ON THE ELECTRICAL RESISTANCE OF CONTACTS BETWEEN SOLID CONDUCTORS

By J. Frenkel
Department of Physics, University of Minnesota

(Received October 30, 1930)

Abstract

A contact between two solid conducting bodies is visualized as a small gap between them. This gap can be described as a potential-hill over which electrons, according to the wave-mechanical theory, can pass even with insufficient kinetic energy. The general expression of the resulting current intensity as function of the potential-difference is obtained and discussed for the case of two identical or different bodies in connection with the resistance of granular structures (thin metallic films) and the rectifying action of certain contacts.

Introduction

The usual picture of an electrical contact between two solid conducting bodies is that their surfaces or part of their surfaces are at a distance of atomic dimensions from each other, so that the electrons can pass through the contact surface in the same way they pass through any surface within the same body.

Now such an intimate contact between two bodies along a large part of their surfaces is probably very rarely realized. Nor is it necessary for the conduction of electricity from one body to the other. Such conduction can also take place through those parts of their surfaces which lie rather far apart from each other, that is, at a distance many times larger than the usual atomic distance. In fact, according to a well-known principle of wave mechanics, which has been used already (and sometimes abused) for the explanation of a great many phenomena, an electron can jump over a "potential-hill" even if it does not have sufficient kinetic energy to do so according to the classical mechanics. Now the gap between two contiguous bodies may be considered as the top of such a hill, with practically vertical slopes at (or rather just beyond) the respective surfaces. There must be in general a steady flow of electrons across the gap in both directions, the difference between the two flows being the actually observed current intensity I. In the case of equilibrium the latter is of course equal to zero. If, however, an additional potential difference ϕ is maintained across the gap, I will be a certain function of ϕ, different from zero.

It will be our first object to determine the general character of this function $I(\phi)$. Before, however, proceeding further let us remark that this function can always be expanded in a power series and that for small values of ϕ one can simply put $I=\alpha_1\phi$ in accordance with Ohm's law, α_1, being the conductivity of the contact, that is the reciprocal of its electrical resistance. For larger values of ϕ one must get

$$I = \alpha_1\phi + \alpha_2\phi^2 + \alpha_3\phi^3 + \cdots .$$

The coefficients of the even powers of ϕ vanish in the case of a contact between two identical bodies. They must be however different from zero in the case of two bodies of different nature. One thus gets in this case, to the second approximation, $I = \alpha_1\phi + \alpha_2\phi^2$, which means that the current is changed in magnitude when the sign of ϕ is reversed. This means a rectification effect of the same type as that given by an electron valve on the curved part of the characteristic, and may be quite large for some particular contacts.

1. General Theory

For the sake of simplicity we shall consider the contiguous surfaces of the two bodies (a, b) as two parallel infinite planes. Their distance apart will be denoted by δ. The potential energy curve will be represented by the full line $MNQRST$ (Fig. 1) $NQ = U_a$ and $SR = U_b$ denote the increase of potential energy of an electron crossing the surface of the respective body (from inside to the outside). The inclined line QR represents a homogeneous electrical field acting between the two bodies. The corresponding change of the poten-

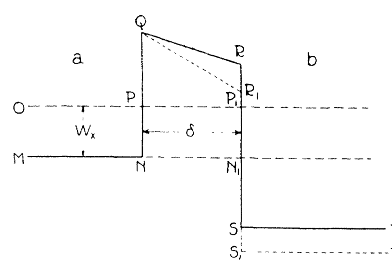

Fig. 1.

tial energy of an electron passing across the gap from a to b will be denoted by V and will be reckoned positive if this energy is diminished (as shown on the figure), that is, if the force F acting on the electron in the gap is directed from a to b ($F = V/\delta$). In the following it will be always supposed that the potential energy of an electron in a is *higher* than in b; the difference represented by the line N_1S_1 will be denoted with U_{ab}. One has the obvious relation $U_b - U_{ab} + V = U_a$ or

$$U_{ab} = U_b - U_a + V \tag{1}$$

$NP = W_x$ represents the kinetic energy of an arbitrarily selected electron in a, or rather that part of its kinetic energy W, which corresponds to the x-component of the velocity, the x-axis being drawn in the direction ab. *Every* electron in a for which this component v_x is positive, will be able to pass through the gap, no matter how small W_x is, in comparison with the height of the potential wall NQ. On the other hand only those electrons of b will be able to jump over the gap to a for which the part of the kinetic energy corresponding to the x-component of the velocity (in the negative direction) is larger than $SN_1 = U_{ab}$.

