Owen W. Richardson

Thermionic phenomena and the laws which govern them

Nobel Lecture, December 12, 1929

In its broadest aspect this subject may be summarized as the branch of Physics which deals with the effect of heat on the interaction between electricity and matter. It is not altogether new. Nearly 200 years ago it was known that air in the neighbourhood of hot bodies conducted electricity. In 1873 Guthrie showed that a red-hot iron ball in air could retain a negative but not a positive charge. In a series of researches extending from 1882 to 1889, Elster and Geitel examined the charge collected on an insulated plate placed near various hot wires in diverse gases at different pressures. The observed effects were very specific and varied, but there emerged a general tendency for the plate to acquire a positive charge at low temperatures and high pressures, and a negative charge at high temperatures and low pressures. The matter became really interesting in 1899 when J. J. Thomson showed that the discharge from an incandescent carbon filament in a vacuum tube was carried by negative electrons. In 1900 McClelland showed that the currents from a negatively charged platinum wire were influenced very little, if at all, by changes in the nature and pressure of the surrounding gas, if the pressure were fairly low. These facts seemed to me to be highly significant, and I resolved to investigate the phenomenon thoroughly.

The view of these effects generally held at that time by people who had thought about them was that the electric discharges were carried by ions and electrons which were generated by the interaction of the neighbouring gas molecules with the hot body. It was left an open question as to whether this action was merely thermal, a matter of kinetic energy, or was chemical, or involved the intervention of radiation. The effects observed in the best vacua were attributed to the residual gas which could not be got rid of. This was, of course, easily possible. I felt, however, that it was very likely that interacting gases had little to do with the main phenomenon, but that the negatively charged electrons and, possibly, the positively charged ions too were coming from the heated solid. This would be reasonable from the point of view of the theories of metallic conduction which had been put forward

between 1888 and 1900 by Thomson, Riecke, and Drude. I decided that the best way to make progress was to get rid of the complications due to the presence of gases and to find out what, if anything, happened when gas effects were excluded.

This was not so easy at the beginning of this century as it would be at the present time. Largely owing to the technical importance of the phenomena under consideration the art of evacuating gases has advanced enormously since then. In those days the gas had all to be got away by hand pumps. As the heating of the tube walls and other parts of the apparatus by the hot wire generates gas from them which continues almost indefinitely this is a most tedious operation. I have often heated a wire in a tube for weeks in succession in order to make sure that the currents observed were stable and not coming from residual gas. There was no ductile tungsten; the most refractory material readily available in a reasonably pure form was platinum. In 1901 I was able to show that each unit area of a platinum surface emitted a limited number of electrons. This number increased very rapidly with the temperature, so that the maximum current i at any absolute temperature T was governed by the law

$$i = AT^{\frac{1}{2}}e^{-w/kT} \tag{1}$$

In this equation k is Boltzmann's constant, and A and w are specific constants of the material. This equation was completely accounted for by the simple hypothesis that the freely moving electrons in the interior of the hot conductor escaped when they reached the surface provided that the part of their energy which depended on the component of velocity normal to the surface was greater than the work function w. In 1903 I showed that the same conclusions could be drawn for sodium and more qualitatively for carbon. Further, that the differences of the work functions of different substances should be equal to their contact potential differences, and the experimental values for platinum and sodium verified this. The results also verified the conclusion that the work functions for different elements should be of the same order of magnitude as ½ (e^2/d), where e is the electronic charge and d the radius of the atom, and also that it should vary roughly as the inverse cube root of the atomic volume. In the same year Wehnelt found that similar phenomena were exhibited by a large number of metallic oxides. The alkaline earths in particular had an exceptionally low work function and were in consequence very efficient emitters of electrons.

