# Nonperturbative evaluation of STM tunneling probabilities from *ab initio* calculations

H. Ness and A. J. Fisher
*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*
(Received 18 March 1997)

We present a method for calculating the rate of electron transfer across the tunnel junction in a scanning tunneling microscope without any perturbation expansion in the tip-sample coupling. The method may be readily combined with separate *ab initio* electronic-structure calculations for the tip and the sample. This involves replacing the asymptotic scattering states by localized initial and final states on either sides of the tunnel junction. We present examples of applications to $C_2H_4$ molecules adsorbed on the Si(001) surface. [S0163-1829(97)06444-8]

## I. INTRODUCTION

The scanning tunneling microscope (STM) (Refs. 1 and 2) is now established as an important tool in the study of surface phenomena. Clean surfaces,[1] and surface processes such as oxidation,[3] adsorbed molecules,[4–6] and their reactions[7] have all been studied, to give only a few examples. Even very large systems such as DNA (Refs. 8 and 9) and thick alkane films[10,11] have been successfully imaged.

Understanding the information conveyed by STM images has itself become a major area of activity in theoretical physics.[12] A particularly important step forward was taken when Tersoff and Hamann[13,14] considered the resolution of an ''ideal'' STM. Working to first order in the tip-sample interaction, using the form of the tunneling matrix element due to Bardeen,[15] and assuming an *s*-like form for the tip wave function, they found that the tunnel current at small bias is simply proportional to the substrate density of states at the center of the tip. This result allows a relatively straightforward interpretation of STM images in a variety of situations. It is particularly powerful when combined with *ab initio* molecular-dynamics methods,[16] which allow the atomic structure and charge densities of complicated surface reconstructions,[17,18] or of adsorbed systems,[19] to be determined simultaneously.

First-order perturbation theory in tip-sample coupling can be extended beyond Tersoff-Hamann theory[20] to allow for the effects of higher angular momentum admixtures in the tip orbital,[21–24] to treat the effects of local variations in the potential,[25] and to incorporate the wave functions of realistic tip models in order to discuss issues such as the tip dependence of images[26,27] and the occurrence of negative differential resistance in the STM.[28]

There remain problems, however, in which perturbation theory is not expected to provide an adequate picture of the tunneling process because the tunneling is, in some sense, strong and cannot be treated as a perturbation. One such example occurs in the transition between the tunneling and point-contact regimes,[29,30] where the transmission coefficient for electrons between the tip and the sample becomes of order unity. This case has recently been extensively studied in the context of pull-off of a STM tip in contact with a surface,[31–33] and in the formation of thin junctions when notched wires are broken apart.[34]

Another case involves tunneling through an adsorbed molecule; here the difficulty is that although the tunneling matrix element between the tip and molecule may not itself be large, the molecule-substrate interaction and the interaction between opposite ends of the molecule introduce additional small energy scales which should be treated on the same footing as the interaction with the tip. An extreme example occurs in studies of tunneling through thick and nominally insulating organic films,[10,11] where one expects that tunneling through the film is at least as unlikely as tunneling across the vacuum gap between the tip and the film. Such problems require a unified treatment of electron transport through the coupled system of tip, substrate, and any intervening molecule.

There have been a number of previous nonperturbative calculations of tunneling in three-dimensional systems. These include treatments of resonant tunneling through a simple barrier[35] and between two planar surfaces, one of which has a protruberance which models a STM tip.[36] The resistance of a single atom adsorbed between two jellium surfaces was computed nonperturbatively by Lang,[30] and successfully compared with the experimental results of Gimzewski and Möller.[29] A complete theory to calculate the tunneling current between two electrodes has been developed by Noguera.[37–39] It is based on the matching of Green's functions on two surfaces defining three domains: the left and right electrodes, and the barrier (where no occupied states are available). To our knowledge, there have been no direct applications of the formalism for realistic three-dimensional tip-sample systems. This is due to the difficulty of exactly calculating the Green's functions in real space on the two surfaces in order to perform the matching. Another model, with a particular tip geometry, was later proposed by Sacks and Noguera.[40] More recently[41] the tunneling current between aluminum and palladium surfaces, represented using muffin-tin orbitals and a tip atom adsorbed on a jellium slab, has been computed by non perturbative solution of the equations of scattering theory.

In a different class of approaches, the electronic structure of the tip and substrate (including adsorbates if present) is mapped to a system described by some Hamiltonian that is relatively simple (at least, simple in comparison with the Hamiltonian of a first-principles electronic-structure calculation). This can either be a tight-binding model,[31,42] where the

tunneling current can be calculated exactly even for a complex three-dimensional system, or a simpler model potential problem.[43] Results have also been obtained using the tight-binding approach for adsorbed molecules;[44] using the model potential approach, the conductance has been calculated as a function of position of an aluminum atom between two aluminum surfaces.[45]

The present paper is motivated by the need for a method which allows a nonperturbative calculation of the tunnel current, but which can be used directly with the self-consistent potential produced by total-energy calculations or *ab initio* molecular dynamics. Such a method allows nonperturbative calculations even for complex systems for which it is difficult to fit a simple approximate form to the true potential, and enables one to calculate the geometry and STM image of a system together from first principles.

We begin the paper with a brief discussion of Landauer's conductance formula (Sec. II A); in particular, we emphasize that it expresses the current carried by the system in terms of the $S$-matrix element (a particular type of Green's-function matrix element) between asymptotic propagating states on either side of the tunnel barrier. We then express the current in terms of a different set of Green's-function matrix elements, those between *localized* states on either side of the barrier (Sec. II B). These matrix elements have the advantage that they can be computed within a total-energy calculation with periodic boundary conditions. We discuss the effect of the presence of many electrons in the system (Sec. II B 2), and show how it may be accounted for through a scheme which treats electron and hole tunneling on an equal footing. Next, we move on to consider how the Green's-function elements in question can be evaluated (Sec. II D). We also provide a scheme to calculate these Green's-function elements for real systems without the necessity of performing an expensive summation over the entire spectrum of the coupled tip-sample system (see the Appendix). In Sec. III we discuss the connection of the present formalism to other existing theories. We then turn to some actual applications of the method (Sec. IV). These concern more especially the STM contrast of adsorbed ethene ($C_2H_4$) molecules on the Si(001) surface, and the effects of the tip-induced electric field on this contrast. An application for the bare graphite surface has been already published elsewhere[46]

## II. TUNNEL CURRENT AND GREEN'S FUNCTIONS

### A. Transmission coefficient

We begin our discussion of tunneling with the Landauer formula[47] for the current in a system containing noninteracting electrons and dominated by elastic scattering. This equation formalizes the intuitively appealing notion that the current carried through a system should be proportional to the product of the rate at which carriers arrive from some source with the probability that those carriers are successfully transmitted through the sample. Originally derived heuristically,[47] this result has been related (with the aid of some delicate mathematics) to the Kubo-Greenwood formula of first-order perturbation theory.[48] It has been generalized to multiterminal systems,[49] to systems in which the electrons undergo arbitrarily strong interactions with each other in a finite part of the system,[50] or to systems in which the potential in the asymptotic region is nonseparable.[51,52] In this paper, we consider only a simple two-terminal STM device, and the two-terminal form of the result is sufficient:

$$I=\frac{e}{h}\int_{E_F}^{E_F+eV} dE\ \mathrm{Tr}[\hat{\rho}_{0\ \mathrm{left}}(E)\hat{t}^{\dagger}\hat{\rho}_{0\ \mathrm{right}}(E)\hat{t}], \tag{1}$$

where $I$ is the dc current, $E_F$ the equilibrium Fermi energy of the system, and $V$ the applied voltage. The density matrices $\rho_{0,\mathrm{left}}(E)$ and $\rho_{0,\mathrm{right}}(E)$ specify the occupancies of the states of the incoming and outgoing leads, while the $t$ matrix is related to the $S$ matrix of the system in the following manner:[53]

$$\hat{S}=1-\frac{2\pi i}{\hbar}\sum_{ij}\ |j\rangle t_{ij}\delta(E_i-E_j)\langle i|. \tag{2}$$

However, $\hat{S}$ is also related to the time-evolution operator of the system in the interaction representation:

$$\hat{S}=\hat{U}_I(-\infty,\infty), \tag{3}$$

where

$$\begin{aligned} i\hbar\,\partial_t\hat{U}_I(t,t_0)&=\hat{V}_I(t)\,\hat{U}_I(t,t_0),\\ \hat{V}_I(t)&=\exp(i\hat{H}_0t/\hbar)\hat{V}\,\exp(-i\hat{H}_0t/\hbar). \end{aligned} \tag{4}$$

Here $\hat{H}_0$ is taken to be the Hamiltonian of the leads, and $\hat{V}$ is the so-called perturbation potential in the total Hamiltonian that couples the incoming and outgoing leads. This implies that the $t$ matrix has the following interpretation: an incoming state $|\phi_i^{\mathrm{left}(+)}(E)\rangle$ at energy $E$ on the left of the sample (for example) is scattered into the combination $\Sigma_j t_{ij}(E)|\phi_j^{\mathrm{right}(+)}(E)\rangle$ on the right.

It is the definition of $\hat{S}$ and $\hat{t}$ in terms of asymptotic states that makes the exact calculation of the transmission so difficult to perform: it involves matching the potentials of the tip and of the sample smoothly onto the potential of an infinite wire or medium to carry the incoming and outgoing current. This is very difficult for a realistic model of a tip and surface containing, for example, adsorbed molecules or defects.

