CHAPTER II

Theory of Electron Emission and Scanning Probe Microscopy

The first section covers the theory related to electron emission in presence of electric field as well as temperature effects and in second section the theory and working principles of scanning probe microscopy (STM & AFM) techniques has be discussed in details.
Chapter II

Section I

2.1.1 Introduction

Field emission is a phenomenon that has a vast technological context. From the time this phenomenon has been understood, researchers have found novel applications for utilizing it. The first section of my thesis is on the experimental study of field emission from different form of carbon based nano structured cathodes such as Carbon nanotubes (CNTs), Carbon nanoflakes (CNFs) and Nanodiamonds (NDs) as also the study of thermionic and thermal-field emission from vertical CNTs. This chapter gives as an introduction to thermionic emission, thermal-field emission, and field electron emission and provides the background theoretical knowledge essential to understand the experiments and the analysis. There are different mechanisms of electron emission with the combination of electric field and temperature viz. thermionic emission, field electron emission, and thermal-field emission. A description of thermal emission is also provided along with thermal-field emission since it is relevant for experiments described later in the thesis. This chapter also describes the numerous past and current practical applications for thermionic, field and thermal-field electron emission and also those envisioned for the future.

2.1.2 Electron Emission Theories

Electron emission can be defined as the liberation of electrons from the surface of a material due to external energy transferred to the electrons. This phenomenon is most frequently observed in metals as there are more free electrons which can gain external energy. The minimum energy (usually measured in electron volts) needed to remove an electron from the Fermi level in a metal to a point finite distance away from the surface is called the work function of that surface [1]. There are various mechanisms through which an electron inside a metal can be emitted from its surface. Based on the source of energy for the emitted electron, the mechanisms are classified as photo emission (energy from light), thermionic emission (energy from heat), secondary electron emission (kinetic
energy from another electron) and field emission (energy from electric field). The mechanisms relevant to this research work are of course thermionic, field electron emission and thermal-field emission. A combined thermal-field emission description is employed when emission is due to both a high temperature and under influence of an electric field [2]. These two relevant electron emission mechanism theories will be discussed in detail below.

2.1.3 Field Electron Emission

The mechanism of field emission has no analogue in the other electron emission mechanisms since it is based on the phenomenon of quantum mechanical tunneling. It was observed a long time back in 1897 by Wood [3] but was first explained correctly by R. Fowler and L. Nordheim [4] in 1928. Fowler-Nordheim (F-N) explained that electrons are emitted as they tunnel through a potential barrier that is lowered and narrowed due to presence of intense electric fields and derived the emission current density. Thus, according to the F-N 1-D model, electrons arrive at the surface of a metal, which is assumed at 0°C, according to Fermi-Dirac statistics and penetrate the potential barrier in front of the surface with a probability given by the Schrödinger equation.

![Figure 2.1. 1-D potential energy barrier for an electron near a metal surface.](image-url)
Chapter II

The shape of the potential barrier is described by the electric field and the presence of image charges. Far outside the metal surface, \((z \to \infty)\), in absence of an electric field, the potential energy is chosen to be zero. Inside the metal the electrons are assumed to have a constant effective potential energy \(-W_a\). Then, in presence of an electric field, \(E\), the potential barrier is described by [5]

\[
V(z) = \begin{cases} 
-W_a & \text{where } z < 0, \\
-\frac{e^2}{4z} - eEz & \text{where } z > 0,
\end{cases}
\]

Figure 1.1 shows the one-dimensional potential energy barrier faced by an electron inside the metal, near the surface. The first term in equation 2.2 comes from the inclusion of image charges. Classical image charge correction is good approximation since it is difficult to exactly calculate the electron potential at the surface from the appropriate exchange and correlation energy terms. The supply function of the electrons is taken from the Summerfield’s theory of electrons in a metal and is equal to the number of electrons with energy within the range \(E\) to \(E+dE\) whose \(z\) part of energy lies in the range \(W\) to \(W+dW\), incident on the surface per second per area. Thus the supply function is given by [6]

\[
N(W, E)dWdE = \frac{4\pi m}{\hbar^2} \frac{dWdE}{\exp\left(\frac{(E-W)}{kT}\right)+1}
\]

This supply function is then multiplied by the barrier penetration probability or the transmission coefficient, \(D(W)\), which is defined as the probability for an electron, with \(z\) part of energy equal to \(W\), that will penetrate the potential barrier. This yields the number of electrons within the range \(W\) and \(W + dW\) that emerge from the metal surface per second per unit area. \(D(W)\) can be calculated using the WKB approximation [7]. For \(W << V\text{max}\) (the apex of the potential barrier) and for the emission range \(W \approx \xi\), where \(\xi\) is the Fermi energy, \(D(W)\) is shown to be [6]

\[
D(W) \approx \exp\left[\frac{-c+(W-E)}{d}\right]
\]

Where \(c = \frac{4(2m\phi^3)^\frac{1}{2}}{3\hbar E}v(y)\)

\(d = \frac{\hbar E}{2(2m\phi)^\frac{1}{2}v(y)}\)
Chapter II

\[ y = \left( \frac{e^2E}{\phi} \right)^{1/2} \]  \hspace{1cm} \text{Equation 2.7}

