

# Theory of scanning tunneling spectroscopy

C. Julian Chen

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

(Received 23 July 1987; accepted 28 October 1987)

A new three-dimensional tunneling theory is introduced for interpreting scanning tunneling spectroscopy (STS) images. By expanding the asymptotic wave function of the acting atom in terms of complete sets of eigenfunctions in spherical coordinates and parabolic coordinates, a derivative rule is derived, which can be applied to interpret scanning tunneling microscopy and STS images immediately. The relation between the observed dynamic conductance and the density of states of the sample is shown in conjunction with a linear bias-distortion correction.

The invention of the scanning tunneling microscope (STM) has revolutionized surface science.<sup>1</sup> For the first time, individual atoms on a surface become directly perceptible. In addition to the topographical information provided by STM, by varying the bias voltage (up to a few volts) between the tip and the sample, a rich body of information can be obtained on the electronic states and energy spectra from the local dynamic conductance  $dI/dV$ .<sup>2-4</sup> This method, known as scanning tunneling spectroscopy (STS), is providing an increasingly large body of information on the details of electronic states on various surfaces.

The theory of STM and STS has been developed along with the advance of instrumentation. A recent review of the current status of the theory is given by Feuchtwang and Cutler.<sup>5</sup> In spite of a number of excellent works published so far,<sup>6-10</sup> they concluded that "a consistent and tractable theory of multidimensional tunneling theory is not yet available."<sup>5</sup> The most convenient formulation, as noted by the authors, is the perturbation theory of Bardeen.<sup>11</sup>

Bardeen's perturbation theory was developed for interpreting tunneling phenomena through a metal-insulator-metal junction.<sup>12</sup> The electrodes are described as free-electron-gas metals with planar surfaces. With a properly chosen unperturbed states of those electrodes, Bardeen proved that the tunneling matrix element can be evaluated by integrating a currentlike operator over a plane lying in the insulator slab. In one-particle form, it is

$$M_{\mu\nu} = -\frac{i\hbar^2}{2m_e} \int [\chi_\nu^* \nabla \psi_\mu - \psi_\mu \nabla \chi_\nu^*] \cdot d\mathbf{S}, \quad (1)$$

where  $\chi_\nu$  and  $\psi_\mu$  are unperturbed wave functions of the electrodes.

As is well known, the accuracy of any perturbation theory depends critically on the choice of the unperturbed wave functions, or the unperturbed Hamiltonians. One of the difficulties of three-dimensional tunneling theories is that the choice of unperturbed Hamiltonians is not unique. This is especially true for the problem of STS, in which the potential in the tunneling gap is not flat.

The first application of Bardeen's formula in interpreting STM images is made by Tersoff and Hamann,<sup>6</sup> which was successful in interpreting the topographic image of gold.<sup>6</sup> In their theory, the tip is modeled as a system of free electron gas confined in the tip body, which has a local radius  $R$  at the end; and the sample wave function is expanded in terms of

planar sinusoidal waves. To simplify calculation, further approximation was made to model the tip as a spherical potential well with radius  $R$ . Using Bardeen's formula, a very simple result was found that the tunneling matrix element is proportional to the magnitude of the sample wave function at the center of the sphere. Their theory was found to have a good agreement with observed image by assuming that the tip has a radius  $R \approx 9 \text{ \AA}$ , a cluster of about 200 atoms.

The atomic nature of STM is demonstrated in the theories of Lang,<sup>10</sup> where the actual wave function of a single atom adsorbed on a jellium surface, obtained from *ab initio* calculation, is used to represent both tip and sample. Using numerical integrations with Bardeen's formula, Lang showed that the operation of STM can be understood in terms of tunneling between a single atom on the tip of a single atom on the sample. Lang also proved that not only the density of states (DOS) of the sample, but also the DOS of the adsorbed atom of the tip, are necessary to be considered for understanding STM and STS.<sup>10</sup> However, the model used for Lang's theory may be too simple for handling complicated situations.

