Scanning Microscopy

Volume 1993
Number 7 *Physics of Generation and Detection of Signals Used for Microcharacterization*

Article 20

1993

Contrast Mechanism of Scanning Tunneling Microscopy: The Two-Sided Point of View

C. Julian Chen
IBM TJ Watson Research Center

Follow this and additional works at: https://digitalcommons.usu.edu/microscopy

 Part of the Biology Commons

Recommended Citation

Chen, C. Julian (1993) "Contrast Mechanism of Scanning Tunneling Microscopy: The Two-Sided Point of View," *Scanning Microscopy*: Vol. 1993 : No. 7 , Article 20.
Available at: https://digitalcommons.usu.edu/microscopy/vol1993/iss7/20

This Article is brought to you for free and open access by the Western Dairy Center at DigitalCommons@USU. It has been accepted for inclusion in Scanning Microscopy by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

Scanning Microscopy Supplement 7, 1993 (Pages 281-299) 0892-953X/94$5.00+.25
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

CONTRAST MECHANISM OF SCANNING TUNNELING MICROSCOPY: THE TWO-SIDED POINT OF VIEW

C. Julian Chen

IBM TJ Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598
Telephone number: (914) 945-2935, FAX number: (914) 945-2141, e-mail: JUCHEN @ YKTVMV

Abstract

It has been known for three decades that tunneling experiments should be explained by the electronic structures of *both* electrodes. Scanning tunneling microscopy (STM) is no exception. Since the development of STM in the early 1980s, experimental facts have repeatedly shown the necessity to explain the STM images and tunneling spectra from a *two-sided point of view*. In other words, the STM images and tunneling spectra should be explained by the interaction of the electron density distributions of the tip *and* the sample (in energy and space). In the early years of STM, due to a scarceness of experimental data and conceptual difficulties, a one-sided view was commonly used, which attributed the STM images and scanning tunneling spectroscopy (STS) data to the electron density distribution of the sample *only*. As experimental findings accumulate and theoretical concepts develop, a consistent two-sided view of STM has been gradually formulated. This review article is a brief account of the two-sided view of STM in a conceptual and historical perspective.

Key Words: Microscopy, tunneling, STM, atoms, electronic states, energy spectrum.

The Scanning Tunneling Microscope

The scanning tunneling microscope (Binnig and Rohrer, 1982, 1987) is shown schematically in Figure 1. A probe tip, usually made of W or Pt-Ir alloy, is attached to a *piezodrive,* which consists of three mutually perpendicular piezoelectric transducers: *x*-piezo, *y*-piezo, and *z*-piezo. Upon applying a voltage, a piezoelectric transducer expands or contracts. By applying a sawtooth voltage on the *x*-piezo and a voltage ramp on the *y*-piezo, the tip raster scans on the *xy*-plane. Using the coarse positioner and the *z*-piezo, the tip and the sample are brought to within a few Ångströms from each other. The electron wavefunctions in the tip overlap electron wavefunctions in the sample surface. A bias voltage, applied between the tip and the sample, causes an electrical current to flow by *tunneling* through the vacuum gap. The tunneling current is amplified by the current amplifier to become a voltage which is compared with a reference value. The difference is then amplified again to drive the *z*-piezo. The phase of the amplifiers is chosen to provide negative feedback. If the tunneling current is larger than the reference value, then the voltage applied to the *z*-piezo tends to withdraw the tip from the sample surface and *vice versa.* Therefore, an equi-

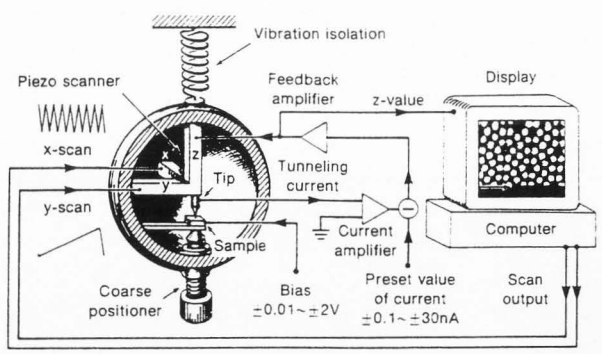

Figure 1. Schematic diagram of a scanning tunneling microscope

librium z-position is established through the feedback loop. As the tip scans over the xy-plane, a two-dimensional array of equilibrium z-positions, representing a contour plot of the equal tunneling-current surface, is obtained and stored. During the last ten years since its debut, the STM has resolved details of local electronic structure of a large variety of conductive surfaces down to atomic scale, which is about 2 Å (Quate, 1986). According to Rohrer (1992), the atomic resolution on rigid surfaces already becomes a must in all STM experiments.

One of the motivations for Binnig *et al.* (1982) to pursue tunneling with a controllable gap was to achieve it in a configuration that allows simultaneously spatially resolved tunneling spectroscopy. By fixing the position of the tip with regard to the sample, varying the bias voltage and recording the tunneling current, a local tunneling spectrum is obtained. By scanning the tip over the sample surface, a two-dimensional array of tunneling spectra is obtained. This technique, commonly called scanning tunneling spectroscopy (STS), was demonstrated by Feenstra and his coworkers (Feenstra and Stroscio, 1987; Feenstra et al, 1987). During the seven years of its invention, a large number of STS data has been accumulated.

Tunneling Junction and the Bardeen Theory

Tunneling spectroscopy with metal–insulator–metal (MIM) tunneling junctions was first demonstrated by Giaever (1960a, 1960b), which provided a direct confirmation and measurement of the energy gap in superconductors, as shown in Figure 2. If both electrodes are normal metals, the density of states (DOS) of both electrodes are almost constant with respect to energy. The $I-V$ curve is a straight line. If one of the metals is superconducting (that is, there is an energy gap and the DOS has a sharp peak), then the $I-V$ curve shows a threshold. If both metals are superconducting (that is, energy gaps exist in the DOS of both), then there are two thresholds with an interval of negative differential conductance in between. To summarize, the observed tunneling spectrum is determined by a convolution of the electronic spectra of *both* electrodes. Although the experimental result contains information of both electrodes, it does not cause confusion. Actually, by using the same right-hand-side electrode and combining with different left-hand-side electrodes, the difference in the observed tunneling spectra should be a measure of the difference in the left-hand-side electrode. A similar argument is valid for the right-hand-side electrode. In all cases, the origin of the observed tunneling spectrum is symmetric with respect to the two electrodes.

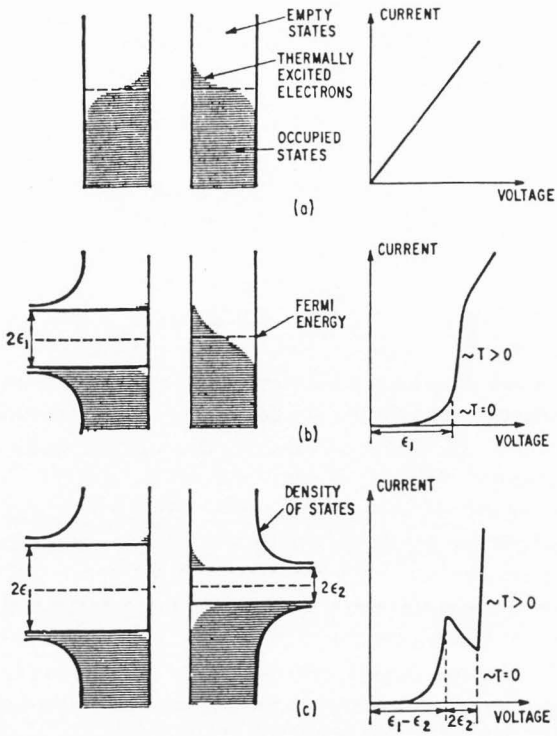

Figure 2. Tunneling spectroscopy in classic tunneling junctions. (a) If both electrodes are metallic, the $I-V$ curve is linear. (b) If one electrode has an energy gap, an edge occurs in the $I-V$ curve. (c) If both electrodes have energy gaps, two edges occur. A "negative differential conductance" exists. (After Giaever and Megerle, 1961).

Tip-State Effects in STM and STS

A logical consequence of the Bardeen theory is that the STM images and the STS spectra are determined by the electronic states of the sample *and* the tip. Since the pioneering work of Binnig and Rohrer (1982), numerous authors observed, reported, and analyzed tip-state effects in STM and STS. Following are some examples.

Demuth et al. (1986) reported an STM image in which the corrugation undergoes a sudden change in the middle of a scan (see Figure 3). The top two-thirds of the image shows a large corrugation, while the lower one-third has a much smaller corrugation. The

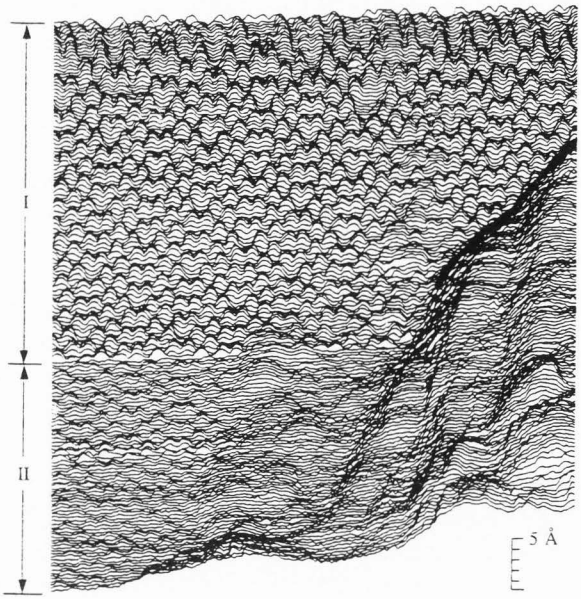

Figure 3. Topographic image of Si(111)-7×7 surface. Due to a spontaneous tip restructuring, the corrugation undergoes an abrupt change from large (region I) to very small (region II). After Demuth et al. (1986).

differences between regions I and II arise from a small change in the structure of the tip during scanning. The ≈4 Å lateral shift and the 0.5 Å change of the vertical position suggest an atomic-scale rearrangement of the tip.

