Tunneling matrix elements in three-dimensional space: The derivative rule and the sum rule

Article *in* Physical review. B, Condensed matter · December 1990

DOI: 10.1103/PhysRevB.42.8841 · Source: PubMed

1 author:

Julian Chen
Columbia University
74 PUBLICATIONS **2,344** CITATIONS

SEE PROFILE

Tunneling matrix elements in three-dimensional space: The derivative rule and the sum rule

C. Julian Chen
IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598
(Received 15 May 1990)

In this paper, a systematic derivation of the tunneling matrix elements in three-dimensional space is presented. Based on a modified Bardeen tunneling theory, explicit expressions for the tunneling matrix elements for localized tip states are derived with use of the Green's-function method. It is shown that by expanding the vacuum tail of the tip wave function in terms of spherical harmonics, the tunneling matrix elements are related to the *derivatives* of the sample wave functions at the nucleus of the apex atom (taken as the center of the spherical-harmonics expansion), in a simple and straightforward way. In addition, an independent derivation based on a general *sum rule* is also presented, which is valid in a number of curvilinear coordinate systems. In spherical coordinates, a general form of the derivative rule follows. In parabolic coordinates, similar results are obtained. Physical meanings of these matrix elements, as well as their implications to the imaging mechanism of scanning-tunneling microscopy, are discussed.

I. INTRODUCTION

Scanning tunneling microscopy (STM) has developed into an important field of experimental physics since its introduction by Binnig and Rohrer[1] in 1982. Its applications have been extended to chemistry, biology, and various engineering sciences.[2] STM routinely resolves individual atoms on virtually every kind of solid surface, including the smoothest ones on the earth: the close-packed metal surfaces[3] with atomic distances 2.5–3 Å. Various authors have developed theories of STM. Garcia *et al.* modeled the tip and the sample as corrugated metallic surfaces.[4] Using a direct method to calculate the tunneling current numerically, a satisfactory agreement with the observed STM image of Au(111) (1×2) was achieved.[4] A systematic method for calculating the current distribution of this model was developed by Stoll *et al.*[5] Tersoff and Hamann[6] showed that the STM images without atomic resolution, such as those on reconstructed Au surfaces, may have a simple explanation. By modeling the tip as a spherical potential well of radius $R=9$ Å, taking the s-wave solution of this macroscopic Schrödinger equation to represent the electronic state of the tip, using Bardeen's perturbation theory of tunneling,[7] at a distance (from the center of the sphere to the sample surface) 15 Å, they showed that the STM image is approximately the Fermi-level local density of states (LDOS) contour of the sample at the center of that sphere.[6] However, for atom-sized features, the corrugation amplitude of the LDOS contour, even at an extremely short tip-sample distance and with an unrealistically small R, is still too small to be observable.[6] Therefore, the Tersoff-Hamann model fails to explain the basic experimental fact of STM: atomic resolution.[3] Baratoff[8] suggested that the atomic resolution in STM may arise from a localized surface state at the tip, in other words, a dangling bond protruding from the tip. Although this concept is very attractive, the mathematical problem of calculating the tunneling current is not trivial.[8] Noguera *et al.* pointed out that because of the strong interaction between the tip and the sample under normal operational conditions, Bardeen's perturbation theory of tunneling phenomena,[7] which is the basis of the Tersoff-Hamann model,[6] is no longer valid.[9] They also show that the s-wave model of Tersoff and Hamann is an oversimplification.[9] Another important fact known to every STM experimentalist is the spontaneous switching of STM instrument resolution during imaging.[10] Very often, the instrument resolution changes all of a sudden without any external operation. Sometimes the corrugation amplitude switches abruptly by one order of magnitude under exactly the same tunneling conditions. Based on this fact, the role of localized surface states at the tip is again proposed.[10]

To have a better understanding of STM, two kinds of knowledge are needed. The first is the knowledge of the specific states on the tip. The second is a theory elucidating the interrelation between the tip states and the images. This paper treats primarily the second issue, i.e., the theoretical connection between tip state and observed image. Such a theory is useful in two ways: to predict the image using the knowledge of the tip state, and to infer the tip state from the observed image.

As realized by many authors,[8,9] because of the short tip-sample separation, Bardeen's tunneling theory is no longer valid for treating transmission phenomena occurring in STM. Actually, under normal experimental conditions of STM, the tunneling barrier is often lower than the Fermi level. In such cases, even the concept of "tunneling" becomes questionable. Based on the uncertainty principle, we have shown[11] that because the barrier occurring in STM is very thin, the distinction between tunneling (classically forbidden transmission) and channeling (classically allowed transmission) disappears. We have also shown that by considering the distortion of the sample wave function due to the existence of the tip and *vise versa*, a modified Bardeen theory, in terms of *distorted* wave functions, is accurate throughout the entire re-

gime of STM operation even if the barrier top becomes lower than the Fermi level. For metals and individual atoms, the distortion due to the existence of another party is a constant, and Bardeen's integral, after such a modification, is still valid.[11]

In this paper, based on the modified Bardeen tunneling theory,[11] we present a systematic derivation of analytic expressions for the tunneling matrix elements arising from localized states commonly found on tips. A brief description of the general results, the *derivative rule* and the *reciprocity principle,* was included in an earlier publication.[11] Based on the derivative rule, we have made a quantitative explanation of the observed atom-resolved images on close-packed metal surfaces.[12] In general, using these tunneling matrix elements, starting with an analytic form for the surface wave function, analytic expressions for the tunneling current distribution with any tip state can be obtained. Thus, analytic expressions for the images can be obtained, which can be conveniently used for comparison with experimental STM images.[12]

The organization of this paper is as follows. Section II is a brief summary of the existing knowledge about tip states. We show that the evidence for the existence and importance of localized tip states is ample. In Secs. III–V, we give a simple derivation of tunneling matrix elements for nine commonly occurring tip states, using the Green's-function method. This derivation is straightforward, and provides an intuitive interpretation of its physical meaning. Explicit results are listed up to $l=2$. A simple *derivative rule* is shown. However, the Green's-function method derives the expression for each tunneling matrix element individually, with a lack of generality. Therefore, a second derivation, which is more elegant and general, is presented. It is based on a sum rule for tunneling matrix elements in three-dimensional space, valid for many curvilinear coordinate systems, as presented in Sec. VI. As a special case of that sum rule, in spherical coordinates, a general expression for the tunneling matrix elements in terms of spherical-harmonic expansion follows directly. Here, the derivative rule reappears (Sec. VII). The implication of the derivative rule for tunneling matrix elements to the imaging mechanism of scanning tunneling microscopy is presented in Sec. VIII. In Sec. IX, the consequence of sum rule in the parabolic coordinate system is briefly presented, which generates similar results as in the spherical coordinate systems. Finally, in Sec. XI, we show that at the macroscopic limit, the microscopic theory presented here reduces to the s-wave model of Tersoff and Hamann.[6] We conclude with some ideas to expand our knowledge about the electronic states on the tip and their role in STM imaging.

II. LOCALIZED SURFACE STATES ON THE TIP

Tungsten and platinum-iridium alloys are almost exclusively used in STM as tip materials. These elements are d-band metals.[13] At the Fermi level, most of the density of states (DOS) is from d states. Therefore, it is natural to consider the role of d tip states in STM.[8,12,14] Table I shows the Fermi-level DOS of three common tip materials:[13] W, Pt, and Ir.

TABLE I. Fermi-level DOS of common tip materials (Ref. 13).

Material	W	Pt	Ir
s state	3.1%	0.77%	0.94%
d state	85%	98%	96%

Yet there is some doubt about the role of d states in STM, which arises from the concern that d states decay much faster than s states.[15] Indeed, this is true for *free atoms.* As an example, in a free tungsten atom,[16] the orbital exponent of a $6s$ wave function is $\zeta=3.1$ Å^{-1}, whereas for a $5d$ electron, $\zeta=6.33$ Å^{-1}. Therefore, even at a distance 1 Å from the center of a tungsten nucleus, the charge density of a $5d$ electron is too small to be effective. Because of the large decay constant, the charge density of the atomic $5d$ state at a normal tip-sample distance is even smaller.[15]

However, for d-type surface states on top of bulk tungsten,[17–20] the situation can be completely different. For example, it is a well-established fact both experimentally and theoretically[17–20] that on a W(100) surface there is a highly localized metallic d_{z^2} surface state at 0.3 eV below the Fermi level (0.4 eV in width). This surface state was first observed in the 1960s by Swanson and Crouser in field-emission experiments,[17] known as "Swanson hump." These localized surface states were later studied extensively using angle-resolved photoemission by Weng *et al.*[18] With angle-resolved photoemission, the energy level and the angular dependence of the surface states can be simultaneously determined. For the surface state 0.3 eV below the Fermi level, angle-resolved photoemission experiments identified that its orbital character is mainly d_{z^2}. The second one is made of $d_{x^2-y^2}$ and $d_{zx,zy}$ orbitals. The third is made of d_{z^2}, s, and p_z orbitals. Table II is a brief summary of the measured energy levels[18] of these three surfaces states on W(100) and Mo(100).

An extensive first-principles calculation of the W(100) surface using the linearized-augmented-plane-wave (LAPW) method including relativistic corrections was executed by Posternak *et al.*[19] The orbital character of the surface state near the Fermi level was identified to be essentially a d_{z^2} state,[19] in good agreement with the results of photoemission experiments.[18] Their calculation was repeated by Mattheis and Hamann[20] using an independent LAPW program, which yields essentially the same conclusion. The absolute value of the electron density of this state can be found from Fig. 8 of Ref. 19. At a

TABLE II. Energy level of surface states (Ref. 18) (eV below E_F).

Material	first	second	third
W	0.3	0.8	4.2
Mo	0.2	0.6	3.3

distance 2.3 Å from the tungsten nucleus on the vacuum side, its charge density is 0.4 electrons per unit cell (31.6 Å^3), or 1.26×10^{-2} Å^{-3}, which is substantially higher than that of a 6*s* electron in a free tungsten atom at the same distance,[16] 1.94×10^{-4} Å^{-3}, and several orders of magnitude greater than that of the atomic 5*d* electrons. Because of its shallow energy level, it decays much slower than the atomic 5*d* states. Actually, from Fig. 8 of Ref. 19, the decay constant is estimated to be about 1.0 Å^{-1}. This value agrees well with the measured work function, $\simeq$4 eV. Because of the relatively small decay constant, that charge density of that d_{z^2} surface state extends much further into the vacuum than the atomic 6*s* and 5*d* states.[16] Figure 1 shows a comparison of the charge-density contour of a surface state[19] with a 6*s* state and a 5*d* state of a tungsten atom.[16] As shown, at a distance about 2 Å from the nucleus, the charge density of the surface state is already much higher than the atomic 6*s* and 5*d* states. As we have mentioned, the decay length of the surface state, $\kappa^{-1}\simeq1$ Å, is much longer than those of the 6*s* and 5*d* atomic states. Hence, at large distances, the charge density of this surface state is many orders of magnitude larger than those of the atomic 6*s* and 5*d* states. (Notice that this is only true on the vacuum side, see Fig. 1. In the bulk, the localized surface state decays even faster than the atomic 6*s* and 5*d* states. Actually, in the angle-resolved photoemission experiment, the measured quantity is determined by the vacuum tail of the surface states. Therefore, the values of the surface state inside the bulk is irrelevant. Similarly, in STM, the effective part of the surface state is its vacuum tail, not the part inside the tip body.)

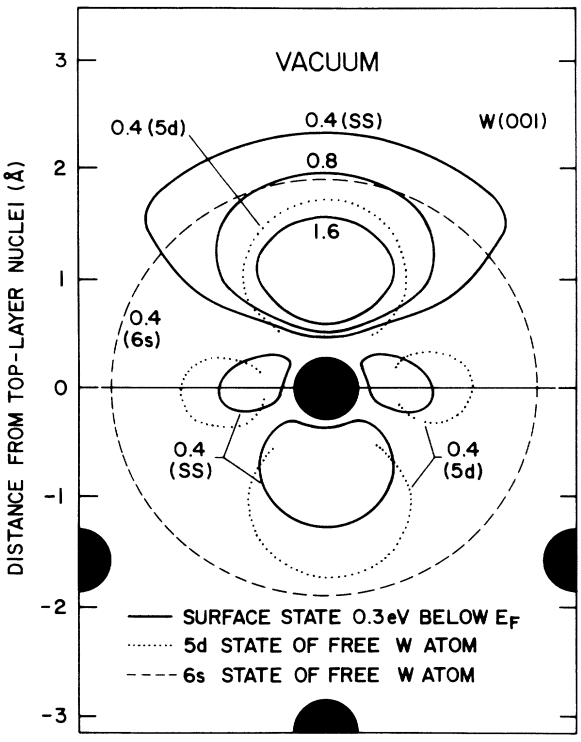

FIG. 1. A comparison of the charge-density contour of a surface state (Ref. 19) with a 6*s* state and a 5*d* state of a tungsten atom (Ref. 16). As shown, at a distance about 2 Å from the nucleus, the charge density of the surface state is already much higher than the atomic 6*s* and 5*d* states. Since the decay constant of the surface state, $\kappa\simeq1$ Å^{-1}, is much smaller than those of the 6*s* and 5*d* atomic states, at larger distances, the charge density of this surface state is many orders of magnitudes larger than those of the atomic 6*s* and 5*d* states.

