Simulation of current in the scanning tunneling microscope

Th. Laloyaux, I. Derycke, J.-P. Vigneron, Ph. Lambin, and A. A. Lucas
Institute for Studies in Interface Sciences, Facultés Universitaires Notre-Dame de la Paix, rue de Bruxelles, 61, B-5000 Namur, Belgium
(Received 15 July 1992; revised manuscript received 1 December 1992)

Considering simple models of the scanning tunneling microscope and metallic samples, we use a finite-element method to solve Schrödinger's equation for the electrons tunneling from the tip to the sample. We plot current-density maps for various geometries of the electrodes: hemispherical or cylindrical tip facing a planar surface or a surface with a Gaussian boss or dip. It can be seen on the current-density maps that the electron flow passes preferentially through the thinnest region of the barrier. From the current density in the case of a planar sample, we investigate the width of the tunnel current beam when it penetrates into the sample. From the dependence of the current on the distance between a hemispherical tip and a Gaussian boss or dip, we show that the corrugation of the sample surface is attenuated by a factor of two in the constant-current image. The effective work function, determined from the logarithmic derivative of the current with respect to the distance, differs from the real work function of the sample and, as an effect of the image potential, decreases when the tip approaches the sample. A comparison between a numerical resolution of the exact Schrödinger equation and the transfer Hamiltonian approximation shows that the latter gives good results, even when the tip is close to the sample.

I. INTRODUCTION

Since its invention by Binnig and Rohrer,[1-3] the scanning tunneling microscope (STM) has attracted a vast amount of theoretical work (Refs. 4–25 and many other publications). Understanding the detailed tunneling mechanism that is taking place during an STM scan is still an important challenge, although it is fair to say that the description of three-dimensional elastic tunneling is experiencing today much practical progress. From a mathematical point of view, the basic difficulty to accurately account for the electron scattering by the tip-sample potential barrier lies in the fact that three-dimensional position coordinates cannot be exactly separated. This is encountered in all realistic STM models because the electrons cross the vacuum gap through a very small protrusion, sometimes reduced to a single atom, located at the tip apex: the microscope resolution is precisely related to the concentration of electron presence close to such an asperity. The radius of curvature of the tip is in practical cases comparable to the electron wavelength in the metal tip or in the conducting sample. It also compares with the electron penetration length in the vacuum. For this reason, a planar model of the barrier is a crude approximation of the real geometry and is relevant only in the spirit of a phenomenological description of the tunnel current.

The present work focuses essentially on the three-dimensional (3D) tunneling problem, putting less emphasis on the usually underlined relation between the STM image and the electronic structure of the sample and the tip. We will be more interested in the description of the electron spreading under a given, assumed—though realistic—potential than on the precise physical origin of this potential. Therefore, in the various models we describe in Sec. II, both the tip and the sample will be seen as simple free-electron metals and the potential in the vacuum will simply contain information about the macroscopic work function and the applied bias. The only source of sophistication we shall allow ourselves for realism is introduction of a classical static image potential. The geometry will reflect the arduous mixing of microscopic and macroscopic elements found in actual STM tip-sample devices, but will be kept simple and schematic enough so that the various simulations we report here can be convincingly described as highly accurate and reliable. In this spirit, we study the impact of the tip shape and location on an STM image and discuss the resolution in terms of the electron flow between the tip and the sample.

Solving the tunneling problem in a multidimensional configuration of the barrier cannot be done analytically. To overcome that difficulty, most authors[4,9,10,14,17,18] used the transfer Hamiltonian approximation (THA) to calculate the current. Our choice is to solve the exact Schrödinger equation using a numerical method. We use a finite-element technique, described in Sec. III.

Our computation yields current-density maps for different models of tips and samples. From these maps, and from the total current, we extract information on the resolution of the STM, on the work-function measurements done with an STM, and on the validity of the THA. These results are presented in Sec. IV.

II. METHODOLOGY

Following Lucas *et al.*,[15,16,26] we consider models where both the tip and the sample are free-electron metals. The potential energy of the electrons is constant inside the materials (14 eV below the vacuum level) and the work function is 4.5 eV. These two numerical values, assumed to be the same for both the tip and the sample, are

typical of aluminum. In the vacuum between the metals, the electrons experience the bias potential applied between the electrodes and the classical image potential due to the response of the conduction electron of the electrodes. Considering the shape of the tip and the sample, we have three types of models (see Fig. 1).

(1) The same model as in Refs. 15, 16, and 26, where the sample is planar and the tip is a hemisphere protruding from a planar electrode. The hemisphere is intended to represent an atom or a cluster protruding from the apex of the tip. The tip radius is varied from 2 to 6 Å and the distance between the tip and the sample from 1 to 5 Å.

(2) A model similar to the previous one, but with the hemisphere replaced by a cylindrical protrusion. This geometry models a close-packed tip terminated by a plateau rather than a single atom. The height of the cylinder is 3 Å; its radius varies from 3 to 6 Å and the distance between the tip and the sample ranges from 1 to 5 Å.

(3) A hemispherical tip (3-Å radius, cf. the first model) facing a sample with a Gaussian boss or a Gaussian dip coaxial with the tip. With this model, we study the influence on the images of the sample relief. The height of the boss (or the depth of the dip) is 3 Å and its standard deviation is 2 Å.

All these models present an axial symmetry by construction. Thus we can work in practice with two dimensions by using the cylindrical coordinates [see Fig. 1(a)]: ρ (the distance from the symmetry axis) and z (the distance from the planar surface of the tip, the planar surface of the sample being at $z=D$). The dependence of the wave function on φ (the angle of rotation around the symmetry axis) can be factored, thanks to the axial symmetry.

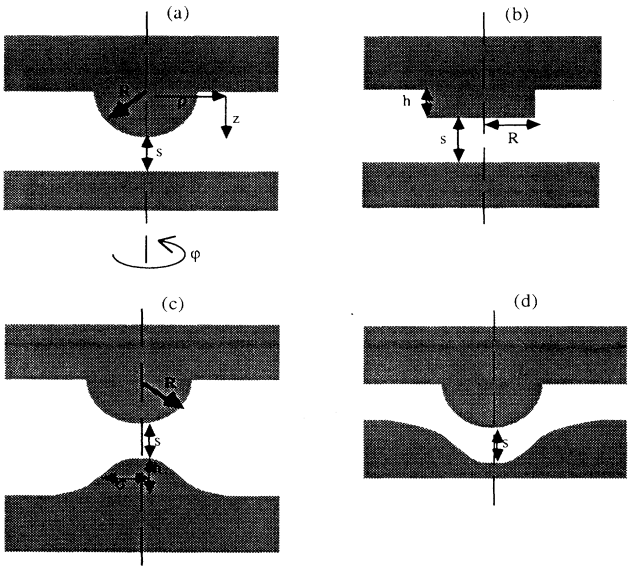

FIG. 1. Models of STM tip and sample. Both the tip and the sample are free-electron metals. The tip is either a hemisphere [(a), (c), and (d)] with a radius R_{tip} ranging from 2 to 6 Å or a cylinder (b) with a height of 3 Å and a radius R_{tip} ranging from 3 to 6 Å, protruding from a planar surface. The sample is either planar [(a) and (b)] or has a Gaussian protrusion (c) or trough (d). The height of the protrusion or the depth of the trough is 3 Å and its standard deviation is 2 Å. The distance s between the tip and the sample varies from 1 to 5 Å.

The models just described are very simple and this simplicity allows an accurate numerical calculation of the tunneling current. Although flat surfaces are irrealistic, they are rather good models for metallic surfaces where atoms can hardly be resolved with an STM, e.g., close-packed surfaces. Our approach does not lead to a simulation of images, but we will use it to investigate other aspects of scanning tunneling microscopy, such as the current beam width, the dependence of the current on the distance, and the validity of the THA.

