Complex oxides studied by Scanning Tunneling Microscopy / Spectroscopy

Kelly, S.J.

Citation

Kelly, S. J. (2012, April 3). *Complex oxides studied by Scanning Tunneling Microscopy / Spectroscopy*. *Casimir PhD Series*. Retrieved from https://hdl.handle.net/1887/18667

Author: Kelly, Simon J.
Title: Complex oxides studied by scanning tunneling microscopy/spectroscopy
Issue Date: 2012-04-03

Chapter 2

Scanning Tunneling Microscopy/Spectroscopy

2.1 Introduction

Scanning Tunneling Microscopy (STM) was the principal tool used for gathering data in this Thesis. As such, a basic description of STM and an introduction to the basic function, basic components, and underlying theory will demonstrate both the advantages and disadvantages of this tool.

STM was used to measure both topography and spectroscopy. Ideally, STM topography maps the position of surface atom cores, and scanning tunneling spectroscopy (STS) measures the local density of states (LDOS) of the sample. As measuring the LDOS is the focus of this research, STS was of primary interest. But because surface topography and the sample LDOS can be strongly correlated, STS was only performed in conjunction with STM topographic measurements.

Spectroscopy, though, is not without difficulties. In its favor, and the most important reason STS is used to measure LDOS, is that STS affords very fine spatial resolution, at the level of single molecules[2] or even single atoms.[3,4] The primary difficulty with STS is that tunneling spectra are always a convolution of sample DOS, tip DOS, and the highly non-linear tunneling barrier. Separating the sample DOS from tunneling spectra is non-trivial and more will be said about this later. Compounding this, as tunneling is limited to within angstrom of the sample surface, tunneling spectra can be dominated by surface electronic structure. If a measurement of the surface electronic structure is desired, then STS is an ideal tool, but for bulk DOS some effort must be made to verify that what is measured is bulk and not surface DOS.

The rest of this Chapter is organized as follows. Immediately following this introduction, some specific details of the STM's used in this study are presented (Sec. 2.2). This is followed by a general introduction to the STM (Sec. 2.3), and then, in greater detail, a discussion of piezoelectric actuators (Sec. 2.4), and technical issues surrounding measuring the tunneling current (Sec. 2.5). This is followed by discussions of implementing STM topography (Sec. 2.6), STS (Sec. 2.7) and details of the theory behind STS (Sec. 2.8).

2.2 STM Instruments and Techniques Used in this Study

The measurements in this Thesis were made almost exclusively with STM's designed and built within the group. Other measurements were made with commercial STM's but these were used only to verify measurements already made using self-made instruments. These two STM's were a flow-cryostat STM and a UHV STM. The flow cryostat and UHV STM's are similar and are both based upon the Pan design.[5]

The UHV STM consists of a home made microscope and UHV chambers connected to a Heliox$^{\text{UHV}}$ ^{3}He refrigerator from Oxford Instruments. The ^{3}He refrigerator is a one-shot type[6] meaning that the ^{3}He must be recondensed after each cooling cycle, limiting the base temperature hold-time to about 25 hours under the STM heat-load. The STM is suspended by a spring from the refrigerator and both are fitted inside the UHV tube.

The UHV tube is surrounded by a cryostat containing about 70 liters of liquid helium used to cool the superconducting magnet (10 T, parallel to STM vertical axis, perpendicular to the sample surface) and for the ^{3}He refrigerator operation. The practical liquid helium hold time, without warming up the superconducting magnet, is about 60 hours.

The Heliox$^{\text{UHV}}$ refrigerator is especially developed for STMs in UHV, which has a number of consequences for its design: the refrigerator itself has been designed to be not extremely rigid, in order to isolate from external mechanical and acoustic noise. The UHV environment does not allows for use of ^{4}He exchange gas for first cooldown, therefore the ^{3}He inside the refrigerator is used for this purpose. Finally, to allow in situ tip/sample exchange, part of the refrigerator can be elongated with a vertical lead screw; for this reason the ^{3}He and ^{4}He capillaries are in large part coiled up to allow extension. When the refrigerator is in fully retracted position [as in Fig. 2.1(a)], the STM is freely suspended in the middle of the superconducting magnet (measurement position). When the refrigerator is in the fully elongated position (insert extends about 50 cm down into the UHV chamber) tip and/or sample can be reached and exchanged with a mechanical hand.

This low temperature STM is designed for best performance in a temperature range between 300 mK and 2.5 K. Nevertheless, by using a heater installed on the ^{3}He pot, the STM can also operate at temperatures up to 180 K, although cooling and warming cycles are very slow in this "high-temperature" mode. The temperature was measured with Cernox™ CX-1030 thermometer calibrated by the manufacturer, Lake Shore Cryotronics, Inc., and mounted on the STM close to the sample. Because the thermometer was not attached directly to the sample, the sample and the thermometer temperatures can differ. This difference can be estimated by measuring a superconducting gap and fitting this gap using a thermally-smeared BCS density of states, with the effective sample temperature typically about ~450 mK warmer

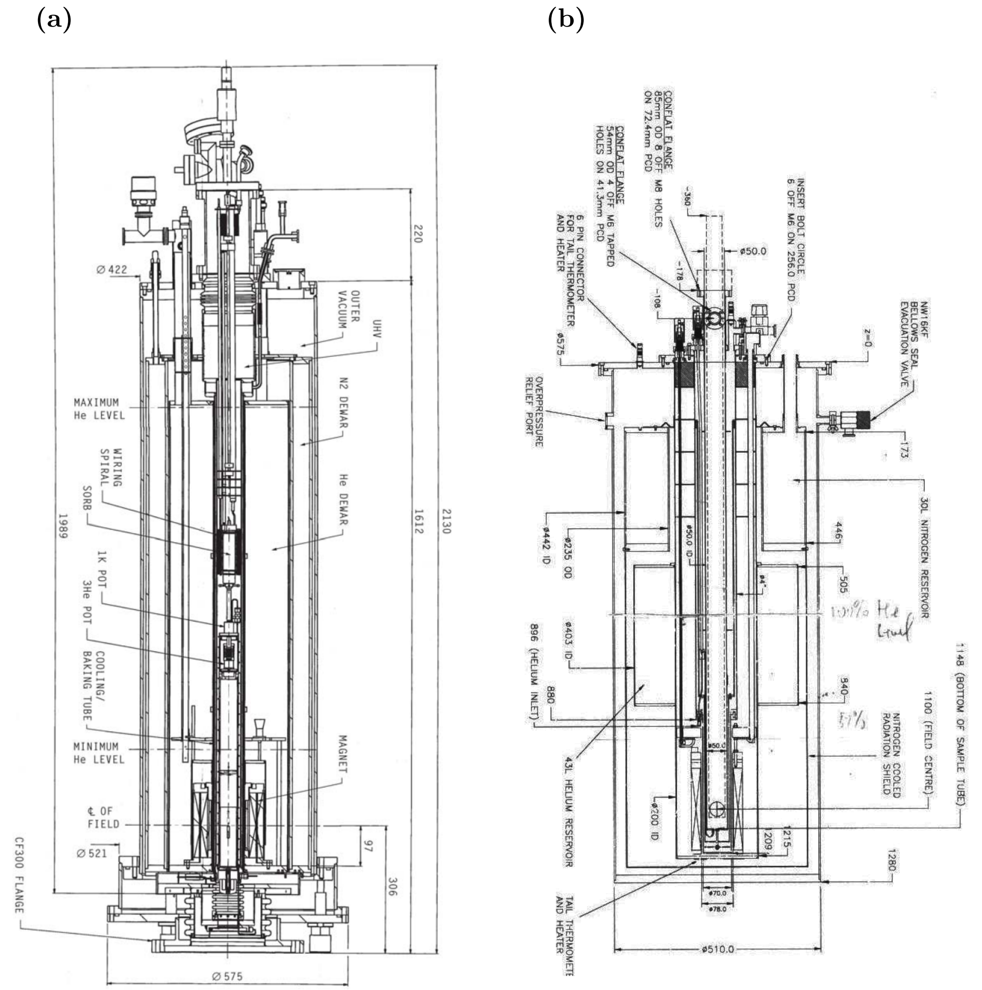

Figure 2.1: (a) Cross-section of UHV Cryostat. (b) Cross-section of VTI Cryostat. See text for further details.

than the thermometer reading at ∼1 K. An example fitted Nb superconducting gap is shown in Ch. 5. The bottom of the cryostat is connected to a Leiden-designed UHV chamber which is pumped by a 300 liters/s ion pump, a titanium sublimation pump, or both. The UHV tube is protected from thermal radiation emanating from the UHV chamber with a liquid nitrogen cooled shield, which opens when the STM is lowered into the UHV chamber. The UHV chamber contains tip/sample storage plus leak-valves and evaporation cells for tip/sample preparation.

Samples and tips are loaded into the main UHV chamber via a hybrid load-lock/preparation chamber. This chamber is evacuated with a 200 l/s turbo pump that is also used for initial, pre-bake system pump-down. For surface and tip cleaning and annealing the load-lock is equipped with a 0.2-5 kV ion-source and a heating stage.

The entire STM is suspended on air-dampers (Kinetic Systems Inc.), which rest on a heavy concrete island further isolated from the rest of the laboratory.

The STM head has been designed and built in Leiden by S. Otte.[7] The Z-approach and Z-scan mechanisms are integrated in one single piezo-motor based on the Pan design, and the XY motion is done with a XY-shear piezo stack to increase the mechanical resonance frequency and decrease the mechanical loop. The STM is connected to two stainless steel coaxial cables for tunneling current and bias ($R = 120\ \Omega$ and $C = 1.2$ nF each) and 5 non-shielded stainless steel thin wires for driving the piezos. The STM feedback and scanning is done with a commercial controller (RHK Technology Inc. SPM 1000) and the current is amplified either by a RHK pre-amplifier at fixed gain (10^9 V/A) or a Femto Gmbh pre-amplifier at variable gain. Both amplifiers have comparable noise level of about 0.2 $\mathrm{pA}/\sqrt{\mathrm{Hz}}$ with the above mentioned sensitivity and the input impedance of the wires. The STM has a XY-range of 300 nm × 300 nm at base temperature and about 5 times more at room temperature. The STM is lateral and vertical stable to about 0.2 Å in 30 minutes.

Even though the STM is suspended on air-dampers and relatively isolated inside the UHV chamber, intermittent external mechanical and acoustic noise sources were able to couple into the tunneling signal at times. Particularly 2 and 11 Hz from the eigen-frequency of the air-dampers and ~20-24 Hz from the air-conditioning system of the laboratory. The first can be reduced by improving the damping of the air-pillars, but the latter would require a full acoustic isolation of the entire setup.

The exchange-gas STM is a home made STM in helium boil-off gas within a variable temperature insert (VTI) from Cryogenic Ltd (see Fig. 2.1) mounted through the bore of a 12 T magnet. Using resistance heating, the sample temperature can be varied between 4.2 K and 340 K. Once inserted into the cryostat, samples were held between 300 K and 340 K while being flushed with dry helium gas to effect desorption of contaminants. To minimize the sample cryopumping while at low temperatures, samples were kept warmer than the surrounding VTI. Ambient STM measurements were made using the same STM as used for helium gas measurements, but without inserting the STM into the VTI. The cryostat was initially suspended with 3 springs on the laboratory ceiling but in a later stage mounted in a heavy frame (stainless steel filled with sand) on top of air-dampers from Newport. The STM head was designed and built in Leiden.[8] The coarse-approach is based on the standard Pan design, and it can be used either as slip-stick or walker (the latter more reliable at low temperatures). This STM offers more XYZ range (2 um at 300 K, 5 times less at 4 K) because it is based on a piezo-tube scanner. Tip and samples cannot be changed in situ. The noise level is somewhat better than for the UHV STM because of the lower input impedance of the wires. The STM is suspended at the end of a relatively floppy insert (resonant frequency about 200 Hz) with a spring of lower frequency ~100 Hz. Also this STM was affected intermittently by acoustic noise from the laboratory ventilation system. The same type of electronic was used as for the UHV STM.

One issue common to both the flow-cryostat STM and the UHV STM is the length of the coaxial cable connecting the STM to the transimpedance amplifier used to measure the tunneling current. This is a problem because the capacitance of a coaxial cable is proportional to its length (∼100 pF/m)[†] and increasing the input capacitance of the transimpedance amplifier, as will be discussed in more detail below, will increase its noise floor. Unfortunately, the length of the coaxial cable was driven by the need to cool the STM to cryogenic temperatures and a concomitant requirement for thermal isolation.[‡]

Both mechanically cut PtIr (90%:10%) and electrochemically etched Pt wires were used as STM tips. No significant difference between these two types of tips was apparent.

2.3 Basic STM Function

The STM tip-sample distance can be controlled with sub-nanometer resolution because of the exponential dependence of tunneling current on tip-sample distance. The tunneling current (I_{tunnel}) as a function of tip-sample separation (z) can be written,[9]

$$I_{tunnel} \propto e^{-2\kappa z}, \tag{2.1}$$

where

$$\kappa^2 = 2m\bar{\Phi}/\hbar^2, \tag{2.2}$$

and where $\bar{\Phi}$ is the average work function, and m the electron mass. Using a typical value for the work function, 5 eV, the tunneling current will change by an order of magnitude for every 1 Å change in the barrier thickness.[9]

In order to operate, the STM tip and sample must first be close enough for electrons to tunnel between them but not so close as to be in contact. This is usually accomplished with a combination of coarse approach and fine-control mechanisms. Typical coarse approaches are capable of motion on the order of millimeters. Within a micron of the surface, the fine control will take over. A schematic showing the basic components of an STM is shown in Fig. 2.2. In this Figure, the tip is mounted onto a piezoelectric tube—fine-control—which itself is mounted onto the coarse approach mechanism. This is the configuration the flow-cryostat and UHV STM's used in this research. Many other arrangements are possible. Much more will be said about both the coarse and fine control mechanisms later in this Chapter.

Operation of an STM begins with the tip and sample separated by a distance on the order of millimeters, which allows the tip or sample to be changed without

---

[†] One STM used in this research used Lakeshore Cryogenics, Inc. Type C coaxial cable with a capacitance of 79 pF/m.

[‡] The coaxial cables for the flow-cryostat STM and the UHV STM were ∼1 m and ∼5 m, respectively.