The "transmission coefficient" for the a-electrons with the energy W_x will be denoted with $D(W_x)$. It is equal to the probability of any such electron passing from a to b (or the fraction of all the electrons succeeding in this enterprise). On the other side of the gap, that is in b, such an electron will have for the x-direction the kinetic energy W_x+U_{ab}. Reversing the direction of its motion we should get the same probability $D(W_x)$ for its getting back to a.[1] We thus see that $D(W_x)$ is the probability of a b-electron having for the negative x-direction the kinetic energy W_x+U_{ab}, to pass across the gap to a. The number of electrons per unit volume having velocity components in dv_x, dv_y, dv_z will be denoted by $f_a dv_x dv_y dv_z$ and $f_b dv_x dv_y dv_z$ for a and b respectively. So far as the *change in the velocity distribution of the electrons in each body*, due to the passage of the electrons to the other body or from the latter, can be neglected, f_a and f_b can be treated as functions of the resulting kinetic energy $W=W_x+W_y+W_z=\frac{1}{2}m(v_x^2+v_y^2+v_z^2)$. In case of two metals these are the well-known Fermi-Pauli-Sommerfeld functions.[2]

$$f_a = 2\left(\frac{m}{h}\right)^3 \frac{1}{e^{W/kT}/A+1}, \quad f_b = 2\left(\frac{m}{h}\right)^3 \frac{1}{e^{W/kT}/B+1}. \tag{2}$$

The number of electrons passing from a to b per unit surface per second is equal to

$$I_1 = \int_0^\infty dv_x \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dv_y dv_z D(W_x) f_a v_x.$$

For the corresponding number of electrons passing from b to a we get

$$I_2 = \int_{v_x^0}^\infty dv_x \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dv_y dv_z D(W_x - U_{ab}) f_b v_x.$$

Where v_x^0 is defined by the condition $\frac{1}{2}m(v_x^0)^2=U$. These expressions may be simplified by introducing instead of v_x the variable $W_x=\frac{1}{2}mv_x^2$ and instead of v_y and v_z the variable $R=W_y+W_z=\frac{1}{2}m(v_y^2+v_z^2)$ and the angle $\phi=\arctan(v_y/v_z)$. It will be remarked that $(2R/m)^{\frac{1}{2}}$ and ϕ are the polar coordinates replacing the rectangular coordinates v_y, v_z. We get then

$$I_1 = \frac{2\pi}{m^2}\int_0^\infty dW_x D(W_x) \int_0^\infty f_a(W_x+R)dR \tag{3}$$

and

$$I_2 = \frac{2\pi}{m^2}\int_{W_x^0}^\infty dW_x D(W_x - U_{ab}) \int_0^\infty f_b(W_x+R)dR$$

or replacing W_x by $W_x{}^1 = W_x - U_{ab}$,

$$I_2 = \frac{2\pi}{m^2}\int_0^\infty dW_x' D(W_x') \int_0^\infty f_b(W_x'+U_{ab}+R)dR. \qquad 4)$$

[1] Cf. J. Frenkel, Einführung in die Wellenmechanik, p. 57.

[2] Cf. A. Sommerfeld, Zeits. f. Physik **47**, 7 (1928).

The resulting flow of electrons from a to b is

$$I = I_1 - I_2 = \frac{2\pi}{m^2}\int_0^\infty dW_x D(W_x)$$
$$\int_0^\infty [f_a(W_x + R) - f_b(W_x + U_{ab} + R)]dR. \quad (5)$$

We shall now suppose that $I = 0$. This means that the two bodies are in a statistical equilibrium with respect to each other, as a result of the existence of a definite contact potential difference between them, corresponding to the drop of potential energy V in the gap.

This state of equilibrium must obviously be independent of the special shape of the function $D(W_x)$, which determines the velocity with which it is established or the time of relaxation. Therefore in the case of equilibrium the coefficient of $D(W_x)$, that is the integral over R in (5), must vanish for any value of W_x, whence it follows that the integrand must vanish. We get thus, as the condition of equilibrium

$$f_a(W) = f_b(W + U_{ab}) \quad (6)$$

This equation can be considered as the direct consequence of the principle of detailed balance.

