Transfer Hamiltonian analytical theory of scanning tunneling spectroscopy

M. Passoni and C. E. Bottani
Center for NanoEngineered MAterials and Surfaces (NEMAS), CNISM, and Dipartimento di Ingegneria Nucleare, Politecnico di Milano, Via Ponzio 34/3, I-20133 Milan, Italy

(Received 5 March 2007; revised manuscript received 30 May 2007; published 7 September 2007)

The transfer Hamiltonian approach to the scanning tunneling spectroscopy (STS) is extended in a twofold direction. First, a theory representing an extension to the case of arbitrary temperature and applied voltage of the work of Chen [Phys. Rev. B **42**, 8841 (1990)] is developed. Within this framework analytical expressions of the tunneling current and its derivative can be obtained under rather general assumptions for the tip density of states. In particular, the situation of a general electronic structure of the tip states is considered. The calculation of theoretical dI/dV curves and conductivity maps, to be compared with experiments and numerical simulations, becomes possible and these results lead also to the best normalization procedure of the current derivative to obtain the desired physical information, namely, the sample local density of states, provided the electronic tip properties are known. Second, a general theoretical description in terms of the system spectral densities is derived, providing a generalization of the approach developed by Feuchtwang and Cutler [Phys. Scr. **38**, 252 (1988)]. We believe that with these achievements the forecast of STS theory gets significantly closer to the experimental results.

DOI: 10.1103/PhysRevB.76.115404 PACS number(s): 73.40.Gk, 73.20.At

I. INTRODUCTION

In the last two decades, after the invention of the scanning tunneling microscope (STM), a huge amount of experimental investigations of solid surfaces has been performed, showing the great potentialities of this kind of probe. One of the most attractive peculiarities of the STM is the possibility of studying the *local* surface-projected electronic density of states, possibly with atomic resolution, performing the so-called scanning tunneling spectroscopy (STS).[1–3]

From a theoretical point of view great efforts have been spent in order to build a satisfactory model description of STM/STS measurements. Generally speaking, to analyze the physics of tunneling in complex systems, three different categories of theoretical formulations are available.[4–6] The simplest one assumes a one-dimensional (or else, a separable) geometry in which the tunneling between two subsystems is described in terms of single particle states of the whole system and it is solved either exactly (if possible) or by using the WKB approximation. The second class is based on the so-called transfer Hamiltonian (TH) formalism, originally derived from time dependent perturbation theory, by which, in principle, multidimensional and many body effects can be included. The third kind of descriptions comprises more general many body theories, such as Landauer-Büttiker formalism,[7,8] nonequilibrium Green's function or Keldish theory,[8–10] and generalized scattering approaches.

All these theories were born and grew up before the invention of STM, and all of them have been further examined and developed in their general foundations, adapted and used, contributing significantly to gain basic understanding of the physics involved in STM/STS experiments.

Among the others, the TH formalism, first derived by Oppenheimer[11] from time dependent perturbation theory and later extended by Bardeen[12,13] to the case of solid junctions, is the most extensively used. The seminal work of Tersoff and Hamann,[14] the subsequent important generalizations achieved by Feuchtwang and Cutler,[15] and, above all, the series of papers of Chen[16,17] have provided a quite satisfactory theoretical framework for STM in the regime of low applied bias. Besides its relative simplicity in the treatment of three dimensional systems (and also of many body and inelastic effects in generalized versions), one of the most attractive features of the TH formulation of tunneling is that the description is given in terms of the properties of the two subsystems, i.e., the sample and the tip in the STM case.

Even though a large number of numerical studies founded on the TH formalism can be found in the STM/STS literature [essentially based on the full numerical resolution of Eqs. (4)–(9) of Sec. II of this paper, see, e.g., Refs. 18–23], it is important to note that, so far, most of the TH-based STM analytical theories have been developed in the limit of vanishing applied bias.

On the other hand, STS measurements are routinely performed exploring finite ranges of negative and positive applied bias. Several important theoretical results have been obtained in this field.[18,20,24–26] A common feature of these analytical STS descriptions is that they are mostly based on the first class of the tunneling theories mentioned above, namely, the one-dimensional (1D)-WKB formalism. In addition to these approaches, several numerical investigations, largely consisting in first principle calculations, based on the third class of tunneling theories suitably adapted to the STM case,[27–33] have been conducted,[30,34–39] while only few analytical treatments falling into this last category exist, for simplified geometries.[40–42] In most works of this last class the emphasis is given mainly on the numerical evaluation of the tunneling current, for the prediction of STM images. Much less attention is given to the analysis of the most relevant quantities for STS, namely, the voltage derivative of the current $dI(V)/dV$, as a function of the applied bias at a given position of the sample, and the so-called conductivity maps, i.e., two-dimensional $dI(V)/dV$ maps of the sample at constant current and applied bias.

On the basis of these considerations it can be concluded that, up to now, a well established and complete analytical theory, capable of both elucidating the essential underlying physics of the STS and interpreting STS experimental observations also quantitatively, is not available yet.

If a better quantitative understanding of STS is desired, the knowledge of the electronic states of the tip is a must. Bearing this crucial point in mind, several efforts have been produced, both experimentally and by computing techniques. Yet this important goal is difficult to achieve since the tip preparation is not a completely controllable procedure. In particular, the tip apex cannot be easily shaped as designed. Moreover, a realistic numerical simulation of such a complicated and not well known system is still a challenge. Nevertheless, general properties have been established and it would be very useful to incorporate this knowledge in a description of the STS.

The principal aim of this paper is to start the development of a general analytical description of the STS based on the TH formalism, exploiting, as much as possible, its attractiveness in terms of relative simplicity and physical transparency.

The paper is organized as follows. In Sec. II the theoretical framework will be discussed. Then, in Sec. III, the theory is developed with the goal of obtaining quite general expressions of the tunneling current and above all its voltage derivative, for arbitrary applied bias, assuming a given behavior of the tip wave functions and density of states. Different cases, which are likely to be found in tips used in experimental situations, will be considered, in order to provide a sufficiently general analysis. It is also possible to derive general considerations about STS by treating the two subsystems, namely, the sample and the tip, in a more symmetric way, following and extending an approach first proposed by Feuchtwang and Cutler; this will be the subject of Sec. IV. Concluding remarks are left to Sec. V, while the Appendix deals with an approximate analytical approach for treating thermal effects in STS.

II. THEORETICAL FRAMEWORK

We start by summarizing the definitions of the physical quantities of major interest and relevance in STM/STS theory, namely, the spectral density $\rho(\mathbf{r},\mathbf{r}',\varepsilon)$

$$\rho(\mathbf{r},\mathbf{r}',\varepsilon) = -2\,\mathrm{Im}\,G^R(\mathbf{r},\mathbf{r}',\varepsilon) = 2\pi\sum_\mu \psi_\mu(\mathbf{r})\psi_\mu^*(\mathbf{r}')\delta(\varepsilon_\mu-\varepsilon), \tag{1}$$

the local density of states $\rho(\mathbf{r},\varepsilon)$ (density of state per unit volume, LDOS)

$$\begin{aligned}\rho(\mathbf{r},\varepsilon) &= -\frac{1}{\pi}\,\mathrm{Im}\,G^R(\mathbf{r},\mathbf{r},\varepsilon)\\ &= \frac{1}{2\pi}\rho(\mathbf{r},\mathbf{r},\varepsilon)\\ &= \sum_\mu |\psi_\mu(\mathbf{r})|^2\delta(\varepsilon_\mu-\varepsilon),\end{aligned} \tag{2}$$

and the density of states $\rho(\varepsilon)$ (DOS) of a system

$$\begin{aligned}\rho(\varepsilon) &= -\frac{1}{\pi}\int \mathrm{Im}\,G^R(\mathbf{r},\mathbf{r},\varepsilon)d\mathbf{r} \equiv -\frac{1}{\pi}\mathrm{Tr}[\mathrm{Im}\,G^R(\varepsilon)]\\ &= \sum_\mu \delta(\varepsilon_\mu-\varepsilon) \equiv \sum_\mu \rho(\varepsilon,\mu).\end{aligned} \tag{3}$$

In the above equations, the expressions of the physical quantities are given in the general form, as derived from Green's function theory, and in the one-particle approximation, respectively. G^R is the retarded Green's function of the system, while $\psi_\mu(\mathbf{r})$, ε_μ, and $\rho(\varepsilon,\mu)\equiv\delta(\varepsilon_\mu-\varepsilon)$ are the eigenfunction, eigenvalue, and partial density of state (pDOS) associated with the μth one-particle level (the set is supposed discrete for simplicity); μ is the appropriate set of quantum numbers needed to describe the state while Tr is the trace operation. We remember that the connection between the two descriptions is established by noting that G^R can always be expanded, if a complete set of eigenfunctions is known, in the form $G^R=\Sigma_\mu\psi_\mu(\mathbf{r})\psi_\mu^*(\mathbf{r}')/(\varepsilon-\varepsilon_\mu+i\eta)$, where η is an infinitesimally small positive constant.[43]

Let us now consider the following expression of the tunneling current I, which will be our starting point for the subsequent analysis:

$$\begin{aligned}I = \frac{4\pi e}{\hbar}\int_{-\infty}^{+\infty}\sum_{\mu,\nu}|M_{s\nu,t\mu}|^2[f_t(\varepsilon-eV)-f_s(\varepsilon)]\\ \times\rho_t(\varepsilon-eV,\mu)\rho_s(\varepsilon,\nu)d\varepsilon;\end{aligned} \tag{4}$$

here, $|M_{s\nu,t\mu}|^2$ is a transmission coefficient which will be discussed later, $f_{t(s)}(\varepsilon)$ is the Fermi distribution function of the tip (sample), indices ν and μ label sample and tip energetic levels, respectively, all energies are referred to the Fermi energy, and V is the bias voltage. Moreover, ρ_s and ρ_t are the sample and tip electronic partial density of states, respectively.

