

NOTES 9: FERMI-SOMMERFELD MODEL OF FREE ELECTRONS

Arguably the most important phenomenon in all of electronics is the natural presence of “free” electrons in solids, particularly metals. “Free” means that they are not bound to any particular atom or defect but, instead, move about the *entire* solid under the random (Brownian) motion of thermal equilibrium, and the deterministic motion under any applied electric or magnetic fields. This mode was put forth shortly after the discovery of the electron in early part of the 20th century to explain the amazing variation in electrical properties between common materials, particularly between dielectrics and metals. It was believed, but not proven rigorously till later using *energy band* principles, that dielectrics have no “free” electrons but metals have very many. In fact, great progress was made in explaining certain metallic properties, such as the electrical conductivity of solids, simply by assuming that there was an integral number of “free” electrons per atom in the solid. The mechanical response of the electrons to an electric field could then be described similar to the response in vacuum, but with the addition of a strong scattering mechanism. This so-called Drude model will be one of the first subjects of transport theory later on.

But similar to the case of phonons, there were discrepancies between theory and experiment in describing the *thermal* properties. Specifically, the heat capacity of common metals was found to be significantly lower at room temperature and below than predicted by classical mechanics combined with Boltzmann probability. The solution, first obtained by A. Sommerfeld and E. Fermi, was to treat both the mechanics and statistical mechanics using quantum principles. As described earlier, this entailed two giant conceptual steps: (1) that the electron density was high enough to make individual electrons indistinguishable by the de-Broglie criterion, and (2) that the fundamental identity of the electrons was subject to the Pauli exclusion principle, requiring the use of Fermi-Dirac statistics as a special case of the Boltzmann probability.

Quantum mechanics

Although the electrons are free to move about the solid, they are rather abruptly confined to the boundaries as shown in the one-dim sketch in Fig. 1(a) along the x axis. A similar drawing can be made along the y and z axes. The solid line shows the potential-energy distribution along one dimension of the “box”. Consistent with common experience, electrons usually do not escape from the solid so there is a binding (potential) energy U_B . Unlike the “particle-in-a-box” problem wherein $U_B \rightarrow \infty$, the binding energy U_B in this model is finite, consistent the longstanding experimental observations that external forces such as static electric fields or light can cause sudden emission of electrons from the surface in a threshold fashion. For the case of electric fields, the threshold is stated in terms of the strength of the electric field. For radiation it is stated in terms of the photon energy $h\nu$, and is an even sharper function. This is the famous *photoelectric effect* – the first direct evidence for the corpuscular nature of electromagnetic radiation, i.e., photons.¹

¹ Interestingly, it was also the citation for A. Einstein’s Nobel prize in 1917

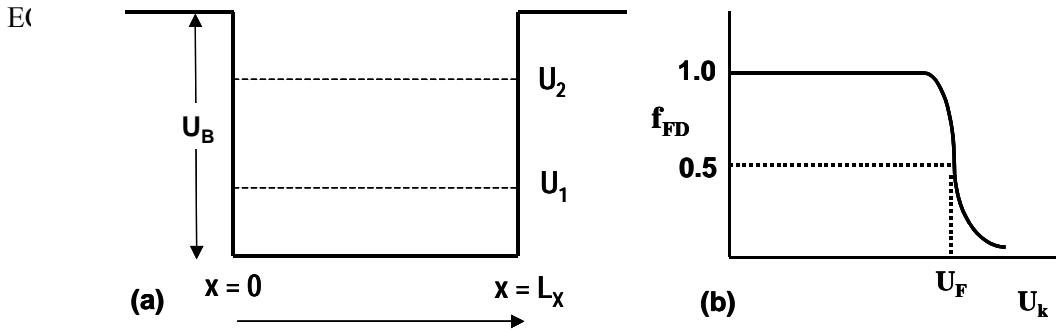


Fig. 1. (a) Energy diagram of electron in one dimension. (b) Fermi-Dirac function in the limit of low temperature where $U_F > 0$

To account for the possibility of a metal, the electron density must be assumed high enough that the electrons are spatially indistinguishable by the de-Broglie criterion. Hence quantum mechanics must be used to solve for the energy states and related effects through the solution to Schrodinger's equation

$$\frac{\hbar^2}{2m} \nabla^2 \psi + V\psi = U\psi \quad \text{or} \quad \nabla^2 \psi + \frac{2m}{\hbar^2} U' \psi = 0 \quad \text{or} \quad \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{2m}{\hbar^2} U' \psi = 0$$

where $U' = U - V$. It is simple to show by the method of separation of variables that this can be reduced to three coupled one-dim differential equations of the form

$$\frac{d^2 \psi}{dq^2} + k_q^2 \psi = 0$$

where $q = x, y, \text{ or } z$ and k is a constant. For each q this is a 2nd-order linear differential equation with constant coefficients for which we know the general solution to be

$$\psi = A e^{jk_q x} + B e^{-jk_q x}$$

where A and B are arbitrary constants. Unlike the "particle-in-a-box" solution given in Chapter 1 where the boundary conditions required ψ to vanish at the walls of the box and everywhere outside, here we allow for the possibility that the box is barely confining and thus ψ does not vanish at the walls. Instead, we take advantage of the fact that the interface can be assumed physically identical at each wall, so that the wave function should be periodic: $\psi(x = 0) = \psi(x = L_x)$; $\psi(y = 0) = \psi(y = L_y)$; $\psi(z = 0) = \psi(z = L_z)$. Under these conditions, we get the condition

$$\psi(q = 0) = A + B = \psi(x = L_q) = A e^{jk_q L_q} + B e^{-jk_q L_q} \quad \text{which can be true if and only if}$$

$$e^{jk_q L_q} = e^{-jk_q L_q} = 1 \quad . \quad \text{In turn, this requires that} \quad k_q = 0, \pm \frac{2\pi}{L_q}, \pm \frac{4\pi}{L_q}, \dots$$

$$U_{|\mathbf{k}|} = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = \frac{(\hbar 2m\pi/L_x)^2}{2m} + \frac{(\hbar 2n\pi/L_y)^2}{2m} + \frac{(\hbar 2o\pi/L_z)^2}{2m}$$

where $m, n, \text{ and } o$ are integers. The allowed values of $|\mathbf{k}|$ have the same form as for lattice waves with one important difference – the maximum relevant value is no longer limited by a Nyquist criterion to π/d as before, where d is the atomic interplanar

separation. With the “box” of electrons, the values of k_x , k_y , or k_z can get arbitrarily large.

