, and the solution is then done automatically by the package subroutines.

2.2.3 Extended systems

The empirical TB framework is easily extended to the case of systems that are periodic in one or more dimensions. The presentation is done here for a 3D crystal, but the formulation for 2D layers such as graphene, or systems such as polymers, is straightforward.

The TB orbitals

$$\phi_{\mathbf{R}j}(\mathbf{r}) = \varphi_j(\mathbf{r} - \mathbf{R} - \boldsymbol{\tau}_j) \tag{2.50}$$

are now indexed by the lattice vector $\mathbf{R}$, which specifies the unit cell, and j, which runs over the M orbitals in the cell. We again assume orthonormality,

$$\langle \phi_{\mathbf{R}i} | \phi_{\mathbf{R}'j} \rangle = \delta_{\mathbf{R}\mathbf{R}'}\, \delta_{ij} \tag{2.51}$$

and that the position matrix have the simplest possible form,

$$\langle \phi_{\mathbf{R}i} | \mathbf{r} | \phi_{\mathbf{R}'j} \rangle = (\mathbf{R} + \boldsymbol{\tau}_j)\, \delta_{\mathbf{R}\mathbf{R}'}\, \delta_{ij}\ . \tag{2.52}$$

The Hamiltonian matrix elements are defined via

$$H_{ij}(\mathbf{R}) = \langle \phi_{\mathbf{R}'i} | H | \phi_{\mathbf{R}'+\mathbf{R},j} \rangle = \langle \phi_{\mathbf{0}i} | H | \phi_{\mathbf{R}j} \rangle \tag{2.53}$$

where $\mathbf{R}$ is now a relative index specifying a hopping from orbital i in cell $\mathbf{R}'$ to orbital j in cell $\mathbf{R}'+\mathbf{R}$, or equivalently because of translational symmetry, from orbital i in the home unit cell $\mathbf{R}=\mathbf{0}$ to orbital j in cell $\mathbf{R}$.

To make the transition to Bloch states, we first construct Bloch-like basis functions and then compute the Hamiltonian matrix elements in this basis. There are, however, two distinct ways of doing this that differ in detail, which we shall refer to as being associated with Convention I or II defined as follows.

In Convention I, Bloch-like basis functions are constructed as

$$|\chi_j^{\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}+\boldsymbol{\tau}_j)}\, |\phi_{\mathbf{R}j}\rangle \tag{2.54}$$

[8] In the raw output shown here, the band index n starts from zero in Python style; on p. 46 it was shifted by one to make the table look more conventional.

with the understanding that the normalization is to a single unit cell, i.e.,

$$\langle\chi|\chi'\rangle \equiv \int_{\text{cell}} d^3r\, \chi^*(\mathbf{r})\, \chi'(\mathbf{r})\,. \tag{2.55}$$

It follows from Eq. (2.51) that

$$\langle\chi_i^{\mathbf{k}}|\chi_j^{\mathbf{k}}\rangle = \delta_{ij}\,. \tag{2.56}$$

The Bloch eigenstates are then expanded as

$$|\psi_{n\mathbf{k}}\rangle = \sum_j C_j^{n\mathbf{k}}\, |\chi_j^{\mathbf{k}}\rangle \tag{2.57}$$

and the Hamiltonian matrix is constructed as

$$H_{ij}^{\mathbf{k}} = \langle\chi_i^{\mathbf{k}}|H|\chi_j^{\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}+\boldsymbol{\tau}_j-\boldsymbol{\tau}_i)}\, H_{ij}(\mathbf{R})\,. \tag{2.58}$$

The eigenvalue equation to be solved is

$$H_{\mathbf{k}} \cdot C_{n\mathbf{k}} = E_{n\mathbf{k}}\, C_{n\mathbf{k}} \tag{2.59}$$

where $H_{\mathbf{k}}$ is the $M \times M$ matrix of elements $H_{ij}^{\mathbf{k}}$ and $C_{n\mathbf{k}}$ is the column vector of elements $C_j^{n\mathbf{k}}$.

