

The Tight-Binding Method*

Introduction

One way of computing the band structure of a solid is to view it as a gas of nearly free conduction electrons, only weakly perturbed by the periodic potential of the ion cores. That sort of approximation then leads to the “free electron gas” or “nearly free electron gas” treatments. A different method is to treat the solid in terms of the covalent bonds between the ion cores. In that approximation, we use the atomic orbitals (wave functions) at the various lattice sites as a starting point, we then construct basis functions that are a linear combination of these atomic orbitals, and finally we solve the Schrödinger equation using these basis functions. This type of solution invariably makes some restriction on the set of atomic orbitals that is used (e.g., just s -orbitals on atoms, or s - and p -orbitals, or whatever is appropriate for the problem at hand). For a given set of orbitals, then a limited number of bands within the band structure are obtained. In addition to restricting the starting set of atomic orbitals, another approximation is to assume that these functions are localized predominantly on only a central atom being considered as well as some small number of neighboring atoms. For example, just nearest-neighbor atoms (or perhaps nearest-neighbor and second-nearest-neighbor atoms) might be included. In many situations, these sorts of restrictions on the set of atomic orbitals and the number of neighboring atoms that are included can nevertheless lead, at least qualitatively, to shapes of the computed bands that are in good agreement with what is known from experiment (or known from much more sophisticated computations).

In these notes we will give a general description of the tight-binding method, although we will write it only for a lattice in 1 dimension. This restriction only slightly simplifies the formalism that is described here, but when applied to specific examples the restriction considerably simplifies the computations (since it is straightforward to think about what the neighbors are and how orbitals on neighboring atoms overlap with each other). Nevertheless, considering just 1-dimensional examples still allows us to apply the method to some important cases of physical interest. We note that, even with this restriction to a 1-dimensional lattice, the theory is still somewhat complicated due to the various types of situations that are dealt with. Hence, we start our discussion by introducing these various situations one at a time, so as to build up a general sort of 1-dimensional theory.

Since we want to form electronic states of the solid utilizing states associated with the atoms, we first consider their atomic states. For example, we if have just *single* atom located within each unit cell of the solid, then we consider just the atomic orbitals (wave functions) of that atom, as obtained from

$$H_{\text{at}} \phi_n(\mathbf{r}) = E_n \phi_n(\mathbf{r}) \quad (1)$$

* The formalism in these notes is based upon lectures in a solid-state physics class that the author attended during 1978-1979 at the California Institute of Technology, taught by Prof. D. L. Smith.

where H_{at} is a Hamiltonian for the atom, having energy eigenvalues E_n and wave functions $\phi_n(\mathbf{r})$. As mentioned above, we restrict our interest to only a small set of such orbitals, let's say $n = 1, 2, \dots, n_a$ of them with n_a being a small integer (such as 1 or 2). In a 1-dimensional lattice, we have unit cells located at $R_j = ja$ with a being the lattice constant and j an integer, $-N/2 < j \leq N/2$ for N atoms in the lattice. For a single atom per unit cell, then the locations of those atoms in the solid can be taken as identical to the locations of the unit cells.

However, more generally, we will be interested in handling problems in which there might be more than one type of atom per unit cell. We refer to these atoms as a "basis", which when periodically repeated, forms the entire lattice. We must add an additional subscript to refer to *each* of these atoms in the basis. We use Greek subscripts such as α or γ to label the basis atoms, with their location within the unit cell denoted by b_α , $\alpha = 1, 2, \dots, n_b$. Considering a unit cell located at the origin of our coordinate system ($R_j = 0$), then each of these basis atoms has its own potential $V_\alpha(\mathbf{r} - b_\alpha)$ associated with it, with associated orbitals $\phi_n(\mathbf{r} - b_\alpha)$. Thus, generalizing Eq. (1) to this situation, we have

$$H_{\text{at},\alpha} \phi_n(\mathbf{r} - b_\alpha) = E_n \phi_n(\mathbf{r} - b_\alpha) \quad (2a)$$

with atomic Hamiltonian for the atom type labeled by α ,

$$H_{\text{at},\alpha} = \frac{-\hbar^2}{2m_e} \nabla^2 + V(\mathbf{r} - b_\alpha) . \quad (2b)$$

