

Room Temperature Gold-Vacuum-Gold Tunneling Experiments

Volume 91

Number 4

July-August 1986

E. Clayton Teague

National Bureau of Standards
Gaithersburg, MD 20899

An experiment has been completed which demonstrated quantum mechanical tunneling of electrons between two gold electrodes separated in vacuum. The tunneling current between the gold electrodes has been measured, for fixed voltages of 0.1 and 0.01 volts, as the electrode spacing was varied from a distance of approximately 2.0 nm down to a point where the electrodes touched. Current changes of over five orders of magnitude were found for electrode spacing changes of approximately 1.2 nm. For the first time, these data enable one to deduce the work function of the electrodes in a tunneling experiment from experimental parameters independent of the tunneling device. Also obtained were current-voltage characteristics for fixed electrode spacings in the direct tunneling region where electrode spacings were less than 2.0 nm. An analysis is given which attempts to deduce an absolute electrode spacing and tunneling area from the nonlinear properties of the I - V data and the cur-

rent versus spacing data. The analysis suggests that van der Waals and electrostatic forces play a major role in determining the I - V characteristics and that the tunneling area may be as small as 10^{-16} m². Along with a review of the theory of work functions and quantum mechanical tunneling, numerical calculations of the tunneling current based on the free-electron model of the electrodes and the barrier, an image-potential reduced barrier, and a WKB approximation for the tunneling probability have been performed and compared with Simmons' theory and with the experimental results.

Key words: metal-vacuum-metal tunneling; point-contact diode; scanning tunneling microscopy; temperature control; tunneling theory; vacuum tunneling; vibration isolation; work function

Accepted: April 24, 1986

Preface

The following paper is a reproduction of the author's dissertation which was completed and submitted to North Texas State University in August 1978. A majority of the experimental work was performed at the National Bureau of Standards. Results from the experiments were first reported at the Spring 1978 APS meeting and later at a Na-

tional Science Foundation Workshop on Opportunities for Microstructures Science, Engineering, and Technology held in Airlie, VA, November 19-22, 1978. While a complete description of the experiments and results is available as a publication from University Microfilms, publication in the more accessible form represented here has been prompted by requests from colleagues and by the significant number of recent references to the work. In accordance with the requests, the Board of Editors of the Journal of Research has agreed to publish the document unchanged except for the ad-

About the Author: E. Clayton Teague is with the Center for Manufacturing Engineering in the NBS National Engineering Laboratory.

Contents

Preface	171
1. Introduction	172
2. Some Basic Concepts of Metal-Vacuum-Metal Tunneling Theory	174
The Work Function and the Image Potential	175
The Barrier Potential Energy Function . . .	180
The Tunneling Process	183
Approximate Analytical Expressions for Tunneling Current	186
Numerical Calculations of Tunneling Current	188
Current Calculations for Spherical Electrodes	195
3. Experimental Apparatus and Method	195
Mechanical and Thermal Requirements . . .	195
Choosing an Electrode Material and Preparing the Electrodes	196
Temperature Control System	199
Vacuum System for the Tunneling Experiment	201
Electrode Holder and Micropositioning Assembly	202
Vibration and Acoustical Isolation	205
Tunneling Electronics	206
4. Experimental Results and Analysis . .	207
Data and Analysis	208
Effects of Mechanical Stress on Tunneling Characteristics	213
Conclusions	215
Acknowledgments	215
Appendices	
I. NUMERICAL CALCULATIONS OF TUNNELING CURRENT	216
II. STATISTICAL FLUCTUATIONS IN THE LENGTH OF A BAR	221
III. VACUUM CONDITIONS TO ACHIEVE AND MAINTAIN A CLEAN SOLID SURFACE	222
IV. AN ELECTRONIC DEVICE UTILIZING QUANTUM MECHANICAL TUNNELING	223
V. ELECTRONICS FOR THE TEMPERATURE CONTROL SYSTEM	225
References	232

dition of footnotes to identify recent work to which it is closely related.

Renewed interest in these experiments has primarily resulted from the success and current widespread use of the scanning tunneling microscope (STM) introduced by Binnig and Rohrer.¹ The findings of this work on vacuum tunneling should be of interest to many who are using or planning use of the STM. The current vs. voltage data, current vs. electrode spacing data, needs for temperature stability, and observations about the potential influences of van der Waals forces in these static *non-servoed tunneling experiments are important* in understanding the servoed operation used for tunneling in the STM. While servoed operation overcomes much of the sensitivity of the STM to effects from vibration, thermal expansions, and materials instabilities in the electrodes, these effects are proving to be significant factors in determining the stability and reproducibility of STM operation.

1. Introduction

According to quantum mechanics, a material particle can pass through a spatial region in which the particle's kinetic energy is less than its potential energy. The phenomenon is known as tunneling. A macroscopic example would be a ball which, when rolled toward a sharply peaked hill with insufficient speed to get over the top, does not roll back, but appears on the other side of the hill without affecting the ball or the hill and continues on its way. The probability of such an event occurring is astronomically small at the macroscopic level. However, for electrons and other fundamental particles and walls formed by nuclear and atomic forces, the probability while still small (one in 10^9 is not unusual) leads to major physical effects [1]². This work pertains to the tunneling of electrons *between two metal plates closely spaced in vacuum*.

Tunneling is not a new idea; it is in fact almost as old as quantum mechanics. In less than two years after Schrödinger [2] had published his now classic papers on "Quantization as a Problem of Proper Values," Oppenheimer [3] had used the wave me-

¹Binnig, G.; H. Rohrer, Ch. Gerber, and E. Wiebel, Surface studies by scanning tunneling microscopy, *Phys. Rev. Lett.* **49** (1982) 57; and Binnig, G., and H. Rohrer, Scanning tunneling microscopy, *Helv. Phys. Acta* **55** (1982) 726.

²Figures in brackets indicate literature references.

chanical concept of tunneling to explain the autoionization of atomic hydrogen, Fowler and Nordheim [4] had explained the main features of electron emission from cold metals in high external fields, and Gamow [5] had described the alpha particle decay process in heavy nuclei. The efforts by Frenkel [6] and by Holm [7] to describe the anomalous temperature dependence of the resistance of electrical contacts were the earliest studies of electron tunneling between metals.

Since these early studies, electron tunneling in solids has remained a very active area of solid state physics. Discoveries over the last 15 years by Giaever [8] that tunneling could be used to measure the properties of superconductors, by Esaki [9] of interband tunneling in semi-conductors, by Josephson [10] of the unique characteristics of tunneling between superconductors, and Lambe and Jakelvic [11] that tunneling could be used as a spectroscopic tool have had a major influence on the development of physics, and all have important technological consequences.

In the simplest form, an experiment to measure electron tunneling between two metals would be performed as shown in figure 1(a). Two parallel plates are separated by an insulator and connected to a variable voltage source. An ammeter and voltmeter are also included to measure the current flow and the applied voltage. Normally, a direct current would not flow through the circuit. However, if the plates are spaced very close together (on the order of 10 atomic diameters), a tunneling current will flow. If the plates could be made sufficiently flat and parallel over an area of one square millimeter, the current flow would be several amperes for an applied voltage of one-tenth volt.

Conventional devices for conducting these experiments are fabricated by evaporating one elec-

trode onto a glass substrate, forming an oxide layer with a fixed thickness of 1 to 3 nm on the electrode and then evaporating a second electrode over the oxide layer. These experiments have been widely used to measure the characteristics of Josephson junction devices, electrons and phonons in superconductors, semiconductor surfaces and junctions, the material used for the insulating barrier, and intentionally introduced impurities at the barrier-electrode interface. While these experiments have made and continue to make valuable contributions to solid state physics and technology, they have been seriously restricted by the small number of electrode materials which tend to form natural oxides suitable for a tunneling barrier. In addition, interpretation of the experimental results has been limited because of the difficulties in characterizing the atomic structure and composition of the insulator and the insulator-electrode interfaces. A major problem resulting from this lack of structure characterization is that the energy-momentum relationship of the electron while it is in the barrier cannot be modeled adequately.

By making a change from these conventional devices to one which uses a vacuum barrier, material restrictions are removed, electron properties in the barrier are well known, structure of the electrode-barrier interface is more defined and, in addition, the electrode spacing may be continuously changed. There are, however, some difficulties with vacuum barrier tunneling which must be weighed along with its advantages.

A schematic energy diagram for a metal-vacuum-metal tunneling structure is shown in figure 1 (b). The energy diagram is based on Sommerfeld's electron theory of metals in which all the conduction electrons of a metal are considered as being placed in a potential energy box and distributed

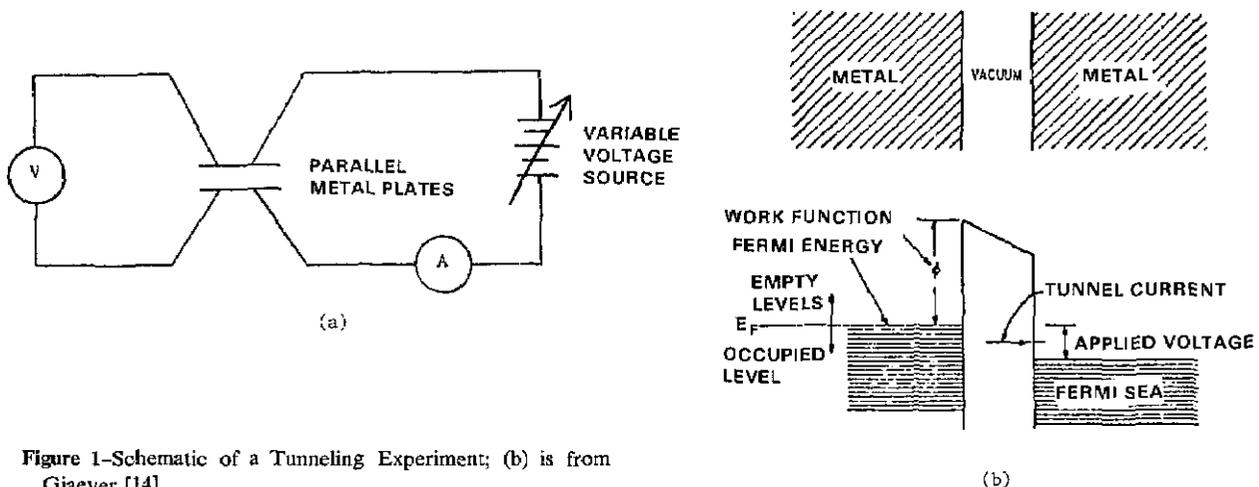


Figure 1-Schematic of a Tunneling Experiment; (b) is from Giaever [14].

over the allowed energy values according to Fermi-Dirac statistics. At zero temperature, the electrons, therefore, fill the energy levels up to a maximum energy, the Fermi energy. The energy required to remove one of the electrons from this energy to a large distance from the metal is known as the metal's work function. As shown in the figure, when a voltage is applied to one of the metals, its Fermi energy is lowered relative to the other; the energy difference is equal to the product of the magnitude of the charge on an electron and the applied voltage.

It will be shown in section 2 that an approximate expression for the tunneling current, I , under these conditions is given by

$$I = C \exp(-aSB^{1/2})AV, \quad (1-1)$$

where $C \approx 3.16 \times 10^{13} \text{ A/m}^2/\text{V}$, $a = 10.25 \text{ (nm)}^{-1} (\text{eV})^{-1/2}$, S is the electrode spacing, B is the average barrier height for an electron at the Fermi energy, A is the tunneling area, and V is the applied voltage. The important consequences of eq (1-1) for the vacuum tunneling experiment are that small changes in electrode spacing or mean barrier height have a very large effect on the tunneling current. A change in electrode spacing of 0.1 nm produces approximately a factor of 10 change in the tunneling current for a mean barrier height of 5 eV. Changes of about half this magnitude can be produced as a result of work function changes produced by the formation of oxide or other films on the electrode surface. Thus great care must be exercised in electrode preparation and in the mechanical and thermal design of the vacuum tunneling device.

Young et al. [12] and Thompson and Hanrahan [13] have previously performed vacuum barrier tunneling experiments. Both of these experiments were primarily concerned with observing the transition from Fowler-Nordheim field emission tunneling to direct vacuum tunneling. Young et al. [12] did report a preliminary observation of a linear current-voltage characteristic for direct tunneling at a spacing of approximately 1.2 nm. Thompson and Hanrahan [13] have not reported any current-voltage characteristics. Uncontrolled fluctuations in electrode spacing limited the exploration of direct vacuum tunneling properties in both experiments.

The objective of this work, which has evolved from that of Young et al., was to extend the previous vacuum barrier investigations to a detailed experimental study of the direct vacuum tunneling region. Such detailed experimental results would

provide a means to test the present theoretical framework of electron tunneling between solids. Some of the unresolved questions in this area: 1) What is the relationship between the classical electrostatic image potential and the quantum mechanical exchange and correlation potential? 2) What is the transit time for a tunneling electron? and 3) How is the distance between the last atomic planes of the two electrodes related to the tunneling barrier thickness?

Explicit goals of this work, which were partially achieved, were first, to measure the current through a vacuum tunneling device for fixed voltages of less than one-tenth volt as the electrode spacing was varied from a distance of approximately 2.0 nm down to a point where the electrodes touched; second, to measure the current-voltage characteristics for fixed electrode spacing in the direct tunneling region; third, to compare the measured results with available theories of electron tunneling; and finally, to establish a basis for using a vacuum tunneling device for use as a probe to measure the position of a surface to a precision of 0.1 nm, a precision 10 to 100 times greater than that of presently available techniques. Contributions to all these goals were obtained and will be described in later sections. The qualification on completely achieving the goals was given because the quality of electrode surfaces and the precision of electrode spacing measurements were not adequate in the area of the first three goals to resolve completely the questions posed about the theoretical framework.

A discussion of some fundamental concepts related to electron tunneling in solids, with particular emphasis on the origin of the work function and the barrier potential energy function, is given in section 2. Section 3 describes the requirements of the experimental apparatus for a vacuum tunneling experiment and the particular solution adopted for this work. Experimental results and an analysis of the relationship between the results and the numerical calculations of section 2 are given in section 4.

2. Some Basic Concepts of Metal-Vacuum-Metal Tunneling Theory

The objective of this section is to calculate, with the use of simple models, the current versus voltage (I - V) characteristics and, for a fixed voltage, the current versus spacing (I_{cv} - S) characteristics of the gold-vacuum-gold tunneling device used in the experiments of this work. The primary results of these calculations are tables and graphs of

the I - V and I_{cv} - S characteristics which can be used for comparison with the measured data to be presented in section 4.

While quite thorough treatments of electron tunneling in conventional metal-insulator-metal junctions are available in the literature [14-23], no numerical results have previously been published which fully accounted for the image potential variation with electrode spacing and applied bias voltage, both of which are essential for comparison with the metal-vacuum-metal tunneling data. Additional motivations for performing the calculations were that 1) almost all of the available theoretical results have been seriously questioned in terms of numerical accuracy and overall validity of the analytical approximations which were employed [24], and 2) even if one were to accept the validity of results which have been obtained, the effort to extrapolate and apply these approximate results to the gold-vacuum-gold tunneling device would be almost as great as carrying out the more accurate numerical calculations.

The content of the following discussion is therefore 1) a description of models of the electron properties of the tunneling device and of the tunneling process, 2) a development based on these models to obtain analytical expressions for the tunneling current which are, within the limitations of the models, "almost" exact, 3) a reduction of the exact expressions with commonly used approximations to obtain analytical solutions for I vs. V and S in terms of experimental parameters, and 4) a description of the procedure employed to carry out numerical computations of I vs. V and S along with a presentation of results from these computations. No attempt will be made to give either a review or a comprehensive survey of the literature that is related to the topics of this chapter. Tutorial coverages of the theoretical and experimental aspects of electron tunneling into semiconductors, metals, and superconductors are given in the book edited by E. Burstein and S. Lundqvist [14]. A comprehensive review and digest of the tunneling literature up through 1969 has been given by C. B. Duke [15]. Recent development in the theories of tunneling are covered in two papers of debate between T. E. Feuchtwang and C. Caroli et al. [21] and references contained therein.

The Work Function and the Image Potential

By its nature, the metal-vacuum-metal tunneling device can be completely represented by a model which accounts for the electron properties of two metal-vacuum interfaces and their interaction. Ne-

glecting for the moment the interaction, the problem is then to obtain a model that adequately describes a uniform metal and its interface to vacuum. In this section properties of the metal-vacuum interface will be developed in terms of the general theory of metal surfaces formulated by Lang and Kohn; LKI [25], LKII [26], LKIII [27]. Their theory will be used as a basis to describe the origins of the metal work function and to obtain a potential energy function for the metal-vacuum interface.

Lang and Kohn's work is based on the density-functional theory of inhomogeneous electron gas systems formulated by Hohenberg, Kohn, and Sham (HKS) [28]. The HKS formalism is a self-consistent-field method and as such its fundamental idea is to replace the actual interactions; electrostatic, exchange, and correlation, between the electrons and the ion-cores by an interaction between one electron and an averaged self-consistent potential produced by the other electrons and the ion-cores. It therefore fully accounts for the effects of electron exchange and correlation and is in principle capable of providing exact results. Since the concepts of this formalism will be used so extensively in the following discussion of the work function and the surface potential energy function, a brief summary of the HKS theory, based in part on a review by Hedin and Lundqvist [29], will now be given.

The theory rests on two hypotheses. One states that the ground-state wavefunction of an interacting many-electron system, and thus all ground-state properties, are functionals of the electron density $N(\mathbf{r})$. The other states that the ground state energy for a given external potential $U(\mathbf{r})$ can be written as

$$E_u[N(\mathbf{r})] = \int U(\mathbf{r})N(\mathbf{r})d\mathbf{r} + F[N(\mathbf{r})], \quad (2-1)$$

where F is a universal functional of N , independent of the external potential U , and that E_u is a minimum for the correct density function $N(\mathbf{r})$.

With these hypotheses as a basis, HKS [28] further show that the total electronic ground-state energy of a many-electron system in an external potential $U(\mathbf{r})$ can be written in the following form:

$$\begin{aligned} E_u[N] = & e \int U(\mathbf{r})N(\mathbf{r})d\mathbf{r} \\ & + (e^2/2) \int \frac{N(\mathbf{r})N(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ & + T_s[N] + E_{xc}[N]. \end{aligned} \quad (2-2)$$

The functional $T_s[N]$ is the kinetic energy of a non-interacting electron system, and $E_{xc}[N]$ represents

the exchange and correlation energy. The first two terms in eq (2-2) represent the electron gas' electrostatic potential energy in the field $U(\mathbf{r})$ and the Coulomb self-energy of the electrons. All the complexities of the many-body interactions are contained in the functional $E_{xc}[N]$.

An effective potential, U_{eff} is then defined as

$$U_{\text{eff}}[N;\mathbf{r}] = U(\mathbf{r}) + e \int \frac{N(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + U_{xc}[N;\mathbf{r}], \quad (2-3)$$

where

$$U_{xc}[N;\mathbf{r}] = \delta E_{xc}[N] / \delta N(\mathbf{r}). \quad (2-4)$$

$U_{\text{eff}}[N;\mathbf{r}]$ is then shown to play the role of an effective single particle potential so that $E_u = T_s[N] + U_{\text{eff}}[N;\mathbf{r}]$. δ denotes "variation of" in the calculus of variations sense. This expression for the total ground-state energy combined with an assumed form of $E_{xc}[N]$ enables one to solve the following self-consistency problem to obtain an exact density distribution of the system of N interacting electrons:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + U_{\text{eff}}[N;\mathbf{r}]\right] \psi_i = e_i \psi_i \quad (2-5)$$

$$N(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \Theta(\mu - e_i), \quad (2-6)$$

m is the electron mass and \hbar is Planck's constant divided by 2π . The sum is over the N lowest-lying orthonormal solutions, ψ_i , of eq (2-5). $\Theta(x)$ is unity for $x < 0$, and zero for $x > 0$, μ is the chemical potential of the metal. The e_i are Lagrange multipliers introduced to fix the number of electrons as the states ψ_i are varied. Their identification as single particle energies is discussed later in this section. At finite temperatures, U_{xc} is defined in terms of the system's free energy and the Θ function is generalized to the Fermi distribution function.

The many-body problem has, therefore, been reduced to a set of one-body problems coupled only through the effective one-body potential $U_{\text{eff}}[N;\mathbf{r}]$. Solution of eqs (2-3 to 2-6) proceeds in the usual manner: one begins with an assumed $N(\mathbf{r})$ and $E_{xc}[N(\mathbf{r})]$, constructs $U_{\text{eff}}[N;\mathbf{r}]$ from (2-3) and then finally calculates a new $N(\mathbf{r})$ with the use of (2-5 and 2-6). The procedure is then repeated until the new and old $N(\mathbf{r})$ differ by acceptable tolerances.

Lang and Kohn (LKI) calculate the electronic structure at a metal-vacuum interface by applying this formalism to the jellium model of a metal. In

this model the positive ion-cores of the metal-lattice are replaced by a uniform, spatially fixed background of charge with density

$$N_+(\mathbf{r}) = N_a, \quad x < 0 \\ = 0, \quad x > 0. \quad (2-7)$$

The conduction or valence electrons with density $N(\mathbf{r})$ are free to move and settle in an equilibrium configuration with respect to this fixed background of positive charge. $N(\mathbf{r})$ is subject to the charge neutrality condition.

$$\int [N_+(\mathbf{r}) - N(\mathbf{r})] d\mathbf{r} = 0. \quad (2-8)$$

For the tunneling experiment this model is a reasonable approximation since the ion-core lattice and its resultant electron band structure are not expected to significantly affect the characteristics of the tunneling current.

The electronic structure of an inhomogeneous electron gas with the inhomogeneity being produced by the background of eq (2-7) is calculated in LKI. They demonstrate by numerical calculations that even though the electron density in the surface region of a metal varies quite rapidly, good agreement with experimental results is obtained if one uses the approximation

$$E_{xc}[N] = \int e_{xc}(N(\mathbf{r})) N(\mathbf{r}) d\mathbf{r}, \quad (2-9)$$

e_{xc} is the exchange and correlation energy per particle of a uniform electron gas. By making use of Wigner's form of $e_{xc}(N)$

$$e_{xc} = -0.458/r_s(N) - 0.44/(r_s(N) + 7.8), \quad (2-10)$$

$$r_s(N) = [3/(4\pi a_B^3 N)]^{1/3}, \quad (2-11)$$

and a $U(\mathbf{r})$ determined by eq (2-7), the functions $N(x)$, $U_{\text{eff}}(x)$, and $V(x)$, the electrostatic part of U_{eff} , are calculated for several values of the electron gas density. In eqs (2-10 and 2-11), e_{xc} is in units of Hartrees. (1 Hartree = 27.2 eV), a_B is the Bohr radius, N is the uniform background density in number/cm³, and r_s is the Wigner-Seitz radius for a density N . Figure 2 is a graph constructed from a table given in LKI of these three functions for an electron density very close to that of electrons in gold, i.e., 5.76×10^{22} electrons/cm³. This density corresponds to an r_s of 3.03.

Equations (2-3 and 2-4) in connection with the results plotted in figure 2 shows that for x less than

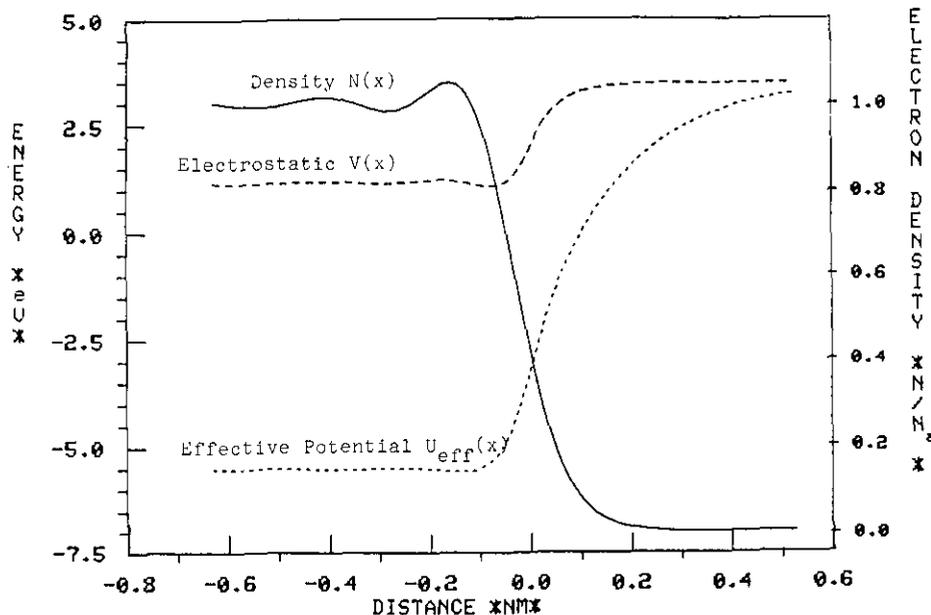


Figure 2—Graph of Electron Density, Effective Electron Potential Energy, and Electrostatic Component of Potential Energy at a Metal-Vacuum Interface. The positive background density corresponds to a Wigner-Seitz radius of 3. The graph is based on data from a table in N. D. Lang and W. Kohn, Phys. Rev. B1 (1970) 4555.

about -1.0 nm, U_{eff} approaches a constant value within the bulk of the material as anticipated. Equation (2-5) then simplifies to the one particle Schrödinger equation

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_E(-\infty) + \mu_{\text{xc}}(N_a)\right] \psi_i = e_i \psi_i. \quad (2-12)$$

$V_E(-\infty)$ is the electrostatic potential energy of an electron in the metal interior; $\mu_{\text{xc}}(N)$ is the exchange and correlation part of the total chemical potential $\mu(N)$ of an infinite uniform gas of density N . From the definitions given earlier for $E_{\text{xc}}[N]$ and e_{xc} , and the definition of the chemical potential,

$$\mu_{\text{xc}}(N) = d((Ne_{\text{xc}}(N)))/dn \quad (2-13)$$

or

$$= e_{\text{xc}}(N) + N d(e_{\text{xc}}(N))/dn. \quad (2-14)$$

If the zero of energy is chosen such that

$$V_E(-\infty) + \mu_{\text{xc}}(N_a) = 0, \quad (2-15)$$

then the equation describing the electron motion in the electrodes at distance greater than 1 nm from the surface is

$$-\frac{\hbar^2}{2m} \nabla^2 \psi_i = e_i \psi_i. \quad (2-16)$$

Therefore, for the case when $N(\mathbf{r})$ is a constant, the electrons in an interacting gas move as free particles.

Interpretation of the eigenvalues of eq (2-5) as the excitation energies of the elementary excita-

tions of the electron gas is not valid for a general $U_{\text{eff}}[N; \mathbf{r}]$ [29-31]. However, arguments are presented by Moruzzi et al. [31] that demonstrate that for most bulk simple metals this interpretation is a very good approximation. Slater [30] gives some numerical estimates of the energy errors involved in using this approximation for self-consistent calculations in atoms. Hedin and Lundqvist [29] consider the problem from the Green function standpoint and one may conclude from a comparison of eq (2-5) with their equation for the excitation energies of quasielectrons that for uniform positive backgrounds the resulting energies and those of (2-5) would be the same. However, nonuniformities in the electron gas such as that generated by the pseudopotentials of the ion-cores and the discontinuity at a metal surface cause all the usual differences between one-electron theory and many-electron theory to become manifest, i.e., self-energy corrections, relaxation times, and different energy-momentum relationships.

The work from eqs (2-1 through 2-16) may appear an overly complex way of obtaining this free-electron result. However, the reader should observe that in addition to the free-electron behavior, this approach has also, by eqs (2-12 and 2-15), very naturally given a means to determine the binding energy of the lowest energy state. The results also clearly demonstrate that the binding energy is due both to the electrostatic energy produced by the existence of a surface and to the energy resulting from exchange and correlation. Finally, the development through the calculated U_{eff} has produced a realistic potential energy function at a metal surface.

The work function, WF , of a metal provides the necessary reference for relating $V(-\infty)$ to the vacuum energy level. It is defined as the minimum energy required to separate an electron from the metal at zero temperature. Thus,

$$WF = (V_E(+\infty) + E_{N-1}) - E_N \quad (2-17)$$

where $V_E(+\infty)$ is the total electrostatic potential energy far from the metal surface and E_M is the ground state energy of the metal with M electrons. Arbitrary zeros of energy for V and E do not affect the result of eq (2-17). The chemical potential μ of a system is defined as $E_N - E_{N-1}$, thus eq (2-17) can also be written as:

$$WF = V_E(+\infty) - \mu. \quad (2-18)$$

Since the electrostatic potential in the interior of a metal with and without a surface differs by $V(-\infty)$, μ for a metal with a surface is

$$\mu = V_E(-\infty) + E_F + \mu_{xc}(N_s), \quad (2-19)$$

E_F is the Fermi energy of a degenerate electron gas of density N_s . WF is therefore

$$WF = [V_E(+\infty) - V_E(-\infty)] - E_F - \mu_{xc}(N_s). \quad (2-20)$$

In a more compact form,

$$WF = \Delta V_E - \bar{\mu}, \quad (2-21)$$

where $\Delta V_E = V_E(+\infty) - V_E(-\infty)$ and

$$\bar{\mu} = E_F + \mu_{xc}(N_s). \quad (2-22)$$

In this form the WF is seen to consist of two components, a surface part ΔV_E , which is the rise in mean electrostatic potential energy across the metal surface and $\bar{\mu}$, a bulk part which is the intrinsic bulk chemical potential of the electrons relative to the mean electrostatic potential energy in the bulk interior. ΔV_E is due to the dipole layer created by the spilling-out of the electrons as shown by the distribution in figure 2. All many-body effects are contained in the exchange and correlation contributions to $\bar{\mu}$ and in their effect on the barrier potential ΔV_E .

Combining eq (2-20) with the zero energy convention of eq (2-15) shows that the lowest energy eigenvalue of eq (2-16) is at an energy of $[\Delta V_E + \mu_{xc}(N_s)]$ less than $V(+\infty)$. Eigenstates of eq (2-16) are then according to the Pauli exclusion principle occupied at zero temperature from this energy up to the maximum value E_F . A diagram of the electron energy as a function of x position in the region of a metal vacuum interface is given in figure 3.

The work function of gold was estimated in LKH by again using eq (2-13) and Wigner's expression for $e_{xc}(N)$ to obtain μ ; ΔV_E was calculated

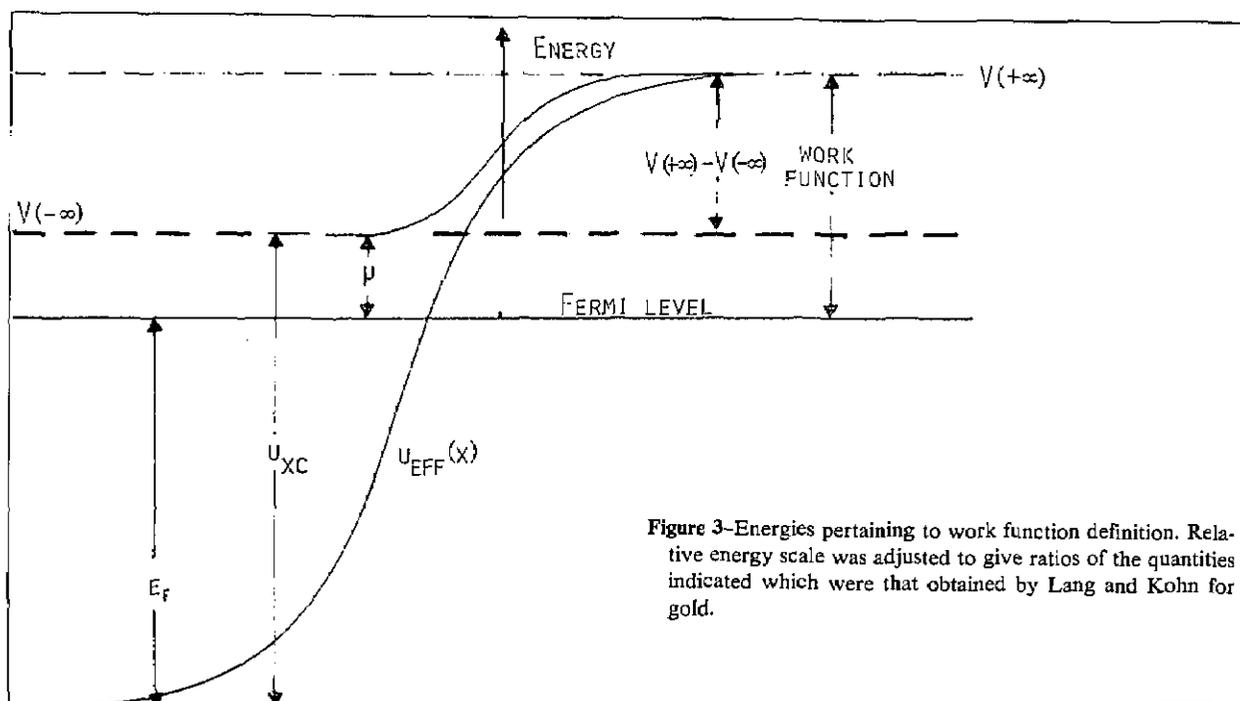


Figure 3—Energies pertaining to work function definition. Relative energy scale was adjusted to give ratios of the quantities indicated which were that obtained by Lang and Kohn for gold.

from the double-layer solution of Poisson's equation

$$\Delta V_E = (e^2/\epsilon_0) \int x [N(x) - N_+(x)] dx \quad (2-23)$$

and the density distributions obtained from self-consistent solutions of eqs (2-5 and 2-6). In eq (2-23), e is the electron charge and ϵ_0 is the free space dielectric constant. The results for gold were (all in units of electron volts)

$$\begin{aligned} E_F &= 5.57, \mu_{xc} = -6.75, \Delta V_E \\ &= 2.32, WF = 3.50. \end{aligned} \quad (2-24)$$

Experimental values of the work function are [32] 5.0 ± 0.2 eV for polycrystalline samples; the uncertainty accounts for the range of values obtained by various experimental techniques. Attempts to achieve better agreement by incorporating lattice effects through a pseudopotential were not successful.

It is not clear how the discrepancy between the theoretical and experimental work functions should be apportioned between μ_{xc} and ΔV_E . But regardless of this difficulty, Lang and Kohn's calculations have shown in a definite way the relative magnitudes of the two major contributing terms to the work function. In addition, this insight into how the work function energy is generated will later prove useful in interpreting the effects of absorbed molecular species on the values of the work function.

There remains one further significant question about the electron structure at a metal-vacuum interface: For purposes of a tunneling experiment, what is the potential energy of an electron 1 to 10 nm outside a metal surface? A related question which must also be answered is: Where is the effective metal surface? Both questions are addressed in LKIII. By solving eqs (2-5 and 2-6) self-consistently, an $N(\mathbf{r})$ is calculated from a U_{eff} determined by the sum of the energies due to the distribution of eq (2-7), a point charge outside the metal surface, and a weak external electric field F . The unperturbed charge distribution produced by the charge density of eq (2-7) is then subtracted from this result to obtain a screening charge density of $N_s(x)$. The potential energy of an electron outside the surface in the presence of a weak uniform electric field pointing out of the metal is then calculated as the interaction energy between the potential of the point charge, q , and its induced screening charge.

The result for a charge located at the position x_1 is to order $(x_1 - x_0)^{-3}$ given by

$$\begin{aligned} V_{\text{IP}}(x) &= \text{const} - q^2/[4\pi\epsilon_0(x_1 - x_0)] \\ &\quad - qE(x - x_0). \end{aligned} \quad (2-25)$$

The center of mass, x_0 , of the induced screening charge is defined by

$$x_0 = \int x N_s(x) dx / \int N_s(x) dx. \quad (2-26)$$

For small fields, the center of mass of the induced screening charge is approximately the same as that for the point charge. The answer to the first question posed is, therefore, that the potential to a good approximation is the classical image potential referred to a plane passing through x_0 . Relative to the edge x_B of the uniform positive-charge background, the position x_0 is determined by $x_0 - x_B = 0.076$ nm for an $r_s = 3$, i.e., gold's electron density. This value was obtained by interpolating between the values of $x_0 - x_B$ given in LKIII for $r_s = 2$ and $r_s = 4$. x_0 and x_B are located schematically in figure 4.

In connection with the problem of determining the potential at a metal surface, field penetration effects are often considered to be a significant factor. As pointed out by Lang and Kohn, the analogy here of penetration effects is just the fact that the induced charge distributions have finite widths.

A definite answer for the second question is not as easily obtained. Locating x_B of the positive background model relative to the lattice of the metal is the problem. The approach taken in LKIII was to locate, somewhat arbitrarily, x_B at a point one-half an interplanar spacing in front of the last plane of ions. A justification for this assignment can be obtained from the self-consistent calculations by Applebaum and Hamann [34] for the Na (100) surface in which the lattice properties and atomic nature of the surface were treated at the same level of detail as is common in bulk calculations. They give a figure which shows the effective potential and charge density in much the same manner as figure 2, with the ion-core locations shown. Taking the point at which the electron charge density has dropped to one-half the bulk value as the proper place to locate the positive-background edge gives the x_B value assumed by LKIII.

