Theoretical study of transport through a quantum point contact

E. Tekman

Department of Physics, Bilkent University, Bilkent 06533, Ankara, Turkey

S. Ciraci*

Department of Physics, Bilkent University, Bilkent 06533, Ankara, Turkey
and IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland
(Received 1 October 1990)

We developed a formalism within the linear-response theory to investigate the transport through a quantum point contact between two electron-gas reservoirs. It is valid for two-terminal conductance through a constriction of a two-dimensional (2D) or 3D potential and has a wide range of applicability covering ballistic as well as tunneling regimes. We studied the quantization of conductance and examined several effects influencing the quantum transmission. Among these effects we found that the simple phase relation results in resonance structures superimposed on the plateaus between two steps of quantized conductance. These resonances are destroyed by the smooth entrance, finite temperature and bias, and variation of the potential. The simulation of adiabatic transmission in constrictions having smoothly varying widths resulted in the conductance with sharp quantum steps without the resonance structure. The quality of quantization is strongly affected by the length of constriction, Fermi-level smearing, the obstacle at the entrance, impurity scattering, nonuniformities of geometry and potential, and in particular by the variation of the longitudinal potential resulting in a sharp saddle-point structure. The quasibound states may occur in a local widening of the width or in a locally lowered potential. These states give rise to a sudden increase of the transmission prior to the opening of a new conduction channel. We present an extensive analysis of this phenomenon and show that it is due to resonant tunneling through these bound states. Owing to enhanced backscattering, the bound states of an attractive impurity in a constriction can yield dips in the conductance at the threshold of channels. In addition to quantized ballistic transport, we extended our method to treat the transport mechanism in scanning tunneling microscopy and in field emission of collimated electrons from an atomic-size source. The issues of current interest in these fields that we treated are (i) the transition from the tunneling to the ballistic regime and the interpretation of conductance oscillations, and (ii) the anomalous corrugation of flat metal surfaces. Our results reveal crucial features of the lateral confinement of the current-transporting states in the constriction of potential between the tip and sample. The effective barriers created from this confinement effect dominate the transmission at small tip-sample distances and influence the apparent barrier height.

I. INTRODUCTION

Advances in the growth techniques and new electronic materials developed therefrom have provided almost defect-free electronic devices, which have dimensions in one or more directions on the quantum scale.[1] New quantum regimes governing such systems of lower dimensionality have led to novel electronic properties with potential applications. Quantum wells, wires, and dots, which have been implemented in the terminology of condensed-matter physics, indicate not only different dimensionality but also exhibit dramatically different electronic properties under both zero and finite magnetic fields. The electronic transport properties in lower dimensionality also have several important features, which have attracted a wide range of both experimental and theoretical interest. Just one example of this is the tremendous progress made recently for the resonant tunneling[2] in a double-barrier structure, which displays a number of interesting properties such as negative differential resistance and bistability.

Some of the transport studies carried out on small devices go beyond the tunneling regime. If the size of the sample (or device) is smaller than the phase breaking length, the transport is not *incoherent*, but may have a *diffusive* character. In this case, electrons have a well-defined phase throughout the device, even though they may experience elastic scattering. The small size of these devices allows the observation of important quantum interference effects such as the Aharonov-Bohm[3] effect and universal conductance fluctuations.[4] Numerous publications on this type of transport have appeared, thus contributing to a field called *mesoscopics*, a term that indicates a new length scale for physical events between macroscopic and microscopic. As early as 1957, Landauer[5] proposed that the conduction in a solid is a scattering event, and that transport is the consequence of the incident current flux. Based on the counting (self-

consistency) arguments of reflections and transmissions, he derived his famous formula for a one-dimensional conductor yielding the conductance $G=(2e^2/h)(T/R)$, with T and $R=(1-T)$ being the transmission and reflection coefficients, respectively. His views have made a great impact on the physics of mesoscopic systems. In an effort to rederive his formula and also to extend it to higher dimensions and to many channels, new concepts were implemented in the field.[6-8] For example, Engquist and Anderson[9] argued that the conductance has close bearing on the type of measurement. Thus, differences between two-, four-, and multiterminal measurements were clarified. Efforts to obtain the Landauer formula from the linear-response theory or from the Kubo formula, however, resulted in different expressions for single[10] and for multichannel systems.[11,12] These theories (which we call two-terminal theories) predict that the conductance is proportional to T [or $\mathrm{Tr}(\tilde{T}_t\tilde{T}_t^\dagger)$, $\tilde{T}_t$ being the matrix of transmission amplitudes of size $N_c \times N_c$ for an N_c-channel system], and conform with Landauer's original formula for a 1D single-channel conductor, as well as its extension to a multichannel system at the limit of $T \ll 1$. However, these theories[10,12] are seemingly at variance with Landauer's formula at the other limit $T \rightarrow 1$ [or for a multichannel system $\mathrm{Tr}(\tilde{T}_t\tilde{T}_f^\dagger) \rightarrow N_c$], and predict the quantum of conductance $2e^2/h$ (or the N_c multiple of it when N_c channels are opened) in the absence of scatterers in the 1D conductors. In response to these arguments, Landauer[13] pointed out that, even if all point scatterers in a 1D conductor were eliminated, the variation of the potential due to the self-consistent charge distribution is still a source of scattering. He also emphasized that the source of discrepancy among various theories lies in the type of probing.[13] Presently, it appears that the new multiprobe generalization proposed by Büttiker[14] seems to yield the correct answer in many controversial issues of ballistic transport. His formula includes not only proper asymmetry in the presence of a magnetic field, but also leads to global Onsager-Casimir symmetry relations.

Recent experiments performed independently by van Wees et al.[15] and Wharam et al.[16] have been a breakthrough in the field of quantum ballistic transport in a quantum point contact (QPC) in a two-dimensional electron gas (2D EG). Using high-mobility GaAs-Al$_{1-x}$Ga$_x$As heterojunctions and the split-gate technique, they imposed a small constriction on the sample. A channel was obtained from this constriction by applying a negative bias to the split gate, and thus by causing the depletion of electrons beneath the gate. Thus, the portion of the 2D EG lying below the gap of the split-gate electrode remains conducting. In their experiments the length of the constriction d is smaller than the electron mean free path l_e, so that electrons are prevented from being scattered in and around the constriction. The width of the constriction w is also in the range of the Fermi wavelength λ_F, whereby quantum size effects[17] become relevant. At very low temperatures ($T \simeq 0.6$ K), the two-terminal conductance of the quantum point contact, G, was found to change with the gate voltage V_g (or equivalently with w) approximately in units of $2e^2/h$.

This observation was interpreted as the *quantization* of the conductance. It was also argued that the predictions of the two-terminal theories for a 1D ballistic conductor are confirmed by this experimental interpretation. The transport through a point contact also emerged in seemingly unrelated fields. For example, the oscillations of the current, which were measured with a scanning tunneling microscope operating with a tip-sample distance that varied in the range of mechanical contact,[18] were first attributed to a quantization of conductance.[19] In the same context, the physics underlying the field emission of the focused electrons[20] was found to be closely related to the transport through a quantum point contact.

Almost three decades ago, Sharvin[21] pointed out the resistance of a ballistic channel (or point contact) and developed a formalism in the semiclassical regime. Earlier, the quantization of conductance in a point contact was not considered as a possibility because it was thought that the quantization is hindered by various intervening geometrical and material effects. It was, however, indigenous to the two-terminal theories.[10-12] Büttiker et al.[22] also pointed out the possibility that the crossing of a transverse level leads to sharp changes in the conductance. The experiments now clearly show that the quantization is achievable with an accuracy of 1% despite the varying system parameters. This is, of course, far from being coincidental. On the other hand, the quantum of conductance can routinely be measured with an accuracy of one part in 10^7 in the experiments related to quantum Hall effects. Hence, 1% deviation from the exact value implies that the quantization in the QPC is highly distorted. The step structure in the conductance curve can easily be visualized in terms of a new conduction channel opened by a subband dipping into the Fermi level, but an elaborate theory was required to reveal the crucial features of the transport. In fact, questions as how the behavior of the conductance depends on temperature and on the variation of the potential in the constriction, and how the geometry of the contact—in particular the form of the connection to the reservoirs—affects the quantization required detailed analysis. This analysis, as well as the quantitative study of a small but finite deviation from the exact quantization, have been addressed by recent theoretical studies, including ours.[23-32]

Several theories have been developed based on the assumption that the transport in a QPC is ballistic as suggested by experiment.[15,16,33] While the adiabatic evolution of the current-transporting states without undergoing intersubband scattering was initially foreseen in a constriction[24] (allowing only very smooth, continuous variation of w over a scale of the order of the electron wavelength), the emphasis has been placed on the critical effects of the boundaries between the reservoirs and the constriction. In fact, it was shown theoretically[23,26-29] that the interference of the coherent electron states reflected elastically from the boundaries gives rise to the resonance structure superimposed on the plateaus (i.e., the constant value of G between two consecutive quantum steps). That this resonance structure was not observed in the experimental data was attributed to the finite temperature effects or to the scattering that des-

troyed the phase coherence.[29] There have even been claims that the transport is not ballistic.[34] In the meantime, the nonadditivity of conductance measured on two consecutive constrictions separated by an electron gas[33] has been taken as clear evidence for the ballisticity of the transport.[26]

The transport in a QPC is the subject of the present study. We developed a formalism by using simple linear-response arguments together with the solution of the Schrödinger equation to provide an explanation for the transport in a QPC. It has a range of applicability covering ballistic as well as tunneling regimes. Some of our studies based on this formalism were published earlier as short communications.[29,35,36] The purpose of this paper is twofold: the first objective is to present a comprehensive description of our formalism and investigate several features of the transport in a QPC, not yet taken into account. Our second objective is to apply our method to the study of some important problems in other fields, which have a close bearing on the transport events studied in this work. For example, scanning tunneling microscopy (STM) (Ref. 37) is based on the fact that the tunneling current is exponentially dependent on the thickness of the potential barrier. Owing to this high sensitivity of the tunneling current to the distance between the probing tip and sample, STM has become an excellent tool for imaging the local electronic density of states, and hence the surface atomic structure, imperfections, and variations of the electric and magnetic fields, etc., without invoking translational periodicity. On the other hand, we point out important features of STM, which pertain to mesoscopic physics. It has now become clear from recent works that the stable atomically sharp tips can be fabricated;[20] the cross section relevant for electron transmission is in the range of atomic dimensions or the Fermi wavelength. At large tip-sample distances the potential energy varies along the z direction and has a finite barrier, and perhaps has an approximately rotational symmetry in the xy plane. As the tip approaches the sample, the effective potential barrier collapses even before the electrodes engage in a mechanical contact,[38,39] so the character of the transport can change from the tunneling to the ballistic regime.[36] Since our approach starts with the potential between the tip and sample surface, it is not restricted to the tunneling regime in contrast to the theories that make use of the electronic states of the bare surface. Also, since our theory has validity in a range covering *tunneling as well as ballistic* regimes, physical events in both regimes and the transition between them can be successfully addressed with the present approach. The effective potential barrier,[36,39,40] generated due to the lateral confinement of states between tip and sample, can easily be visualized within the framework of our method, and is seen to have important implications. In this respect, the present paper is intended to span various fields and create new interest in using STM to study mesoscopic events.

The subjects are treated in the following order. In Sec. II, the basic principles of the theoretical method including the underlying approximations are explained. In the same context, the extension of the formalism to allow self-consistent-field (SCF) calculations are discussed. In the same section, the theoretical frameworks of the other methods are outlined and compared with that of the present method. In Sec. III the expressions of conductance are obtained for a finite, uniform constriction by using two types of constriction potential. An analogy with 1D transport is also established in order to attribute simpler physical meanings to the terms in the expression of conductance. A detailed account of the transmission resonance structure is given, and the effects of finite temperature and voltage on the resonances are examined. The actual form of a split gate and the potential created in it cannot be accessed experimentally, we nevertheless consider possible geometrical features by investigating model nonuniform constrictions in Sec. IV. To deal with the varying widths and potential in the constriction, we first integrate the transfer-matrix method into our theoretical framework. Here we study a wedgelike entrance (or tapering), slowly varying widths leading to adiabatic evolution, surface roughness in the constriction, and quasibound states. The elastic scattering from a point impurity is studied in Sec. V and important results derived therefrom are used in Sec. VI to examine a resonant tunneling effect in a QPC. The effect of the variation of the potential in the constriction is explored by considering two different saddle-point structures in Sec. VII. The extension of the formalism to cover STM and field emission of collimated electrons and the investigation of current problems are presented in Sec. VIII. We propose a model that provides an explanation for different observed behaviors of current (or conductance) measured as a function of tip-sample distance near the mechanical contact, specifically the nature of large-period "quantum" oscillations and saturations at the first plateau. We also address the anomalous corrugation of flat metal surfaces obtained from STM. Our results suggest that the site dependence of the width of the potential barrier becomes pronounced due to the tip-sample interaction. This site dependence is actually imaged by the tip to result in a relatively larger corrugation. Finally, we summarize the important findings and comment on the current issues in the field in our concluding remarks.

II. THEORETICAL APPROACH

In the theoretical models[15,16] initially used to explain the quantization of conductance, the QPC was perceived as a uniform wave guide, and only the events in this wave guide were taken into account. The current-transporting states are laterally confined in the waveguide, the width of which is in the range of the Fermi wavelength. Then, the transverse momenta of these states are quantized, resulting in a *subband structure*. The wave propagation (or transport) in subband n takes place, as long as the minimum of band ε_n is smaller than the Fermi level E_F. Upon application of a small bias voltage, a small difference $\Delta\mu$ between the chemical potentials of two reservoirs connected by a constriction is obtained. Then, each occupied subband contributes to the current by an amount $I_n = e v_n(E_F) D_n(E_F) \Delta\mu$, where $v_n(E_F)$ is the velocity of the electron in the nth subband, and $D_n(E_F)$ is

the density of states, both calculated at the Fermi level. As far as the carrier flow is concerned, the system at hand is effectively one dimensional, and thus the product $v_n(E)D_n(E)$ is independent of the energy E and the subband index n, but is equal to $e/\pi\hbar$. Consequently, each occupied subband contributes an equal amount, leading to the expression of conductivity pertaining to the two-terminal measurement $G = 2e^2N_c/h$, where N_c is the number of occupied subbands. According to this description, the conductance is quantized such that it increases by a quantum of conductance $2e^2/h$ whenever a subband dips below the Fermi level. This can be achieved either by widening w of the QPC (and thus by lowering the subband energies) or by increasing the density of electrons (and thus by raising E_F). Similar quantum size effects in the work function and surface energy of very thin metal films were treated earlier for 2D metals.[41] Note that the above simple model conjectures an ideal step structure and can be valid only if the waveguide is perfectly uniform (as far as the potential is concerned) and *infinitely long*. Hence, the effects originating from nonuniformities, especially from boundaries where the waveguide joins the 2D EG reservoirs, are neglected and the evanescent waves are prevented from contributing to the transport. This certainly has no bearing on the real experimental setup, and the experimental data themselves deviate somewhat from this "ideal step structure" (they are device dependent but reproducible for a given device). Furthermore, experiments were performed with QPC's having a finite length d, in particular with $d < l_e$. This normally gives rise to the reflections of waves from the boundaries. Despite the potential induced by charge depletion in the split gate being smooth, the QPC is by no means uniform in width nor in the variation of the potential along the channel.

Our objective is to develop a formalism that provides a proper treatment of events resulting in a "quantized" conductance in a ballistic QPC, and allows a systematic study of several effects. In our approach, the current-carrying states are obtained by solving the Schrödinger equation in the channel (or waveguide), and also by matching these solutions at the boundaries between adjacent regions. We group our method[29] together with those developed in the same context[23,26] as *boundary-matching techniques*. We now present a detailed description of the method.

