DEGENERATE FERMION SYSTEMS

Lecture Notes

8.322
Quantum-Theory II-2006

R. L. Jaffe

Massachusetts Institute of Technology

Contents

DEGENERATE FERMION SYSTEMS

Many of the systems which command the attention of modern physicists are dense collections of spin-1/2 particles (fermions) whose dynamics are dominated by the Pauli exclusion principle. Famous examples are the electrons in heavy atoms, in metals and in white dwarf stars, the nucleons (protons and neutrons) in nuclei and neutron stars. Even quarks may form such systems deep within collapsed stars. In this state the specific forces between particles are often less important than the general structure impressed on the system by the exclusion principle.

At high temperatures, particles have lots of energy and (as we shall see) many quantum states available to them. On the average, the probability that any quantum state is occupied is rather small ($\ll 1$) and the exclusion principle plays little role. At lower temperatures, particles have less energy, fewer quantum states are available and average occupation number of each state increases. Then the exclusion principle becomes essential: the available levels up to some maximum energy (determined by the density) are, on average, nearly filled; higher levels are, on average, nearly empty. Such systems are then termed "degenerate," hence the title of this section. Actually these statements are strictly true only at zero temperature and when the mutual interactions of the fermions are ignored.

The maximum energy of a filled level is known as the Fermi energy (E_F). A collection of degenerate fermions is often referred to as a Fermi gas, and sometimes, picturesquely, as a "Fermi sea," though the "sea" with its "Fermi surface" dividing filled from unfilled levels, exists in energy space rather than configuration space. Thermal effects (and interactions) smear out the top of the Fermi sea, allowing states which would be filled at zero temperature some probability of being empty and vice-versa. The scale of temperatures is set by the Fermi temperature, $T_F \equiv E_F/k$, where k is Boltzmann's constant. When $T \ll T_F$ thermal effects are small. Often in systems of interest to us the approximation $T = 0$ will be good enough because the high density of the matter makes E_F (hence T_F) very large compared to ambiant temperatures.

Chapter 0

NATURAL UNITS

Most of us are used to the cgs system of units in which *all* physical quantities are measured in combinations of centimeters, grams and seconds. This is a convenient, practical system for most macroscopic applications. When we leave the scale of human dimensions to study very small sizes and very energetic processes, the cgs system is no longer so natural. When relativity and quantum mechanics are important, the fundamental constants $\hbar$ (Planck's constant) and c (the speed of light) set natural scales for action and velocity. There is a particularly simple and elegant system of units which makes use of h and c and which is used by all particle physicists, most nuclear and astrophysicists and theorists of all kinds. It is known, somewhat arrogantly, as "natural" units. We will often use this unit system in this section. This section provides an introduction to natural units for those of you who have never encountered them before.

In the cgs system all quantities can be expressed in terms of fundamental unit of length (ℓ), time (t) and mass (m). It is worth remembering how this is accomplished for such diverse quantities as electric charge, electric and gravitational fields and so forth. Of course, mechanical quantities such as momentum, energy or viscosity are easily expressed as powers of m, ℓ, and t[1]

$$\begin{aligned} [\text{momentum}] &= m\ell t^{-1} \\ [\text{energy}] &= m\ell^2 t^{-2} \\ [\text{viscosity}] &= m\ell t^{-1} \quad . \end{aligned} \tag{1}$$

Electromagnetic quantities can be expressed in terms of m, ℓ and t by means of *dynamical* equations of electromagnetic theory. Thus Coulomb's law:

$$F = e_1 e_2 / r^2 \tag{2}$$

[1]In the subsequent equations $[x]$ is to be read "the dimensions of x."

tells us that

$$[\text{charge}] = [\text{force}]^{1/2}\ell = m^{1/2}\ell^{3/2}t^{-1} \quad . \tag{3}$$

So the basic unit of charge in the cgs system is the $\text{gm}^{1/2}\text{cm}^{3/2}/\text{sec}$(!). It is the charge which produces a force of 1 dyne at a separation of 1 cm from an equal charge. Since this is a cumbersome notation this unit is given its own name: the esu or stat Coulomb, but one should not forget that the unit of charge is a *derived* quantity in the cgs system. [This is not the case in the MKS system, where the Coulomb is defined to be the charge of so many electrons. To accommodate this *ad hoc* definition a constant $(1/4\pi\epsilon_0)$ must be added to Coulomb's law in MKS. For an excellent discussion of electromagnetic units see the Appendix on units in J. D. Jackson, *Classical Electrodynamics*.]

Similar exercises tell us the dimensions of many familiar quantities:

$$\begin{aligned} [\text{resistance}] &= \ell^{-1}t \\ [\text{inductance}] &= \ell^{-1}t^2 \\ [\text{magnetic field}] &= m^{1/2}\ell^{1/2}t^{-2} \\ [\text{compressibility}] &= m^{-1}\ell^{-2}t^2 \quad , \end{aligned} \tag{4}$$

for example.

There is no fundamental reason to use mass, length and time as the basic units of the system. One could choose any three independent quantities as the fundamental objects in which all physical objects are measured. "Natural units" are a case in point. They are obtained by using *action*, *velocity* and *energy* as the fundamental dimensional quantities. Remembering

$$\begin{aligned} [\text{action}] &= m\ell^2t^{-1} \\ [\text{velocity}] &= \ell t^{-1} \\ [\text{energy}] &= m\ell^2t^{-2} \end{aligned} \tag{5}$$

we can easily re-express quantities in terms of some basic units of action, velocity and energy. Just as we obtain the cgs system by taking the centimeter as the unit of length, the gram as the unit of mass and the second as the unit of time; so we obtain the natural system by taking Planck's constant (actually $\hbar = h/2\pi$) as the unit of action, the speed of light (c) as the unit of velocity and the electron volt (eV) as the unit of energy

$$\begin{aligned} \hbar &\equiv \frac{h}{2\pi} = 1.05457266(63) \times 10^{-27}\,\text{gm cm}^2\text{sec}^{-1} \\ c &= 2.99792458 \times 10^{10}\,\text{cm sec}^{-1} \\ \text{eV} &= 1.60217733(49) \times 10^{-12}\,\text{gm cm}^2\text{sec}^{-2} \quad . \end{aligned} \tag{6}$$

With (6) and (7) we can proceed to express physical quantities in natural units. Here are some examples:

$$\begin{aligned} [\text{time}] &= (\text{eV})^{-1}\hbar \\ [\text{length}] &= (\text{eV})^{-1}\hbar c \\ [\text{force}] &= (\text{eV})^{2}\hbar^{-1}c^{-1} \\ [\text{pressure}] &= (\text{eV})^{4}\hbar^{-3}c^{-3} \\ [\text{charge}^2] &= \hbar c \\ [\text{magnetic field}] &= (\text{eV})^{2}\hbar^{-3/2}c^{-3/2} \quad . \end{aligned} \tag{7}$$

I suspect that the advantages of this system are not yet apparent. The great advantage — and the great confusion for non-experts — comes when we *suppress mentioning* the factors of $\hbar$ and c, *leaving all physical quantities measured in units of electron volts*. Such a step could be taken in the cgs system, too. We could, for example, suppress the "cm" and "sec" and measure all quantities as some power of a fundamental unit of mass, the gram. This is not done for two reasons: first, because there is nothing particularly fundamental about one second or one centimeter so we are not eager to suppress the label which tells us that time was measured in seconds and length in centimeters; and second, because we are used to having a *different* set of units for every different physical quantity — thus, for example, momentum and energy (see (2)) have different units in cgs, but they would both be measured in grams is we suppressed cm and sec. If you quoted an answer to a calculation in grams, you would have to tell your reader whether it was a momentum or an energy before he would be able to evaluate it in cgs units.

In the case of natural units the first disadvantage is eliminated: $\hbar$ and c *are* natural units for action and velocity in fundamental physics; and the second disadvantage is outweighed by the great advantage of measuring all quantities in the same units. One must be careful, however, to specify the physical quantity of interest to avoid confusing things measured in the same powers of eV. The problem of converting back from natural units to cgs units is made easier by conversion factors

$$\begin{aligned} \hbar c &= 195.327053(59)\,\text{MeV fm} \\ \hbar &= 6.5821220(20)\times 10^{-22}\,\text{MeV sec} \end{aligned} \tag{8}$$

($1\,\text{MeV} = 10^6\,\text{eV}$, $1\text{ fm} = 10^{-13}\,\text{cm}$). Note $\hbar c$ is equal to unity in natural units.

At this point several examples will (I hope) make the use of natural units clearer and more compelling.

Example 1:

The energy equivalent of the electron's rest mass is 511 keV:

$$m_e c^2 = 511 \text{ keV} \quad .$$

What length is $[511 \text{ keV}]^{-1}$?

Answer:

$$\begin{aligned} \ell_e &= \frac{\hbar c}{m_e c^2} = 197 \,\text{MeV fm}/511 \text{ keV} \\ &= 385 \,\text{fm} = 3.85 \times 10^{-11} \,\text{cm} \quad . \end{aligned}$$

This is the electron's Compton wavelength.

What time is $[511 \text{ keV}]^{-1}$?

Answer:

$$t_e = \frac{\ell_e}{c} = 1.28 \times 10^{-21} \,\text{cm}$$

which is the time it takes light to travel an electron's Compton wavelength.

What pressure is $[511 \text{ keV}]^4$?

Answer:

$$\begin{aligned} P &= [511 \,\text{keV}]^4/(\hbar c)^3 \\ &= [511 \,\text{keV}]^4/[195 \,\text{MeV fm}]^3 \\ &= 8.9 \times 10^{-3} \,\text{eV}/\text{fm}^3 \\ &= 8.9 \times 10^{-3} (1.6 \times 10^{-12} \,\text{erg})/(10^{-39} \,\text{cm}^3) \\ &= 14.2 \times 10^{24} \,\text{dyne}/\text{cm}^2 \quad . \end{aligned}$$

Example 2: An electron with kinetic energy 10 eV scatters at an angle of 0.2 radian from an atom. What length scale structure within the atom does it probe?

Answer: First calculate its momentum:

$$\begin{aligned} p &= \sqrt{2mE} \\ &= (2 \times (511 \,\text{keV}) \times 10 \,\text{eV})^{1/2} \\ &= 3.2 \,\text{keV} \quad . \end{aligned}$$

Use the uncertainty principle:

$$\begin{aligned} \Delta p &\cong 0.2p = 0.64 \,\text{keV} \quad , \\ \Delta x &\cong 1/\Delta p = (0.64 \,\text{keV})^{-1} \quad . \end{aligned}$$

Restore cgs units:

$$\Delta x \cong 197 \,\text{MeV fm}/0.64 \,\text{keV} \cong 3.1 \,\text{Å}$$

Example 3: According to (8), e^2 — the square of the electron's charge — is dimensionless when measured in natural units. What is its value?

Answer:

$$\begin{aligned} e &= 4.803 \times 10^{-10}\,\text{esu} \\ e^2 &= 2.307 \times 10^{-19}(\text{esu})^2 \\ (1\,\text{esu})^2 &= 1\,\text{dyne-cm}^2 = 1\,\text{gm cm}^3/\text{sec}^2 \\ \hbar c &= 3.161 \times 10^{-17}\,\text{gm cm}^3/\text{sec}^2 \\ e^2 &= 2.307 \times 10^{-19}/2.161 \times 10^{-17}(\hbar c) = 1/137(\hbar c) \quad . \end{aligned}$$

$e^2/\hbar c \equiv \alpha$ is known as the "five structure constant." Its measured value is $\alpha = (137.0359895(61))^{-1}$.

Example 4: What is the energy of interaction of the magnetic moment of an electron in the magnetic field of a proton at a distance of 1 Å, when the spins are as shown in Fig. 0.1.

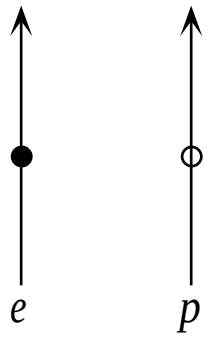


Figure 1: Proton and electron with spins parallel.

Answer:

$$\begin{aligned} E &= -\vec{\mu}_e \cdot \vec{B} = \frac{\vec{\mu}_e \cdot \vec{\mu}_p}{r^3} = -\frac{\mu_e \mu_p}{r^3} \\ \mu_e &= \frac{e\hbar}{2m_e} \\ \mu_p &= 2.793\left(\frac{e\hbar}{2m_p}\right) \\ E &= -2.793\frac{e^2\hbar^2}{4m_p m_e r^3} \quad . \end{aligned}$$

Now quickly:

$$\begin{aligned} e^2 &\rightarrow \alpha \\ \hbar^2 &\rightarrow 1 \\ m_p &\rightarrow 938\,\mathrm{MeV} \\ m_e &\rightarrow 511\,\mathrm{keV} \\ 197\,\mathrm{MeV\ fm} = 1 &\rightarrow 1\mathring{A} = 10^5/197\,\mathrm{MeV} \end{aligned}$$

So

$$\begin{aligned} E &= -2.793\left(\frac{1}{4}\right)\left(\frac{1}{137}\right)\left(\frac{1}{938}\right)\left(\frac{1}{0.511}\right)\left(\frac{197}{10^5}\right)^3\ \mathrm{MeV} \\ &= 8.13\times 10^{-8}\,\mathrm{MeV}\ . \end{aligned}$$

Chapter 1

INTRODUCTION TO THE PROPERTIES OF DEGENERATE FERMION SYSTEMS

The simplest system with which to begin a study of degenerate fermions is a collection of identical, spin-1/2 fermions confined to a cube of side ℓ by a boundary condition $\psi = 0$. We will be able to draw a lot of insight from this simple case and use it to study a very wide variety of problems. First we remind ourselves of the solutions to the single particle Schrödinger equation, then construct the properties of the N-particle system. Later we shall include the effects of relativity, some interactions and non-zero temperature.

1.1 SCHRÖDINGER PARTICLE IN A BOX

Consider the Schrödinger wavefunction in coordinate space, $\psi(\vec{x}, t)$. It obeys the Schrödinger equation

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

with

$$H = \frac{\vec{p}^2}{2m} = -\frac{\hbar^2\vec{\nabla}^2}{2m} \quad .$$

The energy eigenstates are

$$\psi_E(\vec{x}, t) = e^{-iEt/\hbar}\psi_E(\vec{x})$$

obeying

$$H\psi_E = E\psi_E \quad .$$

In a three-dimensional box, $0 \le x_j \le \ell$, $(j = 1, 2, 3)$ the energy eigenstates, subject to $\psi = 0$ on the boundary, are

$$\psi_{n_1 n_2 n_3}(\vec{x}) = \sqrt{\frac{8}{\ell^3}} \sin\frac{n_1 \pi x_1}{\ell} \sin\frac{n_2 \pi x_2}{\ell} \sin\frac{n_3 \pi x_3}{\ell} \qquad n_j = 1, 2, \ldots$$

with eigenenergies

$$E(n_1, n_2, n_3) = \frac{\pi^2 \hbar^2}{2m\ell^2} \left(n_1^2 + n_2^2 + n_3^2\right) \quad . \tag{1.1}$$

The states are orthonormal

$$\int d^3\vec{x}\, \psi^*_{n_1 n_2 n_3}(\vec{x}) \psi_{m_1 m_2 m_3}(\vec{x}) = \delta_{n_1 m_1} \delta_{n_2 m_2} \delta_{n_3 m_3}$$

and complete

$$\sum_{n_1 n_2 n_3} \psi_{n_1 n_2 n_3}(\vec{x}) \psi^*_{n_1 n_2 n_3}(\vec{y}) = \delta^3(\vec{x} - \vec{y})$$

for any $\vec{x}$ and $\vec{y}$ in the box.

It is very useful to define a "density of states" available for a single particle in the box. First, let $N(E)$ be the number of states with energy below E,

$$N(E) = \sum_{n_1 n_2 n_3} \Theta\left(E - E(n_1, n_2, n_3)\right) \quad .$$

$\Theta(x)$ is the step function:

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} .$$

It is the integral of the Dirac δ-function, defined by

$$\begin{aligned} \delta(x) &= 0 \qquad && \text{for } x \neq 0 \\ \int dx\, \delta(x) f(x) &= f(0) \qquad && \text{for a suitably smooth function } f(x) \quad . \end{aligned}$$

To see this, observe that the derivative of $\Theta(x)$ is zero except at $x = 0$, where it is infinite, and the area under $\Theta'(x)$ is $\Theta(+\epsilon) - \Theta(-\epsilon) = 1$. Both $\delta(x)$ and $\Theta(x)$ can be viewed as the limit of a sequence of ordinary functions. For example

$$\begin{aligned} \delta(x) &= \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2} \\ \Theta(x) &= \lim_{\epsilon \to 0} \frac{1}{\pi} \left[\tan^{-1}\frac{x}{\epsilon} + \frac{\pi}{2}\right] \quad . \end{aligned}$$

This representation of the δ-function is known as the "Lorentzian representation." Note $N(E)$ is an integer valued, discontinuous function which jumps in integer steps when each eigenvalue is passed.

We define a density of states per unit energy, $\rho(E)$, by

$$\rho(E) \equiv \frac{dN}{dE} = \sum_{n_1 n_2 n_3} \delta\left(E - E(n_1, n_2, n_3)\right)$$

which is a sum of spikes of unit area (in E). When E is large and the spacing between states, δE, is small, $\rho(E)$ can be approximated by a continuous function. This kind of approximation will emerge naturally and will suffice for all our applications.

Our object is to count the number of states with energy below E. This is facilitated by considering a three-dimensional space with axes labeled by n_1, n_2 and n_3. Each integer point in the positive ($n_1 > 0$, $n_2 > 0$, $n_3 > 0$) octant of this space corresponds to a single state. Each point in turn can be associated with a unit volume. Equation (1.1) defines a sphere of radius $\left(2m\ell^2 E/\pi^2\hbar^2\right)^{1/2} \equiv \xi$ in (n_1, n_2, n_3) space. Ignoring the issue of whether cubes at the sphere's surface or on the planes $n_j = 0$ should be counted we estimate $N(E)$ to be

$$\begin{aligned} N(E) &= \left(\frac{1}{8}\right)\frac{4\pi}{3}\left(\frac{2m\ell^2 E}{\pi^2\hbar^2}\right)^{3/2} \\ &= \frac{1}{6\pi^2} V \left(\frac{2mE}{\hbar^2}\right)^{3/2} \quad . \end{aligned} \tag{1.2}$$

where $V \equiv \ell^3$. Note the factor $1/8$ because we want only the positive octant of the sphere.

The approximation of ignoring the "roughness" at the surface is measured by the surface area divided by the volume, or approximately by

$$\frac{1}{\xi} = \frac{\pi\hbar}{\sqrt{2mE}\,\ell} \quad .$$

$\sqrt{2mE} \equiv p$ is the characteristic momentum of particles we are trying to describe. Although momentum is not conserved for particles in a box $\sqrt{2mE}$ is its rsm *expectation* value. $h/p \equiv \lambda$ is their DeBroglie wavelength. Thus our smooth approximation to $N(E)$ is valid provided $\xi \gg 1$, or equivalently, $\lambda/\ell << 1$. It is certainly not valid for the lowest energy eigenstates, where $\lambda \approx \ell$, but should work well when we discuss energetic single particle states. Later in this section we will explore important corrections to $N(E)$ and look for situations in which they are physically significant.

Return to $N(E)$ and differentiate to obtain $\rho(E)$,

$$\rho(E) = \frac{1}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} V\sqrt{E} \quad . \tag{1.3}$$

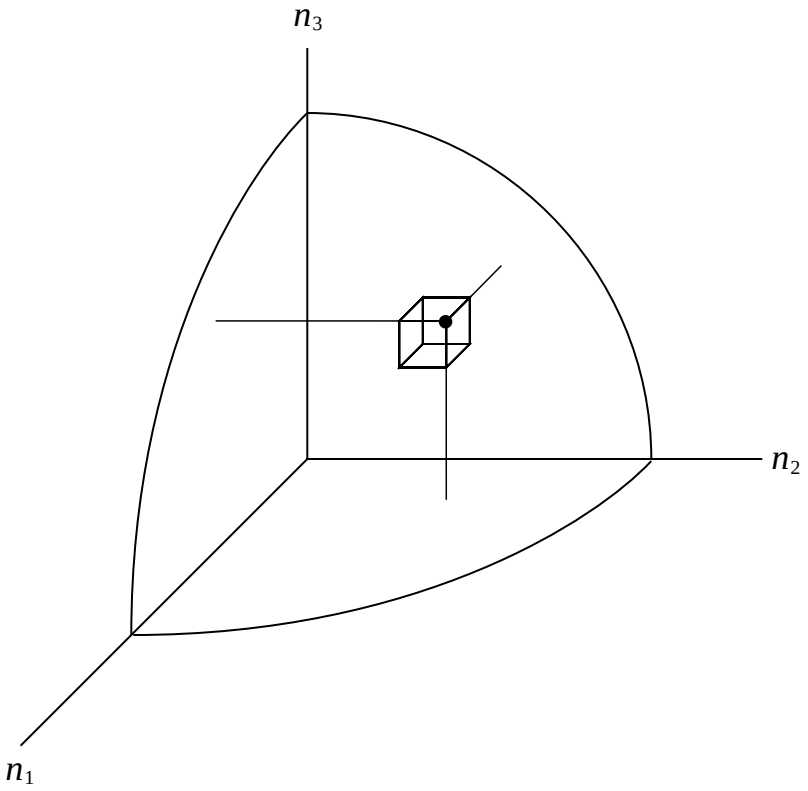


Figure 1.1: Each state of a particle in a box corresponds to a cube of unit volume in (n_2, n_2, n_3) space.

It is often convenient to introduce $p = \sqrt{2mE}$ and the wave number, $k \equiv p/\hbar$ and to define the associated density of states per unit k:

$$\rho(k) \equiv \frac{dN}{dk} = \rho(E)\frac{dE}{dk} = \frac{4\pi k^2}{(2\pi)^3}V \quad . \tag{1.4}$$

Although this result was derived for a cube and for large energy, it suggests a generalization

$$dN = \frac{d^3k\, d^3x}{(2\pi)^3} \tag{1.5}$$

to an arbitrarity shaped domain by means of the replacement

$$\begin{aligned} 4\pi k^2\, dk &\rightarrow d^3k \\ V &\rightarrow d^3x \quad . \end{aligned}$$

Equation (1.5) has a nice, semiclassical ring to it: There is one quantum state per volume h^3 in phase $(d^3x\, d^3p)$ space.

However the jump from (1.4) to (1.5) was hardly convincing. It suggests that the density of states is in some sense universal — independent of the potential energy, independent of the shape of the box, *etc.* In fact, (1.5) *is* universal but only as the *leading* term in an expansion valid at high density. We'll be able to establish this result more rigorously later at least under certain circumstances.

1.2 DYNAMICS AT ZERO-TEMPERATURE

To begin a study of fermion systems we consider zero temperature. For many applications of interest to us the temperatures are "small" and the zero temperature limit is accurate.

At $T = 0$ the ground state of a system of N fermions will have all single particle energy levels filled up to some maximum E (or p) and the remainder empty. This maximum single particle energy is known as the "Fermi energy," E_F. For a localized system momentum will not be conserved. It is convenient nevertheless to define a momentum $p \equiv \sqrt{2mE}$ as in (1.1). The momentum corresponding to E_F, $p_F = \sqrt{2mE_F} \equiv \hbar k_F$, is known as the "Fermi momentum." p_F is determined by the density alone, from (1.4),

$$N = g\int_0^{p_F} \frac{4\pi p^2 dp}{(2\pi\hbar)^3}V \tag{1.6}$$

$$n = \frac{N}{V} = \frac{g p_F^3}{6\pi^2\hbar^3} \tag{1.7}$$

where n is the particle density. Here we have introduced a "degeneracy factor," g, which counts the number of states available with the same momentum and position. The simplest example is spin: for a particle of spin-1/2, two states ($\uparrow$ or $\downarrow$) are available for each $\vec{x}$ and $\vec{p}$. For spin s, $g = 2s + 1$. Other examples occur: "nucleons" come in two species — the proton and neutron — so $g = 2(2s + 1) = 4$ for nucleons; quarks come in three "colors" and several "flavors" leading to a large value of g. It is only useful to introduce g if the interactions do not distinguish between the states at hand. Thus electrons with spin up and down must be treated separately when a sample is placed in a magnetic field.

We can also calculate the total internal energy (U) of the fermion system as a function of k_F. First we consider the case of a non-interacting, non-relativistic system

$$E(p) = \frac{p^2}{2m} \tag{1.8}$$

so

$$\begin{aligned} U &= \int_0^{p_F} dp \frac{dN}{dp} E(p) \\ &= g \int_0^{p_F} \frac{4\pi p^4 dp}{2m(2\pi\hbar)^3} V \end{aligned} \tag{1.9}$$

$$u = \frac{U}{V} = \frac{g\hbar^2 k_F^5}{20m\pi^2} \tag{1.10}$$

where u is the energy density. Clearly, all the intensive properties of the system — density, energy density, pressure, *etc.* — are functions of a single independent variable, which we may take to be k_F.

Now, using the style and method of thermodynamics (although we're at $T = 0$), let's deduce some important dynamical properties of the system. We assume that u and n are functions of k_F alone but we do not limit ourselves to the specific functional form of (1.7) and (1.10). U can be regarded as a function of V and N (cf. (1.7) and (1.10)), so

$$dU = \left(\frac{\partial U}{\partial V}\right)_N dV + \left(\frac{\partial U}{\partial N}\right)_V dN \tag{1.11}$$

The pressure, P, is defined by

$$P = -\left(\frac{\partial U}{\partial V}\right)_N \quad . \tag{1.12}$$

We can also define a "chemical potential," μ, by

$$\mu = \left(\frac{\partial U}{\partial N}\right)_V \tag{1.13}$$

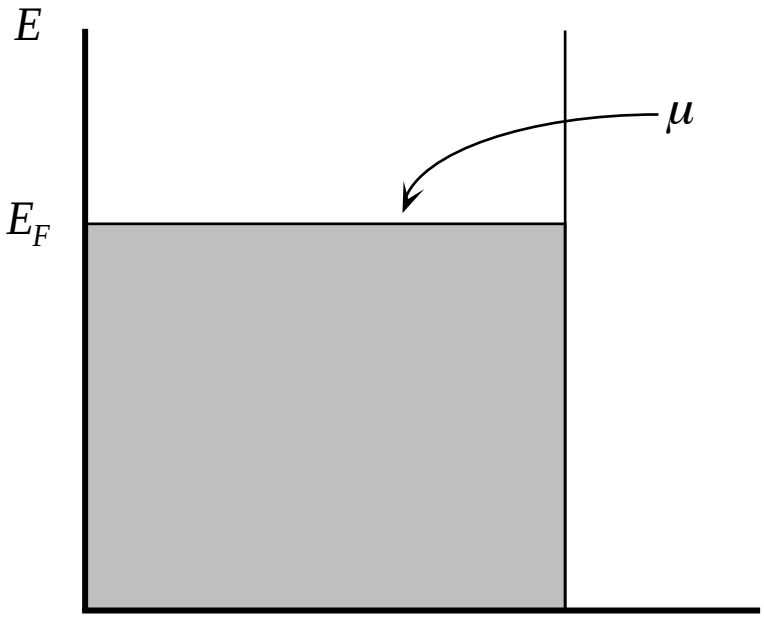


Figure 1.2: The chemical potential, μ, is the energy of the most energetic particle in a degenerate fermi system at zero temperture. All states with $E < \mu$ are occupied; all states with $E > \mu$ are unoccupied.

μ measures the change in internal energy of the system when a single particle is added while V is held fixed, so no mechanical work is done. That is, μ is the energy cost of putting a particle at the top of the Fermi "sea." On the basis of this, we expect $\mu = E_F = p_F^2/2m$. We will see below that this is a correct argument.

Since n and u depend on k_F but not V, we may write

$$N = n(k_F)V \tag{1.14}$$

$$U = u(k_F)V \quad . \tag{1.15}$$

Substituting from (1.14) and (1.15) into (1.11),

$$dU = Vu' \left[\left(\frac{\partial k_F}{\partial N} \right)_V dN + \left(\frac{\partial k_F}{\partial V} \right)_N dV \right] + u dV \tag{1.16}$$

where $'$ denotes d/dk_F. Now from (1.14),

$$dn = n' dk_F = d\left(\frac{N}{V}\right) = \frac{dN}{V} - \frac{N}{V^2} dV$$

whence

$$V \left(\frac{\partial k_F}{\partial N} \right)_V = \frac{1}{n'} \tag{1.17}$$

$$V \left(\frac{\partial k_F}{\partial V} \right)_N = -\frac{n}{n'} \quad . \tag{1.18}$$

Combining these with (1.16), we obtain

$$dU = u'\left(\frac{1}{n'}dN - \frac{n}{n'}dV\right) + udV \quad .$$

So we can read off,

$$\mu = \frac{u'}{n'} \tag{1.19}$$

$$P = \frac{u'n - un'}{n'} \quad . \tag{1.20}$$

While we're at it, let's compute some other useful dynamical properties of the system. First we compute the (inverse) compressibility, dP/dn, relying again on the fact that k_F is the only independent intensive variable,

$$\frac{dP}{dn} = \frac{dP}{dk_F}\frac{dk_F}{dn} = \frac{1}{n'}\frac{d}{dk_F}\left(\frac{u'n - un'}{n'}\right)$$

$$\frac{dP}{dn} = \frac{n}{(n')^3}\left(u''n' - u'n''\right) \quad . \tag{1.21}$$

A very useful dimensionless quantity related to the compressibility is the "adiabatic index," γ, defined by

$$\gamma \equiv \frac{n}{P}\frac{dP}{dn}$$

$$\gamma = \frac{n^2}{n'^2}\left(\frac{n'u'' - n''u'}{nu' - n'u}\right) \quad . \tag{1.22}$$

γ is important in many applications where dynamical stability is obtained by balancing the internal pressure of the gas against an "external" pressure like gravity. If γ is a constant, the definition of γ can be integrated to yield

$$\frac{P}{n^\gamma} = \text{const.}$$

This, you will recognize, is the adiabatic (no heat flow) pressure-volume relation derived in thermodynamics.

Finally, another useful quantity is the speed of sound. The compressibility of the fermion system determines the restoring force while the mass density or energy density (in the relativistic case) determines the inertia. Not surprisingly the general expression for the speed of sound, v, is

$$v^2 = c^2\frac{dP}{d\bar{u}}$$

where c is the speed of light and $\bar{u}$ is the energy density including the particle's rest mass. Returning to earlier work

$$\bar{u} = mc^2 n + u \quad . \tag{1.23}$$

We can apply all these formulas to the case of a non-interacting, non-relativistic Fermi gas, as a simple example,

$$\begin{array}{lllllllll} n & = & \frac{g p_F^3}{6\pi^2\hbar^3} & & P & = & \frac{2}{3}u \ , & & u & = & \frac{g p_F^5}{20 m \pi^2 \hbar^3} \\ \mu & = & \frac{p_F^2}{2m} & & \frac{v^2}{c^2} & = & \frac{1}{3}\,\frac{p_F^2}{m^2} & & \frac{dP}{dn} & = & \frac{2}{3}\mu \\ \gamma & = & \frac{5}{3} & & & & & & & & \end{array} \tag{1.24}$$

The relation $P = \frac{2}{3}u$ captures one of the striking features of degenerate Fermi systems: they exert pressure even at zero temperature. This follows from the exclusion principle which requires the population of high energy levels at high density. "Degeneracy pressure," as it is called, is responsible for the remarkable stability of degenerate Fermi systems and for their prevalence in Nature.

To get some idea of the scale of common fermion systems we estimate the density, Fermi energy and pressure for some classical examples.

1. Electrons in Metals

As we shall see soon, metals can be reasonably approximated as a regular array of atoms permeated by a gas of electrons. Only a few (typically one or two) "valence" electrons per atom are free to move, the rest — "core" electrons — remain localized about each particular atom. At densities of ordinary matter, $\rho \cong 10\,\mathrm{gm/cm^3}$, most of the mass comes from nucleons, $m_N = 1.67 \times 10^{-24}\,\mathrm{gm}$ so the number of density of nucleons is $n_N \approx 6 \times 10^{24}\,\mathrm{cm^{-3}}$ (Arogodro's number). Roughly 1/2 are protons, each screened by an electron, so

$$n_e \approx 3 \times 10^{24}\,\mathrm{cm^{-3}} \quad .$$

To be specific, consider silver with $Z = 47$ and a single valence electron. We estimate an electron gas with a density

$$\begin{aligned} n_e &\approx \frac{1}{47} \times 3 \times 10^{24}\,\mathrm{cm^{-3}} \\ &\approx 6 \times 10^{22}\,\mathrm{cm^{-3}} \quad . \end{aligned}$$

The Fermi momentum is determined from (1.24), $n_e = p_F^3/6\pi^2\hbar^3$

$$\begin{aligned} p_F &= \left(6\pi^2 n_e\right)^{1/3}\hbar \\ p_F &\approx 1.8\,\mathrm{keV/c} \end{aligned}$$

How high an energy is this?

$$E_F = \mu = \frac{p_F^2}{2m} \approx 3\,\mathrm{eV} \quad .$$

Note that E_F is small compared to mc^2 so the electrons are definitely non-relativistic; also $kT \sim 1/40\,\mathrm{eV}$ at room temperature, so $E_F >> kT$. The electron Fermi gas in matter at room temperature is very energetic. Many of the striking properties of metals are explained by the presence of this hidden hot gas of electrons. Note that thermal effects are scaled by $kT/E_F \sim 1/100$ so they may be ignored entirely or included by means of a low temperature expansion.