The wave functions Ψ of the tunneling electrons (and hence the current density across the vacuum between the tip and the sample) are obtained by solving Schrödinger's equation for all energies E at which electron contribute to the current (the Rydberg atomic units are used throughout this paper):

$$-\frac{\partial^2\Psi}{\partial\rho^2}-\frac{1}{\rho}\frac{\partial\Psi}{\partial\rho}-\frac{\partial^2\Psi}{\partial z^2}-\frac{1}{\rho^2}\frac{\partial^2\Psi}{\partial\varphi^2}+V(\rho,z)\Psi(\rho,z,\varphi)=E\Psi(\rho,z,\varphi)\ . \quad (1)$$

We assume the interaction between the electron and the barrier is elastic. Because the potential energy V does not depend on the angle φ, the dependence on this angle can be separated in (1).

The separable solutions have the form

$$\Psi(\rho,z,\varphi)=e^{im\varphi}\psi_m(\rho,z) \quad (2)$$

with m an integer number. The 2D wave function ψ_m obeys the equation

$$-\frac{\partial^2\psi_m}{\partial\rho^2}-\frac{1}{\rho}\frac{\partial\psi_m}{\partial\rho}-\frac{\partial^2\psi_m}{\partial z^2}+\frac{m^2}{\rho^2}\psi_m(\rho,z)+V(\rho,z)\psi_m(\rho,z)=E\psi_m(\rho,z)\ . \quad (3)$$

The quantum number m is the angular momentum canonically conjugated to the angle φ. It is a good quantum number in our model. An electron with an angular momentum $m\neq 0$ rotates around the symmetry axis. Since for each state there is another state with an opposite angular momentum $-m$ having the same energy, there is no net φ component in the total current. Electronic states with opposite angular momenta m and $-m$ obey the same equation (3) and thus have the same solutions $\psi_m(\rho,z)$.

Solving Eq. (3) demands a model of the potential energy V. Inside the metals, V is a constant: V_{in}^1 inside the tip (the chosen value is -14 eV) and V_{in}^2 inside the sample (V_{in}^1 minus the bias applied on the sample). In the gap between the electrodes, V is the sum of two terms: the applied bias and the image potential energy:

$$V=V_b+V_{\text{im}}\ . \quad (4)$$

The bias V_b is the solution of the Laplace equation

$$\nabla^2 V_b=0 \quad (5)$$

with the boundary conditions:

$$V_b = \begin{cases} 0 \text{ (on the tip)}, & (6) \\ -V_{\text{bias}} \text{ (on the sample)}, & (7) \end{cases}$$

where V_{bias} is the voltage applied on the sample. The image potential energy of an electron located at $\mathbf{r}_e$ (between the electrodes) is given by

$$V_{\text{im}}(\mathbf{r}_e) = \tfrac{1}{2} V_{\text{pol}}(\mathbf{r}_e) , \quad (8)$$

where e is the electron charge and V_{pol} is the potential brought about by the response of the electrons in the tip and the sample to the presence of the electron at $\mathbf{r}_e$. V_{pol} is the solution of Laplace's equation.

$$\nabla^2 V_{\text{pol}} = 0 \quad (9)$$

with the boundary condition

$$V_{\text{pol}}(\mathbf{r}) = -\frac{1}{|\mathbf{r}-\mathbf{r}_e|} \text{ (on the surfaces)} . \quad (10)$$

In the case of the plano-hemispherical model, V_b and V_{im} have been computed by Lucas *et al.*[26] in an analytical way. For our other models, we have to resort to a numerical resolution. We use the well-known finite-difference method.[27] The potential barriers in some of our models are plotted in Fig. 2. Notice the effect of the image potential on these barriers: the potential is rounded and lowered, especially where the barrier is the thinnest. It is expected that the tunneling current will flow preferentially through the region of the barrier.

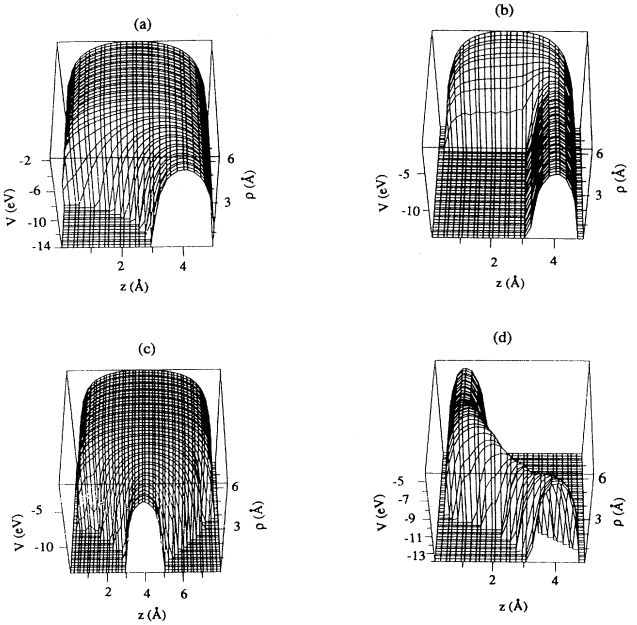

FIG. 2. Potential barriers in the models of Fig. 1. The potential inside the metals is taken constant and equal to 14 eV below the vacuum level. The divergences of the classical image potential at the interfaces have been truncated at this value. The work function of the tip and the sample is 4.5 eV. The radius of the hemispherical tip [(a), (c), and (d)] is 3 Å, the radius of the cylindrical tip (b) is 6 Å, and its height is 3 Å, the distance s measured along the symmetry axis between the apex of the tip and the sample surface is 2 Å in all four cases.

Inside the metals, the potential energy is constant and we can separate the variables ρ and z in Eq. (3). The separable solutions found in this way in the tip ($z \leq 0$) are expressed as

$$\psi_{m,k_\parallel}(\rho,z) = J_m(k_\parallel \rho)\exp\{\pm i[E - V_{\text{in}}^1 - (k_\parallel)^2]^{1/2} z\} , \quad (11)$$

whereas in the sample ($z \geq D$), they take the form

$$\psi_{m,k_\parallel}(\rho,z) = J_m(k_\parallel \rho)\exp\{\pm i[E - V_{\text{in}}^2 - (k_\parallel)^2]^{1/2}\} . \quad (12)$$

The J_m are cylindrical Bessel functions of the first type and $k_\parallel$ is the modulus of the component of the wave vector of the electron perpendicular to z. In the exponent of Eq. (11), the plus sign corresponds to electrons coming from the interior of the tip and moving towards the barrier and the minus sign corresponds to electrons reflected back into the tip. In the exponent of Eq. (12), the plus sign corresponds to electrons leaving the barrier to enter into the sample and the minus sign corresponds to electrons of the sample moving towards the barrier. The wave functions described by Eq. (2) with ψ_m given by (11) [respectively, (12)] form a basis for the free-electron wave functions with energy E in the tip (respectively, in the sample).

Without loss of generality, we can assume that the sample is biased positively. Thus, the electrons tunnel through the barrier from the tip to the sample. The wave function of an electron with energy E and angular momentum m impinging on the barrier with a component of the momentum parallel to the surface equal to $k_\parallel^{\text{inc}}$ is, in the region $z \leq 0$, the sum of an incoming and backscattered waves:

$$\psi_{m,k_\parallel^{\text{inc}}}(\rho,z) = J_m(k_\parallel^{\text{inc}}\rho)\exp\{i[E - V_{\text{in}}^1 - (k_\parallel^{\text{inc}})^2]^{1/2} z\} + \sum_{k_\parallel^{\text{ref}}} B_{k_\parallel^{\text{ref}}} J_m(k_\parallel^{\text{ref}}\rho) \times \exp\{-i[E - V_{\text{in}}^1 - (k_\parallel^{\text{ref}})^2]^{1/2}\} . \quad (13)$$

The tunneling gives rise to transmitted waves. Thus, in the region $z \geq D$, the wave function is

$$\psi_{m,k_\parallel^{\text{inc}}}(\rho,z) = \sum_{k_\parallel^{\text{tr}}} C_{k_\parallel^{\text{tr}}} J_m(k_\parallel^{\text{tr}}\rho) \times \exp\{i[E - V_{\text{in}}^2 - (k_\parallel^{\text{tr}})^2]^{1/2} z\} . \quad (14)$$

The $B_{k_\parallel^{\text{ref}}}$ and the $C_{k_\parallel^{\text{tr}}}$ are still unknown at this point, and so are the values of the wave functions in the region $0 < z < D$, where the potential does not vanish. The resolution of Schrödinger's equation in the region of the barrier, together with the matching conditions of the wave function and its z derivative at $z = 0$ and $z = D$, will yield all these unknown values.

