The measurement of Planck's constant using the visible photoelectric effect

1981 Eur. J. Phys. 2 139

(http://iopscience.iop.org/0143-0807/2/3/003)

Eur. J. Phys. 2 (1981) 139–149. Printed in Northern Ireland

The measurement of Planck's constant using the visible photoelectric effect

R G Keesing

Department of Physics, University of York, Heslington, York Y01 5DD, England

Received 3 December 1980, in final form 30 July 1981

Abstract This paper presents an experimental and theoretical investigation of the sources of error in the measurement of Planck's constant using the photoelectric effect. It points out the approximate nature of the Einstein photoelectric equation and gives details of a method of determining h which avoids the problems of measuring 'cut-off' potentials. The design of a simple photodiode is also given which does not suffer from having a detectable reverse photocurrent or leakage current. Finally there is a brief discussion of the early measurements of h which attempts to show how some important misconceptions have arisen.

Zusammenfassung In dieser Arbeit untersuchen wir die Fehlerquellen bei der Messung der Planck'schen Konstanten mit Hilfe des photoelecktrischen Effekts experimentell und theoretisch. Es zeigt sich, daß die Einsteinsche Gleichung für den photoelektrischen Effekt nur näherungsweise gültig ist. Eine Methode zur Bestimmung von h wird vorgeschlagen, die die Probleme der Messung der Gegenpotentiale vermeidet. Eine einfache Photodiode ohne meßbaren inversen Photostrom oder Leckstrom wird entworfen. Schließlich diskutieren wir kurz einige der frühen Messungen von h um zu zeigen, wie manche wichtige begriffliche Fehler entstanden sind.

1. Introduction

The measurement of Planck's constant using the photoelectric effect has a long history stretching back to the beginning of this century. The work of Millikan (1916a) during the years 1905–1916 appeared to establish Einstein's interpretation of the photoelectric effect in metals upon a sound experimental base. Millikan's experimental apparatus was very sophisticated, this sophistication being apparently justified by the accuracy of his final result which was

$$6.57 \pm 0.03 \times 10^{-34}\ \text{Js}.$$

Simple experiments to measure Planck's constant using the visible photoelectric effect are to be found in many undergraduate teaching laboratories. It is my experience that the results are almost always low and on occasion are negative. It was to understand these effects that the following experimental and theoretical investigation into the design of the apparatus and the interpretation of the photoelectric effect was carried out.

My original understanding of Einstein's theory of the photoelectric effect is contained in his famous photoelectric equation, that is, the maximum energy of a photoelectron in vacuum, E_{max} is related to the energy of the photon $h\nu$ and the work function of the emitter $e\phi$ by:

$$E_{max} = h\nu - e\phi.$$

Einstein argued that there would be electrons ranging in energy from zero to E_{max} from the assumption that electrons could lose energy during their escape from the metal. Further it was expected that E_{max} would be observable and independent of light intensity.

There are two distinct experiments that can be performed to investigate the above relationship, both of which use monochromatic light sources focused upon a cathode in a vacuum photodiode. The first method employs an open circuit diode, and allows the anode to charge negatively in response to the photocurrent from the cathode until no more electrons can reach it. The reverse, self-induced, bias is then measured as a function of frequency. The second method, used by Millikan amongst others, observes the curve of the photocurrent plotted against the retarding voltage and attempts to measure the voltage at which the cur-

0143-0807/81/030139+11$01.50

rent becomes zero, as a function of the frequency of the light.

There are several complicating effects in a practical photodiode which may cause the cut-off energy to be difficult to observe. These include:
(*a*) point to point variation of the work function of the cathode,
(*b*) energy dependent electron reflection at the anode,
(*c*) geometrical effects which cause an electron to have sufficient energy but insufficient momentum to reach the anode,
(*d*) reverse photoelectric and leakage currents and
(*e*) thermal effects.
These effects may cause different problems for the two experiments and thus it is useful to discuss them separately.

2. The two experiments

2.1. The reverse, self-induced bias, method

It was originally thought that this method would lead to better estimates of E_{max} than the Millikan method because it overcame the two problems of variation of work function and imperfect geometry. In the cut-off region, where the reverse voltage has reduced the photocurrent to very small values, only electrons which originate from regions of lowest work function, with trajectories almost parallel to the field lines, contribute to the cut-off voltage. Thus a polycrystalline cathode appears in this region to have a single-valued work function and ideal geometry, in the sense of condition (*c*). It was clear that if the point at which the photocurrent goes to zero is to be observed the reverse current would have to be negligibly small, for if it were not, one would simply detect the point at which the forward and reverse currents balance.

In order to investigate the reverse bias experiment a very low reverse current planar vacuum photodiode was designed. It has a cathode work function of approximately 2.5 eV and anode work function of greater than 4.0 eV. This is achieved by having a nickel cathode and oxodised copper anode. The cathode is caesiated to reduce its work function to about 2.0 eV. The effect of monolayer quantities of caesium on the anode is to leave it with a work function in excess of 4.0 eV. Figure 1 shows the basic layout of the tube, it has a pyrex envelope, the caesium generator is an SAES getter strip, which lies in the plane of the anode and points towards the centre of the cathode, and surrounding the anode lead is a glass sleeve to prevent leakage currents. (See the appendix for details of processing and performance.)

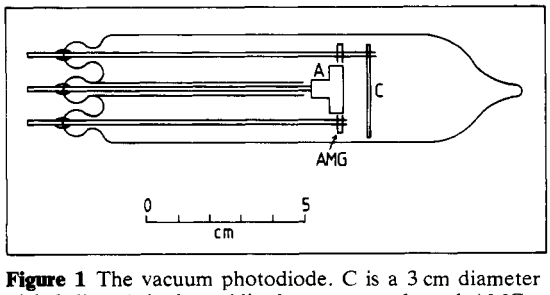

Figure 1 The vacuum photodiode. C is a 3 cm diameter nickel disc; A is the oxidised copper anode and AMG is the caesium generator which lies in the plane of A and behind it.

