Chapter 15

Quantitative Methods for Bandstructure Calculations

15.1 Introduction

In this chapter, we discuss a non-perturbative or *exactly* solvable model of electron bandstructure in a crystal. It is the celebrated Kronig-Penney model. The purpose of the solution is to illustrate much of bandstructure physics and also to develop a bag of useful concepts that permeate much of solid state physics.

15.2 Exact Solution: The Kronig-Penney Model

An exactly solvable periodic potential problem in quantum mechanics for the electron is the Kronig-Penney model. The problem is exactly solvable in all dimensions - we consider the 1D case. The periodic potential is modeled as a series of Dirac-delta functions

$$V(x) = \sum_{n} S\delta(x - na), \qquad (15.1)$$

where a is the lattice constant, and S is the strength of the perturbation. The sum over n runs over all lattice sites. For example, for a 1D closed ring of length L with N = L/a lattice points and lattice constant $a, 0 \le n \le N - 1$. This is schematically represented in Figure 15.1.



FIGURE 15.1: The Kronig-Penney "Dirac" comb periodic potential for a particle on a ring. Left: positive or repulsive potential for S > 0, and Right: Negative or attractive potentials for S < 0.

Now we in the Math primer (**cite**), we have seen the identity $\sum_n \delta(x - na) = \sum_n \frac{1}{a} e^{-i\frac{2\pi}{a}nx}$. Using this relation with $G_n = \frac{2\pi}{a}n$ and substituting in the Schrödinger equation, we get

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{S}{a}\sum_{n}e^{-iG_nx}\right]\psi = E\psi.$$
(15.2)

The wavefunctions are Bloch functions, which are Fourier expanded in $G_m = \frac{2\pi}{a}m$ as

$$\psi_k(x) = e^{ikx} u_k(x) = e^{ikx} \sum_m u_{G_m} e^{iG_m x} = \sum_m u_{G_m} e^{i(k+G_m)x}, \quad (15.3)$$

Note that $\psi(x=0) = \sum_{G_m} u_{G_m}$, the sum of *all* the Bloch coefficients, in other words, as long as G's are reciprocal lattice vectors, $\sum_G u_G = \psi(0)$. Now substituting 15.3 in the Schrodinger equation, we get

$$\sum_{m} \frac{\hbar^2 (k+G_m)^2}{2m} u_{G_m} e^{iG_m x} + \frac{S}{a} \sum_{m} \sum_{n} u_{G_m} e^{i(G_m - G_n)x} = E_k \sum_{m} u_{G_m} e^{iG_m x}, \quad (15.4)$$

Multiplying by e^{-iGx} and integrating over all x, we use to identity $\int_0^L e^{i(G'-G)x} dx = L\delta_{G,G'}$ to get

$$\frac{\hbar^2 (k+G)^2}{2m} u_G + \frac{S}{a} \sum_n u_{G_n+G} = E_k u_G, \qquad (15.5)$$

Solving for u_G yields

$$u_G = \frac{S}{a} \frac{\sum_{G_n+G} u_G}{E_k - \frac{\hbar^2 (k+G)^2}{2m}},$$
(15.6)

Now for a very useful trick: summing both sides over G's cancels the u_G terms because $\sum_{G_n+G} u_G = \sum_G u_G = \psi(x=0)$, leaving us with the identity

$$1 = \frac{S}{a} \sum_{G} \frac{1}{E_k - \frac{\hbar^2 (k+G)^2}{2m}}.$$
(15.7)

This is a rather fancy way of writing unity! Note that this is an *exact* form of the solution of Schrodinger's equation for the periodic potential problem. Inverting it into the form

$$\frac{a}{S} = \sum_{G} \frac{1}{E_k - \frac{\hbar^2 (k+G)^2}{2m}},$$
(15.8)

we are in a position to investigate the aftermath of the solution in Equation 15.7.



FIGURE 15.2: A graphical solution scheme for the repulsive Kronig-Penney Dirac comb. Note that the lowest energy is larger than zero.