2. Nucleons and Electrons at Nuclear Densities

At very high densities matter is stabilized by nuclear forces which are attractive at distances $\sim 1\,\mathrm{fm} \equiv 10^{-13}\,\mathrm{cm}$ and strongly repulsive at very small distances $\sim 0.2\,\mathrm{fm}$. In equilibrium, nuclear matter is characterized by a volume of about $6\,\mathrm{fm}^3$ per nucleon (proton or neutron) corresponding to a single nucleon in a sphere of radius 1.2 fm. Since roughly half the nucleons are protons, the charge density would be half the nucleon density. This positive charge density could be shielded by electrons or eliminated by converting protons to neutrons. That is a dynamical issue we'll discuss later in connection with neutron stars. Here let's estimate the Fermi momentum of nucleons and electrons at nuclear densities

$$n_N \cong \frac{1}{6}\,\mathrm{fm}^{-3} = \frac{1}{6}\times 10^{39}\,\mathrm{cm}^{-3} = \frac{g_N\left(p_F^N\right)^3}{6\pi^2\hbar^3}$$

$g_N = 4 = 2\times 2$ for spin and proton/neutron. Using MeV as our units for energy and MeV/c as our units for momentum

$$p_F^N = \left(\frac{6\pi^2}{4}n_N\right)^{1/3}\hbar \cong 270\,\mathrm{MeV/c} \quad .$$

This is a large momentum — $p_F^N c \sim \frac{1}{4}m_N c^2$ and $m_N c^2 \cong 940$ MeV — but, fortunately the motion is still approximately non-relativistic. At the top of the Fermi sea

$$\left(\frac{v_F}{c}\right)^2 = \frac{\left(p_F^N\right)^2}{M_N^2 c^2} = 0.083 \quad .$$

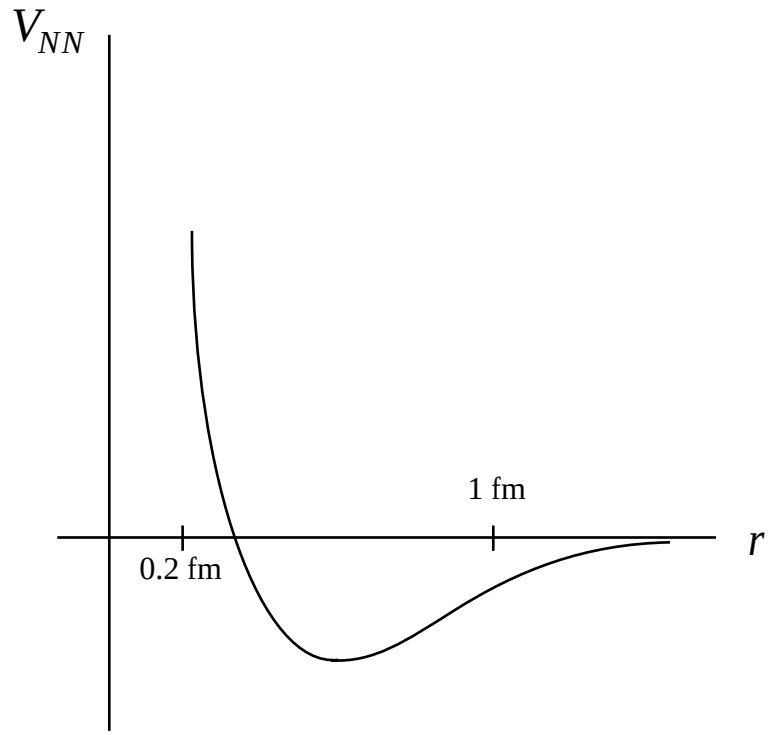


Figure 1.3: A "cartoon" of the nucleon-nucleon force — attractive at distances of order 1 fm., and repulsive at distances of order 0.2 fm.

The electrons are another story: Their density is $\approx \frac{1}{2} n_N$; the degeneracy factor is also down by a factor of two $g_e = \frac{1}{2} g_N$, so $p_F^e \cong p_F^N$; and the electrons are extremely relativistic. So to describe electrons at nuclear matter densities (or even much lower densities) one must use relativistic momentum/energy relations.

3. Quarks at and above Nuclear Matter Densities

At densities so high that internucleon separations are significantly less than 1 fm, a description in terms of nucleons breaks down. Nucleons, it terms out, are bound states with size of order 1 fm made of quarks. Matter at densities exceeding that of nuclear matter will have to be discussed in terms of quarks. The quark Fermi momentum can be measured relative to p_F^N

$$\frac{p_F^q}{p_F^N} = \left(\frac{2}{3} \, \frac{n_q}{n_N} \right)^{2/3} \quad .$$

The factor 2/3 comes because each species of quark has a degeneracy factor 6.

Quark masses differ from one type to another. The commonly accepted values are listed in Table 1.1.

For $n_q \approx n_N$ the different species of quark range from extremely relativistic to very non-relativistic. The s-quark is very interesting — neither limit applies.

Flavor	Name	Spin	Electric Charge	Approximate Mass (MeV)
u	up	$\pm 1/2\hbar$	$2/3$	5 MeV
d	down	$\pm 1/2\hbar$	$-1/3$	10 MeV
s	strange	$\pm 1/2\hbar$	$-1/3$	150 MeV
c	charm	$\pm 1/2\hbar$	$2/3$	1500 MeV
b	bottom	$\pm 1/2\hbar$	$-1/3$	5,000 MeV
t	top	$\pm 1/2\hbar$	$2/3$	$\sim$ 170 GeV

Table 1.1: The names, electric charges and masses of the six quarks.

1.3 RELATIVISTIC FERMI GAS, OTHER EQUATIONS OF STATE AT T=0

As we saw at the conclusion of the last section, the electron Fermi gas at nuclear densities and the quark Fermi gas system are relativistic. In this section we briefly summarize the relativistic Fermi gas, the extreme relativistic case and a simple approximate form which often suffices to connect the non-relativistic and relativistic cases.

1. Extreme Relativistic Case: $E = |\vec{p}|c$

It is easy to repeat the general analysis of Section 1.2 with the results

$$\begin{aligned} n(k_F) &= \frac{g p_F^3}{6\pi^2\hbar^3} \\ u(k_F) &= \frac{g c p_F^4}{8\pi^2\hbar^3} \\ \mu &= p_F c \\ P &= \frac{1}{3}u \\ \frac{dP}{dn} &= \frac{1}{3}p_F c = \frac{1}{3}\mu \\ \gamma &= \frac{4}{3} \ . \end{aligned}$$

Note that the relativistic gas is more compressible than the non-relativistic one

$$\left(\frac{dn}{dP}\right)_{\rm REL} > \left(\frac{dn}{dP}\right)_{\rm NR}$$

provided one is comparing systems at the same density and pressure.

2. General Case: $E = \left(\vec{p}^2c^2 + m^2c^4\right)^{1/2}$

For convenience, use relativistic units, $c = 1$, so $E = \left(\vec{p}^2 + m^2\right)^{1/2}$. In this case the integrals are complicated, though elementary functions,

$$\bar{u}(p_F) = \frac{4\pi g}{(2\pi\hbar)^3}\int_0^{p_F} p^2 dp\,\sqrt{p^2+m^2} \quad .$$

[Note that for $p_F^2 << m^2$ this reduces to the non-relativistic result including the rest mass energy (Eq. (1.23)).]

$$\bar{u} = \frac{4\pi g m^4}{(2\pi)^3}\int_0^{x_F} x^2\,dx\sqrt{x^2+1} \equiv \frac{gm^4}{2\pi^2}f(x_F)$$

where $x_F \equiv k_F/m$. $f(x_F)$ can be evaluated by elementary means,

$$f(x) = \frac{1}{4}x\left(1+x^2\right)^{3/2} - \frac{1}{8}x\left(1+x^2\right)^{1/2} - \frac{1}{8}\ln\left(x+\sqrt{1+x^2}\right)$$

but we can evaluate the compressibility without knowing this integral

$$\frac{dP}{dn} = \frac{1}{3}\,\frac{p_F^2}{\sqrt{p_F^2+m^2}} \tag{1.25}$$

which interpolates between the non-relativistic result, $p_F^2/3m$, and the relativistic result, $p_F/3$. Not surprisingly, the adiabatic index, γ, slowly changes from 5/3 to 4/3 as x_F goes from 0 to ∞.

3. Polytropes

For matters of dynamic stability the adiabatic index, γ, plays a central role. A particularly simple family of equations of state are those characterized by a constant γ

$$\gamma = \frac{n}{P}\,\frac{dP}{dn} = \text{const.}$$

It is easy to show that this requires

$$u = \sigma p_F^{3\gamma}$$

for some constant σ. So with

$$n = \frac{g\,p_F^3}{6\pi^2\hbar^3}$$

we find

$$P = \sigma(\gamma-1)p_F^{3\gamma} = (\gamma-1)u \quad . \tag{1.26}$$

It is easy to verify that $\frac{n}{P}\frac{dP}{dn} = \gamma$, but more interesting perhaps is the speed of sound

$$\begin{aligned} \bar{u} &= u + n\,mc^2 = \sigma p_F^{3\gamma} + \frac{gmc^2 p_F^3}{6\pi^2\hbar^3} \\ v^2 &= c^2\frac{dP}{d\bar{u}} = \frac{\gamma - 1}{1 + \frac{nmc^2}{\gamma\sigma p_F^{3\gamma}}} \quad . \end{aligned}$$

Since $v^2 > 0$, we see $\gamma \geq 1$ is required. Causality requires $v^2 \leq c^2$, so $\gamma \leq 2$ is also required. Thus

$$1 \leq \gamma \leq 2 \quad .$$

For "large" Fermi momentum and $\gamma > 1$, $n/p_F^{3\gamma} \to 0$, so

$$\frac{v}{c} \to \sqrt{\gamma - 1} \quad . \tag{1.27}$$

In particular, for the extreme relativistic case, $\gamma = 4/3$, $v = \sqrt{\frac{1}{3}}c$ is the velocity of sound.

It should be noted that the interest in polytropic equations of state does not appear to come from any fundamental *physical* arguments but instead because, 1) they yield tractable differential equations for stellar structure and 2) they apply to any equation of state in a region where $n \approx$ constant, since $\gamma(n) \to \gamma$ is a good approximation.

1.4 FINITE TEMPERTURE EFFECTS

So far we have been studying the quantum mechanical ground state of the N-fermion system. The particles occupy the lowest energy states available. Those states with energy below E_F have unit probability to be occupied, those with energies above E_F have unit probability to be empty. Now we consider what happens when such a system is brought into thermal equilibrium at a temperature T. These fermions can exchange energy with the heat bath. It is most convenient to describe the thermodynamics of this system in a framework where the number of particles in the system is regarded as variable and is known only as a thermal, or "ensemble" average. This requires an extension of statistical mechanics beyond 8.044 level which is developed in subsection 1 below.

The result of this work is a famous formula for the (thermal) average occupation number of an energy level with energy E:

$$f(E) = \frac{1}{1 + e^{(E-\mu)/kT}} \tag{1.28}$$

where k is Boltzmann's constant and μ is the generalization of the "chemical potential" introduced in §1.b to finite temperature. $f(E)$ behaves as expected as $T \to 0$, namely

$$\lim_{T\to 0} f(E) = \begin{cases} 1 & E < \mu \\ 0 & E > \mu \end{cases} . \tag{1.29}$$

In this limit we can identify μ with the Fermi energy, E_F. For $T > 0$, (1.28) describes the way in which thermal effects smear out the top of the Fermi sea. At very large temperatures, $T \gg E_F/k$, $f(E)$ reduces to the familiar Boltzmann's factor

$$f(E) \underset{E\to\infty}{\sim} e^{-E/kT} \quad . \tag{1.30}$$

$f(E)$ and its derivative are plotted for a variety of values of kT in Fig. 1.4.

According to (1.28), the occupation number of each energy level is specified only on (thermal) average. The chemical potential, μ, is determined by constraining the total number of particles to sum to some value, $\bar{N}$, on (thermal) average:

$$\bar{N} = \int_0^\infty dE\, \rho(E) f(E) \tag{1.31}$$

where $\rho(E)$ is the density of states (1.30). Equation (1.31) should be regarded as an equation for μ as a function of other thermodynamic variables like T, P and $\bar{N}$.

$f(E)$ is a pretty singular function, as can be seen from its behavior as $T \to 0$. It is difficult, in general, to perform the calculations involving integrals of $f(E)$ necessary to extract thermodynamic properties of the Fermi gas. We are primarily interested in low temperatures where Fermi systems remain approximately degenerate and their special features persist. Fortunately, there is an approximation to $f(E)$ for this limit which makes calculation easier. Readers who are familiar with the derivation of $f(E)$ or are willing to accept (1.28) without proof might want to skip directly to §1.d.2 where the approximation is developed. Finally, in §1.d.3 we proceed with developing the thermal properties of a Fermi Gas at low temperatures.

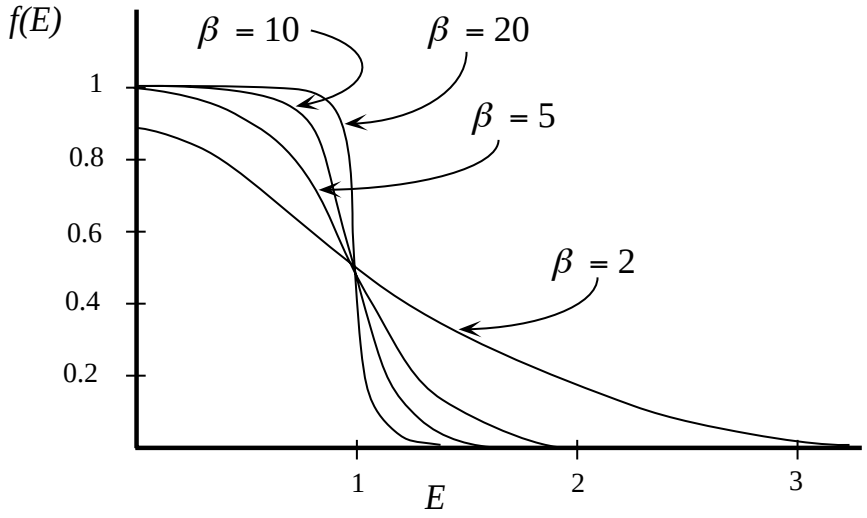
1. The Fermi Distribution Function

In thermodynamics, the properties of a system at fixed volume (V), temperature (T) and particle number ($\bar{N}$) may be summarized by the Helmholtz free energy, F, defined by

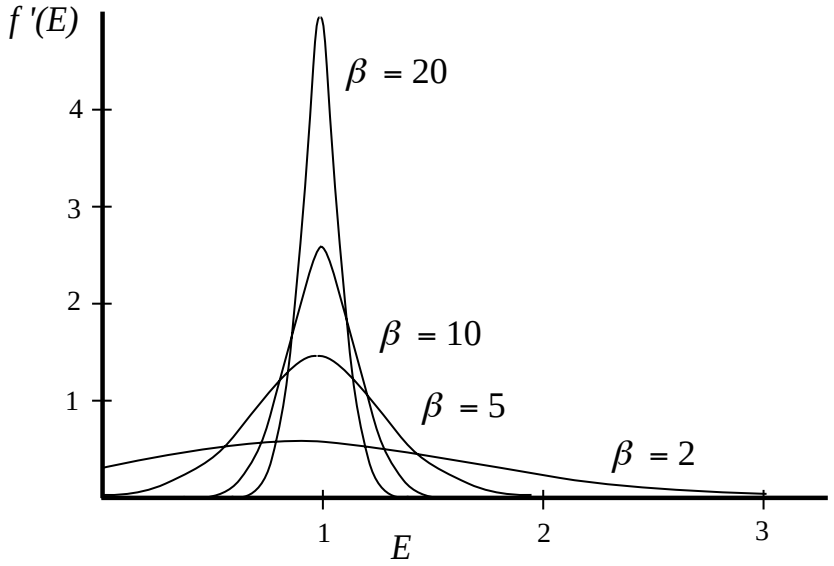
$$F \equiv U - TS \tag{1.32}$$

where U and S are the internal energy and entropy, respectively. For reversible processes

$$dU = -PdV + TdS \tag{1.33}$$



(a)



(b)

Figure 1.4: The Fermi distribution (a) and its derivative (b) plotted for several values of the temperature.

(we consider only the possibility of mechanical, "PdV," work), so

$$dF = -PdV - SdT \tag{1.34}$$

or

$$\left(\frac{\partial F}{\partial V}\right)_T = -P \qquad \text{and} \qquad \left(\frac{\partial F}{\partial T}\right)_V = -S \quad . \tag{1.35}$$

A standard trick in thermodynamics is to trade independent variables by a Legendre transformation. For example, it is inconvenient to use F to study systems at constant pressure because V, not P, is regarded as an independent variable in (1.32) and (1.34). To get around this we *define* the "Gibbs" free energy,

$$G(P,T) \equiv F + PV \tag{1.36}$$

where $P = P(V,T) \equiv -\left(\partial F/\partial V\right)_T$, so that

$$dG = VdP - SdT \tag{1.37}$$

or

$$\left(\frac{\partial G}{\partial P}\right)_T = V \qquad \text{and} \qquad \left(\frac{\partial G}{\partial T}\right)_P = -S \quad . \tag{1.38}$$

$G(P,T)$ is then the appropriate function for the study of mechanical systems at constant pressure. The same transformation is used to trade velocity for momentum in going from the Lagrangian to Hamiltonian form of mechanics.

We now play the same trick with particle number. All of the quantities we have been discussing depend on the number of particles in the system, $\bar{N}$. Up to now we have suppressed that dependence. Now we make it explicit: $F = F(T,V,\bar{N})$, *etc.* We define the chemical potential as the derivative of F with respect to $\bar{N}$,

$$\mu = \left(\frac{\partial F}{\partial \bar{N}}\right)_{V,T} \quad . \tag{1.39}$$

μ measures the change of free energy when an additional particle is added to the system (at fixed V and T). Next we trade the $\bar{N}$ dependence of F for the μ dependence of a new thermodynamic function known as the "thermodynamic potential," Ω,

$$\Omega \equiv F - \mu\bar{N} \tag{1.40}$$

with

$$d\Omega = -PdV - SdT - \bar{N}d\mu \tag{1.41}$$

or

$$\left(\frac{\partial\Omega}{\partial V}\right)_{T,\mu} = -P \ , \qquad \left(\frac{\partial\Omega}{\partial T}\right)_{V,\mu} = -S$$
$$\left(\frac{\partial\Omega}{\partial \mu}\right)_{T,V} = -\bar{N} \ . \tag{1.42}$$

Next we turn to statistical mechanics and relate Ω to the microscopic characteristics of the system. For fixed particle number this connection was made (in 8.044) through the partition function,

$$Z(T,V,N) = \sum_{\nu} \exp\left(-\frac{E_\nu}{kT}\right) \tag{1.43}$$

(where the sum ranges over all states of the system labeled by ν) and through the interpretation of the Boltzmann factor, $\exp -E_\nu/kT$, as the (thermal) average relative probability to find the system in a state with energy E_ν. Thus, for example, the internal energy is given by

$$\begin{aligned} U(T,V,N) &= \langle E_\nu \rangle = \frac{1}{Z}\sum_{\nu} E_\nu \, e^{-E_\nu/kT} \\ &= kT^2 \frac{d}{dT} \ln Z(T,V,N) \ . \end{aligned} \tag{1.44}$$

Similarly, the Helmholtz free energy is given by

$$F(T,V,N) = -kT \ln Z(T,V,N) \ . \tag{1.45}$$

We shall give the generalization to indefinite particle number without proof. [The proof requires introducing and developing the "grand canonical ensemble" and is given in any reasonable introduction to statistical mechanics, *e.g.* Morse, *Thermal Physics*; Huang, *Statistical Mechanics*.] We define the "grand partition function" by

$$\mathcal{Z}(T,V,\mu) \equiv \sum_{N} e^{\mu N/kT} Z(T,V,N) \tag{1.46}$$

and interpret

$$e^{(\mu N - E_\nu)/kT} \tag{1.47}$$

as the relative probability to find the system with *particle number N and energy E_ν*. We then assert that Ω is given by

$$\Omega = -kT \ln \mathcal{Z} \tag{1.48}$$

in analogy to (1.45). Equation (1.46) is not all that obvious. It is, perhaps, reasonable that μN enters into the exponential in a manner analogous to E_ν, since μ is identified with the free energy per particle. However, the sign seems surprising. In fact, it is an entropy effect: configurations with larger N are favored (at fixed T and E) because they can be put together in more ways (higher entropy).

Having balked at deriving (1.46) and (1.47) we shall at least show that they are consistent with the thermodynamic definition of Ω. In particular, we expect $(\partial\Omega/\partial\mu)_{T,V} = -\bar{N}$ from (1.42) and find from (1.47) and (1.48)

$$\left(\frac{\partial\Omega}{\partial\mu}\right)_{T,V} = -\frac{\sum\limits_N N\, e^{\mu N/kT} Z(T,V,N)}{\mathcal{Z}} \tag{1.49}$$

which gives $-\bar{N}$ provided $e^{(\mu N - E_\nu)/kT}$ is interpreted as a relative probability.

The great virtue (and the whole point!) for us to study $\mathcal{Z}$ is that it factors into a product of individual particle contributions for a Fermi gas. Consider a system of fermions with energy levels $E_1, E_2, E_3, \ldots$ available. The quantum state of this system is determined completely by given the occupation numbers $n_1, n_2, n_3, \ldots$ of each level. In this case

$$\begin{aligned} E_\nu &= \sum_j n_j E_j \\ N &= \sum_j n_j \end{aligned} \tag{1.50}$$

and

$$\mathcal{Z} = \prod_j \left(\sum_{n_j} e^{n_j(\mu - E_j)/kT} \right) \tag{1.51}$$

where we have used the familiar fact that the exponential of a sum is the product of exponentials. For fermions the only possibilities are $n_j = 0$ or 1, so (1.51) becomes

$$\mathcal{Z} \equiv \prod_j \mathcal{Z}_j = \prod_j \left(1 + e^{(\mu - E_j)/kT}\right) \tag{1.52}$$

and from (1.48)

$$\Omega = \sum_j \Omega_j = -kT \sum_j \ln\left(1 + e^{(\mu - E_j)/kT}\right) \quad . \tag{1.53}$$

To see how to interpret this result we calculate the mean particle number,

$$\bar{N} = -\left(\frac{\partial\Omega}{\partial\mu}\right)_{V,T} \tag{1.54}$$

and internal energy

$$\begin{aligned} U &\equiv \frac{1}{\mathcal{Z}} \sum_N e^{\mu N/kT} \sum_\nu E_\nu \, e^{-E_\nu/kT} \\ &= \mu \bar{N} + kT^2 \frac{\partial}{\partial T} \ln \mathcal{Z} = \mu \bar{N} + \Omega - T \frac{\partial}{\partial T} \Omega \end{aligned}$$

and find

$$\bar{N} = \sum_j \frac{e^{(\mu - E_j)/kT}}{1 + e^{(\mu - E_j)/kT}}$$

so

$$\bar{N} = \sum_j f(E_j) \tag{1.55}$$

with

$$f(E_j) = \frac{1}{1 + e^{(E_j - \mu)/kT}} \quad . \tag{1.56}$$

Similarly we obtain

$$U = \sum_j E_j f(E_j) \quad . \tag{1.57}$$

Equations (1.55), (1.56) and (1.57) are our principle results: We can apparently interpret $f(E_j)$ as the thermal average occupation number of the single particle state with energy E_j. In the continuum approximation, (1.55) and (1.57) become

$$\bar{N} = \int_0^\infty dE \, \rho(E) f(E) \tag{1.58}$$

$$U = \int_0^\infty dE \, E \, \rho(E) f(E) \tag{1.59}$$

$$\Omega = kT \int_0^\infty dE \, \rho(E) \ln \left(1 - f(E)\right)$$

where $\rho(E)$ is the density of states (1.3), (1.58) is to be understood as an equation for the chemical potential, μ, as a function of $\bar{N}, T$ and V. Equation (1.59) is one example of many which express the thermal properties of a Fermi system as an integral over $\rho(E)f(E)$. Before proceeding with the development of the thermodynamics we introduce an approximation scheme for $f(E)$.

2. Low Temperature Expansion for $f(E)$.

At $T = 0$, $f(E)$ is a step function

$$f(E) = \begin{cases} 1 & 0 < E < \mu \\ 0 & E > \mu \quad . \end{cases} \tag{1.60}$$

Its derivative is a δ-function at $E = \mu$,

$$\frac{df}{dE} = f'(E) = -\delta(\mu - E) \tag{1.61}$$

and $f(E)$ can be reconstructed from (1.61) plus the condition $f(0) = 1$. At low temperatures $f'(E)$ is no longer a δ-function at $E = \mu$, but it is still very strongly peaked near $E = \mu$. $f(E)$ and $f'(E)$ have been plotted in Fig. 1.4 for various values of kT including cases where $kT \ll \mu$. In the limit $kT \ll \mu$, integrals of smooth functions multiplying $f'(E)$ are determined entirely by the properties of the smooth function (its value and its derivatives) in the vicinity of $E = \mu$, where $f'(E)$ varies violently. The simplest way to evaluate such integrals, of the form[1]

$$I[g] = \int_0^\infty dE\, f'(E) g(E) \tag{1.62}$$

for smooth functions $g(E)$, makes use of a "generalized function" expansion for $f'(E)$. This method has wide application elsewhere in physics, so we take the time to develop it here. What we are after is a power series expansion of $I[g]$ in kT/μ, which is a small parameter at low temperature (compared to density). What we will obtain is a power series in kT/μ — technically an asymptotic expansion — which is accurate up to corrections which vanish exponentially, like $e^{-1/T}$, as $T \to 0$.

The idea is to replace $f'(E)$ by a series of δ-functions and derivatives, of the form

$$f'(E) = -\delta(\mu - E) + \sum_{j=1}^{\infty} c_j (kT)^j \delta^{(j)}(\mu - E) \tag{1.63}$$

where $\delta^{(j)}(x)$ is the j^{th} derivative of the δ-function,

$$\delta^{(j)}(x) = \frac{d^j}{dx^j}\delta(x) \quad . \tag{1.64}$$

Some properties of $\delta^j(x)$ are obtained in the problems. For our purpose, it is enough to know:

$$\begin{aligned} \delta^{(j)}(x) &= 0 \quad \text{for} \quad x \neq 0 \\ \int dx\, \delta^{(j)}(x) g(x) &= (-1)^j g^{(j)}(0) \end{aligned} \tag{1.65}$$

where the latter is valid for functions $g(x)$ which are smooth (all derivatives exist) in the neighborhood of $x = 0$ ($g^{(j)}(x) \equiv d^j/dx^j\, g(x)$). Note that Eq. (1.65) is just what one would expect for derivatives of the δ-function combined with integration-by-parts. Let us use (1.63) – (1.65) to evaluate $I[g]$

$$I[g] = -g(\mu) + \sum_{j=1}^{\infty} (-1)^j c_j (kT)^j g^{(j)}(\mu) \quad . \tag{1.66}$$

[1]We shall see that we need only integrals of $f'(E)$, not $f(E)$, for thermal properties.

So the virtue of this method is that the $\{c_j\}$ are properties of $f(E)$. Once they are known, any integral of the form of $I[g]$ can be performed by the (much easier) operation of differentiating $g(E)$.

To calculate the coefficients c_j, we choose very simple forms for $g(x)$ and calculate the integral against $f'(E)$ explicitly. Now

$$f'(E) = -\frac{1}{kT}\ \frac{e^{(E-\mu)kT}}{\left[1+e^{(E-\mu)/kT}\right]^2} \tag{1.67}$$

is easily seen to be *even* in $(E-\mu)$.[2] This tells us that $c_j = 0$ for j-odd, so

$$f'(E) = -\delta(\mu - E) + c_2(kT)^2\delta''(\mu - E) + c_4(kT)^4\delta^{iv}(\mu - E) + \ldots \tag{1.68}$$

To find c_2 consider $g(E) = (\mu - E)^2$. According to (1.67)

$$\int_0^\infty dE f'(E)(\mu - E)^2 = 2c_2(kT)^2 \tag{1.69}$$

and by explicit calculation (see box),

$$c_2 = -\frac{\pi^2}{6}\quad . \tag{1.70}$$

A slightly longer calculation of the same form yields

$$c_4 \cong -1.8940\quad . \tag{1.71}$$

Putting this together with (1.68) we obtain

$$f'(E) = -\delta(\mu - E) - \frac{\pi^2k^2T^2}{6}\delta''(\mu - E) - 1.8940(kT)^4\delta^{iv}(\mu - E) - \ldots \tag{1.72}$$

which is the primary result of this section.
To perform the integral

$$c_2 = \frac{1}{2k^2T^2}\int_0^\infty dE(\mu - E)^2\left\{-\frac{1}{kT}\ \frac{e^{(\mu-E)/kT}}{\left[1+e^{(\mu-E)/kT}\right]^2}\right\}$$

define $x \equiv (E-\mu)/kT$, so

$$c_2 = -\frac{1}{2}\int_{-\mu/kT}^\infty dx\ \frac{x^2e^x}{\left(1+e^x\right)^2}\quad .$$

[2] This is not quite correct since $f'(E) = 0$ for $E < 0$ but not for $E > 2\mu$. By neglecting this we are making errors which are exponentially small ($\sim e^{-\mu/kT}$) at low temperature.

At the cost of an error of order $e^{-\mu/kT}$ we can replace the lower limit by $x = -\infty$ and use $f'(x) = -f'(-x)$,

$$c_2 = -\int_0^\infty dx \frac{x^2 e^{-x}}{(1+e^{-x})^2} \quad .$$

Now, expand $(1+e^{-x})^{-2} = 1 - 2e^{-x} + 3e^{-2x} - 4e^{-3x} \ldots$, so

$$\begin{aligned} c_2 &= -\sum_{a=1}^{\infty} (-1)^{a+1} a \int_0^\infty x^2 \, e^{-ax} \\ &= -2\left[1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} \ldots\right] = -\frac{\pi^2}{6} \end{aligned}$$

where the final step is a standard result found in textbooks and tables. The generalization to arbitrary j is straightforward,

$$c_j = -2\sum_{a=1}^{\infty} \frac{(-1)^{a+1}}{a^j} = -2\left(1 - 2^{1-j}\right)\zeta(j) \quad .$$

Here $\zeta(j)$ is Riemann's zeta function ($\zeta(s) \equiv \sum_{k=1}^{\infty} k^{-s}$) which is tabulated in many books. $\zeta(2n)$ is related to a Bernoulli number, $\zeta(2n) = (2\pi)^{2n}|B_{2n}|/2(2n)!$ and $B_0 = 1$, $B_2 = 1/6$, $B_4 = -1/30 \ldots$ For large j, c_j converges quickly to -2.

3. Thermodynamic Properties at Low Temperature

Let us begin by calculating $\bar{N}$ as a function of μ, V and T. From (1.58)

$$\bar{N} = \int_0^\infty dE \, \rho(E) f(E) \tag{1.58}$$

where $\rho(E) = dN/dE$ is given by (1.3). Integrating by parts,

$$\bar{N} = N(E) f(E)\Big|_0^\infty - \int_0^\infty dE \, N(E) f'(E) \quad . \tag{1.73}$$

The surface term vanishes because $\lim_{E\to\infty} f(E) = 0$ and $N(0) = 0$. Upon substituting for $f'(E)$, we obtain

$$\bar{N} = N(\mu) + \sum_{j=1}^{\infty} (-1)^j c_j (kT)^j N^{(j)}(\mu) \quad . \tag{1.74}$$

At low temperatures we can keep only the first non-trivial term, of order T^2,

$$\bar{N} \cong N(\mu) + \frac{\pi^2 k^2 T^2}{6} N''(\mu) \quad . \tag{1.75}$$

Referring back to (1.3)

$$N(\mu) = \frac{g}{6\pi^2}\left(\frac{2m\mu}{\hbar^2}\right)^{3/2} V \tag{1.76}$$

and

$$N''(\mu) = \frac{g}{8\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\frac{V}{\sqrt{\mu}} \tag{1.77}$$

so

$$\bar{N} \cong \frac{mg}{3\pi^2\hbar^2}\left(\frac{2m}{\hbar^2}\right)^{1/2} V\left(\mu^{3/2} + \frac{\pi^2 k^2 T^2}{8}\,\frac{1}{\mu^{1/2}}\right) \quad . \tag{1.78}$$

To the same accuracy, keeping terms through order T^2, we can invert (1.78) and obtain the chemical potential as a function of the mean density, $\bar{N}/V$ and the temperature:

$$\mu(T) \cong \mu(0) - \frac{\pi^2 k^2 T^2}{12}\left(\frac{1}{\mu(0)}\right) \tag{1.79}$$

where

$$\mu(0) \equiv E_F = \frac{\hbar^2 k_F^2}{2m} = \frac{\hbar^2}{2m}\left(\frac{6\pi^2\bar{N}}{gV}\right)^{2/3} \quad . \tag{1.80}$$

$\mu(T)$ is the energy at which the fermion occupation number is on average one-half. According to (1.79) this measure of the degeneracy of the Fermi system drops with temperature. As we promised in §1, the natural scale of temperatures in the degenerate Fermi system is $T_F \equiv E_F/k$. Degeneracy effects are important at temperatures below T_F.

In a similar fashion we can calculate the internal energy defined by (1.59) and obtain

$$U \cong \frac{gV}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\left\{\frac{2}{5}\mu^{5/2} + \frac{1}{4}\pi^2 k^2 T^2 \mu^{1/2}\right\} \tag{1.81}$$

through order T^2. It is more useful to have U as a function of $\bar{N}$, T and V rather than μ, T and V, so we substitute (1.79) for $\mu = \mu(T, \bar{N}, V)$ and find

$$U \cong \frac{gV}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\left\{\frac{2}{5}\mu(0)^{5/2} + \frac{1}{6}\pi^2 k^2 T^2 \mu(0)^{1/2}\right\} \quad . \tag{1.82}$$

From (1.82) we can easily obtain the heat capacity at constant volume

$$C_V \equiv \left(\frac{\partial U}{\partial T}\right)_{\bar{N},V} \cong \bar{N}k\left(\frac{\pi^2}{2}\,\frac{T}{T_F}\right) \quad . \tag{1.83}$$

Notice that C_V goes to zero at low temperatures and does not approach the classical result ($C_V = \frac{3}{2}\bar{N}k$) until $T \sim T_F$. The small heat capacity of a degenerate Fermi gas can be understood as a consequence of the limited role of fermions below the Fermi surface, $E \cong E_F$. Particles with $|E_F - E| \gg kT$ cannot be excited thermally because very few unoccupied states are available to them. They are effectively "frozen out" of the thermal dynamics, and don't contribute to C_V. In fact, one can estimate C_V by

arguing that the heat capacity per "active" particle is $(3/2)k$, but that only a fraction of the total number of particles, $\Delta N/N$, estimated by

$$\frac{\Delta N}{N} \sim \frac{dN}{dE}\,\frac{\Delta E}{N} \sim \frac{1}{N}\left(\frac{dN}{dE}\right) kT \sim \frac{T}{T_F}$$

contributes. This is an example of a phenomenon known as "Pauli blocking," in which most of the particles in a degenerate Fermi gas do not respond to small external perturbations, in this case thermal.

