Chapter 1

Theory of the STM

Introduction

Since the invention of the scanning tunneling microscope (STM) in 1983 by Binnig and Rohrer [1] and atomic force microscope nearly two decades ago, these instruments have established themselves as the most important techniques in surface investigations. Moreover, they have been used to fabricate surface structures ranging from ~ 100 nm down to the atomic dimensions. In this paragraph the principal way of working, the underlying theory and aspects of resolution and tip condition are discussed.

In the technique of STM a sharp metal tip is brought very close ($< 10\,\text{Å}$) to a conducting surface (*Fig.1*). When a bias voltage is placed across the tip – sample junction, electrons quantum mechanically tunnel across the gap and produce a measurable tunneling current (typically from 10pA to 10nA). This current has an exponential dependence on the tip – sample separation, resulting in atomic resolution of surface features.

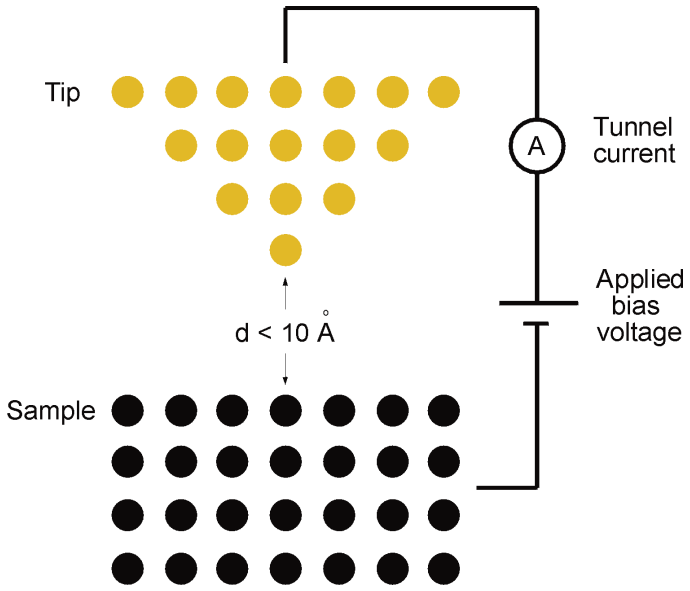

Fig.1

Scheme of an STM.

During image acquisition, the tip scans across the sample using *xy* piezoelectric elements (the tip can be moved in all dimensions with a precision better than 0.01 Å), and a feedback loop adjusts the tip height (*z* piezoelectric element) in order to maintain a constant current.

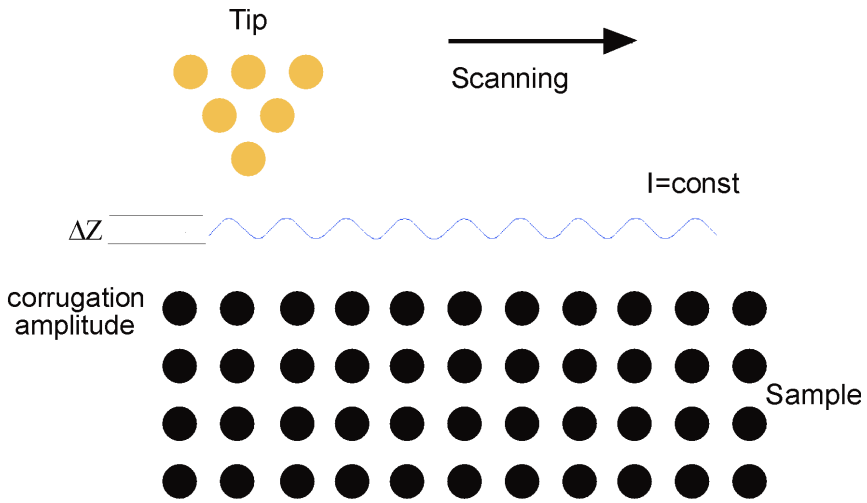

Fig.2

Schematic presentation of the constant-current mode.

Then the tip height signal is displayed resulting in an STM image, which contains both topographic and electronic information.

