Chapter 1
Basic Principles

Abstract This chapter describes the basic principles of Fowler–Nordheim theory of field emission from metals and classical Morgulis-Stratton theory of field emission from semiconductors. Then it discusses the theoretical basics of Müller field emission microscope as the first experimental device which allowed to test the main conclusions of the classical theory of field electron emission. Disadvantages of these classical theories and limits to their applicability are discussed.

1.1 Introduction

Structures comprised of charged particles (electrons or ions), such as solids, liquids gases or plasma, are all natural emitters of charged particles. Of them, it is easy to assess the maximum electron current density for metals. As in normal conditions the electron current onto the surface of metal body has density of $j = e n_0 V$ (where $e = 1.6 \times 10^{-19}$ C is the electron charge; $n_0 = 10^{22} \div 10^{23}$ cm^{-3} is the concentration of electrons; V—velocity, which for electrons of Fermi energy is equal to 10^8 cm/s), the maximum possible value of current density is a colossal $j = 10^{12}$ A/cm^2. But in reality only a small fraction of all the electrons are emitted out of metal due to significant brake forces that emerge on passing of the electron through the surface. For classic metals the work function is about $2 \div 6$ eV, and there are not a lot of electrons that have energy exceeding this work function—in normal conditions, in fact, their amount is negligible. The standard method of increasing the number of emitted electrons is giving additional energy to conducted electrons such that they can overcome the work function.

Such additional energy can be applied by some external impact such as: electromagnetic radiation (photo-electron emission, PEE), bombardment with high-energy particles (secondary electron emission, SEE) etc. Nowadays the most widely used method is, however, the method of thermos-electron emission (TEE),

i.e. emission of electrons from warmed up solids. There is yet another type of electron emission that does not require applying of additional energy, which is thus called field-electron emission (FEE).[1]

The effect of FEE appears when an extremely strong electric field (with field strength $E \sim 10^7$ V/cm) is applied to the surface of a solid resulting in transformation of the potential threshold into a potential barrier which is low and, most importantly, narrow enough to allow the electron tunneling i.e. the electrons are able to phase through this barrier and leave the solid without loss of energy.

The phenomenon of FEE was not theoretically explained up until the discovery of quantum mechanical tunneling effect (e.g., [1, 2]). Soon after this discovery, about 1928–1929, Fowler and Nordheim have constructed the first well-founded theory of FEE in metals [3, 4]. Further evolution of FEE theory was mostly based around development of new models and computation techniques.

As to the theory of semiconductor FEE, one has to admit that up to now such a theory that would be general enough, logically sound and mathematically and physically consistent, does not exist [5–7]. The same applies to FEE from carbon structures [8].

In order to understand the phenomenon of autoelectronemission (same as in case of thermoelectron and other types of emission) it is necessary to understand the properties of the system of electrons in solid crystals that serve as emitters. The most important such properties are the characteristics of the electrons' energy spectrum (zonal structure of the crystal), defined by laws of quantum mechanics, and the energy distribution of electrons in permitted zones that are defined by statistical laws. These questions are thoroughly studied is a series of excellent monographs and textbooks (such as, e.g. [1, 5, 6, 9–12]), and thus we are not going to pursue those problems in too much detail within confines of this book.

Instead we are going to focus mostly on studying and analysis of phenomena on phase boundaries between metal (or semiconductor) and vacuum and between metal (semiconductor) and dielectric, that define the phenomenon of FEE itself and its circumstantial processes.

Basic dependence of emission current density on intensity of the emission-causing impact (temperature, lighting, electric field intensity etc.) for any kind of emission can be calculated if we manage to evaluate the integral

$$ j = e \int\limits_{-\infty}^{+\infty} dp_y \int\limits_{-\infty}^{+\infty} dp_z \int\limits_{-\infty}^{+\infty} f(p_x, p_y, p_z) \mathscr{D}(p_x, E) p_x m^{-1} dp_x, \qquad (1.1) $$

where e is the electron charge, f is the function of electron distribution in the metal by pulse components (p_x, p_y, p_z); $\mathscr{D}(p_x, E)$ is the potential barrier penetrability factor (penetrability) in presence of an external electric field at the metallic surface

---

[1]Other names for the effect of field electron emission (FEE) include autoelectron emission (AEE), cold emission, electrostatic emission etc.

for the electrons falling on the surface of metal with impulses p_x along the normal to the surface coinciding with axis x, and E is electric intensity. If $\mathscr{D}$ depends only on p_x, then integration of (1.1) over p_y and p_z reduces the problem under consideration to one-dimensional one and transition from variable p_x to $\mathscr{E}_x$ results in

$$j = |e| \int\limits_{0}^{+\infty} \mathscr{D}(\mathscr{E}_x, E, \Phi) n(\mathscr{E}_x) d\mathscr{E}_x, \tag{1.2}$$

where $n(\mathscr{E}_x)d\mathscr{E}_x$ is an electron flux with energies in the interval $d\mathscr{E}_x$ about $\mathscr{E}_x$ from within the cathode; Φ is the work function known to be caused by the energy spent to remove an electron from the solid (or the fluid) in the vacuum.

The value of $n(\mathscr{E}_x)$ is calculated from Fermi-Dirac distribution from the ratio

$$n(\mathscr{E}) = \frac{4\pi mkT}{h^3} \ln\left[1 + \exp\left(-\frac{\mathscr{E}}{kT}\right)\right],$$

where h is Plank constant.

As $n(\mathscr{E})$ is given, it is possible to obtain basic emission dependence for the case of FEE $j = j(E)$ via the integral (1.2) if the value of $\mathscr{D}(\mathscr{E}_x, E)$ is known or has been determined.

As noted above, electrons are retained inside a solid, metal or semiconductor by electric forces and cannot leave it by themselves. This situation is formalized by introduction of a concept of a potential threshold/barrier for the solid–vacuum boundary.

The potential barrier can be defined as a region of space where the potential energy of a particle (in this case an electron) assumes a value exceeding that of its total energy. Putting this in mathematical terms, the potential barrier is represented as a functional dependence of the potential energy of particle/electron and its coordinate. Visually the potential barrier can be represented as an arbitrary potential curve, for an instance the one shown in Fig. 1.1 as an instance. The region where energy U is distinct from zero is finite. It stretches from $x = a$ to $x = b$. The region is referred to as a potential barrier ($\mathscr{E} < U$).

In classical mechanics [13] typically a particle cannot get into the potential barrier region without some external work function against potential field forces. In quantum mechanics, however, it is typical for a particle to penetrate a potential

Fig. 1.1 Arbitrary potential barrier
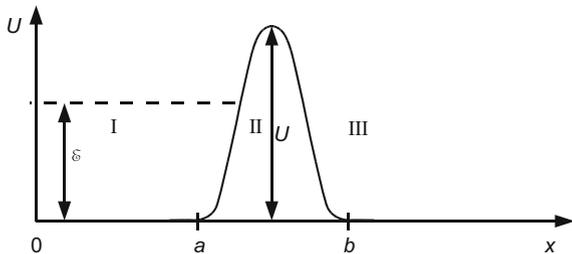
barrier even when it does not get any supplementary energy (tunnel effect). Classic statistical physics often represents thermal energy exchange between particles as an energy source necessary to overcome the potential barrier. The exchange is similar to the one taking place, for example, in thermionic emission. If at the "summit" of the potential barrier $\mathscr{E} = U_{\max}$, the total energy of the particle $\mathscr{E}$ is only an insignificant part of the total energy of the entire closed system in the thermal equilibrium state, the probability of penetration through the potential barrier is directly proportional to the Boltzmann factor $\exp[-(U_{\max} - \bar{\mathscr{E}})/kT]$, where $\bar{\mathscr{E}}$ is the average energy of the particle, T is the absolute temperature, k is the Boltzmann constant [14].

In quantum mechanics [15], penetration through a potential barrier can take place even in a purely conservative system not located in a thermostat nor exchanging energy with any other objects. In the "classically unattainable" region $U > \mathscr{E}$, the wave function of the particle is distinct from zero (although it decreases exponentially towards the potential barrier region). If the region dimensions are finite, the particle can penetrate through it. In a one-dimensional case, when the "*classically unattainable*" (the so-called subbarrier) region is limited with two points x_1 and x_2, whose locations are identified by the equation $U(x) = \mathscr{E}$, the probability of penetration of a particle/electron through a potential barrier (barrier-transmission factor, barrier penetrability) is defined as

$$\mathscr{D} = \exp\left\{ -2\left| \int_{x_1}^{x_2} k(x)dx \right| \right\}, \qquad (1.3)$$

where

$$k = \frac{\sqrt{2m[\mathscr{E} - U(x)]}}{\hbar}$$

is the wave number of the particle, m is its mass and $\hbar$ is the Planck constant. Expression (1.3) holds true on condition that within the distance of one particle wave length from any cusp decomposition of potential energy U into the difference $(x - x_{1,2})$ can be done through linear approximation. Away from the cusps, quasi-classical approximation of quantum mechanics is applicable everywhere [1]. Given these conditions, the transmission factor $\mathscr{D}$ is exponentially small (in classical mechanics it is zero). For this type of potential barrier, the particle wave properties can also be significant in the subbarrier region of values $\mathscr{E} > U$. In classical mechanics, this region would demonstrate free passage of particles over the barrier in the direction of their movement and the reflectivity factor $R = 1 - \mathscr{D}$ would become zero. In quantum mechanics, it remains finite for any value of $\mathscr{E}$, though, generally, it is exponentially small in the suprabarrier region.

Let us go back to Fig. 1.1. Obviously, the shape of the potential barrier shown there is nowhere near the actual situation at the solid-vacuum interface. In absence
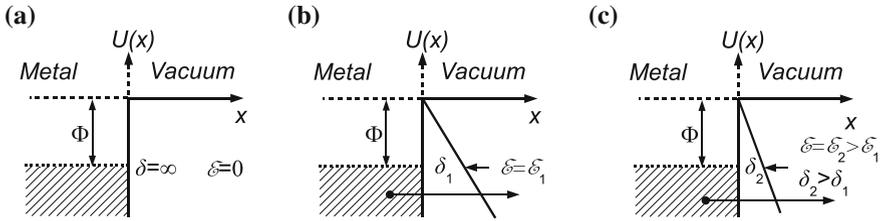
Fig. 1.2 Field electron emission. Change the width of the potential barrier with increasing electric field strength. $U(x)$—the potential energy of the electron

of an external electric field at the boundary there is a potential threshold rather than a potential barrier (cf. Fig. 1.2a). One can also imagine that in this case, electrons are retained in the metal by a semi-infinite potential barrier. Electron emission only appears when electrons can leave the metal by passing the threshold due to the energy received before. In particular, the energy can be received as a result of heating the metal in the same way it happens in thermionic emission. The situation is radically different if the metal/semiconductor surface is subjected to a strong electric field. The potential threshold becomes a potential barrier. The higher the electric field intensity, the narrower is the barrier (Fig. 1.2b, c). As a result, metal electrons have a theoretical possibility to leave the metal due to the tunnel effect [1].

The triangular potential barrier represented in Fig. 1.2b, c, is closer to the actual situation, and Fowler and Nordheim even used it in developing their first theory to evaluate potential barrier penetrability [3]. Nevertheless, it is also highly idealistic. Just like the electrons in the shell of any atom cannot leave the atom under usual conditions (i.e. without an external action) due to the forces of attraction to the nucleus, *free* electrons of a solid cannot (as we already mentioned) leave the solid under usual conditions due to the decelerating field forces at the boundaries of the solid. As established, these forces have a dual origin. Firstly, they are caused by the fact that the electron leaving a solid polarizes electron gas (the term of *polarization* has been used in a general sense, i.e. it includes both electrostatic induction for conduction electrons and intra-atomic polarization for bound electrons). Secondly, these forces are caused by double electric layers on clean surfaces of solids. From the quantum-mechanical point of view, the double layers are explained by "buckling" of ψ-functions on the crystal surface. From the classical point of view double layers are caused by the fact that some *free* electrons leave the crystal surface and develop an equilibrium electron cloud near the surface, thus creating a responding deficiency of electrons on the surface.

Double layers also emerge when alien substances are adsorbed onto the surface of a solid. Depending on their nature, they can form electric double layers of various values and signs.

As escape of an electron obviously needs expenditure of energy to deform the electron gas, there is less energy at the solid-vacuum interface before than after the escape of the electron irrespective of the real shape of the potential barrier.

The value characterizing the height of the potential barrier at the solid boundary is the work function Φ.

