CHAPTER III. MAXWELL'S EQUATIONS AND SPECIAL RELATIVITY

3.1 The Electromagnetic Field Equations

(Reference: Frank "Electricity and Optics", Chap. VIII.)

Thus far we have treated the electric field and the magnetic field as though they were completely independent. This is not true in general because of Faraday's Induction Law, and Maxwell's introduction of displacement current. Faraday showed that if the magnetic field is changing with the time, the electromotive force around a closed path is not zero, but is equal to the time rate of change of the flux of B linked by the path. That is

$$\oint E \cdot ds = -\frac{\partial \Phi}{\partial t} = -\iint \frac{\partial B}{\partial t} \cdot dS \qquad (3.1)$$

In order that the equation of continuity of current be satisfied, Maxwell introduced the idea of displacement current, and showed that the displacement current density is just equal to $\frac{\partial D}{\partial t}$. In stating Ampere's circuital theorem, we must take account of this current density as well as the real current density, J. Hence, Ampere's circuital law becomes,

$$\oint H \cdot ds = \iint \left(J + \frac{\partial D}{\partial t}\right) \cdot dS \qquad (3.2)$$

These equations (3.1) and (3.2) together with

$$\iint_{\text{closed}} D \cdot dS = \rho$$

$$\iint_{\text{closed}} B \cdot dS = 0$$

are the integral equations of the electromagnetic field.

Maxwell deduced from them the equivalent differential equations which are generally called Maxwell's Equations, and are

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu \frac{\partial H_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu \frac{\partial H_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu \frac{\partial H_z}{\partial t} \end{aligned}\right\} \qquad (3.3)$$

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= j_x + \varepsilon \frac{\partial E_x}{\partial t} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= j_y + \varepsilon \frac{\partial E_y}{\partial t} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= j_z + \varepsilon \frac{\partial E_z}{\partial t} \end{aligned}\right\} \qquad (3.4)$$

$$\varepsilon \frac{\partial E_x}{\partial x} + \varepsilon \frac{\partial E_y}{\partial y} + \varepsilon \frac{\partial E_z}{\partial z} = \rho \qquad (3.5)$$

$$\mu \frac{\partial H_x}{\partial x} + \mu \frac{\partial H_y}{\partial y} + \mu \frac{\partial H_z}{\partial z} = 0 \qquad (3.6)$$

In free space, $\mu = \mu_0$, $\varepsilon = \varepsilon_0$ and with no free charges or real currents, these equations become in vector notation,

$$\left.\begin{aligned} \text{curl } E &= -\mu_0 \frac{\partial H}{\partial t} \\ \text{curl } H &= \varepsilon_0 \frac{\partial E}{\partial t} \\ \text{div } E &= 0 \\ \text{div } H &= 0 \end{aligned}\right. \qquad (3.7)$$

By taking the curl of the first equation and the time derivation of the second we have

$$- \text{curl}(\text{curl } E) = \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2}$$

but, $- \text{curl}(\text{curl } E) \equiv - \text{grad}(\text{div } E) + \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2}$

so
$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} \qquad (3.8)$$

since div $E = 0$

Now eq. (3.8) is just the wave equation for the vector E, and the wave is propagated with a velocity equal to $1/\sqrt{\mu_0 \varepsilon_0}$ which we know to be c. An identical wave equation can be obtained for H. Maxwell's work was of the greatest importance since it showed the inherent electromagnetic nature of light, and also predicted the possibility of the radiation of radio waves.

3.2 Moving Coordinate Systems

If we have two coordinate systems, (x,y,z) and (x^*,y^*,z^*) and if the starred coordinate system is moving relative to (x,y,z) at a constant velocity v in the direction of the positive x-axis, then we should certainly expect to be able to express the coordinates of a point (x,y,z) in terms of the coordinates (x^*,Y^*,z^*). If we assume that the two systems coincide at time $t = 0$ then at some later time, we have

$$\begin{aligned} x^* &= x - vt \\ y^* &= y \\ z^* &= z \end{aligned} \qquad (3.9)$$

Such a change of coordinates is called a Galilean transformation, and it is apparent that Newton's laws of motion for a mass point will be the same in either system since there is no relative acceleration. However, if we inquire how Maxwell's Field Equations behave under such a coordinate transformation we find very quickly that their form changes radically. A simple physical example serves to illustrare this point. Suppose we have a uniform magnetic field B at rest in (x,y,z) and a charged particle at rest in (x*,y*,z*). In the unstarred system we would measure a magnetic force since the particle is moving, i.e., $F = qv \times B$. In the starred system the particle is not moving and hence no magnetic force is possible.

It developes that the form of Maxwell's equations and the general electromagnetic force equation (the ponderomotive equation)

$$F = q\left[E + v \times B\right] \qquad (3.10)$$

remains unchanged under quite a different coordinate transformation called the Lorentz transformation. (However, the form of the field vectors must be changed in order for this invariance to hold).

3.3 Einstein's Derivation of the Lorentz Transformation.

In addition to the trouble with Maxwell's equations, by 1905 there were two sets of experimental data which indicated that not all physical laws were constant under Galilean transformations. In the first of these experiments Fizeau set about to find the velocity of light in a moving liquid of refractive index n. If the light propagation was unaffected by the liquid motion one would expect to measure a velocity $w = \frac{c}{n}$. If the light propagation depended entirely on the liquid one would expect to measure a velocity $w = \frac{c}{n} \pm v$ where v is the velocity of the liquid relative to the earth. Actually, Fizeau found that neither answer was correct but rather that

$$w = \frac{c}{n} \pm v(1 - \frac{1}{n^2})$$

an intermediate value.

This strange result led Michelson and Morley to their famous experiment by the following argument. Certainly, neither the earth nor the moving fluid constitute a "preferred" coordinate system in space then we should be able to detect the change in motion of the earth relative to this system by measuring the velocity of light from distant stars at two times, one-half year apart, between which times the relative velocity of the earth would have reversed. The result of this experiment was completely negative; the velocity of light from the stars did not change.

This result led Einstein to make the following famous assumption; <u>the velocity of light in free space is independent of the motion of the observer</u>. We can now follow the consequences of this assumption. Imagine a spherical wave of light starting from the origin of coordinates at time $t = 0$, when the starred and unstarred coordinates are coincident. By Einstein's assumption this spherical wave will be propagated in both systems according to:

$$\left.\begin{aligned} &x^2 + y^2 + z^2 = c^2t^2 \\ \text{and,}\quad & \\ &x^{*2} + y^{*2} + z^{*2} = c^2t^{*2} \end{aligned}\right\} \qquad (3.11)$$

We desire to find the functional relation between the two coordinate systems, but we have not assumed that the time must be the same in each system, as does the Galilean transformation. Although the Galilean transformation seemed correct for the dynamics of mass-points (Newton's Laws) this may have been only because the velocities involved were very much lower than c. Since it is very desirable to have the laws of mechanics as well as electromagnetism obey the same laws we shall require that our new relationship reduce to Galilean transformation for $v << c$.

Let us rewrite equation (3.11) so that

$$x^{*2} + y^{*2} + z^{*2} - c^2t^{*2} = x^2 + y^2 + z^2 - c^2t^2 \qquad (3.12)$$

and assume

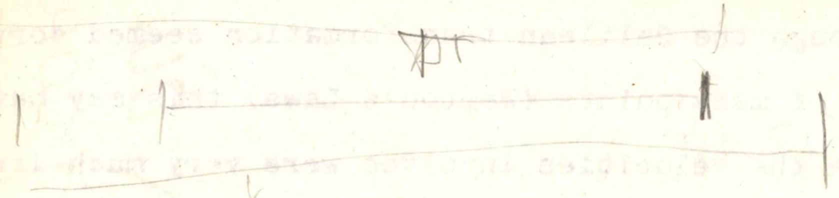
$$\left.\begin{aligned} x^* &= \alpha(x - vt) \\ y^* &= y \\ z^* &= z \\ t^* &= \alpha(\gamma t - \delta x) \end{aligned}\right\} \qquad (3.13)$$

Substituting (3.13) in (3.12) and collecting coefficients of x^2, t^2 and xt respectively, we have

$$\left.\begin{aligned} \alpha^2 - \alpha^2\delta^2c^2 &= 1 \\ \alpha^2v^2 - c^2\alpha^2\gamma^2 &= -c^2 \\ -2\alpha^2v + 2c^2\alpha^2\gamma\delta &= 0 \end{aligned}\right\} \qquad (3.14)$$

Since x and t are independent, their coefficients and those of xt must independently satisfy (3.14).

The solutions of (3.14) are:



$T = \frac{X}{C} \times V = X\frac{V}{C}$ $\quad C = \frac{X}{T}$

$$\left.\begin{aligned} \alpha^2 &= \frac{1}{1-v^2/c^2} \\ \gamma^2 &= 1 \\ \delta^2 &= v^2/c4 \end{aligned}\right\} \text{ or } \left\{\begin{aligned} \alpha^2 &= \infty \\ \gamma^2 &= v^2/c^2 \\ \delta^2 &= 1/c^2 \end{aligned}\right.$$

The second set of these solutions is obviously non-physical. From the first set we will choose the positive roots as corresponding to the velocity direction we have already chosen, and hence we have the well-known Lorentz transformations:

$$\begin{aligned} x^* &= \frac{x - vt}{\sqrt{1-\beta^2}} \\ y^* &= y \\ z^* &= z \\ t^* &= \frac{t - \frac{vx}{c^2}}{\sqrt{1-\beta^2}} \end{aligned} \qquad (3.15)$$

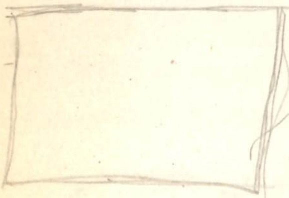
where $\beta^2 = v^2/c^2$

3.4 Kinematics in Special Relativity

The first consequence of these equations (3.15) is that v can never equal or exceed c for a material particle. A second consequence is the apparent contraction of moving bodies. If we place ourselves in the unstarred coordinate system and measure the length of a body (at rest relative to ourselves) whose extremities are at x_2 and x_1 we would achieve the result $x_2 - x_1$. But if this same body is in the starred coordinate system we would find

$$l = x_2 - x_1 = \sqrt{1-\beta^2}\,(x_2^* - x_1^*)$$

and hence the body is foreshortened in the direction of its motion. (The Fitzgerald-Lorentz contraction). Similarly moving clocks apparently run more slowly to a fixed observer.



(It should be pointed out that these effects are very small until $v > c/10$.)

In adding relative velocities we must take account of the transformation laws. Let us assume a third frame of reference (x^{**}, y^{**}, z^{**}, t^{**}) moving with a velocity u relative to (x^*, y^*, z^*, t^*) along the positive x-axis.

$$
\begin{aligned}
x^{**} &= \frac{x^* - \mu t^*}{\sqrt{1-\beta^2}} \\
y^{**} &= y^* = y \\
z^{**} &= z^* = z \\
t^{**} &= \frac{t^* - \frac{u x^*}{c^2}}{\sqrt{1-\beta^2}}
\end{aligned}
\qquad (3.16)
$$

Substituting (3.16) in (3.15), we can show that

$$x^{**} = \frac{x - wt}{\sqrt{1-\beta^2}}$$

$$t^{**} = \frac{t - \frac{wx}{c^2}}{\sqrt{1-\beta^2}}$$

where
$$w = \frac{v + u}{1 + \frac{vu}{c^2}} \qquad (3.17)$$

Again, this law (3.17) for the addition of velocities reduces to $w = v + u$ when $vu \ll c^2$. It can be shown that w can never exceed c, providing, of course, that u or v never exceeds c, and hence, no relative motion can ever exist at a speed greater than the speed of propagation of light in vacuum.

We can also derive the result of Fizeau's experiment using this expression, (3.17). Let v be the velocity of the liquid relative to the earth, and let u be the velocity of light relative to the liquid, i.e., c/n. Then the velocity of light relative to the earth is

$$I = I_0 \sin\omega\left(t^{x} - \frac{x}{c}\right)^{x} = I_0 \sin\omega($$

conventional. relitivistic

$$\text{sourse away} - f_a = \overbrace{f_0\left(1 - \frac{v}{c}\right)} = f_0\left(1 - \frac{v}{c} + \frac{1}{2}\frac{v^2}{c^2}\right)$$

$$\text{tarand } f_T = f_0\left(1 + \frac{v}{c}\right) = f_0\left(1 + \frac{v}{c} + \frac{1}{2}\frac{v^2}{c^2}\right)$$

H ins ⟶

$$\frac{v^2}{c^2} = \frac{6 \times 10^2}{9 \times 10^6} = 10^{-4}$$

30 Kv

$$V = \sqrt{2 \cdot 10^8 \times 3 \times 10} = 2.5 \times 10^6 \text{ m/s}$$

$$\frac{1 \pm v/c \sqrt{1-\beta^2}}{1-\beta^2}$$

$$w = \frac{v + c/n}{1 + \frac{vc}{nc^2}} = \frac{v + c/n}{1 + v/nc}$$

Since $v \ll n\,c$ we can expand and neglect higher powers of v/nc

$$w = \frac{c}{n} + v - \frac{v}{n^2} = \frac{c}{n} + v(1 - \frac{1}{n^2}).$$

Doppler Effect

If a source of light waves is moving relative to the observer the observed frequency will be shifted from the value in the case where observer and source are fixed, even though the observed velocity is constant. Classically the frequency shift is given by

$$f^* = f(1 \pm \frac{v}{c})$$

where the plus sign is to be taken if the source is moving toward the observer. Relativistically this expression must be modified, since there is relative motion of the coordinate systems of the source and the observer, so that

$$I = I_0 \sin\omega\left(T^* - \frac{x^*}{c}\right) = I \sin\omega\left(\frac{T - \frac{xv}{c^2}}{\sqrt{1-\beta^2}} - \frac{\frac{x}{c} - \frac{v}{c}T}{\sqrt{1-\beta^2}}\right)$$

$$f^* = f(1 \pm v/c)/\sqrt{1-\beta^2}$$

If v is small compared to c we can write this, neglecting higher order terms in v/c,

$$f^* \approx f(1 \pm v/c + \tfrac{1}{2}\frac{v^2}{c^2})$$

Hence the relativistic correction is of second order, but its sign is independent of the direction of the motion. always +

Ives made use of this fact in a recent experimental proof of the correctness of the relativistic expression. He measured the shift in frequency of the monochromatic light emitted by hydrogen

$$K.E. \neq \frac{\frac{1}{2}mv^2}{\sqrt{1-v^2/c^2}}$$

$$\boxed{KE = m_0c^2\left(\frac{1}{\sqrt{1-v^2/c^2}}-1\right)}$$

atoms moving in opposite directions in a discharge tube. Classically the average value measured would just be equal to the frequency of light emitted by atoms at rest. Relativistically the average would be shifted from the rest value by an amount equal to $+\frac{1}{2}\frac{v^2}{c^2}f$. The relativistic shift was detected by Ives.

3.5 Relativistic Dynamics

In an isolated system (no external forces), both the momentum and the total energy must remain constant, at least according to Newtonian dynamics. We should like to find the corresponding quantities in relativity which stay constant in isolated systems, and also the law relating these quantities to the forces of the system. We proceed in much the same manner as we did in deriving the transformation laws, but since the algebra is very tedious we will simply state the results here.

The momentum vector becomes

$$p = \frac{m_0 v}{\sqrt{1-\beta^2}} \tag{3.18}$$

and the energy becomes

$$W = \frac{m_0 c^2}{\sqrt{1-\beta^2}} + U \tag{3.19}$$

where U is the potential energy.

We have written m_0 instead of m to imply that m_0 is the measured mass in a system in which the mass point is at rest. We can then regard $m_0/\sqrt{1-\beta^2}$ as the mass when the particle has a velocity v relative to the observer.

The law of motion becomes, then,

$$F = \frac{dp}{dt} = \frac{d(mv)}{dt} = \frac{d\left(\frac{m_0 v}{\sqrt{1-\beta^2}}\right)}{dt} \qquad (3.20)$$

By integrating along a path s and defining the change in potential energy as

$$U_B - U_A = -\int F \cdot ds$$

we find that the total energy W is a constant of the motion. The proof of this statement is left to the problems.

The expression for the total energy W implies that there is a "rest" energy m_oc^2 even when the particle is at rest. Ample proof of the reality of this rest energy has been given of late, and this prediction of rest energy is one of the great triumphs of relativity theory. For the purposes of this course we shall define the kinetic energy as the energy due to the motion of the particle and this is

$$K.E. = mc^2 - m_oc^2 = m_oc^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) \qquad (3.21)$$

It is obvious that the expression for momentum reduces to its classical value when $v << c$. That the kinetic energy also reduces to its classical value can be seen by expanding the term in parentheses in eq. (3.21) by means of the binomial theorm, i.e.,

$$KE = m_oc^2\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} - \cdots\cdots - 1\right)$$

neglecting powers of $\frac{v}{c}$ greater than the second, this becomes,

$$KE \approx m_oc^2\left(\frac{1}{2}\frac{v^2}{c^2}\right) = \frac{1}{2}m_ov^2$$

and hence the relativistic laws give the correct result for law velocities.

A useful relation whose derivation is left to the problems is:

$$\boxed{KE = \sqrt{(m_oc^2)^2 + p^2c^2} - m_oc^2} \qquad (3.22)$$

3.6 Motion in a Magnetic Field

If a charged particle moves in a magnetic field (at relativistic velocities) our equation of motion becomes

$$\frac{d}{dt}\left(\frac{m_0\vec{v}}{\sqrt{1-\beta^2}}\right) = q\,\vec{v} \times \vec{B} \qquad (3.23)$$

The v^2 term in β^2 is just a scalar $(\vec{v} \cdot \vec{v})$ and will not be changed in a magnetic field as we can see by taking the dot product of $\vec{v}$ with both sides of the above equation,

$$\vec{v} \cdot \left[\frac{d}{dt}\left(\frac{m_0\vec{v}}{\sqrt{1-\beta^2}}\right)\right] = q\,\vec{v} \cdot \left[\vec{v} \times \vec{B}\right] = 0$$

The right hand side is $\vec{v} \cdot \vec{F}$, the power or rate of doing work, and it is identically zero. Hence as in the classical case the magnetic field does no work. $d = \frac{dW}{dt}$ $v \cdot F \equiv 0$

Let us assume that initially $\vec{v}$ is perpendicular to $\vec{B}$ and using cylindrical coordinates let us set, in eq. (3.23),

$$\vec{v} = \vec{i}_\varphi\, v_\varphi$$

and

$$\vec{B} = \vec{i}_z\, B_z$$

Then $\vec{v} \times \vec{B} = \vec{i}_r\, v_\varphi B_z$, and since v_φ cannot change with the time (3.23) becomes

$$\frac{m_0 v_0}{\sqrt{1-\beta^2}}\,\frac{d i_\varphi}{dt} = q\,v_\varphi B_z \vec{i}_r$$

By eq. (2.10.7)

$$\frac{d\vec{i}_\varphi}{dt} = \vec{i}_r \frac{d\varphi}{dt} ,$$

and hence

$$\frac{m_0 v_\varphi}{\sqrt{1-\beta^2}} = - q \frac{v_\varphi}{d\varphi/dt} B_z$$

Where the negative sign simply means that the force is centripetal and has no other significance. Since $\frac{v_\varphi}{d\varphi/dt} = R$, the radius of curvature we have,

$$\frac{m_0 v_\varphi}{\sqrt{1-\beta^2}} = qRB_z ,$$

or generally for circular motion in a magnetic field

$$p = \frac{m_0 v}{\sqrt{1-\beta^2}} = qRB \qquad (3.24)$$

classical but with $\frac{m_0}{\sqrt{1-\beta}}$

where p is the magnitdde of the momentum.

3.7 General Remarks

It is difficult to give any general rules concerning when relativistic formulae must be used, i.e., when the classical laws break down seriously. As a sort of general condition it is not too conservative to use relativistic formulae whenever the Kinetic Energy of the particle exceeds 10% of its rest energy, m_0c^2. In certain resonance phenomena, for instance the motion of particles in cyclotrons, 3 or 4% is probably a better figure.

A numerical example may help to illustrate: The rest energy of an electron is

$$m_0c^2 = 9.11 \times 10^{-31} \times 9 \times 10^{+16} = 8.2 \times 10^{-14} \text{ joules}$$

or

$$m_oc^2 = \frac{8.2\times10^{-14}}{1.6\times10^{-19}} = .51\times10^6 \text{ ev. or } .51 \text{ mev.}$$

Hence for kinetic energies of 51,000 ev. we have

$$.51\times10^5 = .51\times10^6\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) \text{ or}$$

$$\frac{1}{\sqrt{1-\beta^2}} = 1+0.1$$

solving $\beta^2 = .1736$ and

$$v = .416\ c$$

And hence even at 50 kv the electron has reached 40% of the speed of light, and its mass has increased by 10%. The proton is 1837 times heavier than the electron so its rest energy is about 935 mev.

For any particle

$$\frac{KE}{m_oc^2} = \frac{1}{\sqrt{1-\beta^2}} - 1 \qquad (3.25)$$

Fig. 3.1 shows the quantity $\frac{KE}{m_oc^2}$ plotted as a function of $\beta^2 = v^2/c^2$. The classical value is given for the sake of comparison.

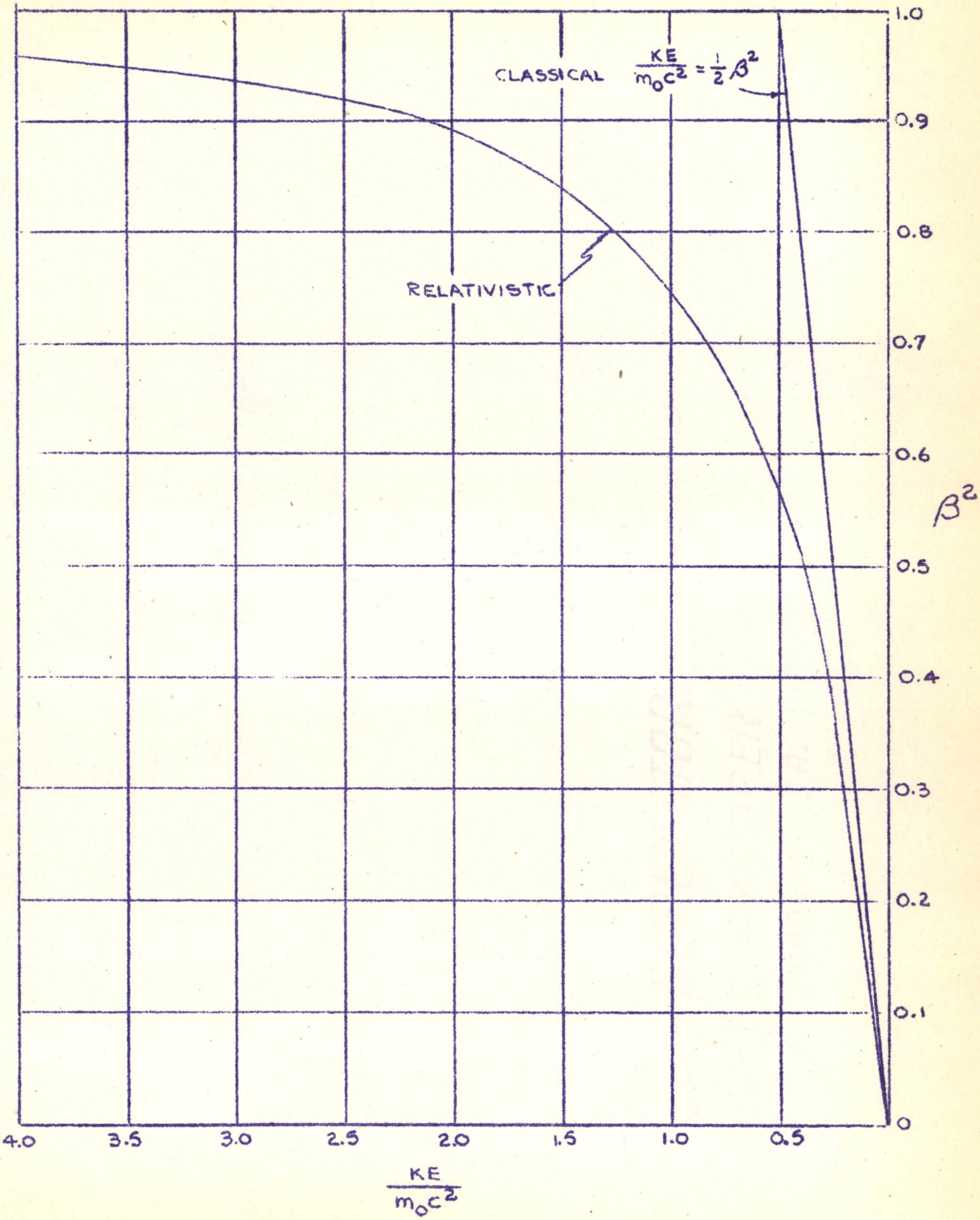

CLASSICAL $\frac{KE}{m_0c^2} = \frac{1}{2}\beta^2$
RELATIVISTIC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
β^2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
$\frac{KE}{m_0c^2}$

Fig. 3.1

CHAPTER IV. HIGH ENERGY PARTICLE ACCELERATORS

In order to study the properties of the atomic nucleus, and to provide extremely high energy x-rays for medical and metallographic uses it is necessary to accelerate charged particles to energies in excess of one million electron volts. Conventional transformer-rectifier type high voltage supplies are difficult to construct for these potentials because of the insulation problems involved, and furthermore such supplies do not give a very stable voltage. Novel methods of acceleration are required, therefore.

These methods of acceleration can be divided into three general types:

1) Direct Voltage
2) Resonance
3) Induction

The very highest energies, over 100 to 500 Mev are obtained by combination of the latter two methods. In this chapter we shall attempt to describe examples of the more widely used accelerators; in any event we shall cover those principles which are fundamental to all particle accelerators.

4.1 The Electrostatic Generator

The most widely used of the direct voltage machines is the electrostatic or Van de Graaff generator. This type of accelerator, while limited to voltages of the order of 10 million, is a remarkably stable source, and is the most precise source of high voltage now existant; stabilities of one part in 10^4 are practicable. With such stability this type of machine is most useful for the accurate determination of nuclear energy levels, although recently, widespread use as a source of high energy x-rays has been made.

The principle of operation is quite simple. The high voltage terminal is a suitably shaped metallic shell (for instance a spherical shell). The high voltage terminal is charged to a high potential by conveying electric charge from ground to the terminal by means of a rapidly moving insulating belt. Fig. 4.1

is a schematic of such a generator. Charge is sprayed from corona points, at about 10 KV potential, on to the moving belt and is removed at the high voltage terminal by an inverse process. Electrons or positive particles are generated inside the high voltage terminal, and can then be accelerated down an evacuated discharge tube and allowed to strike the target material desired. The power required inside the high voltage terminal can be transmitted mechanically by the belt to a dynamo situated inside the terminal.

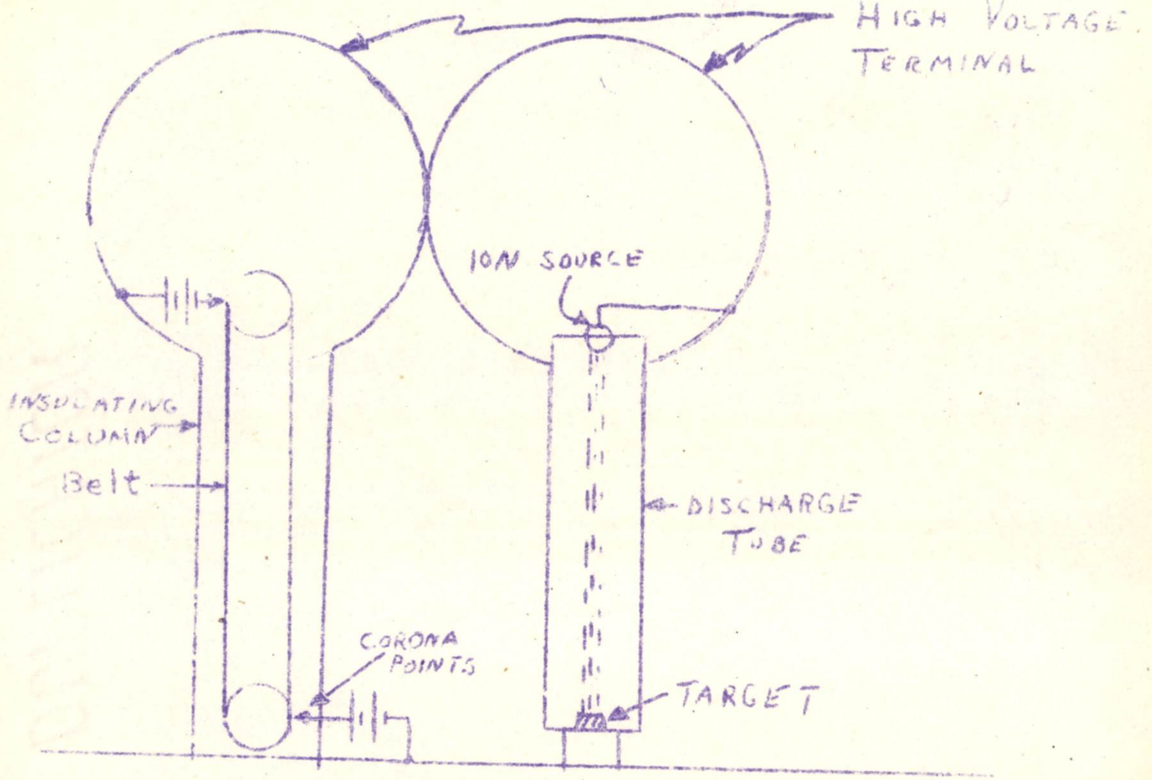


Fig. 4.1

The voltage to which such a generator may be charged depends on the insulating qualities of the supporting column and discharge tube and the atmosphere surrounding the terminal. Usually the discharge tube insulation limits the voltage. The insulating properties of both the column and of the surrounding atmosphere can be enhanced greatly by increasing the pressure of the gas surrounding the terminal and the column. Most modern electrostatic generators are so "pressurized."

In air the breakdown field strength is 3×10^6 volts/meter.

In order to obtain a potential of 10^6 volts it is necessary to have a sphere whose radius,

$$a = \frac{V}{E_B} = \frac{10^6}{3\times 10^6} = \frac{1}{3} \text{ meter.}$$

For stable operation a radius of about two times this is required.

While the inherent stability of an electrostatic generator is quite good, this stability can be greatly improved by the use of "feed-back" systems; i.e., any small change in the voltage can be detected and used to affect the charging current so that the variation tends to be suppressed.

$E = \frac{V}{r}$

The electrostatic generator works equally well for positive or negative, heavy or light particles.

4.2 The Cyclotron

When one tries to increase the potential of a direct voltage device beyond 10 million volts the insulating requirements become almost insurmountable. If we are to obtain particles of energy much in excess of 10^7 e.v. then we must resort to some other means of acceleration. Although in 1928 an elementary form of linear accelerator (see Sec. 4.8) was successfully used, the first important development of resonance acceleration was the cyclotron of E. O. Lawrence. The cyclotron works on the simple principle that the angular velocity of a charged particle moving in a magnetic field is a constant, and is given by,

$$\omega = q/m\, B$$

(ω is not constant, of course, in the relativistic region).

If now we continue to apply an RF field of frequency

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\,\frac{q}{m}\cdot B \qquad (4.1)$$

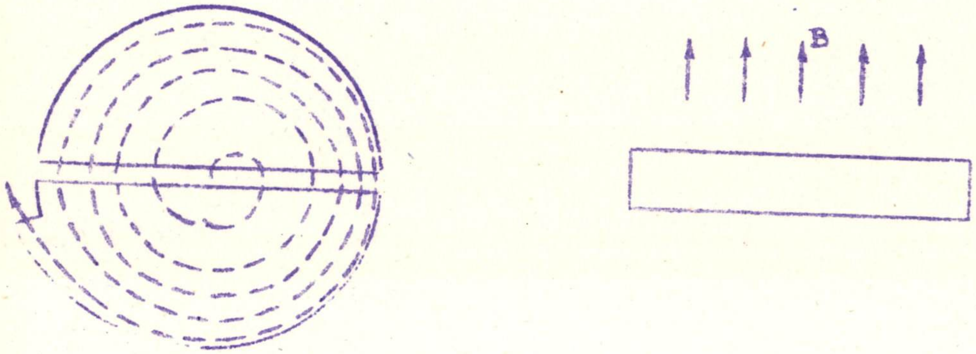
PV

g/m B

W=

in phase with the angular velocity ω of the particles in a magnetic field, we should be able to transform the RF energy into particle kinetic energy.

