Chapter 7

Energy Bands

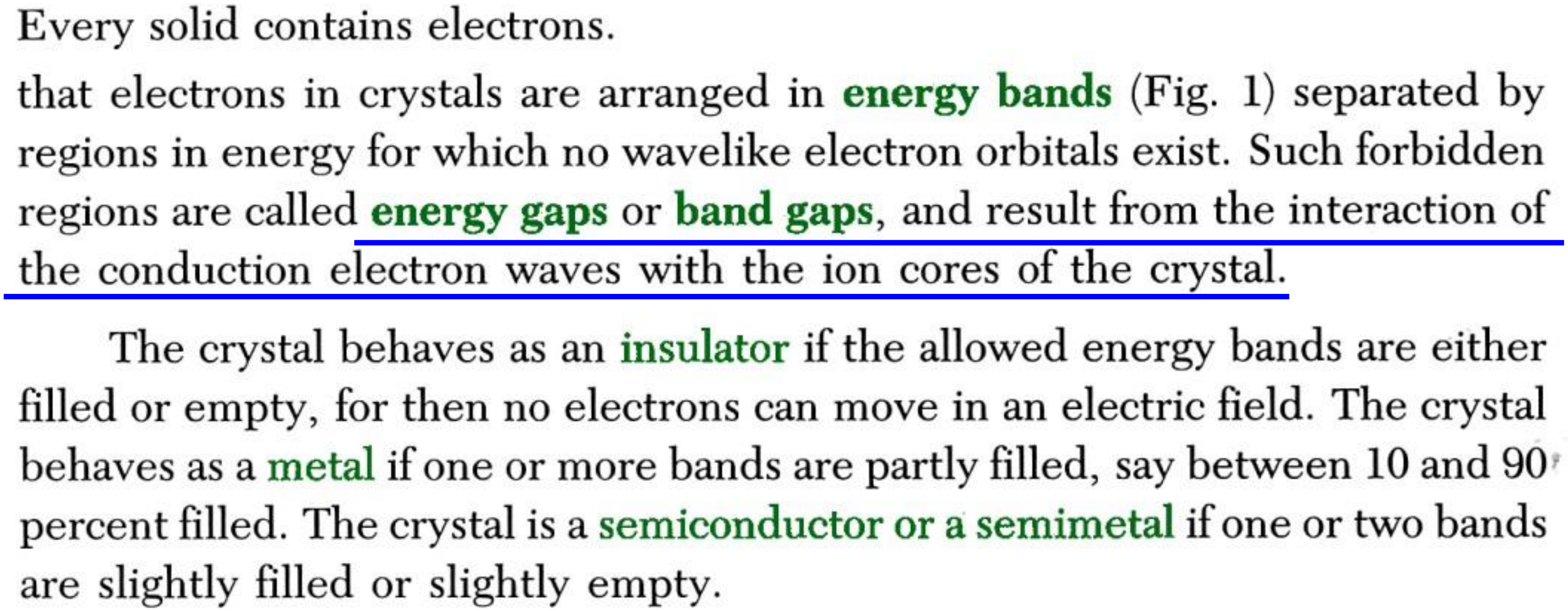
Free Electron Model (Sommerfeld):

Success:

- Heat capacity
- Thermal conductivity
- Magnetic susceptibility
- Electrodynamics of metals

Failure:

- Distinction between metals, semimetals, and insulators.
- Positive Hall coefficient
- Magneto-transport



Every solid contains electrons.

that electrons in crystals are arranged in **energy bands** (Fig. 1) separated by regions in energy for which no wavelike electron orbitals exist. Such forbidden regions are called **energy gaps** or **band gaps**, and result from the interaction of the conduction electron waves with the ion cores of the crystal.

The crystal behaves as an insulator if the allowed energy bands are either filled or empty, for then no electrons can move in an electric field. The crystal behaves as a metal if one or more bands are partly filled, say between 10 and 90 percent filled. The crystal is a semiconductor or a semimetal if one or two bands are slightly filled or slightly empty.

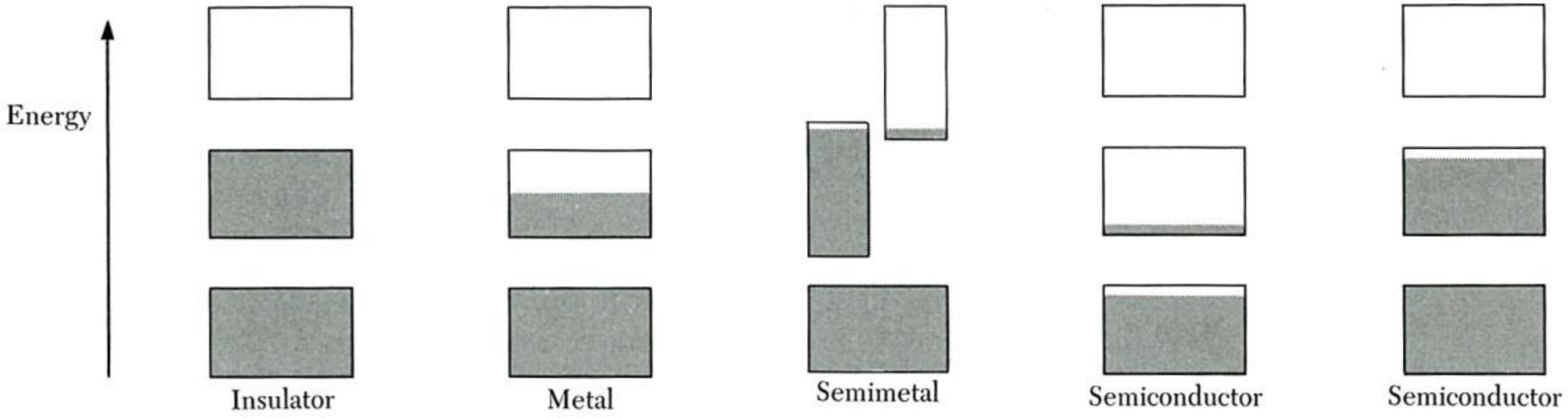
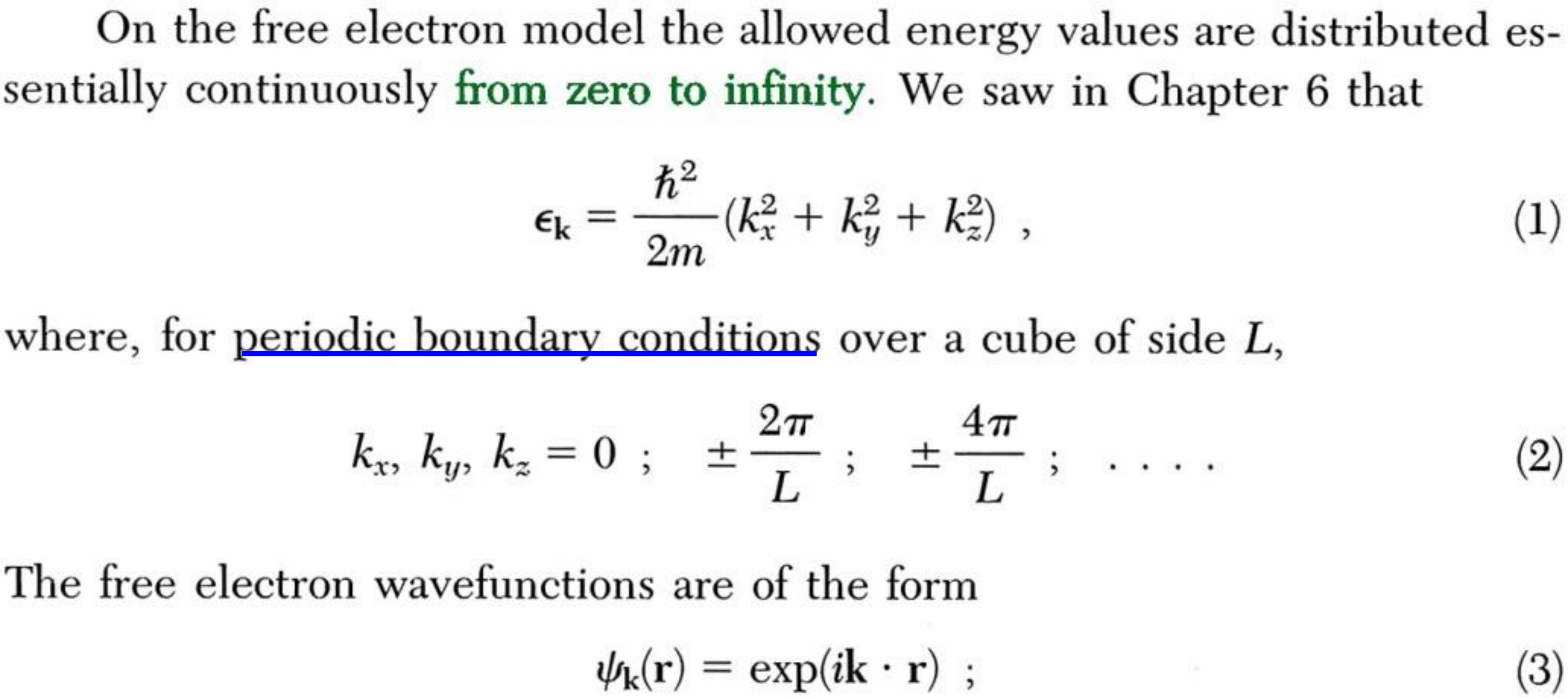