Work function is defined as the minimum work necessary to remove an electron from a metal. In this definition, the work function equals the difference in energies of two states of the entire crystal. In the initial state, an electrically neutral metal with N electrons is in its basic state with energy $\mathscr{E}_N$. In the final state, one electron has been removed from the metal and has only electrostatic energy corresponding to the level of vacuum $\mathscr{E}_{vac}$. The metal with remaining $N-1$ electrons passes into a new state $\mathscr{E}_{N-1}$. Hence, the work function can be described as

$$\Phi = \mathscr{E}_{N-1} - \mathscr{E}_{vac} - \mathscr{E}_N.$$

For thermodynamic change of the state, the difference $\mathscr{E}_{N-1} - \mathscr{E}_N$ can be represented as a derivative of free energy F with respect to the number of particles at constant temperature and constant volume V. The derivative $\left(\frac{\partial F}{\partial N}\right)_{T,V}$ is an electrochemical potential Γ of electrons. At nonzero temperatures it is equal to the Fermi level $\mathscr{E}_F$:

$$\mathscr{E}_{N-1} - \mathscr{E}_N \rightarrow \left(\frac{\partial F}{\partial N}\right)_{T,V} = \Gamma$$

or

$$\left(\frac{\partial F}{\partial N}\right)_{T,V} = \mathscr{E}_F$$

Thus, the work function can generally be expressed as:

$$\Phi = \mathscr{E}_{vac} - \mathscr{E}_F.$$

The work function from clean surfaces mostly comes down to work against polarization forces. As for double layers, they are either small (e.g. in alkaline metals) or compensated by high enough kinetic energy of electrons in the substance. As for semiconductors or dielectrics, we have not found any discussion of double layers on their surfaces in the literature yet. From the classical point of view, the relatively low concentration of "free" electrons makes the double layers very small.

The quantum-mechanical calculations made for metals show that the polarization forces outside the metal come down to the classical "mirror force" (polarization force) introduced by Schottky. According to him, the force retaining an electron in a metal is caused by the fact that the escaping electron induces a distributed positive charge in the metallic surface, and the field thus generated between the electron and the induced charge is fully equivalent to the field between the electron and the positive charge equal to the electron charge and located in the point of the electron image (Fig. 1.3), i.e. $F = \frac{e^2}{4x^2}$. However, according to Schottky considers this only

Fig. 1.3 On the calculation
of the mirror image force
(polarization forces): *e* charge
of the emitted electron, +*e* its
mirror image
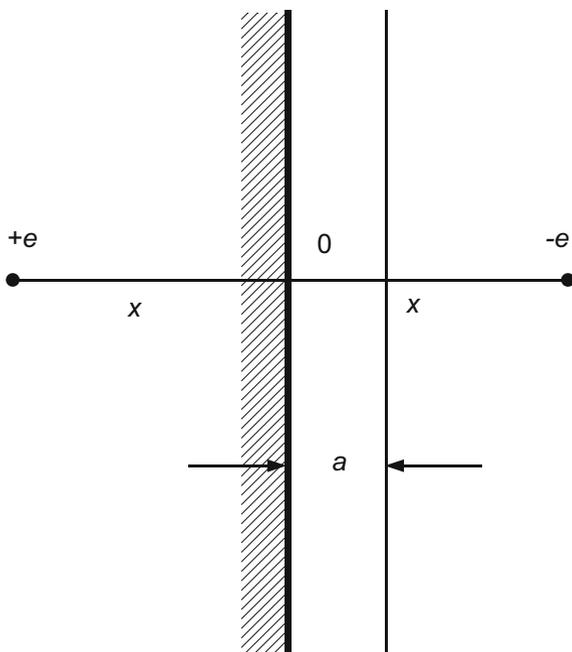
holds starting with some distance $x \geq a$ nearly equivalent to the constant of the metal lattice. At the distances $0 < x < a$, the expression does not hold because of the influence of the atomic structure of the surface. Schottky believes that at distances $0 < x < a$, the polarization force can be regarded as constant and equal to the image force in the point $x = a$. Therefore, the total work function of the metal $\Phi = e\varphi = \int_0^\infty F dx$ turns out to be equal to

$$\Phi = e\varphi = \frac{e^2}{2a}. \tag{1.4}$$

This expression cannot be directly applied to semiconductors and dielectrics.

If an electron escapes from some depth of a dielectric or semi-conductor crystal, then we can show that the force of interaction between the escaping electron and the positive charge it leaves is negligible compared with the force arising due to intra-atomic polarization of the substance by the escaping electron. This force is equivalent to the force of interaction between the electron and its "*dielectric image*". To estimate the work function for this case, we can go along the same line of reasoning and derive the formula

$$\Phi = e\varphi = \frac{e^2}{2a} \frac{\varepsilon - 1}{\varepsilon + 1}. \tag{1.5}$$

Here the value of dielectric constant ε shows only the electronic part of polarization because the orientation part does not have enough time to manifest itself within the electron escape period. The expression (1.5) for semiconductors and dielectrics determines the value of their electron affinity. As it will be clear from the subsequent, it is this value that matters for consideration of field emission.

If the surface of a metal or a semiconductor is covered with an adatom layer, and the adatom is polarized and supplied with a dipole moment either due to interaction between adatoms and the main substance or for some other reasons, then the work function is changed by the value of potential energy surge caused by the dipole layer (Fig. 1.4).

In that case, the work function can be expressed as the sum of two summands. The first summand is connected with the bulk properties of the metal and determined by the electrostatic potential in the depth of the bulk. In this case, the Fermi level counts as zero. In terms of a potential diagram (Fig. 1.5), it is $U(-\infty)$. The second summand is a term sensitive to the surface properties and expressed by the difference of electrostatic potential energies:

$$\Delta U = U(\infty) - U(-\infty) = 4\pi e \int_{-\infty}^{\infty} x[n(x) - n_+(x)]dx,$$

Fig. 1.4 The decrease in the work function $e\varphi$ of the magnitude $\Delta(e\varphi)$ for the adsorption of substances that give a positive dipole moment. The *dashed line* shows the resulting dipole
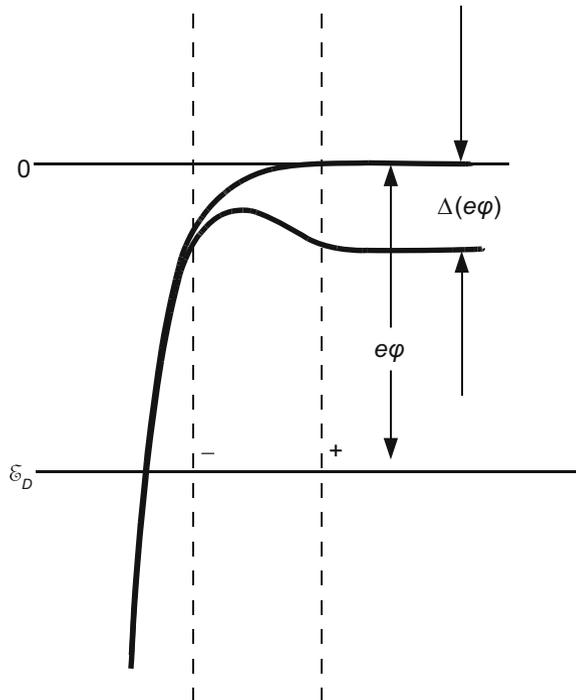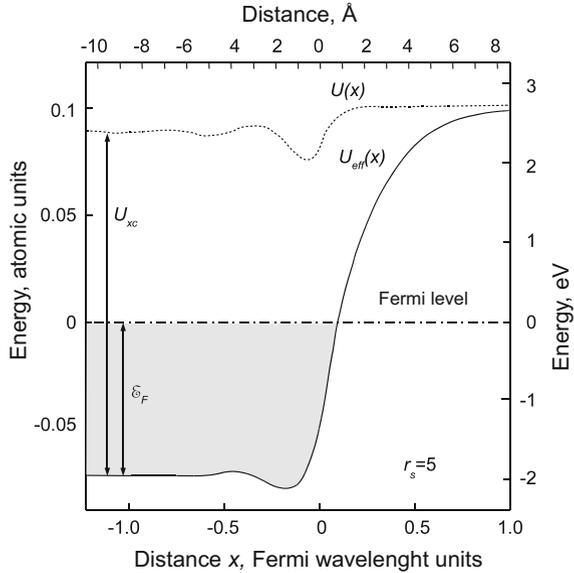
Fig. 1.5 The spatial distribution of the electrostatic potential $U(x)$ (*dashed line*) and the total effective one-electron potential U_{eff} (*solid line*) near the surface in the jelly model for $r_s = 5$ [16]

where $n(x)$ is the electron charge density distribution, and $n_+(x)$ is the positive charge density distribution.

This term is referred to as an electrostatic dipole barrier because it corresponds to the work that has to be done to transfer an electron through the surface dipole layer. The dipole moment value is a surface characteristic, which changes from surface to surface. In one and the same metal, closely packed (atomically smooth) faces usually have a larger dipole moment than loosely packed (atomically-rough) faces do. It is understandable at the qualitative level if we imagine a rough surface as containing some frame ions biased to the region of the charged negative layer, which reduces the dipole moment, and consequently, the work function.

As a rule, the above also means that electron emitter surfaces are not homogeneous in terms of work function. Inhomogeneity is caused either by the fact that the surface is formed by different faces of a single crystal, or the different degree of coverage with adatoms, or the polycrystalline nature of the surface. As a result, the contact potential difference causes development of patches of accelerating and decelerating fields (spot fields) on the surface. However, when a strong enough electric field is applied to the emitter surface, relatively long-range spot fields cease to matter, while the local work functions of particular surface patches remain significant.

Field emission Field emission is emission of electrons in the case when the emitter boundary has a strong external electric field applied to it, which accelerates electrons from the surface. A potential wall at the solid-environment interface is turned into a barrier whose thickness depends on the field applied (the stronger the field, the thinner is the barrier) (see Fig. 1.1). In this case, non-excited electrons can tunnel this barrier, generating current whose density is a characteristic of this

process. It is clear that the field emission current density largely depends on the field intensity at the surface.

Thermal-field emission Strictly speaking, emission of non-excited electrons in the course of tunneling takes place at zero temperature. If $T > 0$, some electrons have levels of energy $\mathscr{E} > \mathscr{E}_0$ For excited electrons, the potential barrier becomes lower and, respectively, narrower, which increases their barrier-transmission factor. This leads to growth of field emission current in comparison with the previous case. So thermal-field emission is caused by both the processes of transmission of electrons through a potential barrier, and the processes of their thermal excitation.

Photo-electron emission By analogy with thermal-field emission, electrons can be additionally excited by irradiation of the emitter surface with light quanta. This is extremely important in case of a semiconductor emitter since this excitation transfers valence-band electrons to the conduction band. Imposition of the electric field results in a combined photo-field emission.

Having considered the main types of electron emission, we can now present a potential barrier model closest to the actual situation and, consequently, most often used to calculate potential barrier penetrability before calculation of the basic dependence of field emission $j = j(E)$ [integral (1.2)].

As noted above, electron emission without pre-excitation (field emission) can only take place as a result of a tunnel effect if a strong electric field is applied to the surface, whereby the potential threshold (an infinite potential barrier) turns into a finite-width potential barrier. A triangular potential barrier is based on the specificity of field emission, i.e. availability of an electric field. However, this barrier model is too rough. By now, the most complete theory of field emission from metals
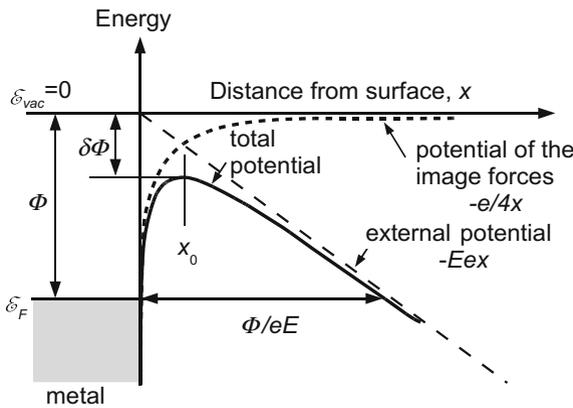

Fig. 1.6 The diagram of the potential energy of the electron near the metal surface in the presence of an external electric field with electric field strength *E*. The total potential (*solid line*) is the sum of the image potential (*dotted line*) and the potential of the applied field (*dashed line*). Φ—work function in the absence of an applied field. Reducing the potential for value $\delta\Phi$ because the Schottky effect. x_0—position of the maximum total potential

has been developed for the so-called polarization force or electrical image-force barrier. A potential diagram explaining this model is presented in Fig. 1.6.

In subsequent paragraphs we will take it upon ourselves to develop a strict quantum-mechanical assessment of the field emission phenomenon. In conclusion of this paragraph, however, we have to mention that this phenomenon can also be explained in terms of classical physics.

Actually, the external-field force acting on an electron is accompanied by an attraction force from its electrostatic image in the metal. Effective potential energy including interaction with the image is formulated as

$$U_{\text{eff}}(x) = U_0 - eEx - \frac{e^2}{4x},$$

where x is the distance to which the electron moves from the metallic surface, the distance between the electron and its image being $2x$. The function of $U_{\text{eff}}(x)$ peaks at the point $x_0 = \sqrt{e/4E}$, the potential energy at the peak being

$$U_{\text{max}} = U_{\text{eff}}(x_0) = U_0 - \sqrt{e^3E}.$$

So inclusion of electrostatic-image forces shows that electric field imposition actually reduces the work function. If we meet the condition that either

$$\mathscr{E}_F \geq U_0 - \sqrt{e^3E}$$

or

$$\sqrt{e^3E} \geq \Phi,$$

where $\mathscr{E}_F$ is the Fermi level, Φ is the work function, then the electrons will be free to leave the metal. According to classical views, the ratio $\sqrt{e^3E} \geq \Phi$ gives us the electric field intensity at which field emission (e.g. for silver) is $E \geq 10^8$ V/cm. Meanwhile, experimental data (e.g. [17]) show that field emission can be observed for fields at least an order of magnitude smaller. Hence, classical views cannot adequately explain the field emission phenomenon.

