3 **Fundamentals of STM and STS**

3.1 **Scanning tunneling microscopy (STM)**

The scanning tunneling microscope was developed by Binnig, Rohrer, Gerber and Weibel [54]. Since STM allows imaging on an atomic scale, it belongs to the most powerful experimental techniques of surface science.

In STM, a small probe (usually an atomically sharp metallic tip) is placed very close to the surface (0.3-2 nm) and a small bias voltage $U$ is applied between the tip and the sample (see Figure 3.1). As a result, a current of electrons $I$ flows between the electrodes through the vacuum gap. This process is a quantum mechanical phenomenon and has been called the “tunneling” effect. The tunneling current is usually the main value measured in the STM experiment and the basic of its theoretical description is given in the next chapter. In the imaging process, the tip scans over the surface and a regulation system attempts to keep the current constant by varying the tip/sample distance ($z$). This mode of imaging is mostly used and is called “constant current topography” mode (CCT). The $z$-regulation signal is recorded during sample scanning. Usually, it is represented in form of a grey level image. Here, the darker tone corresponds to the scanner stretch and the lighter one to its retraction. Thus, information on the surface is obtained in real space.

The tunneling current strongly depends not only on the topography but also on the electronic structure of the surface. Therefore, surfaces with different chemical composition (as for example islands of oxide films on metals) may provide different contrasts in STM. This may complicate the interpretation of the STM images. On the other hand this “chemical” contrast may yield valuable information on the local sample composition. An exact interpretation of the chemical contrast, however, presupposes that additional information on the electronic structure is available.

Information on the electronic structure can be obtained by applying STS. Varying one of the tunneling parameters ($U$, $I$ or $z$), information on the tunneling barrier can be extracted from the STS data. The corresponding spectroscopic techniques are described in more detail in Chapter 3.3.
3.2 The tunneling effect

A theory which describes the tunneling process is generally restricted to elastic tunneling processes. This means that tip/sample interactions which create phonons, plasmons and other losses are ignored. Also some simplification are usually made. For example, the wave functions of the tip is reduced to s-state wave functions.

In general, there are two approaches to derive an equation for the tunneling current in STM. One is based on the scattering theory and is able to yield exact solutions for certain tunneling conditions [55]. The main advantage of this approach consists in the acceptance of different tunneling distances and correspondingly different tip/sample interactions. Usually, a modified Bardeen approach [56] is used which is based on perturbation theory.

The first theoretical description of the tunneling process in STM was presented in 1985 by Tersoff and Hamann [57] which has been applied by many authors in numerous publications. In this theory, the wave functions of the tip and the sample are considered as separated, undisturbed systems and described by means of the Schrödinger equation

\[(T + U)\Psi_\mu = E_\mu \Psi_\mu.\]  (3.1)

Here, \(T\) corresponds to the kinetic energy, \(U\) to the potential energy, \(\Psi_\mu\) to the wave function, and \(E_\mu\) corresponds to the energy of the state \(\mu\). The wave function of the complete system is determined by means of time dependent perturbation theory. For this, the potential of the tip is described as a perturbation \(U_T(t)\) and inserted into a time dependent Schrödinger equation which yields

\[i\hbar \frac{\partial \Psi_\mu}{\partial t} = (T + U_S + U_T(t))\Psi_\mu.\]  (3.2)

Here, \(U_S\) is the potential of the sample and \(\hbar\) the Plank constant.

Using

\[\Psi(x, t) = \Psi(x) \exp\left(-iEt/\hbar\right)\]  (3.3)

as an approach for the wave function, the transition probability \(w_{\mu\nu}\) of an electron tunneling from the state \(\mu\) with the wave function \(\Psi_\mu\) into the state \(\nu\) with the wave function \(\Psi_\nu\) can be written as:

\[w_{\mu\nu} = \frac{2\pi}{\hbar} |M_{\mu\nu}|^2 \delta(E_\nu - E_\mu).\]  (3.4)

Here, \(M_{\mu\nu}\) corresponds to the matrix element of the transition probability. Bardeen [56] has shown that the matrix element can be represented by a surface integral

\[M_{\mu\nu} = \frac{\hbar^2}{2m} \int d\hat{S} (\Psi_\nu^* \nabla \Psi_\mu - \Psi_\mu \nabla \Psi_\nu^*).\]  (3.5)
Here, $m$ is the mass of the electron and $dS$ the surface element. The tunneling current is calculated by summing up over all possible states. Using the density of states $\rho$ of the sample and of the tip, one finally obtains for the tunneling current

$$I = \frac{4\pi e^2}{\hbar} U_b \rho_S(E_F) \rho_T(E_F) |M|^2$$

(3.6)

which is valid for small bias voltages $U_b$. Here, $e$ is the electron charge and $E_F$ the corresponding Fermi level.

