Scanning Probe Microscopy, Theory

AJ Fisher, University College London, UK

Symbols

d	tip–sample separation
e	electronic charge
$f(E)$	occupation probability for electron state with energy E
$\hbar$	Planck's constant divided by 2π
k_{cant}	force constant of SFM cantilever
m_e	mass of electron
M_{tip}	mass of vibrating tip assembly
M_{ts}	matrix element connecting tip and sample states in STM
r_0	centre of curvature of STM tip
R	radius of curvature
V	potential energy of electron
W_{ts}	transition rate between tip and sample states
κ	decay constant for electron wavefunctions ψ
σ	differential conductance
Φ	electrostatic potential
ω	angular frequency of vibration
Ω	normalization volume for tip wavefunction in the Tersoff–Hamann model

Introduction

The three most important scanning probe techniques are

- Scanning tunnelling microscopy (STM);
- Scanning force microscopy (SFM, also known as atomic force microscopy, AFM);
- Scanning near-field optical microscopy (SNOM).

The three methods give different types of information and require correspondingly different theoretical treatments. STM probes the electronic states of a surface; SFM probes the force (or force gradient) between a tip and a surface; while SNOM probes the electromagnetic field near a surface. In addition, magnetic force microscopy (MFM) is now a recent addition to the range of techniques and is covered by a separate article in this encyclopedia.

However, all these techniques share several common features. First, they measure local, not average, surface properties. Any theory must therefore include the local surface properties if it is to be useful. Second, they all lack a simple *inversion theorem*: in no case is it possible to infer directly physical properties of the system from the scanning probe results. Interpretation therefore has to proceed by an indirect 'interpretation cycle':

1. Build a model of the relevant local features (e.g. structure, excitations) of the system under study;
2. Develop a theory of the scanning probe experiment concerned;
3. Combine (1) and (2) to determine the predicted experimental signal from the model adopted;
4. Alter the model if the predictions and the experiment do not match.

In this article we shall examine what type of model of the physical system under study is appropriate under item (1) of the 'interpretation cycle' for each technique, and how a suitable theory of the experiment can be constructed for item (2).

The Scanning Tunnelling Microscope: Electronic Spectroscopy

General Considerations

The fundamental physical process in STM is the tunnelling of electrons between the tip and the sample under study, through the barrier formed by the vacuum between them (see **Figure 1**). The 'height' of this barrier in energy is approximately equal to the work functions of the tip or sample material. In the simplest possible one-dimensional model, we assume that the electron potential energy V takes a constant value V_0 through the tunnelling
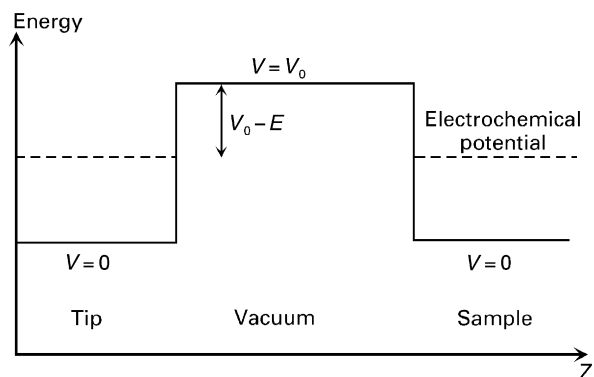
Figure 1 Schematic diagram of an STM junction at zero bias, illustrating the meaning of the symbols defined in the text.

gap; the barrier height is therefore $(V_0 - E)$ where E is the electron energy. The electron wavefunctions then decay in the vacuum as $\exp(-\kappa z)$, where z is the coordinate normal to the surface and

$$\frac{\hbar^2 \kappa^2}{2 m_e} = (V_0 - E) \qquad [1]$$

If $V_0 - E$ is $5 \, \text{eV} = 8.01 \times 10^{-19} \, \text{J}$ then $\kappa = 1.15 \times 10^{10} \, \text{m}^{-1}$

Tunnelling can occur from tip to sample and from sample to tip. If no bias is applied to the system (i.e. if the electrochemical potentials of the electrons in tip and sample far from the junction are equal), the rates of tunnelling in opposite directions are equal, and no net current flows. (Note that if the tip and sample have different work functions, a finite charge transfer will occur at zero bias to establish a dipole layer at the surfaces, and hence an electric field in the vacuum gap; it is the potential difference arising from this field which equalizes the electrochemical potential in the two materials.)