Note that in Eqs. (13) and (14) we have to sum over $k_\parallel$ and not over m. This is because, thanks to the rotational

symmetry of the geometries of the models, m is a good quantum number while $k_\parallel$ is not, due to the presence of the protrusion which breaks the translational symmetry parallel to the surface and scatters the electrons to values of $k_\parallel$ different from the initial value $k_\parallel^{\rm inc}$. The clarity of Eqs. (13) and (14) demands that we reveal which are the allowed values of $k_\parallel^{\rm inc}$ and on which values of $k_\parallel^{\rm ref}$ and $k_\parallel^{\rm tr}$ we make the summations. Equation (13) is subject to the condition that the wave function must be finite at $-\infty$. For the first term (the incident wave), the condition imposes the radicand to be positive, thus

$$k_\parallel^{\rm inc} \le (E - V_{\rm in}^1)^{1/2} . \tag{15}$$

The second term is regular at $-\infty$ whichever value of $k_\parallel^{\rm ref}$ is considered. The terms with $k_\parallel^{\rm ref}$ smaller than $(E - V_{\rm in}^1)^{1/2}$ correspond to reflected oscillating waves, the other terms to evanescent waves decaying towards the bulk. The latter would be unphysical in a bulk material, but they must be included in this problem where interfaces are present. Applying the same reasoning to Eq. (14) leads to the conclusion that all values of $k_\parallel^{\rm tr}$ are allowed, those smaller than $(E - V_{\rm in}^2)^{1/2}$ corresponding to transmitted oscillating waves and the other ones to waves decaying towards the bulk of the sample. A continuum $(0, +\infty)$ of values of $k_\parallel$ actually exist. Nevertheless, in order to keep finite the amount of numerical calculations to be performed (see Sec. III), we confine the calculations within a cylinder $\rho < R$, the wave functions being identically zero outside the cylinder. In our calculations, we took $R = 8$ Å. For the mathematical problem to be well stated, we need a boundary condition on the lateral surface of the cylinder ($\rho = R$):

$$\Psi(R, z, \varphi) = 0 . \tag{16}$$

The boundary condition (16) does not model an aspect of the real physical system. Its physical meaning is that the electrons are confined inside the cylinder, with a hard wall on the surface $\rho = R$. This artificial confinement is not expected to perturb the solutions, provided R is chosen significantly larger than the lateral extend of the tunneling current. We tested this assumption successfully by increasing R: it appeared that no significant change in the tunneling current density occurred.

In the case of separable solutions, Eq. (16) is equivalent to

$$J_m(k_\parallel R) = 0 , \tag{17}$$

which induces a quantization of $k_\parallel$. We also need a consistent choice for the normalization of our solutions:

$$\int_0^R [J_m(k_\parallel \rho)]^2 \rho \, d\rho = 1 . \tag{18}$$

If we choose a larger value for R, the normalization of the Bessel functions will decrease, but the allowed values of $k_\parallel$ [cf. Eq. (17)] will get closer to each other. As a consequence, there will be no effect on the total current for a large enough R.

The quantization (17) also limits the number of values of m to be considered. The smallest value of $k_\parallel$ allowed by (17) increases when $|m|$ is increased. For a large enough $|m|$, this smallest value becomes larger than $(E - V_{\rm in}^1)^{1/2}$ and there are no more allowed incident waves. In our models, the largest allowed value of m was 8.

Every electron gives rise to a current density $\mathbf{j}$ that can be deduced from its wave function Ψ:

$$\mathbf{j} = 2 \, {\rm Im}(\Psi^* \nabla \Psi) . \tag{19}$$

The incident wave in the tip is characterized, in cylindrical coordinates, by the angular momentum m, the modulus $k_\parallel$ of the component of the momentum parallel to the surface and the component k_z perpendicular to the surface. The latter is positive for an incident wave. The energy E can be determined from $k_\parallel$ and k_z. The total current density $\mathbf{j}_{\rm tot}$ is obtained by summing over all values of m, $k_\parallel$, and k_z, and multiplying by 2 for the two spin orientations. Each term in the sum is weighted by the probability that the initial state (in the tip) is occupied and the final state (in the sample) is unoccupied. At zero temperature, these probabilities are given by Heaviside Θ functions,

$$\mathbf{j}_{\rm tot} = 2 \sum_{m=-\infty}^{+\infty} \int_0^{+\infty} k_\parallel dk_\parallel \int_0^{+\infty} dk_z \, \mathbf{j}_{m,k_\parallel,k_z} \Theta((k_\parallel)^2 + (k_z)^2 + V_{\rm in}^1 - (E_F - V_b)) \Theta(E_F - ((k_\parallel)^2 + (k_z)^2 + V_{\rm in}^1)) . \tag{20}$$

For convenience, we use the transformation

$$k_z = [E - (k_\parallel)^2 - V_{\rm in}^1]^{1/2} \tag{21}$$

to replace the integration over k_z by an integration over the energy E:

$$\mathbf{j}_{\rm tot} = \sum_{m=-\infty}^{+\infty} \int_0^{+\infty} k_\parallel dk_\parallel \int_{E_F - V_B}^{E_F} dE \frac{\mathbf{j}_{m,k_\parallel,E}}{k_z} . \tag{22}$$

In practice, the integral over $k_\parallel$ is replaced by a sum over the discrete values of $k_\parallel$ allowed by the relation (17) and the integral on E is replaced by a sum over a mesh of equidistant energies between $E_F - V_B$ and E_F.

III. FINITE-ELEMENT METHOD

The finite-element method (FEM, see, e.g., Ref. 28) is a discretization method instrumental in the resolution of partial differential equations. It is commonly applied in stress-strain calculations, hydrodynamics or aerodynamics. Its use in quantum mechanics is much less frequent.[29–36] The method we describe here is a 2D generalization[37] of the technique we developed for 1D scattering problems.[36]

The two-dimensional area in which Eq. (3) must be solved numerically is defined by $0 \le \rho \le R$ and $0 \le z \le D$. We divide this area in smaller domains: the so-called *finite elements* (FE). We use square elements with a side

$h=0.5$ Å. Now, let us define a basis set adapted to the mesh. In the 1D FEM,[36] the basis set includes two *shape functions* ϕ^0 and ϕ^1 per node z_i of the mesh. Each 1D shape function centered on z_i extends from z_{i-1} to z_{i+1}, i.e., over two adjacent FE's. The transformation

$$z \rightarrow t = |z - z_i| / h \tag{23}$$

maps the finite elements (z_{i-1}, z_i) and (z_i, z_{i+1}) on $(0,1)$, in which the shape functions are expressed as

$$\phi^0(t) = (1-t)^2(1+2t) , \tag{24}$$

$$\phi^1(t) = \mathrm{sgn}(t) h t (1-t)^2 . \tag{25}$$

The shape functions $\phi^0(z)$ and $\phi^1(z)$ are continuous and differentiable, the ϕ's and their derivatives vanish on all the nodes of the mesh, except that $\phi^0(0)=1$ and $d\phi^1(0)/dz = 1$. ϕ^0 and ϕ^1 are plotted in Fig. 3. In general, different choices can be made for the basis set in the FEM. Taking two shape functions per node as described above offers the advantage that the derivative as well as the value of the wave function can be described properly. This is necessary because the probability current depends on both these quantities.