1.2 Fowler–Nordheim Theory of Field Emission from Metals

The classic Fowler–Nordheim theory of field emission from metals (e.g. [3–6, 8, 913–15, 17–20,]) is not only of historical interest. Mainly, this theory describes the process of field emission from metals adequately in terms of quality and precisely enough in terms of quantity despite considerable simplifying assumptions included into the physical and mathematical models of the phenomenon.

Before drilling down the central tenets of the Fowler–Nordheim theory, let us state that any physical theory is a compromise between a strict physical model, adequately selected calculation method and a possibility to obtain a result in the simplest way and an analyzable form and turn our attention to its basic physical prerequisites and assumptions used to simplify evaluations:

(1) the metal—vacuum interface is considered an ideal plane: the evaluation problem becomes one-dimensional, the potential function (energy) $U(x)$ depends only on the x-coordinate, the external field is homogeneous;
(2) inside the metal $U_1(x) = \text{const} = -U_0$; outside the metal, the potential barrier is caused only by activity of image forces $U_2(x) = e^2/4x$, where e is electron charge;
(3) barrier penetrability D is computed by the semi-classic Wentzel-Kramers-Brillouin (WKB) method; subsequent improvements of the theory of field emission from metals were mostly connected to application of penetrability computation methods exceeding the WKB method in validity;
(4) the metal model selected was the Sommerfeld model of free electrons in a potential box evolving degenerated gas obeying Fermi-Dirac statistics;
(5) the Fowler–Nordheim theory was developed for the temperature $T = 0$ K, i.e. for mere field emission.

The current density in the case of field emission can be calculated by the usual formula that is also suitable for calculating current density for other types of emission, such as thermionic, photo-electron etc.:

$$j = e \int_0^\infty n(\mathscr{E}_x)\mathscr{D}(\mathscr{E}_x, E)d\mathscr{E}_x, \tag{1.6}$$

where $n(\mathscr{E}_x)\,d\,\mathscr{E}_x$ is the number of electrons with energy ranging between $\mathscr{E}_x$ and $\mathscr{E}_x d\mathscr{E}_x$ falling from the inside of the metal onto the surface unit per second; $\mathscr{E}(\mathscr{E}_x, E)$ is a potential-barrier transmission factor at the metallic surface equal to the ratio of intensity of the electronic wave passing through the barrier to the incident wave intensity. Naturally, $\mathscr{D}$ depends on both $\mathscr{E}_x$ and the field value E:

$$\mathscr{D}(\mathscr{E}_x, E) = 1 - R(\mathscr{E}_x, E), \tag{1.7}$$

where R is the electron wave reflection factor.

It should also be stated that unlike thermionic emission, when we only have to count the electrons whose energy exceeds the height of the potential barrier at the metal-vacuum boundary, in case of field emission, we have to count all electrons falling on the barrier because theoretically all of them are able to leave the metal, and it is only probability that is different.

In the general case of arbitrary E and T, evaluation of current density j depending on the field E, temperature T and work function ($\Phi = e\varphi$) by the formula (1.6) is a great theoretical challenge. Therefore, it is expedient to break the entire energy

spectrum into several regions and apply particular mathematical methods to each region. Figure 1.7 shows an energy-level diagram for a metal and a potential barrier at the metal-vacuum boundary in absence and presence of electric field. When $T = 0$, energies of all electrons are concentrated in Region A. $T > 0$ causes electrons to appear in the Regions B, C and D. Region B electrons take part in field emission along with Region A electrons. Within Region B, there is a small subgroup B_1, where electron energies get close to the maximum of potential barrier. Obviously, barrier penetrability for Region B electrons is higher than that for Region A electrons. The electron emission caused by electrons in Region B is referred to as thermal-field emission. Region C and D electrons cause thermionic emission, Region C electrons are responsible for the Schottky effect and Region D electrons cause emission without an external electric field. (For the sake of simplicity, Figs. 1.7 and 1.8 show a potential energy curve passing through the point $-U$ at $x = 0$. Actually, it passes through this point at x_k determined from the equality $U = \mathscr{E}_{x_k} + \frac{e^2}{4x_k}$).

Figure 1.8 shows the image force barrier in presence of an external accelerating electric field E. Then, for a potential electron energy V in the presence of an external field, the following ratios are valid

$$
\begin{aligned}
V &= -U && \text{at } x < 0, \\
V &= -\tfrac{e^2}{4x} - eEx && \text{at } x \geq a.
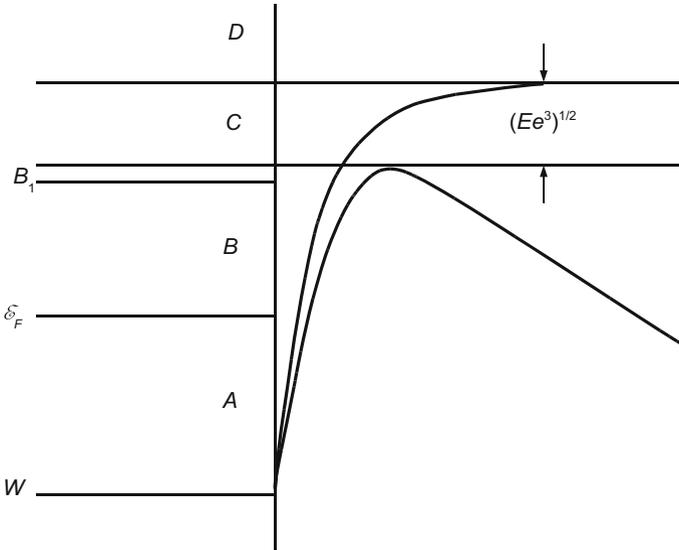\end{aligned}
\tag{1.8}
$$

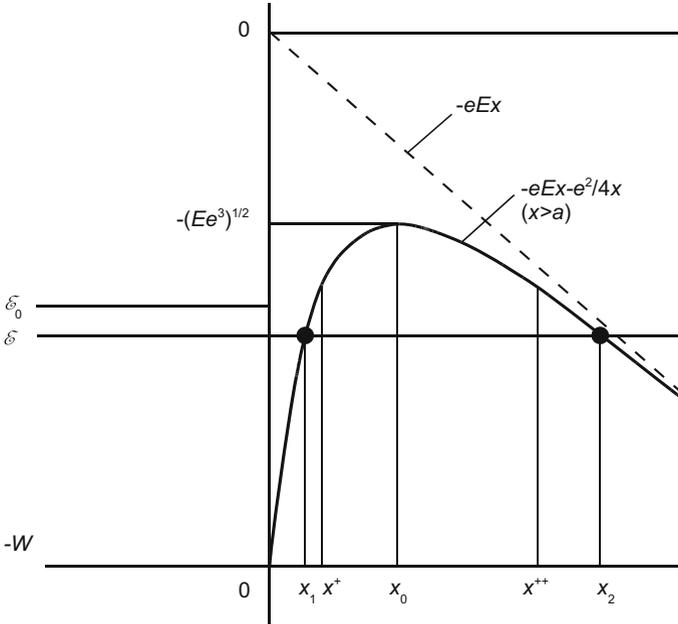**Fig. 1.7** Typical areas of electron energy interval

Fig. 1.8 To derivation of the basic formula of field electron emission

In this case, the barrier summit is located below the vacuum level by the value of $\sqrt{e^3 E}$ and at a distance from the metal; $x_0 = \frac{1}{2}\sqrt{\frac{e}{E}}$; x_1 and x_2 are the distances from the metallic surface at which the set electron energy E is equal to potential energy.

The Schrödinger equation for the barrier shown in Fig. 1.8 is given by

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2\eta^2}\left[\mathscr{E} - \left(-\frac{e^2}{4x} - eEx\right)\right]\psi = 0. \tag{1.9}$$

By definition, the barrier penetrability is

$$\mathscr{D} = \frac{|a_t|^2}{|a_i|^2} \quad \text{or} \quad \mathscr{D} = \frac{|a_i|^2 - |a_r|^2}{|a_i|^2}, \tag{1.10}$$

where a_i, a_r, a_t are the amplitudes of incident, reflected and transmitted waves, respectively. Let

$$\Phi(x) = k_0^2\left[\mathscr{E} - \left(-\frac{e^2}{4x} - eEx\right)\right], \tag{1.11}$$

where $k_0^2 = \frac{2m}{\hbar^2\eta^2}$. The equation

$$\frac{d^2\psi}{dx^2} + \Phi(x)\psi = 0 \tag{1.12}$$

can be integrated, and its solution is:

$$\psi = ve^{ik_0\omega}\left(1 + \frac{\lambda_1}{k_0} + \frac{\lambda_2}{k_0^2} + \cdots\right) \tag{1.13}$$

In this equation, v, ω, λ_1, λ_2 etc. are functions of x. They can be derived by substituting the expression (1.13) into (1.12) and then equating the coefficients at the same powers of k_0. If k_0 is large, which happens occasionally, we can confine ourselves to the first term of the solution (1.13), which can be easily shown to have the form

$$\psi = \left(\frac{\Phi(x)}{k_0^2}\right)^{-\frac{1}{4}} e^{\pm i} \int^x [\Phi(x)]^{\frac{1}{2}}dx \tag{1.14}$$

This limiting form of solution for (1.12) is valid if the barrier (Fig. 1.8) contains points x^+ and x^{++} meeting the following conditions:

$$\left.\begin{array}{l} \left|\frac{\Phi'(x_1)}{\Phi''(x_1)}\right| \gg |x^+ - x_1| \gg \left|\frac{1}{\Phi'(x_1)}\right|^{\frac{1}{3}}, \\[2mm] |\Phi'(x^+)| \ll |\Phi(x^+)|^{\frac{3}{2}}, \\[2mm] \left|\frac{\Phi'(x_2)}{\Phi''(x_2)}\right| \gg |x_2 - x^{++}| \gg \left|\frac{1}{\Phi'(x_2)}\right|^{\frac{1}{3}}, \\[2mm] |\Phi'(x^{++})| \ll |\Phi(x^{++})|^{\frac{3}{2}}. \end{array}\right\} \tag{1.15}$$

The primes denote differentiation with respect to x.

The conditions (1.15) show that the potential barrier should have a "smooth" nature, i.e. the potential function should not have any infinitely large first or second derivatives. Hence, the solution (1.14) is inapplicable to an idealized triangular barrier (Fig. 1.8). It follows from (1.15) that the higher the potential curve rises above points x_1 and x_2, the better are the conditions (1.15) met. In particular, it directly follows from them that

$$\int\limits_{x_1}^{x_2} |\Phi(x)|^{\frac{1}{2}}dx \gg \frac{4}{3}. \tag{1.16}$$

This corresponds to $\mathscr{D} \gg 7\%$. The calculation presented above was proposed by G.F. Vasilev [17]. His method of solution of (1.12) is close to the famous Wentzel-Kramers-Brillouin (WKB) method [21–23] but not exactly the same. The solution (1.12) is obtained in form

$$\psi = a\Phi(x)^{-\frac{1}{4}}e^{i\int\limits^{x}[\Phi(x)]^{\frac{1}{2}}dx}, \tag{1.17}$$

where the constant factor a and the lower limit of integration are chosen arbitrarily. The solution (1.17) is valid given the following conditions:

$$\Phi(x) > 0; \quad \frac{1}{4\Phi}\left|\frac{\Phi''}{\Phi} - \frac{5}{4}\left(\frac{\Phi'}{\Phi}\right)^2\right| \ll 1. \tag{1.18}$$

It can be shown that for the barrier described by (1.8), this condition is met well. Calculating the left-hand side, we get:

$$\frac{3a_0}{8x}\frac{1 + \frac{32}{3}\frac{Ex}{e^2}}{\left(1 + 4\frac{Ex}{e^2}\right)^3} \ll 1. \tag{1.19}$$

In this equation, a_0 is represented with the Bohr radius of a hydrogen atom. It can be seen that for $x \geq 1$ Å, the condition (1.19) holds.

So, in the region $x > x_2$, where $\Phi > 0$, it can be stated that:

$$\psi(x) = a\Phi(x)^{-\frac{1}{4}}e^{i\left(\int\limits_{x_2}^{x}\Phi^{\frac{1}{2}}dx + \frac{\pi}{4}\right)}. \tag{1.20}$$

The expression (1.20) differs from (1.17) by the fact that in the former the lower limit of integration is defined and another wave phase is introduced for convenience: (1.20) represents an electronic wave leaving the metal. It is invalid in the negative kinetic energy field $x_1 < x < x_2$, since $\Phi < 0$ [cf. (1.18)].