As the above simple description implies, the essential feature responsible for the quantization of the conductance in the QPC is the quasi-1D nature of the system, which yields subbands dispersing along the propagation direction. In order to emphasize the similarities with the above description, we explicitly include the quasi-1D character of transport in our formalism. This is done by separating the space into three parts. The leftmost and rightmost parts are the 2D EG's that are connected to the reservoirs. The EG is semi-infinite, and the 2D free-electron wave functions are solutions of the Schrödinger equation. The central part is the constriction, which is characterized by a laterally confining potential. Consequently, the solutions of the Schrödinger equation in the constriction are the subband wave functions arising from

the quantization of the transverse momentum. The separation of the space into the 2D EG and constriction can be represented in the Hamiltonian by using the potential energy $V(y,z)$ given by

$$V(y,z) = [\phi_m(z) + V_c(y,z)]\Theta(z)\Theta(d-z) . \quad (1)$$

In the constriction $(0 < z < d)$, the potential is broken down into two components. The *longitudinal* part $\phi_m(z)$ contains the variation of the minimum value of the potential along the direction of propagation. The *confining* part $V_c(y,z)$, on the other hand, gives rise to the subband structure. The potential is assumed to be zero in both the left- $(z < 0)$ and right-hand side $(z > d)$ 2D EG. Geometrical parameters relevant to a QPC and the longitudinal and confining potentials are schematically illustrated in Fig. 1. For a general constriction potential $V(y,z)$, the above decomposition may not be straightforward. In our analysis, we start with the component potentials in order to clarify the roles and effects of $\phi_m(z)$ and $V_c(y,z)$. The Hamiltonian reduces merely to the kinetic-energy term in the 2D EG. The subband wave functions in the constriction are calculated from the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + \phi_m(z) - \frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial y^2} + V_c(y,z) \right] \xi_{nE(y,z)}$$

$$= E\xi_{nE(y,z)} , \quad (2)$$

where we assumed that the energy spectrum is continuous due to the *free* propagation along the z direction. For the same reasons, the states are twofold degenerate, so that ξ_{nE} and $\bar{\xi}_{nE}$ satisfy Eq. (2) for left- and right-moving current density, respectively. Basically, the partial differential equation in Eq. (2) is not separable, but in principle it can be integrated numerically. Analytical solutions are available only for special potential profiles. An approximation scheme to Eq. (2) will be presented in Sec. IV. We now proceed with the calculation of conductance by assuming that the subband wave functions
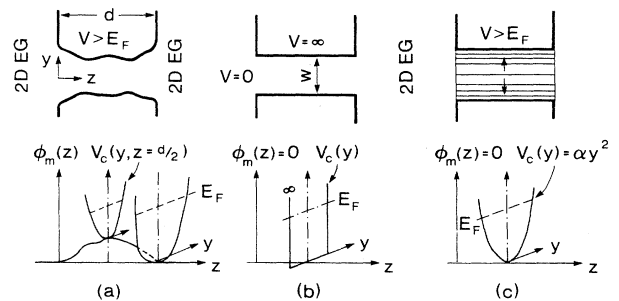

FIG. 1. Top panels: Schematic description of quasi-1D constrictions of potential $V(y,z)$ between two 2D EG's. The z and y directions are propagation and transverse confinement directions, respectively. The length of constriction is d. Bottom panels: variation of the potential inside the constrictions. (a) A general constriction with potential given in Eq. (1). (b) Uniform constriction with an infinite-well confining potential. (c) Uniform constriction with a quadratic confining potential.

$\xi_{nE}(y,z)$ and $\bar{\xi}_{nE}(y,z)$ are known.

Since the electrons in transport measurements originate from one 2D EG and are collected in the other, the subband wave functions obtained from Eq. (2) have to be matched to the 2D plane waves in the EG at the boundaries ($z=0$ and $z=d$), with the usual boundary conditions. The boundary conditions at $z \rightarrow \pm \infty$ fix the incoming and outgoing waves. To determine the current-transporting state in the entire space, we start with an incident plane wave in the left-hand side 2D EG with a wave vector $\mathbf{k}_i = (\kappa_0, k_0)$ and energy $E = \hbar^2 |\mathbf{k}_i|^2 / 2m^*$. Then, the current-transporting wave function $\psi_{\mathbf{k}_i}(y,z)$ is written

$$\psi_{\mathbf{k}_i}(y,z) = \begin{cases} e^{ik_0 z} e^{i\kappa_0 y} + \int d\kappa \, e^{-ik_z(\kappa)z} e^{i\kappa y} A_{\mathbf{k}_i}(\kappa), & z<0 \\ \sum_n [\xi_{nE}(y,z)\Theta_{n\mathbf{k}_i} + \bar{\xi}_{nE}(y,z)\Delta_{n\mathbf{k}_i}], & 0<z<d \\ \int d\kappa \, e^{ik_z(\kappa)z} e^{i\kappa y} B_{\mathbf{k}_i}(\kappa), & d<z, \end{cases} \tag{3}$$

where $k_z^2(\kappa) = 2m^* E / \hbar^2 - \kappa^2$. The imaginary part of $k_z(\kappa)$ is positive in order for the wave functions to assume finite values for $z \rightarrow \pm \infty$. The coefficients $A_{\mathbf{k}_i}(\kappa)$, $B_{\mathbf{k}_i}(\kappa)$, $\Theta_{n\mathbf{k}_i}$, and $\Delta_{n\mathbf{k}_i}$ are obtained from the continuity of $\psi_{\mathbf{k}_i}(y,z)$ and its derivative with respect to z at the boundaries ($z=0, z=d$). The continuity of the derivative along the y direction is guaranteed by the continuity of the wave function itself. At this stage we proceed with the transverse Fourier transform [which is defined as $F(q) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} dy \, e^{-iqy} f(y)$] of the linear equations obtained from continuity conditions to transform in the q space:

$$(2\pi)^{1/2}[\delta(q-\kappa_0) + A_{\mathbf{k}_i}(q)]$$
$$= \sum_n [\Xi_{nE}(q,0)\Theta_{n\mathbf{k}_i} + \bar{\Xi}_{nE}(q,0)\Delta_{n\mathbf{k}_i}], \tag{4}$$

$$(2\pi)^{1/2} e^{ik_z(q)d} B_{\mathbf{k}_i}(q)$$
$$= \sum_n [\Xi_{nE}(q,d)\Theta_{n\mathbf{k}_i} + \bar{\Xi}_{nE}(q,d)\Delta_{n\mathbf{k}_i}], \tag{5}$$

$$(2\pi)^{1/2}[ik_0 \delta(q-\kappa_0) - ik_z(q) A_{\mathbf{k}_i}(q)]$$
$$= \sum_n [\Xi'_{nE}(q,0)\Theta_{n\mathbf{k}_i} + \bar{\Xi}'_{nE}(q,0)\Delta_{n\mathbf{k}_i}], \tag{6}$$

$$(2\pi)^{1/2} e^{ik_z(q)d} ik_z(q) B_{\mathbf{k}_i}(q)$$
$$= \sum_n [\Xi'_{nE}(q,d)\Theta_{n\mathbf{k}_i} + \bar{\Xi}'_{nE}(q,d)\Delta_{n\mathbf{k}_i}], \tag{7}$$

where Ξ and Ξ' indicate the transverse Fourier transform of the subband wave function and its derivative along the z direction, respectively. By eliminating the coefficients of plane waves, $A_{\mathbf{k}_i}(\kappa)$ and $B_{\mathbf{k}_i}(\kappa)$, from Eqs. (4)–(7), one obtains the following relations:

$$(2\pi)^{1/2} 2k_0 \delta(q-\kappa_0) = \sum_n \{ [k_z(q)\Xi_{nE}(q,0)$$
$$-i\Xi'_{nE}(q,0)]\Theta_{n\mathbf{k}_i}$$
$$+ [k_z(q)\bar{\Xi}_{nE}(q,0)$$
$$-i\bar{\Xi}'_{nE}(q,0)]\Delta_{n\mathbf{k}_i} \}, \tag{8}$$

$$\sum_n \{ [k_z(q)\Xi_{nE}(q,d) + i\Xi'_{nE}(q,d)]\Theta_{n\mathbf{k}_i}$$
$$+ [k_z(q)\bar{\Xi}_{nE}(q,d) + i\bar{\Xi}'_{nE}(q,d)]\Delta_{n\mathbf{k}_i} \} = 0. \tag{9}$$

Equations (8) and (9) must be solved to obtain the coefficients Θ_n and Δ_n for a given incident plane wave of $\mathbf{k}_i$. Note that Eq. (8) stands for the transmission of the incident plane wave into the subband states at the entrance of the constriction ($z=0$), and Eq. (9) stands for the reflection of the subband states at the exit of the constriction ($z=d$). This means that by solving the Schrödinger equation for the subband wave functions, the problem reduces to the calculation of the multiple reflections from the edges of the constriction.

Assuming that the current-transporting states are determined, we next deal with the total current. Since the current is conserved, the current calculated at $z=z_0$ is the same as the current passing through the system. One can specifically choose z_0 to lie in the constriction (i.e., $0 < z_0 < d$) so that the final current expression is given in terms of the coefficients of the subband wave functions. This way the quasi-1D nature of the system is incorporated into the current expression, and the current passing through the point contact can be related to the subband occupation, as the above simple explanation as well as the Landauer formulas both conjecture. The current energy density is obtained from the expectation value of the current density operator,

$$J(E) = 2e \int \frac{d\mathbf{k}_i}{(2\pi)^2} \langle \psi_{\mathbf{k}_i}(y,z) | j | \psi_{\mathbf{k}_i}(y,z) \rangle |_{z=z_0} \delta \left[\frac{\hbar^2 |\mathbf{k}_i|^2}{2m^*} - E \right] \Theta(k_0) \tag{10}$$

with the inclusion of spin degeneracy. The δ function selects the states on the Fermi circle and $\Theta(k_0)$ guarantees that the incident wave vector $\mathbf{k}$ is pointing towards the constriction. Using the wave function given in Eq. (3) one can express the current energy density as

$$J(E) = \frac{e}{\pi h} \int_{-k_E}^{k_E} \frac{d\kappa_0}{k_0} \text{Im} \sum_{nm} \left[\Theta_{nk_i}^* \left(\int dy \, \xi_{nE}^*(y,z_0) \frac{\partial}{\partial z} \xi_{mE}(y,z) \bigg|_{z=z_0} \right) \Theta_{mk_i} \right.$$

$$+ \Delta_{nk_i}^* \left(\int dy \, \bar{\xi}_{nE}^*(y,z_0) \frac{\partial}{\partial z} \bar{\xi}_{mE}(y,z) \bigg|_{z=z_0} \right) \Delta_{mk_i}$$

$$+ \Theta_{nk_i}^* \left(\int dy \, \xi_{nE}^*(y,z_0) \frac{\partial}{\partial z} \bar{\xi}_{mE}(y,z) \bigg|_{z=z_0} \right) \Delta_{mk_i}$$

$$+ \left. \Delta_{nk_i}^* \left(\int dy \, \bar{\xi}_{nE}^*(y,z_0) \frac{\partial}{\partial z} \xi_{mE}(y,z) \bigg|_{z=z_0} \right) \Theta_{mk_i} \right] . \tag{11}$$

The total current passing through the constriction is calculated by weighting the current energy density $J(E)$ with the number of electrons moving to the right in excess to those moving to the left in terms of the Fermi-Dirac distribution, and then by integrating over the entire energy range. This is $I = \int_0^\infty dE[f_{FD}(E,T) - f_{FD}(E+eV,T)]J(E)$. As mentioned above, the conductance of the QPC depends on how the voltage in the circuit is measured. We are adopting a two-terminal geometry for the measurements: two reservoirs at $z = \pm\infty$ are connected to the left and right 2D EG, and an infinitesimal difference $\Delta\mu = \mu_L - \mu_R$ is kept between the electrochemical potential of the left (μ_L) and right (μ_R) reservoirs. Beyond the screening length,[13] $\Delta\mu = eV$. The conductance is defined as the ratio of the current passing through the constriction to the difference of voltages measured deep in the reservoirs. At $T = 0$ K, only the states lying at the Fermi level contribute to the current. Hence, the conductance for the infinitesimal bias $\Delta\mu$ is given by $G = I/V - eJ(E_F)$. Since the conductance is measured in a two-terminal configuration in the experiments, the effects of the self-consistent field of the nonequilibrium electrons, and the resulting changes in $\Delta\mu$ between the two sides of the constriction, are not reflected by the results. The theoretical treatment given above is consistent with the experimental results as far as the relevant Landauer formula is concerned. Earlier, Landauer[13] conjectured that the dilution at the wide regions has the effect of a reservoir except for phase randomization. Numerical calculations by Yosefin and Kaveh[32] showed that this effect is present for constrictions with a smoothly varying cross section. The model used in this study, however, has an abrupt and infinite jump in cross section at the edges of the constriction. Therefore, the dilution effect is expected to be even stronger for the present case, and it is possible to take four-terminal measurements only by including the voltage probes in the constriction, i.e., by using a cross geometry,

In principle, the above formalism can be extended to yield self-consistent charge density. This requires the numerical solution of the Schrödinger equation with a general 2D (or 3D) potential, and the matching of solutions to the incoming and outgoing plane wave during each iteration. The wave function can also be expressed by linear combinations of appropriate basis sets. However, the self-consistent potential is strongly dependent on the actual geometry of the constriction, which is unfortunate-

ly not accessible. In this work we therefore use the simple and realistic confining potential yielding analytical transverse wave functions, and focus our efforts on analyzing several effects including the inhomogeneities of potentials.

The calculations by Kirczenow[23] and Szafer and Stone[26] used the same principles, i.e., that of matching the current-transporting wave functions, at the boundaries, except that in the latter[26] the 2D EG has finite dimensions and is thus relevant for a four-terminal configuration.[13] As pointed out above, the quasi-1D character with subband structure, etc., is explicit in the formalism. This makes the interpretation of experimental results more comprehensive. On the other hand, the subband structure is implicitly incorporated in the tight-binding model by Haanappel and van der Marel,[27] the scattering model by Garcia and Escapa,[28] and Anderson's model used by He and Das Sarma.[30] Moreover, the length of the constriction is not a restriction, as in the tight-binding methods. The changes in the potential profile can be treated by the use of transfer-matrix method. Since a mixed basis set that consists of the plane wave and constriction states is used, calculations do not require extensive computational effort; numerical results converge rapidly. The third dimension (i.e., the x axis) can also be implemented in the method to consider a tubelike constriction,[36] which does not change the essential components of the formalism. The versatility of this approach becomes important in studies[36] related to STM. The recent work by Pernas et al.[31] used the Keldysh-Green function within a tight-binding approach, and described the constriction by a chain of atoms[42] to determine the variation of $\mu(z)$ across the chain. After these general aspects of the approach we now treat some special cases.

III. UNIFORM CONSTRICTION

The existing studies showed that simulating the QPC by a uniform constriction connected to two 2D EG's with abrupt junctions bears little similarity to reality. Nevertheless, the uniform constriction is the easiest one to solve and it also has several features relevant to a real system. Here, the word uniform refers to the confining potential, which is the same throughout the constriction. That is, $V_c(y,z) = V_c(y)$ and $\phi_m(z) = 0$ for all $0 < z < d$. In this case, Eq. (2) is separable and its solutions are ex-

pressed by $\xi_{nE}(y,z) = e^{iy_n z}\varphi_n(y)$ where the form of the transverse wave function $\varphi_n(y)$ is obtained from the solution of

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dy^2} + V_c(y) \right]\varphi_n(y) = \varepsilon_n \varphi_n(y) \qquad (12)$$

with the subband energy ε_n. The propagation along the z direction is given by $\gamma_n = [2m^*(E - \varepsilon_n)/\hbar^2]^{1/2}$, where the root with the positive imaginary part is chosen. Equations (8) and (9) are first simplified by using the transverse Fourier transform of the subband wave function[43] and its derivative, $\Xi(q,z)$ and $\Xi'(q,z)$, respectively. Then, they are transformed into the following simple forms by multiplying them from the left by the transverse Fourier transform of the subband wave function $\Phi_n^*(q)$ and integrating over q:

$$(2\pi)^{1/2}2k_0\Phi_m^*(\kappa_0) = \sum_n [(K_{mn} + \delta_{mn}\gamma_n)\Theta_{nk_i}$$

$$+ (K_{mn} - \delta_{mn}\gamma_n)\Delta_{nk_i}], \qquad (13)$$

$$\sum_n [(K_{mn} - \delta_{mn}\gamma_n)e^{i\gamma_n d}\Theta_{nk_i}$$

$$+ (K_{mn} + \delta_{mn}v_n)e^{-i\gamma_n d}\Delta_{nk_i}] = 0, \qquad (14)$$

where we used the orthonormality property $\int dq\, \Phi_m^*(q)\Phi_n(q) = \delta_{mn}$ of the transverse wave functions and define $K_{mn} = \int dq\, \Phi_m^*(q)k_z(q)\Phi_n(q)$. These equations can be expressed in matrix form as

$$(2\pi)^{1/2}2k_0\tilde{\Phi}^\dagger(\kappa_0) = [(\tilde{K} + \tilde{\Gamma})\tilde{\Theta}_{k_i} + (\tilde{K} - \tilde{\Gamma})\tilde{\Delta}_{k_i}], \qquad (15)$$

$$[(\tilde{K} - \tilde{\Gamma})e^{i\tilde{\Gamma}d}\tilde{\Theta}_{k_i} + (\tilde{K} + \tilde{\Gamma})e^{-i\tilde{\Gamma}d}\tilde{\Delta}_{k_i}] = 0. \qquad (16)$$

Here, $\tilde{\Theta}$ and $\tilde{\Delta}$ are the column vectors for the coefficients of the subband wave functions with right-moving and left-moving probability currents, respectively, $\tilde{\Phi}$ is a row vector and $\tilde{K}$ and $\tilde{\Gamma}$ are square matrices with elements K_{mn} and $\delta_{mn}\gamma_n$, respectively. Before obtaining the solution for the coupled linear equations given in Eqs. (15) and (16), we wish to comment on their physical meaning. Let us consider only the junction at $z = 0$ (i.e., between the left-hand side 2D EG and the constriction) and assume that there are no left-moving states occupied in the constriction. In this case Eq. (15) becomes

$$\tilde{t}_{k_i} = (2\pi)^{1/2}2k_0(\tilde{K} + \tilde{\Gamma})^{-1}\tilde{\Phi}^\dagger(\kappa_0). \qquad (17)$$