An independent evidence of the existence of d_{z^2} tip state on tungsten tips was shown by Ohnishi *et al.*[14] To understand the role of surface states on a tungsten tip in STM, they performed a first-principles calculation of electronic states of several kinds of W clusters.[14] Their results show clearly that at the apex atom of both W_4 and W_5 clusters, there is a dangling-bond state near the Fermi level which can be accurately described as a d_{z^2} state.[14] Using these W clusters as the tip, they calculated the tunneling current to a silicon sample using the Green's-function method. They find that the tunneling current is primarily generated by this d_{z^2} dangling-bond tip state.[14]

Another important example of localized surface states on the tip is the dangling-bond state of silicon, as proposed by Demuth *et al.* for interpreting the observed atomic resolution and the spontaneous resolution switching on silicon surfaces.[10] Direct experimental evidence showed that by mildly colliding the tip with the silicon surface, the tip picks up a silicon cluster, and the atomic resolution is frequently resumed.[10] Evidently, a silicon cluster is formed at the apex of the tip.[10] On silicon surfaces, there is no pure *s*-wave state. Oriented sp^3 dangling-bond states exist on a variety of silicon surfaces.[21] Demuth *et al.* proposed[10] that the sp^3 tip state may generate much greater atomic corrugation than the Fermi-level LDOS.

In order to understand the effect of different localized tip states, the corresponding tunneling matrix elements arising from these tip states have to be evaluated explicitly. In the following, we present derivations and results for these tunneling matrix elements. Explicit forms for nine important cases are given. As shown, the final results are extremely simple.[12]

III. TIP WAVE FUNCTIONS

In the 1982 paper of Binnig and Rohrer,[1] the tunneling current is attributed to the overlap of the wave functions of the tip and the sample near the Fermi level at a separation surface located roughly in the middle of the gap. We have shown that the Bardeen tunneling theory,[7] after proper modification, is appropriate for evaluating the tunneling (transmission) matrix elements even when the barrier becomes classically allowed.[11] For metals and individual atoms, the correction to the Bardeen integral is simply a constant independent of relative tip-sample position.[11] A schematic for the perturbation calculation of the tunneling matrix elements is shown in Fig. 2. For the free tip state, the potential on and beyond the separation surface is equal to that of the vacuum. Therefore, on and beyond the separation surface, i.e., on the sample side

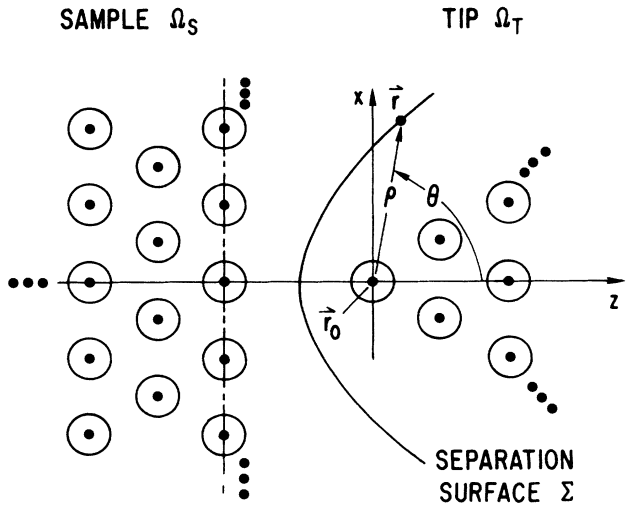

FIG. 2. Schematics for calculating tunneling matrix elements. The z axis is defined by the surface of the sample, with the positive z direction pointing out from the sample. A separation surface is drawn between the tip and the sample, roughly in the middle of the gap. The precise position is not critical.

(see Fig. 2), the tip wave function satisfies Schrödinger's equation in the vacuum,

$$(\nabla^2-\kappa^2)\chi(\mathbf{r})=0 , \tag{3.1}$$

where $\kappa=(2m_e\phi)^{1/2}\hbar^{-1}$ is the decay constant, determined by the work function ϕ. Equation (3.1) is also known as the modified Helmholtz equation. Inside the tip, including the vicinity of the apex atom, Eq. (3.1) is not valid for the tip wave function. This is not a concern because we are not interested in the tip wave function inside the tip body. Similarly, for the free sample state, on and beyond the separation surface, i.e., on the tip side (see Fig. 2), the sample wave function satisfies Eq. (3.1). Inside the sample body, Eq. (3.1) is not valid for the sample wave function. Again, in order to calculate the tunneling matrix element, we only need the sample wave function on and beyond the separation surface, which always satisfies Eq. (3.1). Within the framework of the modified Bardeen tunneling theory,[11] the above condition is valid *by definition.*

In the following, we will derive a general expression for the tip wave function on and beyond the separation surface. Because the solution of Dirichlet's problem of Eq. (3.1) is unique, the values of tip wave function in the entire region beyond the separation surface (i.e., the sample side) are determined by the values of the tip wave function on the separation surface. We expand the tip wave function on and beyond the separation surface into spherical-harmonic components:

$$\chi(\mathbf{r})=\sum_{l,m} C_{lm}f_{lm}(\kappa\rho)Y_{lm}(\theta,\phi) , \tag{3.2}$$

where $\rho=|\mathbf{r}-\mathbf{r}_0|$, $\mathbf{r}_0$ is the center of the apex atom, and Y_{lm} is spherical harmonics.[22,23] Substituting Eq. (3.2) into Eq. (3.1), we obtain the differential equation for the functions $f_{lm}(u)$:

$$\frac{d}{du}\left[u^2\frac{df(u)}{du}\right]-[u^2+l(l+1)]f(u)=0 . \tag{3.3}$$

The functions only depend on l. There are two standard linearly independent solutions for Eq. (3.3): the spherical modified Bessel function of the first kind,

$$i_l(u)=\left[\frac{\pi}{2u}\right]^{1/2}I_{l+1/2}(u) , \tag{3.4}$$

and of the second kind,[22]

$$k_l(u)=\left[\frac{2}{\pi u}\right]^{1/2}K_{l+1/2}(u) . \tag{3.5}$$

These special functions are actually *elementary functions.*[22] Explicitly, these are

$$i_l(u)=u^l\left[\frac{d}{u\,du}\right]^l\frac{\sinh u}{u} \tag{3.6}$$

and

$$k_l(u)=(-1)^l u^l\left[\frac{d}{u\,du}\right]^l\frac{\exp(-u)}{u} . \tag{3.7}$$

The most important properties of these functions are listed in Appendix A. Obviously, the functions $i_l(u)$ diverge at large u, which does not meet the boundary condition for tip wave functions. The functions $k_l(u)$ are regular at large u, which satisfies the required boundary condition. Therefore, a component of tip wave function in the vacuum region with quantum numbers l and m must have the general form

$$\chi_{lm}(\mathbf{r})=C_{lm}k_l(\kappa\rho)Y_{lm}(\theta,\phi) . \tag{3.8}$$

In the absence of magnetic field, it is convenient to write those tip wave functions in real form, as listed in Table III. The constant C_{lm} is to be determined by comparison with the results of first-principles calculations of actual tip states. For convenience, we eliminate the rather complicated normalization factors of the spherical harmonics in these expressions by redefining the constant C_{lm}.

It is clear from Eq. (3.8) that only the *angular dependence* of the tip wave function on and beyond the separation surface matters. For example, according to the ex-

TABLE III. Tip wave functions.

State	Wave function
s	$C(\kappa\rho)^{-1}\exp(-\kappa\rho)$
p_z	$C[(\kappa\rho)^{-1}+(\kappa\rho)^{-2}]\exp(-\kappa\rho)\cos\theta$
p_x	$C[(\kappa\rho)^{-1}+(\kappa\rho)^{-2}]\exp(-\kappa\rho)\sin\theta\cos\phi$
p_y	$C[(\kappa\rho)^{-1}+(\kappa\rho)^{-2}]\exp(-\kappa\rho)\sin\theta\sin\phi$
d_{z^2}	$C[(\kappa\rho)^{-1}+3(\kappa\rho)^{-2}+3(\kappa\rho)^{-3}]\exp(-\kappa\rho)(\cos^2\theta-\frac{1}{3})$
d_{xz}	$C[(\kappa\rho)^{-1}+3(\kappa\rho)^{-2}+3(\kappa\rho)^{-3}]\exp(-\kappa\rho)\sin(2\theta)\cos\phi$
d_{yz}	$C[(\kappa\rho)^{-1}+3(\kappa\rho)^{-2}+3(\kappa\rho)^{-3}]\exp(-\kappa\rho)\sin(2\theta)\sin\phi$
d_{xy}	$C[(\kappa\rho)^{-1}+3(\kappa\rho)^{-2}+3(\kappa\rho)^{-3}]\exp(-\kappa\rho)\sin^2\theta\sin(2\phi)$
$d_{x^2-y^2}$	$C[(\kappa\rho)^{-1}+3(\kappa\rho)^{-2}+3(\kappa\rho)^{-3}]\exp(-\kappa\rho)\sin^2\theta\cos(2\phi)$

perimental results of angle-resolved photoemission,[18] on W(001) and Mo(001), the surface state near the Fermi level has an angular dependence of $Y_{20}(\theta,\phi)$. Therefore, for this localized surface state, the constants in Eq. (3.8) should be all zero except C_{20}. If a silicon cluster is at the tip end, the probability of having an sp^3 dangling-bond tip state is high.[10] To describe this state, we take $C_{10}=\sqrt{3}C_{00}$ and set all other coefficients to zero. The absolute values of the constants can be determined by comparison with the results of first-principles calculations. The validity of the above idealized cases, obviously, depends on the condition that on the separation surface the electron density from other states in the tip can be neglected; and the arrangement of the atoms near the apex atom of the tip satisfies the condition for the dangling-bond state to exist. The probability of having this condition is limited. As shown in Sec. VIII, the observed images depend on the actual states of the tip in a dramatic way. This existence of a large variety of different tip states as a result of different structures near the apex of the tip provides an explanation to the observed frequency of spontaneous switching of instrument resolution.[1,7,10]

IV. GREEN'S FUNCTION AND TIP WAVE FUNCTIONS

The Green's function for the Schrödinger equation, Eq. (3.1), is defined by the differential equation[23]

$$(\nabla^2-\kappa^2)G(\mathbf{r}-\mathbf{r}_0)=-\delta(\mathbf{r}-\mathbf{r}_0) . \tag{4.1}$$

With the boundary condition that it is regular at $|\mathbf{r}-\mathbf{r}_0|\to\infty$, the explicit form of the Green's function is[22]

$$G(\mathbf{r}-\mathbf{r}_0)=\frac{\exp(-\kappa|\mathbf{r}-\mathbf{r}_0|)}{4\pi|\mathbf{r}-\mathbf{r}_0|} , \tag{4.2}$$

which can be verified by direct substitution. Denoting $\rho=|\mathbf{r}-\mathbf{r}_0|$, the Green's function can be written in terms of the spherical modified Bessel function of the second kind,[22]

$$G(\mathbf{r}-\mathbf{r}_0)=\frac{\kappa}{4\pi}k_0(\kappa\rho) . \tag{4.3}$$

Therefore, the s-wave tip wave function is identical to the Green's function up to a constant, *with the center of the apex atom taken as* $\mathbf{r}_0$,

$$\chi_s(\mathbf{r})=\frac{4\pi C}{\kappa}G(\mathbf{r}-\mathbf{r}_0) . \tag{4.4}$$

By taking derivative to both sides of Eq. (4.3) with respect to z_0 and using the relation $k_0'(u)-k_1(u)$ (see Appendix A), we obtain

$$\frac{\partial}{\kappa\,\partial z_0}G(\mathbf{r}-\mathbf{r}_0)=\frac{\kappa}{4\pi}\frac{z-z_0}{\rho}k_1(\kappa\rho) . \tag{4.5}$$

The second fraction in Eq. (4.5) is $(z-z_0)/\rho=\cos\theta$ (see Fig. 2). Therefore, the tip wave function for the p_z states is

$$\chi_{p_z}(\mathbf{r})=\frac{4\pi C}{\kappa}\frac{\partial}{\kappa\,\partial z_0}G(\mathbf{r}-\mathbf{r}_0) . \tag{4.6}$$

Similarly, we have

$$\chi_{p_x}(\mathbf{r})=\frac{4\pi C}{\kappa}\frac{\partial}{\kappa\,\partial x_0}G(\mathbf{r}-\mathbf{r}_0) , \tag{4.7}$$

$$\chi_{p_y}(\mathbf{r})=\frac{4\pi C}{\kappa}\frac{\partial}{\kappa\,\partial y_0}G(\mathbf{r}-\mathbf{r}_0) . \tag{4.8}$$

By taking the derivative with respect to x_0 on both sides of Eq. (4.5), noting that $[k_1(u)/u]'=-k_2(u)/u$ (see Appendix A), the same argument leads to

$$\chi_{d_{xz}}(\mathbf{r})=\frac{4\pi C}{\kappa}\frac{\partial^2}{\kappa^2\,\partial x_0\,\partial z_0}G(\mathbf{r}-\mathbf{r}_0) . \tag{4.9}$$

Similar results can be obtained for the wave functions of d_{yz} and d_{xy} states.