According to Kramers [24], in this region, the individual function grows exponentially from x_2 to x_1 if it behaves as $\Phi^{-\frac{1}{4}}\cos\left(\int\Phi^{\frac{1}{2}}dx + \frac{\pi}{4}\right)$ at $x > x_2$ and decreases along the same direction if cos is replaced with sin. So, in the field $x_1 < x < x_2$

$$\psi(x) = a|\Phi|^{-\frac{1}{4}}\left[\exp\left(\int\limits_{x}^{x_2}\left|\Phi^{\frac{1}{2}}\right|dx\right) + \frac{1}{2}i\,\exp\left(-\int\limits_{x}^{x_2}\left|\Phi^{\frac{1}{2}}\right|dx\right)\right]. \tag{1.21}$$

It is well-known that at $x < 0$ inside the metal, the function ψ has the wave nature again. Therefore, the expression (1.21) can be represented as follows:

$$\psi(x) = a\Phi^{-\frac{1}{4}}e^{Q}\left\{\left(1+\frac{1}{4}e^{-2Q}\right)\exp\left[i\left(\int_{x_1}^{x}\Phi^{\frac{1}{2}}dx+\frac{\pi}{4}\right)\right]\right.$$
$$\left.+\left(1-\frac{1}{4}e^{-2Q}\right)\exp\left[-i\left(\int_{x_1}^{x}\Phi^{\frac{1}{2}}dx+\frac{\pi}{4}\right)\right]\right\}. \tag{1.22}$$

where

$$Q = \int_{x_1}^{x_2}\Phi^{\frac{1}{2}}dx. \tag{1.23}$$

As inside the metal at a sufficient distance from the surface,

$$\Phi(x) = k_0^2(\mathscr{E}+W) = \chi^2 = \text{const}, \tag{1.24}$$

then

$$\psi(x) = \frac{a}{\sqrt{\chi}}e^{Q}\left[\left(1+\frac{1}{4}e^{-2Q}\right)e^{i(\chi x+\gamma)}+\left(1-\frac{1}{4}e^{-2Q}\right)e^{-i(\chi x+\gamma)}\right], \tag{1.25}$$

i.e. inside the metal, the function ψ is the sum of the incident wave and the reflected wave of almost identical amplitude. By definition,

$$\mathscr{D} = 1 - R = 1 - \frac{\left(1-\frac{1}{4}e^{-2Q}\right)^2}{\left(1+\frac{1}{4}e^{-2Q}\right)^2} = \frac{e^{-2Q}}{\left(1+\frac{1}{4}e^{-2Q}\right)^2} \approx e^{-2Q} \tag{1.26}$$

or according to (1.21) and (1.11)

$$\mathscr{D} = \exp\left[-2\sqrt{\frac{8\pi^2 m}{h^2}}\int_{x_1}^{x_2}\left|\left[\mathscr{E}-\left(-\frac{e^2}{4x}-eEx\right)\right]\right|^{\frac{1}{2}}dx\right]. \tag{1.27}$$

It follows from (1.27) that the barrier penetrability depends on the area of the potential curve located above this energy level $\mathscr{E}$.

The integral in the exponential curve (1.27) is a complete elliptic integral calculated by Nordheim through its transformation to normal elliptic integrals.

If the expression for $\mathscr{D}$ is applicable to the case of a triangular barrier, which is in general case not necessarily true, then it is easy to get:

$$Q_0 = \frac{2\pi\sqrt{2m}}{h}\int_{0}^{x_2}\sqrt{|\mathscr{E}+eEx|}dx = \frac{2}{3}\frac{2\pi\sqrt{2m}}{h}\frac{|\mathscr{E}|^{\frac{3}{2}}}{eE}. \tag{1.28}$$

Table 1.1 Nordheim
function $\theta(y)$ and function $s(y)$

| y	$\theta(y)$	$s(y)$	y	$\theta(y)$	$s(y)$
0	1.0000	1.0000	0.55	0.6351	0.9464
0.05	0.9948	0.9995	0.6	0.5768	0.9366
0.1	0.9817	0.9981	0.65	0.5152	0.9261
0.15	0.9622	0.9958	0.7	0.4504	0.9149
0.2	0.9370	0.9926	0.75	0.3825	0.9030
0.25	0.9068	0.9885	0.8	0.3117	0.8903
0.3	0.8718	0.9835	0.85	0.2379	0.8770
0.35	0.8323	0.9777	0.9	0.1613	0.8630
0.4	0.7888	0.9711	0.95	0.0820	0.8483
0.45	0.7413	0.9637	1	0	0.8330
0.5	0.6900	0.9554			

To calculate Q, we introduce function $\theta(y)$, the so-called Nordheim function:

$$Q = Q_0 \theta(y), \qquad\qquad (1.29)$$

where

$$y = \frac{\sqrt{e^3 E}}{|\mathscr{E}|} \qquad\qquad (1.30)$$

is the ratio of reduction of the work function due to the Schottky effect (i.e. the mirror forces) to the "work function of an electron with given energy". Function $\theta(y)$ was tabulated by Nordheim (cf. Table 1.1), which makes it easy to calculate Q, and, thus, $\mathscr{D}$. Graph of $\theta(y)$ is plotted in Fig. 1.9.

Fig. 1.9 Dependence $\theta(y)$ of
y
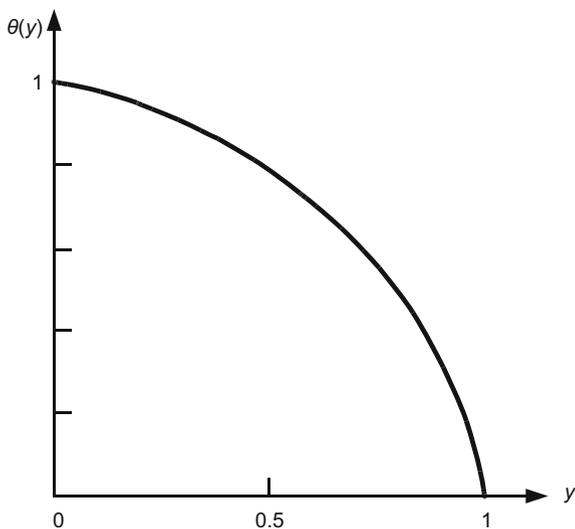
Integration over v_y and v_z results in expression

$$j = \frac{4\pi m^2 ekT}{h^3} \int\limits_0^\infty \ln\left(1 + \exp\left[-\left(\frac{mv_x^2}{2} - \mathscr{E}_0\right)/kT\right]\right)\mathscr{D}(v_x, E)v_x dv_x. \qquad (1.31)$$

Let $\mathscr{E} = \frac{mv_x^2}{2} - \mathscr{E}_0$. When $T \to 0$, the expression becomes much simpler because

$$\ln\left(1 + e^{\frac{-\mathscr{E}}{kT}}\right) = -\frac{\mathscr{E}}{kT} \quad \text{at} \quad \mathscr{E} < 0, \qquad (1.32)$$

and with $\mathscr{E} > 0$, this expression is equal to zero.

Hence, for the case of $T \approx 0$

$$j = -\frac{4\pi me}{h^3} \int\limits_{-\infty}^{0} \mathscr{D}(\mathscr{E}, E)\mathscr{E}d\mathscr{E}. \qquad (1.33)$$

To count the electrons emitted not only from the Fermi level $\mathscr{E}_0$, but also from the levels close to it, we expand the expression (1.29) into series in terms of $\mathscr{E}$. Eventually, that results in the expression

$$\mathscr{D} = \mathscr{D}_0 e^{\lambda \mathscr{E}}, \qquad (1.34)$$

where

$$\mathscr{D}_0 = \exp\left[-\frac{8\pi}{3}\frac{\sqrt{2m}}{h}\frac{\varphi^{\frac{3}{2}}}{eE}\theta\left(\frac{\sqrt{e^3 E}}{\Phi}\right)\right]. \qquad (1.35)$$

In this expression, Φ is the work function for metal and

$$\lambda = \frac{4\pi\sqrt{2m\Phi}}{heE}\left(\theta - \frac{2}{3}y\frac{d\theta}{dy}\right)_{y=\sqrt{\frac{e^3 E}{\Phi}}}. \qquad (1.36)$$

The bracketed expression is later disregarded because it is very close to one. As a result of substitution of (1.36), (1.35) and (1.34) into (1.33) and integration of (1.33), we get:

$$j \approx \frac{e^3 E^2}{8\pi h\Phi}\exp\left[-\frac{8\pi}{3}\frac{\sqrt{2m}}{h}\frac{\Phi^{\frac{3}{2}}}{eE}\theta\left(\frac{\sqrt{e^3 E}}{\Phi}\right)\right]. \qquad (1.37)$$

The expression (1.37) is the basic formula of field emission, i.e. the Fowler−Nordheim formula. If work function is measured in eV and electric field in V/cm, then

$$j = 1.55 \times 10^{-6} \frac{E^2}{\Phi} \exp\left[-\frac{6.85 \times 10^7 \Phi^{\frac{3}{2}}}{E} \theta\left(\frac{3.62 \times 10^{-4} \sqrt{E}}{\Phi} \right) \right]. \qquad (1.38)$$

Formula (1.38) shows that field emission depends on the electric field in the same way as thermionic emission depends on temperature, i.e. it is greatly sensitive to the value of the field. Just as $\lg \frac{j}{T^2} = f\left(\frac{1}{T}\right)$ is a straight line in case of thermal emission, the relation $\lg \frac{j}{E^2} = f\left(\frac{1}{E}\right)$ here is an almost straight line as well. Little deviations from the straight line are caused by presence of function $\theta(y)$ in the exponential quantity, but usually for the experimentally studied fields values of $\theta(y)$ vary insignificantly.

Field emission currents are already quite considerable at rather small $\mathscr{D}$ (approximately 10^{-10}). When fields $E = 6 \times 10^7$ V/cm and $\Phi = 4.5$ eV, $j \approx 10^7$ A/cm^2.

1.3 Classical Morgulis-Stratton Theory of Field Emission from Semiconductors

The theory of field emission from semiconductors is based on the research by Morgulis [25] and Stratton [26–28]. In this work we are not intending to describe the Morgulis-Stratton theory in great detail, instead limiting ourselves to presenting brief summary of the main results, because this theory will be shown below as having a limited sphere of applicability. Those wishing to study Morgulis' and Stratton's research papers in more detail we can point to references [5–7, 9, 17, 26–28].

Let us focus on some key differences between field emission from semiconductors and field emission from metals.

The Fermi level in a metal lies mostly by a few electron-volts higher than the lower boundary of the conduction band. That indicates high concentration of free electrons. In the semiconductor, the Fermi level lies in the bandgap. Respectively, the semiconductor conduction band displays a very low concentration of quasi-free electrons. This is the most significant moment to be taken into account when developing the theory of field emission from semiconductors.

Along with emission from the semiconductor conduction band, there may also be emission from the valence band. Its probability increases with decrease in the width of the bandgap (Figs. 1.10 and 1.11) [6].

The value of field emission current from the conduction band depends mostly on electron concentration in the surficial region of a space charge. The concentration
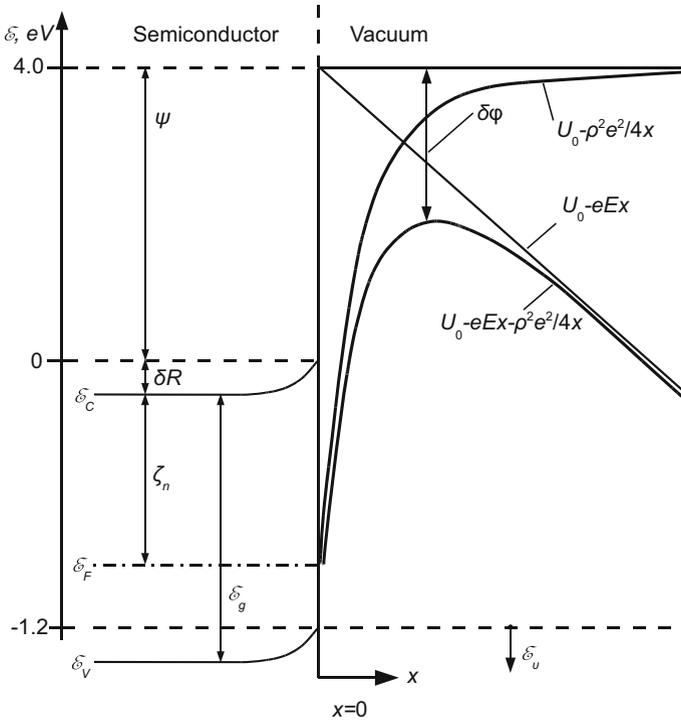
Fig. 1.10 The diagram of the potential energy of an electron near the surface of the semiconductor in the non-degenerate case fo electron gas. The value $\mathscr{E}_v = -(\mathscr{E} + \mathscr{E}_g)$—energy, measured down from the top of the valence band $\mathscr{E}_v$ on the surface. Numerical values are given for silicon

Fig. 1.11 The diagram of the potential energy of an electron near the surface of a semiconductor with a degenerate electron gas
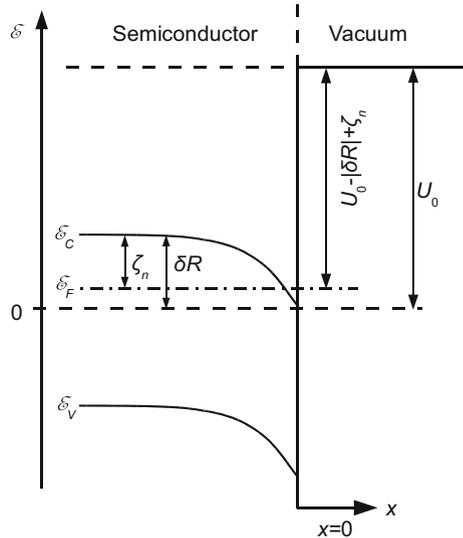
value is affected by both the electric field penetrating into the semiconductor from the outside and the surface charges. This effect is usually measured by the band δR bending value. In this case, the distribution function is given by

$$f(\mathscr{E}) = \left[1 + \exp\left(\frac{\mathscr{E} + \zeta_n + \delta R}{kT} \right) \right]^{-1}, \zeta_n > 0, \tag{1.39}$$

where ζ_n is the energy-level separation from the Fermi level to the lower boundary of the conduction band $\mathscr{E}_c$ in the bulk of the semiconductor. Both the Morgulis-Stratton theory and later theoretical papers usually assumed that the bands bend only in respect of the Fermi level, and the Fermi level remains invariable up to the semiconductor surface (Figs. 1.10 and 1.11). Band bending is positive when the bands on the surface are bent upwards, i.e. the electron concentration in the surficial region of the semiconductor space charge is less than inside the bulk.