### B. Tunneling between localized states

The aim of this paper is to suggest that for the purpose of calculating the tunnel current through a STM junction the matrix elements of the $S$ matrix between incoming and outgoing propagating states may be conveniently replaced by the matrix elements of the quantum-mechanical propagator between two localized electronic states, one on each side of the tunnel junction. This means that the coupling of the states near the junction into the asymptotic scattering states (and the associated spreading resistance) is not described exactly. On the other hand, the essential physics of the tunneling process, including multiple scattering and tunneling through resonant surface or molecular states, is accurately described. We will come back to the interpretation of the present formulation and its connection to the $S$-matrix formalism and the Landauer formula in Secs. III B and III C.

### 1. A single electron

Motivated by this observation, we consider the evolution of an electron which at $t=0$ is localized in an initial state $|i\rangle$ on one side of the junction. Let us calculate the probability it may be found in a second localized state $|f\rangle$ on the other side of the junction at a later time $t$. This is given by

$$P(t)=|\langle f|\exp(-i\hat{H}t/\hbar)|i\rangle|^2 = \left|\sum_n \langle f|n\rangle\langle n|i\rangle\exp(-i\epsilon_n t/\hbar)\right|^2, \tag{5}$$

where the $|n\rangle$'s are eigenstates of the Hamiltonian $\hat{H}$ of the entire system, with eigenvalues $\epsilon_n$. In general this quantity will be a highly oscillatory function of $t$; we can, however, obtain a simpler measure of the rate of electron transfer across the junction by averaging over a time $\tau$. It is convenient to do this by introducing an exponentially decaying weighting function:

$$\overline{P(\tau)}\equiv\frac{1}{2}\frac{\int_0^\infty P(t)e^{-t/\tau}dt}{\int_0^\infty e^{-t/\tau}dt}. \tag{6}$$

(The factor of $\frac{1}{2}$ in the above definition has been introduced to make it consistent with others that will be introduced shortly for a many-particle system, where hole tunneling must also be taken into account.) We can simplify this expression by noting that the numerator can be written as

$$\hbar^2\int_{-\infty}^{\infty}|G_{fi}(t)|^2e^{-t/\tau}dt, \tag{7}$$

where $G_{fi}(t)$ is the causal Green's function connecting states $|i\rangle$ and $|f\rangle$:

$$G_{fi}(t)\equiv-\frac{i}{\hbar}\theta(t)\langle f|\exp(-i\hat{H}t/\hbar)|i\rangle, \tag{8}$$

where $\theta$ is the Heaviside step function. This Green's function can itself be written, from its Fourier transform, as

$$G_{fi}(t)=\int_{-\infty}^{\infty}\frac{dE}{2\pi\hbar}\,\tilde{G}_{fi}(E)\exp(-iEt/\hbar), \tag{9}$$

with

$$\tilde{G}_{fi}(E)\equiv\lim_{\delta\to0+}\sum_n\frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\delta}. \tag{10}$$

This result enables us to make a spectral decomposition of the average tunneling probability $\overline{P(\tau)}$ in the form

$$\overline{P(\tau)}=\frac{\eta}{\hbar}\int_{-\infty}^{\infty}dt\int_{-\infty}^{\infty}\frac{dE}{2\pi}\int_{-\infty}^{\infty}\frac{dE'}{2\pi}\sum_n\frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\eta}\sum_m\frac{\langle i|m\rangle\langle m|f\rangle}{E'-\epsilon_m+i\eta}\exp[-i(E-E')t/\hbar]$$
$$=\eta\int_{-\infty}^{\infty}\frac{dE}{2\pi}\left|\sum_n\frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\eta}\right|^2=\eta\int_{-\infty}^{\infty}\frac{dE}{2\pi}|\tilde{G}_{fi}(E+i\eta)|^2, \tag{11}$$

where $\eta=\hbar/2\tau$, and we have recognized the spectral representation of the $\delta$ function in performing the integrals over $t$ and $E'$.

It is worth noticing that the introduction of the exponential decaying weighting function in Eq. (6) is reminiscent of what Lippmann and Schwinger, in their original paper about scattering processes, called the adiabatic decrease of the perturbation when the time $t$ goes to infinity. This was achieved by introducing the factor $\exp(-\eta|t|/\hbar)$ into the expression of the scattering states $|\phi^{(+)}(E)\rangle$ [see, for example, Eq. (1.51) in Ref. 54]. On the other hand, this can be understood as a convenient way to introduce the finite imaginary part $\eta$ in the energy of the Green's function, in order to remove their singularities and make their energy integrals tractable.

### 2. Many electrons

The evaluation of $\overline{P(\tau)}$ using Eq. (11) is the central feature of the present approach. In one important respect, however, the real systems differ from the model we used to derive the result: states up to the Fermi energy $E_F$ are already filled with electrons, and the Pauli principle therefore forbids the introduction of an electron into these states. Instead, conduction through energy levels with $\epsilon<E_F$ takes place via holes.

We may formalize this by replacing the causal Green's function in Eq. (8) by a time-ordered Green's function,

$$G_{fi}(t)=-\frac{i}{\hbar}\langle 0,N|T[\hat{c}_f(t)\hat{c}_i^\dagger(0)]|0,N\rangle$$
$$\equiv-\frac{i}{\hbar}\theta(t)\langle 0,N|\hat{c}_f(t)\hat{c}_i^\dagger(0)|0,N\rangle$$
$$+\frac{i}{\hbar}\theta(-t)\langle 0,N|\hat{c}_i^\dagger(0)\hat{c}_f(t)|0,N\rangle, \tag{12}$$

where $|0,N\rangle$ is the $N$-particle ground state and the operators $\hat{c}_i^\dagger$ and $\hat{c}_f$, respectively, create an electron in single-particle state $|i\rangle$ and annihilate an electron in single-particle state $|f\rangle$. The first term in Eq. (12) describes propagation of an electron from state $|i\rangle$ to state $|f\rangle$, while the second term describes the propagation of a hole from $|f\rangle$ to $|i\rangle$. For the one-electron case discussed above, the second term vanishes.

A suitable measure of the time-averaged tunneling probability is then

$$\overline{P(\tau)}=\hbar^2\frac{\int_{-\infty}^{\infty}|G_{fi}(t)|^2e^{-|t|/\tau}dt}{\int_{-\infty}^{\infty}e^{-|t|/\tau}dt}. \tag{13}$$

The Fourier transform of the time-ordered Green's function can be written as

$$\widetilde{G}_{fi}(E)=\lim_{\delta\to 0+}\left[\sum_n \frac{\langle 0,\,N|\hat{c}_f|n,N+1\rangle\langle n,N+1|\hat{c}_i^{\dagger}|0,\,N\rangle}{E-(E_{n,N+1}-E_{0,N})+i\delta} + \sum_n \frac{\langle 0,\,N|\hat{c}_i^{\dagger}|n,N-1\rangle\langle n,N-1|\hat{c}_f|0,\,N\rangle}{E+(E_{n,N-1}-E_{0,N})-i\delta}\right] \tag{14}$$

where $|n,N+1\rangle$ and $|n,N-1\rangle$ are the many-particle eigenstates of the $N-1$ and $N+1$ particle systems, and $E_{n,N+1}$ and $E_{n,N-1}$ are the corresponding many-particle eigenvalues. Let us assume that the many-particle states are well approximated by a set of Slater determinants of single-particle eigenstates $|n\rangle$ with eigenvalues $\epsilon_n$. Then we can write

$$\widetilde{G}_{fi}(E)=\sum_n (1-f_n)\frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\delta}+\sum_n f_n\frac{\langle i|n\rangle\langle n|f\rangle}{E-\epsilon_n-i\delta}, \tag{15}$$

where $f_n$ is the occupation number of $|n\rangle$ and will be a Fermi function if the system is in thermal equilibrium. Similar manipulations to those used above then give the spectral decomposition of the time-averaged tunneling probability as

$$\overline{P(\tau)}=\eta\int_{-\infty}^{\infty}\frac{dE}{2\pi}\left|\sum_n (1-f_n)\frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\eta} + \sum_n f_n\frac{\langle i|n\rangle\langle n|f\rangle}{E-\epsilon_n-i\eta}\right|^2. \tag{16}$$

This equation corresponds to Eq. (11) in the single-particle system.

Equations (11) and (16) show that, for small values of $\eta$, the contribution to $\overline{P}$ from energy $E$ is overwhelmingly dominated by eigenstates $|n\rangle$ with $\epsilon_n\approx E$. In other words, the contribution to the tunneling probability at energies $E<E_F$ is essentially determined only by filled states, and at energies $E>E_F$ it is determined only by the empty states. The difference between Eqs. (11) and (16) is confined to a narrow region of width $\eta$ around the Fermi energy. We therefore make a very small error by using the causal Green's function and Eq. (11) to calculate the tunneling probability, instead of the time-ordered Green's function and Eq. (16).

#### *3. Reservoirs*

Up to this point we considered the propagation of the electron only once it has been introduced into state $|i\rangle$, by some as-yet-unspecified procedure. This implicitly assumes that we are always able to supply electrons to state $|i\rangle$ and remove them from state $|f\rangle$ at whatever rate is required to keep up with their propagation through the STM tunneling junction. This is, of course, an idealization; in reality, the supply and removal of the electrons is determined by connections to reservoirs in which the electrons can reach equilibrium by inelastic processes. If these reservoirs have the same electrochemical potential, so that all the states with a given energy are equally occupied, *no* net charge flows through the system. This is simply a consequence of the unitarity of the operator $\exp(-i\hat{H}t/\hbar)$; propagation from $|f\rangle$ to $|i\rangle$, once the electron has been introduced into $|f\rangle$, is just as likely as propagation from $|i\rangle$ to $|f\rangle$ once the electron has been introduced into $|i\rangle$.

