SOVIET PHYSICS JETP VOLUME 23, NUMBER 5 NOVEMBER, 1966

# ON THE THEORY OF FIELD EMISSION FROM METALS

F. I. ITSKOVICH

Kharkov Military Engineering College

Submitted to JETP editor December 22, 1965

J. Exptl. Theoret. Phys. (U.S.S.R.) 50, 1425–1437 (May, 1966)

Field emission from a metal single crystal is investigated for an arbitrary electron dispersion law. If the Fermi surface is intersected by an axis $p_z$ perpendicular to the emitting surface of the sample, the free electron theory formula for the field emission current remains valid except for the pre-exponential factor. Otherwise the conservation of the tangential quasimomentum of an electron emitted from the metal leads to the result that the work function w in the exponential must be replaced by a larger quantity W. The distance between the Fermi surface and the $p_z$ axis can be estimated from the difference $W - w$, which thus yields definite information concerning the electron spectrum of the metal.

In recent years the influence of complex electron dispersion on different metal properties has been studied extensively. The theory of field emission, however, has undergone no change since the original work of Fowler and Nordheim[1] that was based on the free electron model. This omission has stimulated the present work.

The field emission current is calculated with the aid of Bloch wave functions, but the essential results depend not on these functions but only on the dispersion law. The sensitivity of the field emission current to the dispersion law is not sufficiently great to provide a basis for reconstructing the Fermi surface. In some cases we find no essential difference from the free electron theory. Nevertheless, the information derived from field emission is of definite value insofar as it pertains to the position of the Fermi surface in **p** space whereas other methods, such as the investigation of quantum oscillations in magnetic fields, yield only information concerning the shape of the Fermi surface.1)

## 1. POTENTIAL BARRIER TRANSMISSION COEFFICIENT FOR ELECTRONS IN A METAL

We know that field emission represents the tunneling penetration of electrons through the surface potential barrier of a metal placed within a high electric field ($F \sim 10^6$ V/cm or higher) oriented towards the metal. The shape of the barrier is highly dependent on the polarization of the metal by the emerging electrons. Beginning at a distance of the order of the lattice constant d between the electron and the metal surface the polarization can be computed in the form of an "image force," so that the potential energy of an electron is

$$U(z) = -eFz - e^2/4z \qquad (z \gg d) \tag{1}$$

where the metal fills a half-space $z < 0$, and e and F are positive. In a narrow layer $0 < z \lesssim d$ Eq. (1) is incorrect and U depends also on x and y (with the periodicity of the metal lattice). However, we shall neglect this and shall write the electron wave functions outside the metal as products of plane waves in the x and y directions multiplied by functions of z. Similarly, the wave functions inside the metal will be superpositions of Bloch waves.

For the purpose of determining the transmission coefficient of the barrier we must construct a wave function that for $z > 0$ contains only waves emerging from the metal, for $z < 0$ a single wave function impinging on the metal surface ($v_z > 0$), reflected Bloch waves ($v_2 < 0$), and also, for a bounded metal, Bloch waves that decay exponentially inside the metal. Let the incident Bloch wave belong to the s band, with the quasi-wave vector **k** and the energy $E = \mathscr{E}_s(\mathbf{k})$:

$$\psi_{s\mathbf{k}} = \sum_{\mathbf{q}} b_{\mathbf{q}}^{s\mathbf{k}} e^{i(\mathbf{k}+\mathbf{q})\mathbf{r}},$$

where **q** represents reciprocal lattice vectors multiplied by $2\pi$. The continuity of $\psi$ and $\partial\psi/\partial z$ at the metal surface makes it obvious that the penetrating, reflected, and exponentially decaying interior waves must be linear combinations of expressions whose dependence on x and y has the

---

1) The photoeffect in magnetic fields is an exception to some extent.[2]

form $\exp\{i(\mathbf{K}+\mathbf{Q})\cdot\mathbf{R}\}$; the capital letters denote projections of the corresponding vectors on the xy plane. The metal surface as a crystallographic plane is perpendicular to some reciprocal lattice vector. Therefore an infinite number of vectors q exist having a given projection **Q**, and the **Q** vectors form a two-dimensional lattice.

Consequently, the wave function of interest outside the metal is

$$\sum_{\mathbf{Q}} c_{\mathbf{Q}} e^{i(\mathbf{K}+\mathbf{Q})\mathbf{R}} f_{\mathbf{Q}}(z) \quad (z>0), \tag{2a}$$

where $f_{\mathbf{Q}}(z)$ is the solution of a one-dimensional Schrödinger equation with the potential energy (1) and the energy

$$E_{\mathbf{Q}}^{(z)} = E - \hbar^2(\mathbf{K}+\mathbf{Q})^2/2m_0 \tag{3}$$

($m_0$ is the mass of a free electron) representing the motion of an electron in the positive z direction. The set of reflected and exponentially damped waves must contain all Bloch functions with the same E and **K** as the incident wave, so that inside the metal we have the wave function

$$\psi_{s\mathbf{K}k_z} + \sum_r a_r \psi_{r\mathbf{K}\varkappa_r} \quad (z<0); \tag{2b}$$

here $\kappa_r$ is a root of the equation $\mathscr{E}_r(\mathbf{K}, \kappa) = E$; among the real roots we must take those corresponding to waves with $v_z < 0$, and among the complex roots we must take those with negative imaginary parts.[2)]