Demuth et al. (1986) also made an explanation of this phenomenon based on an educated guess. They argued that a cluster of tungsten (or silicon) atoms at the apex of the probe tip may have been atomically rearranged by a very slight contact to the surface or by the transfer of a chemically active atom which reacts to change the tip configuration. Actually, a similar observation was reported in an early paper of Binnig and Rohrer (1982). Their observation motivated Baratoff (1983, 1984) to propose a theory of the contrast mechanism of STM. Baratoff proposed that the large corrugation observed by the STM is due to the presence of a dangling bond protruding from the tip apex. His proposal was based on the observation of a sudden improvement of experimental resolution accompanied by a retraction of the tip while the set tunneling current remains unchanged, as described by Binnig and Rohrer (1982).

Klisner et al. (1990) reported and analyzed an interesting observation. During the imaging of a Ge(111)-c2×8 surface, they observed a double-tip effect. Each feature on the surface was duplicated to become a double feature. The tunneling spectra of each point on the surface with the two microtips were recorded. They observed that the two tunneling spectra originating from the two microtips at the same point of the surface were substantially different. Especially for the section of tunneling spectra between −0.5 V and −2 V, a very large difference was observed. Clearly, the electronic structures of the two microtips on the same tip shank are very different.

Klitsner et al. (1990) made the following analysis. From the point of view of tunneling theory, the information obtained from STM is a convolution of electronic states in the sample with electronic states in the tip. Although the electronic structure of *bulk* metal tips usually resemble a free-electron metal, for atomically sharp tunneling tips, it could be very different. the electronic structure of a microtip made of a small metal cluster may be very different from that of a bulk made of the same metal atoms. In general, the sharpest, highest-resolution tips are the most likely to have highly structured electron energy spectra. They concluded that tip electronic states must be taken into account in interpreting STM and STS data.

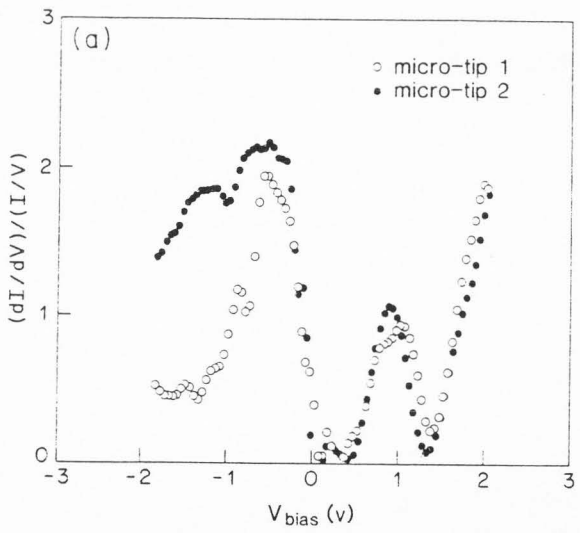

Figure 4. Effect of tip electronic states on tunneling spectra. Tunneling spectra of the same spot on the sample surface with two different microtips are shown. Those microtips are on the same tip shank, about 9 Å apart. Especially on the negative-bias region, the tunneling spectra from two microtips are very different. Obviously, the electronic structure of the two microtips are very different. After Klitsner et al., (1990).

Similar observations have been reported few years earlier by Tromp et al. (1987) and Stroscio et al. (1987). Based on the observation that very sharp tips usually have highly structured electron energy spectrum, Feenstra et al. (1987) developed a method to generate tips with electronic structure that resembles a free-electron metal. It provides reliable tunneling spectra, albeit with low spatial resolution.

To summarize, in STM and STS, the image resolution and tunneling spectra are determined not only by the electronic states of the sample surface but also by the electronic states of the tip. The latter is determined by the configuration of a few atoms at the apex of the tip. A slight change of the chemical identity and geometrical structure of a few atoms at the tip apex would dramatically change the STM images and local tunneling spectra. At atomic resolution, the tip electronic structure usually deviates significantly from the electronic property of the tip bulk and seldomly resembles a free-electron metal.

The One-Sided View of STM

In the early years of STM, a generally accepted theory was the *s*-wave-tip model, proposed by Tersoff and Hamann (1983, 1985). The *s*-wave-tip model attributes the STM images and STS data to the electron density distribution of the sample *only,* independent of tip electronic structure. The motivation and spirit of this one-sided view of STM are clearly stated by an original author in the theoretical part of a well-known review article (Hansma and Tersoff, 1987):

> The theory of tunneling makes no distinction between the surface and the tip. However, in STM, this distinction is crucial. Ideally, one would like to relate the STM image directly to a property of the surface, whereas in any exact analysis the current involves a complicated convolution of the electronic spectra of surface and tip.
>
> Tersoff and Hamann observed that the tip properties can be taken out of the problem by considering a particular model of the tip, motivated as follows. The ideal scanning tunneling microscope would have the greatest possible resolution, and would measure an intrinsic properties of the unperturbed surface, rather than a property of the joint surface-tip system. These goals would be best achieved by a tip whose potential and wave functions were arbitrarily localized. In this case, in the limit of small voltage, the tunneling conductance is
>
> $$\sigma \propto \rho(\mathbf{r}_t, E_F), \tag{1}$$
>
> where $\mathbf{r}_t$ is the tip position, E_F is the Fermi energy, and
>
> $$\rho(\mathbf{r}_t, E_F) \equiv \sum_{\nu} |\Psi_\nu(\mathbf{r})|^2 \delta(E_\nu - E) \tag{2}$$
>
> is the surface local density of states (LDOS) at point $\mathbf{r}$ and energy E.

The primary promise of the *s*-wave-tip model is to predict the greatest possible resolution of STM. However, experimentally, the resolution of STM has greatly exceeded the predictions of the *s*-wave-tip model. Fermi-level local density of states (LDOS) is a familiar quantity in surface science. For many surfaces, its contours have been calculated from first-principles and measured by various methods, such as atom-beam scattering. For nearly free-electron metal surfaces, for example Al(111), at 3 Å from the plane of the top-layer nuclei, the corrugation amplitude of Fermi-level LDOS is less than 0.03 Å. It is impossible to place a tip much closer than 3 Å from the nuclei of the Al atoms of the sample surface. Therefore, the *s*-wave-tip model predicts no atomic resolution on close-packed metal surfaces (Tersoff and Hamann, 1985; Hansma and Tersoff, 1987).

Experimentally, atomic resolution has been observed on every clean close-packed metal surfaces, with nearest-neighbor atomic distance less than 3 Å, such as Au(111), Au(110), Au(001), Al(111), Cu(110), Cu(001), Pt(001), Pt(111), Ru(0001), Ni(110), Ni(001), etc. (see the review article by Behm in *Scanning Tunneling Microscopy and Related Methods,* edited by Behm et al. (1990)). On Al(111), corrugation as large as 0.3 Å is routinely observed. The largest corrugation observed was 0.8 Å, about 20 times greater than the greatest possible corrugation predicted by the *s*-wave-tip model (Wintterlin et al., 1989). In some cases, even the sign of corrugation is reversed from the predictions of the *s*-wave-tip model (Barth et al., 1990, Burne, 1992). In other words, the atomic sites appear as depressions (minima) on the topographic image, rather than protrusions (maxima). The amplitude of such inverse corrugation can be one order of magnitude greater than the greatest possible corrugation amplitude

predicted by the *s*-wave-tip model; thus it cannot be explained by a multiple *s*-wave-tip effect.

The basic assumption in the original papers of Tersoff and Hamann (1983, 1985) is that the tip can be described as a piece of free-electron metal with a hemispherical end. Experimental observations (see last section) have shown that when atomic resolution is observed, the electronic structure of the tip often does not resemble that of a free-electron metal. More precisely, the electronic structure of a tip with atomic resolution often does not resemble that of any bulk metal.

Besides discrepancies with experiments, there is a theoretical difficulty in the *s*-wave-tip model: the Fermi-level LDOS contour is taken at a geometrical point which represents the tip. When atomic resolution is to be explained, the physical definition of that geometrical point becomes obscure. For a macroscopic free-electron-metal tip with a radius of curvature R, the position of the tip $\mathbf{r}_t$ is the center of curvature. It is well defined if the radius of curvature is large enough. For example, in their original paper (Tersoff and Hamann, 1983, 1985), the radius of curvature was 9 Å and the distance from $\mathbf{r}_t$ to the sample surface was 15 Å. At such a distance, atomic resolution is impossible. If the radius is comparable to the size of an atom, the position of the tip, $\mathbf{r}_t$, becomes ill-defined. An alternative definition of the tip position is the position of the nucleus of the apex atom. In this case, the electronic states of the apex atom become important. If the apex atom is an alkali or alkali earth atom, such as Na or Ca, the tip state is almost a pure *s*-wave state, and the STM images would resemble that of the *s*-wave-tip model (Lang 1986). However, Tersoff and Lang (1990) concluded that real tips do not have an electronic structure similar to that of Na or Ca. Rather, the tips are made of transition metals, probably contaminated with atoms from the surface (Si and C are common sample materials). For a Si-atom tip, the *p* state dominates the Fermi-level LDOS of the tip. For a Mo-atom tip, while the *p* contribution is reduced, this is more than compensated for by the large contribution from states of *d* like symmetry. Tersoff and Lang's calculations (1990) show that on an atomic scale, the STM images from a Si, C, or Mo tip can be dramatically different from the images predicted by the *s*-wave-tip model. They show that while atom-sized features on the surface are imaged, the tip electronic structure cannot be taken out from the problem.

Tip Treatment

The importance and the role of tip electronic states in STM are known to experimentalists through the process of *tip treatment.* From the beginning of the STM experimentation, Binnig and Rohrer (1982) realized that the atomic resolution in STM was due to the existence of a few rather sharp microtips generated through the the tip formation process. The extreme sensitivity of the tunneling current versus distance then selects the microtip closest to the sample surface for tunneling. They also reported (Binnig and Rohrer, 1982) several in-situ tip sharpening procedures: by gently touching the tip with sample surface, the resolution was often improved, and the tips thus formed were quite stable; by exposing the tip to high electric fields, of the order of 10^8 V/cm, the tips are often sharpened.

On the contrary, tungsten tips, carefully prepared by electrochemical etching, with a perfectly smooth end of very small radius observed by scanning or transmission electron microscopy, would not provide atomic resolution immediately. Atomic resolution might happen spontaneously by repeated tunneling and scanning for an unpredictable time duration (Demuth et al., 1988).

Various tip sharpening procedures have been demonstrated. The high-field treatment has been commonly used (Wintterlin et al (1989). During a scan, the bias is raised suddenly to −7.5 V (at the sample) and left at this voltage for approximately four scan lines. The tip responds to the voltage jump by a sudden withdrawal by ≈ 30 Å. Subsequently the bias voltage is reduced to its initial value of −500 mV. The tip does not return to its former z position but remains displaced from that by about 25 Å. It is obvious that the tip actually gets longer by about 25 Å. This process turns out to be completely reproducible and in most cases results in tips achieving atomic resolution.

