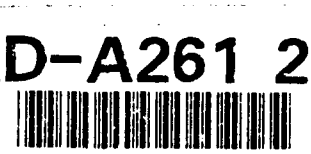
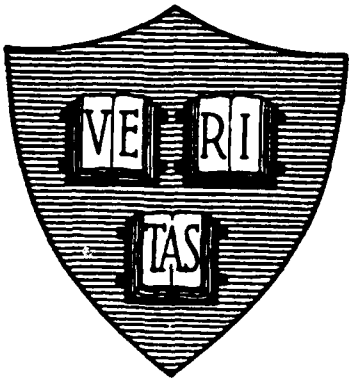


Office of Naval Research

National Sciences Foundation Grant DMR-89-12927
National Sciences Foundation Grant DMR-89-20490
Grant N00014-89-J-1023
Grant N00014-89-J-1565

SINGLE ELECTRON CHARGING EFFECTS IN MESOSCOPIC SYSTEMS



By

Ashraf Elfar Hanna

October 1992

Technical Report No. 34

Division of Applied Sciences
Harvard University · Cambridge, Massachusetts

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Unclassified/Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 34	5. MONITORING ORGANIZATION REPORT NUMBER(S) Office of Naval Research
6a. NAME OF PERFORMING ORGANIZATION Harvard University 6b. OFFICE SYMBOL (*if appicable*)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (*City, State, and ZIP Code*) Division of Applied Sciences Harvard University Cambridge, MA 02138	7b. ADDRESS (*City, State, and ZIP Code*) 800 N. Quincy Street Arlington, VA 22217
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (*if appicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-89-J-1023
8c. ADDRESS (*City, State, and ZIP Code*) 800 N. Quincy St. Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (*Include Security Classification*)
Single Electron Charging Effects in Mesoscopic Systems

12. PERSONAL AUTHOR(S)
Ashraf Elfar Hanna

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Interim (Technical)	FROM_____TO_____	October 1992	157

16. SUPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

We report the investigation of two ultra-small tunnel junctions in series. These junctions have capacitances on the order of 10-18 – 10-19F each. In this regime the charging energy associated with a single electron tunneling event (e2/2C) can be made significantly larger than thermal fluctuations (kBT) at liquid helium temperatures. As such, measurements of the tunneling current as a function of voltage show unique features associated individual electron tunneling events.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ■ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Prof. Michael Tinkham, Director	22b. TELEPHONE (*Include Area Code*) (617) 495-3735 22c. OFFICE SYMBOL

19. ABSTRACT (continued)

In this thesis we present high-resolution data from measurements on a two ultra-small tunnel junctions in series. The junctions are created by using a scanning tunneling microscope vacuum tunneling into a small (10-100Å) gold sphere which is separated from a metal electrode by an oxide layer. Our measurements show sharp features, such as the Coulomb blockade and Coulomb staircase, in the I(V) data of two junction systems.

We explain our data by use of the orthodox theory of tunneling. The features in the data displayed overwhelming agreement with the predictions of the orthodox theory, which correctly produces the location of linear voltage onsets, current steps and step slopes. The high resolution of our data also allowed the categorizing of the I(V) curves according to the onset of certain features. The orthodox theory, combined with our data, has also shed understanding on the nature of Q_0, the "fractional" charge on the middle electrode.

Higher order tunneling processes associated with macroscopic quantum tunneling of charge, q-MQT, have also been observed in our data. This observation is not only of fundamental physical interest, but is an effect which could set the limit of future devices operating on the transfer of single electrons.

This research was performed under the direction of Professor Michael Tinkham.

Office of Naval Research

National Sciences Foundation Grant DMR-89-12927
National Sciences Foundation Grant DMR-89-20490
Grant N00014-89-J-1023
Grant N00014-89-J-1565

SINGLE ELECTRON CHARGING EFFECTS IN MESOSCOPIC SYSTEMS

by

Ashraf Elfar Hanna

Technical Report No. 34

October 1992

The research reported in this document was made possible through support extended the Division of Applied Sciences, Harvard University, by the National Sciences Foundation Grants DMR-89-12927 and DMR-89-20490 and by the Office of Naval Research, under Grants N00014-89-J-1023, N00014-89-J-1565.

Division of Applied Sciences
Harvard University · Cambridge, Massachusetts

ABSTRACT

We report the investigation of two ultra-small tunnel junctions in series. These junctions have capacitances on the order of 10^{-18} - 10^{-19}F each. In this regime the charging energy associated with a single electron tunneling event ($e^2/2C$) can be made significantly larger than thermal fluctuations (k_BT) at liquid helium temperatures. As such, measurements of the tunneling current as a function of voltage show unique features associated individual electron tunneling events.

In this thesis we present high-resolution data from measurements on a two ultra-small tunnel junctions in series. The junctions are created by using a scanning tunneling microscope vacuum tunneling into a small (10-100Å) gold sphere which is separated from a metal electrode by an oxide layer. Our measurements show sharp features, such as the Coulomb blockade and Coulomb staircase, in the I(V) data of two junction systems.

We explain our data by use to the orthodox theory of tunneling. The features in the data displayed overwhelming agreement with the predictions of the orthodox theory, which correctly produces the location of linear voltage onsets, current steps and step slopes. The high resolution of our data also allowed the categorizing of the I(V) curves according to the onset of certain features. The orthodox theory, combined with our data, has also shed understanding on the nature of Q_0, the "fractional" charge on the middle electrode.

Higher order tunneling processes associated with macroscopic quantum tunneling of charge, q-MQT, have also been observed in our data.

This observation is not only of fundamental physical interest, but is an effect which could set the limit of future devices operating on the transfer of single electrons.

This research was preformed under the direction of Professor Michael Tinkham.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

The late twentieth century ushered in a plethora of new electronic products with decreased size yet increased performance. At the heart of this advancement was the ability to greatly miniaturize transistors and to increase the number of transistors per unit area on a chip. This decrease in circuit size not only increased performance, but allowed electronics production to become less expensive. As we near the end of the twentieth century, where practically every product contains some sort of electronic device, the push for even smaller circuits continues. However, as the size of electronic circuits decrease, the effects of single electrons can become significant. This inevitable limit of "classic" electronics is also the beginning of a new age of "quantum" electronics, where the flow and control of individual electrons is the standard.

Advances in electron-beam lithography have allowed the creation of several solid state systems which demonstrate behavior that is affected by the quantum mechanical properties of single electrons. For example, quantum conductance fluctuations resulting from interference of electron wave functions, and unique to each sample, have been observed in metallic wires and semiconducting devices [Washburn and Webb, 1986; Smith, 1989]. Also, there have been observations of discrete electron states in GaAs quantum well structures known as "quantum dots".

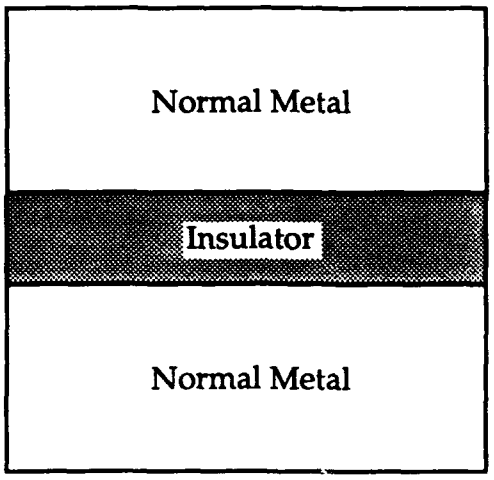


Figure 1.1, Normal-metal tunnel junction. A normal metal tunnel junction consists of two sections of a normal-metal separated by an insulator. The thickness of the insulator must be small enough so that a measurable finite probability exists for electron tunneling across the insulator.

These systems are labeled *mesoscopic*, that is, though they are composed of a macroscopic number of particles, they are small enough so that their behavior is directly attributable to microscopic processes. This work is an experimental study of such a mesoscopic system: two normal-metal tunnel junctions in series.

A normal-metal tunnel junction is simply two normal-metal electrodes separated by a thin insulating barrier, see Fig 1.1. Though tunneling phenomena have been known for quite some time, in our experiment the junction area is small enough (100 - 1000Å^2), yielding junction capacitances on the order of 10^{-18} - 10^{-19} F, that the electrostatic charging energy involved in the tunneling of a single electron cannot be ignored as is usually done. This corresponds to such a large single electron electrostatic charging, 10 - 500 meV, that a current versus voltage measurement, I(V), no longer displays the expected linear shape; rather, it shows peculiar zero conductance region followed by a series of discrete steps. Each step is associated with the singular increase of the number of electrons on the junction. Such an I(V) measurement of a two-junction system is shown in Fig. 1.2, where each step is approximately 0.5 volts wide.

Much of similar recent work on small two-junction systems has been done on lithographically patterned tunnel junctions and on the two-dimensional electron gas in GaAs/AlGaAs heterostructures. Charging effects have also been observed in sheet resistance of granular thin films consisting of small metal particles embedded in an oxide layer. We have used a scanning tunneling microscope (STM) to observe charging effects, forming one of the tunneling barriers between the metal STM tip and a small metal particle, and another tunnel junction between the particle and an underlying metal plane.

The advantages of using the STM setup are: (i) the ability to vary the parameters of one of the junctions (changing the tip to particle or grain distance), (ii) extremely small capacitances giving a charging energy much

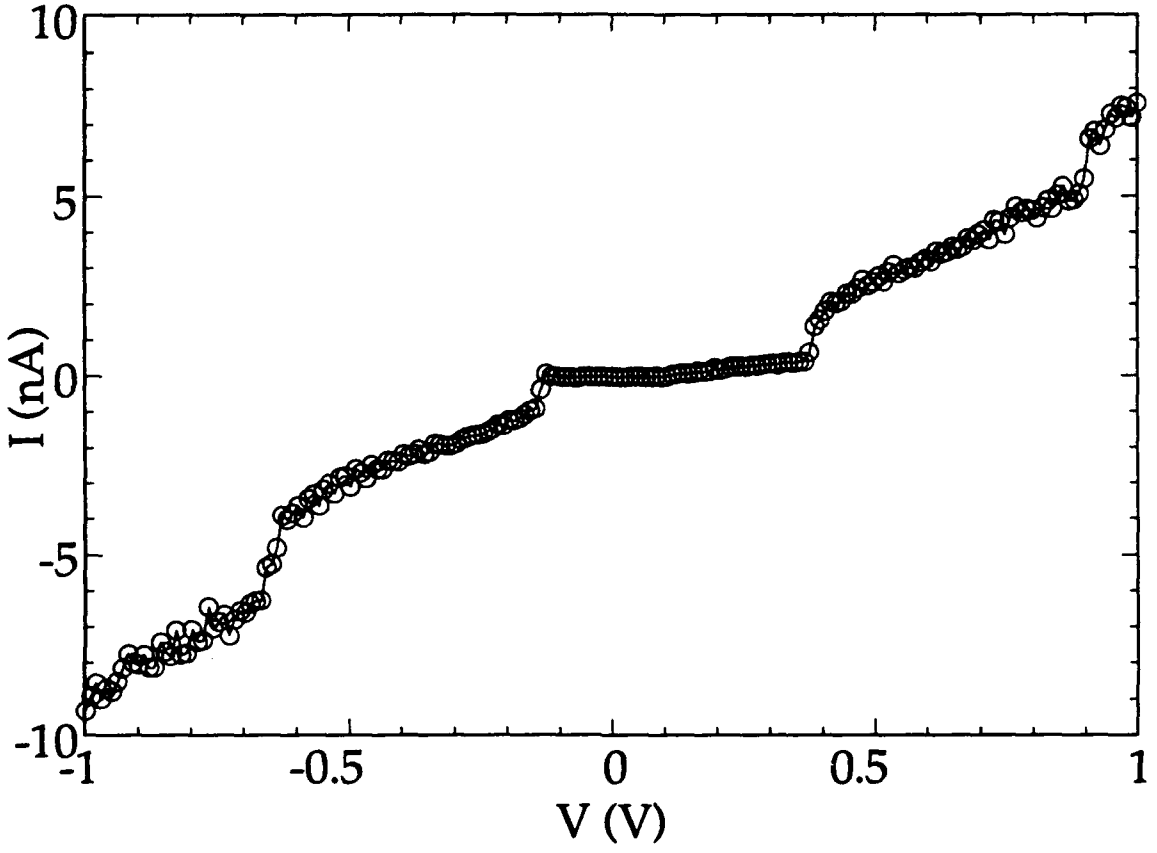


Figure 1.2, Coulomb staircase. This data shows the measured current - voltage characteristics I(V) on a two-junction system. Classically a linear I(V) is expected, however, in our system where the capacitance of the junction is on the order of 10^{-19} F, the quantized nature of charge is evident. The zero conductance region in the center is the "Coulomb blockade." The following steps are called the "Coulomb staircase," and are due to the increase of the number of electrons on the middle electrode by one.

greater than k_BT, even at 4.2K,† (iii) higher carrier density in a mental systems yielding a larger, easier-to-measure current, and (iv) the measurements are on a single particle, not averaged over an ensemble which is done in some systems mentioned above.

These advantages allowed us to take much higher resolution data than previously observed. We were able not only to see clear steps as shown in Fig. 1.2, but our data also showed sharp new features such as certain asymmetries and current onsets as shown in Fig. 1.3. Extracting the exact location of certain features allowed us to carefully investigate proposed theories explaining our system. A numerical simulation of the semi-classical or so-called "orthodox" theory of tunneling was found to be in excellent agreement with our data. In the orthodox picture the only quantum effect is the ability of the electron to tunnel through the classically forbidden insulating barrier.

The source of the interesting physics of this system is that the charge state of the junction may be viewed as continuously variable, but the discharge process is discrete because only one electron crosses the junction during each tunnel event. That there may be a fractional electron charge on the junction requires careful discussion. The electron fluid in the wires and the junction is best thought of as being dominated by the long range Coulomb interaction. This nearly incompressible fluid can be pressed against the junction plate by a continuously adjustable force, the bias electric field. This

† The other methods mentioned above are usually operated at mK temperatures.

produces a polarization charge which increases continuously with the applied bias and in effect corresponds to a fractional charge on the junction. Though the electron gas is made of discrete quanta of charge e, that does not alter the fact that the polarization charge can be adjusted continuously. In fact, the charge on the junction, Q, may be an arbitrarily small fraction of the elementary charge e, caused by a small shift of the electrons in the junction electrodes with respect to the positive ionic background. Tunneling, though, may only occur with discrete electrons carrying a quantized charge e. In the orthodox model, the environment is treated classically, hence there is no *quantum* uncertainty in the polarization charge. There is uncertainty in the polarization charge due to thermal fluctuations, however.

The basic physical requirements necessary for observing single electron effects are: (i) the tunneling resistance being much greater than the quantum resistance, $R_T \gg R_Q$, and (ii) the charging energy being much larger than thermal fluctuations, $E_C \gg k_BT$. Ignoring for the moment the special case of a single junction, the condition $R_T \gg R_Q$, ensures that the wave function of an added electron on a metallic island is well localized there. If the system has a lower tunneling resistance, charges can be transferred through the barrier without paying the Coulomb energy penalty since delocalized states with lower Coulomb energy are available for the transport. The condition $E_C \gg k_BT$ ensures that thermal activation of electrons does not wash out Coulomb energy effects. Externally applied voltages can then influence the Coulomb energy in order to manipulate the individual electrons.

When the junction is voltage biased, the I(V) curve will show a simple Coulomb blockade and staircase, Fig. 1.2. However, when the junction is

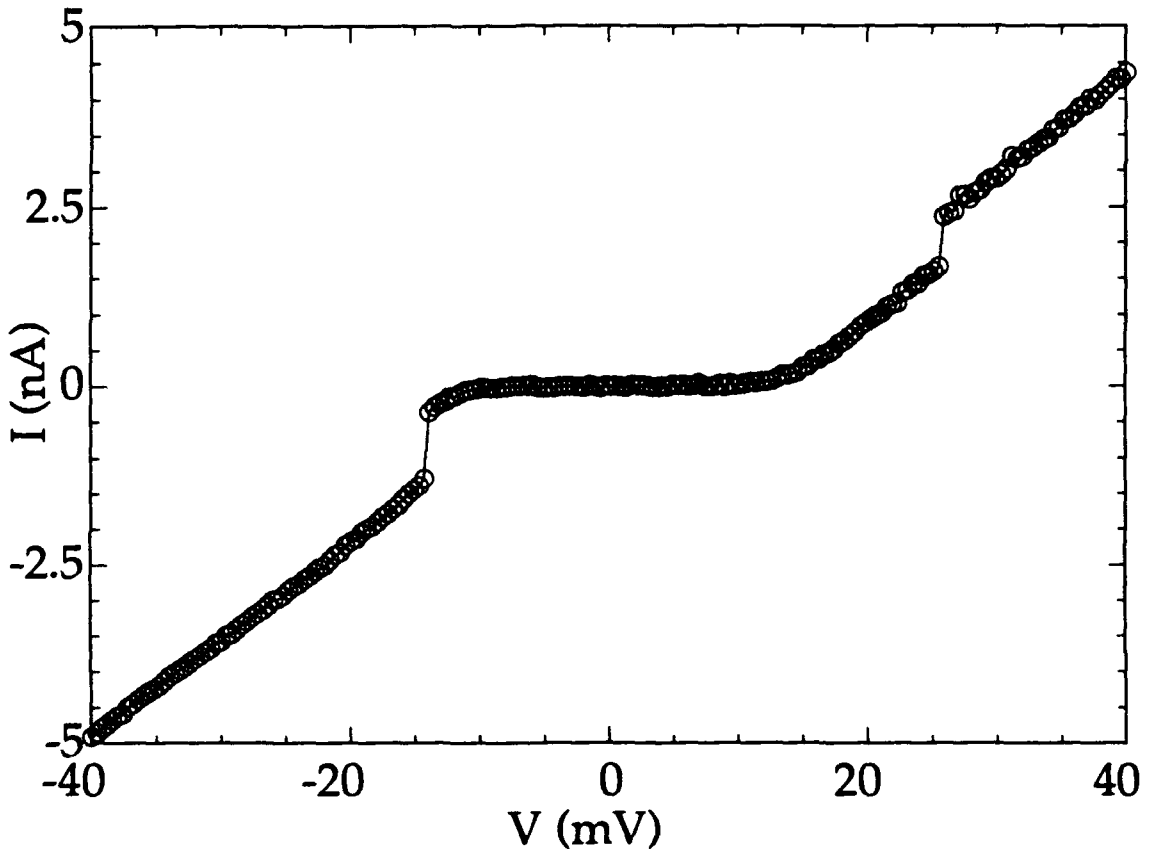


Figure 1.3a, I(V) characteristics I. These I(V) data show precise voltages at which the current rises linearly and locations at which there are distinct jumps. There is also an asymmetry between positive and negative voltages.

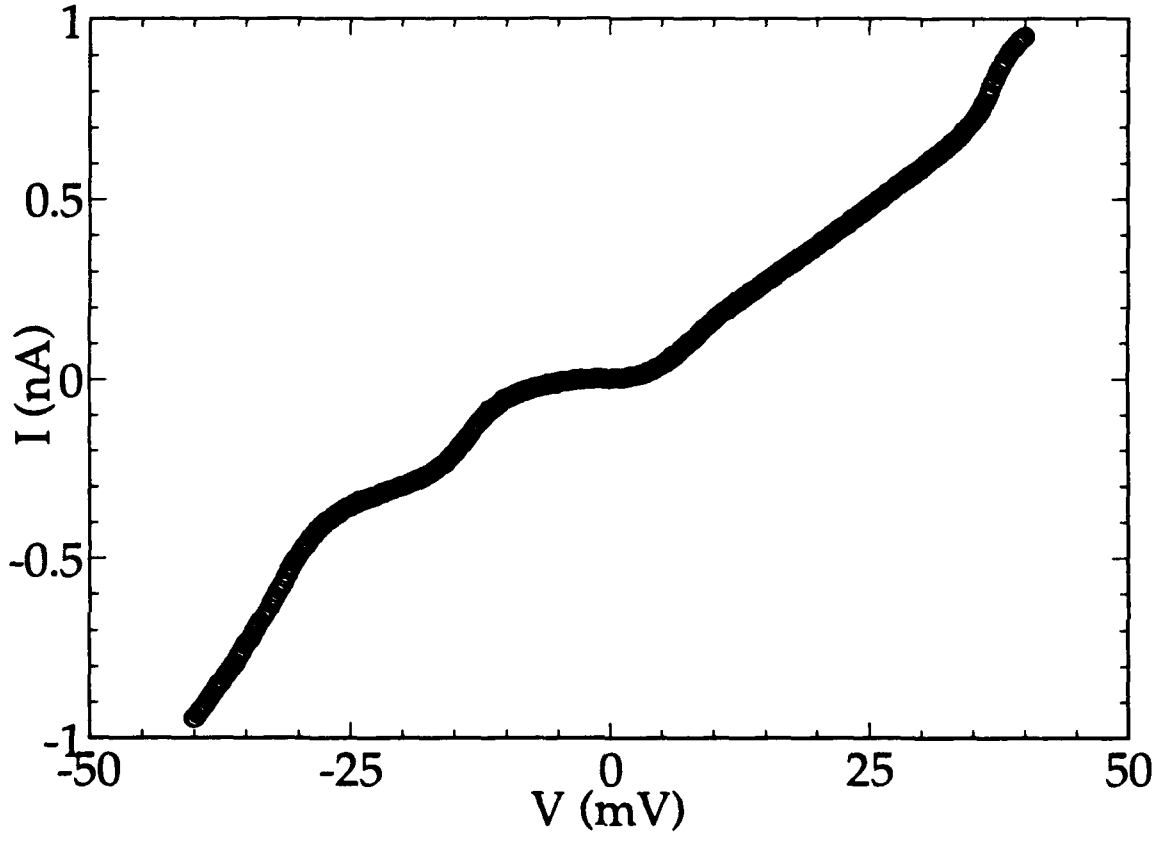


Figure 1.3b, I(V) characteristics II. These data illustrate the unusual variety of I(V) characteristics obtainable from this two-junction system. It shows asymmetric features, which despite the unusual appearance, can be well explained by the orthodox theory

current biased certain dynamical features are expected to be seen. In a current biased circuit, the current will charge up the junction until the voltage across the junction reaches a value at which it is energetically favorable to tunnel one electron. At this point the voltage will drop (by e/C), and the current will continue to charge up the junction until it once more becomes energetically favorable to tunnel one electron. For a constant current I, this will produce electrons tunneling across the junction with frequency f=I/e. This phenomenon is referred to as single electron tunneling (SET) oscillations.

The Coulomb blockade, Coulomb staircase and SET oscillations are all features of tunnel junction systems which are predicted by the orthodox theory. In chapter two, we present a semi-formal derivation of the orthodox theory of tunnel junctions. From the basic Hamiltonian we derive the effects observed above. We also go one step further than orthodox theory, which is first order perturbation theory, and calculate the next higher order processes. These processes, labeled "macroscopic quantum tunneling of charge" (q-MQT), show that there is a small conductance within the Coulomb blockade, where the orthodox theory predicts none.

In chapter three, the experimental apparatus is discussed. A description of the low temperature STM is given, and sample creation is discussed. Data measurement techniques are also explained in detail.

Chapter four begins by taking the limit of the orthodox theory equations in the range of our experiment. This gives a pair of simple analytic equations which we use to fit our data. The data is seen to be in excellent agreement with the calculations.

In our data, however, we observe a finite conductance within the Coulomb blockade region; it is presented in chapter five. This is direct contradiction with orthodox theory which predicts a definite zero conduction in this regime. By analyzing the data further, a strong case is made for the observation of *elastic* macroscopic quantum tunneling of charge. In theory, there are two types of q-MQT, elastic and inelastic; these are distinguished in chapter two. We observe both; however, the observation of elastic q-MQT is predicted to be much more difficult, due to the small signal and the necessary physical requirements. The parameters of our experiment allow detection of such a small effect. The large interest in small q-MQT effects, is that other than being of fundamental interest, q-MQT sets important limitations on the performance and accuracy of practical devices based on single electron effects.

Finally, in chapter six, we present our conclusions, and ideas for further research.

CHAPTER TWO

THEORETICAL OVERVIEW

2-1 INTRODUCTION

This chapter is devoted to elucidating the theoretical basis on which our data are interpreted. The chapter is not founded on detailed quantum mechanical formalism, rather, the major issues are motivated by physical arguments and references are provided for any further facts which the reader may desire to ascertain.

The principal analysis of the system is accomplished using perturbation theory. The Hamiltonian for the system is written down, and the term due to electron tunneling is considered as perturbing the Hamiltonian of isolated electrodes (without tunneling, so-called tunnel junctions are actually capacitors). The lowest order perturbation theory, labeled "orthodox theory", predicts the Coulomb blockade, Coulomb staircase and a dynamical property of tunnel junctions: "single electron tunneling (SET)" oscillations. The theory is quite successful in explaining much of our data.

Section 2-2 will describe in detail the orthodox theory derivation for the simplest case of a single current biased tunnel junction. A classical "master equation" is obtained, which when solved predicts the system properties. In section 2-3, the two junction system is discussed. Due to the similarity with the single junction case, only the extra terms in the two-junction Hamiltonian will be discussed, and the two-junction master

equation derivation will be less detailed. Since this is the system of most interest to us, the predicted junction properties, such as the Coulomb blockade, will be described in detail.

The next non-vanishing higher order perturbation theory, labeled "macroscopic quantum tunneling of charge (q-MQT)", predicts effects not obtainable from orthodox theory. In particular, a finite conductance within the Coulomb blockade region is predicted. This directly contradicts the prediction of zero current within the Coulomb blockade by orthodox theory. In section 2-4, I will explain the ideas motivating q-MQT, and how q-MQT exhibits a finite conductance in regions where orthodox theory predicts none.

2-2 ORTHODOX THEORY

The orthodox theory is the name given to the first order quantum mechanical perturbation theory treatment of tunnel junctions, in which the Hamiltonian term due to electron tunneling perturbs the Hamiltonian of the isolated electrodes.

The usefulness of the orthodox theory is that under certain limits, which are typical experimental constraints, the junction can be treated in terms of a classical probability density, $\sigma(Q,t)$, where Q is the charge on the junction. The equation of motion for σ yields a "master equation," which when solved predicts the junction properties.

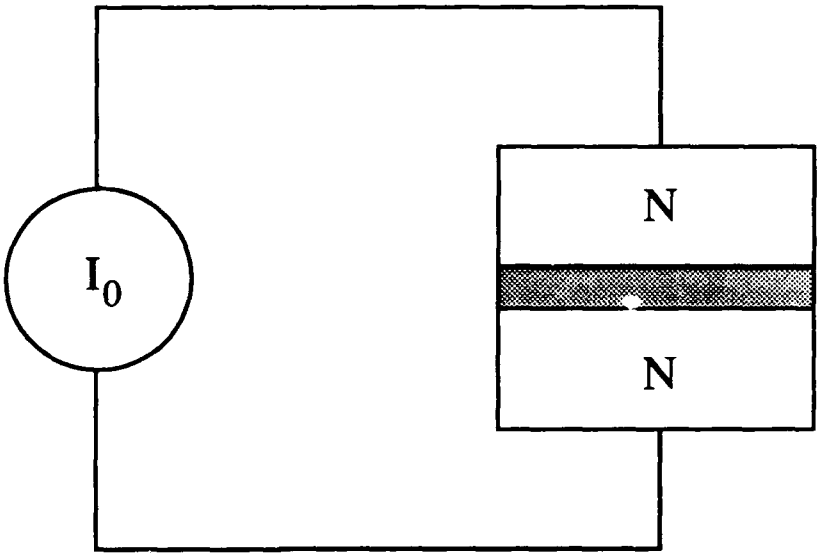


Figure 2.1, Current biased single junction. This figure shows the simplest case for analysis: a single tunnel junction which is DC current biased.

2-2.1 Current-biased Single Junction

Hamiltonian

This section will detail orthodox throry predictions for a single normal-metal tunnel junction with an external bias I_0, as shown in Fig. 2.1. The

orthodox theory starts with the following Hamiltonian,[Averin and Likharev, 1986, 1991; Likharev, 1988]†

$$H = H_0 + H_T - I\Phi \tag{2.1}$$

$$H_0 = H_1\{k_1\} + H_2\{k_2\} + \frac{Q^2}{2C} \tag{2.2}$$

H_0 represents the unperturbed Hamiltonian and H_T and $I\Phi$ are the perturbing terms. H_1 and H_2 represent the energy of the internal degrees of freedom, $\{k_1\}$ and $\{k_2\}$, of the two junction electrodes respectively. The discreteness of the transfer of electric charge is apparent in H_T,

$$H_T = H_+ + H_-, \quad H_+ = \sum_{k_1,k_2} T_{k_1k_2} c^\dagger_{k_1} c_{k_2}, \quad H_- = H_+^\dagger \tag{2.3}$$

which is a sum of independent contributions from transfers of single electrons from states $\{k_2\}$ inside one electrode to states $\{k_1\}$ inside the other, and vice versa. $c^\dagger_k$ and c_k are electron creation and annihilation operators for state k.

The operator Q represents the electric charge on the junction, and can be expressed using the same creation and annihilation operators as H_T,

$$Q = \frac{e}{2}\left(\sum_{k_1} c^\dagger_{k_1} c_{k_1} - \sum_{k_2} c^\dagger_{k_2} c_{k_2}\right) \tag{2.4}$$

---

† There is also a term due to the shunt current which I have neglected in this presentation, (i.e., $G_S = 0$).

so that H_T and Q do *not* commute. Using Fermi commutation rules, it can readily be shown for an arbitrary function B(Q),

$$H_{\pm}B(Q) = B(Q \pm e)H_{\pm} \tag{2.5}$$

Generally, H_1 and H_2 do not necessarily commute with Q either. However, in any imaginable junction the number of electron states in the electrodes, N, is so large that a realistic charge $Q \ll eN$ does not change their internal properties. In this case, the approximation

$$[H_{1,2},Q] \approx 0 \tag{2.6}$$

is valid. The current I, is simply given by, $I = I_0(t)$, and Φ is the operator defined as,

$$\Phi \equiv \int_{-\infty}^{t} V\, dt' \tag{2.7}$$

where $V = Q/C$ is the operator for the voltage across the junction. The operators Φ and Q are related by the commutation relation (2.8) [Anderson, 1964],

$$[\Phi,Q] = i\hbar \tag{2.8}$$

However, one can see that this model neglects several factors:

(i) dimensions and shape peculiarities of the tunnel junction. This model is, in fact, zero dimensional.

(ii) a non-vanishing duration of the the electron tunneling process. $c_k^{\dagger}$ and c_k are taken at coinciding points in time.

(iii) a non-vanishing duration of the electric charge redistribution inside the electrodes after the tunnelling event. (An instant formation of the surface charge distribution).

(iv) the quantized energy levels in the electrodes. The energy spectra of the operators H_1 and H_2 are taken to be continuous.

(v) the quantum nature of the source current $I_0(t)$.

Despite these shortcomings, the orthodox theory has proved to be quite effective for calculating the tunneling current in small junctions. It is also relatively simple and has physically interpretable results.

Master Equation

The basic equations (2.2), (2.5), (2.6), and (2.8) allow one to calculate all the properties of the junction without further utilization of the explicit expressions (2.3) and (2.4). The calculation is performed using a density matrix written in the Q-basis, $\rho(Q,Q',t)$. This is done because the charge state of junction is the important variable, and not the particular collection of electron energy eigenstates which make up the charge state. Since an individual charge state can correspond to several different occupation arrangements of electron energy eigenstates, the density matrix in the charge basis is obtained by tracing over the internal states of the von Neumann density matrix.§

---

§ $\rho(k_1,k_2,t) = \langle k_1|\Psi(t)|k_2\rangle$

$$\rho(Q,Q',t) = \mathrm{Tr}_{k_1 k_2}\, \rho(k_1,k_2,t) \tag{2.9}$$

In the case of small tunneling conductances, $G_T \ll R_Q^{-1}$ ($R_Q = \pi\hbar/2e^2$) and low tunnel current $I \ll e/R_QC$, which classically correspond to slowly varying the charge on the junction, Q, (i.e., the terms which change Q, tunneling and current, are both small), the density matrix $\rho(Q,Q',t)$ evolves slowly in time. In the interaction picture, where operators evolve in time using the unperturbed Hamiltonian and wave functions evolve using the Hamiltonian perturbation, the evolution of $\rho(Q,Q',t)$ can be described by,

$$\dot{\rho} = F_I + F_T \tag{2.10}$$

The term F_I describes the effects of the bias current and F_T describes tunneling effects. Their contributions to $\dot{\rho}$ can be calculated independently in the first non-vanishing order of standard perturbation techniques [Averin and Likharev, 1986].

$$F_I = -\frac{iI_0(t)}{\hbar}[\Phi(t),\rho(t)] \tag{2.11}$$

$$F_T(Q,Q,'t) = \frac{-1}{\hbar^2}\int_{-\infty}^{t} dt'\, \mathrm{Tr}_{k_1,k_2}\{[H_T(t),[H_T(t'),\rho(Q,Q',t')f]]\} \tag{2.12}$$

f is the thermal-equilibrium density matrix of the electrodes,

$$f\{k_1,k_2\} \equiv Z^{-1}\exp\{-(H_1+H_2)/k_BT\} \tag{2.13}$$

$$Z \equiv \mathrm{Tr}_{\{k_1\},\{k_2\}}\exp\{-(H_1+H_2)/k_BT\}. \tag{2.14}$$

Substituting for H_T in integral (2.12) shows that F_T relates only those terms $\rho(Q,Q',t)$ of the density matrix whose arguments differ by $\pm e$ [Averin

and Likharev, 1986]. That is, F_T relates only elements with equal deviation from the principal diagonal Q = Q′. Hence, single-electron tunneling *cannot* induce non-diagonal elements of the density matrix. Similar analysis shows that F_I has exactly the same property. Thus, if $\rho(Q,Q',t)$ is a diagonal matrix at $t = -\infty$ then it will continue to be diagonal at all later times. Consequently, the resulting expression for $\rho(Q,Q',t)$ is then given by a diagonal matrix,[1]

$$\rho(Q,Q',t) = \sigma(Q,t)\delta(Q-Q'). \tag{2.15}$$

The condition $G_T \ll R_Q^{-1}$ allows the approximation[2] $\rho(t) = \rho(t')$. The integrals in Eqn. (2.12) can now be solved [Rogovin and Scalapino, 1974]. One can thus reduce the set of equations for ρ, to a differential master equation [Averin and Likharev, 1986] for the classical probability density $\sigma(Q,t)$.

$$\frac{\partial\sigma(Q)}{\partial t} = -I(t)\frac{\partial\sigma}{\partial Q} + F_T(Q) \tag{2.16}$$

$$F_T(Q) = \Gamma^+(Q-e)\sigma(Q-e) + \Gamma^-(Q+e)\sigma(Q+e) - [\Gamma^+(Q) + \Gamma^-(Q)]\sigma(Q) \tag{2.17}$$

Γ^+ is the tunneling rate *increasing* Q by e (an electron tunneling from electrode 1 to 2) and Γ^- is the tunneling rate *decreasing* Q by e (an electron

---

[1] In the case of a finite shunt conductance, the assumption that $\rho(Q,Q',t)$ is diagonal at $t=-\infty$ is unnecessary, since any non-diagonal elements can be shown to decay with a time constant of $\tau_s = C/G_S$ if $G_S R_Q \ll 1$.

[2] This is the Markoffian approximation, which states that the tunneling creation and annihilation operators are evolving in time much faster than ρ is evolving in time (i.e., tunneling time is much shorter than quantum fluctuation time). Thus, the system dynamics are mostly caused by those operators' evolution, and as such, ρ can be taken as constant, to solve the integral in Eqn. (2.10).

tunneling from electrode 2 to 1). Equation (2.17) can thus be read as: 'the total rate of change of the probability of occupying state Q equals the rate of transitions to state Q from the (Q - e) state plus the rate of transitions to state Q from the (Q + e) state minus the rate of transitions from the state Q to the (Q ± e) state.'

A basic "golden-rule" calculation is used to solve for the general tunneling rate $\Gamma^{\pm}$ across a normal-metal tunnel junction,

$$\Gamma^{+} = \int_{-\infty}^{+\infty} \frac{2\pi}{\hbar} |T(E)|^2 D_1(E - E_1) f(E - E_1) D_2(E - E_2)[1 - f(E - E_2)] dE \qquad (2.18)$$

where $D_1(E)$ and $D_2(E)$ are the normal metal density of states on electrodes 1 and 2 of the junction. E_1 and E_2 are the Fermi energies on the electrodes 1 and 2. And, f(E) is the Fermi distribution function. Γ^{-} is obtained by interchanging the indices 1 and 2. Under the assumptions that D_1 = constant, D_2 = constant, and $|T(E)|$ = constant, the above integral (2.18) can be solved [Lambe and Jakelvic, 1968] to obtain,

$$\Gamma^{\pm}(Q) = \frac{V}{eR_T} \left[1 - \exp\left(-\frac{\Delta E^{\pm}}{k_B T} \right) \right]^{-1} \qquad (2.19)$$

The tunneling resistance R_T incorporates all the constants, ΔE is the change in electrostatic energy due to a tunneling event,

$$\Delta E^{\pm} = \frac{Q^2}{2C} - \frac{(Q \pm e)^2}{2C} = \mp \frac{e}{C}(Q \pm e/2) \qquad (2.20)$$

and the relation $I_T(V) = V/R_T$ has been used for a normal-metal junction.