Let us now assume that V is increased by the amount V_1, corresponding to an additional (external) potential difference $\phi = V_1/e$ (e = charge of an electron). Instead of the initial potential energy curve Fig. 1, we shall get in this case the curve $MNQR_1S_1T_1$ (partially dotted line) if V_1 is positive (which it, of course, need not be). This will alter to some extent the transmission coefficient $D(W_x)$ replacing it by $D_1(W_x)$ say, and what is more important, change the potential energy difference U_{ab} replacing it by $U_{ab} + V_1$. As a result I_2 will now be smaller than I_1 (if $V_1 > 0$), and we shall have a current flowing through the gap in the direction of the applied electrical force. Taking account of the condition (6) we can determine this current by the formula

$$I = \frac{2\pi}{m^2}\int_0^\infty dW_x D_1(W_x)$$
$$\int_0^\infty [f_b(W_x + U_{ab} + R) - f_b(W_x + U_{ab} + V_1 + R)]dR \quad (7)$$

For sufficiently small values of V_1, we can put

$$f_b(W_x + U_{ab} + R) - f_b(W_x + U_{ab} + R + V_1) = -V_1\frac{\partial f_b(W_x + U_{ab} + R)}{\partial R}.$$

This reduces the inner integral in (7), in view of $f_b(\infty) = 0$ to $V_1 f_b(W_x + U_{ab})$, so that neglecting the difference between D and D_1, which is immaterial so far as second powers of V_1 are neglected, we get to a first approximation

$$\frac{I}{V_1} = \frac{2\pi}{m^2}\int_0^\infty dW_x D(W_x) f_b(W_x + U_{ab}) \equiv \alpha_1. \quad (8)$$

This expression, or rather its product with e^2, may be defined as the reciprocal of the resistance of the contact (per unit surface). Putting $V_1 = eE\delta$ where E is the (additional) electrical field in the gap, we can define the quantity

$$\frac{eI}{E} = \frac{e^2 I\delta}{V_1} \equiv \sigma_g$$

that is

$$\sigma_g = \frac{2\pi e^2 \delta}{m^2} \int_0^\infty dW_x D(W_x) f_b(W_x + U_{ab}) \tag{9}$$

as the specific "conductivity" of the gap forming the contact.

Proceeding to the second approximation, we get

$$f_b(W_x + U_{ab} + R) - f_b(W_x + U_{ab} + R + V_1) = -V_1 \frac{\partial f_b}{\partial R} - \frac{V_1^2}{2} \frac{\partial^2 f_b}{\partial R^2}$$

$$D_1 = D + \frac{\partial D}{\partial V_1} V_1$$

(the arguments being in both cases those corresponding to $V_1 = 0$) whence

$$I = \alpha_1 V_1 + \alpha_2 V_1^2 \tag{10}$$

with the previous value of α_1 and

$$\alpha_2 = \frac{2\pi}{m^2} \int_0^\infty dW_x \left[\frac{1}{2} D(W_x) f_b'(W_x + U_{ab}) + \frac{\partial D}{\partial V_1} f_b(W_x + U_{ab}) \right] \tag{11}$$

f_b' denoting the first derivative of f_b.

It must be emphasized that the above results hold for the case only that $U_{ab} + V_1$ remains positive. If $U_{ab} + V_1 < 0$ the role of the bodies a and b will be exchanged.

2. Application to the Case of Two Identical Metals and to Granular Structures (thin films)

The last remark applies in particular to the case of a contact between two identical metals, which is characterized by U_{ab} (as well as V) being equal to zero. The second term in (10) will then vanish, and I will be an odd function of V_1.

Introducing f for f_b in one of the expressions (2) we get in this case for the "specific conductivity" of the gap

$$\sigma_g = \frac{4\pi m e^2 \delta}{h^3} \int_0^\infty dW_x \frac{D(W_x)}{e^{W_x/kT}/A + 1} \cdot \tag{12}$$

It will be interesting to compare this expression with that of the usual specific conductivity of the corresponding metal σ as derived from Sommerfeld's theory. The latter expression can be put in the form

$$\sigma = \frac{e^2 l n}{m v_0} \tag{13}$$

where n denotes the number of electrons in unit volume, v_0 their maximum velocity for $T=0$, and l the mean free path of the electrons having this velocity.[3] Putting

$$n = 2\left(\frac{m}{h}\right)^3 \int\int\int_{-\infty}^{+\infty} \frac{dv_x dv_y dv_z}{e^{W/kT}/A + 1} = 8\pi\left(\frac{m}{h}\right)^3 \int_0^{\infty} \frac{v^2 dv}{e^{W/kT}/A + 1}$$
$$= \frac{8\pi m^2}{h^3}\int_0^{\infty} \frac{v dW}{e^{W/kT}/A + 1}$$

we see that the integral

$$\frac{8\pi m^2}{h^3}\int_0^{\infty} \frac{dW D(W)}{e^{W/kT}/A + 1} \tag{14}$$

can be considered as the mean value of $D(W)/v$ for all the electrons (irrespective of the direction of their velocity). We thus get, according to (12)

$$\sigma_g = \frac{1}{2}\,\frac{e^2\delta n}{m}\left(\frac{\overline{D}}{v}\right) \tag{15}$$

or approximately—putting $(\overline{D/v}) = D/v_0$

$$\frac{\sigma_g}{\sigma} = \frac{1}{2}\,\frac{\delta\overline{D}}{l} \tag{15a}$$

This relation shows, that with respect to its conductivity the gap can be treated as a metal, where the free electrons have a mean free path of the order of magnitude of $\delta\overline{D}$. One can of course use it in the opposite way and treat a metal as a series of gaps. This interpretation roughly corresponds to the theories of Bloch and Peierls, where the electrons are considered as bound to the separate atoms, but still capable of jumping from one atom to the next one over the potential hill separating them.

