Queen's University Belfast

School of Mathematics and Physics

PHY3012 SOLID STATE PHYSICS

A T Paxton, November 2012

Books

The primary textbook for this course is
H Ibach and H Lüth, *Solid State Physics,* Springer, 4th Edition, 2009

Additionally, I recommend
J R Hook and H E Hall, *Solid State Physics,* Wiley, 2nd Edition, 2000
S Elliott, *The Physics and Chemistry of Solids,* Wiley, 1998
C Kittel, *Introduction to Solid State Physics,* Wiley, 8th Edition, 2005

Slightly more advanced texts that I recommend are
M P Marder, *Condensed Matter Physics,* Wiley, 2000
N Ashcroft and N D Mermin, *Solid State Physics,* Holt-Sanders, 1976

Contents

There follows a list of the most commonly used quantities, with occasional comparison with the usage in the textbooks, Ibach and Lüth (IL), Kittel (K), Hook and Hall (HH) and Elliott (E). I will always put vectors in bold type face, and their magnitude in normal typeface, *e.g.*, $p = |\mathbf{p}|$. The complex conjugate of a complex number is indicated by a bar over it, for example, $\bar{\psi}$.

a	spacing between lattice planes, lattice constant, acceleration	
$\mathbf{a}_i$	basis vectors of the direct lattice	$\mathbf{a}, \mathbf{b}, \mathbf{c}$ (HH, E)
a, a^+	destruction and creation operators for phonons	
A	force constant matrix, cubic coefficient in heat capacity, exchange integral	
$\mathbf{b}_i$	basis vectors of the reciprocal lattice	$\mathbf{g}_i$ (IL), $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ (HH, E)
B	magnetic induction	
c	speed of sound	
C	spring constant, heat capacity	
C_e	heat capacity due to electrons	
c_e	heat capacity per unit volume due to electrons	
C_ph	heat capacity due to phonons	
c_ph	heat capacity per unit volume due to phonons	
e	charge on the electron is $-e$	
e	base of the natural logarithms	
E	eigenvalue, especially of an electron state	E (IL), ϵ (K), ε (HH), $\mathcal{E}$ (E)
E_F	Fermi energy	
E_c	energy at conduction band edge	
E_v	energy at valence band edge	
E_g	semiconductor energy gap	
$\boldsymbol{\mathcal{E}}$	electric field	$\boldsymbol{\mathcal{E}}$ (IL), $\mathbf{E}$ (HH, K, E)
$f(E), f(\mathbf{k}), f_\mathrm{FD}$	Fermi–Dirac function	
F	force	
$\mathbf{g}$	reciprocal lattice vector	$\mathbf{G}$ (IL, K, E), $\mathbf{G}_{hkl}$ (HH)
$\mathrm{g_e}$	electron g-factor	
$\hbar$	reduced Planck constant $\hbar = h/2\pi$	
h	hopping integral	
H	hamiltonian, magnetic field	
i	$\sqrt{-1}$	
i, j, l	labels of atoms	
I	Stoner parameter	
$\mathbf{J}_\mathrm{e}$	electric current density	
$\mathbf{J}_\mathrm{Q}$	heat current density	
$\mathbf{k}$	wavevector, vector in reciprocal space	
k_B	Boltzmann constant	
k_F	Fermi wavevector	
L	Lagrangian, length	
ℓ, m, n	direction cosines	
m	mass of the electron, magnetic moment per atom / μ_B $m = n_\uparrow - n_\downarrow$	
m^*	effective mass	
M	mass of an atom	

n, m, p	integers	
$n(\omega), n(E)$	density of states	D (K), $V \cdot D$ (IL), g (E, HH)
$n_s(E)$	density of states per spin per atom	
n_e	number of electrons per unit volume, electron density	
n_h	number of holes per unit volume, hole density	
$n_\uparrow, n_\downarrow$	number of up or down spin electrons per atom	
N_c	number of unit cells in the crystal	
N_at	number of atoms in the crystal	
N_e	number of electrons in the crystal	
N_s, N_d	number of s, d-electrons per atom	
$\mathbf{p}$	momentum	
$\tilde{\mathbf{q}}_\mathbf{k}$	polarisation vector (vector amplitude)	
q, q_k	position coordinate; normal mode coordinate	
$\mathbf{q}$	scattering vector	$-\mathbf{K}$ (IL)
$\mathbf{r}$	a vector in direct space	
$\mathbf{r}_j, \mathbf{R}$	a vector denoting the crystallographic position of atom j	$\mathbf{r}$ (IL)
$\delta\mathbf{r}_j$	the vector displacement of atom j from its equilibrium position	$\mathbf{u}$ (IL, K, E, HH)
S	overlap integral, entropy	
$\mathcal{S}$	thermopower	
t	time	
T	absolute temperature	
T_F	Fermi temperature	
U	internal energy	E (HH)
v_F	Fermi velocity	
$\mathbf{v}_d$	drift velocity	
$\mathbf{v}, v_g$	group velocity	
V	volume	
W	potential energy, d-band width	
α	index, expansion coefficient, variational parameter, spin function	
β	spin function	
γ	linear coefficient in heat capacity	
δ_{ij}	Kronecker delta	
ϵ_0	permittivity of free space ($\epsilon_0\mu_0 = 1/c^2$)	
θ_D	Debye temperature	
θ_E	Einstein temperature	
κ	thermal conductivity	
λ	wavelength, mean free path	
μ	chemical potential, carrier mobility	
μ_0	permeability of free space ($4\pi \times 10^{-7}$ N A^{-2})	
μ_B	Bohr magneton	
μ_e	electron moment $\approx \mu_B$	
$\mathit{\Pi}$	Peltier coefficient	
σ	conductivity	
τ	relaxation time	
χ	magnetic susceptibility	
ψ, ϕ, φ	wavefunction	
ω	angular frequency	
ω_D	Debye frequency	
ω_E	Einstein frequency	

Introduction

We are concerned with the physics of matter when it is in the solid, usually crystalline, state. You are faced with the complex problem of describing and understanding the properties of an assembly of some $10^{23} - 10^{24}$ atoms, closely bonded together. The subject can be divided into two sub disciplines.

1. The nature of the bonding (mechanical properties *etc.*)
2. The physics of solids (spectroscopy, energy bands, transport properties *etc.*)

We will deal largely with the second. Much of the physics of solids can be understood by regarding the solid as matter containing certain *elementary excitations*. These interact with externally applied fields (including electromagnetic radiation) as well as each other to produce the observable phenomena which we measure. Examples of elementary excitations are

electrons and holes
phonons
plasmons
magnons and spin waves

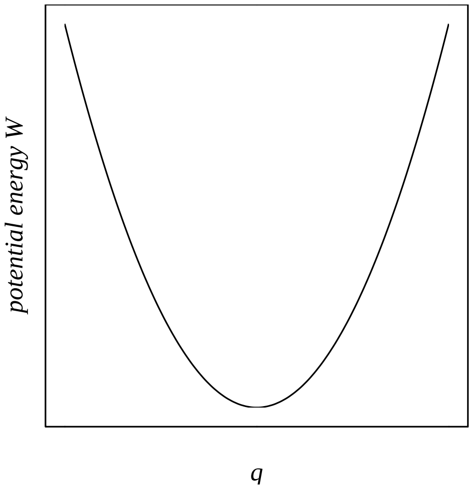
We will study electrons and phonons. Generally we will describe the mechanics of a single particle at a time. Many particle physics is usually beyond the scope of an undergraduate physics programme.

0. Lagrangian and Hamiltonian mechanics

A body with mass M moves in one dimension. The position coordinate at time t is denoted q. By Newton's second law, we have

$$M\frac{\mathrm{d}^2q}{\mathrm{d}t^2} \equiv M\ddot{q} = F(q)$$

where $F(q)$ is the force exerted on the body when it is at position q. We take it that this force arises from the fact that the body is moving in some potential W. For example, here is a simple harmonic potential, $W \propto q^2$:



The force is, therefore,

$$F(q) = -\frac{\mathrm{d}W}{\mathrm{d}q}$$

and the kinetic energy is

$$K = \frac{1}{2}mv^2 = \frac{1}{2}M\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)^2 \equiv \frac{1}{2}M\dot{q}^2$$

where v is the velocity of the body.

We define a function called the Lagrangian as

$$L(\dot{q}, q) = K - W \tag{0.1}$$

the difference between the kinetic and potential energies. If we differentiate L with respect to $\dot{q}$ we get

$$\frac{\partial L}{\partial \dot{q}} = \frac{\mathrm{d}}{\mathrm{d}\dot{q}}\left(\frac{1}{2}M\dot{q}^2\right) = M\dot{q} \tag{0.2}$$

from which it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} = M\ddot{q} \qquad \text{(mass} \times \text{acceleration)}$$

On the other hand differentiating L with respect to q you find

$$\frac{\partial L}{\partial q} = -\frac{\mathrm{d}W}{\mathrm{d}q} = F \qquad \text{(force)}$$

Clearly, then, Newton's second law (force = mass × acceleration) can be written in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0$$

Whereas the Lagrangian is an energy functional that allows us to deduce the equation of motion, the Hamiltonian allows us to write down the *total energy* in terms of q and $\dot{q}$,

$$H = K + W$$

Normally, though, we write

$$K = \frac{p^2}{2M}$$

where p is the momentum "conjugate" to q, and $p = Mv = M\dot{q}$ is equivalent to writing

$$p = \frac{\partial L}{\partial \dot{q}}$$

from equation (0.2). In order to go over to *quantum* mechanics we do just two things.

1. Make the substitution

$$p \longrightarrow -\mathrm{i}\hbar\frac{\partial}{\partial q}$$

   from which appears the commutation relation

$$[q, p] = \mathrm{i}\hbar$$

   The conjugate observables p and q are known as *complementary* in quantum mechanics.

2. The resulting *Hamiltonian operator*

$$\begin{aligned} H &= \frac{p^2}{2M} + W \\ &= \frac{\hbar^2}{2M}\frac{\partial^2}{\partial q^2} + W \end{aligned}$$

   is then put into the time dependent Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\psi = H\psi$$

   which is the equation of motion of a quantum particle described by the *wavefunction* ψ. Stationary states of the particle are described by the time independent Schrödinger equation

$$H\psi = E\psi$$

   and ψ is then an *eigenfunction* of H having the *eigenvalue* E.

1. Lattice vibrations and Phonons

Phonons are *quantised* lattice vibrations. We will pursue the following steps.

1. Set up the problem of lattice waves using classical, Newtonian mechanics and the elementary properties of waves.
2. Solve the problem for a simple model. This is a typical approach in solid state physics.
3. Describe the important consequences of crystal periodicity.
4. *Quantise* the waves.

1.1 Vibrations in the monatomic lattice

Vibrations of the atoms in a crystal take the form of waves, standing or travelling, whose amplitude is the largest displacement of an atom from its equilibrium position. In addition to its amplitude a wave is characterised by its wavevector, $\mathbf{k}$, and angular frequency, $\omega = 2\pi \times$ frequency.

The atoms' official, crystallographic positions are denoted $\mathbf{r}_j$. j runs from 1 to N_c, where N_c is the number of unit cells in the crystal; and to make the notation simple, we consider the case where there is only one atom per unit cell. The atomic displacements are denoted $\delta\mathbf{r}_j$ ('δ' meaning 'a small change in...') and can be written

$$\delta\mathbf{r}_j = \frac{1}{\sqrt{N_\mathrm{c}}}\,\tilde{\mathbf{q}}_\mathbf{k}\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}_j - \omega t)} \qquad (1.1.1)$$

$\tilde{\mathbf{q}}_\mathbf{k}$ is called the *polarisation vector* or *vector amplitude*; $1/\sqrt{N_\mathrm{c}}$ is only for normalisation. The wavelength is

$$\lambda = \frac{2\pi}{k}$$

The angular frequency is a function of the wavelength, thus

$$\omega = \omega(\mathbf{k})$$

and it is essentially this "dispersion relation" we shall be looking for; t is the time.

These waves are excited by temperature or externally applied fields or forces. The crystal has lowest potential energy when all the $\delta\mathbf{r}_j$ are zero. We call this energy

$$W = W(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_\mathrm{c}})$$

and then expand the potential energy of the crystal with lattice waves in a Taylor series,

$$U_\mathrm{pot} = W + \sum_j \frac{\partial W}{\partial \mathbf{r}_j}\delta\mathbf{r}_j + \frac{1}{2}\sum_{ij}\delta\mathbf{r}_i\,\frac{\partial^2 W}{\partial\mathbf{r}_i\partial\mathbf{r}_j}\,\delta\mathbf{r}_j + \ldots \qquad (1.1.2)$$

where the partial derivatives are evaluated at the equilibrium positions of the atoms.[†] Now the first term can be taken as zero, and this *defines* our zero of energy as the potential energy of the crystal at rest. The second term is also zero because the crystal is in equilibrium with respect to small atomic displacements. If we truncate the Taylor series after the second order terms we are said to be working within the harmonic approximation.

Let us write down the hamiltonian in this approximation,

$$\begin{aligned} H &= \text{kinetic energy} + \text{potential energy} \\ &= \sum_j \frac{p_j^2}{2M} + \frac{1}{2}\sum_{ij} A_{ij}\delta\mathbf{r}_i\delta\mathbf{r}_j \end{aligned} \tag{1.1.3}$$

$A_{ij} = \frac{\partial^2 W}{\partial \mathbf{r}_i \partial \mathbf{r}_j}$ is called the force constant matrix. It is the change in potential energy when atom i is moved to $\mathbf{r}_i + \delta\mathbf{r}_i$ and atom j is simultaneously moved moved to $\mathbf{r}_j + \delta\mathbf{r}_j$. M is the mass of each atom. (Note, A_{ij} really should have further indices indicating the directions in which atoms i and j are moved. A typical matrix element is $A_{ix,jy}$ for example. I can leave these out for simplicity and anyway we'll only end up dealing with one-dimensional cases.)

The force constant matrices can only be determined if we have a complete knowledge of the bonding in the solid.

Next, we want the equation of motion for atom j. In classical, hamiltonian mechanics, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_j = -\frac{\partial H}{\partial \mathbf{r}_j} = M\frac{\mathrm{d}^2}{\mathrm{d}t^2}\delta\mathbf{r}_j$$

rate of change of momentum = *force acting upon atom j* = *mass* × *acceleration*

Differentiating equation (1.1.3) we get

$$M\frac{\mathrm{d}^2}{\mathrm{d}t^2}\delta\mathbf{r}_j = -\frac{\partial H}{\partial \mathbf{r}_j} = -\sum_i A_{ij}\delta\mathbf{r}_i$$

We can easily differentiate equation (1.1.1) twice with respect to time:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\delta\mathbf{r}_j = -\omega^2\frac{1}{\sqrt{N_\mathrm{c}}}\,\tilde{\mathbf{q}}_\mathbf{k}\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}_j - \omega t)}$$

[†] As a matter of notation, when I write, say,

$$\sum_i \frac{\partial W}{\partial \mathbf{r}_i}\,\delta\mathbf{r}_i$$

I really mean

$$\sum_{i\alpha} \frac{\partial W}{\partial r_{i\alpha}}\,\delta r_{i\alpha}$$

and $\alpha = 1, 2, 3$ (or x, y, z) is an index to each component of the vector.

We now insert equation (1.1.1) for $\delta\mathbf{r}_j$ and equate the two right hand sides:

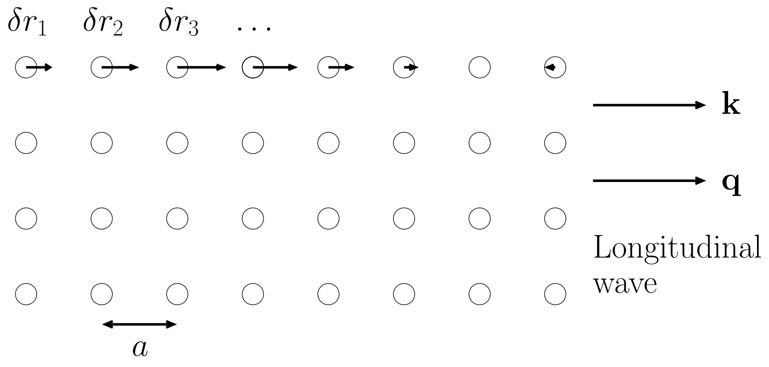
$$\sum_i A_{ij} \frac{1}{\sqrt{N_\mathrm{c}}}\, \tilde{\mathbf{q}}_\mathbf{k}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}_i - \omega t)} = M\omega^2 \frac{1}{\sqrt{N_\mathrm{c}}}\, \tilde{\mathbf{q}}_\mathbf{k}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}_j - \omega t)}$$

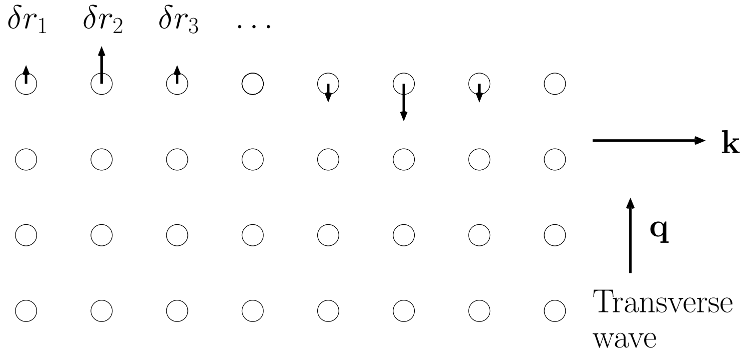
Cancel $\mathrm{e}^{-\mathrm{i}\omega t}/\sqrt{N_\mathrm{c}}$ from both sides and our equation of motion for atom j is, finally,

$$M\omega^2 \tilde{\mathbf{q}}_\mathbf{k} = \sum_i A_{ij}\, \tilde{\mathbf{q}}_\mathbf{k}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_j - \mathbf{r}_i)} \tag{1.1.4}$$

I have left in the $\tilde{\mathbf{q}}_\mathbf{k}$ on either side to emphasise that there is an equation of motion for each of the three directions of polarisation of the wave. If there were more than one atom per unit cell, then there would be an equation like (1.1.4) for each atom type.

In general we can't solve this to get the dispersion relation unless we know the force constants, which is a hard problem. But we can invoke a simple model for the crystal's force constants for which we can solve the equation of motion. Consider a cubic crystal with one atom per unit cell. We also consider those 'modes' of vibration whose wavevector is perpendicular to a simple lattice plane and for which the displacement is either parallel or perpendicular to the wavevector. Under these circumstances whole planes of atoms move in unison and the problem becomes confined to one dimension (see Kittel, chapter 5).





We only need to solve equation (1.1.4) for one of the N_c atoms in the crystal since they are equivalent by translational symmetry. Of course, as you can see in the diagram, the displacement of an atom is not the same as one of its neighbours at any one time, but they differ only in their phase. If a is the lattice constant then

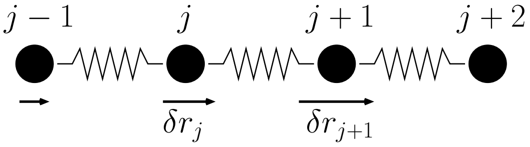
$$r_{j+1} = r_j + a.$$

Inserting this into equation (1.1.1), I immediately get

$$\delta r_{j+1} = \delta r_j \, \mathrm{e}^{\mathrm{i}ka} \tag{1.1.5}$$

which is a mathematical statement of the same assertion.

The simple model that we can solve requires us to suppose that there are only forces between neighbouring planes of atoms. This is a lousy approximation, but it will serve to illustrate much of the essential physics of lattice waves. It's as if the atoms were connected by springs with a spring constant (stiffness) C:



In this diagram, each solid dot represents a whole plane of atoms.

Considering only neighbouring planes, there will be contributions to the potential energy, proportional to the difference in displacement squared, from each spring.

$$U_{\mathrm{pot}}^{j,j+1} = C\left(\delta r_j - \delta r_{j+1}\right)^2$$

We sum this over all planes and divide by two (otherwise we've counted each plane twice)

$$\begin{aligned}
U_{\mathrm{pot}} &= \frac{1}{2}C\sum_j \left(\delta r_j - \delta r_{j+1}\right)^2 \\
&= \frac{1}{2}C\sum_j \left(\delta r_j^2 + \delta r_{j+1}^2 - 2\delta r_j \delta r_{j+1}\right) \\
&= C\sum_j \left(\delta r_j \, \delta r_j - \delta r_j \delta r_{j+1}\right)
\end{aligned}$$

You can see how I got from the second to the third line by noticing that the first two terms in the second line are identical—a sum of squares of all the δr_j. Now compare the above equation with equation (1.1.3) which states, in one dimension,

$$U_{\mathrm{pot}} = \frac{1}{2}\sum_{ij} A_{ij} \, \delta r_i \, \delta r_j$$

You see that we can identify <u>all</u> the A_{ij} for our simple model:

$$\begin{gathered}
A_{ii} = 2C \\
A_{i,i+1} = A_{i,i-1} = -C \\
\text{all other } A_{ij} = 0
\end{gathered}$$

so we can solve the equation of motion (1.1.4) for longitudinal modes

$$M\omega_L^2 = \sum_i A_{ij}\,\mathrm{e}^{\mathrm{i}k(r_j - r_i)}$$

There are only three non zero terms on the right hand side; k is parallel to r, and let us write the distance between neighbouring atoms as $r_j - r_i = a$, the lattice constant. We get

$$M\omega_L^2 = C\left(2 - \mathrm{e}^{\mathrm{i}ka} - \mathrm{e}^{-\mathrm{i}ka}\right)$$

and therefore by elementary trigonometry, and remembering that $\mathrm{e}^{i\theta} = \cos\theta + \mathrm{i}\sin\theta$

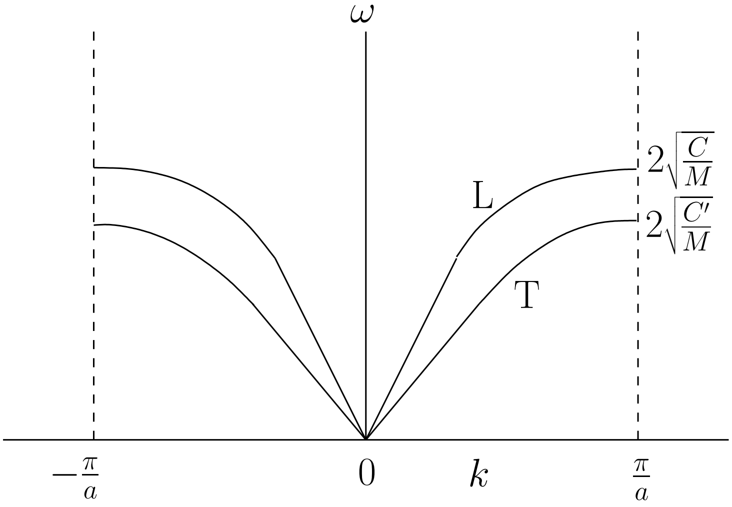
$$\boxed{\omega_L = \sqrt{\frac{C}{M}}\,2\left|\sin\frac{1}{2}ka\right|}$$

This is the dispersion relation for longitudinal waves in our model crystal. The vertical bars denote "absolute value," and we don't expect ω to be negative.

Transverse waves will have exactly the same solution but with a different, presumably smaller, spring constant C'

$$\omega_T = \sqrt{\frac{C'}{M}}\,2\left|\sin\frac{1}{2}ka\right|$$

We can sketch our dispersion relation, roughly like this:



1.2 Notes on section 1.1

Here are some very important points.

1. As $k \to 0$ (in the long wavelength limit) we have

   $$\omega_T = \sqrt{\frac{C'}{M}}\, ka$$
   $$\omega_L = \sqrt{\frac{C}{M}}\, ka$$

   The quantity $\mathrm{d}\omega/\mathrm{d}k$ is called the group velocity of the wave and for long waves these are the velocities of sound waves in the crystal. The slope of the dispersion is related to the elastic constants of the crystal. This is why you measure elastic constants in single crystals by measuring the velocities of longitudinal and transverse sound waves.

2. As $k \to \pi/a$ the velocity tends to zero. We say that "at the Brillouin zone boundary" the lattice waves are standing waves; they have been "Bragg reflected."

3. I have only plotted $\omega(k)$ in the range of k between $-\pi/a$ and π/a. If k is between $-\pi/a$ and π/a it is said to be "within the first Brillouin zone." There is no physical significance to k-values outside this range. For if I consider eq (1.1.1),

   $$\delta r_j = \frac{1}{\sqrt{N_\mathrm{c}}}\, \tilde{q}_k\, \mathrm{e}^{\mathrm{i}(nka-\omega t)}$$

   where $r_j = na$, and add to k an amount $p \times 2\pi/a$ where p is an integer, I get

   $$\begin{aligned}\delta r_j &= \frac{1}{\sqrt{N_\mathrm{c}}}\, \tilde{q}_k\, \mathrm{e}^{\mathrm{i}(nk+2\pi p/a)a}\, \mathrm{e}^{-\omega t} \\ &= \frac{1}{\sqrt{N_\mathrm{c}}}\, \tilde{q}_k\, \mathrm{e}^{\mathrm{i}(nka-\omega t)}\, \mathrm{e}^{\mathrm{i}2\pi p}\end{aligned}$$

   I have simply multiplied by one! $\left(\mathrm{e}^{\mathrm{i}2\pi p} = 1\right)$. I get no new displacements, no new physics. It is important to understand this point. The keen-eyed will have noted that I've *assumed* that $q_{k+2\pi p/a} = q_k$; we can prove this but not until section 1.6.2 when we do the Fourier transformation of the displacements.

   Another way to see this is to remark that eq (1.1.1) is a wave *only defined* at atom positions. The atomic displacement is not a continuous function of position in the usual sense. (See, for example, fig. 5, chapter 4 in Kittel; or p. 222 in Elliott.)

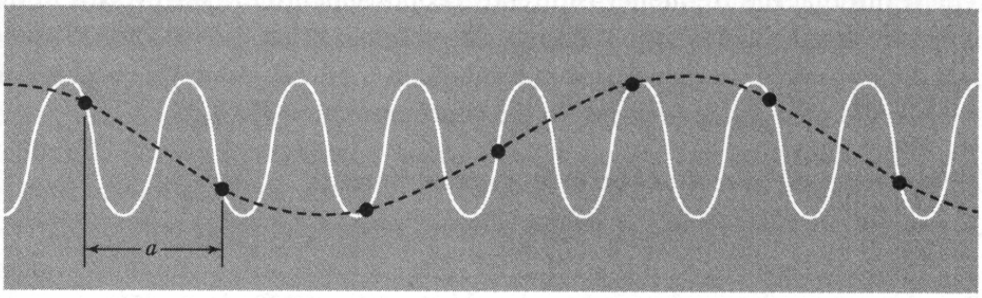


Figure 5 The wave represented by the solid curve conveys no information not given by the dashed curve. Only wavelengths longer than $2a$ are needed to represent the motion.

Both waves carry the same information about the atom displacements. Only wavelengths longer than $2a$ are needed to represent the motion.

4. Only lattice waves with wavevectors in the first Brillouin zone are physically meaningful. Are all values of k in this range allowed? No. There is a finite number that depends on the size of the crystal, but not on the boundary conditions. *What does this mean?*

   Suppose our crystal has length L in the direction of k and $L = N_{\mathrm{at}}a$ with $N_{\mathrm{at}} = N_{\mathrm{c}}$ the number of planes. We are essentially considering a one dimensional line of atoms.

What shall we say about the displacements at the surfaces of the crystal, *i.e.*, atoms 1 and N_{at}? We could clamp them (*i.e.*, set $\delta r_1 = \delta r_{N_{\mathrm{at}}} = 0$) rather like the boundary conditions on the wavefunctions in an infinite potential well (*i.e.*, $\psi(r_1) = \psi(r_{N_{\mathrm{at}}}) = 0$).

Or, we could impose *cyclic*, or *periodic*, or *Born–von Kármán* boundary conditions; *i.e.*, set

$$\delta r_1 = \delta r_{N_{\mathrm{at}}} \qquad \text{not necessarily zero}$$

According to eq (1.1.5)

$$\delta r_{j+1} = \delta r_j \, \mathrm{e}^{\mathrm{i}ka}$$

therefore

$$\delta r_{j+2} = \delta r_{j+1} \, \mathrm{e}^{\mathrm{i}ka} = \delta r_j \, \mathrm{e}^{\mathrm{i}2ka}$$

and if I keep this up, we get

$$\begin{aligned} \delta r_{N_{\mathrm{at}}} &= \delta r_1 \, \mathrm{e}^{\mathrm{i}N_{\mathrm{at}}ka} = \delta r_1 \, \mathrm{e}^{\mathrm{i}kL} \\ &= \delta r_1 \qquad \text{according to our boundary conditions} \end{aligned}$$

Thus

$$\mathrm{e}^{\mathrm{i}kL} = 1$$

which is only true if

$$kL = 2\pi n \tag{1.1.6}$$

where n is any integer, and so the allowed values of k are those for which

$$k = \frac{2\pi n}{L}$$

lying between $-\pi/a$ and π/a.

They form a dense set of points separated by an interval $2\pi/L$ which decreases as the size of the crystal increases. But there is a finite number of allowed values of k (allowed states).

We can write down the number of states per unit k from (1.1.6)

$$\frac{\mathrm{d}n}{\mathrm{d}k} = \frac{L}{2\pi}$$

and the total number of states is this times the allowed range of k,

$$\frac{\mathrm{d}n}{\mathrm{d}k} \times \frac{2\pi}{a} = \frac{L}{2\pi} \times \frac{2\pi}{a} = \frac{L}{a} = N_{\mathrm{at}}$$

The number of allowed states is equal to the number of atoms in the line.

Very often we want to know the number of states in a range of frequency, or the density of states,

$$\begin{aligned} n(\omega) &= \frac{\mathrm{d}n}{\mathrm{d}\omega} = \frac{\mathrm{d}n}{\mathrm{d}k} \times \frac{\mathrm{d}k}{\mathrm{d}\omega} \\ &= \frac{L}{2\pi} \times \frac{1}{\mathrm{d}\omega/\mathrm{d}k} \end{aligned}$$

$\mathrm{d}\omega/\mathrm{d}k$ is the group velocity and is only known once the we know the dispersion relations.

1.3 Boundary conditions and density of states in three dimensions

Density of states is a very important concept in solid state physics for all elementary excitations as is the evaluation of the number of states, so let's do the thing again, now in 3-D.

What do we imply by imposing periodic boundary conditions? In general we deal with a piece of crystal of some shape, bounded by a number of crystal surfaces. But bulk properties must be independent of the details of the surface and its environment. Therefore

I can surely concentrate upon a part of the whole crystal, whose shape may as well be the shape of the crystallographic unit cell, buried deep inside the specimen. Any bulk property cannot depend upon the choice of boundary conditions that I impose on this "embedded" crystal as long as it is of macroscopic dimensions. Also these bulk properties cannot depend upon details of the specimen surfaces as long as these are sufficiently far from the "embedded" crystal.

Therefore suppose the "embedded" crystal contains N_c unit cells and has the same shape as a primitive unit cell. If the lattice vectors are $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, the crystal dimensions are

$$N_1\,\mathbf{a}_1 \times N_3\,\mathbf{a}_2 \times N_3\,\mathbf{a}_3$$

Now, atoms separated by any lattice vector $(n_1\,\mathbf{a}_1,\ n_2\,\mathbf{a}_2,\ n_3\,\mathbf{a}_3)$ are indistinguishable by translational symmetry in terms of any measurable physical quantity. This means that the displacements arising from a lattice wave can differ by, at most, a phase factor as we have seen in eq (1.1.5).

This is true also for electron wavefunctions. If $\psi_\mathbf{k}(\mathbf{r})$ is the wavefunction of an electron at $\mathbf{r}$, with wavevector $\mathbf{k}$, we must have

$$\begin{aligned}
\psi_\mathbf{k}(\mathbf{r}+n_1\mathbf{a}_1) &= \psi_\mathbf{k}(\mathbf{r})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot n_1\mathbf{a}_1}\\
\psi_\mathbf{k}(\mathbf{r}+n_2\mathbf{a}_2) &= \psi_\mathbf{k}(\mathbf{r})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot n_2\mathbf{a}_2}\\
\psi_\mathbf{k}(\mathbf{r}+n_3\mathbf{a}_3) &= \psi_\mathbf{k}(\mathbf{r})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot n_3\mathbf{a}_3}
\end{aligned}$$

(NB, this is the only way to ensure that

$$|\psi_\mathbf{k}(\mathbf{r}+n_1\mathbf{a}_1)|^2 = |\psi_\mathbf{k}(\mathbf{r})|^2 \qquad \textit{etc.}$$

which is the observable probability density.)

Periodic boundary conditions require that the displacements or wavefunctions are equal at opposite faces of the "embedded" crystal:

$$\begin{aligned}
\psi_\mathbf{k}(\mathbf{r}+N_1\mathbf{a}_1) &= \psi_\mathbf{k}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}N_1\mathbf{k}\cdot\mathbf{a}_1}\psi_\mathbf{k}(\mathbf{r})\\
\psi_\mathbf{k}(\mathbf{r}+N_2\mathbf{a}_2) &= \psi_\mathbf{k}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}N_2\mathbf{k}\cdot\mathbf{a}_2}\psi_\mathbf{k}(\mathbf{r})\\
\psi_\mathbf{k}(\mathbf{r}+N_3\mathbf{a}_3) &= \psi_\mathbf{k}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}N_3\mathbf{k}\cdot\mathbf{a}_3}\psi_\mathbf{k}(\mathbf{r})
\end{aligned}$$

We see that the exponentials must equal unity so that

$$\left.\begin{aligned}
N_1\mathbf{k}\cdot\mathbf{a}_1 &= 2\pi m_1\\
N_2\mathbf{k}\cdot\mathbf{a}_2 &= 2\pi m_2\\
N_3\mathbf{k}\cdot\mathbf{a}_3 &= 2\pi m_3
\end{aligned}\right\}\mathbf{k}\cdot\mathbf{a}_j = 2\pi\sum_i \frac{m_i}{N_i}\delta_{ij}$$

m_1, m_2 and m_3 being integers. (On the right is just a neat compact way of writing the three equations on the left.) Now remember the definition of the reciprocal lattice vectors, $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$,

$$\mathbf{b}_i\cdot\mathbf{a}_j = 2\pi\delta_{ij}$$

in which δ_{ij} is the “Kronecker delta” (equal to one if $i = j$ and zero otherwise).

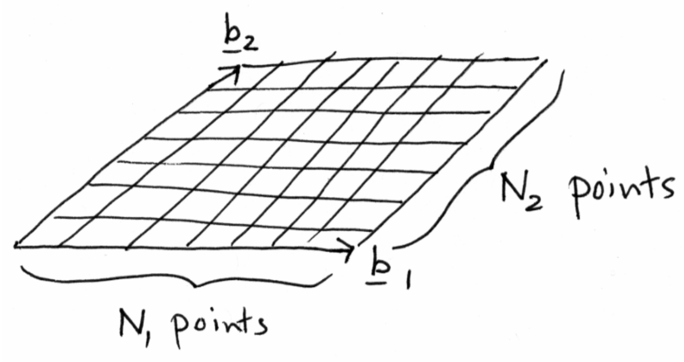
To satisfy the three conditions, **k** must be some fraction of a reciprocal lattice vector

$$\mathbf{k} = \frac{m_1}{N_1}\mathbf{b}_1 + \frac{m_2}{N_2}\mathbf{b}_2 + \frac{m_3}{N_3}\mathbf{b}_3$$

You can see that this does indeed meet the three conditions—take the scalar product of **k** with any of the direct lattice vectors:

$$\mathbf{k}\cdot\mathbf{a}_j = 2\pi \sum_i \frac{m_i}{N_i}\delta_{ij}$$

Thus the allowed values of **k** form a *fine mesh of points in the reciprocal lattice.*



But we also know that **k** must remain within the first Brillouin zone and this has the same volume in **k**-space as the parallelepiped formed from the vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$.

Hence all the allowed values of m_1, m_2 and m_3 are given by the conditions

$$0 < m_1 \le N_1$$
$$0 < m_2 \le N_2$$
$$0 < m_3 \le N_3$$

Therefore the number of allowed states is

$$N_1 N_2 N_3 = N_{\mathrm{c}}$$

the *number of unit cells in the crystal.* We already got this resut in the 1-D case.

Let us denote the volume of the unit cell by v_{c}. The volume of the “embedded” crystal is

$$V = N_{\mathrm{c}} v_{\mathrm{c}}$$

and the volume of the first Brillouin zone is

$$(2\pi)^3 \frac{1}{v_{\mathrm{c}}}$$

If there are N_c $\mathbf{k}$ vectors allowed in the first Brillouin zone the "volume per state" in $\mathbf{k}$-space is

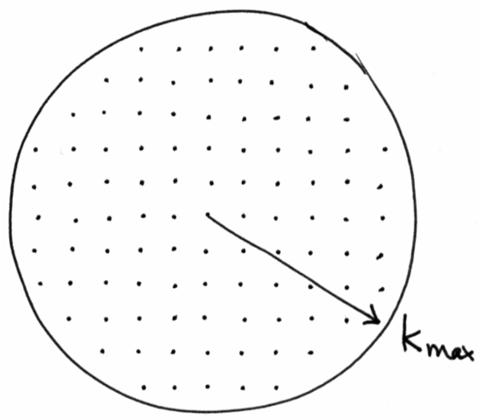
$$\frac{1}{N_c}\frac{(2\pi)^3}{v_c} = \frac{(2\pi)^3}{V}$$

The "spottiness of $\mathbf{k}$-space" is such that the allowed states are a mesh of points each occupying a volume

$$\frac{(2\pi)^3}{V}$$

(Compare this with the result $\Delta k = 2\pi/L$ that we got in 1D.)

The number of allowed states having wavevector magnitude smaller than some $k_{\max}$ is the number of states enclosed by a sphere of radius $k_{\max}$.

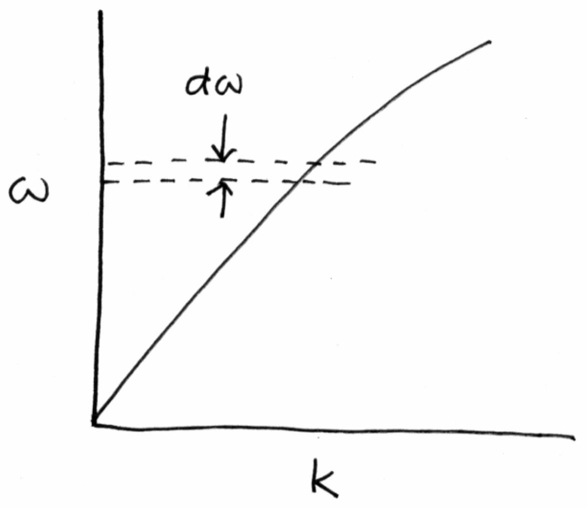


That is,

number of allowed states with wavevector less than $k_{\max}$	=	volume of sphere	×	number of states per unit volume

$$\begin{aligned} N_{\mathrm{st}} &= \frac{4\pi}{3}\,k_{\max}^3\,\frac{V}{(2\pi)^3} \\ &= \frac{1}{6}\,\frac{1}{\pi^2}\,V k_{\max}^3 \end{aligned} \qquad (1.3.1)$$

We define the density of states for lattice waves in terms of the number of states in a narrow range of frequencies dω about ω.



This number is, clearly, inversely proportional to the slope of the dispersion curve. Now,

$$n(\omega) \equiv \frac{\mathrm{d}N_{\mathrm{st}}}{\mathrm{d}\omega}$$

so that the total number of states with wavevector less than k_{max} is

$$N_{\mathrm{st}} = \int_0^{k_{\mathrm{max}}} n(\omega)\mathrm{d}\omega$$

Using our expression for N_{st}, we get

$$\boxed{\begin{aligned} n(\omega) &= \frac{\mathrm{d}N_{\mathrm{st}}}{\mathrm{d}\omega} = \frac{\mathrm{d}N_{\mathrm{st}}}{\mathrm{d}k} \times \frac{\mathrm{d}k}{\mathrm{d}\omega} \\ &= \frac{1}{2}\frac{1}{\pi^2} V k^2 \frac{1}{\mathrm{d}\omega/\mathrm{d}k} \end{aligned}} \tag{1.3.2}$$

and $\mathrm{d}\omega/\mathrm{d}k$ is the group velocity.

This definition is appropriate as long as the dispersion relation $\omega(\mathbf{k})$ is the same for all directions of $\mathbf{k}$, so that

$$\omega(\mathbf{k}) = \omega(k)$$

Generally the $\mathbf{k}$-vectors corresponding to a given frequency form a non spherical surface in $\mathbf{k}$-space. If we denote this constant energy surface by S_ω, then the general expression for the density of states is

$$n(\omega) = \frac{V}{(2\pi)^3} \int \frac{\mathrm{d}S_\omega}{v_g}$$

where the group velocity is

$$\mathbf{v}_g = \boldsymbol{\nabla}_{\mathbf{k}}\, \omega(\mathbf{k})$$

You will find a simple derivation in Kittel, chapter 5; or Elliott, page 231. See also Ibach and Lüth, section 5.1. We will only need eq (1.3.2).