The final step in our derivation is therefore to connect the initial and final states to reservoirs with electrochemical potentials $\mu_i$ and $\mu_f$, respectively. This has the consequence that electrons can be supplied to state $|i\rangle$ only at energies up to $\mu_i$, and removed from state $|f\rangle$ only at energies above $\mu_f$. Only in the energy region between $\mu_i$ and $\mu_f$ is there a contribution to the current. The transmission probability from $|i\rangle$ to $|f\rangle$ becomes

$$\overline{P(\tau)}=\eta\int_{\mu_f}^{\mu_i}\frac{dE}{2\pi}|\widetilde{G}_{fi}(E+i\eta)|^2. \tag{17}$$

We identify the average rate of electron transfer over time $\tau$ with the quantity

$$\frac{\overline{P(\tau)}}{\tau}=2\frac{\eta^2}{h}\int_{\mu_f}^{\mu_i}dE|\widetilde{G}_{fi}(E+i\eta)|^2. \tag{18}$$

Recognizing that each electron carries charge $e$ from one $|i\rangle$ to $|f\rangle$, so the current is $J\equiv e\overline{P(\tau)}/\tau$, and that the electrochemical potential is related to the voltage by a factor of $e$, we find that the differential conductance of the system is

$$\sigma=\frac{\partial\overline{J}}{\partial\mu_i}=e\frac{\partial\overline{P/\tau}}{\partial\mu_i}=\frac{2e^2}{h}\eta^2|\widetilde{G}_{fi}(E+i\eta)|^2. \tag{19}$$

This has the familiar form of the quantum of conductance, $2e^2/h$, multiplied by a dimensionless factor (the so-called dimensionless or normalized conductance $\sigma_N$).

### C. Choice of initial and final states

One important issue to be addressed in a method such as this is the choice of initial and final states between which to calculate the Green's functions. While the use of localized states $|f\rangle$ and $|i\rangle$ is important in that it enables us to obtain an expression for the tunnel current that can be evaluated in a system with periodic boundary conditions, we need to be sure that it does not introduce errors. The principal difficulty we found in applying this method occurs when there are several degenerate (or nearly degenerate) states which contribute to the tunneling at a particular energy. In these circumstances, the overlaps of the eigenstates $|n\rangle$ with $|f\rangle$ and $|i\rangle$ play an important role in ensuring that those states which most efficiently bridge the gap between tip and substrate contribute most to the formation of the STM image; on the other hand, one must ensure that the overlaps do not go too far in singling out some eigenstates of the degenerate subspace at the expense of the others.

We propose some procedures to circumvent such undesirable effects. Although the terms ''initial'' and ''final'' states are rather misleading since, as already pointed out in Sec. II B 3, the transport at any given energy is symmetric, i.e., $|\widetilde{G}_{if}|^2=|\widetilde{G}_{fi}|^2$ and the tunneling direction is only introduced by the applied bias. Therefore, we consider in the following the case of electron transport form the sample to the tip, and we concentrate for simplicity our discussion only on the ef-

fects of the overlap elements $\langle n|i\rangle$. However the solutions we propose hereafter are also applicable for the overlap elements $\langle f|n\rangle$.

The first possible solution consists in calculating an averaged differential conductance $\overline{\sigma}$ from Eq. (19) over a set of random localized $|i\rangle$ states in the sample. With such a method, it is expected that no particular state $|n\rangle$ among the states relevant for the tunneling energy will play a drastically more important role over another one.

The second possibility is to go beyond the consideration of localized state. This can be done by considering for the initial state a state coming from another *ab initio* calculation in which, for example, the sample is represented by a slab. In this case, the arbitrary choice of the localization of the $|i\rangle$ state on the values of the overlap $\langle n|i\rangle$ does not occur anymore. The values of such overlap matrix elements is strongly dependent on the remaining orthogonality between the $|i\rangle$ and $|n\rangle$ states. An automatic selection of the important $|n\rangle$ states relevant for the tunneling energy is then ensured. In practice, the $|i\rangle$ states are chosen to be the eigenstates of the corresponding isolated surface slab. The $|n\rangle$ are then the eigenstates of such a surface ''perturbed'' by the presence of adsorbates and/or by the presence of the tip. We comment further on the significance of such a choice for the $|i\rangle$ and $|n\rangle$ states in the next sections; the effects of this choice on the tunnel current are shown in Sec. IV.

### D. Calculation of the Green's function

Another important point in the theory presented in Sec. II B is that it involves the Green's function $\widetilde{G}(E)$ for the coupled tip-sample system. A direct calculation of $\widetilde{G}(E)$ via its representation in Eq. (10) would in principle involve the complete spectrum of the single-particle states of this coupled system.

Even in principle, a plane-wave basis set is not sufficient to describe the real-space representation $\widetilde{G}(\mathbf{r},\mathbf{r}',E)$ of the Green's function. Because of the finite size of this basis set, the Green's function cannot be described for energies above the cutoff energy corresponding to the plane-wave basis, and does not have the correct analytic properties. To reduce this error, a procedure has been proposed for correcting the Green's function (see, for example, Refs. 55 and 56). However, although, in principle, the plane-wave representation is not complete for the full $\widetilde{G}(\mathbf{r},\mathbf{r}',E)$, the calculation of the matrix elements $\langle f|\widetilde{G}(E)|i\rangle$ *will* be accurate, provided that the states $|i\rangle$ and $|f\rangle$ are themselves plane-wave converged.

In practice, we will show that only relatively few eigenfunctions make significant contributions to the tunneling current at a particular energy. In particular, we now concentrate our discussion on the convergence of the conductance with respect to (i) the number of states $|n\rangle$ included in the summation, and (ii) the value of the imaginary part $\eta$ of the energy. With respect to point (i), it is also possible to avoid the determination of the eigenvalue spectrum altogether in the calculation of the tunneling current by using an iterative method. A technique for doing this based on the recursion method (or Lanczos algorithm) is presented in the Appendix.

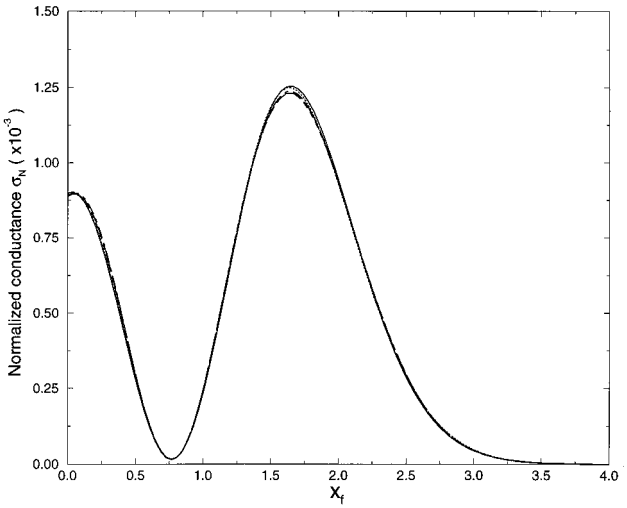



FIG. 1. Values of the normalized conductance $\sigma_N$ vs the position $x_f$ of the final Gaussian state $|f\rangle$. The Green's function $\widetilde{G}_{fi}$ is calculated by including $n_{\max}$ states. The different values of $n_{\max}$ are 8 (solid line), 10 (dotted line), 15 (dashed line), 20 (long dashed line), and 30 (dot-dashed line). The tunneling energy is $E/\hbar\omega = 2.6$ close to the eigenvalue $\epsilon_{n=2} = 2.5\hbar\omega$ and $\eta/\hbar\omega = 0.1$. The fixed position of the initial state is $x_i = -0.22$, and its corresponding Gaussian width is $\beta_i = 9.1$. Finally the final state moves along the $x$ axis, and its Gaussian width is $\beta_f = 7.1$.

#### 1. Direct calculation

In this section, we show that it is not necessary to have the complete eigenvalue spectrum representation for the Green's function Eq. (10) when one wants to calculate the tunneling current or the differential conductance as given by Eq. (19). The influence of the imaginary part $\eta$ of the energy on the conductance is also considered.

For this, we consider a simple one-dimensional model case. The eigenstates $|n\rangle$ and eigenvalues $\epsilon_n$ are chosen to be those of a harmonic oscillator. The Green's function is given by Eq. (10), where the index summation $n$ goes in principle to infinity. We take localized Gaussian functions for the initial and final states.

First of all, let us rewrite the expression of the differential conductance in a more convenient form. Starting from the definition of $\sigma(E)$, we can write the normalized conductance

$$\sigma_N(E) = \sigma(E)/(2e^2/\hbar) = \eta^2 |\widetilde{G}_{fi}(E+i\eta)|^2 = \sum_n \frac{|\mathcal{O}_n|^2}{1+\Delta_n^2} + 2\sum_{n,m>n} \mathcal{O}_n \mathcal{O}_m^* \frac{1+\Delta_n\Delta_m}{(1+\Delta_n^2)(1+\Delta_m^2)}, \tag{20}$$

where $\mathcal{O}_n = \langle f|n\rangle\langle n|i\rangle$ and $\Delta_n = (E-\epsilon_n)/\eta$. One can distinguish the individual contribution to the conductance of each state $|n\rangle$ [first summation term on the right-hand side of Eq. (20)] and a ''mixing'' term (second summation term). From this equation, we can analyze the effect of changing first the number of states ($n_{\max}$) included in the summations and, second, the value of the energy imaginary part $\eta$.