Figure 15.2 shows a graphical solution of Equation 15.8 plotted as a function of the energy E_k for two values of k. When the strength of the potential S > 0, a/S > 0, and is the constant shown in red in the Figure. The RHS is a complex function of energy E_k , with a number of poles located at $E_k = \frac{\hbar^2(k+G)^2}{2m}$, where the RHS diverges. There are several points of intersection - one of which is highlighted. The energies E_k corresponding to these intersection points are the only allowed eigenvalues for the

problem. There are several allowed eigenvalues: in fact, there are exactly N distinct eigenvalues corresponding to n = 0, 1, ..., N - 1 values of $G_n = \frac{2\pi}{a}n$.

If we turn the strength of the potential down by taking $S \to 0$, the red line corresponding to a/S goes off to $+\infty$, and the intersections of the RHS and LHS then are exactly at the N energies for which the RHS blows up. Clearly, these energy eigenvalues are at $E_k = \frac{\hbar^2(k+G)^2}{2m}$, and we have recovered the nearly free-electron model of the electron.

If the strength is made very large, ...

If the strength is made negative by letting S < 0, it is clear that the red line a/S < 0, and there is an energy intersection for energy that is *negative*, i.e., $E_k(min) < 0$. This is a "bound" state... or weakly mobile...



FIGURE 15.3: The solid lines show the bandstructure for repulsive (left) and attractive (right) Kronig-Penney potentials. The nearly-free electron bandstructure $E(k) = \frac{\hbar^2(k+G)^2}{2m}$ is shown as dashed lines. The allowed energy bands are indicated in gray along with the energy gaps.

Figure 15.3 shows the calculated energy bandstructures for S > 0 (left) and S < 0 (right). The axes are in units of $F = \frac{\hbar^2}{2m} \cdot (\frac{\pi}{a})^2$ for energy, and $\frac{2\pi}{a}$ for k. The solid lines in the figure represents several important features of any bandstructure in the presence of a non-zero periodic potential. This is superposed on the dashed line plot of bandstructure when the periodic potential is turned off (S = 0), but the electron wavefunction is still required to satisfy the lattice periodicity and symmetry, the 'nearly' free-electron (NFE) model with $E = \frac{\hbar^2(k+G)^2}{2m}$. Note that for a repulsive potential with S > 0, the Kronig-Penney bandstructure energies are *higher* than the NFE values at *all* values of k except at the Brillouin zone center and edges $k = 0, \pm \frac{\pi}{a}n$. The highest eigenvalue of each Kronig-Penney band is degenerate with the NFE eigenvalues of $E(k = n\frac{\pi}{a}) = n^2 \cdot F$, where n = 1, 2, ..., locating energy eigenvalues F, 9F, ... at $k = \pm \frac{\pi}{a}$ at the BZ edge, and 4F, 16F, ... at k = 0 as the maxima of the corresponding bands.

That the energy eigenvalues for S > 0 are higher (or equal to) than the NFE values is guaranteed by the Hellmann-Feynman theorem. The Hellmann-Feynman theorem states that the eigenvalues E_k of any Hamiltonian \hat{H} satisfy $\frac{\partial E_k}{\partial \lambda} = \langle k | \frac{\partial \hat{H}}{\partial \lambda} | k \rangle$. Imagine the Kronig-Penney potential as a perturbation to the NFE Hamiltonian $\hat{H} = \hat{H}_0 + \lambda \hat{W}$ where $W(x) = S \sum_n \delta(x - na)$, and $\hat{H}_0 | k \rangle = E_k^0 | n \rangle$, the eigenvalues of the NFE model $E_k^0 = \frac{\hbar^2 (k+G)^2}{2m}$ shown by the dashed lines in Figure 15.3. Then, we must have $\frac{\partial E_k}{\partial \lambda} = \langle k | \frac{\partial (\hat{H}_0 + \lambda W)}{\partial \lambda} | k \rangle = \langle k | W | k \rangle = \int dx | \psi_k(x) |^2 W(x) = SN |u_k(0)|^2 \geq 0$, and the perturbed eigenvalue $E_k \geq E_k^0$. This remains true at all points in k-space except at points of degeneracy, as indicated by an arrow in the left figure of Figure 15.3. At k-points were eigenvalues are degenerate, the splitting is such that for S > 0, one eigenvalue increases, while the other stays put. The lowest energy allowed is $E_{min}^+ > 0$ for S > 0, and the lowest band is rather narrow. This means the electron is 'sluggish' in this band, and it has a large effective mass. As we move up to higher energies, the points of degeneracy develop sharper curvatures and the bands become wider, making the electron effective mass lighter.