By taking derivative with respect to z_0 on both sides of Eq. (4.5), an extra term containing $k_l(\kappa\rho)/\rho$ is generated. Considering that $3k_1(u)/u=k_0(u)-k_2(u)$ (see Appendix A), we obtain

$$\chi_{dz^2}(\mathbf{r})=\frac{4\pi C}{\kappa}\left[\frac{\partial^2}{\kappa^2\,\partial z_0^2}G(\mathbf{r}-\mathbf{r}_0)-\tfrac{1}{3}G(\mathbf{r}-\mathbf{r}_0)\right] . \tag{4.10}$$

For the wave function of a $d_{x^2-y^2}$ tip state, the extra term generated by taking the second derivative cancels. Therefore,

$$\chi_{d_{x^2-y^2}}(\mathbf{r})=\frac{4\pi C}{\kappa}\left[\frac{\partial^2}{\kappa^2\,\partial x_0^2}-\frac{\partial^2}{\kappa^2\,\partial y_0^2}\right]G(\mathbf{r}-\mathbf{r}_0) . \tag{4.11}$$

We will use these relations to evaluate tunneling matrix elements in the next section.

V. THE DERIVATIVE RULE: INDIVIDUAL CASES

In this section, we derive the tunneling matrix elements M for different tip states using the modified Bardeen formula.[11] As we have shown, in all cases related to STM, the tunneling (transmission) matrix element can be evaluated as a surface integral on the separation surface Σ (see Fig. 2) between the tip and the sample,[11]

$$M=\frac{\hbar^2}{2m_e}\int_\Sigma(\chi^*\nabla\psi-\psi\nabla\chi^*)\cdot d\mathbf{S} , \tag{5.1}$$

where ψ is the sample wave function, χ is the tip wave function, and $d\mathbf{S}$ is the surface element on Σ, see Fig. 2.

Consider first the s-wave tip state. Using Eq. (4.4),

$$M_s\equiv\frac{\hbar^2}{2m_e}\int_\Sigma(\chi_s\nabla\psi-\psi\nabla\chi_s)\cdot d\mathbf{S}$$

$$=\frac{2\pi C\hbar^2}{\kappa m_e}\int_\Sigma[G(\mathbf{r}-\mathbf{r}_0)\nabla\psi-\psi\nabla G(\mathbf{r}-\mathbf{r}_0)]\cdot d\mathbf{S} . \tag{5.2}$$

Using Green's theorem, it can be converted into a volume integral over Ω_T, i.e., the volume of the tip or the space to the right of the separation surface (Fig. 2):

$$M_s=\frac{2\pi C\hbar^2}{\kappa m_e}\int_{\Omega_T}[G(\mathbf{r}-\mathbf{r}_0)\nabla^2\psi-\psi\nabla^2G(\mathbf{r}-\mathbf{r}_0)]d\tau . \tag{5.3}$$

Since the sample wave function ψ satisfies Schrödinger's equation (3.1) in Ω_T, the first term of the integrand becomes $G(\mathbf{r}-\mathbf{r}_0)\kappa^2\psi$. On the other hand, the Green's function satisfies Eq. (4.1). Therefore, the second term of the integrand becomes $-\psi[\kappa^2 G(\mathbf{r}-\mathbf{r}_0)-\delta(\mathbf{r}-\mathbf{r}_0)]$. Hence,

$$M_s=\frac{2\pi C\hbar^2}{\kappa m_e}\psi(\mathbf{r}_0)\ . \tag{5.4}$$

Mathematically, this is the result of Tersoff and Hamann.[6] (Notice that the original meaning of the s wave in the theory of Tersoff and Hamann is the representation of the wave function of a macroscopic spherical potential well, not as an s-wave component of an atom-like orbital.) As shown, the simplicity of this result is a consequence of choosing the Green's function to represent the tip wave function. As Tersoff and Hamann derived this very simple result using a rather cumbersome plane-wave expansion method,[6] the present proof is much more straightforward. Besides, this proof is valid on any surface separating the tip and the sample, not necessarily a plane. More importantly, using this method, the tunneling matrix elements for other components of the tip wave function can be easily obtained, in virtue of the relation between tip wave functions and the Green's function, as shown in the previous section. For example, by taking the derivative with respect to z_0 on both sides of Eqs. (5.1) and (5.2), noticing that z_0 is a parameter in the integral (which does not involve in the integration process), and using the expression of the p_z tip wave function, Eq. (4.6), we find

$$\begin{aligned}\frac{2\pi C\hbar^2}{\kappa m_0}\frac{\partial\psi(\mathbf{r}_0)}{\partial z_0}&=\frac{2\pi C\hbar^2}{\kappa m_0}\frac{\partial}{\partial z_0}\int_\Sigma[G(\mathbf{r}-\mathbf{r}_0)\nabla\psi-\psi\nabla G(\mathbf{r}-\mathbf{r}_0)]\cdot d\mathbf{S}\\&=\frac{2\pi C\hbar^2}{\kappa m_0}\int_\Sigma\left[\frac{\partial}{\partial z_0}[G(\mathbf{r}-\mathbf{r}_0)]\nabla\psi-\psi\nabla\frac{\partial}{\partial z_0}[G(\mathbf{r}-\mathbf{r}_0)]\right]\cdot d\mathbf{S}\\&=\frac{\hbar^2}{2m_e}\int_\Sigma[\chi_{p_z}(\mathbf{r})\nabla\psi-\psi\nabla\chi_{p_z}(\mathbf{r})]\cdot d\mathbf{S}\\&=M_{p_z}\ .\end{aligned} \tag{5.5}$$

Therefore, we find that *the tunneling matrix element for a p_z tip state is proportional to the z derivative of the sample wave function at the center of the apex atom.*

Using the expressions for other tip wave functions in terms of Green's functions in the previous section, we immediately obtain the tunneling matrix elements for all the tip states listed in Table III. Table IV is a summary of the results.

The tunneling matrix elements listed in Table IV can be summarized in terms of an extremely simple *derivative rule.* By writing the angle dependence of the tip wave function in terms of x, y, and z, then replacing them with the following rule:

$$x\rightarrow\partial/\partial x\ ,$$

$$y\rightarrow\partial/\partial y\ ,$$

$$z\rightarrow\partial/\partial z\ ,$$

and acting on the sample wave function, we obtain all the tunneling matrix elements immediately, except that for $d_{3z^2-r^2}$. Actually, since $\partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2=\kappa^2$, a direct consequence of the Schrödinger equation (3.1), we find that all the tunneling matrix elements in Table IV follow the derivative rule.

TABLE IV. Tunneling matrix elements.

State	$M\propto$ value at $\mathbf{r}_0$
s	ψ
p [z]	$\frac{\partial\psi}{\partial z}$
p [x]	$\frac{\partial\psi}{\partial x}$
p [y]	$\frac{\partial\psi}{\partial y}$
d [zx]	$\frac{\partial^2\psi}{\partial z\,\partial x}$
d [zy]	$\frac{\partial^2\psi}{dz\,\partial y}$
d [xy]	$\frac{\partial^2\psi}{\partial x\,\partial y}$
d [$z^2-\frac{1}{3}r^2$]	$\frac{\partial^2\psi}{\partial z^2}-\frac{1}{3}\kappa^2\psi$
d [x^2-y^2]	$\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial y^2}$

VI. THE SUM RULE

The simple derivation of the nine tunneling matrix elements presented in the previous section generates the tunneling matrix elements individually. In the following sections, we present an alternative proof, which is more elegant and provides a general formula of the tunneling matrix element for an arbitrary tip state. This derivation is based on a sum rule, which is valid in most curvilinear coordinate systems. This sum rule is a consequence of the analytic property of the Schrödinger equation in free space, or the modified Helmholtz equation.[22,23] We state the sum rule first, then prove it.

We consider an orthogonal curvilinear coordinate sys-

tem ξ_1, ξ_2, and ξ_3, see Fig. 3, assuming that Schrödinger's equation, Eq. (3.1), is separable in this coordinate system,[22,23] which means that every solution of Eq. (3.1) can be expanded in terms of basic solutions of Eq. (3.1), as a product of functions of individual coordinates. Since both the vacuum tail of the tip wave function and the vacuum tail of the sample wave function satisfy Eq. (3.1), they can also be expanded in terms of these basic functions. The sum rule states that the tunneling matrix element is the sum of the product of corresponding coefficients in these two expansions, multiplied by a universal constant.

Following is the proof. To make it easy to read, we follow the notation in the textbook of Morse and Feshbach.[23] Let the surface $\xi_1=\xi_{10}$ be the separation surface. Region $\xi_1<\xi_{10}$ denotes the region of the tip, and $\xi_1>\xi_{10}$ denotes the region of the sample, see Fig. 3. Bardeen's integral, in terms of curvilinear coordinates,[7,11]

$$M=-\frac{\hbar^2}{2m_e}\int_{\xi_1=\xi_{10}}\frac{h_2h_3}{h_1}\left[\psi\frac{\partial\chi^*}{\partial\xi_1}-\chi^*\frac{\partial\psi}{\partial\xi_1}\right]d\xi_2\,d\xi_3\ . \tag{6.1}$$

In terms of curvilinear coordinates, Eq. (3.1) has the form

$$\frac{1}{h_1h_2h_3}\sum_{j=1}^{3}\frac{\partial}{\partial\xi_j}\left[\frac{h_1h_2h_3}{h_j^2}\frac{\partial\Psi}{\partial\xi_j}\right]-\kappa^2\Psi=0\ . \tag{6.2}$$

Equation (6.2) is called *separable* if any solution of it can be expanded as

$$\Psi=\sum_{n,m}C_{nm}X_{nm}^{(1)}(\xi_1)X_{nm}^{(2)}(\xi_2)X_m^{(3)}(\xi_3)\ , \tag{6.3}$$

where each of the factors satisfies a Sturm-Liouville equation,[22,23]

$$\frac{d}{d\xi_n}\left[p_n(\xi_n)\frac{dX^{(n)}}{d\xi_n}\right]+[q_n(\xi_n)+\lambda_n r_n(\xi_n)]X^{(n)}(\xi_n)=0\ . \tag{6.4}$$

For most of the separable coordinate systems, one of the coordinates can be separated first.[23] We denote it as

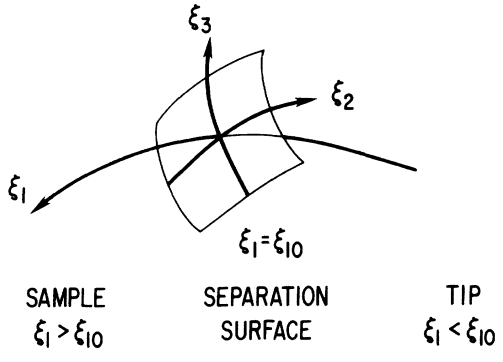

FIG. 3. Derivation of the sum rule. The separation surface is the coordinate surface $\xi_1=\xi_{10}$, a constant.

