REVIEWS OF
MODERN PHYSICS

VOLUME 21, NUMBER 2 APRIL, 1949

Thermionic Emission

CONYERS HERRING*
University of Texas, Austin, Texas, and Bell Telephone Laboratories, Murray Hill, New Jersey
AND
M. H. NICHOLS**
Princeton University, Princeton, New Jersey

TABLE OF CONTENTS

* Present address, Bell Telephone Laboratories, Murray Hill, New Jersey.
** Most of the work was done while this author was a National Research Fellow at Princeton University. The work was completed at the University of Michigan where he is presently located.

INTRODUCTION

NO broad review of the field of thermionic emission has been published in English since 1936.*** Since that time our understanding of many thermionic phenomena has advanced considerably, as a result of improved experimental techniques and more careful thought on theoretical interpretations. Now many of the principles involved in the interpretation of experiments on thermionic or photoelectric emission from contaminated metals, oxide cathodes, or other surfaces are already clearly illustrated in the case of clean metals, for which the various phenomena can usually be more unambiguously interpreted than in the other fields. This fact suggests that the need for an up-to-date review of thermionic emission may be filled more suitably by a detailed and critical review devoted mainly to thermionic emission from clean metals, rather than by a more sketchy review of a wider field. The present article is therefore restricted to clean metals, with the hope that by emphasizing basic principles it will assist workers in the various other fields to avoid misinterpretations of experimental results such as have often occurred in the literature. To this end an attempt will be made throughout to point out what modifications, if any, must be made when applying the concepts under discussion to fields other than that of clean metals. The important field of space-charge limited emission will not be treated here, as it has been excellently summarized by Langmuir and Compton (L5) and by Becker (B8).

Perhaps foremost among the developments of recent years has been the realization that nearly all experiments involving polycrystalline emitters are greatly influenced by the occurrence of differences in work function between the different patches of the cathode surface. The theory of some of these so-called "patch effects" was elaborated by Becker (B8) in his 1935 review, which also gives an account of the earlier literature. Developments of the past decade have confirmed Becker's picture and supplied quantitative values for some of the parameters entering into it. Thus by now there is need for a more extensive discussion of the role of patches in all the main types of thermionic experiments. A major portion (Chapter II) of this article will be devoted to this end, omitting, however, any discussion of the interplay of patch and space-charge effects.

Prerequisite to an understanding of phenomena involving patchy surfaces is of course a clear comprehension of the principles underlying emission from uniform surfaces. The thermodynamic aspects of these principles were explored fairly thoroughly two or three decades ago, but unfortunately there exist few treatments of this subject which are both sound and thorough, and none of these is concise. To fill this gap the present review will commence with a brief exposition of this classical material, with such modifications as have been suggested by recent theoretical developments. The reader interested in a lenghthier discussion of some of the material covered may be referred to the article of Schottky and Rothe (11) and the book of Bridgman (2).

The remaining two chapters of this article are devoted, respectively, to a review of data on thermionic emission from clean metals which have been published since Becker's review (B8), and to a survey of modern quantum-theoretical developments bearing on thermionic and related phenomena. The latter survey is intended primarily for the theoretical physicist interested in this field, but we hope it will be of value to many experimentalists also.

Because of the length of this review, an effort has been made to present the material in such a way that any of the four chapters, and in many cases any of the main sections, can be read by itself with a minimum of consultation of other parts of the article. Also, to avoid digressions from the main thread of the presentation, much important material has been relegated to appendices.

The appended bibliography lists all papers since 1935 which have come to the authors' attention and which bear on the fields covered by this review. References to earlier papers have been included as far as is necessary to an orderly exposition of the subject matter, but without any attempt at comprehensiveness.

Although all portions of this article have been planned and discussed by the authors jointly, the principal responsibility for Chapters I and IV and Appendices I, II, III, V, and VI rests with the first author (C.H.), that for Chapters II and III and Appendix IV with the second (M.H.N.).

CHAPTER I. THERMODYNAMICS OF EMISSION FROM UNIFORM SURFACES

I.1 Introductory Remarks on the Application of Thermodynamics to Thermionic Phenomena

We shall be concerned in this chapter with a variety of phenomena such as contact potential difference, the variation of saturation current with temperature, the cooling effect accompanying emission, etc., between which a number of relationships can be deduced. These relationships can be deduced in a variety of ways, each of which has advantages of its own. To aid in visualizing the

*** The following treatises and reviews listed in the bibliography at the end of this article contain a general discussion of thermionic emission from metals: Richardson (10), (1921); Bloch (1) (1923); Schottky and Rothe (11) (1928); Dushman (D7) (1930); Suhrmann (17) (1934); Becker (B8) (1935); Reimann (9) (1935); Rukop, Schottky, and Suhrmann (R20) (1935); Jones (8) (1936); Herrmann and Wagener (6) (1943).

physical processes underlying them it is sometimes useful to derive the relationships by assuming a particular statistical model for the electrons inside a conductor. This method has the disadvantage, however, that it does not tell us which of the relationships we deduce are valid only for the particular model we have assumed, and which are of universal validity, valid for any model obeying the basic laws of statistical mechanics. An alternative method of attack is to try to deduce as many relationships as possible from the principles of thermodynamics, and to introduce detailed assumptions regarding the mechanism of emission only when it proves impossible to reach the desired conclusion without them. This procedure, which is of course equivalent to using statistical mechanics with an absolute minimum of particularizing assumptions, has the advantage of showing clearly the extent to which any given theoretical formula can be relied on. This chapter will be devoted to theoretical considerations of this latter type, leaving most of the discussion of particular models and specific mechanisms to Chapter IV.

The starting point of the thermodynamic treatment is the assumption that the phase under consideration, be it metal, semiconductor, or space-charge filled vacuum, can be divided up into regions each of which can be treated as a thermodynamic system. By this is meant that each such region must be sensibly uniform in temperature and composition and must have substantially the same properties as it would have in thermal equilibrium. In some applications of the theory it suffices to treat the entire cathode as a single such region and a large but finite volume of the electron vapor as another single region. In other applications, however, one wishes to take account of nonuniformity of conditions inside the cathode—a factor likely to be important when applying the theory to semiconductors—or of variations in the electrostatic potential in the space-charge region outside. In such cases it is necessary to choose the elementary regions to be small, and it is pertinent to inquire how small they may be chosen. A rather lengthy discussion would be necessary to answer fully this question as to when it is permissible to treat small regions of metal, semiconductor, or vapor as thermodynamic systems. To avoid burdening the present section with too many details, we shall omit such discussion for the present and refer the reader interested in this question to Appendix I and to the article of Schottky and Rothe ((15), Kap. III) and shall merely quote the results of these discussions where they are needed.

I.1a *The Electrochemical Potential*

For our purposes the most convenient thermodynamic quantity to use in describing the state of any solid, liquid, or gaseous phase containing electrons is the so-called "electrochemical potential" of the electrons in it. To define this quantity, consider first the case of a single, isolated body containing n electrons; let its total energy be U, its absolute temperature T, its entropy S, and its volume v. Then the electrochemical potential $\bar{\mu}$ of the electrons in it is defined as****

$$\bar{\mu} = (\partial F/\partial n)_{T,v}, \tag{I.1.1}$$

where $F = U - TS$ is the Helmholtz free energy.† The dependence of S on n is made free from ambiguity by the requirement that for any condensed phase in equilibrium $(\partial S/\partial n)_{T,v} \to 0$ as $T \to 0$, in conformity with the convention of taking the entropy of a perfect crystal as zero at $T = 0$. If we are dealing with electrons in the vapor phase, this specification of the zero of entropy is equivalent to taking for the entropy the value calculated from quantum statistics ((18), #135). The dependence of U on n can be made similarly free from ambiguity by agreeing to take the energy of an electron at rest at infinity as having some specified value, e.g., zero.

The definition (I.1.1) is easily seen to be equivalent to the following: If a small number Δn of electrons are brought up from infinity and added isothermally to the body in question, keeping the latter at constant volume, the work required is $\bar{\mu}\Delta n$ minus the free energy which these Δn electrons originally had as vapor at infinity, the latter quantity to be computed with the above conventions regarding the zeros of energy and entropy.

So far we have considered only the case of an isolated body. When, as in most practical applications, there are a number of bodies in our system at different temperatures and potentials, the definition (I.1.1) may appear at first sight to break down. For because of the long range electrostatic forces, we cannot usually break the total energy up into a sum of contributions from the various bodies individually. However, we can still define $\bar{\mu}$ for any one of the bodies by the definition given in the preceding paragraph. This equivalent to writing for the electrochemical potential of the electrons in body i,

$$\bar{\mu}_i = [\partial F'/\partial n_i]_{\text{all } T\text{'s, } v\text{'s, other } n\text{'s}} \tag{I.1.2}$$

with $F' = U - \sum T_j S_j$, where U is the total energy

**** In this article $\bar{\mu}$ and the related quantity μ defined by (1.1.3) are expressed in units of energy per electron. This differs from the usual convention employed in the literature according to which the same symbols are used for energy per mole. Since the difference is merely one of units it seems proper to use the same symbols in both cases.

† Equivalent definitions (see Eq. (I.7.1) below) can be given in terms of the Gibbs free energy G, the energy U, or the heat content H:

$$\bar{\mu} = (\partial G/\partial n)_{T,p} = (\partial U/\partial n)_{S,v} = (\partial H/\partial n)_{S,p}.$$

of all bodies together, and T_j, S_j refer to the temperatures and entropies of the various bodies.

I.1b *The Electrostatic Potential and the Chemical Potential*

It is clear from the above definitions that any change $\Delta\Phi$ in the electrostatic potential Φ of the region to which electrons are added will change $\bar{\mu}$ by $-e\Delta\Phi$. Since one can change the potential of the interior of a conductor by shifting charges outside the conductor, by depositing a layer of dipoles on its surface, etc., this means that the $\bar{\mu}$ of a body is not a function of the internal state of the body alone, but depends on the condition of its surface and on external conditions. However, we may define a quantity μ, called simply the "chemical potential," by

$$\mu=\bar{\mu}+e\Phi. \tag{I.1.3}$$

This quantity will then be a function only of the local internal state of the body, and independent of surface and external conditions. The quantity Φ occurring in (I.1.3) may be defined in a variety of ways, since only changes in it have any macroscopic physical significance; however, the most convenient definition for our purposes is to relate it to the atomic picture of the body or region in question by defining Φ as the space average of the microscopic electrostatic potential over a volume containing many atoms but still small on a macroscopic scale (see (19), Chapter I).

I.1c *Properties of the Electrochemical Potential*

The property of the electrochemical potential $\bar{\mu}$ which makes it particularly useful is that the electrochemical potentials of the electrons in any two regions 1 and 2 in thermal equilibrium with each other must be equal. This is easily shown from the fact that the free energy of the whole system must be a minimum with respect to all other configurations which could be obtained by transferring electrons isothermally from the one region to the other: this means that the infinitesimal increment of free energy resulting from any such transfer must vanish, i.e., by (I.1.2)

$$\bar{\mu}_1=\bar{\mu}_2 \text{ in equilibrium.} \tag{I.1.4}$$

It is helpful in visualizing the significance of the electrochemical potential to relate it to the concepts of the usual statistical model of the free electrons in a body. In this model it is assumed that there are a large number of quantum states with different energies ϵ which are populated by non-interacting electrons. The distribution of electrons among these states follows the Fermi law, and it can be shown (e.g. (4), Chapter II) that the quantity $\bar{\mu}$ defined for this model by (I.1.1) or (I.1.2) is identical with the energy parameter occurring in the Fermi distribution function, so that

$$\text{probability that a state of energy } \epsilon \text{ is occupied}=\frac{1}{\exp(\epsilon-\bar{\mu}/kT)+1} \tag{I.1.5}$$

provided that the energy ϵ is measured from a zero representing an electron at rest at infinity. If ϵ is measured from a zero coinciding with the local mean electrostatic potential energy $-e\Phi$, then of course $\bar{\mu}$ in (I.1.5) should be replaced by μ. While this gives a useful picture of the meaning of $\bar{\mu}$, it must be emphasized that the original definitions of $\bar{\mu}$ and μ do not depend on the validity of any such simple model.

In the preceding discussion we have spoken of the various bodies or regions for which $\bar{\mu}$ is defined as if they were each of macroscopic size and uniform in temperature and other properties. The discussion of Appendix I alluded to at the beginning of this section shows, however, that it is meaningful to consider $\bar{\mu}$ defined as a function of position inside any conductor even in the presence of electrical and thermal gradients, provided these are not unreasonably large. The accuracy with which $\bar{\mu}$ can thus be localized is such that in practically all thermionic applications it is meaningful to speak of differences in the value of μ between points a few hundred angstroms apart. This is true whether the points considered are in a metal, a semiconductor, or a space-charge region in the vacuum.

I.1d *The Image Force and the Motive*

The behavior of electrons in the space just outside the surface of a conductor is greatly influenced by the attraction of each electron to its image in the conducting surface. If the surface is flat and perfectly conducting this image force has the magnitude $e^2/4x^2$, where x is the distance of the electron from the surface; the electron therefore behaves as if it were moving in an electrostatic field whose potential is the actual electrostatic potential plus $e/4x$. The name "motive" has been suggested for this fictitious potential, and will be employed in this article: motive is defined in general as a quantity whose gradient at any point gives $1/e$ times the force on an electron at that point. It is noteworthy that this conception of an image potential is not limited in its application to ranges and energies for which the concepts of classical physics are valid: Bardeen (B3) has shown that it is legitimate to use the image potential in the Schrödinger equation to calculate the wave function of an electron outside the surface of a metal. When the surface in question is that of a semiconductor the image force will of course be smaller

than for a metal. For an insulator of effective dielectric constant κ_1 the image force would be $(e^2/4x^2)\cdot(\kappa_1-1/\kappa_1+1)$; for thermal electrons at large values of x it would be reasonable to take for κ_1 the static dielectric constant κ of the material, while at small values of x the value κ_0 characteristic of near infra-red frequencies would be more appropriate. The electrical conductivity of a semiconductor will of course increase the image force to something intermediate between the value for a dielectric and that for a metal.

I.2 Definition of the True Work Function and Application to Contact Potentials

The "true work function" $e\varphi$ of a uniform surface of an electronic conductor is defined as the difference between the electrochemical potential $\bar{\mu}$ of the electrons just inside the conductor and the electrostatic potential energy $-e\Phi_a$ of an electron in the vacuum just outside it. Thus,

$$\varphi=-\Phi_a-(\bar{\mu}/e)=\Phi_c-\Phi_a-(\mu/e) \qquad (I.2.1)$$

by (I.1.3), where Φ_c is the electrostatic potential inside the conductor.

The first two terms on the right of (I.2.1) represent the potential difference between the inside and outside of the conductor and depend on the condition of the surface as well as upon the structure of the interior. Any change, Δ, in the dipole moment per unit area of the surface changes $(\Phi_c-\Phi_a)$ by $4\pi\Delta$. Since this dipole moment may be expected to be different for different faces of a single crystal, the work function will vary from face to face. The chemical potential μ, on the other hand, is a volume property independent of the structure of the surface; since it is one of the two terms in the expression (I.2.1) for the work function, it is sometimes called the "inner work function."

The true work function defined by (I.2.1) is, of course, in general a function of the temperature; one would expect, in fact, that both the inner work function and the potential jump should vary with temperature and by comparable amounts. (See the discussion in Section IV.3 below.) Actually, of course, the total change in φ over the range from absolute zero to incandescent temperatures is, at least for clean metals, only a very small fraction of φ itself. The only restriction on the temperature variation of φ which can be deduced from thermodynamic principles is that resulting from Nernst's third law. As applied to our problem this law states that as $T\to 0$ the entropy S of a crystal approaches a limiting value (which may be taken to be zero) independent of the concentration of electrons in it. Since $S=-\partial F/\partial T$, we have from (I.1.1)

$$\frac{\partial S}{\partial n}=-\frac{\partial^2 F}{\partial n\partial T}=-\frac{\partial\bar{\mu}}{\partial T}\to 0 \text{ as } T\to 0. \qquad (I.2.2)$$

This must hold for an isolated block of conductor of arbitrary shape. If we are interested in the work function of any particular crystal face, we may apply (I.2.2) to a thin slab cut parallel to this face, and have the result that

$$\partial\varphi/\partial T\to 0 \quad \text{as} \quad T\to 0. \qquad (I.2.3)$$

When two electronic conductors 1,2, are kept at the same temperature and connected electrically through a circuit containing no source of e.m.f., electrons will flow from the one to the other until an equilibrium state is reached. In this equilibrium state, according to (I.1.4), the electrochemical potentials of the electrons in the two conductors must be equal. By (I.2.1) this implies that there must be a difference of potential between a point just outside conductor 1 and a point just outside conductor 2, given by the difference of the two work functions:

$$\Phi_{a1}-\Phi_{a2}=\varphi_2-\varphi_1. \qquad (I.2.4)$$

This is called the "contact potential difference" between 1 and 2; the existence of such a difference is called the Volta effect, after its discoverer. It is to be emphasized that the validity of (I.2.4) rests on the assumption that the two conductors are in equilibrium at the same temperature. If their temperatures are different, a thermoelectric e.m.f. must be added to the right of (I.2.4). This will be discussed in Section I.6a.††

It is especially important to note that the contact potential difference (I.2.4) will in general be different from zero even when 1 and 2 refer merely to different crystal faces of the same conductor, or to different grains of the surface of a polycrystalline conductor. Most of Chapter II of this article will be devoted to a consideration of the effects of these "patch fields" on the measured properties of polycrystalline cathodes.

A few words of caution are needed concerning the various meanings which are to be found in the literature for the term "work function." This term is often used indiscriminately to denote quantities derived from a Fowler plot of photoelectric data, a Richardson plot of thermionic emission current, the cooling effect accompanying thermionic emission, and various types of contact potential measurement. Moreover, nearly all the experiments in the literature have been made on specimens with polycrystalline surfaces, whose properties, as we shall see in Chapter II below, may be related to those of the various single crystal faces in a complicated way. There are thus a variety of different though related physical quantities to which the

†† Incorrect derivations are sometimes encountered in textbooks which yield an additional term on the right of (I.2.4) equal to the Peltier heat at the junction of 1 and 2. Actually, (I.2.4) is exactly correct when the temperatures are the same.

name "work function" may be applied; in this article we shall endeavor to distinguish these with different names and symbols. For definiteness we shall restrict the use of the term "true work function," defined above, to homogeneous surfaces, although the definition (I.2.1) and the contact potential Eq. (I.2.4) can be made to apply to polycrystalline surfaces (see Section II.4).

I.3 The Richardson-Laue-Dushman Equation for the Thermionic Emission Current from a Uniform Surface

The thermodynamic derivation of the formula for the saturation emission current from a uniform surface involves the following steps:

(1) A calculation of the density of electron vapor a short distance away from the surface when in equilibrium with the emitter.

(2) A calculation of the number of electrons crossing a plane at this distance from the surface in unit time in either direction, in terms of the density of the vapor.

(3) The assumption that when "saturation current" is being drawn no electrons cross the plane going inward.

(4) The assumption that under these saturation conditions the number of electrons which cross the plane in unit time going outward is equal to the number calculated in (2) for equilibrium conditions minus a fraction of this number, representing ingoing electrons which are reflected at the surface of the conductor.

These steps illustrate the advantage mentioned in the first section which thermodynamic derivations possess over derivations which assume a particular statistical model. Here as much of the reasoning as possible involves processes taking place in the vapor phase outside the conductor, and the only properties of the latter which need to be assumed are those involved in steps (1) and (4).

1.3a *Equilibrium Density and Flux of Electrons*

Step (1) may be carried out by setting the electrochemical potential of the electrons in the vapor equal to the value $\bar{\mu}$ possessed by the electrons in the conductor, and assuming that the relation between the density of the vapor and its chemical potential μ_v is that given by the quantum statistics for a perfect gas of particles of statistical weight 2. For a region of the vapor phase well away from boundary surfaces, this relation is known to be (see for example (12), p. 151)

$$n_1=[2(2\pi mkT)^{\frac{3}{2}}/h^3]\exp(\mu_v/kT),$$

where n_1 is the number of electrons per unit volume in the vapor, and the other symbols have their standard meanings. We can extend the validity of this equation to regions where the image potential is appreciable by setting, in place of μ_v, the quantity $\bar{\mu}+eM_v$, where M_v is the motive at the given point in the vapor, thus

$$n_1=2[(2\pi mkT)/h^3]^{\frac{3}{2}}\exp(\bar{\mu}+eM_v/kT). \quad (I.3.1)$$

This assumption of perfect gas characteristics can be shown to be entirely justified for almost all attainable densities of the vapor.††† This may be demonstrated, for example, by employing the Debye-Hückel theory of electrolytes to evaluate the correction $\Delta\mu$ to the chemical potential of the electrons resulting from non-ideality of the electron vapor. According to this theory, ((4), Chapter IX) which is asymptotically valid when the density of electrons is sufficiently small,††††

$$\Delta\mu=-(e^2/2)[4\pi e^2 n_1/kT]^{\frac{1}{2}}, \quad (I.3.2)$$

where n_1, the number of electrons per unit volume of the vapor, may be expressed in terms of the equivalent saturation current density j by Eq. (I.3.3) below, assuming for definiteness that the reflection coefficient $\bar{r}$, discussed in connection with step (4) below, is zero. If this is done we find, for example, that even for $j=1$ amp./cm^2, $T=1000°$, $\Delta\mu/kT$ is only -0.004, while for other temperatures and currents this ratio varies as $j^{\frac{1}{2}}T^{-7/4}$.

Step (2) involves merely an integration over the Maxwellian distribution of electron velocities, and gives

$$\text{number of electrons crossing unit area in either direction in unit time}=n_1(kT/2\pi m)^{\frac{1}{2}}. \quad (I.3.3)$$

Assumption (3) may be interpreted as merely a definition of saturation, provided the plane in question is drawn far enough from the surface to be well beyond the maximum in the motive produced by the superposition of the applied field and the image field.

I.3b *Correspondence between Equilibrium and Saturation Conditions, and the Reflection Coefficient*

Step (4) contains the major assumptions of the derivation. In the first place, it involves assuming that the transition from equilibrium to saturation conditions does not change the state of the emitting

††† Historically the question of the legitimacy of treating the electron vapor as a perfect gas has been a rather controversial one, closely related to the question of the legitimacy of treating very small regions of the vapor as thermodynamic systems. The most satisfactory of the early discussions of these subjects is that of Schottky ((S4), partially summarized in (11), #4. Kap. IV, where additional references are also given). The criterion of ideality employed in this work was that the mean potential energy of two nearest neighbor electrons be $\ll kT$. The Debye-Hückel theory, developed after this work of Schottky, makes possible a much more precise treatment, and is therefore used here.

†††† Strictly speaking, (I.3.2) is accurate only if the region occupied by the electrons is electrically neutral, as would be the case if the electronic space charge were neutralized by a uniformly distributed positive charge. Fortunately, this space-charge free condition is just what we wish to assume in step (1).

conductor sufficiently to have an appreciable effect on the rate at which electrons leave it. This assumption will be valid for metals, and usually for semiconductors also, provided we use reasonable prudence in choosing which of the various conceivable equilibrium states of the system is to be put into correspondence with a given set of saturation conditions. For example, if we wish to compute the saturation current when a given strong collecting field E is applied, we should use in step (1) not an equilibrium state in which there is no field at the surface of the conductor, but rather one in which a field of the same strength E exists from the surface out as far as the plane used in step (2); to achieve an equilibrium state in the latter case it will, of course, be necessary to assume that the direction of the field reverses somewhere further out, so that electrons which start ouward do not continue in that direction forever. If the drawing of current from the conductor changes its temperature (because of the cooling effect discussed in Section I.7 below) or the chemical potential of the electrons near its surface (as may be the case for semiconductors), the equilibrium state to which step (1) is applied should be one with the same temperature and chemical potential as obtained under saturation conditions, rather than those which would obtain with the same arrangement of electrodes, heating current, etc. if the emission current were stopped. In extreme cases, e.g., for a semiconductor of low work function but low conductivity, it may happen that even the equilibrium state which best approximates the condition of the conductor when emission current is being drawn approximates it so poorly that it cannot safely be used to compute the rate at which electrons leave the surface.

If in the equilibrium state we follow the subsequent paths of those electrons which cross the plane going inward, some will be found to enter the interior of the conductor and some, say a fraction $\bar{r}_v$ of the total, to return outward across the plane. It is usually assumed that there are no appreciable number of collisions between electrons in the vapor in the region between the plane in question and the conductor. This assumption has the convenient consequence that the reflection coefficient $\bar{r}_v$ is not dependent on the density of the vapor, and is simply a property of the surface and the field near it, which could in principle be measured in a separate experiment. Moreover, some such assumption is necessary in order that the number of electrons leaving the conductor and passing outward across the plane be unaffected by the difference in density of the vapor between equilibrium and saturation conditions.

In order to justify this assumption it is desirable to place the plane as close as possible to the surface, subject to the restriction that it be beyond the range of the image force, in order that assumption (3) be valid. The condition for validity of the present assumption is therefore that the mean free path of an electron in the vapor be large compared with the range of the image force. Now a rough estimation of the mean free path given by Schottky and Rothe ((11), #6 Kap. IV) gives a value of the order of 10^{-3} cm for a saturation current density j of 1 amp./cm^2 and a temperature of 1000°K, and proportional to $T^{4/3}j^{-2/3}$ for other conditions. Since the image potential has fallen to 1 percent of kT at 1000° when the distance from the surface is only 4×10^{-5} cm, we may conclude that our assumption is adequate in all ordinary cases.

I.3c *The Emission Equation and Its Meaning*

The result of the four steps (1) to (4) is obtained by combining the reflection effect discussed above with (I.3.3) and (I.3.1). Expressed in terms of the saturation current density j, it is

$$j=A(1-\bar{r}_v)T^2\exp(\bar{\mu}+eM_v/kT), \quad \text{(I.3.4)}$$

where M_v is the motive on the plane used in the derivation, $\bar{r}_v$ is the average reflection coefficient for a Maxwellian distribution of electrons incident inward across this plane, and where the universal constant A has the value

$$A=(4\pi mk^2e/h^3)=120 \text{ amp./cm}^2\text{/deg.}^2. \quad \text{(I.3.5)}$$

As will be shown in the next subsection, the dependence of (I.3.4) on the distance at which the plane is drawn is only apparent: Changes in this distance will merely change $(1-\bar{r}_v)$ and $\exp(eM_v/kT)$ by reciprocal amounts. If a fairly weak field will suffice to produce saturation, a plane just outside the range where image forces are appreciable will still be at a potential practically the same as that just outside the surface of the conductor, and using the definition (I.2.1) of the true work function, (I.3.4) reduces to the simpler and more familiar form

$$j=A(1-\bar{r})T^2\exp(-e\varphi/kT), \quad \text{(I.3.6)}$$

where $\bar{r}$ is now the reflection coefficient $\bar{r}_v$ of the surface for zero applied field.# The effect of stronger collecting fields will be considered in Section I.4.

The significance of (I.3.6) should not be misunderstood. Since φ and $\bar{r}$ may be functions of the temperature, this equation by itself does not tell us exactly how j should vary with T, although of

Equation (I.3.6) appears superficially to be different from the usual equation derived by statistical mechanics for the thermionic emission current from a semiconductor in which an exponential appears with a coefficient proportional to $T^{5/4}$ and to the square root of the density of impurity centers. It is easily shown, however, that substitution into (I.3.6) of the temperature dependent expression for φ, appropriate to the statistical model being assumed, gives exactly the other formula (see (3), #11.62).

course the fact that φ varies only slightly with T insures that j will increase rapidly as the temperature rises. What (I.3.6) does do is to relate j to $\bar{r}$ and φ, which quantities can, in principle at least, be measured in independent experiments. The exact variation of j with T for the present case of a uniform surface will be discussed in Section I.5, while emission from a patchy or polycrystalline surface will be taken up in Sections II.7 and II.8.

It may seem at first sight a little startling that the emission current which we have derived involves a mean reflection coefficient $\bar{r}$ for electrons impinging on the surface from without, whereas any consideration of the details of the emission process leads one to think in terms of a reflection coefficient for electrons coming out of the metal. The latter type of reflection coefficient does indeed enter in the usual statistical theory of the emission process (see, e.g., (3), Chapter 11); however, it is easily shown from the first principles of wave mechanics that the reflection coefficient for an electron of a given energy crossing a barrier is the same for both directions, so that whenever the assumptions of the usual statistical theory are valid the $\bar{r}$ of that theory and the $\bar{r}$ we have used here are the same. However, if such complicating features as inelastic collisions between the electron and the metal are of importance, it is easier to think in terms of entering electrons as we do here, and because of the universal validity of the principle of detailed balance, the results will always be correct.

I.3d *The Velocity Distribution*

It is worth noting at this point that the principle of detailed balance which we have used in step (4) applies not only to the total electron current, but also to the electrons in any particular range of velocities. Thus the number of electrons emitted from a uniform surface per unit time in an element $d\omega$ of solid angle in any given direction and in any given range dv of velocities must be equal to the number which in equilibrium pass inward with velocities in the diametrically opposite range, times a factor $(1-\gamma)$, where γ is the fraction of those electrons in the range $dvd\omega$ of outward velocities under equilibrium conditions which have come from reflection of electrons incident from outside. One can easily show that in the absence of magnetic fields γ equals the reflection coefficient r of the surface for electrons incident from without with velocities diametrically opposite to those in $dvd\omega$. This equality is obvious when r depends only on the normal component of velocity; a general proof is given in Appendix Va. Since in thermal equilibrium the distribution of velocities is Maxwellian, the fraction dv of the emitted electrons with velocities in the range $dvd\omega$ is

$$dv=(1-r)v_xf(v)dvd\omega\Big/\int\int_{v_x>0}(1-r)v_xf(v)dvd\omega,$$

where $f(v)$ is the Maxwell distribution function and the x direction is normal to the surface. Using the definition of $\bar{r}$ we can write more simply, if f is normalized over the half-space $v_x>0$,

$$dv=(v_x/\bar{v}_x)(1-r)/(1-\bar{r})fdvd\omega, \qquad (I.3.7)$$

where $\bar{v}_x=(2kT/\pi m)^{\frac{1}{2}}$ is the mean v_x of a half-Maxwellian distribution. Thus the deviation of the emitted electrons from the Maxwellian velocity distribution is determined simply by the velocity dependence of the reflection coefficient r, and is negligible if r is small or if r is insensitive to velocity.

I.3e *Alternative Derivations of the Emission Equation*

In concluding this section a few words should be said regarding the relationship between the derivation given here and the more commonly used argument based on the Clausius-Clapeyron vapor pressure equation. Mathematical details of this relationship are given in Appendix II. The derivation most commonly given yields an equation of the form of (I.3.6), with the same A but with φ replaced by a temperature-independent quantity. This result is based on the approximation that a quantity usually called "specific heat of free electrons" in the metal is zero. A more careful derivation, given by Bridgman ((B25) or (2), Chapter IV) shows that this approximation amounts to assuming that the specific heat of an isolated metal block is unchanged by removing electrons from it; if a change in the specific heat does occur, the work function φ contains an integral involving the values of this specific heat change over the temperature range 0 to T. Now if this specific heat change per mole of electrons could be equated simply to the change in the electronic contribution to the molar specific heat of the metal, as measured for example at low temperatures where the vibrational specific heat is negligible, it would be a good approximation to neglect it, at least for most metals (see the discussion of the quantity $\Delta_T\mu$ in Section IV.3e below). However, as will be shown in detail in Section IV.3c, the removal of electrons from a metal block may be expected to change not only the electronic specific heat, but also the forces which bind the atoms together, so that the vibrational specific heat will be changed. Moreover, the electric field produced at the surface of a metal because of the positive charge it acquires on losing electrons will modify the motion and hence the specific heat of the surface atoms. Usually these secondary

effects on the specific heat will be considerably larger than the direct effect on the electronic specific heat. The theories summarized in Section IV.3 indicate that for typical high melting metals at incandescent temperatures the temperature-dependent part of the work function will probably be of the order of kT.

I.4 The Schottky Effect

In this section we shall investigate the form which the emission equation (I.3.4) takes for a uniform surface when the electric field strength E at the surface is not negligible.## For this case the graph of the potential energy, $-eM$, against distance from the surface looks as shown in Fig. 1. It has a maximum at a distance x_0 from the surface, and the height of this maximum is an amount $\Delta\varphi$ below the value $-e\Phi_a$ which the non-image part of the potential assumes immediately outside the surface. If the surface is metallic, the image potential is $-e^2/4x$ and we have

$$x_0=\frac{1}{2}\left(\frac{e}{E}\right)^{\frac{1}{2}}=\frac{1.90\times10^{-4}}{E^{\frac{1}{2}}}\text{ cm}, \qquad (I.4.1)$$

$$\Delta\varphi=(eE)^{\frac{1}{2}}=3.79\times10^{-4}E^{\frac{1}{2}}\text{ volt}, \qquad (I.4.2)$$

if E is in volt/cm. If the surface is that of a semiconductor the image field will of course be weaker, as was explained in Section I.1d; however, if the effective dielectric constant is fairly large the image force will be almost as great as for a metal, and (I.4.1) and (I.4.2) will be roughly correct.

One might suppose that the simplest way to derive the emission for this case would be to apply the considerations of the preceding section to a plane at the distance x_0 from the surface. This can in fact be done if the collecting field E is fairly weak: for this case assumptions (3) and (4) should be valid on a plane at the distance x_0, and so we may use (I.3.4) with M_v set equal to the potential Φ_a just outside the surface plus $\Delta\varphi$, and with a value of $\bar{r}_v$ which should be almost independent of E and equal to the $\bar{r}$ occurring in (I.3.6). If this is done and $\bar{\mu}$ expressed in terms of the true work function by (I.2.1) we get

$$j=A(1-\bar{r})T^2\exp\left[\frac{-e[\varphi-(eE)^{\frac{1}{2}}]}{kT}\right] \qquad (I.4.3)$$

so that a plot of $\log j$ against $E^{\frac{1}{2}}$—the so-called Schottky plot—should give a straight line. The logic of this derivation becomes questionable, however, when E is large, since then the motive will vary so rapidly with x near its maximum that we cannot neglect its variation over the minimum distance within which a thermal electron can be localized (see Appendix I).

A better though slightly lengthier way of applying (I.3.4) to emission with strong collecting fields is to apply it to a plane drawn some distance outside the maximum of the potential energy, and to imagine that the field E decreases gradually from the large value which it has at the surface of the conductor to a very small value in the neighborhood of this plane, increasing again further out so that no appreciable space charge will accumulate. For the usual saturation field strengths the emission under these fictitious conditions will be the same as if the collecting field remained constant at all distances from the surface; however, the introduction of a region of constant motive now makes it legitimate to apply the thermodynamic methods of the preceding section to this region. In applying (I.3.4) it is only necessary to evaluate the fraction $\bar{r}_v$ of a Maxwellian distribution of electrons which start toward the surface from the region just mentioned, which are reflected outward again. Now to a first approximation we may say that this fraction $\bar{r}_v$ consists of all electrons whose initial normal energies $[(mv_x^2/2)-eM_v]$ are less than the potential energy $-eM(x_0)$ at the top of the hump, plus an additional fraction $\bar{r}$ of the remainder, representing electrons which cross the hump but get reflected at the surface of the conductor. Since in a Maxwellian distribution the number of electrons in the element $dv_xdv_ydv_z$ of velocity space is proportional to $\exp[\{-m(v_x^2+v_y^2+v_z^2)\}/2kT]$, the fraction for which $[(mv_x^2/2)-eM_v]>-eM(x_0)=-e(\Phi_a+\Delta\varphi)$ is just $\exp[[-eM_v+eM(x_0)]/kT]$ and so

$$(1-\bar{r}_v)\approx(1-\bar{r})\exp\left[\frac{-eM_v+e\Phi_a+e\Delta\varphi}{kT}\right]. \qquad (I.4.4)$$

Inserting this into (I.3.4) and using (I.2.1) and (I.4.2) gives just the previous result (I.4.3); this illustrates the fact that the value of (I.3.4) is independent of where the plane is drawn to which M_v and $\bar{r}_v$ refer, as long as it is outside the potential maximum. However, it must be emphasized that the expression (I.4.4) for the transmission coefficient $(1-\bar{r}_v)$ is only an approximation, although usually a very good one, and that in interpreting such phenomena as the periodic deviations from the Schottky line to be discussed in Section IV.5 below it is necessary to insert a better approximation into (I.3.4). In this example it is necessary to take account of the fact that, according to wave mechanics, some electrons with normal energies sufficient to pass over the potential hump will be reflected by it, while some of energies too low to pass over it may even "tunnel" through it.

In Section II.5 below we shall see how the simple

For an account of the early history of the image force concept and its effect on electron emission see (11) Chapter XI.

relation (I.4.3) between emission current and collecting field is greatly modified when the emitting surface consists of patches or crystal grains having different work functions. The analysis of the present section does not apply to such cases, because the contact potential differences between the patches produce strong and irregular electrostatic fields near the surface, which may make the dependence of motive on position very different from that shown in Fig. 1.

I.5 The Apparent Work Function and Emission Constant of a Uniform Surface###

The "apparent work function" of an emitting surface is obtained by measuring the emission at various temperatures, extrapolating the emission at each temperature to zero field by an empirically determined Schottky coefficient, and plotting the resulting $\log(j/T^2)$ against $(1/T)$. The apparent work function φ^* is a measure of the slope of this Richardson plot:

$$\varphi^* = -(k/e)[d/d(1/T)]\ln(j/T^2). \quad \text{(I.5.1)}$$

If the emitting surface is uniform, the extrapolated zero-field emission will obey the Richardson-Laue-Dushman Eq. (I.3.6), and the value of (I.5.1) will be

$$\varphi^* = \varphi - T(d\varphi/dT) - [kT^2/e(1-\bar{r})](d\bar{r}/dT). \quad \text{(I.5.2)}$$

For clean metals it is very probable that $\bar{r}$ is small and insensitive to temperature,#### hence that the last term of (I.5.2) is negligible. The term $T(d\varphi/dT)$ is probably small for clean metals also, though not negligible: the rather dubious direct measurements of this quantity suggest that for metals like Ta, W, and Mo, it is of the order of a tenth of a volt (K4, P3), and this is consistent as to order of magnitude with the values which one can surmise from the apparent emission constants of these metals, which involve both $d\varphi/dT$ and the patch structure of the surface (see Eq. (I.5.4) and Sections II.7 and II.8). However, the absolute magnitude of this term may be much larger for semiconductors or for surfaces covered with adsorbed films whose density may change with temperature.

The "apparent emission constant," A^*, is defined similarly in terms of the intercept of the Richardson plot of the zero-field currents:

$$\log A^* = \text{value of } \log(j/T^2) \text{ extrapolated linearly to } 1/T = 0. \quad \text{(I.5.3)}$$

and Convection of Entropy by an Electric Current

In this section we shall consider two related questions. The first concerns the way in which the

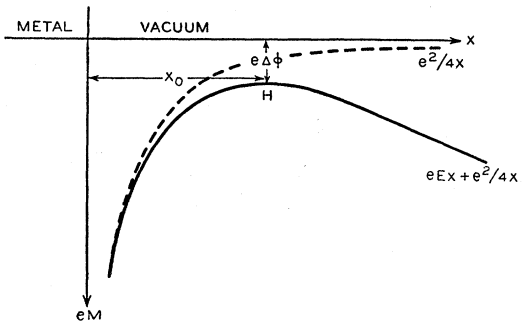


FIG. 1. Plot of motive against distance from a uniform metal surface, in the presence of a normal field E.

For a uniform surface this gives,

$$A^* = A(1-\bar{r})\exp[(e\varphi^* - e\varphi)/kT]$$
$$= A(1-\bar{r})\exp\left[-\frac{e}{k}\frac{d\varphi}{dT} - \frac{T}{(1-\bar{r})}\frac{d\bar{r}}{dT}\right], \quad \text{(I.5.4)}$$

where A is the universal constant defined by (I.3.5). Here again we may expect the term in $d\bar{r}/dT$ to be negligible for clean metal surfaces.

The apparent work function and emission constant of a surface consisting of patches or crystal grains of different work functions will be discussed in Sections II.7 and II.8 below, and related to the emitting properties of the various types of surface present in the patches. There is another type of non-uniformity of a surface, however, which is worth mentioning briefly here, namely small-scale surface roughness. If it is possible to draw a plane at such a distance from the surface that the motive is practically constant over the plane and does not become repulsive inside it by an amount which is appreciable compared with kT/e, and if the reflection coefficient of the surface is small, then the arguments of Section I.3 can be applied to the electron current crossing this plane, and the emission per unit of *projected area* will be given by (I.3.6), independently of the degree of roughness. At the other extreme, if the collecting field is so large that at a distance x_1 from the surface comparable with the scale of the roughness $Ex_1 \gg kT$, then each portion of the surface will emit according to (I.4.3) with the proper local value of E, and the total current will be obtained by integrating this over the total area of the rough surface. In the remaining cases, i.e., if the reflection coefficient is large and $Ex_1 \lesssim kT$, the surface excess correction is more complicated to evaluate.

I.6 Thermoelectric E.M.F.'s and Convection of Entropy by an Electric Current

In this section we shall consider two related questions. The first concerns the way in which the

The first clear exposition of the facts pointed out in this section is due to Becker and Brattain (B9). Because the quantity $\varphi - Td\varphi/dT$ occurring in (I.5.2) is related by (II.1) to the latent heat of vaporization, they suggest that it be called the "heat function."

See the discussion of reflection coefficient in Section IV.4 below.

electrochemical potential $\bar{\mu}$ of the electrons in a conductor varies with position when the conductor is not in equilibrium, i.e., when temperature gradients or electric currents are present. A knowledge of the variation along a temperature gradient is desirable, for example, for the interpretation of experiments which measure the contact potential difference between a hot conductor and a cold one. The second question concerns the flow of heat, or more properly of entropy, in a conductor carrying a current. This will, for example, affect the relationship between the work function and the magnitude of the cooling effect accompanying thermionic emission, which will be discussed in the next section.

I.6a *Variation of $\bar{\mu}$ in the Presence of Electrical and Thermal Gradients*

Consider first the variation of $\bar{\mu}$ with position inside a conductor at uniform temperature, when an electronic current of density j is flowing. If the state of the conductor is uniform except for the potential gradient producing the current, $\mu=\bar{\mu}+e\Phi$ will be constant and Ohm's law will apply so that‡ $\rho j=E=-\nabla\Phi$, where ρ is the resistivity and Φ the electrostatic potential, and where j is the vector of current density in the usual sense, i.e., pointing in the equivalent direction of flow of positive charge. When the state of the conductor is not uniform, so that μ changes from point to point, j will contain a contribution independent of $\nabla\Phi$, which may be interpreted as due to a diffusion of electrons from places of high electron density to places of lower density. To obtain the appropriate generalization of Ohm's law to this case we may write, whenever the departure from equilibrium conditions is infinitesimal,

$$j=-(1/\rho)\nabla\Phi+\alpha\nabla\mu+\beta\nabla T+\sum\gamma_i\nabla Q_i, \quad \text{(I.6.1)}$$

where α, β, and the γ_i, like ρ, are functions of temperature, composition, etc., and where the Q_i are any set of variables independent of Φ and μ whose gradients might conceivably affect the value of j. Now when the temperature is uniform we know from (I.1.4) that whenever $j=0$, $\bar{\mu}$ is constant, no matter how Φ, μ, and the Q's may vary individually. Therefore, we must have all $\gamma_i=0$ and, by (I.1.3), $\alpha=1/\rho e$, so that

$$e\rho j=\nabla\bar{\mu} \quad \text{whenever } T \text{ is constant.} \quad \text{(I.6.2)}$$

Now consider a conductor within which there is a temperature gradient but in which no current is flowing. A very simple argument can be constructed to show that $\bar{\mu}$ must change as we go along the temperature gradient in such a specimen. Consider the thermoelectric circuit shown in Fig. 2, consisting of two different metals a and b and a source of e.m.f., say a battery, at temperature T' in the b part of the circuit. If this e.m.f. is chosen equal and opposite to the e.m.f. E_{ba} of the thermoelectric circuit, no current will flow. In this state the difference between the $\bar{\mu}$'s of the electrons in metal b just to the right and just to the left of the battery will be $-eE_{ba}$. Since $\bar{\mu}_b=\bar{\mu}_a$ at each junction, we may write

$$\int_{T'}^{T_2}\frac{d\bar{\mu}_b}{dT}dT+\int_{T_2}^{T_1}\frac{d\bar{\mu}_a}{dT}dT+\int_{T_1}^{T'}\frac{d\bar{\mu}_b}{dT}dT=-eE_{ba},$$

and differentiating with respect to T_1 we find

$$\frac{d\bar{\mu}_a}{dT_1}-\frac{d\bar{\mu}_b}{dT_1}=-e\frac{dE_{ba}}{dT_1}=e\int_0^{T_1}\left(\frac{\sigma_a}{T}-\frac{\sigma_b}{T}\right)dT, \quad \text{(I.6.3)}$$

where σ_a, σ_b, are the Thomson coefficients of the two metals, which in the usual thermodynamic theory of thermoelectricity are related as indicated to the thermoelectric power.‡‡ The relation (I.6.3) must hold for any pair of metals whatever, and this strongly suggests that in each metal

$$\frac{d\bar{\mu}}{dT}=e\int_0^T\frac{\sigma}{T}dT \text{ whenever } j=0. \quad \text{(I.6.4)}$$

This relation has in fact been shown to be a consequence of certain basic assumptions regarding the reversibility and linearity of thermoelectric effects (H8), and there can be little doubt of its correctness. From (I.6.4) and (I.6.2) one can, if one wishes, obtain the value of β which should be inserted into (I.6.1).

I.6b *The Reversible Thermoelectric Heating Effects*

Let us now consider the way in which the reversible heating and cooling effects in a conductor carrying current may be correlated. It is known that an electric current, besides producing an irreversible Joulean heating proportional to the square of the current, produces a reversible linear Peltier or Thomson heat in a body of inhomogeneous composition or temperature. It has been shown that the latter effects amount to a convection of entropy by the electric current (E2, W1, H8).‡‡‡ We shall not attempt here to establish the legitimacy of this concept of convection of entropy or even to formulate it in its most general terms; these matters are

‡ For simplicity this and the following equations assume the conductor to be electrically isotropic, as is the case for most thermionically important metals and semiconductors. For crystals of lower than cubic symmetry more general forms of the equation must be used (H8).

‡‡ Alternative derivations of (I.6.3) have been given by Bridgman ((2), Chapter IV) and Wagner (W1).

‡‡‡ In crystals of lower than cubic symmetry the vector of entropy flow need not be parallel to the current vector, however.

covered in the references just cited. However, it will be worth while to give a simple derivation of the entropy convection coefficient for a cubic metal or semiconductor, starting from the assumption that a current of density $\mathbf{j}$ carries with it an entropy flux $C\mathbf{j}$. In the presence of a temperature gradient certain amounts of heat must be added or withdrawn at various places in the specimen in order to maintain a steady state in the absence of a current, and certain additional amounts are required to neutralize the Joulean heating, which amounts to ρj^2 per unit volume. Besides all this, an amount of heat q_0, proportional to j, must be added reversibly per unit volume per unit time to maintain a steady state in the presence of a current of density $\mathbf{j}$. This amount is due to the Thomson effect, and by the definition of the Thomson coefficient σ we have

$$q_0 = \sigma \mathbf{j} \cdot \nabla T.$$

On the other hand, our convection assumption requires that

$$q_0 = T\nabla \cdot (C\mathbf{j}) = T\frac{dC}{dT}\mathbf{j} \cdot \nabla T.$$

To make this consistent with the preceding for arbitrary temperature and current distributions, we must have

$$\sigma = T(dC/dT).$$

Now as $T \to 0$ the third law of thermodynamics requires that $C \to 0$ and $\sigma/T \to 0$, so C must have the value

$$\text{entropy convected per unit current} = C = \int_0^T (\sigma/T)dT. \quad \text{(I.6.5)}$$

This is identical with the quantity sometimes called the "absolute thermoelectric power;" notice that the same expression has already been encountered in (I.6.4).‡‡‡‡

The three relations (I.6.2), (I.6.4), and (I.6.5) which we have derived in this section differ in an essential way from the thermodynamic relations derived in previous sections. For the Thomson coefficient σ is not a property of a conductor in thermal equilibrium, but is rather a property associated with the transport of electrons. For example, the electron theory of metals yields a value for σ which involves the rate of change of the mean free path of an electron with its velocity (see (12), p. 179) and mean free path has no essential connection with the energy band structure which determines all equilibrium properties.