The practical consequences of (1.83) are well-known in everyday life. The small heat capacity of the electrons in good conductors is an example we shall encounter shortly.

Finally let us construct the equation of state of a free Fermi gas. We calculate the pressure from (1.42), (1.46), (1.48) and (1.53)

$$\begin{aligned} P \qquad &= -\left(\tfrac{\partial\Omega}{\partial V}\right)_{\mu,T} = -\tfrac{\partial}{\partial V}\left\{-kT\ln\mathcal{Z}\right\} \\ &= kT\,\tfrac{\partial}{\partial V}\int_0^\infty dE\,\rho(E)\ln\left(1+e^{(\mu-E)/kT}\right) \quad . \end{aligned} \tag{1.84}$$

The only volume dependence in (1.84) (at fixed T and μ) is in $\rho(E)$, which is linearly proportional to V, so

$$P = \frac{kT}{V}\ln\mathcal{Z} = -\frac{\Omega}{V} \quad . \tag{1.85}$$

To evaluate PV, we apply the low temperature expansion to Ω

$$\begin{aligned} \Omega \;&=\; -kT\int_0^\infty dE\frac{dN}{dE}\ln\left(1+e^{(\mu-E)/kT}\right) \\ &=\; -kT\left\{N(E)\ln\left(1+e^{(\mu-E)/kT}\right)\Big|_0^\infty + \frac{1}{kT}\int_0^\infty dE\,N(E)f(E)\right\} \quad . \end{aligned} \tag{1.86}$$

The surface term vanishes at both limits, so

$$\Omega = \int_0^\infty dE\,N(E)f(E) \quad . \tag{1.87}$$

Defining $N(E) = d\mathcal{N}/dE$ and integrating by parts again

$$\Omega = -\int_0^\infty dE\,\mathcal{N}(E)f'(E) \quad . \tag{1.88}$$

Using (1.72) for $f'(E)$ and (1.21) for $N(E)$ we find

$$-\Omega = PV = P_0V\left(1+\frac{5}{12}\,\frac{\pi^2k^2T^2}{E_F^2}\right) \tag{1.89}$$

where $P_0 = 2/3\left(gp_F^5/20m\pi^2\hbar^2\right)$ is the pressure we calculated in §1.1 for zero temperature. Once again, the effect of quantum mechanics and the exclusion principle are all important. The equation of state looks nothing like the classical ideal gas: the pressure does not vanish at low temperatures but instead reduces to the quantum degeneracy pressure. Thermal effects increase P quadratically with T, at least for temperatures low in comparison with E_F/k.

Chapter 2

MATTER AT HIGH DENSITY

Strong attractive forces are able to raise matter to very high densities at which the degeneracy of fermions plays a major role in providing stability. The classic examples are gravitationally bound systems: white dwarf and neutron stars in which the degeneracy pressure of electrons and neutrons, respectively, balances the attractive force of gravity. Another example may be quark matter: a degenerate Fermi gas of quarks held together either by the strong chromodynamic forces between quarks or by strong gravitational forces in the cores of neutron stars. In this section we apply the methods of §1 to the study of these systems.

Our first task is to study the composition and the equation of state of cold matter at densities above ordinary ones. This discussion can be carried out without regard to how matter is brought to such extreme conditions. After tracing matter from densities of a few grams per cm^3 up to enormous densities of order 10^{20} gm/cm^3, we turn to the issue of how such extraordinary densities might arise in Nature. First we study white dwarf stars in some detail, then we look very briefly at neutron stars. Much of this discussion is based on the presentation in two excellent texts: Landau and Lifshitz, *Statistical Physics* and Shapiro and Teukolsky, *Black Holes, White Dwarfs and Neutron Stars*. Finally we explore the dynamics of cold quark matter.

2.1 A BRIEF SURVEY OF THE COMPOSITION OF MATTER AT HIGH DENSITY

We wish to follow the history of matter as a function of the density for high densities and low temperatures. We ignore for now the external agent which has brought the system to high densities. We also ignore thermal effects, which is justified provided

$T \ll T_F$, where T_F is the Fermi temperature appropriate to whatever constituents and density we are studying.

Ordinary matter remains qualitatively unchanged as the density is increased until the volume per atom becomes smaller than atomic dimensions. Then atoms lose their individuality and matter begins to resemble a gas of electrons and nuclei. A measure of this transition is the density (p_1) at which the Fermi energy of an electron gas is comparable to the electrostatic interaction between electrons and nuclei:

$$\langle \frac{p_F^2}{2m_e} \rangle \approx \langle \frac{Z\, e^2}{r} \rangle \quad , \tag{2.1}$$

where Z is the nuclear charge. Using (1.7)

$$n = \frac{1}{3\pi^2}\, \frac{p_F^3}{\hbar^3} \tag{1.7}$$

and estimating very crudely

$$\langle \frac{1}{r} \rangle \approx \left(\frac{4\pi n}{3} \right)^{1/3} \tag{2.2}$$

we find

$$n \approx 4 \left(\frac{2m_e e^2}{3\pi\hbar^2} Z \right)^3 \quad . \tag{2.3}$$

A more sophisticated estimate based on the Thomas–Fermi approximation (see §4) takes Coulomb interactions into account in determining $\langle 1/r \rangle$ and gives

$$n \approx \left(\frac{m_e\, e^2}{\hbar^2} \right)^3 Z^2 \quad . \tag{2.4}$$

To obtain the corresponding mass density (ρ) we must introduce a parameter Y_e, which measures the electron to nucleon ratio in the material

$$\rho = \frac{m_p n}{Y_e} \quad . \tag{2.5}$$

$Y_e = 1$ for ordinary hydrogen, $Y_e = 1/2$ for nuclei with equal numbers of protons and neutrons and $Y_e = 0.46$ for ^{56}Fe, the most stable atomic nucleus. [In (2.5) we ignore the small effects of nuclear binding on the mass of the system.] Combining (2.4) and (2.5)

$$\rho_1 \approx \frac{m_p}{Y_e} \left(\frac{m_e e^2}{\hbar^2} \right)^3 Z^2 \tag{2.6}$$

$(\hbar^2/me^2) = 0.529 \times 10^{-9}$ cm is the Bohr radius. As an example, for $^{56}_{26}$Fe, $\rho_1 \approx 1.66 \times 10^4$ gm/cm^3. This density corresponds to an electron Fermi energy of ≈ 10 KeV and a nuclear Fermi energy (for iron) of ≈ 1 eV. So we have reason to believe electrons form

a degenerate Fermi gas at densities in excess of ρ_1 provided the ambient temperature is below 10 KeV, whereas the gas of nuclei is not degenerate unless the temperature is far lower. At these densities the electrons exert significant degeneracy pressure, from (1.24),

$$P = \frac{p_F^5}{15m\pi^2\hbar^3} = \frac{3}{5}\left(3\pi^2\right)^{2/3}\frac{\hbar^2}{m_e}n^{5/3} \quad . \tag{2.7}$$

The nuclei contribute insignificantly to the pressure because their masses are so much larger than m_e. On the other hand, the nuclei dominate the mass density of the system.

As the density is increased the electron Fermi energy increases and the electron Fermi gas becomes relativistic. A measure of the density at which relativistic effects become important for electrons can be obtained by setting $p_F = m_e c$. The corresponding mass density is

$$\rho_2 = \frac{m_p}{3\pi^2 Y_e}\left(\frac{m_e c}{\hbar}\right)^3 \quad . \tag{2.8}$$

Again taking ${}^{56}_{26}$Fe as an example, $\rho_2 = 2.13 \times 10^6$ g/cm^3.

At yet higher densities it becomes energetically favorable for electrons at the top of the Fermi sea to react with nuclei via reactions which absorb electrons. If ${}_Z A$ is a nuclear species with charge Z and mass number A, the sort of reaction we have in mind is

$$e^- + {}_Z A \rightarrow {}_{Z-1}A + \nu_e \quad , \tag{2.9}$$

a typical weak interaction known as inverse β-decay. The massless neutrino is not bound to matter and is radiated away. Exactly which nuclear reactions occur depends on the composition of the material. For the sake of definiteness let's study the case of a gas of protons, neutrons and electrons. These can react with one another via β-decay

$$n \rightarrow p + e^- + \bar{\nu}_e \tag{2.10}$$

and its inverse

$$p + e^- \rightarrow n + \nu_e \quad . \tag{2.11}$$

Because protons are lighter than neutrons,

$$Q \equiv (m_n - m_p)c^2 = 1.293\,\mathrm{MeV} \quad , \tag{2.12}$$

at low densities the system consists entirely of protons and electrons. Neutrons first appear when the electron Fermi energy exceeds Q, for then it is energetically favorable

for an electron at the top of the Fermi sea to combine with a proton via (2.11). This corresponds to a matter density of

$$\rho_3 = \frac{m_p}{3\pi^2 Y_e}\frac{(Q^2 - m_e^2c^4)^{3/2}}{\hbar^3} = 1.23 \times 10^7\,\mathrm{gm/cm}^3 \quad . \tag{2.13}$$

The exact composition of a real material depends on the Q-values of thousands of nuclear β-decay reactions. The first serious attempt to model this system was made by Harrison and Wheeler in the 1950's. Their work was based on a simple, semi-empirical mass formula for nuclei. Later Salpeter and Baym, Pethick and Sutherland (BPS) improved on the Harrison–Wheeler formalism. The result of their work is as follows: At densities above ρ_3 electrons are absorbed by nuclei forming neutron rich nuclei. There are, however, limits to the neutron-to-proton ratio of nuclei above which nuclei become unstable to neutron emission, a phenomenon known as neutron drip. This limit is reached, according to BPS, at a density of

$$\rho_4 \approx 2.4 \times 10^{11}\,\mathrm{g/cm}^3 \quad . \tag{2.14}$$

At densities above ρ_4, matter consists of a gas of ultra-relativistic electrons, non-relativistic neutrons and nuclei at the neutron rich limit of stability. Eventually, the neutron degeneracy pressure replaces the electron degeneracy pressure as the dominant pressure. The two are equal at

$$\rho_5 \approx 10^{12}\,\mathrm{gm/cm}^3 \quad . \tag{2.15}$$

This marks the beginning of a density regime in which matter may be regarded as essentially a degenerate non-relativistic gas of neutrons. While ρ_5 is an enormous density by ordinary standards, it is still far below the density of ordinary nuclei: The radii of ordinary nuclei are well-described by the simple rule $R(A) = 1.2\,A^{1/3} \times 10^{-13}$ cm corresponding to a density of

$$\rho_6 \approx 2.3 \times 10^{14}\,\mathrm{gm/cm}^3 \quad . \tag{2.16}$$

So it is reasonable to treat matter in the range $\rho_5 < \rho < \rho_6$ as a diffuse, weakly interacting gas of neutrons, at least as a first approximation. Even at nuclear matter densities this gas of neutrons is not very relativistic: The Fermi momentum corresponding to ρ_6 is

$$p_F^n c \approx 320\,\mathrm{MeV}$$

so

$$\frac{p_F^2}{m_n^2c^2} \approx \frac{v^2}{c^2} \approx 0.12 \quad .$$

The description of matter in terms of neutrons and protons begins to break down when the separation between neutrons becomes small compared to their intrinsic size, which is known to be of order 10^{-13} cm. We must then turn to the quarks which compose protons and neutrons. At densities above 1 nucleon per cubic fermi (1 fermi $\equiv 10^{-13}$ cm), or

$$\rho_7 \sim 1.7 \times 10^{15}\,\mathrm{fm/cm}^3 \quad , \tag{2.17}$$

a description in terms of quarks appears necessary. Quarks come in several varieties with rest mass (we quote mc^2) ranging from about 5 MeV and 10 MeV for the up and down quark, upwards through $\sim$ 150 MeV for the strange quark, to $\sim$ 1,500 MeV, $\sim$ 5,000 MeV and 1.7×10^5 MeV for the charm bottom and top quarks. We believe, then, that matter at densities exceeding ρ_7 consists of a degenerate Fermi gas of quarks in equilibrium with the weak interactions — which turn quarks of different types into one another. This description of matter is believed to be correct up to Fermi energies of at least hundreds of GeV (1 GeV = 10^3 MeV) corresponding to densities of 10^{20} gm/cm^3!

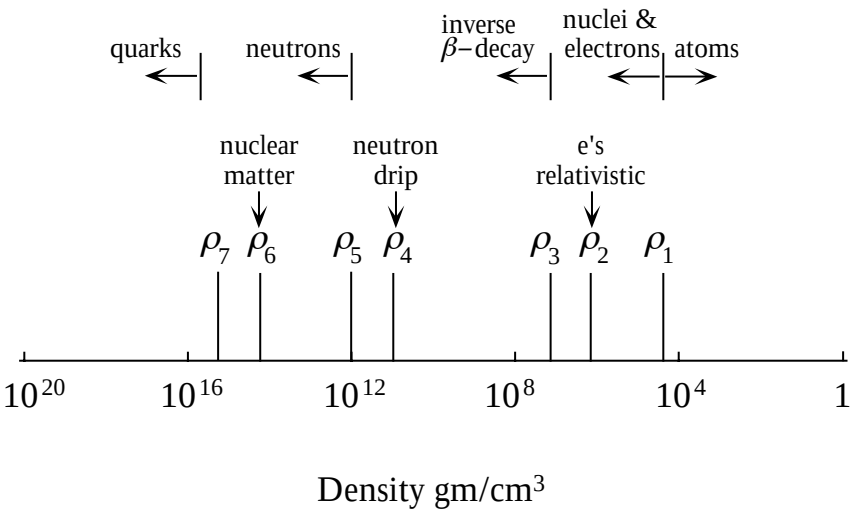


Figure 2.1: Landmarks in the description of matter as a function of density, ranging from a gas of quarks at arbitrarily high density to a diffuse gas of atoms at arbitrarily low density.

This rather grand survey of the character of dense matter is summarized in Fig. 2.1. Which, if any, of the scenarios introduced above actually occurs in Nature depends on the dynamical history of stars or other astrophysical objects capable of producing the gravitational pressure necessary to raise matter to such enormous density. There is ample evidence that relativistic electrons provide the degeneracy pressure which support white dwarf stars and that the neutrons do the same for neutron stars.

2.2 WHITE DWARFS — INTRODUCTION

The actual path history of stellar evolution is a complex and not entirely understood subject. We will give only a cartoon sketch to motivate our study of white dwarfs and neutron stars. The primary energy source for young stars is hydrogen fusion. The thermal pressure this generates counteracts the attractive forces of gravity. When its hydrogen supply is depleated a star begins to contract. If sufficient heat is generated from gravitational potential energy then the stellar core temperature will rise to the point when helium burning begins. Helium is present both as a primordial constituent of stars and as a product of hydrogen burning. Cycles of burning and contraction can, in principle, continue until the star's nuclear fuel is exhausted, that is, until it consists mainly of iron, the most stable (as measured in binding energy per nucleon) of nuclei. During this time, as the star contracts, electron degeneracy pressure increases. If the degeneracy pressure is great enough it can oppose further contraction and arrest further cycles of nuclear burning and contraction. It appears that relatively low mass stars ($M \lesssim 1.4\, M_\odot$) stop burning nuclear fuel at some stage before complete conversion to iron. When they reach this state, stars slowly cool as they radiate away their residual energy. These stars are known as "white dwarfs"; they have masses of order $M_\odot$, radii of order 5,000 km and mean densities of order 10^6 gm/cm^3, in the range where their dynamics is dominated by electron degeneracy pressure.

The history of the discovery and understanding of white dwarfs is a particularly interesting one. A very brief summary is given in the material transcribed below from Shapiro and Teukolsky's book, *Black Holes, White Dwarfs and Neutron Stars*:

> White dwarfs are stars of about one solar mass with characteristic radii of about 5,000 km and mean densities of around 10^{-6} g cm^{-3}. These stars no longer burn nuclear fuel. Instead, they are slowly cooling as they radiate away their residual thermal energy.
>
> We know today that white dwarfs support themselves against gravity by the pressure of degenerate electrons. This fact was not always clear to astronomers, although the compact nature of white dwarfs was readily apparent from early observations. For example, the mass of Sirius B, the binary companion to Sirius and the best known white dwarf, was determined by applying Kepler's Third Law to the binary star orbit. Early estimates placed its mass M in the range from $0.75\, M_\odot$ to $0.95\, M_\odot$. Its luminosity L was estimated from the observed flux and known distance to be about 1/360 of that of the sun. In 1914 W. S. Adams (1915) made the surprising discovery that the spectrum of Sirius B was that of a "white star," not very different from its normal companion, Sirius. By assigning an effective temperature of $8{,}000\, K$ to Sirius B from

these spectral measurements and using the equation for blackbody emission, $L = 4\pi R^2 \sigma T_{\rm eff}^4$, a radius R of 18,800 km could be inferred for the star. (This is about four times bigger than the modern value.)

Referring to Sirius B in his book *The Internal Constitution of the Stars*, the great astrophysicist Sir Arthur Eddington (1926) concluded that "we have a star of mass equal to the sun and of radius much less than Uranus." He also reported in his book the extraordinary new measurements by W. S. Adams (1925) of the gravitational redshifts of several spectral lines emitted from the surface of Sirius B. By applying the theory of general relativity, the ratio M/R could be inferred from the measured redshifts. As the mass was already known from the binary orbit, the radius of Sirius B could be determined. The redshifts obtained by Adams, though crude, confirmed the previous estimates of R and the compact nature of the white dwarf. Eddington (1926) thus wrote in his book that "Prof. Adams has killed two birds with one stone; he has carried out a new test of Einstein's general theory of relativity and he has confirmed our suspicion that matter 2,000 times denser than platinum is not only possible, but is actually present in the Universe."[1]

Eddington (1926) went on to argue that although only three white dwarfs could be firmly established at that time, white dwarfs are probably very abundant in space, since the known ones were all very close to the sun. But regarding the means by which white dwarfs supported themselves against collapse, Eddington could declare only that "it seems likely that the ordinary failure of the gas laws due to finite sizes of molecules will occur at these high densities, and I do not suppose that the white dwarfs behave like perfect gas."

In August of 1926, Dirac (1926) formulated Fermi–Dirac statistics, building on the foundations established only months earlier by Fermi. In December 1926, R. H. Fowler (1926), in a pioneering paper on compact stars, applied Fermi–Dirac statistics to explain the puzzling nature of white dwarfs: he identified the pressure holding up the stars from gravitational collapse with *electron degeneracy pressure*.

Actual white dwarf models, taking into account special relativistic effects in the degenerate electron equation of state, were constructed in 1930 by Chandraskhar (1931a,b). In the course of this analysis, Chandrasekhar (1931b) made the momentous discovery that white dwarfs had a *maximum* pass of $\sim 1.4\, M_\odot$, the exact value depending on the composition of the matter. This

[1] Adams' value for the redshift, and also the separate value obtained by Moore, agreed with general relativity, although using an incorrect observational value for the radius. The modern situation is described in Shapiro and Teukolsky Section 3.6.

maximum mass is called the *Chandrasekhar limit* in honor of its discoverer. Chandrasekhar (1934) was immediately aware of the important implication of his finding, for he wrote in 1934: "The life history of a star of small mass must be essentially different from the life history of a star of large mass. For a star of small mass the natural white-dwarf stage is an initial step towards complete extinction. A star of large mass cannot pass into the white-dwarf stage and one is left speculating on other possibilities."

In 1932, L. D. Landau (1932) presented an elementary explanation of the Chandrasekhar limit. He applied his argument several months later to neutron stars when he learned of the discovery of the neutron (Section 9.1).

The role of general relativity in modifying the mass-radius relation for massive white dwarfs above $1\,M_\odot$ was first discussed by Kaplan (1949). He concluded that general relativity probably induces a dynamical instability when the radius becomes smaller than 1.1×10^4 km. The general relativistic instability for white dwarfs was discovered independently by Chandrasekhar in 1964.[2]

2.3 WHITE DWARFS — DYNAMICS

In a cold white dwarf, electron degeneracy pressure balances gravity. Because electrons are so much less massive than nucleons ($m_e \cong (1/2{,}000)m_{p,n}$, the electrons contribute negligibly to the star's mass and the nuclei contribute negligibly to the degeneracy pressure. The densities of electrons and nuclei scale together according to (2.5), however, because electrostatic forces will keep the material locally neutral on average. For simplicity we assume spherical symmetry and ignore stellar rotation, so that n and ρ are functions of r, the distance from the star's center, alone.

In equilibrium, the gravitational force on a small volume element must be countered by Fermi pressure. Consider the volume element shown in Fig. 2.2, consisting of a slice between r and $r + dr$ with cross section A. The gravitational force on this element is

$$dF = -\frac{Gm(r)\rho(r)}{r^2}Adr \tag{2.18}$$

where $m(r)$ is the mass enclosed within a radius r:

$$m(r) = 4\pi \int_0^r dr' r'^2 \rho(r') \quad . \tag{2.19}$$

[2]See Chandrasekhar and Tooper (1964).

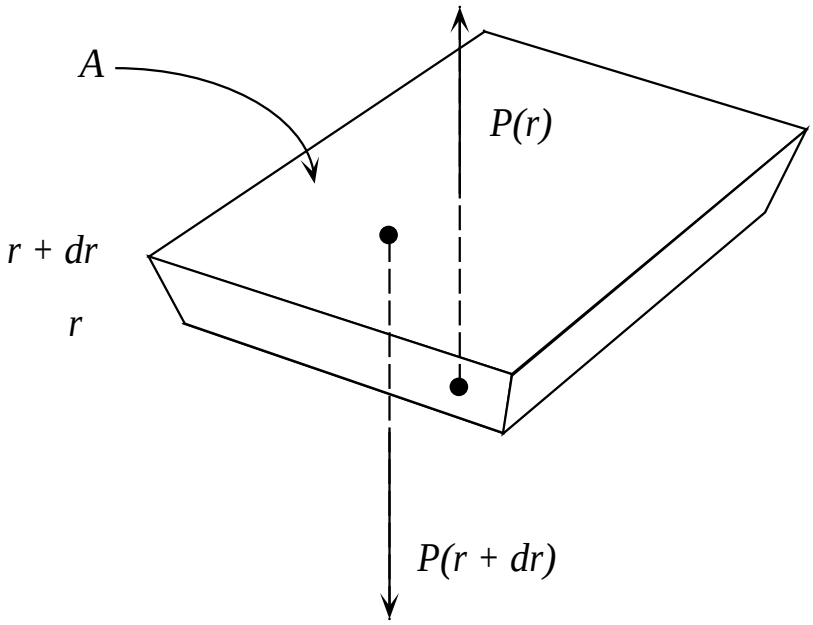


Figure 2.2: Hydrostatic equilibrium — the force of gravity is balanced by a pressure gradient in equilibrium.

Let $P(r)$ be the pressure profile of the star. The outward force on the area A due to this pressure at r is $F(r) = P(r)A$; the inward force at $r+dr$ is $F(r+dr) = P(r+dr)A$. The net outward force must balance gravity:

$$\frac{dP}{dr} = -\frac{Gm(r)\rho(r)}{r^2} \quad . \tag{2.20}$$

Equations (2.19) and (2.20) can be combined into a single equation involving $P(r)$ and $\rho(r)$:

$$\frac{1}{r^2}\,\frac{d}{dr}\,\frac{r^2}{\rho(r)}\,\frac{d}{dr}P(r) = -4\pi G\rho(r) \quad . \tag{2.21}$$

The relation between the pressure, $P(r)$, and the mass density, $\rho(r)$, depends on the equation of state of the degenerate electron gas,

$$\frac{dP}{dr} = \frac{dP}{d\rho}\,\frac{d\rho}{dr} = \frac{Y_e}{m_p}\,\frac{dP}{dn}\,\frac{d\rho}{dr} \quad . \tag{2.22}$$

And dP/dn is given by (1.25)

$$\frac{dP}{dn} = \frac{1}{3}\,\frac{p_F^2 c}{\sqrt{p_F^2 + m_e^2 c^2}} \tag{2.23}$$

where

$$p_F = (3\pi^2 n)^{1/3}\hbar \quad . \tag{2.24}$$

Equation (1.25) applies to a completely degenerate ($T = 0$), free electron gas. Corrections due to interactions and finite temperature effects are small at low temperature and high density. It's convenient to reexpress dP/dn in terms of $\rho(r)$ and, since it has units of energy, measure dP/dn relative to $m_e c^2$,

$$\frac{dP}{dn} = \frac{1}{3} m_e c^2 \frac{(\rho/\rho_0)^{2/3}}{\left[1 + (\rho/\rho_0)^{2/3}\right]^{1/2}} \tag{2.25}$$

where

$$\rho_0 \equiv \left(\frac{m_e c}{\hbar}\right)^3 \frac{m_p}{3\pi^2 Y_e} \tag{2.26}$$

is the density corresponding to a mass of order m_p in a volume of order the electron's Compton wavelength. It was called ρ_2 in the survey of § 2.a. dP/dn has simple non-relativistic and relativistic limits

$$\frac{dP}{dn} = \begin{cases} \frac{1}{3} m_e c^2 \left(\frac{\rho}{\rho_0}\right)^{2/3} & \text{Non} - \text{Relativistic} \\ \frac{1}{3} m_e c^2 \left(\frac{\rho}{\rho_0}\right)^{1/3} & \text{Relativistic} \quad . \end{cases} \tag{2.27}$$

It is straightforward to substitute dP/dn from (2.25) back onto (2.21) and integrate it numerically. Many of the important features of white dwarfs structure can be obtained by analytic means if we approximate dP/dn by a function which is easier to integrate. Since both limits in (2.27) are special cases of the polytropic form of dP/dn, we will try the general case,

$$\frac{dP}{dn} = \frac{1}{3} m_e c^2 \left(\frac{\rho}{\rho_0}\right)^{\gamma - 1} \quad . \tag{2.28}$$

Finally, it is useful to let $(\rho/\rho_0)^{\gamma-1}$ be the independent variable, so we define

$$\kappa\theta(r) \equiv \left(\frac{\rho}{\rho_0}\right)^{\gamma - 1} \tag{2.29}$$

and choose the constant κ so that $\theta = 1$ at $r = 0$, *i.e.*

$$\kappa = \left(\frac{\rho_c}{\rho_0}\right)^{\gamma - 1} \tag{2.30}$$

where ρ_c is the central density. Substituting all of this back into (2.21) we obtain

$$\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} \theta = -\frac{12\pi(\gamma - 1) G m_p \rho_0 \kappa^{\left(\frac{2-\gamma}{\gamma-1}\right)}}{Y_e m_e c^2} \theta^{1/\gamma - 1} \quad . \tag{2.31}$$

To simplify (2.31) we define a scaled length variable

$$r \equiv a\xi \tag{2.32}$$

and choose

$$a^2 \equiv \frac{\pi\hbar^3 Y_e^2}{4(\gamma-1)m_e^2 m_p^2 Gc\kappa^{\frac{2-\gamma}{\gamma-1}}} \tag{2.33}$$

so that

$$\frac{1}{\xi^2}\frac{d}{d\xi}\xi^2\frac{d}{d\xi}\theta = -\theta^{\frac{1}{\gamma-1}} \quad . \tag{2.34}$$

This is known as the Lane-Emden equation. For $\gamma = 5/3$ and $4/3$ it describes the non-relativistic and ultra-relativistic cases precisely. Otherwise, it's a polytropic approximation.

We need the boundary conditions on (2.34). One is already known from (2.29) and (2.30),

$$\theta(0) = 1 \quad . \tag{2.35}$$

Since (2.34) is second order, we need another condition to uniquely fix a solution. To find it, integrate (2.34) once to obtain

$$\theta'(\xi) = -\frac{1}{\xi^2}\int d\xi' \xi'^2 \theta^{\frac{1}{\gamma-1}}(\xi') \tag{2.36}$$

Since θ is finite as $\xi \to 0$ and since $\gamma > 1$ (see §1), we see θ' vanishes at $\xi = 0$,

$$\theta'(0) = 0 \quad . \tag{2.37}$$

Equations (2.35) and (2.37) together uniquely specify the solution to the Lane-Emden equation.

2.4 WHITE DWARFS — CHARACTERISTIC SOLUTIONS

Before discussing solutions in detail we can obtain their gross features from dimensional considerations. For the polytropic equation of state the solution to (2.34) subject to $\theta(0) = 1$, $\theta'(0) = 0$ determines the profile of the star in dimensionless variables. Equations (2.26) and (2.33) restore the natural scales to the problem. The only significant free parameter is the central density, or κ as defined by (2.30). We have already defined a characteristic density

$$\rho_0 = \left(\frac{m_e c}{\hbar}\right)^3 \frac{m_p}{3\pi^2 Y_e} = \frac{1}{Y_e} 9.81 \times 10^5 \,\mathrm{gm/cm}^3 \tag{2.38}$$

and a characteristic length scale

$$R_0 = \left[\frac{3\pi\hbar^3}{4Gm_p^2m_e^2c}\right]^{1/2} Y_e = 7.72\, Y_e \times 10^3\,\mathrm{km} \quad . \tag{2.39}$$

Here we have set $\gamma = 4/3$ and $\kappa = 1$ in (2.33) to have a representative length scale. Finally from (2.19) we obtain a representative mass

$$M_0 \equiv 4\pi R_0^3 \rho_0 = \left(\frac{3\pi\hbar^3c^3}{G^3}\right)^{1/2} \frac{Y_e^2}{m_p^2} = 5.66\, Y_e^2 \times 10^{33}\,\mathrm{gm} \tag{2.40}$$

which is of order $M_\odot$. Equations (2.38) – (2.40) give us an indication of the scale of a collapsed object supported by electron degeneracy pressure. Notice that the mass is particularly simple — it depends only on m_p and on fundamental constants of Nature ($\hbar$, c and G).

The solutions to the Lane–Emden equation, $\theta(\xi)$, decrease monotonically with ξ and vanish at $\xi \equiv \xi_1$ provided $\gamma > 6/5$. Then $R = a\xi_1$ defines the surface of the star.

$$R = \left[\frac{1}{3(\gamma-1)\kappa^{\left(\frac{2-\gamma}{\gamma-1}\right)}}\right]^{1/2} R_0\xi_1 \quad . \tag{2.41}$$

Similarly the mass can be expressed in terms of M_0, ξ_1, κ and γ,

$$\begin{aligned} M &= 4\pi \int_0^R r^2\, dr\, \rho(r) \\ &= 4\pi a^3 \rho_0 \kappa^{1/\gamma-1} \int_0^{\xi_1} d\xi\, \xi^2 \theta(\xi)^{\frac{1}{\gamma-1}} \\ &= -4\pi a^3 \rho_0 \kappa^{\frac{1}{\gamma-1}} \left[\xi_1^2 \theta'(\xi_1)\right] \end{aligned} \tag{2.42}$$

where, in the last step, we used (2.36). Substituting for a, we find

$$M = \left[\frac{\kappa^{3\gamma-4/\gamma-1}}{3(\gamma-1)}\right]^{1/2} M_0 \left[-\xi_1^2\theta'(\xi_1)\right] \quad . \tag{2.43}$$

Values of ξ_1 and $-\xi_1^2\theta'(\xi_1)$ for various γ are given in Table 2.1. Together with

$$\rho_c = \rho_0 \kappa^{1/\gamma-1} \tag{2.44}$$

these equations determine the characteristics of polytropic white dwarfs. The unknown central density parameterized by κ can be eliminated from (2.42) and (2.43) to give a relation between M and R:

$$M \sim \frac{1}{R^{3\gamma-4/2-\gamma}} \tag{2.45}$$

γ	ξ^1	$-\xi_1^2 d\theta/d\xi\Big\vert_{\xi_1}$
5/3	3.65375	2.71406
5/4	4.35287	2.41105
7/5	5.35528	2.18720
4/3	6.89685	2.01824

Table 2.1: Parameters of solutions to the Lane–Emden equation

which gives $M \sim 1/R^3$ for the non-relativistic limit, $\gamma = 5/3$.

This connection between M and R is the key to understanding several important features of white dwarfs. Let us reexamine these results and their implications from a very physical point of view proposed by Landau in 1932. First suppose the electron Fermi gas were always non-relativistic. It is an easy exercise using the methods of §1 to calculate its total kinetic and gravitational potential energies as a function of N (the total number of nucleons — protons and neutrons). We have, for the total kinetic energy

$$U = uV = \frac{\hbar^2 k_F^5}{10m\pi^2} \propto N\frac{p_F^2}{2m} \propto N\left(\frac{N}{V}\right)^{2/3}$$

so

$$U \propto \frac{N^{5/3}}{R^2} \quad , \tag{2.46}$$

and for the gravitational potential energy,

$$\begin{aligned} V_G &= -G\int \frac{m(r)dm(r)}{r} = -G\int_0^R dr\, 4\pi r\rho(r)m(r) \\ V_G &\propto -\frac{N^2}{R} \quad . \end{aligned} \tag{2.47}$$

Comparing (2.46) and (2.47) we see that the sum of the two expressions always has a minimum and that the minimum occurs when $R \sim N^{-1/3}$. Since the total mass scales linearly with N we find $R \sim M^{-1/3}$.