Here we replaced $\tilde{\Theta}$ by $\tilde{t}$, since the problem under consideration corresponds to transmission into the constriction for an incident wave from the 2D EG, and the coefficients of the subband wave functions are just the corresponding transmission probability amplitudes. The vector $\tilde{t}$ is analogous to the transmission amplitude for a 1D infinite barrier $t = 2k/(k + k')$. The only difference is that the presence of more than one subband has to be taken into account, and all the relevant quantities in the *strictly* 1D problem have to be converted into matrices by using $\Phi_m(q)$ as a basis. Next, consider only the junction at $z = d$ and assume that the occupation of all the right-

moving states in the constriction is given by an arbitrary vector. Equation (16) then yields

$$\tilde{r} = (\tilde{K} + \tilde{\Gamma})^{-1}(\tilde{\Gamma} - \tilde{K}), \qquad (18)$$

where the square matrix $\tilde{r}$ contains the reflection probability amplitudes back into the constriction for the incident subband wave functions. Similar to the transmission amplitude, the reflection amplitude is analogous to the 1D equivalent $r = (k - k')/(k + k')$. Clearly this analogy to the 1D case holds for the entire subsequent formulation. For example, the wave function for a finite length constriction can be visualized by using the result for a finite rectangular barrier in 1D. The corresponding amplitude for the right-moving wave in the barrier is given by $\vartheta = t[1 - r^2\exp(2ik'd)]$ with the above t and r. Apparently, the analogy does not allow us to write the solution for all cases without making any calculations, due to the fact that matrix multiplication is not commutative. We found this 1D picture useful in understanding the underlying physics and to interpret the formalism as well as the data. Returning to our original problem, i.e., the conductance of a uniform, finite constriction, Eqs. (15) and (16) can be solved simultaneously to yield

$$\tilde{\Delta}_{k_i} = e^{i\tilde{\Gamma}d}\tilde{r}e^{i\tilde{\Gamma}d}\tilde{\Theta}_{k_i}, \quad \tilde{\Theta}_{k_i} = (\tilde{I} - \mathbf{r}e^{i\tilde{\Gamma}d})^{-1}\tilde{t}_{k_i}. \qquad (19)$$

Using the above wave function [the coefficients of which are given in Eq. (19)], we use Eq. (11) to calculate the corresponding conductance of a uniform constriction. Since $\{\varphi_n(y)\}$ form an orthonormal complete set, this equation is further simplified to obtain the following expression for conductance:

$$G = \frac{e^2}{\pi h}\int_{-k_F}^{k_F}\frac{d\kappa_0}{k_0}[\tilde{\Theta}_{k_i}^\dagger\tilde{\Gamma}_R\tilde{\Theta}_{k_i} - \tilde{\Delta}_{k_i}^\dagger\tilde{\Gamma}_R\tilde{\Delta}_{k_i}$$

$$+ 2\,\mathrm{Im}(\tilde{\Theta}_{k_i}^\dagger\tilde{\Gamma}_I\tilde{\Delta}_{k_i})], \qquad (20)$$

where $\tilde{\Gamma}_R = \tilde{\Gamma}_R + i\tilde{\Gamma}_I$. This expression is not reminiscent of the relevant Landauer formula,[10-12] $G \sim T$, since the cross section of the entire system under consideration changes discontinuously, and is infinite for $z < 0$ and $z > d$. In contrast to methods proposed in Refs. 24, 25, and 31, the contributions of various types of states are explicitly given in our formalism. In Eq. (20), the first and second terms in square brackets are related to the right-moving and left-moving states, respectively. The contribution of the evanescent (or tunneling) states is expressed by the third term. This is the feature that distinguishes quantum from classical transmission. While for each subband below E_F a channel of propagation is opened, the subbands above the Fermi level contribute to a smaller extent by tunneling. The contribution of tunneling becomes significant at small d, and the sharp rises in conductance due to the opening of a new channel are smoothed out by the evanescent states.

Using the above formalism, we consider a uniform constriction characterized by its length and the form of the confining potential $V_c(y)$, as illustrated in Figs. 1(b) and 1(c). Here we consider two types of confining potentials, for which analytical solutions are known. The first one is

an infinite-well potential expressed as $V_c(y)=0$ if $|y| \leq w/2$, otherwise it is infinity. This yields subband energies $\varepsilon_n = \hbar^2(2\pi/w)^2/2m^*$ and corresponding transverse wave functions $\varphi_n(y)$, which are known to be zero if $|y| > w/2$. The second type of potential is parabolic[44] of width $w=(\hbar/m^*\omega)^{1/2}$. The solutions are 1D harmonic-oscillator eigenstates with energies $\varepsilon_n = \hbar\omega(n+\frac{1}{2})$. Figures 2(a) and 2(b) illustrate the variation of the calculated conductance, G, with w for these potentials. The step structure is common for confining potentials, even though the spacing between steps may differ. It is also noted from both figures that the length is a crucial parameter of a QPC. After these general comments, we now examine the $G(w)$ curves more closely.

Earlier, using a semiclassical treatment for a very short constriction, Sharvin[21] showed that the conductance is independent of any material properties but is solely determined by the geometry (or area) of contact and electron density. The expression of contact conductance he obtained (which is referred to as Sharvin's conductance) is given by $G_s=(2e^2/h)2w/\lambda_F$. It varies linearly with w and goes to zero as $w \rightarrow 0$. Based on the full quantum treatment, we find that the behavior of conductance curves is similar for both confining potentials when $d=0$, except that these curves differ from Sharvin's linear conductance curve by the superimposed weak oscillations, due to the quantum interference effects. Also, the entire curve is slightly displaced from the origin. This can be explained by the Heisenberg uncertainty principle imposing the condition that $\Delta w \Delta p_y \sim h$. Since the transverse momentum Δp_y cannot exceed the Fermi momentum (i.e., $\Delta p_y \leq \hbar k_F$), the transmission is suppressed for very small Δw. To distinguish it from the semiclassical case, the conductance of a very short QPC in the quantum limit is named Sharvin's quantum conductance. As d increases, the contribution of evanescent waves decreases and quantum oscillations evolve into a steplike structure. The larger the value of d is, the sharper the steps are and

the closer their values are to the integer multiples of $2e^2/h$. For $d>5\lambda_F$, steps occur almost exactly at the integer multiples of quantum of conductance.

A. Resonance structure

The interference of waves reflected from the ends of the constriction yields the resonance structure superimposed on the flat plateaus of $G(w)$. That is, the conductance in the plateau oscillates between the quantized value (resonances) and minima (antiresonances). In Eq. (19), the matrix $e^{i\tilde{\Gamma}d}$ for the occupied subbands consists of phases that change with w. By neglecting the off-diagonal terms of $\tilde{K}$ and expressing $\tilde{r}$ and $\tilde{t}$ accordingly, we see that the resonances occur when $\gamma_n(w)d \approx \text{integer} \times \pi$. This is characteristic of the long, but finite, uniform constriction. The position of the mth resonance on the nth plateau in Fig. 2(a) is estimated to be $w_{mn} \simeq n\lambda_F[4-(m\lambda_F/d)^2]^{-1/2}$. Similarly, the number of resonances on the nth plateau can be estimated and is given by the simple expression $M_n \simeq (2d/\lambda_F)[(2n+1)/(n+1)^2]^{1/2}$. From these approximate expressions we deduce the results that the number of resonances on a plateau increases with increasing d but decreases with increasing w (or increasing subband index n). Within the same approximation, we also find that δG_{mn} (i.e., the difference in conductance between the mth resonance and the subsequent antiresonance on the nth plateau) decreases as either m or n increases. As for fixed m and n, the larger the value of d, the greater is that of δG_{mn}. The analysis based on $G(w)$ curves, which are produced with a relatively finer mesh, showed that the envelope of antiresonances is approximately independent of the length of the constriction d.

The experimental data lack the resonance structure. Moreover, the sharp corners of the step structure in the
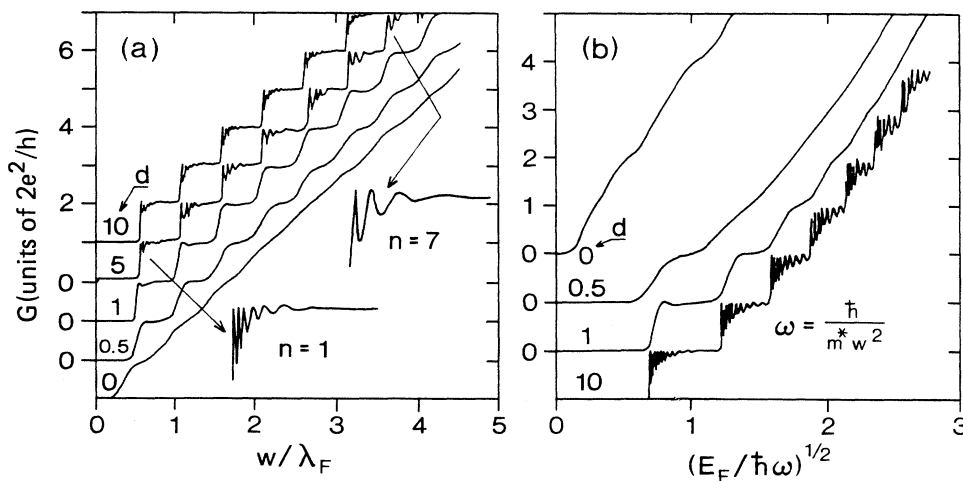

FIG. 2. Conductance G of a quantum point contact (QPC) with uniform constriction between two 2D EG's calculated at $T=0$ K for various lengths d in units of λ_F. (a) Infinite-well (hard-wall) and (b) parabolic confining potentials. The resonance structures of the first and seventh channels obtained for the infinite-well potential in a uniform constriction are magnified and shown in the inset.

experiment are rounded to display a smooth curve, due to both lithography and fringing fields. Evidently, the predictions of the uniform constriction and the experimental data are at variance. Assuming that $l_e > d$ and, thus, inelastic scattering by phonons is negligible, there are a number of effects (such as longitudinal variation of potential in the constriction and the saddle-point effect, nonuniform width, scattering from defects, finite temperature and bias, etc.) that can account for the absence of resonances in the experiment. Some of these effects can be easily controlled and their contribution causing G to deviate from the exact quantum values is minimized. In the rest of this section we will examine two effects that can destroy the simple phase relation, in spite of the uniform and perfect constriction.

B. Effects of finite temperature and bias voltage

Finite temperature increases the probability of inelastic scattering by phonons. This causes the mobility and, hence, l_e to decrease. In addition, the states are averaged in the energy range of $\sim 4k_B T$ around the Fermi circle. Here, we omit the former effect, assuming that $l_e > d$ still, and consider the latter effect (i.e., smearing out of the sharp Fermi level). Specifically, we examine the temperature range in which the resonance structure is destroyed. The conductance at finite temperature is calculated from the integral

$$G(w;T{\neq}0) = \int dE\, G(w;T=0)[-\partial f_{FD}(E,T)/\partial E] ,$$

where the term in the square brackets becomes a δ function at $T=0$ K. Since the difference between consecutive subband energies is approximately proportional to $\sim E_F/n$, the effect of temperature on quantization is independent of d, but increases with increasing subband index n. However, this is not true for resonances, since energy spacing between two adjacent resonances is proportional to d^{-2}. In Fig. 3(a) we illustrate the behavior of

$G(w)$ calculated for different d and T values. Even at temperatures as low as $T=5$ K, the resonance structure completely disappears, and higher-lying steps are smeared out. In compliance with the above discussion, the effect of finite temperature on resonances depends on the length of constriction. The resonance peaks are widely spaced for small d, and thus they persist, in spite of the energy spreading due to the finite temperature. In contrast to this, the resonance structure of the long constriction $(d > 10\lambda_F)$ is closely spaced, and thus they can easily be eliminated even at $T=0.6$ K. On the other hand, it is known that the quantization begins to disappear[45] owing to voltage fluctuations[46] even before for $d \geq 10\lambda_F$. In the temperature range within which the experiments are performed $(T \leq 0.1$ K), the resonance structure was maintained if the length of constriction was small (i.e., $d < 2.5\lambda_F$) and, of course, if the geometry of constriction was suitable to maintain the simple phase relation. The experiments[47] carried out below 100 mK displayed some irregular features reminiscent of resonance structure. Now the consensus is that the temperature effects alone are not sufficient to explain the experimental results.

The conductance of a uniform constriction under a finite bias voltage is calculated by the following expression:

$$G(w;V{\neq}0) = \frac{1}{eV}\int_{E_F-eV/2}^{E_F+eV/2} dE\, G(w(E/E_F);V=0) ,$$

(21)

which is also present in Fig. 3(b). Similar to the effect of the finite temperature, the finite bias voltage also affects the resonance structure. For example, a basis of $eV=0.05E_F$ is sufficient to destroy the resonance structure. For a bias of $eV=0.5E_F$, the quantized steps begin to disappear for high-index subbands (or channels). Note that in the experiment, the temperature effects are more important than those of the finite bias, since the applied bias was kept low $(eV \ll k_B T)$ to prevent heating due to hot electrons. However, experiments[47] performed to ex-
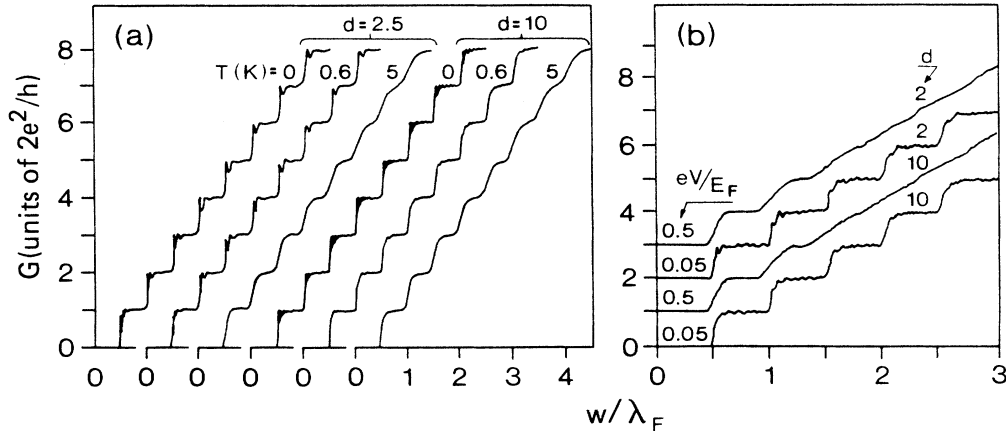

FIG. 3. Conductance $G(w)$ of the quantum point contact with a uniform constriction and infinite-well confining potential between two 2D EG's, calculated for (a) finite temperature and (b) finite bias V. The length d is given in units of λ_F.

plore the nonlinear conductance of the QPC showed that the constant conductance on the plateaus is no longer valid above a critical bias. The nonlinear conductance due to the finite bias was also treated theoretically.[48]

IV. NONUNIFORM CONSTRICTION

The openings to 2D EG reservoirs are expected to be either flared or smooth, owing to the limitations in the device fabrications. In addition, the electrostatic depletion region created by the split gate cannot be as sharp as the lithographic geometry. Here, we deal with the conductance of a wedgelike, tapered constriction, and with random nonuniformities of the width and surface roughness, as a first step towards a realistic QPC. In all these cases, the width of the constriction depends on the longitudinal coordinate z; namely, it is given by $w(z)$. As a result, the subband wave functions and their energies will also depend on the longitudinal coordinate. We adopt the transfer-matrix method[29] to calculate the current by multiple boundary matching, without invoking the numerical integration of the Schrödinger equation in these nonuniform constrictions.

A. Transfer-matrix method

While a rectangular barrier is the 1D analog of the uniform constriction, a nonuniform constriction can be identified with the general barrier potential of the 1D case, which is treated using the transfer-matrix method. To this end, one divides the space into a number of segments and also assumes that the potential is constant in each of these segments. The approximate solution is then obtained by the usual boundary matching of wave functions at each interface between the adjacent segments. The number of segments used in the calculations is determined according to the longitudinal variation of $V(\mathbf{r})$. However, this number cannot be arbitrarily increased since numerical accuracy diminishes with multiple matrix multiplication. In line with the above discussion, the nonuniform constriction is also divided into a number of segments. In each segment, $\phi_m(z)$ and $V_c(y,z)$ are assumed to be constant. Thus, the solutions for the subband wave functions in the ith segment are the same as that for a uniform constriction with confining potential $V_c(y,z_i)$, where the energy zero is shifted by $\phi_m(z_i)$. Then the current-carrying states at z_i read as

$$\psi_{\mathbf{k}_i}(y,z) = \sum_n [e^{i\gamma_n(z_i)z}\Theta_{n\mathbf{k}_i}(z_i) + e^{-i\gamma_n(z_i)z}\Delta_{n\mathbf{k}_i}(z_i)]\varphi_n(y,z_i), \quad z_{i-1} < z < z_i , \tag{22}$$

where the propagation constant $\gamma_n(z_i) = \{2m^*[E - \varepsilon_n(z_i)]\}^{1/2}$ is expressed in terms of z-dependent subband energy. For the nonuniform constriction, the boundary conditions at $\{z_i\}$ have to be taken into account, together with those at $z=0$ and $z=d$. The transfer matrix for the ith interface can be written as

$$\widetilde{T}_{i,i+1} = \frac{1}{2}\begin{bmatrix} e^{-i\widetilde{\Gamma}(z_i)z_i}[\widetilde{S}_{i,i+1} + \widetilde{\Gamma}(z_i)^{-1}\widetilde{S}_{i,i+1}\widetilde{\Gamma}(z_{i+1})]e^{i\widetilde{\Gamma}(z_{i+1})z_i} & e^{-i\widetilde{\Gamma}(z_i)z_i}[\widetilde{S}_{i,i+1} - \widetilde{\Gamma}(z_i)^{-1}\widetilde{S}_{i,i+1}\widetilde{\Gamma}(z_{i+1})]e^{-i\widetilde{\Gamma}(z_{i+1})z_i} \\ e^{i\widetilde{\Gamma}(z_i)z_i}[\widetilde{S}_{i,i+1} - \widetilde{\Gamma}(z_i)^{-1}\widetilde{S}_{i,i+1}\widetilde{\Gamma}(z_{i+1})]e^{i\widetilde{\Gamma}(z_{i+1})z_i} & e^{i\widetilde{\Gamma}(z_i)z_i}[\widetilde{S}_{i,i+1} + \widetilde{\Gamma}(z_i)^{-1}\widetilde{S}_{i,i+1}\widetilde{\Gamma}(z_{i+1})]e^{-i\widetilde{\Gamma}(z_{i+1})z_i} \end{bmatrix}, \tag{23}$$

in terms of the overlap matrix $\widetilde{S}$, which has elements $(S_{i,i+1})_{mn} = \int dy\, \varphi_m^*(y,z_i)\varphi_n(y,z_{i+1})$ between the ith and $(i+1)$th intervals. This matrix continues the solution from the ith segments to the $(i+1)$th segments. The solutions in the first and the last segments are, in turn, connected to each other by

$$\begin{bmatrix} \widetilde{\Theta}_{\mathbf{k}_i}(z_1) \\ \widetilde{\Delta}_{\mathbf{k}_i}(z_1) \end{bmatrix} = \widetilde{T}_{1,N}\begin{bmatrix} \widetilde{\Theta}_{\mathbf{k}_i}(z_N) \\ \widetilde{\Delta}_{\mathbf{k}_i}(z_N) \end{bmatrix}, \tag{24}$$

where $\widetilde{T}_{1,N} = \prod_i^{N-1}\widetilde{T}_{i,i+1}$ is the product of all transfer matrices along the constriction. The continuity equations at $z=0$ and $z=d$ together with Eq. (24) can be solved simultaneously to obtain the wave function in the Nth segment. The wave function in any other segment can be calculated thereof by using the transfer matrix. The conductance of the constriction can be calculated by using Eq. (10) and the wave function given in Eq. (22). The result is exactly in the same form as Eq. (20). The

segment in which $\widetilde{\Gamma}$, $\widetilde{\Theta}_{\mathbf{k}_i}$, and $\widetilde{\Delta}_{\mathbf{k}_i}$ are calculated does not matter since the current along the constriction is conserved.