This principle may be readily understood by referring to Fig. 4.2 which shows horizontal and vertical sections through the accelerating chambers. The electrodes or "Dees" can be thought of as the two halves of a hollow metallic pill box split along a diameter. Charged particles, generally protons or deuterons but sometimes α-particles, are created near the center of the electrodes by direct ionization of the gas in the chamber or by a suitable ion gun. If now an RF field is applied to the Dees these particles will find themselves acted upon by an electric force which will tend to accelerate them. When the particles pass into the hollow dee they will no longer be acted on by the RF field, but the magnetic field will bend them in a circular arc traversed with angular velocity ω. If the particle is favorably accelerated on its first traversal of the gap, and if $f = \frac{\omega}{2\pi}$, it will again be accelerated on its second and subsequent traversals and will follow the spiral path shown, gaining energy on each traversal of the gap. Particles which start in an unfavorable phase will soon be lost.



HORIZONTAL SECTION

VERTICAL SECTION

Fig. 4.2

Since, $2\pi f = q/m\ B = \omega$

and

$$KE = \frac{1}{2} mv^2 = \frac{1}{2} m\omega^2 r^2$$

$$KE = \frac{1}{2} m(\frac{q}{m})^2 B^2 r^2$$

is the kinetic energy at any radius r. The kinetic energy is independent of the RF voltage and depends only on the size and intensity of the magnetic field. The energy of the emerging particles, at radius r_o, is then

$$KE = \frac{1}{2} m(\frac{q}{m})^2 B^2 {r_o}^2 \qquad (4.2)$$

Equations (4.1) and (4.2) are the fundamental relations for the cyclotron. It is apparent that the resonance condition (4.1) cannot be satisfied at high speeds because of the relativistic increase of mass, and in fact constant field and frequency cyclotrons are limited to kinetic energies of about 1.5% of the rest energy m_oc^2, which is about 30 Mev for deuterons. It is obvious from this that cyclotrons have no practical use for accelerating light particles (electrons) since the limiting energy would be of the order of 10 Kev.

4.3 Resonant measurement of e/m

However, an interesting application of the cyclotron resonance principle to the precision determination of the electric charge to mass ratio was made by Dunnington (then a student of Lawrence) about 1933. Referring to Fig. 4.3 an evacuated chamber is placed in a magnetic field, perpendicular to the page, as shown. Electrons are emitted from a source S and can be collected at C. The segment SC is insulated from the main can A, and a fixed radio frequency field is applied between them. An appropriate slit system is provided as shown. Electrons leaving S will be accelerated by the RF field in crossing the first gap. (Electrons which emerge during a decellerating phase will be quickly lost and need not con-

$KE = m$

$KE = \frac{1}{2}mv^2$

$v = \sqrt{\frac{2KE}{m_1}}$

cern us here). Initially the device acts as a velocity

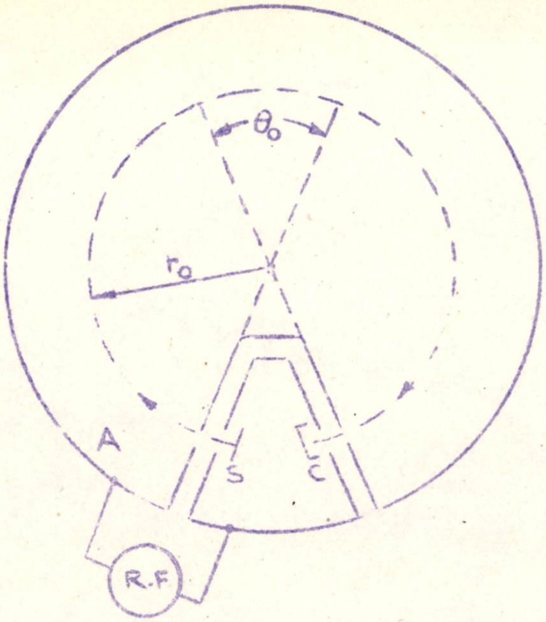


Fig. 4.3

selector since only those electrons which satisfy

$$\frac{mv_0}{r_0} = eB \qquad (4.3)$$

can arrive at C. Now if the applied frequency f is initially low so that the electrons move from the first slits to the final slits in less than a period then the decellerating force at the final slit system is less than the accelerating force at the initial slits. Hence, all the electrons of correct v_0 are collected. If the frequency is adjusted so that the time of flight is exactly one period then the decellerating force is equal to the accelerating force. Further increase of the frequency will then cause a sharp drop in the current at C.

The condition for this is,

$$v_0 = \frac{(2\pi - \theta_0)r_0}{T_0} = (2\pi - \theta_0)r_0f_0 \qquad (4.4)$$

Combining (4.3) and (4.4) we have

$$\frac{e}{m} = \frac{(2\pi - \theta_o)f_o}{B}$$

where f_o is the frequency at which the sharp drop in current begins. This method is capable of good precision and has probably given the best dynamical value of e/m which has been obtained.

4.4 The FM Cyclotron

If we are to accelerate particles to kinetic energies greater than 1 or 2% of their rest energy by cyclotron principles it is evident from eq. (4.1) that we must vary either B or f in order to make up for the relativistic mass change. At first sight it would seem proper to shape the magnetic field B so that it is increased as r increased in just the same manner as the mass increases with radius. This would involve no changes with time and the cyclotron would remain a continuously acting accelerator. Unfortunately, for stability reasons, it is necessary to decrease B slightly as r increases. Increasing B with r would lead to defocussing in the vertical direction (the direction of B). Hence, the only possibility is to vary the frequency with time.

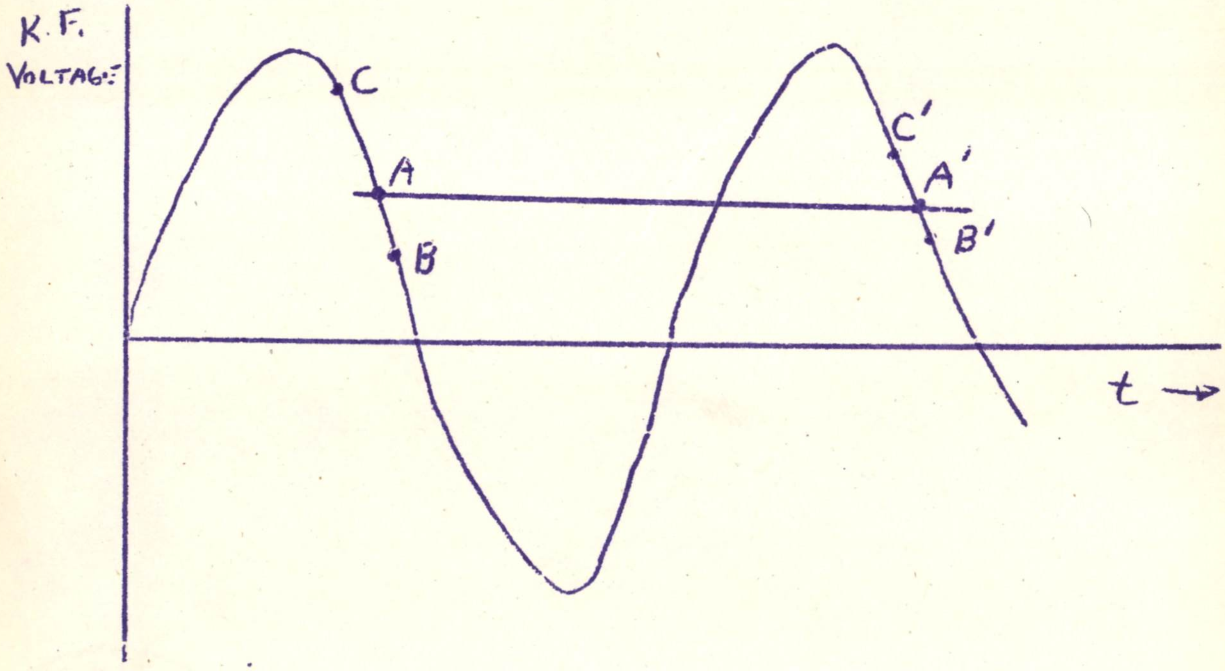
Now if the frequency varies with the time, two conditions will be imposed on the cyclotron operation that are not imposed on the operation of the constant frequency cyclotron. Firstly, only those particles which start from rest when $f = \frac{1}{2\pi}\frac{q}{m_o}B$ willbe accelerated, and consequently the cyclotron operation will not be steady but will be pulsed. Secondly, the frequency change must just match (inversely) the change in mass of the particle. But the time rate of change of the particle energy, and therefore mass, will depend on the RF field when the particle crosses the gap between the dees. Thus it would seem that the oscillator frequency must be externally synchronized with the rate at which the particle gains mass. Fortunately for cyclotron constructors such is not the case, within limits, as will be seen.

The Principle of Phase Stability

We will endeavor to show that under the action of a frequency-modulated field the particles will tend to remain in the right phase for acceleration to take place and for the relativistic resonance condition to be satisfied,

$$f = \frac{1}{2\pi} \frac{e}{m_o} B \frac{1}{1 + \frac{K.E.}{m_o c^2}} \qquad (4.5)$$

where f and K.E. are the instantaneous values of frequency of field and kinetic energy of the particles



Fig, 4.4

In Fig. 4.4 let us assume that a particle which comes through the accelerating gap at A receives just the right energy to come through the gap a cycle later at A' in the identical phase relation.

A particle which arrives at the gap a little later than the "correct" particle, phase B, will receive less "kick" from the RF field and will tend to catch up, phase B'. At first this sounds backwards but we must remember that the time required for a particle to make a revolution is equal to,

$$T = \frac{2\pi}{\omega} = \frac{2\pi mB}{q}$$

and hence the late particles will gain less energy and therefore mass, and will tend to make the next revolution in less time and will tend to catch up to the correct particles. Similarly, particles arriving too soon will receive more kick, will take longer for a revolution, and will tend to be caught by the correct particles, phases C and C'. Hence, the motion on the right side of the positive cycle is stable in phase, and will remain so if the frequency is not varied at too rapid a rate. Particles on the left-hand side of the positive cycle are in an unstable phase and will soon fall out of resonance.

Hence, the beam in an F.M. cyclotron will consist of several groups of particles bunched in phase which in turn will constitute a pulse. The number of bunches per pulse will depend on the rate of change of frequency with time.

Such an F.M. cyclotron has been in operation at Berkeley for about two years, producing 200 Mev deuterons and 400 Mev α-particles for a total frequency change of about 18%. The energy in such a cyclotron is given by the relativistic expression,

$$KE = \sqrt{(m_o c^2)^2 + B^2 r_o^2 q^2 c^2} - m_o c^2 \qquad (4.6)$$

These two equations (4.5)and (4.6) are the fundamental relations for F.M. cyclotrons.

4.5 The Betatron

We have already examined the cases of the direct voltage accelerator (electrostatic machine) and the resonance accelerator (cyclotron). There remains a third class, the induction accelerator, of which the so-called betatron is the principal, if not only, member. As its name implies the betatron is used as a source of high energy electrons (Beta rays) and works on the principle of acceleration by induced electromotive force.

Consider an electron which is moving in a circular orbit of radius r_o. If this orbit links magnetic flux which is changing with the time there will be an emf developed which will produce a force on the electron. Furthermore, if the orbit and the flux are both symmetric about a common axis, this force will be directed along the circular path (in the azimuthal direction), and will accelerate the electron if we choose the sense of the rotation to satisfy Lenz' law. The orbit itself will remain circular providing

$$p = B_o e r_o \qquad (4.7)$$

where B_o is the magnetic field at the orbit and must be a function of the time, since p is to increase. The emf developed is

$$\text{emf} = \frac{d\phi}{dt} \qquad (4.8)$$

where ϕ is the total flux linked by the orbit. This emf will do work on the electron and the work per revolution is,

$$\text{Work} = e\,\frac{d\phi}{dt}$$

This is the same as having a force F acting on the electron and we may write

$$\text{Work} = e\,\frac{d\phi}{dt} = F \cdot 2\pi r_o$$

or

$$F = \frac{e}{2\pi r_o} \times \frac{d\phi}{dt} \qquad (4.9)$$

Since $F = \frac{dp}{dt}$, we have

$$\frac{dp}{dt} = \frac{e}{2\pi r_o} \times \frac{d\phi}{dt} \qquad (4.10)$$

integrating, we have,

$$p - p_1 = \frac{e}{2\ r_o} (\phi - \phi_1) \qquad (4.11)$$

where p_1, ϕ_1 are the initial values of the momentum and the flux. If we start the electrons in such a manner that

$$p_1 = \frac{e}{2\pi r_o} \times \phi_1$$

then

$$p = \frac{e}{2\pi r_o} \times \phi \qquad (4.12)$$

for the duration of the acceleration. Using eq. (4.7) we have

$$B_o e r_o = \frac{e}{2\pi r_o} \times \phi$$

or

$$B_o = \frac{1}{2\pi r_o^2} \times \phi$$

by definition $B_{AV} = \frac{1}{\pi r_o^2} \times \phi$, and hence,

$$B_o = \frac{1}{2} B_{AV} \qquad (4.13)$$

or the field at the orbit must be 1/2 the average field enclosed by the orbit in order that accelerated motion at constant radius be possible. Equations (4.7), (4.12) and (4.13) may be regarded as the fundamental relations for the betatron, when taken together with the relativistic energy equation

$$K.E. = \sqrt{(m_o c^2)^2 + (B_o e r_o c)^2} - m_o c^2 \qquad (4.14)$$

For very high energies ($\sim$100 Mev) the $m_o c^2$ terms in (4.14) may be neglected and we have,

$$KE = B_o e r_o c \qquad (4.15)$$

and it is seen that the product $B_o r_o$ is a measure of the kinetic energy attainable. For large energies we want large fields and this implies the use of iron cores. Such cores saturate around 16,000 gauss. The energy, however, is determined by B_o the orbital value of B, and this cannot exceed one-half the saturation value. Hence, the betatron is limited in energy (for a given weight of iron) to values corresponding to 4 - 6,000 gauss.

Actually, closer examination of the fundamental equations allows somewhat higher values. It is actually the time rate of change of the field at the orbit which must be one-half the time rataeof change of the average field. By appropriate D.C. biasing the magnetic field at the orbit may reach higher values, but not the normal saturation values.

Since the magnetic field is changing with the time laminated iron is necessary to cut down eddy current losses. Even with laminated iron, however, the frequency most widely used is 60 cycles, although, as we shall see, a higher frequency would permit higher energies. The frequency of 60 cycles is a good practical compromise between particle energy and power requirements.

In practice the magnet and a condenser bank from a resonance circuit whose free oscillation frequency is the desired 60 cycles. With the magnet disconnected the condenser bank is charged from the line or from an auxiliary generator. The connection to the magnet is then made and simultaneously low energy electrons are injected into the orbit. The first quarter cycle of the discharge is used for acceleration; afterward the stored magnetic energy is returned to the condenser bank. The average power required from the line is then jusst that function, roughly 20%, which is lost due to resistance in the connectors and condensers, eddy currents and hysteresis.

The acceleration process is a long one (in the time scale of the electrons) requiring one-240th of a second. Since the electrons are moving approximately with the speed of light most of the time, they travel about 10^6 meters or 150,000 revolutions of 1 meter radius, which is the radius of the G.E. 100 Mev betatron. The energy acquired per revolution is thus about 600 ev. (Comparing this energy gain per revolution with $e\frac{d\phi}{dt}$ is left to the problems.)

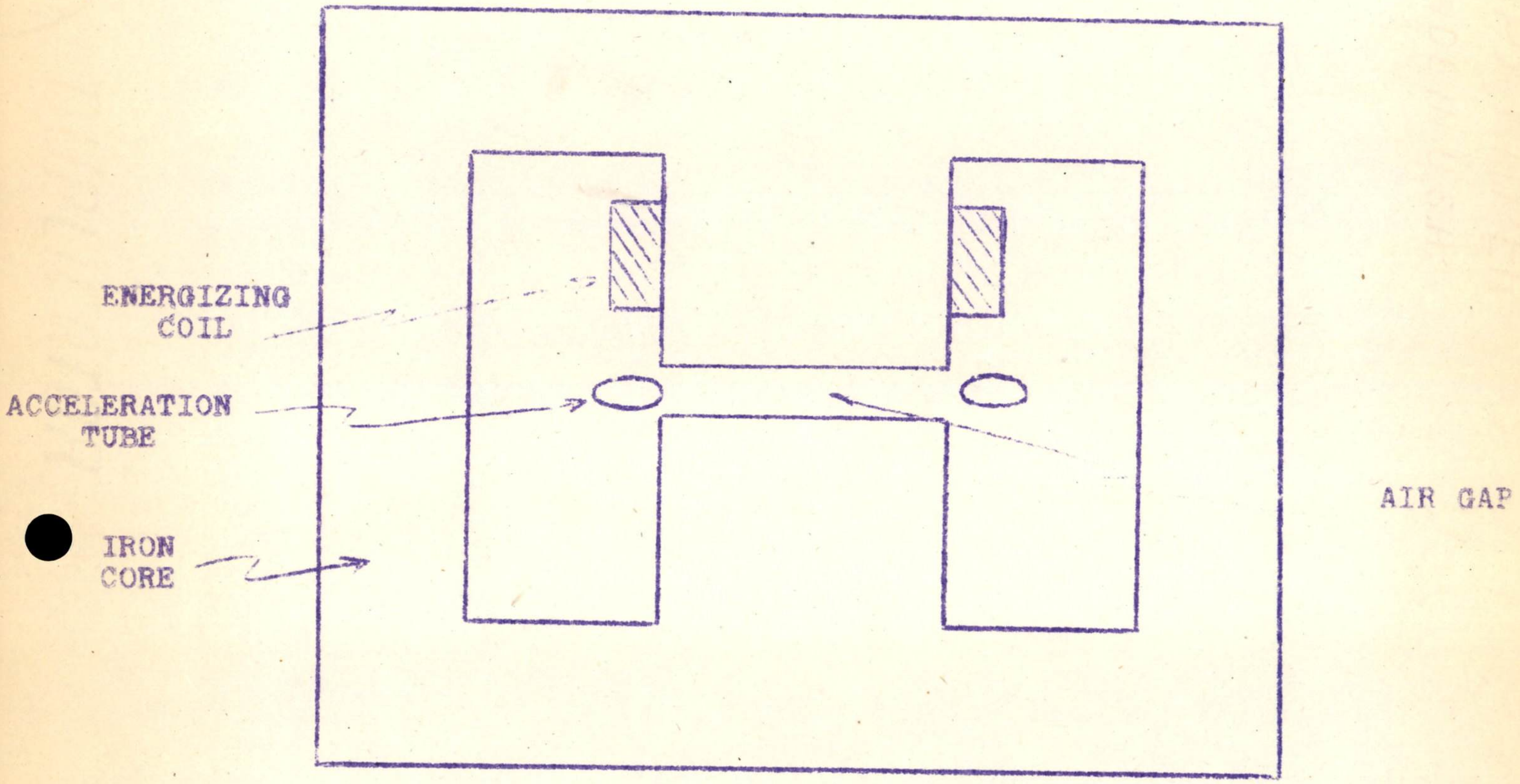
Very many betatrons in the range 2-20 Mev have been constructed for x-ray use. The G.E. 100 Mev betatron is the largest now operating, although one for the 300 Mev range is under construction at the University of Illinois.

There are, however, better ways of achieving high energies. In addition to the high cost and the unfavorable condition imposed by eq. (4.13), there is another very serious limitation to the maximum energy to be obtained from the betatron. It is well known that accelerated charges radiate. The azimuthal acceleration is small (for electrons), but the centripetal acceleration is about 10^9 times larger. Theory predicts radiation along the forward direction of motion of the electrons at a rate which is proportional to the fourth power of the kinetic energy. At energies of the order of 100 Mev this radiation becomes appreciable, and has actually been observed visually at G.E., but at energies of the order o of 500 Mev it becomes limiting. That is, the electrons radiate energy as fast as they receive it from the induction field. The only way to make up for this is to vary the magnetic field at a faster rate in order to increase the rate of acquisition of energy by induction. Such an increase in $\frac{d\phi}{dt}$ is impractical for iron magnets but might be practical for an air-core magnet However, as we shall see in Section 4.7 there are better ways over these difficulties.

Fig. 4.5 shows a schematic cross-section of a typical betatron. The air gap is necessary in order to get the proper field at the orbit. In addition to the requirement that

$$\frac{dB_o}{dt} = \frac{1}{2}\frac{dB_{AV}}{dt}$$

there is another requirement, for stability reasons, on the rate of change of B with r at the orbit, and this will be discussed in the next section.

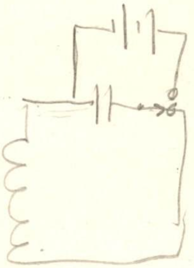


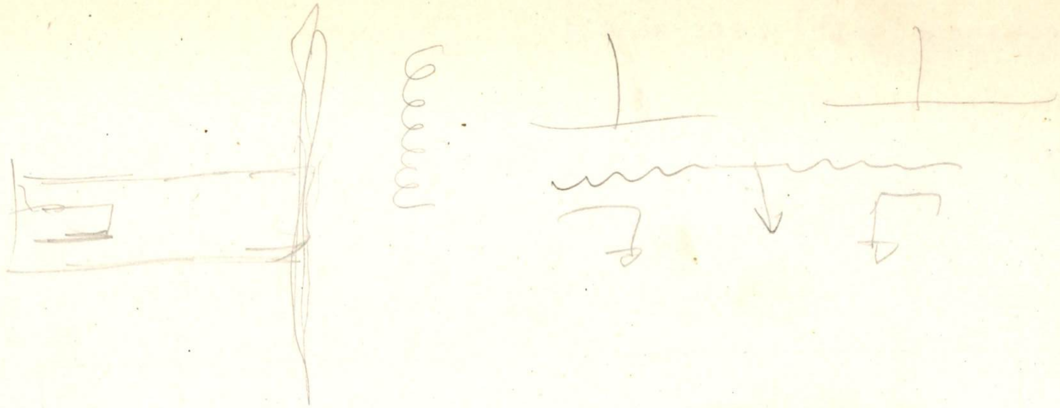
$$KE = \sqrt{\cancel{m_0 c^2} + B^2 r^2 e^2 c^2} \cancel{+ m_0 c^2}$$ negligible

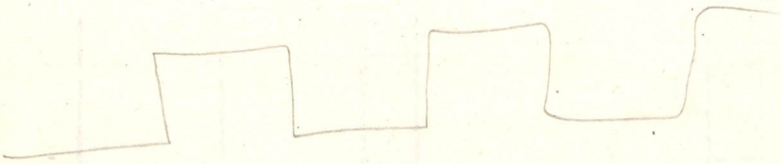
$$\simeq B_0 r_0 e c$$

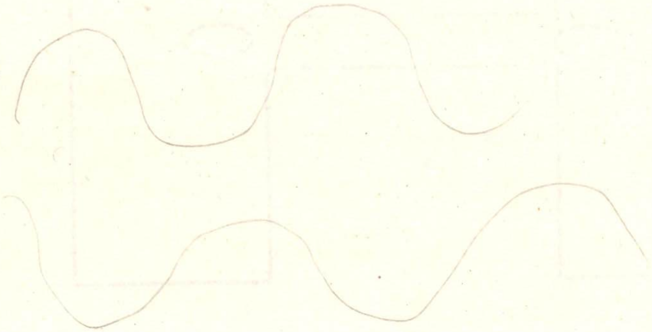
Section through Betatron

Fig. 4.5









4.6 Stability Considerations in Magnetic Accelerators

We have seen that the principle of phase stability is necessary for the successful operation of resonant magnetic accelerators if the RF field changes with time. If the magnetic field were to change with the time with the RF field constant (the synchrotron) or if both fields change with the time (the FM synchrotron), this principle is still valid and is essential for operation. However, we have not yet considered stability requirements to keep the particles at the right radius and to prevent them from colliding with the top and bottom of the accelerating chamber, i.e., r and z stability. The following discussion is valid for all magnetic accelerators.

Let us assume that the field is not quite uniform and that besides its usual z component B_z, it has as well an r component B_r. Then the equations of motion for the r and z directions are, for a negatively charged particle,

$$\left.\begin{aligned} \frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 &= -\frac{e}{m} B_z \cdot r \cdot \frac{d\theta}{dt} \\ \frac{d^2z}{dt^2} &= \frac{e}{m} B_r \cdot r \cdot \frac{d\theta}{dt} \end{aligned}\right\} \qquad (4.16)$$

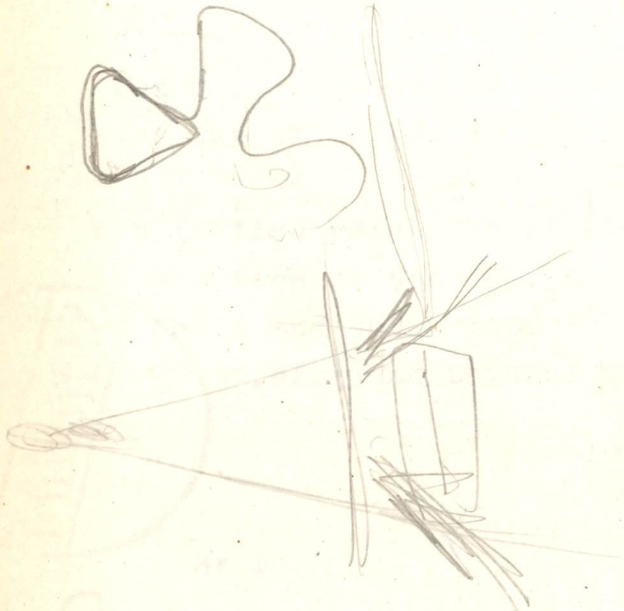
The above equations are based on the assumption that the field is cylindrically symmetric, that $B_\phi = 0$, and that time variations in B can be neglected since these are very slow. That is, we wish to investigate the variations of r and z from their equilibrium values, r_o and 0, respectively, in short times. In this approximation we may consider the kinetic energy is constant so that

$$\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 = v_o^2 \qquad (4.17)$$

where v_o is constant.

The equilibrium solution of eqs. (4.16) is then, for $r = r_o$, $z = o$, $B_r = o$

$$\frac{d\theta}{dt} = \omega_o = \frac{e}{m} B_z, \quad \frac{dz}{dt} = 0, \quad \frac{dr}{dt} = 0 \qquad (4.18)$$



and $v_o = r_o \omega_o$.

We wish to investigate small departures from the equilibrium solutions. Let us set

$$r = r_o + r', \quad \frac{d\theta}{dt} = \omega_o + \omega', \quad z = z \qquad (4.19)$$

where r', ω' and z are small quantities.

Let us assume

$$B_z = B_o\left(\frac{r}{r_o}\right)^{-n}, \text{ then } \frac{\partial B_z}{\partial r} = -\frac{n}{r} B_o \qquad (4.20)$$

By Taylor's expansion,

$$B_z = (B_z)_{r=r_o} + \left(\frac{\partial B_z}{\partial r}\right)_{r=r_o} \cdot r' + \dots \approx B_o - \frac{nr'}{r_o} \cdot B_o \qquad (4.21)$$

and

$$B_r = \left(\frac{\partial B_r}{\partial z}\right)_{r=r_o, z=0} \cdot z + \dots$$

by Maxwell's eq. (3.4), we have*

$$\frac{\partial B_r}{\partial z} = \frac{\partial B_z}{\partial r} \qquad (4.22)$$

and hence,

$$B_r \approx \left(\frac{\partial B_z}{\partial r}\right)_{r=r_o, z=0} \cdot z = -\frac{nz}{r_o} \cdot B_o \qquad (4.23)$$

Let us now examine the energy relation (4.17) and substitute our assumed solutions (4.19) in it. We have

---

* Actually we must first transform eq. (3.4) to cylindrical coordinates. Then neglecting real current and $\partial E_\phi / \partial t$, which is consistent with our present approximation, we have, $\frac{\partial H_r}{\partial z} = \frac{\partial H_z}{\partial r}$, from which (4.21) follows.

$$\left[\frac{d}{dt}(r_o + r')\right]^2 + (r_o + r')^2(\omega_o + \omega')^2 + \left(\frac{dz}{dt}\right)^2 = v_o^2$$

but

$$\frac{dr_o}{dt} = o, \text{ and } \left(\frac{dr'}{dt}\right)^2 + r_o^2\omega_o^2 + \left(\frac{dz}{dt}\right)^2 = v_o^2$$

and hence, we have

$$r_o\omega' + r'\omega_o = 0$$

or

$$\omega' = -\frac{r'}{r_o}\omega_o \qquad (4.24)$$

when we have neglected products of the small quantities, r' and ω'.

Let us now substitute our expressions (4.19), (4.21), (4.23) and (4.24) in the equations of motion (4.16). We have for the r motion,

$$\frac{d^2r'}{dt^2} = (r_o + r')(\omega_o + \omega')^2 - \frac{e}{m}\left(B_o - \frac{nr'}{r_o}\cdot B_o\right)(r_o + r')(\omega_o + \omega')$$

remembering that $\frac{e}{m}B_o = \omega_o$, and neglecting products of small quantities this expression becomes,

$$\frac{d^2r'}{dt^2} = r_o\omega_o\left(\omega' + n\omega_o\cdot\frac{r'}{r_o}\right)$$

Using equation (4.24), we have

$$\left(D^2 + (1-n)\omega^2\right)r' = 0$$

$$\frac{d^2r'}{dt^2} + (1-n)\omega_o^2 r' = 0 \qquad D = \pm\sqrt{(1-n)}\,\omega_0 \qquad (4.25)$$

$$r' = C_3 \sin Dt + C_4 \cos Dt$$

And hence the r motion is stable if $n < 1$.

Let us now examine the equation for z.

$$\frac{d^2z}{dt^2} = \frac{e}{m}\left(-\frac{nz}{r_o}B_o\right)(r_o + r')(\omega_o + \omega') = -\frac{nz}{r_o}\omega_o(r_o + r')(\omega_o + \omega')$$

which becomes on neglecting products of small quantities (remember z itself is small),

$$\frac{d^2z}{dt^2} + n\omega_o^2 z = 0 \qquad (4.26)$$

and hence for stable z motion we must have $n > 0$.

Taking eqs. (4.25) and (4.26) together we require that

$0 < n < 1$.

It is customary to take n about .75 for betatrons and synchrotrons. This tends to reduce the vertical oscillations, at the expense of the radial oscillations to be sure, but the vertical direction is the direction of increasing air gap.

4.7 The Synchrotron

The synchrotron is primarily a type of resonance accelerator in which the magnetic field is varied with the time. In the ordinary (electron) synchrotron the particles move in a circular orbit of radius r_o, and are accelerated by an R.F. field. Let us assume that 2 Mev electrons are injected into such an orbit. The electrons are already moving with a velocity of .98c so that the acceleration, if you please, process takes place at almost constant velocity, i.e., the mass and energy increase although the velocity increases but slightly. If we apply an R.F. field across a pair of electrodes, through which the electrons pass, the electrons will stay in resonance with the field providing,

$$f = \frac{v}{2\pi r_o} \doteq \frac{c}{2\pi r_o} \qquad (4.27)$$

In order for the electrons to remain in a stable orbit of radius r_o, we must have

$$\omega_o = \frac{v_o}{r_o} \doteq \frac{c}{r_o} = \frac{e}{m} B_o = 2\pi f \qquad (4.28)$$

Since m increases with the time we must have B_o increasing with the time also.

The usual type of accelerating chamber for a synchrotron is shown in Fig. 4.6.