I.7 The Cooling Effect Accompanying Thermionic Emission from a Uniform Surface¶

The evaporation of electrons from the surface of a hot conductor, like the evaporation of molecules from a liquid, causes a cooling of the surface, and if a steady state is to be maintained, heat must be supplied to the surface from elsewhere, either by conduction or by other means. Measurements of this effect are almost always made under saturation conditions; let the heat loss per emitted electron for this case be $l_{\text{sat.}}$. To calculate theoretically the relation between $l_{\text{sat.}}$ and other measurable quantities it is advantageous to proceed in two steps. The first is to compute the heat loss per electron, l_{slow}, which would occur for a very gradual quasi-static withdrawal of electrons from the conductor. This calculation may be based on the principle that the entropy fed in per unit time at the surface in a steady-state quasi-static vaporization must equal the difference between the entropy fluxes in the vapor and the metal, or, in more familiar language, must equal $1/T$ times the negative Peltier heat developed at the metal-vapor junction. The second step is to compute the difference $(l_{\text{sat.}} - l_{\text{slow}})$ from energy considerations.

I.7a *The Heat Loss in a Quasi-Static Vaporization of Electrons*

Let us therefore consider a uniform emitting surface forming one boundary of a container whose volume is slowly increased, e.g., by drawing back a piston. Let the emitting conductor be connected to an external circuit so that it does not become charged as more and more electrons are vaporized. Draw a plane parallel to the surface just beyond the range at which image forces are appreciable. The entropy per electron of the electron gas near this plane depends on its density, hence on the work function of the surface. To compute this entropy we may note that in any process the change in the

‡‡‡‡ The use of the third law to obtain an expression like (I.6.5) is at first sight somewhat paradoxical in the case of semiconductors: according to the usual statistical model of a semiconductor the absolute thermoelectric power is essentially $1/T$ times the energy difference between the Fermi level $\bar{\mu}$ and the bottom of the conduction band (or the top of the filled band, in the case of defect conductors); this apparently becomes infinite rather than zero as $T \to 0$. On closer examination, however, one can see that the integration in (I.6.4) and (I.6.5), or equivalently the differentiation with respect to T which occurs in the Thomson relations, must be taken over the sequence of states which occur in an unevenly heated conductor when no current is flowing. It is not possible for all such states to be electrically neutral, and as $T \to 0$ a condition will eventually be reached where the Fermi level lies in the conduction band (or the filled band). From about this point on all the thermoelectric constants will decrease instead of increasing as $T \to 0$.

¶ The first theoretical prediction of the cooling effect was given by Richardson, (R8); references to subsequent treatments of the problem by thermodynamics are given in the article of Schottky and Rothe ((11), Kap. VI). The last term in Eq. (I.7.4) for the cooling effect seems to have been given first by Wagner, (W1), and later independently by Herring (H8).

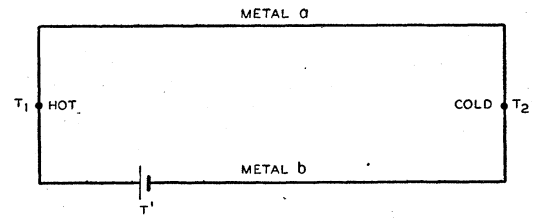


FIG. 2. Thermoelectric circuit.

Helmholtz free energy $F=U-TS$ of the electron gas is

$$dF=-SdT-pdv+\bar{\mu}dn,$$

(see the definition of $\bar{\mu}$, (I.1.1)), hence that the change in the Gibbs function $G=F+pv=U-TS+pv$ is

$$dG=-SdT+vdp+\bar{\mu}dn,$$

so that

$$\bar{\mu}=(\partial G/\partial n)_{T,p}. \qquad (I.7.1)$$

Now the gradual increase of volume to which we are subjecting the vapor phase changes the number, n, of electrons outside the plane but does not change the temperature or pressure. If space charge either is negligible or is compensated by a suitable distribution of positive charge the system of electrons outside the plane will be at constant potential, and treating it as a perfect gas (see Section I.3a), we have

$$\left(\frac{\partial U}{\partial n}\right)_{T,p}=\tfrac{3}{2}kT-e\Phi_v, \quad \left(\frac{\partial(pv)}{\partial n}\right)_{T,p}=kT,$$

so that (I.7.1) gives

$$\bar{\mu}=(5/2)kT-e\Phi_v-T(\partial S/\partial n)_{T,p},$$

where Φ_v is the electrostatic potential of the region of vapor considered. Using the definition (I.2.1) of the true work function, with $\Phi_v=\Phi_a$, we have finally

$$\text{entropy per electron}=(\partial S/\partial n)_{T,p}=e\varphi/T+(5/2)k. \qquad (I.7.2)$$

The quantity (I.7.2) represents the ratio of the entropy passing outward across the plane which we have drawn to the net number of electrons crossing it. Since our quasi-static vaporization is a reversible process, the entropy l_{slow}/T which must be fed in, at, or near the emitting surface to maintain a steady state is equal to the number of electrons vaporized times the difference between (I.7.2) and e times (I.6.5). Therefore,¶¶¶

$$l_{\text{slow}}=e\varphi+\frac{5}{2}kT+eT\int_0^T(\sigma/T)dT. \qquad (I.7.3)$$

I.7b *The Heat Loss under Saturation Conditions*

The difference $(l_{\text{sat.}}-l_{\text{slow}})$ must be equal to the difference, between saturation and quasi-static conditions, in the average kinetic energy per electron crossing a plane parallel to the metal surface. For quasi-static conditions this average energy is $\frac{3}{2}kT$ plus the work per electron done against the piston, or a total of $\frac{5}{2}kT$. For saturation conditions the average kinetic energy will depend on the velocity distribution of the emitted electrons. If the reflection coefficient, r, of the surface is the same for electrons of all velocities in the region of thermal energies, as the theoretical considerations of Section IV.4 suggest it should be, the velocity distribution will be Maxwellian, and the average kinetic energy per electron crossing the plane will be

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty}\exp[-(m/2)(v_x^2+v_y^2+v_z^2)/kT]\cdot(m/2)(v_x^2+v_y^2+v_z^2)\cdot v_x dv_x dv_y dv_z}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty}\exp[-(m/2)(v_x^2+v_y^2+v_z^2)/kT]v_x dv_x dv_y dv_z}=2kT.$$

Thus, if r is constant we have for a uniform surface¶¶

$$l_{\text{sat.}}=e\varphi+2kT+eT\int_0^T(\sigma/T)dT. \qquad (I.7.4)$$

This formula applies to zero-field emission; if a strong collecting field E is being used it is easy to see that φ should be replaced by $[\varphi-(eE)^{\frac{1}{2}}]$.

The magnitude of the last term of (I.7.4) is not reliably known for any metal at incandescent temperatures; however, it is probably quite small, of the order of a few hundredths of a volt or less, and so will only be significant in the most accurate experiments. For semiconductors this term may be orders of magnitude larger than for metals; however, in any experiment where the cathode consists

¶¶ In a derivation whose correctness has been disputed by Nottingham (N10), Fleming and Henderson (F4, F5) have obtained an expression for $l_{\text{sat.}}$ differing from (I.7.4) through replacement of the last term by $1.2kT$. We do not understand their reasoning and cannot agree with their conclusion.

¶¶¶ It must be emphasized that this equation applies to a quasi-static vaporization of electrons only if the space charge is somehow compensated. As Wagener ((6), #9) has pointed out, the heat loss under space-charge limited conditions depends upon the current density only and is independent of the work function of the cathode.

of a thin layer of semiconductor on a metal base, this term will be almost entirely compensated by the Peltier heat $e\Pi$ developed at the junction between metal and semiconductor, since

$$e\Pi = eT\int_0^T \frac{(\sigma_{\text{metal}} - \sigma_{\text{semicond.}})}{T} dT.$$

I.7c *The Surface Heat of Charging*

This section would not be complete without a brief discussion of an apparent paradox which one may encounter in attempting to relate the heat loss accompanying emission to other thermodynamic quantities. Consider again the experiment in which the volume of the electron vapor near a conductor is slowly increased; this time, however, let the conductor be electrically isolated, so that a charge builds up on its surface as electrons leave it. Then we have for the heat loss l_i per electron evaporated

$$l_i = \frac{L}{N_0} = T\left(\frac{\partial S_{\text{gas}}}{\partial n}\right)_{p,T} - T\left(\frac{\partial S_{\text{cond.}}}{\partial n}\right)_{p,T},$$

where L/N_0 is the heat of evaporation per electron and where S_{gas} and $S_{\text{cond.}}$ are the entropies of electron vapor and conductor, respectively. We may substitute (I.7.2) in the first term, while for the second we have

$$\left(\frac{\partial S_{\text{cond.}}}{\partial n}\right)_{p,T} = -\left(\frac{\partial^2 G_{\text{cond.}}}{\partial n \partial T}\right)_p = -\left(\frac{\partial \bar{\mu}}{\partial T}\right)_{p,n}$$

by (I.7.1). Thus, remembering the definition (I.2.1) of φ,

$$l_i = L/N_0 = e\varphi + (5/2)kT - eT(d\varphi/dT), \quad \text{(I.7.5)}$$

which differs from (I.7.3) in the last term. The difference between (I.7.3) and (I.7.5) is clearly due to the fact that in the present case the conductor becomes positively charged as electrons are removed, while in the former case a steady state was assumed involving replacement electrons fed in through the conductor. This difference has accordingly been named the "surface heat of charging" ((2), Chapters III and IV). The two terms in it are readily understandable: because of the absence of replacement electrons, the last term of (I.7.3) should naturally be absent from (I.7.5), and because the electron concentration in the metal is being changed in the one case but not in the other, the term $-T(\partial S_{\text{cond.}}/\partial n)_{p,T}$ occurs in (I.7.5) but not in (I.7.3). It is sometimes useful to divide the latter term into two parts, first, the change in $S_{\text{cond.}}$ which would occur if the electron concentration in the volume of the metal could be changed without giving the metal a net charge (e.g., by adding electrons and at the same time adding an imaginary continuous distribution of positive charge); second, the further change which would occur if the metal could be charged without changing its electron content. The former is a volume effect, the latter an effect of the field at the surface. This division is closely related to the division of $d\varphi/dT$ into volume and surface parts which will be made in Section IV.3 below.

CHAPTER II. ELECTRON EMISSION FROM NON-UNIFORM SURFACES

Up to this point the discussion has been purely theoretical and has dealt with uniform surfaces. The purpose of this chapter is to summarize the experimental evidence for non-uniformity of the properties of the different crystal surfaces of the same metal and to indicate how this non-uniformity must be taken into account in the interpretation of surface phenomena.

II.1 Experimental Evidence for Non-Uniformity

Since the last review article on thermionic emission, a number of important experiments designed to investigate non-uniformity of surface properties have been published. The purpose of this section is to enumerate briefly the results of these important experiments.

II.1a *Experimental Evidence for Non-Uniformity of Thermionic Emission from Clean Metals*

The bulk of this evidence comes from electron projection tube studies. The results of several such experiments have been published (B28, B29, J6, Y1, J4, M10, N1), but for the purposes of this section, the most interesting and informative investigation is the one carried out by Martin (M10). In this work a tungsten sphere one centimeter in diameter made up largely of one single crystal was mounted at the center of a spherical fluorescent screen. This apparatus made possible the study of the thermionic emission in all crystal directions from clean tungsten surfaces and from tungsten surfaces coated with caesium or barium. All emission patterns showed the proper symmetry for the cubic system into which tungsten crystallizes. Thermionic emission from clean tungsten varied smoothly over the surface of the sphere showing maxima centered around the normal emergence of the (001) and (111) crystal directions (Miller indices used) and minima around the (112) and (110) directions. Equivalent emission was observed from all spots (on the same or different crystals) having the same crystal direction normal to their macroscopic surfaces. These results distinctly show that the thermionic emission from the surface of the tungsten sphere is characterized by the crystal-

lographic direction normal to the macroscopic surface.

As the result of certain additives, such as alkali oxides and silica, the "non-sag" type of tungsten wire re-crystallizes into crystals occupying the entire cross section of the wire (13, R12).[a] If the recrystallization schedule is favorable, single crystals several centimeters long can be grown (R12). As the result of mechanical stress during the drawing process, the crystals normally have a face diagonal very closely aligned with the axis of the wire. If the wire is round and smooth, this orientation causes all crystal directions which have Miller indices of the form (*hhk*) to emerge normal to the surface of the wire. As stated above, Martin found that for uncontaminated surfaces all maxima and minima of thermionic emission are centered about directions having this form. Making use of this circumstance, Nichols (N1) has made quantitative measurements of thermionic emission along the important crystal directions for clean tungsten surfaces. Figure 3 is a polar plot of the distribution of emission around a single crystal wire. The maxima and minima occur along the same principal directions as found by Martin. Apparently Martin's heat treatment was not sufficient to produce the relative minimum along the (001) direction as compared to the (116) direction. The empirical thermionic constants given in Table I were determined for each of the maxima and minima and will be discussed in a later section. The measurements showed a variation of thermionic work function of at least 0.3v.

The well-established low voltage deviation from the Schottky line for thermionic emission from polycrystal surfaces of clean metals indicates a variation in thermionic properties of the various surfaces of the crystals. This effect is discussed at length in Section II.5.

II.1b *Experimental Evidence for Non-Uniformity of Field Emission from Clean Metals*

Electron projection tubes have also been used to study the dependence of field emission upon crystallographic direction. The technique of obtaining clean surfaces is difficult in the case of field emission because during observation the temperature must be sufficiently low to rule out thermionic emission. This permits residual gases to be adsorbed more readily and decreases the rate of evaporation of impurities from the surface. However, very careful experiments (J3, M25) have resulted in patterns which are probably characteristic of clean metals such as tungsten, molybdenum, and nickel. These patterns show large variation in emission and have the symmetry of the crystal structure of the metal from which the emitter is made. In the case of tungsten, these patterns are identical in character to the corresponding thermionic emission patterns (M10, N1). Müller (J3) has shown that the non-emitting areas are not merely the result of depressions in an otherwise approximately spherical surface. There is no fundamental change in field emission patterns from clean smooth surfaces over the experimental range of voltage gradients.

II.1c *Experimental Evidence for Non-Uniformity of Photoelectric Emission from Clean Metals*

Because of difficulties of intensity, scattering, etc., few photoelectric studies using projection tubes have been published.[b] However, experiments with large single crystals of zinc, silver, copper, and tungsten have been reported (L10, D3, N3, U1, M13, F2). These large specimens are very difficult to clean up. In several cases, the reported photoelectric work functions for the different surfaces exposed differed to the extent of several tenths of an electron volt. For example, Mendenhall and DeVoe (M13) reported a work function of 4.50 ev for a (112) direction and 4.35 ev for a (310) direction of a single crystal of tungsten. Although the crystal had been heated for many hours at 2200 degrees Kelvin, evidence of surface impurities on the surface normal to the (310) directions still remained.

The Schottky phenomena connected with photoelectric emission also indicate a variation in the photoelectric properties of the various crystal surfaces of the same metal. This is briefly discussed in Section II.12.

II.1d *Experimental Evidence for the Non-Uniformity of the Adsorptive Properties of Metal Surfaces*

Projection tube experiments for thermionic[c] and field emission (J3) show strong preferential adsorption over the various crystal surfaces of the same specimen. For example, in addition to the projection tube study of thermionic emission from the clean tungsten crystal sphere, Martin (M10) also investigated thermionic emission from caesium and barium adsorbed on the tungsten sphere. Relatively simple adsorption patterns were obtained when the surfaces were clean, but exceedingly complex patterns resulted from caesium on

[a] We have been informed by Mr. George Moore, of the Bell Telephone Laboratories, that there is a difference in the properties of prewar "non-sag" type of tungsten wire and the corresponding postwar production in that it is very difficult to produce large crystals in the latter. This may be due to different impurities occurring in the ores or manufacturing processes.

[b] For example, see (P2, M5). Brüche (B27) compares the thermionic electron projection tube with the photoelectric projection tube.

[c] (K2, B28, B29, S1, S2, B8, M2, M3, B30, B31, B32, J6, B33, B34, A1, M14, J4, M10, M16, B10, F8.) See also bibliography given by Blewett (B16, B17).

slightly contaminated surfaces. *However, all the patterns always showed the proper symmetry.* Under different conditions of contamination, different directions having high Miller indices showed properties distinctly different from those of nearby regions. As the temperature was changed, certain regions of good emission moved smoothly into regions of poor emission and vice versa. Similar effects have been reported for field emission (J3).

Both photoelectric and thermionic emission (B8, L11, N7) from composite surfaces show more low voltage deviation from the Schottky line than do clean surfaces. This also indicates preferential adsorption resulting in a variation of emitting properties over the various crystal surfaces exposed.

II.1e *Experimental Evidence for the Non-Uniformity of the Contact Potential*

Several measurements of contact potential between different crystal surfaces of single crystals of metals such as silver, bismuth, and copper have been published (F2, Z1, K5, F1, R14, A6). These experiments require rather large crystals. Consequently it is very difficult to obtain clean surfaces, particularly with the less refractory metals. Contact potentials of the order of tenths of volts have been reported. Contact potential between solid and molten phases of the same metal has also been measured (D5, K6). In some of the measurements the Kelvin method was used and in others the volt-ampere characteristic of photoelectric or thermionic emission with the specimen as one of the electrodes was used. These methods will be discussed in Sections II.4 and II.11.

In the case of silver, (F2), the contact potential difference determined by the Kelvin method agreed well with photoelectric data from the same crystals.

II.1f *Experimental Evidence for the Non-Uniformity of the Surface Structure of Crystals*

If metallic single crystals are heat treated in vacuum or inert gas, surface configurations characteristic of the crystal directions normal to the underlying surface and the type of treatment evolve (S12, J5, S3, S11). If the crystals are heated in a chemically active gas or are etched chemically or electrolytically under proper conditions, different reaction rates for the various crystal surfaces result in growth of certain crystal surfaces at the expense of others (B26, J5, S3, S11, 13). These phenomena indicate that the various kinds of crystal surfaces differ appreciably from one another in such properties as surface energy, chemical activity and perhaps evaporation rate (see Section II.2).

II.2 Atomic Arrangement at a Crystal Surface

The growth of certain crystal faces at the expense of others when a tiny metal crystal is heated, which we have mentioned in Section II.1f, shows that the different crystal faces have different surface tensions, i.e., different values of the free energy per unit area of surface. Similarly the occurrence of preferential etching in the presence of an active gas or solution indicates that reaction rates also depend upon the orientation of the surface relative to the lattice. It would be logical at this point to discuss the atomic interpretation of these phenomena, in an effort to infer what some of the emitting surfaces used in electronic experiments would look like if we were able to see every atom. But since an adequate treatment of this subject requires a lengthy discussion of ideas and experiments in fields otherwise unrelated to thermionics, we have placed the detailed treatment in Appendix III, and shall merely summarize here some of the principal results. Although our present state of knowledge does not suffice to predict with any certainty what a metal surface should look like on an atomic scale, it is likely that, when with the aid of new discoveries a complete understanding of surface arrangements is reached, this understanding will be based in large part on the theoretical and experimental evidence elaborated in Appendix III. We discuss this matter in some detail there because of the importance of the details of surface structure for the intelligent interpretation of experiments. Not only must these details be known for the construction of specific theoretical models of the emission process, but they may have a bearing on such matters as the reproducibility of experimental results.

Two conflicting models of a crystal surface have been proposed.:[d]

(1) The surface of the metal specimen may, in a region where the macroscopic surface has a given orientation, always approximate the ideal crystal plane of the corresponding orientation, i.e., a surface formed from an infinite crystal lattice by removing all atoms whose centers lie on one side of the plane.[e] This would give steps and risers of atomic dimensions only (M10).

(2) The surface may, on the other hand, have a hill-and-valley structure on a scale considerably larger than atomic dimensions, so that no matter what the orientation of the macroscopic surface, the microscopic surface will consist of facets of simple structure (L4, T4).

As is shown in Appendix III, the choice between alternatives such as these usually depends upon the relative rates of various competing processes or tendencies. For example, it is shown in Appendix IIIa that if allowed to approach thermal equilibrium, surfaces of certain orientations will remain smooth while all other surfaces will develop a

[d] Although mechanical polishing may produce a somewhat amorphous surface layer on a specimen at room temperature, this will undoubtedly recrystallize at the temperatures necessary for thermionic emission studies. Thus only crystalline surfaces will be considered here.

[e] Of course, the atomic spacings near the surface might differ slightly from those in the interior.

TABLE I. Thermionic constants for maxima and minima of Fig. 3. From Nichols.*,**

Miller indices	φ**	A**
111	4.39	35
112	4.69	125
116	4.39	53
001	4.56	117
110	4.68	15

* See reference (N1) in the Bibliography.
** Because of a numerical error, the values of φ** in Table I of this reference are 0.03 volt too low. The values in the above table are the corrected ones.

hill-and-valley structure; however, it is shown in Appendix IIIc that it is sometimes possible for one-way processes such as uncompensated evaporation to produce a smoothing effect which works counter to the tendency just mentioned. Experimental evidence bearing on these points is discussed in Appendix IIId. Observations have shown the existence of still another important type of process, not yet understood theoretically, the "d.c. etch." This effect, first pointed out by Johnson (J5), has been studied in some detail by Schmidt (S3). It consists in a step-like structure observed on wires which have been subjected to prolonged heating on direct current and near the support hooks of wires heated on alternating current, where strong longitudinal temperature gradients occur. Figure 4 shows some etch structures of this sort photographed by Schmidt. Since tungsten and tantalum wires heated on a.c. in vacuum develop surfaces which appear smooth (J5, M19), and since the evidence discussed in Appendix III suggests that these surfaces may well be atomically smooth, it seems likely that use of a.c. instead of d.c. for heating will in many cases eliminate a troublesome complication from the interpretation of experimental results.

In the later sections of this article we shall try to take cognizance of the possible inadequacy of hypothesis (1) by using the term "crystal surface" or "patch" for the emitting surface of a specimen whose macroscopic surface has a specified crystallographic orientation, rather than the term "crystal plane." Also, in the later sections the equations are set up in terms of discrete patches. If the surface properties vary smoothly over the surface, the sums over the patches can be replaced by integrals.

II.3 Surface *versus* Volume Effects as Causes of Non-Uniformity of Emission

As discussed in Section I.2, the work function of a clean metal is made up of two parts: the inner work function and the surface double layer of charge. The double layer of charge at the surface is a result of an unsymmetrical charge distribution around the surface ions, the nature of which depends upon the arrangement of the atoms at the surface. The atomic theory of the double layer is discussed in Section IV.2. As shown there, it is to be expected that the moment of the double layer will be different for different crystal faces, and this variation will cause non-uniformity of electron emission properties.

There is another effect, essentially a volume one, which has been suggested as a possible explanation for the non-uniformity of electron emission from clean metals (M23, B12). This effect is caused by the periodicity of the crystal lattice which results in forbidden energy regions occurring for those electron momenta which satisfy the conditions for Bragg reflection along the various crystal directions. If such a region exists at or just above the top of the surface barrier in the case of thermionic emission, or at the top of the Fermi level in the case of field emission, some of the electrons one would normally expect to emerge from the surface will be absent since for the corresponding directions of motion inside the crystal, the energy in question will lie in the forbidden region. The magnitude of the reduction in emission will depend upon the range of energies and directions of motion for which this occurs. This possibility will be discussed in Section IV.4b for the case of thermionic emission.

The observed contact potential between different crystal surfaces of the same clean metal cannot be explained by the volume effect, and must be due to the double layer effect inasmuch as the contact potential is proportional to the difference in the moments of the double layers (see Section I.2). On the basis of the double layers, the striking correspondence of thermionic emission patterns to field emission patterns from single crystals of tungsten would be explained by the similar way in which the work function enters into Richardson's equation and into the Fowler-Nordheim equation for field emission. The Fowler-Nordheim equation, including the correction for the Schottky effect, can be written (G10):

$$j=(1.55\times10^{-6}E^2/\theta^2\varphi)\exp(-6.838\times10^{7}\theta\varphi^{\frac{3}{2}}/E)\ \text{amp./cm}^2, \quad \text{(II.3.1)}$$

where E is the surface field strength in volts per centimeter, φ is the work function in volts, and θ is a slowly varying function of $(E^{\frac{1}{2}}/\varphi)$. The general features of this equation have been verified by experiment (J3) and in particular the $\varphi^{\frac{3}{2}}$ dependence has been verified by Haefer (H2) in which he determined the size of the points by use of the electron microscope. On the basis of the work function differences noted in Table I used in connection with Eq. (II.3.1), Benjamin and Jenkins (B12) were unable to account for their estimated relative values of emission from the various crystal surfaces of tungsten. This hinges on the surface field strengths

at which the observations were made [see Haefer (H2)]. More recently, Müller (M24) in measuring the velocity distribution of field electrons from tungsten has found that on the basis of Eq. (II.3.1) the relatively low field emission from the (110) surface could be accounted for quantitatively by a work function of about 6.0 volts for that surface. As explained in Section II.7b there is also evidence from thermionic data from tungsten single crystals for the existence of a considerably higher work function surface than is indicated in Table I.

In order to account for this correspondence of thermionic emission patterns to field emission patterns from tungsten by the volume effect, it would be necessary for the electron momenta to satisfy the conditions for Bragg reflection both at the top of the surface barrier and also at the Fermi level for the (110), (112), and (001) directions. Now the band structure of tungsten is complex with numerous overlapping bands. Results of Manning and Chodorow (M8) indicate that in tungsten allowed bands exist both at the Fermi level and at the top of the surface barrier. Although the approximations involved in these calculations are many, they at least suggest that in tungsten the volume effect may be of no consequence either for thermionic or for field emission. This question is further discussed in Section IV.4b.

The observed non-uniformity (see Section II.1d) of the emitting properties of layers of foreign material adsorbed on polycrystalline surfaces can be explained only by the surface effect since the adsorption of atomic layers on the surface cannot influence the motions of electrons inside the metal. The principal surface effect is the formation of double layers of charge although it is not unlikely that in some cases the adsorbed layers may alter the reflection effect at the surface. The double layers of charge result from the distortion of the charge distribution in the neighborhood of the adsorbed atoms or ions and may be negative or positive depending upon the nature of the materials. For low concentrations of adsorbed materials, such as barium, thorium, or caesium on tungsten strong non-uniformity of emission is observed (see Section II.1d). It is reasonable to conclude that this is caused principally by a variation of the concentration of the adsorbed materials from one type of crystal surface to another which results from a corresponding variation of binding energy of the adsorbed material by amounts comparable to, or greater than, kT.

II.4 Effect of Patches on the Contact Potential

Consider an electronic conductor whose surface is made up of different types of patches having different double layers and as a result, having different work functions. Assume that there are no externally applied electric fields. From Eq. (I.2.1) the electrostatic potential outside the ith type of patch a distance very small compared to the patch dimensions is given by

$$\Phi_{ai} = -\bar{\mu}/e - \varphi_i, \tag{II.4.1}$$

where φ_i is the work function of the ith type of patch (see Section I.2) and $\bar{\mu}$ is the electrochemical potential inside the conductor. Thus the electrostatic potential just outside the patches is not constant but varies from one type of patch to another in accordance with the differences in the work functions. In the absence of applied fields, the electrostatic potential at a distance from the surface sufficiently large compared to the dimensions of the patches becomes constant (see Appendix IV). This constant potential is given by

$$\bar{\Phi}_a = \sum_i f_i \Phi_{ai} = -\bar{\mu}/e - \bar{\varphi}, \tag{II.4.2}$$

where f_i is the fraction of the surface occupied by the ith type of patch and

$$\bar{\varphi} = \sum_i f_i \varphi_i. \tag{II.4.3}$$

Now consider two conductors A and B, both with patchy surfaces and both at the same temperature connected electrically by a circuit containing no sources of e.m.f. Let the distance of separation of the conductors be sufficiently large compared to the patch size so that the patch fields in the space centrally located between the plates are negligibly small. By the arguments of Section I.2, the equilibrium state will be that in which the electrochemical potentials in the two conductors are equal. Then from Eq. (II.4.2)

$$\bar{\Phi}_{aA} - \bar{\Phi}_{aB} = \bar{\varphi}_B - \bar{\varphi}_A. \tag{II.4.4}$$

Thus a difference of potential exists between the outer neighborhoods of the two patch surfaces

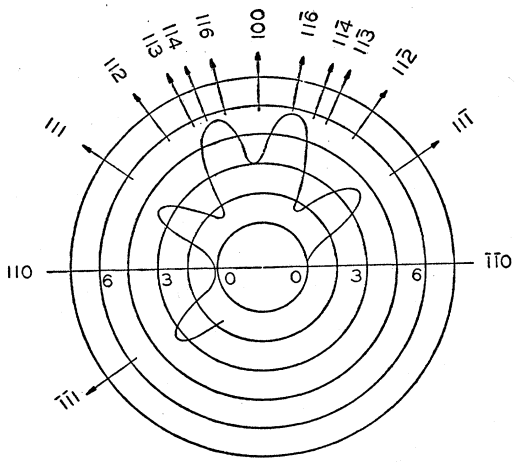


FIG. 3. Polar plot of thermionic emission current from single crystal tungsten wire. Temperature 1480°K.

equal to the difference of their average work functions.

In order to have the outer neighborhoods of the two conductors at equal potential, it is necessary to insert a battery of potential $V_0=\bar{\varphi}_B-\bar{\varphi}_A$ into the connection between the two conductors. If $\bar{\varphi}_A<\bar{\varphi}_B$, then the positive terminal of the battery must be connected to conductor B. If a battery of voltage V different from V_0 is inserted into the connection between the two conductors, there will be a net charge on one of the conductors given by $(V-V_0)/C$ where C is the capacity between the two conductors. A convenient way of experimentally determining V_0 is to adjust V so that a change in the spacing of the conductors—i.e., a change in the capacity C—does not cause a flow of charge. When this is realized, $V=V_0$. This constitutes the well-known Kelvin method for determining the contact potential and yields the value $V_0=\bar{\varphi}_B-\bar{\varphi}_A$. Other methods for determining the contact potential will be discussed in Section II.11. If temperature gradients exist anywhere in the circuit it is necessary to make corrections for thermoelectric effects (see Section I.6).

II.5 Effect of Patches on the Current Voltage Relationship in an Accelerating Collecting Field

The Schottky theory of the current voltage relationship for emission from *uniform surfaces* has been reviewed in Section I.4. Generally, the experimental Schottky plots deviate from the theory of Section I.4 in the low voltage region. As a result of the work of Linford (L11), Becker (B8), Nottingham (N7), and others, it is now understood that this low voltage deviation is the result of patchy emitting surfaces used in the experiments. Because of the importance of the Schottky plot as a means of investigating patchy emitting surfaces, a brief qualitative discussion of this topic is included.

As shown in the previous section, the potential outside a metal surface which has patches of different double layers is not constant even in the absence of applied fields but varies in such a way that the differences between the potentials just outside the different patches are equal to the differences in the work functions of the patches. The resulting field outside the metal is called the patch field. If the patches are small compared to the radius of curvature of the surface, the patch field dies off exponentially so that at a distance from the surface large compared to the patch size, the patch field is very weak (see Appendix IV). Thus the motive outside a patchy surface is equal to that outside a uniform surface defined in Section I.1 minus the term $V_p(xyz)$ representing the potential of the patch field.

II.5a *Qualitative Discussion of the Patch Theory for Metals*

For purposes of discussion, the problem can be divided into three cases.

Case 1. Collecting field large compared to the patch field.—According to the Schottky mirror image theory of emission from a uniform surface, the electrons pass over a potential energy maximum at a distance x_0 from the surface given by Eq. (I.4.1). After passing over this maximum, the electrons are free from the metal. The lowering of the work function, $\Delta\varphi$, as a result of the accelerating collecting field is given by Eq. (I.4.2). In the case of a patchy metal surface, the same theory will apply if the collecting field E is so strong that the absolute value of the work done against the normal component, E_p, of the patch field between $x=0$ and $x=x_0$ is small compared to $e\Delta\varphi$. In equation form, this condition is

$$\left|e\int_0^{x_0}E_p dx\right|\ll e\Delta\varphi. \qquad \text{(II.5.1)}$$

By letting $\bar{E}_p$ be the average value of E_p over the limits on x and using Eqs. (I.4.1) and (I.4.2) this becomes

$$|\bar{E}_p|\ll 2E. \qquad \text{(II.5.2)}$$

If this equation is satisfied, the emission from each patch is independent of the neighboring patches. Now the effective lowering of the work function of each of the individual patches is given by Eq. (I.4.2). Since under these conditions the emission from the entire surface is the sum of the currents from each of the individual patches, it follows from Eq. (I.4.3) that the emission from the entire surface contains the factor $\exp[e(eE)^{\frac{1}{2}}/kT]$. This is further discussed in Section II.7. Therefore the Schottky plot, as defined in Section I.4, will give a straight line of the same slope as that for a uniform surface. Roughly

$$\bar{E}_p\sim\text{differences in work functions of the patches/patch diameter}\sim\delta\varphi/\Delta y, \qquad \text{(II.5.3)}$$

where $\delta\varphi\sim$maximum difference in the work functions and $\Delta y\sim$patch diameter. Approximately, Eq. (II.5.2) becomes

$$E\gg\delta\varphi/\Delta y. \qquad \text{(II.5.4)}$$

Schematically, the situation in the case of a patchy emitter with collecting fields satisfying Eq. (II.5.4) is as indicated in Fig. 5a.

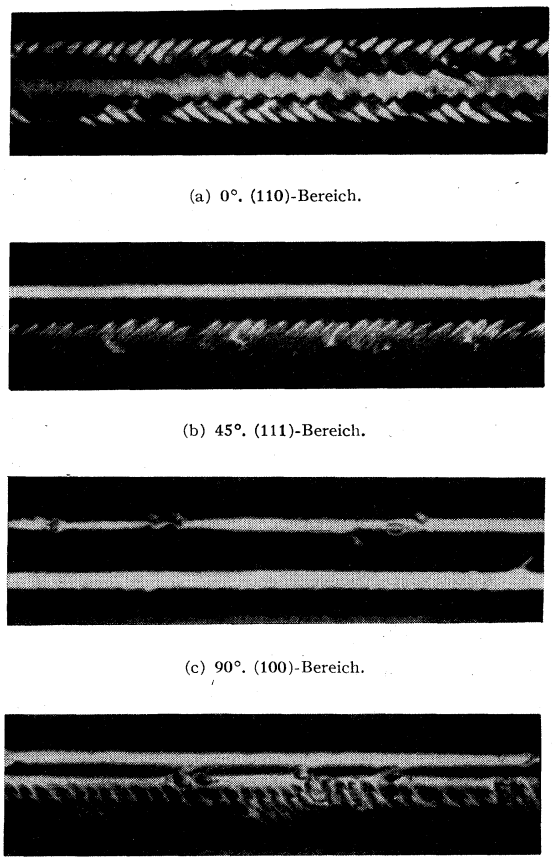
Case 2. Collecting field so small that x_0 lies essentially outside the patch field.—This requires that $E]_{x_0}\gg|E_p]_{x_0}|\sim\delta\varphi/\Delta y\exp(-2\pi x_0/\Delta y)$.[f] This means that the patch field has died off to unim-

[f] The coefficient of $2\pi x_0/\Delta y$ depends upon the shape of the patches; it is of the order of unity. See Appendix IV.

portance so that there is a motive maximum at x_0 as illustrated in Fig. 5b. The reason for the motive maximum at x_0 is that the patch field, because of its exponential nature, dies off faster than the image forces. In general, going out from the surface, the motive slopes "upward" from the lower work function surfaces and "downward" outside the higher work function patches as is indicated schematically in Fig. 5b. Thus the individual electron trajectories in the patch field are complicated and will be discussed, for the case of electron emission into retarding fields, in later sections. For the discussion in this section it is assumed that the effects of initial tangential energies average out so that only electron velocities normal to the surface are considered.

In the absence of applied fields, the motive, as indicated in Fig. 5b, for sufficiently large values of x approaches the value $-\bar{\Phi}_a - e/4x$ where $\bar{\Phi}_a$ is given by Eq. (II.4.2). The effect of sufficiently weak applied fields is to produce a motive maximum outside the patch field whose value is $-\bar{\Phi}_a - \Delta\varphi$ where $\Delta\varphi$ is given by Eq. (I.4.2). Now all electrons from patches whose work functions are less than $\bar{\varphi}$ (see Eq. II.4.3) must pass over a motive maximum no higher than $-\bar{\Phi}_a - \Delta\varphi$ which is at x_0 so that if only normal energies are considered, the effective work function of these patches becomes $\bar{\varphi} - \Delta\varphi$. Electrons from patches whose work functions are higher than $\bar{\varphi}$ must pass over a barrier higher than the one at x_0 so that the emission from these patches is little affected by the applied fields of this case. If the surface is more or less evenly divided between the high and low work function patches, most of the emission comes from the low work function patches. As a first approximation then, one can ascribe to the entire surface an effective work function of $\bar{\varphi} - \Delta\varphi$. (This is further discussed in Section II.8.) To within this approximation the Schottky plot for this case will also be a straight line with the same slope as that from a uniform surface.

Case 3. Collecting field same order of magnitude as patch field.—This is the transition region between cases 1 and 2. In this region the variation of the patch field potential $V_p(xyz)$ is very important. For these collecting fields intermediate between those of cases 1 and 2, the Schottky plot will not be straight, and its average slope will be considerably greater than that for a uniform surface. The order of magnitude of this slope in this "anomalous" region can be estimated simply. As the collecting field E is increased from a value $\ll|\bar{E}_p|$ to a value $\gg|\bar{E}_p|$, the motive hump in Fig. 5 outside the patches of low work function changes from $-\bar{\Phi}_a - \Delta\varphi$ to $-\Phi_{a\ \min} - \Delta\varphi$. $\bar{E}_p$ is given by Eq. (II.5.3). Most of this change will come in a range of values of E which is a few times $\bar{E}_p$ in width. Now in case 1, most of the emission comes from the low work function patches. For a first approximation it is therefore sufficient to ascribe to the entire surface, in Case 1, an effective work function equal to the lowest work function in the surface, provided the surface is reasonably evenly divided between the low and high work function surfaces. (A more accurate expression for the apparent work function of a patchy surface is derived in Section II.7.) Thus if the surface is reasonably evenly divided between the low and high work function surfaces, the approximate difference between the effective work function for the entire surface corresponding to Case 2 and that corresponding to Case 1 is about $\delta\varphi/2$, where $\delta\varphi$ is the difference between the highest and lowest work functions appearing in the surface.



(a) 0°. (110)-Bereich.

(b) 45°. (111)-Bereich.

(c) 90°. (100)-Bereich.

(d) 135°. (111)-Bereich.

FIG. 4. Photographs of a tungsten wire showing the effect of prolonged heating in vacuum by direct current. The wire, a single crystal oriented with a (110) direction along its axis, was photographed from a number of directions around its circumference, corresponding to the normal emergence of the (110), (111), (100), and (111) directions, respectively. Taken from Schmidt (S3).*

* This figure was taken from Vol. 120 (1943) of the Zeits. f. Physik, published by Verlag Julius Springer of Berlin. The use of this figure in this article is by permission of the Attorney General of the United States in the public interest under License No. JA-1316.

To this approximation

$$\bar{\Phi}_a - \Phi_{a\,\min} = \delta\varphi/2. \qquad \text{(II.5.5)}$$

Thus as the collecting field is increased from a value $\ll|\bar{E}_p|$ to a value $\gg|\bar{E}_p|$ the logarithm of the thermionic emission changes roughly by an amount

$$\Delta \ln j = e\delta\varphi/kT. \qquad \text{(II.5.6)}$$

Using Eq. (II.5.3) the slope of the Schottky plot over this range is given roughly by

$$\Delta \ln j/(\text{a few times } |\bar{E}_p|)^{\frac{1}{2}} \sim \text{a few times less than } e(\delta\varphi\Delta y)^{\frac{1}{2}}/kT. \qquad \text{(II.5.7)}$$

For a uniform surface, the slope of the Schottky plot is given by

$$d \ln j/dE^{\frac{1}{2}} = e^{\frac{3}{2}}/kT. \qquad \text{(II.5.8)}$$

The ratio of Eq. (II.5.7) to (II.5.8) is the anomaly factor in the slope of the Schottky plot.

$$\text{Anomaly factor} \sim \text{a few times less than } (\delta\varphi\Delta y/e)^{\frac{1}{2}}. \qquad \text{(II.5.9)}$$

Thus if $\delta\varphi$ is several tenths of a volt, or perhaps one or two volts in case of adsorbed materials, and if Δy is of the order of 10^{-4} cm, the expression (II.5.9) may be of the order of tens.

Figure 6 shows the schematic Schottky plot which is expected from these considerations for a surface made up of fine patches. The voltage range in which the transition takes place depends upon the patch size, work function differences and the configuration of the emitter and collector. The general characteristic of Fig. 6 has been found for fine grain specimens such as thoriated tungsten (see Fig. 7).[g]

II.5b *Quantitative Calculations of the Effect of Patches on the Schottky Plot*

Compton and Langmuir (C7), Linford (L11), Becker (B8), Nottingham (N7), and others have investigated the effect of patches on the Schottky plot. Compton and Langmuir rejected the patch theory, as an explanation of experimental results, because they felt that the required patch size was too large. By the time of Becker's review on thermionic emission, electron projection tube experiments had indicated patch sizes adequate to explain the experimental results. As mentioned in Section II.1, subsequent experiments on the non-uniformity of electron emission have substantiated the patch theory for clean surfaces as well as for adsorbed surfaces.

To calculate quantitatively the current voltage relation for a patchy surface with applied accelerating fields, Becker and the others start with a plane surface (yz plane) made up of a regular array of patches. The motive, consisting of the mirror image potential, the applied potential, and the patch potential is determined. For each value of the applied field, the motive maximum outside the surface occurs at a distance x_0 which is a function of the corresponding y and z values. The value of the motive at this maximum is then determined as a function of y and z for the various values of the applied field. This defines a motive "sky-line" for each value of the applied field which the electrons must pass over. In order to simplify the computation of the electron current, it is assumed that the current crossing each element of projected area, $dydz$, of the skyline is proportional to $\exp[-e\varphi'(y,z)/kT]$ where $\varphi'(y,z)$ is the height of the sky-line relative to the electrochemical potential. This amounts to assuming that the electron distribution emerging from each patch is Maxwellian and that the y and z momenta average out so the problem reduces to one of normal kinetic energy only. The latter assumption is questionable, for reasons discussed below in Section II.6b, particularly at low applied fields. The other important assumptions are that the applied electric field does not change the thermodynamic state of the interior of the surface or the double layer at the surface, and that space-charge effects are negligible. These assumptions are reasonable at least for metals at sufficiently low current densities. The theoretical results, when the patch size and patch double-layer differences are properly chosen, show reasonably good agreement with experiment. For example, Fig. 7 is taken from Nottingham (N7) who used a strip distribution of patches whose work function varied sinusoidally. This variation was assumed for mathematical simplicity instead of the more or less discontinuous variation of work functions from patch to patch which one expects a polycrystalline surface to show. The period of variation is z_0 and the amplitude of variation is a. The values of a and z_0 were adjusted to fit the experimental data at 1160°K which are from thoriated tungsten whose state of activation is Z. Using these constants, a reasonably good fit for the 929°K data was obtained. These results are similar to those obtained by Becker and others.

In connection with Fig. 7, it is to be noted that the theoretical curves approach the asymptotic Schottky line more rapidly than do the experimental curves. This is also characteristic of the results of Becker (B8) who worked with a checkerboard rather than a strip array of patches. This is probably due to the fact that a regular arrangement of patches with sinusoidally varying work functions was used for the calculation whereas in the experi-

[g] We have been informed by Dr. LeRoy Apker, of the General Electric Research Laboratory, that he has obtained photoelectric data which definitely show a straightening out on the low-voltage side of the anomalous region.

ment, the patches are of various sizes and shapes and the work functions vary more or less discontinuously. The neglect of the tangential components of the electron momenta in the calculation of the current over the skyline barrier is probably most important in the low voltage range (see Section II.6b).

II.5c *Patch Theory for Semiconductors*

The considerations just outlined, which determine the effect of patches on the emission from metals, become somewhat more complicated when the emitter is a semiconductor. For in this case it is frequently not permissible to assume that the electrochemical potential $\bar{\mu}$ of the electrons inside the conductor is the same under all points of the emitting surface, or that the chemical potential μ is unaffected by the applied field. Lack of constancy of $\bar{\mu}$ may occur when the resistivity of the semiconductor is appreciable. If two patches of the surface have different work functions and therefore emit different current densities, the ohmic potential drops between the core metal and the two surface patches will be different. Moreover, even a surface of uniform work function can manifest patch effects if the core from which the current is supplied to the semiconductor is a wire mesh, or otherwise nonuniform. As for the variation of μ with applied field, this may result from the fact that application of a field—or in fact merely the patch fields normally present at the surface—necessitates an accumulation of charge in the surface layers of the semiconductor. Since the density of free electrons originally present is small, a moderately strong field may produce an appreciable modification in the density of free electrons, hence in μ. These phenomena are closely related to those treated in the theory of rectifying contacts (S6).

II.5d *The Schottky Plot as a Tool for Investigating Surfaces*

It is clear that the shape of the Schottky curve gives information as to the patch nature of the emitting surface. Thus by investigating the dependence of emission on collecting field strength one can derive information regarding the sizes and other properties of the patches of which a nonuniform surface is composed. If the patches are large enough to be seen with a visual or electron microscope using low power, it will not require a very great field to bring the emission into the region where the Schottky plot has the normal slope, and it may be hard to study the anomalous region. But if the patches have dimensions smaller than say a micron or so, the Schottky plot can give information which may be difficult to obtain in other ways.

II.6 Effect of Patches on the Current Voltage Relationship in a Retarding Field

As was reviewed in Section I.3d, the deviation of thermally emitted electrons from the Maxwellian velocity distribution is determined by the dependence of the reflection coefficient, r, upon the velocity and is negligible if r is small or if r is only slightly dependent on the velocity. Theoretical considerations indicate that in many cases clean and uniform metal surfaces should have a small reflection coefficient and that metals such as tungsten and tantalum are probably of this type. This is confirmed by the small magnitude of the periodic Schottky effect for these metals (see Section IV.4d). In Appendix V it is shown that in the case of a patchy surface, the effect of the different emitting properties of the patches can be simulated for retarding fields by a uniform surface of work function $\bar{\varphi}$ of Eq. (II.4.3) with a velocity dependent reflection coefficient henceforth called the *associated reflection coefficient.* Depending upon the nature of the patches, the associated reflection coefficient may be quite large even though the reflection coefficient of the surfaces of the individual patches may be small.

Germer, (G2), Nottingham (N7), and others have investigated the velocity distribution of electrons thermally emitted from polycrystalline surfaces by means of retarding potentials. Nottingham's experiments indicate that for energies greater than about 0.5 ev, the emitted electrons follow the Maxwell distribution, but for energies less than this, there are fewer electrons than would be expected on the basis of a Maxwell distribution from a uniform surface. Gimpel and Richardson (G3), and Davisson, Ahearn, and Teal (D1)[h] have investigated the reflection of slow electron beams from polycrystalline surfaces. Although the experimental difficulties in this type of experiment are very great, the results indicate an appreciable reflection of electrons whose energies are of the order of 0.1 ev or less. Since theoretical and other experimental evidence as mentioned above, points to only a small reflection effect for uniform surfaces, such as tungsten and tantalum and probably others, it is reasonable to look for the explanation of these results in the patch effect at the electrodes.

In the published work on the calculation of the effect of patches, to be referred to later, it was assumed that the effect of the initial energies, tangential to the emitting surface, average out so that the problem reduces to that of normal energies alone. The results using this normal energy approximation are not in agreement with the interpretation of the experimental results. However, before

[h] We are indebted to Drs. Davisson, Ahearn, and Teal for discussion of their work prior to its publication.

these experimental results are interpreted as a quantitative measure of the reflection of electrons at the metallic surface, the effect of patches should be determined in a more rigorous manner for it is certain that the presence of patches influences the observed electron energy distribution. Inasmuch as reflection plays an important role in the interpretation of emission phenomena and inasmuch as a rigorous patch theory for retarding fields has not been published, the authors feel that a brief and more or less qualitative discussion of patch effects should be included in this paper as follows.

It is convenient to divide the discussion into the effects of patches on the contact potential, the effects on electrons at the emitter, and the effects on electrons at the collector. The effect of patches on the contact potential has been discussed in Section II.4 where it was pointed out that the electric field in the neighborhood of a surface having patches of different double layers is never everywhere zero. Figure 8 represents the potential between two plane parallel patchy electrodes illustrating in a schematic way the potential variation outside the maximum and minimum work function patches. In this figure it has been assumed that the electrodes are sufficiently separated so that the patch field in the space centrally located between the electrodes is essentially zero. Now zero field in this centrally located space occurs when the applied voltage equals $\bar{\varphi}_B - \bar{\varphi}_A = V_0$, where $\bar{\varphi}_B$ is the surface average work function of the emitter defined by Eq. (II.4.3), and $\bar{\varphi}_A$ is the corresponding average for the collector. The discussion of the other effects follows in the next three sub-sections.