The joining of the two expressions for the desired wave function leads to the following system of equations for $a_r$ and $c_a$:

$$c_{\mathbf{Q}} f_{\mathbf{Q}}(0) = B_{\mathbf{Q}}^{s\mathbf{K}k_z} + \sum_r a_r B_{\mathbf{Q}}^{r\mathbf{K}\varkappa_r},$$

$$-ic_{\mathbf{Q}} f'_{\mathbf{Q}}(0) = k_z B_{\mathbf{Q}}^{s\mathbf{K}k_z} + A_{\mathbf{Q}}^{s\mathbf{K}k_z} + \sum_r a_r (\varkappa_r B_{\mathbf{Q}}^{r\mathbf{K}\varkappa_r} + A_{\mathbf{Q}}^{r\mathbf{K}\varkappa_r});$$

$$B_{\mathbf{Q}} = \sum_{q_z} b_{\mathbf{Q}q_z}, \quad A_{\mathbf{Q}} = \sum_{q_z} q_z b_{\mathbf{Q}q_z}. \tag{4}$$

A general solution of this system is impossible, aside from the fact that the Fourier coefficients of the Bloch waves are also actually unknown. Nevertheless, important information can be derived from (4) through a study of the function $f_{\mathbf{Q}}(z)$.

In the quasiclassical region to the right of the turning point $z_2$ (Fig. 1) we can write f(z) as

$$f(z) = \frac{1}{\sqrt{p_z(z)}} \exp\left\{i\left[\frac{1}{\hbar}\int_{z_2}^{z} p_z(z)\,dz + \frac{\pi}{4}\right]\right\} \quad (z - z_2 \gg \Delta z_2),$$

$$p_z(z) = \{2m_0[E^{(z)} - U(z)]\}^{1/2}$$

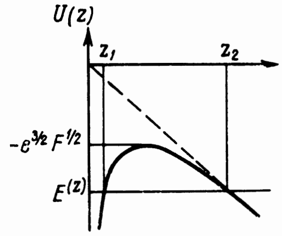

FIG. 1. Electron potential energy outside a metal. The dashed line represents the form of the barrier neglecting the image force.

with the index **Q** omitted from all quantities temporarily. Therefore, if the incident Bloch wave is normalized to a single electron per unit volume ($\sum_{\mathbf{q}} |b_{\mathbf{q}}^{s\mathbf{k}}|^2 = 1$) the transmission coefficient is

$$D = \sum_{\mathbf{Q}} |c_{\mathbf{Q}}|^2 / m_0 v_z, \tag{5}$$

where $v_z = (1/\hbar)\partial\mathscr{E}_s/\partial k_z$ is the electron velocity component in the state $\psi_{s\mathbf{k}}$.

All subsequent calculations will involve the assumption

$$\xi(E^{(z)}) \equiv \frac{2}{\hbar}\int_{z_1}^{z_2} |p_z(z)|\,dz \gg 1. \tag{6}$$

We know that this assumption makes f(z) quasiclassical inside the barrier (except near the turning point) and makes the penetration factor exponentially small. We are therefore limiting ourselves to low emission currents (far below the maximum). We shall also assume that the temperature is not too high, so that the main contribution to the current comes from electrons having energies below the Fermi level (field emission); with the given energy normalization this level is $-w$ [w (the work function) ~ 1 eV].

Because of the foregoing assumptions, $z_2 \gg d$ and Eq. (1) holds true almost everywhere within the barrier. This gives

$$\xi(E^{(z)}) = \frac{4}{3}\frac{(2m_0)^{1/2}(-E^{(z)})^{3/2}}{e\hbar F}\theta\left(\frac{e^{3/2}F^{1/2}}{-E^{(z)}}\right); \tag{7}$$

the expression preceding $\theta$ is the value of $\xi$ for an acute-angled barrier, i.e., neglecting the image force. Figure 2 is a graph of $\theta$, which was introduced and first tabulated by Nordheim; exact tables can be found in [3], Appendix I. The quantity $e^{3/2}F^{1/2}$ represents the lowering of the potential barrier that results from the image force (Fig. 1) and is responsible for the Schottky effect. It follows from (7) that the condition (6) corresponds to fields $F \ll F_0$, where $F_0$ is the field in which $\theta$ vanishes, i.e., the potential barrier disappears for an electron having the energy $E^{(z)}$, with $-E^{(z)} = 1 - 6.5$ eV, $F_0 = e^{-3}E^{(z)2} = 7 \times 10^6 - 3 \times 10^8$ V/cm. Since $\theta$ varies strongly when F is near $F_0$, the inequality $F \ll F_0$ can actually fail to be a very

---

[2)] If the equation $\xi_r(\mathrm{K}, \varkappa) = E$ has several roots of the indicated kinds, the summation in (2b) must include a term corresponding to each one.

severe condition: $\xi$ reaches the value 10 when F is smaller than $F_0$ by only a factor of two or three.

Using known formulas (in [4], for example) relating the quasiclassical wave functions on both sides of the turning point, we represent f(z) in a broad region considerably to the left of $z_2$ by

$$f(z) = \exp\{{}^1/_2\xi(E^{(z)})\}g(z) \quad (z_2 - z \sim z_2), \tag{8}$$

where g(z) is a function which in the quasiclassical region inside the barrier has the asymptotic form

$$g(z) = \frac{1}{\sqrt{|p_z(z)|}}\exp\left\{-\frac{1}{\hbar}\int_{z_1}^{z}|p_z(z)|\,dz\right\}$$

$$(z - z_1 \gg \Delta z_1,\ z_2 - z \gg \Delta z_2).$$

We shall be interested ultimately in the values of $f_{\mathbf{Q}}$ for z = 0, in Eq. (4). An analytic expression for g(z) can be given in the neighborhood to the left of $z_1$ only when a linear region of U(z) near this point is superimposed on the quasiclassical region; in this case g(z) is expressed in terms of an Airy function. In the general case we are limited to $g(0) \sim 1/\sqrt{p_z^0}$, where $p_z^0$ is a characteristic momentum (at the potential barrier) that does not contain large parameters which could compete with $e^{\xi/2}$. Any attempt to learn more than the order of magnitude of g(0) would go beyond the limits of accuracy since the potential actually depends on x and y near the metal surface.