Suppose now that a finite bias potential $\Delta \Phi$ is applied to the system (see **Figure 2**), of a sign which raises the electrochemical potential for electrons on the left of the junction by $|e| \Delta \Phi$ relative to those on the right. Over a range of energies, electrons are now more likely to tunnel from left to right than viceversa, and a net current flows from right to left. If the difference in chemical potentials is small so that current is dominated by electrons with a single energy E, we can use the fact that the current is proportional to the tunnelling probability and hence to the absolute square of the wavefunction to deduce that it will vary with the tip–sample separation d as $\exp(-2\kappa d)$, with κ given by eqn [1]. Taking the value of κ we estimated earlier, we obtain the often-quoted rule of thumb that the tunnel current should reduce by roughly a factor of ten whenever the tunnel gap is increased by $1 \, \text{Å} = 10^{-10} \, \text{m}$.

This is an approximate theory of the tunnelling process, but it says nothing about the contrast to be expected when the STM tip is moved across the surface. A better
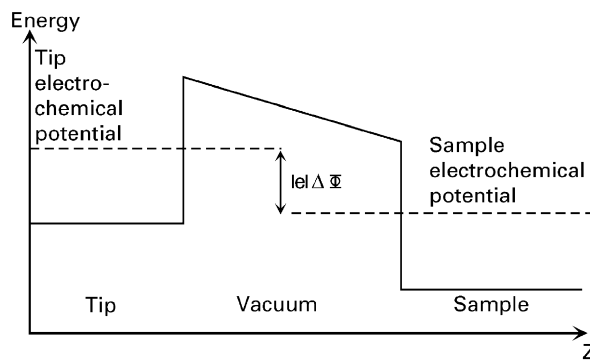
theory must take account of the atomistic structure of the tip and the surface, as well as a better theory of the tunnelling between them. In doing this, it is important to realize that the energy of the tunnelling electrons being used to probe the system is very similar to the energy of electrons in the bonding orbitals holding the atoms together. There is therefore a very close relationship between the tunnelling process, the electronic structure, and the atomic (or chemical) structure of the system.
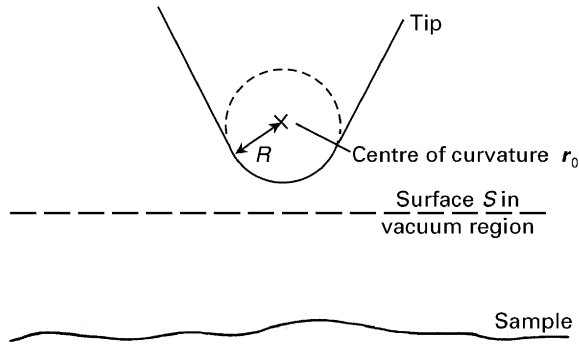
Step (1) of the 'interpretation cycle' for STM must therefore involve a model of the atomic and electronic structure of the surface, including any adsorbates or surface defects. In practice this is most often obtained numerically using density-functional theory, in which the total energy of the electrons in the system is calculated from the electronic charge density, rather than from the full many-electron wavefunction. The Hartree–Fock method, which employs an approximate form for the many-electron wavefunction which neglects the correlations between the motions of the electrons, is also used. Such calculations are now relatively standard, and many can be found in the literature for surfaces of different types. Step (2) must involve a three-dimensional theory of electron tunnelling between the surface (represented in this way) and the tip; we now turn to this more difficult step.

Perturbation Theory

The interpretation of many spectroscopies (for example, optical spectroscopy) proceeds by the identification of a well-defined 'perturbation' which is applied to the system when the experiment is performed. This is both convenient (because the response of the system to the perturbation is not too difficult to evaluate in terms of the matrix elements of the perturbation) and conceptually useful (because it allows a clear separation between the 'system' and the 'probe').