The corrugation amplitude measured in STM is a quantity which is defined as the difference between the largest and smallest tip-sample distances in a constant-current experiment. Due to the exponential dependence of the tunneling current on the width of the barrier, i.e. of the tip-sample distance, such an experimental setup allows a high resolution vertical to the surface. Combined with the high accuracy of the positioning of the tip parallel to the surface, images with a corrugation amplitude smaller than 0.01 Å can be obtained. Steps and islands are mapped easily and if the experimental setup is stable enough one can achieve atomic

resolution. In images with atomic resolution single atoms in the surface topmost layer are observed.

As a first approximation a single electron approach to the tunneling problem will be used. In the simplest 1D model the electron with mass m is considered to be incident on a rectangular potential barrier (*Fig.3*).

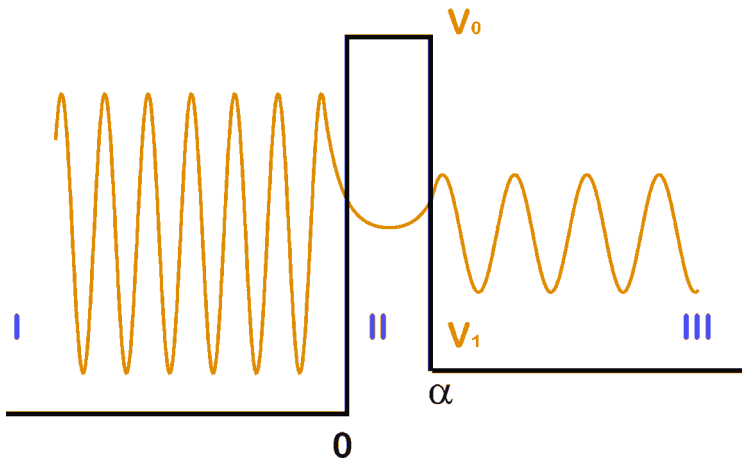

Fig.3

Wave function tunneling through 1D rectangular potential barrier of width α.

One needs to calculate the probability for the electron to penetrate from the right side of the barrier. For this the Schrödinger equation needs to be solved:

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+V(x)\right)\Psi(x)=E\Psi(x) \tag{1}$$

The wave function solutions can be written for the regions I, II and III for the energy of the electron E lower than the barrier potential V_0 :

$$\Psi_I = e^{ikx}+\mathrm{Re}^{-ikx}\,, \qquad k=\sqrt{\frac{2mE}{\hbar^2}} \tag{2}$$

$$\Psi_{II} = Ae^{\omega x}+Be^{-\omega x}, \qquad \omega=\sqrt{\frac{2m(V_0-E)}{\hbar^2}} \tag{3}$$

$$\Psi_{III} = Te^{iqx}, \qquad q = \sqrt{\frac{2m(E - V_1)}{\hbar^2}} \tag{4}$$

The coefficients R, T, A, B are determined from the boundary conditions and the requirements that the wave function and the derivatives should be continuous.

One can define the current density (flux) as

$$j = \frac{\hbar}{2im}\left(\Psi^* \frac{\partial \Psi}{\partial x} - \frac{\partial \Psi^*}{\partial x}\Psi\right) \tag{5}$$

and considering $D = \frac{j_{III}}{j_I}$ the calculations lead to

$$D = \frac{(2\omega)^2 kq}{\omega^2 (k+q)^2 + (\omega^2 + k^2)(\omega^2 + q^2)\sinh \omega\alpha} \tag{6}$$

When the model is used to describe the STM tunnel junction, the barrier height is of the order of the work function and the barrier width corresponds to the tip-sample distance. For a barrier of typically 4eV height and $5\,\text{Å}$ width, one finds that the sinh term dominates, leading to $D \approx e^{-2\omega\alpha}$. This means that the transmission is very sensitive to the barrier width and height. The exponential dependence on the barrier width is common to all tunneling problems (for small enough width).