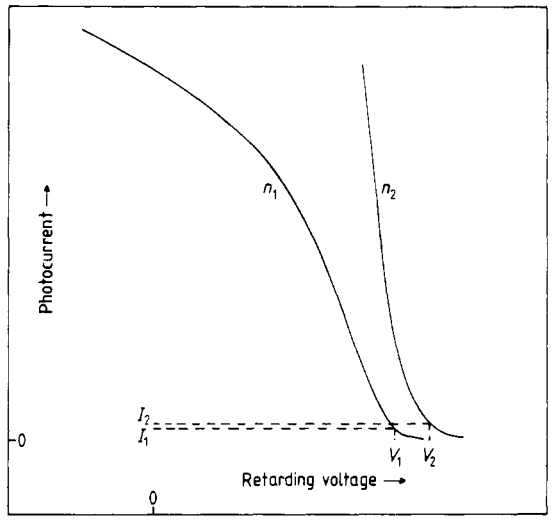

Figure 2 Schematic diagram of two photocurrent retarding curves for different fluxes (n_1 and n_2) of the same energy photons showing the shift in measured cut-off voltage with light flux.

The tube was in the process of manufacture when it was realised that the reverse bias experiment was basically unsound. The problem is this: if there is absolutely no reverse current from whatever source, the reverse standing voltage, due to the photocurrent, increases without limit. This arises from the thermal distribution of the free electron energies in the cathode, for it will be recalled that for all finite temperatures T and energies E the free electron energy distribution function, which is proportional to $E^{1/2}/(1+\exp(E-E_f)/kT)$, is finite. As long as one waits a sufficient time there will always be electrons which can surmount any finite reverse bias.

Consider the effect of measuring the self-induced reverse bias. Placing a voltmeter across the diode will cause the unknown reverse bias to settle down to some measured value V; if the voltmeter has an internal resistance R, a photocurrent V/R must flow. The measured value V will depend upon light intensity in a non-linear way. Consider figure 2 where two schematic current–voltage curves are

shown for two different intensities of the same wavelength light: V_1 and V_2 are the two measured self-induced reverse biases and I_1 and I_2 are the two measuring photocurrents, $V_1/I_1 = V_2/I_2 = R$, the instrumental resistance. The actual shapes of these functions will be discussed in §3, however, the point to note is that increasing the light intensity at any wavelength causes the cut-off voltage to increase.

As the measured cut-off voltage increases with light intensity it is possible for the cut-off voltage for an intense red line to exceed that for a weak blue line and thus indicate a negative value of Planck's constant. Figure 3 shows the reverse self-induced bias (cut-off voltage) as a function of light intensity measured with one of the special diodes, an argon ion laser, tuned to 514.5 nm and a Keithley 616 dmm with input resistance set at $10^{11}\ \Omega$. as the laser power was increased from 0.005 to 2.0 W the observed cut-off voltage increased from 0.22 to 1.17 volts.

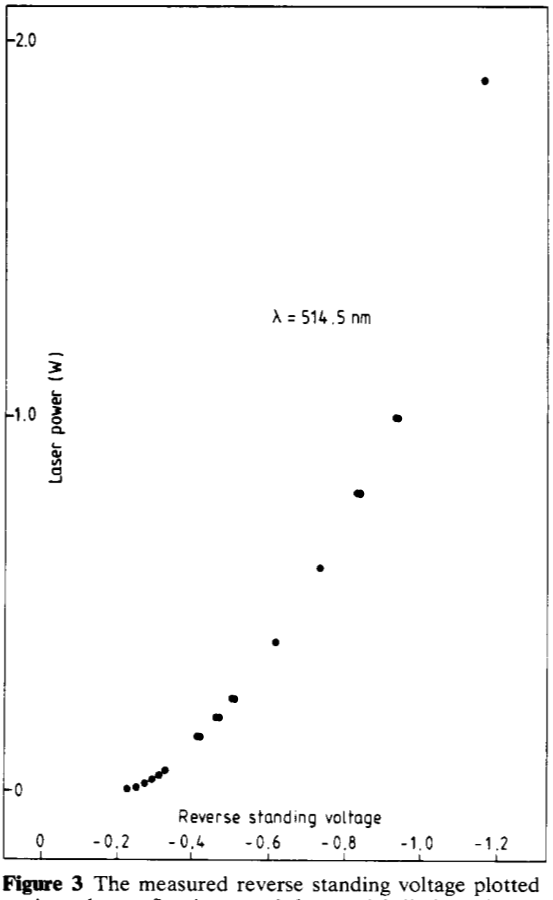

Figure 3 The measured reverse standing voltage plotted against photon flux in one of the special diodes using a Keithley 616 digital multimeter with $10^{11}\ \Omega$ input resistance.

This method of determining Planck's constant has not been pursued further.

2.2. The Millikan method

The photodiode designed for the previous experiment allows the observation of the $(I-V)$ curves without the complicating effects of a reverse current, however all the other effects mentioned in the introduction are expected to be present. Thus it is expected that no cut-off voltage will be observable because of the finite temperature of the cathode. It

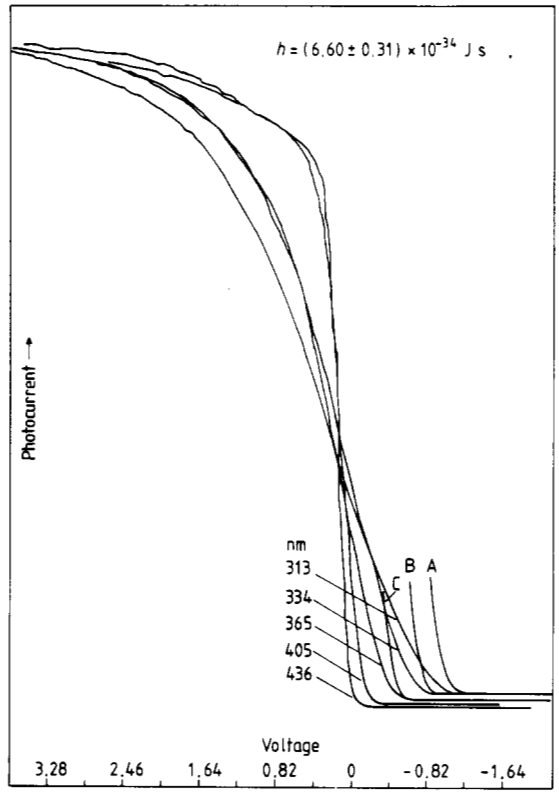

Figure 4 Observations of five photocurrent retarding curves with the onset regions of the 313, 334 and 365 nm curves magnified and replotted as A, B and C to be parallel to the 436 nm curve. In order that the replotting procedure should be successful there must be a very large forward to reverse current ratio.

is also expected that the energy at which the photocurrent becomes negligible with respect to the system background will increase with increasing light intensity. Figure 4 shows five $(I-V)$ curves obtained with one of the special diodes for wavelengths varying between 313 and 436 nm. The curves were obtained with a mercury arc lamp and monochromator, and the light was focused to a

small spot at the centre of the cathode. The curves have been normalised to display the high current regions on the same scale; the feet of the curves have been displaced slightly in order to show them clearly.