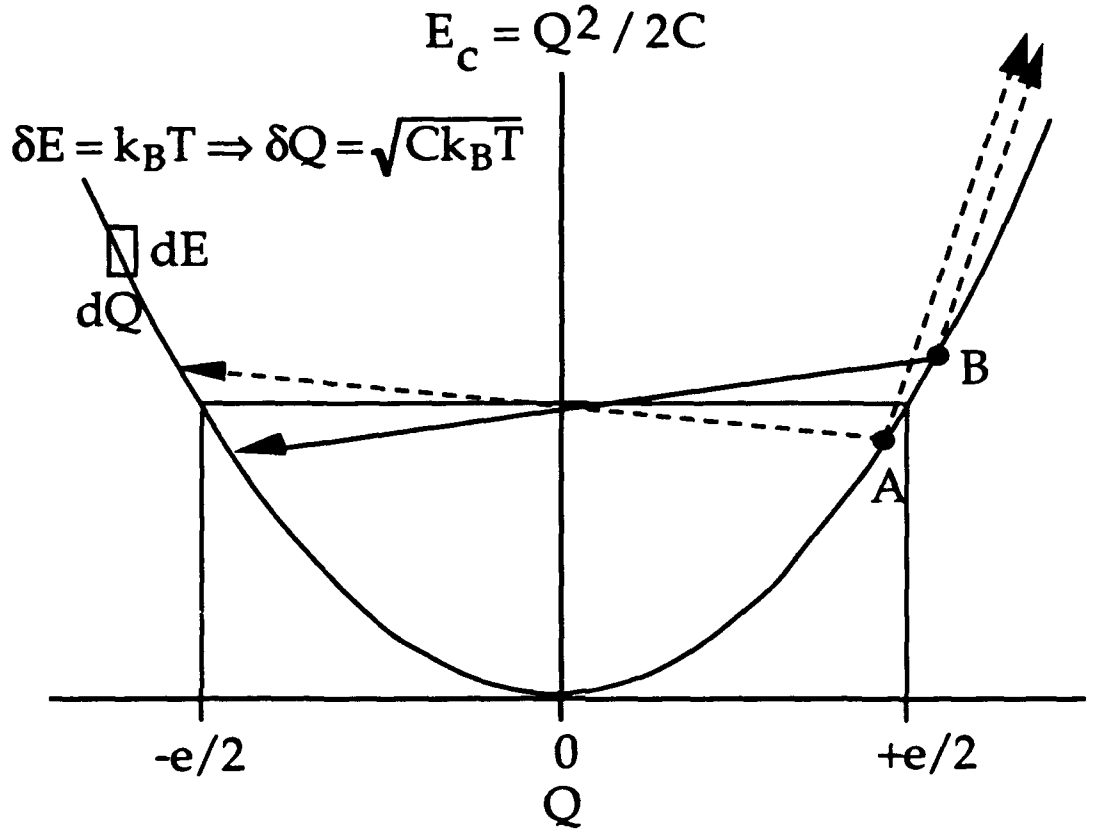


Figure 2.2, Coulomb blockade of single tunnel junction. The above diagram shows the effect of a single electron transfer from two initial charge states, A and B. The first, A, is when Q is less than e/2; the addition or subtraction of a single electron is energetically forbidden (dashed lines), thus the junction has no current flow. This region, [-e/2,e/2], is called the Coulomb blockade region. The second case, B, is when Q is larger than e/2; here, it is energetically favorable for one electron to tunnel through the junction (solid line). As can be calculated from the diagram the voltage thresholds for current flow are $V_T = \pm e/2C$.

From these relations, it can be shown that the solution for $\sigma(Q,t)$ is a Gaussian centered around a value Q_c and with width $\delta Q = (Ck_BT)^{1/2}$, where $\dot{Q}_c = I(t)$. The width δQ is obtained by relating thermal energy fluctuations, $\delta E = k_BT$, to the corresponding charge fluctuation as given by Eqn. (2.20). For single electron effects to be observable, the condition $\delta Q \ll e$, must be satisfied. In the case of $I(t) = 0$, the physical origin of the Coulomb blockade of single-electron tunneling is apparent. The energy change (2.20) is positive for $\Delta Q = \pm e$ if the initial charge Q is within the limits $|Q| \leq e/2$, thus at low temperatures ($k_BT \ll \Delta E$) single-electron tunneling is not allowed ($F_T = 0 \Rightarrow \dot{\sigma}$, thus no current flow). This situation is depicted in Fig. 2.2.

When $I(t) \neq 0$ the average value of Q_c will move until it reaches Coulomb blockade boundary, $\pm e/2$. At this point F_T will grow sharply providing a rapid transfer of the probability packet to a new, lower energy, location close to the other boundary. Physically, this transfer is the tunneling of a single electron. The new packet enters the blockade range, moves through it with speed $\dot{Q}_c = I(t)$ until it reaches the other boundary where the whole process is repeated. This periodic charge build-up and tunneling process is depicted in Fig. 2.3, and has been labeled single electron tunneling oscillations (SET) [see Averin and Korotkov, 1990]. The frequency of the oscillations in Q, and corresponding voltage oscillations, is just,

$$f_{SET} = \frac{I}{e} \tag{2.21}$$

The origin of the SET oscillations is of course the Coulomb blockade. In a current-biased junction with small capacitance C, a tunneling event leads to a noticeable change of the voltage V across the junction, ($\Delta V = \pm e/C$). If this

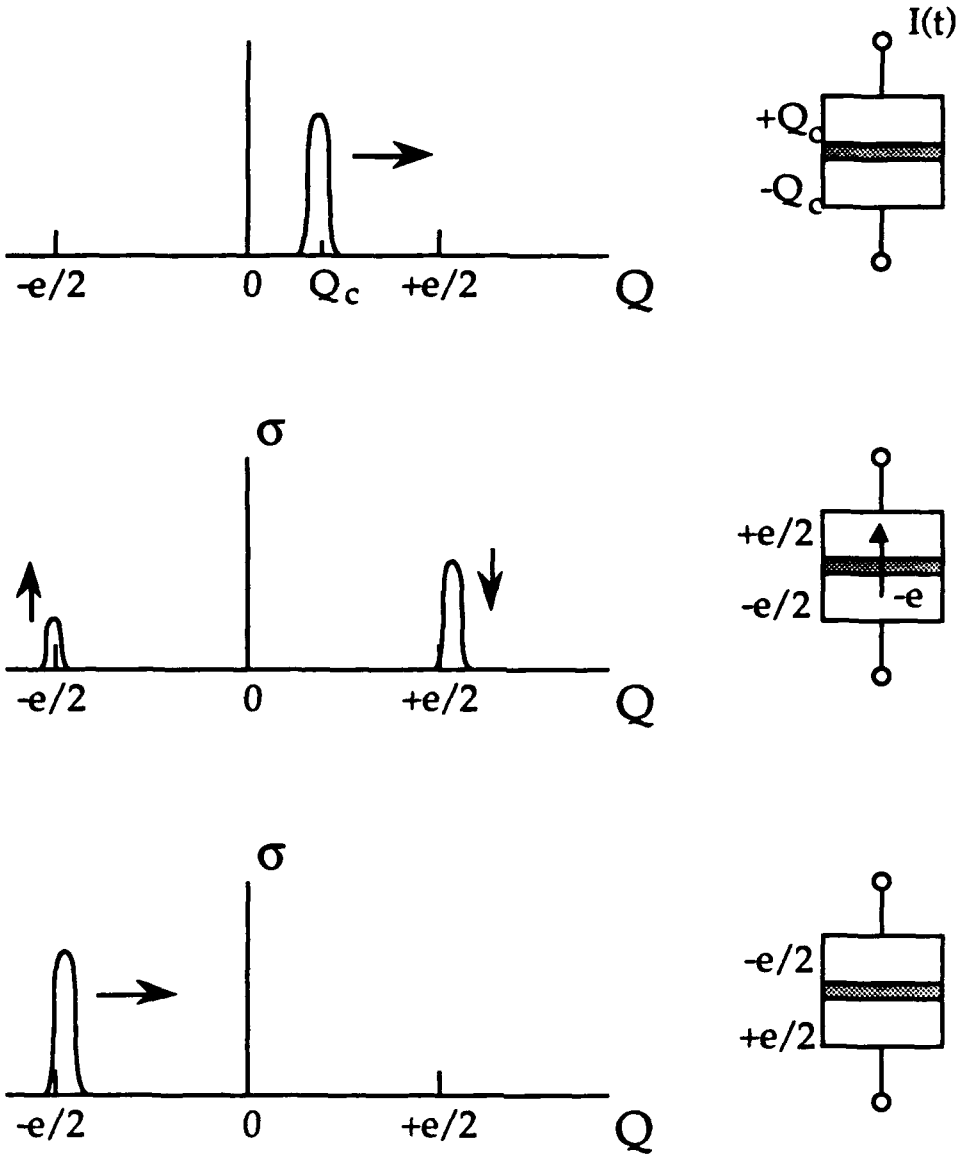


Figure 2.3, Single electron tunneling (SET) oscillations. A current biased single normal-metal junction should exhibit voltage oscillations. This is due to the applied current charging up the junction until the Coulomb threshold voltage (Fig. 2.2) is reached, at which point an electron can tunnel and the voltage is changed by e/C. The frequency of the the voltage oscillations is given by Eqn. (2.21).

change is larger than the masking thermal fluctuations, the other electrons are affected by this change, and a strong correlation of tunneling events is established. After one tunneling event, the other electrons are blocked until an additional charge of ±e is added through the continuous current source. This correlation results in the SET oscillations with a transfer of exactly one electron through the junction per given time period.

For a single junction, the Coulomb blockade and SET oscillations are the significant single electron tunneling effects predicted by orthodox theory. These effects are, however, quite difficult to detect because the capacitance used in the above equations is essentially the total capacitance between the electrodes, and should include not only the junction capacitance C_J, but also the stray capacitance of the leads C_{leads}, which may be quite large. In fact, to observe single-electron effects at liquid helium temperatures, the total capacitance should be of order 10^{-16}F. This corresponds to lead lengths of about 10μm, which is difficult to realize experimentally.

The difficulties associated with a single junction system can be overcome by using a two junction system. In the next section, the master equation for a two junction system is calculated and its expected properties predicted.

2-2.2 Double Junction

In a two junction system, shown in Fig. 2.4, the effects of large lead capacitances do not matter. In this system it is the net capacitance between the middle electrode and its environment, C_Σ, which enters into the charging energy equations. C_{leads} does not contribute to C_Σ, where $C_\Sigma = C_1 + C_2$, $C_{1,2}$

being the capacitance of the individual junctions 1 and 2, respectively. Accordingly, this system should allow easier observation of single-electron effects.

Hamiltonian

The Hamiltonian for this system is [Likharev, 1988; Averin and Likharev, 1991],

$$H = G + \sum_{i=0,1,2} H_i\{k_i\} + \sum_{j=1,2} (H_T)_j \tag{2.22}$$

$$G = \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} - Q_m V = \frac{Q^2}{2C_\Sigma} - Q_m V + \text{constant} \tag{2.23}$$

$H\{k_i\}$ describes the electrodes, H_{T_i} describes the tunneling across each junction (as given in Eqn. (2.2) for a single junction), $Q_{1,2}$ describe the charge on junctions 1 and 2, respectively, and Q_m is charge passed through the system (i.e., the charge passed through the source). Q_m can be expressed in terms of n_1 and n_2 the numbers of electrons passed through the junctions,

$$Q_m = e\left(n_1 \frac{C_2}{C_\Sigma} + n_2 \frac{C_1}{C_\Sigma}\right) \tag{2.24}$$

The constant in Eqn. (2.24) is to account for any other possible source of charge, these sources will be explored in detail below. These same numbers, n_1 and n_2, are used to calculate the net charge on the middle electrode, Q,

$$Q = en + Q_0, \qquad n = n_1 - n_2 \tag{2.25}$$

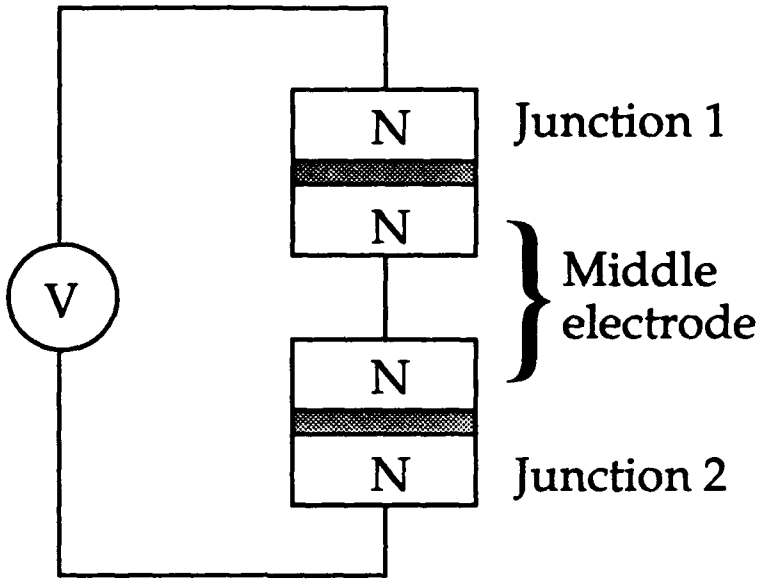


Figure 2.4, Two normal-metal tunnel junction in series. A linear chain of two junctions, which is voltage biased, is the simplest system in which to observe single electron processes without the destructive effects of the lead capacitance.

Q_0 is the due to the sum of the effects from the background charge which may be due to charge impurities in the tunnel barrier and the contact potential of the different electrode materials. The precise nature and significance of Q_0 will be discussed in more detail in chapter four, but what is important is that both components of Q_0 are continuous on the scale of the electron charge, e, so that Q_0/e can be a fraction. The Hamiltonian (2.22) is

very similar to that discussed in the previous section and thus for $G_1, G_2 \ll R_Q$, similar perturbation techniques may be used to obtain the properties of interest.

Master Equation

Following similar procedures outlined in section 2-1, I obtain the following master equation for the probability density $\sigma(n,t)$,

$$\frac{\partial\sigma(n)}{\partial t} = \sum_{j=1,2}\{\Gamma_j^+(n-1)\sigma(n-1) + \Gamma_j^-(n+1)\sigma(n+1) - [\Gamma_j^+(n) + \Gamma_j^-(n)]\sigma(n)\} \qquad (2.26)$$

where $\Gamma_j^{\pm}$ are the rates of the single-electron tunneling through the jth junction as given by Eqn. (2.29). The change in energy ΔE in Eqn. (2.20) is replaced in Eqn. (2.23) by $\Delta G_j^{\pm}$, the change in G due to the a single tunneling event ($\Delta n_j = \pm 1$). $\Delta G_j^{\pm}$ are given by,

$$\Delta G_1^{\pm} = \Delta U^{\pm} \pm \frac{eC_2}{C_\Sigma}V \qquad (2.27a)$$

$$\Delta G_2^{\pm} = \Delta U^{\pm} \mp \frac{eC_1}{C_\Sigma}V, \qquad (2.27b)$$

where

$$\Delta U^{\pm} = \frac{Q^2}{2C_\Sigma} - \frac{(Q \pm e)^2}{2C_\Sigma}. \qquad (2.27c)$$

The contribution from $\Delta U^{\pm}$ is the change in the charging energy of the middle electrode, and the contribution from the bias across the junctions is the work done by the voltage source. Fig. 2.5 shows the phase diagram of this

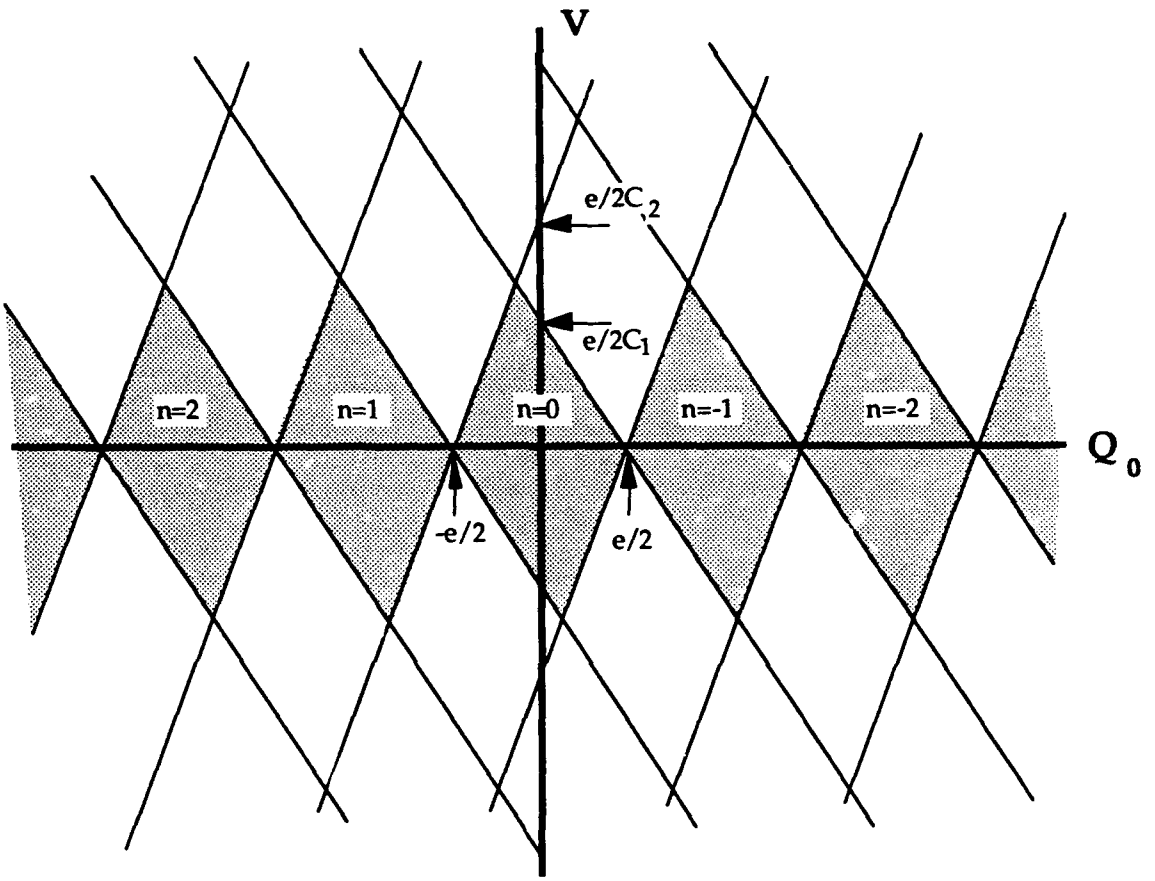


Figure 2.5, Two junction phase space. This diagram shows the equipotential lines where $\Delta G^{\pm}_{1,2} = 0$ for $T = 0$. The shaded sections are Coulomb blockade regions where there is no current flow despite an applied voltage.

double junction system, where the straight lines correspond to equations $\Delta G^{\pm}_{1,2} = 0$, for various values of n.

Within the rhombic-shaped regions covering the Q_0 axis, there exists a value of n which provides the equilibrium state of the junction ($\Delta G^{\pm}_{1,2} < 0$). Inside these regions, $\Gamma^{\pm}_{1,2}(n)$ all vanish (for $k_BT \ll e^2/2C_\Sigma$) so that the current through the junctions,

$$\bar{I}_j(V) = e \sum_{n=-\infty}^{\infty} \sigma(n,t)\left[\Gamma_j^+(n) - \Gamma_j^-(n)\right] \tag{2.28}$$

vanishes.[3] These zero current regions correspond to the Coulomb blockade of tunneling, similar to that discussed for a single junction earlier. For the two junction system, in order for an electron to travel through both junctions (to a lower energy state), it must first enter the middle electrode and hence increase its energy by $\sim \Delta U \approx e^2/2C_\Sigma$. For low voltages, this charging is energetically unfavorable, and as such forbidden at low temperatures. This situation is depicted in Fig. 2.6.

When the applied voltage V exceeds the Coulomb blockade thresholds, $V_C^{\pm}$, which are functions of Q_0 as shown in Fig. 2.5, a constant transfer of single electrons through the junctions begins. This situation is depicted in Fig. 2.7. In this situation, there is a high degree of mutual correlation of the tunneling events; as soon as an electron enters the middle electrode through

---

[3] Eqn. (2.26) is the general equation for the current through the system. The bar over the I_j in is to average over time scales larger than the tunneling time for individual events.

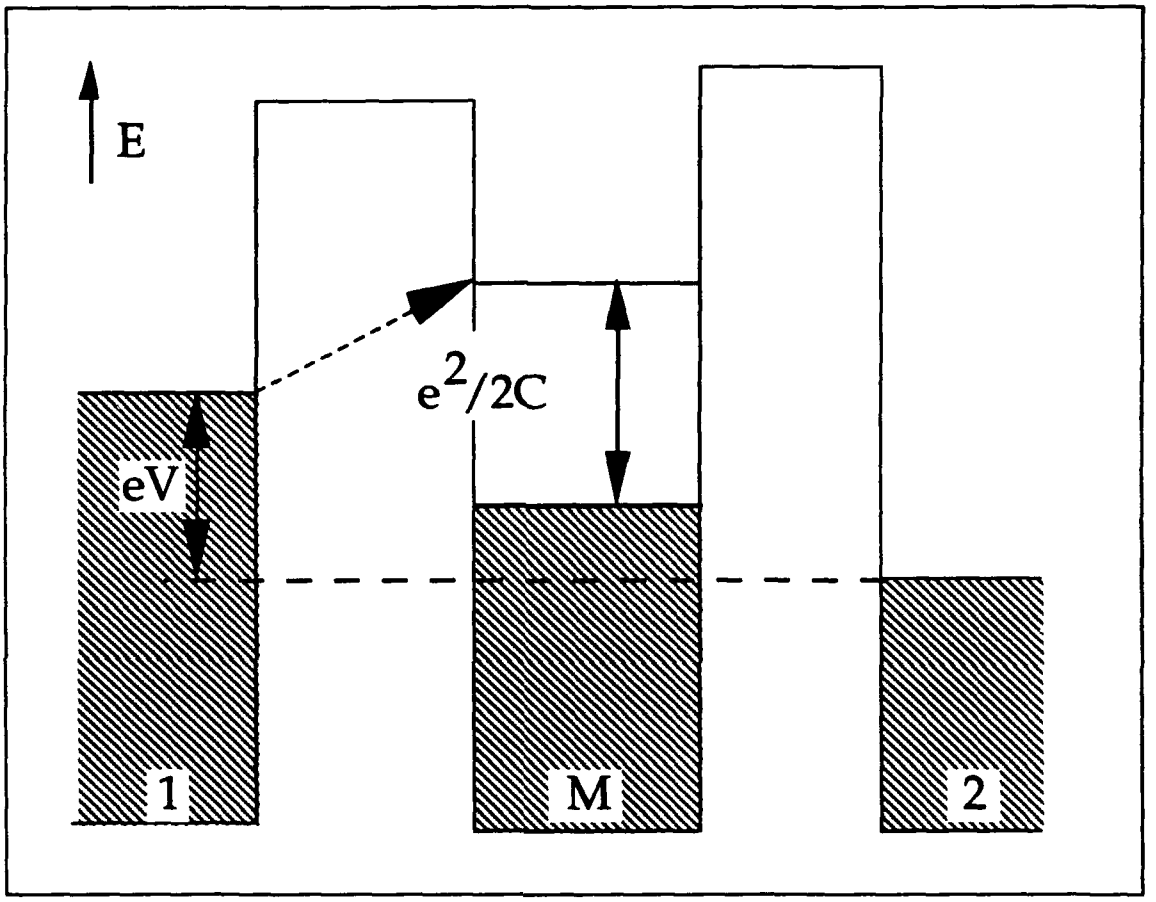


Figure 2.6, Coulomb blockade state. This figure shows energy level diagrams for a two junction system biased in the Coulomb blockade regime. The tunneling of a single electron onto the middle electrode is energetically unfavorable, despite an applied non-zero voltage, V. The two electrodes are labeled 1 and 2, and the middle electrode is labeled by M. The hatched areas represent filled electron states.

one of the junctions, it acquires a large probability rate Γ for immediately leaving the electrode through the other junction. At the same time there is a very little time correlation between events; there is no particular frequency of the electron tunneling events. This is because in the voltage biased situation considered, the transfer of a single electron across the junctions does not change the voltage V across the junctions and hence does not affect the consequent tunneling events. Coherent oscillations (similar to SET oscillations in a single junction) would, of course, appear if the system were DC current biased [Likharev, 1988; Averin and Likharev, 1991].

The mutual correlation of tunneling events causes the number of electrons on the middle electrode to be sharply defined. The variation of this number is responsible for a characteristic staircase pattern in the I(V), as well as a strong influence of Q_0 on the I(V). Fig 2.8 shows simulated I(V) curves for integer and half-integer values of Q_0, in a two junction system. The type of I(V) shown in Fig. 2.8 has been labeled the "Coulomb staircase", because each step in the I(V) corresponds to the average value on the number of electrons on the middle electrode, <n>, differing by one. The detailed effects of Q_0 on the I(V) will be discussed more in Chapter four.

2-3 Q-MQT Theory

As mentioned in the previous section, a voltage biased series of two tunnel junctions exhibits a Coulomb blockade of tunneling in which the current flow through the system is blocked by the energy cost associated with the single tunneling event of adding (or subtracting) one electron to the

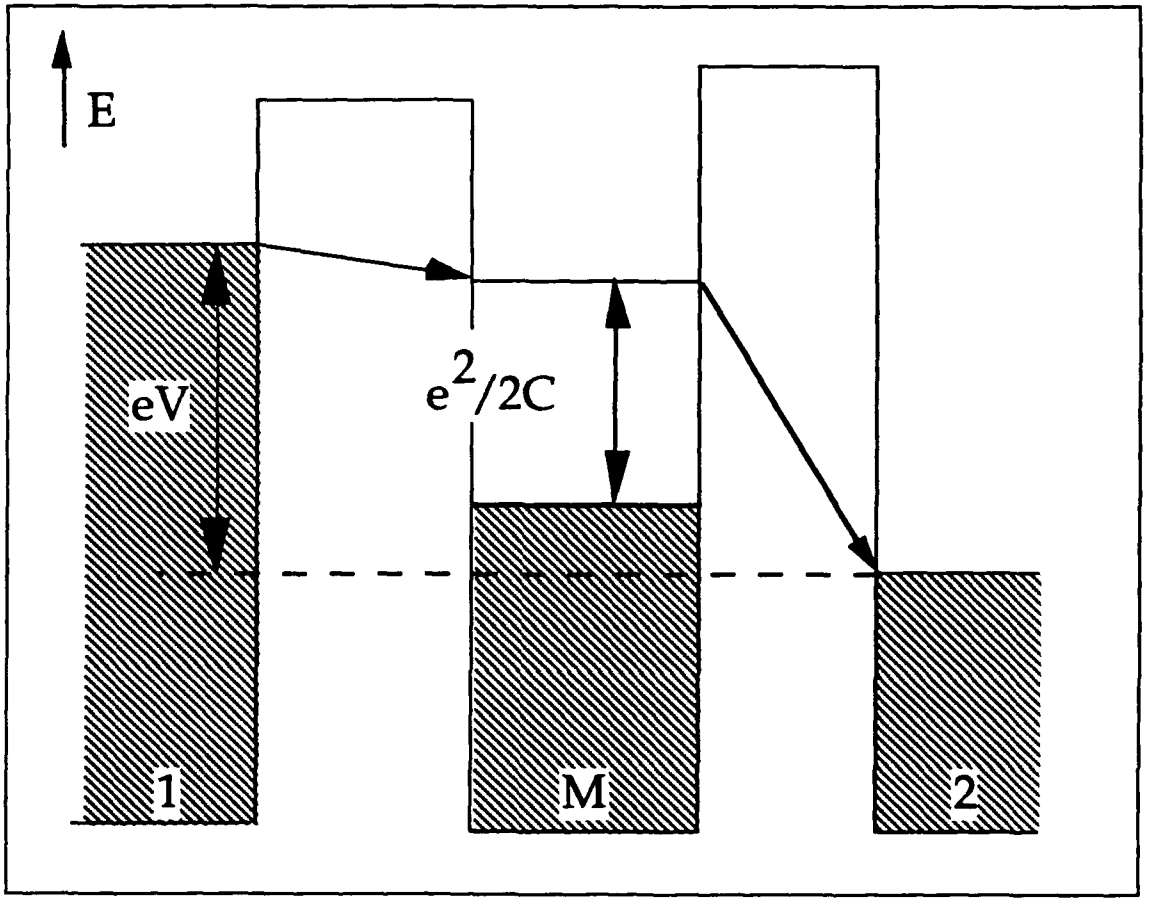


Figure 2.7, Overcoming the Coulomb barrier. In the situation depicted above the applied voltage is large enough so that it is energetically favorable for the electron to enter onto the middle electrode. Once there, the electron has is a large probability to tunnel once more to the other electrode, resulting in a current flow through the entire device. The two electrodes are labeled 1 and 2, and the middle electrode is labeled by M. The hatched areas represent filled electron states.

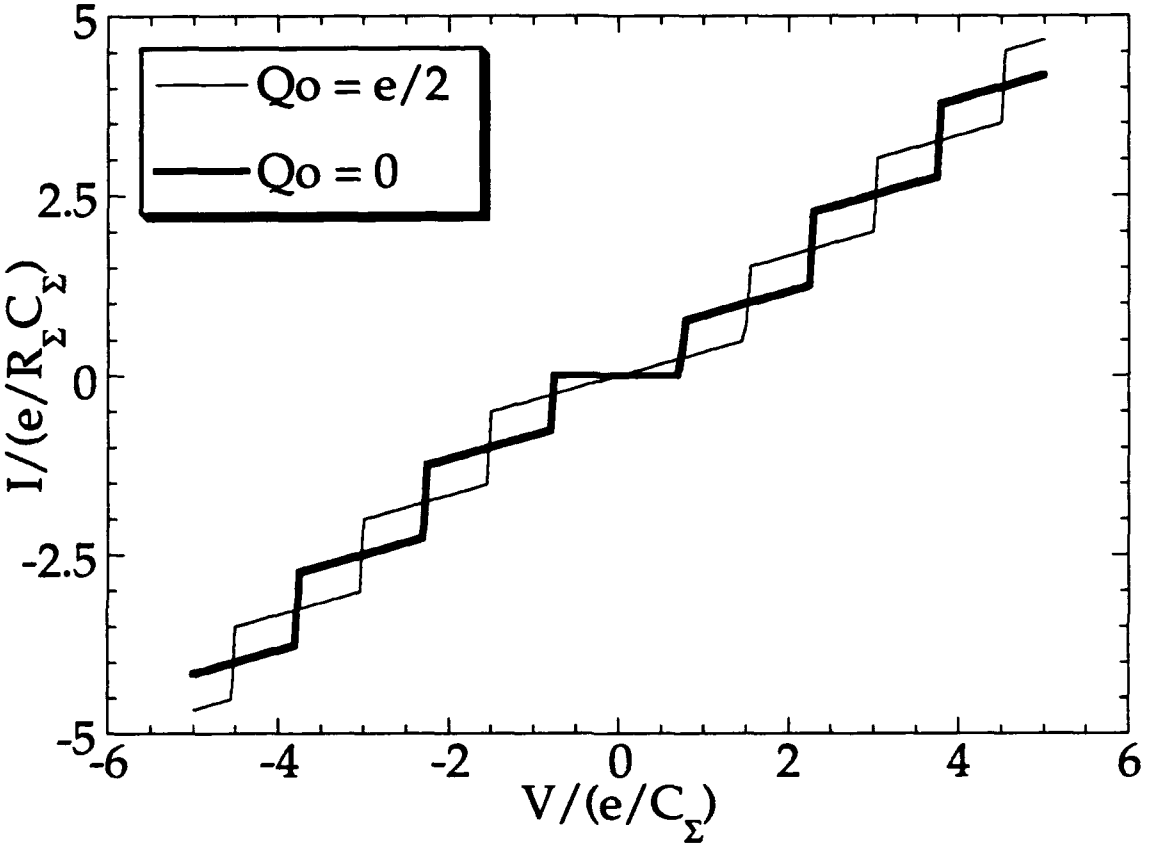


Figure 2.8, Coulomb staircase structure. The above is a simulated I(V) calculated from the orthodox theory for a two junction system. The most striking feature is the staircase structure, which has a strong dependence on the value of Q_0, particularly within the low voltage region. Each step on the staircase corresponds to a shift of the average number of electrons on the middle electrode by one.

middle electrode. It is, however, energetically favorable for the electron to tunnel through the whole system. That is, an electron could tunnel through the barrier associated with each junction and also through the effective barrier arising due to the Coulomb interaction. As a result of the much longer tunneling path, this second sort of tunneling, which has been labeled macroscopic quantum tunneling of charge (q-MQT), has a much lower rate in comparison with measured single junction tunneling rates. In this section, I will outline the calculation for the expected rate of q-MQT events, which give rise to a finite current within the Coulomb blockade regime.

The q-MQT current arises from quantum fluctuations of the macroscopic variable Q, the electric charge on the middle electrode. An electron can tunnel though the whole system via a virtual intermediate state with increased electrostatic energy. This process gives rise to two components to the q-MQT current: (i) an *elastic* component, in which the tunneling electron maintains its phase information between the two tunneling events, and (ii) an *inelastic* component, in which the phase information of the tunneling electron is lost between the individual tunneling events.

The Hamiltonian of the the two junction system is still that given in Eqn (2.22). The tunneling Hamiltonian for each junction, H_{T_i}, is given by

$$H_{T_i} = H_i^+ + H_i^-, \quad H_i^+ = \sum_{m,n} T^i_{m,n} c^\dagger_m c_n, \quad H_i^- = (H_i^+)^\dagger \tag{2.29}$$

and $H_T \equiv H_{T1} + H_{T2}$. Again H_{T_i} can be treated as a perturbation because of the small conductances considered, $G_i \ll 1/R_Q$. Since one act of q-MQT involves

two electron tunneling events, the current associated with this tunneling is of fourth order in H_T [Averin and Nazarov, 1990].

$$I = \frac{2e}{\hbar^4} \mathrm{Re}\left(\int_{-\infty}^{t} d\tau \int_{-\infty}^{\tau} d\tau' \int_{-\infty}^{\tau'} d\tau'' \left[\begin{array}{l} \left\langle H_I^{\dagger}(t) H_T(\tau) H_T(\tau') H_T(\tau'') \right\rangle - \left\langle H_T(\tau'') H_T(\tau') H_T(\tau) H_I^{\dagger}(t) \right\rangle \\ + \left\langle H_T(\tau'') H_T(\tau') H_I^{\dagger}(t) H_T(\tau) \right\rangle - \left\langle H_T(\tau) H_I^{\dagger}(t) H_T(\tau') H_T(\tau'') \right\rangle \end{array} \right] \right) \tag{2.30}$$

where the time dependence of H_T is determined by H_0 and the average is taken over the equilibrium density matrix corresponding to H_0. As in the previous section, the next step is to substitute in for H_T and H_I in Eqn. (2.30), and to collect the terms that describe the q-MQT process. Applying Wick's theorem to the averages, two distinct contributions to the tunneling current are obtained. For example [Averin and Nazarov, 1990],

$$T_{km}^{(1)} T^{*(1)}_{k'm} T_{nl}^{(2)} T^{*(2)}_{nl'} \left\langle c_l^{\dagger} c_k c_l c_{k'}^{\dagger} \right\rangle \rightarrow - \left| T_{km}^{(1)} \right|^2 \left| T_{nl}^{(2)} \right|^2 \left\langle c_l^{\dagger} c_l \right\rangle \left\langle c_k c_k^{\dagger} \right\rangle + T_{km}^{(1)} T_{lm}^{*(1)} T_{nk}^{(2)} T_{nl}^{*(2)} \left\langle c_k^{\dagger} c_k \right\rangle \tag{2.31}$$

where the indices m, n refer to energy eigenstates of the external electrode, and k, l refer to those on the middle electrode. The terms obtained from substitution into Eqn. (2.30) which resemble the first term on the right in Eqn. (2.31), only depend on the absolute value of the tunneling amplitudes $T^{(i)}$ and hence describe tunneling without any coherence between tunneling events in the two junctions. Basically, two independent tunneling events occur in the two junctions† One tunneling event has an electron enter the middle electrode above the Fermi level, and the other has an electron leave

† Roughly, this means two different electrons tunnel across the two junctions.