Note the differences for the attractive delta potential (S < 0) band structures highlighted by the right panel in Figure 15.3, and drawn at exactly the same scale for easy comparison. The lowest energy allowed now is $E_{min}^- < 0$ for S < 0, i.e.. it is *negative* in stark contrast to the situation for S > 0. The Hellmann-Feynman theorem now guarantees that the eigenvalues are *lower* than the NFE case. At the k-points of degeneracy, the splitting is such that one eigenvalue stays put again, but the other is pushed *down*, exactly opposite to the case of S > 0. The lowest eigenvalue of each Kronig-Penney band is degenerate with the NFE eigenvalues of $E(k = n\frac{\pi}{a}) = n^2 \cdot F$ again, where n = 1, 2, ...,locating energy eigenvalues F, 9F, ... at $k = \pm \frac{\pi}{a}$ at the BZ edge, and 4F, 16F, ... at k = 0as now the *minima* of the corresponding bands.

15.3 Tight-binding models emerge from Kronig-Penney

We will now see that an approximate method to calculate bandstructures called the tight-binding method emerges naturally from the exact Kronig Penney model. Apply the trigonometric identity $\cot(x) = \sum_{-\infty}^{+\infty} \frac{1}{n\pi+x}$ on the right hand side of the central Kronig-Penney eigenvalue equation 15.8, using the fact $G_n = n\frac{2\pi}{a}$. A few trigonometric identities later, equation 15.8 transforms into:

$$\cos(ka) = \cos(qa) + \frac{mSa}{\hbar^2} \cdot \frac{\sin(qa)}{qa}, \qquad (15.9)$$

where $q = \sqrt{2mE_k/\hbar^2}$. This is still an *exact* solution of the Schrodinger equation. Now the values of E_k that satisfy this equation will form the energy bandstructure E_k for each k. The left hand side is limited to $-1 \leq \cos(ka) \leq +1$, but the RHS of equation 15.9 can reach values up to $1 + \frac{mSa}{\hbar^2} = 1 + C$ which can clearly exceed unity. This restricts the allowed values of q for real energy eigenvalues $E = \frac{\hbar^2 q^2}{2m}$ for each k. Figure 15.4 shows the 'bands' of q where the RHS lies between $-1 \leq RHS \leq +1$, and real energy eigenvalues are allowed.

Now the zeroes of $\frac{\sin(x)}{x}$ occur at $x = n\pi$ where $n = \pm 1, \pm 2, \ldots$ It is clear that a band of q-values, and corresponding energies are allowed near the zeroes of the RHS as indicated in Figure 15.4 (left). Let us find an approximate solution for the first band $E_1(k)$ by expanding the RHS for a large strength, or $C = \frac{mSa}{\hbar^2} >> 1$ near the first zero at n = 1, around $qa = \pi$. Using $\delta = \pi - qa$, the expansion yields $\cos(qa) + C \cdot \frac{\sin(qa)}{qa} \approx -1 + \frac{C}{\pi}\delta$, which when used in equation 15.9 yields

$$E_1(k) \approx E_0 - 2J(1 + \cos ka),$$
 (15.10)

where $E_0 = \frac{\pi^2 \hbar^2}{2ma^2}$ coincides with the NFE energy at $k = \frac{\pi}{a}$, and the "hopping" or tunneling term is $J = \frac{\pi^2 \hbar^4}{2m^2 a^3 S} = \frac{E_0}{C}$. This is clearly in the form of a tight-binding model! Now we really don't need to stop at the first root - expanding around $qa = n\pi$, and retaining only the linear expansion terms, we get a more comprehensive tight-binding bandstructure of the n^{th} band as:

$$E_n(k) \approx n^2 E_0 \left[1 - \frac{1}{C} + \frac{(-1)^n}{C} \cos(ka) \right]^2.$$
 (15.11)

Figure 15.4 shows a plot of the first three bands for the dimensionless strength C = 10. Note that the energy eigenvalues at the BZ edges co-incide with the free-electron values. This is similar to the case for S > 0 in the Kronig Penney model in Figure 15.3 (Left).