This is not straightforward in STM. There are two reasons: (i) the tip and the sample may be very close, and hence strongly coupled together, and (ii) even when this is not the case, there are mathematical difficulties in separating the Hamiltonian into parts describing a non-interacting tip and sample, because the kinetic energy operator for the electrons appears in both parts. Nevertheless problem (ii) has been solved, and it has been shown that a sensible perturbation theory can be constructed in which the appropriate matrix element is that of the electron current density operator, evaluated over a surface S separating the tip and the sample (see **Figure 3**). We write

$$M_{ts} = \frac{\hbar^2}{2 i m_e} \int_S d^2 \boldsymbol{r} [\psi_t \nabla \psi_s - \psi_s \nabla \psi_t] \qquad [2]$$

Energy

Tip electro-chemical potential

Sample electrochemical potential

$|e| \Delta \Phi$

Tip   Vacuum   Sample

Z

Figure 2   Schematic diagram of an STM junction at finite bias.

Figure 3 The important quantities in the Tersoff–Hamann model of STM. The matrix element is evaluated on the surface S; the conductance is proportional to the sample density of states at the tip centre of curvature r_0.

where ψ_t is a one-electron state of the tip (in the absence of the sample) and ψ_s is a state of the sample (in the absence of the tip). Note that in order to derive this result, one has to assume that *both* states are valid solutions of the Schrödinger equation in the neighbourhood of the surface S; this implies that the potential for electrons at S must be equal to the vacuum potential. The transition rate for electrons from state t to state s (or vice versa) can then be written

$$W_{ts} = \frac{2\pi}{\hbar}|M_{ts}|^2\delta(E_t - E_s) \quad [3]$$

and the total current from tip to sample as

$$I = \sum_{ts}\frac{2\pi}{\hbar}|M_{ts}|^2\delta(E_t - E_s)f_t(E_t)[1 - f_s(E_s)] \quad [4]$$

where $f_t(E)$ and $f_s(E)$ are the occupation probabilities for electron states with energy E in the tip and the sample respectively.

The most commonly used model in interpreting STM data is the Tersoff–Hamann model, in which the analysis is carried a step further. It is assumed that the tip wavefunction is an s-wave, and decays into the vacuum as

$$\psi_t(r) = \Omega^{-1/2}\kappa R\exp(\kappa R)\frac{\exp(-\kappa|r - r_0|)}{\kappa|r - r_0|} \quad [5]$$

where Ω is a normalization volume, r_0 is the centre of the curvature of the tip, R is the radius of curvature, and κ is as defined earlier. In this special case the integral in eqn [2] can be evaluated exactly (under the assumption that ψ_s obeys the free-space Schrödinger equation), and one finds in the limit of small bias that the differential

conductance σ of the STM is

$$\sigma = \frac{32\pi^3}{\hbar}e^2 N_t(E_F)R^2\kappa^{-4}\exp(2\kappa R)$$
$$\sum_S|\psi_s(r_0)|^2\delta(E_s - E_F) \quad [6]$$

Here $N_t(E_F)$ is the total tip density of states at the Fermi energy. This is a very simple and important result; it tells us that the tunnelling conductance measures the sample density of states at the Fermi energy, evaluated at the centre of curvature of the tip (i.e. some distance outside the sample surface). This is relatively straightforward to calculate, and easy to interpret in simple chemical terms. The model disregards all details of the tip; they are absorbed into the values of the constants R and Ω. The (usually unknown) structure of the end of the tip can therefore be disregarded, at the cost of sacrificing any information about the absolute value of the conductance. It is largely because of these advantages that the Tersoff–Hamann model is so popular.

The approximations leading to eqn [6] are valid only if there is no electric field in the vacuum. Nevertheless, the Tersoff–Hamann model is often used to interpret images taken at finite bias voltage $\Delta\Phi$, or even data from the 'spectroscopic' mode of the STM in which the tip position is held fixed and the bias varied. The density of states involved in eqn [6] is projected onto a 'window' of energies $\Delta E = e\Delta\Phi$, rather than onto a single energy. There is no theoretical justification for this, as the true states of the system are bound to be modified by the addition of such a bias voltage, but it has proved useful as a way of qualitatively rationalizing STM data provided the bias is not too large.