In the study of 2D tunneling,[37] a natural choice for the shape functions is to take products of one-dimensional shape functions of each of the coordinates. As a consequence, four shape functions are centered on each node of the 2D mesh (see Fig. 4). They are localized on the four elements containing the node (except for the nodes lying on the boundary of the domain). Considering the way we derived the 2D shape functions from the 1D ones, it is straightforward that their values, their derivatives with respect to z, their derivatives with respect to ρ, and their second derivatives with respect to z and ρ vanish on all nodes of the mesh, except that for each of these shape functions, one of these four quantities equals 1 at the node on which it is centered. As for 1D tunneling, the choice we made for the shape functions allows us to represent properly the probability current density.

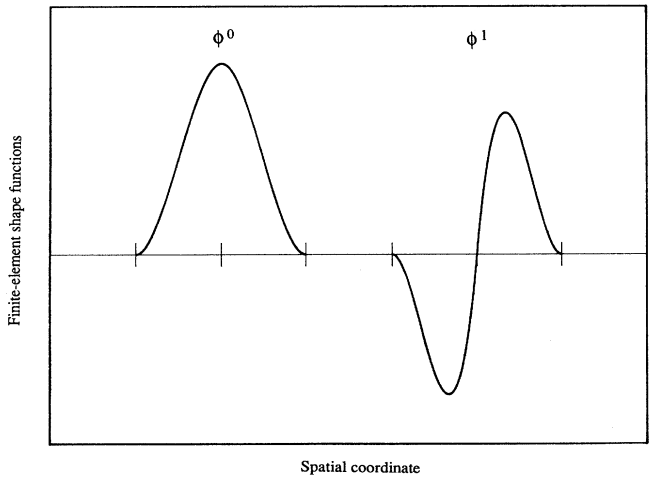

FIG. 3. The two types of shape functions used in the 1D FEM. For the clarity of the picture, we plotted shape functions centered on different nodes of the mesh.

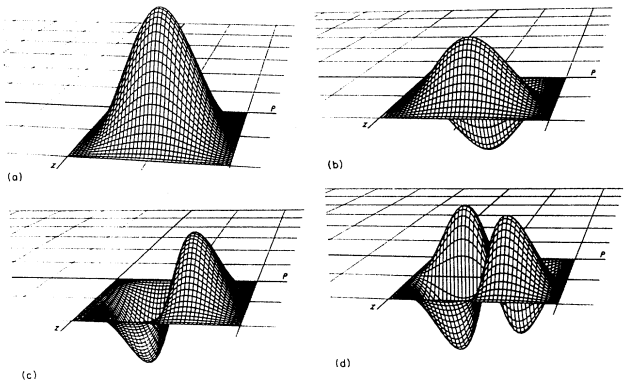

FIG. 4. The four types of shape functions used in the 2D FEM.

The basis set includes the following functions.

(i) The shape functions centered on all nodes, except the nodes with $z_i=0$ or $z_i=D$ (for the nodes lying on the axis or the external surface of the cylinder, only the shape functions satisfying the boundary conditions are included in the basis set).

(ii) The incident and reflected waves defined by (11), extending in the region $z \le 0$ and continued smoothly in the first row of FE's ($0 \le z \le h$) by the appropriate combination of shape functions associated with the nodes $z_i=0$ (this is how the matching conditions at the interface between the tip and the barrier are implemented).

(iii) The transmitted waves defined by (12), extending in the region $z \ge D$ and continued smoothly in the last row of FE's ($D-h \le z \le D$) by the appropriate combination of shape functions associated with the nodes $z_i=D$.

The number of reflected and transmitted waves to be included in the basis set is kept finite by the quantization (17) and by the FE discretization scheme: since the J_m are oscillating functions, they can be represented satisfactorily in a FE basis only if the distance between two consecutive zeros remains larger than the discretization length h or, equivalently, if $k_\parallel$ is not too large. In practice, we compute the J_m and the corresponding values of $k_\parallel$ by solving Bessel's equation using the 1D FEM. These numerical J_m are well adapted to the FE representation of the wave functions and their number is equal to twice the number of elements in $(0,R)$. All basis functions are continuous and differentiable and so will be the wave functions we will find.

Using the basis set described above, we can express any wave function ψ as

$$\psi(\rho,z) = \psi_{\mathrm{inc}}(\rho,z) + \sum_i \psi_i \phi_i(\rho,z) , \tag{26}$$

where ψ_{inc} is the incident wave characterizing the electronic state (actually the basis function obtained by a continuation of this wave, as explained above) and the ϕ_i are the elements of the basis set, excluding the incident waves. The ψ_i are the (so far unknown) discretization parameters, corresponding either to a coefficient B or C of

Eqs. (13) and (14) or to the value, the ρ or z derivative or the second ρz derivative of ψ on a node of the mesh. Using the Hamiltonian operator H and the form (26) of ψ, we write Schrödinger's equation (3) as

$$\sum_i \psi_i (H-E)|\phi_i\rangle = -(H-E)|\psi_{\text{inc}}\rangle \ . \tag{27}$$

In order to solve (27) for the ψ_i's we need as many independent equations as the number of unknown quantities. We obtain them using the so-called *Galerkin method* by imposing that the projection of both sides of (27) on every basis function $|\phi_j\rangle$ be equal. So, for all j,

$$\sum_i \psi_i \langle \phi_j|(H-E)|\phi_i\rangle = -\langle \phi_j|(H-E)|\psi_{\text{inc}}\rangle \ . \tag{28}$$

Equations (28) form a linear system called the *weak form* of Eq. (27).

The explicit form of the matrix elements in Eq. (28) is

$$\langle \phi_j|(H-E)|\phi_i\rangle = \int_{-\infty}^{+\infty} dz \int_0^R \rho\, d\rho\, \phi_j(\rho,z) \left[-\frac{\partial^2}{\partial\rho^2} - \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{\partial^2}{\partial z^2} + \frac{m^2}{\rho^2} + V(\rho,z) - E \right] \phi_i(\rho,z) \ . \tag{29}$$

The integration domain in (29) is at most four FE's whichever functions ϕ_i and ϕ_j we take. The reason for this is that the shape functions are localized on at most four FE's and that the incident reflected and transmitted waves which extend to $z=-\infty$ or to $z=+\infty$ are solutions of Schrödinger's equation in the region $z \le 0$ or $z \ge D$, making the integrand in (29) equal to zero in that region. Equation (29) involves second derivatives of the basis functions, whereas these are in general not twice differentiable. This problem may be overcome by using integration by parts and the properties of the basis functions on the boundaries of the integration domains:

$$-\int_{-\infty}^{+\infty} \phi_j \frac{\partial^2 \phi_i}{dz^2} dz = \int_{-\infty}^{+\infty} \frac{\partial \phi_j}{\partial z}\frac{\partial \phi_i}{\partial z} dz \ , \tag{30}$$

$$-\int_0^R \phi_j \left[\frac{\partial^2 \phi_i}{d\rho^2} + \frac{1}{\rho}\frac{\partial \phi_i}{d\rho} \right] \rho\, d\rho = \int_0^R \frac{\partial \phi_j}{\partial \rho}\frac{\partial \phi_i}{\partial \rho} \rho\, d\rho \ . \tag{31}$$

The integrals are computed using a three-point Gauss-Legendre[27] integration formula:

$$\int_{\text{FE}} f(x)dx \approx \sum_{i=1}^{3} w_i f(x_i) \ , \tag{32}$$

which, with a suitable choice of the weights w_i and the sampling points x_i, is exact if the integrand $f(x)$ is a polynomial with a degree not larger than 5. Formula (32) is a good approximation for higher degree polynomials or for integrands other than polynomials. Its use requires the knowledge of the integrand (in particular the potential) only on the sampling points, i.e., in the 2D case, on nine points per FE.

