Microscopic view of scanning tunneling microscopy

C. Julian Chen
IBM Thomas J. Watson Research Center Yorktown Heights, New York 10598

(Received 6 August 1989; accepted 20 July 1990)

We present a theory of the atomic resolution in scanning tunneling microscopy (STM) in terms of localized surface states on the tip. The tunneling matrix elements arising from these tip states are evaluated with the *derivative rule*. For example, a p_z surface state on the tip generates a tunneling matrix element proportional to $[\partial\psi/\partial z]$ at the nucleus of the apex atom, and a $d_{3z^2-r^2}$ tip state generates a tunneling matrix element proportional to $[3\partial^2\psi/\partial z^2 - \kappa^2\psi]$, ($\psi$ is the sample wave function, κ is the decay constant of surface wave function, $\kappa = \sqrt{2m_e\phi/\hbar}$). To obtain analytic results of theoretical STM images, we further developed a simple independent-orbital model to describe the wave functions of the sample surface. With this model, we present qualitative and quantitative explanations of the observed atomic resolution on metals and semiconductors, the spontaneous switching of instrument resolution during imaging, and various tip-sharpening procedures.

I. INTRODUCTION

To the science community, the value of scanning tunneling microscopy (STM) lies in its unique capability to visualize individual atoms and individual localized electronic states on solid surfaces. Since its invention by Binnig and Rohrer,[1] STM experimentalists have observed atomic details of literally every kind of conducting solid surfaces,[2] including the "smoothest" ones.[3] The physics of its imaging mechanism is an intriguing scientific puzzle by itself, as Binnig and Rohrer wrote in 1982[1]:

> The STM graphs are not simply reproductions of surface–atom corrugation. They rather reflect the corrugation of the electron wave functions near the Fermi level, roughly in the middle of the vacuum gap. To make a quantitative connection with the corrugation of the surface cores, a detailed tunnel theory on the atomic scale is required, ...

The experimental observation of sharp atomic images with STM is infrequent. It requires an extraordinary "sharp tip," whose formation is not completely reproducible, and whose nature is not well understood.[4] Immediately after the first STM experiments, Baratoff remarked that its atomic resolution is probably due to *a single localized surface state, or dangling bond, at the apex of the tip.*[5] For example, tungsten, the most important tip material, has a strong tendency to form highly localized d_{z^2} dangling bonds on its surfaces.[6,7] On the sample side, it is known that on each top-layer atom at clean silicon surfaces, there is an sp^3 dangling bond.[8] Therefore, the atom-resolved STM image on silicon is a result of an interplay between a single dangling bond at the tungsten tip and the two-dimensional array of dangling bonds on the silicon surface, as shown in Fig. 1. In principle, the STM image can be calculated based on this concept. For simple crystalline surfaces, the image can be expressed as a product of a profile function which depends on the structure of the surface, and a corrugation amplitude Δz which is a function of the tip–sample distance z (Fig. 1).

In spite of the attractiveness of this concept, its quantification takes several years to work out due to the conceptual and mathematical difficulties.[5] In this paper, we will discuss the concept of localized surface state at the tip, and develop a simple formalism to calculate the theoretical STM images, which can be used as a connection between first-principle calculations of surfaces of interest and experimental STM images. Details are presented for an elemental surface with hexagonal symmetry, probably the most frequently encountered surface structure, such as close-packed metal surfaces, many layered materials, Si(111), Si(111)–Al $\sqrt{3}\times\sqrt{3}$, etc.

II. LOCALIZED SURFACE STATES ON THE TIP

Every experimentalist knows that atomic resolution in STM is infrequent. Even if the most careful preparation of a tip is made, for example, by chemical etching or mechanical methods, atomic resolution is not guaranteed. In addition,

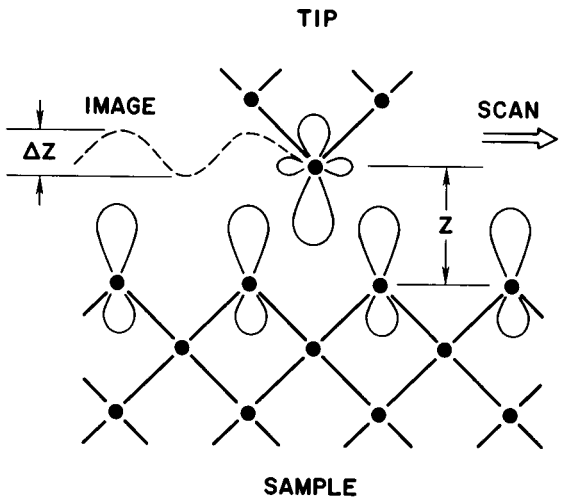

FIG. 1. Microscopic view of STM imaging mechanism. A tip with a dangling-bond state scans over the sample surface, composed of a two-dimensional array of atomiclike states. The overlap of the tip state with the atomic-like states on the sample generates a tunneling conductance, which depends on the relative position of the tip state and the sample state. By keeping the tunneling conductance constant, a topographic image is generated. The atomic corrugation depends on the spatial distribution and the type of atomiclike states on the tip as well as on the surface.