Standard numerical packages can easily be used to solve this eigenvalue problem, which amounts to diagonalizing $H_{\mathbf{k}}$, thereby giving the TB solution for the energy eigenvalues and eigenvectors. If only the eigenvalues are needed, these can be obtained from the secular equation

$$\det(H_{\mathbf{k}} - E_{n\mathbf{k}}) = 0\,. \tag{2.60}$$

Of course, this TB solution only produces M bands, where M is the number of TB basis orbitals per cell, representing an approximation to the L bands of the crystal that are built from these TB orbitals (usually these are the L lowest valence and conduction bands).

In Convention II, by contrast, the phase factor $e^{i\mathbf{k}\cdot\boldsymbol{\tau}_j}$ is not included in the definition of the Bloch-like basis functions. Using tilde'd quantities to denote objects defined in Convention II, we get

$$|\widetilde{\chi}_j^{\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, |\phi_{\mathbf{R}j}\rangle\,, \tag{2.61}$$

$$|\psi_{n\mathbf{k}}\rangle = \sum_j \widetilde{C}_j^{n\mathbf{k}}\, |\widetilde{\chi}_j^{\mathbf{k}}\rangle\,, \tag{2.62}$$

$$\widetilde{H}_{ij}^{\mathbf{k}} = \langle\widetilde{\chi}_i^{\mathbf{k}}|H|\widetilde{\chi}_j^{\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\, H_{ij}(\mathbf{R})\,, \tag{2.63}$$

and the secular equation is

$$\widetilde{H}_{\mathbf{k}} \cdot \widetilde{C}_{n\mathbf{k}} = E_{n\mathbf{k}} \widetilde{C}_{n\mathbf{k}} \, . \tag{2.64}$$

The quantities in the two conventions are related via

$$\widetilde{H}^{\mathbf{k}}_{ij} = e^{i\mathbf{k}\cdot(\boldsymbol{\tau}_i - \boldsymbol{\tau}_j)} \, H^{\mathbf{k}}_{ij} \tag{2.65}$$

and

$$\widetilde{C}^{n\mathbf{k}}_{j} = e^{i\mathbf{k}\cdot\boldsymbol{\tau}_j} \, C^{n\mathbf{k}}_{j} \, . \tag{2.66}$$

The two conventions are essentially just related by a unitary rotation in the M-dimensional space.

When reading the literature, it is worth taking care to determine which convention the authors are using. Convention II is probably the more common one, because the extra factors of $e^{i\mathbf{k}\cdot\boldsymbol{\tau}_j}$ can be ignored, and the relation between the Hamiltonian matrix elements expressed in Eq. (2.63) is that of a simple discrete Fourier transform. However, Convention I is in many ways more natural for our purposes, as we shall see later, and it is the convention adopted in the PythTB code.

One way of understanding the distinction between the two conventions is to draw an analogy between the Bloch function $\psi_{n\mathbf{k}}(\mathbf{r})$ and $\widetilde{C}^{n\mathbf{k}}_{j}$, and between the cell-periodic Bloch function $u_{n\mathbf{k}}(\mathbf{r})$ and $C^{n\mathbf{k}}_{j}$. Recalling that

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \, , \tag{2.67}$$

and temporarily adopting the change of notation $C^{n\mathbf{k}}_{j} \to C_{n\mathbf{k}}(j)$ and similarly for $\widetilde{C}$, we can write

$$\begin{aligned} |\psi_{n\mathbf{k}}\rangle &= \sum_{\mathbf{R}} \int_{\text{cell}} d^3r \; u_{n\mathbf{k}}(\mathbf{r}) \, e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{r})} \, |\mathbf{R}+\mathbf{r}\rangle \, , \\ &= \sum_{\mathbf{R}} \sum_{j} C_{n\mathbf{k}}(j) \, e^{i\mathbf{k}\cdot(\mathbf{R}+\boldsymbol{\tau}_j)} \, |\phi_{\mathbf{R}j}\rangle \end{aligned} \tag{2.68}$$

while

$$\begin{aligned} |\psi_{n\mathbf{k}}\rangle &= \sum_{\mathbf{R}} \int_{\text{cell}} d^3r \; \psi_{n\mathbf{k}}(\mathbf{r}) \, e^{i\mathbf{k}\cdot\mathbf{R}} \, |\mathbf{R}+\mathbf{r}\rangle \, , \\ &= \sum_{\mathbf{R}} \sum_{j} \widetilde{C}_{n\mathbf{k}}(j) \, e^{i\mathbf{k}\cdot\mathbf{R}} \, |\phi_{\mathbf{R}j}\rangle \, . \end{aligned} \tag{2.69}$$