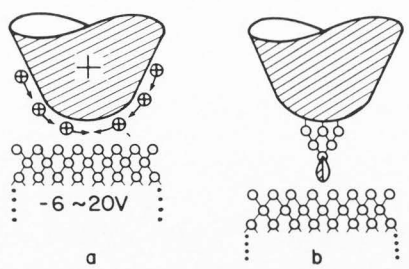

Figure 5. Mechanism of tip sharpening by an electrical field. (a) W atoms on the tip shank walk to the tip apex due to the nonuniform electrical field. (b) A microtip is formed.

At the beginning, the mechanism of such a tip-sharpening process was not well understood (Wintterlin et al., 1989). Two hypotheses were proposed: it is either a restructuring of the tip itself or a transfer of material from the sample to the tip. If the latter is the correct mechanism, its result should depend on the sample material. Later, the same tip sharpening procedure was successfully applied on Ru(0001), Ni(100), NiAi(111), and Au(111), indicating that this phenomenon is not specific for certain surfaces (Behm, 1990). Therefore, there must be a restructuring of the tip, that is, the W atoms move from the shank surface to the apex, as shown in Figure 5.

Tip Electronic Spectrum

The nature of the tip treatment procedure was studied under controlled conditions by Binh and Garcia (1992), see Figure 6. By applying a strong electrical field to a moderately heated W tip, W atoms emit from the tip apex. The direction of the field is the same as in the in-situ tip treatment. The emitted W atoms form an image on the fluorescent screen, similar to those formed by inert gas atoms in the field-ion microscopy. As a result of this atomic metal ion emission process, tips ending with a single W atom are generated.

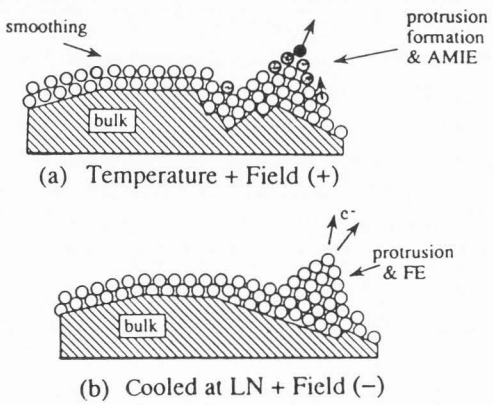

Figure 6. Atomic metallic ion emission and microtip formation. (a) By applying a high field (1.2–1.8 V/Å,), the W atoms move to the protrusions. The apex atom has the highest probability to be ionized and leave the tip. The W ions form an image of the tip on the fluorescence screen. A well-defined pyramidal protrusion, often ended with a single atom, is formed. (b) By cooling down the tip and reversing the bias, a field-emission image is observed on the fluorescence screen. The patterns are almost identical. (After Binh and Garcia, 1992).

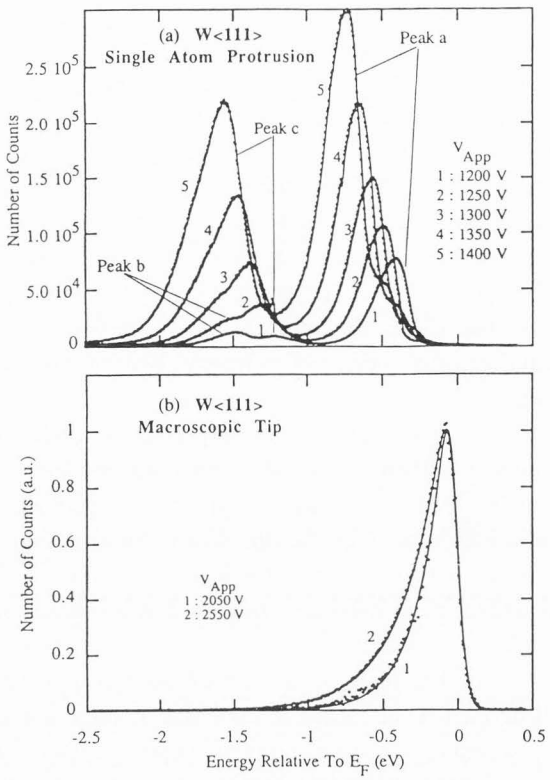

Figure 7. Experimental field-emission spectra (FES) of a W tip with a single protruded atom. (a) The FES of a tip with single-atom protrusion. Well-separated peaks are observed. The position of the peaks vary with the applied voltage. (b) The FES of a macroscopic tip after the single-atom protrusion is destroyed by heating. The FES shows a free-electron-metal behavior. The edge of the peak is always at the Fermi level, independent of the applied voltage. (After Binh et al., 1992).

The electronic structure of the single-atom tips was studied by field-emission spectroscopy (Binh et al., 1992). The results are shown in Figure 7(a). The major findings are:

1. The spectra of the single-atom tips are composed solely of well-separated peaks. The actual position and intensity of the peaks depends on the details of the atomic structure near the apex atom. For comparison, with a macroscopic tip, only one peak is observed right below the Fermi level, see Figure 7(b).
2. The positions of the peaks shift with applied voltage, as shown in Figure 7(a). For comparison, the edge of the single peak for a free-electron

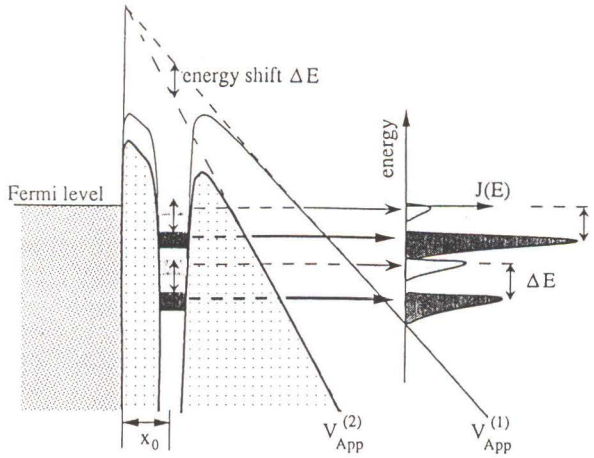

Figure 8. A simplified model of field emission from single-atom tips. The lightly shaded bands signify the positions of two bands for $V_{app}^{(1)}$, and the two darkly shaded bands are the same bands after they were shifted by increasing the applied voltage to $V_{app}^{(2)}$. After Binh et al. (1992).

metal tip does not change with applied voltage, see Figure 7(b).

The field-electron spectra of the single-atom tips strongly suggest that the electrons do not tunnel directly from the Fermi sea to the vacuum. Rather, they come solely from the localized states of the protruding atom. The shift of the peak positions can be understood by a model shown in Figure 8. The direction and the order of magnitude of energy level shift expected from this model are consistent with experimental observations.

Table 1. Electronic configurations of some 5d metals

Atom	Z	5s	5p	5d	6s
Lu	71	2	6	1	2
Hf	72	2	6	2	2
Ta	73	2	6	3	2
W	74	2	6	4	2
Re	75	2	6	5	2
Os	76	2	6	6	2
Ir	77	2	6	7	2
Pt	78	2	6	9	1

Finally, we make some remarks on the nature of these discrete electronic states. Table 1 shows the electronic configurations of eight atoms with adjacent atomic numbers. For W, the 6s levels are below some of the 5d levels. Four of the ten 5d states are occupied. When a W atom is brought to the vicinity of the tip end, the energy levels of the 5d states split, shift in energy, and are broadened. According to the details of the atomic structure, one of the 5d states (m=0, m=1, or m=2) should dominate the Fermi-level LDOS.

Tip Electronic States and STM Resolution

As we have discussed in Section 1, experimentally, atomic resolution has been observed on a large variety of metal and semiconductor surfaces. In order to resolve single atoms, a lateral resolution of 2 Å is required. The importance of the STM — the feature that sets it apart from other instruments — is that it can resolve details in the vicinity of a single atom, otherwise it would not have created the excitement that now surrounds it (Quate, 1986).

The atomic resolution in STM and the large variety of STM images observed for the same sample surface can be explained by the tip electronic structure. Based on the early STM experiments on Si(111)-7 × 7 (Binnig et al. 1982), Baratoff (1983, 1984) proposed that the atomic resolution in STM is probably due to a single dangling bond protruding from the tip. Many transition-metal surfaces, such as W(100), Mo(100), have a strong tendency to form highly localized surface

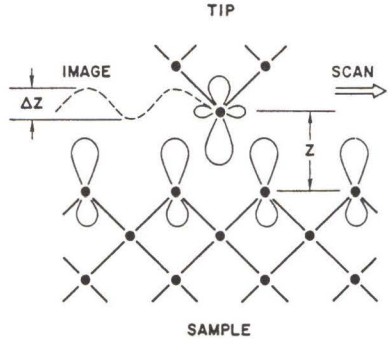

Figure 9. Microscopic view of STM imaging mechanism. An atomic state at the tip end, exemplified by a d_{z^2} dangling bond on a W tip, interacts with a two-dimensional array of atomic states, exemplified by the sp^3 states on Si surface. This results in a highly corrugated tunneling current distribution.

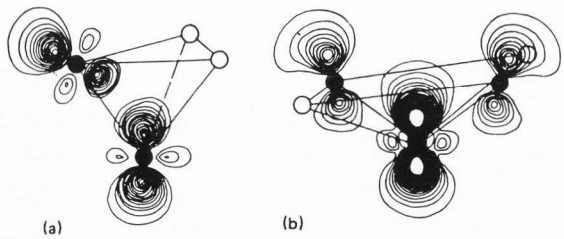

Figure 10. Electronic states on W clusters. The electronic states near the Fermi level on tungsten clusters, W_4 and W_5, calculated by Ohnishi and Tsukuda (1989). At low bias, these *d*-like tip states contribute more than 90% of the tunneling current.

states. (see Weng et al., 1978, and references therein). Especially, the d_{z^2}-like surface state on W(001) surface, is located around the Fermi level. Those localized surfaces states were discovered experimentally by Swanson and Cruiser (1966, 1967) and studied extensively by many authors in the seventies and eighties (see Posternak et al., 1980; Mattheiss and Hamann, 1984; Drube et al., 1986).

