RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 40, NUMBER 17 15 DECEMBER 1989-I

Method for the calculation of scanning tunneling microscope images and spectra

J. Tersoff

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

(Received 27 September 1989)

Most popular methods for calculating electronic structure cannot be readily used in quantitative calculations of scanning tunneling microscope (STM) images, because they are inappropriate for describing the wave-function tails far outside a surface. Moreover, the inclusion of a finite voltage between surface and tip, which is crucial in STM spectroscopy, is generally not feasible in calculations for real surfaces. Here, an approximate method is proposed which overcomes both of these problems. Any standard technique for calculating surface electronic structure, such as those using plane-wave or local-orbital bases, can be easily adapted to incorporate this approach.

In any microscopy or spectroscopy, it is crucial to be able to calculate the expected image or spectrum for a proposed model of the structure, in order to compare with experiment. For scanning tunneling microscopy[1] (STM), such comparisons have been hampered by the difficulty of calculating the tails of the wave functions far outside the surface.[2] For spectroscopy, the necessity of including the effect of the electric field on the wave function renders the problem intractible at present, except in simple models.[3,4]

Here, a method is proposed which solves both of these problems. The tails of the wave functions at arbitrary distances are obtained by a relatively accurate extrapolation procedure, which is easily added on to any desired electronic-structure method. Moreover, the effect of applied voltage can be included in the extrapolation, permitting the calculation of voltage-dependent images and tunneling spectra. It is our hope that the calculation of STM images and spectra will become routine in the future, because of the convenience of this approach.

Most electronic-structure methods use a fixed basis set, and transform Schrödinger's differential equation into an eigenvalue problem. This approach works well for most physical properties of interest, such as total energy. However, serious problems arise in treating the wave-function tails at distances which are relevant for STM, typically 5 Å or more.

A local-orbital basis typically does not have the correct asymptotic behavior outside a surface, and so can give qualitatively incorrect results at large distances. On the other hand, a plane-wave basis is in principle completely general; but in practice, any tractible number of plane waves can describe the exponential decay accurately only over very short distances. The linear augmented-plane-wave method[5] (LAPW) has been applied successfully to STM,[6] but is not in such wide use as other methods.

We first describe a method for circumventing these problems in the case of a free surface, permitting the calculation of STM images at low voltage. The extension to larger voltages, for application to STM spectroscopy, is given below.

Calculating STM images or spectra is equivalent to calculating the current $J(\mathbf{r}_t;V)$ as a function of tip position $\mathbf{r}_t$ and voltage V. In perturbation theory,

$$J=\frac{2\pi e}{\hbar}\sum_{\mu,\nu} f(E_\mu)[1-f(E_\nu)]\,|M_{\mu\nu}|^2\delta(E_\mu+V-E_\nu) \qquad (1)$$

at low temperature, where $f(E)$ is the Fermi function, V is the applied voltage (in units of energy), E_μ is the energy of the state ψ_μ, relative to the Fermi level of that electrode, and subscripts μ and ν refer to the respective electrodes.

The only difficult part is the calculation of the matrix element $M_{\mu\nu}$, which Bardeen showed[7] could be calculated directly from the wave functions,

$$M_{\mu\nu}=\frac{\hbar^2}{2m}\int_T d\mathbf{A}\cdot(\psi_\mu^*\nabla\psi_\nu-\psi_\nu\nabla\psi_\mu^*)\,, \qquad (2)$$

the integral being over a surface T in the barrier.[8]

For small voltages, and for the tip and surface well separated, it is a good approximation to use wave functions of the free tip and surface. Then, once these wave functions have been obtained, the calculation of the current from (1) and (2) is quite straightforward.[6]

The surface wave functions can be readily calculated in simplified models of the surface,[9-11] which have proven very useful in studying STM. However, to date, only the LAPW method[5] has been used successfully to quantitatively calculate STM images for real surfaces, and at realistic distances.

The LAPW method generates the basis functions outside the surface by integrating Schrödinger's equation, and so gives wave functions which are accurate even far outside the surface. Unfortunately, that method, while very accurate, is often considered difficult to implement and expensive to use, so it has not become widely available. The method here is inspired in part by the very successful treatment of the wave-function tails in the LAPW method.

To calculate the wave-function tails, we assume that the surface electronic structure has already been calculated, with a method which is accurate *except* for its inability to describe the wave functions far from the surface. In particular, the potential should be reasonably accurate everywhere, with errors being small relative to the work function.[12]

We then divide space into two regions, A and B, as shown in Fig. 1. The wave functions already calculated are accurate throughout region A, including the surface S dividing regions A and B.