If we identify $\int_{\text{cell}} d^3r \, ... \, |\mathbf{R}+\mathbf{r}\rangle$ in the continuum framework with $\sum_j \, ... \, |\phi_{\mathbf{R}j}\rangle$ in the TB one, we can see that Eq. (2.68) identifies $C^{n\mathbf{k}}_{j}$ (Convention I) as the TB analogy of the cell-periodic function $u_{n\mathbf{k}}(\mathbf{r})$, while Eq. (2.69) identifies the $\widetilde{C}^{n\mathbf{k}}_{j}$ (Convention II) with the ordinary Bloch function $\psi_{n\mathbf{k}}(\mathbf{r})$.

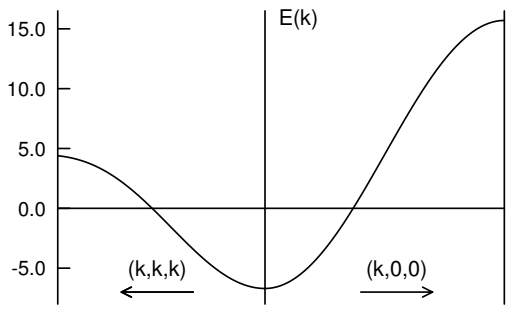


Figure 2.4 Band structure obtained from nearest-neighbor tight-binding model of s orbitals in bcc Li. The line at $E = 0$ indicates the approximate position of the Fermi level.

As hinted earlier on p. 42, the cell-periodic functions $u_{n\mathbf{k}}$ will play a more central role than the Bloch functions $\psi_{n\mathbf{k}}$ in the formulation of Berry-phase quantities such as electric polarization in the following chapters. It is largely for this reason that we have adopted Convention I for use here and for the PythTB implementation.

2.2.4 Examples

We first consider the dispersion of the s-like states of an alkali metal in the bcc crystal structure, taking Li (lattice constant $a = 3.5$ Å) for definiteness. We construct the simplest possible tight-binding model with one s orbital with site energy E_s on each atom, and a hopping of amplitude t between each atom and its eight nearest neighbors. The lattice vectors can be chosen as $\mathbf{a}_1 = (a/2)(-1,1,1)$, $\mathbf{a}_2 = (a/2)(1,-1,1)$, and $\mathbf{a}_3 = (a/2)(1,1,-1)$, in which case the nearest neighbors are located at $\pm\mathbf{a}_1$, $\pm\mathbf{a}_2$, $\pm\mathbf{a}_3$, and $\pm(\mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3)$, with $V_{\text{cell}} = a^3/2$. Since we only have one orbital per unit cell, we discard indices ij in the formalism of Sec. 2.2.3. The real-space Hamiltonian of Eq. (2.53) is just $H(\mathbf{0}) = E_s$, $H(\mathbf{R}) = t$ for the eight nearest-neighbor vectors enumerated above, and zero otherwise. Then $H_{\mathbf{k}}$ of Eq. (2.58) becomes

$$H_{\mathbf{k}} = E_0 + 2t\cos[(-k_x + k_y + k_z)a/2] + 3 \text{ more terms} \tag{2.70}$$

Since $H_{\mathbf{k}}$ is a 1×1 matrix, Eq. (2.60) implies that $E_{\mathbf{k}} = H_{\mathbf{k}}$, and we have obtained a closed-form solution for the energy bands.

Since it is not easy to visualize a function of three variables (k_x, k_y, k_z), it is conventional to plot the energy bands along some high-symmetry lines in reciprocal space. Let's do this first along the line $\mathbf{k} = (k, 0, 0)$, where we find $E(k) = E_0 + 8t\cos(ka/2)$. This is plotted for $k \in [0, 2\pi/a]$ in the right

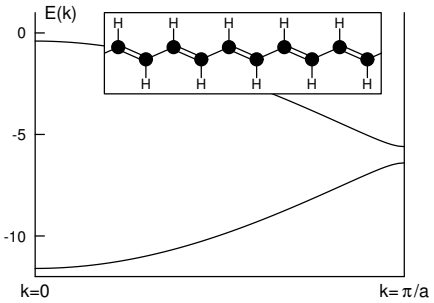


Figure 2.5 Dispersion of bands obtained from the tight-binding model of p_z orbitals on C atoms in $(CH)_x$ (polyacetylene) for parameter values of $E_p = -6.0\,\text{eV}$, $t = -2.8\,\text{eV}$, $\Delta = 0$, and $\delta = -0.2\,\text{eV}$ (see text). Inset shows a plan view of the molecule, which lies in the x-y plane; the C-C bonds alternate between double and single bonds as indicated in the sketch.

