

SOLID STATE PHYSICS

Advances in
Research and Applications

Editors

Frederick Seitz
Rockefeller University, New York, New York

David Turnbull
*Division of Engineering and Applied Sciences, Harvard University
Cambridge, Massachusetts*

Henry Ehrenreich
*Division of Engineering and Applied Physics, Harvard University
Cambridge, Massachusetts*

The following monographs are published within the framework of the series:

1. T. P. DAS and E. L. HAHN, *Nuclear Quadrupole Resonance Spectroscopy*, 1958
2. WILLIAM LOW, *Paramagnetic Resonance in Solids*, 1960
3. A. A. MARADUDIN, E. W. MONTROLL, and G. H. WEISS, *Theory of Lattice Dynamics in the Harmonic Approximation*, 1963
4. ALBERT C. BEER, *Galvanomagnetic Effects in Semiconductors*, 1963
5. ROBERT S. KNOX, *Theory of Excitons*, 1963
6. S. AMELINCKX, *The Direct Observation of Dislocations*, 1964
7. JAMES W. CORBETT, *Electron Radiation Damage in Semiconductors and Metals*, 1966
8. JORDAN J. MARKHAM, *F-Centers in Alkali Halides*, 1966
9. ESTHER M. CONWELL, *High Field Transport in Semiconductors*, 1967
10. C. B. DUKE, *Tunneling in Solids*, 1969
11. M. CARDONA, *Modulation Spectroscopy*, 1969

TUNNELING IN SOLIDS

C. B. Duke

GENERAL ELECTRIC RESEARCH AND
DEVELOPMENT CENTER
SCHENECTADY, NEW YORK

now at

DEPARTMENT OF PHYSICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

1969



ACADEMIC PRESS New York and London

QC
174.1
D76
PHYSICS-MATH

Foreword

This book continues the supplementary treatises of the Solid State Physics serial publications. In it Dr. Duke reviews, comprehensively and critically, a field which has become in the past fifteen years one of the most important in solid state research. This field also is one which has some important ramifications in technology. The scope and viewpoint are clearly set forth by Dr. Duke in his introductory material.

June, 1969

FREDERICK SEITZ

DAVID TURNBULL

HENRY EHRENREICH

COPYRIGHT © 1969, BY ACADEMIC PRESS, INC.
ALL RIGHTS RESERVED
NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM,
BY PHOTOSTAT, MICROFILM, RETRIEVAL SYSTEM, OR ANY
OTHER MEANS, WITHOUT WRITTEN PERMISSION FROM
THE PUBLISHERS.

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
Berkeley Square House, London W.1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 55-12299

PRINTED IN THE UNITED STATES OF AMERICA

I. Introduction

The definition of the concept of tunneling and observation of its experimental manifestations can be regarded as one of the early triumphs of the Schrödinger wave mechanics. The concept arises from the quantum-theory prediction that an initial state, prepared in one classically allowed region of configuration space, has a nonzero probability of penetrating through a classically forbidden region into a second classically allowed region. The first application of this prediction was made in 1928 by Oppenheimer¹ to describe the autoionization of excited states of atomic hydrogen in a strong electric field. Subsequent applications have led to substantial consequences in the fields of nuclear physics, chemical physics, and solid state physics as well as atomic spectroscopy. A history of these consequences may be divided naturally into a recounting of events in two eras. During the prewar era of 1928–1940, the theoretical applications of the independent-electron model of tunneling were developed almost to their modern form. Unambiguous experimental confirmations of the qualitative validity of this model (and hence of wave mechanics) quickly resulted from analyses of the alpha decay of heavy nuclei (1928) and of electric-field broadened spectral emission from atomic hydrogen (1930). The last major advance in the prewar period was the invention of the field-emission microscope (1937) and the subsequent confirmation of the tunneling model (1928) of field emission from metals.