FIGURE 15.4: The left figure shows a plot of the RHS of Equation 15.9 with x = qa, and the LHS is limited to $-1 \leq LHS \leq +1$. The narrow bands within which the two sides are equal are highlighted; each leads to an allowed energy band. Because the intersections are near $x = n\pi$ where $n = \pm 1, \pm 2, ...,$ an approximate analytical expression of all the bands can be obtained (see Equation 15.11). This first three tight-binding bandstructures are plotted in the right panel. Compare with Figure 15.3.

Now we can write down a more general tight-binding model by starting from orbitals that are localized at each lattice point, and by trying linear combinations of such orbitals in the Bloch-form to coax out the E(k) eigenvalues and corresponding eigenfunctions. We write the linear combination of atomic orbitals (LCAO) ansatz wavefunction as

$$|\psi\rangle = \sum_{m=1}^{N} \frac{e^{i\mathbf{k}\cdot\mathbf{R}_{m}}}{\sqrt{N}} |m\rangle, \qquad (15.12)$$

where we have initially assumed just one orbital per lattice site, and $\psi(\mathbf{r}) = \langle \mathbf{r} | \psi \rangle$ and $\phi_m(\mathbf{r}) = \langle \mathbf{r} | m \rangle$ is the orbital centered at site m. This way or writing the ansatz ensures it is indeed a Bloch function, which is verified by checking $\mathbf{r} \to \mathbf{r} + \mathbf{R}$ leads to $\psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi(\mathbf{r})$. We feed this ansatz into the Schrodinger equation, cancel \sqrt{N} from each side, and get the relation for energy eigenvalues $E(\mathbf{k})$ for each value of \mathbf{k} :

$$\hat{H}\sum_{m=1}^{N}e^{i\mathbf{k}\cdot\mathbf{R}_{m}}|m\rangle = E(\mathbf{k})\sum_{m=1}^{N}e^{i\mathbf{k}\cdot\mathbf{R}_{m}}|m\rangle, \qquad (15.13)$$

We note that \hat{H} does not affect $e^{i\mathbf{k}\cdot\mathbf{R}_m}$, it acts only on $|m\rangle$. Next we project both vectors on to the ansatz $\sum_{n=1}^{N} e^{i\mathbf{k}\cdot\mathbf{R}_n} \langle n|$, and rearrange to get the central result of the tight-binding energy bandstructure:

$$E(\mathbf{k}) = \frac{\sum_{n,m=1}^{N} e^{i\mathbf{k}\cdot(\mathbf{R}_m - \mathbf{R}_n)} \langle n | \hat{H} | m \rangle}{\sum_{n,m=1}^{N} e^{i\mathbf{k}\cdot(\mathbf{R}_m - \mathbf{R}_n)} \langle n | m \rangle},$$
(15.14)

where ... Now it is clear that there are N^2 terms in the double sum. Out of these, the N'diagonal' terms are obtained when n = m, for which we have $e^{i\mathbf{k}\cdot(\mathbf{R}_m-\mathbf{R}_n)} = e^{i\mathbf{k}\cdot(0)} = 1$ and the diagonal matrix elements are all equal: $\langle n|\hat{H}|n\rangle = E_0$. This energy is slightly *lower* in energy than the original 'atomic' orbital energy because each electron orbital also sees nearby atomic potentials. In the denominator, the diagonal sum gives just $\sum e^{i\mathbf{k}\cdot(0)}\langle n|n\rangle = N$. Let us now look at the rest $N^2 - N$ off-diagonal terms in the numerator and denominator.

Now considering a 1D lattice of lattice constant a, for the 1st nearest neighbors, we have N terms for which the terms in the numerator take the form $\sum_{n=1}^{N} (e^{+ika} \langle n | \hat{H} | n + 1 \rangle + e^{-ika} \langle n - 1 | \hat{H} | n \rangle) = -2Nt_1 \cos(ka)$, where the hopping integral $t_1 = \langle n | \hat{H} | n + 1 \rangle$. Similarly, the denominator gives the sum of the N 1st nearest neighbor terms as $+2Ns_1 \cos(ka)$, where $s_1 = \langle n | n + 1 \rangle$ is clearly very small because of decaying wavefunctions that are tightly bound to the lattice sites.