The exponent is a measure for the apparent barrier height and can be determined with STM by recording the tunneling current as a function of tip – sample distance. Measurements of the apparent height e.g. on the Au (110) surface can be found in [2]. Spatial variations of the apparent barrier height on steps, defects, and adsorbates are of current research interest. Here we present I-Z spectroscopy data taken with a low temperature STM. The measurements were performed on Ni (111) and Ag (111) surfaces. These measurements revealed for the first time a discontinuous change of the apparent barrier height in vacuum tunneling junction on Ni

(111) surface. Both Ni (111) and Ag (111) surfaces were cleaned with sputter and anneal cycles before transferring them in the low temperature STM. The measurements were taken at 5 K. On the Ni (111) surface mechanically cut Au and $Pt_{.8}Ir_{.2}$ tips were used. The tip apex was formed in situ by the repeatingly touching the surface with 1.5 V until a sharp tip was created. It turned out that in both cases material was transferred to the surface. From this one can conclude that the measurements were done with metallic Au or PtIr tips. On the Ag (111) surface only PtIr tips were used.

These measurements were performed by choosing an initial height and then retracting the tip while recording the current with a constant voltage. Then the voltage was varied and another spectrum was taken, and so on. The initial height at each spectrum was either the same or determined by the chosen voltage and the set current. From the I-Z spectra the slope $\kappa = -\frac{1}{2}\frac{\partial \ln I}{\partial z}$ was determined (see *Fig.4*).

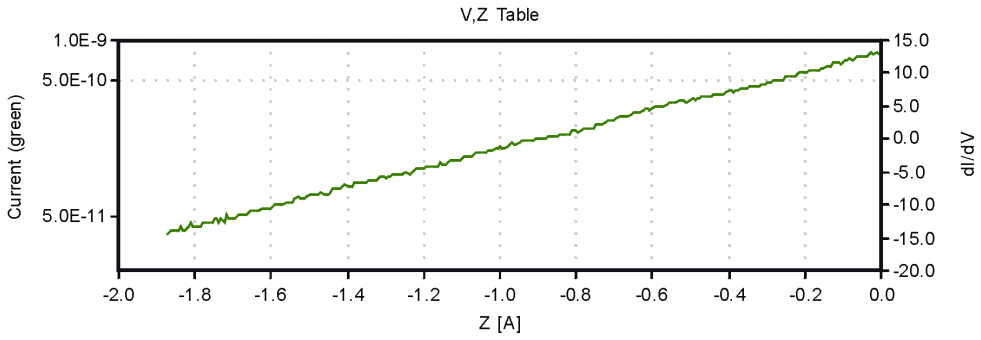

Fig.4

I-Z spectrum taken on a Ag (111) surface with a PtIr tip at U=50 mV.

The slope κ as a function of energy on Ni (111) surface is shown in *Fig.5*. In the spectrum one can see a discontinuous change across the Fermi energy. The gap is observed in both cases: in the case of the constant initial height measurements and voltage dependent initial height measurements.

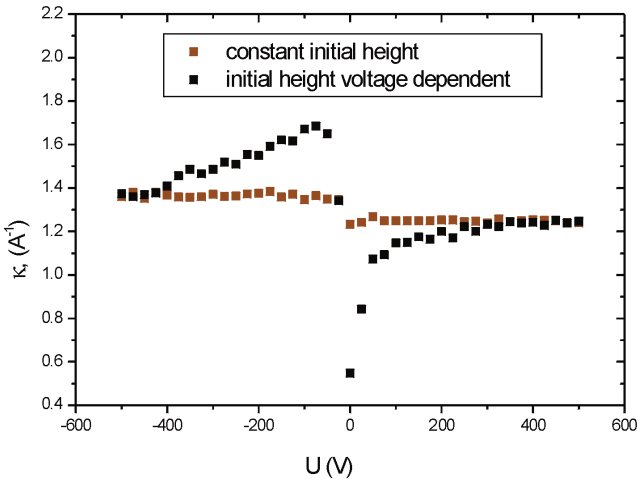

Fig.5

κ as a function of voltage on Ni (111) surface measured with an Au tip.