It is possible to extend perturbation theory beyond the Tersoff–Hamann model, for example by including tunnelling to or from states of non-zero angular momentum on the tip, or by using states explicitly calculated from a particular atomistic model to find the matrix element in eqn [2]. However, both these approaches require additional information about the geometry of the tip and the electronic states it supports. This is generally not available from experiment, as a tip will be modified by the forces acting in the course of the experiment (as discussed in more detail below); even if the tip is well characterized before use (for example, by electron microscopy or field-ion microscopy), this information will become out-of-date once the experiment starts.

Another extension of this type of perturbation theory is to the case where there is some additional electronic order in the tip or the sample – for example, magnetic or superconducting order. In the case of magnetic order one is led to consider separate currents of spin-up and spin-down electrons, proportional to the spin-resolved components of the density of states. For a superconductor, the

tunnel current depends on the quasiparticle density of states.

Beyond Perturbation Theory

Perturbation theory of this kind leads to an appealing picture of STM. Nevertheless it is not always justified; here we list some of the reasons why it may break down.

- A substantial redistribution of charge and potential takes place, so the effective one-electron Schrödinger equation is altered. This effect has been predicted theoretically when the tip–sample separation drops below about $3\,\text{Å}$; it tends to result in a lowering of the potential energy for an electron in the vacuum and a collapse of the tunnelling barrier.
- The electron tunnelling probability between tip and sample is not small. In practice this occurs only when the electron transport is no longer dominated by tunnelling – either because a physical contact or *nanojunction* has been formed between the two, or because the tunnel barrier has collapsed completely (see above). The signature of this state of affairs is that the STM conductance becomes of the order of the quantum of conductance, e^2/h.
- Although small, the tunnelling matrix element through the vacuum is not the smallest energy scale in the problem. This can occur when, for example, a highly insulating molecule is adsorbed on a surface; tunnelling through the molecule can then be just as difficult as tunnelling through the vacuum, so it is not appropriate to treat the vacuum tunnelling as a perturbation.

For all these problems, the theoretical cure is the same: one must perform a single coupled calculation for the electron states in the whole system (tip plus adsorbate – if any – plus sample) under a non-zero bias, allowing a current to flow.

However, this has proved to be very difficult without additional simplifications. In the elastic scattering quantum chemistry (ESQC) method developed by Joachim and Sautet, there is no self-consistency in the Hamiltonian for the electrons and only a relatively small basis set, giving very limited flexibility to the electron wavefunctions. In another approach, pioneered by the group of Tsukada, a more detailed numerical representation of the wavefunction is adopted: the wavefunctions are calculated on a mesh of points and full self-consistency is achieved between the wavefunctions and the electronic potential. The simplification in this case is that the wavefunctions far from the tunnel junction are those of a fictitious 'jellium' in which the positive charge of the nuclei is smeared out into a uniform background. In yet a third approach the conductance is calculated in a non-

perturbation manner between two localized states, rather than between the true bulk states of the tip and sample.

Other Factors

There are also other factors that are known to be important in STM. One of these is the mechanical interaction between the tip and the sample; the forces that arise can distort the tip, with the result that the displacement of the tip apex is not the same as that recorded from the piezoelectric actuators controlling the tip (see **Figure 4a**). This effect was revealed by careful measurements of the *corrugation* of the STM image of adsorbates on metals as a function of the conductance.