1.4 A useful identity

The use of periodic boundary conditions and Brillouin zones allows us to prove a useful identity. If we consider two lattice waves or electron wavevectors, $\mathbf{k}$ and $\mathbf{k}'$ and make a sum over all unit cells we find

$$\begin{aligned}\sum_{j=1}^{N_\mathrm{c}} \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_j} &= \begin{cases} N_\mathrm{c} & \text{if} \quad \mathbf{k}=\mathbf{k}' \\ 0 & \text{if} \quad \mathbf{k}\neq\mathbf{k}' \end{cases} \\ &= N_\mathrm{c}\delta_{\mathbf{k}\mathbf{k}'}\end{aligned}$$

The proof is fairly easy in one dimension. If $k = k'$

$$\sum_{j=1}^{N_\mathrm{c}} \mathrm{e}^{\mathrm{i}(k-k')\cdot r_j} = \sum_{j=1}^{N_\mathrm{c}} 1 = N_\mathrm{c}$$

If $k \neq k'$ we write $r_j = ja$ (a is the lattice spacing)

$$k = \frac{2\pi n}{L}\,; \qquad k' = \frac{2\pi n'}{L}\,; \qquad k - k' = \frac{2\pi(n-n')}{L} = \frac{2\pi m}{N_\mathrm{c} a}$$

Then

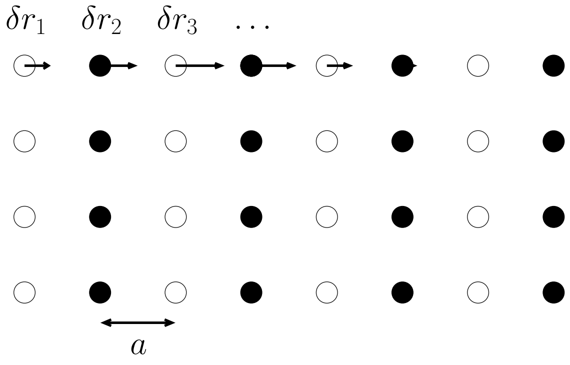
$$\begin{aligned}\sum_{j=1}^{N_\mathrm{c}} \mathrm{e}^{\mathrm{i}(k-k')\cdot r_j} &= \sum_{j=1}^{N_\mathrm{c}} \mathrm{e}^{\mathrm{i}2\pi m j/N_\mathrm{c}} \\ &= \sum_{j=1}^{N_\mathrm{c}} \left(\mathrm{e}^{\mathrm{i}2\pi m/N_\mathrm{c}}\right)^j \\ &= \frac{1-\mathrm{e}^{\mathrm{i}2\pi m}}{1-\mathrm{e}^{\mathrm{i}2\pi m/N_\mathrm{c}}} \times \mathrm{e}^{\mathrm{i}2\pi m/N_\mathrm{c}} \\ &= 0\end{aligned}$$

because $\mathrm{e}^{\mathrm{i}2\pi m} = 1$. We have used the identity

$$\sum_{j=1}^{N} x^j = \frac{1-x^N}{1-x}\,x$$

1.5 Vibrations in the diatomic lattice

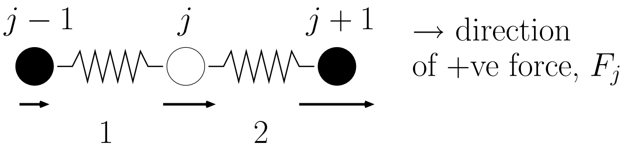
We get some new physics if our crystal has two types of atom. Let us extend our simple model to two atoms per primitive unit cell with masses M_1 and M_2



The mass of an atom in plane j is M_1 if j is odd, and M_2 if j is even. Longitudinal mode displacements are

$$\delta r_j = \frac{1}{\sqrt{N_\mathrm{c}}} \begin{cases} \tilde{q}_1 \, \mathrm{e}^{\mathrm{i}(kja-\omega t)} & j \text{ odd} \\ \tilde{q}_2 \, \mathrm{e}^{\mathrm{i}(kja-\omega t)} & j \text{ even} \end{cases}$$

Here, $\tilde{q}_1$ and $\tilde{q}_2$ are parallel vectors, so we write them as scalars and drop the subscript k, but the amplitudes will be different at the different types of atomic planes. We will do the problem this time in terms of forces (rather than potential energy) just to be different. The force F_j on the plane of atoms j is proportional to the stretch of the bond. We have to be careful about the sign:



$$\begin{aligned} &\text{Bond 2: if } \delta r_{j+1} - \delta r_j > 0 \text{ bond is stretched,} \quad F_j > 0 \\ &\qquad\qquad\qquad\qquad\quad < 0 \text{ bond is compressed,} \; F_j < 0 \\ &\text{Bond 1: if } \delta r_{j-1} - \delta r_j < 0 \text{ bond is stretched,} \quad F_j < 0 \\ &\qquad\qquad\qquad\qquad\quad > 0 \text{ bond is compressed,} \; F_j > 0 \end{aligned}$$

Using these rules, after some algebra, we see that

$$\begin{aligned} F_j &= C\left[(\delta r_{j+1} - \delta r_j) + (\delta r_{j-1} - \delta r_j)\right] \\ &= \frac{1}{\sqrt{N_\mathrm{c}}} \begin{cases} C\left(\tilde{q}_2 \, \mathrm{e}^{\mathrm{i}ka} + \tilde{q}_2 \, \mathrm{e}^{-\mathrm{i}ka} - 2\tilde{q}_1\right) \mathrm{e}^{\mathrm{i}kja} \, \mathrm{e}^{\mathrm{i}\omega t} & j \text{ odd;} \\ C\left(\tilde{q}_1 \, \mathrm{e}^{\mathrm{i}ka} + \tilde{q}_1 \, \mathrm{e}^{-\mathrm{i}ka} - 2\tilde{q}_2\right) \mathrm{e}^{\mathrm{i}kja} \, \mathrm{e}^{\mathrm{i}\omega t} & j \text{ even.} \end{cases} \end{aligned}$$

The equation of motion is

$$F_j = M_j \frac{\mathrm{d}^2}{\mathrm{d}t^2} \, \delta r_j$$

and using $\mathrm{e}^{\mathrm{i}ka} + \mathrm{e}^{-\mathrm{i}ka} = 2\cos ka$ we get two simultaneous equations,

$$\left. \begin{aligned} 2C\left(\tilde{q}_2 \cos ka - \tilde{q}_1\right) &= -M_1\omega^2 \tilde{q}_1 \\ 2C\left(\tilde{q}_1 \cos ka - \tilde{q}_2\right) &= -M_2\omega^2 \tilde{q}_2 \end{aligned} \right\}$$

The solution to these is

$$\omega^2 = \frac{2C}{M_1 M_2}\left(\frac{1}{2}\left(M_1 + M_2\right) \pm \sqrt{\frac{1}{4}\left(M_1 - M_2\right)^2 + M_1 M_2 \cos^2 ka}\right) \tag{1.5.1}$$

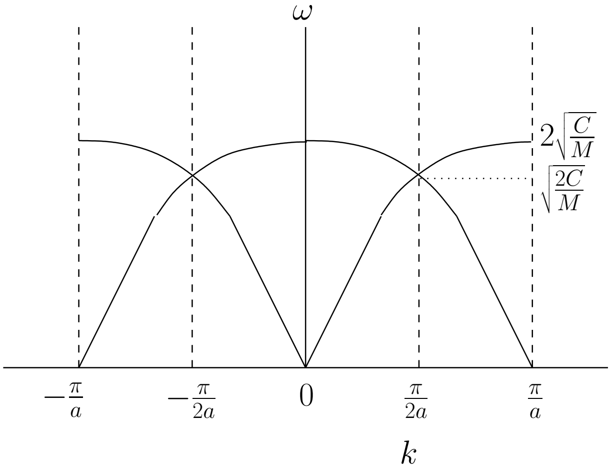
Of course, if $M_2 = M_1 = M$ the problem reverts to our original one having just one type of atom, and equation (1.5.1) becomes

$$\omega^2 = \frac{2C}{M}\left(1 \pm \cos ka\right)$$

There are two solutions corresponding to the plus and minus signs:

$$\omega = \sqrt{\frac{C}{M}}\begin{cases} 2\sin\frac{1}{2}ka & \text{(plus sign)} \\ 2\cos\frac{1}{2}ka & \text{(plus sign)} \end{cases}$$

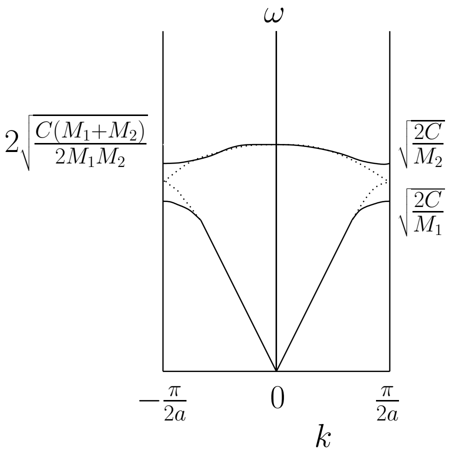
Why are there now twice as many solutions? Let us sketch them as before.



We can do one of two things:

1. Discard the cosine solutions. In fact, they merely duplicate the sine solutions. (It's like saying you can take either the real or the imaginary part of equation (1.1.1) for the displacements.) This is called the extended zone scheme.
2. Recognise that in the two atom unit cell our lattice constant is really $2a$, so the Brillouin zone is *half as big*. Then we keep both curves, but only within the first Brillouin zone; *i.e.*, k is restricted to the range $-\pi/2a$ to $\pi/2a$. This is called the reduced zone scheme.

When the masses are different, then we get the following dispersion curves in the reduced zone, by sketching equations (1.5.1) with $M_1 > M_2$.

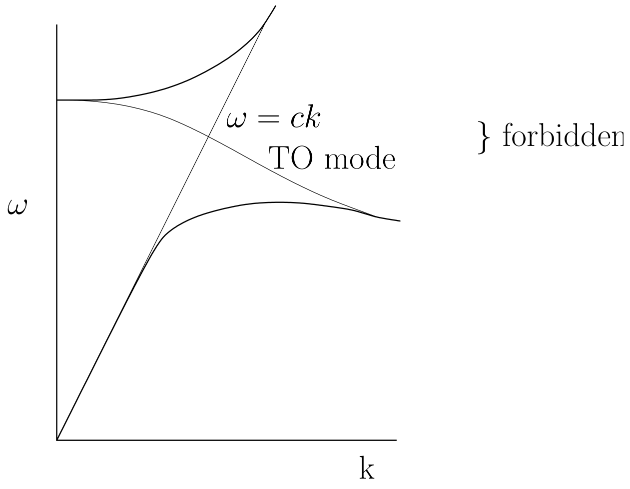


The dotted lines show the case $M_1 = M_2$. Notice how having different masses *opens up a gap* in the allowed values of ω. This is entirely analogous to the band gap in the electronic states in a semiconductor; the origin, though, is quite different—obviously all electrons have the same mass.

Each line in the $\omega(k)$ dispersion relation is called a band. A three dimensional crystal has three bands having $\omega \to 0$ as $k \to 0$, one longitudinal and two transverse modes, and these are called acoustic mode bands. This is because as $k \to 0$, *i.e.*, in the long wavelength limit, they have a constant velocity and become the longitudinal and transverse sound waves.

In general, there are $3n_\mathrm{c}$ bands, where n_c is the number of atoms in the primitive unit cell. The remaining $3n_\mathrm{c} - 3$ bands have $\omega \neq 0$ at $k = 0$ similar to the upper band in the diagram. These are called optical mode bands because at $k = 0$ they couple to electromagnetic radiation.

The interaction of light with transverse optical modes can be illustrated in a dispersion diagram (or band structure) as follows.



In this figure the k axis is magnified: the speed of light c being what it is, the slope of the dispersion of light is much much steeper than the slope of the acoustic lattice waves. At this scale the acoustic branch would appear horizontal, just as in the previous diagram, the light branch would appear vertical.

The dispersion of light is linear: $\omega = ck$. In the absence of interaction we get the thin lines. But the lattice waves and light waves behave like two coupled harmonic oscillators (the coupling arises from the interaction between the electromagnetic field and the dipole moment of the oscillating ions) so we get solutions (the thick lines) as if they were two coupled pendulums.

The elementary excitation produced is called a polariton. See how the lattice waves strongly modify the dispersion of light leading to a forbidden gap where the light is strongly attenuated when passing through the solid. This effect can be observed in what's called 'Raman spectroscopy.'

1.6 Phonons

Let us remark that features common to the elementary excitations in solids such as forbidden gaps, a density of states, spottiness of **k**-space have emerged as a consequence of a classical, Newtonian treatment of lattice waves. You may have thought these only arose in a quantum theory.

We do need to use quantum mechanics if we want to understand how lattice waves interact with, in particular, electrons. For example, the so-called electron–phonon coupling is responsible for superconductivity in metals, and "polarons" in insulators (these are distortions of the ionic lattice around electrons in the conduction band). It also acts to limit the electrical conductivity of metals.

So *what are phonons?* Quantum mechanics can turn waves into particles and vice versa (the wave–particle duality, if you like). When we quantise lattice waves, we get *particles* called phonons. Here are the steps we need to take.

1. Transform the displacements $\delta\mathbf{r}_j$ into normal modes.
2. Find the hamiltonian for these modes.
3. See at once that it is the hamiltonian for a set of independent (*i.e.*, uncoupled) simple harmonic oscillators.
4. All we need then is the solution of one of the standard problems in quantum mechanics: the simple harmonic oscillator.

In view of step 4, since we will not wish to break step at that point, we will now digress to solve the quantum mechanical simple harmonic oscillator.

1.6.1 The simple harmonic oscillator

The simple harmonic oscillator is one of the standard problems in quantum mechanics. The classical hamiltonian is

$$H = \frac{1}{2M}p^2 + \frac{1}{2}M\omega^2 q^2$$

To "quantise" we replace p with $-\mathrm{i}\hbar\,\mathrm{d}/\mathrm{d}q$ which results in the "commutation relation" $[q,p] = pq - qp = \mathrm{i}\hbar$

The time independent Schrödinger equation is

$$H\psi_n(q) = E_n\psi_n(q)$$

with solutions for $n = 0, 1, 2\ldots$ and $n = 0$ being the ground state. We have to solve

$$\left(-\frac{\hbar^2}{2M}\frac{\mathrm{d}^2}{\mathrm{d}q^2} + \frac{1}{2}M\omega^2 q^2\right)\psi_n(q) = E_n\psi_n(q)$$

Let us define a dimensionless coordinate to simplify the formulas,

$$x = \sqrt{\frac{M\omega}{\hbar}}\,q$$

Then the Schrödinger equation is

$$\frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)\psi_n(x) = E_n\psi_n(x)$$

How do we solve this? Well, note that

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)$$

is a difference of squares like

$$(-A^2 + B^2) = (-A + B)(A + B)$$

so let's try writing

$$\begin{aligned}\left(-\frac{\mathrm{d}}{\mathrm{d}x} + x\right)\left(\frac{\mathrm{d}}{\mathrm{d}x} + x\right) &= \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right) - \left(\frac{\mathrm{d}}{\mathrm{d}x}x - x\frac{\mathrm{d}}{\mathrm{d}x}\right)\\ &= \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right) - 1\end{aligned}$$

An extra "minus one" appears as a consequence of the commutation relation, and hence

$$\begin{aligned}\frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)\psi_n(x) &= \frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}}{\mathrm{d}x} + x\right)\left(\frac{\mathrm{d}}{\mathrm{d}x} + x\right)\psi_n(x) + \frac{1}{2}\hbar\omega\psi_n(x)\\ &= \hbar\omega\left(a^+a + \frac{1}{2}\right)\psi_n(x)\end{aligned}$$

having defined

$$\begin{aligned} a^+ &= \frac{1}{\sqrt{2}}\left(-\frac{\mathrm{d}}{\mathrm{d}x}+x\right) \\ a &= \frac{1}{\sqrt{2}}\left(\frac{\mathrm{d}}{\mathrm{d}x}+x\right) \end{aligned}$$

We easily see that

$$aa^+ - a^+a = 1$$

or

$$\left[a, a^+\right] = 1$$

Let us examine the properties of the operators a and a^+. (Note that a^+ is pronounced "a dagger" and in some texts is denoted $a^\dagger$or a^*). We have three ways to write the hamiltonian now:

$$\begin{aligned} H &= -\frac{\hbar^2}{2M}\frac{\mathrm{d}^2}{\mathrm{d}q^2} + \frac{1}{2}M\omega^2 q^2 \\ &= \frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}+x^2\right) \\ &= \left(a^+a+\frac{1}{2}\right)\hbar\omega \end{aligned}$$

So

$$\begin{aligned} Ha &= \left(a^+aa+\frac{1}{2}a\right)\hbar\omega \\ &= \left(\left(aa^+-1\right)a+\frac{1}{2}a\right)\hbar\omega \\ &= \left(aa^+a-\frac{1}{2}a\right)\hbar\omega \\ &= a\left(a^+a-\frac{1}{2}\right)\hbar\omega \\ &= a\left(H-\hbar\omega\right) \end{aligned}$$

In the second line, we have used the commutation relation. Operating on an eigenfunction ψ_n we get

$$\begin{aligned} H\left(a\psi_n\right) &= a\left(H-\hbar\omega\right)\psi_n \\ &= \left(E_n-\hbar\omega\right)\left(a\psi_n\right) \end{aligned}$$

since $H\psi_n = E_n\psi_n$. Hence $a\psi_n$ is an eigenfunction with the energy lowered by $\hbar\omega$. This amount of energy is called a quantum, and a is called the *quantum destruction operator.* Similarly you can see that

$$H\left(a^+\psi_n\right) = \left(E_n+\hbar\omega\right)\psi_n$$

and a^+ is an operator that *creates* a quantum of energy.

There cannot be negative eigenvalues (since the hamiltonian is a sum of squares of hermitian operators). There must exist a ground state ψ_0 with energy $E_0 > 0$. Applying the destruction operator to this state we get

$$H\left(a\psi_0\right) = \left(E_0 - \hbar\omega\right)\left(a\psi_0\right)$$

which imples that $a\psi_0$ is an eigenfunction with energy *lower* than E_0. Since we have defined E_0 as the lowest energy, to avoid this paradox it must be true that

$$a\psi_0 = 0$$

Therefore

$$\begin{aligned} H\psi_0 &= \left(a^+ a + \frac{1}{2}\right)\hbar\omega\psi_0 \\ &= \frac{1}{2}\hbar\omega\psi_0 \end{aligned}$$

So the ground state has energy $\frac{1}{2}\hbar\omega$. This is called the zero point energy.

Note, in passing, that since

$$H = \frac{1}{2M}p^2 + \frac{1}{2}M\omega^2 q^2$$

then the expectation value of the energy in the ground state is

$$\frac{1}{2}\hbar\omega = \frac{1}{2M}\left\langle p^2\right\rangle + \frac{1}{2}M\omega^2\left\langle q^2\right\rangle$$

Now if three real numbers are $a + b = c$ then $ab \le \frac{1}{4}c^2$. Cancelling the $\frac{1}{2}$'s we see that

$$M\omega^2\left\langle q^2\right\rangle \frac{1}{M}\left\langle p^2\right\rangle \le \frac{1}{4}\left(\hbar\omega\right)^2$$

that is,

$$\sqrt{\left\langle q^2\right\rangle}\sqrt{\left\langle p^2\right\rangle} \le \frac{1}{2}\hbar$$

The left hand side is the root mean square position times the r.m.s. momentum and is smaller than or equal to $\frac{1}{2}\hbar$. But by the uncertainty principle

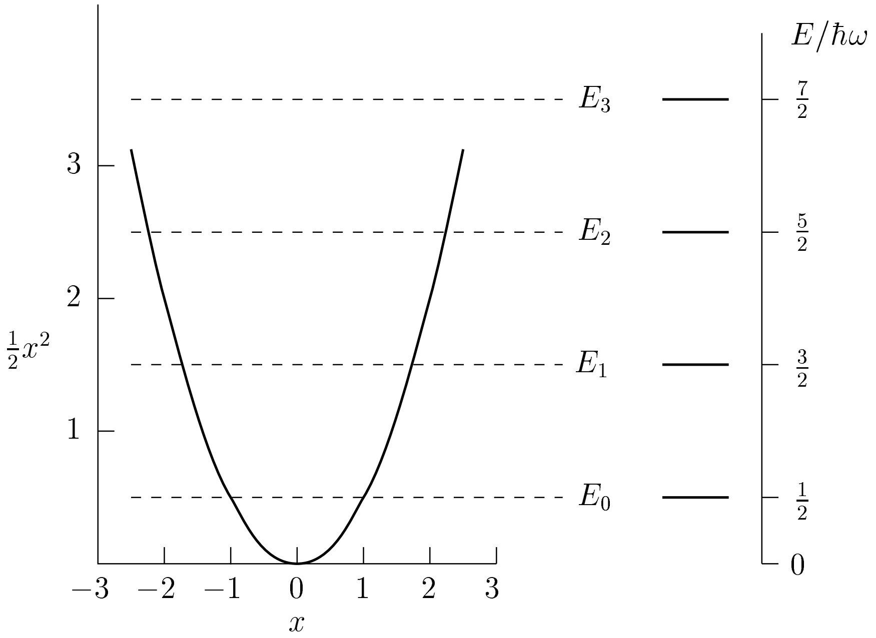
$$\Delta q \Delta p \ge \frac{1}{2}\hbar$$

so $\frac{1}{2}\hbar\omega$ is the *lowest energy the ground state can have without violating the uncertainty principle.*

What are the energy levels of the harmonic oscillator? We know two things:

1. The lowest energy level has energy $E_0 = \frac{1}{2}\hbar\omega$.
2. The operator a^+ operating on any eigenfunction ψ_n with energy E_n results in the eigenfunction ψ_{n+1} having the next highest energy E_{n+1}. Successive energy levels are separated by the quantum of energy $\hbar\omega$.

These levels are illustrated below. The diagram shows plotted the classical potential energy $\frac{1}{2}\hbar\omega x^2$ in which the oscillator moves (in units of $\hbar\omega$). On the right are the energy levels, also in units of $\hbar\omega$.



Clearly

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega$$

and the solution to the Schrödinger equation is

$$\begin{aligned} H\psi_n &= \left(a^+a + \frac{1}{2}\right)\hbar\omega\psi_n \\ &= \left(n + \frac{1}{2}\right)\hbar\omega\psi_n \end{aligned}$$

You can see that ψ_n is an eigenfunction of the operator a^+a having eigenvalue n,

$$a^+a\psi_n = n\psi_n$$

a^+a is called the number operator because it counts the number of quanta of energy possessed by the oscillator when it's in the n^{th} excited state.

1.6.2 Quantisation of the lattice waves

Now we come back to the solid state problem of the quantisation of the lattice waves. The mathematics is quite complicated. We will confine ourselves the one dimensional situation with one atom per unit cell. First we define normal modes. To begin with, I write

$$\delta r_j = \frac{1}{\sqrt{N_\mathrm{c}}}\sum_k q_k\,\mathrm{e}^{\mathrm{i}kr_j} \qquad (1.6.1)$$

Note that

$$q_k = \tilde{q}_k \, \mathrm{e}^{\mathrm{i}\omega t}$$

from eq (1.1.1), that is, the amplitude times the phase factor $\mathrm{e}^{\mathrm{i}\omega t}$. What are these quantities, q_k? Multiply (1.6.1) on both sides by $\mathrm{e}^{-\mathrm{i}k'r_j}/\sqrt{N_\mathrm{c}}$ and sum each side over all atoms,

$$\begin{aligned}\frac{1}{\sqrt{N_\mathrm{c}}}\sum_j \delta r_j \, \mathrm{e}^{-\mathrm{i}k'r_j} &= \frac{1}{N_\mathrm{c}}\sum_k q_k \sum_j \mathrm{e}^{\mathrm{i}(k-k')r_j} \\ &= \sum_k q_k \delta_{kk'} \\ &= q_{k'}\end{aligned}$$

because in the second line we can use our nifty identity from section 1.4, *viz.*,

$$\sum_j \mathrm{e}^{\mathrm{i}(k-k')r_j} = N_\mathrm{c}\delta_{kk'}$$

So, we have

$$q_k = \frac{1}{\sqrt{N_\mathrm{c}}}\sum_j \delta r_j \, \mathrm{e}^{-\mathrm{i}kr_j} \tag{1.6.2}$$

Note the symmetry between (1.6.1) and (1.6.2). q_k is the discrete Fourier transform of the displacements, δr_j; it is called a normal mode of wavevector k. You can easily see that

$$q_{-k} = \bar{q}_k$$

Now I want to write the hamiltonian in terms of normal modes rather than displacements. In terms of displacements,

$$H = \sum_j \frac{1}{2}M\left(\frac{\mathrm{d}}{\mathrm{d}t}\delta r_j\right)^2 + \frac{1}{2}\sum_{ij}\delta r_i A_{ij}\delta r_j \tag{1.6.3}$$

and A_{ij} is the force constant matrix. We just need to insert (1.6.1) into this to get it in terms of the q_k. We'll do the potential energy (second term in (1.6.3)) first.

$$\begin{aligned}\frac{1}{2}\sum_{ij}\delta r_i A_{ij}\delta r_j &= \frac{1}{2}\sum_{ij}\left(\frac{1}{\sqrt{N_\mathrm{c}}}\sum_k q_k \, \mathrm{e}^{\mathrm{i}kr_i}\right) A_{ij}\left(\frac{1}{\sqrt{N_\mathrm{c}}}\sum_{k'} q_k \, \mathrm{e}^{\mathrm{i}k'r_j}\right) \\ &= \frac{1}{2N_\mathrm{c}}\sum_{kk'} q_k \, q_{k'} \sum_{ij} \mathrm{e}^{\mathrm{i}kr_i} A_{ij} \, \mathrm{e}^{\mathrm{i}k'r_j} \\ &= \frac{1}{2N_\mathrm{c}}\sum_{kk'} q_k \, q_{k'} \sum_{ij}\left(\mathrm{e}^{\mathrm{i}k(r_i-r_j)}A_{ij}\right)\mathrm{e}^{\mathrm{i}(k+k')r_j}\end{aligned}$$

This will simplify nicely if we are careful. Take the sum over i and j and put the sum over j first:

$$\sum_j \mathrm{e}^{\mathrm{i}(k+k')r_j}\left\{\sum_i \mathrm{e}^{\mathrm{i}k(r_i-r_j)} \, A_{ij}\right\}$$

The term in braces is the same for any atom labelled j because by translational symmetry all the atoms are identical (there's only one atom per unit cell in our simple case here). So we can define a quantity

$$\lambda_k = \sum_i \mathrm{e}^{\mathrm{i}k(r_i - r_j)}\, A_{ij} = M\omega_k^2$$

which is the equation of motion for any atom j; compare with eq (1.1.4) in section 1.1. We call λ_k the dynamical matrix; it's actually a diagonal matrix: one entry for each k. By transforming to normal modes, *we have diagonalised the force constant matrix.*

Then, using our identity from section 1.4,

$$\sum_j \mathrm{e}^{\mathrm{i}(k+k')r_j}\, \lambda_k = N_\mathrm{c}\lambda_k\delta_{-kk'}$$

so the potential energy in terms of normal mode coordinates is simply

$$\begin{aligned}\frac{1}{2}\sum_{ij} A_{ij}\delta r_i \delta r_j &= \frac{1}{2N}\sum_{kk'} q_k\, q_{k'}\; N_\mathrm{c}\lambda_k\delta_{k',-k}\\ &= \frac{1}{2}\sum_k \lambda_k\, q_k\, q_{-k}\end{aligned}$$

See how the Kronecker delta $\delta_{k',-k}$ *picks out* from the double sum of k and k' only those terms for which $k = -k'$. We have indeed diagonalised A_{ij}.

The kinetic energy is easier. Using again our identity from section 1.4,

$$\begin{aligned}\sum_j \delta\dot{r}_j\, \delta\dot{r}_j &= \frac{1}{N_\mathrm{c}}\sum_j\sum_{kk'} \dot{q}_k\, \dot{q}_{k'}\; \mathrm{e}^{\mathrm{i}(k+k')r_j}\\ &= \sum_k \dot{q}_k\, \dot{q}_{-k}\end{aligned}$$

Therefore in terms of normal coordinates the hamiltonian is

$$H = \sum_k \left(\frac{1}{2}M\,\dot{q}_k\,\dot{q}_{-k} + \frac{1}{2}\lambda_k\, q_k\, q_{-k}\right)$$

To find the momentum conjugate to q, I write the Lagrangian as, equation (0.1),

$$\begin{aligned}L &= K - W\\ &= \sum_k \left(\frac{1}{2}M\,\dot{q}_k\,\dot{q}_{-k} - \frac{1}{2}\lambda_k\, q_k\, q_{-k}\right)\end{aligned}$$

and define

$$p_k = \frac{\partial L}{\partial \dot{q}_k} = M\dot{q}_{-k}$$

Finally I have

$$H = \frac{1}{2}\sum_k \left(\frac{1}{M} p_k^2 + \lambda_k \, q_k^2 \right)$$

Doesn't that look a lot neater than eq (1.6.3)?

When you compare this with the hamiltonian for a single simple harmonic oscillator you will see that we have recast the classical Newtonian hamiltonian of the complex motions of the atoms making up lattice waves into a sum of hamiltonians of simple harmonic oscillators. Here are two very important points.

1. The separation of the vibration into *independent* normal modes is permitted only because we are working in the harmonic approximation. But no other approximation is invoked, for example there is no restriction on the nature of the interatomic forces, such as the ball and spring models we used earlier.
2. Because the hamiltonian is a *sum* over independent simple harmonic oscillators, we can immediately quantise the lattice waves.

So, we can write down at once the hamiltonian in second quantisation describing N_c simple harmonic oscillators in each band,

$$H = \sum_k \hbar\omega_k \left(a_k^+ a_k + \frac{1}{2} \right)$$

This is exactly the same as the hamiltonian we had for the simple harmonic oscillator previously. The only difference is that instead of just one oscillator having a (single mode of) natural frequency ω, there are N_c oscillators each having a natural frequency ω_k so we have to label it, and its creation and destruction operators also, with a label k. Remember that for the simple harmonic oscillator we said that a^+a was the "number operator" whose eigenvalue is the *number of quanta* of energy possessed by the oscillator when it is the n^{th} excited state. Similarly here we say that for each normal mode of vibration, labelled k,

$$n_k = \left\langle a_k^+ a_k \right\rangle$$

is the average, or *expectation* value of the number operator and is the number of quanta, or phonons that are excited in the k^{th} mode. This is the proper way to understand what a phonon is. It is an elementary excitation in the vibrations of the atoms in the solid. The more a particular mode is excited, say by heating the crystal, the more quanta, or phonons, there are possessed by each of the modes, labelled k.

So the expectation value of the total energy is

$$\langle H \rangle = \sum_k \hbar\omega_k \left(n_k + \frac{1}{2} \right)$$

and we say that there are n_k phonons present in the crystal having wavenumber k and frequency ω_k. These are of course related through the dispersion relation, $\omega = \omega(k)$.

Now,

$$\sum_k \frac{1}{2}\hbar\omega_k$$

is the zero point energy of the crystal. The zero point energy cannot do work. However it does contribute to the cohesive energy of the solid, and can be observed by X-ray scattering.

As the temperature of a crystal is raised, more phonons are created, and phonons of higher frequency are excited. Remember there is a maximum k and hence a maximum ω_k and hence a maximum energy $\hbar\omega_k$ that a phonon can have. At a certain temperature, called the Debye temperature, all allowed modes contain lots of phonons.

We need to know how many phonons there are of energy $\hbar\omega_k$ as a function of temperature. Phonons are distributed under Bose–Einstein statistics because they are *bosons*—firstly since their creation and destruction operators *commute* rather than *anti*commute as do, say electron operators, secondly because they possess integral spin, *viz.,* zero. Therefore we know at once that the number of phonons in state k is

$$n_k = \frac{1}{\mathrm{e}^{\hbar\omega_k/k_\mathrm{B}T} - 1}$$

where k_B is the Boltzmann constant (1.38×10^{-23} J K^{-1} or 8.62×10^{-5} eV K^{-1}) and T is the absolute temperature.

So as we heat the crystal, we put energy into the lattice vibrations so that the internal energy as a function of temperature is

$$\begin{aligned} U &= \sum_k \hbar\omega_k \left(\frac{1}{2} + n_k\right) \\ &= \sum_k \hbar\omega_k \left(\frac{1}{2} + \frac{1}{\mathrm{e}^{\hbar\omega_k/k_\mathrm{B}T} - 1}\right) \end{aligned} \qquad (1.6.4)$$

This is a useful formula. We can use it to work out the heat capacity and thermal conductivity due to phonons among other things.

But it's not easy to do the summation even if we know the dispersion relation. Let's take the high temperature limit. If $\hbar\omega_k/k_\mathrm{B}T \ll 1$ then, because $\mathrm{e}^x = 1 + x$ for small x we have

$$n_k = \frac{k_\mathrm{B}T}{\hbar\omega_k} \gg 1 \qquad \text{lots of phonons in each mode} \qquad (1.6.5)$$

and so

$$U = \sum_k \hbar\omega_k \left(\frac{1}{2} + \frac{k_\mathrm{B}T}{\hbar\omega_k}\right) \qquad \text{at high temperature}$$

and leaving out the zero point energy we have

$$\begin{aligned} U &= \sum_k k_\mathrm{B} T \\ &= k_\mathrm{B} T \sum_k 1 \end{aligned}$$

so we need to know what is the number of terms in the sum over k so we can add all the "one's" and multiply that by $k_\mathrm{B}T$. In other words how many allowed values of k are there? You see now why we had to study this carefully in sections 1.2 and 1.3.

In a crystal containing N_at atoms and one atom per unit cell there are three branches or *bands*—one longitudinal and two transverse—and we therefore get

$$U = 3N_\mathrm{at} k_\mathrm{B} T$$

which is the energy you would get from the *equipartition theorem*: each degree of freedom contributes $\frac{1}{2}k_\mathrm{B}T$ of potential and $\frac{1}{2}k_\mathrm{B}T$ of kinetic energy.

The heat capacity (due to phonons) at constant volume is hence

$$\begin{aligned} C_\mathrm{ph} = \frac{\mathrm{d}U}{\mathrm{d}T} &= 3N_\mathrm{at} k_\mathrm{B} \qquad && \mathrm{J\ K^{-1}} \\ &= 3R && \mathrm{J\ mol^{-1}\ K^{-1}} \end{aligned}$$

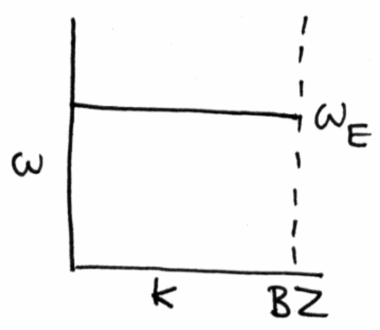
in which R is the gas constant, 8.314 J mol^{-1} K^{-1}. In fact the formula

$$\boxed{C = 3R}$$

is the law of Dulong and Petit, which you have just derived from first principles and which describes the heat capacity per mole of a large number of solids at high temperature.

Unfortunately this does not hold at low temperature. Below the Debye temperature quantum effects appear because not all the modes are excited. At low temperature we are stuck; we cannot do the summation (1.6.4) for a general dispersion relation, $\omega(\mathbf{k})$. As is usual in solid state physics, we proceed by making a model; in fact two models which are now described.

1.6.3 The Einstein model



Here we suppose that there is no dispersion: all atoms simply vibrate with a single frequency, $\omega = \omega_E$. Then the total energy due to phonons is

$$U = 3N_{\rm at}\hbar\omega\left(\frac{1}{2} + \frac{1}{{\rm e}^{\hbar\omega/k_{\rm B}T} - 1}\right)$$

there's no summation of $\mathbf{k}$ anymore, instead the energy at $\omega = \omega_{\rm E}$ multiplied by the number of allowed modes, $3N_{\rm at}$. Again we omit zero point energy and

$$U = 3N_{\rm at}\frac{\hbar\omega}{{\rm e}^{\hbar\omega/k_{\rm B}T} - 1}$$

so the heat capacity is

$$C_{\rm ph} = \frac{{\rm d}U}{{\rm d}T} = 3N_{\rm at}k_{\rm B}\left(\frac{\hbar\omega}{k_{\rm B}T}\right)^2 \frac{{\rm e}^{\hbar\omega/k_{\rm B}T}}{\left({\rm e}^{\hbar\omega/k_{\rm B}T} - 1\right)^2}$$

Let us examine the high and low temperature limits. First we define

$$x = \frac{\hbar\omega}{k_{\rm B}T}$$

so that

$$C_{\rm ph} = 3N_{\rm at}k_{\rm B}\frac{x^2{\rm e}^x}{({\rm e}^x - 1)^2}$$

$$\text{as} \quad T \to \infty \ , \qquad x \to 0 \ , \qquad C_{\rm ph} \to \frac{0}{0}$$

$$\text{as} \quad T \to 0 \ , \qquad x \to \infty \ , \qquad C_{\rm ph} \to \frac{\infty}{\infty}$$

To work out "indeterminate forms" we use L'Hôpital's rule; that is differentiate numerator and denominator with respect to x until the ratio becomes determinate:

$$\frac{x^2{\rm e}^x}{({\rm e}^x - 1)^2} \longrightarrow \frac{2x{\rm e}^x + x^2{\rm e}^x}{2\,(e^x - 1)\,{\rm e}^x} \longrightarrow \frac{2 + 2x}{2{\rm e}^x} \longrightarrow \frac{1}{e^x}$$

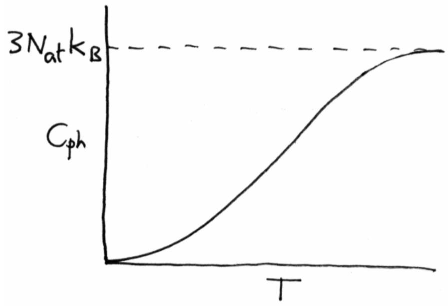
so

$$\lim_{T \to \infty} C_{\rm ph} = 3N_{\rm at}k_{\rm B}$$

which is the law of Dulong and Petit. Also

$$\begin{aligned}\lim_{T \to 0} C_{\mathrm{ph}} &= 3N_{\mathrm{at}}k_{\mathrm{B}} \lim_{x \to \infty} C_{\mathrm{ph}} = 3N_{\mathrm{at}}k_{\mathrm{B}}\frac{1}{x} \\ &= 3N_{\mathrm{at}}k_{\mathrm{B}}\mathrm{e}^{-\hbar\omega/k_{\mathrm{B}}T} \\ &\longrightarrow 0\end{aligned}$$

so, for the heat capacity as a function of temperature we have this.



and C_{ph} goes to zero exponentially with the temperature. The Eintein model resolves a problem that *classically* C_{ph} should be $3N_{\mathrm{at}}k_{\mathrm{B}}$ at all temperatures, but is observed to go to zero as the temperature is reduced. This was one of the triumphs of the "old quantum theory."

However, experimentally C_{ph} does not go to zero exponentially but as T^{-3}. See figure 9, chapter 5 in Kittel.

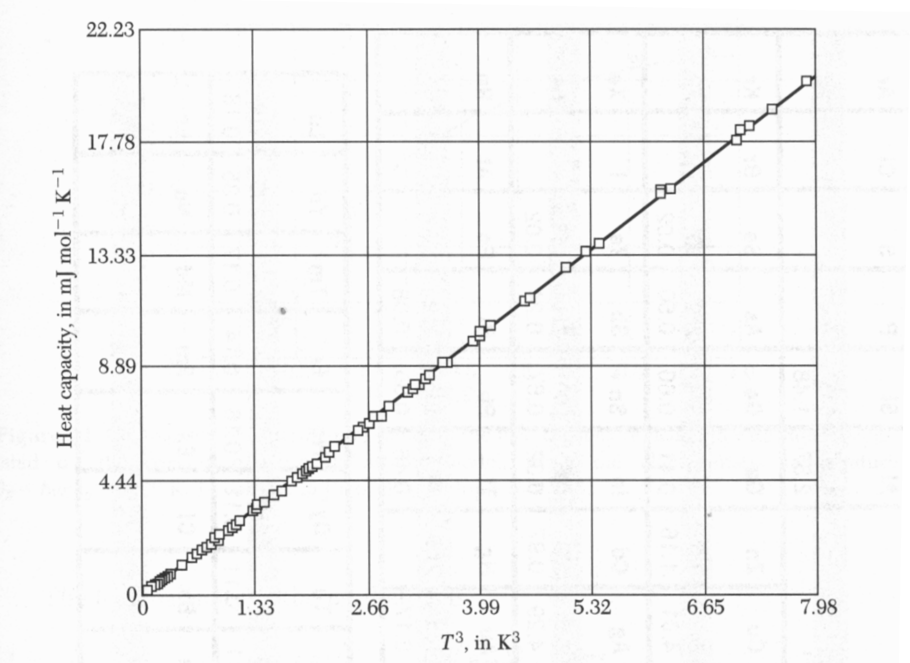


Figure 9 Low temperature heat capacity of solid argon, plotted against T^3. In this temperature region the experimental results are in excellent agreement with the Debye T^3 law with $\theta = 92.0$ K. (Courtesy of L. Finegold and N. E. Phillips.)

1.6.4 The Debye model

The Debye model fixes up the low temperature behaviour. Remember that in the Einstein model we use the approximation that the crystal has just one frequency of vibration, ω_{E}, so the dispersion curve is just a horizontal line, $\omega(k) = \omega_{\mathrm{E}} = \text{constant}$. In the Debye model we take the next level of approximation and use a linear dispersion (like light in a vacuum—usually this behaviour is called *non dispersive*), namely $\omega(k) = ck$ which is the dispersion for low frequency acoustic waves, and c is the velocity of sound.

We now introduce a very important concept. The internal energy, U, is

$$U = \int \hbar\omega \, \frac{1}{\mathrm{e}^{\hbar\omega/k_{\mathrm{B}}T} - 1} \, n(\omega) \, \mathrm{d}\omega$$

that is, the energy of a phonon mode with frequency ω times the number of phonons of that frequency at temperature T times the number of states in the interval $\mathrm{d}\omega$ about ω, integrated over all values of ω. *Be sure you understand this.*

Now, there is a cut-off in the values of k and ω. In the isotropic case we are considering there is a maximum value of k, k_{max}, and the number of modes with wavevector less than k_{max} is, equation (1.3.1),

$$\frac{1}{6}\,\frac{1}{\pi^2}\,V\,k_{\mathrm{max}}^3 = N_{\mathrm{at}}$$

(remember N_{at} is the number of atoms in the crystal of volume V; we are allowing just one atom per unit cell for simplicity). So

$$k_{\mathrm{max}}^3 = 6\pi^2 \frac{N_{\mathrm{at}}}{V}$$

and therefore the cut-off frequency is (using our linear dispersion)

$$\omega_{\mathrm{D}}^3 = (c\,k_{\mathrm{max}})^3 = 6\pi^2\,c^3\,\frac{N_{\mathrm{at}}}{V}$$

The density of states from equation (1.3.2) is

$$\begin{aligned} g(\omega) &= \frac{1}{2}\,\frac{1}{\pi^2}\,V\,k^2\,\frac{1}{v_g} \\ &= \frac{1}{2}\,\frac{1}{\pi^2}\,V\,\frac{\omega^2}{c^3} \end{aligned}$$

since $\omega = ck$ and the group velocity is $v_g = \mathrm{d}\omega/\mathrm{d}k = c$, and so the internal energy is

$$\begin{aligned} U &= 3\int_0^{\omega_{\mathrm{D}}} \frac{1}{2}\,\frac{1}{\pi^2}\,V\,\frac{\omega^2}{c^3}\,\frac{\hbar\omega}{\mathrm{e}^{\hbar\omega/k_{\mathrm{B}}T} - 1}\,\mathrm{d}\omega \\ &= \frac{3V\hbar}{2\pi^2 c^3}\int_0^{\omega_{\mathrm{D}}} \frac{\omega^3}{\mathrm{e}^{\hbar\omega/k_{\mathrm{B}}T} - 1}\,\mathrm{d}\omega \end{aligned}$$

where we have included a factor of 3 because there are three allowed modes for each k: one longitudinal and two transverse; for simplicity we take just one velocity c for all modes.