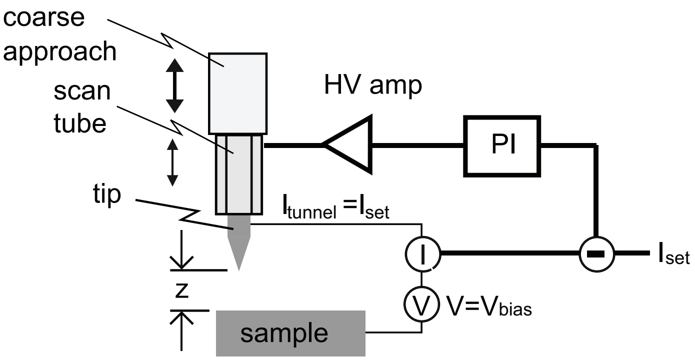

Figure 2.2: Basic STM components plus feedback loop. Adapted from Ref. 10.

damaging either of them. A fixed bias voltage (V_{bias}) is then applied between the tip and sample. The tunneling current between the tip and the sample is measured by a transimpedance amplifier. With an initial tip-sample separation of millimeters, there will be no detectable current. The tip and sample are then brought together slowly by stepping the coarse approach. Each step of a typical STM coarse approach will consist of retracting the fine-control mechanism, stepping the coarse mechanism, and then extending the fine-control mechanism while testing for a tunneling current. This cycle is repeated until I_{tunnel} rises to the current set-point (I_{set}). Once the coarse approach has halted, a feedback mechanism will assume control of the tip-sample separation.

The purpose of the STM feedback mechanism is to maintain tunneling current set-point by adjusting the tip-sample distance with the fine-control mechanism, usually a piezoelectric actuator. This feedback mechanism is needed because the tip-sample distance can change because of changes in the sample topography as the tip is scanned laterally, due to thermal drift, or other errors such as piezoelectric creep. At the heart of the feedback mechanism is an analog or digital proportional-integral or PI controller and a high-voltage (HV) amplifier. The PI controller adjusts the tip-sample distance in response to any difference between I_{tunnel} and I_{set}. If $I_{tunnel} < (>) I_{set}$, then the PI controller will decrease (increase) z using the fine-control mechanism causing an increase (decrease) in I_{tunnel} until $I_{tunnel} = I_{set}$. The HV amplifier is necessary because the strain coefficient of the piezoelectric materials used in the fine-control mechanism is small, requiring a large field to produce any appreciable motion. Due to the exponential dependence of I_{tunnel} on z, small changes in z will generate large changes in I_{tunnel} [see Eq. (2.1)]. It is this sensitivity and the small strain coefficient of piezoelectric materials that permits the PI controller to effectively control the tip to within Å's of the surface.[10]

The final component needed for a functioning STM is a way to scan the tip laterally across the sample surface. In the design described above, a piezoelectric tube will provide this function. In one STM used in this research, scanning was accomplished

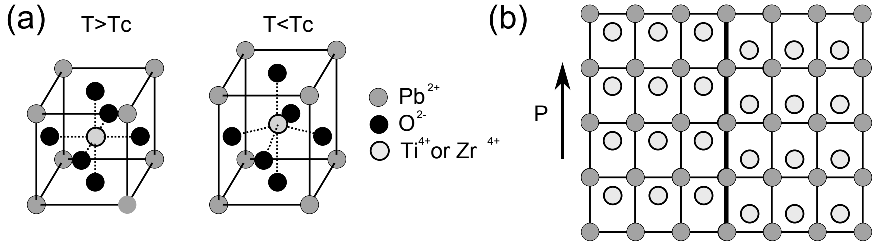

Figure 2.3: (a) Structure of PZT is cubic above T_c and tetragonal below. (b) Domain structure of PZT in tetragonal/ferroelectric phase. Adapted from Refs. 12,13

with an actuator comprised of two shear piezoelectric actuators, stacked one on top of the other, with their shearing motions directed orthogonally. Other arrangements are possible.[11]

2.4 Piezoelectric Actuators

In this research, all of the STMs used piezoelectric actuators for the fine-control, scan, and coarse approach functions. Because of this, an introduction to piezoelectricity and piezoelectric actuators, including a discussion of their advantages and disadvantages follows.

2.4.1 Basics of Piezoelectric Materials

Piezoelectric materials are commonly used for STM actuators. The most common material is PZT (lead zirconate titanate, $\mathrm{Pb[Zr}_x\mathrm{Ti}_{1-x}\mathrm{]O}_3$). Among the advantages of piezoelectric materials, and in particular PZT, are that they have high sensitivity (on the order of 10 nm motion per applied volt); that they can be made UHV compatible (minimal outgassing, bakeable); that they do not generate nor are they affected by magnetic fields; and that they can operate at temperatures from mK to well above room temperature. PZT is a solid solution of two perovskites, PbTiO_3 and PbZrO_3, with mixtures of containing roughly equal parts of Ti and Zr showing the largest piezoelectric strain coefficients (see below).[12]

The structure of PZT is shown in Fig. 2.3(a). Above the Curie temperature (T_c) PZT is cubic and paraelectric, but below T_c it is tetragonal and ferroelectric. To form the tetragonal structure, four of the Pb^{2+} ions will shift along, what is designated, the c-axis, elongating the unit cell. Simultaneously, the Ti^{4+} or Zr^{4+} ion, formerly at the center of the unit cell, will be displaced along the same axis, and together with the O^{2-} ions, this will create a ferroelectric dipole parallel to the c-axis. Because the elastic energy will favor adjacent unit cells having parallel c-axes, domains will form with parallel dipoles. However, because this polarization will also create a surface charge with its associated depolarization field, large domains will break up into

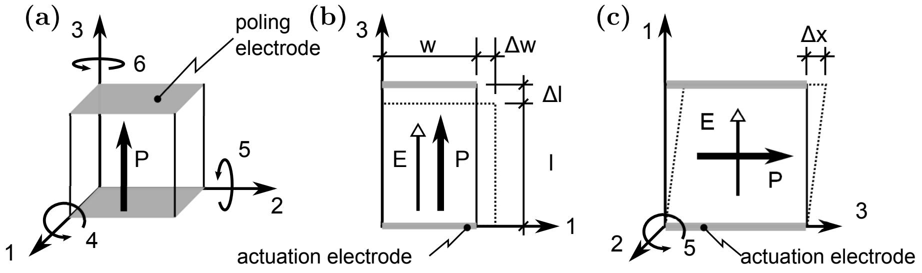

Figure 2.4: Relationship between poling axis, applied field, and actuation. (a) The axes are defined in relation to the poling axis. (b) A field applied parallel to poling axis will cause both a parallel and a perpendicular response. (c) A field applied perpendicular to the poling axis will cause shearing parallel to the poling axis. Adapted from Ref. 12

smaller domains with anti-parallel polarization [see Fig. 2.3(b)].[13] Defects and strain will cause further domain generation, resulting in domains with randomly oriented polarizations. Given this and that commercial PZT is polycrystalline, virgin PZT will have no net polarization. To be useful, PZT must be poled in a high electric field at, possibly, an elevated temperature to reorient the domains in one predominant direction.[12,14,15] PZT is ferroelectric and, therefore, a piezoelectric material. This is true of all ferroelectrics. A piezoelectric material, however, need not be ferroelectric. Quartz is a prominent example of a non-ferroelectric piezoelectric.

The piezoelectric effect is the generation of an electric field in response to strain. Piezoelectric actuators rely on the inverse piezoelectric effect, where an applied electric field induces strain. For the inverse piezoelectric effect in one-dimension, the strain (S) is related to the applied field (E) by,

$$S = d_1 E, \tag{2.3}$$

with the piezoelectric constant, d_1, in unit of meters per volt. At constant stress T (the force per unit area), this constant is defined by,

$$d_1 = \left(\frac{\partial S}{\partial E}\right)_T. \tag{2.4}$$

In general, a piezoelectric material will also be subject to stress, and so, in three-dimensions, the equation of state relating the inverse piezoelectric and elastic strain is given by,

$$\mathbf{S} = \mathbf{s^E}\mathbf{T} + \mathbf{d}\mathbf{E}, \tag{2.5}$$

where $\mathbf{S}$ and $\mathbf{T}$ are now tensors, $\mathbf{d}$ is a matrix, and $\mathbf{s^E}$ is the elastic compliance matrix defined in constant electric field.[12]

The actuation direction of a piezoelectric is determined by the relationship between the poling axis (P) and the electric field axis. When $E \parallel P$, the material responds by straining both parallel and perpendicular to the field axis [Fig. 2.4(b)]. When $E \perp P$, the material responds by shearing perpendicular to the field axis [Fig. 2.4(c)]. The directions of a piezoelectric are labeled 1, 2, and 3, forming a right-handed orthogonal coordinate system [Fig. 2.4(a)]. By convention, axis 3 denotes the poling direction. The piezoelectric properties are given in relation to this coordinate system. The strain coefficient (d_{ij} [m/V]) gives the strain in direction-j with a field applied in direction-i. For STM actuators, the d_{33} (parallel), d_{31} (perpendicular), and d_{15} (shear) coefficients are the most important. Writing Eq. 2.5 explicitly,[12]

$$\begin{bmatrix} S_{11} \\ S_{22} \\ S_{33} \\ 2S_{23} \\ 2S_{31} \\ 2S_{12} \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{21} & s_{22} & s_{23} & 0 & 0 & 0 \\ s_{31} & s_{32} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{bmatrix} \begin{bmatrix} T_{11} \\ T_{22} \\ T_{33} \\ T_{23} \\ T_{31} \\ T_{12} \end{bmatrix} + \begin{bmatrix} 0 & 0 & d_{31} \\ 0 & 0 & d_{32} \\ 0 & 0 & d_{33} \\ 0 & d_{15} & 0 \\ d_{15} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{2.6}$$

This equation can be considerably simplified if the piezoelectric is unconstrained and not subject to external stress ($T_i = 0$). Then for an applied field E_3 only (which means $E \parallel P$), Eq. 2.6 will reduce to,

$$S_{11} = d_{31}E_3 \text{ and } S_{22} = d_{32}E_3, \tag{2.7}$$

which is a strain along both axes perpendicular to the applied field, and

$$S_{33} = d_{33}E_3, \tag{2.8}$$

a strain along the same axis as the applied field. Similarly, if a field is applied along axis 1, E_1, then the piezoelectric will shear. From Eq. 2.6,

$$\gamma = 2S_{31} = d_{15}E_1, \tag{2.9}$$

where γ is the total shear strain.

Equations 2.7 through 2.9 can be rewritten in terms of applied voltage, V, and the piezoelectric dimensions. If V is applied across a piezoelectric of length l and in the same direction as the poling direction, then the displacement parallel to the field is,

$$\Delta l = d_{33}\left(\frac{V}{l}\right)l, \tag{2.10}$$

while the displacement perpendicular to the field is,

$$\Delta w = d_{31}\left(\frac{V}{l}\right)w. \tag{2.11}$$

Similarly, when potential V is applied across a piezoelectric of length l, but perpendicular to the poling direction, the piezoelectric will shear, with the total shear given by,

$$\Delta x = d_{15}\left(\frac{V}{l}\right)w. \tag{2.12}$$

Piezoelectric materials can also generate a force. For a field applied along axis 3, consider only the strain along axis 1 and ignore the strain along axis 3. In this case, Eq. 2.6 will reduce to,

$$S_{11} = s_{11}T_{11} + d_{31}E_3. \tag{2.13}$$

This is Eq. 2.7 generalized to include an additional term, $s_{11}T_{11}$, accounting for any force applied to the piezoelectric along axis 1. This could be an externally applied force, or the force generated as the piezoelectric attempts to expand, contract, or shear against a constraint. Consider a simplified case of a piezoelectric material contained within perfectly rigid constraint along axis 1. This constraint is the same length as the piezoelectric before the field is applied so it will initially be unstrained. Because of the constraint, though, the piezoelectric cannot expand or contract along axis 1. This means that $S_{11} = 0$ and Eq. 2.13 can be rewritten as,

$$T_{11} = -d_{31}E_3/s_{11}, \tag{2.14}$$

where T_{11} is the stress (or force per unit cross-section) generated. The force generated is simply the stress multiplied by the cross-sectional area perpendicular to axis 1,

$$F = A_1T_{11} = -A_1d_{31}E_3/s_{11}. \tag{2.15}$$

This is the maximum force a piezoelectric can generate for a particular applied field. Piezoelectric actuator manufacturers will often list a blocking force as one of an actuators specifications. The blocking force is the maximum force that a piezoelectric actuator can generate at the maximum rated voltage. Increasing the voltage above the rated maximum will in principle increase the maximum force, but doing so will reduced actuator lifetime, and at high enough field cause dielectric breakdown. Operating the actuator at less than the maximum rated voltage will result in a blocking force less than the maximum. As can be seen from Eq. 2.15, the blocking force is linear in applied field.

Similar equations can be found for shear piezoelectrics. The shear strain for a piezoelectric with a field applied along axis 1 will be,

$$\gamma = 2S_{31} = s_{55}T_{31} + d_{15}E_1. \tag{2.16}$$

Now consider the same shear piezoelectric without an externally applied stress but with a perfectly rigid constraint resisting the piezoelectric from shearing. This will mean that $\gamma = S_{31} = 0$, and Eq. 2.16 will be reduced to,

$$T_{31} = -d_{15}E_1/s_{55}. \tag{2.17}$$

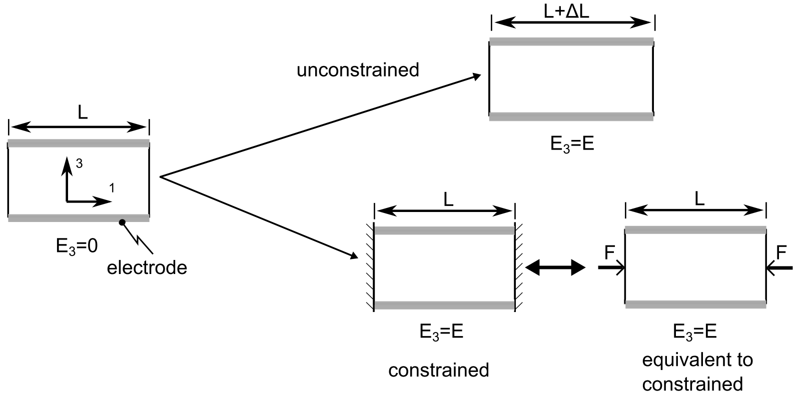

Figure 2.5: Piezoelectric unconstrained (giving rise to a change in the length L); or constrained along axis 1 with electric field applied along axis 3.