To further investigate the STM imaging mechanism, Ohnishi and Tsukada (1989) made an extensive calculation of the electronic states for a number of W clusters. From the calculations, they found that on the apex atom of many W clusters, there is a d_{z^2}-like state protruding from the apex atom, energetically very close to the Fermi level. Using Green's function methods, they also found that the tunneling current is predominately contributed by this *d*-state. Ohnishi and Tsukada (1989) proposed that such an orbital would be advantageous for a sharp STM image. Figure 10 shows the electronic states near the Fermi level on W_4 and W_5 clusters.

Demuth et al. (1988) analyzed the effect of electronic states on the tip based on a number of experimental facts. They emphasized that the tip is one half of the STM experiment and is more difficult to control than the sample surface. Even the best prepared clean tungsten tips usually do not immediately produce the highest resolution on Si surface. When there is no atomic resolution, an effective procedure to achieve atomic resolution is to mildly collide the tip with the Si surface. After such a controlled crashing, a crater is found on the Si surface, which shows that a Si cluster has been picked up by the tip. Atomic resolution is then often achieved. Demuth *et al.* (1988) proposed that at the end of the Si cluster, there is a p_z-like dangling bond protruding from the tip end, which provides atomic resolution.

The effect of p_z or d_{z^2} dangling bonds on STM resolution can be understood in the light of the *reciprocity principle*, which is the fundamental microscopic symmetry between the tip and the sample: by interchanging the "acting" electronic state of the tip and the sample state under observation, the image should be the same. The discrepancy between the sharp STM image and the low corrugation of the charge density on Al(111) can be intuitively explained in the light of the reciprocity principle. Figure 11 shows a qualitative explanation of the effect of a d_{z^2} tip state. For an *s*-wave tip state, the STM image of a metal surface is the charge-density contour, which can be evaluated using *atomic-charge superposition*, i.e., as a sum of the charge densities of individual atoms, each made of s-states. According to the reciprocity principle (see Figure 11), with a d_{z^2} tip state, the tip no longer traces the contour of the Fermi-level LDOS. Instead, it traces the charge-density contour of a *fictitious surface* with a d_{z^2} state on each atom. Obviously, this contour exhibits much stronger atomic corrugation than that of the Fermi-level LDOS.

The Modified Bardeen Approach

Besides the scarceness of experimental data, the lack of a consistent tunneling theory also prevented the formulation of the two-sided view of STM. First, because the tip-sample distance under normal operational conditions of STM is very short, the original Bardeen approach (Bardeen, 1960) is no longer appropriate.

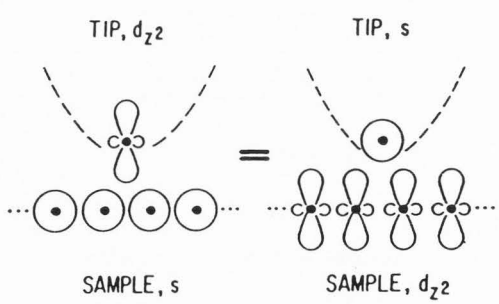

Figure 11. Origin of atomic resolution on metal surfaces. According to the reciprocity principle, the image taken with a d_{z^2} tip state (which exists on a W tip) on a free-electron metal surface is equivalent to an image taken with a point tip on a fictitious sample surface with a d_{z^2} state on each top-layer atom, which obviously has a strong corrugation.

Second, the tunneling matrix elements in three-dimensional space must be evaluated explicitly. Third, systematic methods for calculating STM images have to be developed. After years of trial, a consistent theoretical approach for the two-sided view of STM was gradually formulated.

In this section, we briefly describe a theoretical method for treating the combined effect of tip electronic states and sample electronic states in the imaging process. First, it is shown that by introducing proper modifications, the Bardeen approach of tunneling phenomenon (Bardeen, 1960) can be extended to cases where a strong interaction between the tip and sample exist. A modified Bardeen approach was established. The tunneling current at a bias V is determined by the tunneling matrix elements $|M|$ through the integral

$$I = \frac{4\pi e}{\hbar} \int_0^{eV} \rho_S(E_F - eV + \epsilon)\, \rho_T(E_F + \epsilon)\, |M|^2\, d\epsilon, \quad (3)$$

where $\rho_S(\epsilon)$ is the DOS of the sample, and $\rho_T(\epsilon)$ is the DOS of the tip. Those matrix elements can be expressed as a surface integral, evaluated from the wavefunctions of the tip and the sample (with proper modifications) on a separation surface between them,

$$M = -\frac{\hbar^2}{2m} \int_\Sigma (\chi^* \nabla \psi - \psi \nabla \chi^*) \cdot d\mathbf{S}, \quad (4)$$

where ψ_μ is a wavefunction of the sample, modified by the field of the tip; and χ_ν is a wavefunction of the tip, modified by the field of the sample. By expanding the tip wavefunctions into its spherical harmonic components, we obtain the tunneling matrix elements for all the tip states up to l=2, as shown in Table 2 (Chen, 1993).

The results can be summarized as the *derivative rule:* write the angle dependence of the tip wavefunction in terms of x, y, and z. Replace them with the simple rule,

$$x \to \frac{\partial}{\kappa \partial x}; \quad y \to \frac{\partial}{\kappa \partial y}; \quad z \to \frac{\partial}{\kappa \partial z}; \quad (5)$$

and, acting on the sample wavefunction, up to a constant, the tunneling matrix element is obtained.

Table 2. Tunneling matrix elements

Tip state	Matrix element
s	$C\psi(\mathbf{r}_0)$
p_z	$C\frac{\partial \psi}{\partial z}(\mathbf{r}_0)$
p_x	$C\frac{\partial \psi}{\partial x}(\mathbf{r}_0)$
p_y	$C\frac{\partial \psi}{\partial y}(\mathbf{r}_0)$
d_{zx}	$C\frac{\partial^2 \psi}{\partial z \partial x}(\mathbf{r}_0)$
d_{zy}	$C\frac{\partial^2 \psi}{\partial z \partial y}(\mathbf{r}_0)$
d_{xy}	$C\frac{\partial^2 \psi}{\partial x \partial y}(\mathbf{r}_0)$
$d_{z^2-\frac{1}{3}r^2}$	$C\left[\frac{\partial^2 \psi}{\partial z^2} - \frac{1}{3}\kappa^2 \psi\right](\mathbf{r}_0)$
$d_{x^2-y^2}$	$C\left[\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial y^2}\right](\mathbf{r}_0)$

The STM Images

A one-dimensional case

In this section, we discuss the simplest case: a metal surface of a one-dimensional periodicity a with a reflection symmetry at $x = 0$. The general formula for the electron charge density distribution is

$$\rho(x, z) = C_0 e^{-\alpha z} + C_1 e^{-\beta z} \cos^2(qx), \quad (6)$$

where

$$q = \frac{1}{2} g = \frac{\pi}{a}, \quad (7)$$

and g is the primitive reciprocal lattice vector (Tersoff and Hamann, 1985). The constants C_0, C_1, α, and β are determined by fitting with results from first-principles calculations. The uncorrugated term comes mainly from the $\bar{\Gamma}$ point in the Brillouin zone. The wavefunction, to the lowest Fourier component, is

$$\psi_{\bar{\Gamma}} \propto e^{-\kappa z}, \tag{8}$$

which makes the first term $C_0 e^{-2\kappa z}$. Therefore, we identified the first decay constant in Equation (6),

$$\alpha = 2\kappa. \tag{9}$$

The Bloch functions near the $\bar{K}$ points have a long decay length and contribute to the second term of Equation (6). The lowest-order symmetric Fourier component of the Bloch function near $\bar{K}$ is:

$$\psi_{\bar{K}} \propto e^{-\sqrt{\kappa^2 + q^2}\, z} \cos qx. \tag{10}$$

The charge density is proportional to $|\psi_K|^2$. We then find the Fermi-level LDOS

$$\begin{aligned} \rho(\mathbf{r}, E_F) &= \frac{1}{\epsilon} \sum_{E_\mu = E_F}^{E_F + \epsilon} |\psi_\mu(\mathbf{r})|^2 \\ &= C_0 e^{-2\kappa z} + C_1 e^{-2\sqrt{\kappa^2 + q^2}\, z} \cos^2 qx. \end{aligned} \tag{11}$$

The corrugation amplitude of the Fermi-level LDOS for a metal surface with one-dimensional corrugation can be obtained directly from Equation (11),

$$\Delta z = \frac{C_1}{2\kappa C_0} e^{-\gamma z}, \tag{12}$$

where

$$\gamma \equiv \beta - \alpha = 2\sqrt{\kappa^2 + q^2} - 2\kappa. \tag{13}$$

Using the expressions of the tunneling matrix elements listed in Table 1, theoretical STM images can be calculated. For an s-wave tip state, we recover the result of Tersoff and Hamann (1983, 1985):

$$I \propto \rho(\mathbf{r}_0, E_F). \tag{14}$$

The topographic image is

$$\Delta z(x) = \frac{C_1}{2\kappa C_0} e^{-\gamma z} \cos^2 qx. \tag{15}$$

According to the derivative rule, the tunneling matrix element for surface wavefunction at $\bar{\Gamma}$ from a p_z tip state is identical to that from an spherical-tip tip state. However, for a surface wavefunction at $\bar{K}$, the tunneling matrix element from a p_z tip state is:

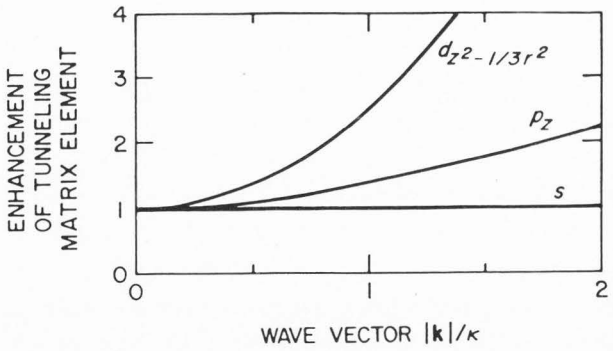

Figure 12. Corrugation enhancement due to p and d tip states. Solid curves, enhancement of tunneling matrix elements arising from $l > 0$ tip states. The tunneling current is proportional to the square of the tunneling matrix element. Therefore, the enhancement factor for corrugation amplitude is the square of the enhancement factor for tunneling matrix element, dotted curves.

$$M_{\bar{K}} \propto \left(1 + q^2/\kappa^2\right)^{1/2} \psi_{\bar{K}}, \tag{16}$$

and the topographic image arising from a p_z tip state is:

$$\Delta z(x) = \left(1 + \frac{q^2}{\kappa^2}\right) \frac{C_1}{2\kappa C_0} e^{-\gamma z} \cos^2 qx. \tag{17}$$

Therefore, the corrugation amplitude arising from a p_z tip state gains a factor of $[1 + (q^2/\kappa^2)]$ over that of the charge density contour, see Figure 12.