Figure 1 Schematic electron occupancy of allowed energy bands for an insulator, metal, semimetal, and semiconductor. The vertical extent of the boxes indicates the allowed energy regions; the shaded areas indicate the regions filled with electrons. **In a semimetal** (such as bismuth) one band is almost filled and another band is nearly empty at absolute zero, but a pure **semiconductor** (such as silicon) becomes an insulator at absolute zero. The left of the two semiconductors shown is at a finite temperature, with carriers excited thermally. The other semiconductor is electron-deficient because of impurities.

Brief Review:

Free Electron Model



On the free electron model the allowed energy values are distributed essentially continuously from zero to infinity. We saw in Chapter 6 that

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2) \ , \tag{1}$$

where, for periodic boundary conditions over a cube of side L,

$$k_x,\ k_y,\ k_z = 0 \ ; \quad \pm\frac{2\pi}{L} \ ; \quad \pm\frac{4\pi}{L} \ ; \quad \dots \ . \tag{2}$$

The free electron wavefunctions are of the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) \ ; \tag{3}$$

To understand the difference between insulators and conductors, we must extend the free electron model to take account of the periodic lattice of the solid. The possibility of a **band gap** is the most important new property that emerges.

The electrons respond to applied electric or magnetic fields as if the electrons were endowed with an effective mass m^*, which may be larger or smaller than the free electron mass, or may even be negative. Electrons in crystals respond to applied fields, as if endowed with negative or positive charges, $-e$ or $+e$, and herein lies the explanation of the negative and positive values of the Hall coefficient.

Nearly Free Electron Model

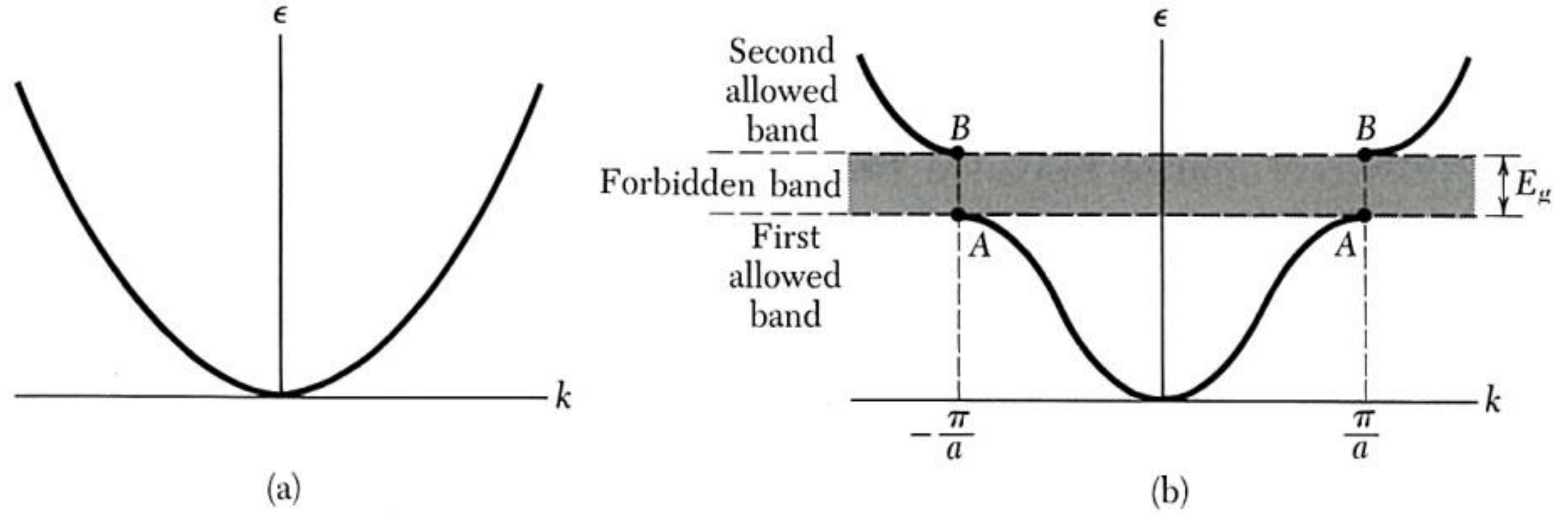
The band structure of a crystal can often be explained by the nearly free electron model for which the band electrons are treated as perturbed only weakly by the periodic potential of the ion cores. This model answers almost all the qualitative questions about the behavior of electrons in metals.

We know that Bragg reflection is a characteristic feature of wave propagation in crystals. ***Bragg reflection of electron waves in crystals is the cause of energy gaps.*** (At Bragg reflection, wavelike solutions of the Schrodinger equation do not exist, as in Fig. 2.) These energy gaps are of decisive significance in determining whether a solid is an insulator or a conductor.