In the prewar era, predictions also were made of tunneling through thin insulating films between metals and semiconductors, and of dielectric breakdown in solids. However, the observation of these effects awaited the development of materials technology during the 1940's and early fifties. The postwar period was initiated in the late fifties by the invention of the p-n tunnel diode (1957) and the discovery of tunneling into superconductors (1960). The latter discovery led to the prediction (1962) and observation (1963) of weak coupling between superconductors. Many of the phenomena characterized as being "tunneling in solids" have their theoretical heritage in the prewar era. Nevertheless,

which are identical to the original one-electron Hamiltonian characterized by the potential shown in Fig. 1.

The simplest, and most easily utilized, interpretation of the single-particle tunneling process is the stationary-state theory of weakly quantized (or quasi-stationary) "bound-states in the continuum."² A semiclassical version of this theory first was introduced to describe alpha decay.^{5,6} In the one-dimensional potential shown in Fig. 1, the eigenvalue spectrum is continuous and nondegenerate for all values of E . However, for $E < -E_m$ the eigenfunctions are linear combinations of a growing and a decaying solution for $x_1(E) \leq x \leq x_2(E)$. Those states for which the eigenfunction is purely decaying with increasing x in the forbidden region are called "weakly quantized" or quasi-stationary states. They occur at quantized energies $E = -E_n$ and are the analogs of the *bona fide* bound states in zero field (which they become in the limit that the field goes to zero⁷). The probability that electrons in these states tunnel from the interior to the exterior region is commonly taken to be

$$P = |\psi_{-E_i}[x_2(-E_i)]/\psi_{-E_i}[x_1(-E_i)]|^2.$$

It often is calculated using the WKB method².

Another common measure of the change in the character of the eigenvalue spectrum as a function of energy is the spectral density,

$$S_e(E) \equiv |\psi_E(0)|^2,$$

$$S_o(E) \equiv |d\psi_E(0)/dx|^2,$$

defined for even (e) and odd (o) states in the absence of a field. These spectral densities show sharp peaks in the neighborhood of $E = -E_i$ for wave functions normalized to constant current in the exterior region.^{2,7} The stationary-state point of view has the decided advantage that it is applicable even for strong fields in which the tunnel probability cannot be treated as a perturbation.⁷⁻¹⁰ The correspondence between the transfer-Hamiltonian (i.e., Oppenheimer's time-dependent method) and stationary-state formulations of a one-electron theory of tunneling has been studied by Lanczos¹¹ for the case of the free-field ionization of atomic hydrogen, and by Rice¹² for the application to the quantum theory of chemical reactions.

The inequivalence of the transfer-Hamiltonian and stationary-state formulations for calculating the tunneling probability has caused controversy in papers on a variety of physical phenomena. Another early application of the idea of tunneling was to the description of chemical reactions, especially those, like molecular disassociation, which could

be viewed as transitions from a bound state into the continuum. The stationary-state model of chemical reactions was developed by Langer.¹³ The transfer-Hamiltonian model was originated by Wentzel¹⁴ and refined by Rice.¹⁵ A tension between the two formulations also arose in calculations of the electric dipole layer of atoms adsorbed on metal substrates. The stationary-state model was developed by Gurney¹⁶ in the early years of quantum mechanics. In the postwar period, the transfer-Hamiltonian method has been applied to this problem by several authors whose work is reviewed by Gadzuk.¹⁷ Entirely postwar versions of this topic have arisen in connection with metal-insulator-metal tunneling,^{3,4} and with the theory of localized magnetic states of transition-metal impurities in simple metals.¹⁸⁻²¹

The first application of the idea of tunneling to describe emission from a bulk material was the calculation of field emission from a free-electron metal made by Fowler and Nordheim²² in 1928. They considered a semi-infinite slab of metal, with work function ϕ , terminating abruptly at $z = 0$ as illustrated in Fig. 2. Their original calculation of the tunneling current through a triangular potential barrier into the vacuum has been extended in several directions. The image-potential rounding of the metal-vacuum potential has been included in some calculations of the current,^{23,24} but little effort²⁵ has been made to include refined details of the potential near the metal boundary.²⁶⁻²⁸ The influence of adsorbed impurities²⁹⁻³² and the effects of the metallic band structure³³⁻³⁴ on field emission have been predicted and (tentatively) observed in the postwar era.