It follows from (8) that the desired quantities $c_{\mathbf{Q}}$ appear in (4) only in the product form $c_{\mathbf{Q}} \exp\{{}^1/_2\xi(E_{\mathbf{Q}}^{(z)})\}$. The small exponential factor can be separated by means of the substitution

$$c_{\mathrm{Q}} = \exp\{-{}^1/_2\xi(E_{\mathrm{Q}}^{(z)})\}\tilde{c}_{\mathrm{Q}} \tag{9}$$

The new unknowns $\tilde{c}_{\mathbf{Q}}$ are determined from equations that differ from (4) through the replacement of $f_{\mathbf{Q}}(0)$ and $f'_{\mathbf{Q}}(0)$ by $g_{\mathbf{Q}}(0)$ and $g'_{\mathbf{Q}}(0)$ and therefore contain no parameters that could compete with $e^{-\xi/2}$. Equations (9) and (7) show that $c_{\mathbf{Q}}$ increases extremely sharply as $E_{\mathbf{Q}}^{(z)}$ increases. This is fully understandable physically since $E_{\mathbf{Q}}^{(z)}$ is the energy of electron motion along the z axis.

It is easily seen that $E_{\mathbf{Q}}^{(z)}$ in Eq. (3) has a maximum as a function of **Q**. If **K** designates the reduced [lying in the central (first) Brillouin zone of the two-dimensional lattice of **Q** vectors] tangential quasiwave vector, then this maximum is reached at **Q** = 0. Therefore if the smallest nonvanishing vector **Q** is not too small (for which it is required that the crystallographic indices of the z = 0 plane be not too large), all $c_{\mathbf{Q}}$ with $\mathbf{Q} \neq 0$ will be exponentially small compared with $c_0$.[3] Consequently a single term is strongly dominant in the superposition of waves going out of the metal: Eq. (2a) is practically reduced to

$$c_0 e^{i\mathbf{KR}} f_0(z) = \tilde{c}_0 e^{i\mathbf{KR}} g_0(z)\cdot \quad (z > 0).$$

In other words, among all the equivalent values of the tangential quasimomentum (it will henceforth be more convenient to use the quasimomentum **p** rather than the quasi-wave vector **k**) the potential barrier "releases" from the metal only the smallest value as the quasimomentum is transformed into momentum. The formula (5) for the transmission coefficient is simplified correspondingly:

$$D_s(\mathbf{p}) = C_s(\mathbf{p}) \exp\{-\xi(E - P^2/2m_0)\},$$

$$C_s(\mathbf{p}) = |\tilde{c}_0|^2/m_0 v_z. \tag{10}$$

The exponential in the transmission coefficient (10) is actually the same as that given by the free electron model. This results from the fact that the exponential is determined by the behavior of the wave function inside the barrier (outside of the metal). However, the energy of electron motion along the z axis, appearing in the exponential, is expressed for the general case in terms of E and **P** values that are conserved in escape from the metal ($E_0^{(z)} = E - P^2/2m$); for a free electron the energy can also be expressed in terms of $p_z$ within the metal. We note that the conservation of the reduced tangential electron quasimomentum, on which, as will be shown, all essential results of the present work are based, is derived from from the translational symmetry of the Hamiltonian in the xy plane, i.e., it is not associated with the assumed schematization of the potential form in the surface layer and requires only that the metal surface be an ideal atomic plane.

To determine the factor preceding the exponential in (10) we must solve (4) and also know $v_z$; this is impossible for the general case. A calculation is possible only in the nearly free electron approximation (of course, neglecting the dependence of the potential on x and y for $z \lesssim d$). For states whose tangential momentum lies in the two-dimensional central Brillouin zone, we then obtain $C \sim 1$;[4] for all other states C contains a small

---

[3] Only when **K** is close to the boundary of the two-dimensional central Brillouin zone will one or more values of $c_{\mathrm{Q}}$ be of the same order of magnitude as $c_0$; this is not reflected in the subsequent results.

[4] The exact value of C depends, of course, on the shape of the potential barrier extremely close to the metal surface. Specifically, if there is a quasiclassical region to the left of the turning point $z_1$ and this region is superposed on the linear behavior of U(z) near $z_1$, then we have C = 1.

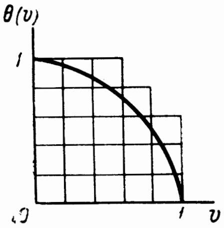

FIG. 2

parameter in the approximation of loosely bound electrons. Consequently, we can accept $C_s \sim 1$ for the general case.

## 3. CALCULATION OF THE FIELD EMISSION CURRENT

To calculate the current emitted from a metal we must sum the electron currents [using the Fermi function f(E)] over all states in a completely orthonormalized system of functions. Although Bloch waves for different values of s and **p** are mutually orthogonal, the wave functions (2a)–(2b) constructed in the preceding section do not comprise an orthogonal system as a general rule. Indeed, if we have several different Bloch waves with $v_z > 0$ and identical values of E and **P** (due to the superposition of bands or simply the fragmented form of the constant-energy surface), the corresponding wave functions will contain the same reflected wave [Eq. (2b)] and will therefore not be mutually orthogonal as a general rule.