In summary, the electron potential outside a gold surface is given by eq (2-25) with x_0 located at a point one-half an interplanar spacing plus 0.076 nm in front of the last plane of ions (see fig. 4). This along with U_{eff} is sufficient information to deter-

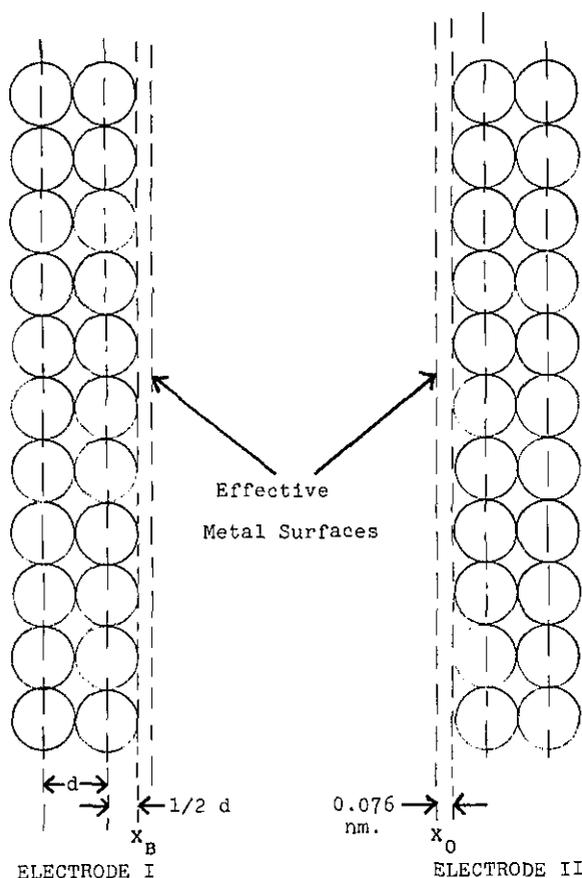


Figure 4—Location of effective metal surfaces. Diagram illustrates simple model of two identical gold electrodes with the atoms spaced at the nearest-neighbor distance in a gold crystal. The electrode spacing shown would correspond to a separation of the respective last ion planes of approximately 1.73 nm; the nearest-neighbor distance was assumed to be 0.288 nm.

mine the potential between two gold electrodes in vacuum.

However, the effective spacing between the electrodes for tunneling is reduced even more than twice the sum just given since the tunneling distance will, as shown later, be determined primarily by the distance between the points of the electrodes at which $U_{\text{eff}}=E_F$. For gold (see fig. 2) the effective tunneling surface, by this convention, would be located 0.105 nm plus one-half an interplanar spacing in front of the last plane of ions. This is the distance obtained from figure 2 at which $U_{\text{eff}}=0$; x_B in figure 2 is at the ordinate 0. For an interplanar spacing, d , the location of the gold surface for tunneling, x_s , relative to the location of the last plane of ions, x_i , is therefore

$$x_s = x_i + (x_B - x_i) + (x_0 - x_B) + t \quad (2-27)$$

or,

$$x_s = x_i + 0.5d + 0.076 \text{ nm} + t, \quad (2-28)$$

where t is the distance from x_0 to where $U_{\text{eff}}=E_F$; thus, $t=0.105-0.076=0.029$ nm. The positive x direction is out of the metal surface.

At large, positive x , $U_{\text{sc}}[N(\mathbf{r})]$ is thought to become the classical image potential [$U_{\text{sc}}(x) \rightarrow (e^2/4\pi\epsilon_0 x)$] [25,33]. However, no form of the exchange and correlation potential has as yet been put forth which has the expected asymptotic functional dependence on x . The transition of an electron from inside a metal in the presence of many other electrons where energy changes result from electrons of like spin avoiding each other (exchange) and electrons of either spin avoiding each other to minimize the coulomb energy (correlation) to a region in vacuum where at least the correlation is an insignificant factor has been extremely difficult to model. The slowly varying density approximation for $e_{\text{xc}}[N]$ given in eq (2-9) leads to an effective exchange and correlation potential which vanishes exponentially as x goes to infinity (see LKI). A control calculation was performed in LKI by checking the self-consistent results against that obtained when the image potential ($1/x$) form was used for x values greater than the point where $U_{\text{eff}}=E_F$. Differences in $N(x)$ obtained with the two potentials were no greater than 0.3% for $r_s=2.5$. The more commonly used Slater $X-\alpha$ approximation for the exchange and correlation potential varies as $N(x)^{1/3}$. With reasonable $N(x)$'s it also vanishes much more rapidly than $1/x$ as x goes to infinity [34].

The Barrier Potential Energy Function

In this section and the following section on the tunneling process, the tunneling device is assumed to be composed of two semi-infinite metal electrodes with plane parallel faces separated by a vacuum space. The parallel faces are taken to be normal to the x axis with electrode II located further along the positive x direction than electrode I.

Assume first that the electrodes are in equilibrium with each other. Then, the chemical potentials of the two electrodes will be equal [35] and eq (2-19) gives

$$V_E^I(-\infty) + E_F^I + \mu_{\text{xc}}^I(N_a^I) = V_E^{II}(-\infty) + E_F^{II} + \mu_{\text{xc}}^{II}(N_a^{II}). \quad (2-29)$$

The superscripts denote the electrode labels. When the electrodes are not contacting this condition

combined with the definition of the work function in eq (2-20) requires that

$$V_V = V_E^I(+\infty) - V_E^{II}(+\infty) = WF^I - WF^{II}. \quad (2-30)$$

Thus, if the two electrodes have different work functions, the equilibrium condition is for there to exist an electrostatic potential difference between their surfaces. This potential difference is usually known as the Volta or contact potential [36]. Recent work by Bennett and Duke [37] has, however, shown that the name contact potential would be inappropriate since they conclude that when the metals are in contact, the junction potential, V_J , is

$$V_J = E_F^I + \mu_{xc}^I(N_a^I) - [E_F^{II} + \mu_{xc}^{II}(N_a^{II})]. \quad (2-31)$$

This result interpreted in terms of eq (2-29) means that when the electrodes are in contact, their background electrostatic potentials, $V(-\infty)$, are constrained to be equal; a very reasonable conclusion. Bennett and Duke's [37] arguments are not based on eq (2-29). Their results follow from solutions to Schrödinger's and Poisson's equations with appropriate boundary conditions.

The problem of describing the transition between the conditions specified by eqs (2-30 and 2-31) has, unfortunately for use in a general vacuum tunneling configuration, not been investigated. For the particular configuration, i.e., with almost identical gold electrodes, used in the experiments of this work the effects from both V_J and V_V were minimal. In the following discussion, V_V and V_J are assumed to be zero for the gold-vacuum-gold tunneling experiment.

Assuming equal Fermi energies and work functions in the two electrodes, the barrier potential energy function, $V_B(x)$ at equilibrium is, following the conventions of figure 3 given by

$$V_B(x) = E_F + WF + V_{IP}(x), \quad (2-32)$$

where $V_{IP}(x)$ is the image potential of a point charge between two conductors with plane parallel faces. The energy zero is assumed to be such that $V_E^I(-\infty) + \mu_{xc}^I(N_a^I) = 0$. The classical image potential (2-25) is employed because of the difficulties given earlier about the incorrect extrapolations of available forms of the exchange and correlation potentials. In vacuum, the classical double image potential energy function is [38]

$$V_{IP}(x) = -\frac{e^2}{8\pi\epsilon_0} \left(\frac{1}{2x} + \sum_{n=1, \infty} \frac{ns}{n^2s^2 - x^2} - \frac{1}{ns} \right), \quad (2-33)$$

where s is the distance between the faces of the two conductors and x is the coordinate relative to an origin at the face of electrode I. To facilitate numerical calculations, a simpler approximate form of eq (2-33) obtained by Simmons [38] will be used. The approximate form is

$$V_{IP}(x) = -1.15(\ln 2)e^2 \frac{s}{8\pi\epsilon_0 x(s-x)}. \quad (2-34)$$

With s and x expressed in nm and $V_{IP}(x)$ in electron volts (eV), eq (2-34) reduces to

$$V_{IP}(x) = -0.2867 \frac{s}{x(s-x)}. \quad (2-35)$$

Plots of $V_B(x)$ with the V_{IP} of eq (2-35) and assumed electrode Fermi energies of 5.5 eV and work functions of 5.0 eV for s values ranging from 0.2 nm to 5.0 nm are given in figures 5 and 6. Note that in these figures, the vacuum level is at an energy of 10.5 eV. These graphs demonstrate that as the electrode spacing decreases, the maximum barrier height drops by more than one electron volt when $s = 1.0$ nm; when $s = 0.23$ nm, the maximum barrier height has dropped to the Fermi energy of 5.5 eV. This is coincidentally about equal to the nearest-neighbor distance for the gold lattice.

For the barrier potential energy function plotted in figures 5 and 6 to correspond with the surface location described by eq (2-28) and the corrections to the classical image potential given in eq (2-25), the $x = 0$ and $x = s$ positions would correspond to the x_0 described in these equations. While t , the distance from x_0 to the x value at which $U_{eff} = E_F$, was found in LKIII to be 0.029 nm for gold, the value obtained from figure 5 is approximately 0.06 nm. Neglecting for the moment the possible effects of nonzero t , the modifications of eqs (2-25 and 2-28) mean that tunneling currents calculated from the barrier potential with a spacing parameter s should correspond to a true physical spacing, s_T ,

$$s_T = x_i^{II} - x_i^I = s + d + 0.152 \text{ nm}. \quad (2-36)$$

The importance of eq (2-36) will be discussed in section 4 when an attempt is made to relate the calculated current values to measurements.

When the positive terminal of a battery of potential V is connected to electrode II and the negative terminal to electrode I, the mean electrostatic potential, for negative charges, of electrode II is lowered by an energy of eV relative to that in electrode I. Since the E_F and μ_{xc} are fixed by the

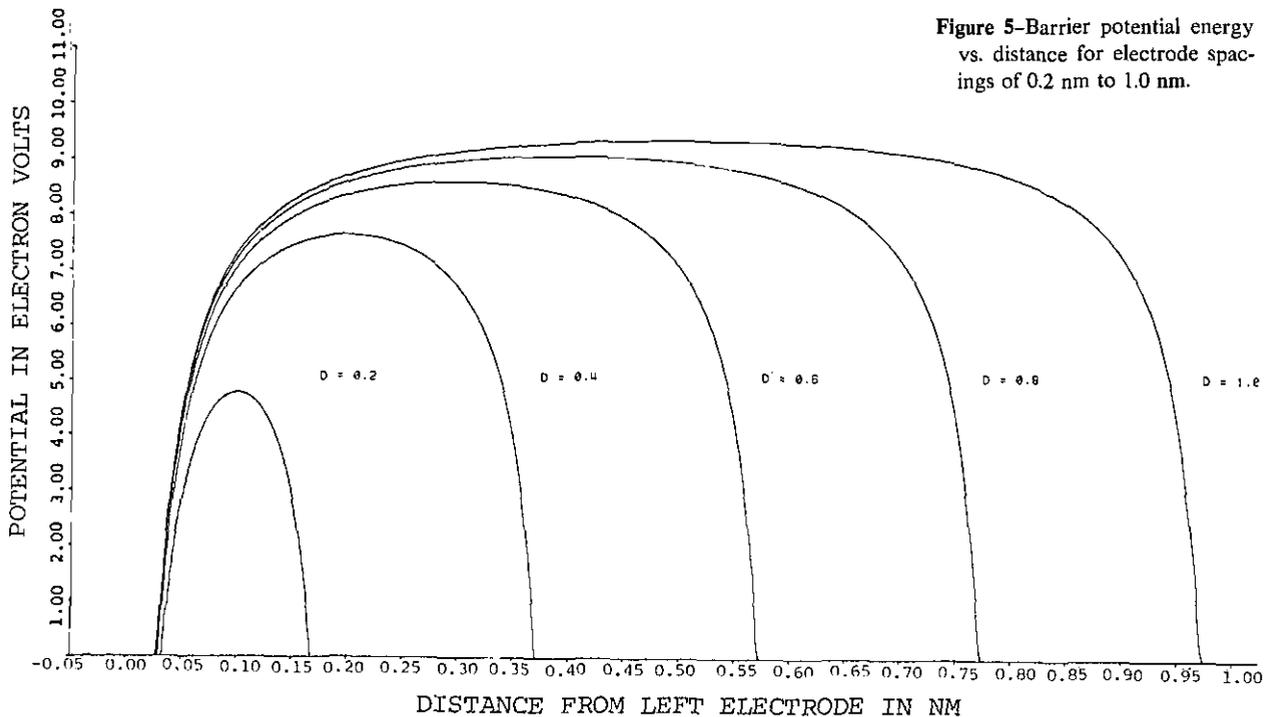


Figure 5—Barrier potential energy vs. distance for electrode spacings of 0.2 nm to 1.0 nm.

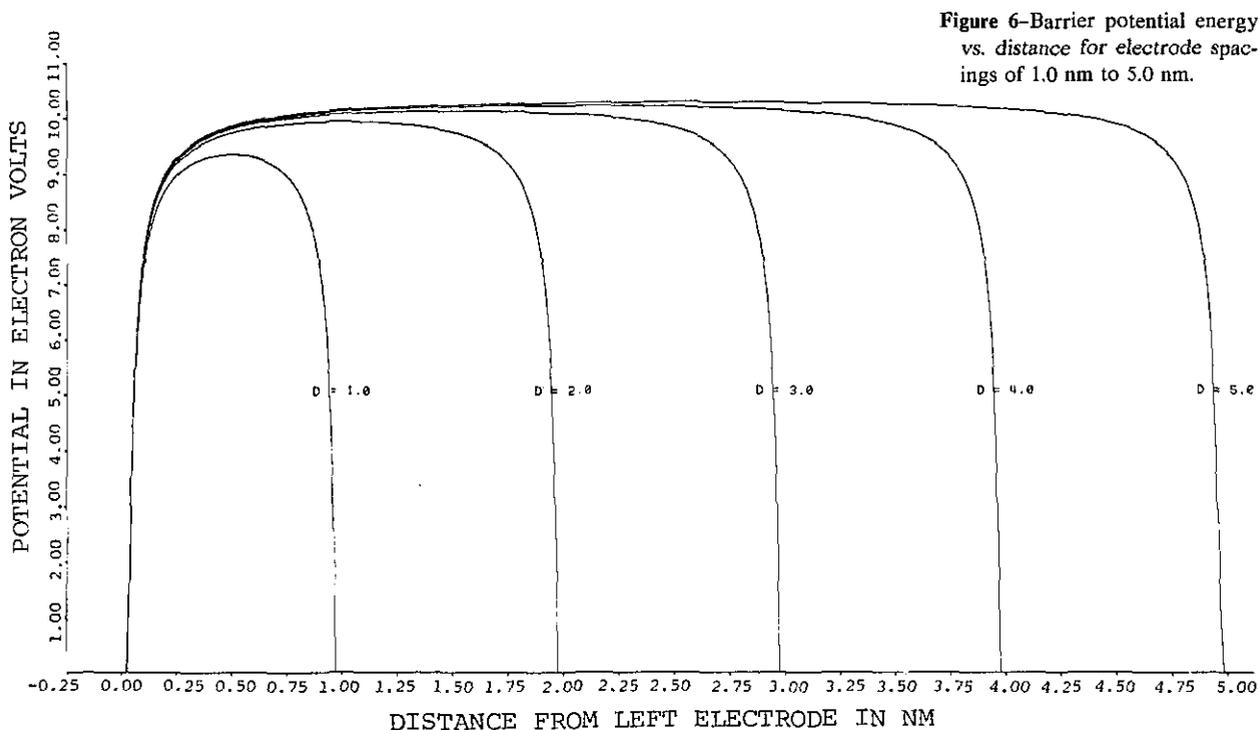


Figure 6—Barrier potential energy vs. distance for electrode spacings of 1.0 nm to 5.0 nm.

electron density, the Fermi level is also lowered by an energy eV . Combining the barrier potential energy function with this effect and assuming the field between the electrodes is uniform gives a potential energy function for the metal-vacuum-metal system which can be subdivided into three areas of interest:

Region I: $V_T(x)=0 \quad x \leq x^I$

Region II: $V_T(x)=-eV \quad x^I \geq x$

Region III: $V_B(x)=E_F+WF - eVx/s$

$+V_{IP}(x)x^I > x > x^II, \quad (2-37)$

x^I and x^{II} are the values for which $V_B(x)$ is equal to 0 and $-eV$, respectively. Figure 7 is a schematic illustration of eq (2-37) combined with the concepts of figure 3 concerning the work function. The areas of the diagram up to the Fermi energy represent eigenstates that are occupied when the electrodes are at zero temperature.

The Tunneling Process

From this point onward in the theoretical development no further attempt will be made to account for many-electron effects on the calculated tunneling current. All the previous work to account for many-electron effects was required since, as shown, the dominant sources of the work function originate from electron-electron interactions and their effects on the electron distribution at a metal surface. A fully independent-electron model of the electrodes and of the tunneling process will be used in the remainder of this section. There were two major reasons for this decision. First, at the resolutions of I - V measurements anticipated for this experiment, the small changes in current due to electron-electron interactions, electron-phonon interactions, and self-energy changes produced by transition through the barrier would not be observable. The changes are usually less than one percent. Second, within the limitations of the independent electron model, the stationary state method of current-carrying states is a well-defined and convenient procedure for calculating current flow in quantum systems [39,40].

For an electron described by one of the solutions, $\psi_k(r)$, to the time-independent Schrödinger equation for the potential $V_T(x)$

$$\left[-\frac{\hbar^2}{2m} + V_T(x) \right] \psi_k(r) = E(k) \psi_k(r), \quad (2-38)$$

the electron current density flow is given by [40]

$$\mathbf{J}(k, r) = \frac{-ie\hbar}{2m} (\psi_k \nabla \psi_k^* - \psi_k^* \nabla \psi_k), \quad (2-39)$$

where * denotes the complex conjugate operation. The standard interpretation of eq (2-39) is that $\mathbf{J}\cdot\mathbf{n}$ is the probability per unit time that a charge $-e$ will flow through the unit area with normal \mathbf{n} . Since the electrons occupying the states $\psi_k(r)$ are assumed to not interfere with each other, the total current density is just the sum of the current density contributions from each eigenstate of the metal-vacuum-metal system which is allowed by the Pauli exclusion principle.

The non-equilibrium aspects of a biased junction are usually treated by assigning different thermal occupational probability functions to each metal with the reference energies offset by the applied voltage and differences in Fermi energies. Thus, with $V^I(-\infty) + \mu_{xc}^I$ as the reference energy zero and a positive polarity voltage, V , connected to electrode II, the energies in the two electrodes would be related by the energy conservation equation:

$$E_I = E_{II} + E_{FI} - E_{FII} - eV. \quad (2-40)$$

For a net current to flow, occupied states on one side of the barrier must exist at the same energy as empty states on the other. In figure 7, the applied voltage will establish, at zero temperature, empty states in electrode II opposite filled states with the same energy in electrode I. Therefore, in this instance, electrons would dominantly flow from electrode I to electrode II by tunneling through the barrier, i.e., via a current-carrying stationary state which existed in all three regions of the system. The current density is computed as the difference between the current density flowing from I to II and II to I.

The thermal occupational probabilities are given by the Fermi-Dirac distribution

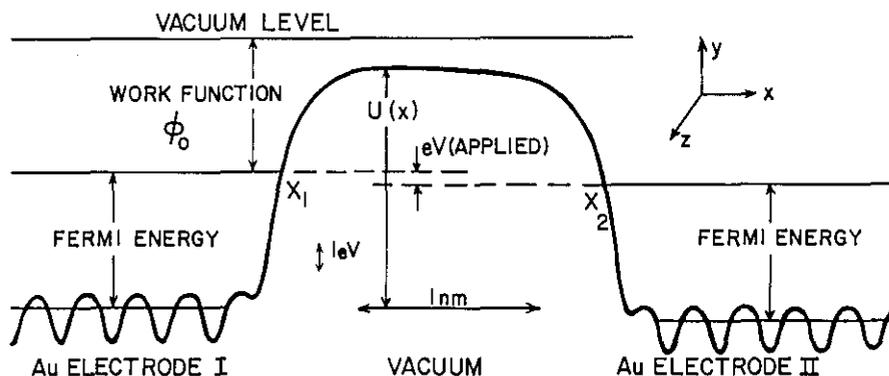


Figure 7—Energy Diagram for Metal-Vacuum-Metal Tunneling. Energy and distance in the diagram are drawn to scale for the quantities represented. The small oscillations of the energy at the bottom of the occupied states represent the magnitude and spacing of the ion-core potentials. X_1 and X_2 are the classical turning points for an electron near the Fermi energy.

$$f(E) = \left(1 + \exp\left[\frac{E - E_F}{k_B T}\right]\right)^{-1} \quad (2-41)$$

where E is the energy eigenvalue, k_B is Boltzmann's constant and T is the system temperature. Equation (2-40) combined with eq (2-41) gives the result

$$f_{II}(E) = f(E) \quad (2-42)$$

Again, taking the plane of the electrodes to be normal to the x -direction, the current density, $\mathbf{J}_{I \rightarrow II}$, from I to II is then given by

$$\mathbf{J}_{I \rightarrow II} = 2 \sum_{\mathbf{k}} \mathbf{J}_{I \rightarrow II}^x(\mathbf{k}, \mathbf{r}) f_I(E) [1 - f_{II}(E)]. \quad (2-43)$$

The factor of 2 is to account for the additional states produced by the two electron spin orientations. $\mathbf{J}_{I \rightarrow II}^x$ is the x component of $\mathbf{J}(\mathbf{k}, \mathbf{r})$. Similarly,

$$\mathbf{J}_{II \rightarrow I} = 2 \sum_{\mathbf{k}} \mathbf{J}_{II \rightarrow I}^x(\mathbf{k}, \mathbf{r}) f_{II}(E) [1 - f_I(E)]. \quad (2-44)$$

Since $V_T(x)$ is a real potential, the solutions to Schrödinger's equation must obey the general reciprocity theorem of scattering processes. Blatt and Weisskopf [41] give a succinct statement of the theorem: "This theorem states that the probability for a transition proceeding one way in time is equal to the probability for the same transition, but with the sense of time reversed." The sense of time in eqs (2-43 and 2-44) is manifest only in the direction of the wave vector, thus, $\mathbf{J}_{I \rightarrow II}^x(\mathbf{k}, \mathbf{r}) = \mathbf{J}_{I \rightarrow II}^x(-\mathbf{k}, \mathbf{r}) = \mathbf{J}_{II \rightarrow I}^x(\mathbf{k}, \mathbf{r})$. The expression for the net current density is therefore

$$\mathbf{J} = 2 \sum_{\mathbf{k}} \mathbf{J}^x(\mathbf{k}, \mathbf{r}) [f(E) - f(E + eV)]. \quad (2-45)$$

To calculate $J_x(\mathbf{k}, \mathbf{r}) = J^x(\mathbf{k}, \mathbf{r})$ (a subscript notation is adopted hereafter), explicit solutions to eq (2-38) for all three regions of $V_T(x)$ must be obtained. The solutions given here will be based on the results obtained by Merzbacher [42] with the use of the Wentzel, Kramers, Brillouin (WKB) approximation technique for solving Schrödinger's equation. If periodic boundary conditions are imposed in the y and z directions, the solutions to eq (2-38) may be written as

$$\psi_{\mathbf{k}}(\mathbf{r}) = A^{-1/2} \exp(i\mathbf{k}_p \cdot \mathbf{P}) M_{k_x}(x), \quad (2-46)$$

where \mathbf{k}_p and \mathbf{P} are the wave vector and displacement, respectively, parallel to the electrode faces, k_x is the wave vector component normal to the

electrode faces, and A is the area of the tunneling device; $M_{k_x}(x)$ is defined below. The eigenvalues $E(\mathbf{k})$ are

$$E(\mathbf{k}) = \frac{\hbar^2}{2m} (k_p^2 + k_x^2) = \frac{\hbar^2 k^2}{2m}. \quad (2-47)$$

Let $x_L(E)$ and $x_R(E)$ be the classical turning points for the energy E at electrode I and electrode II, respectively, then with Merzbacher's conventions

$$\begin{aligned} M_{k_x}(x) &= A k^{-1/2}(x) W_+ \\ &\quad + B k^{-1/2}(x) W_- \quad x \leq x_L(E) \\ M_{k_x}(x) &= C K^{-1/2}(x) Z_- \\ &\quad + D K^{-1/2}(x) Z_+ \quad x_L < x < x_R \\ M_{k_x}(x) &= F k^{-1/2}(x) q_+ \\ &\quad + G k^{-1/2}(x) q_- \quad x_R(E) \leq x, \end{aligned} \quad (2-48)$$

where

$$\begin{aligned} W_{\pm} &= \exp\left[\pm i \int_{x_L}^x k_x(x) dx\right] \\ Z_{\pm} &= \exp\left[\pm \int_{x_L}^x K(x) dx\right] \\ q_{\pm} &= \exp\left[\pm i \int_{x_R}^x k_x(x) dx\right] \end{aligned} \quad (2-49)$$

and

$$K(x) = \left[\frac{2m}{\hbar^2} \{V_B(x) - (E(k) - \frac{\hbar^2 k_p^2}{2m})\}\right]^{1/2}. \quad (2-50)$$

With these conventions, Bohm [43] then shows in detail that with the use of connection formulas

$$A = [(1/2)(2\Theta + 1/(2\Theta))]F \quad (2-51)$$

where

$$\Theta = \exp\left(\int_{x_L}^{x_R} K(x) dx\right). \quad (2-52)$$

If now the solution in region I is assumed to be

$$M_{k_x}^I(x) = e^{ik_x x} + R e^{-ik_x x} \quad (2-53)$$

with R being adjusted for normalization, then the solution in region II is, assuming $G=0$, i.e., no wave is incident from right,

$$M_{k_x}^{II}(x) = 2[2\Theta + 1/(2\Theta)]^{-1} e^{ik_x(x_L - x_R)} e^{ik_x x}. \quad (2-54)$$

The conservation of k_p and total energy in the tunneling process was implicitly assumed in the calcu-

lations from eqs (2-46 to 2-54). Since $M_k^{\text{II}}(x)$ is the stationary state solution in region II and probability current is "conserved," it may be used as well as the solution in the other two regions to calculate the current density. Substituting eq (2-54) into eq (2-46) and then the result into eq (2-39) gives

$$J_x(\mathbf{k}, \mathbf{r}) = 4[2\Theta + 1/(2\Theta)]^{-2} e \hbar k_x / m. \quad (2-55)$$

Let

$$D(E, \mathbf{k}_p) = 4[\Theta + 1/(2\Theta)]^{-2}, \quad (2-56)$$

then

$$J_x(\mathbf{k}, \mathbf{r}) = D(E, \mathbf{k}_p) e \hbar k_x / m. \quad (2-57)$$

$D(E, \mathbf{k}_p)$ may be interpreted as a transmission coefficient for the part of the solution in region I into region II.

Substituting eq (2-57) into eq (2-45) gives the complete expression for the current density J as

$$J = 2e \hbar / m \sum_{\mathbf{k}} k_x D(E, \mathbf{k}_p) [f(E) - f(E + eV)]. \quad (2-58)$$

The summations are most conveniently carried out by changing to integrations over the variable E_x defined as

$$E_x = \hbar^2 k_x^2 / 2m = E(k) - \hbar^2 k_p^2 / 2m. \quad (2-59)$$

The conversion is performed with the following three relations: the first is the well known sum to integral conversion in k -space, the second one is based on the free-electron properties expressed in eq (2-47) and the third is the differential of eq (2-59):

$$\begin{aligned} \sum_{\mathbf{k}} &\rightarrow \int d^3k / (2\pi)^3 \\ d^2k_p &= (2\pi m / \hbar^2) dE_p \\ k_x dk_x &= (m / \hbar^2) dE_x. \end{aligned} \quad (2-60)$$

With these relationships, eq (2-58) may be written in the following form:

$$J = \int_0^{\infty} dE_x D(E_x) N(E_x) \quad (2-61a)$$

$$N(E_x) = (4\pi m e / h^3) \int_0^{\infty} [f(E) - f(E + eV)] dE_p. \quad (2-61b)$$

In words this states that the net current density is equal to the sum of the number of electrons with the "x-part of their energy" within dE_x that are incident on the surface of the negatively biased electrode per unit time per unit area, $N(E_x)dE_x$, times the tunneling probability of an electron at this energy, $D(E_x)$.

This is the same form of the equations obtained by Duke [44] and as such is an acknowledgment that his work was frequently consulted in developing eqs (2-61). Another form of eq (2-61a) which is more useful for interpreting the energy aspects of the tunneling process is

$$J = \frac{2e}{\hbar} \int dE [f(E) - f(E + eV)] \int 2\pi^{-2} d^2k_p D(E, k_p). \quad (2-61c)$$

Equation (2-61c) facilitates the physical interpretation of the origins of the tunneling current in terms of the overlapping of energy surface projections onto the tunneling plane and enables direct connection with energy diagrams such as that given in figure 7. However, calculation of the tunneling current, particularly for the free-electron model of the electrodes, is most easily accomplished with the use of eq (2-61a).

Equation (2-61b) may be integrated by parts to obtain

$$N(E_x) = (4 m e k_B T / h^3) \ln (FDI / FDII),$$

$$FDI = 1 + \exp [(E_F - E_x) / k_B T]$$

$$FDII = 1 + \exp [E_F - E_x - eV] / k_B T. \quad (2-61d)$$

Note that the Fermi energies of the two electrodes have been assumed equal. Collecting all the definitions involved in eq (2-56) together for reference gives

$$\begin{aligned} D(E_x) &= 4[2\Theta + 1/(2\Theta)]^2, \\ \Theta &= \exp \left(\int_{x_L}^{x_R} K(x) dx \right), \\ K(x) &= [2m / \hbar^2 (V_B(x) - E_x)]^{1/2}, \\ V_B(x) &= E_F + WF - eVx / s \\ &\quad - (\text{const})s [x(s - x)]^{-1}, \end{aligned}$$

$$\begin{aligned} \text{const} &= 1.15 (\ln 2) e^2 / 8\pi \epsilon_0 \\ &= 0.2867 (\text{eV nm}). \end{aligned} \quad (2-61e)$$

The accuracy of the WKB approximation for calculating the transmission coefficient has been

extensively investigated in the literature of this field. One of the most thorough checks was performed by Politzer [45], who compared values computed with the form

$$D(E) = \exp\left[-2 \int_{x_L}^{x_R} K(x) dx\right], \quad (2-62)$$

of the WKB transmission coefficient with those obtained from a numerical solution of Schrödinger's equation using essentially the same V_T as given in eq (2-37). Her results show that the WKB result is valid at the fields and energies involved in most configurations of the vacuum tunneling experiment, i.e., fields less than 10^8 V/cm and energies much less than barrier height.

The major difference between the exact results and the WKB approximation was that the exact $D(E)$ was an oscillatory function of energy for energies and fields which resulted in an effective bound state potential well above the top of the barrier. The amplitude of the oscillations in the exact $D(E)$ was relatively small and such that

$$0.5 < \text{Log}(D(E)_{\text{WKB}}/D(E)_{\text{NUM}}) < 0.8 \quad (2-63)$$

for all fields and for energies more than 0.5 eV from the top of the potential barrier. Near the Fermi energy, at which the major contribution to the tunneling current occurs, the ratio's mean value was about -0.2 . The mean difference between the tunneling current for the two calculations of $D(E)$ was not calculated by Politzer, and barrier heights applicable to the vacuum tunneling experiment were not investigated.

Approximate Analytical Expressions for Tunneling Current

In this section, the qualitative properties of the tunneling current will be derived from the zero temperature, low voltage solutions to eqs (2-61 a to e). The overall character of the tunneling current will be shown to be approximately ohmic with a small cubic dependence on voltage having importance at higher voltages. At all voltages, the tunneling current is found to be exponentially dependent on the electrode spacing. Finally, the zero temperature solutions are extended to finite temperatures and estimates of the effects of room temperature on the tunneling current are given.

The tunneling characteristics determined by eqs (2-61) are most easily understood by first investigating the zero temperature behavior. At zero temperature

$$\begin{aligned} N(E_x) &= A_T eV, & 0 \leq E_x \leq E_F - eV \\ &= A_T (E_F - E_x), & E_F - eV < E_x < E_F \\ &= 0, & E_F \leq E_x \end{aligned} \quad (2-64)$$

A_T is the "thermal" constant $4\pi me/h^3$. Thus, eq (2-61a) takes the form

$$\begin{aligned} J &= A_T [eV \int_0^{E_F - eV} D(E_x) dE_x \\ &+ \int_{E_F - eV}^{E_F} (E_F - E_x) D(E_x) dE_x]. \end{aligned} \quad (2-65)$$

For large Θ in eq (2-61e), which for most barriers in tunneling experiments is a true condition, $D(E_x)$ can be approximated by the form given in eq (2-62). Simmons [46] has proposed an average barrier height method to further simplify $D(E_x)$. His approach is to approximate the integral in eq (2-61e) over $K(x)$ with an average barrier height, B , above the Fermi level of the negatively biased electrode, i.e., he defines B as

$$B = W^{-1} \int_{x_{LF}}^{x_{RF}} V_B(x) dx, \quad (2-66)$$

where x_{LF} and x_{RF} are the classical turning points for $E_x = E_F$, and $W = x_{RF} - x_{LF}$ is the width of the barrier at this energy.

With this definition, Simmons [46] shows that for voltages of less than about one volt

$$D(E_x) = \exp[-aW(E_F + B - E_x)^{1/2}], \quad (2-67)$$

$a = (4\pi/h)(2m)^{1/2} = 10.24(\text{eV})^{-1/2} \text{ nm}^{-1}$. By utilizing this $D(E_x)$, an approximate low voltage tunneling current density can be calculated from eq (2-65). Let $N = E_F - E_x$ and $p = aWB^{1/2}$, then eq (2-65) may be written as

$$\begin{aligned} J &= A_T \left\{ -eV \int_{E_F}^{eV} \exp[-p(1+n/B)^{1/2}] dN \right. \\ &\left. - \int_{eV}^0 N \exp[-p(1+N/B)^{1/2}] dN \right\}. \end{aligned} \quad (2-68)$$

For small voltage, N/B , will be much less than one, so the approximation $(1+N/B)^{1/2} = (1+N/2B)$ may be used. The integrals may then be evaluated to obtain the result:

$$\begin{aligned} J &= A_T \left[\left(4e^{-p}/(aW)^2 \right) \right. \\ &\left. \times \left(1 - \exp[-aWeV/(2B^{1/2})] \right) \right]. \end{aligned} \quad (2-69)$$

Only a factor of $\exp(-aWE_F/B^{1/2})$ was neglected to obtain eq (2-68). If one takes just the leading term in the series for the exponential, the following

result for the tunneling current density at low bias voltages and zero temperature is obtained:

$$J_L(T=0) = J_{0L} \exp(-p)V, \quad (2-70)$$

where $J_{0L} = [(2m)^{1/2}/W](e/h)^2 B^{1/2}$. If B is expressed in eV and W in nm, J_{0L} is

$$J_{0L} = 3.16 \times 10^{13} B^{1/2} W^{-1} (\text{A}/\text{m}^2/\text{V}). \quad (2-71)$$

By evaluating eq (2-68) in a more exact manner, Simmons [46] derives a general formula for the current density at zero temperature. The formula and its associated relations are given below in what has been found to be a consistent algebraic form, which due to misprints, etc., is not correct in published forms. It is given here because it will be used for comparing with the numerical results of the next section. With the same definitions given in eqs (2-66 and 2-67), the formula is

$$J(T=0) = J_0(B \exp(-aWB^{1/2}) - (B + eV) \exp[-aW(B + eV)^{1/2}]),$$

$$J_0 = e(2\pi h)^{-1} W^{-2}$$

$$= 6.2 \times 10^{12} W^{-2} (\text{A nm}^2)/(\text{m}^2 \cdot \text{eV}). \quad (2-72a)$$

Based on eq (2-66), the analytic approximation for B including image potential effect is

$$B = WF - (V/2S)(S_1 + S_2) - [0.286/(S_2 - S_1)] \times \ln[S_2(S - S_1)/(S_1(S - S_2))] \quad (2-72b)$$

S = electrode spacing in same sense as previously discussed.

$$S_1 = 0.3/WF$$

$$S_2 = S[1 - 2.3/(3WFS + 10 - 2VS)] + S_1.$$

In eq (2-72b), the work function should be expressed in electron volts and all distances involved in nanometers.

Equation (2-72a) reduces quite easily to eq (2-70) by expanding both exponentials in a series and neglecting terms containing V^2 and higher orders. However, if one keeps all terms out through V^3 , it is found that the quadratic terms cancel to give the next higher order approximate result [19]. For identical electrodes

$$J(T=0) = J_{0L}[V + (j_3 S^2/B)V^3], \quad (2-73a)$$

with

$$j_3 = (\pi^2 m e / h^2) = 3.27 \text{ V}^{-1} \text{ nm}^{-2}. \quad (2-73b)$$

This value of j_3 was taken from the results of Brinkman et al. [19] who expanded eqs (2-61) in powers of the voltage and obtained approximate expressions for the coefficients of the linear and quadratic terms of the conductance. They considered eq (2-73b) to be accurate to within 10% for trapezoidal shaped barriers and where S is greater than 1 nm. An earlier value obtained by Simmons [47] for the same conditions was $j_3 = 1.15 \text{ V}^{-1} \text{ nm}^{-2}$. No calculations of j_3 with image potentials corrections are available in the literature. In spite of the relatively unknown values for this constant, the qualitative behavior predicted by eq (2-73a) has been fairly well established in many different experiments [48,49,19].