Figure 1 shows the values of $\sigma_N(E,x)$ for different values of $n_{\max}$ versus the position $x_f$ of the center of the final state. For a tunneling energy $E$ that is close to an eigenvalue $\epsilon_n$,

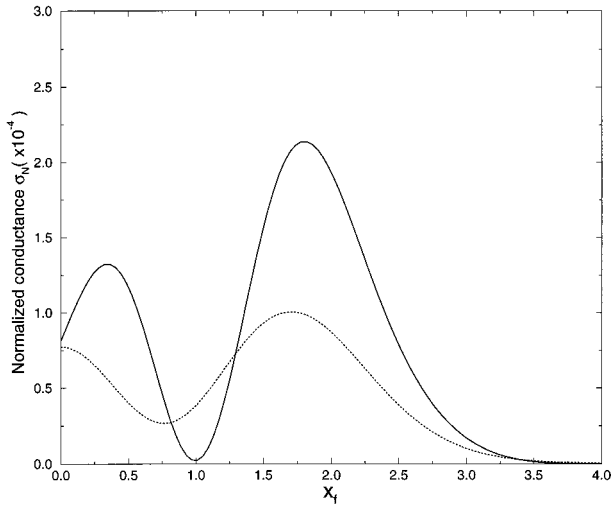



FIG. 2. Normalized conductance vs $x_f$ (solid line) calculated for $n_{\max}=10$. The individual contribution (dotted line) of each state [first summation term in Eq. (20)] is not the most important quantity contributing to $\sigma_N$ when the tunneling energy $E$ is between two eigenvalues, for instance, $E/\hbar\omega=3.0$. The values of the other parameters are the same as for Fig. 1.

the dependence of the conductance on $n_{\max}$ is negligible, provided that the states close to the tunneling energy are included in the summation. In this case, the individual contribution of the eigenstate $|n=2\rangle$ to the conductance is the most important. For a tunneling energy between two different eigenvalues, the individual contributions of the corresponding eigenstates to the conductance are not the most important terms and the ''mixing'' term has a non-negligible contribution (Fig. 2).

The global shape of the conductance is not drastically affected for small values of $\eta$ (Fig. 3). For large values, the spreading of the Lorentzians $(1+\Delta_n^2)^{-1}$ becomes too large, and all eigenstates contribute almost equally to the conductance; the energy selectivity of the relevant states for the tunneling is then lost. Furthermore the conductance becomes strongly dependent on $n_{\max}$. This essential physics will also hold for cases in which the eigenstates $|n\rangle$ are represented by much more complex wave functions, such as the one for a surface expanded in augmented plane waves in Sec. IV.

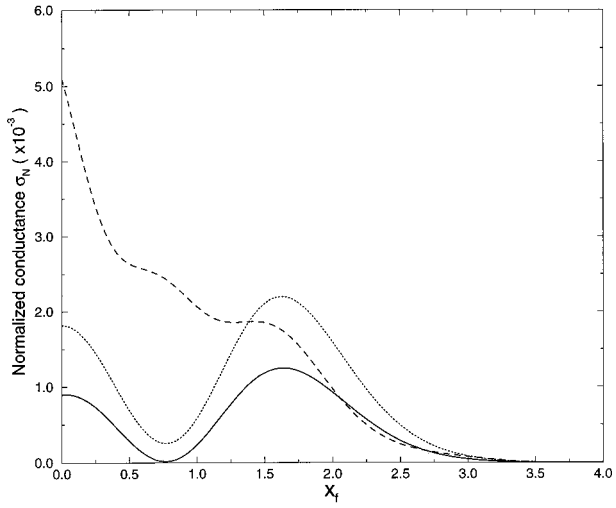



FIG. 3. Influence of $\eta$ on the normalized conductance $\sigma_N$ vs $x_f$; $\eta=0.10$ (solid line), $\eta=0.30$ (dotted line), and $\eta=1.1$ (dashed line). The values of the other parameters are identical to those used for Fig. 2.

## III. CONNECTION TO EXISTING THEORIES

### A. Tersoff-Hamann formula

It is interesting to notice how Tersoff-Hamann theory[13,14] can be obtained in the present context. As already mentioned in Sec. I, the Tersoff-Hamman model is based on first-order perturbation theory using the unperturbed sample surface wave functions, and a spherical symmetry for the unperturbed tip wave function. Let us neglect the effect of the tip potential, so that the eigenstates $|n\rangle$ appearing in Eqs. (10) and (19) are simply the eigenstates of the sample. In the limit of zero bias, we need to consider only the states close to the Fermi level $E_F$. In this condition, as shown in Sec. II D 1, and for an infinitesimally small value of $\eta$, the differential conductance can be written as [the ''mixing'' term in Eq. (20) is negligible for a tunneling energy close to an eigenvalue of the Hamiltonian]

$$\sigma_N(E)\approx\sum_n |\langle f|n\rangle|^2|\langle n|i\rangle|^2\delta(E_F-\epsilon_n). \tag{21}$$

Next, the initial state $|i\rangle$ is taken to be a state localized on or near the tip apex at some position $\mathbf{r}_0$. Then the scalar product $\langle n|i\rangle$ is a convolution between the surface states $\langle\mathbf{r}|n\rangle=\psi_n(\mathbf{r})$ and the tip wave function which closely resembles $\psi_n(\mathbf{r}_0)$ if the spatial extent of $\langle\mathbf{r}|i\rangle$ is not too large compared to the features of $\psi_n(\mathbf{r})$ [for the extreme case in which the initial state is a $\delta$ function on the tip, $\langle n|i\rangle$ is simply equal to $\psi_n(\mathbf{r}_0)$]. The exact analytical calculation of $\langle n|i\rangle$ is only possible by assuming a particular form for the surface wave function.

In a perturbative approach such as Tersoff-Hamann, all states at [or close to, for a broadened $\delta(E_F-\epsilon_n)$ function] the Fermi energy participate in tunneling with a weighting that is determined only by the matrix element of the transfer Hamiltonian. In our formalism, this implies choosing the same weight for each $|n\rangle$ state, disregarding the overlap with the final $|f\rangle$ state, i.e., the scalar product $\langle f|n\rangle$ is taken to be a constant. In these conditions, one recovers the result of Tersoff and Hamann, i.e., the differential conductance $\sigma_N(E_F)$ is proportional to the surface local density of states $\rho_S(\mathbf{r}_0,E_F)=\Sigma_n|\psi_n(\mathbf{r}_0)|^2\delta(E_F-\epsilon_n)$.

It is interesting that from this point of view, the major effects disregarded in the Tersoff-Hamann approach but included in Eq. (19) are the coherence between different eigenstates participating in the transport, and the weighting of the different eigenstates according to their ability to carry current between the initial and final states.

It should also be possible to derive an equivalent expression to the Tersoff-Hamann result by another route. This consists in partitioning the system into two subsystems (the unperturbed tip and sample). Then the full Green's function $\tilde{G}(E)$ of the system could be found from the unperturbed tip $\tilde{G}_T$ and sample $\tilde{G}_S$ Green's functions by matching these Green's functions on the surface separating the two sub-

systems. Such an approach would essentially duplicate Noguera's discussion[37] of the validity of the transfer Hamiltonian approximation, so we do not pursue it here.

### B. Scattering theory

In this section, we establish a connection between the formalism described in Sec. II B and the standard formulation of the current in scattering theory obtained from the generalized Ehrenfest theorem (for example, see Ref. 57). We consider the case of time-independent scattering. We also discuss the separation of the total Hamiltonian into the noninteracting system, and the so-called perturbation potential used in scattering theory. It should be noticed at this point that our formalism is not based on this separation and deals directly with the total Hamiltonian of the system.

We do not rederive the theory of scattering which can be found in several textbooks. Instead we start directly from the formalism described in Secs. 36 and 37 of Ref. 58. The noninteracting system, consisting of the left and right sides of the STM tunneling junction, is described by the Hamiltonian $\hat{H}_0$. The basic aim of scattering theory is to express the transition probability from eigenstates of $\hat{H}_0$ on one side to the other side of the junction under the action of a perturbation (or scattering) potential $\hat{V}$. The precise definition of such a potential in the case of a STM junction is discussed in more detail below. From the time-dependent point of view, the scattering occurs as follows: an initial $\phi_\alpha(\mathbf{r},t)$ eigenstate of $\hat{H}_0$ is scattered by $\hat{V}$ into a scattering state $\psi_\alpha^+(\mathbf{r}',t')$. The relation between these two states can be obtained via the general form of the space-time propagator (or Green's function) $G^+(\mathbf{r}',t';\mathbf{r},t)$ of the total Hamiltonian $\hat{H}=\hat{H}_0+\hat{V}$:

$$\psi_\alpha^+(\mathbf{r}',t')=i\int G^+(\mathbf{r}',t';\mathbf{r},t)\,\phi_\alpha(\mathbf{r},t)\,d^3\mathbf{r}. \tag{22}$$

It should be noted that, in order to obtain transitions between eigenstates of $\hat{H}_0$, the pertubation potential should be initially absent, then switched on, and off again after the scattering of the state concerned has been completed. Otherwise, if the perturbation potential is always present, the system does not ''evolve;'' it just stays in an eigenstate of the total Hamiltonian $\hat{H}$. From Eq. (22), a Lippman-Schwinger equation for the scattering state $\psi_\alpha^+$ can be obtained from the Dyson equation of the $G^+$ Green's function, and the integral properties of such a Green's function.