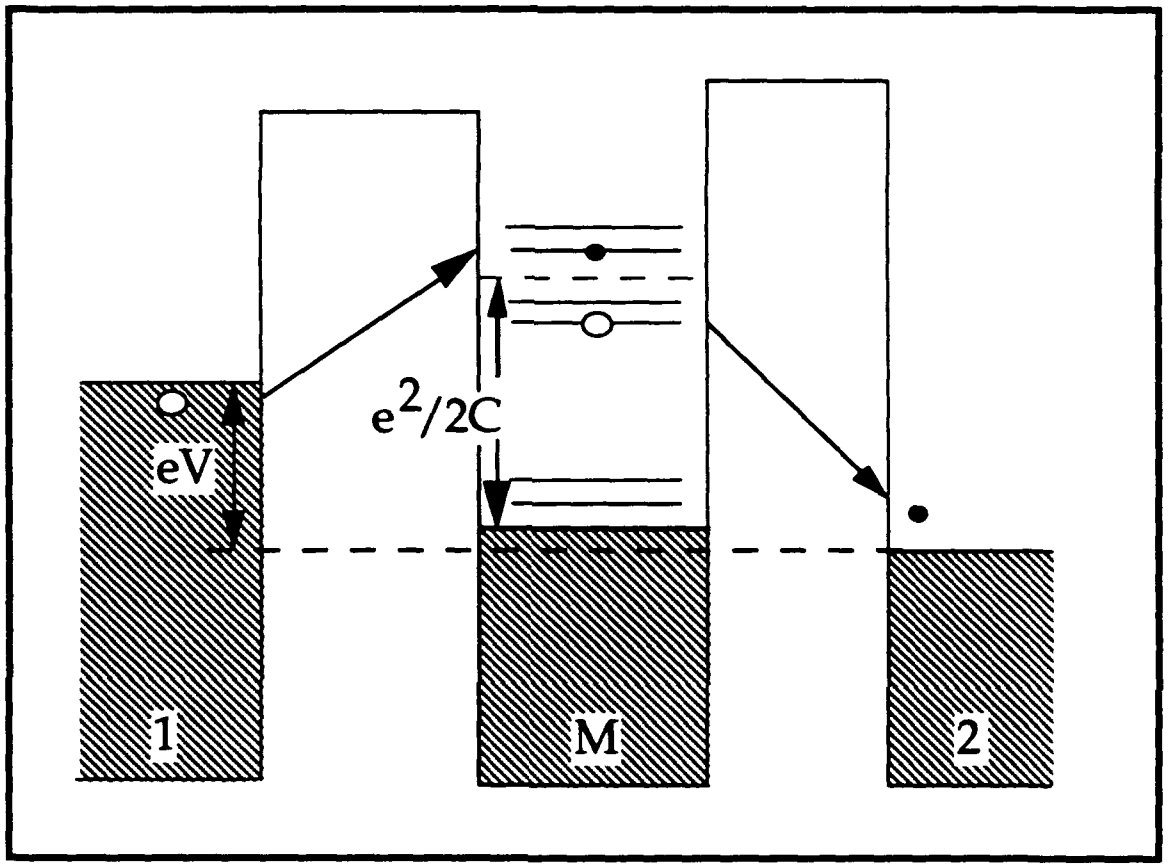


Figure 2.9, Inelastic q-MQT process. This diagram represents the process of inelastic q-MQT. An electron (solid circle) from electrode 1 will enter the middle electrode above the Fermi surface, despite the increase in energy. An electron below the Fermi surface of the middle electrode will tunnel to electrode 2, leaving an electron-hole excitation on the middle electrode.

from below its Fermi level. This tunneling process involves the creation of an electron-hole pair and is the cause of the label "inelastic".

The terms obtained from the substitution into Eqn. (2.30) which resemble the second term on the right in Eqn. (2.31), do not create such excitations and are accordingly labeled elastic tunneling events. This implies that the same electron tunnels through both junctions. In the elastic case, it becomes necessary to go beyond the zero dimensional approximation of the junctions and to consider electron scattering in the middle electrode. Such scattering can change the phase and energy of the electron, which makes the elastic tunneling rate very sensitive to the motion of the electron inside the middle electrode. The information about the electron motion inside the middle electrode is contained in the phase factors of the tunneling amplitudes, and it will be necessary to take them into consideration for elastic q-MQT current calculations.

Carrying out the transformations of Eqn. (2.30) into Eqn. (2.31) quantitatively, the q-MQT tunneling rate γ is obtained [Averin and Odintsov, 1989; Averin and Nazarov 1990],

$$\gamma^{\pm} = \gamma_{in}^{\pm} + \gamma_{el}^{\pm} \tag{2.32}$$

$$\gamma_{in}^{+} = \frac{\hbar G_1 G_2}{2\pi e^4} \int d\varepsilon_1 d\varepsilon_2 d\varepsilon_3 d\varepsilon_4 f(\varepsilon_1)[1-f(\varepsilon_2)]f(\varepsilon_3)[1-f(\varepsilon_4)] \times \left(\frac{1}{\varepsilon_2 - \varepsilon_1 + E_1} + \frac{1}{\varepsilon_4 - \varepsilon_3 + E_2} \right) \delta(eV + \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4) \tag{2.33}$$

where $f(\varepsilon)$ is the Fermi distribution function,[4] and

$$E_1 = \frac{e}{C_\Sigma}\left(\frac{e}{2} + Q_0 - VC_2\right) \qquad E_2 = \frac{e}{C_\Sigma}\left(\frac{e}{2} - Q_0 - VC_1\right) \tag{2.34}$$

For small voltages, $eV \ll E_i$, the inelastic contribution to the tunnel current can be solved explicitly for small ($k_BT \ll E_i$) but non-zero T [Averin and Odintsov, 1989],

$$I_{in} = e(\gamma_{in}^{+} + \gamma_{in}^{-}) = \frac{\hbar G_1 G_2}{12\pi e^3}\left(\frac{1}{E_1 + E_2}\right)^2 [(eV)^2 + (2\pi k_B T)^2]V \tag{2.35}$$

The elastic contribution to the q-MQT tunneling rate γ_{el}^{+} is,

$$\gamma_{el}^{+} = \frac{2\pi}{\hbar}\sum_{m,n,k,l} T_{km}^{(1)} T^{*(1)}_{k'm} T_{nl}^{(2)} T^{*(2)}_{nl'} f(\varepsilon_m)[1 - f(\varepsilon_n)]F(\varepsilon_l, \varepsilon_m, \varepsilon_n)F(\varepsilon_k, \varepsilon_m, \varepsilon_n)\delta(eV + \varepsilon_m - \varepsilon_n) \tag{2.36}$$

$$F(\varepsilon, \varepsilon_m, \varepsilon_n) = \frac{1 - f(\varepsilon)}{\varepsilon - \varepsilon_m + E_1} + \frac{f(\varepsilon)}{\varepsilon_n - \varepsilon + E_2} \tag{2.37}$$

The calculation for the elastic current $I_{el} = e(\gamma_{el}^{+} + \gamma_{el}^{-})$ is carried out over a coordinate representation which takes into account the phases of the

---

[4]The reverse tunneling rate γ_{in}^{-} is obtained from Eqn. (2.30) with the substitutions $V \rightarrow -V$, and $E_i \rightarrow E_i + eV$.

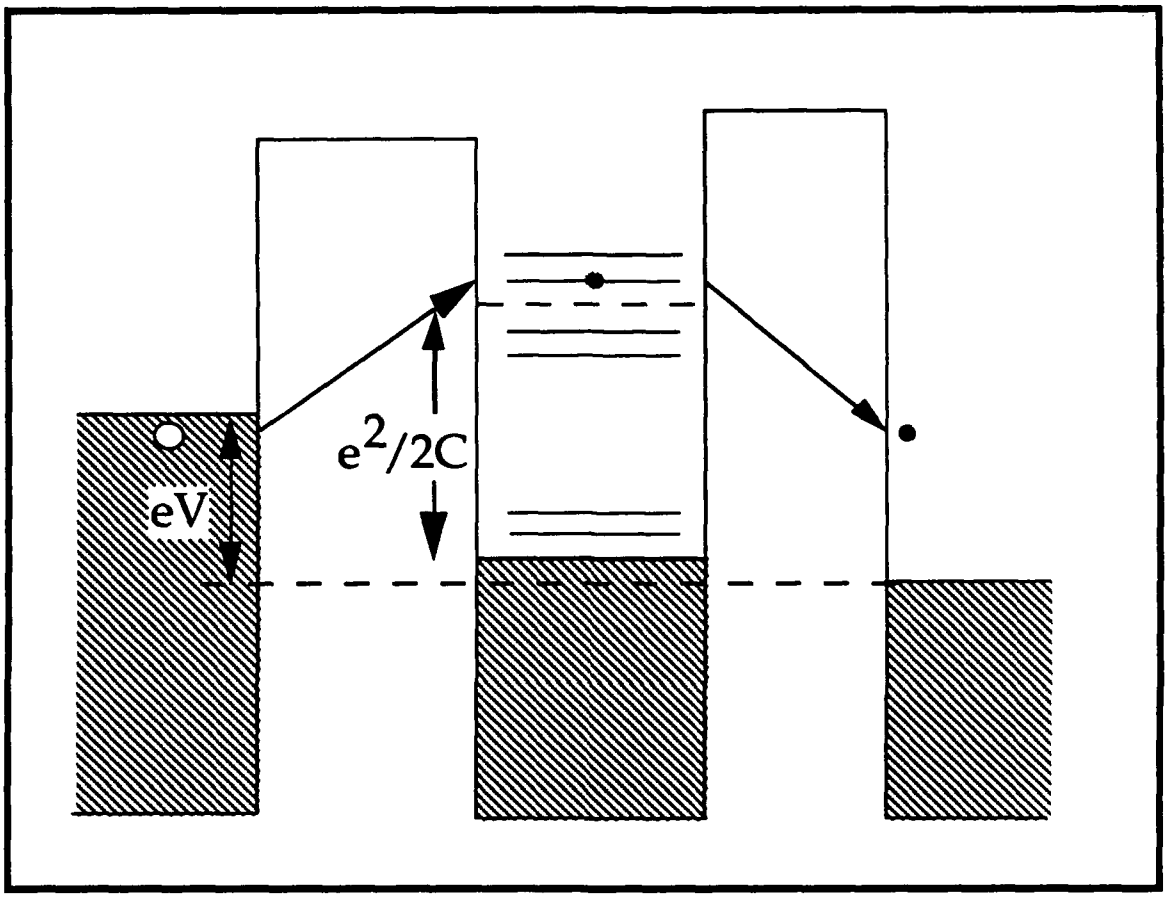


Figure 2.10, Elastic q-MQT process. An electron (solid circle) from electrode 1 will tunnel into the middle electrode, despite the increase in energy. It will then tunnel onto electrode 2. There is no electron-hole excitation left on the middle electrode. This elastic process is possible as long as the electron does not lose its phase information while on the middle electrode (i.e., no inelastic scattering).

tunneling amplitudes [Averin and Nazarov, 1990]* . In general, the calculated current, I_{el}, will depend on the geometry of the junctions.

It is, however, possible to obtain an equation independent of specific junction geometry, by considering the limit where the characteristic tunneling time, $\hbar/E_C$, is much larger than the classical time, L^2/D, of the diffusion through the middle electrode[5] (L is the characteristic mean dimension of the middle electrode, and D is the diffusion coefficient). This limit is where the the probability of finding an electron with a particular momentum is a constant. That is, the electron can sample all the Fermi surface with equal probability, as well as, all the particle volume with equal probability. In this limit, irrespective of the shape of the middle electrode, it can be shown that [Averin and Nazarov 1990],

$$I_{el} = \frac{\hbar G_1 G_2 \Delta}{4\pi e^2}\left(\frac{1}{E_1} + \frac{1}{E_2}\right)V \tag{2.38}$$

As shown in Eqn. (2.38) , the elastic tunneling rate is smaller by a factor of Δ/E_0 than the inelastic one, where Δ is the energy level spacing on the middle electrode and $E_0^{-1} = E_1^{-1} + E_2^{-1}$.

---

* Since it can be shown that I_{el} is linear in V, it is therefore sufficient to calculate $G_{el} = dI_{el}/dV$ at V=0.

[5]Despite the virtual character of the the excitation on the middle electrode, it can be described in terms of real electron motion inside the particle, as long as the classical time of motion within the particle is less than the virtual-electron-excitation time, $\hbar/E_C$.

The predictions of Eqns. (2.35) and (2.38) are the main results of this section. They predict a finite current within the Coulomb blockade due to higher-processes where an electron can virtually charge the middle electrode. The total q-MQT curre .t was calculated, and separated into elastic and inelastic components, both of which contribute to the current. The elastic component dominates only at very small voltages and temperatures, and with a very small middle electrode so that Δ is large.

CHAPTER THREE

EXPERIMENT

The instrument used to obtain our data is a low temperature scanning tunneling microscope This chapter is devoted to both the general features of STMs and the specific operation of our particular instrument. The first section explains the fundamental theory behind scanning tunneling microscopy. The next three sections respectively detail the mechanical, electrical and computer characteristics of the instrument built for this study. The final section describes overall instrument operation, and discusses sample preparation and measurement.

3-1 STM THEORY

The development of the STM by Binnig and Rohrer [1982] marked the beginning of a remarkable ability to image surfaces with atomic resolution. So powerful is this device that it has since led to manipulation of atoms, and investigation and creation of single electron sensitive circuits. The fundamental idea behind STMs is the ability to position and maintain a sharpened metal tip approximately ten angstroms from the sample surface. This is done by piezoelectric materials which expand or contract their crystal lattice when an electric field is applied. By maintaining a fixed distance between tip and sample while laterally moving the tip, a topographic image of the surface can be obtained. (This is a simplification; what is actually imaged is the three-dimensional variation of the charge density at the sample surface.)

The tip to sample distance is maintained by monitoring the vacuum tunneling of electrons across the gap. The vacuum tunneling phenomenon has been known since the 1920s. Three-dimensional vacuum tunneling has often been idealized by using a one-dimensional planar tunneling model, which is much simpler to calculate. From the solution of a one-dimensional barrier with plane waves [Simmons, 1963], the tunneling current can be written as

$$I = c/d^2\{(\bar{\Phi} - eV/2)\exp[-A(\bar{\Phi} - eV/2)^{1/2}d] - (\bar{\Phi} + eV/2)\exp[-A(\bar{\Phi} + eV/2)^{1/2}d]\} \quad (3.1)$$

where $A = 4\pi(2m)^{1/2}/h \approx 1.025(eV)^{-1/2}$ Å^{-1}. $\bar{\Phi}$ is the average of the barrier height between the two electrodes, V is the bias potential between the tip and the sample, d is the width of the vacuum gap, m is the mass of an electron, h is Plank's constant, and c is a constant. When tunneling between metals, if V is greater than approximately 5V, one enters the field emission regime; however Equation (3.1) is still valid, (field emission is not forbidden classically unlike tunneling). When the gap is small and the voltage is low, the current-gap distance relation can be simplified to

$$I \propto (V/d)\exp(-A\bar{\Phi}^{1/2}d) \quad (3.2)$$

Equation (3.2) indicates that a 1Å change in the gap distance produces roughly an order of magnitude change in the tunneling current, with $\bar{\Phi} \approx$ 4eV.

The exponential dependence of the tunneling current versus gap width has been measured by several investigators, [Young et al., 1971 and 1972]. Young et al., realized that surface topography could be mapped by using a

feedback system to maintain a constant tunneling current. This early device, called the topografiner, used x, y, and z piezoelectric drivers to scan the surface and maintain a constant tunneling current; however, it failed to give atomic resolution. This was most likely due to the large gap distance used, 100-1000Å, and insufficient vibration isolation. The scanning device later built by Binnig and Rohrer was maintained under ultra-high vacuum conditions, contained much greater vibration isolation, and used a tunneling gap of about 10Å. This resulted in atomic resolution of the samples.

The common general features of STMs are schematically shown in Figure 3.1. When a bias voltage is applied between the tip and the sample, electrons can tunnel across the vacuum gap. The feedback loop which controls the tunneling gap width, maintains a fixed tunneling current and hence a fixed tunneling distance. As the STM tip is scanned laterally across the surface by the x and y piezo drivers, the change in z piezo driver voltage, which is controlled by the feedback loop, reflects the changes in surface height. The plot of the z piezo voltage as a function of x and y piezo voltage yields atomic resolution contours of the surface.

3-1.1 STM Tips

It is well known among STM experimenters that atomic resolution is infrequent and difficult. This is due to the extraordinarily "sharp" tip required, whose formation is not completely reproducible and whose nature is not well understood. Even if the tip is carefully prepared by, for example, chemical etching or mechanical methods, atomic resolution is not guaranteed. The tip of an STM is crucial in the instrument's performance.

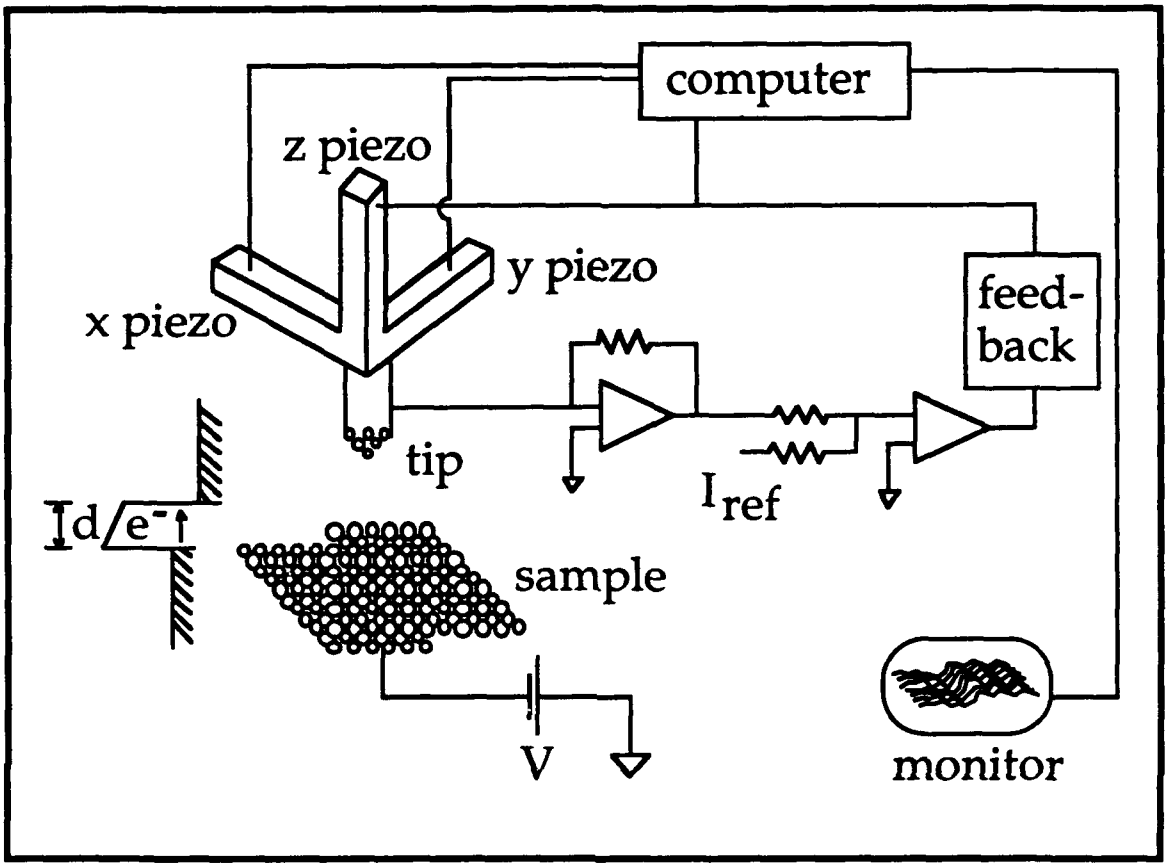


Figure 3.1, STM Schematic. This figure schematically represents the operation of an STM. The sample is maintained at bias voltage with respect to the tip. When the tip approaches to within a few angstroms of the sample, a current arises due to electrons tunneling across the vacuum gap. This tunnel current is measured and maintained at a set value by varying the tip to sample distance. Due to the exponential dependence of the tunnel current on the vacuum gap width, by maintaining a set current, the gap width is also precisely maintained. The tip is then scanned over the the sample by the x and y piezo, while the z piezo moves the tip so as to maintain the gap width. A plot of the z piezo voltage as a function of the x and y voltages yields a three-dimensional image of the sample.

During the early STM experiments, it was observed that the instrument resolution undergoes unexpected switching for no apparent reason. The resolution could change by as much as a factor of ten during a single scan. Experimenters often waited for hours or days for a "sharp tip" to occur and for "good images" to be taken. Spontaneous switching of instrument resolution was characteristic of STMs.

The resolution flipping indicated that subtle changes in the tip produced large differences in image resolution. This behavior is now interpreted in terms of localized states because the conditions for localized surface states to occur are equally subtle. Also, recent techniques for tip-sharpening can be understood as methods of forming localized surface states on the tip. It is strongly believed that atomic resolution is due to localized surface states or dangling bonds at the tip apex [Baratoff, 1984], as shown in Figure 3.2.

There are two basic methods for tip-sharpening. Tip-sharpening in this context refers to the procedures used during STM tunneling and imaging, to obtain atomic resolution. The actual physical tip is usually mechanically and possibly chemically sharpened before use in the STM. These preliminary procedures are necessary but are usually insufficient for immediately obtaining atomic resolution. The first method of tip-sharpening is applying a high-voltage bias. The second, is by picking up sample clusters. These techniques are described in detail below.

As reported by Wintterlin et al. [1989] "the most significant single parameter for the observed corrugation apparently related to the specific procedure of tip preparation" was to raise the bias voltage at the

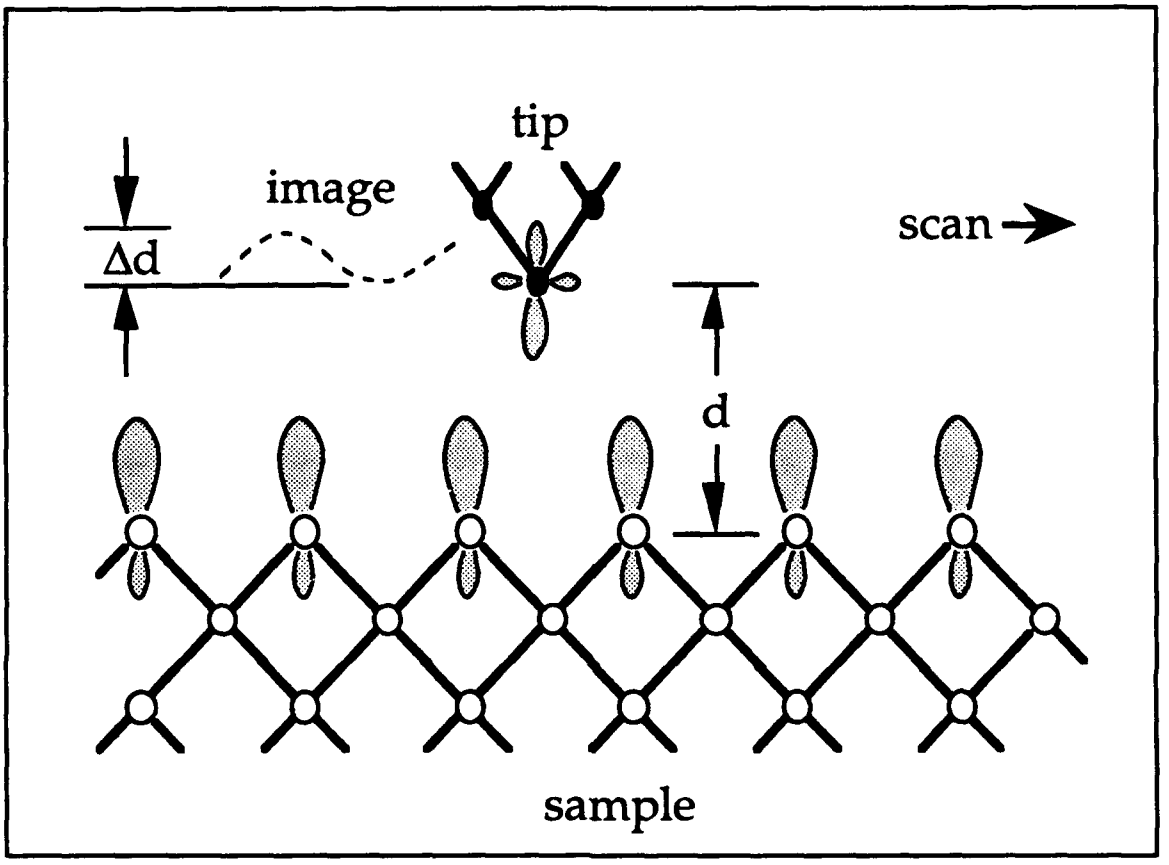


Figure 3.2, Imaging. This microscopic view of STM imaging mechanism shows a tip with a dangling bond state scanning over a sample surface. The overlap of the dangling states from the tip and the atoms on the surface create an increased tunneling conductance and allow for atomic resolution.

sample to -7.5 volts and to maintain this for a few seconds. The feedback circuit responded to the voltage and withdrew the tip approximately 30Å. Afterwards, atomic resolution was achieved. This only worked with a positive bias at the tip. The explanation of this process is shown in Figure 3.3.

When a positive bias is applied on the tip, the positive tungsten ions feel a force aligned with the electric field vector. For the atoms in the bulk, this force is completely offset by the force of the "electron wind" which points in the opposite direction. But at the surface the situation is different. The electron density is less at the surface and thus the electron wind is reduced. The tungsten atoms now exhibit a net positive charge and due to the polarity of the electric field are attracted towards the apex of the tip. This results in a tungsten cluster forming at the tip, usually with one atom at its apex.

Using first principle methods, Ohnishi et al. [1989] have calculated the electronic states of tungsten clusters. They have shown that at the apex atom of a tungsten cluster, there is a localized metallic dangling bond state. By taking this cluster as a tip, they also show that the tunneling conductance is predominantly generated by this dangling state protruding from the apex atom.

Another method for tip-sharpening was reported by Demuth et al. [1988], and is shown in Figure 3.4. Upon gently colliding the tip with the silicon surface, they observed that the tip picked up a cluster of silicon atoms. This was evidenced by a hole on the silicon surface. After the collision, atomic resolution was achieved. Demuth proposed that the sp^3 dangling bond from the apex silicon atom created the 'atomically sharp' tip. Other experimenters have used similar methods where tip material was deposited

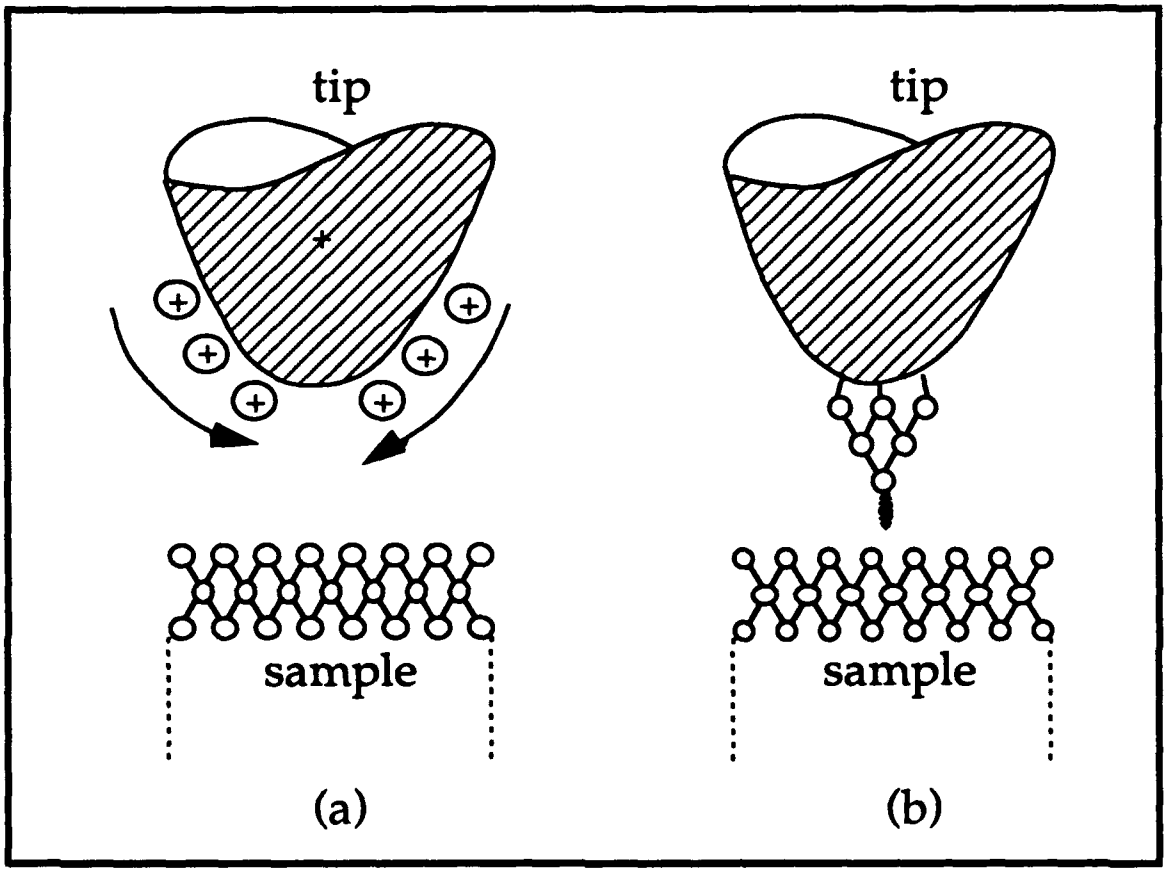


Figure 3.3, Tip sharpening by high voltage. (a) by applying a high positive voltage to the tungsten tip, some tungsten atoms are caused to move to the end of the tip. (b) a tungsten cluster forms with a dangling bond at the apex atom.

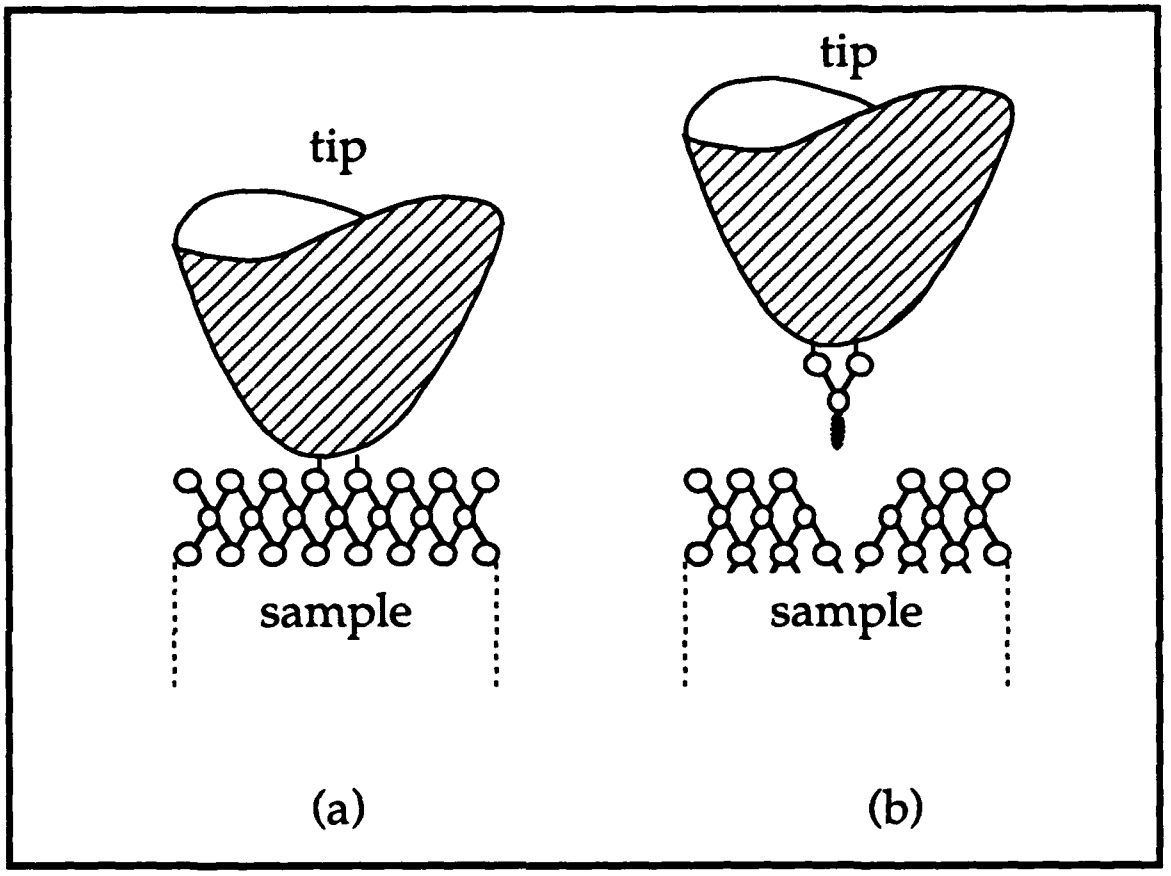


Figure 3.4, Tip sharpening by contact. (a) lightly colliding the tip with the surface. (b) a group of the surface atoms is picked up by the tip. The cluster formed at the tip has a dangling bond.

on the sample and created a "sharp" tip. For both cases, numerous first principle calculations have shown the existence of a dangling bond.

3-2 INSTRUMENTATION.

The STM used in our experiments has several features which made for straightforward construction and operation. First, its small size allowed unsophisticated and inexpensive vibration isolation methods to be extremely effective. Also, it allowed for quick and simple refrigeration methods down to 4.2K. Second, the relatively elementary design of the electronics and computer system gave stable and reliable service. The next three sections explain in more detail the mechanical, electrical and computer characteristics of this instrument.

3-2.1 Mechanical

The particular STM used in our experiments is shown in Figure 3.5. Though it appears large, the actual nucleus of the STM mechanical section is only 2 inches long, less than 1 inch in diameter and weighs about 10 ounces. This central section contains the tip, including the piezo controller, and the sample. The rest of the gear essentially serves to allow the STM to operate in vacuum and to be lowered into liquid helium. These structures are only complicated by the fact that they must be vacuum tight, and carry high voltage electrical connections and movable mechanical connections to the central region.

A schematic of the central region is show in Figure 3.6. It is made of stainless steel. Stainless steel is chosen because of its relatively low thermal

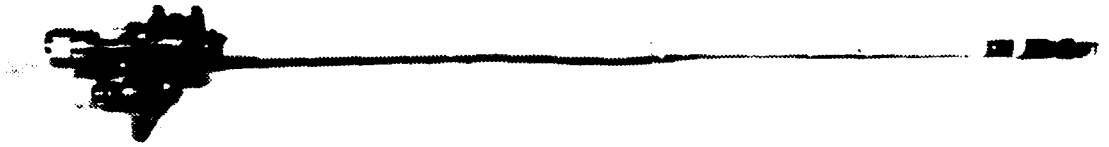


Figure 3.5a Photograph of Tinkham group STM. This photograph show the actual instrument used in our study. The entire aparatus is about 48" long in order to imerse the tunneling section into a liquid helium dewar. The tunneling section is shown in the following picture.

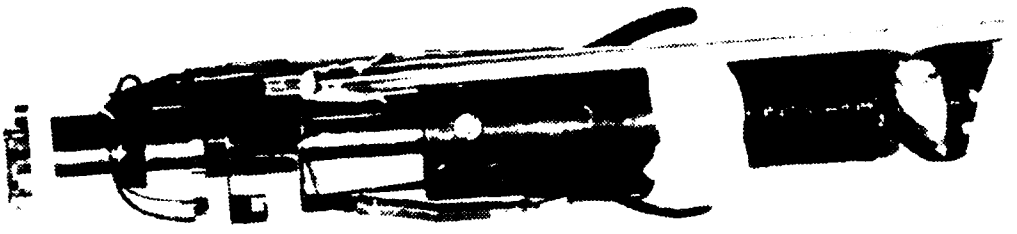


Figure 3.5b Photograph of STM center section. Above is shown the tunneling section of the instrument. The dimensions of the unit can be obtained from Fig. 3.6 where its compact size is evident. Photographs of the accompanying electronics and computer hardware can be found in the STM Operation Guide.

coefficient of expansion as well as its low thermal conductivity. It also has the added benefits of being a comparatively clean and inexpensive metal. The central region is divided into two parts, the tip holder, and the sample holder. They are held together by four stainless steel screws.

The tip holder is a simple piece of stainless steel designed to hold the piezo and tip in relative immobility with respect to the sample. The fine angstrom movement of the tip is done by use of a cylindrical piezo tube which has the tip attached at one end with a piece of macor and a small stainless tube. Macor is a machinable ceramic and is also a superior insulator. The piezo is nickel plated on the inside and the outside. The outside plating is cut into quarters, as shown in Figures 3.7 (a) and (b). A voltage applied, to the inner plating, relative to the set of four outer plates, causes the piezo to expand in the longitudinal, z, direction. Applying voltage across opposite plated quarters on the outside of the cylinder, causes the piezo to bend in the x or y direction. This is how the tip is moved. Typical deflection characteristics of the piezo at room temperature, are 50Å per volt in the x and y direction, and 20Å per volt in the z direction. These values are reduced by about a factor of 5 at 4.2K. The piezo cylinder is attached to the tip at one end and to the stainless steel tip holder at the other end, Figure 3.7 (c). Both attachments are made by macor spacers. Electrical connections are made to the piezo through small gold pins mounted in the macor spacer. Mating gold pins are mounted in another piece of macor which is on the end of the wires leading out of the rig. To lower noise and inductance, the leads are twisted pairs.