Using the expression for the transmission matrix element of a d_{z^2} tip state, for a sample wavefunction at $\bar{\Gamma}$, it picks up a factor 2/3, whereas for a sample wavefunction at $\bar{K}$ it picks up a factor $[(2/3) + (q^2/\kappa^2)]$. Similar to the case of p_z tip state, we find the topographic image to be:

$$\Delta z(x) = \left(1 + \frac{3q^2}{2\kappa^2}\right)^2 \frac{C_1}{2\kappa C_0} e^{-\gamma z} \cos^2 qx. \tag{18}$$

The enhancement for the tunneling matrix element is shown in Figure 12. The enhancement factor for corrugation amplitude, $[1 + (3q^2/2\kappa^2)]^2$, could be sub-

stantial. For example, on most close-packed metal surfaces, $a \approx 2.5\text{Å}$, which implies $q \approx 1.25\text{Å}^{-1}$. An enhancement of 11.2 is expected. Most of the commonly used tip materials are d-band metals, for example, W, Pt, and Ir. As we have shown in Section 5, localized d_{z^2} states often occur on the surfaces. These states can enhance the corrugation amplitude by more than one order of magnitude.

Surfaces with hexagonal symmetry

Probably, the most commonly encountered surfaces in STM experiments belong to the hexagonal lattice system (Chen, 1990, 1992). The close-packed metal surfaces and many cleaved surfaces of layered materials belong to this category. In Figure 13, the structure of a close-packed metal surface is shown. The large dots represents the atoms in the top layer. The circles represents the atoms in the second layer. The small dots are those in the third layer. However, experimentally, it was found that only the atoms in the first layer are observed. Therefore, the surface has a hexagonal symmetry, *p6mm*, which is the highest symmetry in all plane groups. The high symmetry makes the treatment much simpler, since the basic features of the images with the lowest non-trivial Fourier components are determined by symmetry only. In this case, the charge density should have a hexagonal symmetry, i. e., invariant with respect to plane group *p6mm* (see Figure 13). Up to the lowest non-trivial Fourier components, the most general form of surface charge density with a hexagonal symmetry is:

$$\rho(\mathbf{r}) = \sum_{E_F - \Delta E}^{E_F} |\psi(\mathbf{r})|^2 \approx a_0(z) + a_1(z)\, \phi^{(6)}(k\mathbf{x}), \quad (19)$$

where $\mathbf{x} = (x, y)$ and $k = 4\pi/\sqrt{3}\, a$ is the length of a primitive reciprocal lattice vector. A hexagonal cosine function is defined for convenience,

$$\phi^{(6)}(\mathbf{X}) \equiv \frac{1}{3} + \frac{2}{9} \sum_{n=0}^{2} \cos \mathbf{u}_n \bullet \mathbf{X}, \quad (20)$$

where $\mathbf{u}_0 = (0, 1)$, $\mathbf{u}_1 = (-\frac{1}{2}\sqrt{3}, -\frac{1}{2})$, and $\mathbf{u}_2 = (\frac{1}{2}\sqrt{3}, -\frac{1}{2})$, respectively. By plotting it directly, it is clear that the function $\phi^{(6)}(k\mathbf{x})$ has maximum value 1 at each atomic site, and nearly 0 in the space between atoms. The function $[1 - \phi^{(6)}(k\mathbf{x})]$ has minimum value 0 at each atomic site, and nearly 1 in the space between atoms, which describes an inverted corrugation (see Figure 14).

The $a_0(z)$ term in Equation (19) comes mainly from the Bloch functions near $\bar{\Gamma}$, whose lowest Fourier component is:

$$a_0(z) \propto e^{-2\kappa z} \quad (21)$$

The Bloch functions near the $\bar{\mathrm{K}}$ points have the longest decay length, which are the dominating contribution to the second term in Equation (19). In general, a surface Bloch function at that point has the form:

$$\psi_{\bar{\mathrm{K}}} = \sum_{\mathbf{G}} a_{\mathbf{G}} e^{-\sqrt{\kappa^2 + |\mathbf{k}_1 + \mathbf{G}|^2}\, z}\, e^{i(\mathbf{k}+\mathbf{G})\bullet \mathbf{x}} \quad (22)$$

with $|\mathbf{k}_1| \equiv q = k/\sqrt{3}$. By inspecting Equation (22) and Figure 13, one finds that the only slow-decaying symmetric Fourier sums of the Bloch functions near $\bar{\mathrm{K}}$ are:

$$\psi_1 = B\, e^{-\frac{1}{2}\beta z} \sum_{n=0}^{2} \cos(q\mathbf{U}_n \bullet \mathbf{x}), \quad (23)$$

and

$$\psi_2 = B\, e^{-\frac{1}{2}\beta z} \sum_{n=0}^{2} \sin(q\mathbf{U}_n \bullet \mathbf{x}), \quad (24)$$

$$a_0(z) \propto e^{-2\kappa z} \quad (21)$$

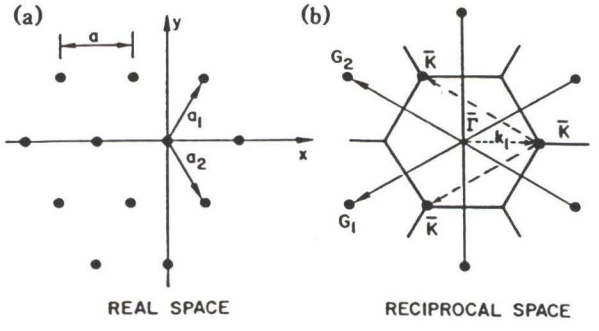

Figure 13. Geometrical structure of a close-packed metal surface. (a) The real space. The top layer exhibits a sixfold symmetry. (b) The reciprocal space. The lowest non-trivial Fourier components of LDOS arise from Bloch functions near the $\bar{\Gamma}$ and $\bar{\mathrm{K}}$ points.

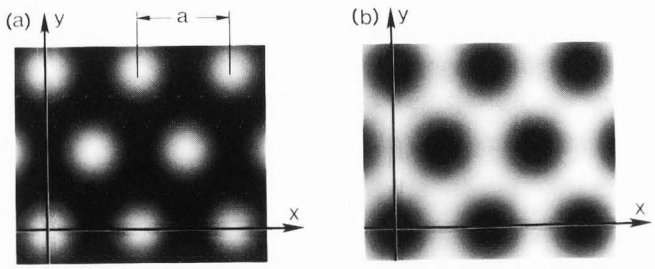

Figure 14. The hexagonal cosine function and its complementary function. The hexagonal cosine function defined by Equation (20), $\phi^{(6)}(k\mathbf{x})$, has maximum value 1 at each atomic site, and nearly 0 in the space between atoms. The function $[1-\phi^{(6)}(k\mathbf{x})]$ has minimum value 0 at each atomic site, and nearly 1 in the space between atoms, which describes an inverted corrugation.

where $\mathbf{U}_0 = (1, 0)$, $\mathbf{U}_1 = (-\frac{1}{2}, \frac{1}{2}\sqrt{3})$, $\mathbf{U}_2 = (-\frac{1}{2}, -\frac{1}{2}\sqrt{3})$; B is a real constant; and $\beta = 2(\kappa^2 + q^2)^{1/2}$ is the corresponding decay constant. The charge density is the sum of Equation (21) and the sum of the charge density proportional to $|\psi_1^*\psi_1| + |\psi_2^*\psi_2|$. A straightforward calculation gives

$$\rho(\mathbf{r}) \propto \sum_{E_F-\Delta E}^{E_F} |\psi(\mathbf{r})|^2 = \Delta E\left[C_0 e^{-2\kappa z} + C_1 e^{-\beta z}\phi^{(6)}(k\mathbf{x})\right], \tag{25}$$

where C_0, C_1 are constants. The corrugation charge-density contour, Δz, as a function of z, can be obtained from Equation (25)

$$\Delta z(x) = \frac{C_1}{2\kappa C_0} e^{-\gamma z}\phi^{(6)}(k\mathbf{x}). \tag{26}$$

Similarly, $\gamma = \beta - 2\kappa$. The ratio (C_1/C_0) can be determined by comparing Equation (26) with the charge-density contours obtained from first-principle calculations or atom-scattering experiments. Following the procedure for the one-dimensional case, the STM image for the p_z tip state is

$$\Delta z(\mathbf{x}) = \left(1 + \frac{q^2}{\kappa^2}\right) \frac{C_1}{2\kappa C_0} e^{-\gamma z}\phi^{(6)}(k\mathbf{x}), \tag{27}$$

and the STM image for the d_{z^2} tip state is

$$\Delta z(\mathbf{x}) = \left(1 + \frac{3q^2}{2\kappa^2}\right)^2 \frac{C_1}{2\kappa C_0} e^{-\gamma z}\phi^{(6)}(k\mathbf{x}). \tag{28}$$

A comparison of the theory with experiments is shown in Figure 15. For Al(111), $a = 2.88\,\text{Å}$, $\phi = 3.5\,eV$, it follows that $\kappa = 0.96\text{Å}^{-1}$, $\beta = 3.48\text{Å}^{-1}$. The slope of the $\ln \Delta z \sim z$ curve from Equation (26) through Equation (28) fits well with experimental data. The absolute tip-sample distance is obtained from curve fitting, which also fits well with the results of first-principle calculations and experimental measurements.

Corrugation inversion

In this sub-section, we discuss the interpretation of the corrugation inversion, reported by Barth *et al.* (1990) and Brune (1992). On the images of clean metal surfaces obtained by STM, in many cases, the sites of metal atoms are minima rather than maxima. Often, with a sudden change of the tip, under the same tunneling conditions, the atomic corrugation switches abruptly from negative to positive. Furthermore, while the atomic corrugation is inverted, the average contour of the large reconstruction remains unchanged. We show that this phenomenon can be explained by the effect of $m \neq 0$ tip states (Chen, 1992).