If Fermi level lies in the conduction band (Fig. 1.11 shows the degenerate case), the Fermi distribution function is used (1.39). If Fermi level lies in the forbidden band up to the surface (Fig. 1.10, non-degenerate case), one can use a simplified expression (the Boltzmann distribution function):

$$f(\mathscr{E}) \approx \exp\left(-\frac{\mathscr{E} + \zeta_n + \delta R}{kT} \right), \zeta_n + \delta R \gg kT, \tag{1.40}$$

where energy $\mathscr{E}$ is counted from the lower conduction band boundary on the semiconductor surface. Formulas (1.39) and (1.40) regard δR as a parameter. The band bending calculation methods have been written up in [6].

In case of semiconductors with a large specific resistance and at high field emission current densities, thermodynamic equilibrium may be disturbed. In these cases, we cannot use the distribution function given by (1.39).

Another important issue in development of the theory is selection of expression for penetrability D in the equation similar to the one used to calculate dependence $j = j(E)$ for metals (1.2) and taking account of influence of the semiconductor band structure on field emission.

The Morgulis-Stratton theory has been developed only for n-type low-resistance semiconductors and for the region of relatively weak fields and currents in the way quite similar to the one for metals. It is based on the assumption of thermodynamic stationarity of the emission process and, hence, constancy of the chemical potential level in the sample. The latter is only valid if just a few electrons falling on the potential barrier tunnel in vacuum. The summed current through the sample is much less than the drift electron current to the surface and the diffusion electron current from the surface into the depth of the sample, so the respective approximation is referred to as zero current approximation. In particular, zero current approximation means that voltage drop IR of the cold emitter is zero, I being the total emission current, R being resistance of the pointed emitter. Let us mention that for metals, zero current approximation is valid in arbitrary strong fields, but for semiconductors, the region of its applicability is considerably narrower.

Fig. 1.12 Schematic band diagram for the semiconductor surface, Φ—work function, χ—electron affinity, eV_s—band bending, $\mathscr{E}_v$—the top of the valence band, $\mathscr{E}_c$—the bottom of the conduction band, $\mathscr{E}_F$—Fermi level, $\mathscr{E}_d$—donor level
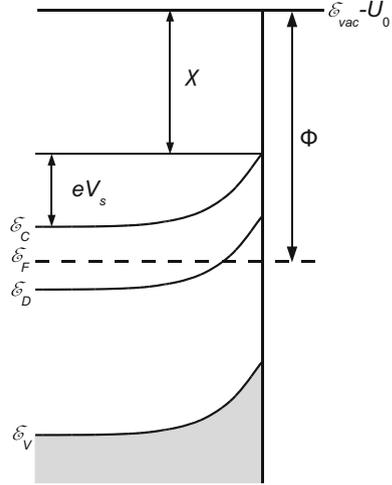
With zero current approximation, to solve the problem of the structure of the surficial region of the space charge and field distribution in the sample means to solve the Poisson equation with set boundary conditions.

However, when using the ratio (1.39) to calculate field current from the semiconductor, one has to bear in mind that the Nordheim function argument is to be further multiplied by

$$v = [(\varepsilon - 1)/(\varepsilon + 1)]^{1/2},$$

where ε is the static dielectric constant of the semiconductor.

Construction of the theory of field emission from semiconductors should involve not only the work function concept, but also the electron affinity concept (Fig. 1.12).

A semiconductor experiences an additional band bending effect and the total work function is given by

$$\Phi = \chi + eV_s + (\mathscr{E}_c - \mathscr{E}_F),$$

where χ is electron affinity, eV_s describes band bending, and the term $(\mathscr{E}_c - \mathscr{E}_F)$ denotes the difference in energies between the Fermi level and the conduction band bottom in the bulk. The dipoles caused by adsorbates affect the first two terms, while semiconductor doping affects the third term.

The integrand in (1.2) has a sharp maximum located either in the neighborhood of $\mathscr{E} = \mathscr{E}_F$ or near the conduction band bottom. In the neighborhood of this maximum, the integrand is expanded into the Taylor series and integrated term by term. Taking into account that electron energy distribution of the conduction band $N(\mathscr{E})$ can be determined in several approximations, Stratton calculation results can be summed up as follows:

1. If electron gas in the conduction band is not degenerate, penetration of the external field in the sample is neglected, then

$$N(\mathscr{E}) = n_\infty \left(\frac{1}{2\pi mkT}\right)^{1/2} \exp\left[-\frac{\chi_{af} - \mathscr{E}}{kT}\right], \tag{1.41}$$

where n_∞ is the equilibrium concentration of electrons in the bulk, χ_{af} is the affinity barrier.

For this approximation, we can get:

$$j = en_\infty \left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left[-\frac{8\pi\sqrt{2m}\chi_{af}^{3/2}}{3heE}\Theta\left\{\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^{1/2}\frac{e^{3/2}E^{1/2}}{\chi_{af}}\right\}\right]. \tag{1.42}$$

Where Θ is the Nordheim function, ε is the dielectric constant.

2. If electron gas in the conduction band is not degenerate, we should take the shallow-penetration field into account. To take the field penetration into account, we should consider not the equilibrium concentration of electrons in the bulk, but surface concentration n.

We assume that emission current does not practically disturb this distribution, which then can be given according to the Boltzmann formula:

$$n = n_\infty \exp\left[\frac{\Delta R}{kT}\right], \tag{1.43}$$

where $\Delta R = 2kT\, arsh\left[\frac{eE}{2\varepsilon kT}\left(\frac{\varepsilon kT}{8\pi he^2}\right)^{1/2}\right]$ [6]. Similarly to the first case, we derive a formula for current density, where bulk concentration is replaced with the surface one.

3. If electron gas in the conduction band is degenerate due to field penetration, $N(\mathscr{E})$ is determined in the same way as for a metal. However, here we deal with constantly incrementing surface electron concentration due to incrementing band bending. So we propose to replace the semiconductor work function with the one determined according to the formula

$$\chi' = \chi_{cp} - b(eE)^{4/5}, \tag{1.44}$$

where b is a constant dependent on the semiconductor properties. In this approximation given $\chi_{af} > b(eE)^{4/5}$ we get the following expression for field emission current density:

$$j = \frac{e}{8\pi h} \frac{(eE)^2}{\chi_{af}} \exp\left[-\frac{8\pi\sqrt{2m}\chi_{af}^{3/2}}{heE} \Theta\left(\frac{(eE)^{1/2}}{\chi_{af}} \right) \right] \times \left[1 - \left(1 + 2\frac{\kappa_0 b\chi_{af}}{(eE)} \right) \right]$$

$$\times \exp\left[-2\frac{k_0 b\chi_{af}^{1/2}}{eE^{1/5}} \right].$$

$$(1.45)$$

4. Stratton also derived the emission current formula with account of surface states. We are not going to present it in its complexity, but rather illustrate the main results of the Stratton theory, including those attained with account of emission from surface states (Fig. 1.13). Figure 1.13 prompts that, firstly, current voltage characteristics of density of field emission current lg $j = f$ (l/E) from an electron semiconductor are shown with straight lines in all approximations, except for the case when surface states have to be taken into account. Figure 1.13 shows a few calculated current-voltage characteristics for SiC (curve 1 according to (1.42), curve 2 according to (1.44), curve 3 according to the formula taking surface states into account). The bending of current-voltage characteristics arising in the presence of surface states is caused by the following. When fields are small, field emission becomes lower due to the inner decelerating barrier. In the process of cancellation of this barrier by a penetrating field, emission begins growing in an accelerated way. Along with the external field growth, the emission growth causes a simultaneous increase in both potential-barrier penetrability and electron concentration in the surface layer.

According to (1.42), the field emission current density is proportional to the electron concentration in this layer, i.e. $j \sim n$, if the electron gas of the conduction band in the surficial layer is non-degenerate.

If field penetration is disregarded, $n = n_\infty$. While if we take it into account, n depends on n_∞ according to (1.43), i.e. in a more complex way. According to (1.43), field emission current also depends on temperature [mainly, due to the dependence $n = f$(T)]. For degenerate electron gas according to (1.44), j is not an explicit function of the temperature.

To conclude this brief summary of the results derived in zero-current approximation, we should say that for the case of field emission from low-resistance n-semiconductors and in the field of weak fields and currents, the Morgulis-Stratton theory is satisfactory for description of experimental facts (linearity of current-voltage characteristics, absence of photosensitivity and weak thermosensitivity) and allows calculation of emission-current density. However, experimental research has not identified any dependences lg $j = f$(1/E), e.g. curve 3 in Fig. 1.13.

The straight current-voltage characteristic slope constructed in coordinates lg j, E^{-1} (Fig. 1.13) is determined by affinity to the electron, the bandgap width and degeneration parameter. In a considerable concentration of surface states, the current-voltage characteristics of n-type semiconductors can display characteristic bending (Fig. 1.13). Their effect is that when intensity of external field E is low, the
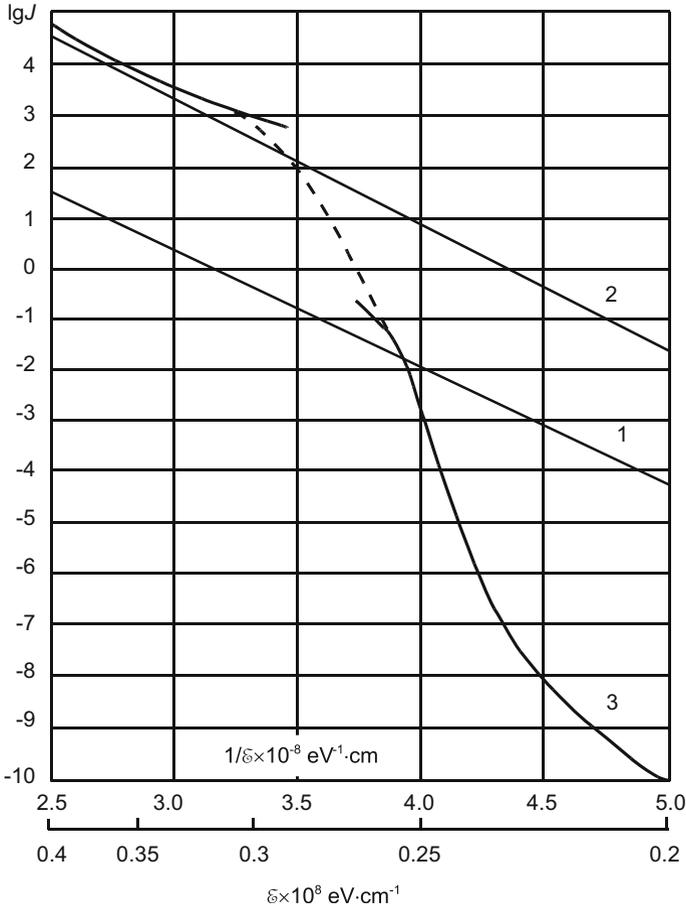
Fig. 1.13 The calculated current-voltage characteristics for SiC: 1—of formula (1.42), 2—of formula (1.45), 3—calculation that takes into account the surface states

surface-charge field bends the allowed energy bands upwards, and emission initially comes from the valence band. As soon as the field reaches the value when the surface-charge field is canceled, the bands start bending contrariwise, conductance electron concentration goes up and field emission from the conduction band becomes prevailing. On the current-voltage curve, emission moves from the valence band to the conduction band, which is accompanied by upsurge of current.

First attempts to develop a theory of field emission from n-type high-resistance semiconductors and p-type semiconductors were not made until after it had been established that current-voltage characteristics in this case are non-linear, and emission is photo- and heat-sensitive (e.g. [5–7, 29, 30]).