Now consider the wave functions in a Laue representation,

$$\psi_\nu = \sum_Q a_{\nu Q} \exp(i\mathbf{Q}\cdot\mathbf{x}) u_{\nu Q}(z)\,, \tag{3}$$

where $\mathbf{Q} = \mathbf{k}_\parallel + \mathbf{G}$, $\mathbf{k}_\parallel$ is the two-dimensional Bloch wave vector, and the sum is thus over reciprocal-lattice vectors $\mathbf{G}$. Substituting (3) into Schrödinger's equation gives the following equation for the functions $u_{\nu Q}(z)$:

$$\sum_{Q'} a_{\nu Q'} V_{Q-Q'}(z) u_{\nu Q'} + a_{\nu Q}\left[-E_\nu + \frac{\hbar^2}{2m}\left(Q^2 - \frac{d^2}{dz^2}\right)\right] u_{\nu Q} = 0\,, \tag{4}$$

where $V_Q(z)$ is the two-dimensional Fourier transform of the potential, and the argument z of $u_{\nu Q}(z)$ has been suppressed for brevity.

Choosing u to be normalized to unity on the plane S, the coefficients $a_{\nu Q}$ are given by

$$a_{\nu Q} = \int_S \psi_\nu(\mathbf{x},z) \exp(-i\mathbf{Q}\cdot\mathbf{x})\, d\mathbf{x}\,, \tag{5}$$

using the ψ_ν which have already been calculated. The integral over the two-dimensional unit cell is easily carried out with the method of Monkhorst and Pack.[13]

If necessary, the $u_{\nu Q}$ could be determined by solving the coupled one-dimensional equations (4) in region B, subject to the boundary condition of vanishing at large z and being unity on S. Note that $u_{\nu Q}$ automatically has the same logarithmic derivative as the true wave function across S, so there is no need to match the slope explicitly.

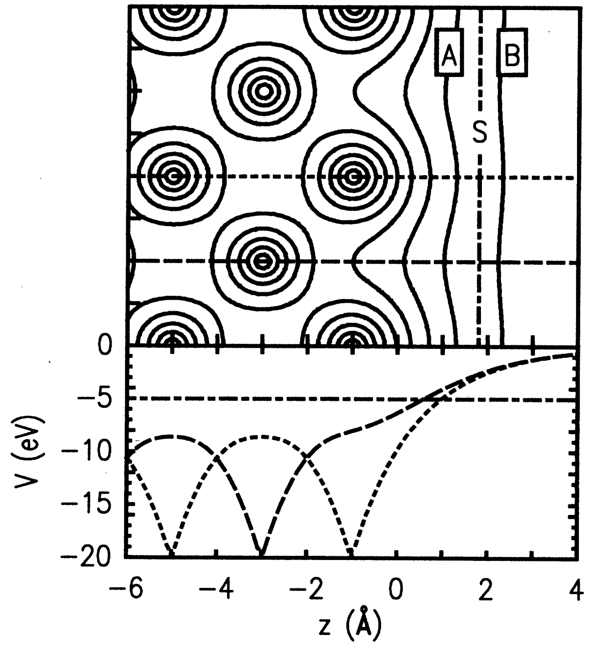

FIG. 1. Schematic illustration of potential V at a surface (based on a simple model, not a real calculation). A (100) plane through an (001) surface is shown. The upper panel shows surface S (dashed-dotted line) dividing regions A and B; solid lines are contours of constant potential. Lower panel shows $V(z)$ along two lines perpendicular to surface; dotted and dashed curves correspond, respectively, to paths shown as dotted and dashed in upper panel, with dashed-dotted line indicating Fermi energy, and zero being the vacuum level.

The important simplification is that in region B, if $V_Q(z)$ is negligible for $Q \neq 0$, the equations decouple, giving

$$\left[V_0(z) - E_\nu + \frac{\hbar^2}{2m}\left(Q^2 - \frac{d^2}{dz^2}\right)\right] u_{\nu Q} \simeq 0\,, \tag{6}$$

which can be immediately integrated to give $u_{\nu Q}$. Note that in (6), $u_{\nu Q}$ depends on ν only through the energy E_ν, so it suffices to calculate u_Q for several values of energy in the relevant range.

Thus there are two levels of approximation which can be used here. The simplest, and the one which we anticipate will be most useful in practice, is (6), which renders the problem almost trivial. After the usual surface calculation, it is merely necessary in addition to calculate the coefficients (5), and to tabulate the new energy-dependent basis functions (6).