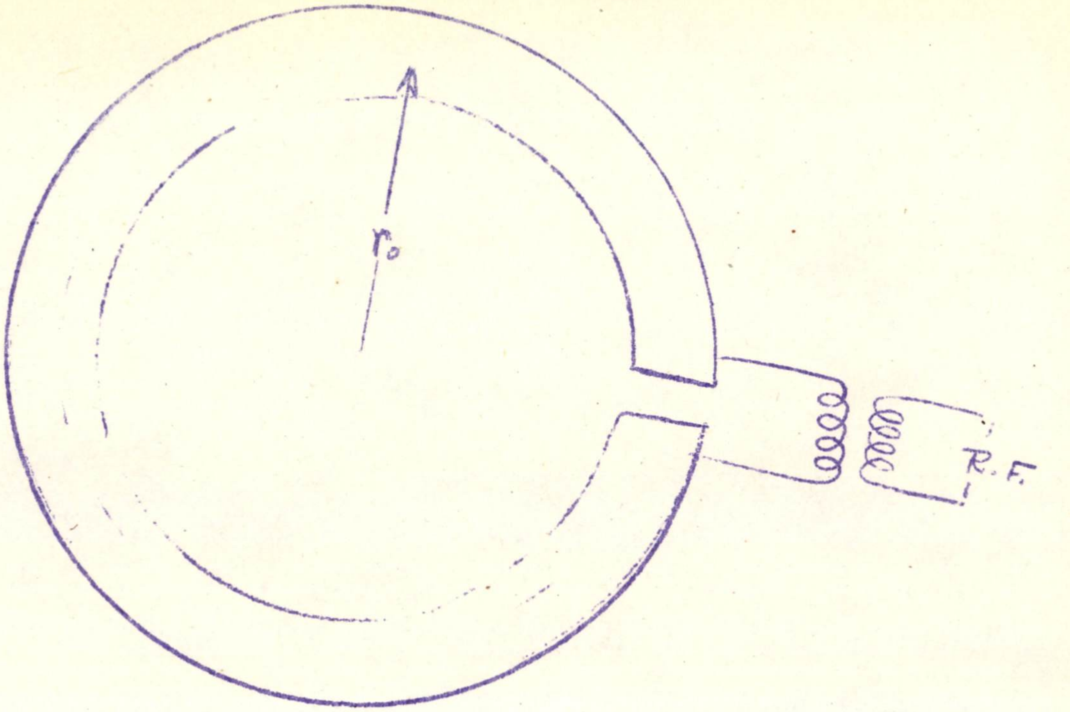


Fig. 4.6

The chamber consists of a hollow metal doughnut with a gap across which the RF is applied. Except for the time of passage across the gap the electrons are moving in electric-field free space. Since the magnetic field must increase with the time, just as in the betatron, the acceleration process is slow, and only moderate RF voltages are required of the order of 5,000 volts for a 300 Mev machine. While there is always some betatron action, i.e., induction acceleration, most of the acceleration comes from the RF field. The synchrotron has three distinct advantages over the betatron;

1) The relation $B_o = \frac{1}{2} B_{AV}$ no longer is required.

2) No field is required inside the orbit, so the iron may be used economically.

3) Radiation by the electrons and the concurrent loss

of energy can be made up for by increasing the R.F. voltage.

In practice all synchrotrons now operating and all but one now building start with betatron action, since the time varying magnetic field is present. When the electrons have attained relativistic energies, about 2 Mev, the field inside the orbit is made to saturate and the RF field is turned on. Since electron synchrotrons are used only for energies very much greater than m_oc^2 the final energy is very nearly given by,

$$K.E. \approx B_o r_o ec$$

where B_o is the final value of the orbital field, and values very near to the saturation value for iron can be used.

According to eq. (4.28) the magnetic field must change at the same rate as the mass changes. No device is required to insure this, however, since the principle of phase stability applies.

Although losses from radiation are much less important in the synchrotron than in the betatron, nevertheless the acceleration process cannot be carried out indefinitely. At 10^3 Mev the radiation losses are such that 70 Kev of RF energy must be added per revolution to compensate. Hence, RF insulation provides the practical limitation in energy, which is probably about three times as great as that in the betatron.

4.8 The FM Synchrotron

In order to accelerate heavy particles to energies equal to their rest energy (935 Mev for the proton) or greater, we must go to machines of considerable size, and considerations of cost become extremely important. The FM cyclotron in principle would be quite workable in the 1000 Mev range, but in point of fact it is much more expensive than a synchrotron. This is because the cyclotron needs a high magnetic field over the entire area enclosed by the final orbit while the synchrotron needs a high magnetic field only at the orbit.

Now the particles in all magnetically operated machines in the very high energy region have energies approaching $B_o r_o ec$. B_o is limited to something around 15,000 gauss so for practical purposes the energy is proportional to the radius. This means that if we double the radius of a cyclotron to get double the energy we must multiply the volume, the mass and hence the cost by almost eight times. In the synchrotron, however, this latter factor is between two and four.

In order to use the synchrotron with heavy particles we must change the frequency since

$$f = \frac{v}{2\pi r}$$

and v will not be effectively constant until energies, for protons of the order of 4000 Mev are attained.

With both the magnetic field and the frequency changing with the time, we cannot rely on the principle of phase stability alone, but we must control the rate of frequency change so that it is properly adjusted to the rate of change of B_o. The particles will then settle in stable orbits and stable phases.

4.9 The Linear Accelerator

The Linear Accelerator is a resonance device in which the accelerating force is provided by an RF field traveling down a wave guide. We must see then that he velocity of the particles is equal to the phase velocity of the electromagnetic wave. The principle of phase stability will permit the particles to extract energy from the wave, but careful attention must be given to focussing problems.

A mathematical discussion of the linear accelerator is beyond the scope of these notes.

ch.5

Space Charge

Ch. 10 - Frank

Ch 7 - M&S

CHAPTER FIVE. ATOMIC STRUCTURE

Ref. M. and S. Chap. IX

One of the most potent bits of evidence that atoms possess an internal structure is the fact that, under proper conditions, they emit radiation in the form of spectrum lines. The problem of atomic structure is to devise a model which, on the basis of known physical laws, will account for these lines. At first sight on the basis of classical mechanics, we might suppose that if an atom emits a spectrum line of frequency f then it must contain within it an oscillator of this same frequency. But we soon run into difficulties with such a theory, as is now well known.

The essential feature of atomic spectra, which was early discovered in a purely empirical manner, is that it is possible to discover a series of numbers or so-called spectral terms, such that by taking differences between these numbers we arrive at the frequencies of this observed spectral lines. The number of such spectral terms required to explain a given spectrum is, in general, much smaller than the number of observed spectrum lines. We shall not here go into this matter in any detail, such as the way these spectral terms fall into groups, with certain selection rules to say which terms do or do not combine, but for the present content ourselves with the existence of these spectral terms. The problem of atomic spectra will therefore be solved if we can devise a model which will account for the spectral terms and give a correct set of selection rules stating how these terms are to be combined.

Various pieces of evidence, notably Rutherford's experiments on the angular distribution of α-particles scattered by atoms, point to the fact that an atom consists of a small, massive core carrying a positive charge Ze, surrounded by Z electrons. These electrons are thus acted on by very strong electrostatic forces of attraction by the nucleus and in order to prevent collapse of such an atom it is necessary to assume that the electrons are in motion. This leads to a fundamental difficulty since these orbital electrons will be undergoing acceleration; we have seen that according to classical electromagnetic theory a charged particle moving with acceleration will radiate energy. As the electron radiates, its energy will change and so too will its frequency of motion. Such an atom should therefore emit a continuous spectrum, contrary to observation. We must, therefore, postulate that when an electron is bound to an atom it does not, in general, radiate energy even though it is moving with acceleration.

More generally, Bohr in 1913 laid down the following fundamental postulates.

1. Atoms exist in "stationary states", without radiating.
2. When an atom changes from one stationary state $\overset{W}{\cancel{E}}_1$ to another $\overset{W}{\cancel{E}}_2$ of lower (higher) energy it emits (absorbs) a quantum of energy, i.e., electromagnetic radiation, of frequency f where

$$\boxed{hf = \overset{W}{\cancel{E}}_1 - \overset{W}{\cancel{E}}_2} \qquad (5.01)$$

The constant h is the one introduced by Planck some years before this to explain the spectrum of a black body, and successfully used by Einstein to account for the photoelectric effect. These postulates of Bohr are extremely important. Atomic theories differ in the way energy states are calculated, or in the atom model used, but Bohr's postulates remain unaltered.

In order to calculate the energy states, $\overset{W}{\cancel{E}}_1$ it is necessary to set up an atom model and make one additional postulate which introduces the constant h, i.e., quantizes the energy states. In the Bohr theory it is the angular momentum of the electron which is quantized. We (arbitrarily) say that the angular momentum is an integral multiple of $h/2\pi$ and will attempt to make this choice more plausible latter. As an example let us consider the simplest case of a single electron in the field of a nucleus of charge $+Ze$. For simplicity we will consider only the case of circular orbits. The electrostatic force between particles is given by

$$F = -\frac{Ze^2}{4\pi\varepsilon_o r^2} \qquad (5.02)$$

and the potential energy is

$$\boxed{V = -\frac{Ze^2}{4\pi\varepsilon_0 r}} \qquad (5.03)$$

We have already seen that gravitational forces can be neglected in comparison with electrostatic ones in problems such as this. Since the nucleus is very much more massive than the electron ($M \sim 1840Zm$) we can assume the electrom to revolve about a fixed center (infinite nuclear mass). Then we shall have

$v = \omega r$
$\omega^2 r$

$$F = -\frac{m\,v^2}{r} = -\frac{Ze^2}{4\pi\varepsilon_0 r^2} \qquad (5.04)$$

where v is the orbital velocity of the electron. But

$$\overset{W}{E} = \frac{1}{2}mv^2 + V$$

$$= \frac{1}{2}mv^2 - \frac{Ze^2}{4\pi\varepsilon_0 r} \qquad (5.05)$$

Note that from Eqs. (5.04) and (5.05) we have

$$\frac{1}{2}mv^2 = \frac{Ze^2}{8\pi\varepsilon_0 r} = -\frac{1}{2}V \qquad (5.06)$$

so

$$\boxed{\overset{W}{E} = -\frac{1}{2}mv^2 = \frac{1}{2}V = -\frac{Ze^2}{8\pi\varepsilon_0 r}} \qquad (5.07)$$

The minus sign in Eqs.(5.07) means that work must be done to remove the electron. The angular momentum, mvr, is to be quantized, so

$$mvr = nh/2\pi \qquad (5.08)$$

$\frac{Ze^2}{2\varepsilon_0 v^2} = nh$

From (5.04) and (5.08) we get

$$r = \frac{Ze^2}{4\pi\varepsilon_0 mv^2} = \frac{nh}{2\pi mv} \qquad (5.09)$$

whence

$$\boxed{v = \frac{Ze^2}{2\,\varepsilon_0 nh}} \qquad (5.10)$$

$$\boxed{r = \frac{\varepsilon_o n^2 h^3}{\pi m Z e^2} = n^2 r_1} \qquad (5.11)$$

and

$$\boxed{W_n = -\frac{mZ^2e^4}{8\varepsilon_o^2 n^2 h^2}} \qquad (5.12)$$

We have thus obtained the energy states, W_n, for a hydrogenic atom, i.e., hydrogen ($Z = 1$); ionized helium ($Z = 2$); doubly ionized lithium ($Z = 3$), etc. The lowest energy state is given by $n = 1$, and this is the normal, or ground state, of the atoms. The frequency of the radiation emitted in a transition from a state E_{n_1} to a state E_{n_2} ($n_1 > n_2$) is thus given by

$$f_{12} = \frac{1}{h}\left(E_{n_1} - E_{n_2}\right) = \frac{mZ^2e^4}{8\varepsilon_o^2 h^3}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) \qquad (5.13)$$

It is customary to measure frequencies in wave numbers (reciprocal wavelengths) and when so measured we shall denote them by $\tilde{f}$. Thus

$$\tilde{f} = \frac{1}{\lambda} = \frac{f}{c} \qquad (5.14)$$

so that if we put

$$R = \frac{me^4}{8\varepsilon_o^2 h^3 c} \qquad (5.15)$$

we shall have, for the case of hydrogen ($Z = 1$)

$$\boxed{\tilde{f} = Z^2 R\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)} \qquad (5.16)$$

Now just exactly this formula with $n_2 = 2$ and $n_1 = 3, 4, \ldots$ had been found empirically by Balmer to fit the lines in the so-called Balmer series in the spectrum of atomic hydrogen. Balmer found that the constant R (the Rydberg constant) had the value

$$\boxed{R_H = 10{,}967{,}758 \text{ m}^{-1}} \qquad (5.17)$$

If we put the usual values of the constant in Eq. (5.15) we obtain

$$R = \frac{9.107\times 10^{-31}\times (1.601\times 10^{-19})^4}{8\times(8.85\times 10^{-12})^2(6.610\times 10^{-34})^3\times 2.9978\times 10^8}$$

$$= 1.0974\times 10^7 \ \mathrm{m}^{-1} \qquad (5.18)$$

which is a remarkable argument. The corresponding spectrum of ionized helium ($Z = 2$) would be given by:

$$\tilde{f} = 4R\left(\frac{1}{1^2} - \frac{1}{n^2}\right) \quad n = 2,3,4,\ldots \qquad (5.19)$$

Empirically we find

$$R_{He} = 10{,}972{,}226 \ \mathrm{m}^{-1} \qquad (5.20)$$

which is slightly greater than R_H. This, too, is understandable when we recall that we assumed the atomic nucleus to be fixed. Actually, both nucleus and electron move about their common center of mass and the effect of this is to replace m in all of the above equations by

$$m' = \frac{mM}{M+m} \qquad (5.21)$$

where m' is the reduced mass of the electron and depends on M. With this refinement, theory and experiment are in almost perfect agreement.

Equation (5.11) tells us the "size" of a hydrogenic atom. For hydrogen in the ground state, i.e., $Z = 1$ and $n = 1$ we get

$$\boxed{r = \boxed{\frac{\varepsilon_0 h^2}{\pi m e^2}} = \frac{8.85\times 10^{-12}\times(6.61\times 10^{-34})^2}{\pi\times 9.1\times 10^{-31}\times(1.60\times 10^{-19})^2}}$$

$$= 5.27\times 10^{-11} \ \mathrm{m} = .527 \ \mathrm{A} \qquad (5.22)$$

as the radius of the first Bohr orbit in hydrogen. Thus a normal hydrogen atom has a diameter of about one angstrom.

Returning to Eq. (5.16) we see that a number of other spectral series are possible. Some of these are

Lyman Series	$n_2 = 1;\ n_1 = 2,3,4,\ldots$
Paschen Series	$n_2 = 3;\ n_1 = 4,5,6,\ldots$
Brackett Series	$n_2 = 4;\ n_1 = 5,6,7,\ldots$

Note that for each of these series the frequencies of successive lines in the series come closer and closer together as n_1 increases, until we reach the series limit corresponding to $n_1 = \infty$. It is convenient to represent the term values, or energy levels, on an energy level diagram, Fig. (5.1).

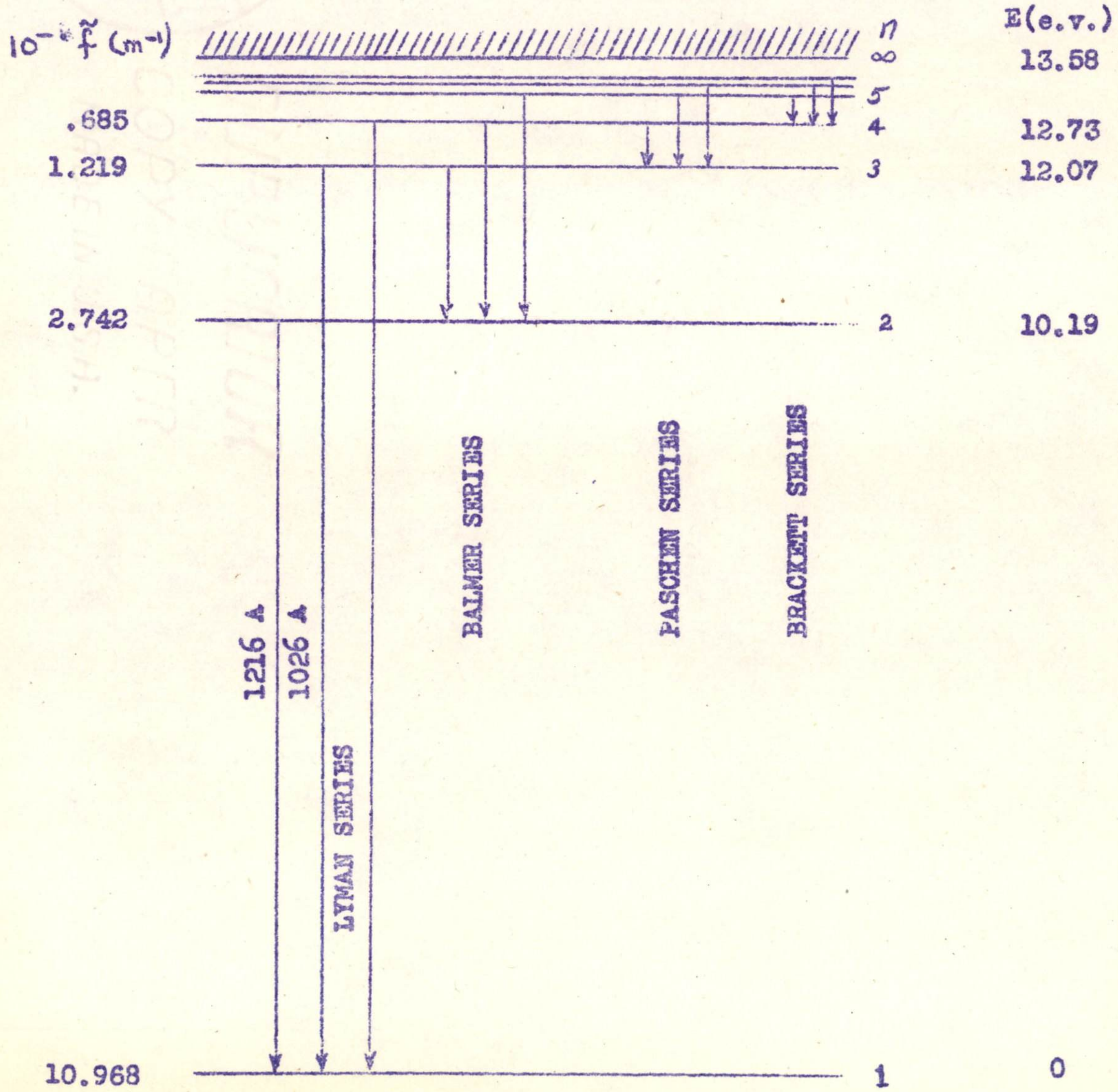


Fig. 5.1
ENERGY LEVEL DIAGRAM FOR HYDROGEN

The numbers on the left side of the diagram represent the term values in wave number units, and on the right the corresponding energy is given in electron volts measured from the ground state. The quantum number, n, refers to the corresponding orbit. The emission of a spectrum line then corresponds to a transition downward from one energy level to another. E.g. the Hα line of the Balmer series corresponds to the transition from the state $n=3$, $-1,218,600$ m^{-1} (-1.508 ev) to the state $n=2$ $-2,742,000$ m^{-1} (-3.394 ev) and the wave number of this line is then $1,523,400$ m^{-1} (1.886 ev). The wavelength is given by ($1A = 10^{-10}$ m).

$$\lambda = \frac{1}{\tilde{f}} = 6.583 \times 10^{-7} \text{ m} = 6583 \text{ A} \qquad (5.23)$$

Other lines in the various series of the hydrogen spectrum are indicated in Fig. 5.1.

The treatment of the hydrogen atom given above is overly simplified, but it does bring out the fundamental soundness of Bohr's theory. Among the aspects neglected is the possibility of other than circular orbits. The general orbit of a particle of negative energy in an inverse square attractive field is an ellipse. It was shown by Sommerfeld that essentially the same energy levels occur for elliptic orbits unless the relativistic variation of the mass of the electron is taken into account. If this is done the energy levels are split, giving rise to a "fine structure". We will not go further into this subject at the present time.

The Bohr theory has many weaknesses. We assume that the laws of classical mechanics hold for an electron in one of its stationary states, but we subject these same laws to the arbitrary postulate of quantization of the angular momentum. Although the theory can be made to give a satisfactory explanation of the spectra of atoms with a single radiating electron, and thus includes in addition to the hydrogenic spectra, the spectra of the alkalis Li, Na, K, etc. where the single valence electron is more or less loosely bound and on the "outside" of the atom, it fails completely for the case of the second most simple atom.

namely helium, where it definitely predicts incorrect values for the energy levels. The resolution of this difficulty requires a complete revision of the laws of atom mechanics into the formal structure of quantum mechanics, or wave-mechanics. We shall not go into this matter other than to point out that a wave-mechanical treatment leads to the conclusion that a particle of momentum p has associated with it a wavelength Λ given by

$$\Lambda = \frac{h}{p} \qquad (5.24)$$

This implies that a beam of electrons, for example, can be diffracted by a grating or crystal in the same way that light or x-rays can. Such indeed is the case. It is interesting, though perhaps not too fundamental, to note that the Bohr quantum condition for a circular orbit, Eq. (5.08), can be written in the form:

$$\boxed{2\pi r = \frac{nh}{mv} = n\Lambda} \qquad (5.25)$$

so that the electron forms a standing wave around the orbit.

A more detailed treatment of atom mechanics would lead us to the conclusion that we must assign not one but four quantum numbers to each orbital electron. The number n used above is the principal quantum number, and is unrestricted as to (positive) value. The other numbers, which we shall not discuss here, do not have such freedom; but the point of importance is that no two electrons in a given atom can have identical sets of quantum numbers. This is known as the "Pauli's Exclusion Principle" and is of great importance in understanding the electronic structure of the elements. As a consequence of this principle it turns out that the electrons form a series of groups ,or shells, about the atomic nucleus. The innermost shell can contain only two electrons, each with $n = 1$ and with spin quantum numbers (i.e., intrinsic angular momentum) of opposite sign. The next shell, $(n = 2)$ can contain 8 electrons; the next $(n = 3)$ 18 electrons, and so on. In addition, the electrons in a shell of given n fall into sub-groups. If the outer shell of an atom is filled with electrons then the atom is chemically inert (the noble gases); i.e., it is difficult to remove an electron from a closed shell. This formation of closed shells leads to a "periodicity" in atomic properties, so that the elements can be arranged in the familiar periodic table. Since we are not primarily concerned with atomic structure in this course we shall not go into further detail in the subject. The important facts of concern to us are the existence of fixed energy states characteristic of each type of atom, and selection rules which state the possibility or impossibility of transitions between pairs of these states. The case of the hydrogenic atom treated above does not involve any special selection rules unless fine structure is taken into account, as spectrum lines occur corresponding to transitions between all pairs $n_2 \rightarrow n_1$, with $n_2 > n_1$.

For more complicated spectra certain transitions are forbidden, both in emission and absorption. In order for a transition from a higher to a lower state to occur the atom must first be raised from the ground state to an "excited" energy state, or perhaps even ionized by entirely removing the outer electron. The process of excitation or ionization can be accomplished in a number of ways. By electron impact (inelastic collision); by impact with another atom or ion; by absorption of radiation. We shall discuss each of these processes briefly.

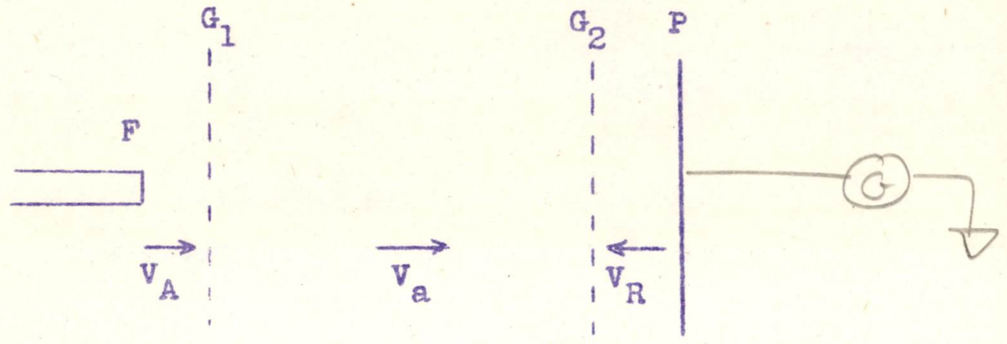
When an electron (or other particle) collides with an atom in a gas the collision is in general one of two kinds. An elastic collision is one in which the kinetic energy of the system atom-plus-electron is conserved. We must of course conserve momentum, in accordance with Newton's laws. The atom is in the same energy state after the collision as it was before. Its kinetic energy will in general be different and so, too, will that of the electron, but the sum is unaltered. Such collisions tell us nothing about atomic energy states. In an inelastic collision kinetic energy is not conserved. Inelastic collisions are of two types. In a "collision of the first kind" kinetic energy of the system is absorbed by the atom, which becomes either excited, i.e., raised to a higher energy state, or perhaps ionized, so that it loses its most loosely bound electron. The relation expressing conservation of energy becomes

$$(\mathrm{K.E.})_{\mathrm{Atom}} + (\mathrm{K.E.})_{\mathrm{elec.}} \rightarrow (\mathrm{K.E.})^{1}_{\mathrm{Atom}} + (\mathrm{K.E.})^{1}_{\mathrm{elec.}} + \Delta E_{\mathrm{atom}} \qquad (5.26)$$

where Δ E is the increase in internal energy of the atom. In a "collision of the second kind" an already excited atom undergoes a collision with an electron, or other particles. As a result of this collision the atom drops to a lower energy state and the kinetic energy of the system is increased by the lost energy of excitation. Momentum is of course conserved, and Eq. (5.26) applies, if the arrow is reversed. Any collision in which an excited atom loses its energy of excitation without the emission of radiation is, in fact, called a collision of the second kind.

A study of energy losses suffered by electrons undergoing inelastic collisions gives us information regarding atomic energy states. For a perfectly elastic collision the energy of the scattered electron is almost the same as that of an incident electron. This is because the mass of the atom is so much greater than that of the electron. If energy, as well as momentum, is conserved in a collision the fraction of the electron's energy transferred to the atom is equal to the mass ratio of electron to atom, neglecting the initial thermal energy of the atom which is quite small. Thus, if the energy difference of an electron before and after an atomic collision is measured the difference must correspond closely to a difference in energy states. This method of measuring atomic energy states directly was demonstrated by Franck and Hertz in a classical experiment carried out with mercury vapor in 1913 shortly after Bohr's first theory was developed. They found that if a mercury atom and an electron undergo a collision, the collision is always elastic as long as the energy of the impinging electron is less than 4.9 ev. But above this value some of the electrons lost 4.9 ev of energy which was taken up

as energy of excitation of the mercury atom. That this is so is shown by the fact that a line of the mercury spectrum is excited at the same time. This is caused by the excited mercury atoms dropping back to the ground state with the emission of a quantum of energy of 4.9 ev (2537A). The method of Franck and Hertz was refined by Franck and Einsporn, and their apparatus is shown schematically in Fig. 5.2. Electrons emitted from the filament F are accelerated through the variable



Apparatus for Measuring Excitation Potentials

Fig. 5.2

potential difference V_A between F and the grid G_1. A small potential difference V_a(~ .1 volt) is applied between the grids G_1 and G_2, so that in this region where most of the collisions occur all the colliding electrons have essentially the same energy. A small retarding potential V_R of about .5 volts is applied between the grid G_2 and the collecting plate P, the current to which is measured by the galvanometer Ga, for various values of V_A. The filament F operates below saturation so that, in the absence of collisions the current to P will increase with V_A. But when V_A just exceeds a critical potential corresponding to an energy transition in an atom of the gas in the tube and if such an electron makes an inelastic collision, the electron will lose almost all of its energy and so will not be collected by P and there will be a sharp drop in the current to P whenever a critical potential is reached. The current will not increase again until V_A exceeds the critical potential plus the retarding potential. The apparatus shown in Fig. 5.2 is capable of resolving critical potentials which are very close together. The table below lists some of the critical potentials for mercury vapour. The corresponding terms values $\left[\boxed{\tilde{f} = \frac{e}{hc} V;}\right.$ and $\frac{e}{hc} = 8.079 \times 10^5$ m^{-1} volt^{-1}] are also listed; these are taken from spectroscopic data and so are much more accurate. Note that critical potentials are measured above the ground state, term values below the ionization potential.

critical potential (volts)	term values (m^{-1})	
10.39	0	ionization potential
8.86	1,238,610	
8.61	1,451,910	
8.53	1,529,450	
8.38	1,631,660	
7.93	2,025,310	
7.73	2,183,080	
6.71	3,011,280	
5.46	4,013,830	metastable
4.88	4,476,890	
4,66	4,653,620	metastable
0	8,417,850	ground state

It will be seen from the table on page 90 that a more careful investigation with higher resolution discloses the lowest excitation potential for mercury vapour to be 4.66 volts, whereas no line in the mercury spectrum corresponds to a transition from this state to the ground state. I.e., a transition from this state to the ground state is "forbidden" by the selection rules. Such a state is called "metastable". A metastable state is a state from which the atom cannot spontaneously drop to a lower state with the emission of radiation. The question then arises as to how an atom, once excited to a metastable state, can ever return to normal, and we see that this can only be done by means of a collision of the second kind, or else by further excitation, followed by an allowed transition to a lower state. The lifetime of a metastable state is thus much longer than that of an ordinary excited state and further, depends rather strongly on the pressure (mean free path). Ordinarily, an excited atom returns to its usual state, with the emission of radiation, in about 10^{-9} to 10^{-7} seconds. The life time of a metastable state under usual circumstances is about 10^{-4} to 10^{-3} seconds, since the atom must wait for a collision in order to be able to change its energy of excitation.

Suppose a mercury atom to have been excited to the state of 4.88 ev energy. All it can do is return to the ground state, with the emission of a quantum of energy of this amount, and a single spectrum line of frequency f, or wavelength λ, given by

$$\boxed{\lambda = \frac{c}{f} = \frac{1}{\tilde{f}} = 2537A} \qquad (5.26)$$

is emitted. That is, we can excite this single line of the mercury spectrum by bombarding mercury vapour with a beam of electrons of energy greater than 4.9 ev but less than 5.46 ev (the next excited state). Similarly, by proper choice of energy of bombarding electrons we can arrange it so that only a certain few lines of the complete spectrum are emitted. Of course, the possibility of multiple collisions does not preclude further excitation, but except for further excitation from a metastable state the probability of this occuring is extremely small, on account of the very short mean lifetime of an ordinary excited state. For example, suppose the bombarding electrons have an energy of 5 ev. A mercury atom excited to the state of 4.88 ev energy would immediately drop back to the usual state and emit the 2536A line, before it could undergo a further collision and be excited to a state of energy between 4.88 ev and (4.88+5) ev; but an atom excited to the metastable state of energy 4.66 ev could very well wait around long enough to suffer an additional inelastic collision and be further excited (but not ionized, since $10.39 > 9.88$). Since the 5.46 ev energy level is also metastable we see that some ionisation will occur whenever the bombarding electrons have an energy exceeding this value, since $10.39 - 5.46 < 5.46$. When the available energy of the bombarding electrons exceeds the ionisation energy then in general all lines of the spectrum are emitted copiously. However, an ionized atom can be further excited (or ionised) by raising one of the remaining electrons to an excited state, i.e., a state of higher energy than that in which it normally finds itself. We shall not be much concerned with multiple ionisation, or spectra of ionised atoms and so will not pursue the matter further.

Another method of exciting or ionising atoms is by impact with positive ions, or in fact by bombardment with particles of any kind. All of the energy considerations discussed above will still apply, of course, but the probability of such processes taking place is much less than in the case of electrons of the same energy. Thus if the fundamental process under investigation involves electron bombardment, any positive (or negative) ions so formed in the process will play a minor role in the overall picture. The reason for this is because the probabilities of excitation or ionization depend more nearly on the velocity of the impacting particles rather than on the energy. Thus an electron and a positive ion of the same velocity will have roughly the same probability of ionizing by collision.