Calculations of field emission from a bulk sample differ from those of autoionization probabilities in that the eigenvalue spectrum of the Schrödinger equation is continuous before ($F=0$ in Fig. 2) as well as after the imposition of the external field. This fact led to the exclusive

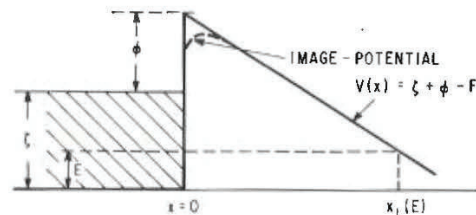


FIG. 2. Schematic illustration of the potential energy at a metal-vacuum interface. The solid line shows the triangular barrier used by Fowler and Nordheim, whereas the dashed line indicates the rounding due to the image potential. The Fermi energy of the free-electron metal is denoted by ζ .

Evidently, not all of these barriers are relevant for field emission. However, they all have been used in the analysis of some type of tunnel junction and are listed together for convenience. The emission current through these barriers, and any barriers that may be decomposed into segments of these barriers, may be calculated at an arbitrary temperature by quadrature of Eq. (5.13a). For most other barriers, either an approximate form of $D(E_x)$ can be used, or D itself may be calculated by numerical integration of Eq. (5.1). When evaluating claims made by authors who use the former procedure, one should keep in mind the rough correspondence between exact solutions of Eq. (5.1) for approximate potential barriers and approximate solutions of Eq. (5.1) for more refined barriers.

The most common approximation for D is the use of the semiclassical WKB form

$$D(E_x) = \exp[-2J(E_x)], \quad (5.17a)$$

$$J(E_x) = \int_{x_1(E_x)}^{x_2(E_x)} \kappa(x, E_x) dx, \quad (5.17b)$$

$$\kappa(x, E_x) = [2m(V_e(x) - E_x)/\hbar^2]^{1/2}. \quad (5.17c)$$

The quantities x_1 and x_2 are the classical turning points of an electron of energy E_x at the edges of the potential barrier as shown in Fig. 10. The approximation is adequate only if E_x is not near or above the top of the barrier ($E_x \ll V_m$) and if the sides of the barrier at x_1 and x_2 are gently sloping.⁸¹

We classify calculations reported in the literature according to the method used to evaluate $J(E_x)$. Four possibilities occur:

- (1) $J(E_x)$ is evaluated exactly.
- (2) $J(E_x)$ is evaluated approximately by expanding $\kappa(x, E_x)$ as a power series in $(E_x - \zeta)/\zeta$.
- (3) $J(E_x)$ is evaluated approximately by expansion of $\kappa(x, E_x)$ about some value of x [usually taken to be a local maximum of $V_e(x)$].
- (4) $J(E_x)$ is evaluated approximately by the simultaneous (bilinear) expansion of $\kappa(x, E_x)$ about fixed values of x and E_x .

Exact evaluations of $J(E_x)$ have been carried out for the Schottky image potential^{23,24,107} [given by Eq. (5.15)], the trapezoidal barrier,¹³⁴ the triangular barrier,¹³⁵ and various parabolic barriers.^{127,138,139} The expansion of $J(E_x)$ about $E_x = \zeta$ was initiated by Fowler and Nordheim²² and subsequently has been used in field emission calculations by almost all authors^{107,108,131,140} who do not evaluate the integral in (5.13a) numerically. Approximation methods 3 and 4 have been employed

primarily to calculate the tunneling current in metal-insulator-metal junctions (see Section 7) and the thermionic emission current over the potential barrier at the surface (see below).