In this paper, a new three-dimensional tunneling theory of STS is presented, which is both atomic and analytic, with a built-in finite bias. The basic idea is similar to that of Bardeen.<sup>11</sup> By trying to make a rigorous treatment for the three-dimensional case with finite bias, we found that the unperturbed Hamiltonian of the sample and the unperturbed Hamiltonian of the tip have to be different in the gap region, as shown in Fig. 1. In other words, the unperturbed wave functions are solutions of the following Schrödinger's equations (Fig. 1):

$$[-(\hbar^2/2m_e)\nabla^2 + U_S - E_\mu]\psi_\mu = 0, \quad (2)$$

$$[-(\hbar^2/2m_e)\nabla^2 + U_T - E_\nu]\chi_\nu = 0. \quad (3)$$

The choice of unperturbed potentials is different from the original Bardeen theory.<sup>11</sup> Following Oppenheimer's method of error estimation,<sup>13</sup> it is found that our choice of unperturbed Hamiltonian minimizes the error introduced by neglecting the higher terms in the perturbation series. Another advantage of our choice is that the tip states are invariant as the bias changes, so to simplify calculations. Finally, it simplifies the estimation of bias distortion. Because in our theory, the bias only affects the sample wave function, thus can be treated with perturbation methods.

The tunneling is considered to be into or from the acting

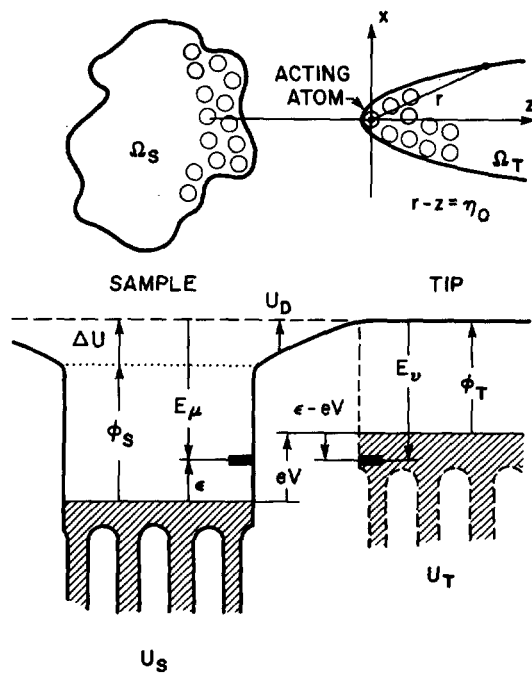


FIG. 1. Schematics.

atom on the tip. To calculate tunneling conductance, the wave function of the acting atom is expanded in terms of a complete set of eigenfunctions. Two kinds of expansions are found to be convenient and useful, namely, the expansions in spherical coordinates and in parabolic coordinates. Both methods were used at the very beginning of quantum mechanics, i.e., in the 1926 papers of Schrödinger on hydrogen atom.<sup>14</sup> The expansion in spherical coordinates is appropriate for describing an acting atom loosely bonded on the tip, whereas the expansion in parabolic coordinates is appropriate for describing an acting atom tightly bonded to the rest of the tip body. From a conceptual point of view, the expansion in parabolic coordinates gives a natural classification of the tip states for describing the images, whereas the expansion in spherical coordinates is convenient for identifying the atomic origin of the tip states. The description of the sample wave function is completely arbitrary. It works even in regions where a plane-wave expansion does not exist.

It seems that a considerable mess might be created by calculating the tunneling matrix elements from all components in the expansion of the acting-atom states. However, the final result found here is surprisingly simple. It can be summarized as a derivative rule, that is, the tunneling matrix element to or from an  $l$ th (or  $m$ th) component in the expansion is proportional to the  $l$ th (or  $m$ th) derivative of the sample wave function at the center of the acting atom.

Modifying Bardeen's theory to a more realistic geometry for STM/STS one can reformulate the tunneling theory in many curvilinear coordinate systems in which the Schrödinger equation in free space, i.e., the Helmholtz equation, is separable. In Fig. 1, it is exemplified in parabolic coordinates  $(\xi, \eta, \phi)$ , defined as  $x = \sqrt{\xi\eta} \cos \phi$ ,  $y = \sqrt{\xi\eta} \sin \phi$ , and  $z = (\xi - \eta)/2$ . The parabolic coordinate system has a cen-

ter (the center of the acting atom) and a direction (the  $z$  axis of scanning, which are different from  $x$  and  $y$ ), which provides a natural reference framework for describing tip states. By going through a procedure similar to those of Oppenheimer<sup>13</sup> and Bardeen,<sup>11</sup> we find the following expression of the tunneling matrix element:

$$M_{\mu\nu} = -\frac{i\hbar^2}{2m_e} \int_{\eta=\eta_0} \left[ \psi_\nu \frac{\partial \chi_\nu^*}{\partial \eta} - \chi_\nu^* \frac{\partial \psi_\mu}{\partial \eta} \right] \frac{h_\xi h_\phi}{h_\eta} d\xi d\phi, \quad (4)$$

where  $h_\xi, h_\phi$ , and  $h_\eta$  are scaling factors. Although the above expression is similar to Bardeen's,<sup>11</sup> it has some differences. First, the wave functions in that expression are the eigenfunctions of the unperturbed Hamiltonians of the tip and the sample, which are different in the gap region. Second, the expression is valid only on the paraboloid that is the boundary of the tip body, not in the entire barrier region.