The transmission coefficient D in the case of two identical metals $U_a = U_b = U$, that is for an energy-curve of the shape shown by the full line line of Fig. 2, is given as a function of the energy $W_x = \frac{1}{2}mv_x^2$ so long as the

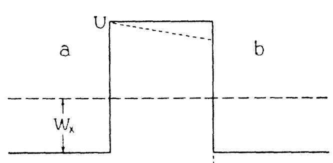

Fig. 2.

latter is smaller than U_0, by the formula

$$D = \left[\operatorname{csh}^2 \beta\delta + \frac{1}{4}\left(\frac{\beta^2}{\alpha^2} - \frac{\alpha^2}{\beta^2}\right)^2 \sinh^2 \beta\delta\right]^{-1} \tag{16}$$

[3] A. Sommerfeld, Zeits. f. Physik **47**, 1 (1928). Formulas (48c) and (42a).

where[4]

$$\alpha^2 = \frac{8\pi^2 m}{h^2} W_x, \quad \beta^2 = \frac{8\pi^2 m}{h^2}(U - W_x). \tag{16a}$$

For values of $\beta\delta$ which are large compared with 1, this reduces approximately to

$$D = 4e^{-2\beta\delta}\left[1 + \left(\frac{U}{2W_x}\,\frac{U - 2W_x}{U - W_x}\right)^2\right]^{-1}. \tag{17}$$

It may be convenient to write β in the form $\beta = 2\pi/\lambda$ where λ can be defined as the wave-length of an electron moving with the positive kinetic energy $U_0 - W_x$. If expressed in volts this energy is equal to ϕ, then

$$\lambda = \frac{1.1 \times 10^{-7}}{\phi^{1/2}} \text{ cm} = \frac{11}{\phi^{1/2}} \text{ Angstrom units.}$$

For the electrons with velocity $v_x = v_0$, that is, the maximum velocity at zero-point of temperature, which approximately corresponds to the maximum of the Fermi distribution curve in the region of usual temperatures, the difference $U_0 - W_x$ is just equal to the work function of the metal (as measured directly in the Richardson effect).[5] Putting $\phi \cong 4$ volts we get for these electrons (using Angstrom units for λ and for the distance δ):

$$\lambda \cong 5, \ \beta \cong 1.2, \ 2\beta\delta \cong 2.4\delta.$$

Since the expression in brackets in (17) is of the order of magnitude 1, we get as a rough estimate of the transmission coefficient D for $W_x = \frac{1}{2}mv_0^2$

$$D_0 \cong e^{-2.4\delta}.$$

If the mean value of D entering in (15a) could be identified with D_0, we should have

$$\frac{\sigma_g}{\sigma} \cong \frac{\delta}{l} e^{-2.4\delta}$$

that is for $\delta = 10A$ with $l \cong 100$ (which roughly corresponds to the mean free path of the electrons at room temperature)

$$\sigma_g/\sigma \cong 10^{-10}.$$

To get absolute figures we note that for good conductors the specific resistance $1/\sigma$ is of the order of 10^{-5} ohms. The resistance of the contact $r = \delta/\sigma_g$ reckoned per unit surface (in cm^2 of course) thus turns out to be of the order of $10^{-7} \cdot 10^5 = 10^{-2}$ ohms. For a twice larger gap with $\delta = 20A = 2 \times 10^{-7}$ cm we should get in the same way a resistance about 10^{10} times larger than the previous one, that is, about 10^8 ohms. Further increase of δ

[4] Cf. J. Frenkel, reference 1, p. 59.

[5] A. Sommerfeld, reference 3, formula (53b).

would mean practically complete disappearance of current; for with $\delta = 20A$ it would require an electric field of 10 million volts per cm across the gap, corresponding to a potential difference of 2 volts, to obtain a current of the order of 10^{-8} amp/cm^2.