The sample holder is similarly a piece of stainless steel designed to rigidly support the sample stage. The stage can be flexed several microns in

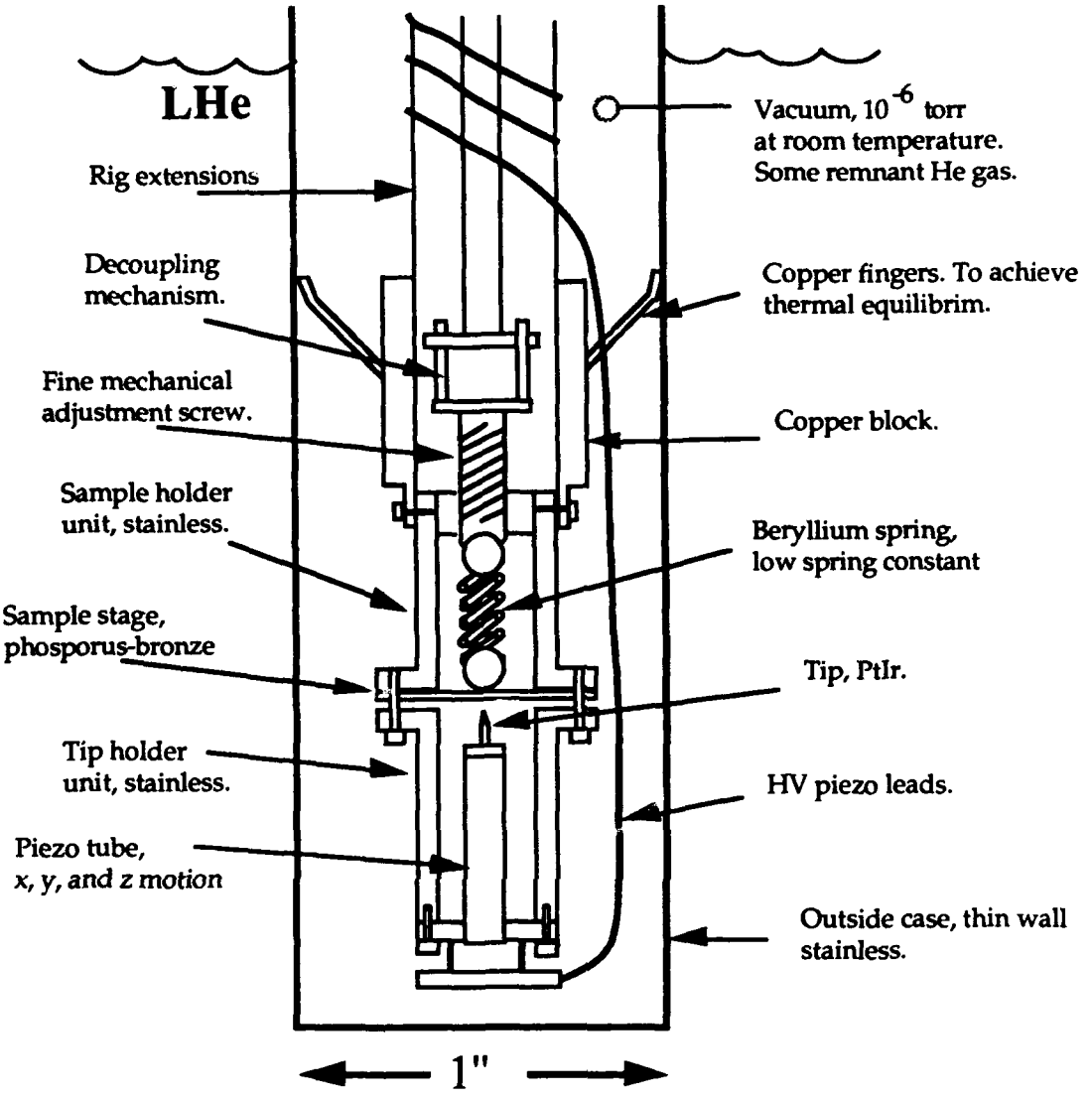


Figure 3.6, Schematic of Tinkham group STM. The diagram shows all the main components of the STM used in our experiments. The 1" mark gives the scale for all component sizes. The STM section shown above is connected at the end of a 45" stainless steel tube, which allows us to submerge the entire apparatus in LHe. The fine mechanical screw is turned from above through a vacuum fitting. This compresses the spring and flexes the sample stage. The movement of the sample stage is several microns per revolution of the screw and thus allows us to bring the sample within tunneling distance from the tip. Once there is tunneling, the z movement of the piezo maintains the tunneling distance.

the z direction by compressing the spring held in the sample holder. This is the fine mechanical adjustment of the system and allows the tip and sample to come within tunneling distance. The coarse mechanical adjustment of the system is accomplished when the sample is mounted by bringing the tip and sample within several microns of each other as observed under an optical microscope. The coarse and fine approach of the tip will be discussed later in the sample preparation sections.

3-2.2 Electronics

The electronics of the STM are the heart of the instrument. They are basically responsible for measuring the tunnel current, outputting the proper voltage to the z piezo, so as to maintain the vacuum tunnelling distance, and communicating with the computer. The electronics are independently capable of operating the instrument without a computer. That is, the electronics can maintain a set tunneling current, or equivalently a fixed distance between the tip and sample, as well as move the tip in the x-y plane. A computer is only needed for coordinating and storing the information of z(x,y), to create an image, and to obtain conductance data, I(V). The electronics are primarily analog.

The most important and complex part of the electronics is the *feedback* board, it outputs a voltage to the z piezo so as to maintain a fixed tunneling current. The x and y *controller* boards, which allow the computer to control the x-y motion of the tip with minimal CPU involvement, are next in complexity. And lastly, there are several *amplifier* and *buffer* boards as well as a *precision voltage reference* board which are relatively simple circuits. It

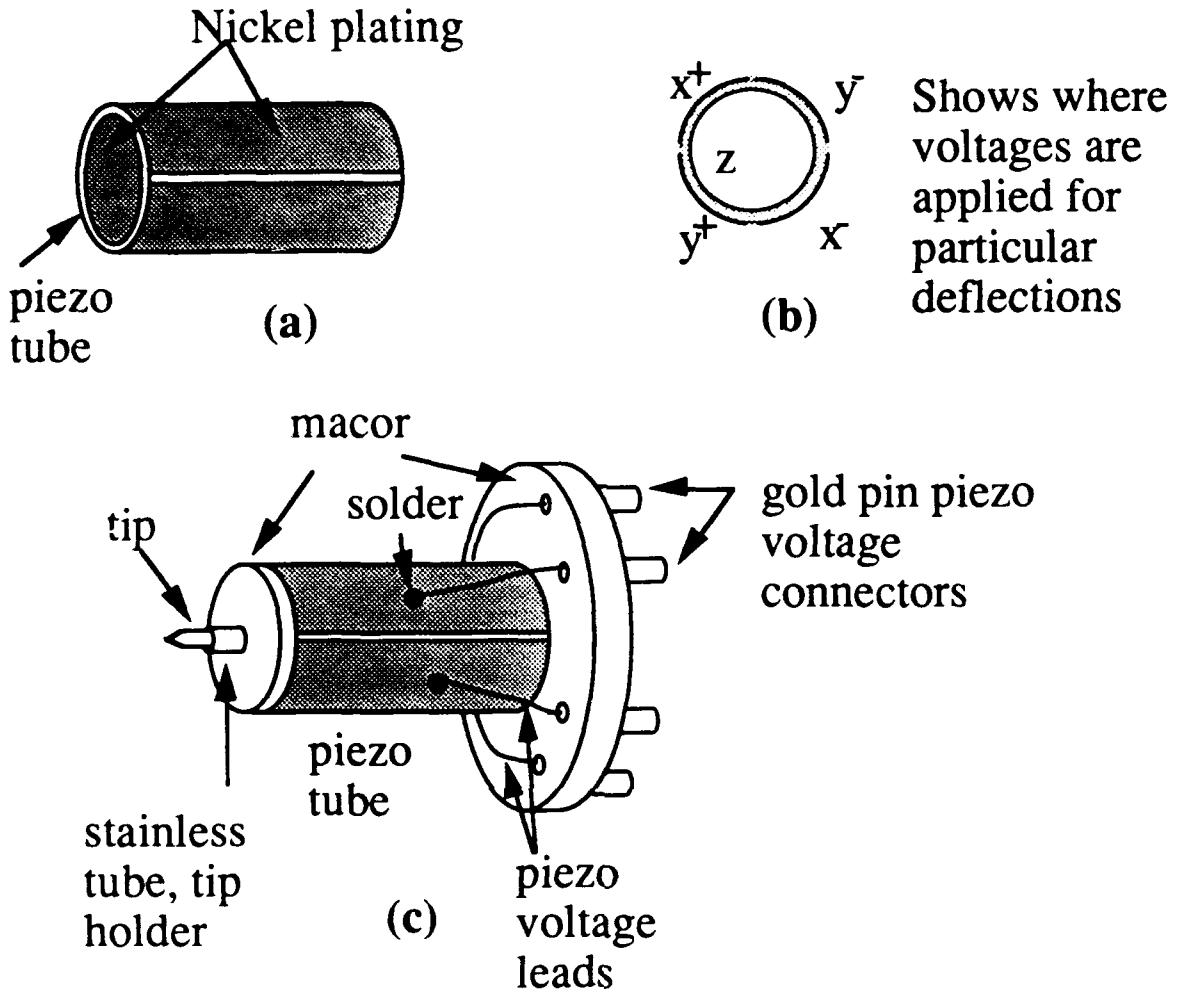


Figure 3.7, Piezo tube scanner. (a) shows the piezo tube as received from the manufacturer. It is one inch long and 1/8" outer diameter, (drawing is not to scale). It is completely coated on the inside with nickle. The outside is also coated; however it is separated into four symmetric quadrants as shown in (b). (b) also shows where the voltages are applied so as to move the piezo in the desired direction. Applying voltages across the opposite plated quarters causes the piezo to bend in the x and y directions. Applying a voltage, on the inner coating, relative to the outer coatings, causes the piezo tube to expand or contract in the z (axial) direction. (c) shows how the piezo tube is connected to the tip, the voltage leads and the rest of the apparatus.

should be noted that the above characterization is a relative comparison; all the circuits are quite standard, being at the level of an introductory undergraduate electronics course. All the circuits were made by etching copper on a fiberglass board which was then tinned. The printed circuit boards were one sided and made in house at Harvard. The layout was done on a Macintosh, using Quick Circuit, and the negatives were processed at Charrette. Once the layout was done, the whole process of creating a negative and then making the board took only about ten days and cost around $25. The benefits of a printed circuit board, as compared with a wired board, are usually higher signal-to-noise ratio and a much more reliable performance. The replacement and diagnosis of damaged chips is also much easier on a printed circuit board.

Feedback Board

The schematic diagram of the feedback circuit is shown in Figure 3.8. The input signal is first rectified so that the direction of the tunneling current will not matter. The AD630 is a balanced modulator/demodulator. In our setup the input signal is used as its own reference in the balanced modulator topology; the AD630 thus acts as a precision rectifier. The high frequency performance is far superior to that which can be achieved with diode feedback and op amps (operational amplifiers). There are no diode drops which the op amp must jump over with the commutating amplifier. The output of the AD630, which is the absolute value of the measured tunneling current, is then subtracted from a reference signal, which is the desired tunneling current (The desired tunnel current is set by a potentiometer on the front panel). The difference is amplified through a variable-gain low-noise

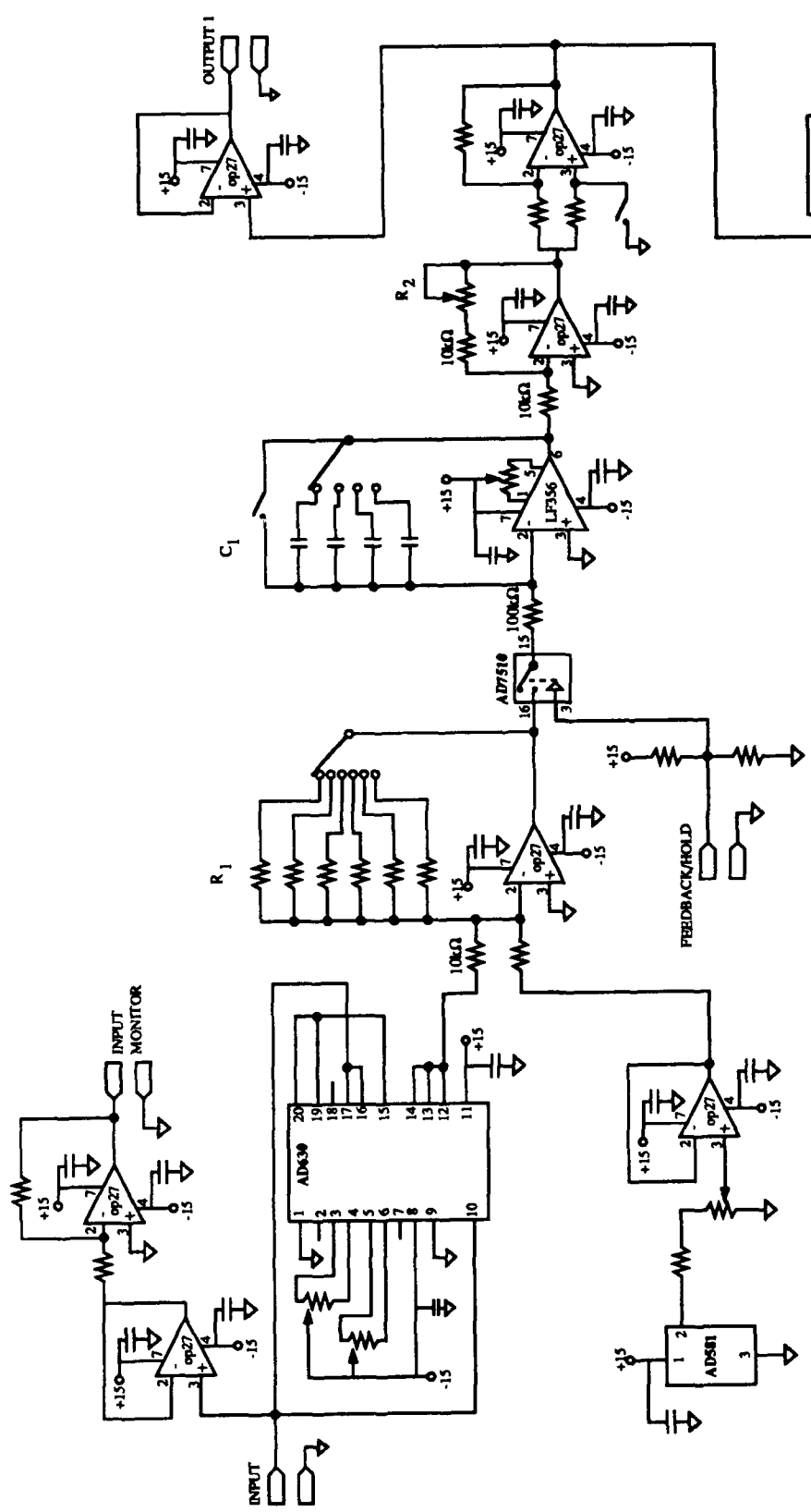


Figure 3.8 Feedback board. The above schematic diagram details the design of the feedback board, which maintains a constant tunneling current between the STM tip and the sample. The input comes from a current amplifier and is proportional to the tunnel current. The INPUT MONITOR is the INPUT x 10. The FEEDBACK/HOLD is a switch which either has the feedback operating or holds the final value sent to the z piezo. This is necessary for taking I(V) data. The OUTPUT is given by Equation (3.3). It is amplified and then sent to the z piezo.

amplifier and continuously integrated. This integrated difference is amplified once more and sent to a ±1 gain switchable amplifier. This allows us to change the polarity of the output voltage to suit our requirements. The signal is buffered one more time before proceeding to a BNC output. Equation (3.3) describes the explicit function performed by the feedback board.

$$V_{out} = \frac{\pm 1}{(100k\Omega)C_1} \left\{ \int (V_{ref} - |V_{signal}|) \frac{R_1}{10k\Omega} dt \right\} \frac{(10k\Omega + R_2)}{10k\Omega} \qquad (3.3)$$

R_1, R_2, and C_1 are variable in our system. To optimize the output voltage we try to achieve the greatest gain and shortest possible RC time without having the circuit become unstable and oscillate.

The feedback board integrates the input in order to obtain stable performance. For example, consider the most simple choice for a feedback circuit; one which outputs a voltage directly proportional to the the difference between V_{ref} (a constant) and V_{signal}, as shown below.

$$V_{out} = g(V_{ref} - V_{signal})$$

The proportionality constant g is the gain of the circuit. By choosing the appropriate sign of g, this circuit will provide negative feedback and try to maintain $V_{signal} = V_{ref}$. However it will never be able maintain this condition because $V_{signal} = V_{ref}$ is not an equilibrium point with respect to time, i.e.,

$$\frac{dV_{out}}{dt} = -g \frac{dV_{signal}}{dt} \neq 0 \quad \text{at } V_{signal} = V_{ref}$$

The last statement is true except in the special case with the initial conditions of $V_{signal} = V_{ref}$ and $(dV_{signal}/dt) = 0$.

In comparison, consider the circuit with the following output voltage,

$$V_{out} = g\int (V_{ref} - V_{signal})\, dt$$

In this case, we obtain,

$$\frac{dV_{out}}{dt} = g(V_{ref} - V_{signal}) = 0 \qquad \text{at } V_{signal} = V_{ref}$$

which means, $V_{signal} = V_{ref}$ is an equilibrium point with respect to time. It is also seen that this point is stable under deviations of V_{signal} from V_{ref}. That is, the sign of (dV_{out}/dt) is different depending on whether V_{signal} is greater than or less than V_{ref}. Thus, this circuit is able to obtain and continuously maintain the stable point $V_{signal} = V_{ref}$.

X-Y Controller Boards

The x-y controller boards are two identical boards, one for x and one for y, which allow the computer to control the output signal for x and y piezo movement with minimal CPU involvement. The schematic for the controller boards is shown in Figure 3.9. The input of the controller boards is a square wave clock signal sent out from the computer's clock board. This signal enters a series of four 4-bit counters wired in series, which essentially performs as a 16-bit counter. The output of the four counters, a 16-bit digital number, is sent to a bipolar 16 bit digital-to-analog converter. This output voltage is then added to a variable offset, buffered, and finally sent out to a

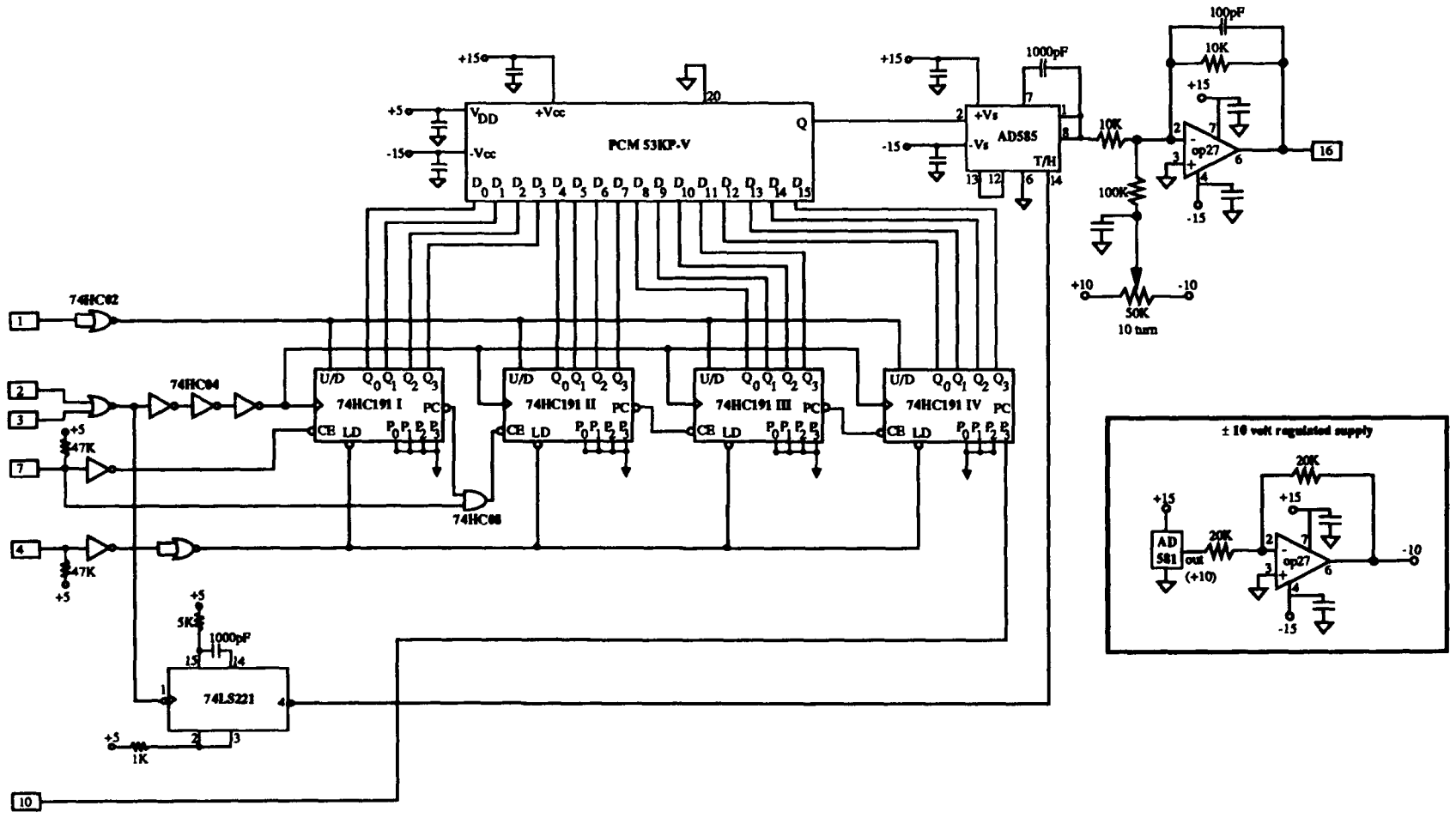


Fig 3.9, X-Y Controller board. The above schematic details the electronics used in controlling the x and y movement of the piezo. One board is needed for each direction. The function of the board is to take a clock signal and convert it into a voltage ramp for the piezos. Thus the computer only has to set the duration and speed of the clock signal for each direction (four bytes) to move the tip from one point to the next.

BNC connector. This will eventually be amplified once more and sent to the x and y piezos.

The usefulness of this architecture is that the square wave signal coming from the computer clock board only requires that its frequency and duration be set. For example, the CPU will tell the clock board to output 2^7 cycles at 100 Hz, and the clock board will perform the operation without further input from the CPU. This would result in the controller board ramping from zero to +5V ($+10v \cdot 2^7 / 2^8$) in a time period of 1.28 sec (2^7 / 100 Hz). By setting the frequency and duration bits for the x-controller and the corresponding bits for the y-controller, one can move the tip to any position on the $2^{16} \times 2^{16}$ grid with only four bytes and a comparable amount of CPU operations. This *vectoring* method of moving the tip is much less CPU intensive than the standard methods, in which when moving from (x_0, y_0) to (x_1,y_1), the CPU calculates and outputs the coordinates of all the intermediate positions on the $2^{16} \times 2^{16}$ grid. Figure 3.10 shows a graphic comparison of the two methods.

This vectoring method of moving the tip is particularly useful when taking an image. This is because the tip will raster across the sample with minimal CPU input allowing the CPU to read the z piezo voltage, or equivalently the height of the sample surface. Having the primary responsibility of the CPU be only to read and store height information allows images to be taken much faster than is usually possible with conventional tip control electronics. For our system an image of a 1000Å x 1000Å area with a 160 pixel x 160 pixel resolution takes about 30 seconds.

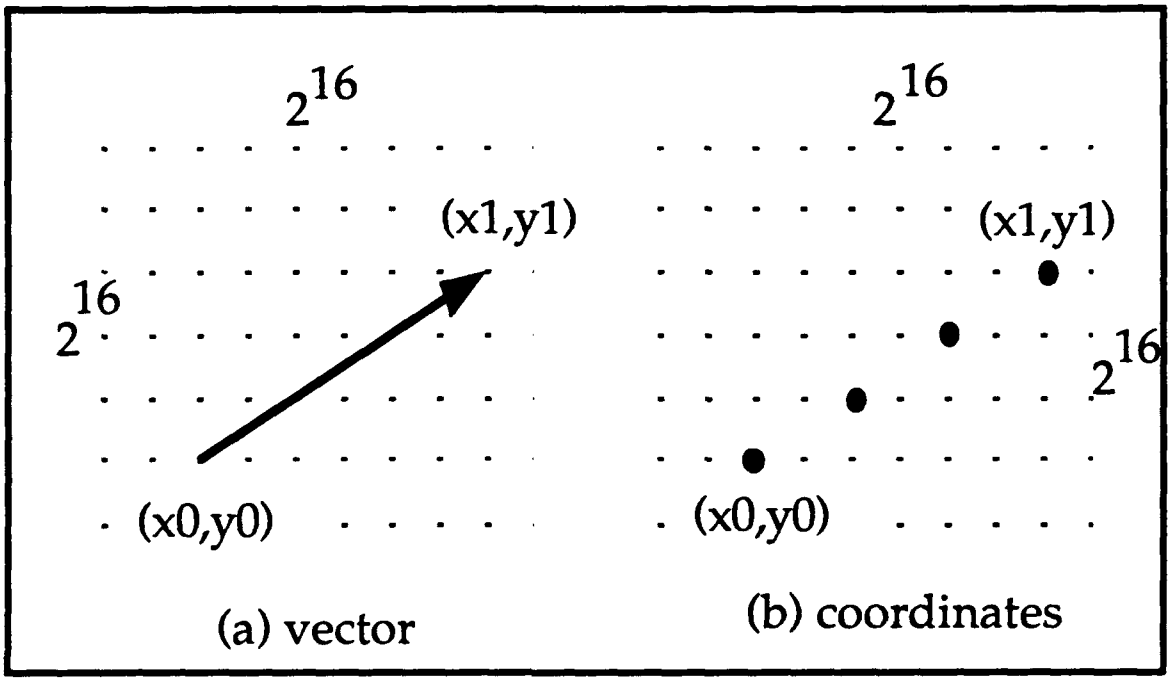


Figure 3.10, X-Y motion. This diagram illustrates the significant saving of CPU (computer) involvement by using the "vectoring" method instead of sending out the actual position map. It compares the two methods in moving from (x0, y0) to (x1,y1). In (a) the computer only has to calculate the distance to move in the x and y directions and how fast to move in the x and y directions. The ratio of the y to x velocity will determine the slope of the movement. Thus the computer only had to make four calculations. In (b), the standard method of x-y STM control, the computer calculates the coordinates of all points between the endpoints and outputs them to a D/A converter. The number of calculations in this method is on average about 256. The importance of saving computer computations is that it allows much faster tip movement, and hence much faster imaging as well as the ability to run the STM with a relatively slow and inexpensive computer.

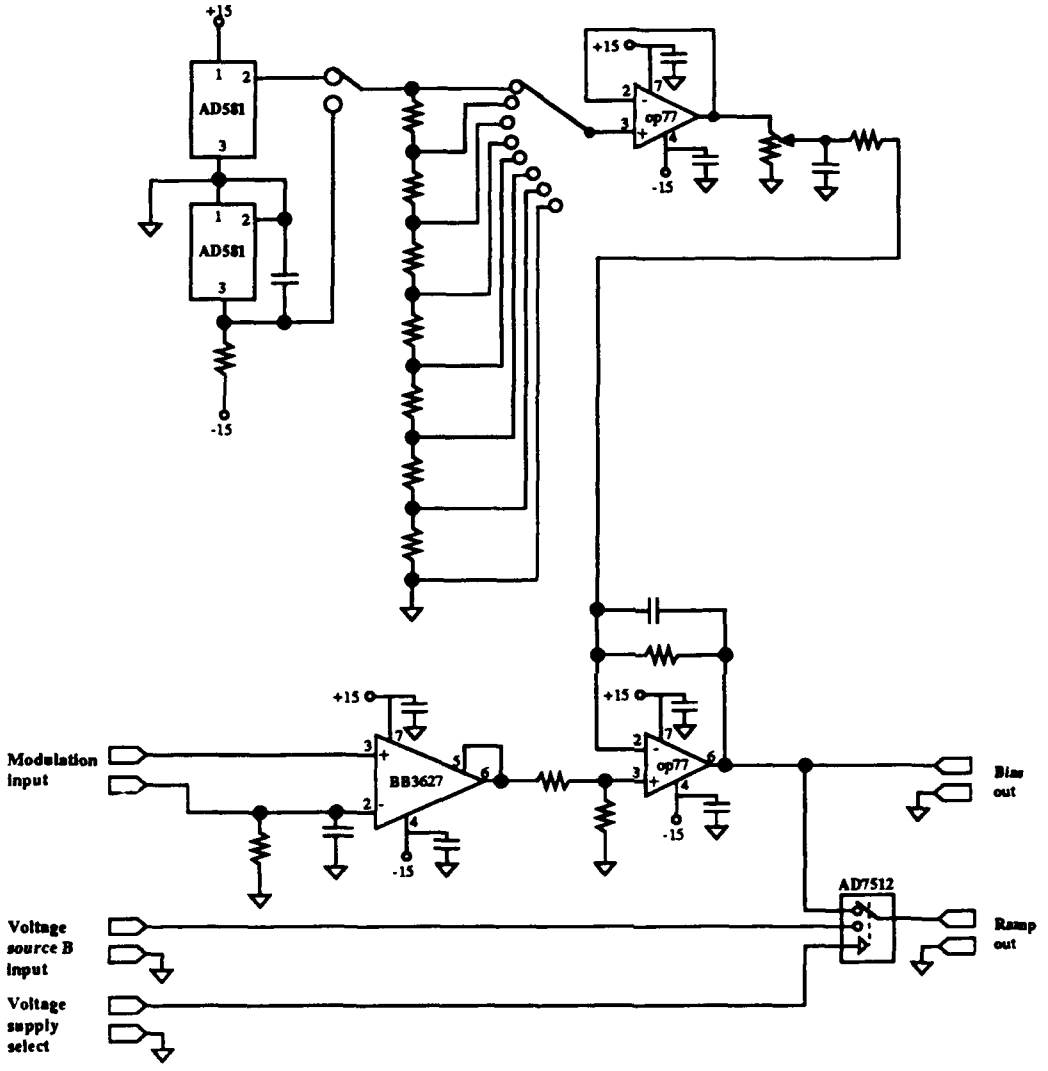


Figure 3.11, Bias board. This board outputs a regulated voltage which is used to provide bias across the tip to sample gap.

Precision Voltage Board

A very stable and precise voltage output board was designed for the STM. This board maintains the bias voltage across the vacuum gap between the tip and sample. The heart of this board is two AD581 High Precision 10V Reference ICs. This board, Figure 3.11, is quite elementary, using voltage dividers and op amps to give a variable output voltage. The board also contains an electronic switch to allow the output to be disconnected from the constant voltage source and connected to another input. (This other input is usually sent a voltage ramp signal, which is used to take I(V) data, as will be explained below).

Amplifiers

There are several amplifier boards in our circuitry. The underlying requirement in all of them is 'low noise.' The three largest amplifier boards are for the x, y and z piezo signals. These high voltage boards amplify the signal from the x-y controller board and the feedback board before it is sent to the piezo. The three boards are identical except that the z amplifier board has gains from 10 to 100, while the x and y signal amplifiers have gains from 1 to 10. The maximum output voltage for all these boards is ±120V. The main component in these boards is a PAR4A High Voltage Amplifier. These high voltage amplifier chips allow the board to be simple but they should be heat sunk to control heating.

Other amplifier circuits include a D/A interface board which reads the output voltage of the feedback board, amplifies it if necessary, and sends it to the computer. This signal is what is utilized to make a topographic image.

Unity gain buffer circuits have been built for x and y signals but are not currently used.

Power Supply

To operate all the above electronics, a power supply box has been built. The box outputs voltage at +5, ±15, and ±120 volts. Current outputs are 1A for ±15V and 0.5A for +5V and ±120V. The voltage supplies are regulated and air cooled.

3-2.3 Computer

The STM is controlled by a DELL 310, which is an IBM compatible PC based on the Intel 386 processor (CPU), running at 20MHz. Our unit also contains an Intel 387 math coprocessor, 1 Mb of standard RAM, and 1 Mb of extended RAM. A Metrabyte CTM-05 clock board and a Data Translation 2823 Analog and Digital I/O system are also in the computer. The CTM board has a variable-frequency square wave output, which is sent to the controller boards for x-y motion. The 2823 is used to read the topographic information from the feedback board as well as to output a voltage ramp and read the tunneling current when taking I(V)s.

Imaging software used by the STM was primarily created by Jeremy Avigad and Eric Ganz in Professor Golovchenko's group at Harvard University. Certain sections have been greatly modified to operate with the different hardware which we use, and we have also added the ability to take I(V) traces; however the majority of the software is the same. The programs are written in C with the main program, "SCAN", occupying about 300Kb after compilation.

Function	Program
Image surface	xyscan.c ; xyscan2.c ; xyfile.c ; xybox.c
Take I(V)	didv.c ; didv2.c ; dt.c
Take line scan	linescan.c ; line2.c
Process images	process.c
Test electronics	tester.c
General	counter.c; control.c; intrpts.c; vga.c; tip.c; loadtmp.c; misc.c

Table 3.1, STM operation software. The above table shows which C programs specifically run the desired function. They are all linked under the program "scan.c".

The program performs five distinct functions: (1) image surface, (2) take I(V), (3) take line scan, (4) process images, and (5) test electronics. The parent program is labeled "scan.c"; it provides the main menu and links all the different programs required to perform the five functions. The source code is written in Microsoft C. Table (3.1) shows the programs required for each function. The last row shows C programs which are used by several of the functional programs. These are more general programs which, for example, communicate with the VGA graphics, and other such necessities of computer operation.

The programs are stored on hard drive C: on the Dell. At the writing of this thesis, Microsoft windows is installed on the Dell and all necessary programs can be found in the folder labeled "Applications". The "SCAN" program can be run by starting the STM program from windows. Also, the entire hard disk C: is periodically backed up onto floppy disks which can be found in the computer desk. The information on them has been compressed, and can not be read directly. The "Fastback" program must be used to read these disks.

3-3 OPERATION

Figure 3.12 shows an overall wiring diagram of the STM. The two basic operating features of our instrument are (i) the ability to image samples on a few angstroms scale, and (ii) the ability to make localized measurements of tunneling current as a function of bias voltage, I(V). These two features are discussed in detail below.

3-3.1 Imaging

As was explained above, images are obtained by having the feedback loop maintain a constant tunneling current, hence a constant distance, between the tip and the sample. Then as the tip is moved laterally in the x-y plane, the voltage sent to the z piezo to maintain a constant distance is recorded. A map of this voltage versus the x-y position of the tip produces the topographic image.

Figure 3.13 is a picture of an evaporated gold film, taken at 4.2K. The sequence of photos shows the angstrom control we have in investigating different areas of the sample. This is further simplified by the addition of mouse control, which allows one to just point at a section of the sample on the screen and have the tip move there. These capabilities permitted easily searching for and probing the specific features we were interested in on each sample. Figure 3.14 shows images of graphite planes obtained at room temperature. This is a raw image of what is believed to be carbon atoms. The small size of the STM allows for simple yet effective vibration isolation, which allows such images to be obtained.

3-3.2 Conductance Measurement

The second feature of the STM is its ability to make very localized (~ 10Å) measurements of tunneling current as a function of bias voltage, I(V). Our instrument is designed to do this in two ways: i) static measurements, and ii) dynamic measurements. The basic difference is that in static measurements of I(V), the feedback loop is turned off, while in dynamic measurements the feedback loop is kept on.