panel of Fig. 2.4 for parameters $E_s = 4.5\,\text{eV}$, $t = -1.4\,\text{eV}$. Then again, we can look along $\mathbf{k} = (k, k, k)$, and obtain $E(k) = E_0 + 6t\cos(ka/2) + 2t\cos(3ka/2)$, which is plotted from right to left in the left panel of Fig. 2.4 for $k \in [0, \pi/a]$. Each Na atom contributes one valence electron per unit cell, and since each spin-degenerate band can accommodate two electrons per cell according to Sec. 2.1.4, the band is half-occupied. The horizontal line at $E = 0$ in Fig. 2.4 indicates the approximate position of the Fermi level.

This model gives a qualitatively correct description of the Fermi surface, which is roughly spherical, filling half the volume of the BZ. However, it overestimates the departures from sphericity and fails to give the correct energy and degeneracy for states at the BZ boundary. These features can be improved by including more atomic orbitals in the basis; for example, the addition of p and d orbitals on each site would give a nine-band model whose lowest band would resemble that in Fig. 2.4, but with a dispersion in better agreement with experiment.[9]

As a second example we consider a tight-binding model of the highest occupied and lowest unoccupied bands of polyacetylene, which is the planar 1D polymer composed simply of CH units as shown in the inset of Fig. 2.5. Because carbon prefers to be fourfold coordinated, the chain adopts a dimerized structure of alternating "double" and "single" bonds as sketched in the figure. In physical terms, this translates to "short" and "long" C-C bonds. Since the entire molecule has M_z mirror symmetry, the Bloch bands cleanly separate into those arising from states of even M_z symmetry (principally H

[9] Actually, the nearly-free electron model, described in any solid-state physics text, gives a more accurate and concise description of the energy bands of alkali metals such as Li. We have focused on the TB approach here for pedagogical purposes.

s, C s, and C p_x and p_y) and those of odd M_z symmetry (mainly C p_z). The states of even M_z symmetry form Bloch bands corresponding to covalent C-C and C-H bonds, with a large band gap between the lower-energy σ bonding and higher-energy σ-antibonding bands. On the other hand, the π manifold made out of the p_z orbitals provide states that are nearly metallic, as we shall see.

We therefore concentrate on constructing a TB model only for these p_z states. We assign nearest-neighbor hopping strengths $t_\mathrm{s} = t + \delta$ and $t_\mathrm{l} = t - \delta$ to the short and long bonds respectively, where δ reflects the difference in hopping strength arising from the dimerization. We also allow the assignment of different site energies $E_1 = E_p + \Delta$ and $E_2 = E_p - \Delta$ to the two C p_z orbitals, where Δ is a measure of the asymmetry between the two C sites. This is not chemically motivated – we would have $\Delta = 0$ for ordinary polyacetylene – but a model that includes the asymmetry Δ will prove to be useful to us later. (You can think of this asymmetry as caused by the application of an electric field along the y direction if you like.)

We take $\mathbf{a}_1 = a\hat{x}$ and assume the C atoms in the home unit cell are located at positions $\boldsymbol{\tau}_1 = (-a/4, b/2, 0)$ and $\boldsymbol{\tau}_2 = (a/4, -b/2, 0)$. (Note that we have neglected the dimerization in setting the atomic positions, assuming for the sake of simplicity that it is already adequately represented by the δ term in the Hamiltonian.) There are two orbitals per cell, so that $H(\mathbf{R})$ and $H_\mathbf{k}$ now become 2×2 matrices, and we find

$$H(\mathbf{0}) = \begin{pmatrix} E_p + \Delta & t + \delta \\ t + \delta & E_p - \Delta \end{pmatrix},$$