It must be remarked that for such high values of the field intensity—or rather of the potential difference—the current would no longer be proportional to the latter, but would increase much faster, gradually assuming the character of Millikan's "field-currents." It can be easily shown that this character, corresponding to a practically unidirectional flow of electrons (from a to b, that is in the direction of the applied force only), would be acquired for potential differences of the same order of magnitude as that corresponding to the potential jump at the surface of the metal (U_a/e).

We must now come back and test the validity of our assumption that the mean value of the transmission coefficient $\overline{D(W_x)}$ can be identified with its value D_0 for the electrons with the velocity $v_x = v_0 (= v_{\max}$ for $T = 0)$.

To do this we must find out the maximum of the function

$$F(W) = \frac{D(W)}{e^{W/kT}/A + 1}$$

which enters the integral (14) defining the mean value of $D(W)/v$. Leaving aside the case of extremely high temperatures we can put $A = e^{-W_0/kT}$ where $W_0 = \frac{1}{2}mv_0^2$. Neglecting the variations of the denominator of (17) when compared with that of the numerator we can further put, according to (16a),[6]

$$D(W) \cong e^{-[(U-W_0)/kT_1]^{1/2}} \tag{18}$$

the "effective temperature" T_1, being defined by

$$1/kT_1 = 32\pi^2 m\delta^2/h^2. \tag{18a}$$

This gives

$$F(W) = \frac{e^{-[(U-W)/kT_1]^{1/2}}}{e^{(W-W_0)/kT} + 1}$$

or with

$$\frac{W_0}{kT} = \xi_0, \ \frac{W}{kT} = \xi, \ \frac{U}{kT} = \xi_1, \ \frac{U - W}{kT_1} = \gamma(\xi_1 - \xi), \ \gamma = \frac{T}{T_1}$$

$$F(W) = \frac{e^{-[\gamma(\xi_1 - \xi)]^{1/2}}}{e^{\xi - \xi_0} + 1} \cdot \tag{19}$$

The maximum of this function corresponds to the minimum of its reciprocal. Putting $\partial F^1/\partial\xi = 0$ we get

$$1 + e^{-(\xi - \xi_0)} = 2[(\xi_1 - \xi)/\gamma]^{1/2}. \tag{19a}$$

It can be easily shown[7] that this equation has either two solutions $\xi = \xi' < \xi_0$ and $\xi = \xi'' > \xi_0$, or none, depending upon the value of the parameter γ. The

[6] This implies, of course the limitation to the case $W < U$, see below.

[7] For instance, graphically, by tracing the exponential curve $Y = 1 + e^{\xi - \xi_0}$ and one branch of the parabola $Y = 2[(\xi_1 - \xi)/\gamma]^{1/2}$ (for $\xi < \xi_1$).

limiting value of this parameter is approximately equal to ξ_1. When $\gamma \gg \xi_1$, the equation (19a) has no solution, which means that the function (19) steadily increases, as ξ increases from 0 to ξ_1. In the opposite case $\gamma \ll \xi_1$, we get approximately

$$e^{\xi_0 - \xi'} \cong 2[(\xi_1 - \xi_0)/\gamma]^{1/2} \text{ and } 2[(\xi_1 - \xi'')/\gamma]^{1/2} \cong 1 \tag{20}$$

the first solution corresponding to a sharp maximum of $F(W)$ in the neighborhood of $W = W_0$, and the second to a faint minimum in the neighborhood of $W = W_1$. In order to see what case we have to deal with in practice, we must introduce numerical values.

If δ is measured in Angstroms, then it follows from (18a) $T_1 = 1.2 \times 10^4/\delta^2$ Thus the "effective temperature" is high for $\delta = 1A$ which means an actual contact between the two metals, of the order of magnitude of the room temperature $= 120°$K for $\delta = 10A$ and becomes very small as δ increases beyond this value.

Assuming U_a to be equivalent to 14 volts and W_0 to 10 volts, and taking $T = 300°$K, which corresponds with respect to the thermal energy kT to about 0.02 volts, we get $\xi_0 \cong 500$ and $\xi_1 \cong 700$.

We thus see, that in the above considered case of a gap $\delta = 10A$, the parameter γ is approximately equal to 2.5, that is, extremely small compared with its limiting value 700. For this case the first of the equations (20) gives approximately $\xi_0 - \xi' \cong 20$ and the second $\xi_1 - \xi'' \cong 2$. It can be easily verified that the maximum of $F(W)$ at $W = W' = kT\xi'$ is actually so sharp that $D(W)/v$ is practically equivalent to $D(W')/v'$ which is only very slightly different from the value $D(W_0)/v_0$ assumed above.