This is the shear stress that will be felt by the constraint. The force generated is just the shear stress scaled by the cross-sectional area of the piezoelectric perpendicular to the poling direction,

$$F = A_3 T_{31} = -A_3 d_{15} E_1 / s_{55}. \tag{2.18}$$

The forces that piezoelectric actuators generate can be understood in a different but equivalent way by comparing a piezoelectric without constraint to the same piezoelectric with constraint. Consider the piezoelectric shown in Fig. 2.5. Before E_3 is applied, the piezoelectric has length l. After the field is applied, the unconstrained piezoelectric will have expanded by Δl along the 1-axis, while the constrained piezoelectric will remain the same length.[†] Viewed from a alternate but equivalent prospective, consider the force needed to compress the unconstrained piezoelectric of length $l + \Delta l$ down to length l. For a small strain and within the elastic range of a material, the force required to uni-axially compress a body is given by Hooke's Law, $F = EA\Delta l/(l + \Delta l)$, where E is the elastic modulus of the material and A is the area perpendicular to the strain. PZT, the piezoelectric material used in most STM actuators, is an approximately Hookean material, and so this equation applies. This force is exactly the same as the force that the constrained piezoelectric will impose on the constraints.

2.4.2 Scan Tube

Topography is typically measured by scanning a tip over surface a while the feedback loop adjusts z to maintain $I_{tunnel} = I_{set}$. For this a method of moving laterally is

---

[†] This is true for a perfectly rigid constraint. If the constraint is compliant, the piezoelectric will lengthen to some degree, and the force imposed on the constraint by the piezoelectric will be less than that for a perfectly rigid constraint.

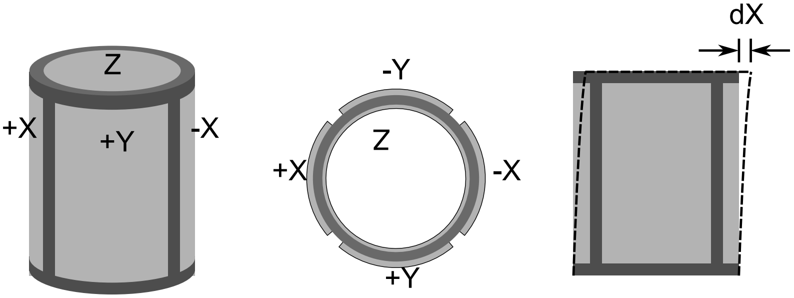

Figure 2.6: 5-Electrode Scan Tube.

needed. This is usually accomplished with one or more piezoelectric actuators. A common configuration, and one that has been used since the early days of STM,[16] is a single piezoelectric tube with one electrode on the inner surface and four electrodes on the outer surface (Fig. 2.6). If poled radially, this single, 5-electrode tube is sufficient provide 3-axis motion.

For z-axis motion, a voltage is applied between the inner diameter electrode, Z, and the four outer diameter electrodes, $+X$, $-X$, $+Y$, and $-Y$, with the same voltage applied to all four outer electrodes. This will subject the tube to a field parallel (or antiparallel) to the poling direction. The tube will respond both parallel ($d_{33}E_3$) and transverse ($d_{31}E_3$) to the field. The parallel response will cause an increase or decrease in the tube diameter, but, because of axis-symmetry, this strain will play no role in the STM tip's motion.[†] The transverse response will cause the tube to extend or retract in proportional to the applied field. Using the standard methods of continuum mechanics, the relationship between the change in applied voltage (ΔV) and the change in length (Δz) for an unconstrained tube has been shown to be,[17]

$$\Delta z = \frac{d_{31}\Delta V L}{h}, \tag{2.19}$$

with d_{31}, as before; L the tube length; and h the tube wall thickness.[‡] This can be rewritten as,

$$\Delta z = A_z \Delta V, \tag{2.20}$$

where $A_z = d_{31}L/h$. Because of broad manufacturing tolerances (i.e., dimensions plus material properties), A_z is usually determined experimentally. Piezoelectric material

---

[†] This is also because in most STMs with five-electrode scan tubes, the tip is mounted on a rigid plate attached to the end of the tube. Any radial expansion of the tube will be resisted by the tip mounting. This was not the case with the original five-electrode scan tube where the tip was affixed directly to the inner diameter of the tube.[16]

[‡] Eq. 2.19 assumes that $h \ll D$, where D is the tube diameter. This is not actually true for the scan tube used in most STM's. For instance, the scan tube from the STM used for some of the research in this Thesis has dimensions $h = 0.5$ mm and $D = 8$ mm. But violating this assumption is not important since STMs are calibrated before use.

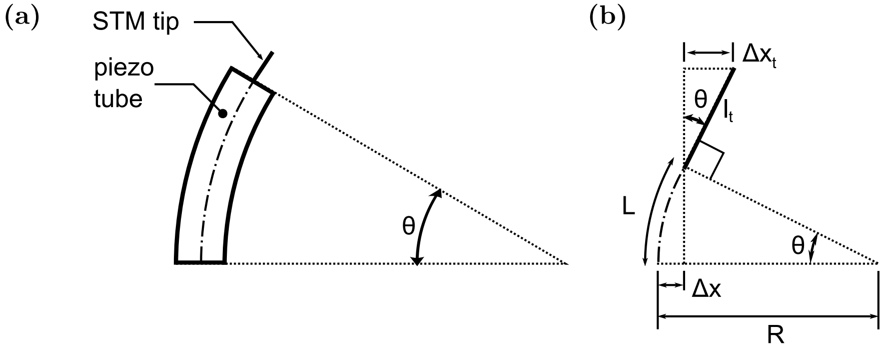

Figure 2.7: (a) Unequal voltages applied to the outer electrodes of the piezoelectric tube will cause it to bend. Because of this the ends of the tube will no longer parallel and the STM tip will be tilted. (b) Schematic of bent piezoelectric tube showing the contribution to lateral motion from bent tube (ΔX) and tilted tip (ΔX_t).

properties, including d_{31}, are temperature dependent. Typical strain coefficients for the piezoelectric material used in STM scan tubes, type PZT-5A, are $d_{31} = -1.73$ Å/V at room temperature, and $d_{31} = -0.3$ Å/V at liquid-helium temperature.[18] Again because of manufacturing tolerances, d_{31} is not known accurately at any temperature, and the STM scan tube should be calibrated as a variety of temperatures. Calibrations are usually made by measuring a known sample. Commonly used reference samples include highly oriented pyrolytic graphite, HOPG,[19] Au(111),[20] and even the surface reconstructions of silicon, Si(111)-7x7 and Si(001)-2x1.[21]

Lateral motion is accomplished by applying unequal voltages to the outer electrodes. This will cause it to bend. Analysis of tube bending is more complicated than the analysis of z-axis motion. Consider the case where the inner electrode is grounded, one of the outer electrodes has a voltage is applied to it, and the other three outer electrodes are grounded. As the voltage is applied, the quadrant subjected to the field will attempt to respond by lengthening or shortening axially and by expanding or contracting radially, but will be partially constrained by the two quadrants adjacent to it. A full analysis of this can only be accomplished numerically and this has been done with the finite elements method.[22] However, by making some reasonable simplifying assumptions, an analytical expression can be found.[17,23] First consider how a scan tube is typically used in an STM. One end of the tube is rigidly attached to a base while a tip holder is attached to the free end. The effect of the base and tip holder is to constrain the tube ends to remain planar (although not necessarily parallel to each other).[†] Now the force generated by one quadrant will be transmitted to all four quadrants by the base and tip holder and all four quadrants

---

[†] Without the tip holder, the free end of the piezoelectric tube will not remain planar if unequal voltages are applied to the electrodes. Finite element analysis of a piezoelectric tube with voltage applied to only one quadrant shows that the quadrant subject to voltage will be extended while the adjacent quadrants will be distorted. This can be seen in Fig. 1 of Carr.[22]

will be strained. But, as the force is exerted by only one quadrant, the force will not be distributed axis-symmetrically, and a moment will be generated. This moment will cause the tube to bend. Applying a voltage to either the $+X(Y)$ or $-X(Y)$ electrode will bend the tube along the x(y)-axis. Most tube scanners, though, are operated in a symmetric mode. In this mode voltages are applied to pairs of electrodes, either $+X$ and $-X$ or $+Y$ and $-Y$. A voltage is applied to one electrode while a voltage of equal amplitude but opposite polarity is applied to the other electrode. This will double the tube's deflection per volt. In symmetric mode, the lateral tube deflection will be,[17]

$$\Delta X \text{ or } \Delta Y = \frac{2\sqrt{2}d_{31}VL^2}{\pi Dh}, \tag{2.21}$$

where V is the voltage applied to one quadrant with $-V$ applied to the opposite quadrant. Note that this only accounts for the deflection of the tube. Because the tube deflects by bending, the free end of the tube–where the tip is mounted–will be tilted (see Fig. 2.7). As it is tilted, the end of the tip will be shifted laterally by more than the end of the scan tube. The tilt angle of the tip can be found from the curvature of bending of the tube. Using the same simplifying assumptions as were used to calculate the tube deflection, the curvature of bending of the tube, R, is given by,

$$R = \frac{\pi Dh}{4\sqrt{2}d_{31}V}. \tag{2.22}$$

From simple geometry, the angle between the top and base of the scan tube, θ, is found to be,

$$\theta = \frac{360L}{2\pi R}. \tag{2.23}$$

Using this angle, the additional lateral deflection of the tip can be found using simple geometry. That is,[23]

$$\Delta X_t = l_t \sin(\theta_x), \tag{2.24}$$

$$\Delta Y_t = l_t \sin(\theta_y), \tag{2.25}$$

where l_t is the tip length, and θ_x and θ_y are the bending along the x- and y-axes respectively.

For a typically STM scan tube, R is very large while θ is very small. This can be illustrated with a scan tube from one of the STM's used in this research. The relevant parameters for this tube are $L = 20$ mm, $D = 8$ mm, $h = 0.5$ mm, and $d_{31} = -0.173 \times 10^{-9}$ m/V. With a voltage of 100 V applied symmetrically, the tube deflection will be ΔX or $\Delta Y = 1.56 \times 10^{-6}$ m, with $R = 1.28 \times 10^{2}$ m and $\theta = 8.92 \times 10^{-2}$ °. Assuming a tip of length $l_t = 6$ mm, the additional lateral deflection due to the tilted tip would be ΔX_t or $\Delta Y_t = 9.35 \times 10^{-7}$ m. That is a

60% increase over the scan tube deflection alone. Although the tip itself is usually shorter than this, the length relevant for this calculation is the distance from the end of the scan tube to the end of the tip, including the thickness of the tip holder, ground plane, etc. Although a calibration can be performed, this calibration should be repeated after every tip change. From Fig. 2.7 it would also appear that tube bending will shift the tip by $\Delta Z_t = l_t(1 - \cos(\theta))$. However, because angle θ is so small, ΔZ_t will be negligible for any practical STM scan tube (for the above example, $\Delta Z_t = 7.3 \times 10^{-9}$ m). A more complete analysis, including the tip being offset from the center of the scan tube has been performed by Yang *et al.*[23]

An alternative to the scan tube is to combine two shear piezoelectric actuators and one extension piezoelectric actuator to form an actuator capable of three-axis motion—a three-axis shear/extension actuator. In this scheme, the two shear piezoelectric actuators are mounted with their shear motions oriented orthogonally. Such actuators are sold commercially. They are much more rigid than tube actuators and so have higher first mode frequencies. Finite element calculations show that the first vibration mode for a three-axis shear/extension actuator can easily be an order of magnitude larger than that for scan tube of comparable scan range ($\approx$ 10 kHz versus $\approx$ 1 kHz). Unfortunately, this rigidity comes at the cost of higher capacitance. For example, one manufacturer offers a three-axis shear/extension actuator with a room-temperature scan range of 10μm with a capacitance of 100 nF per axis while their tube scanner of similar scan range has a capacitance of only 10 nF per axis.[24] Such large capacitances render shear/extension actuator unsuitable for ultra fast-scanning STM's. As high scan rate is unimportant for STS, one of the STM's used in this research successfully used a shear/extension actuator for scanning.

2.4.3 Coarse Approach

The purpose of the coarse approach mechanism is to bring the tip and sample close enough that electrons can tunnel between them. To make tip and sample changes convenient they should be separated by at least a millimeter. The coarse approach must move this distance in steps smaller than the maximum z-range of the scan tube to avoid crashing the tip into the sample, as has already been explained.

The requirements for an STM coarse approach are quite demanding. As was stated above, the coarse approach should be able to move over distances of millimeters in sub-micrometer steps.[25] It must be rigid enough that any compliance is undetectable in either topography or spectroscopy. In general, and especially for variable temperature STM's, the coarse approach should be designed to minimize thermal drift. For use in a high magnetic field, as some of the STM's used in this study were, the coarse approach should also be non-magnetic. Operation in UHV requires that it be resistant to bakeout temperatures of up to 150 ℃, and also that it functions in UHV.[†] If it is to operate below room temperature it must accommodate thermal expansion and

---

[†] At room temperature in normal humid air, most surfaces would be expected to be covered in water. This can act at a lubricant. In UHV, some or all of this water will be desorbed. How this effects friction depends upon the surface.[25,26]

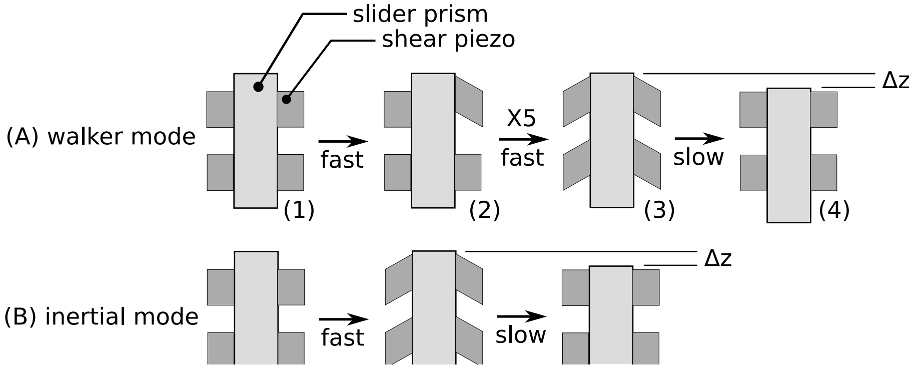

Figure 2.8: Schematic of walker (a) and inertial (b) coarse approach mechanism. Note that the prism is three-sided, so that there are a total of six piezo-actuators.

contraction, and it must not produce more heat than the cooling system can remove. This is particularly true of very low temperature STM's which are usually cooled by very low power systems such as ^{3}He or dilution refrigerators. Any electrical or mechanical power input must also not act as a conduit for excessive heat flow into the cooler parts of the system.