In summary, the ideal tunneling characteristics are then as follows:

1) Current density varies exponentially with electrode spacing, with an exponent of $-10.25 \times$ mean barrier height in electron volts \times electrode spacing in nanometers.

2) The current-voltage characteristic should be symmetric for identical electrodes. (This just serves as a check on the formalism or the experimental method since it must be true if the electrodes are truly identical.)

3) For low voltages, the current-voltage characteristic is approximately ohmic.

4) For voltages up to the work function, the non-ohmic conduction should be dominantly a cubic function of the applied voltage, with higher order terms becoming more important as the voltage is increased up to the work function.

The extension of the previous results to finite temperatures is the last topic to be discussed in this section. If the approximations which led to eq (2-68) are substituted into eqs (2-61), the following result is obtained:

$$J(T) = A_T k_B T \exp(-aWB^{1/2}) \times \int_{E_F} \ln(FDI/FDII) \exp(-aWN/2B^{1/2}) dN, \quad (2-74)$$

where again $N = E_F - E_x$. When $k_B T$ is less than $2B^{1/2}/aW$, the integral may be evaluated by the techniques of Murphy and Good [50] to obtain

$$a \quad J(T) = A_T Q^{-2} [Q k_B T / \sin(\pi Q k_B T)] \times \exp(-p) [1 - \exp(-Q eV)], \quad (2-75)$$

where $Q = (1/2) a W B^{-1/2}$. For T approaching zero, the factor in the first bracket goes to one, and Simmons [46] shows that the remainder of the terms are equal to $J(T=0)$. Therefore,

$$J(T)/J(T=0) = Q k_B T / \sin(\pi Q k_B T). \quad (2-76)$$

Q is typically less than $4(eV)^{-1}$. The argument of the sine is therefore much less than one at room temperature. The series expansion gives

$$J(T)/J(T=0) = 1 + (1/6)(\pi Q k_B T)^2$$

or in practical units (2-77)

$$J(T) = J(T=0) \left(1 + [3 \times 10^{-11} (eV/nm^2/K^2) (WT)^2 / B] \right).$$

For $W = 1$ nm, $T = 300$ K, and $B = 2$ eV; $J(T) = 1.013 J(T=0)$. Such small changes in the tunneling current would not be detectable with the present experiment. No significant effect from conducting the experiment at room temperature was therefore anticipated.

Numerical Calculations of Tunneling Current

In order to facilitate the analysis of measured $I-V$ and $I_{cv}-S$ tunneling data, the following calculations, based on eqs (2-61) were performed. The results are given in accompanying figures and tables.

1) The product of $D(E_x)N(E_x)$ as a function of E_x for four voltages; 0.001, 0.010, 0.10, and 1.0 V, at each of three electrode spacings; $S = 0.3, 1.0,$ and 2.0 nm. See figures 8, 9, and 10.

2) For a simple model of the tunneling between spherical electrodes, described below, calculations of the tunneling current density and current were made for a fixed bias voltage of 0.1 V as a function of electrode spacing and work function. Computation times were long for the current calculations, 15 seconds per data point, so only a limited number were obtained. These results are given in table 1 and figures 11 through 17. The graphs in figures 11 and 12 were at first plotted in a normalized form, but after finding that much information was lost because of the large orders of magnitude changes between graphs, the four separate graphs were chosen as the more useful representation. Figure 11 shows the changes in curvature or lack thereof in the $I-V$ for a spacing of 2000 pm. Figure 12 shows the increase in curvature of the $I-V$ characteristic as a function of electrode spacing for a fixed work function of 4.5 eV. Figures 13-17 are graphs of the data in table 1 and enable a comparison of the cur-

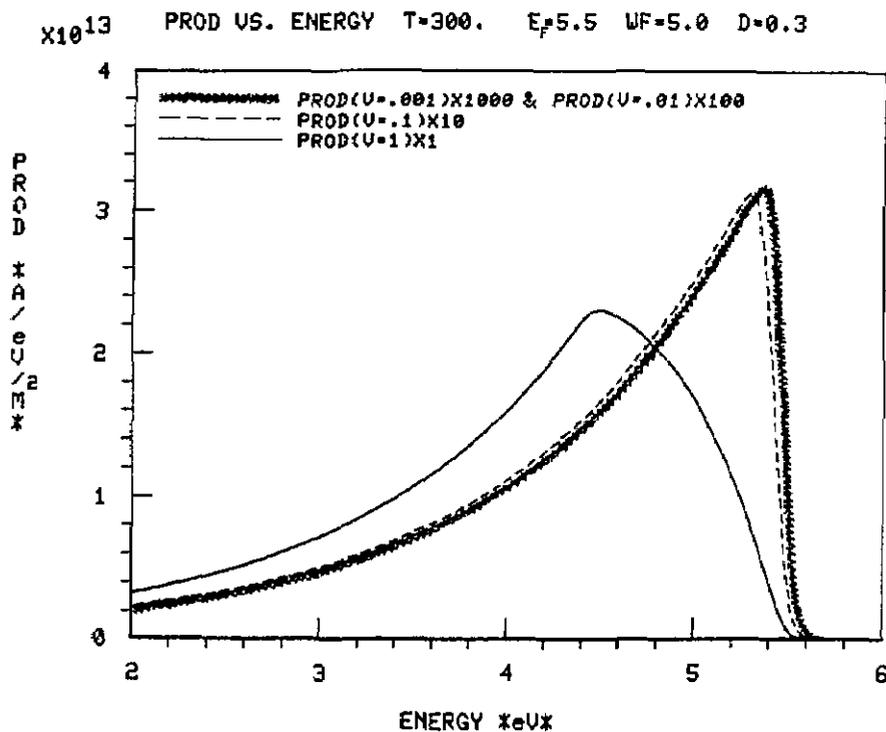


Figure 8-Distribution of energy associated with normal component of tunneling electron momentum; electrode spacing of 0.3 nm. The Fermi energy, E_F , and the work function, WF , are in units of electron volts.

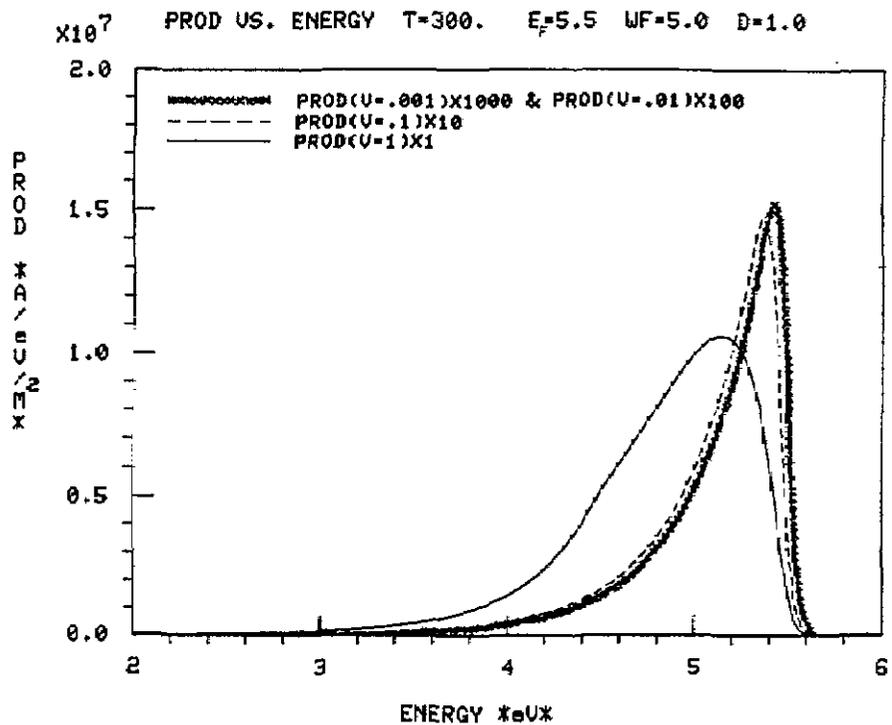


Figure 9—Normal-energy distribution of tunneling electrons; electrode spacing of 1.0 nm.

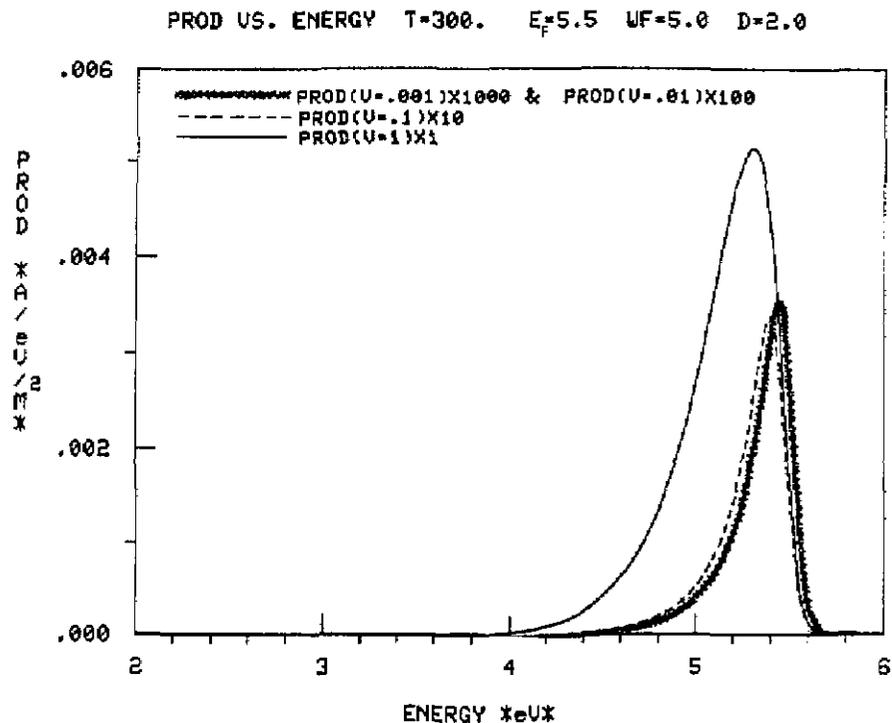


Figure 10—Normal-energy distribution of tunneling electrons; electrode spacing of 2.0 nm.

rent and current-density so that the tunneling area may be estimated.

3) The tunneling current density for a wide range of voltage and electrode spacing; $V=N \times 0.05$ volts, $N=1$ to 12 and $S=(0.3+M \times 0.1)\text{nm}$, $M=0$ to 27. See table 2.

4) As a cross-check on the correctness of the numerical calculations and to determine the correct form of Simmons' equations, eqs (2-72 and 2-77) were used to compute the tunneling current density as a function of electrode spacing and bias voltage. The results are given in table 3 and

Table 1. Tunneling Between Identical Spherical Electrodes; Current and Current Density Versus Electrode Spacing and Work Function.

Bias Voltage = 0.1 Volt Temperature = 300 Degrees Kelvin Fermi Energy = 5.5 Electron Volts								
Electrode Spacing (nm)	—WF=4.0 eV—		—WF=4.5 eV—		—WF=5.0 eV—		—WF=5.5 eV—	
	Current (A)	Current Density (A/m ²)	Current (A)	Current Density (A/m ²)	Current (A)	Current Density (A/m ²)	Current (A)	Current Density (A/m ²)
.3	.256+01	.924+13	.155+01	.627+13	.955+00	.424+13	.601+00	.287+13
.4	.420+00	.168+13	.210+00	.918+12	.110+00	.515+12	.601-01	.297+12
.5	.610-01	.253+12	.262-01	.117+12	.120-01	.567+11	.576-02	.287+11
.6	.852-02	.357+11	.319-02	.143+11	.128-02	.609+10	.546-03	.273+10
.7	.118-02	.495+10	.384-03	.173+10	.136-03	.648+09	.514-04	.258+09
.8	.161-03	.680+09	.460-04	.207+09	.144-04	.685+08	.481-05	.242+08
.9	.218-04	.926+08	.547-05	.247+08	.151-05	.721+07	.449-06	.226+07
1.0	.295-05	.126+08	.649-06	.294+07	.158-06	.756+06	.418-07	.211+06
1.1	.398-06	.169+07	.766-07	.347+06	.165-07	.790+05	.388-08	.196+05
1.2	.534-07	.228+06	.902-08	.410+05	.171-08	.823+04	.360-09	.181+04
1.3	.716-08	.306+05	.106-08	.482+04	.178-09	.856+03	.333-10	.168+03
1.4	.956-09	.409+04	.124-09	.566+03	.185-10	.889+02	.307-11	.155+02
1.5	.127-09	.545+03	.146-10	.663+02	.191-11	.921+01	.283-12	.143+01
1.6	.170-10	.727+02	.170-11	.775+01	.198-12	.954+00	.261-13	.132+00
1.7	.226-11	.966+01	.199-12	.906+00	.205-13	.986-01	.240-14	.122-01
1.8	.299-12	.128+01	.232-13	.106+00	.211-14	.102-01	.221-15	.112-02
1.9	.397-13	.170+00	.270-14	.123-01	.218-15	.105-02	.203-16	.103-03
2.0	.526-14	.226-01	.314-15	.143-02	.225-16	.108-03	.187-17	.946-05

demonstrate that his theory gives values which are surprisingly close to those of the numerical calculations³.

The product $D(E_x)N(E_x)$ is known as the normal-energy distribution [51], and is the distribution of tunneling electrons in energy associated with the component of momentum normal to the electrode faces. The normal-energy distributions were calculated and plotted to give insight into the thermal distributions and barrier transmission aspects of the tunneling process. Interpretation of figures 8, 9, and 10 in terms of the Pauli exclusion principle and the energy diagram of figure 7 is difficult. For a total energy distribution, the graphs should have a sharp peak at the Fermi level of the negatively biased electrode and have a width equal to the bias voltage with some small amount of thermal smearing at both edges. Plots of the total energy distribution for tunneling in Al-GaSe-Au structures are

given by Kurtin et al. [23]. The utility of the normal-energy distribution may be seen by recalling that $N(E_x)$ for zero temperature is given by eq (2-64); $N(E_x)$ at room temperature is only slightly different with small spreading at $E_x = E_F$ and rounding at $E_x = E_F - eV$. Combining this truncated triangular shape with the graphs of figures 8, 9, and 10 enables one to readily see the exponential changes in the distribution due to changes in the tunneling probability as $E_F - E_x$ and $(E_F - eV) - E_x$ are changed. Other observations from figures 8, 9, and 10 are: 1) that the sources of non-linearity of the distributions for different voltages are apparent in the increasingly unequal heights as the electrode spacing increases, and 2) that the thermal smearing of about 25 mV at room temperature results in the distributions for 1 and 10 mV being almost identical.

The other figures are reasonably self-explanatory and, while no further discussion of them will be given here, they will be referred to in the analysis of experimental data.

The computer programs used to obtain the data described in items 1 through 4 are given in Appendix I, along with annotations which explain the major mathematical difficulties encountered while

³The negative current density values for an electrode spacing of 0.3 nm result from the approximations used in Simmons' theory. In his theory the current density is expressed as the difference between two terms (see the equations in the computer program listing in Appendix I) whose magnitudes become approximately equal at electrode spacings of 0.3 nm and work functions of 5 eV.

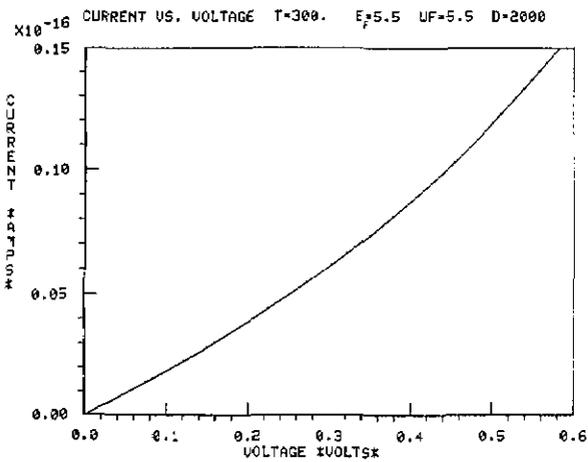


Figure 11a—Theoretical current-voltage characteristic; E_F (shown as FE) and WF are in eV; D is in pm.

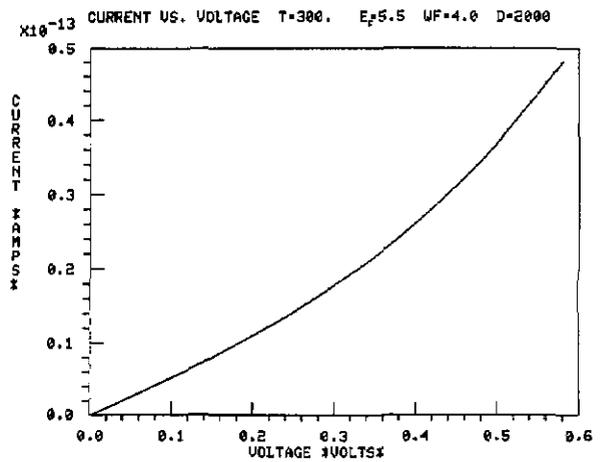


Figure 11d—Theoretical current-voltage characteristic; E_F (shown as FE) and WF are in eV; D is in pm. Repeat of figure 11a with WF of 4.0 eV.

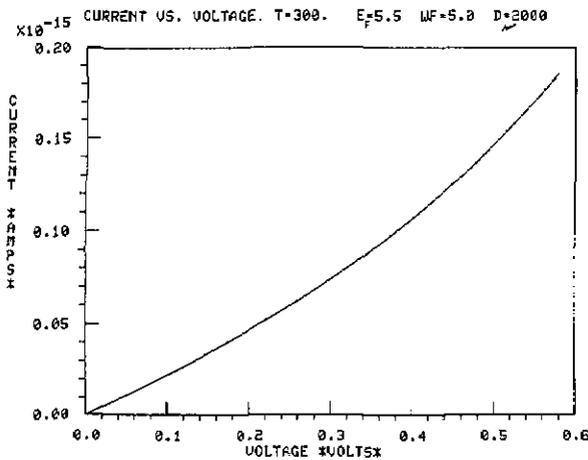


Figure 11b—Theoretical current-voltage characteristic; E_F (shown as FE) and WF are in eV; D is in pm. Repeat of figure 11a with WF of 5.0 eV.

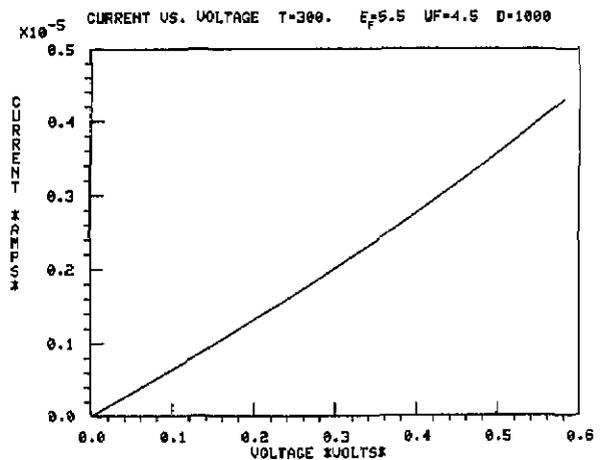


Figure 12a—Theoretical current-voltage characteristic; E_F (shown as FE) and WF are in eV; D is in pm.

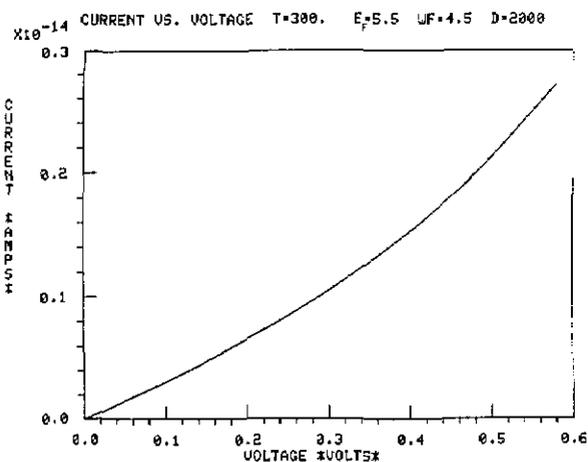


Figure 11c—Theoretical current-voltage characteristic; E_F (shown as FE) and WF are in eV; D is in pm. Repeat of figure 11a with WF of 4.5 eV.

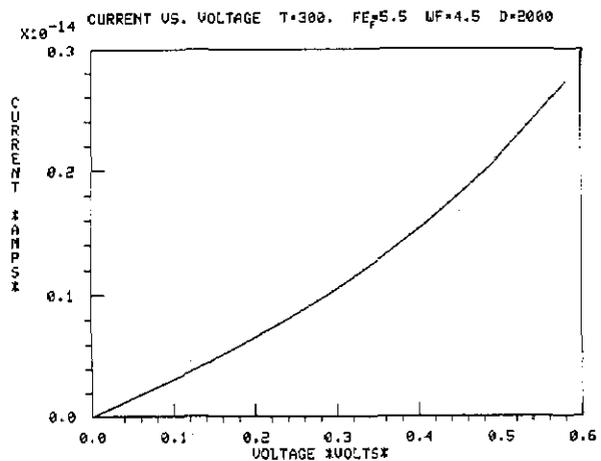


Figure 12b—Theoretical current-voltage characteristic. Repeat of figure 12a with an electrode spacing of 2000 pm.

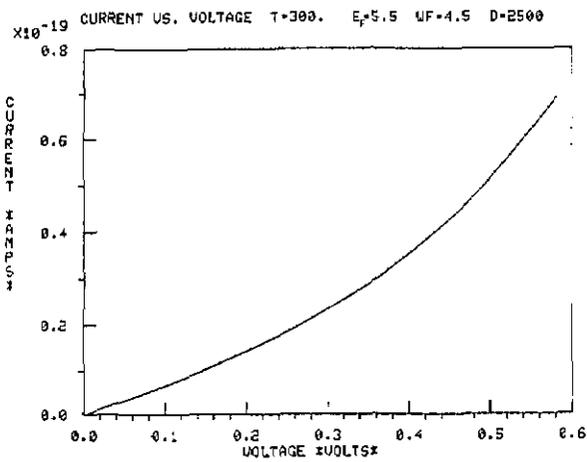


Figure 12c-Theoretical current-voltage characteristic. Repeat of figure 12a with an electrode spacing of 2500 pm.

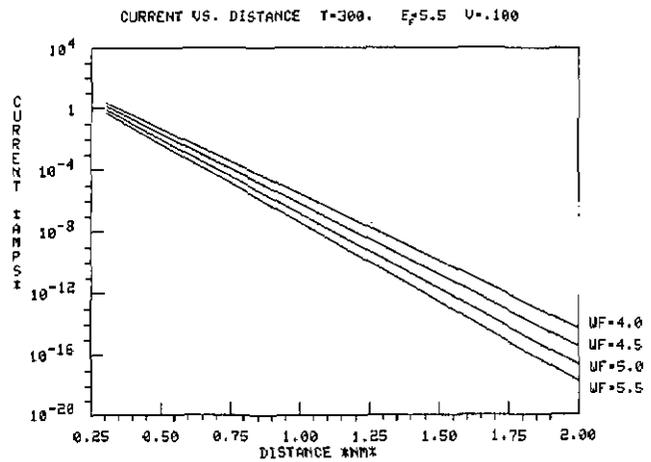


Figure 14-Theoretical log current-spacing characteristics for four work functions; bias voltage of 0.1 volt. Work functions are in units of eV.

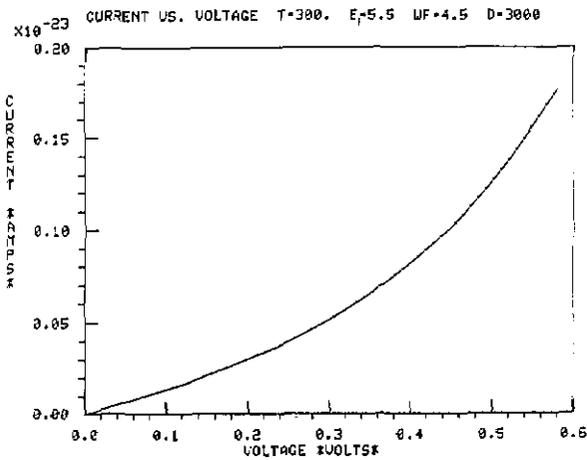


Figure 12d-Theoretical current-voltage characteristic. Repeat of figure 12a with an electrode spacing of 3000 pm.

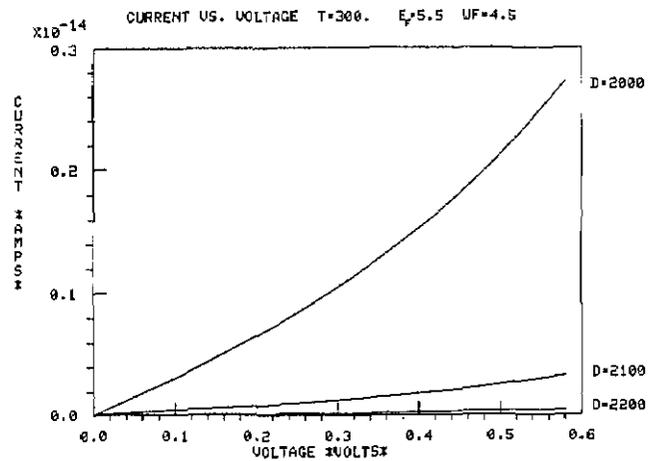


Figure 15-Theoretical current-voltage characteristics; work function of 4.5 eV, electrode spacings of 2.0, 2.1, and 2.2 nm.

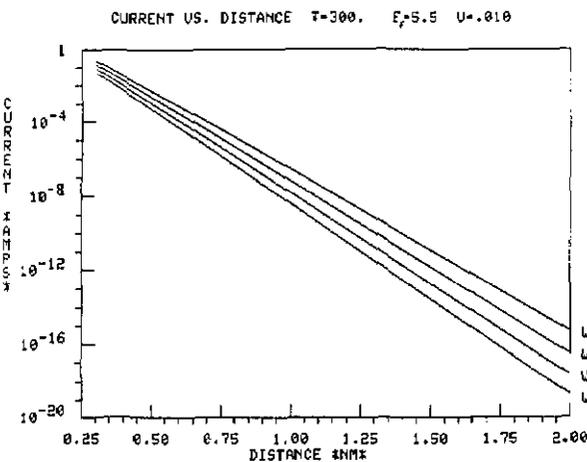


Figure 13-Theoretical log current-spacing characteristics for four work functions; bias voltage of 0.01 volt. Work functions are in units of eV.

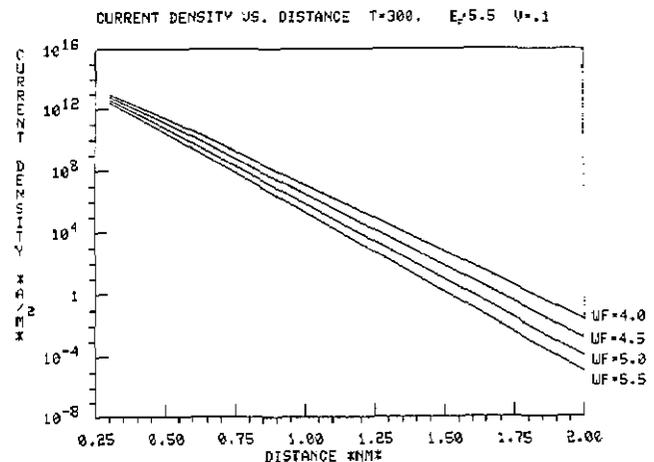


Figure 16-Theoretical log current density-electrode spacing characteristics; work functions of 4.0, 4.5, 5.0, and 5.5 eV.

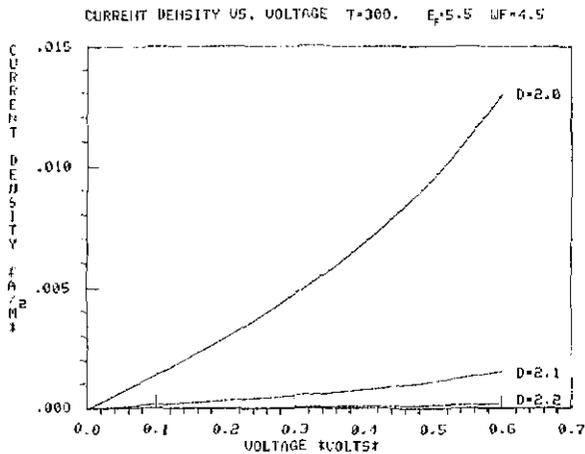


Figure 17—Theoretical current density-voltage characteristics; work function of 4.5 eV and electrode spacings of 2.0, 2.1, and 2.2 nm.

bringing the programs up to a correctly operating state. There are, however, several additional points about the physics involved which should be described here. To calculate all the constants involved, the following values of the fundamental constants were used:

$$m = \text{mass of the electron} = 9.11 \times 10^{-31} \text{ kg}$$

$$e = \text{charge on the electron} = 1.60 \times 10^{-19} \text{ C}$$

$$k_B = \text{Boltzmann's constant} = 1.38 \times 10^{-23} \text{ J/K} = 86.2 \text{ } \mu\text{eV/K}$$

$$h = \text{Planck's constant} = 6.625 \times 10^{-34} \text{ J s.}$$

With these values,

Table 2. Tunneling Between Identical Electrodes: Current Density Versus Electrode Spacing and Voltage (Current Density in Units of A/m²).

Electrode Spacing (nm)	Voltage (Volts)											
	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60
Temperature = 300 Degrees Kelvin												
Work Function = 5.0 Electron Volts												
Fermi Energy = 5.5 Electron Volts												
.3	.212+13	.424+13	.636+13	.849+13	.106+14	.128+14	.149+14	.171+14	.192+14	.214+14	.236+14	.258+14
.4	.257+12	.515+12	.773+12	.103+13	.129+13	.155+13	.182+13	.208+13	.235+13	.262+13	.289+13	.317+13
.5	.283+11	.567+11	.852+11	.114+12	.143+12	.172+12	.201+12	.231+12	.260+12	.291+12	.321+12	.353+12
.6	.304+10	.609+10	.916+10	.122+11	.154+11	.185+11	.217+11	.249+11	.282+11	.315+11	.349+11	.383+11
.7	.324+09	.648+09	.975+09	.130+10	.164+10	.197+10	.232+10	.266+10	.302+10	.338+10	.375+10	.413+10
.8	.342+08	.685+08	.103+09	.138+09	.173+09	.209+09	.246+09	.283+09	.322+09	.361+09	.402+09	.444+09
.9	.360+07	.721+07	.109+08	.145+08	.183+08	.221+08	.260+08	.300+08	.342+08	.384+08	.428+08	.474+08
1.0	.377+06	.756+06	.114+07	.153+07	.192+07	.233+07	.274+07	.317+07	.362+07	.408+07	.456+07	.506+07
1.1	.394+05	.790+05	.119+06	.160+06	.201+06	.244+06	.288+06	.334+06	.382+06	.432+06	.484+06	.539+06
1.2	.410+04	.823+04	.124+05	.167+05	.211+05	.256+05	.303+05	.351+05	.403+05	.456+05	.513+05	.573+05
1.3	.426+03	.856+03	.129+04	.174+04	.220+04	.267+04	.317+04	.369+04	.424+04	.482+04	.544+04	.610+04
1.4	.442+02	.889+02	.134+03	.181+03	.229+03	.279+03	.332+03	.387+03	.446+03	.508+03	.576+03	.648+03
1.5	.458+01	.921+01	.139+02	.188+02	.238+02	.291+02	.347+02	.406+02	.469+02	.536+02	.609+02	.688+02
1.6	.474+00	.954+00	.144+01	.195+01	.248+01	.303+01	.362+01	.425+01	.492+01	.565+01	.644+01	.731+01
1.7	.490-01	.986-01	.149+00	.202+00	.257+00	.315+00	.378+00	.444+00	.517+00	.596+00	.682+00	.777+00
1.8	.506-02	.102-01	.155-01	.209-01	.267-01	.328-01	.394-01	.465-01	.543-01	.628-01	.721-01	.825-01
1.9	.522-03	.105-02	.160-02	.217-02	.277-02	.341-02	.411-02	.486-02	.570-02	.661-02	.763-02	.877-02
2.0	.538-04	.108-03	.165-03	.224-03	.287-03	.354-03	.428-03	.509-03	.598-03	.697-03	.808-03	.933-03
2.1	.554-05	.112-04	.170-04	.232-04	.297-04	.368-04	.446-04	.532-04	.628-04	.735-04	.856-04	.993-04
2.2	.570-06	.115-05	.175-05	.239-05	.308-05	.382-05	.465-05	.556-05	.659-05	.775-05	.906-05	.106-04
2.3	.586-07	.118-06	.181-06	.247-06	.319-06	.397-06	.484-06	.582-06	.692-06	.817-06	.960-06	.112-05
2.4	.603-08	.122-07	.186-07	.255-07	.330-07	.412-07	.504-07	.608-07	.726-07	.862-07	.102-06	.120-06
2.5	.619-09	.125-08	.192-08	.263-08	.341-08	.428-08	.525-08	.636-08	.763-08	.909-08	.108-07	.128-07
2.6	.636-10	.129-09	.198-09	.272-09	.353-09	.444-09	.547-09	.665-09	.802-09	.960-09	.115-08	.136-08
2.7	.654-11	.133-10	.204-10	.281-10	.366-10	.461-10	.570-10	.696-10	.842-10	.101-09	.122-09	.145-09
2.8	.671-12	.136-11	.210-11	.289-11	.378-11	.479-11	.594-11	.728-11	.886-11	.107-10	.129-10	.155-10
2.9	.689-13	.140-12	.216-12	.299-12	.391-12	.497-12	.619-12	.762-12	.931-12	.113-11	.137-11	.166-11
3.0	.707-14	.144-13	.222-13	.308-13	.405-13	.516-13	.646-13	.798-13	.980-13	.120-12	.146-12	.177-12

Journal of Research of the National Bureau of Standards

Table 3. Tunneling Between Identical Electrodes; Current Density Versus Electrode Spacing and Voltage. Calculated from Simmons Theory (Corrected). (Current Density in Units of A/m²).

Temperature = 300 Degrees Kelvin Work Function = 5.0 Electron Volts										
Electrode Spacing (nm)	Voltage (Volts)									
	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
.3	-.704+12	-.142+13	-.216+13	-.292+13	-.371+13	-.452+13	-.538+13	-.627+13	-.721+13	-.821+13
.4	.101+12	.204+12	.308+12	.413+12	.520+12	.628+12	.738+12	.849+12	.962+12	.108+13
.5	.182+11	.367+11	.556+11	.749+11	.946+11	.115+12	.135+12	.156+12	.178+12	.200+12
.6	.236+10	.478+10	.725+10	.978+10	.124+11	.150+11	.178+11	.206+11	.235+11	.265+11
.7	.279+09	.564+09	.856+09	.116+10	.146+10	.178+10	.211+10	.245+10	.280+10	.316+10
.8	.314+08	.636+08	.966+08	.131+09	.166+09	.202+09	.239+09	.278+09	.319+09	.361+09
.9	.345+07	.698+07	.106+08	.144+08	.183+08	.223+08	.265+08	.308+08	.354+08	.401+08
1.0	.372+06	.755+06	.115+07	.156+07	.198+07	.242+07	.288+07	.336+07	.386+07	.439+07
1.1	.398+05	.807+05	.123+06	.167+06	.212+06	.260+06	.310+06	.362+06	.417+06	.476+06
1.2	.421+04	.855+04	.130+05	.177+05	.226+05	.277+05	.331+05	.387+05	.447+05	.511+05
1.3	.444+03	.901+03	.138+04	.187+04	.239+04	.293+04	.351+04	.412+04	.477+04	.547+04
1.4	.465+02	.946+02	.144+03	.197+03	.251+03	.309+03	.371+03	.437+03	.507+03	.583+03
1.5	.486+01	.988+01	.151+02	.206+02	.264+02	.325+02	.391+02	.461+02	.537+02	.619+02
1.6	.506+00	.103+01	.158+01	.215+01	.276+01	.341+01	.411+01	.486+01	.568+01	.656+01
1.7	.526-01	.107+00	.164+00	.224+00	.288-00	.357+00	.431+00	.511+00	.599+00	.695+00
1.8	.545-02	.111-01	.170-01	.233-01	.300-01	.373-01	.451-01	.537-01	.631-01	.735-01
1.9	.565-03	.115-02	.177-02	.242-02	.313-02	.389-02	.472-02	.563-02	.664-02	.776-02
2.0	.584-04	.119-03	.183-03	.251-03	.325-03	.405-03	.493-03	.590-03	.698-03	.819-03
2.1	.603-05	.123-04	.189-04	.260-04	.337-04	.422-04	.515-04	.618-04	.734-04	.865-04
2.2	.621-06	.127-05	.196-05	.270-05	.350-05	.439-05	.537-05	.647-05	.771-05	.912-05
2.3	.640-07	.131-06	.202-06	.279-06	.363-06	.456-06	.560-06	.677-06	.810-06	.962-06
2.4	.659-08	.135-07	.208-07	.288-07	.376-07	.474-07	.584-07	.708-07	.851-07	.101-06
2.5	.678-09	.139-08	.215-08	.298-08	.389-08	.492-08	.608-08	.741-08	.894-08	.107-07
2.6	.697-10	.143-09	.221-09	.307-09	.403-09	.511-09	.634-09	.775-09	.939-09	.113-08
2.7	.716-11	.147-10	.228-10	.317-10	.417-10	.530-10	.660-10	.810-10	.986-10	.119-09
2.8	.735-12	.151-11	.235-11	.327-11	.431-11	.550-11	.687-11	.847-11	.104-10	.126-10
2.9	.755--13	.155-12	.242-12	.338-12	.446-12	.571-12	.716-12	.886-12	.109-11	.133-11
3.0	.774-14	.159-13	.249-13	.348-13	.461-13	.592-13	.746-13	.927-13	.114--12	.140-12

$$N(E_x) = 1.3585 \times 10^{10} T \ln [] \text{ A/eV K}^2 \quad (2-78)$$

The values of $a = 10.25(\text{eV})^{-1/2} (\text{nm})^{-1}$ and $1.15e^2 \ln 2 / (8\pi\epsilon_0) = 0.2867 (\text{eV nm})$ were also obtained with these values for the fundamental constants.