II.6a *The Normal Energy Approximation for a Patchy Emitter and Uniform Collector with Retarding Collecting Fields*

Consider first the effect of a patchy emitter and uniform collector. If it is assumed that the effect of initial tangential energies average out, so that the problem reduces to that of normal kinetic energies only, the determination of the reflection coefficient associated with the patchy emitter is simple and straightforward. In order to arrive at an associated reflection coefficient in this case, the problem is to determine what sort of reflection coefficient a uniform surface of work function $\bar{\varphi}_B$ must have so as to behave in the same way as the patchy surface it is to represent in the region of retarding collecting fields. Now on this normal energy picture, all emitter patches whose work functions, φ_{Bi}, are such that $\varphi_{Bi} > \bar{\varphi}_B$ are areas of complete reflection for all electrons whose normal energy is less than $e\varphi_{Bi}$. For example, a checkerboard emitter with only two types of patches of work function φ_{B1} and φ_{B2} with $\varphi_{B2} > \varphi_{B1}$ and with these patches occupying equal areas will have an associated reflection coefficient of 0.5 up to electron normal energy of $e\varphi_{B2} = e(\bar{\varphi}_B + \delta\varphi/2)$ where $\delta\varphi = \varphi_{B2} - \varphi_{B1}$. For normal energies greater than $e\varphi_{B2}$, the associated reflection coefficient is zero. In terms of normal kinetic energy ϵ after emission from a uniform surface of work function $\bar{\varphi}_B$ the associated reflection coefficient for the two patch checkerboard is given by Fig. 9. If, instead, a strip arrangement of patches in the yz plane with a sinusoidally varying work function given by $\bar{\varphi}_B + a\cos 2\pi z/z_0$ is used, the reflection coefficient associated with this surface is given by Fig. 10. This curve was obtained by calculating the fraction of the area having work function greater than $e\bar{\varphi}_B + \epsilon$. This is essentially the same result obtained by Nottingham (N7). In order to give a comparison with experimental results, the solid line curve of Fig. 11 is taken from Nottingham (N7) and represents the transmission coefficient which he deduced from his experimental data on polycrystalline thoriated tungsten (by methods to be discussed in Chapter III). The dash-dot and dotted curves are transmission coefficients taken from Fig. 10 for $a = 0.17$ and $a = 0.46$, respectively, values which Nottingham found gave agreement with the characteristic for accelerating fields (see Section II.5b). The agreement between theory and experiment is not very satisfactory particularly in the low energy range, as might be expected partly because the normal energy approximation has not been justified, particularly in the low energy range, and also because the experimental data in the low energy range are difficult to evaluate (see Chapter III). Also, the assumptions as regards the patch nature of the surfaces represented in Figs. 9 and 10 were very crude and in general do not represent accurately an actual surface (see Section II.1 and references thereto). For example, if it is assumed that the exposure of the various crystal surfaces in the case of the polycrystalline wire used is more or less the same as in the case of a sphere carved from a single crystal of the metal, then results such as those of Martin (M10) indicate that the work function distribution over the surface is not sinusoidal for the sinusoid puts too much weight on the low and also perhaps too much on the high work function surfaces. The inadequacy of the sinusoidal varying distribution to represent the symmetry of cubic crystals has been pointed out by Recknagel (R1). Of course, in the case of a polycrystalline emitter with patches of various shapes and sizes and arrangements, the use of a regular arrangement for calculation purposes may yield results considerably different from the actual behavior of the emitter even if the calculation is rigorous.

Having the reflection coefficient associated with the patch field, it is easy to calculate, for example,

the retarding potential characteristic for thermionic emission in the usual way in which the transmission coefficient associated with the patch effect is included along with the surface transmission coefficient. Figure 18 which illustrates a method for determining contact potential from the volt ampere characteristic makes use of a retarding potential characteristic for a two patch checkerboard plane emitter calculated by use of Fig. 9 under the assumption that the velocity distribution of electrons emitted from the patches is Maxwellian.

II.6b *The Effect of Tangential Energies in the Patch Field of the Emitter*

The assumption that the effect of initial tangential energies of the emitted electrons averages out so that only normal kinetic energy need be considered is questionable. As mentioned in Section II.5a, the patch potential for an electron slopes "upward" outside a low work function patch and "downward" outside a high work function patch. Thus the trajectory of the electrons is in general a complicated pattern of excursions up and down the sides of these potential "channels" with frequent interchanges of kinetic and potential energy. A rigorous treatment of the problem has not been published.

As indicated in Appendix V, considerable insight into the problem can be gained by considering the equivalent case of electrons *incident* on a patchy surface. From purely geometrical reasoning, it is shown there that for a two-dimensional array of patches the associated reflection coefficient, averaged over the tangential momenta of a Maxwellian distribution, must approach zero with zero slope as indicated in Fig. 33. This feature is more in line with the experimental curve in Fig. 11 than the normal energy approximation results. As indicated in Sections II.3 and II.7b, there is evidence for work function differences between crystal surfaces of clean tungsten perhaps as large as one volt. In the case of adsorbed surfaces the work function differences may be greater. Accordingly one would expect an associated reflection coefficient different from zero for normal kinetic energies equal to and less than the same order of magnitude as these work function differences.

II.6c *The Effect of Patches on the Collector*

In the previous discussion of thermionic emission in a retarding field, the collector surface was considered uniform. Consider now the effect of a patchy collector in which the patch size is small compared to the size of the anode and the anode-cathode spacing. In the first place, the average work function $\bar{\varphi}_A$ must be used in the expression for the contact potential. Secondly, the effect of reflection in the collector patch field must be taken into account. For purposes of discussion, consider a collector whose surface is a checkerboard of two kinds of patches of work function φ_{A1} and φ_{A2} with $\varphi_{A2} > \varphi_{A1}$. If only normal energies are considered, electrons in the normal energy range $e(V+\bar{\varphi}_A)$ to $e(V+\varphi_{A2})$ relative to the electrochemical potential inside the emitter are unable to reach the φ_{A2} patches and are reflected back toward the emitter. If plane parallel electrodes are used, these reflected electrons return to the emitter. Then if only normal energies are considered, and if it is assumed that none of the electrons reflected back to the emitter subsequently reach the collector, the collector current can be computed by adding the currents collected by the two types of collector patches. In this case the current collected by the φ_{A2} patches, which occupy half of the collector area, is obtained from the characteristic with uniform collector surface multiplied by one half with $V_0 = \bar{\varphi}_B - \varphi_{A2}$. The current collected by the φ_{A1} patches is obtained in like manner but with $V_0 = \bar{\varphi}_B - \bar{\varphi}_A$. Thus the result is the sum of two characteristics with one shifted a distance $\delta\varphi_A/2$ along the $V - V_0$ axis where $\delta\varphi_A = \varphi_{A2} - \varphi_{A1}$.

On the other hand, if cylindrical or spherical electrodes with emitter diameter small compared to collector diameter are used, the electrons in the energy range $eV + e\bar{\varphi}_A$ to $eV + e\varphi_{A2}$ (relative to the electrochemical potential inside the emitter) which are reflected from the collector have a very good chance of missing the emitter and approaching the collector at a different place. If the re-approach is made in the vicinity of a low work function patch, these electrons can be collected. This phenomenon thereby greatly increases the probability of electrons in this energy range being collected. This effect has been pointed out by Hill (H12). Thus in the usual experimental case in which a cylindrical collector and a fine wire emitter are used, the neglect of reflection in the patch field of the collector is probably justified.

II.7 The Effect of Patches on the Strong Field Richardson Plot

II.7a *General Theory*

Consider a surface which is made up of patches of different emitting properties. Let a surface field E be applied which is much stronger than the patch field or more precisely, of sufficient strength to satisfy the requirement of Case 1 of Section II.5a. Then each patch emits independently of the nature of its neighboring patches. From Eq. (I.4.3) the emission per unit area from the ith patch is given by

$$j_i = A(1-\bar{r}_i)T^2 \exp[-e(\varphi_i - (eE)^{\frac{1}{2}})/kT], \quad \text{(II.7.1)}$$

where φ_i is the zero field work function of the ith patch and $\bar{r}_i$ is the average reflection coefficient for the ith patch. The total current per unit area is given by

$$j=\sum_i f_i j_i = AT^2\sum_i f_i(1-\bar{r}_i)\exp(-e\varphi_i/kT) \times\exp[e(eE)^{\frac{1}{2}}/kT], \quad \text{(II.7.2)}$$

where f_i is the fraction of the surface occupied by the ith type of patch. It will be assumed that f_i is independent of temperature as is generally the case for clean metals but not always for composite surfaces. Define the zero field current j_0 by

$$j_0=AT^2\sum_i f_i(1-\bar{r}_i)\exp(-e\varphi_i/kT). \quad \text{(II.7.3)}$$

This definition of j_0 amounts to an extrapolation of the Schottky plot for Case 1, Section II.5a ,to zero field. In analogy with Eqs. (I.5.1) and (I.5.2) the corresponding zero-field apparent work function, φ^{**}, for this surface is given by

$$\varphi^{**}=-(k/e)d\ln(j_0/T^2)/d(1/T)=\sum_i w_i\varphi_i^*, \quad \text{(II.7.4)}$$

where w_i is the fraction of the total zero field emission current from the ith type of patch, and φ_i^* is the zero field apparent work function of the ith type of patch. Or

$$\delta\varphi^{**}=\sum_i w_i\delta\varphi_i^*,$$

where

$$\delta\varphi_i^*=\varphi_i^*-\varphi^*_{\min.}, \quad \text{(II.7.5)}$$
$$\delta\varphi^{**}=\varphi^{**}-\varphi^*_{\min.}.$$

Here $\varphi^*_{\min.}$ is the lowest apparent work function. From Eq. (I.5.3) the corresponding A^{**} is given by

$$\ln A^{**}=\ln(j_0/T^2)+e\varphi^{**}/kT=\ln\sum_i f_i A_i^* \times\exp(-e\delta\varphi_i^*/kT)+(e/kT)\sum_i w_i\delta\varphi_i^*. \quad \text{(II.7.6)}$$

These equations in essentially this form have been published by Recknagel (R1). Similar reasoning has also been applied by Heinze and Wagener (H5), King (K1), and Gysae and Wagener (G13).

From Eq. (II.7.4) it can be seen that $\varphi^{**}\geqslant\varphi^*_{\min.}$. The relationship giving the upper limit of φ^{**} is complicated but the upper limit is certainly less than the $\bar{\varphi}-Td\bar{\varphi}/dT$ of Eq. (II.4.3) because in Eq. (II.7.4) the low work functions are more heavily weighted than the high work functions. Also, from Eq. (II.7.6) it follows that $A^{**}\geqslant f_{\min.}A^*_{\min.}$ where $f_{\min.}$ and $A^*_{\min.}$ correspond to the lowest work function type of patch. If all A_i^* are equal, $A^{**}\leqslant A_i^*$ as can be seen from Fig. 13. Also in Eq. (II.7.4) the w_i are temperature dependent. At very low temperatures $\varphi^{**}\to\varphi^*_{\min.}$ from above and at very high temperatures $\varphi^{**}\to\sum_i A_i^* f_i\varphi_i^*/\sum_i A_i^* f_i$ from below. This causes the Richardson plot from a non-uniform surface to be concave upward and is discussed in Section II.7c. It can be shown in an identical manner that Eqs. (II.7.4), (II.7.5), and (II.7.6) are also valid for a surface made up of subgroups of patches, in which case w_i is the fraction of the total emission from the ith subgroup and φ_i^* is replaced by φ_i^{**}, the zero field apparent work function for the subgroup which is determined by Eq. (II.7.4) summed over all patches of the subgroup.

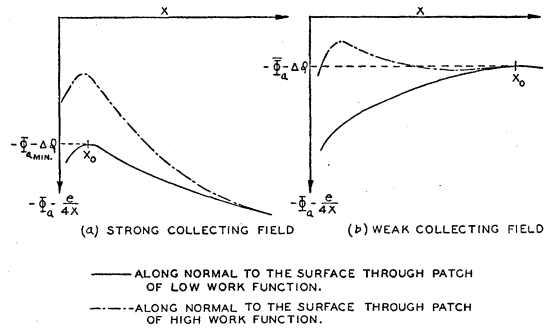


FIG. 5. Schottky effect for a non-uniform surface. The motive (see Section I.1d) is plotted against the distance from the surface.

II.7b *Strong Field Thermionic Emission Constants from Clean Tungsten Single Crystals*

As pointed out in Section II.1a, one quantitative projection tube experiment (N1) with a single crystal of clean tungsten has resulted in the data of Table I giving the thermionic constants for the directions of maxima and minima of Fig. 3. Perhaps the most striking thing about this table is the wide variation in the A^{**} values reported. In the case of the (110) direction, the Schottky plot indicated a patchy surface inasmuch as it deviated from the Schottky line at a surface gradient of about 5×10^4 volts per cm ($x_0=90\times10^{-8}$ cm). Microscopic examination showed a shingle-like structure normal to the (110) direction (see Section II.2). The Schottky plots of the other important crystal directions did not deviate from the Schottky line down to surface gradients of about 3.0×10^4 volts/cm ($x_0=110\times10^{-8}$ cm), below which the projection properties of the tube were not satisfactory. However, microscopic examination showed a d.c. etch (see Section II.2) which extended from the shingle-like region on the (110) surface to the (111) surface. The other surfaces were reasonably smooth although some smooth hills and valleys and some small surface crystals existed (see Fig. 7, reference N1).

It is therefore not certain that the emission constants in Table I are characteristic of uniform surfaces; in fact, it is quite certain that those for the (110) are influenced by patches. For this reason, the thermionic constants in Table I are labeled with double stars. Consider the data for the (110) surface. In the first place, the φ^{**} for this patch surface is essentially the same as that for the (112)

surface. Therefore, it appears reasonable to assume that the (110) surface is made up of a small fraction of (112) surface plus one or more other patches of considerably higher work function. In the second place, it seems unlikely that the A^* value for a uniform (110) face could be as small as 15 amp./cm^2 deg.^{2} when measured A^{**} values for other surfaces run as high as 120. Differences in the A_i^* of different crystal faces can be due only to differences in the reflection coefficients or in the temperature derivatives of the double layer moments and theoretical considerations (see Sections IV.3d and IV.4) suggest that neither of these effects is likely to be large.

Suppose, therefore, that the (110) surface is made up of two different types of patches, one being the (112) surface of $\varphi_1^{**}=4.68$ and $A_1^{**}=120$ and the other having $\varphi_2^{**}=4.68+\delta\varphi$ and $A_2^{**}=120$. Figure 12 is a rough plot of the resulting A^{**} and $\delta\varphi^{**}$ for a two patch surface *versus* the fraction, f, of the surface taken up by the low work function patches for various values of $\delta\varphi$ assuming that both types of patches have a constant $A^*=120$. It is important to notice that a $\delta\varphi$ of at least about 0.7 ev is required to get an A^{**} value of 15 as given in Table I. For $\delta\varphi=0.7$, a value of $f=0.05$ is required to give an A^{**} of 15. The corresponding value of φ^{**} is 4.71 which is a little high so that a $\delta\varphi$ of around 1.0 ev would be better, in which case f is about 0.12 and φ^{**} about 4.68. Thus it is reasonable to assume that in addition to the three essentially different work functions appearing in Table I, there is at least another which is perhaps about 5.35 ev or greater. As pointed out in Section II.3 there is also evidence from field emission data for a work function of about 6.0 ev for the 110 surface.

II.7c *The Effect of Patches on the Strong Field Thermionic Emission Constants from Polycrystalline Tungsten*

The purpose of this section is to describe a convenient method of estimating the effect of patches on the strong field thermionic emission constants.

In order to estimate the effect of patches it is necessary to have information as to the φ_i^* and A_i^* values of the individual types of patches and the relative area occupied by each type of patch. In many cases it is desirable to include numerous types of patches of different φ_i^* and A_i^* in the estimation . It is therefore convenient to make use of the fact that Eqs. (II.7.4) and (II.7.5) also apply to subgroups of patches as follows. First a basic set of a few different representative types of patches is chosen and curves of φ^{**} and A^{**} calculated for different f_i values. Then by means of the curves, combinations of these basic types of patches are chosen to represent the different types of patches to be included in the estimation. This results in a set of f_i values for each type of patch to be included and the sums of the f_i values are then used again with the curves to estimate the average constants for the entire surface.

In order to illustrate this method for tungsten, a basic set of four types of patches was chosen. As a result of the discussion in Section II.7b four types of patches of work functions 4.35, 4.56, 4.68, and 5.35 were used. This gives four $\delta\varphi_i^*$ values of 0.00, 0.21, 0.33, and 1.00. For convenience, and inasmuch as the (112) and (001) surfaces in Table I show A^{**} values of about 120, the A^* values of the four basic patches were all more or less arbitrarily taken to be 120.

Figure 13 gives the $\delta\varphi^{**}$ and A^{**} values calculated for surfaces made up of various amounts of the four patches. The quantities f_1, f_2, and f_3 refer, respectively, to the fraction of the total surface occupied by the 4.35-, 4.56-, and 4.69-ev type of patches. The quantity f_4, the fraction of the surface occupied by the 5.35-ev type of patch is given by $1-f_1-f_2-f_3$.

The over-all $\delta\varphi^{**}$ from the patchy surface was computed from the slope of a straight Richardson line passed through a point computed for 1000°K and a point computed for 2000°K. The corresponding A^{**} was computed from the value of $\delta\varphi^{**}$ and the point at 2000°K.

Although Eq. (II.7.4) indicates that the φ^{**} is temperature dependent, Richardson plots from polycrystalline tungsten are straight lines, to within experimental error, over the experimental temperature range. The experimental temperature range is sufficiently limited so that the calculated Richardson plots from the four patch surfaces of Fig. 13 are as straight as the experimental ones as long as the fraction of the surface made up of low work function patches is not too small. Figure 14

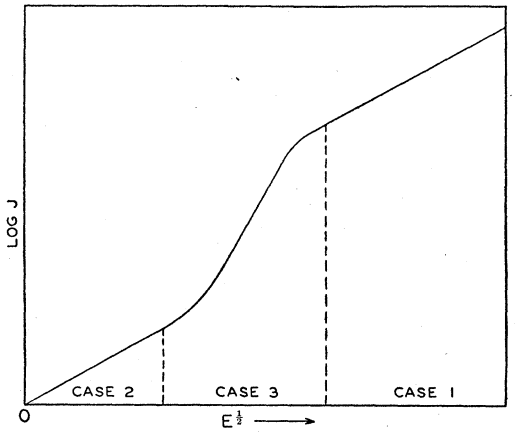


FIG. 6. Schematic Schottky plot giving the logarithm of the current expected from a patchy emitter as a function of the square root of the applied field at the surface of the emitter.

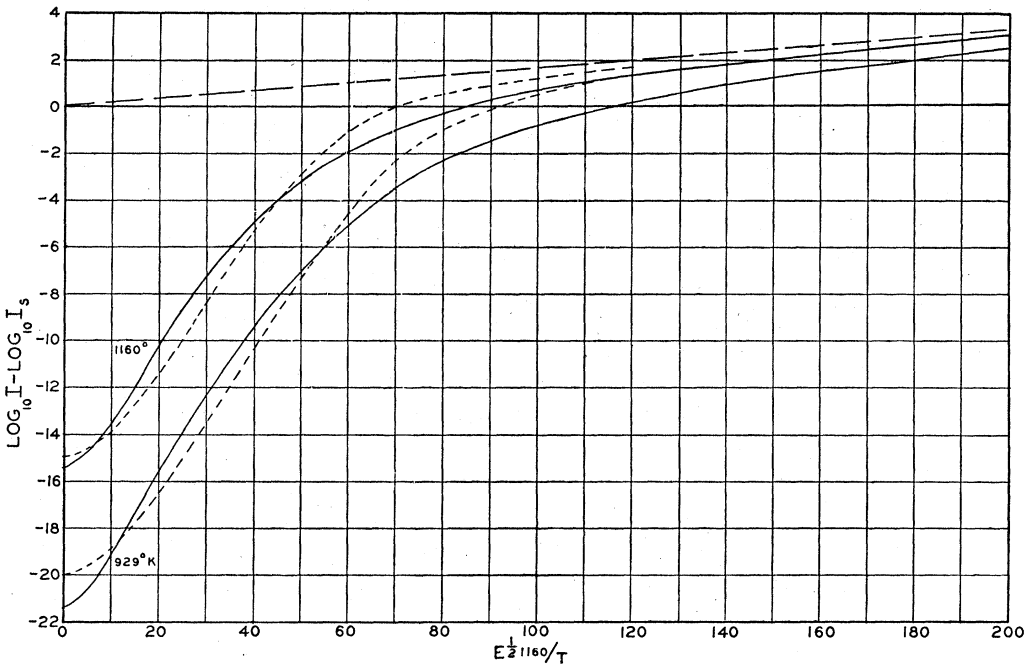


FIG. 7. Computed and experimental Schottky curves for thoriated tungsten, activation $Z=0.25$. Dotted lines computed with $z_0=5\times10^{-4}$ cm and $a=0.465$ volt. Asymptotic Schottky line shown dashed. From Nottingham (N7).

shows Richardson plots for several representative combinations of the four patches where four points were calculated and straight lines fitted to them by inspection. As the lines are drawn, the points lie above the line for the 1000°K temperature. Thus the patch effect tends to make the Richardson lines concave upward. In experiment this would tend to be offset by the end loss of heat from the specimen and sometimes by adsorption of residual gases or diffusion of impurities to the surface. As long as the Richardson plots for the combination of patches are reasonably straight lines, the constants A^{**} and φ^{**} of Fig. 13 are essentially the same as those calculated from Eqs. (II.7.5) and (II.7.6).[i]

[i] Equation (II.7.6) shows that if all A_i^* are assumed equal, the A^{**} is directly proportional to A_i^*. Therefore, Figs. 12 and 13 can be used for any A_i^* which is the same for all patches by multiplying the A^{**} values given by the curves by the constant $A_i^*/120$. Also in Eq. (II.7.5) the individual w_i remain constant as long as the individual values of $\delta\varphi_i^*/T$ are unchanged. Therefore results calculated from Eq. (II.7.5) for a given set of $\delta\varphi_i^*$ and a given temperature can be used for new sets of $\delta\varphi_i^*$ and temperature which are different from the old by the same factor. In this case the old $\delta\varphi^{**}$ must be divided by this factor to give the new $\delta\varphi^{**}$. There is no change in the A^{**} values as can be seen from Eq. (II.7.6). The $\delta\varphi^{**}$ values in Figs. 12 and 13 were computed from Richardson lines drawn through points calculated at 2000°K and 1000°K and not by Eq. (II.7.5). Therefore, one would not expect the above relationship to hold exactly between the various curves of the family even though the corresponding values of temperature were in this range. However, as long as the f values of the low work function patches are not too small—i.e., as long as the Richardson plots such as those of Fig. 14 are reasonably straight lines—the relationship holds fairly well between the various curves of the family.

The next section illustrates the method by estimating the thermionic constants of two types of tungsten wire, one with preferred crystal orientation and the other with random crystal orientation. It is evident that there is considerable arbitrariness in the choice of the combinations of the four basic types of patches used to represent the types of patches included in the estimation. However, this arbitrariness is reduced to some extent by the choice of conveniently large f_i values for the low work function basic patches in each case. This is in order to assure a reasonably straight Richardson line for each type of patch included in the estimation. Also, as will be pointed out later, the low field thermionic constants can serve as a check on the choice of the f_i values.

II.7d *Average High Field Thermionic Constants for Polycrystalline Tungsten Wire*

In the case of polycrystalline specimens, the average thermionic constants depend upon the orientation of the crystals. If the crystals are oriented at random, the result may be different from the case in which there is some measure of preferred orientation.

Tungsten wire which has been properly doped, develops upon heat treatment crystals with a strong preference for alignment of a (110) axis parallel to the wire axis. Such wire is commercially known as "non-sag wire," (see Section II.1a). It

was from this type of wire that Fig. 3 and Table I were obtained. In order to estimate the average thermionic constants for this type of wire by means of Fig. 13, it is necessary to estimate the fraction of the surface taken up by each of the four basic patches. This is done by the use of Table I and Fig. 3. Figure 15 shows a basis for an estimate in which it is assumed that the A_i^{**} and $\delta\varphi_i^{**}$ values vary more or less smoothly between the values given in Table I for the maxima and minima of emission. To get one of the many possible sets of values for f_1, f_2, f_3, and f_4, a combination of patches is chosen for each block of Fig. 15 by use of Fig. 13 in such a way as to give the A_i^{**} and $\delta\varphi_i^{**}$ of that block. In each choice, the fraction of low work function surface is kept as large as convenient in order to keep the resulting Richardson line reasonably straight. Table II gives such a set of values for f_1, f_2, f_3, and f_4. To get the average constants for the wire, the f values are added and with the use of Fig. 13, the corresponding φ^{**} and A^{**} values are obtained. For the values given in Table II, the average φ^{**} value is 4.46 and the average A^{**} is 45. This agrees well with experiments carried out by the author (see (53) Table IV; this does not agree with entry 58), using prewar non-sag wire in which the average current is measured and the φ^{**} and the A^{**} determined from a Richardson plot.

It is to be noticed that the φ^{**} value of 4.46 and the A^{**} value of 45 are somewhat lower than the previously accepted values of about 4.53 and 60 for pure tungsten wire. One possibility which is worth investigating as a cause for this discrepancy is that previously accepted constants may have been obtained from wire with a smaller amount of preferred orientation of the crystals than non-sag wire. It is therefore of interest to calculate the average constant to be expected on the basis of the four patch model with random crystal orientation.

Only a very crude estimate can be made of the f values for random crystal orientation because of lack of quantitative data. The following is such an estimate which is based on Martin's work (M10) using the single crystal sphere because ideally in this case all crystal surfaces have equal exposure and thus corresponds to random crystal orientation in a wire. Inspection of Martin's photographs indicates that the pattern from clean tungsten can be approximated by circular patches of good emission centering on the (116) and (111) directions and patches of very poor emission around the (110) directions. This is corroborated by field emission data (J3). If it is assumed that the emission peaks of Fig. 3 are cross sections of the circular patches, the angles subtended at the center of the sphere by these patches can be approximated by the blocks of Fig. 15. For example, to compute the fractional area of the circular patch centered about the (111) direction, to which is assigned φ^{**} of 4.41 ev and A^{**} of 44, the area of the circular patch on a unit sphere which is cut out by a right circular cone with apex at the center of the sphere and half angle of 6° is calculated. It is assumed that this circular patch is surrounded by an annulus lying between this cone and another cone of half angle 14°, having A^{**} of 80 and φ^{**} of 4.56. Since there are eight (111) directions cutting the sphere, these areas must be multiplied by eight. To get fractional areas these areas are divided by 4π. The same procedure is applied to the (116) direction using a circular patch cut by a cone of 11.5° half-angle with A^{**} of 54 and φ^{**} of 4.39 and an annulus between this cone and another cone of half-angle 16.5° with A^{**} 95 and φ^{**} of 4.56. Since there are twenty-four (116) directions cutting the sphere these areas must be multiplied by 24. Also there are twelve (110) directions with cones of half-angle 12° and A^{**} of 15 and φ^{**} of 4.69. It is assumed that all the rest of the area is described by A^{**} of 120 and φ^{**} of 4.69. Table III summarizes these results. The composite A^{**} is 54 and the composite φ^{**} is 4.49. This is, of course, only a rough estimate. For example, if the total $f_1=0.05$, the total $f_2=0.4$, and the total $f_3=0.4$, the result would have been composite $A^{**}=63$ and composite $\varphi^{**}=4.51$.

Thus the estimate for random orientation gives values more in line with the previously accepted values of A^{**} and φ^{**} for pure polycrystalline tungsten specimens. The literature pertaining to the crystal orientation of pure tungsten polycrystalline wires which have undergone heat treatment typical of thermionic specimens—i.e., flashing to 2900°–3000°K for several minutes—does not seem very conclusive. For example, Jeffries (J2) reported that crystals in hard drawn tungsten wire of 0.0005 inches diameter retained the preferred orientation of the raw wire (110 axis aligned with the wire axis) when heated to around 1500°K. He included x-ray Laue photographs confirming this. On the other hand, Goucher (G5) stated that pure tungsten wire of about 0.008 inches diameter, if flashed for several minutes at 2900°–3000°K, produces crystals oriented at random. Since Goucher's objective was to study the behavior of tungsten crystals under tension, it is possible that he was referring to a random orientation about a (110) axis which was parallel to the wire axis rather than to a completely random orientation. The x-ray Laue photograph which he included is not conclusive in this respect because only a strip cut at right angles to the axis of the wire was included. In order to settle this question, an x-ray analysis should be made using pure tungsten wires which have been given the heat treatment normally given thermi-

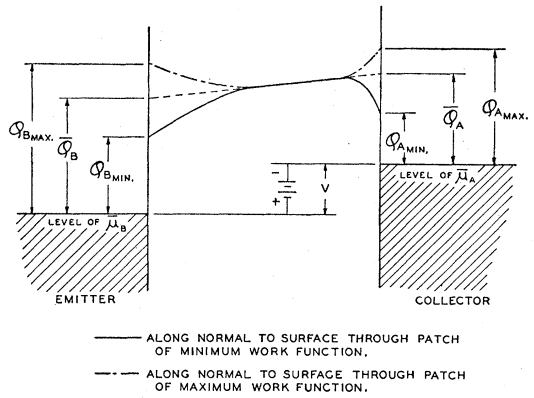


FIG. 8. Schematic of potential variation between a plane patchy emitter and a parallel patchy collector.

onic specimens with wire diameters normally used in thermionic experiments.

The above results show that in the detailed interpretation of thermionic data even from clean metals it is necessary to know in detail the crystalline nature of the emitting surface.

II.8 The Effect of Patches on the Weak Field Richardson Plot

Consider the thermionic emission from a patchy surface collected by fields so small that the motive maximum lies outside the patch field or, more precisely, that Case 2 of Section II.5a is satisfied.

II.8a *Normal Energy Considerations*

If only normal energies are considered, then all patches with work function less than the $\bar{\varphi}$ defined in Eq. (II.4.3) will act as one patch whose work function is $\bar{\varphi}$ and whose fractional area is the sum of the fractional areas of all patches whose work function is less than $\bar{\varphi}$ provided that reflection effects at the metallic surface are reasonably independent of velocity. Then the theory of Section II.7 is applicable. Let all φ_i for $i<k$ be equal to

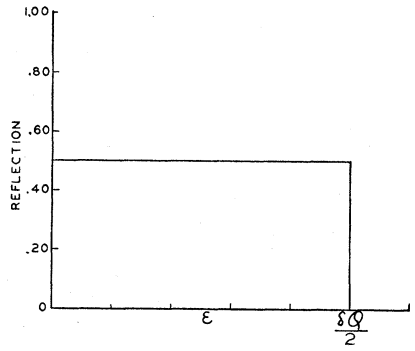


FIG. 9. Associated reflection coefficient for two-patch checkerboard surface as a function of normal energy in electron volts using the normal energy approximation.

or less than $\bar{\varphi}$. The $\Delta\varphi$ of Eq. (I.4.2) is very small because the field strengths are small. If $\Delta\varphi$ is neglected and if it is assumed that $\bar{r}_i$ is zero, then the slope of the Richardson plot gives

$$\bar{\varphi}^{**}=\sum_{i\geq k} w_i\varphi_i^*+w\bar{\varphi}^*, \qquad \text{(II.8.1)}$$

where w_i is the fraction of the total emission from the ith patch. Also w is the fraction of the total emission from all patches whose $\varphi_i<\bar{\varphi}$ and is given by $w=fAT^2\exp(-e\bar{\varphi}/kT)/j_0$, where j_0 is the total current per unit area and $f=\sum_{i<k}f_i$. Also $\bar{\varphi}^*=\bar{\varphi}-Td\bar{\varphi}/dT$. Thus $\bar{\varphi}^{**}$ is a little larger than $\bar{\varphi}^*$ and is therefore larger than the corresponding φ^{**} for high fields. The value of $\bar{A}^{**}$ under these conditions will also be higher than the corresponding A^{**} for high fields because all patches having work function less than $\bar{\varphi}$ act as one patch of work function $\bar{\varphi}$ and fractional area $f=\sum_{i<k}f_i$.

II.8b *Tangential Energy Considerations*

As before, the problem is complicated when tangential as well as normal energies are taken into account. Since the associated reflection coefficients must be averaged over a Maxwellian distribution of normal velocities to get the emission currents, it follows from Fig. 33 of Appendix V that if the associated reflection coefficient is principally a result of the patch effect, the Richardson plot when tangential energies are taken into account will be steeper than with the corresponding normal energy approximation provided, of course, that in the tangential energy case the reflection coefficient does not decrease appreciably with decreasing energy. Therefore, the apparent work function from this case will be greater than that for the normal energy approximation. The corresponding emission constant will also be larger.

II.9 Relative Values of the Strong Field and Weak Field Thermionic Constants for a Patchy Surface

Although the detailed interpretation of the weak field constants is complicated, the approximate value of the weak field work function of a patchy surface is a little larger than $\bar{\varphi}$. On the other hand, the corresponding high field work function is heavily weighted on the low work function side. If the different patches occupy comparable amounts of emitter area and if the A^* do not vary too widely, then the difference between the low field work function and the high field work function is the same order of magnitude as the maximum difference in work function of the patches. In this connection, it is of interest to compare the approximate weak field work functions calculated by using Eq. (II.4.3) and the f values given in Tables II and III with the corresponding calculated strong field

work functions. Application of Eq. (II.4.3) gives 5.05 ev and 4.81 ev, respectively, for the approximate weak field work functions as compared to 4.46 and 4.49 for the high field work functions. The differences between the respective low and high field work functions are the same order of magnitude as the difference in work functions of the assumed fundamental patches. Thus a check on the existence of the high work function patch—i.e., the 5.35-ev patch postulated in Section II.7b—could be obtained from an experimental determination of both the high field and the low field work functions of the same polycrystalline specimen under conditions in which space charge effects are negligible. In this connection it is interesting to note that Nottingham (N7) has reported a zero field work function for clean tungsten of 4.75 volts (see Section III.2).

The dependence of the thermionic constants on the patch nature of the emitting surface and the collecting fields is well illustrated by experimental data from composite surfaces. The general theory is the same as that of the previous sections but the work function differences, especially for small coverage, may be larger. In addition, the nature of the patches and the amount of adsorbed material is dependent upon the temperature so that straight Richardson plots are obtained only over a limited range of temperature. The published data from composite surfaces such as thorium on tungsten, vary greatly. This is undoubtedly due in part to variations in crystal structure and purity of the specimens as well as to differences in vacuum technique. However, the general trend of the data as functions of collecting fields is illustrated by Fig. 16 taken from Brattain and Becker (B23). (See also Rose (R13).) This figure is a plot of approximate data on the dependence of φ^{**} and A^{**} upon collecting voltage for a thoriated polycrystalline tungsten wire of 0.004-inch diameter at various stages of activation. Here the activation is designated by f which is the ratio of activation time to the activation time required to produce maximum emission with high collecting fields. In accordance with the above paragraphs, it is to be noticed that both A^{**} and φ^{**} increase with decreasing field strength. It is important to notice that the increase is less rapid for the higher f values. This means that as f approaches unity, the surfaces become more uniform. This has an important bearing on the interpretation of the A^{**} and φ^{**} from polycrystalline thoriated tungsten wire. For example, Fig. 17 is a plot of Table I from Brattain and Becker (B23) which summarizes their data for the case of thorium evaporated onto a tungsten ribbon $0.035\times0.0036\times11.1$ cm and which gives A^{**} and φ^{**} for 100-volt collecting potential as a function of the activation, f. The authors state that this potential gives

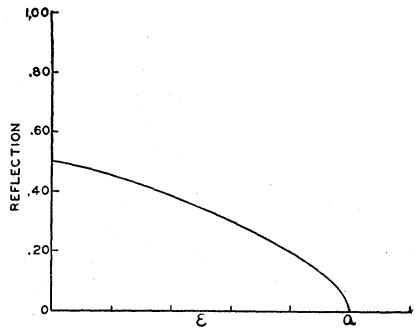


FIG. 10. Associated reflection coefficient for a sinusoidal strip surface as a function of normal energy in electron volts using the normal energy approximation.

saturated emission. Perhaps on account of the tendency for (001) surfaces of tungsten to predominate in the surface of the ribbon (A7), the A^{**} values are higher than those published by Dushman (D7) for thoriated tungsten wire. However, the trend is the same. Now the voltage dependence, Fig. 16, indicates that the surface becomes relatively uniform for values of activation about unity. Thus it is difficult to account for the low values of A^{**} by the patch effect alone. However, the variation of φ^{**} with f is reasonable on the basis of Eq. (II.7.4) because as the activation is increased, the areas covered become larger and the work function of the covered areas decreases to a minimum value. In order to better interpret the data from thoriated tungsten, it would be desirable to have quantitative data on single crystal thoriated surfaces.

The data of Fig. 16 show a difference of about 0.7 volt between the low field and high field work functions for low activation. This is the same order of magnitude as the difference of 0.6 ev between patches reported by Ahearn and Becker (A1). The difference between the low field work function and high field work function and the patch size can both be estimated from a Schottky plot which includes the anomalous region because the anomalous region is the transition between the high fields of case 1, Section II.5a and the low fields of Case 2 of that Section.

II.10 The Effect of Patches on the Calorimetrically Determined Work Function

The cooling effect accompanying thermionic emission has been discussed in Section I.7 for a uniform surface. The heat loss per electron for zero field emission is given by Eq. (I.7.4). Consider as before a patchy surface with collecting fields much stronger than the patch field so as to satisfy Case 1 of Section II.5. Then the heat loss per electron extrapolated to zero collecting field is given by (H8)

$$l_{\text{sat.}}{}^{**}=\sum_i w_i l_{\text{sat.}}{}^{(i)}, \qquad \text{(II.10.1)}$$

where as before, w_i is the fraction of emission from the ith patch and $l_{\text{sat.}}^{(i)}$ is the zero-field heat loss per electron for a uniform surface of the ith type. Using Eqs. (I.5.2), (II.7.4), and (II.10.1), and assuming that $\bar{r}_i$ is independent of temperature, this can be written in terms of φ^{**} of Eq. (II.7.4) (see reference (H8))

$$e\varphi^{**}=l_{\text{sat.}}{}^{**}-\sum_i Tw_i dl_{\text{sat.}}{}^{(i)}/dT+e\sigma T. \quad \text{(II.10.2)}$$

It follows from Section I.7 that the last term is small, of the order of hundredths ev or less, both for metals and semiconductor coatings on metal cores.

The second term is of the order of kT if σ and $d\varphi_i/dT$ are small because from Eq. (I.7.4)

$$Tdl^{(i)}/dT=eTd\varphi_i/dT+e\sigma T$$
$$+eT\int_0^T(\sigma/T)dT+2kT. \quad \text{(II.10.3)}$$

II.11 The Effect of Patches on the Measurement of Contact Potential by Means of Volt-Ampere Characteristics

In addition to the Kelvin method described in Section II.4, several other methods making use of the volt-ampere characteristics of thermionic or photoelectric emission have been worked out. The principal methods are discussed in this section.

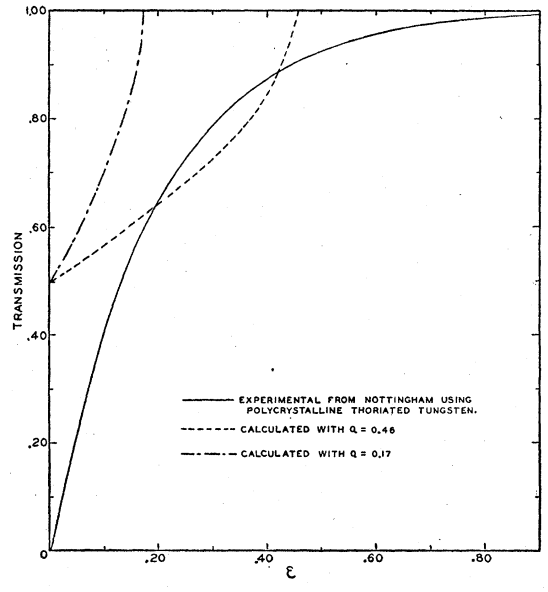


FIG. 11. Experimental transmission coefficient for polycrystalline thoriated tungsten and calculated associated transmission coefficients using the normal energy approximation for a sinusoidal strip distribution, all plotted as functions of normal energy. From Nottingham (N7).

II.11a *The Breakpoint of the Retarding Potential Characteristic*

For uniform plane cathodes in the absence of space charge, the retarding potential characteristic would have an abrupt change in slope at $V=V_0$ which is the condition for zero field between the electrodes. In the case of patchy cathodes, this breakpoint is somewhat smoothed out particularly if the cathode has a wide variation of patch size, patch work function, and patch distribution. This makes the determination of the break point from an inspection of the experimental curve more difficult and less accurate. Also, if the patch size is not too small, the anomalous Schottky Effect (Case 3, Section II.5a) sets in at small accelerating potentials which can obscure the breakpoint.

II.11b *The Intersection of the Tangent to the Retarding Potential Curve and the Tangent to the Saturated Emission Curve*

Consider first the case of plane parallel electrodes with a patchy cathode and uniform anode. It follows from Section II.6 that for values of $V-V_0$ which are much larger than any difference in work functions occurring in the cathode, the $\log(j)$ *versus* $V-V_0$ plot approaches a straight line (as before, $V_0=\bar{\varphi}_B-\bar{\varphi}_A$). This line can be projected to intersect with a tangent drawn to the $\log(j)$ *versus* $V-V_0$ curve for accelerating potentials sufficiently large to give saturation or, more precisely, sufficiently large so that the requirements of Case 1 in Section II.5a are satisfied. Actually the latter curve varies only slowly with $(V-V_0)$ so that if the patch size is not too small, the extrapolation of the tangent will not involve much error. (If the patch size is very small, the extrapolation must extend over a large $V-V_0$ range thereby entailing somewhat more error.) This is illustrated schematically in Fig. 18 for a two-patch checkerboard using the normal energy approximation. Generally, this method makes possible the determination of a characteristic point with considerably more precision than the determination of the breakpoint by inspection. It is clear that the intersection for saturated emission lies to the right of $V-V_0=0$ in Fig. 18.

Heinze and Wagener (H5) have analyzed this method for a patchy cathode. Using the terminology of the previous sections their analysis is briefly as follows. Equation (II.7.3) gives the saturated emission current, j_0,—i.e., the current collected with high accelerating fields so that Case 1 of Section II.5a is satisfied—extrapolated to zero field. Define a mean work function $\tilde{\varphi}$ such that

$$j_0=AT^2\exp(-e\tilde{\varphi}/kT)\sum_i f_i 1-\bar{r}_i. \quad \text{(II.11.1)}$$

Hence from Eq. (II.7.3)

$$\tilde{\varphi}=\left(\frac{kT}{e}\right)\ln\left[\frac{\sum_i f_i(1-\bar{r}_i)}{\sum_i f_i(1-\bar{r}_i)\exp(-e\varphi_i/kT)}\right]. \quad \text{(II.11.2)}$$

Neglecting reflection, the $\tilde{\varphi}$ so defined is related to φ^{**} of Section II.7 by the equation

$$\varphi^{**}=\tilde{\varphi}-Td\tilde{\varphi}/dT. \quad \text{(II.11.3)}$$

Now consider the case of plane parallel electrodes with a patchy cathode and a uniform anode of work function φ_A. Apply a retarding voltage $V-V_0 > \varphi_{B\,\max}-\bar{\varphi}_B$, where $\varphi_{B\,\max}$ is the highest work function patch appearing in the cathode and $\bar{\varphi}_B$ is the average work function of the cathode as defined by Eq. (II.4.3). Then if only normal energies are considered, the emission from each patch reaching the collector is independent of the presence of the neighboring patches. If the velocity distribution of the emitted electrons is assumed to be Maxwellian,—i.e., $\bar{r}_i$ is independent of velocity—the current per unit area of the ith type of patch reaching the collector is given by

$$j_i=AT^2(1-\bar{r}_i)\exp[-e(V+\varphi_A)/kT], \quad \text{(II.11.4)}$$

where V is the applied retarding potential—i.e., the positive side of V is connected to the cathode. Therefore the total current reaching the anode from all the patches of the cathode per unit area of the cathode is given by

$$\ln j=\ln[AT^2\textstyle\sum_i f_i(1-\bar{r}_i)]-e(V+\varphi_A)/kT. \quad \text{(II.11.5)}$$

The intersection of the projection of the straight line portion of the retarding potential curve and the tangent to the saturation current curve occurs when the current j of Eq. (II.11.5) equals the current j_0 of Eq. (II.11.1). To within the approximations outlined in the first paragraph of II.11b, this gives at the intersection point

$$V=\tilde{\varphi}_B-\varphi_A. \quad \text{(II.11.6)}$$

In terms of the V_0 defined in Section II.4, the intersection occurs at

$$V-V_0=\tilde{\varphi}_B-\bar{\varphi}_B. \quad \text{(II.11.7)}$$

Therefore since generally $\tilde{\varphi}_B<\bar{\varphi}_B$ the intersection occurs to the right of the origin of $V-V_0$ in Fig. 18. In the case of a two-patch checkerboard emitter, $\tilde{\varphi}_B-\bar{\varphi}_B\sim-\delta\varphi/2$ where, as before, $\delta\varphi$ is the difference in work function of the two patches. In connection with this method it is important that both $\tilde{\varphi}_B$ and $\bar{\varphi}_B$ are different temperature dependent averages of the individual patch work functions.

If the patch size of the cathode is sufficiently small so that Case 2 of Section II.5a is satisfied over an observable range of small accelerating potentials, a tangent may be drawn to this portion of the characteristic as indicated in Fig. 18. This case is more difficult to analyze but a rough estimate is as follows. If only normal energies are considered and if the emitted electrons have a Maxwell velocity distribution, the current per unit area for small accelerating potentials satisfying Case 2 can be written

$$\ln j=\ln\bar{A}_B{}^{**}T^2-e\bar{\varphi}_B{}^{**}/kT, \quad \text{(II.11.8)}$$

when $\bar{A}_B{}^{**}$ and $\bar{\varphi}_B{}^{**}$ are defined in Section II.8a. At the intersection point

$$\ln\bar{A}_B{}^{**}T^2-e\bar{\varphi}_B{}^{**}/kT=\ln AT^2\textstyle\sum_i f_i(1-\bar{r}_i) -e(V+\varphi_A)/kT. \quad \text{(II.11.9)}$$

This can be written

$$V-V_0=\bar{\varphi}_B{}^{**}-\bar{\varphi}_B +(kT/e)\ln[A\textstyle\sum_i f_i(1-\bar{r}_i)/\bar{A}_B{}^{**}]. \quad \text{(II.11.10)}$$

If $\bar{r}_i$ is small, then $V-V_0$ at the intersection point has a small positive value and hence the intersection occurs slightly to the left of $V-V_0=0$ in Fig. 18. In the case of the two patch checkerboard with $\delta\varphi=0.5$, $A_1{}^*=A_2{}^*=120$ and $\bar{r}=0$, it can be seen from Fig. 12 that $\bar{A}_B{}^{**}$ is about 90 and $\bar{\varphi}_B{}^{**}=\bar{\varphi}_B+0.03$. If $T=1160$°K, then $V-V_0=0.06$ as indicated in Fig. 18.

The above discussion has been based on the normal energy approximation. Since it is natural to presume that the associated reflection coefficient is zero for normal kinetic energies well above $e(\varphi_{B\,\max}-\bar{\varphi}_B)$ even when tangential energies are taken into account, the above results in the case of saturated emission—i.e., Case 1—are valid, but the results for the low field case—i.e., Case 2—must be corrected.

In the above analysis, it has been assumed that the collector surface is uniform. If the collector surface is patchy, φ_A should be replaced by $\bar{\varphi}_A$ defined by Eq. (II.4.3) and reflection in the patch field of the anode should be corrected for in a manner analogous to that outlined in Section II.6c. If cylindrical geometry with cathode diameter small compared to anode diameter is used, Section II.6c indicates that reflection in the anode patch field can probably be neglected. In the cylindrical case, the above analysis must be modified to take into account the curvature of the retarding potential characteristic for uniform surfaces. To allow for this, Heinze (H4) has applied a numerical correction thereby permitting the use of the above results.

This method with cylindrical geometry has been used to observe work function changes in the anode or in the cathode. If the $\bar{\varphi}_A$ of the anode is constant, a shift in the intersection point (using saturated emission) indicates a change in the $\tilde{\varphi}_B$ of the cathode. For example, Heinze and Wagener (H5) have

used this method to observe changes in $\bar{\varphi}_B$ during the activation of oxide cathodes. On the other hand, if the $\bar{\varphi}_B$ of the cathode is held constant, then the shift of the intersection point is due to a change in the $\bar{\varphi}_A$ of the anode. This has been used by Rothe (R15) and others to follow changes in $\bar{\varphi}_A$.

If this method is used to follow changes in the cathode, the geometry can be easily arranged for convenience in making Richardson plots of the cathode emission. This is one of the main advantages of this method.