We conclude that a non-relativistic and self-gravitating degenerate Fermi gas always has an equilibrium configuration. It does not undergo gravitational collapse. Of course, when M gets very large and R gets very small, the system becomes very dense and the non-relativistic approximation breaks down. Before studying what happens in this case, it is worth remarking that the result $R \sim M^{-1/3}$ implies a maximum size for cold self-gravitating matter. When M is very small atomic forces keep the

density relatively constant, so $R \sim M^{1/3}$. But when M is very large we've just shown that $R \sim M^{-1/3}$. Somewhere in between R takes on a maximum, $R_{\max}$. Landau and Lifshitz give a simple estimate leading to $R_{\max} \sim 10^5 Y_e/Z^{1/3}$ km, where Y_e and Z characterize the material of interest.

To learn more about non-relativistic white dwarfs, we return to (2.45) and supply the missing proportionality factors for $\gamma = 5/3$:

$$MR^3 = \frac{9\pi^2\hbar^6}{128G^3m_p^5m_e^3}Y_e^5\left[-\xi_1^5\theta'(\xi_1)\right] \tag{2.48}$$

$$= \frac{91.9\hbar^6}{G^3m_e^3m_p^5}Y_e^5 \quad . \tag{2.49}$$

For $Y_e \approx 0.46$,

$$MR^3 \approx 0.9 \times 10^{60}\,\text{gm cm}^3$$

or

$$\frac{M}{M_\odot}\qquad\left(\frac{R}{7{,}700\,\text{km}}\right)^3 = 1 \tag{2.50}$$

This simple mass radius relation governs white dwarfs from the lowest masses, where electrostatic interactions modify it, up to large masses, where the electrons become relativistic. It is shown as the curve marked "Non-relativistic White Dwarfs" in Fig. 2.3. For comparison in Fig. 2.3 we show a realistic plot of M versus R (marked A) in which the composition is assumed in equilibrium with the weak interactions [taken from Shapiro and Teukolsky Fig. 3.1].

Now let us consider the opposite extreme: the completely relativistic, degenerate electron gas. The gravitational potential energy remains the same, $V_G \propto -N^2/R$, but the kinetic energy now scales as $N^{4/3}/R$. Since both terms in the energy scale with the same power of the radius equilibrium is not possible: either the kinetic energy dominates and the system expands until it loses its extreme relativistic behavior, or the potential energy dominates and the system collapses until new dynamics takes over. Since the potential energy goes with a higher power of N than the kinetic energy, larger systems tend toward collapse. The value of N at which the two terms, U and V_G, are equal is a measure of the maximum mass a white dwarf can have. The proper way to calculate this maximum mass is to leave the polytropic approximation and return to the actual equation of state defined by (2.23). The problem must be solved numerically (see Chandrasekhar); however, the result is that near the maximum mass the electrons are ultra-relativistic throughout nearly all of the star. So we can

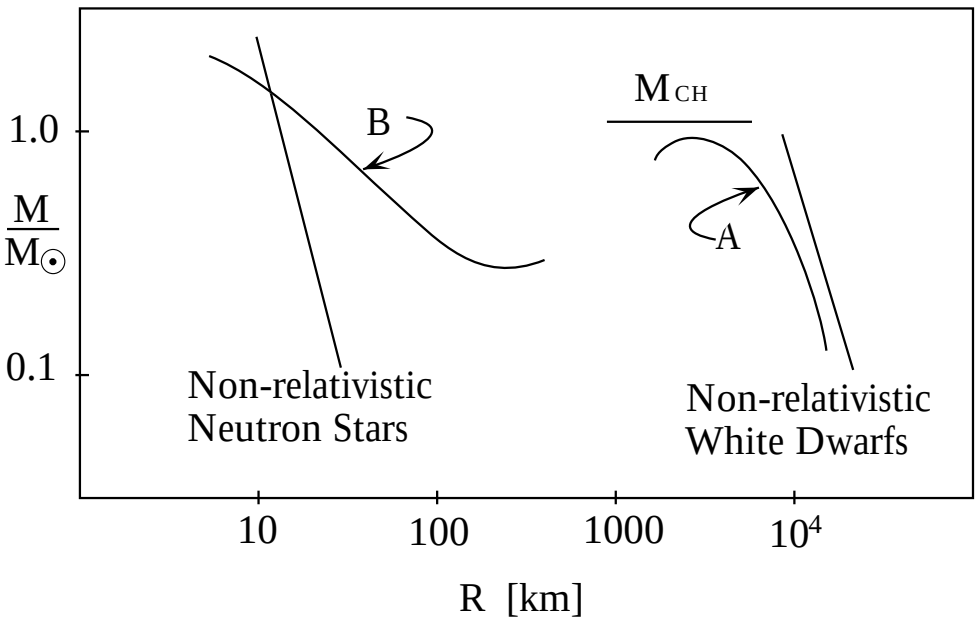


Figure 2.3: Idealized and more sophisticated descriptions of white dwarfs and neutron stars. The curves labeled "A" and "B" were taken from Shapiro and Teukolsky.

estimate the maximum mass by solving the Lane–Emden equation for $\gamma = 4/3$. The result of that simple numerical calculation is $\xi_1 = 6.897$ and $\xi_1^2\theta'(\xi_1) = -2.018$, so

$$R = 3.35 \times 10^4 \,\mathrm{km} \left(\frac{\rho_c}{9.81 \times 10^5 \,\mathrm{gm/cm}^3} \right)^{-1/3} (2Y_e)^{2/3} \tag{2.51}$$

and

$$M = 1.44(2Y_e)^2 M_\odot \quad . \tag{2.52}$$

Equation (2.52) was first derived by Chandrasekhar and is known as the Chandrasekhar limit. It is marked as M_{CH} in Fig. 2.3. The results of integration of the dynamical equation for the exact Fermi gas equation of state can be found in Chandrasekhar's 1939 paper. They agree with the polytropic results for $\gamma = 5/3$ and $4/3$ in the case of very light and very heavy white dwarfs as expected. The very simple dimensional argument of Landau gives a result very close to (2.52). The dynamics of realistic white dwarfs are complicated by many features we have ignored. To mention a few:

1. Electrostatic interactions between electrons and nuclei and electrons and one-another.

2. Thermal processes: the temperature is not zero, nor is it uniform. Heat is transported throughout the system and it is not in thermal equilibrium.

3. The composition is not uniform and changes with time. Inverse β-decay becomes important at large masses.

These and others can be found in the monograph by Shapiro and Teukolsky. The results of a more sophisticated analysis are shown in Fig. 2.3. As is clear from the figure, the simple Fermi gas analysis we have pursued gives a pretty good first approximation for white dwarfs.

2.5 AN INTRODUCTION TO NEUTRON STARS

A star with mass greater than the Chandrasekhar limit cannot form a white dwarf. Either it expels enough mass during some cataclysmic event (a nova) or it collapses beyond the white dwarf domain to higher density regime. The possibility that there might be another regime of stable stars at density for greater than that of a white dwarf was perceived by Landau shortly after the discovery of the neutron. He apparently applied the stability argument he had developed for white dwarfs to a cold, degenerate neutron gas. Credit for the idea of neutron stars goes, however, to Baade and Zwicky, who in 1934 not only described their properties in detail but suggested that they would be formed in supernova explosions. Once again we quote liberally the historical background given in Shapiro and Teukolsky:

> In 1934 Baade and Zwicky proposed the idea of neutron stars, pointing out that they would be at very high density and small radius, and would be much more gravitationally bound than ordinary stars. They also made the remarkably prescient suggestion that neutron stars would be formed in supernova explosions.[3]
>
> The first calculation of neutron star models was performed by Oppenheimer and Volkoff (1939), who assumed matter to be composed of an ideal gas of free neutrons at high density. Work on neutron stars at this time focused mainly on the idea the neutron cores in massive normal stars

[3]Baade and Zwicky (1934): "With all reserve we advance the view that supernovae represent the transitions from ordinary stars into *neutron stars*, which in their final stages consist of extremely closely packed neutrons."

According to Rosenfeld (1974), on the day that word came to Copenhagen from Cambridge telling of Chadwick's discovery of the neutron in 1932, he, Bohr, and Landau spend the evening discussing possible implications of the discovery. It was then that Landau suggested the possibility of cold, dense stars composed principally of neutrons. Landau's only publication on the subject was concerned with neutron cores (Landau, 1938).

might be a source of stellar energy. When this motivation faded as the details of thermonuclear fusion become understood, neutron stars were generally ignored by the astronomical community for the next 30 years. However this was by no means universally so. For example, the papers of Harrison, Wakano, and Wheeler (1958), Cameron (1959a), Ambartsumyan and Saakyan (1960), and Hamada and Salpeter (1961) contain detailed discussions of the equation of state and neutron star models, and the book by Harrison, Thorne, Wakano and Wheeler (1965) contains an extensive discussion. A reason often given for the neglect of the neutron star idea is that because of their small area, their residual thermal radiation would be too faint to observe at astronomical distances with optical telescopes.

However, the discovery of cosmic, nonsolar X-ray sources by Giacconi *et al.*[4] in 1962 did generate a great flurry of interest in neutron stars. A sizeable number of theorists independently speculated that the X-ray telescope was observing a young, warm neutron star, and they fervishly began to calculate the cooling of neutron stars. The identification of the first "quasi-stellar object" (QSO, or quasar) by Schmidt at Mt. Palomar in 1963 triggered further interest in neutron stars. This interest stemmed from the possibility that the large redshifts of spectral lines observed for quasars might be attributed to the gravitational redshift at the surface of a compact object. Arguments showing that the largest quasar redshift already exceeded the maximum gravitational redshift from a stable neutron star soon dispelled any connection between quasars and (isolated) neutron stars.

In any case, with the discovery of X-ray sources and quasars, dozens of theoreticians focused their attention on the equilibrium properties of compact stars and on star collapse. But in spite of this mounting theoretical effort, most physicists and astronomers did not take the possibility of neutron stars (let along black holes!) very seriously. Probably the vast extrapolation from familiarly known physics was the most important reason for their attitude!

All his changed when pulsars were discovered in late 1967. Gold (1968) proposed that they were rotating neutron stars, and this is generally accepted today (see Chapter 10).

[4]Ciaconni, Gursky, Paolini and Rossi (1962).

Since 1968, there has been much theoretical work on properties of neutron stars. This was further stimulated by the discovery of pulsating compact X-ray sources ("X-ray pulsars") by the UHURU satellite in 1971. These are believed to be neutron stars. Although the idea of accreting binary systems for X-ray sources had been proposed earlier, the first conclusive evidence for periodicity was found in the sources Cen X-3 and Her X-1.

The near simultaneous discoveries, of the Crab and Vela pulsars in the late fall of 1968, both of which are situated in supernova remnants, provided evidence for the formation of neutron stars in supernova explosions. The crab nebula, for example, is the remnant of the supernova explosion observed by Chinese astronomers in 1054 A.D.

Optical and X-ray observations of binary X-ray sources allow one to determine the neutron star masses in some of these systems. The discovery of the first binary pulsar by Hulse and Taylor (1975) also provides an opportunity to measure the mass of a neutron star and, as we shall see later, to test for the existence of gravitational radiation.

At the time of this writing, about 350 pulsars are known, three of which are in binary systems. Over 300 compact X-ray sources are known, about 19 of which show periodicity, and so are probably in binary systems.

The analysis we performed for white dwarfs is easily adapted to an idealized "neutron star" consisting of a degenerate, non-relativistic gas of self-gravitating neutrons. This is a mere "cartoon" of a neutron star, ignoring as it does the rich nuclear physics and general relativistic effects which occur in neutron stars. We will nevertheless capture enough of the correct physics to see why neutron stars are stable and what are their size and mass scales. Replacing the electron gas by a neutron one, in §2.c. we obtain

$$\begin{aligned} \rho_0^n &= \frac{m_n}{3\pi^2}\left(\frac{m_n c}{\hbar}\right)^3 = 6.1\times 10^{15}\,\mathrm{gm/cm^3} \\ R_0^n &= \left[\frac{3\pi^2\hbar^3}{4Gm_n^4 c}\right]^{1/2} = 4.20\,\mathrm{km} \\ M_0^n &= \left(\frac{3\pi^2\hbar^3 c^3}{G^3}\right)^{1/2}\frac{1}{m_n^2} = 5.66\times 10^{33}\,\mathrm{gm} \end{aligned} \tag{2.53}$$

and (for $\gamma = 5/3$)

$$\begin{aligned} \rho_c^n &= \rho_0\kappa^{3/2} \\ M^n &= \frac{\kappa^{3/4}}{\sqrt{2}}M_0\left[-\xi_1^2\theta'(\xi_1)\right] \\ R^n &= \frac{1}{\sqrt{2}\,\kappa^{1/4}}R_0\xi_1 \quad . \end{aligned} \tag{2.54}$$

Eliminating κ we obtain

$$MR^3 = \frac{91.9\hbar^6}{G^3 m_n^5} = 7.2 \times 10^{51}\,\mathrm{gm\ cm^3} \tag{2.55}$$

or

$$\frac{M}{M_\odot}\left(\frac{R}{15\,\mathrm{km}}\right)^3 \approx 1 \quad . \tag{2.56}$$

These results suggest that neutron stars have masses comparable to white dwarfs but radii a factor of $m_e/m_n \approx 2{,}000$ smaller and density of order $(m_n/m_e)^3 \sim 6 \times 10^9$ greater than white dwarfs.

Next we turn to the issue of whether or not neutron stars are stable. Once again, answering this question properly requires time and analysis beyond the scope of this course. We can get an indication of the issue, however, by following a simple analysis due to Landau. Consider the non-relativistic "neutron star" we have just constructed. At the outer limits of this object the pressure must vanish. Thus it is clear that the outer region of a neutron star consist of a shell of electrons and nuclei at low pressure and density. The thickness of this shell may be large, but its density is much lower than the neutron core, so the core accounts for nearly all of the neutron star's mass. Let us compare the energy of neutrons at the core radius (R^n) with the energy of the nuclei and electrons into which they would disassociate if they were transported to the outer limit of the star (R'). At the core radius

$$E \approx m_n c^2 - \frac{GMm_n}{R^n} \tag{2.57}$$

is the energy per neutron, and at the surface of the star

$$E' \approx \left(\frac{M_A + Zm_e}{A}\right)c^2 - \frac{GMm_n}{R'} \quad . \tag{2.58}$$

If the star is to be stable we must require $E < E'$. In comparing (2.57) and (2.58) we may ignore $1/R'$ compared to $1/R$, so we obtain

$$\frac{GMm_n}{R} > \Delta \tag{2.59}$$

where $\Delta = [m_n - (1/A)\,(M_A + Zm_e)]\,c^2$. Comparing (2.59) with the mass-radius relation (2.56) we find numerically $M > 0.17\,M_\odot$ for oxygen and $M > 0.18\,M_\odot$ for iron. This calculation gives a lower limit on M for a neutron star to be stable against a differential change — transporting a small amount of material from a region of high density to low density. However it would not detect an instability requiring a global change in the configuration of the star. For this reason the neutron star is only metastable in this mass region. A more general analysis shows neutron stars

to be truly stable for $M \gtrsim 0.3\, M_\odot$. The mass-radius relation for a non-relativistic neutron star is graphed in Fig. 2.3 along with the results of more realistic studies. The "realistic" curve (marked B) is taken from a schematic diagram in Shapiro and Tuekolsky [Fig. 6.3]. The failure of the neutron model results from the omission of the electrons which greatly increase the radius at low mass.

To determine an upper limit on the masses of neutron stars we need to incorporate the effects of general relativity. The most obvious modification of the hydrostatic analysis of §2.c. is that energy also gravitates — energy here includes both the kinetic energy of the neutron Fermi gas and the gravitational potential energy itself. The equations of hydrostatic equilibrium for a spherical self-gravitating body in the context of general relativity were worked out by Oppenheimer and Volkoff. They should be compared with (2.19) and (2.20)

$$\frac{dm}{dr} = 4\pi r^2 \rho \tag{2.60}$$

$$\frac{dP}{dr} = -G\frac{m\rho}{r^2}\left(1+\frac{P}{\rho c^2}\right)\left(1+\frac{4\pi P r^3}{mc^2}\right)\left(1-\frac{2mG}{rc^2}\right)^{-1} \quad . \tag{2.61}$$

General relativistic effects become important when $P \sim \rho c^2$ which is certainly the case at densities of order ρ_0^n (Eq. (2.53)). The potential singularity suggested by the last term in (2.61) is associated with gravitational collapse;

$$r_s = \frac{2mG}{c^2}$$

$r_s = 2.95$ km for $M = M_\odot$, so this singularity is significant for neutron star stability.

Consider, first a pure, ideal Fermi gas of neutrons.[5] Direct integration of the Oppenheimer–Volkoff equations gives $M_{\max} = 0.7\, M_\odot$, $R = 9.6$ km and $\rho_c = 5 \times 10^{15}$ gm/cm^3. Had we ignored general relativity the result would have been quite different: the Chandrasekhar limit for a neutron gas ($\gamma = 4/3$ polytrope) is $5.73\, M_\odot$. General relativity lowers this for two principle reasons :1) The maximum occurs at a value of ρ_c well below the ultra-relativistic limit as assumed by the $\gamma = 4/3$ polytrope; 2) $5.73\, M_\odot$ includes the neutron rest mass but not the (negative) gravitational binding energy. The mass-central density relation for white dwarfs and ideal neutron stars is displayed in Fig. 2.4 (borrowed from Shapiro and Teukolsky).

More realistic calculations differ in their treatment of neutron and nuclear interactions at high densities and give upper limits to neutron star masses which range from $\approx 1.5\, M_\odot$ up to about $3\, M_\odot$. At this point we must abandon this brief sketch of neutron stars and return to the main body of the course. For a complete and readable survey of neutron stars see Shapiro and Teukolsky's excellent book.

[5] Taken from Shapiro and Teukolsky.

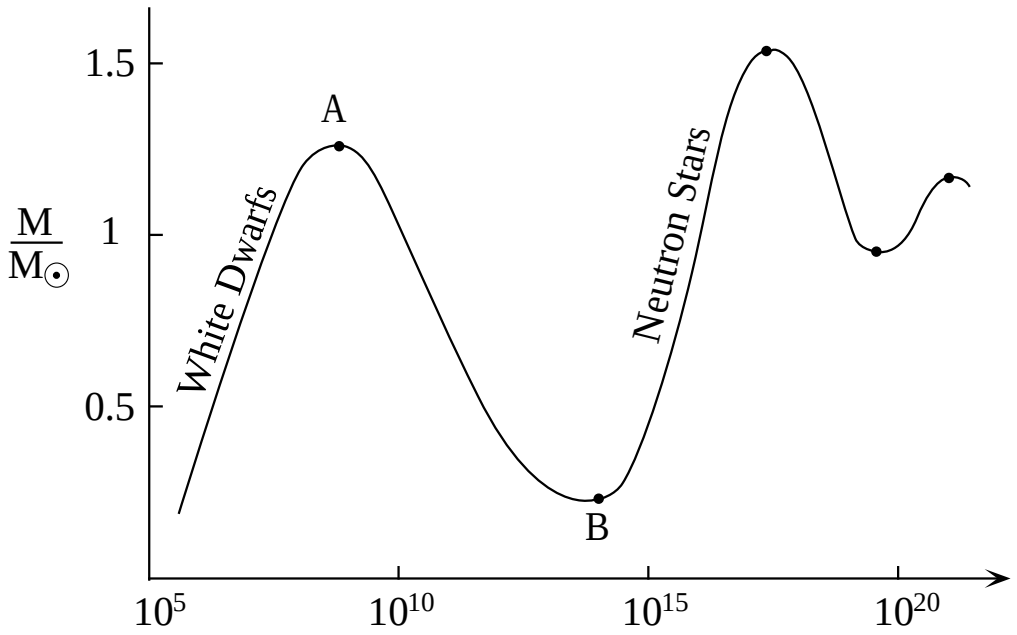


Figure 2.4: The mass of a compact star as a function of its central density ρ_c in [gm/cm^3].

2.6 QUARKS AND QUARK MATTER

We have already alluded to the possible importance of quark matter in the dynamics of neutron star cores, where the density is so great that the description of matter in terms of nuclei and nucleons breaks down. In this section, we wish to describe another, highly speculative and bizarre possibility: that a degenerate Fermi gas of quarks may be the time ground state of matter at zero external pressure. Like all speculative ideas, this one is most likely wrong. But in this case as in many others, the fruit of speculation is a significant insight into the workings of Nature in the ordinary world.

To begin we have to say something about quarks and their interactions. We shall be concerned with the three lightest species, or "flavors" of quarks, known as the u-, d- and s-quarks, for "up," "down" and "strange." They are spin-1/2 fermions with rest masses (mc^2) of about 5 MeV, 10 MeV and 150 MeV, respectively. Quarks are never observed in isolation so their masses are a subject of considerable uncertainty. These numbers may be off by as much as a factor of two in either direction, although the ratios of quark masses are much better known. The three quarks have different (and unusual) electromagnetic charges: $Q_u = 2/3|e|$, $Q_d = -1/3|e|$ and $Q_s = -1/3|e|$. In addition, quarks carry another quantum number, "color," which is essential in understanding the strong interactions which bind quarks together into protons, neutrons and other strongly interacting particles known collectively as "hadrons." Each

quark comes in three colors and two spins for a total degeneracy factor of six. Some important properties of quarks were summarized in Table 1.1.

The interactions which confine quarks to the interior of hadrons depend on the color quantum number. In many ways they resemble electrodynamics and are known as *chromo*dynamic forces. For reasons which are not yet totally understood the forces between quarks grow very strong if one attempts to isolate a single quark. This phenomenon, known as "quark confinement" is the focus of much current research. Only systems of quarks which are neutral with respect to color appear to be allowed as real particles. Color neutrality differs from electromagnetic charge neutrality — understanding the difference requires an excursion into group theory which we will take later in the course. For example, a bound state of a quark and antiquark may be color neutral, much in the same way that an electron-positron bound state is charge neutral. In addition, a bound state of three quarks may be color neutral — in contrast to a state of three electrons which has electric charge -3. All color neutral states consist of multiples of three quarks and/or quark-antiquark pairs. Hadrons with no excess of quarks over antiquarks are known as mesons. Hadrons with an excess of quarks (or antiquarks) are baryons (or antibaryons). Protons, neutrons and nuclei are examples of baryons. The net excess of quarks over antiquarks is conserved by all known interactions and (divided by three) is known as baryon number, A,

$$A = \frac{1}{3}\left(N_Q - N_{\bar{Q}}\right) \quad . \tag{2.62}$$

Once quarks are confined into color neutral states the residual interactions among the quarks appear to be rather weak. This is one of the most puzzling aspects of quark dynamics and is only partially explained by a famous property of chromodynamics known as "asymptotic freedom." This catchy slogan refers to the fact that interquark interactions weaken at very short distances. Roughly speaking it's as though the charge which enters Coulomb's law decreases logarithmically as $r \to 0$

$$V_{12}(r) \sim \frac{\alpha_0}{r|\ln r|} \quad . \tag{2.63}$$

In practice, quark interactions within hadrons are even weaker than this behavior would suggest. A very simple and successful model of hadrons ignores them entirely. This model, developed here at MIT, and known as the "Bag Model," supposes that the vacuum exerts a universal pressure on any color neutral collection of quarks, preventing them from escaping to large distances. The dynamics of hadrons is governed by the balance between this vacuum pressure, B, and the Fermi degeneracy pressure of the quarks. Estimates of B center around 1.0×10^{29} atmospheres, or in units more suitable to quark physics

$$B \approx 62.5\,\mathrm{MeV/fm}^3 \quad . \tag{2.64}$$

This huge pressure is responsible for the tiny size of hadrons. The fact that the units of pressure may be expressed either as force/area of energy/volume helps us understand the physical origin of the bag model. Apparently there is a cost (in energy) in opening up a region of space in which quarks may live: quarks are sources of strong color fields. The vacuum, in its normal state cannot support color fields — they are quenched much as ordinary magnetic fields are quenched inside a superconductor. Very near the quarks, the color fields are so intense that they alter the conformation of the vacuum and create a region in which they can live more or less normally. Toward the edges of this region the color field weakens and eventually the normal vacuum reasserts itself quenching the fields to zero. Magnetic fields inside superconductors behave the same way: strong enough fields drive out the condensate of Cooper pairs and create a "normal region." The boundary of this normal region comes when the fields are not strong enough and the Cooper pair density and associated supercurrents arise and quench the magnetic field to zero. The bag constant, B, is the cost — energy per unit volume — of changing the conformation of the vacuum to allow for the presence of quarks and their associated color fields. These arguments are illustrated graphically in Fig. 2.5.

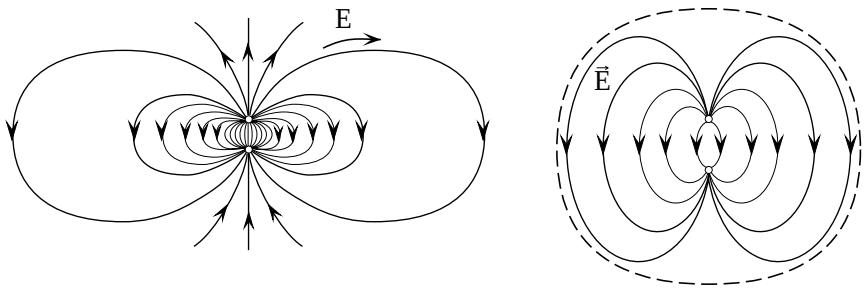


Figure 2.5: The lines of magnetic field surrounding a point magnetic dipole in vacuum (left) and inside a bulk superconductor (right). We believe that the same pictures apply to quarks bound in hadrons with magnetic fields $\leftrightarrow$ color electric fields.

Quark matter can be described in the Bag Model as a degenerate Fermi gas of quarks confined to finite volume by the universal pressure B. B also contributes to the energy of the system:

$$E = T + BV \quad . \tag{2.65}$$

where T is the kinetic energy of the quark Fermi gas. With the problem reduced to this simple model let's explore a few applications.

Protons, neutrons and nuclei are composed entirely of up and down quarks. A look at Table 2.2 shows that the rest masses of the u- and d-quarks are very small compared to the scale of nucleon and nuclear masses. Apparently most of the mass

of the nucleons comes from the kinetic energy of relativistic confined quarks. In fact, to a good approximation we can ignore the rest masses of the quarks. Let us now consider a collection of u- and d-quarks large enough so that they can be described as a Fermi gas ignoring surface effects (see §3 for a discussion of surface effects). Each quark species will be characterized by a Fermi momentum, p_F^j $(j = u, d)$. Since the average of the u- and d-quarks charge is positive, electrons will be present to ensure overall charge neutrality. Using the ultra-relativistic limit from §1.c.,

$$\begin{aligned} n^u &= \frac{\left(p_F^u\right)^3}{\pi^2\hbar^3} \quad , \qquad u^u = \frac{3c\left(p_F^u\right)^4}{4\pi^2\hbar^3} \\ n^d &= \frac{\left(p_F^d\right)^3}{\pi^2\hbar^2} \quad , \qquad u^d = \frac{3c\left(p_F^d\right)^4}{4\pi^2\hbar^3} \\ n^e &= \frac{\left(p_F^e\right)^3}{3\pi^2\hbar^3} \quad , \qquad u^e = \frac{c\left(p_F^e\right)^4}{4\pi^2\hbar^3} \quad . \end{aligned}$$

Charge neutrality requires

$$\frac{2}{3}n^u - \frac{1}{3}n^d - n^e = 0 \tag{2.66}$$

and equilibrium with respect to the weak interaction

$$d \rightarrow u + e^- + \bar{\nu}_e \tag{2.67}$$

requires

$$p_F^d = p_F^u + p_F^e \quad . \tag{2.68}$$

Finally, the total energy density is given by

$$u = u^u + u^d + u^e + B \quad . \tag{2.69}$$

The ground state configuration of the system is determined by minimizing the energy per quark subject to the constraints (2.66) and (2.68). This is equivalent to minimizing the total energy subject to the constraint that the total number of quarks is conserved. We parameterize p_F^u, p_F^d and p_F^e by

$$\begin{aligned} p_F^d &\equiv p \\ p_F^u &= p\cos\theta \\ p_F^e &= p(1-\cos\theta) \end{aligned} \tag{2.70}$$

then charge neutrality (2.66) constrains θ

$$\cos^3\theta - \cos^2\theta + \cos\theta - \frac{2}{3} = 0 \tag{2.71}$$

so

$$\theta = \cos^{-1}(0.7959) \approx \cos^{-1}\frac{1}{2^{1/3}} \quad . \tag{2.72}$$

For this value of θ, n^e is negligibly small,

$$\frac{n_e}{n_d} \approx \frac{1}{3}(1-\cos\theta)^3 \approx 2.8\times 10^{-3} \tag{2.73}$$

so we can ignore the electrons in the rest of the calculation. The energy per quark is given by

$$\frac{u}{n} = \frac{(3/4)pc(1+\cos^4\theta) + (\pi^2\hbar^3 B/p^3)}{(1+\cos^3\theta)} \tag{2.74}$$

and p is determined so that u/n is a minimum,

$$p = \left[\frac{4\pi^2\hbar^3 B}{c(1+\cos^4\theta)}\right]^{1/4} \tag{2.75}$$

and at the minimum

$$\begin{aligned} u &= 4B \\ n &= \left[\frac{4\pi^2 B}{\hbar c(1+\cos^4\theta)}\right]^{3/4}\frac{(1+\cos^3\theta)}{\pi^2} \\ \frac{u}{(1/3)n} &= \frac{U}{A} = 3\left(4\pi^2 B\hbar^3 c^3\right)^{1/4}\frac{(1+\cos^4\theta)^{3/4}}{1+\cos^3\theta} \quad . \end{aligned} \tag{2.76}$$

Some comments are in order: $u = 4B$ is a general "virial theorem" for the ultra-relativistic system subject to an external pressure. $u/(1/3)n^v = U/A$ is the energy per baryon. If we take $B = 62.5\,\mathrm{MeV/fm^3}$ then $U/A = 965$ MeV, which is only slightly greater than the mass per baryon of a nucleus ($Mc^2/A \approx 930\,\mathrm{MeV}$). So we conclude that there is another configuration of the quarks which go together to make up a nucleus. Unlike a normal nucleus, which consists of protons and neutrons in non-relativistic motion, this "quark matter' state consists of an ultra-relativistic degenerate Fermi gas of u- and d-quarks in one big bag. The quark matter state has never been seen because it is slightly ($\sim 4\%$) more energetic than the normal nuclear configuration. Earlier on, I remarked that B is not a very well-known parameter of chromodynamics. With a somewhat smaller value of B, quark matter consisting of u- and d-quarks would have been bound. This possibility can be excluded experimentally: It if were bound, ordinary nuclei like $^{56}_{26}$Fe would decay quickly into lumps of quark matter. We have lots of evidence that ordinary nuclei are composed of protons and neutrons, not a Fermi gas of quarks. So this possibility is ruled out by experiment.

There remains the very interesting possibility that quark matter becomes energetically favored compared to nuclear matter under extreme conditions of temperature or pressure. Much effort has been put into exploring these ideas both experimentally and theoretically. On the theoretical side, the temperature dependence of quantum chromodynamics has been studied using models and large scale computer calculations. There's considerable evidence at a temperature of order 200 MeV ($\equiv kT$) chromodynamics undergoes a phase transition from a phase in which quarks are confined to hadrons to one in which quarks and gluons form a sort of plasma. Less is known about the density dependence of chromodynamics. There may be a phase transition from the "low"density nuclear phase to a high density quark matter phase at some critical density ρ_0, or perhaps the change is gradual much like the transition in which the electrons and ions disassociate from one another in a gas at high temperature or density. Phase diagrams for chromodynamics showing both possibilities are presented in Fig. 2.6. On the experimental side, the aim is to briefly raise nuclei to high temperature and density by colliding two of them together. Some of the kinetic energy of the colliding nuclei goes into heating them up; collisions raise the system to higher density. A hypothetical trajectory followed by the colliding nuclei is shown in Fig. 2.6. Whether the collisions will produce enough heating and/or overpressure to produce quark matter is as yet not known.

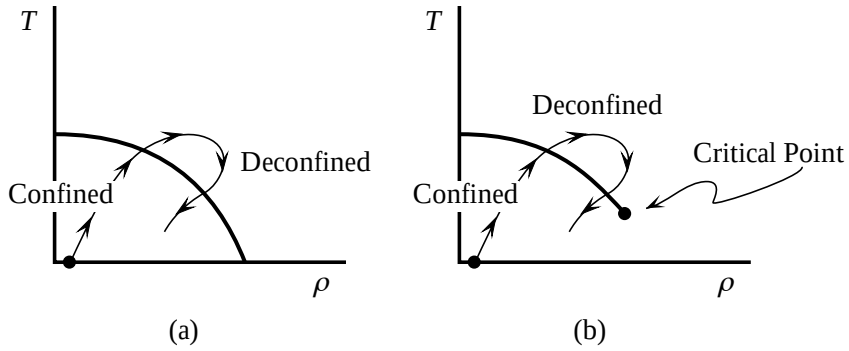


Figure 2.6: Possible phase diagrams for QCD. At low temperature and pressure quarks are confined in hadrons. At high temperature and pressure quarks are deconfined. The abruptness of the transition is not known. Two possibilities are shown. The trajectory shown might describe what happens to nuclei in heavy ion collisions.

There is another, even more exciting and more speculative possible role for quark matter in Nature. To see its origin, return to the results obtained in (2.76) for u- and d-quark matter. The Fermi energy (chemical potential) for a u- or d-quark is just its Fermi momentum, $\mu = p_F c$, because we are ignoring quark rest masses.

$$\mu_u = \mu_d = \left[\frac{4\pi^2\hbar^3 c^3 B}{1+\cos^4\theta}\right]^{1/4} \quad . \tag{2.77}$$

For $B = 62.5\,\mathrm{MeV/fm}^3$ and $\cos\theta = 0.7959$ we find $\mu_u = \mu_d = 340$ MeV. This Fermi energy is well-above the mass of the strange quark $m_s c^2 \approx 150$ MeV, so quark matter could lower its energy if it was possible to convert up and/or down quarks into strange quarks. This is an example of a common phenomenon known as "symmetry energy" in Fermi systems involving several species of fermion: At chemical potentials large compared to the difference in rest mass of the fermion species, interactions which transform are species into another will tend to equalize the numbers of different species. For the case of u-, d- and s-quarks, the weak interactions provide the mechanism via the reactions

$$\begin{aligned} e^- + u &\longleftrightarrow \nu_e + s \\ d + u &\longleftrightarrow u + s \quad . \end{aligned} \tag{2.78}$$

These "strangeness changing" weak interactions have a natural time scale of order $\tau \equiv 10^{-10}$ sec. So a lump of non-strange quark matter held together for times of order τ will spontaneously develop strangeness and lower its total energy (radiating away neutrinos and photons). Perhaps it will radiate away so much energy that it actually becomes stable!