ξ_3. The boundary condition for its Sturm-Liouville equation is that the function $X^{(3)}(\xi_3)$ must be regular in the entire range of ξ_3. It leads to an eigenvalue problem, which gives a series of eigenvalues for λ_3 and a series of eigenfunctions which are orthonormal with the weighting function $r_3(\xi_3)$:

$$\int X_m^{(3)}r_3(\xi_3)X_{m'}^{(3)}\,d\xi_3=\delta_{mm'}\ . \tag{6.5}$$

The Sturm-Liouville equation for ξ_2 generates another series of eigenfunctions,

$$\int X_{nm}^{(2)}r_2(\xi_2)X_{n'm}^{(2)}\,d\xi_2=\delta_{nn'}\ , \tag{6.6}$$

and eigenvalues for λ_3. Both $X^{(2)}(\xi_2)$ and $X^{(3)}(\xi_3)$ are bounded.[22,23]

The Sturm-Liouville equation for $X^{(1)}(\xi_1)$ deserves special attention. Since the parameter λ_1 in this Sturm-Liouville equation is already fixed by the energy level of Schrödinger's equation, i.e., κ^2, it is no longer an eigenvalue problem. By the analytic nature of the modified Helmholtz equation, i.e., Schrödinger's equation for the vacuum, the solution cannot be regular on the entire range of ξ_1. Actually, as we show in Appendix B, if the absolute value of a solution of Eq. (3.1) is finite in the entire space, then it must be zero everywhere. This is a theorem very similar to Liouville's theorem in the theory of analytic functions of complex variables.[23] The actual wave function of the tip satisfies Eq. (3.1) (and is regular) *only in the sample region,* i.e., only for $\xi_1>\xi_{10}$ (see Fig. 3). In the region $\xi_1<\xi_{10}$, the actual tip wave function does not satisfy Eq. (3.1). The Sturm-Liouville equation of the tip wave function $X^{(1)}(\xi_1)$ in the region $\xi_1<\xi_{10}$ defines the *vacuum continuation* of the vacuum tail of the tip wave function into the tip body, a concept similar to the analytic continuation in the theory of the functions of complex variables. In analogy to the situation in the theory of functions of complex variables, the tip wave function $X^{(1)}(\xi_1)$ must have at least one singularity, or divergence, in the tip body. We denote this solution as $X_{nm}^{(1T)}(\xi_1)$. Similarly, in the sample body, i.e., as $\xi_1>\xi_{10}$ (see Fig. 3), the actual sample wave function does not satisfy Eq. (3.1). The Sturm-Liouville equation for $X^{(1)}(\xi_1)$ in the region $\xi_1>\xi_{10}$ defines the vacuum continuation of the vacuum tail of the sample wave function into the sample body. We denote the second solution as $X_{nm}^{(1S)}(\xi_1)$. Since these two functions are linearly independent solutions of the same Sturm-Liouville equation, their Wronskian[22,23] must be proportional to $1/p_1(\xi_1)$. We choose the normalization of this pair of functions such that the constant in the Wronskian is unity. In other words, we define the normalization of these functions such that

$$X_{nm}^{(1S)}\frac{dX_{nm}^{(1T)}}{d\xi_1}-X_{nm}^{(1T)}\frac{dX_{nm}^{(1S)}}{d\xi_1}=\frac{1}{p_1(\xi_1)}\ . \tag{6.7}$$

It is interesting to note that these two solutions, if they exist, must be unique up to a constant. In fact, since any solution of that Strum-Liuoville equation must be a linear combination of these two, a function with two nonzero coefficients must diverge in both regions.

Therefore, we obtain the most general form of the wave functions of the tip and the sample, as follows:

$$\psi=\sum_{n,m}\alpha_{nm}X_{nm}^{(1S)}(\xi_1)X_{nm}^{(2)}(\xi_2)X_m^{(3)}(\xi_3)\ , \tag{6.8}$$

$$\chi=\sum_{n,m}\beta_{nm}X_{nm}^{(1T)}(\xi_1)X_{nm}^{(2)}(\xi_2)X_m^{(3)}(\xi_3)\ . \tag{6.9}$$

In Appendix C, we prove that for most of the curvilinear coordinate systems, the following *elimination equation* is valid:

$$h_2h_3/h_1=p_1(\xi_1)r_2(\xi_2)r_3(\xi_3)\ . \tag{6.10}$$

Substituting Eqs. (6.8) and (6.9) into Eq. (6.1), using the orthonormal relations, Eqs. (6.5), (6.6), and the Wronskian, (6.7), we obtain an extremely simple result, the sum rule:

$$M=-\frac{\hbar^2}{2m_e}\sum_{n,m}\alpha_{nm}\beta_{nm}^*\ . \tag{6.11}$$

That is, the tunneling matrix element is the sum of the corresponding coefficients in the expansions, Eq. (6.8) and Eq. (6.9), multiplied by a universal constant. All the coordinate dependences are eliminated by virtue of Eq. (6.10).

As an immediate application of the sum rule, we will derive a general expression for the tunneling matrix elements in spherical coordinates, and recover the derivative rule. Another example, in parabolic coordinates, is outlined in Sec. IX.

VII. THE DERIVATIVE RULE: GENERAL CASE

In this section, we apply the above sum rule to the case of spherical coordinates ρ, θ, and ϕ, and derive a formula for tunneling matrix elements for a general tip wave function. For the angular coordinates, θ and ϕ, the Strum-Liouville equations, (6.4), lead to spherical harmonics. The differential equation with respect to ρ is Eq. (3.3), which leads to solutions for the tip and the sample in terms of spherical modified Bessel functions.[22] The most general forms of those wave functions are

$$\psi=\sum_{l,m}\alpha_{lm}i_l(\kappa\rho)Y_{lm}(\theta,\phi) \tag{7.1}$$

and

$$\chi=\sum_{l,m}\beta_{lm}k_l(\kappa\rho)Y_{lm}(\theta,\phi)\ , \tag{7.2}$$

where $i_l(x)$ and $k_l(x)$ are spherical modified Bessel functions of the first and second kind, respectively,[22] see Appendix A. The Wronskian of these functions is[22]

$$i_l(x)k_l'(x)-i_l'(x)k_l(x)=-x^{-2}\ . \tag{7.3}$$

The sum rule, Eq. (6.11), immediately gives the general expression for the tunneling matrix elements,

$$M=\frac{\hbar^2}{2m_e\kappa}\sum_{l,m}\alpha_{lm}\beta_{lm}^*\ . \tag{7.4}$$

The coefficients β_{lm} are determined by fitting the tip wave function on and beyond the separation surface using Eq. (7.2). Inside the tip body, the actual tip wave function does not satisfy Eq. (3.1), and the expansion (7.2) does not represent the actual tip wave function. Instead, it represents the *vacuum continuation* of the vacuum tail of the tip wave function into the tip body. According to the general discussion in Sec. VI, such a function must have at least one singularity in the tip body. Indeed, at the origin of the spherical coordinates, which is the nucleus of the apex atom, every term in expansion (7.2) diverges. It is a singularity in the tip wave function as well as Eq. (3.1). Nevertheless, in the entire region of the tip body except this singular point, both the vacuum continuation of the tip wave function (7.2) and the sample wave function satisfies Schrödinger's equation in the vacuum, Eq. (3.1). Under this condition, the actual separation surface for calculating tunneling matrix elements using Bardeen's formula is irrelevant, as long as it encloses this singular point. [Actually, it is easy to show that for a Bardeen integral on a closed surface, if both χ and ψ satisfy Eq. (3.1) in the entire volume enclosed by this surface, then the integral is zero.] Therefore, for evaluating the integral using the vacuum continuation of the vacuum tail of the tip wave function, one may take any sphere around the singularity, which is the nucleus of the apex atom.

An observation on Eq. (7.4) reveals immediately that in spherical coordinates, the sum rule simply means that for each component of the tip wave function with angular dependence characterized by l and m, the tunneling matrix element is proportional to the corresponding component of the sample wave function with the same angular dependence.

In the following, we show that the coefficients α_{lm} in Eq. (7.1) are related to the derivatives of the sample wave function with respect to x, y, and z at the center of the apex atom (i.e., the center of the spherical coordinate system) in an extremely simple way. We thereby obtain the derivative rule again, from a completely different point of view.

The key of the proof is the properties of the spherical modified Bessel function of the first kind, $i_l(\kappa\rho)$. For small values of ρ, the function $i_l(\kappa\rho)$ has the following form:

$$i_l(\kappa\rho)=\frac{(\kappa\rho)^l}{(2l+1)!!}[1+O(\rho^2)]\ . \tag{7.5}$$

By multiplying it with a spherical harmonics of order l, this term becomes a homogeneous polynomial of x, y, and z or order l. By taking a partial derivative of $i_l(\kappa r)$ with respect to x, y, and z with (summed) order n, and taking the value at $r=0$, all the terms with powers of x, y, and z of $l>n$ drop off. In particular, for the cases of $l=0$ and $l=1$, there is only one term left in the derivative at $\rho=0$, i.e., a term containing only one coefficient in Eq. (7.4). For $l=2$, the derivative may contain the second term in $i_0(\kappa\rho)$, which should be subtracted off to obtain the coefficient for an $l=2$ component.

We start our derivation by writing down the explicit

form of the vacuum asymptote of a tip wave function (as well as its vacuum continuation in the tip body). As we have explained in Sec. III, for the simplicity of relevant mathematics, the rather complicated normalization constants of the spherical harmonics are absorbed in the expression of the sample wave function. Up to $l=2$, we define the coefficients of the expansion by the following expression:

$$\chi=\beta_{00}k_0(\kappa\rho)+[\beta_{10}(z/\rho)+\beta_{11}(x/\rho)+\beta_{12}(y/\rho)]k_1(\kappa\rho)$$
$$+[\beta_{20}(z^2/\rho^2-\tfrac{1}{3})+\beta_{21}(xz/\rho^2)+\beta_{22}(yz/\rho^2)+\beta_{23}(xy/\rho^2)+\beta_{24}(x^2/\rho^2-y^2/\rho^2)]k_2(\kappa\rho)\ . \tag{7.6}$$

Similarly, for the sample wave function, up to the lowest significant term in the power expansion of the spherical modified Bessel function of the first kind, $i_l(\kappa\rho)$,

$$4\pi\psi=\alpha_{00}[1+\tfrac{1}{6}(\kappa\rho)^2]+\alpha_{10}\kappa z+\alpha_{11}\kappa x+\alpha_{12}\kappa y+\alpha_{20}(3\kappa^2/4)(z^2-\tfrac{1}{3}\rho^2)$$
$$+\alpha_{21}(\kappa^2 xz)+\alpha_{22}(\kappa^2 zy)+\alpha_{23}(\kappa^2 xy)+\alpha_{24}[\kappa^2(x^2-y^2)]+O(\rho^3)\ . \tag{7.7}$$

The factor 4π is introduced for convenience. Now, it is straightforward to obtain a relation between the coefficients α_{lm} and the derivatives of sample wave functions. For example, because of $\rho^2=x^2+y^2+z^2$, we have

$$(4\pi/\kappa^2)(\partial^2/\partial z^2)\psi(\mathbf{r}_0)=\alpha_{20}+\tfrac{1}{3}\alpha_{00}\ . \tag{7.8}$$

Noticing that

$$4\pi\psi(\mathbf{r}_0)=\alpha_{00}\ , \tag{7.9}$$

we obtain

$$\alpha_{20}=4\pi(\kappa^{-2}\partial^2/\partial z^2-\tfrac{1}{3})\psi(\mathbf{r}_0)\ . \tag{7.10}$$

The derivations for other components are very straightforward. Therefore, the tunneling matrix element for an arbitrary tip state, up to $l=2$, is

$$M=\frac{2\pi\hbar^2}{m_e\kappa}\left[\beta_{00}\psi+\beta_{10}\frac{\partial\psi}{\kappa\partial z}+\beta_{11}\frac{\partial\psi}{\kappa\partial x}+\beta_{12}\frac{\partial\psi}{\kappa\partial y}+\beta_{20}\left[\frac{\partial^2\psi}{\kappa^2\partial z^2}-\tfrac{1}{3}\psi\right]+\beta_{21}\frac{\partial^2\psi}{\kappa^2\partial x\,\partial z}+\beta_{22}\frac{\partial^2\psi}{\kappa^2\partial y\,\partial z}\right.$$
$$\left.+\beta_{23}\frac{\partial^2\psi}{\kappa^2\partial x\,\partial y}+\beta_{24}\left[\frac{\partial^2\psi}{\kappa^2\partial x^2}-\frac{\partial^2\psi}{\kappa^2\partial y^2}\right]\right]\ . \tag{7.11}$$

The values of the derivatives are taken at $\mathbf{r}_0$.