B. Wedgelike entrance and tapered constriction

The variation of the conductance with the narrowest width of a wedgelike entrance is calculated by using the transfer-matrix method. The geometry and parameters relevant to this type of constriction are described as an inset in Fig. 4. Our results are based on calculations for the wedge angle α ranging from 0° (Sharvin case) to 90° (uniform constriction). Until the wedge angle reaches a certain value ($\alpha \simeq 50°$), the conductance curves do not deviate significantly from that of Sharvin's quantum conductance, corresponding to $\alpha=0°$. For wedge angles exceeding $\sim 50°$, the quantum steps start to develop but only become apparent for $\alpha > 75°$. At a particular wedge angle ($\alpha \simeq 60°$), the quantum effects are emphasized as d
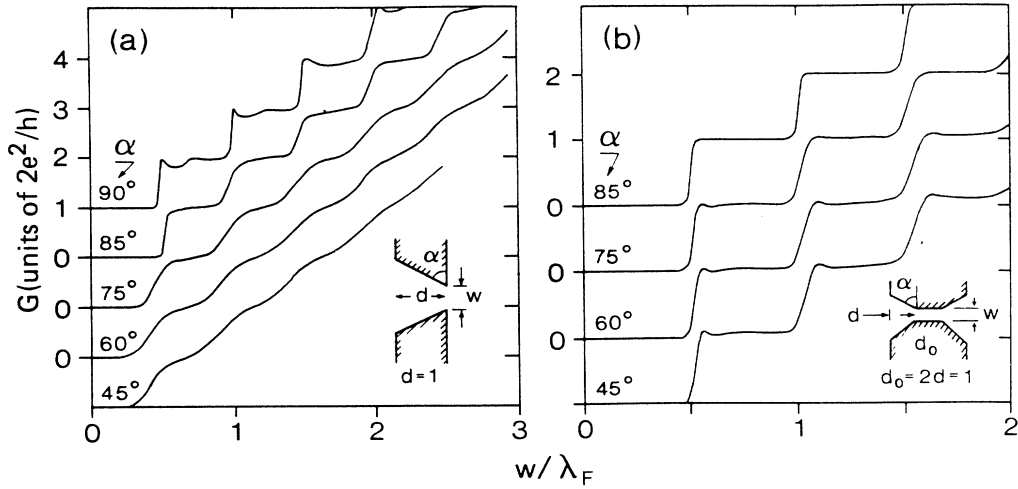
FIG. 4. Conductance $G(w)$ of a quantum point contact for (a) a wedgelike entrance and (b) a flared constriction as described by the insets. The confinement in the transverse direction is the infinite-well potential. The length d is given in units of λ_F.

increases. We also note that, in spite of the apparent step structure at large α, the resonance structure does not occur, owing to the phase mixing caused by the interaction between different subbands in the aperture.

A tapered constriction with a finite uniform part at the center (d_0) and flared openings to 2D EG reservoirs (characterized by α and d) is described as an inset in Fig. 4(b). If $d_0 = 0$, the conductance G is similar to Sharvin's quantum conductance even for $\alpha \sim 45°$. Weak oscillations change gradually to the step structure as α increases. The step structure with flat plateaus and quantized conductance however, even appears for $\alpha \sim 45°$ if a uniform part of length $d_0 \simeq \lambda_F$ is put between two taperings.

C. Adiabatic evolution of states in the QPC

The tapered constriction discussed above may be reminiscent of the special case considered by Glazman and his co-workers.[24] They treated a constriction between two large circles, the width of which varies very slowly, and obtained quantized conductance without resonance structure. They explained such a behavior by the adiabatic evolution of current-carrying states without reflections. In fact, as illustrated in Fig. 4(b), the steps become sharper and, concomitantly, the resonance structure in the conductance curve calculated for the tapered structure with $d_0 = \lambda_F$ and $85° \lesssim \alpha < 90°$ disappears. By contrast, one would expect the resonance structure to become pronounced and sharper, since the tapering changes into the uniform constriction as $\alpha \rightarrow 90°$, and its length increases from d_0 to $d_0 + 2d$. This unexpected behavior of G has close similarities with the model of Glazman et al.,[24] and is explained by the adiabatic evolution of the current-carrying states. Since $|\partial w/\partial z|$ is small for $\alpha \gtrsim 85°$, a state entering the tapered entrance evolves without changing the quantum number n associated with the transverse wave function, but the eigenstate $\varepsilon_n(z)$

slowly varies with z. In this case, the motion along the z direction can be considered by a 1D Schrödinger equation with the potential $\phi_m(z) + V_{n,\text{eff}}(z)$. The effective potential $V_{n,\text{eff}}(z)$ is essentially the slowly varying, z-dependent subband energy $\varepsilon_n(z)$, with a small correction term. As a result, $\xi_{n,E}(y,z) = \xi_{n,E}(z)\varphi_n(y,z)$ and the subband wave functions $\varphi_n(y,z)$ belonging to different subbands are decoupled, each satisfying its own "effective" Schrödinger equation. In general, this can be achieved by slow variation of w or $V_c(y,z)$ in the length scale of λ_F. The quantization of the conductance originates from the transmission at the narrowest portion (or at the neck). The resonances are lacking because intersubband scattering and intraband reflections are suppressed, due to smooth variation of $w(z)$ or $V_c(y,z)$. For the transmission from wide to narrow w, $\varepsilon_n(z)$ is lowered and the momentum k_y is increased, while the momentum k_z in the direction of propagation is decreased. A reverse situation occurs at the exit to the 2D EG, if the state continues to evolve adiabatically. It appears that the condition for the adiabatic evolution of a state in a special QPC is satisfied for the geometrical parameters corresponding to $d_0 = \lambda_F$ and $85° \lesssim \alpha < 90°$. Note that, owing to the adiabatic change, the quantization of conductance is not affected in any essential manner but that, owing to the suppression of reflections at the ends constriction, the resonance structure disappears. Other types of geometry, for example, constrictions obtained by two sine or cosine modulations, which also provide adiabatic change, were investigated earlier.[34] The adiabatic approximation and its limits are thoroughly investigated by Yacoby and Imry.[48] They showed that the adiabaticity effects are important even for finite constrictions, which have abrupt junctions to the 2D EG. The corrections due to the abruptness of connections are exponentially small. This is exactly what was found earlier, using the transfer-matrix method.[35] In Sec. VIII we will return to the problem of adiabaticity in STM and related fields.

D. Quasibound states in a QPC

While the geometry described above has hornlike openings towards the 2D EG reservoirs, a finite constriction, which is relatively narrower at both ends, gives rise to spatially varying subband energies $\varepsilon_n(z)$, which are lowered towards the center. These subbands can be viewed as potential wells, in which waves are confined and form quasi-0D states. A similar confinement leading to bound states can also occur even if the local widening of w is abrupt, or the potential is lowered locally as with the attractive impurity potential inside the constriction. The latter situation will be the subject of the following section. Here we can present a simple physical picture for the (quasi-) bound states in the constriction: the solution of the Schrödinger equation in the region where either the size (width) or potential differ significantly from those in adjacent parts of the constriction may yield states with relatively lower energies. If these states cannot find matching partners they decay into the adjacent regions of the constriction, and their charge density increases in the region of localization. Depending upon the extent of the adjacent regions (i.e., the distance between the 2D EG and the center of localization), these states are either totally confined or can match in the continuum of states in the 2D EG to form resonances. The Coulomb blockade can be important for the transport through strongly confined 0D states. For states occurring above the threshold of the first channel, strong confinement is not expected due to mixing with subband states. The occurrence of quasibound states in local widening in a constriction was predicted in Ref. 35. If the local widening
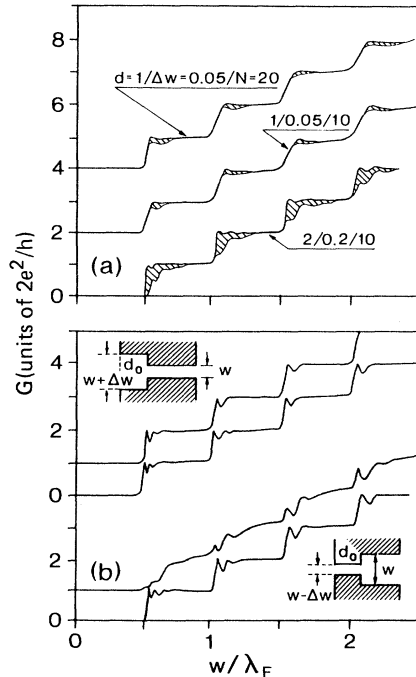

FIG. 5. Conductance $G(w)$ of a quantum point contact with the infinite-well potential confinement. (a) Rough surface. (b) Obstacles at the entrance. See Sec. IV for more details.

or lowering of the potential is repeated in the constriction and, at the same time, adjacent constrictions are allowed to couple, the states of the individual wells combine to give either a bonding and antibonding combination, or a miniband structure. The latter occurs if the wells are periodically repeated. The above picture is identical with that developed for single and multiple-quantum-well structures (or the Krönig-Penney model within the effective-mass approximation) in semiconductor heterostructures or superlattices.[1,49] Recently, a miniband structure was observed[50] in an artificial, finite 1D crystal produced by a sequence of quantum dots in a similar setup to the one described above, leading to the quantization of conductance. Castano et al.[5] presented a theoretical study of the periodic modulation of the potential inside the ballistic constriction. They considered periodically repeating zero-potential (potential-well) and adjacent finite-potential (barrier) regions. They showed that, under the applied voltage and in the high field regime, transmission and reflection from the miniband structure gives rise to nonlinear transport characteristics and negative differential conductance.[52] This is similar to the resonant tunneling behavior observed in semiconductor superlattices with thin barriers.[53] An interesting effect brought about by the quasibound states, namely resonant tunneling in a QPC structure, will be studied extensively in Sec. VI.

E. Surface roughness

The roughness originating from the quality of the split gate gives rise to irregular variations in the potential of a QPC. We simulate the effect of surface roughness by irregular changes of the width of the infinite-well confining potential. The effect of the potential fluctuations inside the constriction can also be revealed from this simulation. The random modulation of w is characterized by two parameters. These are the length of roughness, i.e., $\delta d = d/N$, and its amplitude Δw. At each step i ($i = 1, \ldots, N$) along the constriction, w is varied by $\chi \Delta w$, where the value of χ ($0 \leq \chi \leq 1$) is taken at random. Thus, a histogram profile $\delta w(z)$ is superimposed over the uniform width w. Finally, the conductance as a function of width is calculated for various $\delta w(z)$ profiles, and is traced on the same plot with respect to the average width $\bar{w} = w + d^{-1} \int_0^d \delta w(z) dz$. These plots are presented in Fig. 5(a), where $G(\bar{w})$ lies in the shaded region for a given profile characterized by δd and Δw. This simply indicates that $G(\bar{w})$ fluctuates in the shaded area when the surface of the constriction varies within the limits set by δd and Δw. The important conclusion is that the extent of the lateral variations in the width of the constriction, that is Δw, is the crucial parameter. As Δw increases, the deviation from quantized values becomes more significant, and the interference resonances become less visible. In the figure, a weak resonance structure is still seen, since the Δw value used is not large enough to destroy all phase coherence. The longitudinal variation δd affects the quantization to a lesser extent.

In this context, we will touch upon two other forms of roughness. These are regarded as obstacles, and can be

implemented in our model by changing w by $\pm \Delta w$ for a length of d_0 at the entrance to a uniform constriction of length d [see Fig. 5(b)]. The obstacle, which is wider than the uniform constriction, provides a sufficient number of states (or enough freedom) to link the 2D EG states to those of the main constriction. As a result, the conductance is not affected significantly. In contrast, the "real" obstacle (or narrower entrance) lacks the appropriate states to link the 2D EG to the rest of the constriction. In this case, the openings of channels are shifted by Δw (in units of λ_F) and, hence, the sharp step structure is disturbed and flat plateaus disappear with increasing Δw. We find that even a small reduction in the longitudinal width at the entrance (for example, $\Delta w = 0.05 \lambda_F$) results in important effects and drastic deviations from the ideal $G(w)$ curve. This effect would not be as dramatic if the same obstacle were located near or at the center of the constriction. These results imply that the openings of the channels would be delayed and, consequently, the quantization would be disturbed by any shrinkage of the potential due to the accumulation of charge at the openings of the constrictions. Therefore, a hornlike, smooth entrance is expected to be necessary in obtaining sharp step structures.

V. ELASTIC SCATTERING FROM IMPURITIES

The elastic scattering by impurities in a QPC has been of interest recently.[27,54–58] The prime motivation is the influence of the impurities on the quantization of conductance. Earlier, we studied the elastic scattering caused by impurities and its effect on the conductance for different constriction geometries.[58] The form of the model potential was more realistic compared to previous models,[27,54,55] and enabled us to obtain exact solutions of scattering events. Thereafter, it was appropriate to carry out a systematic analysis to reveal the effect of the location, lateral extent, and type (attractive versus repulsive) of the impurity potential. In this paper, we outline our previous results and concentrate on the states bound in the attractive impurity potential. The enhanced transmission over these states and the physical picture developed are used in the following section in connection with resonant tunneling effects in the ballistic constriction.

To develop a formalism, we first consider an infinite constriction: the potential of the impurity located at (y_I, z_I) in this ballistic constriction is represented by

$$V_I(y,z) = (\hbar^2 \beta / m^*) \exp(-q|y - y_I|) \delta(z - z_I) , \tag{25}$$

which is the δ Dirac function in the z direction, but which decays exponentially along the y direction with a decay length of q^{-1}. The magnitude of β sets the magnitude of the impurity potential, which may be attractive $(\beta < 0)$ or repulsive $(\beta > 0)$. Then, the eigenstates of the uniform constriction in the presence of the above scattering potential are expressed by

$$\psi_j(y,z) = e^{i\gamma_j^z} \varphi_j(y)$$
$$+ \int dy \, dz \, g(z - z'; y, y') V_I(y, z') \psi_j(y, z') . \tag{26}$$

The first term on the right-hand side represents the incident wave, which is the unperturbed solution for the jth subband with the wave function $\varphi_j(y)$, the eigenenergy ε_j and the corresponding wave vector γ_j along the z direction described in the preceding sections. The above expression is the well-known Lippman-Schwinger equation adapted to quasi-1D systems with the retarded Green's function g. For the impurity potential given in Eq. (25), the equation for the scattered wave function is exactly solvable and the $\tilde{T}$ matrix is given by

$$\tilde{T}_{mj}(k) = (2\pi)^{-1/2} \tilde{\Omega}_{mj} \exp[-i(k - \gamma_j)z_I] , \tag{27}$$

where $\tilde{\Omega} = \tilde{u}(\tilde{\Gamma} + i\tilde{u})^{-1}\tilde{\Gamma}$, with $\Gamma_{ij} = \delta_{ij}\gamma_j$ and $\tilde{u}_{ij} = \beta \int dy \, \varphi_i(y)\varphi_j(y) \exp(-q|y - y_I|)$. In line with previous discussions, the expression of conductance for the infinite, uniform constriction with an impurity is given in terms of the above matrices,

$$G = 2\frac{e^2}{h} \sum_j^{\varepsilon_j < E_F} \left[1 + 2\frac{\mathrm{Im}(\tilde{\Omega})_{jj}}{\gamma_j} + \frac{\mathrm{Re}(\tilde{\Omega}^\dagger \tilde{\Gamma}^{-1} \tilde{\Omega})_{jj}}{\gamma_j} \right] . \tag{28}$$

Important results obtained from the calculations of impurity scattering in an infinite constriction are outlined: if the potential of the impurity is weak (for example, $|\beta| < 0.5k_F$ and $q = 10\lambda_F^{-1}$), $G(w)$ will still reflect a staircase structure with rounded steps and plateau values close to the quantized values, $2e^2 N_c/h$ (where N_c represents the number of conduction channels). Also, in compliance with the first-order Born approximation, the sign of the scattering potential does not show pronounced differences if the scatter is weak.