At this point many students ask the good questions, how can such a simple model describe such a complex system as a large population of “free” electrons mixed with the dense atomic distribution of a real solid? And what happened to the potential energy associated with (Coulombic) interaction between the electrons, and between the electrons and the atoms? The Fermi model makes it possible to describe such complicated behavior by one simple augmentation, the introduction of the effective mass m^* such that the energy states are described by

$$U_k = \frac{\hbar^2 k^2}{2m^*} \quad (1)$$

In other words, the effect of the complicated interactions is embedded in m^* . The proof of this simple but profound expression for metals was not made until the 1950s by Landau and others. Surprisingly, it was proven to apply even earlier to a revolutionary class of solids known as semiconductors. This will be addressed with the discussion of electron *band theory*.

Aside on Quasiparticles and Collective Excitations

A common description of an electron (or any particle for that matter) that can be described by (1) is *quasiparticle*. By definition, a quasiparticle is a fundamental particle interacting with other particles (or forces) such that its “free” particle behavior is essentially maintained and the interaction is added as a “dressing” to the mass. As such, if the interactions are virtually “turned off”, the quasiparticle “returns” to the real fundamental particle. This definition is convenient for distinguishing quasiparticles from other multiparticle phenomena such as the Cooper pairs of (low- T_C) superconductivity. The Cooper pairs involve two electrons with opposite spin that can bind because of a peculiar interaction with the phonons. When the phonon interaction is virtually “turned off”, what remains is two electrons, not one. So the Cooper pair is not a quasiparticle.

By the same token, phonons are not quasiparticles either because they do not “return” to one particle when the atomic interaction is “turned off”. But because they represent one of a large set of independent harmonic oscillators in which each atom plays the same fundamental role, they are often called *collective excitations*. This label probably would not have become popular if not for the discovery in the middle 20th century of a multitude of interactions between collections of particles that could be represented as independent harmonic oscillators. Two other examples are the plasmons, discussed later in this chapter, and the magnons discussed in the chapter on ferromagnetism. The plasmons are the collective (Coulombic) interaction of “free” electrons with the background lattice, and the magnons are the collective (Magnetic) interaction of “fixed” spins with each other.

Statistical Mechanics

Since the spin of electrons is $\frac{1}{2}$, they must abide by the Pauli exclusion principle so their statistical behavior is described by the Fermi-Dirac distribution. We use the

wave vector $\mathbf{k} = k_x \mathbf{x} + k_y \mathbf{y} + k_z \mathbf{z}$, or its magnitude $k \equiv |\mathbf{k}|$ to signify each quantum state. And the mean number of electrons in each state is

$$\langle n_{\mathbf{k}} \rangle = \frac{g}{\exp(U_{\mathbf{k}} - \mu) / k_B T + 1} \equiv \frac{g}{\exp(U_{\mathbf{k}} - U_F) / k_B T + 1} = g \cdot f_{FD}$$

where μ is the chemical potential, U_F is the *temperature-dependent* Fermi energy, and f_{FD} is the Fermi-Dirac function, plotted in Fig. 1(b). The chemical potential is the customary quantity in physics and chemistry. The *temperature-dependent* Fermi energy U_F is the customary quantity in electrical engineering.² Both depend on temperature. The factor $U_{\mathbf{k}}$ is the spin degeneracy factor = 2 for spin 1/2. As for phonons $\langle n_{\mathbf{k}} \rangle$ is to be thought of as the mean number of electrons in the wave represented by \mathbf{k} , this time a particle (de-Broglie) wave rather than a lattice wave.

The mean total energy of a population of “free” electrons is given by:

$$\langle U \rangle = \sum_{\mathbf{k}} \langle n_{\mathbf{k}} \rangle U_{\mathbf{k}} \quad \text{where} \quad U_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m^*}$$

As with phonons, because $\langle n_{\mathbf{k}} \rangle$ is an explicit function of U , it is easiest to evaluate the sum as an integral over U :

$$\langle U \rangle = \int_0^\infty \langle n_{\mathbf{k}} \rangle U_{\mathbf{k}} D(U) dU = \int_0^\infty 2f(U) U D(U) dU$$

where $D(u)$ is the density of states, i.e., the number of \mathbf{k} states per unit energy interval.

We can calculate $D(U)$ in a similar fashion to phonons using the chain rule of differential calculus

$$D(U) dU = \frac{dN(\mathbf{k})}{dU} dU = \frac{dN(\mathbf{k})}{dk} \frac{dk}{dU} dU$$

where $N(\mathbf{k})$ is the number of states between $k = 0$ and k in “ \mathbf{k} space”. In a 3-dim solid,

$$N(k) = \left(\frac{L}{2\pi}\right)^3 \cdot \frac{4}{3} \pi k^3 = \frac{V}{6\pi^2} k^3$$

Dim	$N(\mathbf{k})$	$D(U)dU$	Description
3	$Vk^3/(6\pi^2)$	$[V(m^*)^{3/2}/(2)^{1/2} \pi^2 \hbar^3](U)^{1/2} dU$	Bulk
2	$Ak^2/(4\pi)$	$Am^*/(2\pi \hbar^2) dU$	Quantum Well
1	$Lk/(2\pi)$	$[L(m^*/8)^{1/2}/\hbar](U)^{-1/2} dU$	Quantum Wire

So,
$$\frac{dN(k)}{dk} = \frac{V}{2\pi^2} k^2$$

And from the energy expression:
$$U = \frac{\hbar^2 k^2}{2m^*}, \quad \frac{dU}{dk} = \frac{\hbar^2 k}{m^*} = \frac{dk}{dU} = \frac{m^*}{\hbar^2 k}$$

So,

² This is often a source of confusion because the physics and chemistry literature usually defines U_F as the low-temperature limit of μ . Here we will define that low temperature limit as U_{F0} to stay consistent with electrical engineering, particularly the plentiful literature on semiconductor devices.

$$D(U)dU = \frac{V}{2\pi^2} k^2 \left(\frac{m^*}{\hbar^2 k} \right) dU = \frac{m^* V dU}{2\pi^2 \hbar^2} \left(\frac{\sqrt{2m^*U}}{\hbar} \right) = \frac{V(m^*)^{3/2} \sqrt{U}}{\sqrt{2}\pi^2 \hbar^3} dU$$

Very similar expressions can be derived in one- and two-dimensions using the same expression for $U(k)$ but the modified expressions for $N(k)$ listed in Table I. The one- and two-dimensional expressions are very useful for free electrons confined to “quantum wires” and “quantum wells”, respectively. Note that in all cases the expressions are a factor of 2 lower than occur many solid-state physics books (e.g., Kittel) which combine the spin degeneracy into the density of states. Since the spin degeneracy is broken by the presence of a suitably large external magnetic field, it is best to keep $D(U)$ functions independent of spin, as we will see later in the topic of magnetism. So here we include the spin degeneracy in $\langle n_k \rangle$ and any other statistical averages, where it belongs.