II.11c *The Space Charge Limited Characteristic*

If the emission is space-charge limited and if initial electron velocities are taken into account, it turns out that the potential outside the cathode surface is depressed so that electrons which reach the anode must pass over a barrier higher than the surface barrier of the cathode. Gysae and Wagener (G12) using Langmuir's theoretical results for a Maxwellian distribution of emitted electrons have pointed out that the current collected by the anode depends mostly on the applied voltage and the $\bar{\varphi}_A$ of the anode and only slightly, if at all, on the work function of the cathode. They verified this result experimentally by using first a clean tungsten cathode and then the same cathode with a layer of barium oxide on it. No shift of the characteristic was observed. Gysae and Wagener (G13) have also carried out an experiment in which a patchy anode was simulated and verified in this manner that the current collected depends on the average anode work function $\bar{\varphi}_A$. The effect of a change in $\bar{\varphi}_A$ is to shift the accelerating volt-ampere characteristic a corresponding distance along the V axis. Since in this case accelerating potentials are used, reflection effects at the anode do not occur. Thus this method is applicable to the measurement of changes in anode work function $\bar{\varphi}_A$ or to the measurement of the difference of the $\bar{\varphi}_A$ of different materials used as anodes with the same geometry. This method has been used by Riemann (R3), Gysae and Wagener (G12), I. Langmuir and Kingdon (L6), and others. A method similar to this but not space-charge limited has also been used (see for example (M15) and (L2)). This method has a disadvantage, compared to the space-charge limited method, in that the results are directly dependent upon the constancy of cathode conditions during measurements.

II.11d *The Electron Beam Method*

A modification of the above methods has been used by Anderson (A2) who determined contact potential differences between two surfaces by plotting the retarding potential characteristics of a beam of low energy electrons using first a plane collector of one surface and then the other surface. By an inspection of the characteristics he chooses a collector current, j_c, in the steepest portion of the characteristic and determines the corresponding potentials. The difference in these two potentials is taken to be the contact potential between the two plane surfaces. If the collector patch size is sufficiently small compared to the size of the electron beam and the electrode spacing, the general reasoning of the preceding sections can be applied to estimate the effect of patches. If only electron energies normal to the collector are considered, the total collector current can be considered as the sum of the currents collected by the various types of patches each acting as a uniform surface independent of the neighboring patches except that, as before, those patches whose work functions are less than $\bar{\varphi}_A$ act as one patch of work function $\bar{\varphi}_A$ and fractional area given by $\sum f_i$ summed over-all patches whose $\varphi_{Ai}<\bar{\varphi}_A$. The shape of the volt-ampere characteristic for each type of patch depends on the normal energy distribution of the electrons in the beam and is the same shape for each type of patch. However, the volt-ampere characteristic for each type of patch is displaced along the voltage scale by an amount $\varphi_{Ai}-\bar{\varphi}_A$ for all patches whose $\varphi_{Ai}>\bar{\varphi}_A$.

For purposes of discussion, consider again a two-patch checkerboard surface on the collector whose checker size is small compared to the beam size and electrode spacing and let $\varphi_{A2}>\varphi_{A1}$. Let the volt-ampere characteristic of the current per unit area collected by the φ_{A2} patches be somewhat as indicated by the dashed line in Fig. 19. The current per unit area collected by the φ_{Ai} patches will then be the same shape but displaced a distance $\delta\varphi_A/2$ to the left on the voltage scale and is indicated by the dash dot line. The over-all characteristic per unit area will be half the sum of the currents from the individual types of patches and is indicated by the solid line. Let the reference current, j_c, for the case of a uniform surface of work function $\bar{\varphi}_A$ occur at $V-V_0=V_c$, where $V_0=\bar{\varphi}_B-\bar{\varphi}_A$. Then for the patch surface, if $\delta\varphi_A/2$ is appreciably smaller than the range of $V-V_0$ over which the characteristic is essentially linear, the reference current occurs for $V-V_0=V_c+\delta\varphi_A/4$. Now if the above collector is replaced by a collector having values $\bar{\varphi}_A'$ and $\delta\bar{\varphi}_A'$, the reference current occurs for $V'-V_0'=V_c+\delta\varphi_A'/4$. Since $V_0-V_0'=\bar{\varphi}_A'-\bar{\varphi}_A$, the difference $V-V'$ in the applied potentials between these two cases is

$$V-V'=\bar{\varphi}_A'-\bar{\varphi}_A'+\tfrac{1}{4}(\delta\varphi_A'-\delta\varphi_A) \quad \text{(II.11.11)}$$

provided that the $\bar{\varphi}_B$ of the cathode remains constant. Thus on the normal energy picture, the difference $(V-V')$ involves not only the difference in the surface average of the work functions but also the work function difference of the patches.

If tangential kinetic energies in the electron beam are taken into account, the problem is much more difficult to analyze. Consider for example, two collector surfaces with equal $\bar{\varphi}_A$ but with different associated reflection coefficients and let the distribution of velocities in the electron beam be Maxwellian for the electrons which reach the neighborhood of the collector. Then it is easy to show that

$$V - V' = kT/e \ln[(1-\bar{r})/(1-\bar{r}')], \quad \text{(II.11.12)}$$

where $\bar{r}$ and $\bar{r}'$ are the associated reflection coefficients averaged over a Maxwell distribution (see Appendix V). In general, when tangential energies are taken into account, the average reflection coefficients, $\bar{r}$, will be a complicated function of the patch arrangements and work function differences.

II.11e *The Magnetron Method*

Oatley (O1) has developed a method of determining contact potential by the use of a magnetic field parallel to the axis of a filament with concentric anode. His method is to plot the square of the magnetic field, H_0, required to reduce the current, in the absence of space charge, to one-half its field free value *versus* the applied voltage V. The result is approximately a straight line and the intercept on the V axis, after correction for initial velocities, is the contact potential. The previous sections indicate that this is the difference in $\bar{\varphi}_A$ and $\bar{\varphi}_B$ of the anode and cathode if the local effects on the anode are neglected. This neglect may not be entirely justified because with the applied magnetic field a significant fraction of the electrons "graze" the anode so if patches of appreciable size exist on the anode, the "grazing" electrons can be appreciably affected.

The main advantage claimed for this method over the intersection point method is that larger anode currents are drawn at considerably higher electron energy so that a beneficial cleaning of the anode results by this electron bombardment. This method can be used for observations of the contact potential difference of the anode against the cathode as well as for changes in $\bar{\varphi}$ of either the cathode or anode.

II.11f *Comparison of Methods*

All the above methods, with the exception of the space charge limited characteristic, involve the difficult problem of evaluating patch effects at the emitter or collector. As pointed out in Section II.11b in connection with the intersection point method, the effect of reflection in the patch field of the collector can be minimized and can perhaps be neglected, if the collector is cylindrical (or spherical) with diameter very large compared to that of the emitter. If this geometry is used, the intersection point method can give relatively reliable measurement of changes in the $\bar{\varphi}_A$ of the collector. In this connection it should be emphasized that the Kelvin method described in Section II.4 can give reliable measurement of the difference in the $\bar{\varphi}$ of the two electrodes free from the problem of motion of electrons in the patch fields. If temperature gradients exist anywhere in the circuit,

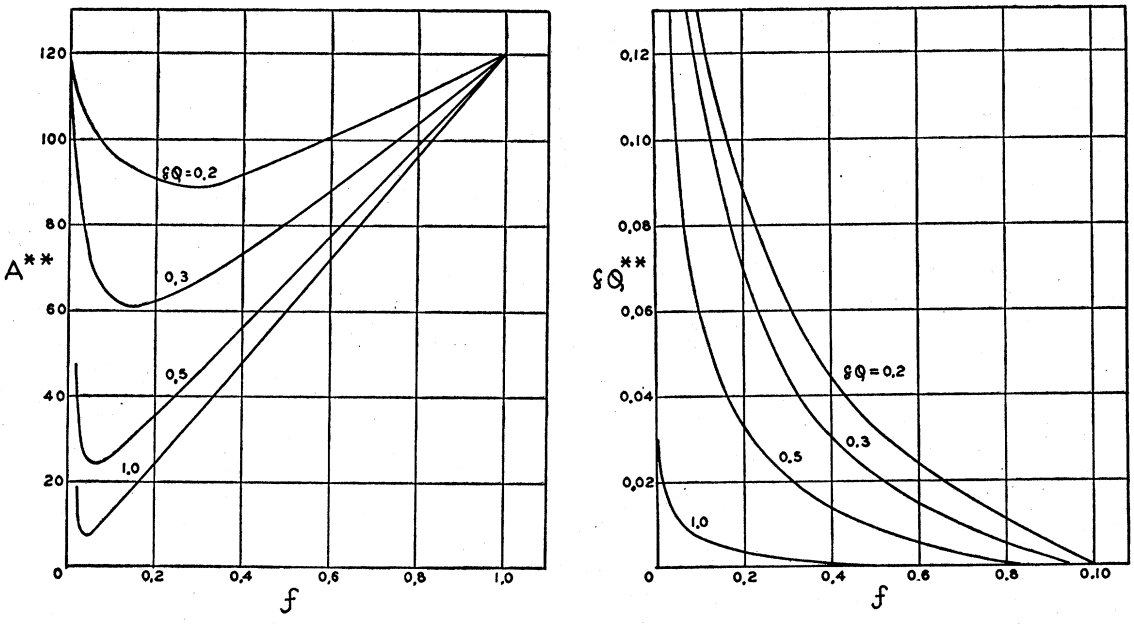


FIG. 12. Computed values of A^{**} and $\delta\varphi^{**}$ for a two-patch surface with high collecting fields, where f is the fraction of the surface occupied by the low work function patches.

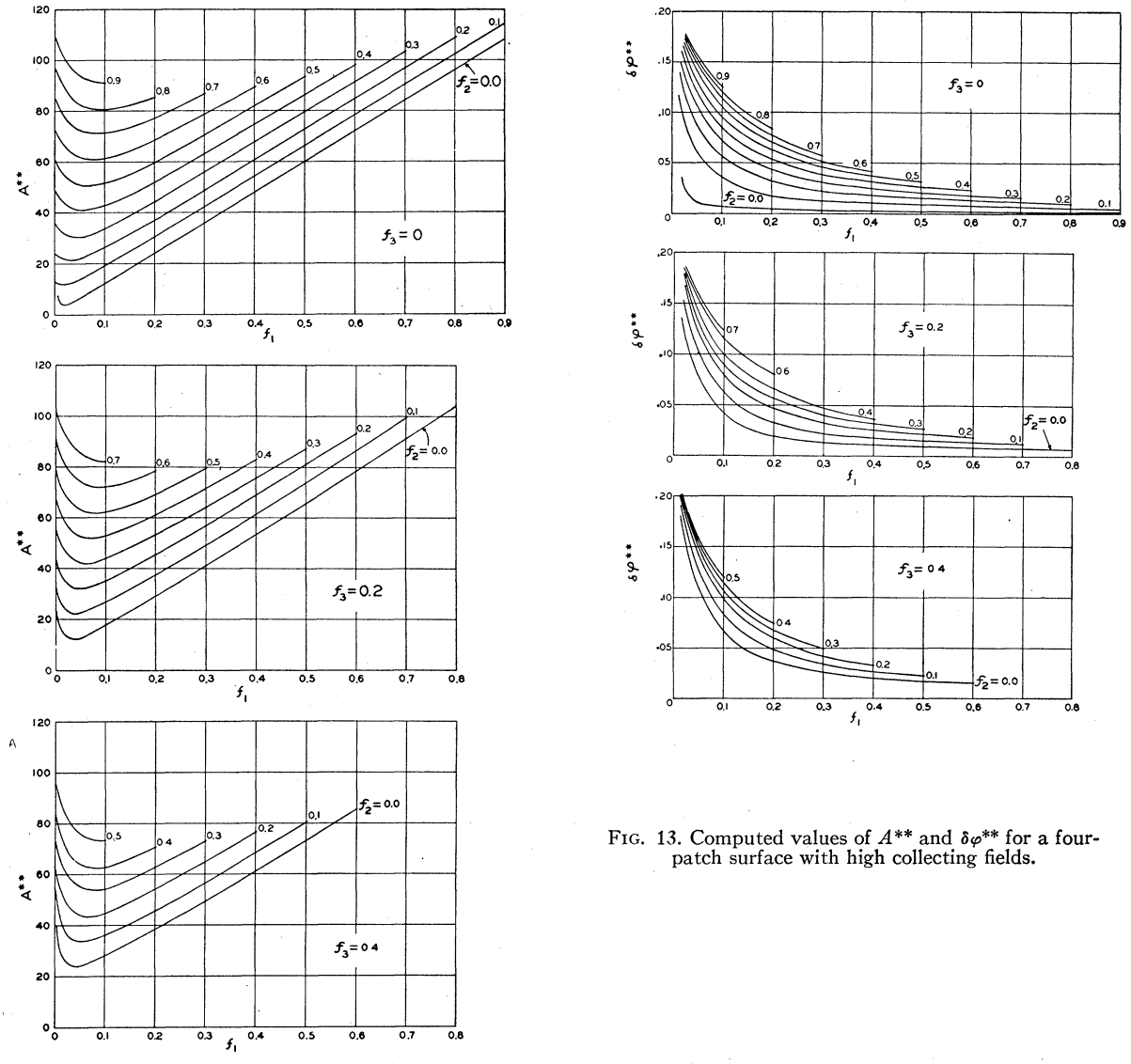


FIG. 13. Computed values of A^{**} and $\delta\varphi^{**}$ for a four-patch surface with high collecting fields.

it is necessary to correct for thermoelectric effects (see Section I.6).

II.12 Photoelectric Emission

Many of the results of the above section carry over into photoelectric emission. Although it is not the purpose of this review to include the field of photoelectric emission, the important photoelectric patch effects will be briefly discussed in this section because of their close relationship with the preceding sections.

The photoelectric emission from patchy surfaces as a function of accelerating collecting fields strength has been investigated in numerous papers.[i] The results are in good agreement with Section II.5. The experimental determination of the velocity distribution of photoelectrons has also been the subject of numerous researches.[k] In the low energy regions, these measurements are subject to the same difficulties as the measurements of the velocity distribution of thermally emitted electrons if the surfaces of the emitter and collector are not of uniform work function.

The most accurate determinations of the photoelectric work function are obtained from an analysis of the shape of the spectral distribution curve near the long wave-length limit. A method for analyzing the data has been worked out by Fowler (F6) who derived, under simplifying assumptions, the following expression for the spectral distribution curve for a uniform surface at temperature T:

$$I = HT^2 p[(h\nu - e\varphi)/kT], \qquad \text{(II.12.1)}$$

[i] See for example (L11), (S19), (S20), (B21), (I1), (K1).

[k] See for example (N6), (M12), (L8), (O4), (H14), (H12).

where I is the emission per unit area extrapolated to zero field. H is a constant, ν the frequency of the radiation, and φ the work function at the temperature T. The function p is a series whose numerical values have been tabulated. This expression is meant to hold only near the threshold. Equation (II.12.1) can be written

$$\log(I/T^2) = C + P(x), \qquad \text{(II.12.2)}$$

where

$$C = \log H, \quad P = \log p \quad \text{and} \quad x = (h\nu - e\varphi)/kT.$$

In Fig. 20 the curve marked $f = 0.5$ and $\delta\varphi = \infty$ is a plot of Eq. (II.12.2).

Fowler's method of determining the true work function from the spectral distribution curve of photoelectric emission from a uniform surface at the temperature T is as follows. Plot the measured values of $\log(I/T^2)$ near the threshold *versus* the quantity $h\nu/kT$. Using the same graphical units, plot the function $P(x)$. Superimpose the two curves by a simple translation of axes. The shift along the $\log(I/T^2)$ axis represents the constant C. The shift along the $h\nu/kT$ axis gives $e\varphi/kT$.

An alternative method has been worked out by DuBridge (D6) in which the frequency of the incident light is held constant and the temperature of the emitter varied. The data are plotted as $\log(I/T^2)$ *versus* $\log(1/T)$ where I is the current per unit area and T the temperature. Then a plot of $\log P(x)$ *versus* $\log x$, using the same graphical units is superimposed on the experimental data by a simple translation of axes. The shift along the $\log(1/T)$ axis gives $\log(h\nu - e\varphi)/k$. Since ν, the frequency of the radiation is known, φ can be computed. The shift along the $\log(I/T^2)$ axis involves the constant H, the intensity of the light, etc. Linford (L11) has pointed out that if H is temperature dependent, difficulty in superposition of the theoretical curve on the data may result. The same can be said for φ.

Still another method which has often been used is that of the complete photoelectric effect,—i.e., measurement of the emission from a photoelectric cathode when illuminated by black body radiation of various temperatures. Since some misconceptions are prevalent regarding the thermodynamic significance of the emission-temperature relation, a theoretical discussion of this method is given in Appendix VI; no further discussion will be included in this chapter.

II.12a *The Effect of Patches on the Determination of the Strong Field Photoelectric Work Function*

If the applied field strength is much greater than the patch fields—i.e., large enough to satisfy Case 1, Section II.5a—the emission from each patch is independent of its neighbors. Assuming that Eq. (II.12.1) is correct for each patch alone, the extrapolated zero-field emission from unit area of a polycrystalline surface is

$$I = \sum_i f_i H_i T^2 p[(h\nu - e\varphi_i)/kT], \qquad \text{(II.12.3)}$$

where H_i and φ_i are characteristic of the ith type of patches. The rest of the notation is the same as in the preceding sections. Figure 20 is the plot of the logarithm of Eq. (II.12.3) *versus* $(h\nu - e\varphi_1)/kT$ for a surface having two different kinds of patches of equal area. The work functions of the patches are φ_1 and φ_2; $\delta\varphi = \varphi_2 - \varphi_1$. It is assumed that $H_1 = H_2$ and that $T = 273°\text{K}$.

Actually, polycrystalline emitting surfaces are generally made up of a variety of patches having different work functions with size and arrangement more or less random so that the result will be a more or less smooth deviation from the Fowler curve for a uniform surface similar to the deviation in Fig. 20. Inspection of Fig. 20 shows that if the work function differences range over a few tenths of a volt, an approximate fit can be obtained by displacing the fowler curve for a uniform surface ($\delta\varphi = \infty$) upward and to the right an amount corresponding to about several hundredths of an electron volt at room temperature. This corresponds to an average photoelectric work function for the composite surface several hundredths of a volt higher than the work function of the lowest work function type patch in the surface, provided, of course, that

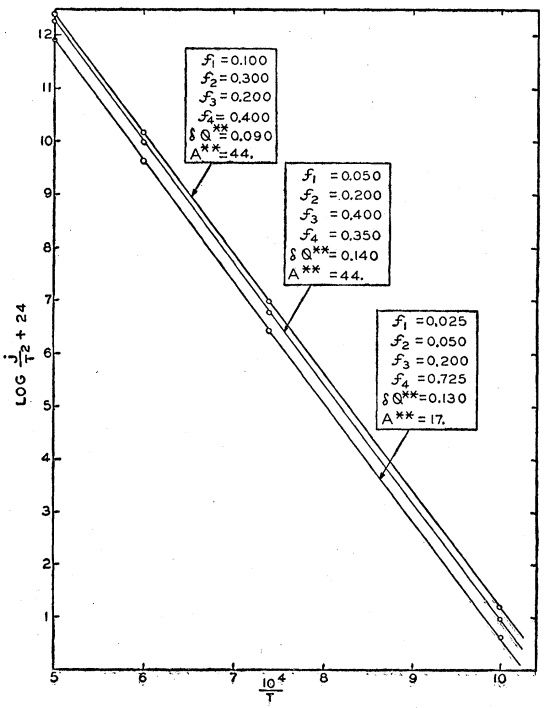


FIG. 14. Calculated Richardson plots from four patch surfaces of Fig. 13.

TABLE II. Patch estimates for preferred crystal orientation.

Angles	φ^{**}	A^{**}	Fraction of surface	f_1	f_2	f_3	f_4
−43° to −40°	4.56	120	0.03		0.030		
−40° to −17°	4.39	54	0.29	0.070	0.067	0.058	0.096
−17° to −12°	4.56	95	0.06		0.050		0.010
−12° to 0°	4.69	120	0.13			0.130	
0° to 8°	4.56	80	0.09	0.002	0.045	0.036	0.017
8° to 20°	4.41	44	0.13	0.015	0.017	0.026	0.072
20° to 28°	4.56	80	0.09	0.002	0.045		0.040
28° to 35°	4.69	120	0.08			0.080	
35° to 47°	4.69	15	0.13			0.013	0.117
			Sum	0.089	0.254	0.343	0.352

this type of patch occupies an appreciable portion of the area. It is of importance to notice that the effect of patches on the Fowler plot depends on the temperature because the work function differences enter into the function $P(x)$ divided by kT so that for higher temperatures the deviation from the Fowler plot for uniform surfaces sets in at lower values of x. This means that for higher temperatures the average photoelectric work function obtained by the method outlined above may deviate less from the work function of the lowest work function type of patch and therefore may decrease with increasing temperatures. In the thermionic case, the average apparent work function increases with increasing temperature relative to the work function of the lowest work function type of patch. This may explain in part the results of King (K1).

Figure 21 is a plot of Eq. (II.12.3) using the DuBridge method in which $\log I/T^2$ is plotted against $\log(x)$ for different values of $\delta\varphi$. In the case that the theoretical curve for a uniform surface, represented by the curve for $\delta\varphi=\infty$, is superimposed to give the best approximate fit to the average data from a patchy surface, Fig. 21 indicates that the effect is a shift to the left and down thereby giving an average work function lower than that of the lowest work function type of patch. Inspection of Fig. 21 shows that the average determined by the method may be several hundredths of a volt lower than the work function of the lowest work function patch depending on how close the illuminating frequency is to the threshold frequency of the lowest work function patch. However, if accurate data are taken at the lower temperature and at frequencies near the threshold of the lowest work function patches and if the theoretical curve is superimposed on the data for only the lower temperatures, Fig. 21 indicates that the patch effect can be reduced. The main reason for this is that the plot by the DuBridge method has appreciable curvature for low values of x.

II.12b *The Effect of Patches on the Determination of the Weak Field Photoelectric Work Function*

If the field strength is weak enough to give motive maximum outside the patch field—i.e., weak enough to satisfy Case 2 of Section II.5a—and if only normal energies are considered, the emission per unit area is given by Eq. (II.12.4) in which $\varphi_i<\bar{\varphi}$ for all $i<k$, $\bar{H}$ is the effective constant for the emission from the patches whose work functions are less than $\bar{\varphi}$ and

$$f=\sum_{i<k} f_i$$

$$I=f\bar{H}T^2 p[(h\nu-e\bar{\varphi})/kT] + \sum_{i>k} f_i H_i p[(h\nu-e\varphi_i)/kT]. \quad \text{(II.12.4)}$$

Assuming that the constant H is the same for all patches and is independent of the velocity of the emitted electrons, the plot of the logarithm of Eq. (II.12.4) for a surface made up of two different kinds of patches would be the same as Fig. 20 except that the φ_1 of Fig. 20 becomes for this case $\bar{\varphi}$ and the $\delta\varphi$ marked on the curves must be multiplied by two. The curves, Fig. 21, for the alternative method of DuBridge would follow in the same way.

II.12c *Agreement with Experiment*

Jamison and Cashman (J1) and Farnsworth and Winch (F2) have published Fowler plots which they recognized as distinctly showing the effect of

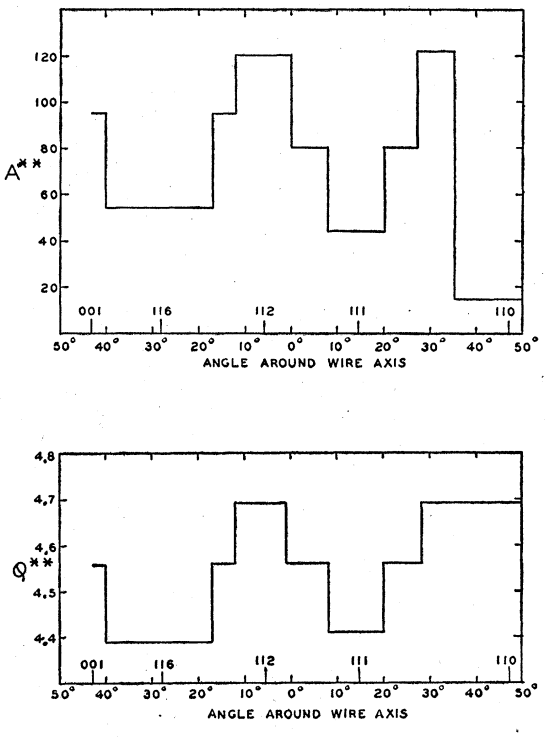


FIG. 15. Approximate values of A^{**} and φ^{**} of single crystal tungsten wire (prewar non-sag variety) as a function of the angle measured about the wire axis.

patches. In both cases the shape of the plots was in agreement with that of Fig. 20. Mann and DuBridge (M7) have also published a Fowler plot which clearly shows patch effects. In numerous other cases the experimental data could be fitted as well by a Fowler curve taking into account the patch effect as by the Fowler curve for a uniform surface. One of the reasons for this is that the Fowler curve is more or less linear at the higher values of x with most of the curvature at the lower values of x. However, most of the experimental points are often obtained at the higher values of x because of the convenience in measuring the larger currents and perhaps also because the percentage change in currents due to unstable surface conditions is smaller at the higher values of x. Now from Fig. 20 it can be seen that under these conditions, the superposition of the theoretical curve for a uniform surface and the data by simple translation of axes entails an amount of arbitrariness which may obscure the patch effect. Another point is that photoelectric data are often taken with electrode geometry which results in rather low applied collecting field strengths at the surface of the emitter so that in many cases the weak collecting field results of Section II.12b apply. In these cases, the work function differences must be taken relative to $\bar{\varphi}$ of the emitter. In any case, the photoelectric Schottky effect if observed over a sufficiently wide range of collecting field strengths, will indicate the magnitude of patch effect to be expected in connection with the Fowler plots.

II.12d *Comparison of the Average Thermionic and Photoelectric Emission Constants and the Contact Potential from Non-Uniform Surfaces*

In order to obtain information on such things as the temperature dependence of the work function, comparison of work functions obtained from thermionic, photoelectric, and contact potential data have often been made. (See for example (B9), (C2).) If the surfaces in question are not uniform, it is clear from the above that the thermionic and photoelectric work functions involve averages different from each other and from the averages entering into the contact potential. These averages also depend on the temperature in a different way. In the case of collecting fields strong compared to the patch field, the thermionic work function obtained from a Richardson plot and the photoelectric work function obtained from a Fowler plot may differ by several hundredths of a volt for a metal like tungsten. If the photoelectric and thermionic workfunctions are determined with different collecting field strengths—i.e., for example the photoelectric work function at low collecting field strength compared to the emitter patch field (as is often the case with the geometry used in photoelectric experiments, because of the convenience of large emitting areas) and the thermionic work function at high collecting field strengths compared to the patch field—the results may differ by an amount of the same order of magnitude as the work function differences appearing in the emitting surface provided, of course, that the surface area is relatively evenly divided between the high and low work function patches.

TABLE III. Patch estimates for random crystal orientation.

Surface	Fraction of surface	φ**	A**	f_1	f_2	f_3	f_4
116 spot	0.240	4.39	54	0.058	0.050	0.048	0.084
	0.252	4.56	95		0.210		0.042
111 spot	0.022	4.41	44	0.003	0.003	0.004	0.012
	0.097	4.56	80	0.001	0.068		0.028
110 spot	0.120	4.69	15			0.015	0.105
Balance	0.269	4.69	120			0.269	
			Sum	0.062	0.331	0.336	0.271

Now in order to take into account in a quantitative way the effect of patches on the average data, it is first necessary to know in detail the emission constants, arrangement, and relative areas of the patches in the surface. This would require something like a quantitative electron projection tube study of the surface using both thermionic and photoelectric emission in which case the average emission data from the non-uniform surface have their value mainly as a check on the electron projection tube data. Certain types of surfaces such as single crystal surfaces, surfaces formed by condensation of metal vapors on amorphous or on single crystal surfaces (A6), or surfaces of ribbons formed by rolling of such metals as tungsten which show preference for the (100) surfaces to lie in the surface of the ribbon (A7), may exhibit small patch effects. As pointed out before, an estimate of the extent of the patch effects can be determined by electron projection tube studies of the surface or by a Schottky plot over sufficiently wide range of surface collecting fields.

CHAPTER III. SURVEY OF RECENT EMISSION DATA

The purpose of this chapter is to list the experimentally determined emission constants for clean metals published since the last review and to discuss briefly certain aspects of recent important papers relating to thermionic emission on which the authors feel that the material in the other chapters of this paper may shed some light.

III.1 Measured Emission Constants

Since the last review article, numerous measurements of the emission constants of clean metals have been published. In order to bring the sum-

mary of such measurements up to date, Tables IV and V have been included. Table IV summarizes the data taken in connection with polycrystalline surfaces[1] and Table V summarizes the data taken from single crystal surfaces.

In connection with the polycrystalline specimens listed in Table IV, it is unfortunate that in many cases the voltage dependence of emission was not included in the published report. Most of these cases were for photoelectric studies in which the specimens were in the form of surfaces evaporated onto a base material and were of such size and geometry that it is likely that the collecting fields were small enough to approximate those of Case 2 of Section II.5. For those cases for which the collecting conditions are not clearly indicated in the reference, the probable conditions are stated. Since it has been shown that in many cases the crystalline nature of the surface influences the results, each entry includes a comment on the nature of the specimen.

In Table V are listed the results from single crystal surfaces. With exception of entry 5, the measurements were made on large metallic specimens which are difficult to clean up by heat treatment. Also undesirable thermal etching is encountered. Anderson's work, entry 5, on silver evaporated onto rock salt is of interest because of its apparent reproducibility but more investigation is required. For the thermionic constants from a single crystal of tungsten, see Table I. Since the measurements of Table I were made with a fine single crystal wire, it was possible to clean the wire thoroughly by heat treatment; however, surface etching on certain crystal surfaces was observed. These measurements should be repeated.

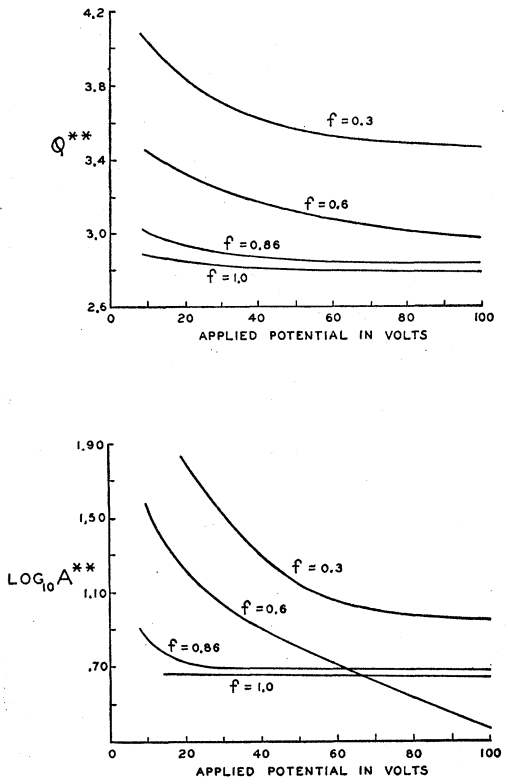


FIG. 16. Dependence of A^{**} and φ^{**} from polycrystalline thoriated tungsten wire upon collecting voltage for various degrees of activation. Taken from Brattain and Becker (B23).

[1] The extensive measurements of the contact potential of many different metals in purified argon reported by O. Klein and E. Lange, Zeits. f. Elektrochemie **44**, 542 (1938), have not been included in this table.

III.2 The Velocity Distribution of Thermally Emitted Electrons and the Reflection Effect

Since the last review, an extensive and careful study of the velocity distribution of thermally emitted electrons has been carried out by Nottingham (N7). Measurements of the volt-ampere characteristics for retarding potentials were made on thoriated polycrystalline tungsten wire and also on clean tungsten wire but the latter were stated by Nottingham to be less reliable and were not reported in detail.

Nottingham found that in the low energy region the retarding potential curves deviated on the low side from that expected on the basis of a Maxwell distribution. After choosing zero field points on the retarding potential curves by inspection, he found that the emission into a retarding field from the thoriated polycrystalline tungsten surfaces could be simulated for *all* states of activation by a uniform surface with the *same* associated reflection coefficient (see Section II.6), namely $\exp(-mv_x^2/2\omega)$, where v_x is the velocity in the direction normal to the surface and where ω is a constant whose value Nottingham determined as 0.191 ev. When the values of the logarithm of the currents at the zero-field points for the various states of activation, as determined by Nottingham, are plotted against the applied potential, they all fall nearly on a straight line.

It seems very likely that patch fields were present even for the state of activation which Nottingham designated as the "dispersed state" because Nottingham's Schottky plots show the low field deviation characteristic of patchy surfaces and also because electron projection tube data from clean polycrystalline tungsten and from thoriated polycrystalline tungsten show pronounced patch effects. (See Sections II.1a and II.1d.) The voltage dependence of emission constants from thoriated tungsten wire measured by Brattain and Becker (B23) and Rose (R13) indicates patch effects which change with activation (see Fig. 16). Therefore, it is at first surprising that Nottingham's experimen-

tally determined associated reflection coefficient is independent of activation since one would expect from Section II.6 that the associated reflection coefficient would change if the work function differences between the patches changed with activation.

Since Nottingham's experimental data were accurately taken, we would like to suggest two possible explanations for the independence of Nottingham's measured associated reflection coefficient from activation. One is an ambiguity in the determination of true zero field. That such an ambiguity may have been present seems plausible to us because in fitting the experimental data with computed curves of log current against retarding potential, it is difficult to choose between one computed curve and another if both horizontal and vertical shifts are allowed in the fitting. For example, over the range of retarding potentials up to at least $10kT$ the curve calculated at 1330°K from the assumption $r=\exp(-\epsilon_x/0.191 \text{ ev})$ can by proper horizontal and vertical shifts be fitted to within a maximum difference of about 0.015 on the $\log_{10}j$ scale by the curve calculated from $r=0.75 \exp(-\epsilon_x/0.3 \text{ ev})$ or by that calculated from $r=0.75 \exp(-\epsilon_x/0.191 \text{ ev})$. In this connection, consider a surface having two types of patches and let the work function difference between the patches be $\delta\varphi$. Then the associated reflection coefficient will always be the same function of $mv_x^2/2e\delta\varphi$, $mv_y^2/2e\delta\varphi$, and $mv_z^2/2e\delta\varphi$. It follows that changing $\delta\varphi$ should have essentially the same effect on the retarding potential characteristic as changing ω. Thus it appears that for different states of activation the work function differences in Nottingham's surfaces could change appreciably, perhaps by as much as 50 percent, without materially affecting the fit of the data. However, the horizontal shift required to fit the data implies a shift in the zero-field points. Since it would be difficult to improve upon the accuracy of Nottingham's measurements, this possible ambiguity in the interpretation of the experimental results can probably only be eliminated by use of an independent method of determining true zero field.

The other explanation is that for different states of activation, the distribution of work function differences between the various patches remains more or less the same even though the work function differences between two given types of patches may change appreciably. (As pointed out in Chapter II, there is evidence for work function differences even in clean tungsten of perhaps as much as one volt and Ahearn and Becker (A1) measured work function differences of about 0.7 volt in the case of thoriated tungsten ribbon.)

Possibly the true explanation of the independence of Nottingham's measured associated reflection coefficient from activation is a combination of the two above explanations. In fact, the accuracy of Nottingham's measurements is such that the zero-field uncertainty would probably not suffice to explain the results if the variation in the degree of patchiness of the surface as the activation is varied were as great in this experiment as is suggested by the electron microscope data of Ahearn and Becker (A1). However it should be mentioned that Ahearn and Becker worked with thoriated tungsten ribbons in which there may well have been a strong preferential crystal orientation (A7) different from that of Nottingham's wires. Vacuum conditions may also have been significantly different.

If the data are, in fact, represented for all states of activation by the same associated reflection coefficient, then it follows that the plot of the logarithm of the zero-field current against contact potential should be a straight line as is indicated by Nottingham's data.

Nottingham considered the effect of patches and on the basis of the normal energy approximation concluded that it was incapable of accounting for the shape of his empirical associated reflection co-

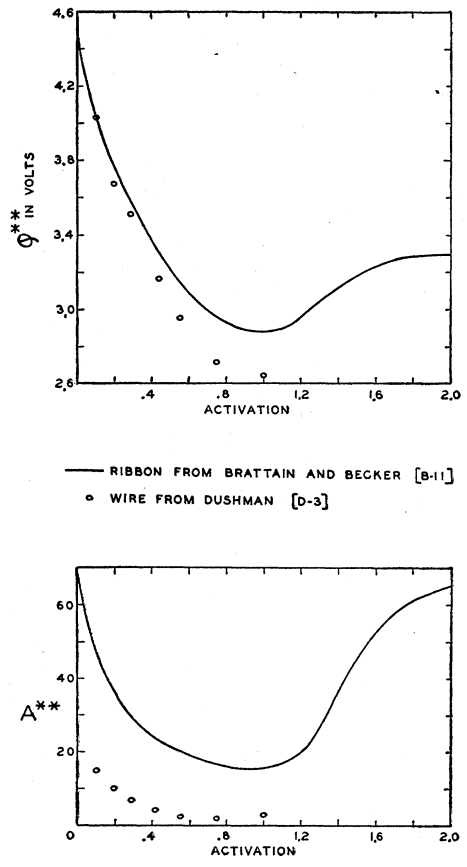


FIG. 17. Dependence of A^{**} and φ^{**} from polycrystalline thoriated tungsten wire upon activation.

efficient. As is indicated in Appendix Vb especially Fig. 33, the shape of the associated reflection coefficient, when tangential energies are taken into account, is not necessarily inconsistent with Nottingham's experimental results.

By use of the empirical associated reflection coefficient determined from the retarding potential characteristics, Nottingham was able to translate his measured zero-field currents into currents that would have been observed from the associated uniform surface without the associated reflection. Under these conditions the apparent zero-field work function should be $\bar{\varphi}-Td\bar{\varphi}/dT$, where $\bar{\varphi}$ is defined by Eq. (II.4.3). Now for clean tungsten it is reasonable to assume that $Td\bar{\varphi}/dT$ is small so that the apparent work function from such a zero-field plot should be closely $\bar{\varphi}$. From Section II.9 one would expect that $\bar{\varphi}$ for clean polycrystalline tungsten should be several tenths of a volt larger than the high field work function—i.e., about 4.7–5.0 volts. From Figs. 10 and 11 of Nottingham's paper (N7) the zero-field work function for clean tungsten corrected for reflection is about 4.75 volts. Although Nottingham points out that this value must be taken as provisional because of experimental difficulties, it is entirely consistent with the results of Section II.9.

In conclusion, it appears to the authors that the empirical reflection coefficient deduced by Nottingham must have been considerably influenced by patchiness of the emitting surfaces. How much of this reflection is caused by patches and how much by reflection at the metallic surface remains undetermined. Inasmuch as theory and the periodic Schottky effect both indicate at most a small reflection effect for clean tungsten (see Section IV.4), and since it follows from above that the effect of patches seems capable of accounting for the results, the authors believe that most of the effect must be due to patches.

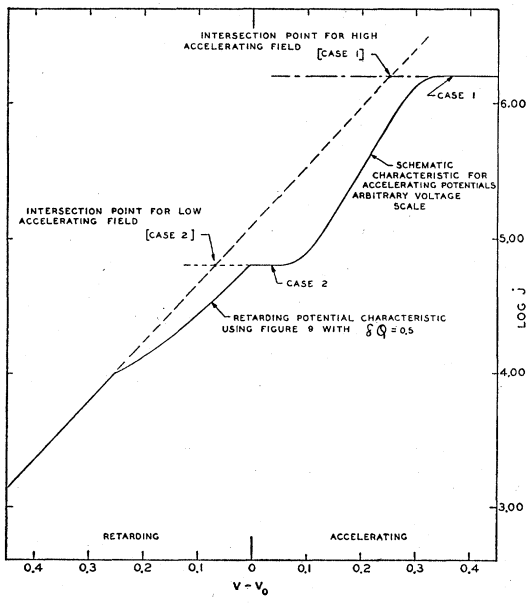


FIG. 18. Characteristics illustrating the intersection point method, using the normal energy approximation with a two-patch plane checkerboard.

III.3 The Cooling Effect Accompanying Thermionic Emission

Two papers on the cooling effect of thermionic emission from clean metals have been published recently. The first to be commented on is that of Fleming and Henderson (F4). By means of a thermocouple in the emitting cathode, they were able to measure the heat loss per electron. Since their emitter was polycrystalline, Eq. (II.10.2) should be used to relate the average calorimetric to average thermionic constants. If, lacking information as to the nature of their surface, Eq. (I.7.4) is used instead and the term involving the Thomson coefficient is neglected, the work function determined from their measured heat loss is about 4.64 ± 0.09 volts whereas their value calculated from their equation (in notation of Chapter I) $l_{\text{sat.}}=e\varphi+3.2kT$ is 4.46 ± 0.09 volts. As pointed out in Section 1.7b, the correctness of their equation has been disputed. For a metal like tungsten, it is likely that the error in neglecting the terms in Eqs. (II.10.2) and (I.7.4) involving the Thomson coefficient is not larger than several hundredths of a volt. However, for other metals, such as palladium, the effect may be several times larger [see Lander (L1)].

In addition to measuring the cooling effect of thermionic emission, Fleming and Henderson also measured the thermal effect of field emission. To within their stated sensitivity, they were unable to detect any thermal effects. On the basis of the experimentally measured and theoretically expected velocity distribution of field electrons (M25), one would expect a *heating* effect, caused by transitions of conduction electrons into the states left vacant by the field electrons, measurable with Fleming and Henderson's stated sensitivities.

The second publication is that of Krüger and Stabinow (K4) who compensated the heat loss resulting from evaporation of electrons by increasing the heating current to keep the resistance constant using a resistance bridge as the null instrument. They measured the heat loss for tungsten, tantalum, and molybdenum. Using Eq. (I.7.4) and neglecting the term involving the Thomson coefficient, Krüger and Stabinow obtained values for the φ of Eq. (I.7.4) as follows: for molybdenum a mean value of 4.40 volts, for tungsten 4.565 volts at 2100°K, and 4.600 volts at 2700°K, for tantalum 4.24 volts at 2550°K and 4.21 volts at 2050°K. In

the case of tantalum and tungsten this implies a temperature coefficient of about 0.6×10^{-4} volt/degree. Neglecting patch effects and assuming reflection effects to be small, Krüger and Stabinow calculated the value of A^* from Eq. (I.5.4) to be expected on the basis of their temperature coefficient for tungsten and tantalum. Their value which involves an error is 0.66 amp./cm^2 deg.^{2} whereas the correct value calculated from their temperature coefficient is 60 amp./cm^2 deg.^{2}. It should be pointed out, of course, that the use of Eqs. (I.7.4) and (I.5.4) is valid only for uniform surfaces. Since Krüger and Stabinow used polycrystalline surfaces, Eq. (II.10.2) should be used to relate the calorimetric to the thermionic work functions (H8). On the basis of the work function differences observed in polycrystalline tungsten surfaces, at least an appreciable part of Krüger and Stabinow's temperature coefficient can be accounted for by use of Eq. (II.10.1) even if the $l_{\text{sat.}}^{(i)}$ are assumed independent of temperature. As before, the neglect of the terms involving the Thomson coefficient probably involves an error of not more than several hundredths of a volt for these metals.

III.4 Determination of the Temperature Coefficient of the Work Function by the Temperature Dependence of the Contact Potential

Two papers on the temperature dependence of the contact potential of clean tungsten have appeared since the last review. The first of these is by D. Langmuir (L2). His method of determining contact potential differences was that of Section II.11c and his currents were not space charge limited. With a tube which showed contamination effects, Langmuir found a negative temperature coefficient for clean polycrystalline tungsten of about 4×10^{-5} volt/deg. Langmuir also studied the contact potentials of thoriated tungsten. In connection with Section III.2 it is of interest that when he plotted the logarithm of his zero field currents for different states of activation against the contact potential he obtained a straight line in agreement with Nottingham's results (N7). However, it is not obvious that Langmuir's method of adjusting the potential of his central guard ring gives the correct experimental results—see Fig. 3 of Langmuir's paper.

Potter (P3) has used the Kelvin method to determine the contact potential between two polycrystalline tungsten wires which could be held at different temperatures. His results are also subject to contamination effects but he deduced a mean positive temperature coefficient of the work function of about 6.5×10^{-5} volt/deg. Potter points out that this temperature coefficient, as opposed to Langmuir's, can account for the discrepancy between the theoretical A value of 120 and the experimental A^{**} value of about 60. As pointed out by Nottingham (N9), the patch effect alone can account for this discrepancy in the case of tungsten.

In both of the above papers the thermoelectric e.m.f. (see Section I.6) has been neglected. The effect of this neglect has been discussed by Herring (H8) and an approximate correction would change Potter's result to about 5×10^{-5} volt/deg. and Langmuir's result to about -5.5×10^{-5} volt/deg.

III.5 Miscellaneous Publications and Remaksr

Haberlund and Walcher (H1) and later Winkler (W18) have investigated the thermionic emission from platinum, nickel, and iron at temperatures as low as 700°K by making use of ion multiplication devices similar to Geiger counters. Haberlund and Walcher used hydrogen and Winkler used air as filler gases. If this method were developed to use a sufficiently purified rare gas, it would no doubt be a useful tool in emission studies.

The periodic Schottky effect reported by Seifert and Phipps (S9), Turnbull and Phipps (T5), and Nottingham (N8) has been interpreted by Mott-Smith (M18) and Guth and Mullin (G8) in terms of reflection effects on emitted electrons. Since this is of considerable theoretical importance it is discussed at length in Sections IV.4d and IV.5.

Three papers on the work function-evaporation energy cycle have appeared since the last review (F3, W19, and W4). Using work functions defined in a manner analogous to Eq. (I.2.1) these authors investigated the relationship $\varphi_-+\varphi_+=\varphi_0+I$ where φ_- is the electronic work function for a uniform surface, φ_+ the positive ion work function for a uniform surface, φ_0 the evaporation work function for neutral atoms for a uniform surface, and I is the ionization potential. Although they worked with polycrystalline surfaces and used the apparent work functions $\varphi_-{}^{**}$, $\varphi_+{}^{**}$, and $\varphi_0{}^{**}$, obtained from Richardson type plots (see Section II.7a), some measure of agreement was obtained. Fiske (F3) working with tantalum found that the cycle failed to close by 1.1 ev which he states is outside his experimental error. Wright (W19) working with molybdenum reported that his cycle closed to within 0.1 ev.

Using the method of Section II.11e, Oatley (O2) has investigated the effect of bombarding platinum anode surfaces with ions of argon and oxygen. With a suitable bombardment schedule, he obtained surface average work functions for the platinum anode surface which agreed numerically with the high field thermionic work function reported in Table IV, entry 37. This may indicate that suitable ion bombardment is an effective way of cleaning surfaces and may be useful in investigating surfaces of the less refractory metals.

CHAPTER IV. ATOMIC THEORIES OF THERMIONIC CONSTANTS AND PROCESSES

In this chapter a review will be presented covering the applications of the modern quantum theory to the calculation of quantities relating the thermionic properties of clean metal surfaces. Some of these quantities, the work function in particular, are equally closely related to other fields, such as photoelectric emission; however, the mechanisms of non-thermionic processes such as photoelectric and field emission will not be covered.

Among general references in the field of this chapter may be mentioned especially, the article of Sommerfeld and Bethe (15), the book of Seitz (12), and a review article by Nordheim (N5).

IV.1 Value of the Inner Work Function at $T=0$

The arguments to be sketched in this section constitute an attempt to construct a quantitatively valid extension of the often-used energy level diagram of Fig. 22. In using this diagram qualitatively it is customary to ignore the interaction of the free electrons with one another and to treat the interior of the metal as a region of constant potential; for this model the work function $e\varphi$ can be represented as the difference (W_a-W_i), where W_a/e is the difference in electrostatic potential between the inside and the outside of the metal, and W_i is the kinetic energy of the highest filled level in the metal at the absolute zero. This picture requires but little modification to take account of the non-uniformity of potential inside the metal, provided the mutual interactions of the electrons are ignored or replaced by an average "self-consistent field:" in such case we can replace W_a by the difference between the potential energy of an electron outside the metal and the energy of the lowest state for a conduction electron in the interior and replace W_i by the difference in energy between this lowest level and the highest occupied energy level. As we shall see below, however, one cannot expect to compute a good value for φ without taking account of the fact that the free electrons in the metal try to avoid one another, and hence have a smaller interaction energy than they would have if they were statistically independent. It is therefore best to give up the picture of independent electrons in individual quantum states, each making a specific contribution to the total energy, and to base all calculations on the original definition of the work function (Section I.2) which at $T=0$ reduces simply to

$$e\varphi=-(\partial U/\partial n)_v-e\Phi_a, \qquad \text{(IV.1.1)}$$

where U is now simply the energy of the lowest quantum state of a crystal containing n electrons, and where as before Φ_a is the electrostatic potential in the vacuum just outside the crystal face whose work function is being calculated.