Here, \( \Phi \) is the work function of the metal surface and \( t(y) \) and \( v(y) \) are slowly varying functions. Now, the number of electrons in the given energy range penetrating the barrier is given by \( N(W;E)D(W)dWdE = P(W;E)dWdE \). The total energy distribution, \( P(E)dE \) is then calculated by integrating over the energy range \( E \) to \( -W_a \). This integration is facilitated by setting the limit \( -W_a \) equal to \(-\infty\) to obtain

\[ P(E)dE = \int_{W=-\infty}^{W=-W_a} N(W,E)D(W)dWdE \]  \hspace{1cm} \text{Equation 2.8}

\[ = \frac{4\pi m d}{h^3} \exp \left[ -\frac{c}{d} \right] \times \frac{e^{E/d}}{\exp \left[ \frac{E}{kT} \right] + 1} dE \]  \hspace{1cm} \text{Equation 2.9}

And, finally the total emitted current density is given by \( e \int P(E)dE[6] \). Thus,

\[ J = e \int_{-\infty}^{\infty} P(E)dE = \frac{4\pi m d}{h^3} \exp \left[ -\frac{c}{d} \right] \times \frac{e^{E/d}}{\exp \left[ \frac{E}{kT} \right] + 1} dE \]  \hspace{1cm} \text{Equation 2.10}

After some manipulations this can be put in standard form. The solution is valid only when \( d > kT \). The reduced equation is then written as

\[ J = \frac{e^2E^2}{8\pi \hbar \Phi t^2(y)} \times \exp \left[ -\frac{8\pi(2m)^{1/2}E^{3/2}}{3heE} \frac{1}{\sin(\pi kT/d)} \right] \frac{\pi kT/d}{\sin(\pi kT/d)} \]  \hspace{1cm} \text{Equation 2.11}

For \( T \to 0, \frac{\pi kT/d}{\sin(\pi kT/d)} = 1 \). and so on

\[ J = \frac{Ae^2}{Bt^2(y)} \exp \left[ -\frac{Bv(y)\Phi^3}{E} \right] \]  \hspace{1cm} \text{Equation 2.12}

Where

\[ A = \frac{e^3}{8\pi \hbar} \approx 1.541434 \times 10^{-6} \text{eV}^{-2} \]  \hspace{1cm} \text{Equation 2.13}

And

\[ B = \frac{8\pi \sqrt{2m}}{3he} \approx 6.830890 \text{eV}^{-1/2} \]  \hspace{1cm} \text{Equation 2.14}

Equation 2.12 is known as the standard F-N equation for current density due to cold field electron emission and the constants \( A \) (2.13) and \( B \) (2.14) are known as the first and second F-N constants.

A more generalized equation has been proposed in recent times that include various physical correction factors. In the standard form of F-N equation, the slowly varying functions \( t(y) \) and \( v(y) \) are replaced by their approximate numerical values. This has been shown to under-predict \( J \) values, often by a factor of 100 [8]. Hence, in the general form, the functions \( t^{-2} \) of \( y \) and \( v(y) \) are replaced by parameters, \( \lambda \) and \( \mu \), whose
forms depend on the type of approximation made. The parameter \( \lambda \) includes effects from
the Tunneling pre-factor emerging from the JWKB treatment for calculating transmission
probability \( D(W) \). It also includes temperature effects and electronic band structure
effects. The parameter \( \mu \) contains information of the barrier shape [8]. The F-N equation
has been successfully able to predict emission currents for a very large range of electric
fields and current densities and works surprisingly well at non-zero temperature.
However, this simple equation fails at very large current densities where space charge
effects start to dominate as well as high temperatures and low fields where thermal
emission dominates. The next section describes thermal field emission in more detail.

### 2.1.4 Thermionic Emission

In thermionic emission, (and photoemission) as opposed to field emission, the potential
barrier in not deformed, but the electrons are given sufficient energy to overcome the
barrier. This energy comes from heating the metal until sufficient electrons acquire
kinetic energies \( \geq \Phi + \xi \). The emission current density can be estimated by Richardson's
Law [9] (also known as RLD equation)

\[
J = A_G T^2 \exp \left( \frac{-\Phi}{kT} \right)
\]  

Here, \( A_G = \lambda_R A_0 \) where \( \lambda_R \) is a material specific correction factor and \( A_0 \) is a universal
constant given by

\[
A_0 = \frac{4\pi mk^2 e}{\hbar^3} = 1.20173 \times 10^6 \text{ A/m}^2\text{K}^2
\]  

The derivation of this equation is less complex. The same supply function is used as in
equation 1.3, however the transmission coefficient is determined in the following way: If
the electron's z directed energy, \( W < V_{\text{max}} \) then \( D(W) = 0 \) where as, for \( W > V_{\text{max}} \),
\( D(W) = 1 \). These criteria can be used to easily obtain equation 2.15.

When there is an external electric field applied between the cathode and the
anode, electron emission cannot be explained on the basis of the RLD alone. This is
frequently called field enhanced thermionic emission and in this case the RLD equation is
corrected for the Schottky effect. The lowering of the potential barrier at the surface of a
metal due to presence of an electric field is known as the Schottky effect [10]. This effect
Chapter II

is incorporated by adding image charges outside the metal surface. The “effective” work function is then reduced by an amount \( \Delta \phi = \sqrt{\frac{eE}{4\pi\varepsilon_0}} \) and the current density is then given by

\[
J = A_0 T^2 \exp \left[ -\frac{\phi - \Delta \phi}{kT} \right]
\]

However, even this correction is valid only for electric fields lower than \( 10^8 \) V/m. For higher electric fields, a combined thermal-field emission theory is more appropriate as this does not consider a simplistic transmission coefficient as in case of thermionic emission.