Again, we obtain the "derivatives rule" from a completely different point of view. The second derivation gives a general formula to calculate the tunneling matrix element for an arbitrary tip wave function, with its vacuum tail expanded in terms of spherical harmonics.

The present approach is convenient for treating a single localized surface state at the tip. If the tip wave function is more complicated, the expansion (7.2) may converge very slowly or not converge at all. An example is the case of two tip atoms at almost the same vertical distance to the sample surface, but with a large horizontal distance in x,y. This is a commonly occurring pathological condition, the double tip. In this case, the total tunneling current can be considered as the sum of the two components of tunneling current, each from one of these two tip atoms, evaluated separately using the derivative rule.

VIII. EFFECT OF DIFFERENT TIP STATES

Using the expressions of the tunneling matrix elements derived in the previous sections, theoretical predictions for STM images can be made. Instead of discussing actual examples of solid surfaces, in this section we illustrate the effect of different tip states by considering a simple metal surface with a one-dimensional periodicity a as well as a reflection symmetry at $x=0$. In the following discussion, we will use the method of Harris and Liebsch[24] to relate the surface charge density as well as the STM images to the Bloch waves at several symmetric points. This method has been used extensively in analyzing helium scattering data.[24] The general form of the surface charge density, up to the lowest nontrivial Fourier components, is

$$\rho(x,z)=\sum_{E\le E_F}|\psi(\mathbf{r})|^2\simeq a_0(z)+a_1(z)\cos^2(\tfrac{1}{2}qx)\ , \tag{8.1}$$

where $q=2\pi/a$ is the length of the primitive reciprocal-lattice vector, and the sum is carried over the entire valence band (Fig. 4). The origin of z is defined as the plane of the top layer nuclei of the sample. The square of the cosine is more convenient than the cosine itself because it is easier to be correlated with corrugation amplitudes as well as surface wave functions.

To correlate the surface change density with the sur-

face Bloch functions, we consider the contributions from different regions in the first Brillouin zone, similar to the method of Harris and Liebsch.[24] To simplify the notation, all the constants that are independent of coordinates are denoted as const. As is shown in Fig. 4, the term $a_0(z)$ in Eq. (1) originates from the constant term in the Bloch functions, i.e., the Bloch functions near point $\bar{\Gamma}$, whose lowest Fourier component is

$$\psi_{\bar{\Gamma}}=\text{const}\times\exp(-\kappa z)\ , \tag{8.2}$$

which makes

$$a_0(z)=\text{const}\times\exp(-2\kappa z)\ . \tag{8.3}$$

The contribution to the second term of Eq. (8.1) comes from Bloch functions near the $\bar{K}$ points.[24] In general, a surface Bloch function at that point has the form

$$\psi_{\bar{K}}=\text{const}\times\exp(\tfrac{1}{2}iqx)\times\sum_n \exp\{-[\kappa^2+(\tfrac{1}{2}q+nq)^2]^{1/2}z\}\exp(inqx)\ , \tag{8.4}$$

where $(\frac{1}{2}q)$ is the magnitude of the wave vector at point $\bar{K}$ in the reciprocal space (Fig. 4). In addition to the term with $n=0$, the term with $n=-1$ has the same decay length, and thus the same magnitude. Also, the Bloch function that generates the symmetric charge density must also by symmetric. The only possible symmetric lowest-order Fourier sum of the Bloch function near point $\bar{K}$ is

$$\psi_{\bar{K}}=\text{const}\times\exp[-(\kappa^2+\tfrac{1}{4}q^2)^{1/2}z]\cos(\tfrac{1}{2}qx)\ . \tag{8.5}$$

The charge density is proportional to $|\psi_{\bar{K}}|^2$. Combining with Eq. (8.3), the total charge density is then

$$\begin{aligned}\rho(\mathbf{r})&=\sum_{E\le E_F}|\psi(\mathbf{r})|^2\\&=C_0\exp(-2\kappa z)\\&\quad+C_1\exp[-2(\kappa^2+\tfrac{1}{4}q^2)^{1/2}z]\cos^2(\tfrac{1}{2}qx)\ ,\end{aligned} \tag{8.6}$$

where C_0, C_1 are constants. As shown below, for predicting theoretical images, the only relevant quantity is C_1/C_0.

In the following, we discuss the intercorrelation between tip states and images one by one.

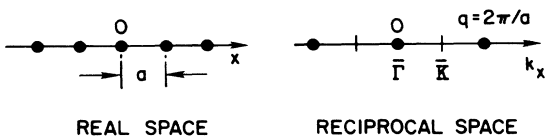

FIG. 4. A metal surface with one-dimensional periodicity, with a reflection symmetry at $x=0$. See the discussion of the effect of different tip states on the images. The reciprocal space is also shown.

A. *s*-wave tip state

This case has been discussed extensively by Tersoff and Hamann.[6] For metals, the tunneling current is

$$\begin{aligned}I&=\text{const}\times\sum_{E_F\le E\le E_F+eV}|\psi(\mathbf{r}_0)|^2\\&\propto C_0\exp(-2\kappa z)+C_1\exp[-2(\kappa^2+\tfrac{1}{4}q^2)^{1/2}z]\cos^2(\tfrac{1}{2}qx)\ .\end{aligned} \tag{8.7}$$

In the *s*-wave model of Tersoff and Hamamm, $\mathbf{r}_0$ is the center of the spherical potential well with radius R, which is used to represent the tip.[6] For free-electron metals, the local density of states near the Fermi level is proportional to the total valence-electron charge density. Therefore, up to an overall constant depending on bias V, the tunneling current is

$$I=\text{const}\times\rho(x,z)\ . \tag{8.8}$$

The *corrugation amplitude* (or simply *corrugation*[1-5]) defined as the difference of the maximum z and the minimum z (over the entire surface under consideration) with a constant tunneling current, follows immediately from Eq. (8.6):

$$\begin{aligned}\Delta z_s&=(C_1/C_0)(2\kappa)^{-1}\exp\{-2[(\kappa^2+\tfrac{1}{4}q^2)^{1/2}-\kappa]z\}\\&\quad+O(\exp\{-4[(\kappa^2+\tfrac{1}{4}q^2)^{1/2}-\kappa]z\})\ .\end{aligned} \tag{8.9}$$

Thus, for small corrugations, the theoretical topographic image arising from an *s*-wave tip state is

$$z(x)=\Delta z_s\cos^2(\tfrac{1}{2}qx) \tag{8.10}$$

and the corresponding constant-z current image is

$$\begin{aligned}I(x)&=I_0(1+(C_1/C_0)\exp\{-2[(\kappa^2+\tfrac{1}{4}q^2)^{1/2}-\kappa]z\}\\&\quad\times\cos^2(\tfrac{1}{2}qx))\ .\end{aligned} \tag{8.11}$$

By substituting the actual numbers into Eq. (8.1), we recover this well-known fact: Even at an extremely short tip-sample distance, e.g., 3 Å, and with an unrealistically small tip radius R, e.g., zero, for low-Miller-index metal surfaces, the corrugation amplitude of such a Fermi-level LDOS is smaller than 0.03 Å, which is at least one order of magnitude smaller than what has been observed expermentally.[3]

B. p_z tip state

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

$$M_{\bar{K}}=\text{const}\times(1+\tfrac{1}{4}q^2/\kappa^2)^{1/2}\psi_{\bar{K}} \tag{8.12}$$

and the topographic image arising from a p_z tip state is

$$z(x)=(1+\tfrac{1}{4}q^2/\kappa^2)\Delta z_s\cos^2(\tfrac{1}{2}qx)\ . \tag{8.13}$$

Therefore, the tunneling matrix element arising from a p_z

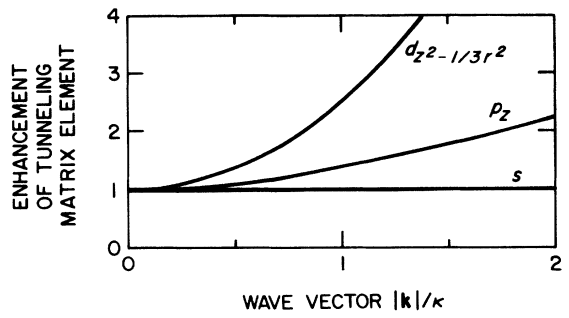

FIG. 5. Enhancement of tunneling matrix elements arising from nonspherical tip states. The tunneling current is proportional to the square of the tunneling matrix element. Therefore, the enhancement factor for the tunneling current is the square of the enhancement factor for the tunneling matrix element.

state gains a factor of $(1+\frac{1}{4}q^2/\kappa^2)^{1/2}$ over the s-wave tip state, and the corrugation amplitude arising from a p_z tip state gains a factor of $(1+\frac{1}{4}q^2/\kappa^2)$ over that of the charge-density contour (see Fig. 5). This is the quantitative explanation of the resolution enhancement due to p-like localized tip states.[10]

C. d_{z^2} tip state

According to the derivative rule, for a d_{z^2} tip state, the tunneling matrix element for a sample wave function at point $\bar{\Gamma}$ picks up a factor $\frac{2}{3}$, whereas for a sample wave function at point $\bar{K}$ picks up a factor $(\frac{2}{3}+\frac{1}{4}q^2/\kappa^2)$. Similar to the case of the p_z tip state, the topographic image is

$$z(x)=(1+\tfrac{3}{8}q^2/\kappa^2)^2\Delta z_s\cos^2(\tfrac{1}{2}qx)\ . \tag{8.14}$$

The enhancement for the tunneling matrix element is shown in Fig. 5. The enhancement factor for corrugation amplitude $(1+\frac{3}{8}q^2/\kappa^2)^2$ could be substantial. For example, on most close-packed metal surfaces, $a\simeq 2.5$ Å, which implies $q\simeq 2.5$ Å^{-1}. An enhancement of 11.2 is expected. As we have mentioned, all the commonly used tip materials are d-band metals:[13] for example, W, Pt, and Ir. Experiments and theories have shown that localized d_{z^2} states often occur on the surfaces.[8,12,17–20] The derivative enhancement of corrugation amplitude, due to these d_{z^2} tip states, is probably the prime origin of atomic resolution observed by STM.[3,12]

D. p_x tip state

For a pure p_x tip state, the tunneling matrix element to the sample wave function at point $\bar{\Gamma}$ is negligible, whereas for sample wave functions at point $\bar{K}$, a phase shift of 90° is obtained. The tunneling current is

$$I=\text{const}\times(\tfrac{1}{4}q^2/\kappa^2)\sin^2(\tfrac{1}{2}qx)\ . \tag{8.15}$$

The meaning of this result was discussed[11] in conjunction with Tersoff's model for STM images of semiconductors and semimetals,[25] where the relevant surface wave function may occur only at the edge of the first Brilluoin zone, i.e., the $\bar{K}$ point. In one-dimensional models, for the lowest Fourier component, a single sinusoidal wave may dominate the surface wave function.[25] By considering the s-wave tip state only, Tersoff argued that the nodal structure of this state could give rise to very large corrugations,[25] e.g., $\Delta z>3$ Å. Lawunmi and Payne[26] made a careful analysis of Tersoff's model. They show that in any realistic case, on a real tip, the existence of a small component of the p wave is always expected. With any realistic value of the p_x component, the corrugation will reduce to less than 1 Å. Therefore, Lawunmi and Payne concluded[26] that Tersoff's s-wave model for the observed large corrugation on graphite[25] cannot be correct.

E. d_{x^2} tip state

A straightforward calculation using the tunneling matrix elements listed in Table IV shows that the d_{x^2} state results in a very large corrugation amplitude on metal surfaces, because the tunneling matrix element for the sample wave function at the $\bar{\Gamma}$ point vanishes. Similar to the previous case, the existence of other states must be taken into account. Also, since there is very little study of the existence of these localized states on the tip, we will leave it for further investigation.