As expected the curves approach zero current asymptotically, there being no well defined cut-off energy. The small curve marked A is the foot of the 313 nm curve replotted by increasing the current gain by a factor of 3. This is equivalent to increasing the light intensity by the same factor and shows the shift in energy at which the photocurrent becomes effectively zero, with light intensity. The detailed shapes of the tails of these curves may be due to any or all of the effects mentioned in the introduction, however, as there is no observable cut-off energy, there is no way the Einstein theory can be used to obtain Planck's constant from them.

3. A theoretical analysis of the $I-V$ curves in a planar photodiode

In order to obtain an understanding of the relative importance of the various broadening mechanisms of the $I-V$ curves the following analysis has been performed. It is only expected to be successful when the Fermi energy is large compared with the most probable energy of photoemission.

We wish to know the photocurrent as a function of energy and angle for various photon energies in order to compute the $(I-V)$ curves of the diode. Thus consider a small volume $d\tau$ in the cathode and let the free-electron number density be n. Referring to figure 5 the number of free electrons in $d\tau$ is

$$nr^2 \sin\theta \, dr \, d\theta \, d\varphi.$$

Let $F(E)$ be the density of occupied electron states, then the number of electrons in $d\tau$, with energies between E and $E+dE$, $dN(E)$, is

$$dN(E) = nF(E)r^2 \sin\theta \, dr \, d\theta \, d\varphi \, dE. \qquad (1)$$

For the alkali metals it is reasonable to assume that $F(E)$ is the free-electron energy distribution function, however, for the transition metals, if accurate results are to be obtained for the higher photon energies, the appropriate density of occupied states function for the metal concerned must be used.

Letting the photon flux through $d\tau$ be $P\,s^{-1}\,m^{-2}$ and the photon absorption cross section be $Q(E)$; the number of photoelectrons excited in $d\tau$ per second in the energy range E to $E+dE$ is equal to

$$PQ(E)nF(E)r^2 \sin\theta \, dr \, d\theta \, d\varphi \, dE. \qquad (2)$$

Assuming that these electrons are isotropically and elastically scattered out of $d\tau$ the number arriving at dS per second is

$$(1/4\pi)PQ(E)nF(E)\exp(-r/\lambda(E')) \times \sin\theta \cos\theta \, dr \, dS \, d\varphi \, d\theta \, dE \qquad (3)$$

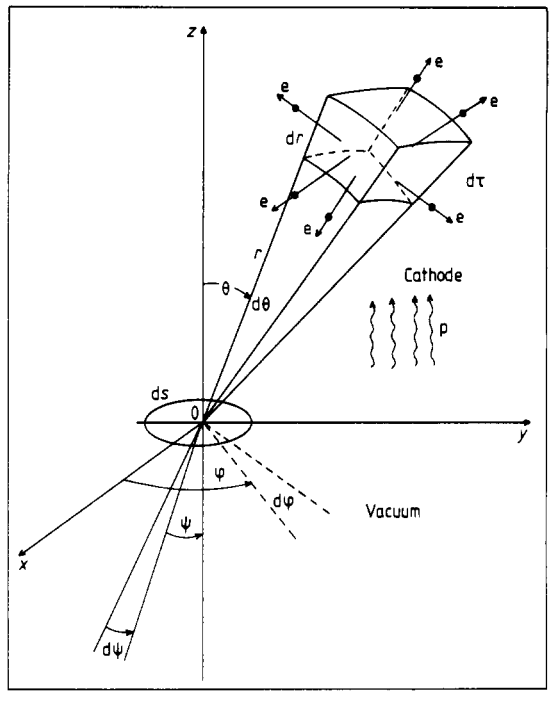

Figure 5 Photoelectrons are scattered out of $d\tau$, some of which arrive at ds without further scattering.

where $E' = E + h\nu$, the energy of the photoelectrons in the metal and $\lambda(E')$ is the energy-dependent mean free path for electron scattering in the metal. Over the escape depth of the photoelectrons the photon flux is sensibly constant because the mean free path for photon scattering is some forty times greater than for electron scattering. Integrating the above expression over r and φ gives the flux of photoelectrons coming from angles θ to $\theta+d\theta$ in the energy range E to $E+dE$ as

$$\tfrac{1}{2}PQ(E)\lambda(E')nF(E)\sin\theta\cos\theta \, d\theta \, dE \, dS. \qquad (4)$$

For an electron to escape into the vacuum its kinetic energy, $\frac{1}{2}mv^2$, angle of approach θ, and the potential barrier Φ, must be related by: $\frac{1}{2}mv^2\cos^2\theta \geq e\Phi$, $(e\Phi = E_F + e\phi$ i.e. the Fermi energy plus the work function). The kinetic energy of the electron is equal to the energy it possessed from the energy distribution plus the energy of the photon; thus $(E+h\nu)\cos^2\theta \geq e\Phi$. For an electron to escape, its approach angle $\theta \leq \cos^{-1}[e\Phi/(E+h\nu)]^{1/2}$. The number of electrons which can escape is then found by integrating equation (4) over θ to give

$$\tfrac{1}{4}PQ(E)\lambda(E')T(E',\theta)nF(E) \times\{1-[e\Phi/(E+h\nu)]\}\,dE\,dS \qquad (5)$$

where $T(E',\theta)$ is the barrier transmission factor.

These electrons emerge into the electric field of the planar diode with an energy $E+h\nu-e\Phi$ and at

all angles between $0 \rightarrow \frac{1}{2}\pi$. Using equation (4) together with the condition that the transverse velocity is unmodified during escape, the differential flux of photoelectrons in the vacuum between ψ and $\psi + d\psi$ is found to be

$$dN(E, \psi) = \tfrac{1}{2}PQ(E)\lambda(E')T(E', \theta)nF(E) \times [1 - e\Phi/(E + h\nu)] \sin\psi \cos\psi \, dE \, dS \, d\psi. \quad (6)$$

For a plane diode with an infinite ratio of anode radius x_0 to anode–cathode spacing y_0 the maximum angle of photoemission for collection at the anode at a retarding potential difference V and energy E is

$$\psi_{max} = \cos^{-1}\left(\frac{V}{E + h\nu - e\Phi}\right)^{1/2} = \cos^{-1}(V/E_v)^{1/2}. \quad (7)$$