We explain **physically the origin of energy gaps** in the simple problem of a linear solid of lattice constant *a* for a 1-D linear lattice.

Free Electron Model

Nearly Free Electron Model



(a) (b)

Figure 2 (a) Plot of energy ϵ versus wavevector k for a free electron. (b) Plot of energy versus wavevector for an electron in a monatomic linear lattice of lattice constant a. The energy gap E_g shown is associated with the first Bragg reflection at $k = \pm\pi/a$; other gaps are found at $\pm n\pi/a$, for integral values of n.

The Bragg condition $(k + G)^2 = k^2$ for diffraction of a wave of wavevector **k** becomes in one dimension

$$k = \pm \frac{1}{2} G = \pm n\pi/a, \qquad \textit{Solutions in 1-D} \qquad (4)$$

Where $G = 2\pi n/a$ is a reciprocal lattice vector and ***n*** is an integer. The first reflections and the first energy gap occur at $k = \pm\pi/a$. The region in **k** space between $-\pi/a$ and π/a is the **first Brillouin zone** of this lattice. Other energy gaps occur for other values of the integer ***n***.

The wavefunctions at $k = \pm\pi/a$ are not the traveling waves $\exp(i\pi x/a)$ or $\exp(-i\pi x/a)$ of free electrons. At these special values of k the wavefunctions are made up of equal parts of waves traveling to the right and to the left. When the Bragg reflection condition $k = \pm\pi/a$ is satisfied by the wavevector, a wave traveling to the right is Bragg-reflected to travel to the left, and vice versa. Each subsequent Bragg reflection will reverse the direction of travel of the wave. A wave that travels neither to the right nor to the left is a standing wave: it doesn't go anywhere.

The time-independent state is represented by standing waves. We can form two different standing waves from the two traveling waves $\exp(\pm i\pi x/a)$, namely

Standing wave solutions

$$\begin{aligned}\psi(+) &= \exp(i\pi x/a) + \exp(-i\pi x/a) = 2\cos(\pi x/a)\ ;\\ \psi(-) &= \exp(i\pi x/a) - \exp(-i\pi x/a) = 2i\sin(\pi x/a)\ .\end{aligned} \tag{5}$$

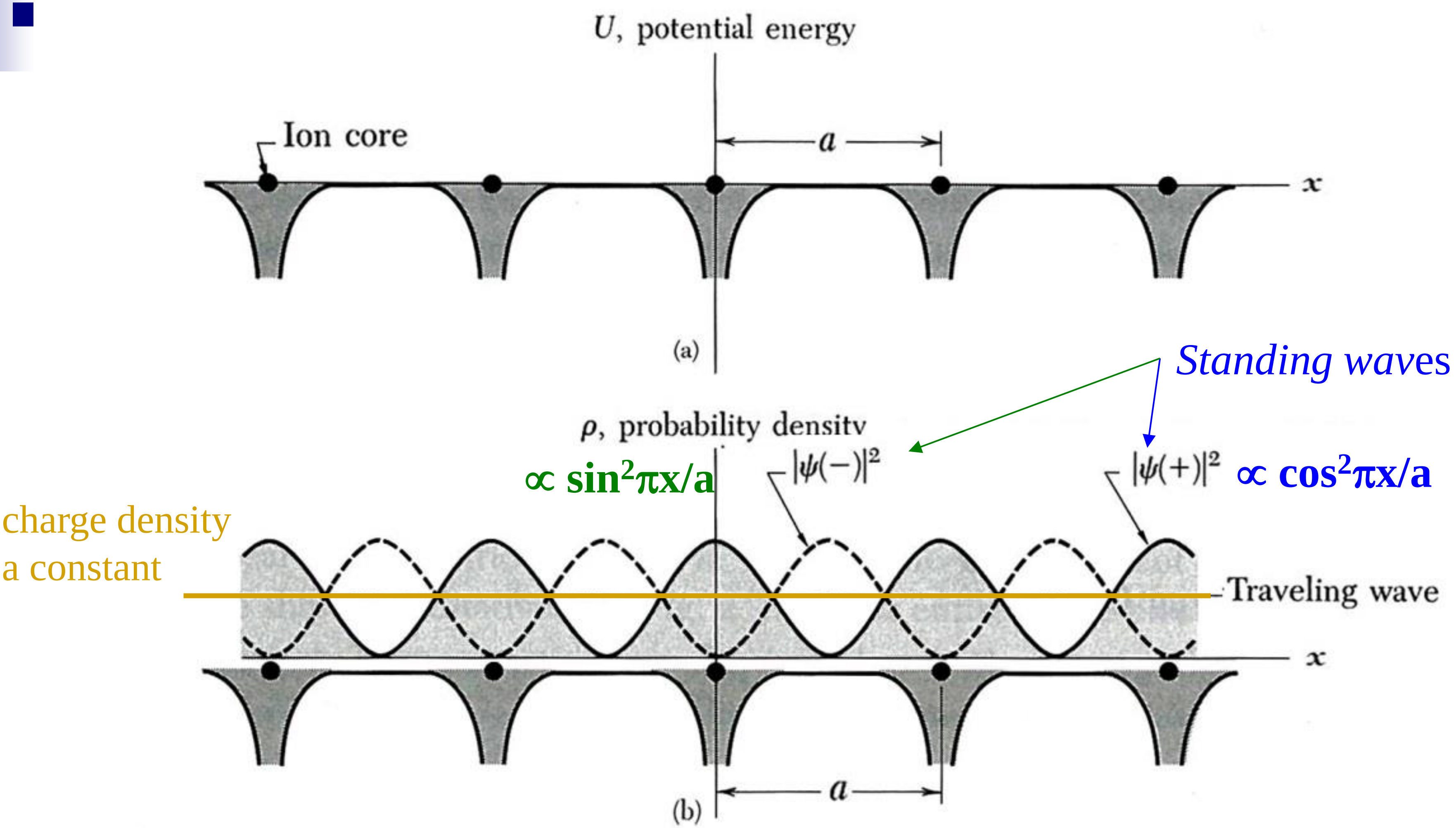
The standing waves are labeled (+) or (−) according to whether or not they change sign when $-x$ is substituted for x. Both standing waves are composed of equal parts of right- and left-directed traveling waves.

Origin of the Energy Gap

The two standing waves $\psi(+)$ and $\psi(-)$ pile up electrons at different regions, and therefore the two waves have different values of the potential energy. This is the origin of the energy gap. The probability density ρ of a particle is $\psi^*\psi = |\psi|^2$. For a pure traveling wave $\exp(ikx)$, we have $\rho = \exp(-ikx)\exp(ikx) = 1$, so that the charge density is constant. The charge density is not constant for linear combinations of plane waves. Consider the standing wave $\psi(+)$ in (5); for this we have

$$\rho(+) = |\psi(+)|^2 \propto \cos^2 \pi x/a \ .$$

This function piles up electrons (negative charge) on the positive ions centered at $x = 0, a, 2a, \ldots$ in Fig. 3, where the potential energy is lowest.