The condition $\gamma > \xi_1$ can be realized at $T = 300°$K for very broad gaps only with a width $\delta > 170A$. In this case the main part of the electric current —*for sufficiently small values of the potential differences* $\phi = V_1/e$, that is, for very small field intensities $E = \phi/\delta$—should be due to electrons having a kinetic energy larger than U. For these electrons the transmission coefficient $D(W)$ is of the order 1, whereas their number for usual temperatures is extremely small. The electric current between a and b would have in this case the character of a thermionic current (and not of a field current) whose strength can be calculated by using the general expression (15a) for the effective conductivity of the gap with[8]

$$D \cong e^{-(\xi_1 - \xi_0)} = e^{-(U - W_0)/kT}. \tag{21}$$

Thus in this case the electrical resistance of the gap $r = \delta/\sigma_g$ should be independent of its width δ and should vary with the temperature as $e^{-(u-W_0)/kT}$ It would have an appreciable magnitude only for very high temperatures lying in the same range as the temperatures for which thermionic currents are observed. For usual temperatures, gaps of such width could no longer be treated as contacts, whereas in the case of shorter gaps with $\delta \cong 10$A their resistance would be practically independent of the temperature and would vary exponentially with increase of δ.

[8] This being (approximately) the relative number of electrons with a kinetic energy larger than U.

The above results may have interesting applications to the question about the electrical resistance of granular structures, such as metallic powders and probably also extremely thin metallic films obtained by means of cathode sputtering. As well known, the latter possess an abnormally high specific resistance, which for films with a thickness of about 10^{-6} cm and lower, may be 20 times larger than that of the same metal in block, and further abnormally small temperature coefficient of resistance, which in fact can become negative (decrease of resistance with increase of temperature). It has been often assumed[9] that the high value of the specific resistance of very thin films is explained by the fact that their thickness may be smaller than the normal mean free path of the electrons (the latter being supposed to be scattered irregularly from both surfaces of the film). It would follow from this idea, that the "critical thickness" d for which the specific resistance should begin to increase, must be approximately equal to the mean free path l, and therefore must vary with the temperature in the same way as does the latter. According to the modern wave-mechanical theory of metallic conduction l varies inversely with T (or still faster in the region of very low temperatures), whereas as a matter of fact d remains practically independent of the temperature.

If on the other hand we adopt the equally often advocated granular theory of the constitution of thin films, and substitute for the usual conception of metallic contacts (which has been a serious obstacle for this theory) the conception developed in this paper, then the main properties of these films, distinguishing them from the metal in block, receive a satisfactory explanation. To say nothing of the abnormally high specific resistance, the smallness (or even the negative sign) of its temperature coefficient may be explained by the fact that the width of the gaps between adjacent grains is diminished, as a consequence of their thermal dilatation, with increase of temperature. Account should be taken of course of the thermal dilatation of the dielectric base upon which the film is deposited. But a comparison of the thermal dilation coefficients shows that they are as a rule larger in the case of the metals. This relation can be illustrated by the fact that the gaps between adjacent rails in a railway line decrease in the summer and increase in the winter time, and not vice versa. Denoting the length of a rail or grain with L and its effective dilation coefficient with α, we see that when the temperature is raised by ΔT the width of the gap is decreased by $\Delta\delta = -L\alpha\Delta T$. The relative decrease $\Delta\delta/\delta = -(L/\delta)\alpha\Delta T$ may be quite large even for a very small value of $\alpha(\leqq 10^{-5})$ if L is sufficiently large with respect to δ. And since the resistance of a gap varies exponentially with δ, this means a marked decrease of resistance, partially compensated by the normal increase of the resistance of the separate grains.

It is further well known that the resistance of thin films depends very largely upon the gas or gases present during their preparation. On our theory these gases must make thin monomolecular adsorbed layers on the surface of the separate grains of which the film is built up, thus changing

[9] An assumption that has been worked out mathematically by J. J. Thomson long ago.

the potential energy U which determines the transmission coefficient according to (18) and consequently the resistance of the gaps between the grains. It may be remarked that this change of resistance must be quite parallel to the change of the thermionic emission of the corresponding metal owing to the presence of the adsorbed layer. In fact the difference $U-W_0$ in equation (18) may be identified with the "work function" $U-W_0$ which determines the thermionic emission in Sommerfeld's theory. It follows then from (15a) that the logarithm of the specific conductivity of a gap (contact) must vary with U, as the square root of the logarithm of the thermionic emission of the same metal for the "effective" temperature.

3. Application to the Case of Two Different Conductors and to the Phenomenon of Rectification

Turning now to the consideration of a contact between two different conducting bodies (a, b) we shall first consider them as metals; the case of a non-metallic body (semi-conductor) may be obtained perhaps (see below) by taking the extreme form of the Fermi distribution law for a small concentration of free electrons which is nothing else but Maxwell's distribution law.

