Theory for a quantum modulated transistor

Fernando Sols[a)]
Department of Physics and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

M. Macucci, U. Ravaioli, and Karl Hess
Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 24 March 1989; accepted for publication 20 June 1989)

We present a theoretical study of semiconductor T-structures which may exhibit transistor action based on quantum interference. The electron transmission through a semiconductor quantum wire can be controlled by an external gate voltage that modifies the penetration of the electron wavefunction in a lateral stub, affecting in this way its interference pattern. The structures are modeled as ideal two-dimensional electron waveguides and a tight-binding Green's function technique is used to compute the electron transmission and reflection coefficients. The calculations show that relatively small changes in the stub length can induce strong variations in the electron transmission across the structure. Operation in the fundamental transverse mode appears to be important for applications. We also show that a bound state of purely geometrical origin nucleates at the intersection between waveguide and stub. The performance of the device can be improved by inserting additional stubs of slightly different lengths. Taking into account the applicable scaling rules, we give estimates of the experimental parameters that optimize the transmission characteristics and speed of operation of the proposed transistor.

I. INTRODUCTION

Recent advances in nanometer-scale technology have made possible the fabrication of structures having sizes below the length scale over which the electron wave preserves its phase coherence. In general, these so-called mesoscopic systems[1] cannot be described by classical transport theory, since the phase coherence length L_ϕ, particularly at low temperature, is larger than the sample size and the wave nature of the electrons needs to be taken explicitly into account. Universal conductance fluctuations[2,3] and Aharonov–Bohm oscillations of the magnetoresistance[4,5] in solid-state systems constitute clear evidence of this new transport regime, where reproducible sample-specific behavior emerges. This is in contrast with the classical transport regime, where the loss of electron phase memory statistically averages the response of many microscopic configurations. Anderson localization,[6] weak localization,[7] and resonant tunneling phenomena[8] are other manifestations of regimes where the preservation of electron phase coherence at specific length scales plays a crucial role. The advent of the scaling theory of localization[9] has revived the interest in the 1957 paper by Landauer,[10] who studied for the first time the relation between the electron scattering properties and the conduction properties of a sample. Landauer's theories[10–12] have proven to be fundamental for theoretical research on mesoscopic systems. The conductance of multiprobe samples has been studied extensively by Büttiker.[13]

Most of the initial work on electron transport in small systems has dealt with metallic samples. More recently, increasing attention has been paid to semiconductor-insulator structures, which can now be fabricated and controlled on a very small scale.[5,14,15] In the absence of impurities, the sample boundaries become the only source of electron scattering. The short wavelength of the Fermi electrons in metals validates the use of a semiclassical picture where the electron is reflected by the internal walls of the sample and moves ballistically between collisions. On the other hand, the electron wavelength in semiconductor quantum wires may be comparable to the size of the sample, and in particular to the effective width of the wire. In this regime, one can view the electron motion as the propagation of a wave in a guide, and quantum wires may appropriately be called "electron waveguides." This notion has recently found clear experimental verification.[16]

As a result of the development of nanostructure technology, there has been great interest in the possibility of realizing electronic devices that are directly based on quantum mechanical effects, in the sense that the wave nature of the electron plays an essential role in their operating principle. Conductivity oscillations, due to quantum interference, were experimentally measured as a function of gate voltage in Si n-channel metal-oxide-semiconductor field-effect transistors (MOSFETs) where a tungsten grating was used as gate, causing the formation of a surface superlattice and a periodic potential for the quasi-two-dimensional electron gas at the interface.[17] A similar effect was observed for an AlGaAs/GaAs "washboard transistor" by varying the backgate voltage.[18] Strong negative differential conductivity for gate voltage near pinchoff was measured in high electron mobility transistors with the gate replaced by a grid structure which creates a lateral surface superlattice.[19] Fowler has proposed a transistor based on the electrostatic Aharonov–Bohm effect in a gated ring structure[20] and Bandyo-

[a)] Permanent address: Departamento de Física de la Materia Condensada, Universidad Autónoma de Madrid, E-28049 Madrid, Spain.

 0021-8979/89/203892-15$02.40

padhyay *et al.* have proposed the fabrication of a transistor based on the same effect in a layered semiconductor structure.[21] We have proposed a new transistor principle in which the electron transmission in a quantum wire is controlled by an external gate voltage that modifies the penetration of the electron wavefunction in a lateral stub, affecting in this way the interference pattern.[22] Independently, Fowler has suggested a similar T-structure transistor,[23] and Datta[24] has proposed an analogous device where the control voltage is applied in a stub inserted in the elbow of a quantum wire bent at a right angle. Very recently, experimental evidence of the presence of quantum interference effects in this type of structures has been presented by Miller *et al.*[25]

The structure of the quantum modulated transistor (QMT) that we have proposed is schematically shown in Fig. 1(a). For convenience, we use the terminology of the field-effect transistors (FET), and refer to the electron reservoirs as source (S) and drain (D), while the electrical termination of the stub is the gate (G). We want to stress, however, that the principle of operation of the proposed QMT differs substantially from the one of standard FETs. The nonlocality of electron waves is exploited directly in the QMT and the drain current can be modified by remote control, rather than by depletion of the conducting channel. In this regard, the quantization of the longitudinal conductance, recently observed in small point contacts,[26] can be viewed as the quantum regime of an FET.