In the following, we will show briefly how a very simple result, the derivative rule, can be obtained. As shown in Eq. (4), what is needed for calculating the tunneling matrix elements is the wave functions on the boundary of the tip. Consider first the tip wave function  $\chi_\nu$ . On and outside the tip boundary, the tip wave function satisfies the Schrödinger's equation in free space (see Fig. 1):

$$(\nabla^2 - \kappa^2)\chi_\nu = 0, \quad (5)$$

where  $\kappa = \sqrt{-2mE}/\hbar$  is the inverse decay length. Expanded in terms of eigenfunctions of the other two coordinates  $\xi$  and  $\phi$  (in the case of spherical coordinates, it is the spherical harmonics), Eq. (5) reduces to a set of ordinary differential equations with boundary condition to be regular at  $r \rightarrow \infty$ . Thus, the contribution of the tip wave function is determined only by its asymptotic values. The details of the tip wave functions near the center of the acting atom are not important. On the other hand, the sample wave function  $\psi_\mu$  satisfies the same differential equation, Eq. (5), on and inside the tip body. Expand the wave function in terms of the same eigenfunctions of  $\xi$  and  $\phi$ ; the differential equation of the sample wave function reduces to the same ordinary differential equation as the tip wave function. The boundary condition is to be regular at the center of the acting atom. Thus, it is determined only by the values of the sample wave function in the vicinity of the center of the acting atom. The details of the sample wave functions outside the tip body is not important. The integration in Eq. (4) is now the orthonormal condition of the eigenfunctions in  $\xi$  and  $\phi$ . What is left from the integrand is the Wronskian of the ordinary differential equation, which just cancels the factor  $h_\xi h_\phi h_\eta$ . As a result, we find a final expression for the tunneling matrix element as a sum of the products of the expansion coefficients. The coefficients for the expansion of the sample wave functions can be determined by its derivatives at the center of the acting atom. In terms of both spherical coordinates and parabolic coordinates, the relation between the coefficients and the derivatives at the center of the acting atom is very simple; a derivative rule is found.

In terms of spherical coordinates, the expansions are

$$\psi_\mu = \sum_{l,m} \alpha_{lm} i_l(\kappa r) Y_{lm}(\theta, \phi), \quad (6)$$

$$\chi_v = \sum_{l,m} \beta_{lm} k_l(\kappa r) Y_{lm}(\theta, \phi), \quad (7)$$

where  $i_l(x)$  and  $k_l(x)$  are spherical modified Bessel functions of first and second kind, respectively.<sup>15</sup> The tunneling matrix element is

$$M_{\mu\nu} = \frac{\hbar^2}{2m\kappa} \sum_{l,m} \alpha_{lm} \beta_{lm}^*. \quad (8)$$

Using the properties of modified spherical Bessel functions,<sup>15</sup> the relation between the coefficients and the derivatives is found to be

$$\alpha_{lm} = \frac{\sqrt{4\pi}(2l+1)!!}{\kappa^l \sqrt{(2l+1)(l+m)!(l-m)!}} \frac{\partial^l \psi_v(\mathbf{r}_0)}{\partial z^{l-|m|} \partial x^{|m|}}, \quad (9)$$

where  $x_{\pm} = (x \pm iy)/2$ , and  $\mathbf{r}_0$  is the center of the acting atom. The sign is determined by the sign of  $m$ .

In terms of parabolic coordinates, the expansions are

$$\psi_{\mu} = \sum \alpha_{nm\pm} e^{-\kappa(\xi+\eta)/2} (\xi\eta)^{m/2} \times M(n+m+1, m+1, \kappa\eta) L_n^m(\kappa\xi) e^{\pm im\phi}, \quad (10)$$

$$\chi_v = \sum \beta_{nm\pm} e^{-\kappa(\xi+\eta)/2} (\xi\eta)^{m/2} \times U(n+m+1, m+1, \kappa\eta) L_n^m(\kappa\xi) e^{\pm im\phi}, \quad (11)$$

where  $M(a, b, z)$  and  $U(a, b, z)$  are Kummer's functions, and  $L_n^m(z)$  is Laguerre's polynomial.<sup>16</sup> By expanding the sample and tip wave functions in terms of those functions, the tunneling matrix element is found to be

$$M_{\mu\nu} = -\frac{\pi\hbar^2}{2m\kappa} \sum_{n,m,\pm} \frac{(n+m)!}{n!} \alpha_{nm\pm} \beta_{nm\pm}^*. \quad (12)$$