Assuming that only s-like wave functions contribute to the tunneling from a tip of radius $R$, the DOS of the sample can be written as

$$\rho_S(r, E_F) = |\Psi_S(r)|^2 - e^{-2k(R + d_e)}$$

(3.7)

where

$$k = \frac{\sqrt{2m\phi}}{\hbar}$$

(3.8)

is the decay coefficient of the wave function into the vacuum and $\phi$ the height of the energy barrier. Assuming a point-like tip and summing up equations (3.6) and (3.7) finally yields

$$I = U_b \rho_S(E_F) e^{\frac{2m\phi}{\hbar e}}.$$  

(3.9)

According to (3.9), the tunneling current $I$ near $E_F$ is proportional to the bias voltage $U_b$ and the density of states $\rho_S$ of the sample. More important, however, is the exponential dependence on the distance $z$ between the tip and the sample which is the main reason for the high sensitivity of the STM perpendicular to the surface.
3.3 Scanning tunneling spectroscopy (STS)

STS provides valuable information which is complementary to the information obtained in the conventional STM. STS can be accomplished in different ways. The main idea of all STS experiments is to measure one of the tunneling parameters \( I \) or \( z \) whereas one of the others (usually \( U \) or \( z \)) is changed (e.g., \( I(U), I(z), z(U) \)). The STS implementation can vary depending on the energy range accessed, the amount of spectroscopic detail required, and which degree of spatial resolution is simultaneously needed.

3.3.1 \( I(U) \) spectroscopy

Most of the STS investigations are performed under the condition of a constant tip/sample separation, which is accomplished by shortly interrupting the feedback controller. Then, the bias voltage is ramped over the desired interval and the tunneling current \( I(U) \) is measured during the voltage ramp. For correlating the tunneling spectra to the local topography (and thus to take full advantage of the high spatial resolution of STM), the \( I(U) \) measurement must be performed pixel for pixel during the topography measurement. \( I(U) \) STS is the most used technique to measure the surface electronic structure. However, the interpretation of the STS data is complicated by two main facts. First, the DOS of the tip is usually unknown. Second, the voltage dependence of the tunneling probability \( T(E,U_b) \) is usually also unknown. The first problem is typically investigated by comparing measurements taken at identical surface locations and by ensuring that all results are reproducible using different tips. Although the electronic structure of the tip is unknown, it is at least constant, for stable STS conditions. Thus, in the tunneling spectra obtained at different locations, the tip electronic structure will only contribute a constant background to the measurement. Consequently, the spatially dependent variations in the electronic structure will not be influenced by the tip states. The second problem, i.e., the unknown \( T(E,U_b) \) dependence, is managed in different ways depending on the details of the situation. In some cases, the \( T(E,U_b) \) dependence on the tunneling probability can be minimized by a special data representation, i.e., by plotting \((dI/dU)/(I/U)\) vs. \( U \) (or equally \( d(ln I)/d(ln U) \) vs. \( U \)). This becomes more obvious if an approximate expression of the tunneling current is used [58].

\[
I = \int_{0}^{eU} \rho(E)T(E, eU_b)dE
\]  
(3.10)

Here, \( T(E,U_b)=\exp(-2kz) \) is the tunneling probability of the electron (see also equations (3.7) and (3.8)). After differentiation, one obtains

\[
\frac{dI}{dU} = e\rho(eU_b)T(eU, eU) + e \int_{0}^{eU} \rho(E)\frac{d}{d(eU)}[T(E, eU)]dE.
\]  
(3.11)

From equation (3.11) one finally finds
Since $T(E, eU)$ and $T(eU, eU)$ appear as ratios in this equation, their exponential dependencies on $z$ and $U$ cancel out approximately. The first term in the numerator of equation (3.12) is the surface DOS of the sample. The denominator of the equation can be considered as a “normalization” function of the surface DOS. The second term in the numerator is a “background” term arising from the fact that the tails of the wave-functions are affected by the value of the electric field in the tip/vacuum/sample junction. Thus, a change of the voltage changes also the electric field [58]. The relative magnitude of the various terms depend on the polarity of the voltage. For a positively biased sample ($U>0$), $T(E, eU) \leq T(eU, eU)$ and the maximum transmission occurs at $E=eU$. In this case, each term in equation (3.12) is of the same order of magnitude, and one obtains a “normalized” measure of DOS together with a slowly varying background. If the sample is negative ($U<0$), $T(E, eU) \geq T(eU, eU)$ the maximum transmission occurs at $E=0$. Then, the background term has the same order of magnitude as the denominator. However, they are both larger than the DOS term by a factor of $(T(0, eU))/(T(eU, eU))$. The measure of DOS is, therefore, reduced by this factor, which can be quite large for large $z$ and $U$. The reason for this reduction is simply that the transmission probability is strongly increased for states near the Fermi level of the negatively biased electrode. Thus, it should be difficult to observe low-lying occupied surface states with STS.