 0734-2101/91/010044-07$01.00

from the very beginning of STM experiments, it was found that the STM resolution undergoes unexpected switching for no obvious reason.[4,5] Sometimes, the corrugation suddenly changes by a factor of two, four, or even one order of magnitude during a single scan. In earlier years, a STM experimentalist may have waited for several hours or even days for a "good tip" to occur and a "good image" to be observed. This phenomenon, the *spontaneous switching of instrument resolution*, is characteristic of STM. The frequency of such resolution flipping indicated that a very subtle change in the tip would induce a tremendous difference in instrument resolution. In terms of localized surfaces state, this behavior can be easily understood, because the conditions for localized surface states to occur is equally subtle.

In the recent years, many STM experimentalists have reported tip-sharpening procedures.[3,4] In the following, we will show that these procedures can be interpreted as methods to form localized surface states on the tip.

First case is the tip-treatment procedure reported by Wintterlin *et al.*[3] As emphasized by the authors, "the most significant single parameter for the observed corrugation apparently related to the specific procedure of tip preparation" was to raise the bias voltage to -7.5 V (at the sample) and to leave for approximately four scan lines. The tip responded to the voltage and jumped by a sudden withdrawal by approximately 30 Å. After that, atomic resolution was frequently achieved. An important observation is that this tip sharpening only works with a positive bias at the tip. An explanation of this fact is shown in Fig. 2. In the bulk, the positive tungsten ions are imbedded in a negative electron cloud. By applying an electrical field in the metal, the ions feel a force aligned with the electrical field intensity vector. However, the electron cloud exerts an "electron wind force" of equal magnitude but opposite direction. Therefore, the tungsten atoms in the bulk feel a zero net force.[9] For tungsten atoms on the surface, the situation is different. Because the electron density is more diluted, the electron wind force is reduced, the tungsten atoms exhibit a net positive charge. Therefore, the electrical field in the gap, with polarity as indicated in Fig. 2, attracts tungsten ions towards the apex. As a result, a tungsten cluster is formed on the tip, often with a single atom at its apex. Another way of explaining this phenomenon is to conceive the gap under a strong electrical field as a miniature plasma. The electrons and positive metal ions are the carriers. The W^+ ions move towards the sample, and the electrons move towards the tip. The possibility to form negative aluminum ions is very slim, so a flow of aluminum to the tip is unlikely. The cluster thus formed should not be very stable, which is consistent with the experimental observations.[3] The electronic states of tungsten clusters have been calculated using first-principle methods[7] by Ohnishi *et al.* They show that at the apex atom of tungsten clusters W_4 and W_5, there is a localized metallic dangling bond state which can well be described as a d_{z^2} state. Using Green's function methods, they have also shown that by taking this tungsten cluster as the tip, the tunneling conductance is predominately generated by that d_{z^2} state protruding from the apex atom.

Another case is the tip-sharpening procedure reported by Demuth *et al.*,[4] as shown in Fig. 3. By gently colliding the tip with silicon surface, they observed that the tip picked up a silicon cluster, as evidenced by a hole on the silicon surface at the contact location, and atomic resolution is frequently achieved. Demuth *et al.*[4] also proposed that an sp^3 dangling bond is formed at the apex silicon atom thus to make the tip probably even sharper than a geometrical point. Numerous first-principle calculations and STM/STS experiments have shown the existence of that dangling bond. The energy levels of those dangling bonds are shown to cross the Fermi level.[8]

III. CONCEPTUAL BACKGROUND

In spite of the attractiveness of this model of STM imaging (in terms of localized surface states on the tip and on the sample), in the early years, there had been a number of conceptual and mathematical difficulties preventing the establishment of a tractable and consistent theory. We now address these issues briefly as follows.

First, the tip–sample distance under normal STM operation is very short. As determined by direct measurements, it is 1–4 Å before a mechanical contact.[10] In this range, the potential barrier between the tip and the sample is often lower than the Fermi level. The transport process is no longer tunneling. Tunneling theories, such as Wentzel–Kramers–Brillouin (WKB) and transfer-Hamiltonian, are no longer adequate.[11] New theories to calculate the current are needed. One solution to this problem is as follows.[12] In STM, the

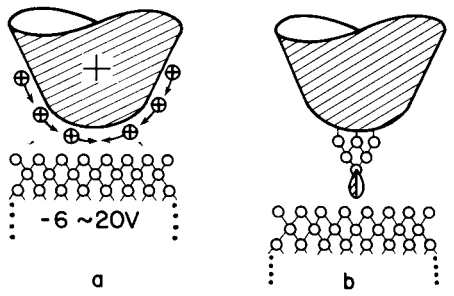

FIG. 2. Tip sharpening by applying a high-voltage bias. (a) tungsten atoms migrate to the end of the tip, 2(b), a tungsten cluster with a d_{z^2} dangling bond at the apex atom.