The linear system (28) is solved using the Gaussian triangular factorization (LU) technique.[38] This yields the unknown coefficients ψ_i in the development (26), hence the representation of the wave function ψ in the FE basis. This representation is continuous and differentiable at every point. Thus, the current density associated to any tunneling state ψ can be extracted using Eq. (19).

It results from the Schrödinger equation (1) that the current defined by (19) obeys the continuity equation, i.e., that the probability density (and the charge) is conserved. The numerical solutions we obtain by the FEM do not strictly have the same property. Nevertheless, for infinitely small FE ($h \to 0$), the approximate FE solution must converge towards the exact solution. Thus, testing the charge conservation in our solutions may be used to check their accuracy. For an electron with an energy of 9.5 eV tunneling through a one-dimensional parabolic barrier 14 eV high and 5 Å wide, the deviations from current conservation amount to 6% if the barrier is divided into five FE's, but they reduce to 0.9% and 0.1% with 10 or 20 FE's, respectively. In the 2D computations, we take $h=0.5$ Å. If we integrate the current on various cross sections of the cylinder, we find deviations from charge conservation of the order of 1%.

IV. RESULTS AND DISCUSSION

Figure 5 displays current-density maps in the plano-hemispherical model for two sets of parameters. The current flows in a narrow beam from the apex of the tip, which is consistent with the atomic resolution of the STM. Results for the plano-cylindrical model are shown in Fig. 6. Between the flat surfaces, the current pattern is similar to the current in a planar junction, except near the edges of the cylinder, where it diverges. When the tip is far enough from the sample in both Figs. 5 and 6, current loops appear on the side of the barrier where the reflected wave interfere with the incident waves. Because of those curls, it may also happen [Fig. 5(b)] that the current density leaves the tip from its planar part, enters the hemisphere through its side, travels in the interior of the tip to its apex, and, finally, tunnels to the sample

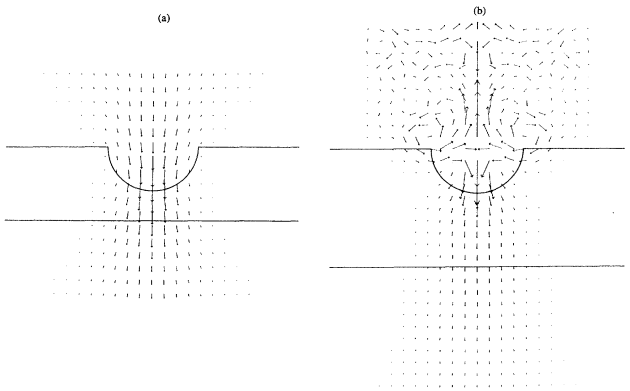

FIG. 5. Maps of the current density between a hemispherical tip with a 3-Å radius and a flat surface located (a) 2 Å and (b) 5 Å from the tip. The applied bias is 10 mV.

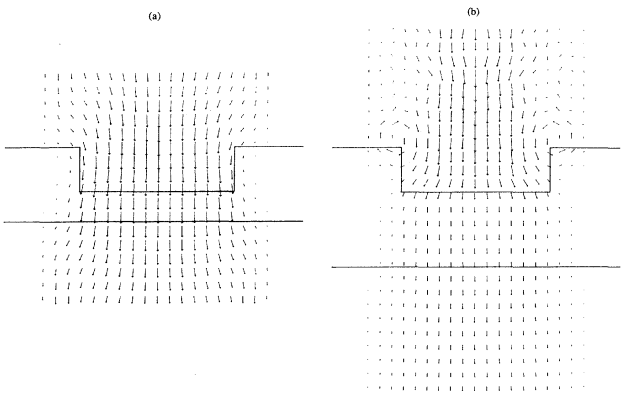

FIG. 6. Maps of the current density between a cylindrical tip with a 5-Å radius and a 3-Å height and a flat surface located (a) 2 Å and (b) 5 Å from the tip. The applied bias is 10 mV.

through the thinnest part of the barrier. When the electrodes are moved infinitely apart from each other, the curls still remain, which proves that they are caused by the reflection on a corrugated barrier rather than by tunneling. It might be conjectured than the curls are caused by the sharp kinks between the hemisphere (or the cylinder) and the planar surface. Figure 7, however, which displays the current between a Gaussian tip and a planar surface, demonstrates that even with a smooth tip, the current loops show up.

For sample surfaces with a relief (Fig. 8), the current flow depends strongly on the kind of site above which the tip is located. The electrons cross the barrier where it is the thinnest. This results in a very narrow beam when the tip is above a boss and, when the tip is above a trough, in an annular beam between the tip and the sides of the dip.

It is interesting to examine the contribution of each value of the angular momentum m to the current. For example, in the plano-hemispherical junction with $R=3$ Å, $s=2$ Å, and $V_{\text{bias}}=10$ mV, the $m=0$ states carry 62% of the current, and the $m=\pm1$ states, 33%. A resolution of the same model by the Green-function method[15] concluded that the $m=0$ contribute 90% of the current. This is explained by the fact that, in Eq. (3), the term involving m acts as a centrifugal potential that keeps the electron away from the axis $\rho=0$. As a matter of fact, only the $m=0$ states do not vanish on the axis. Now, the thinnest part of the barrier is on the axis. Thus, in the plano-hemispherical geometry, the $m\neq0$ states have to cross a thicker barrier and the probability they tunnel through is smaller. For a cylindrical tip, the apex is a plateau and the thin part of the barrier is wider. For that reason, the contribution to the current from the $m=0$ states is not so dominant as for hemispherical tips: they contribute 35% of the current, while the contribution of the $m=\pm1$ states is 36% and that of the $m=\pm2$ states is 26%. For a hemispherical tip above a Gaussian boss, the tunneling electron beam is so narrow that the $m=0$ states contribute 90% of the current and almost all the rest is carried by $m=\pm1$ states. When the tip is above a dip, the current flows essentially from the sides of the tip: as a result, the current is distributed over a larger number of angular momenta: 17% for $m=0$, 30% for $m=\pm1$, 21% for $m=\pm2$, 12% for $m=\pm3$, etc.

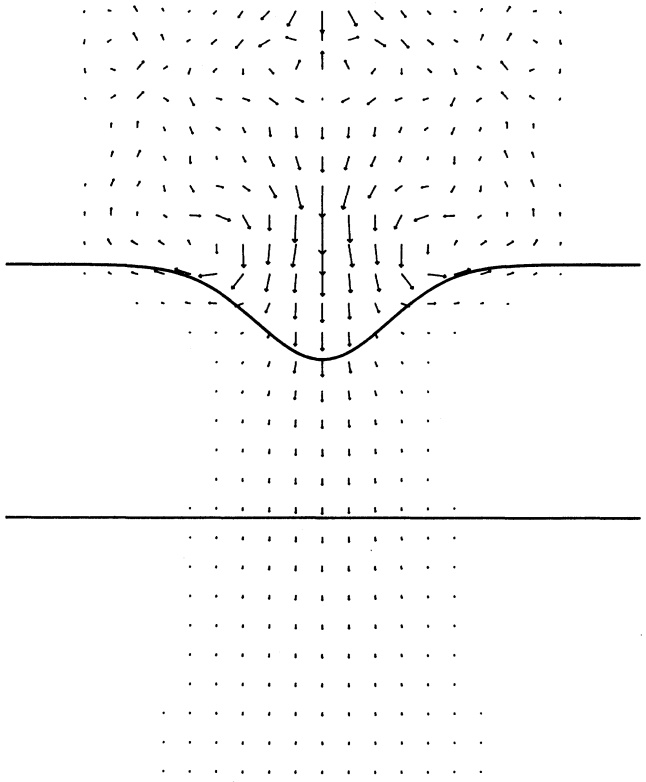

FIG. 7. Map of the current density between a Gaussian tip with a 3-Å height and a 2-Å standard deviation and a flat sample located 5 Å from the tip. The applied bias is 10 mV, the inner potentials are 14 eV below the vacuum level, and the work functions are 4.5 eV.