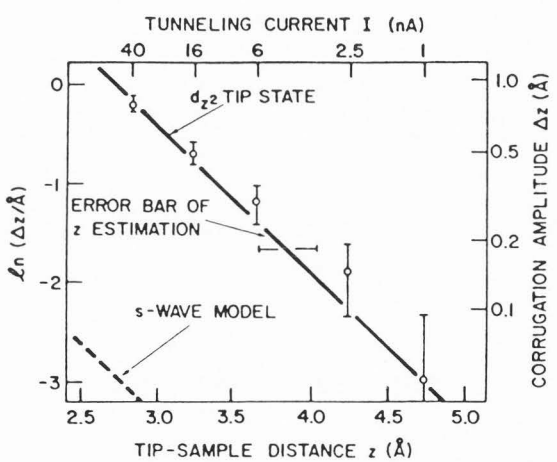

Figure 15. Interpretation of the STM corrugation observed on Al(111). The predicted corrugation amplitude with a d_{z^2} tip state agrees well with experimental data (circles with error bars). The corrugation from an s-wave tip state (dashed curve), is included for comparison (after Chen 1990).

First, we present a heuristic picture of the effect of m≠0 tip states. For simplicity, we assume that the tip has an axial symmetry. In other words, the two m=1 states, xz and yz, are degenerate. Similarly, the two m=2 states, xy and x^2-y^2, are also degenerate. The LDOS of those tip states are shown in Figure 16. To describe the gold surface, we use the elementary picture of Tersoff and Hamann (1985) by assuming that each Au atom has only s-wave states near the Fermi level and that the tunneling current from each Au atom is additive. According to the reciprocity principle, the tip state and the sample state are interchangeable. Consider the current distribution from one Au atom on the sample and a certain tip state. Because the sample atom has s-state only, the tunneling current distribution is proportional to the tip-state LDOS, measured at the center of that Au atom. For a d_{z^2} tip state, it has a sharp peak centered at the Au atom site. The total current distribution is the sum of tunneling current for all the Au atoms at the surface. The sharpness of the tunneling current distribution for the d_{z^2} tip state, compared with that of the s-wave tip state (Figure 16), again illustrates why the d_{z^2} tip state enhances image corrugation. The m=1 and m=2 tip states exhibit a ring-shaped LDOS, as shown in Figure 16. The tunneling current distribution for a single Au atom should be proportional to the tip LDOS, which is ring-shaped (Figure 16). The total current distribution is the sum of the tunneling current for all the Au atoms at the surface. Therefore, with an m≠0 tip state, an inverted STM image for each Au atom should be expected. In other words, with an m≠0 tip state, every site of Au atom at the surface should appear as a depression rather than a protrusion in the STM image.

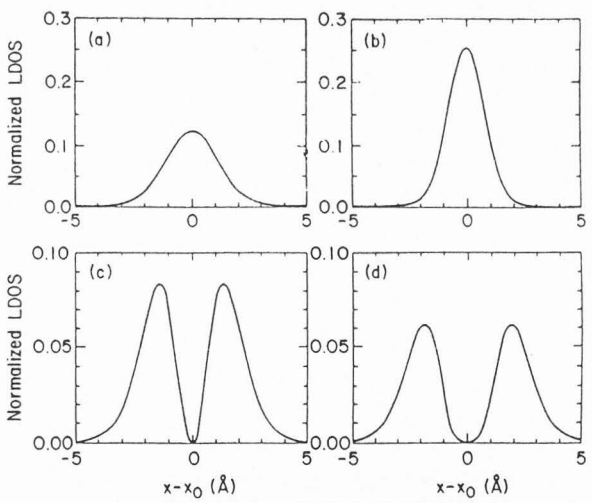

Figure 16. LDOS of several tip electronic states. Evaluated and normalized on a plane $z_0 = 3$Å from the nucleus of the apex atom. An axial symmetry is assumed. (a) s-state. (b) l=2, m=0 state ($d_{3z^2-r^2}$). (c) l=2, m=1 states (d_{xz} and d_{yz}). (d) l=2, m=2 states ($d_{x^2-y^2}$ and d_{xy}).

The general expression for the tunneling current can be obtained using the explicit forms of tunneling matrix elements listed in Table 2. To make the five d-states on the equal footing, normalized spherical harmonics are used. The wavefunctions and the tunneling matrix elements are listed in Table 3.

Up to a constant, the tunneling current is

$$\begin{aligned} I &= 4\left|D_0B_0\right|^2 e^{-2\kappa z} \\ &+ 9\left|D_0B_1\right|^2 e^{-2\kappa_1 z}\left(3(\kappa_1/\kappa)^2-1\right)^2\phi^{(6)}(k\mathbf{x}) \\ &+ 54\left|D_1B_1\right|^2 e^{-2\kappa_1 z}\left(q\kappa_1/\kappa^2\right)^2\left(1-\phi^{(6)}(k\mathbf{x})\right) \\ &+ (27/2)\left|D_2B_1\right|^2 e^{-2\kappa_1 z}\left(q/\kappa\right)^4\left(1-\phi^{(6)}(k\mathbf{x})\right). \end{aligned} \tag{29}$$

The first term in Equation (29) represents the uncorrugated tunneling current, which decays much more slowly than the corrugated terms. Therefore, if D_0 is not too small, the corrugation of the topographic image is

$$\begin{aligned} \Delta z = \Bigg[&\left(\frac{3\kappa_1^2}{2\kappa^2}-\frac{1}{2}\right)^2 - \frac{3}{2}\left|\frac{D_1}{D_0}\right|^2\left(\frac{q\kappa_1}{\kappa^2}\right)^2 \\ &- \frac{3}{8}\left|\frac{D_2}{D_0}\right|^2\left(\frac{q}{\kappa}\right)^4\Bigg]\Delta z_0, \end{aligned} \tag{30}$$

where

$$\Delta z_0 = \frac{9}{2\kappa}\left|\frac{B_1}{B_0}\right|^2 e^{-2(\kappa_1-\kappa)z}\phi^{(6)}(k\mathbf{x}) \tag{31}$$

is the corrugation of the Fermi-level LDOS of the sample. The ratio $|B_1/B_0|$ is determined by first-principles calculations or independent experimental measurements, such as helium atom scattering. For Au(111), $a = 2.87$Å, $q = 1.46$Å^{-1}, $\kappa = 0.96$Å^{-1}, and $\kappa_1 = 1.74$Å^{-1}. From Equation (30), we obtain

Table 3. Wavefunctions and tunneling matrix elements for different d-type tip states. The tip is assumed to have an axial symmetry. For brevity, a common constant factor is omitted.

State	Tip wavefunction	Tunneling matrix element
$3z^2 - r^2$	$D_0\, k_2(\kappa r)\left[3\cos^2\theta - 1\right]$	$D_0\left[\frac{3}{\kappa^2}\frac{\partial^2}{\partial z^2} - 1\right]\psi(\mathbf{r}_0)$
xz	$D_1\, k_2(\kappa r)\sqrt{3}\ \sin 2\theta \cos\phi$	$D_1\left[\frac{2\sqrt{3}}{\kappa^2}\frac{\partial^2}{\partial x \partial z}\right]\psi(\mathbf{r}_0)$
yz	$D_1\, k_2(\kappa r)\sqrt{3}\ \sin 2\theta \sin\phi$	$D_1\left[\frac{2\sqrt{3}}{\kappa^2}\frac{\partial^2}{\partial y \partial z}\right]\psi(\mathbf{r}_0)$
$x^2 - y^2$	$D_2\, k_2(\kappa r)\sqrt{3}\ \sin^2\theta \cos 2\phi$	$D_2\left[\frac{\sqrt{3}}{\kappa^2}\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}\right)\right]\psi(\mathbf{r}_0)$
xy	$D_2\, k_2(\kappa r)\sqrt{3}\ \sin^2\theta \sin 2\phi$	$D_2\left[\frac{2\sqrt{3}}{\kappa^2}\frac{\partial^2}{\partial x \partial y}\right]\psi(\mathbf{r}_0)$

$$\Delta z = \left[19.6 - 11.4\left|\frac{D_1}{D_0}\right|^2 - 2.0\left|\frac{D_2}{D_0}\right|^2\right]\Delta z_0. \quad (32)$$

The enhancement factor E, i.e., the quantity in the square bracket of Equation (32), is displayed in Figure 17. Because the corrugation amplitude depends only on the relative intensities of different components, we normalize it through

$$|D_0|^2 + |D_1|^2 + |D_2|^2 = 1. \quad (33)$$

Naturally, the results can be represented by a diagram similar to a three-component phase diagram, as shown in Figure 17. Several interesting features are worth noting. First, when the $m = 0$ or d_{z^2} state dominates, a large, positive enhancement is expected. The condition for a substantial enhancement is quite broad. For example, when the condition $|D_0|^2 > 1.2|D_1|^2 + 0.2|D_2|^2$ is satisfied, the positive enhancement should be greater than 10, or a full order of magnitude. It is about 15% of the total phase space. To have an enhancement of more than 5, one-third of the total phase space is available. Therefore, the experimental observation of large positive corrugation enhancement should be frequent. Second, when m≠0 states dominate, an inverted corrugation should be observed. Again, the probability for a negative image to occur is large. Actually, when the condition $|D_0|^2 < 0.58|D_1|^2 + 0.1|D_2|^2$ is fulfilled, the image corrugation is inverted. This is about 43% of the total phase space. To have negative corrugations with an enhancement factor of 5 or more, 14% of the total phase space is available. Third, from Equation (32) and Figure 17, it is apparent that the effect of m=1 states in generating inverted corrugation is much stronger than that of m=2

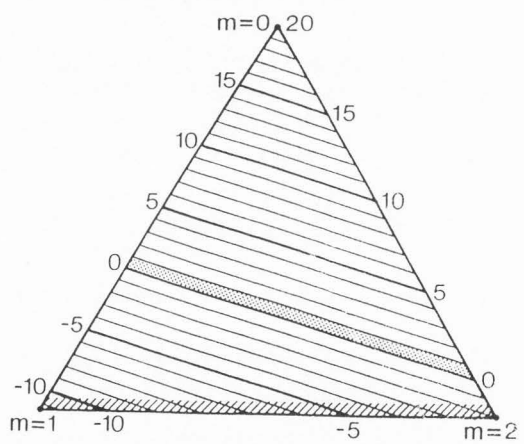

Figure 17. Enhancement factor E for different tip states. The shaded area near E=0 is the area where the corrugation amplitude is within the limit of the Fermi-level LDOS contours. In the hatched area near the bottom, the theoretical amplitude of the negative corrugation shows a spurious divergence.