We have seen that when between two metals "in contact" that is at a small distance from each other, a potential difference V, determined by (1) and (6) is established, there will be no current flowing between them. Putting in (2) $A=e^{-W_a/kT}$ and $B=e^{-W_b/kT}$, we get, according to (6) $W-W_a=W-W_b+U_{ab}$ or according to (1).

$$V=(U_a-W_a)-(U_b-W_b). \tag{22}$$

This formula shows that the contact potential difference between two metals is equal to the difference of their respective work functions, as of course it should be.[10]

If a small additional potential difference V_1 is introduced, we must get a current determined by the "specific conductivity" (9). By the same argument as in the preceding paragraph we easily get

$$\sigma_g \cong \frac{1}{2}\,\frac{e^2\delta n_b D(W_g)}{mv_g}$$

where $W_g+U_{ab}=W_b$, W_b being (practically) the maximum kinetic energy of of the electrons in the body b at $T=0$. Since, according to the equilibrium condition $W_b-U_{ab}=W_a$, we get further, dividing σ_g by $\sigma_b=e^2l_bn_b/mv_b$ the specific conductivity of the metal b,

$$\frac{\sigma_g}{\sigma_b} \cong \frac{1}{2}\,\frac{\delta}{l_b}\,\frac{v_b}{v_a}D(W_a) \tag{23}$$

where $v_a=v_g$ is the largest velocity of the electrons in a at $T=0$. So far as v_a and v_b are of the same order of magnitude (which is the case for all metals)

[10] Cf. C. Eckart, Zeits. f. Physik **38** (1928) and J. Frenkel, Zeits. f. Physik (1928).

we do not have to discuss this expression in detail, since it is practically the same as in the case of two identical metals.

A few words should be added with regard to the character of the function $D(W_x)$.

For the case of the energy curve $MNQRST$ of Fig. 1 with an inclined top, this function cannot be evaluated in a simple way. We can, however, simplify the problem by replacing the inclined top line QR by a horizontal line $Q'R'$ passing through its center. We shall thus get a practically equivalent energy curve $MNQ'R'ST$ (Fig. 3), for which the function D is given

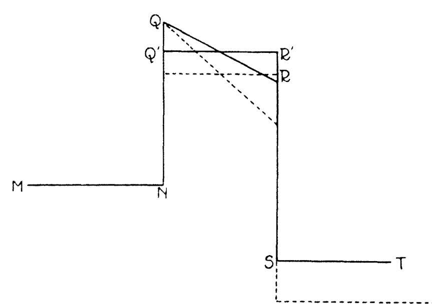

Fig. 3.

by the formula

$$D = \frac{4\alpha_a\alpha_b}{(\alpha_a + \alpha_b)^2 \operatorname{csh}^2 \beta\delta + (\beta - \alpha_a\alpha_b/\beta)^2 \sinh^2 \beta\delta}$$

where

$$\alpha_a^2 = \frac{8\pi^2 m}{h^2} W_x, \quad \alpha_b^2 = \frac{8\pi^2 m}{h^2}(W_x + U'), \quad \beta^2 = \frac{8\pi^2 m}{h^2}(U_a' - W_x),$$

with $U_a' = NQ' = U_a - \frac{1}{2}V$, $U_b' = SR' = U_b + \frac{1}{2}V$, $U' = U_{ab}$. For sufficiently large values of $\beta\delta$, with which we are here concerned, the above expression reduces approximately to

$$D = \exp\left[-\left\{(U_a - W_x - \tfrac{1}{2}V)/kT_1\right\}^{1/2}\right] \tag{24}$$

with the previous definition of T_1. In substituting this into (23) we must put $W_x = W_a$. We thus get the same expression for D as in the case of two identical metals, with the only difference that the "work function" $U_a - W_a$ is replaced by

$$U_a - W_a - \tfrac{1}{2}V = \tfrac{1}{2}[(U_a - W_a) + (U_b - W_b)] \tag{25}$$

that is by the arithmetic mean of the work functions of the two metals.