In evaluating (IV.1.1) by means of specific theoretical expressions for U, it is customary to set $(\partial U/\partial n)_v$ equal to the change in U which would occur, per electron added, if the added electrons were uniformly distributed throughout the volume of the metal. This is legitimate, in spite of the fact that the electrons added will actually all be localized at the surface of the metal block. To see this, we may imagine the addition of the electrons to take place in two steps. First let the electrons be added together with an imaginary distribution of positive charge just sufficient to neutralize the Δn added electrons and distributed uniformly over the metal. In this step the added electrons will, of course, distribute themselves uniformly, throughout the volume of the metal. Then let the positive charge be removed, permitting the excess electrons to migrate to the surface. Now the energy change involved in the first step is just the value of $(\partial U/\partial n)_v\Delta n$, calculated for a uniform distribution of the added electrons, plus the energy of interaction of the positive charge with the metal, including the added electrons. The energy change involved in the second step will differ from the value of this latter interaction energy only by a slight amount, due to the rearrangement of the Δn electrons while the positive charge is being removed. This difference will be of the order of magnitude of $(e\Delta n)^2$ divided by the linear dimension of the metal block and may therefore be neglected as $\Delta n\to 0$.

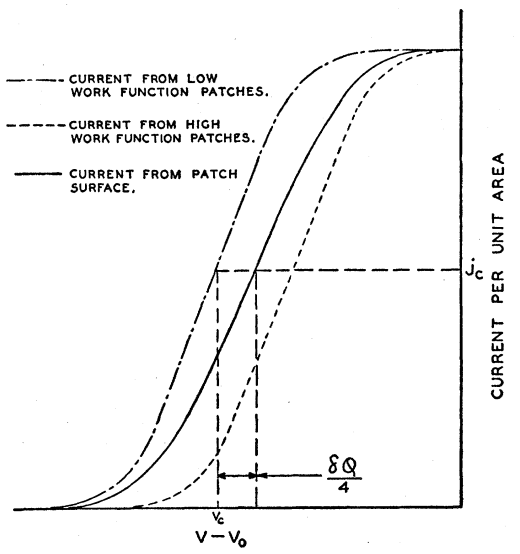


FIG. 19. Characteristics illustrating the electron beam method, using the normal energy approximation with a two-patch checkerboard.

IV.1a *The Method of Wigner and Bardeen*

An evaluation of the work function from (IV.1.1) can of course, be based on any theory of metallic

binding which permits the energy of the metal to be calculated for an arbitrary state of charge. The only such theories which have been really successful to date, however, are the various modifications of that developed by Wigner and Seitz, which have been applied by them and others to metals at the left side of the periodic table. The application to the calculation of the work function is due to Wigner and Bardeen (W15). Without attempting to discuss the details of this method, which have been fully expounded elsewhere (e.g., (12), Chapters IX and X), we may state its philosophy briefly thus: A self-consistent field solution is determined for the electrons of the metal, and its energy is evaluated. Then the assumption is made that the difference between the actual energy of the ground state of the metal and that of the self-consistent field solution is practically the same as it would be for a gas of free electrons moving in a constant external potential and having the same average density as the free electrons in the metal. An approximate value of this difference, or "correlation energy," has been calculated for the latter case by Wigner (W12, W14). The principal limitation of the theory is that both in the correlation energy and in the Coulomb and exchange terms to be discussed below it makes use of the assumption that the wave functions of the free electrons resemble plane waves over the greater part of the volume of each unit cell. This assumption appears to be justified for metals on the left side of the periodic table, but is certainly not valid for transition metals. The analysis to follow will therefore apply only to metals of the former type.

The theory just outlined may be formulated mathematically in several ways, all essentially equivalent to one another. For our purposes the most convenient formulation is the one most closely akin to that used for self-consistent fields in atoms. Let the quantities ϵ_i by the energy parameters for a single free electron moving in the field of the ion cores plus the Coulomb field of all the electrons. Then the total energy of a block of metal at $T=0$, relative to the state of separated ion cores and electrons, is

$$U=\sum \epsilon_i - \text{Coulomb energy} - \text{exchange} - \text{correlation} + E_I, \quad \text{(IV.1.2)}$$

where E_I is the interaction energy of the ion cores with one another, which at any fixed volume may be assumed independent of the population of the free electron levels. Note that the classical Coulomb self-energy of the distribution of free electrons comes in with a negative sign, because twice this energy has already been included in $\sum\epsilon_i$.[m] Explicit

[m] The form (IV.1.2) corresponds to the Hartree type of calculation, rather than to the Slater-Fock type in which the one-electron wave functions are eigenfunctions of a Hamiltonian which includes exchange. If the latter method were used the exchange term would be added instead of subtracted. For monovalent metals, at least, the eigenfunctions and hence the final energy are nearly the same for both methods.

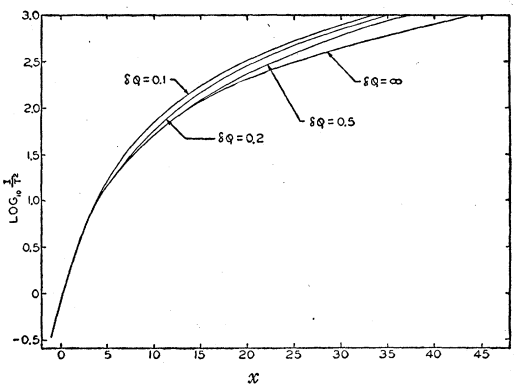


FIG. 20. Calculated Fowler plots for a two-patch checkerboard surface for various work function differences between the patches, with high collecting fields.

expressions for the exchange and correlation energies may be given if we assume them the same as for a free electron gas of the same density. The former is ((12), #75)

$$\text{exchange energy per electron} = 0.458e^2/r_e, \quad \text{(IV.1.3)}$$

where $r_e=(3/4\pi n_1)^{\frac{1}{3}}$ is the radius of a sphere whose volume equals the volume of the metal per free electron. The numerical results of Wigner (W12) on the correlation energy can be represented, over the range of electron densities which occur in metals, by the expression

$$\text{correlation energy per electron} = 0.29e^2/(r_e+5.1a_H), \quad \text{(IV.1.4)}$$

where a_H is the Bohr radius $\hbar^2/me^2=0.531$A. Multiplying these by the number n of free electrons, inserting into (IV.1.2), and differentiating with respect to n gives the value of $\partial U/\partial n$ to be inserted into (IV.1.1). The derivative of the Coulomb energy with respect to n, i.e., the change in this quantity when an electron is added at the top of the Fermi distribution, is easily shown to be equal to the change in the Coulomb potential averaged over the original electronic charge distribution. Now if we assume the wave functions of the states originally occupied to change only negligibly with addition of the new electron, the quantity $\sum\partial\epsilon_i/\partial n$ will be just the sum over all occupied states of the change in Coulomb potential times the charge distribution of the ith state, hence will exactly cancel the change in the Coulomb energy term. This

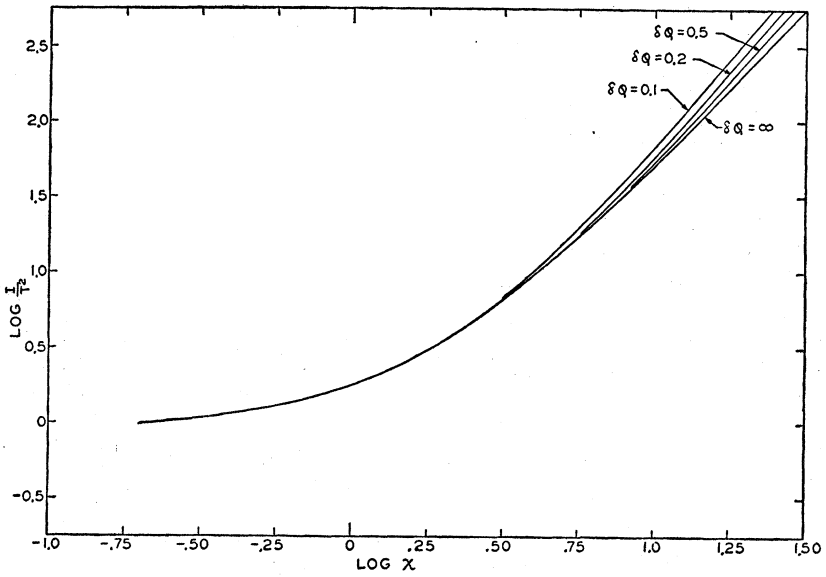


FIG. 21. Calculated DuBridge plot for a two-patch checkerboard for various work function differences between the patches with high collecting fields.

being the case we obtain

$$e\varphi = -\left(\frac{\partial U}{\partial n}\right)_v - e\Phi_a = (-e\Phi_a - \epsilon_{\max}) + \frac{0.611e^2}{r_e} + \frac{0.29e^2}{r_e+5.1a_H} + \frac{0.10e^2r_e}{(r_e+5.1a_H)^2}. \quad \text{(IV.1.5)}$$

Here the first term represents the energy difference between the top of the Fermi distribution and the potential energy $-e\Phi_a$ of an electron just outside the surface, and thus corresponds to the $(W_a - W_i)$ of the simple picture of Fig. 22. This term depends, of course, on the magnitude of the double layer at the surface of the metal. The remaining terms of (IV.1.5) are the exchange and correlation corrections.

A convenient alternative expression for the work function of a metal, based on the same assumptions, can be obtained by subtracting $-U/n$ from $e\varphi$, i.e., by combining (IV.1.2) and (IV.1.5). The resulting expression can be made more useful by noting that the difference (E_I — Coulomb energy) occurring in (IV.1.2) can be reduced to a summation over the various atomic cells of the metal crystal, in which most of the terms are negligibly small. This is a consequence of the fact that in all the simple types of crystals each atom can be surrounded by a cell of almost spherical shape. If we consider any two of these, the electrostatic interaction of the ions in these two cells would be exactly equal to that of the electronic charges if the cells were spherical and non-overlapping, so to a good approximation the electrostatic contribution to (E_I — Coulomb energy) will vanish. If this vanishing could be assumed for every pair of cells, the only contributions remaining would be the non-electrostatic interactions of the ion cores, as a result of van der Waals and repulsive forces, and the electrostatic self-energy of the electron distribution in each single atomic cell. If we approximate the cell by a sphere of equal volume filled with a uniform distribution of free electron charge, the latter is

$$\text{self-energy of spherical cell} = -0.6e^2X^2/r_s, \quad \text{(IV.1.6)}$$

where r_s is the radius of the sphere whose volume equals the atomic volume, and X is the number of valence electrons per atom. A more elaborate calculation shows that for metals on the left side of the periodic table this approximation is quite a good one for the contributions from all the pairs of cells in the interior of the metal (W16). At the surface, however, we may expect the charge distribution around each ion to be asymmetrical, so that the surface cells will have a dipole moment. The effect of this on the term (E_I — Coulomb energy) can readily be computed directly; however, we may obtain the correct final formula for the work function more easily by first combining (IV.1.2) and (IV.1.5) under the assumption that the surface cells are identical with the interior cells, and then using the fact established earlier (Section I.2) that addition of a dipole moment of magnitude $D/4\pi$ per unit area to the surface changes the work function by D. This gives finally

$$e\varphi = -\frac{U}{n} - (\epsilon_{\max} - \epsilon_{Av}) + \frac{E_I'}{n} - \frac{0.6e^2X}{r_s} + \frac{0.153e^2}{r_e} + \frac{0.10e^2r_e}{(r_e+5.1a_H)^2} + eD, \quad \text{(IV.1.7)}$$

where ϵ_{Av} is the average energy of the occupied free

electron states and where E_I' is the van der Waals and repulsive energy of the ion cores, which is usually negligible for metals to which the present method applies. Here the quantity $-U/n$ is simply the energy per electron required to decompose the metal into ion cores of charge Xe and free electrons; it can be therefore be evaluated empirically from the heat of sublimation of the metal and the ionization potentials of the atoms. The quantity $(\epsilon_{max}-\epsilon_{Av})$ depends upon the distribution of levels in energy; if this distribution is a parabolic one corresponding to an effective mass m^*, an approximation which should be fairly good for monovalent metals, we have ((12), #26)

$$(\epsilon_{max}-\epsilon_{Av}) \approx \frac{1}{5}\frac{h^2}{m^*}\left(\frac{3n_1}{8\pi}\right)^{\frac{2}{3}} = \frac{0.737 a_H e^2}{r_e^2}\left(\frac{m}{m^*}\right). \quad \text{(IV.1.8)}$$

The remaining terms in (IV.1.7) are likewise easily computed functions of the density, except the term eD, which depends upon the wave functions of electrons near the surface (see Section IV.2). Let it be emphasized again that (IV.1.7) is expected to apply only to metals on the left of the periodic table.

IV.1b *The Modification of Hellmann and Kassatotschkin*

An interesting simplification of the method just outlined has been proposed by Hellmann and Kassatotschkin (H6, H7). These authors calculate binding energies, work functions, etc., for a number of metals at the left of the periodic table, by a method which differs from that of Wigner and Seitz only in that the energies of the various states of the conduction electrons are obtained in a different and simpler way. In the Wigner-Seitz method one integrates the wave equation for the conduction electrons in the field of the ion cores and obtains wave functions which have, generally, several nodal surfaces near each nucleus; these nodes are necessary if the wave functions of the conduction electrons are to be orthogonal to those of the core electrons. Hellmann and Kassatotschkin use wave functions for the conduction electrons which have no such nodes near the nuclei, and which can therefore be approximated fairly well by single plane waves; they show that it is possible to find an empirical ion core field with the property that the energies of such "nodeless" wave functions should agree rather well with the energies of the wave functions calculated by the Wigner-Seitz method. Their procedure beings, therefore, by using a trial-and-error method to find an empirical field for an isolated metal ion, in which the $1s$ eigenfunction will have the same energy as the lowest s eigenfunction for a valence electron attached to the ion (e.g., the $4s$ for potassium), the $2p$ the same energy as the lowest valence electron p state (e.g., the $4p$ for potassium), the $2s$ the same energy as the next lowest valence electron state (e.g., the $5s$ for potassium), etc. With an empirical field of this sort surrounding each ion of the metal, the energies of the eigenfunctions for the conduction electrons are easily calculated, at least to a rough approximation, and various energetic properties of the metal can be obtained, including the work function from (IV.1.5) or (IV.1.7).

IV.1c *Numerical Calculations*

Table VI gives values of $(\varphi-D)$ which have been calculated from (IV.1.7) and (IV.1.8), using either the usual Wigner-Seitz model or the Hellmann-Kassatotschkin (H-K) modification. The rows for which the column headed "Reference" is left blank have been calculated by one of us (C.H.) using the assumptions given in the "Remarks" column; since the values of m/m^* used here for Li and Na are probably better than those used by Wigner and Bardeen, the latters' values for these metals are not given. The accuracy which the calculated values may be expected to have can be gauged roughly by comparing the second and third columns, since the difference between the figures in these two columns measures the difference between the observed binding energy[n] and that computed from (IV.1.2). However, it must be remembered that the various assumptions and approximations of the theory enter into the work function and the binding energy in rather different ways. The reliability of the calculated values of m^* or its equivalent varies from case to case, the most reliable values being those in the first two rows of Table VI. The last column lists the values which would have to be assumed for the moment D of the surface double layer to make the work function computed from the observed binding energy agree with the observed work function. While it is unlikely, especially in the extreme case of Be, that the entire discrepancy between the third and sixth columns of the table is due to the double layer alone, the figures given suggest strongly that D increases rapidly with increasing density of conduction electrons and perhaps with increasing chemical valence as well. We shall see in Section 4.2a that Bardeen's theoretical calculation of D agrees roughly with the value in Table VI for sodium; unfortunately, no comparable calculation has been carried out for a higher density of electrons.

[n] The values used are those compiled by Seitz ((12), Chapter I).

TABLE IV. Experimentally determined emission constants from clean polycrystalline metals.

Entry No.	Element	Photoelectric work function	Work function obtained from contact potential	Thermionic φ^{**}	Thermionic A^{**}	Reference	Comments
1	Al	4.39				G1	Evaporated onto glass surface. Probably contaminated. Determined threshold by inspection. Probably Case 2 collecting fields.
2			4.08			B22	Contact potential against sodium determined by retarding potential curves. Sodium work function taken to be 2.40.
3	Ba	2.48				M11	Evaporated onto glass surface. Probably Case 2 collecting fields. Determined by Fowler plot.
4		2.49				C4	Evaporated onto nickel plate. Probably Case 2 collecting fields. Determined by Fowler plot.
5		2.51				J1	Evaporated onto glass plate. Probably Case 2 collecting fields. Determined by Fowler plot.
6			2.39±0.05			A2	Evaporated onto tungsten ribbon and determined against tungsten by method of Section II.11d. *See note 1.*
7	Be	3.92				M7	Evaporated onto a metal surface. Probably Case 2 collecting fields. Determined by Fowler plot.
8	Bi	4.22 to 4.25				J7	Evaporated onto glass surface. Probably Case 2 collecting fields. Determined by Fowler plot.
9		4.46				W8	Evaporated onto glass surface. Determined by DuBridge method. Possibly contaminated. Probably Case 2 collecting fields.
10		4.76 to 5.06				W7	Evaporated onto glass surface. Determined by DuBridge method. Probably Case 2 collecting fields. Probably contaminated.
11		4.32				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.
12	Ca	2.71				J1	Evaporated onto glass surface. Probably Case 2 collecting fields. Determined by Fowler plot.
13		3.21				L8	Evaporated onto tantalum sheet. Determined by inspection from energy distribution curve.
14	C			4.34	30	R4	Cylindrical filament. Collecting fields probably Case 1.
15				4.39	15	B24	Specimen heated in oven the walls of which served as anode.
16	Cr			4.60	48	W2	Round filament prepared by plating a brass tube. Case 1 collecting fields.
17		4.37				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.
18	Co			4.41±0.10	41	W3	Flat strips rolled before heat treatment. Probably Case 1 collecting fields.
19	Cb			4.01	37	R6	Wire. Case 1 collecting fields.
20	Fe			4.48±0.06	26	W3	Below $\beta-\gamma$-transition. Strip filament. Probably Case 1 collecting fields.
21				4.21±0.05	1.5	W3	Above $\beta-\gamma$-transition. Strip filament. Probably Case 1 collecting fields.
22		approx. 4.63				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.

IV.1d *Use of the Fermi-Thomas Model*

A number of papers have been published on the problem of computing the work functions of metals by use of the Fermi-Thomas statistical model. As an excellent summary of this work up to 1933 is given in the article of Sommerfeld and Bethe ((15), #14 and #15), only a brief discussion will be given here. Although some of these calculations are interesting and illustrate the basic physical principles which influence the value of the work function, we do not believe that one can expect quantitative accuracy from theories of this kind, since the conditions for validity of the Fermi-Thomas method

TABLE IV.—*Continued*

Entry No.	Element	Photoelectric work function	Work function obtained from contact potential	Thermionic φ^{**}	Thermionic A^{**}	Reference	Comments
23	Mg	3.60				C5	Evaporated onto molybdenum plate. Probably Case 2 fields. Determined threshold by inspection.
24		3.67±0.02				C3	Evaporated onto glass. Probably Case 2 collecting fields. Determined by Fowler plots.
25		3.79				C3	Evaporated onto nickel. Probably Case 2 collecting fields. Determined by Fowler plots.
26		3.68				M7	Evaporated onto glass surface. Determined by Fowler plot. Probably Case 2 collecting fields.
27			3.78			A4	Evaporated onto glass plate and determined against barium by method of Section II.11d using average work function of barium of 2.52 volts.
28	Mo			4.20±0.02	55	W19	Strip filament cut from specially purified sheet metal. Probably Case 1 collecting fields.
29				4.37±0.02	115	W19	Commercial wire. Probably Case 1 collecting fields. *See note 2.*
30				4.17	51	W4	Strip filaments presumably cut from sheet metal. Probably Case 1 collecting fields. *See note 3.*
31		approx. 4.35				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.
32				4.33	24.6	F10	Wire. Probably Case 1 collecting fields.
33		4.12				C5	Sheet. Possibly Case 2 fields. Threshold determined by inspection.
34				$l^{**}/e=4.40$		K4	Determined calorimetric work function l^{**} of Eq. (II.10.1). Wire. Case 1 collecting fields. See Section III.3.
35	Ni			4.61±0.05	30	W3	Ribbon filaments. Probably Case 1 collecting fields.
36		4.87				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.
37	Pt			5.32	32	W10	Wire and wire rolled into ribbon. Probably Case 1 collecting fields. *See note 4.*
38		greater than 6.2				R7	Sputtered onto glass surface. Probably Case 2 collecting fields. Some contamination likely. Threshold determined by inspection.
39	K	2.26±0.02				M12	Evaporated onto platinum foil. Probably Case 2 collecting field. Threshold determined by inspection.
40	Rh			4.80	33	W6	Strip filament. Probably Case 1 collecting fields.
41	Si			3.59	8	B24	Specimen heated in oven, the walls of which served as anode.
42	Ag		4.33±0.05			A3	Evaporated onto glass plate and determined against barium by method of Section II.11d using average work function of barium of 2.39 volts.
43			4.47			A6	Evaporated onto glass plate and determined against barium by method of Section II.11d using average work function of barium of 2.52 volts. Therefore agrees well with previous entry. See also entry 5 in Table V.

are usually not adequately fulfilled. Moreover, most of these calculations make no attempt to separate the inner work function from the contribution of the double layer; since we know that the work function is different for different crystal faces and is modified by adsorbed films, no theory can be really acceptable which attempts, as these the theories usually do, to express the work function of a metal entirely in terms of volume properties such as binding energy or Fermi energy.

If the Fermi-Thomas model is used in its simplest form, with neglect of exchange and correlation

TABLE IV.—*Continued*

Entry No.	Element	Photoelectric work function	Work function obtained from contact potential	Thermionic φ^{**}	Thermionic A^{**}	Reference	Comments
44	Na	2.29				M7	Evaporated onto glass surface. Probably Case 2 collecting fields. Determined from Fowler plot which indicates patch effects.
45		2.28				M11	Evaporated onto glass surface. Probably Case 2 collecting fields. Determined from Fowler plot.
46		2.40				B22	Evaporated onto aluminum. Probably Case 2 collecting fields. Determined by Fowler plot that did not fit very well.
47	Ta			4.10	37	C1	Shape of filament not specified. Probably Case 1 collecting fields.
48				4.19±0.02	55±5	F3	Wire slightly flattened by rolling. Case 1 collecting fields.
49		4.05				C5	Sheet. Possibly Case 2 collecting fields. Threshold determined by inspection.
50				$l^{**}/e=4.24$ at 2550°K		K4	Determined the calorimetric work function l^{**} of Eq. (II.10.1). Wire. Case 1 collecting fields. See Section III.3.
51	W			4.55	75	W5	Wire. Probably Case 1 collecting fields. *See note 5.*
52				4.52	72	R5	Wire. Case 1 collecting fields.
53				4.45	38	N2	Prewar "non-sag" wire polished smooth. Case 1 collecting fields. *See note 6.*
54				4.53	22	F10	Wire. Probably Case 1 collecting fields.
55		4.49±0.02				A7	Ribbon. X-ray analysis showed that most crystallites were oriented with (100) directions within 10° of the perpendicular to the ribbon surface. Case 1 collecting fields. Determined by Fowler and DuBridge plots.
56				$l^{**}/e=4.60$ at 2700°K		K4	Determined the calorimetric work function l^{**} of Eq. (II.10.1). Wire. Case 1 collecting fields. See Section III.3.
57				$l^{**}/e=5.01$ at 2000°K		F4	Determined the calorimetric work function l^{**} of Eq. (II.10.1). Wire. Case 1 collecting fields. See Section III.3.
58				4.52	60	N6a	Prewar "non-sag" wire. Case 1 collecting fields.
59	U			3.27±0.05	about 6	H13	Strip and oval filament. Case 1 collecting fields. Purity of specimens questionable.
60	Zn		4.28±0.02			A5	Evaporated onto glass plate and determined against barium using average work function of barium of 2.52.
61		4.24				D2	Evaporated onto molybdenum strip. Collecting fields uncertain.

Note 1: The value 2.39 for the work function of barium was obtained by subtracting the difference in the potentials V_c corresponding to the reference current j_e (see Section II.11d) from the thermionic φ^{**} of tungsten which was taken to be 4.52. It follows from Section II.11d that the surface average $\bar{\varphi}$ would have been more correct. Since $\bar{\varphi}$ is somewhat larger than φ^{**} the value 2.39 is not necessarily inconsistent with entries 3, 4, and 5.

Note 2: This author states that he was unable to reproduce the increase in work function caused by rolling as reported by Wahlin and Reynolds (see reference W4). See note 3.

Note 3: The values of φ^{**} of 4.17 and A^{**} of 51 are from two samples of pure molybdenum from the Fansteel Company. These were about 1 mm wide and 0.05 mm thick presumably cut from sheet metal. A strip of Mo obtained from another source was cold rolled to uniform thickness and gave φ^{**} of 4.38 and A^{**} of 175. A sample of the Fansteel metal was cold rolled lightly and gave φ^{**} of 4.30 and A^{**} of 96. The authors were unable to detect by spectroscopic analysis any difference in the composition of the specimens. The authors suggested that the variation in values may be due to differences in crystal orientation but did not investigate the matter further.

Note 4: Two platinum wires were rolled flat, one before heat treatment and one after. No appreciable differences between the round and flat filaments were detected.

Note 5: After the specimen was flashed to 3100°K the values rose to 4.63 and 212. This effect is not explained although the authors suggest that it may be caused by a change in the crystalline nature of the surface.

Note 6: Reworking of the data revealed numerical errors in the values reported in the reference. The values given in the table are the corrected ones.

effects, the calculated work function φ can easily be shown to be zero, for the same reason that the calculated ionization potential is zero when this approximation is applied to a free atom. For the boundary of the electronic charge distribution is the locus on which the electrostatic energy $-e\Phi$ equals the energy $E_{\max}$ of the highest occupied electronic state. If there is to be no charge on the metal, this equipotential surface must have the same potential as the surface at infinity, i.e., zero, and therefore $\varphi=E_{\max}=0$. More refined methods of calculating φ can be developed, however, based on (IV.1.1) and involving an expression for the total energy U which takes some sort of account of the fact that the electrons are not statistically independent of one another. One such method, due to Tamm and Blochinzev[o] and to Fröhlich (F13), is based on the

[o] We have been unable to locate any account of this work other than that given in the article of Sommerfeld and Bethe ((15), #15).

assumption that the Fermi-Thomas model gives the correct density and mean kinetic energy for the electrons. From this the total energy can be derived by using the virial theorem and turns out to be different from the value one would calculate directly from the Fermi-Thomas potential, treating the electrons as statistically independent. The authors mentioned conclude that the work function should be essentially equal to the Fermi energy W_i. A more satisfactory method is one due to Bartelink (B4) and further developed by Mrowka and Recknagel (M21). These authors calculate the energy required to remove an electron of zero kinetic energy from an idealized metal consisting of electrons plus homogeneously distributed positive charge; they than assume that the work function is obtained simply by subtracting from this energy the maximum kinetic energy of the electrons as given by the usual Sommerfeld formula. In carrying out the calculation they take account of the statistical correlation between electrons by recognizing that the electron being removed is surrounded, as long as it is in the metal, by a region in which there is a deficiency of other electrons. This deficiency is calculated by solving the Fermi-Thomas equation for the metal with a negative point charge of strength e at a certain point inside it. The contribution which this polarization effect makes to the energy of removal is then obtained by removing the charge e in infinitesimal units dq and reassembling them outside the metal, taking the work involved in removing each dq as the difference in electrostatic potential between the interior and exterior positions.

It will be recognized that the procedure just described amounts to an approximate calculation of the correlation energy of a free electron gas, although the references just cited use the computed polarization energy as if it represented the sum of the exchange and correlation energies. The limitations of this method of calculation can be seen by comparing the results with Wigner's wave mechanical calculation of the correlation energy (W12, W14): over the range of electron densities ordinarily encountered in metals the equations of Mrowka and Recknagel (M21) give values of the correlation energy which are several times too high, while at high electron densities they yield a correlation energy approaching infinity, whereas Wigner's calculation indicates that the correlation energy should approach a finite limit. These results are to be expected, since the statistical model neglects the wave mechanical penetration of one electron into the classically inaccessible region close to another.

IV.1e *Quasi-Empirical Theories*

A number of papers have been published attempting empirically to correlate work functions with such quantities as atomic number, atomic volume, valence, compressibility, etc. of which we attempt to cite only the more recent ones (R16, R17, B19, B20, C6). The subsequent development of reasonably sound quantum mechanical theories of the work function, as outlined above, has decreased the significance of such attempts; moreover, the empirical data on which they are founded are very unreliable, because so few metals can be obtained with clean surfaces.

IV.2 Theory of the Surface Double Layer

In the two following subsections we shall be concerned with the theoretical estimation of the double layer moment D at the surface of a metal at the absolute zero. We define D, just as in Section IV.1a, as 4π times the dipole moment per unit area due to the deviation of the charge distribution of the surface atoms from the symmetrical form found in each unit cell of the interior. The unit cell can of course be drawn in a variety of ways; however, the definition of D can be freed from ambiguity by specifying that the unit cell be drawn in such way as to have no dipole moment (see the discussion in Section IV.3b below).

To calculate D rigorously would be a formidable problem in wave mechanics, and so existing theories have been based on radical simplifications. The force field in which the electrons move has always been assumed one-dimensional, except in the calculations of Smoluchowski (S15), discussed in Section IV.2b, which are, however, based on almost equally idealized assumptions. The methods which have been used can be divided into those involving calculation of different wave functions for all the different electrons of the metal and statistical methods in which only the total electronic charge density enters the calculation.

Before starting to discuss detailed calculations on the double layer, it will be worth while to say a few words about the physical principles which determine its magnitude. Figure 23 shows how it is natural to expect the electronic charge distribution to spread out beyond the limits to which it would be confined if the distribution in each unit cell about the surface atoms were the same as in the unit cells of the interior. It will be seen that this spreading out produces a dipole moment $D/4\pi$ per unit area, whose sign may correspond to having negative charges on the outside, positive on the inside, as shown in B and C of the figure, or, in extreme cases, to positive charges outside and negative inside, as shown in D. This spreading out of the electron distribution occurs because, as is to be expected quantum-mechanically, it lowers the kinetic energy of the electrons; however, it is limited by the increase of potential energy which accompanies it. This increase of potential energy is

TABLE V. Emission constants from single crystal surfaces.

Entry No.	Element	Crystal direction normal to surface	Photoelectric work function	Work function from contact potential	Reference	Comments
1	Cu	111	4.89		U1	Measurements troubled by thermal etching and contamination. Thresholds determined by inspection. Probably Case 2 collecting fields.
2		100	5.64		U1	
3	Ag	111	4.75±0.01		F2	Difference checked to 0.01 volt by contact potential measurement. Determined by Fowler plots. Probably Case 2 collecting fields. Contamination and deviation from desired surface structure encountered.
4		100	4.81±0.01		F2	
5		100		4.79	A6	Evaporated onto (100) surface of rock salt and determined against barium by method of Section II.11d using average work function of barium of 2.52. No studies of surface structure of specimens made.
6	W	112	4.50		M13	Measurements troubled by instability of surfaces particularly the 310 surface. Determined by Fowler plots. Probably Case 2 collecting fields.
7		310	4.30		M13	

partly a result of the fact that in the spread-out configuration, B, of Fig. 23 the electrons are on the average in positions of higher electrostatic potential energy than in A. But there is an important additional, though less obvious, increase in potential energy, namely, that due to the exchange and correlation effects. The average interaction energy of an electron with all the other electrons is less than the Coulomb energy because the electrons avoid one another, and this decrease in interaction energy is the more marked the higher the electron density (see Eqs. (IV.1.3) and (IV.1.4)). Thus in Fig. 23B, where the electron density is lower than in A, the electron interaction energy will be higher. It is interesting to note that it is just this statistical correlation between the positions of electrons which gives rise to the image force $e^2/4x^2$ which acts on an electron at large distances outside the surface, and that the exchange and correlation energies of an electron in the interior can in a certain sense be identified with the total work done against forces of the image type in getting the electron out of the metal. This subject will be discussed further in the following subsection.

IV.2a *Wave Mechanical Theories*

The first attempt to calculate wave mechanically the electron distribution near a metal surface seems to be that of Tamm and Blochinzev (T1, T2). These authors attempted to approximate the self-consistent field solution for an idealized metal by determining the different wave functions for an electron in a potential field containing one or more parameters, these parameters being varied until the total energy of the determinantal wave function of all the electrons was minimized. This procedure, if followed through with a sufficient number of parameters in the potential field, could probably be made to give a fair approximation to the solution of the Slater-Fock self-consistent field equations, although the approximation could not be made arbitrarily good since the wave functions in the latter system are not eigenfunctions in any single potential field. The actual calculations which Tamm and Blochinzev carried through were, however, based on a box-like potential function having as its only adjustable parameter the height of the potential jump at the boundary. The metal they assumed to consist simply of a continuously distributed positive charge, of constant density ρ_+ for $x<0$ and bounded by the plane $x=0$, together with sufficient electrons to make the metal neutral. With this model and with electronic wave functions satisfying the Schrödinger equation in a potential field $U(x)=0$ for $x<s$, $U(x)=W_a$ for $x>s$, they calculated the total energy of the metal as a function of W_a, choosing s for each W_a at the proper value to make the surface neutral. The total energy was found to decrease monotonically with decreasing W_a,[p] and as the minimum possible value of W_a is W_i, the kinetic energy of the fastest electrons in the interior, it was concluded that the surface double layer must be such as to produce a potential jump D equal to W_i. It was further concluded that no self-consistent field solution is possible at the metal surface. These conclusions are refuted by the more refined calculations of Bardeen (B1), which we shall now describe, and so one must conclude that Tamm and Blochinzev placed too much reliance on the oversimplified calculation which they made.

The method adopted by Bardeen (B1) was the

[p] Tamm and Blochinzev made an error in their calculation of the energy, due to their working with a semi-infinite, rather than a bounded crystal, a procedure which necessitates subtracting infinities to get the surface energy. However, correction of this error does not change the qualitative features of their calculation.

more straightforward one of constructing a solution of the Slater-Fock self-consistent field equations by trial and error, comparing the Coulomb and exchange fields due to any computed set of wave functions with the Coulomb and exchange fields assumed in the wave equation from which these wave functions were computed. Like Tamm and Blochinzev, Bardeen assumed the metal to consist of a uniform distribution of positive charge bounded by the plane $x=0$, plus sufficient electrons to neutralize it. It is not unreasonable to expect the surface dipole moment, D, calculated for this model to be roughly correct for an actual metal, provided that the wave functions of the conduction electrons inside the metal are very nearly plane waves, that the electron density in the interior of the model is chosen the same as for the metal, and that the boundaries of the surface cells fit together to form a plane surface. The first conditions is unusually well fulfilled for sodium, and so Bardeen made his numerical calculations for the electron density found in this metal. The third condition is of course not fulfilled, but a correction for its nonfulfillment can be made.

With the aid of certain reasonable approximations in the treatment of the exchange field, Bardeen was able to get an approximately self-consistent solution of the Slater-Fock equations, which yielded a double layer moment D of about 1.0 volt, as shown in the first row of Table VII. He noted, however, that one must expect the actual double layer moment of the idealized metal to be appreciably less than this, because of the correlation energy which the self-consistent field neglects. The correlation term in the inner work function is given by the last two terms of (IV.1.5), and amounts to nearly one ev for sodium. This means that, for any given charge distribution at the surface, the energy required to take an electron all the way out of the metal is actually about a volt less than the energy which would be computed from the Coulomb and exchange fields above. This being the case, one must assume that the energy required to get an electron even part of the way out of the metal will also be greater than in the self-consistent field approximation, and that therefore there will be less spreading out of the electron distribution beyond the boundary $x=0$ of the positive charge distribution. This effect is the analogue of the fact that self-consistent field calculations always give slightly too large diameters for atoms, because they neglect polarization effects. Bardeen attempted to calculate this effect quantitatively by introducing into the Schrödinger equation of each electron a "correlation potential," in addition to the Coulomb and exchange fields already present in the self-consistent field approximation. This correlation potential was so chosen as to be asymptotic to the image potential $e^2/4x$ at large distances from the surface,[q] and to approach a constant value in the interior, the constant value being chosen differently for electrons of different energies, in analogy with the known velocity dependence of the exchange energy of free electrons (see (12), #75), but in such a way as to give the correct value for the average correlation energy for all the electrons in the interior. With these modified wave equations a modified self-consistent field solution was obtained, which yielded a value of 0.4 volt for D, as shown in the second row of Table VII. Comparison with the value of 1.0 volt obtained without the correlation effect shows that correlation plays a vital part in limiting the spreading of the electron distribution, and cannot be neglected in any quantitative calculation.

Table VII shows also the values of the work function φ for this idealized metal with a continuous positive charge distribution, calculated with and without correlation effects. In the first row the correlation terms of (IV.1.5) have been omitted from the inner work function, as well as in the calculation of D, while in the second column they have been included. It will be noticed that the correlation affects D and the inner work function in opposite directions.

To apply these results to the actual metal sodium, D must be corrected for the fact that the unit cells forming a crystal face of sodium would not have a common plane boundary but would have a pebbly or roofed appearance, as indicated in Fig. 23. Bardeen calculated the correction for the (110) face on the assumption that the electron distribution on this face is just the same as it would be for an idealized metal with a plane boundary in the position which averages the peaks and valleys of the actual surface. His result gave a value of D for sodium which was 0.25 volt smaller than for the idealized metal, as shown in the third row of Table VII. Smoluchowski's calculations, discussed in the next section, suggest that the difference may not be as great as this.

IV.2b *Theories Using the Fermi-Thomas Model*

The earlier attempts to calculate work functions from the Fermi-Thomas model were usually so formulated that the contributions from the interior and from the surface double layer were calculated together. For accounts of these earlier theories the reader is referred to the article of Sommerfeld and Bethe ((15), #15) and to the brief discussion we have already given in Section IV.1d. Mrowka and Recknagel (M21) have computed by the ordi-

[q] The relationship of the classical image potential to the quantum-mechanical correlation effect has been described in another paper by Bardeen (B3).

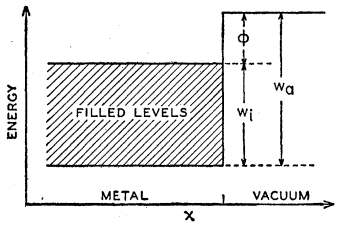


FIG. 22. Simplified energy diagram for metal and vacuum.

nary Fermi-Thomas method the electron distribution near the surface of an idealized metal consisting of electrons plus a uniformly distributed positive charge with a plane boundary. However, the results are, as one would expect, of little quantitative significance, since the potential jump D resulting from the surface double layer always comes out by this method to be equal to the maximum kinetic energy W_i of the internal electrons (see the discussion in Section IV.1d); this is over four volts for the electron density found in sodium while the actual value of D is only about 0.4 volt according to the calculations of Bardeen summarized in the preceding section.

In an attempt to improve the utility of the Fermi-Thomas type of approach Smoluchowski (S15) has made use of an improved expression, as a result of the work of Weizsäcker (W9), for the kinetic energy of a distribution of electrons in terms of their spatial density. Whereas the ordinary expression of the Fermi-Thomas theory for the kinetic energy, proportional to an integral over space of the 5/3 power of the electron density, is valid only if the electron density is very high, referred to the scale of its spatial variation, Weizsäcker's expression is asymptotically correct not only at high densities but for all ordinary cases of low densities as well. This removes the most troublesome defect in the use of the Fermi-Thomas model. Unfortunately, however, Smoluchowski's calculations took the potential energy of the distribution to be merely the Coulomb energy of the total charge distribution and as the improvement in the kinetic energy expression only makes it more advantageous than before for the electrons at the metal boundary to spread out into the vacuum, the double-layer moment which results from minimizing the total energy comes out even larger than in the ordinary Fermi-Thomas calculation, which as we have seen in the preceding paragraph gives far too large values. The reason for this unsatisfactory result is obvious from a comparison with the wave mechanical calculations of Bardeen, which we have discussed in Section IV.2a: none of the Fermi-Thomas calculations so far made has taken into account the exchange and correlation fields, which exert a powerful influence to make the overlapping electrons hug the boundary of the metal.

In spite of this defect, the calculations of Smoluchowski are of special interest because they represent the only attempt so far to estimate the effect of the differences in atomic arrangement from one crystal to another. Smoluchowski idealized the metal as consisting of a number of polyhedral unit cells, arranged into a crystal as shown schematically in Fig. 23, each polyhedron being filled with positive charge of uniform density. In describing the electron distribution near the surface, Smoluchowski distinguished two effects, which he called "smoothing" and "spreading." The smoothing effect is measured by the smoothness of the contours of constant electron density, the spreading effect by the distance normal to these contours within which the density decreases from its interior value to a negligible value outside. In Fig. 23, A illustrates no smoothing and no spreading, D illustrates complete smoothing and no spreading while B illustrates spreading and partial smoothing together. It is clear that smoothing gives a double layer with the positive charge outward, hence a negative contribution to D, while spreading gives a positive contribution to D. Smoluchowski attempted to calculate both effects by assuming the electronic charge density to be a function involving two parameters, one determining the steepness of the essentially exponential decay of electron density with distance normal to the surface, and the other determining the height of the hills and valleys in the surfaces of constant density. These two parameters were chosen so as to minimize the total energy. It was found that the first parameter, which measures the spreading effect, was almost constant from one crystal face to another. The second parameter, however, which measures the smoothing, showed considerable variation, the smoothing being, as one would expect, the more complete the shorter the spacing between the peaks and valleys of the original surface. Thus, no attempt to calculate theoretically the variation of work function from one crystal face to another can be quantitatively valid unless the variations in the amount of smoothing are taken into account. However, Smoluchowski's calculations confirm that there should be a strong tendency for crystal faces on which the surface atoms are densely packed to have higher work functions than those on which the surface arrangement is more open, since a given amount of smoothing gives a smaller negative contribution to D on a dense face than on a less dense one.

It is interesting to note in passing that the type of calculation just described gives figures for the specific surface energies of the different crystal faces of a metal. Smoluchowski's results for these,

like his computed work function differences, are probably numerically unreliable because of his neglect of exchange and correlation effects; however, they suggest that there is a general tendency for crystal faces of high work function to have low surface energies, and *vice versa.*

IV.3 Temperature Derivative of the Work Function and Its Bearing on A^*

A number of attempts have been made to compute the temperature derivative of the true work function φ from some assumed model describing the electrons in a metal. According to Eq. (I.5.4), a positive value of $(d\varphi/dT)$ implies that the effective emission constant A^* for a uniform surface should be less than the universal constant $A=120$ amp./cm^2/deg.^{2}. It has often been assumed, erroneously of course, that the observed emission constant A^{**} of a polycrystalline metal is the same as the A^* of a uniform surface, and that therefore the fact that A^{**} is about one-half A for W, Ta, and Mo implies that $d\varphi/dT$ is of the order of 6×10^{-5} volt/deg. By coincidence, a value of this order can actually be obtained from a simple theoretical argument (H11, R2). However, as we shall see, there are several effects which may be expected to make contributions of comparable magnitude to $d\varphi/dT$, and most of the attempts to compute this quantity, including the ones just cited, overlook some of these effects.

It will be convenient to start by classifying the various physical effects which may cause φ to vary with T, and then to try to estimate the magnitude of each effect by using modern theories of metallic structure; some of the possible effects will, of course, turn out to be negligible. First of all we may break φ up into the inner work function and the potential jump at the surface. Let Φ be defined as the volume average of the electrostatic potential in the interior of the metal (see Section I.1b) and let the zero of potential be chosen to be the potential Φ_a just outside the surface, a choice which will simplify the notation in the discussion to follow. Then from (I.2.1)

$$e(d\varphi/dT)=e(d\Phi/dT)-d\mu/dT. \quad \text{(IV.3.1)}$$

We may distinguish three contributions to the last term by setting

$$\mu(v,T)=\mu(v,0)+\Delta_T\mu+\Delta_\Theta\mu. \quad \text{(IV.3.2)}$$

Here $\Delta_T\mu$ represents the small change which would occur in μ if the temperature were raised from 0 to T while keeping the volume constant and holding the atoms of the metal fixed in their equilibrium positions, and $\Delta_\Theta\mu$ represents the additional change which results from the thermal vibration of the atoms. The mechanism of this latter change will be elucidated briefly below. The quantity Φ may be broken up in a similar manner:

$$\Phi(v,T)=\Phi_0(v)+\Delta\Phi+D. \quad \text{(IV.3.3)}$$

Here $\Phi_0(v)+\Delta\Phi(v,T)$ is defined to be the value which the inner potential would have at temperature T if the mean charge distribution around the surface atoms were the same as that around the atoms in the interior (see Section IV.2); $\Delta\Phi(v,T)$ is defined to be the temperature-dependent part of this potential, vanishing at $T=0$. The quantity D thus represents the amount by which the inner potential is altered by the differences between the mean charge distribution in the surface cells and in those of the interior, and it therefore represents the double layer moment according to the definition employed in Section IV.2.

A number of terms result when (IV.3.2) and (IV.3.3) are inserted into (IV.3.1) with the relation

$$d/dT=(\partial/\partial T)_v+\alpha v(\partial/\partial v)_T,$$

where α is the volume coefficient of expansion. These may be grouped as indicated in the first column of Table VIII. The remaining columns of the table summarize the discussion to be given in the following paragraphs; the references are listed in the approximate order of the thoroughness with which they treat the individual effects listed.

The results summarized in Table VIII, though rather qualitative, indicate the range of values within which it is likely for the effective emission constant of an ideal metal surface to lie. According to Eq. (I.5.4), the A^* of a uniform surface depends upon $d\varphi/dT$ and upon the behavior of the reflection coefficient. Unless r is very close to unity or varies rapidly with temperature, alternatives which the arguments of Section IV.4 suggest are unlikely for uncontaminated metal surfaces, reflection will not affect the order of magnitude of A^*. And since the two largest entries in Table VIII are probably usually of opposite signs and each of the order of a few times k for metals with low expansion coefficients, one can conclude that these metals, if free from contamination, should probably have A^* values not more than a few times larger or a few times smaller than $A=120$ amp./cm^2/deg.^{2}. Moreover, by Section II.7 the emission constant A^{**} of a polycrystalline surface should be smaller than the A^* of the patches from which most of the emission comes. Thus when A^{**} values are observed as large as say 100A, it seems more reasonable to suspect a temperature-dependent contamination, giving a large negative $d\varphi/dt$, than to accept them uncritically as characteristic of a clean metal surface. For example, early experiments on such metals as platinum and nickel gave very large values for A^{**}, while more recent experiments employing more careful efforts to achieve clean surfaces have

TABLE VI. Values of $(\varphi - D)$ obtained from (IV.1.7).

Metal	Computed $(\varphi-D)$, volts (at 0°K unless otherwise noted) Using theoretical U	Using observed U	Reference*	Remarks*	Observed φ at room temperature, volts*	Inferred double layer moment D, volts
Li	1.96	2.17		$m/m^*=0.656$ (B2)	2.28 (O3)	0.0–0.1
		2.24		same, but with room temp. r_s		
Na	1.80	1.92		$m/m^*=1.07$ (B2)	2.29 (M7)–2.28 (M11)	0.3
		2.02		same, but with room temp. r_s		
	1.96		(H6)	H–K method		
K		2.20	(W15)	m/m^* assumed $=1$, room temp. r_s	2.24 (O3)–2.26 (M12)	0.0–0.1
	1.95		(H6)	H–K method		
Rb		2.20	(W15)	m/m^* assumed $=1$, room temp. r_s	2.16 (O3)	0.0
	1.90		(H6)	H–K method		
Cs		2.15	(W15)	m/m^* assumed $=1$, room temp. r_s	1.87 (O3)	−0.3
	1.86		(H6)	H–K method		
Cu		3.68		$m/m^*=1.0$ (F14)	4.9 (111 face) 5.55 (100 face) (U1)	1.2–1.9
Be	−1.90	−1.2	(H10)	Based on directly computed energy levels, rather than on effective mass concept.	3.92 (M7)	5.1
Mg	1.25		(H6)	H–K method	3.67–3.79 (M7, A4, C3)	2.4–2.5
Ca		1.0		Based on directly computed energy levels (M9)	2.71 (J1)–3.21 (L8)	1.7–2.2
	1.9		(H6)	H–K method		

* References referred to in parentheses are found in the Bibliography.

yielded values of around 30 amp./cm²/deg.². (See entries 35 and 37 in Table IV, Section III.1.) A similar suspicion of contamination should be entertained when values of A^{**} as small as say 0.01A are obtained; however, in this case there may also be a possibility of attributing the low A^{**} to very small isolated patches of low work function, or, in the case of metals with a large expansion coefficient, to the principal thermal expansion effect itself. Obvious checks could be applied to either hypothesis.

Note that for the (100) and (211) faces of tungsten the data of Table I, Section II.1, suggest that $|ed\varphi/dt| \ll k$, i.e., that the various effects enumerated in Table VIII cancel one another fairly accurately. This cancellation, though doubtless fortuitous, is quite consistent with the views presented here.