The solution that best accommodates all of these requirements is a coarse approach using piezoelectric actuators. Since the invention of the STM in 1982,[27] many piezoelectrically-driven coarse approach mechanisms have been used successfully. Underlying most of these designs are two types of motion: walker and inertial. The walker mode is similar to animal locomotion, relying on sets of piezoelectric actuators moving out of phase with each other. The inertial mode uses alternating fast and slow piezoelectric actuation and the inertia of a slider mass for motion.

To illustrate the walker and inertial coarse approach modes, the basic operation of the Pan design will be described.[28] The design is based on a prism-shaped slider. Further details of this design will be given later. Both modes are illustrated in Fig. 2.8. The Pan design relies on a total of six shear piezoelectric actuators in two rings around the slider. For clarity, only four of them are shown in this Figure. The walker mode functions by moving each actuator individually and in sequence. While the other five remain fixed, one actuator is stepped by applying a rapidly increasing voltage to cause that piezoelectric to shear quickly enough to overcome static friction and slide along the prism [Fig. 2.8(A.2)]. While this actuator is shearing, the prism is held in place by the five stationary actuators. After sliding, the actuator is then held fixed while the other actuators are stepped in turn [Fig. 2.8(A.3)]. Once all six actuators have been sheared, they are then simultaneously retracted slowly back to their starting position [Fig. 2.8(A.4)]. Each cycle will move the prism by the distance the actuators were able to slide along the prism. The prism will only move a small distance with each cycle, typically less than 1 μm. Moving a millimeter or more, as

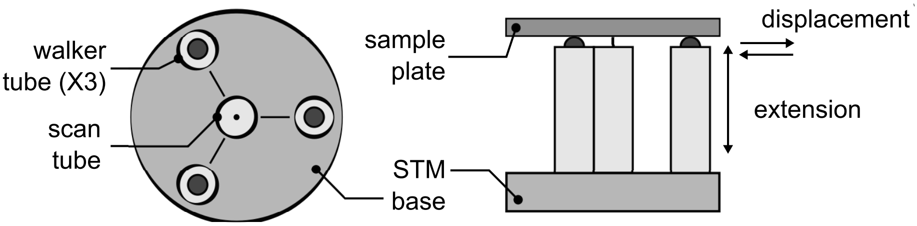

Figure 2.9: Schematic of first incarnation of beetle walker, the Besocke beetle.[30] This design provided lateral but no vertical coarse motion.

will be required after a tip or sample change, will require of the order of 1000 cycles.

In the inertial coarse approach mode, all of the actuators are sheared simultaneously. The inertial mode consists of a repeated cycle of alternating fast and slow actuation. During the fast phase, the actuators are sheared quickly enough that the inertial force of the prism will exceed the static friction between the prism and the actuators and the actuators will slide along the prism [Fig. 2.8(B.2)]. Following this, all six actuators are slowly retracted back to their starting position [Fig. 2.8(B.3)]. Because the actuators slid along the prism during the fast phase, there will be a net movement of the prism. The step size will be the distance the actuator slide along the prism, and, as with the walker mode, each step will only move the prism less than 1μm. Because the actual motion of the actuator and prism is more complicated than described here, the step size will not necessarily be the same as the distance the actuators were sheared.[†]

Survey of Coarse Approach Designs

The original STM coarse approach presented by Binnig *et al* was a horizontal walker.[16,27] The design, 'the louse', consisted of three feet connected via a piezoelectric body. Walking was performed by clamping two trailing feet, extending the body, clamping the leading foot, releasing the other feet, and contracting the body. Although this system operated successfully, because of issues with thermal drift, it is not the best choice for a variable temperature STM.[29] Finally, it is relatively large and cannot be accommodated easily into the bore of a superconducting solenoid.

Another design that appeared in the early days of STM, but, in modified form, is still uses in both homebuilt[31,32] and commercial[33] STM's is the Besocke beetle design.[30] In its first incarnation, it was not capable of vertical coarse motion but, rather, was designed to provide horizontal coarse motion. The design relied upon four 5-electrode piezoelectric tubes, with three tubes used as legs for coarse motion, and one used for scanning (see Fig. 2.9).[‡] The three coarse motion legs form an

---

[†] For example, in walker mode, the five stationary actuators are compliant and they will deflect somewhat as the walker leg deflects. Similarly, in inertial mode the prism can stop sliding before the actuator stroke is complete, or can continue after the actuators stops.

[‡] Later incarnations have replaced the 5-electrode walker tubes with shear piezoelectric actua-

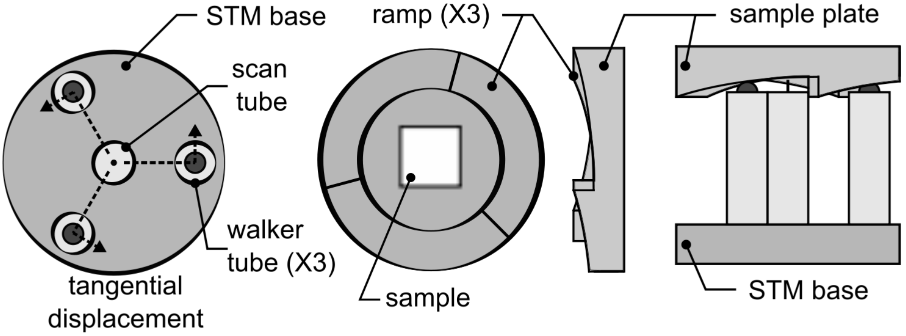

Figure 2.10: Schematic of beetle STM with ramps for coarse approach motion.

equilateral triangle with the scan tube at its center. The sample is mounted onto a flat sample plate which sits on top of the legs. Horizontal motion begins by extending all three legs to separate the tip and sample. All three legs are then rapidly deflected horizontally. If the deflection is rapid enough, then the sample plate's inertia will resist its motion, and the piezoelectric legs will slide under the sample plate. After sliding, the legs are then deflected horizontally back to their starting position slowly enough so as not to slide. The sample plate will now be shifted in relation to the tip by the distance the legs slid. To move further, the cycle can be repeated. To recommence scanning, the walker legs are retracted back until the tip and sample are close enough for tunneling.

The original beetle design can function as a coarse approach if the sample is tilted from horizontal. However a modified version of the beetle was soon conceived by Frohn *et al* that permitted vertical coarse motion with the sample held horizontally.[35] The modification involved machining three circular ramps, one for each leg, into the underside of the sample plate (Fig. 2.10). These ramps translate horizontal into vertical motion. The leg deflection must now be tangential to the circle defined by the three coarse motion legs, with the scan tube at its center. The leg deflection will now rotate rather than translate the sample plate. As the sample plate rotates, the legs will walk up or down the ramps, increasing or decreasing the tip-sample separation. The maximum coarse approach motion will be limited because each ramp can only occupy 120°and they cannot be steeper than the legs can climb. The original Frohn design design was limited to a 0.3 mm vertical approach motion.[35] The great advantage of the beetle design is that, by making the legs and scan tube from the same piezoelectric material, it is thermally stable and particularly well suited to variable temperature operation.

Soon after the beetle design, a linear inertial STM design was published by Lyding *et al.*[29] Like the beetle design, this design was also thermally compensated, but using two concentric piezoelectric tubes, with one for scanning and the other for the coarse

---

tors.[34]

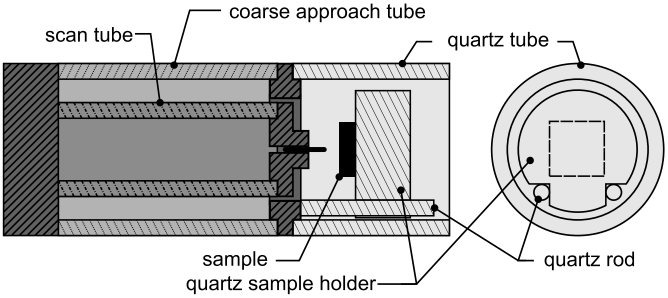

Figure 2.11: Schematic of the STM with linear inertial coarse approach designed by Lyding *et al.*[29]

approach (see Fig. 2.11). The tip is mounted to the inner piezoelectric tube while a quartz sample holder is clamped to quartz rods connected to the outer tube. Coarse motion is accomplished by applying a sawtooth voltage to the outer piezoelectric tube, making is oscillate in the z direction—fast in one direction, and slow in the other. During this cycle, when the tube is accelerating rapidly, the inertia of the sample holder will cause it to slide along the quartz rods. This design is thermally compensated because both piezoelectric tubes the same length and because quartz, with its low thermal expansion coefficient, is used for the sample holder and rods. Finally, the thermal expansion of the tungsten tip is compensated for by adding a copper beryllium spacing collar to the outer piezoelectric tube. In this original form, the coarse approach was designed to operate horizontally, requiring the STM to be mounted so as to accommodate this. As with the louse discussed above, this orientation is inconvenient for high magnetic field operation where the diameter of the superconducting solenoid limits the maximum horizontal dimension. Reorienting this design to operate vertically proved to be troublesome. Renner *et al* found that a coarse approach of similar design would not operate vertically with a sawtooth-profile piezoelectric driving voltage.[36] They developed a cycloidal voltage profile that would drive their linear inertial motor vertically. In general, the piezoelectric driving voltage profile is critical to the functioning of piezoelectric motors and this will be discussed in more detail below. A later incarnation of the linear inertial motor was capable of x, y, z motion.[37] This design combined a slider mounted on three orthogonal rails with a five-electrode piezoelectric tube capable of 3-axis driving motion.

The final STM design to be discussed here features a coarse approach capable of both inertial and walker motion and is, in modified form, the coarse approach mechanism chosen for the homebuilt STM's used in this research. The design, invented by Pan *et al*,[28] consists of a triangular cross-section prism (with truncated vertices) and three pairs of shear piezoelectric actuators, one pair in contact with each face (see Fig. 2.12). Two pairs of these actuators are attached to a rigid STM body. The remaining pair is attached to a plate and pressed against the prism with a spring. The normal force between the actuators and the prism is varied by adjusting the spring

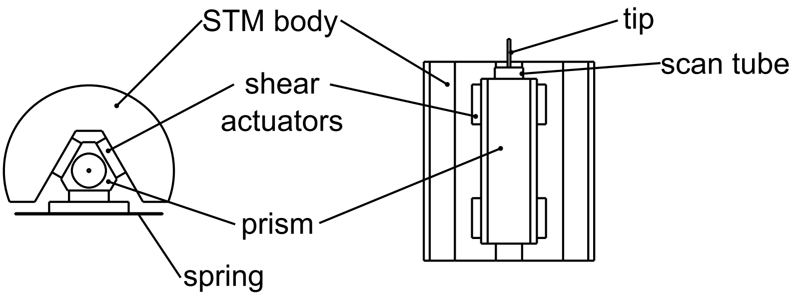

Figure 2.12: Schematic of the Pan-design STM coarse approach.[28]

force. In the original publication this coarse approach was operated in only the walker mode. The advantages of this design when used in the walker mode include reliability and suitability for vertical operation.[28] Because the walker mode does not rely on inertia, the step size when walking against gravity or with gravity is about the same. Nevertheless, this design can also be used in inertial mode, which is how this design was used in this research. To operate in walker mode, six pairs of wires are needed, one pair for each piezoelectric actuator. In inertial mode, only one pair of wires is required as all six piezoelectric actuators are driven simultaneously.

Coarse Approach Motors at Low Temperatures

Low temperature poses severe challenge to STM coarse approach motors. This challenge stems primarily from weakened piezoelectric actuators, but can be exacerbated by increased friction caused by frozen contaminants. Piezoelectric actuators are weaker at low temperatures because the piezoelectric coefficient is reduced. For example, at liquid helium temperature (4.2 K), the piezoelectric coefficient, d_{31}, of PZT-5A is reduced to about $\approx$17% of its room temperature (293 K) value (from -1.73 Å/V to -0.30 Å/V).[18] This means that at 4.2 K, for a given applied voltage, a piezoelectric actuator will only be able to move by $\approx$17% of the distance it would move at room temperature but also be only able to exert a force $\approx$17% of what it could at room temperature (see Eq. 2.15).

This reduced force can be a significant issue for both inertial and walker modes where the piezoelectric actuators must be able to overcome static friction. During the sliding portion of the cycle, the shear piezoelectric actuators of a Pan-design STM need to exert a force in excess of the static friction between them and the prism. For Coulomb friction between two surfaces not in relative motion, the (static) friction force is $F_f \leq \mu_s F_N$, where μ_s is the coefficient of static friction, and F_N is the normal force between the surfaces. This static friction will inhibit the surfaces sliding against each other until the force applied tangential to the surface, F, exceeds the maximum static friction, $F_f^{max} = \mu_s F_N$. Once the surfaces begin to slide, the friction force will be reduced to $F_f = \mu_k F_N$, where μ_k is the coefficient of kinetic friction, and where

$\mu_k < \mu_s$. Sliding will continue until $F < \mu_k F_N$. Because experiments have shown that for some material combinations in contact under high vacuum, including sapphire against sapphire, μ_s does not change significantly between room temperature and 4.2K,[26] and the piezos will still need to overcome the same static and kinetic friction as at room temperature. Further compounding this, at low temperatures, contaminants can freeze onto the sliding surfaces and, possibly, dramatically increase μ_s and μ_k. A combination of weakened piezoelectric actuators and frozen surface contaminants can severely reduce the reliability of low-temperature STM coarse approaches.