Due to the nature of $N(E_x)$, which was explained earlier, eq (2-61a) was decomposed into three parts:

$$J_1 = A_7 k_B T \int_0^{E_F - c - eV} \ln(FDI/FDII) dE_x \quad (2-79a)$$

$$J_2 = A_7 k_B T \int_{E_F - c - eV}^{E_F + WF - \text{ETOP}} \ln(FDI/FDII) dE_x \quad (2-79b)$$

$$J_3 = A_7 k_B T \int_{E_F + WF - \text{ETOP}}^{E_F + 80k_B T} \ln(FDI/FDII) dE_x \quad (2-79c)$$

with $J = J_1 + J_2 + J_3$. c was chosen as $40k_B T$ and ETOP is the top of the barrier potential function. Neglecting small changes in the maximum barrier height with applied bias voltages, ETOP is given by

$$\text{ETOP} = [0.575 e^2 \ln 2 / (8\pi\epsilon_0)] S^{-1} \quad (2-80)$$

J_3 was not included in the calculations unless ETOP was less than $(E_F + 80k_B T)$.

The form for $D(E_x)$ used in the calculations was

$$D(E_x) = (1 + \exp[2 \int_{x_L}^{x_R} K(x) dx])^{-1} \quad (2-81)$$

$K(x)$ is defined in eq (2-61e). This form may be approximately obtained from eq (2-61e) by factoring $4\Theta^2$ from the second term and performing a series expansion. The result is $D(E_x) = (1/2 + \Theta^2)^{-1}$.

Equation (2-81) has been obtained by Miller and Good [52] by analytic continuation and was shown to hold for all energies, including energies greater than the barrier height; one determines complex turning points in this instance. The Miller and Good form of the WKB transmission enabled the calculations to proceed smoothly up to the top of the barrier, where $x_L = x_R$ and $D(E_x) = 0.5$. The value for the exact transmission coefficient of a particle with energy at the top of a rectangular shaped barrier is 0.2. Since one would expect the image potential rounded barrier to be less reflective, the 0.5 value is at least intuitively correct. To avoid the complexities of determining complex turning points, a simple linear interpolation between 0.5 and 1 was used from ETOP to the vacuum level.

Current Calculations for Spherical Electrodes⁴

Current flow was obtained from the current density calculations by assuming that in the almost planar electrode structure, the current flow was parallel to the axis of the two spheres. The incremental tunneling areas are then rings around this axis of radius r with spacing greater than the closest spacing between the electrodes. The current flow in this approximation is

$$I = \int_{x_0}^{x_m} J(x) dA, \quad (2-82)$$

$J(x)$ is the current density for an electrode spacing x , x_0 is the on-axis electrode spacing and x_m is the spacing between points off-axis at which $J(x)$ has dropped to a negligible value. For r much smaller than the sphere radius R ,

$$r(x) = [R(x - x_0)]^{1/2}. \quad (2-83)$$

With $dA = 2\pi r dr$, eq (2-82) is therefore

$$I = \pi R \int_{x_0}^{x_m} J(x) dx. \quad (2-84)$$

The 10^{-12} factor employed in the programs in Appendix I results from the use of x in nm and R in mm.

3. Experimental Apparatus and Method

The rule-of-thumb expression for the tunneling current I , at small voltage, V

$$I = C \exp(-aSB^{1/2})AV, \quad (3-1)$$

and its ramifications are the major factors which determine all aspects of the apparatus design, construction, and use of apparatus for metal-vacuum-metal tunneling. C is a constant, A is the device tunneling area, a is $10.25 \text{ (eV)}^{-1/2} \text{ (nm)}^{-1}$, S is the electrode spacing, and B is the mean barrier height. Some insight into the experimental problems which had to be addressed may be obtained by observing that according to eq (3-1) a change in electrode spacing of 0.1 nm, about one-half an atom diameter, will generate a factor of 10 change in the tunneling current, assuming a barrier height of 5 eV. In a metal-vacuum-metal tunneling device, an electrode spacing stability of one pm is therefore required to achieve a noise level of 2% or less in the tunneling current for a fixed voltage. This corresponds to a signal-to-noise ratio of only 34 dB.

Mechanical and Thermal Requirements

The electrode spacing stability is in turn determined by the mechanical and thermal properties of the electrodes and the assembly into which they are placed. As representative values, let the effective length, L , for thermally generated changes in electrode spacing be 5 mm and the expansion coefficient, k_T , of the materials involved be 10 ppm/K. Then, the required temperature stability, dT , to achieve a 1 pm stability, dL , in electrode spacing is

$$dT = dL(Lk_T)^{-1} = 20 \mu\text{K} \quad (3-2)$$

The laboratory in which the experiment was to be performed had a temperature stability of ± 0.5 K over a 24-hour period. However, when local sources of heat such as that of the body heat of the experimenter are taken into account, the local vari-

⁴The high lateral resolution implied by the imaging of the Si (111) surface obtained by G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel (Phys. Rev. Lett. 50 (1983) 120) has stimulated much theoretical investigation of the effect of electrode shape on the tunneling current and effective tunneling area. The strong exponential dependence of the tunneling current on electrode spacing tends to confine the tunneling current to very small regions of one of the electrodes if it is sharpened. Some of the investigations of the shape of the electrodes on the tunneling current and tunneling area are: N. M. Miskovsky, S. J. Shepherd, P. H. Cutler, A. A. Lucas, and T. E. Sullivan, Effect of Geometry and Multiple Image Interactions on Tunneling and I-V Characteristics of MVM Point-Contact Junctions, App. Phys. Lett. 37 (1980) 189; J. Tersoff and D. R. Hamann, Theory and Application for the Scanning Tunneling Microscope, Phys. Rev. Lett. 50 (1983) 1998; N. Garcia, C. Ocal, and F. Flores, Model Theory from Scanning Tunneling Microscopy: Application to Au (110)(1X2), Phys. Rev. Lett. 50 (1983) 2002; T. E. Feuchtwang, P. H. Cutler, N. M. Miskovsky, Theory of Vacuum Tunneling Microscopy, Phys. Lett. 99A (1983) 67; A. Baratoff, Theory of Scanning Tunneling Microscopy - Methods and Approximations, Physica 127B (1984) 143.

ations in environmental temperature are more represented by ± 1 K. An approximate attenuation of 10^{-5} of the environmental temperature variations must, therefore, be provided by a temperature control system.

Requirements for mechanical stability were determined by measuring the amplitude and frequency composition of building vibrations at the planned laboratory location of the experimental apparatus. The transducer used for these measurements was a Texas Instruments S-36 very low frequency seismic detector whose sensitivity was 1.97 V/cm/s. The transducer's response was approximately 10% greater at its resonance frequency of 2 Hz and dropped by about 15% at 1 Hz. Figure 18 shows the velocity amplitude spectrum of the building vibrations in the National Bureau of Standards Reverberation Chamber. Measurements were performed here because it was one of the quietest potential locations for conducting the vacuum tunneling experiment. In addition, it was the only laboratory which allowed convenient access to the Fourier transform electronics necessary for obtaining real-time vibration spectra. Vibration spectra from this location were compared with other possible locations for the experiment with the use of oscillographs. These oscillographs showed that the vibrations present on a vibration table with a resonant frequency of 1 Hz and located in a basement laboratory were similar in amplitude and frequency content to those measured in the well-isolated NBS Reverberation Chamber. The NBS Reverberation Chamber is a room-within-a-room structure with an inside room volume of about 400 m³. The inner room is supported on springs to produce a room-spring resonance frequency of 3.5 Hz.

Depending on the frequency, the vibration amplitudes which must be accounted for, even with a

vibration isolation table, are therefore between 4 and 25 nm. The amplitudes are deduced from figure 8 with the 25 nm amplitude corresponding to frequencies of about 3.5 Hz, and the 4 nm to a frequency of about 35 Hz. Figure 18 only shows the vertical component of vibration displacements; measured horizontal components were similar in amplitude, but with major contributions existing only at frequencies near 35 Hz and below 10 Hz. The mechanical design of the electrode holder must, therefore, attenuate these vibrations by about 10^{-4} to produce the 1 pm stability.

A second factor which must be accounted for in designing for mechanical stability is acoustical pressure variations in the air around the experimental apparatus. No measurements of the magnitude and frequency content of the acoustical pressure variations were made since reliable estimates could be obtained from the literature. Hunter [53] gives a graph of the noise amplitude as a function of frequency for an average room. For a mean noise level of 40 dB referenced to 2×10^{-5} N/m², the amplitude at 20 Hz was 40 dB and dropped almost linearly to -5 dB at 5000 Hz. Within the laboratory used for this experiment, there were no distinctive noise sources, so this description of the acoustical environment was taken as representative.

Choosing an Electrode Material and Preparing the Electrodes

Variations in the mean barrier height also produce large changes in the tunneling current. For small changes in B , dB, eq (3-1) may be used to obtain

$$dI/I = -1/2 a S B^{1/2} \text{ dB.} \quad (3-3)$$

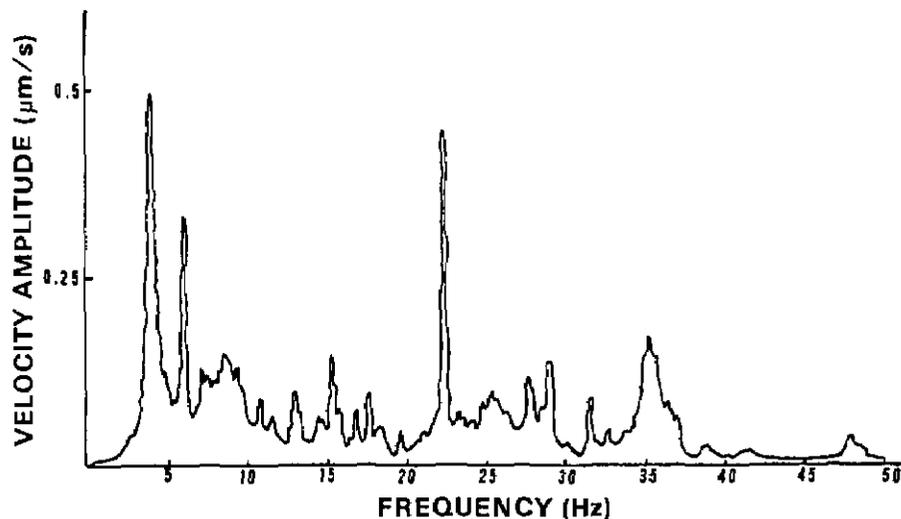


Figure 18—Measured velocity amplitude spectrum of building vibrations in the National Bureau of Standards Reverberation Chamber.

With $S=1$ nm, $B=4$ eV, and $\text{dB}=0.1$ eV, $dI/I = -0.26$, B must be stable to 0.008 eV to achieve a 34 dB signal to noise ratio.

Changes in work function and thus barrier height are produced by the adsorption and chemisorption of any kind of foreign atom on a metal surface [27,32]. The change in the work function WF , WFC , is a local microscopic effect, but in terms of a spatially averaged change, it can formally be related to an effective dipole moment p per adsorbed particle via the equation [32,54]

$$WFC = 4pN, \quad (3-4)$$

where N is the number of adsorbed atoms or molecules per unit area. Adsorption is viewed as changing the $[V(+\infty) - V(-\infty)]$ term of the WF described in section 2. Changes in WF range from $+0.08$ eV for a 1/3 monolayer of selenium on nickel to -3.1 eV for a monolayer of cesium on tungsten (100) [32,54]. Larger increases in the WF are also found, but usually no greater than 1 eV. WFC 's due to more common adsorbates such as oxygen, nitrogen, and hydrogen are typically about 1 eV, and may be of either sign depending on the substrate-adsorbate combination.

This demand for stable clean electrode surfaces must be carefully balanced with the opposing need for high temperature stability. There are two major reasons for the conflict between these two requirements. First, to achieve atomically clean metal surfaces, such that the work function would not be changed by any contaminants, some form of in-situ vacuum cleaning process must be employed. The most common and reliable methods for in-situ cleaning all require large energy inputs either to simply evaporate or decompose all surface and near surface contaminants, e.g., high temperature flashing of W or high temperature cycling in oxidizing and reducing atmospheres, e.g., Ni in O_2 and H_2 . Second, the time for a contaminating layer to form on a clean surface is comparable with estimated times for an electrode assembly to thermally equilibrate to the stabilities demanded by the vacuum tunneling experiment.

The time, T , in which an initially clean surface will become covered with a monolayer of adsorbate when the surface is exposed to a gas at pressure P Torr is (see Appendix III)

$$T \cong 2 \times 10^6 / P \text{ seconds.} \quad (3-5)$$

Equation (3-5) is only a rough estimate, but it is useful in determining the vacuum conditions necessary for this particular application. A minimum

time for a system that has been raised to 1300 K to obtain a $10 \mu\text{K}$ stability is about 24 hours or 8.6×10^5 seconds. Thus, maintaining a clean surface over this period of time would require that the partial pressures of all possible adsorbates be less than 10^{-12} Torr. Vacuum pressures at this level are at the limits of modern technology and are achieved with only cryopumping or by very careful use of vacuum materials.

As a compromise between the demands of temperature stability and the requirement for clean surfaces, gold was chosen as the electrode material because none of the common adsorbates, including molecular oxygen, have been found to chemisorb at room temperature and it could be conveniently formed into electrodes with clean, smooth surfaces prior to installation into the temperature control and micropositioning assembly.

The approach taken was to fabricate the gold electrodes in one UHV system, to store and transport the fabricated electrodes in desiccated containers, and finally to assemble the electrodes into an electrode assembly in the laboratory atmosphere. As discussed in the next paragraph, this procedure certainly does not result in an ideally clean surface. However, after the electrodes are placed in a UHV environment, the surfaces should be altered by at most a monolayer of hydrocarbon and require no further processing before the vacuum tunneling experiment could reliably be performed at a stable temperature.

The most definite study of gold's resistance to adsorbing molecular species was performed by Chesters and Somorjai [55]. They found that of the species most likely to be encountered in a laboratory environment, oxides, sulphides, hydrocarbons, nitrides, and water vapor, the only ones which were observed to adsorb on a gold (111) crystal face or a stepped crystal face were heavy hydrocarbons. Their investigations were with the crystal's temperature less than 500°C and partial pressures less than 10^{-6} Torr. Naphthalene adsorbed on both types of gold surfaces and the adsorption behavior indicated dissociative chemisorption. Surface steps did not play a critical role in the adsorption properties of gold, in contrast to that of the similar chemical material, platinum. Earlier work by Trapnell [56] using polycrystalline gold films showed that reversible molecular adsorption of CO , C_2H_4 , C_2H_2 , and other light hydrocarbons did occur with low heats of adsorption (0.4 eV/molecule) for high pressures. These low heats of adsorption allow the adsorption process to be reversible and Trapnell [56] shows that at low pressures and room temperature, the

light hydrocarbons desorb with no measurable, i.e., less than 2% coverage at pressures less than 10^{-6} Torr.

Work function changes, *WFC*, in gold due to adsorption have received very little study, in part because gold is the most common reference electrode for measurements of the work function [32]. In particular, no work to determine the *WFC*'s due to the adsorption of heavy hydrocarbons was found in the literature. However, *WFC*'s in platinum have been studied for a wide variety of molecular species. While the type and behavior of molecular species which adsorb on gold surfaces is markedly different from that on platinum surfaces, the *WFC*'s observed in the studies of adsorption on platinum at least set an upper bound on the *WFC*'s to be expected for adsorption on gold. The values of the *WFC*'s of the platinum do reflect the effective dipole moments of the adsorbed molecules on a dominantly *s*-electron like surface and should give a valid estimate of the dipole moment of the molecule when it is adsorbed on gold. The major

difference between platinum and gold from the standpoint of chemical properties is that platinum has an unfilled 5d electronic subshell. For platinum, the *WFC*'s were measured by Gland and Somorjai [57] for a wide range of heavy hydrocarbons. *WFC*'s were found to be between -1.0 and -2.0 eV depending on the type of adsorbate and the degree of coverage of the platinum surface.

The gold electrodes were formed by melting the end of a 1.25 mm gold wire in UHV with electron bombardment. The apparatus used for forming the electrodes is shown and described in figure 19. Gold wire with 99.99% purity was used. Following degreasing with a trichloroethylene, acetone, ethanol sequence, the wires were etched in boiling aqua-regia, i.e., 3.5 parts HCl and 1 part HNO₃; both acids were fresh ACS grade chemicals. After rinsing the wires thoroughly in running tap water, they were ultrasonically cleaned in de-ionized, filtered water and finally blown dry with hot air. The wires were then installed in the electron bombard-

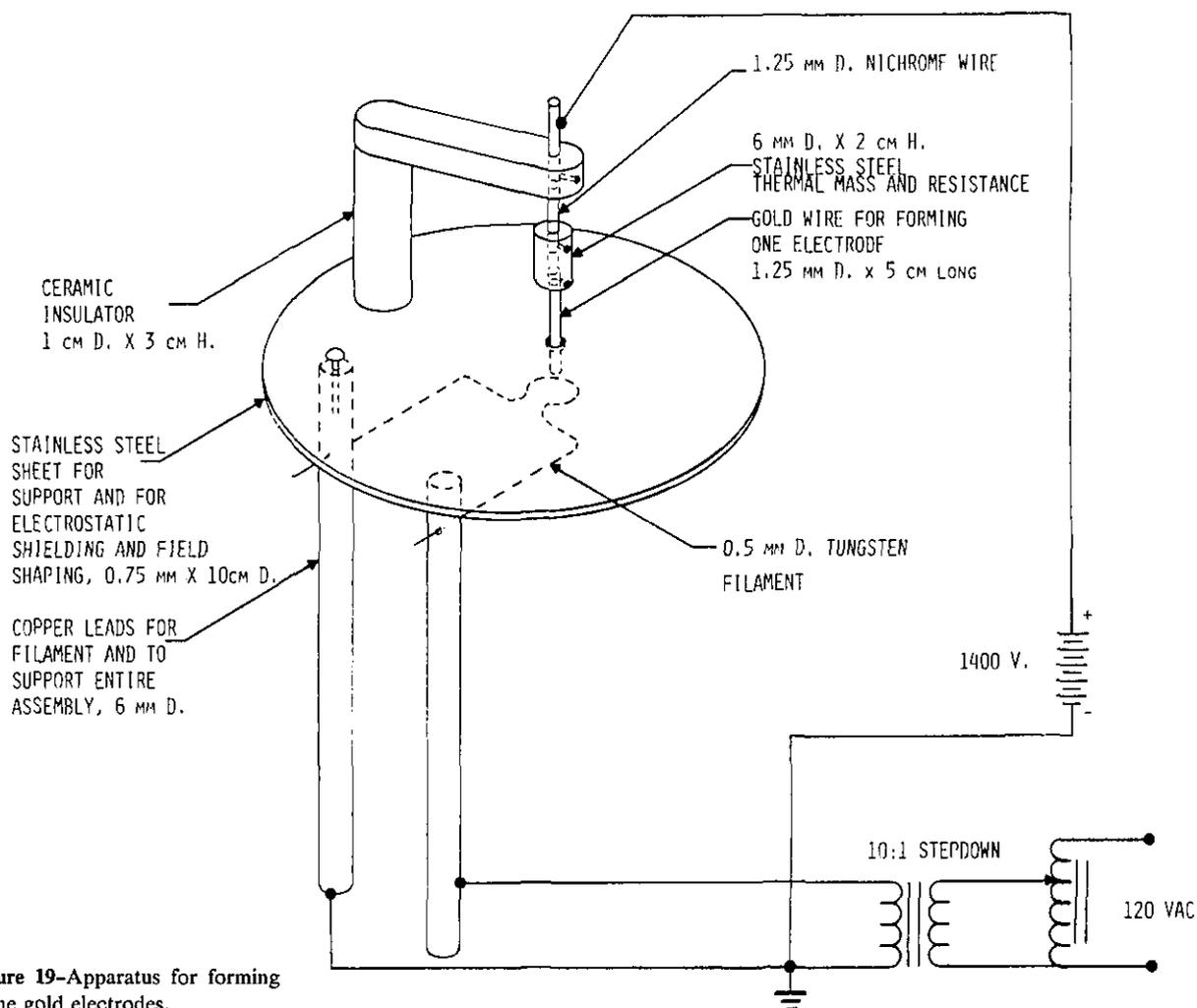


Figure 19—Apparatus for forming the gold electrodes.

ment apparatus and vacuum pumping was begun.

The vacuum system used for the melting operation utilized a sublimation-ion pump combination which maintained pressures of 2 to 4×10^{-8} Torr while the gold was molten. With the 1.4 kV indicated in figure 19, only about 15 mA of bombardment current was required to melt the end of a gold wire.

The stainless steel mass-thermal resistance is an essential component of the melting apparatus. This component combined with the nichrome wire provides sufficient thermal resistance to allow an equilibrium between input and dissipated power to be obtained by convenient manual adjustment of the filament heating current. Without these components, the equilibrium is difficult to obtain; the nonequilibrium result in this case is that the molten region of the wire moves up the length of the wire, forms a sphere and drops off the wire. After several of these occurrences, the configuration shown in figure 19 was developed to avoid this problem. With this design, a molten sphere could be maintained on the end of the wire with the support of the liquid's surface tension for 10 to 20 minutes. These long periods in the molten state in vacuum provided a means to evaporate many contaminants from the surface and were sufficient for the surface tension forces to determine the microscopic surface topography.

After melting, the electrode surfaces were inspected with an optical microscope at a magnification of 300X. The process from etching in aqua regia to melting was then repeated until no surface contamination was visible. For the pair of electrodes used in the experiment, the process was repeated four times. The resulting surfaces were very smooth with no visible topography or contaminants. The electrodes were finally examined with a scanning electron microscope (SEM) after the experiment was completed. The SEM micrographs of the electrode surfaces are shown in figures 20 and 21. The surfaces were sufficiently smooth that in both the electron and optical microscopes good focus was difficult to obtain.

The anticipated electrode surfaces for this experiment were, therefore, clean gold surfaces altered by at most a monolayer of adsorbed heavy hydrocarbons to produce work functions which could range from gold's work function, 5.0 eV, down to a minimum of 3.0 eV.

Temperature Control System

In order to provide the required stability of ± 1 pm for the interelectrode spacing, the design goal

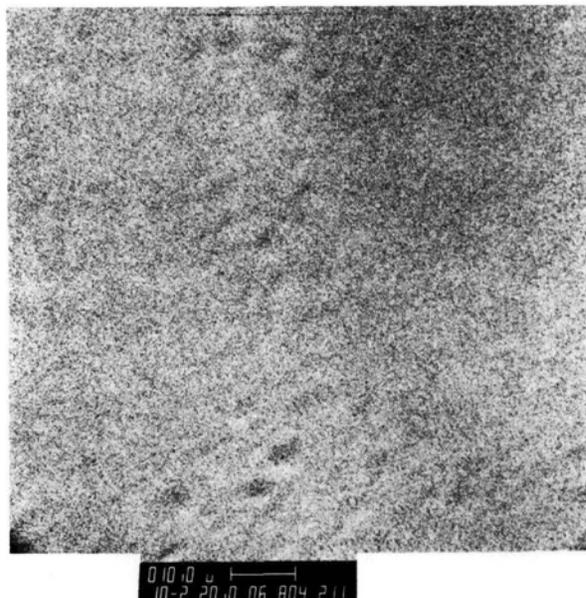


Figure 20—Scanning-electron-microscope photomicrograph of gold electrode surface; view normal to surface.

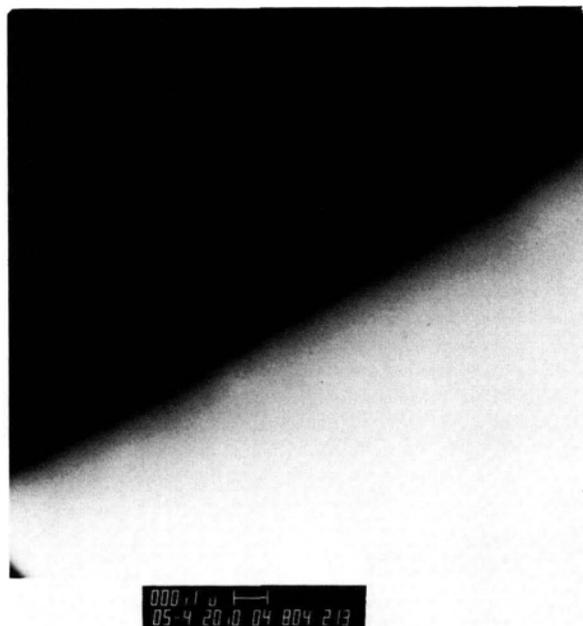


Figure 21—Scanning-electron-microscope photomicrograph of gold electrode; surface projection.

for the temperature control system was to produce a temperature stability of ± 20 μ K. Other workers [58,59] have shown that by careful design, temperature stabilities at this level can be achieved with the use of a two-stage control system. This approach was therefore adopted. The first and outer stage was designed to operate at a temperature of 5 K above the laboratory temperature with a stability of ± 1 mK. The second and inner stage was designed to operate at a temperature 0.5 K above the

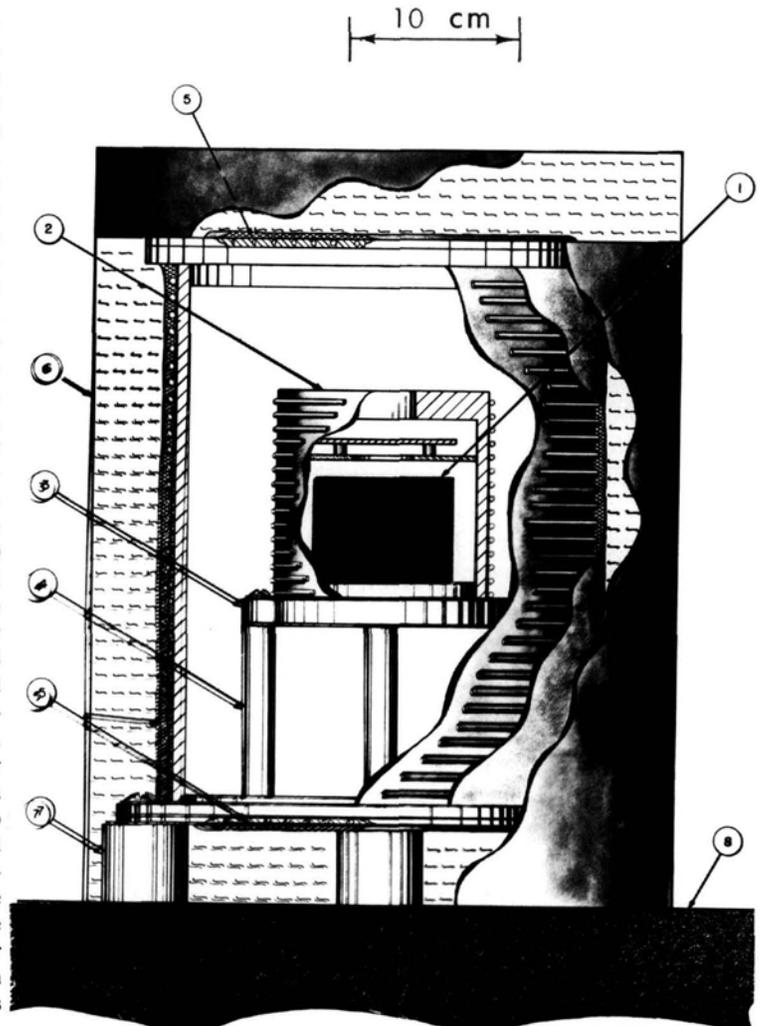
outer stage's temperature with the specified design goal stability.

The apparatus used for the combined purposes of temperature control, vibration isolation, and acoustical isolation is shown and described in figure 22. Vibration and acoustical isolation aspects of the apparatus are described in a later section. Several features of this apparatus should be emphasized: 1) Bifilar windings were used for both control heaters to insure that stray magnetic fields in the region of the electrode assembly would be minimal. 2) Radiation baffling was provided at the vacuum pumping part of the inner heater enclosure to minimize the radiant coupling between the outer heater's fluctuations and the electrode assembly. The baffle design insures that any radiation entering the pumping port must be scattered at least five times from the surface of the inner heater enclosure. 3) All joints in the polyurethane insulation and openings for the various vacuum parts were taped with

building construction duct tape to minimize convection losses. 4) Electrical leads passing from the laboratory through the first stage and from the first stage through the second stage to the electrode assembly were in both cases solidly heat-sunk to the enclosure of each respective stage. For the outer stage, thermal contact for each electrical lead was obtained by soldering one point of the lead to a copper post which was mounted to the outer enclosure with the same high thermal conductivity epoxy used to mount the heater wires; see figure 22. Thermal contact for leads passing through the inner stage enclosure was obtained by using copper feed-throughs in the base of the enclosure, item 3 in figure 22, which were electrically insulated with beryllium oxide washes. Total contact area for each lead was 3 cm^2 .

A block diagram of the temperature control system employed for both control stages is shown in figure 23. The basic design approach of both stages

Figure 22—Apparatus for vibration isolation, acoustical isolation, and temperature control, consisting of the following components: 1) Electrode holder and micropositioning assembly. These are described in figure 25. 2) Inner stage temperature control enclosure (6061 Aluminum) and heater. The heater wires shown are only schematic, actual wires were of #30 nichrome in a bifilar configuration which almost completely covered the enclosure's cylindrical surface. The radiation baffle is the cross-hatched piece located immediately above the electrode assembly. 3) Aluminum (6061 alloy) base for inner enclosure 4) Alumina standoffs. Three were used to provide thermal isolation between the two temperature control stages. 5) Outer stage temperature control enclosure, heater and vacuum chamber walls. The cylinder wall thickness was 3.2 mm, top and bottom thicknesses were 1.9 cm. Heater wires were #31 copper spaced approximately as indicated in the drawing. An epoxy coating approximately 3.0 mm thick with a thermal conduction coefficient of $0.01 \text{ (cal cm)/(sec cm}^2 \text{ }^\circ\text{C)}$ [60], was used to bond the heater wires to all surfaces of the outer enclosure and to provide acoustical damping. 6) 8 cm thick polyurethane foam. This foam provided both thermal and acoustical isolation of the system enclosure and 7) Epoxy standoffs, three, were used to provide thermal insulation of outer enclosure and vacuum system from the vibration isolation table. These standoffs were bonded to the vibration isolation table top and permitted rigid coupling of the entire apparatus to the isolation table top. And 8) Granite isolation table. The table was $61 \text{ cm} \times 25 \text{ cm}$ and had an approximate mass of 409 kg. It was supported with four damped steel springs from a simple aluminum frame. The resonant frequency of the system was less than 1 Hz for vertical motion.



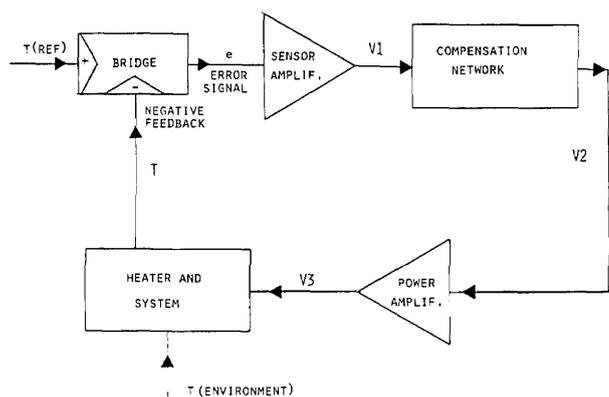


Figure 23—Block diagram of temperature control electronics.

was to use the output of a bridge to drive a high gain amplifier and compensation network to account for the lags of the detectors and stage-masses, then to close the control loop by incorporating a thermistor detector as one element of the bridge. A full discussion of the electronics of both control stages is given in Appendix V. There it is shown that one of the most important parameters of the control loop is the time response of the thermistors with respect to changes in power input to the heaters. For this reason, the thermistors were mounted in a manner to allow the closest possible coupling between the heaters and the thermistors.

The principal difference between the two stages of temperature control was that a dc bridge and a sensor amplifier were used for the outer stage while an ac excited bridge and lock-in amplifier were used for the inner stage. The compensation

network, power amplifier and heaters were tailored for the requirements of each stage. Values for the heat capacities, thermal resistances, and masses of each stage are given in Appendix V.

Both controllers were designed to allow convenient and accurate changes in operating temperature. The outer stage's setpoint could be changed in 2 mK increments with the use of a switched resistor network. Setpoints for the inner stages were changed with a seven-digit ratio transformer which permitted 10 μ K increment changes.

Performance of the outer control stages was checked by using the inner stage detector system as a temperature monitor. When this stage's compensation network was adjusted as described in Appendix V, temperature deviations from the setpoint, measured over a 10-hour period, were no greater than ± 0.5 mK. The only check on the inner control stage's performance was to monitor the stability of the interelectrode spacing. The results of this check are given in section 4 where it is shown that no changes in electrode spacing due to temperature variations were observed. A major indication of the sensitivity of the inner stage controller was that a setpoint change of 30 μ K would cause an observable change in the output of this stage's sensor amplifier. The response of both stages to setpoint changes is described in Appendix V.

Vacuum System for the Tunneling Experiment

A schematic diagram of the vacuum system used for the tunneling experiment is given in figure 24.

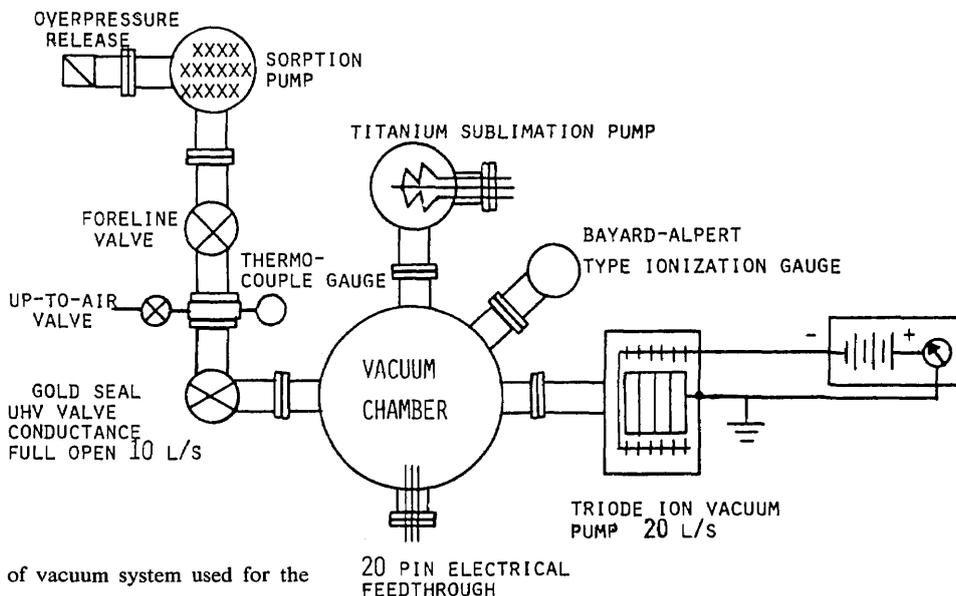


Figure 24—Schematic diagram of vacuum system used for the tunneling experiment.

The vacuum chamber and pumping system were constructed from 304 stainless steel with standard UHV copper gasket flanges employed for all connections between system components. The system was pumped from atmospheric pressure down to the operating pressure using only the three pumps shown in figure 24 in order to avoid exposing the gold electrodes to any unnecessary hydrocarbon gases. Base pressure of the vacuum chamber as determined by a crosscheck of the ion-pump current and the Bayard-Alpert gauge was 5×10^{-9} to 1×10^{-8} Torr after several weeks of pumping. The

Bayard-Alpert gauge was not operated during the critical temperature control parts of the experiment.