2.1.5 Thermal-Field Emission

The most prominent contribution to a combined theory of thermionic and field electron emission is probably the one given by Murphy and Good in 1956 [5]. They developed a set of equations for thermionic emission regime, field emission regime and an intermediate emission regime. The calculations were based on a general expression for emitted current as a function of temperature, field, and work function, in the form of a definite integral. This general equation is formed using Fermi-Dirac free electron distribution in the metal and classical image charge barrier at the surface. The transmission coefficient, \( D(W) \), is still considered to be 1 for \( W > W_l \), where the limiting value, \( W_l = -\frac{1}{2} \sqrt{2eE} \). Although this is not accurate, it simplified calculations a lot and the results are relatively accurate for the range of applicability. The general emission current equation is then given by [5]

\[
J(E, T, \phi) = e \int_{-W_a}^{W_l} D(E, W) N(T, \phi, W) dW
\]

\[
= \frac{4\pi m k T e}{h^3} \int_{-W_a}^{W_l} \ln[1 + \frac{-(W + \phi)}{kT}] dW
\]

\[
+ \frac{4\pi m k T e}{h^3} \int_{W_l}^{\infty} \ln[1 + \exp[-(W + \phi)/kT]] dW
\]

This equation (2.19) can be made to look better in terms of Hartree units.
Chapter II

The technique for evaluating the integral in equation (2.19) is using different approximations depending on the conditions of temperature and field. Thus, for thermionic emission regime the conditions are given by

\[ \ln \left[ \frac{1-d}{d} \right] - \frac{1}{d(1-d)} > -\pi E F \left( \phi - E F \right) \]  
2.20

\[ \ln \left[ \frac{1-d}{d} \right] - \frac{1}{1-d} > -\pi E F \]  
2.21

Where, \( d = \frac{E F^3}{\pi kT} \)

Now, the approximation used is the first term in an expansion of the logarithm above the Fermi energy and the first term in an expansion of the exponent in the denominator about the peak of the barrier. This leads to an integral which can be evaluated in terms of elementary functions. Without going into all the detailed step, which can be found in reference [5], the final expression for current density due to thermionic emission is given by

\[ J = \frac{1}{2} \left( \frac{\pi}{\lambda} \right)^2 \left( \frac{\pi d}{\sin \pi d} \right) \exp \left[ -\frac{\phi - E F}{\pi kT} \right] (A/m^2) \]  
2.22

This equation (2.22) can be seen to be similar to the RLD equation (2.15) apart from the difference of the Hartree units used for defining energy in place SI units. Hartree unit is a unit of energy defined as \( E_h = \frac{\hbar^2}{m_e a_0^2} \) where \( a_0 \) is the Bohr radius. In parallel with the treatment of thermionic emission, the approximations used for the field emission regime is to use the first term in an expansion of the denominator factor below the peak of the potential barrier and the first two terms in an expansion of the denominator-exponent about Fermi energy. The limits of this approximation and the applicability of the field emission equation are given by

\[ \phi - E F > \frac{E F^3}{\pi} + \frac{kT}{1-ckT} \]  
2.23

\[ 1 - ckT > (2\sqrt{2}E^{-1}\phi^2 \pi)/(\phi^2 - E) \]  
2.24

Where, \( c = 2\sqrt{2}E^{-1}\phi^2 \pi/(\phi^2 - E) \)

and \( f = \frac{1}{2} \sqrt{2}E^{-1}\phi^2 (\phi^2 - E)^{-1} \pi(y) \)
Chapter II

Then, using the above mentioned approximations the current density in the field emission regime is given by

$$ J = \frac{E^2}{16\pi^2\phi T^2(y)} \left( \frac{nckT}{\sin nckT} \right) \exp \left[ \frac{4\sqrt{2} \phi \varepsilon(y)}{3E} \right] \tag{2.25} $$

Again, it can be noticed that this equation in the limit for low temperature is similar to (besides the Hartree units) F-N equation (2.12).

In the intermediate emission regime, which cannot be modeled by either pure field emission or pure thermionic emission, a saddle point approximation is used by Murphy and Good. The conditions of this approximation are

First

$$ \left( \frac{E^2}{\eta} \right)^{-1} > 1 + \frac{E^4}{\pi(d-1)} \tag{2.26} $$

Where

$$ d = 2\sqrt{2}t_\eta \pi^{-1} \left( -\frac{E^2}{\eta} \right)^{\frac{1}{2}} \right) , $$

$$ t_\eta = t \left( -\frac{E^2}{\eta} \right) , $$

$$ \eta = -\frac{E^2}{8(kT)^2} t_\eta^2 \quad \text{and} $$

Second

$$ -\frac{E^2}{8(kt)^2} > -\phi + \frac{kT}{1-E(2\sqrt{2} \phi \varepsilon(y))^{-1}} \tag{2.27} $$

The final expression for emission current density in the intermediate regime is then given by

$$ J = \frac{E}{2\pi} \left( \frac{kt_\eta}{2\pi} \right)^{\frac{1}{2}} \exp \left[ -\frac{\phi}{kT} + \frac{E^2\Omega}{2\pi(kT)^3} \right] \tag{2.28} $$

Where

$$ \Omega = \frac{3}{t_\eta^2} - \frac{2\varepsilon(y)}{t_\eta^3} $$

Thus, the set of equations 2.22, 2.25 and 2.28 together describe combined thermal-field emission of electrons and are frequently called the Murphy-Good (M-G) equations. The bounding region of validity of these equations is shown in figure 2.2 Jensen has published methods to combine thermal and field emission regimes for better accuracy and more range of validity using a method to find best approximated expansion point numerically and then use analytical approximation methods at that point to get the unified solution [2] [11].
The current density equation is calculated analytically by making certain assumptions depending on the potential barrier, temperature or electric field and so forth.