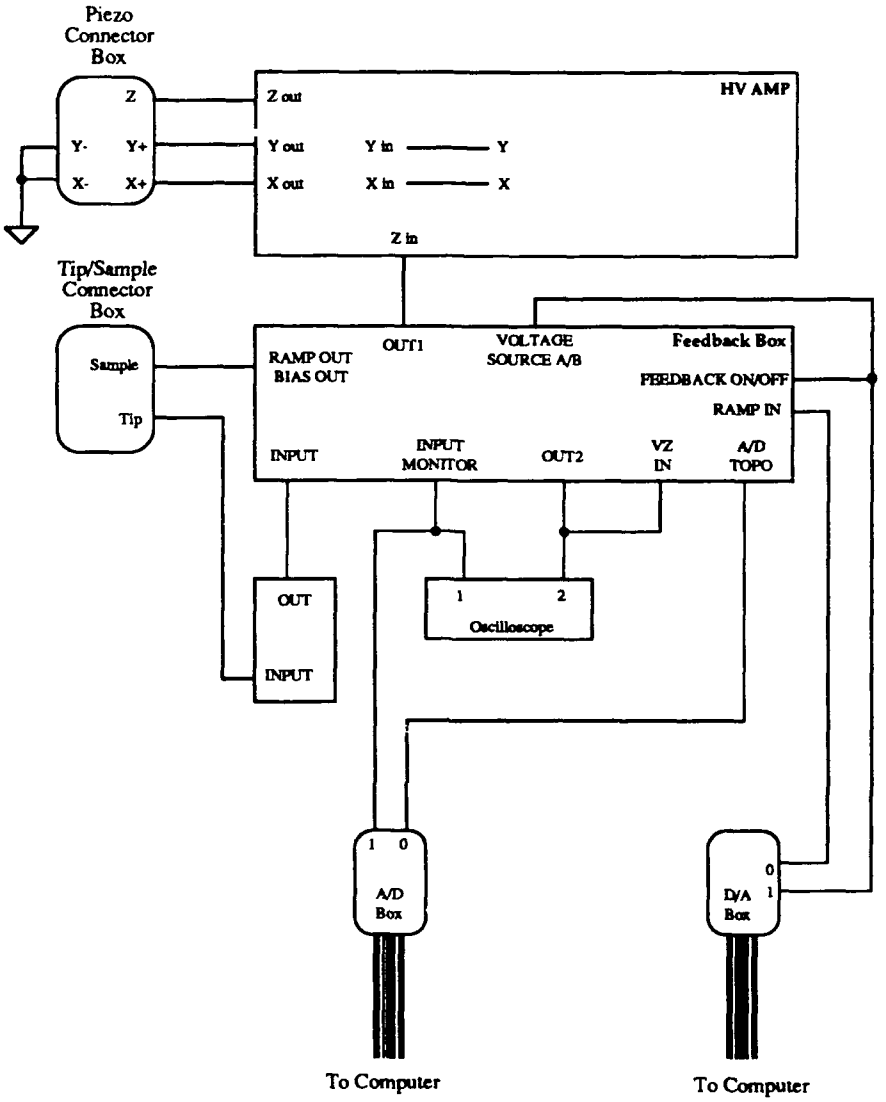


Figure 3.12, STM Wiring. The above diagram shows the overall wiring of the STM. The cables are coax, except for the computer connections which are ribbon cable.

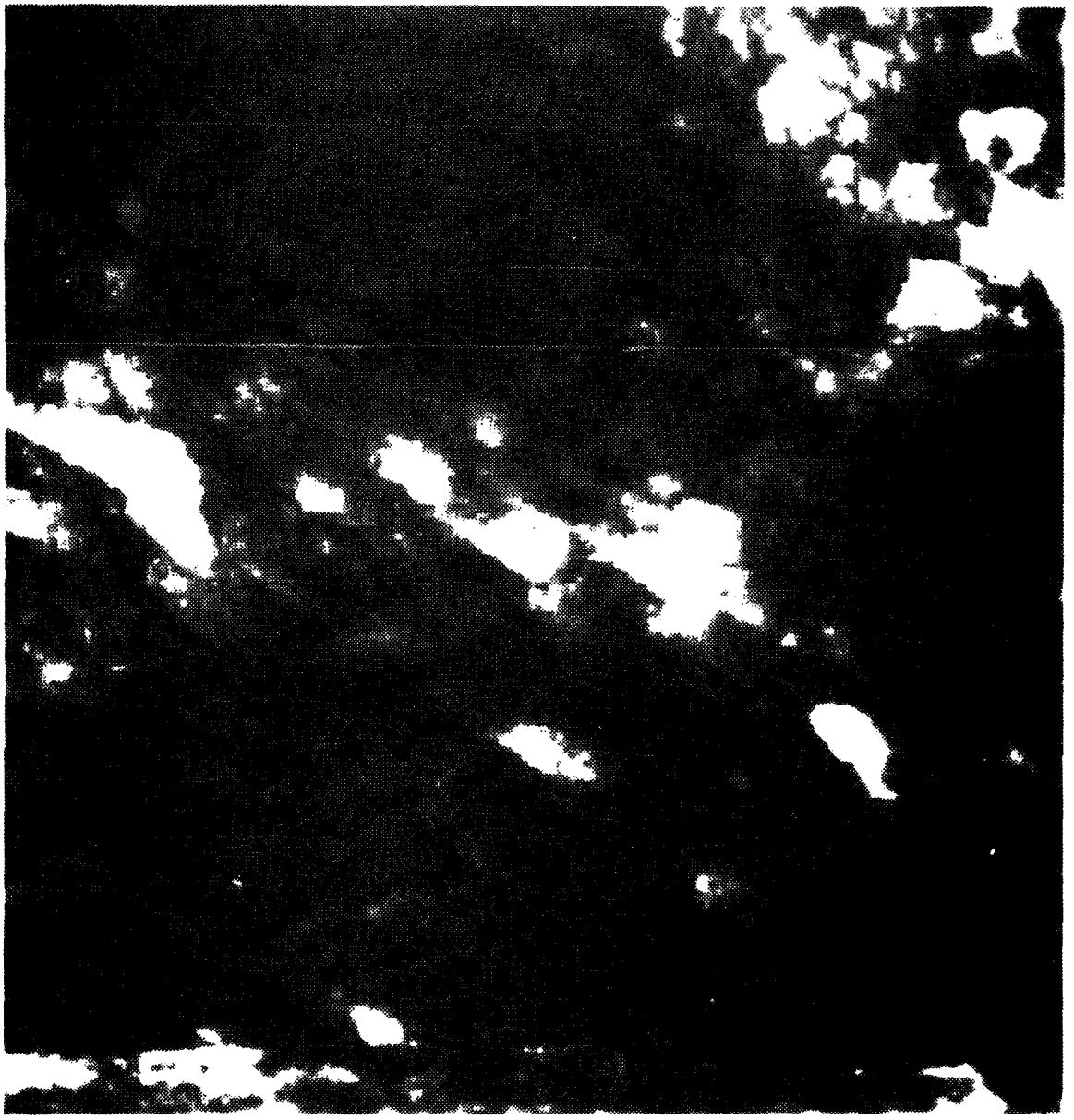


Figure 3.13a, Gold film image sequence. The following sequence of photographs illustrates the flexibility with which the STM scans the surface. A particular feature can be sought out and slowly approached by using only a computer mouse. The surface image above is 10,000 Å x 10,000 Å. The three following pages show progressivly closer ups images of the feature at the center of this image.

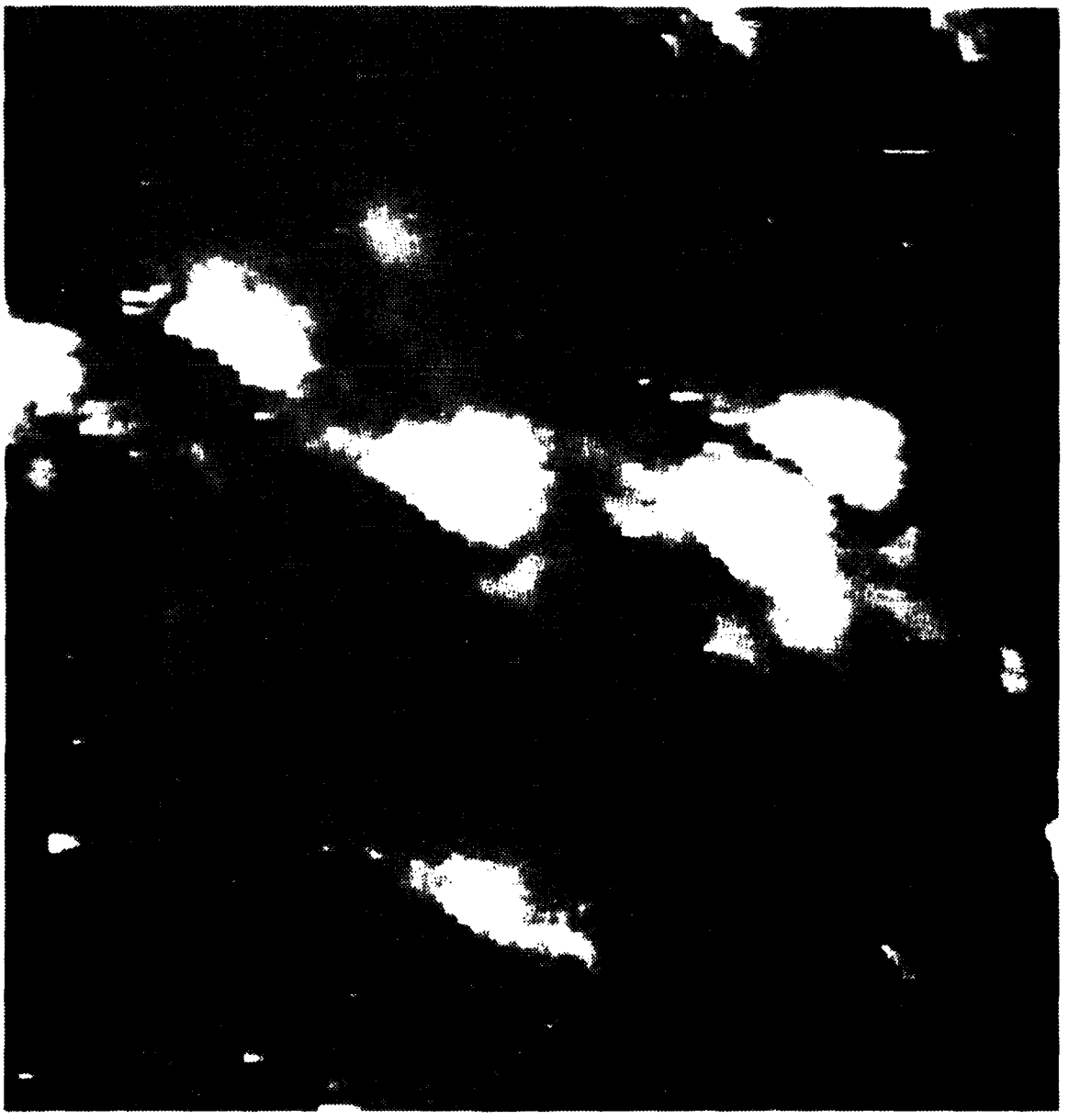


Figure 3.13b, Gold film image sequence. This STM surface image is a close up of the image shown in Fig. 3.13a. This image is 5,000Å x 5,000Å. Notice the enlargement of the same features in the center of the previous figure.

Figure 3.13c, Gold film image sequence. This STM surface image is a close up of the image shown in Fig. 3.13b. This image is 2,000Å x 2,000Å. Notice the enlargement of the same features in the center of the previous figure.

Figure 3.13d, Gold film image sequence. This STM surface image is a close up of the image shown in Fig. 3.13c. This image is 500Å x 500Å. Notice the enlargement of the same features in the center of the previous figure

Figure 3.14a, Graphite images. These photographs reveal the differences in surface structure between graphite and the gold shown in Fig. 3.13. Here, the graphite displays a layered planar structure (a), which upon magnification shows the apparent underlying atomic structure. See Fig. 3.14b.

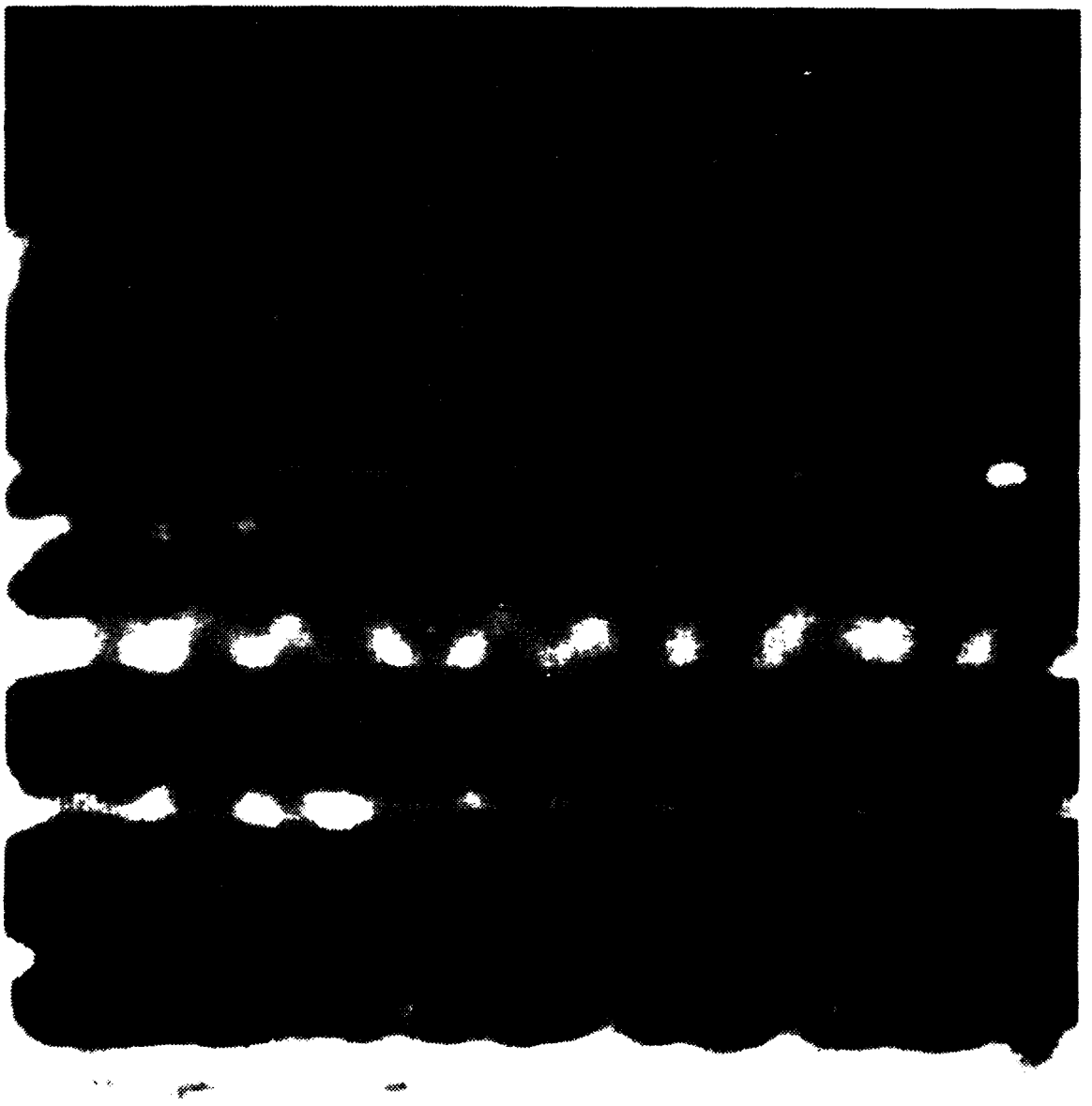


Figure 3.14b, Atomic graphite images. This image is of a graphite sample and is 85Å x 85Å. It is a close up of the graphite surface image in Fig. 3.14a. Possible atomic structure is visible.

Static I(V) Measurement

In static measurements of I(V), the tip is positioned over the area where the measurement is to be performed. The feedback circuit is allowed to equilibrate, and the tunneling current to become quiet. Depending on the sample and environment this may take from less than a second to a few minutes. The feedback circuit is then turned off, the bias voltage is ramped through the desired voltage range as the tunneling current is recorded by the computer, then the feedback circuit is turned on again. This whole process, which is completely automated, take less than 1/2 second for a 400 point I(V). This process is illustrated in Figure 3.15.

An added benefit of the static I(V) measurement is that the conductance, $\sigma \equiv dI/dV$, can be measured simultaneously. This is done by adding a high frequency, > 20KHz, ac modulation to the bias voltage ramp wave. A lock-in amplifier is then used to measure the response of the tunnel current to the voltage modulation. In this way one obtains a direct measurement of the $\sigma(V)$. For our data, however, $\sigma(V)$ was obtained from taking the numerical derivative of the measured I(V).

The benefits of the static I(V) measurement are the relative simplicity of the hardware and software necessary to make this measurement. Also, the interpretation of the data is very straightforward, because this is a direct measurement of the tunneling current as a function of the bias voltage between the tip and the sample. The only criticism of static I(V) measurements is that there could be drift in the tip-to-sample distance during the time period when the feedback is turned off, which could lead to artifacts in the data. Although this is a concern, the observed voltage output to the

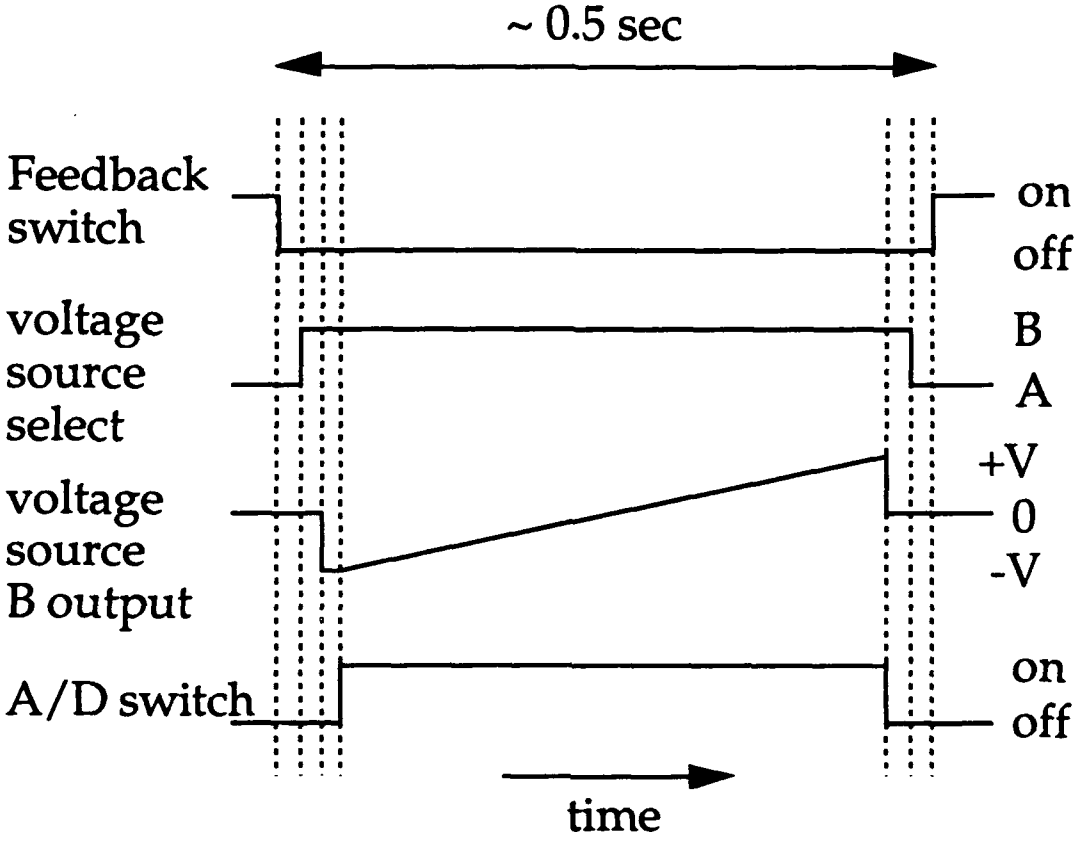


Figure 3.15, Static I(V). The above diagrams show the sequence of events for taking a static I(V) curve. First the feedback is turned off. The voltage select switch then flips from source A, the constant bias voltage, to source B, which is a D/A ramp. B outputs -V, waits to allow transients to settle, and then linearly ramps to +V. The A/D is switched on as the B output voltage begins to ramp so as to measure the tunnel current. The current measurement is then turned off, the voltage source is returned to A, and the feedback is turned back on. The entire sequence take about 1/2 second for a 400 point I(V). The addition of a small oscillating voltage (dV) to the output of source B, a lock-in amplifier and another A/D channel, allows for the simultaneous measurement of the conductance dI/dV as a function of voltage. The time between events is not to scale; rather, this diagram is to illustrate the sequence of events.

z piezo before and after the a static I(V) measurement, was identical to within 0.25 mv ($\Delta z \approx 0.001$Å), implying that the tip to sample distance had not changed significantly during the measurement. This is quite expected since the feedback loop is off for only a short time period, and because our measurements are taken at 4.2K, where mechanical drift is very slow. Nonetheless, in anticipation of situations where the static method of measuring I(V) may be inappropriate, our instrument incorporates an ability to take this measurement another way, as described below.

Dynamic I(V) Measurement

The second method by which to measure I(V) is labeled a dynamic I(V) measurement. This process does not involve turning off the feedback circuit and as such does not suffer from the criticism of a static measurement. This method is however more difficult, and the data are not quite as easy to interpret.

The measurement depends on the fact that the feedback circuit has an adjustable time constant which determines its response time. The feedback circuit does not respond to variations in the tunneling current which occur at frequencies above a set threshold. Usually, one tries to set the feedback circuit to have the shortest time constant (highest frequency) possible without becoming unstable and oscillating. The feedback circuit response time is set on the front panel of the feedback box, by selecting the resistance and capacitance used in the feedback circuit.

The dynamic measurement of I(V) depends on the fact that the feedback circuit can be set so as to not respond to a tunneling current above a

certain frequency. In this way by making the bias voltage oscillate at a frequency larger than the response frequency, the feedback circuit will not respond to each cycle but rather only respond to a time average of several cycles. The length of the time average is the set response time of the feedback circuit. Note that the first section of the feedback circuit is a rectifier; thus the bias voltage and therefore the actual tunneling current can oscillate around zero, while the feedback circuit would see a non-zero tunnel current because it averages the *magnitude* of the tunnel current.

When the frequency of the bias voltage and the response time of the feedback circuit are adjusted properly, as described above, then the feedback circuit maintains a fixed distance between the tip and sample by maintaining the average magnitude of the tunneling current. In this situation, with the feedback on, the tip to sample distance is precisely maintained while each cycle of the oscillating bias voltage provides an I(V) sweep. By using a digitizing oscilloscope or computer controlled A/D board one can obtain I(V) information with the feedback circuit on. Figure 3.16 shows the dynamic technique for obtaining I(V) measurements.

The shortcoming of this procedure is that in order to have a reliable feedback circuit, it is desirable to have it operate with a small time constant, ≈ 5 ms. In this case, the bias voltage would have to oscillate with a frequency of ~ 2KHz to assure no feedback response to individual voltage cycles. To collect data one would have to typically digitize 400 points in 0.25ms (one half cycle), for an A/D throughput of 1600KHz. This throughput is higher than the capability our digitizing system, which is 100KHz A/D throughput. In actuality it is not necessary to digitize the data. One can display the I(V) trace

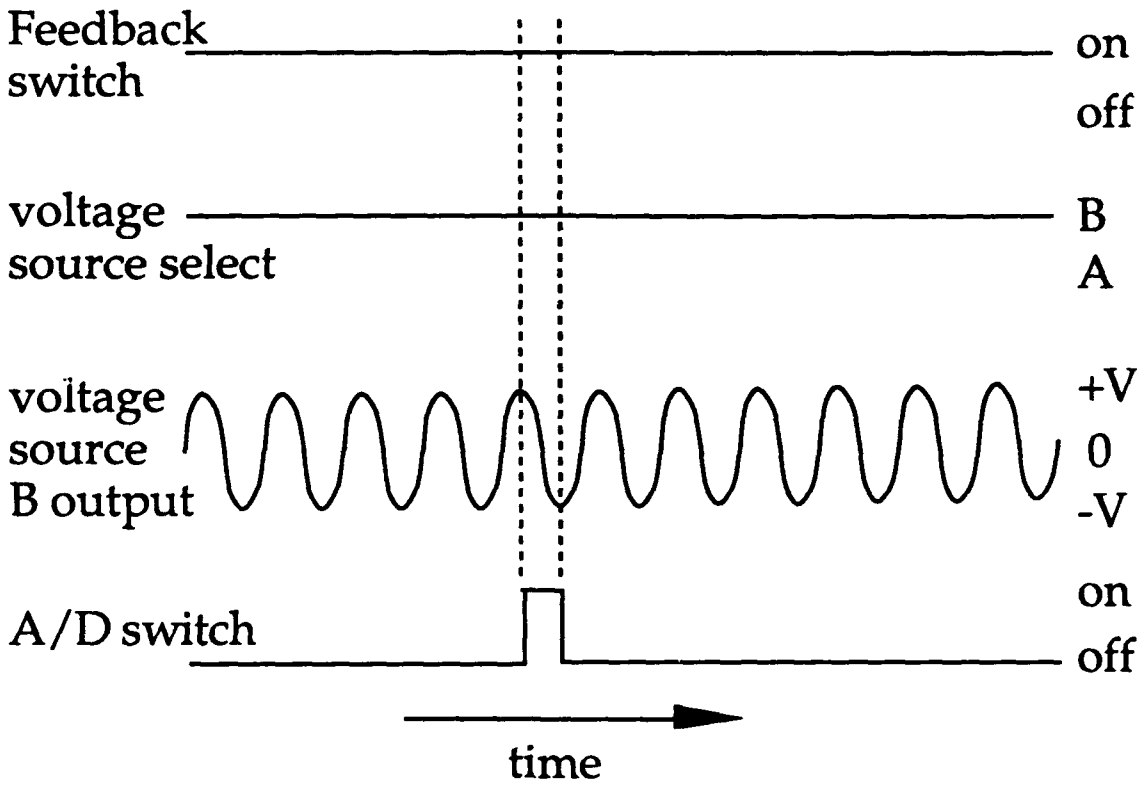


Figure 3.16, Dynamic I(V). The above sequence illustrates the process for obtaining I(V) by the dynamic method. The bias voltage source is not the standard low-noise DC supply (port A), but rather is constantly switched to the variable supply (port B) which has a high frequency, ~ 2KHz, oscillating voltage. The feedback circuit is never switched off because it is set to not respond to such high frequency tunnel current changes. The feedback circuit only responds to the time average of the magnitude of the tunnel current. As such, it constantly maintains the tunnel gap distance. To get the I(V) data, however, the data must be digitized during one half cycle of the voltage oscillation. This requires quite fast electronics, which is the reason we do not employ this method.

on an analog oscilloscope, and photograph the image. For data analysis, the images could then be digitized by a scanner or a manual digitizer.

By increasing the feedback response time, one can lower the usable bias voltage oscillation frequency and make data taking easier, but this could lead to the same concern as in the static measurement case, the variation in the tip to sample distance. In this case it would not be because the feedback is off, but rather because of its slow response time.

3-4 SAMPLES

Samples for STM imaging must be prepared in extremely clean condition if atomic resolution is to be commonly expected. Certainly the most remarkable images have been obtained from STMs in ultra-high vacuum chambers, where the sample was prepared in situ. For our instrument, samples may spend as much as an hour at atmosphere and room temperature before being loaded into the vacuum chamber of the STM. Accordingly, we choose samples which do not oxidize or tend to collect other surface contaminants. Examples are gold and carbon. Below we describe typical sample preparation for our STM and measurement parameters in our system.

3-4.1 Preparation

To fit onto our STM sample stage the sample must be from 1mm x 1mm to 2mm x 4mm in size, and no more than 3mm thick. The sample is glued onto the sample stage with GE varnish. A sheet of mica a few microns

thick goes between the sample and the stage for electrical insulation. A short magnet wire lead is put alongside the sample, and is connected with a drop of silver paint, Figure 3.17. It is then allowed to set for about 1 hour. The other end of the lead is connected by a miniature connector to a micro-coax cable for transmission up to the electronics. The stage is secured to the sample holder unit by two stainless steel screws. They should be made snug, but *not* tightened. Tightening prevents the stage from flexing during the fine mechanical approach.

It is important that the fine adjustment screw is now turned to its extremes so as to test the flexibility of the stage. It may happen that the sample will fracture when the stage is flexed, and it is better to find this out before all the later steps. To reduce the likelihood of sample fracture during fine mechanical approach, it is suggested that the sample be made the smallest practical size. The strains felt by a small sample in the middle of the flexing stage are much less than those felt by a sample glued along the entire length of the stage.

Once the stage is flexed to both extremes, the sample holder unit is ready to be attached to the tip holder unit. This is done with four stainless socket head screws. They should be tightened securely. Now, the coarse adjustment unit is slipped on the tip holder unit as shown in Figure 3.18, and under a stereo microscope the coarse tip-to-sample approach is made. The tip should be brought to within a few microns of the sample. This is most easily done by causing the tip to approach the sample until, under the highest magnification (eyepiece at 15x • objective at 7x), the reflection of the tip is clearly seen in the sample.

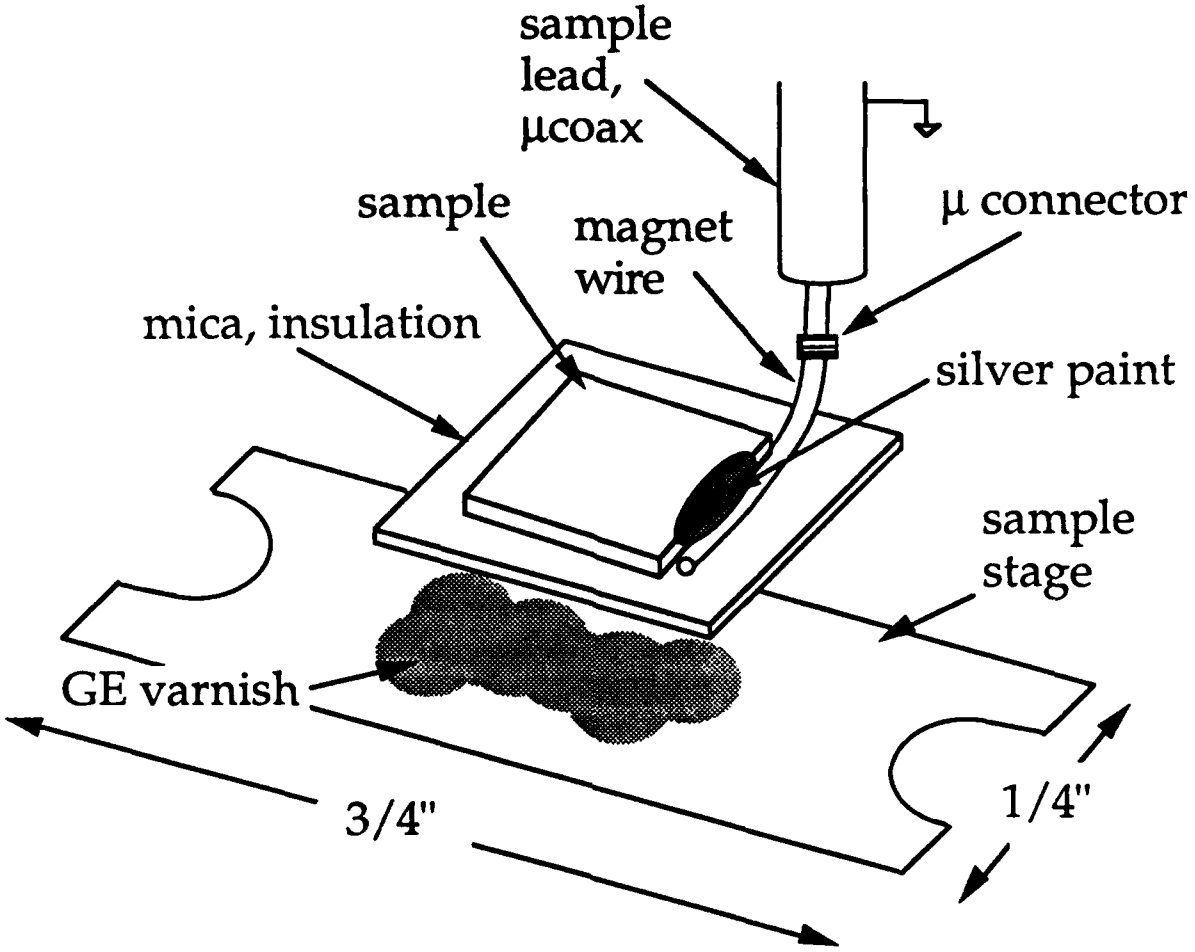


Figure 3.17, Sample mounting. The above diagram illustrates how an individual sample is mounted onto the sample stage. A thin sheet of mica is first glued to the sample stage to prevent any short circuits of the sample to the grounded sample stage. The sample is then glued to the mica and the lead is glued along side. Finally a drop of silver paint is used to electrically connect the sample and the lead.

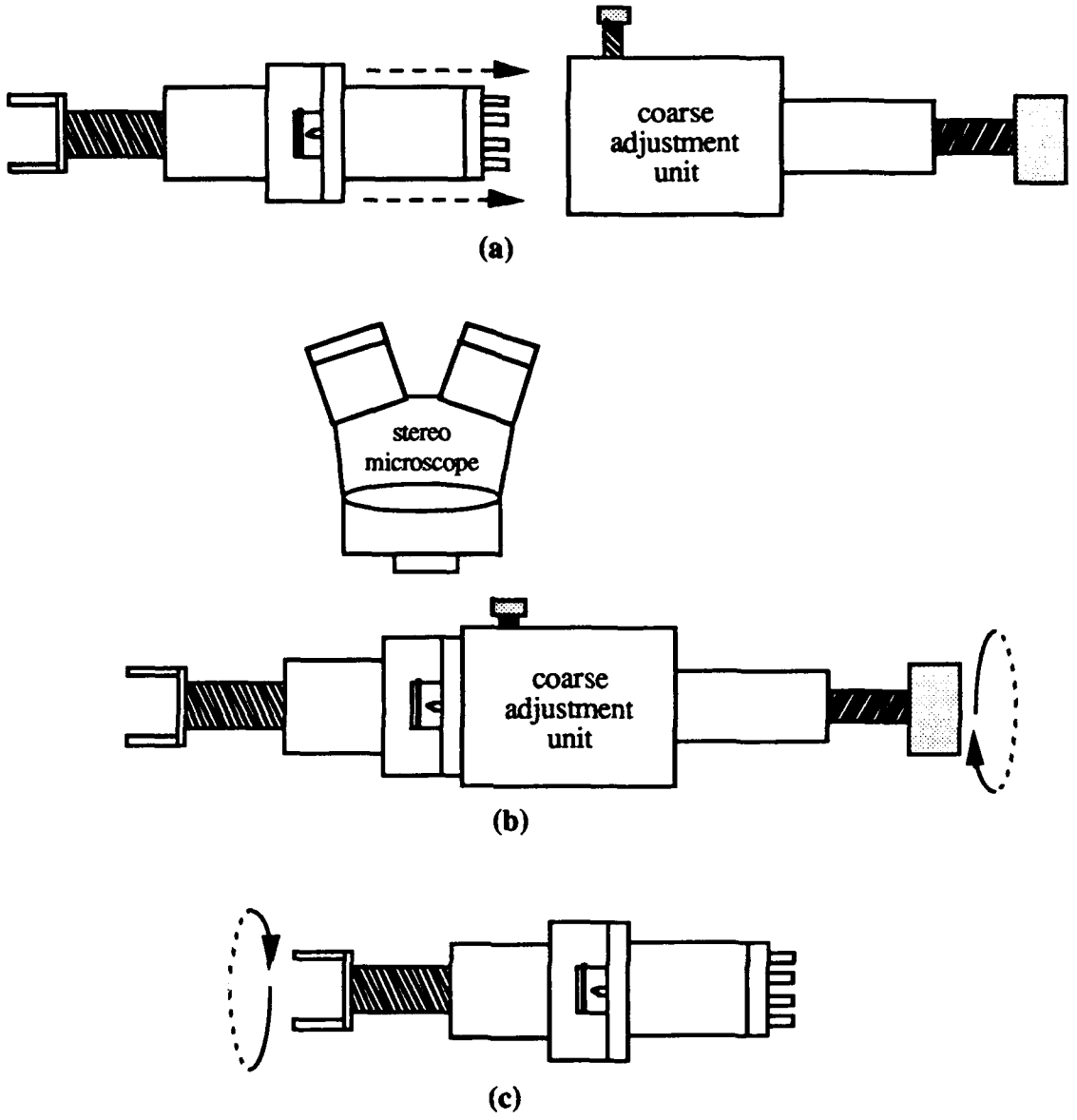


Figure 3.18, Coarse tip approach. Once the sample is mounted, the tip holder unit and the sample holder unit are secured together. The entire piece is then slid into the coarse adjustment unit as shown in (a). The coarse adjustment unit pushes the tip out of its supporting tube so as to bring it to within a few microns of the sample. The procedure, shown in (b), is observed under a stereo microscope. Caution must be taken not to crash the tip. Once this is complete, the coarse adjustment unit is removed from the tip holder. At this point the tip is still much too far for tunneling; the remaining distance is reduced by the fine mechanical approach shown in (c). However, this is only done once the whole unit is mounted (cooled if desired) and operating, not independently as shown in (c).

Once the coarse adjustment is made, the coarse adjustment unit is carefully removed from the tip holder. The entire tip and sample unit is now mounted in the copper block at the end of the stainless steel extensions, and the leads for the sample, the tip, and the piezo are attached. After wrapping the entire piece in Teflon tape for safety and to protect the leads, the unit is lowered into the outer vacuum tube and secured.