The meaning of the functions in Eqs. (10) and (11) can be made clear by writing them in terms of Cartesian coordinates. For  $n=0$ , up to  $m=2$ , the sample wave function near the center of the acting atom can be written as<sup>17</sup>

$$\psi = e^{-\kappa z} [\alpha_{\sigma} + \alpha_{\pi x} x + \alpha_{\pi y} y + \alpha_{\delta(x^2-y^2)} (x^2 - y^2) + 2\alpha_{\delta xy} xy]. \quad (13)$$

The corresponding components of the tip wave functions have asymptotes in the gap region as follows<sup>17</sup>:

$$\begin{aligned} \chi_{\sigma} &\simeq e^{-\kappa r}/2r, \\ \chi_{\pi x} &\simeq e^{-\kappa r} x/4r^2, \\ \chi_{\pi y} &\simeq e^{-\kappa r} y/4r^2, \\ \chi_{\delta(x^2-y^2)} &\simeq e^{-\kappa r} (x^2 - y^2)/8r^3, \\ \chi_{\delta xy} &\simeq e^{-\kappa r} xy/4r^3, \end{aligned} \quad (14)$$

Apparently, those asymptotes of tip wave functions are identical to the atomic wave functions of spherical case with  $s$ ,  $p$ , and  $d$  angle dependences, which describe the extension of atomic wave functions beyond the tip head. However, when there is an intimate connection with the rest of the tip, the spherical symmetry of the atom is lost. The quantum number  $l$  is no more appropriate for describing tip states. Nevertheless, the center of the acting atom and the  $z$  axis of the scanning determine a well-defined axis. It provides a natural framework for classifying the tip states in terms of quantum

number  $m$ , i.e., states  $\sigma, \pi, \delta$ , etc., same as the states in a linear molecule.<sup>18,19</sup>

The derivative rule is now a direct consequence of Eqs. (9) and (13). A short list is as follows:

Tip state (parabolic)	Tip state (spherical)	Tunneling matrix element (up to a constant)
$\sigma$	$s, p_z, d_{zz}$	$\psi(\mathbf{r}_0)$
$\pi x$	$p_x, d_{xz}$	$[\partial\psi/\partial x](\mathbf{r}_0)$
$\pi y$	$p_y, d_{yz}$	$[\partial\psi/\partial y](\mathbf{r}_0)$
$\delta(x^2 - y^2)$	$d(x^2 - y^2)$	$[\partial^2\psi/\partial x^2 - \partial^2\psi/\partial y^2](\mathbf{r}_0)$
$\delta(xy)$	$d(xy)$	$[\partial^2\psi/\partial x\partial y](\mathbf{r}_0)$

In Fig. 2, some current images calculated from the derivative rule are shown. The notation of states are referred to the axis  $z$  and the center of the acting atom. The tip is moving in a plane  $z = Z_0$  with respect to the center of the surface atom. It is assumed that the ionization energy of that state is about 4 eV and the distance between the focus of the tip and the center of the sample is  $\sim 10$  Å. Thus, the size of the image is about  $\sqrt{Z_0/\kappa} \simeq 3$  Å.<sup>17</sup>

In spite of the seemingly unilateral look of the derivative rule, the reciprocity of the tip state and the sample state is preserved fairly accurately. The reciprocity principle, which is an implicit premise in Lang's theories,<sup>10</sup> is a basic microscopic symmetry of STM and STS. If the "acting" electronic state of the tip and the sample state under observation are interchanged, the image should be the same. An alternative wording of the same principle is, an image of microscopic scale may be interpreted either as by probing the sample state with a tip state or by probing the tip state with a sample state. Figure 2 shows that the result of calculation and the derivative rule is in accordance with the reciprocity principle.