Finally, a further generalization of the problem is to consider having different *types* of atoms in the unit cell (with these atoms necessarily having different locations). These different types of atoms will have different potential and atomic orbitals, so we require a subscript α not only for the location of the atom but also for its potential, its wave functions, and the associated eigenvalues,

$$H_{\text{at},\alpha} \phi_{n,\alpha}(\mathbf{r} - b_\alpha) = E_{n,\alpha} \phi_{n,\alpha}(\mathbf{r} - b_\alpha) \quad (3a)$$

with atomic Hamiltonian

$$H_{\text{at},\alpha} = \frac{-\hbar^2}{2m_e} \nabla^2 + V_\alpha(\mathbf{r} - b_\alpha) . \quad (3b)$$

Here, the atom located at b_α has a given type (labeled by α), so it has a potential $V_\alpha(\mathbf{r} - b_\alpha)$ and there is a set of states (one state for each value of n) associated with the atom, with wave functions $\phi_{n,\alpha}(\mathbf{r} - b_\alpha)$ and eigenvalues $E_{n,\alpha}$.

Having defined our atomic states $\phi_{n,\alpha}(\mathbf{r} - b_\alpha)$ in this manner, we can now proceed to form a set of basis states of the entire lattice; we will then take linear combinations of the basis states in order to solve for the eigenstates for our problem. (Note that we're using the word "basis" in two ways, for the basis atoms in each unit cell and for the set of basis states that we use to construct complete eigenstates; the distinction between these two usages should be clear from the context). These basis states that we construct must have Bloch form, which for the 1-dimensional lattice that we are considering then leads to basis states of

$$\psi_{n,\alpha,k}(\mathbf{r}) = \frac{1}{N^{1/2}} \sum_j e^{ik(\mathbf{R}_j + \mathbf{b}_\alpha)} \phi_{n,\alpha}(\mathbf{r} - \mathbf{R}_j - \mathbf{b}_\alpha) . \quad (4)$$

(The Bloch nature of these states is easiest to see if we write them in a fully 3-dimensional form, using vectors \mathbf{k} , \mathbf{R}_j , and \mathbf{b}_α , so that

$$\psi_{n,\alpha,k}(\mathbf{r}) = \frac{1}{N^{1/2}} \sum_j e^{i\mathbf{k} \cdot (\mathbf{R}_j + \mathbf{b}_\alpha)} \phi_{n,\alpha}(\mathbf{r} - \mathbf{R}_j - \mathbf{b}_\alpha) \quad (5a)$$

$$= \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{N^{1/2}} \left[\sum_j e^{i\mathbf{k} \cdot (\mathbf{R}_j + \mathbf{b}_\alpha - \mathbf{r})} \phi_{n,\alpha}(\mathbf{r} - \mathbf{R}_j - \mathbf{b}_\alpha) \right] \quad (5b)$$

with the second line following immediately from the first, and where the periodic nature of the function within the square brackets of the second line establishes the Bloch nature of these basis functions for the lattice). From this set of basis states, a total wave function for the lattice is formed by using linear combinations

$$\sum_{n,\alpha} \chi_{n,\alpha}(k) \psi_{n,\alpha,k}(\mathbf{r}) \quad (6)$$

with the coefficients $\chi_{n,\alpha}(k)$ to be determined. Associated with each such solution there will be an energy eigenvalue, which we denote $\epsilon(k)$. The number of such solutions is equal to the number of atoms within the unit cell (n_b) times the number of atomic orbitals per atom (n_a) that we utilize (actually, we don't need to assume the same number of atomic orbitals for each atom in the unit cell, so some atoms might have fewer than n_a states on them, but anyway the maximum number of solutions is the product $n_a n_b$).