A second effect is that of the tip electric field. This can be very large: fields above $10^9\,\text{V m}^{-1}$ can be obtained when a potential difference of a few volts is dropped over a narrow tunnelling gap. This can have two results: first it distorts the atomic structure of the surface, causing
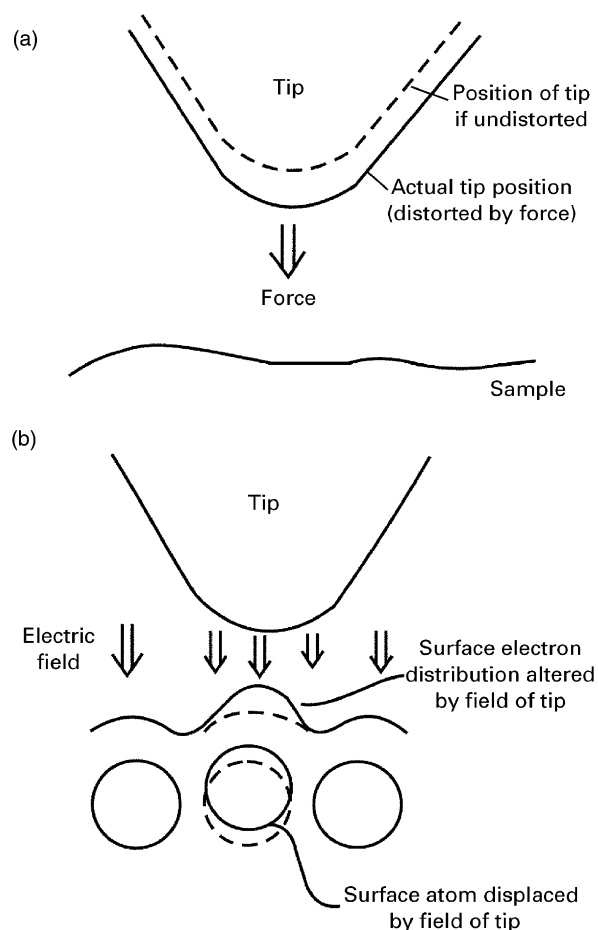
Figure 4 Two physical phenomena which alter STM images from those predicted by simple theory: (a) Tip–sample forces distort the actual change in separation from that measured by the piezoelectric transducers. (b) Electric fields from the tip cause motion of surface atoms and distortion of surface electron states.

movements of a few tenths of an angstrom in the surface atoms, and second it distorts the electronic structure, changing the tunnelling probability at different points on the surface (see **Figure 4b**). These effects have been shown to be important in images of the Si(001) surface.

A third complicating effect is the inelastic scattering of electrons from other excitations during tunnelling. These other excitations may be electronic; the most important example is surface plasmons, which may be found in either the tip or the sample. Scattering from surface plasmons produces electromagnetic disturbances near the tunnel gap which can result in the emission of electromagnetic radiation from the STM. The frequency distribution of the emitted light is then characteristic of the surface plasmon spectrum. The distortion of the surface plasmons can also have other, more subtle, effects since it determines whether or not the electron experiences the effect of the 'image interaction' outside a surface. This can in turn have a large effect on the electron potential and the tunnelling current.

Alternatively, the scattering excitations may be atomic vibrations. These may result in phonon-assisted sidebands around resonant tunnelling peaks, corresponding to the absorption or emission of phonons. In extreme cases the transfer of electronic energy to atomic motion may produce atom transfer between tip and sample, or even desorption; this is a form of DIET (desorption induced by electronic transitions) and may be used to break bonds selectively on surfaces.

The Scanning Force Microscope: Force Spectroscopy

In order to interpret these experiments one needs to bear in mind the different types of forces that can act between the tip and the sample.

- At large distances the force most commonly present is the Van der Waals force. Between two atoms the Van der Waals force energy decays with separation z according to the well-known z^{-7} law, but for a sphere above a planar surface (one simple model for the tip–surface system) the decay is only as z^{-2}. This relatively slow fall-off tells us that in SFM, unlike STM, the large-scale structure of the tip is important.
- If the sample is an insulator, it may be locally charged. The interaction between these local 'patch charges' and the tip also decays like a power law in the tip–sample separation. The patch charges are difficult to control; the highest-resolution SFM results are generally obtained on conducting samples.
- At smaller distances (of the order of 3–5 Å separation) local interactions between the closest atoms of the tip and sample start to become important. These include the onset of covalent bonding, and local electrostatic forces.

- As the tip–sample separation drops below the sum of the atomic radii of the atoms, the Pauli exclusion principle raises the energy of the overlapping electron distributions, producing a repulsive force. If the tip and sample are forced together beyond this point, atomic deformations (first elastic, then plastic) occur.