2.1.6 Range of Validity of the Various Emission Theories

The expressions given in the preceding sections provide an almost complete theoretical determination of the emitted current density and of the energy distribution functions for arbitrary values of cathode temperature and applied electric field. The figure 2.3 illustrates schematically the range of validity of the various approximations given above, in the case of a tungsten cathode. As shown, there are three major boundaries. The first boundary CC' corresponds to the condition $q \leq 1$, i.e.

$$F \leq F_1(T) = \left( \pi m^{1/2} kT / \hbar e^{1/4} \right)^{3/4}$$  \hspace{1cm} 2.33

$F_1$ depends on the temperature but not on the work function of the cathode; in practical units,

$$F_1 \approx 1100T^{3/4} (V/cm)$$  \hspace{1cm} 2.34

Where, $F$ and $T$ are in $V/cm$ and $^{0}K$ respectively. As long as the applied field is low enough so that condition (2.33) is satisfied, i.e. below the boundary line CC', the emission
Chapter II

is predominantly thermionic in character. Below the boundary AA’ (i.e. for $F < 0.15 F_1$) Eqs. (A.5) and (A.6) based on the Simple Schottky theory apply to a good approximation, e.g. to within 10% for the total emitted current density.

Figure 2.3. Temperature-Field domains for various electron emission mechanisms.

Between the boundaries AA’ and CC’, the emission will be referred to as the “extended Schottky emission,” and the more general expressions (A.12) to (A.14) must be used; these expressions break down completely for $F \geq F_1$, but appear fairly accurate for $F \leq 0.75 F_1$. The boundary BB’, corresponding to $q = 0.5$ or $F = 0.4 F_1$, is of interest as it separates the region (below) where the larger fraction of the emitted current is contributed by electrons emitted over the top of the barrier ($E > E_0$) from the region where the majority of emitted electrons escape through the potential barrier by the tunnel effect.

The secondary important boundary DD’ corresponds to the condition $p \leq 1$, i.e.

$$F_2$$

depends on both temperature and work function; in practical units,
Chapter II

\[ F_2 = 9.4 \times 10^3 \phi^{1/2} T \text{ (V/cm)} \]  \hspace{1cm} 2.36

Above the boundary DD’ the emission is of a field emission rather than thermionic character. Above boundary GG’, i.e. for \( F > 4.2 \ F_2 \) the equations of field emission apply to a good approximation (e.g. to within 10% for \( J \)), whereas the T-F emission theory, corresponding to equations of T-F emission would be used between boundaries GG’ and DD’; the latter expressions break down completely when \( F \leq F_2 \), but are fairly accurate for \( F \geq 1.3 \ F_2 \). The boundary EE’ (corresponding to \( p = 1/2 \) or \( F = 2F_2 \)) marks the separation between regions where the major fraction of the emitted electrons have initial total energies either above or below the Fermi energy.

Finally, the upper boundary HH’ corresponds to an applied field:

\[ F_3 = \frac{\phi^2}{e^3} \approx 7 \times 10^6 \phi^2 \text{ (V/cm)} \]  \hspace{1cm} 2.37

Above this boundary the field emission or T-F emission expressions do not apply because the top of the potential barrier is reduced below the Fermi energy. This region corresponds to emitted current densities of the order of \( 10^{10} \text{ A/cm}^2 \), and is well beyond the range which can be investigated experimentally at present. There unfortunately exists a gap between the regions of validity of the T-F emission theory and the extended Schottky emission theory; this gap corresponds in Fig. A.1 to the shaded area between boundaries CC’ and DD’. In fact these analytical expressions become inaccurate near these boundaries, and the actual region where an analytical expression has not yet been developed is somewhat wider than the shaded area, extending approximately from 0.75 \( F_1 \) up to 1.3 \( F_2 \) as indicated earlier. To illustrate these considerations, figure 4.2 shows the emitted current density \( J(F) \) for \( \Phi = 4.5 \text{ eV} \) and 4 values of cathode temperature.
Chapter II

Figure 2.4. Estimated current density vs. electric field for four values of cathode temperature; the solid curves \( J_{ES} \) and \( J_{TF} \) are derived from the extended Schottky and T-F theories.

The solid curves \( J_{ES} \) and \( J_{TF} \) are derived from the extended Schottky and T-F theories, which appear accurate respectively to the left of points AA'A" and to the right of points BB'B". Since the actual emitted current density must be a smoothly varying function of F, it is estimated by interpolation in the intervals AB, leading to the dotted portions of the complete \( J(F) \) curves. For each cathode temperature, there is a region where the average total energy of the emitted electrons varies rapidly with the applied field, from a value near the top of the barrier to a value near the Fermi energy [12 Ref. Advances in electronics and electron physics].