Now if V is increased by $V_1 \ll V$ the transmission coefficient is changed by the amount $(\partial D/\partial V)V_1$ where

$$\frac{\partial D}{\partial V} = \frac{D}{4[\frac{1}{2}kT_1(U_a - W_a + U_b - W_b)]}.$$

Introducing this expression and the expression (24) in (11) and putting for the sake of brevity

$$[2kT_1(U_a - W_a + U_b - W_b)]^{1/2} = \theta \tag{26}$$

we have

$$\alpha_2 = \frac{2\pi}{m^2}\frac{1}{2\theta}\int_0^\infty dWD(W)[f_b(W + U) + \theta f_b'(W + U)]$$

or so long as θ is small compared with $W+U$

$$\alpha_2 = \frac{\pi}{m^2\theta}\int_0^\infty dWD(W)f_b(W + U + \theta). \tag{27}$$

Comparing this (8) we see that

$$\alpha_2 = \alpha_1'/2\theta \tag{28}$$

where α_1' is the value taken by the coefficient α_1, if U is increased by θ. It may be remarked that α_1, is connected with the "specific conductivity" of the gap σ_g by the relation

$$\sigma_g = e^2\delta\alpha_1.$$

Now α_1', is but slightly different from α_1, since θ is assumed to be small, so that we can finally put $\alpha_2 \cong \alpha_1/2\theta$. This relation could be of course obtained directly by neglecting the first term in the integral (11) which does not take account of the change of the transmission coefficient D caused by the introduction of the additional potential difference V_1. It can be easily shown that the ratio of this term to the second term, which just characterized this change, is approximately equal to

$$\frac{\theta}{U_a - W_a + U_b - W_b} = \left(\frac{2kT_1}{U_a - W_a + U_b - W_b}\right)^{1/2}$$

that is, remains very small for values of T_1 corresponding to gaps of the width $\delta = 10A$ or even less than that.

The "characteristic curve" of our contact, considered as a rectifier, is thus the parabola

$$I = \alpha_1 V_1(1 + V_1/2\theta)$$

This equation holds, of course, for sufficiently small values of V_1 only. It follows from it that the rectifying action of the contact becomes prominent for values of $|V_1|$ which are of the same order of magnitude as θ. Using the previous value of $T_1 = 120°$ which corresponds to 0.01 volts and assuming for the mean work function of the two metals $\frac{1}{2}(U_a - W_a + U_b - W_b)$ a value corresponding to 4 volts, we get for θ about 0.4 volts.

This is a rather large figure, which explains the fact that metallic contacts cannot be used as detectors for radio-oscillations of small amplitude.

It seems possible, however, on our scheme to explain the detector action of contacts between metals and some semi-conducting "crystals" used for this purpose in simple radio receivers, by assuming that in these bodies we have to deal with a distribution of electrons different from that of Fermi.

The simplest assumption would be to replace the Fermi distribution by the usual Maxwellian one, corresponding to a relatively small number of free electrons per unit volume. This can be considered as a particular case of the preceding theory, since the Maxwell distribution is the limiting case

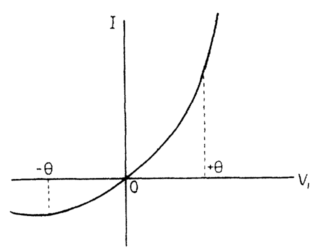

Fig. 4.

of the Fermi distribution which may be specified by putting the maximum kinetic energy of the electrons at $T=0$ equal to zero. It is clear that in this case the semi-conductor will play the role of the body a and the metal that of the body b, so that we shall have $W_a=0$, $V=U_a-(U_b-W_b)$, etc.[11]

It seems at first sight that this will leave our formulas for α_1, and α_2 unaltered since they depend upon the distribution function of the body b only and not on that of the body a. That this is not so, is clear, however, from equation (23), which with $W_a=0$ and $v_a=0$ would give $\sigma_g=\infty$. This shows that some of the approximations used in the evaluation of α_1, and α_2 no longer hold in the limiting case we are now considering.

We shall not try to adjust our calculations to this case, for the implied picture of a semi-conductor as of a box enclosing a rarefied electron gas seems hardly adequate enough to deserve a quantitative treatment. It is, however, directly apparent without any calculations whatsoever, that the rectifying effect of a contact between two bodies, so far as it depends upon their dissimilarity with respect to the concentration and the velocities of the electrons, must increase for a given absolute value of V_1 as this dissimilarity becomes more pronounced.

A satisfactory extension of the above theory to the case of contacts between a metal and a semi-conductor, or between two semi-conductors will be possible only after an at least crude electron theory of such semi-conduc-

[11] We have designated with a that body for which the potential energy (inside) is higher than for the other when equilibrium is reached, that is, when the straight lines representing the kinetic energies W_a and W_0 lie on the same level (that is, coalesce with each other).

tors shall have been developed. It will be further necessary to take account of the fact that in actual contacts the distance δ between contiguous surfaces does not remain constant, but varies in a more or less periodic manner, within a certain range, and that the adjacent (curved) surfaces of the two bodies need not be equipotential surfaces, as is the case if they are far apart.