For a finite anode radius x_0 there exists a minimum value for V/E_v at which all electrons which satisfy (7) strike the anode. In order to obtain ψ_{max} one must solve the equations of motion for an electron leaving the cathode at an angle ψ into a plane retarding field V/y_0, having some initial energy E_v. The maximum angle, ψ_{max}, with which an electron can leave the cathode and arrive at the anode, of radius x_0, having surmounted a plane potential barrier V is given by

$$\sin^2\psi_{max} = \frac{(1 - V/2E_v) + (1 - V/E_v - x_0^2V^2/4y_0^2E_v^2)^{1/2}}{2(y_0^2/x_0^2 + 1)}. \quad (8)$$

For any energy E_v there will be a range of values of V for which ψ_{max} is defined by (7), beyond this range it is necessary to use (8). Writing $\sin^2\psi_{max} = 1 - V/E_v$ equation (8) becomes

$$(1 - V/2E_v) + \left(1 - \frac{V}{E_v} - \frac{x_0^2}{y_0^2}\frac{V^2}{4E_v^2}\right)^{1/2} - 2(1 - V/E_v)(1 + y_0^2/x_0^2) = 0. \quad (9)$$

Solving (9) for various values of x_0/y_0 gives the critical values of V/E_v for various anode radii to diode spacings. The flux of electrons striking the anode originating from the energy distribution at an energy $E \rightarrow E + dE$ is then

$$dN(E) = \{\tfrac{1}{2}PQ(E)\lambda(E')T(E', \theta)nF(E) \times [1 - e\Phi/(E + h\nu)] \, dE \, dS\} \times (1 - V/E_v) \quad (10a)$$

or

$$\times \frac{(1 - V/2E_v) + (1 - V/E_v - x_0^2V^2/y_0^2 4E_v^2)^{1/2}}{2(1 + y_0^2/x_0^2)}. \quad (10b)$$

One uses (10*a*) for values of V/E_v between unity and the critical value found from solving (9), and (10*b*) for values V/E_v smaller than this critical value.

In order to obtain the differential current collected by the anode it is necessary to multiply (10) by the electronic charge and the anode collection efficiency $C(E_a)$; E_a being the energy of the electron when it strikes the anode. For the case of an accelerating voltage in the diode a different critical value for V/E_v is found by setting $\sin^2\psi = 1$ in (8) and solving for the various values of anode radius to diode spacing, x_0/y_0, of interest. To form the photocurrent as a function of potential difference, (10) is integrated from $E_v = V$ to ∞, in the retarding case, and zero to infinity in the accelerating case.

4. Some experimental and theoretical results

The photodiodes have a caesiated nickel cathode and an anode radius to diode spacing of approximately unity. The Fermi energy for nickel is about 9.0 eV; the potential barrier of caesiated nickel is taken to be 12 eV, measured from the bottom of the conduction band, i.e. −3 eV from the vacuum; T is taken to be room temperature.

The following approximations have been used in order to make (10) tractable. Over the range of electron energies in the conductor which take part in photoemission, $Q(E)$, the photon absorption cross section, $\lambda(E')$, the electron mean free path and $T(E', \theta)$, the electron transmission coefficient are all assumed constant. For a photon energy of 4 eV the energy range for photon absorption to produce photoemission is 8–9 eV, plus a few kT, and for internal photoelectron scattering it is 12–13 eV. In both cases the energy range is small compared with the actual energy and thus both $Q(E)$ and $\lambda(E')$ are expected to be reasonably constant. The angular effects of the transmission coefficient $T(E', \theta)$ are expected to be small because of the small angular range involved in near threshold photoemission.

The transmission coefficient $T(E')$ and the anode collection efficiency $C(E_a)$ are both expected to be strongly energy dependent in the threshold region; however, no account of this behaviour has been taken in evaluating the theoretical $I - V$ curves.

The density of occupied electron states, $F(E)$, for nickel has been taken from the experimental measurements of Pierce and Spicer (1972). The rapidly falling, high-energy part of the measured density of occupied states function has been replaced by the equivalent part of the free-electron function, for a temperature of 293 K. This procedure has been adopted in an attempt to remove the effect of finite-energy resolution from the function and to give it its almost inevitable thermal tail.

The results of experimental photocurrent–voltage, $(I - V)$ curves, for photon energies of 2.85 and 3.96 eV are shown in figure 6 compared with the theoretical curves for 3.2 and 4.0 eV photons. The experimental curves have been obtained using a high-pressure mercury arc lamp, Spex Minimate

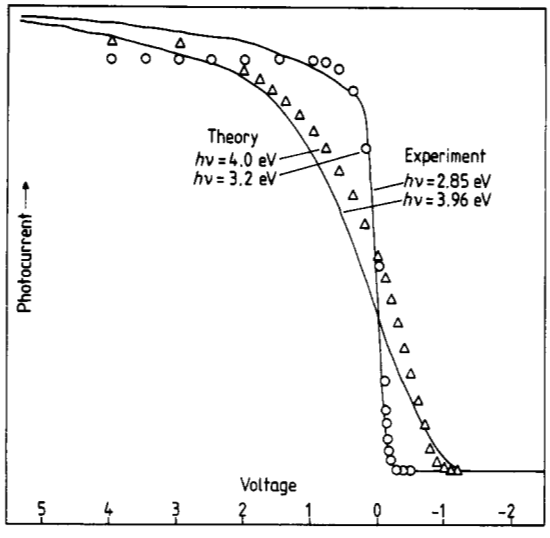

Figure 6 Comparison of experimental (——) and theoretical (○, △) photocurrent retarding curves for a planar nickel/copper diode with anode radius to diode spacing ratio of 1:1; at room temperature.

monochromator with 1 mm diameter 'target' slits and a Keithley 616 dmm. The photocurrents are in the range $3\times10^{-9}\rightarrow10^{-10}$ A. The voltage is swept with a Feedback generator and the current is automatically plotted against voltage with a Hewlett–Packard pen recorder. The experimental curves have been plotted so that they have approximately the same displacement on the current axis. The theoretical curves have been scaled to equal the experimental current at the voltage at which the theoretical currents saturate.