There are N more terms for the 2nd nearest neighbors characterized by the hopping integral $t_2 = \langle n | \hat{H} | n + 2 \rangle$ and the overlap integral $\langle n | n + 2 \rangle = s_2$. And then for the 3rd nearest neighbor, and so on... It is intuitively clear that the successive terms t_n and s_n decay fast. Now we can write the expression for the tight-binding bandstructure as:

$$E(k) = \frac{E_0 - 2t_1 \cos(ka) - 2t_2 \cos(2ka) - 2t_3 \cos(3ka)...}{1 + 2s_1 \cos(ka) + 2s_2 \cos(2ka) + 2s_3 \cos(3ka)...} \approx E_0 - 2t_1 \cos(ka).$$
(15.15)

If instead of 1D, we are in 2D or 3D, then there are more nearest neighbors and the bands acquire more "structure".

15.4 **Point defects in Kronig-Penney Models**

Now imagine that in the Kronig-Penney model, only one of the N sites has a potential that is different from the other sites. Let us call this difference in the strength U_0 , meaning at this particular site, the delta-function strength is $S + U_0$ instead of S, where



 U_0 can be positive or negative. What is the effect on the energy eigenvalues and the eigenstates due to the presence of this 'defect'?

FIGURE 15.5: Figures showing the effect of defect states on the allowed energy eigenvalues as a function of the defect potential strength. The left figure shows the graphical solution to the Kronig-Penney type solution, and in particular illustrates the splitting off of one eigenvalue - the highest eigenvalue of the band for positive defect potentials, and the lowest energy eigenvalues for negative defect potentials. This is further high-lighted in the figure on the right, where the eigenvalue spectrum is plotted as a function of the defect potential.

This problem can now be solved because the exact solution of the Kronig-Penney model without the defect has given us the eigenvalues for each k-state in the BZ as $E_{KP}(k)$ - for example - shown in Figure 15.3. Then, we go through exactly the same procedure that led to the Kronig-Penney solution in Equation 15.8, and end up with the new solution

$$\frac{Na}{U_0} = \sum_k \frac{1}{E_k - E_{KP}(k)},\tag{15.16}$$

where k are the allowed states in the 1st BZ, N is the number of lattice sites, and therefore Na = L is the macroscopic length. Clearly, in the absence of the defect, $U_0 \rightarrow 0$, and the LHS $\rightarrow \infty$. This happens exactly N times in the RHS when the allowed energies $E_k = E_{KP}(k)$, i.e., we recover the original Kronig-Penney solution without the defect, as we should. But when $U_0 \neq 0$, the allowed energies E_k must deviate from $E_{KP}(k)$ to satisfy the exact solution above. To illustrate the solution graphically, we plot the RHS and the LHS in Figure 15.5. We will see in the next section that the RHS of Equations 15.16 and 15.8 are actually the Trace of the Green's function matrix of the problem, i.e., $\sum_k \frac{1}{E_k - E_{KP}(k)} = \text{Trace}[\hat{G}(E)]$. The plot in Figure 15.5 for a few-site chain shows the effect of the defect on the eigenvalue spectrum clearly. The figure on the right illustrates the movement of the eigenvalues as the strength of the defect is tuned from zero to large positive and large negative. The eigenvalues at $U_0 = 0$ constitute the band without the defect. When the defect strength is +ve and strong, the LHS L/U_0 line moves closer to the x-axis (left figure), and it is clear that one of the intersections - at the *top* of the energy band splits off from the band rapidly, whereas all other eigenvalues do not change as much. Any change is positive, as guaranteed by the Hellmann-Feynman theorem. This is a characteristic feature - similarly, for a negative U_0 , the lowest eigenvalue of the band splits off and leaves other eigenvalues mostly unchanged.

We will see later that $U_0 > 0$ 'defects' explain the formation of acceptor states at the top of valence bands, and are designed such that the splitting energy is less than kT for room-temperature generation of holes. Similarly, the bottom of the band with $U_0 < 0$ models donor states and electron doping at the bottom of the conduction band of semiconductors.

15.5 Green's functions and Kronig-Penney for higher-dimensions

We noted the repeated appearance of sums over the Brillouin zone of the kind $\sum_k \frac{1}{E-E(k)}$ which have units of (energy)⁻¹. This may be thought of as a function of the variable E, or energy. The reason why such sums permeate exact solutions of problems will now become clear: and will lead us to define Green's functions.