Section II: Introduction of Scanning Probe Microscopy (SPM)

A. Scanning Tunneling Microscopy (STM)

2.2.1 Introduction
Chapter II

In the decade since the inventions of the scanning tunneling microscope (STM) in 1983 by Binnig Rohrer [13] and atomic force microscope [14], these instruments have established themselves as the most important techniques in surface investigations. With little sample preparation, very high vertical and lateral resolutions, and ability to work in various environment like vacuum, air and fluid etc., these techniques has made a dramatic impact in fields as diverse as material science, semiconductor physics, biology, electrochemistry, tribology, biochemistry, surface thermodynamics, organic chemistry, lithography etc. The reason for its nearly instantaneous acceptance as a characterization tool is that STM provides three-dimensional, real space images of surfaces at high spatial resolution. When the sample is clean and flat, even atoms can be imaged. The main disadvantage of this technique is that it cannot be used to study non-conducting samples. This disadvantage was overcome after the invention of Atomic Force Microscope (AFM) [14]. In this chapter the working principle of STM and its theory will be discussed. Theory of tunneling spectroscopy will also be looked upon. The later part of the chapter deals with the principle, working of AFM and theory of AFM.

2.2.2 Principle of STM

The STM consists of a sharp metal tip, often made of Pt-Rh or tungsten (W) and a conducting or semiconducting planer sample surface (Fig.2.1). When the tip is brought very close to the sample (within a few Å) and a small bias is applied between the two, tunneling current flows because of the phenomenon of quantum mechanical tunneling. This current has an exponential dependence on the tip-sample separation, a small change in the distance (1 Å), and results into one order of magnitude change in tunnel current, resulting in atomic resolution of surface features. The tip is scanned over the sample in raster pattern by means of piezo scanner and variations in the tunneling current (I) are plotted which are essentially a magnified view of the surface topography.

Mechanism of tunneling can be understood from the one dimensional potential barrier problem in elementary quantum mechanics as shown in Fig. 2.2. State of an electron with energy E moving in a potential $U(z)$ ($E < U(z)$) is described by wave function $\Psi(z)$ which satisfies Schrödinger equation,
Chapter II

\[ -\frac{\hbar^2}{2m} \frac{d^2\psi(z)}{dz^2} + U(z)\psi(z) \quad 2.40 \]

The solution of this equation is

\[ \psi(z) = \psi(0)e^{-kz} \quad 2.41 \]

Where

\[ k = \frac{\sqrt{2m(E-U)}}{\hbar} \quad 2.42 \]

Eq.2.41 describes the state of the electron decaying in the positive z-direction. The probability density of finding an electron across the barrier is \(|\psi\psi^*|\) which is non-zero from eq. 2.41.

Starting from this, the tunneling mechanism in STM can be explained. Consider a metal vacuum metal junction as shown in Fig. 2.41. The width of the potential barrier is decided by the distance between the tip and the sample and the height of the barrier is decided by the work function \(\Phi\) of the materials.

Figure 2.5 A schematic showing principal of STM.
Chapter II

Figure 2.6 tunneling through one dimensional potential barrier.

For simplicity, it is considered that the work function of the tip and the sample are same. If the vacuum level is taken as the reference point of energy then, \( E_f = \Phi \). Electrons can tunnel from tip to sample or sample to tip if a small bias is applied between the two. This current is proportional to the number of states present between \( E \) and \( eV \) (Fig. 2.2). Thus,

\[
I \propto \sum_{E_n = E_f - eV}^{E_f} |\psi_n(0)|^2 e^{-2kz}
\]

Where

\[
k = \frac{\sqrt{2m\Phi}}{h}
\]

Figure 2.7 One dimensional metal-vacuum-metal tunnel junctions. The sample, left, and the tip, right are modeled as semi-infinite pieces of free electron model. When the distance \( z \) is small, the vacuum tail of a sample state can penetrate into the
region of the tip. By applying a bias voltage $eV$, the sample states between energy level $E_f - eV$ and $E_f$ can tunnel to the tip, generating a tunneling current proportional to the bias voltage $V$. Once the atomic configuration of an STM junction is fixed, the tunneling conductance $G=I/V$ is fixed.

The local density of states (LDOS) is defined as the number of electrons per unit volume per unit energy at a given point in space at a given energy. Thus at a location $z$ with energy $E$, LDOS $\rho_s(z,E)$ of a sample is defined as

$$\rho_s(z,E) = \frac{1}{V} \sum_{E_n=E-V}^E |\psi_n(z)|^2$$  \hspace{1cm} (2.45)

From equation 2.43 and 2.45 tunneling current can be written in terms of LDOS of the sample as [15]

$$I \propto V \rho_s(0,E_f)e^{-kz}$$  \hspace{1cm} (2.46)

By substituting the value of $k$ from equation 3.5 and taking typical value of work function as $\Phi = 4 \text{ eV}$ the decay constant comes out to be $k \approx 1 \text{ Å}^{-1}$. This expression shows that 1Å change in $z$ will produce an order of magnitude change in tunnel current ($I$). This is the reason for high vertical resolution in STM. This distance dependence is used in STM to get the information about the topography of the sample. The tunneling current $I$ is also proportional to the LDOS (Eq. 2.46). Thus, the spectroscopic data (I-V curves) gives information about the local density of states of the sample. This will be discussed in details in the next section.