Two of the STM's used in this research were required to function at low temperatures. While one of these STMs was used exclusively in UHV, the other was operated in flowing helium boil-off gas. Despite similar coarse approach designs, the UHV STM was more reliable at low temperatures (T<40 K) than the He-gas STM. The most likely explanation is that there were more frozen contaminants on the sliding surfaces of the He-gas STM than on the sliding surfaces of the UHV STM. This is simply because the UHV STM is baked out before it is cooled and the He-gas STM is not. Although helium boil-off gas is likely free of a significant concentration of contaminants, neither the He-gas STM nor the variable temperature insert (VTI) in which it was operated were as free of contaminants, especially water, as the UHV STM and UHV chamber. Instead of a bake-out—which was not possible, the He-gas STM was inserted into the VTI and warmed to ≈350 K for up to several hours while enveloped in flowing helium boil-off gas. The combination of dry helium boil-off gas and warming will remove some of the adsorbed water on the STM sliding surfaces, but not as much as would be removed by a UHV bake-out.[†]

Beyond their different operating conditions, the most significant difference between the UHV STM and the He-gas STM was in the materials used for their sliding surfaces. While the UHV STM used a titanium carbide-coated prism in contact with alumina, the gas-flow STM used a sapphire prism in contact with sapphire. However, this difference is unlikely to be the cause of the He-gas STM's poor low-temperature performance. A third STM was built with the same sliding surfaces as the UHV STM. This STM performed especially poorly in He-gas.[‡] This leaves surface contamination, probably water, combined with piezoelectric actuators weakened at low-temperature as the most likely causes of poor low-temperature performance of the gas-flow STM coarse approach. A UHV bake-out will remove some of the contaminants and most likely increase the reliability of a low-temperature Pan-style STM.

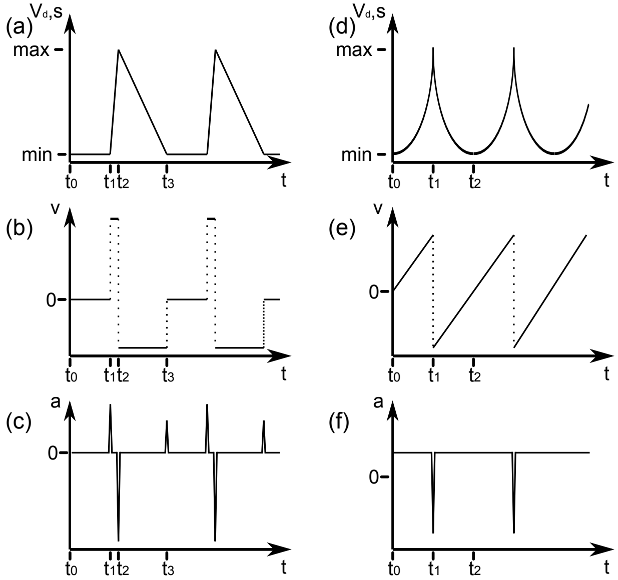

Figure 2.13: Illustration of the saw-tooth [(a),(b), and (c)] and cycloidal [(d),(e), and (f)] voltage profiles (V_d), and corresponding actuator position (s), velocity (v), and acceleration (a).

Coarse Approach Driving-Voltage Profile

The voltage profile used to drive a piezoelectric actuator in an STM coarse approach is critical to its function at low temperature, and particularly when operating vertically against gravity. The first inertial STM coarse approach, the Lyding STM discussed above, was driven by a saw-tooth profile [Fig. 2.13(a)], but could only function horizontally.[29] This design was then adapted for vertical operation by Renner *et al* by, amongst other things, changing the voltage profile used.[36] This modification was deemed necessary because a saw-tooth profile has three sharp turning points that will cause high acceleration and thus—possibly—sliding motion in both directions. This is a rather subtle point and will be discussed in more detail in what follows.

An illustration of a saw-tooth profile of the driving voltage (V_d) and the corresponding actuator position (s), velocity (v) and acceleration (a) are shown in Fig. 2.13(a), (b), and (c). The V_d cycle consists of three phases: a quiescent phase be-

---

[†] Baking out involves pumping the system down to about 10^{-8} mBar, then heating the chamber to above 100 °C for several days while pumping continuously. Water is the main contaminant removed during bake-out. Adsorbed water will be desorbed very quickly from the surface, while absorbed water must first diffuse to the surface. As diffusion is thermally activated, UHV chambers are heated to expedite the process.

[‡] During preliminary tests in the boil-off gas within the neck of a helium Dewar, the coarse approach of this STM failed to function even while still above 200 K.

tween t_0 and t_1; a fast ramp phase between t_1 and t_2; and a slow ramp phase between t_2 and t_3. During the quiescent phase the actuator does not move. At t_1 the fast ramp begins and V_d is ramped up to the maximum voltage as fast as the driving electronics will permit. At t_2, the slow ramp begins and V_d is decreased at a rate lower than during the fast ramp. During both the fast-ramp and the slow ramp the actuator will be moving at an approximately constant speed.[†] At t_1, t_2, and t_3 the actuator must apply a brief force or impulse to the slider mass to change the its velocity. This is illustrated in Fig. 2.13(c). During these three impulses the actuator will be accelerating and thus applying a force to the slider. Between impulses the slider will be moving at an approximately constant velocity and the actuator will not be applying a force to it. Only when the actuator is applying a force can the slider begin to slide.[‡] Thus sliding can commence during any or all of the three impulses of each cycle. Two of these impulses apply a force in one direction while the third applies a force in the opposite direction. The slider can therefore slide in both directions during each cycle. Ideally the velocity profile is tuned such that only during the larges impulse, the one at t_2, will the actuator apply force sufficient to overcome static friction and cause sliding.

Other voltage profiles have been proposed to sidestep the careful tuning required for the saw-tooth profile. After failing to operate an inertial coarse approach oriented vertically with a saw-tooth voltage profile, Renner *et al* developed a quadratic, cycloidal voltage profile to drive their piezoelectric actuators [Fig. 2.13(d)]. They hypothesized that during each saw-tooth cycle the coarse approach moves forward at t_2 but backwards at t_1. When combined with gravity—as would be the case with the course approach mounted vertically, the course approach ceases to function. In contrast, the quadratic profile has, except at the cusps, a constant, low acceleration. At the cusps, the actuator velocity is reversed by an impulse [Fig. 2.13(f)]. The cycloidal voltage profile must be tuned such that during the constant acceleration phase the actuator force is less than F_f^{max}, while during the impulse the actuator force exceeds it. Because there is only one impulse in each cycle, there will only be sliding in one direction, and therefore the coarse approach should move more efficiently.

Reversing the coarse approach direction can be accomplished in variety of ways. For the saw-tooth profile, reversing the profile in time or polarity will work. For the cycloidal profile, the polarity must be reversed.

Mariotto *et al* studied a Pan-style coarse approach in both inertial and walker modes, mounted vertically, but only at room temperature.[25] They applied a voltage profile consisting of a voltage step followed by a quarter sine-wave decrease to zero voltage (see Fig. 2.14). They found that the in the walker mode the step size per cycle was greater than in the inertial mode. They also found that the actual slope of the voltage step, limited by the high voltage amplifier slew rate and the capacitance being

---

[†] At driving frequencies less than the natural frequency of the piezoelectric actuator-coarse approach system, the actuator position will follow the driving voltage reasonably closely and, therefore, $s \propto V_d$.

[‡] The force must also exceed the maximum static friction, F_f^{max}, between the actuator and the slider for sliding to occur.

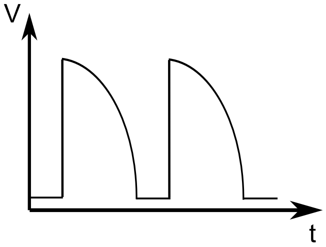

Figure 2.14: Quarter sine voltage profile used by Mariotto *et al.*[25]

driven, determined the step size in both inertial and walker modes. That the voltage slope is critical is not entirely obvious because for any sliding to occur, whether in the inertial of walker mode, the piezoelectric must accelerate and a linear voltage increase has zero acceleration. The authors attribute the slope dependence to the viscoelastic properties of the epoxy used in assembling the shear piezoelectrics.

For the STMs used in this study, when at room temperature, even if mounted vertically, the particular voltage profile used was not critical, and a variety of amplitudes and frequencies could be used to drive the coarse approach. Once cooled below room temperature, though, the voltage profile, amplitude, and frequency become critical. Even small changes to the driving frequency or amplitude can result in the coarse approach not functioning. This indicates that the simple picture of a piezoelectrically driven coarse approach described above is incomplete.

2.5 Measuring the Tunneling Current

The tunneling current needs to be measured for both feedback control and for tunneling spectroscopy. Typically, a transimpedance amplifier is used. This type of amplifier outputs a voltage that is proportional to the input current.

2.5.1 Amplifier Basics

The basic transimpedance amplifier circuit is conceptually remarkably simple and involves only one resistor and one operational amplifier (op-amp) [Fig. 2.15(a)].

Assuming an ideal op-amp and ignoring the feedback resistor parasitic capacitance (C_{fb}) the output voltage from the circuit shown in Fig. 2.15(a) will be,

$$V_o = -I_i R_{fb}, \tag{2.26}$$

where V_o is the output voltage, I_i is the input current, and R_{fb} is the feedback resistor. Although this circuit works, in order to measure the very small tunneling currents usually encountered in STM, the feedback resistor needs to be very large. The tunneling resistance in an STM typically ranges from about 100 MΩ to 10 GΩ, while the bias voltage ranges from about 100 mV to 1 V. This will mean that a typical

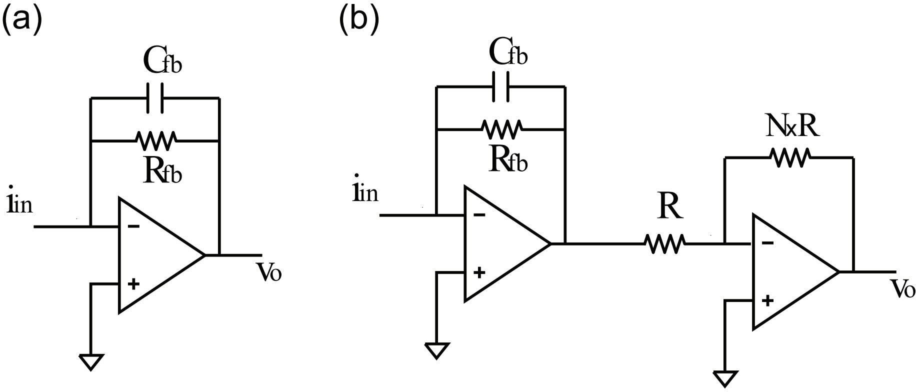

Figure 2.15: Schematic of (a) transimpedance amplifier and (b) transimpedance amplifier with second stage of gain N. Adapted from Refs. 10,33,38.

tunneling current will range from about 10 nA to 100 pA. To produce a convenient output voltage, $V_o = 1$ V, the amplifier gain needs to be between 10^8 and 10^{10} V/A, requiring a feedback resistor of between 100 MΩ and 10 GΩ.

Using such large resistances offer both advantages and a disadvantages. On the plus side, the thermal current noise from the feedback resistor, the major source of noise in transimpedance amplifiers,[10] decreases as the resistance increases. The thermal current noise (or Johnson noise), I_n, is given by,[10,38]

$$I_n = \sqrt{\frac{4k_BTB}{R_{fb}}}, \tag{2.27}$$

where k_B is the Boltzmann constant, T is the resistor temperature, and B is the bandwidth over which the noise is measured. It can be seen that, for a given bandwidth, increasing the resistance will decrease the current noise. This suggest that R_{fb} should be maximized to minimize noise. Unfortunately, as we will now see, because of parasitic capacitance, increasing R_{fb} will also reduce the amplifier bandwidth.

All real resistors suffer from parasitic capacitance, C_{fb}. This capacitance is equivalent to a capacitance in parallel with the feedback resistor [Fig. 2.15(b)]. At higher frequencies, C_{fb} will short out R_{fb}, reducing the effective feedback resistance, and, as can be deduced from Eq. 2.26, reduce the gain of the amplifier. Although parasitic capacitances are typically small, on the order of 1 pF,[10] when combined with very large resistances, the effect becomes significant. Including C_{fb} in parallel with R_{fb} the gain, G, becomes,

$$G = \frac{V_o}{I_i} = \frac{R_{fb}}{\sqrt{1 + (2\pi C_{fb}R_{fb})^2}}, \tag{2.28}$$

The addition of the parasitic capacitance has made the amplifier into a low-pass filter with a -3 dB point at frequency,[10]

$$f = \frac{1}{2\pi R_{fb} C_{fb}}. \tag{2.29}$$

Substituting typical values for an STM transimpedance amplifier, $R_{fb} = 100$ MΩ and $C_{fb} = 1$ pF, results in a bandwidth of $f \approx 1.5$ kHz, which is quite limiting in practical use.

A narrow transimpedance amplifier bandwidth imposes severe restrictions on both topographic and spectroscopic measurements. These include requiring more time to measure topography and spectroscopy and, possibly, making modulation techniques infeasible.

Topographic measurements will take longer because the feedback mechanism uses the tunneling current for error correction. To avoid loosing fine detail from topography and to avoid crashing the tip into the sample the scanning speed must be reduced.

Spectroscopic measurements will also take longer. To compensate for a narrow bandwidth, the bias voltage sweep-rate must be reduced, increasing the IV curve measurement time. This can be a significant problem for dense spectroscopy maps requiring hundreds of IV measurements, particularly if averaging is necessary and multiple measurements must be made at each point. Increased measurement time also exacerbated problems associated with drift. During a fixed-gap IV curve measurement the bias voltage is swept with the feedback disengaged.[†] During this time the tip-sample distance can change because of thermal drift. Because of the exponential dependence of tunneling current on tip-sample distance, a small change in tip-sample distance can cause a large change in tunneling current, large enough to distort or obscure the spectroscopic information contained in the measurement.

Lastly, if modulation techniques are used, a narrow amplifier bandwidth will require that either the modulation frequency be decreased to within the bandwidth or that the modulation amplitude be increased to compensate for the reduced gain at the modulation frequency. Neither option is desirable. The ideal modulation frequency is above the feedback bandwidth but within the transimpedance amplifier's bandwidth. If the modulation frequency is within the feedback bandwidth, then the tip height will also be modulated. If the modulation frequency is above the feedback bandwidth and if the modulation amplitude is increased to compensate for this, then the spectroscopic data will be broadened, and fine detail being lost. For modulation amplitudes of more than a few percent of the bias voltage, the measurement will no longer be a measure the of the IV curve slope.