It is necessary to say a word or two in parenthesis about the positive ioniza-tion which is frequently observed. This is due to an emission of positive ions which arises in various ways. When any ordinary sample of a solid is first heated, it gives rise to a copious emission of positive ions which decays (and sometimes recovers) with time in a manner which resembles superficially that of radioactive substances. This effect is due to impurities. After this has been got rid of, there may be another more stable emission characteristic of the substance itself. There is a third type which is a direct result of interaction between the heated solid and the surrounding gas. I devoted a good deal of time between 1904 and 1912 to the investigation of these effects. The results were interesting, but there is not time to consider them in any detail. I will only mention that all three types of positive emission, when stable, were found to obey the same temperature law $AT^{1/2e-b/T}$ as the electronic emission but, of course, with different constants A, b; that the carriers of the char-acteristic emissions were charged atoms of the metallic constituent; and that the carriers of the temporary effect were singly charged atoms of sodium or potassium, the latter usually predominating, which are present as contam-inants.

The central idea which lies behind the theory summarized in Eq. (1) is that of an electron gas evaporating from the hot source. If this idea is correct, the thermionic currents should be able to flow against a small *opposing* elec-tromotive force because the kinetic energy of the heat motion of the electron gas molecules, in other words the electrons, will carry some of them through it. Furthermore, we could at that time find out a great deal more about what the electrons in an electron gas were doing than we could about the mol-ecules of an ordinary gas. Owing to the fact that they are electrically charged, their motion can be controlled by an external electric field. By measuring the electronic current which flows against various directly opposing fields it is possible to ascertain the proportion of the emitted electrons which have a value of the component of their velocity perpendicular to the emitting sur-face between any assigned limits. By making observations of the spreading of the electrons sideways under different small accelerating fields it is possible to deduce similar information about the components of velocity parallel to the surface. By experiments of this kind made in 1908-1909, partly with the help of F. C. Brown, I was able to show that the distribution of velocity among the emitted electrons was identical with the Maxwell distribution for a gas, of equal molecular weight to that of the electron, at the temperature of the metal. The identity was shown to hold for each velocity component.

Apart from its interest in connection with electrons, this was the first experimental demonstration of Maxwell's law for any gas, although the law was enunciated by Maxwell in 1859.

There were two other matters which required urgent investigation before the theory of electron emission could be regarded as securely founded. The first was this. If the electrons are really coming out of the hot body by virtue of their heat energy being able to overcome the work function w, the hot body should be cooled by this process. It is like the cooling of water by evaporation. I published a calculation of the magnitude of this effect in 1903, but the first experimental investigation was made by Wehnelt and Jentzsch in 1909. They observed a cooling effect, but the magnitude did not agree with the theory. In 1913 H. L. Cooke and I devised an improved experimental method of attacking this question, redetermined this cooling effect, and showed that it agreed with the value of the work function deduced from the variation of the thermionic currents with the temperature. Our conclusions have since been confirmed by the very accurate experiments of Davisson and Germer made in 1922.

The other matter to which I referred is the converse of this. If a stream of electrons flows into a conductor from outside, there should be a development of heat which does not depend either on the temperature of these electrons or on the magnitude of the small potential differences used to drive them. H. L. Cooke and I devised and put into operation an apparatus for detecting and measuring this effect in 1910-1911. The results showed a satisfactory agreement with the value of the work function obtained by the other two methods.

Despite the steadily accumulating mass of evidence to the contrary, some of which I have briefly outlined, the view had been fairly commonly held up to about 1913 that thermionic emission was not a physical phenomenon but a secondary effect of some chemical reaction between the hot body and the surrounding gas. The advent of ductile tungsten enabled me, in 1913, to get very big currents under better vacuum conditions than had hitherto been possible and to show that the mass of the electrons emitted exceeded the mass of the chemicals which could possibly be consumed. This experiment, I think, ended that controversy so far as it could be regarded seriously.