At this stage, the unit is lowered into a glass dewar, where it can be cooled with LN_2 or LHe, or kept at room temperature. The electrical connections are made by the vacuum tight connectors at the top of the STM extensions. The system is designed so that the piezo connector, and the sample and tip connector are not interchangeable. That is, there is no way to connect the electronics to the rig so as to destroy the sample by sending it the high piezo voltages.

3-4.2 Measurement

To measure the sample, the computer and electronics should now be turned on. Set the sample (or tip) voltage, to a high voltage (5-10 volts). Slowly turn the fine approach mechanism until a tunnel current registers on the oscilloscope screen. The z piezo voltage will jump back, in response to the mechanical approach. Turn the fine approach mechanism back one-eighth turn so that it disengages from the sample. Turn down the voltage to about 40mV, to bring the tip closer to the sample. The STM is now tunneling and can proceed with measurements.

To obtain an image of the surface run the "x-y scan" section of the program. An I(V) can be obtained by following the directions of the "di/dv"

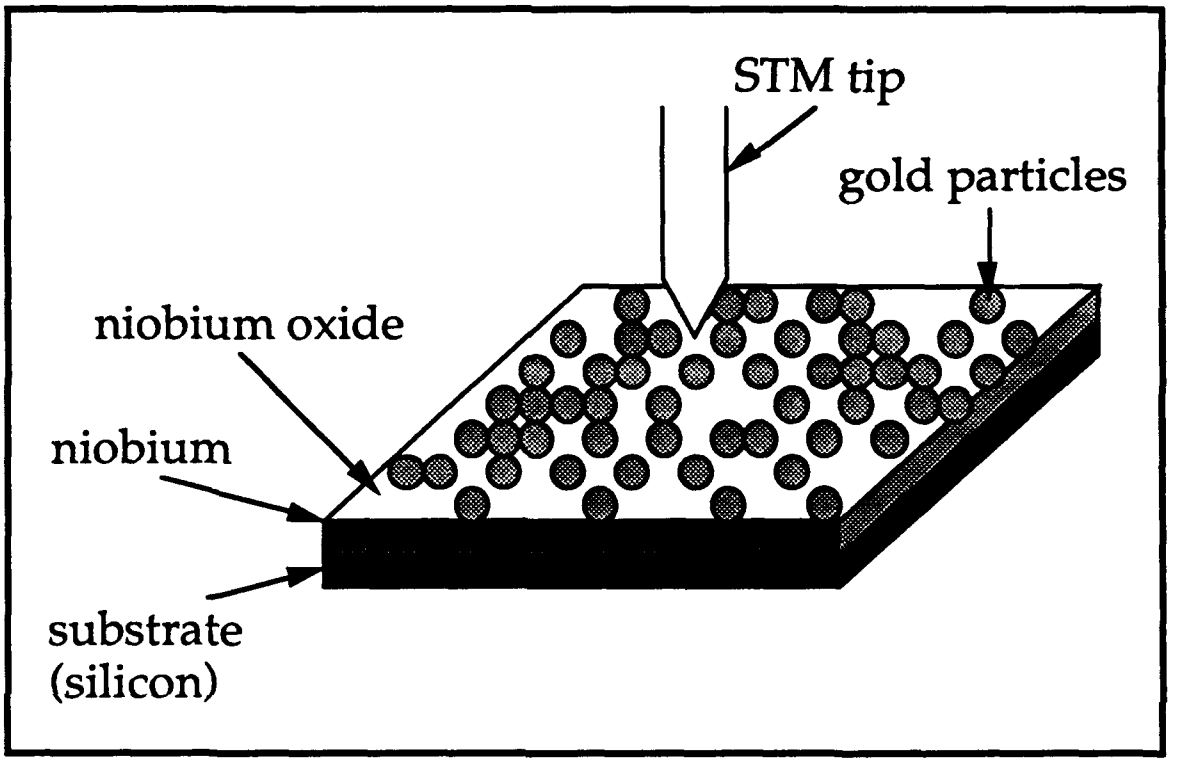


Figure 3.19, STM tunnel junctions. This diagram shows how we form the two serial tunnel junction system for studying single electron charging effects. The first tunnel junction is the vacuum gap between the STM tip and the gold particle. The second is the oxide layer between the gold particle and the niobium under-layer. The gold particles are approximately 100Å in diameter and thus have a charging energy of several millivolts. We have observed the Coulomb blockade and the Coulomb staircase in this system.

section of the program. Basically, the directions on the screen explain the operation of the unit. Detailed instructions are given in the STM Operation Guide.

For the charging experiments reported, we fabricated samples containing electrically isolated gold spheres. This was done by sputtering 5000Å niobium onto a substrate and then sputtering on a thin layer of gold, 300Å, in situ, Figure 3.19. An oxide forms on the surface of the niobium, insulating it from the gold. Under the STM, electrons tunnel from the tip to a gold particle and then tunnel to the niobium. This series of two tunnel junctions is what was used to investigate single electron charging effects.

The two junction system is shown in Figure 3.20. The niobium-oxide was a relatively low resistance tunnel junction, kΩ, compared with the vacuum tunnel junction resistance, MΩ. The asymmetry in the resistance allowed the single electron features, such as the Coulomb staircase, to become sharper and more easily described than if the two tunneling resistances were approximately equal, as will be discerned in chapter four. The asymmetry in the junction resistances is also necessary for stable STM operation. If the resistance between the gold particle and the substrate is much greater than the resistance of the junction between the tip and gold particle, then the change in tunnelling current flowing through the two junction system would not vary greatly with the distance between the tip and sample. Accordingly the feedback circuit could not effectively maintain a fixed distance between the tip and the sample.

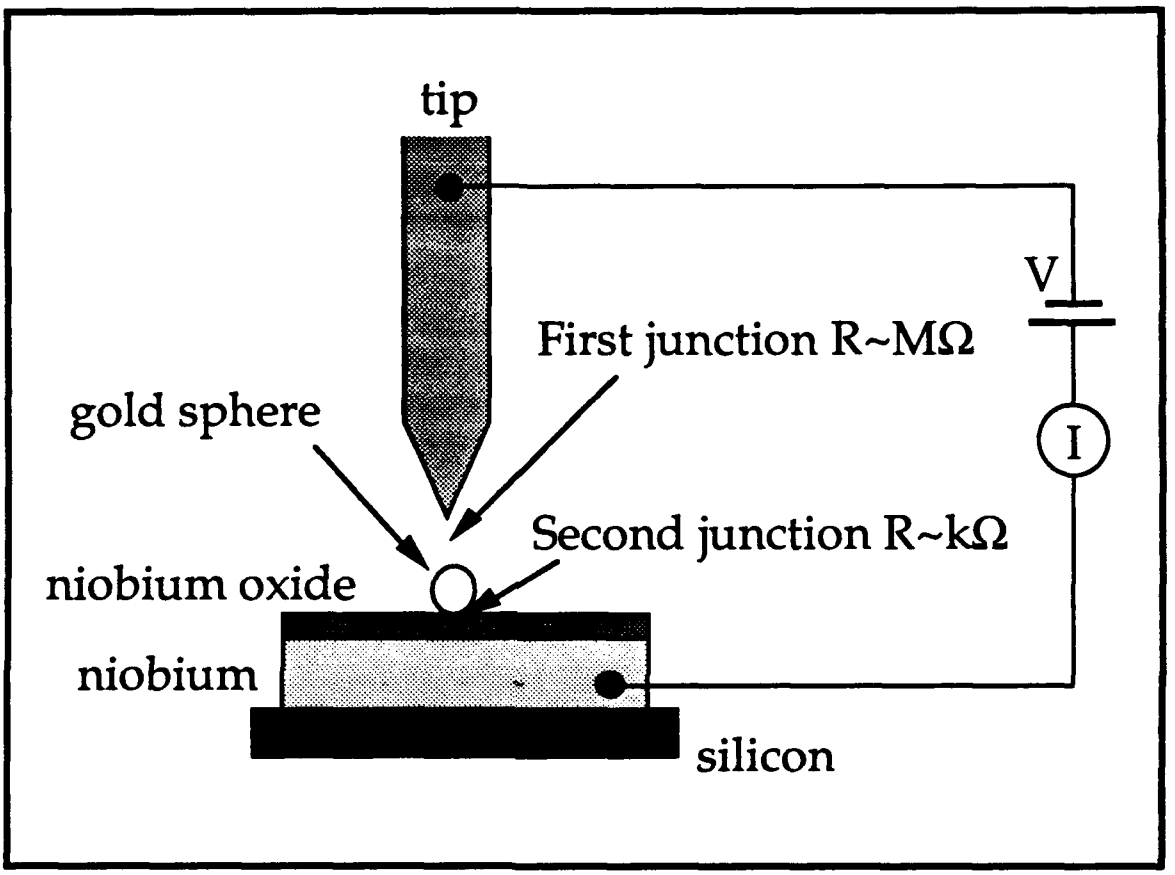


Figure 3.20, STM two junction system. The above diagram illustrates how we form the two junction system in our STM. The asymmetric junction resistances allow for more detailed features of the Coulomb staircase to be observed.

CHAPTER FOUR

SEMI-CLASSICAL JUNCTION EFFECTS

In this chapter we present a simple analytic equation for I(V) which depends on the parameters of the junctions, $C_{1,2}$, $R_{1,2}$ and Q_0, and allows us to obtain them directly from the data. An expression is obtained for the current in a voltage-biased two junction system in which the ratio of tunneling resistances $R_2/R_1 \gg 1$. Within this framework we show how I(V) curves can be separated into four distinct cases, depending on the capacitance ratio C_2/C_1 and the fractional residual charge Q_0. We explain how the individual parameters of the system can be obtained from the I(V) curve and how asymmetries are achieved when $Q_0 \neq 0$. The width of the Coulomb blockade region is also shown to depend on Q_0 and vary from the often quoted value e/C_Σ.

We also report the measurement of high-quality Coulomb staircase data which confirm our mathematical analysis. The data were taken with a low temperature STM and a granular gold film as described in chapter three. The fractional residual charge Q_0 on the center electrode was varied without an external electrode. The full orthodox theory simulations incorporating the junction parameters are in remarkable agreement with our data. The asymmetric gap-like features in the I(V) curve are seen to vary with Q_0, and can be well understood by use of the C_2/C_1 - Q_0 phase diagram.

4-1 ANALYTIC THEORY

The simple analytic theory used for understanding the data is based on the orthodox theory (chapter two) of correlated electron tunneling, and follows the derivation presented by Averin and Likharev [1991]. Figure 4.1 shows a schematic of the system. (A recent paper by Amman et al., [1991] gives a description of the two junction system, and provides a general solution; however, that solution is not a simple analytic equation and is not intuitive by inspection). The analysis begins by taking the appropriate limit, $R_2/R_1 \gg 1$, in the orthodox theory. This is the limit where we operate the experiment, and which is necessary for stable tunneling of the STM.

The particle tunneling rate for the jth junction is represented by $\Gamma^{\pm}_j(n)$, where the "+/-" refers to electrons tunneling on/off the center electrode ($n \rightarrow n\pm1$). $\Gamma^{\pm}_j$ can be easily obtained from a basic golden rule calculation [Averin and Likharev, 1991]:

$$\Gamma_j^{\pm}(n) = \frac{1}{R_j e^2}\left(\frac{-\Delta G_j^{\pm}}{1-\exp(\Delta G_j^{\pm} / k_B T)}\right) \tag{4.1}$$

where ΔG is the energy change of the system when the electron tunnels across the barrier and R_j is the tunneling resistance of the jth junction. The equations for ΔG are obtained from the electrostatic energy considerations below. For convention we take $e > 0$.

Consider two capacitors in series, biased at total voltage V. The total energy stored in the system is,

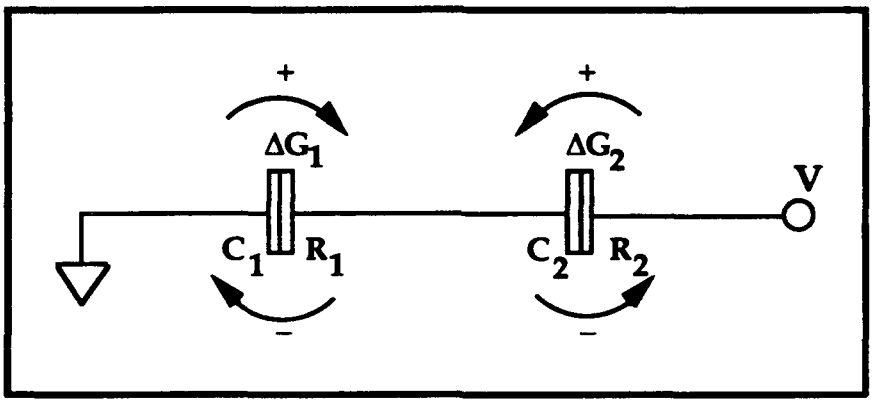


Figure 4.1, Two junction system. This diagram schematically represents our experimental situation of two normal metal tunnel junctions in series. The Orthodox theory of correlated tunneling is numerically solved to obtain the expected I(V) characteristics of the system.

$$\frac{(V-V')^2 C_2}{2} + \frac{V'^2 C_1}{2} = E_{tot} \tag{4.2}$$

C_j is the capacitance of the jth junction, $C_\Sigma = (C_1 + C_2)$, and V' is the voltage across C_1. We can solve for V' by minimizing the total energy with respect to V'. However, there exists the boundary condition that the sum of the induced charge equals the net excess charge on the center island, Q, as shown in Eqn. (4.3).

$$Q = V'C_1 - (V-V')C_2 \tag{4.3}$$

The equations. are in general most easily solved by using a Lagrange multiplier, which in this case yields,

$$V' = \frac{C_2 V + Q}{C_\Sigma} = \frac{C_2 V + ne - Q_0}{C_\Sigma} \tag{4.4}$$

In the second term we have set $Q = (ne - Q_0)$, where n is an integer and Q_0 is a possible fixed fractional component which will be discussed later. Note that when solving for V' above, Q is a constant.

Now consider one electron which tunnels across the second junction, onto the center electrode. The energy gain of the electron is its final energy minus its initial energy. The initial energy (with respect to the chemical potential when V=0) is simply $E_i = eV$. The final energy of the electron is somewhat more complicated to calculate. The energy of the added electron will of course be eV', however, it is not clear whether to use the value of V' before the addition of the new electron or after. As shown in Eqn. (4.4), V' will change by e/C_Σ with the addition of a single electron. Follow the classical approach of adding the charge ($n \rightarrow n+1$) in infinitesimally small quantities, gives:

$$E_f = \int_{ne}^{(n+1)e} V'(ne)\ d(ne) = \frac{C_2}{C_\Sigma} eV + \frac{(ne - Q_0)e}{C_\Sigma} + \frac{e^2}{2C_\Sigma} \tag{4.5}$$

This, as would be expected, is just the average of eV'(ne) and eV'(ne+1). Solving for the change in electron energy as it tunnels across junction 2, yields:

$$\Delta G_2^+ = E_f - E_i = \frac{e}{C_\Sigma}\left(\frac{e}{2} + (ne - Q_0) - C_1 V\right) \tag{4.6}$$

Following similar procedures, the general equations for ΔG are given by:

$$\Delta G_1^\pm = \frac{e}{C_\Sigma}\left(\frac{e}{2} \pm (ne - Q_0) \pm C_2 V\right) \qquad \Delta G_2^\pm = \frac{e}{C_\Sigma}\left(\frac{e}{2} \pm (ne - Q_0) \mp C_1 V\right) \tag{4.7}$$

The current can be obtained from:

$$I(V) = e\sum_{n=-\infty}^{\infty}\sigma(n)\left[\Gamma_2^+(n) - \Gamma_2^-(n)\right] = e\sum_{n=-\infty}^{\infty}\sigma(n)\left[\Gamma_1^-(n) - \Gamma_1^+(n)\right] \tag{4.8}$$

where $\sigma(n)$ is the ensemble distribution of the number of electrons on the center electrode. The expression on the far right in Eqn. (4.8) gives the net current through junction 1; this is equal to the net current through junction 2, the middle expression, as long as the there is no charging of the center electrode. That is, there is a steady state situation. This steady state condition, described by Eqn. (4.9), requires that the net probability for making a transition between any two adjacent states in steady state is zero (i.e., the number of electrons entering the particle per unit time equals the number of electrons leaving the particle per unit time); thus:

$$\sigma(n)\left(\Gamma_1^+(n) + \Gamma_2^+(n)\right) = \sigma(n+1)\left(\Gamma_1^-(n+1) + \Gamma_2^-(n+1)\right) \tag{4.9}$$

Since the $\Gamma^\pm_j$ are known from (4.1) and (4.7), this allows one to solve for the distribution $\sigma(n)$, subject to the normalization condition,

$$\sum_{n=-\infty}^{\infty}\sigma(n)=1. \tag{4.10}$$

We can thus numerically solve for I(V) from (4.8), as will be discussed below.

To obtain a simple analytic expression for I(V) we consider the limit where $R_2/R_1 \gg 1$. The most probable number of electrons on the center electrode, n_0, i.e., the value for which $\sigma(n_0) \geq \sigma(n_0 \pm 1)$, is then primarily determined by junction 1 because $\Gamma_1 \gg \Gamma_2$ in Eqn. (4.9). Combining Eqn. (4.9) and the maximum probability condition requires that

$$\frac{\Gamma_1^-(n_0+1)}{\Gamma_1^+(n_0)} \geq 1 \qquad\qquad \frac{\Gamma_1^+(n_0-1)}{\Gamma_1^-(n_0)} \geq 1 \tag{4.11}$$

Which yields

$$\frac{1}{e}(-C_2V+Q_0-e/2) \leq n_0 \leq \frac{1}{e}(-C_2V+Q_0+e/2) \tag{4.12}$$

in the limit where $|\Delta E_1| \gg k_BT$, because in this limit Γ_1 is given by Eqn. (4.13).

$$\Gamma_1^{\pm}(n_0)=\begin{cases}\dfrac{1}{R_1C_\Sigma}\left(-\dfrac{1}{2}\mp(n_0-Q_0/e)\mp C_2V/e\right) & \text{for } \Delta E_1^{\pm} < 0 \\ 0 & \text{for } \Delta E_1^{\pm} > 0\end{cases} \tag{4.13}$$

$\sigma(n)$ is expected to be sharply peaked if $|\Delta E_1| \gg k_BT$. That is, $\sigma(n) \approx \delta_{n,n_0}$ at low temperatures. The net current (4.8) is then given by,

$$I(V)=e\left[\Gamma_2^+(n_0)-\Gamma_2^-(n_0)\right] \tag{4.14}$$

Note that junction 1, having a much higher tunneling rate, determines n_0, while junction 2, with the smaller tunneling rate, responds to this n_0 as a constant and adjusts the current correspondingly. For low temperatures such that $|\Delta E_2^{\pm}(n_0)| \gg k_B T$, we can simplify $\Gamma_2(n_0)$ to obtain

$$\Gamma_2^{\pm}(n_0) = \begin{cases} \dfrac{1}{R_2 C_\Sigma}\left(-\dfrac{1}{2} \mp (n_0 - Q_0/e) \pm C_1 V/e\right) & \text{for } \Delta E_2^{\pm} < 0 \\ 0 & \text{for } \Delta E_2^{\pm} > 0 \end{cases} \tag{4.15}$$

This means that I(V) = 0 when

$$(-e/2 + n_0 e - Q_0)/C_1 \le V \le (e/2 + n_0 e - Q_0)/C_1 \tag{4.16a}$$

This is the so-called Coulomb Blockade. Outside of this voltage range I(V) is given by [Averin and Likharev, 1991]:

$$I(V) = \frac{1}{R_2 C_\Sigma}\left(-(n_0 e - Q_0) + C_1 V - \frac{e}{2}\,\mathrm{sgn}(V)\right) \tag{4.16b}$$

where n_0 is obtained from (4.12). Eqns. (4.12) and (4.16) form the basis for interpreting our data in the limit of $R_2/R_1 \gg 1$. They are used qualitatively to identify the "case" to which a given I(V) curve belongs and to determine parameter values from specific features in the data. These parameter values are then used in Eqn. (4.8), the exact I(V) solution, for quantitative comparison with the data.

In a two junction system, the most commonly observed feature of a Coulomb Blockade is an offset in the asymptote of the I-V curve [van Bentum et al., 1988]. Our analysis gives more detail. Eqn. (4.12) shows that

$Q = (n_0 e - Q_0)$ jumps from $(-C_2V - e/2)$ to $(- C_2V + e/2)$ each time n_o changes by one. From (4.16b) the corresponding currents are V/R_2 and $(V - e/C_\Sigma)/R_2$, for $V>0$. Thus (4.16b) describes a series of sloping steps within an envelope with slope $dI/dV = 1/R_\Sigma \approx 1/R_2$, and *average* offset $e/2C_\Sigma$. When the capacitances are as small as 10^{-18} - 10^{-19} F, however, the asymptotic limit is not well defined in our range of measurement, which is only a few volts. The data in this relatively small voltage region show detailed features about the conduction onset in a two junction system. In particular, the data show that there are distinct cases which differ according to whether the onset of conduction is caused by: (a) n_0 changing, because V reaches the limits in Eqn. (4.12) before reaching those in Eqn. (4.16a), or, (b) overcoming the Coulomb blockade of junction 1, because V reaches the limits in Eqn. (4.15a) before those in Eqn. (4.12). The former causes a discrete jump in current while the latter causes the onset of a linear increase in current. Taking account of positive and negative voltages separately, we distinguish four cases which are illustrated schematically in Figure 4.2.

The C_2/C_1 - Q_0 phase diagram can also be directly obtained by considering the set of values of Q_0 and V which give $\Delta G_1 = 0$ or $\Delta G_2 = 0$. These values are obtained from Eqn. (4.7) by setting $\Delta G_i = 0$, as shown in Eqn. (4.17) below.

$$\pm C_2V = -\frac{e}{2} \mp (ne - Q_0) \quad \text{for } \Delta G_1^{\pm} = 0 \tag{4.17a}$$

$$\mp C_1V = -\frac{e}{2} \mp (ne - Q_0) \quad \text{for } \Delta G_2^{\pm} = 0 \tag{4.17b}$$

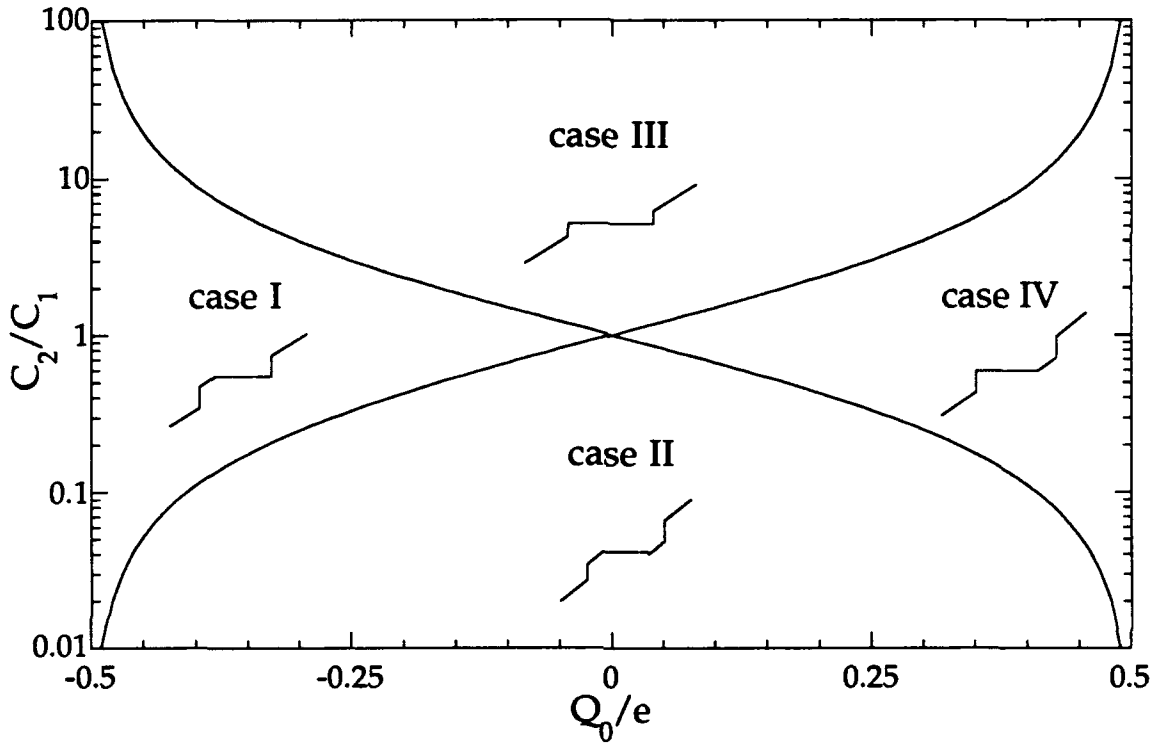


Figure 4.2, I(V) phase diagram. The above diagram shows the four qualitatively different cases of I(V) curves. Within each case is shown a small trace which is representative of data in that region, although the shape of the trace can be varied greatly within each case by changing Q_0 or C_2/C_1. It is the order of the conduction onsets which is fixed in each case.

Fig. 4.3 shows the contours where $\Delta G_1 = 0$ or $\Delta G_2 = 0$ for several different values of n, as a function of Q_0 and V. Note that the $\pm$ in ΔG_i does not change this diagram since that only shifts the contour line ($\Delta G_i = 0$) vertically by an amount equal to changing n by 1, i.e., the set of $\Delta G^+_i = 0$ are identical with the $\Delta G^-_{i+1} = 0$ contours. Each diamond-shaped section is a location with a fixed n which yields the lowest energy.

The C_2/C_1 - Q_0 phase diagram is obtained by considering the conduction onsets as one varies V for a particular Q_0. For example, in Figure 4.3 consider $Q_0 = -0.1e$, as one positively ramps the voltage (moving up along the V axis) the first intersection is with a $\Delta G_1 = 0$ contour, giving a positive current step. The first intersection when negatively ramping the voltage (moving down along the V axis) is also with a $\Delta G_1 = 0$ contour giving a negative current step. However, if $Q_0 = -0.3e$ then the the first contour crossed when the voltage is positively ramped is a $\Delta G_2 = 0$ contour, which yields a linear increase in the current, while the first contour crossed when the voltage is negatively ramped is with a $\Delta G_1 = 0$ contour, giving a negative voltage step. The point at which the transition from one qualitative type of I(V) to the other is obtained by calculating the Q_0 where the ΔG_1 and the ΔG_2 contours intersect. From Figure 4.3 and basic geometry, it is clear that for $Q_0 < 0$, this point, Q_0^-, occurs at,

$$Q_0^- = \left(\frac{e}{2}\right)\left(\frac{C_2 - C_1}{C_2 + C_1}\right) \tag{4.18a}$$

Similarly, for $Q_0 > 0$, the transition point occurs at,

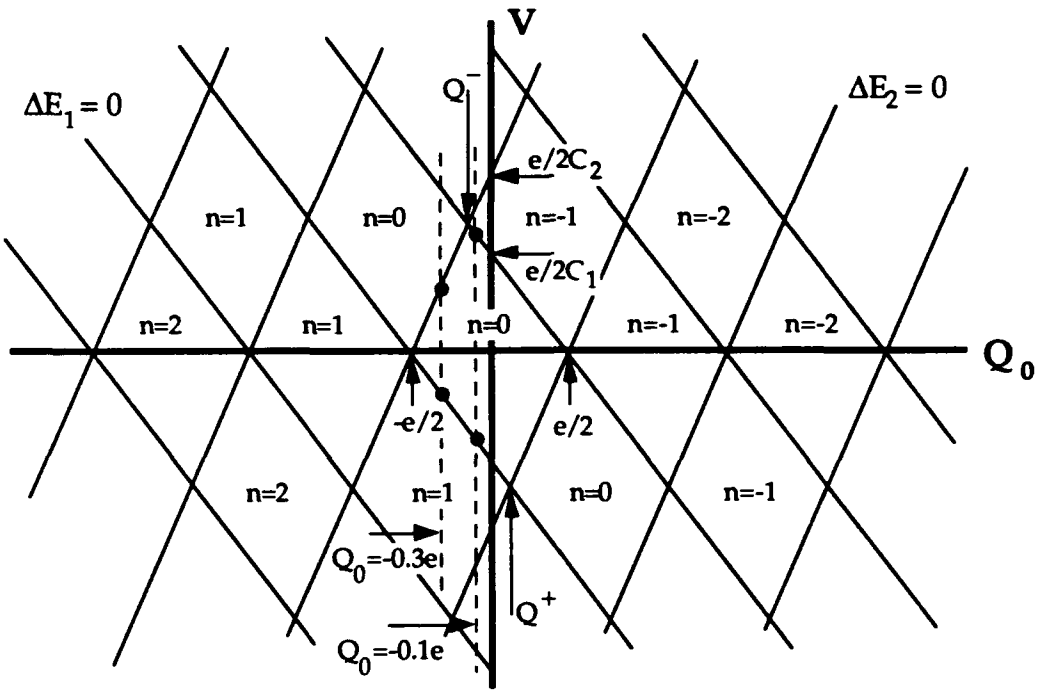


Figure 4.3, Two junction energy contours. Each line in the above diagram is a contour for a fixed n, in the V-Q_0 phase space where $\Delta G_1 = 0$ or $\Delta G_2 = 0$. Each plaquet represents a stable state with a fixed value of n. In our case where $R_2/R_1 \gg 1$, the value of n is well defined (the distribution of n is narrow), the diagram above shows the distribution of n values across the phase space. It is easily observed that n only changes its value, causing a step in the I(V), when a $\Delta G_1 = 0$ contour is crossed. That is, junction 1 determines n, as expected from previous analysis.

$$Q_0^+ = \left(\frac{e}{2}\right)\left(\frac{C_1 - C_2}{C_1 + C_2}\right) \qquad (4.18b)$$

The graph of Q_0 versus C_2/C_1 in Eqns. (4.18a) and (4.18b) give the boundaries between the different cases shown in Figure 4.2.

The above analysis had $C_2/C_1 < 1$ (this is implicit in Fig. 4.3). If $C_2/C_1 > 1$, then the Eqns. for Q_0^- and Q_0^+ would be interchanged. This is of course expected from symmetry. Nonetheless, the I(V) case boundaries would be the same. Lastly, Figure 4.3 shows that as one continuously increases V, further intercepts with the $\Delta G_1 = 0$ contour lines will cause the a repeated current steps (Coulomb staircase) in the I(V).

4.2 DATA

Our experimental setup consisted of a low temperature scanning tunneling microscope, tunnelling into a gold surface. The STM used a chemically etched platinum-iridium tip. Images of the gold surface show a very grainy surface. It is believed that the two-junction system was created by the presence of a gold grain between the tip and the bulk of the sample. The sample was voltage-biased and the data were taken at 4.2K.

The four cases identified above are exactly what is observed in our experiments. By noting the sequence of conduction onsets, one can identify the case to which a measured I(V) belongs, and then proceed to experimentally determine the model parameters, including Q_0. In our approximation, where $R_2 \gg R_1$ and $T \approx 0$, the Eqns. (4.12) and (4.15) give

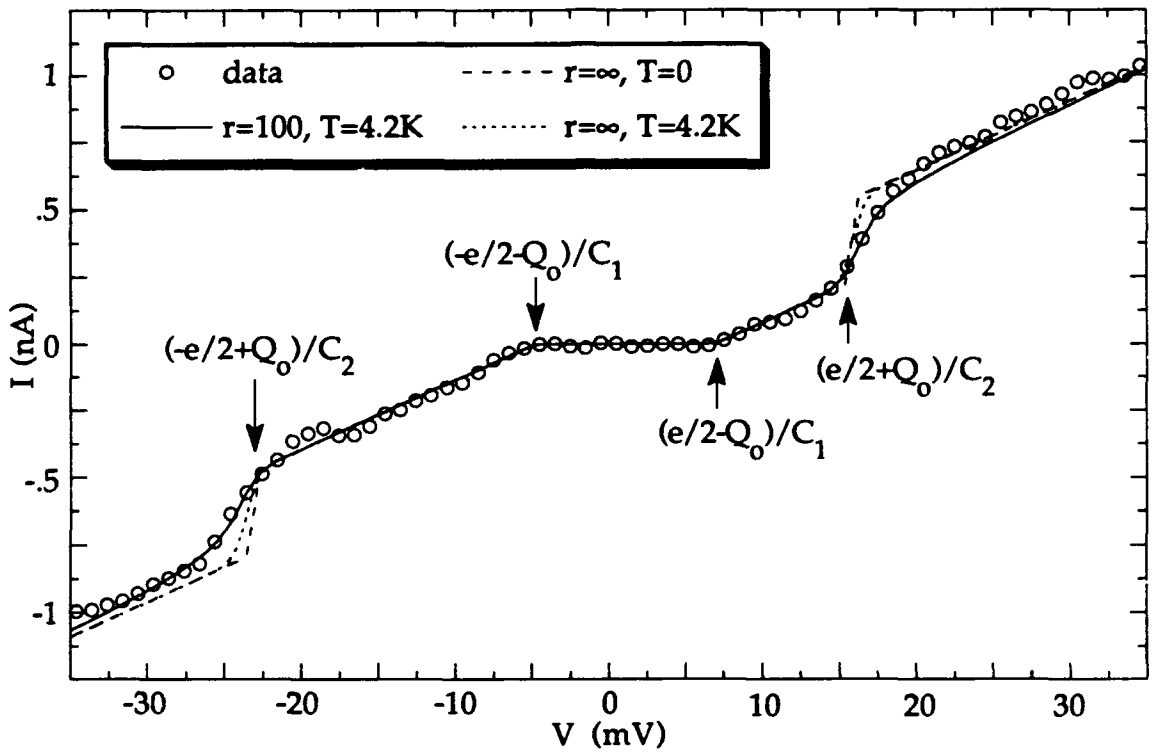


Figure 4.4, I(V) curve for case II. This trace is representative of case II, where the linear conduction limits are reached before the step conduction limits. The dashed curve is a graph of Eqn. (4.16), in the limit of T=0 and r=∞. The dotted curve shows the effect of a finite temperature, T=4.2K. The solid curve is the full numerical solution of the orthodox theory Eqn. (4.8), using the parameters: C_1=1.36x10^{-17}F, C_2=4.05x10^{-18}F, R_1=0.3MΩ, R_2=29.3MΩ and Q_0=-0.096e. These were obtained from the data features noted in the diagram above.

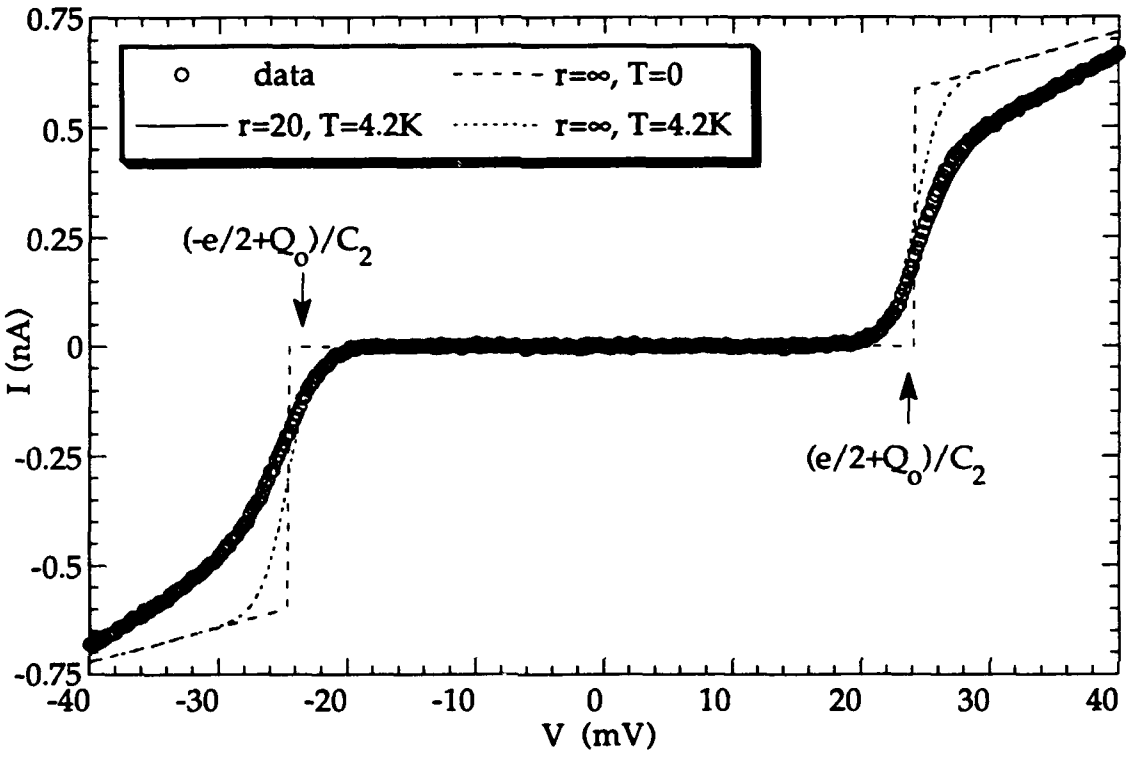


Figure 4.5, I(V) curve for case III. The above trace is representative of the case II region of Figure 4.2. The dashed and dotted curves are different levels of approximation as explained earlier. The solid line is the full numerical solution of the orthodox theory Eqn. (7), and lies directly on our measured data. There is a strong resemblance between I(V) traces in case II and possible superconducting gap measurements made with STMs by other researchers. *The measured parameters are:* C_1=1.64x10^{-18}F, C_2=3.28x10^{-18}F, R_1=2.0MΩ, R_2=39.2MΩ and Q_0=-0.005e, which were used in the orthodox theory simulation.