When an electron undergoes an inelastic collision with an atom it can transfer a part of its energy to the atom, thereby exciting it, and carry away the rest of the energy. In view of the well-established photon (quantum) nature of the radiation process, as well as of the photoelectric effect, it might be thought that essentially the same sort of behaviour would occur when an atom undergoes a collision with a quantum, i.e., is exposed to radiation. Thus it might be thought that any quantum having an energy greater than 4.66 ev could excite a amercury atom, transferring this much of its energy, the remainder going on as a quantum of lower energy. While such a process is energetically possible the probability of its occuring is practically negligible unless the quantum loses exactly all of its energy. Thus excitation of mercury to the 4.88 ev state by absorption of radiation can take place only by absorbing a quantum of mercury radiation, 2536A, emitted by another mercury atom on dropping from the 4.88 ev state to the ground state (or by absorbing the same energy quantum from a continuous spectrum). Thus mercury vapour will strongly absorb the 2536A line of the mercury spectrum, thereby exciting the mercury atoms which then re-emit this same radiation, which is therefore called resonance radiation. Similarly, the 1849A line in the mercury spectrum is a resonance line, since it is emitted by a transition from the 6.71 ev energy level to the ground state, and so will be strongly absorbed by normal mercury vapour. In general, any spectrum line which is emitted by an atom in a transition from any excited state to the ground state is a resonance line for that atom, and will be strongly absorbed by vapour of that atom. On the other hand, mercury vapour, for example, is quite transparent to radiation emitted by some other atom (as a line spectrum). If a vapour be exposed to a beam of radiation possessing a continuous spectrum, it will strongly absorb those wavelengths corresponding to its own resonance emission lines. Absorption of other than resonance lines requires the presence of already excited atoms and the probability of the process occuring is correspondingly less.

If the energy of the incident quantum is greater than the ionisation energy then the process is a bit different. The quantum is still absorbed in its entirety, with consequent ionization. The energy of the ejected electron is given by the Einstein photoelectric equation

$$\frac{1}{2}mv^2 = hf - W_i \qquad (5.27)$$

where W_i is the energy required to ionize the atom, e.g., 10.39 ev for the outermost electron in the case of mercury vapour. Photoelectric ionization can occur, however, from any energy state of the atom, provided the incident quantum has the requisite energy. The atom then may emit a more complicated spectrum, corresponding to transitions between the deeper energy levels. The x-ray photoelectric effect falls within this category. Of course, inner electrons can also be ejected by electron bombardment if we use an incident beam of high energy electrons. The method described for determining excitation potentials does not distinguish between excitation and ionization, since all it measures is the energy loss of the impacting electron. However, a slight modification in the method of using the same apparatus easily distinguishes between the two processes. This is illustrated in Figure 5.3.

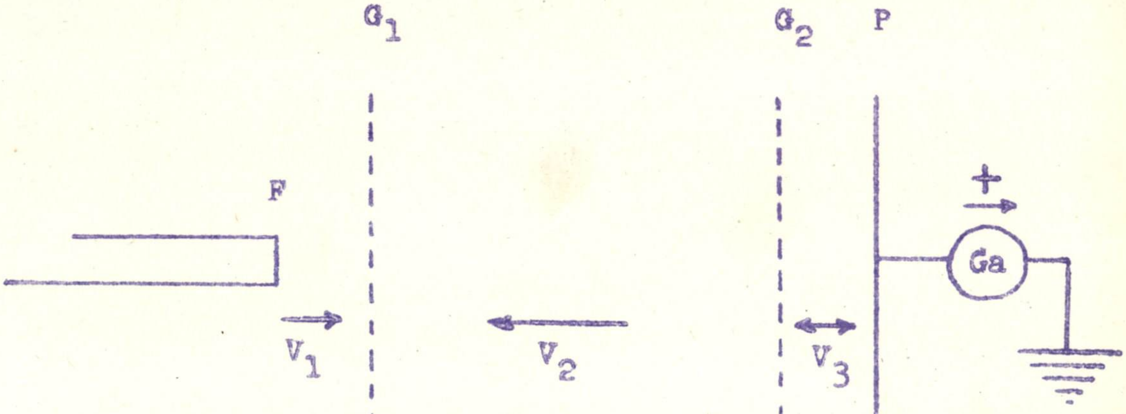


Figure 5.3
Apparatus for Determining Ionization Potentials

V_1 is the (variable) accelerating potential.
V_2 is a retarding potential, greater than V_1.
V_3 is a small constant potential whose direction can be reversed.
The pressure is such that the mean free path of the electrons is larger than the space between G_2 and P, but smaller than the separation of G_1 and G_2. The galvanometer current is determined as a function of V_1 for each direction of V_3 and a curve such as that shown in Figure 5.4 is obtained. Since $V_2 > V_1$ electrons leaving F cannot reach G_2, and so do not reach P. The current

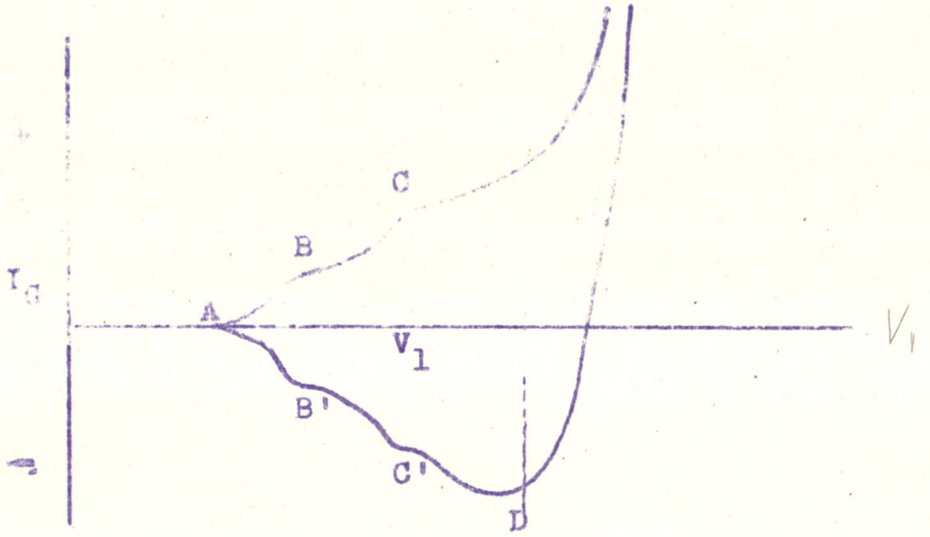


Figure 5.4

is zero for either direction of V_3. But if V_1 is great enough to permit an inelastic collision exciting an atom of gas to a resonance level (not a metastable level) resonance radiation will be excited (point A). This radiation may excite photoelectrons from the grid G_2 and the plate P. With V_3 acting $G_2 \rightarrow P$ so that P is at the higher potential, electrons from G_2 reach P, and the galvanometer current is ~~negative~~. If V_3 acts $G_2 \leftarrow P$ photoelectrons from P travel toward G_2 and the galvanometer current is ~~negative~~ +. This is the portion from A to B or B'. Excitation of a second level will cause radiation of another wavelength to be emitted so that there will be a change in shape of the current curve, BC, BC', and so on. Suppose, however, that an atom is ionized. The original electrons from

$$Al = V$$

$$\frac{n}{V} = N = \frac{A}{V\sigma} \qquad n = \frac{A}{\sigma}$$

$$\frac{1}{\sigma} \qquad l = \frac{V}{A} = V$$

$$V = Al$$

$$\boxed{n = \frac{A}{6}}$$

$$N = \frac{n}{V} = \frac{A}{6V}$$

$$l = \frac{v}{}$$

F as well as the electrons removed in the ionization process will be returned toward F but any positive ions formed will be accelerated toward P and so will produce a positive current. If V_3 is small enough the magnitude of the current will not change much on reversing V_3 and a sharp break occurs at D, the ionization potential.

So far we have discussed possible processes for exciting or ionizing an atom. An equally important question is: what is the probability of each of these processes taking place. Consider the case of electron bombardment of an atom. If the energy of the incident electron is less than the energy of the lowest excited state then all collisions are perfectly elastic and no excitation or ionization is possible. IF the energy of the electron just exceeds the critical energy necessary for excitation then it is quite probable that such an inelastic collision will take place. Although any greater energy can also cause this same excitation, the probability decreases with increasing energy, until at very high energies this process is not likely to occur to any extent. The probability of a given type of collision occuring is expressed quantitatively in terms of the collision cross section. Let us suppose that P_c is the probability that a particle will make a given type of collision in traveling a unit distance. We recall from kinetic theory that the mean free path of a particle is given by

$$\boxed{l = \frac{1}{N\sigma}} = 1/P_c \qquad (5.28)$$

where N is the number of particles per unit value and σ the cross-sectional area of the particle. We also recall that the value to use for σ depends on the process involved, i.e., the cross section is different for different processes. Now the mean free path is the mean distance a particle travels between successive collisions. Hence, the mean number of collisions per unit path length, P_c, is just the reciprocal of the mean free path. Thus

$$\boxed{P_c = \frac{1}{l} = N\sigma} \qquad (5.29)$$

If we know σ, and the number of particles per unit volume then P_c is determined. Actually, equation (5.29) is usually used to determine σ from measured values of P_c. We recall, too, from kinetic theory that if ℓ be the mean free path for a given particle in a gas, e.g., an electron traveling through any gas, then the fraction of particles which travel a distance x <u>without</u> making a collision is given by

$$\frac{N}{N_o} = e^{-x/\ell} \qquad (5.30)$$

If we understand P_c to mean the number of collisions per unit path length at unit pressure and recall that ℓ is inversely proportioned to the pressure, then we shall have

$$\frac{1}{\ell} = p_o P_c \qquad (5.31)$$

where p_o is the pressure, measured in the same units as those used in defining P_c. (It is usual to specify P_c per centimeter of path at a pressure of 1 millimeter of mercury.) Equation (5.30) then becomes

$$N = N_o \, e^{-p_o P_c x} \qquad (5.32)$$

By measuring the fraction of particles which have not undergone a given type of collision, as a function of the distance traveled we can determine P_c, and knowing the pressure the collision cross section σ is thus determined. From the way in which P_c is defined it follows that the probability of making <u>any</u> type of collision is the sum of the probabilities of making each of the various <u>possible</u> types of collision. For example,

$$P_{total} = P_{elastic} + P_{inelastic} \qquad (5.33)$$

where $P_{inelastic}$ can be further broken up into the various possible types of inelastic collisions, excitation to various levels, ionization, etc.

A method for determining P_c for electrons (or other charged particles) in a gas is illustrated in Figure 5.5. Electrons emitted from the cathode, K, are accelerated through

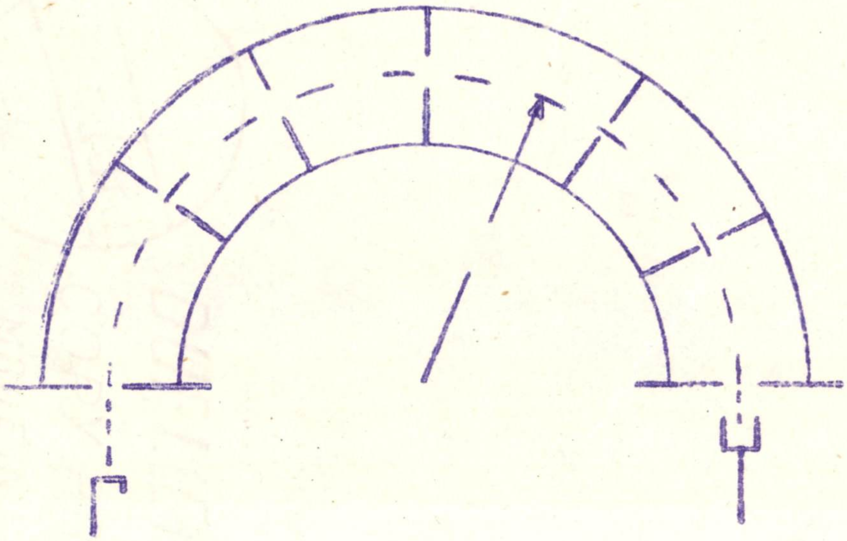


Fig. 5.5
Apparatus for Determining Collision Probabilities

a potential difference V_A to the anode A. The space between K and A is evacuated, so that practically no collisions can occur in this region. Some of the electrons pass through the slit s_1, into the collision chamber containing the gas under investigation. A uniform magnetic field B, perpendicular to the plane of the diagram, is of such a value as to cause electrons of energy eV_A to be bent into an arc of radius R. The baffles b prevent electrons of other energy from reaching the collector C through the slit s_2. Further, any electrons even though of the proper energy, will not reach s_2 if they are at any point of the path deviated from the proper directinn. Thus, if an electron makes a collision, elastic or inelastic, after leaving s_1 it will not reach s_2. The current to C, therefore, measured the number of electrons which have traveled a distance $2\pi R$ without making a collision. This method enables us to determine the total collision cross section as a function of electron energy and gas pressure. Further refinements are necessary in order to determine separately the various contributions to the total cross section.

5.2 Some Additional Facts of Electron Physics

A. The Compton Effect

We have seen that fundamental particles can have both corpuscular and wave properties, and that electromagnetic radiation can have cropuscular (quantum) aspects as well as wave properties. This is not a contradiction;;rather modern physics states that both fundamental particles and electromagnetic radiation exhibit their corpuscular aspects on emission or absorption, and their wave aspects in transmission from source to absorber. Although earlier in these notes we treated particles as corpuscles in discussing their transmission we were working in the same approximation that one makes in applying geometrical optics to light transmission. It is only occasionally that the physical optics of particles becomes important.

In treating collision phenomena one may use either the corpuscle or the wave aspect. A. H. Compton in 1923 gave one of the clinching proofs of the inherent quantum nature of light by considering the effect of the collision of a light quantum (hf) and an electron. Before developing the theory of this effect, let us consider the momentum associated with a beam of light.

If n quanta of frequency f are absorbed by a material body (carbon black, for example) then this represents an energy transfer of nhf. The rate of doing work is then $\frac{d}{dt}(nhf)$ and this should equal CF, where F is a force we associate with this process. Then

$$\boxed{\frac{d}{dt}(nhf) = CF = C\frac{dp}{dt}}$$

from which the momentum of the light beam is nhf/c and the momentum of each quantum is hf/c. This effect was predictad classically and explains the experimentally known fact of radiation pressure. (The fact that a comet's tail always points away from the sun is explained by radiation pressure).

Let us first see if a free electron can absorb a quantum. The laws of conservation of energy and momentum state, assuming the electron is initially at rest,

$$mc^2 - \cancel{m_oc^2} = hf$$

$$mv = hf/c$$

dividing the first by the second, we have,

$$\frac{c^2}{v}(1 - \frac{m_o}{m}) = C$$

or

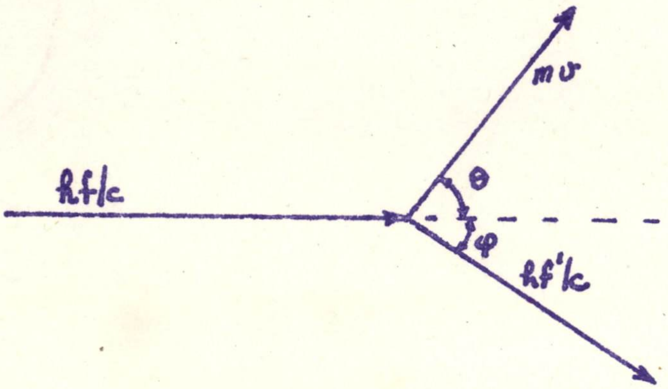
$$1 - \sqrt{1 - v^2/c^2} = \frac{v}{c}$$

from which

$$(1 - \frac{v}{c})^2 = 1 - \frac{v^2}{c^2}$$

and this result is obviously impossible.

However, Compton showed that an electron may scatter a quantum hf providing the scattered radiation takes on a slightly different frequency f'. Consider the collision problem in Fig. 5.6.



$\frac{hf}{c} = mv + \frac{hf}{c}$

$\frac{1}{2}mv^2 =$

Fig. 5.6

The conservations laws are:

$$\left.\begin{aligned} hf + m_o c^2 &= hf' + mc^2 \\ hf/c &= hf'/c \cos \phi + mv \cos \theta \\ hf'/c \sin \phi &= mv \sin \theta \end{aligned}\right\} \qquad (5.34)$$

From the first of equations (5.34) we have,

$$\frac{1}{\sqrt{1 - v^2/c^2}} = 1 + \frac{hf - hf'}{m_o c^2}$$

or setting $\varepsilon = \dfrac{hf - hf'}{m_o c^2}$

$$\frac{1}{\sqrt{1 - v^2/c^2}} = 1 + \varepsilon$$

from which, (5.35)

$$v = \frac{c}{1+\varepsilon} \sqrt{\varepsilon^2 + 2\varepsilon}$$

Now in order to confirm any conjectures we may desire from theory by experiment we must consider which variables we can most easily observe. These variables are f' and φ. Hence, we proceed to eliminate v and θ from our equations.

Substituting (5.35) in the 2nd and 3rd of equation (5.34) and solving for sin θ and cos θ we have,

$$\sin \theta = \frac{hf'}{m_o c^2} \sin \phi \cdot \frac{1}{\sqrt{\varepsilon^2 + 2\varepsilon}}$$

$$\cos \theta = \frac{hf - hf \cos \phi}{m_o c^2} \cdot \frac{}{\sqrt{\varepsilon^2 + 2\varepsilon}}$$

Squaring and adding, we have,

$$1 = \left(\frac{h}{m_o c^2}\right)^2 \frac{f^2 + f'^2 - 2ff' \cos \phi}{\varepsilon^2 + 2\varepsilon}$$

Substituting for $\mathcal{E}$ in terms of f and f' and rearranging terms, we have,

$$f - f' = \frac{h}{m_o c^2} ff' (1 - \cos \emptyset) \qquad (5.36)$$

from which

$$f' = f \cdot \frac{1}{1 + \frac{hf}{m_o c^2} (1 - \cos \emptyset)}$$

remembering that $f = c/\lambda$, $f' = c/\lambda'$, we have

$$\lambda' - \lambda = \frac{h}{m_o c} (1 - \cos \emptyset) \qquad (5.37)$$

The constant $h/m_o c$ is equal to 2.4×10^{-12} m or just .024 A, and is usually called the Compton wavelength. Such a shift of wave length is most easily observed in the x-ray region.

B. Magnetic Moment and Electron Spin

Let us consider an electron in a Bohr orbit in a hydrogen atom. Such an electron is equivalent to a current of $ev/2\pi r$ and should give rise to a magnetic moment.

$$\mu = iA = \frac{ev}{2\pi r} \cdot \pi r^2 = \frac{evr}{2}$$

By Bohr's first quantum condition, equation (5.08)

$mvr = \frac{nh}{2\pi}$ $vr = \frac{nh}{2\pi m}$, and hence

$$\mu = \frac{neh}{4\pi m}$$

For the ground state n = 1 and

$$\mu = \frac{eh}{4\pi m} \qquad (5.38)$$

(This expression is often taken as defining an atomic unit of magnetic moment, called the Bohr magneton).

One would expect such a magnetic moment to interact with applied magnetic fields and to shift the atomic energy levels and hence the frequency of the spectral lines. Such a is indeed the case; the Zeeman Effect.

In addition to the angular momentum and magnetic moment associated with the orbital motion of electrons, it is necessary to assume that each electron <u>inherently</u> has an angular momentum (spin) and magnetic moment in order to explain all the facts of atomic spectra. Classically, this is the equivalent of saying that the electron has structure, and that its charge and mass are so distributed throughout this structure as to give rise to the observed values of spin angular momentum and magnetic moment. The inherent magnetic moment of the electron has been found equal to the Bohr magneton, equation (5.38), but the inherent spin angular momentum is always numerically equal to $\frac{1}{2} \cdot \frac{h}{2\pi}$. In other words the spin is quantized but only one value, $\frac{1}{2}$, is allowed. Actually, the spin has two allowed values $\pm \frac{1}{2}$, since, if a magnetic field is applied the inherent electron magnetic moment can either line up with the applied field or against the applied field.

Independently of the spectral data, Stern and Gerlach showed the existence of the electron magnetic moment, and the two allowed states of the spin quantum number, by the famous experiment which bears their name. They allowed a well collimated beam of neutral silver atoms to pass through a region in which an inhomogenous magnetic field existed. (Inhomogenous, since we wish to operate on a magnetic moment, or dipole, which is unaffected by a uniform field). After passing through the field the displacement caused by the field is measured. It was found that the beam was displaced by equal amounts on both sides of the undisplaced position, in accordance with the theory developed from atomic spectra.

C. Quantum Numbers and Pauli Principle.

Since an electron in an atom has three degress of freedom, and in addition has spin angular momentum, we should expect that it requires four quantum numbers to completely specify the (quantised) state of the atom. Pauli's Principle states that no two electrons in a single atom can be in one and the same quantum state, i.e., can have identical sets of quantum numbers. No exceptions to this rule have ever been observed.

Free electrons also have four quantum numbers associated with their three components of momentun and their one component of spin. The extension of the Pauli Principle to free electrons will be dealt with in connection with the Fermi-Dirac Statistics.

D. Heisenberg Uncertainty Principle

The correct quantum mechanical formulation of the laws governing the behavior of atoms and electrons is essentially statistical in nature. That is, we can talk about the probability of an event occurring, but we cannot say with certainty whether any event whith is energetically possible will or will not occur. Heisenberg first pointed out that there is a fundamental uncertainty in any physical measurement we choose to make such that the product of the uncertainty in the position measurement and the uncertainty in the momentum measurement is of the order of Planck's constant h, i.e.,

$$\boxed{\Delta p_x \Delta x \sim h}$$

or in terms of the simultaneous measurement of time and energy

$$\boxed{\Delta W \Delta t \sim h.}$$

Such uncertainty is very small compared to ordinary physical measurement and in general does not effect the accuracy of experiments in any way. However, the implications of this uncertainty principle are very important. For instance, in discussing the natural breadth of spectral lines, or of the stationary states from which they arise, one can write,

$p = mv = m\frac{dx}{dt}$

$w = \frac{dR}{dt} dT$

$\Delta m\left(\frac{dx}{dt}\right)\Delta x \geq h$

$$\Delta W \Delta t \sim h$$

$$\Delta W \sim h \Delta f$$

$$h \Delta f \Delta t \sim h$$

$$\Delta f \sim \frac{1}{\Delta t}$$

Hence, the natural breadth (the uncertainty in the frequency measurement) is of the order of the reciprocal of the uncertainty in the time that the emission takes place (the lifetime of the state). This result is also true for classical radiators although we have no uncertainty principle in classical physics.

The uncertainty principle also has very important implications regarding the behaviour of dense assemblages of free electrons, for instance, the conduction electrons in metals. In fact, we must use this principle in order to understand the behavior of metallic electrons as we shall see later in connection with the Fermi-Dirac statistics.

CHAPTER VI. COLLISION PHENOMENA AND THE GASEOUS DISCHARGE

(Reference: Millman and Seely, Chap. X)

6.1 Collision Phenomena

In dealing with the motion of charged particles under the action of electric or magnetic fields we have assumed that no forces, other than those due to these fields, act on such particles. This means, for example, that collisions do not take place between the particle in question and any other particle. We have excluded this possibility by assuming that the particles move in a "vacuum". But from the point of view of kinetic theory, we need only require that the pressure be so low that the mean free path be large compared with the dimensions of the apparatus. This is an essential requirement for the problems considered hitherto, such as the cathode ray tube, particle accelerating machines, etc. But there are cases where we not only cannot make this assumption but where the occurence of collisions is an essential feature of the problem as in most of the last chapter. With this in mind we now go to the other extreme and suppose the pressure to be such that the mean free path is small compared with the dimensions of the apparatus. In this chapter we shall consider some of the properties of a gas in the conducting state and the production and removal of the ions which cause this conductivity, as well as some of the phenomena of the gaseous discharge.

6.1.1 Ion Mobilities

Consider a particle of charge q, mass m, moving through a gas under the action of an electric field E. The ion will be moving with its random thermal velocity and will make collisions. We are not concerned with the random motion, but only the superposed motion due to the action of the field. It is found experimentally that under such conditions a given type of ion acquires a drift velocity proportional to the field E. In the absence of collisions we know that such an ion would move with constant acceleration. This difference in behavious can be accounted for roughly, as follows. Consider an ion between collisions. The force in this ion is

$$F = Eq \tag{6.01}$$

If this force acts for a time τ the ion will acquire a momentum component given by

$$mv = F\tau = Eq\tau \tag{6.02}$$

where v is the velocity acquired under the action of the field. Let us suppose that this additional velocity is completely lost at each collision. (This is not quite true). If we now let τ be the mean time between collisions then, on the average, the drift velocity acquired by the ion is $\frac{1}{2}v$ and so

$$\frac{1}{2}v = \frac{1}{2}E\frac{q}{m}\tau \tag{6.03}$$

For a given type of ion, moving through a given gas at fixed pressure and temperature, the quantities q, m, and τ are constants. Thus the drift velocity is proportional to the field E, or

$$v' = kE \tag{6.04}$$

The quantity k is called the mobility of the ion, and the above crude calculation shows that it should be of the order of magnitude

$$\boxed{k \sim \frac{1}{2}\frac{q}{m}\tau = \frac{1}{2}\frac{q}{m}\frac{\ell}{v}} \sim \bar{v} \tag{6.05}$$

where ℓ is the mean free path of the ion in the gas and $\bar{v}$ the mean (thermal) speed. This formula does indeed give values for k of the right order of magnitude ($k \sim 10^{-4}$ m/s per v/m) but cannot be expected to be correct in detail.

The concept of the mobility of an ion is important in discussing ionic conduction in a fluid medium. Let us suppose a potential difference V to be applied across a pair of electrodes immersed in the fluid a distance d apart. Let n_j be the concentration (ions per unit volume) of ions of type j, each carrying a charge Q_j and having a mobility k_j. In the presence of an electric field E such ions will acquire a drift velocity v_j, parallel to the field E. In a time dt the charge dQ crossing an area ΔA will be

$$dQ = \Delta A(\sum q_j n_j v_j)\,dt = \Delta A(\sum q_j n_j k_j)E\,dt \qquad (6.06)$$

The current i is thus

$$i = \frac{dQ}{dt} = \Delta A \sum (q_j n_j k_j) \cdot E \qquad (6.07)$$

At a given temperature and pressure k_j is constant and if the ions maintain their identity and type, q_j is constant as far as variations in E, and hence of the applied potential V, are concerned. The behaviour of n_j is, however, subject to other circumstances. In the first place, the passage of a current necessarily means the removal of ions. This means that the n_j depend on the current i, in general. However, if we suppose that the current is so small that the fraction of ions withdrawn is negligible then the n_j are constant if we neglect recombination. Or, if we suppose a uniform source of ionization to be present then the n_j are constant, the constant value being determined by the rate of recombination. If we further suppose that E is uniform, and given by V/d then

$$i = \Delta A(\sum n_j q_j k_j)\,\frac{V}{d} \qquad (6.08)$$

and we see that the current obeys Ohm's law. At higher field intensities the relation between current and voltage becomes much more complicated.

6.1.2 Recombination of Ions

Suppose that we have a source of ionizing radiation incident on a gas and let us suppose that it produces P pairs of ions per unit volume, per unit time. For simplicity we will suppose only two kinds of ions produced; that is, we assume a single gas in the ionization chamber and that only one process of ionization occurs. Let n_+ and n_- be the concentrations of free positive and negative ions, respectively.

Collisions will occur, and in particular, when a positive and negative ion collide there is a certain probability that they will recombine and form a neutral molecule. The number of collisions between positive and negative ions is proportional to the concentration of each and so the number of recombinations occuring is proportional to $n_+ n_-$. Hence, if we neglect the effect of all other types of collisions, e.g., ions with neutral molecules, the increase in the number of positive (or negative) ions in time dt is given by

$$dn_+ = dn_- = (P - \theta \cdot n_+ n_-)dt \tag{6.09}$$

where the coefficient θ is called the coefficient of recombination, and depends on the particular ions in question, as well as on the pressure, but is independent of n_+ and n_-. Rewriting equation (6.09) we have

$$\boxed{\begin{aligned} \frac{dn_+}{dt} &= P - \theta\, n_+ n_- \\ \frac{dn_-}{dt} &= P - \theta\, n_+ n_- \end{aligned}} \Bigg\} \tag{6.10}$$

from which it follows that the difference $n_+ - n_-$ is a constant, so that if the gas is initially uncharged we shall have

$$n_+ = n_- = n \tag{6.11}$$

or

$$\frac{dn}{dt} = P - \theta n^2 \tag{6.12}$$

The ionic concentratiin will reach a steady state when $dn/dt = 0$, i.e.,

$$n' = \sqrt{P/\theta} \tag{6.13}$$

where n' is the equilibrium concentration corresponding to the source strength P. If an un-ionized gas be exposed to a source of strength P then the concentration will build up to the equilibrium value n' according to the equation

$$n = n' \frac{e^{2\sqrt{P\theta}\,t} - 1}{e^{2\sqrt{P\theta}\,t} + 1} \qquad (6.14)$$

and so does not approximate to the steady state until $2\sqrt{P\theta}\ t_1 >> 1$, or until

$$t_1 >> \frac{1}{2\sqrt{P\theta}} \qquad (6.15)$$

t_1 is thus a measure of the time required to reach a steady state, and is thus long for a weak source.

Suppose, on the other hand, that a steady state has been reached and that the source is then removed. How does the concentration decay with time? The equation (6.12) now becomes

$$\frac{dn}{dt} = -\theta n^2 \qquad (6.16)$$

with $n = n'$ at $t = 0$. Thus

$$\frac{1}{n} - \frac{1}{n'} = \theta t \qquad (6.17)$$

or

$$n = \frac{n'}{1 + n'\theta t} \qquad (6.18)$$

The extension of these results to cases where more than one type of positive and negative ion is present follows immediately from equation (6.10). We have assumed in this treatment that ions are removed by recombination only and that no external field is present. If the pressure is not too high the coefficient of recombination is roughly proportional to the pressure. Although the value of θ depends on the gas, typical values lie in the range

$$\theta \sim 1\text{-}2 \times 10^{-12}\ m^3/sec \qquad (6.19)$$

at a pressure of one atmosphere. Now we know from kinetic theory that the number of collisions made per unit time by a molecule is given by

$$Z = 2\sqrt{2}\, n\sigma\bar{v} \qquad (6.20)$$

where n is the concentration $\bar{v}$ the mean speed and σ the moleculer cross section. Thus the <u>total</u> number of collisions made per unit volume per unit time is nZ and for such gases as oxygen, nitrogen, etc. this is

$$nZ \sim 1\text{-}2 \times 10^{-16}\, n^2 \qquad (6.21)$$

at room temperature. Hence, collisions between oppositely charged ions are about 10,000 times more frequent than collisions between neutral molecules.

6.1.3 Diffusion of Ions

In addition to being removed by recombination ions can disappear from a gas by diffusion to the walls of the containing vessel, where they may become neutralized. If this happens there will exist a concentration gradient in the vicinity of the walls, so that diffusion will persist. Let Ψ be the flux of particles in question, e.g., positive ions. Ψ is the number of such particles crossing unit area per unit time. Then

$$\Psi_+ = -D_+\nabla n_+ \qquad (6.22)$$

defines the diffusion coefficient D_+ for these ions in the gas in question; D depends both on the kind of ion and the gas through which it diffuses. Consider a volume τ bounded by a surface S. The total flux through the surface is

$$\int_S \Psi\, dS = -D\int_S \nabla n \cdot dS = -D\int_\tau \nabla\cdot(\nabla n)\, d\tau = -D\int \nabla^2 n\, d\tau \qquad (6.23)$$

But the total flux through the surface is just equal to the rate at which the concentration within the volume τ is decreasing. That is

$$\int_S \Psi\, dS = -\int_\tau \frac{dn}{dt}\, d\tau \qquad (6.24)$$

So

$$\int_\tau \frac{dn}{dt}\, d\tau = D \int_\tau \nabla^2 n \, d\tau \qquad (6.25)$$

Since the volume τ is arbitrary we must have

$$\frac{dn}{dt} = D \nabla^2 n \qquad (6.26)$$

which is the differential equation of diffusion. We have assumed that the only cause of a change in n is diffusion. If there is a source of ionization of strength P present, and if recombination be taken into account, then on referring to equation (6.12) we see that equation (6.26) is to be replaced by

$$\boxed{\frac{dn_1}{dt} = P + D \nabla^2 n_1 - \theta\, n_1 n_2} \qquad (6.27)$$

with similar equations for each type of ion.