Quantum efficiency measurements show that the work function of the cathode is about 2.65 eV, giving electrons photoemitted at the Fermi energy with a photon of energy 2.85 eV, a vacuum energy of 0.2 eV. As the value of the work function used in the theoretical calculations is 3.0 eV, electrons excited at the theoretical Fermi energy by photons of energy 3.2 eV will also have a vacuum energy of 0.2 eV. Thus the experimental 2.85 eV curve should be directly comparable with the theoretical 3.2 eV curve. The 3.96 eV experimental curve should lie 0.31 eV to the right of the 4.0 eV theoretical curve. The overall agreement between experiment and theory is surprisingly good. The theory shows that the major difference between curves for different photon energies is due to the difference in accelerating voltage required to collect high- and low-energy photoelectrons in the diode. The photoelectrons are emitted into the vacuum at all angles, thus the accelerating voltage required to cause all large-angle, high-energy electrons to strike the anode is much higher than that for the large-angle low-energy electrons. The tail in the retarding region of both curves is in reasonable agreement with the theory, showing that this region is dominated by the effects of temperature and geometry. No account of any variation in work function of the cathode has been taken in evaluating the theoretical $(I-V)$ curves and thus it may be concluded that it is unimportant in this particular diode.

Figure 7 shows the theoretical $(I-V)$ curves without any scaling; for equal fluxes of 3.5 and 4.0 eV photons. A point of interest is that the two curves appear parallel over a significant part of the foot of the retarding region. Figure 8 shows the threshold region on a much enlarged energy scale with the 3.5 eV curve plotted in its correct energy position and also displaced to the right by the difference in photon energies, i.e. 0.5 eV. It will be noticed that the curves are very reasonably parallel over about 0.2 V around the tails and as they tend to zero current they become coincident. Thus if the $(I-V)$ curves were measured with *the same photon flux for both photon energies*, h could be estimated reasonably accurately by measuring the energy displacement of the parallel retarding curves in the observable onset region. 'Parallel' is here used to describe the property of 'continuous equidistance': it does not imply any rectilinear behaviour.

In hindsight one might expect the tails of the various $(I-V)$ curves to be displaced, one from another, by approximately $h(\nu_i-\nu_j)$. This occurs because these tails are dominated by the region of

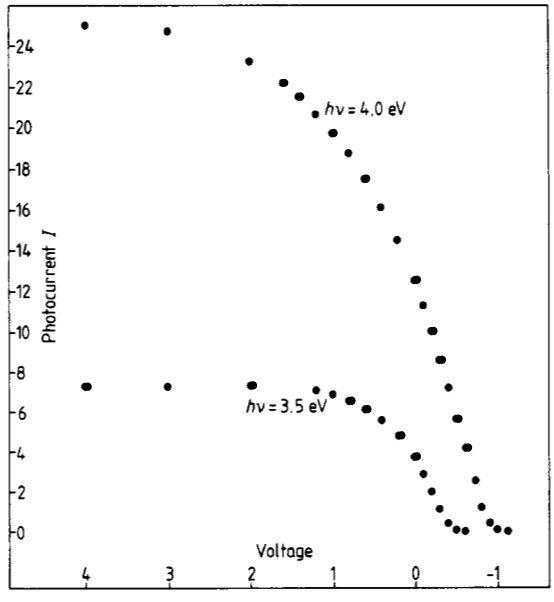

Figure 7 Theoretical photocurrent retarding curves for anode radius to diode spacing ratio of unity at room temperature, showing the onset regions for equal photon fluxes, in a nickel/copper diode.

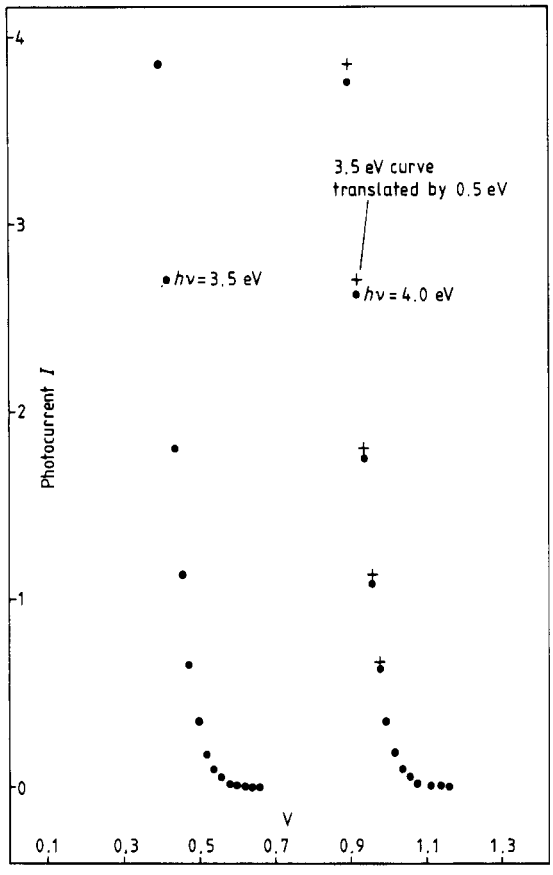

Figure 8 Enlarged threshold regions of figure 7 with the retarding curve for a photon energy of 3.5 eV plotted in its correct position and also shifted by 0.5 eV, for direct comparison with the curve for 4.0 eV photon energy.

rapid fall of the energy distribution in the metal. In escaping from the metal all these electrons simply increase in energy by $h\nu - e\phi$. The energy distribution is modified in the process of transmission by geometrical and quantum mechanical factors, however in a planar diode the geometrical modification upon transmission is compensated, to a large extent, by the collection geometry of the diode. The quantum mechanical effects are of secondary importance except possibly for the barrier transmission coefficient $T(E', \theta)$ which appears, from experiment, to be insignificant. That the curves should be accurately parallel for equal photon fluxes, over an extended energy range in the region of the tail is an accident of planar geometry.

There are two difficulties in making measurements with identical photon fluxes; firstly the experimental apparatus requires a calibrated radiometer; secondly as the quantum efficiency may vary by a factor of twenty or more with wavelength, displaying the curves on the same current scale may make those near threshold difficult to observe. It is possible to simulate the effect of measurement using equal photon fluxes, for the y gain can be altered and the threshold regions replotted so that they all appear parallel. The effect of this procedure is shown in figure 4 as the small curves A, B and C; it eliminates the problem of determining the 'cut-off point' and only requires direct measurements of lateral displacements between parallel curves. The value of h obtained using this procedure on these particular results was $(6.60 \pm 0.31) \times 10^{-34}$ Js.

It will be evident that by arbitrarily choosing the system y gain or photon flux in a line, the relative positions of the retarding curves for various wavelengths is, to some extent, arbitrary. The 'geometrical theory' gives guidance upon how to display the retarding curves if a reasonable value of h is to be obtained, however the approximations inherent in the evaluation may lead to errors which are very difficult to estimate.