To get the heat capacity we need to differentiate with respect to T. T only appears in the integrand and the limit ω_D does not depend on T. Indeed ω_D *defines* the Debye temperature θ_D if we write

$$k_\mathrm{B}\theta_\mathrm{D} = \hbar\omega_\mathrm{D} \tag{1.6.6}$$

from which

$$\theta_\mathrm{D} = \frac{\hbar c}{k_\mathrm{B}}\left(\frac{6\pi^2 N_\mathrm{at}}{V}\right)^{\frac{1}{3}}$$

Now

$$\frac{\mathrm{d}}{\mathrm{d}T}\,\frac{\omega^3}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1} = \frac{\omega^3}{\left(\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1\right)^2}\;\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}\;\frac{\hbar\omega}{k_\mathrm{B}T^2}$$

and so

$$C_\mathrm{ph} = \frac{3V\hbar^2}{2\pi^2 c^3 k_\mathrm{B} T^2}\int_0^{\omega_\mathrm{D}} \frac{\omega^4 \mathrm{e}^{\hbar\omega/k_\mathrm{B}T}}{\left(\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1\right)^2}\,\mathrm{d}\omega$$

We simplify the integral by making the substitution

$$x = \frac{\hbar\omega}{k_\mathrm{B}T} \quad ; \qquad \mathrm{d}\omega = \frac{k_\mathrm{B}T}{\hbar}\mathrm{d}x$$

so that

$$\begin{aligned} C_\mathrm{ph} &= \frac{3V\hbar^2}{2\pi^2 c^3 k_\mathrm{B} T^2}\int_0^{\hbar\omega_\mathrm{D}/k_\mathrm{B}T}\left(\frac{k_\mathrm{B}T}{\hbar}\right)^4 \frac{x^4\,\mathrm{e}^x}{\left(\mathrm{e}^x-1\right)^2}\,\frac{k_\mathrm{B}T}{\hbar}\,\mathrm{d}x \\ &= 9N_\mathrm{at}k_\mathrm{B}\left(\frac{T}{\theta_\mathrm{D}}\right)^3\int_0^{\theta_\mathrm{D}/T}\frac{x^4\,\mathrm{e}^x}{\left(\mathrm{e}^x-1\right)^2}\,\mathrm{d}x \end{aligned}$$

You can easily see this by gathering terms and using

$$\frac{N_\mathrm{at}}{\theta_\mathrm{D}^3} = \frac{v k_\mathrm{B}^3}{\hbar^3 c^3 6\pi^2}$$

from our definition of the Debye temperature (1.6.6).

It is not at all easy to see that as $T \to \infty$ the heat capacity tends to the Law of Dulong and Petit. The best way is to evaluate the function numerically. You can see it plotted in Kittel, p124, where you can also see demonstrated that at low temperature the heat capacity becomes

$$\boxed{C_\mathrm{ph} = \frac{12\pi^4}{5}\,N_\mathrm{at}\,k_\mathrm{B}\left(\frac{T}{\theta_\mathrm{D}}\right)^3}$$

This is the important Debye "T^3 Law" which says that at low temperature the heat capacity of crystals is proportional to T^3. This, rather than the Einstein exponential drop to zero, is what is observed.

The fact that the heat capacity falls to zero at low temprature is a *quantum mechanical effect.* The Einstein model has just one frequency ω_{E}. When the temperature falls below about $\theta_E = \hbar\omega_E/k_{\mathrm{B}}$ then the probability of that frequency being excited becomes exponentially small, as predicted by the Bose–Einstein distribution. The Debye model fixes this since the $w = ck$ dispersion allows modes that have frequencies that approach zero. There is a simple way to understand the T^3 Law. At high temperature all modes up to frequency ω_{D} are excited and these are enclosed within a sphere in **k**-space of radius $k_D = \omega_{\mathrm{D}}/c$. At some low temperature T we suppose that only modes with energies less than $ck = \hbar\omega = k_{\mathrm{B}}T$ will be excited. These occupy a sphere of radius k so instead of there being all $3N_{\mathrm{at}}$ modes excited, only a fraction k^3/k_D^3, the ratio of the volumes of these spheres, are excited. Hence there are

$$3N_{\mathrm{at}}\left(\frac{k}{k_D}\right)^3 = 3N_{\mathrm{at}}\left(\frac{T}{\theta_{\mathrm{D}}}\right)^3$$

excited modes and so the internal energy $3N_{\mathrm{at}}k_{\mathrm{B}}T$ is reduced by the same factor to

$$U \approx 3N_{\mathrm{at}}k_{\mathrm{B}}T\left(\frac{T}{\theta_{\mathrm{D}}}\right)^3$$

and hence the heat capacity is

$$C_{\mathrm{ph}} \approx 12N_{\mathrm{at}}k_{\mathrm{B}}\left(\frac{T}{\theta_{\mathrm{D}}}\right)^3$$

The main reason the numerical factor doesn't agree with the proper limit (*i.e.*, $12\pi^4/5 \neq 12$) is the assumption that the internal energy is found by multiplying the high temperature value by the fraction of modes excited. This is because the remaining low temperature modes will not behave classically, each carrying an energy of $k_{\mathrm{B}}T$.

1.7 Particle–particle interactions—phonon scattering

We observe elementary excitations such as phonons through their interactions with other, test particles in scattering experiments. (See Ibach and Lüth, chapter 3.)

According to de Broglie, quantum wave particle duality asserts that particle/waves have an energy

$$E = \hbar\omega = h\nu$$

and a momentum

$$p = \hbar k$$

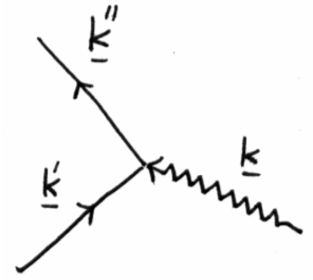
Crystal periodicity introduces an important feature as we have seen, that we can replace the wavevector $\mathbf{k}$ by $\mathbf{k}+\mathbf{g}$ if $\mathbf{g}$ is a vector of the reciprocal lattice ($\mathbf{g} = \sum_i m_i\mathbf{b}_i$) without any detectable change in the physics. Thus the momentum of a phonon is undetermined to within an additive amount $\hbar \mathbf{g}$

$\hbar \mathbf{k}$ is called the crystal momentum or pseudomomentum. The *physical* momentum is

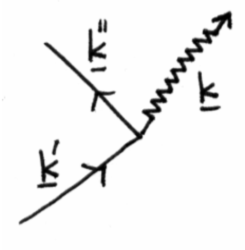
$$\sum_j M \frac{\mathrm{d}}{\mathrm{d}t} \delta \mathbf{r}_j$$

and this is zero for all phonons except those at $\mathbf{k} = 0$ whose displacements describe a uniform displacement of the lattice.

When particles interact, energy and momentum are conserved. The simplest is a three particle interaction, *e.g.*, an electron scattered by a phonon.



phonon absorption



phonon emission

Here a straight line represents an electron and a zig-zag line represents a phonon (see Kittel, chapter 4, figure 1).

Energy conservation requires

$$\begin{aligned} E_{\mathbf{k}'} + \hbar\omega_{\mathbf{k}} &= E_{\mathbf{k}''} && \text{absorption} \\ E_{\mathbf{k}'} &= \hbar\omega_{\mathbf{k}} + E_{\mathbf{k}''} && \text{emission} \end{aligned}$$

Consider just the phonon emission (creation) process. Energy conservation would imply (the electrons have momentum $\hbar\omega_{\mathbf{k}'}$ and $\hbar\omega_{\mathbf{k}''}$)

$$\hbar\omega_{\mathbf{k}'} = \hbar\omega_{\mathbf{k}''} + \hbar\omega_{\mathbf{k}}$$

and momentum conservation requires

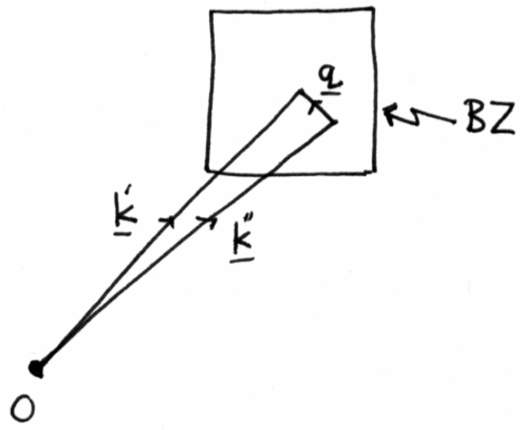
$$\mathbf{k} = \mathbf{k}' - \mathbf{k}'' \equiv \mathbf{q}$$

$\mathbf{q}$ is called the scattering vector or momentum transferred to the scatterer—*i.e.*, the crystal. Note well, that Ibach and Lüth in their chapter 3 use an opposite sign so their $\mathbf{K}$ is our $-\mathbf{q}$ but this is very unusual in scattering physics.

If $\mathbf{k}'$ and $\mathbf{k}''$ are such that their difference $\mathbf{q}$ *lies within the first Brillouin zone* then the wavevector of the phonon created is

$$\mathbf{k} = \mathbf{q}$$

For example



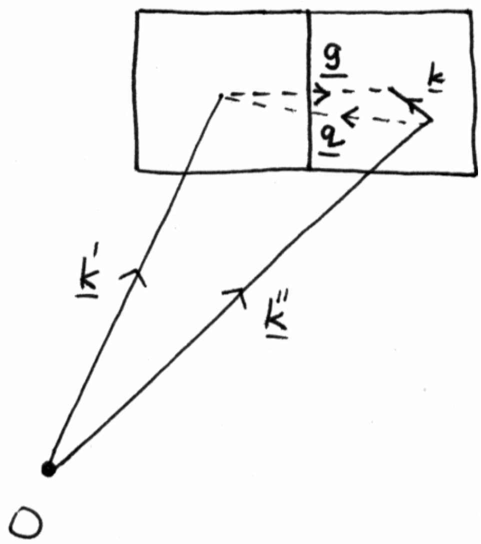
This is called a Normal scattering process, or *N process.* However if the difference **q** is larger than the dimension of the Brillouin zone this implies that the emitted phonon wavevector lies outside the first Brillouin zone which is not permitted in the theory. But we can always *reduce* this wavevector by subtracting any number of reciprocal lattice vectors until the remaining wavevector *does* lie inside the first zone. The phonon wavevector resulting is not **q** but **q** − **g** and the momentum balance is written

$$\mathbf{k}' - \mathbf{k}'' = \mathbf{q} = \mathbf{k} + \mathbf{g}$$

or

$$\mathbf{k} = \mathbf{q} - \mathbf{g}$$

For example



This is called an Umklapp process or *U process.* The scattering vector (momentum transfer) is larger than in the Normal case by an amount $\hbar\mathbf{g}$ but the emitted phonon has the same wavevector as before. We have not violated momentum conservation; instead we say that the lattice can always provide a momentum $\pm\hbar\mathbf{g}$ in a scattering process. You can think of an Umklapp process as the creation or destruction of a phonon accompanied by a Bragg reflection—the phonon with wavevector $\mathbf{q}$ is reflected to $\mathbf{q}-\mathbf{g}$ by the lattice planes whose normal is $\mathbf{g}$.

What we said about electron scattering applies equally well to the scattering of other particles. We need to look at some typical energies and wavelengths in experimentally produced beams (see Ibach and Lüth, figure 3.9).

	energy (eV)	$\lambda = 2\pi/k$ (Å)
Phonons	$\sim \hbar\omega_\mathrm{D} = k_\mathrm{B}\theta_\mathrm{D}$ $\approx 0.025(\mathrm{eV})$ for θ_D = Room Temp.	$\infty \to \sim 10$
Electrons	10–1000	4–0.5
Photons	1000–100 000	20–0.1
Thermal neutrons	0.01–1	3–0.4

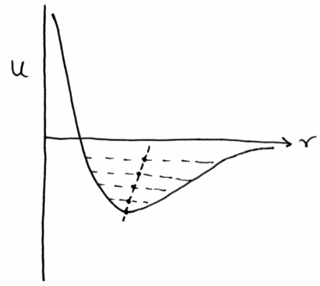
All these beams have wavelengths on the order of the lattice spacing and can be used in the study of solids. But only neutrons have sufficiently low energy that the energy change due to phonon scattering can be detected. Hence thermal neutrons are used to measure phonon dispersion curves (see Ibach and Lüth, figure 4.4).

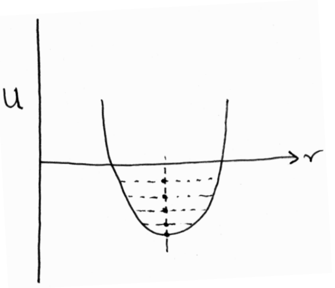
1.8 Anharmonic effects

In equation (1.1.2) we expanded the potential energy in a Taylor series up to second order in the displacements of the atoms from their "official" positions. This is called the harmonic approximation. As we saw when we quantised the lattice waves, a consequence is that the phonons behave as independent simple harmonic oscillators—they cannot interact or scatter against each other. This precisely true for *photons*, they live forever if they are not scattered by matter: hence the 4°K cosmic background radiation.

Errors arising from the harmonic approximation are,

1. No thermal expansivity. Atoms are vibrating in a quadratic potential so at all temperatures their mean position is the same:





As temperature increases for an oscillator in an anharmonic potential well the mean position increases since the occupation of higher energy excited states increases. For details of how to use this to calculate the free energy as a function of T, and to estimate the thermal expansion coefficient, see Ibach and Lüth, section 5.5. It is notable that *some* crystals in *some* (usually low) temperature ranges shrink upon heating (at least in one of more of the unit cell lattice vector directions). This negative thermal expansion, or rather zero thermal expansion, is exploited by ceramic materials used in cooker hobs.

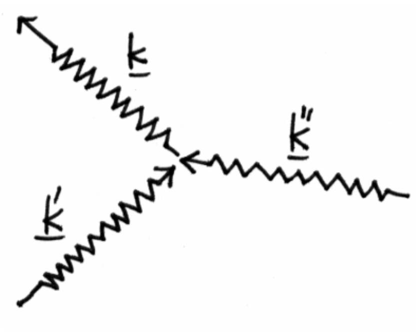
2. In the harmonic approximation the heat capacities at constant volume, C_v and constant pressure C_p are identically equal. While they are very different in gasses (remember for one mole of ideal gas $C_p - C_v = R$) they are anyway almost equal in a solid.
3. In the harmonic approximation the elastic constants are independent of temperature, volume or pressure.

1.8.1 Phonon–phonon scattering

You may not be surprised that after including third order terms in the potential energy the second quantised hamiltonian has extra terms such as

$$a^+_{\mathbf{k}} a_{\mathbf{k}'} a_{\mathbf{k}''}\, \delta_{\mathbf{k}+\mathbf{k}'+\mathbf{k}'',\mathbf{g}}$$

Thus phonon scattering is permitted. The above expression implies that phonons of momentum $\hbar\mathbf{k}'$ and $\hbar\mathbf{k}''$ are destroyed while creating a phonon of momentum $\hbar\mathbf{k}$. The Kronecker delta assures momentum conservation since it is zero unless $\mathbf{k} + \mathbf{k}' + \mathbf{k}'' = \mathbf{g}$.



If $\mathbf{g}$ is zero this is a Normal process.

2. Electrons

In what sense is an electron an *elementary excitation*? In low energy physics electrons are not created or destoyed. However, we will see that electrons in the ground state of a crystal belong to a Fermi sea of electrons and almost all of these are frozen into their states by the Pauli exclusion principle. They only therefore become interesting in terms of phenomena when those nearest to the surface of the sea, the Fermi surface, "break away" into excited states, creating thereby an electron–hole pair. It is these elementary excitations that we study when we probe the solid with lasers.

In condensed matter and molecular physics, or again if you like, low energy physics, we always distingush *core* and *valence* electrons. The latter, which originate from the outer shell of the atomic structure, are those which take part in bonding and the response to perturbations. The lower-lying core electrons remain bound to the nucleus barely aware that their parent atoms have participated in the formation of a molecule or solid. The bonding that holds these atoms together is a consequence of the motions of the valence electrons which is described by the Schrödinger equation. In this the potential takes two parts; the first is that due to the combined Coulomb force of the nucleus and core electrons, oddly enough called the *external potential.* The other term is the Coulomb interaction between all the valence electrons themselves. If we can we like to simplify this last term by postulating that each electron moves in an average electric field due to all the other valence electrons. This is called the one electron approximation and allows us to write down a single, independent Schrödinger equation for each electron. Why should this work? Why do electrons appear to be "independent?" The electron–electron interaction is strong; why can we ignore it? We often approximate the electron gas in a solid, especially a metal, as *jellium*; that is a uniformly dense electron gas of $n_\mathrm{e} = N_\mathrm{e}/V$ electrons per unit volume moving in a *fixed*, *uniform* positive background charge intended to mimic the nuclei and core electron charge. As each electron moves it keeps away from the other electrons because of both the Coulomb repulsion and the Pauli exclusion—this is called correlation and exchange. So at the microscopic level the electron gas is not uniform, each electron digs itself an *exchange and correlation hole*, so as to expose a patch of the bare background positive charge. You can show that the amount of positive charge exposed is exactly minus the charge on the electron. So as it moves, the electron drags with it its exchange and correlation hole, and this object or *quasiparticle* is overall neutral and so does not interact strongly with other quasiparticles. So when we pretend that we are dealing with independent electrons, we are actually dealing with these quasiparticles; they may have an "effective mass" different from m, but they do each obey an independent particle Schrödinger equation. The theory behind all this is called density functional theory and has revolutionised quantitative calculations in condensed matter and molecular physics. Oddly enough you won't find it mentioned in textbooks; the exception being the book by Marder.

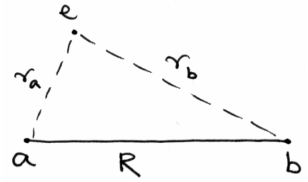
At each nucleus an electron sees an infinitely deep potential energy well. So why do we assert that the free electron model amounts to a reasonable model for the electrons in a solid? How can we imagine that a uniform positive background is in any sense a plausible representation of the periodic array of potential wells? The answer lies in the core electrons. These occupy a set of sharp energy levels in the Coulomb potential, rather like the energy levels of an infinite square well. These are occupied according to the *Aufbau* rule and are not available to the valence electrons because of the Pauli principle. In this way the presence of the core electrons provides a repulsive potential that, in so called *sp*-bonded solids at least, practically cancels the attractive potential from the nucleus. The result is that the electrons see only a weak, sometimes even repulsive, potential at the atomic cores. This is the subject of pseudopotential theory and again you don't often find it in textbooks although it provides the foundation for some of the most spectacular theoretical predictions in the last 20 years. The free electron approximation is therefore for many purposes a good one. We will see that it is a problem that can be solved; historically this was done in the very early days of the new quantum mechanics by Arnold Sommerfeld in the late 1920's and the free electron gas is hence often called the "Sommerfeld model." Again historically, this was of vast importance; as with Einstein's theory of the heat capacity 20 years earlier it cleared up an outstanding inconsistency between observation and classical physics. In the one case the question was, why does the heat capacity deviate from the Dulong and Petit law at low temperature? In the other, why do the electrons not contribute a further $\frac{1}{2}k_{\mathrm{B}}T$ of energy from each degree of freedom to the heat capacity? In other words, in a monovalent metal why is the high temperature heat capacity $3k_{\mathrm{B}}$ per atom and not $(9/2)k_{\mathrm{B}}$?

2.1 The hydrogen molecule

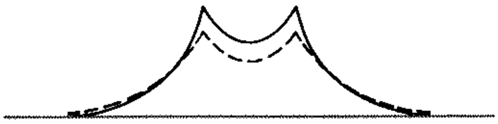
We are now going to make a wide digression to discuss the hydrogen molecule. This is not really solid state physics. On the other hand it is the starting point for the description of the quantum mechanics of the chemical bond; moreover it is fundamental to the understanding of magnetism as it introduces the electron property of spin, the Pauli principle, exchange and correlation, Hartree and Hartree–Fock approximations and configuration interaction.

2.1.1 The H_2^+ molecular ion

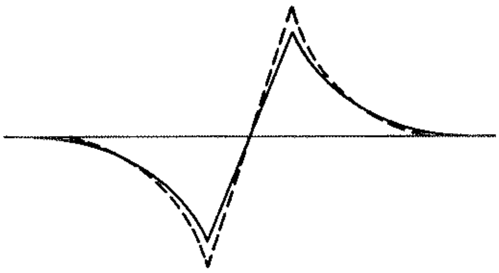
We start with the H_2^+ molecular ion because this is one of the rare quantum mechanical problems that can be solved *exactly*, although the mathematics is complicated. The molecule and its electron are sketched as follows to define the distances $r_a = |\mathbf{r} - \mathbf{R}_a|$ and $r_b = |\mathbf{r} - \mathbf{R}_b|$ between the two protons, a at position $\mathbf{R}_a$ and b at position $\mathbf{R}_b$, and the electron at position $\mathbf{r}$.



In this section 2.1 we will work in atomic units of length, $a_0 = 1$, so distances are in units of the Bohr radius, $a_0 = 4\pi\epsilon_0\hbar^2/me^2$. At a bond length $R = 2$ the two lowest energy solutions of the Schrödinger equation look like this,



$1\sigma_g$



$1\sigma_u$

compared with a linear combination of atomic orbitals (LCAO)—the dotted line

$$\psi(1\sigma_g) = \mathrm{e}^{-\alpha r_a} + \mathrm{e}^{-\alpha r_b} \tag{2.1.1a}$$
$$\psi(1\sigma_u) = \mathrm{e}^{-\alpha r_a} - \mathrm{e}^{-\alpha r_b} \tag{2.1.1b}$$

If $\alpha = 1$ each term is the exact $1s$ orbital solution to the Schrödinger equation for the hydrogen atom and it's their linear combination that is shown in the dotted line in the figure. We introduce α as a "variational parameter." Note that the solution to the Schrödinger equation is drawn in the figure but I have not derived the wavefunction or even given you its mathematical form (the maths is just too difficult). The dotted line is the LCAO with $\alpha = 1$. If I did a *variational* calculation, using (2.1.1) as trial functions and varying α to minimise the energy, then the result would be indistinguishable by eye from the exact result. This is the power of the variational method: we arrive at a very precise result but still in the simple mathematical form of the LCAO.

It's not too hard to normalise (2.1.1), although you need a three dimensional integral using cylindrical symmetry. Having done that, we can write (2.1.1) in terms of normalised molecular orbitals,

$$u_g = \left[\left(\frac{2\pi}{\alpha^3}\right)(1+S))\right]^{-\frac{1}{2}} \left(\mathrm{e}^{-\alpha r_a} + \mathrm{e}^{-\alpha r_b}\right) \tag{2.1.2a}$$
$$u_u = \left[\left(\frac{2\pi}{\alpha^3}\right)(1-S))\right]^{-\frac{1}{2}} \left(\mathrm{e}^{-\alpha r_a} - \mathrm{e}^{-\alpha r_b}\right) \tag{2.1.2b}$$

where

$$S = \frac{\alpha^3}{\pi} \int \mathrm{e}^{-\alpha r_a}\, \mathrm{e}^{-\alpha r_b}\, \mathrm{d}\mathbf{r}$$

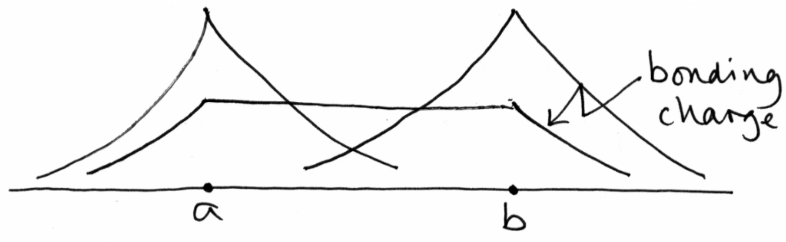
S, of course, depends on the distance between the nuclei R and in fact you can do the integral to get

$$S = \mathrm{e}^{-\alpha R}\left(1 + \alpha R + \frac{1}{3}\alpha^2 R^2\right)$$

Note that the electron probability density, or charge density in the H_2^+ ion is

$$u_g^2 = \frac{1}{2(1+S)} \frac{\alpha^2}{\pi} \left(\mathrm{e}^{-2\alpha r_a} + \mathrm{e}^{-\alpha(r_a+r_b)} + \mathrm{e}^{-2\alpha r_b}\right)$$

This is a superposition of two reduced atomic charge densities of amount $1/2(1+S)$ and a bonding charge of amount $S/1+S$. Note that the effect of bonding is to remove charge from the atomic orbitals and pile it up between the atoms. *This is an operational definition of a chemical bond.*



2.1.2 The H_2 molecule

Now we come to study the H_2 molecule, which is indeed the H_2^+ molecular ion with an added electron. This is really a many electron problem; and indeed it's the simplest such problem and hence worth spending some time on. Very often when wondering how to proceed in a complicated question of electron interactions, the great solid state physicist Volker Heine would say, "let us go back to the hydrogen molecule..."

There are two distinguished approaches to solving approximately the Schrödinger equation for the hydrogen molecule. These are the molecular orbital method of Hund and Mulliken and the valence bond method of Heitler and London.

2.1.2.1 The molecular orbital approach

Hund and Hartree independently introduced the notion of a self consistent field. The solution to the Schrödinger equation must be one in which which each electron moves in the electrostatic field of the nuclei plus the average or mean field of the other electron(s). The solution of the Schrödinger equation is self consistent if the wavefunction squared is a solution of the Poisson equation for the mean electrostatic field that appears in the Schrödinger equation.

The wavefunctions u_g and u_u (2.1.2) of the H_2^+ molecular ion are self consistent in the sense that they are symmetric or antisymmetric in a plane bisecting the line joining the nuclei and hence, when squared, will reproduce correctly the symmetry of the electrostatic potential.

So it makes sense to start with these; and since we may put two electrons into each state (having opposite spins) we could form the product wavefunction

$$u_g(s_1\mathbf{r}_1)\, u_g(s_2\mathbf{r}_2)$$

which is a solution of the sum of two independent H_2^+ hamiltonians for electrons 1 and 2.

Here, $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions of the electrons and s_1 and s_2 their spins. Since the spins must be of opposite sign we write this as

$$u_g(1)\alpha(1)u_g(2)\beta(2) \tag{2.1.3}$$

using the shorthand $\mathbf{r}_1 \to 1$, $s_1 \to 1$ *etc.* The *spin functions* α and β are defined as follows. An electron spin s is either plus or minus one (in units of $\frac{1}{2}\hbar$) and the function $\alpha(i) \equiv \alpha(s_i)$ is defined to be one if electron i has spin $+\frac{1}{2}\hbar$ and zero if it has spin $-\frac{1}{2}\hbar$; conversely $\beta(i) \equiv \beta(s_i)$ is zero if electron i has spin $+\frac{1}{2}\hbar$ and one if it has spin $-\frac{1}{2}\hbar$. In short,

$$\alpha(1) = 1 \quad \alpha(-1) = 0 \quad \beta(1) = 0 \quad \beta(-1) = 1$$

The wavefunction (2.1.3) would correspond to a self consistent field in the Hartree approximation. However this wavefunction does not respect the antisymmetry of fermions with respect to exchange of position and spin coordinates. If we want to combine functions like $u_g(1)\alpha(1)$ and $u_g(2)\beta(2)$ into products and *guarantee* antisymmetry, this can be done by forming Slater determinants of these. Thereby we replace (2.1.3) by

$$\frac{1}{\sqrt{2}}\begin{vmatrix} u_g(1)\alpha(1) & u_g(2)\alpha(2) \\ u_g(1)\beta(1) & u_g(2)\beta(2) \end{vmatrix} = u_g(1)u_g(2)\Xi_1 \tag{2.1.4}$$

where

$$\Xi_1 = \frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2) - \beta(1)\alpha(2)\right]$$

is the singlet spin function. The wavefunction is a singlet because the total spin is zero. Equation (2.1.4) is the self consistent field solution in the Hartree–Fock approximation.

We will write the atomic s-orbitals for electrons 1 and 2 alternatively at nuclei a and b (including the variational parameter α) as

$$\begin{aligned}
\psi_a(\mathbf{r}_1) &= a(1) = \sqrt{\frac{\alpha^3}{\pi}}\,\mathrm{e}^{-\alpha|\mathbf{r}_1-\mathbf{R}_a|} \\
\psi_b(\mathbf{r}_2) &= b(2) = \sqrt{\frac{\alpha^3}{\pi}}\,\mathrm{e}^{-\alpha|\mathbf{r}_2-\mathbf{R}_b|} \\
\psi_a(\mathbf{r}_2) &= a(2) = \sqrt{\frac{\alpha^3}{\pi}}\,\mathrm{e}^{-\alpha|\mathbf{r}_2-\mathbf{R}_a|} \\
\psi_b(\mathbf{r}_1) &= b(1) = \sqrt{\frac{\alpha^3}{\pi}}\,\mathrm{e}^{-\alpha|\mathbf{r}_1-\mathbf{R}_b|}
\end{aligned} \tag{2.1.5}$$

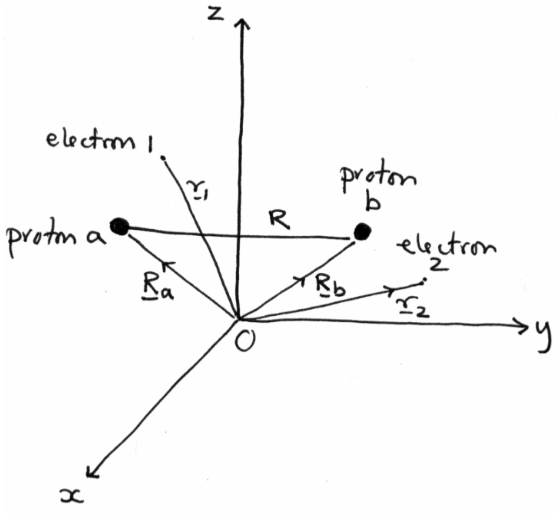
We can then see that the spatial part of the wavefunction in the molecular orbital is

$$\begin{aligned}
u_g(1)u_g(2) &= \frac{1}{2(1+S)}\Big(\left(a(1)+b(1)\right)\left(a(2)+b(2)\right)\Big) \\
&= \frac{1}{2(1+S)}\Big(\underbrace{a(1)a(2)+b(1)b(2)}_{\text{“ionic function”}} + \underbrace{a(1)b(2)+b(1)a(2)}_{\text{“H–L function”}}\Big)
\end{aligned} \tag{2.1.6}$$

In the Hartree–Fock molecular orbital the electron spends half its time in an “ionic function,” *i.e.*, having both electrons associated with the same atom. The wavefunction is said to be uncorrelated: the electrons don't seem to mind being attached to the same nucleus. One might want to question this assumption.

2.1.2.2 The valence bond approach

Let us carry through the Heitler–London theory from the beginning. The geometry of the problem is as in the figure. It looks more complicated than it is. The nuclei (protons) are labelled a and b and are placed at positions $\mathbf{R}_a$ and $\mathbf{R}_b$ with respect to the origin of some cartesian coordinate system. Similarly electrons labelled 1 and 2 are at positions $\mathbf{r}_1$ and $\mathbf{r}_2$.



The hamiltonian for the system of electrons and fixed nuclei is:

$$
\begin{aligned}
H = & \left(-\frac{\hbar^2}{2m}\nabla_1^2 - \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_a|}\right) + \left(-\frac{\hbar^2}{2m}\nabla_2^2 - \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_b|}\right) \\
& - \frac{1}{4\pi\epsilon_0}\left(\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_b|} - \frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_a|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{e^2}{R}\right) \\
& \equiv H_a + H_b + V_{\text{int}}
\end{aligned}
$$

1. The hamiltonian has electron (but not nucleus, because they're fixed) kinetic energy, electron-nucleus, electron-electron and nucleus-nucleus terms.
2. H_a and H_b, the quantities in large parentheses, are the hamiltonians for isolated hydrogen atoms labelled a and b.

We now use first-order perturbation theory for V_{int}. We will only need to know the wavefunctions of the *un*perturbed system; namely the eigenfunctions of $H_a + H_b$ which are *products* of the eigenfunctions of the free hydrogen atoms, equations (2.1.5), such as $a(1)b(2)$ and $a(2)b(1)$. Let us for now simply require that

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\Psi(\mathbf{r}_2, \mathbf{r}_1)|^2$$

as our fermion condition and include the spin functions later. Then the wavefunction of the unperturbed system must be

$$\Psi(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}\Big[a(1)b(2) \pm a(2)b(1)\Big]$$

Our best estimate of the perturbation energy as a function of R is

$$E(R) = \frac{\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\, \overline{\Psi} V_{\text{int}} \Psi}{\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\, \overline{\Psi}\Psi}$$

The bar indicates complex conjugate. But the wavefunctions in this case are real. Now using

$$\hat{V}_{\text{int}} = \frac{1}{4\pi\epsilon_0}\left(-\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_b|} - \frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_a|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{e^2}{R}\right)$$

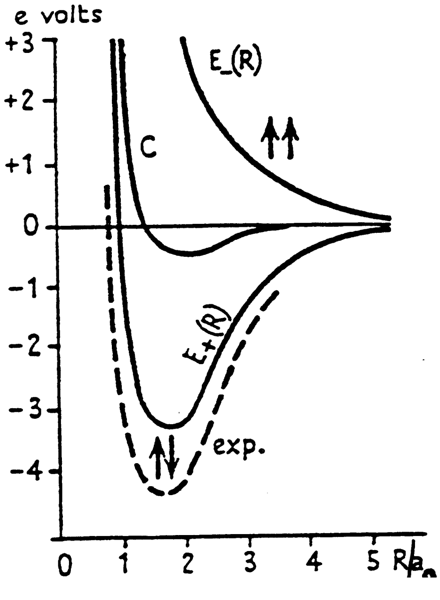
we get two solutions corresponding to the plus and minus signs respectively:

$$E_+(R) = \frac{C+A}{1+S^2} \qquad \text{and} \qquad E_-(R) = \frac{C-A}{1-S^2}$$

where

$$\begin{aligned}
C &= \frac{1}{4\pi\epsilon_0}\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\, \bar{a}(1)a(1)\left(-\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_B|} - \frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_A|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{e^2}{R}\right)\bar{b}(2)b(2) \\
A &= \frac{1}{4\pi\epsilon_0}\iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\, \bar{a}(1)b(1)\left(-\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_B|} - \frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_A|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{e^2}{R}\right)\bar{b}(2)a(2) \\
S^2 &= \iint \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\, \bar{a}(1)b(1)\bar{b}(2)a(2)
\end{aligned}$$

The binding energy as a function of R is shown here.



1. The picture is superficially the same as in the molecular orbital theory. There is an attractive "bonding" solution and a repulsive solution—called "antibonding" in MO theory.
2. Neglecting the overlap S^2 (*i.e.*, to first order in S), the "bonding–antibonding splitting" is a consequence of the appearance of the integral A which is negative and much larger in magnitude than C.

C is easily interpreted: it is the interaction due to $\hat{V}_{\text{int}}$ between the charge clouds $\psi_a^2(\mathbf{r}_1)$ and $\psi_b^2(\mathbf{r}_2)$ on the unperturbed atoms. C is called the "Coulomb integral." Note that there is a self interaction (often called the "Hartree energy") and interactions with the two nuclei.

The integral A which leads to the bonding is similarly interpreted as the interaction between the "exchange charge densities" $\overline{\psi}_a(\mathbf{r}_1)\psi_b(\mathbf{r}_1)$ and $\overline{\psi}_b(\mathbf{r}_2)\psi_a(\mathbf{r}_2)$. This is a wholly quantum mechanical effect arising from the indistiguishability principle.

Actually the condition $\Psi^2(\mathbf{r}_1\mathbf{r}_2) = \Psi^2(\mathbf{r}_2\mathbf{r}_1)$ is not the whole story. The Pauli exclusion principle states that the wavefunction for fermions is *antisymmetric* under exchange of position *and* spin coordinates. That is,

$$\Psi(s_1\mathbf{r}_1 s_2\mathbf{r}_2) = -\Psi(s_2\mathbf{r}_2 s_1\mathbf{r}_1)$$

This is a more restrictive requirement. To satisfy the Pauli principle, we must have for the wavefunction corresponding to $E_+(R)$

$$\Psi(\mathbf{r}_1 s_1 \mathbf{r}_2 s_2) = \Big(a(1)b(2) + b(2)a(1)\Big)\Big(\alpha(1)\beta(2) - \alpha(2)\beta(1)\Big) \tag{2.1.7}$$

and for that corresponding to $E_-(R)$

$$\Psi(\mathbf{r}_1 s_1 \mathbf{r}_2 s_2) = \Big(a(1)b(2) - b(2)a(1)\Big)\left\{\begin{array}{l}\alpha(1)\alpha(2)\\ \beta(1)\beta(2)\\ \alpha(1)\beta(2) + \alpha(2)\beta(1)\end{array}\right\} \tag{2.1.8}$$

Note that the symmetric orbital function is multiplied by an antisymmetric spin function and *vice versa*. Spin factors with a single term *e.g.*, $\alpha(1)\alpha(2)$ indicate that electrons 1 and 2 both have the same spin $(+1)$—the total spin for these states is ± 1 in units of $\hbar$. Two term spin factors are interpreted as mixtures of states in which the two electrons have opposite spins and the total spin is zero. Equation (2.1.7) is called a "singlet" state. Equation (2.1.8) has three degenerate levels having total spin $-1, 0, +1$ in units of $\hbar$, but whose degeneracy is lifted in a magnetic field. It is called a "triplet" state.

The essential conclusion from all this is that the *symmetric* orbital $a(1)b(2) + b(2)a(1)$ must have an *antisymmetric* spin function in which the spins have opposite signs and the *antisymmetric* orbital $a(1)b(2) - b(2)a(1)$ must be occupied by electrons with parallel spins. *Only hydrogen atoms whose electrons have opposite spins will form a molecule.*

In this description of the H_2 molecule, the "exchange energy" A is identified as the origin of the bonding. This is a quantum mechanical effect. The general definition of the exchange energy is half the energy difference between antiparallel and parallel spin states of a quantum mechanical two particle system,

$$A_x \equiv \frac{1}{2}\left(E_{\uparrow\downarrow} - E_{\uparrow\uparrow}\right).$$

You could show that for the Heitler–London H_2 molecule, neglecting S^2, the exchange energy is

$$\begin{aligned} &\frac{1}{2}\left(\, E_- - E_+ \,\right) \\ &\quad = \frac{1}{4\pi\epsilon_0}\iint \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2\, \bar{a}(1)\bar{b}(2)\left(-\frac{e^2}{|\mathbf{r}_1 - \mathbf{R}_b|} - \frac{e^2}{|\mathbf{r}_2 - \mathbf{R}_a|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}\right) a(2)b(1) \\ &\quad \equiv A_H. \end{aligned}$$

A_H is called the "Heisenberg exchange energy." A positive exchange energy favours antisymmetric or antiferromagnetic spin alignments. Ferromagnetism occurs when the exchange coupling is negative. More on that later when we discuss magnetism.

2.1.2.3 Molecular orbital and valence bond pictures compared

I want to discuss the contrast between the Heitler–London and molecular orbital pictures of the H_2 molecule. We have called $a(1)$ the wavefunction of electron 1 near the isolated hydrogen nucleus a and similarly $b(2)$. Of course, when the molecule is formed from its atoms, electron 1 may become associated with hydrogen nucleus b and *vice versa* in which case we will need $a(2)$ and $b(1)$ defined in equations (2.1.5). In the *molecular orbital* picture of Hund and Mulliken, we begin by recalling the bonding molecular orbital of the H_2^+ molecular ion, equation (2.1.2a). We could construct the *product* or "Hartree" wavefunction

$$u_g(1)\,\alpha(1)\,u_g(2)\,\beta(2)$$

by putting electrons 1 and 2 into a bonding orbital, having opposite spins, but this wavefunction does not have the fermion antisymmetry property. A better wavefunction is the Slater determinant,

$$\frac{1}{\sqrt{2}}\begin{vmatrix} u_g(1)\,\alpha(1) & u_g(2)\,\alpha(2) \\ u_g(1)\,\beta(1) & u_g(2)\,\beta(2) \end{vmatrix} = u_g(1)\,u_g(2)\,\Xi_1,$$

where

$$\Xi_1 = \frac{1}{\sqrt{2}}\left[\alpha(1)\,\beta(2) - \beta(1)\,\alpha(2)\right]$$

is the *singlet* spin function; that is, the total spin is zero. The point is that the MO picture amounts to the Hartree–Fock approximation since the wavefunction is a single Slater determinant.