In the stationary case, the wave functions are expressed as $\phi_\alpha(\mathbf{r},t)=u_\alpha(\mathbf{r})e^{-\mathrm{i}E_\alpha t/\hbar}$, $|u_\alpha\rangle$ being the eigenfunction of the time-independent Hamiltonian $\hat{H}_0$ and $\psi_\alpha^+(\mathbf{r},t)=u_\alpha^+(\mathbf{r})e^{-\mathrm{i}E_\alpha t/\hbar}$. The same exponential factor occurs in the expressions for $\phi_\alpha$ and $\psi_\alpha^+$, because the energy is conserved in elastic scattering. This can been seen in more detail in Ref. 54 [for example, Eq. (1.71) in this reference]. Then it can be shown that the matrix elements of the propagator or Green's function are written as follows:

$$e^{iE_\beta t'/\hbar}\langle u_\beta|i\hat{G}^+(t',t)|u_\alpha\rangle e^{iE_\alpha t/\hbar} = \delta_{\alpha\beta}-\frac{i}{\hbar}\langle u_\beta|\hat{T}|u_\alpha\rangle\int g(t_1)e^{i(E_\beta-E_\alpha)t_1/\hbar}dt_1, \tag{23}$$

where $\hat{G}^+(t',t)=-(i/\hbar)\,\theta(t'-t)\exp[-i\hat{H}(t'-t)/\hbar]$ is the causal Green's function obtained from the total Hamiltonian, $g(t)$ is the time dependence function of the perturbation potential $\hat{V}(\mathbf{r},t)=\hat{V}(\mathbf{r})g(t)$, and the $T$-matrix elements are expressed in the common form

$$\langle u_\beta|\hat{T}|u_\alpha\rangle=\langle u_\beta|\hat{V}|u_\alpha^+\rangle=\int d^3\mathbf{r}\;u_\beta^*(\mathbf{r})V(\mathbf{r})u_\alpha^+(\mathbf{r}). \tag{24}$$

We now choose the form of the time dependence of the perturbation potential, $g(t)$; let it be unity during a period $\Delta T$ and go to zero outside this period (i.e., when $t\to\pm\infty$) over a time interval $\Delta t$ [cf. Fig. (34) in Ref. 58]. Then the Fourier transform of $g(t)$ is not strictly a $\delta$ function but merely a function $\tilde{g}(E_\alpha-E_\beta)$ strongly peaked around $E_\alpha-E_\beta=0$ in an energy range of approximately $\hbar/\Delta t$. $\Delta t$ can be chosen as large as needed to reduce arbitrarily the energy interval in which $\tilde{g}$ is nonzero.

Now the transition probability defined by the left-hand side of Eq. (23),

$$P(t'-t)=|\langle u_\beta|\hat{G}^+|u_\alpha\rangle|^2, \tag{25}$$

is directly proportional to the square modulus of the $T$-matrix elements when the initial state $|u_\alpha\rangle$ is orthogonal to the final state $|u_\beta\rangle$.

If we now apply the same time-averaging formalism as developed in Sec. II B to Eq. (25), using Eq. (23) for the Green's function, we obtain the following equivalence for the time-averaged tunneling current:

$$\overline{J}=e\frac{\overline{P(\tau)}}{\tau}\equiv\frac{2\pi e}{\hbar}\,|\langle u_\beta|\hat{T}|u_\alpha\rangle|^2\,\delta(E_\alpha-E_\beta), \tag{26}$$

since the $T$-matrix elements are independent of time. Note that, for convenience, we replaced the $\tilde{g}$ function by the equivalent $\delta$ function when $E_\alpha=E_\beta$.[59] Then Eq. (26) shows the equivalence between our formalism and the expression of the tunneling current obtained from the generalized Ehrenfest theorem. This is a ''strict'' equivalence when and only when the averaged transition rate $P(\tau)$ is defined from the eigenstates of the Hamiltonian $\hat{H}_0$ of the noninteracting system.

Note again that, in the present formalism, we deal directly with the total Hamiltonian $\hat{H}$; there is no need to know explicitly the form of the perturbation potential $\hat{V}$. $\hat{H}_0$ has to be known only to determine its corresponding initial and final eigenstates.

Now it is essential to discuss the definition of $\hat{H}_0$ and the meaning of $\hat{V}$ for a practical application to scanning tunneling microscopy. In a realistic description of the tunneling junction, $\hat{H}_0$ should ideally represent the Hamiltonian of the two separate noninteracting sides of the junction. These can be the sample side and the tip side. Then the total Hamiltonian should include all the effects when the tip is brought close to the sample surface, for example the effects of the tip-induced electric field, the possible chemical bonding between the tip atoms and surface atoms, the possible modifications of the atomic structure of the tip and surface, etc. These effects are quite subtle, and the definition of $\hat{V}$ is not

straightforward.[51] Hence the present formalism using only the full Hamiltonian seems to be a convenient method for treating such a problem.

However the calculation of the corresponding eigenstates of $\hat{H}_0$ is a very hard problem in *ab initio* (plane-wave-like) methods, since both parts of the subsystems are represented by semi-infinite systems. To our knowledge, calculations of the tunneling current for such cases are only possible by introducing some approximations for the asymptotic propagating states. This has been done by using a tight-binding-like basis set[31,42] or using an analytical form for the asymptotic propagating states which assumes there is a jellium on one[41] or both sides[60–62] of the tunneling junction.

Therefore it is convenient to introduce a model description of the tunneling junction to perform calculations of the current from *ab initio* (plane-wave-like) methods. This can be done by representing the surface by a slab, and the tip by an isolated cluster or a cluster deposited on another slab. Then our definition of the current $J$ must be seen as an averaged electron transition rate from one slab state to another slab state ''located'' on the other side of the tunneling junction.

Note also that, when the transition rate is calculated via our originally derived Green's-function matrix elements $\tilde{G}_{fi}(E)$ (i.e., when the states $|u_\alpha\rangle$ and $|u_\beta\rangle$, rather than being eigenstates of $\tilde{H}_0$, are substituted by arbitrary localized initial $|i\rangle$ and final $|f\rangle$ states on both sides of the tunneling junction), the current is not exactly determined. In this case, the calculation should be seen as a ''sampling'' of the exact $T$ matrix.

Finally, there is still a possibility of matching of the wave functions of the slab states onto the leads (the remaining of the semi-infinite part for a realistic description of the tunneling junction). For such a purpose, the formalism used by Weir and Wingreen to determine a generalized Landauer formula, briefly described in Sec. III C, can be helpful.

### C. Generalized Landauer formula of Meir and Wingreen

Now we come back to the Landauer formulation of the tunneling current, as generally described by Eq. (1). Meir and Wingreen[50] derived a generalized Landauer formula for the tunneling current flowing between two leads through an intermediate region in which the electrons may interact. In the case of independent (or noninteracting) electrons, the current flowing from the left to the right ($\mu_L > \mu_R$), is expressed as

$$J = \frac{e}{h} \int d\epsilon [f_L(\epsilon) - f_R(\epsilon)] \mathrm{Tr}[\hat{G}^a \hat{\Gamma}^R \hat{G}^r \hat{\Gamma}^L], \qquad (27)$$

where $f_{L,R}$ are the unperturbed Fermi-Dirac distribution functions of the leads, and $\hat{G}^{r,a}$ are, respectively, the retarded and advanced Green's functions of the complete intermediate region. The $\hat{\Gamma}$ operators represent the coupling between the intermediate region and the leads. For example, $\hat{\Gamma}^L$ for the left side is $\Gamma^L_{n,m} = 2\pi \Sigma_{\alpha \in L} \rho_\alpha(\epsilon) V_{\alpha,n}(\epsilon) V^*_{\alpha,m}(\epsilon)$, where $V_{\alpha,n}$ is the potential coupling the incoming $\alpha$ channel (with density of states $\rho_\alpha$) to the $n$th single-particle state of the intermediate region. By noticing that the transmission coefficient $t_{\beta,\alpha}$ from the left (channel $\alpha$) to the right (channel $\beta$) is given by $t_{\beta,\alpha} \equiv 2\pi \Sigma_{n,m} V^*_{\beta,n} G^r_{n,m} V_{\alpha,m}$, it can be shown[50] that Eq. (27) reduces to Eq. (1). In fact, under the conditions of noninteracting electrons, zero temperature, and a finite bias between the two electrodes, Eq. (27) is strictly equivalent to the expression for the tunnel current discussed by Caroli *et al.*[63] In all these treatments, the differential conductance is ultimately proportional to a product of two one-particle Green's functions, but also involves quantities describing the leads connecting the system to the reservoirs.[50,52,63,64]

In our case, the coupling between the so-called intermediate region and these leads is not treated exactly. We consider the propagation of an electron introduced (somehow) into the $|i\rangle$ state, and removed (somehow) from the state $|f\rangle$. This means that we approximate the expressions for the $\hat{\Gamma}$ operators $\hat{\Gamma} = 2\pi \Sigma_\alpha \hat{V}|\alpha\rangle \rho_\alpha \langle\alpha|\hat{V}^\dagger$ (where $\hat{\Gamma} = \hat{\Gamma}^{L,R}$ if $\alpha$ indexes a channel in the left or right lead, respectively).

For example, if we consider only one incoming channel with a normalized density of states, we suppose that the effect of $\hat{V}$ on the incoming state is simply to transform $|\alpha\rangle$ into $|i\rangle$, i.e., $\hat{V}|\alpha\rangle \propto |i\rangle$. The $\hat{\Gamma}^L$ operator takes the form of a projector onto the $|i\rangle$ states; similarly, $\hat{\Gamma}^R$ is a projector onto the $|f\rangle$ states). This expression can be written as $\hat{\Gamma}^L = \Sigma_{\alpha \in L} |i\rangle |V_{i\alpha}(\epsilon)|^2 \langle i| \delta(\epsilon - \epsilon_\alpha)$. Then, if we take $V_{i\alpha}$ to be independent of the energy, Eq. (27) becomes proportional to our expression for the tunneling current $J = eP(\tau)/\tau$.