Standing waves

$\propto \sin^2 \pi x/a$

$\propto \cos^2 \pi x/a$

charge density a constant

Figure 3 (a) Variation of potential energy of a conduction electron in the field of the ion cores of a linear lattice. (b) Distribution of probability density ρ in the lattice for $|\psi(-)|^2 \propto \sin^2 \pi x/a$; $|\psi(+)|^2 \propto \cos^2 \pi x/a$; and for a traveling wave. The wavefunction $\psi(+)$ piles up electronic charge on the cores of the positive ions, thereby lowering the potential energy in comparison with the average potential energy seen by a traveling wave. The wavefunction $\psi(-)$ piles up charge in the region between the ions, thereby raising the potential energy in comparison with that seen by a traveling wave. This figure is the key to understanding the origin of the energy gap.

Figure 3a pictures the variation of the electrostatic potential energy of a conduction electron in the field of the positive ion cores. The ion cores bear a net positive charge because the atoms are ionized in the metal, with the valence electrons taken off to form the conduction band. The potential energy of an electron in the field of a positive ion is negative, so that the force between them is attractive.

For the other standing wave $\psi(-)$ the probability density is

$$\rho(-) = |\psi(-)|^2 \propto \sin^2 \pi x/a \ ,$$

which concentrates electrons away from the ion cores. In Fig. 3b we show the electron concentration for the standing waves $\psi(+)$, $\psi(-)$, and for a traveling wave.

When we calculate the average or expectation values of the potential energy over these three charge distributions, we find that the potential energy of $\rho(+)$ is lower than that of the traveling wave, whereas the potential energy of $\rho(-)$ is higher than the traveling wave. We have an energy gap of width E_g if the energies of $\rho(-)$ and $\rho(+)$ differ by E_g. Just below the energy gap at points A in Fig. 2 the wavefunction is $\psi(+)$, and just above the gap at points B the wavefunction is $\psi(-)$.

Magnitude of the Energy Gap

The wavefunctions at the Brillouin zone boundary $k = \pi/a$ are $\sqrt{2}\cos \pi x/a$ and $\sqrt{2}\sin \pi x/a$, normalized over unit length of line. We write the potential energy of an electron in the crystal at point x as

roughly as $U(x) = -U\cos 2\pi x/a$ (negative, with a period of ***a***)

The first-order energy difference between the two standing wave states is

$$E_g = \int_0^{\overset{a}{1}} dx\, U(x)\, [|\psi(+)|^2 - |\psi(-)|^2] \qquad \boxed{\text{x } -1}$$

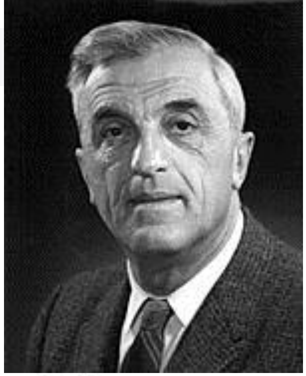
$$= -2\int dx\, U\cos(2\pi x/a)\,(\cos^2 \pi x/a - \sin^2 \pi x/a) = U \qquad \boxed{-U} \tag{6}$$

See Fig. 2

We see that the gap is equal to the Fourier component of the crystal potential.

Awarded the 1952 Nobel Prize for "their development of new ways and methods for nuclear magnetic precision measurements

Felix Bloch
(1905-1983, Swiss)



BLOCH FUNCTIONS

F. Bloch proved the important theorem that the solutions of the Schrödinger equation for a periodic potential must be of a special form:

$$\boxed{\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \ ,} \tag{7}$$

where $u_{\mathbf{k}}(\mathbf{r})$ has the period of the crystal lattice with $\underline{u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{T})}$. The result (7) expresses the Bloch theorem:

> The eigenfunctions of the wave equation for a periodic potential are the product of a plane wave $\exp(i\mathbf{k} \cdot \mathbf{r})$ times a function $u_{\mathbf{k}}(\mathbf{r})$ with the periodicity of the crystal lattice.

A one-electron wavefunction of the form (7) is called a Bloch function and can be decomposed into a sum of traveling waves, as we see later. Bloch functions can be assembled into localized wave packets to represent electrons that propagate freely through the potential field of the ion cores.

We give now a restricted proof of the Bloch theorem, valid when ψ_k is nondegenerate.

We consider N identical lattice points on a ring of length Na. The potential energy is periodic in a, with $\underline{U(x) = U(x + sa)}$, where s is an integer.

We are guided by the symmetry of the ring to look for solutions of the wave equation such that

$$\psi(x + a) = C\psi(x) \quad , \tag{8}$$

where C is a constant. Then, on going once around the ring,

$$\psi(x + Na) = \psi(x) = C^N \psi(x) \quad ,$$

because $\psi(x)$ must be single-valued. It follows that C is one of the N roots of unity, or

$$C = \exp(i2\pi s/N) \;; \qquad s = 0,\ 1,\ 2,\ \ldots,\ N - 1 \ . \tag{9}$$

We see that

$$\underline{\psi(x) = u_k(x) \exp(i2\pi sx/Na)} \tag{10}$$

satisfies (8), provided that $u_k(x)$ has the periodicity a, so that $\underline{u_k(x) = u_k(x + a)}$. With $\boxed{k = 2\pi s/Na}$, we have the Bloch result (7). For another derivation, see (29).

Demonstration of the Energy Gap !

KRONIG-PENNEY MODEL

A periodic potential for which the wave equation can be solved in terms of elementary functions is the square-well array of Fig. 4. The wave equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = \epsilon\psi \ , \qquad (11)$$

where $U(x)$ is the potential energy and ϵ is the energy eigenvalue.

Total period = a + b

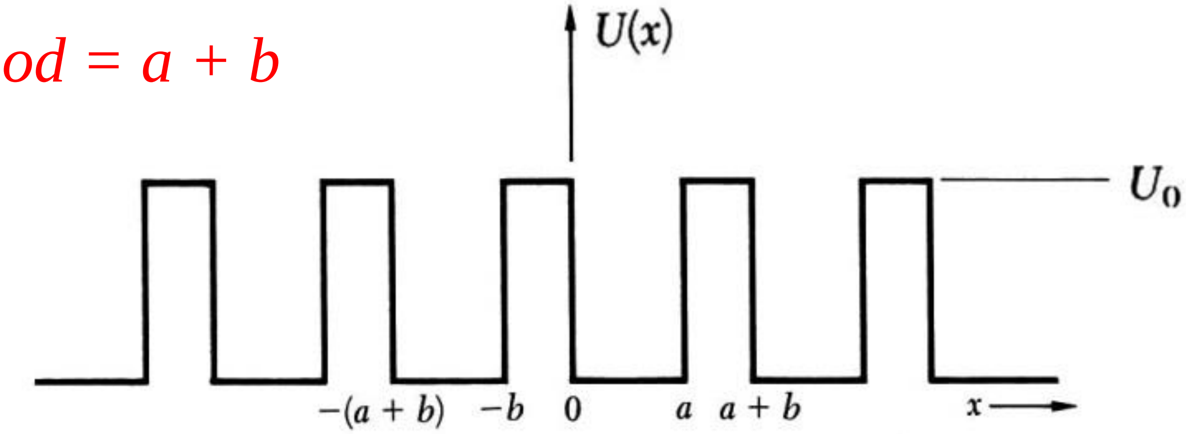


Figure 4 Square-well periodic potential as introduced by Kronig and Penney.