### 2.2.3 Theory of STM

Before the STM was invented lot of work was done on the tunneling spectroscopy of metal-insulator-metal (MIM) junctions [16]. Study on tunneling in MIM junction is useful in understanding the tunneling phenomenon in STM as well as in understanding tunneling spectroscopy (STS). Bardeen’s approach [16] of time dependent perturbation theory in understating MIM is extensively used. In this approach two separate subsystems are considered for tip and the sample as shown in Fig. 2.3. Electronic states of tip and sample are obtained by solving stationary Schrödinger equations. The rate of transfer of an electron from one electrode to another is calculated by time dependent perturbation
Bardeen showed that the amplitude of electron transfer (matrix element $M$) is determined by a surface integral on a separation surface between two electrodes, $z = z_0$,

$$M = \frac{\hbar}{2m} \int_{z=z_0} (\Psi^* \frac{\partial \chi}{\partial z} - \psi^* \frac{\partial \chi}{\partial z})$$  \hspace{1cm} (2.47)

Where $\Psi$ and $\chi$ are wave functions of the two electrodes of the sample and tip respectively. From the Fermi-golden rule the probability of electron in energy level $E$ with energy $E$ ($\Psi$) is given by

$$W = \frac{2\pi}{\hbar} |M|^2 \delta(E_{\Psi} - E\chi)$$  \hspace{1cm} (2.48)

The $\delta$ function in the equation shows that the tunneling will occur between only those states which are equal in both tip and the sample.

Thus, the tunneling current can be evaluated by summing over only those states which are present in both i.e. states between the Fermi levels of the two systems (equal to the applied bias voltage, between 0 and $eV$ Fig. 2.3)

At finite temperature electrons follow Fermi distribution. Thus the total current is

$$I = \frac{4\pi}{h} \int_{-\infty}^{\infty} [f(E_f - eV + \epsilon) - f(E_f + \epsilon)] \rho_s (E_f - eV + \epsilon) \rho_t (E_f + \epsilon) |M|^2 \ d \epsilon$$  \hspace{1cm} (2.49)
Chapter II

Where

$$f(E) = \frac{1}{1 + \exp \left( \frac{E - E_f}{k_B T} \right)}$$

is the Fermi distribution function and \( \rho_s \) and \( \rho_T \) are the density of states of sample and tip respectively.

If \( K_B T \) is smaller than the energy resolution required in the measurement then the Fermi distribution function can be approximated as a step function. And if we assume that the matrix element does not change much in the interval of interest. Then eq. 2.10 reduces to

$$I = \int_0^{eV} \rho_s \rho_t \, dE$$

If the tip is free electron metal tip or if we assume that the tip density of states are constant then

$$\frac{dl}{dV} \propto \rho_s (E_s - eV)$$

Thus one can get information about the sample density of states in the localized area by taking tunneling current versus bias voltage (I-V) characteristics [17].

There are no spatial resolution criteria in STM. The vertical resolution is governed by the stability of the tunnel junction, since height details of the sample surface smaller than the tip to sample vibration amplitudes are usually masked. The lateral resolution is determined by width of tunnel current. The approximate expression for tunnel current \( I \propto V \exp (-2\kappa_0 s) \) can be used to get an order of magnitude estimate of the vertical and horizontal resolution.

2.2.4 Imaging modes

There are two main modes of STM imaging constant height mode and constant current mode.

I. Constant Height Mode

In this mode the tip is scanned at constant height for fixed bias across tunnel junction and tunnel current is monitored for imaging. This method is also called current imaging. This has an obvious advantage that relatively high scanning speeds (1KHz) can be used, enabling real-time video display of the surface [18]. The disadvantage is that the method
can only be used for atomically flat surfaces, where the surface roughness is less than the tunneling gap.

II. Constant current mode

This is the most widely used method for acquiring STM images. In this mode the tip is scanned across the surface at fixed bias voltage, the feedback controller keeps the tunnel current constant by changing the voltage to z-piezo. The feedback voltage of z piezo, which is proportional to z displacement of the tip is recorded as a function of x, y coordinates of the tip and therefore generates a topographical image of the surface. The scan rate in this method is typically low (less than 10 Hz) and is limited by response time of the feedback loop and the inertia of motion.

2.2.5 Tunneling Spectroscopy

One of the most exciting aspects of STM is that in addition to providing topographic information of the surface it has a lot of information about the electronic structure of the surface. STM is capable of providing information about the occupied and unoccupied states of the electron. As discussed earlier the tunneling current can be represented in terms of the density of states of the sample and the tip. Thus if the feedback loop is interrupted and the variation of tunnel current with respect to the applied bias voltage is plotted it will give information about the density of states of the tip and the sample (eq. 2.12). If the tip density of state is flat then the information which one gets from the I-V data represents the density of states of the sample. Compared with the other spectroscopic techniques such as Ultraviolet Photoemission Spectroscopy (UPS), Electron Energy Loss Spectroscopy (EELS) etc STM has a unique advantage that it is a local probe. All the techniques mentioned above provide information averaged over a large area of surface where as the tunneling current can give spectroscopic data with atomic resolution [18]. The main merits of performing tunneling spectroscopy using STM are (1) it is a local probe and can be used to probe the electronic properties ranging from individual ad-atoms [19] on the surface to spatial properties [20]. (2) It can be performed on a
preselected position using the scanning ability. (3) It can be performed under various conditions like UHV, air, organic medium etc. at different temperatures.

B. Atomic Force Microscopy (AFM)

2.3.1 Introduction

STM can be used only for the conducting or semi-conducting surfaces. After the invention of Atomic Force Microscope (AFM) it was possible to look at non-conducting samples [14]. In AFM a tip is attached to a cantilever and brought close to the sample. Interaction force between tip and sample causes deflection in the cantilever. These deflections are detected using appropriate electronics to get the topography of the sample.

Figure 2.9. Schematic of AFM.