Fermi Energy

In the application of the Fermi-Dirac distribution to “free” electrons in any dimension, it is important to realize that U_F is really the chemical potential – a macroscopic thermodynamic variable. In other words, it must be defined in some way at the *macroscopic* level independent of the quantum mechanics and statistics. The simplest definition follows from fixing the total number of electrons in the solid such to a value N_e . This is rather obvious in an isolated solid, but needs to be revised when dealing with solids in which the number of electrons changes under external influences.³ For fixed N_e we have the following *constraint*

$$\langle N_e \rangle = \int_0^\infty 2D(U) f_{FD}(U) dU, \tag{10}$$

a simple-looking definite integral but having the following subtle aspect. To be true at all temperatures, something inside the integral must change to counteract the effect of temperature on f_{FD} . Clearly, the only parameter that can adjust itself is U_F . So this equation becomes the implicit definition of N_e vs U_F .

Like the case of phonons, it is relatively easy to evaluate (10) in the limit of high and low temperature. The low-temperature case is facilitated by the fact that as $T \rightarrow 0$, $U_F > 0$ and thus f_{FD} behaves like a unit (i.e., Heaviside) step function θ as clear from Fig. 1(b):

$$f_{FD}(U) = \theta(U_F - U) \quad ; \quad U_{F0} = \lim_{T \rightarrow 0} \{U_F\}$$

In other words, U_F is the maximum energy an electron can have in the limit of zero temperature. Substitution of the step function yields into (10) yields

$$N_e = \int_0^{U_F} 2D(U) dU = \frac{\sqrt{2} V m^{3/2}}{\pi^2 \hbar^3} \frac{U^{3/2}}{3/2} \Bigg|_0^{U_{F0}} = \frac{V(2m^*U_{F0})^{3/2}}{3\pi^2 \hbar^3}$$

³ This occurs in semiconductor devices, and forces the introduction of the *quasi-Fermi* energy as described later in Chapter XX.

$$U_{F0} = \frac{[3\pi^2 \hbar^3 N_e/V]^{2/3}}{2m^*} = [3\pi^2 \rho]^{2/3} \frac{\hbar^2}{2m^*} \quad (11)$$

where ρ is the electron bulk density. Similar expressions can be derived for U_{F0} in the lower dimensions, which both are proportional to $\hbar^2/2m^*$ as shown in Table II.

Table II. Expressions for the low-temperature Fermi energy U_{F0} vs density of electrons		
Dimension	Density	U_{F0}
3	Bulk, ρ	$(3\pi^2 \rho)^{2/3} \cdot \hbar^2 / 2m^*$
2	Sheet, ρ_S	$2\pi \rho_S \cdot \hbar^2 / 2m^*$
1	Line, ρ_L	$(\rho_L)^2 \cdot (\pi \hbar)^2 / 2m^*$

The high –temperature limit of (10) is facilitated by the following approximation (proven in a homework problem),

$f_{FD} = [\exp(U-U_F)/k_B T + 1]^{-1} \approx [\exp(U-U_F)/k_B T]^{-1} = \exp(U_F/k_B T)\exp(-U/k_B T)$.
 The proof, carried out numerically on $F_{1/2}(U_F)$, rests on the fact in the high-T limit $U_F < 0$ but $|U_F| \gg k_B T$, so that $\exp(U_F/k_B T)$ is always greater than one, no matter what U is considered between 0 and ∞ . Hence

$$\langle N_e \rangle \approx 2e^{U_F/k_B T} \int_0^\infty D(U) e^{-U/k_B T} dU = \frac{(2)^{1/2} (m^* k_B T)^{3/2}}{\pi^2 \hbar^3} e^{U_F/k_B T} \int_0^\infty (x)^{1/2} e^{-x} dx \quad (12)$$

The latter integral is evaluated as a Gamma function by the following identity from applied mathematics

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$$

Two very useful evaluations of $\Gamma(n)$ are $\Gamma(3/2) = (\pi)^{1/2}/2$ and $\Gamma(5/2) = 3(\pi)^{1/2}/4$. Substitution of $\Gamma(3/2)$ into(12) results in

$$\frac{\langle N_e \rangle}{V} \approx \frac{e^{U_F/k_B T}}{(2)^{1/2}} \left(\frac{m^* k_B T}{\pi \cdot \hbar^2} \right)^{3/2} \quad (13)$$

For arbitrary temperatures we must resort to the generic form of (10)

$$\begin{aligned} \langle N_e \rangle &= \frac{2(m^*)^{3/2} V}{\sqrt{2} \pi^2 \hbar^3} \int_0^\infty f_{FD}(U) U^{1/2} dU = \frac{2(m^*)^{3/2} (k_B T)^{3/2} V}{\sqrt{2} \pi^2 \hbar^3} \int_0^\infty \frac{x^{1/2}}{\exp[x-x_F]+1} dx \\ &\equiv \frac{2(m^*)^{3/2} (k_B T)^{3/2}}{\sqrt{2} \pi^2 \hbar^3} F_{1/2}(x_F) \end{aligned}$$

where $x \equiv U/k_B T$ and $x_F \equiv U_F/k_B T$. $F_{1/2}(x_F)$ is one of a class of important quantities in solid-state theory called Fermi integrals.

$$F_y(x_F) = \int_0^\infty \frac{x^y}{\exp[x-x_F]+1} dx$$

the most useful ones being $F_{1/2}$ and $F_{3/2}$. Unfortunately, these are some of the most difficult integrals in mathematical physics to approximate over a wide range of temperature. So several generations of scientists and engineers have used look-up tables. As we shall see, the widespread availability of computational tools such as Matlab, make this a relatively easy exercise by numerical integration. The results from one integration routine are the subject of a homework problem.