There is a very close relationship between thermionic and photoelectric phenomena. The photoelectric threshold frequency, the least frequency v_0 which will eject an electron from a given substance, is connected with the thermionic work function w_0 by the simple relation

$$w^{0} = h\, v^{0}$$

where h is Planck's constant. This was established by experiments made by
K. T. Compton and myself in 1912. We know that any body in thermal
equilibrium at any temperature T is surrounded by a bath of radiation in
which the frequency distribution is given by Planck's formula. This formula
puts no finite limit on the magnitude of the frequencies occurring; so that
there will always be some frequencies present for which v_o is greater than
w_0/h. Such frequencies will eject electrons by photoelectric action; so that
the temperature radiation alone will, by a kind of photoelectric effect in-
tegrated over the whole spectrum, give rise to an electronic emission which
should increase with the temperature. In 1912 I showed that it followed from
the principles of thermodynamics that this integrated photoelectric emission
would follow Eq. (I) exactly with, possibly, a different value for the con-
stant A. This conclusion was established by direct experiment later by W.
Wilson, in 1917. Thermionic emission might thus well be an integrated pho-
toelectric emission; only the absolute magnitude could decide. In 1912 there
were no known data which would enable the magnitude of this integrated
photoelectric effect to be ascertained, so, with the collaboration first of K.
T. Compton and later of F. J. Rogers, I set about to determine the absolute
values of the photoelectric yields of various substances as a function of fre-
quency. With the help of these absolute values I was able in 1916 to calculate
the electron emission from platinum at 2,000° K due to its complete black-
body spectrum. The result showed that thermionic emission is at least 5,000
times, and almost certainly 100 million times, as large; so that thermionic
emission cannot be merely an integrated photoelectric effect, although it has
the same thermodynamic properties. A photochemical theory of chemical
reactions based on considerations analogous to these has been put forward
independently by Penrrin and seems to have met with very similar difficulties.
If we have to make a decision now, the verdict must be, on the facts at pres-
ent revealed, that the part of these effects which is of radiational origin is
comparatively unimportant. I am not sure, however, that the end has been
heard of this matter. I have a feeling that there is something coordinating
these radiational and mechanical or chemical effects which at present is con-
cealed from us.

I will now say a few words about the relation between thermionic phe-
nomena and theories of metallic conduction. In so far as it can be regarded as

a serious contribution to scientific knowledge, thermionics was born at the same time as the theories of metallic conduction associated with J. J. Thomson, Riecke, Drude, and Lorentz, and it grew up with them. The dominant feature of these theories is the assumption that the currents in metals are carried by electrons which are moving freely and which possess the same average amount of kinetic energy as that of the molecule of a monatomic gas at the same temperature. Since all the thermionic facts which I have outlined received a ready explanation on these theories, there came to be a presumption that they favoured them rather than others, such as that put forward by Lindemann, which supposed the electrons in metals to be normally at rest. The fact that the experiments confirmed the requirement of the former theories, that the emitted electrons should have a Maxwell distribution of kinetic energy, especially seems to have led to the spreading of this opinion. It is a requirement of classical dynamics that this distribution should hold for electrons in any part of a system in thermal equilibrium, and as it is found to be true for the external electrons, the only part of the system accessible to experimental investigation, there is a presumption that it will also be true of the internal electrons. But this presumption has no validity apart from classical dynamics. Except for the considerations dealt with in the next paragraph which were perhaps still somewhat uncertain, the ascertained facts of thermionic emission did not favour one type of theory of metallic conduction rather than another until 1922, when Davisson and Germer made a very accurate comparison of the experimental value of the work function deduced from the cooling effect with that deduced from the temperature emission formula, at different temperatures, using the same tungsten filament. An analysis of their results showed that the experimental evidence was definitely against the classical theory of metallic conductors and in favour of a type of theory which makes the kinetic energy of the internal electrons practically independent of the temperature.

In 1911 as a result of pursuing some difficulties in connection with the thermodynamic theory of electron emission I came to the conclusion that

$$i = AT^2 e^{-w/kT} \tag{2}$$

was a theoretically preferable form of the temperature emission equation to Eq. (1) with, of course, different values of the constants A and w from those used with (1). It is impossible to distinguish between these two equations by experimenting. The effect of the T^2 or $T^{1/2}$ term is so small compared with