$$H(\mathbf{a}_1) = \begin{pmatrix} 0 & 0 \\ t - \delta & 0 \end{pmatrix},$$

$$H(-\mathbf{a}_1) = \begin{pmatrix} 0 & t - \delta \\ 0 & 0 \end{pmatrix}.$$

We then evaluate Eq. (2.58) to get $H_{ij}(k)$ in the 1D k-space. For the off-diagonal elements, we use that the $(t_j - t_i)_x$ separation in Eq. (2.58) is $a/2$ for the $t(1+\delta)$ term in $H(\mathbf{0})$ and $-a/2$ for the $t(1+\delta)$ term in $H(-\mathbf{a}_1)$, so that $H_{12}(k) = 2t\cos(ka/2) + 2i\delta\sin(ka/2)$ and H^k_{21} is the complex conjugate of this. We thus get

$$H(k) = \begin{pmatrix} E_p + \Delta & 2t\cos(ka/2) + 2i\delta\sin(ka/2) \\ \text{c.c.} & E_p - \Delta \end{pmatrix}. \tag{2.71}$$

Before plugging into the secular equation of Eq. (2.60), it is instructive to

rewrite this in terms of the 2×2 identity and Pauli matrices

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.72}$$

in the canonical form

$$H = f_0\, I + f_x\, \sigma_x + f_x\, \sigma_y + f_x\, \sigma_z = f_0\, I + \mathbf{f} \cdot \boldsymbol{\sigma}\,. \tag{2.73}$$

An elementary calculation shows that $\det(H - E) = 0$ yields eigenvalues $E = f_0 \pm \sqrt{f_x^2 + f_y^2 + f_z^2}$. In our case we have $f_0 = E_p$, $f_x(k) = 2t\cos(ka/2)$, $f_y(k) = -2\delta\sin(ka/2)$, and $f_z = \Delta$, so that

$$E(k) = E_p \pm \sqrt{4t^2\cos^2(ka/2) + \Delta^2 + 4\delta^2\sin(ka/2)}\,. \tag{2.74}$$

The first term in the square root is the dominant one, since we have in mind that Δ (the site asymmetry) is small or zero, and δ (the bond asymmetry) is weaker than the bond itself. At the zone boundary $k = \pi/a$, however, it vanishes, and $E(\pi/a) = E_p \pm \sqrt{\Delta^2 + 4\delta^2}$. An electron counting argument shows that this band subspace is half-filled in $(\mathrm{CH})_x$, so it follows that the system is an insulator with a global gap only if Δ and/or δ is nonzero. A numerical solution for the case of $E_p = -6.0\,\mathrm{eV}$, $t = -2.8\,\mathrm{eV}$, $\Delta = 0$, and $\delta = -0.2\,\mathrm{eV}$ is shown in Fig. 2.5.

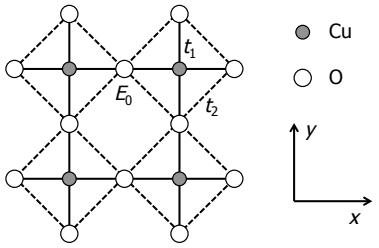
Once again, this simple tight-binding model has its limitations. A more careful treatment including non-orthogonality or further-neighbor overlaps would modify the shape of these bands slightly, and more importantly, the inclusion of σ bonding and antibonding orbitals of even mirror symmetry would give additional occupied and unoccupied bands that overlap to some degree with the two π bands. Nevertheless, the model provides a good qualitative and semiquantitative description of the low-energy excitations controlled by the gap and the valence and conduction band edges, and is the foundation for the well-known model of Su et al. (1979), which also includes electron-phonon interactions to describe the spontaneous dimerization and the formation of defects in this interesting system.

Exercises

Exercise 2.2.1 Take the program `benzene.py` on p. 159 of App. A and modify it by considering the presence of a uniform electric field along $\hat{x}$ or $\hat{y}$ (your choice), by letting the site energies be raised or lowered in a way that is linear in their spatial coordinates. Make a plot of the six eigenvalues versus the field strength. Does the high-field behavior make sense? Explain why.

Exercise 2.2.2 Same as Ex. 2.2.1, but instead of an external field, augment the model by adding an alternation of the bond strengh such that bonds 0-1, 2-3 and 4-5 have strength $t + \delta$ while bonds 1-2, 3-4 and 5-0 have strength $t - \delta$. Make a plot of the six eigenvalues versus δ. Does anything special happen when $\delta = t$? Explain why.