Other Useful Metrics for the Fermi Gas

By definition U_{F0} is the maximum kinetic energy of an electron in the limit of “low” temperature. But how low is “low”? A simple answer to this question is found by defining a new quantity, the Fermi temperature T_F , by $U_{F0} \equiv k_B T_F$. A related question is: how much kinetic energy do the most energetic electrons have in the Fermi gas? The answer to this is found through the definition $U_{F0} = (1/2)m^*(v_F)^2$ where v_F is the Fermi velocity. Both of these quantities can be computed quickly from (11) once ρ is known for a given Fermi gas. The values for several common metals are listed in Table III. The density scales rather closely with the number of “free” electrons donated per atom in the solid. But U_{F0} , v_F , and T_F vary more slowly with ρ , consistent with a sublinear functional dependence.

Table III. Fermi-gas parameters of various common metals					
Material	#free electrons per atom	ρ [$\times 10^{22}$ cm ⁻³]	U_{F0} [eV]	v_F [$\times 10^8$ cm/s]	T_F [$\times 10^4$ K]
Cu	1	8.4	7.0	1.6	8.1
Au	1	5.9	5.5	1.4	6.4
Be	2	24.2	14.1	2.2	16.4
Zn	2	13.1	9.4	1.8	10.9
Al	3	18.1	11.6	2.0	13.5
Ga	3	15.3	10.3	1.9	12.0

Physically, the quantities T_F and v_F are to be used with some caution because they are not thermodynamic nor statistical quantities. Rather, T_F represents the temperature much below which the Fermi gas must be described by the Fermi-Dirac distribution rather than simple Boltzmann statistics. The Fermi gas in this case is often called “degenerate”, particularly in the context of semiconductor physics. The opposite limit of $T \gg T_F$ is often called “non-degenerate.”

The Fermi velocity is, indeed, a physically meaningful velocity but only for electrons having kinetic energies at or near U_{F0} . From Table III it is clear that this is indeed a very high velocity, for example in gold and copper $v_F/c \sim 0.5\%$ where c is the speed of light in vacuum. But as we shall see later in addressing electron transport, the Fermi gas is still random in the sense that for every electron moving in a given direction at v_F there is probably an electron moving in the opposite direction at v_F . So the effect of these “speedy” electrons on the overall transport is not as great as one might first assume.

A third quantity of importance to ac electrical and optical response is the Fermi wave vector, defined simply by equivalent momentum relations $m^*v_F = \hbar k_F$. In essence $2\pi/k_F$ is the de-Broglie wavelength of an electron having U_{F0} .

Mean Energy and Heat Capacity of Free Electrons

To obtain other thermodynamic and transport properties of the Fermi gas one must compute higher-order Fermi integrals. For example, the mean energy per unit volume is given by

$$\begin{aligned} \frac{\langle U \rangle}{V} &= \frac{2(m^*)^{3/2}}{\sqrt{2\pi^2\hbar^3}} \int_0^\infty f_{FD}(U)U^{3/2}dU = \frac{2(m^*)^{3/2}(k_B T)^{5/2}}{\sqrt{2\pi^2\hbar^3}} \int_0^\infty f_{FD}(x)x^{3/2}dx \quad (14) \\ &\equiv \frac{2(m^*)^{3/2}(k_B T)^{5/2}}{\sqrt{2\pi^2\hbar^3}} F_{3/2}(x_F) \end{aligned}$$

Like the density this has a much simpler form in the limit of low temperature $T \ll T_F$, found by using the step-function property of f_{FD}

$$\frac{\langle U \rangle}{V} \approx \frac{2(m^*)^{3/2}}{\sqrt{2\pi^2\hbar^3}} \int_0^{U_{F0}} U^{3/2}dU = \frac{(2m^*)^{3/2}(U_{F0})^{5/2}}{5\pi^2\hbar^3} = \frac{3}{5}\rho U_{F0}$$

The calculation of specific heat capacity is, in one sense, straightforward since all of the energy is kinetic and thus contributes to the total heat Q . So $C_V = \Delta Q/\Delta T \approx d\langle U \rangle/dT$. But when operated on (14), this derivative is complicated by the fact that $F_{3/2}(x_F)$ is itself implicitly a function of temperature. So like the phonon heat capacity, the C_V for electrons becomes relatively simple in the high-temperature and low-temperature limits. The high-temperature limit is most easily approximated starting with the same approximation as before, $f_{FD}(U) \approx \exp(U_F/k_B T)\exp(-U/k_B T)$, so that

$$\frac{\langle U \rangle}{V} = \frac{2(m^*)^{3/2}}{\sqrt{2\pi^2\hbar^3}} \int_0^\infty f_{FD}(U)U^{3/2}dU \rightarrow \frac{2(m^*)^{3/2}(k_B T)^{5/2}e^{U_F/k_B T}}{\sqrt{2\pi^2\hbar^3}} \Gamma(5/2)$$

Since $\Gamma(5/2) = 3(\pi)^{1/2}/4$, this becomes

$$\frac{\langle U \rangle}{V} = \frac{3(m^*)^{3/2}(k_B T)^{5/2}e^{U_F/k_B T}}{2\sqrt{2\pi^3}\hbar^3} = \frac{3}{2}\langle N_e \rangle k_B T \quad (15)$$

where $\langle N_e \rangle$ is given by Eqn (13).

From (15) we get $C_V \approx d\langle U \rangle/dT = (3/2)\langle N_e \rangle k_B$. This is just the electronic analog of the Dulong-Petit law for phonons. And like the phonon case, it can be deduced from the classical law of equipartition – that there is $(1/2)k_B T$ of energy per “degree-of-freedom”. In the case of electrons there are three degrees-of-freedom – three Cartesian components of the kinetic energy. For the phonons, there were six – three Cartesian components of the kinetic and potential energies.