At temperatures sufficiently high that $N(E_x)$ gives appreciable probabilities of thermal occupation of states at $E_x \sim V_m \sim [\phi + \zeta - (e^3 F)^{1/2}]$, the WKB formula (5.17a) no longer provides an adequate approximation for the $D(E_x)$ associated with the barrier illustrated in Fig. 10. Two approaches have been devised to develop a more adequate approximations for D in the energy region $E_x \gtrsim V_m$. The first¹⁴¹ (and still most satisfactory) technique used to calculate the transition probability of electrons passing over the top of the barrier in Fig. 10 is the direct solution to Eq. (5.1) obtained using a parabolic approximation to $V_e(x)$ near its maximum. The solutions to the parabolic potential are joined to WKB solutions in regions of space away from the maximum. The original application of this technique by Guth and Mullin¹⁴¹ subsequently has been refined by several authors.¹⁴²⁻¹⁴⁶ It underlies the current interpretation^{147,148} of thermionic emission from metals. The second approximation method is based on an extension^{149,150} of the WKB approximation in which Eq. (5.17a) is replaced by

$$D(E_x) = \{1 + \exp[2J(E_x)]\}^{-1}, \quad (5.18)$$

and the use of complex values of x_1 and x_2 for $E_x > V_m$ is incorporated in the evaluation of $J(E_x)$. However, authors who utilize both this method^{108,131} and the conventional WKB method^{107,151} [Eq. (5.17a)] either set^{107,131,151} $D(E_x) = 1$ for $E_x \geq V_m$ or use a parabolic-barrier formula¹⁰⁸ for $D(E_x)$ in this energy range.

We conclude this section on field emission by noting three relatively recent extensions of the uniform-field free-electron model. The first is the extension of the model, following the phenomenological point of view, to incorporate the consequences of the band structure in both metals^{33,34} and semiconductors.^{108,152-156} The semiconductor analyses by Stratton^{153,154,156} and Fischer¹⁵⁵ are based on the expansion of $\kappa(x, E_x)$ in the E_x variable using the WKB form (5.17a) in Eq. (5.13a). Errors by Stratton¹⁵⁴ and Fischer¹⁵⁵ are corrected by Stratton.¹⁵⁶ The extended WKB formula (5.18) is used by Christov¹⁰⁸ with various approximations for $D(E_x)$ being utilized in different ranges of values of E_x . The second extension of the simple theory of field emission is motivated by the observation that the model potentials considered so far always have contained a single uniform field F outside the metal (or semiconductor) surfaces. However, experiments carried out using field-emission microscopes measure the current from a metal tip on which

Two features of Eqs. (7.30) and (7.31) are significant in determining the magnetic-field dependence of the tunnel current. First, D depends on the Fermi energy, ζ_L , which itself is an oscillatory function of H if the average number of carriers in the left-hand metal is regarded as a thermodynamic constant.^{260,263} This oscillatory dependence of $\ln[D]$ on H can cause oscillations in j_t as a function of H for fixed bias (but not in j_t as a function of bias for fixed magnetic field). The oscillations in $\zeta_L(H)$ as a function of H are appreciable fractions of $\zeta_L(H=0)$ only if $\hbar\omega_c \sim \zeta_L$.²⁶³ However, Kulik and Gogadze²⁵⁹ propose that the use of pulsed magnetic fields permits such oscillations to be observed for an asymmetric metal-insulator-metal junction. They invoke the relative magnetic-field induced shift of the Fermi energies on the two sides of the junction as the mechanism responsible for creating an observable oscillatory voltage across the metal-insulator-metal "capacitor." Therefore, their prediction is not a direct consequence of the appearance of $\zeta_L(H)$ in the exponent in Eq. (7.31) for D (see, however, the discussion of Bernard *et al.*³⁰⁴ in Section 8b).