Equation (4), originally derived under simplifying assumptions or from time dependent perturbation theory, has been extensively discussed in the theoretical literature (see, for example, Refs. 13 and 44). Formula (4) provides a quite general form for the electric current transmitted between two quantum systems, as can be understood on the basis of the following considerations.[8] Under very general assumptions, it is shown that I can be written as follows (Landauer-Büttiker formula):

$$I = \frac{2e}{h}\int \bar{T}(\varepsilon)[f_1(\varepsilon)-f_2(\varepsilon)]d\varepsilon, \tag{5}$$

where Eq. (5) has a very transparent physical meaning: a quantum of current $2e/h$ can be carried by an electron in a single energetic channel (e/h is about 40 nA/meV while the factor 2 account for spin degeneracy, which will be assumed in this work); the energetic interval of the allowed states in which a nonzero net current can flow due to the exclusion principle is selected by the difference between the two probability distribution functions f_1 and f_2 of the two subsystems; $\bar{T}(\varepsilon)$ is the average transmission probability that an electron injected at one side of the barrier will transmit to the other

side. If inelastic processes and incoherent transport can be neglected, $\bar{T}(\varepsilon)$ can be calculated using S matrix and Green's function theory, obtaining the following, rather abstract, but compact form:[8]

$$\bar{T} = \mathrm{Tr}[\Gamma_t(\varepsilon) G_b^R(\varepsilon) \Gamma_s(\varepsilon) G_b^A(\varepsilon)], \tag{6}$$

where $G_b^{R(A)}(\varepsilon)$ is the retarded (advanced) Green's function of the vacuum barrier and $\Gamma_{t(s)}(\varepsilon)$ is the so-called contact function of the tip (sample). An explicit calculation procedure (for example, the so-called discrete or tight binding Hamiltonian method[8,43]) provides a (more transparent) expression

$$\bar{T} = \mathrm{Tr}[\Gamma_t(\varepsilon) G_b^R(\varepsilon) \Gamma_s(\varepsilon) G_b^A(\varepsilon)] = \mathrm{Tr}[\rho_t(\varepsilon) \bar{M}(\varepsilon) \rho_s(\varepsilon) \bar{M}(\varepsilon)^\dagger], \tag{7}$$

where $\rho_{t(s)}(\varepsilon)$ are the spectral functions of the isolated tip (sample) (the notation here is abbreviated by omitting the spatial arguments, for the sake of simplicity) and the matrix element $\bar{M}(\varepsilon)$ depends on the Green's function of the barrier and its coupling to the tip and sample. Since the spectral function of a system has a physical meaning which is directly related to its local density of states [see Eq. (2) and also the discussion in Sec. IV A], a comparison among Eqs. (5)–(7) and Eq. (4) provides the following useful identification:[8]

$$\bar{T} = \mathrm{Tr}[\rho_t(\varepsilon) \bar{M}(\varepsilon) \rho_s(\varepsilon) \bar{M}(\varepsilon)^\dagger] \leftrightarrow 4\pi^2 \sum_{\mu,\nu} |M_{s\nu,t\mu}|^2 \rho_t(\varepsilon - eV, \mu) \rho_s(\varepsilon, \nu). \tag{8}$$

From these considerations, it is then clear that the quantity $|M_{s\nu,t\mu}|^2$ in Eq. (4) represents the transmission coefficient between the νth state of the sample and the μth state of the tip through the vacuum barrier.

In the TH formalism, the matrix element $M_{s\nu,t\mu}$ is given by[12,13]

$$M_{s\nu,t\mu} = -\frac{\hbar^2}{2m} \int_S \left[\psi_{s\nu}^*(\mathbf{r_t} - \mathbf{r}) \frac{\partial}{\partial n_t} \psi_{t\mu}(\mathbf{r_t}) - \psi_{t\mu}(\mathbf{r_t}) \frac{\partial}{\partial n_t} \psi_{s\nu}^*(\mathbf{r_t} - \mathbf{r}) \right] dS, \tag{9}$$

where $\psi_{s\nu}(\mathbf{r_t} - \mathbf{r})$ and $\psi_{t\mu}(\mathbf{r_t})$ are one-electron stationary states of the uncoupled sample and tip, respectively. According to the TH theory,[12,13] the integral is performed over a generic surface S located well inside the tunneling barrier and separating the sample and tip region; $\mathbf{r_t}$ is a coordinate referred to a frame fixed on the tip (whose origin, in a frame fixed on the sample, is located in $\mathbf{r}_0 \equiv -\mathbf{r}$) and $\partial/\partial n_t$ is the projection of the gradient operator on the direction of the outer normal $\mathbf{n_t}$ to the surface S (see Fig. 1).

The limits of validity of the TH tunneling theory have been deeply investigated, both before and after the advent of the STM. Among the others, Refs. 13, 42, and 44–47 provide an extensive investigation of this issue, actually very important for practical purposes and also extremely interesting from a fundamental perspective since it involves profound and subtle concepts (see also Sec. IV A of this paper). Here, we only say that the TH formalism has proven to be a very powerful theoretical tool in many physical situations, including the description of STM images. Generally speaking, it always gives a good description in the case of high and thick barriers; besides this, a modified version of the original Bardeen approach, still based on expressions similar to Eq. (9) (but in which the two sets of functions have a different meaning,[45] taking into account the reduction of the barrier height determined by the interaction between the two subsystems and the electric field dependence in the vacuum) further extends the limit of applicability of the formalism also to more general barrier configurations which can be quite often realized in STM/STS experiments (see also Ref. 40 for a critical analysis of this issue).

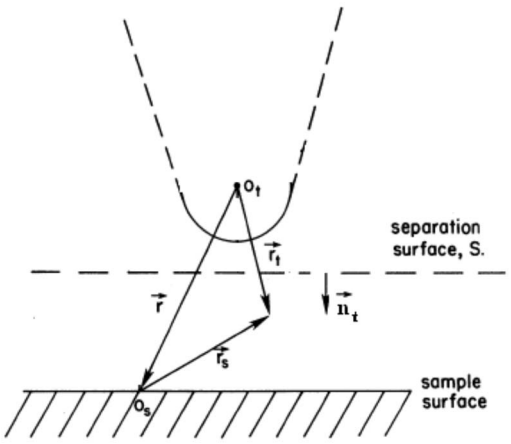

FIG. 1. Schematic view of the coordinate systems and corresponding relevant notations used in this paper.

In any case, even though the TH formalism is less general than other theories [namely, those based on Eqs. (5)–(7)], it has some very attractive features, like its more transparent physical interpretation and its relative simplicity, which opens the possibility for analytical investigations.

Two different theoretical approaches based on the TH formalism, useful for constructing an analytical description of an STS measure, will be developed in the next sections. In Sec. III, specific assumptions about the tip properties will be considered: in particular, the voltage derivative of the current dI/dV will be calculated on the basis of a knowledge of the angular symmetry of the tip states and by modeling the energetic dependence of the tip DOS. In Sec. IV, no specific assumptions about the tip will be made; we will show that in this case general expressions for the dI/dV can be derived, in terms of the sample and tip spectral densities. In both cases, a comparison with more simplified treatments, usually adopted in the literature for STS analysis, will be presented.

III. SCANNING TUNNELING SPECTROSCOPY CURRENT DERIVATIVE FOR FINITE VOLTAGE AND ARBITRARY ELECTRONIC TIP STRUCTURE

In this section, we will neglect thermal effects and approximate the Fermi functions with simple step functions,

with the consequence that in Eq. (4) the integration is performed over the interval $[0,eV]$. A general approach for the approximate treatment of thermal effects in expressions of the kind of Eq. (4) is given in the Appendix.

Since in Eqs. (4) and (9) the wave functions in general depend on the corresponding eigenvalue, $\psi_{t\mu}=\psi_{t\mu}(\mathbf{r_t},\varepsilon_\mu)$ and $\psi_{s\nu}=\psi_{s\nu}(\mathbf{r_t}-\mathbf{r},\varepsilon_\nu)$ must be calculated at $\varepsilon_\mu=\varepsilon-eV$ and $\varepsilon_\nu=\varepsilon$, respectively. Moreover, if the applied bias is not negligible, the wave functions must be calculated *in the presence of V*.[48] Therefore, $M_{s\nu,t\mu}$ in Eq. (9) will depend in general on both ε and V, $M_{s\nu,t\mu}=M_{s\nu,t\mu}(\varepsilon_\nu,\varepsilon_\mu)=M_{s\nu,t\mu}(\varepsilon,\varepsilon-eV)$ (for the sake of simplicity, the further explicit dependence in $\psi_{t\mu}$, $\psi_{s\nu}$, and $M_{s\nu,t\mu}$ on V is understood, unless explicitly indicated).

Let us focus on a single channel contribution $I_{\nu\mu}$ to the current, coming from a particular choice of μ and ν. If we calculate the first derivative of the current with respect to the voltage in Eq. (4), we obtain

$$\frac{\hbar}{4\pi e}\frac{dI_{\nu\mu}}{d(eV)} = |M_{s\nu,t\mu}(eV,0)|^2\rho_t(0,\mu)\rho_s(eV,\nu) + \int_0^{eV}\rho_s(\varepsilon,\nu)\frac{\partial}{\partial(eV)} \times[|M_{s\nu,t\mu}(\varepsilon,\varepsilon-eV)|^2\rho_t(\varepsilon-eV,\mu)]d\varepsilon. \tag{10}$$

It is evident that in Eqs. (4), (9), and (10) the sample and the tip are described symmetrically and, generally speaking, this already shows the great importance of the tip properties in determining the information contained in an STS measure. Actually, by keeping this symmetry in the description of the system, several interesting general conclusions can be obtained; this approach will be followed in Sec. IV. On the other hand, the two subsystems, the sample and the tip, are actually characterized by very different physical properties and it would be very useful to consider them in order to develop an STS theory. In particular, the distinctive features of the tip geometry and LDOS play a crucial role and a lot of experimental and theoretical work has been conducted in order to obtain a better understanding of the tip properties (see, for example, Sec. IV B of Ref. 49 and references therein). Unfortunately, little detailed experimental information is available, essentially because of uncertainty on tip fabrication and characterization down to the atomic scale. As a consequence it is also very difficult to perform reliable numerical simulations of the tip system. However, several general results have been obtained on the tip LDOS which are fundamental for the qualitative and quantitative understanding of the STM/STS. First of all, the angular symmetry of the tip apex states is essential in determining the spatial resolution achievable. While one of the most common assumptions usually adopted is that of a single *s*-like tip state which contribute to the tunneling current,[14] in Refs. 16 and 17 Chen put forward strong arguments according to which only p_z- or d_{z^2}-like tip states are able to explain the observed atomic resolution in STM. The tip material is usually a transition metal, which is dominated by *d*-type contributions, especially in the vacuum region, the relative importance of *s*, *p*, and *d* states also depending on the energetic interval considered. Therefore, it is important to evaluate the transition matrix element $M_{s\nu,t\mu}$ for the various angular states which can characterize a realistic tip, also taking into account the fact that a number of different tip states can in principle significantly contribute to the tunneling current. Another crucial issue is actually the energy dependence of the tip DOS. The simplifying assumption of constant tip DOS is predominantly used in the theoretical interpretation of STS measurements. On the other hand, on the basis of experimental measurements and numerical simulations,[16,29,50] there is a strong indication that the real DOS of a tip is characterized by highly localized states too, in the energy intervals of interest for STS applications (up to several eV around the Fermi energy). In order to include this kind of information in a theory of STS, we will consider an idealized model for describing the tip system, namely, a tip with a DOS characterized by three well defined peaks, here approximated by δ functions, $[\rho_t(\varepsilon,\mu)=\delta(\varepsilon-\varepsilon_\mu)$ with $\mu=1,2,3]$ in $\varepsilon_1<0$, $\varepsilon_2=0$, and $\varepsilon_3>0$, superimposed on an almost constant background $\bar{\rho}_{t\mu}$; later we will generalize the DOS to the case of an arbitrary number of localized peaks. We will also assume that all these components of the tip DOS are characterized by a well defined angular symmetry, as described below. This model tip allows the analysis of the different qualitative situations which can be often found in a real tip, in which the DOS can have a number of localized states over a smooth background. Our idealized tip DOS model is sketched in Fig. 2, where it is also indicated how it is representative of the realistic physical situation.