The low-temperature limit of the heat capacity is not so simple but amenable to the following approximation method developed first by W. Pauli. It takes advantage of the fact that the kinetic energy of the Fermi gas, unlike phonons or any classical system, does not vanish as $T \rightarrow 0$. Hence, the heat capacity can be evaluated by the following approximation,

$$C_V \equiv \frac{\delta Q}{\delta T} \approx \frac{d\langle U \rangle}{dT} = \frac{d[\langle U(T=0) \rangle + \Delta\langle U(T) \rangle]}{dT} = \frac{d\{\Delta\langle U(T) \rangle\}}{dT}$$

since $U(T=0) \equiv U_{F0}$ is independent of T . The ΔU term in the numerator can be written

$$\Delta U = \langle U(T) \rangle - \langle U(T=0) \rangle = \int_0^\infty 2f_{FD}(U)UD(U)dU - \int_0^{U_{F0}} 2UD(U)dU \quad (16)$$

The second integral is not zero for free electrons since the Pauli exclusion principle prevents $\langle U(T=0) \rangle$ from vanishing. Following Pauli, (16) can be calculated using the clever addition-and-subtraction of a common term. The common term,

$$\int_0^{U_{F0}} 2U_{F0}D(U)dU, \text{ is chosen in light of the fact that } (U-U_F)\frac{df_{FD}(U)}{dT} \text{ behaves like a}$$

Dirac delta function when $T < T_F$ (see homework problem). This leads to

$$\Delta U = \int_0^\infty 2f_{FD}(u)UD(U)dU - \int_0^{U_F} 2UD(U)dU + \int_0^{U_{F0}} 2U_{F0}D(U)dU - \int_0^{U_{F0}} 2U_{F0}D(U)dU$$

But from the definition of N_e , $\int_0^{U_{F0}} 2U_{F0}D(U)dU = \int_0^\infty 2U_{F0}D(U)f_{FD}(U)dU = U_{F0}N_e$, so we can write

$$\Delta U = \int_0^\infty 2f_{FD}(U) (U-U_{F0})D(U)dU - \int_0^{U_{F0}} 2(U-U_{F0}) D(U)dU$$

The second integral is independent of temperature, so

$$C_V \approx \frac{d\langle \Delta U \rangle}{dT} \approx \int_0^\infty 2(U-U_{F0})D(U) \frac{df(U)}{dT} dU \approx \int_0^\infty 2D(U)\delta(U-U_{F0})dU \\ \approx 2D(U_{F0}) \int_0^\infty (U-U_{F0}) \frac{df_{FD}(U)}{dT} dU$$

The temperature derivative is

$$\frac{df_{FD}(U)}{dT} = \frac{+\exp[(U-U_F)/k_B T] (U-U_F)/k_B T^2}{\{\exp[(U-U_F)/k_B T] + 1\}^2}$$

So by defining $x \equiv (U-U_F)/k_B T$, we get

$$C_V \approx 2D(U_{F0}) k_B \times k_B T \int_{-U_F/k_B T}^\infty x^2 \frac{e^x}{(e^x + 1)^2} dx$$

Since we are in the low temperature limit, $U_F/k_B T \gg 1$, and this integral becomes

$$C_V \approx 2D(U_{F0}) k_B^2 T \int_{-\infty}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx$$

The definite integral has the value $\pi^2/3$, so that

$$C_V \approx \frac{2\pi^2 D(U_{F0})}{3} k_B^2 T$$

It is simple to show $D(U) = \frac{3N_e}{4U}$ so $D(U_F) = \frac{3N_e}{4U_F}$. And since $U_F \equiv k_B T_F$, we get

E

Metal	$N_e/V(\text{cm}^{-3})$	$U_F(\text{eV})$	T_F	$\frac{C_v(\text{electron})}{C_v(\text{phonon})}(300\text{K})$
Cu	8.45×10^{22}	7.00	81200	6.0×10^{-3}
Au	5.9×10^{22}	5.51	63900	7.7×10^{-3}
Al	18.06×10^{22}	11.63	134900	3.6×10^{-3}

$$C_v \approx \frac{2\pi^2}{3} \frac{3N_e}{4U_F} k_B^2 T = \frac{\pi^2}{2} N_e k_B \left(\frac{T}{T_F}\right) \quad (17)$$

This simple expression has several interesting properties. First it looks remarkably similar to the classical result $(3/2)\langle N_e \rangle k_B$ derived from (15). But for $T \ll T_F$, it is significantly less than the classical result for the following reason. In this degenerate limit, all the electrons occupying energy levels well below the Fermi energy are prohibited by Pauli exclusion from making any small energy changes with a change of temperature because all nearby levels are already occupied. Second, it is independent of the solid dimensionality so applies just as well to a bulk sample as it does to a quantum well or quantum wire. Third, it is useful over a very wide temperature range in common metals, particularly around 300 K or below, since $T_F > 10^4$ K in these materials.

Since in metals the the number of free electrons tends to be an integral times the number of atoms, we can contrast (17) to the Dulong-Petit law for phonon $C_v \approx 3N_c k_B$ where N_c is always the number of primitive unit cells . In the special case of one free electron per atom, $N_e = N_c$, and we get

$$\frac{C_v(\text{electron})}{C_v(\text{phonon})} \approx \frac{\pi^2}{6} \frac{T}{T_F} = 1.64 \left(\frac{T}{T_F}\right)$$

This suggests that the electronic contribution to specific heat capacity is insignificant except, perhaps, at low temperatures where the T^3 dependence for phonons drops to levels comparable to the T dependence for electrons. Experimentally, metals usually display a heat capacity given by the expression

$$C = \gamma T + AT^3,$$

where $\gamma T \rightarrow$ is the electronic term, and AT^3 is the phonon term consistent with the Debye model. For temperatures $T \approx T_D$ but $T \ll T_F$, the phonons generally dominate the heat capacity of a solid, even in metals.

It is interesting to contrast these results against the common experience with the thermal properties of metals and insulators. For example, a 1 cm^3 sample of copper “feels” much colder to the human touch than a 1 cm^3 cube of crystal quartz, for example. But this “feel” of cold, like many interactions between solids and outside “forces”, is dominated not by the heat capacity but by the thermal conductivity. And as we shall see later in the coverage of transport theory, the thermal conductivity depends on the velocity of the energy carriers, be it the electrons or phonons. Phonons have a maximum velocity of the speed of sound, typically between 5,000 and 10,000 m/s for the common metals in Table III. Because this is so much smaller than the average electron velocity in metal, the phonons in a crystalline insulator are typically much less effect in heat *transport* than

the electrons in a metal, in spite of their much higher specific heat capacity. We will return to this important issue later.