the exponential factor that a small change in A and w will entirely conceal it. In fact, at my instigation K. K. Smith in 1915 measured the emission from tungsten over such a wide range of temperature that the current changed by a factor of nearly 10^{12}, yet the results seemed to be equally well covered by either (1) or (2). It is, of course, very satisfactory to know that either formula will do this. There are not many physical laws which have been tested over so wide a range. The great advantage of Eq. (2) is that it makes A a universal constant; so that there is only one specific constant for each substance, namely w. The first time I mentioned explicitly that A was a universal constant was in 1915. Here I came to it as a result of a thermodynamic argument about electron emission. In 1914 I had already come to it by a different route. I had come to the conclusion that the classical statistics were not applicable to the electrons inside conductors. There was no means of ascertaining what the correct statistics were, so I endeavoured to avoid this difficulty by adopting some quantum ideas previously used by Keesom to calculate the specific heat of helium at low temperatures. In this way I determined the constant A as $0.547 \ mk^2e/h^3$ (m and e being the mass and charge of the electron, k and h Boltzmann and Planck's constants). These calculations have since been improved upon by others, but there still seems to be some doubt about the pure number factor which I made out to be 0.547. The most probable value of it seems to be 4π. Amongst those whose writings have made important contributions to this question since 1915 are von Laue (1918), Tolman (1921), Dushman (1923), Roy (1926), Sommerfeld (1927), and R. H. Fowler (1928).

By 1924 it was easy to prove that all the existing theories of metallic conduction were wrong, but just where they went wrong it was impossible to say. None of them were able to unite in a straightforward and satisfactory way such diverse facts as the law of Wiedemann and Franz, the large number of free electrons and the mean free paths required by the optical properties of metals, their known crystal structures, their small specific heats, the variation of conductivity with temperature and the existence of supraconductivity, and the relation between the thermionic cooling effect and the temperature.

This great problem was solved by Sommerfeld in 1927. Following up the work of Pauli on the paramagnetism of the alkali metals, which had just appeared, he. showed that the electron gas in metals should not obey the classical statistics as in the older theories, such as that of Lorentz for example, but should obey the new statistics of Fermi and Dirac. This makes a pro-

found change in the distribution of velocities among the electrons when their concentration is very great, as in the interior of a metal, but it makes little or no difference when the concentration is small, as in the external electron atmospheres. It thus allows us to retain the experimentally established Maxwell distribution for the external electrons. On this theory the energy of the internal electrons is the energy of their Schrödinger proper values. If the concentration of the electrons is large, as in a metal, it is fixed almost entirely by the density of the electrons and has little to do with their temperature. It is, in fact, a kind of zero-point energy. This feature immediately accounts for the very small contribution of the electrons to the specific heats of metals, which was so great a difficulty for the older theories. It appears also to be capable of accounting for the other serious difficulties.

There is one other feature of Sommerfeld's theory which I must mention as it affects the interpretation to be put on the thermionic work function w. Before the advent of this theory w was interpreted as the work required to remove a free electron at rest inside the metal to a point outside. According to Sommerfeld's theory w is equal to the difference between this work and the maximum energy of the internal electrons. The value of this is

$$\frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{\frac{2}{3}}$$

if n is the number in unit volume. For most metallic conductors this quantity is equivalent to about 10 volts. As w is generally somewhere about 4 volts, this means that the difference between the electrostatic potential energy of a free electron inside and outside a metal is some 3 or 4 times as large as was formerly supposed. Direct experimental evidence that this is correct is furnished by the recent experiments of Davisson and Germer on the diffraction of electrons by nickel crystals.

We have seen that the classical theories of metallic conduction gave a pretty good account of those thermionic phenomena which I have so far referred to. The only clear exceptions which emerged were the magnitude of the work function in relation to temperature as deduced from the cooling effect and the calculation of the actual magnitude of the absolute constant A which enters into the $AT^{2}e^{-w/kT}$ formula. As this contains Planck's constant h its elucidation necessarily involved some form of quantum theory. As a historical fact, however, it was chiefly on other difficulties with the properties of metals that the older theories wrecked themselves. I come now to some

thermionic phenomena with which the older theories were not so successful.