In the region $0 < x < a$ in which $U = 0$, the eigenfunction is a linear combination,

$$\psi = Ae^{iKx} + Be^{-iKx} \quad , \tag{12}$$

of plane waves traveling to the right and to the left, with energy

$$\epsilon = \hbar^2 K^2 / 2m \quad . \tag{13}$$

In the region $-b < x < 0$ within the barrier the solution is of the form

$$\psi^* = Ce^{Qx} + De^{-Qx} \quad , \tag{14}$$

with

Note $\epsilon < U_o$

$$U_0 - \epsilon = \hbar^2 Q^2 / 2m \quad . \tag{15}$$

We want the complete solution to have the Bloch form (7). Thus the solution in the region $a < x < a + b$ must be related to the solution (14) in the region $-b < x < 0$ by the Bloch theorem: by translation of $r = a+b$

$$\psi(a < x < a + b) = \psi(-b < x < 0)\, \boxed{e^{ik(a+b)}} \ , \tag{16}$$

which serves to define the wavevector k used as an index to label the solution.

The constants A, B, C, D are chosen so that ψ and $d\psi/dx$ are continuous at $x = 0$ and $x = a$. These are the usual quantum mechanical boundary conditions in problems that involve square potential wells. At $x = 0$,

$\psi(0) = \psi^*(0)$
$d\psi(0)/dx = d\psi^*(0)/dx$

$$A + B = C + D \ ; \tag{17}$$

$$iK(A - B) = Q(C - D) \ . \tag{18}$$

At $x = a$, with the use of (16) for $\psi(a)$ under the barrier in terms of $\psi(-b)$,

$$Ae^{iKa} + Be^{-iKa} = (Ce^{-Qb} + De^{Qb})\, e^{ik(a+b)} \ ; \tag{19}$$

$$iK(Ae^{iKa} - Be^{-iKa}) = Q(Ce^{-Qb} - De^{Qb})\, e^{ik(a+b)} \ . \tag{20}$$

$\psi(a) = \psi^*(-b) \exp[ik(a+b)]$
$d\psi(a)/dx = d\psi^*(-b)/dx \exp[ik(a+b)]$

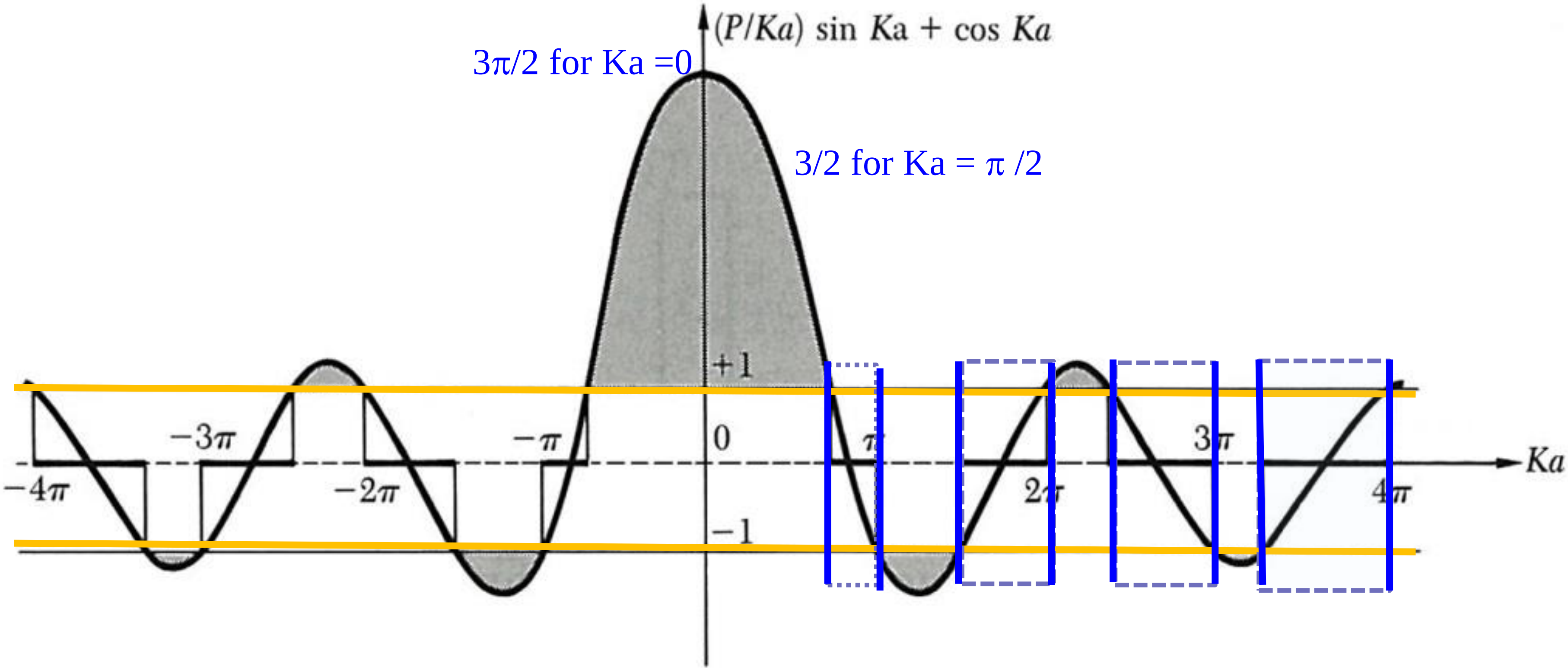
The four equations (17) to (20) have a solution only if the determinant of the coefficients of A, B, C, D vanishes, or if

$$[(Q^2 - K^2)/2QK] \sinh Qb \sin Ka + \cosh Qb \cos Ka = \cos k(a + b) \quad . \tag{21a}$$

The result is simplified if we represent the potential by the periodic delta function obtained when we pass to the limit $b = 0$ and $U_0 = \infty$ in such a way that $Q^2ba/2 = P$, a finite quantity. In this limit $Q \gg K$ and $Qb \ll 1$. Then (21a) reduces to

$$\boxed{(P/Ka)\sin Ka + \cos Ka = \cos ka \quad .} \tag{21b}$$

The ranges of K for which this equation has solutions are plotted in Fig. 5, for the case $P = 3\pi/2$. The corresponding values of the energy are plotted in Fig. 6. Note the energy gaps at the zone boundaries. The wavevector k of the Bloch function is the important index, not the K in (12), which is related to the energy by (13).



Certain allowed values of ***Ka****, thus giving rise to allowed values of* ϵ

Figure 5 Plot of the function $(P/Ka) \sin Ka + \cos Ka$, for $P = 3\pi/2$. The allowed values of the energy ϵ are given by those ranges of $Ka = (2m\epsilon/\hbar^2)^{1/2}a$ for which the function lies between ±1. For other values of the energy there are no traveling wave or Bloch-like solutions to the wave equation, so that forbidden gaps in the energy spectrum are formed.