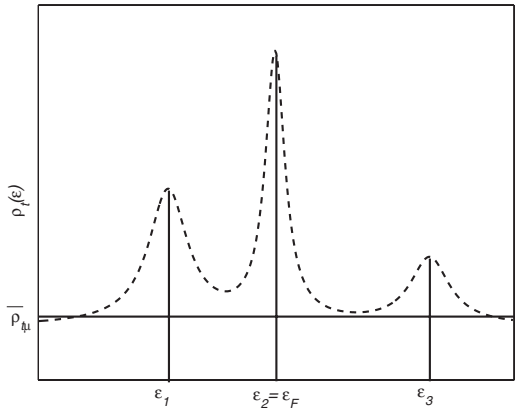

FIG. 2. Pictorial view of the basic qualitative features which can be found in a realistic tip density of states (dashed line) and its approximation as adopted in this work (solid lines).

In this paper, we will mainly investigate the case of a positive applied bias V (sample positive with respect to the tip): in this situation the sample contribution is due to the initially unoccupied, excited levels, while the occupied tip states which contribute to the current are in the interval $[-eV,0]$. At first, we neglect the contribution coming from the constant background, considering its effect subsequently. Therefore, from the sum over μ we have two contributions, coming at $\varepsilon=\varepsilon_1$ and at $\varepsilon=0$. From Eq. (10) we obtain

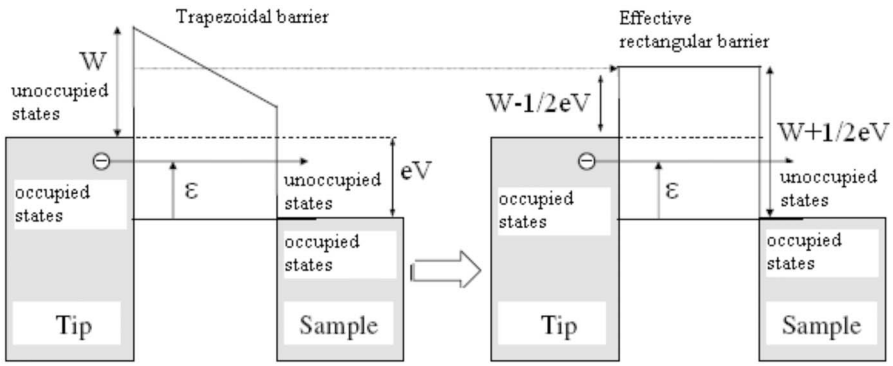

FIG. 3. Left: the application of a positive bias V to the sample produces a trapezoidal barrier to be overcome by an electron which starts from an initially occupied tip state of energy ε and arrives to the corresponding initially unoccupied sample state (only the elastic contribution is considered). W is the common work function of the two subsystems. Right: the approximation of the trapezoidal barrier with an effective rectangular one produces a V-dependent barrier height both for the tip (height equal to $W-1/2eV$) and for the sample ($W+1/2eV$).

$$\frac{\hbar}{4\pi e}\frac{dI_\nu}{d(eV)} = |M_{s\nu,t0}(eV,0)|^2\rho_s(eV,\nu)\delta(0) + \int_0^{eV}\rho_s(\varepsilon,\nu)\frac{\partial}{\partial(eV)}[|M_{s\nu,t1}(\varepsilon,\varepsilon-eV)|^2\delta(\varepsilon-eV-\varepsilon_1)]d\varepsilon$$

$$= |M_{s\nu,t0}(eV,0)|^2\rho_s(eV,\nu)\delta(0) + \int_0^{eV}\rho_s(\varepsilon,\nu)\delta[\varepsilon-(eV+\varepsilon_1)]\frac{\partial}{\partial(eV)}[|M_{s\nu,t1}(\varepsilon,\varepsilon-eV)|^2]d\varepsilon$$

$$-\int_0^{eV}\rho_s(\varepsilon,\nu)\left[\frac{\partial}{\partial\varepsilon}\delta(\varepsilon-(eV+\varepsilon_1))\right]|M_{s\nu,t1}(\varepsilon,\varepsilon-eV)|^2 d\varepsilon, \tag{11}$$

where it has been also used the fact that $\partial\delta(x-y)/\partial x=-\partial\delta(x-y)/\partial y$. Upon exploiting the properties of the δ functions, it is easy to calculate the integrals

$$\frac{\hbar}{4\pi e}\frac{dI_\nu}{d(eV)} = |M_{s\nu,t0}(eV,0)|^2\rho_s(eV,\nu)\delta(0) + \rho_s(\varepsilon_1+eV,\nu)\left[\frac{\partial}{\partial(eV)}|M_{s\nu,t1}(\varepsilon_1+eV,\varepsilon_1)|^2 + \frac{\partial}{\partial\varepsilon}|M_{s\nu,t1}(\varepsilon,\varepsilon-eV)|^2|_{\varepsilon=\varepsilon_1+eV}\right]$$

$$+\left[|M_{s\nu,t1}(\varepsilon,\varepsilon-eV)|^2\frac{\partial}{\partial\varepsilon}\rho_s(\varepsilon,\nu)\right]_{\varepsilon=\varepsilon_1+eV}. \tag{12}$$

For an evaluation of this expression fully in terms of the physical properties of the system, it is required to calculate the matrix elements $M_{s\nu,t\mu}(eV,0)$ and $M_{s\nu,t\mu}(\varepsilon,\varepsilon-eV)$, which depend on the spatial sample structure and tip wave functions. This can be done through an extension of the so-called Chen's derivative rules,[16,17] by assuming definite angular properties of the tip states. As explained above, it is particularly interesting to consider tip states with $\mu=s$-, p_z-, and d_{z^2}-like behavior. In the limit of negligible applied bias, Chen[16] showed that the matrix elements are given by the following derivative rules:

$$\mu = s, \quad M_{s\nu,ts} = \frac{2\pi C_{ts}\hbar^2}{m\kappa_W}\psi_{s\nu}(\mathbf{r_0}), \tag{13}$$

$$\mu = p_z, \quad M_{s\nu,tp_z} = \frac{2\pi C_{tp_z}\hbar^2}{m\kappa_W^2}\frac{\partial}{\partial z}\psi_{s\nu}(\mathbf{r_0}), \tag{14}$$

$$\mu = d_{z^2}, \quad M_{s\nu,td_{z^2}} = \frac{2\pi C_{td_{z^2}}\hbar^2}{m\kappa_W^3}\left(\frac{\partial^2}{\partial z^2}-\frac{1}{3}\kappa_W^2\right)\psi_{s\nu}(\mathbf{r_0}), \tag{15}$$

where the $C_{t\mu}$ are normalization constants related to the amplitude of the given tip state, $\kappa_W=\sqrt{2mW/\hbar}$ is the decay constant for states at the Fermi energy, and W is the work function of the system. These relations must be extended in order to include the energy and voltage dependence of the matrix elements in the general case of finite applied bias, as explained above. It is then necessary to consider the Schrödinger equation of the tip in the tunneling region. In this paper, we will introduce the presence of V in the simplest way,[6] namely, by approximating the almost trapezoidal vacuum barrier with an effective rectangular one, as indicated in Fig. 3 (for the sake of simplicity, the work functions of the tip and the sample are assumed to be equal). We obtain

$$M_{s\nu,ts}(\varepsilon,\varepsilon-eV)=\frac{2\pi C_{ts}\hbar^2}{m\kappa_{ts}(\varepsilon-eV)}\psi_{s\nu}(\mathbf{r_0},\varepsilon)$$

$$\rightarrow M_{s\nu,ts}(eV,0)=\frac{2\pi C_{ts}\hbar^2}{m\kappa_{ts}(0)}\psi_{s\nu}(\mathbf{r_0},eV), \tag{16}$$

$$M_{s\nu,tp_z}(\varepsilon,\varepsilon-eV)=\frac{2\pi C_{tp_z}\hbar^2}{m\kappa^2_{tp_z}(\varepsilon-eV)}\frac{\partial}{\partial z}\psi_{\nu s}(\mathbf{r_0},\varepsilon)$$

$$\rightarrow M_{s\nu,tp_z}(eV,0)=\frac{2\pi C_{tp_z}\hbar^2}{m\kappa^2_{tp_z}(0)}\frac{\partial}{\partial z}\psi_{s\nu}(\mathbf{r_0},eV), \tag{17}$$

$$M_{s\nu,td_{z^2}}(\varepsilon,\varepsilon-eV)$$

$$=\frac{2\pi C_{td_{z^2}}\hbar^2}{m\kappa^3_{td_{z^2}}(\varepsilon-eV)}\left(\frac{\partial^2}{\partial z^2}-\frac{1}{3}\kappa_{td_{z^2}}(\varepsilon-eV)^2\right)\psi_{s\nu}(\mathbf{r_0},\varepsilon)$$

$$\rightarrow M_{s\nu,td_{z^2}}(eV,0)$$

$$=\frac{2\pi C_{td_{z^2}}\hbar^2}{m\kappa^3_{td_{z^2}}(0)}\left(\frac{\partial^2}{\partial z^2}-\frac{1}{3}\kappa^2_{td_{z^2}}(0)\right)\psi_{s\nu}(\mathbf{r_0},eV), \tag{18}$$

where now $\kappa_\mu(\varepsilon_\mu)=\sqrt{2m(\varphi-\varepsilon_\mu)}$ represents the decay constant for a generic tip state of energy ε_μ, $\varphi=W-eV/2$ being the effective barrier height with an applied bias for the tip. These general expressions allow a full computation of Eq. (12) in terms of the physical properties of the sample. We can start drawing some general observations assuming the following general expression for the spatial dependence of the sample wave functions $\psi_{s\nu}$:

$$\psi_{s\nu}(\mathbf{r},\varepsilon_\nu,V)=C_{s\nu}\varphi_{s\nu}(\mathbf{r}_\parallel,\varepsilon_\nu)d_{s\nu}(z,\varepsilon_\nu,V). \tag{19}$$