IV.3a *The Principal Thermal Expansion Effect*

The quantity in the first row represents the effect which has received the most attention in the literature. The possibility that thermal expansion may cause φ to vary with T was pointed out long ago by Richardson (R9), but an attempt to calculate the magnitude of the effect from modern theories seems first to have been made by Herzfeld (H11), and later independently by Reimann (R2). On the basis of the simple Sommerfeld free electron model of Fig. 22, the maximum kinetic energy W_i of the occupied electronic states varies as the $-\frac{2}{3}$ power of the volume, so that if W_a remains constant the expansion of the metal with increasing temperature should increase the work function of this model. This is the effect calculated by Reimann, and gives values of $ed\varphi/dT$ of the order of k for metals such as W, Ta, and Mo. Actually there is

no reason why W_a should not change. Herzfeld identified W_a with $-e\Phi$, the space average of the electrostatic potential, and computed its variation with volume for Ni by straightforward methods on the assumption that the charge distribution around the surface atoms is the same as around those of the interior. In this case the potential $-e\Phi$ was found to become numerically less as the volume increases, by an amount sufficient to outweigh the decrease of W_i and give a negative temperature coefficient for $\varphi = W_a - W_i$. A computation of the variation of Φ with volume has also been made by Blochinzev and Drabkina (B18), using a Fermi-Thomas model developed by Tamm and Blochinzev (T2). The validity of this model is questionable (see Section IV.1d), however, it agrees with Herzfeld's calculation in that the variation of $-e\Phi$ outweighs that of W_i. A calculation by Seely (S8), on the other hand, yields a variation of $-e\Phi$ which is smaller than that of W_i. Seely assumes the potential at any point inside the metal to be given by

$$\Phi = e \sum_n (Z - S(r_n))/r_n, \qquad \text{(IV.3.4)}$$

where Z is the atomic number, $S(r)$ is a screening function, and r_n is the distance from the point in question to the nucleus of the nth atom of the crystal. This assumption thus has the property of making the charge distribution around the surface atoms different from that around the atoms of the interior, in such manner as to give a positive value to D. Seely assumed the variation of the function S with lattice constant to be such that the values $S(r_n)$ remain unchanged by thermal expansion. This assumption makes D increase as the crystal expands, by an amount sufficient to make the sign of the computed $d\varphi/dT$ positive.

Much more reliable than any of the preceding is the method used by Wigner (W13) to evaluate the thermal expansion effect. This method consists in using the Wigner-Bardeen expression (IV.1.7) for the work function. The quantity $[e\Phi_0 - \mu(v,0)]$ is represented by the right side of (IV.1.7) with the last term omitted. When this is differentiated with respect to volume the term in dU/dv drops out, since the energy is a minimum at the volume which the metal has at the absolute zero. The quantity in the first row of Table VIII is therefore given by a sum of simple functions of the lattice constant, plus a term involving the volume derivative of $(\epsilon_{\max} - \epsilon_{Av})$. Wigner evaluated this term by assuming the electrons to have the same effective mass as free electrons, i.e., by assuming that an electron of wave vector k has energy $\epsilon_k = \epsilon_0(v) + (h^2/2m)k^2$. This assumption is known to be good for Na (B2), probably good for Cu (F14), but poor for Li (B2) and for Be (H10). It is not certain to what extent this assumption is valid for thermionic metals such as Ta, W, and Mo; the calculations of Manning and Chodorow (M8) suggest that the assumption is probably not justified for these metals. If this free electron assumption is used nevertheless, for want of a better one, the term $-(\epsilon_{\max} - \epsilon_{Av})$ gives a positive contribution to $d\varphi/dT$. Since the sum of the other terms of (IV.1.7), exclusive of eD, also increases with increasing volume, the quantity in the first row of Table VIII comes out positive, and the numerical value turns out to be of the order of a few times k, or more for metals with especially large expansion coefficients (see the entries for Li and Na in Table VI, Section IV.1c).

IV.3b *The Internal Electrostatic Effect of Atomic Vibrations*

The second and third entries in Table VIII, taken together, represent the effect of atomic vibrations in the interior; the quantity $(e\Delta\Phi - \Delta_\Theta\mu)$ measures the amount by which the work function would change if one could eliminate thermal vibrations by holding all the atoms of the interior of the metal fixed in their equilibrium positions, without altering the volume, electron temperature, or distribution of charge at the surface. In the treatment given here this effect has been separated into two terms because it is more convenient to calculate the $\Delta\Phi$ and $\Delta_\Theta\mu$ effects separately. Wigner, however, has treated the two effects together (W13).[r]

To show that the thermal vibrations of the atoms in the interior of a metal do, in fact, alter the mean electrostatic potential, consider the potential Ψ at

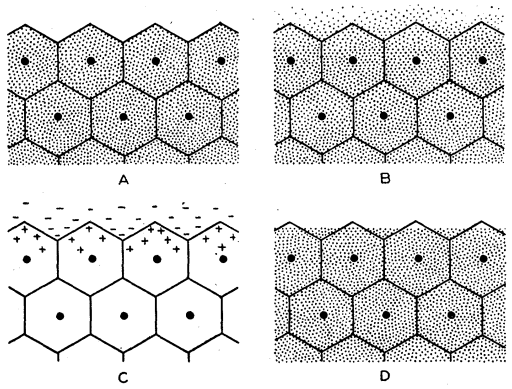


FIG. 23. Formation of a double layer at the surface of a metal by spreading out of the electron distribution. A: Electron distribution as it would be if the charge distribution about the surface atoms were the same as that about the interior atoms. B: Actual electron distribution, mean density of electrons being indicated by density of stippling. C: Charge density for B minus that for A, equivalent to a double layer. D: Illustration of complete "smoothing" with no "spreading" (see Section IV.2b). In all four sketches the vacuum is above, the metal below; cell boundaries and nuclear positions are drawn in merely to provide points of reference.

[r] We wish to thank Professor Wigner for discussing these effects with us and clarifying the relationship between his approach and ours.

TABLE VII. Double-layer moments calculated by Bardeen.

	Double-layer moment D, volts	Work function φ, volts
Idealized metal, correlation omitted	1.0	2.0
Idealized metal, correlation included	0.4	2.35
(110) face of sodium correlation included, complete smoothing assumed	0.15	2.17*

* From the value 2.02 volts given in Table VI for $(\varphi - D)$.

the various points within a unit cell of the crystal. The mean potential Φ is the volume average of Ψ, which can be expressed as an integral over the volume Ω_0 of a single unit cell. Now Ψ can be expressed as a sum of contributions from the charge distributions in the various unit cells of the crystal:

$$\Psi = \sum \Psi_j, \qquad \text{(IV.3.5)}$$

where Ψ_j represents the potential field produced by all the charges in the jth unit cell.

If we assume the charge distribution in the surface cells to be the same as in the interior cells, the average potential will, according to the definition accompanying (IV.3.3), be $\Phi_0+\Delta\Phi$. Thus

$$\Phi_0+\Delta\Phi = \frac{1}{\Omega_0}\sum_j \int_{\text{cell } o} \Psi_j d\tau. \qquad \text{(IV.3.6)}$$

Now if the suffix $-j$ is used to identify the cell whose position relative to the zeroth cell is the inverse of the position of the jth cell,

$$\int_{\text{cell } o} \Psi_j d\tau = \int_{\text{cell } -j} \Psi_0 d\tau,$$

and so (IV.3.6) becomes

$$\Phi_0+\Delta\Phi = \frac{1}{\Omega_0}\int_{\text{crystal}} \Psi_0 d\tau, \qquad \text{(IV.3.7)}$$

and since under the present assumption each cell is electrically neutral, the limits of integration may be extended to infinity without appreciable error. To express (IV.3.7) in terms of the charge density ρ we may write

$$\int_\infty \Psi_0 d\tau = \int_\infty \Psi_0 \frac{\nabla^2 r^2}{6} d\tau,$$

and if the unit cell has been so drawn as to have zero dipole moment we may transform this expression by Green's Theorem and discard the surface integral over the sphere at infinity, with the result

$$\Phi_0+\Delta\Phi = \frac{1}{6\Omega_0}\int_\infty r^2\nabla^2\Psi_0 d\tau = -\frac{2\pi}{3\Omega_0}\int_{\Omega_0} r^2\rho d\tau. \qquad \text{(IV.3.8)}$$

Setting $\rho=\rho_0+\Delta\rho$, where ρ_0 is the charge distribution at the absolute zero and $\Delta\rho$ the statistical average of the charge resulting from thermal vibrations, we have finally

$$\Delta\Phi = -\frac{2\pi}{3\Omega_0}\int_{\Omega_0} r^2\Delta\rho d\tau. \qquad \text{(IV.3.9)}$$

Note that (IV.3.8) and (IV.3.9) are independent of the choice of origin for r, by virtue of the neutrality of the unit cell and the fact that it has been drawn so as to have no dipole moment.

A correct numerical evaluation of (IV.3.9) would of course have to take account of the contribution to $\Delta\rho$ caused by distortion of the free electron distribution by the vibrating ion cores, as well as of the contribution from the ion cores themselves. However, one can probably get the correct order of magnitude of $\Delta\Phi$ by computing the latter effect alone and assuming the ion cores to vibrate as rigid units. For the shear modes of the alkali metals it is in fact probably fairly accurate to treat the ion cores as rigid and to assume that the charge distribution of the free electrons is unaffected by thermal vibrations; this is shown, for example, by the successful calculation of the shear constants of the metals by Fuchs (F15, F16), using these assumptions. Even for alkalis, however, it is not legitimate to neglect the distortion of the valence electron distribution in compressional modes of vibration, so the value which we shall compute for (IV.3.9) will be at best only a rough approximation to the truth.

If the valence electron distribution is unchanged by the vibrations and if the lattice is monatomic and composed of rigid ions of charge Ze, the integral in (IV.3.9) will equal $\overline{\xi^2}Ze$, where $\overline{\xi^2}$ is the mean square distance of the ion from its equilibrium position. At high temperatures this quantity can be shown to be proportional to the mean, over all normal modes of vibration of the lattice, of the reciprocal of the square of the frequency. If the frequency spectrum of the lattice is of the Debye form a straightforward calculation gives

$$\overline{\xi^2} = 9\hbar^2 kT/M(k\Theta)^2, \qquad \text{(IV.3.10)}$$

where M is the ionic mass, Θ the Debye temperature, and T is assumed well above Θ. Inserting (IV.3.10) into (IV.3.9) and expressing the result in numerical form, there results, for the idealized case considered here,

$$e\frac{\partial\Delta\Phi}{\partial T} = -\left[\frac{1.47\times 10^8 Z}{M\Theta^2\Omega_0}\right]k, \qquad \text{(IV.3.11)}$$

if M is in atomic weight units, Θ in degrees, and Ω_0 in cubic angstroms.

IV.3c *The Effect of Atomic Vibrations on the Chemical Potential*

The third entry in Table VIII represents an effect which can be described physically in either of two ways, which are of course equivalent to each other thermodynamically. The first and more direct way of picturing the mechanism involved is illustrated in Fig. 24. Here the energies of the various levels which the different electrons of the metal occupy are plotted as functions of any one of the many possible coordinates which describe the displacements of the atoms from their equilibrium positions. If the crystal lattice is one of high symmetry, as is the case for most metals, the pattern of curves will be symmetrical, although the individual curves need not be. For simplicity only symmetrical curves are shown. The important thing is merely that they have a curvature. For the case shown in the figure, the average energy of each electronic level increases with increasing amplitude of thermal vibration of the atoms, and this effect will tend to make the height of the Fermi level increase with temperature. It is perfectly possible, of course, for the Fermi level to decrease with temperature, instead of increase. All we can say with certainty is that the curve of total energy of the metal against amplitude of vibration must be concave upward, and this does not require that the uppermost curves in Fig. 24 be also concave upward, though it makes it probable that this is the case for most modes of vibration.

If the energy ordinates are measured from some absolute level, this change in the mean height of the uppermost filled levels will measure the temperature derivative of the electrochemical potential $\bar{\mu}$; however, we have already estimated the temperature derivative of the mean electrostatic potential in the preceding section, so we shall suppose the ordinates of Fig. 24 to be measured relative to the mean electrostatic potential, so that we may interpret the curvature of the curves for the uppermost filled levels in terms of the temperature variation of the chemical potential μ. This vibration-dependent part of μ is what we have called $\Delta_\Theta\mu$.

To obtain a rough quantitative estimate of $\Delta_\Theta\mu$ and its temperature variation, however, it is more convenient, as Wigner (W13) has shown, to use a different approach based on a thermodynamic identity. We shall start from the relation

$$(\partial\mu/\partial T)_v = (\partial^2 F/\partial T\partial n)_{v,\ \text{charge}}.$$

Here the subscript "charge" on the right means that $\partial/\partial n$ is to be interpreted as the change which would occur if one electron were added and if an equal positive charge, uniformly distributed over the interior of the crystal, were imagined to be added at the same time. Without the positive charge the change would give $\bar{\mu}$ instead of μ. Interchanging the order of differentiation gives

$$(\partial\mu/\partial T)_v = -(\partial S/\partial n)_{v,\ T,\ \text{charge}}. \quad \text{(IV.3.12)}$$

Now the entropy, S, can, to a good approximation, be divided into an entropy of vibration and a generally much smaller entropy of electronic excitation. The contribution of the vibrational entropy to (IV.3.12) gives $\Delta_\Theta\mu$, while the contribution of the electronic entropy gives the other term, $\Delta_T\mu$, introduced in (IV.3.2) above. Following Wigner (W13) we may use a Debye model for the vibrations and write for the vibrational entropy

$$S_{\text{vib.}} = \text{function of } (T/\Theta),$$

where the Debye temperature, Θ, may depend on the volume and on the electron concentration. Thus

$$\frac{\partial S_{\text{vib.}}}{\partial n} = -\frac{T}{\Theta}\frac{\partial\Theta}{\partial n}\frac{\partial S_{\text{vib}}}{\partial T} = -\frac{C_v}{\Theta}\frac{\partial\Theta}{\partial n},$$

giving finally for the third row of Table VIII

$$-(\partial\Delta_\Theta\mu/\partial T) = -(C_v/\Theta)(\partial\Theta/\partial n)_{v,\ T,\ \text{charge}}, \quad \text{(IV.3.13)}$$

where, if C_v is to be interpreted as the molar specific heat, $\partial\Theta/\partial n$ must be interpreted as the change in Θ when one electron per mole is added to the metal with a compensating imaginary positive charge.

The problem is thus reduced to that of estimating how the effective Debye temperature Θ is affected by changes in the electron concentration. A rough lower limit to $\partial\Theta/\partial n$ for tantalum and tungsten can be obtained by comparing the values of Θ for these metals. If tungsten had the same nuclear charge, atomic weight, and atomic volume as tantalum and differed only in having one electron per atom more, compensated by a uniform positive charge in each unit cell, then the difference between the two Θ values could be interpreted as $N\partial\Theta/\partial n$ where N is Avogadro's number. Actually, of course, the compensating positive charge is localized at the tungsten nucleus instead of being spread over the cell. This will make the electronic wave functions hug the nucleus more closely and probably reduce Θ. Corrections for the differences in atomic weight and atomic volume are easily applied, since Θ varies inversely as the square root of the nuclear mass and since the variation of Θ with volume is related to the volume expansion coefficient α in the Debye approximation by

$$\alpha = -\frac{KC_v}{v}\frac{\partial\log\Theta}{\partial\log v},$$

where K is the compressibility, v the molar volume,

TABLE VIII. Contributions to $ed\varphi/dT$.

Term	Probable order of magnitude	Reference	Remarks
$\alpha v(d/dv)[e\Phi_0-\mu(v,0)]$ (principal thermal expansion effect)	probably positive, a few times k for high melting metals; rather larger for metals such as Fe which have larger α	Wigner (W13)	Wigner-Seitz-Bardeen model, including exchange and correlation.
		Seely (S8)	free electron model, no exchange or correlation
		Herzfeld (H11)	ditto
		Reimann (R2)	free electron model, ignoring change in Φ
		Blochinzev and Drabkina (B18)	Fermi-Thomas model
$e(\partial\Delta\Phi/\partial T)$ (internal electrostatic effect of atomic vibrations)	probably negative, of the order of a few times k	text Wigner (W13)	
$-(\partial/\partial T)\Delta_\Theta\mu$ (effect of atomic vibrations on the chemical potential)	probably only a fraction of k for most metals with d shells in contact; probably somewhat larger and negative in sign for other metals	text Wigner (W13) Herzfeld (H11)	
$e(dD/dT)$ (temperature variation of double-layer moment)	probably negative and not more than a fraction of k	text	principally due to thermal expansion
		Blochinzev and Drabkina (B18)	show effect of thermal energies of electrons to be negligible
$-(\partial/\partial T)\Delta_T\mu$ (effect of electronic specific heat)	may be of either sign but is negligible unless Fermi level lies very close to the top or bottom of a band	Sun and Band (S21)	
$\alpha v(\partial/\partial v)(e\Delta\Phi-\Delta_T\mu-\Delta_\Theta\mu)$ (minor expansion effects)	probably negligible		

and C_v the molar specific heat at constant volume. When the lower limit to $\partial\Theta/\partial n$ calculated in this way is inserted into (IV.3.13), the result is obtained that for tantalum and tungsten $-\partial\Delta_\Theta\mu/\partial T$ should be negative and at least as large as $\frac{1}{5}$ or $\frac{1}{6}$ of k. A similar calculation can of course be carried out for any adjacent pair of metals in the periodic table which have the same crystal structure; for ruthenium and rhodium and for iridium and platinum the resulting upper limit for $-\partial\Delta_\Theta\mu/\partial T$ comes out positive and close to $\frac{1}{3}$ of k.

Thus it seems likely that $-\partial\Delta_\Theta\mu/\partial T$ is only a fraction of k for these transition metals, and it is possible that it may be of either sign. This is what one would expect if the interatomic "stiffness" on which Θ depends is principally determined by the d shells, and if the properties of these d shells are not very radically altered by addition of a single electron. This situation might be expected to prevail for those metals, such as transition metals and noble metals, which have the d shells of neighboring atoms in contact. For other metals, especially those of low valence, a stronger dependence of Θ on μ, and hence a larger (presumably negative) value of $-(\partial\Delta_\Theta\mu/\partial T)$ might occur.

IV.3d *Temperature Variation of the Double Layer Moment*

Let us turn now to the fourth entry in Table VIII. The most reasonable cause for an appreciable temperature variation of the surface double layer, D, is again thermal expansion: presumably if the calculations of D described in Section IV.2a were made for different values of the lattice constant, appreciably different results would be obtained, so that thermal expansion should change D. We may make a crude estimate of the magnitude of this effect by noting that if the atoms of a metal could be squeezed together or pulled apart at will, a plot of D against lattice constant ought to look something as shown in Fig. 25. The upward concavity is suggested by Table VI; near the observed lattice constant a variation of D with a power of the lattice constant of the order of unity would be reasonable. Taking D to be of the order of one to several volts for a metal such as tungsten, we get a negative value of $\alpha ve(dD/dv)$ of the order of some tenths of k. This is rather small compared to the volume effect in the first row of the table; however, it is entirely possible that the volume derivative of D may be several times larger than assumed. Moreover, the expansion of the surface layers of a crystal need not be identical with the expansion of the body of the lattice. Other effects besides thermal expansion might conceivably contribute to the temperature derivative of D, but it is hard to think of any other effect which would be any larger. One such effect would be the roughening at the surface by thermal vibrations. Another, which can be shown to be negligible (B18), would be an

increase in the extension of the electron cloud outside the surface due to the thermal energies of the electrons.

IV.3e *The Effect of Electronic Specific Heat*

As was explained in connection with Eq. (IV.3.12) the fifth entry in Table VIII represents the effect of the normal "free election" contribution to C_v. If the electrons in the metal are treated as non-interacting and as occupying a continuum of levels of density $\nu(\epsilon)$ per unit energy, the levels being independent of temperature, the standard theory for an almost completely degenerate Fermi gas (e.g., (15), #16) gives

$$\mu(T)=\mu_0-\frac{\pi^2}{6}\left(\frac{d\ln\nu}{d\epsilon}\right)_{\epsilon=\mu_0}(kT)^2+O(T^4), \quad \text{(IV.3.14)}$$

whence

$$-\frac{\partial\Delta_T\mu}{\partial T}=\frac{\pi^2}{3}kT\left(\frac{d\ln\nu}{d\epsilon}\right)_{\epsilon=\mu_0}\cdot k. \quad \text{(IV.3.15)}$$

Now for perfectly free electrons $\nu\propto\epsilon^{\frac{1}{2}}$, so the coefficient of k in (IV.3.15) becomes $(\pi^2/6)(kT/\mu_0)$, which is ordinarily very small. For transition metals, however, the existence of narrow, partially filled d bands may make (IV.3.15) considerably larger, and the coefficient of k may become of the order of unity if the Fermi level μ_0 happens to be within a distance of the order of kT from the top or bottom of a band. Some calculations for this case have been made by Sun and Band (S21), with an attempt to allow for the overlapping of the s and d bands. The values of $-(\partial/\partial T)\Delta_T\mu$ which they calculate are probably too large, since the neglected term in (IV.3.14) is positive and quite appreciable for the cases they consider.

IV.3f *Minor Expansion Effects*

The last entry in Table VIII, the effect of thermal expansion on $\Delta\Phi$, $\Delta_T\mu$, and $\Delta_\Theta\mu$, could only be appreciable if one of these quantities were to change by an appreciable fraction of its value under the small change of volume produced by thermal expansion from 0°K to T. This is very unlikely.

IV.4 The Reflection Coefficient for a Uniform Surface at Zero Field

It would be very difficult to calculate rigorously the reflection coefficient of a crystal surface for an electron incident from outside. For a rigorous calculation should take account both of the atomicity of the potential field acting on the electron at the surface and in the interior, and of the possibilities of electron-electron collisions and inelastic collisions of the electron with the crystal lattice. Because of the difficulty of solving wave mechanical problems with more than one degree of freedom, however, most of the detailed calculations which have been published so far have assumed a simplified model of a metal, for which the potential acting on the electron is taken to be a function only of the coordinate normal to the surface. These will be summarized in Section IV.4a. The effect of the periodic field inside the crystal on the reflection coefficient can in some cases be to produce total reflection, in others merely to increase or decrease the partial reflection. Total reflection, which in the absence of inelastic impacts should occur when the entering electron finds itself in a forbidden energy region inside the crystal will be discussed in Section IV.4b, while the influence of the lattice structure on partial reflection will be taken up in Section IV.4c. The discussion given in these sections makes it appear highly probable, *a priori*, that for metals such as clean tungsten the reflection coefficient is of the same order as that given by the one-dimensional calculations of Section IV.4a, *viz.*, 0.05 or so; nevertheless, there is a possibility that the reflection coefficient may be much higher, and this possibility is quite a likely one for certain other metals. It is therefore fortunate that some fairly convincing evidence for a small reflection coefficient for tungsten is provided by the phenomenon of periodic deviations from the Schottky line; this will be discussed in Section IV.4d. Section IV.4e, finally, will contain a few remarks about the effects of inelastic impacts.

Throughout the whole of this section only crystal surfaces which are practically atomically smooth will be considered. Surfaces which are non-uniform on a scale large compared with the wave-length of a thermal electron will show reflection coefficients which are dominated by the patch field effects discussed in Section II.6; non-uniformities on a scale of the order of tens of angstroms, however, may produce reflection effects which are more difficult to estimate.

IV.4a *One-Dimensional Theory with Constant Interior Potential*

The best one-dimensional approximation to the effective potential acting on an electron colliding with a metal surface would probably be a function similar to that shown in the full curve of Fig. 26. This potential function follows the image law,

$$V=-e^2/4x,$$

at large distances outside the surface, deviates from the image law at smaller distances in a manner similar to the effective potential used by Bardeen (B1) in calculating the charge distribution in the double layer, and shows a periodic structure in the

interior similar to that which the actual potential would show if averaged over y and z.

One or two facts regarding the choice of this optimum one-dimensional potential are worth pointing out here. One is the fact, already noted in Section IV.2a, that because of the velocity dependence of exchange and correlation effects, electrons of different velocities will move in different potential fields. As long as one is dealing only with electrons whose energies outside the crystal are of the order of kT, this difference can be ignored, since a change as small as this in the energy will not affect the effective potential very appreciably. However, it should be borne in mind that the effective potential for these thermionic electrons will not be the same as for higher energy electrons—e.g., those encountered in electron diffraction or in secondary emission—or for lower energy electrons, e.g., those encountered in field emission.

For electrons of the speeds encountered in thermionic emission the correctness of the image law at large distances is indicated experimentally by the fact that the field dependence of the emission from metals follows the Schottky law (I.4.3). Moreover, from the purely theoretical point of view Bardeen (B3) has treated the many-electron problem of the motion of a slow electron outside the surface of a metal containing free electrons and has shown that it is a good approximation to use a wave equation for the external electron in which the potential energy is simply the image function.

It has sometimes been surmised, especially for composite surfaces, that there may be a hump in the potential curve near the metal surface. This seems a plausible possibility for composite surfaces, although there seems to be no reason to expect any hump at all for a clean metal surface. The existence of a hump extending above the level of the potential at infinity can of course be ruled out whenever the dependence of emission on field strength follows the Schottky law (I.4.3) though a hump with its peak significantly lower than the potential at infinity would not be inconsistent with approximate fulfillment of this law. However, if the field dependence of the emission can be measured sufficiently accurately, the presence of even the latter type of hump should become noticeable through an increase in the amplitude of the periodic deviations from the Schottky line, as compared with a clean metal (see Section IV.4d below).

The best theoretical estimates of the reflection coefficient which have been published to date have been made assuming a potential of the form

$$V=-e^2/4x \quad \text{for} \quad x>x_1,$$
$$V=-e^2/4x_1=-W_a \quad \text{for} \quad x<x_1. \qquad \text{(IV.4.1)}$$

This is shown by the dashed curve of Fig. 26. The solutions of the Schrödinger equation in this potential field are confluent hypergeometric functions outside x_1, and, of course, plane waves inside x_1. Using series expressions for the confluent hypergeometric functions involved, Nordheim (N4) and MacColl (M1) have calculated reflection coefficients for various values of W_a and of the normal energy $\epsilon_x=mv_x^2/2$, where v_x is the component of velocity of the electron normal to the surface when it is at an infinite distance outside the metal. The results of MacColl are reproduced in Fig. 27 as full curves.

It is interesting to compare these exact calculations of MacColl with previous calculations using the same potential function (IV.1.1) but employing the WKB approximation. According to Bethe ((15), #18) and Frank and Young (F9), a straightforward calculation can be made using this method for the wave functions to the right of point B, Fig. 26, and joining these wave functions to plane waves in the region left of B; the resulting reflection coefficient is

$$r=\frac{W_a^4}{16(\epsilon_x+W_a)^3+W_a^4}, \qquad \text{(IV.4.2)}$$

where W_a and ϵ_x are to be expressed in Rydberg units. Values computed from this approximate formula for $W_a=10$ ev and 20 ev are plotted in the dotted curves of Fig. 27.

Calculations using potential functions other than (IV.4.1) need be mentioned only briefly. Eckart (E1) has calculated the reflection coefficient for a barrier represented by a certain analytical expression involving three parameters. This expression is capable of representing barriers with a potential hump at the surface, as well as monotonic barriers such as those of Fig. 26; however, the Eckart potential approaches zero exponentially with increasing distance outside the surface and so is not capable of approximating an image law in this region. For this reason the reflection coefficient computed with the Eckart potential always approaches unity as the normal energy, ϵ_x, of the electron approaches zero. This behavior is well known in the case of the discontinuous step potential (see (15), #18), which is a limiting case of the Eckart potential. However, calculations we have made show that in order to get from Eckart's formulae a reflection coefficient of even 0.10 for an electron of energy $\epsilon_x=0.01$ ev, one would have to assume a potential whose initial drop on approaching the surface from outside would be much more abrupt than the image law, so much so as to be quite out of the question.

IV.4b *Total Bragg Reflection*

"Bragg reflection" has often been cited as a likely cause for a high reflection coefficient for electrons incident on a metal surface. The simplest

illustration of this is when the energy of the incident electron lies in the forbidden region between two allowed bands of energies in the interior; in such case the electron cannot continue to move into the metal, and so must be reflected with a reflection coefficient $r=1$. The same effect can occur even when the energy bands in the metal overlap: all that is necessary is that the given energy be forbidden for electrons having certain directions of propagation. Let us assume the surface of the metal to be an ideal crystal plane normal to the x direction, and let the incident electron have total energy ϵ and correspond to a wave function $\exp(i\mathbf{k}\cdot\mathbf{r})$ at a large distance outside the metal. Then if the electron is not to be totally reflected, there must exist a wave function of the Bloch or running-wave type inside the metal, having the same total energy as the incident electron and having wave numbers k_y' and k_z' which either coincide, respectively, with the incident k_y and k_z, or else differ from these by 2π times some vector of the two-dimensional lattice reciprocal to that of the identity periods of the crystal surface. Now imagine a surface to be drawn in the three-dimensional space of k_x', k_y', k_z', representing the locus of all wave vectors k' corresponding to the particular energy value ϵ inside the metal. It may or may not happen that this surface is intersected by one of the lines corresponding to allowed values of k_y' and k_z'. If there is no such intersection, the incident electron cannot continue to move into the metal, and barring inelastic collisions it must be totally reflected.

To predict when total reflection of this sort will occur, it would, of course, be necessary to have detailed calculations of the energy band structure of the metal concerned. In only a few cases have such calculations been made for the range of electron energies of interest in thermionic emission, and even for these cases the methods of calculation are sometimes questionable. Moreover, a correction must be applied to the electronic energy values as they are usually calculated, to take account of exchange and correlation effects. For the energy values usually given represent values of a Hartree energy parameter; an incident electron of essentially zero energy will, after entering the crystal, have a Hartree energy parameter lying considerably closer to the Fermi level than would correspond to the work function φ. This is because the negative exchange and correlation contributions to the interaction energy of the excited electron with the other electrons will be appreciably smaller than for an electron at the Fermi level.

Bearing these facts in mind, one can surmise from the calculations of Tibbs (T3) on the energy bands of copper and silver that slow electrons incident on a (111) plane of either of these metals might well suffer total reflection, if inelastic collisions are neglected. On the other hand, the calculations of Manning and Krutter (M9) on calcium and of Slater (S13) and Von der Lage and Bethe (V3) on sodium suggest that total reflection is unlikely for slow electrons incident on these metals. The calculations of Manning and Chodorow (M8) on tungsten are of course of particular interest, and because of the similarity of the structures of tungsten and tantalum the results should be at least roughly applicable to the latter metal. If these calculations can be trusted, total reflection should not occur for these metals for electrons of thermionic energies, no matter what the crystallographic orientation of the surface. Moreover, this absence of total reflection should persist over a moderate range of higher and lower energies, covering those expected for photoelectrons and field electrons.

It is appropriate to discuss here a paradox which arises in the statistical theory of thermionic emission when reflection effects are ignored. It is easily shown (e.g., (15), #18) that if the electrons in a metal are represented by Bloch waves, each energy band which contains electrons of the proper energies and directions of motion for thermionic emission will, in the absence of reflection at the surface, contribute a term $AT^2\exp(-e\varphi/kT)$ to the thermionic emission current. This result must hold whatever is the dependence of the electronic energy ϵ_k on the wave vector $\mathbf{k}$, provided that in the region of interest ϵ_k is a single-valued function of k_x, the component of $\mathbf{k}$ normal to the surface. Since it is perfectly possible for two or more such energy bands to overlap, it has been suggested ((15), #18) that overlapping bands may cause the emission current density to be two or more times the value $AT^2\exp(-e\varphi/kT)$. But according to the general reasoning of Section I.3, the emission can only be

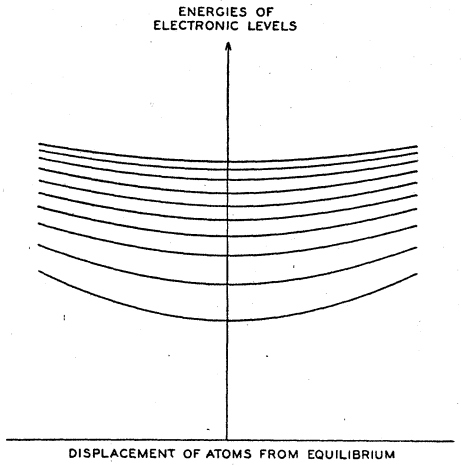


FIG. 24. Schematic variation of electronic energy levels with amplitude of displacement of nuclei from their equilibrium positions.

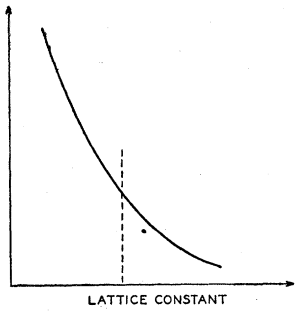


FIG. 25. Schematic variation of double layer moment D with lattice constant.

less than this quantity, never more! Otherwise the principle of detailed balance cannot be satisfied when the metal is in equilibrium with electron vapor. The explanation must be that whenever there are n different wave functions with the same k_y and k_z and the same energy, and each with a value of k_x corresponding to motion toward the surface, it must be possible to find $(n-1)$ orthogonal linear combinations of them which suffer total reflection at the surface and so remain confined within the metal. Indeed, it can easily be proved directly that this is the case. Suppose, for example, that ψ_1 and ψ_2 are two such Bloch waves inside the metal. Then there will be two complete wave functions which have, respectively, the forms

$$\chi_1=\psi_1+a_1\psi_1' \text{ inside}, \quad =b_1\exp(i\mathbf{k}\cdot\mathbf{r}) \text{ outside},$$
$$\chi_2=\psi_2+a_2\psi_2' \text{ inside}, \quad =b_2\exp(i\mathbf{k}\cdot\mathbf{r}) \text{ outside},$$

where ψ_1' and ψ_2' are reflected waves moving inward from the surface and where the wave vector $\mathbf{k}$ outside the crystal is determined by k_y, k_z, and the energy. Then the linear combination $(b_1\chi_2-b_2\chi_1)$ will vanish outside the crystal.

IV.4c *Other Effects of the Band Structure in the Interior*

Even when total reflection does not occur it may of course happen that the wave vector $\mathbf{k}'$ of the electron after it has entered the crystal lies so close to the boundary of a Brillouin zone that a high reflection coefficient r results. For a slight change in conditions would suffice to make $r=1$, and it is not to be expected that the change in r would be discontinuous. The only published calculations on this gradual rise in r as the condition for total reflection is approached are those of Morse (M17) and of Kronig and Penney (K3). These authors, who were more interested in the problem of electron diffraction than in the reflection of thermal electrons, used approximations to the potential field of the metal which were discontinuous on the boundary plane of the metal. For this reason the reflection coefficient calculated by their methods approaches unity as the normal energy of the impinging electron approaches zero, even when the conditions are so chosen that the wave vector of the electron inside the crystal is nowhere near the boundary of a Brillouin zone. We have seen in Section IV.4a that in the one-dimensional case the more correct image barrier gives a much lower reflection coefficient than the discontinuous rectangular barrier; for this reason an attempt to estimate the importance of reflection caused by the band structure in the interior of a metal can be carried out more easily and safely on the basis of calculations using an image barrier. We shall therefore devote most of this section to a discussion of some unpublished calculations of MacColl,[8] using a one-dimensional potential of the form

$$V=-e^2/4x \text{ for } x>x_1,$$
$$V=-V_0+V_1\sin[2\pi(x-x_1)/\lambda] \text{ for } x<x_1, \quad \text{(IV.4.3)}$$

where V_0, V_1, and λ are constants to be chosen and where to make V continuous x_1 must have the value $e^2/4V_0$.

All the calculations were made for a single value of λ, *viz.*, $\lambda=2$ angstroms. Figure 28 shows a few curves of reflection coefficient against normal energy; these have been selected from the many calculated by MacColl because they correspond to cases where the upper edge of a forbidden band of energies lies within one or two volts above or below the zero of normal energy, *viz.*, the energy of an electron of zero normal velocity far outside the metal. When the forbidden band extends from the negative into the positive energy range, the reflection coefficient is, of course, unity up to a certain value of normal energy, where a sharp decrease sets in; when the top of the forbidden band lies below the zero of normal energy, the curves are joined by dotted lines—physically meaningless, of course—to a point on the line $r=1$ with an abscissa corresponding to the band edge. The width of the forbidden band in each of the cases plotted is almost exactly equal to the value of V_1. As the behavior of the curves is not of a sort which can be predicted without computation, all the computed points are shown. The curve labeled (10,0) is of course identical with the curve $W_a=10$ of Fig. 27.

It will be noticed that the reflection coefficient decreases very rapidly as the normal energy of the electron recedes from the band edge, and that in all cases the reflection coefficient has sunk below 0.1 when the distance from the band edge is only a fraction of the band width. However, over a range of several volts from the band edge the reflection coefficient can be depressed considerably *below* the value for a constant inner potential ($V_1=0$).

[8] We wish to thank Dr. MacColl for discussing his work with us and supplying the results for use in this review.

Figure 29, which we have constructed by a rather rough interpolation from MacColl's results, shows this effect more clearly for the case where the width of the energy gap is 2 ev. The fact that the curve for $V_1=2$ ev continues to be below that for $V_1=0$, even at large distances from the band edge, may be due merely to the smoother behavior of the potential near x_1, and if so is hardly to be considered as due to the band structure; however, the pronounced dip in the reflection coefficient which commences about 4 or 5 volts from the band edge is probably due to the periodic potential, i.e., it would probably be absent if a constant potential were used in the interior.

The tentative conclusion to be drawn from these results is that if the effects of inelastic impacts can be neglected, reflection coefficients for thermal electrons should less frequently be found to lie between unity and 0.1 or 0.2 than below the latter value or at the former (total reflection). However, if the reflection coefficient is small, its value can easily depart widely from the value predicted by simple theories such as those of Section IV.4a.

Blochinzev and Drabkina (B18) have investigated the complementary problem of the reflection of electrons at a barrier of the form

$$V(x,y,z)=0 \text{ for } x>0,$$
$$=-V_0+\sum_g V_g \exp(ig_y y+ig_z z) \text{ for } x<0,$$

where the vectors $\mathbf{g}/2\pi$ run over the two-dimensional reciprocal lattice of the surface. Using methods similar to the more general ones of Morse (M17), they derive an expression for the reflection coefficient for electrons trying to leave the metal. This reflection coefficient can become fairly large if the V_g are large and if V_0 is large enough so that for electrons of positive total energy there are two or more allowable values for the normal kinetic energy inside the metal corresponding to given values of the total energy and of k_y and k_z, the tangential components of the reduced wave vector. This is not surprising, since we have seen in the prededing section that when the latter situation occurs the *sum* of the transmission coefficients for all the Bloch waves in question cannot exceed unity.

Mrowka ((M20), see also (S5)) has published a preliminary notice of some calculations which relate the crystalline anisotropy of the emission constants of a metal to the anisotropy of the energy band structure in the interior. The results seem to conflict with the views we have presented in this article, but since no explicit account of Mrowka's calculations has been published, no critical comments can be given here.

IV.4d *Evidence Provided by Periodic Deviations from the Schottky Line*

In 1939 Seifert and Phipps (S9) and Turnbull and Phipps (T5) discovered that when careful measurements are taken on the variation of emission current, j, with the strength, E, of the collecting field, at constant temperature, the Schottky plot of $\log j$ against $E^{\frac{1}{2}}$ is not straight, as is demanded by the simple theory expressed in Eq. (18), but shows small oscillations about a straight line, which increase in amplitude and period as the field E is increased. This was later fully confirmed by Nottingham (N8). According to a suggestion of Mott-Smith (M18), worked out in detail by Guth and Mullin (G8, G9, G10) this phenomenon is due to the fact that according to wave mechanics the potential hump H of Fig. 1, produced by the superposition of the collecting field and the image potential, is capable of reflecting a small fraction of the electrons which have normal energies sufficient to pass over it. The electron waves reflected from the hump can interfere with the waves reflected from the surface of the metal, and as the phase of this interference will depend on the distance of the hump from the surface, hence on the value of E, the average reflection coefficient $\bar{r}$ may be expected to show a slight periodic variation with E. The detailed theory of Guth and Mullin is in fairly satisfactory agreement with the experimental results of Turnbull and Phipps (T5) and of Nottingham (N8). This agreement constitutes a pleasing verification of wave mechanics. But this is not the only interest of the phenomenon and its theory: as will be shown in the following paragraphs and in Section IV.5, the comparison of theory and experiment enables one to conclude with some certainty that the reflection coefficient of the metal surface used was small and of the order of magnitude given by Fig. 27. Moreover, this result suggests the possibility of using Schottky plots as a tool to investigate such topics as the effect of adsorbed layers on the reflection coefficient and the possible occasional

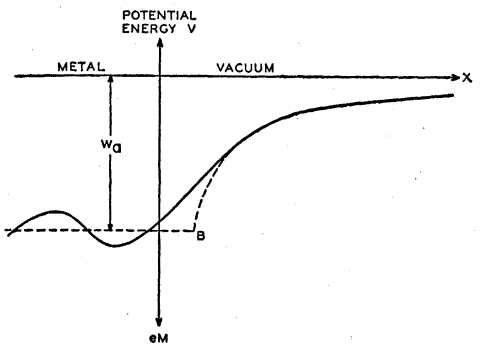


FIG. 26. One-dimensional approximation to the effective potential for an electron crossing the surface of a metal.

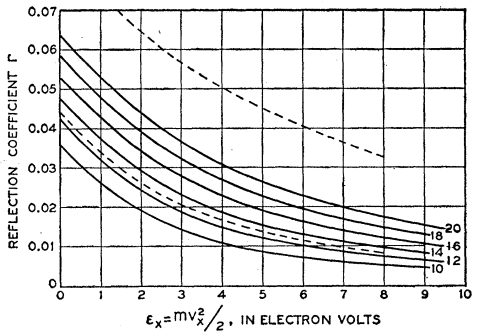


FIG. 27. Reflection coefficient of the one-dimensional image barrier (IV.4.1). The full curves are plotted from the exact calculations of MacColl for the values $W_a = 10$, 12, 14, 16, 18, and 20 electron volts, as shown. The lower and upper dotted curves give the values calculated from the WKB expression (IV.4.2) for $W_a = 10$ and 20 electron volts, respectively.

occurrence of high reflection coefficients for certain crystal faces of certain metals due to the electronic band structure in the interior.

As the mathematical basis of the theory will be developed in the next section, we shall give only a qualitative discussion here. To begin with, let us review the meaning of the term "reflection coefficient" in connection with thermionic emission. The most natural way of visualizing the effect of surface reflection on emission is to picture the electrons as moving through the crystal toward the surface, and these being either reflected or transmitted. This is in essence what actually occurs, of course. But it is a little easier to reason clearly regarding the various complicating factors which the development of the theory encounters if we make use of the principle of detailed balance in the way described in Sections I.3b and I.4, and so express the saturation emission current in terms of the reflection coefficient $\bar{r}_v$ of the surface barrier for electrons incident from well outside the barrier. In this way it is possible to describe the reflection coefficient in terms of the properties of electron waves outside the boundary of the metal, and the properties of the interior of the metal need to be introduced only in one place, *viz.*, as a boundary condition satisfied by the external waves.

The problem is thus to replace Eq. (I.4.4) by an expression for $\bar{r}_v$ which will take account of the interference between the electron waves reflected at the potential hump H of Fig. 1 and waves reflected at the metal surface. This will be done in Section IV.5. Here, however, we are interested only in the dependence of $\bar{r}_v$ on the individual reflecting powers of the two barriers involved. Suppose that for ingoing electron waves incident on the outer hump, H, the reflected wave amplitude in the absence of a second reflecting surface (i.e., the boundary of the metal) would be λ times the incident amplitude. Similarly let μ be the ratio of reflected to incident amplitude for waves incident on the metal surface proper from without. If λ is not too great, the reflected wave resulting from superposition of the waves reflected from the two barriers will have an amplitude $(\lambda+\mu)$ times the incident amplitude; the more detailed treatment of Section IV.5 shows that regardless of the values of λ and μ this ratio of amplitudes is $(\lambda+\mu)/(1+\lambda^*\mu)$, which is nearly $(\lambda+\mu)$ if either λ or μ is small. Here λ and μ are, of course, complex numbers, since the phase relation between the two interfering waves must be taken into account. Thus the reflection coefficient, r_v, of the composite barrier for electrons of any given initial velocity is given approximately by

$$r_v \approx |\lambda+\mu|^2 = |\lambda|^2 + |\mu|^2 + 2R(\lambda^*\mu). \quad \text{(IV.4.4)}$$

Here the first two terms represent the reflection coefficients which the image hump and the metal surface would have, respectively, each in the absence of the other. The third term, twice the real part of $\lambda^*\mu$, depends upon the phase difference between λ and μ, and thus upon the energy of the electron and upon the distance x_0 of the image hump from the metal surface, hence upon the field strength E.

Since most of the electrons which escape from the metal have energies within a few kT's above the top of the image hump, we are interested in the values of λ and μ only in this range of energies. As we shall see in Section IV.5, $|\lambda|$ must go to zero when the wave-length of the electron when it is at the top of the barrier becomes small compared with the distance the electron must move to double its kinetic energy, a condition which is satisfied at fields of the order of 10^5v/cm when the energy of the electron is only a few hundredths of an electron volt above the top of the barrier. The situation is quite different with $|\mu|$, however. For any electron capable of crossing the image hump will have a kinetic energy of the order of volts by the time it gets within an atomic diameter of the metal surface, and it is very unlikely that a change of a few tenths of a volt in this value can make a radical change in $|\mu|$. Thus it is probably legitimate to treat $|\mu|$ as a constant in averaging (IV.4.4) over a Maxwellian distribution of incident electrons to get $\bar{r}_v$.

Since only the last term of (IV.4.4) contributes to the periodic variation of $\bar{r}_v$ with E, the amplitude of the periodic deviations from the Schottky line will be proportional to $|\mu|$. Now the calculations of Guth and Mullin (G8) which we shall describe in Section IV.5 are based on assumptions which seem safe enough as far as the calculation of λ is concerned, while for the calculation of μ they employ assumptions equivalent to those of the WKB calculations discussed in Section IV.4a, assumptions which one might reasonably question on

several grounds, especially on the ground of the use of the one-dimensional approximation. However, the fact that the amplitude of the periodic deviations predicted by these calculations agrees with the observed amplitude for tungsten indicates that the value of $|\mu|$ computed by Guth and Mullin is essentially correct, since no other factor in the computed amplitude is likely to be much in error. Since we have seen that $|\mu|$ is not likely to depend sensitively on the energy of the electron over a range of a few tenths of an electron volt, the reflection coefficient $|\mu|^2$ computed for the metal surface at zero field by the one-dimensional methods of Section IV.4a is probably not far from the truth, at least for tungsten. Similar experiments on other surfaces would obviously yield very useful information.

IV.4e *Inelastic Impacts*

In the preceding sections the reflection of the electron by the metal surface has been treated as elastic, i.e., the electron has been treated as a wave moving in a fixed potential field. Now as has been pointed out by Compton and Langmuir (C7) and by Becker and Brattain (B9), the occurrence of inelastic impacts may considerably decrease the reflection coefficient of a solid surface. For, because of the image force, an electron which has only thermal velocity when at a large distance from the surface will enter the metal with a kinetic energy of the order of volts; only a small fraction of this need be lost to prevent the electron from again escaping to infinity. Unfortunately, inelastic impact has received little attention in the theoretical literature, so only a few general remarks can be made here.

Inelastic impact can occur either through excitation of lattice vibrations, or through loss of energy to the free electrons already present in the metal. The latter effect is probably the more serious for metals. Several attempts have been made to estimate the mean free path of an electron whose energy is a few volts above the Fermi level. From photoelectric measurements on thin films of platinum Compton and Ross (C8) and later Goldschmidt and Dember (G4) estimated that the mean free path of the photoelectrons was of the order of tens of angstroms; this result, however, was based entirely on the assumption that the photoelectrons were produced throughout the whole thickness of the film, an assumption which conflicts with the more modern view that photoelectrons are produced chiefly at the surface. Direct measurements of the transmission of low energy electron beams through thin nickel foils have been made by Becker (B7), but it is not clear how his results should be interpreted. He found that for incident energies in the range 6 to 100 ev only about 10^{-4} of the incident electrons were able to penetrate a foil 0.02μ thick, but that increasing the thickness to 0.04μ merely reduced the transmission by an additional factor of twenty. If one attempts to reconcile the two results by assumption of inhomogeneities in the foils, one must assume a mean free path, with respect to some kind of elastic or inelastic collision, of not more than about 20A. A figure of this order seems not unreasonable theoretically (see (15), p. 472) and is plausible in view of the known rate of energy loss of electrons of somewhat higher energy; however, more clear-cut experimental evidence is much to be desired.