IX. RESULTS IN PARABOLIC COORDINATES

As we mentioned in Sec. VI, the sum rule is valid for a number of curvilinear coordinate systems. In a certain sense, the spherical coordinate expansion of the tip wave functions may not be the most convenient one for STM. In STM, there is a center (the apex atom) as well as a direction (the sample surface). For treating such problems, for example, a hydrogen atom in an external electrical field, i.e., the Stark effect, the parabolic coordinate system (ξ,η,ϕ) is a natural choice. It is defined as[27]

$$x=\sqrt{\xi\eta}\cos\phi\ , \tag{9.1}$$

$$y=\sqrt{\xi\eta}\sin\phi\ , \tag{9.2}$$

$$z=(\xi-\eta)/2\ , \tag{9.3}$$

and it follows that

$$r=(\xi+\eta)/2\ . \tag{9.4}$$

The separation surface is described by a paraboloid $r-z=\eta_0$. The interior and the exterior of the tip correspond to regions $\eta<\eta_0$ and $\eta>\eta_0$, respectively.

The Schrödinger equation in the vacuum, Eq. (3.1), in terms of parabolic coordinates, is

$$\frac{4}{\xi+\eta}\left[\frac{\partial}{\partial\xi}\left[\xi\frac{\partial\Phi}{\partial\xi}\right]+\frac{\partial}{\partial\eta}\left[\eta\frac{\partial\Phi}{\partial\eta}\right]\right]+\frac{1}{\xi\eta}\frac{\partial^2\Phi}{\partial\phi^2}=\kappa^2\Phi\ . \tag{9.5}$$

With the substitution[27]

$$\Phi=e^{-\kappa(\xi+\eta)/2}(\xi\eta)^{|m|/2}f(\eta)g(\xi)e^{im\phi}\ , \tag{9.6}$$

one finds that both $f(\eta)$ and $g(\xi)$ satisfy the confluent hypergeometric equation, or Kummer's equation.[28] The boundary condition for $g(\xi)$ is that it should be regular at

both $\xi=0$ and $\xi=\infty$. It leads to an eigenvalue problem and yields the well-known associated Laguerre polynomials,[28] which appear in the problem of hydrogen atom:

$$g(\xi)=L_n^{|m|}(\kappa\xi) , \tag{9.7}$$

where n is an integer. The equation for $f(\eta)$ is

$$\kappa\eta f'(\kappa\eta)+(1+|m|-\kappa\eta)f(\kappa\eta)-(n+|m|+1)f(\kappa\eta)=0 . \tag{9.8}$$

It does not have any solution which is regular at both $\eta=0$ and $\eta=\infty$. Therefore, different solutions have to be assigned for the sample and the tip. The sample wave function should be regular at $\eta=0$, which is the first Kummer function:[28]

$$f_S(\eta)=M(n+|m|+1,|m|+1;\kappa\eta) . \tag{9.9}$$

It diverges at large η, which means that inside the sample body, there should be some singularities, which are as expected. On the other hand, the tip wave function should be regular at $\eta=\infty$, i.e., on the sample side, which implies that Kummer's second function is the right choice,[28]

$$f_T(\eta)=U(n+|m|+1,|m|+1;\kappa\eta) . \tag{9.10}$$

It diverges as $\eta\rightarrow 0$, which means that at the focus of the tip, there should be some singularity. Since the focus of the paraboloid is the center of the acting atom, the singularity is also as expected. The basic sample wave functions and basic tip wave functions in terms of parabolic coordinates are as follows:

$$\psi_{nm}=e^{-\kappa(\xi+\eta)/2}(\xi\eta)^{|m|/2}M(n+|m|+1,|m|+1;\kappa\eta) \times L_n^{|m|}(\kappa\xi)e^{im\phi} , \tag{9.11}$$

$$\chi_{nm}=e^{-\kappa(\xi+\eta)/2}(\xi\eta)^{|m|/2}U(n+|m|+1,|m|+1;\kappa\eta) \times L_n^{|m|}(\kappa\xi)e^{im\phi} . \tag{9.12}$$

Those Kummer functions, although unfamiliar at first sight, provide a reasonable and understandable description of the tip wave functions and sample wave functions, especially when they are converted into Cartesian coordinates.

According to the general method described in Sec. IV, the most general forms of the sample wave function and the tip wave function are

$$\psi=\sum_{n,m}\alpha_{nm}\psi_{nm}, \quad \eta\le\eta_0 \tag{9.13}$$

$$\chi=\sum_{n,m}\beta_{lm}\chi_{nm}, \quad \eta\ge\eta_0 . \tag{9.14}$$

According to the general sum rule, remember that the Wronskian of the standard Kummer functions is[28]

$$W(M,U)=-\frac{|m|!}{(n+|m|)!}e^{\kappa\eta_0}(\kappa\eta_0)^{-|m|-1} . \tag{9.15}$$

We find the tunneling matrix element

$$M_{\mu\nu}=-\frac{\pi\hbar^2}{2m_e\kappa}\sum_{n,m}\frac{(n+|m|)!}{n!}\alpha_{nm}\beta_{nm}^* . \tag{9.16}$$

In Appendix D, the lower-order tip and sample wave functions in the parabolic coordinate system are listed. For clarity, the arguments of the wave functions are converted into Cartesian coordinates using the definition of parabolic coordinates, Eqs. (9.1)–(9.3). For the sample wave functions, every component has a common factor $\exp(-\kappa z)$, which is multiplied by simple polynomials of x,y,z.

Consider the polynomials with $n=0$ first. Up to $m=2$, the sample wave function near the focus of the tip is

$$\psi=e^{-\kappa z}[\alpha_\sigma+\alpha_{\pi_x}x+\alpha_{\pi_y}y+\alpha_{\delta_{x^2-y^2}}(x^2-y^2)+2\alpha_{\delta_{xy}}xy] . \tag{9.17}$$

The subscripts are assigned according to the irreducible representations of the point group $C_{\infty v}$. The coefficients can be evaluated by taking derivatives of the wave function with respect to x,y. Denoting the center of the apex atom, i.e., the focus of the parabolic coordinate system, in a Cartesian-coordinate system fixed on the sample as $\mathbf{r}_0$, the coefficients are

$$\alpha_\sigma=e^{\kappa z}\psi(\mathbf{r}_0) ,$$

$$\alpha_{\pi x}=\frac{\partial}{\partial x}[e^{\kappa z}\psi(\mathbf{r}_0)] ,$$

$$\alpha_{\pi y}=\frac{\partial}{\partial y}[e^{\kappa z}\psi(\mathbf{r}_0)] ,$$

$$\alpha_{\delta_{x^2-y^2}}=\frac{1}{2}\left[\frac{\partial^2}{\partial y^2}-\frac{\partial^2}{\partial y^2}\right]e^{\kappa z}\psi(\mathbf{r}_0) ,$$

$$\alpha_{\delta_{xy}}=\frac{1}{2}\frac{\partial^2}{\partial x\,\partial y}[e^{\kappa z}\psi(\mathbf{r}_0)] .$$

Those coefficients have a direct relation with the tunneling matrix elements.

Next, we consider the meaning of the tip states. Since the tunneling is due to the tail of tip wave functions outside the tip in the $-z$ direction, we are interested in the asymptote of those wave functions for large $-z$. For the $n=0$ states,

$$\chi_\sigma\simeq\frac{e^{-\kappa r}}{2r} ,$$

$$\chi_{\pi_x}\simeq\frac{e^{-\kappa r}x}{4r^2} ,$$

$$\chi_{\pi_y}\simeq\frac{e^{-\kappa r}y}{4r^2} , \tag{9.18}$$

$$\chi_{\delta_{x^2-y^2}}\simeq\frac{e^{-\kappa r}(x^2-y^2)}{8r^3} ,$$

$$\chi_{\delta_{xy}}\simeq\frac{e^{-\kappa r}xy}{4r^3} .$$

Apparently, those asymptotes of tip wave functions are identical to the atomic wave functions of a spherical potential well with s,p,d symmetry, which describe the extension of atomic wave functions beyond the apex of the tip.

TABLE V. Derivative rule in the parabolic coordinate system.

Tip state	Tunneling matrix element (up to a constant factor)
σ	$\psi\ (\mathbf{r}_0)$
π_x	$\frac{\partial\psi}{\partial x}\ (\mathbf{r}_0)$
π_y	$\frac{\partial\psi}{\partial y}\ (\mathbf{r}_0)$
$\delta_{x^2-y^2}$	$\left[\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial y^2}\right]\ (\mathbf{r}_0)$
δ_{xy}	$\frac{\partial^2\psi}{\partial x\,\partial y}\ (\mathbf{r}_0)$

Using the sum rule for paraboloidal coordinate system, Eq. (9.16), and the explicit expression of coefficients α's, Eq. (9.18), we obtain the derivative rule for tip states in the parabolic coordinate system, as shown in Table V.

X. MACROSCOPIC LIMIT

In this section, we show that the s-wave model[6] is the macroscopic limit of the microscopic theory under very general conditions. In this *macroscopic limit,* the personality of different tip states disappears, and the image becomes the macroscopic contour of the surface, i.e., the contour of the local density of states.[6]

The macroscopic limit can be approached from two different points of view: either by considering only the large features on the surface, i.e., features with linear dimensions much larger than the typical diameter of an atom, or by assuming a large tip-sample separation such that the atomic information fades away.

To illustrate the first point of view, we consider again the Fourier expansion of the surface wave functions. A component with lateral wave vector $\mathbf{q}$ is

$$\psi(\mathbf{q})=C(\mathbf{q})\exp[-(\kappa^2+|\mathbf{q}|^2)^{1/2}z]\exp(i\mathbf{q}\cdot\mathbf{x})\ , \qquad (10.1)$$

where $\mathbf{x}=(x,y)$. The large features contain only long-wavelength Fourier components, which satisfy the validity condition of the Tersoff-Hamann model:[6]

$$(1+|\mathbf{q}|^2/\kappa^2)\simeq 1\ . \qquad (10.2)$$

Using the derivative rule, we find immediately that for those large features (1) the tunneling matrix elements for $l\neq 0$, $m=0$ states are identical to that for an s-wave tip state, and (2) the tunneling matrix elements for any $m\neq 0$ tip state is much smaller than that for an s-wave tip state. Therefore, by considering the long-wavelength components only, the tunneling current is proportional to the LDOS, or the charge density (for metals), at the center of the tip.[6]

Let us examine the meaning of the condition Eq. (10.2). In terms of the wavelength L, related to the lateral wave vector $\mathbf{q}$, Eq. (10.2) becomes

$$L\gg 2\pi/\kappa\ . \qquad (10.3)$$

Since κ^{-1} is the thickness of the electron tail on the surface, Eqs. (10.2) and (10.3) simply mean that the feature scale under consideration is much larger than the typical thickness of the electron tail. In this case, there is a uniquely defined contour on the surface, that is, the LDOS contour. This contour is exactly the macroscopic (jellium) contour of the metal surface. This point can be further clarified by considering the z dependence of the corrugation amplitude. Under the condition $|\mathbf{q}|\ll\kappa$, the corrugation amplitude varies with z as[6]

$$\Delta z=\text{const}\times\exp(-\tfrac{1}{4}\kappa^{-1}|\mathbf{q}|^2 z)\ . \qquad (10.4)$$

Since $|\mathbf{q}|\ll\kappa$, the corrugation amplitude varies much more slowly with z, compared with the variation of the absolute value of current with z. In other words, the image represents a contour that is roughly independent of the distance from the solid surface within a reasonable distance.[6]

The typical value of κ is 1 Å^{-1}. Therefore, Eq. (10.3) means

$$L\gg 6\ \text{Å}\ . \qquad (10.5)$$

Since the typical diameter of an atom is 2–5 Å, condition (10.5) simply means to disregard atom-scale features from the very beginning.[6]

The second way to approach the macroscopic limit is to assume a large tip-sample distance, for example, larger than 6 Å.[15] Again, from an observation of Eq. (10.1) one finds that the Fourier components with $|\mathbf{q}|$ larger than or comparable with κ have much larger decay constants than Fourier components with $|\mathbf{q}|\ll\kappa$. Therefore, at the large-distance limit, only the Fourier components with $|\mathbf{q}|\ll\kappa$ could survive. An interesting fact is that for *all* long-wavelength Fourier components, the decay constant is roughly the same, κ, whereas for the short-wavelength components, the decay constants $(\kappa^2+|\mathbf{q}|^2)^{1/2}$ vary from component to component. In other words, the corrugation amplitudes of atom-size features decay at much faster rates with z than the long-wavelength components, and the rates differ from feature to feature according to their dimension. On the other hand, the equal-LDOS contours of the macroscopic features, i.e., the long-wavelength components in the Fourier expansion of the surface wave function, vary only slightly within the entire operational distance of STM. Considering the macroscopic features only, the exact definition of the "center" of the tip is not critical. Therefore, the conclusion of Tersoff and Hamann[6] that the STM image is directly connected to the LDOS contour at *the center of curvature of the tip* is valid, *only in the macroscopic limit.* Actually, the basic parameters in the s-wave model, *the radius of curvature* and *the center of curvature* of the tip,[6] have no physical meaning from a microscopic point of view.