$\epsilon = \hbar^2 K^2/2m$

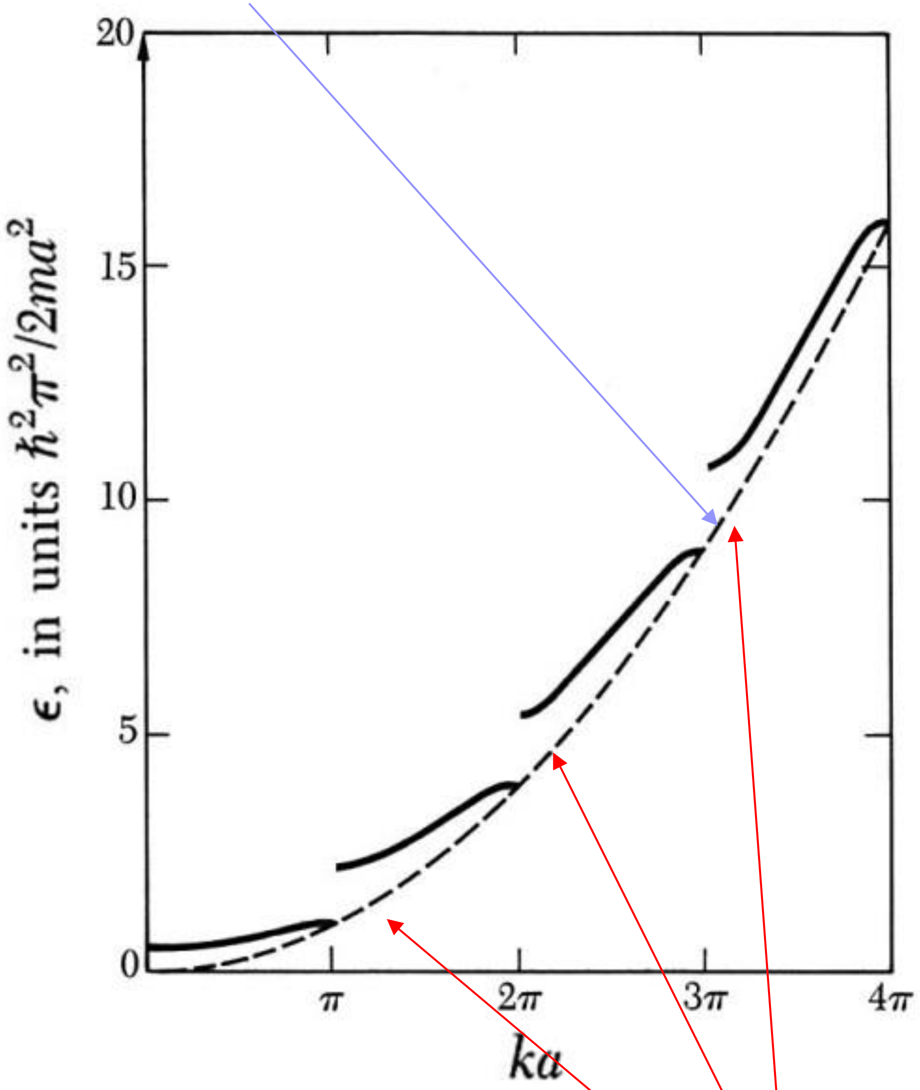


Figure 6 Plot of energy vs. wavenumber for the Kronig-Penney potential, with $P = 3\pi/2$. Notice the energy gaps at $ka = \pi, 2\pi, 3\pi \ldots .$

Forming energy gaps in ϵ *at* $\boldsymbol{Ka = n\pi}$ *!*

Exact Proof of the Bloch Theorem

We considered in Fig. 3 the approximate form we expect for the solution of the Schrödinger equation if the wavevector is at a zone boundary, as at $k = \pi/a$. We treat in detail the wave equation for a general potential, at general values of k. Let $U(x)$ denote the potential energy of an electron in a linear lattice of lattice constant a. We know that the potential energy is invariant under a crystal lattice translation: $U(x) = U(x + a)$. A function invariant under a crystal lattice translation may be expanded as a Fourier series in the reciprocal lattice vectors G. We write the Fourier series for the potential energy as

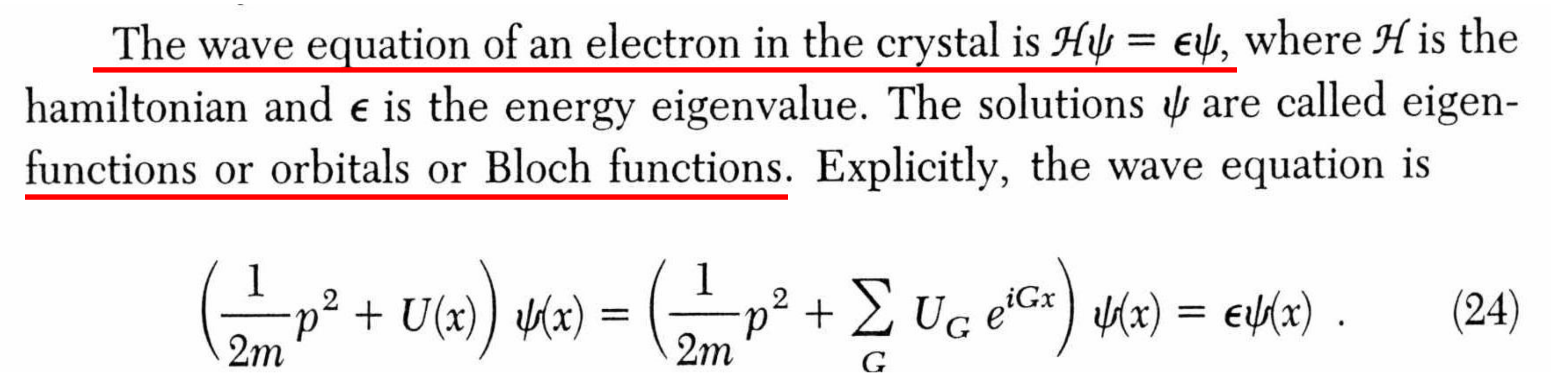
$$U(x) = \sum_G U_G \, e^{iGx} \; . \tag{22}$$

The values of the coefficients U_G for actual crystal potentials tend to decrease rapidly with increasing magnitude of G. For a bare coulomb potential U_G decreases as $1/G^2$.