From *Fig.5* one can conclude that the gap size depends on the initial height, but a clear tendency could not be observed reproducibly (further experimental data needs to be taken). Such kind of a gap was observed in most cases on Ni (111) surface but never on Ag (111) surface. The gap is sensitive to the tip apex, in some cases even an inverse gap was observed.

Fig.6 shows a number of results taken on the Ni (111) and Ag (111) surface with different tips and different tip materials over a wide energy range. At high voltages field emission resonances can be observed. There is a small signature at the Fermi level also in Ag (111) measurements, but this is due to the low current which can not be resolved properly by the A/D converter with this gain setting. The gap has been observed reproducibly with the different tips and on different sample locations. It is also possible to observe it on dI/dV spectra, although it is difficult to resolve it since it is only a 5-10% effect.

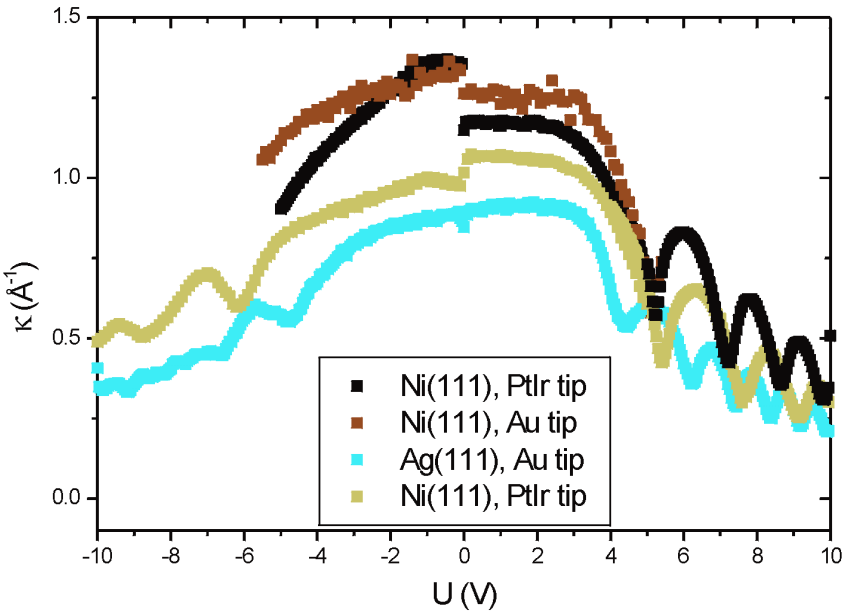

Fig.6

κ as a function of voltage measured on Ni (111) and Ag (111) surfaces with PtIr and Au tips.

To conclude the measured values for κ are within the expected range and show the expected voltage dependence. The discontinuity at the Fermi level on the Ni surface has to be considered as unique and requires a detailed theoretical analysis.

In this section a general formula for elastic tunneling will be derived. A well known result from first-order time-dependent perturbation theory is Fermi`s golden rule, which states that the transmission rate from the initial state $|i\rangle$ to a final state $|f\rangle$ is given by

$$R_{i\to f} = \frac{2\pi}{\hbar}\left|M_{fi}\right|^2 \delta\left(\mathrm{E}_i - \mathrm{E}_f\right) \tag{7}$$

Here M_{fi} is the matrix element of the perturbation potential between the initial and final states, the δ-function ensures energy conservation.