This has the consequence, as already mentioned in Sec. II B, that the absolute values of our tunneling current are not correct because they do not include the effects of the spreading resistance due to the coupling with the leads. Nevertheless the essential physics of the tunneling processes inside the intermediate region is correctly described, since we use the full Green's functions $\hat{G}^r$ and $\hat{G}^a$ for the intermediate region in the expression of the tunneling current; these Green's functions are obtained, for example, within the density-functional-theory formalism. A more accurate description of the coupling $\Gamma$ operators would be possible by approximating the asymptotic propagating states of the leads by, for example, a coupled chain of Wannier functions.

## IV. APPLICATION

In this section, we present an application of the technique for a specific surface. Calculations of the STM image for the bare graphite surface, using the present technique, have been published elsewhere.[46] The surface we propose to study here is more complex than the bare graphite surface because it is heterogeneous. It is a silicon surface on which small organic molecules are deposited.[65] Such systems are interesting for understanding the fundamental adsorption mechanisms as well as for applications in film growth. In this paper, we consider ethene ($C_2H_4$) molecules adsorbed on the Si(001) surface.[66,67]

In the STM images presented below, the presence of the tip has been approximated by particular tip states and simple potentials. This simplification is not due to any restriction in principle on the method, but should be seen as a first stage of calculation. This, of course, should be improved and completed by further calculations including a more realistic atomic description of the tip.

The method used to determine the ground state of the system is the local-density-functional based[68,69] projector-augmented wave (PAW) method developed by Blöchl.[70] The main features of the PAW method are as follows: (i) it is an all-electron method which includes the full valence-electron wave functions, representing them by augmented plane-waves; (ii) it employs a generalization of the pseudopotential concept using (localized) projectors to represent the nonlocal part of the atomic potentials;[71] and (iii) Car-Parrinello-like Lagrangian[16,72] dynamics is implemented in the PAW framework and permits us to determine simultaneously the electronic and atomic ground-state configurations of the systems considered.

Experimental images have been obtained for different coverages of $C_2H_4$ adsorbed on Si(001), both at positive and negative sample biases. These results can be found in, for example, Refs. 66, 67, and 73. It is found that, in the images at negative sample bias (the current flows then from the sample to the tip), the isolated molecules adsorbed on the surface appear slightly darker than the bare dimers of the clean parts of the surface. Preliminary calculation of constant current scans have shown that it is necessary to take explicit account of the tip-induced electric field in the calculations in order to obtain, even qualitatively, the correct STM contrast.[73] This result can be understood in terms of the behavior, under the presence of an external electric field, of electronic states from a surface presenting different parts characterized by different polarizabilities.[74]

The details of the calculations of the structure[19] and of the surface polarization effects due to our model tip-induced electric field[75] have been already described elsewhere. We concentrate our discussion here on the influence of the choice for the initial states $|i\rangle$ on the tunneling current and more especially on the STM contrast. As already mentioned in Sec. II C, owing to the approximations made in the coupling of the tunneling junction to the leads, different choices for the $|i\rangle$ and $|f\rangle$ states are possible. Here we choose two alternatives for the initial (sample) states: (i) a set of three-dimensional Gaussian states localized randomly in the surface, and (ii) a set of Kohn-Sham eigenstates obtained from the PAW method for the same conditions of calculation as in Ref. 75 for the bare Si(001) surface. As mentioned at the beginning of this section, the final (tip) state is an approximate one, chosen for computational convenience. It is chosen to be a localized three-dimensional Gaussian state. This state moves in the vacuum space above the surface as the tip scans this surface. This means that the tip potential is only taken partially into account in the calculations via the tip-induced electric field.

The differential conductance is calculated for a set of $M$ different energies $E_\alpha$ close to the top of the valence band. Then the conductance values are summed over the corresponding energy window $\Delta E=\mu_i-\mu_f=eV$ to give an approximation for the integral defined by Eq. (18): $J=V\sigma_{\mathrm{av}}=V\Sigma_\alpha\sigma(E_\alpha)/M$. The precise value of the $\Delta E$ appropriate to the experiments is somewhat uncertain, since the exact value of the bias across the tunneling junction is strongly affected by the screening and long-ranged band bending below the surface. However, no strong modifications of the experimental contrast were observed for different negative bias voltages. This suggests that the choice of $\Delta E$ is not a critical parameter in the present case. We take an integration energy window $\Delta E=0.9$ eV, as this range includes all the occupied states which are strongly localized near the surface. Furthermore this choice seems to be reasonable since the band gap of silicon is $\approx 1.1$ eV and the Si substrate in the experiments is $n$ type; the Fermi level in the bulk is therefore just below the bottom of the conduction band. The experimental images were taken for a negative sample bias of 2 V; neglecting band-bending effects, this corresponds to tunneling in the top 0.9 eV of the valence band.

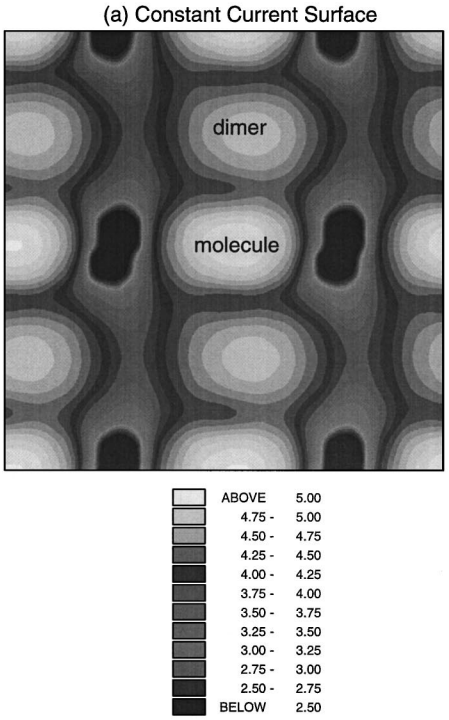
Constant-current images of the surface are shown on Figs. 4 and 5 for both choices of the initial states. It is found that a contrast in agreement with the experiments (i.e., the molecules appear darker than the bare dimers) can only be obtained with the presence of the tip-induced electric field. The most important result regarding the use of the method is that the STM contrast of the adsorbed molecules on the surface is not strongly or qualitatively dependent on the two different choices we made for the initial states. In other words, in the present case the approximation made in the coupling of the tunneling junction to the leads does not play an important role on the STM contrast. Instead, the surface polarization effects due to the tip-induced electric field are mainly responsible for the observed contrast. For a more complete understanding of the tip-sample interactions, further calculations should be performed with the presence of a more realistic description of the tip states and potential.

## V. CONCLUSIONS

We have presented a method which allows electron tunneling in a model STM junction to be tackled in a nonperturbative manner. The calculation follows the progress of an electron which is injected into a state on one side of the tunnel junction, and adopts as a measure of the tunnel current the time-averaged rate of transfer to a second state localized on the other side. The method can be used directly with the self-consistent potential and wave functions obtained by total-energy calculations or *ab initio* molecular dynamics.

We have shown that the electron transfer rate can be calculated in a manner which does not involve the unacceptable computational expense of obtaining the full spectrum of single-particle eigenfunctions for the coupled tip-sample system. We have presented some applications of the technique here. The STM contrast calculated for adsorbed ethene ($C_2H_4$) molecules on the Si(001) surface appears to be in agreement with the corresponding experimental images.

These preliminary results are encouraging, but the present version of the technique needs to be developed and improved to include a better representation of the tip states and potential. This can be done, for example, by including in the *ab initio* calculation an atomic cluster tip close to the surface, as already done by other authors.[76–79] Finally, in order to obtain exact absolute values of the tunneling current, the coupling of the tunnel junction with the leads carrying the current on both sides of the junction should be treated exactly. Such a coupling appears to be possible with the help of real space embedding potentials[80,81] and tricks used in the so-called $O(N)$ electronic structure methods.[82]



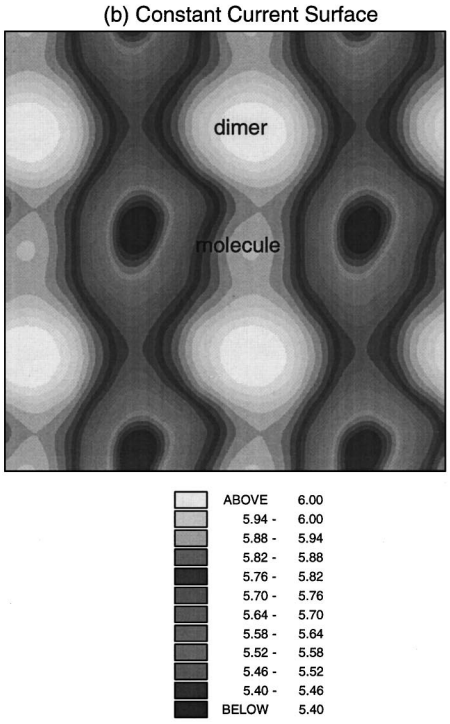
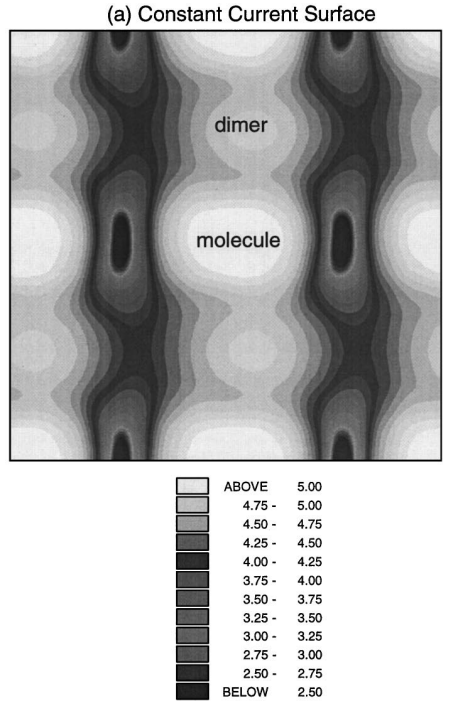




FIG. 4. Grey-scaled constant-current image above the surface unit cell. The corresponding conductance value is $5 \times 10^{-5}\,[2e^2/h]$. The conductance is calculated using a set of random localized Gaussian initial states. The ''tip-sample'' distance is given in Å, and corresponds to the position of the center of the Gaussian final state. (a) Without the tip-induced electric field, the molecules appear brighter than the bare dimers. (b) In the presence of the tip-induced electric field, the contrast is inverted, and the dimers appear brighter than the molecules, in qualitative agreement with the experiences.