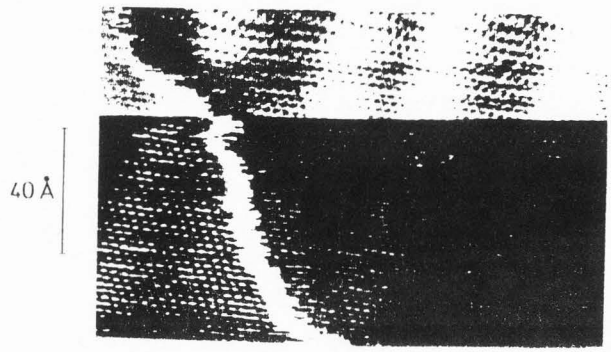

Figure 18. Au(111) topograph taken while a change in the tip state reversed the corrugation. The upper part exhibits a positive corrugation, whereas the lower part exhibits a negative corrugation. Individual Au atoms on both parts are clearly resolved. After Barth et al. (1990).

states. This is expected from Figure 16. The m=1 states have a much sharper rim than the m=2 states. Finally, there is a small region in which an almost complete cancellation of the positive enhancement and the negative enhancement can occur, as indicated by the shaded area near zero corrugation. In this case, the image is similar to the prediction of the *s*-wave model. The observed image corrugation in this case should be equal to or smaller than the corrugation of the Fermi-level LDOS. From Equation (32) and Figure 17, the available phase space is about 2.8% of the total phase space. Therefore, the probability is small. Practically, when this situation occurs, an almost flat image is observed. The experimentalist explains it as a bad tip. A tip sharpening procedure is then conducted until a large corrugation is observed, which is explained as having a good tip.

An experimental observation of the corrugation reversal during a scan is shown in Figure 18. Owing to a sudden change of the tip state, the image switched from positive corrugation to negative corrugation. An interpretation is that before the tip restructuring, an m=0 tip state dominates; after the tip restructuring, an m≠0 tip state dominates.

The corrugation inversion due to m≠0 tip states is a universal phenomenon in the STM imaging of low-Miller index metal surfaces. For example, the same phenomenon was observed on Al(111) by Burne (1992). For most metals (except several alkali and alkali earth metals, which have rarely been imaged by STM), the nearest-neighbor atomic distance $a \approx 3$Å. Consequently, the numerical coefficients on Equation (32) are very close to those for Au(111).

The images of Si(111)2X1

The STM images of the Si(110)2 × 1 were obtained by Feenstra et al. (1987). The structure of the Si(111) 2 × 1 surface is shown in Figure 19. The unit cell dimensions are $a_1 = 6.65\text{Å}$ and $a_2 = 3.84\text{Å}$, in the $[2\bar{1}1]$ and $[01\bar{1}]$ directions, respectively. The lengths of the primitive reciprocal lattice vectors are

$$b_1 = \frac{2\pi}{a_1} \approx 0.945\text{Å}^{-1}, \tag{34}$$

in the $[2\bar{1}1]$ direction, and

$$b_2 = \frac{2\pi}{a_2} \approx 1.636\text{Å}^{-1}, \tag{35}$$

in the $[01\bar{1}]$ direction, respectively.

We now present a simple model for the Si(111) 2 × 1 surface. At a positive polarity, the electrons tunnel into the empty dangling bonds on the A atoms. On each A atom, there is a $3p_z$ state. The origin of the coordinate system is set at one of the A atoms. Three tip states are considered: the *s* state, the p_z state, and the d_{z^2} state. By keeping the leading exponential term only, simple explicit expressions for the Fourier coeffi-

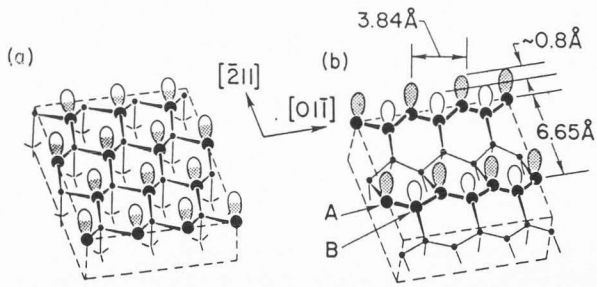

Figure 19. The nascent Si(111) surface and its reconstruction. (a) The nascent Si(111) surface has a three-fold symmetry, with nearest-neighbor atomic distance 3.84Å. (b) The Si(111) surface reconstructs immediately at room temperature to a metastable Si(111) 2 × 1 surface, which has a lower symmetry. Two rows of dangling bond states are formed: one is filled, another one is empty.

cients are obtained. The corrugation functions are then derived.

For convenience, the following parameters are introduced. The decay constant κ of the average (uncorrugated) tunneling current depends on the average energy level of the empty dangling bond states. If it is higher than the Fermi level by ΔE, the decay constant is

$$\kappa = \frac{\sqrt{2m_e(\phi - \Delta E)}}{\hbar}. \tag{36}$$

In the following calculation, the value $\kappa \approx 0.96\,\text{Å}^{-1}$ is taken. The decay constant β of the corrugation component of the tunneling current in the $[01\bar{1}]$ direction and that in the $[2\bar{1}\bar{1}]$ direction are different. We denote them as

$$\beta_1 = \sqrt{4\kappa^2 + b_1^2}, \tag{37}$$

and

$$\beta_2 = \sqrt{4\kappa^2 + b_2^2}. \tag{38}$$

It is also convenient to introduce two more parameters,

$$\gamma_1 = \beta_1 - 2\kappa, \tag{39}$$

and

$$\gamma_2 = \beta_2 - 2\kappa. \tag{40}$$

For an s-wave tip state, the corrugation is

$$\Delta z(x, y) = \frac{16\kappa^2}{\beta_1^3} e^{-\gamma_1 z} \cos^2 \frac{b_1 x}{2} + \frac{16\kappa^2}{\beta_2^3} e^{-\gamma_2 z} \cos^2 \frac{b_2 y}{2}. \tag{41}$$

For a p_z tip state,

$$\Delta z(x, y) = \frac{4}{\beta_1} e^{-\gamma_1 z} \cos^2 \frac{b_1 x}{2} + \frac{4}{\beta_2} e^{-\gamma_2 z} \cos^2 \frac{b_2 y}{2}. \tag{42}$$

And for a d_{z^2} tip state,

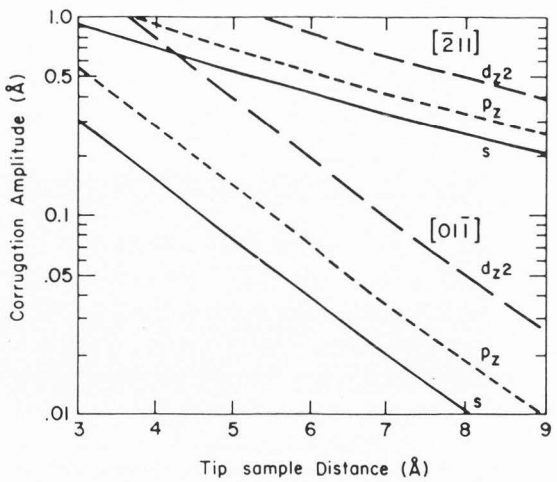

Figure 20. Corrugation amplitudes of the STM images of Si(111)2X1 surface. The dependence of corrugation amplitudes with tip-sample distances are calculated using the independent-state model. The corrugation in the $[2\bar{1}\bar{1}]$ direction is much easier and much less dependent on tip electronic states than the corrugations in the $[01\bar{1}]$ direction.

$$\Delta z(x, y) = \frac{16\kappa^2}{\beta_1^3}\left[\frac{3}{2}\left(\frac{\beta_1}{2\kappa}\right)^2 - \frac{1}{2}\right]^2 e^{-\gamma_1 z}\cos^2\frac{b_1 x}{2} + \frac{16\kappa^2}{\beta_2^3}\left[\frac{3}{2}\left(\frac{\beta_2}{2\kappa}\right)^2 - \frac{1}{2}\right]^2 e^{-\gamma_2 z}\cos^2\frac{b_2 y}{2}. \tag{43}$$

The corrugation amplitudes in the $[01\bar{1}]$ direction and in the $[2\bar{1}\bar{1}]$ direction are displayed in Figure 20. Some general features are worth noting: First, the corrugation in the $[2\bar{1}\bar{1}]$ direction is much easier to observe and much less dependent on tip states than the corrugation in the $[01\bar{1}]$ direction. Second, the decay constants for the corrugations in the two directions are quite different, but are independent of the tip state.

Conclusions

We have shown that in order to understand the contrast mechanism of STM with atomic resolution and the observed scanning tunneling spectroscopy, a two-sided view must be taken. In other words, the tip electronic states must be considered as one half of the problem. We briefly present a theory of the contrast mechanisms of STM in terms of specific electronic states at the tip apex. It provides a quantitative under-

standing of the observed atomic resolution in STM, and provides an insight to the conditions for the atomic resolution to occur.

Acknowledgements

The author wishes to acknowledge inspiring discussions with A. Baratoff, I.P. Batra, N. Garcia, R.J. Hamers, R.M. Feenstra, and J.V. Barth.

References

Bardeen J (1960) Tunneling from a many-body point of view. Phys. Rev. Lett. **6**, 57-59.

Baratoff A (1983) Theory of scanning tunneling microscopy. Europhysics Conference Abstracts **7b**, 364. (Available from Alexis Baratoff, Univ. Basel, Inst. Physik, CH-4056 Basel, Switzerland).

Baratoff A (1984) Theory of scanning tunneling microscopy - Methods and approximations. Physica (Amsterdam) **127B**, 143-150.

Barth JV, Brune H, Ertl G, Behm RJ (1990) Scanning tunneling microscopy observations on the reconstructed Au(111) surface: Atomic structure, long-range superstructure, rotational domains, and surface defects. Phys. Rev. **B 42**, 9307-9318.

Behm RJ (1990) Scanning tunneling microscopy: Metal surfaces, adsorption and surface reactions. In: Scanning Tunneling Microscopy and Related Methods. Behm RJ, Garcia N, Rohrer H (eds.). Kluwer, Dordrecht, 173-210.

Binh VT, Garcia N (1992) On the electron and metallic ion emission from microtips fabricated by field-surface-melting technique: experiments on W and Au tips. Ultramicroscopy **42-44**, 80-90.