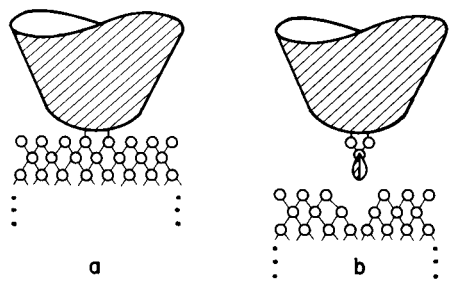

FIG. 3. (a) Tip sharpening by mildly colliding the tip with silicon surface, (b), a silicon cluster with a p_z dangling bond is formed, leaving a hole on the silicon surface.

potential barrier is always thinner than the nominal de Broglie wavelength, as a consequence of the uncertainly principle, the distinction between tunneling and channeling disappears. After the distortion of the sample wave function due to the existence of the tip is taken into account, Bardeen's surface integral for tunneling matrix element,[11]

$$M = -\frac{\hbar^2}{2m_e}\int [\chi^*\nabla\psi - \psi\nabla\chi^*]\cdot d\mathbf{S} \tag{1}$$

becomes valid even if the barrier becomes lower than the Fermi level.[12] In Eq. (1), ψ and χ are the corrected (or distorted) sample and tip wave function, respectively, and the integral is carried over a separation surface roughly in the middle of the gap. To be exact, it is a *transmission* matrix element. The distortion correction to the free sample and tip wave functions, for metals, is a constant over the entire range of normal STM operation. Therefore, except for special cases such as graphite, we don't have to worry about it too much.[12]

Second, the mathematical problem of evaluating the Bardeen integral between two sets of localized surface states is not trivial. Actually, even if a single dangling bond at the tip is considered, at least the nine lowest components ($s, p_x, p_y, p_z, d_{3z^2-r^2}, d_{xy}, d_{xz}, d_{yz}$, and $d_{x^2-y^2}$) have to be included. Recently, it was proved that the transmission matrix elements are determined by the derivatives of the sample wavefunction ψ at the center of the tip $\mathbf{r}_0$ following an extremely simple *derivative rule*.[13] For example, a p_z tip state generates a transmission matrix element as

$$M \propto \kappa^{-1}\partial\psi(\mathbf{r}_0)/\partial z \tag{2}$$

and a $d_{3z^2-r^2}$ tip state generates a transmission matrix element as

$$M \propto \kappa^{-2}\partial^2\psi(\mathbf{r}_0)/\partial z^2 - (1/3)\psi(\mathbf{r}_0). \tag{3}$$

The decay constant κ is determined by the work function ϕ

$$\kappa = \sqrt{2m_e\phi}/\hbar. \tag{4}$$

Based on this derivative rule, an interpretation of the atom-resolved images observed on Al(111) is made, using surface Bloch wave methods.[14] In this paper, we will present a simple model of STM imaging following the intuitive concept of the interplay between one localized surface state at the tip and a two-dimensional array of localized surface states on the sample.

IV. IMAGING A LOCALIZED SURFACE STATE

In this section, we present an approximate method of describing the localized surface states on the sample, and calculate the tunneling conductance distribution for a single localized surface state on the sample. The Slater functions[14] are used to describe the localized surface states

$$\psi_{nlm} = Cr^{n-1}e^{-\zeta r}Y_{lm}. \tag{5}$$

Here, n is the principle quantum number, l and m are angular quantum numbers, and ζ is the orbital exponent. In this paper, we deal with $m = 0$ states only. The spherical harmonics Y_{lm} reduces to Legendre polynomials, $P_l(\cos\theta)$, where $\cos\theta = z/r$. For example,

$$P_0(\cos\theta) = 1, \tag{6}$$

$$P_1(\cos\theta) = \cos\theta, \tag{7}$$

$$P_2(\cos\theta) = \tfrac{1}{2}[\cos^2\theta - 1/3]. \tag{8}$$

For surface states near the Fermi level, ζ equals to the decay constant κ we introduced in Eq. (4).