Plasma Oscillations and Waves

One of the essential aspects of the Fermi electron model in solids is that it maintains space-charge neutrality. The number of free-electrons is counterbalanced by an equal number of positively-charged atomic cores. In other words, the solid is a *neutral plasma* and, as such, can display some interesting effects. One of the most interesting and timely for electronics and photonics is the plasma oscillations. Physically, they arise by displacing the Fermi sea of electrons from its time-averaged position with an external electric or magnetic force, and then releasing the force. What results is a decaying collective oscillation of the Fermi sea that is reminiscent of lattice waves. Hence, the total energy derived for the Fermi sea is just the static term, analogous to cohesive energy of the atomic cores. The energy associated with the collective motion of the electrons is analogous to the lattice waves. And like the lattice waves, the amplitude of the plasma waves can be so small that quantum mechanics must be used to predict the magnitude, leading to the plasmons.

Electrostatic Interaction and Plasma Waves

The discussion of plasmons must start with issue of internal electric fields and the deviation of any material from space-charge neutrality. To simplify the analysis and glean the key concepts, we address the slab sample in Fig. 2 in which the Fermi gas is displaced by a distance Δr uniformly to one side of the fixed atomic cores. If Δr is small, there will be a net sheet density ρ_s of (negative) charge on this side, and an equal density of (positive) charge on the opposite side. This represents a violation of space-charge neutrality that, from basic electrostatics, must create an electric field. The relation between the two is given by Poisson's equation:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon} \quad (18)$$

where ρ is the net charge density, total positive minus total negative, and ϵ is the dielectric constant. This is elegantly solved for the given geometry by Gauss' divergence theorem over the "pillbox" drawn in Fig. 2(a). Integrating the right side of (18) over the pillbox, we get

$$\int_V \frac{\rho}{\epsilon} dV = \frac{Q}{\epsilon} \quad (19)$$

because of our assumption of a uniform ρ . To integrate the left side over the pillbox, we apply two arguments: (1) Gauss' divergence theorem, and (2) a surface electric field oriented perpendicular to the slab. Gauss' theorem follows from very general considerations about continuous vector fields, so applies as well in a solid-state (neutral) plasma as in any other medium. The perpendicular electric field follows by symmetry considerations when the diameter of the pillbox is much less than the width of the slab. The application of both leads to

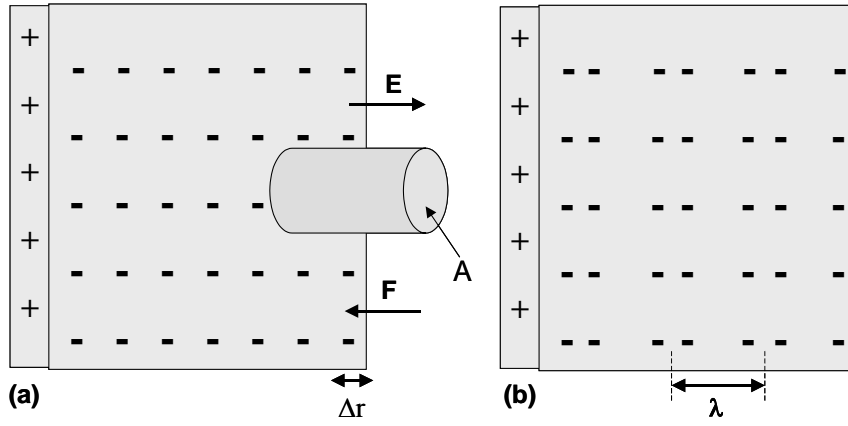


Fig. (a) Schematic view showing displacement of Fermi gas of electrons from the positively charge and static atoms, creating a surface charge, electric field, E and force F . The displacement shown has infinite wavelength. (b) Same sample as (a) but now with non-uniform displacement of electronic charge, creating a longitudinal plasma excitation of finite wavelength

$$\int_V \vec{\nabla} \cdot \vec{E} dV = \oint_S \vec{E} \cdot d\vec{S} = E \cdot A \quad (20)$$

where \vec{E} is the electric field vector, E is its magnitude, V is the volume of the pillbox, \vec{S} is the surface vector, and A is the area of the end cap. Equating (19) and (20) we get

$$E = Q/(\epsilon A) \equiv -e\rho_s / \epsilon, \quad (21)$$

where ρ_s is the sheet electron density.

Eqn (21) is a well-known result from electrostatics, and must be associated with the restoring force shown in Fig. 2(a) arising from Coulomb's law,

$$\vec{F} = Q \cdot \vec{E} \quad (22)$$

which points opposite to the electric field in the electron layer since Q is negative there.

Now as in the case of lattice waves, we apply Newton's law to the entire electron layer (where force exists)

$$F = Q \cdot E = M^* \frac{d^2 \Delta r}{dt^2}$$

where M^* is the effective mass for all the electrons in the layer. A single-dependent variable is created by yet another consequence of the uniform charge assumption: $\rho_s = \rho \cdot \Delta r$ where ρ is the bulk density. This allows us to write an oscillation equation

$$\frac{Q \cdot \rho \cdot \Delta r}{\epsilon} = M^* \frac{d^2 \Delta r}{dt^2} \quad \text{or} \quad \frac{d^2 \Delta r}{dt^2} + \frac{e^2 \cdot \rho \cdot \Delta r}{m^* \epsilon} = 0$$

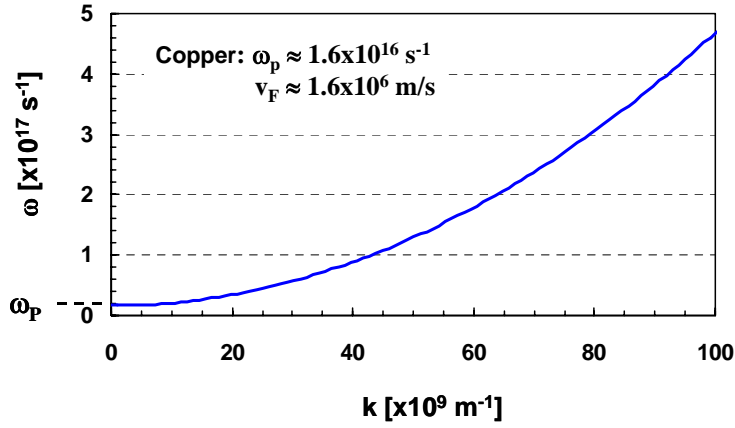


Fig. 3. Dispersion curve for solid-state (neutral) plasma waves.

since $Q = -N_e \cdot e$ and $M^* = N_e \cdot m^*$. The latter expression is a linear 2nd order differential equation with constant coefficients so is expected to have sinusoidal solutions. Either a $A\sin(\omega t)$ or $B\cos(\omega t)$ trial solution leads to the relation

$$-\omega^2 + \frac{e^2 \cdot \rho}{m^* \epsilon} = 0 \quad \text{or} \quad \omega_p = \left(\frac{e^2 \rho}{m^* \epsilon} \right)^{1/2} \quad (23)$$

where ω_p is the circular plasma frequency. The solution $A\sin(\omega_p t)$ or $B\cos(\omega_p t)$ describes a plasma oscillation, in this case a uniform one.