In contrast, the Heitler–London or *valence bond* wavefunction for the H_2 singlet ground state is

$$[\, a(1)\, b(2) + a(2)\, b(1)\,]\, \Xi_1, \tag{2.1.9}$$

and the first excited, triplet, states are

$$[\, (a(1)\, b(2) + a(2)\, b(1)\,] \times \begin{cases} \alpha(1)\, \alpha(2), \\ \alpha(1)\, \beta(2) - \beta(1)\, \alpha(2), \\ \beta(1)\, \beta(2). \end{cases}$$

The physical picture is that if electron 1 is associated largely with nucleus a, then electron 2 will be found near nucleus b, and *vice versa.* So instead of the electrons sharing a bonding MO, they are keeping well apart *in a correlated state.* Mathematically, this is reflected in the fact that (2.1.9) cannot be written as a single determinant. Indeed, consider the four determinants

$$\frac{1}{\sqrt{2}} \begin{vmatrix} a(1)\, \chi_a(1) & b(1)\, \chi_b(1) \\ a(2)\, \chi_a(2) & b(2)\, \chi_b(2) \end{vmatrix}$$

arising when (*i*) χ_a and χ_b are both α; (*ii*) both β; (*iii*) α, β and (*iv*) β, α respectively. You will find that (2.1.9) is the sum of determinants (*iii*) and (*iv*). The triplet states which are degenerate in the absence of a magnetic field are, respectively: determinant (*i*); the difference of (*iii*) and (*iv*); and (*ii*).

So, the valence bond wavefunction is correlated, having the right properties upon dissociation that each nucleus ends up with one electron of a different spin. This is always the best starting point for studying a correlated system, and the wavefunction cannot be written as a single Slater determinant.

The molecular orbital $u_g(1)\, u_g(2)$ can be expanded as in (2.1.6) to give

$$u_g(1)\, u_g(2) = \frac{1}{2(1+S)} \Big(a(1)\, b(2) + b(1)\, a(2) + a(1)\, a(2) + b(1)\, b(2) \Big) \tag{2.1.10}$$

so the electron spends equal amounts of time in the "covalent" state as in the "ionic" state and there's no guarantee that on dissociation each nucleus will get one electron: there's an equal probability that both electrons will end up on the same nucleus. One way to "fix up" the MO picture is to include some amount of the antibonding MO determinant $u_u(1)\, u_u(2)$ from (2.1.2b) in a variational calculation. This is called *configuration interaction.*

If we depart from hydrogen-like orbitals $\mathrm{e}^{-\alpha r}$ and think of these as unknown functions with the properties that there are hopping integrals

$$\begin{aligned} h &= \langle a(1)\, | H |\, b(1) \rangle \\ &= \langle a(2)\, | H |\, b(2) \rangle < 0 \end{aligned}$$

overlap integrals

$$S = \langle a(1) \mid b(1) \rangle = \langle a(2) \mid b(2) \rangle$$

and a Coulomb repulsion when two electrons occupy the same site

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0}\left\langle a(1)a(2)\left|\frac{e^2}{|\mathbf{r}_1-\mathbf{r}_2|}\right|a(1)a(2)\right\rangle \\ &= \frac{1}{4\pi\epsilon_0}\left\langle b(1)b(2)\left|\frac{e^2}{|\mathbf{r}_1-\mathbf{r}_2|}\right|b(1)b(2)\right\rangle \end{aligned}$$

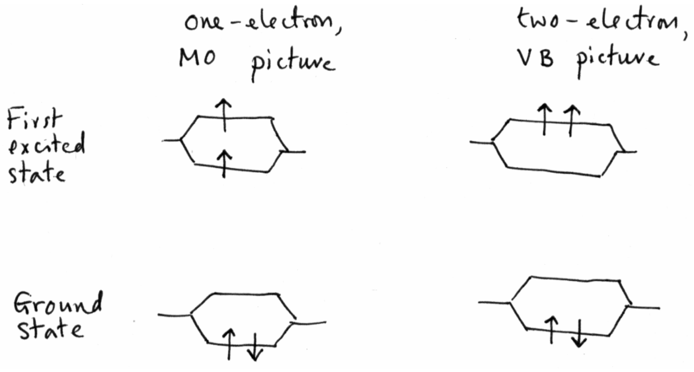
then solving the 2 × 2 secular equations between the bonding and antibonding molecular orbitals of equations (2.1.2) $u_g(1)\,u_g(2)$ and $u_u(1)\,u_u(2)$, as in equation (2.1.10), one finds that the ground state spatial wavefunction is

$$\Psi(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2(1+S^2)}}\left(\underbrace{a(1)b(2)+b(1)a(2)}_{\text{“H–L function”}} - \frac{2h}{U}\frac{1-S^2}{1+S^2}\left(\underbrace{a(1)a(2)+b(1)b(2)}_{\text{“ionic function”}}\right)\right)$$

as long as $U \gg |h|$. This results in a tight binding model (section 2.4) in which, by choosing parameters t and U we can go from the uncorrelated MO picture to the correlated VB picture.

The most important point to take away is that the MO picture is a one electron approximation. The valence bond description on the other hand is a two electron, or many electron theory. The way to tell is that in the former, all integrals are over the single electron coordinate of one or other of the electrons. In the many electron approach the wavefunction is a function of both or all the electron coordinates and integrals are many-dimensional. This is of course much harder and we try always to cast our physics into the one particle description; a typical instance being the physics of semiconductors, energy bands and so on. But there are some realms of physics in which the many particle treatment is essential. In solid state physics the outstanding example is superconductivity—there is simply no single particle picture for this. And hence, reluctantly, I cannot include superconductivity in this lecture course. On the whole, magnetism also falls into this category of phenomena. However, we will study this; and where we can, use the one electron approximation.

A clear way to see the difference between the one and two electron pictures is the following diagram, taken from figure 8.3 in Ibach and Lüth. The two energy level diagrams for the H_2 molecule seem to look the same. However in the MO description these are *single particle levels* and, like in the *Aufbau* principle in chemistry you populate these by adding one electron after the other. In this way the Pauli principle is not fundamental—you build it in by hand as you populate the one electron levels. In contrast energy levels in the many body picture are energies of the *combined* system of all the electrons. Either both electrons are in one or the other state. In fact it looks at first sight as if the Pauli principle is being violated in the right hand diagram. Do be sure you understand this vital difference between the one electron picture which we use every day, but which is a simplified abstraction, and the many electron picture which is actually the right physics. We have discussed in the introduction to this section 2, why and under what circumstances we are permitted to view the electronic structure problem in a single particle picture.



2.2 Free electron gas

Back to solids, we now describe in detail the "free electron gas," within the Sommerfeld model. We imagine a cubical box containing N_e electrons of side L neutralised by a uniform background of positive charge. The potential energy is constant everywhere and we apply periodic boundary conditions at the walls of the box. The Schrödinger equation, if the electrons are independent, is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi_\mathbf{k}(\mathbf{r}) = E_\mathbf{k}\psi_\mathbf{k}(\mathbf{r})$$

where m is the mass of the electron, and we have labelled the solutions with a quantum numbers

$$\mathbf{k} = (k_x,\, k_y,\, k_z)$$

The solutions are the plane waves

$$\psi_\mathbf{k}(\mathbf{r}) = \frac{1}{\sqrt{V}}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \frac{1}{\sqrt{V}}\,\mathrm{e}^{\mathrm{i}k_x x}\mathrm{e}^{\mathrm{i}k_y y}\mathrm{e}^{\mathrm{i}k_z z}$$

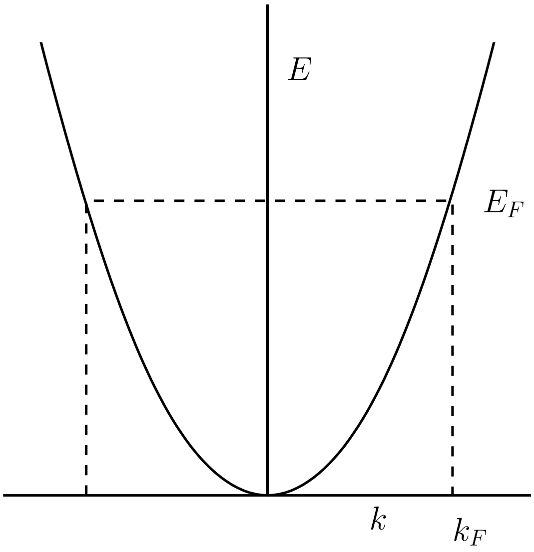
and

$$\boxed{E(\mathbf{k}) = \frac{\hbar^2}{2m}\left(k_x^2+k_y^2+k_z^2\right) = \frac{\hbar^2 k^2}{2m}} \qquad (2.2.1)$$

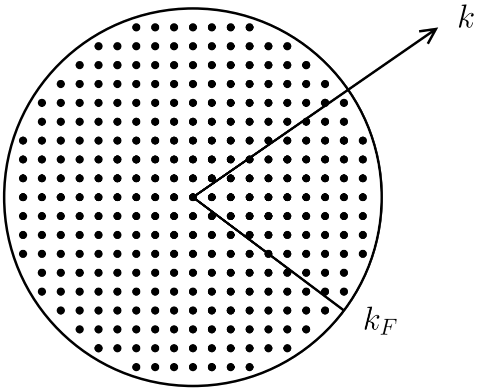
Periodic boundary conditions as discussed in section 1.3 place restrictions on the allowed values of the wavevector $\mathbf{k}$, namely

$$\mathbf{k} = \frac{2\pi}{L}\,(n_1,\, n_2,\, n_3) \qquad n_1,\, n_2,\, n_3 = 1, 2 \ldots \infty$$

so, in contrast to the lattice waves, there's an infinite number of allowed states, although they are discretely spaced out so that the density of points in $\mathbf{k}$-space is $V/(2\pi)^3$ points per unit volume of reciprocal space—the same as for lattice waves. Our picture is similar to the problem of lattice waves. We have an *isotropic dispersion relation*, $E_\mathbf{k} = E(k)$ which is, in this case, a parabola:



and a spotty **k**-space:



I have drawn a sphere in **k**-space for the following reason. Electrons are fermions. By the Pauli principle, each state can be occupied by at most two electrons of opposite spin. If there is a total of N_e electrons then the ground state at 0°K has all states in a sphere of radius k_F occupied and all other states unoccupied. The sphere in **k**-space separating these states is called the *Fermi sphere*[†] and k_F is the *Fermi wavevector*. We have

$$\begin{aligned} N_\mathrm{e} &= \begin{matrix}\text{Volume inside}\\ \text{Fermi sphere}\end{matrix} \div \begin{matrix}\text{Volume}\\ \text{per state}\end{matrix} \times 2 \text{ for spins} \\ &= \frac{4\pi}{3}\, k_\mathrm{F}^3 \times \frac{V}{(2\pi)^3} \times 2 \\ &= \frac{1}{3\pi^2}\, k_\mathrm{F}^3 V \qquad (2.2.2) \end{aligned}$$

that is,

$$k_\mathrm{F} = \left(3\pi^2 n_\mathrm{e}\right)^{\frac{1}{3}}$$

[†] In a crystal with anisotropic dispersion, this is called the *Fermi surface*.

where

$$n_{\mathrm{e}} = \frac{N_{\mathrm{e}}}{V}$$

is the electron density (number of electrons per unit volume). The lowest energy states with k increasing from 0 to k_{F} are occupied. The energy of the highest occupied state is called the *Fermi energy*:

$$\begin{aligned} E_{\mathrm{F}} &= \frac{\hbar^2 k_{\mathrm{F}}^2}{2m} \\ &= \frac{\hbar^2}{2m}\left(3\pi^2 \frac{N_{\mathrm{e}}}{V}\right)^{\frac{2}{3}} \end{aligned} \tag{2.2.3}$$

Now we can work out the density of states and hence the kinetic energy of the electron gas. The number of states with wavevector less than some k is, from eq (2.2.2),

$$\frac{1}{3\pi^2}\, k^3\, V$$

and from eq (2.2.1)

$$k^3 = \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{3}{2}} \tag{2.2.4}$$

and so the number of states with energy less than some E is

$$N(E) = \frac{V}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{3}{2}}$$

The density of states is the derivative of this with respect to energy, exactly as we saw in the case of lattice waves,

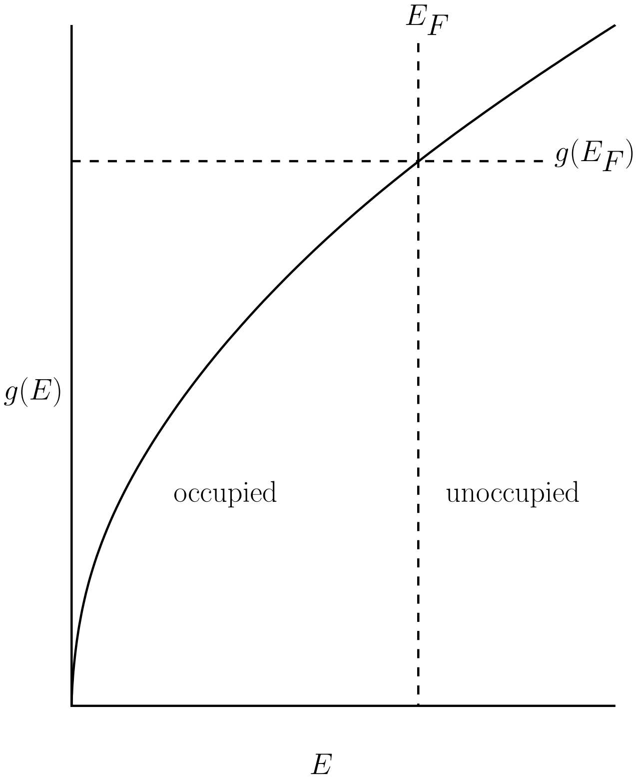
$$\boxed{\begin{aligned} n(E) &= \frac{\mathrm{d}N}{\mathrm{d}E} \\ &= \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \frac{V}{2\pi^2}\sqrt{E} \\ &= \frac{3}{2}\frac{N(E)}{E} \end{aligned}} \tag{2.2.5}$$

(In fact you can obtain the same result using the dispersion relation (2.2.1) and the general relation (1.3.2)) and so the density of states at the Fermi level, a quantity that we will use a lot in these notes is

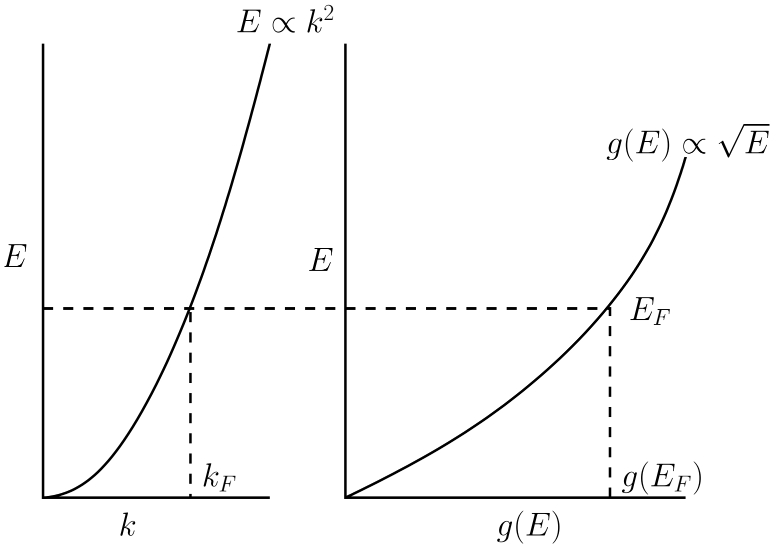
$$n(E_{\mathrm{F}}) = \frac{3}{2}\frac{N_{\mathrm{e}}}{E_{\mathrm{F}}} \tag{2.2.6}$$

(Be careful because Ibach and Lüth use $D(E)$ to denote the density of states *per unit volume.*)

We plot the density of states either in this way…



…or like this, to emphasise how the states are filled up, like a bath, to the Fermi level:



(I'm sorry, I've used $g(E)$ rather than $n(E)$ for the density of states.)

The internal energy at 0°K is obtained as we did in the Debye model for phonons:

$$\int_{\substack{\text{occupied}\\\text{states}}} \quad (\text{energy} \times \text{density of states})$$

It is purely <u>kinetic</u> energy because the potential energy is zero. The kinetic energy per electron is therefore

$$\begin{aligned} E_{\mathrm{kin}} &= \frac{1}{N_{\mathrm{e}}} \int_0^{E_{\mathrm{F}}} E\, n(E)\, \mathrm{d}E \\ &= \frac{V}{N_{\mathrm{e}}} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \frac{1}{2\pi^2}\, \frac{2}{5}\, E_{\mathrm{F}}^{\frac{5}{2}} \end{aligned}$$

But from eqns (2.2.4) and (2.2.2) we have

$$\frac{V}{N_{\mathrm{e}}} = \frac{3\pi^2}{k_{\mathrm{F}}^3} = 3\pi^2 \left(\frac{\hbar^2}{2m}\right)^{\frac{3}{2}} E_{\mathrm{F}}^{-\frac{3}{2}}.$$

Nearly everything cancels and we get

$$E_{\mathrm{kin}} = \frac{3}{5}\, E_{\mathrm{F}}$$

An electron at the Fermi level has a kinetic energy equal to E_{F} even at 0°K. Its group velocity is

$$v_g = \frac{\mathrm{d}(E/\hbar)}{\mathrm{d}k} = \frac{\hbar k_{\mathrm{F}}}{m} \tag{2.2.7}$$

When you work this out for a typical metal, you find a speed of order 10^6 ms^{-1}—about 100$^{\text{th}}$ the speed of light in a vacuum. Even at 0°K the electrons are really zipping about.

Away from 0°K, the electron states are occupied according to Fermi–Dirac statistics. The probability that a state of energy E will be occupied at temperature T is

$$f(E) = \frac{1}{\mathrm{e}^{(E-\mu)/k_B T} + 1}$$

μ is called the chemical potential of the electrons and must be chosen in such a way that the total number of electrons is N_{e}:

$$N_{\mathrm{e}} = \int_{-\infty}^{\infty} f(E)\, \mathrm{d}E$$

At 0°K, $\mu = E_{\mathrm{F}}$. The Fermi function is close to a step function at all reasonable temperatures (see fig 3, chap 6, Kittel).

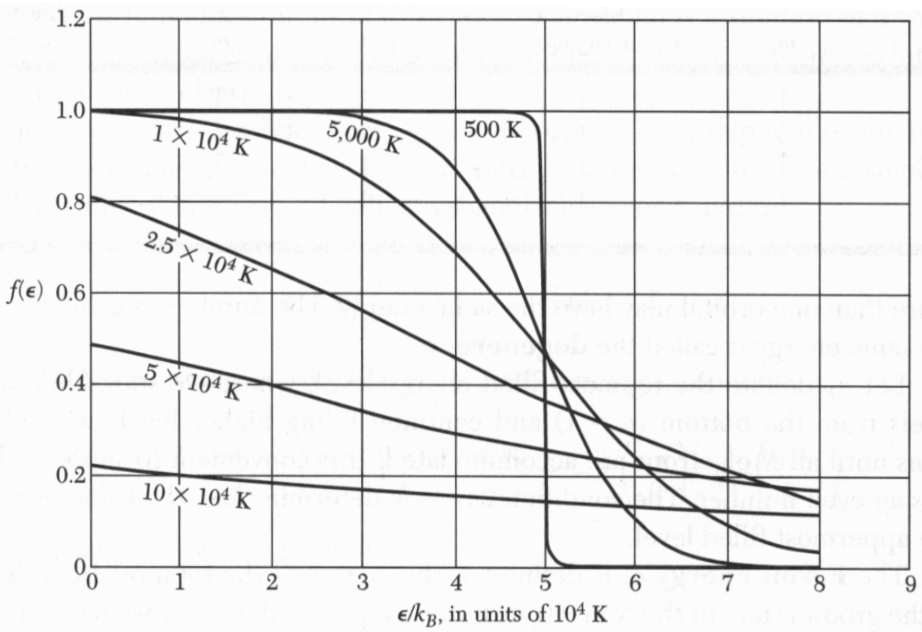


Figure 3 Fermi-Dirac distribution function (5) at the various labelled temperatures, for $T_F \equiv \epsilon_F/k_B = 50{,}000$ K. The results apply to a gas in three dimensions. The total number of particles is constant, independent of temperature. The chemical potential μ at each temperature may be read off the graph as the energy at which $f = 0.5$.

At non zero temperature some states below the Fermi level are unoccupied and some above become occupied. However this "smearing" is only over a small fraction of the energy range. If we define a Fermi temperature from

$$k_B T_F = E_F$$

the smearing is in a range of width $k_B T$ such that

$$\frac{k_B T}{E_F} = \frac{T}{T_F}$$

which is small since typical values of E_F calculated from equations (2.2.2) and (2.2.3) are as shown in this table

	valence	E_F/eV (expt.)	T/T_F at room temp.
Na	1	3.2 (2.8)	0.008
Mg	2	7.1 (7.6)	0.004
Al	3	11.6 (11.8)	0.002

whereas $k_B T$ at room temperature is only $\frac{1}{40}$eV (it's useful to remember this "rule of thumb"). The electron gas is said to be highly degenerate (*entarted* in German); only a tiny fraction (a few thousandths) of the electrons can be involved in thermal or transport processes. The rest are "frozen in" by the Pauli principle. This explains why the electron contribution to the heat capacity is small. Classically you'd expect each electron to supply an internal energy of $\frac{3}{2}k_B T$ to the crystal ($\frac{1}{2}k_B T$ of kinetic energy for each degree of freedom; the potential energy is zero) so the heat capacity should be $3R$ per mole from the phonons plus $\frac{3}{2}R$ from the electrons, which is not in accord with the observations of Dulong and Petit. But because of quantum mechanics (*viz.,* the Pauli principle) at some temperature T only a fraction of the electrons, about T/T_F of them, can be excited to above the Fermi level. So you'd expect the heat capacity due to the electrons to be roughly

$$C_\mathrm{e} \approx \frac{3}{2} N_\mathrm{e} k_\mathrm{B} \frac{T}{T_\mathrm{F}}$$

An exact calculation (see Kittel, chapter 6, or Ibach and Lüth, section 6.4) results in

$$C_\mathrm{e} = \frac{1}{3}\pi^2 k_\mathrm{B}^2 n(E_\mathrm{F}) T$$

Not surprisingly this depends on the density of states at the Fermi level, which is, using (2.2.6),

$$n(E_\mathrm{F}) = \frac{3}{2}\frac{N_\mathrm{e}}{E_\mathrm{F}}$$

Hence,

$$C_\mathrm{e} = \frac{1}{2}\pi^2 T N_\mathrm{e} \frac{k_\mathrm{B}^2}{E_\mathrm{F}}$$

In terms of $T_\mathrm{F} = E_\mathrm{F}/k_\mathrm{B}$, therefore,

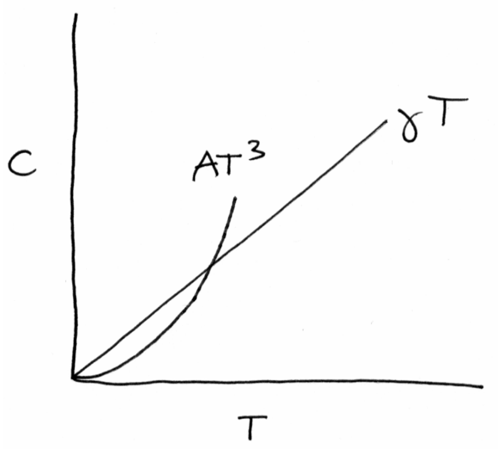
$$C_\mathrm{e} = \frac{1}{2}\pi^2 N_\mathrm{e} k_\mathrm{B} \frac{T}{T_\mathrm{F}} \tag{2.2.8}$$

So our initial estimate was out by a factor of $\pi^2/3$ but we *did* get the linear dependence on temperature.

Thus at low temperature in metals the heat capacity has the form

$$C = \gamma T + A T^3 \tag{2.2.9}$$

with the first term coming from the electrons, and the second from the phonons.



This can be confirmed from figure 9, chapter 6, Kittel in which equation (2.2.9) is shown plotted for potassium.

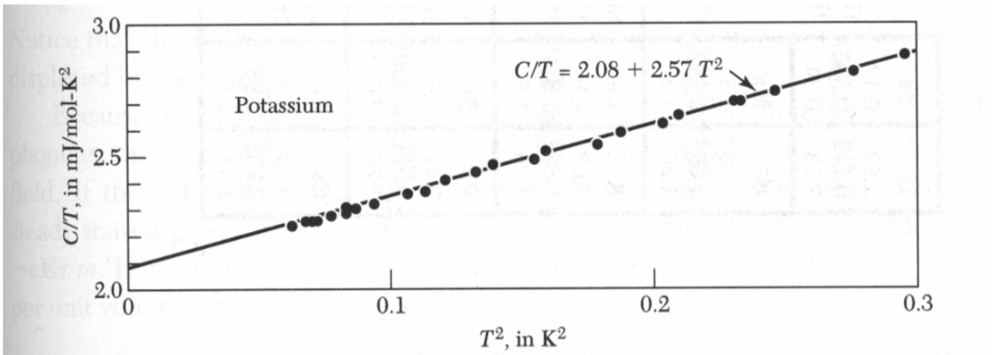


Figure 9 Experimental heat capacity values for potassium, plotted as C/T versus T^2. (After W. H. Lien and N. E. Phillips.)

From equation (2.2.2) we have

$$k_F = \left(3\pi^2 n_e\right)^{\frac{1}{3}}$$

and for a given metal we know n_e, the number of valence electrons per unit volume and hence we can predict the constant

$$\begin{aligned}\gamma &= \frac{1}{2}\pi^2 N_e \frac{k_B}{T_F}\\ &= \frac{\pi^2}{k_F^2}\frac{m}{\hbar^2}k_B^2\end{aligned}$$

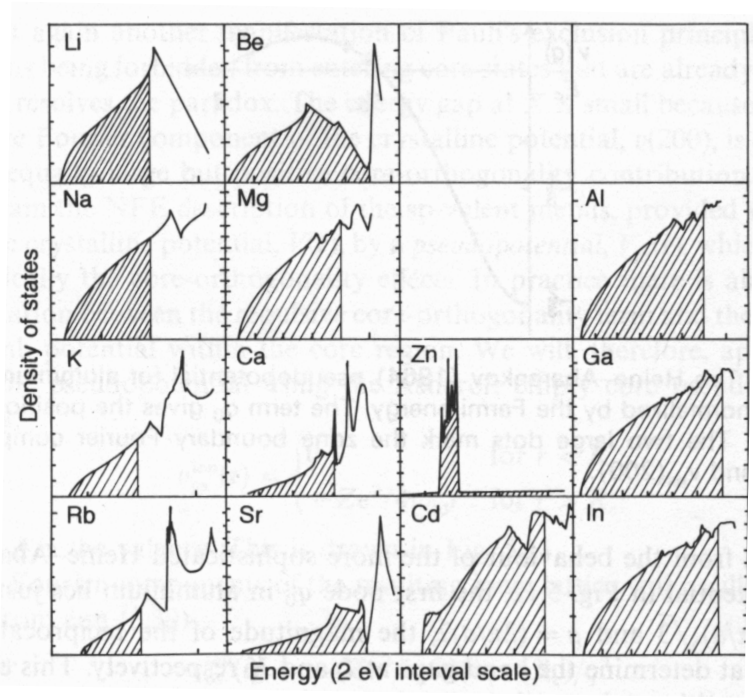
and compare it with experiment, γ_{exp}. If we write

$$\gamma_{\text{exp}} = \frac{\pi^2}{k_F^2}\frac{m^*}{\hbar^2}k_B^2$$

we may call m^* the "thermal effective mass" of the real electrons and the ratio m^*/m will tell us to what extent the electrons in a metal deviate from the free electron gas. In so called normal metals the ratio is pretty close to one.

metal	valence	m^*/m
Cu	1	1.38
Ag	1	1.0
Mg	2	1.33
Ca	2	1.9
Zn	2	0.85
Cd	2	0.75
Hg	2	1.88
Al	3	1.48
Ga	3	0.58
In	3	1.37
Sn	4	1.37
Pb	4	2

The extent to which a metal is "normal" can also be seen by comparing its actual, calculated density of states to the free electron square root form. This is illustrated below.



2.3 Effects of the crystal lattice

We replace the smeared out positive background in the Sommerfeld model with nuclear point charges at stationary lattice positions (*i.e.*, we neglect electron–phonon coupling).

Each nucleus is associated with a number of core electrons so that the deep attractive potential due to the nuclei is nearly cancelled by a repulsive potential due to the presence of the core electrons.

In *sp*-bonded metals (*e.g.*, Na, Li, Mg, Al, Ga, In) valence electrons see a weak, sometimes even repulsive, potential at the crystal lattice sites. How does this affect the free electron gas picture?

Crystal periodicity imposes two conditions upon the electron wavefunctions.

1. Bloch's theorem:

$$\boxed{\psi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\, u_{\mathbf{k}}(\mathbf{r})} \tag{2.3.1}$$

   This states that the free electron plane wave solutions are modulated by a function $u_{\mathbf{k}}(\mathbf{r})$ which has the periodicity of the crystal lattice,

$$\boxed{u_{\mathbf{k}}(\mathbf{r} + \mathbf{r}_j) = u_{\mathbf{k}}(\mathbf{r})}$$

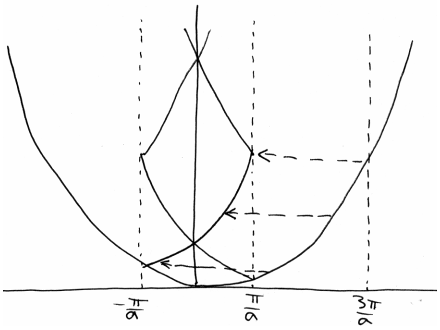
if $\mathbf{r}_j$ is a direct space lattice vector. That means $\mathbf{r}_j = \sum_i m_i \mathbf{a}_i$. We have already described Bloch's theorem indirectly in section 1.3. Essentially, quantum mechanics does not require the wavefunction to have the symmetry of the lattice but it *does* require $|\psi_\mathbf{k}(\mathbf{r})|^2$ to.

2. The wavefunction and energy eigenvalues are periodic in **k**-space,

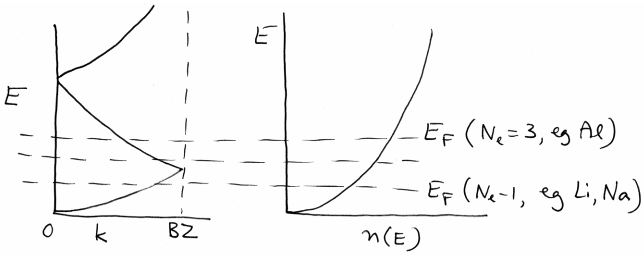
$$\boxed{\begin{aligned} \psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) &= \psi_\mathbf{k}(\mathbf{r}) \\ E(\mathbf{k}+\mathbf{g}) &= E(\mathbf{k}) \end{aligned}}$$

   This is a consquence of the periodicity of the crystal electrostatic potential and the properties of Fourier series. For a complete derivation of these results, 1 and 2, see Elliott, section 5.2.1.

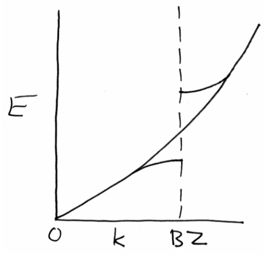
As with phonons this means that we need only deal with wavevectors that lie in the first Brillouin zone. In contrast with the phonon case, there is no maximum wavevector for electrons. We can make the following construction.



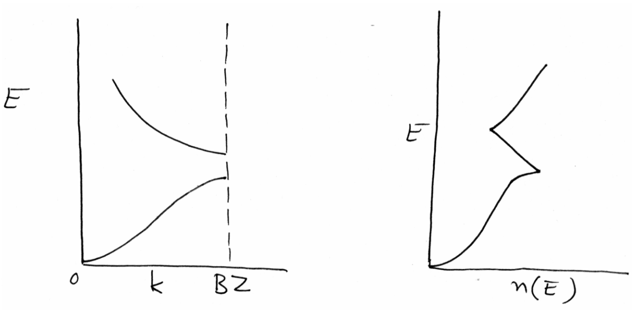
Here I have cut loose segments of the parabola that lie outside the first Brillouin zone and translated them into the first zone by the appropriate multiple of $\pm 2\pi/a$. In this reduced zone scheme we have energy bands and density of states for the free electron gas in a crystal, the so called "empty lattice."



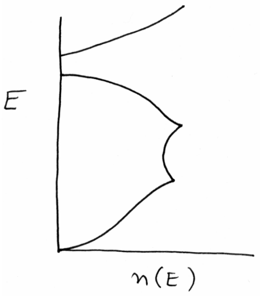
Now, what is the effect of turning on the weak crystal potential?



The electrons are Bragg reflected and form standing waves. A gap of "forbidden" energies opens up at the Brillouin zone boundary. In the reduced zone scheme, it's rendered like this:



If a gap opens up throughout the Brillouin zone then the density of states may look like this:



and we see the origin of the difference between metals, semiconductors and insulators, at least in as much as the independent electron picture is valid.

2.4 The tight binding picture

We can see how energy bands are built up, maybe more readily than in the "nearly free electron approximation," by imagining a crystal being created by bringing together atoms from infinity. Or, imagine a crystal with a huge lattice constant that we proceed to reduce to its observed value. Suppose each atom has just one valence s-electron (*e.g.*, H or Na).

We will write the electron wavefunction in the crystal as

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{N}}} \sum_i \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_i} \varphi_s(\mathbf{r} - \mathbf{r}_i) \tag{2.4.1}$$

where $\varphi_s(\mathbf{r} - \mathbf{r}_i)$ is an s-orbital centred at an atom labelled i at position vector $\mathbf{r}_i$. The phase factor guarantees that Bloch's theorem is satisfied since if $\mathbf{r}_j$ is another atomic site, then

$$\begin{aligned}\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{r}_j) &= \frac{1}{\sqrt{\mathcal{N}}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_j} \sum_{i'} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}_{i'}}\ \varphi_s(\mathbf{r} - \mathbf{r}_{i'}) \\ &= \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_j}\ \psi_{\mathbf{k}}(\mathbf{r})\end{aligned}$$

if $\mathbf{r}_{i'} = \mathbf{r}_i - \mathbf{r}_j$. If $\psi_{\mathbf{k}}(\mathbf{r})$ is an eigenstate of the Schrödinger equation

$$H\psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r})$$

then its eigenvalue will be

$$E(\mathbf{k}) = \frac{\int \bar{\psi}_{\mathbf{k}}(\mathbf{r}) H \psi_{\mathbf{k}}(\mathbf{r}) \mathrm{d}\mathbf{r}}{\int \bar{\psi}_{\mathbf{k}}(\mathbf{r}) \psi_{\mathbf{k}}(\mathbf{r}) \mathrm{d}\mathbf{r}}$$

Let us write the hamiltonian as

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$$

and approximate the electrostatic potential energy as a sum of overlapping atomic-like potentials, one at each atomic site:

$$V(\mathbf{r}) = \sum_l v(\mathbf{r} - \mathbf{r}_l)$$

Now we have

$$\begin{aligned}\int \bar{\psi}_{\mathbf{k}}(\mathbf{r}) H \psi_{\mathbf{k}}(\mathbf{r}) \mathrm{d}\mathbf{r} &= \frac{1}{\mathcal{N}} \sum_i \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \times \\ &\int \bar{\varphi}_s(\mathbf{r} - \mathbf{r}_j) \left[-\frac{\hbar^2}{2m}\nabla^2 + \sum_l v(\mathbf{r} - \mathbf{r}_l) \right] \varphi_s(\mathbf{r} - \mathbf{r}_i) \mathrm{d}\mathbf{r}\end{aligned} \tag{2.4.2}$$

and

$$\int \bar{\psi}_{\mathbf{k}}(\mathbf{r}) \psi_{\mathbf{k}}(\mathbf{r}) \mathrm{d}\mathbf{r} = \frac{1}{\mathcal{N}} \sum_i \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \int \bar{\varphi}_s(\mathbf{r} - \mathbf{r}_j) \varphi_s(\mathbf{r} - \mathbf{r}_i) \mathrm{d}\mathbf{r} \tag{2.4.3}$$

$\mathcal{N}$ is a normalisation constant. This could form the basis of a precise solution to the one-electron problem, but in the tight binding approximation we do three things.

1. Neglect *three centre integrals* in (2.4.2), *i.e.*, those for which $\mathbf{r}_i \neq \mathbf{r}_j \neq \mathbf{r}_l$.
2. Neglect *overlap integrals*, (2.4.3) except those for which $\mathbf{r}_i = \mathbf{r}_j$ (those integrals are one by virtue of the choice of normalisation).
3. We don't even attempt to calculate the remaining integrals. Instead we treat them as adjustable parameters that are non zero only between nearest neighbours. This

is exactly analogous to the model we made for lattice waves in section 1.1 (*i.e.*, we chose to model the interatomic forces with some “spring constants.”)

We then find

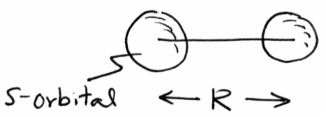
$$E(\mathbf{k}) = E_s + \underbrace{\int \bar{\varphi}_s(\mathbf{r}) \left[\sum_{\mathbf{R}\neq\mathbf{0}} v(\mathbf{r}-\mathbf{r}_i)\right] \varphi_s(\mathbf{r})\mathrm{d}\mathbf{r}}_{\text{crystal field energy}} + \sum_{\substack{i\\ \mathbf{r}_i\neq\mathbf{0}}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_i} \underbrace{\int \bar{\varphi}_s(\mathbf{r}) v(\mathbf{r}) \varphi_s(\mathbf{r}-\mathbf{r}_i)\mathrm{d}\mathbf{r}}_{\text{hopping integral, } h(R)} \tag{2.4.4}$$

Here, E_s is the free-atomic energy level because equation (2.4.2) when $\mathbf{r}_i = \mathbf{r}_j = \mathbf{r}_l$ is the eigenvalue in the isolated atom. We usually neglect the second term in equation (2.4.4) which is called the *crystal field energy*; this effectively causes a shift in the orbital energy E_s due to the potential at neighbouring atoms—it is important in *ligand field theory*, the chemistry of transition metal ions. So this all simplifies nicely to

$$E(\mathbf{k}) = E_s + \sum_{\mathbf{R}\neq\mathbf{0}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\, h(R) \tag{2.4.5}$$

where the sum is over all position vectors $\mathbf{R} = \mathbf{r}_i$ that are not zero and $h(R)$ is the transfer integral, or bond integral, or hopping integral. In this simple s-band model, we write

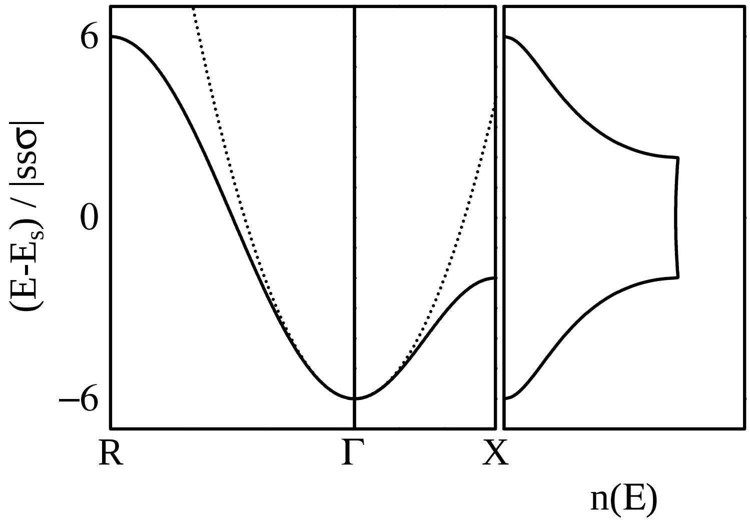
$$h(R) = h = ss\sigma$$



We can now make a bandstructure. Consider the simple cubic lattice; if only nearest neighbour bonds are allowed (again, rather like in our spring model for the lattice waves) then there are six vectors in the sum (2.4.5): $[\pm a, 0, 0]$, $[0, \pm a, 0]$, $[0, 0 \pm a]$, so

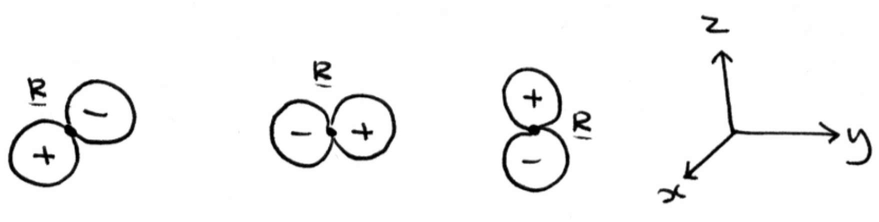
$$E(\mathbf{k}) = E_s + 2ss\sigma\left(\cos k_x a + \cos k_y a + \cos k_z a\right)$$

This function can be plotted along lines of high symmetry in the first Brillouin zone. By convention in the simple cubic lattice the reciprocal lattice points are given the symbols $\mathit{\Gamma}$ for $(0, 0, 0)$, R for $\frac{\pi}{a}(111)$ and X for $\frac{\pi}{a}(100)$.



To compare with the free electron dispersion, I have included the curve for $E_k = \hbar^2 k^2/2m$, equation (2.2.1), as a dotted line.