5. Experiments with a guard ring diode and triode

There are several approximations in the theory which make the level of agreement with experiment surprising. It has been assumed that a plane field exists between the anode and cathode, however it is expected that there will be serious field fringing and thus the electron trajectories will not be parabolic. In order to overcome this problem a guard ring was

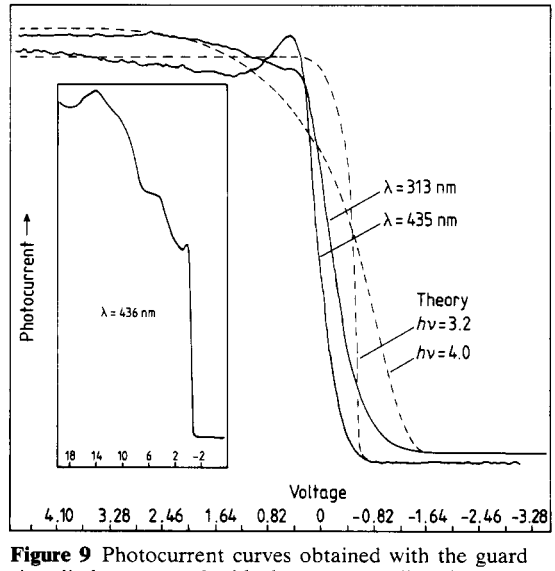

Figure 9 Photocurrent curves obtained with the guard ring diode compared with the corresponding theoretical results. The effect of the energy-dependent anode reflection coefficient, $C(E_a)$, is very marked. The inset shows the effect of the variation of the anode reflection coefficient over an enlarged energy range.

incorporated in the plane of the anode of three times its diameter. Figure 9 shows that the agreement between experiment and theory was rather fortuitious. Now that the field fringing has been removed the effect of the energy-dependent anode collection efficiency, $C(E_a)$, becomes evident. It can be seen that the field fringing was partially compensated by the energy dependence of $C(E_a)$. Inset in figure 9 is a plot of anode current over an energy range of −6–20 V. The structure that is seen agrees well with that of McRae and Caldwell (1976) for the reflection coefficient for electrons on oxidised single crystals of copper.

It is possible to remove the complicating effects of $C(E_a)$ if all the electrons are collected at effectively the same energy, independent of the diode field. Thus to test the limits of applicability of the simple theory a phototriode was constructed. A 1 cm diameter 750 line per inch, oxidised copper, electroform mesh grid separated by 1 cm from the plane nickel cathode defined the active area of the diode. The grid is at the centre of a 3 cm diameter copper disc to avoid the problem of field fringing. A Faraday cup type collector resides behind the grid which is kept at 2 V positive with respect to the cathode.

If an electron strikes the grid it will be collected or reflected with an efficiency that depends upon energy. For the case of a retarding field, if an electron is reflected there is only a small chance that it can return to the grid for a second attempt at transmission. The reason is that there is expected to be reasonably isotropic scattering at the grid and cathode, and the chance that an electron is scattered at both the grid and cathode into the small solid angle required for return to the grid is small. Thus the collected current should be dominated by that which is directly transmitted by the grid.

Figure 10 shows the experimental $(I-V)$ curves for photons of two different energies together with the theoretical results for an anode radius to diode spacing ratio of 1:2. The theoretical curves have been plotted so that their amplitude is equal to the experimental curve at the theoretical saturation voltage. The general behaviour of the theoretical curves is reasonably well represented by experiment, the agreement being better in the retarding than the accelerating region. The disagreement in the accelerating region is due, at least in part, to the energy-dependent reflection coefficient of the grid, for it will be noticed that the experimental curves have a broad maximum in the region 9–15 V, where one would expect a region of monotonically increasing current.

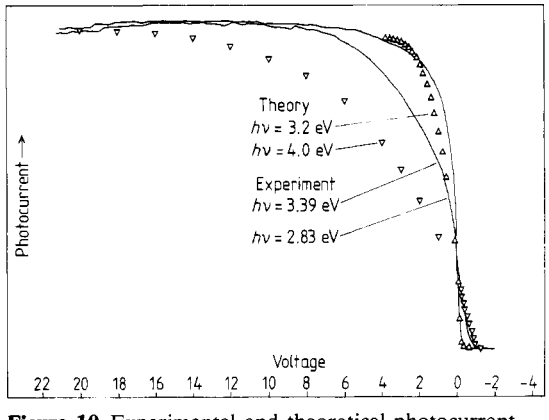

Figure 10 Experimental and theoretical photocurrent curves for the phototriode. The effective anode radius to diode spacing ratio is 0.5.

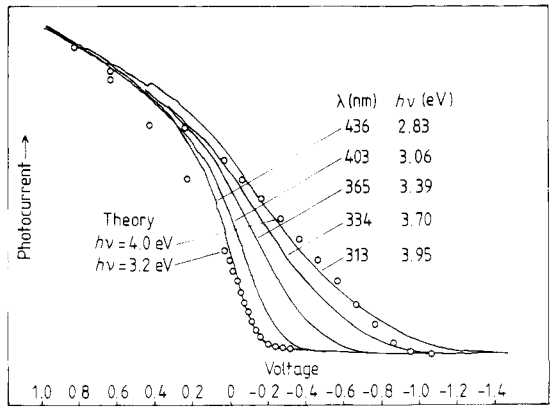

Figure 11 Detail of the threshold region of figure 10 together with other retarding curves showing the good agreement between theory and experiment for the λ = 436 nm curve.

The cut-off region of the curves is shown in more detail in figure 11 in comparison with the theoretical curves. There is good agreement for the lowest energy photoelectrons; this is to be expected for the approximations of constant excitation cross section $Q(E)$ and mean free path $\lambda(E')$ together with the small escape cone angle are best met in the threshold region. However the energy-dependent barrier transmission coefficient $T(E')$ would be expected to be most important in modifying the energy distribution for this group of electrons. Comparison of the measured relative quantum efficiency against wavelength with predictions of the theory seems to confirm the lack of significant energy variation of $T(E')$ at 0.2 eV above threshold.

There is known to be a spread of 0.3 eV in the work function of nickel with crystal orientation. An effect of this size would be easily seen in these measurements if it were present. It must be concluded that the present nickel surfaces are dominated by one particular crystal orientation.

In a careful single measurement using the 434 and 366 nm mercury lines, assuming the theoretical condition of parallelism of the $(I-V)$ curves in the threshold region for equal fluxes of photons, h came out to be $6.79 \pm 0.20 \times 10^{-34}$ Js.