2.5.2 Practical Transimpedance Amplifier Design

Because of the limitations imposed by the narrow bandwidth, practical STM transimpedance amplifiers are usually a modified version of the simple, one large resistor

---

[†] If it were not disengaged the feedback would adjust the tip-sample distance to compensate for changes in bias.

amplifier. Three variants of the one-resistor transimpedance amplifier will be discussed here: a two-stage amplifier, a T-network feedback amplifier, and a compensating second-stage amplifier.

Two-Stage Amplifier

The two-stage amplifier consists of a transimpedance amplifier followed by a voltage amplifier. A simplified circuit diagram of this scheme is shown in Fig. 2.15(b). Because the voltage amplifier will boost the overall amplifier gain, the transimpedance amplifier gain can be reduced by using a smaller feedback resistor. This smaller feedback resistor will reduce the impact of parasitic capacitance and increase the amplifier bandwidth (see Eq. 2.29). The overall gain can be made as large as the overall gain of a single-stage transimpedance amplifier with a larger feedback resistor. The two-stage cascade amplifier is the solution favored by RHK, the manufacturer of the some of the commercial amplifiers used in this study.[33] These amplifiers combine a transimpedance amplifier with a 1 MΩ[†] feedback resistor followed by a 100X gain voltage amplifier. This is an overall gain of 10^8V/A. The disadvantage of this scheme is that the resistor thermal noise of the first stage will increase because of the smaller feedback resistor and the second stage will amplify this extra noise. To address this issue, one variant of this amplifier features a second stage with a variable bandwidth, low-pass filter. This allows the bandwidth to be trimmed to the minimum required for scanning and spectroscopy, cutting off higher frequencies that contribute only to the noise floor.

T-Network Feedback Amplifier

A second way to expand the transimpedance amplifier bandwidth is to replace the single large feedback resistor with an impedance T-network [see Fig. 2.16(a)]. This T-network consists of three impedances, one of which is connected to ground, with each impedance composed of either a resistor, a capacitor, or a resistor in parallel with a capacitor. Two variants of this will be discussed here: a resistor network replacing a single large feedback resistor,[10] or directly compensating for the parasitic capacitance in the feedback of the transimpedance amplifier itself with a resistor-capacitor (R-C) combination,[10,39,40] Circuit schematics of the resistor network and R-C compensating feedback are shown in Fig. 2.16(b) and (c). Both of these techniques can be understood as special cases of a more general impedance network feedback.

Fig. 2.16(a) shows a simplified circuit diagram of a transimpedance amplifier with a T-network feedback of three impedances, Z_1, Z_2, and Z_3, each of which can represent a resistor, a capacitor, or a parallel resistor and capacitor. Its transfer function is defined as the ratio of the output voltage to the input current, $G = v_o/I_{in}$. To find this, first sum the currents at the common connection point of the three

---

[†] We modified this amplifier to increase the gain by substituting a 10 MΩ or 100 MΩ feedback resistor to boost the overall gain to 10^9 V/A and 10^{10} V/A respectively.

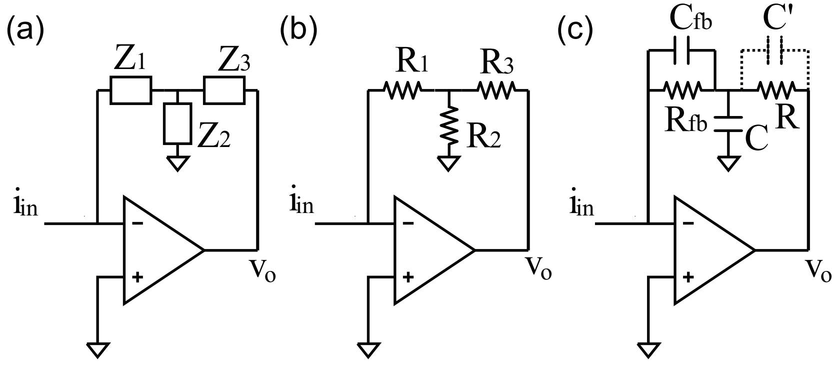

Figure 2.16: Schematic of transimpedance amplifier with (a) impedance network feedback, (b) resistor network feedback, and (c) feedback compensation. Schematics (b) and (c) are special cases of (a). Adapted from Refs. 10,38–40.

impedances,[38]

$$\frac{(0-V_1)}{Z_1}+\frac{(0-V_1)}{Z_2}+\frac{(V_o-V_1)}{Z_3}=0. \tag{2.30}$$

If the op-amp is assumed to be ideal, no current will flow into the $V_{(-)}$ input, and all of the input current must flow through the feedback. For an input current of I_i, therefore, the voltage drop across Z_1 will be,

$$(0-V_1)=I_iZ_1. \tag{2.31}$$

This can then be combined with Eq. 2.30 to solve for the amplifier gain,

$$G=\frac{V_o}{I_i}=-\left[1+\frac{Z_3}{Z_1}+\frac{Z_3}{Z_2}\right]Z_1. \tag{2.32}$$

The gain of the two special cases is calculated by substituting specific values for the three impedances.

The resistor network consists of three resistors connected in a T-network. As will be discussed below, the resistance of each of the three resistors can be smaller than that needed for a single-feedback resistor amplifier, and, because of this, the parasitic capacitance will not reduce the bandwidth as significantly. For a resistor network with $Z_1=R_1$, $Z_2=R_2$, and $Z_3=R_3$,[†] Eq. 2.32 gives,

$$\frac{V_o}{I_i}=-\left[1+\frac{R_3}{R_1}+\frac{R_3}{R_2}\right]R_1=-\left[R_3+R_1\left(1+\frac{R_3}{R_2}\right)\right]. \tag{2.33}$$

[†] In this analysis, impedance Z_1 is taken as a purely resistive, without a parallel parasitic capacitance. This is because for smaller resistances, the -3 dB point frequency, which is inversely proportional to the product of resistance and capacitance (Eq. 2.29), should be beyond the bandwidth of interest. To analyze the resistor with parallel capacitance, the impedance of a resistor in parallel with a capacitor can be substituted for Z_1.

This equation shows how the resistor network boosts the gain. In addition to a small gain from R_3 directly, the voltage divider formed by R_2 and R_3 acts to multiply the gain of R_1. This is because the op-amp must increase v_o to compensate for the voltage divider which reduces the voltage that R_1 is subject to (V_1). The resistance of R_1 can be chosen so that it and its parasitic capacitance will lead to the desired bandwidth. Then R_2 and R_3 can be chosen so that their ratio, R_3/R_2, provides the necessary boost to the gain. As with the cascade amplifier discussed above, the drawback of this design is that, because the feedback resistors are smaller, the thermal noise current noise will be larger.

The R-C compensating feedback consists of a feedback resistor, R_{fb}, with parasitic capacitance in parallel, C_{fb}, and a second resistor, R, and capacitor, C, that act to compensate for this parasitic capacitance. The third capacitance, C', shown in Fig. 2.16(c) is optional and its significance will be discussed below. For the R-C feedback compensation, R_{fb} is sized to provide all of the gain needed while the R-C portion of the feedback is sized to compensate for the parasitic capacitance and expand the bandwidth. Again, Eq. 2.32 can be used to analyze this circuit. Impedance Z_1 will now be resistor, R_{fb}, in parallel with a capacitor, C_{fb}. Thus,

$$Z_1 = \frac{R_{fb}}{1 + i\omega R_{fb} C_{fb}}. \tag{2.34}$$

Now set $Z_2 = 1/i\omega C$ and $Z_3 = R$ and substitute into Eq. 2.32

$$G = \frac{V_o}{I_i} = -\left[R + R_{fb}\frac{(1\ +\ i\omega RC)}{(1 + i\omega R_{fb} C_{fb})}\right] \tag{2.35}$$

If $RC = R_{fb}C_{fb}$ then C_{fb} will be perfectly compensated for by R and C, and the amplifier frequency response will be flat.[10,40] A second capacitor can be added in parallel to R to reduce the bandwidth to the minimum needed and thereby limit the amplified noise.[40] A disadvantage of this design is that it is unstable for high input capacitance.[41] As long coaxial cables are unavoidable for the low-temperature, high magnetic field STMs used in this research, R-C feedback compensation would be impractical.

Compensating Second Stage

A third method to compensate for the bandwidth-limiting parasitic capacitance is to follow the transimpedance amplifier with a high-frequency boosting second stage.[41,42] A simplified circuit diagram of the design proposed by Carla *et al* is shown in Fig. 2.17.[42] In this scheme, the first stage consists of a transimpedance amplifier with a single feedback resistor large enough to provide sufficient gain and to achieve low thermal current noise. As has already been discussed, a large resistor in parallel with its its parasitic capacitance will reduce the bandwidth of the amplifier. Including parasitic capacitance, the transfer function of the first stage is,

$$H_1(\omega) = \frac{R_{fb}}{1 + i\omega R_{fb} C_{fb}}. \tag{2.36}$$

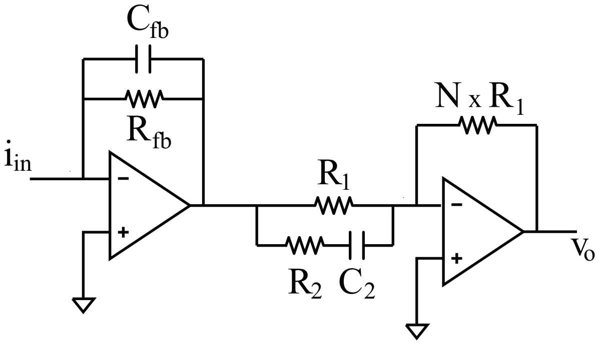

Figure 2.17: Schematic of transimpedance amplifier with high-frequency boosting second stage. Adapted from Ref. 42

This is a low-pass filter with bandwidth $\omega_{h1} = 1/R_{fb}C_{fb}$. To expand the bandwidth, the second stage must invert this and flatten out the frequency response. The second stage proposed by Carla et al is just a voltage amplifier with a resistor, R_2, and capacitor, C_2, in parallel to R_1. The parallel resistor-capacitor will act to short out R_1, increasing the second-state gain at higher frequencies. The transfer function of the second stage is,

$$H_2(\omega) = N\frac{1 + i\omega C_2(R_1 + R_2)}{1 + i\omega R_2C_2}. \tag{2.37}$$

where N is the second stage low-frequency gain. If $C_2(R_1 + R_2)$ is set equal to $R_{fb}C_{fb}$, then the transfer function of the combined two-stage amplifier will be,

$$H(\omega) = N\frac{R_{fb}}{1 + i\omega R_2C_2}. \tag{2.38}$$

The overall amplifier bandwidth is $\omega_h = 1/R_2C_2$ while the overall low-frequency gain is NR_{fb}. The second stage has perfectly compensated for the parasitic capacitance of the first-stage feedback resistor, but still has a limited bandwidth. Nevertheless, because it is the sum of $R1$ and R_2 multiplied by C_2 that is used to compensate for the parasitic capacitance, there remains some flexibility in choosing R_2 and C_2, allowing some control over the bandwidth, ω_h. R_{fb} can now be set so as to minimize the thermal noise current. Other researchers have proposed similar high-frequency boosting stages.[41,43]

2.5.3 Amplifier Noise

The feedback resistor Johnson noise is only one of a number of amplifier noise sources. The op-amp itself contributes noise and is exacerbated by the particular way that transimpedance amplifiers are used in STM current measurements. There exists a rich

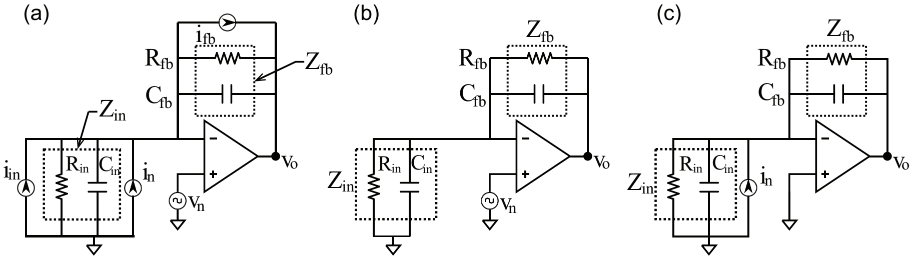

Figure 2.18: Noise model of STM transimpedance amplifier. (a) Noise model showing all noise sources. (b) Noise model showing only the op-amp voltage noise (v_n). (c) Noise model with noise current (i_{eq}) equivalent to v_n. Adapted from Refs. 38,42,44,45.

literature of analyzing noise in transimpedance amplifiers. Much of this analysis concerns photodiode applications, but it can easily be adapted to STM transimpedance amplifiers.

With noise sources included, a STM transimpedance amplifier can be modeled as shown in Fig. 2.18(a).[38,42,44,45] These noise sources are the Johnson noise of the feedback resistor (i_{fb}), the op-amp current noise (i_n), and op-amp voltage noise (v_n). The resistor current noise is produced by the random thermal motion of current carriers and was discussed above (Eq. 2.27). This noise is modeled as a current source in parallel with a noiseless resistor. Johnson noise is the minimum noise possible noise for a resistor of a given value. Real resistors are noisier, though metal resistors come close to the minimum thermal noise.[10] The op-amp noises, v_n and i_n, are generated by the op-amp itself and are usually specified in data sheets issued by the manufacturer; v_n is specified as a voltage at positive input of the op-amp, $v_{(+)}$, while the current noise is specified as a current in parallel with the input current.

Also shown in Fig. 2.18(a) are an input resistance (R_{in}) and an input capacitance (C_{in}). This input impedance (Z_{in}) is the sum of the capacitance across the input terminals of the op-amp, the capacitance between the tip and sample, and the capacitance of the coaxial cable connecting the STM to the amplifier. The input capacitance of op-amps used for STM current measurements is typically about 3 pF,[†] and while the tip-sample capacitance is negligible,[46] the coaxial cables connecting the tip/sample to the transimpedance amplifier have capacitances of $\sim$100 pF per meter. The amplifier input coaxial cable need not be very long for the capacitance of the cable to dominate the other sources of capacitance. As will be shown, large input capacitances lead to substantial noise. Finally, the feedback impedance (Z_{fb}) consisting of R_{fb} and C_{fb} in parallel, will be included in the model.

To understand the relative importance of the noise sources, the equivalent input current spectral density of each of the noise sources should be calculated. In accord

---

[†] Three op-amps commonly used for STM transimpedance amplifiers are the OPA111[41] and OPA656 from Texas Instruments and the AD549[42] from Analog Devices. All have input capacitances of less than 3 pF.

with the principle of superposition, each noise can be treated separately, then summed quadratically. When modeling each noise, the op-amp itself will be treated as noiseless. In all of the following analysis, the virtual short-circuit approximation will be assumed. This will mean that the positive and negative inputs of the op-amp will be at the same potential, and that no current can flow into or out of either of them.