Models for the Forces

Of these interactions, the Van der Waals attraction and the Pauli repulsion are universal; the presence of the others depends on the nature of the material. The combination of Van der Waals and Pauli interactions is often captured by the simple '6-12' Lennard-Jones interatomic potential

$$V(r) = \varepsilon \left[\left(\frac{r}{\sigma} \right)^{-12} - \left(\frac{r}{\sigma} \right)^{-6} \right] \qquad [7]$$

in which the attractive r^{-6} term represents the Van der Waals force and the repulsive r^{-12} term the Pauli force. Simulations of generic interatomic interactions are often performed using this potential, although it cannot be expected to be realistic for anything other than interactions between the simplest rare-gas solids. More realistic calculations include approximate forms for the electrostatic and covalent interactions between the atoms, or (better still) find these forces directly from the electronic structure of the materials involved.

High-Resolution SFM Operation

With this in mind, let us examine the most common modes of SFM operation when high-resolution information about the surface is required.

- **Non-contact mode**. In this mode the tip is kept at a distance from the sample in the attractive part of the force–distance curve; usually it is then scanned across the sample, and the tip–sample distance adjusted to keep the cantilever displacement (and hence the force) constant. This procedure keeps the tip in the region where the tip–sample force is (relatively) well understood, but with the price that the force is determined by the cumulative effect of a large number of atoms – hence the resolution of individual atomic-scale features is seldom possible.
- **Contact mode**. Here, by contrast, the tip is allowed to penetrate into the repulsive regime of **Figure 5**. This has the advantage that one expects a large component of the force to be determined by a relatively small number of atoms near the tip apex, but the disadvantage that the force becomes dependent on complex atomic processes involving the irreversible deformation of the tip–sample junction. Images with apparent atomic resolution can be seen in contact mode on simple crystalline materials such as alkali

Force

Repulsion

a

b

Distance

d

c
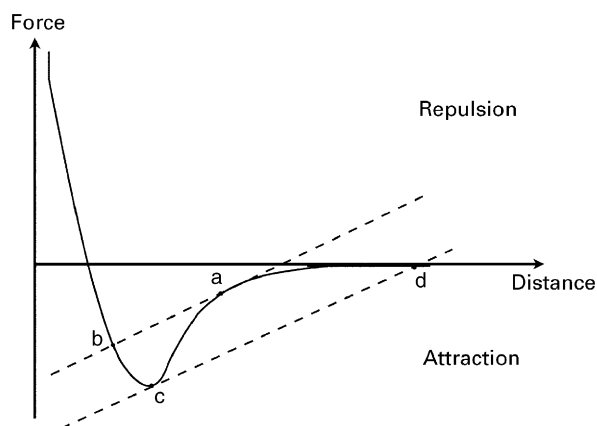
Attraction

Figure 5 Schematic force–distance curve for an SFM experiment. On approach, the tip jumps from point a to point b; on retraction, it jumps from c to d. The dotted lines have a slope equal to the cantilever force constant k_{cant}.

halides, but the conclusion of careful simulations is that the atomic-scale features are not, in fact, correlated with the positions of atoms in the surface. This theoretical conclusion is reinforced by the failure to resolve atomic defects (known to be present on the surface) in experiments.

One might think that a technique intermediate between contact and non-contact modes could be devised simply by bringing the tip close to the surface, but not in contact with it. In fact this is very difficult because of the 'jump-to-contact' phenomenon: a static tip held above a surface by an SFM cantilever with a given force constant k_{cant} can be stable only as long as the force gradient from the tip–sample interaction is less than k_{cant} (see **Figure 5**). The force gradient of a Van der Waals interaction between a tip and a flat surface diverges as the separation between them is reduced, so this condition is always violated and the tip snaps into contact with the sample. If the tip is pulled off the surface, a similar jump out of contact occurs (although between different values of tip–sample separation).