We shall first consider the case in which this degeneracy is absent from filled levels. The emission current density is then

$$j_z = -\frac{2e}{h^3} \sum_{s} \int_{(v_{zs}>0)} v_{zs}(\mathbf{p}) D_s(\mathbf{p}) f(\mathscr{E}_s(\mathbf{p}))\, d^3\mathbf{p}$$

integrated over a unit cell of the reciprocal lattice. A transformation to the variables $p_x$, $p_y$, an and E gives

$$j_z = -\frac{2e}{h^3} \int \Phi(E) f(E)\, dE, \quad \Phi(E) = \int_{\Sigma(E)} D(E, \mathbf{P})\, d^2\mathbf{P}, \tag{11}$$

where Σ(E) is the portion of the projection of the constant-energy surface on the $p_x$–$p_y$ plane that lies in the central Brillouin zone of the two-dimensional lattice; Σ(E) is symmetric about the coordinate origin. These formulas remain valid in the case of a degeneracy if the wave functions (2a)–(2b) form an orthogonal system, but in this case D(E, **P**) must be understood to mean $\Sigma_l D^{(l)}(E, \mathbf{P})$ summed over all states with the given values of E and **P**. Finally, in the general case, by orthogonalizing the wave functions corresponding to determinate values of E and **P** we again obtain (11); here D(E, **P**) is quadratic in the quantities $[D^{(l)}(E, \mathbf{P})]^{1/2}$ and its matrix is determined by the reflection coefficients $a_r$ of the orthogonalized functions. Since for fixed values of E and **P** all $D^{(l)}(E, \mathbf{P})$ contain the same exponential [see Eq. (10)], this also appears in D(E, **P**):

$$D(E, \mathbf{P}) \approx \exp\{-\xi(E - P^2/2m_0)\}; \tag{12}$$

the pre-exponential factor that has been omitted here is of the same order of magnitude as $C_s$ in (10). Thus in the most general case the field emission current is determined by (11), (12), and (7).

The exponential in D(E, **P**) has such a large absolute value [see Eq. (6)] that the essential contribution to the integral Φ(E) comes from a small region of Σ(E) where P is close to its minimum $P_{min}(E)$.[5] Linearizing the exponential in this region and integrating over polar coordinates, we obtain

$$\Phi(E) \approx S(E) \exp\{-\xi(E - \Delta(E))\},$$

$$\Delta(E) = P_{min}{}^2(E)/2m_0, \quad S(E) = m_0 \varepsilon(E) \varphi(E),$$

$$\varepsilon(E) = -\frac{1}{\xi'(E - \Delta(E))}$$

$$= \frac{2}{3} \eta\left(\frac{e^{3/2} F^{1/2}}{-E + \Delta(E)}\right) \frac{-E + \Delta(E)}{\xi(E - \Delta(E))}, \tag{13}$$

$$\eta(v) = \left[1 - \frac{2}{3} \frac{\theta'(v)}{\theta(v)} v\right]^{-1}, \quad \varphi(E) = \int_0^{t_{max}(E)} \varphi_E(\{2m_0[\Delta(E) + \varepsilon(E)t]\}^{1/2}) e^{-t} dt,$$

$$t_{max}(E) = [P_{max}{}^2(E) - P_{min}{}^2(E)]/2m_0\varepsilon(E); \tag{14}$$

here $\varphi_E(P)$ is the sum of parts of a circle (in radians) having the radius P from its center at the coordinate origin; the summed parts lie within Σ(E) (Fig. 3a). Since $\eta(v) \approx \theta(v)$ [as v increases from 0 to 1 the ratio $\eta(v)/\theta(v)$ decreases from 1 to 0.9] we have $\epsilon(E) \ll -E + \Delta(E)$.

Let us consider the function S(E). In the case of small groups (pockets), when $t_{max} \ll 1$, S equals the area of Σ. In the opposite case S equals in order of magnitude the area of the part of Σ that is cut out by a circle of radius $\sqrt{2m_0(\Delta + \epsilon)}$ about

---

[5] The only exception is the case in which the function $\varphi_E \sqrt{2m_0(\Delta + \epsilon t)}$ of Eq. (14) increases steeply for $t \gg 1$. This occurs when Σ(E) is reduced to a few regions of which those closest to the origin are extremely small. Furthermore, since these small regions make only a small contribution to Φ(E) they can be neglected, thus reducing the calculation to the ordinary case. In subsequent formulas $P_{min}$ will then be taken as the value $\tilde{P}_{min}$ pertaining to the large regions that are more distant from the coordinate origin (Fig. 3b, c).

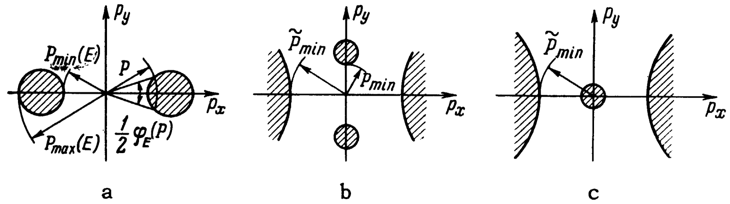

FIG. 3. Projections $\Sigma(E)$ of the constant-energy surface on the $p_x$-$p_y$ plane (shaded).

the origin. The requirement of a small pocket can be written as

$$P_{max} - P_{min} \ll a\sqrt{2m_0} \quad a = \sqrt{\Delta + \varepsilon} - \sqrt{\Delta} \leqslant \sqrt{\varepsilon}. \tag{15}$$

It follows that the number of electrons (holes) per atom in this pocket is much smaller than $\nu = 10^{-2}\alpha^3\gamma$, where $\gamma$ is the ratio between the volume occupied by the pocket in p space and $(P_{max} - P_{min})^3$; the concentration of atoms is taken to be $10^{23}$ cm$^{-3}$, and $\Delta$ and $\epsilon$ are measured in eV.