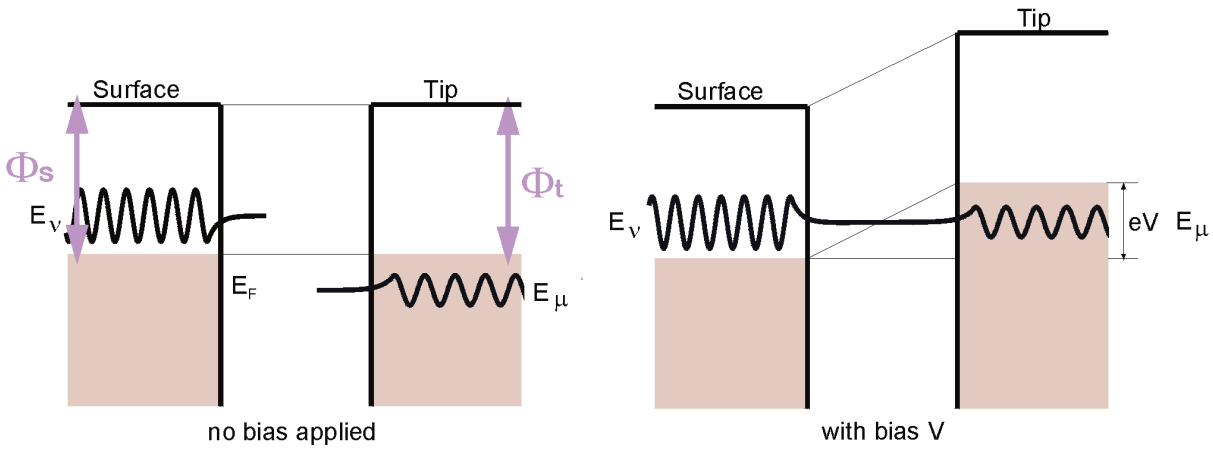

Fig.7

Schematic tip-sample energetics. Energy scheme of tip and surface for the case of a positive voltage applied to the surface.

According to Fermi's golden rule one can write the current flowing from the tip to the sample and correspondingly from the sample to the tip as:

$$I_{t\to s} = \frac{2\pi e}{\hbar}\int |M_{ts}|^2 N_t(E-eV)N_s(E)f_t(E-eV)[1-f_s(E)]dE \tag{8}$$

$$I_{s\to t} = \frac{2\pi e}{\hbar}\int |M_{ts}|^2 N_t(E-eV)N_s(E)[1-f_t(E-eV)]f_s(E)dE \tag{9}$$

Here $f(E)=[\exp((E-E_F)/k_BT)+1]^{-1}$ denotes the Fermi-Dirac distribution, $N(E)$ denotes the density of states. The occupied states (N_t, f_t) can tunnel into unoccupied sample states $(N_s[1-f_s])$. The common energy scale is chosen to be that of the sample, and thus the tip energy scale is shifted by $\Delta E = eV$.

The total current flowing from the tip to the sample is:

$$I = I_{t\to s} - I_{s\to t} = -\frac{2\pi e}{\hbar}\int |M_{st}|^2 N_t(E-eV)N_s(E)[f_s(E)-f_t(E-eV)]dE \tag{10}$$

At T=0 the Fermi-Dirac distribution is a step function and one can write the total current flowing from the tip to the sample as:

$$I(T=0) \approx \int_{E_F}^{E_F+eV} |M_{ts}|^2 N_t(E-eV) N_s(E) dE \qquad (11)$$

Depending on the polarity of the voltage, filled tip states tunnel into empty sample states (V>0), or filled sample states tunnel into empty tip states (V<0). In general, the energy levels can change in the case when the bias voltage is increased.

The Bardeen approach

The next step in calculating the tunneling current will be to evaluate the matrix element M_{ts}. This evaluation must be related to the overlap of tip-sample wave functions, which have the exponentially decaying tails in the barrier, leading to the exponential dependence of tip-sample distance (barrier width).

Bardeen´s approach [3] first applies time dependent perturbation theory to the problem and then makes some additional approximations. In this scheme the system consists of two independent regions: One for the tip and the other for the sample.