Another interesting observation from Fig. 2 is that whenever the tip state and the sample state have the same symmetry and the same orientation, the current image has a single peak at the point the axes coincide, just as an  $s$ -wave sample state probed by an  $s$ -wave tip state. If the tip state and the sample state have different symmetry or mutually orthogonal orientation, at the center of the pattern (when the axes coincide), there is no tunneling current. This can be understood from group-theoretical considerations. Actually, the image belongs to an irreducible representation decomposed from a direct product of the irreducible representations to which the tip states and the sample states belong.<sup>18</sup>

The tunneling from the  $m \neq 0$  state was neglected by Tersoff and Hamann because those states "give a node towards the surface."<sup>6</sup> Their argument was correct for the case in their paper, i.e., a microscopically large spherical tip ( $R \simeq 9$  Å) and a metal surface with an almost flat image. Nevertheless, if the tip is considered as a single atom and the sample is semiconductor or semimetal, the  $m \neq 0$  terms may give a tunneling conductance comparable to the  $m = 0$  term. An example is Tersoff's model of Si(111) ( $2 \times 1$ ) and graphite surfaces,<sup>20</sup> where the surface wave function is represented by a sine wave,  $\Psi_k = \sin(k_f x) \exp(-\alpha_F z)$ , which has a node at  $x = 0$  (or  $x = \pm 2n\pi/k_f$ ). For a  $\sigma$ -tip state, the tunneling conductance at  $x = 0$  is zero. For a  $\pi x$ -tip state, at  $z = 0$ , the

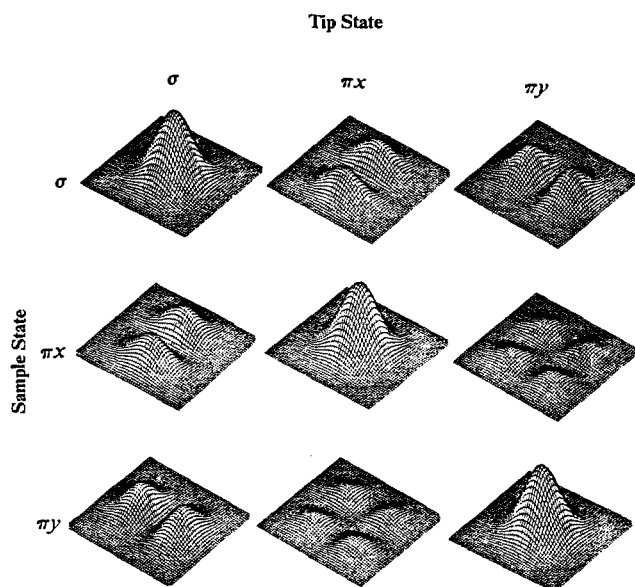


FIG. 2. Current images calculated from the derivative rule.

node of the tip state coincides with that of the sample wave function. Therefore, the tunneling conductance for a  $\pi x$ -tip state reaches maximum at  $x = 0$ . Since  $k_f$  and  $\alpha_f$  have the same order of magnitude, the tunneling conductance from  $\sigma$ -tip state and  $\pi x$ -tip state are of the same order of magnitude.

Now, the tunneling current can be calculated from the well-known formula<sup>21</sup>:

$$I(x, z) \propto \int_{-\infty}^{\infty} \rho_S(E + eV) \rho_T(E) dE |M(E)|^2 \times [f(E) - f(E + eV)], \quad (15)$$

where  $f(E) = \{1 + \exp[(E - E_F)/k_B T]\}^{-1}$  is the Fermi distribution function. In the low-bias limit, i.e.,  $eV \ll k_B T$ , Eq. (15) reduces to

$$I(x, z) \propto \int_{-\infty}^{\infty} \rho_S(E) \rho_T(E) dE |M(E)|^2 \times \{\cosh[(E - E_F)/2k_B T]\}^{-2}. \quad (16)$$

If  $eV \gg k_B T$ , the Fermi distribution function can be approximated by a step function. In this case, the bias distortion becomes important. By treating the bias distortion as a perturbation, we find finally that the derivative of the logarithm of current with respect to voltage is a good representation of the local density of states with only a constant correction,<sup>17</sup>

$$\frac{d \ln |I|}{dV} = -\frac{e\kappa L}{2\phi_S} + \frac{\rho_S(eV) |M(eV)|^2}{\left| \int_0^{eV} \rho_S(\epsilon) |M(\epsilon)|^2 d\epsilon \right|}, \quad (17)$$

where  $L$  is the average distance from the center of the acting atom to the center of the surface atoms of the sample, and  $\phi_S$  is the work function of the sample. Since the current images at lower bias are taken with a shorter gap distance and vice versa, the quantity suggested by Feenstra *et al.*<sup>4</sup> ( $d \ln I /$

$d \ln V$ ) is a good quantity to describe the spectra. However, if the relative distance  $L$  can be estimated from experimental conditions, the correction term, i.e., the first term in the expression, can be eliminated completely.