6. Systematic error due to geometry

Figure 12 shows in detail the theoretical threshold regions for photon energies of 3.2 and 4.3 eV as a function of R, the ratio of anode radius to diode spacing for a caesiated nickel cathode in a planar diode. The 3.2 eV photon curves have been translated 1.1 eV to higher energies for ease of comparison. For $R < 1.3$ the gradient of the 3.2 eV photon curve is less than that for the 4.3 eV photon curve over the whole energy range. For $R = 1.5$ the curves cross at 1.19 eV and for $R > 4$ the gradient for the 3.2 eV curves is greater than that for 4.3 eV curves over the whole of the retarding region of the 3.2 eV curve. This behaviour can lead to systematic errors in the estimation of h.

A popular method of determining the cut-off point is that of constructing tangents to reasonably straight parts of the retarding curves. In the unlikely event of a light source having equal photon fluxes in all lines the resulting percentage error would be $+0.7\%$ for $R = 0.5$, decreasing to -2% and $R = 1$ and stabilising at -1% for $R > 4$.

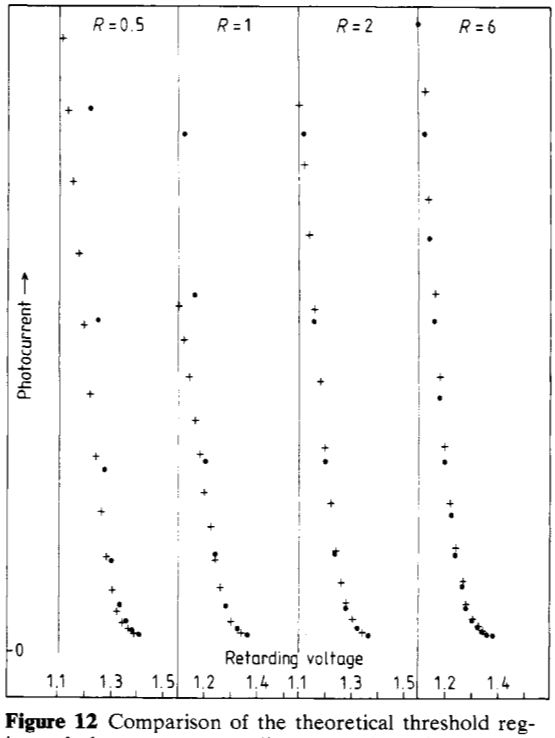

Figure 12 Comparison of the theoretical threshold regions of photocurrent retarding curves in a planar diode for several different anode radii to diode spacing ratios, R, at room temperature. The + data are for 3.2 eV. photons replotted 1.1 eV to higher energies for direct comparison with the 4.3 eV photon curves (·).

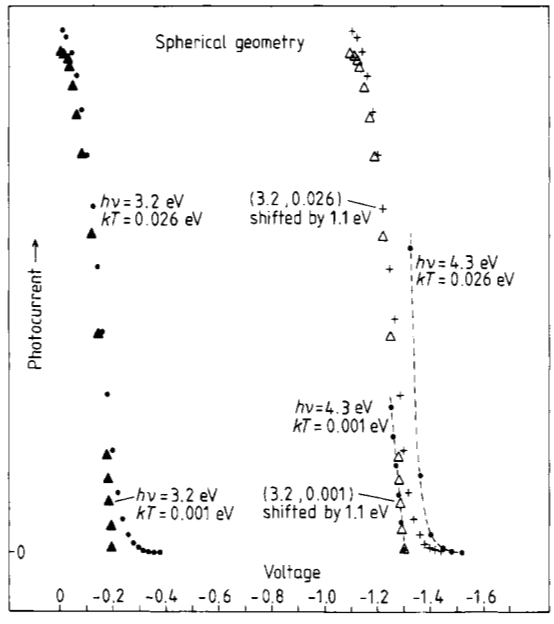

Figure 13 Theoretical photocurrent retarding curves in spherical geometry at room temperature and 8 K for equal fluxes of 3.2 and 4.3 eV photons. The 3.2 eV curves have been plotted in their correct position and also shifted by 1.1 eV to higher retarding energies. The thermal tail at 8 K is too small to be plotted on this scale. The effect of temperature upon this geometry is very marked.

The computed curves for a point caesiated nickel cathode at the centre of a spherical anode at room temperature and at 8 K are shown in figure 13. If one were to measure h in this geometry at room temperature using equal fluxes of 3.2 and 4.3 eV photons and estimate the cut-off energy by constructing tangents to reasonably straight portions of the retarding curves it would lead to an error of $+5\%$. As the temperature is reduced to zero the error reduces to zero. This error will depend upon the ratio of photon fluxes in the two lines. If one had a very strong red line, say an order of magnitude more intense than a blue line, the error would be much reduced and could even become negative.

The effect of temperature upon the errors in plane geometry is much less marked. At room temperature the tail of the $I-V$ curve is about equally due to geometry and temperature, and remains concave to the energy axis at all temperatures.

7. The early experimental work upon the photoelectric effect

The critical experiment in establishing the Einstein interpretation of the photoelectric effect was that of Millikan (1916a). The experiments of Hughes (1912) and of Richardson and Compton (1912) showed that there appeared to be a linear relation between frequency and stopping potential but the gradient of the function was not h. Hughes' value was about 20% low and that of Richardson and Compton had a mean value of 5.4×10^{-34} Js, but varied with the metal being observed. Millikan did not only obtain an extremely good straight line relation between frequency and stopping potential, but for the first time the gradient gave h in very good agreement with the accurate value already known from black body radiation measurements.

Millikan's results are interesting because he explicitly states that the $(I-V)$ curves 'plunge sharply into the potential axis and do not approach it asymptotically', Millikan (1961b). In fact the measured curves exhibit long asymptotic tails, however he established experimentally that these tails were due *solely* to the presence of stray light. Once this had been demonstrated Millikan (1916a, p 370) felt justified in removing them and 'cut the feet off'.

The absence of tailing is surprising on three counts. Firstly for there to be no thermal tail to the $(I-V)$ curves the cathodes would have had to be at a very low temperature indeed, this is known not to be the case. In discussing thermal effects Millikan states that the photoelectrons do not take part in the energies of thermal agitation as might be expected. Secondly the function which he observed has a property which is unique in experimental physics, that is, it is not continuously differentiable. Thirdly the observation of a discontinuity in the retarding curve requires the retarding grid energy analyser to have infinite energy resolution.