A. Conductance dips in quantized channels

Ignoring the intersubband interaction, which is included via the nondiagonal elements of $\tilde{u}$, one is left with the strictly 1D problem for each subband. Clearly, an attractive δ-function potential has a bound state in one dimension. Since the bound states do not contribute to the transport in an infinitely long (or even finite but long) system, the effects of these bound states would not appear in the conductance curve. This is the situation illustrated by curve a in Fig. 6, where the intersubband interaction is negligible due to long decay length q^{-1}. For a relatively stronger impurity potential $(0.5k_F \lesssim |\beta| \lesssim k_F)$, not only are steps smoothed, but also the values of the plateaus exhibit significant deviations from the perfect quantized values. In this range of β the difference between the attractive and repulsive impurity becomes pronounced. In particular, for strongly attractive and laterally confined (i.e., large q) impurity potentials, dips in the $G(w)$ curve appear below certain steps. The values of G at the minima are found to be $\sim 2e^2(N_c - 1)/h$. The occurrence of these dips is closely related to the position of the impurity
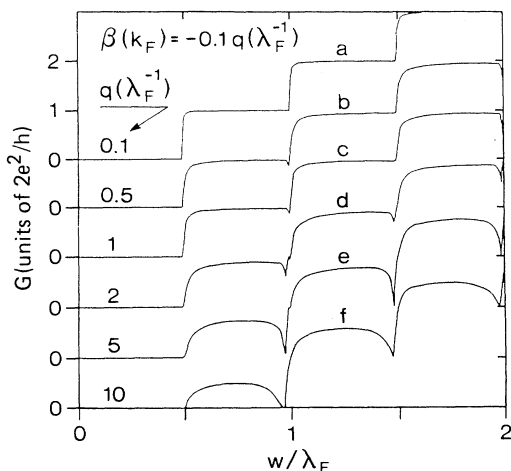
FIG. 6. Conductance $G(w)$ of the infinite, uniform constriction with an attractive impurity at $y_I = 0.13\lambda_F$. The impurity potential is described in Eq. (25) and $\beta/q = -0.1$. The confinement in the transverse direction is the infinite-well potential.

and hence, to the symmetry of the interaction. These dips originate from the enhanced backscattering due to the bound states of an attractive impurity potential. As seen in Fig. 6, curves b and c, some dips begin to emerge just below the propagation threshold of the subband $n \geq 1$. For intermediate values of q (Fig. 6, curves d and e) there are both dips and peaks, and for large q values (Fig. 6, curve f) the peaks merge into the propagation threshold, leaving only dips. These dips are the sign of enhanced backscattering resulting from the subband interaction. When the coupling between the subbands is not negligible, one has to deal with the full quasi-1D Schrödinger equation. The 1D bound states described above are degenerate with the continuum of propagating states from lower-lying subbands and they have to be mixed. Such a mixing give rise to two types of states. The first type are the quasibound states with no net current, reminiscent of the bound states. These do not contribute to the current for infinite constriction. The second type are complete backscattering states, which have a unity reflection probability. The wave function for these states has a standing wave form in the lower-lying subband and a quasibound-state contribution from the next subband. For very large values of q (Fig. 6, curve f), the first type of states disappears since the 2D δ function has no bound states. Recently, Faist, Guéret, and Rothuizen[59] observed the dips and absence of plateaus in conductance curves of highly disordered samples. Although the arguments given here are not directly applicable to the experiment, the underlying physics is simply the enhancement in backscattering due to presence of localized states.

B. Impurity scattering in a finite constriction

The above formalism can be extended to calculate the conductance for a finite constriction with an impurity po-

tential described in Eq. (25). To this end we assumed that $v_I(y,z)$ is zero outside the QPC region, $0 \leq z \leq d$. This assumption simplifies the solution, since elastic scattering takes place only in the constriction. An expression of $G(w)$, which is similar to Eq. (20) in Sec. III, is obtained in terms of the vectors of coefficients $\bar{\Theta}$ and $\bar{\Delta}$. Note, however, that in the present case the coefficients $\bar{\Theta}$ and $\bar{\Delta}$ depend on the parameters of the impurity potential; namely, $\bar{\Omega}$ and z_I, as well as the parameters of the constriction.

As mentioned earlier, the sharp changes in the conductance of the infinite constriction are smoothed out in the impurity-free, finite constriction, due to the contribution of the evanescent waves, especially for $d \leq \lambda_F$. For relatively longer constrictions this effect decreases, but the resonance structure due to the interference of the left- and right-moving waves becomes pronounced. In the presence of elastic scattering by a single impurity, neither a phase breaking due to the inelastic scattering nor a phase averaging due to a large number of scatterers can take place. As a result, the resonance structure will be present in any finite constriction having an elastic-scattering center. The dramatic effect of the impurity, however, is seen in $G(w)$, which can exceed the conductance for the infinite constriction with a single impurity, as well as that of the impurity-free, finite constriction (excluding the contribution due to the tunneling). This arises due to constructive combination of the scatterings from the impurity, and from the boundaries at $z = 0$ and $z = d$. Here, the point to emphasize is that the heights and positions of the resonances and antiresonances in $G(w)$ are strongly dependent on the position of the impurity along the z direction. This implies an important feature of the mesoscopic transport; the conductance will oscillate by moving the impurity along the channel. The magnitude and period of these oscillations are related to the length and width of the constriction, in addition to the properties of impurity. In fact, such an effect was observed by Ralls and Buhrman.[60] A summary of our results for the finite constriction having a single elastic impurity is illustrated in Fig. 7.

In the final part of this section we address two questions, which are actually interrelated. (i) Is the 1D analogy still valid for impurity scattering? (ii) Are the expressions of conductance related to the relevant Landauer formula? The conductance in the presence of elastic scattering from the impurity can be obtained by exploiting the analogy between the strictly 1D and quasi-1D systems, discussed in Sec. III. In the strictly 1D case, the solution is known to be obtainable from the characteristic discontinuity of the derivative of the wave function due to the δ function. We can use the same procedure within the formalism for a uniform constriction developed in Sec. III. To this end, the subband wave function can be written by using the transmission matrix element between the nth and mth subband states: $\xi_{nE}(y,z) = e^{i\gamma_n z}\varphi_n(y)$ $+ \sum T_{mn}e^{-i\gamma_m z}\varphi_m(y)$. Using this wave function in the Schrödinger equation, which includes the impurity potential, one observes the continuity of the wave function and the discontinuity in the first derivative. From the continuity of the wave function at z_I, and also from the in-
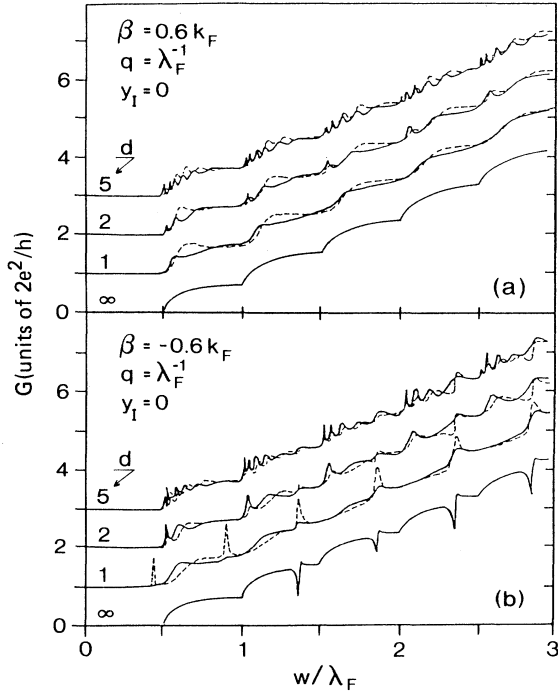
FIG. 7. Conductance $G(w)$ of a finite, uniform constriction having an impurity potential with $q = \lambda_F^{-1}, y_I = 0$ and for (a) $\beta = 0.6 k_F$ and (b) $\beta = -0.6 k_F$. Solid (dashed) curves denote $z_I = 0.2\lambda_F$ ($z_i = 0.5\lambda_F$) and are offset vertically by $(2e^2/h)$ for clarity. The length d is given in units of λ_F.

tegration of the Schrödinger equation between z_I^- and z_I^+, one obtains

$$\tilde{T} = e^{-i\tilde{\Gamma}z_i}(\tilde{\Gamma} + i\tilde{u})^{-1}\tilde{\Gamma}e^{-i\tilde{\Gamma}z_i}$$

and (29)

$$\tilde{R} = -ie^{-i\tilde{\Gamma}z_i}(\tilde{\Gamma} + i\tilde{u})^{-1}\tilde{u}e^{-i\tilde{\Gamma}z_i} \ .$$

These are analogous to the transmission, $t = k/(k + i\beta)$, and reflection, $r = -i\beta/(k + i\beta)$, probability amplitudes for a δ-function scatterer in a strictly 1D system. Using the transmission matrix $\tilde{T}_t$, which includes only the occupied subbands in the Landauer formula $G = (2e^2/h)\text{Tr}(\tilde{T}_t^+\tilde{T}_t)$, one obtains the conductance for an infinite constriction. It can be shown that this form is exactly the same as the expression in Eq. (28), obtained by the Green's-function formalism. A similar procedure can be applied for a finite constriction, but it uses more complicated expressions.

VI. RESONANT TUNNELING IN A QPC

In the adiabatic approximation, $\varepsilon_n(z)$ is viewed as an effective potential for the motion along the z direction. If the constriction becomes locally wider, or the potential $V(\mathbf{r})$ is locally lowered, the effective potential $\varepsilon_n(z)$ acts as a quantum well between two potential barriers. As explained in Sec. IV, under certain circumstances, states

can be confined to this quantum well. Earlier, we showed[35,58] that the confined (or quasibound) states can give rise to resonant tunneling (RT) effects, similar to that found in the quantum well between two double barriers produced by semiconductor heterostructures.[2,52,53] Owing to this effect, a RT peak is formed in the $G(w)$ curve. In this section we present an analysis of RT in a QPC with the assumption that the effect of the Coulomb blockade is negligible. The characteristics of RT transmission are illustrated by curves a–d in Fig. 8. Below the steps, prior to the opening of a new channel of transmission, i.e., $w < n\lambda_F/2$, there are sharp peaks with amplitudes approaching $2e^2/h$. Since the nth channel is not propagating, transport occurs via tunneling for this particular channel. However, this cannot be an ordinary tunneling process because the conductance does not increase uniformly. In addition to normal tunneling through the effective barrier, a RT effect begins, as depicted by the peaks in the $G(w)$ curve. For certain values of w, the resonance condition is satisfied, yielding a unity transmission. By further increasing w, the resonance condition is destroyed and G rapidly drops to its ordinary tunneling value.

An attractive impurity can also lead to a RT effect: the $\tilde{T}$ matrix defined in Eq. (27) becomes singular whenever the determinant of $(\Gamma + i\tilde{u})$ is zero. Clearly, $\tilde{u}$ is real and $\tilde{\Gamma}$ is purely imaginary when all subbands are unoccupied, i.e., $w < \lambda_F/2$. Despite this, for a given negative β, this determinant vanishes when $w = w_b(\beta, q, y_I)$. This corresponds to a bound (0D) state, for which the wave function decays exponentially, as $\exp(-|\gamma_b||z - z_I|)$, away from the impurity in an infinite constriction of $w = w_b$. Clearly, this bound state does not give rise to a RT effect for the infinite constriction. However, it can interact with the continuum of states in the 2D EG when d is
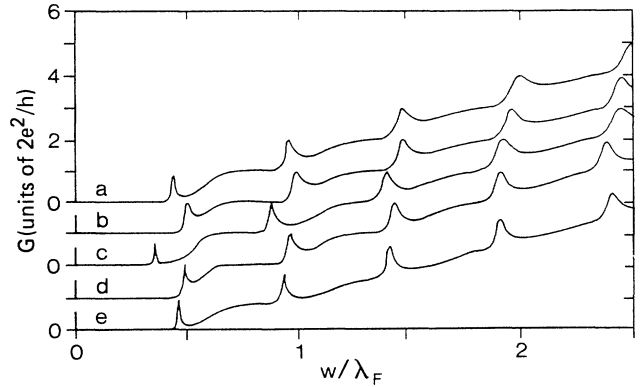

FIG. 8. Resonant tunneling effects on the conductance of a QPC of length $d = \lambda_F$ due to the widening of w, the potential well or attractive impurity at the center of the constriction. a, the channel becomes wider by $\Delta w = 0.2\lambda_F$ for $0.25\lambda_{FJ} \leq z \leq 0.75\lambda_F$. b, the potential well of $\Delta\phi_m(z) = -0.2E_F$ for $0.25\lambda_F \leq z \leq 0.75\lambda_F$. c, cosine modulation to enlarge the channel by $\Delta w = 0.3\lambda_F$. d, same as b, but with $\Delta\phi_m = 0.3E_F$. e, an attractive impurity at $y_I = 0.13\lambda_F$, $z_I = 0.5\lambda_f$ with strength $\beta = -0.5k_F$ and lateral decay length $q^{-1} = 2\lambda_F$.

finite, and then RT will occur whenever its energy is aligned with E_F. In order to have unity transmission, the interactions of the bound state with the left- and right-hand sides of the 2D EG has to be equal. For a symmetric QPC this is only possible when the impurity is placed at the center, i.e., $z_I = d/2$. The RT effect due to an attractive impurity is illustrated by curve e in Fig. 8. We note that, since there are propagating subbands for $w > \lambda_F/2$, $\tilde{\Gamma}$ will have both real and imaginary elements. As a result, the truly bound states cannot exist in this situation.

The above RT event is examined in detail in Fig. 9(a), where the variation of G below the first step is shown as a function of the impurity position z_I. Apparently, below the resonance, the conductance increases exponentially and becomes independent of z_I. This is ordinary tunneling, where the position of the impurity does not affect the conductance in any essential manner. Near the resonance, conductance rises rapidly and approaches its maximum value G_{res} at $w_{res} \simeq w_b$. For $w > w_{res}$, the conductance decreases again, and at $w = \lambda_F/2$ the first subband becomes propagating. The important observation is that the value of G_{res} is exactly equal to the quantum of conductance when the impurity is at the center, but it decays if the impurity moves away from the center. Also note that above w_{res} the conductance depends on z_I. This occurs because w_{res} is close to the propagation threshold for the first subband, and the tunneling probability

through the QPC is large. Two important quantities extracted from this analysis are w_{res} and the sharpness of resonance. The sharpness is quantified by the full width at half maximum (FWHM). For a decreasing resonance peak value, the FWHM increases and w_{res} shifts to larger values. Both of these effects are results of the increasing interaction of the bound state with the 2D EG on one side. The larger this interaction is, the larger is the escape probability, and hence the wider is the resonance (i.e., the larger is the FWHM). In addition, renormalization of the bound state due to interaction with 2D EG states increases and w_{res} deviates from w_b and shifts to larger values of w.

Figure 9(b) shows w as a function of the strength β of the impurity. Note that the impurity is located at the center $(z_I = d/2)$, so that $G_{res} = 2e^2/h$. Naturally the bound states become deeper as the magnitude of the strength is increased. On the other hand, w_{res} increases with increasing q. As pointed out earlier, the integrated strength of the impurity $\sim \beta/q$ becomes smaller for large q and the bound state gets closer to the propagation threshold, i.e., $w = \lambda_F/2$. For very small q, the impurity potential is almost flat in the transverse dimension, so that $\bar{u}$ is diagonal. For this case, the bound states are the exact solutions of the 1D problem, having a single δ-function scatterer with $w_b = \lambda_F/\{4[(\beta/k_F)^2 + 1]\}^{1/2}$ and binding energy $E_b = -\hbar^2\beta^2/2m^*$ (which is the energy difference between the bound state and the first subband from which it is split off). For this special case, all subbands have real bound states below them. For finite q, that is, for nonvanishing intersubband interaction, it is not possible to have a completely localized solution for subbands other than the first one. This leads to the formation of resonance states below the higher-lying steps. The actual bound states (lying below the first subband) are used in the analysis of the RT effects in order to focus on the interaction of the bound state with the 2D EG states, and to eliminate the interactions between the resonance state and the propagating states in the channel. One last observation about w_{res} is that, for certain attractive impurities, no resonances are observed. In two dimensions, all attractive potentials have bound states, so a value of w_b can always be identified for attractive impurities. Nevertheless, since the binding energy is extremely small for very small scatterers, the renormalization overcomes the binding, and the bound state does not seem to be split off the subband from which it originates. They can be resolved for longer QPC's, however, due to reduced renormalization effects.