Exercise 2.2.3 High-T_c superconductors contain two-dimensional planes of Cu and O orbitals arranged in a square lattice (3 orbitals/cell) as shown:



(The orbitals are really Cu $3d_{x^2-y^2}$ and oxygen $2p_z$ orbitals, but please ignore these details, and just treat them as illustrated in the figure.) Let $E_{\mathrm{Cu}} = 0$ by convention, and let the parameters of the model be the oxygen site energy E_{O}, the nearest-neighbor Cu–O hopping t_1, and the further-neighbor O–O hopping t_2. Write a Python program to calculate the bands for this model. Show your program and a sample output for some parameter values chosen by you (realistically, t_1 should be quite a bit larger than the other parameters).

2.3 Linear response theory

Before concluding this chapter, we briefly introduce linear response theory. This is essentially just a slight extension of ordinary first-order perturbation theory, but is usually omitted in elementary quantum mechanics texts.

We start with ordinary perturbation theory as it applies to a state $|n(\lambda)\rangle$ satisfying $H(\lambda)|n(\lambda)\rangle = E_n(\lambda)|n(\lambda)\rangle$ for a Hamiltonian depending smoothly on a parameter λ. Hereafter we drop the explicit λ dependence and write $(E_n - H)|n\rangle = 0$. Taking the first derivative $\partial_\lambda = d/d\lambda$ of this expression and keeping terms at first order in the perturbation yields

$$(E_n - H)|\partial_\lambda n\rangle = \partial_\lambda (H - E_n)\,|n\rangle\,. \tag{2.75}$$

where $\partial_\lambda H$ represents the perturbation (often written as V or H_1 in quantum mechanics texts). But $\partial_\lambda E_n = \langle n|\partial_\lambda H|n\rangle$, so we can write this as

$$(E_n - H)|\partial_\lambda n\rangle = \mathcal{Q}_n\,(\partial_\lambda H)\,|n\rangle \tag{2.76}$$

where

$$\mathcal{Q}_n = \sum_{m \neq n} |m\rangle\langle m| = 1 - \mathcal{P}_n \tag{2.77}$$

is the projection operator onto all states other than state n, i.e., the complement of the projector $\mathcal{P}_n = |n\rangle\langle n|$.

Equation (2.76), sometimes known as a *Sternheimer equation*, is an inhomogeneous linear equation for the perturbed wavefunction $|\partial_\lambda n\rangle$. Its solution is

$$|\partial_\lambda n\rangle = -iA_n|n\rangle + \sum_{m \neq n} \frac{|m\rangle\langle m|}{E_n - E_m} (\partial_\lambda H)\,|n\rangle \tag{2.78}$$

for any real A_n, as can be checked by plugging into Eq. (2.76) and multiplying on the left by an arbitrary eigenstate $\langle m|$. This exercise shows that $A_n = i\langle n|\partial_\lambda n\rangle$, an object known as a "Berry connection" that we shall come to know well in the next chapter. One way to think about this term is to observe what happens if H actually has no dependence on λ at all; then the last term in Eq. (2.78) vanishes and we find a solution like $|n(\lambda)\rangle = e^{-iA_n\lambda}|n_0\rangle$. In this case the *physical state* does not change with λ, but its *representation as a vector* changes because of a phase rotation in the definition of $|n\rangle$ as a function of λ. Because it does not represent a physical change of state, this term is dropped in most elementary quantum texts, and we shall do so here as well. However, to clarify that we have done so, we include a $\mathcal{Q}_n$ factor explicitly in rewriting Eq. (2.78) as

$$\mathcal{Q}_n\,|\partial_\lambda n\rangle = \sum_{m \neq n} \frac{|m\rangle\langle m|}{E_n - E_m} (\partial_\lambda H)\,|n\rangle \tag{2.79}$$

We can simplify the notation further by introducing the definition

$$T_n = \sum_{m \neq n} \frac{|m\rangle\langle m|}{E_n - E_m}\,, \tag{2.80}$$

in which case we arrive at a very concise expression for the perturbed wavefunction, namely

$$\mathcal{Q}_n|\partial_\lambda n\rangle = T_n\,(\partial_\lambda H)\,|n\rangle\,. \tag{2.81}$$