A second significant feature of Eq. (7.31) is the dependence of $\ln[D]$ on $\hbar\omega_c(n + \frac{1}{2})$. The effective barrier height for all the electrons increases because of the $\hbar\omega_c/2$ term. Furthermore, electrons in the lower Landau levels (e.g., $n=0$) have larger tunneling probabilities. The effect on the current of this dependence of the barrier height on n can be seen by calculating the current assuming ζ_L is fixed. It is conventional^{260a} to linearize the exponent in (7.31) as a function of $n\hbar\omega_c/V_B$. In the high-bias (Fowler-Nordheim) limit that $eV > V_B + \zeta_L$, we obtain

$$\ln[D/D_0] = -(\alpha/\zeta_L)[\zeta_L - E + (n + \frac{1}{2})\hbar\omega_c], \quad (7.32a)$$

$$D_0 = \exp\left[-\frac{4}{3}\left(\frac{2mV_B}{\hbar^2}\right)^{1/2} w\left(\frac{V_B}{eV}\right)\right], \quad (7.32b)$$

$$\alpha = -2(2mV_B/\hbar^2)w\zeta_L/eV. \quad (7.32c)$$

In this limit, all of the electrons can tunnel from the left-hand to the right-hand electrode. If H is increased at zero temperature and fixed bias, V , then, for $\alpha \lesssim 2$, the current oscillates^{260a} because the number of Landau levels contributing to the current (i.e., occupied) diminishes with increasing H . However, if $\alpha > 2$, the tunneling probability for tunneling from the $n \neq 0$ Landau levels becomes so small that the current appears to be a monotonically decreasing function of the magnetic field.^{260a} Although the analytical expression for the current [Eq. (6) given by Goldberg and Pollack^{260a}] formally still exhibits oscillations, they are too small to appear on sensible numerical graphs. A similar result is obtained for the current in the case that $\zeta_L(H)$ is an oscillatory

function of H due to the imposition of the constant carrier-number boundary condition.^{260a}

An extension of the current calculation to nonzero temperatures has been given by several authors.^{259,261} Unfortunately, in their analysis they seem to assume that the dependence of the barrier-penetration factor on the magnetic quantum number, n , is negligible relative to variations in density-of-states factors. Such an assumption is incorrect (for any barrier) if $\alpha > 2$ in the linearized expression for $\ln[D]$ analogous to Eqs. (7.32).

Another type of quantum size effect is predicted from planar tunnel junctions when the dimensions of one or more of the metallic electrodes become small relative to the wavelength of an electron at the Fermi energy. Using an average-barrier approximation for D , Gogadze and Kulik¹¹¹ predict an oscillatory component in the current due to density-of-states factors resulting from quantization in the electrodes of the electronic motion normal to the junction interface. Although some preliminary measurements of this effect have been reported for bismuth electrodes,¹¹² no detailed descriptions and analyses of these experiments have been given. Thus, for example, the separation of tunneling from transport^{264,265} effects has not been demonstrated. Also, the effects of fluctuations in the thickness of an electrode have not been discussed.

Our last topic is that of composite junctions. A new type of quantum size effects can occur in metal-insulator-metal-insulator-metal junctions when the intermediate metal becomes atomically thin. For sharp barriers, the one-dimensional theoretical model¹¹⁰ predicts transmission resonances analogous to those actually observed in field ionization of hydrogen near metallic surfaces.^{7,224} Similar resonances would be predicted if the intermediate "metal" were a semiconductor or semimetal. These effects, like the high-bias resonances in trapezoidal¹³⁴ and more complex²³⁶ metal-insulator-metal barriers, have not been observed.

Another type of composite junction that has been studied is the metal-insulator-metal "injection-electrode" used as a unit to inject electrons into some fourth material.^{266,267} Despite several experimental efforts,²⁶⁶⁻²⁷⁵ no careful (theoretical or experimental) studies of the operation of these emitters have been performed. An example of the lack of detailed understanding of the operation of these units is the assertion by Hall²⁷⁶ that the "tunnel injection" of electrons into germanium by a Au-Al₂O₃-Al emitter, as reported by Spratt *et al.*,²⁶⁸ is merely "pin-hole" emission through short circuits in the Al₂O₃.

Finally, a potentially significant utilization of composite junctions is for the production of injection electroluminescence. As in the case of metal-insulator-metal injection electrodes, only a few preliminary