A normalization of the differential conductance $(dI/dU)$ to the total conductance $(I/U)$ provides a very conventional measure of the surface DOS for the case of metallic or small-band-gap surfaces [58],[59]. However, for large-band-gap surfaces, the measure of $(dI/dU)/(I/U)$ diverges at the band edges, since the current approaches zero faster than the conductance [58]. Inside the gap, $(dI/dU)/(I/U)$ is undetermined because $I=0$. In the case of thin insulating films, however, the substrate states exponentially decay through the film, and a finite density of states remains at the insulator surface [60]. Thus, the tunneling current inside the band gap remains finite and the normalization problems of the bulk insulators are avoided.

A model of tunneling through a thin-film insulator (based on established approaches for planar tunneling) was developed by Viernow et al. [60]. In this model the tunneling current is expressed as in equation (3.10) with the assumption that the DOS of tip and sample are unaffected by each other. This reduced the problem to the calculation of the transmission probability $T(E, U)$, which consists of the $T_{\text{vac}}(E, U)$ term for the vacuum barrier and the $T_{\text{ins}}(E, U)$ term for the insulator film

$$T(E, U) = T_{\text{vac}}(E, U)T_{\text{ins}}(E, U), \quad (3.13)$$
where

\[
T_{\text{vac}}(E, U) = \exp[-2\frac{\sqrt{2m(\Phi - (1/2)eU - E)}}{\hbar} z],
\]

and

\[
T_{\text{ins}}(E, U) = \begin{cases} 
\exp[-2\frac{\sqrt{2m(E_{\text{CBM}} - eU - E)}}{\hbar} d_{\text{ins}}} & \text{for } eU + E < E_{\text{CBM}} \\
1 & \text{for } eU + E > E_{\text{CBM}}
\end{cases}
\]

Here, \(\Phi\) represents the average work function between tip and insulator, and \((\Phi - (1/2)eU)\) is the average barrier across the vacuum gap, \(E_{\text{CBM}}\) is the energy of the conduction band minimum of the insulator relative to its Fermi level, \(z\) is the distance of the tunneling tip from the insulator surface, and \(d_{\text{ins}}\) the thickness of the insulator film. Any voltage drop across the insulator is neglected.

Using this model, Viernow et al. [60] were able to calculate the \((dI/dU)/(I/U)\) spectra and fit them to the experimental data adjusting only the \(E_{\text{CBM}}\) term. The authors have also shown that the calculated spectra are insensitive to the tip distance, but they depend on the film thickness.

This model was also used in the present work and will be applied when the STS data analysis will be presented.

### 3.3.2 \(I(z)\) spectroscopy

For the simple one-dimensional model, a tunneling current can be written as in (3.9). Within this model, CCT images can be interpreted as contour maps of the constant DOS. The quantity \(\rho_s\) depends on the local chemical nature of the surface. Therefore, its measurement allows to distinguish between different surface elements, which can be difficult to perform in a CCT mode.

From equation (3.9), one can evaluate an apparent barrier height \(\Phi_{\text{ap}}\) from

\[
\Phi_{\text{ap}} = \left[ \frac{1}{1.025} \cdot \frac{d\ln(I/U_b)}{dz} \right]^2.
\]

This quantity has been generally accepted as an experimental measure characterizing the chemical nature of the sample surface.