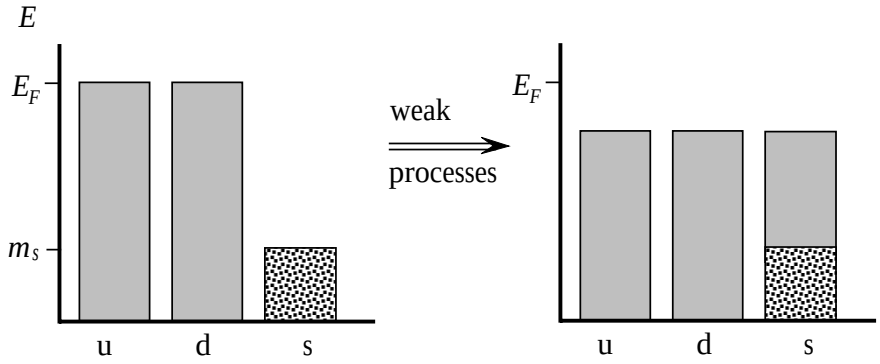


Figure 2.7: Quark fermi seas. With strangeness changing interactions "turned off", u and d quarks equilibrate (left). When strangeness changing interactions are "turned on", the system can lower its energy if $E_F > m_s$ (right).

The startling proposition, then, is that *strange quark matter* or "strange matter" as it is known may actually be the true ground state of collections of quarks at zero temperature and external pressure. We define "strange matter" to be quark matter in equilibrium with the weak interactions which change flavor. The process of equilibration is illustrated in Fig. 2.7. The reader might protest that the existence of nuclei is *prima facia* evidence against the absolute stability of any form of quark matter, since the time scale for quarks to rearrange themselves within a nucleus

is $\sim 10^{-24}$ second. [This is the time it takes light — or any signal — to cross a nucleon.] This argument was generally accepted until the mid-1980's when Witten pointed out an elementary flaw: A weak interaction is suppressed by a factor of order 10^{-14} compared to the natural time scale of nuclear motions. One weak interaction *à la* (2.78) converts one u- or d-quark into an s-quark. A single weak interaction inside a nucleus converts a nucleon into a strange nucleon known as a hyperon and known to be much heavier than a proton or neutron. The symmetry energy discussed above cannot be realized until a significant number of nucleons convert to hyperons at which time it is then energetically favorable for the many nucleons in a nucleus to convert collectively into strange matter. Thus, for example, it might be that ${}^{56}_{26}\mathrm{Fe}$ is unstable against decay to a quark Fermi gas containing on the order of 50 s-quarks, however one, two or even several weak interactions will not adjust the quark flavor composition to the point where the reorganization into quark matter will occur. Each weak interaction brings with it a suppression factor of roughly 10^{14}. The likelihood of many strangeness changing weak interactions occurring over the same brief time period within a single nucleus is so small that the lifetime of a nucleus like ${}^{56}_{26}\mathrm{Fe}$ against decay into strange quark matter far exceeds the age of the universe.

It could be that strange matter, though the true ground state of matter, has never been made in the history of the universe. On the other hand it might be that deep in the cores of neutron stars, neutrons are raised to such high densities that they coalesce into non-strange quark matter, which then converts at its leisure (so to speak) into strange quark matter. Whether strange matter is made in neutron stars and how it might subsequently get distributed through the universe are open issues in the new and active field of particle astrophysics.

The properties of strange matter: a degenerate Fermi gas of u-, d- and s-quarks have been worked out in a series of papers over the last decade. Time doesn't permit us to treat this subject here, but a few of the more striking properties of strange matter are explored in the problems.

Chapter 3

GREEN'S FUNCTIONS AND THE DENSITY OF STATES

The method Green's functions is a powerful tool for the analysis of differential equations. In general application the method allows one to

a. Incorporate information about boundary conditions into the problem in a natural way.

b. Solve the homogeneous ("isolated") and inhomogeneous ("externally driven") problems with the same method, often at the same time.

c. Incorporate constraints like causality naturally into the propagation of signals.

d. Study, at one time and in a single function, the generic properties of all solutions to a problem.

Here we are primarily interested in *d*.

The price of this power is a somewhat more complicated formalism. Here we introduce only what we need to study the density of states. Later (*e.g.* in the unit on scattering theory) we develop more aspects of Green's function theory.

3.1 GREEN'S FUNCTIONS

Consider the time-independent Schrödinger equation with Hamiltonian H:

$$(E - H)\psi_E(\vec{x}) = 0 \qquad (\text{in } D) \quad . \tag{3.1}$$

(3.1) is valid in some domain D bounded by a surface ∂D on which we assume

$$\psi(\vec{x}) = 0 \qquad (\text{on } \partial D) \quad . \tag{3.2}$$

The region D may have any shape. It may be compact or infinite in extent. H may contain a potential energy $V(\vec{x})$.

Now consider the related problem of the same differential operator with the addition of a point source at the point $\vec{y}$:

$$(E - H)G(\vec{x}, \vec{y}, E) = i\hbar\delta^3(\vec{x} - \vec{y}) \qquad (\text{in } D) \tag{3.3}$$

and subject to

$$G(\vec{x}, \vec{y}, E) = 0 \qquad (\vec{x} \text{ on } \partial D) \quad . \tag{3.4}$$

G is called the Green's function for the differential operator $E - H$. [In other cases the form of (3.3) can be motivated as the actual response of the system to a sharp impulse at some time or place. That interpretation doesn't work for $E - H$ because we wouldn't know how to interpret $(E - H)\psi(\vec{x}) = f(\vec{x})$ physically. So the Green's function method is more formal in this case.]

Using the completeness and orthonormality of energy eigenstates it is easy to construct $G(\vec{x}, \vec{y}, E)$. The solutions to (3.1) and (3.2) obey

$$\int d^3x\, \psi_n^*(\vec{x})\psi_{n'}(\vec{x}) = \delta_{nn'} \qquad (\text{orthonormality}) \tag{3.5}$$

$$\sum_n \psi_n(\vec{x})\psi_n^*(\vec{y}) = \delta^3(\vec{x} - \vec{y}) \qquad (\text{completeness}) \quad . \tag{3.6}$$

where n is a short-hand for all labels necessary to specify a unique state. These forms hold if D is compact. If not, it is necessary to use continuum normalization. Using (3.5) and (3.6) it's easy to see

$$G(\vec{x}, \vec{y}, E) = i\hbar \sum_n \frac{\psi_n(\vec{x})\psi_n^*(\vec{y})}{E - E_n} \tag{3.7}$$

satisfies both the differential equation and the boundary condition required of G. (3.7) is singular whenever $E = E_n$ — one of the energy eigenvalues. Analysis of there singularities is a key part of Green's function theory but it is not necessary for study of the density of states, so we postpone it until a later section.

Some examples are in order:

Example 1: A free particle in unbounded space

For this special, simple case we call the Green's function G_0. H reduces to

$$H_0 = \frac{\vec{p}^2}{2m} = -\frac{\hbar^2}{2m}\vec{\nabla}^2 \quad ,$$

so

$$\left(E + \frac{\hbar^2}{2m}\vec{\nabla}^2\right) G_0(\vec{x},\vec{y},E) = i\hbar\delta^3(\vec{x}-\vec{y})$$

or

$$\left(\vec{\nabla}^2 + k^2\right) G_0(\vec{x},\vec{y},E) = \frac{2mi}{\hbar}\delta^3(\vec{x}-\vec{y}) \tag{3.8}$$

where $k^2 \equiv 2ME/\hbar^2$. Now, we know a solution to a related problem from electrostatics:

$$\vec{\nabla}^2\left(\frac{1}{r}\right) = -4\pi\delta^3(\vec{x}) \quad , \tag{3.9}$$

the potential due to a point charge. It is easy to see that a slight modification yields a solution for G_0:

$$G_0(\vec{x},\vec{y},E) = -\frac{2mi}{4\pi\hbar}\,\frac{e^{\pm ikr}}{r} \tag{3.10}$$

where $r \equiv |\vec{x}-\vec{y}|$.

The same result can, of course, be obtained from the general solution, (3.7), with the replacements

$$\begin{aligned} \psi_n(\vec{x}) &\rightarrow \psi_{\vec{p}}(\vec{x}) = \frac{1}{(2\pi\hbar)^{3/2}} e^{i\vec{p}\cdot\vec{x}/\hbar} \\ \delta_{nn'} &\rightarrow \delta^3\left(\vec{p}-\vec{p}'\right) \\ \sum_n &\rightarrow \int d^3p \end{aligned}$$

so

$$G \rightarrow G_0(\vec{x},\vec{y},E) = i\hbar\int \frac{d^3p}{(2\pi\hbar)^3}\,\frac{e^{i\vec{p}\cdot(\vec{x}-\vec{y})/\hbar}}{E-p^2/2m} \quad . \tag{3.11}$$

To evaluate this integral we must learn how to handle the singularity at $p^2 = 2mE$. It turns out that there are two choices — corresponding to closing a contour in the upper or lower half complex p-plane — yielding after some elementary analysis, Eq. (3.10) with either sign.

Example 2: A free particle in a bounded domain

Consider next the case of a particle in an arbitrary domain, D, with no potential energy term in H. We can then write

$$G = G_0 + \tilde{G} \tag{3.12}$$

where G_0, given in (3.7), is the free Green's function and upon substitution in (3.3), $\tilde{G}$ is found to obey

$$\begin{aligned} (E-H)\tilde{G}(\vec{x},\vec{y},E) &= 0 \qquad (\vec{x} \text{ in } D) \\ \tilde{G} &= -G_0 \qquad (\vec{x} \text{ on } \partial D) \quad . \end{aligned} \tag{3.13}$$

In some cases $\tilde{G}$ is easy to find. Consider, for example, the case of a half space, $x_3 < 0$, bounded by the infinite plane at $x_3 = 0$. Then an argument borrowed from the method of images in electrostatics tells us,

$$\tilde{G}(\vec{x},\vec{y},E) = -G_0\left(\vec{x},\tilde{\vec{y}},E\right) \tag{3.14}$$

where $\tilde{\vec{y}}$ is defined as follows: If $\vec{y} = (y_1, y_2, y_3)$ then $\tilde{\vec{y}} = (y_1, y_2, -y_3)$. The geometry is shown in Fig. 3.1. Clearly $\tilde{G}$ obeys $(E-H)\tilde{G} = 0$ for both $\vec{x}$ and $\vec{y}$ in D (since $\vec{x}$ cannot coincide with $\tilde{\vec{y}}$). To show $\tilde{G} = -G_0$ for $\vec{x}$ on ∂D note that $|\vec{x}-\vec{y}| = |\vec{x}-\tilde{\vec{y}}|$ when $x_3 = 0$. Since $G_0(\vec{x},\tilde{\vec{y}},E)$ depends only on $|\vec{x}-\tilde{\vec{y}}|$ we have established

$$\tilde{G}(\vec{x},\vec{y},E) = -G_0\left(\vec{x},\tilde{\vec{y}},E\right) = -G_0(\vec{x},\vec{y},E) \quad ,$$

for $\vec{x}$ on ∂D.

Example 3: Particle in a potential $V(x)$, no boundaries

$$H = H_0 + V$$

So G obeys

$$(E - H_0 - V)\, G\,(\vec{x},\vec{y},E) = i\hbar\delta^3(\vec{x}-\vec{y}) \quad . \tag{3.15}$$

This is apparently satisfied by the infinite series

$$\begin{aligned} G(\vec{x},\vec{y},E) &= G_0(\vec{x},\vec{y},E) + \frac{1}{i\hbar}\int d^3z_1\, G_0(\vec{x},\vec{z}_1,E)V(\vec{z}_1)G_0(\vec{z}_1,\vec{y},E) \\ &+ \left(\frac{1}{i\hbar}\right)^2 \int d^3z_1\, d^3z_2\, G_0(\vec{x},\vec{z}_1,E)V(\vec{z}_1) \\ &\qquad \times G_0(\vec{z}_1,\vec{z}_2,E)V(\vec{z}_2)G_0(\vec{z}_2,\vec{y},E) \\ &+ \ \ldots \end{aligned} \tag{3.16}$$

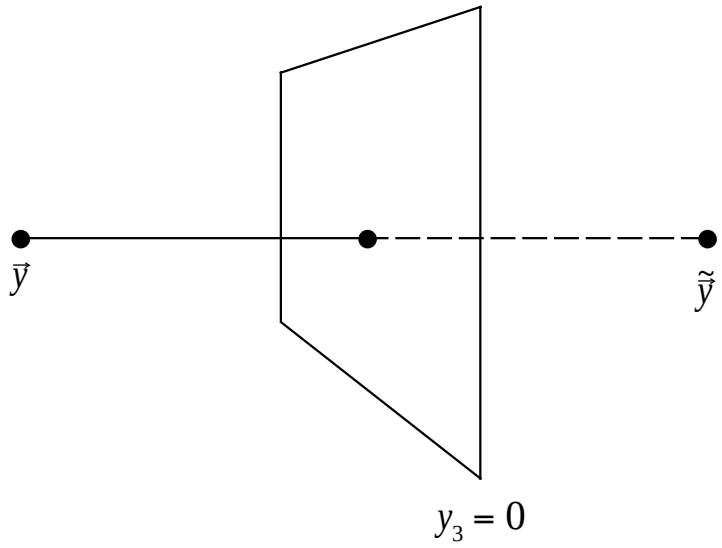


Figure 3.1: The point $\vec{y}$ and its reflected point $\tilde{\vec{y}}$.

as can be seen by direct substitution.

Here is a derivation.

$$(E - H_0 - V(\vec{z}))\, G(\vec{z}, \vec{y}, E) = i\hbar\delta^3(\vec{z} - \vec{y})$$

Multiply by $G_0(\vec{x}, \vec{z}, E)$ and integrate d^3z,

$$\int d^3z\, G_0(\vec{x}, \vec{z}, E)\, (E - H_0 - V(\vec{z}))\, G\,(\vec{z}, \vec{y}, E) = i\hbar G_0(\vec{x}, \vec{y})$$

Integrate ∇_z^2 (in H_0) by parts and use

$$\left(E + \frac{\hbar^2}{2m}\nabla_z^2\right) G_0(\vec{x}, \vec{z}, E) = i\hbar\delta^3(\vec{x} - \vec{z})$$

[Note the Laplacian acts on $\vec{z}$, not $\vec{x}$ as in the definition of G_0. This result follows from the explicit form of G_0 (3.10) which is clearly symmetric in $\vec{x} \leftrightarrow \vec{z}$.] This leaves an integral equation

$$i\hbar G(\vec{x}, \vec{y}, E) = i\hbar G_0(\vec{x}, \vec{y}, E) + \int d^3z G_0(\vec{x}, \vec{z}, E) V(\vec{z}) G(\vec{z}, \vec{y}, E)$$

which yields (3.16) by iteration. This series solution is only useful when it converges, which is determined by how small and how smoothly varying is $V(\vec{z})$. Generally speaking if V is smooth and small the expansion is useful (more on this later).

3.2 RELATION OF GREEN'S FUNCTION TO THE DENSITY OF STATES

Consider $N(E)$, the number of states with energy less than E,

$$N(E) = \sum_n \Theta(E - E_n)$$

and its derivative

$$\rho(E) = \sum_n \delta(E - E_n) \quad .$$

Let us use the Lorentzian representation for the δ-function introduced in §1.1

$$\delta(x) = \frac{1}{\pi} \lim_{\epsilon \to 0} \frac{\epsilon}{x^2 + \epsilon^2}$$

so

$$\rho(E) = \frac{1}{\pi} \sum_n \lim_{\epsilon \to 0} \frac{\epsilon}{(E - E_n)^2 + \epsilon^2} \quad .$$

As a way of smoothing out this sum over δ-functions, we propose to keep the width of these Lorentzian's finite while we study the sum. In effect we are replacing the sharp spikes by smoothed peaks of unit area.

$$\begin{aligned} \rho(E) \to \rho_\epsilon(E) &= \frac{1}{\pi} \sum_n \frac{\epsilon}{(E - E_n)^2 + \epsilon^2} \\ \rho_\epsilon(E) &= \frac{1}{\pi} \Im \sum_n \frac{1}{E - E_n - i\epsilon} \quad . \end{aligned} \tag{3.17}$$

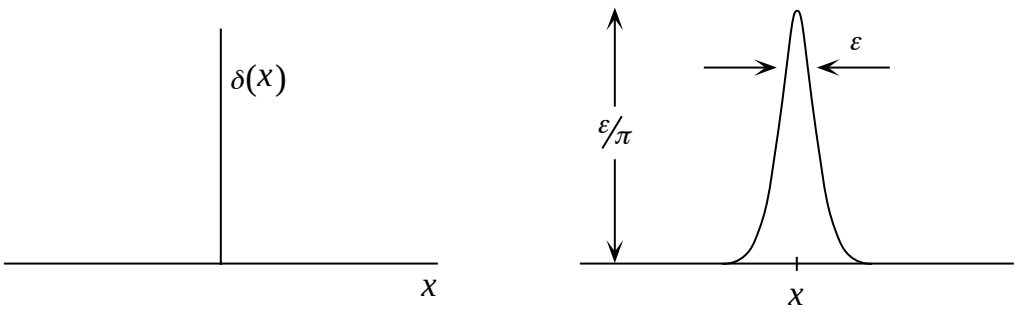


Figure 3.2: A delta function and the smoothed out, "Lorentzian" approximation, $\frac{\epsilon}{\pi} \frac{1}{\epsilon^2 + x^2}$.

We might have taken (3.17) as our starting definition of the smoothed density of states. Comparing (3.17) with the expression, (3.7) for the Green's function, we see

$$\rho_\epsilon(E) = -\frac{1}{\pi\hbar} \Re \int d^3x \, G(\vec{x}, \vec{x}, E - i\epsilon) \tag{3.18}$$

This is the fundamental connection we were seeking. Let us now exploit it to learn more about the density of states for various systems. We will prove useful results by going over the three examples for which we constructed Green's functions.

3.3 EXAMPLES

Example 1: Free particle

$$G_0(\vec{x},\vec{y},E-i\epsilon) = i\hbar \int \frac{d^3p}{(2\pi\hbar)^3}\ \frac{e^{i\vec{p}\cdot(\vec{x}-\vec{y})/\hbar}}{E-i\epsilon-p^2/2m}$$

The angular integration is easily carried out with the result,

$$G_0(\vec{x},\vec{y},E-i\epsilon) = \frac{i}{2\pi^2\hbar r}\int_0^\infty dp \frac{p\sin pr/\hbar}{E-i\epsilon-p^2/2m}$$

Now change the integration variable to $k \equiv p/\hbar$, and analyze the structure of the integrand in the complex plane,

$$\begin{aligned}
\frac{1}{E-i\epsilon-k^2\hbar^2/2m} &= \frac{i\epsilon}{\left(E-k^2\hbar^2/2m\right)^2+\epsilon^2} + \text{Real part} \\
&= i\pi\delta\left(E-\hbar^2k^2/2m\right) + \text{Real part} \\
\Re G_0(\vec{x},\vec{y},E-i\epsilon) &= -\frac{\hbar}{2\pi r}\int_0^\infty kdk \sin kr\ \delta\left(E-\frac{\hbar^2k^2}{2m}\right) \\
&= -\frac{m}{2\pi\hbar}\ \frac{\sin kr}{r}
\end{aligned}$$

with

$$k \equiv \sqrt{\tfrac{2mE}{\hbar^2}}\ .$$

Since $\lim_{x\to 0}\sin x/x = 1$,

$$\Re G_0(\vec{x},\vec{x},E-i\epsilon) = -\frac{mk}{2\pi\hbar}$$

and from (3.18),

$$\boxed{\rho_\epsilon(E) = \frac{mk}{2\pi^2\hbar^2}\times \text{Volume}} \tag{3.19}$$

which is independent of ϵ, so ϵ can be taken to zero. This agrees with our earlier, naive calculation, $\rho(E) = \frac{4m\pi}{(2\pi\hbar)^3}\sqrt{2mE}\,V$.

Example 2: Free Particle Near a Plane Boundary

Choose the boundary to be the plane at $x_3 = 0$. From our previous work

$$G(\vec{x},\vec{y},E) = G_0(\vec{x},\vec{y},E) - G_0\left(\vec{x},\tilde{\vec{y}},E\right)$$

where $\tilde{y} = (y_1, y_2, -y_3)$. Using results from the previous example,

$$\Re G(\vec{x}, \vec{y}, E - i\epsilon) = -\frac{m}{2\pi\hbar}\left(\frac{\sin kr}{r} - \frac{\sin k\tilde{r}}{\tilde{r}}\right)$$

where $r \equiv |\vec{x} - \vec{y}|$ and $\tilde{r} = \left|\vec{x} - \tilde{\vec{y}}\right|$. The first term gives our previous result, so

$$\rho_\epsilon(E) = \frac{m}{2\pi^2\hbar^2}\left(kV - \int d^3x \frac{\sin k\tilde{r}}{\tilde{r}}\Big|_{\vec{x}=\vec{y}}\right) \quad .$$

Let's calculate $\tilde{r}$ at $\vec{x} = \vec{y}$:

$$\tilde{r} = \left[(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 + y_3)^2\right]^{1/2} \to |2x_3| \quad \text{when} \quad \vec{x} = \vec{y} \quad .$$

So we need the integral

$$\begin{aligned} I &\equiv \int d^3x \frac{\sin 2kx_3}{2x_3} = S\int_{-\infty}^{0} dx_3 \frac{\sin 2kx_3}{2x_3} \\ &= \frac{S}{2}\int_{-\infty}^{0} \frac{du}{u}\sin u \quad . \end{aligned}$$

This is a well-known integral which is $\pi/2$. Here, S is the area of the bounding surface. So

$$\rho(E) = \frac{m}{2\pi^2\hbar^2}\left(kV - \frac{\pi}{4}S\right) \tag{3.20}$$

(having taken $\epsilon \to 0$). Thus the density of states is modified near a plane surface on which the Schrödinger wave function vanishes. The modification is proportional to the surface area.

This is a very interesting result with many applications. Here are a series of comments.

1. Physical origins: Because $\psi = 0$ at the boundary, the states available to a particle are depleted near the boundary, hence the minus sign in (3.20). This is also the reason that the effect is proportional to the surface area.

2. A simpler form: Replace $\rho(E) = dN/dE$, by $\rho(k) = dN/dk = \rho(E)dE/dk$

$$\begin{aligned} \rho(k) &= \frac{\hbar^2 k}{m}\rho(E) \\ \rho(k) &= \frac{1}{2\pi^2}\left(k^2V - \frac{\pi k}{4}S\right) \quad . \end{aligned} \tag{3.21}$$

3. Our derivation assumed a plane boundary, but we found that the modification of $\rho(k)$ comes from distances of order $1/k$ from the boundary. From close enough, any smooth surface looks like a plane, so we suspect — and indeed it is the case — that the surface modification of $\rho(k)$ in (3.21) is accurate for any smooth surface when k is large enough. More complicated aspects of the geometry of the boundary surface give rise to corrections to $\rho(k)$ which grow more slowly or even fall with k. These dependences have been worked out [Ref. R. Balian and C. Bloch, *Ann. Phys.* (NY) **60**, 401 (1970)] and give decreasing powers of k multiplying progressively more detailed aspects of the geometry

$$\rho(k) = \frac{1}{2\pi^2}\left(k^2V - \frac{\pi}{4}kS + \frac{1}{3}\oint d^2s \frac{1}{2}\left(\frac{1}{R_1}+\frac{1}{R_2}\right) + \ldots\right) \quad . \tag{3.22}$$

Here R_1 and R_2 are the principal radii of curvature defined at each point on the surface. To see the usefulness of this expansion consider a region characterized by a typical length, ℓ. Successive terms are down by powers of $(1/k\ell)$, so the expansion is useful in a fixed region at high density or at fixed density in the limit of large size. [At each point (x_0, y_0) a surface can be approximated by $z = \frac{1}{a}(x-x_0)^2 + \frac{1}{b}(y-y_0)^2$, which is a paraboloid ($a, b > 0$ or $a, b < 0$) or a hyperboloid ($a > 0$, $b < 0$ or vice versa). Note, the x and y axis are to be oriented in the tangent plane oriented in a way such that no $x - y$ cross terms appear. Then $R_1 = \sqrt{|a|}$, $R_2 = \sqrt{|b|}$. For a sphere $R_1 = R_2 = R$. For more detail see standard introductions to differential geometry.]

4. Generality: Our result is actually a property of the wave equation, $(\nabla^2 + k^2)\phi = 0$, but the surface modifications depend on the form of the boundary condition. A "Neumann boundary condition," $\hat{n} \cdot \vec{\nabla}\phi = 0$, yields a different surface term. In fact, the coefficient of S is positive in this case. This is to be expected: for this boundary condition, ϕ tends to have a maximum at the surface, so the availability of states is enhanced on account of the surface.

   The first term, $\rho(k) \sim \frac{1}{2\pi^2}k^2V$, is asymptotically ($k \to \infty$) exact provided the surface satisfies appropriate smoothness conditions (shown by Hermann Weyl (1911)). All other terms are boundary dependent.

   The general subject of the relation between the geometry of a manifold and the distribution of eigenvalues of the wave equation is a subject of current research in advanced mathematics as well as one of practical interest in acoustics, structural engineering and music!

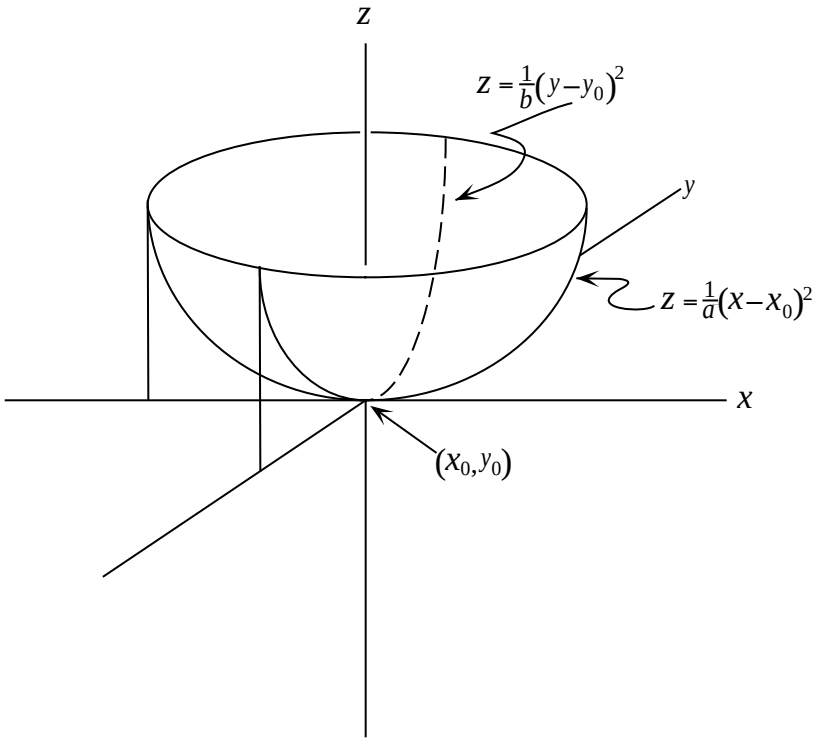


Figure 3.3: Parabolic approximation to a surface at x_0 defines the principal radii of curvature.

Example 3: Particle in a Potential $V(\vec{x})$

We can use the expansion of $G(\vec{x},\vec{y},E)$ (3.16) to find the effect of the presence of a potential $V(\vec{x})$ on the density of states. We shall study a specific case, namely an external potential whose volume integral exists so we can define the volume average of $V(x)$:

$$\langle V(x)\rangle \equiv \frac{1}{V}\int d^3x\, V(\vec{x}) \quad .$$

We require

$$\rho_\epsilon(k) = -\frac{\hbar k}{\pi m}\Re\int d^3x\, G(\vec{x},\vec{x},E-i\epsilon)$$

where now

$$G(\vec{x},\vec{y},E) = G_0(\vec{x},\vec{y},E) + G_1(\vec{x},\vec{y},E) + \ldots$$

as given in (3.16). We compute the effect of the first interaction dependent term, G_1. Using the explicit form of G_0 (3.11), setting $\vec{x}=\vec{y}$ and performing the coordinate space integrations,

$$\int d^3x\, G_1(\vec{x},\vec{x},E-i\epsilon) = \frac{i\hbar\int d^3x\, V(x)}{(2\pi)^3}\int d^3k\left(E-i\epsilon-\frac{\hbar^2k^2}{2m}\right)^{-2} \quad .$$

If we now write $\left(E - i\epsilon - \frac{\hbar^2 k^2}{2m}\right)^{-2} = -\frac{d}{dE}\left(E - i\epsilon - \frac{\hbar^2 k^2}{2m}\right)^{-1}$ the real part can be recognized as a δ-function, and with a bit of algebra we obtain

$$\rho(k) = \lim_{\epsilon \to 0} \rho_\epsilon(k) = \frac{k^2 V}{2\pi^2}\left(1 - \frac{m\langle V(x)\rangle}{\hbar^2 k^2} + \ldots\right)$$

where the ellipses denote the higher terms in the expansion of $G(\vec{x}, \vec{y}, E)$, which are further suppressed at large k.

The conclusion of this exercise is that we are justified in ignoring these sorts of interactions at large k (high density). In fact we can consistently keep surface effects $\mathcal{O}(1/k)$ while ignoring interactions $\mathcal{O}(1/k^2)$. Of course other sorts of interactions, such as two-body forces between fermions remain to be analyzed and (as far as I know) require the full apparatus of many-body theory which is beyond the syllabus of this course.

3.4 SURFACE TENSION OF NUCLEI AND QUARK MATTER

The modification of the density of states due to the boundary generates a surface area dependent term in the energy. To study this in detail we would have to treat the statistical mechanics of a system dependent on three extensive variables, volume, V, surface area S, and particle number N. Rather than go off in this direction, we will do a couple of simple examples to illustrate the importance of this effect.

The surface energy of nuclei

Consider *nuclei*, which behave roughly like a gas of Z non-relativistic protons and N non-relativisitic neutrons ($Z + N = A$), bounded by a surface on which $\psi = 0$, and with k_F independent of A at large A. The dynamics behind this is the short-range repulsion between nucleons which keeps the density from growing large. We calculate the energy as a function of k_F (which we view as a parameter) and A, the particle number:

$$A = g\int_0^{k_F} \frac{dA}{dk} dk = \frac{g}{2\pi^2}\left(\frac{1}{3} V k_F^3 - \frac{\pi}{8} S k_F^2\right) \tag{3.23}$$

$$U = g\int_0^{k_F} \left(\frac{\hbar^2 k^2}{2m}\right)\frac{dA}{dk} dk = \frac{g\hbar^2}{4\pi^2 m}\left(\frac{1}{5} V k_F^5 - \frac{\pi}{16} S k_F^4\right) \tag{3.24}$$

The surface depletes the number of states at a given energy, thus the more surface, the higher the energy for fixed k_F. To minimize the surface/volume ratio a large

nucleus will (other things being equal) adjust to a spherical configuration: $V = \frac{4}{3}\pi R^3$, $S = 4\pi R^2$, so

$$A \equiv n_V R^3 - n_S R^2 + \mathcal{O}(R) \tag{3.25}$$

$$U = u_V R^3 - u_S R^2 + \mathcal{O}(R) \tag{3.26}$$

with

$$n_V = \frac{2g}{9\pi} k_F^3 \qquad u_V = \frac{g}{15\pi} \frac{\hbar^2 k_F^5}{m} \tag{3.27}$$

$$n_S = \frac{g}{4} k_F^2 \qquad u_S = \frac{g}{16} \frac{\hbar^2 k_F^4}{m} \tag{3.28}$$

We solve in the approximation $k_F R >> 1$, *i.e.* large nuclei, $A >> 1$. From (3.25)

$$R \cong R_0 + \delta R = \left(\frac{A}{n_V}\right)^{1/3} + \frac{1}{3}\left(\frac{n_S}{n_V}\right) + \mathcal{O}\left(\frac{1}{A^{1/3}}\right) \tag{3.29}$$

which can be substituted into (3.26),

$$U = \left(\frac{u_V}{n_V}\right) A + \left(\frac{u_V n_S - u_S n_V}{n_V^{5/3}}\right) A^{2/3} + \mathcal{O}(A^{1/3}) \quad . \tag{3.30}$$

The second term is a surface correction to the energy which vanishes in comparison with the leading term at large A. Upon substituting from (3.27) and (3.28) we obtain

$$U = \frac{\hbar^2 k_F^2}{2m}\left(\frac{3}{5}A + \frac{9\pi}{80}\left(\frac{2g}{9\pi}\right)^{1/3} A^{2/3}\right) + \mathcal{O}(A^{1/3}) \quad .$$

For the case of a nucleus: As determined in Section 1, $p_F \approx 270$ MeV/c, $g = 4$ ($p, n \uparrow$ & $\downarrow$), $mc^2 \cong 940\,\mathrm{MeV}$. So we find $E_{\mathrm{surf}} \approx 9.74\, A^{2/3}\,\mathrm{MeV}$. This should be compared to the surface area

$$S \approx 4\pi R_0^2 = 4\pi\left(\frac{A}{n_V}\right)^{2/3} = 4\pi\left(\frac{9\pi}{2g}\right)^{2/3} \frac{1}{k_F^2} A^{2/3} = 16 A^{2/3}\,\mathrm{fm}^2 \quad ,$$

So we find a "surface tension" (at constant k_F or bulk density) of

$$\frac{E_{\mathrm{surf}}}{S} = \frac{9.74}{16}\,\frac{\mathrm{MeV}}{\mathrm{fm}^2} \cong 0.61\,\frac{\mathrm{MeV}}{\mathrm{fm}^2} \quad .$$

The masses of nuclei are fairly well described by a Fermi gas model with corrections for Coulomb interactions and a "pairing force" which distinguishes between nuclei with even and odd numbers of protons and neutrons. This is all summarized by a *semi-empirical mass formula* first proposed by Weizsäcker:

$$M(Z,N,A) = ZM_p + NM_n - a_1 A + a_2 A^{2/3} + a_3 \frac{Z^2}{A^{2/3}} + a_4 \frac{(Z-N)^2}{A} + \delta(A). \tag{3.31}$$

The first two terms count the masses of the Z protons and N neutrons. The third term includes both the kinetic energy of the "gas" of protons and neutrons and their attractive interactions energy, which also grows like A. The fifth term is the Coulomb interaction energy. The sixth is the "symmetry" energy (see the problem sets) and the last is the pairing energy which is different for nuclei with Z and N even or odd. A nice discussion of the Weizsäcker mass formula can be found in A. de Shalit and H. Feshbach *Theoretical Nuclear Physics* in Section II.3, where the current values of the coefficients a_j are also given. The coefficient a_2 is the "surface energy" to be compared with our calculation $E_{\rm surf} \approx 9.74\, A^{2/3}\,$MeV. The value quoted by Feshbach is $a_2 = 18.56$ MeV, so the surface dependence of the density of states accounts for about half of the surface energy of nuclei. The other half presumably comes from the effect of the surface on the internucleon interaction. [By the way, De Shalit and Feshbach and others calculate the surface energy of a Fermi gas incorrectly and obtain a result too large by a factor of two, thereby giving a surface energy in excellent agreement with the data.]