The effects of the bound-state–continuum interactions are analyzed in Figs. 9(c) and 9(d). In Fig. 9(c), the width w_{res} is compared to w_b, which is determined by the zero of the determinant of $(\tilde{\Gamma} + i\bar{u})$. As long as the bound state is deep, $|\gamma_b|$ is large, so the prefactor for hybridization of the bound state with the continuum, $\exp(-|\gamma_b|d)$, is negligible. Thus, the resonant transmission occurs exactly at the same width as does the bound state, i.e., $w_{res} = w_b$. As the binding energy decreases, so does $|\gamma_b|$, and then the hybridization energy may be comparable with E_b. For such a case, the presence of 2D
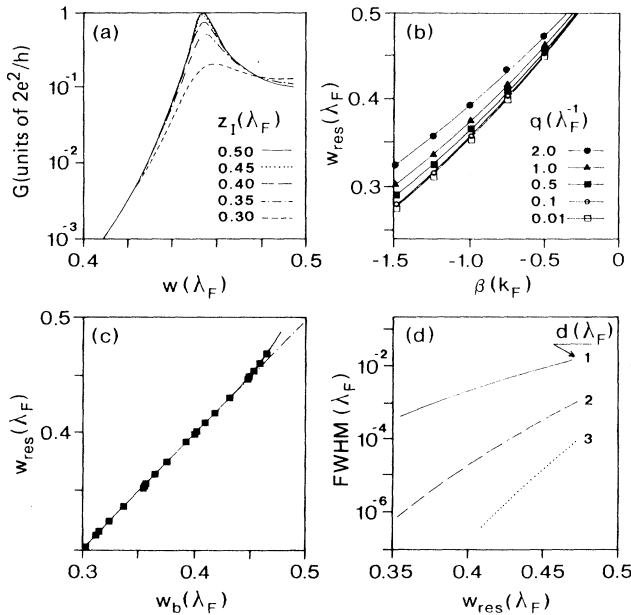

FIG. 9. (a) $G(w)$ for different longitudinal positions of the attractive impurity z_I in a uniform constriction of length $d = \lambda_F$, $y_I = 0.13\lambda_F$, $\beta = -0.5k_F$, and $q^{-1} = 2\lambda_F$. (b) The width w_{res} at which the RT occurs vs β for different values of q. (c) w_{res} vs the width w_b at which the infinite channel has a bound state. The length $d = \lambda_F$ and $z_I = d/2$. Deviation from the dashed-dotted curve is the normalization energy. (d) FWHM as a function of w_{res} for varying length d of the constriction.

EG's leads to a shift of the resonance with respect to w_b towards larger w values, that is, towards smaller E_b. The deviation of w_{res} from w_b is proportional to $\exp(-|\gamma_b|d)$, as expected. The sharpness of the resonances are displayed in Fig. 9(d) for different lengths d. The FWHM decreases as either the binding energy or the length is increased. This is in agreement with the above arguments for the escape rate and tunneling probability. Since the bound states shown in the figure are far below the propagation threshold, the renormalization effects are not important and the wave function varies with $\exp(-|\gamma_b||z-d/2|)$. Therefore, the FWHM is proportional to $\exp(-c|\gamma_b|d)$, which is in agreement with the numerical results in Fig. 9(d) obtained over more than four decades. The exponent c is approximately equal to 1.5, but an analytical calculation of c is quite complicated due to the effects of the quasi-1D nature.

Earlier, Smith et al.[61] reported the observation of quasi-0D states formed between two potential barriers created by gate depletion. However, their results did not provide clear evidence for a RT event, since the resistance but not the conductance curve has peaks. Subsequently, Moreno and Smith[62] studied the effect of the longitudinal potential profile on ballistic transport and especially on resonance tunneling. Their approach, which considers an infinitely long constriction and neglects the 2D EG, differs from the present study. For the same reason, the resonance peak below the first subband, which is studied extensively in this section, is missing in their results. Recently, McEuen et al.[63] observed RT peaks due to a single impurity in a QPC. Their results are in agreement with RT through a quasi-0D state coupled to a 2D EG continuum via a Breit-Wigner-type resonance interaction.

VII. SADDLE-POINT EFFECTS IN THE CONSTRUCTION

Most of the preceding analyses were based on the assumption that in Eq. (1) all of the z-dependent part, i.e., $\phi_m(z)$, can be separated from the total potential $V(\mathbf{r})$. Furthermore, it is also assumed that $\phi_m(z)=0$. Of course, this is an oversimplification, and the actual lateral potential varies normally along the constriction. Earlier, the contour plots of potential obtained from our SCF pseudopotential calculations[38,39] for STM operating near a mechanical contact revealed that $V(\mathbf{r})$ has a saddle-point structure near the plane bisecting the gap. The form of the potential in the constriction between tip and surface in STM, and also in the field emission of focused electrons[20] from a sharp tip, is three dimensional, but has a saddle-point structure between the transverse and longitudinal (propagation) directions. Earlier, we developed a tubelike model potential, which contains all the essential ingredients of SCF calculations (such as the saddle-point structure).[36] In the following section we will make further use of the 3D model potential.

In Fig. 10 we present the contour plots of the potential for the Al tip and Al(111) surface, calculated using the SCF pseudopotential method. In these calculations, the tip was represented by a pyramid consisting of four
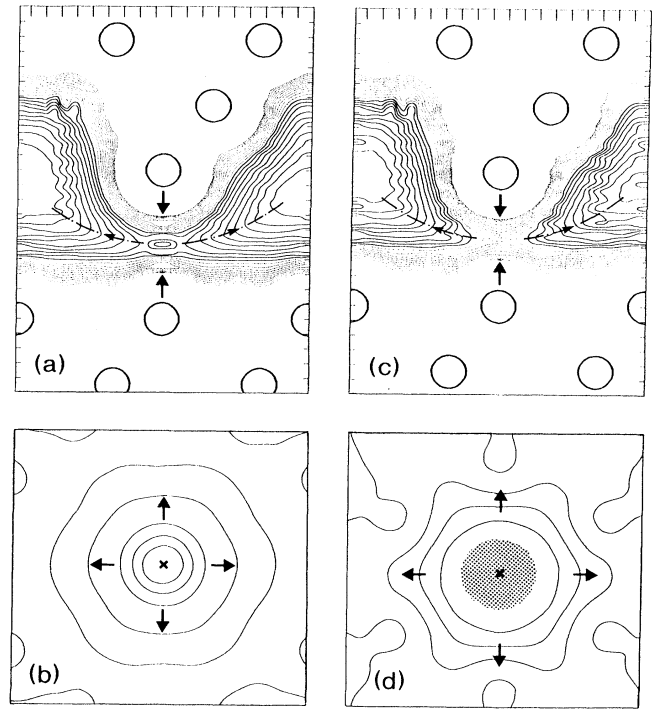

FIG. 10. Potential energy of a QPC calculated by using the SCF pseudopotential method described in Ref. 62. The QPC created by an atomically sharp, pyramidal Al tip and Al(111) surface, which are separated by d. (a) Contour plots of the potential in a longitudinal (y,z) plane, with circles indicating Al atoms, and (b) in a transverse (x,y) plane near the saddle point for $d=4.8$ Å. Plots (c) and (d) correspond to $d=5.3$ Å. Solid and dotted curves correspond to $V(\mathbf{r})>E_F$ and $V(\mathbf{r})<E_F$, respectively. The potential increases in the directions indicated by small arrows.

atoms, which is attached to the base Al(111) slab. This pyramidal tip is periodically repeated resulting in a (3×3) tip array. The artificial periodicity produced by a supercell consisting of 49 Al atoms was used to represent the wave function by a basis set of ~2000 plane waves. Since the lateral periodicity is large $(\sim9$ Å$)$, the intertip interactions have no significant effects on the results. The details of these calculations will be published elsewhere.[64] It is clearly seen that, for $d=5.3$ Å (10 a.u.), the saddle-point value of the potential, $V_{sp}=\max[\phi_m(z)]$, occurs near the plane bisecting d, and that $V_{sp}>E_F$, whereas for $d=4.8$ Å (9 a.u.), the saddle point dips into the Fermi level. In the transverse plane the potential displays an approximately quadratic variation.[44] More recently, Büttiker[65] presented a comprehensive analysis of the saddle-point structure by using adiabatic approximations and the quadratic potential, $V=-\beta z^2+\alpha y^2$. He showed that the quality of quantization of the conductance for a saddle-point constriction increases with an increase in the value of β/α. To complete this analysis, it is also necessary to clarify the range of the ratio β/α in which the adiabatic approximation is valid. Note that $\alpha/\beta\rightarrow\infty$ if $\beta\rightarrow0$, i.e., $\phi_m(z)\rightarrow0$ (or a constant value). This is the sit-

uation we described for the quantization due to the parabolic confining potential in Fig. 2(b). In the other extreme case, as justified by the results of Büttiker,[65] $\alpha/\beta \to 0$ corresponds to the quantum Sharvin regime.

Apparently, the saddle-point structure of the potential in the QPC is a reality. We have further investigated[66] this effect here by using two types of saddle-point structure. In this way we hope to clarify this effect within the framework of the present formalism, and to prepare the ground for further study in the following section. At the same time, our numerical results obtained from the transfer-matrix method will be compared with those of Büttiker[65] obtained analytically. Since we do not use any restriction, such as the adiabatic approximation, the present results have more general validity but less transparency.

The first type of saddle-point structure we consider has not been studied previously. The full-potential $V(\mathbf{r})$ is not separable, and is expressed as $V = \alpha(z)y^2$ in the region $0 \le z \le d$. Its lateral extent at any point z in the constriction is given by $w(z) = (\hbar^2/2m^*\alpha)^{1/4}$. The variation of width, $w(z) = w_0 + \eta\lambda_P[2(z-d/2)/d)]^2$, produces an effective potential barrier with a saddle-point structure. The second type is similar to that in Fig. 10, as well as to that used by Büttiker,[65] and is expressed for a finite constriction of d as

$$V(\mathbf{r}) = \eta E_F \left[1 - \left[\frac{z-d/2}{d/2} \right]^2 \right] + \alpha y^2 . \tag{30}$$

The first quadratic term on the right-hand side corresponds to $\phi_m(z)$ and reaches the maximum value ηE_F at the center of constriction, $z = d/2$. Note that by including an additional quadratic term, such as αx^2 one can obtain a 3D, tubelike constriction along the z direction. Figure 11 outlines our numerical results. In compliance with the above discussion concerned with adiabaticity, the resonance structure due to reflections from the boundaries disappears. This demonstrates that not only the tapered entrance, but also the saddle-point structure can destroy the resonance structure. In the curves, the features that are dependent on the form of the potential [or $w(z)$] are the sharpness of the transition between the plateaus[67] (or the extent of the plateaus) and the value of the conductance at the center of the plateaus. For small values of η, which correspond to smoother longitudinal variations, the steps are relatively sharp and thus the transition region is small. Also, the value of the conductance is close to the quantum values. This implies that the effective barrier due to the lateral confinement is significant, and the contribution of tunneling is negligible. However, as η increases, the steps become smoother with the increasing contribution from evanescent waves and the flat plateaus, and thus quantization gradually disappears. Very large values of η correspond to Sharvin's quantum conductance. These conclusions are in line with those obtained by the analytical study.[65] At this point it is in order to emphasize that the transfer-matrix method, as applied in the present study, suits our purpose of investigating potentials that are neither separable nor analytically solvable.
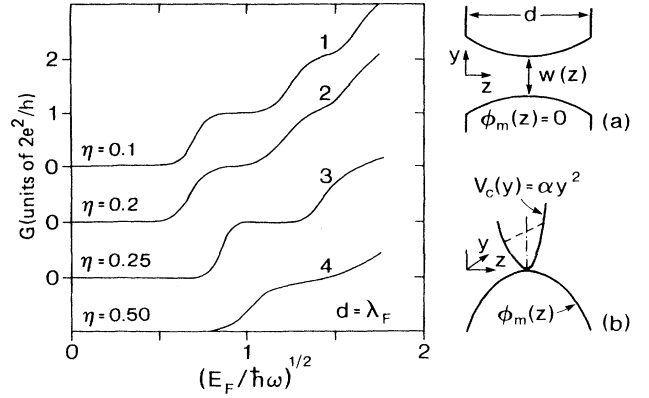

FIG. 11. Conductance G of a QPC with a saddle-point structure described by insets (a) and (b). The transverse confining potential is parabolic, $V_c(y) = \alpha y^2$, so that $w = (\hbar^2/2m^*\alpha)^{1/4}$ and $\alpha = \omega^2/2$. Curves 1 and 2 correspond to the a-type saddle-point structure; curves 3 and 4 correspond to the b-type saddle-point structures.

VIII. APPLICATIONS IN SCANNING TUNNELING MICROSCOPY

The link between STM and the electronic properties of the electrodes can be provided by the calculation of the current[68] as a function of the lateral and vertical position of the tip under a given bias voltage. If the tip-sample interaction is negligible, the tunneling current can be calculated by using the transfer Hamiltonian approach,[69] based on first-order time-dependent perturbation theory, and by replacing the many-body wave functions with the one-electron wave functions of the independent electrodes. Of course, even this approach requires detailed knowledge of the electronic states. As the tip approaches the sample, the potential barrier Φ_b is gradually lowered, leading to gradually increasing tip-sample interactions.[38,39,68] The electronic charge is rearranged and the ions are displaced to attain the lowest total energy at the preset tip position. Owing to significant overlap at small tip-sample distances d, the states of bare electrodes are strongly disturbed. They may even be combined to form site-specific resonance states.[38,39,70] Owing to the displacement of the outermost tip atoms in the attractive force field, the actual vertical motion of the tip deviates from that recorded in the topographic mode.[38] This implies that theories based on first-order time-dependent perturbation theory can no longer be used to calculate the tunneling current at small values of d.

Upon closer approach of the tip, the potential barrier collapses. This happens in the decreasingly attractive force range (acting on the outermost tip atom)[71] before mechanical contact occurs. For example, this critical distance[39] for an Al tip and Al sample is 5 Å (9.5 a.u.), as presented in Fig. 10. However, since the wave function is laterally confined between the tip and sample, the energy ε_1 of the lowest propagating state (which is the solution of the Schrödinger equation in this region) may occur above E_F, even if Φ_b has collapsed. This state acts as an

effective barrier[39,40,68] $\Phi_{eff} = \epsilon_1 + \Phi_b - E_F$ for the incident current-transporting states near E_F. SCF calculations, carried out for an Al tip and Al sample,[39] predict strongly site-dependent features for the effective barrier in this range of d values (~ 4.5 Å). The estimated values at the top (T) and hollow (H) sites were $\Phi_{eff} \simeq 2.3$ and 1 eV, respectively. If the energy of the lowest state quantized in the orifice between the tip and sample (i.e., the lowest subband energy, ϵ_1) dips into the Fermi level, a point contact is initiated, and the character of the transport changes from a tunneling to a ballistic regime. This corresponds to the opening of the first channel in Fig. 2(a). Even if the lateral extent of the point contact is small for certain atomically sharp tips, and hence $\Phi_{eff} > 0$, ballistic transport can eventually occur due to induced plastic deformation. The gradual collapse of the barrier as d decreases and its delayed effect on the tunnel current have been illustrated by studies of the transition from tunneling to essentially single-atom contact.[18]

The above description of the physical events taking place in STM points to the fact that there is always a constriction between electrodes. Depending on the value of the narrowest diameter of the constriction measured in terms of λ_F, and/or on the maximum value [or saddle-point value $V_{sp}(d)$] of the potential in this constriction, the transport occurs either in the tunneling or in the ballistic regime. In this respect, the geometry of the constriction and the underlying physics have close bearings on the transport through QPC's treated in the preceding sections. This section will clarify that the *confinement* of current-transporting states in the *constriction* is crucial for STM operating at small d. The second important fact to emerge from the SCF calculations is that tip-sample interactions at small d are on the atomic scale. On the other hand, the calculation of the tunnel current by the *ab initio* methods, which can provide almost a full treatment of these atomic-scale interactions, are not feasible due to conditions requiring extremely fine sampling of wave functions in the Brillouin zone.[38] Earlier, the tunnel current was calculated by evaluating the current operator with respect to the wave functions of the jellium electrodes, and important aspects of STM were thus revealed.[72] However, the full treatment of the atomic-scale interactions could not be included in those calculations.

The method[36] we have developed to calculate the current in STM-related systems aims to treat the atomic-scale interactions by using the potential $V(\mathbf{r})$ obtained from detailed SCF pseudopotential calculations. In addition, some features of STM (which are relevant to the mesoscopic transport but have previously gone unnoticed) will be revealed within the framework of the present theory. In this context, the aspects of STM related to quantum ballistic transport are emphasized. After all, the scanning tunneling microscope is a mesoscopic device.

We modeled the tip-sample system by two jellium electrodes separated by a potential barrier depending on d. Only at the point of the contact is a constriction (or an orifice) formed. Since the SCF calculations on a sharp tip facing the metal-sample surface predict an approximately quadratic lateral potential in the gap, the potential of the

constriction for a given d is represented by

$$V(d,\mathbf{r}) = \phi_m(d,z)$$
$$+ \alpha(d,z)\rho^2 \Theta(z + d_j/2)\Theta(z - d + d_j/2) , \qquad (31)$$

where $\rho = (x^2 + y^2)^{1/2}$. The term $\phi_m(d,z)$ can be calculated, either from the jellium model, or obtained from SCF calculations. Note that the maximum of $\phi_m(d,z)$ is equal to $V_{sp}(d)$, and thus that $\max[\phi_m(d,z)] = \Phi_b$ if $V_{sp}(d) > E_F$.[73] Twice the distance between the first atomic plane and the jellium edge is denoted by d_j. Subsequent to the collapse of the barrier the radius of the orifice at E_F, $\rho_m = \{[E_F - V_{sp}(d)]/\alpha\}^{1/2}$, increases as d decreases. Note that the material and tip-specific features of the contact are represented by $\alpha(d,z)$.