It became apparent at a very early stage that the emission of electrons from conductors at a given temperature was very susceptible to the influence of foreign substances and particularly to gaseous contaminants. This was not surprising in itself, as the phenomenon is essentially a surface phenomenon but some of the observed effects were unexpected. In 1903 H. A. Wilson made the important observation that the emission from platinum could be enormously increased by the presence of small quantities of hydrogen. In 1908 he showed that in certain circumstances this emission was a function of the pressure of the hydrogen at a fixed temperature. If the pressure was kept fixed, the currents still followed the $AT^{2}e^{-b/T}$ formula, but with changed values of A and b. These parameters were now functions of the pressure of such a kind that they obeyed an equation

$$b = c \, l \, o \, g \, A + d \qquad\qquad (3)$$

where c and d are new constants independent of both pressure and temperature. In 1913 Langmuir observed that the emission from tungsten was affected by various gases, hydrogen having a particularly depressing effect on this substance, although the effect may in reality be caused by water vapour. In 1915 I pointed out that in all these cases of contamination the currents still obeyed the $AT^{2}e^{-b/T}$ formula, with changed values of the parameters, but that all the values, including those for the pure metals, satisfied Eq. (3) with the same constants c and d. In 1925 A. F. A. Young and I extended the list to include potassium contaminated in a large number of ways. The number of substances which subscribe to Eq. (3) has recently been added to very considerably by several American investigators. This effect is not small, it is large. The parameter A can change by a factor of 10^{12} as a result of contamination.

Since 1915 I have felt that this result must be important both on account of its generality and of its magnitude, but I have never been able to arrive at any satisfactory reason for it. It seems now that this is one of those phenomena which are only to be accounted for with the help of the new waves of L. de Broglie. The solution of the problem we owe to R. H. Fowler and Nordheim (1928). Their explanation is similar in principle to that by which Gamow and Gurney and Condon explain the disintegration of the radioactive nucleus. They take the conventional simplified picture of a metal as a sharply bounded region of low potential energy densely packed with elec-

trons. But as de Broglie has shown us, an electron can be regarded as a train of waves, or a wave packet, having a wavelength equal to h divided by the momentum of the electron. If such a wave is incident on the surface of the metal, it may be either reflected or transmitted. Thus the problem of the emission of electrons by a conductor may be looked upon as the problem of the reflection of the corresponding de Broglie waves at the hill of potential gradient which exists at the boundary. If the height of this hill is H, then on my old theory none of the electrons reaching the boundary would escape if their normal component N of kinetic energy were less than H, whereas all would escape for which N exceeded H. In the wave reflection problem it is still true that there is total reflection for N less than H, but the sharp discontinuity at $N = H$ has disappeared. It is found that the proportion transmitted is a continuous function of N and H, whose value tends to unity as the difference between N and H increases. This, however, makes very little difference; and when the calculations are completely carried out, it is found that Eq. (2) is still valid with the magnitude of the universal constant A unaltered in any essential way.

This result, however, depends essentially on the assumption that the potential energy increases to a permanent maximum as the electron crosses the surface. No doubt this is the correct picture for a pure metal, but for a contaminated one we may expect something different. If the contaminant is a thin layer, it may be only a few molecules thick, of a more electropositive substance, we should expect the hill to rise to a maximum height, let us say H_1, and then fall to a permanently lower level at a height H_2. On the old ideas the condition for escape would be that N should exceed the maximum height H, but in the wave problem it is possible for some of the waves to penetrate the hump H_1-H_2 provided its thickness is not large compared with the wavelength. There is a well-established optical analogue of this in the failure of total reflection when the thickness of the reflecting medium becomes comparable with the wavelength of the light. When the transmission of the de Broglie waves is calculated for this more complicated potential distribution, it is found that the emission formula (2) still holds good, but with new constants A and b, which are connected together by a relation which is equivalent to (3). I am not claiming that all the facts in this department of thermionics have been' completely coordinated by these theories of Fowler and Nordheim, but it is satisfactory that we have begun to understand something about this intractable subject.