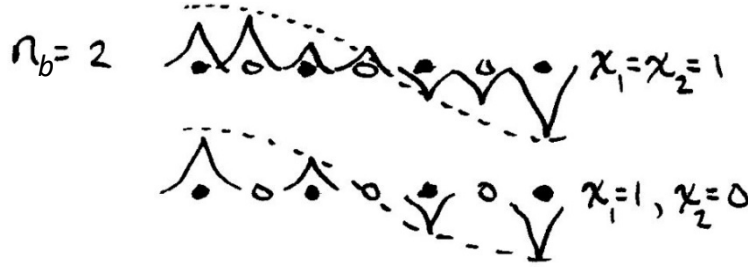
With n_a and n_b being small integers, then in the tight-binding method we often end up with quite small matrices, e.g. 1×1 for a single atom per unit cell and just one state per atom, or 2×2 if we either have two atoms per unit cell and a single state per atom, or one atom per unit cell and two states on that atom, etc. In all cases, we will treat that various overlap integrals involving wave functions and atomic potentials just as parameters in these matrices, with the main output from the analysis being the shape of bands that result for various assumptions concerning those parameters. Most importantly, for a given set of atomic orbitals, we can say something at the outset concerning the sign of the various matrix elements (as we will see in example computations), which then significantly constrains what the resulting bands look like.

Before describing the full solution for the electronic states of the solid, let us consider a few examples of the basis states of Eq. (5) and the resulting total wave functions of Eq. (6). For the simplest possible case of a single atom in each unit cell ($n_b = 1$), and one s -state on each atom ($n_a = 1$), the basis state can be schematically drawn as



where the solid dots here represent atoms in a 1-dimensional lattice, and the peaked features on each atom schematically show the profile of an s -state associated with each atom. These atomic states are then weighted by an $\exp(ikR_j)$ function, as in Eq. (4) (with $b_\alpha = 0$). In this case, the problem just reduces to a 1×1 matrix (there are no coefficients $\chi_{n,\alpha}(k)$ as in Eq. (6) to be solved for), so the basis state pictured here is then the complete eigenstate of our 1-dimensional solid. However, there is still a nontrivial eigenvalue function $\epsilon(k)$ that is obtained.

A more complicated situation occurs if we have two atoms in the per unit cell, $n_b = 2$. Then, we have basis states analogous to the one above (for $n_b = 1$) but now for *each* of the atoms in the unit cell. These basis states are combined to form a total wave function, as in Eq. (6). Two possible combinations that form totals wave function are



Here we are picturing two types of atoms per unit cell ($n_b = 2$), shown by filled and open dots, with an s -state on each atom ($n_a = 1$). Two resultant total wave functions are shown, one with $\chi_1 = \chi_2 = 1$, and the other with $\chi_1 = 1$ and $\chi_2 = 0$ (since we have just one type of atomic state here, then the n subscript of $\chi_{n,\alpha}$ in Eq. (6) is not used, so we are discussing χ_α with the subscript referring to the location of the atoms within the unit cell). These values of χ_α arise from a detailed solution of the problem, for a particular value of k . Since we have two atoms in the unit cell, then we have a 2×2 problem that we end up solving, leading to 2 bands in the electronic band structure.

Formalism

In order to solve for the coefficients $\chi_{n,\alpha}(k)$ in Eq. (6) and also obtain the associated $\epsilon(k)$ eigenvalues, we must solve the Schrödinger equation for the entire solid, with potential of

$$V(\mathbf{r}) = \sum_{j,\alpha} V_\alpha(\mathbf{r} - R_j - b_\alpha) \quad (7a)$$

and Hamiltonian

$$H = \frac{-\hbar^2}{2m_e} \nabla^2 + V(\mathbf{r}) \quad (7b)$$

(we are neglecting electron-electron interactions entirely, treating this problem just in a single-particle viewpoint). We use the restricted set of basis functions formed in Eq. (5), with the number of such functions being $n_b n_a$, the number of atoms within each unit cell times the number of

atomic orbitals per atom. We assume that all atomic orbitals are orthogonal (not only the ones on the same atom, but also ones on different atoms). This assumption then leads to orthogonality of the basis functions of Eq. (5),

$$\langle \psi_{n\alpha k} | \psi_{n'\alpha' k'} \rangle = \delta_{nn'} \delta_{\alpha\alpha'} \delta_{kk'} . \quad (8)$$