Equations (19) and (20) will be used to show that reasonable values of the current correspond to $\xi \sim 10$ and accordingly $\epsilon \sim 10^{-2} - 10^{-1}$ eV, so that $\nu < (10^{-5} - 10^{-3})$. On this basis quadratic dispersion can be assumed for small pockets; it follows that

$$S(E) = \pm 2\pi nm(E - E_g), \tag{16}$$

where n is the number of projections of identical pockets on the $p_x$–$p_y$ plane at a minimum distance from the origin, $m = \sqrt{m_1 m_2}$ is the effective mass, and $E_g$ is the energy at which the pockets become degenerate (disappear). Here and henceforth the upper sign refers to electrons and the lower sign to holes. We note that for isotropic pockets $P_{max} - P_{min} = \sqrt{\pm 8m(E - E_g)}$ in (15).

In the case of large groups ($t_{max} \gg 1$, or $P_{max} - P_{min} \gg \alpha\sqrt{2m_0}$) only the interval $0 < t \lesssim 1$ makes a large contribution to the integral $\varphi(E)$, i.e., S is determined by a small part of the region $\Sigma$ near $P_{min}$. If in this interval $\varphi_E\sqrt{2m_0(\Delta + \epsilon t)} = \text{const}$ (Fig. 4), then $\varphi(E) = \varphi_E$. If $\Sigma(E)$ includes a sufficiently large area in the neighborhood of the origin (Fig. 4a), then $\varphi(E) = 2\pi$.

In another typical situation the boundary of $\Sigma$ in the region that is important for S can be approximated by arcs of two symmetric circles, whose radius R will be taken as positive when they are convex toward the origin; the case of $R < 0$ is shown in Fig. 5. For this approximation $\beta = R(R + P_{min})/m_0\epsilon$ must be large as a general rule; for $R > 0$ this is equivalent to the condition $t_{max} \gg 1$. Consequently, at least one of two inequalities must hold true: $\beta \gg \Delta/\epsilon$ or $\Delta/\epsilon \gg 1$. If $\beta \gg \Delta/\epsilon$ (which is equivalent to $|R| \gg P_{min}$) we have

$$\varphi(E) = 2\pi e^{\Delta/\varepsilon}\left[1 - \Psi\left(\sqrt{\frac{\Delta}{\varepsilon}}\right)\right] \simeq \begin{cases} 2\sqrt{\pi\varepsilon/\Delta} & (\Delta/\varepsilon \gg 1), \\ 2\pi & (\Delta/\varepsilon \ll 1), \end{cases}$$

$$\Psi(v) = \frac{2}{\sqrt{\pi}}\int_0^v e^{-t^2}\,dt. \tag{17}$$

If $\Delta/\epsilon \gg 1$, then

$$\varphi(E) = 2[\pi\varepsilon R / \Delta(R + P_{min})]^{1/2} \ll 1 \tag{18}$$

and the reverse is also true: If $\varphi(E) \ll 1$, then $\Delta/\epsilon \gg 1$. For $\beta \ll \Delta/\epsilon$, $R \ll P_{min}$ (if $R > 0$) or $-(R + P_{min}) \ll P_{min}$ (if $R < 0$).

We now proceed directly to calculate the current. At zero temperature we write the first equation of (11) as

$$j_z = -\frac{2e}{h^3}\int_{E_{min}}^{-w} \Phi(E)\,dE$$

(recalling that w is the work function). Because the integrand varies extremely steeply due to the large absolute value of its exponential [see Eq. (6)], the essential contribution to this integral comes from a small interval $\sim \delta E$ around the maximum of $\Phi(E)$. Therefore the field emission current is

$$j_z = -(2e / h^3)\Phi(E_M)\delta E, \tag{19}$$

where

$$\Phi(E_M) = \max_{E \leqslant -w} \Phi(E) \approx S(E_M)e^{-\xi(-W)},$$

$$\xi(-W) = \frac{4}{3}\frac{(2m_0)^{1/2}}{e\hbar F}W^{3/2}\theta\left(\frac{e^{3/2}F^{1/2}}{W}\right), \quad -W = E_M - \Delta(E_M).$$

The location of the maximum of $\Phi(E)$ is determined practically only by its exponential index

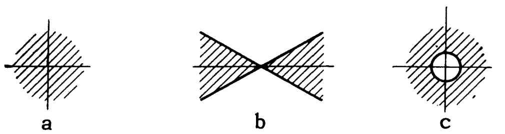

FIG. 4. Different forms of $\Sigma(E)$ for which $\varphi_E = \text{const}$.

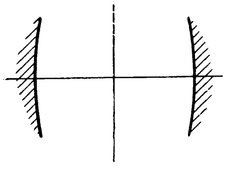

FIG. 5. The form of $\Sigma(E)$ for $R < 0$.