$\varphi_{s\nu}(\mathbf{r}_\parallel,\varepsilon_\nu)$ is a function of the coordinate $\mathbf{r}_\parallel$ belonging to the surface and $d_{s\nu}(z,\varepsilon_\nu,V)$ describes the barrier- and z-dependent exponential decay of the considered electronic state. This is a quite general expression which can be adopted for atomically flat surfaces. Since, from the separation of the sample Schrödinger equation in the coordinates $\mathbf{r}_\parallel,z$, it follows that $(\partial^2/\partial z^2)d_{s\nu}=\alpha(z,\varepsilon_\nu,V)d_{s\nu}$ (where α is the appropriate function obtained by the Schrödinger equation of the sample), from Eqs. (16)–(18) we see that the general matrix elements are in any case proportional to $\psi_{s\nu}(\mathbf{r}_0)$, $M_{s\nu,t\mu}=K_\mu(\varepsilon,V)\psi_{s\nu}(\mathbf{r}_0)$, where K_μ depends on the angular symmetry of the tip functions. The derivative of the current can then be expressed as a function of the sample pDOS and partial local density of state (pLDOS) $\rho_s(\mathbf{r},\varepsilon,\nu)\equiv|\psi_{s\nu}(\mathbf{r})|^2\rho_s(\varepsilon,\nu)$ as

$$\frac{\hbar}{4\pi e}\frac{dI_\nu}{d(eV)}=A_{1\mu}\rho_s(\mathbf{r_0},eV,\nu)+A_{2\mu}(\mathbf{r_0})\rho_s(\varepsilon_1+eV,\nu)$$

$$+A_{3\mu}(\mathbf{r_0})\rho_s'(\varepsilon,\nu)|_{\varepsilon=\varepsilon_1+eV}, \tag{20}$$

where the functions $A_{1\mu}$, $A_{2\mu}$, and $A_{3\mu}$, which depend on ε_1, V, and W, can be completely calculated by using Eqs. (16)–(18). Equation (20) tells us that the derivative of the current with respect to the applied bias is not simply proportional to the sample LDOS calculated in $\varepsilon=eV$: other terms related to the sample LDOS in $\varepsilon=\varepsilon_1+eV(<eV)$ arise. The importance of this distortion can be determined by evaluating the relative weight of the factors $A_{i\mu}$ and, above all, the exponential damping term $d_{s\nu}(z,\varepsilon_\nu,V)$ in the wave function $\psi_{s\nu}$ [see Eq. (19)], which depends on energy. In general it is possible to say that, since $\varepsilon_1<0$, the disturbing terms, proportional to $|d_{s\nu}(z,\varepsilon_1+eV,V)|^2$, are more damped than the desired signal (proportional to the sample LDOS at eV), if $V>0$.

Let us now discuss the contribution to I and $dI/d(eV)$ coming from a constant background $\bar{\rho}_t$ in the tip DOS. From Eqs. (10) and (16)–(18), it is evident that, even though in this case ρ_t does not depend on V, the second integral term in $dI/d(eV)$ is in general not zero since $M_{s\nu,t\mu}$ depends on V because of the decaying function $d_{s\nu}(z,\varepsilon_\nu,V)$ and the normalization factor $C_{s\nu}$ contained in $\psi_{s\nu}$. We saw that $M_{s\nu,t\mu}=K_\mu(\varepsilon,V)\psi_{s\nu}(\mathbf{r}_0)$. We now assume that $[\partial/\partial(eV)]d_{s\nu}=\beta(z,\varepsilon_\nu,V)d_{s\nu}$ so that $[\partial/\partial(eV)]d^2_{s\nu}=2\beta(z,\varepsilon_\nu,V)d^2_{s\nu}$ and $[\partial/\partial(eV)]C_{s\nu}=\gamma(\varepsilon_\nu,V)C_{s\nu}$, with β and γ suitable functions which can be calculated explicitly for specific cases. From these considerations it follows that

$$\frac{\partial}{\partial(eV)}|M_{s\nu,t\mu}|^2=\left[\frac{\partial}{\partial(eV)}(K^2_\mu)+2K^2_\mu(\beta+\gamma)\right]|\psi_{s\nu}(\mathbf{r})|^2$$

$$\equiv D_\mu(\varepsilon,V,\nu)|\psi_{s\nu}(\mathbf{r})|^2.$$

The contribution to $dI/d(eV)$ due to a constant tip DOS then reads

$$\frac{\hbar}{4\pi e}\frac{dI_{\nu\mu}}{d(eV)}=\bar{\rho}_t\left[K^2_\mu(eV)\rho_s(\mathbf{r_0},eV,\nu)\right.$$

$$\left.+\int_0^{eV}D_\mu(\varepsilon,V,\nu)\rho_s(\mathbf{r_0},\varepsilon,\nu)d\varepsilon\right]. \tag{21}$$

The first term is of the same kind of the first term in Eq. (20) and can be added to it; the second is a background term which depends on all the sample states contained in the interval $[0,eV]$, weighted by the function D_μ which describes the effect induced by a finite applied bias on the tip and sample states. Generally speaking, when V becomes an appreciable fraction of the work function, we expect that it will be important to take these effects into account for a better quantitative description of the tunneling current.[48] When the action of the applied bias on the system wave functions and energy levels are neglected D_μ is zero, the second term in Eq. (21) vanishes, and then the approximation of constant tip DOS would lead to a direct proportionality between $dI_{\nu\mu}/d(eV)$ and the sample partial density of states at the energy eV, which is the most generally adopted description.

We can now generalize the theory to the case of a tip DOS with an arbitrary number of peaks in the interval $[-eV,0]$ and a constant background, each state having its own angular symmetry. This tip determines a $dI/d(eV)$ expressed as a sum $dI/d(eV)=\Sigma_\nu dI_\nu/d(eV)$, with

$$\frac{\hbar}{4\pi e}\frac{dI_\nu}{d(eV)} = A_1(eV)\rho_s(\mathbf{r_0}, eV, \nu) + \int_0^{eV} D(\varepsilon, V, \nu)\rho_s(\mathbf{r_0}, \varepsilon, \nu)d\varepsilon + \sum_\mu [A_{2\mu}(\mathbf{r_0}, \varepsilon_\mu, V)\rho_s(\varepsilon_\mu + eV, \nu) + A_{3\mu}(\mathbf{r_0}, \varepsilon_\mu, V)\rho_s'(\varepsilon, \nu)|_{\varepsilon=\varepsilon_\mu+eV}]. \quad (22)$$

Equation (22) is one of the main results of this paper. It shows in which way the key quantity directly measured in an STS experiment, the derivative of the current, is related to the sample density of states for a given tip DOS. In particular, $dI/d(eV)$ when $V>0$ is a sum of various channels $dI_\nu/d(eV)$, each of which contains in general three groups of terms: the first is proportional to the real quantity of interest, namely, the sample pLDOS evaluated at the tip (center of curvature) position and at the energy eV; the second is a background term which is a weighted integral of the sample pLDOS in the interval $[0, eV]$; and the third kind of contribution comes from possible localized peaks in the occupied tip DOS at the energies $\varepsilon_\mu<0$, which produce corrections related to the sample DOS and its derivative evaluated at an energy shifted with respect to eV, namely, at $\varepsilon_\mu+eV(<eV)$. In Eq. (22), the functions $A_1 \equiv \Sigma_\mu A_{1\mu}$, $D \equiv \Sigma_\mu D_\mu$ (where D describes the possible presence of background with different angular symmetry), $A_{2\mu}$, and $A_{3\mu}$ can be calculated explicitly by using Eqs. (16)–(18), depending on the nature of the tip DOS used. The total current derivative is the sum over the various sample channels, each labeled by the set of quantum numbers ν. The first term then becomes the total sample LDOS, while the others in general exhibit a more complicated structure, which can be completely determined for specific cases. All these contributions present a strong spatial z dependence, which can be evaluated explicitly in order to investigate the more appropriate data-normalization techniques for extracting the sample LDOS at the surface, as it will be discussed in the next section. These considerations generalize those obtained using a simple WKB expression of the current, and Eq. (22) allows a quantitative estimation of these effects. Moreover, using Eq. (22) it is possible to include also the spatial dependence of the signal, thus enabling, for example, the calculation of conductivity maps, to be compared with experiments. It must be noted that a residual dependence on V is still contained in the functions A_i; A_1 is also proportional to the tip LDOS at the Fermi energy, while $A_{2\mu}$ and $A_{3\mu}$ are proportional to the intensity of the peak located in $\varepsilon=\varepsilon_\mu$. If ε_μ is not too far from the Fermi energy and the peak intensity is relevant, the resulting dI/dV as a function of V can be significantly different from the actual sample LDOS.

Let us now comment on the possible applicability of this description to the interesting case of a sample which is not simply an atomically flat surface but a cluster or molecule adsorbed on it. In order to analytically evaluate the various transition matrix elements arising for different tip states, we made the assumption of sample states which can be written in a separable form like in Eq. (19). This enabled us to establish a direct proportionality between the general transition matrix element and the sample wave function considered, which leads to a transparent physical interpretation of the final result contained in Eq. (22). In the case of adsorbates on the sample surface (or, more generally, for nonatomically flat surfaces), strictly speaking, such an assumption is no longer possible. However, if the size of the structures in the direction normal to the sample surface is small compared to the tip-sample distance, it is still possible to assume a separable form of the sample wave function, the presence of the adsorbate resulting only in a modification of the functions φ and d in Eq. (19). In this sense, Eq. (22) can be applied also to these situations. If the adsorbate cannot be considered as a small perturbation, in general we can conclude that the quantitative details in the evaluation of the general transition matrix element change, even if the general features of the problem as described by the structure of Eq. (22) can still provide a useful qualitative insight of the physical picture.