Another way to think about the plasma wave is that the uniform charge assumption limits the solution for the Δr oscillation to a wave vector $k = 0$. It is simple to imagine how plasma waves having nonzero wave vectors could exist, similar to the case of lattice waves, but not limited to discrete (Nyquist) sampling and the associated Brillouin-zone. If the gradient in the charge density is parallel to the direction of propagation, one has a longitudinal plasma wave, similar to the longitudinal lattice wave. But a detailed analysis shows that the waves with larger wave vector would have commensurately larger frequency ω . The resulting plasma-wave dispersion curve is shown in Fig. (3) and is approximated by the following form up to a power of k^2

$$\omega \approx \omega_p \left(1 + \frac{3k^2 v_F^2}{10\omega_p^2} + \dots \right)$$

The longitudinal plasma wave differs from the longitudinal acoustical lattice wave in two important aspects: (1) ω does not go to zero as k goes to zero, but rather approaches a minimum value of ω_p , and (2) ω is concave-up versus frequency for plasmons, whereas ω is always concave down for acoustical phonons. These aspects are, in fact, related to the fact that there is a cut-off nature to the propagation of plasmons, similar to the cut-off condition of electromagnetic modes in a hollow metal pipe (i.e., waveguide). The reason for the cutoff frequency is simply that the plasma displacement Δr leads to an oscillation no matter what the wavelength. In contrast, a uniform lattice displacement cannot oscillate because there is no potential energy added to the “springs”, i.e., the bonds between adjacent atoms. In fact, a uniform lattice displacement is equivalent to just a linear *translation* of the entire solid. An external force is required to

produce the translation, but after the force is “turned off”, the solid simply remains in the new position and no oscillation occurs.

Plasmons

We chose the label Δr for the displacement of the electron gas from the background atoms to remind us of lattice waves, for which we chose the same label to represent the deviation of each atom from its equilibrium position. And just as with lattice waves, we can calculate a total energy for a plasma wave in terms of Δr . The kinetic energy is straightforward, $U_{KE} = (1/2)M*(d\Delta r/dt)^2$. For an assumed (long wavelength) plasma wave of the form $\Delta r = A\sin(\omega_p t)$, we get the instantaneous form

$$U_{KE} = (1/2)M*\omega^2 A^2 \cos^2(\omega_p t) \quad (20)$$

The potential energy term is not so obvious but can be recalled from electrostatics wherein one of the most useful results is the (potential) energy density of an electric field

$$U_{PE} = (1/2) \epsilon E^2$$

Substitution of our surface charge layer parameters leads to

$$U_{PE} = \frac{e^2 \rho^2}{2\epsilon} A^2 \sin^2(\omega_p t) \quad (21)$$

Just as for the lattice waves Eqns (20) and (21) represent, for a given ω , a classical harmonic oscillator. The telltale sign is the *quadrature* relationship (i.e., 90° phase difference) between the instantaneous kinetic and potential energies.

Given the harmonic oscillator representation, it is natural to ask, just as for lattice waves, what value of plasma-wave amplitude can be resolved experimentally. And this quickly leads to the realization that quantum mechanics must be applied in solving for A. The quantities to be quantized are the plasma wave amplitude and its conjugate momentum, both done by the same *second quantization* procedure as for phonons. For brevity we will not go over this procedure again, but jump to the logical conclusion. That is, for each allowed plasma wave on the dispersion curve of Fig. 3, there exists a harmonic oscillator ladder of quantum states given by

$$U_k = (n_k + 1/2)\hbar\omega_k$$

In this case n_k is the number of plasmons, and represents the amplitude of the plasma wave of wave vector k .

Statistical mechanics can be applied via the Boltzmann criterion to compute the mean energy $\langle U_k \rangle$ as

$$\langle U_k \rangle = (\langle n_k \rangle + 1/2)\hbar\omega_k$$

And just as with phonons, the “subsystem” is the distributed plasma wave, or “mode”, represented by k . As such, the subsystem has no mass and, therefore, no chemical potential. It should therefore be governed by the same Planck statistics as the phonons,

$$\langle n_k \rangle = \frac{1}{\exp(\hbar\omega_k/k_B T) - 1}$$

This means that the same statistical averages, and the associated rise up the solid-state hierarchy back to the macroscopic level, can be carried out for plasmons, just as for phonons. For example, we can compute the temperature-dependent part of the mean energy for the entire plasma as

$$\langle U \rangle = \int_0^{\infty} \frac{D(\omega_p) \hbar \omega_p d\omega}{\exp(\hbar \omega_p / k_B T) - 1}$$

where $D(\omega)$ is the plasmonic density-of-states. The computation of this interesting quantity will be left to a homework exercise, but it is obvious that the cutoff effect $\omega \geq \omega_p$ displayed in Fig. 3 adds some new subtlety to the calculation. Specifically, there is a singularity in $D(\omega)$ since $dk/d\omega$ diverges at $k = 0$. This turns out to be an integrable singularity, as first worked out by the theoreticians that developed the plasmon as a viable collective excitation of the solid-state – D. Pines and Nozieres – around 1950.

In addition and of great practical importance, the common metals of Table III clearly satisfy $\hbar \omega_p \gg k_B T$, so that $\langle n_k \rangle$ for all plasmonic modes in common metals is very small at room temperature. Plasmons can always be excited, typically by energetic electrons or other charged particles, but are not significantly present in thermal equilibrium. This picture changes significantly when the free electron concentration and the associated ω_p drops to the point where ω_p approaches $k_B T$. This occurs in semimetals, for example, and creates rather spectacular effects on the optical properties in the visible and infrared portions of the electromagnetic spectrum. This is just one of many interesting effects in the emerging field of research called “plasmonics.”