Working Principle

The AFM consists of a microscale cantilever with a sharp tip (probe) at its end that is used to scan the specimen surface. The cantilever is typically silicon or silicon nitride with a tip radius of curvature on the order of nanometers. When the tip is brought very close to the sample surface, forces between the tip and the sample lead to a deflection of the cantilever according to Hooke's law.
Chapter II

\[ F = -kX \]  

Where \( F \) is force acting on the cantilever, \( X \) is displacement of the cantilever due to the force acting on it and \( k \) is force constant of the cantilever. A detector measures the cantilever deflection as the tip is scanned over the sample, or the sample is scanned under the tip (Figure 2.9). The measured cantilever deflections allow a computer to generate a map of surface topography.

2.3.2 Theory of AFM

The AFM operates in two basic modes. One is contact or static mode and the other is non-contact or dynamic mode AFM. There are various forces which act on the tip when it is in the interaction region with the sample. The main forces are long range attractive forces, contact and short range repulsive forces, capillary and adhesion forces, electrostatic forces etc. The attractive and repulsive forces between two molecules or atoms are represented by Lennard-Jones potential given by

\[ U(r) = -U_0 \left( \frac{r_0}{z} \right)^{12} - \left( \frac{r_0}{z} \right)^6 \]  

Where \( U_0 \) is energy in balanced condition, \( r_0 \) is interatomic distance and \( z \) is interatomic spacing. Figure 2.10 shows the shape of the potential. Thus when the tip is at larger distance from the sample there is an attractive force acting on it and when the tip is brought closer repulsive force acts on it.
2.3.3 Force versus distance curve

In the repulsive region, the force gradient is high and thus, the sensitivity is more. Since the tip is held very close to the sample the resolution is also good. AFM operating in this region is called contact or static mode AFM. In the contact mode, the cantilever is stationary and brought towards the sample. A typical force-distance curve for the contact mode is shown in Figure 2.11. The initial part of the curve represents the cantilever free of interaction forces. As the tip is brought near the sample the cantilever bends due to the force acting on it. When the force gradient acting on the tip due to the sample becomes greater than the force constant of the cantilever the cantilever jumps to contact with the sample as shown in the Figure 2.11. When the tip is still pushed towards sample contact repulsive force acts on it, which is represented as the linear region in the force-curve. This is the region in which the contact mode AFM is operative. The retracing curve shows hysteresis due to adhesive forces between the tip and the sample and finally, the cantilever snaps back to its equilibrium position.
2.3.4 Modes of operation

Three regimes are identifiable in Figure 2.10: contact, non-contact and tapping. The contact regime is established in the positive force portion; in this region the tip “touches” the surface generating repulsion. This regime is usually used with hard surfaces because there is a high risk of damaging or modifying the surface. In the non contact regime the cantilever oscillates at a frequency close to the resonance. In this case the force between tip and surface is negative, i.e., they are attracted to each other. Although there is no contact with the surface and, consequently, no damage can occur, in the non contact regime the interaction originates only a weak signal; furthermore, under this regime the tip also maps the material deposited onto the surface, for instance water. Therefore, the tapping mode is often preferred. Also in this regime the cantilever oscillates, but the oscillations are wider (resonance frequency) and the tip slightly touches the surface. The tapping mode is good for soft surfaces.

2.3.5 Contact mode AFM
Chapter II

In contact-AFM mode, also known as repulsive mode, an AFM tip makes soft "physical contact" with the sample. The force on the tip is repulsive with a mean value of $10^{-9}$ N. This force is set by pushing the cantilever against the sample surface with a piezoelectric positioning element. In contact mode AFM the deflection of the cantilever is sensed and compared in a DC feedback amplifier to some desired value of deflection. If the measured deflection is different from the desired value the feedback amplifier applies a voltage to the piezo to raise or lower the sample relative to the cantilever to restore the desired value of deflection. The voltage that the feedback amplifier applies to the piezo is a measure of the height of features on the sample surface. It is displayed as a function of the lateral position of the sample.

2.3.6 Noncontact mode AFM

Non-contact AFM (NC-AFM) or dynamic force microscopy (DFM) is one of several vibrating cantilever techniques in which an AFM cantilever is vibrated near the surface of a sample. The spacing between the tip and the sample for NC-AFM is on the order of tens to hundreds of Angstroms. This spacing is indicated on the van der Waals curve of Figure 2.10 as the non-contact regime. NC-AFM is desirable because it provides a means for measuring sample topography with little or no contact between the tip and the sample. Like contact AFM, non-contact AFM can be used to measure the topography of insulators and semiconductors as well as electrical conductors. The total force between the tip and the sample in the non-contact regime is very low, generally about $10^{-12}$ N. This low force is advantageous for studying soft or elastic samples.

In this mode, the cantilever with a sharp tip at its free end is oscillated at or near resonance. The resonance frequency of the cantilever shifts upon approach [21]. The shift in resonance frequency which varies with mean tip-sample separation is tracked using a phase lock loop and used as input to the feedback controller. The sample is then scanned to generate surface topography. This mode is usually referred to as Frequency Modulation Atomic Force Microscopy (FM-AFM) [22]. If the excitation frequency is kept constant, this shift in resonance frequency induces change in amplitude of oscillation. The variation of amplitude with tip sample separation is also used in imaging [21].
called Amplitude Modulation Atomic Force Microscopy (AMAFM). FMAFM is the method of choice for working in Ultra High Vacuum (UHV) conditions while AMAFM is preferred for operation in ambient conditions and in liquids. The theory developed by Martin et al. is useful to understand the concept of NC-AFM.