XI. CONCLUSION

We have derived and listed the tunneling matrix elements for tip states up to $l=2$ using two completely independent methods, the Green's-function method and through the sum rule. We have shown that the tunneling matrix elements are related to the *derivatives* of the sample wave functions at the center of the apex atom in a

simple and straightforward way. Physical meanings of these matrix elements as well as implications to imaging mechanism of scanning tunneling microscopy are discussed. The results can be used to predict the image using the knowledge of the tip state, and to infer the tip state from the observed image.

It is clear that in order to understand the atomic resolution achieved by STM, as well as the spontaneous switching of instrument resolution during imaging, the knowledge about the electronic states at the tip, especially the localized surface states, is essential. Yet there is not enough knowledge about these states. The most common tip material is tungsten. The only first-principles calculation of the electronic states of a STM tip is the W cluster calculation by Ohnishi *et al.*[14] Because of the limited size of clusters, only discrete states are obtained. The results are important, but not conclusive. The first-principles calculation of the W(001) surface is relevant,[19,20] but not conclusive either. A more relevant calculation will probably be an all-electron first-principles calculation of a W slab of certain orientation with an extra W atom (or Si atom) adsorbed on an appropriate site. Thus, the wave function as well as the density of states can be obtained simultaneously. This kind of calculation is not beyond the reach of current computational physics. For example, all-electron calculations of a W(100) slab and a Mo(100) slab with a Cs atom adsorbed have been performed.[29] The second most important tip material is platinum. Since the requirement for calculating localized surface states on platinum structures is very similar to that for tungsten, reliable information can be obtained as well.

Note added in proof. In some cases, the macroscopic limit of the microscopic theory presented here does not agree with the Tersoff-Hamann model, especially for the case of graphite.[15] Recently, the case of graphite has been treated using the first-principles method.[32] In the Tersoff-Hamann model, for the case of graphite, at low bias, the only sample states involved in the imaging process are the six sine waves at the corners of the surface Brillouin zone, and the corrugation amplitude is infinite.[15] The *s*-wave model also predicts that, on graphite, the average tunneling current should decay about 50–60 times per angstrom.[15] Using an improved *s*-wave model, the predicted corrugation is finite and *independent of* z.[33] Therefore, even at the large-distance limit, the STM image of graphite predicted by the *s*-wave models has a constant corrugation amplitude.[15,33] The present theory is based on a modified Berdeen approach, which includes the distortion of the sample wave function due to the existence of the tip. For graphite, such tip-induced local states (TILS) have dramatic implications.[32] Using an analytic treatment of TILS, it is shown that the corrugation amplitude of the STM images of graphite in ultrahigh vacuum should decay about 4–5 times per angstrom, and the average tunneling current should decay about 8–9 times per angstrom. This prediction has been verified experimentally.[34]

ACKNOWLEDGMENTS

The authors wishes to thank J. E. Demuth, H. Rohrer, A. Baratoff, I. P. Batra, P. Marcus, J. Tersoff, N. D. Lang, and V. Moruzzi for inspiring discussions, as well as M. Gutzwiller and J. Slonczewski for a critical reading of the manuscript and helpful suggestions, and E. Marino for proofreading the final manuscript.

APPENDIX A: SPHERICAL MODIFIED BESSEL FUNCTIONS

It is difficult to find appropriate references for spherical modified Bessel functions, in spite of their simplicity. Probably, the only extensive source is in Ref. 28, pp. 443–445. However, the definition found therein is inconvenient for applications to tunneling problems, and the notation is confusing. In the following, we provide a list of important formulas of those functions following the definition and notation of Arfken.[22]

The spherical modified Bessel function of the first kind,

$$i_n(z)=(\pi/2z)^{1/2}I_{n+1/2}(z)\ , \tag{A1}$$

and of the second kind,

$$k_n(z)=(2/\pi z)^{1/2}K_{n+1/2}(z)\ , \tag{A2}$$

are linearly independent solutions of the equation

$$\frac{d}{dz}\left[z^2\frac{df(z)}{dz}\right]-[z^2+n(n+1)]f(z)=0\ . \tag{A3}$$

Function $i_n(z)$ is the only solution of Eq. (A3) that is regular at $z=0$, and function $k_n(z)$ is the only solution that is regular at $z=\infty$. These so-called special functions are actually *elementary functions* with the following general expression:

$$i_n(z)=z^n\left[\frac{d}{z\,dz}\right]^n\frac{\sinh z}{z} \tag{A4}$$

and

$$k_n(z)=(-1)^n u^n\left[\frac{d}{z\,dz}\right]^n\frac{\exp(-z)}{z}\ . \tag{A5}$$

The first three pairs of these functions are

$$i_0(z)=\frac{\sinh z}{z} \tag{A6}$$

$$i_1(z)=-\frac{\sinh z}{z^2}+\frac{\cosh z}{z}\ , \tag{A7}$$

$$i_2(z)=-\left[\frac{3}{z^3}+\frac{1}{z}\right]\sinh z+\frac{3}{z^2}\cosh z\ , \tag{A8}$$

$$k_0(z)=\frac{1}{z}e^{-z}\ , \tag{A9}$$

$$k_1(z)=\left[\frac{1}{z}+\frac{1}{z^2}\right]e^{-z}\ , \tag{A10}$$

$$k_2(z)=\left[\frac{1}{z}+\frac{3}{z^2}+\frac{3}{z^3}\right]e^{-z}\ . \tag{A11}$$

The function $i_n(z)$ has the following power-series ex-

pansion near $z=0$:

$$i_n(z)=z^n\sum_{k=0}^{\infty}\frac{1}{k!(2n+2k+1)!!}\left[\frac{z^2}{2}\right]^k . \tag{A12}$$

The first term is proportional to z^n:

$$i_n(z)\simeq\frac{z^n}{(2n+1)!!} . \tag{A13}$$

An $i_n(z)$ of even order only has an even power of z, and an $i_n(z)$ of odd order only has an odd power of z. Those properties are essential in a complete derivation of the derivative rule.

On the other hand, the functions $k_n(z)$ have the following exact general expression:

$$k_n(z)=\frac{e^{-z}}{z}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!}\frac{1}{(2z)^k} . \tag{A14}$$

The following recursion relations are found to be useful, which apply to both i_n and k_n:

$$(2n+1)f_n(z)=zf_{n-1}(z)-zf_{n+1}(z) , \tag{A15}$$

$$\frac{d}{dz}f_n(z)=f_{n+1}(z)+\frac{n}{z}f_n(z) . \tag{A16}$$

Finally, the Wronskian of the pair is simply

$$W\{i_n(z),k_n(z)\}\equiv i_n(z)k_n'(z)-i_n'(z)=-z^{-2} . \tag{A17}$$

APPENDIX B: LIOUVILLE'S THEOREM FOR MODIFIED HELMHOLTZ EQUATION

In this Appendix, we show that if the absolute value of a solution $\chi(\mathbf{r})$ of the modified Helmholtz equation,

$$(\nabla^2-\kappa^2)\chi(\mathbf{r})=0 , \tag{B1}$$

is bounded in the entire space, in other words, $|\chi(\mathbf{r})|<M$ (a constant), then $\chi(\mathbf{r})$ must be zero everywhere in the entire space. Since this statement is very similar to Liouville's theorem in the theory of analytic functions of complex variables,[23] it can be considered as a variation of Liouville's theorem.

It is known that the solution of the Dirichlet problem for Eq. (B1) is unique.[30] In other words, if the value of $\chi(\mathbf{r})$ on the boundary S of a volume Ω is given, then the value of $\chi(\mathbf{r})$ at any point in the volume Ω is uniquely determined by Eq. (B1). Actually, the value of $\chi(\mathbf{r})$ at any point $\mathbf{r}_0$ can be represented by the following Green's-function solution:[30]

$$\chi(\mathbf{r}_0)=-\int_S\chi(\mathbf{r}_S)[\nabla G(\mathbf{r}_0,\mathbf{r})]_{\mathbf{r}=\mathbf{r}_S}\cdot d\mathbf{S} , \tag{B2}$$

where $\mathbf{r}_S$ denotes the points on the boundary. The Green's function is defined by[30]

$$G(\mathbf{r}_0,\mathbf{r})=(4\pi|\mathbf{r}_0-\mathbf{r}|)^{-1}\exp(-\kappa|\mathbf{r}_0-\mathbf{r}|)+g(\mathbf{r}_0,\mathbf{r}) , \tag{B3}$$

and the function g is the solution of Dirichlet's problem of Eq. (B1) with boundary condition

$$g(\mathbf{r}_0,\mathbf{r})|_{\mathbf{r}=\mathbf{r}_S}=-(4\pi|\mathbf{r}_0-\mathbf{r}_S|)^{-1}\exp(-\kappa|\mathbf{r}_0-\mathbf{r}_S|) . \tag{B4}$$

To prove Liouville's theorem for Eq. (B1), we take a sphere of radius R centered at $\mathbf{r}_0$ as the boundary. Using Eqs. (B3) and (B4), it is easy to show that

$$g(\mathbf{r}_0,\mathbf{r})=-\frac{\exp(-\kappa R)}{4\pi\sinh(\kappa R)}\frac{\sinh(\kappa|\mathbf{r}-\mathbf{r}_0|)}{|\mathbf{r}-\mathbf{r}_0|} . \tag{B5}$$

A direct calculation gives the value of the Green's function on the boundary:

$$[\nabla G]_S=-(\kappa/4\pi R)\sinh(\kappa R) . \tag{B6}$$

Substituting Eq. (B6) for Eq. (B2) and taking the absolute value on both sides, we obtain

$$|\chi(\mathbf{r}_0)|\le M\kappa R/\sinh(\kappa R) . \tag{B7}$$

Letting $R\to\infty$, we find that for every point $\mathbf{r}$ in the entire space,

$$\chi(\mathbf{r})\equiv 0 . \tag{B8}$$

Thus, Liouville's theorem for the modified Helmholtz equation is proved.

APPENDIX C: THE ELIMINATION EQUATION

In this Appendix, we will find the conditions for the validity of the elimination equation, and show that for most of the separable curvilinear coordinate systems, the elimination equation is valid.