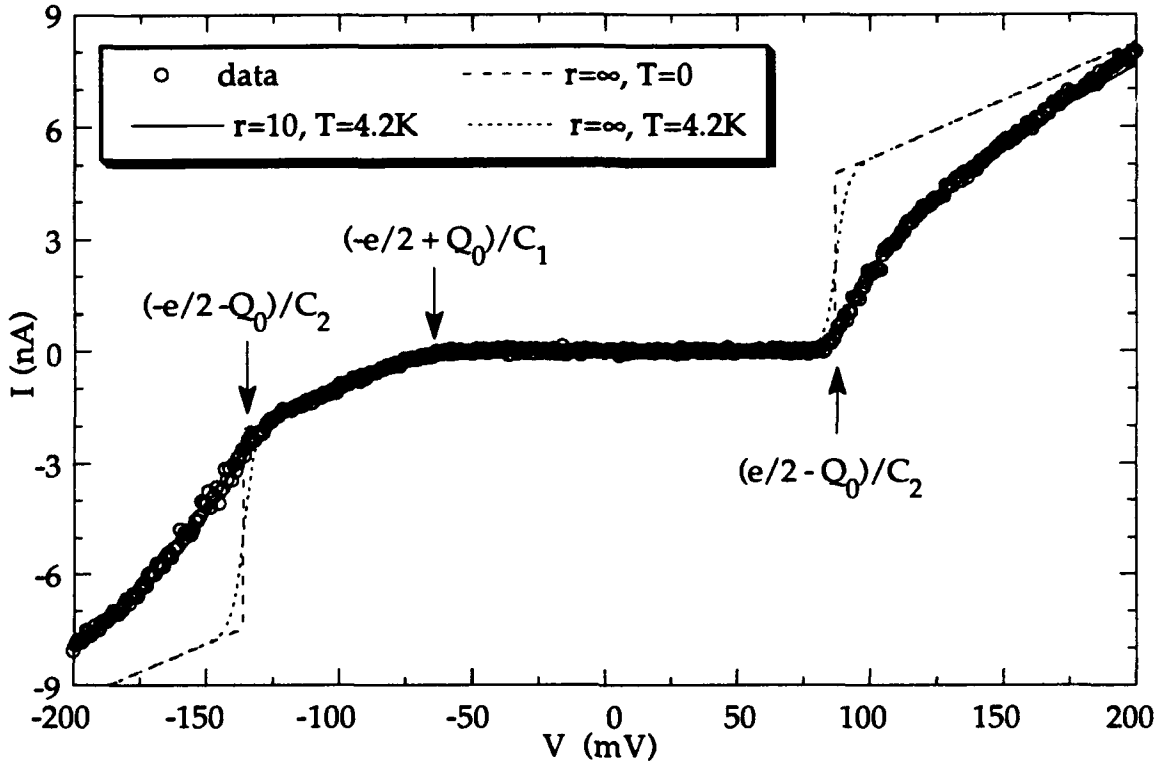


Figure 4.6, I(V) curve for case I. This diagram shows a case I trace, which has the interesting feature of being strongly asymmetric. In this case the voltage onset for linear conduction is reached first when applying a negative voltage, while the onset for step conduction is reached first when applying a positive voltage. The dashed and dotted curves are different levels of approximation as described earlier, and the solid curve is the full numerical solution of the orthodox theory Eqn. (7). Our data are in full agreement with the orthodox prediction. The measured parameters are: C_1=7.2x10^{-19}F, C_2=4.05x10^{-19}F, R_1=1.7MΩ, R_2=16.6MΩ and Q_0=-0.11e.

simple expressions for distinct features in the data. For example, from (4.12) we obtain that the difference between the threshold voltages for the first positive step and first negative step is e/C_2, while their sum is $2Q_0/C_2$. Similarly, Eqn. (4.15) shows that the plateau slope on a step (n_0 constant) is C_1/R_2C_Σ. Some of these quantities are noted in Fig. 4.4, 4.5 and 4.6. These preliminary parameter values are refined for finite temperatures by using Eqn. (4.8), to compute the dotted curve which has $r \equiv R_2/R_1 = \infty$ and T=4.2K. Since, the position of the noted features are *not* affected be changing r, we then simply decrease r until the best fit is obtained. This is the solid theoretical trace (usually hidden by the experimental data points). The dashed curve is Eqn. (4.16) calculated with the same junction parameters except with $r=\infty$ and T=0, to show explicitly the nature of the finite temperature effects in our data.

Rounding in the steps arises from two sources, a non-zero temperature and a finite R_2/R_1. Finite temperature causes a symmetric broadening of the current step jumps. In our experiment, at 4.2K, $k_BT < 0.02(e^2/C_\Sigma)$ and thermal effects are small. The finite ratio R_2/R_1 leaves the linear conduction onsets very sharp, but the step onsets ($n_0 \rightarrow n_0 \pm 1$) are no longer vertical jumps connecting plateaus with universal slope C_1/R_2C_Σ, as in the simplified result Eqn. (4.16). Rather, the onsets have a finite slope and the step top rounds toward a plateau slope of C_1/R_2C_Σ, see Fig. 4.7. The step onset can be distinguished from a linear increase by noting that the slope of the linear increase is the same as the slope of all the step plateaus, while the step onset will have a much larger slope. This distinction can be clearly seen in Figs. 4.4 and 4.7.

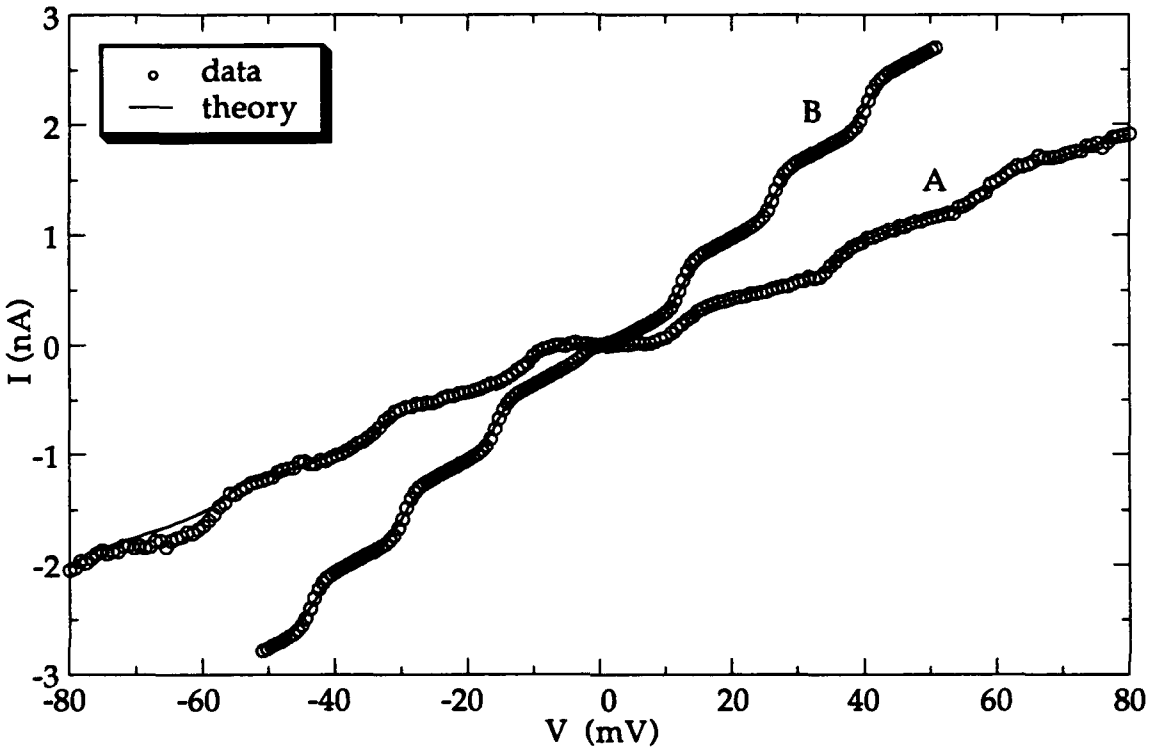


Figure 4.7, Coulomb staircase data. The numerical solution of the orthodox theory continues to fit our data outside of the central region, where several steps are seen. Note that the junctions' parameters used in the simulation were obtained only from the features in the central voltage region as described earlier. By moving the STM tip closer to the sample, we go from A to B, decreasing the total resistance of the system (asymptotic slope).

Fig. 4.7 shows data in a wider voltage range where successive steps become apparent, forming the Coulomb staircase. From Eqns. (4.12) and (4.16) it is clear that the width of the individual steps is e/C_2, and that the slope on the step plateau is C_1/R_2C_Σ, in all cases; this agrees extremely well with our data, where the values of C_1 and C_2 are already uniquely obtained from the central region. Also, as Q_0 approaches $\pm e/2$ in any case, the width of the zero conductance region goes to zero as shown in Fig. 4.8. This is the single electron transistor effect in which by changing Q_0 from 0 to $\pm e/2$ we change the conductance of the center region from zero to a finite value. A close look at measurements reported by other researchers shows that their results are well accounted for in our model. In particular some unexplained asymmetric traces are perfect examples of cases I and IV [McGreer et al., 1989; Wilkins et al., 1990; van Bentum et al., 1988]. It is also worth noting how closely class III traces resemble superconductor-normal metal STM measurements reported as showing large superconducting gaps [Wang et al., 1990; Viera et al., 1989].

4-3 Numerical Simulation

The orthodox theory which is compared with our data is obtained by numerically solving Eqn. (4.8). The program outputs the current I in units of $e/(R_\Sigma C_\Sigma)$ as a function of V in units of e/C_Σ. The input information is R_2/R_1, C_2/C_1, T in units of $k_BT/(e^2/C_\Sigma)$ and Q_0 in units of e.

The state of the particle is described by the variable n, the charge on the particle. Because of finite temperatures, there are fluctuations in n and the state of the particle is better described by a distribution of several values of n.

(There are also quantum fluctuations in n, which will be addressed in the next chapter). This is where the numerical simulation begins. For given input parameters we solve for n_{max} and n_{min} which are the values of n which to increase or decrease by one requires a $\Delta G^{\pm}$ of 20 times k_BT, respectively. That is,

$$n_{max} = n \quad \ni \quad \Delta G^{+}_{1,2}(n+1) \geq 20k_BT \tag{4.19a}$$

$$n_{min} = n \quad \ni \quad \Delta G^{-}_{1,2}(n-1) \geq 20k_BT \tag{4.19b}$$

They are given by,

$$n_{max} = \max\begin{cases} \text{int}[VC_1 + Q_0 - 1/2 + 20T + 1] \\ \text{int}[-VC_2 + Q_0 - 1/2 + 20T + 1] \end{cases} \tag{4.20a}$$

$$n_{min} = \min\begin{cases} \text{int}[VC_1 + Q_0 + 1/2 - 20T - 1] \\ \text{int}[-VC_2 + Q_0 + 1/2 - 20T - 1] \end{cases} \tag{4.20b}$$

This form for n_{min} and n_{max} is obtained by using Eqn. (4.8) for $\Delta G^{\pm}_{1,2}$, and solving Eqn. (4.19). We have used the normalized units shown above and have set $e = C_\Sigma = 1$, e.g., the $e^2/2C_\Sigma$ term in Eqn. (4.8) yields the 1/2 in Eqn. (4.19). The **int** function gives the largest integer contained in a number x, and is defined as

$$\text{int}[x] = n \quad \ni \quad n \in \mathbf{Z} \quad \text{and} \quad 1 > x - n \geq 0 \tag{4.21}$$

Having set the limits on n, $\Gamma(n)$ is then calculated for each value of n between n_{min} and n_{max}. Using Eqn. (4.11) we then solve for n_0. Setting $\sigma(n_0)=1$, the relative values for $\sigma(n)$ are then obtained by using the steady state

Eqn. (4.9). We normalize $\sigma(n)$ by the condition shown in Eqn. (4.10), where the sum runs from n_{min} to n_{max}. The current is now obtained by using Eqn. (4.8). This entire process is repeated for each value of the voltage in the voltage range, to yield an entire I(V) sweep. The time required to calculate a 400 point I(V) is about fifty seconds.

This numerical simulation program is labeled "transs0" and can be found on the E: drive of the Dell computer, under the SET directory. It is written in Fortran, and was originally created by A. Karatkov and K. K. Likharev. A revised version, created by the author, which is faster and simpler, resides on drive F: This version is more suitable for having the results transfered to the Macintosh computers for data analysis and graphing.

4-4 Q_0 Dependence

Q_0 is the fractional component of Q, the charge on the center electrode. Of course, the charge on the middle electrode is quantized in units of e, however, the Fermi level of the middle electrode may be shifted by an outside electric field. The effect of the electric field shifting of Fermi level is continuous and can be incorporated into our analysis by adding a fractional charge term to the center electrode. The effect of Q_0 on the conductance properties of the two junction system is seen in the series of simulation traces shown in Figs. 4.8 and 4.9. The most prominent effects of Q_0 are limited to the small voltage region, where the value of Q_0 determines the voltage of the conduction onsets. Also, the zero bias conductance changes from zero to a finite quantity, when Q_0 is changed from 0 to $\pm e/2$. This can be seen in Fig. 4.8

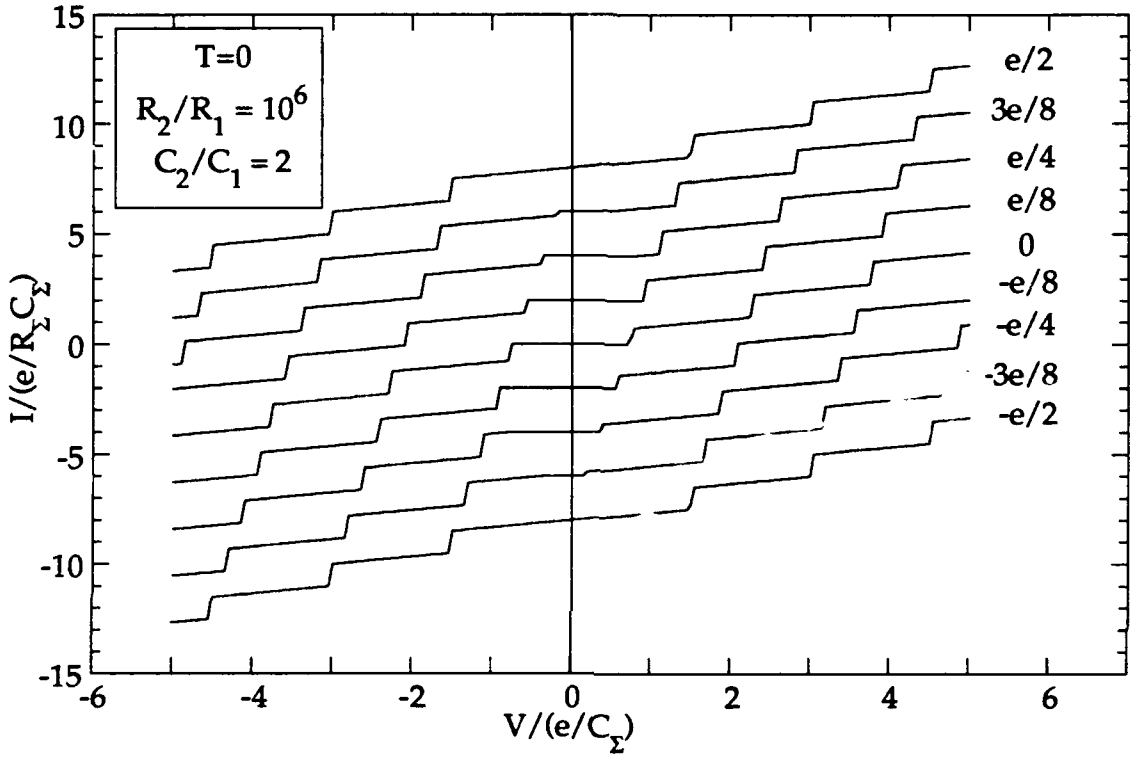


Figure 4.8, Q_0 effects I. This series of I(V) traces shows the expected adjustment in conduction properties of a two junction system with $C_2/C_1 > 1$, as Q_0 is varied. As anticipated from Figure 4.3, we observe I(V)s from Case I, Case III, and Case IV as Q_0 is varied from -e/2 to e/2. Traces are offset vertically.

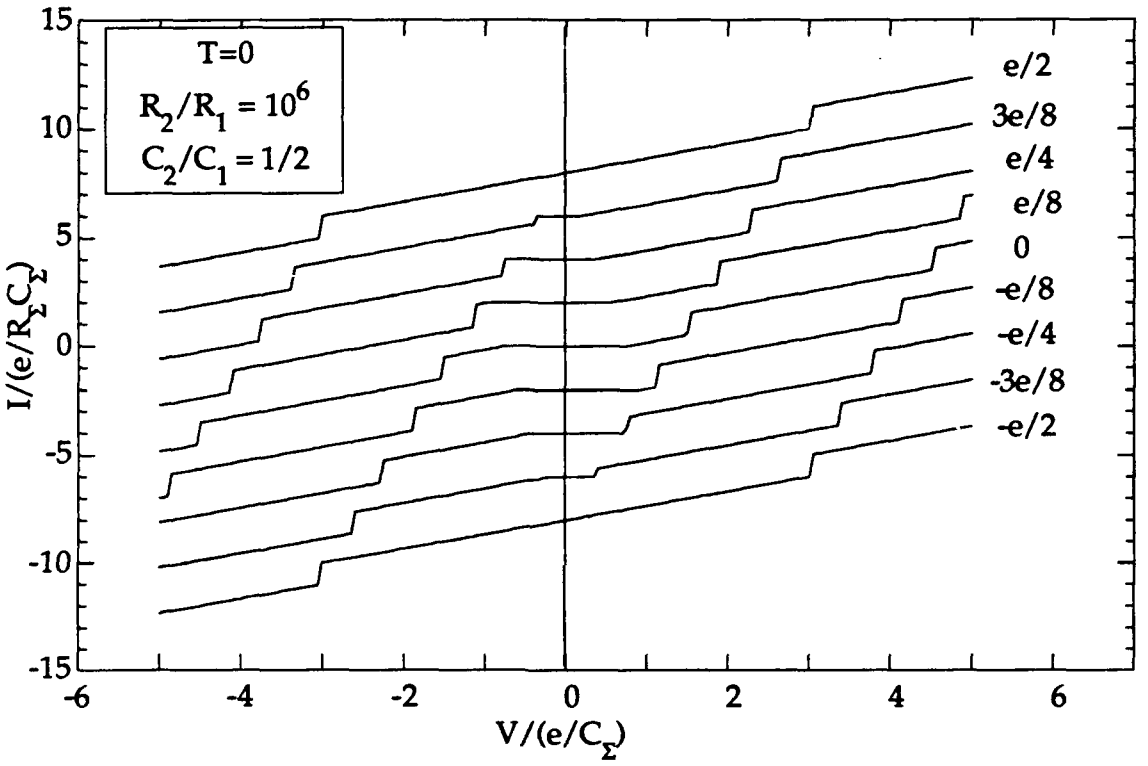


Figure 4.9, Q_0 effects II. This series of I(V) traces shows the expected adjustment in conduction properties of a two junction system with $C_2/C_1 < 1$, as Q_0 is varied. As anticipated from Figure 4.3, we observe I(V)s from Case I, Case II, and Case IV as Q_0 is varied from -e/2 to e/2. Traces are offset vertically.

and 4.9, and is an important property of the two junction system. This effect has been labeled the "single electron transistor" effect.

In order to vary Q_0, a third electrode is usually fabricated to capacitively couple to the center island (particle). By applying a voltage to this electrode Q_0 can thus be continuously varied. This method of varying Q_0 is unfeasible in our experiment because the length scales used are much smaller than current fabrication techniques allow. However, in our measurements we do observe the effects of a finite Q_0, and in fact are also able to vary it continuously without using a third electrode. This is described in detail below.

In our experiment Q_0 is thought to originate from the difference in work functions of the different metals used in the junctions. If so, the more fundamental variable is the contact potential across the junction. Q_0 (mod e) is obtained from,

$$Q_0 = \frac{1}{e}[C_1 \Delta\varphi_1 - C_2(d) \Delta\varphi_2] \tag{4.22}$$

where the $\Delta\varphi_1$ and $\Delta\varphi_2$ are the contact potentials across junctions 1 and 2 respectively, and d is the distance between the tip and sample. By changing d, we can change the capacitance C_2 and hence change Q_0 as well. This allows us to move systematically around the phase space of Figure 4.2, taking data in the various sections.

Figure 4.10 shows two measured I(V) traces where we varied the width of the zero conduction region (Coulomb blockade) by changing d, which not only changes the tunneling resistances, but also changes C_2 and hence Q_0, as

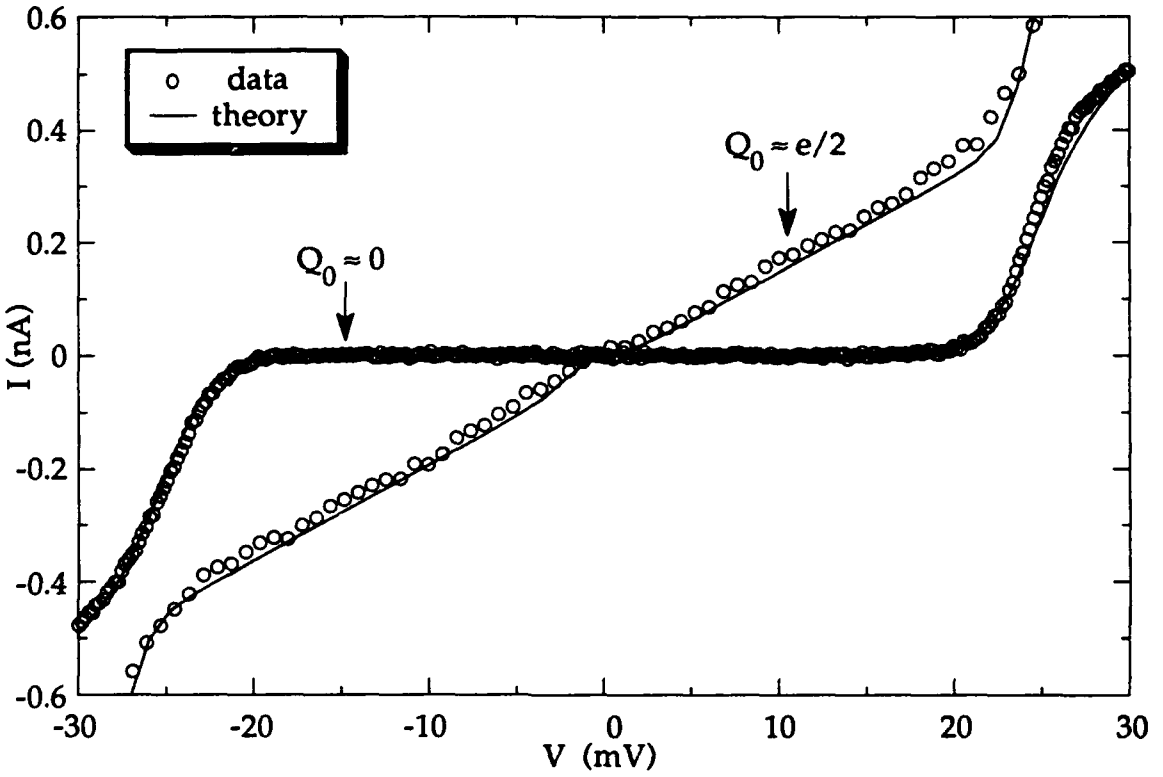


Figure 4.10, Single electron transistor effect. As described in Eqn. (4.22), by changing C_2, Q_0 is varied because of a non-zero contact potential. C_2 is varied by changing the distance between the STM tip and sample.

predicted above. It is the change in Q_0 which gives the different low voltage conduction properties shown in Fig. 4.10.

By continuously retracting the tip (increasing resistance), we can vary C_2 sufficiently to change Q_0 through multiples of e. This should cause oscillation in the width of the zero conduction region. Figure 4.11 shows the observation of this effect (data is shown in Table 4.1). The data shown in Figures 4.10 and 4.11 are, we believe, the first observations of the single electron transistor effect where Q_0 was varied without the use of a third electrode. This is also a confirmation of Eqn. (4.22), which was predicted by Averin and Likharev [1991]. Data similar to ours have recently been reported by other groups as well [Rong et al., preprint; Schönenberger et al., preprint].

Fitting our data by the orthodox theory as described earlier, we can extract Q_0 and $C_{1,2}$ of each of the I(V) traces. The plot of Q_0 as a function of tunneling resistance is shown in Fig. 4.12. It does not exhibit a regular cyclic behavior. This is expected because the contact potential for a gold sample and platinum tip is slightly greater than 0.5 volts [CRC, 1983], which means that a variation in capacitance of about 3×10^{-19} will change Q_0 by one electron. The capacitance in the table above, C_2, obtained from the orthodox theory fit ranges in value from 3×10^{-18} to 7×10^{-17}. Hence we expect that when moving the tip between each data point, Q_0 cycles through its entire range several times. (Due to the large capacitances listed, thermal effects were more noticeable in these I(V), and thus it was harder to differentiate the I(V) case, causing some uncertainty in our reported capacitance values). Our data in Fig. 4.12 may thus be interpreted as random points along a 'high frequency'

I(V)Trace Number	Resistance R_Σ (MΩ)	Capacitance C_Σ (aF)	Gap width (mV)	Q_0 (e)
189	4.0	5.9	27.3	-0.03
188	4.2	7.0	22.7	-0.04
187	4.4	8.5	18.9	-0.05
186	4.7	12.6	12.7	-0.09
185	5.0	23.5	6.8	-0.16
184	5.3	undetermined	0	0.50
183	5.7	undetermined	0	0.50
182	6.2	6.8	23.4	-0.06
181	6.7	9.7	16.54	-0.05
180	7.3	27.1	5.9	-0.13
179	8.0	16.8	9.5	-0.13
171	10.0	4.5	33	0.03
174	16.0	10.4	24.7	-0.02
170	20.0	28.6	42.5	-0.04
173	22.9	39.0	15.4	-0.11
172	26.7	72.7	4.1	-0.41
169	40.0	3.9	2.2	-0.03
176	53.3	3.8	35.8	0.01
177	80.0	6.5	5.6	-0.07
178	160.0	4.8	40.95	0.18

Table 4.1 Data on single electron transistor effect. The above table shows the data of twenty I(V) traces on the same gold particle, where the tunneling resistance between the tip and the particle varied. In each case, Q_0 was obtained by making an orthodox theory fit to the data. The data for each I(V) curve were taken only on the center section, so that in some cases not all junction parameters could be determined.

variation of Q_0 as a function of C_2. In that case, we would not expect to observe a clear cyclic variation of Q_0 in this measurement.

4-5 SUMMARY

In this chapter we presented simple analytic equations for I(V) based on the orthodox theory of tunneling. These equations were obtained in the limit of $R_2/R_1 \gg 1$, which is the limit appropriate for the regime of operation of the STM in this experiment. From the analytic equations we were able to solve for the location of distinct features in the I(V) trace, and from this we could obtain the parameters of the system. This framework showed how the I(V) traces could be distinguished in four distinct cases depending on the capacitance ratio C_2/C_1 and the fractional residual charge Q_0.

Our data, using an STM tunneling into small gold spheres, confirmed our analysis. A numerical solution of the full orthodox theory was in overwhelming agreement with our data. The variation of Q_0 was also theoretically investigated in the orthodox theory, and our data was shown to be in qualitative agreement with predictions.

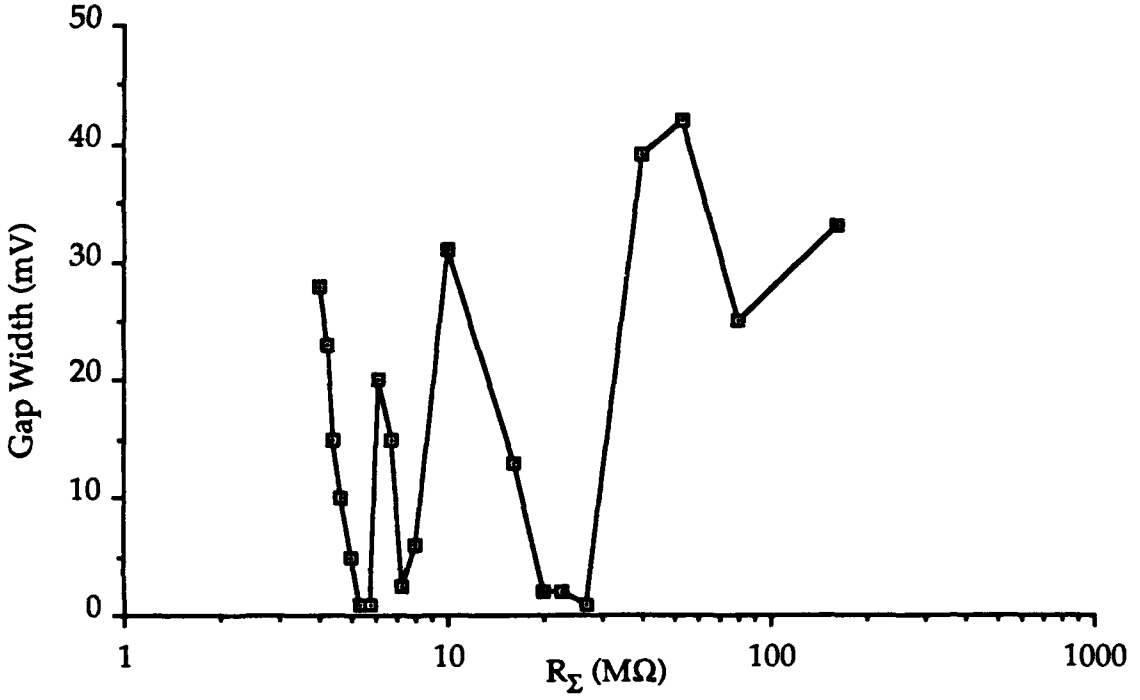


Figure 4.11, Coulomb blockade width oscillations. This plot shows the raw data: the width of the zero conduction region (Coulomb gap) versus the total tunneling resistance (R_Σ). The oscillatory behavior is due to Q_0 which originates with contact potential between the tip and s ample as shown in Eqn. (4.22). By continuously retracting the tip away from the sample (increasing resistance), we vary C_2 sufficiently so that Q_0 changes through multiples of e. This causes oscillations in the width of the zero conduction region, as shown above.

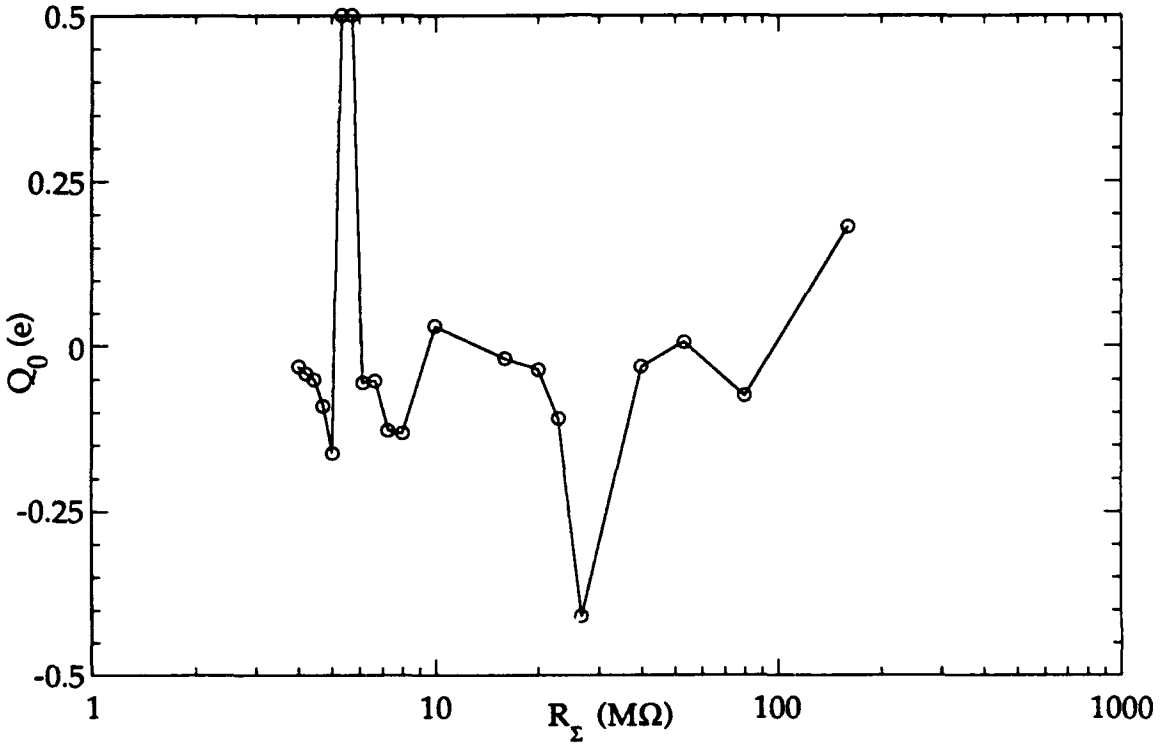


Figure 4.12, Q_0 oscillations I. This plot shows how Q_0, obtained by an orthodox theory fit of an I(V) trace, varies as the tip to sample tunneling resistance is increased. Although the tunneling resistance generally increases as the tip to sample distance is increased, when the tunneling distance involved is of the order of 1-10Å there could be tunneling resonances which decrease the resistance for certain distances. This would complicate observation of Q_0 oscillations on this plot, hence we extract C_2 by the same fit and plot $Q_0(C_2)$ in Figure 4.13.

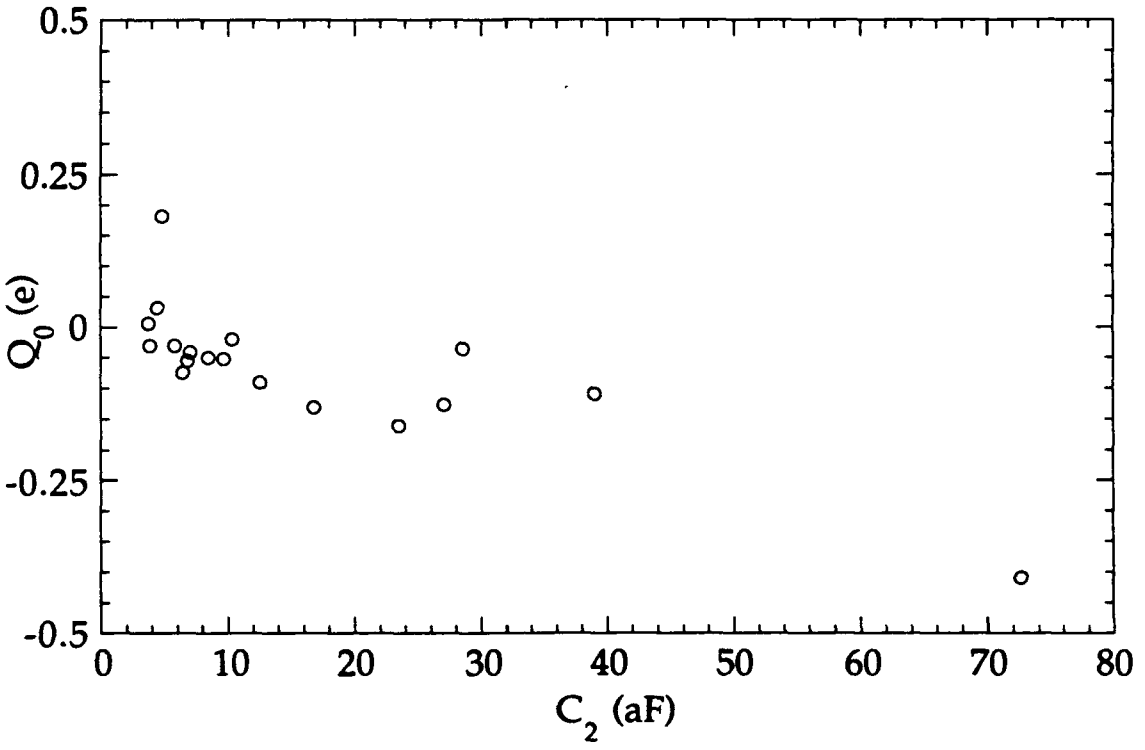


Figure 4.13, Q_0 oscillations II. This plot shows how Q_0 varies as the capacitance between tip and sample is increased. Because the contact potential in this system is on the order of a volt, the oscillation of Q_0 through one cycle (Q_0 changes by one electron) requires a change in capacitance of only about 3×10^{-19}F. Thus in moving from one data point to the next, Q_0 may cycle through several periods, and accordingly regular periodic oscillations are not expected to be observed.