We used our results for the models with a planar sample surface to investigate how the width of the current beam depends on the tip shape and size. The fact that

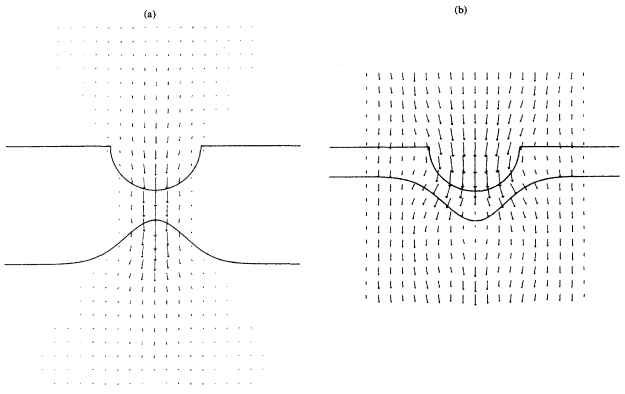

FIG. 8. Maps of the current density between a hemispherical tip with a 3-Å radius and a Gaussian (a) boss or (b) trough with a 3-Å height and a 2-Å standard deviation. The distance between the tip and (a) the top or (b) the bottom of the Gaussian is 2 Å. The applied bias is 10 mV.

the current flows as a localized beam between the tip and the surface is a consequence of the nonplanarity of the barrier. Therefore, the investigation of the beam is a point in which our method and our models can be instrumental, since we give an exact treatment of tunneling through corrugated barriers. The quantity we use to characterize the beam is the half width at half maximum (HWHM) of the distribution of the normal component of the current (j_z) versus the distance from the axis (ρ), i.e., the distance at which j_z is half its value on the axis.

In Fig. 9, we plot the HWHM for the plano-hemispherical model versus the distance s between the apex of the tip and the surface, for different values of the tip radius R. The dependence of the HWHM on the tip radius is rather small: the HWHM increases by about 0.2 Å per Å increase in the radius. When the tip is moved towards the sample, the HWHM first decreases and below a value s (depending on the radius and about 2 Å) it increases. This is due to the fact that, for short tip-to-sample distances, the multiple image potential near the apex of the tip lowers the barrier by such an extent that there is no more barrier for the electrons at the Fermi level. This is the *electronic contact* regime.[12] In that regime, approaching the tip towards the sample makes the hole in the barrier larger and thus increases the width of the electron channel.

When the tip is a plateau of atoms (Fig. 10), the HWHM is slightly larger than the radius of the tip. This is obviously explained by the fact that the current is normal to the surfaces except near the edges of the tip, where it fans out. The distance s has no strong influence on the HWHM.

The ability of the STM to resolve a relief on a metal surface can be investigated using the models with a boss or a trough. In Fig. 11, the total current is plotted versus the distance between the apex of the tip and the top of the boss or the bottom of the dip. This enables us to compare the measured corrugation to the real corrugation. Let us assume that the boss and the dip are present on the same surface and far enough from each other to be imaged separately by the STM. The corrugation of the sample surface is thus 6 Å peak to valley. Now, it can be seen on Fig. 11 that, for instance, the current is the same when the tip is 2 Å above the hill as when it is 4.6 Å above the dip. Thus, the measured corrugation in the constant current mode will be 3.4 Å, i.e., twice smaller than the sample corrugation. This is consistent with the experimental fact that a corrugation is only exceptionally resolved on a clean and atomically flat metal surface.

The exponential dependence of the current with

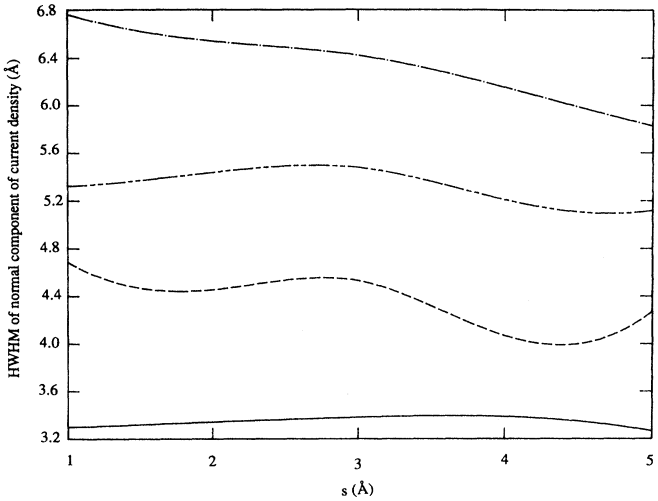

FIG. 10. Half width at half maximum (HWHM) of the normal component of the current density on the sample surface vs the distance (s) between the cylindrical tip and the sample for various values of the tip radius (——, $R=3$ Å; — — —; $R=4$ Å; ----, $R=5$ Å; —·—·—·, $R=6$ Å).

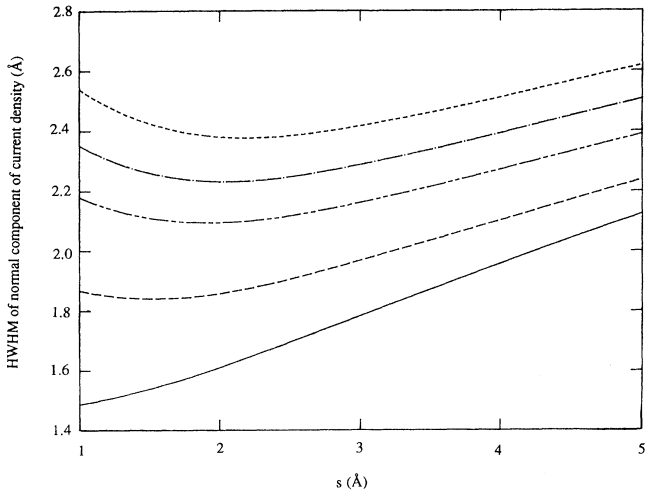

FIG. 9. Half width at half maximum (HWHM) of the normal component of the current density on the sample surface vs the distance (s) between the hemispherical tip and the sample for various values of the tip radius (——, $R=2$ Å; — — —, $R=3$ Å; ----, $R=4$ Å; —·—·—·, $R=5$ Å, · · · ·, $R=6$ Å).