Consider a freely oscillating cantilever. The distance between the tip and the sample is very large so that the interaction between them can be neglected. In such a case the amplitude of the oscillation of the tip oscillating at the cantilever resonance frequency $\omega$ can be written in the form of Lorentzian:

$$A = \frac{A_0(w_0/w)}{\sqrt{1+Q^2(w/w_0-w_0)}}$$  \hspace{1cm} (2.56)

Where

$$W_0 = c\sqrt{k}$$  \hspace{1cm} (2.57)

is the resonant frequency, $c$ is a function of cantilever mass, $k$ is the spring constant, $A_0$ is the amplitude of resonance, $Q$ is the quality factor. As the tip approaches the sample Van der Waals forces should be taken into account. If the vibration amplitude of the cantilever is small enough forces due to tip-sample interaction causes additional spring type force $f$ on the cantilever. Its derivative $f'$ normal to the surface reduces the spring constant $k$ and shifts the resonant frequency (from eq. 2.57) by

$$w'_0 = c\sqrt{k} - f'$$  \hspace{1cm} (2.58)

Thus, the reduced oscillation amplitude $A'$ can be found out by substituting eq. 2.58 in equation eq. 2.56. In the above discussion the assumption that the amplitude of oscillation is small imposes two conditions. First, there is no energy transfer between the tip and the sample and the frequency shift is solely caused by the integration forces between the tip and the sample. And second is the force gradient $f'$ doesn't excite other modes of vibration of the cantilever. This condition makes sure that the eq. 2.58 is valid. Thus, the resonant frequency shift due to the interaction forces between the tip and the sample causes reduction in amplitude. To calculate the biggest change in the lever amplitude for a given shift in the resonant frequency the steepest portion or the portion with highest slope in the amplitude versus frequency curve is calculated. This occurs at

$$w_m = w_0(1 \pm 1/\sqrt{3Q})$$  \hspace{1cm} (2.59)
Chapter II

This shows that maximum sensitivity i.e. maximum change in the amplitude for a given frequency will be achieved not at the resonant frequency but at the frequency slightly greater than the resonant frequency of the cantilever. A force gradient \( f' \) produces frequency shift \( \Delta \omega = \omega_0 f'/2k \), and the change is amplitude

\[
\Delta A = \frac{(2A_0Q)/3\sqrt{3}k}{f'}
\]  

From the above equation the smallest detectable force gradient is found out by equating \( \Delta A \) to the noise. The main noise sources are optical and thermal. The optical noises are mainly caused due photons of the laser which bounce on the cantilever. Its effect can be neglected. The cantilever vibrates due to the thermal noise with the amplitude \( A_T = \sqrt{2k_B T/k} \). This produces vibration noise given by

\[
N = \sqrt{4k_B TQB/kw_m}
\]

Where \( \omega_m \) is the frequency of oscillation and \( B \) is the bandwidth. Thus the smallest detectable force gradient is given by

\[
f'_m = \frac{1}{A_0} \sqrt{\frac{27k_B T B}{Qw_0}}
\]

From eq. 2.62 several parameters can be optimized in order to get maximum sensitivity like decreasing the cantilever stiffness, increasing the oscillation amplitude of the cantilever, increasing quality factor and increasing the natural resonant frequency of the cantilever.

Later on two modifications were done in the above approach. First the amplitude of oscillation of the cantilever was increased and the cantilever stiffness was increased to 40 N/m. With these two conditions the cantilever is oscillated and approached towards the sample. As the amplitude is very large during each oscillation the cantilever almost touches the sample. This increases the resolution in non-contact mode AFM as the tip goes very close to the sample during each oscillation. Also since the stiffer cantilevers are used this avoids jump to contact instability which happens in contact mode and thus avoids the damage to the sample.

2.3.7 Intermittent-contact AFM
Intermittent-contact atomic force microscopy (IC-AFM) is similar to NC-AFM, except that for IC-AFM the vibrating cantilever tip is brought closer to the sample so that at the bottom of its travel it just barely hits, or "taps," the sample [23]. The ICAFM operating region is indicated on the van der Waals curve in Figure 2.10. As for NC-AFM, for IC-AFM the cantilever's oscillation amplitude changes in response to tip-to-sample spacing. An image representing surface topography is obtained by monitoring these changes. Some samples are best handled using IC-AFM instead of contact or non-contact AFM. IC-AFM is less likely to damage the sample than contact AFM because it eliminates lateral forces (friction or drag) between the tip and the sample [24]. In general, it has been found that IC-AFM is more effective than NC-AFM for imaging larger scan sizes that may include greater variation in sample topography. IC-AFM has become an important AFM technique since it overcomes some of the limitations of both contact and non-contact AFM.

### 2.4 Conducting Atomic Force Microscopy (C-AFM)

The operations of Conductive AFM and STM are identical in which a bias voltage is applied between the tip and the sample. The technique differs from STM in that the probe is positioned using normal force feedback, which decouples probe positioning from the sample conductivity and facilitates interpretation of the tip location with respect to the sample (i.e., in contact or out of contact). In C-AFM mode [25-29], a conductive AFM tip scans the surface while in contact. C-AFM is able to image both the topography and the conductivity of the surface at the same time. The current flows between the tip and the sample allowing Conductive AFM mode to measure the surface conductivity of a sample. Contact topography image is generated by using feedback loop to maintain the constant tip deflection and C-AFM image is generated by measuring the current flow. The C-AFM can be operated in the imaging mode and spectroscopic mode.
Chapter II

2.5 References

9. O. W. Richardson, the emission of electricity from hot bodies. Longmans, Green and co., 2 ed., (1921).
Chapter II