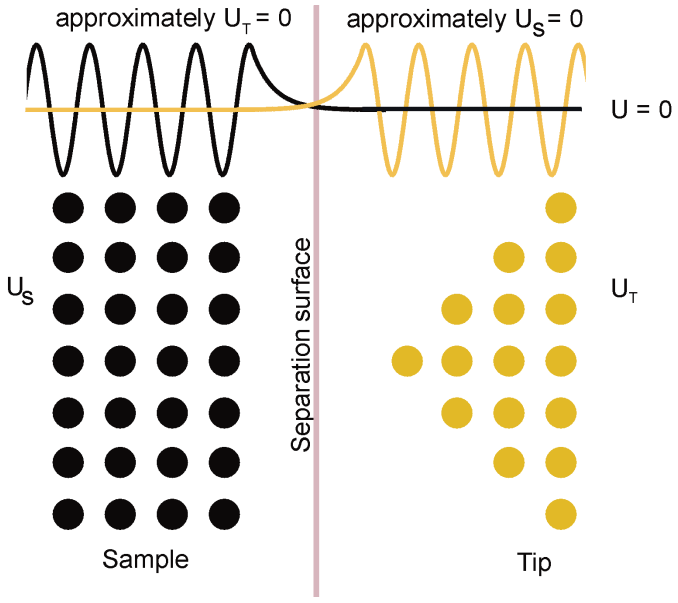

Fig.8

Setup of the Bardeen approach. The two regions: tip region and sample region are separated by the separation surface. In the sample region the tip potential is zero and correspondingly in the tip region the sample potential is zero.

The total separation of the tip and sample systems leads to the two independent Schrödinger equations for both of the systems:

$$(T+U_s)\Psi_s = E_s\Psi_s \tag{12}$$

$$(T+U_t)\Psi_t = E_t\Psi_t \tag{13}$$

T is the operator of the kinetic energy of the single electron, U_t and U_s are the potentials of tip and sample respectively. The single-particle wave function of the entire system is determined by the total Hamiltonian.

$$\mathrm{H} = T+U_t+U_s \tag{14}$$

The actual wave function Ψ can be expanded in any of the basic sets Ψ_S or Ψ_T, like $\Psi = \sum_t a_t\Psi_t$. One can apply time-dependent perturbation theory to describe the tunneling of an electron from the sample to the tip. The case of an electron tunneling from the tip to the sample may be treated completely analogously. This is used in a time-dependent perturbation calculation, where the electron at initial time $\tau = 0$ is taken to be in a sample state. The wave function at the time τ then is written in the form:

$$\Psi(\tau) = \Psi_s e^{-i\omega_s\tau} + \sum_t a_t(\tau)\Psi_t e^{-i\omega_t\tau} \tag{15}$$

where $a_t(0) = 0$ and $E = \hbar\omega$. Now one can insert Ψ in a time-dependent Schrödinger equation:

$$\mathrm{H}\Psi = (T + U_t + U_s)\Psi = i\hbar\Psi \tag{16}$$

From these calculations the effective matrix element M_{ts} is

$$M_{ts} = \langle \Psi_t | U_t | \Psi_s \rangle = \int \Psi_t^* U_t \Psi_s dr \tag{17}$$

To evaluate the matrix element, Bardeen introduced an additional approximation. He assumed the potential U_s should be zero in the tip region of space. Similar a tip potential should be zero in the sample region (*Fig.8*). A separation surface S was introduced. This surface separates the regions in which the two potentials differ from zero. In any region of space it can be written down as $U_t U_s = 0$. In general, the approximation will become better if the potentials U_t and U_s are reasonably small at and beyond the separation surface. This will be the case if the separation surface is located far out in the vacuum. So one can say that Bardeen's approximation is valid for tunneling processes through a wide barrier (large tip-sample distances) and becomes invalid for very small barriers (very small tip-sample distances).

Since $U_t = 0$ outside the tip region, the integration needs to be performed only in the tip region volume Ω_t, and assuming that $U_s = 0$ in this region one obtains:

$$M_{ts}^* = \int_{\Omega_t} \left(\Psi_s (E_t - T) \Psi_t + \Psi_t (T - E_s) \Psi_s^* \right) dr \tag{18}$$

Finally, integrating partially this volume integral can be rewritten as a surface integral

$$M_{ts} = \frac{\hbar^2}{2m} \int_S \left(\Psi_s \nabla \Psi_t^* - \Psi_t^* \nabla \Psi_s \right) \cdot dS \tag{19}$$

where S is the surface separating tip and sample regions. To evaluate the effective matrix element M_{ts} one must know the tip and sample wave functions. But the exact atomic structure of the tip and usually also the chemical nature of the tip apex is unknown, so one needs to assume a model tip wave function which is done in the following section.