However, for highly corrugated surfaces, it might be impractical to take S sufficiently far from the surface to ensure that $V_Q \simeq 0$ for $\mathbf{Q} \neq 0$; the original calculation might not be accurate out to such distances. In that case, one could work directly with (4). Note that only the lowest few $\mathbf{Q}$ need be retained, however; even if the corrugation is substantial in region B, higher Fourier components are virtually absent. This fact immensely simplifies the calculation.

Alternatively, for a highly corrugated surface, one could match the wave functions on a nonplanar surface S, which follows the corrugation of the surface. In that way it becomes feasible to position S so that $V_Q \simeq 0$ for $Q \neq 0$ in region B, and so (6) can still be used. Taking a set of "special points" $\mathbf{x}$ within the unit cell,[13] and a corresponding finite set of $\mathbf{G}$, one can write

$$\psi_\nu(\mathbf{x}, z_S[\mathbf{x}]) = \exp(i\mathbf{k}_\parallel\cdot\mathbf{x}) \sum_G C_{xG} a_{\nu G}\,,$$

where from (3),

$$C_{xG} = \exp(i\mathbf{G}\cdot\mathbf{x}) u_{\nu G}(z_S[\mathbf{x}])\,.$$

Here, C is implicitly energy dependent, and $z_S[\mathbf{x}]$ is the z coordinate of the surface S at $\mathbf{x}$. The determination of $a_{\nu G}$ then merely requires inversion of the matrix C_{xG}.

Again, as in the alternative approach of solving (4), the smoothness of the functions in region B is essential in rendering the calculation quite modest. Because the surface S follows the actual surface topography in this implementation, the wave functions are quite smooth on S. Therefore quite a coarse grid of $\mathbf{x}$, and a correspondingly small set of $\mathbf{G}$, can be used, so the matrix C is small.

We now consider the problem of tunneling spectroscopy. Here, inclusion of the applied voltage is crucial. However, self-consistent calculations for real surfaces in the presence of applied fields are not generally feasible.

An approximate method for calculating tunneling current at finite voltage, using free-surface wave functions, has been described by Lang,[3] but is accurate only for small voltages. At the other extreme, Lang *et al.* have implemented an approach which is essentially exact at any voltage,[4] but only for a model pair of electrodes.

Here, we use the usual trick of decomposing the actual system, with applied voltage, into two fictitious noninteracting surfaces. Each surface has an artificial potential, which reproduces the real potential up to the plane T used in (2), while satisfying the conditions for the applicability of the Bardeen approach.

The decomposition of the potential is illustrated in Fig. 2. The full potential, Fig. 2(b), is decomposed into two free surfaces, Fig. 2(c). Each of these is identical to the full potential up to the plane T. The current can be calculated by applying (2) to the wave functions of the fictitious free surface at the plane T.

Unfortunately, it is not practical to calculate the potential of the full system self-consistently, including applied voltage, except in simple models.[4] Therefore we assume that the electric field is fully screened in region A, and completely unscreened in region B. In that case, the potential in region A is simply the original self-consistent potential for the unperturbed free surface; in region B, the original potential has added to it a term linear in distance z, going from zero to $V+\Delta\phi$ in the region between the planes S of the respective surfaces. (Here V is the voltage, and $\Delta\phi$ is the work-function difference between surface and tip.) Figure 2(b) illustrates how the zero-field potential of Fig. 2(a) is modified, in this approximation, by the electric field.

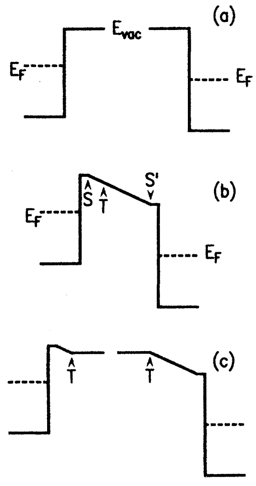

FIG. 2. (a) Two neutral surfaces give a zero-field (not zero-voltage) junction; $E_{\rm vac}$ is the vacuum level. (b) This initial potential is assumed to be modified simply by a linear variation between the planes S and S'. (c) The potential is then divided into two effective free-surface potentials, each of which equals the original potential only up to the plane T.

This assumption should be particularly accurate when the tunneling gap is wide, since then the voltage drop occurs in the low-density region, and the precise position of S relative to the surface atoms is unimportant. But in any case, this approach is much more accurate than the best approximation suggested to date, that of Ref. 3. In particular, the approach is not confined to small voltage. However, at high voltages the plane T should be taken nearer to the negative electrode, to ensure that the corresponding fictitious free surface still has all the relevant states well confined.

Having specified the potential for our fictitious surface, it remains only to calculate (6) with the appropriate $V_0(z)$. This is the only calculation which must be repeated for different voltages; the original calculation of electronic structure, and (5), need be done only once.