In the linear-response theory we also consider how the expectation value of some Hermitian observable $\mathcal{O}$ changes as a result of the perturbation. We

have

$$\begin{aligned}\partial_\lambda \langle \mathcal{O} \rangle_n &= \partial_\lambda \langle n|\mathcal{O}|n\rangle \\ &= \langle \partial_\lambda n|\mathcal{O}|n\rangle + \langle n|\mathcal{O}|\partial_\lambda n\rangle \\ &= 2\mathrm{Re}\,\langle n|\mathcal{O}|\partial_\lambda n\rangle \\ &= 2\mathrm{Re}\,\langle n|\mathcal{O}\mathcal{Q}_n|\partial_\lambda n\rangle \end{aligned} \tag{2.82}$$

where the last equality follows because the contribution from the first term in Eq. (2.78) is pure imaginary. If we wish we can combine this with Eq. (2.81) to get

$$\partial_\lambda \langle \mathcal{O} \rangle_n = 2\mathrm{Re}\,\langle n|\mathcal{O}\,T_n\,(\partial_\lambda H)|n\rangle \tag{2.83}$$

at the expense of introducing the sum over states encoded in T_n. Both Eqs. (2.82) and (2.83) are valid and equivalent expressions.

Two comments are in order. First, in the context of an independent-particle treatment of an electron system such as a molecule, the expression

$$\partial_\lambda \langle \mathcal{O} \rangle = 2\mathrm{Re} \sum_n^{\mathrm{occ}} \langle n|\mathcal{O}\mathcal{Q}_n|\partial_\lambda n\rangle \tag{2.84}$$

for the total change to $\langle \mathcal{O} \rangle$ coming from all occupied states can be replaced by the more convenient expression

$$\partial_\lambda \langle \mathcal{O} \rangle = 2\mathrm{Re} \sum_n^{\mathrm{occ}} \langle n|\mathcal{O}\mathcal{Q}|\partial_\lambda n\rangle\,, \tag{2.85}$$

where

$$\mathcal{Q} = \sum_m^{\mathrm{unocc}} |m\rangle\langle m| = 1 - \sum_n^{\mathrm{occ}} |n\rangle\langle n| \tag{2.86}$$

is the projection onto the unoccupied-state manifold. This follows because

$$\begin{aligned} 2\mathrm{Re}\,\textstyle\sum_n^{\mathrm{occ}} \langle n|\,\mathcal{O}(\mathcal{Q}_n - \mathcal{Q})|\partial_\lambda n\rangle &= 2\mathrm{Re} \sum_{n\neq n'}^{\mathrm{occ}} \langle n|\mathcal{O}|n'\rangle\langle n'|\partial_\lambda n\rangle \\ &= \sum_{n\neq n'} \Big(\langle n|\mathcal{O}|n'\rangle\langle n'|\partial_\lambda n\rangle + \langle n'|\mathcal{O}|n\rangle\langle \partial_\lambda n|n'\rangle \Big) \\ &= \sum_{n\neq n'} \Big(\langle n|\mathcal{O}|n'\rangle\langle n'|\partial_\lambda n\rangle - \langle n'|\mathcal{O}|n\rangle\langle n|\partial_\lambda n'\rangle \Big) \\ &= 0 \end{aligned} \tag{2.87}$$

where the second line is obtained from $\mathrm{Re}\,(z) = z + z^*$, the third comes from $\partial_\lambda \langle n'|n\rangle = \partial_\lambda \delta_{nn'} = 0$, and the last follows from the interchange of dummy labels n and n' in the last term. Intuitively, insofar as the λ perturbation induces only a unitary rotation between occupied states $|n\rangle$ and $|n'\rangle$, this has

no influence on the overall $\langle\mathcal{O}\rangle$; it is only when the perturbation induces an occupied state $|n\rangle$ to acquire character on an unoccupied state $|m\rangle$ that $\langle\mathcal{O}\rangle$ is perturbed. As a result, we do not really care about $\mathcal{Q}_n|\partial_\lambda n\rangle$; it is enough to calculate $\mathcal{Q}|\partial_\lambda n\rangle$, which is the component of the perturbed wavefunction $|n\rangle$ that projects into the unoccuped space; this is the only piece that enters the linear response theory. It is also convenient to replace Eq. (2.80) by a revised definition

$$T_n = \sum_m^{\text{unocc}} \frac{|m\rangle\langle m|}{E_n - E_m}\,, \tag{2.88}$$

which we will take as the definition of T_n henceforth. Then

$$\mathcal{Q}|\partial_\lambda n\rangle = T_n\,(\partial_\lambda H)\,|n\rangle\,. \tag{2.89}$$

are the first-order wave-function perturbations that are needed for insertion into Eq. (2.85).[10]