The spirit of Slater's approximate atomic wave function is to use a single power factor r^{n-1} to represent the entire algebraic factor in the hydrogen-like wave functions. The reason is that except for the very center of the atom, only the highest power of r in the algebraic factor is significant. In treating STM-related problems, we are only interested in the values of atomic wave functions a few angstroms away from the nucleus of the atom, not the values in the core. Therefore, the Slater functions should provide an adequate theoretical description. For the same reason, we can derive all the Slater functions from a single function

$$\psi_{000} = Cr^{-1}e^{-\kappa r} \tag{9}$$

through differentiation *acting only on the exponential factor.* This rule conforms exactly to the spirit of Slater's approximation. Actually, any differentiation to the algebraic factor with r^{n-1} creates a term with r^{n-2}, which is appreciable only near the center of the atom, so it can be neglected. For example, using Eqs. (6)–(8), it is easy to verify that:

$$\psi_{n00} \simeq [-\partial/\partial\kappa]^n\psi_{000}, \tag{10}$$

$$\psi_{n10} \simeq (-\partial/\kappa\partial z)[-\partial/\partial\kappa]^n\psi_{000}, \tag{11}$$

$$\psi_{n20} \simeq \tfrac{1}{2}[3(\partial/\kappa\partial z)^2 - 1][-\partial/\partial\kappa]^n\psi_{000}. \tag{12}$$

The sign $\simeq$ indicates that the differentiation is acting on the exponential factor only.

Now, we use the derivative rule to calculate the tunneling matrix elements from these Slater wavefunctions. The tunneling conductance $g(\mathbf{r})$ is proportional to the square of the tunneling matrix element[11]:

$$g(\mathbf{r}) \propto |M|^2. \tag{13}$$

Because the equal-conductance contour is independent of an overall constant, we neglect it for convenience. Table II lists the results for three types of $m = 0$ tip states, and three types of sample states. The relevant quantity for predicting experimental images is the apparent size of the atom. By apparent size we mean the apparent radius of the image of the sample state of interest. It is convenient to define the apparent radius by

$$R = \frac{\partial g(\mathbf{r})}{\partial z}\left(\frac{\partial^2 g(\mathbf{r})}{\partial x^2}\right)^{-1}. \tag{14}$$

For a spherical distribution, this definition coincides with the actual radius. As shown in Table I, the apparent radius is reduced for p and d states on the tip as well as on the sample. In other words, with p and d states, the images of an atom looks much sharper than s states, which is expected. The images for mixed states, e.g., sp^3 states, can be treated using the same method.

Not surprising, we observed the *reciprocity principle* again:[13] by interchanging the tip state and the sample state, the conductance distribution, and consequently, the apparent size of the image, is unchanged.

TABLE I. Conductance distribution function $g(\mathbf{r})$ for different tip states and sample states (as single localized surface states). For the definition of the apparent radius of the atom, see the text. A relation $\cos\theta = z/r$ is implied.

Tip state	Sample state	Conductance distribution	Apparent radius
s	s	$\exp(-2\kappa r)$	z
s	p	$\cos^2\theta\exp(-2\kappa r)$	$z/[1+1/\kappa z]$
s	d	$[\cos^2\theta-(1/3)]^2\exp(-2\kappa r)$	$z/[1+3/\kappa z]$
p	s	$\cos^2\theta\exp(-2\kappa r)$	$z/[1+1/\kappa z]$
p	p	$\cos^4\theta\exp(-2\kappa r)$	$z/[1+2/\kappa z]$
p	d	$[\cos^2\theta-(1/3)]^2\cos^2\theta\exp(-2\kappa r)$	$z/[1+4/\kappa z]$
d	s	$[\cos^2\theta-(1/3)]^2\exp(-2\kappa r)$	$z/[1+3/\kappa z]$
d	p	$[\cos^2\theta-(1/3)]^2\cos^2\theta\exp(-2\kappa r)$	$z/[1+4/\kappa z]$
d	d	$[\cos^2\theta-(1/3)]^4\exp(-2\kappa r)$	$z/[1+6/\kappa z]$

V. IMAGING CRYSTALLINE SURFACES

In this section, we present an analytic model for crystalline surfaces by assuming that the localized surface states on different atom sites are independent. In other words, we assume that the total tunneling current can be approximated as the sum of the tunneling currents from individual orbitals. This *Ansatz* is not always valid. For example, it does not work for solids exhibiting charge density waves. However, for many metal and semiconductor surfaces, it provides an adequate description.