First consider the op-amp noise voltage, v_n. The amplifier with all other noise sources removed and no input current is illustrated in Fig. 2.18(b). Because this noise is assumed to be connected at the positive input of the op-amp, $v_{(+)} = v_n$. This would cause $v_{(-)} \neq v_{(+)}$ and the op-amp will respond by driving v_o. With $v_o \neq 0$, a current will flow from the output of the op-amp and through Z_{fb}. Because no current can flow into the negative input of the op-amp, all of the current will then flow through Z_{in} and the current will be,

$$I = -v_o \frac{1}{(Z_{in} + Z_{fb})} \tag{2.39}$$

The voltage drop across the input impedance will be,

$$0 - v_{(-)} = IZ_{in} \tag{2.40}$$

Eliminating I from Eq. 2.39 with Eq. 2.40, and solving for v_0 gives,

$$v_o = -v_{(-)} \frac{(Z_{in} + Z_{fb})}{Z_{in}} \tag{2.41}$$

Since the op-amp is assumed to be ideal, v_o will be driven until $v_{(-)} + v_{(+)} = 0$. This and Eq. 2.41 can be used to solve for the output voltage resulting from a noise voltage at the op-amp's positive input terminal,

$$v_o = -v_n \frac{(Z_{in} + Z_{fb})}{Z_{in}} \tag{2.42}$$

In order to compare it to other noise sources, v_0 must be converted to an equivalent input noise current at the op-amp input. Illustrated in Fig. 2.18(c) is the equivalent circuit with v_n removed and replaced by an equivalent current, i_{eq}. By the virtual-short approximation, $v_{(-)} = v_{(+)} = 0$ V, and there is no voltage drop across Z_{in}, and no current can flow through it. All of the current, therefore, will flow through Z_{fb},

$$v_o = i_{eq} Z_{fb} \tag{2.43}$$

This can be combined with Eq. 2.42 to eliminate v_o and solve for i_{eq},

$$i_{eq} = -v_n \left(\frac{Z_{in} + Z_{fb}}{Z_{in} Z_{fb}} \right) = -v_n \left(\frac{1}{Z_{in}} + \frac{1}{Z_{fb}} \right) \tag{2.44}$$

With this, the equivalent input spectral density current can be calculated for the three noise sources considered in this analysis,

$$i^2 = i_n^2 + v_n^2 \left(\frac{1}{Z_{in}} + \frac{1}{Z_{fb}} \right)^2 + \frac{4k_B TB}{R_{fb}} \tag{2.45}$$

Now the relative noise contribution from each sources can be compared. The current noise (i_n) intrinsic to the op-amp is fixed for each op-amp and a matter of choosing an op-amp with lower input current noise. Reducing the noise current caused by v_n can be accomplished by either minimizing v_n itself or maximizing Z_{in} and Z_{fb}.[42] Once again, v_n is fixed for each op-amp, and can only changed by selecting another op-amp. Both Z_{in} and Z_{fb} are combinations of resistors and capacitors in parallel. The impedance for a resistor (R) in parallel with a capacitor (C) is,

$$Z(\omega) = \frac{1}{1/R + i\omega C}. \tag{2.46}$$

Thus, to increase the impedance, increase R and decrease C. Substituting Eq. 2.46 into Eq. 2.45 gives,

$$i^2 = i_n^2 + v_n^2 \left(1/R_{in} + i\omega C_{in} + 1/R_{fb} + i\omega C_{fb}\right)^2 + \frac{4k_B TB}{R_{fb}} \tag{2.47}$$

The input resistance is a combination of circuit board isolation, wiring insulation, and tip or sample isolation. With proper design, $R_{in} \gg R_{fb}$, and R_{in} can be neglected. The feedback impedance, C_{fb}, cannot be reduced below the intrinsic parasitic capacitance of the feedback resistor,[†] but, if almost any length of coaxial cable connects the STM to the transimpedance amplifier input, then $C_{in} \gg C_{fb}$, and C_{fb} can be neglected. Eq. 2.47 reduces to,

$$i^2 = i_n^2 + v_n^2 \left(i\omega C_{in} + 1/R_{fb}\right)^2 + \frac{4k_B TB}{R_{fb}} \tag{2.48}$$

To decrease the output noise current, decrease C_{in} and increase R_{fb}.

The final noise source is the Johnson noise of the feedback resistor. As can be can be seen from Eq. 2.45, increasing R_{fb} will directly reduce the feedback resistor noise. In addition, because of the parasitic capacitance, increasing the feedback resistance will narrow the amplifier bandwidth, and this will also decrease the amplifier noise.

Now a note concerning the current noise spectral density versus the voltage noise spectral density of the feedback resistor Johnson noise. Rewriting the current noise spectral density in terms of voltage,

$$v_{thermal} = i_n R_{fb} = \sqrt{\frac{4k_B TB}{R_{fb}}} R_{fb} = \sqrt{4k_B T R_{fb} B}, \tag{2.49}$$

It seems that, in contrast to the current noise spectral density (Eq. 2.27), the thermal voltage noise will be larger for larger resistors. Thus, because $v_o = v_{thermal}$, the noise voltage at the output will also increase with increasing R_{fb}. This is true, but it is not the absolute size of the output noise that is important, but rather the signal to

---

[†] Large resistors with very low parasitic capacitance are available. Ciofi *et al* report of commercially available 1 GΩ-plus resistors with parasitic capacitances of ∼100 fF.[43]

noise ratio (SNR). The amplifier output voltage (v_s) for an input signal current (I_s) is given by,

$$v_s = R_{fb} I_s. \tag{2.50}$$

Thus the signal will increase as $v_s \sim R_{fb}$ whereas, from Eq. 2.49, the Johnson noise will increases as $v_n \sim \sqrt{R_{fb}}$. The SNR will therefore increase as SNR $\sim \sqrt{R_{fb}}$.

2.6 Topography

STM topographic images are measured by scanning the tip across the the sample in a square or rectangular pattern. The two raster patterns, x-fast and y-fast, used during measurements for this Thesis are shown in Fig. 2.19. The scans are denoted x-fast or y-fast based upon which direction the lines are measured. This example shows a 512×512 pixel image, the resolution typically used. With the feedback engaged, the tip is scanned along a line parallel to either the x or y direction (line 1). Along this line, the topography is recorded at 512 equally spaced points or pixels. At the end of the line, the tip is scanned back along the same line with another 512 points of topography recorded (line 2). Thus the topography along each line is measured twice, once forward and once in reverse. Once a forward and reverse scan is completed, the tip is moved one pixel in the orthogonal scan direction (3), and the process is repeated. This continues until 512 lines have been measured.

In constant current mode, the feedback mechanism will respond to changes in topography by adjusting the tip height to keep the tunneling current at the current set-point, I_{set}. It does this by varying the voltage applied to the z-direction actuator,V_z. As has been discussed, this actuator is usually a piezoelectric actuator. V_z is recorded at each pixel and changes in topography are calculated using z-axis calibration, A_z. Similarly, the tip lateral position is calculated using the input voltages to the x- and y-axis actuators, V_x and V_y, and the calibration of those actuators, A_x and A_y.

Even with an accurate calibration, the measured topography may not be the true topography if that is defined as the position of atoms at the surface. STM, though,

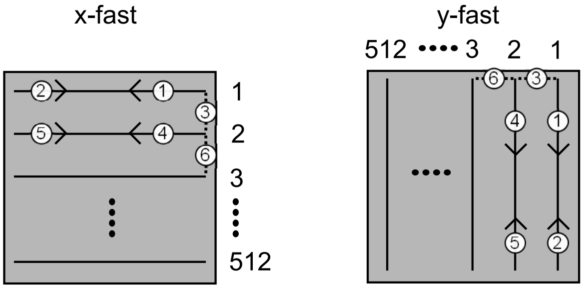

Figure 2.19: Raster scan for STM topography and lock-in conductance map.

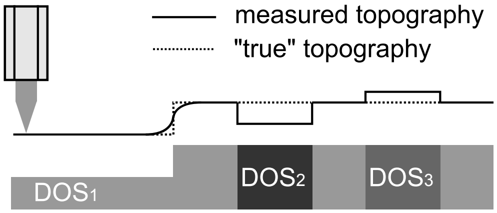

Figure 2.20: STM topography on electronically inhomogeneous sample with step edge.

does not measure the position of surface atoms directly. Instead an STM tip follows a contour of the surface LDOS.[10,47,48] The tunneling current is a convolution of the tunneling barrier and the tip and sample DOS and is given by,

$$I \propto \int_0^{eV} T(S,V,E)\rho_s(E)\rho_t(E-eV)dE, \tag{2.51}$$

where $T(S,V,E)$ is the tunneling barrier, $\rho_s(E)$ is the sample DOS, and $\rho_t(E-eV)$ is the tip DOS. Usually the tip material is chosen so as to have a flat DOS at the Fermi energy, so Eq. 2.51 reduces to,

$$I \propto \int_0^{eV} T(S,V,E)\rho_s(E)dE. \tag{2.52}$$

Thus the tunneling current is proportional to the integrated sample LDOS within eV of the Fermi energy (E_F). This can have some rather unexpected consequences since it makes the current bias-dependent in a non-trivial way. A well-known example is the 7×7 reconstruction on the Si(111) surface, where the unit cell image is different for different bias voltages.[49] On the (110) surface of GaAs, only the Ga or the As atoms are revealed in STM topography, depending upon the polarity of the sample bias voltage.[50] On surfaces with nanometer-scale electronic inhomogeneities, variation in the LDOS can cause features in the measured topography that are the artifacts of electronic variations and not of the true topography. A schematic of a non-homogeneous sample is shown in Fig. 2.20. If, for instance, the DOS is lower in one area of a sample, then the feedback will decrease the tip-sample distance when tunneling in that area, leading to a 'false' dip in the apparent topography.

2.7 Scanning Tunneling Spectroscopy

Mapping the local density of states, LDOS, on material surfaces with an STM is at the core of this research. Information about the LDOS is derived from STS measurements. Two methods to map LDOS predominate in the literature. The first and most complete method is to measure full IV curves at every point. The second method is

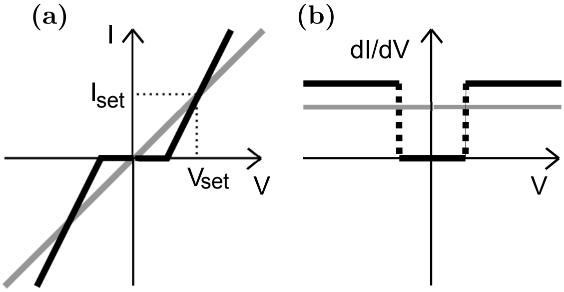

Figure 2.21: Schematic (a) IV and (b) dI/dV curves.

to measure conductance maps with a lock-in amplifier. While a full IV curve maps provide more information, they can be prohibitively time consuming. For this reason, the lock-in method was mostly used in this research.

The full IV method requires that individual IV curves be measured at each point on a grid. Typically the area of interest will be divided up into an n by m grid, with n and m being anywhere from 10 to 64 or more. The tip is scanned over the surface stopping at each point on the grid to measure one or more IV curves. A full map can easily consist of 4096 (64 x 64) or more IV curves.

Measuring an IV curve begins after the STM tip has been positioned at a particular grid-point. After a short delay to permit any vibrations to decay, the feedback mechanism is disengaged, fixing the tip-sample separation and therefore the tunneling barrier. At this point, with the bias voltage remaining at V_{bias}, the tunneling current will still be I_{set}, the value it was before the feedback was disengaged. Now the bias voltage is swept through a range of voltages, with both the applied voltages and corresponding measured tunneling currents recorded. Schematic IV curves are shown in Fig. 2.21(a).

More than one IV curve is usually measured at each grid point and averaged to reduce white noise. The average IV curve from each grid point is then numerically differentiated to calculate the differential conductance (dI/dV) which is then used as a proxy for the sample LDOS. A schematic dI/dV spectra is shown in Fig. 2.21(b).

In the literature and in this Thesis IV curves are often described as being metallic or insulating, a statement that should be used with care. The tunneling current is a convolution of the sample and tip DOS, and a tunneling transmission probability. Because the tunneling transmission probability can dominate the shape of an IV curve, only curves measured *at the same set-point*—the same V_{bias} and I_{set}—should be compared. Consider the IV curves shown in Fig. 2.21(a). At the set-point, the grey line has a lower conductance yet can be said to be more metallic than the black line, especially since the zero-bias conductance is higher.[†] If they were not measured at the same set-point, then if would be difficult to separate differences due

---

[†] This can be verified by noting that IV curves measured at the same set-point must intersect at both $V_{bias} = 0$ and $V_{bias} = V_{set}$

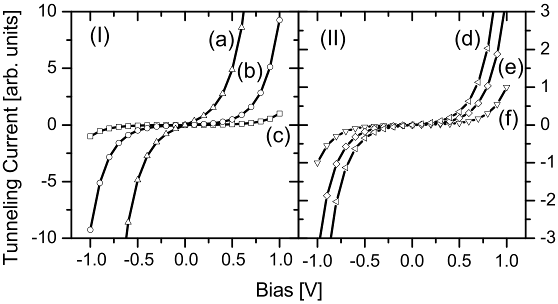

Figure 2.22: Simulated *IV* curves calculated using Eq. 2.66. (I) The effect of increasing the tip-sample in increments of 1 Å. (II) The effect of varying I_{set} or V_{set}. See text for more detail.

to the differing sample DOS's from those due to the differing tunneling transmission probabilities.