The next question is, how do such inelastic interactions of a slow impinging electron with the atoms of a metal affect the reflection coefficient? Slater (S14) has pictured the effect, as far as electrons of the original energy are concerned, as equivalent mathematically to the introduction of an imaginary term into the potential energy operator of the Schrödinger equation. He has shown that this will broaden the Bragg reflection peaks and reduce their height, so that the reflection coefficient never reaches unity and may be much less, even at its maximum. Since Slater's calculations refer to energies of the order of 50–100 ev and do not take explicit account of the surface of the metal, the numerical details of his results are not of interest here. His method, however, could easily be applied to the problem of electron reflection by a simple image barrier such as the dotted curve of Fig. 26; it can easily be estimated that if this were done, using a complex potential consistent with the estimates of the preceding paragraph on the frequency of inelastic collisions, the resulting reflection coefficient would not differ very greatly from that computed with a real potential as in Section IV.4a. Thus it is quite conceivable that the effect of inelastic impacts might merely be to decrease greatly any Bragg reflection peaks without radically affecting the reflection elsewhere.

For insulators and many semiconductors it is likely that inelastic impacts by incident thermal electrons are less probable than for metals, since there are no conduction electrons capable of absorbing small amounts of energy, and since the incident electrons will in most cases not have sufficient energy to excite the bound electrons. This fact has been suggested by Bethe (B14) as a reason for the fact that insulators have higher secondary electron yields than metals.

IV.5 Mathematical Theory of Periodic Deviations from the Schottky Line

We have already outlined, in Section IV.4d, the nature of the phenomenon of periodic deviations from the Schottky line and the physical principles

involved in explaining it. In the present section we shall take up the mathematical problem of the solution of the Schrödinger equation for an electron approaching or leaving the surface of a metal in the presence of a strong electric field. In doing so our main concern will be to examine the adequacy of the one-dimensional methods of calculation which have to be used in problems of this sort. Omitting computational details we shall then summarize briefly some of the results obtained by Guth and Mullin (G8, G9, G10) on the way in which the transmission coefficient of the surface region varies with the energy of the electron and the strength of the electric field.

Following the procedure outlined in Section IV.4d, let us set up an expression for the reflection coefficient of the combined barrier formed by the image hump and the metal surface, as experienced by electrons incident from outside. Now there are three regions within which it is meaningful to speak of electrons of a given energy as being either "ingoing" or "outgoing" (see Fig. 1):

(1) The region well outside the image hump, i.e., to the right of H.
(2) A region between the image hump and the metal surface, not too close to either.
(3) The region well inside the metal.

Near the image hump itself, and near the metal surface, the WKB approximation breaks down, and in these regions one cannot classify the solutions of the Schrödinger equation into ingoing and outgoing waves. The argument to be presented here is based on the assumption that in regions (1) and (2) the potential can be taken to be a function of x, the distance normal to the surface, and is independent of y and z. It will be further assumed that in regions (1) and (2) the wave function can be adequately represented by a single product of the form $\exp(ik_y y+ik_z z)$ times a function of x. This is a reasonable assumption. For the most general possible wave function having tangential wave numbers k_y and k_z and energy ϵ is

$$\exp(ik_y y+ik_z z)\sum_g f_g(x)\exp(i\mathbf{g}\cdot\mathbf{r}), \quad \text{(IV.5.1)}$$

where the vectors $\mathbf{g}/2\pi$ run over the two-dimensional lattice reciprocal to the identity periods of the metal surface. If the energy, ϵ, lies only slightly above the top of the image hump, the large tangential kinetic energy of all the terms of (IV.5.1) except the term $\mathbf{g}=0$ will necessitate normal energies for the $f_g(x)$ lying far below the image hump. These $f_g(x)$ will therefore decrease exponentially with increasing x, and will probably become negligible before the outer boundary of region (2) is reached.

To compute the over-all reflection coefficient for electrons having given y and z components of momentum, corresponding to the wave numbers k_y and k_z, it is merely necessary to find what linear combination of outgoing and ingoing waves in region (1) corresponds to a purely ingoing wave in region (3). Let

$$\chi_i(x)\exp(ik_y y+ik_z z)$$

be the solution of the Schrödinger equation in region (2) which represents ingoing waves having the given energy and wave numbers k_y and k_z. Similarly let

$$\psi_i(x)\exp(ik_y y+ik_z z)$$

be the ingoing solution in region (1) for the same energy, k_y, and k_z. Corresponding expressions formed with the complex conjugate functions χ_i^*, ψ_i^*, will of course represent outgoing waves, with the same flux density as the ingoing ones. The phases and amplitudes of χ_i and ψ_i need not be specified at the moment; a convenient choice of the phase of ψ_i will be made presently. Using the symbol $\rightarrow$ to mean "joins on to" we may define μ, c, and λ by

$$\text{ingoing wave } (k_x,k_y,\epsilon) \text{ in region (3)}\rightarrow (\chi_i+\mu\chi_i^*)\exp(ik_y y+ik_z z), \quad \text{(IV.5.2)}$$

$$\chi_i\rightarrow c(\psi_i+\lambda\psi_i^*). \quad \text{(IV.5.3)}$$

By proper choice of the phase of ψ_i, c can be made real; from now on it will be supposed that this is the case. From (IV.5.2) and (IV.5.3),

$$\text{ingoing wave } (k_x,k_y,\epsilon) \text{ in region (3)}\rightarrow c[(\psi_i+\lambda\psi_i^*)+\mu(\psi_i^*+\lambda^*\psi_i)]\exp(ik_y y+ik_z z),$$

and by comparing the coefficients of ψ_i and ψ_i^* the reflection coefficient is seen to be

$$r_v=|\lambda+\mu/1+\mu\lambda^*|^2. \quad \text{(IV.5.4)}$$

This is the equation upon which the arguments of

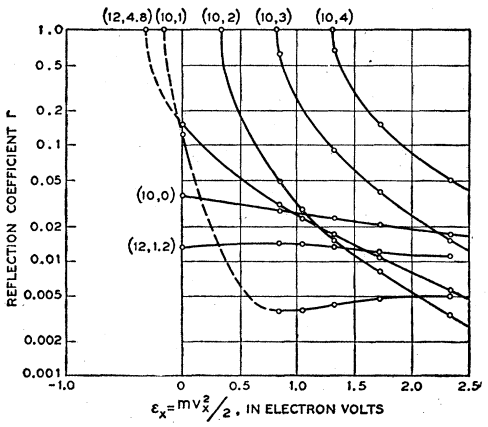


FIG. 28. Examples of the behavior of the reflection coefficient near the edge of a forbidden band of energies. (Based on MacColl's calculations using the potential (IV.4.3).) The values in parentheses by each curve are the values of V_0 and V_1, respectively, for that curve, measured in electron volts.

Section IV.4d were based. Note that no assumption has been made regarding separability of the wave equation in region (3), or the nature of the potential field in the transition region at the metal surface.

To compute r explicitly as a function of the applied electric field strength, E, and the energy, ϵ, of the electron, it is necessary to get explicit expressions for λ and μ. Now as has already been pointed out in Section IV.4d, μ is not likely to be appreciably affected by E or by changes of the order of kT in the electronic energy ϵ; $|\mu|^2$ may in fact be identified with the reflection coefficient r at zero field, and any of the one-dimensional methods of calculation discussed in Section IV.4a will yield a value for μ although because of the three-dimensional nature of the problem at and inside the metal surface, it is conceivable that this value may not be correct. The calculation of λ, however, is less simple than the one-dimensional calculation of μ. Guth and Mullin (G8) have calculated λ as a function of E and $\epsilon_x=\epsilon-(\hbar^2/2m)(k_y^2+k_z^2)$; their procedure was to join on WKB solutions in regions (2) and (1) to more accurate solutions of the Schrödinger equation near the top H of the image hump, where the WKB approximation of course breaks down for electrons whose energies are only slightly above the top of the hump. To obtain these more accurate solutions in the region near the top of the hump, they integrated exactly the Schrödinger equation for a parabolic potential $V=\text{constant}\cdot(x-x_0)^2$, so chosen as to osculate the actual potential at the hump H.

Although Guth and Mullin did not express their results in terms of reflection amplitudes λ and μ of the sort employed above, it is possible with a little labor to translate their expressions into the language used here. The results, which we shall give without proof, give an illuminating picture of the way in which the periodic variations in r_v come about. The results for $|\lambda|$ and $|\mu|$ are

$$|\mu|\approx W_a^{\frac{1}{2}}/4, \qquad \text{(IV.5.5)}$$

$$|\lambda|\approx(\exp[2\pi(x_0^3/2)^{\frac{1}{2}}\epsilon_x]+1)^{-\frac{1}{2}}, \qquad \text{(IV.5.6)}$$

where W_a is the difference in potential energy between the inside and outside of the metal at zero field (see Fig. 26), expressed in Rydberg units, x_0 is the distance of the image hump from the surface in units of the Bohr radius $\hbar^2/me^2$, and ϵ_x is the normal part of the energy of the electron, expressed in Rydbergs and measured from a zero coinciding with the height of the top H of the image hump (see Fig. 1). Note that (IV.5.5) is consistent with (IV.4.2), as it should be, since both are derived by a WKB calculation using the dotted potential of Fig. 26. Regarding $|\lambda|$, it will be noticed that its maximum value, $2^{-\frac{1}{2}}$, occurs for $\epsilon_x=0$, and that

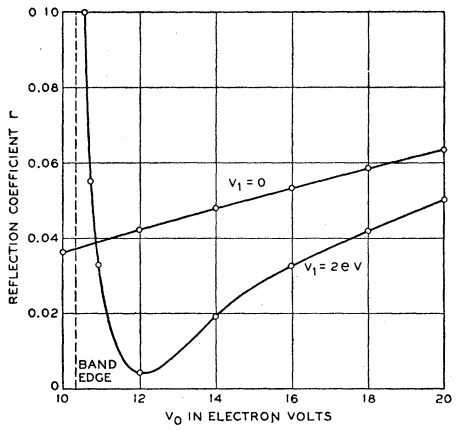


FIG. 29. Examples of the effect on the reflection coefficient produced by introducing a periodic term in the interior potential. (Based on MacColl's calculations using the potential (IV.4.3).)

with increasing electron energy it goes rapidly to zero, as one would, of course, expect from the fact that the image hump is so smoothly rounded. For example, if $E=2.5\times10^5$ volts/cm and $\epsilon_x=0.02$ ev, $\alpha=0.64$, and so $|\lambda|=0.13$. Thus the periodic variation of $\bar{r}_v$ with field is due almost entirely to those electrons which cross the image hump with much less than average thermal velocity.

The phase difference between λ and μ turns out to be a rather complicated function of ϵ_x and E, which Guth and Mullin were forced to approximate in a rather crude way in order to facilitate numerical calculation. Fortunately, however, for the small values of ϵ_x where $|\lambda|$ is large, this phase is not very sensitive to ϵ_x and rough approximations to its dependence on ϵ_x are adequate. For larger values of ϵ_x a large error in the phase does little harm, since $|\lambda|$ is very small. The calculation of $(1-r_v)$ given by Guth and Mullin in their first paper (G8) amounts essentially to inserting this approximate phase together with (IV.5.5) and (IV.5.6) into (IV.5.4).

Having calculated $(1-r_v)$ as a function of ϵ_x and the applied field E, it is merely necessary to average this quantity over a Maxwellian distribution of values of ϵ_x in order to get the mean transmission coefficient $(1-\bar{r}_v)$, which is proportional to the observed emission currents for the various values of E. As can be seen from Fig. 3 of the first paper of Guth and Mullin (G8), the agreement between the theoretical and observed Schottky plots is rather good, both in the amplitude and in the phase of the periodic deviations.

Guth and Mullin have extended their calculations to give predictions of periodic variations of photoelectric current with applied field (G9) and have also discussed field emission and the tunnelling of thermally excited electrons through the upper por-

tion of the image barrier, which gives rise to a gradual transition from thermionic to field emission (G10).

APPENDIX I. LIMITS OF APPLICABILITY OF THERMODYNAMIC CONCEPTS

There are many situations in which it is desirable to know on how fine a scale it is meaningful to define the chemical potential of electrons as a function of position and to assume it to depend only on the local composition, electron density, and temperature. This is particularly true when one is dealing with semiconductors, since beneath the surface of a semiconductor there is often a space-charge layer within which the concentration of free electrons varies rapidly with depth; moreover, semiconducting cathodes often have a very small crystal grain size, and the penetration of the resulting patch fields into the interiors of the grains gives rise to an additional inhomogeneity. Similar questions arise for metals too, as for example, when dealing with a metal containing tiny inclusions of some precipitate. They arise again for the space-charge cloud outside a cathode, where the potential changes rapidly with position.

In these and other situations there are often departures from thermal equilibrium, the commonest such departures being those resulting from temperature inhomogeneities or to the flow of electronic current. The latter flow may be spatially quite non-uniform, as in the case of an oxide cathode whose emission comes preferentially from certain patches of a fine-scaled distribution. In such cases one would like to know whether it is legitimate to define a chemical potential for the electrons at each point which will be the same function of temperature, electron density, etc., as in thermal equilibrium, and whether, for example, it is correct to use the chemical potential so defined in transport equations such as (I.6.2) and (I.6.4).

We shall start by discussing the legitimacy of treating the chemical potential of electrons as a function of position within a system which is in thermal equilibrium but which contains small-scale inhomogeneities in such properties as chemical composition or electrostatic potential. Now it is easy enough to define a chemical potential, μ, as a function of position in such a case: the macroscopic system as a whole must have a definite electronic electrochemical potential $\bar{\mu}$, defined as in Section I.1a, and so for any tiny region for which an electrostatic potential Φ can be defined we can define μ to be $\bar{\mu}+e\Phi$. However, in order that the quantity so defined be a useful concept, it is desirable that it be the same function of local state—e.g., of temperature, electron density, chemical composition, etc.—as the μ of a large homogeneous specimen. The question to be answered is thus: how fine-scaled can the spatial variations in local state become, before the relation between μ and the other local state parameters starts to deviate appreciably from the relation which would obtain for a large homogeneous body? In answering this we shall arbitrarily adopt the criterion of considering a deviation in μ to be "appreciable" if it is comparable with kT or larger.

When speaking of μ as a function of local state it is of course necessary that the elementary regions be large enough so that the other quantities involved in the specification of local state can be unambiguously defined, yet small enough so that these other quantities have essentially constant values over a region. The former condition can usually be satisfied for regions as small as a few tens of angstroms across, or even smaller, provided all the regions are in thermal equilibrium with each other and that conditions are not changing too rapidly with time. For if, for example, the average electrostatic potential over a unit cell of a crystal varies sufficiently gradually with position so that it can be considered constant over a region of the size just mentioned, this average can be adopted as the electrostatic potential Φ entering into the definition of μ. Again, the electron density in a tiny region can be defined in terms of the time average of the number of electrons in the region, even when the region is so small as to contain on the average far less than one electron. Chemical composition and state of strain can be defined with no greater difficulty.

There are two types of effects which under conditions of thermal equilibrium can cause the relation between μ and local state in the presence of small-scale variations in local state to differ from that for a homogeneous body. One is that the potential energy of interaction of the electrons in one small region with their surroundings may depend on conditions in neighboring regions in a more complicated way than through the effect of these regions on the mean electrostatic potential in the first region. The other is associated with the Heisenberg uncertainty principle: it makes no sense to talk about a small difference between the μ's of two neighboring regions if the act of localizing electrons in either region would introduce so large an uncertainty in their kinetic energy as to eclipse the small difference in the two μ's. The limitations resulting from these two effects will be discussed in turn, and will be followed by a discussion of the limitations due to departures from thermal equilibrium.

Ia. Potential Energy Limitations

We shall consider two types of variation of local state with position which may affect the interaction energy of one region with the electrons in a neigh-

boring region, *viz.*, variations in dielectric constant and variations in electron density.

The order of magnitude of the effect on μ caused by differences in the dielectric constants of neighboring regions is easily estimated from the image law. An electron located in a medium of dielectric constant κ_1, and at distance x from a plane boundary separating this medium from one of dielectric constant κ_2 has a potential energy of interaction with its image equal to

$$(e^2/4\kappa_1 x)(\kappa_1-\kappa_2)/(\kappa_1+\kappa_2).$$

This image law will break down when x becomes less than some length λ of the order of a couple of angstroms. Thus, the average interaction energy over a region of linear dimension, d, obtained by integrating on x, will be $\ll kT$ if

$$d\gg(e^2/4\kappa_1 kT)(\kappa_1-\kappa_2)/(\kappa_1+\kappa_2)\ln(d/\lambda). \quad \text{(I.1)}$$

For the space-charge region in the vacuum outside a metal cathode one must of course set $\kappa_1=1$, $\kappa_2=\infty$. If the minimum distance of the region in question from the second medium is not zero, but some positive value d_0, then d_0 should be substituted for λ in (I.1).

Turning to the effect of inhomogeneities in electron density, we may note that the effect on the μ on one region caused by changing the electron density in a neighboring region is due entirely to the fact that that mean interaction energy of an electron in the first region with the electrons in the second differs slightly from what one would compute from the mean electrostatic potential in the first region caused by the electrons in the second. This difference, sometimes called the "correlation energy," can be roughly evaluated from the Debye-Hückel theory of electrolytes ((4), Chapter IX) when the electron density is low, and from the calculations of Wigner (W12) when the electrons form a degenerate gas, as in a metal. For the former case, which applies to electrons in a semiconductor or a vacuum, we have for the total correlation contribution to μ, i.e., the difference between the actual μ of a dilute electron gas and the μ which an electron gas of the same density would have if the motions of the different electrons were statistically independent, the value (4, #913)

$$\Delta\mu=-(e^2/2\kappa)[4\pi e^2 n_1/\kappa kT]^{\frac{1}{2}}, \quad \text{(I.2)}$$

where n_1 is the number of electrons per unit volume and κ is the dielectric constant, which was taken as unity when we used this expression in Section I.3a. The criterion for the validity of (I.2) is most conveniently expressed by introducing the characteristic length ξ defined by

$$\xi=[\kappa kT/4\pi e^2 n_1]^{\frac{1}{2}}, \quad \text{(I.3)}$$

and sometimes called the "Debye length;" (I.2) is then valid whenever

$$n_1\xi^3\gg 1. \quad \text{(I.4)}$$

When the electron density n_1 is too large for this to be satisfied, (I.2) probably gives an upper limit to $|\Delta\mu|$.

We wish to know what restrictions must be imposed on the linear dimension d of a cubical region, in order that the part of (I.2) attributable to the interaction of electrons inside the cube with those outside it be $\ll kT$. Now whenever (I.4) is satisfied, (I.2) is $\ll kT$, since

$$\Delta\mu/kT=1/8\pi n_1\xi^3,$$

and for such cases the present considerations give no lower limit for d. At higher electron densities, where $\Delta\mu$ becomes comparable with kT, d must be taken large enough so that when $|\Delta\mu|$ is multiplied by the fraction of the electrons in the cube which are close enough to the boundary to polarize the electron distribution outside, the product will be $\ll kT$. Taking $3n_1^{-\frac{1}{3}}/d$ for this fraction and taking (I.2) as an upper limit for $|\Delta\mu|$, the restriction on d comes out to be that any d satisfying

$$d\gg n_1^{-\frac{1}{3}}(3/8\pi n_1\xi^3) \quad \text{(I.5)}$$

will be allowable, but that this restriction may be unnecessarily stringent.

For a numerical example take the case $\kappa=5$, $T=300°$. Then (I.4) is satisfied if $n_1^{\frac{1}{3}}\ll 1.9\times 10^7$ cm^{-1}, and in this range the present considerations impose no minimum size for a region. For the same κ and T (I.5) becomes $d\gg 6.2\times 10^{-9} n_1^{1/6}$ cm, if n_1 is in cm^{-3}. For the large κ and high T which one encounters for incandescent oxide cathodes, (I.2) will be $\ll kT$ for any reasonable electron density.

Turning now to metals, it is known that if the conduction electrons are quite "free" the correlation contribution to μ is roughly one ev and does not vary much with electron density or temperature. Local variations of electron density within a metal of homogeneous composition are of coures never great enough to affect this correlation energy appreciably; if we wish to consider tiny regions of two different metals in contact, the restriction which the linear dimensions d of the regions must satisfy to make their correlation interaction $\ll kT$ per electron is roughly

$$(3n_1^{-\frac{1}{3}}/d)\cdot(1\text{ ev})\ll kT,$$

or

$$d\gg(1000/T)\cdot 35 n_1^{-\frac{1}{3}}. \quad \text{(I.6)}$$

This limit is of the same order as that imposed by the uncertainty principle, which we shall now discuss.

Ib. Uncertainty Principle Limitations

It will suffice to consider how much the chemical potential μ corresponding to a given mean electron

density would be changed by isolating a small cubical region of side d by means of walls impenetrable to electrons but capable of conducting heat and transmitting electrostatic fields. In such a cubical box the energy level of a free electron will be quantized to values which can be represented, to an approximation adequate for the present purpose, by $\epsilon_i=\hbar^2k_i^2/2m^*$, where m^* is an effective mass for the electrons and the three components of the vectors k_i are restricted to be nonzero integral multiples of the unit $\delta k=\pi/d$. It is not hard to calculate rigorously the change $\delta\mu$ in μ resulting from introduction of the walls while keeping the electron density constant; for the present purpose, however, it will suffice to calculate merely its order of magnitude. This can be done by noting that if k-space is divided into cubes of side δk, the allowed states in the presence of the walls can be put in one-to-one correspondence with the cubes by taking for k_i the corner of each cube farthest from the origin. The continuous distribution of k values which occurs in the absence of the walls is roughly equivalent to assigning to each volume d^3 a set of k_i' values corresponding to the centers of the cubes. Thus we may expect $\delta\mu$ to be of the order of magnitude of the average value of $\hbar^2(k_i^2-k_i'^2)/2m^*$ for thermal electrons. For a semiconductor we may take these to be electrons for which $\hbar^2k_i^2/2m^*=3kT/2$, moving in the direction of a cube diagonal. Then if $\delta k\ll k_i$,

$$\hbar^2(k_i^2-k_i'^2)/2m^*=3^{\frac{1}{2}}\hbar^2k_i\delta k_i/2m^*$$
$$=(3\pi\hbar/2d)(kT/m^*)^{\frac{1}{2}},$$

giving

$$\delta\mu/kT\sim\frac{40}{d}\left(\frac{m}{m^*}\cdot\frac{1000}{T}\right)^{\frac{1}{2}}, \qquad \text{(I.7)}$$

if d is in angstroms, the result being valid only when it is rather smaller than unity. For a metal we may make the same calculation for an electron with kinetic energy W_i, obtaining

$$\delta\mu/kT\sim\frac{120}{d}\left(\frac{m}{m^*}\right)^{\frac{1}{2}}\left(\frac{1000}{T}\right)W_i^{\frac{1}{2}}, \qquad \text{(I.8)}$$

if d is in angstroms and W_i in electron volts. Thus d may have to be of the order of hundreds of angstroms to make (I.7) or (I.8) small compared with unity.

Ic. Limitations Resulting from Current Flow and Thermal Gradients

Consider first the case of a body at uniform temperature through which an electronic current is flowing. The statistical picture of such a body suggests a simple criterion for the legitimacy of defining the chemical potential μ to be the same function of local temperature and electron density as in the absence of a current, and using this μ in transport equations such as (I.6.2) and the Richardson Eq. (I.3.4). This criterion is that the statistical distribution of the conduction electrons in any region should not differ appreciably from the distribution which would obtain in the same region in the absence of a current but at the same temperature and mean electron density. Here an "appreciable" difference between the two electron distributions may, in accordance with the standard adopted above, be taken to mean any difference which for any electronic states of importance fails to be small compared with the change in the distribution which would result from a change in μ of the order of kT.

Thus for a semiconductor in which the electrons or holes have a mobility v_1 and an effective mass m^* the requirement in a homogeneous region of the semiconductor is that the electric field, E, satisfy

$$v_1E\ll(kT/m^*)^{\frac{1}{2}}=1.7\times10^7\left(\frac{m}{m^*}\cdot\frac{1000}{T}\right)^{\frac{1}{2}}. \qquad \text{(I.9)}$$

In regions where inhomogeneities or neighboring surfaces cause E to be different from zero even in thermal equilibrium, E in (I.9) should be replaced by ρj, where j is the current density caused by conduction and diffusion, and ρ is the resistivity.

For a metal a similar criterion may be set up, replacing the mean thermal velocity on the right of (I.9) by the change in velocity which an electron at the Fermi surface must acquire in order to change its energy by kT. This crude criterion for the validity of Ohm's law (I.6.2) agrees roughly with the detailed statistical calculations of Guth and Meyerhöfer (G7), who found that for the best metallic conductors current densities of the order of 10^9 amp./cm^2 would be required to give a 1 percent deviation from Ohm's law.

When there are temperature inhomogeneities the mean free path of the electrons becomes a limitation on the extent to which μ can be localized. In such cases the limitation on the applicability of thermodynamic concepts is most conveniently expressed as an indeterminacy in the temperature: the temperature to be assigned to any locality is uncertain by an amount of the order of the amount by which the temperature changes in a mean free path. "Temperature" can of course only be defined in such cases as the temperature of the equilibrium electron distribution whose properties most closely approximate those of the actual distribution in the region under consideration. Note that of all the limitation discussed in this Appendix, this is the only one which involves the mean free path.

A case analogous to the case of strong thermal inhomogeneities might arise for a semiconducting thermionic cathode whose conduction band lies close to or above the energy level of an electron at

rest in the vacuum outside the surface. Drawing saturation current from such a cathode would, of course, cause a serious departure from thermal equilibrium conditions for a distance of several mean free paths below the surface. However, we do not know of any cathode materials known to be of this type.

APPENDIX II. SOME THERMODYNAMIC RELATIONS

In Section I.7c it is shown that the latent heat of evaporation of electrons from an isolated conductor is

$$L=N_0[T(\partial S_{\text{gas}}/\partial n)_{p,T}-T(\partial S_{\text{cond.}}/\partial n)_{p,T}]$$
$$=Ne[\varphi-T(d\varphi/dT)]+(5/2)RT, \quad \text{(II.1)}$$

where φ is the true work function defined by (I.2.1), S_{gas} and $S_{\text{cond.}}$ represent, respectively, the entropies of electron vapor and conductor, and n is the number of electrons in either phase. For the temperature variation of L we can write, using the symbol $(\partial/\partial T)_e$ to denote $(\partial/\partial T)_p+(dp/dT)(\partial/\partial p)_T$,

$$\left(\frac{\partial L}{\partial T}\right)_{e,n}=N_0\left[\left(\frac{\partial^2(TS_{\text{gas}})}{\partial T\partial n}\right)_e-\left(\frac{\partial^2(TS_{\text{cond.}})}{\partial T\partial n}\right)_e\right]$$
$$=N_0\left[\frac{\partial(M_{\text{gas}}C_{p\text{ gas}})}{\partial n}-\frac{\partial(M_{\text{cond.}}C_{p\text{ cond.}})}{\partial n}\right],$$

where the M's refer to the number of moles of material present in the two phases, and the C_p's are the respective specific heats. Thus since $M_{\text{cond.}}$ and $C_{p\text{ gas}}$ are independent of the n's,

$$N_0\left[\frac{\partial(M_{\text{gas}}C_{p\text{ gas}})}{\partial n}\right]_{T,p}=-\frac{5}{2}R,$$

$$L(T)=L(0)+\frac{5}{2}RT$$
$$-N_0M_{\text{cond.}}\int_0^T\left(\frac{\partial C_{p\text{ cond.}}}{\partial n}\right)_{T,p}dT. \quad \text{(II.2)}$$

Inserting (II.1) and (II.2) into the Clausius-Clapeyron equation

$$d\ln p/dT=L/RT^2,$$

and integrating with use of the value of the chemical constant given by the quantum statistics it is fairly easy to show that

$$\ln p=-\frac{L(0)}{RT}+\frac{5}{2}\ln T$$
$$-\int_0^T\left[\frac{1}{\tau^2}\int_0^\tau\frac{M_{\text{cond.}}}{R}\left(\frac{\partial C_{p\text{ cond.}}}{\partial n}\right)_{T,p}d\tau'\right]d\tau$$
$$+\ln\frac{2(2\pi m)^{3/2}k^{5/2}}{h^3} \quad \text{(II.3)}$$

from which the emission Eq. (I.3.6) can be derived in the usual way, with

$$e\varphi=-\frac{L(0)}{N}-\frac{T}{N}\int_0^T$$
$$\times\left[\frac{1}{\tau^2}\int_0^\tau M_{\text{cond.}}\left(\frac{\partial C_{p\text{ cond.}}}{\partial n}\right)_{T,p}d\tau'\right]d\tau. \quad \text{(II.4)}$$

APPENDIX III. SURFACE STRUCTURE OF CRYSTALS

IIIa. Relative Thermodynamic Stability of Different Surface Arrangements

The first question one is inclined to ask in attempting to decide between hypotheses such as (1) and (2) of Section II.2 is: Which arrangement has the lower free energy? For a piece of metal must, if held for a long time at constant temperature and in equilibrium with its vapor, undergo progressive changes in its surface structure in the direction of lower and lower free energy. However, even under conditions where an equilibrium arrangement of minimum free energy would ultimately be achieved—to within the minor thermal fluctuations which must always occur[t]—the structure actually obtained in any finite time under given experimental conditions will depend on the rate of approach to the equilibrium state. Moreover, most experiments are conducted under such conditions that complete equilibrium would never be achieved—i.e., the metal is usually evaporating away all the time—and the atomic distribution in a steady state will then depend upon the relative speeds of various competing processes, and may differ greatly from the arrangement in thermodynamic equilibrium. Thus, after answering this first question, one must still inquire about the rates of the processes involved.

Because atoms at a free surface are bound by fewer neighbors than atoms in the interior of a substance, the energy, or more conveniently the free energy, of a given amount of material depends upon the area a of surface which it exposes. Thus one may, to a good approximation, set the free energy of a drop of liquid equal to a function proportional to the total mass, plus a term γa, where the coefficient γ is called the surface tension. For a solid the surface term must be written in the form $\sum\gamma_i a_i$, where γ_i is the specific surface free energy of the i-th crystal plane, and a_i is the area of surface exposing this type of plane.

It was shown many years ago by Wulff ((W20), esp. p. 512; see also (L9), (L7)) that the configuration which minimizes $\sum\gamma_i a_i$ for a fixed total volume is a polyhedron constructed in the following way:

[t] An interesting, though perhaps not entirely satisfactory discussion of these fluctuations on a crystal surface has been given by Frenkel (F12).

Let γ_i be plotted radially as a function of the direction of the normal to the surface; a two-dimensional cross section of this plot will look something like the outer curve in Fig. 30, with cusped minima in certain directions corresponding to surfaces of particularly simple structure. At each point of this polar plot construct a plane perpendicular to the radius vector at that point. Then the volume which can be reached from the origin without crossing any of the planes is geometrically similar to the ultimate equilibrium shape for the crystal. This construction is shown by the dashed line in the figure.

An approach to this equilibrium shape can, of course, only be realized in practice for very small crystals, and it is now generally accepted that the shapes of large crystals are usually determined by kinetic factors dependent on conditions of growth (16). We must expect, in fact that a crystal held in the presence of its saturated vapor will traverse a succession of configurations of successively lower free energies, but that these changes will practically cease when it becomes necessary to make a very large number of atomic displacements to produce an appreciable further lowering of the free energy. A change in the shape of a specimen of macroscopic size is quite inefficient in the latter sense, and so is not observed. However, it may in some cases be possible with far fewer displacements to decrease the free energy of an initially ideally smooth surface by rearranging the atoms into hills and valleys, many atoms on a side, but still small compared with macroscopic dimensions. It can be shown from simple free energy considerations that if the macroscopic surface of a crystal has the same orientation, relative to the crystal axes, as one of the boundary planes of Wulff's equilibrium polyhedron, no hill-and-valley structure can have a lower free energy than a flat surface, but that if the macroscopic surface does not coincide in orientation with some one of these planes, there will always exist a hill-and-valley structure of lower free energy than the flat surface (H9). The proof of the latter statement is based on the assumption that the free energy of the surface can be represented as a sum of the form $\sum\gamma_i a_i$, i.e., that the contributions of edges and corners to the free energy can be neglected in comparison with those from the surfaces. This assumption must be valid for a hill-and-valley structure of sufficiently large amplitude, but may not be valid if the structure is only a few atoms high. It is understood, of course, that all the surfaces considered will have a slight irregularity as a result of thermal fluctuations.

Thus, if the shape of the equilibrium polyhedron were known, it could be predicted which crystal surfaces of a metal would ultimately acquire a hill-and-valley structure when held at high temperature in equilibrium with the metal vapor, and which ones would remain essentially atomically smooth. However, care must be used in attempting to draw an inference of this kind from experiments. For example, the relative stability of different crystal planes may be changed by the presence of adsorbed gas. Again, the tendency of the surface to seek configurations of lower and lower free energy may be obliterated if other kinds of irreversible changes are taking place simultaneously. As we shall see below, such perturbing processes include not only evaporation into vacuum, but also the migration of metal atoms along the surface under the influence of electrical or thermal gradients, and possibly even the migration caused by inhomogeneous curvature of the surface.

There is, however, another limitation on the applicability of the statements made above, which applies even under equilibrium conditions and which it is therefore appropriate to discuss briefly in this section. The conclusions reached above were based on the assumption that the surface free energy of a metal specimen is given by $\sum\gamma_i a_i$, i.e., depends only on the areas of the various types of crystal faces which are exposed, and not on the nature or extent of the edges and corners where these faces join. This assumption becomes more and more nearly correct the larger the scale of the surface structure, but for very small specimens or for hill-and-valley structures of small amplitude it will certainly not be valid. Crystals with non-ionic binding would not be expected to have absolutely sharp corners, since the corner atoms ought to be less tightly bound than the average, and hence in thermal equilibrium should move to other sites. Stranski and his collaborators (16, S16) have discussed this fact in some detail in its relation to crystal habit. For our present purposes, however, the important question is, how much rounding-off will occur on the corners of the equilibrium polyhedron or of a quasi-equilibrium hill-and-valley structure? The answer to this question turns out to be closely related to the characteristics of the Wulff diagram of Fig. 30 (H9); however, a satisfactory discussion of this matter would be too lengthy to present here. Although quantitative predictions are difficult, it seems, at the present writing, very unlikely that the rounding of the equilibrium polyhedron for tungsten could be as complete as that observed for field emission points (see IIId of this Appendix).

IIIb. Mechanisms and Rates of Rearrangement of Surface Atoms

Mechanisms which have been suggested for the atomic rearrangement necessary in any change of the surface structure of a crystal include evaporation and condensation, surface migration, self-

diffusion in the volume of the metal, and plastic flow. It seems likely that under the proper conditions any one of these can effect changes under the motivation of surface tension forces, and in fact except for plastic flow there is at least indirect evidence for the participation of each of these mechanisms in surface changes caused by heating. Thus, for example, the fact that when a metal is held at sufficiently high temperature in the presence of its vapor evaporation and condensation can predominate over other mechanisms of rearrangement, may be inferred from the fact that tungsten (J5) and tantalum (M19) filaments develop a different surface appearance according to whether they are heated in vacuum or in an inert gas where most of the evaporated atoms return to the surface. Because of the large latent heat required for evaporation, however, it is to be expected that the importance of evaporation and condensation will rapidly decrease as the temperature is lowered, and that eventually processes having lower activation energies, such as volume or surface diffusion, will predominate. Surface diffusion or migration of foreign atoms on metal surfaces has been directly observed in many thermionic and field emission studies, of which only a few need be cited (A1, M22, B23, S7, B11). Since in some of these cases the foreign atoms were practically as tightly bound to the metal surface as atoms of the base metal itself, it is to be expected that a clean surface of the base metal will show self-migration of comparable speed at comparable temperatures. Although it should be easy with modern radioactive tracer techniques to measure self-migration on high melting metals, the only conclusive demonstrations of this phenomenon which have come to our attention to date have involved the growth of crystals of low melting substances by condensation (e.g., (V2, V1)), and under these conditions there is a possibility that the phenomena observed were greatly influenced by contaminations on the surface. It has been presumed that surface migration is responsible for the d.c. etching phenomenon (J5, S3, M19) and for changes in shape which have been observed on the sharp points used in field emission studies (M22, B12, B13, J3); however, it is hard to be absolutely certain that volume transport is not involved in these phenomena. Volume diffusion, finally, has been suggested as the mechanism primarily responsible for the sintering of metal powders (P1),[u] and this hypothesis receives some support from studies of the temperature dependence of sintering rates (K4a).[v]

[u] An earlier theory advanced by Frenkel (F11) attributes sintering to a viscous flow of the solid caused by vacancies in the crystal lattice. This is quite different from the diffusion process considered by Pines (P1). Frenkel's mechanism contradicts the established view that creep cannot be attributed to the motion of lattice vacancies (S10), and we do not believe it can be correct.

IIIc. Smoothing by Evaporation and Related Effects

Most thermionic experiments on metals are conducted in vacuum and involve heating of the cathode surface, at least for short times, to temperatures where appreciable evaporation will occur. This one-way evaporation will not generally bring about the same type of surface structure as results when crystal and vapor are allowed to approach thermal equilibrium. For example, it is sometimes supposed that evaporation proceeds at significantly different rates from the different crystal surfaces; this is perfectly possible, and there need be no particular connection between these rates of evaporation and the surface tensions which determine the shape of the equilibrium polyhedron. For the equilibrium vapor pressure must be the same over all kinds of crystal surfaces, provided they are flat, and by a simple appeal to the principle of detailed balance it can be shown that the rate of evaporation into vacuum is proportional to the equilibrium vapor pressure multiplied by the "sticking coefficient," α, which measures the fraction of the atoms impinging on the surface from the vapor phase stick and become incorporated into the crystal ((20), Kap. 2). It is conceivable that α might vary from one crystal surface to another, and these variations need not be correlated with any equilibrium property of the crystal. There is both theoretical and empirical evidence, however, that α is often practically unity for metal surfaces (L3); in these cases it is thought that surface migration plays an important part, the impinging atoms migrating over the surface so rapidly that they become incorporated into a partially-completed plane of atoms before they have an appreciable chance of being re-evaporated ((20), Kap. 2). When this happens there will be no difference in the evaporation rates from different crystal surfaces, provided the term "evaporation rate" is understood to refer to evaporation from a surface which is flat over a sufficiently large area. We shall presently see, however, that even when $\alpha=1$ effects can occur which have the same appearance as preferential evaporation.

It is necessary at this point to say a few words about the mechanism of evaporation which seems most likely in the light of present knowledge ((20), Kap. 2). This is shown schematically in Fig. 31. The atoms, idealized in the drawing as cubes, are arranged for the most part in complete and partially completed layers. The upper left portion of the sketch shows a partially completed layer,

[v] We are indebted to Dr. Kuczynski for discussions of his work prior to its publication.

bounded by a partially completed row which ends with the atom C. This atom will be easier to remove than other atoms such as D or E. An atom in a position like that of C can be removed by placing it in a position such as B, adjoining the edge of the incomplete layer, or it can be placed in an isolated position on the surface, as shown for A. The step $C \rightarrow B$ may be expected to require less work than $C \rightarrow A$; since there are many more positions of type A than of type B and the latter are in turn much more numerous than those of type C, we may expect that under conditions where no atoms return to the surface from the vapor phase the usual itinerary of an atom which evaporates will be of the type $C \rightarrow B \rightarrow A \rightarrow$vapor. When the crystal is in equilibrium with its vapor, on the other hand, this may or may not be the case. Atoms which condense on the crystal will usually land in positions of type A. If surface migration is sufficiently rapid in comparison with the rate of incidence of condensing atoms, these will usually become incorporated into the incomplete layer before they are re-evaporated, following itineraries of the type $A \rightarrow B \rightarrow C$; in such case the sticking coefficient α mentioned above will be nearly unity. If on the other hand surface migration is unable to keep up with the arrivals of new atoms on the surface, the concentration of atoms of type A will become significantly greater than it would be in the absence of the vapor phase, and many of these A atoms will re-evaporate before reaching positions of the type C; in this case α will be significantly less than unity. Since the activation energy for migration may be expected to be considerably less than the binding energy, the former situation ($\alpha \approx 1$) may be expected always to prevail at sufficiently low temperatures.

Consider now the process of evaporation into vacuum from the vicinity of an edge or corner of a crystal, which as we have seen in the previous section may be expected to be somewhat rounded even in thermal equilibrium. Per unit area of this rounded edge or corner there will be more of the comparatively loosely bound atoms, such as C of Fig. 31, than on the flat faces, and there will be fewer of the tightly bound atoms such as D or E. Therefore, the number of atoms per unit area which detach themselves from positions of the C type in unit time will be greater for the curved edge or corner region than on a flat face, and consequently this one-way evaporation will eat away the edge and corner regions preferentially; however, the atoms thus removed from edges and corners will usually make their final jumps into space from positions of type A on the flat faces, perhaps at some distance from the edge or corner where they originated. When the crystal is in equilibrium with its vapor this preferential rate of detachment of atoms from edges and corners is of course compensated by the fact that newly arrived atoms migrating over the surface are much more likely to find a place to stick on the curved portion than on an equal area of the flat portion.

The type of argument just given suggests further conclusions. If the range of surface migration is sufficiently great, it may happen that an atom which gets detached from an incomplete layer on one face of a small crystal will be more likely to migrate to another face than to re-incorporate itself into an incomplete layer on the original face. In such cases those faces of a small crystal which have a high concentration of positions of type C, Fig. 31, will be eaten away more rapidly by evaporation than will those with a low concentration of C positions; this can happen even when the sticking coefficient α is nearly unity. This preferential evaporation could, of course, not occur for a very large crystal when $\alpha \approx 1$; when the size of the crystal face is much larger than the range of surface migration a higher density of C atoms on one face of the crystal will not only cause a higher rate of detachment of C atoms, but will at the same time cause a larger percentage of the A atoms to return to C positions, so that the density of A atoms and hence the rate of evaporation will not be any higher than for the other faces.

Finally, we may note that the preferential eating away of edges and corners should occur not only when a metal evaporates into vacuum, but also in any other thermal process in which there is a net transport of material away from the region of the specimen where these edges or corners are located. Any process involving loss of material from a region of the surface should thus have a smoothing effect on this region, and if the rate of loss is sufficiently rapid this smoothing may even succeed in obliterating the hill-and-valley structure which some crystal surfaces tend to develop under equilibrium conditions.

Although the smoothing effect of evaporation should probably predominate over all other effects caused by heat treatment in vacuum when the temperature is high enough for evaporation to predominate over migration, we might expect that at temperatures where migration predominates heat treatment of a polished surface of macroscopic radius would cause the surface to progress toward thermal equilibrium, hence, for some crystal surfaces, to develop the hill-and-valley structure discussed in Section IIIa. However, it may take a very long time for such a structure to develop with appreciable amplitude. In the first place, it may conceivably happen that a hill-and-valley structure of small amplitude may have a higher free energy than the flat surface, even though one of large amplitude would have a lower free energy. In such case an "activation energy" would

have to be surmounted to produce the hill-and-valley structure. Even in the absence of such activation energy, the rate at which the sizes of the hills grow ought to become rapidly smaller as the hills become bigger, so that during any actual heat treatment only a small amplitude might result.

IIId. Observational Evidence

Observations of the appearance of heat treated metal surfaces have been made both optically[w] and by means of electron microscopy. There are several ways in which electron-optical techniques can be used to give the high magnification necessary for a study of the fine details of surface structure; information to be discussed below has been obtained from projection tube studies (J3), from shadowgrams (H2, J3), and from transmission studies of replicas of surfaces (M4, M6, H3) a method which promises to give more detailed information than any other. Important evidence on the question of surface structure has also been obtained in indirect ways from studies of thermionic emission from single crystals (M10, N1) and field emission from sub-microscopic single crystal points (J3, H2, M24).

We shall begin with a discussion of the appearance of tungsten filaments after various kinds of heat treatment (J5, S3); generally similar phenomena have been observed in less detail for tantalum (M19). Following Johnson (J5) we may summarize the facts as follows:

(1) Filaments which had been heated on a.c. in vacuum remained perfectly round and smooth, to within the resolution of the optical microscope, even after a period of incandescence sufficient to evaporate away 17 percent of the metal in the filaments. This fits in with the expectation developed in the preceding section that uncompensated evaporation should have a smoothing effect.

(2) Filaments which had been heated on a.c. in inert gas suffered "thermal etching" so that they exposed large (100) faces and smaller (110) faces. Both types of facets showed some concavity, plausibly attributed by Johnson to secondary effects connected with temperature inhomogeneities. The faceted structure may be presumed to represent a configuration with lower surface free energy than the smooth wire, and its form suggests that the (100) plane has the lowest surface tension.

(3) Filaments which had been heated on d.c. in vacuum showed a totally different step-like structure identical with a structure observed near the support hooks on the a.c. vacuum wires, where there was of course a strong temperature gradient. This, the "d.c. etch" effect mentioned in Section II.2 and shown in Fig. 4, is apparently caused by preferential surface migration in one direction along the wire, by a mechanism not yet understood.

(4) Filaments heated on d.c. in gas showed a combination of the effects (2) and (3).

Wires which have not been polished of course show scratches and die marks; the smoothing effect of evaporation in vacuum causes these to become gradually obliterated under prolonged heating. Wires which have undergone mechanical polishing, however, seem to be initially slightly roughened by heating, and then become smooth again. This may be due to recrystallization of a distorted surface layer produced by the polishing.

The assumption that evaporation into vacuum has a smoothing effect also for other metals seems to be consistent with experiments on copper (E3) and silver (S11), which have shown that the presence of oxygen or oxide is necessary for the appearance of a "thermal etch" on these metals, and that they become smooth when purified and heated in vacuum. However, the fact that heating in nitrogen also smooths silver (S11) necessitates caution in interpreting the vacuum smoothing as due to the process described in the preceding section.

From observation (2) above and the theoretical result of IIIa of this Appendix one would expect that most of the portions of a macroscopically curved specimen of tungsten would acquire a hill-and-valley structure if allowed to approach thermal equilibrium, but that probably the (100) and (110) faces would remain smooth. One might expect an approach to thermal equilibrium to result from heating at temperatures where migration or diffusion is appreciable but evaporation is negligible. However, as is discussed in the review by Jenkins (J3) all the observations which have been made on the sharp tungsten points used in field emission studies have shown that these points, whose radii are of the order of 10^{-4} cm, become smoothly rounded when heated for a short time at temperatures of the order of 1500°–1700°K or higher in the absence of a field; the same rounding occurs irrespective of the previous history of the point. Similar results have been found for molybdenum. The smoothness of the points has been confirmed in several ways: by electron microscope shadowgrams (H2, J3); by the smooth boundary observed for the area of lowered work function when barium is evaporated onto the point from one side (M22); by the absence of any bright spots in the field emission pattern, such as would be produced by field enhancement if there were a hill-and-valley structure as much as a few tens of atoms high (M25); and by the agreement of observed field emission currents (H2) with values calculated from

[w] The literature contains very many observations of metal surfaces, of which only some of the more recent ones need be cited here, with emphasis on those made on high melting metals. Papers devoted to particular metals include: tungsten (B15, J5, S3); tantalum (M19); copper (E3, D4, G11); silver (S11). L. Graf (G6) has published observations of surface structure for a number of metals, with many references to the earlier literature. Most experiments of this sort have involved heating in air, and the structures observed may often be due to oxidation or to the effects of adsorbed gas.

the Fowler-Nordheim theory (F7, N4, G10) for points of known radius, an agreement which would be spoiled if there were appreciable field enhancement at the peaks of a hill-and-valley structure. It is rather puzzling that these points become rounded and that instead of developing a polyhedral or hill-and-valley structure on heating they actually lose such structure when it has been artificially produced prior to the heating. A possible explanation is that the tendency of surface migration to bring about thermal equilibrium is outweighed by a smoothing effect of the type discussed in the preceding section, caused by a net flow of atoms by migration or diffusion away from the end of the point toward the less curved parts of the wire. That something of this sort must occur is indicated by the occurrence of changes in the shape of the point on heating, and one would expect it to produce a smoothing effect similar to that of one-way evaporation. The alternative explanation that the rounded shape is one characteristic of thermal equilibrium cannot be entirely excluded, but, as explained in IIIa of this Appendix, it seems unlikely.

Another type of evidence on the structure of metallic surfaces, less direct than the preceding, is provided by projection tube studies of preferential adsorption. Elaborate patterns have been obtained from barium, alkali metals, etc. deposited on field emission points (J3) and on a polished tungsten sphere emitting thermionically (M10). Martin (M10) has argued that the sharpness and complexity of the variations in adsorptive affinity with the orientation of the crystal surface in the latter experiments are evidence against the hypothesis of a hill-and-valley structure many atomic spacings in amplitude. For if this hypothesis were correct for most of the crystal surfaces, adsorption on these surfaces would take place on the sides of the hills, which would presumably consist of crystal planes of low indices; thus over a considerable portion of the surface of the sphere the only things which would vary would be the relative amounts of the area covered by the various hill sides, and a smoothly varying adsorption pattern should result.