Internal Electrostatic Variation: Screening

So far our interaction between the electrons has assumed a collective long-wavelength response of the electron gas to the fixed distribution of atomic positive charge, which itself was assumed to be uniform. What happens if atomic distribution is non uniform, as often occurs with charged impurities or defects, or by intentional inhomogeneous doping in semiconductor devices? This leads to another important response of the Fermi electron gas called *screening*, which is even more prevalent than plasma waves since it is not limited to an ω_p cutoff frequency. In short, screening is the non-oscillatory collective reaction of the electron gas to internal charges, positive or negative, and always tending to achieve space-charge neutrality.

When the electron gas density is very high as in metals, the screening is very strong. Hence to 1st order the macroscopic electric field in any metal is practically zero and the internal electrostatic potential Φ , related by $\vec{E} = -\vec{\nabla}\Phi$, must be uniform. In other words, the metal is an electrical *isopotential*. This is arguably one of the most important effects in all of electronics and one that we often take for granted. It allows us to apply voltage bias to electronic devices from remote power supplies. Even more fundamentally, it allows us to define a *ground plane (or planes)* in devices, integrated circuits, and electromagnetic components of all types. Without such reference planes, it would be very difficult if not impossible to solve for the ac voltage and current through a semiconductor amplifier, or the electromagnetic modes on transmission lines and in waveguides.

The solution to the screening problem is difficult, in general, and requires the techniques of many-body physics and/or quantum-field theory. But it becomes tractable when $\Phi(x)$ is slowly varying and relatively weak compared to $k_B T$. Under this condition, we can think of $\Phi(x)$ as a weak perturbation of the thermodynamic equilibrium in the solid such that the chemical potential, or temperature-dependent Fermi energy, becomes

$$U_F(x) = U_F + e\Phi(x)$$

Consequently, the Fermi-Dirac function is modified as

$$\langle n_k(x) \rangle = \frac{1}{\exp[(U - U_F - e\Phi(x))/k_B T] + 1} \quad (30)$$

This allows us to make the giant step of combining the statistical mechanics with the electrostatics of Poisson's equation through one of the most important equations in solid-state,

$$\vec{\nabla} \cdot \vec{E} = -\nabla^2 \Phi = \frac{\rho_{\text{tot}}}{\epsilon} = \frac{-e\langle \rho_e(x) \rangle - \rho_A}{\epsilon} \quad (31)$$

where ρ_e is (statistically-averaged) spatially-dependent concentration of electrons, and ρ_A is the fixed density of background positive charge. The pair of equations (30) and (31) forms a *self-consistent* pair when combined with the generalization of (10)

$$\langle \rho_e(x) \rangle = \int_0^\infty 2D(U) \langle n_k(x) \rangle U \cdot dU \quad (32)$$

The *self-consistent* solution is generally very difficult, but becomes remarkably simple and interesting in the special case of a non-degenerate Fermi gas. In that case we can write

$$\langle n_k(x) \rangle \approx \exp[U_F + e\Phi(x)/k_B T] \exp(-U/k_B T)$$

and (32) takes on the form

$$\langle \rho_e(x) \rangle \approx \rho_0 \exp[e\Phi(x)/k_B T] \quad (33)$$

where ρ_0 is a spatially dependent concentration corresponding to (13)

$$\rho_0 \approx \frac{e^{U_F/k_B T}}{(2)^{1/2}} \left(\frac{m^* k_B T}{\pi \cdot \hbar^2} \right)^{3/2}$$

Substitution into (31) leads the remarkably simple Poisson equation

$$\nabla^2 \Phi = (e/\epsilon) \{ \rho_0 \exp[e\Phi(x)/k_B T] - \rho_A \}$$

Now because $\Phi(x)$ is assumed weak, we can Taylor expand the exponential to get

$$\nabla^2 \Phi \approx (e/\epsilon) \{ \rho_0 [1 + e\Phi(x)/k_B T] - \rho_A \}$$

The final step is to suppose that the electrons and background density of positive charge is, to first order, neutral, so that $\rho_0 = \rho_A$, and thus

$$\nabla^2 \Phi(x) \approx e^2 \rho_0 \Phi(x) / (\epsilon k_B T) \quad (34)$$

Being a 2nd order linear differential equation with constant coefficients (at a given temperature), (34) is particularly easy to solve and has solutions along each direction q of a Cartesian space given by

$$\Phi(q) = A\exp(\alpha q) + B\exp(-\alpha q) \tag{35}$$

By substitution back into (34), we find α has the form

$$\alpha = [(e^2\rho_0/(\epsilon k_B T))]^{1/2}$$

The first term of (35) can be discarded on the grounds of possibly yielding infinite potential and therefore, being nonphysical. The second term yields an exponential decay of the potential in space over a characteristic 1/e decay length of

$$\beta_D = [(\epsilon k_B T/ e^2\rho_0)]^{1/2} \tag{36}$$

This important quantity is called the Debye length after its creator.⁴

Physically, the Debye solution means that any small disturbance of a non-degenerate Fermi gas will be neutralized over the length scale of approximately β . The inverse dependence on ρ_0 in (36) is rather intuitive, but the direct dependence on T is not so clear. To elucidate this effect, we can recast (36) in a different form using (23) and the result derived in Chapter 1 for the rms voltage velocity of a Maxwellian distribution of particles, $v_{rms} = (3k_B T/m^*)^{1/2}$, here generalized for the effective mass. We get the fascinating expression,

$$\beta = \frac{v_{rms}}{(3)^{1/2}\omega_p} \tag{37}$$

Now we see that the temperature dependence in (36) is really a velocity, and the greater the velocity, the longer the screening length. Intuitively, this makes sense since “speedier” electrons are less “dedicated” to the job of screening the internal charge than slower ones.⁵

Finally, we note that the self-consistent set (30)-to-(32) can also be solved straightforwardly in the degenerate, or low-temperature, limit of the Fermi gas and for weak $\Phi(x)$. Although more laborious than the Debye method and requiring a dielectric response analysis, it also yields a characteristic length for the screening effect:

$$\beta_{TF} = \frac{v_F}{(3)^{1/2}\omega_p} \tag{38}$$

This is called the Thomas-Fermi screening length and is the subject of a homework problem. The obvious similarity between (37) and (38) is a great example of the profound nature of the Fermi gas model – how such a simple-looking model can yield such amazing and powerful predictions !

⁴ the same genius who created the Debye model for phonons, amongst other great discoveries of science

⁵ The analog in the human world is an impatient worker.