$-\xi(E-\Delta(E))$, and since $-\xi$ is an increasing function of its argument, this position is determined by the quantity $E-\Delta(E)$.[6] Therefore

$$-W = \max_{E\leqslant -w}[E-\Delta(E)] = \max_{T=0} E^{(z)}. \tag{20}$$

Equations (19) and (20) have very clear physical meanings; the principal contribution to field emission comes from electrons whose energy of motion in the z direction is nearly maximal. It is easily seen from (20) that

$$w \leqslant W \leqslant w + \Delta(-w), \tag{21}$$

so that $W = w$ only when $\Delta(-w) = 0$, i.e., $\Sigma(-w)$ contains the origin (particularly in the free electron model) and $S(-w)$ is extremely small.[7] Therefore the principal difference between (19) for the field emission current obeying any dispersion law and the special case of free electrons consists in replacing the work function w of the exponential by an "effective work function" W (with $W = w$ under the aforementioned conditions).

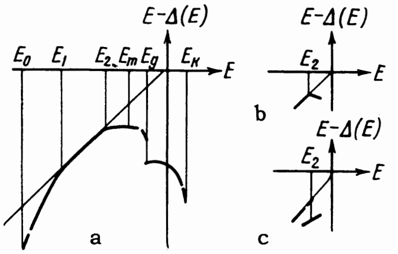

FIG. 6. $E-\Delta(E)$ vs. E in a single band. The dashed segments of the curve represent the result obtained by neglecting the considerations in footnotes 5) and 6).

To determine the effective work function we shall first investigate the variation of the quantity $E-\Delta(E)$ within a single band (Fig. 6). At the beginning of the filled band $E = E_0$ the form of $\Sigma(E)$ grows in size from merely one or a few points; at the end of the band ($E = E_K$) it contracts again to one or a few points, so that each point of the two-dimensional central Brillouin zone is included in $\Sigma(E)$ within a definite energy interval. Let $[E_1, E_2]$ be the interval within which $\Sigma(E)$ includes the coordinate origin and $\Delta(E) \equiv 0$. (In a particular case $E_1$ can coincide with $E_0$ and $E_2$ with $E_K$). A decrease of $\Delta(E)$ obviously occurs in the interval $[E_0, E_1]$; therefore $E-\Delta(E)$ increases in the range $E_0 < E < E_2$. Beginning at $E_2$, $\Delta(E)$ increases; then $E-\Delta(E)$ can either have a maximum at $E_2$ [if $\Delta'(E_2) > 1$ (Fig. 6b) or if $E_2 = E_g$ (Fig. 6c)], or can continue to increase [if $\Delta'(E_2) < 1$ (Fig. 6a)]. In the latter case $E-\Delta(E)$ passes through a maximum at some value $E = E_m < E_K$, corresponding to hole pockets or an open energy surface, because at the disappearance point of the hole pocket $\Delta'(E_g) = \infty$ [if $\Delta'(E_g) \neq 0$]. Thus the maximum of $E-\Delta(E)$ within the band is reached either at $E_2$ or $E_m$, or at $-w$ if the band is partially filled and $\Delta'(-w) < 1$. To determine $E_M$ we need only to select the largest of the maxima of all bands containing electrons (taking footnote [6] into account); we emphasize that this maximum does not necessarily lie in the highest band.

We shall now consider two possible variants.

1. $E_M = -w$, $W = w + \Delta(-w)$. If $\Delta(-w) = 0$ [$\Sigma(-w)$ contains the origin], then, as already mentioned, $W = w$, i.e., the exponential in the current expression (19) is the same as in the free electron model, independently of the dispersion law. The pre-exponential factors are determined by the following formulas: If $\Sigma(-w)$ includes a sufficiently large neighborhood of the coordinate origin, then $S(-w) = 2\pi m_0\epsilon$ and $\delta E = \epsilon$ [$\epsilon$ denotes $\epsilon(-w)$ here and henceforth].[8] If there is a small pocket ($\mu\zeta \ll \epsilon$), then $S(-w) = 2\pi m\zeta$ and $\delta E = {}^1\!/_2\,\zeta$ (for $\zeta \ll \epsilon$), $\delta E \gg \epsilon$ (for $\zeta \gg \epsilon$) for electrons; $\delta E = \epsilon(\epsilon/\zeta + 1)/(1+\mu)$ for holes. [Here $\zeta = \pm(-w - E_g)$ is the chemical potential measured from the bottom of the group; $\mu = m/m_0$; quadratic dispersion is assumed in the entire energy interval that is important for the current; $\Sigma(E)$ is a circle about the origin.]

If $\Delta(-w) \neq 0$ then $\Delta(E)$ has a square root singularity at $E = E_g$ and two cases must be distinguished for the pre-exponential factor:

a) In the interval $-w - E \lesssim \delta E$ we have $S(E) \approx \text{const}$ and $\Delta(E)$ is expanded in powers of $E + w$. For a group with isotropic quadratic dispersion this occurs when $(1+\mu)\zeta + 2\sqrt{\mu\Delta(-w)\zeta} \gg \epsilon$ (for electrons) or $(1+\mu)\zeta \gg \epsilon$ (for holes); this is always fulfilled for large groups, while for small groups it can be fulfilled only for electrons with $\mu \ll 1$. We also have $\delta E = 1/\chi'(-w)$ if $\chi'(-w) \gg [\chi''(-w)]^{1/2}$, and $\delta E = [-\pi/2\chi''(-w)]^{1/2}$ if $\chi'(-w) \ll \sqrt{-\chi''(-w)}$, $\chi''(-w) < 0$. [If $\chi''(-w) > 0$

---

[6] The pre-exponential function $S(E)$ plays an important role only when the maximum of the exponential is found in the vicinity of $E_g$, where $S(E)$ is extremely small [see Eq. (16)]. However, this maximum of the exponential is not a maximum of $\Phi(E)$; to be rid of the former automatically, in calculating max $[E-\Delta(E)]$ we exclude the energy intervals where $S(E) \approx 0$.