As for the theory of thermofield emission (TFE) from semiconductors, which was most fully developed by Vasiliev [31], it has no practical sense, as the

experiment shows (e.g. [29]) that heating of a semiconductor field emission cathode to quite low temperatures $\sim 200 \div 400$ °C results in significant distortion of its initial properties. For this reason, presentation of even some particular elements of the theory of thermal-field emission from semiconductors is of no special sense, which is why we are not going to do that.

1.4   Basics of Müller Field Emission Microscope Theory

The main element of a field emission projection microscope, an outstanding invention by Erwin Müller (e.g. [5, 8, 17, 18, 9, 32–35]) that enabled to verify key conclusions of the first theory of field emission from metals, is an electron beam source, viz. a field emission cathode (emitter). E. Müller suggested shaping it as a sharp tip. That shape of the field emission cathode enabled E. Müller and people continuing his work not just to achieve intensive electron emission, but also to examine a surface from tip emitters in an electron microscope projector with resolution $\leq 20 \div 30$ Å (E. Müller's later invention, a field ion microscope [35], allowed to increase the resolution up to $2 \div 3$ Å) and visually observe the processes of adsorption, migration, diffusion etc. within the first and the second monoatomic layers at relatively low voltages ranging from one-to-two digit kilovolts.

A wide application of the field emission microscope in studies of surfaces of metals as well as of other materials (most importantly, semiconductors) is caused by obvious advantages of the field emission method in comparison with other methods.

(1) In principle, this technique does not apply any restrictions on the working temperature and enables to study an object at the temperatures ranging from 0 K up to the melting point.
(2) A field emission microscope can directly produce an image of the sample tip surface. As mentioned above, the resolution reaches $20 \div 30$ Å. At present, the identical resolution power has been achieved in a transmission electron microscope, as well as in the scanning tunneling and atomic force microscopes. As compared with field emission methods, those devices are unable to keep track of kinetics of the processes taking place at the surface, they can only register some intermediate states of the process. Some detailed surface structure information can also be received by the low-energy electron diffraction (LEED) method. A drawback of this method as opposed to field emission technique is that it can study only one crystallographic plane at a time, and this plane has to have a long-range order structure.
(3) The field emission method enables to monitor changes of these values due to exponential dependence of emission current on the work function and the intensity of the electric field. That makes the field emission microscope especially advantageous for studies of adsorption and surface migration.

(4) A strong external electric field and high temperature change the emitter shape, and that is easy to detect both by emission pattern and by change of the emission current. By the emission pattern one can directly see the exit of impurities out of the bulk of the crystal.

(5) A field emission microscope can be used to study the phenomena of epitaxy, crystal irregularities, emergence of dislocations etc.

(6) Field emission microscopy allows to evaluate the work functions, i.e. to make quite accurate measurements of relative changes in work functions for various crystallographic planes, to evaluate surface tension factors, activation energy for migration, energies of adsorption and desorption and to determine energy-distribution half-widths with high resolution etc.

However, all those advantages of the field emission method do not come without drawbacks:

(1) Field emission results from the tunnel effect in electric fields of about 10^7 V/cm. As a rule, the emitter is shaped as a very thin tip with a curvature radius of $10 \div 100$ nm on the top to provide the electric field intensities necessary for emission. That makes it impossible to take measurements of the same object by other methods and compare the results.

(2) The phenomena taking place on the field-emitter surface can only be observed under very strong electric fields. Extrapolation of the results to the case of infinitely small field is most often incorrect.

Figure 1.14 shows a diagram of a model of the most primitive field emission microscope (a field emission diode). The anode is almost spherical as the shape of the anode is of no importance. The emitter tip (terminus) is situated in curvature center of the anode. Electrons emitted from the termination have very little initial kinetic energy and their paths are determined by spatial distribution of the applied acceleration field, this distribution being almost spherical in immediate proximity to the hemispherical surface of the tip. That is the case for a metal emitter, whose surface is an equipotential of an electric field and is directed perpendicular to its power lines. The field value decreases in a close proportion to the squared radial distance from the terminus. So the electrons acquire their terminal velocity at a distance of several peak radii from the emitting surface, going further to the anode along the radial paths [23, 36]. The anode is phosphor-coated, which turns it into a luminescent screen. Falling electrons activate the screen and it generates a magnified image of the terminus surface. The magnification factor is proportional to the ratio of the distance between the terminus and the screen to the tip radius; its practical value is usually in an order of 10^5.

In Fig. 1.15 one can a simple representation of a field electron motion trajectory as a continuous curve. Let us note that this diagram is not in a true scale. In reality, the distance from the emitter terminus to the screen equals several centimeters, which is almost 10^5 times as long as the radius of the terminus. The radial and tangential components of the initial electron velocity are designated with u_r and u_t. Since u_t is non-zero, the electron reaches the screen in a point biased by the value of

Fig. 1.14 The scheme of
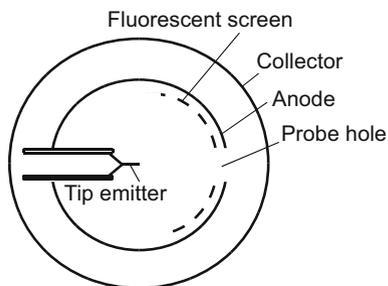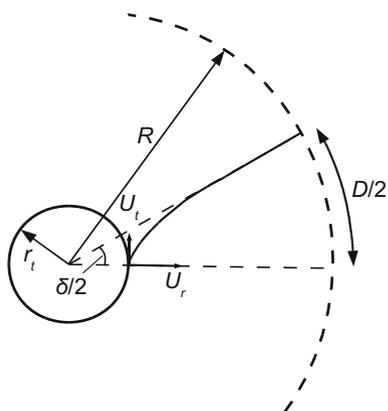field emission diode

Fig. 1.15 The trajectory of
the field emission electron

$D/2$ from the radial projection of the point of field-electron escape from the surface of the cathode. The calculations made by R. Homer [37] showed that

$$D/2 \approx 2\langle u_t \rangle t, \tag{1.46}$$

where $\langle u_t \rangle$ is a mean tangential component of the initial velocity of emitted electrons and t is given as

$$t \approx x(2eV/m)^{-1/2}, \tag{1.47}$$

where x is the distance from the terminus to the screen, V is the voltage applied between the emitting cathode and the anode. $\langle u_t \rangle$ can be estimated as

$$\langle u_t \rangle = (2\langle \mathscr{E}_t \rangle/m)^{1/2}, \tag{1.48}$$

where $\langle \mathscr{E}_t \rangle$ is a mean tangential component of field-electron energy on the emitting surface. Direct calculation of this value (in electron-volts) on the basis of the Fowler–Nordheim field emission theory gives
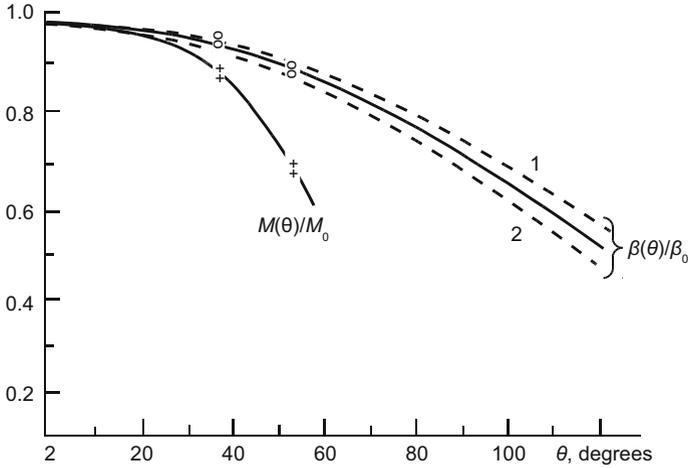
Fig. 1.16 Dependence of the local field factor β (*circles*) and the multiplication M (*crosses*) of the azimuthal angle θ at $\theta = 0$, $\beta = \beta_0$, $M = M_0$. The *dashed curves* are calculated for two models of the sphere on the orthogonal cone, simulating two different emitter profiles: with a pronounced (1) and small (2) compression. The *solid curve* corresponds to the average emitter

$$\langle \mathscr{E}_t \rangle = 0.433 E \upsilon^{-1} \left(3.79 F^{1/2} / \Phi \right) \Phi^{-1/14}, \tag{1.49}$$

where υ is the elliptic function, Φ is the effective work function, E is the field on the tip surface (in volts per ångström), which can be found from the ratio

$$E = \beta V \equiv V/k r_t, \tag{1.50}$$

$k \approx 5$ at the emitter terminus and grows with increase of the polar angle (Fig. 1.16).

Figure 1.14 shows that the observer looking at the screen perceives an arbitrary small segment on the tip surface as stretched by the value of

$$\delta/2 = D/2M, \tag{1.51}$$

where M is a field emission microscope magnification factor. If the emitter terminus was perfectly spherical, factor M would be equal to x/r_t. In practice, presence of the emitter trunk impresses the field lines towards the termination axis, thereby reducing the magnification effect. It has been established that for real emitters

$$M = x/\gamma r_t \tag{1.52}$$

where $\gamma \approx 1.5$. From (1.47) to (1.52) we can find that the resolution of the field emission microscope (in ångströms) is

$$\delta \approx 2,62\gamma \left(r_{\mathrm{t}}/kv\Phi^{1/2} \right)^{1/2}, \tag{1.53}$$

where r_{t} is the tip radius (in ångströms). It is estimated that for a typical emitter it is about $\delta \approx 25$ Å.

Along with (1.53), there is another condition that constraints the resolution of a field emission microscope. It is caused by the Heisenberg uncertainty principle, from which follows that if an electron is emitted from the region of width δ_0 it has to have the minimum tangential component of velocity of at least $u_{\mathrm{t}} \approx \hbar/2m\delta_0$. This principle imposes an irremovable limitation on resolution of any microscope. However, the estimates show that in most cases this limitation can be neglected as the resolution estimate is not much different from the estimate derived from the ratio (1.53).

As to identification of the basic crystallographic faces displayed on the luminescent screen of the field emission microscope (Fig. 1.17), it is carried out with standard stereographic (or pseudo-stereographic) projections [18]. For the basic stereographic structures (BCC, FCC, HCP), they can be found in monographs and textbooks (e.g. [38]) or manuals on crystallography. When deciphering structure images, they have to be calculated on computer. In the field emission microscope, a sample (a tip field emission cathode) is shaped and deformed in the process of surface image acquisition and the deformation nature can be of two kinds, both elastic and plastic. If the field meeting the conditions of the best mapping of the sample surface equals E_0, then the crystalline lattice of the sample is elastically extended and deformation

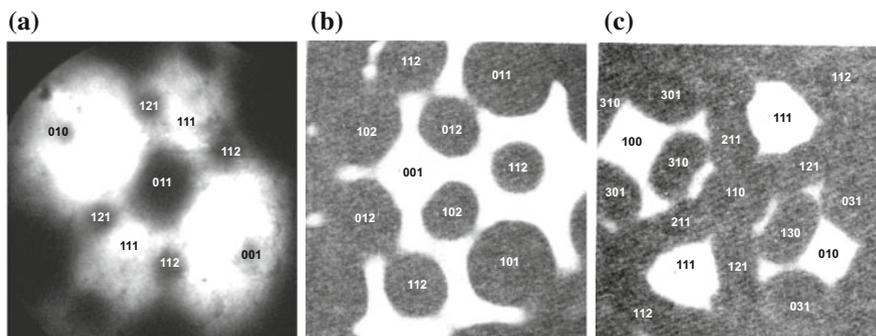$$\varepsilon_0 \approx \sigma E_0/K, \tag{1.54}$$

Fig. 1.17 Examples of typical field emission images of the tip surface of tungsten (**a**, **b**) and niobium (**c**): **a**—the surface is smoothed high-temperature heating ($T_0 \cong 0.7\ T_{\mathrm{melting}}$) without a field; **b**, **c**—surface reconstructed with a slight warming ($T_0 \cong 0.2\ T_{\mathrm{melting}}$) in the presence of an electric field

where K is the volumetric modulus of elasticity usually equal to a few percent. The respective value of the elastic energy accumulated in the sample per atom equals

$$\mathscr{E}_{at} = \Omega \sigma^2 / 2K, \tag{1.55}$$

where Ω_a is atomic volume. Since $\mathscr{E}_{at}$ is usually of the order of 0.1 eV, i.e. comparable with activation energy of low-temperature diffusion processes, impression of an electric field in the field ion microscope can considerably change the nature of pin-hole defect annealing.

Approximating the tip shape with a lathe object having symmetry axis z and neglecting particular atomic and electronic surface structures acting on the plane $z = z_0$, we can express force F_z as

$$F_z = \int \int [F^2 / (8\pi \cos \alpha)] dS_e, \tag{1.56}$$

where dS_e is an element of the tip surface, α is an angle between the normal line to this element and axis z. Integration is performed over the entire surface, i.e. from the tip terminus to the chosen section plane. As a result, (1.56) can be represented as follows:

$$F_z = \int\limits_0^{r_0} \frac{Fdr}{8\pi dl} 2\pi dl = \int\limits_0^{F_0} \frac{F^2}{8\pi} rdr, \tag{1.57}$$

where dl is the linear element of the surface, r_0 is the distance from the tip centre to z_0.