The surface energy of a quark gas

The same method which we have just applied to nuclei can also be applied to a Fermi gas of up, down and strange quarks. The most important difference is a surprising dependence on the quark mass. The coefficient of S in (3.22) actually depends on the ratio of k to the fermion mass. Equation (3.22) was derived for non-relativistic fermions for which $k << m$ (actually $\hbar k << mc^2$). If the quarks are relativistic, the boundary condition they obey at the surface of the bag in which they are confined is no longer the Schrödinger condition, $\psi = 0$. The resulting modification of (3.22) is

$$\rho(k) = \frac{1}{2\pi^2}\left(k^2 V - \frac{\pi}{4} k S \left\{1 - \frac{2}{\pi}\tan^{-1}\frac{k}{m}\right\} + \ldots\right) \tag{3.32}$$

Note that in the extreme relativistic limit, $k/m >> 1$, the density of states becomes independent of the surface, so the surface tension of the quark gas vanishes. Up and down quarks are so light that they are ultra-relativistic in all cases of interest, so they do not contribute to the surface tension of quark matter at all. The mass of the strange quark is large enough so that strange quarks, if present, do contribute to the surface tension. The surface tension of quark matter in equilibrium with the weak interactions (as described in §2f) therefore varies rapidly as a function of density. At low density, $k_F << m_s$, so there are essentially no strange quarks present, and the surface tension vanishes. As the density increases, the k_F becomes comparable to m_s and strange quarks appear in the quark gas, so the surface tension increases. Eventually, when $k_F >> m_s$, even the strange quark mass is negligible and the surface tension again becomes small. This effect is illustrated in Fig. 3.4, where the surface tension of quark matter is plotted versus the mass of the strange quark for a fixed value of the energy per unit of baryon number in the $V \to \infty$ limit.

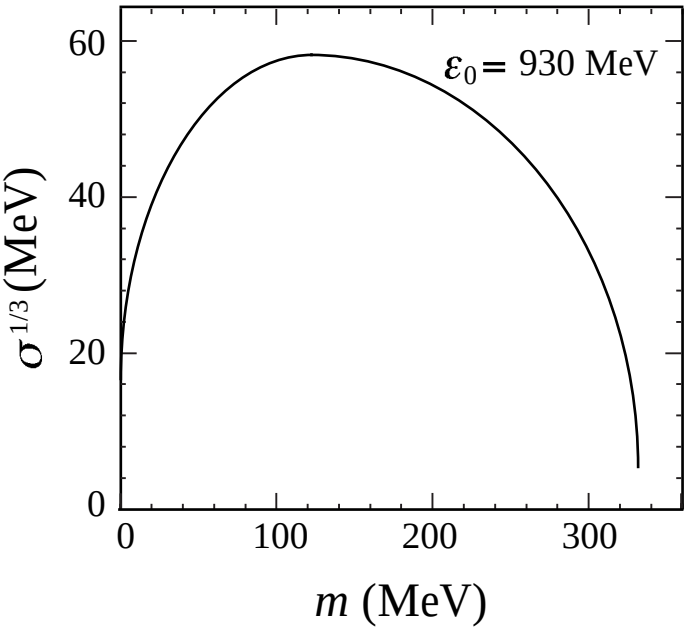


Figure 3.4: The surface tension of quark matter as a function of the strange quark mass.

Chapter 4

THE THOMAS–FERMI MODEL OF ATOMS

It is an experimental fact that atoms do not grow appreciably in size as we go through the periodic table. The physical reason for this is the increasing nuclear charge, Z, which pulls the innermost electrons inward. [The radius of the first Bohr orbital in an atom with nuclear charge Ze is $a_0^Z = \hbar^2/mZe^2$.] These inner electrons shield the outer electrons from the full nuclear charge. The outermost electron (we'll define this more carefully in the course of this section) sees a "nuclear" charge of order e and has an orbit with radius of order $a_0^{Z=1}$ ($\equiv a_0$).

Thus the electron cloud in an atom becomes dense at high Z and it might make sense to apply degenerate fermion system methods to atomic electrons. To see if this is so let us estimate $k_F R$. As we saw in § 2, we expect to be okay if $k_F R \gg 1$. Assuming the electrons are non-relativistic and ignoring interactions we approximate,

$$n \cong \frac{k_F^3}{3\pi^2}$$

from (1.7) ($g = 2$). From the geometry, $n \cong 3Z/4\pi R^3$, so

$$k_F R \cong \left(\frac{9\pi}{4} Z\right)^{1/3} \quad .$$

For $Z = 20$, $k_F R \cong 5.2$; for $Z = 50$, $k_F R \cong 7.1$; for $Z = 90$, $k_F R \cong 8.6$, so the approximation looks okay even for moderately small atoms.

The goal of the Thomas–Fermi (TF) approach is to characterize an atom and its properties in terms of the electron density, $\rho(\vec{r})$, which can be estimated using degenerate fermion methods. The advantage of this method is its simplicity and physical origins. The disadvantage is that a certain amount of the quantum mechanical information contained in the phase of the wavefunction is lost. It is relatively easy to

include the electrons' kinetic energy and their interaction with the nuclear charge. Likewise, the electrostatic repulsion between electrons can be expressed as a function of $\rho(\vec{r})$. Dirac showed how to include exchange forces between electrons, but other, more subtle effects like spin-orbit coupling are lost. The result is an approximate, generic description of an atom, lacking many subtleties such as shell closures, but very useful for first estimates of atomic properties.

4.1 ATOMIC ENERGY AS A FUNCTION OF THE ELECTRON DENSITY

Our aim here is to obtain an expression for the energy of an atom, $E[\rho]$, as it depends on the function $\rho(\vec{r})$. E, a number, depends on the function, $\rho(\vec{r})$. Such a relation is known mathematically as a "functional." Functionals figure in Lagrangian mechanics and the calculus of variations. Once we have $E[\rho]$, we will find the function $\rho(\vec{r})$ which minimizes it subject to the constraint that $\int d^3r\,\rho(r) = N$, where N is the number of electrons in the atom.

Consider a small volume element, Ω, at a distance r from a nucleus. We treat the electrons *in this volume element* as a degenerate Fermi gas, characterized by a Fermi momentum, $k_F(r)$ ($\equiv p_F(r)/\hbar$) and density, $\rho(r)$, which can vary with r. Notice that we take ρ and k_F to be functions of $r \equiv |\vec{r}|$ alone. Although the electrons interact, we know from § 3 that to a good approximation we may use the free density of states, so

$$\rho(r) = \frac{k_F^3(r)}{3\pi^2} \tag{4.1}$$

where we've remembered the degeneracy factor of 2 (for spin up and down). The electrons' kinetic energy density, $t(r)$ can be expressed in terms of $k_F(r)$ and from (4.1), in terms of $\rho(r)$:

$$t(r) = \frac{\hbar^2 k_F^5(r)}{10m\pi^2} = \frac{\hbar^2}{10m\pi^2}\left(3\pi^2\rho(r)\right)^{5/3} \quad . \tag{4.2}$$

The total electron kinetic energy, the integral of $t(r)$, is a functional of ρ:

$$T[\rho] \equiv \int d^3x\, t(r) = \int d^3x \frac{\hbar^2\left(3\pi^2\rho\right)^{5/3}}{10m\pi^2} \tag{4.3}$$

The electron's potential energy consists of two terms,

$$V[\rho] = -\int d^3x \frac{Ze^2\rho(r)}{r} + \frac{e^2}{2}\int d^3x_1\, d^3x_1 \frac{\rho(r_1)\rho(r_2)}{|\vec{x}_1 - \vec{x}_2|} \tag{4.4}$$

the first due to the interaction with the nucleus, the second from interelectron repulsion. So we can express the Hamiltonian (energy) as a functional of ρ:

$$H[\rho] = T[\rho] + V[\rho] \quad . \tag{4.5}$$

Notice that to this order the only quantum mechanics which has entered this calculation is the Exclusion Principle.

4.2 THE THOMAS–FERMI EQUATION

We assume that the ground state of an atom is obtained by minimizing $H[\rho]$ subject to the constraint that N electrons are present. We enforce the constraint by means of a Lagrange multiplier, $H[\rho] \to H[\rho, \lambda]$

$$H[\rho, \lambda] = T[\rho] + V[\rho] - \lambda \left(N[\rho] - N \right) \tag{4.6}$$

where

$$N[\rho] \equiv \int d^3x \, \rho(r) \quad . \tag{4.7}$$

We look for extrema, $\delta H/\delta\rho(r) = 0$[1] and $\partial H/\partial\lambda = 0$, which will turn out to be minima. From $\delta H/\delta\rho = 0$,

$$\frac{\hbar^2}{2m} \left(3\pi^2\rho\right)^{2/3} - \frac{Ze^2}{r} + e^2 \int d^3x_2 \frac{\rho(r_2)}{|\vec{x} - \vec{x}_2|} = \lambda \tag{4.8}$$

and from $\partial H/\partial\lambda = 0$,

$$\int d^3x \, \rho(r) = N \quad . \tag{4.9}$$

Equation (4.8) can be viewed from another, quite enlightening point of view. The first term is $k_F^2/2m$, the kinetic energy of an electron at the top of the Fermi sea in the volume Ω. The next two terms are the potential energy of that electron. Thus, the left-hand side is the Fermi energy, the energy of an electron at the top of the Fermi sea in Ω. (4.8) requires the Fermi energy to be independent of r. This is a reasonable equilibrium constraint since if E_F were not independent of r, electrons would move from regions of high Fermi energy to low, thereby reducing the energy of the atom.

Although these equations look difficult to solve, they may be converted to a simple differential equation for the electrostatic potential, $V(r)$, defined by

$$V(r) = -\frac{Ze^2}{r} + e^2 \int d^3x_2 \frac{\rho(r_2)}{|\vec{x} - \vec{x}_2|} \quad . \tag{4.10}$$

[1] $\delta/\delta\rho$ denotes a functional derivative. It is a shorthand for the sort of manipulations familiar from the calculus of variations (described in texts on classical mechanics such as Goldstein).

$V(r)$ obeys Poisson's equation,

$$\nabla^2 V(r) = -4\pi e^2\,\rho(r) \qquad (\text{for } r \neq 0) \quad . \tag{4.11}$$

It's convenient to work with $U(r)$ defined by

$$U(r) \equiv V(r) - \lambda \tag{4.12}$$

which also obeys (4.11) — since λ is a constant.

Notice, now, that (4.8) relates $\rho(r)$ to $U(r)$

$$U(r) = -\frac{\hbar^2}{2m}\left(3\pi^2\rho\right)^{2/3} \quad . \tag{4.13}$$

So, combining (4.11) – (4.13), and using $\nabla^2 U = 1/r\ d^2/dr^2\, rU$ (for a function of r alone),

$$\frac{1}{r}\ \frac{d^2}{dr^2} rU = -\frac{4\pi\, e^2}{3\pi^2}\left(-\frac{2mU}{\hbar^2}\right)^{3/2} \quad . \tag{4.14}$$

To simply this equation we introduce a new dependent variable,

$$\Phi(r) = -\frac{r}{Z\, e^2} U(r) \tag{4.15}$$

and a new scale for distance,

$$r \equiv bx \tag{4.16}$$

with the constant b to be determined. Substituting into (4.14) we find

$$\Phi'' = \frac{1}{\sqrt{x}}\Phi^{3/2} \tag{4.17}$$

where $'$ denotes d/dx, provided we choose

$$b \equiv \frac{(3\pi)^{2/3}}{2^{7/3}}\left(\frac{\hbar^2}{m\, e^3}\right) Z^{-1/3} = 0.885\, a_0 Z^{-1/3} \quad . \tag{4.18}$$

Equation (4.17) is the "Thomas–Fermi Equation." It is a remarkable, and at first sight, puzzling result, for it suggests that the electron distribution in atoms is universal except for a scale factor, b, which varies slowly with Z. This is not quite correct, for we have not yet found the boundary conditions necessary to obtain specific solutions to (4.17). Nevertheless, some of this universality will be preserved.

4.3 BOUNDARY CONDITIONS AND SOLUTIONS TO THE THOMAS–FERMI EQUATION

From its definition, (4.12), and the equilibrium condition, (4.8), it is clear that $U(r)$ is negative, and $\Phi(x)$ must be positive:

$$\Phi(x) \geq 0 \quad . \tag{4.19}$$

Further, as $r \to 0$, inside the electron cloud, the second term in (4.10) becomes negligible compared to the first, so $rV(r) \to -Z\,e^2$ as $r \to 0$, which means

$$\Phi(0) = 1 \tag{4.20}$$

The Thomas–Fermi equation is a non-linear, second order ordinary differential equation. A second boundary condition, beyond $\Phi(0) = 1$, is required to uniquely specify a solution. To understand the origins of this condition consider integrating the TF equation outward from $x = 0$ with $\Phi(0) = 1$ and with various choices of $\Phi'(0) \equiv -\mu$, which we use to parameterize solutions. Since $\Phi > 0$, the solutions to the Thomas–Fermi equation are convex, *i.e.* $\Phi'' \geq 0$. For large values of μ, $\Phi(x)$ decreases rapidly to zero at a finite value of x, $x_0(\mu)$. We call solutions which vanish at some $x_0(\mu)$ Class I Solutions. As μ is decreased, $x_0(\mu)$ grows until at $\mu = \mu_0 = 1.5880710\ldots$ a special solution is obtained for which $\Phi(x)$ tends to zero only as $x \to \infty$. This special solution, labelled II in the figure will be denoted $\Phi_0(x)$. For $\mu < \mu_0$ the solutions grow without bound (Class III).

To characterize these solutions let us integrate $\rho(r)$ out to some finite radius $R \equiv bX$

$$\int_{r<R} d^3x\,\rho(r) = N(R) = 4\pi \int_0^R dr\,r^2 \left(-\frac{1}{4\pi e^2}\;\frac{1}{r}\;\frac{d^2}{dr^2} rU(r) \right) \quad . \tag{4.21}$$

After substituting Φ and scaling to X we find

$$N(X) = Z\,[x\Phi' - \Phi]_0^X \tag{4.22}$$

or

$$N(X) = Z\,(X\Phi'(X) - \Phi(X) + 1) \quad . \tag{4.23}$$

The quantity $X\Phi'(X) - \Phi(X)$ has a simple physical interpretation: it is related to the radial electric field $E_r = -e\frac{dV}{dr}$,

$$E_r \propto \frac{1}{x^2}\,(\Phi - x\Phi'(x)) \quad . \tag{4.24}$$

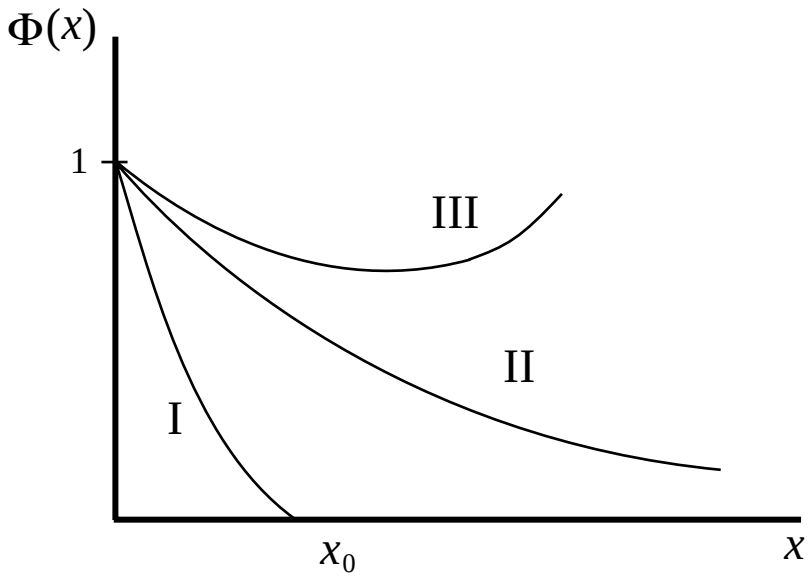


Figure 4.1: Solutions to the Thomas-Fermi Equation

Electrons should not bind to an atom in this approximation unless E_r is positive (attractive to an electron) at the surface, so we add the restriction

$$\Phi(x) - x\Phi'(x) \geq 0 \tag{4.25}$$

to our list of conditions on the Thomas–Fermi equation. This constraint is trivially satisfied by solutions of Class I and II ($\Phi > 0$ and $\Phi' < 0$ everywhere), but it is significant for Class Eqs. III solutions, where it determines the maximum x ($\equiv x_1$), obeying

$$\Phi(x_1) = x_1\Phi'(x_1) \tag{4.26}$$

to which the equation can be integrated. x_1 can be determined graphically by the construction shown in Fig. (4.2), as the point where the tangent to $\Phi(x)$ passes through the origin. Armed with this information let's explore the properties of the three classes of solutions, beginning with Class II.

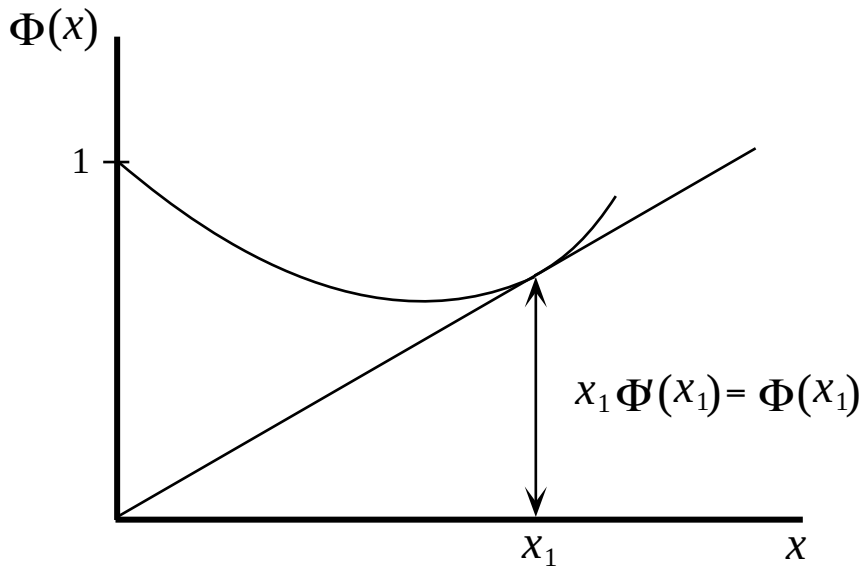


Figure 4.2: Graphical construction for finding a specific solution to the TF equation of Class III

Class II Solution: Neutral Atoms

Here $\lim_{x\to\infty}\Phi(x) = \Phi'(x) = 0$ so from Eq. (4.23) we find $N = Z$, *a neutral atom*. The electronic properties of this neutral Thomas–Fermi atom are entirely specified by $\Phi_0(x)$ (together with the definitions which relate Φ and x back to physical observables). The asymptotic form of $\Phi_0(x)$ at large x is given by

$$\Phi_0(x) \sim \frac{144}{x^3} \quad . \tag{4.27}$$

In fact (as can be checked by direct substitution) $144/x^3$ is an exact solution to the TF equation but is not physically interesting because it fails to satisfy $\Phi(0) = 1$. The approximation (4.27) is of limit practical usefulness since it deviates from $\Phi_0(x)$ by as much as 40% for x-values as large as 100. For small x, $\Phi_0(x)$ is well-approximated by the power series expansion,

$$\Phi_0(x) \cong 1 - \mu_0 x + \frac{4}{3}x^{3/2} + \mathcal{O}\left(x^{5/2}\right) \quad .$$

$\Phi_0(x)$ can be obtained numerically directly by integrating the Thomas–Fermi equation with $\mu = \mu_0$. Values of $\Phi_0(x)$ are listed in Table 4.1 and shown in Fig. (4.3):[2]

The electron density $\rho(r)$ falls very slowly in the TF approximation to a neutral atom,

$$\rho \sim (-U)^{3/2} \sim \left(\frac{\Phi}{r}\right)^{3/2} \sim \frac{1}{r^6}$$

[2]From Landau and Lifshitz.

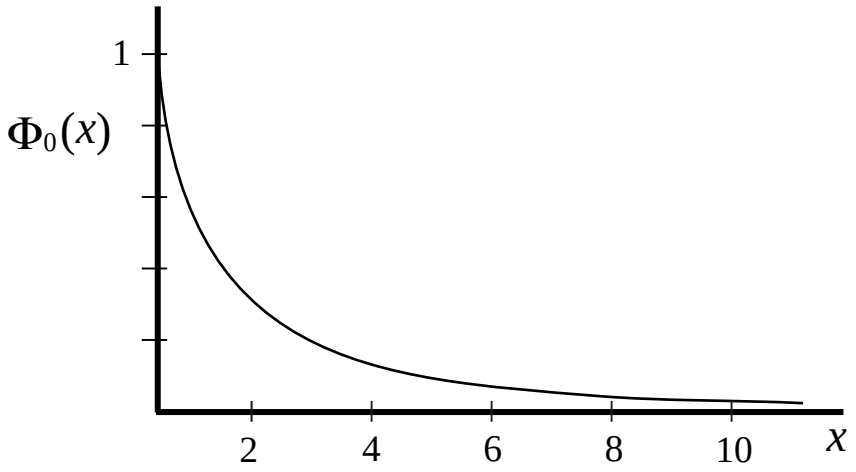


Figure 4.3: The Thomas–Fermi function $\Phi_0(x)$.

whereas experiment and direct solutions to the Schrödinger equation tell us $\rho(r)$ falls exponentially at large r. [An electron bound by an energy B has an $\exp -\sqrt{2mB}\, r$ tail to its wavefunction at large r.] One should not expect the TF model to work well at large r where $\rho(r)$ is small since the TF approximation is, by its very nature, a high density one. It does better with aggregate descriptions of atomic properties. For example, the fact that all neutral atoms are described by the same scaled equation predicts that properties of neutral atoms scale with powers of b. Thus the root mean square charge radius should scale like b

$$\sqrt{\langle r^2 \rangle} \sim b \sim Z^{-1/3} \quad .$$

Atoms should get *smaller* with increasing Z, and they do. Similarly, the total binding energy of a Thomas–Fermi atom should scale like $Z^{7/3}$,[3]

$$E = -20.93\, Z^{7/3}\, \mathrm{eV}$$

which interpolates fairly well the energies of high Z atoms.

Of course the Thomas–Fermi model is not sensitive to the shell structure which is evident in so many atomic properties. If, for example, one plots the electron density for a heavy element calculated in some model which incorporates shell structure, one sees that the Thomas–Fermi model smoothly interpolates the density coming from the more detailed model. The curves shown in Fig. (4.4)[4] show the comparison for Mercury.

[3]For this application of the Thomas–Fermi model, see S. Flügge, *Practical Quantum Mechanics*, § 174.

[4]From A. S. Davydov, *Quantum Mechanics*.

x	Φ_0	x	Φ_0	x	Φ_0
0.00	1.000	1.4	0.333	6	0.0594
0.02	0.972	1.6	0.298	7	0.0461
0.04	0.947	1.8	0.268	8	0.0366
0.06	0.924	2.0	0.243	9	0.0296
0.08	0.902	2.2	0.221	10	0.0243
0.10	0.882	2.4	0.202	11	0.0202
0.2	0.793	2.6	0.185	12	0.0171
0.3	0.721	2.8	0.170	13	0.0145
0.4	0.660	3.0	0.157	14	0.0125
0.5	0.607	3.2	0.145	15	0.0108
0.6	0.561	3.4	0.134	20	0.0058
0.7	0.521	3.6	0.125	25	0.0035
0.8	0.485	3.8	0.116	30	0.0023
0.9	0.453	4.0	0.108	40	0.0011
1.0	0.424	4.5	0.0919	50	0.00063
1.2	0.374	5.0	0.0788	60	0.00039

Table 4.1: Values of the Function $\Phi_0(x)$

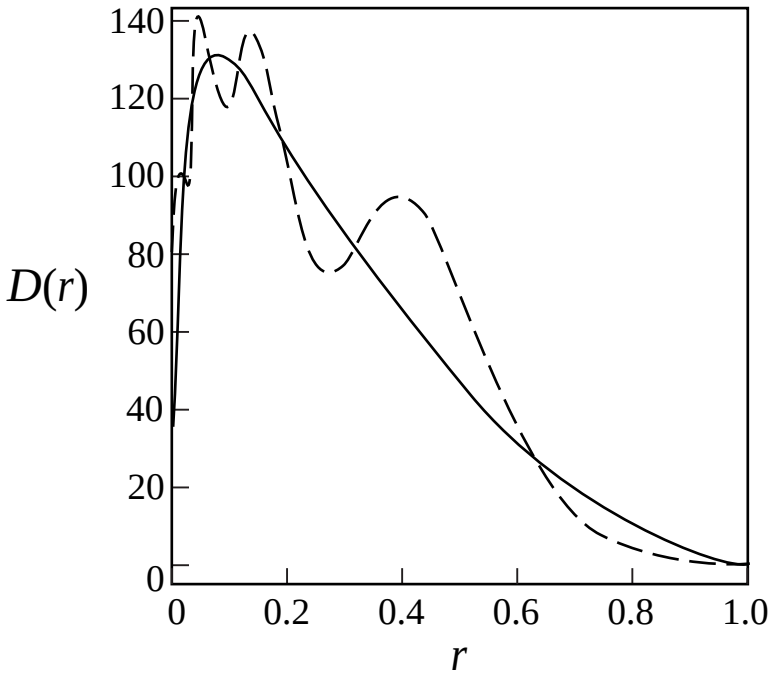


Figure 4.4: The electron density in Mercury. The solid curve is the TF approximation. The dashed curve is a self-consistent field approximation. r is measured in units where $a_0 = \hbar^2 \ \ me^2 = 1$. [From A. S. Davydov, Quantum Mechanics.]

Class I Solutions: Positive Ions

Here $\Phi_0(x) = 0$ at some finite x_0 where $\Phi'(x_0) < 0$. According to (4.23),

$$N \equiv N(x_0) = Z\,(1 + x_0\Phi'(x_0)) < Z \quad . \tag{4.28}$$

Apparently these solutions correspond to positive ions ($N < Z$). The degree of ionization then is specified by

$$x_0\Phi'(x_0) = \frac{N - Z}{Z} \equiv -\frac{z}{Z} \quad .$$

Note that, in contrast to neutral atoms, positive ions are finite in size in the TF approximation. Their properties are determined by the parameter z/Z as well as the scale parameter b.

Class III Solutions: Neutral Atoms Under Pressure

For these solutions one integrates $\Phi(x)$ out until $E_r = 0$, *i.e.* out to x_1 determined by (4.26). Then from (4.23)

$$N \equiv N(x_1) = Z$$

we see that these solutions, like those of Class II, correspond to neutral atoms. However $\rho(r)$ does not vanish at $R = bx_1$, so these "atoms" differ qualitatively from the

ones studied earlier. The difference can be understood by calculating the pressure exerted by the electrons at $r = R$: From § 1

$$P = \frac{\hbar^2 k_F^5}{15m\pi^2} = \left(\frac{\Phi(x_1)}{x_1}\right)^{5/2} \quad . \tag{4.29}$$

This non-zero electron pressure can be sustained in static equilibrium if it is countered by some external pressure whose origins might for example be gravitational.

Notice that there are no solutions to the TF model corresponding to negative ions.

4.4 APPLICATIONS OF THE THOMAS–FERMI MODEL

The Thomas–Fermi model provides a reasonable description of both the electron density, $\rho(r)$, and the electrostatic potential, $V(r)$, in regions where the electron density is large. Near the boundaries of atoms the electron density is low and TF does not work so well (as we have just seen in the case of neutral atoms). It works best to describe average properties of atoms, such as the total binding energy and the charge distribution which can be measured by electron scattering from atoms. A TF calculation of the total binding energy is given by Flügge (problem 174). The TF treatment of electron scattering is described below in the section on scattering theory.

One interesting application of the TF picture of atoms makes use of the expression for the electrostatic potential, $V(r)$, which emerges from the model. $V(r)$ can be regarded as an approximation to the potential in which the individual electrons in the atom move. With this in mind one can, for example, estimate the value of the nuclear charge (Z) at which electronic bound states with a given value of orbital angular momentum ℓ first appear. An electron moving in the TF potential $V(r)$ obeys the radial Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2}{dr^2}u(r) + \left(V(r) + \frac{\ell(\ell+1)\hbar^2}{2mr^2}\right)u(r) = Eu(r) \quad . \tag{4.30}$$

Since a bound state corresponds to $E < 0$ and since the kinetic term, $-\frac{\hbar^2}{2m}u''$, will contribute positively to the energy, a bound state is not possible unless the "effective potential" is negative,

$$V_{\text{eff}}(r) \equiv V(r) + \frac{\ell(\ell+1)\hbar^2}{2mr^2} < 0 \quad , \tag{4.31}$$

for some range of r. Equation (4.31) is a necessary condition for a bound state. It is also close to sufficient. [The argument mimics the proof that attractive one-dimensional potentials always have bound states.] Converting to the scaled TF variables, (4.31) becomes

$$0.885\, Z^{2/3} 2x\Phi(x) > \ell(\ell+1) \quad . \tag{4.32}$$

A numerical study of the function $2x\Phi_0(x)$ shows that it has a rather broad maximum near $x \cong 2$ [see figure] with $\max[2x\Phi_0] \cong 0.973$, so we have the condition

$$Z > 1.26\,[\ell(\ell+1)]^{3/2} \tag{4.33}$$

so we get for $\ell = 1, 2, 3, 4$, $Z > 2.8$, 18.5, 62.4, and 112.7, respectively. These can be compared with the observed Z values at which the p, d and f shells begin to fill, namely $Z(p) = 5$, $Z(d) = 21$ and $Z(f) = 58$. Not bad for such a simple model.

The Thomas–Fermi potential is a good starting place for another, more sophisticated treatment of atomic structure: the "self-consistent field method" known also as the "Hartree–Fock" method. This method is described in some detail in a later section of the course. The basic idea is that the wavefunctions of atomic electrons give rise to an effective single electron potential in which the solutions to the Schrödinger equation are the wavefunctions themselves. Self-consistency is achieved by iteration. It is very useful to begin with a set of simple electron wavefunctions motivated by the problem (as opposed to some generic set like harmonic oscillator eigenstates). Just such a set is provided by the solutions to the Schrödinger equation for the TF potential $V(r)$. In fact, even without iteration, the TF single particle wavefunctions and energies provide a rather good approximation to atomic energy levels (especially when some of the improvements described in the next section are included).

4.5 IMPROVEMENTS OF THE THOMAS–FERMI MODEL

The Thomas–Fermi model just described may appear quite crude, however, it has a better pedigree than our discussion would seem to imply. In fact it can be shown that the TF model yields the exact solution to the non-relativistic N-electron problem in the field of a charge Ze nucleus as $Z \to \infty$. [This is only a *formal* limit, however, since the non-relativistic approximation breaks down as $Z \to \hbar c/e^2 \approx 137$.]

The model can be improved by the inclusion of various corrections. The subject has been reviewed (in a rather formal context) by E. Lieb (*Rev. Mod. Phys.* **53** (1981) 603). Here we mention two of the more important corrections.

The first, suggested by Amaldi and Fermi, is to remember that the N^{th} electron does not act on itself. Thus the potential $V(r)$, which determines the motion of one electron, should obey a Poisson equation with that electron excluded, so Eq. (4.11) is modified to read

$$\nabla^2 V(r) = -4\pi e^2 \left(\frac{N-1}{N}\right) \rho(r) \quad .$$

This change can be absorbed into a redefinition of parameters so the TF equation remains unchanged. However the boundary condition changes: the outer limit of a neutral atom is defined by the condition $V(R) = -e^2/R$, $dV/dR = e^2/R^2$, since there remains one unit of nuclear charge not screened by the $Z-1$ electrons which contribute to $V(r)$. This correction has the effect of making atoms finite in size. For details see Flügge, § 173.

A second significant correction was suggested and calculated by Dirac. The resulting approximation is known as the Thomas–Fermi–Dirac (TFD) approximation. The Hamiltonian $H[\rho]$ (see (4.3) – (4.5) on which the TF method is based ignores one important Coulomb effect: the exchange energy between identical electrons. The exchange energy is given by

$$\Delta E \equiv -\frac{1}{2} e^2 \sum_{i,j} \delta m_{si}, m_{sj} \int d^3x_1 \, d^3x_2 \frac{\psi_i^*(\vec{x}_1)\psi_j(\vec{x}_1)\psi_j^*(\vec{x}_2)\psi_i(\vec{x}_2)}{|\vec{x}_1 - \vec{x}_2|} \quad . \tag{4.34}$$

The sum on i and j ranges over all electrons. The Kronecker δ-symbol is 1 if the two electrons are in the same spin state, zero otherwise. ΔE can be expressed in terms of the "density matrix"

$$\rho(\vec{x}_1, \vec{x}_2) = \sum_j \psi_j(\vec{x}_1)\psi_j^*(\vec{x}_2) \tag{4.35}$$

but not, in general, in terms of the electron density $\rho(\vec{x})$.