It is well known that the Helmholtz equation in curvilinear coordinates,

$$\frac{1}{h_1h_2h_3}\sum_{n=1}^{3}\frac{\partial}{\partial\xi_n}\left[\frac{h_1h_2h_3}{h_n^2}\frac{\partial\psi}{\partial\xi_n}\right]+\kappa^2\psi=0 , \tag{C1}$$

can be separated into three ordinary differential equations in which the function ψ is expressed as a product of three one-variables functions,

$$\psi=X_1(\xi_1)X_2(\xi_2)X_3(\xi_3) . \tag{C2}$$

The most general form of those ordinary differential equations is

$$\frac{1}{f_n}\frac{d}{d\xi_n}\left[f_n\frac{dX_n}{d\xi_n}\right]+(\kappa^2\Phi_{n1}+\lambda_2\Phi_{n2}+\lambda_3\Phi_{n3})X_n=0 , \tag{C3}$$

where $n=1$, 2, and 3. The nine functions $\Phi_{nk}(\xi_n)$ $(n=1-3,k=1-3)$ are Stäckel functions, and the determinant $\det|\Phi_{nk}|$ is the Stäckel determinant. The Helmholtz equation, Eq. (C1), is separable if and only if the Robertson condition[23] is fulfilled:

$$h_1h_2h_3=\det|\Phi_{nk}|f_1(\xi_1)f_2(\xi_2)f_3(\xi_3) . \tag{C4}$$

A consequence of the Robertson condition is

$$h_2h_3/h_1=(\Phi_{22}\Phi_{33}-\Phi_{23}\Phi_{32})f_1f_2f_3 . \tag{C5}$$

In the following, we will show that if at least one of the off-diagonal Stäckel functions is zero, then the elimination equation is valid. Without losing generality, we assume that $\Phi_{32}=0$. Thus, Eq. (C5) reduces to

$$h_2h_3/h_1=\Phi_{22}\Phi_{33}f_1f_2f_3 . \tag{C6}$$

The ordinary differential equation with respect to ξ_3 is

$$\frac{1}{f_3}\frac{d}{d\xi_3}\left[f_3\frac{dX_3}{d\xi_3}\right]+(\kappa^2\Phi_{31}+\lambda_3\Phi_{33})X_3=0 . \quad \text{(C7)}$$

The boundary condition that X_3 be regular over the entire range of ξ_3 makes Eq. (C7) an eigenvalue problem. Since there is only one parameter, λ_3, by comparing it with the standard form of Sturm-Liouville equation, we find

$$p_3=f_3 , \quad \text{(C8)}$$

$$q_3=\kappa^2 f_3\Phi_{31} , \quad \text{(C9)}$$

$$r_3=f_3\Phi_{33} . \quad \text{(C10)}$$

The ordinary differential equation with respect to ξ_2 is then

$$\frac{1}{f_2}\frac{d}{d\xi_2}\left[f_2\frac{dX_2}{d\xi_2}\right]+(\kappa^2\Phi_{31}+\lambda_2\Phi_{22}+\lambda_3\Phi_{23})X_2=0 . \quad \text{(C11)}$$

Again, it is an eigenvalue problem. Since λ_3 is already fixed, the only parameter for the eigenvalue is λ_2. A similar comparison with the standard form of Sturm-Liouville equation gives

$$p_2=f_2 , \quad \text{(C12)}$$

$$q_2=f_2(\kappa^2\Phi_{21}+\lambda_3\Phi_{23}) , \quad \text{(C13)}$$

$$r_2=f_2\Phi_{22} . \quad \text{(C14)}$$

Substituting (C10) and (C14) for (C6), we find the elimination equation

$$h_1h_2/h_3=p_1(\xi_1)r_2(\xi_2)r_3(\xi_3) . \quad \text{(C15)}$$

An extensive list of Stäkel functions can be found in the classical paper by Eisenhart.[31] As shown, except for conical coordinates, ellipsoidal coordinates, and paraboloidal coordinates, the elimination equation (C15) is valid for all separable coordinate systems. To enumerate, they are oblate spheroidal coordinates, prolate spheroidal coordinates, circular cylinder coordinates, elliptic cylinder coordinates, spherical coordinates, parabolic coordinates, and parabolic cylinder coordinates. Those coordinate systems cover a large number of cases of practical interest.

APPENDIX D: SAMPLE AND TIP WAVE FUNCTIONS IN PARABOLIC COORDINATES

In this Appendix, we are going to discuss the general properties of the sample wave functions and tip wave functions in the parabolic coordinates. Consider the sample wave functions first. From Eq. (9.11),

$$\psi_{nm}=e^{-\kappa(\xi+\eta)/2}(\xi\eta)^{|m|/2}M(n+|m|+1,\ |m|+1;\kappa\eta)\times L_n^{|m|}(\kappa\xi)e^{im\phi} . \quad \text{(D1)}$$

Using the well-known Kummer transformation and the definition of associated Laguerre polynomials,[28] Kummer's function, Eq. (D1), can be converted into

$$M(n+m+1,\ m+1;\kappa\eta)=e^{\kappa\eta}M(-n,\ m+1;-\kappa\eta)=e^{\kappa\eta}\frac{n!\,m!}{(n+m)!}L_n^m(-\kappa\eta) . \quad \text{(D2)}$$

With the expansion formula of Laguerre's polynomials,[19]

$$L_n^m(u)L_n^m(v)=\frac{(n+m)!}{n!}\sum_{k=0}^{n}\frac{L_{n-k}^{m+2k}(u+v)(uv)^k}{(m+k)!k!} , \quad \text{(D3)}$$

the sample wave function can be written as a polynomial of the Cartesian coordinates times a common exponential factor:

$$\psi_{nm}=e^{-\kappa z}(x\pm iy)^{|m|}\,|m|!\sum_{k=0}^{n}\frac{(-1)^k[\kappa^2(x^2+y^2)]^k}{(|m|+k)!\,k!}\times L_{n-k}^{|m|+2k}(2\kappa z) . \quad \text{(D4)}$$

The sign in Eq. (D4) is the sign of m. The first ten real wave functions are listed as

n,m	ψ,
0,0	$e^{-\kappa z}$,
0,1	$e^{-\kappa z}x$,
0,1	$e^{-\kappa z}y$,
0,2	$e^{-\kappa z}(x^2-y^2)$,
0,2	$e^{-\kappa z}2xy$,
1,0	$e^{-\kappa z}[1-2\kappa z-\kappa^2(x^2+y^2)]$,
1,1	$e^{-\kappa z}x\,[2-2\kappa z-\kappa^2(x^2+y^2)]$,
1,1	$e^{-\kappa z}y[2-2\kappa z-\kappa^2(x^2+y^2)]$,
1,2	$e^{-\kappa z}(x^2-y^2)[3-2\kappa z-\frac{1}{3}\kappa^2(x^2+y^2)]$,
1,2	$e^{-\kappa z}2xy[3-2\kappa z-\frac{1}{3}\kappa^2(x^2+y^2)]$.

Next, the tip wave functions. The general form is Eq. (9.12):

$$\chi_{nm}=e^{-\kappa(\xi+\eta)/2}(\xi\eta)^{|m|/2}U(n+|m|+1,|m|+1;\kappa\eta)L_n^{|m|}(\kappa\xi)e^{im\phi} . \quad \text{(D5)}$$

For $n=0, m=0$, Kummer's second function is the exponential integral function[28]

$$U(1,1,;x)=e^xE_1(x) , \quad \text{(D6)}$$

where

$$E_1(x)\equiv\int_x^\infty\frac{e^{-t}}{t}dt . \quad \text{(D7)}$$

The rest of the functions can be calculated by the recursion relations of Kummer functions.[28] One of them gives

$$U(m+1,\ m+1;u)=\frac{(-1)^m}{m!}\frac{d^m}{du^m}[e^uE_1(u)] . \quad \text{(D8)}$$

A list of first five tip wave functions is as follows:

$$
\begin{array}{ll}
n,m & \chi, \\
0,0 & e^{-\kappa r}e^{-\kappa\eta}E_1(\kappa\eta), \\
0,1 & e^{-\kappa r}x\left[\frac{1}{\kappa\eta}-e^{-\kappa\eta}E_1(\kappa\eta)\right], \\
0,1 & e^{-\kappa r}y\left[\frac{1}{\kappa\eta}-e^{-\kappa\eta}E_1(\kappa\eta)\right], \\
0,2 & e^{-\kappa r}\frac{x^2-y^2}{2}\left[\frac{1}{(\kappa\eta)^2}-\frac{1}{\kappa\eta}+e^{-\kappa\eta}E_1(\kappa\eta)\right], \\
0,2 & e^{-\kappa r}xy\left[\frac{1}{(\kappa\eta)^2}-\frac{1}{\kappa\eta}+e^{-\kappa\eta}E_1(\kappa\eta)\right].
\end{array}
$$

We are interested only in the values of tip wave functions outside the tip body, i.e., the region where $z\simeq -r$ or $\eta\simeq 2r$. Using the asymptotic expansion of the exponential integral functions,[28]

$$E_1(8)=e^{-u}\left[\sum_{k=0}^{n}\frac{(-1)^k k!}{u^{k+1}}+o\left[\frac{1}{u^{n+1}}\right]\right], \qquad \text{(D9)}$$

we find that the tip wave functions are reduced to

$$
\begin{array}{ll}
n,m & \chi, \\
0,0 & \frac{e^{-\kappa r}}{2r}, \\
0,1 & \frac{e^{-\kappa r}x}{4r^2}, \\
0,1 & \frac{e^{-\kappa r}y}{4r^2}, \\
0,2 & \frac{e^{-\kappa r}(x^2-y^2)}{8r^3}, \\
0,2 & \frac{e^{-\kappa r}xy}{4r^3}.
\end{array}
$$

As seen, they are similar to atomic wave functions.

[1]G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982); G. Binning and H. Rohrer, Helv, Phys. Acta **55**, 726 (1982).

[2]See, for example, Proceedings of the Third International Conference on Scanning Tunneling Microscopy [J. Microsc. **152** (1988)].

[3]V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, and R. J. Wilson, Phys. Rev. Lett. **59**, 2879 (1987); J. Wintterlin, J. Wiechers, H. Burne, T. Gritsch, H. Höfer, and R. J. Behm, *ibid.* **62**, 59 (1989); P. H. Lippel, R. J. Wilson, M. D. Miller, Ch. Wöll, and S. Chiang, *ibid.* **62**, 171 (1989).

[4]N. Garcia, C. Ocal, and F. Flores, Phys. Rev. Lett. **50**, 2002 (1983); N. Garcia and F. Flores, Physica B+C (Amsterdam) **127B**, 137 (1984).

[5]E. Stoll, A. Baratoff, A. Selloni, and P. Carnevali, J. Phys. C **17**, 3073 (1984).

[6]J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983); Phys. Rev. B **31**, 805 (1985).

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

[8]A. Baratoff, Physica B+C (Amsterdam) **127B**, 143 (1984).

[9]C. Noguera, J. Microsc. **152**, Part 1, 3 (1988); W. S. Sacks and C. Noguera, *ibid.* **152**, Part 1, 23 (1988); C. Noguera, J. Phys. (Paris) **50**, 2587 (1989).

[10]J. E. Demuth, U. Koehler, and R. J. Hamers, J. Microsc. **151**, 299 (1988).

[11]C. J. Chen, J. Vac. Sci. Technol. A **6**, 319 (1988); IBM Research Report RC 15 291, 1989 (unpublished).

[12]C. J. Chen, Phys. Rev. Lett. **65**, 448 (1990); J. Vac. Sci. Technol. A (to be published).

[13]D. A. Papaconstantopoulos, *Handbook of the Structure of Elemental Solids* (Plenum, New York, 1986).

[14]S. Ohnishi and M. Tsukuda, Solid State Commun. **71**, 391 (1989).

[15]N. D. Lang, Phys. Rev. Lett. **55**, 230 (1985); **56**, 1164 (1986); **58**, 45 (1987).

[16]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).

[17]L. W. Swanson and L. C. Crouser, Phys. Rev. Lett. **16**, 389 (1966); **19**, 1179 (1967).

[18]S.-L. Weng, E. W. Plummer, and T. Gustafsson, Phys. Rev. B **18**, 1718 (1978).

[19]M. Posternak, H. Krakauer, A. J. Freeman, and D. D. Koelling, Phys. Rev. B **21**, 5601 (1980).

[20]L. F. Mattheiss and D. R. Hamann, Phys. Rev. B **29**, 5372 (1984).

[21]J. Northrup, Phys. Rev. Lett. **57**, 154 (1986).

[22]G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1968), Vol. 16.5.

[23]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 511–515.

[24]J. Harris and A. Liebsch, Phys. Rev. Lett. **49**, 341 (1982).

[25]J. Tersoff, Phys. Rev. Lett. **57**, 440 (1986).

[26]D. Lawunmi and M. A. Payne, J. Phys. Condens, Matter **2**, 3811 (1990).

[27]The definition of parabolic coordinates is not consistent in the literature. Our notation is in agreement with the following sources: E. Schrödinger, Ann. Phys. (Leipzig) [Folge 4] **80**, 438 (1926); L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958), pp. 130–132; H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Plenum, New York, 1977), pp. 27–32.

[28]*Handbook of Mathematical Functions,* edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972), pp. 505–510.

[29]E. Wimmer, A. J. Freeman, M. Weinert, H. Krakauer, J. R. Hiskes, and A. M. Karo, Phys. Rev. Lett. **48**, 1129 (1982); E. Wimmer, A. J. Freeman, J. R. Hiskes, and A. M. Karo, Phys. Rev. B **28**, 3074 (1983); S. R. Chubb, E. Wimmer, A. J. Freeman, J. R. Hiskes, and A. M. Karo, *ibid.* **36**, 4112 (1987).

[30]V. I. Smirnov, *A Course of Higher Mathematics* (Pergamon, Oxford, 1964), Vol. IV, pp. 694–700.

[31]L. P. Eisenhert, Phys. Rev. **45**, 427 (1934).

[32]S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **41**, 2763 (1990).

[33]J. Tersoff, Phys. Rev. B **41**, 1235 (1990).

[34]C. V. Chen and F. Bozso (unpublished).