We want the potential energy $U(x)$ to be a real function:

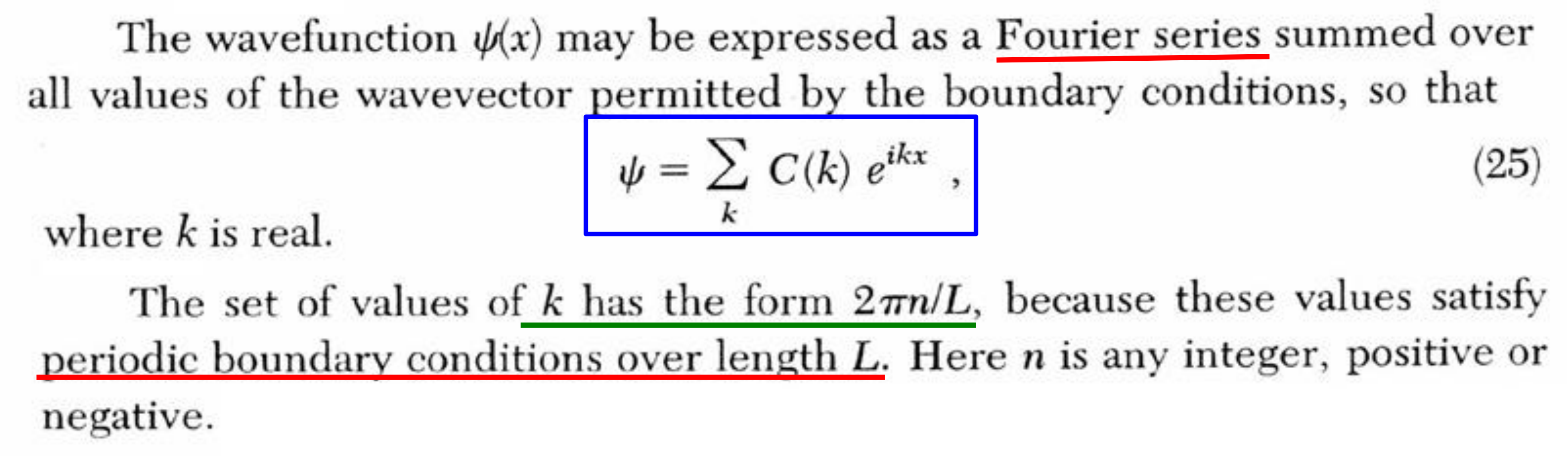
$$U(x) = \sum_{G>0} U_G(e^{iGx} + e^{-iGx}) = 2 \sum_{G>0} U_G \cos Gx \; . \tag{23}$$

For convenience we have assumed that the crystal is symmetric about $x = 0$



The wave equation of an electron in the crystal is $\mathcal{H}\psi = \epsilon\psi$, where $\mathcal{H}$ is the hamiltonian and ϵ is the energy eigenvalue. The solutions ψ are called eigenfunctions or orbitals or Bloch functions. Explicitly, the wave equation is

$$\left(\frac{1}{2m}p^2 + U(x)\right)\psi(x) = \left(\frac{1}{2m}p^2 + \sum_G U_G\, e^{iGx}\right)\psi(x) = \epsilon\psi(x) \ . \tag{24}$$



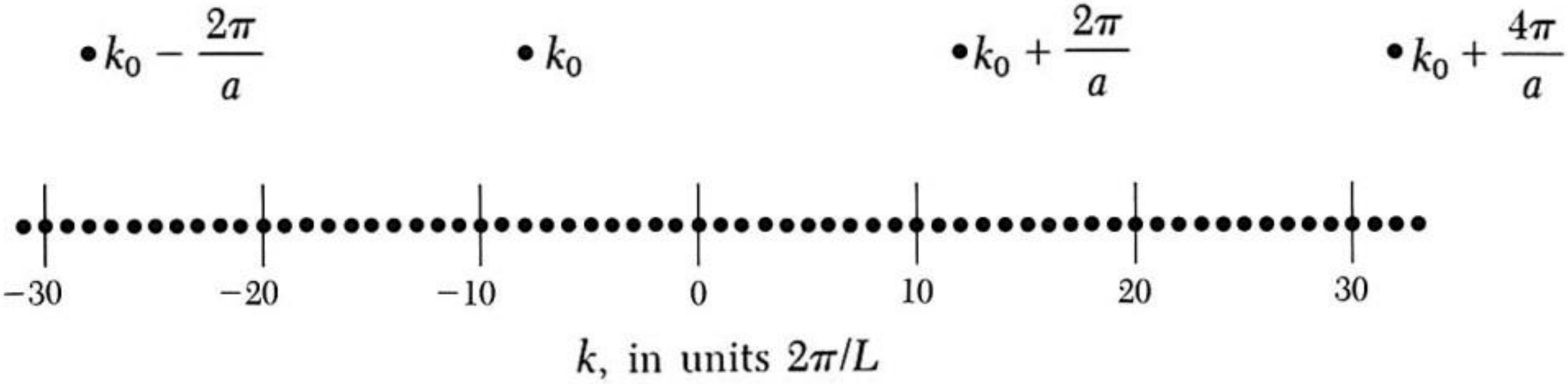
The wavefunction $\psi(x)$ may be expressed as a Fourier series summed over all values of the wavevector permitted by the boundary conditions, so that

$$\boxed{\psi = \sum_k C(k)\, e^{ikx} \ ,} \tag{25}$$

where k is real.

The set of values of k has the form $2\pi n/L$, because these values satisfy periodic boundary conditions over length L. Here n is any integer, positive or negative.

We can label a wavefunction ψ that contains a component k as ψ_k or, equally well, as ψ_{k+G}, because if k enters the Fourier expansion then $k + G$ may enter.



$L = 20\ a$

Figure 7 The lower points represent values of the wavevector $k = 2\pi n/L$ allowed by the periodic boundary condition on the wavefunction over a ring of circumference L composed of 20 primitive cells. The allowed values continue to $\pm\infty$. The upper points represent the first few wavevectors which may enter into the Fourier expansion of a wavefunction $\psi(x)$, starting from a particular wavevector $k = k_0 = -8(2\pi/L)$. The shortest reciprocal lattice vector is $2\pi/a = 20(2\pi/L)$.

We shall usually choose as a label for the Bloch function that k which lies within the first Brillouin zone.

To solve the wave equation, substitute (25) in (24) to obtain a set of linear algebraic equations for the Fourier coefficients. The kinetic energy term is

$$\frac{1}{2m}p^2\psi(x) = \frac{1}{2m}\left(-i\hbar\frac{d}{dx}\right)^2\psi(x) = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = \frac{\hbar^2}{2m}\sum_k k^2C(k)\,e^{ikx} \;;$$

and the potential energy term is

$$\left(\sum_G U_G\,e^{iGx}\right)\psi(x) = \sum_G\sum_k U_G\,e^{iGx}C(k)\,e^{ikx} \;.$$

The wave equation is obtained as the sum:

$$\sum_k \frac{\hbar^2}{2m}k^2C(k)\,e^{ikx} + \sum_G\sum_k U_GC(k)\,e^{i(k+G)x} = \epsilon\sum_k\; C(k)\,e^{ikx} \;. \tag{26}$$

Each Fourier component must have the same coefficient on both sides of the equation. Thus

$$\boxed{(\lambda_k - \epsilon)C(k) + \sum_G U_GC(k-G) = 0 \;,} \tag{27}$$

with the notation

$$\lambda_k = \hbar^2k^2/2m \;. \tag{28}$$

The Central equation to solve the C(K), and ε

there are, in principle, an infinite number of $C(k-G)$ to be determined. In practice a small number will often suffice, perhaps two or four.