Another "concurrent" process is bombardment of the sample surface with low-energy (several dozen kiloelectronvolts) ions of residual gases. In absence of an electric field, the tip surface experiences ξ_0 gas atom impacts per time unit caused by their mere thermal motion:

$$\xi_0 = S_0 P_g \left(2\pi M k T_g \right)^{-1/2}, \tag{1.58}$$

where S_0 is the tip surface area; M is the gas particle mass; P_g is the gas pressure and T_g is its temperature. In the electric field of intensity E, the gas-molecule flow from the tip surface increases $n*$ times, which is connected with polarization of the molecules. $n*$ can be approximated from the equation

$$n* \approx \pi \alpha_g E^2 / \left(2kT_g \right), \tag{1.59}$$

where α_g is gas-molecule polarisability.

As is known [18], the number of impacts ξ_u experienced with the entire tip working surface per time unit can be estimated by the formula

$$\xi_u = \sum_i \xi_{u,i} = \sum_i \left[1,44 \frac{P_i I_{\text{FE}}}{k T_0 e} \bar{R}_0 \sigma_i(\Psi) \right], \tag{1.60}$$

where P_i is the partial pressure of the residual gas components; k is the Boltzmann constant; T_0 is the tip temperature; e is the electron charge; $\bar{R}_0$ is the mean radius of the tip surface curvature; $\sigma_i(\Psi)$ is the section of ionization equal to

$$\sigma_i(\Psi) = \frac{N_e \pi a_0^2}{V_{\text{ion}}^2} 2.66 \frac{\Psi - 1}{\Psi^2} \ln(1.25\Psi); \tag{1.61}$$

where N_e is the number of electrons in the outer shell of the ionizing atom of the residual gas, a_0 is the first Bohr orbit radius, V_{ion} is the atom ionization potential, and $\Psi = \mathscr{E}_e / V_{\text{ion}}$, where $\mathscr{E}_e$ is the electron energy equal to the value of the potential in the atom ionization point.

The bombardment of the field-emitter surface with low-energy ions of residual gases can have both negative and positive value. In the practice of operation of cathode devices on the basis of field emission (in the same way as in the case of research with field emission microscope), it can result in a change of the surface structure, pulverization or redistribution of the component (film) activating electron emission, vacuum breakdown and a total outage of the emitter (sample). At the same time, the very possibility to irradiate the sample in the field emission microscope looks undoubtedly fruitful and gives a wide range of opportunities.

Speaking about the bombardment of tip surfaces with low-energy gas ions functioning in the field emission mode, it has to be noted that such a bombardment is among the main causes of low-frequency emission noises. As such, let us also dwell on the issue of the noises a bit.

The current emitted by the sample in the field emission microscope is somehow labile: in addition to the usual "*schrot*" noise component natural for any kind of particle emission, the field emission current has another relatively low-frequency noise component. The change of the spectral power density $W(f)$ of the latter component is close to dependence $1/f\gamma$, where γ ranges between 0.8 and 1.2. That is the so-called flicker noise of field emission.

As is known, when studying fluctuation of the current, the main values to be analyzed were the correlation function (auto-correlation) $f^*(\tau)$ and the noise-power spectral density function $W(f)$:

$$\begin{aligned} f^*(\tau) &= \langle I(t)I(t+\tau) \rangle, \\ W(f) &= \langle dI^2 \rangle / df. \end{aligned} \tag{1.62}$$

These values are brought together in the Wiener–Khinchin formula

$$W(f) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} f * (\tau) \exp(-2\pi I f \tau) d\tau. \tag{1.63}$$

In addition, the ratio between the mean square of the current intensity fluctuation $(\overline{\Delta I})^2$, the correlation function and the noise-power spectral density function is given by the expression

$$\langle \Delta I^2 \rangle = f * (0) = \int\limits_{0}^{\infty} W(f) df. \tag{1.64}$$

It is widely agreed that the main cause of field emission flicker noise is the presence of adsorbed particles on the field-emitter surface and their irregular surface movement [39].

Another "*concurrent*" process that has already been mentioned is the thermal-field change of the tip surface shape. The joint action of heat and the electric field (the latter manifesting itself through surface atom polarization and movement of the resulting dipoles in the locally-enhanced field region, whereas heating of tips enables, facilitates and boosts such mobility) can significantly change the tip surface shape. This change has to be controlled and taken into account when studying these or those phenomena in the field emission microscope.

Real surface geometry. Various authors successfully resorted to field emission microscopy to analyze both the equilibrium shape of the metal microcrystal surface at different temperatures in and out of an electric field, and to study the proper atom condensation processes on an approximately hemispherical substrate. Below we'll give a consistent consideration of the most important results attained in these aspects.

To begin with, the field emission microscopic image itself easily enables to establish the mean curvature radius $\bar{R}_0$ of the tip terminus. In case of field emission microscopy, $\bar{R}_0$ is determined by measuring of dependence of field emission current I_{FE} on tip potential V_0. Having expressed the Fowler−Nordheim equation in terms of I_{FE} and V_0 as

$$\frac{I_{FE}}{V_0^2} = \frac{a}{\bar{R}_0^2} \exp\left[-\frac{b\bar{R}_0}{V_0} v \left(\frac{cV_0^{1/2}}{\bar{R}_0^{1/2}} \right) \right], \tag{1.65}$$

where parameters a, b and c are constant for this material and connected with the value of the mean work function $\bar{\Phi}$, v is a tabulated function, we can see that dependence $\ln(I_{FE}/V_0^2)$ of $1/V_0$ is close to the straight line up to $\ln a/(\bar{R}_0)^2$ and its slope equals $-b\bar{R}_0 v$. At $v \approx 1$, value $\bar{R}_0$ can be approximately determined from the graph $\ln(I_{FE}/V_0^2) = f(1/V_0)$ by the line slope. The determined value R_0 can be later

used to calculate the argument of function υ and then value υ can be determined from the table. Having inserted the new value υ in the ratio $-b\bar{R}_0\upsilon$ equal to the straight line slope, we can determine a more precise value $\bar{R}_0$ etc. Precision of this method of finding $\bar{R}_0$ is $\pm 15\%$. It cannot be improved in view of essential variations of the work function and, consequently, I_{FE} along the surface of the tip terminus.

As shown in [40], impact of temperature leads to spheroidisation in tip apex.

Let us note that the mass transfer leading to evolution of the tip terminus profile is most probably caused [18] by surface self-diffusion. This seems feasible given the small size of tip apex.

At the same time, we can see an absolutely different pattern of evolution of tip terminations when they are heated in the presence of an external electric field [18]. In this case, apart from the capillary forces we have to take into account the ponderomotive ones as well. In this case, the chemical potential of atoms at the needle tip can be given by:

$$\mu = \mu_0 + \omega(2\gamma/R - E^2/8p), \tag{1.66}$$

The minimum value of voltage V_0 (tip potential) at which the electric field can cause a considerable change in the tip shape against the background of capillary forces can be estimated with the following expression:

$$V_0* \approx (16\pi\gamma R/\chi)^{1/1.4}. \tag{1.67}$$

Dependence $V_0* = f(R)$ means that in the course of needle elongation under the influence of the field, the termination curvature radius decreases and, consequently, the difference between the set value of potential V_0 and its critical value increases, which means that elongation of the tip has to be a self-accelerating process.

Analysis of thermal-field and condensation growth patterns of metal crystal tips by field emission microscopy is presented in more detail in [39–41].

1.5   Disadvantages of Classical Field Emission Theories and Limits to Their Applicability

Simplicity of presuppositions of the Fowler–Nordheim theory and at the same time quite satisfactory description of the real phenomenon by this theory caused both serious efforts to analyze its appropriateness [5–9, 19, 20, , 37, 42, 43] and numerous attempts to extrapolate it to less idealized models. Essentially, all researchers involved considered the limits to applicability of the Fowler–Nordheim theory and possibilities to work around them with some improvements to the theory. Most papers improving the Fowler–Nordheim theory focus only on the extremely high current-density region. This region has some specific physical

distinctions, and, at the same time, it is extremely interesting from the standpoint of some important applications of field emission. Because of this, we are going to analyze it in greater detail.

In addition, the classical Fowler–Nordheim theory considers only the case of "pure" field emission (which, strictly speaking, only holds for $T = 0$ K). It does not analyses theoretically interesting and practically important cases of thermofield emission for $T > 0$ K nor for $T \gg 0$ K.

Noteworthy is the following often neglected circumstance. The basic formula of the Fowler–Nordheim theory does not link the voltage and current, but rather the electric intensity and current density, i.e. basically unmeasurable values, especially if the tip dimensions are microscopic. The most common workaround for this probles is to assume that the field intensity value is proportional to voltage and the current value is proportional to its density, i.e.

$$E = \beta V; \quad I = Sj, \qquad (1.68)$$

where β is the so-called field factor and S is the emitting area. The first relation does not usually cause doubt if values β and E are understood as their local values in an emitting surface point. As for the second equation, it is more precisely written down as

$$I = \int_{(S)} j(x, y, z)dS, \qquad (1.69)$$

where S is the total surface of the emitting tip.

This issue is of both academic and quite practical nature because these values determine functioning of major vacuum-microelectronic devices, e.g. a flat display [13, 43]. In particular, [44] proposes to determine the line inclination tangent and the segment cut by the line on the ordinate axis with a line by Fowler–Nordheim approximation in the coordinates $1/V$, $\ln(I/V^2)$ and then calculate the emission area with a margin error of at least 15% and the field factor with even lower precision by the ratios

$$S = 4 \times 10^{-17}\left(I/V^2\right)_{1/V=0} \times p^2, \qquad (1.70)$$

where $p = -2.82 \times 10^9 (\Phi^{3/2}/\beta)$ is a Fowler–Nordheim line inclination tangent. The current is expressed in amperes, the length in meters and the work function in electron-volts. The first ratio is based on approximate constancy of value $\Phi^2 \exp(9.81\Phi^{-1/2})$. The second one needs knowledge of value Φ itself, which results in decrease in precision as the local work function value is measured with great difficulty and low precision.

Nevertheless, most theories assume strict observance of these equations and scrutinize only the Fowler−Nordheim equation. The equation is criticized in three aspects. Firstly, the physical prerequisites of its deduction seem to cause some

doubts that are not entirely unbased [45, 46]. Analysis of this type of criticism is beyond the purpose of our study, though some issues concerning the physical prerequisites of tunneling calculation with the WKB method will be considered in further chapters. Secondly, correctness of substitution of the elliptic integrals as components of the full formulas with their approximate values is open questionable and improved expressions are being proposed [47, 48]. Thirdly, doubt is caused by the formula parameter determination methods based on construction of current-voltage characteristics in coordinates $\{\ln(I/V^2), I/V\}$ with approximation by the least square method [44].

The critics' arguments are actually quite feasible. Indeed, application of the least square method is connected with a particular (Gaussian) kind of random error distribution [49]. It is known that any transformation of variables violates the initial formula for the distribution function [50]. However, it still remains to be seen how significant this violation is. This question is even more important if we note that there are a lot of approximations initially built into the model, which may cause much more serious errors. It is also important that according to the Neumann-David theorem, the estimate of the linear-regression coefficient by the least-square method is asymptotically unbiased and effective irrespective of a particular formula of the error-distribution function [51]. Certainly, modification of the formula of the error-distribution function causes change in the estimates of confidence intervals. However, in this case, we should speak not about the biased estimate of the physical model parameters, but rather about determination of the confidence interval for this estimate. We cannot but note that the latter is very seldom, if ever, done.

The papers we know of have not payed enough attention to influence of an error in the independent variable. The classical least square method assumes that a random error burdens the measurements only in case of a dependent variable, while the independent one is considered to be measured in an absolutely precise way. Certainly, that is not the case. In practice, voltage measurements also bear a random error comparable with a current-measurement error (of course, in relative terms). Thus the usual formula of the least square method is mathematically unreasonable here. What we need is a more complex model that's still based on minimization of residual sum of squares but taking into account the errors of both variables. So the first stage of work should be analysis of applicability of Fowler−Nordheim formulas or their improved analogues, comparative estimate of efficiency in terms of computer time consumption and, finally, definition of the physical sense of the values obtained. The last stage is also important if the results attained are not the end in itself or a reference point for relative comparisons but the initial data for further mathematical or natural simulation or even practical use.

For such analysis, we developed an information expert system [20] and carried out a natural experiment on measuring of dependence $I(V)$ in a wide range of current voltages and densities, the latter exceeding five orders of magnitude. The results attained were processed in various ways:

1. Construction of a line in coordinates $\{\ln(I/V^2), 1/V\}$ by the usual least-square method [52].
2. Construction of a line in coordinates $\{\ln(I/V^2), 1/V\}$ by the least-square method with account of an error in the independent variable ($1/V$) [14].
3. Minimization of value χ^2 in direct approximation of the functional dependence (this was carried out in *Origin* software package) of non-linear regression [16]. In this case, the trial function was represented by both the classical Fowler–Nordheim expression and more complex expressions.
4. Application of nonparametric-regression methods, which do not require from the error distribution function to have any particular formula [53], except for its symmetry about zero.