As shown in Fig. 4, the total conductance G of a crystalline surface, considered as a two-dimensional array of localized orbitals, is a periodic function with primitive vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. Therefore, it can be expanded into a two-dimensional Fourier series,

$$G(\mathbf{r}) = \sum_{n,m=-\infty}^{\infty} g(\mathbf{r}+n\mathbf{a}_1+m\mathbf{a}_2) = \sum_{j,k=-\infty}^{\infty} \tilde{G}_{jk}(z)\exp[i(j\mathbf{b}_1+k\mathbf{b}_2)]. \quad (15)$$

The Fourier coefficients are

$$\tilde{G}_{jk}(z) = [\mathbf{a}_1\cdot\mathbf{a}_2]^{-1} \times \int d^2\mathbf{x}\, G(\mathbf{r})\exp[-i(j\mathbf{b}_1+k\mathbf{b}_2)\cdot\mathbf{x}], \quad (16)$$

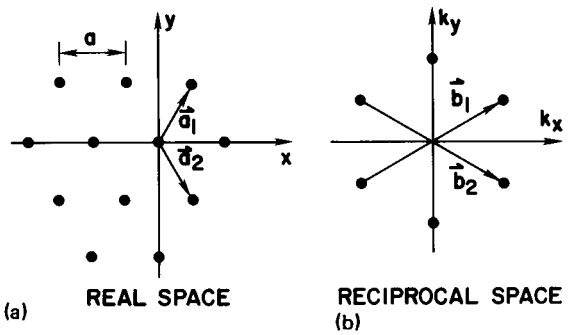

FIG. 4. Elemental surface with hexagonal symmetry. (a) real space, (b) reciprocal space. The length of the primitive reciprocal vector is $b = 4\pi/\sqrt{3}a$.

where $\mathbf{x} = (x,y)$, and the integration extends to the entire surface.

In the following, we will concentrate on the case of elementary crystalline surfaces with hexagonal symmetry, in other words, simple structures belonging to plane group $p6mm$.[16] Only the lowest nontrivial Fourier components are considered. An inspection into Fig. 4 reveals that only the Fourier components at $\mathbf{k} = 0$ and the six equivalent points are significant. Because of the axial symmetry of the conductance function $g(\mathbf{r})$ on each site, the Fourier coefficients at these six points, i.e., $\tilde{G}_{1,0}(z)$, $\tilde{G}_{-1,0}(z)$, $\tilde{G}_{0,1}(z)$, $\tilde{G}_{0,-1}(z)$, $\tilde{G}_{1,1}(z)$, and $\tilde{G}_{-1,-1}(z)$, are equal. We denote it as $\tilde{G}_1(z)$. Up to this term,

$$G(\mathbf{r}) = \tilde{G}_0(z) + (9/2)\tilde{G}_1(z)\phi^{(6)}(b\mathbf{x}), \quad (17)$$

where b is the magnitude of the primitive reciprocal vectors, $\mathbf{b}_1$ and $\mathbf{b}_2$. An inspection to Fig. 4 gives $b = 4\pi/\sqrt{3}a$. A hexagonal cosine function is defined for convenience as

$$\phi^{(6)}(\mathbf{X}) \equiv \frac{1}{3} + \frac{2}{9}\sum_{j=0}^{2}\cos\omega_j\cdot\mathbf{X}, \quad (18)$$

where $\omega_0 = (0,1)$, $\omega_1 = \left(-\frac{1}{2}\sqrt{3}, -\frac{1}{2}\right)$, and $\omega_2 = \left(\frac{1}{2}\sqrt{3}, -\frac{1}{2}\right)$, respectively. It is easy to show that the function $\phi^{(6)}(b\mathbf{x})$ has maximum value 1 at each atomic site, and minimum value 0 at the center of each atomic triangle. [The constant 1/3 in Eq. (18) is introduced for convenience. It makes a slight shift of the origin of z, which is negligible in this approximation.]

Now, we look for a theoretical expression of the topological STM image in the form:

$$z = z_0 + \Delta z\phi^{(6)}(b\mathbf{x}). \quad (19)$$

Since the second term in Eq. (17) is much smaller than the first term, it is easy to show that:

$$\Delta z(z_0) = -\frac{9\tilde{G}_1(z_0)}{2\partial\tilde{G}_0(z_0)/\partial z_0}. \quad (20)$$

Therefore, the problem of calculating the topological images reduces to the problem of evaluating the Fourier coefficients, Eq. (16). From Table I, we observed that what we need to calculate are the Fourier coefficients for the functions

$$f(\mathbf{r}) = r^{n-1}\cos^m\theta\exp(-2\kappa r). \quad (21)$$

We start with the mathematical identity

$$I \equiv \int\int dx\, dy\, r^{-1}\exp(-2\kappa r)\exp(i\alpha x + i\beta y)$$

$$= 2\pi\gamma^{-1}\exp(-\gamma z), \tag{22}$$

where $r^2 = x^2 + y^2 + z^2$ and $\gamma^2 = 4\kappa^2 + \alpha^2 + \beta^2$. A proof of this mathematical identity is given in the Appendix.