With this result, then the matrix elements of H between the basis functions turn out to be

$$\begin{aligned} \langle \psi_{n\alpha k} | H | \psi_{n'\alpha' k'} \rangle &= \frac{1}{N} \sum_{ij} e^{i[k'(R_i + b_{\alpha'}) - k(R_j + b_{\alpha})]} \langle \phi_{n\alpha}(\mathbf{r} - R_j - b_{\alpha}) | H | \phi_{n'\alpha'}(\mathbf{r} - R_i - b_{\alpha'}) \rangle \\ &= \frac{1}{N} \sum_i e^{i(k' - k)R_i} \sum_R e^{ik'b_{\alpha'}} e^{ik(R - b_{\alpha})} M_{n\alpha; n'\alpha'}(R) \\ &= \delta_{kk'} \sum_R e^{ik(R + b_{\alpha} - b_{\alpha'})} M_{n\alpha; n'\alpha'}(R) \end{aligned} \quad (9a)$$

with

$$M_{n\alpha; n'\alpha'}(R) = \langle \phi_{n\alpha}(\mathbf{r} - b_{\alpha}) | H | \phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'}) \rangle . \quad (9b)$$

Note that in Eq. (9a), whereas on the first line we have a double sum with both sums ranging over all of the unit cells in the solid, in the final line we have just a single such sum (the sum over R or the sums over R_i or R_j all denote the same thing – they are sums over all unit cells in the lattice – although mathematically, going from the first to the second line of Eq. (9a) involves the new variable $R \equiv R_i - R_j$). Furthermore, the integral of Eq. (9b) contains an atomic orbital located in the unit cell that is at the origin of the coordinate system, along with an orbital located in a unit cell that is a specified distance, R , away. This form permits us to truncate the sum over R in the final line of Eq. (9a) such that it only extends over some chosen number of neighboring atoms.

Let us examine the matrix elements $M_{n\alpha; n'\alpha'}(R)$. Inserting the Hamiltonian of Eq. (7b), we have

$$M_{n\alpha; n'\alpha'}(R) = \left\langle \phi_{n\alpha}(\mathbf{r} - b_{\alpha}) \left| \frac{-\hbar^2}{2m_e} \nabla^2 + \sum_{j,\gamma} V_{\gamma}(\mathbf{r} - R_j - b_{\alpha}) \right| \phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'}) \right\rangle . \quad (10)$$

For large R we expect $M(R)$ to be small, since the orbitals in the matrix element are far separated. Further considering this dependence on R we recognize that there are two sorts of terms that we must deal with:

(i) $R = 0$ and $\alpha = \alpha'$: In this case we have orbitals on the same atom, and

$$\begin{aligned} M_{n\alpha; n'\alpha}(0) &= \left\langle \phi_{n\alpha}(\mathbf{r} - b_{\alpha}) \left| \frac{-\hbar^2}{2m_e} \nabla^2 + V_{\alpha}(\mathbf{r} - b_{\alpha}) \right| \phi_{n'\alpha}(\mathbf{r} - b_{\alpha}) \right\rangle \\ &\quad + \sum'_{j,\gamma} \left\langle \phi_{n\alpha}(\mathbf{r} - b_{\alpha}) \left| V_{\gamma}(\mathbf{r} - R_j - b_{\gamma}) \right| \phi_{n'\alpha}(\mathbf{r} - b_{\alpha}) \right\rangle \end{aligned} \quad (11)$$

with the prime symbol on the sum over j, γ indicating that the term with $j = 0$ and $\gamma = \alpha$ should not be included, since it has already been treated in the first term on the right-hand side. Note that

that first term is simply equal to an atomic energy when $n = n'$, or it is 0 otherwise. Thus, this first term equals $E_{n\alpha}\delta_{nn'}$. For the second term on the right-hand side, it involves two orbitals on the same site, with a potential term on some other site (the other site either being in a different unit cell, or on a different basis atom, or both). When $n = n'$, this second term produces a small shift in the atomic energy (i.e. this shift arising from the potential of neighboring atoms), whereas when $n \neq n'$ the term still produces a small, nonzero energy although this energy is not a shift of an atomic eigenvalue. We denote the values of this second term by $V_{nn'\alpha}$, and with that we have

$$M_{n\alpha;n'\alpha'}(0) = [E_{n\alpha}\delta_{nn'} + V_{nn'\alpha}]\delta_{\alpha\alpha'} \quad (12)$$