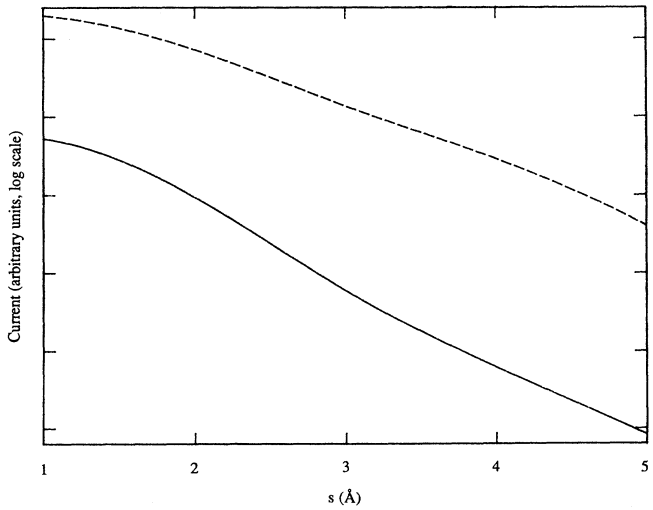

FIG. 11. Tunnel current (log scale, one decade per tick) vs the distance s between a hemispherical tip and the top of the Gaussian boss (solid line) or the bottom of a Gaussian dip (dashed line).

respect to the distance between the tip and the sample is the fingerprint of tunneling. The decay length of the current is proportional to the square root of the work function of the sample. In Figs. 12 and 13, we plot the logarithm of the current versus the distance s, respectively, for the plano-hemispherical and the plano-cylindrical model, and different values of the tip radius. In all cases, the behavior of the current is not strictly exponential: the effective work function decreases when the tip comes closer to the sample. From the slope of the curves, we determine the work function as 2.5 eV for $s=1$ Å and 3.9 eV for $s=5$ Å in the case of the hemispherical tip and 0.7 eV ($R=3$ Å) to 1.0 eV ($R=6$ Å) for $s=1$ Å and 6.8 eV for $s=5$ Å in the case of the cylindrical tip. The real work function we put in the calculations is 4.5 eV (for both the tip and the sample). We interpret the variation of the effective work function with distance as an effect of the interaction of the tunneling electrons with their images. According to a theory by Binnig *et al.*,[5] the image potential should not influence the measured work function. Our calculation is different in that it takes into account the three-dimensional nature of tunneling in the STM, i.e., the complicated shape of the barrier and the fact that some electrons have a nonvanishing $k_{\parallel}$. For those electrons, the effective barrier height is larger than for electrons with a momentum perpendicular to the sample surface. As a consequence, the dependence of the current on the distance is by no means trivial.

Many theories published so far about STM rely on the THA. Since the present work does not, it can be used to test the validity of the THA. For that purpose, we compute the total current in the plano-hemispherical model with the THA and compare it with our essentially exact results. In the THA, the total current $i_{\rm tot}$ for a small bias V_b is given by the golden rule:

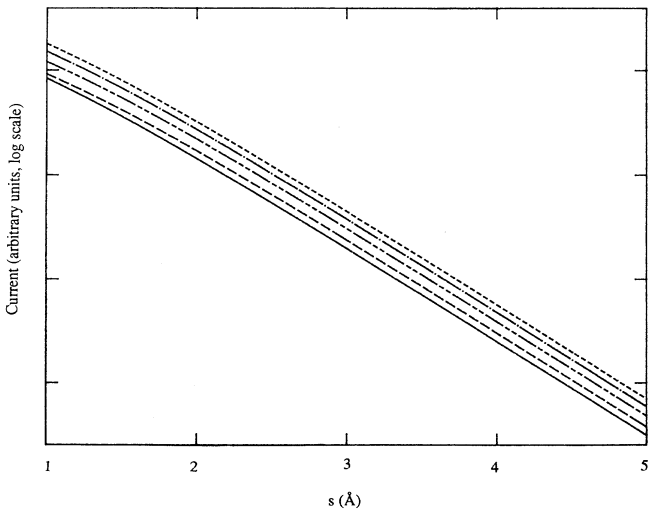

FIG. 12. Tunnel current (log scale, one decade per tick) vs the distance s between a flat sample and a hemispherical tip with different radii (——, $R=2$ Å; — — —, $R=3$ Å; ----, $R=4$ Å; —·—·—·, $R=5$ Å; · · · ·, $R=6$ Å).

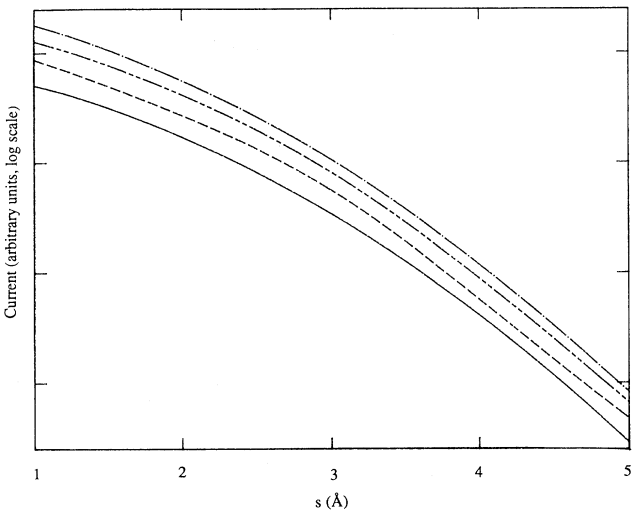

FIG. 13. Tunnel current (log scale, one decade per tick) vs the distance s between a flat sample and a cylindrical tip with a 3-Å height and different radii (——, $R=3$; — — —, $R=4$ Å; ----, $R=5$ Å; —·—·—·, $R=6$ Å).

$$i_{\rm tot}=\frac{2\pi e^2 V_b}{\hbar}\sum_{m=-\infty}^{+\infty}\sum_{k_{\parallel}^1}\sum_{k_{\parallel}^2}\rho_1(E_F,k_{\parallel}^1)|M_{m,k_{\parallel}^1,k_{\parallel}^2}|^2 \times\rho_2(E_F,k_{\parallel}^2)\ , \tag{33}$$

where ρ_1 and ρ_2 are the densities of states with given energy and parallel momentum in the tip and the sample, respectively, and the tunneling matrix element M is obtained from the wave functions by the formula of Bardeen:[39]

$$M_{m,k_{\parallel}^1,k_{\parallel}^2}=\frac{\hbar^2}{2m}\int_S(\psi^*_{m,k_{\parallel}^1}\nabla\psi_{m,k_{\parallel}^2}-\psi_{m,k_{\parallel}^2}\nabla\psi^*_{m,k_{\parallel}^1})\cdot d\mathbf{S}\ . \tag{34}$$

In Eq. (34), the integral is evaluated on a plane S parallel to the sample surface and lying in the vacuum halfway between the sample and the apex of the tip. The wave functions are computed separately for each electrode, using the FEM.

As can be seen from the top two curves of Fig. 14, the difference between the THA and our numerical method is very small. The two curves even cross at $s=1.3$ Å. This agreement is surprising, since it was expected that the THA breaks down for short distances. We interpret this unexpected agreement as the accidental cancellation of two errors made when using the THA.

(1) Equation (33) only give the current at the first order of perturbation.

(2) When we consider electrodes infinitely apart from each other, we neglect the multiple images in the tip and sample surfaces, keeping only one image in each electrode.

To test this hypothesis, we repeated our calculations, neglecting completely the image potential, both in the exact and the THA treatments. The lower two curves in Fig. 14 are the results of this test and are nearly undistinguishable. The THA always overestimates the tunnel

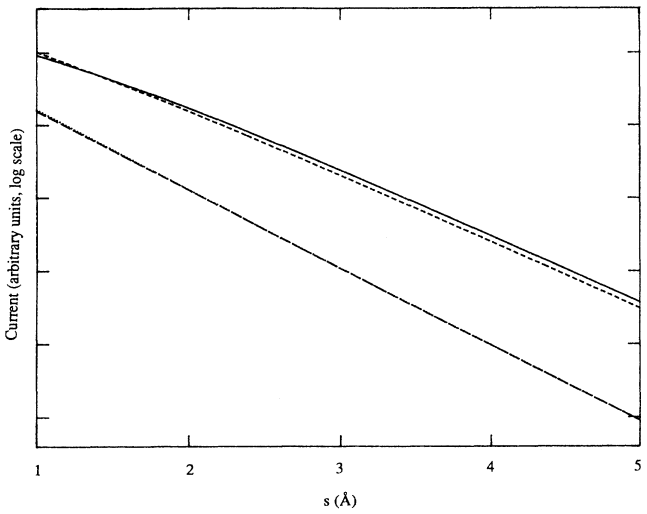

FIG. 14. Tunnel current (log sale, one tick per decade) vs the distance s in the model of Fig. 1(a) and a hemisphere radius of 3 Å. The top two curves give the current when the image potential is present, in the exact formalism (full line) and for the THA (short-dashed line). The lower two curves give the current obtained by neglecting the image potential. The long-dashed curve corresponds to the exact treatment and the dotted line to the THA.

current. In the calculations including the image potential, the current is about one order of magnitude larger because the image potential lowers the barrier but the THA underestimates this lowering and, for that reason, it underestimates the current for short distances. We conclude that the THA gives a good approximation of the current, even at low distances, and that the main error made in the THA is neglecting the multiple images. It must, however, be noted that even when we used the THA, we computed the wave functions of each electrode very accurately, whereas other authors[4,10,14,17,18] took modeled states (e.g., spherical states) and, sometimes, a single state for the tip.