By taking $\partial/\partial\kappa$ on both sides of Eq. (22), we obtain:

$$\int\int dx\, dy\, r^{n-1}\exp(-2\kappa r)\exp(i\alpha x + i\beta y)$$

$$= 2\pi(-\partial/2\partial\kappa)^n[\gamma^{-1}\exp(-\gamma z)]. \tag{23}$$

By taking derivatives with respect to z, a factor $x/r = \cos\theta$ is generated. Within the same approximation of the Slater atomic wave functions, the differentiation acts on the exponential factor only. Therefore,

$$\int\int dx\, dy\, r^{n-1}\cos^m\theta\exp(-2\kappa r)\exp(i\alpha x + i\beta y)$$

$$= 2\pi(-\partial/2\partial\kappa)^n(-\partial/2\kappa\partial z)^m[\gamma^{-1}\exp(-\gamma z)]. \tag{24}$$

Using Eqs. (20)–(24) and the conductance distribution functions listed in Table I, the corrugation amplitudes for an elemental hexagonal surface with different tip states and surface states are obtained, as shown in Table II. As shown, with p and d states on the tip as well as on the sample, substantial corrugation enhancements should be observed. In the last column of Table II, the ratios of the corrugation amplitudes with respect to the s-wave case are given for a surface with atomic distance $a = 2.88$ Å and work function $\phi = 3.5$ eV. The relevant quantities are: $\kappa = 0.96$ Å^{-1}, $b = 2.52$ Å^{-1}, and $\gamma = 3.17$ Å^{-1}. For many cases, the enhancements can be greater than one order of magnitude. In Table II, only cases with $n = 1$ and $m = 0$ are listed. It is easy to extend the results for $n > 1$ and $m > 0$ cases, as well as mixed states, such as sp^3 states.

Figure 5 is a comparison of the above results with the first-principle calculation of the Al(111) surface[17] as well as the experimental results of STM images on Al(111) by Wintterlin *et al.*[3] A very simple model of Al(111) surface is used: on each surface Al atom, there is an independent 1s state near the Fermi level. The charge density contour, i.e., the image with an s-wave tip state, agrees with the extrapolated corrugation amplitudes of the first-principle calculation.[17] The theoretical STM image with a d_{z^2} tip state agrees well with the "best" observed STM image on Al(111).[3] The tip–sample distance is taken from the measured tip–sample distance, 1 Å from a mechanical contact,[10] and the calculated distance[18] of a mechanical contact of a Al–Al system, 2.3 Å. Considering that there should be a p_z component on each Al site, the theoretical curve should be moved up slightly, and the fitting can be further improved.

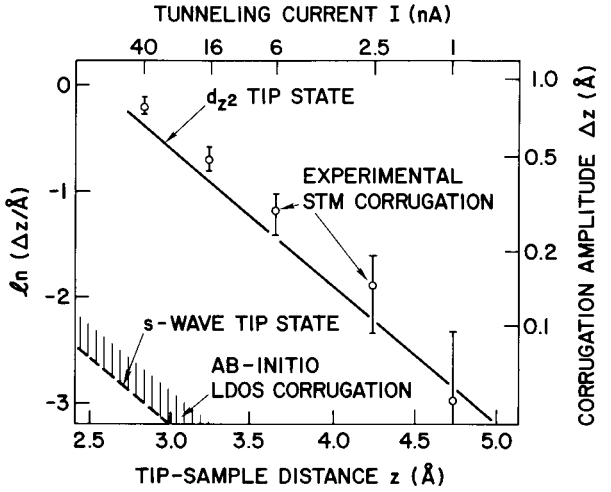

FIG. 5. Theoretical corrugation amplitude for an s and a $d_{3z^2-r^2}$ tip states, on a close-packed metal surface with $a = 2.88$ Å and $\phi = 3.5$ eV. The orbitals on each metal atom on the sample is assumed to be 1s type. Measured STM corrugation amplitudes are from the data of Wintterlin *et al.* (Ref. 3). First-principle calculation of Al(111) is taken from Ref. 17.

TABLE II. Corrugation amplitude $\Delta z(z)$ of a simple crystalline surface with hexagonal symmetry for different tip states and sample states. The last column shows the ratio of Δz vs the s/s case with the following parameters: atomic distance $a = 2.88$ Å, work function $\phi = 3.5$ eV. Quantities involved: decay constant $\kappa = \sqrt{2m_e\phi}/\hbar = 0.93$ Å^{-1}, length of primitive reciprocal vector $b = 2\pi/\sqrt{3}a = 2.52$ Å^{-1}, and $\gamma = \sqrt{4\kappa^2 + b^2} = 3.17$ Å^{-1}.