If the surface is periodic, we can further elaborate the general expression of the sample wave functions, Eq. (19), in the following way, starting from the observation that $\varphi_{s\nu}(\mathbf{r}_\parallel)$ in this case is a Bloch function and each state is then labeled by the wave vector $\mathbf{k}_\parallel$. We can expand the function $\varphi_{\mathbf{k}_\parallel}(\mathbf{r}_\parallel)d_{\mathbf{k}_\parallel}[z, \varepsilon(\mathbf{k}_\parallel)]$ into a complete set of basis functions $a^n_{\mathbf{k}_\parallel}\exp(i\mathbf{G}^n_\parallel\cdot\mathbf{r}_\parallel)d^n_{\mathbf{k}_\parallel}[z, \varepsilon(\mathbf{k}_\parallel)]$,

$$\psi_{s\mathbf{k}_\parallel}(\mathbf{r}, \varepsilon, V) = C_{\mathbf{k}_\parallel}\sum_n a^n_{\mathbf{k}_\parallel}\exp[i(\mathbf{G}^n_\parallel + \mathbf{k}_\parallel)\cdot\mathbf{r}_\parallel]d^n_{\mathbf{k}_\parallel}[z, \varepsilon(\mathbf{k}_\parallel)], \quad (23)$$

where $\mathbf{G}^n_\parallel$ are the set of reciprocal lattice vectors associated with the two-dimensional periodic surface. Also, as already explained, we use the simplest model for the sample wave function in vacuum, describing the vacuum potential as a constant barrier of height $W+eV/2$. With these assumptions, the corresponding quantity in Eq. (19) becomes

$$d^n_{\mathbf{k}_\parallel}(\mathbf{r}, \varepsilon, V) = \exp\{-z\sqrt{2m[W + eV/2 - \varepsilon(\mathbf{k}_\parallel)]/\hbar^2 + |\mathbf{k}_\parallel + \mathbf{G}^n_\parallel|^2}\}. \quad (24)$$

A further simplification can be achieved if we consider that an STM measurement probes only a small part of the sample first Brillouin zone, namely, that in correspondence of the Γ point, $\mathbf{k}_\parallel \simeq 0$.[22,25] This is due to the fact that the states with these values of the wave vector decay more slowly in vacuum and consequently $M_{\mathbf{k}_\parallel=\mathbf{0}}$ is the most dominant contribution in the sum of Eq. (4) on the sample states. Also the nontranslational symmetry operations of the two-dimensional spatial group of the surface can be used to greatly simplify the analysis.[22,45] Using the nontranslational symmetry, some of the reciprocal lattice vectors $\mathbf{G}^n_\parallel$ become equivalent, forming a *star* of reciprocal vectors, each *star* being identified by a representative vector $\mathbf{G}^S_\parallel$. Accordingly, the expansion over the reciprocal lattice vectors can be reduced into a sum over symmetrized *star functions*. If we consider only the contribution coming from the lowest stars [associated with the star vectors $\mathbf{G}^0_\parallel=(0,0)$ and $\mathbf{G}^1_\parallel$ which depends on the particular surface], we can use the following approximate expression of the sample LDOS:

$$\rho_s(\mathbf{r}_\parallel,z,\varepsilon) \simeq \rho_{s,\mathbf{k}_\parallel=\mathbf{0}}(\mathbf{r}_\parallel,z,\varepsilon) \simeq b^0(\varepsilon)|d^0_{\mathbf{k}_\parallel=\mathbf{0}}[z,\varepsilon(\mathbf{k}_\parallel=\mathbf{0})]|^2 + b^1(\mathbf{r}_\parallel,\varepsilon)|d^1_{\mathbf{k}_\parallel=\mathbf{0}}[z,\varepsilon(\mathbf{k}_\parallel=\mathbf{0})]|^2, \quad (25)$$

where the z-dependent decay of the two stars is described by the functions $d^0_{\mathbf{k}_\parallel=\mathbf{0}}=\exp[-z\sqrt{2m[W+eV/2-\varepsilon(\mathbf{k}_\parallel=\mathbf{0})]/\hbar^2}]$, $d^1_{\mathbf{k}_\parallel=\mathbf{0}}=\exp\{-z\sqrt{2m[W+eV/2-\varepsilon(\mathbf{k}_\parallel)]/\hbar^2+|\mathbf{G}^1_\parallel|^2}\}$, respectively, while the coefficients b^0 and b^1 can be expressed in terms of the relevant coefficients $a^n_{\mathbf{k}_\parallel}\exp[i\mathbf{G}^n_\parallel\cdot\mathbf{r}_\parallel]$ of Eq. (23) (namely, those made equivalent by the nontranslational symmetry), if available,[22,48] or can be used as fitting parameters in comparison with first principle numerical calculations or with experimental data. Then, using Eq. (25), together with Eq. (22), the derivative of the current can be calculated and expressed in terms of the electronic structure of a periodic surface.

A. Local density of states vs dI/dV or vs $(dI/dV)/(I/V)$ or vs $(dI/dV)/?$

In the literature STS experiments are usually interpreted by means of relations which are simpler than either Eq. (10) or Eq. (22), or those derived in Sec. IV. It is interesting to compare the expressions usually considered in the literature with those obtained in this section, with three goals: (i) to better understand the limits of the simplified theories as suggested by the TH formalism, (ii) to interpret the proper physical meaning of quantities used in the simplified descriptions to the light of the results of the TH formalism, and (iii) to find an appropriate treatment of an STS measure in order to obtain quantitative information about the sample investigated.

One of the most commonly used expressions for the tunneling current in the presence of a non-negligible bias potential applied across a solid-vacuum-solid interface is (see, for example, Ref. 25)

$$I=\frac{4\pi e}{\hbar}\int_0^{eV} T[E,U(\mathbf{r_0})]\rho_s(\varepsilon)\rho_t(\varepsilon-eV)d\varepsilon, \quad (26)$$

where the quantities $\rho_s(\varepsilon)$ and $\rho_t(\varepsilon-eV)$ are usually referred to as "densities of states" (of the sample and tip, respectively), while the transmission coefficient T is normally calculated in a simplified way, for example, in the 1D-WKB approximation as follows:

$$T[E,U(\mathbf{r}_0)]=\exp\left(-2\sqrt{\frac{2m}{\hbar}}\int_0^{z_0}\sqrt{\varepsilon-U(z)}dz\right), \quad (27)$$

where z_0 represents the width of the tunneling barrier U. From Eq. (26), the corresponding expression for the first derivative of the current is

$$\frac{dI}{d(eV)}=T[eV,U(z_0)]\rho_t(0)\rho_s(eV) + \frac{4\pi}{\hbar}\int_0^{eV}\rho_s(\varepsilon)\frac{d[T(E,U(z_0))\rho_t(\varepsilon-eV)]}{d(V)}d\varepsilon. \quad (28)$$

The consequences of Eqs. (26)–(28) are discussed in great detail in Refs. 25 and 26, where it is shown that, because of the properties of the transmission coefficient, most of the information about sample "density of states" comes only in the region $V>0$, from the first term on the right hand side. Also, starting from Eq. (28), an approximate analysis of the different normalization techniques for the extraction of the surface-projected sample DOS can be conducted (see below). On the other hand, this kind of description is limited for several reasons. First of all, being basically a one-dimensional description, the notion of DOS and LDOS cannot be clearly distinguished, the quantities involved cannot be directly related to those of the physical system, with the consequence that it is also not possible to perform a quantitative analysis of the conductivity maps of a system. Also, the spatial properties of the tip are not considered. Moreover, the WKB approximation cannot describe correctly thin and low barriers, which are often encountered in STM experiments.

The theory as developed in this section allows several significant generalizations in the analysis, in various directions.

First, the TH description allows the inclusion of the spatial effects, leading to the fundamental improvement of expressing the derivative of the current in terms of the proper physical quantities of the sample, namely, its pDOS and LDOS [see Eq. (22)]. Besides improving the interpretation of the experimental data, this fact opens the way for a theoretical evaluation of the so-called conductivity maps, as already pointed out.

Moreover, since the aim of an STS measurement is the estimation of the sample LDOS at the energy eV with the quantity $dI(V)/d(eV)$, it is possible to study in more detail the relative importance of other disturbing terms which in general can be present, even if a featureless tip is assumed [compare the last term in Eq. (28) with the last terms of Eq. (22)].

A crucial issue is the exponential behavior of STS data for finite applied voltage, a signature of the tunnel effect which can substantially hide the information about the sample LDOS *on* the surface, the real quantity of interest. In order to overcome this difficulty, several normalization procedures have been proposed. By far, the most commonly used in the interpretation of experimental data is the normalization to the (experimental) quantity I/V. On the other hand, on the basis of Eqs. (26)–(28), it can be concluded that normalizing to the WKB-derived transmission coefficient T is a more founded approach which can better unveil the physical properties of the sample.[25,26] This approach has been exploited very rarely in the experimental literature.[51,52] Very recently a further refinement of this normalization procedure, within the framework of the same 1D-WKB model, has been proposed.[53]

Our method can be used also to find out the most proper normalization procedure. We saw that the *unperturbed* sample LDOS at the surface, $\rho_{s0}(\mathbf{r}_\parallel,z=0,\varepsilon)$, is altered by the presence of an applied bias since the sample eigenfunctions change if $V\neq 0$. The *perturbed* sample LDOS $\rho_s(\mathbf{r}_{\parallel\mathbf{0}},z_0,\varepsilon=eV)$ is the quantity contained in the first term of the right hand side of Eq. (22). The aim is to extract from this term of dI/dV an estimate for $\rho_{s0}(\mathbf{r}_\parallel,z=0,\varepsilon)$. Besides being a quan-

tity which depends on the sample LDOS perturbed by the presence of the applied bias, dI/dV differs from $\rho_{s0}(\mathbf{r}_\parallel, z=0,\varepsilon)$ for at least two other reasons: (i) dI/dV as given by Eq. (22) is the sum of various terms, only the first of which, $A_1(eV)\rho_s(\mathbf{r}_0, eV)$, can be directly related to the sample LDOS at $\varepsilon=eV$; (ii) dI/dV is calculated at a finite tip-sample distance, $z_0 \neq 0$, which induces an exponentially increasing behavior on the applied bias (at least in the proper tunneling regime, which in turn depends both on V and on $z_0 \neq 0$). The most suitable normalization procedure should take into account these two issues. Each sample pLDOS $\rho_s(\mathbf{r}, \varepsilon, \nu)$ has a z dependence described by the function $|d_{s\nu}(z, \varepsilon_\nu, V)|^2$. For a given z_0, $d_{s\nu}$ in general is a function which increases exponentially with the difference $\varepsilon - \varepsilon_\parallel$, $\varepsilon_\parallel$ being the component of the total energy of the state associated with the motion parallel to the surface. Then, the sample pLDOS with the lowest possible $\varepsilon_\parallel$ component is heavily weighted by $d_{s\nu}$ in the sum over all the sample pLDOS which gives the total LDOS at a given ε (see also the discussion at the end of the previous subsection in the case of a periodic surface). Therefore, a natural normalization function for dI/dV is $|d_{s\nu}(z_0, \varepsilon_{\bar{\nu}}, V)|^2$ where $\bar{\nu}$ identifies the state with zero $\varepsilon_\parallel$. By applying these considerations to the case of a periodic surface, the best normalization factor is the z-dependent weighting factor of the lowest nontrivial (that is, non constant in $\mathbf{r}_\parallel$) star function, namely, $|d^1_{\mathbf{k}_\parallel=0}|^2$ in the notation used in the previous subsection. Moreover, since the other disturbing terms present in dI/dV have a weaker exponential behavior than the first, this kind of normalization is also able to produce an overall damping of all undesired contributions. From the present discussion it is also evident that a careful treatment of the z behavior of the sample eigenstates, for example, considering a better description of the barrier than those usually adopted (e.g., the exact solution associated with a trapezoidal barrier), would be helpful in enhancing the best interpretation of an STS measurement.