Electrode Holder and Micropositioning Assembly

The final design of this critical part of the vacuum tunneling experiment is shown in figure 25. A photograph of the electrode assembly, along with the temperature control heaters and vacuum system, is given in figure 26. The assembly as shown

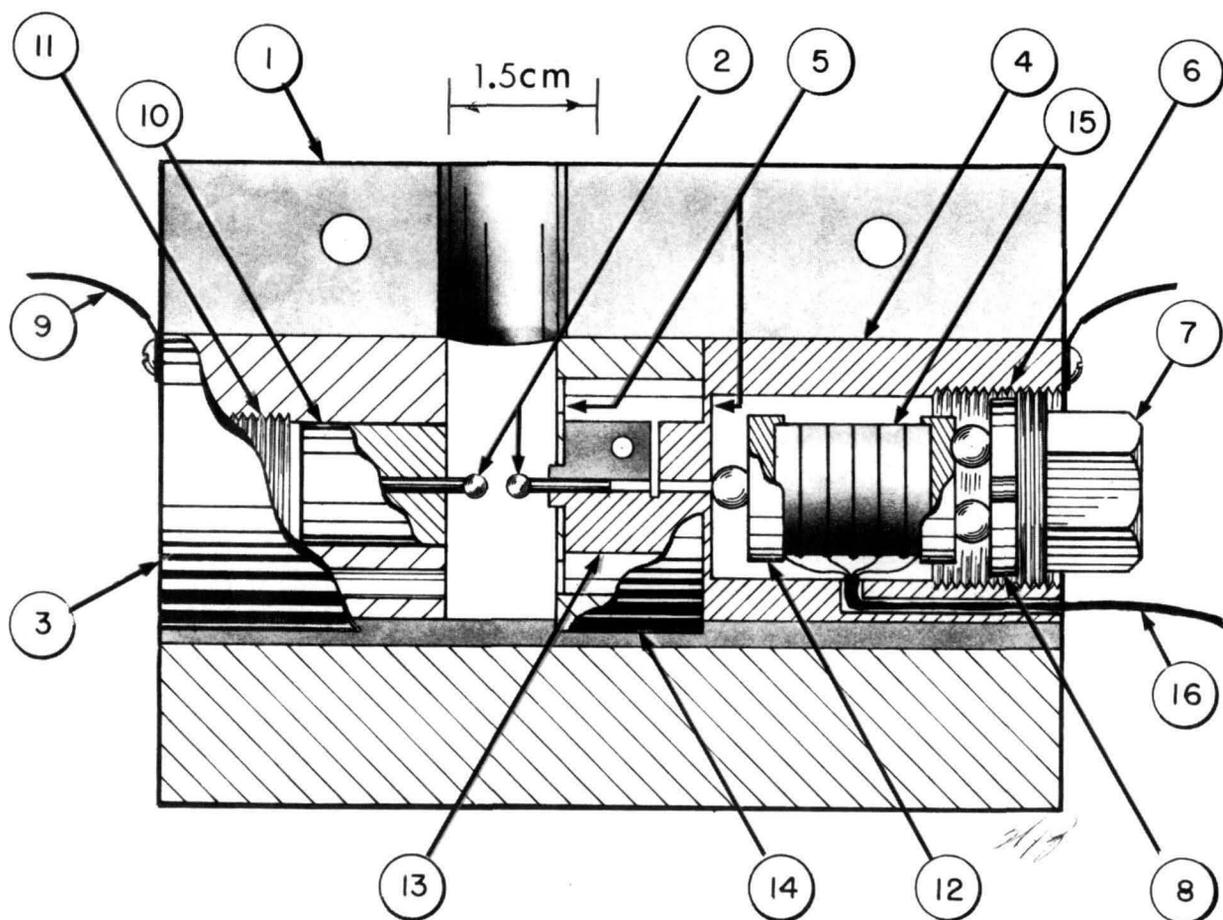


Figure 25—Electrode holder and micropositioning assembly. (1) Housing for electrode holders. This housing is basically an aluminum (6061 alloy) block $8.9 \text{ cm} \times 6.35 \text{ cm} \times 6.35 \text{ cm}$ with a 3.17 cm hole bored and reamed to accept the electrode holders. The diameters of the electrode holders were machined to match the hole in the housing to a clearance no greater than 0.01 mm . (2) Gold electrodes. In the experiment, both gold wires were inserted into their respective holders as far as the spherical ends would allow. (3) OFHC copper mount for the differential-screw mechanism and main body of left electrode holder. (4) 0.025 mm thick polyimide film [63] to provide electrical insulation between the electrode holders and the housing. (5) Diaphragm flexures. The thickness of both flexures was 0.38 mm . (6) Threads to allow prestressing of the piezoelectric element-diaphragm flexure system. (7) Prestressing nut. (8) Hardened steel washer to prevent motion transfer between prestressing nut and the piezoelectric element. (9) Electrical connections to gold electrodes. (10) Clamp-holder for left electrode, OFHC copper. (11) Threads to allow mounting of the differential-screw mechanism. (12) Hardened steel end caps to distribute stress uniformly over piezoelectric element's surface. (13) Clamp-holder for right electrode. (14) Support for outer diaphragm flexure, OFHC copper. (15) PZ displacement element.

in figure 26 provides a system which is compatible with ultra-high vacuum, has mechanical rigidity to maintain electrode spacings to stabilities of approximately 1 pm, and permits the electrodes to be assembled into a tunneling structure on a laboratory scale. In addition, the assembly is capable of producing interelectrode spacings of 1 to 2 nm and has the facility to change the electrode spacing in a rapid and convenient manner.

The basic approach of this design is that a precision-differential screw mechanism is inserted at point 11 in figure 25 to position the left electrode clamp, item 10, which contains one gold electrode until the electrode spacing is less than the distance which can be spanned by the piezoelectric (PZ) element. At this point, all components of the assembly are clamped rigidly in place, the differential-screw mechanism is removed, and final fine tuning of the electrode spacing is achieved by temperature adjustment and piezoelectrically generated displacements of the second electrode holder.

With reference to figure 25, the left electrode holder is composed of the electrode clamp, and a differential-screw-mount. Both were made of OFHC copper to obtain minimum thermal resis-

tance between the electrode and the holder housing and minimum electrical resistance between the electrode and its electrical connection. The electrode clamp was essentially two half-cylinders connected together by a 0.25 mm \times 0.5 mm flexure along the length of the clamp and located next to the hole for the electrode. A separate screw is provided for clamping the electrode into place. The mount for the differential-screw had a similar form. All other clamping is performed with the electrode-holder housing which is cut only in the shaded area of the drawing.

Each vernier division of the differential-screw-mechanism corresponded to an output displacement of about 25 nm. This resolution was such that approximately one-quarter turn of the differential-screw's input covered the PZ element's 500 nm range.

The right electrode holder consisted of an electrode clamp mounted to a double diaphragm-flexure support to insure axial motion of the electrode, a PZ element, and a mechanism to prestress the driving element-flexure combination so that microscopic motion losses would be minimized. End caps for the PZ element and the prestressing mechanism were made from hardened steel. The di-

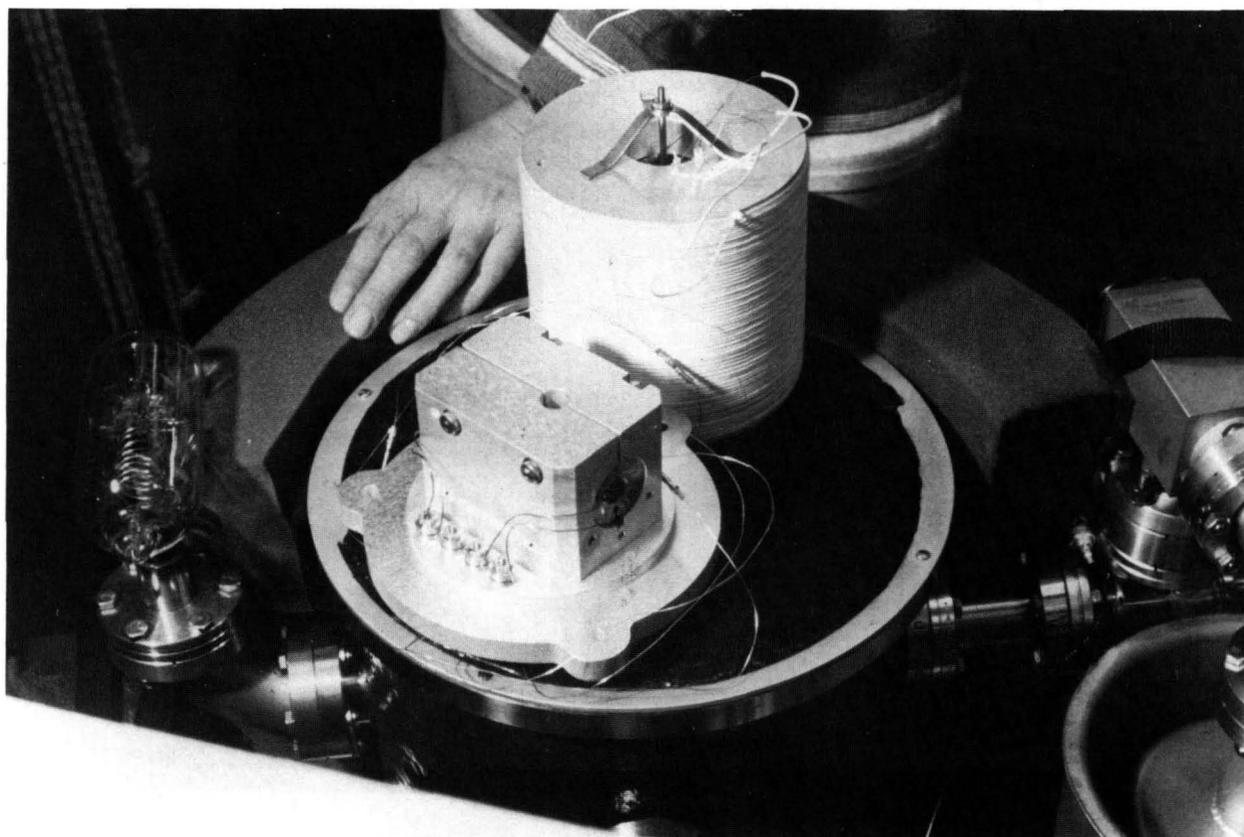


Figure 26—Electrode assembly, temperature control heaters, and vacuum system.

aphragm-flexures and piezoelectric element were prestressed with a force of approximately 890 N (200 pounds). All other parts of the electrode holder, except the 304 stainless steel fasteners, were machined from OFHC copper. The PZ element was electrically insulated from the rest of the holder with four sapphire spheres.

The PZ element consisted of six ceramic discs, each 1.27 cm in diameter by 0.25 cm thick. The ceramic was a barium-titanate-zirconate composition which has very low hysteresis, linear characteristics [61]. The relevant expansion coefficient for the material is $d_{33}=225$ pm/V for fields along the poled direction of the ceramic; here perpendicular to the disc faces. The individual disks and end caps were bonded into one element with the use of a low viscosity, electrically conductive epoxy. The epoxy's electrical resistivity was less than 0.002 ohm-cm [51]. After baking in air at 100 °C for one hour, the outgassing rate was sufficiently low that the 10^{-10} Torr base pressure of a bell jar system with a 50 l/s ion pump was unaffected by the presence of the epoxy bonded piezoelectric element.

Displacement characteristics of the completely assembled micropositioning system were measured before and after the tunneling experiments described in section 4. Before the experiments, the displacements were measured with a precise but uncalibrated mechanical gauge to be $(.56 \pm .06)$ nm/V. After the experiment, the displacements were measured with a Hewlett-Packard 5526A laser interferometer measurement system in combination with a 10565B remote interferometer and retroreflector. This measurement system enabled displacements to be measured with an accuracy of

± 25 nm. For both expansion and contraction of the piezoelectric element, the displacements were measured as $(0.47 \pm .03)$ nm/V. Thus, taking into account the uncalibrated state of the mechanical gauge, the displacement characteristics of the micropositioning system were stable throughout the experiment.

The need for 1 pm electrode spacing stability combined with the 470 pm/V sensitivity of the PZ element requires that the high voltage power supply used to drive the PZ element have a very low voltage noise level, less than 2 mV. The power supply used for this purpose had a noise level of less than 75 μ V peak-to-peak and the capability of being switched in 1 mV increments from -1100 volts to +1100 volts [64]. To enable continuous voltage sweeps over ± 10 volts, a ramp generator was added in series with this power supply. The circuit diagram of the ramp generator and a filter circuit to suppress switching transients are shown in figure 27.

Two other experimental techniques for holding and positioning the tunneling electrodes were thoroughly explored before finally adopting the electrode holder and micropositioning assembly design just described. The technique first explored was to fabricate the tunneling structure as a microelectronic device with the use of photolithographic methods. This approach was not adopted because the technology to which the author had access could not meet the dimensional requirements of the tunneling device. However, as a byproduct of this work, the approach may result in an electronic device with a very high transconductance; values up to 1 mho appear feasible. A patent application

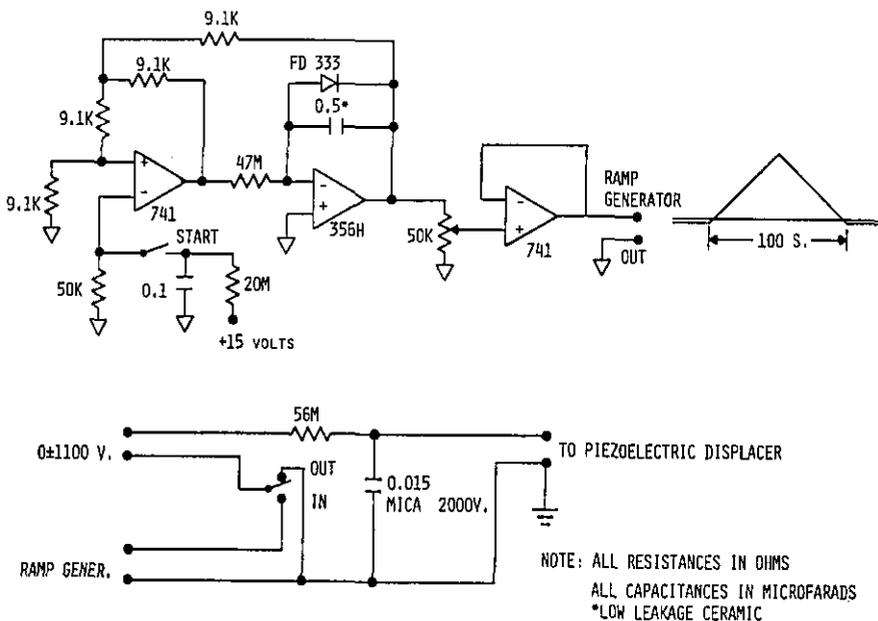


Figure 27—Schematic diagram of ramp generator and high voltage power supply for the piezoelectric element.

on the concept is being submitted to the National Bureau of Standards Patent Review Committee. The concept and a prototype device are described in Appendix IV.⁵

Lever amplification of the piezoelectric element's length changes was also explored as a possible means to bridge the gap between laboratory scale distances and the microscopic dimensions required by the tunneling experiment. While this technique proved to have excessive vibration levels for the tunneling experiment, it has characteristics which are very useful for scanning-electron and optical microscope stages [65].

Vibration and Acoustical Isolation

In the previous metal-vacuum-metal tunneling experiment [12] similar to this work and in all earlier attempts made in connection with this work, vibrations of the tunneling device structure at the 1 nm level have limited a successful execution of the experiment. Every effort was made, therefore, to carefully examine the mechanical design of each part of the apparatus to insure that all known sources of vibrational energy were attenuated to levels which would not result in detrimental changes in the electrode spacing. The vibrational properties of what was thought to be the most likely contributors to changes in electrode spacing will now be described in terms of one-dimensional models.

Starting with the electrode holder and micropositioning assembly, the most likely contributor is vibrational motion of the right electrode holder as the entire assembly is vibrated along the assembly's axis. This motion could be produced either from acoustical pressure variations on the bottom of the vacuum chamber or from building vibrations driving the isolation table top and, thus, the entire vacuum chamber. As a model of this motion, let a mass, m , be connected by a spring with force constant k , and a viscous damper characterized by the constant C , to a vibrating wall whose displacement is given by $x_w = A_0 \sin \omega t$. Then the equation of motion of m in terms of x , measured relative to the wall is

$$m(\ddot{x} + \ddot{x}_w) + C\dot{x} + kx = 0. \quad (3-6)$$

The steady state solution to eq (3-6) with the assumed form of x_w is

⁵A variation on the concept proposed here has recently been developed by J. Moreland and J. W. Elkin, Electron Tunneling Experiments Using Nb-Sn 'Break' Junctions, J. Appl. Phys. 58 (1985) 3888. They use a bending beam to produce and vary the spacing between electrodes.

$$x = R^2[(1-R^2)^2 + 4V^2R^2]^{-1/2} x_w, \quad (3-7)$$

$R = \omega/\omega_0$, $\omega_0 = (k/m)^{1/2}$, and $V = 1/2C(mk)^{-1/2}$. If ω is much less than ω_0 , then

$$x = R^2 x_w. \quad (3-8)$$

Thus, if ω_0 is the lowest resonant mode frequency of the micropositioning assembly for axial vibrations and the full amplitude of the isolation table top's horizontal motion is transmitted along the axis to the electrode holder housing, $f_0 (= \omega_0/2\pi)$ must be greater than 6 kHz. This follows with the use of an ω of $(2\pi)30$ radians/s, an x_w of 40 nm, and a requirement that x be less than 1 pm. The only mechanical element with a resonant frequency near this value is the diaphragm flexures, which were designed to have a resonant frequency of 10 kHz, when unstressed. When stressed, the frequency would be even higher and motion would then be determined almost entirely by the pre-stressing mechanism.

For motions normal to the axis of the assembly all resonant frequencies are much higher than 10 kHz. The relatively large gold wires were used to avoid low resonant frequencies in those directions. For example, even with a very high modulus of elasticity material such as tungsten, the transverse vibrational resonant frequency of a cantilever only 2 mm long and .025 mm in diameter is 4 kHz [66].

To minimize the affects of any acoustically generated vibrations, the electrode holder and micropositioning assembly are only connected mechanically to the laboratory environment by the bottom plate of the vacuum chamber. This plate is heavily damped both by the epoxy layer for attaching the heater wires and by the polyurethane foam.

The four lowest eigenfrequencies of the bottom plate are approximately: 884 Hz, 1.85 kHz, and 3.0 kHz, and 3.5 kHz [66]. These neglect any mass loaded decreases in the frequencies due to the mass of the tunneling assembly, which is supported by the plate and any increases in the frequencies due to the epoxy damping. The lowest of these frequencies again is the most likely source of transmitting acoustical pressure variations to the electrode assembly. As a safety factor for design, a 40 dB white noise spectrum was assumed to exist up to one kHz, i.e., pressure variations, P_w , with amplitudes of 2×10^{-3} N/m². For frequencies much less than the eigenfrequencies, the vibrational amplitudes, Y , of the plate at the center of the plate would be [55]

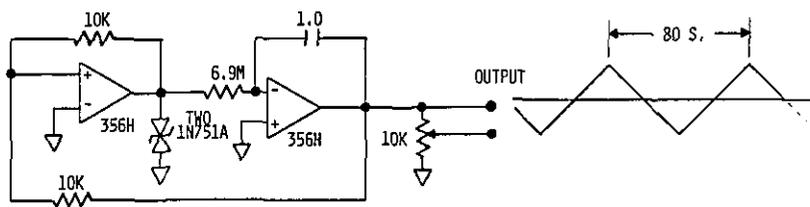
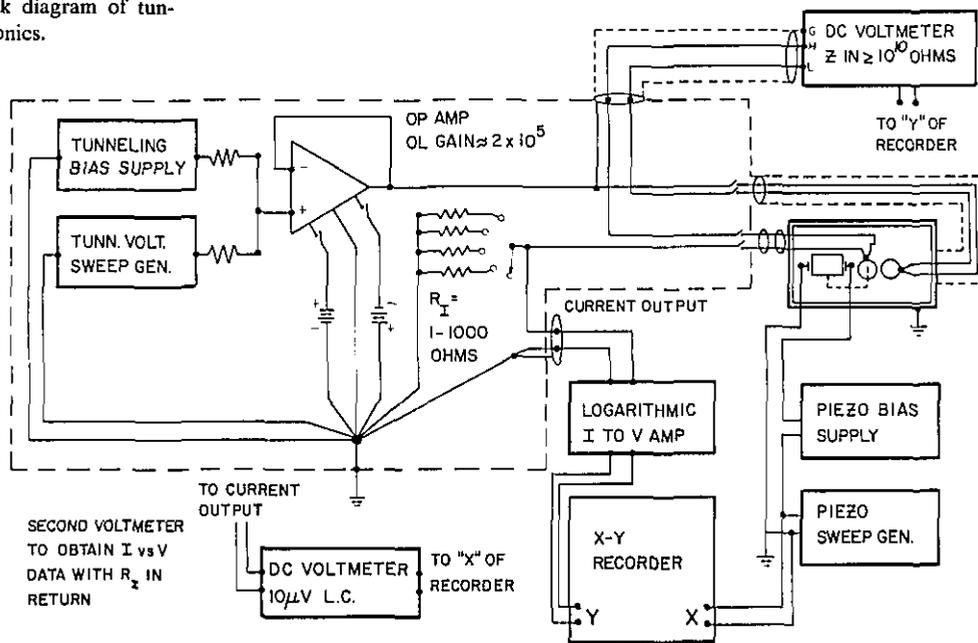
$$Y = 3P_w(1-s^2)a^4/(4tQ), \quad (3-9)$$

s = Poisson's ratio = 0.33, a = radius of the plate = .127 m, t = thickness of the plate = 1.9 cm, and Q = modulus of elasticity of the plate = 1.86×10^{10} N/m². With the indicated P_ω , Y is only 8.7 μ m. For frequencies on resonance, the amplitudes of vibration should be no larger than 10 times this value with the epoxy and polyurethane damping [67]. Since the assumed pressure variations were larger than expected and did not account for any attenuation due to the polyurethane foam, no further stiffening of the plate was performed.

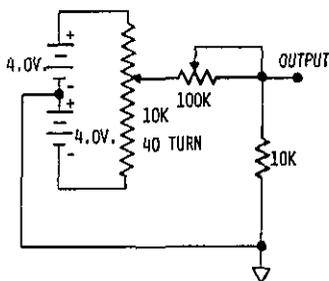
Tunneling Electronics

The electronics used to measure the conduction properties of the tunneling devices are shown in figures 28 and 29. Requirements of the electronics were that it permit one to test devices whose impedances could range from 10^{-2} to 10^{11} ohms with dc test voltages of typically 10 mV. The design shown in figure 28 utilized an operational amplifier in a follower configuration to insure that, other than voltage drops in the current leads, the

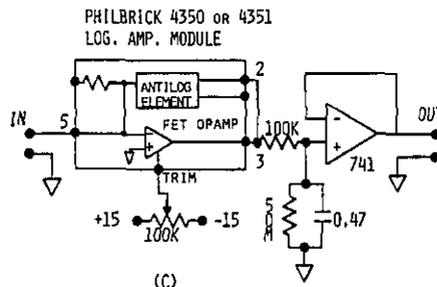
Figure 28—Block diagram of tunneling electronics.



(A)



(B)



(C)

NOTE: ALL RESISTANCES IN OHMS
ALL CAPACITANCES IN MICROFARADS

Figure 29—Schematics for components of the tunneling electronics. (A) Ramp generator for sweeping bias voltage. (B) Manual sweep and bias supply. (C) Logarithmic amplifier.

output impedance of the voltage source would be less than 0.005 ohm. The operational amplifier was a 356H integrated circuit device. The four-wire lead system was carried through to the connection points on the electrode holders shown in figure 25.

With all systems of the tunneling experiment in operation, i.e., vacuum system, temperature control system, high voltage power supply for the micropositioning assembly, and the tunneling electronics, the total system was carefully debugged for ground loop problems and 60 Hz pickup. The noise current in the tunneling electronics after this process was about 10^{-11} A. This value was measured with a 10^8 ohm test resistor substituted for the tunneling device and obtained from both the I - V characteristics of the resistor and the output of the logarithmic current to voltage converter. The major source of this noise was the operational amplifier and could have been reduced with a more elementary resistive voltage generator. However, the level was adequate for this experiment, and the alternative was a greatly increased source resistance.

Both dc voltmeters employed with the tunneling electronics had floating-guarded inputs with dc common-mode rejection ratios of greater than 140 dB. The voltmeter used to measure the bias voltage applied to the tunneling device had a gain of 10 with readout resolution to 10 μ V. The voltmeter for measuring the tunneling current, in the I - V configuration of the electronics, had variable gains up to 10^6 with a meter noise of about 150 nV p-to-p.

The logarithmic current-to-voltage amplifier output was trimmed according to the manufacturer's specifications. After warmup, the offset was less than 100 μ V. Specifications for the amplifier modules were that the output voltage would be a linear function of the logarithm of the input current over a range from 10^{-9} to 10^{-3} A to within 1%. Measurements with test resistors showed that, in addition, the nonlinearity was only 5% for the positive polarity module and less than 2% for the negative polarity module for currents of 10^{-10} A. To reduce the acoustical noise level of the chart recorder when the tunneling device formed an open circuit, a lowpass filter with a cutoff frequency of 32 Hz (see figure 29c) was installed.

In operation, the electronics were changed from the Log I_{cv} - S configuration shown in figure 28 to the configuration for I - V measurements by first shorting the input of the Log module with the 1 ohm resistor, disconnecting the Log module and the piezoelectric sweep generator and finally connecting the dc voltmeter for current detection.

Changing back to the Log I_{cv} - S configuration was performed in a similar manner.

4. Experimental Results and Analysis

Taking into account the extreme delicacy of the interelectrode spacing adjustments, the apparatus described in section 3 has proved to be very reliable and to give reproducible results. The data which will be presented were obtained with the same essential characteristics in two separate experiments with data being obtained over a period of several weeks in both instances.

After several unsuccessful attempts at getting the apparatus to perform as anticipated, a standard sequence of operations was evolved which permitted the experiment to be carried out in a consistent manner. Following formation of the electrodes and their installation into the holder and micropositioning assembly, as described in section 3, the assembly was placed on a temperature-controlled plate of copper. The temperature of the plate was maintained at (300 ± 0.05) K which was the planned operating temperature for the experiment. The electrode spacing was then observed with a low power stereo-microscope and adjusted with the differential-screw mechanism until the electrodes appeared as shown in figure 30.

Concurrent with this adjustment the current flow through the electrodes for an applied voltage of 10 mV was monitored with the tunneling electronics. Final adjustment of the left electrode in the sense of figure 25, was such that with the PZ element at the maximum extension possible in air the electrodes would just make contact. Since the extension of the PZ element in air was only 0.7 of that possible in vacuum, slight misadjustments could be corrected after installation in the temperature-controlled vacuum system. Misadjustments out of range of the PZ element, either in contraction or expansion, could be corrected with small, ± 0.5 K, changes in the operating temperature. Use of the temperature-controlled copper plate was required to overcome the expansion and contraction generated by unknown changes in temperature resulting from body heat during the adjustment process.

All parts of the electrode holder and micropositioning assembly were firmly clamped in place at this time. The assembly was then allowed to equilibrate with "hands-off." (Plastic gloves over cotton gloves were actually used at all times during the assembly to provide thermal insulation and to insure the cleanliness required for UHV.) After con-

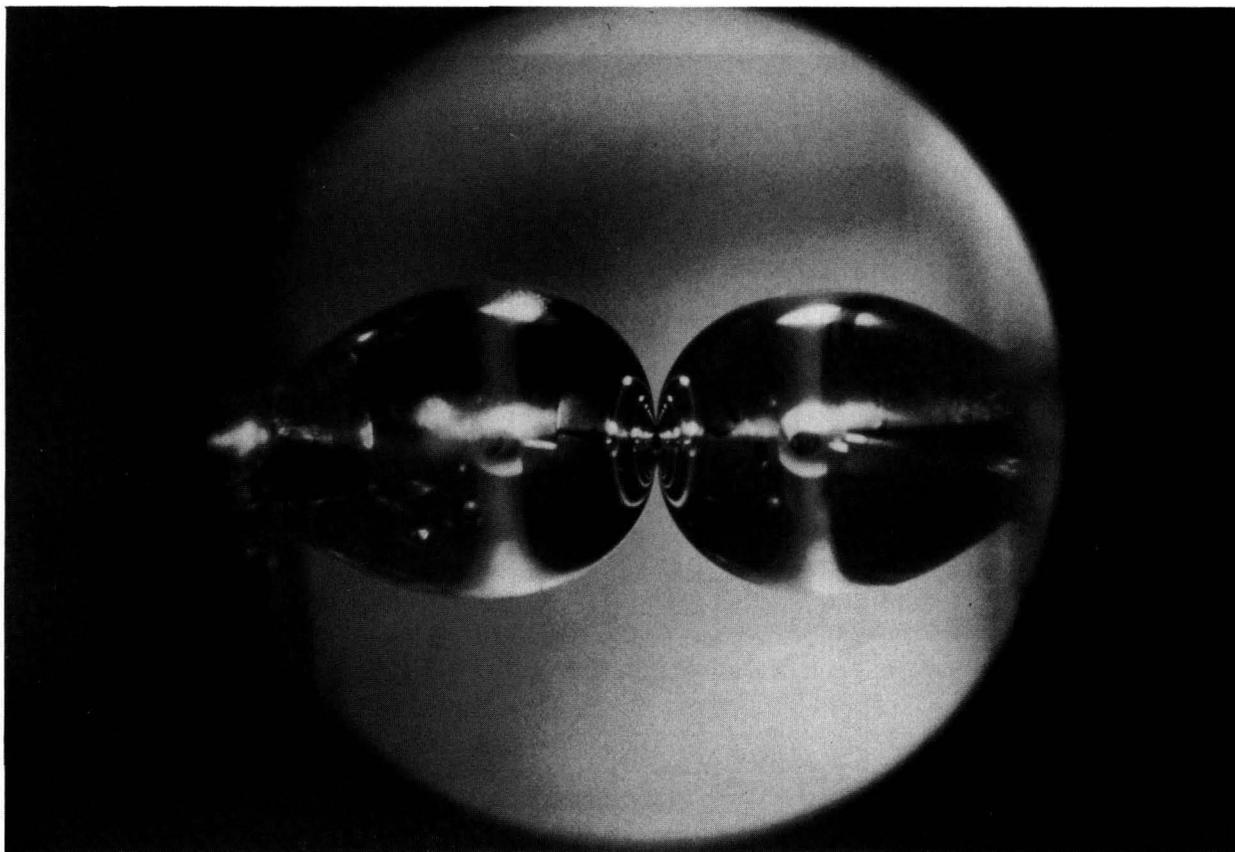


Figure 30—Gold electrodes mounted in electron assembly. All the apparent structure and cuts in the electrodes are images formed by the mirror like properties of the electrode surfaces.

firming that an open and a short could be produced by contracting and expanding the PZ element, the electrode assembly was installed in the vacuum system and pumped to a pressure of approximately 1×10^{-8} Torr. This usually required several weeks with the small ion pump which was used on the vacuum system.

Determining the PZ element voltage and operating temperature at which the electrodes were positioned to within 1 to 2 nm of each other was accomplished by searching for the abrupt change in tunneling current as the electrode spacing was swept through the tunneling region. With the circuit shown in figure 27 the sweep rate for a range of 5 nm was approximately 0.1 nm/s. The search was greatly facilitated with the logarithmic current amplifier. In fact, all attempts at finding the required PZ voltage and temperature settings without the use of a logarithmic amplifier were unsuccessful. After completing this procedure to locate the tunneling region the first time, successive searches were not difficult.

Data and Analysis

The first sets of $\text{Log } I_{cv} - S$ data, while exhibiting the character of many orders of magnitude change in current over 1 to 2 nm change in electrode spacing, contained much structure akin to noise; see figure 31. However, some of the structure may be due to the oscillations in the tunneling probability predicted by numerical calculations based on an exact solution to Schrödinger's equation [45]. Regardless of the possibilities, the exploration of the phenomenon was prevented by an accidental electrical discharge as an attempt was being made to measure the constant-current characteristics which were measured in both previous vacuum tunneling experiments [12].

The series of measurements which was made subsequent to this accidental discharge of the constant-current power supply will be referred to as the first experiment. Representative $\text{Log } I_{cv} - S$ data obtained during this experiment are given in figures 32a and 32b. A comparison of figures 32 and 31 shows that much of the structure present in figure 31 was probably due to fluctuating surface conditions. The larger scale oscillations in figure 31 were not found in any subsequent $\text{Log } I_{cv} - S$ measurements even though they were consistently

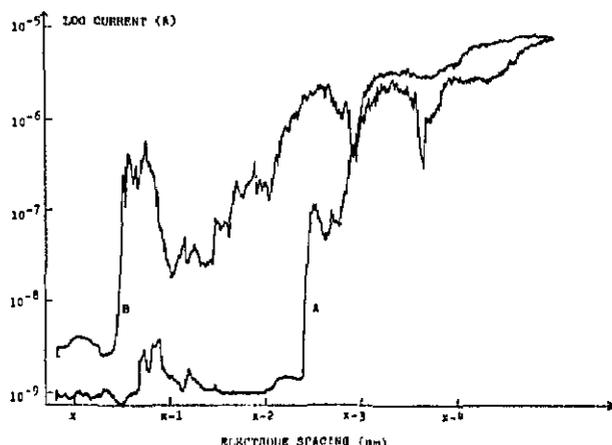


Figure 31—One of the first measured log current-spacing characteristics. Curve A was measured as the PZ element was expanded then curve B as it contracted. The hysteresis is a combination of hysteresis in the PZ element and mechanical relaxation of the electrodes. Cycling gave reasonable reproduction of both parts of the graph.

present in all measurements prior to the accidental discharge. Exploration of the meaning of the oscillations and the conditions necessary for their observation was left for future experiments.⁶

The overall shape and features of the data shown in figure 32 were reproduced about 50 times over a period of several weeks. Small changes in average slope and local details were observed as the electrodes were brought into contact during repeated closing and opening of the electrode spacing. All the $\text{Log } I_{cv}-S$ data had three characteristic regions. Expressed in terms of the tunneling current, the first of these regions was the initial transition into the current range of the Log amplifier. This region extended up to currents of approximately 10^{-10} A and covered the non-linear part of the Log amplifier's response. The direct tunneling region usually extended from currents of 10^{-10} A up to 10^{-5} A for bias voltages of 10 mV. The final region occurred as the spacing of the electrodes was such that multiple portions of the electrode surface, other than the previously dominant part, began to make significant contributions to the total current

⁶Subsequent investigations of this phenomenon have proved very interesting. R. S. Becker, J. A. Golovchenko, and B. S. Swartzentruber (Electron Interferometry at Crystal Surfaces, *Phys. Rev. Lett.* **55** (1985) 987) and G. Binnig, K. H. Frank, H. Fuchs, N. Garcia, B. Reihl, H. Rohrer, F. Salvan, and A. R. Williams (Tunneling Spectroscopy and Inverse Photoemission: Image and Field States, *Phys. Rev. Lett.* **55** (1985) 991) demonstrated oscillations in the tunneling voltage as a function of electrode spacing for a constant tunnel current. Figure 31 is a graph of tunneling current as a function of electrode spacing for a constant bias voltage. The scale of the structure here is similar to that shown in the two references cited.

and as the initially dominant portions made contact.⁷

A second attempt at obtaining constant-current measurements was made with a different power supply. In this attempt the transient discharge occurred again and was sufficient to microscopically weld the electrodes. Efforts to melt the bridge resulted in broken lead wires. The entire tunneling system was therefore disassembled, new lead wires installed, repositioned as described earlier, reassembled and all of the set-up operation repeated. Once the vacuum was at the 10^{-8} Torr level, the electrodes were cleaned with controlled capacitive discharges. A representative $\text{Log } I_{cv}-S$ graph from this second experiment is given in figure 33. Five discharges of both polarities obtained from a $0.47 \mu\text{fd}$ capacitor charged to 14 V were used for the cleaning and/or shaping of the electrode surfaces. The surface condition of the electrodes for the data of figure 33 is obviously different from that of figure 32, but the average slopes in the direct tunneling region of the two sets of data are approximately equal. The data in figure 33 were also reproduced a large number of times over several weeks.