Tip state	Sample state	Corrugation amplitude $\Delta z(z)$	Ratio
s	s	$[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	1
s	p	$[\gamma/2\kappa]^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	2.73
s	d	$\{(3/2)[(\gamma/2\kappa)^2-(1/3)]\}^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	12.9
p	s	$[\gamma/2\kappa]^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	2.73
p	p	$[\gamma/2\kappa]^4[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	7.45
p	d	$[\gamma/2\kappa]^2\{(3/2)[(\gamma/2\kappa)^2-(1/3)]\}^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	35.2
d	s	$\{(3/2)[(\gamma/2\kappa)^2-(1/3)]\}^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	12.9
d	p	$[\gamma/2\kappa]^2\{(3/2)[(\gamma/2\kappa)^2-(1/3)]\}^2[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	35.2
d	d	$\{(3/2)[(\gamma/2\kappa)^2-(1/3)]\}^4[9\kappa/\gamma^2]\exp\{-[\gamma-2\kappa]z\}$	166

VI. DISCUSSIONS

Shortly after Baratoff[5] proposed the dangling-bond tip-state concept, Tersoff and Hamann[19] proposed the *s*-wave model, which asserts that the STM image is simply a contour of the Fermi-level local density of states (LDOS).[19,20] In this section, we will make a comparison between the present theory and the *s*-wave tip-state model or Fermi-level LDOS-contour model, in terms of agreement with basic experimental facts as well as internal logical consistency.

Theoretically, the Fermi-level LDOS-contour model is a *macroscopic model.*[19] The tip is modeled as a spherical potential well with local radius R. Obviously, the model might be valid only with a radius much greater than the size of an atom, such that the atomic details of the tip can be neglected. With $R = 9$ Å and a distance (from the center of the tip to the top-layer nuclei of the sample) 15 Å, at low bias, Tersoff and Hamann show that the center of curvature of the tip follows the contour of the Fermi-level LDOS of reconstructed gold surfaces.[19] "In any case," they emphasized, "the *s*-wave treatment here is not intended as an accurate description of a real tip, but rather as a useful way of parametrizing the effect of finite tip size." The crucial approximation of the Fermi-level LDOS-contour model, i.e., evaluating the matrix element only for an *s*-wave tip wave function, requires a stringent condition, as stated clearly in Tersoff and Hamann's original papers:[19]

$$(1 + q^2/\kappa^2)^{1/2} \simeq 1, \tag{25}$$

where q is the wave vector of a relevant Fourier component of surface wave function. In terms of a characteristic wavelength or feature size $L = 2\pi/q$, Eq. (25) means

$$L \gg 2\pi\hbar(2m_e\phi)^{-1/2} \simeq 12.3\phi^{-1/2}(\text{Å}), \tag{26}$$

where the work function ϕ is in electron volts. Taking a typical value for the work function, $\phi \simeq 3.5$ eV, Tersoff and Hamann's condition means $L \gg 6.6$ Å. For large features on reconstructed surfaces, e.g., Au(110) 2×1 and 3×1, the typical feature size is 8–12 Å, Tersoff and Hamann's condition is satisfied. For atom-scale features, where the typical length scale is 2–4 Å, Tersoff and Hamann's condition is not satisfied. Therefore, the absence of atomic resolution in the *s*-wave model is a premise, a postulate.

Although the absence of atomic resolution is an *assumption* in the *s*-wave model, it claims a *theoretical conclusion* that true atomic resolution in STM is impossible. "As such," the authors concluded,[19] STM "can resolve isolated steps, defects, and impurities. ... Conversely, STM provides little information for relatively smooth low-Miller-index surfaces, the only kind which are presently susceptible to accurate calculation of the vacuum charge." For the observed atom-scale images on Si(111)2×1, layered materials, and graphite, the *s*-wave model asserted that these images are but an electronic effect: "The image has no direct relation to the positions of atoms within the unit cell. ... Thus the image in this case contains no information whatsoever on the positions of individual atoms."[20] Interestingly enough, the validity criterion as well as the predicted resolution of the *s*-wave model is exactly the resolution limit of some commercially available scanning electron microscopes.[21] As is well known, the value of STM lies in its ability to resolve true atomic details beyond the resolution limit of scanning electron microscope.

Therefore, the starting point and conclusion of the present theory, that atomic resolution in STM is possible, but depends on the existence of orientated localized tip states,[14] are qualitatively different from the starting point and conclusions of the Fermi-level LDOS-contour model.[19,20] In words, it is a microscopic or atomic point of view versus a macroscopic or continuum point of view.

Recently, Lawunmi and Payne[22] made a careful analysis of the Fermi-level LDOS-contour model, especially for the case of graphite.[20] They concluded[22] that the Fermi-level LDOS-contour model "cannot be correct."