It is shown in kinetic theory that the diffusion coefficient is related to the mean speed $\bar{v}$ and the mean free path l for the same particle, or ion, through the equation

$$\boxed{D \sim \frac{1}{3} l \bar{v}} \qquad (6.28)$$

and is thus inversely proportional to the pressure. Diffusion coefficients differ for positive and negative ions and from gas to gas, but are of the order of magnitude of $1\text{-}50 \times 10^{-6} m^2/sec$ at a pressure of one atmosphere.

6.1.4 Ionization by Collision

Consider a plane parallel electrode ionization chamber, across which a potential difference V is applied. Suppose that ions are produced at one of the electrodes, e.g. photoelectrons liberated from the cathode by illuminating it with a suitable source of radiation. If now for a given strength source, i.e. a given number of electrons produced per second at the cathode, we measure the current through the ionization chamber as a function of the applied voltage across the plates it will be found that the current-voltage relation resembles that shown in

Fig. 6.1. At low voltages the current is nearly proportional

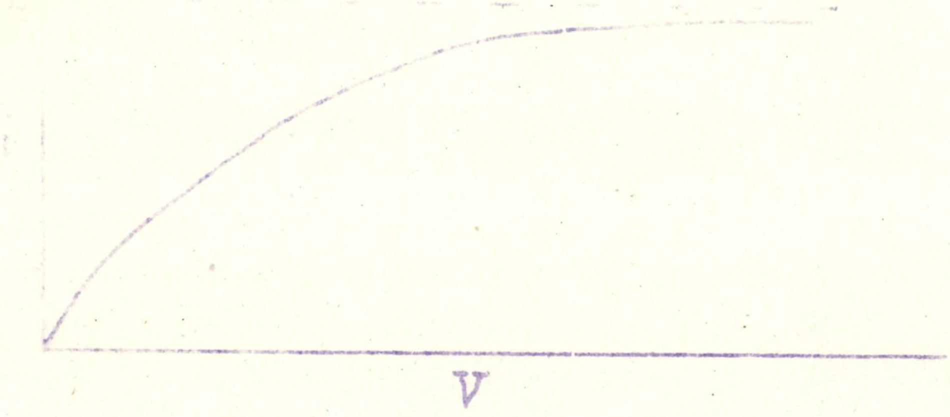


Figure 6.1

Current Voltage Relation in an Ionization Chamber

to the voltage, but as the voltage is increased the current apparently reaches a saturation value independent of V. We wish to determine the current as a function of the applied field and separation of the electrodes. It is found experimentally that for constant field the current increases exponentially with plate separation provided the field is not too great. This was explained by Townsend by assuming that an electron on striking a gas molecule can, if its energy be great enough, ionize the atom thus releasing an additional electron. This process is known as ionization by collision and plays an important role in gaseous conduction. Townsend assumed that each electron in traveling a distance dx produced αdx new ions. The coefficient α is known as the First Townsend Coefficient. Thus, if across a plane at a distance x from the cathode these pass n electrons then a number $n+dn$ will pass across a plane at a distance $x+dx$, where if we neglect recombination and diffusion dn is the number of electrons released by the collision process. According to Townsend's hypothesis we shall have

$$dn = \alpha n dx \qquad (6.29)$$

and integrating

$$n = n_o e^{\alpha x} \qquad (6.30)$$

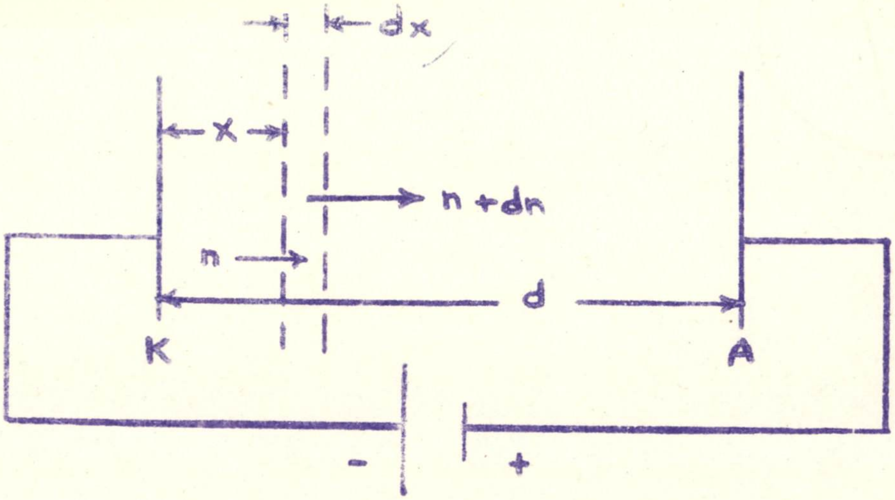


Figure 6.2
Ionization by Collision

On multiplying both sides of equation (6.30) by the electronic charge e we shall thus have for the current reaching the anode

$$\boxed{i = i_o e^{\alpha d}} \qquad (6.31)$$

In this equation i_o corresponds to the electronic current leaving the cathode, and is the value that the saturation current would attain if ionization by collision did not occur. The equation of continuity, of course, requires that the current be the same at all values of x. This means that the positive ion current will so adjust itself that

$$i_+ + i_- = i \qquad (6.32)$$

at each value of x, so that at the anode, for example, the positive ion current vanishes while it is a maximum at the cathode.

The quantity α, for a given gas, depends on the field strength and the mean free path (pressure). Since α is the number of ionizing collisions per unit path length it is, among other things, proportional to the pressure. In addition, it will be some function of the energy of the ionizing particle (electron), i.e. of its velocity. If the energy of the electron is less than the ionization energy we would not expect ionization to occur, since the electron must transfer to the molecule sufficient energy to ionize it. The mean energy acquired by the electron from the field will be $Ee\ell$, where ℓ is the mean free path, and so we should expect α to be of th the form

$$\alpha = p \cdot f_1(Ee\ell) \qquad (6.33)$$

where f_1 is a function which vanishes when the energy is less than the ionization energy. Since ℓ is inversely proportional to the pressure we can also write

$$\boxed{\alpha/p = f(E/p)} \qquad (6.34)$$

so that if α be measured for a given gas at various pressures and for various field strengths we should expect the values to fall on a smooth curve if we plot α/p against E/p. Such is indeed found to be the case, if the field is not too great.

6.1.5 Breakdown

If equation (6.31) is valid we should expect to obtain a straight line of slope α if we plot log i against d, where i is the current for a fixed field strength and gas pressure. This is found to be the case for small values of d, but the current increases more rapidly than equation (6.31) would indicate, for larger values of d until finally, as is well known, the gas breaks down and a spark occurs. This means that there must be some other source of ionization. Townsend at first assumed that the psotive ions formed in the process of ionization by collision were also capable of ionizing by the collision process, and introduced the Second Townsend

Coefficient β, i.e., βdx is the number of ions produced by collision when a single positive ion travels a distance dx through the gas. Although this assumption leads to an equation which, in a formal way, accounts for the observed results it is now known from direct measurement that the quantity β is negligible at the energies involved in the usual cases. That is, ionization by collision of positive ions is negligible in most practical cases. This is because, as we have seen in chapter V, the probability of ionization depends more on the velocity of the ionizing particles than on its energy. A much more likely process is the liberation of secondary electrons at the cathode by positive ion impact. Let us suppose that, on the average, each positive ion striking the cathode releases γ secondary electrons (γ will usually be a small fraction). As before, let us suppose that n_o electrons are released from the cathode by the primary process, e.g., n_o photoelectrons, per unit time. These electrons, under the action of the applied field, will move toward the anode, ionizing by collision and thus producing positive ions. The positive ions will migrate toward the cathode where they will release secondary electrons. Let n_1 be the number of electrons arriving at the anode per unit time. Each of the n_o electrons will ionize by

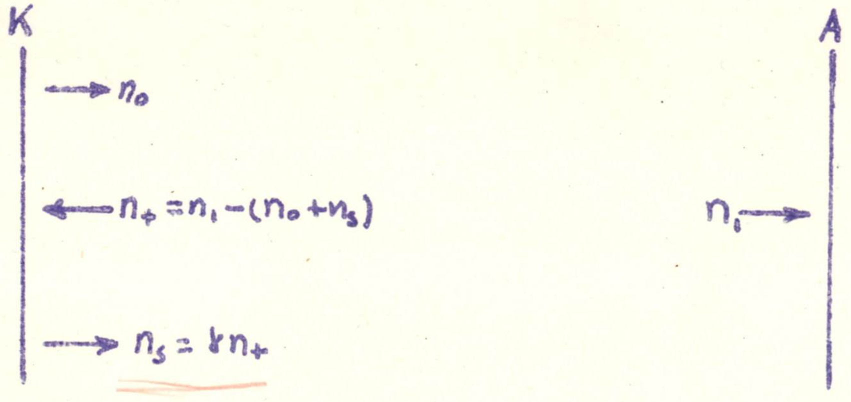


Figure 6.3
The Effect of Secondary Electrons

collision, and on this account $n_o e^{\alpha d}$ electrons will reach the anode. In this process $n_o(e^{\alpha d} - 1)$ positive ions will be produced and on arriving at the cathode will release $\gamma n_o(e^{\alpha d} - 1)$ secondary electrons, which will also ionize by collision, thus supplying $\gamma n_o(e^{\alpha d} - 1)e^{\alpha d}$ additional electrons at the anode. There will thus be an additional number $\gamma n_o(e^{\alpha d} - 1)^2 e^{\alpha d}$ of positive ions reaching the cathode, etc. Hence, the total number of electrons reaching the anode is

$$n_1 = n_o \sum \gamma^j (e^{\alpha d} - 1)^j \cdot e^{\alpha d}$$

$$\boxed{= \frac{n_o e^{\alpha d}}{1 - \gamma(e^{\alpha d} - 1)}} \qquad (6.35)$$

Alternatively, referring to Figure 6.3, if n_+ be the <u>total</u> number of positive ions reaching the cathode

$$n_+ = n_1 - (n_o + n_s) \qquad (6.36)$$

where n_s is the <u>total</u> number of secondary electrons from the cathode. By definition of γ

$$n_s = \gamma n_+ \qquad (6.37)$$

and so

$$n_s = \gamma\left[n_1 - (n_o + n_s)\right] \qquad (6.38)$$

But

$$n_1 = (n_o + n_s)e^{\alpha d} \qquad (6.39)$$

since <u>all</u> electrons leaving the cathode surface, whether primary or secondary, ionize by collision. On combining equation (6.38) and (6.39) we arrive at equation (6.35).

Equation (6.35) indicates that when

$$1 - \gamma(e^{\alpha d} - 1) = 0 \qquad (6.40)$$

the current becomes "infinite", i.e. the current is limited only by the resistance in the external circuit. Physically, this means that a spark occurs. Let us denote the sparking potential by V_1. The field is thus V_1/d. We have seen that α is of the form

$$\alpha = pf(E/p) = pf(V_1/pd) \qquad (6.41)$$

Further, γ will depend on the energy which the positive ion possesses when it strikes the cathode. If we assume an ion to lose all of its energy at each collision then we need only consider the energy acquired in the last mean free path before reaching the cathode. This is equal to $Ee\ell$. Thus

$$\gamma = F_1(Ee\ell) = F(E/p) = F(V_1/pd) \qquad (6.42)$$

Thus equation (6.40) is of the form

$$F(V_1/pd)\left[e^{pdf(V_1/pd)} - 1\right] = 1 \qquad (6.43)$$

Equation (6.43) is an implicit relation between V_1, the sparking potential, and pd where d is the spark length at a pressure p for this potential difference. If we assume equation (6.43) solved for V_1 we shall have

$$\boxed{V_1 = \phi(pd)} \qquad (6.44)$$

Thus, V_1 depends on the product pd only, and not on either alone. This fact is known as Paschen's Law and is well established experimentally. Physically, Paschen's Law means that the sparking potential is a function only of the total number of molecules in the space between the electrodes. The form of the function ϕ is similar to that shown in Figure 6.4.

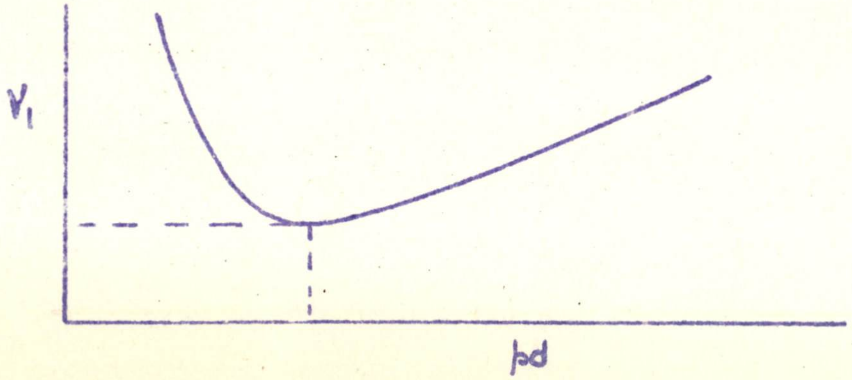


Figure 6.4
Breakdown Potential vs. pd.

The breakdown potential is high at very small or very large values of pd and reaches a minimum somewhere in between. The right hand portion of the curve is nearly linear. Qualitatively this sort of behaviour is understandable. The condition for breakdown is that just enough positive ions must be formed, as a result of ionization by collision, to produce one secondary electron from the cathode by ion impact, for each primary electron. That is,

$$\gamma n_o(e^{\alpha d} - 1) = n_o \qquad (6.45)$$

This is equivalent to equation (6.40). Now at very low values of pd there are so few molecules present that a high field intensity is necessary in order to produce the requisite number of secondary electrons. At the other extreme, for large values of pd the energy gained per mean free path is small unless the field is high. Notice that, regardless of the value of p or d it is impossible to produce breakdown below a certain minimum *voltage* regardless of the *field* intensity. There is evidence that this is not true at extremely small values of d. For air this minimum breakdown potential is about 340 - 350 volts and occurs at a value of pd ~ .6 if p is measured in millimeters of mercury and d in centimeters. The accompanying table gives data on the minimum for a number of gases. Accurate determination of these values is extremely difficult.

Table 6.1

Sparking Data

Gas	Minimum Sparking Potential (volts)	(pd) minimum mm of Hg x cm Spark
Air	325 - 350	.56
Hydrogen	275 - 310	1.2 - 1.4
Oxygen	440 - 460	.72 - .48
Nitrogen	250 - 300	.67 - .75
Helium	150 - 260	4.0 - 2.7
Argon	140 - 230	.9 - .76
Carbon Dioxide	~ 420	~ .51
Sulfer Dioxide	~460	~ .33
Nitrous Oxide	~420	~ .50
Hydrogen Sulfide	~415	~ .60

6.2 The Gaseous Discharge

The phenomenon of the passage of an electric current through a gas between two electrodes across which a potential difference is applied is known as a gaseous discharge. For a given gas, pressure, and geometrical configuration an important characteristic of the discharge is the voltage-current curve which exhibits some remarkable features. A typical voltage-current curve for a low pressure discharge is shown in Fig. 6.5.

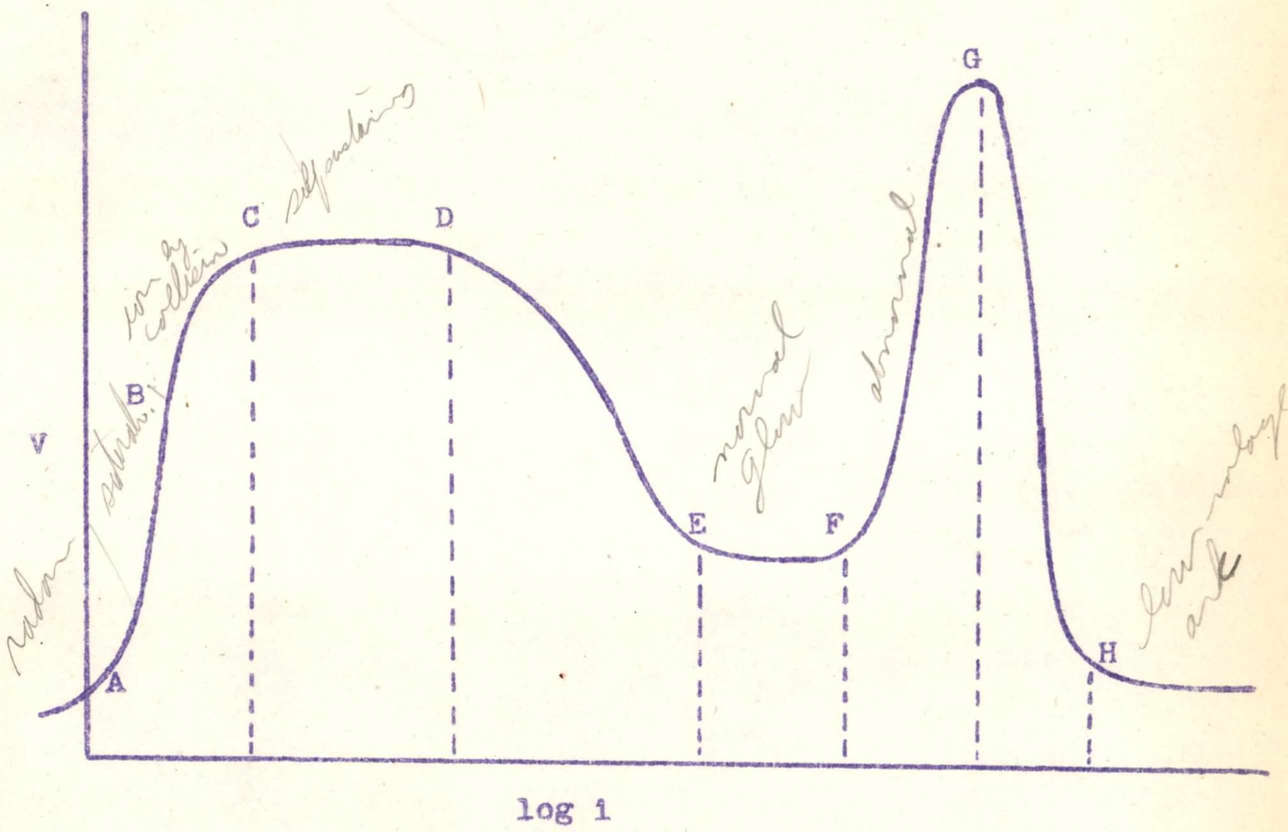


Figure 6.5
Voltage-Current Characterisc of a Discharge Tube

Since for moderate ranges of potential difference across the electrodes the current ranges through many orders of magnitude it is convenient to use a logarithmic scale for the current.

At very low voltages the current is small, gradually approaching an (apparent) saturation value in the region AB (Figure 6.5). Beyond here ionization by collision becomes appreciable until, at C, we reach the sparking potential and the discharge becomes self-maintained, as we have seen previously. In the region C-D the voltage drop across the tube is practically independent of the current. At higher currents the voltage decreases until we reach the region of the normal glow discharge (E-F). In this region the voltage drop is again independent of the current, which is limited by the resistance of the external circuit. It is in this region that glow discharge voltage regulating tubes (VR-105, VR-150, etc.) operate. Just beyond this region the voltage rises rather steeply (F-G) in the region of the abnormal glow, and then quickly drops to a very low value in the region of the low voltage arc, beyond H.

6.2.1 The Normal Glow Discharge

The ordinary low pressure gas discharge tube operates in the region of the normal glow. This region is characterized by a low current density and a moderately high voltage drop. The region between the cathode and anode is broken up into a number of rather definite sections which are readily distinguished visually if the separation of electrodes is not too small. These are, starting at the cathode, 1) the Aston dark space, adjacent to the cathode and very narrow; 2) the cathode glow, which appears to be "floating" on the cathode surface since the Aston dark space is so narrow; next to the cathode glow is the Crookes dark space, 3) whose width increases with decreasing pressure and depends somewhat on the tube current. The luminosity in this region is much less than in the immediately following negative glow, 4) which has a rather sharp boundary on the side toward the cathode and gradually fades off into the next region, 5) called the Faraday dark space, or "second negative dark space", where the luminosity is very low. Following this is 6) the positive column or plasma whose length depends almost solely on the length of the discharge tube. This region extends practically to the anode, from which it is separated by the anode dark space, which is a very narrow region of relatively low luminosity. For certain ranges of current and pressure the positive column breaks up into a series of bright and dark bands, called "striations".

The variations in luminosity along the length of the discharge are accompanied by variations in the electric field intensity. If the distribution of the electric field is known then the potential distribution and the distribution of (total) free charge can be determined. An ingenious method for determining the electric field distribution, due to J. J. Thomson and used by Aston, is illustrated in Figure 6.6. The

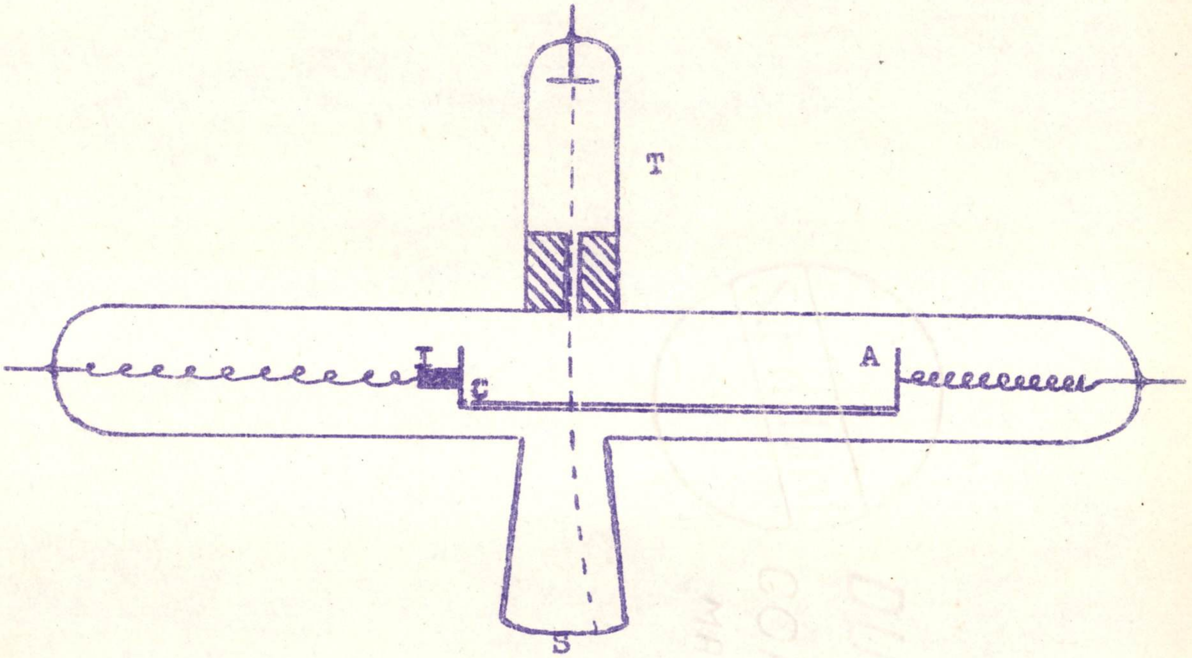
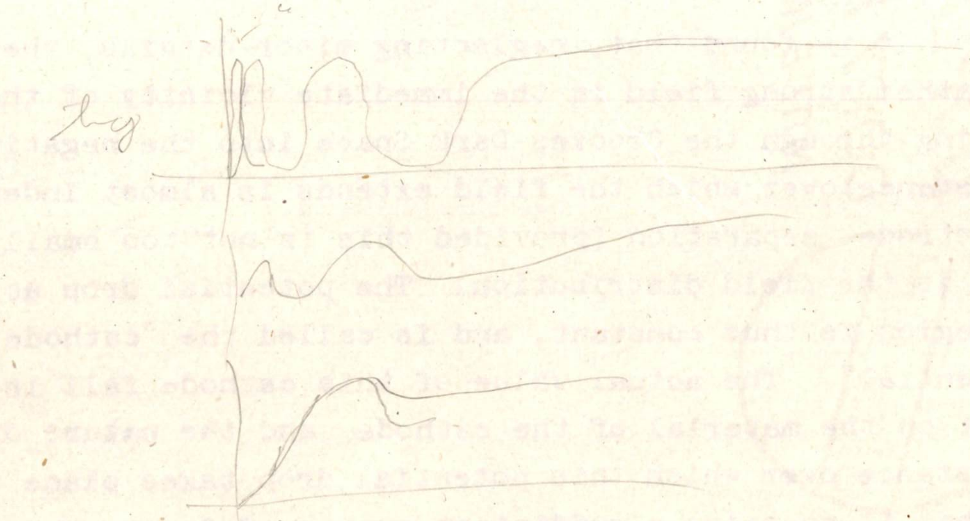


Figure 6.6

Measurement of Electric Field Intensity in a Discharge Tube

cathode and anode, C and A, are rigidly coupled together by means of a frame of glass rods, and this entire structure can be moved from the cathode by means of an iron slug I. An extremely fine beam of electrons from an auxiliary tube T at right angles to the main discharge tube passes through the discharge under investigation and onto a fluorescent screen S. The presence of an electric field E in the main discharge will cause the auxiliary beam to be deflected, and this deflection is a measure of E. In order to minimize the effect of the auxiliary discharge on the distribution of E in the main tube the auxiliary tube is completely shielded electrically from the rest of the apparatus. By using a very small current in the probe beam and a fairly high voltage, the ionization produced by the passage of this beam through the main discharge can be made negligible. This arrangement is particularly suitable for measuring the field in the regions near the cathode where the fields are strongest. It is possible by this method to measure the field to within a very small distance from the cathode itself.

It is found that, neglecting minor details, there is a rather strong field in the immediate vicinity of the cathode, extending through the Crookes Dark Space into the negative glow. The distance over which the field extends is almost independent of electrode separation (provided this is not too small) and so too is the field distribution. The potential drop across this region is thus constant, and is called the "cathode fall of potential". The actual value of this cathode fall is dependent on the material of the cathode and the nature of the gas. The distance over which this potential drop takes place is such that it contains a sufficient number of free paths so that the secondary electrons emitted from the cathode by positive ion bombardment can produce enough electrons by ionization by collision to maintain the discharge. This distance varies from a few tenths to one millimeter at a pressure of one millimeter of mercury, for a wide range of cathodes and gases. The cathode fall of potential is of the order of magnitude of, but usually somewhat less than, the minimum sparking potential of the gas in question, i.e., it is in the range 50-500 volts. The remainder of the potential drop across the discharge takes place, more or less uniformly in the absence of striations, across the positive column. The length of the positive column is usually enormous in comparison with the distance over which the cathode fall occurs. Thus for any reasonable voltage across the discharge tube the field in the positive column is much less than in a region of the cathode fall. The method of Thomson and


hg

Aston is therefore not suitable for field measurements in the positive column, but another method, involving the Langmuir probe, must be used. This will be discussed in section 6.2.2. The probe measures, among other things, the potential distribution in the positive column from which the field can be determined. Anticipating this, we show in Figure 6.7 the results of measurements of light intensity, potential, field, charge and current distributions in a typical discharge. The problem of explaining these results in a quantitative way is almost hopelessly complex, but it is not difficult to see in a qualitative manner why some of the phenomena are as they are.

In the region of the cathode fall we have essentially the mechanism of the self-maintained discharge taking place. We have seen (see equation 6.40) that the condition for this is $1 + \gamma = \gamma e^{\alpha d}$. More generally, since the field will not be uniform, we should write in the present case

$$1+\gamma = \gamma e^{\int_0^d \alpha \, dx} \tag{6.46}$$

where γ is the value appropriate to the field at the cathode. The d obtained from this equation will play a fundamental role and will be very close to the length of the region over which the cathode fall occurs. In this region the positive ions acquire an appreciable energy and thus produces the requisite secondary electrons by cathode bombardment. Although it is conceivable that electrons could also be emitted photoelectrically, independent experiments show that the number of such is inappreciable. We have already noted that the number of secondaries produced by collisions of positive ions with neutral molecules is also negligible. The fundamental process is then the liberation of secondary electrons by bombardment of the cathode with positive ions. These secondary electrons are emitted with very small velocities and consequently give rise to a net negative space charge in the immediate vicinity of the cathode, This is quickly overcome by the formation of positive ions and the acceleration of the electrons so that most of the region of the cathode fall is characterized by a predominantly positive space charge of rather high density. Beyond that, in the region of the negative glow still more electrons appear, since their mobility is so much greater than that of the positive ions. If the electron density exceeds the ion density, as in the case

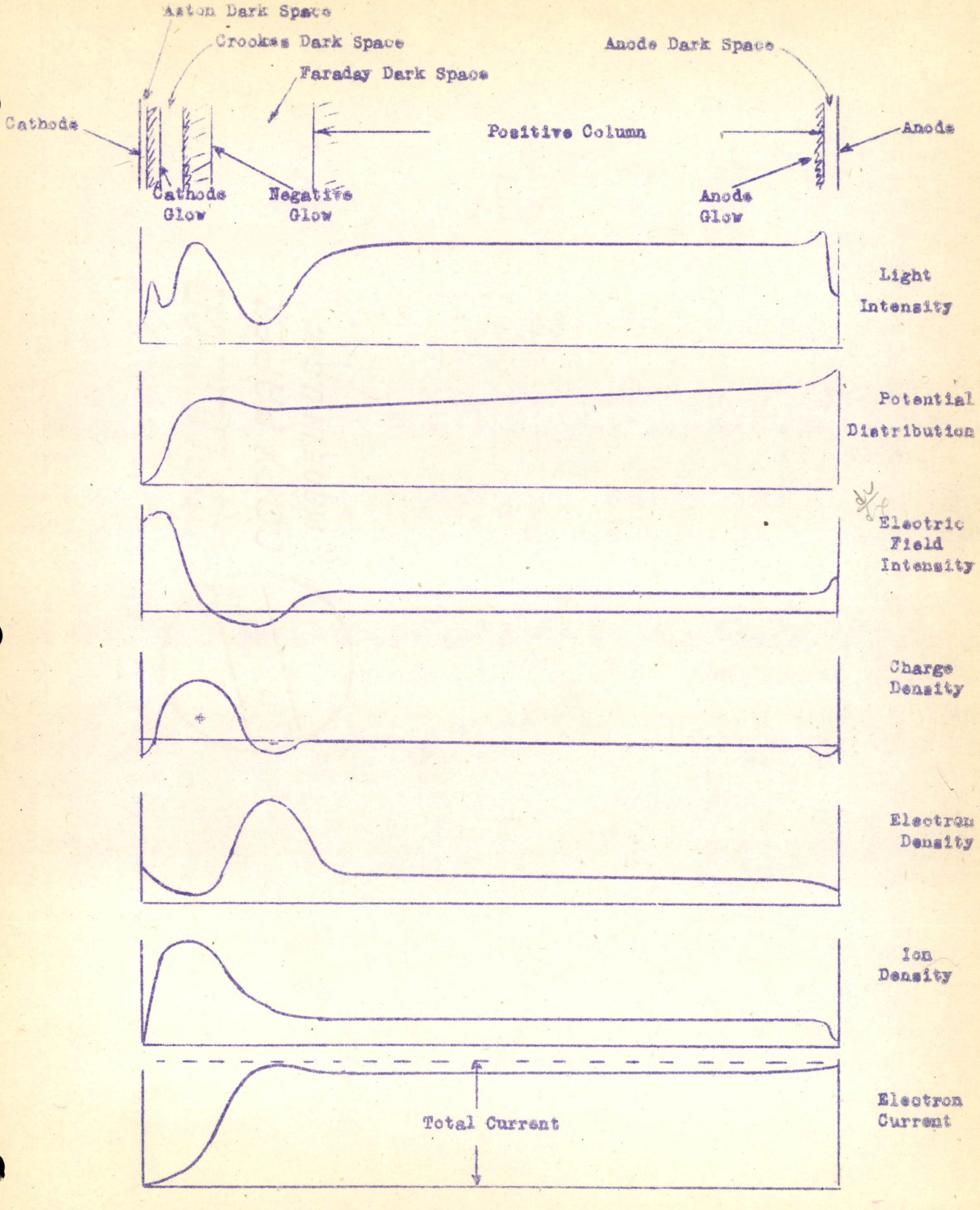


Figure 6.7 Properties of the Gaseous Discharge. (Loeb)

illustrated in Figure 6.7, the potential will pass through a maximum and undergo a small dip on the anode side of the negative glow, in the Faraday Dark Space. From here on the potential rises nearly uniformly almost to the anode and the net space charge thus vanishes. That is, in the positive column the density of positive ions is practically the same as that of the electrons. But the electrons are moving with much greater velocities and so constitute almost the entire current in this region. Finally, in the immediate vicinity of the anode there is a small decrease in electron density, due to removal of electrons, and a larger decrease in positive ion density, resulting in a net negative charge density here and a slight rise in field strength. This corresponds to the region of the anode glow followed by the anode dark space. This latter dark space is extremely narrow, amounting to about one free path for ionization.