(ii) $R \neq 0$ or $\alpha \neq \alpha'$ (or both): In this case we have integrals involving orbitals from different atoms, with $M_{n\alpha;n'\alpha'}(R)$ given by

$$\left\langle \phi_{n\alpha}(\mathbf{r} - b_\alpha) \left| \frac{-\hbar^2}{2m_e} \nabla^2 + V_\alpha(\mathbf{r} - b_\alpha) + \sum'_{j,\gamma} V_\gamma(\mathbf{r} - R_j - b_\gamma) \right| \phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'}) \right\rangle \quad (13)$$

where the prime symbol on the sum over j, γ indicates that the term with $j = 0$ and $\gamma = \alpha$ should not be included (since it is included in the term before), and also the final $\phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'})$ wave function in this integral should not have both $R = 0$ and $\alpha' = \alpha$ (since otherwise we would be dealing with situation (i) above). Now, a portion of Eq. (13) can be evaluated as

$$\langle \phi_{n\alpha}(\mathbf{r} - b_\alpha) | (-\hbar^2/2m_e)\nabla^2 + V_\alpha(\mathbf{r} - b_\alpha) = E_{n\alpha} \langle \phi_{n\alpha}(\mathbf{r} - b_\alpha) | \quad (14)$$

(since $(-\hbar^2/2m_e)\nabla^2 + V_\alpha(\mathbf{r} - b_\alpha)$ is a Hermitian operator, and it can be thought of as operating to the left, thereby producing the atomic energy $E_{n\alpha}$). However, when we also include the $|\phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'})\rangle$ on the right-hand side of Eq. (14) we then get 0 since there is zero overlap between $\phi_{n\alpha}(\mathbf{r} - b_\alpha)$ and $\phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'})$ when $R \neq 0$ or $\alpha \neq \alpha'$ (or both). Thus, we arrive at

$$M_{n\alpha;n'\alpha'}(R) = \langle \phi_{n\alpha}(\mathbf{r} - b_\alpha) | \sum'_{j,\gamma} V_\gamma(\mathbf{r} - R_j - b_\gamma) | \phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'}) \rangle \quad (15)$$

were, again, the prime symbol on the sum over j, γ indicates that the term with $j = 0$ and $\gamma = \alpha$ should not be included, and the final $\phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'})$ wave function should not have both $R = 0$ and $\alpha' = \alpha$. We are therefore considering *both* the potential and the right-hand atomic orbital to be centered at different locations than for the left-hand atomic orbital. This situation can be pictured as



where we are showing two locations of atoms, each with an s -orbital on them, and then a potential at some other, far away, location. For the matrix overlap integral of Eq. (15), the largest result will be obtained when the potential is centered at the atom location α' , corresponding to



We define a new overlap matrix element to refer specifically to this case,

$$W_{n\alpha; n'\alpha'}(R) \equiv \langle \phi_{n\alpha}(\mathbf{r} - b_\alpha) | V_{\alpha'}(\mathbf{r} - R_j - b_{\alpha'}) | \phi_{n'\alpha'}(\mathbf{r} - R - b_{\alpha'}) \rangle \quad (16)$$

and we thus approximate Eq. (15) as

$$M_{n\alpha; n'\alpha'}(R) \approx W_{n\alpha; n'\alpha'}(R) . \quad (17)$$

Therefore, for the matrix elements of the Hamiltonian, Eq. (9a), we have (taking $k = k'$)

$$\langle \psi_{n\alpha k} | H | \psi_{n'\alpha' k} \rangle \approx [E_{n\alpha} \delta_{nn'} + V_{nn'\alpha}] \delta_{\alpha\alpha'} + \sum'_R e^{ik(R+b_\alpha-b_{\alpha'})} W_{n\alpha; n'\alpha'}(R) \quad (18)$$

where the sum over R does not include the $R = 0$ term if $\alpha = \alpha'$. Evaluating the terms on the right-hand side (maintaining all the various parameters, but determining the sign of the $W_{n\alpha; n'\alpha'}(R)$ where possible) then yields a matrix that can be diagonalized in order to determine the energy eigenvalues (i.e. the band structure) and the associated eigenvectors (which provide the total wave function).