There are several possibilities to use \(I(z)\) STS to determine the apparent barrier height. Most known is the measurement of the tunneling current during the tip movement towards the surface. For such a measurement, the tip has to be stabilised. This happens in a constant current mode. Hence, the zero point will correspond to the original tunneling settings. After that, the feedback is deactivate and the tip is moved towards the surface by \(\Delta z\) (tip displacement, usually in the nm range). Simultaneously, the current signal is recorded. After that, the obtained spectra can be evaluated according to formula (3.16). In addition, a lock-in technique can be used to measure \(dI/dz\) which yields tunneling barrier images [61]. An alternative method for obtaining tunneling barrier images is the measurement of the tunneling conductance vs. the applied bias voltage as has been demonstrated by Olesen et al. [62].
3.3.3 \( z(U) \) spectroscopy

Using STM in the field emission regime (in which the applied bias voltage exceeds the averaged work function) makes the energy region above the vacuum level accessible. For this region, Gundlach [63] has calculated both the transmission coefficient and the tunneling current as a function of the applied bias voltage. For metal/insulator/metal systems, it was found that the calculated tunneling current can have superimposed oscillations. This oscillations can be explained by considering the large electron mean free path in the dielectric layer and the specular reflection at the barrier boundaries. Nowadays it is known that the electrons at the Fermi level (which dominantly contribute to the tunneling current) will be partly reflected at the surface and then reflected again at the rising potential within the tunneling barrier. Thus, the electron standing waves can be excited in the vacuum gap in front of the sample surface (see Figure 3.2a). The condition for standing wave formation depend on the electronic potential in the gap, the energy of the field-emitted electrons, and the electron reflectance at the surface. Different surfaces with their own electronic structures will have different reflectances. Hence, the analysis of the distance/voltage characteristics allows a characterization of different surfaces and an estimation of tunneling barriers, respectively.

An approximate description of the experimentally observed high-order resonance states was presented by Coombs and Gimzewski [64]. The resonance condition is fulfilled if an integer number of half wavelengths of the electron wave function will fit between the sample and the rising potential up to the classical turning point \( Z_T \) (see Figure 3.2a) if tunneling from the tip to the sample is regarded. The magnitude of the electron wave vector \( K(z) \) at a point \( z \) between tip and sample is given by

\[
K(z) = \frac{2\pi(2me)^{1/2}}{h} \cdot \left( \frac{U}{z_0} \cdot (z - \phi) \right)^{1/2}.
\]  

Here, \( z_0 \) is the tip/surface separation and \( \phi \) the work function. Integration between the classical turning point \( z=Z_T \) and the sample surface \( z=z_0 \) gives a total phase change of \( \pi n \), where \( n \) is the order of the resonance:
Scanning tunneling spectroscopy

\[ \pi n = \int \left( \frac{(2me)^{1/2}}{\hbar} \cdot \left( \frac{U}{z_0} \cdot z - \phi \right)^{1/2} \right) dz, \]  
(3.18)

or

\[ \pi n = 2\pi \left( \frac{2me}{\hbar} \right)^{1/2} \cdot \frac{2z_0}{3U} \cdot (U - \phi)^{3/2} \]  
(3.19)

where \( U_n \) is the voltage of the \( n \)th state, \( F \) is the field strength in volts per Ångström (Å) and \( \alpha = 4.39 \) V\(^{1/3}\) Å\(^{2/3}\) is a constant. This equation is almost the same as that obtained by Gundlach [63] except that due to his exact inclusion of the boundary conditions, a factor of \( (n-1/4)^{2/3} \) is obtained instead of \( n^{2/3} \). Image forces are neglected here. They will distort the barrier mainly in the regions near to the boundaries and will affect only oscillations with a small state number \( (n) \).

Measurements of the separation/voltage characteristics \( z(U) \) permit to study the oscillatory behaviour of the tunneling current in the near-field regime. The oscillations itself can be directly observed in \( dI/dU \) spectra measured by means of a lock-in technique simultaneously with \( z(U) \) curves as shown in [64]. But one has to consider that in the CCT mode

\[ \left( \frac{\partial I}{\partial z} \right)_z \frac{dz}{dU} + \left( \frac{\partial I}{\partial U} \right)_z = 0, \]  
(3.21)

and since \( ((\partial I)/(\partial s))_U \) is a slowly varying function, \( ((\partial I)/(\partial U))_z = -dz/dU \). Therefore, \( ((\partial I)/(\partial U))_z \) will be mirrored by \( dz/dU \) and steps in \( z(U) \) will be transformed into peaks in \( dz/dU \) (see Figure 3.2b) and \( ((\partial I)/(\partial U))_z = dI/dU \) [65].

In contrast to the above described \( I(z) \) method, the feedback should be kept active during the \( z(U) \) spectra acquisition for keeping the tunneling current constant. If the applied bias voltage increases, the tip will be withdrawn from the sample by the feedback loop to maintain a constant current. The execution of this condition is very important if the STM is used in the near-field emission regime. This is not only necessary for suppressing the electric field strength in vacuum and for preventing a tip crash at the sample surface. It is also necessary for supplying a nearly constant electric field over the sample surface in a vacuum gap, which guarantees defined measurement conditions.