This brief description is not intended to be an explanation of what takes place in a gaseous discharge, but rather a description of the results of measurements of the quantities involved. We have assumed throughout that we are in the region of the "normal glow discharge", and we have noted that this region is characterized by a "low" current density. We shall now make this vague statement a bit more precise by considering the actual currents involved. As seen in Figure 6.5, the current in the normal glow is independent of the voltage, which means that the current is determined by the resistance of the external circuit. Suppose that, starting at the point E, the external resistance is lowered. The current will increase toward F and the voltage drop across the discharge will be practically unaffected. But it will be observed that the area of the cathode covered by the cathode glow increases as the current is increased, until at the end of the region of the normal glow discharge the entire cathode area is covered. Further, the current density over the cathode is approximately constant throughout this change. We have noted (see also Figure 6.7) that the region of the cathode fall of potential is characterized by a high density of positive ions which, neglecting details near the cathode, starts at a high value and drops off with distance toward the Faraday Dark Space. To a rough approximation we have the situation of a space charge limited positive ion current in this region. If this is the case then the (positive ion) current density at the cathode will be of the form

$$J = \frac{4\varepsilon_o}{9}\sqrt{\frac{2e}{M}}\ \frac{V_c^{3/2}}{d^2} \qquad (6.47)$$

where V_c is the cathode fall of potential and d the distance over which this takes place. The length d is of the order of magnitude of, but somewhat less than, the distance at which sparking occurs at the minimum sparking potential, for the pressure in question. Although the details implied by this formula are not to be taken too seriously, it does give results of the right order of magnitude for the normal current density, and dimensionally we must have a dependence of this sort. Once the tube current exceeds the value $i = JA$ where J is the normal current density and A the cathode area we cannot have a normal glow discharge. The region of the abnormal glow now starts and the voltage drop across the tube increases with increasing current, as indicated by the portron FG in Figure 6.7. If the current be still further increased we quickly pass through the (unstable) region GH and into the low voltage arc. The situation here is quite different; the potential drop across the arc may be even less than the ionization potential of the gas in the discharge tube so that, at first sight, one might think that a self-maintained discharge could not exist. But the existence of metastable states and collisions of the second kind shows us that this situation is indeed possible. We will not go further into this subject at present.

6.2.2 The Langmuir Probe

Let us consider again the positive column, or plasma. We have noted that this is a region which is almost electrically neutral, since the density of positive ions is practically equal to the electron density. But the electrons, being much lighter will diffuse more rapidly through the gas. Consider those electrons which diffuse radially to the walls of the tube. They will cause the (glass) wall to become charged so that its potential is somewhat negative with respect to the plasma at each point. This negative potential difference will attract positive ions and since the net current to the wall must be zero the density of positive ions will be greater than that of the electrons since their velocity is much less. The walls will thus be covered with a sheath of electrons; that is, the net current to the walls is zero, but the walls carry a charge which is not zero. An analogous situation will hold for any surface in the region of the plasma. In fact, a confining surface, such as the walls of the discharge tube is almost essential for the formation of a positive column. A positive column does not form if the low pressure discharge takes place in a large spherical enclosure. Rather, the discharge and ionization diffuse over the whole volume and the concentration of excited atoms becomes quite small, so that the Faraday Dark Space extends almost to the anode, where the anode glow becomes more prominent than usual.

Suppose now that a probe, that is, an auxiliary electrode P, be inserted in the discharge as shown in Figure 6.8,

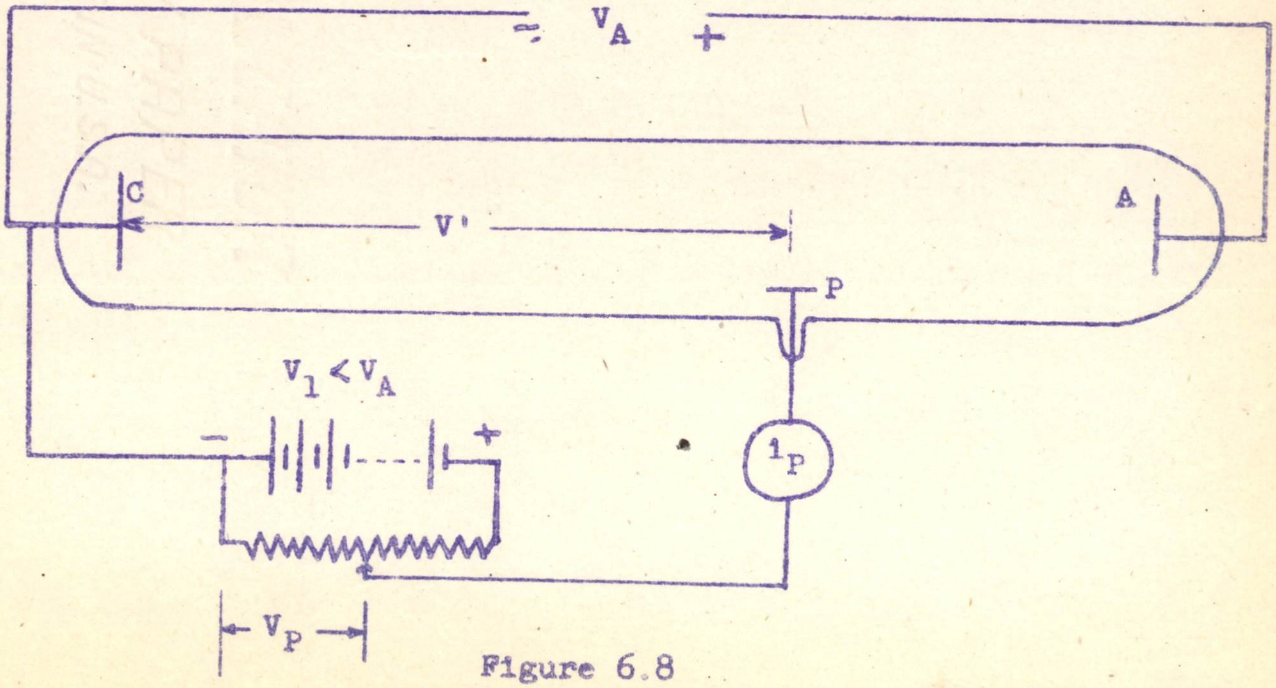


Figure 6.8
Probe Arrangement

radiation ions $j_+ = 1/4\, N_+ e \bar{v}_+$

and that the current to the probe be measured as a function of the potential difference V_p between the probe and cathode. Let V' be the potential in the plasma, measured with respect to the cathode, and let A be the area of the probe surface. Suppose first that $V_p \ll V'$ so that the probe is highly negative with respect to the plasma. Then the probe surface will be covered with a sheath of positive ions of thickness d_s and a space-charge-limited positive ion current will flow to the probe. The thickness d_s will depend on $V' - V_p$ and the ion density in the plasma, but is generally in the range from a few tenths to a few millimeters. The current density of positive ions into this sheath is the random ion current density, J_+, and is equal to the ionic charge times the number of ions crossing unit area per unit time. Thus we have from kinetic theory

$$J_+ = \frac{1}{4} N_+ e \bar{v}_+ \qquad (6.48)$$

where N_+ is the concentration of positive ions in the plasma and $\bar{v}_+$ the mean speed of these ions, if the ions have a Maxwellian distribution characterized by the temperature T_+. (This will be somewhat greater than the gas temperature in the discharge tube) Then

$$\bar{v}_+ = \sqrt{\frac{8KT}{\pi M}} \qquad (6.49)$$

From kinetic theory and so

$$J_+ = N_+ e \sqrt{\frac{KT_+}{2\pi M_+}} \qquad (6.50)$$

J_+ can be measured, but in general N_+ and T_+ are quantities we wish to determine. If we write

$$V_s = V' - V_p \qquad (6.51)$$

then the current density J_+ can be written

$$J_+ = \frac{4\epsilon_0}{9}\sqrt{\frac{2e}{M_+}}\,\frac{V_s^{3/2}}{d_s^2} \qquad (6.52)$$

Suppose now that the probe potential, V_p, be made more positive. Then a few of the faster moving electrons will be able to penetrate the sheath, (which also becomes thinner with decreasing V_s), and the measured current will be reduced. The ion current will be unaffected at first but later, when V_s has changed sign, the slower ions will be repelled and the sheath will be one of electrons rather than positive ions. When V_p is highly positive with respect to V' then only electrons can reach the probe and we have a space charge limited electron current across an electron sheath of thickness d_s. That is

$$J_- = \frac{4\epsilon_0}{9}\sqrt{\frac{2e}{M}}\,\frac{V_s^{3/2}}{d_s^2} \qquad (6.53)$$

Note that in both equations 6.52 and 6.53 V_s denotes the magnitude of the sheath potential .

$$J_- = \frac{1}{4} N_- e\, \bar{v}_- \qquad (6.54)$$

where $\bar{v}_-$ is the mean speed of the electrons and again, we have assumed a Maxwellian energy distribution for the electrons, though their mean energy will generally be much larger than that of the ions. Hence the "electron temperature" will be much greater. If we denote it by T_- then

$$\bar{v}_- = \sqrt{\frac{8KT_-}{\pi m}} \qquad (6.55)$$

so

$$J_- = N_- e\sqrt{\frac{KT_-}{2\pi m}} \qquad (6.56)$$

For values of the probe potential between the two extreme cases just considered the situation is more complicated, and this we now proceed to investigate. The curve in Figure 6.9 shows the current to the probe as a function of the probe

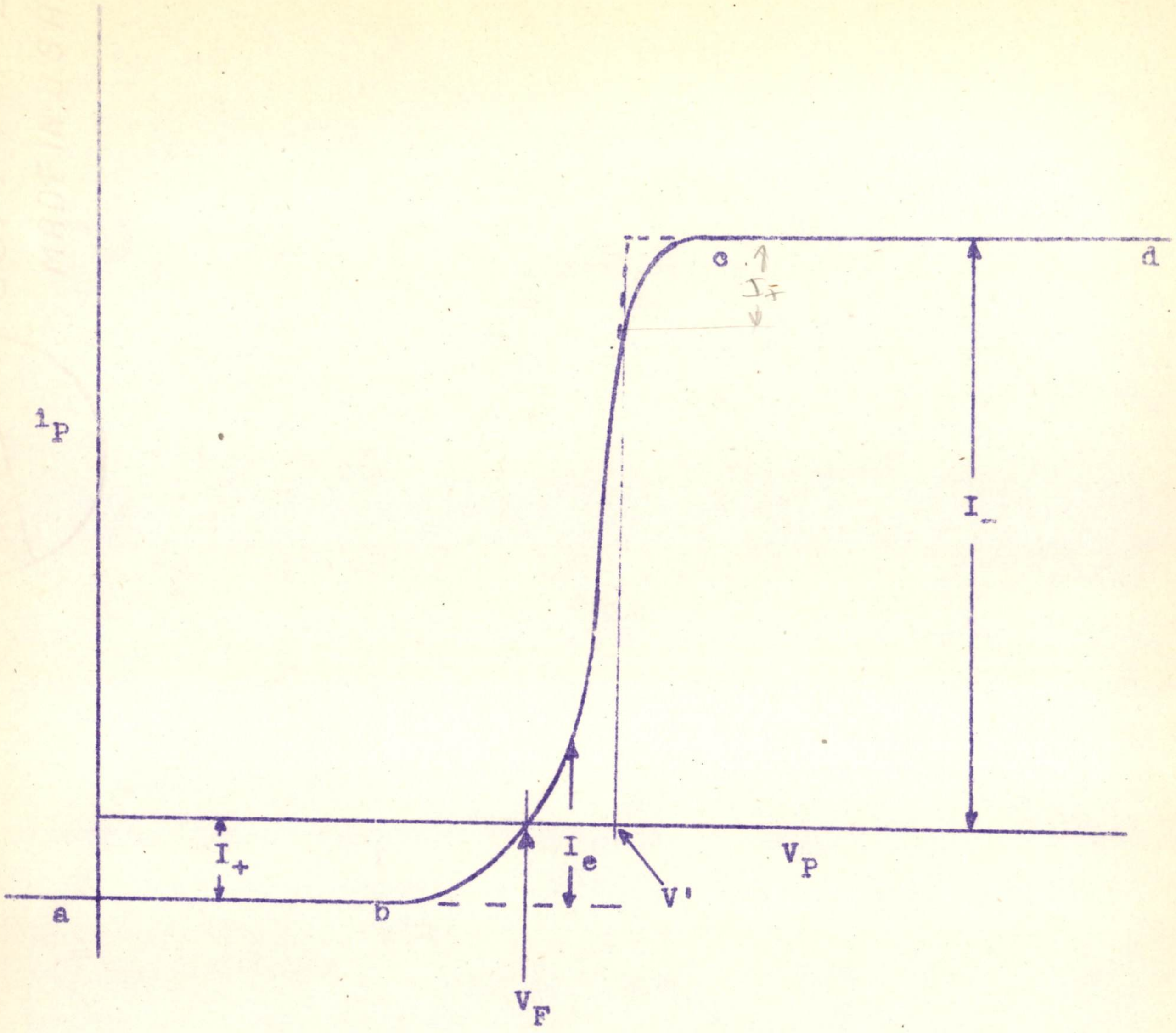


Figure 6.9

Probe Current vs. Probe Potential

potential. The portion ab corresponds to the first case considered, namely a space-charge-limited positive ion current, so that

$$I_+ = AJ_+ \qquad (6.57)$$

where A is the probe area (actually the area of the outside surface of the positive ion sheath) and J_+ is given by equation (6.50) or (6.52). The portion cd corresponds to the second case, a space-charge-limited electrun current, and,

$$I_- = AJ_- \qquad (6.58)$$

where J_- is given as above. The portion of the curve just to the right of b corresponds to the case where we are still collecting a positive ion current, but in addition some electrons are being collected. If we assume that the electrons in penetrating the positive ion sheath do not change their energy distribution and do not alter the positive ion concentration then the positive ion current will be the same as before and by subtracting it off we can measure the electron current I_e, as indicated in the figure. Further, we can determine the energy distribution of the electrons producing the current since the electron current, I_e, collected will be I_- diminished by the effect of the retarding potential from the plasma through the sheath to the probe, that is, the sheath potential. Thus

$$I_e = I_- \, e^{-V_s e/kT_-} \qquad (6.59)$$

or, on using equation (6.51)

$$I_e = I_- e^{-(V' - V_p) e/kT}$$

$$I_e = I_- \, e^{-V'e/kT_-} \cdot e^{+V_p e/kT_-} \qquad (6.60)$$

Thus if we plot $\ln I_e$ as a function of V_p we should obtain a straight line of slope e/kT_- if the electrons have a Maxwellian distribution characteristic of the temperature T_-. (If the curve so obtained is not a straight line, within experimental error of course, it means that the distribution is not Maxwellian). In particular the electron current will just equal the positive ion current, that is the probe current will be zero, for a certain probe potential V_F called the floating potential. It is the potential which an insulated probe would acquire if placed in the plasma. This is easily read from the graph. But for this value of the probe potential the sheath potential is $V_s = V' - V_F$ and so we must have

$$I_{+} = I_{-} e^{-(V' - V_{p})/kT_{-}} \tag{6.61}$$

This equation enables us to determine V', the plasma potential since I_{+}, I_{-} and V_{p} are determined directly and T_{-} is given by equation (6.60). Another, and perhaps more accurate way of determining V' is indicated in the figure. This is to extrapolate the straight line portion cd and the essentially straight line portion where the current is rising sharply, until they intersect at $V_{p} = V'$. The positive ion temperatures T_{+} is yet to be determined. One procedure would be to take this to be the same as the gas temperature, though it is usually somewhat greater than this. In principle, we might follow a method analogous to that used to determine T_{-}. That is, the current measured below cd in the vicinity of the knee of the curve represents the positive ion current I diminished by the effect of the retarding potential. If we denote this ion current by I_{1} then

$$I_{1} = I_{+} e^{-V_{s}e/kT_{+}} \tag{6.62}$$

or

$$I_{1} = I_{+} e^{V'e/kT_{+}} e^{-V_{p}e/kT_{+}} \tag{6.63}$$

so that if we plot $\ln I_{1}$ against $-V_{p}$ we should get a straight line of slope e/kT_{+}. Having determined these quantities, we can now get the ion and electron concentrations from equations (6.50) and (6.56). Probe measurements thus determine the following quantities: the plasma potential, the positive ion concentration, the electron concentration, the electron temperature, the random electron current density, the random ion current density, and the floating potential.

—— o ——

$3/2\ KT = .05$ ev

CHAPTER VII. THE STATISTICAL BEHAVIOR OF ELECTRONS IN METALS

(Ref. M. and S., Chapter IV)

7.1 The Experimental Basis for a new Statistics

In dealing with extremely dense assemblies of particles, such as the free or conduction electrons in a metal, it is necessary to resort to statistical methods; otherwise the mathematical complexity is imppssible to handle. One might at first think that the classical (Maxwell-Boltzmann) statistics for the particles of a gas might hold equally well for the electrons inside a metal, and indeed there is some justification for this point of view since the electrons emitted thermionically from a metal do obey M-B statistics. This fact, however, is fortuitous as we shall see later.

Measurements of the conductivity of metals, the slight dependence of the conductivity on small amounts of impurities and the high reflecting power of metals for light waves all lead to the conclusion that the number of free (conduction) electrons in a metal is about equal to the number of atoms; for light metals this is of the order of 5×10^{22} electrons per cc. On the basis of classical statistics each of these electrons should contribute 3/2 k to the specific heat, C_v. Experimentally the specific heat of metals is explained completely by the heat energy of the atoms, and the electrons are known to contribute very little if any to the specific heat.

While the energy distribution of thermionically emitted electrons seems to follow the M-B laws, the energy distribution of photoelectrically emitted electrons does not. On the basis of M-B statistics one would expect a spread in energy of the photoelectrons of about 3/2 kT or 0.05 e.v. at room temperature. Actually the energy spread is of the order of several volts (depending on the frequency of the light). The scattering of electrons by single metallic crystals, the Davisson-Germer experiment, not only demonstrated the truth of the prediction by de Broglie that electrons have wave properties, and wavelengths $\Lambda = h/p$, but also showed that the binding energy of electrons in metals was equal to the work function of the metal plus about 5 e.v. Classically, we should expect this binding energy to equal the work function plus 3/2 kT. Other experimental evidence such as the anomalous Hall effect and the emission of electrons by strong fields would not be explained classically.

7.2 The Theoretical Foundation for Quantum Statistics

The statistics which apply to free electrons in metals was derived independently by Fermi and Dirac about 1925. The theoretical arguments used follow.

1) We shall assume that the electrons are in thermodynamic equilibrium with each other (this assumption is basic for any kind of statistics) and with the atoms of the metal.

2) We shall assume that there are no ordinary forces, forces which depend on position only, acting on the electrons in the body of the metal, and hence that the potential energy is everywhere constant. There are, of course, forces at the surface which keep the electrons in the metal, but these forces may be handled independently of the statistics. (These forces for the moment may be regarded as strictly analogous to the walls of a box containing gas molecules obeying the classical statistics.)

3) We shall assume that the most probable state corresponds exactly to the actual physical state.

4) We shall assume that the electrnns obey the general laws of quantum mechanics which are: (a) Pauli Exclusion principle, (b) Heisenberg uncertainty principle, and (c) the principle that each electron is completely indistinguishable from every other electron.

Assumptions (1), (2) and (3) could lead equally well to Maxwell's formulation of the classical statistics. It is our problem to introduce assumptions 4(a), 4(b) and 4(c) into the usual statistical procedure, in order to obtain the quantum (Fermi-Dirac) statistics.

7.3 Cells and Phase Space

Let us consider the number of electrons n_1 which fall into a particular quantum state chacterized by energy w_1 in the small range dw_1. (dw_1 is perfectly arbitrary, but when we deal with large numbers of particles it can be treated exactly as a differential). We shall call g_1 the number of states of energy w_1 in the range dw_1 per unit volume of the metal. We can evaluate g_1 by means of the Heisenberg uncertainty principle.

According to Heisenberg the minimum uncertainty in the simultaneous measurement of the momentum, p_x, and the position, x, of a particle is of the order of Planck's constant, h; that is, $\Delta p_x \Delta x \cong h$. We can formulate this in three dimensions,

$$\boxed{\Delta p_x \Delta p_y \Delta p_z \Delta x \, \Delta y \, \Delta z \cong h^3}$$

Now since this is the limit of accuracy with which we can specify the state of a particle in position and momentum, we will assume that this is the size of the unit cell in the combined "space" of momenta and position coordinates. (This is generally called the "phase-space"). If the precision with which we intend to specify the particle is determined by the differentials dp_x, dp_y, dp_z, dx, dy, dz then certainly the product $dp_x dp_y dp_z dx \, dy \, dz$ is greater than h^3. The number of available cells in phase space is then just

$$\boxed{g_1' = \frac{dp_x dp_y dp_z \, dx \, dy \, dz}{h^3}}$$

We should like to determine g_1 in terms of the energy range dw_1 rather than the combined space-momentum range. To do this we first integrate over the space coordinates (nothing depends on x, y and z) obtaining

$$g_1'' = \frac{V_o}{h^3} dp_x dp_y dp_z \qquad (7.1)$$

where V_o is the total volume of the metal we are considering. Now by equation (7.1) the cells are uniformly distributed over momentum space, and the total momentum p .

$$p^2 = p_x{}^2 + p_y{}^2 + p_z{}^2 \qquad (7.2)$$

Hence a sphere of radius p in momentum space is a surface of constant total momentum and the number of cells of momentum p in the range dp is just equal to the number of cells between two spherical surfaces of radii p and p + dp. This number of cells is equal to the "volume" of the spherical shell $4\pi p^2 dp$ multiplied by the density of states, V_o/h^3. Hence,

$$g_1''' = 4\pi p^2 \frac{V_o}{h^3} dp \qquad (7.3)$$

In order to transform (7.3) to the desired form we need only remember

$$w_1 = \frac{p^2}{2m}, \quad p = \sqrt{2mw_1}$$

$w = \frac{1}{2}mv^2$

$p = mv$

$$dw_1 = \frac{p}{m} dp$$

$$p^2 dp = (2m^3)^{\frac{1}{2}} w_1^{\frac{1}{2}} dw_1$$

and hence, evaluating (7.3) in terms of energy w_1, we have

$$\boxed{g_1 = \frac{2\pi}{h^3} (2m)^{3/2} w_1^{\frac{1}{2}} dw_1}$$

We have divided through by V_o since we originally defined g_1 as the number of states per unit volume of the metal.

7.4 The Fermi-Dirac Statistics

Our problem now is to determine how to divide our particles among the various allowed cells. Statistically we wish to determine the probability that n_1 particles are in the state g_1. Now certainly the probability that such a state exists is proportional to the number of ways in which we can form that state, under the quantum mechanical conditions that no two particles can be in the same quantum state, and that the particles are completely indistinguishable; that is, that exchanging particles A and B does not constitute a new state.

Pauli principle as applied to metals states that no two electrons can occupy the same cell in momentum (or energy) space unless they have different spin quantum numbers. Since the spin quantum number has only two possible values this means that two and only two electrons can occupy the same cell in momentum space. Mathematically this is exactly the same as though we had doubled the number of cells available to the electrons, and we will redefine our g_1 accordingly, i.e.,

$$g_1 = \frac{\overset{2\times}{4\pi}}{h^3}(2m)^{3/2}\, w_1^{\frac{1}{2}}\, dw_1 \qquad (7.4)$$

We now inquire how many ways are there to divide n_1 particles into g_1 states. This clearly is the same as asking how many ways are there to take g_1 things n_1 at a time, and we will set the probability, P_1, of this state proportional to this number, getting,

$$P_1 \propto \frac{g_1!}{(g_1 - n_1)!\, n_1!} \qquad (7.5)$$

To see this let us use the following argument: if we have g_1 things we can take any one of them in g_1 different ways and so on until we have but $g_1 - n_1 + 1$ left which we can take in one way only. The total number N_1 is the product of the individual numbers, and so,

$$N_1 = (g_1)(g_1 - 1)(g_1 - 2)\ \ldots\ (g_1 - n_1 + 1) = \frac{g_1!}{(g_1 - n_1)!}$$

However, we have assumed that each of the n_1 ways is distinguishable from any other way and this is the same as saying that the electrons are distinguishable. To correct this we must realize that we must divide by the number of ways of taking n_1 particles n_1 at a time (assuming distinguishability). This is just $n_1!$ and dividing N_1 by this factor we have equation (7.5).

In order to proceed further we recall two theorems.

One is Boltzmann's theorem which states that the entropy of a system is proportional to the logarithm of the probability,

$$S = k \ln P$$

This is reasonable since the most probable state of a system is the state of greatest entropy by the second Law of Thermodynamics. Also since the probability of two events occuring simultaneously is just the product of the probabilities of each occuring independently, while the entropy of the sum of two systems is just the sum of the individual entropy, the logarithmic connection seems correct.

The second theorem we need is due to Stirling and states that

$$\ln n! \cong n \cdot \ln(n) - n$$

as n becomes very large.

Let us now return to equation (7.5) which gives the probability that n_i electrons will be in the i^{th} state with energy w_i in the range dw_i. If we consider all the electrons in a system they can distribute themselves in many states of different w. The probability of any one state occuring for the system is then the product of the probabilities for all the individual states, hence

$$P \propto \prod^{i} \frac{g_i!}{n_i!(g_i - n_i)!}$$

and the entropy of this system is then

$$S \propto \sum_i \left\{ \ln g_i! - \ln n_i! - \ln(g_i - n_i)! \right\} \qquad (7.6)$$

applying Stirling's theorem

$$S \propto \sum_i \left\{ g_i \ln g_i - g_i - n_i \ln n_i + n_i - (g_i - n_i)\ln(g_i - n_i) + g_i - n_i \right\}$$

$$S \propto \sum_i \left\{ g_i \ln g_i - n_i \ln n_i - (g_i - n_i)\ln(g_i - n_i) \right\} \qquad (7.7)$$

Now the probability and the entropy are a maximum for the most probable state. Hence, let us take a small variation in S, δS, as the individual n_i's change by δn_i, and set δS equal to zero for a maximum. The g_i's will remain constant as given by equation (7.4) as long as we keep dw, the energy range under consideration, constant.

$$\delta S \propto \sum_i \left\{ -\delta n_i \ln n_i - 1 + \delta n_i \ln(g_i - n_i) + 1 \right\} = 0$$

$$\sum_i \left\{ \ln(g_i - n_i) - \ln n_i \right\} \delta n_i = 0 \qquad (7.8)$$

We also require that the total number of particles $N = \sum_i n_i$, and the total energy $W = \sum_i n_i w_i$ remain constant and hence that δN and $\delta W = 0$. Then,

$$\sum_i \delta n_i = 0 \qquad (7.9)$$

$$\sum_i w_i \delta n_i = 0 \qquad (7.10)$$

We now use LaGrange's undetermined multipliers to aid our solution. Since all the summations are over the same index, i, and since δn_i is a common factor, let us multiply equation (7.9) by $-\alpha$ and (7.10) by $-\beta$ and add. (α and β will be determined later from the physical aspects of the problem.

$$\sum_i \left\{ \ln \frac{g_i - n_i}{n_i} - \alpha - \beta w_i \right\} \delta n_i = 0 \qquad (7.11)$$

Now the δn_i's are completely independent so we must require that each term in brackets is individually zero. Hence,

$$\ln \left(\frac{g_i}{n_i} - 1 \right) = \alpha + \beta w_i \qquad (7.12)$$

Rearranging terms and inverting the logarithm, we have,

$$n_1 = \frac{g_1}{1 + e^{\alpha + \beta w_1}}$$

Now g_1 is a differential (or nearly so) and our notation should be changed to express n_1 as a differential also. Setting $n_1 = n(w)dw$, dropping the subscripts, and using (7.4),

$$n(w)dw = \frac{4\pi}{h^3}(2m)^{3/2} \frac{w^{\frac{1}{2}}\, dw}{1 + e^{\alpha + \beta w}} \qquad (7.13)$$

n(w)dw is then the number of particles of energy w in the range dw per unit volume. n(w) is the "density" function; that is, it is the number of electrons of energy w per unit volume per unit energy range.

7.5 Determination of α and β

In order to determine β let us examine the behavior of equation (7.13) when $\alpha + \beta w$ is appreciably greater than unity. The exponential is then much greater than 1 and we can rewrite equation (7.13) as approximately,

$$\left.\begin{aligned} n(w)dw &\cong A\, w^{\frac{1}{2}} \cdot e^{-(\alpha + \beta w)}\, dw \\ \text{or} \qquad & \\ n(w)dw &\cong A'\, w^{\frac{1}{2}}\, e^{-\beta w}\, dw \end{aligned}\right\} \qquad (7.14)$$

where A and A' are constants. Now these expressions are of the same form as the corresponding M-B functions. We shall see later that one condition for $\alpha + \beta w$ being greater than 1 is that the density of the electrons is small. Physically we should expect electrons to obey M-B statistics if the density is small since the effect of the Pauli exclusion principle would be much less pronounced. It is reasonable then to assume that β has its classical value of 1/kT.

Let us now set $\alpha = -w_M/kT$ and proceed to determine

w_M. Equation (7.13) becomes,

$$\boxed{n(w)dw = \frac{4\pi}{h^3}(2m)^{3/2} \frac{w^{\frac{1}{2}}\,dw}{1+\exp\left(\frac{w-w_M}{kT}\right)}} \qquad (7.15)$$

Let us evaluate n(w) at the absolute zero of temperature. If we set $T = 0$ in (7.15) the exponential is zero for $w < w_M$ and hence n(w) is proportional to $w^{\frac{1}{2}}$. For $w > w_M$ the exponential term goes infinite and n(w) drops sharply to zero. Hence, physically w_M, at absolute zero, is just the maximum energy of the electrons. Classically we should expect the energy of all electrons at $T = 0°K$ to be zero; Pauli principle denies this.

At 0°K w_M is easily evaluated. Let us integrate n(w)dw over all possible values of w. The integral of the left hand side of (7.15) is just N, the number of free electrons per unit volume. Because of the behavior of the exponential term at 0°K, the integral of the right hand side is just,

$$\frac{4\pi}{h^3}(2m)^{3/2}\int_0^{w_{MO}} w^{\frac{1}{2}}\,dw$$

and hence

$$N = \frac{8\pi}{3h^3}(2m)^{3/2}\, w_{MO}^{3/2}$$

from which

$$\boxed{w_{MO} = \frac{h^2}{2m}\left(\frac{3N}{8\pi}\right)^{2/3}} \qquad (7.16)$$

where we have set w_{MO} equal to the value of w_M at $T = 0°K$. w_M can be evaluated at any value of T by integrating over (7.15) for $T \neq 0$. The result can be expressed as a power series in $\left(\frac{kT}{w_{MO}}\right)^2$, and is,

$$\boxed{w_M = w_{MO}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{w_{MO}}\right)^2 + \dots\right\}} \qquad (7.17)$$

For all real metals w_{MO} is between 2 and 10 e.v. while kT is a few tenths of an electron volt even at elevated temperatures. Hence, w_M changes very little even for wide ranges of temperature; the change in N due to expansion of the metal is certainly as important as the temperature dependence in (7.17). For our purposes we can take w_M as constant and equal to w_{MO}.

The density function n(w) is plotted in Fig. 7.1a. The solid curve shows n(w) at $T \cong 0$. The dotted curve is for a higher temperature; for real metals the dotted curve corresponds to $T \cong 2500°K$.