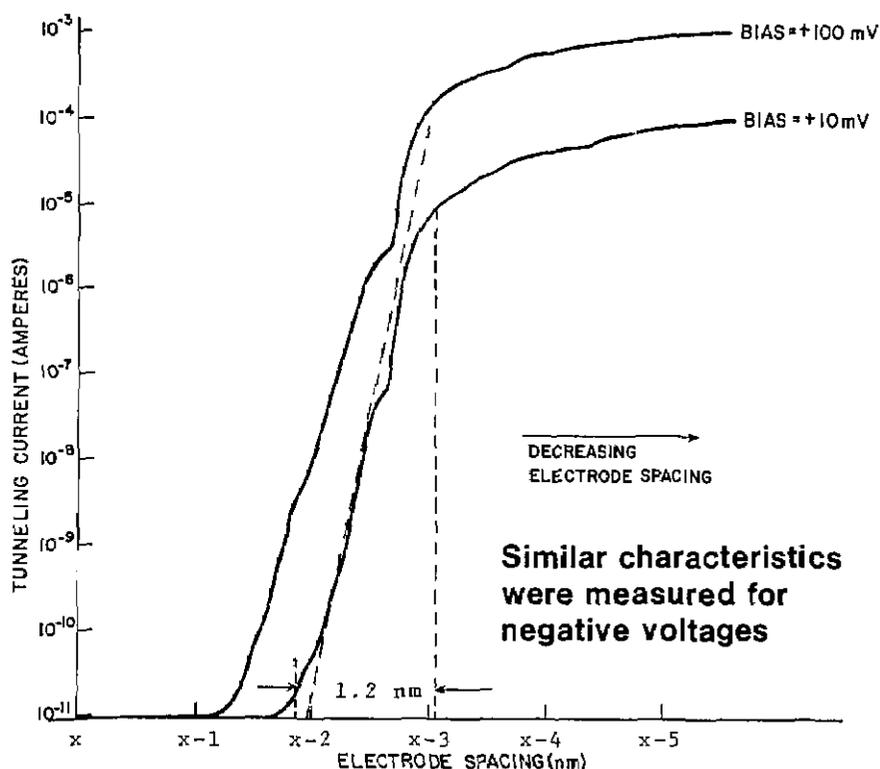
The electrode surfaces were examined in detail in a SEM after the completion of both experiments. No evidence of the discharges could be found.

To confirm that the shape and magnitude of the $\text{Log } I_{cv}-S$ graphs were not being significantly affected by the time response of the Log amplifier, a very slow scan of the PZ element's voltage was performed. The data shown in figure 34 were obtained at a scan rate of 0.01 nm/s. One may conclude from a comparison of figures 33 and 34 that the shape of the Log curves was not affected.

A $\text{Log } I_{cv}$ versus time measurement was also performed to check for temperature and any other instabilities in the experiment. Results are shown in figure 35. None of the changes in current appears to be the result of temperature variations since the transitions occur so rapidly. Changes such as the ones shown in figure 35 could be due to either sudden changes in surface conditions generated by the relatively large electric fields involved or to mo-

⁷There have been two subsequent observations of the exponential variation of tunneling current as a function of electrode spacing for a constant bias voltage: G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, *Tunneling Through a Controllable Vacuum Gap*, *Appl. Phys. Lett.* **10** (1982) 178; and C. F. Quate, *Low-Temperature Vacuum Tunneling Spectroscopy*, *Appl. Phys. Lett.* **45** (1984) 1240. The capability for making such observations is implicit in the work by Becker, et al. cited in footnote (d) and in the work by Moreland, et al., cited in footnote (g), but no explicit graphs of current versus distance or equivalent are given.

Figure 32a—Measured log current-spacing characteristics. The sloping dashed line corresponds to a work function of approximately 3.0 eV. The 100 mV curve was measured immediately following the 10 mV curve. Horizontal placement of the two curves was determined by the zeros of the PZ element's ramp generator.



Similar characteristics were measured for negative voltages

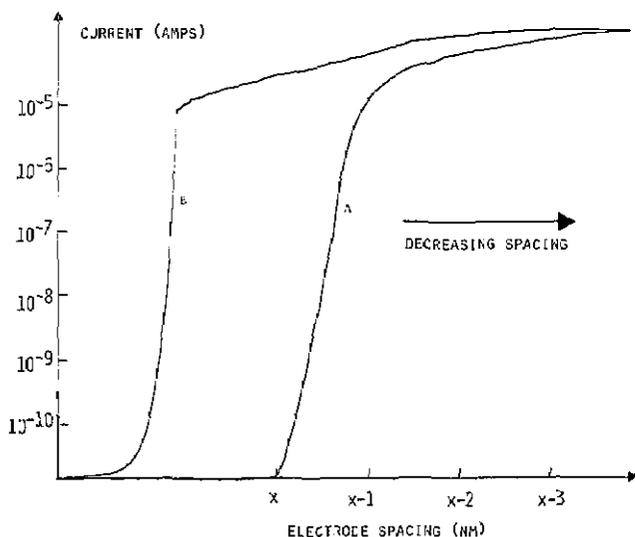


Figure 32b—Measured log current-spacing characteristics, bias voltage=10 mV. Curve A was measured as the PZ element expanded, then curve B.

tion of the electrode surfaces resulting from lattice relaxation from the strains produced by the mechanical stress of van der Waals forces and electrostatic forces. The magnitude of these mechanical stresses is discussed in the next section.

The work functions of the electrodes may be obtained directly from the Log $I_{cv} - S$ data given in figures 32 and 33. The slope of these curves in the direct tunneling region may be related to the work

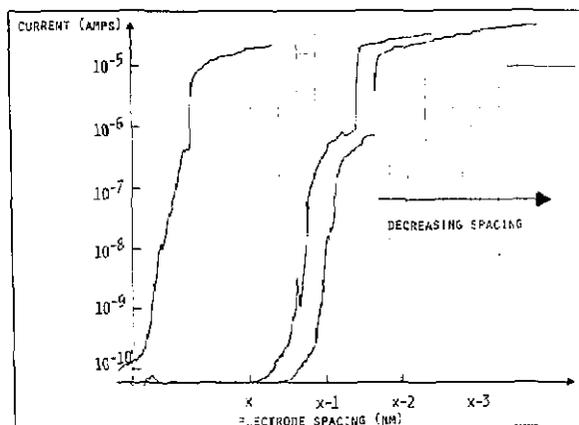


Figure 33—Measured log current-spacing characteristics, bias voltage=10 mV.

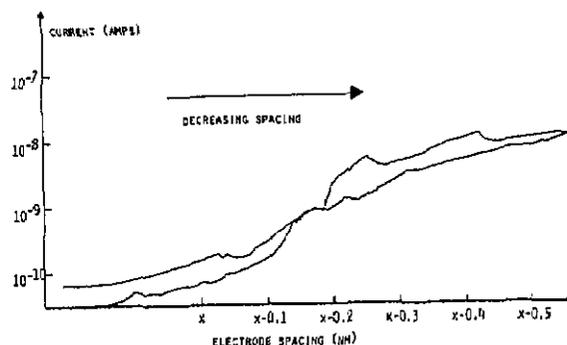


Figure 34—Measured log current-spacing characteristics; expanded spacing scale, bias voltage=10 mV.

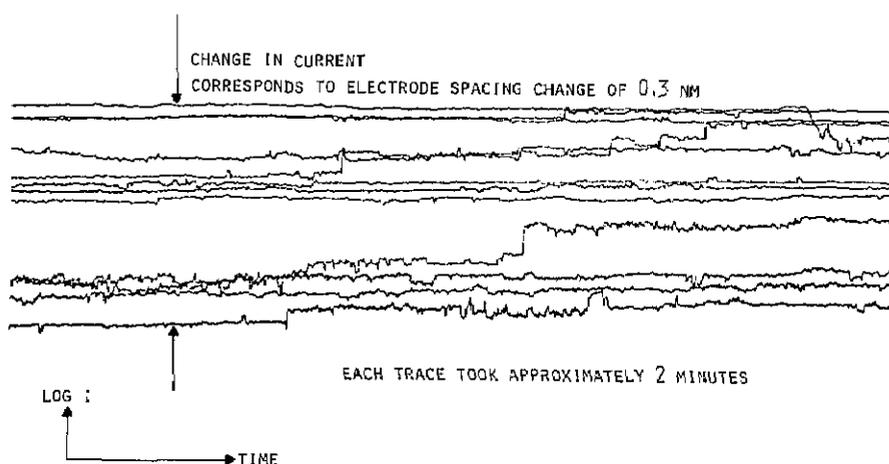


Figure 35—Log current vs. time for a fixed bias voltage of 10 mV.

function with the use of eq (2-70). By taking the ratio at two different spacings, one obtains

$$(\ln I_1 - \ln I_2) / (W_1 - W_2) \cong -a B^{1/2}$$

where $a = 10.25 \text{ (eV)}^{-1/2} \text{ (nm)}^{-1}$, W_i = barrier widths at the Fermi level, and B = average barrier height. The equation only serves to qualitatively illustrate the method relating the slope to the WF since B is a function of WF . The exact relation between the slope of $\text{Log } I_{cv} - S$ data and the WF can be obtained from the calculations of section 2.⁸ The sloping dashed line in figure 32 was drawn with the use of tables 1 and 2 and figure 13. Based on the discussion of the changes in WF with adsorbed hydrocarbons, the WF value of 3.0 eV is very reasonable. Note that this WF value was obtained without the use of any assumed property of the device or the electrode structure. The only input information was the piezoelectric element expansion coefficient and measured currents. This is the first time that such measurements have been performed. All previous measurements employed separate tunneling devices with thickness distributed over the values allowed by the fabrication process. Determinations of the tunneling barrier thickness were also made from capacitance measurements and therefore were not independent of the device characteristics.

Current-voltage measurements of the tunneling device were made during the second experiment, i.e., condition B. The results for an approximate

⁸A recent examination of many of the topics discussed in section 2 and of the validity of measuring work functions from the slope of the $\ln I_{cv}$ versus electrode spacing data has been performed by M. C. Payne and J. C. Inkson, Measurement of Work Functions by Tunneling and the Effect of the Image Potential, *Surface Science* **159** (1985) 485. In addition to the $\ln I_{cv}$ versus electrode spacing calculations given in section 2, they also use a formula derived by Simmons to compute the log of the tunneling voltage as a function of the electrode spacing for a constant tunnel current density.

electrode spacing of 1.2 nm are shown in figure 36. The approximate spacing of the S curve was obtained by determining the PZ element voltage at which touching occurred, assumed to be at the first major drop in slope of the $\text{Log } I_{cv} - S$ curve, then contracting the PZ element the desired distance. The other curves were obtained by changing the applied PZ voltages by the amount to generate 0.1 nm changes in electrode spacing, i.e., 0.2 V.

Many attempts were made to obtain a continuous series of $I-V$ graphs with electrode spacings varying from 2 nm to that when the electrodes made contact. In all attempts, the state of the electrodes did not remain sufficiently stable for the measurements to be performed. A series of measurements at two to three electrode spacings such as shown in figure 36 was obtained four times.⁹

Relating the $I-V$ data to the theoretical calculations of section 2 requires an estimate of the tunneling area and the absolute electrode spacing. (The deduction of WF 's from the $\text{Log } I_{cv} - S$ data only required knowledge of changes in electrode spacing.) Ideally a consistent approach to obtaining both the area and the absolute spacing would be first to deduce an absolute spacing with the use of eq (2-73) which relates the curvature of the $I-V$

⁹The author is not aware of any other published results which show the variation of metal-vacuum-metal tunneling $I-V$ curves with small changes in electrode spacing such as is shown in figure 36. The data given by J. Moreland, S. Alexander, M. Cox, R. Sonnenfeld, and P. K. Hansma (Squeezable Electron Tunneling Junctions, *Appl. Phys. Lett.* **43** (1983) 387) for silver-vacuum-silver tunneling exhibit similar properties in that their curves have an ohmic region near zero bias with non-linear behavior at larger bias voltages i.e., greater than about 0.2 V. $I-V$ data for structures with lead electrodes are also given which demonstrate behavior consistent with superconducting electrodes. $I-V$ data is also given by S. A. Elrod, A. L. de Lozanne, and C. F. Quate for small bias voltages applied to a $W/\text{vacuum}/\text{Nb}_3\text{Sn}$ structure, in the paper Low-temperature Vacuum Tunneling Microscopy, *Appl. Phys. Lett.* **45** (1984) 1240.

Figure 36a—Measured current-voltage characteristics for three electrode spacings. S was approximately 1.2 nm. The S , $S-0.1$, and $S-0.3$ nm curves were drawn manually from the A, B, and C curves of the "raw data" of figure 36b, respectively. The uncertainty in S is much larger than that of the difference between the indicated spacings. The indicated noise level was a function of voltage, increasing with increasing voltage. The source of the noise was apparently changing surface condition.

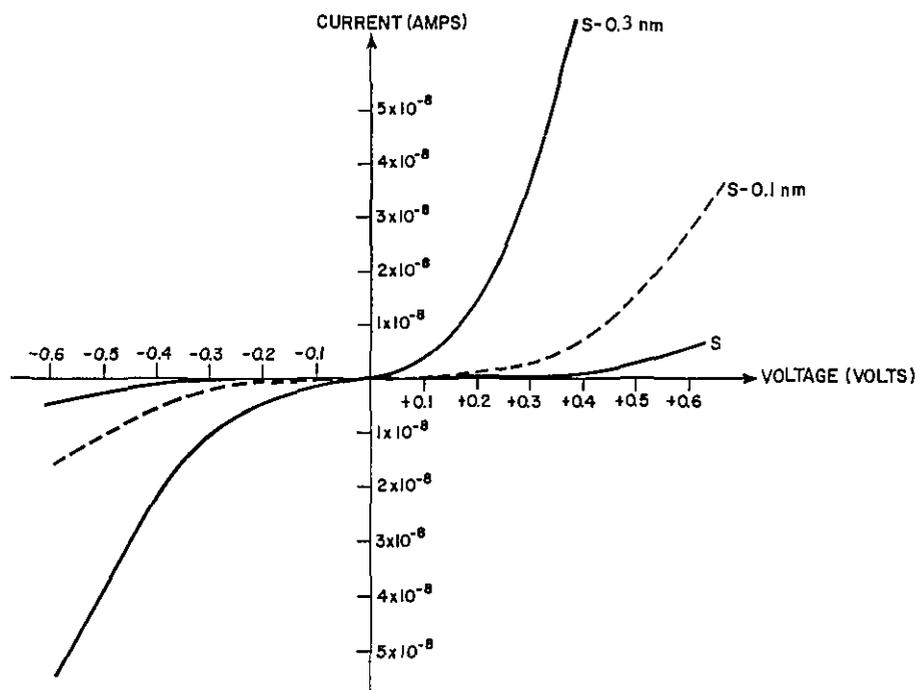
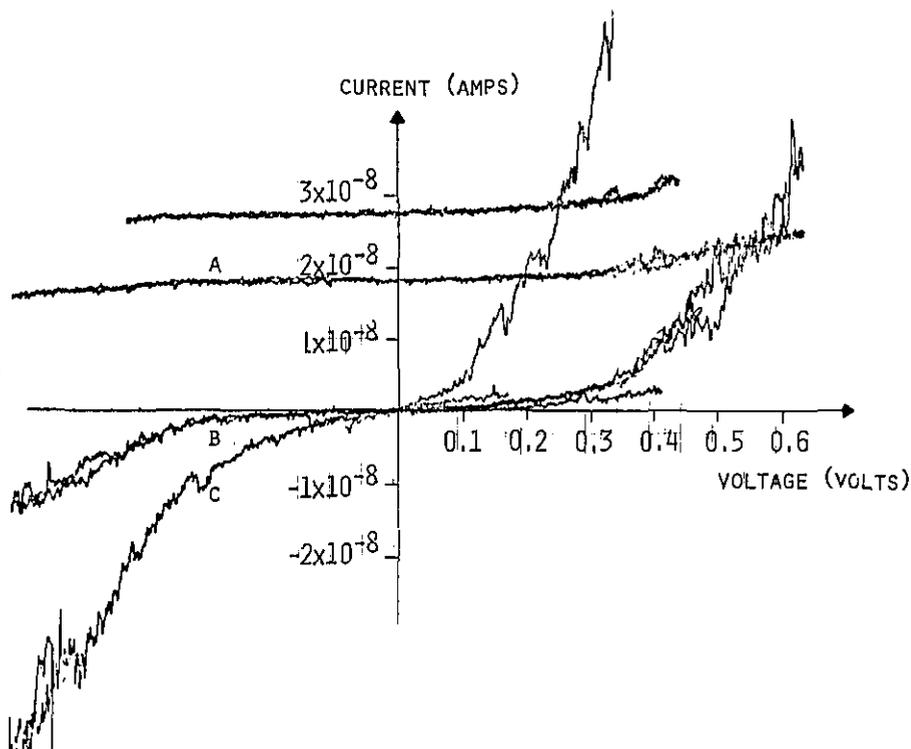


Figure 36b—Measured current-voltage characteristics for three electrode spacings. Curves A and B were measured with electrode spacings differing by 0.1 nm; curve C differs in spacing by 0.2 nm from curve B.



characteristics to the barrier height and spacing. Since the barrier height can be obtained from the slope of the $\text{Log } I_{cv} - S$ data, the absolute spacing would then be the only undetermined parameter. Once it is obtained, the current density calculations could then in combination with the corresponding

$\text{Log } I_{cv} - S$ data and/or the $I - V$ data be utilized to calculate an average tunneling conduction area.

Let $G(V) = dI(V)/dV$, then carrying out this procedure for the $S-0.1$ nm curve of figure 36 gives $G(+0.5)/G(0) = 21.42$ and $G(-0.5)/G(0) = 12.23$. The corresponding S parameters are $S_+^2 = 25.13 B \text{ nm}^2$ and $S_-^2 = 15.0 B \text{ nm}^2$. With the B values calculated from the $\text{Log } I_{cv} - S$ data in figure 32, both values are greater than 6 nm. Based on

the procedure outlined earlier for obtaining these data, S should have been less than 2 nm. A possible explanation of this discrepancy in terms of mechanically generated strains is given in the next section.

Since this method for obtaining an absolute spacing was not reliable, an alternative means to calculate the area and spacing was tried. The second approach was to check for correspondence between the area calculated from the ratio of theoretical current and current density calculations of section 2 and the area estimated from the Log $I_{cv}-S$ data with the use of table 1 or eq (2-70). From the results given in table 1 a conduction area of approximately $2.4 \times 10^{-13} \text{ m}^2$, with a slight variation with work function and electrode spacing, is obtained. If a spacing parameter S of 1.2 nm, a measured current of 10^{-9} A , and a work function of 3.0 eV are assumed, then one may estimate a tunneling conduction area of approximately $5 \times 10^{-16} \text{ m}^2$ by extrapolating the values given in table 1. This area corresponds to a region of 20 nm radius or about $1/20^{\text{th}}$ that of the radius expected for perfect spheres. Such an area is possible and could be produced by microscopic projections from the spherical surfaces.

However, a second possible hypothesis to explain this difference between measured and predicted currents is that the assumed point of contact occurred when S was not zero. Several processes discussed in section 2 would produce this effect. First, the transition from equal Fermi energies to equal bulk electrostatic potentials, even though a small effect for the almost identical electrodes of this equipment, could produce a change in slope of the Log $I_{cv}-S$ curve that would be interpreted as contact. Second, adsorbed or chemisorbed molecules would produce a nonzero S at contact. Third, a change in slope at some value of S will be produced by the onset of constriction resistance [7] dominated current flow. For this conduction process the electrical resistance, R , between two large conductors connected by a small conduction area of radius, a , is given by $R = \rho / (4a)$, where ρ is the conductors' resistivity. Gold's resistivity is $2.3 \times 10^{-8} \Omega\text{-m}$. Thus with the area resulting from the first hypothesis, the expected conduction resistance would be approximately 0.3 Ω . The measured resistance at "contact" was, however, approximately $10^4 \Omega$. Constriction resistance for the conditions encountered in this experiment, therefore, does not appear to be a current limiting factor.

No consistent interpretation of the data given in figures 32, 33, and 36 has been found which gives an absolute electrode spacing and a tunneling conduction area. Future experiments will be required

to resolve the unknown relationships between the barrier parameter S and the true physical spacing, and the difficulties of unknown conduction area.

Somewhat as a last resort the $I-V$ characteristics of the tunneling device were explored at large electrode spacings. Figure 37 shows the breakdown behavior which was typically observed for voltages greater than about 1.0 V, even with relatively large electrode spacings.

Effects of Mechanical Stress on Tunneling Characteristics

There are two likely sources of attractive forces between the tunneling electrodes which could, by producing surface deflections, affect both the Log I vs. S and the I vs. V characteristics. These are the van der Waals type forces between two microscopic bodies and the force arising from the electrostatic field produced by the applied bias voltage.

The van der Waals type forces between two microscopic bodies spaced a distance D apart in a vacuum may be calculated from a general microscopic theory of dispersion forces proposed by Lifshitz [68,69]. When D is less than the smallest characteristic absorption wavelength of the material of the bodies, the force per unit area, S_{VDW} , between two planar bodies is given by

$$S_{VDW} = \hbar\omega / (8 \pi^2 D^3) \quad (4-1)$$

$\hbar\omega$ is known as the Lifshitz-van der Waals interaction constant. By using optical data to deduce the dielectric functions for metals and the Lifshitz theory, Krupp, Schnabel, and Walter [70] computed $\hbar\omega$ for the case when both bodies were gold; $\hbar\omega$ was found to be approximately $1.9 \times 10^{-18} \text{ J}$. With

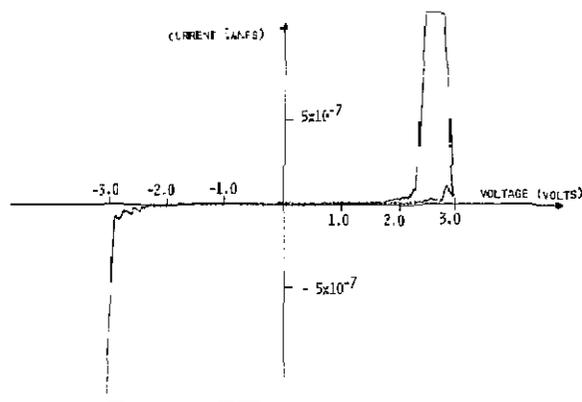


Figure 37—Current-voltage characteristic for electrode spacing of approximately 3.0 nm. The breakdown behavior was typical of all attempts to obtain $I-V$ measurements for bias voltages greater than about 1.0 V.

D expressed in nm, S_{VDW} between gold planes in vacuum is therefore given by

$$S_{VDW}^{Au} = 2 \times 10^7 / D^3 \text{ N nm}^3 / \text{m}^2. \quad (4-2)$$

The stress, T , produced by an electric field, E , perpendicular to a metal surface is [4]

$$T = 1/2 \epsilon_0 E^2, \quad (4-3)$$

$\epsilon_0 = 8.85 \times 10^{-12} \text{ FD} / \text{m}^2$. For a voltage V and electrode spacing D (in nm), T is therefore given by

$$T = 4.42 \times 10^6 (V/D)^2 \text{ N nm}^2 / (\text{V}^2 \text{ m}^2). \quad (4-4)$$

Lifshitz's theory is based on the existence of a fluctuating electromagnetic field in the region between the two bodies and the resultant component of the Maxwell stress tensor. Thus, both types of these forces may be viewed as being the static and dynamic manifestation of the same force.

A complete treatment of the strains generated in the gold electrodes by these stresses would be an extremely difficult problem. One would need to account for the nonplanarity of the two electrodes in terms of its effect on the van der Waals forces and the effect of the electric field not being parallel to the axis of the electrodes. However, since the radii of the effective area is small compared to the sphere radius, a reasonable approximation is that all the forces act parallel to the axis of the two spherical electrodes. Further assume that the equation derived by Timoshenko and Goodier [72] for the deflection, $W(r, \Theta)$, at a point (r, Θ) on a boundary plane of a semi-infinite solid, produced by a stress, $q(r', \Theta')$, over the area $r' dr' d\Theta'$ holds for the surfaces of the spherical electrodes. Timoshenko and Goodier's relation between W and q is

$$W(r, \Theta) = [(1-p^2)/(\pi Y)] q(r', \Theta') dr' d\Theta', \quad (4-5)$$

p is Poisson's ratio for the solid and Y is the elastic modulus. Equation (4-5) is based on the assumption that the solid's elastic properties are isotropic.

Let D_0 be the minimum spacing between the electrodes assuming no mechanical stresses were present. Then, for the small deviations from the axis of the electrodes, the spacing, D , at a distance r from the axis is given by $D = D_0 + r^2/R$; R is the radius of the electrodes. With the assumption about the force directions made earlier, the maximum deflection will occur on the axis. Here eq (4-5) gives

$$W = [2(1-p^2)/Y] \int_0^{R_{\max}} q(r') dr'. \quad (4-6)$$

Substituting the sum of the stresses described by eqs (4-2 and 4-4) for q into eq (4-6) gives

$$W = a \int_0^{R_{\max}} [b D(r')^{-3} + CV^2 D(r')^{-2}] dr', \quad (4-7)$$

$$a = 2(1-p^2)/Y, \quad b = 2 \times 10^7 \text{ N nm}^3 / \text{m}^2, \quad \text{and} \\ C = 4.42 \times 10^6 \text{ N nm}^2 / \text{V}^2 \text{ m}^2.$$

Let $y = D(r')/D_0$ then $dr = 1/2 (D_0 R)^{1/2}$

$(y-1)^{-1/2} dy$ and,

$$W(D_0) = a/2 (R/D_0^3)^{1/2} \int_1^{y_{\max}} [(b/D_0) y^{-3} \\ (y-1)^{-1/2} + CV^2 y^{-2} (y-1)^{-1/2}] dy. \quad (4-8)$$

Both parts of the integrand are standard forms. The integral evaluates to:

$$W(D_0) = (a/2) (R/D_0^3)^{1/2} [b(y-1)^{1/2} / (2D_0 y^2) \\ + (3b/4D_0 + CV^2)(y-1)^{1/2} / y \\ + \tan^{-1} (y-1)^{1/2}]_{y_{\max}}. \quad (4-9)$$

Let y_{\max} go to infinity, then

$$W(D_0) = (a/2) (R/D_0^3)^{1/2} \pi/2 [(3b/4D_0) \\ + CV^2]. \quad (4-10)$$

Y for Gold is $8.3 \times 10^9 \text{ N/m}^2$, $p = 0.33$ and $R = 1.5 \times 10^6 \text{ nm}$; with these values, eq (4-10) reduces to

$$W(D_0) = 2.1 \times 10^{-7} D_0^{-3/2} \\ (3b/4D_0 + CV^2) \text{ nm}. \quad (4-11)$$

For the voltages used to obtain the Log I_{cv} vs. S data, 10 and 100 mv, the dominant term is that from the van der Waals force. If D_0 is 1 nm, then $W = 3.1 \text{ nm}$; if $D_0 = 2 \text{ nm}$, then $W = 0.56 \text{ nm}$. Taking into account the relatively gradual and controlled behavior of most of the Log I_{cv} vs. S data, the "avalanche" magnitudes obtained from this calculation are very puzzling. The history of the electrodes offers a possible explanation. During the process to obtain the first sets of data for Log I_{cv} vs. S , typified by figure 31, an abrupt transition such as one would expect from the predictions of eq (4-11) was observed. However, after the accidental and purposeful electrode cleaning by capacitive discharges, the abrupt or almost step function behavior was no longer observed, e.g., figure 32. If the capacitive discharges produced localized regions of smaller curvature, the magnitude of the

deflections, from both type forces, would be greatly reduced, i.e., if $R = 10^3$ nm, then W would be reduced by a factor of approximately 35 from that calculated from eq (4-11).

The density of electromagnetic modes for the spherical surfaces would be slightly different from that of the flat planes used by Lifshitz. Other microscopic geometrical deviations such as plateaus raised above a relatively smooth background would also affect the mode density. However, the geometry in both cases so closely approximates that of two planes over the areas which would contribute, i.e., about 10 μm in radius, that the changes in the mode densities should not affect the result given in eq (4-11).

The effects predicted by the voltage dependent part support the previous proposal that the anomalously large curvature in the I vs. V data for a fixed spacing was due to mechanical stresses. The difference between the theoretically calculated value of I for a particular voltage and spacing, e.g., figures 11 and 12, and that measured, e.g., figure 36 would correspond approximately to a voltage produced deflection of 0.08 nm for voltages of 0.35 V. These values were obtained from the $s = 0.3$ nm curve of figure 36 as estimates of deviations from a linear I - V characteristic. The change in W between $V = 0$ and $V = 0.35$ V with $D_0 = 1.5$ nm as obtained from eq (4-11) is 0.06 nm.¹⁰

Conclusions

Albeit crudely, an experiment has been performed which demonstrates quantum mechanical tunneling of electrons through a vacuum region of continuously variable thickness. Within experimental error the results indicate that an image potential reduced barrier with the modifications suggested by Lang and Kohn gives a close approximation to the true barrier for metal-vacuum-metal tunneling. For the first time, the work function for the electrodes in a tunneling experiment has been deduced from experimental parameters independent of the tunneling device.

The device which was developed for this experiment demonstrates that vacuum tunneling experiments are feasible and can reliably be performed.

¹⁰Further exploration and utilization of the effects of van der Waals forces has recently been initiated to understand the operation of the scanning tunneling microscope (J. H. Coombs and J. B. Pethica, IBM Europe Institute, STM Workshop, Oberlech, Austria, July 1-5, 1985, IBM J. Res. and Dev. (to be published)) and to enable the investigation of insulator surfaces (G. Binnig, C. F. Quate, and Ch. Gerber, The Atomic Force Microscope, Phys. Rev. Lett. 56 930 (March 1986).

With this demonstration many new areas of research may be explored. A particularly direct application of the vacuum tunneling technique is that of locating the true position of a surface to a precision of 0.1 nm, 20 to 100 times greater than presently available techniques. Without the limitation to electrode materials which form oxides that are good barriers, vacuum tunneling experiments will be a very versatile tool for performing inelastic electron tunneling spectroscopy.

Some contributions toward accomplishing the objectives given in section 1 have therefore been made, but much remains to be done. The problems which have been discovered, e.g., strains generated by field generated stresses and the unknown relationship between physical and theoretical spacing parameters, will need resolution before further work can be performed. However, the work here does give direction for solving the problems and once they are solved the full potential of vacuum tunneling experiments can be realized.

Acknowledgments

The material that has been presented here is the result of many contributions by colleagues at the National Bureau of Standards and North Texas State University. I particularly wish to express my thanks to: James Mackey for his consistent encouragement and supportive friendship from the time he agreed to act as my major professor up to the completion of the dissertation almost eight years later; Russell Young for letting me share in the exploration of his new world of vacuum tunneling; Russell Young, John Simpson, and Karl Kessler for providing the financial and management support at NBS; Fredric Scire for his creative assistance with much of the mechanical design and equipment construction; Charles Reeve for assistance with much of the computing required for the theoretical calculations; Charles Summers for his skilled machining of the apparatus; Dwight Maxson for his design and construction of much of the early electronics, Tom Leedy for assisting with the attempt to fabricate a tunneling device with photolithographic techniques, Robert Hocken for sharing his knowledge and designs of temperature control systems, the late Henry Zoranski for the artwork depicting the key components of the apparatus, Robert Dragoset for a very helpful and thorough reading of this manuscript; and Vita Gagne and Sally Morris for preparation of the final manuscript for this publication.

The author gratefully thanks his wife, Ellen, for her many years of support during the period to complete this work and for her valuable assistance in preparing the original manuscript.

Appendices

Appendix I. NUMERICAL CALCULATIONS OF TUNNELING CURRENT

The computer programs, all in FORTRAN, which were used to calculate the data for the graphs and tables given in section 2, are contained in this appendix. The programs were executed on the National Bureau of Standards 1108A Computing Facility. The main programs of these computations are CURDEN, to calculate the tunneling current density; CURNT, to calculate the tunneling current between spherical electrodes and a program to calculate the tunneling current density according to Simmins' theory.

Aside from the problems of roundoff and overflow errors, the major difficulty which had to be overcome was long computational times. In an earlier version of these programs, the Newton-Raphson method was used to find the classical turning points for the required energy values and a standard iterative Simpson's 1/3 rule integration was used for all integrations. With these algorithms, the computational time for one current density value ranged from 20 to 30 seconds, with the longer times corresponding to larger electrode spacings.

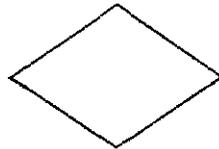
The process of determining the turning points is deeply nested within the chain of integrals to calculate the current and current density. Its computational time is therefore a major component of the overall computational time.

By changing to the indicated tanh rule for integrating over $K(x)$ and using an explicit solution for the roots of a cubic equation, the computational times were cut by a factor of 10 to 15 [73]. The iterative method of increasing the number of partitions until the difference between the result for n and $2n$ partitions was less than a specified error was also dropped. After determining what number was required for the needed accuracy, a fixed number of 64 or 128 was used in all integrations. The particular number is specified in the programs involved.

With the programs as indicated, the time to compute one current density value was two seconds;

the time for one current value was 15 seconds. The resulting values, within the limitations of the physical model, should be accurate to five and four significant digits, respectively, since doubling the number of partitions did not change the least significant digit.