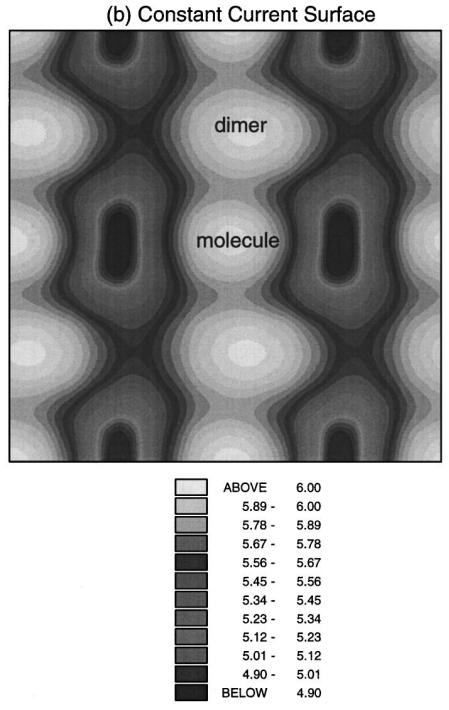




FIG. 5. Grey-scaled constant-current image above the surface unit cell. The corresponding conductance value is $5 \times 10^{-5}\,[2e^2/h]$. The conductance is calculated using a set of initial states coming from a slab calculation of the bare Si(001) surface. The ''tip-sample'' distance is given in Å, and corresponds to the position of the center of the Gaussian final state. (a) Without the tip-induced electric field, the molecules appear still brighter than the bare dimers. (b) With the tip-induced electric field, the contrast is inverted and the molecules appear slightly darker than the bare dimers. Such a contrast is in better agreement with the experiences, especially when one considers the dimer rows. However, the contrast is only slightly affected by the choice of the initial states (compare to Fig. 4).

## ACKNOWLEDGMENTS

We are grateful for discussions with Dr. P. E. Blöchl and Dr. G. A. D. Briggs. We acknowledge financial support from the U. K. Engineering and Physical Sciences Research Council (A.J.F.) and support from Grant No. GR/K80495.


## APPENDIX: ITERATIVE CALCULATION OF THE GREEN'S-FUNCTION MATRIX ELEMENTS

We can avoid the calculation of the eigenvalue spectrum altogether by a trick similar to those used to avoid sums over intermediate states in linear response theory in atoms[83] and solids.[84] It can be seen that the Green's function in question can be written as

$$\tilde{G}_{fi}(E+i\eta)=\sum_n \frac{\langle f|n\rangle\langle n|i\rangle}{E-\epsilon_n+i\eta}=\langle f|(E+i\eta-\hat{H})^{-1}|i\rangle. \tag{A1}$$

This in turn can be written

$$\tilde{G}_{fi}(E+i\eta)=\langle f|\tilde{i}\rangle, \tag{A2}$$

where the state $|\tilde{i}\rangle$ is the solution of the linear equation

$$(E+i\eta-\hat{H})|\tilde{i}\rangle=|i\rangle. \tag{A3}$$

This equation may be efficiently solved by an iterative method (such as the recursion method) with, in general, many fewer operations than are necessary to find the complete spectrum of the Hamiltonian $\hat{H}$. In the present situation, a particularly useful iterative scheme is provided by the Lanczos algorithm.[85–88] It is convenient for our purposes because it involves generating a tridiagonalization form for the Hamiltonian. Once this has been found, it is straightforward to compute the solution $|\tilde{i}\rangle$ for any desired values of $E$ relevant for the tunneling.

The calculation proceeds as follows. The tridiagonalization is generated by a sequence of recursion vectors $|u_i\rangle$ as usual.[86] After $N_L$ Lanczos iterations, the Hamiltonian in the basis of the Lanczos vectors takes the standard tridiagonal form $H_{ii}=\langle u_i|\hat{H}|u_i\rangle=\alpha_i$ and $H_{i,i+1}=H_{i+1,i}=\beta_i$. The space spanned by the vectors $|u_i\rangle$ (for $i=1,\ldots,N_L$) is known as the $N_L$th Krylov space.

The solution of the linear equation (A3) is also the solution to the problem of finding the stationary values of the quadratic from

$$f(|\tilde{i}\rangle)\equiv\langle\tilde{i}|(E+i\eta-\hat{H})|\tilde{i}\rangle-\langle i|\tilde{i}\rangle. \tag{A4}$$

If $|\tilde{i}_m\rangle$ denotes the stationary point of $f$ in the $m$th Krylov space, then the solutions in successive Krylov spaces can be constructed iteratively.[85] The construction of the successive Krylov spaces guarantees that the sequence of $|\tilde{i}_m\rangle$ converges to the true stationary point $|\tilde{i}\rangle$ of $f$. It can be shown[85] that the set $\{|\tilde{i}_m\rangle\}$ satisfies the iterative relation

$$|\tilde{i}_m\rangle=|\tilde{i}_{m-1}\rangle+\rho_m|c_m\rangle, \tag{A5}$$

with

$$|c_m\rangle=|u_m\rangle-\mu_{m-1}|c_{m-1}\rangle,$$

$$\rho_m=-\frac{\mu_{m-1}d_{m-1}\rho_{m-1}}{d_m},$$

$$d_m=E+i\eta-\alpha_m+\beta_{m-1}\mu_{m-1},$$

$$\mu_{m-1}=-\frac{\beta_{m-1}}{d_{m-1}}, \tag{A6}$$

and the initial conditions

$$|\tilde{i}_1\rangle=\rho_1|u_1\rangle,$$

$$|c_1\rangle=|u_1\rangle,$$

$$\rho_1=\frac{1}{d_1}=\frac{1}{E+i\eta-\alpha_1}. \tag{A7}$$

One final simplification arises because we are not concerned with the state $|\tilde{i}\rangle$ for its own sake, but only with the overlap $\langle f|\tilde{i}\rangle$. We can therefore replace the vector manipulations in Eqs. (A5) and (A6) by the simpler scalar formulas involving the elements $\langle f|\tilde{i}_m\rangle$, $\langle f|c_m\rangle$, and $\langle f|u_m\rangle$. Once the chain of the $N_L$ Lanczos vectors has been constructed, this enables $\langle f|\tilde{i}\rangle$ to be found for each energy with only $O(N_L)$ operations.

Some STM images have already been calculated using an iterative method such as that used to determine the Green's-function elements. They have been obtained for a model tip potential.[46] More work needs to be done to take account of a more realistic tip modeling. Note that in such iterative calculations the convergence of the results with the number of iteration steps $N_L$ must be carefully examined, as must the asymptotic values of the $\alpha_i$ and $\beta_i$ recursion coefficients. In some cases, the loss of orthogonality between the Lanczos vectors $|u_i\rangle$, due to the use of finite-precision arithmetic, can cause problems with the convergence of the algorithm and lead to ''multiple copies'' of eigenvalues in the spectrum; however, it is still possible to impose orthogonality after each iteration step,[89] with some loss of efficiency. This method is therefore promising, and computationally efficient for large and complex tip-sample systems; work on optimizing it is still in progress.


[1] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982).

[2] G. Binnig and H. Rohrer, Rev. Mod. Phys. **59**, 615 (1987).

[3] P. Avouris, I.-W. Lyo, and F. Bozso, J. Vac. Sci. Technol. B **9**, 424 (1991).

[4] H. Ohtani, R. J. Wilson, S. Chiang, and C. M. Mate, Phys. Rev. Lett. **60**, 2398 (1988).

[5] D. P. E. Smith, H. Hörber, Ch. Gerber, and G. Binnig, Science **245**, 43 (1989).

[6] M. Hara, Y. Iwakabe, K. Tochigi, H. Sasabe, A. F. Garito, and A.