Implicitly, we have introduced an approximation here, by neglecting the effect of the field in region B on the wave functions in region A. This means that, strictly speaking, the logarithmic derivatives no longer automatically match across S. However, since the change in the potential is significant only in the region where the wave function has little amplitude, this should be reasonably accurate.

Finally, a word about the treatment of the STM tip is in order. So far we have implicitly assumed that the tip could be treated in the same manner as the surface, and have made no distinction. For spectroscopy, the tip could indeed be modeled as a planar metal surface. However, for calculating STM images, one is interested in a sharp tip.

At low voltages, the simplest option is to use the s-wave tip model of Tersoff and Hamann.[6] In that case, the image can be inferred directly from the calculated electronic structure of the bare surface, rendering the problem particularly simple. This approximation is rather accurate in most cases, and the conditions for its applicability have recently been discussed.[14]

There is no problem in extending the s-wave tip model to higher voltages, if one is willing to treat the electric field as still planar. In particular, one can simply take the plane T right at the tip, again avoiding the need to do any actual calculation other than for the bare surface.

In fact, neglecting the lateral variation in the field may not be a bad approximation. While the exact structure of the tip is unknown, a reasonable model is a single atom uppermost on a surface with a radius of curvature of $R \sim 500$ Å. Because of the exponential nature of tunneling, the single atom dominates in the current. But for the long-ranged electrostatics, it is the larger radius of curvature which is most relevant. Since the tunneling gap is much smaller than R, the electric field is not so different from a planar junction in the region of interest.

To summarize, the STM image or spectrum is readily calculated from (1) and (2), if the wave-function tails outside the surface are accurately known. These tails can be calculated most conveniently by solving (6), and matching the solutions $u_{\nu Q}$ to the solutions obtained with any convenient standard method.

Alternatively, for particularly corrugated surfaces, the method can be made more accurate either by performing the matching on a nonplanar surface which follows the surface topography, or by solving the coupled equations (4). In either case, only a few Fourier components need be retained.

The same approach can be used to model the case of an applied voltage, by adding to the free-surface potential a linear "ramp" in the vacuum region, as illustrated in Fig. 2. Since only the simple calculation (6) need be repeated for different voltages, the numerical calculation of tunneling spectra represents no problem.

The additional calculations required are minor compared with the unavoidable initial calculation of the surface electronic structure. Because the method is as simple numerically as it is conceptually, it should make feasible the routine use of theoretical calculations in the interpretation of STM data.

---

[1] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982). For reviews of STM, see G. Binnig and H. Rohrer, Rev. Mod. Phys. **59**, 615 (1987); P. K. Hansma and J. Tersoff, J. Appl. Phys. **61**, R1 (1987).

[2] J. Tersoff, in *Basic Concepts and Applications of Scanning Tunneling Microscopy and Related Techniques,* edited by editors, NATO Advanced Study Institute, Series B Physics (Academic, New York, in press).

[3] N. D. Lang, Phys. Rev. B **34**, 5947 (1986).

[4] N. D. Lang, A. Yacoby, and Y. Imry, Phys. Rev. Lett. **63**, 1499 (1989).

[5] O. Jepsen, J. Madsen, and O. K. Andersen, Phys. Rev. B **18**, 605 (1978); H. Krakauer, M. Posternak, and A. J. Freeman, *ibid.* **19**, 1706 (1979); D. R. Hamann, Phys. Rev. Lett. **46**, 1227 (1981).

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

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

[8] The exchange-correlation potential is sufficiently long-ranged that for reasonably small sample-tip separations, the conditions for the formal validity of the Bardeen approach are violated sufficiently to give modest errors in the tunneling current; see M. C. Payne, J. Phys. C **19**, 781 (1986). However, as discussed in Ref. 2, these errors have little effect on the STM image.

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

[10] W. Sacks, S. Gauthier, S. Rousset, J. Klein, and M. A. Esrick, Phys. Rev. B **36**, 961 (1987).

[11] N. Garcia, C. Ocal, and F. Flores, Phys. Rev. Lett. **50**, 2002 (1983); E. Stoll, A. Baratoff, A. Selloni, and P. Carnevali, J. Phys. C **17**, 3073 (1984).

[12] The calculated potential presumably includes the exchange-correlation potential within the local-density approximation (LDA). While LDA does not give the correct asymptotic form for this potential, the errors are apparently nowhere large, see Ref. 4 above. Therefore, especially in view of the relative insensitivity to errors discussed in Ref. 2, the image potential causes no problem, except that discussed in Ref. 8 above.

[13] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[14] J. Tersoff (unpublished).