Consider the Schrodinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \to [i\hbar\frac{\partial}{\partial t} - \hat{H}]\Psi = 0.$$
 (15.17)

Let us think of the equation as the product of the operator (or matrix) $i\hbar\frac{\partial}{\partial t} - \hat{H}$ with Ψ . For this product to be zero, either $i\hbar\frac{\partial}{\partial t} - \hat{H}$ or Ψ , or both should be zero. The only interesting case here is when we actually have a quantum object with a nonzero wavefunction, $\Psi \neq 0$. Thus, $i\hbar\frac{\partial}{\partial t} - \hat{H}$ should be zero. Now we have learnt that if the quantum object is in a state of definite energy, $i\hbar\frac{\partial}{\partial t}\Psi_n = E_n\Psi_n$, Ψ_n , and E_n is a real eigenvalue representing the energy of the state. Let us generalize this and write $i\hbar\frac{\partial}{\partial t} = E$,

where E is a variable. We can then write the Schrodinger equation as $[EI - \hat{H}]\Psi = 0$, where I is an identity operator, or the identity matrix when the equation is written out for any chosen basis. However, the equation in this form does not hold true for all E, but only for certain $E = E_n$ - only when the variable E matches up with an allowed eigenvalue. Now let us think of $EI - \hat{H}$ as a function of E. When we vary E, this function has very sharp responses when $E = E_n$: the function is a 'detector' of eigenvalues - it detects an eigenvalue by vanishing. At those sharp energies, $\Psi = \Psi_n \neq 0$ is an eigenfunction, so the function provides the eigenfunction as its 'residue'. Now with this qualitative picture in mind, let us solidify the concept of the Green's function of the system.

We like detectors to 'scream' when they detect, rather than to go silent. So, can we find a function \hat{G} that instead of solving the equation $[EI - \hat{H}]\Psi = 0$, solves the equation $[EI - \hat{H}]\hat{G} = I$ instead? Formally, the function is clearly $\hat{G} = [EI - \hat{H}]^{-1}$. This function clearly blows up when $E = E_n$, and is indeed the screaming detector we are looking for. It is the Green's function for the Hamiltonian \hat{H} . Let us assume that we know all the eigenvalues of a particular Hamiltonian \hat{H}_0 to be E_n and the corresponding eigenfunctions are $|n\rangle$. The Green's function can then be written out as a matrix form

$$\hat{G}_{0}(E) = \sum_{n} [EI - \hat{H}]^{-1} |n\rangle \langle n| = \sum_{n} \frac{|n\rangle \langle n|}{E - E_{n}}.$$
(15.18)

It is clear that the Green's function is actually a matrix, and sums of the kind that appeared earlier in the solution of the Kronig-Penney and the defect problems are the sum of the diagonal terms in a diagonal basis. Now it turns out that the sum of the diagonal terms is *invariant* with what basis one writes the matrix - which is why it goes by a name - the Trace. Thus, we have a very important relation

Trace
$$[\hat{G}(E)] = \sum_{k} \frac{1}{E - E_0(k)}$$
 (15.19)

where $E_0(k)$ are the allowed eigenvalues of the system. The solution of the Kronig-Penney model is thus very compactly written in the formal way as $\operatorname{Trace}[\hat{G}_0(E)] = \frac{a}{S}$, where $\hat{G}_0(E) = (EI - \hat{H}_0)^{-1}$, and $\hat{H}_0|k\rangle = E_0(k)|k\rangle$ are the nearly-free electron eigensystem, with $E_0(k) = \frac{\hbar^2(k+G)^2}{2m}$. The solution of a single-site defect state of strength S_0 is then written as $\operatorname{Trace}[\hat{G}(E)] = \frac{Na}{S_0}$, where now the Green's function is for the exactly solved Kronig-Penney eigensystem $\operatorname{Trace}[\hat{G}(E)] = (EI - \hat{H})^{-1}$, where $\hat{H}|k\rangle = E_{KP}(k)|k\rangle$, and $E_{KP}(k)$ are the Kronig-Penney eigenvalues.

More on Green's functions - relation to DOS, etc to be written...