B. Example

The following simple model system, though pathological in some sense, allows an almost complete analytical treatment. We consider a one-dimensional jellium of finite length L for describing the sample electronic states, while the tip is characterized by a constant s-like DOS. In the absence of applied bias, the height of the barrier is equal to the work function W. The unperturbed sample wave functions inside the sample ($z>0$, region I) and in vacuum region ($z<0$, region II) are

$$\psi^{\mathrm{I}}_{sk} = \left(\frac{2}{L}\right)^{1/2} \sin[k(\varepsilon)z - \phi(k)], \quad z>0,$$

$$\psi^{\mathrm{II}}_{sk} = \left(\frac{2}{L}\right)^{1/2} \frac{k(\varepsilon)}{k_W} \exp\left[z\left(\frac{2m}{\hbar}(W-\varepsilon)\right)^{1/2}\right], \tag{29}$$

respectively. In the above equations, the wave number k is related to the energy level ε by the simple free-electron expression $k=\sqrt{2m\varepsilon}/\hbar$, $k_W=\sqrt{2mW}/\hbar$, and $\phi(k)=\arcsin(k/k_W)$. In this one-dimensional system degeneracy is absent, thus pLDOS and LDOS coincide. Since the density of states in the k space is L/π, the unperturbed sample LDOS $\rho_{s0}(z,\varepsilon)$ in the region $z \leqslant 0$ is given by $|\psi^{\mathrm{II}}_{sk}|^2(L/\pi)(dk/d\varepsilon)$. Introducing now the dimensionless quantities $\varepsilon/W \to \varepsilon$, $k_W z \to z$, and $(k_W/W)^{-1}\rho_s(\varepsilon,z) \to \rho_s(\varepsilon,z)$, we have

$$\rho_{s0}(z,\varepsilon) = \frac{\varepsilon^{1/2}}{\pi} \exp[2z(1-\varepsilon)^{1/2}]. \tag{30}$$

When the bias V is applied to the sample, the quantity W is changed into $W+eV/2$ and the dimensionless sample LDOS becomes

$$\rho_s(z,\varepsilon) = \frac{1}{\pi}\frac{\varepsilon^{1/2}}{1+V/2} \exp[2z(1+V/2-\varepsilon)^{1/2}], \tag{31}$$

where the dimensionless potential $eV/W \to V$ has been introduced.

Similar considerations apply to the tip, for which the effective barrier is $W-eV/2$ (see Fig. 3). From the Schrödinger equation for the tip the quantity κ_s of Eq. (16) is equal to $\sqrt{2m(W-eV/2-\varepsilon)}/\hbar = k_W\sqrt{1-V/2-\varepsilon}$. The above relations can be used to determine the function D in Eq. (21): it reads

$$D(\varepsilon,z_0,V) = D_0\left[-\frac{1}{(1+V/2-\varepsilon)^2} + \frac{z_0}{2(1+V/2-\varepsilon)^{3/2}} - \frac{4}{(1+V/2)(1+V/2-\varepsilon)}\right], \tag{32}$$

where $D_0 \equiv (\pi^2 C_{ts}^2 \hbar^4)/(m^2 W^2 k_W^2)$. Finally, using Eqs. (30)–(32) and Eq. (21), $dI/d(eV)$ calculated for this system (where the "1D current" has the dimension of a current density) is given by

$$\frac{\hbar}{4\pi e}\frac{dI}{d(eV)} = \bar{\rho}_t k_W D_0\left[\frac{4}{1-V/2}\rho_s(z_0,V) + \int_0^V D(\varepsilon,z_0,V)\rho_s(z_0,\varepsilon)d\varepsilon\right]. \tag{33}$$

Before showing the quantitative aspects of the result, let us refer to the general discussion developed above in this section. The unperturbed sample LDOS at the surface is $\rho_{s0}(0,\varepsilon)=\varepsilon^{1/2}/\pi$. It is different from the sample LDOS when a voltage is applied, which is [Eq. (31)] $\rho_s(0,\varepsilon)=\varepsilon^{1/2}/[\pi(1+V/2)]$. This last is the only quantity which, in principle, is made accessible by an STS measurement. Therefore, it is crucial to evaluate how large is the distortion induced on the sample LDOS by the presence of the applied bias.[48] Obviously, the distortion becomes a function of V which produces a first contribution to the functional dependence of the measured $dI/d(eV)$ vs V. We also saw that, even assuming a featureless tip, $dI/d(eV)$ is the sum of two contributions, only the first of which being directly proportional to the (distorted) sample LDOS evaluated in $\varepsilon=V$, though even the complete functional dependence on V in the first term of dI/dV is in general changed. In the present example these aspects are described quantitatively by Eqs. (32) and (33).

In STS z_0 is clearly always different from zero. The z-dependent term in the sample LDOS also varies with the

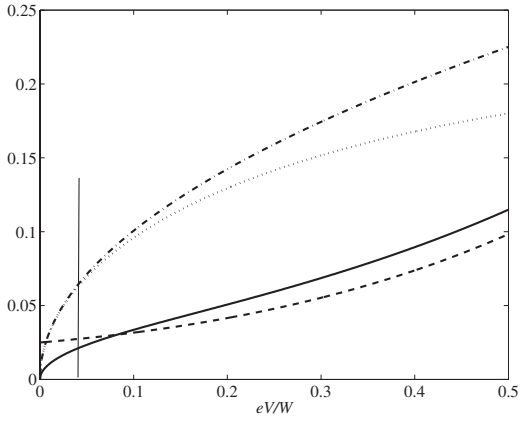

FIG. 4. $(dI/dV)/T_n$ (solid line), $\rho_{s0}(0,V)$ (dot-dashed line), $\rho_s(0,V)$ (dotted line), and $(dI/dV)/(I/V)$ (dashed line), corresponding to the example of Sec. III B, are displayed vs V for the case $z_0=2$. The maxima of the normalized curves are scaled to the maximum value of the function $\rho_{s0}(0,V)$. A generic Fermi energy of the sample is also indicated with a vertical line.

applied bias and, for fixed z and varying V, it determines the exponential growing behavior of the current derivative. It is this term that a proper normalization procedure should in principle compensate. From Eq. (31) it is clear that this quantity is $\exp[2z_0(1+V/2-\varepsilon)^{1/2}]_{\varepsilon=V}=\exp[2z_0(1-V/2)^{1/2}]$. In the region $V>0$ the second background term in the normalized current derivative will contain a net exponentially decreasing dependence on the voltage, helping in the extraction of the quantity of interest, in accordance with the considerations developed in Sec. III.

These aspects are summarized in Fig. 4. According to the considerations developed in the previous subsection, the most suitable normalization factor in this case is the function $T_n \equiv \exp[2z(1-V/2)^{1/2}]$. In Fig. 4 $(dI/dV)/T_n$ (solid line), $\rho_{s0}(0,V)$ (dot-dashed line), and $\rho_s(0,V)$ (dotted line) are displayed vs V; also, the dashed line corresponds to the quantity $(dI/dV)/(I/V)$. The maxima of the normalized curves are scaled to the maximum value of the function $\rho_{s0}(0,V)$. It can be seen that the normalization to T_n preserves the qualitative features of the unperturbed sample DOS, at least in the lower bias interval. This does not happen with the usual (I/V) normalization. Going to higher values of the applied bias, even the quantity $(dI/dV)/T_n$ has properties which differ significantly from the physical quantity of interest (there is a spurious change in the curvature).

IV. CURRENT FOR FINITE VOLTAGES IN TERMS OF SPECTRAL DENSITIES

It is possible to derive other very general relations for the tunneling current, expressed in terms of the spectral densities of the tip and sample. These expressions can be used to obtain some general conclusions when no specific assumptions are made on the tip states and for more realistic numerical simulations of an STS experiment.

Coming back to Eq. (9), it must be observed that the wave functions used in the TH theory are assumed to well approximate the one-electron wave function of the coupled system, in the tip and sample, respectively, as well as inside the potential barrier, up to the integration surface. For consistency, it is then necessary that the energies of the wave functions $\psi_\nu^*(\mathbf{r_t}-\mathbf{r})$ and $\psi_\mu(\mathbf{r_t})$ are the same, and moreover that the probability density and probability density current are conserved across the surface S. This implies the continuity of the logarithmic derivative of the functions as follows:

$$\left.\frac{1}{\psi_\nu}\frac{\partial\psi_\nu}{\partial n_t}\right|_S = \left.\frac{1}{\psi_\mu}\frac{\partial\psi_\mu}{\partial n_t}\right|_S. \tag{34}$$

If we now assume a semiclassical behavior in the vicinity of S, that is if $\psi_\nu(\mathbf{r_t}) \approx A(\mathbf{r_t})\exp(i\varphi(\mathbf{r_t}))$ with $|\nabla A/A| \ll 1$, then we have $(1/\psi_\nu)\nabla\psi_\nu \approx -(1/\psi_\nu^*)\nabla\psi_\nu^*$. This allows us to express the consistency relation given by Eq. (34) in the following, approximated, convenient form:[15]

$$-\left.\frac{1}{\psi_\nu^*}\frac{\partial\psi_\nu^*}{\partial n_t}\right|_S \approx \left.\frac{1}{\psi_\mu}\frac{\partial\psi_\mu}{\partial n_t}\right|_S; \tag{35}$$

in this way, the quantity $|M_{s\nu,t\mu}|^2$ reads

$$|M_{s\nu,t\mu}|^2 = \frac{\hbar^4}{m^2}\int_{SS'}\psi_\nu^*(\mathbf{r_t}-\mathbf{r})\psi_\nu(\mathbf{r_t'}-\mathbf{r}) \times \frac{\partial^2}{\partial n_t \partial n_t'}[\psi_\mu(\mathbf{r_t})\psi_\mu^*(\mathbf{r_t'})]dSdS'. \tag{36}$$

Upon substituting Eq. (36) in Eq. (4) and recalling the definition of the spectral density valid in the case of noninteracting electrons, $\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon)=2\pi\Sigma_\mu\psi_\mu(\mathbf{r_t})\psi_\mu^*(\mathbf{r_t'})\delta(\varepsilon_\mu-\varepsilon)$ (see Sec. II), the resulting expression reads

$$I = \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2\frac{1}{\pi^2}\int_{-\infty}^{+\infty}[f(\varepsilon-eV)-f(\varepsilon)] \times \int_{SS'}\rho_s^*(\mathbf{r_t}-\mathbf{r},\mathbf{r_t'}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t\partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon-eV)d\varepsilon dSdS', \tag{37}$$

or also, by noting that $\rho_s^*(\mathbf{r_t},\mathbf{r_t'},\varepsilon)=\rho_s(\mathbf{r_t'},\mathbf{r_t},\varepsilon)$

$$I = \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2\frac{1}{\pi^2}\int_{-\infty}^{+\infty}[f(\varepsilon-eV)-f(\varepsilon)] \times \int_{SS'}\rho_s(\mathbf{r_t'}-\mathbf{r},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t\partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon-eV)d\varepsilon dSdS'. \tag{38}$$