Now let's make bands out of p-orbitals. Instead of s-orbitals at each site we will put three p-orbitals:



p_x p_y p_z

$\varphi_x(\mathbf{r}-\mathbf{R})$ $\varphi_y(\mathbf{r}-\mathbf{R})$ $\varphi_z(\mathbf{r}-\mathbf{R})$

The expansion of the wavefunction (2.4.1) is now

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{N}}} \sum_{\alpha=x,y,z} c_\alpha \sum_{\mathbf{R}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\, \varphi_\alpha(\mathbf{r}-\mathbf{R})$$

in which c_α are three expansion coefficients to be determined. We put this into the Schrödinger equation

$$H\psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r})$$

Using the same approximations as before *viz.*, 1, 2 and 3 on p 61, the Schrödinger equation becomes a secular equation,

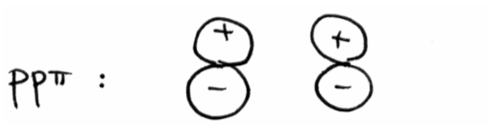
$$|(E_p - E(\mathbf{k}))\, \delta_{\alpha\alpha'} + T_{\alpha\alpha'}| = 0$$

where

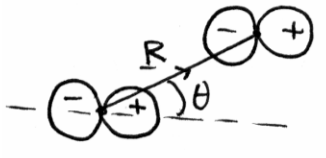
$$T_{\alpha\alpha'} = \sum_{\mathbf{R}\neq\mathbf{0}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \underbrace{\int \bar{\varphi}_\alpha(\mathbf{r}) v(\mathbf{r}) \varphi_{\alpha'}(\mathbf{r}-\mathbf{R}) \mathrm{d}\mathbf{r}}_{h(\mathbf{R})}$$

Note that α here is an index meaning x, y or z, *i.e.*, it labels one of the p-orbitals. We now have new bond integrals (compare equation (2.4.5)) which depend on the orientation of the orbitals. The two new fundamental integrals are

and



which form so called σ-bonds and π-bonds. For a general connecting vector $\mathbf{R}$ the bond integral will be some combination of these,



$$h(\mathbf{R}) = pp\sigma \cos^2\theta + pp\pi \sin^2\theta$$

If ℓ, m, n are direction cosines of $\mathbf{R}$ then

$$\int \bar{\varphi}_x v \varphi_x \mathrm{d}\mathbf{r} = \ell^2\, pp\sigma + \left(1 - \ell^2\right) pp\pi$$
$$\int \bar{\varphi}_x v \varphi_y \mathrm{d}\mathbf{r} = \ell m\, pp\sigma - \ell m\, pp\pi$$
$$\int \bar{\varphi}_x v \varphi_z \mathrm{d}\mathbf{r} = \ell n\, pp\sigma - \ell n\, pp\pi$$

and so on, by permuting x, y and z. These are called Slater–Koster transformations.

Normally the task would be to diagonalise the secular matrix, but in the simple cubic lattice $T_{\alpha\alpha'}$ is *already* diagonal,

$$T_{\alpha\alpha'} = 0 \qquad\qquad \text{for } \alpha \neq \alpha'$$

and

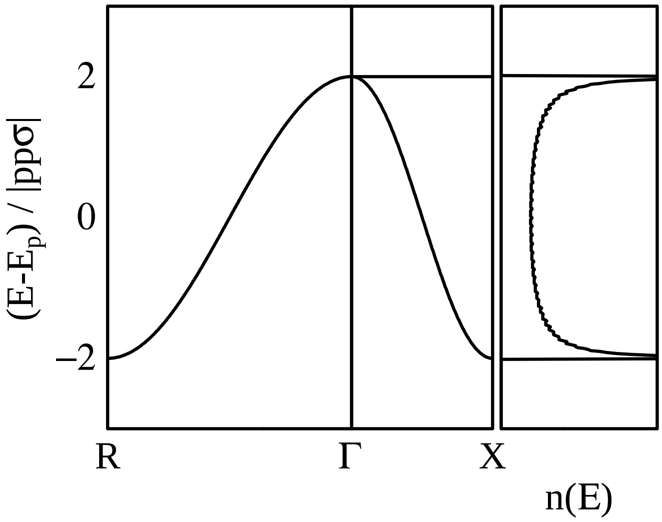
$$\begin{aligned} T_{xx} &= 2pp\sigma \cos k_x a + 2pp\pi \left(\cos k_y a + \cos k_z a\right) \\ T_{yy} &= 2pp\sigma \cos k_y a + 2pp\pi \left(\cos k_z a + \cos k_x a\right) \\ T_{zz} &= 2pp\sigma \cos k_z a + 2pp\pi \left(\cos k_x a + \cos k_y a\right) \end{aligned}$$

In the (100) direction, $\mathbf{k} = (k, 0, 0)$, we have three bands,

$$E(\mathbf{k}) = E_p + \begin{cases} 2pp\sigma\cos ka + 4pp\pi \\ 2pp\pi\cos ka + 2\left(pp\sigma + pp\pi\right) \\ 2pp\pi\cos ka + 2\left(pp\sigma + pp\pi\right) \end{cases}$$

$$\approx E_p + \begin{cases} 2pp\sigma\cos ka \\ 2pp\sigma \\ 2pp\sigma \end{cases}$$

after neglecting $pp\pi$, which I have plotted below.



Finally, we can make bands with d-orbitals. We make linear combinations of five d-Bloch sums (as in equation (2.4.1)) and get a 5×5 secular equation,

$$|D_{\alpha\alpha'} - E(\mathbf{k})\delta_{\alpha\alpha'}| = 0$$

with

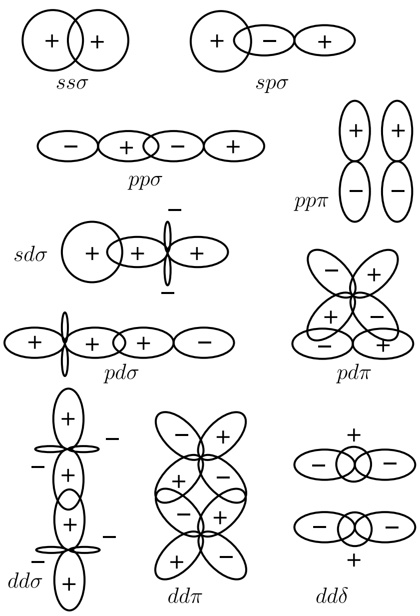
$$D_{\alpha\alpha'} = E_\alpha\delta_{\alpha\alpha'} + \sum_{\mathbf{R}\neq\mathbf{0}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \int \bar{\varphi}_\alpha(\mathbf{r})v(\mathbf{r})\varphi_\alpha(\mathbf{r}-\mathbf{R})\mathrm{d}\mathbf{r}$$

with

$$\alpha = \left(xy, yz, xz, x^2 - y^2, 3z^2 - r^2\right)$$

This reduces to three fundamental bond integrals, $dd\sigma$, $dd\pi$ and $dd\delta$.

Naturally you may wish to make bands out of s and p orbitals, or other combinations. The complete set of 10 fundamental bond integrals is illustrated below.

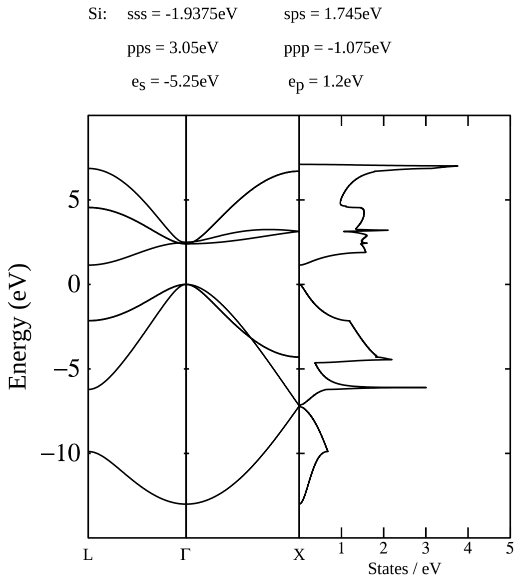


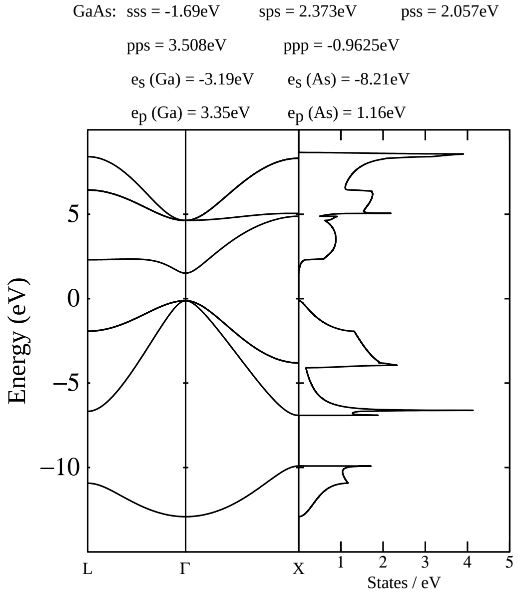
2.4.1 Case study—Si and other semiconductors

Homopolar semiconductors (*e.g.*, diamond, Si, Ge) and heteropolar semiconductors (*e.g.*, GaAs, ZnTe, SiC) occur in the diamond or zincblende crystal structures and thus have two atoms per primitive unit cell. They are *sp*-bonded compounds so in the simplest approximation we need one *s* and three *p*-orbitals on each atom, making an 8×8 secular equation involving the fundamental bond integrals

$$ss\sigma, sp\sigma(ps\sigma), pp\sigma, pp\pi$$

If suitable values are empirically chosen one can construct energy bands easily with a computer:

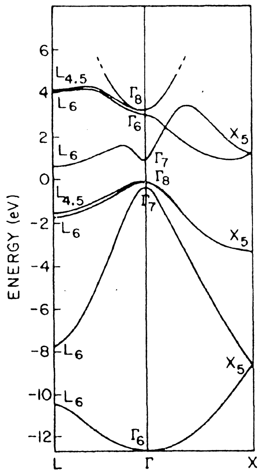


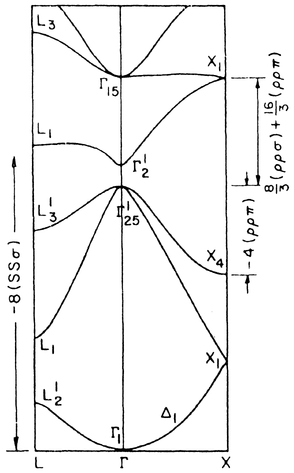


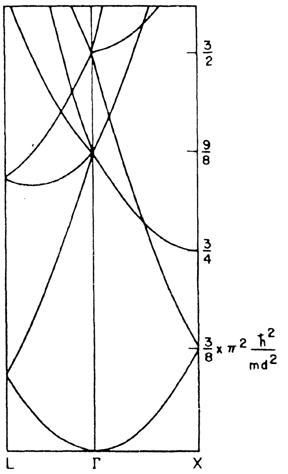
Points to note in these bandstructures are

1. There are *optical* gaps in both Si and GaAs; GaAs also has a *polar* gap.
2. At some high symmetry points in the Brillouin zone the eigenstates are purely of either s or p-character.
3. GaAs has a *direct* optical gap. In real Si the gap is indirect from Γ to a point along $\Gamma - X$ nearest to X. Accurate bands for Si using the density functional theory are shown below on p 68, where you can see the indirect gap clearly.

Actually the density functional theory in the so called "local density approximation" is not very good at getting bandstructures right. A better method (now abandoned in modern electronic structure theory) is the empirical pseudopotential. Here are bands for Ge calculated using it, as well as bands from the tight binding approximation to compare with the free electrons, or "empty lattice."



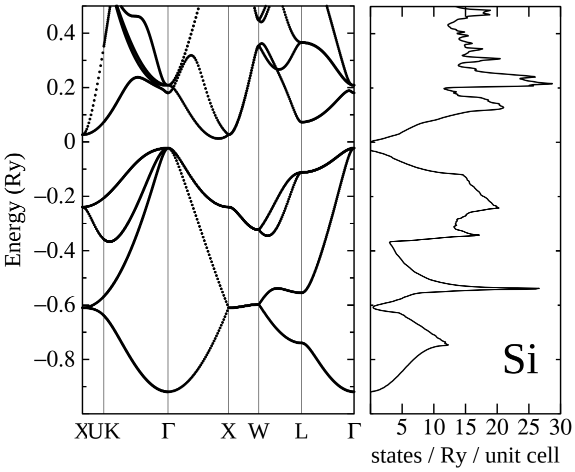




A mathematically equivalent approach in the tight binding method for semiconductors is to make Bloch sums from from linear combinations of s and p-orbitals which have a special chemical significance. Before constructing the secular matrix we can define four <u>sp^3 hybrids</u>, each one pointing along one of the four covalent bonds in the diamond cubic structure:

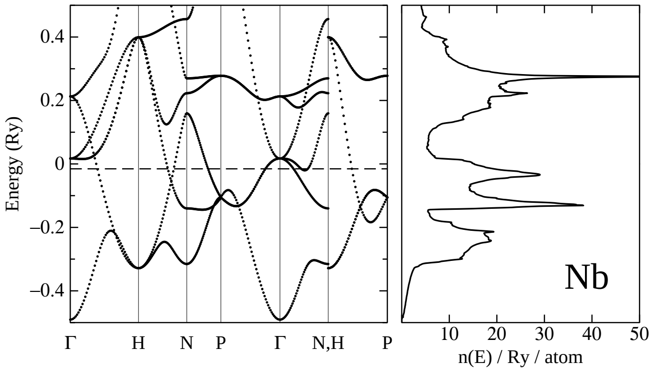
$$\begin{array}{ll}
 & \underline{\text{orientation}} \\
\phi_1 = \frac{1}{2}\left(\varphi_s + \varphi_x + \varphi_y + \varphi_z\right) & [111] \\
\phi_2 = \frac{1}{2}\left(\varphi_s + \varphi_x - \varphi_y - \varphi_z\right) & [1\bar{1}\bar{1}] \\
\phi_3 = \frac{1}{2}\left(\varphi_s - \varphi_x + \varphi_y - \varphi_z\right) & [\bar{1}1\bar{1}] \\
\phi_4 = \frac{1}{2}\left(\varphi_s - \varphi_x - \varphi_y + \varphi_z\right) & [\bar{1}\bar{1}1]
\end{array}$$

Next I'll show you a more accurate calculation of the energy bands of Si, using the "local density approximation" to "density functional theory."

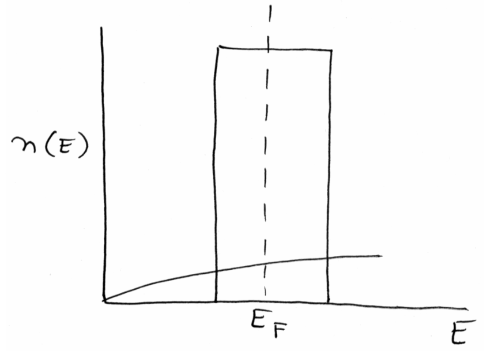


2.4.2 Case study—transition metals

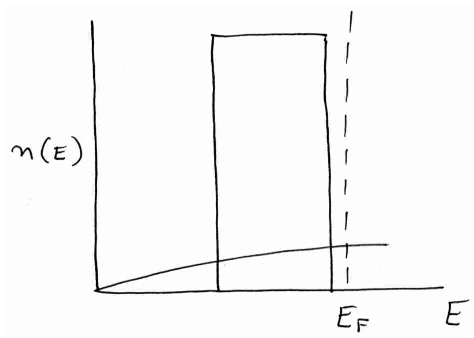
While we're looking at bandstructures, here's niobium, also calculated using the local density approximation to density functional theory.



Transition metals are, approximately, *sd*-bonded. You may think of the *s*-electrons as occupying a free electron, parabolic band. The *d*-bands are flat, or narrow, reflecting the localised nature of *d*-orbitals. A sensible approximation to the density of states is a rectangle of N_d occupied states superimposed on a free electron square root $\sqrt{E}$ density of states occupied by approximately one electron.

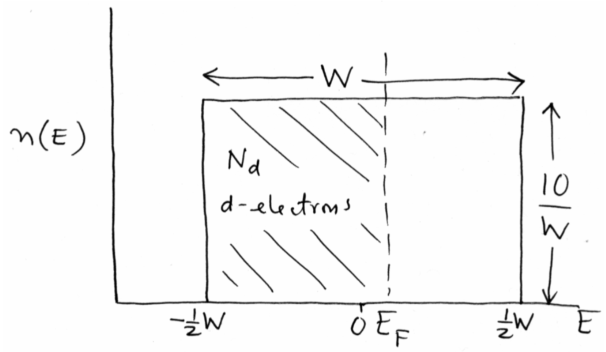


So N_s is roughly fixed and N_d increases across the transition series from one in Sc, Y and La to ten in the "noble metals" Cu, Ag and Au. Inasmuch as you can distinguish between s and d electrons these have different rôles to play in the properties of transition metals: s-electrons are mobile and are responsible for transport (thermal and electrical conductivity). d-electrons are responsible for cohesion as we'll see in the Friedel model to follow); their density of states is much greater so they dominate the electronic contribution to the heat capacity, and also optical properties. Noble metals have a full d-band.



The colours of the noble metals can be explained through the onset of interband (*i.e.*, $3d \rightarrow 4s$) absorption transitions; the onset depending on the energy between the top of the d-band and the Fermi level. In Cu and Au this is much less than in Ag, hence Cu and Au can absorb in the green and hence their reflected light contains more red and yellow.

The Friedel model is the simplest possible description of the properties of transition metals. Only d-electrons are included and they are supposed to have a density of states that is constant between the top and bottom of the band:



The width of the band is W and hence the height of the density of states is $10/W$ so that the total number of electrons is 10. A transition metal is characterised by its number of d-electrons which determines the position of the Fermi energy with respect to the centre of the band,

$$\begin{aligned} N_d &= \int_{-\frac{1}{2}W}^{E_\mathrm{F}} n(E)\,\mathrm{d}E \\ &= \int_{-\frac{1}{2}W}^{E_\mathrm{F}} \frac{10}{W}\,\mathrm{d}E \\ &= \frac{10}{W}\left(E_\mathrm{F}+\frac{1}{2}W\right) \end{aligned}$$

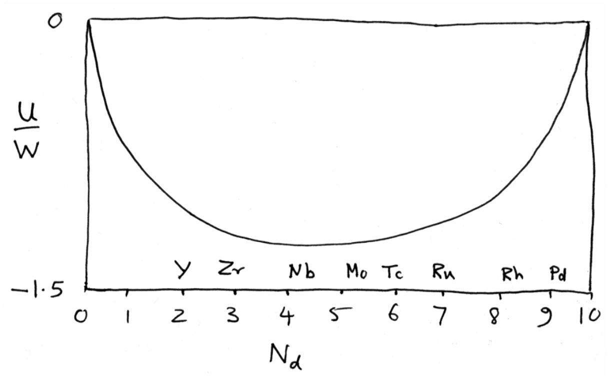
The internal energy, as usual, is

$$\begin{aligned} U &= \int_{-\frac{1}{2}W}^{E_\mathrm{F}} n(E)\,E\,\mathrm{d}E \\ &= \int_{-\frac{1}{2}W}^{E_\mathrm{F}} \frac{10}{W}E\,\mathrm{d}E \\ &= \frac{5}{W}\left(E_\mathrm{F}^2-\frac{1}{4}W^2\right) \end{aligned}$$

putting these two formulas together, we get

$$U = \frac{1}{20}W N_d\,(N_d - 10)$$

This is a parabola,



and this is exactly the trend seen in cohesive energy, sublimation energy and melting point along all three transition metal series in the periodic table.

2.5 Magnetism

According to the Bohr–van Leeuwen theorem, in *classical* mechanics a system of charged particles moving in a constant magnetic field and in thermal equilibrium has zero magnetic moment. Niels Bohr made this discovery as a PhD student in 1912. This means that all magnetic phenomena in matter—diamagnetism, paramagnetism and ferromagnetism—must have their origins in quantum mechanical behaviour. Indeed magnetism arises from the instrinic spin that is carried by elementary particles, namely electrons. This is a kind of angular momentum very loosely analogous to the angular momentum of a spinning solid sphere having some radius r. But since elementary particles are point objects they cannot be "spinning" in any classical sense. So the existence of the magnetic moment of an electron is essentially quantum mechanical in origin. What it is that motivates electrons collectively to align their spins in condensed matter has no explanation in classical physics. So that also calls for a quantum mechanical explanation, and to get to the bottom of this, it is best to start with the simplest system; or to put it another way, "let us go back to the hydrogen molecule. . . "

2.5.1 The localised picture

In the treatment of the Heitler–London H_2 molecule a negative exchange energy favours the singlet state with antiparallel spins. In fact this is a fairly general result. If we write $A_x = \frac{1}{2}\left(E_{\uparrow\downarrow} - E_{\uparrow\uparrow}\right) = \frac{1}{2}\left(E_1 - E_3\right)$, the energy levels of the H_2 molecule can be written

$$\begin{aligned} E &= \frac{1}{2}\left(E_1 + E_3\right) - \frac{1}{2}\kappa\left(E_1 - E_3\right), \qquad \kappa = \pm 1 \\ &= \frac{1}{2}\left(E_1 + E_3\right) - \frac{1}{2}\kappa 2A_x. \end{aligned} \tag{2.5.1}$$

Now, you remember from your quantum mechanics that the eigenvalue of the spin operator $\hat{\mathbf{S}}$ squared is $S^2 = S(S+1)$, in units of $\hbar$. For the electron having spin one-half, $s_i^2 = s_i(s_i+1) = 3/4$. The total spin operator for the two electron system is $\hat{\mathbf{S}} = \hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2$, and the eigenvalue of the total spin squared is,

$$S^2 = s_1^2 + s_2^2 + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = \frac{3}{2} + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = S(S+1).$$

Therefore if we write down the operator

$$\left(\frac{1}{2} + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2\right)$$

we can see immediately that it has an eigenvalue $[S(S+1) - 1]$, which is $+1$ for $S = 1$, the triplet state and -1 for $S = 0$, the singlet state. It therefore has the same properties as the number κ in equation (2.5.1) and can be substituted into that equation to give

$$\begin{aligned} E &= \frac{1}{2}\left(E_1 + E_3\right) - \frac{1}{2}\left(\frac{1}{2} + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2\right)\left(E_1 - E_3\right) \\ &= \frac{1}{2}\left(E_1 + E_3\right) - A_x\left(\frac{1}{2} + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2\right) \\ &= E_0 - 2A_x\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2, \end{aligned} \tag{2.5.2}$$

where $E_0 = (E_1 + 3E_3)/4$ is the average energy of all four spin states of the system.

Equation (2.5.2) is nothing else except a rewriting of the equation we originally got for the Heitler–London H_2 molecule:

$$E_\pm = E_{1,3} = \frac{C \pm A}{1 \pm S^2}$$

which has the singlet and triplet solutions that we found. But it exposes nicely the shifts away from the mean that occur as a result of *exchange splitting.* In fact equation (2.5.2) can be readily generalised to the many-electron case to give the "Heisenberg exchange hamiltonian:"

$$H_{\text{ex}} = E_0 - 2\sum_{i<j} A_{ij}\hat{\mathbf{s}}_i \cdot \hat{\mathbf{s}}_j .$$

Here, the electrons are labelled i, and A_{ij} are exchange coupling integrals. This model is the starting point for the treatment of "localised" magnetism, the Ising model, and so on. You can see that this is essentially a picture of spins localised on atomic sites. In fact it is built in to the Heitler–London wavefunction—equation (2.1.9)—that while electron 1 is associated with atom 1 then electron 2 is associated with atom 2, or while electron 1 is associated with atom 2 then electron 2 is associated with atom 1; that is to say a highly correlated state.

Unfortunately, the Heisenberg hamiltonian is not suitable for describing ferromagnetism in metals. In fact, whereas $A < 0$ in the Heitler–London H_2 molecule—which has the antiferromagnetic ground state—you can see that you will need $A > 0$ to stabilise the ferromagnetic state. It turns out that invariably A is *negative* when calculated between one-electron states on neighbouring atomic sites, and it is difficult to see how this picture can describe ferromagnetism.

Actually, A *is* positive when calculated between states on the same atom. The same method we used for H_2 applied to He, results in $A > 0$. Whereas in the Heitler–London H_2 molecule, $A < 0$ and the ground state energy is $C + A$ (to first order) requiring the spin function to be antisymmetric and hence a singlet; in He, $A > 0$ and the first excited state energy is $C - A$ and the wavefunction calls for the triplet spin function. This result is the basis for the well-known Hund's rule which states that in an unfilled atomic shell, the spins will align parallel, as far as possible, to maximise the benefit from the exchange energy gain. (The ground state of He is, of course, a singlet as two electrons occupy a $1s$ orbital.)

Our problem, in α-Fe for example, is: given we can understand why the spins align within the atom, why do they also couple from atomic site to site collectively to align parallel?

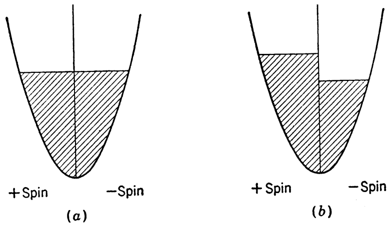
Another way to see why the antiferromagnetic state is expected between two atomic sites, is that an electron will prefer to hop to a neighbouring atom if, when it gets there, it finds the electron already there has opposite spin (then it is not repelled by the Pauli principle). So there is a greater amplitude for hopping between atomic sites whose electrons are are aligned antiparallel. This greater amplitude smears out the position of the electron and lowers its kinetic energy according to the Heisenberg uncertainty principle. Stronger hopping also leads to stronger covalent bonding as is clear from the usual tight binding or molecular orbital picture. You can expect there always to be a competition between ferromagnetism and covalency.

All these considerations have got us so far: we don't yet have a model that can describe magnetism in metals, but we have convinced ourselves that *exchange*, the quantity we have described at such length up till now, is at the bottom of magnetic phenomena. So where do we go from here to get a description of the ferromagnetic state?

2.5.2 The itinerant picture

How does the spin alignment propagate itself throughout a crystal? The mechanism that gives rise to localised spins in f-metals (rare-earths) is not overlap of the atomic f-electron states because that is very small. Instead, the picture is one in which the moment in the atomic f-shell polarises the electron gas in its neighbourhood and this propagates the exchange interaction to the next atom. It is called "indirect exchange." So here the important exchange integrals are between the localised atomic f-states and the free-electron like s and p band states. This is an appropriate description when the electrons carrying the moment are in completely localised states, but is still not what we want for transition metals where we know that the d-electrons are in approximately one-electron bands—they are "itinerant."

So let us first ask the following important question: can the free-electron gas become spontaneously spin polarised at any density? In other words, what is the ground state in the following figure, (a) or (b)?



The figure shows, back-to-back, the free-electron densities of states for plus spin ($\uparrow$) and minus spin ($\downarrow$) electrons. If we transfer $\frac{1}{2}m$ spin-down electrons into the spin-up band (*i.e.*, flip the spins of $\frac{1}{2}m$ electrons), there will be m more spin-up than spin-down electrons—a magnetic moment of $\mu_B m$, where $\mu_B = e\hbar/2m$ is the Bohr magneton. The system will gain exchange energy as a result of the lowering due to exchange, but the process will cost kinetic energy since higher energy band states need to be occupied. For the case of Jellium, both contributions can be worked out in the Hartree–Fock approximation and it turns out that only at very low densities does the exchange term win out, and the Jellium becomes ferromagnetic. The critical density is so low, however, that it is only realised in very few metals (for example Cs) but at these low densities correlation effects that have been ignored become overwhelmingly important.

The d-electrons in transition metals are neither fully localised as are f-electrons nor free-electron like as are the s and p band states. But we need to treat the d-electrons in a band, or itinerant picture. As in the previous paragraph we have to think of the energy bands as two degenerate sets, spin-up and spin-down, which can split under the action of a magnetic or exchange field. The "exchange field" is that which comes from *within* the crystal. This brings us, then, to the "Stoner model" of itinerant ferromagnetism that draws on the ideas given up to now. We write n for the number of electons *per atom*—this is between zero and ten in the d-band. As in the previous paragraph, we imagine flipping the spins of $\frac{1}{2}m$ electrons. The magnetic moment per atom (in units of μ_B) is then

$$m = n_\uparrow - n_\downarrow$$

where

$$n_\uparrow = \frac{1}{2}(n+m) \tag{2.5.3a}$$

is the number of up spins per atom and

$$n_\downarrow = \frac{1}{2}(n-m) \tag{2.5.3b}$$

the number of down spins per atom. The easiest way to calculate the change in the kinetic energy is to consider a rectangular, *i.e.*, constant, density of states $n_s(E) = n_s$ per atom *per spin* as we did in the Friedel model in section 2.4.2. (NB, we use $n_s(E)$ rather than $n(E)$ because $n_s(E)$ is the density of states per atom per spin—in the unpolarised electron gas, of course, $n_s(E) = \frac{1}{2}n(E)/N_{\text{at}}$.)

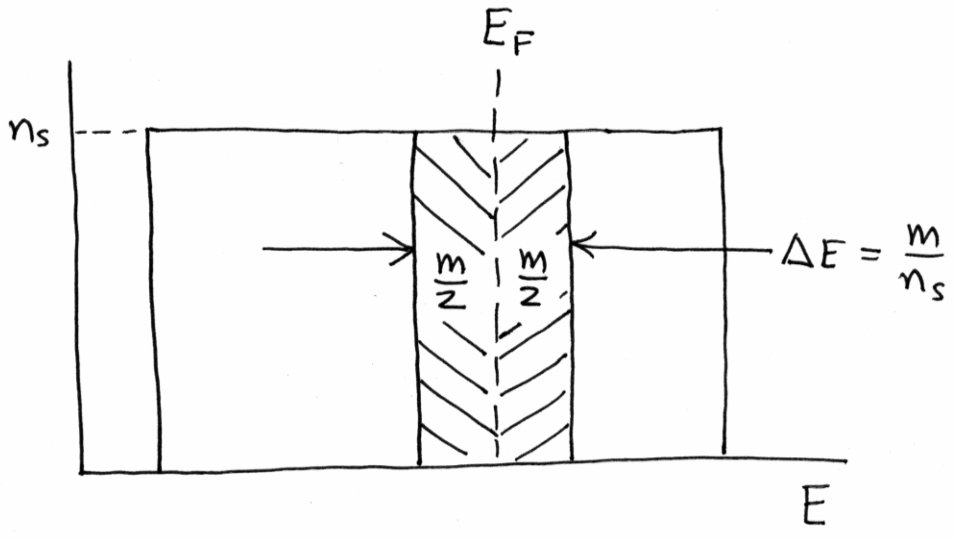
We need to calculate

$$\int^{E_\mathrm{F}-\Delta E} E n_s(E)\mathrm{d}E + \int^{E_\mathrm{F}+\Delta E} E n_s(E)\mathrm{d}E - 2\int^{E_\mathrm{F}} E n_s(E)\mathrm{d}E$$

which is simply

$$n_s(\Delta E)^2 = \frac{1}{4}\frac{m^2}{n_s(E_\mathrm{F})}$$

and this is also approximately true if $n_s(E)$ does not vary too much in the range of energy splitting $2\Delta E$ around E_F, the Fermi level.



This is an energy penalty; what about the energy gain from exchange? Here we postulate a repulsive energy between pairs of *un*like spins, $In_\uparrow n_\downarrow$ where I represents roughly the exchange energy penalty when unlike spin electrons find themselves, say, together in the d-shell of the same atom. It cannot necessarily be written as an exchange integral like A_x: as we have seen these are usually negative. Within this model, it is easy to work out the energy gain from flipping $\frac{1}{2}m$ spins: $I\frac{1}{2}(n+m)\frac{1}{2}(n-m) - I\frac{1}{2}n\frac{1}{2}n = \frac{1}{4}Im^2$. This is an energy gain, so the total (kinetic plus exchange) energy change on flipping $\frac{1}{2}m$ spins is

$$\frac{1}{4}\frac{m^2}{n_s(E_\mathrm{F})} - \frac{1}{4}Im^2 = \frac{1}{4}\frac{m^2}{n_s(E_\mathrm{F})}\left[1 - In_s(E_\mathrm{F})\right]$$

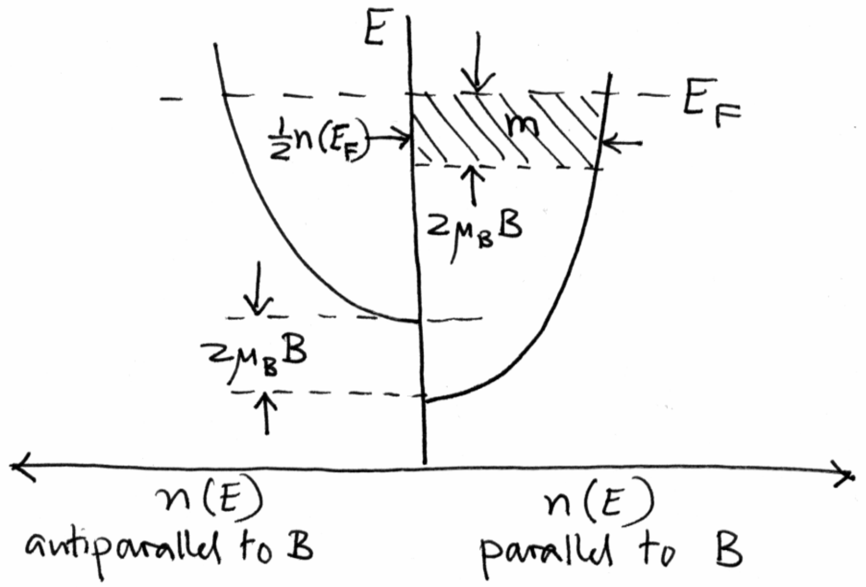
which leaves a *nett* energy *gain* if

$$\boxed{In_s(E_\mathrm{F}) > 1} \tag{2.5.4}$$

which is called the Stoner criterion for ferromagnetism. Note that the Stoner criterion calls for a large I *and* a large density of states at the Fermi level. We see here the competition between magnetism and covalency. Central bcc transition metals characteristically have the Fermi level at a *minimum* in the density of states separating occupied bonding from unoccupied antibonding states—a typical covalent picture, and non magnetic. α-Fe, on the other hand, has the Fermi level at a peak in the density of states and is stabilised in the bcc structure not by covalency but by the exchange energy—a typical picture of ferromagnetism.

2.5.3 Pauli paramagnetism

We have seen that the free electron gas is not magnetic. But what happens to it when we apply a magnetic field? Under a uniform magnetic induction of magnitude $B = \mu_0 H$ ($\mu_0 = 4\pi \times 10^{-7}\ \mathrm{JA^{-2}m^{-1}}$) an electron's energy will be raised or lowered by an amount $\pm\frac{1}{2}\mathrm{g_e}\mu_B B$ (where $\mathrm{g_e} = 2.0023\ldots$ is the electron g-factor) depending whether they are aligned parallel or antiparallel to the field. In what follows I will *always* use $\mathrm{g_e} = 2$. So the two free electron parabolas for the up and down spins are shifted up and down the energy axis so that they are displaced by an by an amount $\mathrm{g_e}\mu_B B$. Since the Fermi energy is same for both spins, there is now a difference in the numbers of spin up and spin down electrons given, approximately, by $\frac{1}{2}n(E_\mathrm{F})\mathrm{g_e}\mu_B B$. It is approximate because the density of states area shown below is not quite rectangular.



So the magnetic moment is

$$
\begin{aligned}
m &= \mu_e \mu_B B n(E_F) \\
&= \mu_B^2 B n(E_F)
\end{aligned}
$$

where we have taken the magnetic moment of the electron as $\mu_e = \frac{1}{2} g_e \mu_B \approx \mu_B$. In that case the magnetic moment per unit volume, the magnetisation, is

$$
\begin{aligned}
M &= m \frac{1}{V} \\
&= \frac{1}{V} \mu_B^2 B n(E_F)
\end{aligned}
$$

The magnetic susceptibility is, by definition, the (linear) response of the electron gas to an applied magnetic field, $H = B/\mu_0$, and we neglect the change in magnetic field due to the magnetisation of the material,

$$
\begin{aligned}
\chi_p &= \mu_0 \frac{\mathrm{d}M}{\mathrm{d}B} \\
&= \mu_0 \mu_B^2 \frac{1}{V} n(E_F) \qquad (2.5.5) \\
&= \mu_0 \mu_B^2 \frac{3}{2V} \frac{N_e}{E_F}
\end{aligned}
$$

known as the *Pauli paramagnetic susceptibility.* (We have used equation (2.2.6).) You note that this is independent of temperature, which is obvious because I have done a zero degree Kelvin calculation. But in fact the paramagnetic susceptibility in metals is observed to be independent of temperature and this is because the electron gas is degenerate—it's the same story as for the heat capacity: only electrons within $k_B T$ of the Fermi surface can flip their spins. Indeed if the gas were *not* degenerate (that is, if the Pauli principle did not apply) then at some temperature T in a magnetic induction B the probability that the electron will be found parallel to the field rather than antiparallel will be about $\mu_B B/k_\mathrm{B}T$ and as there are N_e/V electrons per unit volume each with a magnetic moment μ_B the total magnetisation will be

$$M_\text{classical} = \frac{N_\mathrm{e}}{V}\frac{\mu_B^2 B}{k_\mathrm{B}T}$$

leading to a classical estimate of the susceptibility of $N_\mathrm{e}\mu_0\mu_B^2/Vk_\mathrm{B}T$. This is the well known result for a classical paramagnetic gas.† We could use the same qualitative argument as we used for the heat capacity due to electrons and say that only a fraction T/T_F will take part in the magnetisation. Then we will get

$$\chi \approx \frac{N_\mathrm{e}}{V}\frac{\mu_0\mu_B^2}{k_\mathrm{B}T_\mathrm{F}}$$

which is temperature independent and apart from the factor 3/2 is the same as our quantum mechanical result if we recall that $E_\mathrm{F} = k_\mathrm{B}T_\mathrm{F}$.

In addition, all condensed matter has a diamagnetic response to a magnetic field due to the electrons' orbital motions and Lenz's law, and if this is added (with opposite sign) we obtain the total magnetic susceptibility of the free electron gas,

$$\chi = \mu_0\mu_B^2\frac{1}{V}n(E_\mathrm{F})\left[1 - \frac{1}{3}\left(\frac{m}{m^*}\right)^2\right]$$

where m^* is the electronic *effective mass* which we will come to in section 3. In the free electron gas, $m^* = m$ and so

$$\boxed{\begin{aligned}\chi &= \mu_0\mu_B^2\sqrt{\frac{2m}{\hbar^2}}\left(\frac{n_\mathrm{e}}{3\pi}\right)^{\frac{2}{3}}\\ &= \frac{1}{V}\mu_0\mu_B^2\frac{N_\mathrm{e}}{E_\mathrm{F}}\end{aligned}}$$

in which n_e is the number of electrons per unit volume (see equation (2.2.3)). So you notice that the magnetic susceptibility in the free electron approximation depends only on the electron density and fundamental contants.

2.5.4 The free electron band model of ferromagnetism and Stoner enhancement

We now go back and look at applying the Stoner model to the free electron gas, rather than the rectangular band model. This is quite tricky but it will expose some very

---

† Note that this is not in contradiction to the Bohr–van Leeuwen theorem which refers to a gas of moving *charged* particles. A gas of magnetic dipoles will line up with the field in thermal equilibrium.

fundamental principles in magnetism of solids. We have seen that in the Stoner model an electron prefers to be together with electrons of the same spin. The physical reason for this is that the exchange and correlation energy is lower because the two electrons more naturally remain apart because of the Pauli principle and so their electrostatic repulsion is on average smaller than the case of two otherwise identical unlike-spin electrons. Of course, this is the origin of Hund's first rule.

And so, the electrostatic potential seen by an electron is more attractive at an atomic site, or in an energy band, if that site or band *is already occupied* by other like-spin electrons. In the Stoner model we assume that the energy shift is proportional to that number of electrons, the proportionality constant being the Stoner parameter I. So the energy of an up spin electron is lowered with respect to the free electron value $E(k) = \hbar^2k^2/2m$ by an amount proportional to $n_\uparrow$, remembering that $n_\uparrow$ and $n_\downarrow$ are the number of spin up or spin down electrons *per atom*,

$$\begin{aligned} E_\uparrow(k) &= E(k) - In_\uparrow - \mu_B B \\ &= E(k) - \frac{1}{2}I(n+m) - \mu_B B \\ &= \underbrace{E(k) - \frac{1}{2}In}_{\bar{E}(k)} - \frac{1}{2}\underbrace{(Im + 2\mu_B B)}_{\Delta E} \end{aligned}$$

We have used equations (2.5.3) and in addition to the effect of the internal field on the energy levels, I have included the effect of an externally applied magnetic field, B. Strictly speaking, the shift is $-\frac{1}{2}\mathrm{g_e}\mu_B B$ but, again, I am taking the g-factor for the electron to be exactly two. The similar argument applies to the spin down electrons, so we have

$$\begin{aligned} E_\uparrow(k) &= \bar{E}(k) - \frac{1}{2}\Delta E \\ E_\downarrow(k) &= \bar{E}(k) + \frac{1}{2}\Delta E \end{aligned}$$

with

$$\Delta E = Im + 2\mu_B B \tag{2.5.6}$$

Now the states at temperature T are occupied in accordance with the Fermi–Dirac statistics,

$$f(E) = \frac{1}{\mathrm{e}^{(E-E_\mathrm{F})/k_\mathrm{B}T} + 1}$$

and so we can write the magnetic moment *per atom* in units of μ_B as,

$$\begin{aligned} m = n_\uparrow - n_\downarrow &= \frac{1}{N_\mathrm{at}}\sum_k \left[f_\uparrow(E(k)) - f_\downarrow(E(k))\right] \\ &= \frac{1}{N_\mathrm{at}}\sum_k \left[\frac{1}{\mathrm{e}^{(\bar{E}(k)-\frac{1}{2}\Delta E - E_\mathrm{F}))/k_\mathrm{B}T} + 1} - \frac{1}{\mathrm{e}^{(\bar{E}(k)+\frac{1}{2}\Delta E - E_\mathrm{F}))/k_\mathrm{B}T} + 1}\right] \end{aligned}$$

This is an example of a *self consistent* problem. The moment depends on the shifts in the energy levels; but those shifts depend on the moments! In this rare case the self consistent problem, as exposed by m appearing on both sides of the above equation, can be solved in closed form.

Now we can see by making a Taylor expansion of the Fermi–Dirac function to third order that

$$f(E - \frac{1}{2}\Delta E - E_{\mathrm{F}})) - f(E + \frac{1}{2}\Delta E - E_{\mathrm{F}})) = -\Delta E f'(E) - \frac{1}{3}\left(\frac{1}{2}\Delta E\right)^3 f'''(E)$$

Using this we now find, in the case $B = 0$,

$$m = -\frac{1}{N_{\mathrm{at}}} Im \sum_k \underbrace{\frac{\mathrm{d}f}{\mathrm{d}E}}_{<0} - \frac{1}{24} I^3 m^2 \underbrace{\sum_k \frac{\mathrm{d}^3 f}{\mathrm{d}E^3}}_{>0} \qquad (2.5.7)$$

We divide through by m and we note that the last term has got to be greater than zero. Therefore in order for there to be a non zero magnetic moment *in the absence of a magnetic field* we must have

$$-1 - \frac{I}{N_{\mathrm{at}}} \sum_k f' > 0$$

I will tell you without proof the that at $0°K$ (Ibach and Lüth, section 8.4)

$$\frac{1}{N_{\mathrm{at}}} \sum_k f' = -n_s(E_{\mathrm{F}})$$

($n_s(E)$, not $n(E)$, because the sum is only over one set of spins). So we come back again to the Stoner criterion (2.5.4)

$$In_s(E_{\mathrm{F}}) > 1$$

for there to be a spontaneous spin polarisation of the electron gas.