Yamada, Nature (London) **344**, 228 (1990).
[7] W. M. Heckl and D. P. E. Smith, J. Vac. Sci. Technol. B **9**, 1159 (1991).
[8] G. Binnig and H. Rohrer, in *Trends in Physics*, edited by J. Janta and J. Pantoflicek (European Physical Society, The Hague, 1984), pp. 38–46.
[9] M. G. Youngquist, R. J. Driscoll, T. R. Coley, W. A. Goddard, and J. D. Baldeschwieler, J. Vac. Sci. Technol. B **9**, 1304 (1991).
[10] B. Michel, G. Travaglini, H. Rohrer, C. Joachim, and M. Amrein, Z. Phys. B **76**, 99 (1989).
[11] G. Travaglini, M. Amrein, B. Michel, and H. Gross, in *Scanning Tunneling Microscopy and Related Methods*, Vol. 184 of *NATO Advanced Study Institute, Series E: Applied Sciences*, edited by R. J. Behm *et al.* (Kluwer, Dordrecht, 1990), p. 335.
[12] J. Tersoff, in *Scanning Tunneling Microscopy and Related Methods* (Ref. 11), p. 77.
[13] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1985).
[14] J. Tersoff and D. R. Hamann, Phys. Rev. B **31**, 805 (1985).
[15] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
[16] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[17] I. Štich, M. C. Payne, R. D. Kingsmith, J. S. Lin, and L. J. Clarke, Phys. Rev. Lett. **68**, 1351 (1992).
[18] K. D. Brommer, M. Needels, B. E. Larson, and J. D. Joannopoulos, Phys. Rev. Lett. **68**, 1355 (1992).
[19] A. J. Fisher, P. E. Blöchl, and G. A. D. Briggs, Surf. Sci. **374**, 298 (1997).
[20] A. Baratoff, Physica B **127B**, 143 (1984).
[21] D. Lawumni and M. C. Payne, J. Phys.: Condens. Matter **2**, 3811 (1990).
[22] J. Tersoff and N. D. Lang, Phys. Rev. Lett. **65**, 1132 (1990).
[23] J. Tersoff, Phys. Rev. B **41**, 1235 (1990).
[24] C. J. Chen, Phys. Rev. Lett. **65**, 448 (1990); Phys. Rev. B **42**, 8841 (1990); Phys. Rev. Lett. **69**, 1656 (1992).
[25] J. H. Wilson, D. A. McInnes, J. Knall, A. P. Sutton, and J. B. Pethica, Ultramicroscopy **42–44**, 801 (1992).
[26] M. Tsukada, K. Kobayashi, and S. Ohnishi, J. Vac. Sci. Technol. A **8**, 160 (1990).
[27] M. Tsukada, K. Kobayashi, and N. Isshiki, Surf. Sci. **242**, 12 (1991).
[28] M. Tsukada, K. Kobayashi, N. Shima, and S. Ohnishi, J. Vac. Sci. Technol. B **9**, 494 (1991).
[29] J. Gimzewski and R. Möller, Phys. Rev. B **36**, 1284 (1987).
[30] N. D. Lang, Phys. Rev. B **36**, 8173 (1987).
[31] T. N. Todorov and A. P. Sutton, Phys. Rev. Lett. **70**, 2138 (1993).
[32] A. G. Scherbakov, E. N. Bogachek, and U. Landman, Phys. Rev. B **53**, 4054 (1996).
[33] L. Olesen, E. Laegsgaard, I. Stengaard, F. Besenbacher, J. Schiotz, P. Stoltze, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett. **72**, 2251 (1994).
[34] C. J. Muller, J. M. van Ruitenbeek, and L. J. de Jongh, Phys. Rev. Lett. **69**, 140 (1992).
[35] V. Kalmeyer and R. B. Laughlin, Phys. Rev. B **35**, 9805 (1987).
[36] A. A. Lucas, H. Morawitz, G. R. Henry, J. P. Vigneron, P. Lambin, P. H. Cutler, and T. E. Feuchtwang, Phys. Rev. B **37**, 10 708 (1988).
[37] C. Noguera, J. Phys. (France) **50**, 2587 (1990).
[38] C. Noguera, Phys. Rev. B **42**, 1629 (1990).
[39] C. Noguera, in *Scanning Tunneling Microscopy III*, Vol. 29 of *Springer Series in Surface Science*, 2nd ed., edited by R. Wiesendanger and H. J. Güntherodt (Springer-Verlag, Berlin, 1996), p. 51.
[40] W. Sacks and C. Noguera, Phys. Rev. B **43**, 11 612 (1991).
[41] G. Doyen, D. Drakova, and M. Scheffler, Phys. Rev. B **47**, 9778 (1993); G. Doyen, in *Scanning Tunneling Microscopy III* (Ref. 39), p. 23.
[42] C. Joachim and P. Sautet, in *Scanning Tunneling Microscopy and Related Methods* (Ref. 11), p. 377.
[43] E. Tekman and S. Ciraci, Phys. Rev. B **43**, 7145 (1991).
[44] P. Sautet and C. Joachim, Chem. Phys. Lett. **23**, 23 (1991).
[45] S. Ciraci, E. Tekman, A. Baratoff, and I. P. Batra, Phys. Rev. B **46**, 10 411 (1992).
[46] A. J. Fisher and P. E. Blöchl, in *Computations for the Nano-Scale*, Vol. 240 of *NATO Advanced Study Institute, Series E: Applied Sciences*, edited by P. E. Blöchl *et al.* (Kluwer, Dordrecht, 1993), p. 185.
[47] R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).
[48] A. D. Stone and A. Szafer, IBM J. Res. Dev. **32**, 384 (1988).
[49] M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).
[50] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).
[51] J. B. Pendry, A. B. Prêtre, and B. C. H. Krutzen, J. Phys.: Condens. Matter **3**, 4313 (1991).
[52] T. N. Todorov, G. A. D. Briggs, and A. P. Sutton, J. Phys.: Condens. Matter **5**, 2389 (1993).
[53] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).
[54] B. A. Lippmann and J. Schwinger, Phys. Rev. **50**, 469 (1950).
[55] J. R. Trail, D. M. Bird, S. Crampin, and R. James (unpublished).
[56] R. James and S. M. Woodley, Solid State Commun. **97**, 935 (1996).
[57] G. Doyen, J. Phys.: Condens. Matter **5**, 3305 (1993).
[58] L. I. Schiff, in *Quantum Mechanics*, 3rd ed. (McGraw-Hill, London, 1968), Sec. 36, p. 298; Sec. 37, p. 312.
[59] Strictly speaking, the right-hand side of Eq. (26) should contain $|\tilde{g}(E_\alpha - E_\beta)|^2$. This can be seen from the square modulus of Eq. (23). For the form chosen for $g(t)$, its Fourier transform $\tilde{g}(E)$ is strongly peaked around $E=0$ in the energy interval $\hbar/\Delta t$ and $\tilde{g}(0)$ is proportional to $\Delta t$. Then, when $\Delta t$ is large enough, and $\Delta T$ being larger, we can approximate $|\tilde{g}|^2 \approx (\Delta t/2\pi\hbar)\tilde{g}(E_\alpha - E_\beta)$. If $\Delta t$ is related to the time-averaging factor $\tau$ (i.e., $\Delta t = \tau$) then we obtain the expression of Eq. (26) with the correct units.
[60] K. Hirose and M. Tsukada, Phys. Rev. Lett. **73**, 150 (1994).
[61] K. Hirose and M. Tsukada, Phys. Rev. B **51**, 5278 (1995).
[62] N. Kobayashi, K. Hirose, and M. Tsukada, Jpn. J. Appl. Phys. **35**, 3710 (1996).
[63] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C **4**, 916 (1971).
[64] R. Combescot, J. Phys. C **4**, 2611 (1971).
[65] M. J. Bozack, P. A. Taylor, W. J. Choyke, and J. T. Yates, Jr., Surf. Sci. Lett. **177**, L933 (1986); J. Yoshinobu, H. Tsuda, M. Onchi, and M. Nishijima, J. Chem. Phys. **87**, 7332 (1987); C. C. Cheng, W. J. Choyke, and J. T. Yates, Jr., Surf. Sci. **231**, 289 (1990).
[66] A. J. Mayne, A. R. Avery, J. Knall, T. S. Jones, G. A. D. Briggs, and W. H. Weinberg, Surf. Sci. **284**, 247 (1993).
[67] A. J. Mayne, Ph.D. thesis, University of Oxford, 1994.
[68] P. C. Hohenberg and W. Kohn, Phys. Rev. **136**, B684 (1964).
[69] W. Kohn and J. Sham, Phys. Rev. **140**, A1133 (1965).
[70] P. E. Blöchl, Phys. Rev. B **50**, 17 953 (1994).
[71] P. E. Blöchl, Phys. Rev. B **41**, 5414 (1990).
[72] R. Car and M. Parrinello, in *Simple Molecular Systems at Very*

*High Density*, edited by A. Polian *et al.* (Plenum, New York, 1988), p. 455.
[73] H. Ness, A. J. Fisher, and G. A. D. Briggs, Surf. Sci. **380**, L479 (1997).
[74] H. Ness and A. J. Fisher, J. Phys.: Condens. Matter **9**, 1793 (1997).
[75] H. Ness and A. J. Fisher, Phys. Rev. B **55**, 10 081 (1997).
[76] K. Cho and J. D. Joannopoulos, Phys. Rev. Lett. **71**, 1387 (1993).
[77] R. Perez, M. C. Payne, and A. D. Simpson, Phys. Rev. Lett. **75**, 4748 (1995).
[78] K. Cho and J. D. Joannopoulos, Jpn. J. Appl. Phys. **35**, 3714 (1996).
[79] J. C. Caulfield and A. J. Fisher, Appl. Phys. A: Solids Surf. (to be published February 1998).
[80] J. E. Inglesfield, J. Phys. C **14**, 3795 (1981).
[81] M. I. Trioni, G. P. Brivio, S. Crampin, and J. E. Inglesfield, Phys. Rev. B **53**, 8052 (1996).
[82] A good review of the so-called $O(N)$ methods can be found, for example, in G. Galli, Curr. Opinion Solid State Mater. Sci. **1**, 864 (1996).
[83] G. D. Mahan, Phys. Rev. A **22**, 1780 (1980).
[84] S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **1861**, 1861 (1987).
[85] G. H. Golub and C. F. Van Loan, in *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996), pp. 470–507.
[86] R. Haydock, Solid State Phys. **35**, 215 (1980).
[87] T. J. Godin and R. Haydock, Phys. Rev. B **38**, 5237 (1988).
[88] T. J. Godin and R. Haydock, Europhys. Lett. **14**, 137 (1991).
[89] B. N. Parlett, in *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980), p. 257.