Since a very interesting range of tip–surface separations is rendered unavailable by the jump to contact, it would be desirable to eliminate it. To date, this has been done in two ways. First, a dynamical approach is used: the cantilever is vibrated above the surface with an amplitude of several hundred angströms, in such a way that its point of closest approach is only a few angströms from the surface. The difference from before is that the tip is accelerating rapidly away from the surface as it approaches; this suppresses the jump to contact. One way of expressing this is to say that the effective cantilever force constant is increased from k_{cant} to $k_{cant} + M_{tip}\omega^2$, where M_{tip} is the total mass of the vibrating tip and ω is the

angular frequency of vibration. The tip is usually scanned while keeping the vibrational period constant; this corresponds approximately to a scan of constant force gradient. Atomic resolution has been obtained using this technique, initially on the Si(111)–7×7 surface but now also on others. It seems this resolution can be understood in terms of the interaction between the tip and the surface near the point of closest approach, but the theory is complicated because the vibration of the tip samples all the different regions of the potential surface described above during a cycle, so a unified model containing all of them must be used.

A second approach is to control the force on the tip directly, generally by means of a small magnet mounted on the back. This removes the need to model a complicated tip oscillation, but imposes stringent demands on the response and stability of the electronics controlling the force. Direct measurements of tip–sample potential curves have now been reported using this technique, but comparison with theory is still in its infancy.

Measurements of Elastic Properties

If local but not ultra-high-resolution measurements are required to probe the elastic properties of a hard material, there are advantages in using high-frequency measurements.

The Scanning Near-Field Optical Microscope-Optical Spectroscopy

The theory of scanning near-field optical microscopy is somewhat similar to that of STM, with the transport of light (or photons) replacing the transport of electrical current (electrons). Instead of the Schrödinger equation, the Maxwell equations for the electromagnetic field must be solved near the tip and the sample, taking into account the local electromagnetic properties of each medium. In some respects this is easier, because (in the absence of non-linear media) the Maxwell equations are truly linear and no self-consistency (of the type needed between effective one-electron wavefunctions and the potential) is needed. Also, since the characteristic wavelengths and decay lengths for optical photons are much larger than atomic dimensions, a continuum treatment of the tip and sample materials is almost always sufficient. On the other hand, the Maxwell equations require treatment of two coupled vector fields.

As in the STM case, the equations must in practice be solved numerically. Perturbation theory is seldom employed, and most calculations make a direct solution for the optical modes at a fixed frequency, either by a transfer matrix approach, or by using the Dyson equation to obtain the solution from that of an exactly soluble system (for example, free space).

Conclusions

STM appears to be the most subtle of the scanning probe methods, relying as it does on quantum-mechanical tunnelling. In fact the theory for this technique is the best developed of all the scanning probe family, but much progress remains to be made in accounting correctly for the nature of the tip and for tip–sample interactions. The theory of near-field optical microscopy is similar in spirit, and in some ways more straightforward. The understanding of SFM data is very incomplete, particularly for experiments with resolution on the atomic scale.

See also: Magnetic Force Microscopy, Scanning Probe Microscopes, Scanning Probe Microscopy, Applications, Surface Plasmon Resonance, Applications, Surface Plasmon Resonance, Instrumentation,Surface Plasmon Resonance, Theory.

Further Reading

Briggs GAD and Fisher AJ (1999) Molecules on semiconductor surfaces: STM experiment and atomistic theory hand in hand. *Surface Science Reports* 246: 1–81.

Blöchl PE, Joachim C, and Fisher AJ (eds.) (1993) *Computation for the Nanoscale*. Dordrecht: Kluwer Academic.

Chen J (1993) *Introduction to Scanning Tunnelling Microscopy.* Oxford: Oxford University Press.

Datta S (1995) *Electronic Transport in Mesoscopic Systems*. Cambridge: Cambridge University Press.

Israelachvili J (1992) *Intermolecular and Surface Forces*, 2nd edn. London: Academic Press.

Sautet P (1997) Images of adsorbates with the scanning tunnelling microscope: theoretical approaches to the contrast mechanism. *Chemical Reviews* 97: 1097–1116.

Wiesendanger R (1994) *Scanning Probe Microscopy and Spectroscopy – Methods and Applications*. Cambridge: Cambridge University Press.

Wiesendanger R and Güntherodt H-J (eds.) (1996) *Scanning Tunnelling Microscopy III*, 2nd edn. Berlin: Springer-Verlag.