[7] In accordance with footnote [6], if $S(-w) \approx 0$ the inequality $W \leqslant w + \Delta(-w)$ may be incorrect.

[8] $\delta E < \epsilon$ only for an electron group with $\zeta \lessapprox \epsilon$, $m \gg m_0$ (an unreal case).

and $\chi'(-w) \ll \sqrt{\chi''(-w)}$, then $E_M \neq -w$.] Here $\chi(E) = -\xi(E - \Delta(E))$, so that $\chi' = (1 - \Delta')/\epsilon$, $\chi'' \approx \Delta''/\epsilon$; for isotropic quadratic dispersion

$$\Delta' = \mp\sqrt{\mu\Delta/\zeta}, \quad \Delta'' = (\mu + \sqrt{\mu\Delta/\zeta})/2\zeta > 0.$$

b) The energy interval making the essential contribution to the current is found in the region of small pockets.[9] For electrons $\delta E = {}^1\!/_2\,\zeta$ when

$$(1+\mu)\zeta + 2\sqrt{\mu\Delta(-w)\zeta} \ll \varepsilon$$

while $\delta E = \epsilon(\epsilon/\zeta + 1)$ for holes.

2. $E_M = E_m$, $W = -E_m + \Delta(E_m)$. For hole groups with isotropic quadratic dispersion

$$E_m = E_g - \mu\Delta(E_m) = E_g - \mu\Delta(E_g)/(1+\mu)^2.$$

a) $E_m$ is in the large-group region. For hole groups with isotropic quadratic dispersion this occurs if $\Delta(E_g) \gg (1 - 1/\mu)\epsilon$. Also,

$$\delta E = \sqrt{\frac{\pi}{-2\chi''(E_m)}} \quad \text{for} \quad -w - E_m \ll \frac{1}{\sqrt{-\chi''(E_m)}}$$

and

$$\delta E = \sqrt{\frac{2\pi}{-\chi''(E_m)}} \quad \text{for} \quad -w - E_m \gg \frac{1}{\sqrt{-\chi''(E_m)}}$$

b) The energy interval making the important contribution to the current is in the region of small hole pockets (see footnote [9]). The current is

$$j_z = -\frac{2e}{h^3}\Phi(\tilde{E}_m)(\varepsilon + \zeta)e^{1-\zeta/\varepsilon} \sim$$
$$-(2e/h^3)2\pi nm\varepsilon^2 \exp\{\chi(E_m)\},$$

and, since $\zeta \ll \epsilon$, we have $\exp\{\chi(E_m)\} \approx \exp\{\chi(-w)\} \approx \exp\{\chi(E_g)\}$.

3. $E = E_2$, $W = -E_2$. Different cases are possible for the topology of $\Sigma(E)$ at the point $E_2$ (Fig. 7): a) An opening can appear in $\Sigma(E)$ having the shape of an ellipse centered at the origin, if the energy surface has a conical point at $E_2$ and the axis of the cone is the $p_z$ axis; in this case $\delta E > \epsilon$, with $\delta E \sim \epsilon$ if $\mu - 1 \gtrsim 1$ ($\mu_1 = m_1/m_0$, $m_1$ is the smaller effective mass of the ellipse);
b) a gap in the hyperbolic bridge connecting the two parts of $\Sigma(E)$ at the origin, if at $E_2$ the energy surface has a conical point through which the $p_z$ axis passes outside of the cone; c) "sliding" of the parts of $\Sigma(E)$ so that they overlap at the origin; there can be any even number of these parts, rather than only two as shown in Fig. 7c;
d) disappearance of the part of $\Sigma(E)$ containing the origin if $E_2$ is the point of "collapse" of the corresponding hole group; this case has actually been considered in variant 1 [$\Delta(-w) = 0$]. We note that cases a)–c) correspond to Fig. 6b, and case d) to Fig. 6c. The formulas for the pre-exponential factors are not given here because they are too numerous and complicated.

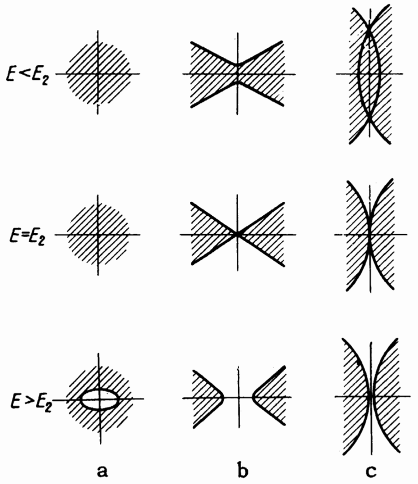

FIG. 7

### 3. DISCUSSION OF RESULTS

We shall now summarize our results. The principal contribution to field emission consists of electrons occupying a small region in **p** space where the energy $E^{(z)}$ of motion along the z axis is near its maximum W at T = 0. Therefore W is the effective work function in the exponential of the current expressed by Eq. (19). Because of the conservation of the reduced tangential electron quasimomentum during emission from a metal, W coincides with the work function w only when the projection of the Fermi surface on the $p_x$–$p_y$ plane includes a sufficiently large neighborhood of the coordinate origin (a situation that occurs in the free electron model). In the opposite case we have W > w and the electrons that will be emitted are not necessarily adjacent to the Fermi level, but can be found much lower, even outside of the conduction band (cases 2 and 3 of the preceding section), because $E^{(z)}$ can reach its maximum nonsimultaneously with the total energy E.