I come now to a phenomenon which is not exactly thermionic, as it is

independent of temperature, but in some ways it is intimately related to thermionic effects. It has been suspected for a long time that electrons-could be pulled out of metals without the co-operation of gases by sufficiently strong electric fields. The effects seemed very erratic and difficult to investigate. The reality of the phenomenon has, however, been firmly established by the work of Gossling, of Millikan and Eyring, and of Rother, during or a little prior to 1926, and by that of various experimenters since then. These currents are carried by electrons and they may be quite large. The magnitude is independent of the temperature of the emitting substance, but at the same time is a continuous function of the applied electric field. The theory of this effect was discussed at length by Schottky in 1923 and more briefly, but in relation to the new experimental data, by Millikan and Eyring in 1926. It does not seem to have been realized, however, that no rational treatment of the old particle theories would get the electrons out in a way which made the emission a continuous function of the field without at the same time being a function which was sensitive to the temperature. I noticed this important point in 1927, and accordingly I attacked the problem from a new point of view by regarding it as a Schrödinger wave problem of an electron in the field of force at the conducting surface. Perhaps the last word has not been said on this matter, but it now looks as though in this attempt I attributed too much importance to the mirror-image attraction of the electron in the surface. Whatever its ultimate importance may be, this paper first drew attention to essential physical aspects of this phenomenon and indicated in a general way the nature of its connection with thermionic effects. In 1928 the problem was attacked by Oppenheimer and, more completely, by Fowler and Nordheim, who succeeded in putting it into an exceedingly simple form. They treat it in the same way as they treated the problem of thermionic emission, namely as a problem in the reflection of de Broglie waves at a potential barrier. The only essential difference between the two problems is that the potential, instead of being constant outside the metal, now falls off as a linear function of the distance from the surface. Their solution of this problem leads to a formula which so far as I am able to judge, is in excellent agreement with the ascertained facts in this domain.

The existence of this field extraction phenomenon has a number of interesting consequences, one of which I will now mention. If we consider an evacuated enclosure containing a number of bodies having different thermionic work functions w_1, w_2, etc., they will not be in electrical equilibrium unless their surfaces are charged. The reason for this is that those with lower

work functions would emit electrons at a more rapid rate than those with higher work functions. The condition for equilibrium to a first approximation, and one which covers the essential features of the phenomenon, is that there should be a certain field of electric force between the different bodies. This is such that, if the potential difference between any point just outside the body with suffix I and any point just outside the body with suffix 2 is V_{12}, then $eV_{12} = w_1 - w_2$. V_{12} is the contact potential difference between the bodies I and 2. There is nothing essential to the thermionic argument which depends on the shape, size or relative position of the bodies, and the result should be the same whether they are interconnected by other conductors or insulated from each other. The quantities such as V_{12} are thus intrinsic potential differences which are characteristic properties of the materials of which the conductors are made.

The field extraction phenomenon requires a modification of this conclusion. To simplify the argument I consider only two bodies, those with suffixes 1 and 2. Some portion of each of them is bounded by a plane surface, and the bodies are arranged so that these plane surfaces are parallel to one another and a distance x apart. The more distant parts of the bodies may be united by an electric circuit which includes a galvanometer. When x is considerable, there is equilibrium and no current passes through the galvanometer, because the excess electrons emitted by the more electropositive body are kept back by the potential difference V_{12} and this equilibrium is practically unaffected by the small force eV_{12}/x. But now suppose x to become very small, let us say comparable with atomic dimensions. The force eV_{12}/x now becomes large and will begin to extract electrons from the more electronegative body. This upsets the equilibrium, which is restored by a current passing through the galvanometer. But this is a perpetuum mobile: the current can be made to do useful work. It consumes nothing and the apparatus has no moving parts. If it is argued that it may be tapping some source of heat, at least it must be a perpetuum mobile of the second kind, since it works at a constant temperature. What is the answer to this riddle? I say it is this: the contact potential difference V_{12} is not completely independent of the distance between the two bodies. When this distance becomes small, V_{12} diminishes, and this diminution takes place in such a way that the additional electron current from the more electropositive body which reaches the more electronegative body owing to the reduced value of V_{12} is just equal to the electron current which is extracted from the more electronegative body by the field. In particular when the bodies are in contact, V_{12} falls to zero or at

any rate to a quantity of the order of the thermoelectric magnitudes. Well, this seems to correspond to the actual properties of the contact difference of potential, and I think it clears up an old difficulty in connection with it.