Tersoff and Hamann's model

The task of calculating Bardeen's matrix element M_{ts} from ab initio results can be extremely simplified if one makes some assumptions regarding the tip. This method was suggested by Tersoff and Hamann in 1983 and it is still widely used in the interpretation of the STM images [4]. Tersoff and Hamann suggested to replace the unknown electronic structure of the tip by a simple model.

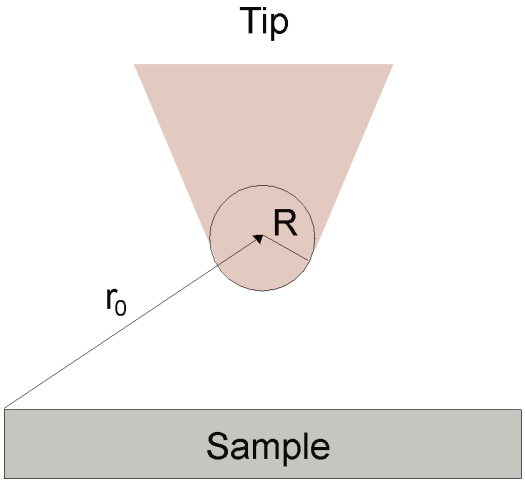

Fig.9

Tip center r_0 and tip radius of curvature R defined by Tersoff and Hamann.

In the system suggested by Tersoff and Hamann the wave function of the outermost tip atom is assumed to be an atomic s-wave-function. Since the tunneling current depends on the overlap of the wave-functions of the tip and the sample, and since the wave-function decays exponentially into the vacuum, only the orbitals localized at the outermost tip atom will be of importance for the tunneling process.

In the limit of low (or zero) temperature and low bias voltage the total current flowing from the tip to the sample can be written as (compare to equation (11)):

$$I = \frac{2\pi e}{\hbar} V \sum_s \left|M_{ts}\right|^2 N_t(E_F)\delta(E_s - E_F) \tag{20}$$

To calculate the matrix element, the surface wave function is written as a 2D Bloch expansion:

$$\Psi_s(r) = e^{ik_{II}\cdot r_{II}} \sum_k a_k(z) e^{iK\cdot r_{II}} \tag{21}$$

where K is a 2D reciprocal lattice vector. After inserting the wave function into the Schrödinger equation and assuming that the potential V is independent of r, it was found that the wave function can be rewritten as:

$$\Psi_s(r) = \sum_K a_K(0) e^{-\sqrt{\kappa^2+(k_{II}+K)^2}z} e^{i(k_{II}+K)\cdot r_{II}} \tag{22}$$

Here $\kappa^2 = \frac{2m}{\hbar^2(V-E)}$. For $E = E_F$ then $\kappa^2 = \frac{2m\phi}{\hbar^2}$, ϕ is the work function of the surface.

The main point of the Tersoff and Hamann theory is the choice of the wave function of the tip. Since there is usually no particular knowledge of the exact tip structure, it can be modelled as a locally spherical potential (*Fig.9*), with curvature R about the center r_0. So asymptotically the tip wave function is chosen to have the form of an s-wave:

$$\Psi_t(r) = \kappa R e^{\kappa R} \frac{e^{-\kappa|r-r_0|}}{\kappa|r-r_0|} \tag{23}$$

If to write the tip wave function as a 2D Fourier sum and let the separation surface S be a plane parallel with the surface, then one finds that the matrix element is proportional to the sample wave function evaluated at the tip center of curvature:

$$M_{ts} \propto \Psi_s(r_0) \tag{24}$$

Then the total current flowing from the tip to the sample can be written as:

$$I \propto VN_t(E_F) \sum_s |\Psi_s(r_0)|^2 \delta(E_s - E_F) \tag{25}$$

In this expression the sum is the local density of states of the sample (LDOS) at the Fermi level evaluated at the tip center. Then the last expression can be rewritten as

$$I \propto VN_t(E_F)LDOS(r_0, E_F) \tag{26}$$

This equation gives an easy tool for interpretation of the STM images. The basic quantity which is imaged is the local density of states of the sample. Even though this model has been very successful it fails in some cases. It turns out that the corrugation amplitudes of closed packed metal surfaces predicted by this model are too small to explain the measured atomic corrugations. Chen [5] pointed out that this is partially due to the restriction to s-like tip orbitals.