VII. CONCLUSIONS

We have established a microscopic theory of STM imaging mechanism in terms of a localized surface state on the tip and a two-dimensional array of independent localized states on the sample. We show that with p and d states on the tip as well as the sample, the corrugation amplitudes of STM images can be orders of magnitude greater than the Fermi-level LDOS, which is the origin of atomic resolution in STM. Analytic expressions for the apparent sizes of atoms as the image of a single atomic state on the sample as well as the corrugation amplitudes of a crystalline surface are given. The theory provides a quantitative interpretation to the atom-resolved images observed on close-packed metal surfaces.

ACKNOWLEDGMENTS

The author wishes to acknowledge J. E. Demuth, H. Rohrer, R. J. Hamers, A. Baratoff, I. P. Batra, H. Ohnishi, N. Garcia, and J. Tersoff for helpful discussions. This paper was presented in part at the Topical Conference on Nanometer Scale Properties of Surfaces and Interfaces, Boston, MA, 23–26 October 1989.

APPENDIX: PROOF OF A MATHEMATICAL IDENTITY

Equation (22) is a direct consequence of two well-known identities:

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{x^2+y^2}} \exp(-\kappa\sqrt{x^2+y^2}) \times \exp(i\alpha x) = 2K_0(y\sqrt{\kappa^2+\alpha^2})$$

and

$$\int_{-\infty}^{\infty} dx\, K_0(\alpha\sqrt{x^2+y^2})\exp(i\beta x) = \frac{\pi \exp(-y\sqrt{\alpha^2+\beta^2})}{\sqrt{\alpha^2+\beta^2}}.$$

These identities, in terms of trigonometric functions, can be found from, e.g., in Ref. 23.

[1] G. Binnig and H. Rohrer, Helv. Phys. Acta **55**, 726 (1982).
[2] For example, J. Vac. Sci. Technol. A **8**, 153 (1990).
[3] V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, and R. J. Wilson, Phys. Rev. Lett. **59**, 2879 (1987); Ch. Wöll, S. Chiang, R. J. Wilson, and

P. H. Lippel, Phys. Rev. B **39**, 7988 (1989); J. Wintterlin, J. Wiechers, H. Burne, T. Gritsch, H. Höfer, and R. J. Behm, Phys. Rev. Lett. **62**, 59 (1989).
[4] J. E. Demuth, U. Koehler, and R. J. Hamers, J. Microsc. **151**, 299 (1988).
[5] A. Baratoff, Europhysics Conference Abstracts, 3rd General Conference of the Condensed Matter Division of the EPS, 28–30 March 1983, P25-103.1; Physica B **127**, 143 (1984).
[6] S.-L. Weng, E. W. Plummer, and T. Gustafsson, Phys. Rev. B **18**, 1718 (1978); M. Posternak, H. Krakauer, A. J. Freeman, and D. D. Koelling, *ibid.* **21**, 5601 (1980); L. F. Mattheiss and D. R. Hamann, *ibid.* **29**, 5372 (1984).
[7] S. Ohnishi and M. Tsukuda, Solid State Commun. **71**, 391 (1989).
[8] J. Northrup, Phys. Rev. Lett. **57**, 154 (1986).
[9] C. Bosvieux and J. Friedel, J. Phys. Chem. Solids **23**, 123 (1962); R. P. Gupta, *ibid.* **47**, 1057 (1986).
[10] U. Dürig, J. K. Gimzewski, and D. W. Pohl, Phys. Rev. Lett. **57**, 2403 (1986); U. Dürig, O. Züger, and D. W. Pohl, J. Microsc. **152**, 259 (1988).
[11] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
[12] C. J. Chen, Mod. Phys. Lett. B (in press).
[13] C. J. Chen, J. Vac. Sci. Technol. A **6**, 319 (1988); Phys. Rev. B **42**, 8841 (1990).
[14] C. J. Chen, Phys. Rev. Lett. **65**, 448 (1990).
[15] C. Zener, Phys. Rev. **36**, 51 (1930); J. C. Slater, *ibid.* **36**, 57 (1930); E. Clementi, D. L. Raimondi, and W. P. Reinhardt, J. Chem. Phys. **47**, 1300 (1967); A. D. McLean and R. S. McLean, At. Data Nucl. Data Tables **26**, 197 (1981).
[16] International *Tables for Crystallography*, edited by T. Hahn (Reidel, Dordrecht, 1987), Vol. A.
[17] K. Mednick and L. Kleinman, Phys. Rev. B **22**, 5768 (1980).
[18] S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **42**, 7618 (1990).
[19] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983); Phys. Rev. B **31**, 805 (1985).
[20] J. Tersoff, Phys. Rev. Lett. **57**, 440 (1986); Phys. Rev. B **39**, 1052 (1989); **41**, 1235 (1990).
[21] For example, JOEL JSM-890 and Hitachi S-900.
[22] D. Lawunmi and M. C. Payne, J. Phys. Condens. Matter **2**, 3811 (1990).
[23] I. S. Gradsteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980), pp. 498, 73.