Equation (19) shows that the effective work function is the only parameter of the electron spectrum that can be determined fairly accurately from field emission measurements; for this purpose it is convenient to utilize the electric field (F) dependence of the current (see [3], Appendix II). We shall

---

[9] For isotropic hole pockets this occurs when $\mu \ll 1$, $\mu\Delta(E_g) \ll \epsilon$. Under these conditions the point $E = E_m \approx E_g - \epsilon$ at which $\Phi(E)$ has a maximum lies in the small-pocket region. In this case $\tilde{E}_m$ does not coincide with $E_m$, since the behavior of $\Phi(E)$ is strongly influenced by the steeply decreasing pre-exponential function S(E) (compare with footnote [6]).

now discuss what information can be derived from this quantity when the work function is known.

When W equals w the Fermi surface is intersected by the $p_z$ axis.10) When W > w the Fermi surface is not intersected by the $p_z$ axis; the smallest distance between them is

$$p_{min}(-w) \geqslant \sqrt{2m_0(W-w)}$$

[see Eq. (21)].11) Equality occurs in case 1; in case 2, $W = -E_m + \Delta(E_m)$; in case 3, $W = -E_2$.12) Thus by using samples having different orientations of the emitting surface relative to the crystallographic axes,13) we can acquire definite information about the Fermi surface of the metal. This information is, of course, quite limited and is not alone sufficient for a plot of the Fermi surface, but the knowledge can be used to test and improve any existing model of that surface. On the other hand, the data derived from an investigation of field emission possess a decided advantage insofar as they characterize the arrangement of the Fermi surface relative to the axes of **p** space (and therefore its topology to some extent), whereas other methods determine only its shape (see footnote 1), however).

We shall now present some examples of the interpretation of experimental field emission data. If we have W = w for all z directions, we can only affirm that the Fermi surface is intersected by all straight lines passing through the coordinate origin. In addition to the simplest case, where the Fermi surface has a closed cavity around the origin, other variants are possible, such as Fig. 8a. If W > w for a few directions of the z axis and W = w for all other directions, the Fermi surface has the form shown in Fig. 8b (as in the case of copper); the width of the neck can be computed from the difference W − w. If W = w for numerous directions of the z axis and W > w for the remaining directions, the Fermi surface consists of small pockets (Fig. 8c).

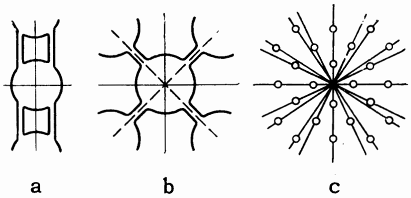

FIG. 8. Examples of Fermi surface topology reconstructed from experimental field emission data. The solid lines are the directions of the $p_z$ axis for which W = w, and the dashed lines for W > w.

The usual experimental technique for investigating field emission when the emitter is a small single crystal emitting electrons from different faces is very suitable for studying the directional dependence of the work function. In tungsten, molybdenum, tantalum, nickel, and rhenium considerable anisotropy is observed (pages 109–111 of [3], where the original studies are referred to, and p. 117 with regard to calculations of W for several faces of tungsten). Since the theory of field emission based on the free electron model arrives at no difference between the effective and ordinary work functions, the observed anisotropy has been ascribed to the work function. In actuality this anisotropy could result mainly from differences between the effective work function and w depending on the form of the Fermi surface.14)

Since the interpretation of field emission data is impossible without knowledge of the work function for the different faces, the latter quantity must be determined independently. One method of accomplishing this consists in measuring the contact potential between the metal of interest and a standard metal for specific faces of both metals. The required work function of the standard metal can be determined by means of the same field emission by selecting a face where W = w (for which purpose the Fermi surface must be known). It will be shown in our next communication that thermionic emission and the threshold frequency of the photoeffect are characterized by values of the effective work function that differ in general from both w and W.

In conclusion, I wish to thank I. M. Lifshitz and M. Ya. Azbel' for discussions of this work, and G. E. Zil'berman and I. O. Kulik for their interest and for valuable discussions.

---

10) We consider here and henceforth a Fermi surface that is extended periodically in p space. In accordance with footnotes 5) and 6), we disregard cavities in the Fermi surface on the $p_x$-$p_y$ plane.

11) According to footnote 7) (or 10)) this inequality may be incorrect if the projection of the Fermi surface on the $p_x$-$p_y$ plane is extremely small.

12) An independent source of information must be available to determine which one of cases 1 – 3 occurs. For example, if it known that a metal contains only electron groups case 1 exists.

13) We recall that the Miller indices of the plane that is the metal surface should not be too large.

14) The Fermi surfaces of tungsten, molybdenum, and chromium constructed by the augmented plane wave method [5] have a cavity around the origin. If this represents reality, W should coincide with w for all faces of these metals.

---

[1] R. H. Fowler and L. W. Nordheim, Proc. Roy. Soc. (London) **A119**, 173 (1928); L. W. Nordheim, Proc. Roy. Soc. (London) **A121**, 626 (1928) and Physik. Z. **30**, 177 (1929).

[2] G. E. Zil'berman and I. O. Kulik, JETP **38**, 1188 (1960), Soviet Phys. JETP **11**, 859 (1960).

[3] M. I. Elinson and G. F. Vasil'ev, Avtoélektronnaya émissiya (Field Emission), Fizmatgiz, 1958.

[4] D. Bohm, Quantum Theory, Prentice-Hall, New York, 1951 (Russ. transl., Fizmatgiz, 1961).

[5] T. L. Loucks, Phys. Rev. **139**, A1181 (1965).

Translated by I. Emin
171