7.6 Other Distribution Functions

Besides the density of electrons per unit energy range we should like to know the density per unit speed range or per unit velocity range. By a procedure similar to that of section 7.3 we can easily show that the distribution of electrons with velocity components v_x, v_y and v_z in the range $dv_x\, dv_y\, dv_z$ is

$$n(v_x, v_y, v_z)dv_x dv_y dv_z = \frac{2m^3}{h^3} \frac{dv_x\, dv_y\, dv_z}{1 + \exp\left(\frac{w - w_M}{kT}\right)} \qquad (7.18)$$

and that the distribution of electrons with speed v in the range dv is

$$n(v)dv = \frac{8\pi m^3}{h^3} \frac{v^2 dv}{1 + \exp\left(\frac{w - w_M}{kT}\right)} \qquad (7.19)$$

In expressions (7.18) and (7.19) the functional dependence of w on the speed and velocities components is understood:

$$w = \frac{1}{2} m(v_x^2 + v_y^2 + v_z^2) = \frac{1}{2} mv^2$$

The density functions $n(v_x, v_y, v_z)$ and $n(v)$ are plotted against w in Fig. 7.1,b and c.

In the study of the emission of electrons from metals there is one other distribution function of some importance, namely, the function that gives us the distribution of electrons of particular velocity component v_z in the range dv_z with no restrictions on the values of the other two components of velocity. This function is certainly obtainable from (7.18) by integration over all possible values of v_x and v_y. Hence,

$$n(v_z)dv_z = dv_z \int_0^\infty\int_0^\infty n(v_x, v_y, v_z)dv_x dv_y$$

$$= dv_z \int\int_{}^{\infty\,\infty} \frac{2m^3}{h^3} \frac{dv_x\, dv_y}{1 + \exp\left(\frac{\frac{1}{2}m(v_x^2 + v_y^2 + v_y^2) - w_M}{kT}\right)}$$

In order to integrate this let us transform to polar coordinates such that

$$\rho^2 = v_x^2 + v_y^2$$

$$\rho d\rho d\varphi = dv_x dv_y$$

We have then

$$n(v_z)dv_z = dv_z \int_0^\infty\int_0^{2\pi} \frac{2m^3}{h^3} \frac{\rho d\rho d\varphi}{1 + \exp\left(\frac{\frac{1}{2}m\rho^2 + \frac{1}{2}mv_z^2 - w_M}{kT}\right)}$$

Integrating over φ, we have,

$$n(v_z)dv_z = \frac{4\pi m^3}{h^3} dv_z \int_0^\infty \frac{\rho d\rho}{1 + \exp\left(\frac{\frac{1}{2}m\rho^2 + \frac{1}{2}mv_z^2 - w_M}{kT}\right)}$$

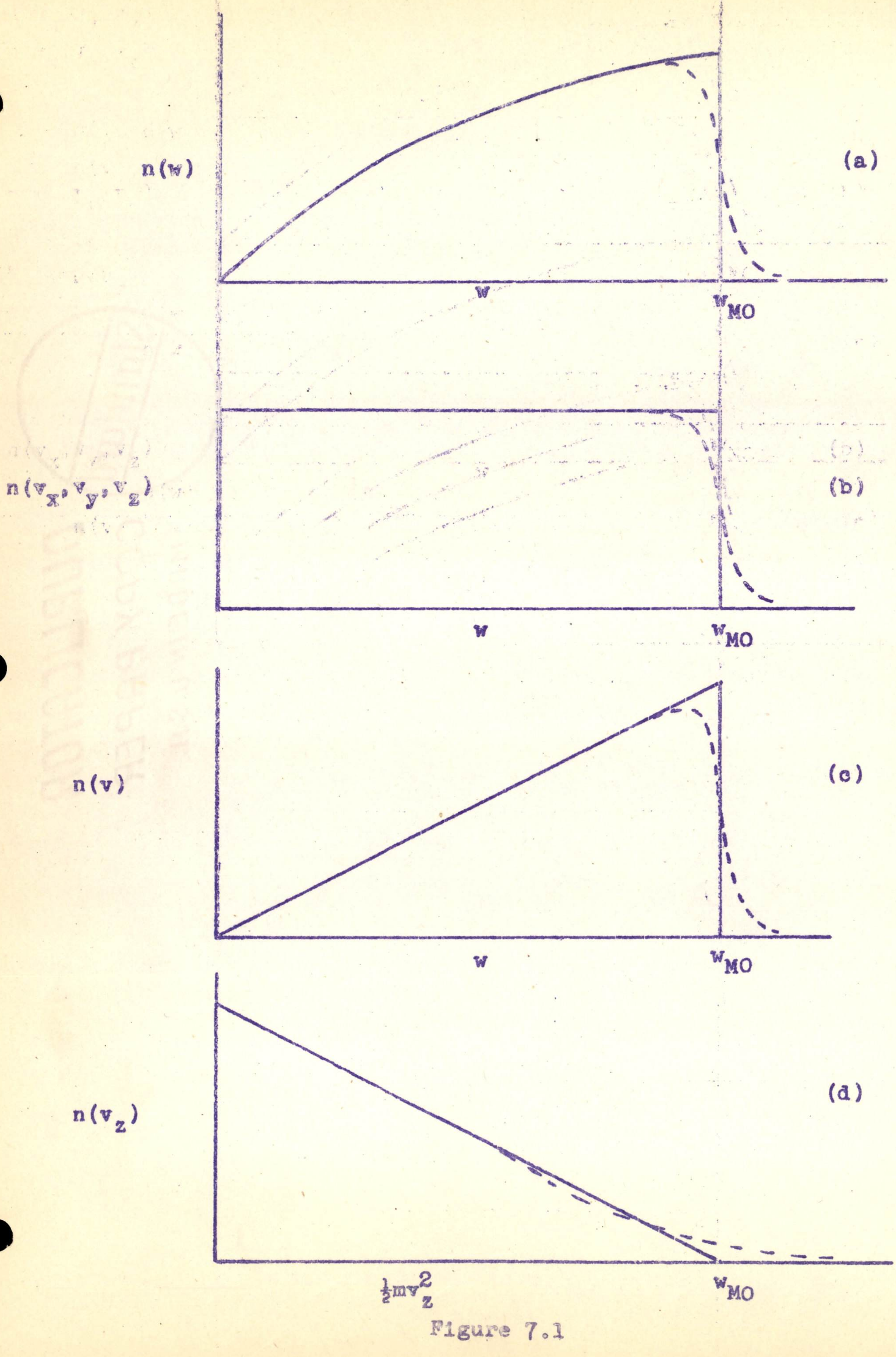


Figure 7.1

Now let

$$\eta = \frac{\frac{1}{2} m \rho^2 + \frac{1}{2} m v_z^2 - w_M}{kT}$$

$$d\eta = \frac{m}{kT} \rho \, d\rho$$

$$n(v_z) dv_z = \frac{4\pi m^2}{h^3} \times kT dv_z \int_{\eta_0}^{\infty} \frac{d\eta}{1 + e^{\eta}} \qquad (7.20)$$

where

$$\eta_0 = \frac{\frac{1}{2} m v_z^2 - w_M}{kT} \quad .$$

The indefinite integral is standard and is given by

$$\int \frac{d\eta}{1+e^{\eta}} = \int \frac{e^{-\eta} d\eta}{1 + e^{-\eta}} = - \ln(1 + e^{-\eta})$$

and therefore,

$$n(v_z) dv_z = \frac{4\pi m^2}{h^3} \; kT \; \ln \left\{ 1 + \exp\left(\frac{w_M - \frac{1}{2} m v_z^2}{kT}\right) \right\} \qquad (7.21)$$

The low temperature behavior of $n(v_z)$ can readily be seen. If T is very small then

$$\frac{w_M - \frac{1}{2} m v_z^2}{kT}$$

is very large, except near $w_M = \frac{1}{2} m v_z^2$, and we can certainly neglect the 1 in comparison to the exponential term. Equation (7.21) then becomes,

$$n(v_z) dv_z \cong \frac{4\pi m^2}{h^3} (w_M - \frac{1}{2} m v_z^2) dv_z$$

This function plotted against $\frac{1}{2} m v_z^2$ is shown in Fig. 7.1(d) as the solid line; the dotted line shows $n(v_z)$ for a temperature in the neighborhood of 2500°K for a real metal.

7.7 The Specific Heat of the Electron Gas

We may use the distribution functions to compute average values of energy, speed etc. For instance, the average value of the energy is just,

$$\bar{w} = \frac{1}{N} \int_0^\infty w\, n(w)dw = \frac{\int_0^\infty w\, n(w)dw}{\int_0^\infty n(w)dw}$$

At 0°K these integrals can be evaluated easily and we have

$$\bar{w} = \frac{3}{5} w_{MO} \qquad (7.22)$$

At temperatures other than 0°K the integral can be evaluated with some difficulty by series expansion (see M. and S, Appendix IV) and the result to the first order in $(kT/w_{MO})^2$ is,

$$\bar{w} \cong \frac{3}{5} w_{MO} \left\{ 1 + \frac{5\pi^2}{12} \left(\frac{kT}{w_{MO}}\right)^2 \right\} \qquad (7.23)$$

Since w_{MO} is of the order of 5 e.v. and kT is about 0.03 e.v. (for room temperatures) the correction term in (7.23) is about 1 part in 5000 or 0.02%.

The specific heat <u>per free electron</u> is then

$$c_v = \frac{dw}{dT} = \frac{\pi^2}{2} k\left(\frac{kT}{w_{MO}}\right)$$

At room temperatures this is about 2% of the classical value of (3/2)k. Hence, although the F-D statistics gives a much higher average energy of the electrons, the temperature dependence is extremely low, and agrees with the experimental observation that the electron gas does not contribute to the specific heat.

The experimental fact that thermionically emitted electrons obey M-B statistics can be readily understood. These electrons evidentally come from the high energy region (high temperature tail) of the F-D distribution. It is here that the exponential term in (7.15) predominates and the distribution of <u>emitted</u> electrons is of the form of (7.14).

If we now re-examine the facts regarding the photoelectric energy distribution and the large potential energy demanded by the Davisson-Germer experiment we see that the large spread in energy of the free electrons in the metal, required by the F-D statistics, agrees with the experimental facts.

8.1 The Metallic Model

In order to fix our ideas about metals we must make some assumptions about the electric fields and potentials existing at the surface and in the body of the metal. In deriving the F-D statistics we have already assumed that no forces act on the electrons in the body of the metal; that is, that the potential energy of the electrons is constant inside the metal. That this is not too bad an assumption, especially for emission problems, is amply proved by experiment. Of course, a detailed examination of the atomic structure of the metal shows that this assumption is untrue. However, the mathematics for dealing with a more exact model becomes hopelessly complicated.

Having decided upon a workable assumption for the interior of the metal we must now consider the forces on the electron very near the surface of the metal. Certainly there are strong forces at the surface; otherwise the electrons would readily escape and we know that they do not escppe except at elevated temperatures or when subjected to other strong stimuli such as illumination by light of high-frequency, high electric fields, etc. The simplest assumption, and one serviceable for some purposes, is that there is a discontinuity in potential at the surface, the "square barrier". A somewhat better assumption, and one which is completely confirmed experimentally from about 50 A on out, is the image field barrier. That is, we know that a charge, $-e$, situated at a distance z from a large metal surface, is acted upon by an image force equal to $-e^2/16\pi\varepsilon_0 z^2$ (directed toward the surface).

A potential energy can be associated with this force and is equal to $-e/16\pi\varepsilon_0 z$. If we arbitrarily set the potential energy at large distances from the metal equal to zero; and set the constant value inside the metal equal to $-W_B$, there exists some uncertainty about the exact value of z at the surface, since according to the image force law we should have an infinity at $z = 0$. Let us arbitrarily define z_0 as the point at which the image potential energy becomes equal to the constant value $-W_B$. We have

$$-W_B = \frac{-e^2}{16\pi\varepsilon_0 z_0}$$

From which

$$z_0 = \frac{e^2}{16\pi\varepsilon_0 W_B} = .36 \times 10^{-10}\ \text{m} = .36\text{A}$$

for $W_B = 10$ e.v.

Hence, the error we can make in allowing the potential energy to become constant at $z = z_0 \doteq .36$A is small in dealing with distances from the surface of 10A or greater. The potential energy as a function of z for both the image barrier and square barrier is shown in Fig. 8.1.

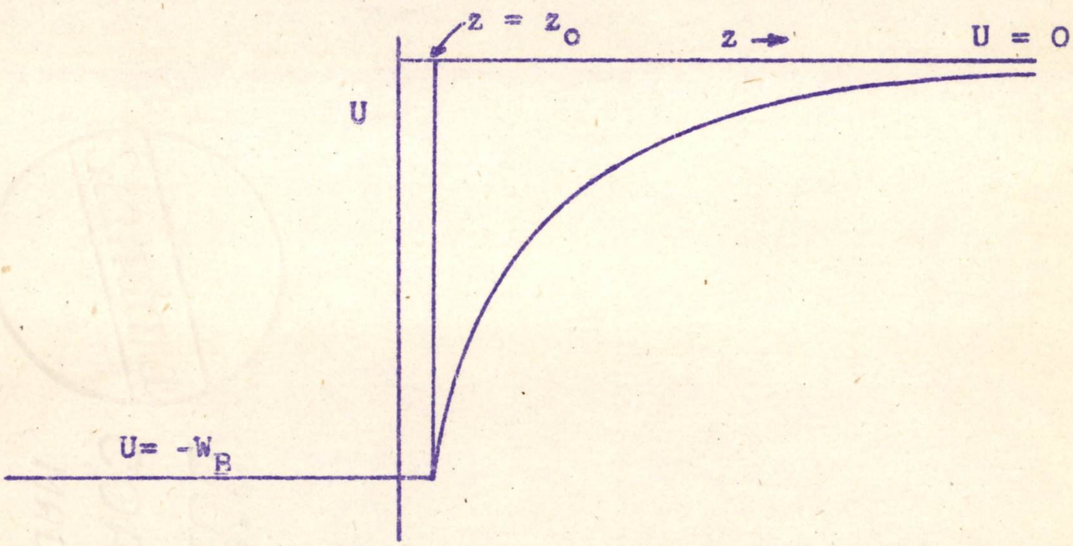


Fig. 8.1

Hence for the square barrier

$$U = -W_B,\ z < z_0,\ U = 0,\ z > z_0$$

and for the image barrier

$$U = -W_B,\ z < z_0,\ U = -\frac{e^2}{16\pi\varepsilon_0 z},\ z > z_0$$

We can also show the distribution of the electrons in energy (for convenience at T = 0°K) on such a plot as shown in Figure 8.2. In Figure 8.2(a) the density of electrons in indicated by the density of the lines. In Figure 8.2(b) the density is indicated by plotting n(w) to scale.

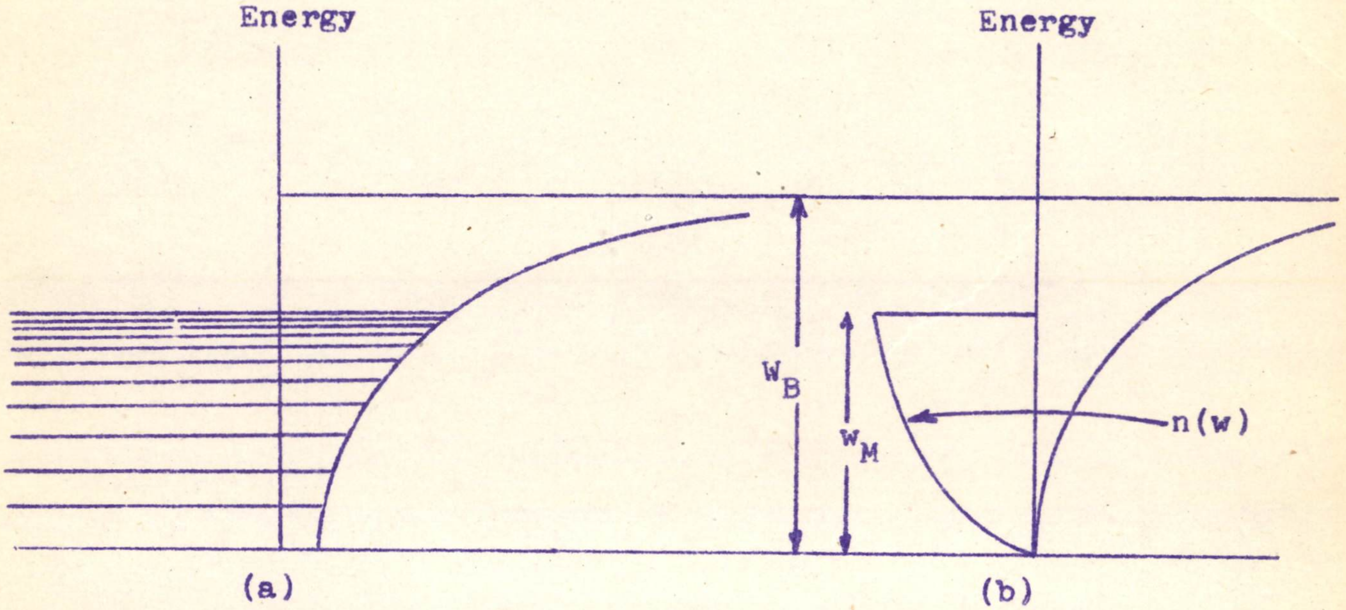


Figure 8.2

For our purposes we shall suppose that the image field barrier correctly represents the facts, and modify this supposition where necessary.

8.2 Thermionic Emission

In order to obtain an expression for the emission of electrons by temperature excitation, we must calculate the number of electrons which have energies large enough, and velocities so directed, that they may surmount the potential barrier at the surface. Since the surface is perpendicular to the z direction, and the image force and potential barrier is in the z direction, we are clearly interested in the number of electrons for which

$$v_z \geq \sqrt{\frac{2W_B}{m}}$$

However, let us consider one point; we are interested in a current, not a charge density. Hence, the number of electrons per unit volume, as given by equation (7.21), having velocity component v_z in the range dv_z is not the right function. Rather we want the number of electrons per second coming to unit area of the surface with velocity component v_z in the range dv_z. If we call this function $n'(v_z)dv_z$ it is related to $n(v_z)dv_z$ by,

$$n'(v_z)dv_z = v_z \cdot n(v_z)dv_z \tag{8.1}$$

In order to order to see this consider all the electrons having velocity components between v_z and $v_z + dv_z$ in the volume of length v_z (numerically) and unit cross-sectional area, as shown in Figure 8.3.

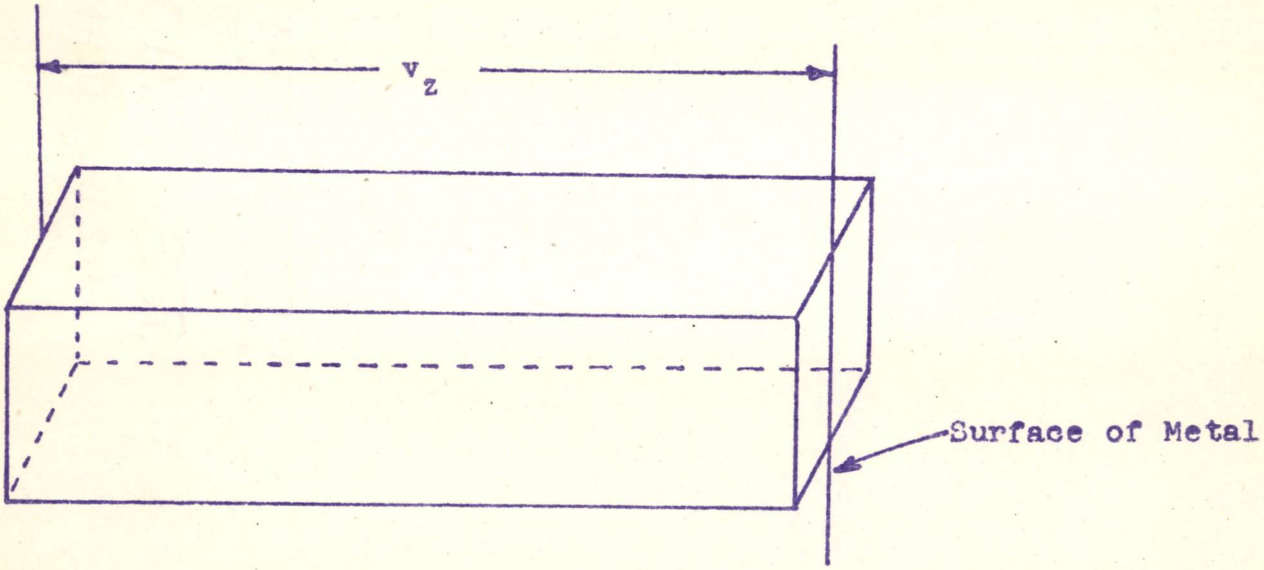


Figure 8.3

Clearly all of the electrons in this volume will arrive at unit area of the surface per second, and it remains to evaluate those which can escape from the metal and contribute to the thermionic current. The thermionic current density is then

$$J_{TH} = e \int_{\sqrt{\frac{2W_B}{m}}}^{\infty} n'(v_z)\,dv_z = e \int_{\sqrt{\frac{2W_B}{m}}}^{\infty} v_z \cdot n(v_z)\,dv_z \qquad (8.2)$$

Using equation (7.21),

$$J_{TH} = \frac{4\pi m^2 e}{h^3}\, kT \int_{\sqrt{\frac{2W_B}{m}}}^{\infty} v_z \,\ln \left\{ 1 + \exp\left(\frac{w_M - \frac{1}{2} mv_z^2}{kT}\right)\right\} dv_z \qquad (83)$$

This would be a rather formidable integral were it not known experimentally that those electrons which can escape (for which $\frac{1}{2} mv_z^2 \geqslant W_B$) have values for $\frac{1}{2} mv_z^2$ about 2-6 e.v. greater than w_m. Hence even for T = 2900 K,

$$\frac{w_M - \frac{1}{2} mv_z^2}{kT} \cong -8$$

or smaller. Hence, the exponential in the logarithmic term is small compared to 1. Recalling that

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} \cdot \cdot \cdot \quad \text{for } x \ll 1$$

we can replace the logarithmic term in (8.3) by

$$\exp\left(\frac{w_M - \frac{1}{2} mv_z^2}{kT}\right)$$

and the first neglected term is $\frac{1}{2} e^{-8}$ or roughly 10^{-4} times as large as this term. Hence we can write,

$$J_{TH} = \frac{4\pi m^2 ekT}{h^3} \int_{\sqrt{\frac{2W_B}{m}}}^{\infty} v_z \exp\left(\frac{w_M - \frac{1}{2} mv_z^2}{kT}\right) dv_z \qquad (8.4)$$

Setting

$$u = \frac{w_M - \frac{1}{2} mv_z^2}{kT}$$

$$du = \frac{-m}{kT} v_z dv_z$$

(8.4) becomes

$$J_{TH} = \frac{4\pi mek^2}{h^3} \int_{\infty}^{-(W_B - w_M)/kT} e^u \, du$$

the value of which is

$$J_{TH} = \frac{4\pi mek^2}{h^3} T^2 e^{-\frac{W_B - w_M}{kT}} \qquad (8.5)$$

Setting $A_o = \frac{4\pi mek^2}{h^3}$ and $\frac{W_B - w_M}{e} = \varphi$

we have a well-known relation generally called the Richardson equation.

$$J_{TH} = A_o T^2 e^{-\frac{\varphi e}{kT}} \qquad (8.6)$$

φ is called the work function, and $e\varphi$ is clearly the minimum energy which must be added to an electron at 0°K in order for it to escape from a metal.

In order to test this equation against experiment, we can plot the experimental values of $\log_{10} \frac{J_{TH}}{T^2}$ against $\frac{1}{T}$. The experimental points should fall on a straight line of slope $- .4343\varphi \frac{e}{k}$ and should intercept the vertical axis at $\log_{10} A_0$. This serves to determine the value of φ, which cannot be predicted by statistical means. However, A_o depends only on universal constants and should equal 1.2×10^6 amp/m^2 - deg^2 or 120 amp/cm^2 - deg^2. Experimentally this value is seldom, if ever, attained for clean metals. The value of 60 is obtained for pure tungsten and several other metals, values of A between 30 and 200 have been reliably observed for many metals.

8.3 Departures of A from Theoretical

Since the value $A_o \cong 120$ amp/cm^2 - deg^2 is not obtained experimentally it will be better to rewrite our equation (8.6) as,

$$J_{TH} = AT^2 \, e^{-\varphi e/kT} \tag{8.7}$$

where A is the experimental value and since $e/k \cong 11{,}600^\circ$K/volt it is sometimes convenient to write

$$J_{TH} = AT^2 \, e^{-b/T} \tag{8.8}$$

where b $11{,}600\,\phi$.

Previous to the development of the quantum (Fermi-Dirac) statistics in addition to the equation (8.8) another form of Richardson's equation was in vogue, namely,

$$J_{TH} \cong AT^{\frac{1}{2}} \, e^{-b/T} \tag{8.9}$$

Experimentally, it is difficult, or impossible, to determine whether the T^2 term of (8.8) or the $T^{\frac{1}{2}}$ term of (8.9) is correct. This seems strange until we realize that the termperature dependence of the exponential term is so strong that variations in thermionic emission due to the $T^{\frac{1}{2}}$ or T^2 term are rather well masked over the range of temperature and types of metals avaialble to us. The T^2 term is generally used because of its basis in quantum theory.

Just before the introduction of F-D statistics a quantum theory of thermionic emission was developed which led to the same form as equation (8.8), but the value of 60 amp/cm^2 - deg^2. This deviation came about because the effect of electron spin was not fully realized. (Compare the factor 2 in equation (7.18)). But because several metals gave values of $A \cong 60$ and none gave $A \cong 120$ the validity of the quantum statistics in this application was suspect for sometime. Nowadays it is generally believed that the experimental values of 60 are completely fortuitous.

If this is indeed the case we must ask ourselves why the theory does not give the observed values of A. There are at least three independent, and probably equally valid, reasons why this is possible, and probably combinations of these three effects are present for all metals.

8.3.1 Temperature Dependence of Work Function.

The most intellectually pleasing reason for $A \neq 120$ is the temperature dependence of the work function. Since metals undergo changes of volume and of crystal structure with changes in temperature we might reasonably expect that the work function should also depend (slightly at least) on T. Setting

$$\varphi = \varphi_0 + \alpha T$$

in equation (8.6) we have

$$J_{TH} = A_0 T^2 e^{-(\varphi_0 + \alpha T)e/kT} = (A_0 e^{-\alpha e/k}) T^2 e^{-\varphi_0 e/kT}$$

The constant in front is then what we would measure as A and φ_0 is what we would measure as the work function. For $A = \frac{1}{2} A_0 = 60$, we have

$$e^{-\alpha e/k} = \frac{1}{2}$$

$$e^{-11,600\alpha} = \frac{1}{2}$$

or

$$\alpha \approx 10^{-4}/{}^\circ K,$$

a reasonable value. Further this effect would explain values of $A > 120$ since α could be negative.

8.3.2. Reflection of Electrons at Surface

Classically, electrons having sufficient energy to climb the barrier at the surface will assuredly contribute to thermionic emission. On the basis of quantum theory this is not strictly true; since the electrons have wave properties there is always a chance that some of them will be reflected at the surface and hence will not leave the metal. If the percentage reflected is r, and if r is independent of v_z, we would have to modify A_0 by

$$A = A_0(1 - r)$$

This can explain measured values of A less than 120, but not greater.

We cannot compute r by statistical means, but only by the methods of quantum theory. Strangely, the image barrier can produce but little reflection; the square barrier appreciable. In fact any potential barrier that varies more rapidly than $\frac{1}{z}$ (the image potential) can produce large reflection. Since there is independent evidence of the existance of reflection, it may be that the field varies rapidly near the surface but becomes similar to the image field at distances greater than several atomic diameters.

8.3.3. Effect of Patchiness of the Surface.

Perhaps the most important effect is the fact that no metal surface is completely uniform. Even if the surface is free of impurities different crystal faces more than likely have different work functions. Hence our experimental curve should be the sum of several curves such as given by equation (8.6), and although actually a curved line rather than a straight line results from a plot of $\log_{10} \frac{I}{T^2}$ vs. $\frac{1}{T}$ it is not difficult to see that over a limited range of T the composite curve should look like a straight line. Because the work function appears in exponential, the dominating term, we would in general measure the work function of the surface of lowest work function, but the measured value of A would very likely be less than 120.

8.4 Energies of Escaped Electrons

We should expect that the emitted electrons would follow a Maxwell-Boltzmann distribution and indeed measurements confirm this. The resenn for this is that the emitted electrons all come from the top of the distribution curve where the exponential term is large. For this case the 1 that occurs in the denominator for most of the F-D functions can be neglected compared to the exponential, and hence the form reduces to the M-B case as was shown in Section 7.5.

Very careful measurement by Nottinghaam and others have shown that the observed curve differs somewhat from the predicted curve. If one takes the ratio of the observed to the predicted distribution function $n(v_z)$ for various values of v_z, and calls this ratio $r(v_z)$ one determines the reflection coefficient of the barrier for electrons of specific velocity perpendicular to the surface. The results thus obtained for $r(v_z)$ are similar to those predicted by the square barrier; actually one need only assume a barrier varying more rapidly than $1/z$. The experimental side of measuring energy distribution will be left to the discussion of the photoelectric effect.

Since the electrons come from the tail of the distribution function, and since this tail has M-B characteristics one could expect that the average energy of the escaping electrons is of the order of kT. Exact analysis shows this average energy is 2kT. Since this represents energy carried away from the surface of amount

$$P_o = \frac{J_{TH}}{e} 2kT \quad \text{watts/m}^2$$

which is numerically equal to about 4×10^{-2} watts/cm^2, for $T = 2000°K$ and $J = 100$ ma/cm, this energy transport should produce a definite cooling of the cathode. Such a cooling effect has been confirmed experimentally.

8.5 Contact Potential Difference

It has been a long established experimental fact that if we place two dissimilar metals in contact, there will be a difference of potential between their faces which are not in contact. This is called the contact potential difference, CPD. Experimentally, we know that this difference is the same always for two given metals, A and B. Experimentally, let us form a condenser by placing two metal plates A and B a very small distance apart and let us connect them through a light flexible wire andaa ballistic galvanometer as shown in Figure 8.4.

Now if a CPD exists charge will be induced on the

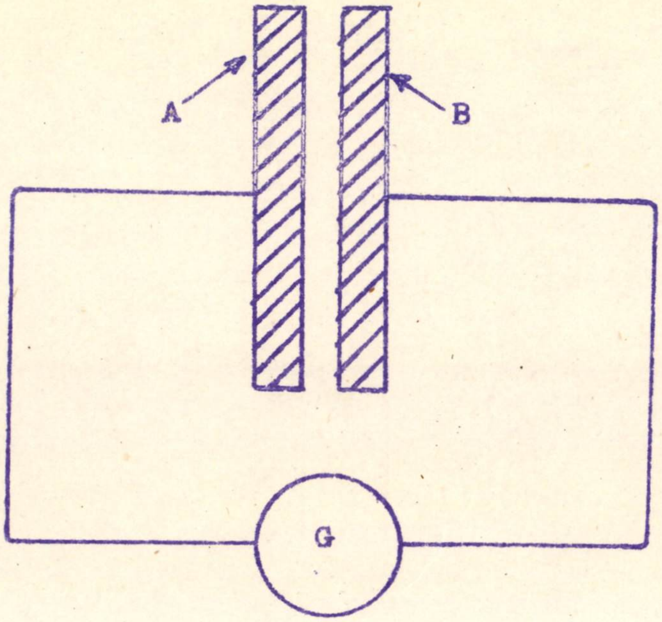


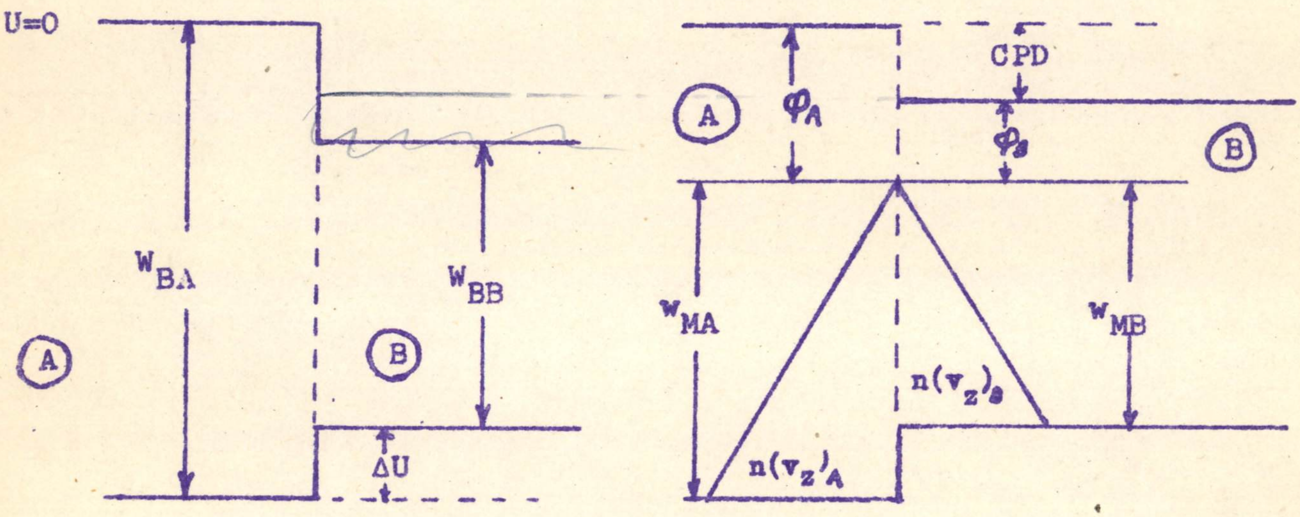
Figure 8.4

condenser plates of amount

$$Q = C_o(CPD)$$

where C_o is the capacity of the plates. If now one plate is suddenly pulled away from the other plate, the capacity drops to almost zero and the charge Q will flow through the galvanemeter. The CPD thus determined is of the order of a volt (0-3 volts) for all dissimilar metal pairs.