However, the density of states of the free electron gas is not large enough to meet this criterion, and the so called "normal metals" are not ferromagnetic. However we can pursue this argument to find their magnetic susceptibility. First we put back the applied magnetic field which splits the energy levels according to (2.5.6) and we neglect the third order term in (2.5.7) to get the magnetic moment, in units of μ_B, of the free electron gas as a function of applied magnetic field,

$$m = (Im + 2\mu_B B) n_s(E_{\mathrm{F}})$$

which leads to

$$m = \mu_B B \frac{2n_s(E_{\mathrm{F}})}{1 - In_s(E_{\mathrm{F}})}$$

The total magnetic moment per unit volume, the magnetisation, M, is

$$\begin{aligned} M &= \frac{1}{V} N_{\mathrm{at}} m \mu_B \\ &= \frac{N_{\mathrm{at}}}{V} \mu_B^2 B \frac{2n_s(E_{\mathrm{F}})}{1 - In_s(E_{\mathrm{F}})} \end{aligned}$$

and so the susceptibility, χ, is

$$\begin{aligned}\chi &= \mu_0 \frac{\mathrm{d}M}{\mathrm{d}B} \\ &= \frac{N_{\mathrm{at}}}{V} \mu_0 \mu_B^2 \frac{2n_s(E_{\mathrm{F}})}{1 - In_s(E_{\mathrm{F}})}\end{aligned}$$

$$\boxed{\chi = \frac{\chi_p}{1 - In_s(E_{\mathrm{F}})}} \qquad (2.5.8)$$

Here I have used $2n_s(E) = n(E)/N_{\mathrm{at}}$; and our earlier result (2.5.5) for the Pauli susceptibilty.

So the result of employing the Stoner parameter, which represents atomic electron–electron interactions is two fold.

1. We can predict <u>ferromagnetism</u> in metals with both a large value of I and a large density of states at the Fermi level. Only three pure metals meet the Stoner criterion, namely Co, Fe and Ni—all transition metals with a large d-electron derived density of states.
2. In the normal, free electron like, metals the electron–electron interaction causes an <u>enhancement</u> in the one electron approximation Pauli susceptibility: from (2.5.8) $\chi > \chi_p$.
3. In addition, you note from (2.5.8) that as $In_s(E_{\mathrm{F}})$ increases towards the value one, the denominator vanishes and the susceptibilty diverges to infinity; this is a sign of the phase transition to ferromagnetism—the magnetisation is huge, independent of any applied magnetic field.

3. Thermal and electrical conductivity—transport

In sections 1 and 2, we examined the stationary states of the Schrödinger equation for phonons and electrons respectively. Here, in section 3 we study transport; that is, the flow of electrons and phonons under the effects of uniform electric fields and temperature gradients.

3.1 Scattering

We will see that ultimately electrical and thermal conductivity are limited by the scattering of carriers. In this section we will only consider electron scattering. We'll come to phonons later. In general, an electron in an occupied Bloch state (or strictly wavepacket) of wavevector $\mathbf{k}$ is scattered into an empty state of wavevector $\mathbf{k}'$. What matters is the *probability per unit time* that this event happens. This is extraordinarily difficult either to measure or to calculate, but formally it's given by Fermi's Golden Rule (actually due to Dirac)

$$w_{\mathbf{k}\to\mathbf{k}'} = \frac{2\pi}{\hbar} n_{\mathbf{k}'} \left| \int \bar{\psi}_{\mathbf{k}'}(\mathbf{r})\, U(\mathbf{r})\, \psi_{\mathbf{k}}(\mathbf{r})\, \mathrm{d}\mathbf{r} \right|^2 \tag{3.1.1}$$

Here $U(\mathbf{r})$ is the "scattering potential energy"—it is that term in the hamiltonian that couples the states $\psi_{\mathbf{k}}(\mathbf{r})$ and $\psi_{\mathbf{k}'}(\mathbf{r})$; $n_{\mathbf{k}'}$ is the "density of final states."

Also, formally, you may note that the cross section for this event is

$$\sigma = \frac{1}{I_0}\, w_{\mathbf{k}\to\mathbf{k}'}$$

where I_0 is the flux of initial states.

The only thing we may need to know from this is that

$$w_{\mathbf{k}\to\mathbf{k}'} = w_{\mathbf{k}'\to\mathbf{k}} \tag{3.1.2}$$

which follows from its definition (3.1.1). This is a kind of "microscopic reversibility."

We can separate scattering events into elastic, where the electron suffers no change in energy eigenvalue, and inelastic, when energy and momentum are delivered to the scatter.

Examples of electron scatterers are

1. Lattice defects, *e.g.*, impurities; point, line and planar defects

   point—vacancies, interstitials

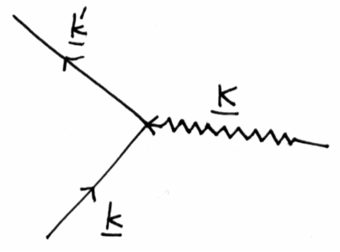
   line—dislocations

   planar—grain boundaries, stacking faults, hetero-interfaces, surfaces

Defects scatter elastically, *i.e.*,

$$E(\mathbf{k}') = E(\mathbf{k})$$

2. Phonons: the vibrating atoms make the scattering potential time dependent and the scattering is inelastic. We have already studied these processes in section 1.7. In a process such as



the matrix element is proportional to

$$\int \underbrace{\bar{u}_{\mathbf{k}'}(\mathbf{r})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}}_{\substack{\text{final}\\ \text{Bloch state}}}\;\underbrace{u_{\mathbf{k}}(\mathbf{r})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}}_{\substack{\text{initial}\\ \text{Bloch state}}}\;\overbrace{U_0\,\mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}}^{U(\mathbf{r})}\,\mathrm{d}\mathbf{r}$$

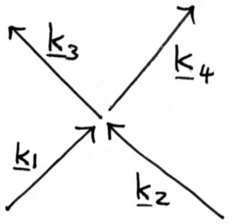
$$= U_0 \int \bar{u}_{\mathbf{k}'}(\mathbf{r})\,u_{\mathbf{k}}(\mathbf{r})\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\,\mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}\,\mathrm{d}\mathbf{r}$$

where $\mathbf{q}$ is the scattering vector $\mathbf{k} - \mathbf{k}'$, and $\mathbf{K}$ is the phonon's wavevector. Because the $u_{\mathbf{k}}(\mathbf{r})$ are periodic (Bloch's theorem) they can be expanded in a Fourier series and the integral vanishes by the property of Fourier integrals unless

$$\mathbf{K} + \mathbf{q} = \mathbf{g}$$

a reciprocal lattice vector, which is the conservation law of crystal momentum. If $\mathbf{g} \neq \mathbf{0}$ this is an Umklapp process.

3. Other electrons. (See Ibach and Lüth, fig 9.3 and accompanying text). In an event like this,



The incoming electron in state $\mathbf{k}_1$ is above, but within $k_{\mathrm{B}}T$ of, the Fermi level and suppose it scatters an electron below E_{F}. Then by energy conservation,

$$E(\mathbf{k}_1) + E(\mathbf{k}_2) = E(\mathbf{k}_3) + E(\mathbf{k}_4)$$

or, in shorthand,

$$E_1 + E_2 = E_3 + E_4 \tag{3.1.3}$$

with

$$E_1 > E_\mathrm{F} \qquad E_3 > E_\mathrm{F} \qquad E_4 > E_\mathrm{F} \qquad E_2 < E_\mathrm{F}$$

Then from (3.1.3)

$$E_1 + E_2 > 2E_\mathrm{F}$$

and rearranging this,

$$|E_2 - E_\mathrm{F}| < E_1 - E_\mathrm{F}$$

That means electron 2 is yet closer to the Fermi surface than is electron 1— the incoming one. Again, from (3.1.3)

$$\begin{aligned}\overbrace{E_3 - E_\mathrm{F}}^{>0} + \overbrace{E_4 - E_\mathrm{F}}^{>0} &= \overbrace{E_1 - E_\mathrm{F}}^{>0} + \overbrace{E_2 - E_\mathrm{F}}^{<0} \\ &< E_1 - E_\mathrm{F} + |E_2 - E_\mathrm{F}|\end{aligned}$$

which implies

$$E_3 - E_\mathrm{F} < E_1 - E_\mathrm{F} + |E_2 - E_\mathrm{F}|$$

and

$$E_4 - E_\mathrm{F} < E_1 - E_\mathrm{F} + |E_2 - E_\mathrm{F}|$$

so electrons 3 and 4 are also within $\sim k_\mathrm{B}T$ of the Fermi surface. So how does the scattering rate depend on temperature? We take it that electron 1—the incoming one—is thermally activated (*i.e.*, not a fast, "hot" electron) so it's within $k_\mathrm{B}T$ of E_F. So electron 2 must be chosen from those states within an energy width $k_\mathrm{B}T$ below the Fermi surface; this multiplies the rate by a factor T/T_F. But the choice of electron 3 must also be from states within $k_\mathrm{B}T$ of E_F, introducing another factor T/T_F. The energy and momentum of electron 4 are now fully determined, so no other factor of T/T_F arises and the scattering rate in the degenerate electron gas is reduced by the Pauli principle by a factor

$$\left(\frac{T}{T_\mathrm{F}}\right)^2 \sim 10^{-6}$$

It is therefore very hard to detect electron–electron scattering—it's a *rare event* and we do not expect it to limit transport processes significantly in metals.

3.2 Electrical conduction in semiconductors

We all know that a semiconductor is an insulator with a small enough band gap that at ambient or slightly elevated temperatures electron–hole pairs are generated that act as carriers. In an intrinsic semiconductor, we have

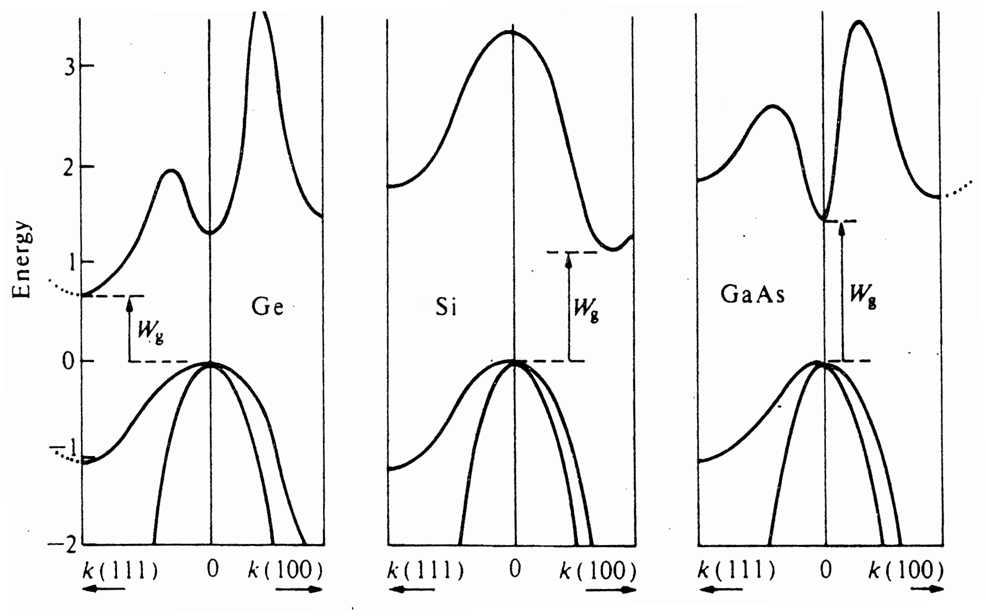
$$n_\mathrm{h} = n_\mathrm{e}$$

which is the number of carriers *per unit volume.* We also know that extrinsic doping can introduce additional carriers and can pin the Fermi level at either donor or acceptor levels. In an intrinsic semiconductor, the thermodynamic chemical potential of electrons, or "Fermi level" is at mid gap. The carriers in semiconductors are non degenerate, they exist in an energy range around $k_{\mathrm{B}}T$ of the band edges and so are not subject to Pauli principle restrictions in their excitations.

Note that the intrinsic carrier concentration in Si at 300°K is

$$n_i = n_{\mathrm{h}} = n_{\mathrm{e}} = 1.6 \times 10^{10}\mathrm{cm}^{-3}$$

while the atomic density is 5×10^{22} atoms per cm^3, so only one in about 10^9 atoms is "ionised." In heavily doped semiconductors, the carrier concentration is about 10^{18} cm^{-3}.



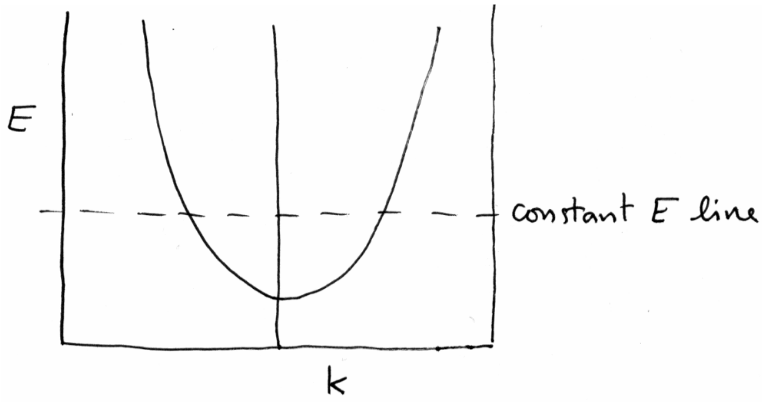
The above figure shows that semiconductor bands can be understood close to the band edges as free electron parabolas,

$$E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m^*} \tag{3.2.1}$$

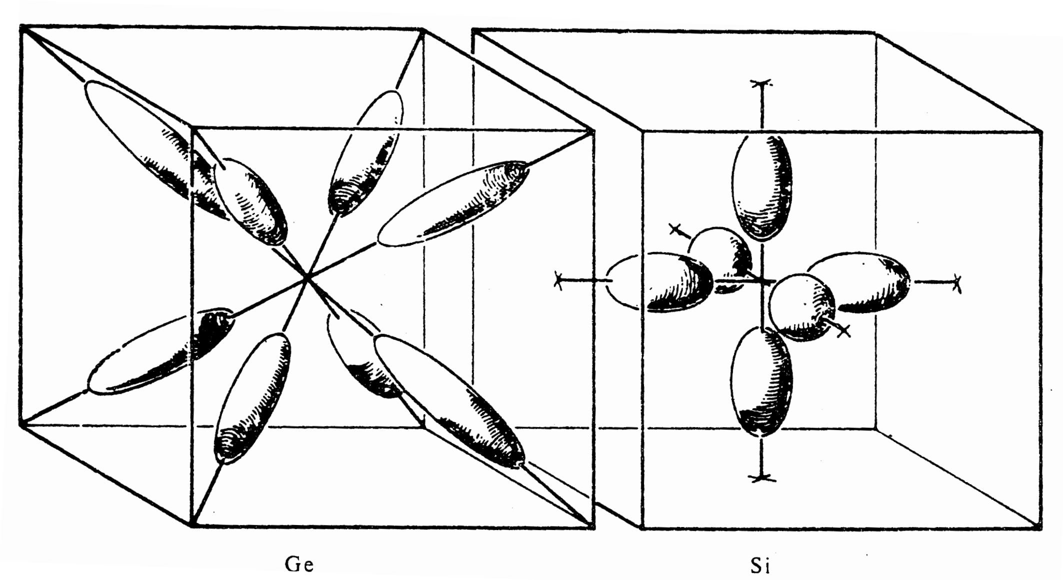
where the effective mass, m^*, accounts for the variation in curvature, with respect to the free electron bands. We will use m_{e} and m_{h} for electron and hole effective masses. Typical values are

	Si		Ge		GaAs	
m_{e}/m	0.98		1.64		0.07	
m_{h}/m	0.45	(heavy)	0.28	(heavy)	0.45	(heavy)
	0.16	(light)	0.04	(light)	0.08	(light)

We cannot plot bands in three dimensions, but we can draw constant energy surfaces. In one dimension a certain energy will intersect a band at a *point.*



In two dimensions you would get a line. If a band was parabolic and isotropic—*i.e.*, m^* is the same for all directions in the Brillouin zone—then a constant energy surface would be a sphere, for instance the *Fermi sphere* if the energy in question is E_F. Below I show constant energy surfaces for electron carriers just above the conduction band edge; they are *ellipsoids* because the effective mass is not the same in all directions of $\mathbf{k}$. However the symmetry of the crystal is observed. You can tell from the figure on p 84 as well as from the figure below, which shows constant energy surfaces in reciprocal space near the conduction band minima, that both Ge and Si are *indirect gap* semiconductors. The valence band edge is at $\mathbf{k} = (000)$ in both but the conduction band edge is at $\pi/a(111)$ ("the L-point") in Ge and near $2\pi/a(100)$ ("the X-point") in Si.



For an isotropic band, you can see from equation (3.2.1) that

$$m^* = \hbar^2 \left(\frac{\mathrm{d}^2 E}{\mathrm{d}k^2} \right)^{-1}$$

that is, inversely proportional to the *curvature* of the band. Hence "heavy" holes occupy the flatter of the two valence bands. If the band is not isotropic then the effective mass is a matrix m^*_{ij} and the reciprocal effective mass tensor is

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\mathrm{d}^2 E}{\mathrm{d}k_i \mathrm{d}k_j}$$

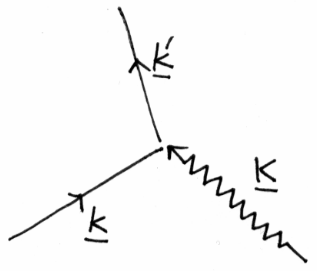
where $\mathbf{k} = (k_1, k_2, k_3)$. Such a 3×3 symmetric tensor can always be diagonalised by the appropriate choice of principal axes, so in general there are three effective masses,

$$\left(\frac{1}{m^*}\right)_{ij} = \begin{pmatrix} \frac{1}{m^*_x} & 0 & 0 \\ 0 & \frac{1}{m^*_y} & 0 \\ 0 & 0 & \frac{1}{m^*_z} \end{pmatrix}$$

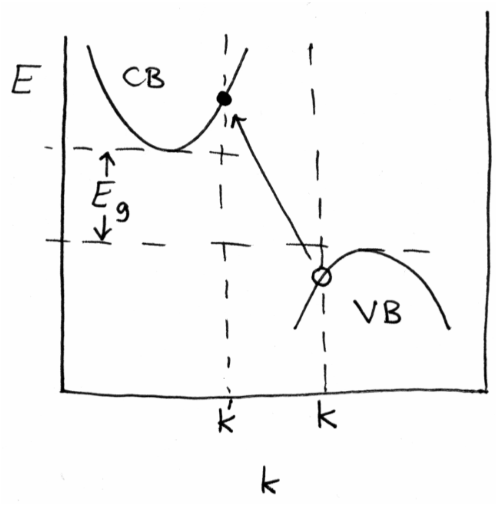
For simplicity we will treat semiconductor bands as isotropic.

3.2.1 Fermi level and number of carriers in a semiconductor

At finite temperature we expect to find electron–hole pair generation like this:



with $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ and $\mathbf{K} + \mathbf{q} = \mathbf{g}$.



We need to use the Fermi–Dirac distribution function $f(E)$ to calculate the density of carriers in each band. The number of electrons per unit volume in the conduction band will be the integral of the density of states per unit volume (see section 2.2) times the occupation probability,

$$n_{\mathrm{e}} = \int \frac{1}{V}\, n(E) f(E) \mathrm{d}E$$

If we assume that the conduction band is parabolic then we need the free electron gas density of states (2.2.5), per unit volume, measured from the bottom of the band,

$$\frac{1}{V}\, n(E) = \left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}} \frac{1}{2\pi^2} (E - E_c)^{\frac{1}{2}}$$

with m_{e} the effective mass in the conduction band and the energy taken relative to the bottom of the conduction band E_c. At normal temperatures, and a few tenths of an electron volt (eV) above the Fermi level, the Fermi–Dirac distribution is indistinguishable numerically from the simpler classical Maxwell–Boltzmann distribution,

$$\frac{1}{\mathrm{e}^{(E-\mu)/k_{\mathrm{B}}T} + 1} \longrightarrow \mathrm{e}^{-(E-\mu)/k_{\mathrm{B}}T}$$

so we have

$$\begin{aligned} n_{\mathrm{e}} &= \int \frac{1}{V}\, n(E) f(E) \mathrm{d}E \\ &= \int_{E_c}^{\infty} \left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}} \frac{1}{2\pi^2} (E - E_c)^{\frac{1}{2}}\, \mathrm{e}^{-(E-\mu)/k_{\mathrm{B}}T} \mathrm{d}E \\ &= \frac{1}{2\pi^2\hbar^3} (2m_{\mathrm{e}})^{\frac{3}{2}}\, \mathrm{e}^{\mu/k_{\mathrm{B}}T} \int_{E_c}^{\infty} (E - E_c)^{\frac{1}{2}}\, \mathrm{e}^{-E/k_{\mathrm{B}}T} \mathrm{d}E \end{aligned}$$

This integral can be done[†] and we get

$$n_{\mathrm{e}} = \frac{2}{h^3} (2\pi m_{\mathrm{e}} k_{\mathrm{B}} T)^{\frac{3}{2}}\, \mathrm{e}^{-(E_c-\mu)/k_{\mathrm{B}}T}$$

electrons per unit volume. Here, $h = 2\pi\hbar$ is Planck's original constant. It is interesting that it appears in this way more often in statistical mechanics than in *quantum mechanics.* The density of holes in the valence band is found in exactly the same way by integrating the density of states downwards from the top of the valence band,

$$n_{\mathrm{h}} = \frac{2}{h^3} (2\pi m_{\mathrm{h}} k_{\mathrm{B}} T)^{\frac{3}{2}}\, \mathrm{e}^{-(\mu-E_v)/k_{\mathrm{B}}T}$$

[†] By substitution, $u = E - E_c$, $\mathrm{d}u = \mathrm{d}E$, choose $E_c = 0$;

$$\begin{aligned} I &\equiv \int_{E_c}^{\infty} (E - E_c)^{\frac{1}{2}}\, \mathrm{e}^{-(E-\mu)/k_{\mathrm{B}}T} \mathrm{d}E \\ &= \int_{0}^{\infty} u^{\frac{1}{2}} \mathrm{e}^{-(u+E_c)/k_{\mathrm{B}}T} \mathrm{d}u \\ &= \mathrm{e}^{-E_c/k_{\mathrm{B}}T} \int_{0}^{\infty} u^{\frac{1}{2}} \mathrm{e}^{-u/k_{\mathrm{B}}T} \mathrm{d}u \\ &= \mathrm{e}^{-E_c/k_{\mathrm{B}}T} \frac{1}{2}\sqrt{\pi} (k_{\mathrm{B}}T)^{\frac{3}{2}} \end{aligned}$$

(Gradshteyn and Ryzhik, art. 3.371)

holes per unit volume, where m_{h} is the valence band effective mass and E_v is the top of the valence band. We can define

$$A_c = \frac{2}{h^3}\left(2\pi m_{\mathrm{e}} k_{\mathrm{B}} T\right)^{\frac{3}{2}}$$
$$A_v = \frac{2}{h^3}\left(2\pi m_{\mathrm{h}} k_{\mathrm{B}} T\right)^{\frac{3}{2}}$$

as effective conduction and valence band densities of states. Unlike the actual density of states these are temperature dependent, and may be thought of as the density of states if all the carriers were located in energy at the band edges. Then simply,

$$n_{\mathrm{e}} = A_c \mathrm{e}^{-(E_c-\mu)/k_{\mathrm{B}}T}$$
$$n_{\mathrm{h}} = A_v \mathrm{e}^{(E_v-\mu)/k_{\mathrm{B}}T}$$

and by multiplying,

$$n_{\mathrm{e}} n_{\mathrm{h}} = A_c A_v \mathrm{e}^{-E_g/k_{\mathrm{B}}T}$$

where $E_g = E_c - E_v$ is the band gap energy. Note that the product $n_{\mathrm{e}} n_{\mathrm{h}}$ depends only on temperature, effective masses and the band gap energy so must be the same for all doping levels and dopant types (n-, or p-type) because it doesn't depend on the Fermi level. We conclude that the product $n_{\mathrm{e}} n_{\mathrm{h}}$ is constant for a given semiconductor and a given temperature independent of the dopant concentrations or the Fermi level. We have seen that in an intrinsic semiconductor,

$$n_{\mathrm{e}} = n_{\mathrm{h}} = n_i$$

hence

$$n_{\mathrm{e}} n_{\mathrm{h}} = n_i^2 \qquad (3.2.2)$$

is true for all semiconductors, doped or not. If we know the donor and acceptor concentrations, N_D and N_A and assume they are all ionised, then charge neutrality requires

$$n_{\mathrm{e}} + N_A = n_{\mathrm{h}} + N_D \qquad (3.2.3)$$

and (3.2.2) and (3.2.3) are sufficient to compute the carrier concentration.

We might also take the ratio,

$$\frac{n_{\mathrm{h}}}{n_{\mathrm{e}}} = \frac{A_v}{A_c}\,\mathrm{e}^{(E_c+E_v-2\mu)/k_{\mathrm{B}}T}$$

from which the chemical potential or "Fermi level" is

$$\mu = \frac{1}{2}\left(E_v + E_c\right) - \frac{1}{2}k_{\mathrm{B}}T\ln\frac{n_{\mathrm{h}}}{n_{\mathrm{e}}} - \frac{3}{4}k_{\mathrm{B}}T\ln\frac{m_{\mathrm{e}}}{m_{\mathrm{h}}}$$

In intrinsic semiconductor the "Fermi level" is mid gap to within a small correction due the ratios of the effective masses. If the dopants are all donors (n-type) then we have, if they are all ionised,

$$n_{\mathrm{e}} = N_D$$

hence

$$N_D = A_c \mathrm{e}^{-(E_c-\mu)/k_{\mathrm{B}}T}$$

so that

$$E_c - \mu = k_{\mathrm{B}}T \ln \frac{A_c}{N_D} \approx 0$$

In Si at room temperature, $N_D = 10^{19}\ \mathrm{cm}^{-3}$ and $A = 2.48 \times 10^{19} \left(\frac{m^*}{m}\right)^{\frac{3}{2}}\ \mathrm{cm}^{-3}$, so that the "Fermi level" μ is pinned at the conduction band edge. Similarly in p-type semiconductor

$$\mu - E_v = k_{\mathrm{B}}T \ln \frac{A_v}{N_A} \approx 0$$

and the "Fermi level" is pinned just above the valence band edge.

3.2.2 Conductivity of a semiconductor

Later, we will consider the velocity of an electron in the absence of scattering, and we call this the ballistic velocity. Elastic collisions serve to randomise the direction of travel so the average velocity in equilibrium is zero. If an electric field is applied then the electron will acquire a small non zero average amount we will call the drift velocity. We assume each collision completely randomises the velocity, but in between collisions the electron is accelerated in the direction of the field. We are interested in separating out that part of the motion, due to the field, that is superimposed upon the motion in zero field. Suppose a collision occurs at $t = 0$, then after a further time t the electron will have travelled a distance

$$d = \frac{1}{2}at^2$$

in the direction of the field, where a is the acceleration. The average drift velocity is obtained by taking the average value of d and dividing by the average value of the time t between scattering events,

$$v_d = \frac{\langle d\rangle}{\langle t\rangle} = \frac{1}{2}a\frac{\langle t^2\rangle}{\langle t\rangle}$$

Later, in section 3.4.2, we will discuss how to calculate these averages under the assumption of a "Poisson distribution" of random collisions. We will say that if a particle is chosen at random at $t = 0$ then the probability density (probability per unit time) for the time of the next collision event is

$$P(t) = \frac{1}{\tau}\,\mathrm{e}^{-t/\tau}$$

where the time constant τ serves to make the exponent dimensionless. The factor $1/\tau$ is present to provide normalisation as you can easily show by integration,

$$\int_0^\infty \frac{1}{\tau}\,\mathrm{e}^{-t/\tau}\mathrm{d}t = 1$$

so the probability of a collision happening some time after $t = 0$ is one. Now, by further integrations you can find that the average value of t is

$$\langle t\rangle = \int_0^\infty t\,P(t)\,\mathrm{d}t = \tau$$

and the average value of t^2 is

$$\langle t^2 \rangle = \int_0^\infty t^2\, P(t)\, \mathrm{d}t = 2\tau^2$$

And so we find that the drift velocity is

$$v_d = a\tau$$

If the magnitude of the electric field is $\mathcal{E}$ and q is the charge of the carrier ($q = \pm e$) then in classical mechanics, that is, using Newton's second law,

$$m^* a = q\mathcal{E}$$

and

$$v_d = \frac{q\tau\mathcal{E}}{m^*}$$

We call τ the "relaxation time;" $1/\tau$ is the scattering rate or number of collisions per second. τ is the mean time between collisions. We then write

$$v_d = \mu\mathcal{E}$$

where $\mu = q\tau/m^*$ is the carrier mobility. The current density is related to the electric field by Ohm's law

$$J_\mathrm{e} = \sigma\mathcal{E}$$

where σ is the conductivity. If there are n_c carriers per unit volume, of charge q then

$$\begin{aligned} J_\mathrm{e} &= n_\mathrm{c} q v_d \\ &= n_\mathrm{c} q \mu \mathcal{E} \end{aligned}$$

hence

$$\boxed{\begin{aligned} \sigma &= n_\mathrm{c} q \mu \\ &= \frac{n_\mathrm{c} q^2 \tau}{m^*} \end{aligned}} \qquad (3.2.4)$$

If both holes and electrons contribute to the current, then

$$\sigma = e\left(n_\mathrm{e}\mu_\mathrm{e} + n_\mathrm{h}\mu_\mathrm{h}\right)$$

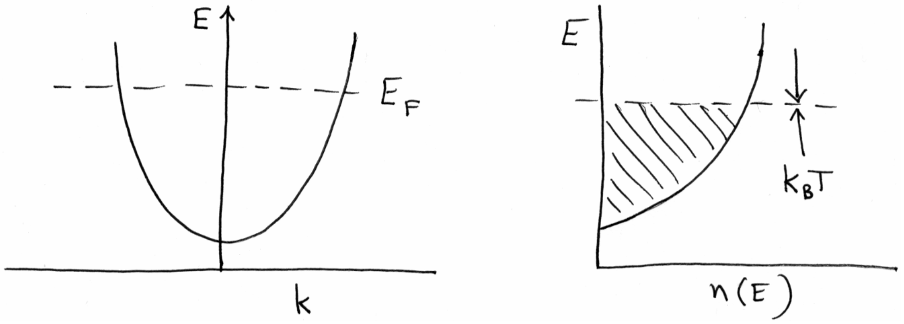
Here is a table of some semiconductor mobilites.

	Si	Ge	GaAs
μ_e ($\mathrm{m^2V^{-1}s^{-1}}$)	0.14	0.39	0.65
μ_h ($\mathrm{m^2V^{-1}s^{-1}}$)	0.05	0.19	0.04

3.2.3 Comparison between a metal and a non degenerate semiconductor

Here I want to make it clear when and how we distinguish between a degenerate and a non degenerate electron gas.

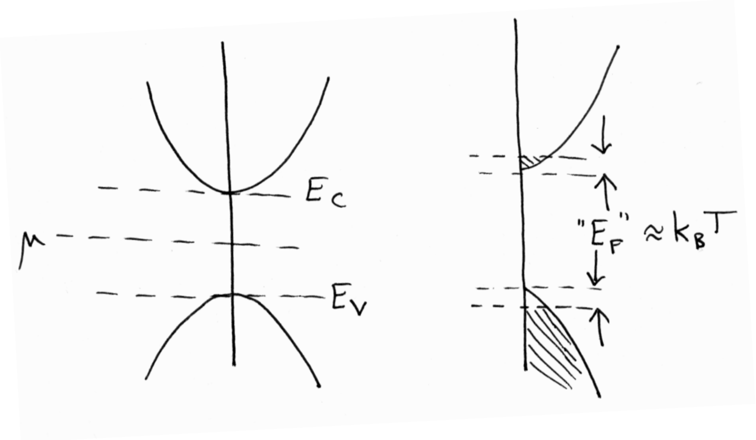
3.2.3.1 A metal



	valence	E_F (eV)	n (cm^{-3})
Al	3	11.7	0.77×10^{24}
Na	1	3.2	0.41×10^{24}

The effective mass approximation is not usually valid because we're dealing with electrons at the top of the band.

3.2.3.1 A non degenerate semiconductor



The effective mass approximation is very good because we are dealing with electrons at the bottom of the band.

	n_i at room temperature (cm^{-3})
Si	1.6×10^{16}
Ge	2.5×10^{13}
GaAs	1.2×10^{7}

3.3 Electron dynamics in metals

Equation (3.2.4) is equivalent to the Drude formula and we will see that it can also be applied in metals in which the electron gas is degenerate and so only a tiny fraction of the total number of electrons contribute to the conductivity. Nevertheless, quite surprisingly, the total electron density n will still appear as in (3.2.4). But we must proceed with proper caution.

3.3.1 Wavepackets

Firstly, an electron in a Bloch state is spread out over the whole crystal. A simple way to localise it is to form a wavepacket. We do it like this. We have already established in equation (2.3.1) that the Bloch state is a solution of the time independent Schrödinger equation in the crystal potential,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\, u_{\mathbf{k}}(\mathbf{r}) \tag{2.3.1}$$

By the rule of superposition of solutions of a linear differential equation, any linear combination of these is also a solution and we take a sum over all $\mathbf{k}$ modulated by a Gaussian function to build a collection of waves centred about a wavevector $\mathbf{k}_0$. This is just how a collection of water waves is assembled when a stone is dropped into a still pond, allowing a single wavepacket to propagate away from the centre. Considering just a single band, we construct a combination of Bloch states,

$$\begin{aligned}\psi(\mathbf{r}) &= \sum_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\, \mathrm{e}^{-\alpha(\mathbf{k}-\mathbf{k}_0)^2} \\ &= \mathrm{e}^{\mathrm{i}\mathbf{k}_0\cdot\mathbf{r}} \sum_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})\, \mathrm{e}^{-\alpha(\mathbf{k}-\mathbf{k}_0)^2}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}} \end{aligned} \tag{3.3.1}$$

and to see how it evolves with time we multiply each term by the usual phase factor of the stationary solution of the time dependent Schrödinger equation,

$$\mathrm{e}^{-\mathrm{i}E(\mathbf{k})t/\hbar}$$

Now we want the energy of the the wavepacket in terms of the eigenvalue at $\mathbf{k}_0$. So we make a Taylor expansion of $E(\mathbf{k})$ about $\mathbf{k}_0$,

$$E(\mathbf{k}) = E(\mathbf{k}_0) + (\mathbf{k}-\mathbf{k}_0)\cdot \left.\frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}}\right|_{\mathbf{k}=\mathbf{k}_0}$$

Here

$$\frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}}$$

is a shorthand for the vector

$$\boldsymbol{\nabla}_{\mathbf{k}}E = \frac{\partial E}{\partial k_x}\hat{\mathbf{x}} + \frac{\partial E}{\partial k_y}\hat{\mathbf{y}} + \frac{\partial E}{\partial k_z}\hat{\mathbf{z}}$$

with $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ the cartesian unit vectors. We now have

$$\begin{aligned}\psi(\mathbf{r},t) &= \mathrm{e}^{\mathrm{i}\mathbf{k}_0\cdot\mathbf{r}}\sum_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})\,\mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\mathrm{e}^{-\alpha(\mathbf{k}-\mathbf{k}_0)^2}\mathrm{e}^{-\mathrm{i}[E(\mathbf{k}_0)+(\mathbf{k}-\mathbf{k}_0)\cdot\boldsymbol{\nabla}_{\mathbf{k}}E]t/\hbar} \\ &= \mathrm{e}^{\mathrm{i}\mathbf{k}_0\cdot\mathbf{r}-\mathrm{i}E(\mathbf{k}_0)t/\hbar}\sum_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})\,\mathrm{e}^{-\alpha(\mathbf{k}-\mathbf{k}_0)^2}\,\mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}_0)\cdot[\mathbf{r}-\hbar^{-1}(\boldsymbol{\nabla}_{\mathbf{k}}E)t]} \qquad (3.3.2)\end{aligned}$$

Compare this carefully with (3.3.1) which is the wavepacket at $t = 0$. They have similar form; the main difference is, in (3.3.1) we have the term

$$\mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}$$

while at time t we now have the term

$$\mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}_0)\cdot[\mathbf{r}-\hbar^{-1}(\boldsymbol{\nabla}_{\mathbf{k}}E)t]}$$

Clearly the centre of the wavepacket has shifted in a time t from $\mathbf{r} = 0$, say, to $\mathbf{r} = \hbar^{-1}(\boldsymbol{\nabla}_{\mathbf{k}}E)t$. Therefore the group velocity of the wavepacket is

$$\mathbf{v}(\mathbf{k}_0) = \frac{1}{\hbar}\boldsymbol{\nabla}_{\mathbf{k}}E(\mathbf{k}_0) = \frac{1}{\hbar}\left.\frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}}\right|_{\mathbf{k}=\mathbf{k}_0} \qquad (3.3.3)$$

We can claim that this is the velocity of the electron in the state $\mathbf{k}_0$ in the crystal under the assumptions that

1. The Taylor expansion is sufficient.
2. The Bloch functions $u_{\mathbf{k}}(\mathbf{r})$ do not change significantly as the wavepacket moves.

This figure, taken from Ibach and Lüth, shows a graph of a wavepacket evolving in time. You note by comparing the factors preceeding the summation signs in (3.3.1) and (3.3.2) that the *shape* of the wavepacket does not remain constant in time. Indeed it spreads out.

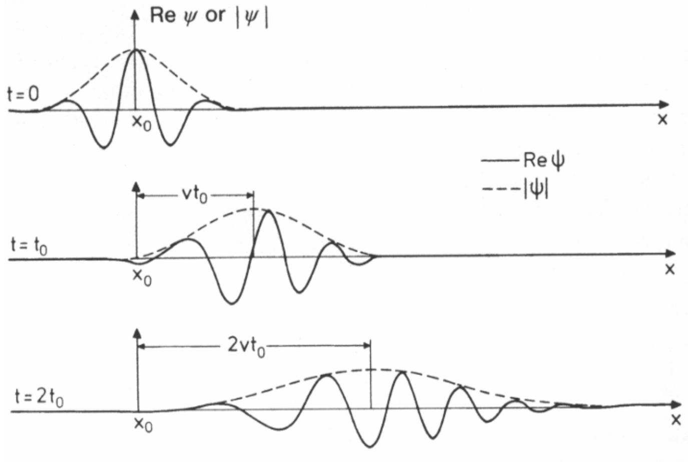


Fig. 9.1. Real space representation of the wave packet describing the motion of a spatially localized free electron at times $t = 0, t_0, 2t_0 \ldots$ (Re$\{\psi\}$: ——; $|\psi|$: – – –). The center of the wave packet, i.e., in the particle picture the electron itself, moves with the group velocity $v = \partial\omega/\partial k$. The halfwidth of the envelope increases with time. As the wave packet spreads, the wavelength of the oscillations of Re$\{\psi\}$ becomes smaller at the front and larger at the rear

Recalling Hamiltonian mechanics, section 0, velocity is

$$\dot{q} = \frac{\partial H}{\partial p}$$

Since $E(\mathbf{k})$ is the expectation value of the hamiltonian H and $\mathbf{p} = \hbar\mathbf{k}$ equation (3.3.3) is consistent with classical mechanics. This is an example of the Ehrenfest theorem of quantum mechanics which states that the expectation values of observables in quantum mechanics obey the same equations of motion, that is follow the same trajectories, as the corresponding observables in classical mechanics.

3.3.2 Velocity of free electrons

According to equation (2.2.7), for free electrons whose dispersion is

$$E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m}$$

the group velocity is

$$v = \frac{\hbar}{m}\, k$$

(By the way, the phase velocity is $\hbar k/2m$—half the group velocity). At the Fermi surface, the velocity is

$$v_\mathrm{F} = \frac{\hbar}{m}\, k_\mathrm{F}$$

called the Fermi velocity. As we remarked earlier, even at 0°K electrons at the Fermi surface are travelling at speeds of order one hundredth of the speed of light *in vacuo.*

If, on the other hand, we treated the electrons as classical particles, then according to the theorem of equipartition the kinetic energy would be

$$\frac{3}{2}k_\mathrm{B}T = \frac{1}{2}mv_\mathrm{T}^2 \tag{3.3.4}$$

hence the "thermal velocity" is

$$v_\mathrm{T} = \sqrt{\frac{3k_\mathrm{B}T}{m}} \approx 10^5\ \ \mathrm{ms}^{-1}\ \text{at } 300^\circ\mathrm{K}$$

In a simple metal like Al, we would have

$$\begin{aligned} v_\mathrm{F} &= \left(\frac{\hbar}{m}\right) k_\mathrm{F} \\ &= \left(\frac{\hbar}{m}\right) \frac{\sqrt{2mE_\mathrm{F}}}{\hbar} \\ &= \sqrt{\frac{2E_\mathrm{F}}{m}} \\ &= 2 \times 10^6\ \ \mathrm{ms}^{-1}\ \text{at } 0^\circ\mathrm{K} \end{aligned}$$

which is equivalent to a kinetic energy

$$E_\mathrm{F} = \frac{1}{2} m v_\mathrm{F}^2 \tag{3.3.5}$$

In a semiconductor, we think of the conduction band as a free electron band occupied up to $\sim k_\mathrm{B}T$, that is, non degenerate, so the carrier velocity is (noting that $\frac{1}{2}mv^2 = \frac{3}{2}k_\mathrm{B}T$)

$$\frac{\hbar}{m^*}\left(3\pi^2 n_\mathrm{e}\right)^{\frac{1}{3}} \approx \sqrt{\frac{3k_\mathrm{B}T}{m^*}}$$

so the thermal and quantum velocities are about the same.