The magnitude of the room temperature thermal tail in Millikan's experiment is expected to be approximately the same as that shown in the theoretical $(I-V)$ curves for the spherical diode, see figure 13 ($h\nu=3.2$, $kT=0.026$ eV). After filtering out all traces of short-wavelength light from the spectrometer his observations on lithium showed no appreciable tailing down to 0.3% of saturation current, whereas one would expect this effect to be noticable at 5–10%.

I have no convincing explanation for Millikan's results but it may be significant that the lithium results look reminiscent of $(I-V)$ curves which are composed of a large forward and small reverse photocurrent, there being some mechanism in the apparatus which supresses the detection of the reverse current.

Lukirsky and Prilezaev (1928) used a spherical geometry to investigate the Einstein photoelectric equation, however, their experiment suffered from having a large reverse photocurrent and consequently they did not observe any thermal tailing. However, they found 'exact cut-off points' indirectly and obtained a value for h of $6.543\pm0.01\times10^{-34}$ Js. Olpin (1930) performed the experiment in cylindrical geometry and measured the 'cut-off potentials' against frequency to obtain a value for h of $6.541\pm0.006\times10^{-34}$ Js.

By 1932 Hughes and Dubridge could lay the foundation of the photoelectric effect in metals upon a wealth of experimental evidence in *exact* agreement with Einstein's theory—a very satisfactory state of affairs.

Acceptance of the strange results obtained by Millikan and others up until 1930 have had a number of unfortunate consequences, which include the following.

(i) The use of Einstein's theory of photoemission from metals to determine Planck's constant, when in fact it is not applicable because of the unobservable nature of the stopping potential E_{max}.

(ii) For some time it was believed that photoelectrons and conduction electrons originated from different states in the metal, however, this is now generally accepted not to be the case.

(iii) Several generations of undergraduate textbook have made claims about the photoelectric effect which are not borne out by direct experiment and are incompatible with other branches of physics.

8. Summary and conclusion

The major errors encountered in an estimation of Planck's constant in the experiment described in this paper stem from sensitivity of the position and gradient of the $I-V$ curves to light intensity. Further the Einstein theory is not strictly applicable to photoemission from metals, for a determination of Planck's constant, at any finite temperature because of the unobservable nature of the parameter E_{max}.

The mechanism proposed by Einstein to account for the energy spread of the photoelectrons is not the dominant one. To a large extent the energy spread in the vacuum reflects that in the metal. The reason for this is that, in near threshold photoemission, the escape cone angle in the metal is small and an inelastic collision would cause the probability for escape to become very small.

That the thermal tail causes the cut-off energy to be unobservable has been known for a long time. Fowler (1966) pointed this out when discussing the work of Morris (1931) on photoemission at very low temperatures. Ritchmyer *et al* (1956) mentioned it in connection with thermionic emission. Cohen *et al* (1957) state; 'Unfortunately the difficulty here is that V_0 (the cut-off energy) is ill defined experimentally; the curve of photoelectric current versus retarding potential, V, does not present a sharp intercept with the zero axis but rather has an asymptotic transition to zero'. However,

there has been considerable diffidence in coming to the unavoidable conclusion that Einstein's equation for photoemission from metals is simply incorrect. The cardinal point of the theory, that is the postulate that $E = h\nu$ is without doubt correct; the irony of the situation is that the application of the postulate to photoemission from metals and the experimental verification were both in error.

Finally, may I suggest that the theory of photoemission from metals be introduced to students in the following way.

If an electron of energy E absorbs a photon of energy $h\nu$ and it subsequently escapes from the metal into the vacuum, without energy loss in the metal, it will have an energy E_v given by

$$E_v = E + h\nu - e\Phi$$

where $e\Phi$ is the work the electron must do to escape from the metal. At, and only at, absolute zero temperature there is a very well defined maximum energy for electrons in a metal, it is known as the Fermi energy E_F. Thus the maximum energy of a photoelectron originating from a metal at absolute zero is

$$E_{v_{max}} = E_F + h\nu - e\Phi$$

$E_F - e\Phi$ is known as the work function $e\phi$. Thus at absolute zero

$$E_{v_{max}} = h\nu - e\phi \qquad \text{Einstein's famous equation.}$$

Acknowledgments

I am pleased to acknowledge the time-consuming help given to me by Mr S Moehr in the manufacture of the various phototubes. To my colleagues in the Physics Department for their willingness to listen and particularly to Dr R C Greenhow and Dr J A D Matthew for many discussions, I extend my thanks. I also wish to thank Dr N Anderson of the Department of Mathematics for his continued interest in the problem, and his assistance with computing.

Appendix

The anode and the cathode are conventionally cleaned and then blasted with fine glass beads so that they have matt surfaces. The tube is evacuated and baked to about 130 °C which produces a base pressure of 5×10^{-8} Torr. The cathode is then heated to approximately 900 °C using an induction heater until it is thoroughly degassed. The diameter of the cathode is made large so that the caesium generator and anode are screened from the field of the induction coil.

The tube is sealed off from the vacuum system and the cathode caesiated while the photoemission is monitored. A guide to the caesium exposure can be given roughly as 5 A for 1 min through an SAES NF/3.9/12 generator. The leakage resistance is above $10^{14}\ \Omega$, while the forward to reverse photocurrent ratio is in the range 5×10^2–5×10^3 at 313 nm and rises to 5×10^3–5×10^5 at 436 nm.

References

Cohen E R, Crowe K M and Dumond I W M 1957 *Fundamental Constants of Physics* (London: Interscience)

Fowler R H 1966 *Statistical Mechanics* (Cambridge: Cambridge University Press) p 359

Hughes A L 1912 *Phil. Trans. R. Soc.* **212** 205

Hughes A L and Dubridge L A 1932 *Photoelectric Phenomena* (New York: McGraw-Hill)

Lukirsky P and Prilezaev S 1928 *Z. Phys.* **49** 236

McRae E G and Caldwell C W 1976 *Surface Sci.* **57** 77–92

Millikan R A 1916a *Phys. Rev.* **7** 355

—— 1916b *Phys. Rev.* **7** 18

Morris L W 1931 *Phys. Rev.* **37** 1263

Olpin A R 1930 *Phys. Rev.* **36** 251

Pierce D T and Spicer W E 1972 *Phys. Rev.* B **6** 1787

Richardson O W and Compton K T 1912 *Phil. Mag.* **24** 575

Ritchmyer F K, Kennard E H and Lauritsen T 1956 *Introduction to Modern Physics* (New York: McGraw-Hill)