CHAPTER FIVE

MACROSCOPIC QUANTUM TUNNELING OF CHARGE

5-1 INTRODUCTION

In this chapter the tunneling characteristics of a two junction system created with extremely small metal particles is studied in the voltage region within the Coulomb blockade. As explained in chapter two, recent theoretical works have predicted that q-MQT through the Coulomb energy barrier should be possible [Averin and Odintsov, 1989; Averin and Korotkov 1990]. For a two junction system this occurs via an intermediate state where an electron virtually occupies the center electrode. In our data, we observe a linear I(V) with conductance orders of magnitude lower than outside the blockade, as predicted for elastic macroscopic quantum tunneling of the charge variable Q.

There are two channels for q-MQT: (i) an *elastic* channel where one electron tunnels across the first barrier, virtually occupies the center electrode, then tunnels across the second barrier, and (ii) an *inelastic* process where in a sense a different electron tunnels across the first and second barriers, leaving an electron-hole excitation. Both cases yield a small, but finite, current for voltages below the Coulomb threshold however their I(V) characteristics are different. Observation of inelastic q-MQT has been reported by Geerligs et al., [1990], where in a two junction system they measure tunneling currents which scale as V^3 and as the product of the junctions' conductances. Their data are in reasonable agreement with the theory of inelastic q-MQT. Elastic q-MQT, on the other hand, is predicted to give a current which is nearly

linear in V, a signature which is more distinct from non-linear thermal effects. In this chapter we present the first data showing elastic q-MQT [Hanna et al., 1992], thereby extending the verification of q-MQT to the elastic case.

5-2 Q-MQT Analytic Theory

As detailed in chapter two, the current due to q-MQT has been calculated by Averin and Nazarov, [1990]. The inelastic contribution,

$$I_{in} = \frac{\hbar G_1 G_2}{12\pi e^2}\left(\frac{1}{E_1}+\frac{1}{E_2}\right)^2\left[(eV)^2+(2\pi k_B T)^2\right]V \qquad (5.1)$$

is calculated by neglecting the coherence between overlapping wave functions of electrons tunneling in different junctions. It thus only depends on the *product of the absolute value of the individual tunneling amplitudes,* as expected if two electrons tunnel independently in the two junctions. As before, in (5.1), G_1 and G_2 are the individual tunnel conductances, and E_1 and E_2 are the system energy increase upon tunneling one electron across junctions left and right, respectively. They are given in Eqn. (2.34), where again C_1 and C_2 are the capacitances of the individual junctions, and $C_\Sigma = C_1 + C_2$. In our setup, Q_0 is the residual fractional electron charge on the center electrode due to contact potential differences between the two metals used to form the junctions [Hanna and Tinkham, 1991; Rong et al.; Wilkins et al., 1989]

For the elastic contribution one must not only consider the coherence of the wave function of the electrons tunneling in different junctions but also the process of electron propagation inside the central electrode. In the regime

appropriate for our samples, Averin and Nazarov [1990] showed that the current is expected to be nearly linear in V and can be given by:

$$I_{el} = \frac{\hbar G_1 G_2 \Delta}{4\pi e^2}\left(\frac{1}{E_1}+\frac{1}{E_2}\right)V \qquad (5.2)$$

where Δ^{-1} is the density of states on the center electrode, and E_1 and E_2 are as above. Details for Eqns. (5.1) and (5.2) can be found in chapter two and the references.

Eq. (5.2) is a limit of the general equation for the case where: the diameter of the central electrode is less than the electron mean free path in the bulk ($D \ll l$), there is diffuse surface scattering, and $E_C < v_F\hbar/D$, where $E_C^{-1} = E_1^{-1} + E_2^{-1}$, and v_F is the Fermi velocity. In the regime where $eV \leq \min(E_1/2, E_2/2)$, I(V) is very nearly linear.

A simple way to delineate the essential difference between elastic and inelastic q-MQT, and the resulting different voltage dependences, is shown in Fig. 1. In the inelastic case, Fig. 1a, we require $E_L+E_{M_1}+E_{M_2}+E_R = eV$, to conserve energy. The E_i refer to the magnitude of the excitation energy of the electron or hole relative to the chemical potential in the corresponding electrode. In the calculation of I_{in} one would include all the E_i which satisfy the above energy condition, as shown schematically in (5.3a).

$$I_{in} \propto \int dE_L dE_{M_1} dE_{M_2} dE_R \;\; \delta(E_L + E_{M_1} + E_{M_2} + E_R - eV) \propto V^3 \qquad (5.3a)$$

In the elastic case $E_{M_1} = 0$ and $E_{M_2} = 0$, so that the energy conservation equation becomes $E_L+E_R = eV$, as illustrated in Fig. 5.1b. This condition further restricts the counting of the paths available for I_{el}, as given by (5.3b).

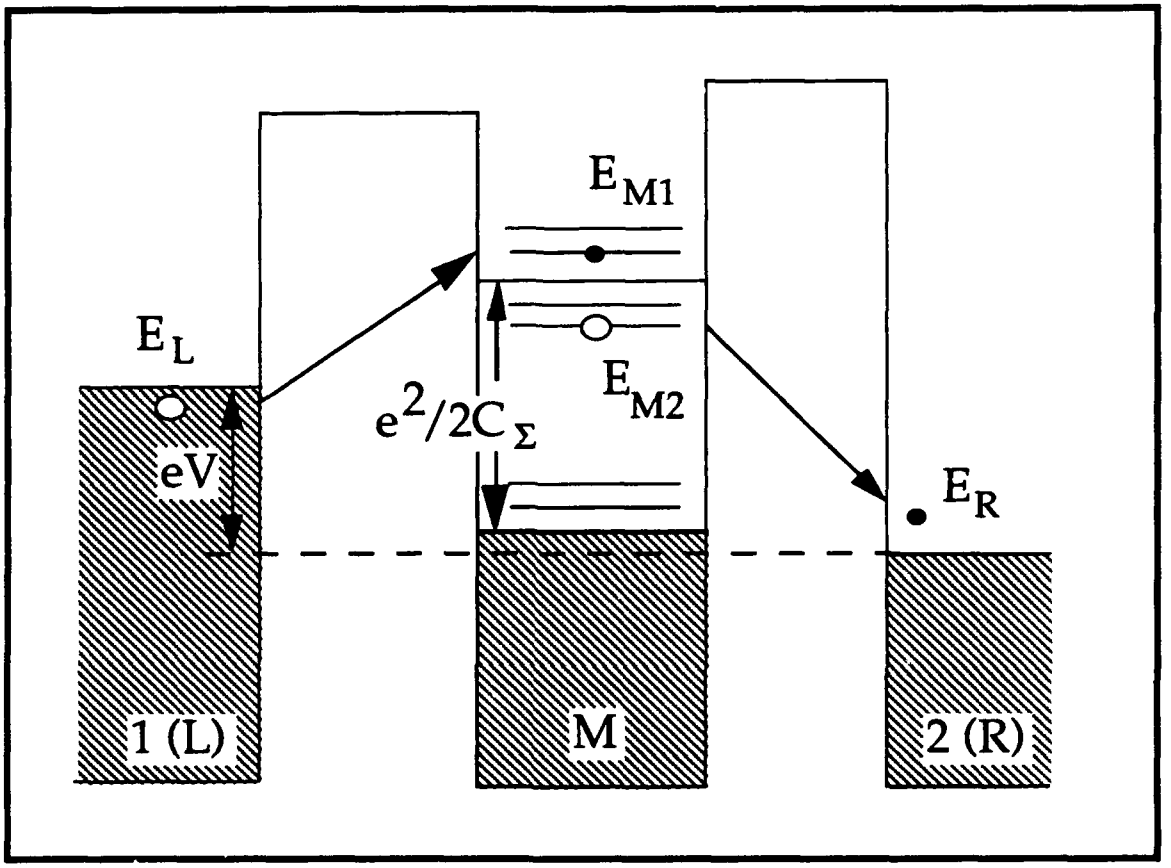


Figure 5.1a, Inelastic q-MQT channel. The diagrams show the lowest order inelastic channel for MQT in a two junction system. Inelastic MQT leaves a characteristic electron-hole excitation on the center electrode and gives an $I(V) \propto V^3$.

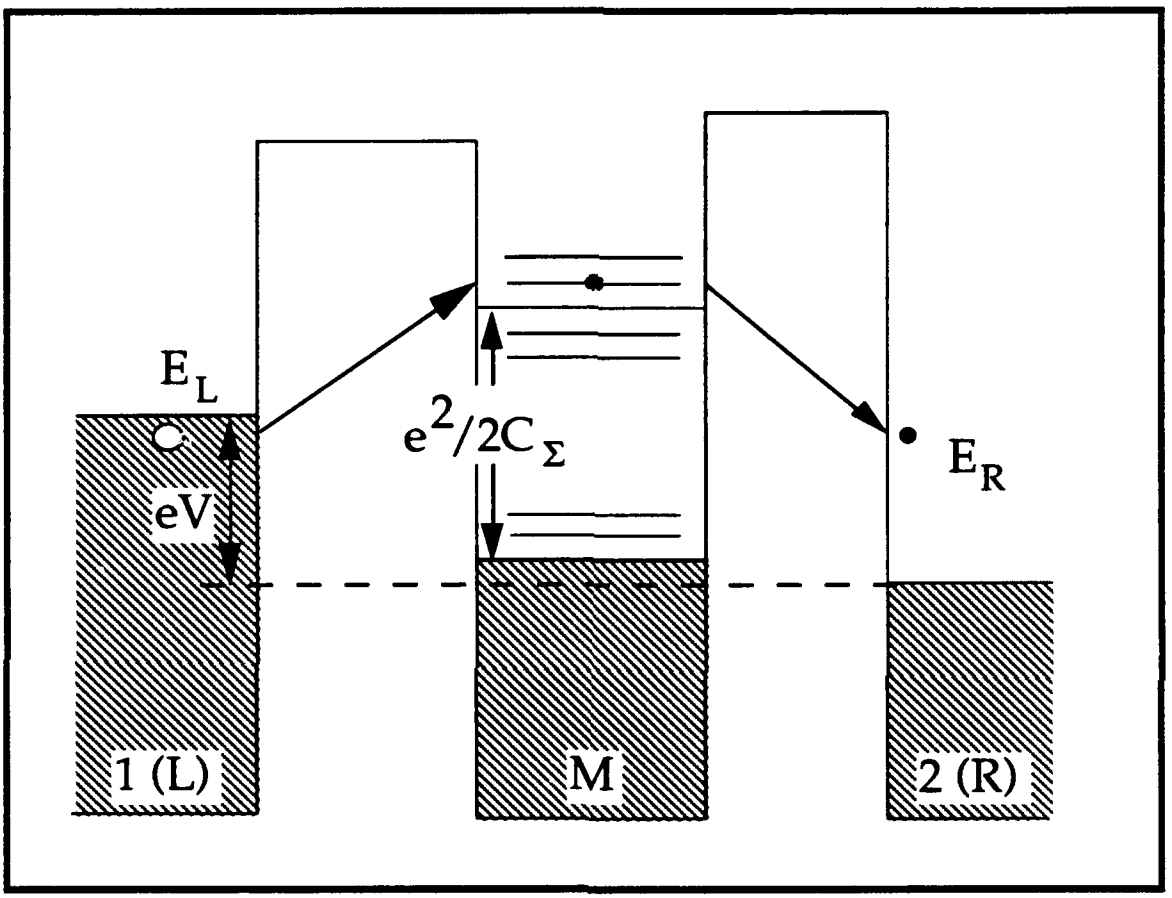


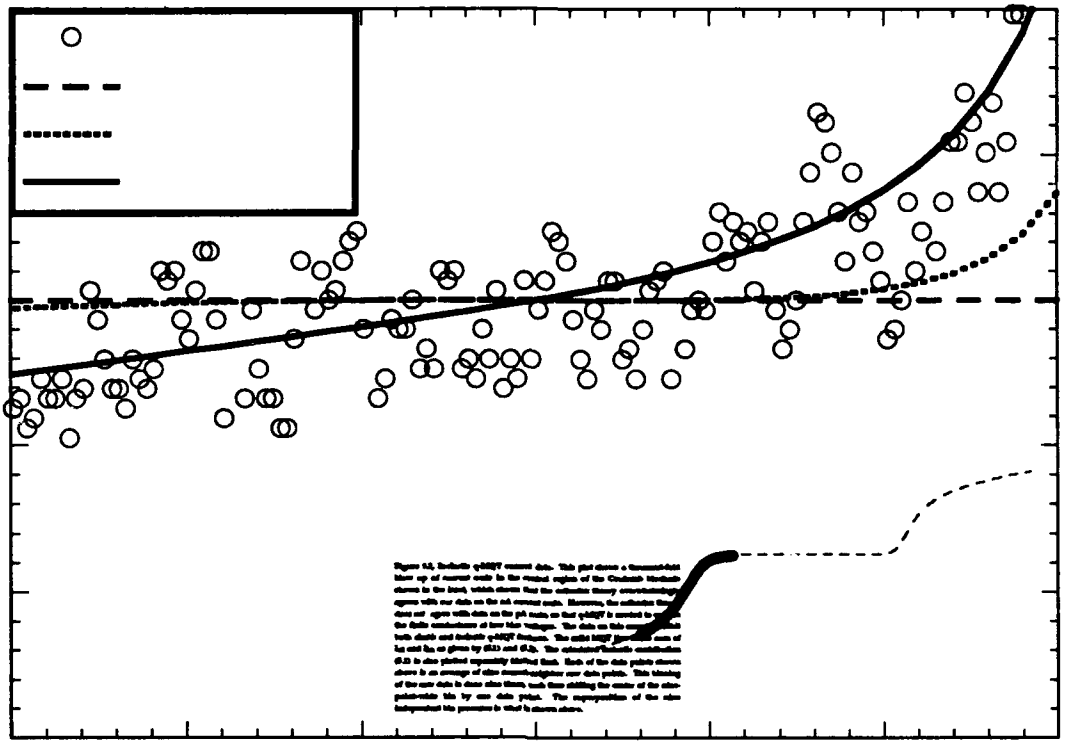
Figure 5.1b, Elastic q-MQT channel. This diagram illustrates the lowest order elastic channel of MQT in a two junction system. This process leaves no electron-hole excitations on the center electrode and has $I(V) \propto V$.

$$I_{el} \propto \int dE_L dE_R \ \delta(E_L + E_R - eV) \propto V \qquad (5.3b)$$

The above simple analysis gives the correct voltage dependences, and the prefactors in (5.1) and (5.2) can be seen to originate from a basic Golden Rule calculation. Clearly, at sufficiently low V, the linear elastic term should dominate over the V^3 term.

5-3 Q-MQT DATA

Our two junction system was created by using a scanning tunneling microscope (STM) tunneling into an individual small gold particle isolated from a Nb film plane by an insulating oxide layer. The data were taken at 4.2K. We confirmed that we had a two-junction system by looking for a Coulomb blockade and staircase in the I(V) trace. As we reported in chapter four [Hanna and Tinkham, 1991], typical traces with this system show clean staircase structures, which are fitted with extraordinary precision by the orthodox theory of tunneling. (Similar, but generally less clean, data have also been reported by others [van Bentum et al., 1988; McGreer et al., 1989; Wilkins et al., 1990]) To search for evidence of elastic q-MQT, the data reported here were taken on particles selected to be very small. By selecting particles with diameters as small as ~10Å, we obtained charging energies $e^2/2C$ (C is the capacitance of the particle) as large as half a volt, as shown by Coulomb blockades of this width. The size of the particles is estimated from the implied capacitance; they are consistent with STM images of the particles on the surface of the underlying film. We expect such particles to satisfy the conditions for (5.2) because: (i) the diameter of the particle D ~ 10-30Å is much



less than the bulk mean free path ℓ, which is estimated [Kadereit, 1967; Larson, 1971] to be ~ 1 mm; (ii) diffuse surface scattering is expected since λ_F is of the order of the surface roughness ~2Å; (iii) $E_C < v_F h/D$, as shown in Table 5.1. The importance of using very small particles is shown by the fact that to obtain $I_{el} > I_{in}$ according to (5.1) and (5.2), the inequality

$$[(eV)^2 + (2\pi k_B T)^2] < 3E_C D \tag{5.4}$$

must be satisfied, where the level spacing D is inversely proportional to particle volume.

To get better statistical accuracy in search of a small q-MQT current, we took more dense data within the Coulomb blockade region, where the current should nominally be zero. Such data are shown in Figs. 2 and 3. As noted earlier [Hanna and Tinkham, 1991] and confirmed in these figures, the orthodox theory shows an extremely convincing fit on a large current scale as shown in the insets of Fig. 5.2 and 5.3; however it can not account for the *finite conductance near zero bias within the Coulomb blockade,* as shown in Figs. 5.2 and 5.3.

The solid lines in Figs. 5.2 and 5.3, show the q-MQT calculation of the tunneling current. Using the junction parameters, C_i, G_i, and Q_0 obtained from the orthodox theory fit, I_{in} is calculated using (5.1), and plotted as the dotted curves in Figs 5.2 and 5.3. The remaining current is then ascribed to I_{el} and from fitting it to (5.2), we extract the parameter Δ_{ex}. The noise in the data is random scattering from making measurements at the sub-picoamp level, and does not appear to reflect any fine structure.

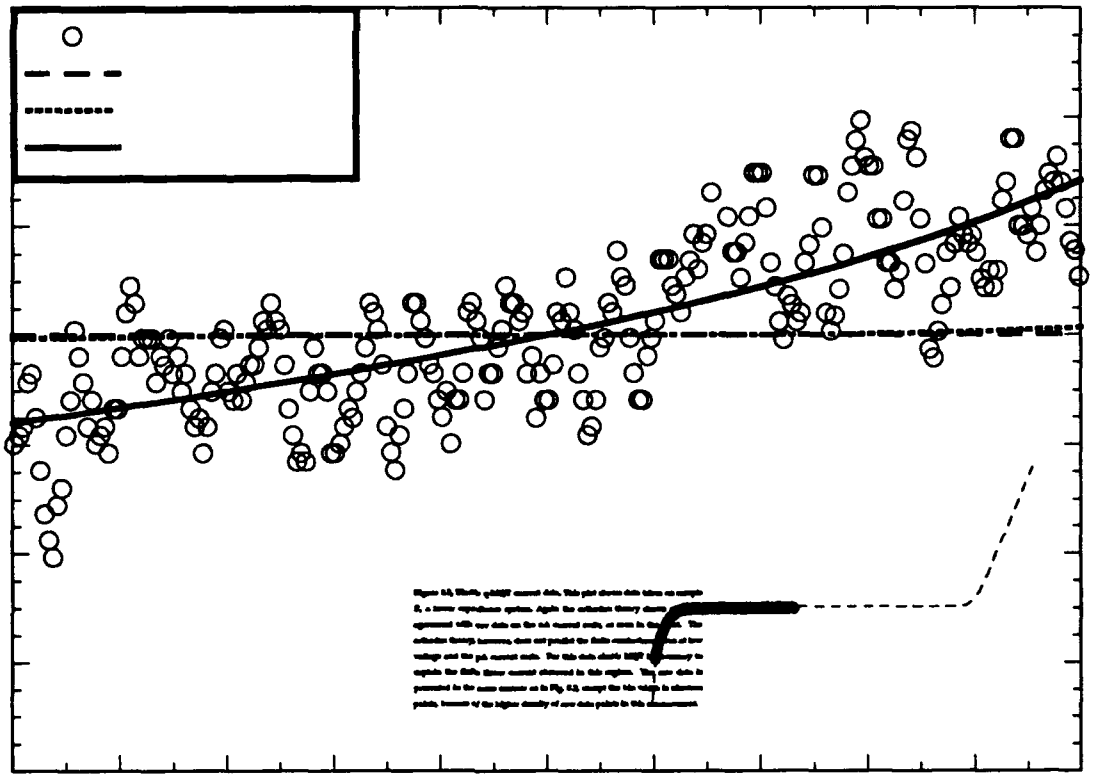
To determine the significance of the trends in our data, a least square fit of the function $a + bV + cV^3$ to the central section of the data was performed to determine the coefficient b of the linear term and c of the cubic term, as well as the uncertainty in these coefficients. For the data shown in Fig 5.2, the least square fit has a linear coefficient of 0.02 ± 0.007 pA/mV and the cubic coefficient is $(7 \pm 3) \times 10^{-5}$ pA/mV^3. For the data in Fig. 3, the coefficient of the linear term is 0.7 ± 0.1 pA/V and the cubic coefficient is 0.5 ± 0.2 pA/V^3. (A fit to the function $a + bV + cV^2 + dV^3$ did not significantly change the value of the coefficients obtained above for the V and V^3 terms)[1]. Thus, in both sets of data the value of the linear coefficient is several standard deviations away from zero, strongly indicating the presence of a linear I(V) within the Coulomb blockade region.

The data also show evidence of the contribution from *inelastic* q-MQT (which predicts a small non-linear current and does not involve the poorly known level separation D), in addition to the *elastic* q-MQT which is the novel feature reported here. In this case the majority of the current within the Coulomb gap is expected from *elastic* q-MQT, as predicted by (5.1) and (5.2), and this is confirmed in the data shown in Fig. 5.3.

Since our high resolution data show only the center section of a Coulomb staircase, the orthodox theory fit cannot determine *all* the junction

---

[1]The coefficient of the V^2 term had a 40% uncertainty for the data in Fig. 2 and a 100% uncertainty for the data in Fig. 5.3. In the absence of any physical model suggesting a V^2 dependence in the tunneling current at this low voltage level, we ignore these ill-defined V^2 terms, and fit to the *actual form predicted* by (5.3), in which higher than linear V dependences in the elastic term arise from the voltage dependence of E_L and E_R.



parameters exactly, because that would require more features (steps) to be accurately fitted [Hanna and Tinkham, 1991]. However, from the data we do observe, we can determine C_1, C_2, Q_0, and R_Σ exactly and then put limits on the others such as R_1 and R_2. It is important to note that there was no way to obtain a finite linear current near zero bias from the orthodox theory within our determined parameter ranges. Thus, some other mechanism is necessary to explain the linear I(V) at small bias voltages.

Thermally excited quasiparticles give some conduction at small bias voltages. However, in our system, where $e^2/C \gg 100k_BT$, these effects, which fall exponentially, are orders of magnitude too small to account for the observed tunnel current. In principle, a linear current could arise from a leakage resistance in our experimental measuring system. However, because we obtained a different conductance in different samples, we believe such an *instrumental* effect can be excluded.

A compelling confirmation that we observe q-MQT effects and not some sort of trivial parasitic linear conduction path is that our data show the *asymmetry* between positive and negative voltage which is predicted by the (independently determined) non-zero Q_0. This asymmetric I(V) contrasts with the small current that might arise by tunneling from the tip to the base electrode through the vacuum, around the gold particle, and through the oxide to the Nb film. The identifying characteristics for this latter sort of tunneling are that the I(V) is: (i) *linear* in the small voltage region of our measurement (no barrier bending), and (ii) *symmetric* in voltage. These characteristics were experimentally confirmed by tunneling directly into the oxide barrier. Our data in Figs. 5.2 and 5.3 show both asymmetry and non-

SAMPLE	D (Å)	C_L (aF)	C_R (aF)	R_L (MΩ)	R_R (MΩ)	Q_0 (e)
A	8	0.095	0.066	0.42	630	0.055
B	30	0.57	3.4	3.5	21	-0.070

SAMPLE	E_L (eV)	E_R (eV)	E_C (eV)	Δ_{EX} (eV)	Δ_{TH} (eV)	$\hbar v_F/D$ (eV)
A	.5	0.4	0.2	2	0.2	1
B	.02	.02	0.01	0.04	0.005	0.3

Table 5.1, Sample parameters. The parameters C_i, R_i and Q_0 are obtained from the orthodox theory fit of the data. D is estimated by assuming that total capacitance is that of a sphere embedded in an insulator with dielectric constant 2. The values for E_i are obtained from (5.2) at V=0, and Δ_{ex} is the fitting parameter used for I_{el}. $\Delta_{th} = 4E_F/\pi nD^3$ is the calculated level spacing for free electrons in a sphere of diameter D. It is important to note that the values for Δ_{ex} are *upper bounds* because decreasing R_L will not greatly affect the orthodox theory fit in the region where we have data; however it will significantly change Δ_{ex} since that depends on the product of R_L and R_R. The values shown above are the values used to calculate the orthodox theory curves shown in Figures 5.2 and 5.3.

linearity, and thus cannot be explained by a leakage current around the gold particle. The data presented here are representative of several ultra-small capacitance junctions where we have observed a tunneling current with a linear component for voltages below the Coulomb blockade.

5-4 SUMMARY

In this chapter we have presented I(V) data taken with a two junction STM apparatus which shows a conductance within the Coulomb blockade which is $\approx$ 1000 times smaller than the conductance outside the blockade. This non-zero current is contrary to orthodox theory predictions. After discounting other possible mechanisms, this current was accounted for by q-MQT theory. Both channels of q-MQT were observed, inelastic and elastic.

In particular a definite *linear* conductance was observed in the low-voltage region within the Coulomb blockade. This could only be attributed to *elastic* q-MQT, complementing the earlier observations of inelastic q-MQT [Geerligs et al., 1990].

CHAPTER SIX

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

Since predictions of single electron charging effects in small junctions first received notice, experimenters have attempted to verify the predictions by several methods. Usually this required mK temperatures because fabricated junctions had capacitances which were quite large. With the development of the STM probing structures of much smaller size became easily realizable. In this research we built and used a low temperature STM (4.2K) to investigate single electron charging of small (10-100Å) particles.

Our data reports high-quality measurements which showed sharp features in the behavior of two junction systems. The features in the data displayed overwhelming agreement with the predictions of orthodox theory, which correctly produces the location of linear voltage onsets, current steps and step slopes. The high resolution of our data also allowed the categorizing of the I(V) curves according to the onset of certain features.

The orthodox theory, combined with our data have also shed understanding on the nature of Q_0, the fractional charge on the middle electrode. The data have also shown the higher order tunneling processes associated with macroscopic quantum tunneling of charge, q-MQT. This observation is not only of fundamental physical interest, but is an effect which could set the limit of future devices operating on single electron transfer.

The low temperature STM, with some minor changes, has the capability to perform some very sensitive I(V) measurements, and confirm some recent ideas about elastic q-MQT theory [Glazman, private communication]. In particular, finer details of elastic q-MQT predict that there should be a certain amount of scatter in the I(V) data. This is because the tunneling electron can virtually occupy any one of several energy eigenstates on the particle. The expected I(V) for an electron virtually occupying one of energy eigenstates is I_{el}, as was calculated in chapter two, which is approximately linear in voltage. The distinction of having many states which could be virtually occupied, is that each state still gives a linear contribution, however, the conductance (slope of I(V)) is different for each state. Thus, the total elastic current will be the sum (or average) of all these linear I(V) with different slopes, which would appear in the data as scatter. The form of the scatter should be theoretically predictable, and a high resolution measurement of elastic q-MQT may be able to observe this phenomenon.

Another area of possible experimental interest is photon-assisted tunneling. The Tinkham group STM has built in design capabilities to couple the sample under investigation to microwaves. In preliminary measurements, 20GHz microwave coupling to a two junction system just showed the classical effect of being rectified by the asymmetric I(V) curve. With minor modifications to the system, infra-red radiation could be coupled to the sample. In this energy regime, we would expect to see more quantum effects, particularly photon-assisted tunneling, where absorbed photon energy is used by electrons to overcome the Coulomb barrier.

There are also possible experiments which combine lithographically patterned junctions with the STM, using the STM tip as one of the electrodes. For example, quasiparticle injection in a superconducting two junction system, could offer some interesting physics. Less charge related ideas include the investigation of multi-layer N-S samples, to explore the spatial variation of the superconducting energy gap.

In conclusion, the addition of a low temperature STM to the Tinkham group laboratory has opened up whole new areas of possible research and investigation. Taking images, as well as creating and measuring properties of a system on an angstrom scale, is a capability which can be extremely useful. I believe that this STM can continue to be a productive experimental tool in future years.

REFERENCES

Amman, M., R. Wilkins, E. Ben-Jacob, P.D. Maker and R.C. Jaklevic, Phys. Rev. B **43**, 1146 (1991).

Anderson, P. W., Lectures in Many-Body Problems, E. R. Caianiello, ed. (Academic Press, New York, 1964), Vol. 2, p. 113.

Averin, D. V., and A. A. Odintsov, Phys. Lett. A **140**, 251 (1989).

Averin, D. V., and A. N. Korotkov, J. of Low Temp. Phys. **80**, 173 (1990).

Averin, D. V., and K. K. Likharev, J. of Low Temp. Phys. 62, 345 (1986).

Averin, D. V., and K. K Likharev, *Mesoscopic Phenomena in Solids*, Eds. B.L. Altshuler, P.A. Lee and R.A. Webb, (Elsevier Science Publishers B.V., Amsterdam, 1991), p. 169.

Averin, D. V., and Yu. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990).

Baratoff, A., Physica B 127, 143 (1984); Europhysics Conference Abstracts, 3rd General Conference of the Condensed Matter Division of the EPS, 28-30 March 1983, P25-103.1.

Binnig, G., H. Rohrer, Ch. Gerber, and E. Weibel, Appl. Phys. Lett. **40**, 178 (1982); Phys. Rev. Lett. 49, 57 (1982); Physica 109/110b, 2075 (1982).

Chen Wang, B. Giambattista, C.G. Slough, R.V. Coleman and M.A. Subramanian, Phys. Rev.

CRC Handbook of Chemistry and Physics, 63rd edition, CRC Press (1985).

D. C. Larson, Physics of Thin Films v. 6, edited by M.H. Francombe and R. W. Hoffman (Academic Press, New York, 1971), p. 81.

Demuth, J. E., U. Koehler, and R. J. Hamers, J. Microsc. **151**, 299 (1988).

Geerligs, L. J., D. V. Averin, and J. E. Mooij, Phys. Rev. Lett. **65**, 3037 (1990).

Hanna, A. E., and M. Tinkham, Phys. Rev B **44**, 5919 (1991).

Hanna, A., M. T. Tuominen, and M. Tinkham, Phys. Rev. Lett. **68**, 3228 (1992).

Kadereit, H. G., Thin Solid Films 1, 109 (1967).

Lambe, J., and R. C. Jaklevic, Phy. Rev. **165**, 821 (1968)

Likharev, K. K., and A. B. Zorin, J. Low Temp. Phys. **59**, 347 (1985).

Likharev, K. K., IBM J. Res. Dev. **32**, 144 (1988).

McGreer, K.A., J-C. Wan, N. Anand and A.M. Goldman, Phys. Rev. B **39**, 12260 (1989). The asymmetric trace (d) in figure 3 shows linear conductance on the negative bias side of the central region. This is characteristic of case I.

Ohnishi, S., and M. Tsukuda, Solid State Commun. **71**, 391 (1989).

Rong, Z.Y., A. Chang, L.G. Cohen, and E. L. Wolf, IEEE Trans. Mag., preprint.

Schönenberger, C., H. van Houten, and H. C. Donkersloot, preprint.

Simmons, J.G., J. Appl. Phys. **34**, 1793 (1963).

Smith, Walter F., Technical Report No. 28 (Tinkham series, Div. of Appl. Sci., Harvard University, 1989).

van Bentum, P. J. M., H. van Kempen, L.E.C. van de Leemput and P.A.A. Teunissen, Phys.

van Bentum, P. J. M., R. T. M. Smokers and H. van Kempen, Phys. Rev. Lett. **60**, 2543 (1988). Trace (a) in figure 4, is asymmetric due to a linear conductance on the positive bias side of the central region. This is characteristic of case IV.

van Bentum, P. J. M., R. T. M. Smokers, and H. van Kempen, Phys. Rev. Lett. **60**, 2543 (1988).

Vieira, S., J. G. Rodrigo, M.A. Ramos, K.V. Rao and Y. Makino, Phys. Rev. B **40**, 11403 (1989).

Wang, Chen, B. Gambattista, C. G. Slough, R. V. Coleman, and M. A. Subramanian, Phys. Rev. B **42**, 8890 (1990).

Washburn, S., and R. A. Webb, Advances in Physics **35**, 375 (1986).

Wilkins, R., E. Ben-Jacob, and R. C. Jaklevic, Phys. Rev. Lett. **63**, 801 (1989).

Wilkins, R., M. Amman, E. Ben-Jacob and R.C. Jaklevic, Phys. Rev. B 42, 8698 (1990). Trace D in figure 2, is asymmetric due to a linear conductance on the positive bias side of the central region. This is characteristic of case IV.

Wintterlin, J., J. Wiechers, H. Burne, T. Gritsch, Höfer, and R. J. Behm, Phys Rev. Lett. **62**, 59 (1989).

Young, R. D., J. Ward, and F. Scire, Phys. Rev. Lett. **27**, 922 (1971).

Young, R., J. Ward, and F. Scire, Rev. Sci. Instrum. **43**, 999 (1972).

ACKNOWLEDGEMENTS

Many people have made significant contributions to this work both directly and indirectly. Foremost is my advisor, Professor Michael Tinkham. He has been an outstanding professor for which to work. His insight into our measurements, coupled with the ability to produce clear phenomenalogical explanations, has been a strength on which our whole group relies. I am particularly grateful for his support for my diverse interests outside of physics. Most importantly, his sense of humor, kindness, and frequent dinner parties, have made the last four years not only educational, but very enjoyable.

My family, Galil, Cecile, and David, has been understanding and supportive throughout my life, but never more so than the last several years. This work is dedicated to them. My father's patience and gentle methods of guidance were essential during my graduate school education. David, my brother, has brought a new joy into my life and was a great distraction from school. To my mother I am especially thankful. This research is certainly a direct result of her encouragement for me to pursue my individual interests. However, I am most grateful to her for raising me within a strong family, community and faith, and for always reminding me that these are the most important achievements in life.

In our group, I would like to thank Marco Iansiti and Professor Chris Lobb for introducing me to the Tinkham group, and initiating my involvement on the STM project. I would like to thank Charlie Johnson for explained to me charging effects and helped me prepare for the qualifying exams. My good friend Tom Tighe was consistently a source of inspiration

and sound judgment, both in the laboratory and in personal matters. I always enjoyed our lunch time discussions as we strolled through Harvard square.

I would also like to thank Lydia Sohn, who always perked up our group with her personality and her outstanding culinary creations. Delft University doesn't realize the full capabilities of the post-doc they have hired. I have received much guidance from Dr. Mark Tuominen and Dr. Mark Rzchowski. They have always been available for questions and assistance, especially when things were not going well. Rich Fitzgerald has been a source of apparently unlimited knowledge, a constructive critic, and an entertaining tennis partner. Niraj Anand was always pleasant company during the late-night experimental runs. Dave Carter, who has transformed sample creation from an art into a science, has been a great asset for the lab. We have all benefitted from his presence, both professionally and personally. And Jack Hergenrother, the lab's rising star, has taken the time to carefully think about and resolve many particulars of charging effects. His laboratory methods will provide much success in his research. I would also like to thank my friend Hans Guido, with whom I shared the late-night distress of numerous homework problems.

There are many people at Harvard who turn research ideas into reality and I have become indebted to many of them during my stay. Louie Defoe is the extraordinary machinist responsible for actually creating the STM. I am obligated to him for being able to engineer my hazy drawings into real products. Armand Dione is responsible for all the photographs in this thesis as well the negatives required to create the etched electronics boards. Elizabeth Heminway who maintains order in our building, provides us with some very needed relaxation during her teas. And, I would like to thank

Ralph Generazzo, who saved us much time by handling our purchases and dealer relationships.

Many families and friends have been very influential and encouraging throughout my life. I am very fortunate to know them. In particular, Father Roufial Michael has been a crucial influence in my life. He was always available for discussions and assistance at any time of day, and continues to fill that role for me today. His strength has been a pillar on which I have often rested. I would also like to thank his entire family who has provided me with kind care as long as I remember.

Also, in Michigan, I would like to thank the Abd el Messiah, Assad, Attia, Elias, and Isaac families, who made a large positive impression on my younger years. My most fond memories are of our community gatherings and picnics. Their love and generosity is a gift for which I will always be thankful. I would also like to thank Mrs. Wafia Matta, a teacher and a guide in academic, personal and spiritual matters.

The Bastawros, Menias, and Mowad were some of the families whose kindness supported me in Chicago. They made my move from home to college much less traumatic. Amir Bastawros and Alex Borione were the finest roommates and friends I could have met in college. I will always be thankful for the wonderful times and experiences we shared.

In Boston, I am exceedingly grateful to the Hanna, Labib, Sidhom, and Zaki families. They have treated me as a family member and have filled my weekends and holidays with warmth. Also, the Andrawis, Bishay, Elshama, Kaddis, and Makar are among the many families have made me feel at home in Boston. I am quite honored to be considered an EPB member.

I would finally like to express my appreciation to the National Science Foundation which has supported my research through grants DMR89-12927 and DMR89-20490, and to the Office of Naval Research for grants N00014-89-J-1565 and N00014-89-J-1023.