In conclusion, in this article, the new three-dimensional tunneling theory is introduced for interpreting scanning tunneling microscopy and spectroscopy (STM/STS) images. The theory is atomic in nature and has a built-in bias-distortion field. By expanding the wave function of the acting atom of the tip in spherical coordinates and parabolic coordinates, a derivative rule is proved, which is immediately applicable for interpreting STS images. The theory is shown to be in accordance with the basic symmetry of STM and STS, the reciprocity principle. Finally, the relation between the observed dynamic conductance and the density of states of the sample is shown in conjunction with bias-distortion correction.

**Acknowledgments:** It is my pleasure to acknowledge J. Demuth, J. Tersoff, R. Tromp, R. J. Hamers, B. Orr, N. D. Lang, R. Feenstra, R. Landauer, and K. K. Chin for simulating discussions, and P. Marcus for his extensive comments after carefully reading the first script of this paper.

<sup>1</sup>G. Binnig, H. Rohrer, Ch. Gerber, and E. Weiber, *Phys. Rev. Lett.* **49**, 57 (1982); **50**, 120 (1983).

<sup>2</sup>R. S. Becker, J. A. Golovchenko, D. R. Hamann, and B. S. Swartzentruber, *Phys. Rev. Lett.* **55**, 2032 (1985); G. Binnig and H. Rohrer, *IBM J. Res. Dev.* **30**, 355 (1986); W. J. Kaiser and R. C. Jaklavić, *ibid.* **30**, 411 (1986).

<sup>3</sup>R. J. Hamers, R. M. Tromp, and J. E. Demuth, *Phys. Rev. Lett.* **56**, 1972 (1986); E. J. van Loenen, J. E. Demuth, R. M. Tromp, and R. J. Hamers, *ibid.* **58**, 373 (1987). For a very readable review, see R. M. Tromp, R. J. Hamers, and J. E. Demuth, *Science* **234**, 304 (1986).

<sup>4</sup>J. A. Stroscio, R. M. Feenstra, and A. P. Fein, *Phys. Rev. Lett.* **57**, 2579 (1986); R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, *ibid.* **58**, 1192 (1987).

<sup>5</sup>T. E. Feuchtwang and P. H. Cutler, *Phys. Scr.* **35**, 132 (1987).

<sup>6</sup>J. Tersoff and D. R. Hamann, *Phys. Rev. Lett.* **50**, 1998 (1983); *Phys. Rev. B* **31**, 805 (1985).

<sup>7</sup>T. E. Feuchtwang, P. H. Cutler, and N. M. Miskovsky, *Phys. Lett. A* **99**, 167 (1983).

<sup>8</sup>N. García, C. Ocal, and F. Flores, *Phys. Rev. Lett.* **50**, 2002 (1983); N. García and F. Flores, *Physica* **127B**, 137 (1984).

<sup>9</sup>A. Baratoff, *Physica* **127B**, 143 (1984); E. Stoll, A. Baratoff, A. Selloni, and P. Carnevali, *J. Phys. C* **17**, 3073 (1984).

<sup>10</sup>N. D. Lang, *Phys. Rev. Lett.* **55**, 230 (1985); *Phys. Rev. B* **34**, 5947 (1986); *Phys. Rev. Lett.* **58**, 45 (1987).

<sup>11</sup>J. Bardeen, *Phys. Rev. Lett.* **6**, 57 (1961).

<sup>12</sup>I. Giaever, *Phys. Rev. Lett.* **5**, 147 (1960); **9**, 147 (1962).

<sup>13</sup>J. R. Oppenheimer, *Phys. Rev.* **13**, 66 (1928).

<sup>14</sup>E. Schrödinger, *Ann. Phys. (Paris)* **4-80**, 438 (1926).

<sup>15</sup>G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1968); pp. 415-416.

<sup>16</sup>M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), pp. 505-510.

<sup>17</sup>C. J. Chen (to be published).

<sup>18</sup>B. L. van der Waerden, *Group Theory and Quantum Mechanics* (Springer, Berlin, 1974).

<sup>19</sup>G. Herzberg, *Molecular Spectra and Molecular Structure* (Van Nostrand Reinhold, New York, 1966), Vols. II and III.

<sup>20</sup>J. Tersoff, *Phys. Rev. Lett.* **57**, 440 (1986).

<sup>21</sup>C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969), pp. 207-213.