Restatement of the Bloch Theorem

Once we determine the C's from (27), the wavefunction (25) is given as

$$\psi_k(x) = \sum_G C(k-G)\, e^{i(k-G)x} \quad , \tag{29}$$

which may be rearranged as

$$\underline{\psi_k(x) = \left(\sum_G C(k-G)\, e^{-iGx}\right) e^{ikx} = e^{ikx} u_k(x) \quad ,}$$

with the definition

$$\boxed{u_k(x) \equiv \sum_G C(k-G)\, e^{-iGx} \quad .}$$

Because $u_k(x)$ is a Fourier series over the reciprocal lattice vectors, it is invariant under a crystal lattice translation T, so that $\underline{u_k(x) = u_k(x+T)}$. We verify this directly by evaluating $u_k(x+T)$:

$$u_k(x+T) = \Sigma\, C(k-G)\, e^{-iG(x+T)} = e^{-iGT}[\Sigma\, C(k-G)\, e^{-iGx}] = e^{-iGT}\, u_k(x) \quad .$$

Because $\exp(-iGT) = 1$ by (2.17), it follows that $\underline{u_k(x+T) = u_k(x)}$, thereby establishing the periodicity of u_k. This is an alternate and exact proof of the Bloch theorem and is valid even when the ψ_k are degenerate.

Crystal momentum of an Electron

- Under a crystal lattice translation which carries $\mathbf{r}$ to $\mathbf{r} + \mathbf{T}$ we have

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{T}) = e^{i\mathbf{k}\cdot\mathbf{T}}\, e^{i\mathbf{k}\cdot\mathbf{r}}\, u_{\mathbf{k}}(\mathbf{r} + \mathbf{T}) = e^{i\mathbf{k}\cdot\mathbf{T}}\psi_{\mathbf{k}}(\mathbf{r}) \ , \tag{30}$$

Thus $\exp(i\mathbf{k}\cdot\mathbf{T})$ is the phase factor[1] by which a Bloch function is multiplied when we make a crystal lattice translation $\mathbf{T}$.

- If the lattice potential vanishes, **U= 0, then from eq. (27)** $(\lambda_{\mathbf{k}} - \epsilon)C(\mathbf{k}) = 0$, so that all $C(\mathbf{k} - \mathbf{G})$ are zero except $C(\mathbf{k})$, and thus $u_{\mathbf{k}}(\mathbf{r})$ is constant. We have $\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$, just as for a free electron.

- Thus $\hbar\mathbf{k}$ is called the **crystal momentum** of an electron. If an electron $\mathbf{k}$ absorbs in a collision a phonon of wavevector $\mathbf{q}$, the selection rule is $\mathbf{k} + \mathbf{q} = \mathbf{k}' + \mathbf{G}$. In this process the electron is scattered from a state $\mathbf{k}$ to a state $\mathbf{k}'$, with $\mathbf{G}$ a reciprocal lattice vector.

Solution of the Central Equation

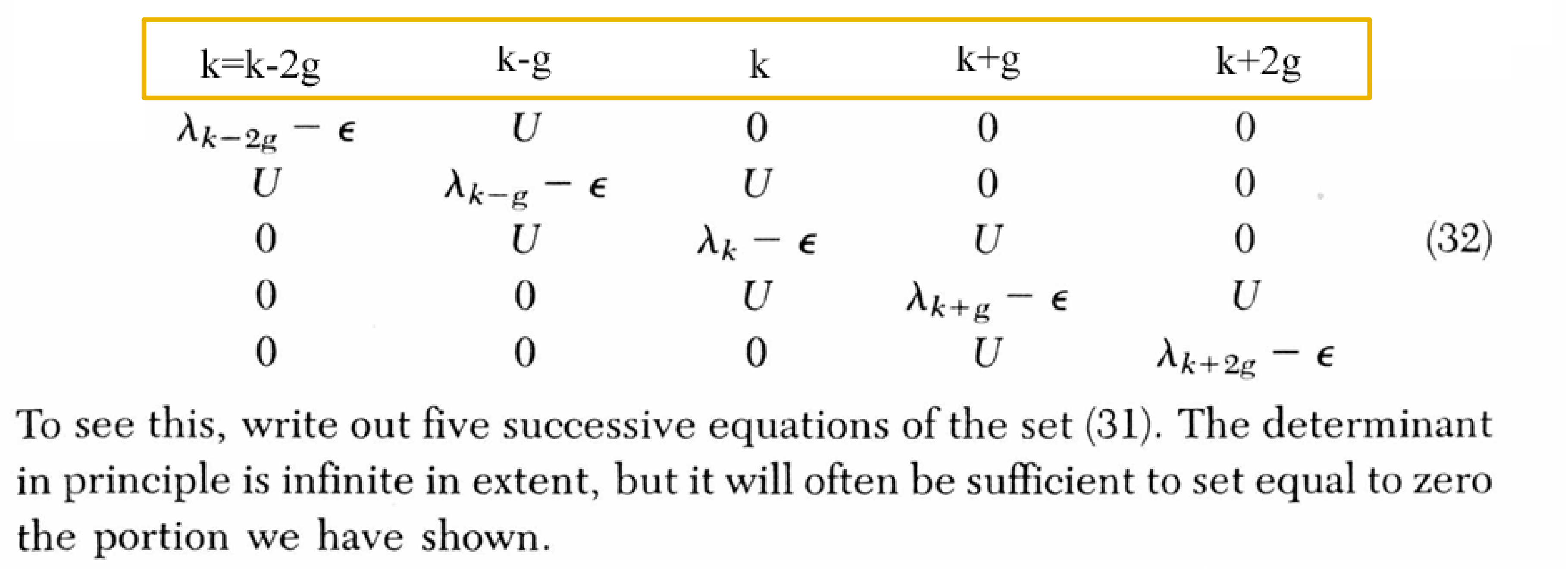
Equation (27) may be called the central equation:

$$\boxed{(\lambda_k - \epsilon)C(k) + \sum_G U_G C(k-G) = 0} \qquad (31)$$

represents a set of simultaneous linear equations that connect the coefficients $C(k-G)$ for all reciprocal lattice vectors G. It is a set because there are as many equations as there are coefficients C. These equations are consistent if the determinant of the coefficients vanishes.

Let us write out the equations for an explicit problem. We let g denote the shortest G. We suppose that the potential energy ***U*(x)** contains only a single

Fourier component ***Ug* = *U-g*** denoted by ***U***. Then a block of the determinant of the coefficients is given by:



$$\begin{array}{ccccc} k{=}k\text{-}2g & k\text{-}g & k & k{+}g & k{+}2g \\ \lambda_{k-2g}-\epsilon & U & 0 & 0 & 0 \\ U & \lambda_{k-g}-\epsilon & U & 0 & 0 \\ 0 & U & \lambda_k-\epsilon & U & 0 \\ 0 & 0 & U & \lambda_{k+g}-\epsilon & U \\ 0 & 0 & 0 & U & \lambda_{k+2g}-\epsilon \end{array} \qquad (32)$$

To see this, write out five successive equations of the set (31). The determinant in principle is infinite in extent, but it will often be sufficient to set equal to zero the portion we have shown.

At a given ***k***, each root ϵ or ϵ_k lies on a different energy band, except in case of coincidence. The solution of the determinant (32) gives a set of energy eigenvalues ϵ_{nk}, where ***n*** is an index for ordering the energies, and ***k*** is the wave-vector that labels ***Ck***.

Most often, ***k*** will be taken in the first zone.