Second, while $\mathcal{Q}$ is easily expressed as a finite sum via $\mathcal{Q} = 1 - \sum_m^{\text{occ}} |m\rangle\langle m|$, the sum over all unoccupied state is the formulas involving T_n is often problematic. One would have to truncate the infinite sum to a finite number of terms, and the resulting calculations are tedious and require careful convergence tests. Fortunately there is an alternative approach, in which the Sternheimer equation of Eq. (2.76), now rewritten as $(E_n - H)|\partial_\lambda n\rangle = \mathcal{Q}\,(\partial_\lambda H)\,|n\rangle$, is solved directly by an iterative algorithm for each occupied state n. The resulting vectors $|\partial_\lambda n\rangle$ are then inserted into Eq. (2.85) to find the change in the expectation of operator $\mathcal{O}$.

Frequently one considers two perturbations corresponding to two operators $A = \partial H/\partial\lambda_A$ and $B = \partial H/\partial\lambda_B$, or equivalently, $H = H_0 + \lambda_A A + \lambda_B B$. Then

$$\frac{\partial^2 E}{\partial\lambda_A\,\partial\lambda_B} = \frac{\partial\langle A\rangle}{\partial\lambda_B} = \frac{\partial\langle B\rangle}{\partial\lambda_A} = \sum_n^{\text{occ}} 2\text{Re}\,\langle n|AT_nB|n\rangle\,. \tag{2.90}$$

For example, λ_A and λ_B could refer to two different atomic displacements in a molecule, in which case Eq. (2.90) yields the force-constant matrix connecting these two displacements, as would be needed for a calculation of the vibrational frequencies. Written explicitly, Eq. (2.90) takes the probably more familiar form

$$\frac{\partial^2 E}{\partial\lambda_A\,\partial\lambda_B} = \sum_n^{\text{occ}} \sum_m^{\text{unocc}} \frac{\langle n|A|m\rangle\langle m|B|n\rangle}{E_n - E_m} \;+\; \text{c.c.} \tag{2.91}$$

[10] In an equivalent formulation, the perturbation of $\langle\mathcal{O}\rangle$ can alternatively be expressed in terms of the perturbation of the one-particle density matrix $\rho = \sum_n^{\text{occ}} |n\rangle\langle n|$, i.e., $\partial_\lambda\langle\mathcal{O}\rangle = \text{Tr}\,[\mathcal{O}\partial_\lambda\rho]$; see, e.g., Thouless (1983). A formulation in terms of Greens functions, which essentially embody the action of Eq. (2.88), is also equivalent and commonly used.

where 'c.c.' indicates the complex conjugate. For $A = B$, the numerator becomes $|\langle n|A|m\rangle|^2$ and Eq. (2.91) reduces to the well-known result of standard second-order perturbation theory. In practice, the infinite sum over unoccupied states is usually avoided by replacing either $T_n B|n\rangle$ by $\mathcal{Q}|\partial n/\partial\lambda_B\rangle$ or $\langle n|AT_n$ by $\langle\partial n/\partial\lambda_A|\mathcal{Q}$, where the iterative Sternheimer approach is used to obtain the needed wavefunction derivative.

The above equations form the basis of linear-response theory as it is implemented in many common code packages, such as ABINT or QUANTUM ESPRESSO in the solid-state community. The above equations are restricted to the case of a single-particle Hamiltonian; when implemented in the context of DFT, additional terms corresponding to the self-consistently induced variation of the Hartree and exchange-correlation potentials also need to be included. For further details on the formalism and computational implementation of this approach, known as "density functional perturbation theory," the reader is referred to the excellent review by Baroni et al. (2001) or the papers of Gonze (1997) and Gonze and Lee (1997). The formalism can also be generalized to describe dynamical response functions, i.e., the ratio of a response $A(\omega)$ to a perturbation $B(\omega)$ at frequency ω, with the energy denominators modified to become $1/(E_n - E_n \pm \hbar\omega)$.

Exercises

Exercise 2.3.1 Something