V. CONCLUSION

We developed a finite-element method instrumental in solving Schrödinger's equation for electrons tunneling between a STM tip and a metal surface. The models we use for the STM and the sample consist in free-electron metallic tip and sample with simple geometries. The simplicity of these models allows us to obtain accurate and reliable values for the current density between the tip and the sample. The electrons cross the barrier preferentially where the distance between the electrodes is small. This results in a very narrow current channel when the tip is hemispherical and faces a planar surface or a protrusion of the sample. When the tip faces a dip in the sample surface, the thinnest part of the barrier is between the sides of the tip and those of the trough and the current channel is ring-shaped. When the probability that the electron is reflected by the barrier is high, the current map exhibits curls on the side of the barrier where the incident waves interfere with the reflected waves.

The lateral resolution of the STM is related to the current beam arriving on the sample. For hemispherical tips, the beam gets slightly wider when the tip radius is increased and gets narrower when the tip is approached to the sample, until the image contribution to the potential is so strong that the barrier is perforated. For multiatomic tips, modeled by a cylinder, the current density is constant and normal to the top of the cylinder and to the sample surface except near the edges of the tip. Thus, the width of the current beam is always slightly larger than the diameter of the tip.

Calculating the tunnel current for a hemispherical tip at different heights above a Gaussian hill or a Gaussian dip on a metal surface, we are led to the conclusion that the corrugation of the metal surface will result in a corrugation twice smaller in the constant current image.

The effective work function calculated from the slope of the $\ln I/s$ curve is found to vary with the tip-sample distance and, to a smaller extent, with the tip radius. This is interpreted as an effect of the image potential and the three dimensionality of our model.

A comparison between our essentially exact treatment of tunneling and the THA shows that the latter yields quite a good approximation of the current, even for thin barriers. The main error introduced by the THA is the fact that for isolated tip and sample only one image in each electrode is considered, neglecting the multiple images.

ACKNOWLEDGMENTS

This work was funded by a program of the Interuniversity Research Projects. Our computations were performed on the Namur Scientific Computing Facility (Namur-SCF), which is funded by FNRS, IBM Belgium, and the Facultés Universitaires Notre-Dame de la Paix (FUNDP). Ph.L. would like to thank the FNRS for financial support.

[1]G. Binnig and H. Rohrer, Helv. Phys. Acta **55**, 726 (1982).
[2]G. Binnig and H. Rohrer, Surf. Sci. **126**, 236 (1983).
[3]G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982).
[4]J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).
[5]G. Binnig, N. Garcia, H. Rohrer, J. M. Soler, and F. Flores, Phys. Rev. B **30**, 4816 (1984).
[6]E. Stoll, Surf. Sci. **143**, L411 (1984).
[7]E. Stoll, A. Baratoff, A. Selloni, and P. Carnevali, J. Phys. C **17**, 3073 (1984).
[8]J. Bono and R. H. Good, Jr., Surf. Sci. **151**, 543 (1985).
[9]N. D. Lang, Phys. Rev. Lett. **55**, 230 (1985); **55**, 2925(E) (1985); **56**, 1164(E) (1986).
[10]A. Selloni, P. Carnevali, E. Tosatti, and C. D. Chen, Phys. Rev. B **31**, 2602 (1985); **34**, 7406(E) (1986).
[11]G. Doyen and D. Drakova, Surf. Sci. **178**, 375 (1986).
[12]S. Ciraci and I. P. Batra, Phys. Rev. B **36**, 6194 (1987).
[13]Balaram Das and J. Mahanty, Phys. Rev. B **36**, 898 (1987).
[14]M. S. Chung, T. E. Feuchtwang, and P. H. Cutler, Surf. Sci. **187**, 559 (1987).

[15]A. A. Lucas, H. Morawitz, G. R. Henry, J. P. Vigneron, Ph. Lambin, P. H. Cutler, and T. E. Feuchtwang, Phys. Rev. B **37**, 10 708 (1988).
[16]A. A. Lucas, H. Morawitz, G. R. Henry, J. P. Vigneron, Ph. Lambin, P. H. Cutler, and T. E. Feuchtwang, J. Vac. Sci. Technol. A **6**, 296 (1988).
[17]C. J. Chen, J. Vac. Sci. Technol. A **6**, 319 (1988).
[18]C. R. Leavens and G. C. Aers, Phys. Rev. B **38**, 7357 (1988).
[19]J. D. Todd and J. B. Pethica, J. Phys. Condens. Matter **1**, 9823 (1989).
[20]G. Doyen, E. Koetter, J.-P. Vigneron, and M. Scheffler, Appl. Phys. A **51**, 281 (1990).
[21]J.-P. Vigneron, M. Scheffler, Th. Laloyaux, I. Derycke, and A. A. Lucas, Vacuum **41**, 745 (1990).
[22]P. Johansson, R. Monreal, and P. Apell, Phys. Rev. B **42**, 9210 (1990).
[23]J. B. Pendry, A. B. Prêtre, and B. C. H. Krutzen, J. Phys. Condens. Matter **3**, 4313 (1991).
[24]I. Derycke, J. P. Vigneron, Ph. Lambin, Th. Laloyaux, and A. A. Lucas, Int. J. Quantum Chem. Symp. **25**, 687 (1991).
[25]A. A. Lucas, J. -P. Vigneron, Ph. Lambin, Th. Laloyaux, and I. Derycke, Surf. Sci. **269/270**, 74 (1992).
[26]A. A. Lucas, J.-P. Vigneron, J. Bono, P. H. Cutler, T. E. Feuchtwang, R. H. Good, Jr., and Z. Huang, J. Phys. (Paris) Colloq. **45**, C9-125 (1984).
[27]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1986).
[28]A. R. Mitchell and R. Wait, *The Finite Element Method in Partial Differential Equations* (Wiley, New York, 1985).
[29]*Finite Elements in Physics,* edited by R. Gruber (North-Holland, Amsterdam, 1987).
[30]A. Askar, J. Chem. Phys. **62**, 732 (1975).
[31]S. Nordholm and G. Backsay, Chem. Phys. Lett. **42**, 253 (1976).
[32]M. Friedman, Y. Rosenfeld, A. Rabinovitch, and R. Thieberger, J. Comput. Phys. **26**, 169 (1978).
[33]M. Duff, H. Rabitz, A. Askar, A. Cakmak, and M. Ablovitz, J. Chem. Phys. **72**, 1543 (1980).
[34]S. Nordholm and G. Backsay, Chem. Phys. Lett. **42**, 259 (1976).
[35]S. Nordholm and G. Backsay, J. Phys. B **11**, 193 (1978).
[36]Th. Laloyaux, Ph. Lambin, J.-P. Vigneron, and A. A. Lucas, J. Comput. Phys. **83**, 398 (1989).
[37]Th. Laloyaux, A. A. Lucas, J.-P. Vigneron, Ph. Lambin, and H. Morawitz, J. Microsc. **152**, 53 (1988).
[38]G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, NJ, 1977).
[39]J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).