Dirac's simplifying assumption was to treat the electrons as plane waves in the spirit of the Thomas–Fermi approximation. Then ΔE can be evaluated (see, for example, Bethe and Jackiw, pp. 92 – 98) in terms of k_F, the local r-dependent Fermi momentum. k_F in turn can be replaced by ρ by means of Eq. (4.1). The result is

$$\Delta E \to \Delta E[\rho] = -\frac{e^2}{4\pi^2} \int d^3x \, k_F^4(r) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int d^3x \, \rho^{4/3}(r) \quad . \tag{4.36}$$

If this term is added to $H[\rho]$, the analysis proceeds much as before with the result:

$$\Phi'' = x\left(\sqrt{\frac{\Phi}{x}} + \beta\right)^3 \tag{4.37}$$

where b is defined by (4.18) and

$$\beta \equiv \frac{1}{\pi}\sqrt{\frac{b}{2a_0\,Z}} \sim \frac{1}{Z^{2/3}} \quad . \tag{4.38}$$

So the effect of exchange interactions is the appearance of the additive constant β in (4.37). Note: $\beta \to 0$ as $Z \to \infty$ so the exchange interaction is negligible at large Z.

Once again, the edge of an atom requires redefinition because of this modification (see Bethe and Jackiw). The resulting TFD model gives a very good approximation to the bulk properties of heavy atoms. The single particle states calculated in the resulting TFD potential give an excellent description of the energy levels of heavy atoms.

Chapter 5

ELECTRONS IN PERIODIC POTENTIALS

Soon after the discovery of the electron by Thompson in 1987, P. Drude developed an explanation of electric conductivity in metals based on the idea that electrons within a metal form a classical gas. By applying the kinetic theory of gases to the electron gas, Drude was able to obtain a fairly good description of the electric and thermal conductivity of metals as well as other properties. Nevertheless, his theory had conspicuous failures. Some of the most glaring failures of the classical theory were eliminated by Sommerfeld (1928) merely by taking into account the degeneracy of the *quantum* electron gas. The resulting Fermi gas model does a surprisingly good job of describing many properties of metals. It too, however, has notable failures (see the end of § 5a for a short list). Late in the 1920's the Swiss physicist Felix Bloch studied the propagation of electrons in periodic potentials such as are found in crystalline solids. He would that the free Fermi gas model is modified in conceptually simple but profound ways. The modifications resolve many features of the behavior of regular solids left unexplained by Sommerfeld's theory. Bloch's work was one of the earliest triumphs of the quantum theory and forms the foundation for modern solid state physics.

In this short discussion we can barely scratch the surface of this rich and complex subject. The interested student can find an embarrassingly (to me) complete presentation on this subject in Ashcroft and Mermin's book *Solid State Physics*. I have chosen to concentrate on the problem of electron motion in a periodic potential. The unusual and non-classical aspects of this motion are important in determining the basic ground rules of solid state physics and also illustrate nice features of the quantum theory of symmetry. Before getting into this technical problem, I will briefly describe some aspects of the Drude–Sommerfeld theory, illustrating both its successes and the shortcomings which were resolved by Felix Bloch's study of motion in periodic

potentials. After studying "Bloch-waves" I will briefly return to the failures of the Drude–Sommerfeld theory and see how Bloch's work resolves them. This will leave you ready to explore semiconductors, devices and a host of applications, but we will move on in other directions.

5.1 A FERMI GAS MODEL OF ELECTRONS IN METALS

We will concentrate on two easily measured and conceptually simple characteristics of metals: the electric and thermal conductivities. The electric conductivity, σ, is defined by Ohm's law

$$\vec{J} = \sigma \vec{E} \tag{5.1}$$

where $\vec{J}$ is the current density and $\vec{E}$, the electric field. The thermal conductivity, κ, is defined by

$$\overrightarrow{\mathcal{J}} = -\kappa \vec{\nabla} T \tag{5.2}$$

where $\overrightarrow{\mathcal{J}}$ is the flux of heat (energy/(time·area)) directly analogous to $\vec{J}$, and T is the temperature. Of course, T is constant in equilibrium and $\overrightarrow{\mathcal{J}}$ is zero. κ is measured under non-equilibrium conditions.

We assume that a metal consists of a regular array of fixed ions which are not efficient at transporting either heat or charge. These are immersed in a Fermi gas of electrons which are in thermal equilibrium at temperature T and which move freely except for collisions characterized by a mean time between collisions, τ.

Elementary kinetic theory relates σ and κ to the properties of the underlying carriers of charge and heat. For σ the relation is obtained as follows. If the medium consists of charges of density n (number/volume), charge e, and a mean velocity $\vec{v}_d$, then

$$\vec{J} = ne\vec{v}_d \quad . \tag{5.3}$$

Random, thermal velocities do not contribute to the flow of current, only the systematic drift induced by $\vec{E}$ does, hence the subscript d, for "drift," on v. To estimate this contribution we assume that the electrons drift in $\vec{E}$ between collisions in which their velocities are randomized. The drift velocity is obtained by integrating Newton's law $m\dot{\vec{v}}_d = e\vec{E}$, so

$$\vec{v}_d = \frac{e}{m}\vec{E}\tau \tag{5.4}$$

where τ is the mean time between collisions. Combining all this we find

$$\sigma = \frac{ne^2\tau}{m} \quad . \tag{5.5}$$

This is not a very convincing derivation: we should have averaged over an ensemble distribution function $f(\vec{v})$ and been more careful about the definition of τ. The more careful arguments yields the same result, (5.5), with a more precise definition of τ.

A slightly more complicated argument relates κ to the mean square electron speed, v^2, τ and the heat capacity per unit volume of the electrons, $c_v = \frac{1}{V}\,\frac{\partial U}{\partial T}$:

$$\kappa = \frac{1}{3} v^2 \tau c_v \quad . \tag{5.6}$$

The 1/3 is a geometrical factor. [For a derivation, see Morse, *Thermal Physics* or Ashcroft and Mermin.]

Not knowing the time between collisions, τ, we can check our picture of metals by dividing out τ between (5.5) and (5.6),

$$\frac{\kappa}{\sigma} = \frac{1}{3}\,\frac{mv^2 c_v}{ne^2} \quad . \tag{5.7}$$

Let's compare the behavior of this quantity for two cases: (1) a classical gas of electrons, (2) a Fermi gas of electrons. For classical electrons

$$\frac{c_v}{n} = \frac{3}{2}k \tag{5.8}$$

is the heat capacity per electron, while

$$\frac{1}{2}mv^2 = \frac{3}{2}kT \tag{5.9}$$

by equipartition (remember v^2 is the (thermal average) mean square speed of the electrons), so

$$\left(\frac{\kappa}{\sigma}\right)_{\text{classical}} = \frac{3}{2}\left(\frac{k^2}{e^2}\right)T \quad . \tag{5.10}$$

for quantum electrons, the heat capacity is given by (1.83)

$$\frac{c_v}{n} = \frac{C_v}{\bar{N}} = \frac{\pi^2 kT}{2T_F} \tag{5.11}$$

and only those electrons near the top of the Fermi sea participate, so

$$\frac{1}{2}mv^2 \approx E_F = kT_F \quad , \tag{5.12}$$

which gives

$$\left(\frac{\kappa}{\sigma}\right)_{\text{Fermi gas}} = \frac{\pi^2}{3}\left(\frac{k^2}{e^2}\right)T \quad . \tag{5.13}$$

Both the classical and quantum theory predict (κ/σ) to vary proportional to T. The coefficient of proportionality, $\kappa/\sigma T$, is known as the "Lorentz number." The predictions are

Table 5.1: Experimental Lorentz numbers $K/\sigma T$ in units of $10^{-8} W \cdot \Omega / K^{2*}$

Metal	0°C	100°C
Ag	2.31	2.37
Au	2.35	2.40
Cd	2.42	2.43
Cu	2.23	2.33
Ir	2.49	2.49
Mo	2.61	2.79
Pb	2.47	2.56
Pt	2.51	2.60
Sn	2.52	2.49
W	3.04	3.20
Zn	2.31	2.33

*From C. Kittel, *Introduction to Solid State Physics*, 2nd edition (John Wiley & Sons, New York, 1965).

$$\begin{aligned} \bullet \text{ Classical:} \quad \frac{\kappa}{\sigma T} &= 1.1 \times 10^{-8} \frac{\text{watt} - \text{ohm}}{^\circ\text{K}^2} \quad , \quad \text{independent of } T^1 \\ \bullet \text{ Fermi gas:} \quad \frac{\kappa}{\sigma T} &= 2.44 \times 10^{-9} \frac{\text{watt} - \text{ohm}}{^\circ\text{K}^2} \quad , \quad \text{independent of } T^2 \end{aligned} \tag{5.14}$$

Data for a variety of conductors are given in Table 5.1 (from Kittel). Clearly, the Fermi gas value agrees quite well with experiment.

Encouraged by this success, let's look at the electrical or thermal conductivity individually. First define a "mean free path,"

$$\ell = \frac{\tau}{v_F} \quad , \tag{5.15}$$

where, in keeping with the Fermi gas picture, we've used a Fermi velocity $v_F = \sqrt{2kT_F/m}$ to set the scale of velocities. Then, from (5.5),

$$\rho \equiv \frac{1}{\sigma} = \frac{m}{ne^2 v_F \ell} \quad . \tag{5.16}$$

[1]Originally, Drude made an error and obtained 2.2×10^{-8}. His fortuitous agreement with data stimulated much of the subsequent work on electron gas theories of metals.

[2]The result we obtained for the Fermi gas survives a more rigorous derivation.

Remember, as discussed in § 1, the Fermi energy of electrons in metals are quite high ($E_F \approx 7\,\mathrm{eV}$ in copper at room temperature), so the system is highly degenerate for temperatures below, say, 500°K. Thus we expect v_F and n to be essentially independent of temperature over a wide range of T up to at least ~ 500°K. Because of this we can use experimental data on resistivity to obtain the electron mean free path from (5.16). Data on copper are shown in Fig. 5.1 (from Serway, Moses and Moyer).

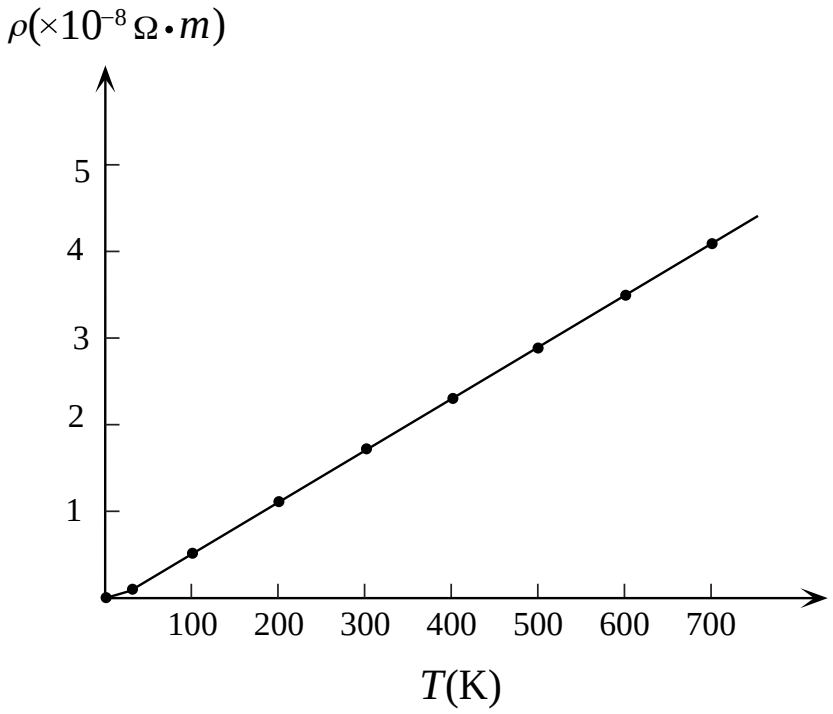


Figure 5.1: The resistivity of pure copper as a function of temperature.

Except at very low temperature, ρ varies linearly with T, so we conclude

$$\ell \sim \frac{1}{T} \quad . \tag{5.17}$$

Furthermore, since the constants in (5.16) are known, we can extract a quantitative estimate of the mean free path. For copper at room temperature ($E_F = 7.05\,\mathrm{eV}$, $n = 8.49 \times 10^{22}\mathrm{cm}^{-3}$) we obtain

$$\ell \approx 390\,\text{Å} \quad . \tag{5.18}$$

This is a remarkable result, roughly 150 times greater than the spacing between atoms. Furthermore, at lower temperature ℓ becomes as much as an order of magnitude larger before leveling off at some very large value as $T \to 0$ (except for the case of

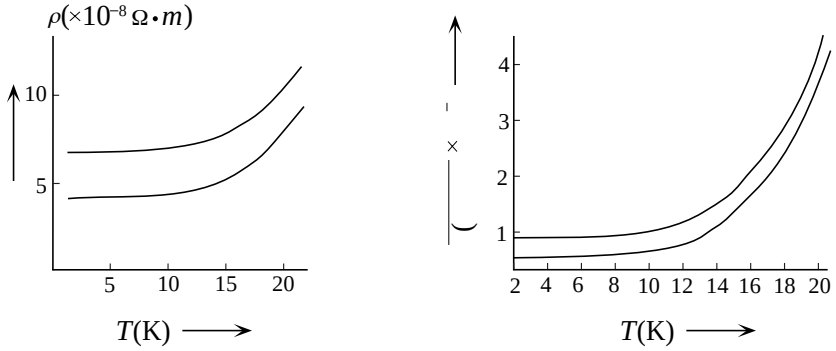


Figure 5.2: (a) Resistivity of silver at low temperature. The upper curve is for a silver sample with a higher concentration of impurity atoms. (From W. J. de Haas, G. J. van den Berg and J. de Boer, *Physica* **2** (1935) 453).) (b) Resistance of sodium as a function of temperature. The upper curve is for a sodium sample with a higher concentration of imperfections/. (From D. K.C. MacDonald and K. Mendelssohn, *Proc. Roy. Soc. (London)* **A202** (1950) 103.

superconductors, for which ρ drops abruptly to zero at some critical temperature). Some data on Ag and Na are shown in Fig. 5.2 (from Serway, Moses and Moyer).

Looking back at (5.16), we see that the small resistivity of metals has two sources unexpected on the basis of classical physics: first, the typical velocities, v_F, are much larger than typical thermal velocities; second, the electron mean free path is surprisingly large, for reasons we will soon understand better. The same effects combine to make the thermal conductivity, κ, anomalously large. Referring to (5.6), the factor $v^2 c_v$ is roughly the same for a classical gas and a degenerate Fermi gas (at the same density, *etc.*). κ is large because τ is large. [The high thermal conductivity of metals (relative to insulators) is a familiar human experience: consider a sauna with metal seats or the prospect of licking a car door handle on a frigid winter day!]

How can we account for the unusually long mean free path of electrons in metals? Of course, the mean free path of particles in a *free* Fermi gas is infinite. But the electrons in metals are not free, they are moving in nearly periodic array of atomic potentials. If electrons behaved as classical particles they would scatter frequently and violently from the atomic lattice. It is easy to imagine that the phenomenon of "Pauli blocking," mentioned in § 1 in connection with the calculation of C_v, suppresses scattering: many states are not available to be scattered into because they are already occupied. This isn't enough, however, since the electrons near the top of the Fermi sea are not affected.

The true cause of the electron's long mean free path is the rigorous quantum mechanical result that *a particle doesn't scatter at all* from a perfectly regular array

of potentials. Its mean free path is infinite. The observed *finite* mean free path of electron's in metals has two sources: scattering from thermal fluctuations of the lattice (responsible for the $1/T$ dependence of ℓ) and from lattice imperfections (responsible for the irreducible resistivity as $T \to 0$, see Fig. 5.2).

In the next few subsections we'll derive and discuss the basic result. The more sophisticated analysis necessary to find the effects of impurities, imperfections and thermal fluctuations as the electron gas is more appropriately the subject of a course on many-body or condensed matter physics.

Before turning to the quantum mechanics of electrons in periodic potentials, let me summarize a few of the problems with the Drude–Sommerfeld free electron theory:

1. It fails to explain the astronomical differences in the electrical resistivity of different simple (crystalline) substances at (say) room temperature. *Insulators* such as quartz and sulphur have huge resistivity, frequently in excess of 10^{15} ohm-cm. Metals have resistivity measured in microhm-cm. Some notable examples are listed in Table 5.2.

2. It doesn't account for the temperature dependence of resistivity. Metals typically become better conductors at lower temperature. Certain poor metals such as silicon and germanium (semiconductors) become better conductors as the temperature is raised.

3. The *Hall effect* enables one to measure the sign of the mobile charges in conductors. While most substances contain *negative* mobile charges in agreement with Drude–Sommerfeld, some have *positive* mobile charges.

4. As already discussed, there is no explanation of the very long mean free path of electrons in good conductors.

[a] Purcell, *Electricity and Magnetism*.
[b] *CRC Handbook of Chemistry and Physics*
[c] $\parallel$ and $\perp$ denote directions with respect to symmetry axis.

Table 5.2: Resistivities

Material	Temperature	Resistivity (ohm-cm)
Sulphur	27°C	$1 \times 10^{17(a)}$
Quartz $(\|\|)^{(c)}$	17°C	$2 \times 10^{14(a)}$
Quartz $(\perp)^{(c)}$	17°C	$2 \times 10^{16(a)}$
Germanium	0°C	$200^{(b)}$
Germanium	227°C	$0.12^{(b)}$
Copper	0°	$1.56 \times 10^{-6(b)}$
Aluminum	20°	$2.828 \times 10^{-6(a)}$
Magnesium	20°	$4.6 \times 10^{-6(a)}$

5.2 ELECTRONS IN PERIODIC POTENTIALS — QUALITATIVE ARGUMENTS

Consider assembling a one-dimensional "crystal" by bringing together a collection of neutral atoms by decreasing the interatomic separation a. We know that the electrons move in orbitals within the separate atomic systems. If the system were classical, nothing much would happen until the classical turning points of electrons in adjacent atoms became interlaced at a separation $a_{\rm cl}$. Quantum mechanics allows richer possibilities — electrons may *tunnel* from atom to atom at separations considerably larger than $a_{\rm cl}$ — and this makes a major difference. To get an idea of the scale of the energies and separations involved, consider Fig. 5.3 where the potential seen by electrons in metallic sodium is shown. It would appear that tunneling is a significant effect for $3s$-electrons but much less important for $2s$-electrons.

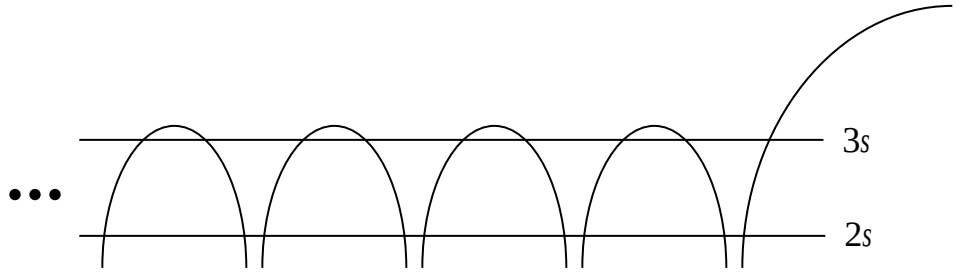


Figure 5.3: Schematic diagram of the potential energy curve seen by an electron in metallic sodium. Also shown are the *atomic* $2s$ and $3s$ levels of the sodium atom.

The energy levels of a widely separated collection of atoms are nearly the same as a single isolated atom. The *degeneracy* of each energy level, however, is vast. For simplicity, consider an "atomic" potential, shown in Fig. 5.4 with a single bound state of energy E_0. An electron moving in a string of such atoms separated by $a \gg d$ has a spectrum with a single energy level, E_0, but N-fold degeneracy. The N distinct wavefunctions of these N degenerate states are obtained by superposing the single "atom" wave functions with appropriate phases so the states are all orthogonal. The simplest case, $N = 2$, is drawn in Fig. 5.4b.

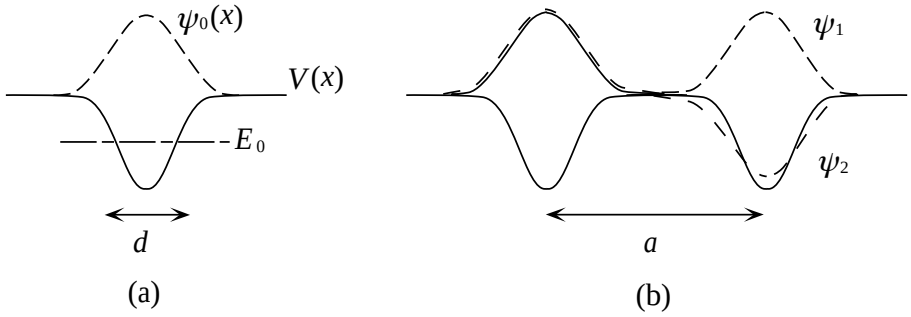


Figure 5.4: (a) A potential with a single bound state; (b) The ground state "bound" in the case of a double potential of the term of $V(x)$ separated by a.

When a decreases and tunneling between adjacent potentials can no longer be ignored, the degeneracy of the energy level, E_0, is lifted. To get an idea what happens consider $N = 2$. The splitting of levels in two well-separated potentials is a classic problem in elementary quantum mechanics: The wavefunctions within the wells change little in the presence of tunneling, so we may regard the states of the system as superpositions of the unperturbed single well "ground states." The symmetric superposition of the individual ground states (ψ_1) shifts slightly down in energy and becomes the true ground state (no nodes). The antisymmetric superposition (ψ_2) shifts slightly up and is an excited state. The generalization to $N > 2$ is more complex. Still the symmetric superposition (with no nodes) is the ground state and the most antisymmetric superposition (with alternating signs and a node in every forbidden region in between wells) is the most energetic state. All the other states ($N = 2$ of them) lie in a *band* between these two. Thus the effect of tunneling is to spread out an energy eigenvalue into a band of energy levels. The band is narrow when tunneling is weak, but it broadens when potential wells overlap more significantly or when one considers excited states within the wells. A schematic diagram of energy vs a/d is given in Fig. 5.5 (from Ashcroft and Mermin). The intervals between bands — energies at which no states are available — are known a *gaps*. As shown in Fig. 5.5 the higher-lying states spread into bands first (as a/d decreases) and bands developed out of different single well states can overlap.

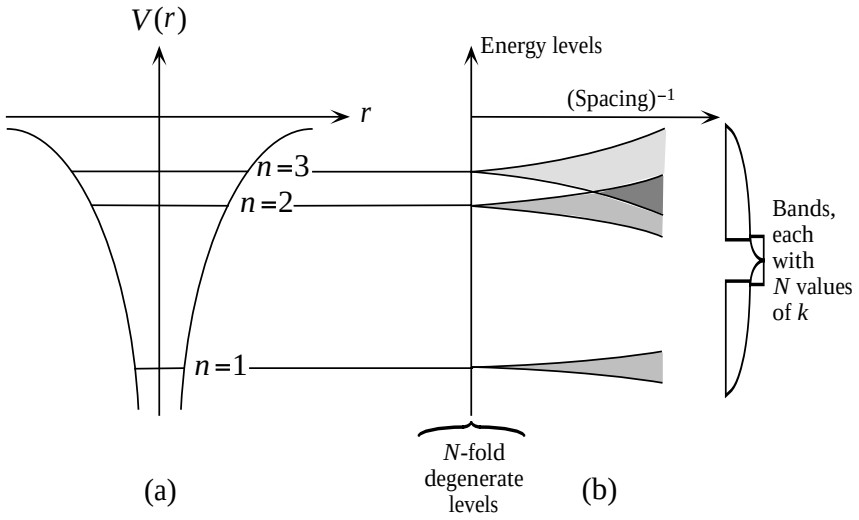


Figure 5.5: (a) Schematic representation of non-degenerate electroc levels in an atomic potential. (From Ashcroft and Mermin.) (b) The energy levels for N such atoms in a periodic array, plotted as a function of mean inverse interatomic spacing. When the atoms are far apart (small overlap integrals) the levels are nearly degenerate, but when the atoms are closer together (larger overlap integrals), the levels broaden into bands. (From Ashcroft and Mermin)

5.3 PERIODIC POTENTIALS IN ONE-DIMENSION I: BLOCH WAVES

Of course true crystals are three-dimensional and have unique properties because of their three-dimensionality. The problem of characterizing the motion of electrons in a periodic three-dimensional potential is not trivial and solving the appropriate Schrödinger equation is beyond analytic methods. Many (but by no means all) of the important features of the problem are preserved by limiting ourselves to motion in one-dimension where the characterization of the motion (Bloch's theorem and Brioullin–Zones) is quite straightforward and the analysis of the Schrödinger equation can be carried quite far analytically.

As a model for a one-dimensional crystal consider a periodic potential $V(x)$ with period a,

$$V(x+a) = V(x) \tag{5.19}$$

So the Hamiltonian for an electron moving in this potential,

$$H = \frac{p^2}{2m} + V(x) \tag{5.20}$$

is unchanged by the *finite translation*, $p \to p$, $x \to x + a$,

$$H(x) = H(x+a) \tag{5.21}$$

so this is a symmetry of the problem. This is similar to the situation in the absence of any potential. Then H is invariant under any translation. One consequence of this is that the momentum operator, p, is conserved and commutes with H

$$i\hbar\dot{p} = [p, H] = 0 \quad . \tag{5.22}$$

According to the usual rules of quantum theory, H and P can be simultaneously diagonalized and a basis of their eigenstates, labeled by their eigenvalues, E and p, can be used to describe the system.

We wish to mimic this procedure in the case where H is invariant only under finite translations. We define an operator, $T(a)$, which translates x by an amount a:

$$T(a)\psi(x) = \psi(x + a) \tag{5.23}$$

in coordinate representation ($\psi(x) \equiv \langle x|\psi\rangle$).[1] We can use the existence of the inner product

$$\langle\varphi|\psi\rangle \equiv \int_{-\infty}^{\infty} dx\, \varphi^*(x)\psi(x) \tag{5.24}$$

to prove $T(a)$ must be a unitary operator,

$$T(a)^\dagger T(a) = 1 \quad . \tag{5.25}$$

To prove this, write

$$\langle\varphi|\psi\rangle = \int_{-\infty}^{\infty} dy\, \varphi^*(y)\psi(y) \tag{5.26}$$

substitute $y = x + a$, use (5.23) and its Hermitian conjugate and change integration variable to x,

$$\langle\varphi|\psi\rangle = \int_{-\infty}^{\infty} dx\, \varphi^*(x)T^\dagger(a)T(a)\psi(x) \quad . \tag{5.27}$$

Since $|\varphi\rangle$ and $|\psi\rangle$ are arbitrary we are left with the operator equation (5.24).

If we now apply the same argument to arbitrary matrix elements of H, we shall prove

$$T^\dagger(a)HT(a) = H \quad . \tag{5.28}$$

First write

$$\langle\varphi|H|\psi\rangle = \int dy\, \varphi^*(y)H(y)\psi(y) \tag{5.29}$$

Follow the same procedure outlined after (5.8) to obtain

$$\langle\varphi|H|\psi\rangle = \int dx\, \varphi^*(x)\, T^\dagger(a)H(x + a)T(a)\psi(x) \quad , \tag{5.30}$$

[1]The precise form of $T(a)$ need not concern us here. In fact $T(a) = e^{iPa/\hbar}$ where $P \leftrightarrow \frac{1}{i\hbar}d/dx$ in coordinate representation.

which, combined with (5.21) yields the desired result. An immediate consequence of (5.25) and (5.28) is that $T(a)$ and H commute and therefore can be simultaneously diagonalized.

Let us try to characterize the eigenstates of $T(a)$:

$$T(a)\psi_\lambda(x) = \lambda(a)\psi_\lambda(x) \quad . \tag{5.31}$$

Since T is unitary we expect the eigenvalue $\lambda(a)$ to be a pure phase

$$\lambda(a) = e^{i\varphi(a)} \quad . \tag{5.32}$$

It's instructive to see how one proves this. Consider the inner product of eigenstates of $T(a)$:

$$\langle\lambda(a)|\lambda'(a)\rangle = \int dx\, \psi^*_\lambda(x)\psi_{\lambda'}(x) \tag{5.33}$$

We can use the translation properties of ψ_λ ((5.27) and (5.31)) to obtain

$$\langle\lambda(a)|\lambda'(a)\rangle = \sum_{n=-\infty}^{\infty} (\lambda^*\lambda')^n \int_0^a dx\, \psi^*_\lambda(x)\psi_{\lambda'}(x) \quad . \tag{5.34}$$

Define $\lambda \equiv e^{i\alpha}$, $\lambda' = e^{i\alpha'}$ where α and α' might, in principle, be complex; then

$$\sum_{n=-\infty}^{\infty} (\lambda^*\lambda')^n = \sum_{n=-\infty}^{\infty} e^{in(\alpha'-\alpha^*)} \quad . \tag{5.35}$$

If α or α' has an imaginary part then this sum is strictly *divergent* and the state $|\lambda(a)\rangle$ is not normalizable even in the continuum sense. So we conclude α and α' must be real, which is what we set out to prove.

Furthermore, when α and α' are real the sum in (5.35) can be expressed in terms of δ-functions

$$\sum_{n=-\infty}^{\infty} e^{inx} = 2\pi \sum_{N=-\infty}^{\infty} \delta(x + 2\pi N) \tag{5.36}$$

Let

$$\sum_{n=-\infty}^{\infty} e^{inx} \equiv f(x) \quad .$$

Clearly $f(x)$ is periodic with period 2π,

$$f(x) = f(x + 2\pi) \quad .$$

So its enough to consider the interval $-\pi \le x \le \pi$. Consider

$$\begin{aligned} f[g] &\equiv \int_{-\pi}^{\pi} dx\, f(x)g(x) \\ &= \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} dx\, e^{inx} g(x) \quad . \end{aligned}$$

But if we write a Fourier series representation for $g(x)$ in the interval $-\pi \leq x \leq \pi$,

$$g(x) = \sum_{n=-\infty}^{\infty} e^{-inx} g_n$$

then the Fourier inverse theorem,

$$g_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx\, e^{inx} g(x)$$

gives

$$f[g] = 2\pi g(0)$$

so

$$f(x) = 2\pi\delta(x)$$

for $-\pi < x < \pi$. Requiring $f(x)$ to have period 2π yields (5.36). □
For a proof see the box. Since α and α' appear only as phases; they are only defined modulo 2π, so (5.34) – (5.36) combine to give

$$\langle \lambda(a) | \lambda'(a) \rangle = 2\pi\delta(\alpha - \alpha') \int_0^a dx\, \psi_\lambda^*(x) \psi_{\lambda'}(x) \tag{5.37}$$

If we now set

$$\alpha \equiv ka \quad ,$$

change the wavefunction labels from λ to k and normalize $\psi_k(x)$ by

$$\int_0^a \frac{dx}{a} |\psi_k(x)|^2 = \frac{1}{2\pi} \quad , \tag{5.38}$$

then

$$\langle k | k' \rangle = \int_{-\infty}^{\infty} dx\, \psi_k^*(x) \psi_{k'}(x) = \delta(k - k') \tag{5.39}$$

which you will recognize as standard *continuum normalization*.

To summarize: we have learned that the eigenfunctions of the finite translation operator $T(a)$, which are also energy eigenstates, are labeled by a real quantum number, k, with

$$T(a)\psi_k(x) = \psi_k(x + a) = e^{ika} \psi_k(x) \quad , \tag{5.40}$$

and

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a} \quad .$$

The states $\{\psi_k(x)\}$ may be thought of as modulated plane waves with the periodic potential doing the modulating. In particular, the $\{\psi_k(x)\}$ obey the continuum normalization condition, (5.39). Equation (5.40) is known as Bloch's theorem and these

states are known as *Bloch waves*. Note that although H is strictly periodic with period a, the energy eigenstates are periodic *only up to a plase*. To make this explicit, define

$$\psi_k(x) \equiv e^{ikx}\varphi_k(x) \tag{5.41}$$

Then from (5.40), φ_k is strictly periodic,

$$\varphi_k(x+a) = \varphi_k(x) \quad . \tag{5.42}$$

Notice that the wave number, k, is only defined modulo 2π. This leads to a certain ambiguity in the kinematics of motion in a periodic potential, which we'll touch upon later. For now, remember that the state labeled by $k + 2\pi N/a$ (for any integer N) is the same state as labeled by k. Note also that k is not a momentum. Indeed, the state $\psi_k(x)$ is not an eigenstate of momentum but only an eigenstate of the finite translation $T(a)$. Nevertheless, k often appears in formulas in a fashion which resembles the free space momentum and is often referred to as the Bloch momentum.