Equation (37), which is the generalization of Eq. (2.18) in Ref. 15, represents the general TH expression of the tunneling current, valid for arbitrary values of the temperature T and the applied bias voltage V and for an arbitrary tip. It can be used in order to calculate a general expression of the first derivative of the current, dI/dV, the key quantity in STS experiments. It is given by

$$\frac{dI}{d(eV)} = \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2 \frac{1}{\pi^2}\left\{\int_{-\infty}^{+\infty}[f(\varepsilon - eV) - f(\varepsilon)]\int_{SS'}\rho_s^*(\mathbf{r_t}-\mathbf{r},\mathbf{r_t'}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\frac{\partial}{\partial(eV)}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV)d\varepsilon dS dS' + \int_{-\infty}^{+\infty}\frac{df(\varepsilon - eV)}{d(eV)}\int_{SS'}\rho_s^*(\mathbf{r_t}-\mathbf{r},\mathbf{r_t'}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV)d\varepsilon dS dS'\right\}, \tag{39}$$

which, again, can be written as follows:

$$\frac{dI}{d(eV)} = \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2 \frac{1}{\pi^2}\left\{\int_{-\infty}^{+\infty}[f(\varepsilon - eV) - f(\varepsilon)]\int_{SS'}\rho_s(\mathbf{r_t'}-\mathbf{r},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\frac{\partial}{\partial(eV)}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV)d\varepsilon dS dS' + \int_{-\infty}^{+\infty}\frac{df(\varepsilon - eV)}{d(eV)}\int_{SS'}\rho_s(\mathbf{r_t'}-\mathbf{r},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV)d\varepsilon dS dS'\right\}. \tag{40}$$

Equation (39) can be somewhat simplified by considering again the limit $T \to 0$, valid if T is well below the Fermi energy. Since $\lim_{T\to 0} df'(\varepsilon - eV)/d(eV) \to \delta(\varepsilon - eV)$ (see also the Appendix) Eq. (39) becomes

$$\frac{dI}{d(eV)} \simeq \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2 \frac{1}{\pi^2}\int_0^{eV}\int_{SS'}\rho_s^*(\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\frac{\partial}{\partial(eV)} \times \rho_t(\varepsilon - eV)d\varepsilon dS dS' + \frac{4\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2 \frac{1}{\pi^2}\int_{SS'}\rho_s^*(eV)\frac{\partial^2}{\partial n_t \partial n_t'}\rho_t(0)dS dS', \tag{41}$$

where, in order to simplify the notations, all spatial arguments have been omitted. We recall that the quantities appearing in these equations are referred to the system in the presence of the applied bias. If the unperturbed spectral densities are known, the variation $\Delta\rho$ of the spectral densities due to a given perturbation V can be obtained by the general formula[48]

$$\Delta\rho = -\frac{1}{\pi}\lim_{\eta\to 0^+}\mathrm{Im}\left[\frac{\partial}{\partial\varepsilon}\ln\mathrm{Det}(1 - VG_0^R)\right], \tag{42}$$

where G_0^R is the retarded Green's function of the unperturbed system (tip and/or sample).

In principle, the theory presented in this section can be used for the interpretations of STS data, with the goal of extracting information about the sample LDOS, but it can also be compared with simpler descriptions. This second approach will be discussed in Sec. IV B.

A. Comparison with more general theories

As already anticipated in Sec. II, the TH formalism is based on some hypotheses not always easy to justify. Moreover, the theory presented is restricted to the case of noninteracting particles. Therefore, in order to check and verify the validity of the TH treatment developed in this section, a comparison of its results with more general theories is certainly appropriate.

Within the TH framework, Appelbaum and Brinkman[54] derived the following expression for the tunneling current in the more general case of interacting particles, using Green's function technique:

$$I = \frac{\pi e}{\hbar}\left(\frac{\hbar^2}{2m}\right)^2 \frac{1}{\pi^2}\int_{-\infty}^{+\infty}[f(\varepsilon - eV) - f(\varepsilon)]\int_{SS'}\rho_s(\mathbf{r_t'}-\mathbf{r'},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV) + \frac{\partial^2}{\partial n_t \partial n_t'}\rho_s(\mathbf{r_t'}-\mathbf{r'},\mathbf{r_t}-\mathbf{r},\varepsilon)\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV) - \frac{\partial}{\partial n_t'}\rho_s(\mathbf{r_t'}-\mathbf{r'},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial}{\partial n_t}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV) - \frac{\partial}{\partial n_t}\rho_s(\mathbf{r_t'}-\mathbf{r'},\mathbf{r_t}-\mathbf{r},\varepsilon)\frac{\partial}{\partial n_t'}\rho_t(\mathbf{r_t},\mathbf{r_t'},\varepsilon - eV)d\varepsilon dS dS'. \tag{43}$$

This would be the expression which can actually be obtained from Eqs. (4)–(9) following the approach described in this section, but without introducing the consistency condition, Eq. (34), which considerably simplifies the final result, Eq. (37). It has to be remarked that this condition, which should be imposed for consistency using the TH approach, is not adopted in most of the TH treatments present in the literature.

We now consider in further detail the problem of the connection between the TH theory and more general descriptions, which has already been introduced in Sec. II. Probably the most complete investigation of the tunneling phenomena in a solid-vacuum-solid interface has been performed before the invention of the STM/STS, in a series of papers by Feuchtwang in the 1970s,[44,46,55–57] making use of the Green's function formalism and, in particular, the Keldish theory for nonequilibrium processes.[8–10] In this approach, the full, nonequilibrium Green's function of the junction is derived and expressed in terms of the Green's functions of the uncoupled systems (which define the zeroth order system) and of a pseudo-Hamiltonian operator whose matrix elements at the lowest order of the perturbation expansion coincide actually with the transition current matrix element of the TH theory, namely, with Eq. (9): In this framework, it is demonstrated that this operator, which appears naturally in the theory, assures the continuity of the full Green's function across the interface. The current is then calculated, starting from its general quantum-field statistical expression, from the Green's functions of the system. In the general three-dimensional case, this theory predicts an expression for the tunneling current which reduces to Eq. (43) if the full Green's function is calculated up to first order in the pseudo-Hamiltonian operator.

In conclusion, it is possible to say that the results of the theory presented in this section are, within the limitations clarified, confirmed by more profound investigations. This means that the implications of our analysis can be retained to be valid also under more general hypotheses, namely, also in the case of interacting particles and without the subtle doubts posed by the transfer Hamiltonian approach.

B. Comparison with one-dimensional WKB-based scanning tunneling spectroscopy theory

We now analyze the meaning of the relations derived in this section and of the physical quantities involved by comparing them, in particular, Eqs. (37), (39), and (41), with the expressions obtained by WKB theory, briefly described in Sec. III A [namely, Eqs. (26)–(28)]. Several interesting conclusions can be derived, which complement the discussion of Sec. III A.

Actually, Eq. (37) shows that in the most general case, the current is not simply given by an energetic convolution between sample and tip density of states, weighted by the transmission coefficient. First of all, instead of the density of states, the physical properties of the system are described by the sample and tip spectral densities, as already previously noted.[15] Moreover, in Eq. (37) a transmission coefficient is not explicitly present at all, because the exponential decay characteristic of tunneling phenomena is, again, implicitly incorporated in the spatial dependence of the sample and tip spectral densities. Finally, in Eqs. (26) and (28), it is evident that the low temperature limit is assumed.

With respect to the low bias potential limit,[15] in the general case it is not possible to uniquely identify an average quantity which can play the role of the mean local density of state for the sample in Eqs. (26) and (28). Since the crucial experimental quantity for STS is the first derivative of the current, it is useful to directly establish a correspondence between Eqs. (28) and (41) introducing the following definition of mean local density of state for the sample at the energy eV:

$$\begin{aligned}\bar{\rho}_s(-\mathbf{r},eV) &\equiv \left[\int_{SS'} \frac{\partial^2}{\partial n_t \partial n_t'} \rho_t(\mathbf{r_t},\mathbf{r_t'},0)dSdS'\right]^{-1} \\ &\times \int_{SS'} \rho_s(\mathbf{r_t'}-\mathbf{r},\mathbf{r_t}-\mathbf{r},eV) \\ &\times \frac{\partial^2}{\partial n_t \partial n_t'} \rho_t(\mathbf{r_t},\mathbf{r_t'},0)dSdS' \\ &\equiv \frac{1}{\bar{\rho}_t(0)} \int_{SS'} \rho_s(\mathbf{r_t'}-\mathbf{r},\mathbf{r_t}-\mathbf{r},eV) \\ &\times \frac{\partial^2}{\partial n_t \partial n_t'} \rho_t(\mathbf{r_t},\mathbf{r_t'},0)dSdS', \end{aligned} \tag{44}$$

where the mean local density of state at the Fermi energy for the tip, $\bar{\rho}_t(0)=\int_{SS'} \frac{\partial^2}{\partial n_t \partial n_t'} \rho_t(\mathbf{r_t},\mathbf{r_t'},0)dSdS'$, has been also defined. In this way, the structure of the two last terms of Eqs. (28) and (41) are similar, keeping in mind that, as already noted, in the TH formalism the exponentially decaying behavior is contained in the properties of the spectral densities.

On the basis of the theory developed and the above discussion, it can be argued that in general in a STS experiment what is possible to measure is a properly defined mean local density of states of the sample, at the energy level selected by the bias potential: This effective sample LDOS is actually a double spatial convolution between sample and tip spectral densities, evaluated at the bias potential and at the Fermi energy, respectively.

V. CONCLUSIONS

The possibility of measuring the local (down to atomic resolution) electronic density of states of a surface [scanning tunneling spectroscopy (STS)] is one among the most attractive features of the scanning tunneling microscope. To this goal a full and thorough knowledge of the quantum states of the system (sample and tip) subjected to a bias voltage and, thus, crossed by a tunneling current, would be required. As such a complete and exact description is not attainable, smart and physically sound approximations must be adopted. Particularly crucial is the feedback of the applied voltage onto the effective interface barrier height and vacuum states. In the first part we have extended the modified Bardeen approach to the case of finite voltage and arbitrary temperature and obtained analytical expressions for the derivative of the tunneling current vs voltage making only general assumptions on the tip density of states. Following this path, we are naturally led to suggest the best normalization for the above derivative. A specific simply workable example shows the relevant differences with respect to previous approaches. In this paper we have considered only the contribution of the electrons occupying the ground state levels of the tip and

flowing into the empty states of the sample through the tunneling current. Even though this is the most favorable situation to extract the sample information from the STS data, also the reverse process (corresponding to a negative bias) should be included in a complete theory, as it is foreseen for a next work. In the second part we have provided a generalization of the results by Feuchtwang and Cutler in terms of theoretical spectral densities and Green's functions of the system. We believe people critically using STS to study real surfaces should benefit from both these methodologies.