The effect of not measuring *IV* curves at the same set-point can be illustrated by calculation. Fig. 2.22 shows *IV* curves calculated using Eq. 2.66 (to be discussed below) for a tip and sample with a flat DOS at E_F and 5 eV average work function, $\bar{\Phi}$. In Fig. 2.22(I), the three curves were calculated with only changes in the tip-sample separation. In curve (a) the tip is closest to the surface while curves (b) and (c) are 1 and 2 Å, respectively, further away. Curve (a) certainly appears more metallic than curve (b) and (c), but the underlying sample DOS used to calculate these three curves is identical, only the tip-sample separation has been modified. Fig. 2.22(II) illustrates the effect of different current and voltage set-points. The effect of a small change in voltage set-point can be seen by comparing curve (d) to curve (e). Both of these curves were calculated with the same current set-point, $I_{set} = 1$, but with differing voltage set-points, $V_{set} = 1.0$ V and $V_{set} = 0.8$ V respectively. The effect of a small change in current set-point can be seen by comparing curve (e) to curve (f). Both of these curves were calculated using the same voltage set-point, $V_{set} = 0.8$ V, but differing current set-points, $I_{set} = 1$ and $I_{set} = 2$ respectively. Even though these curves do not differ in appearance as dramatically as those in Fig. 2.22(I), they still show that small changes in set-point can lead to significant changes in the *IV* curve. Comparing *IV* curves or dI/dV spectra measured with different set-points can lead to misinterpreting data, with the effect of different set-point being mistaken for differences in sample DOS. The safest procedure is to only compare *IV* curves or dI/dV spectra measured at the same set-point.

Although the full *IV* method gathers a very rich data set, this method can be slow, often prohibitively so. For example, a typical 400 point *IV* curve requires at least 100 ms to measure. If the tunneling current is noisy, as it was for many of the samples measured in this work, at every point of a LDOS map, a minimum of 30 to 50 individual *IV* curves must be measured and averaged to achieve an acceptable signal to noise ratio. Between each *IV* curve measurement, because of thermal drift, the feedback should be re-engaged for anywhere from 100 ms to 1 s to reestablish the tip-sample distance and therefore I_{set}. Thus, for a relatively low-resolution 64 by 64 grid-point map, the total measurement time would be more than 9 hours. During the measurement, the temperature must be stable and neither the sample nor the tip should degrade. Additionally, any measurement should be repeated, both on each sample and on others. For samples such as the manganites films measured for this Thesis, where measurements were necessary at a series of temperatures and at a series of magnetic fields at each temperature, measuring LDOS maps with full *IV* curves impractical.

Lock-in conductance maps are made while topography is scanned. The applied bias voltage, V_{bias}, is modulated with a small ac voltage, V_{ac}, at a frequency above the feedback bandwidth. The measured tunneling current will also have an ac component with the same frequency as the bias voltage modulation.[†] The tunneling current is then mixed with V_{ac} using a lock-in amplifier. The output of the lock-in amplifier will be the sample conductance. (Actually the conductance is just the slope of the *IV* curve at V_{bias} and I_{set}.) This is considerably less information than is contained in a full *IV* curve, but, nevertheless, can be used to distinguish variations in LDOS on an inhomogeneous sample. Consider the schematic metallic and insulating *IV* curves shown in Fig. 2.21. At the set point in this illustration, the slopes of the *IV* curves are different and can be used as an indicator of the LDOS. In any actual measurements, neither the shape of the *IV* curve nor the significance of a variation in the slope in known *a priory*. The strategy used in this research was to combine lock-in conductance maps with full *IV* curves taken at selected points. The *IV* curves provide information about the LDOS, while conductance maps provide information about the spatial variation in electronic structure. When measuring conductance maps the scanning speed is typically less than that used when recording topography alone. This is required because of the limitations imposed by the lock-in amplifier. For every pixel measured the measurement time must be at least three times longer than the time constant of the lock-in filter. The time constant is set based upon the need to filter noise and achieve a stable signal. For the measurements in this Thesis, the time constant was between about 0.1 ms to 30 ms, with 1 ms being typical. For a 512 by 512 pixel conductance map, this would result in a scan time of about 26 minutes. This is considerably less than the time to measure a conductance map with full *IV* curves, and the reason that lock-in conductance maps were the preferred technique used in this research.

[†] Because the modulation frequency exceeds the feedback bandwidth, the feedback will not respond to this modulation and will operate as if V_{bias} were a dc signal.

2.8 Scanning Tunneling Spectroscopy Theory

The underlying principle of STM is electrons tunneling between a metal tip and a sample through a vacuum barrier. The great sensitivity afforded by the exponential dependence of tunneling current on the tip-sample distance permits the STM tip to be positioned within Å's of the sample surface. STS similarly makes use of vacuum tunneling as a way to position the tip close to the sample surface but also exploits the properties of tunneling to measure the sample DOS. How tunneling can be used to measure the DOS is the subject of this section.

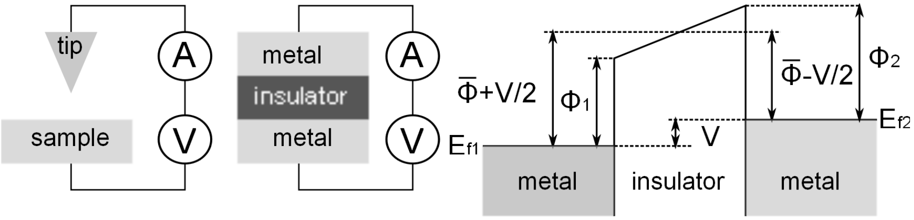

Figure 2.23: Schematic of a NIN planar junction consisting normal metal-insulator-normal metal.

Consider an STM with sample (s) vacuum gap, and tip (t). The current between the tip and sample can be calculated using first-order perturbation theory and is found to be,[9,47,51–53]

$$I = \frac{2\pi e}{\hbar}\sum_{t,s}[f(E_t - eV) - f(E_s)]|M_{t,s}|^2\delta(E_t - E_s), \tag{2.53}$$

with $f(E)$ the Fermi function, V the applied voltage between the electrodes, $M_{t,s}$ the tunneling matrix elements between state Ψ_t of the tip and state Ψ_s of the sample, and E_t and E_s are the energy corresponding to these tip and sample states. All that remains is to calculate the tunneling matrix elements, $M_{t,s}$. By using the wavefunctions of the uncoupled tip and sample and then lowest-order perturbation theory, these can be calculated using,[51,52]

$$M_{t,s} = -\frac{\hbar^2}{2m}\int dS(\Psi_t^*\nabla\Psi_s - \Psi_s^*\nabla\Psi_t), \tag{2.54}$$

with the integral performed entirely within the vacuum between the tip and sample. The electron tunneling between sample (s) and tip (t) can be calculated by assuming the one-dimensional, semiclassical WKB approximation. With this assumption, the tunneling current density is given by, by,[53]

$$J(S,V) \cong \frac{2\pi e}{\hbar}(\frac{\hbar^2}{2m})^2\int_{-\infty}^{\infty} T(S,V,E)[f(E-eV)-f(E)]\rho_s(E)\rho_t(E-eV)dE, \tag{2.55}$$

where $T(S,V,E)$ is the tunneling transmission probability as a function of barrier thickness, S, applied bias voltage, V, and energy, E; $\rho_s(E)$ is the sample density of states and $\rho_t(E-eV)$ is the tip density of states; and $f(E)$ is the Fermi function. At temperatures typical for STM, the Fermi function can be approximated as a step function and so (2.55) can be simplified to,

$$J(S,V) \cong \frac{2\pi e}{\hbar}(\frac{\hbar^2}{2m})^2 \int_0^{eV} T(S,V,E)\rho_s(E)\rho_t(E-eV)dE. \tag{2.56}$$

Assuming a trapezoidal barrier, the tunneling transmission probability can be written in the semiclassical WKB approximation, as,

$$T(S,V,E) \cong exp\bigg\{-2S\bigg(\frac{2m}{\hbar^2}\bigg[\bar{\Phi}+\frac{eV}{2}-E\bigg]\bigg)^{1/2}\bigg\}, \tag{2.57}$$

where $\bar{\Phi}$ is the average of the sample and tip work functions. The next step would be to calculate the tunneling current from (2.56), but as will become apparent, it is easier to extract the sample DOS from the differential conductance. Differentiating (2.56) with respect to bias voltage gives the differential conductance,

$$\begin{aligned}\frac{dI(S,V)}{dV} \cong A\bigg[&eT(S,V,E)\rho_s(E)\rho_t(E-eV)\bigg|_{E=eV} \\ &+\int_0^{eV} T(S,V,E)\rho_s(E)\frac{d\rho_t(E-eV)}{dV}dE \\ &+\int_0^{eV} \frac{dT(S,V,E)}{dV}\rho_s(E)\rho_t(E-eV)dE\bigg],\end{aligned} \tag{2.58}$$

where A is a constant proportional to the effective sample-tip area. But we still cannot solve for $\rho_s(E)$ in (2.58) and further simplifying assumptions must be made. Because the tip is usually chosen to be a metal with a constant DOS E_F the second term in (2.58) can be neglected. From here there are are a number of ways to proceed. Perhaps the most common way is to assume that at low bias voltages $T(S,V,E)$ is constant and the third term in (2.58) can be neglected. Given these assumptions, the tunneling current can be written as,[54,55]

$$I \propto \int_0^{eV} T(E)\rho_s(E)dE, \tag{2.59}$$

and the differential conductance can be written as,

$$\frac{dI}{dV} \approx eT(eV)\rho_s(eV). \tag{2.60}$$

Using these two equations we can calculate the normalized conductance,

$$\frac{dI/dV}{I/V} \approx \frac{eT(eV)\rho_s(eV)}{1/V\int_0^{eV}T(E)\rho_s(E)dE} + \ldots. \tag{2.61}$$

To a first order, the transmission probability can be canceled from the numerator and denominator and this reduces to,

$$\frac{dI/dV}{I/V} \approx \frac{\rho_s(eV)}{1/eV \int_0^{eV} \rho_s(E)dE}. \tag{2.62}$$

In this equation the denominator can be considered a normalization constant, and so,

$$\frac{dI/dV}{I/V} \propto \rho_s(eV). \tag{2.63}$$

Thus the differential conductance normalized by the total conductance is proportional to the sample density of state. In practice, the normalized conductance is calculated from the measured IV curve by dividing the numerical derivative of the tunneling current by the tunneling current at that bias and then multiplying by the bias,

$$\left.\frac{dI/dV}{I/V}\right|_{V_i} \approx \frac{\Delta I_i/\Delta V_i}{I_i/eV_i}. \tag{2.64}$$

This method cannot be used directly on samples with a gapped DOS as the measured current will be zero at non-zero bias voltages and the normalized conductance will diverge. To avoid divergence on insulating or semiconducting samples, the function I/V is smoothed or broadened by convoluted it with a suitable function. One function used for broadening is the exponential function,

$$\overline{I/V} = \int_{-\infty}^{\infty} (I/V')exp\left[\frac{V'-V}{\Delta V}\right]dV'. \tag{2.65}$$

Broadening using this method can distort the normalized conductance if the convolution is too broad. Another issue is caused by noise in the current signal. At small currents, the noise can be of the same order of magnitude as or larger than the signal. Dividing by a small and noisy current leads to an extremely noisy normalized conductance. Averaging multiple curves can alleviate this somewhat, but not completely, whereas curve smoothing can distort the normalized conductance. Because of these issues, noise can be a particular problem if either qualitative or quantitative details of the band gap edge are important.

A second method uses a non-linear fit of the tunneling barrier to extract the sample DOS.[53] The tunneling transmission probability is no longer assumed to be a constant, and thus the third term in (2.58) is not neglected. Substituting $\xi = (E - eV/2)$ into (2.58) and (2.57) will make them symmetric with respect to the tip and sample DOS.

This gives,

$$\begin{aligned}\frac{dI(S,V)}{dV} \cong & \frac{Ae}{2}\Bigg[T'(S,\xi)\rho_s(\xi+eV)\rho_t(\xi-eV)\bigg|_{\xi=eV/2} \\ & + T'(S,\xi)\rho_s(\xi+eV)\rho_t(\xi-eV)\bigg|_{\xi=-eV/2} \\ & + \int_{-eV/2}^{eV/2} T'(S,\xi)\frac{d\rho_s(\xi+eV)}{dV}\rho_t(\xi-eV)d\xi \\ & + \int_{-eV/2}^{eV/2} T'(S,\xi)\rho_s(\xi+eV)\frac{d\rho_t(\xi-eV)}{dV}d\xi\Bigg], \end{aligned} \tag{2.66}$$

and

$$T'(S,\xi) \cong exp\bigg\{-2S\bigg(\frac{2m}{\hbar^2}\Big[\bar{\Phi}-\xi\Big]\bigg)^{1/2}\bigg\}. \tag{2.67}$$

Once again we assume that the tip DOS is constant at E_F, eliminating the fourth term in (2.66). At this point in the analysis, consider only tunneling between two metals. Since the DOS for a metallic sample is also flat then the third term in (2.66) would be eliminated. What remains is an equation that models tunneling between two metallic conductors with flat DOS's. The differential conductance is now,

$$\begin{aligned}\frac{dI(S,V)}{dV} \cong & \frac{Ae}{2}\Bigg[T'(S,\xi)\rho_s(\xi+eV)\rho_t(\xi-eV)\bigg|_{\xi=eV/2} \\ & +T'(S,\xi)\rho_s(\xi+eV)\rho_t(\xi-eV)\bigg|_{\xi=-eV/2}\Bigg], \end{aligned} \tag{2.68}$$

or equivalently,

$$\begin{aligned}\sigma(S,V) \equiv \frac{dI(S,V)}{dV} = & \frac{Ae}{2}\Bigg[exp\bigg\{-2S\bigg(\frac{2m}{\hbar^2}\Big[\bar{\Phi}-eV/2\Big]\bigg)^{1/2}\bigg\} \\ & + exp\bigg\{-2S\bigg(\frac{2m}{\hbar^2}\Big[\bar{\Phi}+eV/2\Big]\bigg)^{1/2}\bigg\}\Bigg]. \end{aligned} \tag{2.69}$$

If we now compare this curve to the differential conductance between a metallic tip and a non-metallic sample, we can take the difference between them as representing the sample DOS scaled by a constant. In practice, a non-linear fit is made of (2.69) to the high-bias part of $\Delta I/\Delta V$, the numerical derivative of the IV curve, with fit parameters A, S, and $\bar{\Phi}$. The deviation of $\Delta I/\Delta V$ from the fitted curve is the sample DOS. At high bias they should be identical, but at low bias, where the sample DOS could be gapped or depleted, there will be a difference. The advantage of this method over the first method described is that the procedure does not involve dividing by the tunneling current and thus avoids dividing by what can be a small number.

Unfortunately, this method is not without problems. As there are three fit parameters and fits to experimental data can lead to widely varying values. Another weakness is that for the procedure to work, the sample DOS must be flat at high bias values where the fit is to be made.