Our model of a metal leads to an explanation of this phenomenon. Let us consider the function where the two metals make contact, and let us assume that the temperature is quite low. We know that if we apply only a slight potential difference across this junction electrons (currents) flow freely pver it. Hence, there is no force at the barrier large enough to prevent electron flow. However, the electrons in the metal have varying kinetic energies from zero to several e.v., and by Pauli Principle only those at the top of theF-D distribution can take up small amounts of energy from an applied electric field. (Those electrons below the top of the F-D distribution cannot acquire energy by increments since the allowed energy states just above them are already filled). Therefore, there can be a force at the contact providing the potential energy difference due to this force does not exceed w_M for either metal. Let us call this potential energy difference ΔU. The potential energy situation at the junction is illustrated in Figure 8.5(a), where the subscripts A and B refer to the



(a) (b)

Figure 8.5

different metals and we have set the potential energy outside metal A equal to zero.

If we take the z direction as normal to the junction the energy conservation law states

$$\frac{1}{2} mv_{zA}^2 = \frac{1}{2} mv_{zB}^2 + \Delta U \qquad (8.10)$$

When no electric field is applied across the junction we know that no net current flows and hence the number of electrons per unit area per sec. leaving A for B must just equal the number of electrons leaving B for A. But since

Pauli's principle applies in both A and B an electron of one particular V_{zA} leaving metal A must be replaced by an electron from metal B which will have the same v_{zA} in metal A; i.e. equation (8.10) will hold. Now this condition can be fulfilled if, and only if, we have

$$n'_A(v_{zA})dv_{zA} = n'_B(v_{zB})\,dv_{zB}$$

or

$$v_{zA}\,n_A(v_{zA})\,dv_{zA} = v_{zB}\,n_B(v_{zB})dv_{zB} \qquad (8.11)$$

That is, the number of electrons of velocity component v_{zA} in the range dv_{zA} coming to unit area of surface A per second must just equal the number of electrons of velocity component v_{zB} in the range dv_{zB} coming to unit area of surface B per second, where the conservation law (8.10) relates v_{zA} and v_{zB}.

Taking the differential of (8.10) and remembering that ΔU is a constant we have,

$$v_{zA}dv_{zA} = v_{zB}\,dv_{zB}$$

Combining this result with (8.11), we have

$$n_A(v_{zA}) = n_B(v_{zB}) \qquad (8.12)$$

Now (8.12) need only be satisfied for these electrons which can be exchanged for energy reasons, i.e., for those in which

$$\frac{1}{2}\,mv_{zA}^2 > \Delta U.$$

Equation (8.12) can only be satisfied if the electrons in both metals at the top of the Fermi band have exactly the same potential energy relative to infinity. Plotting $n(v_z)$ for both metals on our energy diagram this situation is illustrated in Fig. 8.5(b). Because of this an electron just outside metal B has its potential energy relative to infinity depressed by an amount just equal to the difference in the work function, i.e.

$$CPD = \varphi_A - \varphi_B$$

The algebraic sign of the contact potential difference is such that the electrons just outside the metal of lower work function are at a lower potential energy than the electrons just outside the metal of higher work function. That is to say that electrons leaving a metal of work function ϕ_B will be <u>retarded</u> by a potential $\phi_A - \phi_B$, where ϕ_A is the work function of the collector. In the usual experimental set-up the emitter has a lower work function than the collector and the CPD is therefore a retarding potential.

8.6 The Effect of Retarding Fields

If an electric field is applied such that the potential of the collector is lower than that of the emitter, electrons leaving the emitter will be acted upon by a force tending to return them to the emitter. Fig. 8.6 illustrates the potential energy situation for this case.

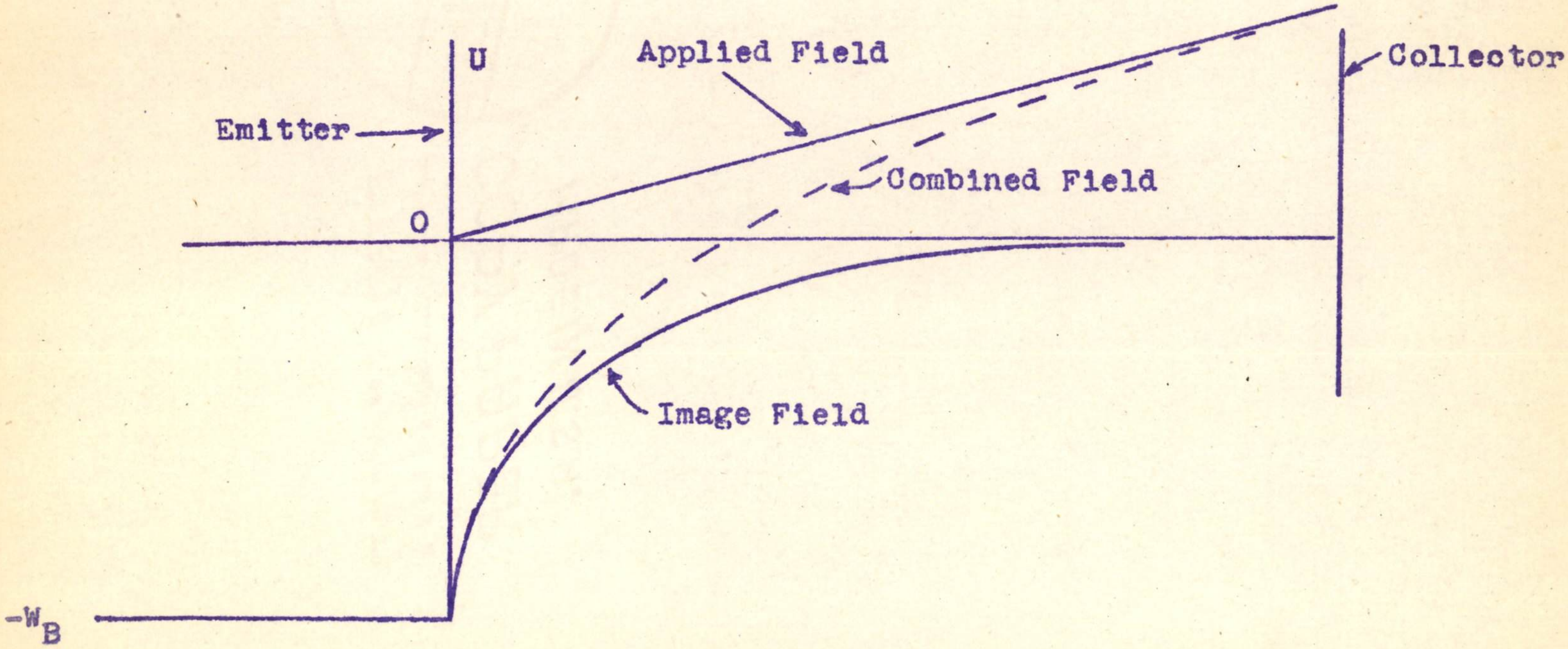


Figure 8.6

The net effect of the applied retarding field is to raise the height of the potential energy barrier over which the electrons must climb. Hence if V_A, the applied potential, is retarding, the net effect is to increase the work function by an amount V_A. Hence, in Richardson's equation (8.7) we should replace ϕ by $\phi + V_A$ and obtain

$$J = J_{TH}\, e^{-\frac{eV_A}{kT}}$$

where J_{TH} is the saturation (zero field) electron current.

Retarding potential methods may be made use of in determining the energy distribution of the emitted electrons. If we have parallel plate geometry and take z as the direction perpendicular to the plates, then the current J(V) measured at a particular retarding potential V represents the current due to all electrons which satisfy the condition.

$$\frac{1}{2} m v_z'^2 \geqslant Ve \qquad (8.13)$$

where v_z' is the velocity component outside the metal. Hence if $n(v_z')dv_z'$ represents the distribution function in v_z' outside the metal

$$I(V) = e \int_{\sqrt{\frac{2eV}{m}}}^{\infty} v_z'\, n(v_z')\, dv_z' \qquad (8.14)$$

is the current due to electrons for which (8.13) is satisfied. It can be shown that

$$n(v_z') = -\frac{dI(V)}{dV}$$

where v_z and V are functionally related by $\frac{1}{2} m v_z'^2 = Ve$. If concentric spherical geometry is used and if the emitter is small compared to the collector, one can measure the total energy (or speed) distribution function.

8.7 The Effect of Accelerating Fields

Schottky Effect

If one applies a field which is favorable to electron emission the situation is not as simple as an additive (or subtractive) change in work function. Fig. 8.7 shows the potential energy diagram for this case.

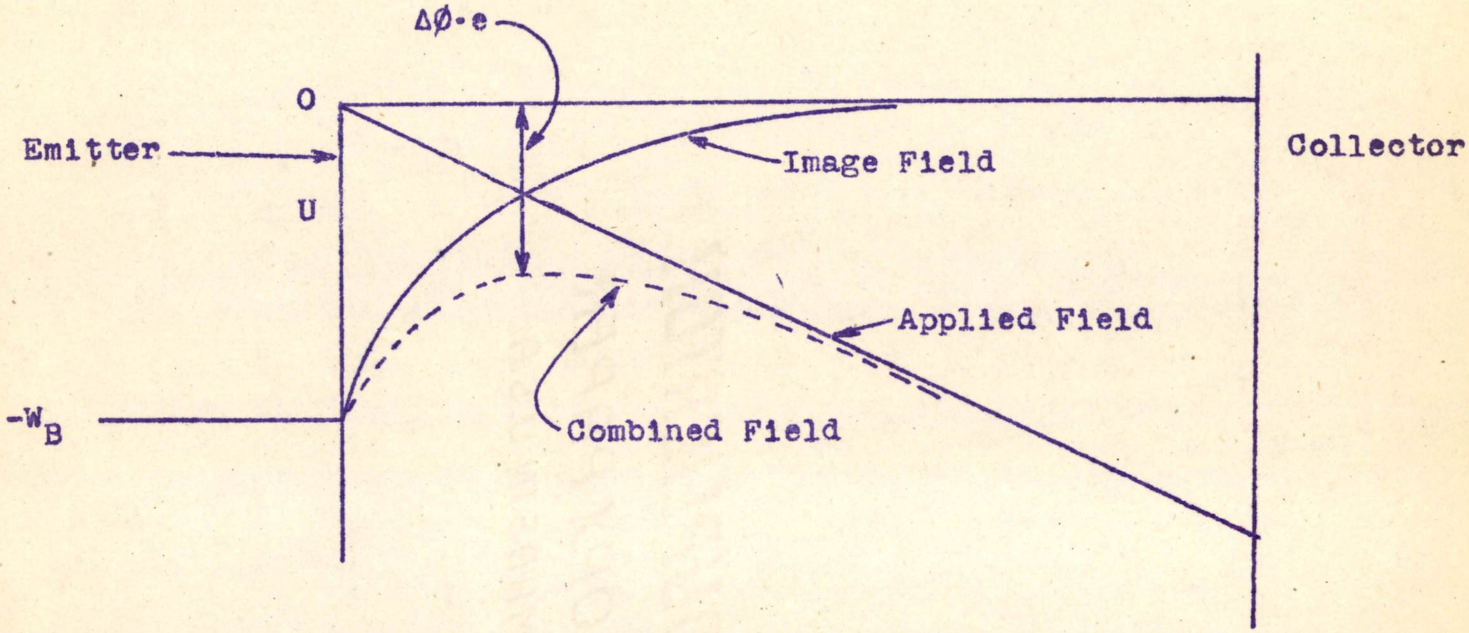


Figure 8.7

The figure illustrates (to an exaggerated scale) the fact that the potential energy barrier at the surface of the metal is reduced, but not by the amount of the applied potential. In fact the reduction of work function is very small, but not unimportant. To compute this reduction let us assume that the geometry is plane parallel so that the applied potential energy is,

$$eV_A = - e\cdot E_A z$$

where E_A is the magnitude of the applied uniform field. The image barrier is

$$U = \frac{- e^2}{16\pi\varepsilon_o z} ,$$

and the net potential energy is

$$U_S = - e E_A z - \frac{e^2}{16\pi\varepsilon_o z} \qquad (8.15)$$

To determine $\Delta\phi$, the lowering of the work function, we must compute the maximum value of the combined potential energy, $\Delta\phi e$. Differentiating U_S with respect to z and setting equal to zero

$$\frac{dU_S}{dz} = - e\cdot E_A + \frac{e^2}{16\pi\varepsilon_o z^2} = 0$$

or

$$z_m = \sqrt{\frac{e}{16\pi\varepsilon_o E_A}}$$

for maximum. Therefore,

$$U_{S(max)} = - 2 e \sqrt{\frac{e E_A}{16\pi\varepsilon_o}}$$

$$\Delta\phi e = U_{S(max)} = - 2e \sqrt{\frac{e E_A}{16\pi\varepsilon_o}}$$

$$\Delta\phi = -\sqrt{\frac{e}{4\pi\varepsilon_o}} \cdot E_A^{\frac{1}{2}} = - 3.79 \times 10^{-5} E_A^{\frac{1}{2}}$$

Hence, for an applied field of 10^6 volts/m the lowering of the work function is only about .04 volt. However, this small charge of work function can produce large changes in the emission current. If in equation (8.7) we replace ϕ by $\phi + \Delta\varphi$ we have

$$J = J_{TH}\, e^{-\frac{\Delta\phi e}{kT}}$$

but

$$-\Delta\phi \frac{e}{k} = 3.79 \times 10^{-5} \times 11{,}600\, E_A^{\frac{1}{2}} = .44\, E_A^{\frac{1}{2}}$$

Hence,

$$J = J_{TH}\, e^{\frac{.44\, E_A^{\frac{1}{2}}}{T}} \qquad (8.16)$$

The validity of this theory, first given by Schottky, can be tested by plotting the logarithm of the emission current against the square root of the applied field (or of the applied voltage for constant geometry). A straight line of slope $\frac{.44}{T}$ should result. This is actually the case ~~of~~ for clean metal surfaces and Schottky's theory is amply justified. In fact, the best measurements of saturation currents are made by means of this theory.

As the temperature of a cathode is raised the emission current density becomes large and space charge effects become appreciable. Hence, one must go to high values of applied potential before collecting all the thermionically emitted electrons. But in going to high potentials one has enhanced the emission by Schottky effect. The only escape is to go to still higher potentials, and then to make a <u>Schottky plot</u> of $\ln J$ vs. $E_A^{\frac{1}{2}}$. By extropolating this back to zero applied field, neglecting the data in the space charge limited region, one arrives at the true value of the saturation current.

8.8 Field Emission (Cold emmission)

If one goes to applied fields of the order of 10^7 volts/m and greater, one finds that the emission is much greater than that predicted by Schottky theory. In fact even if the temperature is very low, appreciable emission is observed. Such an effect has no explanation on the basis of classical theory but can be explained if one considers the wave properties of the electrons. Under very strong fields the potential energy diagram at the surface of the metal looks something like the diagram of Fig. 8.8. (Still somewhat exaggerated). The Fermi levels of the free electrons are indicated on the diagram for a very low temperature.

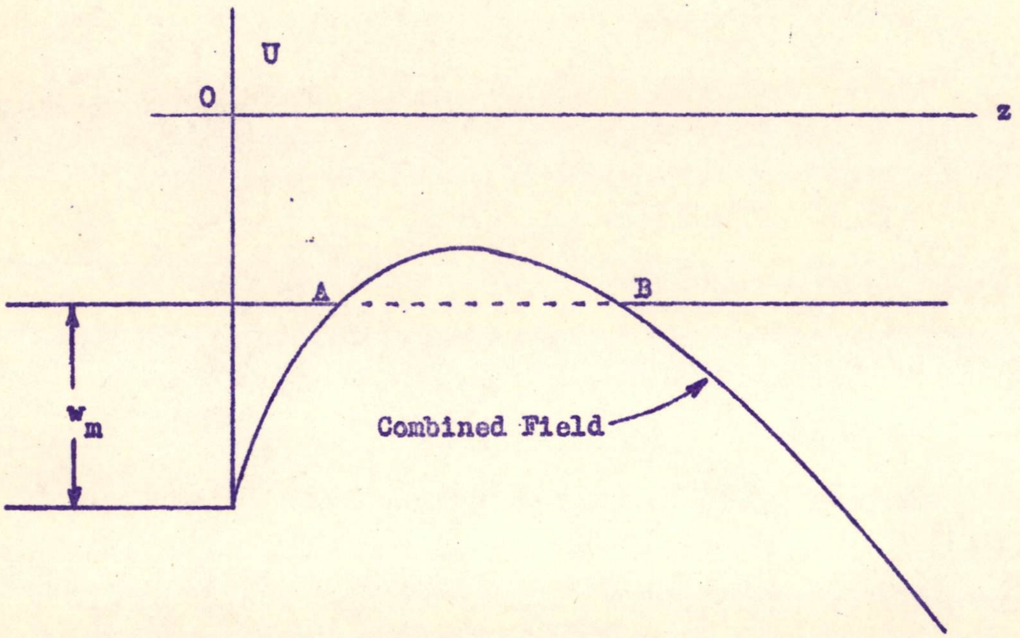


Figure 8.8

Classically the electrons do not have enough energy to surmount the barrier and hence could not escape from the metal. However, electrons with wave properties have a finite chance of penetrating the barrier, from A to B for the most energetic electrons, and appearing outside the metal. Such an effect is called the tunnel effect, for obvious reasons and the detailed theory is due to Fowler and Nordheim. The resulting current density is predicted as,

$$J = C \cdot E_A^2 \, e^{-\frac{D}{E_A}}, \text{ amps/m}^2 \qquad (8.17)$$

where the applied field E_A plays a role analogous to that of the temperature in thermionic emission. The constants C and D are dependent on the particular metal and are

$$C = \frac{6.2 \times 10^{-6}}{W_B} \left(\frac{w_M}{\phi}\right)^{\frac{1}{2}} \quad \text{amps/volt}^2$$

$$D = 6.8 \times 10^{9} \phi^{3/2} \quad \text{volts/m}$$

where W_B snf w_M are in units of electron volts.

8.9 Secondary Emission

If a metallic surface (or insulator for that matter) is bombarded by electrons, one finds that secondary electrons are emitted by the surface, and that this effect is essentially independent of the temperature. Some of the electrons have energies equal to the primary electron and are truly reflected electrons, but for the most part the secondary electrons have energies of the order of 5 e.v., independent of the energy of the primaries. Furthermore, at primary electron energies around 500 e.v. the number of secondaries is about twice the number of primaries for most metals. For Beryllium, or more properly for Beryllium with a thin oxide film, the number of secondaries can reach 4 or 5 times the number of primaries.

The explanation of this effect must be due to successive momentum transfer. A primary electron strikes several of the electrons in the metal and imparts momentum to them, but in a direction away from the surface. These electrons can, however, strike the atoms of the crystal lattice and bounce back, retaining the additional energy received from the primary electron. If this energy is greater than the barrier potential energy the electrons can escape from the metal.

This point of view is strengthened by experiments which have measured the secondary emission as a function of the angle of incidence of the primary beam. These experiments show that the component of velocity perpendicular to the surface is the determining element for secondary emission.

Similar results are obtained when heavy particles (say protons) are used in the primary beam. However, the energy of primary electrons for maximum emission ratio is about 1000 times higher (of the order of the mass ratio) which also confirms the momentum transfer point of view.

8.10 Photoelectric Emission

When metals are illuminated with visible or ultra violet light they emit electrons. This effect was discovered in the late 19th century by Hallwachs, who observed that metals, when subjected to ultraviolet radiation, became positively charged to a few volts. The number of emitted photoelectrons is strictly proportional to the intensity of the light at a given frequency; but the photo current is strongly frequency dependent. The shape of the energy distribution of the emitted electrons depends only on the frequency of the light, and not at all on the intensity, a fact that was completely at odds with classical theory.

Typical experimental curves for the emission from a clean surface of sodium are shown. Fig. 8.9(a) shows the total (saturation) photoelectric current per unit incident intensity as a function of frequency. Fig. 8.9(b) illustrates the distribution in total energy of the emitted photoelectrons at a single frequency

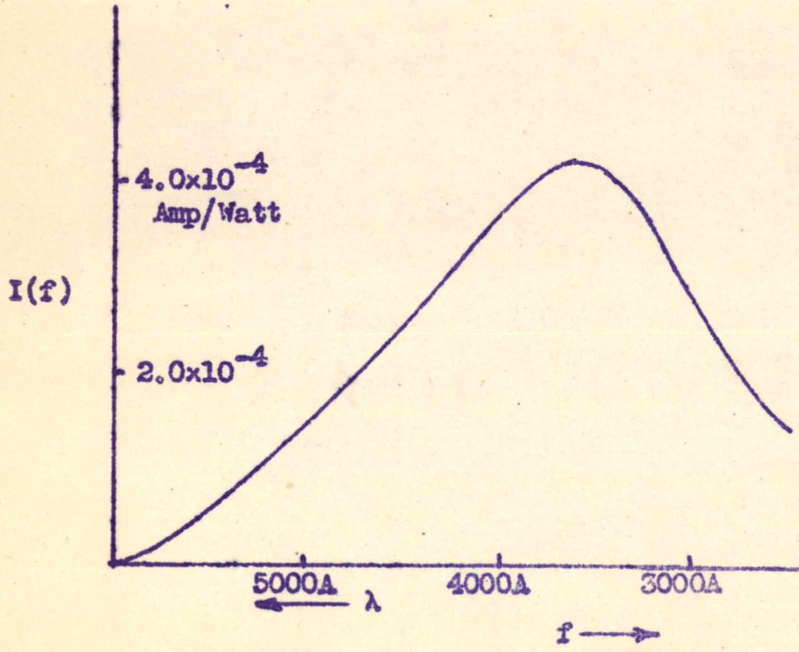


Fig. 8.9(a)

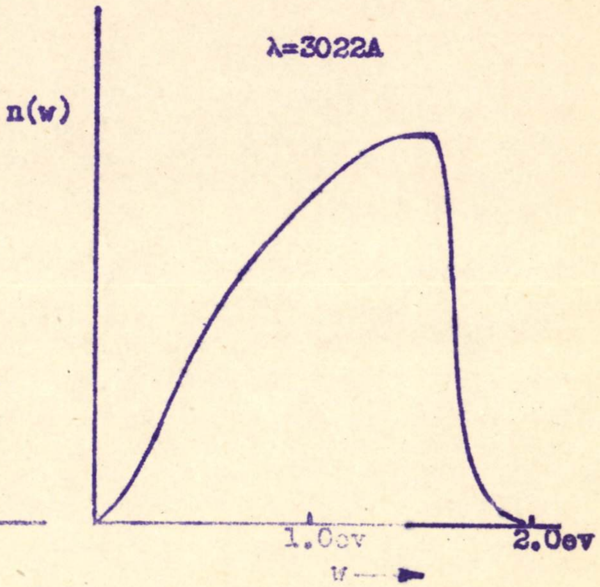


Fig. 8.9(b)

The general explanation of ~~three~~ these results can be obtained from our statistical model of a metal; for a detailed explanation we must use the methods of quantum mechanics. Consider the energy diagram for the electrons near the surface of the metal at a low temperature, in Fig. 8.10.

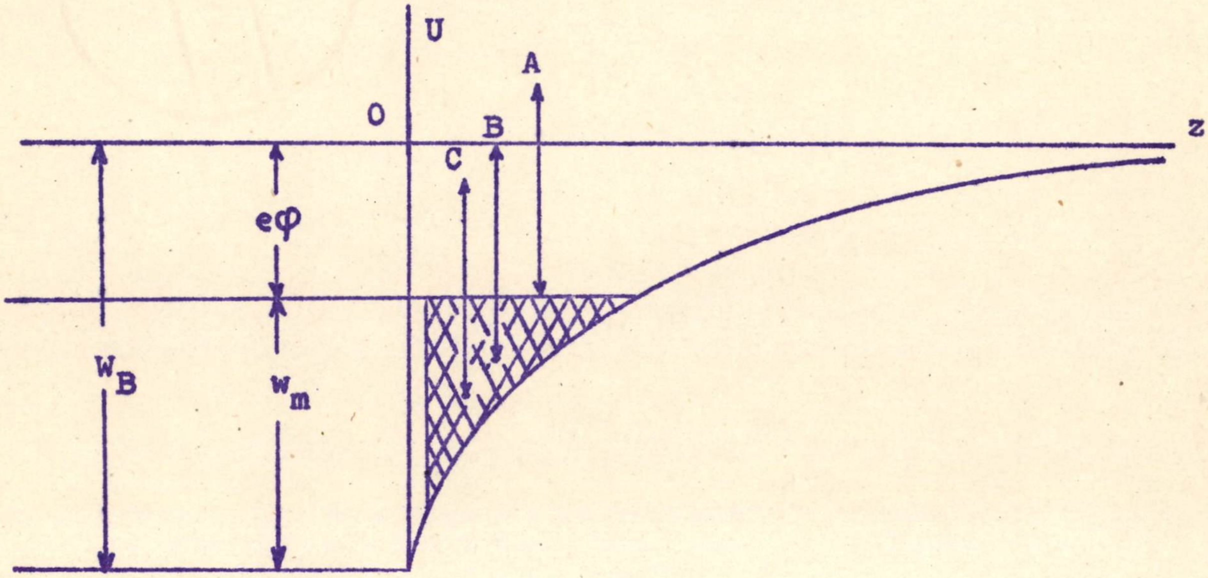


Fig. 8.10

The free electrons cannot absorb quanta (see section 5.2A) and hence only those electrons near the surface, the cross hatched area, can absorb light since such electrons are bound by the image force. Actually, most of the absorption of light near the visible by metals is due to these surface electrons, but only a small proportion of the absorption leads to photoelectrons. From our diagram, if the metal is illuminated by light of frequency f such that $hf < \phi e$, it is apparent that the energy transfer to the Fermi electrons is insufficient to permit their escape. If, however, $hf \geqslant \phi e$ it is still unlikely that all, or even most of the absorbed quanta will lead to photo emission. Even if the electron gains enough energy to escape it may be travelling in the wrong direction. For escape we must have that

$$\frac{1}{2}\, mv_z^2 + hf = \frac{1}{2}\, mv_z'^2 \geq W_B \qquad (8.18)$$

and v_z' must be positive. Also the electron may have low initial energy and hence even after absorption of hf condition (8.18) may not be satisfied. This situation is illustrated by the three vertical arrows in Fig. 8.10. Each represents a quantum of energy hf, but each is interacting with a different electron, indicated by the bottom arrow. Quantum A, absorbed by an electron at the top of the Fermi band, can produce an energetic photoelectron; quantum B, absorbed by an electron of intermediate energy can produce a photoelectron with just enough energy to escape; quantum C, absorbed by an electron near the bottom of the barrier cannot produce a photoelectron.

Quantum A produces the most energetic photoelectron, of energy

$$\frac{1}{2}\, mv_{MAX}^2 = hf - \phi e \qquad (8.19)$$

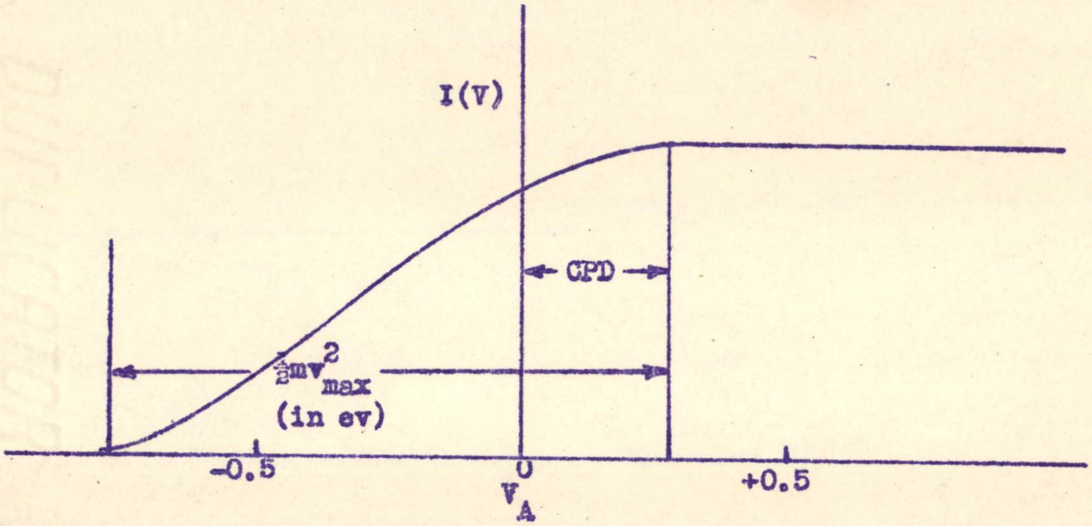
This relation was predicted by Einstein in 1905 (20 years before F-D statistics) and was the first <u>clear</u> statement of quantum theory. In 1915 Millikan measured the maximum energy of photoelectrons as a function of frequency, by measuring the retarding potential, V_A, necessary to stop the fastest photoelectron. A plot of V_A against f yields a straight line of slope h/e, and Millikan's was the first good measurement of this quantity. Actually, of course, $\frac{1}{2}\, mv_{MAX}^2$ is not a unique quantity; at normal temperatures there is always a slight temprature "tail" as shown in Fig. 8.9(b).

In measuring the complete energy distribution function one can use concentric spherical electrodes, where the emitter is about a point source. The photoelectrons emerge radially, and if a retarding potential V_A is applied one collects all photoelectrons of energy eV_A and greater. If the photo current is $I(V_A)$ the energy distribtuion is given by

$$n(w') = - \frac{dI(V)}{dV}$$

where w' is the energy after escape.

Since n(w') can be computed from the known distribution inside, this provides a check on the F-D theory. Very good agreement is obtainable between theory and experiment; there is some divergence which must be attributed to a barrier reflection coefficient. This implies that the barrier departs from the image field picture near the surface. A typical current-voltage relation is shown in Fig. 8.11, where we have plotted the photocurrent against the applied voltage.



Fig, 8.11

The current does not saturate at the zero of applied voltage because of contact potential difference.

The spectral distribution curve of Fig. 8.9(a) cannot be predicted by statistical methods alone, except near the threshold (the long wave length limit) of photoemission. The threshold wave length is related to the work function by the familiar relation

$$\lambda_o = \frac{12,400}{\phi} \qquad (8.20)$$

As the frequency of the incident light is increased from the threshold, we should expect increased emission at first since the number of electrons which can enter into the emission process increases rapidly. However, quantum mechanics shows that the absorption probability falls off with increasing frequency. This decrease in absorption probability predominates after a while and accounts for the maximum in the spectral distribution curve, and the subsequent diminution of the total photocurrent.

¡SELAH!