The fundamental data unit of the 1108A central processing unit is a 36-bit word. Average processing rate is about 900,000 instructions per second; 0.75 microseconds for integer addition to 17.25 microseconds for double precision floating-point division. Single precision floating-point numbers carried about 8 significant decimal digits (27 bits) in the range: 2×10^{-39} to $1 \times 10^{+38}$. Double precision floating-point numbers carried about 18 significant decimal digits (60 bits) in the range: 3×10^{-309} to $8 \times 10^{+307}$.



```

PROGRAM SIMMON
*****
C COMPUTE APPROXIMATE VALUES OF CURRENT DENSITY
C ACCORDING TO SIMMONS THEORY. SEE THE FOLLOWING REFERENCE
C BURSTEIN ELIAS AND LUNDQVIST STIG TUNNELING PHENOMENA
C IN SOLIDS CH. 13 BY J. G. SIMMONS
C EQUATION 11, P. 141 HAS BEEN CORRECTED TO READ AS FOLLOWS
*****
S2 = S1-(23/KS)/(3WF+(10/KS)-2V)+S1
*****
NOTE THAT FOR K = 1, VACUUM BARRIER, W = S2 - S1 RESULTS.
VBI IS EQN 21 AND JO IS EQN 19 IN THE REFERENCE CITED.
*****
REAL JO, JT
DIMENSION J(11)
W(S,V,WF) = S*(1-.23/(3*WF*S+10.-2*V*S))
VB(S,V,WF,S1,S2) = WF*(V(2*S)+S1*S2)
I = (2.86/(S2-S1))*ALOG((S2*S1)/(S1*(S-S2)))
DATA WFA,C/5.0,-1.025,6.2E14/
OPEN6,FILE='CURDUT',
DO 10 I = 3,30
S = I
J(I) = S/10.
DO 20 J = 2,11
V = (J-1)/20.
S1 = J*WF
S2 = W(S,V,WF)-S1
W1 = W(S,V,WF)
V1 = VB(S,V,WF,S1,S2)
V2 = V1 + V
JO = C*(V1*EXP(A*W1*SQRT(V1)))/(W1**2)
J0 = J0 - C*(V2*EXP(A*W1*SQRT(V2)))/(W1**2)
JT(I) = J0*(1.43E-9)/(W1**300)**2/(V1)
CONTINUE
WRITE(6,S00) J(I),J(I)
FORMAT(12X,F4.1,1X,10ES3)
CONTINUE
STOP
END
    
```

Listing of CURDEN

```

C PROGRAM TUNNEL CURRENT DENSITY (CURDEN)
C
C WRITTEN BY E. CLAYTON TEAGUE MARCH 1978
C
C SUBPROGRAMS CALLED: RSM13, PROD
C
    
```

```

C THIS IS A PROGRAM TO CALCULATE THE TUNNELING CURRENT AS A
C FUNCTION OF ELECTRODE SPACING AND APPLIED VOLTAGE. THE INSUL-
C ATING BARRIER IS ASSUMED TO BE VACUUM AS MODIFIED BY THE IMAGE
C POTENTIAL. THE CALCULATIONS ARE BASED ON A FREE ELECTRON MODEL
C FOR THE ELECTRODES WITH A FREE ELECTRON MASS IN BOTH ELECTRODES
C AND IN VACUUM. A SYMMETRICAL STRUCTURE WITH IDENTICAL
C ELECTRODES IS ASSUMED.
C
C THE BASIC TUNNELING EQUATIONS ASSUMING A WKB SOLUTION
C TO THE SCHRÖDINGER EQUATION ARE:
C (C.B. DUKE, TUNNELING IN SOLIDS, ACADEMIC PRESS, NEW YORK, 1969,
C PP. 49-53)
C CURRENT DENSITY(V, D, T,) = INTEGRAL(0 TO INF)[DEX* D(EX, V, D, )*
C N(EX, V, T)]
C
C WHERE:
C
C N(EX, V, T) = (4[PI]MEK/H**3) * LN(XPOBT)
C
C WHERE XPOBT = 1 + EXP[(FE-EX)/KT]/[1 +EXP[(FE - EX - EV)/KT]]
C
C 4[PI]MEK/H**3 = 1.385E10 AMPS/(EV,K)
C
C D(EX, V, D) = (1 + EXP[-J(EX, V, D)])** -1
C
C WHERE J(EX, V, D) = INT [X1(EX, V) TO X2(EX, V)]
C OP [4(PI)(2K)**0.5]/H1*[U(X)- EX]**0.5
C
C THE BOTTOM OF THE CONDUCTION BAND OF THE "LEFT"
C ELECTRODE IS ASSUMED TO BE E = 0 . A VOLTAGE V IS APPLIED TO
C LOWER THE ELECTRON ENERGIES OF THE "RIGHT" ELECTRODE.
C THUS THE TOTAL POTENTIAL ENERGY FUNCTION FOR THE BARRIER IS
C GIVEN BY:
C
C U(X) = 0, X LESS THAN OR EQUAL TO 0
C U(X) = BARRIER HEIGHT - ELECTRIC POTENTIAL -IMAGE POTENTIAL
C 0 < X < D
C U(X) = - EV X GREATER THAN OR EQUAL TO D.
C
C THE BARRIER IS THEREFORE GIVEN BY:
C
C U(X) = FE + WE - (EV/D)X -
C (S**2) D * LN 2 * 1.15 / ((8 PI * EPPO * X(D-X))** -1)
C
C IF D IS IN NM AND V, FE, AND WF ARE IN VOLTS,
    
```

Journal of Research of the National Bureau of Standards

```

C THIS REDUCES TO:
C
C U(X) = FE + WF - (V/D) X - 0.2867 D/ (X[D-X]).
C
C THUS FOR CALCULATION ONE MUST FIRST DETERMINE FOR EACH VOLTAGE
C THE CLASSICAL TURNING POINTS X1(EX, V) AND X2(EX, V) FOR EACH
C RELAVEANT ENERGY BETWEEN 0 AND FE + WF.
C
C THE TUNNELING PROBABILITY IS THEN CALCULATED AS A FUNCTION
C OF ENERGY, VOLTAGE, AND BARRIER THICKNESS. AFTER
C CALCULATING THE SUPPLY FUNCTION "N" AS A FUNCTION
C OF ENERGY, VOLTAGE, AND TEMPERATURE, THE CURRENT DENSITY FLOW
C IS FINALLY CALCULATED AS A FUNCTION OF THE VOLTAGE
C BARRIER THICKNESS, AND TEMPERATURE.
C
C THE USER IS REQUESTED TO SUPPLY THE TEMPERATURE OF THE
C TUNNELING DEVICE, AND THE WORK FUNCTION AND FERMI ENERGY
C OF THE ELECTRODES.
C
C REAL NTGL1,NTGL2,NTGL3
C COMMON T,FE,WF,V,D,FK
C EXTERNAL PROD
C DATA NP /128/
C
C NP2=NP/2
C
C GET TEMPERATURE
C
C WRITE (6,90)
C READ (5,80) T
C
C GET FERMI ENERGY
C
C WRITE (6,100)
C READ (5,80) FE
C

```

Listing of DNSITY

```

C FUNCTION DNSITY (Z)
C
C WRITTEN BY E. CLAYTON TEAGUE MARCH 1978
C
C SEE ANNOTATIONS IN CURRENT DENSITY PROGRAM.
C
C SUBPROGRAMS CALLED: RSIM13, PROD
C
C REAL NTGL1,NTGL2,NTGL3
C
C COMMON T,FE,WF,V,Z1,FK
C EXTERNAL PROD
C DATA NP /64/
C
C NP2=NP/2
C Z1=Z
C BLZCTT=8.6E-5*T
C DLTA=40.*BLZCTT+V
C ETOP=FE+WF-1.1468/Z
C
C SET MAXIMUM ENERGY BARRIER HEIGHT AS WF + DLTA
C
C A=ETOP-WF-DLTA
C
C IF (A.LT.0.) A=0
C B=FE-DLTA
C FK=1.
C
C NTGL1=RSIM13(PROD,A,B,NP2)
C
C A=FE-DLTA
C DIFF=ETOP-(FE+80.*BLZCTT)
C B=ETOP
C
C IF (DIFF.GT.0.1) B=FE+80.*BLZCTT
C FK=2.
C
C NTGL2=RSIM13(PROD,A,B,NP)
C
C IF (DIFF.GT.0.1) NTGL3=0.
C IF (DIFF.GT.0.1) GO TO 10
C A=ETOP
C B=FE+80.*BLZCTT
C FK=3.
C
C NTGL3=RSIM13(PROD,A,B,NP2)
C
C 10 DNSITY=NTGL1+NTGL2+NTGL3
C
C RETURN
C END

```

Listing of PROD

```

C FUNCTION PROD (EM)
C
C WRITTEN BY E. CLAYTON TEAGUE APRIL 1978
C
C SUBPROGRAMS CALLED: RNTOP, TUNNPB
C
C COMMON T,FE,WF,V,D,FK
C
C K=FK
C GO TO (10,20,30), K
C
C NOTE THAT T IN CONSTANT OF RNTOP WAS CANCELLED BY T IN EV/KT.
C
C 10 PROD=(1.385E10)*(V/8.6E-5)*TUNNPB(EM)
C RETURN

```

```

C
C 20 PROD=RNTOP(EM)*TUNNPB(EM)
C RETURN
C
C 30 PROD=RNTOP(EM)*(1+(D/2.2936)*(EM-FE-WF))
C
C RETURN
C END

```

Listing of RNTOP

```

C FUNCTION RNTOP (EM)
C
C WRITTEN BY E. CLAYTON TEAGUE APRIL 1978
C
C DOUBLE PRECISION FERMIL,FERMIR,ALPHA,BETA,BLZCTT
C COMMON T,FE,WF,V
C
C BLZCTT=8.6D-5*DBLE(T)
C ALPHA=DEXP(DBLE(FE-EM)/BLZCTT)
C BETA=DEXP(DBLE(FE-EM-V)/BLZCTT)
C
C AVOID ROUND OFF ERROR IN DOUBLE PRECISION BY USING APPROXIMATION
C THAT LN[(1+A)/(1+B)] = A-B WHEN A AND B ARE MUCH LESS THAN 1.
C
C IF (ALPHA.LT.1D-6) GO TO 10
C
C AVOID OVERFLOW ERRORS BY USING APPROXIMATION THAT
C LN[(1+EXP(A))/(1+EXP(A+B))] = -B WHEN
C A IS MUCH GREATER THAN 1.
C
C IF (BETA.GT.1D6) GO TO 20
C
C FERMIL=1.0D0+ALPHA
C FERMIR=1.0D0+BETA
C RNTOP=1.385D10*DBLE(T)*(DLOG(FERMIL)-DLOG(FERMIR))
C RETURN
C
C 10 RNTOP=1.385D10*DBLE(T)*(ALPHA-BETA)
C
C RETURN
C
C 20 RNTOP=1.385D10*DBLE(T)*DBLE(V)/BLZCTT
C
C RETURN; END

```

Listing of TUNNPB

```

C FUNCTION TUNNPB (EM)
C
C WRITTEN BY E. CLAYTON TEAGUE APRIL 1978
C
C SUBPROGRAMS CALLED: DCUBIC, RNIHF, RTPOT
C
C DOUBLE PRECISION A3,A2,A1,A0,Z1R,Z1I,Z2R,Z2I,Z3R,Z3I
C EXTERNAL RTPOT
C COMMON T,FE,WF,V,D,FK
C
C CALCULATE X1(EM,V,D) AND X2(EM,V,D) TURNING POINTS.
C CUBIC COEFFICIENTS CALCULATED FROM POT (IN SUBPROGRAM RTPOT).
C
C A3=V/D
C A2=-FE-WF+EM-V
C A1=(FE+WF-EM)*D
C A0=-.2867*D
C CALL DCUBIC (A3,A2,A1,A0,Z1R,Z1I,Z2R,Z2I,Z3R,Z3I)
C X1EM=Z2R
C X2EM=Z3R
C
C CALCULATE THE TUNNELING PROBABILITY AS A FUNCTION OF ENERGY
C
C IF (X1EM.LT.0.0.OR.X2EM.GT.D) GO TO 10
C
C CALCULATE TUNNELING INTEGRAL
C
C Y=RNIHF(RTPOT,X1EM,X2EM,16)
C
C USE MILLER AND GOOD'S FORM OF WKB APPROXIMATION
C
C TUNNPB=1/(1+EXP(Y))
C
C RETURN
C
C 10 WRITE (6,20) X1EM,X2EM
C TUNNPB=0.5
C RETURN
C
C
C 20 FORMAT (1X,'TURNING POINTS',1X,F8.5,1X,'AND',1X,F8.5,1X,
C 2 'ARE OUT OF BOUNDS --- TUNNPB SET = 0.5')
C
C END

```

Listing of RTPOT

```

C FUNCTION RTPOT (XEM)
C
C WRITTEN BY E. CLAYTON TEAGUE APRIL 1978
C
C COMMON T,FE,WF,V,D,FK,EM
C
C POT=FE+WF-EM-(V/D)*XEM-0.2867*D/(XEM*(D-XEM))
C IF (POT.LT.0.) POT=0.

```

Journal of Research of the National Bureau of Standards

```
C
RTPOT=10.25*SQRT(POT)
C
RETURN
END
```

Listing of RSIM13

```
FUNCTION RSIM13(F,A,B,N)
-----
RSIM13  WRITTEN BY CHARLES P. REEVE, STATISTICAL ENGINEERING
        DIVISION, NATIONAL BUREAU OF STANDARDS
FOR: NUMERICAL INTEGRATION BY SIMPSON'S 1/3 RULE
SUBPROGRAMS CALLED: USER DEFINED FUNCTION F
CURRENT VERSION COMPLETED APRIL 28, 1978
-----
COMMON T,PE,WF,V,D,FK,X
N2=N/2
IF(2*N2.NE.N) GO TO 5
IF(A.GT.B) GO TO 4
N21=N2-1
H=(B-A)/N
H2=2*H
X=A
RSIM13=F(X)
X=B
RSIM13=RSIM13+F(X)
X=A-H
DO 10 I=1,N2
X=X+H2
RSIM13=RSIM13+4.*F(X)
10 CONTINUE
X=A
DO 12 I=1,N21
X=X+H2
RSIM13=RSIM13+2.*F(X)
12 CONTINUE
RSIM13=H*RSIM13/3.
RETURN
4 WRITE(6,601)
601 FORMAT(1X,'***LOWER LIMIT OF INTEGRATION EXCEEDS UPPER LIMIT***/
```

Listing of RSIM13

```
64X,'VALUE OF INTEGRAL SET = 0.')
RSIM13=0.
RETURN
5 WRITE(6,602)
602 FORMAT(1X,'***THE NUMBER OF PARTITIONS MUST BE EVEN***/4X,
&'VALUE OF INTEGRAL SET = 0.')
RSIM13=0.
RETURN
END
```

RSJM13 WAS A DUPLICATE COPY OF RSIM13

Listing of RNIHF

```
FUNCTION RNIHF(F,A,B,N)
-----
RNIHF  WRITTEN BY CHARLES P. REEVE, STATISTICAL ENGINEERING
C GET WORK FUNCTION
C
WRITE(6,110)
READ(5,80) WF
C
GET MINIMUM ELECTRODE SPACING
C
WRITE(6,120)
READ(5,80) IMIN
C
GET MAXIMUM ELECTRODE SPACING
C
WRITE(6,130)
READ(5,80) IMAX
IDIF=IMAX-IMIN
C
GET INCREMENTS OF ELECTRODE SPACING
C
WRITE(6,60)
READ(5,80) NIES
C
IF (IDIF.LT.NIES) IINC=1
IF (IDIF.LT.NIES) GO TO 10
IINC=IDIF/NIES
C
GET MINIMUM VOLTAGE
C
10 WRITE(6,140)
READ(5,80) JMIN
C
GET MAXIMUM VOLTAGE
C
WRITE(6,150)
READ(5,80) JMAX
JDIF=JMAX-JMIN
C
GET INCREMENTS OF VOLTAGE
C
WRITE(6,70)
READ(5,80) NV
```

```
C
IF (JDIF.LT.NV) JINC=1
IF (JDIF.LT.NV) GO TO 20
JINC=JDIF/NV
C
20 WRITE(6,170) T,WF,PE
WRITE(6,180)
C
DO 50 ID=IMIN,IMAX,IINC
D=ID/1000.
C
DO 40 JV=JMIN,JMAX,JINC
V=JV/1000.
BLZCTT=8.6E-5*T
DLTA=40.*BLZCTT+V
C
IF (E-PE) IS LESS THAN 40KT THEN: LN(XPONT) = EV/KT
C
ETOP=PE+WF-1.1468/D
C
ETOP = TOP OF IMAGE POTENTIAL WITH NO BIAS VOLTAGE.
C
SET MAXIMUM TUNNELING BARRIER HEIGHT AS WF + DLTA
C
1ST INTEGRAL COVERS ENERGY RANGE WHERE LN(XPONT) MAY BE
APPROXIMATED BY EV/KT
C
A=ETOP-WF-DLTA
IF (A.LT.0.) A=0
B=PE-DLTA
FK=1.
NTGL1=RSIM13(PROD,A,B,NP2)
C
2ND INTEGRAL COVERS ENERGY RANGE WHERE FULL VARIATION OF
LN(XPONT) AND TUNNPB MUST BE TAKEN INTO ACCOUNT. NOTE THAT
SINCE TUNNPB INCREASES AS E INCREASES UPPER LIMIT IS EXTENDED
TO 80KT
C
A=PE-DLTA
DIFF=ETOP-(PE+80.*BLZCTT)
B=ETOP
IF (DIFF.GT.0.1) B=PE+80.*BLZCTT
FK=2.
NTGL2=RSIM13(PROD,A,B,NP)
C
3RD INTEGRAL COVERS ENERGY RANGE IN WHICH TUNNPB IS ASSUMED
TO VARY LINEARLY WITH ENERGY, AND WHEN ETOP-PE IS LESS THAN
80KT.
```

```
IF (DIFF.GT.0.1) NTGL3=0.
IF (DIFF.GT.0.1) GO TO 30
A=ETOP
B=PE+80.*BLZCTT
FK=3.
NTGL3=RSIM13(PROD,A,B,NP2)
C
30 RJ=NTGL1+NTGL2+NTGL3
WRITE(6,160) D,V,RJ
C
AT THE NATIONAL BUREAU OF STANDARDS COMPUTER FACILITY THE FOLLOWING
STATEMENT WRITES INTO AN AUTOMATICALLY CREATED TEMPORARY FILE CALLED
7. UPON TERMINATION OF THE RUN FILE 7 MAY BE COPIED INTO THE USERS
OWN FILE FOR PERMANENT STORAGE.
```

```
WRITE(7,160) D,V,RJ
C
CONTINUE
STOP
C
60 FORMAT(1X,'HOW MANY INCREMENTS OF ELECTRODE SPACING?')
70 FORMAT(1X,'HOW MANY INCREMENTS OF VOLTAGE?')
80 FORMAT()
90 FORMAT(' ENTER TEMPERATURE IN DEGREES KELVIN (REAL#)')
100 FORMAT(' ENTER FERMI ENERGY IN ELECTRON VOLTS (REAL#)')
110 FORMAT(' ENTER WORKFUNCTION IN ELECTRON VOLTS (REAL#)')
120 FORMAT(' ENTER MINIMUM ELECTRODE SPACING IN PM. (INTEGER #)')
130 FORMAT(' ENTER MAXIMUM ELECTRODE SPACING IN PM. (INTEGER #)')
140 FORMAT(' ENTER MINIMUM VOLTAGE IN MILLIVOLTS. (INTEGER #)')
150 FORMAT(' ENTER MAXIMUM VOLTAGE IN MILLIVOLTS (INTEGER#)')
160 FORMAT(2F8.3,E12.5)
170 FORMAT(1H1,/,/,8X,' TEMPERATURE = ',1X,F5.1,' DEGREES KELVIN.',/,
2 8X,' WORK FUNCTION = ',1X,F4.1,' ELECTRON VOLTS.',/,/,8X,
3 ' FERMI ENERGY = ',1X,F4.1,' ELECTRON VOLTS.',/,/,/,
180 FORMAT(8X,' COLUMN 1 = ELECTRODE SPACING IN NM.',/,/,8X,
2 ' COLUMN 2 = BIAS VOLTAGE IN VOLTS.',/,/,8X,
3 ' COLUMN 3 = CURRENT DENSITY FLOW IN AMPERES PER SQ.M',/,/,/)
C
END
```

Listing of CURNT

```
C PROGRAM TUNNEL CURRENT (CURNT)
C
C WRITTEN BY E. CLAYTON TEAGUE MARCH 1978
C
C SUBPROGRAMS CALLED: RSJM13, DNSITY
C
COMMON T,PE,WF,V
EXTERNAL DNSITY
DATA NPART /16/
C
GET TEMPERATURE
C
WRITE(6,80)
READ(5,70) T
C
GET FERMI ENERGY
```

Journal of Research of the National Bureau of Standards

```

C      WRITE (6,90)
C      READ (5,70) FE
C
C      GET WORK FUNCTION
C
C      WRITE (6,100)
C      READ (5,70) WF
C
C      GET MINIMUM ELECTRODE SPACING
C
C      WRITE (6,110)
C      READ (5,70) IMIN
C
C      GET MAXIMUM ELECTRODE SPACING
C
C      WRITE (6,120)
C      READ (5,70) IMAX
C      IDIF=IMAX-IMIN
C
C      GET INCREMENTS OF ELECTRODE SPACING
C
C      WRITE (6,50)
C      READ (5,70) NIES
C
C      IF (IDIF.LT.NIES) IINC=1
C      IF (IDIF.LT.NIES) GO TO 10
C      IINC=IDIF/NIES
C
C      GET MINIMUM VOLTAGE
C
C      WRITE (6,130)
C      READ (5,70) JMIN
C
C      GET MAXIMUM VOLTAGE
C
C      WRITE (6,140)
C      READ (5,70) JMAX
C      JDIF=JMAX-JMIN
C
C      GET INCREMENTS OF VOLTAGE
C
C      WRITE (6,60)
C      READ (5,70) NV
C
C      IF (JDIF.LT.NV) JINC=1
C      IF (JDIF.LT.NV) GO TO 20
C      JINC=JDIF/NV
C
C      WRITE (6,160) T,WF,FE
C      WRITE (6,170)
C
C      DO 40 ID=IMIN,IMAX,IINC
C          D=ID/1000.
C
C      DO 30 JV=JMIN,JMAX,JINC
C          V=JV/1000.
C
C      INTEGRATE THE CURRENT DENSITY OVER SPHERICAL
C      SURFACES TO OBTAIN THE NET CURRENT FLOW.
C
C          A=D
C          B=D+0.4
C          CURNT=RSJM13(DNSITY,A,B,NPART)
C
C          CURNT=CURNT*3.14159265*1.5E-12
C
C      WRITE (6,150) D,V,CURNT
C
C      AT THE NATIONAL BUREAU OF STANDARDS COMPUTER FACILITY THE FOLLOWING
C      STATEMENT WRITES INTO AN AUTOMATICALLY CREATED TEMPORARY FILE CALLED
C      7. UPON TERMINATION OF THE RUN FILE 7 MAY BE COPIED INTO THE USERS
C      OWN FILE FOR PERMANENT STORAGE.
C
C          WRITE (7,150) D,V,CURNT
C
C      CONTINUE
C
C      CONTINUE
C
C      STOP
C
C      FORMAT (1X,'HOW MANY INCREMENTS OF ELECTRODE SPACING?')
C      FORMAT (1X,'HOW MANY INCREMENTS OF VOLTAGE?')
C      FORMAT ( )
C      FORMAT (' ENTER TEMPERATURE IN DEGREES KELVIN (REAL#)')
C      FORMAT (' ENTER FERMI ENERGY IN ELECTRON VOLTS (REAL#)')
C      FORMAT (' ENTER WORKFUNCTION IN ELECTRON VOLTS (REAL#)')
C      FORMAT (' ENTER MINIMUM ELECTRODE SPACING IN PM. (INTEGER #)')
C      FORMAT (' ENTER MAXIMUM ELECTRODE SPACING IN PM. (INTEGER#)')
C      FORMAT (' ENTER MINIMUM VOLTAGE IN MILLIVOLTS. (INTEGER#)')
C      FORMAT (' ENTER MAXIMUM VOLTAGE IN MILLIVOLTS (INTEGERS#)')
C      FORMAT (2F8.3,E12.5)
C      FORMAT (1H1,/,/,8X,' TEMPERATURE = ',1X,F5.1,' DEGREES KELVIN.',/,/
C      2 8X,' WORK FUNCTION = ',1X,F4.1,' ELECTRON VOLTS.',/,/,8X,
C      3 ' FERMI ENERGY = ',1X,F4.1,' ELECTRON VOLTS.',/,/,/
C      170 FORMAT (8X,' COLUMN 1 = ELECTRODE SPACING IN NM.',/,/,8X,
C      2 ' COLUMN 2 = BIAS VOLTAGE IN VOLTS.',/,/,8X,
C      3 ' COLUMN 3 = CURRENT FLOW IN AMPERES ',/,/,/)
C
C      END
C
C      DIVISION, NATIONAL BUREAU OF STANDARDS
C      FOR: NUMERICAL INTEGRATION USING THE TANH RULE
C      REFERENCE: HABER, SEYMOUR, THE TANH RULE FOR NUMERICAL
C      INTEGRATION, SIAM J. NUMER. ANAL., VOL. 14, NO. 4,
C      SEPTEMBER 1977
C      SUBPROGRAMS CALLED: USER DEFINED FUNCTION F
C      CURRENT VERSION COMPLETED APRIL 27, 1978
C
C      IF(A.GT.B) GO TO 4
C      H=3.141592654/SQRT(2.*N)-1/(2.*N)
C      X=(B+A)/2
C      RNHF=F(X)
C      DO 10 I=1,N
C          HI=I*H
C          Y=((B-A)/2)*TANH(HI)
C          D=COSH(HI)**2
C          RNHF=RNHF+(F(X+Y)+F(X-Y))/D
C      10 CONTINUE
C      RNHF=((B-A)/2)*H*RNHF
C      RETURN
C      4 RNHF=0
C      WRITE(6,600)
C      600 FORMAT(1X,'***LOWER LIMIT OF INTEGRATION EXCEEDS UPPER LIMIT***/
C      64X,'VALUE OF INTEGRAL SET = 0. IN SUBPROGRAM RNHF')
C      RETURN
C      END
C
C      Listing of DCUBIC
C
C      SUBROUTINE DCUBIC(A3,A2,A1,A0,Z1R,Z1I,Z2R,Z2I,Z3R,Z3I)
C
C      DCUBIC WRITTEN BY CHARLES P. REEVE, STATISTICAL ENGINEERING
C      DIVISION, NATIONAL BUREAU OF STANDARDS
C      FOR: COMPUTING THE ROOTS OF THE CUBIC EQUATION
C      A3 * X**3 + A2 * X**2 + A1 * X + A0 = 0
C      WHERE THE COEFFICIENTS A3, A2, A1, A0 ARE
C      DOUBLE PRECISION.
C      THE METHOD OF SOLUTION IS TAKEN FROM PAGE 17 OF
C      "HANDBOOK OF MATHEMATICAL STATISTICS" BY
C      M. ABRAMOWITZ AND I.A. STEGUN, NBS-AMS-55.
C      SUBPROGRAMS CALLED: KJD
C      CURRENT VERSION COMPLETED MAY 3, 1978
C
C      IMPLICIT DOUBLE PRECISION (A-Z)
C      IF(DABS(A3).LE.1D-8) GO TO 4
C      PI=3141592653589793D-15
C      A2=A2/A3
C      A1=A1/A3
C      A0=A0/A3
C      Q=A1/3-(A2/3)**2
C      R=(A1*A2-3*A0)/6-(A2/3)**3
C      T=Q**3+R*R
C      IF(T.LT.0.) GO TO 2
C      T=DSQRT(T)
C      S1C=DCBRT(R+T)
C      S2C=DCBRT(R-T)
C      GO TO 3
C      2 TCR=0D0
C      TCI=DSQRT(-T)
C      CR=R+TCR
C      CI=TCI
C      A=DATAN(CI/CR)+PI*KJD(CI,CR)
C      A=A/3
C      RR=DSQRT(CR*CR+CI*CI)
C      RR=DCBRT(RR)
C      S1R=RR*DCOS(A)
C      S1I=RR*DSIN(A)
C      CR=R-TCR
C      CI=-TCI
C      A=DATAN(CI/CR)+PI*KJD(CI,CR)
C      A=A/3
C      RR=DSQRT(CR*CR+CI*CI)
C      RR=DCBRT(RR)
C      S2R=RR*DCOS(A)
C      S2I=RR*DSIN(A)
C      3 UCR=- (S1R+S2R)/2-A2/3
C      UCI=- (S1I+S2I)/2
C      VCR=- (DSQRT(3D0)/2)*(S1I-S2I)
C      VCI= (DSQRT(3D0)/2)*(S1R-S2R)
C      Z1R=S1R+S2R-A2/3
C      Z1I=S1I+S2I
C      Z2R=UCR+VCR
C      Z2I=UCI+VCI
C      Z3R=UCR-VCR
C      Z3I=UCI-VCI
C      RETURN
C      4 WRITE(6,600)
C      600 FORMAT(1X,'NO SOLUTION -- COEFFICIENT OF X**3 IS LESS THAN 1E-8')
C      RETURN
C      END
C
C      Listing of KJD
C
C      FUNCTION KJD(T,B)
C
C      KJD WRITTEN BY CHARLES P. REEVE, STATISTICAL ENGINEERING
C      DIVISION, NATIONAL BUREAU OF STANDARDS
C      FOR: DETERMINING THE QUADRANT FOR THE INVERSE TANGENT FUNCTION.
C      THE PROPER MULTIPLE OF PI IS ADDED TO THE PRINCIPLE VALUE
C      OF THE INVERSE TANGENT. THE MODE IS DOUBLE PRECISION
C      SUBPROGRAMS CALLED: -NONE-
C      CURRENT VERSION COMPLETED JUNE 8, 1976
C
C      DOUBLE PRECISION T,B
C      IF(T.LT.0) GO TO 1
C      IF(B.LT.0) GO TO 2
C      KJD=0
C      RETURN
C      1 IF(B.GE.0) GO TO 3
C      2 KJD=1
C      RETURN
C      3 KJD=2
C      RETURN
C      END

```

Appendix II. STATISTICAL FLUCTUATIONS IN THE LENGTH OF A BAR

The fluctuations to be calculated here are those arising from the thermodynamic properties of matter and do not account for any additional quantum fluctuation in the length of the bar. At the temperatures used for this experiment, the quantum fluctuations would be negligible. The calculations are based on the development given by Landau and Lifshitz [74]. Starting from the argument that the probability for a quantity x to have a value in the range x to $x + dx$ is proportional to $e^{S(x)/k}$, where $S(x)$ is the entropy of the system involved and k is Boltzmann's constant, they derive the result

$$\overline{(\Delta V)^2} = -kT \left(\frac{\partial V}{\partial P} \right)_T. \quad (\text{II-1})$$

Here $\overline{(\Delta V)^2}$ is the mean fluctuation of the system volume at the temperature T .

This may be related to the properties of a solid by the bulk modulus, B , which is defined by the relation

$$B = -V \left(\frac{\partial V}{\partial P} \right)_T, \quad (\text{II-2})$$

where V is the volume of the solid. Combining eqs (II-1 and II-2) gives the result:

$$\overline{(\Delta V)^2} = kTV/B. \quad (\text{II-3})$$

If one now assumes that a bar of length, l , and cross-sectional area, A , is constrained such that all

fluctuations are manifest in length changes, then the result

$$\overline{(\Delta l)^2} = (kT/B)(l/A), \quad (\text{II-4})$$

is obtained, where $\overline{(\Delta l)^2}$ is the mean square fluctuation in the length of the bar.

Values for the parameters of eq (II-4) which represent the gold electrode leads are

$$T = 300 \text{ K}$$

$$B = 1.0 \times 10^{10} \text{ dynes/cm}^2$$

$$l = 1 \text{ cm}$$

$$A = 10^{-2} \text{ cm}^2$$

Thus, $[\overline{(\Delta l)^2}]^{1/2} = 2 \times 10^{-15} \text{ m}$. Calculated values for the fluctuations of other parts of the electrode assembly, i.e., the aluminum housing and the copper holders, are also close to this value for the gold electrodes. Since this fluctuations in all the parts of the electrode assembly would be uncorrelated, the net results would be the square root of the sum of the squares of the individual components. The resultant mean square fluctuations in the electrode spacing would then be approximately $4 \times 10^{-15} \text{ m}$. Note that these estimates include all frequency components.

Appendix III. VACUUM CONDITIONS TO ACHIEVE AND MAINTAIN A CLEAN SOLID SURFACE

Let T be defined as the time in which an initially clean surface becomes covered with a monolayer of adsorbate when the surface is exposed to a gas at a pressure P Torr, then the purpose of this approximate expression for T is

$$T = 2 \times 10^{-6} / P \text{ seconds.} \quad (\text{III-1})$$

The following derivation assumes that the surface contains 10^{19} adsorption sites/m² and that the sticking coefficient for adsorption is one, i.e., every adsorbate molecule which strikes the surface is adsorbed.

According to the kinetic theory of gases, the number of molecules striking unit area per unit time, Z , is given by

$$Z = N_A P (2\pi M R_g T)^{-1/2}, \quad (\text{III-2})$$

where N_A is Avagadro's number, M is the molecular weight of the gas, R_g is the universal gas con-

stant, and T is the gas temperature. For Nitrogen, N₂, at 300 °K,

$$Z = 2.873 \times 10^{22} P \text{ molecules/(m}^2\text{s)}. \quad (\text{III-3})$$

Converting the units of P from N/m² to Torr gives

$$Z = 3.83 \times 10^{24} P \text{ molecules/(m}^2\text{s)}. \quad (\text{III-4})$$

Time for one monolayer is density of adsorption sites/ Z , so

$$T = (10^{19}/\text{m}^2)/Z = (2.6 \times 10^{-6})/P \text{ seconds.} \quad (\text{III-5})$$

The general form of eq (III-1) for N adsorption sites/unit area is

$$T = CVN(2\pi M R_g T)^{1/2}/(N_A P), \quad (\text{III-6})$$

where C is a proportionality constant determined by the adsorbate's sticking coefficient.

Appendix IV. AN ELECTRONIC DEVICE UTILIZING QUANTUM MECHANICAL TUNNELING

The idea for this electronic device resulted from attempts to design a metal-vacuum-metal tunneling structure with the use of microfabrication techniques commonly employed by the microelectronics industry. Conceptually, the device would combine the properties of piezoelectric ceramics with the high sensitivity of tunneling to changes in electrode spacing to produce an electronic device with a very high transconductance.

Starting with a highly polished piezoelectric substrate which is polarized as a transverse expander, i.e., fields applied normal to substrate plane result in dimensional changes in the substrate plane, the device would be fabricated as shown in figure 38. After applying electrodes to both faces of the piezoelectric substrate, a thick layer of polycrystalline metal or semiconductor, shown as the dotted layer, would be deposited over the anticipated device area. The next step would be to heavily oxidize the metal or semiconductor layer or by other means form an integral insulating layer, shown as the lightly cross-hatched strips. Following this operation, the tunneling electrodes would be formed on top of the insulating layer using standard evaporation and photomasking technology. The final step in the fabrication process would be to remove the insulating layer using a preferential plasma or chemical etching process to leave a structure similar to that shown in figure 38. A gap within the

range of the piezoelectric substrate would be formed by either highly precise photolithography or by melting a narrow bridge into two parts.

Assuming that one has produced a gap between the two electrodes which is shorter than the transverse dimensional capabilities of the piezoelectric material, the device characteristics should have the following qualitative properties. Let the piezoelectric electrodes be A and B and the tunneling electrodes be C and D. Then, if x_1 is the lateral distance between the bases of the tunneling electrodes and d_{31} is the lateral expansion coefficient of the piezoelectric ceramic,

$$dx_1 = (d_{31}x_1 t^{-1}) dV_{AB}, \tag{IV-1}$$

where t is the ceramic substrate's thickness, and dV_{AB} is the change in voltage across the substrate. For a fabricated gap of 500 nm, the voltage V_{AB}^0 to place the electrodes within tunneling distances for available piezoelectric ceramics would be approximately 50 V for a 0.1 mm thickness.

Considering the device input to be the voltage applied to the piezoelectric ceramic and the output as the tunneling current I_{CD} for a fixed bias voltage, the transconductance is defined as

$$g = (dI_{CD}/dV_{AB}) V_{CD} = \text{const.} \tag{IV-2}$$

Combining eq (IV-1) and (IV-2) with the rule-of-thumb tunneling equation

$$I_{CD} = C_1 \exp(C_2 SB^{1/2}) A V_{CD}, \tag{IV-3}$$

where $C_1 = 3.16 \times 10^{13} \text{ A/m}^2/\text{V}$, $C_2 = 10.25 \text{ (eV)}^{-1/2} \text{ nm}^{-1}$, S is the electrode spacing, B the mean barrier height and A is the tunneling area, gives

$$g = C_1 C_2 B^{1/2} e^{C_2 SB^{1/2}} d_{31} x_1 t^{-1} A V_{CD}. \tag{IV-4}$$

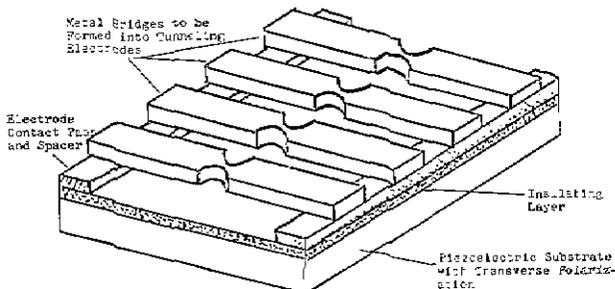


Figure 38—Process to fabricate tunneling devices using photolithographic techniques.

For tunneling areas of 10^{-12} m^2 , $d_{31}=0.25 \text{ nm/V}$, $x_1=4 \text{ mm}$, $t=0.1 \text{ mm}$, $B=1 \text{ ev}$, $S=1 \text{ nm}$, and $V_{CD}=0.1 \text{ V}$.

$$g = 10^{-2} A/V (= \text{mho}). \quad (\text{IV-5})$$

This value is about the maximum transconductance obtainable with conventional electronic devices [75,76]. Increases in tunneling area to 10^{-10} m^2 would raise g to 1 mho.

A photograph of a device resulting from the exploration of this technique for fabricating tunneling structures is shown in figure 39. The approach was not pursued any further, because at the time of the investigation, only limited time was available and the preferential etching process to produce the required electrode length of 2 mm was not sufficiently developed. Shorter lengths place demands on photomasking technology that could not be achieved, i.e., forming lines less than one micrometer in width. However, since the time of these attempted fabrications, April 1976, the capabilities of both technologies have significantly advanced.

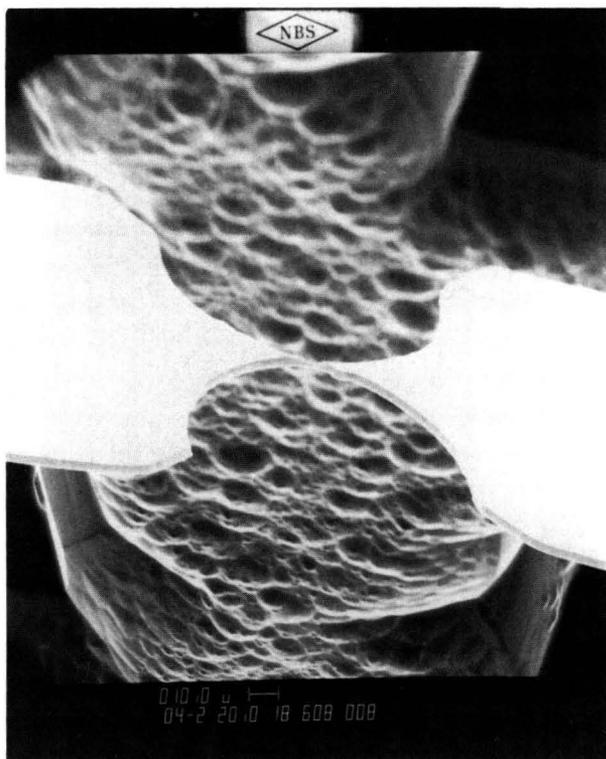


Figure 39—SEM photomicrograph of prototype device fabricated with the use of photolithographic techniques.

Appendix V. ELECTRONICS FOR THE TEMPERATURE CONTROL SYSTEM

The purpose of this appendix is to give a description of the two-stage temperature control system and its associated electronics in terms of the transfer functions of each part of the system, to derive from these transfer functions the response of each stage's control system to setpoint changes and environmental temperature changes, and finally to describe the procedure employed to choose the lead and lag times of the compensation network and to "tune" the network for an acceptable response to setpoint changes. Descriptions of both stage-controllers are given in terms of the block diagram shown in figure 23. The intent here is not to give a detailed analysis of the electronic circuitry, but rather to give a reasonably complete description of the basic properties of the control system in terms of the properties of its subunits. The transfer function of each subunit will now be obtained.