Binh, VT, Purcell ST, Garcia N, Doglini J (1992) Field emission spectroscopy of single-atom tips. Phys. Rev. Lett. **69**, 2527-2530.

Binnig G, Rohrer H (1982) Scanning tunneling microscopy. Helv. Phys. Acta. **55**, 726-735.

Binnig G, Rohrer H, Gerber Ch, Weibel E (1982) Tunneling through a controllable vacuum gap. Appl. Phys. Lett. **40**, 178-180.

Binnig G, Rohrer H (1987) Scanning tunneling microscopy - from birth to adolescence. Rev. Mod. Phys. **56**, 615-625.

Brune H (1992) Structur, Reaktivität und elektronische Eigenschaften von Adsorbaten auf einer Au(111) Oberfläche (Structure, reactivity, and electronic properties of Adsorbates on an Au(111) surface. Doctorate Thesis, Free Univ. Berlin, Germany.

Chen CJ (1990) Origin of atomic resolution on metals in scanning tunneling microscopy. Phys. Rev. Lett. **65**, 448-451.

Chen CJ (1992) Effect of $m \neq 0$ tip states in scanning tunneling microscopy: the explanations of corrugation reversal. Phys. Rev. Lett. **69**, 1656-1659.

Chen CJ (1993) Introduction to Scanning Tunneling Microscopy. Oxford University Press, New York.

Ciraci S, Baratoff A, Batra IP (1990) Site-dependent electronic effects, forces, and deformations in scanning tunneling microscopy of flat metal surfaces. Phys. Rev. **B 42**, 7618-7621.

Demuth JE, Hamers RJ, Tromp RM, Welland ME (1986) A scanning tunneling microscope for surface science studies. IBM J. Res. Develop. **30**, 396-402.

Demuth JE, Koehler U, Hamers RJ (1988) The STM learning curve and where it may take us. J. Microsc. **151**, 289-302.

Doyen G, Kotter E, Vigneron JP, Scheffler M (1990) Theory of scanning tunneling microscopy. Appl. Phys. **A 51**, 281-288.

Drube W, Straub D, Himpsel FJ, Soukiassian P, Fu CL, Freeman AJ (1986) Unoccupied surface states on W(001) and Mo(001) by inverse photoemission. Phys. Rev. **B 34**, 8989-8992.

Dürig U, Züger O, Pohl DW (1988) Force sensing in scanning tunneling microscopy: Observation of adhesion forces on clean metal surfaces. J. Microsc. **152**, 259-267.

Feenstra RM, Stroscio JA (1987) Real-space determination of surface structure by scanning tunneling microscopy. Physica Scripta **T19**, 55-60.

Feenstra RM, Stroscio JA, Fein AP (1987). Tunneling spectroscopy of the Si(111)2 X 1 surface. Surface Sci. **181**, 295-306.

Giaever I (1960a) Energy gap in superconductors measured by electron tunneling. Phys. Rev. Lett. **5**, 147-148.

Giaever I (1960b). Electron tunneling between two superconductors. Phys. Rev. Lett. **5**, 464-466.

Giaever I, Megerle K (1961) Study of superconductors by electron tunneling. Phys. Rev. **122**, 1101-1111.

Hansma PK, Tersoff J (1987). Scanning tunneling microscopy. J. Appl. Phys. **61**, Rl-R23.

Klitsner T, Becker RS, Vickers JS (1990) Observation of the effect of tip electronic states on tunneling spectra acquired with the scanning tunneling microscope. Phys. Rev. **B41**, 3837-3840.

Lang ND (1986) Theory of single atom imaging in scanning tunneling microscopy. Phys. Rev. Lett. **56**, 1164-1167.

Mattheiss LF, Hamann D (1984) Electronic structure of the tungsten (001) surface. Phys. Rev. **B 29**, 5372-5381.

Ohnishi S, Tsukada M (1989) Molecule orbital theory for the scanning tunneling microscopy. Solid State Commun. **71**, 391-394.

Posternak M, Krakauer H, Freeman AJ, Koelling DD (1980) Self-consistent electronic structure of surfaces: Surface states and surface resonanceson W(001). Phys. Rev. **B 21**, 5601-5612.

Quate CF (1986) Vacuum tunneling: A new technique for microscopy. Physics Today, August 1986, 26-33.

Rohrer H (1992) STM: 10 years after. Ultramicroscopy, **42-44**, 1-6.

Soler JM, Baro AM, Garcia N, Rohrer H (1986) Interatomic forces in scanning tunneling microscopy: Giant corrugations of the graphite surface. Phys. Rev. Lett. **57**, 444-447.

Stroscio JA, Feenstra RM, Fein AP (1987) Imaging electronic surface states in real space on the Si(111)2x1 surface. J. Vac. Sci. Technol. **A5**, 838-845.

Swanson LW, Crouser LC (1966) Anomalous total energy distribution for a tungsten field emitter. Phys. Rev. Lett. **16**, 389-392.

Swanson LW, Crouser LC (1967) Total energy distribution of field-emitted electrons and single-plane work functions for tungsten. Phys. Rev. **163**, 622-641.

Tersoff J, Hamann DR (1983) Theory and application for the scanning tunneling microscope. Phys. Rev. Lett. **50**, 1998-2001.

Tersoff J, Hamann DR (1985) Theory of scanning tunneling microscope. Phys. Rev. **B 31**, 805-814.

Tersoff J, Lang ND (1990) Tip-dependent corrugation of graphite in scanning tunneling microscopy. Phys. Rev. Lett. **65**, 1132-1135.

Tromp RM, Hamers RJ, Demuth JE, Lang ND (1988) Tip electronic structure in scanning tunneling microscopy. Phys. Rev. **B 37**, 9042-9045.

Weng SL, Plummer EW, Gustafsson T (1978) Experimental and theoretical study of the surface resonances on the (100) faces of W and Mo. Phys. Rev. **B 18**, 1718-1740.

Wintterlin J, Wiechers J, Brune H, Gritsch T, Hofer H, Behm RJ (1989) Atomic-resolution imaging of close-packed metal surfaces by scanning tunneling microscopy. Phys. Rev. Lett. **62**, 59-62.

Discussions with Reviewers

R.S. Becker: In the **Introduction**, which purports to quickly review the history of the STM, there are some historical inaccuracies which should be included and cited. First, the author should mention that important work with the Topografiner, the immediate ancestor of the vacuum STM, was carried out at NBS in the late 60's by Russell Young and coworkers. In fact, in a 1971 paper [Young, Ward, Scire, Phys. Rev. Lett. **27**, 922 (1971)] these workers reported the first observations of metal-vacuum-metal tunneling, a decade before the refining work of Binnig *et al.* The STM was not invented, but rather developed, by Binnig *et al.*

Author: A complete disposition of its history is out of the scope of this review article. As a former science historian (1978-1979), I have looked into this question seriously. I have scanned through all the available literature, and discussed with several key scientists. The major findings are documented in Section 1.8, Historical Remarks, of my recent book *Introduction to Scanning Tunneling Microscopy* (Chen, 1993). I started that section with a general statement "All major scientific discoveries have prior arts, and STM is no exception." I characterized the Topografiner as the "closest ancestor of the STM". I reproduced a Figure from a 1971 Physical Review Letter of Young *et al.* on page 47 of my book. Although Young and his co-workers did not conduct a tunneling experiment with a controllable gap thus to explore the exponential dependence of tunneling current with gap width, and did not conduct tunneling with scanning, I concluded: "Their results indicate that the implementation of metal-vacuum-metal tunneling with scanning is feasible. It is regrettable that their project was discontinued at that time". However, the statement that "STM was invented by Binnig and Rohrer", which occurs very often in the literature, is still valid. Inventorship is a legal concept. Binnig and Rohrer are recognized by the US Department of Commerce as the inventors of US Patent 4,343,993, entitled "Scanning Tunneling Microscope", issued August 10, 1982. According to the US Patent Law, anyone may challenge the inventorship of a patent during one year of period after the patent is issued. Record shows that no one has challenged their inventorship during that time. Therefore, legally, their inventorship is valid. On the STM Patent, the 1970 paper of Young *et al.* (The Topografiner) is cited as the first reference. Obviously, both Binnig and Rohrer and the US Patent Bureau have scrutinized the work of Young *et al.* and recognized their contribution as the most significant prior art, but not as the inventors of STM.

G. Doyen: The author explains the large experimentally observed corrugation amplitude on Al(111) by postulating tip-sample distances between 2.5 and 4.5 Å (Figure 16). At these distances tip-sample interaction must be of extreme importance. Also, at these distances, the tunnel resistance will be only slightly above the quantum limit of 12906 Ω, whereas in experiment, values between 1 to 10 MΩ apply. There is ample evidence in the literature that tip-sample interaction can change the STM images qualitatively at those close separations.

Author: A tunneling resistance of 1-10 MΩ and a resistance of 0.013 MΩ are qualitatively different. Experimentally, a mechanical contact between the tip and the

sample occurs at a tunneling resistance $R_T \approx 0.1$ MΩ (Dürig *et al.*, 1988). The tunneling resistance changes about one ordèr of magnitude per Å of tip-sample separation. The value 0.1 MΩ clearly divides the operational condition of STM into two regimes. As $R_T > 1$ MQ, the STM is working in the tunneling mode. There is at least a displacement of 1 Å away from a mechanical contact. When RT < 0.1 MΩ, there is no STM, because the tip is in a firm mechanical contact with the sample surface. At a resistance $R_K = 0.013$ MΩ, the tip pushes about 1 Å into the sample surface. There is a large repulsive force between the tip and the sample. Because no scanning can be executed on a rigid surface as Al(111), the phenomena in this regime are unrelated to STM experiments.

In the tunneling regime, where $R_T > 1$ MΩ, tip-sample interactions should still exist. There are two kinds of interactions. First is the polarization (or van der Waals) interaction, which alters the wavefunctions of both parties. It is a long-range interaction which has little effect on atomic resolution. The effect of the van del Waals interaction to the tunneling current can be adequately described by a constant multiplier. The second is the exchange interaction which generates an attractive force. It may cause a substantial deformation on soft surfaces. On metal surfaces, the effect of deformation is much smaller than the effect of different tip states, and does not explain the observed corrugation (Ciraci *et al.*, 1990). On soft surfaces, such as graphite, the deformation of the sample surface is the prime factor of corrugation amplification (Soler *et al.*, 1986). The details of this topic are out of the scope of this review article.