In order to calculate the conductance for the model potential given in Eq. (31), we consider the current-transporting states. These are 3D plane waves in the electrodes. Since ϕ_m and α vary with z, we divide the constriction into discrete segments; in each segment, the wave function consists of the products of the 2D isotopic harmonic-oscillator states (in the xy plane) and 1D plane waves in the propagation direction. Applying the transfer-matrix method and, hence, solving the current-transporting states, we obtain the current (or conductance).

At this point we will also discuss the variants of the above potential, which are applicable to different problems in STM-related systems: for example, modifying $\phi_m(d,z)$, such that it can be lowered below E_F in the vicinity of a specific point z_I. In the same region, $\alpha(d,z)$ is modified to represent the potential of an atom or a substance between the tip and sample surface. As discussed extensively in Secs. V and VI, such an attractive region in the constriction gives rise to bound states corresponding to the states of an atom or molecule in the gap between the tip and sample surface. When one of these bound states coincides with the Fermi energy of the left electrode (which contains the incident electrons), the transmission coefficient increases suddenly, leading to resonant tunneling.[35] Of course, the model potential used to simulate atoms or molecules is too crude. However, even such a simple model is able to reveal some crucial aspects.

Field emission of focused electrons from a sharp tip is another phenomenon to which our theory can be successfully applied. Fink[20] showed that relatively low bias voltage (< 1000 V) applied between an atomically sharp W(111) tip and a screen (separated by centimeters) can yield a low-energy electron beam with a significant emission current of $\sim 10\ \mu$A. Moreover, the beam is well collimated, so that the angle of spread was measured to be only $\sim 2°$ at the screen. A proper understanding of the effects resulting in such a well-collimated and coherent electron beam has been of recent interest. The earliest theoretical work[74] found that the triangular potential barrier (through which the emitted electrons tunnel) is exclusively responsible for tunneling. Later, based on the SCF-jellium calculations, Lang *et al.*[75] argued that a constriction of potential, with a hornlike opening to the vacuum, forms in front of the tip. Low lateral momentum

states selected by this constriction evolve adiabatically in the horn to yield a well-collimated electron beam.

This field emission of electrons differs from STM in two essential points. (i) The sample surface, which is usually ~ 5–10 Å from the apex of the tip in STM, is replaced by the screen, which lies in the range of centimeters from the point source. (ii) The electric field $\mathscr{E}$ is a crucial parameter, so that the potential in Eq. (31) shows a strong field dependence. The potential $\phi_m(z)$ passes through a maximum a few angstroms beyond the atomically sharp tip, is lowered below the Fermi level, is subsequently flattened, and finally reaches the screen. These aspects, which are characteristic of field emission, can be easily be included in our model potential representing the constriction along $d_t < z < d$ in front of the tip,

$$V(\mathscr{E},\mathbf{r}) = \phi_m(\mathscr{E},z) + \alpha(\mathscr{E},z)\rho^2 \Theta(z+d_t)\Theta(d-z) . \quad (32)$$

The bimetallic junction potential $\phi_m(\mathscr{E},z)$ is calculated in the jellium approximation for two jellium edges placed at $z = 0$ (tip) and $z = L$ (screen). The electric field $\mathscr{E}$ is incorporated following Orosz and Balazs.[76] The hornlike apertures to the emitter and to the vacuum at the screen side are simulated by uniformly varying the lateral extent of the confining potential $w = (\hbar^2/2m^*\alpha)^{1/4}$. We note that the precise form of the potential $V(\mathscr{E},\mathbf{r})$ cannot be attained by the state-of-the-art calculations based on the local-density approximation. While rigorous methods have this limitation, our approach allows one to perform a systematical analysis of various effects [such as confinement in the constriction and subband selection through effective barriers, hornlike opening to the vacuum, and increase of the longitudinal momentum in the region where $\Phi_{\mathrm{eff}}(z) < E_F$], all of which contribute to collimation. Calculating the current as described in the preceding sections, and defining the angular spread in terms of the ratio of the expectation value of the transverse to longitudinal momentum, one is able to deduce how the collimation angle depends on the parameters responsible for collimation. A recent study by Tekman, Ciraci, and Baratoff[77] in line with the above theory found that the collimation occurs mainly due to the selection of low transverse momentum states by the *effective barrier* in the constriction. It is improved by the electric field and the hornlike opening to the vacuum, even though the adiabatic approximation is not valid.

A. Transition from tunneling to ballistic regime

We will now consider a specific mode of operation of STM; under the constant bias, starting from the low current level (at large d) the tip gradually approaches the sample surface.[36] In a similar setup, using an Ir tip and a Ag sample, Gimzewski and Möller measured the variation of the current as a function of d. Their $\log_{10}I(d)$ plot showed that the current initially increases exponentially with decreasing d. This is characteristic of the tunneling regime. The discontinuity observed at small d was attributed to the mechanical instability[78] (or hysteric deformation). The value of the conductance G just after the discontinuity was only $\sim 80\%$ of the quantum of conductance. Upon further approach of the tip, I continues to

increase and exhibits an oscillatory behavior. This was one of the unusual modes of behavior observed from the above experiment.[18] Initially, these oscillations were attributed to the quantum effects like those in Sharvin's quantum conductance.[19] Most of the data taken after the hysteric deformation were not reproducible. There were also measurements, in which the current dropped suddenly, near the point of mechanical instability. Subsequently, different studies[79] presented data showing that G is either saturated below the quantum of conductance or increased continuously above it. In our study we analyze observed variations in the behavior of the $\log_{10}I(d)$ curve. In particular, we concentrate on the observed oscillations of $G(d)$ after the mechanical instability.

In order to link theory with experiment, we obtain α by using the diameter of contact given by experiment and by scaling those values calculated for the Al tip and the Al sample as a function of d. Furthermore, we used the electronic parameters of Ag for jellium electrodes. Our results are illustrated in Fig. 12(a). In agreement with previous calculations,[72,80] the conductance associated with a uniform constriction set up by a single atom at the vertex of the tip has a value less than the quantum of conductance. Since the length of the constriction is finite and within the range of internuclear distance a_0 (i.e., the
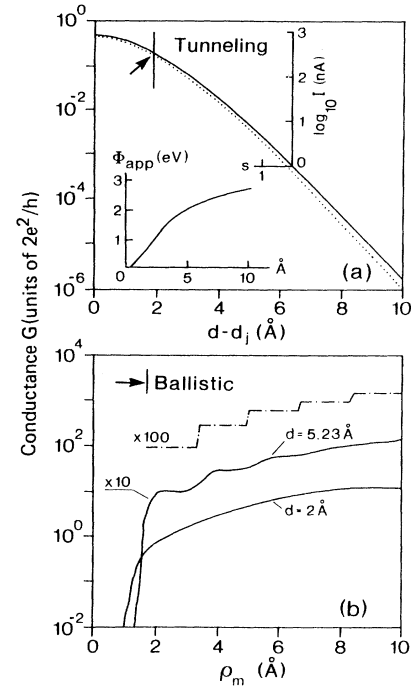

FIG. 12. (a) Conductance $G(d)$ [and $\log_{10}I(d)$] vs tip displacement under the bias voltage $V_b = 20$ mV, calculated for two Ag electrodes. The dotted curve is the contribution of the first channel. The tip-sample distance corresponding to mechanical instability is indicated by an arrow. The inset shows the apparent barrier height Φ_{app} calculated from the current. (b) Conductance vs the diameter of the orifice $\rho_m = \sqrt{E_F}/\alpha$ calculated for $d = \infty$ (i.e., exact quantization is shown by dashed lines); $d = 5.23$ Å ($= \lambda_F$); and $d = 2$ Å ($\sim a_0$). Steps correspond to the opening of conduction channels.

sum of the atomic radii), this result implies that the energy of the first subband ε_1 is still above E_F (i.e., $\Phi_{eff} > 0$) and, hence, the conductance is dominated by tunneling.

From the behavior of G presented in Fig. 12(a), it appears that point contact sets in after the structural instability occurs, at $d > a_0$. While d cannot be smaller than a_0, by pushing the tip further the contact area expands due to increasing plastic deformation followed by adhesion of nearby atoms. The actual form and size of the contact after the point of mechanical instability is uncertain. Several parameters (such as the detailed atomic structure of the apex of the tip, the strength of the intraelectrode and interelectrode interactions, etc.) are expected to influence the contact. For example, the apex atom in the repulsion force range can be displaced in such a way that the aperture of the orifice is reduced, incidentally causing G to decrease temporarily in the course of the approach. Apart from this exceptional case, we assume that the radius of the contact increases upon the approach of the tip. Then, the conductance is related to the radius ρ_m, which increases normally with increased plastic deformation. This is reminiscent of the situation in the QPC shown in Fig. 2(a). As the aperture (diameter) of the contact increases, subbands due to the quantization in the constriction dip sequentially into the Fermi level, causing G to increase by $2e^2 n/h$ each time (n is the degeneracy). As revealed in our analysis related to the QPC in Sec. III, in spite of these expected jumps in G, the perfect quantization with a sharp step structure can only be observed if the length of the contact is longer than λ_F.

To investigate the variation of G in the course of plastic deformation, we calculated the conductance as a function of $\rho_m = \sqrt{E_F/\alpha}$, for $\phi_m(d,z) = 0$ at fixed d. Our results are illustrated in Fig. 12(b). The first channel is opened at a radius as low as $\rho_m = 1.5$ Å followed by a rise of $\sim 2e^2/h$ in G. The pronounced oscillations (or smeared-out step structure) of $\log_{10} G(\rho_m)$ are apparent only for $d > \lambda_F$. However, for $d \simeq a_0$, weak oscillations are washed out in the logarithmic scale. It is seen that the point contact between the tip and sample is not long enough to allow steps or pronounced oscillations in the $\log_{10} I(d)$ curve to form. The observed oscillations possibly originate from the irregular motion of the atoms as the tip is uniformly pushed towards the sample, causing irregular enlargement of the contact area. Also the atoms of a blunt tip may undergo sequential contacts, with each contact opening a new orifice and leading to abrupt changes in the current. Both cases give rise to the variation of $\log_{10} I(d)$ as observed experimentally.

Apart from the contact system relevant to the experiment by Gimzewski and Möller,[18] we also analyzed $G(d)$ by using $\alpha(d)$ values corresponding to different tips. Depending on the shape and material properties of the tip, the form of $G(d)$ may exhibit significant, even *qualitative*, changes from that in Fig. 12(a). For example, G decreases, passing through a maximum[80] ($\sim 2e^2/h$) if ρ_m is allowed to be less than the atomic radius of Ag. On the other hand, if α is small, $G(d)$ may reach a plateau[72] before the point of discontinuity that results from the

mechanical instability. This suggests that Φ_{eff} collapses prior to the hysteric deformation, but the value of G at the plateau may be smaller than the quantum of conductance owing to scattering by the ions in the constriction. Certain contacts may have several subbands close to E_F, each contributing to tunneling. In this case, plateaus do not occur, but $G(d)$ increases almost linearly. For a flat tip (with several atoms at the apex) the contact area is large and, hence, G rises above $2e^2/h$ well before tip-induced plastic deformation sets in. Note that all these arguments are based on the assumption that there is neither an oxide nor a flake between the tip and sample, as this would influence the variation of $G(d)$ dramatically, as will be seen later in this section. These results indicate that the character of transport, and the variation of G as a function of d are not generic, but are strongly tip and sample specific. The plateau can appear before the point of discontinuity only under certain conditions.

B. Anomalous corrugations of flat metal surfaces

One of the peculiar results obtained using STM is that the corrugation of the nominally flat (111) surfaces of the noble[81] and simple[82] metals was much larger than one could deduce from charge density of the free surfaces. STM data obtained on the Al(111) surface have attracted even more interest, because this surface does not have any localized state near E_F that could affect the corrugation. Recent studies have come to different conclusions in attempting to explain this anomalous corrugation. For example, Winterlin et al.[82] argued that the STM-observed corrugation of the Al(111) surface is enhanced by the elastic deformation of the tip which, in turn, is induced by the attractive forces between two electrodes. The calculations by Ciraci, Baratoff, and Batra[39] have been at variance with these arguments. They showed not only that the observed corrugation is reduced by the tip-induced elastic deformation but also pointed out an important feature of STM, i.e., induced site-dependent electronic effects. The proximity of the tip induces site-specific and laterally confined states. In particular, a site-dependent effective barrier sets in, owing to the lateral confinement of states. In a different study, in line with the earlier suggestion by Baratoff,[83] Chen[84] was able to reproduce the observed $I(d)$ curve leading to the huge corrugation of the Al(111) surface by taking a pure d_{z^2} orbital on the outermost W atom of the tip. The theory of Chen[84] has been questioned on the ground that the W tip is covered by the Al atoms from the sample. Even so, his results justify the crucial role played by the p_z-type resonance states (or tip-induced states[70]) proposed by Ciraci et al.[39] These states form owing to the much lowered potential at the apex of the tip and have densities concentrated near and across the tip atom. There have also been other studies proposing different mechanisms. Clearly, the observed anomalous corrugation is a challenging problem of STM, and its interpretation is a critical test in understanding the 3D tunneling between tip and sample.

Since the electronic states of the free electrodes are *modified* by tip-sample interaction in the experiment,

these modifications have to be taken into account in the studies based on the first-order perturbation theory. As we mentioned at the beginning, we take a different approach[85] to this study. We start with a realistic potential (including atomic-scale interactions) rather than with the electronic states of free electrodes to calculate the current. Two features in our formalism, namely the variation of $\phi_m(d,z)$ with lateral position of the tip and the form of $\alpha(d,z)$, are crucial for the present study.

It is known that the jellium approximation alone does not convey any information regarding the corrugation of the sample surface, even though it is appropriate to calculate the tunneling current. Using only the $\phi_m(d,z)$ value from the jellium approximation, one may perhaps obtain an overall picture of the behavior of the tunneling current as a function of d, but not of its variation with the lateral position of the tip for fixed d. The calculation of the tunneling current by state-of-the-art *ab initio* methods is also hindered, because the states are discretized in the k space. Here we implement the corrugation of $\phi_m(l,z)$ [or $\phi_m(d,z)$; see Fig. 13(a) for the relation between l and d] obtained from the SCF calculations into the jellium model, and calculate the tunneling current to infer the STM corrugation. In Fig. 13(a) we show the corrugation of $\phi_m(l,z)$ for various tip-sample distances. We see that the effective width $\zeta(E,l)$ of the potential $\phi_m(l,z)$ at fixed energy is consistently larger at the H site than the T site, but $\Delta\zeta(E,l)=\zeta^H(E,l)-\zeta^T(E,l)$ decreases with increasing d. At large d, $\Delta\zeta \to 0$ as expected. In the
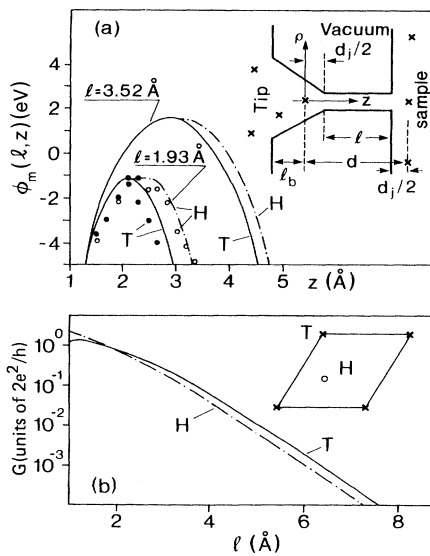

FIG. 13. (a) Variation of the longitudinal potential $\phi_m(l,z)$ in a constriction between the Al tip and Al(111) surface, calculated by the jellium model with the inclusion of corrugation obtained from the SCF pseudopotential calculations in Ref. 62. SCF results are indicated by solid and open circles for the top (T) and hollow (H) sites. The inset shows a schematic description of the constriction. The atomic positions are indicated by crosses. (b) Conductance vs l (i.e., the distance between jellium edges $l=d-d_f$) calculated for the top and hollow sites described in the Al(111) surface unit cell by the inset.

earlier STM studies it was assumed that $\Delta\zeta=0$, always. It is becoming clear that this is true only for very large tip-sample separations.

It is known that the potential $V(\mathbf{r})$ between the surface and the vacuum is dominated by the exchange-correlation potential, whose corrugation near E_F is much smaller than the values given in Fig. 13(a), even with the local-density form ($\sim\rho^{1/3}$). The value of $\Delta\zeta$ can, in fact, be large; this occurs, however, at very low energies, but is not normally accessible by STM. The origin of the enhanced corrugation of the potential barrier or $\Delta\zeta$ is found in the *tip-induced modifications* in the electronic states and hence, in the charge density. Earlier it was shown that, for $d=4.2$ Å (or 8 a.u.), the charge-density saddle-point value of the combined tip-sample system is one order of magnitude larger than twice the value calculated for the bare surface.[39] Moreover, at the same d, the charge density was found to be strongly site dependent for an atomically sharp tip, i.e., 5×10^{-3} and 4×10^{-3} electron/(a.u.)^3 for the T and H sites, respectively. The corrugation of potential, $\Delta\zeta$, is enhanced by the enhanced corrugation of the surface charge density. This important feature was pointed out earlier, in terms of the tip-induced states having significant weight in the vicinity of the tip.[39] This ingredient of the SCF potential[68] is included in our model potential by uniformly widening $\phi_m(l,z)$ (calculated from the jellium approximation) by $\Delta\zeta$ to obtain the $\phi_m(l,z)$ value at the H site.