We may mention finally the evidence from the field dependence of the thermionic emission from single crystal surfaces. In the measurement of the emission in different directions from a single crystal tungsten wire, Nichols (N1) found straight Schottky plots with the theoretical slope over the range of field strengths from below 30,000 to over 100,000 volts/cm, for all the crystal surfaces studied except the (110) surface and closely neighboring ones. According to the theory discussed in Section II.5, this indicates that, except for the latter surfaces, the work function of each surface could not have varied appreciably (say by a few tenths of a volt) from point to point of the surface on a scale anywhere in the range from a few tens of angstroms to a few thousand angstroms. Thus if hill-and-valley structures were present at all, either they must have been very small or very large. The low voltage deviation from the Schottky line which Nichols found for the (110) face indicates a patch structure on a scale of the order of thousands of angstroms or larger, and can be correlated with a shingle-like structure observed microscopically for this part of the wire.

To summarize, we have seen that there is much evidence for the non-occurrence of hill-and-valley structures of many times atomic amplitude on the surfaces of high melting metals which have been subjected to the usual types of heat treatment in vacuum. Little or no evidence contradicts the hypothesis that these surfaces, whatever their macroscopic orientation, are atomically smooth if the heating has been sufficiently severe and when the etching effects of d.c. fields and temperature gradients are eliminated. To reconcile this smoothness with the prediction that a hill-and-valley structure should in many cases be more stable thermodynamically, it is tempting to suppose that a certain activation energy must be surmounted in order to initiate the hill-and-valley structure. At temperatures where evaporation is negligible but migration rapid, this activation energy may be too high to be surmounted in any reasonable time. At higher temperatures, evaporation may be so rapid that the smoothing effect of evaporation will predominate over the tendency to build hills and valleys. Metals heated to a sufficiently high temperature in the presence of their saturated vapor, however, may be able to surmount the activation energy and develop a hill-and-valley structure. This tentative interpretation cannot be relied upon, however, until further experiments have been done on equilibrium surface structures for clean metals and until the mechanism of the d.c. etch is understood.

APPENDIX IV. ELECTROSTATIC POTENTIAL OUTSIDE PATCHY ELECTRODES

The purpose of this appendix is to provide a background for Sections II.4 and II.6. In general, the potential at sufficiently large distances from an uncharged conductor with a double layer spread over its surface approaches a constant value which is a surface average over the potential Φ_a just outside the surface. In the case of electrode geometries such as spherical, cylindrical, or plane, which are commonly used, the potential at sufficiently large distances is just the area surface average of Φ_a. Since the detailed discussion in the above sections has been limited to plane parallel electrode geometry, it has seemed desirable to write down the

complete expression for the potential variation between plane parallel patchy electrodes.

Consider an uncharged plane electrode (xy plane) which has a surface made up of patches of different double layers arranged in a periodic fashion having period X in the x direction and Y in the y direction. Let both X and Y be small compared to the dimensions of the electrode. It is easy to show that the solution of Laplace's equation which gives the potential $\Phi(xyz)$ in the neighborhood of this electrode, neglecting end effects, is

$$\Phi=\bar{\Phi}+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(a_n \sin 2\pi nx/X+b_n \cos 2\pi nx/X)\times(c_m \sin 2\pi my/Y+d_m \cos 2\pi my/Y)\times\exp[-2\pi z(n^2/X^2+m^2/Y^2)^{\frac{1}{2}}]. \quad \text{(IV.1)}$$

The constants $\bar{\Phi}$, a_n, b_n, c_m, and d_m can be evaluated in the usual way to fit the potential Φ_{ai} just outside the individual patches given by Eq. (II.4.1) with the result that

$$\bar{\Phi}=\sum_i f_i\Phi_{ai} \quad \text{(IV.2)}$$

where f_i is the fraction of the area occupied by the ith type of patch. The other components die off exponentially with increasing z so that at a distance from the surface, large compared to the period of the patch arrangement, the patch field is very small and the potential approaches $\bar{\Phi}$. If an external field is applied by means of a parallel uniform electrode placed a large distance away compared to the periodicity of the patch arrangement but a small distance compared to the dimension of the electrodes, a term linear in z should be added to Eq. (IV.1). If the parallel electrode itself has a patchy surface whose periodicity is small compared to the distance of separation, it is necessary to add terms of the same form as those of Eq. (IV.1) with constants valid in the neighborhood of the added electrode. These results are schematically illustrated in Fig. 8.

Generally, a perfectly periodic patch distribution in the case of polycrystalline emitters is never realized with the result that the patch field dies off more slowly than is indicated in Eq. (IV.1). In the case of an isolated conductor of finite size, the potential does nevertheless approach a constant value at sufficiently large distances regardless of the arrangement of the patches.

It should also be pointed out that Eq. (IV.1) is applicable to other than plane electrodes only as long as the periodicity of the patch distribution is very small compared to the radius of curvature and as long as the distance z from the surface is small compared to the radius of curvature. If the first condition is not satisfied, the patch field may die off more slowly with increasing distance from the surface than is indicated in Eq. (IV.1). For example, in the case of cylindrical electrodes a strip array of patches with the strips parallel to the axis of the cylinder can produce a patch potential which dies off only inversely with the radial distance from the surface. (See (14) p. 168.)

APPENDIX V. REFLECTION OF ELECTRONS BY PATCH FIELDS

In the derivation of the velocity distribution Eq. (I.3.7) there is nothing which precludes application to a patchy surface: it is only necessary that the plane across which the flow of electrons of different velocities is measured be drawn at a sufficiently great distance x_1 from the metal surface to be outside the patch field. If the applied field E is small enough so that $Ex_1 \ll kT$, as is usually the case in velocity distribution measurements, then according to Appendix IV the electrostatic potential on this plane will be $-\bar{\varphi}$ relative to the electrochemical potential inside the metal, where $\bar{\varphi}$ is the area average (II.4.3) of the work functions of the different patches. Therefore, if space-charge free emission can be achieved without violating the conditions just assumed the total emission will be given by the usual Richardson expression (I.3.6) with φ replaced by $\bar{\varphi}$ and with an associated reflection coefficient $\bar{r}$ which includes the effect of the patch field; likewise the velocity distribution expression (I.3.7) must still hold if r is interpreted to be the quantity γ, defined in Section I.3d as the fraction of the electrons on the reference plane in the given range $dvd\omega$ of outward velocities which have come

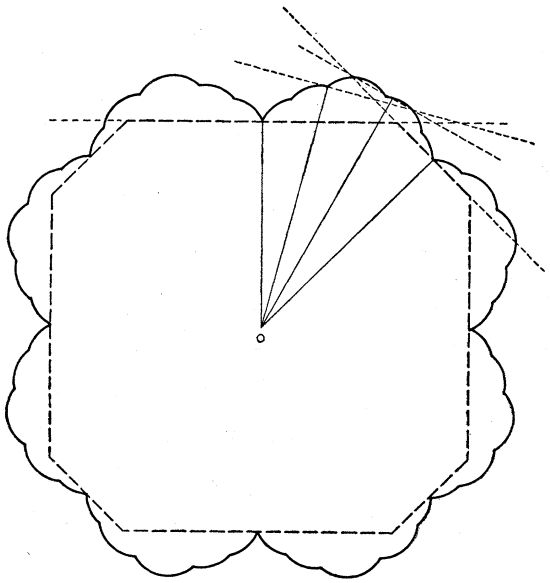


FIG. 30. Construction of the equilibrium polyhedron from the polar plot of surface free energy.

there by reflection of electrons incident from outside. This fraction clearly lies between 0 and 1, but we shall have to prove that it is in fact a reflection coefficient for ingoing electrons. After this proof has been completed we shall derive several qualitative relations regarding the velocity dependence of this r associated with the patch field.

Va. Legitimacy of Using the Mean Reflection Coefficient for Ingoing Electrons in the Velocity Distribution Equation

Consider the trajectories of the electrons mentioned in the preceding paragraph, which cross the reference plane going inward, suffer reflection in the patch field, and return outward across the reference plane with velocities in the range $dvd\omega$. Let y and z be the coordinates in the reference plane $x=x_1$. For certain values of y and z there will exist trajectories of the type just described passing outward across the plane at the point (y,z), while for other points (y,z) no such reflected trajectories will exist. Now by Liouville's theorem ((18), #19) the density of electrons in phase space must be constant along any trajectory when equilibrium conditions prevail. Therefore, the density of electrons in the region $dvd\omega$ of velocity space at any point (x_1,y_0,z_0) which lies on the outgoing part of one of these reflected trajectories must be the same as the density where the ingoing part of this trajectory crosses the reference plane, i.e., at some point (x_1,y_i,z_i) of coordinate space and some particular point of velocity space, concerning which we can say little save that it must correspond to the same kinetic energy as the final outgoing velocity. But according to the Maxwellian distribution law the ingoing density in phase space is a function of this kinetic energy alone, so that as the outgoing coordinates y_0, z_0 are varied the ingoing density in phase space, and hence the reflected density, always remain the same, as long as there exists some reflected trajectory passing through (y_0,z_0). Since the total emergent density in phase space, compounded out of emitted and reflected electrons, has this same Maxwellian value, it follows that for any given point (y_0,z_0) of the reference plane the electrons which emerge in the region $dvd\omega$ of velocity space are either all of reflected origin or all emitted. Therefore, the fraction γ of the total number of electrons in $dvd\omega$, for all values of y_0 and z_0, which have come via reflected trajectories is simply equal to the fraction of the area of the reference plane which is traversed by outgoing reflected trajectories.

In the absence of magnetic fields the electron trajectories will obey the principle of dynamical reversibility, i.e., if there exists a trajectory crossing the reference plane at (y_i,z_i) with inward velocity $\mathbf{v}_i$ which re-emerges at (y_0,z_0) with outward velocity, $\mathbf{v}_0$ then the reverse trajectory starting at (y_0,z_0) with velocity $-\mathbf{v}_0$ will emerge at (y_i,z_i) with velocity $-\mathbf{v}_i$. The fraction of the area of the reference plane for which there exist such reverse trajectories starting with velocity $-\mathbf{v}_0$ is the mean reflection coefficient r for ingoing electrons of velocity $-\mathbf{v}_0$, so we have

$$\gamma(\mathbf{v}_0)=r(-\mathbf{v}_0). \tag{V.1}$$

The classical proof just given is certainly valid for reflection in large-scale patch fields, where the motion of the electrons obeys classical mechanics; to show that (V.1) holds also for reflection phenomena of quantum-mechanical origin, such as the reflection from a uniform metal surface, it is merely necessary to translate the preceding argument into quantum-mechanical language. In quantum mechanics the probability that a system initially in a state i will be found after time t to be in the state j is

$$P(i\rightarrow j;t)=|(\exp[-iHt/\hbar])_{ji}|^2,$$

where H is the Hamiltonian operator of the system. If H is symmetric with respect to reversal of the direction of time (see (W11)), as is the case in the absence of a magnetic field, it is easily shown from this expression that

$$P(i\rightarrow j;t)=P(-j\rightarrow -i;t),$$

where $-j$ and $-i$ refer to the states derived from j and i, respectively, by reversing velocities, which in the absence of spin can be obtained by replacing the wave functions by their complex conjugates. Now if j runs over all states with outward velocity in the given range $dvd\omega$ and positions on the reference plane, there being n_0 such states in all, and if i runs over all ingoing states on the reference plane,

$$\begin{aligned}\gamma(\mathbf{v}_0)&=(1/n_0)\sum_{i,j}\int_0^\infty P(i\rightarrow j;t)dt\\&=(1/n_0)\sum_{i,j}\int_0^\infty P(-j\rightarrow -i;t)dt\\&=r(-\mathbf{v}_0),\end{aligned}$$

in agreement with (V.1).

Vb. Dependence of the Associated Reflection Coefficient on Normal Energy

Considerable insight into the behavior of the reflection coefficient associated with the patch field can be obtained from geometrical considerations in the phase space of the initial coordinates and mo-

menta of the electron, assumed incident on the patch field from without. These geometrical considerations are, of course, easiest to visualize when the number of dimensions involved is least; let us therefore consider first the simplest possible patch field, that arising from parallel strips with work functions φ_1 and φ_2. For this case the motion in the z coordinate can be separated out, and as the average of the reflection coefficient over the initial y and z coordinates of the electron is independent of the initial x coordinate, as long as this is large enough to be outside the patch field, it will suffice to consider the three-dimensional space of p_x, y, p_y, for some fixed value of x. Each point of this space determines an orbit of motion in x and y, and there will be a surface in this space separating points whose orbits strike the metal surface from points whose orbits are reflected in the patch field. We shall assume without proof that this "critical surface" has no discontinuities in slope.

Consider first the case of electrons normally incident, as in the experiments of Davisson, Ahearn, and Teal (D1). For the present case of strip patches the behavior of the measured reflection coefficient as a function of p_x will be determined by the intersection of the critical surface with the plane $p_y=0$. The upper portion of this intersection is shown schematically in Fig. 32. The measured reflection coefficient at any p_x, being an average over y, will be simply the ratio of the range PQ for which the horizontal line at height p_x lies below the critical curve, to the fundamental period Y of the patch field. If the critical curve has a nonvanishing second derivative at its peak, then for values of p_x close to the peak value p_{x0} at which the patch reflection coefficient goes to zero we must have asymptotically

$$r\sim\text{const}(p_{x0}-p_x)^{\frac{1}{2}}\sim\text{const}(\epsilon_{x0}-\epsilon_x)^{\frac{1}{2}}. \quad \text{(V.2)}$$

This is shown by the dotted curve in Fig. 33.

To get the reflection coefficient for a given p_x averaged over a Maxwellian distribution in p_y we must add a third dimension to Fig. 32. If the third coordinate is taken to be

$$\int_0^{p_y}\exp(-p_y^2/2mkT)dp_y,$$

instead of p_y, the desired reflection coefficient will be simply proportional to the area of the plane $p_x=$const. which lies beneath the critical surface. If the critical surface has a maximum at which both its second derivatives exist and are non-vanishing, then for small $(p_{x0}-p_x)$ the area in question will be asymptotically elliptical with each axis proportional to $(p_{x0}-p_x)^{\frac{1}{2}}$. Thus for this case

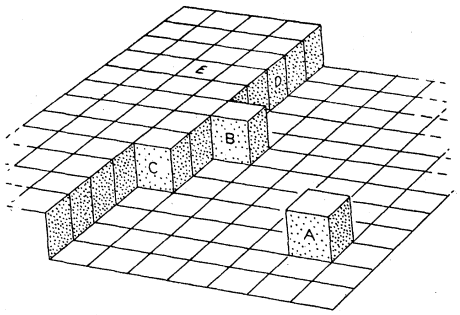


FIG. 31. Schematic illustration of the various kinds of atomic positions at the surface of a crystal.

$$r\sim\text{const}(p_{x0}-p_x)\sim\text{const}(\epsilon_{x0}-\epsilon_x) \quad \text{(V.3)}$$

as shown by the dashed curve in Fig. 33.

Note that the normal energy approximation is equivalent to assuming a discontinuous critical surface, given by $p_x=0$ for a certain range of y values, $p_x=p_{x0}$ for the rest. This gives the discontinuous reflection coefficient shown by the dot-dash line in Fig. 33.

For a more general distribution of patches, e.g., a checkerboard pattern, more dimensions must be used, but the basic ideas are the same. The simplest case is again that of normal incidence, where now the reflection coefficient is given by the fraction of the area of the y-z plane over which the plane $p_x=$const. lies beneath the critical surface. The asymptotic value of this fraction is given by an expression of the form (V.3), since the same geometrical problem is involved as in that case.

For the average reflection coefficient over a Maxwellian distribution of values of p_y and p_z, for fixed p_x, we have for the general patch field to plot the p_x coordinate of the critical surface against y, z, p_y, p_z, or better, against y, z,

$$\int_0^{p_y}\exp\left(-\frac{p_y^2}{2mkT}\right)dp_y,\quad\int_0^{p_z}\exp\left(-\frac{p_z^2}{2mkT}\right)dp_z.$$

The reflection coefficient will then be proportional to the four-dimensional volume of the portion of the hyperplane $p_x=$const. which lies below the critical surface. For small $(p_{x0}-p_x)$, this region will be asymptotically a four-dimensional ellipsoid with each axis proportional to $(p_{x0}-p_x)^{\frac{1}{2}}$. Thus for this case

$$r\sim\text{const}[(p_{x0}-p_x)^{\frac{1}{2}}]^4\sim\text{const}(\epsilon_{x0}-\epsilon_x)^2 \quad \text{(V.4)}$$

as shown by the full line in Fig. 33.

Thus for a two-dimensional regular array of patches the reflection coefficient measured in normal incidence experiments such as those of Davisson, Ahearn, and Teal should look something like the

dashed curve of Fig. 33, raised somewhat everywhere to allow for reflection at the metal surface itself, while the reflection coefficient measured by the velocity distribution method with plane-parallel geometry should look something like the full curve. The reflection coefficient directly measured by the velocity distribution method using cylindrical or spherical geometry is of course not an average reflection coefficient for a given normal energy, but rather an average for given $(\epsilon_x+\epsilon_y)$ or $(\epsilon_x+\epsilon_y+\epsilon_z)$, respectively. The behavior of averages of this sort will differ from that shown in Fig. 33 in that there will be no limiting value of $(\epsilon_x+\epsilon_y)$ or of $(\epsilon_x+\epsilon_y+\epsilon_z)$ above which the reflection coefficient of the patch field is exactly zero.

APPENDIX VI. THE "COMPLETE PHOTOELECTRIC EFFECT"

A method which has sometimes been used to determine the photoelectric threshold is that of the so-called "complete photoelectric effect." This consists in measuring the total photoelectric current obtained from a cold surface when blackbody radiation corresponding to some elevated temperature T is allowed to fall on it. There is a widespread belief that thermodynamic reasoning requires the photoelectric current density j in such an experiment to be given by

$$j=CT^2\exp(-e\varphi/kT) \qquad \text{(VI.1)}$$

where φ is the true work function of the surface as defined in (I.2.1) and C is a constant independent of T. This belief seems to have arisen from a misapplication of some early theoretical speculations of Richardson and others (R9, R10, R11, W17, R18), made in the years when the light quantum concept was relatively new and when all attention was devoted to the qualitative laws of emission rather than to precise measurements of thresholds. Actually, although a formula of the type (VI.1) is capable of representing the variation of j with T fairly well, this formula has no basic thermodynamic significance; by this is meant not merely that C and φ may be temperature dependent, but

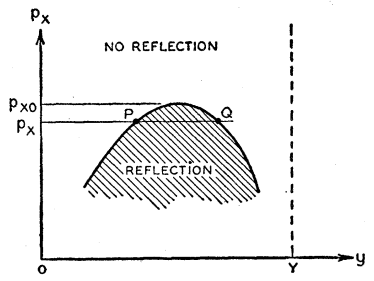


FIG. 32. Regions of reflected and non-reflected trajectories in the phase space of incident electrons.

that there does not seem to be any liklihood of relating the temperature dependence of C to any other quantity which it is feasible to measure, except by the trivial means of computing j from the measured spectral response function by integration over the blackbody spectrum.

All that thermodynamics or the principle of detailed balance can say is that if the emitting surface is in *equilibrium* with blackbody radiation at temperature T—hence itself at temperature T—then

(a) the total number of electrons leaving the surface, whether ejected photoelectrically or thermionically, is given by the Richardson equation, and
(b) the number ejected per unit time by photoelectric emission alone is equal to the number reabsorbed per unit time with simultaneous emission of a quantum of light.

If the latter number is a fraction p of the total number of electrons striking the surface, then the complete photoelectric emission can be shown by the methods of Section I.3 to be simply

$$j=pAT^2\exp(-e\varphi/kT),$$

where A is the Richardson constant 120 amp./cm^2/deg.^{2}. Equation (VI.1) would thus be correct with constant C if it were legitimate to assume that p and φ are independent of temperature, and that for a given temperature T of the light source the photoelectric yield is independent of the temperature of the target being illuminated. The former assumption, regarding the constancy of p, is probably a fair approximation for metals, if T is not too large, but the second assumption will in general be far from the truth. The complete photoelectric yield when the target temperature T_0 coincides with T will be much higher than when $T_0 \ll T$, because of the contribution of electrons which are thermally excited to levels from which they can be liberated by frequencies smaller than $\nu_0=e\varphi/h$. For the number of electrons which can be removed by a frequency $\nu<\nu_0$ will contain a factor $\exp h(\nu-\nu_0)/kT_0$, and since the tail of the Planck function for the density of the radiation contains a factor $\exp(-h\nu/kT)$, the number of such thermally excited electrons which get photoelectrically ejected may become very large, much larger in fact, then the number of electrons which are ejected from the normally occupied states. Thus $j(T,T)$ arises principally from excited electrons, while $j(T_0,T)$ arises mainly from unexcited electrons; one cannot expect any particular relationship between them, and so the hope of justifying (VI.1) by some such assumption as constancy of the ratio of these two currents must be abandoned.

The fact that experimentally Eq. (VI.1) gives reasonable values for the threshold frequency is not hard to understand, however. For consider a target

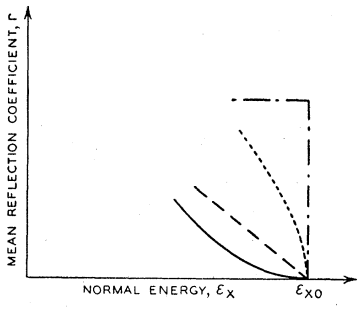


FIG. 33. Theoretical behavior of the associated reflecting coefficient of a patch field as a function of normal energy, in the neighborhood of the value ϵ_{x0} at which the associated reflection coefficient vanishes. The ordinate is the reflection coefficient for given ϵ_x, averaged over the y and z coordinates of the incoming trajectory, and, in the cases indicated, also averaged over a Maxwellian distribution of y and z velocities.

dot-dash curve: normal energy approximation
dotted curve: strip patches, normal incidence
dashed curve: strip patches, Maxwellian average or general patches, normal incidence
full curve: general patches, Maxwellian average

at 0°K, illuminated with black body radiation for which $kT \ll h\nu_0$. Then, if $F(\nu-\nu_0)$ is the spectral sensitivity function for the target, the yield will be, to a good approximation,

$$j=\int_{\nu_1}^{\infty} F(\nu-\nu_0)\cdot(8\pi h\nu^3/c^3)\exp(-h\nu/kT)d\nu,$$

where the second factor is just Wien's approximation to the Planck function. Since for metals F is proportional to $(\nu-\nu_0)^2$ when $(\nu-\nu_0)$ is small, the asymptotic form of j for small T and T_0 will be, as was pointed out long ago by Becker (B5, B6).

$$j\sim\text{const}\cdot T^3\exp(h\nu_0/kT), \qquad \text{(VI.2)}$$

a form which it would be hard to distinguish from (VI.1) on the basis of existing experimental data (S17, S18, S19, R19).

For semiconductors, where the photoelectric work function may be much higher than the thermionic, it is even more obvious that (VI.1) cannot have any thermodynamic significance.

BIBLIOGRAPHY

Books

(1) E. Bloch, *Les Phénomènes Thermioniques* (Les Presses Universitaires, Paris, 1923).
(2) P. W. Bridgman, *The Thermodynamics of Electrical Phenomena in Metals* (The Macmillan Company, New York, 1934).
(3) R. H. Fowler, *Statistical Mechanics*, Second Edition (Cambridge University Press, New York, 1936).
(4) R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, New York, 1939).
(5) E. A. Guggenheim, *Modern Thermodynamics by the Methods of Willard Gibbs* (Methuen and Company, Ltd., London, 1933).
(6) G. Herrmann and S. Wagener, *Die Oxydkathode*, 1 Teil (Johann Ambrosius Barth, Leipzig, 1943; Edwards Brothers, Inc., Ann Arbor, 1946).
(7) A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena*, (McGraw-Hill Book Company, Inc., New York, 1932).
(8) T. J. Jones, *Thermionic Emission* (Methuen and Company, London, 1936).
(9) A. L. Reimann, *Thermionic Emission* (Chapman and Hall, Ltd., London, 1935).
(10) O. W. Richardson, *Emission of Electricity from Hot Bodies* (Longmans Green and Company, London, 1921).
(11) W. Schottky and H. Rothe, *Physik der Glühelektroden*, in *Handbuch der Experimentalphysik*, v. XIII 2 (Akademische Verlagsgesellschaft, Leipzig, 1928).
(12) F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).
(13) C. J. Smithells, *Tungsten*, Second Edition (D. Van Nostrand Company, Inc., New York, 1936).
(14) W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939).
(15) A. Sommerfeld and H. A. Bethe, *Elektronentheorie der Metalle*, in *Handbuch der Physik* v. XXIV 2 (Verlag Julius Springer, Berlin, 1933).
(16) K. Spangenberg, *Wachstum und Auflösen der Kristalle*, in *Handwörterbuch der Naturwissenschaften* v. X Second Edition (Verlag Gustav Fischer, Jena, 1934).
(17) R. Suhrmann, *Elektronenemission metallishcer Leiter*, in Müller-Pouillets *Lehrbuch der Physik* v. IV 4, Chapter 3 (Druck und Verlag von Friedr. Vieweg und Sohn, Akt.-Ges., Braunschweig, 1934).
(18) R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, New York, 1938).
(19) J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities* (Oxford University Press, New York, 1932).
(20) M. Volmer, *Kinetik der Phasenbildung* (Verlag von Theodor Steinkopf, Dresden and Leipzig, 1939; Edwards Brothers, Inc., Ann Arbor, 1945).

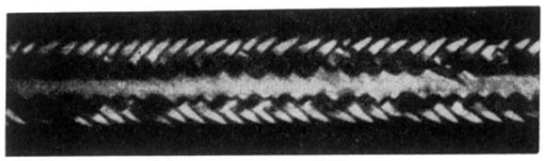
Articles

(A1) A. J. Ahearn and J. A. Becker, Phys. Rev. **54**, 448 (1938).
(A2) Paul A. Anderson, Phys. Rev. **47**, 958 (1935).
(A3) Paul A. Anderson, Phys. Rev. **49**, 320 (1936).
(A4) Paul A. Anderson, Phys. Rev. **54**, 753 (1938).
(A5) Paul A. Anderson, Phys. Rev. **57**, 122 (1940).
(A6) Paul A. Anderson, Phys. Rev. **59**, 1034 (1941).
(A7) L. Apker, E. Taft, and J. Dickey, Phys. Rev. **73**, 46 (1948).
(B1) John Bardeen, Phys. Rev. **49**, 653 (1936).
(B2) John Bardeen, J. Chem. Phys. **6**, 367 (1938).
(B3) John Bardeen, Phys. Rev. **58**, 727 (1940).
(B4) E. H. B. Bartelink, Physica **3**, 193 (1936).
(B5) A. Becker, Ann. d. Physik **60**, 30 (1919).
(B6) A. Becker, Ann. d. Physik **78**, 83 (1925).
(B7) A. Becker, Ann. d. Physik **2**, 249 (1929).
(B8) Joseph A. Becker, Rev. Mod. Phys. **7**, 95 (1935).
(B9) J. A. Becker and W. H. Brattain, Phys. Rev. **45**, 694 (1934).
(B10) J. A. Becker and G. E. Moore, Phil. Mag. **29**, 129 (1940).
(B11) M. Benjamin and R. O. Jenkins, Phil. Mag. **26**, 1049 (1938).
(B12) M. Benjamin and R. O. Jenkins, Proc. Roy. Soc. **A176**, 262 (1940).

(B13) M. Benjamin and R. O. Jenkins, Proc. Roy. Soc. **A180**, 225 (1942).
(B14) Hans A. Bethe, Phys. Rev. **59**, 940(A) (1941).
(B15) Richard P. Bien, Phys. Rev. **47**, 806 (1935).
(B16) John P. Blewett, J. App. Phys. **10**, 831 (1939).
(B17) John P. Blewett, J. App. Phys. **17**, 643 (1946).
(B18) D. Blochinzev and Sch. Drabkina, Physik. Zeits. Sowjetunion **7**, 484 (1935).
(B19) Hans Bomke, Zeits. f. Physik **90**, 542 (1934).
(B20) Hans Bomke, Zeits. f. Physik **91**, 400 (1934).
(B21) R. C. L. Bosworth, Trans. Faraday Soc. **33**, 590 (1937).
(B22) J. J. Brady and V. P. Jacobsmeyer, Phys. Rev. **49**, 670 (1936).
(B23) W. H. Brattain and J. A. Becker, Phys. Rev. **43**, 428 (1933).
(B24) A. Braun and G. Busch, Helv. Phys. Acta **20**, 33 (1947).
(B25) P. W. Bridgman, Phys. Rev. **27**, 173 (1926).
(B26) Arthur A. Brown, Phys. Rev. **60**, 170(A) (1941).
(B27) E. Brüche, Zeits. f. Physik **98**, 77 (1935).
(B28) E. Brüche and H. Johannsen, Physik. Zeits. **33**, 898 (1932).
(B29) E. Brüche and H. Johannsen, Zeits. f. Physik **84**, 56 (1933).
(B30) E. Brüche and H. Mahl, Zeits. f. techn. Physik **16**, 623 (1935).
(B31) E. Brüche and H. Mahl, Zeits. f. techn. Physik **17**, 81 (1936).
(B32) E. Brüche and H. Mahl, Zeits. f. techn. Physik **17**, 262 (1936).
(B33) W. G. Burgers and J. J. A. Ploos van Armstel, Physica **4**, 5 (1937).
(B34) W. G. Burgers and J. J. A. Ploos van Armstel, Physica **4**, 15 (1937).
(C1) Alvin B. Cardwell, Phys. Rev. **47**, 628 (1935).
(C2) Robert J. Cashman, Phys. Rev. **52**, 512 (1937).
(C3) Robert J. Cashman, Phys. Rev. **54**, 971 (1938).
(C4) R. J. Cashman and E. Bassoe, Phys. Rev. **55**, 63 (1939).
(C5) R. J. Cashman and W. S. Huxford, Phys. Rev. **48**, 734 (1935).
(C6) Joseph F. Chittum, J. Phys. Chem. **38**, 79 (1934).
(C7) K. T. Compton and I. Langmuir, Rev. Mod. Phys. **2**, 123 (1930).
(C8) K. T. Compton and L. W. Ross, Phys. Rev. **13**, 374 (1919).
(D1) C. J. Davisson, A. J. Ahearn, and G. K. Teal, as reported by Dr. Davisson at the Oak Ridge Meeting of the American Physical Society, April 9, 1948.
(D2) Charles F. De Voe, Phys. Rev. **50**, 481 (1936).
(D3) John H. Dillon, Phys. Rev. **38**, 408 (1931).
(D4) S. Dobinski and C. F. Elam, Nature **138**, 685 (1936).
(D5) Philip H. Dowling, Phys. Rev. **31**, 244 (1928).
(D6) Lee A. DuBridge, Phys. Rev. **39**, 108 (1932).
(D7) Saul Dushman, Rev. Mod. Phys. **2**, 381 (1930).
(E1) Carl Eckart, Phys. Rev. **35**, 1303 (1930).
(E2) P. Ehrenfest and A. J. Rutgers, Proc. Amst. Acad. **32**, 698 (1929).
(E3) C. F. Elam, Trans. Faraday Soc. **32**, 1604 (1936).
(F1) H. E. Farnsworth and B. A. Rose, Proc. Nat. Acad. **19**, 777 (1938).
(F2) H. E. Farnsworth and R. P. Winch, Phys. Rev. **58**, 812 (1940).
(F3) Milan D. Fiske, Phys. Rev. **61**, 513 (1942).
(F4) G. M. Fleming and J. E. Henderson, Phys. Rev. **58**, 887 (1940).
(F5) G. M. Fleming and J. E. Henderson, Phys. Rev. **59**, 907 (1941).
(F6) Ralph H. Fowler, Phys. Rev. **38**, 45 (1931).
(F7) R. H. Fowler and L. W. Nordheim, Proc. Roy. Soc. **A119**, 173 (1928).
(F8) G. W. Fox and F. M. Bailey, Phys. Rev. **59**, 174 (1941).
(F9) N. H. Frank and L. A. Young, Phys. Rev. **38**, 80 (1931).
(F10) H. Freitag and F. Krüger, Ann. d. Physik **21**, 697 (1934).
(F11) J. Frenkel, J. Phys. U.S.S.R. **9**, 385 (1945).
(F12) J. Frenkel, J. Phys. U.S.S.R. **9**, 392 (1945).
(F13) Herbert Fröhlich, Physik. Zeits. Sowjetunion **7**, 510 (1935).
(F14) K. Fuchs, Proc. Roy. Soc. **A151**, 585 (1935).
(F15) K. Fuchs, Proc. Roy. Soc. **A153**, 622 (1936).
(F16) K. Fuchs, Proc. Roy. Soc. **A157**, 444 (1936).
(G1) E. Gaviola and J. Strong, Phys. Rev. **49**, 441 (1936).
(G2) Lester H. Germer, Phys. Rev. **25**, 795 (1925).
(G3) I. Gimpel and O. W. Richardson, Proc. Roy. Soc. **A182**, 17 (1943).
(G4) H. Goldschmidt and H. Dember, Zeits. f. techn. Physik **7**, 137 (1926).
(G5) Frederick S. Goucher, Phil. Mag. **48**, 229 (1924).
(G6) Ludwig Graf, Zeits. f. Elektrochemie **48**, 181 (1942).
(G7) E. Guth and J. Meyerhöfer, Phys. Rev. **57**, 908 (1940).
(G8) E. Guth and C. J. Mullin, Phys. Rev. **59**, 575 (1941).
(G9) E. Guth and C. J. Mullin, Phys. Rev. **59**, 867 (1941).
(G10) E. Guth and C. J. Mullin, Phys. Rev. **61**, 339 (1942).
(G11) A. T. Gwathmey and A. F. Benton, J. Chem. Phys. **8**, 431 (1940).
(G12) B. Gysae and S. Wagener, Zeits. f. techn. Physik **19**, 264 (1938).
(G13) B. Gysae and S. Wagener, Zeits. f. Physik **115**, 296 (1940).
(H1) H. Haberland and W. Walcher, Zeits. f. Physik **105**, 348 (1937).
(H2) R. Haefer, Zeits. f. Physik **116**, 604 (1940).
(H3) Robert D. Heidenreich, J. Opt. Soc. Am. **35**, 139 (1945).
(H4) W. Heinze, Zeits. f. Physik **109**, 459 (1938).
(H5) W. Heinze and S. Wagener, Zeits. f. Physik **110**, 164 (1938).
(H6) H. Hellmann and W. Kassatotschkin, Acta Physicochimica **5**, 23 (1936).
(H7) H. Hellmann and W. Kassatotschkin, J. Chem. Phys. **4**, 324 (1936).
(H8) Conyers Herring, Phys. Rev. **59**, 889 (1941).
(H9) Conyers Herring, to be submitted to Phys. Rev.
(H10) C. Herring and A. G. Hill, Phys. Rev. **58**, 132 (1940).
(H11) Karl F. Herzfeld, Phys. Rev. **35**, 248 (1930).
(H12) Albert G. Hill, Phys. Rev. **53**, 184 (1938).
(H13) W. L. Hole and R. W. Wright, Phys. Rev. **56**, 785 (1939).
(H14) William V. Houston, Phys. Rev. **52**, 1047 (1937).
(I1) H. E. Ives and H. B. Briggs, J. Opt. Soc. Am. **28**, 330 (1938).
(J1) N. C. Jamison and R. J. Cashman, Phys. Rev. **50**, 624 (1936).
(J2) Z. Jeffries, Trans. Am. Inst. Min. Eng. **70**, 303 (1924).
(J3) R. O. Jenkins, Reports on Progress in Physics **9**, 177 (1943).
(J4) Ralph P. Johnson, J. App. Phys. **9**, 508 (1938).
(J5) Ralph P. Johnson, Phys. Rev. **54**, 459 (1938).
(J6) R. P. Johnson and W. Shockley, Phys. Rev. **49**, 436 (1936).
(J7) Helen Jupnik, Phys. Rev. **60**, 884 (1941).
(K1) Allen L. King, Phys. Rev. **53**, 570 (1938).
(K2) M. Knoll, F. G. Houtermans, and W. Schulze, Zeits. f. Physik **78**, 340 (1932).
(K3) R. de L. Kronig and W. G. Penney, Proc. Roy. Soc. **A130**, 499 (1931).

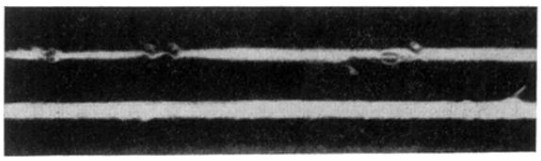
(K4) F. Krüger and G. Stabenow, Ann. d. Physik **22**, 713 (1935).
(K4a) George C. Kuczynski, Phys. Rev. **75**, 344 (1949).
(K5) H. Kurzke and J. Rottgardt, Zeits. f. Physik **100**, 718 (1936).
(K6) H. Kurzke and J. Rottgardt, Zeits. f. Physik **109**, 341 (1938).
(L1) James J. Lander, Phys. Rev. **74**, 479 (1948).
(L2) David B. Langmuir, Phys. Rev. **49**, 428 (1936).
(L3) Irving Langmuir, Phys. Rev. **8**, 149 (1916).
(L4) Irving Langmuir, Phys. Rev. **22**, 357 (1923).
(L5) I. Langmuir and K. T. Compton, Rev. Mod. Phys. **3**, 191 (1931).
(L6) I. Langmuir and K. H. Kingdon, Phys. Rev. **34**, 129, (1929).
(L7) Max v. Laue, Zeits. f. Krist. **105**, 124 (1943).
(L8) Israel Liben, Phys. Rev. **51**, 642 (1937).
(L9) H. Liebmann, Zeits. f. Krist. **53**, 171 (1914).
(L10) Ernest G. Linder, Phys. Rev. **30**, 649 (1927).
(L11) Leon B. Linford, Rev. Mod. Phys. **5**, 34 (1933).
(M1) LeRoy A. MacColl, Phys. Rev. **56**, 699 (1939).
(M2) H. Mahl, Zeits. f. Physik **98**, 321 (1935).
(M3) H. Mahl, Zeits. f. techn. Physik **17**, 653 (1936).
(M4) H. Mahl, Naturwiss. **30**, 207 (1942).
(M5) H. Mahl and J. Pohl, Zeits. f. techn. Physik **16**, 219 (1935).
(M6) H. Mahl and I. N. Stranski, Zeits. f. Metallkunde **35**, 147 (1943).
(M7) M. M. Mann and L. A. DuBridge, Phys. Rev. **51**, 120 (1937).
(M8) M. F. Manning and M. I. Chodorow, Phys. Rev. **56**, 787 (1939).
(M9) M. F. Manning and H. M. Krutter, Phys. Rev. **51**, 761 (1937).
(M10) Stuart T. Martin, Phys. Rev. **56**, 947 (1939).
(M11) Robert J. Maurer, Phys. Rev. **57**, 653 (1940).
(M12) Herbert Mayer, Ann. d. Physik **29**, 129 (1937).
(M13) C. E. Mendenhall and C. F. de Voe, Phys. Rev. **51**, 346 (1937).
(M14) Emery Meschter, Rev. Sci. Inst. **9**, 12 (1938).
(M15) Günther Mönch, Zeits. f. Physik **65**, 233 (1930).
(M16) N. Morgulis, Mem. Phys. Ukr. S.S.R. **8**, 149 (1940).
(M17) Philip M. Morse, Phys. Rev. **35**, 1310 (1930).
(M18) Harold M. Mott-Smith, Phys. Rev. **56**, 668 (1939).
(M19) B. Adalbert Mrowca, J. App. Phys. **14**, 684 (1943).
(M20) B. Mrowka, Physik. Zeits. **38**, 997 (1937) or Zeits. f. techn. Physik **18**, 573 (1937).
(M21) B. Mrowka and A. Recknagel, Physik. Zeits. **38**, 758 (1937).
(M22) Erwin W. Müller, Zeits. f. Physik **108**, 668 (1938).
(M23) Erwin W. Müller, Naturwiss. **27**, 820 (1939).
(M24) Erwin W. Müller, Zeits. f. Physik **120**, 261 (1943).
(M25) Erwin W. Müller, Zeits. f. Physik **120**, 270 (1943).
(N1) Myron H. Nichols, Phys. Rev. **57**, 297 (1940).
(N2) Myron H. Nichols, Phys. Rev. **59**, 944(A) (1941).
(N3) Albert Nitzsche, Ann. d. Physik **14**, 463 (1932).
(N4) Lothar W. Nordheim, Proc. Roy. Soc. **A121**, 626 (1928).
(N5) Lothar W. Nordheim, Physik. Zeits. **30**, 177 (1929).
(N6) Wayne B. Nottingham, Phys. Rev. **41**, 793 (1932).
(N6a) Wayne B. Nottingham, Phys. Rev. **47**, 806(A) (1935).
(N7) Wayne B. Nottingham, Phys. Rev. **49**, 78 (1936).
(N8) Wayne B. Nottingham, Phys. Rev. **57**, 935 (1940).
(N9) Wayne B. Nottingham, Phys. Rev. **58**, 927 (1940).
(N10) Wayne B. Nottingham, Phys. Rev. **59**, 906 (1941).
(O1) C. W. Oatley, Proc. Roy. Soc. **A155**, 218 (1936).
(O2) C. W. Oatley, Proc. Phys. Soc. **51**, 318 (1939).
(O3) Albert R. Olpin, as reported in (7) p. 75.
(O4) Carl F. J. Overhage, Phys. Rev. **52**, 1039 (1937).
(P1) B. Y. Pines, Sci. Pub. Ukranian Physicotechnical Inst. and State Univ., Jan. 15 (1946).
(P2) J. Pohl, Zeits. f. techn. Physik **15**, 579 (1934) or Phys. Zeits. **35**, 1003 (1934).
(P3) James G. Potter, Phys. Rev. **58**, 623 (1940).
(R1) A. Recknagel, Zeits. f. Physik **98**, 355 (1935, 1936).
(R2) A. L. Reimann, Nature **133**, 833 (1934).
(R3) A. L. Reimann, Proc. Roy. Soc. **A163**, 499 (1937).
(R4) A. L. Reimann, Proc. Phys. Soc. **50**, 496 (1938).
(R5) A. L. Reimann, Phil. Mag. **25**, 834 (1938) Supplement.
(R6) A. L. Reimann and C. K. Grant, Phil. Mag. **22**, 34 (1936).
(R7) H. C. Rentschler and D. E. Henry, J. Opt. Soc. Am. **26**, 30 (1936).
(R8) Owen W. Richardson, Phil. Trans. Roy. Soc. **201**, 497 (1903).
(R9) Owen W. Richardson, Phil. Mag. **23**, 594 (1912).
(R10) Owen W. Richardson, Phil. Mag. **24**, 570 (1912).
(R11) Owen W. Richardson, Phil. Mag. **31**, 149 (1916).
(R12) Clark S. Robinson, J. App. Phys. **13**, 647 (1942).
(R13) Albert Rose, Phys. Rev. **49**, 838 (1936).
(R14) Bernhard A. Rose, Phys. Rev. **44**, 585 (1933).
(R15) Horst Rothe, Zeits. f. techn. Physik **6**, 633 (1925).
(R16) F. Rother and H. Bomke, Zeits. f. Physik **86**, 231 (1933).
(R17) F. Rother and H. Bomke, Zeits. f. Physik **87**, 806 (1934).
(R18) S. C. Roy, Phil. Mag. **50**, 250 (1925).
(R19) S. C. Roy, Proc. Roy. Soc. **A112**, 599 (1926).
(R20) H. Rukop, W. Schottky, and R. Suhrmann, Die Physik **3**, 133 (1935).
(S1) Dankwart Schenk, Ann. d. Physik **23**, 240 (1935).
(S2) Dankwart Schenk, Zeits. f. Physik **98**, 753 (1936).
(S3) Richard W. Schmidt, Zeits. f. Physik **120**, 69 (1942).
(S4) W. Schottky, Physik. Zeits. **20**, 220 (1919).
(S5) W. Schottky, Physik. Zeits. **38**, 1024 (1937).
(S6) W. Schottky, Zeits. f. Physik **118**, 539 (1942).
(S7) R. W. Sears and J. A. Becker, Phys. Rev. **43**, 1058 (1933).
(S8) Samuel Seely, Phys. Rev. **59**, 75 (1941).
(S9) R. L. E. Seifert and T. E. Phipps, Phys. Rev. **56**, 652 (1939).
(S10) F. Seitz and T. A. Read, J. App. Phys. **12**, 470 (1941).
(S11) R. Shuttleworth, R. King, and B. Chalmers, Nature **158**, 482 (1946).
(S12) Franz Skaupy, Zeits. f. techn. Physik **5**, 563 (1924).
(S13) John C. Slater, Phys. Rev. **45**, 794 (1934).
(S14) John C. Slater, Phys. Rev. **51**, 840 (1937).
(S15) Roman Smoluchowski, Phys. Rev. **60**, 661 (1941).
(S16) Iwan N. Stranski, Ber. d. d. chem. Ges. **72**, (A)141 (1939).
(S17) Rudolf Suhrmann, Zeits. f. Physik **33**, 63 (1925).
(S18) Rudolf Suhrmann, Zeits. f. Physik **54**, 99 (1929).
(S19) Rudolf Suhrmann, Ergebnisse d. exakt. Naturwiss. **13**, 148 (1934).
(S20) R. Suhrmann and J. von Eichborn, Zeits. f. Physik **107**, 523 (1937).
(S21) N. T. Sun and W. Band, Proc. Camb. Phil. Soc. **42**, 72 (1946).
(T1) I. Tamm and D. Blochinzev, Zeits. f. Physik **77**, 774 (1932).
(T2) I. Tamm and D. Blochinzev, Physik Zeits. Sowjetunion **3**, 170 (1933).
(T3) S. R. Tibbs, Proc. Camb. Phil. Soc. **34**, 89 (1938).
(T4) Lewi Tonks, Phys. Rev. **38**, 1030 (1931).
(T5) D. Turnbull and T. E. Phipps, Phys. Rev. **56**, 663 (1939).
(U1) Newton Underwood, Phys. Rev. **47**, 502 (1935).
(V1) Max Volmer, Trans. Faraday Soc. **28**, 359 (1932).

(V2) M. Volmer and I. Estermann, Zeits. f. Physik **7**, 13 (1921).
(V3) F. C. Von der Lage and H. A. Bethe, Phys. Rev. **71**, 612 (1947).
(W1) Carl Wagner, Ann. d. Physik **6**, 370 (1930).
(W2) Hugo B. Wahlin, Phys. Rev. **73**, 1458 (1948).
(W3) Hugo B. Wahlin, Phys. Rev. **61**, 509 (1942).
(W4) H. B. Wahlin and J. A. Reynolds, Phys. Rev. **48**, 751 (1935).
(W5) H. B. Wahlin and L. V. Whitney, Phys. Rev. **50**, 735 (1936).
(W6) H. B. Wahlin and L. V. Whitney, J. Chem. Phys. **6**, 594 (1938).
(W7) Alfred H. Weber, Phys. Rev. **53**, 895 (1938).
(W8) A. H. Weber and L. J. Eisele, Phys. Rev. **59**, 473(A) (1941).
(W9) C. F. von Weizsäcker, Zeits. f. Physik **96**, 431 (1935).
(W10) Lester V. Whitney, Phys. Rev. **50**, 1154 (1936).
(W11) Eugene P. Wigner, Göttinger Nachichten, p. 546 (1932).
(W12) Eugene P. Wigner, Phys. Rev. **46**, 1002 (1934).
(W13) Eugene P. Wigner, Phys. Rev. **49**, 696 (1936).
(W14) Eugene P. Wigner, Trans. Faraday Soc. **34**, 678 (1938).
(W15) E. Wigner and J. Bardeen, Phys. Rev. **48**, 84 (1935).
(W16) E. Wigner and F. Seitz, Phys. Rev. **46**, 509 (1934).
(W17) William Wilson, Proc. Roy. Soc. **A93**, 359 (1917).
(W18) Ernst Hans Winkler, Zeits. f. Physik **107**, 235 (1937).
(W19) Rufus W. Wright, Phys. Rev. **60**, 465 (1941).
(W20) G. Wulff, Zeits. f. Krist. **34**, 449 (1901).
(Y1) Felix L. Yerzley, Phys. Rev. **50**, 610 (1936).
(Z1) W. A. Zisman and H. G. Yamins, Physics **4**, 7 (1933).

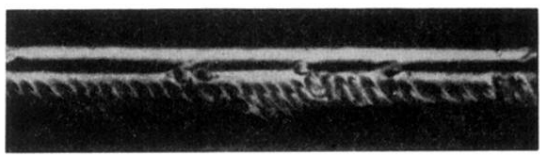


(a) 0°. (110)-Bereich.

(b) 45°. (111)-Bereich.



(c) 90°. (100)-Bereich.



(d) 135°. (111)-Bereich.

FIG. 4. Photographs of a tungsten wire showing the effect of prolonged heating in vacuum by direct current. The wire, a single crystal oriented with a (110) direction along its axis, was photographed from a number of directions around its circumference, corresponding to the normal emergence of the (110), (111), (100), and (111) directions, respectively. Taken from Schmidt (S3).*

* This figure was taken from Vol. 120 (1943) of the Zeits. f. Physik, published by Verlag Julius Springer of Berlin. The use of this figure in this article is by permission of the Attorney General of the United States in the public interest under License No. JA-1316.