The transfer function of the thermistor with respect to ambient changes in temperature, T , can be obtained from the equation for conservation of energy

$$\dot{Q}_E = C \frac{dT_Z}{dt} + K(T_Z - T), \quad (V-1)$$

T_Z = thermistor temperature, \dot{Q}_E = electrical power input, C = thermal capacity of the thermistor, and K = dissipation constant. From this equation, the effects of input power level may also be obtained. At a steady state temperature, T_Z^{SS} , $T_Z^{SS} - T_{REF} = \dot{Q}_E / K$. With the input power levels and dissipation constants described below, $T_Z^{SS} - T_{outer} = 0.1$ mK and $T_Z^{SS} - T_{inner} = 1$ μ K. The Laplace transform of eq (V-1) gives the result

$$T_Z(S) = [1/(1 + ZS)] T(S) + Z/[C(1 + ZS)] \dot{Q}_E(S), \quad (V-2)$$

$Z = K/C$ is the time constant of the thermistor, and S is the Laplace transform variable in the transform domain. The thermistor transfer function with respect to ambient changes in temperature is therefore

$$M(S) = T_Z(S)/T(S) = 1/(1 + ZS). \quad (V-3)$$

The time constant, Z , and dissipation constant, K , for both controllers are strongly dependent on the manner in which the thermistor is mounted. The estimates given below for K were based on the manufacturer's recommendations and the thermal conductivities of the mounting materials. Z must account for the thermistor's response time and the coupling of the thermistor to the heater and system mass. For the thermistor's response, independent of this last factor, Z was approximately 1 to 10 seconds. Accounting for the imperfect coupling of the thermistor and heater, rough estimates for Z based on observations of system behavior are

$$Z_{outer} \cong 10 - 20 \text{ s},$$

and

$$Z_{inner} \cong 100 - 150 \text{ s}.$$

The coupling of the thermistor and heating coils on the inner stage was across 6 mm of the inner aluminum enclosure.

Let Z_a , Z_b , Z_c , and Z_d represent the four impedances in a Wheatstone bridge. Then, if $Z = Z_a = Z_b = Z_d$ and $Z_c = Z + dZ$, with dZ much less than Z , the error signal, e , output by the bridge is [77]

$$e = (E/4) (dZ/Z), \quad (V-4)$$

E is the bridge excitation voltage. Thermistors have the property that $Z = Z(T_Z)$, where T_Z is the device temperature. Over a quite large working range $Z(T_Z)$ for the thermistors used in the controller bridges was given by $Z(T_Z) - Z(T_{REF}) = a(T_Z - T_{REF})$ with $a = -0.04 Z(T_{REF})/K$. This along with the thermal transfer function of eq (V-3) gives

$$e = A(S) (T_{REF} - T_Z), \quad (V-5)$$

where $A(S) = k/(1 + ZS)$, $k = 0.01E$ V/K.

To minimize the temperature difference between the thermistor and its surroundings, the self-heating of the thermistors by the bridge excitation voltage

must be kept to a minimum [78]. The manufacturers recommendation was that the self-heating be less than or equal to 0.1 of the desired control resolution. The dissipation constants of the two thermistors as mounted on the two thermal enclosures were 90 mW/K and 42 mW/K for the outer and inner enclosures, respectively. At bridge balance, one-half of the exciting voltage will be across the thermistor. Thus, the powers which may be dissipated in the thermistors assuming 1 mK stability in the outer stage and 10 μ K stability in the inner stage are 9.0 μ W and 42 nW, respectively. The recommended excitation voltages are, in turn 0.85 V and 0.13 V for the outer and inner stages respectively. Since any variations in these voltages would result in an additional variation of the control signal, stable sources are required. As a compromise between the recommended excitation voltage and the need for stability, a 1.35 V mercury cell was used for the outer-stage bridge. The inner-stage bridge was driven by a stabilized oscillator with an output of 100 mV rms.

The circuit diagrams for both stages are shown in figures 40 and 41. The lock-in amplifier employed for the sensor amplifier was an instrument designed and built in-house at the National Bureau of Standards [79]. It had gain settings from 500 up to 10^6 , time constants available from 0.5 s to 20 s, and operated at a fixed frequency of 1 kilohertz. To a first approximation, the transfer functions of both sensor amplifiers may, for small deviation from bridge balance, be represented by the equation

$$H_1(S) = V_1(S)/e(S) = A/(1 + qS), \quad (V-6)$$

A is the amplifier's dc gain or for the lock-in amplifier its gain when the rms value of the input is not changing, and q is the time constant of the respective amplifier. This transfer function for the lock-in amplifier represents the relationship between the dc output voltage and the rms value of the input ac signal with the selected phase. For the lock-in amplifier, q is simply the time constant setting of the amplifier's low pass filter. Equation (V-6) assumes the lock-in filter is equivalent to one RC stage. For the sensor amplifier of the outer-stage-controller, q was set by the 0.004 μ fd capacitor at 4×10^{-2} s since this was the approximate value expected from the 12 dB/octave gain characteristic and anticipated gain setting.

For both stages, the bridge-sensor amplifier combination was configured to obtain the required negative feedback. With the outer stage controller, this only amounted to choosing the correct polarity of the bridge excitation voltage. With the lock-in amplifier, this was accomplished by choosing the appropriate phase to detect.

A brief note about the particular bridge circuit used for the outer sensor amplifier. This circuit allows the use of a single-ended, chopper-stabilized amplifier and its attendant low drift. Another advantage of the circuit is that the gain is not dependent on the absolute value of the bridge resistors and can be simply adjusted by changing the ratio of the 10 n Ω and 100 Ω resistors.

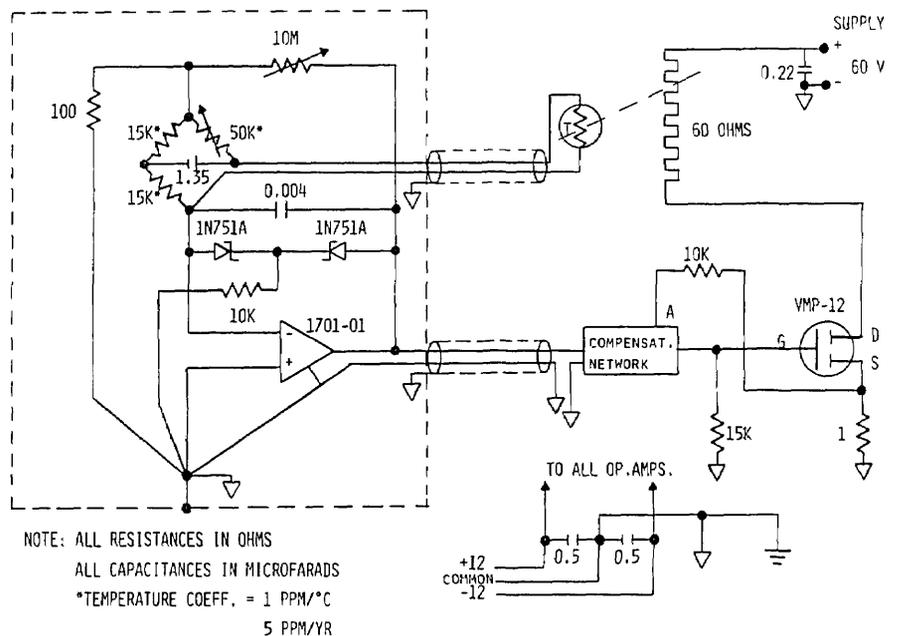


Figure 40—Schematic of electronics for outer temperature control stage.

Since this pole and the pole at $S = 100$ in the proportional channel are so much higher frequency poles than all the other poles of the controllers, their effects, which are minor, will not be included in the following discussion.

The response properties of the power amplifier for both control stages can be understood by noting that a fundamental property of an operational amplifier is to maintain the summing point, i.e., the inverting input, at a "virtual ground" condition [80]. The implication of this property for the power amplifier portion of the controllers is that all current flow to the inverting input of the operational amplifier which drives the MOSFET power transistor, a VMP-12 or 22, must also flow through the 10 K Ω feedback resistor, R_{FB} , connected to the device's source terminal. (This, of course, neglects the small amount flowing through the 100 K Ω resistor to the operational amplifier output.) This condition can be used to relate the voltage applied to the heater, V_3 , and the equivalent voltage output from the compensation network with an equivalent input resistance, R_i , of 100 K Ω s. Thus, with these definitions; $V_2 = I R_i$, where I is the summed current. The voltage at the source terminal will then be maintained at; $V_{FB} = I R_{FB}$. But, $V_{FB} = (V_3 / R_H) R_B$, from a simple application of Ohm's law. So

$$V_3 = (R_{FB} / R_i) (R_H / R_B) V_2. \quad (V-9)$$

Here R_H is the heater resistance and R_B is the bias resistance between common and the device source terminal.

The transfer function of the power amplifier is determined by the dc gain obtained from eq (V-9) and the response of the operational amplifier in the feedback loop. The corner frequency of the MOS device is greater than 100 MHz. Thus, the overall transfer function of the power amplifier would be

$$H_3(S) = V_3 / V_2 = A_p / (1 + pS), \quad (V-10)$$

with $A_p = (R_{FB} / R_i) (R_H / R_B)$ and $p = 6 \times 10^{-3}$ s. As indicated for the compensation network, the effects of this high frequency pole are not accounted for in the following discussion.

To obtain a model of the thermal response of the heaters and their respective masses, a simple lumped parameter model of the heat conduction and radiation was utilized. Referring back to figure 22, let R represent the thermal resistance of all conduction paths between each stage's components and its environment. For the outer stage, these paths include the polyurethane foam, the stainless steel ports to the vacuum pumps, the electrical con-

necting leads, and the epoxy supports used to thermally isolate the outer stage from the granite vibration isolation table. For the inner stage, these paths include the electrical connecting leads and the ceramic supports. By analogy with electrical circuit terminology, R is given by $R = (R_s - T_E) / \dot{Q}_C$, where $T_s =$ temperature of stage, $T_E =$ temperature of the stage's environment and $\dot{Q}_C =$ power loss due to conduction.

Radiation transfer of energy is a major factor which must be considered in the design of the inner control stage. Zemansky [81] shows that the rate of energy transfer \dot{Q}_R , between a body with emissivity, ϵ , and temperature, T , and a surrounding system at temperature, T_E , is given by

$$\dot{Q}_R = A \epsilon \sigma (T^4 - T_E^4). \quad (V-11)$$

Here A is the surface area of the enclosed body and σ is the Stefan-Boltzmann constant.

With the use of this lumped parameter model, the thermal behavior of each stage can be described by the equation for the conservation of energy

$$\dot{Q} = \dot{Q}_{IN} - \dot{Q}_C - \dot{Q}_R \quad (V-12)$$

or

$$C \frac{dT}{dt} = V^2 / r - (T - T_e) / R - A \epsilon \sigma (T^4 - T_E^4), \quad (V-13)$$

where C is the heat capacity of each stage, V is the voltage applied to heater coils of resistance r , and the other symbols are as defined previously. If eq (V-13) is linearized by expanding T and V about their operating point values, T_{REF} and V_0 , at which $dT/dt = 0$, the following equation holds for small deviations about these values

$$\frac{dT}{dt} + \frac{1}{R'C} T = \frac{2V_0}{RC} V + \frac{1}{R'C} T_E, \quad (V-14)$$

where $1/R' = 1/R + 4A \epsilon \sigma T_{REF}^3$. Let $t_T = R'C$ and $t_E = rC$, then the Laplace transform of eq (V-14) gives the result

$$T(S) = G_1(S) V(S) + G_2(S) T_E(S) \quad (V-15)$$

where $G_1(S) = 2V_0 t_T / (t_e (1 + t_T S))$ and $G_2(S) = 1 / (1 + t_e S)$.

The transfer function of the cascaded sensor amplifier compensation network and power amplifier

is given by the product of their individual transfer functions [82]. Therefore, let

$$H(s) = V_3/e = H_1(S) \times H_2(S) \times H_3(S). \quad (V-16)$$

The controllers' feedback responses are then determined by the equations

$$\begin{aligned} e &= A(S)(T_{REF} - T) \\ V &= V_3 = H(S)e \\ T &= G_1(S)V + G_2(S)T_E. \end{aligned} \quad (V-17)$$

The transform variable is not indicated on the dependent variables. Solving this set of equations for T gives

$$\begin{aligned} T &= [G_2(S)T_E + A(S)G_1(S)H(S)T_{REF}] F(S), \\ F(S) &= [1 + A(S)G_1(S)H(S)]^{-1}. \end{aligned} \quad (V-18)$$

The following discussion is presented in terms of changes in T_{REF} . While the transient response of the system is different for changes in T_{REF} and T_E , the stability conditions are the same for both system changes. Calculations were only made in terms of changes in T_{REF} since step changes in T_{REF} were anticipated.

Let $K(S) = T(S)/T_{REF}(S)$, then

$$K(S) = G_1(S)H(S)A(S)F(S). \quad (V-19)$$

$K(S)$ is the overall transfer function describing the response of the closed loop system to changes in the setpoint, T_{REF} . Substitution of the various parameter and function definitions yields the result

$$\begin{aligned} G_1(S)H(S)A(S) \\ = \frac{g(S+S_1)(S+S_2)(S+S_3)}{S(1+t_1S)(1+aS)(1+bS)(1+ZS)(1+qS)}, \end{aligned} \quad (V-20)$$

where $g = 2kAA_pV_3t_T/t_c$.

The zeros S_1 , S_2 , and S_3 are the roots of the equation

$$\begin{aligned} S^3 + [I/p + 1/b + 1/a + D/(abP)]S^2 \\ + [I/bP + I/aP + 1/ab]S + I/abP = 0. \end{aligned} \quad (V-21)$$

This equation reduces to the following equations for the two stages;

outer stage:

$$S^3 + 15S^2 + [5I + 10/D]_1S + 10I/D = 0, \quad (V-22)$$

and inner stage:

$$S^3 + 11S^2 + [20I + 10/D]_2S + 10I/D = 0. \quad (V-23)$$

For eqs (V-22 and V-23), the roots are approximately [83]:

outer stage:

$$S_1 \cong -15, S_2, S_3 \cong -(2S_1)^{-1} []_1 \pm S_1^{-1} []_1^{1/2},$$

and inner stage:

$$S_1 \cong -11, S_2, S_3 \cong -(2S_1)^{-1} []_2 \pm S_1^{-1} []_2^{1/2}.$$

Parameters of the outer-stage controller were

heat capacity = $C = 1.42 \times 10^4 \text{ J/}^\circ\text{C}$
 heater resistance = $r = 60 \ \Omega$
 thermal resistance = $R = 0.8 \text{ }^\circ\text{C/W}$
 steady state voltage = $V_0 = 20 \text{ V}$
 bridge constant = $k = 0.01 \text{ V/}^\circ\text{C}$
 stage time constant = $RC = t_T = 1.1 \times 10^4 \text{ s}$
 power amplifier gain = $A_p = 6$

The poles, other than the high frequency ones not accounted for in eq (V-20), were $a = 0.25\text{s}$, $b = 0.5\text{Ds}$, $Z = 10\text{-}20\text{s}$, and $q = 4 \times 10^{-2}\text{s}$.

Parameters of the inner-stage controller were

heat capacity = $C = 2.9 \times 10^3 \text{ J/}^\circ\text{C}$
 heater resistance = $r = 1.05 \times 10^3 \ \Omega$
 thermal resistance = $R = 8 \text{ }^\circ\text{C/W}$
 steady state voltage = $V_0 = 10 \text{ V}$
 bridge constant = $k = 1 \times 10^{-3} \text{ V/}^\circ\text{C}$
 stage time constant = $RC = 2.3 \times 10^4 \text{ s}$
 power amplifier gain = $A_p = 3$

The poles were located at $a = 2.0\text{s}$, $b = 0.05\text{Ds}$, $Z \cong 100\text{-}150\text{s}$, and $q \cong 0.5\text{-}2\text{s}$. The thermal resistance R of this stage represents the parallel resistance of conduction and radiation.

The poles and zeros of $K(S)$ in the S -plane completely determine the stability of the temperature-control stages and their transient responses to changes in setpoint temperature and environmental temperature [82]. Some guidelines for the design of a feedback system which are usually scattered throughout most texts on control systems are

- A. The system will be unstable if $K(S)$ contains any poles in the right half-plane.
- B. The system's transient response is increasingly oscillatory as the poles of $K(S)$ move toward the imaginary axis.

- C. The system's speed of response is proportional to the distance of the dominant poles from the S -plane origin.
- D. Introduction of negative real zeros in $K(S)$ in general increases system stability.
- E. The relative weighting, R_k , of the k^{th} mode of a system, i.e., a pole or a complex pair of poles, is determined by

$$|R_k| = \frac{\prod_{i=1}^m |P_i - n_i|}{|P_k| \prod_{\substack{i=1 \\ j \neq k}}^n |P_i - P_j|}, \quad (\text{V-24})$$

- where $n_i, i=1, m$ are the zeros of $K(S)$ and $P_j, j=1, n$, are the poles of $K(S)$. Thus, placing zeros near troublesome or oscillatory poles minimizes their effect.
- F. The response of a feedback system (as in fig. 23) to a constant reference input has a zero steady-state error if and only if the open loop transfer function $G_1(S)H(S)A(S)$ has at least one pole at $S=0$. For zero steady-state error with reference inputs of the form, $a t^n$, the requirement is that $G_1(S)H(S)A(S)$ have a pole at $S=0$ with multiplicity m greater than n .
 - G. For disturbance inputs, a similar property holds. Here the pole at $S=0$ must be in the controller transfer function $H(S)$. Therefore, a controller with a pole at $S=0$ insures that a zero steady-state error may be achieved in both cases.

An attempt was made to use these guidelines in adjusting the lead and lag times of the compensation networks. The procedure for adjustment was to set I and D to zero and $P=1$. The overall gain of the feedback loop was then increased by increasing A , the sensor amplifier gain, until the stage-controller broke into oscillation. D and I were then increased to the values calculated below; D to eliminate the instability by introducing zeros near the origin and I to eliminate or minimize any offset by introducing a pole at the origin. The effects of these changes can be visualized in terms of a root-locus diagram. This type of diagram is a graph of the paths traced by the poles of $K(S)$ as g of eq (V-20) is varied from zero to infinity. A root-locus diagram, constructed with the aid of the root-locus properties described by Fortmann and Hitz [84] and Kuo [85], for $K(S)$ of the outer-stage controller is sketched in figure 43. Only the dominant poles and zeros are drawn since the effects of the other ones are principally to change the angles of the asymptotes of the root-loci.

An examination of the diagram in figure 43 will show that before compensation, the two dominant poles are the ones due to the thermal lag of the stage mass and to the time constant of transferring energy from the heater to the thermistor. These are labelled $1/t_r$ and $1/Z$, respectively, in the diagram. The compensation network introduces three zeros and three new poles. The zeros tend to straighten the root-locus branch that causes instability so that the same gain may be used with no oscillations. An equal margin of stability without compensation

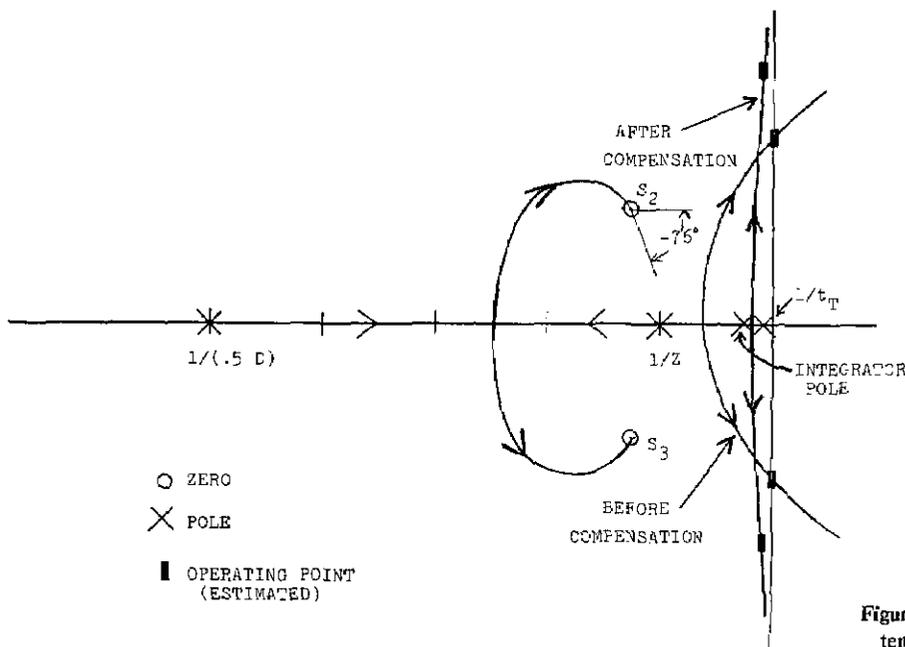


Figure 43—Root-locus diagram for temperature control system.

would have required significantly reduced loop gain and, thus, less attenuation of environmental disturbances.

D and I were initially chosen to bring the zeros, S_2 and S_3 , to be adjacent to the pole at $1/Z$ to approximately cancel its effect. From the roots given for eqs (V-22 and V-23), this condition requires that for the outer stage

$$\frac{1}{30} [5I + 10/D] = 0.1$$

and for the inner stage

$$\frac{1}{22} [20I + 10/D] = 0.01$$

Initial setting for D and I were

for outer stage:

$$I = 0.1 \text{ S}^{-1}, D = 3\text{s}$$

and for inner stage:

$$I = 0.01 \text{ S}^{-1}, D = 50\text{s}$$

After some "tweaking" of I , D , and P to get maximum gain without oscillations, the responses to setpoint changes shown in figures 44 and 45 were obtained. These responses are certainly not optimum, but are acceptable. The sensor amplifier gains and compensation network settings to obtain these responses were: outer stage — $A = 5000 \text{ V/V}$, $I = 0.013 \text{ s}^{-1}$, $D = 4 \text{ s}$, $P = 1$, and inner stage — $A = 5 \times 10^5 \text{ V/V}$, $I = 0.015 \text{ s}^{-1}$, $D = 200 \text{ s}$, $P = 0.75$.

With these sensor amplifier gains, the closed loop gain, including the bridge and thermistor response, were

$$\text{outer stage: } A_{CL} = A A_p k = 3,000 \text{ V}^\circ\text{C}$$

$$\text{inner stage: } A_{CL} = A A_p k = 15,000 \text{ V}^\circ\text{C}$$

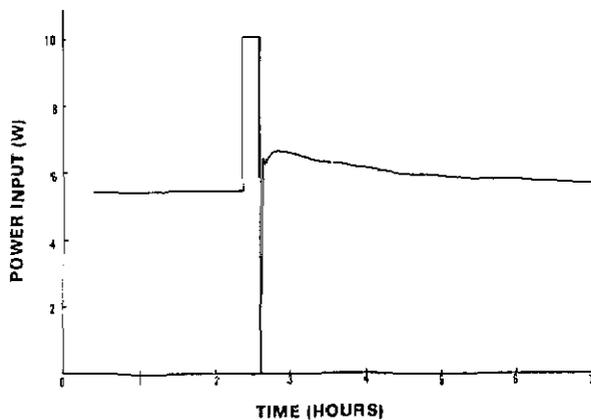


Figure 44—Response of outer temperature control stage to setpoint change.

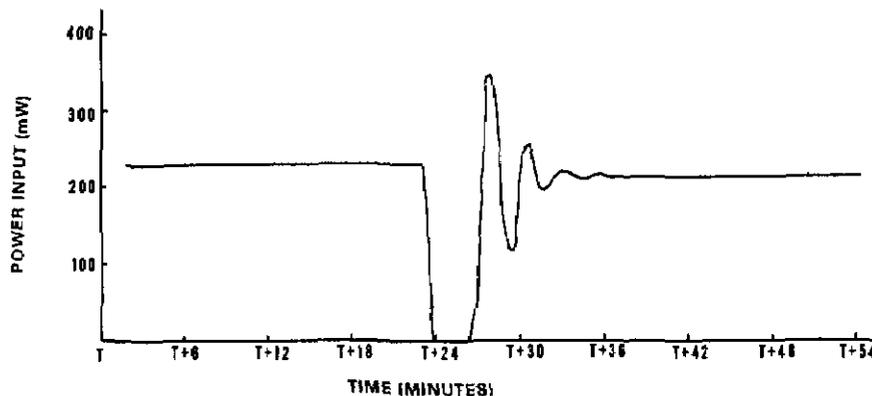


Figure 45—Response of inner temperature control stage to setpoint change.

References

- [1] The example given is similar to an analogy given by Langenberg, D.N. The 1973 Nobel Prize for Physics, *Science* **182** 701 (1973).
- [2] Schrödinger, E., Quantization as a problem of proper values, Part I. *Annalen der Physik*. **79** 361 (1926). Part II, **79** 489 (1926). Part III, **80** 437 (1926). Part IV, **81** 109 (1926).
- [3] Oppenheimer, J.R., Three notes on the quantum theory of aperiodic effects, *Phys. Rev.* **31** 66 (1928).
- [4] Fowler, R.H., and L. Nordheim, Electron emission in intense electric fields, *Proc. Roy. Soc. (London)* **119** 173 (1928).
- [5] Gamow, G., Zur quanten theorie des atomkernes, *Z. Phys.* **51** 204 (1928).
- [6] Frenkel, J., On the electrical resistance of contacts between solid conductors, *Phys. Rev.* **36** 1604 (1930).
- [7] Holm, R., *Electric contacts*. New York: Springer-Verlag (1967).
- [8] Giaever, I., Energy gap in superconductors measured by electron tunneling, *Phys. Rev. Lett.* **5** 147, 464 (1960); and Electron tunneling and superconductivity, *Science* **183** 1253 (1974) Nobel Prize Lecture.
- [9] Esaki, L., Long journey into tunneling, *Science* **183** 1149 (1974) Nobel Prize Lecture; and *Phys. Rev.* **109** 603 (1958).
- [10] Josephson, B.D., Possible new effects in superconductive tunneling, *Phys. Lett.* **1** 251 (1962); and discovery of tunneling supercurrents, *Science* **184** 527 (1974) Nobel Prize Lecture.
- [11] Lambe, J., and R.C. Jaklevic, Molecular vibration spectra by inelastic electron tunneling, *Phys. Rev.* **165** 821 (1968).
- [12] Young, R.D.; J. Ward, and F. Scire, Observation of metal-vacuum-metal tunneling, field emission, and the transition region, *Phys. Rev. Lett.* **27** 922 (1971); and Young, R., J. Ward, and F. Scire, The topografiner: An instrument for measuring surface microtopography, *Rev. Sci. Instrum.* **43** 999 (1972).
- [13] Thompson, W.A., and S.F. Hanrahan, Thermal drive apparatus for direct vacuum tunneling experiments, *Rev. Sci. Instrum.* **47** 1303 (1976).
- [14] Burstein, E., and S. Lundqvist, *Tunneling phenomena in solids*. New York: Plenum Press (1969).
- [15] Duke, C.B., *Tunneling in solids*. New York: Academic Press (1969).
- [16] Simmons, J.G., Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* **34** 1793 (1963).
- [17] Harrison, W.A., Tunneling from an independent-particle point of view. *Phys. Rev.* **123** 85 (1961).
- [18] Stratton, R., Volt-current characteristics for tunneling through insulating films. *J. Phys. Chem. Solids* **23** 1177 (1962).
- [19] Brinkman, W.F.; R.C. Dynes, and J.M. Rowell, Tunneling conductance of asymmetrical barriers, *J. Appl. Phys.* **41** 1915 (1970).
- [20] Applebaum, J.A., and W.F. Brinkman, Theory of many-body effects in tunneling, *Phys. Rev.* **186** 464 (1969).
- [21] Caroli, C.; R. Combescot, D. Lederer-Rozenblatt, P. Nozieres, and D. Saint-James, Comment on a recent theory of tunneling without the transfer Hamiltonian formalism, *Phys. Rev. B* **12** 3977 (1975), and references contained therein.
- [22] Feuchtwang, T.E., Theory of tunneling without transfer-Hamiltonian, *Phys. Rev. B* **12** 3979 (1975), and references contained therein.
- [23] Kurtin, S.L.; T.C. McGill, and C.A. Mead, Direct interelectrode tunneling in GaSe, *Phys. Rev. B* **3** 3368 (1971).
- [24] Duke, C.B.; Reference 15, pp. 60-65.
- [25] Lang, N.D., and W. Kohn, Theory of metal surfaces: charge density and surface energy, *Phys. Rev. B* **1** 4555 (1970).
- [26] Lang, N.D., and W. Kohn, Theory of metal surfaces: work function, *Phys. Rev. B* **3** 1215 (1971).
- [27] Lang, N.D., and W. Kohn, Theory of metal surfaces: induced surface charge and image potential. *Phys. Rev. B* **7** 3541 (1973).
- [28] Kohn, W., and L.J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140** A1133 (1965); and P. Hohenberg and W. Kohn, *Phys. Rev.* **136** B864 (1964). See also Lang, N.D., Density-functional formalism and the electronic structure of metal surfaces, *Solid State Physics* **28** 225 (1973).
- [29] Hedin, L., and B. Lundqvist, Explicit local exchange-correlation potentials, *J. Phys. C* **4** 2064 (1971).
- [30] Slater, J.C., *The self-consistent field for molecules and solids*. New York: McGraw-Hill Book Co. (1974) Chapter 2.
- [31] Moruzzi, V.L.; J.F. Janak and A.R. Williams, Calculated electronic properties of metals. New York: Pergamon Press (1978) p. 9.
- [32] Eastman, D.E., Photoelectric work functions of transition, rare-earth, and noble metals, *Phys. Rev. B* **2** 1 (1970); and G.A. Haas and R.E. Thomas, Thermionic emission and work function, in *Measurement of physical properties*, E. Passaglia, ed., Vol. VI of *Techniques of metal research*, R.F. Bunshah, ed. New York: Interscience Publishers (1972).
- [33] Bardeen, J., Theory of the work function, *Phys. Rev.* **49** 653 (1936).
- [34] Applebaum, J.A., *Electronic structure of solid surfaces*, in *Surface physics of materials*, J.D. Blakely, ed., New York: Academic Press (1975); and Applebaum, J.A., and D.R. Hamann, *Electronic structure of solid surfaces*, *Rev. of Mod. Phys.* **48** 479 (1976).
- [35] Kittel, C., *Elementary statistical physics*. New York: John Wiley (1958) pp. 25-31.
- [36] Spence, E., *Electronic semiconductors*. New York: McGraw-Hill Book Co. (1958) Chapter 10.
- [37] Bennett, A.J., and C.B. Duke, Self-consistent model of bimetallic interfaces, *Phys. Rev.* **160** 541 (1967).
- [38] Simmons, J.G., Image force in metal-oxide-metal tunnel junctions, in Reference 14.
- [39] Duke, C.B., Reference 15, p. 6 and p. 207.
- [40] Landau, L.D., and E.M. Lifshitz, *Quantum mechanics: non-relativistic theory*, 3rd ed. Oxford, England: Pergamon Press (1977) pp. 55-60.
- [41] Blatt, J.M., and V.F. Weisskopf, *Theoretical nuclear physics*. New York: John Wiley and Sons (1952) pp. 528-529.
- [42] Merzbacher, E., *Quantum mechanics*, 2nd Edition, New York: John Wiley and Sons (1970) pp. 116-126.
- [43] Bohm, D., *Quantum theory*. New York: Prentice-Hall, Inc. (1951) pp. 271-277.
- [44] Duke, C.B., Reference 15, p. 59.
- [45] Politzer, B.A., Comparison of a numerical method and the WKB approximation in the determination of transmission coefficients for thin insulating films, *J. Appl. Phys.* **37** 279 (1966).
- [46] Simmons, J.G., Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film, *J. Appl. Phys.* **34** 1793 (1963); and *J. Appl.*

- Phys. 35 2655 (1964). The forms of the image potential given here must be corrected according to the paper by Simmons in Reference 14.
- [47] Simmons, J.G., *J. Appl. Phys.* 34 238 (1963).
- [48] Knauss, H.P., and R.A. Breslow, Current-voltage characteristic of tunnel junctions, *Proc. IRE* 50 1834 (1962).
- [49] Rowell, J.M., Tunneling anomalies—experiment, in Reference 3.
- [50] Murphy, E.L., and R.H. Good Jr., Thermionic emission, field emission, and the transition region, *Phys. Rev.* 102 1464 (1956).
- [51] Young, R.D., Theoretical total-energy distribution of field-emitted electrons, *Phys. Rev.* 113 110 (1959).
- [52] Miller, S.C. Jr., and R.M. Good Jr., A WKB-type approximation to the Schrödinger equation, *Phys. Rev.* 91 174 (1953).
- [53] Hunter, J.L., *Acoustics*. Englewood Cliffs, NJ: Prentice-Hall, Inc. (1962), Chapter 8.
- [54] Rhodin, T.N., and D.L. Adams, Adsorption of gases on solids, Chapter 5 in *Treatise on solid state chemistry: surfaces*, N.B. Hannay, ed. New York: Plenum Press (1976).
- [55] Chesters, M.A., and G.A. Somorjai, The chemisorption of oxygen, water, and selected hydrocarbons on the (111) and stepped gold surfaces, *Surf. Sci.* 52 21 (1975).
- [56] Trapnell, B.M.W., The activities of evaporated metal films in gas chemisorption, *Proc. Roy. Soc.* A218 566 (1953).
- [57] Gland, J.L., and G.A. Somorjai, Low-energy-electron-diffraction and work function studies of adsorbed organic monolayers on the (100) and (111) crystal face of platinum. *Advances in Colloid and Interface Science* 5 205 (1976).
- [58] Hocken, R., Physicist at National Bureau of Standards, private communication.
- [59] Cutkosky, R., and B.F. Field, Standard cell enclosure with 20 μ K stability, *IEEE Trans. Instrum. Meas.* IM-23 295 (1974).
- [60] Emerson and Cuming, Inc.; Canton, MA 02021; Stycast 2850KT with catalyst 24LV for cylindrical walls, catalyst 9 for top and bottom surfaces.
- [61] The particular ceramic used was PZT-8, manufactured by the Vernitron Corporation; Piezoelectric Division; 232 Forbes Road; Bedford, OH 44146. A general description of this class of ceramics is given in *Piezoelectric Ceramics*, by B. Jaffe, W.R. Cook and H. Jaffe, New York: Academic Press (1971).
- [62] Emerson and Cuming, Inc.; Canton, MA 02021; Ecobond Solder V-91.
- [63] The insulating film used was "Kapton" manufactured by E.I. duPont deNemours and Company, Inc.; Plastics Products and Resins Department; Industrial Films Division; Wilmington, DE 19898. Kapton is a mechanically strong, transparent amber-colored film which has a very low outgassing rate and a dielectric strength of 2.75×10^5 V/mm. The manufacturer specifies that no outgassing was revealed by weight loss at 200 °C and 3×10^{-6} Torr pressures. This behavior was confirmed before use in this experiment by observing that a piece .25m \times .25m \times 25 micrometer thick could quickly be pumped down to a pressure of 10^{-5} Torr with a 50 l/s ion pump system.
- [64] Model 341A D.C. Voltage Calibrator manufactured by John Fluke Mfg. Co., Inc.; P.O. Box 43210; Mountlake Terrace, WA 98043.
- [65] Scire, F.E., and E.C. Teague, Piezo-driven 50 micrometer range stage with sub-nanometer resolution, *Rev. Sci. Instrum.* 49 1735 (1978).
- [66] Morse, P.M., and K.V. Ingard, *Theoretical acoustics*. New York: McGraw-Hill Book Co. (1968) Chapter 5.
- [67] Crandall, S.H., The role of damping in vibrational theory, *J. of Sound Vibr.* 11 3 (1970).
- [68] Lifshitz, E.M., The theory of molecular attractive forces between solids, *Soviet Physics, J. of Experimental and Theoretical Physics* 2 73 (1956).
- [69] Israelachvili, J.N., and D. Tabor, Van der Waals forces: theory and experiment, in progress in surface and membrane science, Volume 7, J.F. Danielli, M.D. Rosenberg, and D.A. Cadenhead, eds. New York: Academic Press (1973).
- [70] Krupp, H.; W. Schnabel and G. Walter, *Journal of Colloid and Interface Science* 39 421 (1972).
- [71] Panofsky, W.K.H., and M. Phillips, *Classical electricity and magnetism*. Reading, MA: Addison-Wesley Publishing Co. (1962) pp. 103-107.
- [72] Timoshenko, S., and J.N. Goodier, *Theory of elasticity*. New York: McGraw-Hill Book Co. (1951) pp. 362-372.
- [73] The author is indebted to C.P. Reeve, National Bureau of Standards, for suggesting and assisting with these changes.
- [74] Landau, L.D., and E.M. Lifshitz, *Statistical Physics*. Reading, MA: Addison-Wesley Publishing Co. (1969) pp. 348-352.
- [75] Grove, A.S., *Physics and technology of semiconductor devices*. New York: John Wiley and Sons, Inc. (1967) p. 253.
- [76] Boylestad, R., and L. Nashelsky, *Electronic devices and circuit theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc. (1978).
- [77] Terman and Pettit, *Electronic measurements*. New York: McGraw-Hill Book Company (1952) pp. 74-83.
- [78] *Thermistor Manual*, Bulletin EMC-6, Fenwal Electronics (1974).
- [79] The amplifier was designed by Mark Horowitz, National Bureau of Standards.
- [80] Wait, J.V.; L.P. Huelsman, and G.A. Korn, *Introduction to operational amplifier theory and applications*. New York: McGraw-Hill Book Company (1975) pp. 1-55.
- [81] Zemansky, M.W., *Heat and thermodynamics*. New York: McGraw-Hill Book Company (1957) p. 106.
- [82] Fortmann, T.E., and K.L. Hitz, *An introduction to linear control systems*. New York: Marcel Dekker, Inc. (1977) pp. 159-386.
- [83] Uspensky, J.V., *Theory of equations*. New York: McGraw-Hill Book Company (1948) pp. 82-94.
- [84] Reference 82, pp. 300-307.
- [85] Kuo, B.C., *Automatic control systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc. (1962) pp. 1-170 and 246-300.