The form of $\alpha(l,z)$ determines the lateral confinement of the states between the tip and sample. The larger α is, the stronger the confinement is, and thus the higher the subband energies are. This implies that for an orifice with relatively larger α, the electrons tunnel through an effective barrier, even if Φ_b collapses. We determine α by fitting the SCF potential[59] calculated for various values of d to the model potential given in Eq. (31). For the sake of simplicity, we use the single value of $\alpha(l)$ in the range $d_j/2<z<l+d_j/2$. The calculated G versus l curve is presented in Fig. 13(b). We distinguish three regions in the variation of G. For large l (≥5 Å), $\log_{10}G(l)$ is an approximately straight line with a constant negative slope. This is indicative of tunneling behavior. In this range, the corrugation of the tunneling current in the topographic mode is ~0.3 Å. This value is in agreement with the experiment,[82] which reports a tunneling current of $\sim10-20$ nA for $l\sim5.5$ Å, and for a bias voltage of 50 mV. Note that the corrugation should not be constant for $l>6-7$ Å, as it decreases with increasing l and in the end it diminishes. This region is beyond the range of our calculation, and thus is not treated. In the intermediate region, $2\leq l\leq4$ Å, the corrugation starts to decrease again, in spite of the fact that $\Delta\zeta$ increases continuously with decreasing l. This paradoxical situation occurs because the lateral confinement and the effective barrier derived thereof at the T site become superior to those at the H site, and dominate the current. In the third region, $l<2$ Å, the current at the H site exceeds that at the T site. This is an inversion of the corrugation just before mechanical contact, and indicates that the H site appears as a protrusion (if, of course, it can be observed at all in the presence of mechanical instabilities).

IX. CONCLUDING REMARKS

In this paper we have investigated electron transport through different forms of constriction, which were represented by either quasi-1D or -3D confining potentials, with transmission along one direction. In these constrictions, the lateral size of confinement is in the range of the Fermi wavelength. Moreover, the phase breaking inelastic scattering is negligible owing to low temperature and short transverse distances. In some cases, the effective potential could rise above the Fermi level, and electrons can then tunnel through the barrier. These conditions go beyond the classical regime and require a quantum-mechanical treatment of the transport. The most important quantum effect of the transmission, the quantization of conductance, is now understood to some extent through several earlier works, as well as the present one. There are several factors that may be responsible for the deviation from ideal quantizations, since electrons in such a constriction are scattered by any kind of irregularity in the potential. It is therefore statistical quantities, such as the electron mean free path, that may not be relevant in these quantum systems. As we demonstrated, even the position of the impurity becomes important in the detailed structure of conductance. We could have calculated the charge distribution and the SCF-consistent potential in a quantum point contact. However, we believe that an attempt to develop a self-consistent scheme may not be very interesting, since the actual parameters of a constriction are not accessible. Clearly, the actual distribution of charge is not uniform. Nevertheless, we examined the effect of inhomogeneities such as nonuniform geometries and saddle-point structure on the potential, and revealed significant effects. On the other hand, issues such as the adiabatic approximation, transport under the magnetic field parallel and perpendicular to the transport direction, dimensionality effects such as

transport through a quantum wire as initiated here, the treatment of the full impurity potential, the "voltage drop" in the constriction, quantum collimation effects in 2D and 3D systems, etc., all pose unanswered questions and require further study.

The contributions of this paper are focused mainly on three points. First, we present a detailed formalism, which was lacking in our earlier short publications. Within the framework of this formalism, we discussed several aspects of quantum ballistic transport. We hope that the formalism presented here can be extended to treat cross geometries and more complicated transport events. Second, we presented an extensive discussion of bound states and relevant resonant tunneling phenomena, which were initially investigated by the authors. Third, we extended the approach to treat the transport in STM and related fields, such as the field emission of collimated electrons, with the hope of evolving a new perspective. In particular, the concepts of lateral confinement and the effective barrier created thereof are extensively exploited, as they are relevant to the transport at small tip-sample distances and the effect on the apparent barrier height. These authors believe that further research on STM with interacting tip and sample would reveal interesting features that would promote our knowledge of interatomic interactions and transport in mesoscopic systems.

ACKNOWLEDGMENTS

*Permanent address: Department of Physics, Bilkent University, Bilkent 06533, Ankara, Turkey.

[1]L. Esaki and R. Tsu, IBM J. Res. Develop. **14**, 61 (1970); L. L. Chang and L. Esaki, in *Molecular Beam Epitaxy*, edited by B. R. Pamplin (Pergamon, Oxford, 1980), p. 15; R. Dingle, in *Festkörperprobleme*, Vol. XV of *Advances in Solid State Physics*, edited by H. J. Queisser (Pergamon, Oxford, 1975), p. 21; T. Ando, A. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[2]*Resonant Tunneling*, edited by L. L. Chang and E. E. Mendez (Plenum, New York, 1981).

[3]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[4]S. Washburn and R. A. Webb, Adv. Phys. **35**, 375 (1986); A. G. Aronov and Yu. V. Sharvin, Rev. Mod. Phys. **59**, 755 (1987).

[5]R. Landauer, IBM J. Res. Develop. **1**, 223 (1957); Philos. Mag. **21**, 863 (1970); *Localization, Interaction and Transport Phenomena*, edited by G. Bergman and Y. Bruynseraeda (Springer-Verlag, Heidelberg, 1985), p. 38.

[6]P. W. Anderson, D. J. Thouless, E. Abraham, and D. S. Fisher, Phys. Rev. B **22**, 3519 (1980).

[7]M. Büttiker, IBM J. Res. Develop. **32**, 317 (1988).

[8]A. D. Stone and A. Szafer, IBM J. Res. Develop. **32**, 384 (1988).

[9]H. L. Engquist and P. W. Anderson, Phys. Rev. B **24**, 1151 (1981).

[10]E. N. Economou and C. M. Soukoulis, Phys. Rev. Lett. **46**, 618 (1981).

[11]G. Imry, in *Directions in Condensed Matter Physics*, edited by G. Grinstein and G. Mazenko (World Scientific, Singapore, 1986), p. 101.

[12]D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 6851 (1981).

[13]R. Landauer, IBM J. Res. Develop. **32**, 306 (1988); J. Phys. Condens. Matter **1**, 8099 (1989).

[14]M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[15]B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williams, L. P. Kouwenhowen, D. van der Marel, and C. T. Foxon, Phys. Rev. **60**, 848 (1988).

[16]D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L209 (1988).

[17]Before the discovery of the quantization of conductance, the transport experiments were performed in 1D constrictions or quantum wires with w and d being in the range of those in

Refs. 15 and 16, and the concept of channels was exploited. See G. Timp, A. M. Chang, and P. Mankiewich, Phys. Rev. Lett. **59**, 732 (1987); M. L. Roukes, A. Scherer, and S. J. Allen, *ibid.* **59**, 3011 (1987).

[18]J. K. Gimzewski and R. Möller, Phys. Rev. B **36**, 1284 (1987).

[19]N. Garcia (unpublished).

[20]H. W. Fink, IBM J. Res. Develop. **30**, 460 (1986); Phys. Scr. **38**, 160 (1988); H. W. Fink, W. Stocker, and H. Smith, Phys. Rev. Lett. **65**, 1204 (1990).

[21]Yu. V. Sharvin, Zh. Eksp. Teor. Fiz. **48**, 984 (1965) [Sov. Phys.—JETP **21**, 655 (1965)].

[22]M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985). This effect is, however, mainly due to the decreasing conductivity just below the step (see Ref. 32).

[23]G. Kirczenow, Solid State Commun. **68**, 715 (1988).

[24]L. I. Glazman, G. B. Lesovik, D. E. Khmelnitskii, and R. I. Shekhter, Pis'ma Zh. Eksp. Teor. Fiz. **48**, 238 (1988) [JETP Lett. **48**, 258 (1988)].

[25]I. B. Levinson, Pis'ma Zh. Eksp. Teor. Fiz. **48**, 273 (1988) [JETP Lett. **48**, 301 (1988)]; Zh. Eksp. Teor. Fiz. **95**, 2175 (1989) [Sov. Phys.—JETP **68**, 1257 (1989)].

[26]A. Szafer and A. D. Stone, Phys. Rev. Lett. **60**, 300 (1989).

[27]E. G. Haanappel and D. van der Marel, Phys. Rev. B **39**, 5435 (1989).

[28]L. Escapa and N. Garcia, J. Phys. Condens. Matter **1**, 2125 (1989); N. Garcia and L. Escapa, Appl. Phys. Lett. **54**, 1418 (1989).

[29]E. Tekman and S. Ciraci, Phys. Rev. B **39**, 8772 (1989); in *Science and Engineering of 1- and 0-D Semiconductors*, edited by S. P. Beaumont and C. M. Sotomayer-Torres (Plenum, New York, 1990), p. 99.

[30]S. He and S. Das Sarma, Phys. Rev. B **40**, 1456 (1989).

[31]P. L. Pernas, A. M.-Rodero, and F. Flores, Phys. Rev. B **41**, 8553 (1990).

[32]M. Yosefin and M. Kaveh, Phys. Rev. Lett. **64**, 2819 (1990).

[33]D. A. Wharam, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L887 (1988).

[34]L. Escapa and N. Garcia, Appl. Phys. Lett. **56**, 901 (1990). They calculated the non-self-consistent charge density, showing an accumulation of charge at the abrupt connection of the uniform constrictions. In view of the scattering of multiple reflected states from the corresponding nonuniform potential, they questioned the existence of coherent ballistic transport in a QPC. Certainly, the profile of the charge density is strongly dependent on the convergence and also on the geometry of the constriction. Nevertheless, the SCF charge density is expected to be nonuniform and thus scatters the electrons. Most of the discussion in our earlier work (Refs. 35 and 58) and in the present study are devoted to several scattering effects affecting the quantization.

[35]E. Tekman and S. Ciraci, Phys. Rev. B **40**, 8559 (1989); in *Basic Concepts and Physical Applications of Scanning Tunneling Microscopy and Related Techniques*, edited by R. J. Behm, N. Garcia, and H. Rohrer (Kluwer, Dordrecht, 1990), p. 157.

[36]S. Ciraci and E. Tekman, Phys. Rev. B **40**, 11 969 (1989).

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

[38]S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **41**, 2763 (1990).

[39]S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **42**, 7618 (1990).

[40]N. Lang, Phys. Rev. B **38**, 10 395 (1988).

[41]S. Ciraci and I. P. Batra, Phys. Rev. B **33**, 4294 (1986).

[42]In that work, each atomic chain corresponds to a single channel. There are cases in which a QPC initiated by a single atom may not support even a single channel. For further discussion, see Sec. VIII.

[43]These are relations like $\Xi_{nE}(q,0)=\Phi_n(q)$ and $\Xi_{nE}(q,d)=e^{i\gamma_n d}\Phi_n(q)$ given in terms of the transverse Fourier transform of the subband wave function, $\Phi(q)$.

[44]The work by S. E. Laux, D. J. Frank, and F. Stern, Surf. Sci. **196**, 101 (1988), and the work in Ref. 39 give an indication that the parabolic confining potential is a good approximation. However, when w is large and thus a large number of channels are occupied, the infinite-well potential seems to be a better approximation. These results are verified by D. A. Wharam *et al.*, Phys. Rev. B **39**, 6283 (1989). They showed that the most appropriate form of the confining potential $V_c(y)$ is a quantum well with parabolic walls.

[45]G. Timp, in *Condensed Systems of Low Dimensionality*, edited by J. Beeby (Plenum, New York, 1991).

[46]J. H. Davies and J. A. Nixon, Phys. Rev. B **39**, 3423 (1989).

[47]L. P. Kouwenhoven, B. J. van Wees, C. J. P. M. Harmans, J. G. Williams, H. van Houten, C. W J. Beenakker, C. T. Foxon, and J. J. Harris, Phys. Rev. B **39**, 8040 (1989).

[48]A. Yacoby and Y. Imry, Phys. Rev. B **41**, 5341 (1990).

[49]S. Ciraci and I. P. Batra, Phys. Rev. B **36**, 1225 (1987); **38**, 1835 (1988). An extensive amount of literature related to the miniband formation can also be found in these papers.

[50]L. P. Kouwenhoven, F. W. J. Hekking, B. J. van Wees, C. J. P. M. Harmans, C. E. Timmering, and C. T. Foxon, Phys. Rev. Lett. **65**, 361 (1990).

[51]E. Castano, G. Kirczenow, and S. E. Ulloa, Phys. Rev. B **41**, 5055 (1990).

[52]L. Esaki and L. L. Chang, Phys. Rev. Lett. **33**, 495 (1974).

[53]R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

[54]C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).

[55]J. Masek, P. Lipavsky, and B. Kramer, J. Phys. Condens. Matter **1**, 6395 (1989).

[56]Recently these dips were associated with the quasibound states in the work by P. W. Bagwell, Phys. Rev. B **41**, 10 354 (1990). These states are neither resonance nor quasibound states, since the propagating part of the wave function is as important as the evanescent part.

[57]I. Kander, Y. Imry, and U. Sivan, Phys. Rev. B **41**, 12 941 (1990). This work focuses on the crossover from ballistic to diffusive transport and carries out ensemble averaging over different disorders, using Anderson's model.

[58]E. Tekman and S. Ciraci, Phys. Rev. B **42**, 9098 (1990); E. Tekman and S. Ciraci, in *Resonant Tunneling*, edited by E. E. Mendez and L. L. Chang (Plenum, New York, 1991).

[59]J. Faist, P. Guéret, and H. Rothuizen, Phys. Rev. B **42**, 3217 (1990).

[60]K. S. Ralls and R. A. Buhrman, Phys. Rev. Lett. **60**, 2434 (1988).

[61]C. G. Smith, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L893 (1988).

[62]L. Martin-Moreno and C. G. Smith, J. Phys. Condens. Matter **1**, 5241 (1989).

[63]P. L. McEuen, B. W. Alpenhaar, R. G. Wheeler, and R. N. Sacks, Surf. Sci. **229**, 312 (1990).

[64]S. Ciraci, E. Tekman, A. Baratoff, and I. P. Batra (unpublished). This is a state-of-the-art SCF pseudopotential calculation in the momentum representation carried out within the local-density approximation. In the text, *ab initio* or SCF calculations refer to this type of calculation. We used Hamann-

Schlüter-Chiang pseudopotentials and Ceperly-Alder exchange and correlation potential. For more details, see Refs. 38 and 39.

[65]M. Büttiker, Phys. Rev. B **41**, 7906 (1990).

[66]E. Tekman and S. Ciraci, in *Condensed Systems of Low Dimensionality*, edited by J. Beeby (Plenum, New York, 1991), p. 369.

[67]In Ref. 65, Büttiker used the slope of $G(w)$ between two plateaus, or the flat region of a given plateau obtained from the ratio $\partial T/\partial E$.

[68]S. Ciraci, in *Basic Concepts and Physical Applications of Scanning Tunneling Microscopy and Related Techniques*, edited by R. J. Behm, N. Garcia, and H. Rohrer (Kluwer, Dordrecht, 1990), p. 113.

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

[70]E. Tekman and S. Ciraci, Phys. Rev. B **40**, 10 286 (1989).

[71]The total vertical force acting on the tip may be attractive even if the outermost tip atom is in the repulsive force field forming an atomic-size mechanical contact. See Ref. 38 and J. P. Pethica and W. C. Oliver, Phys. Scr. **19A**, 61 (1987).

[72]N. D. Lang, Phys. Rev. B **36**, 8173 (1987).

[73]Here the value of the potentials, in particular $\phi_m(z)$ and V_{sp}, depends on the tip-sample distance d.

[74]N. Garcia, J. J. Saenz, and H. de Raedt, J. Phys. Condens. Matter **1**, 9931 (1989); H. de Raedt, N. Garcia, and J. J. Saenz, Phys. Rev. Lett. **63**, 2260 (1989); P. Serena, L. Escapa, J. J. Saenz, N. Garcia, and H. Rohrer, J. Microsc. **152**, 43 (1989).

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

[76]L. Orosz and E. Balazs, Surf. Sci. **177**, 144 (1986).

[77]E. Tekman, S. Ciraci, and A. Baratoff, Phys. Rev. B **42**, 9221 (1990).

[78]In this region, the whole tip is in the attractive force field. In the mechanical setup, the tip can flip to another equilibrium position nearer to mechanical contact.

[79]U. Dürig and D. Pohl, Phys. Rev. Lett. **65**, 349 (1990).

[80]J. Ferrer, A. M-Roder, and F. Flores, Phys. Rev. B **38**, 10 113 (1988).

[81]V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, and R. J. Wilson, Phys. Rev. Lett. **59**, 2789 (1987).

[82]J. Winterlin, J. Wiechers, H. Brune, T. Gritsh, H. Höfer, and R. J. Behm, Phys. Rev. Lett. **62**, 59 (1989).

[83]A. Baratoff, Physica B+C **127B**, 143 (1984).

[84]C. J. Chen, Phys. Rev. Lett. **65**, 448 (1990).

[85]E. Tekman and S. Ciraci, Phys. Rev. B **42**, 1860 (1990).