5.4 PERIODIC POTENTIAL IN ONE–DIMENSION II: ELEMENTARY FLOQUET THEORY

Let's examine the Schrödinger equation for a periodic potential

$$-\frac{\hbar^2}{2m}\psi''(x) + V(x)\psi(x) = E\psi(x) \tag{5.43}$$

Our aim is to construct the eigenstates of $T(a)$ and H. Because $V(x)$ is periodic, $\psi(x+a)$ is also a solution to (5.43). Of course, $\psi(x+a)$ obeys

$$-\frac{\hbar^2}{2m}\psi''(x+a) + V(x+a)\psi(x+a) = E\psi(x+a) \tag{5.44}$$

But only if $V(x+a) = V(x)$ can we also write

$$-\frac{\hbar^2}{2m}\psi''(x+a) + V(x)\psi(x+a) = E\psi(x+a) \quad . \tag{5.45}$$

Comparing (5.45) and (5.43) we see that $\psi(x+a)$ obeys the *same* differential equation as $\psi(x)$. Now (5.43) is a second order differential equation and therefore has *two* linearly independent solutions. Let $\psi_1(x)$ and $\psi_2(x)$ be some (arbitrary) choice of

those solutions. Since $\psi_1(x+a)$ obeys the same differential equation (5.45) it must be a linear combination of $\psi_1(x)$ and $\psi_2(x)$:

$$\psi_1(x+a) = A_{11}\psi_1(x) + A_{21}\psi_2(x) \tag{5.46}$$

and likewise for $\psi_2(x+a)$,

$$\psi_2(x+a) = A_{12}\psi_1(x) + A_{22}\psi_2(x) \tag{5.47}$$

The four coefficients A_{ij} can be grouped into a matrix, the *transfer matrix* $\mathbf{A}$.[2] If the basis states $\psi_j(x)$ are real then $\mathbf{A}$ is real.

To find the eigenstates of $T(a)$, we seek linear combinations of ψ_1 and ψ_2 which render the transfer matrix diagonal. Let

$$\phi(x) = \sum_{j=1}^{2} c_j \psi_j(x) \qquad j = 1, 2 \tag{5.48}$$

with the coefficients c_j chosen so that

$$\phi(x+a) = \lambda\phi(x) \tag{5.49}$$

We expect two eigenfunctions $\{\phi_j\}$ and associated eigenvalues $\{\lambda_j\}$, with $j = 1, 2$. In a familiar language: the $\{\phi_j\}$ and the $\{\lambda_j\}$ are eigenfunctions and eigenvalues of the transfer matrix. We know it is possible to find the $\{\phi_j\}$ because we proved $T(a)$ and H commute, which assures that we can diagonalize $T(a)$ among the eigenstates of H.

If we combine (5.46) – (5.49) we obtain an eigenvalue equation

$$\begin{aligned} A_{11}c_1 + A_{12}c_2 &= \lambda c_1 \\ A_{21}c_1 + A_{22}c_2 &= \lambda c_2 \end{aligned} \tag{5.50}$$

or, in matrix notation,

$$\mathbf{A}\vec{c} = \lambda\vec{c} \ . \tag{5.51}$$

The condition for a solution,

$$\det|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{5.52}$$

is particularly simple for a 2×2 matrix,

$$\lambda^2 - \lambda \operatorname{Tr}\mathbf{A} + \det\mathbf{A} = 0 \tag{5.53}$$

where $\operatorname{Tr}\mathbf{A} = A_{11} + A_{22}$ and $\det\mathbf{A} = A_{11}A_{22} - A_{12}A_{21}$. $\operatorname{Tr}\mathbf{A}$ and $\det\mathbf{A}$ are related to λ_1 and λ_2 by $\operatorname{Tr}\mathbf{A} = \lambda_1 + \lambda_2$ and $\det\mathbf{A} = \lambda_1\lambda_2$.

[2] $\mathbf{A}$ is just the (2×2) matrix representation of the operator $T(a)$ restricted to the subspace of H with eigenvalue E.

Since **A** is real (in a basis where ψ_1 and ψ_2 are real), (5.53) is a quadratic equation with real coefficients, hence the two roots (λ_1 and λ_2) are either

$$\begin{array}{lll} 1. & \text{Both real:} & \lambda_1 = \lambda_1^* \ , \quad \lambda_2 = \lambda_2^* \\ 2. & \text{Complex conjugate:} & \lambda_1 = \lambda_2^* \ . \end{array} \tag{5.54}$$

Furthermore, we can show

$$\lambda_1 \lambda_2 = 1 \tag{5.55}$$

as follows: Consider *any* two linearly independent functions ϕ_1 and ϕ_2 obeying (5.43), then

$$\frac{d}{dx}\left(\phi_1 \phi_2' - \phi_2 \phi_1'\right) = 0 \tag{5.56}$$

so

$$\omega(x) \equiv \phi_1 \phi_2' - \phi_2 \phi_1' = c \neq 0 \tag{5.57}$$

the first result, (5.56), follows directly from the Schrödinger equation. The second, (5.57), follows upon integration and the fact that ϕ_1 and ϕ_2 are linearly independent (which requires $c \neq 0$). Since $w(x)$ is a constant

$$\omega(x + a) = \omega(x) \tag{5.58}$$

but if we take ϕ_1 and ϕ_2 to be eigenfunctions of $T(a)$ with eigenvalues λ_1 and λ_2, respectively, then

$$w(x + a) = \lambda_1 \lambda_2 w(x) \tag{5.59}$$

Comparing these two equations,

$$\lambda_1 \lambda_2 = \det \mathbf{A} = 1$$

follows. Equation (5.53) now reads

$$\lambda^2 - \lambda \operatorname{Tr} \mathbf{A} + 1 = 0 \tag{5.60}$$

Now the alternatives of (5.54) are clearer

$$\begin{array}{lll} 1. & \text{Both real and reciprocal:} & \lambda_2 = 1/\lambda_1 \ , \\ 2. & \text{Complex conjugate and unitary:} & \lambda_1 = \lambda_2^* = e^{i\alpha} \ . \end{array}$$

As we established in §5.b, only case 2 corresponds to possible (normalizable) physical states. Consider the modulus of the *sum* of the eigenvalues in the two cases. In both cases $\lambda_1 + \lambda_2$ is real (a theorem on the solutions to the quadratic equation (5.53))

$$\begin{array}{lll} 1. & \text{Case 1:} & |\lambda_1 + \lambda_2| = |\lambda_1 + 1/\lambda_1| \geq 2 \\ 2. & \text{Case 2:} & |\lambda_1 + \lambda_2| = 2\,|\cos\alpha| \leq 2 \end{array}$$

and referring back to (5.53) and using $\lambda_1+\lambda_2 = \mathrm{Tr}\,\mathbf{A}$, the condition for a normalizable state becomes

$$|\,\mathrm{Tr}\,\mathbf{A}| = |A_{11} + A_{22}| \leq 2 \tag{5.61}$$

This is our fundamental result and it deserves some discussion.

First, note that Eq. (5.61) is independent of the original choice of states, ψ_1 and ψ_2, way back to (5.46). The reason is that a change of basis, $\psi \to \tilde{\psi} = \mathbf{S}\psi$ with $\mathbf{S}^\dagger\mathbf{S} = \mathbf{1}$ changes $\mathbf{A}$ by a similarity transform, $\mathbf{A} \to \mathbf{S}\,\mathbf{A}\,\mathbf{S}^\dagger$, and

$$\mathrm{Tr}\,\mathbf{S}\,\mathbf{A}\,\mathbf{S}^\dagger = \mathrm{Tr}\,\mathbf{S}^\dagger\mathbf{S}\,\mathbf{A} = \mathrm{Tr}\,\mathbf{A}$$

For some fixed potential, $V(x)$, the transfer matrix $\mathbf{A}$ and its eigenvalues λ_1 and λ_2 depend parametrically on the energy. Thus, as a function of E, λ_1 and λ_2 execute trajectories in the complex plane. When they lie on the real axis (at reciprocal points) no normalizable energy eigenstates exist. These are the "gaps" mentioned in § 5.b. When they lie on the unit circle, continuum normalized energy eigenstates exist. These are the "bands" mentioned in § 5.b. Notice that the edge of a band occurs when $\lambda_1 = \lambda_2 = \pm 1$, where according to (5.49), the wave function is exactly periodic ($\lambda = 1$) or antiperiodic ($\lambda = -1$).

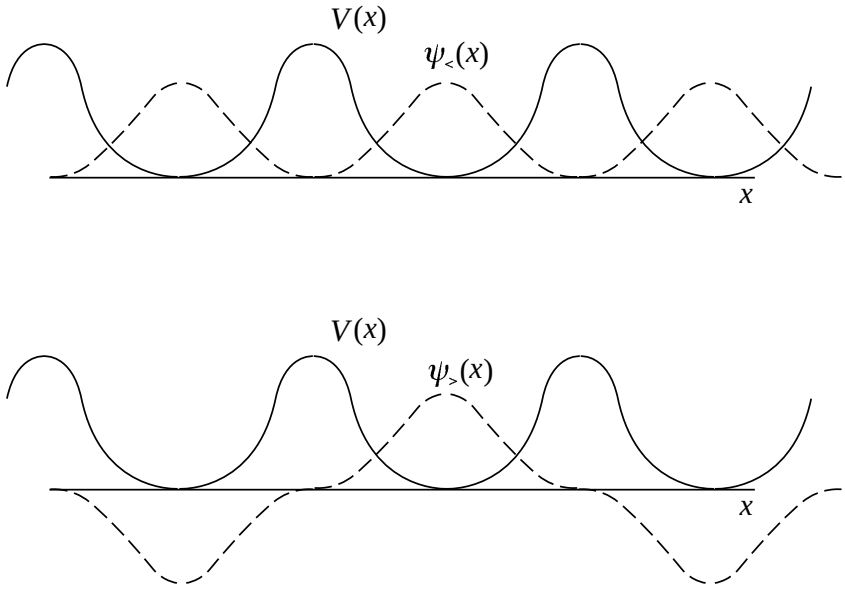


Figure 5.6: The lowest (exactly periodic) and highest (exactly antiperiodic) members of the lowest band in a periodic potential $V(x)$.

Let's combine the analytic analysis we've just completed with the pictorial analysis from the first part of this section to assemble a description of the spectrum. Imagine,

for the sake of illustration, a periodic potential consisting of well-separated wells separated by high barriers (Fig. 5.6). For infinite barrier height the ground state consists of the ground state in each well (ϕ_0) *superposed with arbitrary phases*. The degeneracy parameterized by these phases is broken by tunneling between wells. A famous theorem of quantum mechanics in one dimension tells us that the ground state has no nodes, so the lowest energy eigenstate denoted $\psi_<$ with energy $E_<$ and shown in Fig. 5.6 is constructed by superposing separate-well ground states all with the same relative phase. This solution is strictly periodic so we have learned that $\lambda(E) = +1$ for the lowest allowed energy. The other states in this "ground state band" are obtained by superposing separate-well ground states with various phases until a band edge is reached where we expect a periodic or antiperiodic solution. In this (narrow) band, no wavefunction can have a node in a classically allowed region — corresponding to an excited separate-well state ψ_1 — because this would cost far more energy than placing the node in the forbidden region. A little thought leads to the conclusion that it is not possible to construct another *periodic* wavefunction in this band, but an *anti-periodic* state is easy to construct — by putting a node in each forbidden region. This state, denoted $\psi_>$ with energy $E_>$ is also shown in Fig. 5.6. From this argument it is clear that the eigenvalue $\lambda(E)$ reaches $\lambda(E_<) = 1$ at the bottom of the ground state band, moves along half the unit circle and reaches $\lambda(E_>) = -1$ at the top of the ground state band. This trajectory is shown in Fig. 5.7. [The other eigenvalue traces the reciprocal trajectory.] If we parameterize λ by the "Bloch momentum," k

$$\lambda \equiv e^{ika}$$

we see that ka ranges from 0 to $\pm\pi$ for the two eigenvalues λ and λ^* as E moves through the ground state band. This definition enables us to plot a "dispersion relation," $E(k)$, for the ground state, as in Fig. 5.8.

Notice that we have drawn $E(k)$ as though it varies quadratically with k near the edge of a band. This can be proved by expanding (5.60) about the bound edge: First consider $E = E_< + \Delta E$ and expand $\mathrm{Tr}\,\mathbf{A}$ about its value (2) at $E = E_<$, $\mathrm{Tr}\,\mathbf{A} \cong 2(1 - \sigma\Delta E)$ with $\sigma > 0$ for $0 < \Delta E \ll (E_> - E_<)$. Then the roots of (5.60) are

$$\lambda_\pm = 1 \pm i\sqrt{2\sigma\Delta E} \cong 1 \pm ika$$

so

$$E \cong E_< + \frac{k^2a^2}{2\sigma} \tag{5.62}$$

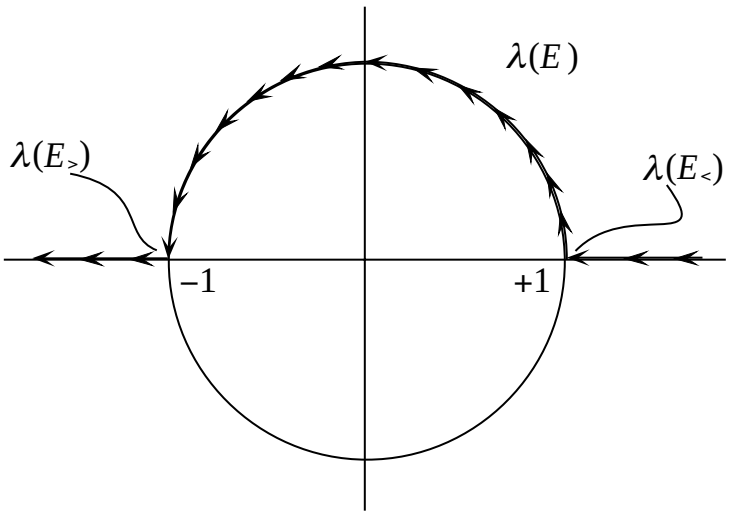


Figure 5.7: The trajectory in the complex plane of one of the eigenvalues, λ, of the transfer matrix as a function of energy. When λ is real, E lies in a gap. When λ is complex, with $|\lambda| = 1$, E lies in a band.

The quantity

$$m^* = \frac{\hbar^2 \sigma}{a^2}$$

where

$$\sigma = -\frac{1}{2}\,\frac{d}{dE}\,\mathrm{Tr}\,\mathbf{A}\Big|_{E=E_<} > 0$$

plays the role of an "effective mass" for electrons near the bottom of a band. As k ranges through the band, the relationship of E to k changes until, as k approaches $\pm\pi/a$, it is once again quadratic but with the opposite sign!

$$E = E_> - \frac{(|ka| - \pi)^2}{2\tau}$$

where

$$\tau = \frac{1}{2}\,\frac{d}{dE}\,\mathrm{Tr}\,\mathbf{A}\Big|_{E=E_>} > 0$$

so electrons have negative effective mass near the top of the band. This is only one aspect of the unusual behavior of electrons in a nearly filled band which we will discuss in some detail in §5.f.

Before leaving this semiquantitative discussion we will extend it to the case of the second band. We left off having found $\psi_>$ at the top of the lowest band. In the approximation of high barriers between attractive wells we expect the next allowed state to be constructed by piecing together wavefunctions corresponding to the first excited state in each well. Since these wavefunctions, ψ_1, have a node within the well, the lowest energy state we can make (least nodes) is *anti-symmetric* (see Figure 5.9).

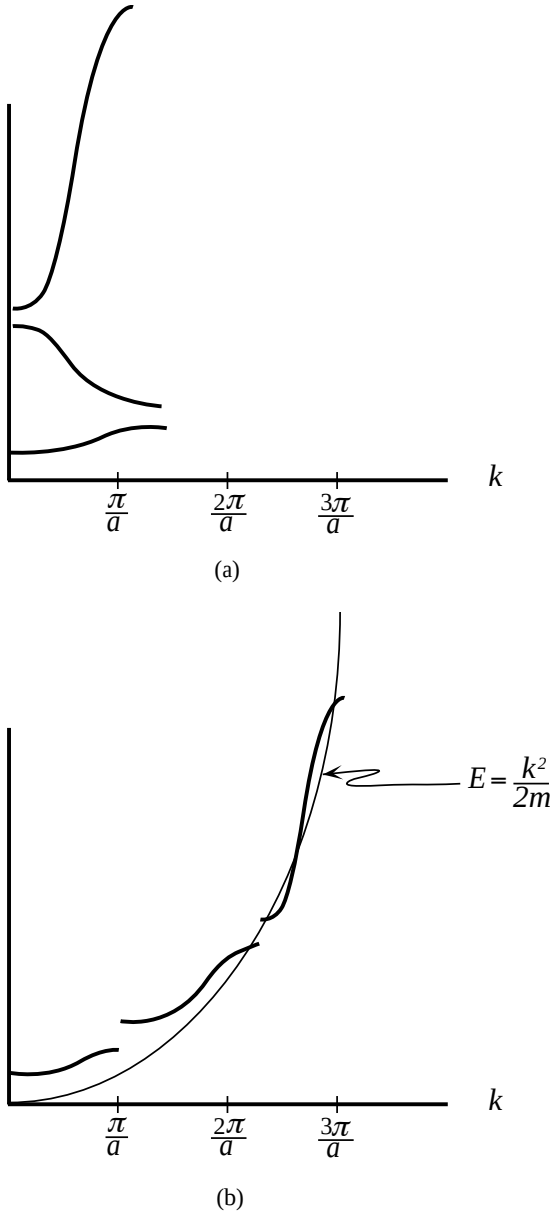


Figure 5.8: Energy as a function of wave number in a series of bands. Adding factors of π/a we see this function approximates the continuum relation $E = \hbar^2 k^2/2m$.

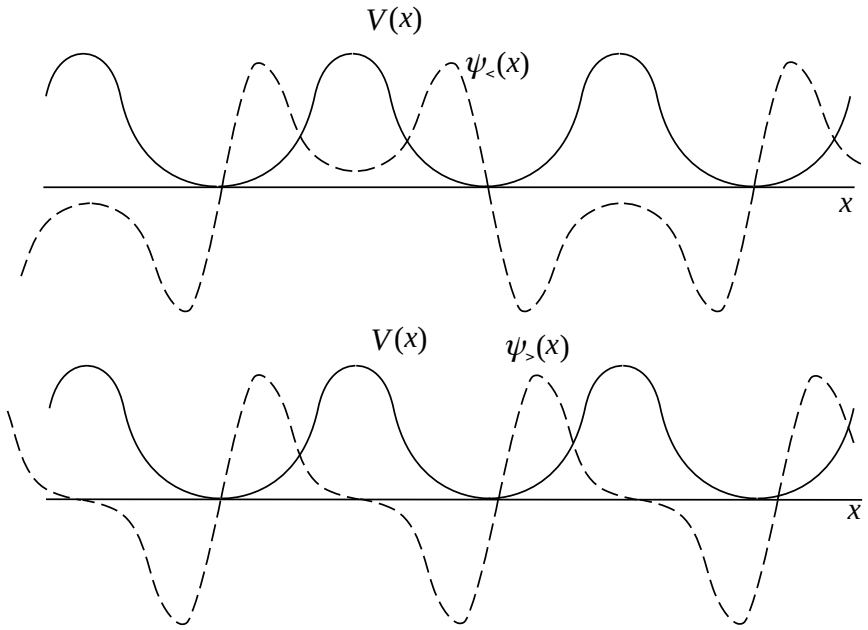


Figure 5.9: The lowest (exactly antisymmetric) and highest (exactly symmetric) states in the second band.

So the band begins with $\lambda_1 = \lambda_2 = -1$. Higher energy states are obtained by superposing ψ_1 in each well with various phases until we reach the upper band edge where the wavefunction has a node in the forbidden region between each well and is once again periodic, so $\lambda_1 = \lambda_2 = 1$. The eigenvalues execute the trajectory shown in Fig. 5.7 in the opposite direction. We'd like to associate a Bloch momentum, k, with the states in this second band. Here we are faced with an ambiguity mentioned earlier. We can always choose k to lie in the interval $-\pi/a < k < \pi/a$, known as the first Brillouin Zone. The resulting dispersion relation, $E(k)$, is shown in Fig. 5.8a. On the other hand, we can add appropriate factors of $2\pi/a$ and make the curve $E(k)$ look single valued, with a jump at $k = \pm\pi/a$, between bands (see Fig. 5.8b). The former is known as the *reduced zone scheme*, the latter is the *extended zone scheme*. Both are useful. Needless to say, the situation is considerable more complex in three dimensions.

This analysis continues in the same spirit for higher bands. The primary modification is that the effects of the barriers between wells become less important at higher energy: the bands become broader in $E(k)$ and the gaps become narrower. Eventually, at high energy, the gaps become negligible and the function $E(k)$ is well-approximated by $k^2/2m$ within the bands. This behavior is sketched in Fig. 5.8.

To make this discussion more concrete we next turn to the analysis of a specific example.

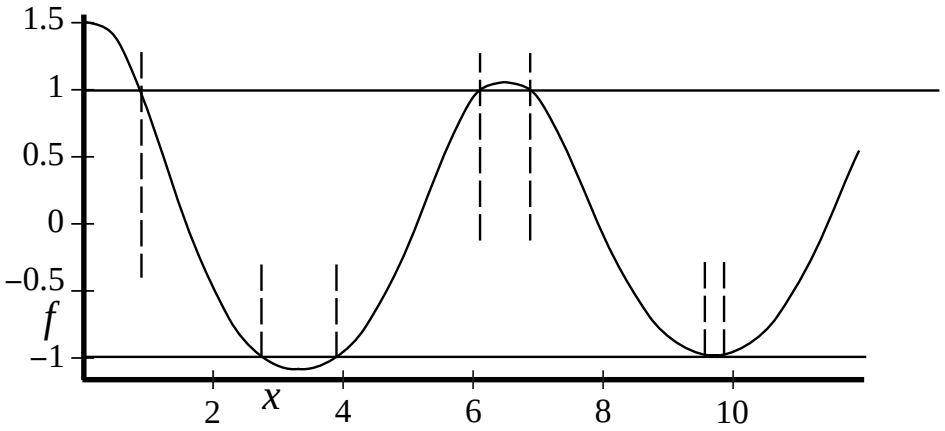


Figure 5.10: Graphical construction for the bands in the Kronig-Penney Model.

5.5 AN EXAMPLE: THE KRONIG–PENNEY MODEL

To illustrate the general method developed in the previous two sections we now turn to an explicit model first devised by Kronig and Penney. We let $V(x)$ be a periodic δ-function:

$$V(x) = V_0 \sum_n \delta(x - na) \tag{5.63}$$

We calculate the transfer matrix $\mathbf{A}$ directly. For $x \neq na$ the solutions to the Schrödinger equation are $\sin qx$ or $\cos qx$ with $q = \sqrt{2mE/\hbar^2}$. At $x = na$ the δ-function interaction requires a discontinuity in ψ':

$$\Delta\psi' \equiv \psi'\Big|_{x=na+\epsilon} - \psi'\Big|_{x=na-\epsilon} = \frac{2mV_0}{\hbar^2}\psi \equiv \frac{1}{d}\psi \tag{5.64}$$

For convenience we choose independent solutions

$$\begin{aligned} \psi_1(x) &= \sin qx \\ \psi_2(x) &= \cos qx \end{aligned} \tag{5.65}$$

in the interval $-a < x < 0$. To calculate $\mathbf{A}$, we make use of the defining relation (5.47)

$$\psi_i(x + a) = A_{ji}\psi_j(x)$$

Since $-a < x < 0$, $0 < x + a < a$, thus we must determine $\psi_i(x + a)$ with the aid of the jump condition (5.64)

$$\begin{aligned} \Delta\psi_1'(0) &= \frac{1}{d}\psi_1(0) = 0 \\ \Delta\psi_2'(0) &= \frac{1}{d}\psi_2(0) = \frac{1}{d} \end{aligned} \tag{5.66}$$

so

$$\begin{aligned} \psi_1(x) &= \sin qx & 0 < x < a \\ \psi_2(x) &= \cos qx + \tfrac{1}{qd}\sin qx & 0 < x < a \end{aligned} \tag{5.67}$$

Now for $0 < x + a < a$,

$$\begin{aligned} \psi_1(x+a) &= \sin q(x+a) = \psi_1(x)\cos qa + \psi_2(x)\sin qa \\ \psi_2(x+a) &= \cos q(x+a) + \frac{1}{qa}\sin q(x+a) \\ &= \left(\frac{1}{qd}\cos qa - \sin qa\right)\psi_1(x) + \left(\cos qa + \frac{1}{qd}\sin qa\right)\psi_2(x) \end{aligned} \tag{5.68}$$

from which we read off

$$\mathbf{A} = \begin{pmatrix} \cos qa & \sin qa \\ \frac{\cos qa}{qd} - \sin qa & \cos qa + \frac{\sin qa}{qd} \end{pmatrix} \tag{5.69}$$

We confirm $\det \mathbf{A} = 1$ and identify

$$\operatorname{tr} \mathbf{A} = 2\cos qa + \frac{\sin qa}{qd} \tag{5.70}$$

so the condition for bands of allowed states is

$$-1 \le \cos qa + \frac{\sin qa}{2qd} \le 1 \tag{5.71}$$

This relation is graphed in Fig. 5.10 for the case of $d > 0$ (repulsive δ-function interaction). Note that the gaps quickly become small as qa increases (relative to qd). The further exploration of this model (determining energies of band edges, effective masses, *etc.* is left to the problems.

5.6 METALS, INSULATORS AND SEMICONDUCTORS

With Bloch's apparatus at our disposal we can return to the qualitative description of crystalline materials and resolve some of the problems of the Drude–Sommerfeld theory summarized in § 5.a.

First consider the problem of mean free path. Bloch waves are stationary states with non-zero average velocity. The velocity operator, defined by $i\hbar\dot{x} = [x, H]$, is not

other than p/m with p represented by $-i\hbar\, d/dx$ in coordinate space. It's expectation value in a Bloch wave state, $\psi_k(x)$, is

$$\langle \dot{x} \rangle = \int dx\, \psi_k^*(x) \left(-\frac{i\hbar}{m}\, \frac{d}{dx} \right) \psi_k(x) \Big/ \int dx\, \psi_k^*(x) \psi_k(x) \quad . \tag{5.72}$$

Using (5.37) and (5.39)

$$\langle \dot{x} \rangle = \frac{\hbar k}{m} + \int_0^a dx\, \varphi_k^*(x) \left(-\frac{i\hbar}{m}\, \frac{d}{dx} \varphi_k(x) \right) \Big/ \int_0^a dx\, \varphi_k^*(x) \varphi_k(x) \tag{5.73}$$

But $\varphi_k(x)$ can be chosen to be real (it's a solution to a real differential equation subject to a real boundary condition), and periodic (5.40)), so the second term in (5.73) vanishes. Thus Bloch waves can be characterized by a constant velocity, $\hbar k/m$, which does not dissipate as the electron propagates through the crystal. Though this may at first see surprising, it is the same effect which allows light to propagate without "resistance" through a refractive medium: the coherent superposition of scattered waves from a regular array of scatters results in unattenuated propagation in the forward direction.

From this point of view, the puzzle becomes why conductors have any resistance at all! The answer is two-fold (justification lies beyond this course): first, the lattice ions in a crystal are not fixed. They vibrate. The vibrations can be viewed as sound waves and quantized as *phonons*, in direct analogy to photons. A crystal lattice generates a (Bose–Einstein) gas of phonons whose density increases with temperature. Electron-photon collisions give rise to a temperature-dependent resistivity in otherwise excellent conductors. Second, no lattice is perfect. Irregularities and impurities — defects in general — scatter electrons and generate a resistivity which fails to vanish even at $T = 0$.

Next, consider how the properties of a material depend on where the Fermi energy falls in relation to the band structure. A given band will have a capacity to hold a certain number of electrons depending on the multiplicity of the original single potential level and the symmetry of the crystal. It is never less than $2N$ whose N is the number of atoms in the crystal and the factor 2 is the electron spin degeneracy. For the sake of simplicity let's consider a band originating in the $3s$ atomic orbital. In metallic sodium this band is half-full: the Fermi energy (defined in the usual way, as the energy of the last filled level) lies in the middle of the band (see Fig. 5.11). The dynamics of electrons near the Fermi surface resembles electrons in a free Fermi gas. In particular, there are plenty of states available just above the Fermi surface. An applied electric field shifts electrons from (say) negative k to positive k (in one dimensions) and generates an electromagnetic current. This is a metal and it is

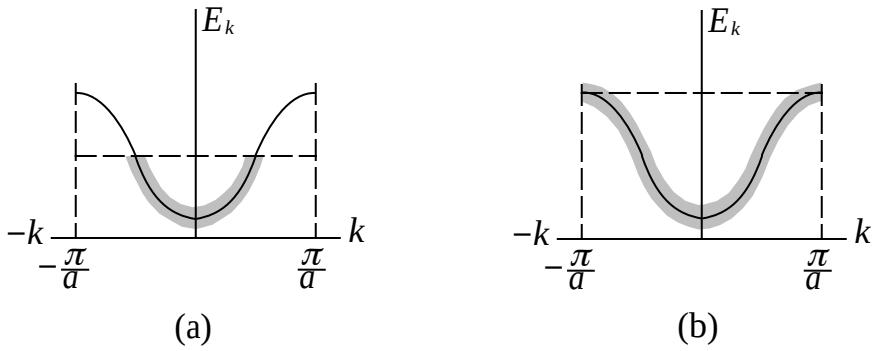


Figure 5.11: Filling of bands by electrons. A solid with one electron per atom has a half-filled band as indicated by the cross-hatching in (a). This material is a metal. If there are two electrons per atom the band is full, as shown in (b), and the substance is an insulator. (From Brehm and Mullin)

well-approximated by the Drude–Sommerfeld theory. From this argument one can generalize that any element with an *odd* number of electrons (and, consequently, E_F falling within a band) will be a conductor.

In contrast suppose we place $2N$ electrons in the $3s$ band. [One might think this applies to magnesium, however that case is more complex and is discussed immediately below.] Now and band is filled (see Fig. 5.11) and insert. There are no nearby levels to populate when an external electric field is turned on so no current develops. The material is an insulator. In fact, this point isn't quite as obvious as it's often portrayed. In the presence of an external field the exact periodicity of the potential is lost so the analysis becomes more complex. A more rigorous argument based on the Liouville theorem is given by Ashcroft and Mermin.

The actual case of magnesium is an interesting one. There are indeed two $3s$ electrons per atom, but a higher band overlaps the $3s$ band in magnesium, so both are partially filled as shown in Fig. 12. So magnesium is a conductor. [Fig. 5.12 is much oversimplified since the three-dimensional momentum space of the periodic magnesium crystal is quite complex and certainly not one-dimensional.]

Next consider what happens if an insulator happens to have an empty band lying only a short interval above the Fermi surface. Say the gap between the bands is Δ. At $T = 0$, the upper band (call it the "conduction" band) is empty and the lower one ("valence" band) is filled. At $T = 0$ the material is an insulator. As the temperature is increased thermal effects diffuse the top of the Fermi sea, according to the Fermi distribution function (1.28). There is a non-vanishing probability to find electrons in the conduction band and "holes" (*i.e.* unoccupied levels) in the lower band (see Fig. 5.13). The electrons give the material a small but significant conductivity (which

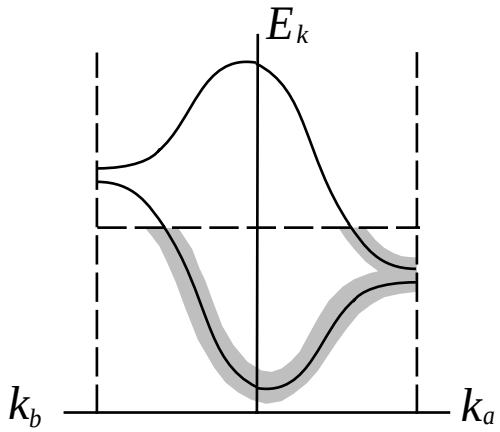


Figure 5.12: Example of overlapping bands in which a substance with two electrons per atom turns out to be a metal instead of an insulator. Generally, k_a and k_b are two distinct directions in three-dimensional k space rather than just $+k$ and $-k$.

increases with temperature). So, in fact, do the holes. An empty state in an otherwise filled band behaves kinematically and electromagnetically like an electron of *positive* charge. It has positive inertia because, if you remember, we showed that levels near the top of a band have negative effective mass. It has positive charge because the absence of an electron moves left when an electric field pulls the electrons to the right. Such a material is known as an *intrinsic semiconductor* — silicon and germanium are good examples.

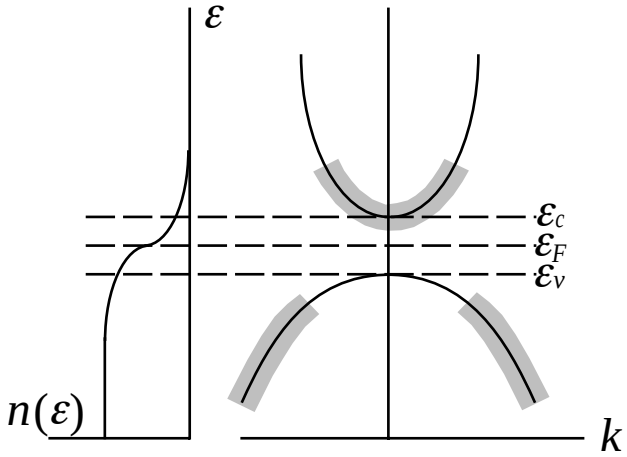


Figure 5.13: Valence and conduction bands in an intrinsic semiconductor. A few electrons are thermally excited out of the valence band. Also shown in the distribution function $n(\epsilon)$. Note that the Fermi energy falls in the band gap.

Another way to obtain a material with semiconducting properties is through "doping" — adding small amounts of impurities to a pure material, typically an intrinsic

semiconductor. The classic example is a crystal of silicon with a small contamination of phosphorus. Phosphorus has an additional valence electron compared to silicon. At zero temperature those valence electrons stay bound to the phosphorus impurities — filling discrete levels in the gap between the filled and empty bands of silicon. The extra electron lies just below the conduction band (in phosphorus, 0.05 eV) so relatively weak fields are sufficient to promote it into the conduction band and produce a current. For obvious reasons phosphorus is called a donor and phosphorus doped silicon is an "n-type" (n for negative) semiconductor. The situation is reversed for aluminum which lacks an electron relative to silicon: electrons in the lower silicon band easily become trapped on aluminum impurities producing holes and electric conduction. Aluminum is an "acceptor" and aluminum (or gallium) doped silicon is a "p-type" semiconductor (p for positive).

Note that the mobile charge carriers in semiconductors may have either sign of electric charge and account for the unusual Hall effect measurements mentioned in § 5.b.

Having now "explained" the four defects of the Drude–Sommerfeld model, it is time to end this lightning survey of solid state physics.