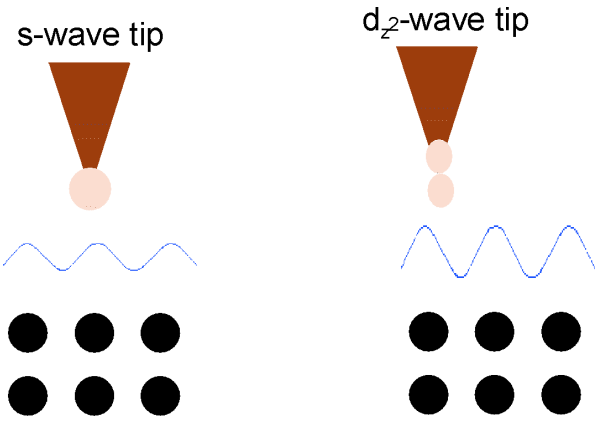

Fig.10
Chen [3] showed that a tip apex wavefunction with d-character greatly enhances the measured corrugation.

It is quite intuitive to assume that a more directed tip state like a d_{z^2}-state pointing towards the sample might lead to the higher values of the corrugation amplitude. Chen used this idea to extend the Tersoff-Hamann model (for other tip states, not only for s-wave tip).

He showed that other tip states can also be related to the Greens' function of the Schrödinger equation. This finally leads to the 'derivative rule' which gives the simple dependence of the matrix element M_{ts} from the tip orbital. The sample wave function essentially have the form:

$$\Psi_s \propto e^{-\kappa z} \tag{27}$$

Here z is the coordinate perpendicular to the surface.

For several tip orbitals matrix elements are listed in the Table 1.

Table1

Matrix elements obtained by use of Chens' derivative rule.

Tip wave function orbital type	Matrix element $M_{ts} \propto$
s	$\Psi_s(r_0)$
$p_i,\ i = x, y, z$	$\frac{\partial}{\partial i}\Psi_s(r_0),\ i = x, y, z$
$d_{z^2-\frac{r^2}{3}}$	$\left(\frac{\partial^2}{\partial z^2} - \frac{\kappa^2}{3}\right)\Psi_s(r_0)$

Including tip orbitals which are directed towards the sample (p_z or $d_{z^2-\frac{r^2}{3}}$) will lead to matrix elements which are still proportional to the wave function at the position of the tip. However, the prefactor is changing resulting in a higher current. This means that by including different tip states, theoretical values of corrugation amplitudes are closer to the experimentally measured ones. On layered materials the elastic deformation of the surface by the tip is known to lead to much higher corrugations (see Addition 2), a mechanism is not included so far in this discussion.

Recently extensive ab initio calculations explained the remaining discrepancy of experiment and theory for metal surfaces by elastic deformation of tip and surface [6]. In the measurement in this work (see Chapter 2) the enhanced corrugation amplitudes are explained with a chemically modified tip which is weakly bounded and therefore easily deformed.

References

[1] G. Binnig, H. Rohrer, Helv. Phys. Acta **55**, 726 (1982).

[2] L. Olesen, E. Lægsgaard, I. Stensgaard, F. Besenbacher, Applied Physics A **66**, S157 (1998).

[3] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).

[4] J. Tersoff, D. Hamann, Phys. Rev. B **31**, 805 (1985).

[5] A. Selloni, P. Carnevali, E. Tosatti, C. Chen, Phys. Rev. B **31**, 2602 (1985).

[6] W. Hoger, A. Fisher, R. Wolkow, P. Grütter, PRL **87**, 236104 (2001).