The results are shown in the figures below. Figure 1.18 shows dependence of a field emission current on voltage in natural coordinates as a semilogarithmic graph and a non-linear-regression curve optimal in terms of χ^2 *minimization*. Table 1.2 shows Fowler–Nordheim line parameters obtained by various methods and meeting the confidence intervals for 95% significance level. For a direct non-linear regression, we obtained estimates of coefficients P_1 and P_2 of the formula $I = P_1 V^2 \exp(-P_2/V)$. The estimates were optimal in terms of χ^2 minimization and we re-calculated them into the respective parameters of the Fowler–Nordheim line for convenience of comparison. The table proves that it is only non-linear regression that gives significantly different parameters of the line in terms of statistics. However, this result is the least satisfactory from the physical point of view.

The matter is that when value χ^2 is minimized, small values of the dependent variable cannot compete with large values. Practically, it is obvious that the regression curve does not take them into account and the differences reach half of
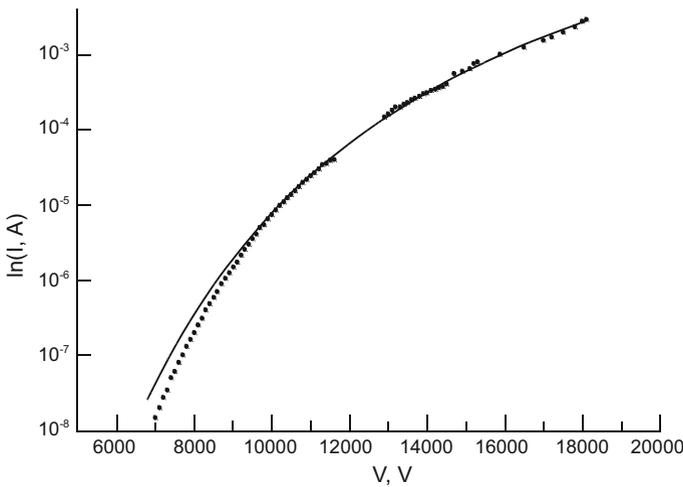

Fig. 1.18 The dependence of the field emission current from the voltage: the experimental points and the curve of non-linear regression

Table 1.2 Regression-curve parameters $\ln(I/V^2) = A \times (I/V) + B$ obtained for the available experimental points by various methods and their confidence intervals with 95% confidence level

Type of regression	A	DA	B	DB
Classical	-1.158×10^5	0.007×10^5	-18.85	0.07
With error V	-1.161×10^5	0.007×10^5	-18.81	0.08
Non-linear	-1.052×10^5	0.02×10^5	-19.67	0.09
Non-parametric	$-1.143 \times 10'$	0.03×10^5	-19.0	0.1

the order of magnitude. It is natural when a dependent variable varies in a large dynamic range. That is why the non-linear regression is unsatisfactory in this application. All other methods are indistinguishable. So it is natural that the one to be preferred should be the simplest and traditional, for which there are standard program packages available. However, comparison with the nonparametric-regression results shows that precision of their estimates is usually overrated. From the procedure of determination of coefficients S and β, especially popular with experimenters, it is obvious that all the methods applied result in the values just slightly different from each other. The maximum difference of the value of emission area S is no more than 20% (calculation by traditional linear methods gives 5.2×10^{-16} m^2 and non-linear regression results in 6×10^{-16} m^2), i.e. the values of errors all lye within the margin of error of the model. By the basic rules of measurement theory, the prospective model accuracy should considerably exceed the accuracy of determination of its parameters [54].

That means that the experiment does not make a statistically significant distinction between different mathematical models of the process. Thus, feom this point onwards we are going to use only the processed Fowler−Nordheim equation in its usual formulation, i.e. a traditional linear regression in coordinates $\{\ln(I/V^2), 1/V\}$. However, the problem of unreasonable extrapolation of local equation to the integral characteristics persists and so does the physical essence of the parameters determined thereby.

Let us turn to discussion of the semiconductor field emission theory described in previous sections. As a preliminary and in addition to what has already been said in 1.3, let us give a brief summary of the basic assumptions underlying this theory.

(1) For the transmission factor $\mathscr{D}$, we used the expression known from the metal field emission theory with account of a more general form of the image-force potential. In this case, we disregarded the possible contribution of relaxation effects to this potential.

(2) The semiconductor electrons were considered to be quasi-free. Further on, we assumed the same dependence $\mathscr{E}(\mathbf{k})$ inside the crystal and on the surface region, which is closely connected with the issue of preservation of crystalline symmetry up to the very surface. However, that assumption about function $\mathscr{E}(\mathbf{k})$ was not verified, e.g., for germanium and silicon [6, 7].

(3) Everywhere, we proceeded from the assumption that electrons are in thermo-dynamic equilibrium with the crystalline lattice.

(4) The influence of the surface charges and field penetration on concentration of electrons in the surface region was described in terms of band bending. Possible quantization of electron states in the bulk-charge band near the surface was not taken into account.

(5) We discussed diversions from thermodynamic equilibrium. The results furnished were attained in the assumption that it is only acoustic-phonon scattering that matters. We considered field emission from *n*-type semiconductors without account of surface states. However, the experiments found out hot-electron emission from *p*-type semiconductors as well [6, 7].

(6) The influence of effective mass was studied in the assumption that emission preserves the tangential component of the wave vector. Essentially, the latter matters when band boundaries in the space of wave vectors **k** lie at **k** $\neq$0, e.g. in the case of germanium and silicon. Preservation of the transverse wave-vector component was admitted both in the semiconductor photoemission theory [9] and in the general consideration of penetrability. Silicon photoemission experiments [9] justify such an assumption for at least a large number of emitted electrons.

For the sake of theory we'd prefer to preserve the transverse wave-vector component unless the translation symmetry of the crystal is broken on the surface. That is observed in the case of a flat boundary, when external forces are perpendicular to the surface [5–7].

Silicon field emission experiments [6] indicate possible non-preservation of the transverse wave-vector component. We gave the following prospective reasons thereof:

- diffused scattering of electrons on the surface due to its roughness;
- violation of periodicity of the crystal potential on the surface;
- interaction with surface phonons.

Stratton explored this issue and found out that only simultaneous presence of first and second factors can result in non-preservation of the transverse wave-vector projection.

It is clear from what has been said that the classical Morgulis-Stratton semiconductor field emission theory, however improved and enlarged in a number of later works it might be, is still connected with a lot of essential simplifications. Therefore, it is no wonder that the theory is still unable to explain a lot of experimental results in terms of quantity. In a lot of cases, qualitative explanation is constrained as well.

References

1. N. Mott, J. Sneddon, Wave Mechanics and Its Applications. M.: Nauka, (1966), 427 p
2. A. Messia, Quantum Mechanics (in 2 volumes). M.: Nauka (1979)
3. R.H. Fowler, L. Nordheim, Electron emission in intense electric field. Proc. Roy. Soc. A. **119** (781), 173–181 (1928)

4. L. Nordheim, Die Theorie der Elektronemission der Metalle. Physikalische Leitschrift. (1929). Bd. 30. N7. S.117–196
5. A. Modinos, Field, thermo and secondary electron spectroscopy. M.: Nauka, (1990), 320 p
6. R. Fischer, H. Neumann, Field emission of Semiconductors. M.: Nauka, (1971), 216 p
7. M.I. Elinson (ed.), The cold cathodes. M.: Sov. Radio, (1974), 336 p
8. E.P. Sheshin, Surface structure and field emission properties of carbon materials. M.: Publishing House of the MIPT. Fizmatkniga (2001), 287 p
9. L.N. Dobretsov, M.V. Gomoyunova, Emission Electronics. M.: Nauka (1964), 364 p
10. M. Green (ed.), The Surface Properties of Solids. M.: Mir, (1972), 432 p
11. A.I. Anselm, Introduction to the Theory of Semiconductors. M.: Nauka (1978), 616 p
12. M. Szilagyi, Electron and Ion Optics. M.: Mir, (1990), 639 p
13. L.D. Landau, E.M. Livshits, Mechanics. M.: "Nauka" (1965), 204 p
14. L.D. Landau, E.M. Livshits, Statistical Physics. Part 1. M.: "Nauka" (1966)
15. Landau L.D., Lifshitz E.M. Quantum Mechanics. Non-relativistic theory. M.: Nauka (1974), 752 p
16. N.D. Lang, W. Kohn. Theory of metal surfaces. Phys. Rev. B. **1**(12), 4555–4568 (1970)
17. M.I. Elinson, G.F. Vasiliev, Field emission. M.: Fizmatgiz (1958), 272 p
18. A.L. Suvorov, The structure and properties of the Surface Atomic Layers of Metal. M. : Energoizdat (1989), 296 p
19. D.A. Ovsyannikov, N.V. Egorov, Mathematical modeling of systems for formation of electron and ion beams. SPb.: Publishing of St. Petersburg State University (1998), 276 p
20. N.V. Egorov, A.G. Karpov, Diagnostic Information and Expert Systems. SPb., St. Petersburg State University Publishing House (2002), 472 p
21. G. Wentzel, Zeits. f. Phys. (1926), 38, 516
22. H.A. Kramers, Zeits. f. Phys. (1926), 39, 828
23. L. Brillouhin, Comptes rendus. (1926), 183, 549
24. H.A. Kramers, Zeit. A. Phys. (1926). Vol. 38. p. 516
25. N.V. Morgulis, To a question about the effect Schottky for compound semiconductor cathodes. ZETP, **16**(11), 959–963 (1946)
26. R. Stratton, Field emission from semiconductor. Proc. Phys. Soc. B **68**(430B), 746–757 (1955)
27. R. Stratton, Theory of field emission from semiconductors. Phys. Rev. **125**(1), 67–82 (1962)
28. R. Stratton, Phys. Rev. **135** (1964)
29. N.V. Egorov, V.R. Tolstyakov, Investigation of the effect of the surface state on the emission characteristics of semiconductor photo field cathodes. Surface (1996). (9), 10–13
30. N.V. Egorov, V.R. Tolstyakov, The effect of multi-particle tunneling in the field electron emission from semiconductors. Surface (1996). (9), 10–13
31. G.F. Vasiliev, Radiotehnika i Elektronika, **3**(7), 962 (1958)
32. E. Mueller, T Field. Tson. Ion microscopy. M.: Metallurgia (1971), 360 p
33. E.V. Mueller, M. Sauton, D. Brandon et al. Field ion microscopy, ed. by J. Ren, S. Ranganathan. M .: Mir (1971), 270 p
34. E.W. Mueller, Phys. Z. (1936). **37**, 838
35. E.W. Mueller, W. Z. Electrochem, (1957). **61**, 43
36. S.A. Fridrihov, S.M. Movnin, Physical Fundamentals of Electronics. M.: Vyshaya Shkola (1982), 608 p
37. R. Gomer, *Field Emission and Field Ionization*. (Harvard University Press, Cambridge, MA, 1961)
38. R. Knox, A. Gold, Symmetry in Solids. M.: Nauka, (1970), 424 p
39. A. Brenac, R. Baptist, G. Chauvet, R. Meyer, Caracteristiques energetiques de cathodes a micropointes a emission de champ. Revue Phys. Appl. **22**, 1819–1834 (1987)
40. C.A. Spindt, I. Brodie, L. Humphrey, E.R. Westerberg, Physical properties of thin film field emission molibdenium cones. J. Appl. Phys. **47**, 5248 (1976)
41. K.L. Jensen, E.G. Zaidman, Field emission from an elliptical boss: Exact versus approximate treatment. Appl. Phys. Lett. **63** (1993)

42. M. Lampert, P. Mark, Injection Currents in Solids. M.: Mir (1973), 416 p
43. R.G. Forbes, Refining the application of Fowler–Nordheim theory. Ultramicroscopy. **79**, 11–23 (1999)
44. E. Hantzsche, Beitrage zur Plazma-physik, **22**, 325 (1982)
45. G.D. Yakovleva, Tables of Airy functions and their derivatives. M.: Nauka (1969)
46. Y.V. Kryuchenko, V.G. Litovchenko, Computer simulation of the field emission from multilayer cathodes. JVST. **B14**, 1934–1937 (1996)
47. M. Hollander, D. Wolfe, Non-parametric statistical methods. M.: Finansy i Statistika (1983), 518 p
48. S.G. Rabinovich, Measurement Errors. M.: Energia (1978), 261 p
49. Y. Linnik. The method of least squares and the foundations of mathematical and statistical evaluation of parameters. M.: Fizmatgiz (1962), 350 p
50. A. Brunetti. A fast and precise genetic algorithm for a non-linear fitting problem. Comp. Phys. Comm. **124**, 204 (2000)
51. Enrico Fermi. Quantum mechanics. M.: Mir (1965). 242 p
52. Bohr. Atomic Physics and Human Knowledge. M.: GIIL (1961)
53. M. Kendall, A. Stewart, The Theory of Distributions. M.: Nauka (1966), 520 p
54. J.N. David, Probability Theory of Statistical Methods (Cambridge, 1951)