Of course an electron doesn't travel very far at about the Fermi velocity before a collision; so equally important is its drift velocity.

3.3.3 Equation of motion of a wavepacket

Applying an electric field $\boldsymbol{\mathcal{E}}$ to a metal, we expect the wavevector $\mathbf{k}_0$ of the wavepacket to change with time although we will exclude the possibility that it makes a transition to a higher energy band. An electron will move with velocity $\mathbf{v}$ under the influence of the field which, using equation (3.3.3), does work on the electron of an amount

$$-e\boldsymbol{\mathcal{E}} \cdot \mathbf{v} = -\frac{e}{\hbar}\,\boldsymbol{\mathcal{E}} \cdot \boldsymbol{\nabla}_\mathbf{k} E(\mathbf{k})$$

per unit time, which we may equate to the rate of change of electron energy,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} \cdot \boldsymbol{\nabla}_\mathbf{k} E(\mathbf{k}) \qquad \left[\textit{i.e.,}\ \frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}}\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} \text{ by the chain rule} \right]$$

We therefore have

$$-\frac{e}{\hbar}\boldsymbol{\mathcal{E}} \cdot \boldsymbol{\nabla}_\mathbf{k} E = \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} \cdot \boldsymbol{\nabla}_\mathbf{k} E \tag{3.3.6}$$

which has the form

$$\mathbf{v}_1 \cdot \mathbf{a} = \mathbf{v}_2 \cdot \mathbf{a}$$

and one is tempted to conclude that by "cancelling" the vectors $\mathbf{a}$ on both sides,

$$\mathbf{v}_1 = \mathbf{v}_2$$

but this may be wrong because

$$\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{b}$$

is also a solution as long as $\mathbf{b} \cdot \mathbf{a} = 0$. But *don't worry*, we will say that (3.3.6) is consistent with the equation of motion

$$\boxed{\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -\frac{e}{\hbar}\,\boldsymbol{\mathcal{E}}} \tag{3.3.7}$$

and we appeal to the Ehrenfest theorem in that we identify

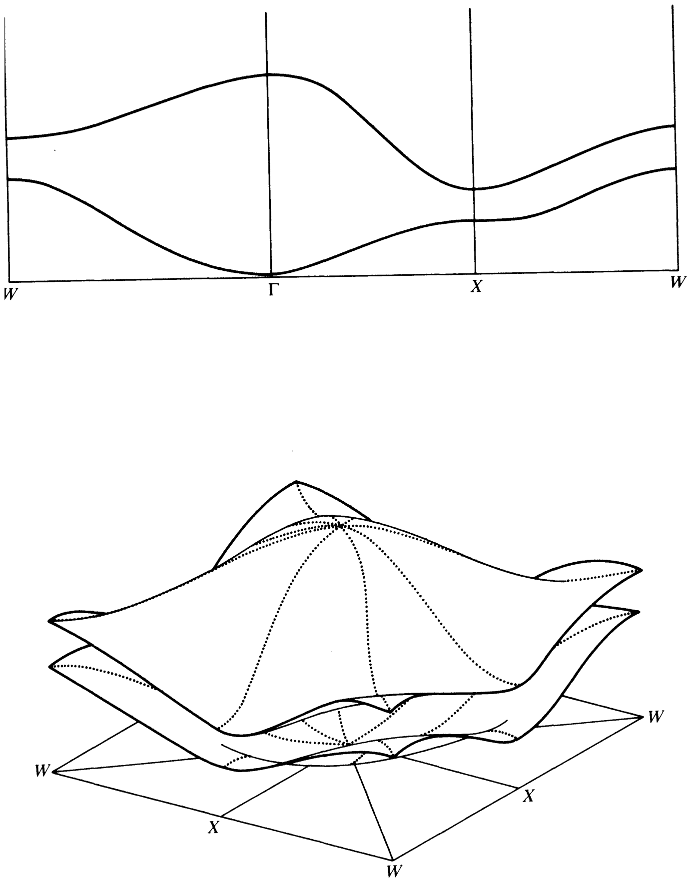
$$\hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e\boldsymbol{\mathcal{E}}$$

(rate of change of momentum = force) which is Newton's second law with momentum identified as $\hbar\mathbf{k}$. In this way we are making a semi classical approach to transport here.

Now for some fun. According to equation (3.3.7) when you apply an electric field the electron's wavevector changes, moving parallel to the field. The velocity is *not necessarily* parallel to the field. In fact equation (3.3.3) is

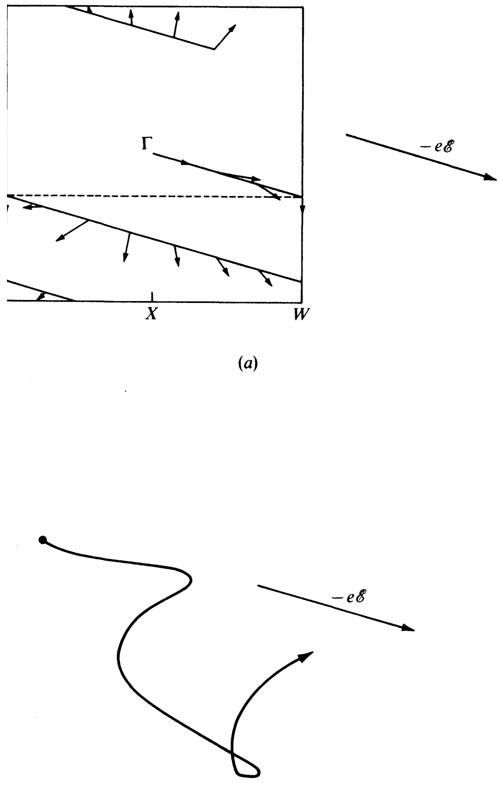
$$\boxed{\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E} \qquad (3.3.8)$$

the velocity changes in a direction that follows the steepest slope of the dispersion relation, the energy bands. Remember that $\mathbf{k}$ is the wavevector in reciprocal space but it's also the propagation direction in real space. This diagram serves to illustrate the point.



(*a*)

Fig. 2.3 Possible energy bands for a two-dimensional square lattice. Across the top of the figure are given the first two bands along symmetry lines, shown in the conventional way. Across the bottom is a sketch of the bands throughout

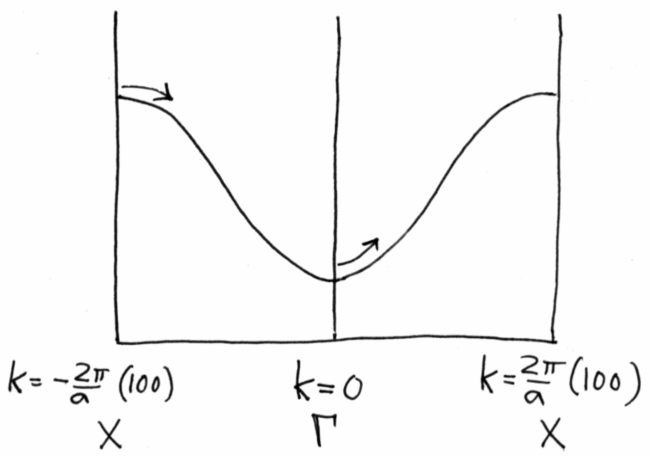


(*a*)

(*b*)

Fig. 2.4 The motion of an electron in a two-dimensional square lattice in the presence of a uniform dc electric field $\mathcal{E}$. (a) The straight-line motion in wave-number space, beginning at Γ. The $\mathrm{d}\mathbf{k}/\mathrm{d}t$ is constant except for transfers across the Brillouin zone. Velocities in real space, $\mathbf{v} = (1/\hbar)\nabla_{\mathbf{k}}E$, are shown as vectors on the trajectory, assuming an energy band such as the first band in Fig. 2.3a. (b) The shape of the corresponding electron path in real space; the scale depends on the strength of the electric field.

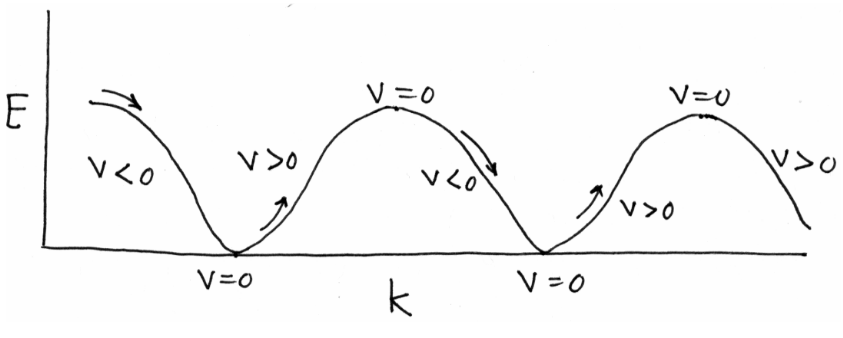
In the very simplest case, say a band along the k_x-direction and a field applied in the x-direction, we have,



an electron initially having $k = 0$ has its wavevector increased by the field along x and as it moves the slope of the dispersion relation increases and its velocity increases as

$$v_x = \frac{1}{\hbar}\frac{\mathrm{d}E}{\mathrm{d}k}$$

after the point of maximum slope it slows down until at X its velocity is zero. The electron reappears at the left and moves back towards $k = 0$ along a band of *negative* slope, hence it *reverses* its velocity. When the wavevector goes back to $k = 0$ the electron has returned to where it started! You may be able to see this better in the extended zone scheme.



This is really odd. It would seem that by applying a DC field, you can get an AC current. However, look at this more closely.

1. What is the period of this so called <u>Bloch oscillation</u>? The rate of change of k is

$$\hbar\frac{\mathrm{d}k}{\mathrm{d}t} = -e\mathcal{E}$$

so by integration

$$k - k_0 = -\frac{e}{\hbar}\mathcal{E}t$$

where k_0 is the integration constant or wavevector at $t = 0$. In a single cycle, k changes by $2\pi/a$ so the period T must be

$$T = \frac{2\pi}{a}\frac{\hbar}{e\mathcal{E}}$$

The field must be small so that the energy bands aren't tilted so much that the electron tunnnels across the bands, and so we must have a potential difference applied much less than the band width. Suppose the potential difference is one volt applied across one centimeter of metal.

$$\mathcal{E} = 1 \text{ V cm}^{-1} = 100 \text{ V m}^{-1}$$

$$\begin{aligned} T &= \frac{2\pi}{0.4 \times 10^{-9}}\frac{1.05 \times 10^{-34}}{1.6 \times 10^{-19}}\frac{1}{100} \\ &\approx 10^{-8} \text{ s} \end{aligned}$$

(I've used the lattice constant of Al.) The Fermi velocity is 2×10^6 m s^{-1} so the electron travels about one centimeter before being Bragg reflected and reversing its direction. It's bound to be scattered before then. The best hope of observing the phenomenon is within a sub band in a semiconductor conduction band in a two dimensional electron gas (see Elliott, p. 717).

2. This is a treatment of a single electron in an otherwise unoccupied band. If there are many electrons (*i.e.*, the Fermi level falls somewhere in the band) the whole sea of electrons is sloshing about collectively. If the band is full there will be as many electrons going one way as the other. Hence a metal must possess a partially occupied band. Let's put this formally. The current density is $-e$ times the number of electrons per unit volume times the velocity of each electron (see section 3.2.2). In unit volume the density of allowed $\mathbf{k}$ values is $1/(2\pi)^3$. Each band contributes the following amount to the current density,

$$\mathbf{J} = \frac{2}{(2\pi)^3}\int_{BZ} f(\mathbf{k})(-e\mathbf{v})\mathrm{d}\mathbf{k} \tag{3.3.9}$$

The factor two is for spin degeneracy; $f(\mathbf{k})$ is the Fermi–Dirac function expressed as a function of $\mathbf{k}$ rather than E,

$$f(\mathbf{k}) = f(E(\mathbf{k}))$$

In a full band, even under an applied electric field, there can be no current. For every state $\mathbf{k}$ having velocity $\mathbf{v}(\mathbf{k})$ there is a state $-\mathbf{k}$ having velocity

$$\mathbf{v}(-\mathbf{k}) = -\mathbf{v}(\mathbf{k})$$

by the symmetry of the energy bands

$$E(-\mathbf{k}) = E(\mathbf{k})$$

3. The one dimensional case is unrealistic. As we have seen the electrons follow the coupled equations (3.3.7) and (3.3.8)

$$\boxed{\begin{aligned}\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} &= -\frac{e}{\hbar}\boldsymbol{\mathcal{E}} \\ \mathbf{v} &= \frac{1}{\hbar}\boldsymbol{\nabla}_{\mathbf{k}}E\end{aligned}}$$

so the electron *doesn't travel in general in the direction of the field.* Do note that the above two equations are probably the most important of section 3. Be sure you understand them.

Once we've described the scattering time in metals we can come back and put this all together into a formula for the electrical conductivity. But let us now pursue the semi classical approach a bit further to obtain another statement of the effective mass. Recall that in section 3.2 we wrote

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\mathrm{d}^2E}{\mathrm{d}k_i\mathrm{d}k_j} \qquad (i,j = x,y,z)$$

where $\mathbf{k} = (k_1, k_2, k_3)$. Furthermore in section 3.3.3, we made the following balance of energy

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{-e}{\hbar}\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\nabla}_{\mathbf{k}}E(\mathbf{k})$$

Now the acceleration of a wavepacket, by differentiating (3.3.8) with respect to t, is

$$\begin{aligned}\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} &= \frac{1}{\hbar}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\nabla}_{\mathbf{k}}E(\mathbf{k}) \\ &= \frac{1}{\hbar}\boldsymbol{\nabla}_{\mathbf{k}}\frac{\mathrm{d}E(\mathbf{k})}{\mathrm{d}t} \\ &= -\frac{e}{\hbar^2}\boldsymbol{\nabla}_{\mathbf{k}}\left[\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\nabla}_{\mathbf{k}}E\right]\end{aligned}$$

If I write the components of the vector $\mathbf{v}$ in full, I get, for constant uniform electric field $\boldsymbol{\mathcal{E}}$,

$$\mathbf{v} = (v_x, v_y, v_z) = (v_1, v_2, v_3)$$

$$\begin{aligned}\frac{\mathrm{d}v_1}{\mathrm{d}t} &= -\frac{e}{\hbar^2}\left(\frac{\partial^2E}{\partial k_1^2}\,\mathcal{E}_1 + \frac{\partial^2E}{\partial k_1\partial k_2}\,\mathcal{E}_2 + \frac{\partial^2E}{\partial k_1\partial k_3}\,\mathcal{E}_3\right) \\ \frac{\mathrm{d}v_2}{\mathrm{d}t} &= -\frac{e}{\hbar^2}\left(\frac{\partial^2E}{\partial k_2\partial k_1}\,\mathcal{E}_1 + \frac{\partial^2E}{\partial k_2^2}\,\mathcal{E}_2 + \frac{\partial^2E}{\partial k_2\partial k_3}\,\mathcal{E}_3\right) \\ \frac{\mathrm{d}v_3}{\mathrm{d}t} &= -\frac{e}{\hbar^2}\left(\frac{\partial^2E}{\partial k_3\partial k_1}\,\mathcal{E}_1 + \frac{\partial^2E}{\partial k_3\partial k_2}\,\mathcal{E}_2 + \frac{\partial^2E}{\partial k_3^2}\,\mathcal{E}_3\right)\end{aligned}$$

which can be assembled into a single formula,

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\frac{e}{\hbar^2}\sum_j\frac{\partial^2E}{\partial k_i\partial k_j}\mathcal{E}_j \qquad (i,j = 1,2,3)$$

This compares with the classical equation for the acceleration of a point charge $-e$,

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\frac{e}{m}\mathcal{E}_i$$

in which the velocity is parallel to the electric field. In the case of the solid state wavepacket, the situation is more general—a component of the electric field $\mathcal{E}_j$ causes an acceleration in the i-direction $\mathrm{d}v_i/\mathrm{d}t$

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -e\left(\frac{1}{\hbar^2}\right)\sum_j \left(\frac{1}{m^*}\right)_{ij}\mathcal{E}_j$$

so $(1/m^*)_{ij}$ is a tensor with components

$$\frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_i \partial k_j}$$

The direction of acceleration, as we've said before, is not necessarily in the direction of the field—it depends on the wavevector of the wavepacket and on the details of the energy band structure. Even under a uniform electric field the wavepacket keeps changing direction as its wavevector changes; see p 96.

And this is still in the absence of scattering. . .

3.4 Scattering using the "kinetic method"—the relaxation time approximation

Ultimately we want to understand transport coefficients, especially in thermal and electrical conductivity. We can't follow an electron or phonon through all the many scattering events as it drifts under the influence of a temperature gradient or electric field (or both). One approach is to borrow concepts such as mean free path and collision rate from the kinetic theory of gasses.

3.4.1 Relaxation time

Random events are often described using the Poisson distribution of elementary statistics and probability. A particle suffers collisions on average at a rate of $1/\tau$ so τ is the mean time between collisions. So what is the probability that it survives a time t without a collision? If the rate is $1/\tau$ then the probability of suffering a collision in a time interval $\mathrm{d}t$ is

$$\frac{\mathrm{d}t}{\tau}$$

So the probability of surviving in this interval is

$$P_s(\mathrm{d}t) = 1 - \frac{\mathrm{d}t}{\tau}$$

Then the probability of surviving a total time of $t + \mathrm{d}t$ is

$$P_s(t + \mathrm{d}t) = P_s(t) \times P_s(\mathrm{d}t)$$

assuming the two probabilities are independent. Then

$$P_s(t + \mathrm{d}t) = P_s(t) \left(1 - \frac{\mathrm{d}t}{\tau}\right)$$

or, rearranging,

$$\frac{P_s(t + \mathrm{d}t) - P_s(t)}{\mathrm{d}t} = -\frac{1}{\tau}\, P_s(t)$$

which means

$$\frac{\mathrm{d}P_s}{\mathrm{d}t} = -\frac{1}{\tau}\, P_s$$

The probability of surviving thus decays in time as minus the scattering rate times the probability of surviving thus far. The solution for $P_s(t)$ is

$$P_s(t) = \mathrm{e}^{-t/\tau} \tag{3.4.1}$$

as you can check by differentiation. The normalised probability density as indeed we saw in section 3.2.2 is

$$P(t) = \frac{1}{\tau}\, P_s(t)$$

3.4.2 The kinetic method

So we have that the collisions of the electrons in a metal follow a Poisson distribution having the properties that the probability density for collisions is

$$\begin{aligned} P(t) &= \frac{1}{\tau}\, P_s(t) \\ &= \frac{1}{\tau}\, \mathrm{e}^{-t/\tau} \end{aligned} \tag{3.4.2}$$

where $P_s(t)$ is the probability that a particle chosen at random at time $t = 0$ will first scatter after a time t. We also have that

$$\frac{\mathrm{d}P_s}{\mathrm{d}t} = -\frac{1}{\tau}\, P_s$$

Now we focus upon an electron having ballistic velocity $\mathbf{v}$, in a uniform electric field $\boldsymbol{\mathcal{E}}$. The electron takes energy from the field of an amount

$$\delta E = -e\mathbf{v} \cdot \boldsymbol{\mathcal{E}}\, t \tag{3.4.3}$$

in a time t. Be careful of the sign. If $\mathbf{v}$ is *in the opposite direction* to the field then $\mathbf{v} \cdot \boldsymbol{\mathcal{E}} < 0$ and the electron speeds up because it's negatively charged. Since by convention $e > 0$, in this case δE is *positive.* (If $\mathbf{v}$ is parallel to $\boldsymbol{\mathcal{E}}$ then the electron slows down so it gives up some kinetic energy to the field and then $\delta E < 0$.) This extra energy causes the electron to *drift*; it acquires a small increment to its otherwise random direction of flight. We write this increment in a formal sort of a way as

$$\delta \mathbf{v} \equiv \delta \mathbf{v}_d = \frac{\delta \mathbf{v}}{\delta E}\, \delta E$$

How much energy can the electron gather from the field before it scatters? It's a matter of summing up over increments of time $\mathrm{d}t$, the increment of energy times the (positive) increment in survival probability, $-\mathrm{d}P_s$. You have seen in section 3.2.2 that the time, t, averaged over $P(t)$ is

$$\langle t \rangle = \int_0^\infty P(t)\, t\, \mathrm{d}t = \tau$$

which you can do by partial integration using (3.4.2). $\mathbf{v}$ and $\boldsymbol{\mathcal{E}}$ are assumed independent of t, so we can write

$$\begin{aligned} E &= \int_0^\infty \delta E \left(-\frac{\mathrm{d}P_s}{\mathrm{d}t} \right) \mathrm{d}t \\ &= -e\mathbf{v} \cdot \boldsymbol{\mathcal{E}} \int_0^\infty \frac{1}{\tau}\, \mathrm{e}^{-t/\tau} t\, \mathrm{d}t \\ &= -e\mathbf{v} \cdot \boldsymbol{\mathcal{E}} \int_0^\infty P(t)\, t\, \mathrm{d}t \\ &= -e\mathbf{v} \cdot \boldsymbol{\mathcal{E}}\, \tau \end{aligned}$$

This is just a very long winded way of taking the average of δE in (3.4.3). The drift velocity is obtained, again rather formally, from

$$\begin{aligned} \mathbf{v}_d &= \frac{E}{\delta E / \delta \mathbf{v}} \\ &= -e\tau \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}E}\, (\mathbf{v} \cdot \boldsymbol{\mathcal{E}}) \end{aligned}$$

You may want to brood over this for a while. The drift velocity of the electron is its charge, times its relaxation time, times the scalar product of its ballistic velocity with the field, times the rate of change of $\mathbf{v}$ with energy. This latter concerns the way in which you change the velocity as you change its energy. You've seen in section 3.3.3 how the velocity changes in a complex way with the eigenvalue through

$$\mathbf{v} = \frac{1}{\hbar} \frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}} \tag{3.3.8}$$

so by changing the electron's energy, you change both its speed and its *direction.* (See p 96.) So the complex nature of the energy bands is accounted for in this formula for the drift velocity.

The electrical current density is of opposite sign to the drift velocity because by convention current is the flow of positive charge.

$$\boxed{\begin{aligned} \mathbf{J} &= -n_\mathrm{e} e \mathbf{v}_d \\ &= n_\mathrm{e} e^2 \tau \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}E}\, (\mathbf{v} \cdot \boldsymbol{\mathcal{E}}) \end{aligned}}$$

In the case of *isotropic bands*.

$$\begin{aligned} J &= n_\mathrm{e} e^2 \tau \frac{\mathrm{d}v}{\mathrm{d}E}\, v\mathcal{E} \\ &= \sigma\mathcal{E} \end{aligned}$$

the last line being Ohm's Law. We can relate the energy to velocity in the free electron gas, since we have from (3.3.8)

$$v = \hbar k/m$$

So

$$E \equiv E(k) = \frac{\hbar^2 k^2}{2m} = \frac{1}{2} m v^2$$

This is just the same as the classical formula for kinetic energy, but m here may be the effective mass. From this,

$$\frac{\mathrm{d}v}{\mathrm{d}E} = \frac{1}{mv}$$

and we end up with the Drude formula for the conductivity,

$$\boxed{\sigma = \frac{n_\mathrm{e} e^2 \tau}{m}} \qquad (3.4.4)$$

3.4.3 Mean free path

From equation (3.4.4) it looks as though the conductivity doesn't depend on the ballistic velocity $\mathbf{v}$, which is fast—some electrons are travelling at speeds near the Fermi velocity. However it *does* because it is the density of scatterers that is most relevant and the relaxation time then depends on $\mathbf{v}$. In fact the average distance between scatterers or rather the average distance an electron or phonon travels between collisions is called the mean free path λ defined such that

$$\lambda = \tau v$$

so that the conductivity in terms of mean free path is

$$\boxed{\sigma = \frac{n_\mathrm{e} e^2 \lambda}{mv}} \qquad (3.4.5)$$

3.4.4 Thermal conductivity

Now let's use the kinetic method to look at thermal conductivity in metals due to electrons. We consider a long bar of metal, hot at one end and cold at the other, in the case where the temperature gradient ∇T is uniform. In one dimension, if the x-direction is chosen to be along the length of the bar then the general formula for the heat flux

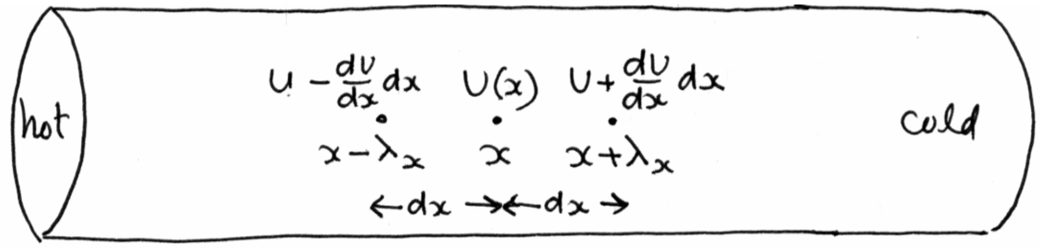
$$\mathbf{J}_Q = -\kappa \nabla T$$

becomes

$$J_Q = -\kappa \frac{\mathrm{d}T}{\mathrm{d}x}$$

and κ is called the thermal conductivity (J s^{-1} m^{-1} K^{-1}).

We assume that at any point in the bar where the temperature is $T(x)$ the electrons possess internal energy $U(T)$ which is the internal energy they *would* possess in a bar of the same metal at uniform temperature T. Now these electrons are travelling in random directions in between collisions at about the Fermi velocity $\mathbf{v}$ and their relaxation time is τ. Now consider a point at position x along the bar. Electrons arrive from the left at velocity $v_x = v$ having left their last collision at a point $x - \mathrm{d}x = x - \lambda_x$ where λ_x is the mean free path along x. They carry energy *to* the point x and their energy is the energy of electrons in the metal at the temperature of the bar at the point $x - \lambda_x$. Electrons also arrive from the right having left their last collision at $x + \lambda_x = x + \mathrm{d}x$ and they bring energy to the point x associated with the temperature in the bar at $x + \mathrm{d}x$.



Electrons arrive from the left carrying energy

$$U - \frac{\mathrm{d}U}{\mathrm{d}x}\mathrm{d}x$$

(assuming $\mathrm{d}x = \lambda_x$ is small enough to take just the linear term in the Taylor expansion of $U(x)$). Since v doesn't depend on temperature (it's the Fermi velocity, not the thermal velocity—see section 3.3.2) half of the total number of electrons arrive from the left bringing a flux of heat equal to

$$\frac{1}{2} n_\mathrm{e} v_x \left(U - \frac{\mathrm{d}U}{\mathrm{d}x}\mathrm{d}x \right)$$

where n_e is the number of electrons per unit volume. The flux from the right is

$$\frac{1}{2} n_\mathrm{e} v_x \left(U + \frac{\mathrm{d}U}{\mathrm{d}x}\mathrm{d}x \right)$$

Therefore the total flux from left to right is the difference of these two,

$$J_Q = -\frac{1}{2} n_{\mathrm{e}} v_x \, 2 \frac{\mathrm{d}U}{\mathrm{d}x} \mathrm{d}x$$

but $\mathrm{d}x = \lambda_x = v_x \tau$ and using the chain rule of differentiation

$$\frac{\mathrm{d}U}{\mathrm{d}x} = \frac{\mathrm{d}U}{\mathrm{d}T} \frac{\mathrm{d}T}{\mathrm{d}x}$$

we get

$$\begin{aligned} J_Q &= n_{\mathrm{e}} v_x^2 \tau \frac{\mathrm{d}U}{\mathrm{d}T} \left(-\frac{\mathrm{d}T}{\mathrm{d}x} \right) \\ &= \kappa \left(-\frac{\mathrm{d}T}{\mathrm{d}x} \right) \end{aligned}$$

Now, $n_{\mathrm{e}} \mathrm{d}U/\mathrm{d}T$ is the electron contribution to the heat capacity *per unit volume*, c_{e}. Also, to generalise to three dimensions we merely have to use

$$v_x^2 = \frac{1}{3} v^2 \tag{3.4.6}$$

(that is, $v^2 = v_x^2 + v_y^2 + v_z^2 = 3v_x^2$) so we get for the thermal conductivity

$$\boxed{\begin{aligned} \kappa &= \frac{1}{3} v^2 c_{\mathrm{e}} \tau \\ &= \frac{1}{3} v \lambda c_{\mathrm{e}} \end{aligned}}$$

a nice simple formula like (3.4.5).

We could do exactly the same excercise in the case that the carriers of thermal energy were phonons. The result we would get is

$$\kappa_{\mathrm{ph}} = \frac{1}{3} v_{\mathrm{ph}} \lambda_{\mathrm{ph}} c_{\mathrm{ph}} \tag{3.4.7}$$

Suppose that the mean free paths are the same for phonons as for electrons; which are the dominant heat carriers in metals? You know that the heat capacity C_{ph} due to phonons is some hundred times larger than that due to electrons C_{e} (see section 2.2), but if you work it out you'll find the phonon velocity is typically 1000 smaller than the Fermi velocity. Hence, in metals, electrons are the dominant carriers of heat. In insulators electrons are largely immobile and phonon thermal conductivity dominates. Diamond has an especially high thermal conductivity. Here are some selected values of κ at room temperature.

	κ (Wm^{-1}K^{-1})
Copper	397
Aluminium	240
Titanium	22
Iron	78
Silicon	150
Diamond	1000–5000[†]
Alumina	26
Glass	1

[†] The higher values for *isotopically pure* diamond

3.4.5 The role of Umklapp processes in phonon thermal conductivity

If thermal conductivity is phonon dominated (*e.g.*, in insulators) then thermal energy is dissipated by the transfer of momentum. There needs to be a mechanism that allows the total phonon wavevector to change so that a region at high temperature can come to equilibrium at a lower temperature. Suppose at some point in the solid there is a total momentum flux,

$$\mathbf{J}_{\mathrm{ph}} = \hbar \sum_{\mathbf{k}} \sum_{j} \mathbf{k} n_j(\mathbf{k})$$

where j labels the phonon bands, acoustic and optic, and $n_j(\mathbf{k})$ is the Bose–Einstein distribution for wavevector $\mathbf{k}$ and band j. If all the collisions are such that

$$\mathbf{k} + \mathbf{k}' = \mathbf{k}''$$

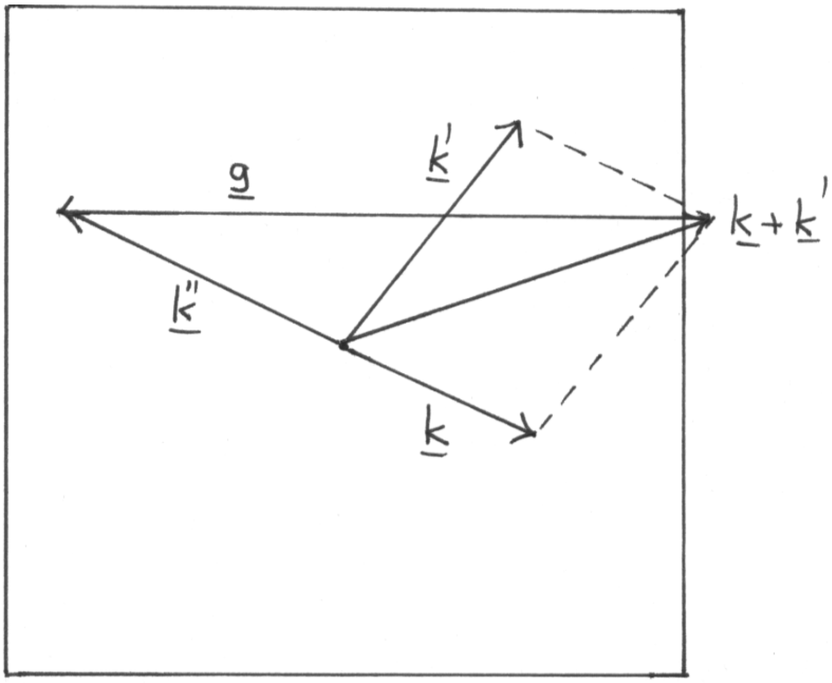
then the total momentum cannot change with time because the sum of the wavevectors $\mathbf{k} + \mathbf{k}' - \mathbf{k}'' = \mathbf{0}$ is unchanging in time. Once such a flux as $\mathbf{J}_{\mathrm{ph}}$ has been established, even in a vanishingly small temperature gradient, the flow will continue for ever. The thermal conductivity will be effectively infinite in the absence of other scattering (*e.g.*, from defects).[†]

To allow phonons to come to equilibrium requires Umklapp processes in which

$$\mathbf{k} + \mathbf{k}' = \mathbf{k}'' + \mathbf{g}$$

with $\mathbf{g} \neq \mathbf{0}$.

[†] The same argument can be applied to an infinitely long tube containing gas. But as long as the tube is finite then equilibrium can be established by gas molecules reversing their directions when they bounce off the closed ends of the tube.



Two phonons carrying momentum to the right disappear and produce a phonon carrying momentum to the left.

At temperatures well below the Debye temperature Umklapp processes are exponentially unlikely. This is called “freezing out of U-processes.” Hence you can believe that the relaxation time increases like

$$\tau \sim \mathrm{e}^{\theta_\mathrm{D}/T}$$

and the conductivity increases like

$$\kappa \sim \mathrm{e}^{\gamma\theta_\mathrm{D}/T}$$

where γ is a factor less than one. This is seen in the thermal conductivity of lithium fluoride crystals. From 100°K down to about 10°K the conductivity increases exponentially. It doesn't increase further because the mean free path becomes as big as the crystal and phonons scatter at the surfaces. Hence at $T < 10°\mathrm{K}$ the conductivity decreases again but now *depends on the size of the crystal*, and increases like T^3 as does the heat capacity.

At temperatures above θ_D we have

$$n_j(\mathbf{k}) \longrightarrow \frac{k_\mathrm{B}T}{\hbar\omega_j(\mathbf{k})}$$

as we have seen earlier in section 1.6.2, equation (1.6.5). So the phonon density increases linearly with temperature. The conductivity is then seen to decrease with temperature according to

$$\kappa \sim \frac{1}{T^\alpha}$$

where α is between one and two. The following figure shows these features.

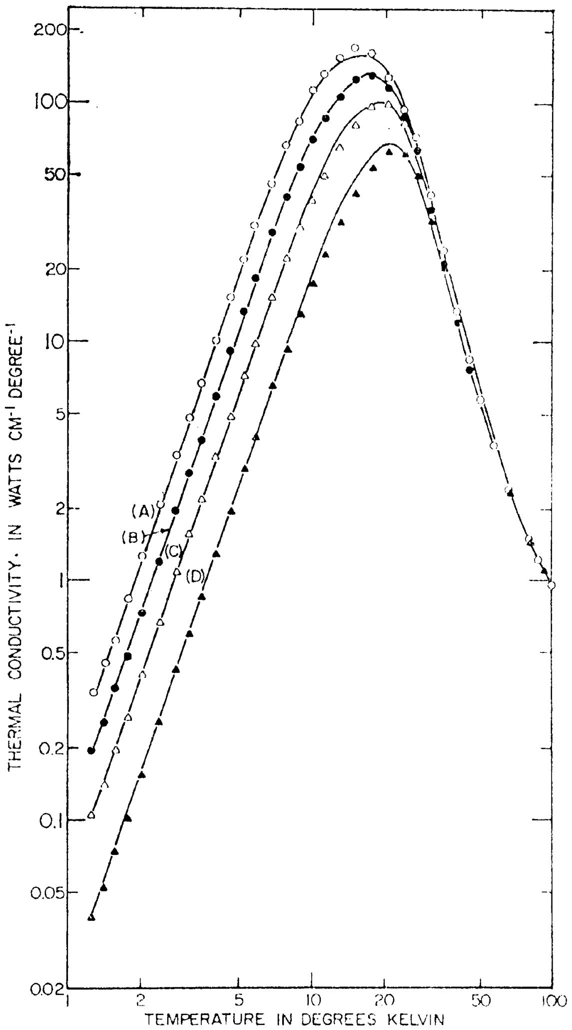


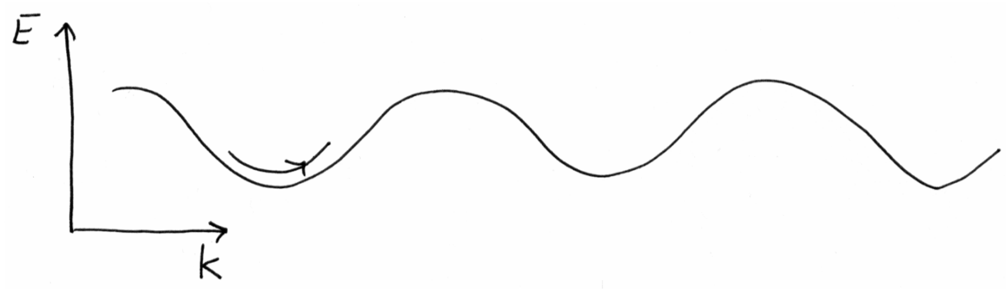
FIG. 1. Thermal conductivity of isotopically pure LiF showing the effect of boundaries for sandblasted crystals. Mean crystal widths: (A) 7.25 mm, (B) 4.00 mm, (C) 2.14 mm, (D) 1.06 mm.

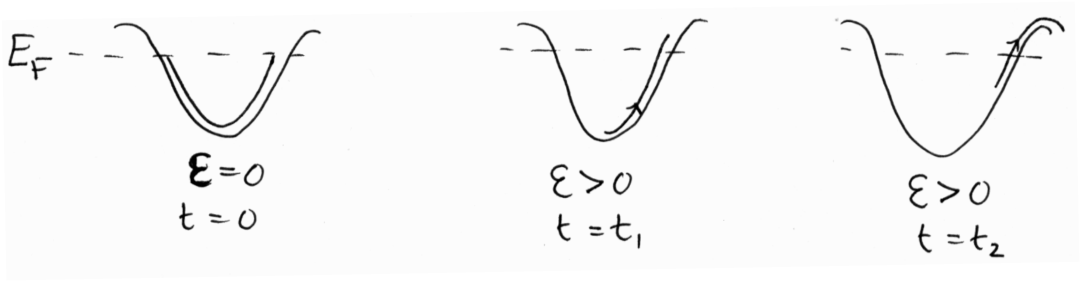
3.5 Boltzmann formulation of electrical conductivity in metals

We have seen that, in the absence of scattering, all electrons under a small constant applied electric field $\boldsymbol{\mathcal{E}}$ in a metal change their wavevectors according to equation (3.3.7)

$$\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -\frac{1}{\hbar}\, e\boldsymbol{\mathcal{E}} \tag{3.3.7}$$

In one dimension, the electron wavevectors move like a train over some hills.





In three dimensions, the whole "Fermi sea" (all electrons enclosed within the Fermi surface) is displaced such that after a time δt, according to equation (3.3.7) all wavevectors are shifted by an amount

$$\delta\mathbf{k} = -\frac{1}{\hbar}\, e\,\boldsymbol{\mathcal{E}}\,\delta t$$

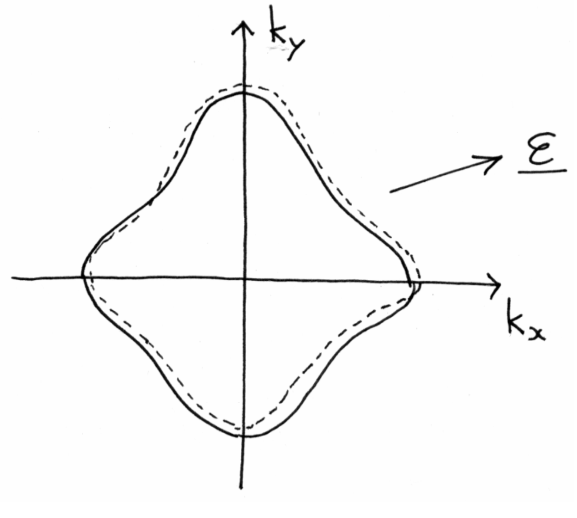
This doesn't continue for ever because the electrons are scattered. Scattering reverses the change in **k** by acting to return the distribution to its equilibrium. In the steady state the rate of change of **k** due to scattering is equal and opposite to that due to the field. To estimate the rate due to scattering, we suppose that the rate of approach to equilibrium is proportional to the deviation *from* equilibium,

$$\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -\frac{\delta\mathbf{k}}{\tau} \tag{3.5.1}$$

the proportionality constant being the scattering rate $1/\tau$. This is the same argument that we used in section 3.4.1 leading to equation (3.4.1), and equating the two rates (3.3.7) and (3.5.1) we get

$$\delta\mathbf{k} = \frac{1}{\hbar}\, e\,\tau\,\boldsymbol{\mathcal{E}}$$

which represents a rigid displacement of the Fermi surface like this.



Let us suppose we have established a steady state using an applied electric field so that the electron probability distribution function is not the Fermi–Dirac function f_{FD} but some stationary, non equilibrium distribution f_{stat}. If I now switch off the electric field then the distribution will relax *towards* f_{FD} at a rate proportional to its deviation *from* equilibrium,

$$\frac{\partial f(\mathbf{r},\mathbf{k},t)}{\partial t}=-\frac{f(\mathbf{r},\mathbf{k},t)-f_{\mathrm{FD}}}{\tau} \tag{3.5.2}$$

where in general the occupation probability depends on $\mathbf{k}$ and on position and on time. The solution to (3.5.2) is

$$f-f_{\mathrm{FD}}=(f_{\mathrm{stat}}-f_{\mathrm{FD}})\mathrm{e}^{-t/\tau} \tag{3.5.3}$$

which you should compare with equation (3.4.1).