ACKNOWLEDGMENTS

The authors wish to thank A. Li Bassi and C. S. Casari for very useful discussions and comments on several aspects of this work and for carefully reading the manuscript.

APPENDIX: APPROXIMATE TREATMENT OF THERMAL EFFECTS

Let us consider the following integral:

$$\int_{-\infty}^{+\infty} f(\varepsilon)H(\varepsilon)d\varepsilon, \tag{A1}$$

here $H(\varepsilon)$ is a generic function of the variable ε while $f(\varepsilon) = [\exp(\varepsilon/k_BT)+1]^{-1}$. It can be shown that the integral can be expressed in terms of the so-called Sommerfeld expansion[58]

$$\int_{-\infty}^{+\infty} f(\varepsilon)H(\varepsilon)d\varepsilon = \int_{-\infty}^{0} f(\varepsilon)H(\varepsilon)d\varepsilon + \sum_{n=1}^{+\infty} a_n(k_BT)^{2n}\frac{d^{2n-1}}{d\varepsilon^{2n-1}}H(\varepsilon)|_{\varepsilon=0}, \tag{A2}$$

where the dimensionless numbers a_n are given by the following intergrals:

$$a_n = \int_{-\infty}^{+\infty} \frac{x^{2n}}{(2n)!}\left[-\frac{d}{dx}(\exp(x)+1)^{-1}\right]dx. \tag{A3}$$

In most cases, the first term of the series, for which $a_1 = \pi^2/6$, is able to describe thermal effects satisfactorily. Consider now the expression

$$\int_{-\infty}^{+\infty} [f(\varepsilon - eV) - f(\varepsilon)]H(\varepsilon)d\varepsilon; \tag{A4}$$

from Eq. (A2), keeping only the first term of the expansion we obtain

$$\int_{-\infty}^{+\infty} [f(\varepsilon - eV) - f(\varepsilon)]H(\varepsilon)d\varepsilon \approx \int_{0}^{eV} H(\varepsilon)d\varepsilon + \frac{\pi^2}{6}(k_BT)^2 \times \left[\left.\frac{dH(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=eV} - \left.\frac{dH(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}\right]. \tag{A5}$$

The first term on the right hand side of this equation gives the usual zero temperature approximation, while the second describes thermal corrections. It can be used in connection with the descriptions developed in this paper for the introduction of finite temperature corrections to the tunneling current I. In particular, upon comparison with Eq. (4) of Sec. II and Eq. (38) of Sec. IV, we can identify the function $H(\varepsilon)$ with the quantities $\Sigma_{\mu,\nu}|M_{s\nu,t\mu}|^2\rho_t(\varepsilon - eV,\mu)\rho_s(\varepsilon,\nu)$ and $\int_{SS'}\rho_s(\mathbf{r}'_\mathbf{t}-\mathbf{r},\mathbf{r}_\mathbf{t}-\mathbf{r},\varepsilon)\frac{\partial^2}{\partial n_t \partial n'_t}\rho_t(\mathbf{r}_\mathbf{t},\mathbf{r}'_\mathbf{t},\varepsilon - eV)dSdS'$, respectively. As a very raw approximation, assume that these expressions are proportional to the sample LDOS (the limits of this sentence are discussed in detail in this paper); then, the inclusion of thermal effects adds to the dI/dV a term which is roughly proportional to $T^2\rho_s''(eV)$. For example, for a Gaussian peak, such a term will produce an enlargement and a reduction of its intensity. The results obtained in this paper can be used to make a quantitative and more general analysis of this issue.

[1] *Scanning Tunneling Microscopy*, edited by J. A. Stroscio and W. J. Kaiser (Academic, Boston, MA, 1993).
[2] C. Bai, *Scanning Tunneling Microscopy and its Applications*, 2nd revised ed. (Springer, New York, 2000).
[3] *Scanning Probe Microscopy and Spectroscopy*, edited by D. Bonnell (Wiley, New York, 2001).
[4] T. E. Feuchtwang and P. H. Cutler, Phys. Scr. **35**, 132 (1987).
[5] G. A. D. Briggs and A. J. Fisher, Surf. Sci. Rep. **33**, 1 (1999).
[6] J. M. Blanco, F. Flores, and R. Pèrez, Prog. Surf. Sci. **81**, 403 (2006).
[7] R. Landauer, IBM J. Res. Dev. **1**, 233 (1957); M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).
[8] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).
[9] L. V. Keldish, Sov. Phys. JETP **20**, 1018 (1965).
[10] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Pergamon, Oxford, 1981).
[11] J. R. Oppenheimer, Phys. Rev. **31**, 66 (1928).
[12] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
[13] C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969).
[14] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983); Phys. Rev. B **31**, 805 (1985).
[15] T. E. Feuchtwang and P. H. Cutler, Phys. Scr. **38**, 252 (1988).
[16] C. J. Chen, Phys. Rev. B **42**, 8841 (1990).
[17] C. J. Chen, Phys. Rev. Lett. **65**, 448 (1990).
[18] N. D. Lang, Phys. Rev. B **34**, 5947 (1986).
[19] M. Tsukada and N. Shima, J. Phys. Soc. Jpn. **56**, 2875 (1987).
[20] G. Hörmandinger, Phys. Rev. B **49**, 13897 (1994).
[21] G. K. Ivanov, M. A. Kozhushner, and I. I. Oleinik, Surf. Sci. **331**, 1191 (1995).
[22] S. Heinze, S. Blügel, R. Pascal, M. Bode, and R. Wiesendanger, Phys. Rev. B **58**, 16432 (1998).

[23] W. A. Hofer and J. Redinger, Surf. Sci. **447**, 51 (2000); W. A. Hofer, J. Redinger, A. Biedermann, and P. Varga, *ibid.* **466**, L795 (2000); W. A. Hofer and A. Garcia-Lekue, Phys. Rev. B **71**, 085401 (2005).
[24] A. Selloni, P. Carnevali, E. Tosatti, and C. D. Chen, Phys. Rev. B **31**, 2602 (1985).
[25] V. A. Ukraintsev, Phys. Rev. B **53**, 11176 (1996).
[26] N. Li, M. Zinke-Allmang, and H. Iwasaki, Surf. Sci. **554**, 253 (2004).
[27] J. Ferrer, A. Martín-Rodero, and F. Flores, Phys. Rev. B **38**, 10113 (1988).
[28] C. Noguera, Phys. Rev. B **42**, 1629 (1990).
[29] W. Sacks and C. Noguera, Phys. Rev. B **43**, 11612 (1991); W. Sacks, *ibid.* **61**, 7656 (2000).
[30] H. Ness and A. J. Fisher, Phys. Rev. B **56**, 12469 (1997).
[31] J. Cerdá, M. A. Van Hove, P. Sautet, and M. Salmeron, Phys. Rev. B **56**, 15885 (1997).
[32] H. Ishida, D. Wortmann, and T. Ohwaki, Phys. Rev. B **70**, 085409 (2004).
[33] K. Palotàs and W. A. Hofer, J. Phys.: Condens. Matter **17**, 2705 (2005).
[34] A. A. Lucas, H. Morawitz, G. R. Henry, J.-P. Vigneron, Ph. Lambin, P. H. Cutler, and T. E. Feuchtwang, Phys. Rev. B **37**, 10708 (1988).
[35] G. Doyen, D. Drakova, and M. Scheffler, Phys. Rev. B **47**, 9778 (1993).
[36] J. Cerdá, A. Yoon, M. A. Van Hove, P. Sautet, M. Salmeron, and G. A. Somorjai, Phys. Rev. B **56**, 15900 (1997).
[37] S. Crampin, J. Electron Spectrosc. Relat. Phenom. **109**, 51 (2000).
[38] W. Kaminski and L. Jurczyszyn, Surf. Sci. **566**, 29 (2004).
[39] M. De Menech, U. Saalmann, and M. E. Garcia, Phys. Rev. B **73**, 155407 (2006).
[40] E. Tekman, Phys. Rev. B **46**, 4938 (1992).
[41] C. Bracher, M. Riza, and M. Kleber, Phys. Rev. B **56**, 7704 (1997).
[42] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **72**, 235113 (2005).
[43] E. N. Economu, *Green's Functions in Quantum Physics* (Springer, New York, 1983); A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (Dover, New York, 2003).
[44] T. E. Feuchtwang, Phys. Rev. B **10**, 4121 (1974).
[45] C. J. Chen, *Introduction to Scanning Tunneling Microscopy* (Oxford University Press, Oxford, 1993).
[46] T. E. Feuchtwang, Phys. Rev. B **13**, 517 (1976).
[47] J. B. Pendry, A. B. Prêtre, and B. C. H. Krutzen, J. Phys.: Condens. Matter **3**, 4313 (1991).
[48] M. C. Desjonquères and D. Spanjaard, *Concepts in Surface Science*, 2nd ed. (Springer, New York, 1996).
[49] W. A. Hofer, A. S. Foster, and A. L. Shluger, Rev. Mod. Phys. **75**, 1287 (2003).
[50] W. A. Hofer, J. Redinger, and P. Varga, Solid State Commun. **113**, 245 (2000); W. A. Hofer, J. Redinger, and R. Podloucky, Phys. Rev. B **64**, 125108 (2001).
[51] M. M. J. Bischoff, T. Yamada, A. J. Quinn, and H. van Kempen, Surf. Sci. **501**, 155 (2002).
[52] T. K. Yamada, M. M. J. Bischoff, G. M. M. Heijnen, T. Mizoguchi, and H. van Kempen, Phys. Rev. Lett. **90**, 056803 (2003).
[53] B. Koslowski, C. Dietrich, A. Tschetschetkin, and P. Ziemann, Phys. Rev. B **75**, 035421 (2007).
[54] J. A. Appelbaum and W. F. Brinkman, Phys. Rev. **186**, 464 (1969).
[55] T. E. Feuchtwang, Phys. Rev. B **10**, 4135 (1974).
[56] T. E. Feuchtwang, Phys. Rev. B **12**, 3979 (1975).
[57] T. E. Feuchtwang, Phys. Rev. B **20**, 430 (1979).
[58] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, PA, 1976).