The Boltzmann equation is a general formula for $f(\mathbf{r},\mathbf{k},t)$. Recall how we did thermal conductivity in section 3.4.4. At time t the occupancy of a wavepacket at $\mathbf{r}$ of wavevector $\mathbf{k}$ is the *same as it was* at $\mathbf{r}-\mathbf{v}\delta t$ when its wavevector was

$$\mathbf{k}-\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}\,\delta t$$

a short time ago. The thinking behind the Boltzmann equation, which we will now derive, is this. An electron wavepacket of wavevector $\mathbf{k}$ is either occupied or unoccupied, and it remains either occupied or unoccupied as it propagates under the influence of applied fields (electric, magnetic or gradients of temperature) *unless it becomes scattered* into a state of a different $\mathbf{k}$. First we find the formula for the occupation probability as a function of time in the absence of scattering; then we add a term to account for scattering, and we try to solve the equation in the relaxation time approximation. We will do this slowly and carefully, starting with some background mathematics of Taylor series.

3.5.1 Derivation of the Boltzmann equation

If $f(x)$ is a function of x, then to first order its Taylor expansion is

$$f(x+h)=f(x)+h\,\frac{\mathrm{d}f}{\mathrm{d}x}$$

and it's understood that the derivative *is evaluated at* x, not $x+h$ or anywhere else. This rule holds throughout what follows. If $h=\mathrm{d}x$ is infinitessimal then

$$f(x+\mathrm{d}x)=f(x)+\frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x$$

If $f(x,k)$ depends on *two* variables x and k then

$$f(x+\mathrm{d}x,k+\mathrm{d}k)=f(x,k)+\frac{\partial f}{\partial x}\,\mathrm{d}x+\frac{\partial f}{\partial k}\,\mathrm{d}k \tag{3.5.4}$$

where now these are *partial* derivatives: other variables being held constant. If I make x and k vectors, $\mathbf{r}$ and $\mathbf{k}$, then a function $f(\mathbf{r})$ now depends on *three* variables, the three components of $\mathbf{r}=x\hat{\mathbf{x}}+y\hat{\mathbf{y}}+z\hat{\mathbf{z}}$. So instead of

$$\frac{\partial f}{\partial x}$$

I need to put

$$\frac{\partial f}{\partial \mathbf{r}}$$

which is a vector having three components. It's given the symbol

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}}$$

To make it clear that the partial derivatives are taken with respect to the components of $\mathbf{r}$ we will write $\boldsymbol{\nabla}$ as $\boldsymbol{\nabla}_{\mathbf{r}}$. Then when we also need to take partial derivatives with respect to the components of $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$ we can use the symbol $\boldsymbol{\nabla}_{\mathbf{k}}$.

We now use $f(\mathbf{r}, \mathbf{k}, t)$ to represent the probability distribution function that we have already described at equation (3.5.2). It depends on $\mathbf{r}$, $\mathbf{k}$ and time, t; hence seven variables. But I can still do a Taylor expansion, as in (3.5.4), but now I'll get seven partial derivative terms,

$$f(\mathbf{r} + \mathrm{d}\mathbf{r}, \mathbf{k} + \mathrm{d}\mathbf{k}, t + \mathrm{d}t) = f(\mathbf{r}, \mathbf{k}, t) + \boldsymbol{\nabla}_{\mathbf{r}} f \cdot \mathrm{d}\mathbf{r} + \boldsymbol{\nabla}_{\mathbf{k}} f \cdot \mathrm{d}\mathbf{k} + \frac{\partial f}{\partial t}\mathrm{d}t \qquad (3.5.5)$$

and here $\boldsymbol{\nabla}_{\mathbf{r}}$ and $\boldsymbol{\nabla}_{\mathbf{k}}$ each gather up three terms into one using the scalar product. In this way, since $\mathrm{d}\mathbf{r} = \mathrm{d}x\hat{\mathbf{x}} + \mathrm{d}y\hat{\mathbf{y}} + \mathrm{d}z\hat{\mathbf{z}}$,

$$\boldsymbol{\nabla}_{\mathbf{r}} f \cdot \mathrm{d}\mathbf{r} = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y + \frac{\partial f}{\partial z}\mathrm{d}z$$

and similarly for the components of $\mathbf{k}$ in the $\boldsymbol{\nabla}_{\mathbf{k}}$ terms:

$$\boldsymbol{\nabla}_{\mathbf{k}} f \cdot \mathrm{d}\mathbf{k} = \frac{\partial f}{\partial k_x}\mathrm{d}k_x + \frac{\partial f}{\partial k_y}\mathrm{d}k_y + \frac{\partial f}{\partial k_z}\mathrm{d}k_z$$

We have

$$\dot{\mathbf{r}} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}, \quad \text{the velocity,} \qquad \text{and} \qquad \dot{\mathbf{k}} = \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}$$

so putting $\mathrm{d}\mathbf{r} = \dot{\mathbf{r}}\,\mathrm{d}t$ and $\mathrm{d}\mathbf{k} = \dot{\mathbf{k}}\,\mathrm{d}t$ into (3.5.5), changing the sign of $\mathrm{d}t$ so that we are looking backwards in time, and rearranging, I get

$$f(\mathbf{r}, \mathbf{k}, t) - f(\mathbf{r} - \mathrm{d}\mathbf{r}, \mathbf{k} - \mathrm{d}\mathbf{k}, t - \mathrm{d}t) = \frac{\partial f}{\partial t}\mathrm{d}t + \boldsymbol{\nabla}_{\mathbf{r}} f \cdot \dot{\mathbf{r}}\mathrm{d}t + \boldsymbol{\nabla}_{\mathbf{k}} f \cdot \dot{\mathbf{k}}\mathrm{d}t \qquad (3.5.6)$$

Now we argue that this is zero in the absence of scattering because the probability of occupancy does not change if the particle arrives at $\mathbf{r}$ having wavevector $\mathbf{k}$ at time t, unscattered since a time $t - \mathrm{d}t$ in the past when it was at $\mathbf{r} - \mathrm{d}\mathbf{r}$ having wavevector $\mathbf{k} - \mathrm{d}\mathbf{k}$. (Its position and wavevector are changing, of course, under the action of our applied fields.) To include the possibility that the particle was indeed scattered in the interval $\mathrm{d}t$, we equate the difference at the left hand side of (3.5.6) not to zero, but to

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} \mathrm{d}t$$

which is the rate of change of the occupation probability due to collisions, multiplied by the time interval, dt, and so equating this to the right hand side of (3.5.6) we get the generalised Boltzmann transport equation,

$$\boxed{\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} = \frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} f \cdot \dot{\mathbf{r}} + \nabla_{\mathbf{k}} f \cdot \dot{\mathbf{k}}} \tag{3.5.7}$$

which in fact applies to any particle, including phonons, if the appropriate distribution function (Fermi–Dirac, Bose–Einstein or Maxwell–Boltzmann) is used for f.

If we use our formulas (3.3.7) and (3.3.8) for $\mathbf{v}$ and $\dot{\mathbf{k}}$ for electron wavepackets, then we get

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} = \frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} f \cdot \frac{1}{\hbar}\nabla_{\mathbf{k}} E + \nabla_{\mathbf{k}} f \cdot \left(-\frac{e}{\hbar}\right)\boldsymbol{\mathcal{E}} \tag{3.5.8}$$

while we omit the effects of magnetic fields for which we would require the *Lorentz force.* Since

$$\left(\frac{\partial f(\mathbf{r},\mathbf{k},t)}{\partial t}\right)_{\text{coll.}}$$

represents scattering transitions between states it must be related to $w_{\mathbf{k}\to\mathbf{k}'}$ in equation (3.1.1). In fact,

$$\begin{aligned}\left(\frac{\partial f(\mathbf{r},\mathbf{k},t)}{\partial t}\right)_{\text{coll.}} &= \frac{1}{(2\pi)^3}\int_{BZ} \{f(\mathbf{r},\mathbf{k}',t) w_{\mathbf{k}'\to\mathbf{k}}\left[1 - f(\mathbf{r},\mathbf{k},t)\right] \\ &\qquad - f(\mathbf{r},\mathbf{k},t) w_{\mathbf{k}\to\mathbf{k}'}\left[1 - f(\mathbf{r},\mathbf{k}',t)\right]\}\,\mathrm{d}\mathbf{k}' \\ &= \frac{1}{(2\pi)^3}\int_{BZ} w_{\mathbf{k}\to\mathbf{k}'}\left[f(\mathbf{r},\mathbf{k}',t) - f(\mathbf{r},\mathbf{k},t)\right]\mathrm{d}\mathbf{k}' \end{aligned} \tag{3.5.9}$$

since $w_{\mathbf{k}\to\mathbf{k}'} = w_{\mathbf{k}'\to\mathbf{k}}$ (see equation (3.1.2)). We read this equation as follows. In the first line we have all the possible scattering events *into* the state $\mathbf{k}$; that is, the transition probability $w_{\mathbf{k}'\to\mathbf{k}}$ times the probability that the initial state $\mathbf{k}'$ is occupied times the probability that the final state $\mathbf{k}$ is *un*occupied. This is integrated over all possible intial states $\mathbf{k}'$. In the second line we have all the possible transitions *out of* state $\mathbf{k}$, with a minus sign, and the sum of these two gives the total rate of change of the occupancy of state $\mathbf{k}$ due to scattering. The purpose of writing this down is not to frighten you, but to close the loop that I started with equation (3.1.1) at the beginning of section 3. It emphasises that the rate of change of f due to collisions is not a completely unknown quantity; it can at least in principle be calculated using perturbation methods in quantum mechanics. Because of equation (3.1.2) it simplifies nicely and comes down to knowing the function $f(\mathbf{r},\mathbf{k},t)$ which is the subject of the generalised equation (3.5.7) and the probability densities $w_{\mathbf{k}\to\mathbf{k}'}$ for all possible scattering events. It is these latter quantities that must be calculated using the advanced, many particle quantum mechanics of electron–phonon and electron–electron interactions.

3.5.2 Solution of the linearised Boltzmann equation

We'd like to solve the Boltzmann equation for f, because then we'd know the positions and wavevectors of all the particles as a function of time—or at least the occupation probabilites, which is all you ever get in quantum mechanics. In the steady state f is unchanging in time, *i.e.*,

$$\frac{\partial f}{\partial t} = 0$$

and we treat a *spatially homogeneous* system, which means there are no variations in chemical composition, or different materials separated by interfaces or junctions, in which case $\boldsymbol{\nabla}_{\mathbf{r}} f = 0$; indeed nothing depends on $\mathbf{r}$.[†] If we are interested in electrons in a metal in a uniform electric field, $\boldsymbol{\mathcal{E}}$, then we have

$$\dot{\mathbf{k}} = -\frac{e}{\hbar}\,\boldsymbol{\mathcal{E}}$$

Finally we make the relaxation time approximation to the rate of scattering, as in equation (3.5.3), so that

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} = -\frac{f - f_{\text{FD}}}{\tau}$$

and (3.5.7) becomes

$$f(\mathbf{k}) = f_{\text{FD}} + \frac{e}{\hbar}\,\tau(\mathbf{k})\,\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\nabla}_{\mathbf{k}} f \qquad (3.5.10)$$

allowing that each wavepacket has its own relaxation time, depending on its wavevector $\mathbf{k}$. This is a differential equation in f and its derivatives with respect to the components of $\mathbf{k}$.

In the form (3.5.8) with equation (3.5.9) for the collision, or *drift* term, the Boltzmann equation is said to be an *integro-differential* equation since f appears both under an integral sign and as a differential. If we make the relaxation time approximation then we obtain (3.5.10) which is a linear differential equation since f appears on the left, and again on the right differentiated with respect to the components of $\mathbf{k}$. To continue to make progress we make a further approximation, namely to *linearise* the right hand side. This is reasonable in the case of small fields where you'd expect the response to be linear, for example as in Ohm's Law. We can then replace $\boldsymbol{\nabla}_{\mathbf{k}} f$ by $\boldsymbol{\nabla}_{\mathbf{k}} f_{\text{FD}}$ in equation (3.5.10) amounting to a first order Taylor expansion of $f(\mathbf{k})$ in which case the formula remains correct to order $\mathcal{E}^2$. Thus the *linearised* Boltzmann equation is

$$\boxed{f(\mathbf{k}) = f_{\text{FD}} + \frac{e}{\hbar}\,\tau(\mathbf{k})\,\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\nabla}_{\mathbf{k}} f_{\text{FD}}} \qquad (3.5.11)$$

[†] This demands also that there are no temperature gradients, since f depends implicitly on T just as do the Fermi–Dirac and Bose–Einstein distributions. If the system were homogeneous except for a uniform temperature gradient in the x-direction, say, $T = T_0 + T'x$ (T_0 and T' constants) and f were the Fermi–Dirac function you can probably work out what is $(\mathrm{d}/\mathrm{d}x)\; f_{\text{FD}}$ and this term could then be retained.

We now recall equation (3.3.9) for the current density from section 3.3.3,

$$\mathbf{J} = \frac{2}{(2\pi)^3} \int_{BZ} f(\mathbf{k})(-e\mathbf{v}) \mathrm{d}\mathbf{k} \tag{3.3.9}$$

and we consider the case of an isotropic metal so that $\mathbf{J}$ is parallel to $\mathbf{v}$, and we will take both to be in the x-direction. We use equation (3.3.8)

$$v_x = \frac{1}{\hbar} \frac{\partial E}{\partial k_x}$$

and

$$\frac{\partial f_{\mathrm{FD}}}{\partial k_x} = \frac{\partial E}{\partial k_x} \frac{\partial f_{\mathrm{FD}}}{\partial E}$$

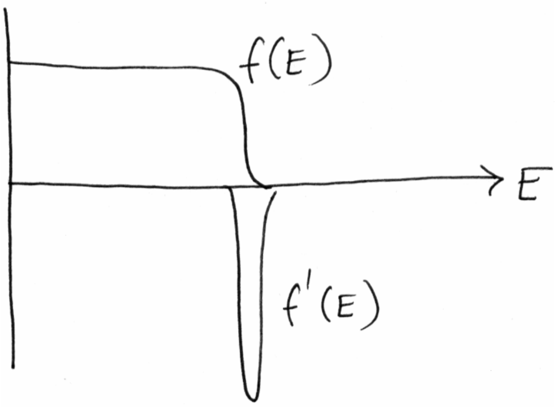
Noting that integrating the first term in (3.5.11) yields zero by symmetry, we now get

$$J_x = -\frac{2}{(2\pi)^3} e^2 \mathcal{E}_x \int_{BZ} \tau(\mathbf{k}) v_x^2(\mathbf{k}) \frac{\partial f_{\mathrm{FD}}}{\partial E} \mathrm{d}\mathbf{k}$$

In the limit of zero temperature

$$\frac{\partial f_{\mathrm{FD}}}{\partial E} = -\delta(E - E_{\mathrm{F}})$$

which is (minus) a "Dirac delta function."

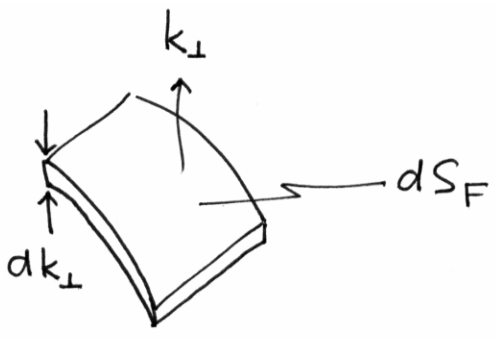


This is non zero only at the Fermi surface—only electrons within $k_{\mathrm{B}}T$ of E_{F} will contribute to the integral. The volume element $\mathrm{d}\mathbf{k}$ is therefore an element of Fermi surface with thickness

$$\mathrm{d}k_\perp = \frac{\mathrm{d}E}{|\boldsymbol{\nabla}_{\mathbf{k}} E|}$$

and so

$$\mathrm{d}\mathbf{k} = \mathrm{d}S_{\mathrm{F}}\, \mathrm{d}k_\perp = \mathrm{d}E \frac{1}{\hbar v(\mathbf{k})}\, \mathrm{d}S_{\mathrm{F}}$$



The term $\mathrm{d}E(\partial f/\partial E)$ reduces the volume integral over the Brillouin zone to a surface integral over the Fermi surface. Hence the conductivity is

$$\begin{aligned}\sigma &= \frac{J_x}{\mathcal{E}} \\ &= \frac{2}{(2\pi)^3}\oint_{E=E_\mathrm{F}} \frac{e^2 v_x^2}{v(\mathbf{k})}\tau(\mathbf{k})\mathrm{d}S_\mathrm{F} \qquad (3.5.12)\end{aligned}$$

In the special case of a parabolic, isotropic, free electron like metal, this will reduce back to the Drude formula! Because in this instance, we have

$$v_x^2(k_\mathrm{F}) = \frac{1}{3}v^2(k_\mathrm{F})$$

and

$$v(k_\mathrm{F}) \equiv v_\mathrm{F} = \frac{\hbar k_\mathrm{F}}{m^*}$$

and the integral

$$\oint_{E=E_\mathrm{F}} \mathrm{d}S_\mathrm{F} = 4\pi k_\mathrm{F}^2$$

is the area of the Fermi sphere. But according to equation (2.2.2) the Fermi wavevector *depends on the* <u>*total*</u> *number of electrons per unit volume*,

$$\frac{N_\mathrm{e}}{V} = n_\mathrm{e} = \frac{k_\mathrm{F}^3}{3\pi^2} \qquad (2.2.2)$$

Putting all these back into equation (3.5.12) we get, for a parabolic, isotropic, free electron like metal,

$$\begin{aligned}\sigma &= \frac{e^2}{4\pi^3\hbar}\frac{1}{3}\frac{\hbar k_\mathrm{F}}{m^*}\tau(k_\mathrm{F})4\pi k_\mathrm{F}^2 \\ &= \frac{e^2\tau(k_\mathrm{F})}{m^*}\frac{k_\mathrm{F}^3}{3\pi^2} \\ &= \frac{n_\mathrm{e}e^2\tau(k_\mathrm{F})}{m^*}\end{aligned}$$

Only electrons at the Fermi surface contribute; not all of them as in the classical picture. Hence the relevant relaxation time is that of the electrons having energy E_F. But the <u>total</u> electron density n_e appears as in the classical Drude formula because it determines the radius k_F of the Fermi sphere.

3.6 The Wiedemann–Franz Law

We now have formulas for the electrical and thermal conductivities of the quantum free electron gas in the relaxation time approximation.

$$\boxed{\begin{aligned} \sigma &= \frac{n_\mathrm{e} e^2 \tau}{m} = \frac{n_\mathrm{e} e^2 \lambda}{mv} \\ \kappa &= \frac{1}{3} c_\mathrm{e} v \lambda \end{aligned}}$$

We take it that the mean free paths are the same for both so they cancel when we take the ratio (but see Hook and Hall, section 3.3.4)

$$\frac{\kappa}{\sigma} = \frac{1}{3} \frac{c_\mathrm{e} m v^2}{n_\mathrm{e} e^2} = \frac{1}{3} \frac{1}{e^2} \frac{c_\mathrm{e}}{n_\mathrm{e}} m v^2$$

However as we saw in section 2.2

$$\frac{c_\mathrm{e}}{n_\mathrm{e}} = \frac{1}{2} \pi^2 k_\mathrm{B}^2 \frac{T}{E_\mathrm{F}} \tag{2.2.8}$$

and we can identify v as the Fermi velocity, as in equation (2.2.7),

$$\frac{1}{2} m v^2 = E_\mathrm{F}$$

Hence

$$\frac{\kappa}{\sigma T} = \frac{1}{3} \pi^2 \left(\frac{k_\mathrm{B}}{e} \right)^2 = 2.43 \times 10^{-8} \ \mathrm{W \Omega K^{-2}}$$

depends only on fundamental constants. This relation between σ and κ is called the *Wiedemann–Franz Law* and has been known from observation for well over 100 years.

Interestingly this ratio is nearly the same in classical physics which severely overestimates the heat capacity but underestimates the velocity. Using the wrong results

$$c_\mathrm{e} \neq \frac{3}{2} n_\mathrm{e} k_\mathrm{B}$$

and

$$\frac{1}{2} m v^2 \neq \frac{3}{2} k_\mathrm{B} T$$

you get

$$\frac{\kappa}{\sigma T} = \frac{3}{2} \left(\frac{k_\mathrm{B}}{e} \right)^2 = 1.11 \times 10^{-8} \ \mathrm{W \Omega K^{-2}}$$

Typical measured values at 0°C are

	$\kappa/\sigma T$ ($\mathrm{W\Omega K^{-2}}$)
Cu	2.22×10^{-8}
Al	2.14×10^{-8}

It is not surprising that people believed the classical point of view of Drude before Sommerfeld's explanation although of course the measured heat capacity can only be explained from the Pauli principle.

3.7 Plausibility argument for the linearised Boltzmann equation

We seek the probability of occupation of a wavepacket of (central) wavevector $\mathbf{k}$ and position $\mathbf{r}$ (subject of course to the uncertainty principle) in a gas of independent fermions subject to the Pauli principle, and under the influence of a weak, uniform electric field $\boldsymbol{\mathcal{E}}$.

We expect this to have the form

$$f(\mathbf{r},\mathbf{k}) = f_{\mathrm{FD}}(\mathbf{k}) + \text{corrections}$$

Now we make two observations about these corrections.

1. In a degenerate electron gas only particles near the Fermi surface contribute to transport. So there must be a factor that is zero above and below E_{F}, for example a gaussian centred at E_{F}. Even better, use (see the figure on page 114)

$$-\frac{\partial f_{\mathrm{FD}}}{\partial E}$$

2. Electrons are negatively charged so they will be accelerated against the field: those travelling with the field will be slowed down and those travelling against the field will speed up. The wavepacket having wavevector $\mathbf{k}$ and velocity $\mathbf{v}(\mathbf{k})$ will have its occupation number changed according to whether $\mathbf{v}(\mathbf{k})$ is parallel or antiparallel to $\boldsymbol{\mathcal{E}}$; so we expect the corrections to contain a factor

$$-e\mathbf{v}(\mathbf{k})\cdot\boldsymbol{\mathcal{E}}$$

We surmise therefore that

$$\text{corrections} \propto \left(-\frac{\partial f_{\mathrm{FD}}}{\partial E}\right)\left(-e\mathbf{v}(\mathbf{k})\cdot\boldsymbol{\mathcal{E}}\right)$$

Since probability is a <u>number</u> this must be dimensionless, so we need to multiply by something with the dimension of time. Call this $\tau(\mathbf{k})$. Then we get

$$f(\mathbf{r},\mathbf{k}) = f_{\mathrm{FD}}(\mathbf{k}) + \tau(\mathbf{k})\frac{\partial f_{\mathrm{FD}}}{\partial E}\, e\mathbf{v}(\mathbf{k})\cdot\boldsymbol{\mathcal{E}}$$

Now, using the chain rule we have,

$$\frac{\partial f_{\rm FD}}{\partial E} = \frac{\partial f_{\rm FD}}{\partial \mathbf{k}} \times \frac{\partial \mathbf{k}}{\partial E} = \boldsymbol{\nabla}_{\mathbf{k}} f_{\rm FD} \times \left(\boldsymbol{\nabla}_{\mathbf{k}} E\right)^{-1}$$

and remembering equation (3.3.8), $\mathbf{v}(\mathbf{k}) = \hbar^{-1}\boldsymbol{\nabla}_{\mathbf{k}} E$, we can "cancel out" the velocity $\mathbf{v}(\mathbf{k})$ to obtain the linearised Boltzmann equation

$$f = f_{\rm FD} + \frac{e}{\hbar}\tau(\mathbf{k})\boldsymbol{\mathcal{E}} \cdot \boldsymbol{\nabla}_{\mathbf{k}} f_{\rm FD}$$

and $\tau(\mathbf{k})$ is the relaxation time.

3.8 Thermoelectric effects

You have spotted that if thermal and electrical conductivity are both mediated by electrons then a temperature gradient may induce an electric current and an electric field may produce a flow of heat. The linearized Boltzmann equation always predicts that the flux of heat or current is *linearly* proportional to the field or temperature gradient. Hence if both gradients are present the proportionality must be expressible as two equations,

$$\mathbf{J}_{\rm e} = L_{\mathcal{E}\mathcal{E}}\boldsymbol{\mathcal{E}} + L_{\mathcal{E}T}\boldsymbol{\nabla} T \tag{3.8.1}$$

$$\mathbf{J}_{\rm Q} = L_{T\mathcal{E}}\boldsymbol{\mathcal{E}} + L_{TT}\boldsymbol{\nabla} T \tag{3.8.2}$$

and we will assume, for simplicity, that the fluxes and fields are parallel as in cubic symmetry crystals. We can then take all quantities as scalars, pointing, say, along the x-direction.

Suppose we measure electrical conductivity while holding the specimen at a constant temperature. Then

$$\boldsymbol{\nabla} T \longrightarrow \frac{{\rm d}T}{{\rm d}x} = 0$$

and by Ohm's law

$$J_{\rm e} = \sigma\mathcal{E}$$

which identifies $L_{\mathcal{E}\mathcal{E}}$ as the electrical conductivity,

$$L_{\mathcal{E}\mathcal{E}} = \sigma \tag{3.8.3}$$

Now suppose we measure thermal conductivity while preventing any current by insulating the specimen electrically. We measure

$$J_{\rm Q} = -\kappa\frac{{\rm d}T}{{\rm d}x}$$

but we cannot identify κ with minus the coefficient L_{TT}. Instead we must write the zero current condition as, from (3.8.1),

$$0 = L_{\mathcal{E}\mathcal{E}}\mathcal{E} + L_{\mathcal{E}T}\frac{{\rm d}T}{{\rm d}x}$$

and hence

$$\mathcal{E} = -\frac{L_{\mathcal{E}T}}{L_{\mathcal{E}\mathcal{E}}}\frac{\mathrm{d}T}{\mathrm{d}x}$$

The heat flow generates an electric field inside the metal. If we now put this result into (3.8.2) we have

$$J_{\mathrm{Q}} = -L_{\mathcal{E}\mathcal{E}}\frac{L_{\mathcal{E}T}}{L_{\mathcal{E}\mathcal{E}}}\frac{\mathrm{d}T}{\mathrm{d}x} + L_{TT}\frac{\mathrm{d}T}{\mathrm{d}x}$$

which identifies the thermal conductivity as

$$\kappa = -L_{TT} + \frac{L_{T\mathcal{E}}L_{\mathcal{E}T}}{L_{\mathcal{E}\mathcal{E}}} \tag{3.8.4}$$

What's happening here is that an electric field is generated to prevent the flow of current and this field in turn opposes the carriers of heat—reducing the thermal conductivity from our expected value given by $-L_{TT}$.

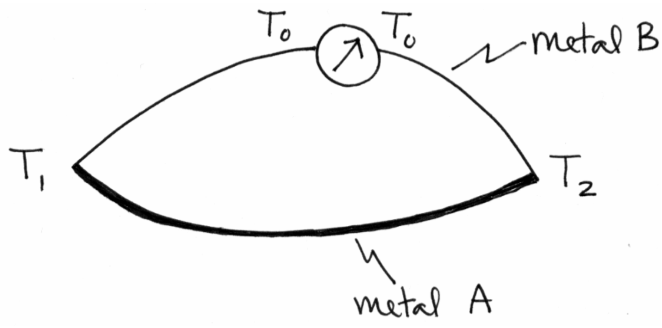
Apparently then a metal in a uniform temperature gradient develops an electric field of amount

$$\mathcal{E} = \mathcal{S}\frac{\mathrm{d}T}{\mathrm{d}x} \tag{3.8.5}$$

where

$$\mathcal{S} = -\frac{L_{\mathcal{E}T}}{L_{\mathcal{E}\mathcal{E}}} \tag{3.8.6}$$

is called the <u>absolute thermoelectric power</u> or <u>thermopower</u>. Suppose we build a circuit like this.



The junctions between the two metals A and B are kept at temperatures T_1 and T_2. A voltmeter is inserted in metal B which is found to be at some intermediate temperature T_0. By elementary electrostatics the voltage measured is given by the line integral of the electric field around the circuit, and using (3.8.5)

$$\begin{aligned}
V &= -\oint \mathcal{E}\mathrm{d}x = -\oint \mathcal{S}\frac{\mathrm{d}T}{\mathrm{d}x}\mathrm{d}x \\
&= \int_{T_0}^{T_1} \mathcal{S}_B \mathrm{d}T + \int_{T_1}^{T_2} \mathcal{S}_A \mathrm{d}T + \int_{T_2}^{T_0} \mathcal{S}_B \mathrm{d}T \\
&= \int_{T_1}^{T_2} (\mathcal{S}_A - \mathcal{S}_B)\mathrm{d}T \\
&= \int_{T_1}^{T_2} \mathcal{S}_A(T)\mathrm{d}T - \int_{T_1}^{T_2} \mathcal{S}_B(T)\mathrm{d}T
\end{aligned}$$

and in the last line I've indicated explicitly that the thermopower is temperature dependent. This experiment then measures the difference in thermopower between metals A and B integrated between the two temperatures T_1 and T_2.

Similar to electrostatic potential, we can define the absolute thermoelectric force of a metal as

$$\Theta(T) = \int_0^T \mathcal{S}(T)\mathrm{d}T$$

The phenomenon we have described is called the Seebeck effect and is the basis upon which a thermocouple is used in the measurement of temperature.

3.8.1 Estimate of the thermopower in metals and semiconductors

If a temperature gradient sets up an electric field we have, in one dimension, equation (3.8.5),

$$\mathcal{E} = \mathcal{S}\frac{\mathrm{d}T}{\mathrm{d}x} \tag{3.8.5}$$

The drift velocity of electrons due to the temperature gradient is obtained using just the same argument as we used to obtain the thermal conductivity in section 3.4.4, (see also Ashcroft and Mermin, p 24)

$$\begin{aligned}
v_{\mathrm{Q}} &= \frac{1}{2}\left[v(x - v\tau) - v(x + v\tau)\right] \qquad (3.8.7) \\
&= -\tau v \frac{\mathrm{d}v}{\mathrm{d}x} \\
&= -\tau \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}v^2\right)
\end{aligned}$$

and in three dimensions we get the factor $\frac{1}{3}$ as in equation (3.4.6) so, using the chain rule,

$$v_{\mathrm{Q}} = -\frac{1}{6}\tau \frac{\mathrm{d}(v^2)}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}x} \tag{3.8.8}$$

The drift velocity due to the field is (see section 3.2.2)

$$v_{\mathrm{e}} = -\frac{e\mathcal{E}\tau}{m} \tag{3.8.9}$$

and the *nett* current is zero so writing $v_{\mathrm{Q}}+v_{\mathrm{e}} = 0$ using (3.8.8) and (3.8.9) and substituting in (3.8.5) we must have

$$\begin{aligned}
\mathcal{S} &= -\frac{1}{3e}\frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{1}{2}mv^2\right) \\
&= -\frac{C_{\mathrm{e}}}{3N_{\mathrm{e}}e} \\
&= -\frac{\pi^2}{6}\frac{k_{\mathrm{B}}}{e}\frac{T}{T_{\mathrm{F}}} \\
&\approx -0.003\frac{k_{\mathrm{B}}}{e} \qquad (3.8.10a) \\
&\approx -0.3\ \mu\mathrm{V\ K}^{-1} \qquad (3.8.10b)
\end{aligned}$$

In the third line I have used equation (2.2.8), taking $\frac{1}{2}mv^2$ to be the internal energy per electron of the electron gas so that its temperature derivative is the heat capacity per electron. Note that the velocities v_{Q} and v_{e} are the slow *drift* velocities here while v in (3.8.7) is the fast Fermi velocity in metals and thermal velocity in semiconductors. You can see that in metals the thermopower increases linearly with temperature; taking $T/T_{\mathrm{F}} = 0.002$ for Aluminium (section 2.2) the thermopower of a metal is seen to be rather small and negative.

Equation (3.8.10*b*) furnishes us with an estimate of the thermopower of a metal. You will find in a semiconductor textbook that the thermopower is

$$\begin{aligned}
\mathcal{S} &= \frac{k_{\mathrm{B}}}{e}\left[2 + \frac{E_c - E_{\mathrm{F}}}{k_{\mathrm{B}}T}\right] \\
&\approx (2+22)\frac{k_{\mathrm{B}}}{e} \qquad (3.8.11)
\end{aligned}$$

for n-type and

$$\mathcal{S} = \frac{k_{\mathrm{B}}}{e}\left[2 - \frac{E_{\mathrm{F}} - E_v}{k_{\mathrm{B}}T}\right]$$

for p-type semiconductors: a "metallic part" plus a much larger part due to the effect of temperature on Fermi level and carrier density. Note that the thermopower is *negative* for positive carriers (holes) so like the Hall effect the thermopower tells what type of carrier is the dominant one. The estimate in (3.8.11) is based on the band gap of Si being 1.1 eV and the fact that $k_{\mathrm{B}}T$ at room temperature is about $\frac{1}{40}$ eV. Comparing (3.8.11) with (3.8.10*a*), the thermopower in a semiconductor or insulator is typically many thousands of times greater than in a metal. It also has the opposite sign in the case of electron carriers.

3.8.2 The Peltier effect

Can I reverse the Seebeck effect? That is, drive a current around the same circuit and obtain a temperature difference between the two junctions. This would act as a useful fridge or heater. Yes, I can; and I get the Peltier effect. When we considered measuring the electrical conductivity we kept the metal at uniform temperature, so that

$$\frac{\mathrm{d}T}{\mathrm{d}x} = 0$$

but according to equation (3.8.2) there will *still* be a thermal current due to the electric field that I apply to measure σ,

$$\begin{aligned} J_{\mathrm{Q}} &= L_{T\mathcal{E}}\mathcal{E} \\ &= \frac{L_{T\mathcal{E}}}{L_{\mathcal{E}\mathcal{E}}}\, J_{\mathrm{e}} \\ &= \mathit{\Pi}_{AB} J_{\mathrm{e}} \qquad (\mathit{\Pi} = T\mathcal{S}\text{—proof below}) \end{aligned}$$

$\mathit{\Pi}_{AB}$ is called the Peltier coefficient and is the heat evolved per unit area of metal A / metal B junction per unit current.

This is the way thermoelectric heaters and coolers work; a device can be switched from heating to cooling simply by reversing the current. The cooler that you plug into your car's cigar lighter can also heat your dinner.

People look for the best thermoelectric materials. A good thermoelectric has a large value of the figure of merit ZT near the temperature of operation, T,

$$ZT = \frac{\mathcal{S}^2\sigma}{\kappa}T$$

where σ and κ are electrical and thermal conductivities and $\mathcal{S}$ is the thermopower.

Bi_2Te_3 has $ZT \approx 1$. Some semiconductors are good candidates, especially those having a large atomic number and a band gap of about $10k_{\mathrm{B}}T$. This implies a small band gap; heavy doping may help. Metals usually have too small $\mathcal{S}$ to be useful.

No really good material has yet been found, say, to replace the conventional refrigerator. But Peltier cooling is used all the time in laboratories, small domestic fridges, and so on.

3.9 Onsager relations

We have four coefficients in (3.8.1) and (3.8.2), $L_{\mathcal{E}\mathcal{E}}$, L_{TT}, $L_{\mathcal{E}T}$, $L_{T\mathcal{E}}$, and have found their relations to four measurable properties, σ and κ, the electric and thermal conductivities; $\mathcal{S}$ and $\mathit{\Pi}$ the thermopower and Peltier coefficient. Now we find a relation between the two coefficients, $L_{\mathcal{E}T}$ and $L_{T\mathcal{E}}$ to expose a relation between $\mathcal{S}$ and $\mathit{\Pi}$. As a by-product you will get a glimpse into the elegant and fascinating subject of irreversible thermodynamics.

Equations (3.8.1) and (3.8.2) are examples of formulas that describe the thermodynamic steady state where forces $X_1, X_2, \ldots, X_n$ generate currents $J_1, J_2, \ldots, J_n$. They are related *linearly* by

$$J_i = \sum_j L_{ij} X_j \tag{3.9.1}$$

According to Onsager's theorem (which we shall not prove here) if the forces and currents have been defined such that

$$\dot{S} = \sum_i X_i J_i \tag{3.9.2}$$

is the rate of entropy production per unit volume per unit time, then we have

$$L_{ji} = L_{ij}$$

To bring them into line with equation (3.9.2) we must make slight adjustments to our definitions of X and J in equations (3.8.1) and (3.8.2).

In the case that *only electric current* is flowing, it is easy to find $\dot{S}$—it is the rate of Joule heating $J_\mathrm{e}\mathcal{E}$ divided by the temperature. So if J_1 is the electric current density, the "conjugate" force must be

$$X_1 = \frac{\mathcal{E}}{T}$$

so that

$$\dot{S} = \frac{J_\mathrm{e}\mathcal{E}}{T} = J_1 X_1 \qquad \text{electric current flowing}$$

as required by Onsager's theorem.

In the case of the heat current we write the rate of entropy production as arising from the creation of heat as described mathematically as the divergence of J_Q/T. That is,

$$\dot{S} = \boldsymbol{\nabla} \cdot \left(\frac{\mathbf{J}_\mathrm{Q}}{T} \right)$$

If $\mathbf{J}_\mathrm{Q}$ is constant as in the steady state we have

$$\dot{S} = \mathbf{J}_\mathrm{Q} \cdot \boldsymbol{\nabla} \frac{1}{T}$$

and in one dimension,

$$\dot{S} = J_\mathrm{Q} \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{1}{T} \right) = J_\mathrm{Q} \left(-\frac{1}{T^2} \right) \frac{\mathrm{d}T}{\mathrm{d}x}$$

So if we identify J_Q with the flux J_2 the conjugate force X_2 must be

$$\mathbf{X}_2 = \boldsymbol{\nabla} \frac{1}{T}$$

or in one dimension,

$$X_2 = -\frac{1}{T^2} \frac{\mathrm{d}T}{\mathrm{d}x}$$

so that

$$\dot{S} = J_2 X_2 \qquad \text{heat current flowing}$$

as required by Onsager's theorem.

We can now rewrite equations (3.8.1) and (3.8.2) in the canonical form of equation (3.9.1),

$$\begin{aligned} J_\mathrm{e} &= L_{\mathcal{E}\mathcal{E}} T \left(\frac{\mathcal{E}}{T}\right) + L_{\mathcal{E}T} T^2 \left(\frac{1}{T^2}\right) \frac{\mathrm{d}T}{\mathrm{d}x} \\ J_\mathrm{Q} &= L_{T\mathcal{E}} T \left(\frac{\mathcal{E}}{T}\right) + L_{TT} T^2 \left(\frac{1}{T^2}\right) \frac{\mathrm{d}T}{\mathrm{d}x} \end{aligned}$$

or equivalently

$$\begin{aligned} J_1 &= L_{11} X_1 + L_{12} X_2 \\ J_2 &= L_{21} X_1 + L_{22} X_2 \end{aligned}$$

Then by Onsager's theorem we have $L_{12} = L_{21}$ or

$$-L_{\mathcal{E}T} T^2 = L_{T\mathcal{E}} T$$

which furnishes us with the relation between the thermopower and the Peltier coefficient which I promised,

$$\mathcal{S} = \frac{\mathit{\Pi}}{T}$$

a result first obtain by William Thomson, later Lord Kelvin, in 1854.

Suppose we have a wire subject to both an electric field and a thermal gradient. If J_e is constant along the wire, then at any point the rate of entropy production is

$$\dot{S} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{J_\mathrm{Q}}{T}\right) + J_\mathrm{e} \frac{\mathcal{E}}{T}$$

I will leave out the algebra, but make $\mathcal{E}$ the subject of equation (3.8.1) and substitite the result into (3.8.2). Use equations (3.8.3), (3.8.4) and (3.8.6) to eliminate the L's, differentiate the product (J_Q/T) with respect to x and use the chain rule to write $\mathrm{d}\mathcal{S}/\mathrm{d}x$ in terms of $\mathrm{d}\mathcal{S}/\mathrm{d}T$ and $\mathrm{d}T/\mathrm{d}x$ and you'll find,

$$\dot{S} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\kappa}{T}\frac{\mathrm{d}T}{\mathrm{d}x}\right) + \frac{J_\mathrm{e}^2}{\sigma T} + J_\mathrm{e}\frac{1}{T}\frac{\mathrm{d}T}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}T}(\mathcal{S}T)$$

The first two terms are, respectively, the entropy production generated by heat flow and Joule heating. The third term is proportional to both the electric and thermal currents and is positive if they are in the same direction and negative if they are opposed. The rate of heat generation due to this effect is

$$T\dot{S} = J_\mathrm{e}\frac{\mathrm{d}T}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}T}(\mathcal{S}T)$$

and can be measured directly in the circuit by measuring the change in heat output when the current is reversed. This is called the Thomson effect (also after Lord Kelvin).

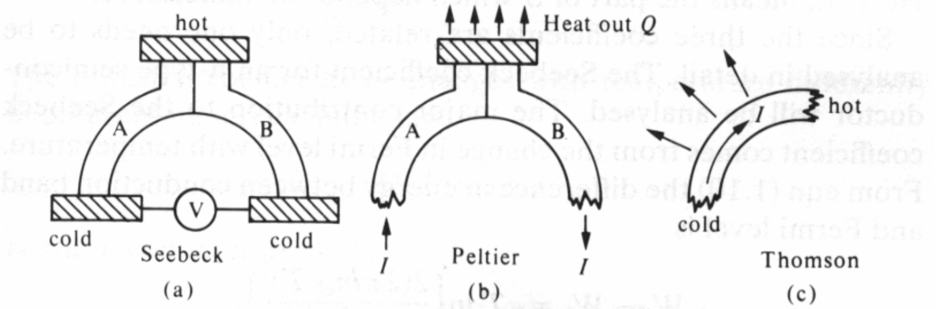


Fig. 2.2. There are three ways in which thermoelectric effects can be demonstrated: (a) the Seebeck effect—a temperature difference across two materials causes an output voltage; (b) the Peltier effect—a current causes heat flow into or out of a junction; (c) the Thomson effect—a current along a temperature gradient causes heat flow into or out of the conductor.

It is important to distinguish the Joule heating from the Thomson heating. Joule heat is given out independently of the direction of the current and is thermodynamically *irreversible*. Conversely the Thomson heating (3.9.5) is reversible and is positive or negative, that is absorbed or emitted according to the direction of the current (the sign of J_{e}). Since this only depends on the thermopower it provides a means for its measurement. Thermopower is central to many investigations in solid state physics as it furnishes direct evidence of the nature of the current carriers.