The principle of operation of the QMT is similar to the one used for frequency measurements in electromagnetic waveguides, where a piston is mechanically moved to vary the resonant frequency of a lateral cavity. In a QMT a voltage is applied to the gate to vary the depletion length associated to the metal-semiconductor junction. In this way, the effective length of the stub is modified and the electron transmission probability can be varied between 0 and 1.[22] That the insertion of a lateral stub could influence the electron transmission through a quantum wire was hinted by Gefen *et al.*[27] and more strongly noted by Landauer.[12] In studies of topological disorder, Guinea and Vergés have analyzed the scattering induced by a finite lateral chain on an infinite tight-binding chain.[28]

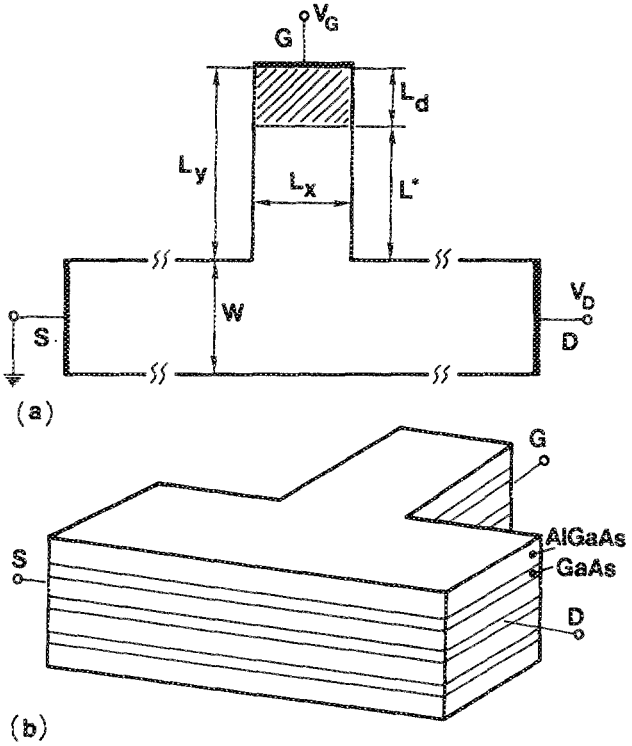

FIG. 1. (a) Schematic view of the quantum modulated transistor structure used in the calculations. (b) Suggested arrangement with several devices stacked on top of each other, obtained by etching a GaAs/AlGaAs superlattice.

As we will see, the current levels in this class of quantum devices are necessarily low, because optimum performance is only achieved in the single channel regime, where all Fermi electrons propagate in the fundamental transverse mode of the electron waveguide[22] and thus respond identically to changes in the effective stub length. Due to the high value of the intrinsic resistance of a single quantum channel ($h/e^2 \cong 25\ \mathrm{k}\Omega$) one may think of stacking in a superlattice many identical QMT structures, as schematically shown in Fig. 1(b). The gate contact may consist of a vertical Schottky barrier metallization, common to all layers, or a single Schottky contact on the top layer, if only a few layers are stacked. Of course, the stacking of many devices will only be useful if the layer-to-layer fluctuations in the response to changes in the gate voltage are sufficiently small, otherwise the overall quantum interference effect could be washed out. The current levels achievable in a single-layer device could nonetheless be sufficient for some applications. In particular, we propose in Sec. IV a logic family based on the quantum interference device presented, which would require low-power consumption. In our model calculations we have assumed that the effective length of the stub can be controlled by varying the depletion length associated to the Schottky gate, so that the electrons see a relatively sharp potential barrier. A more accurate approach which accounts more precisely for the geometrical complications of our particular system may be necessary in the future.

In this paper, we have focused on the problem of electron scattering by the boundaries of the structure shown in Fig. 1(a). We have idealized our system by assuming that the geometry of the structure is delimited by hard wall boundaries, consisting of straight lines and right angles, between which the electron potential is flat. For the numerical calculation of transmission and reflection coefficients the space has been discretized with a sufficiently dense nearest-neighbor square tight-binding (TB) lattice. We have developed a generalization of the recursive Green's function method,[29] by which a portion of wire of uniform width is included at each iteration. This greatly reduces the computation time for systems which can be simply decoupled in subsystems where the propagation is known exactly. Our method is described in Sec. II and complemented in the Appendix. Numerical results for stubs of various lengths and widths are reported in Sec. III. Structures with two or three stubs have also been studied and we show that they can substantially improve the performance. In addition, the existence of a bound state at the intersection of the quantum wires has been verified. This

confirms the findings recently reported by Schult *et al.*[30] In Sec. IV we explore some aspects of the physics related to the operation of the proposed QMT. In view of the results of Sec. III order-of-magnitude estimates are given for the experimental parameters that may lead to a clear observation of this quantum transistor effect. In particular, we estimate the temperature and the carrier density that optimize the transmission characteristics and minimize the switching time. We show that the optimum values of these experimental parameters are closely dependent on the dimensions of the device. Estimates of the electron transit time and of the cutoff frequency f_T suggest that a remarkable performance could be attained.

Some simplifying assumptions have been made to obtain the quantitative information which is presented in this paper. In particular, we have neglected possible complications due to the electron-electron interaction (self-consistency), the presence of impurities (which can be remote in a modulation-doped structure), departures from the idealized geometry, and inelastic scattering. We note that the low dimensionality of quantum wire structures may well imply that some of the physical effects mentioned above introduce qualitative departures from the expected behavior. An example may be the recent observation by Scott-Thomas *et al.*[15] of a periodic dependence of the conductance on gate voltage in a quantum wire obtained from a Si metal-oxide-semiconductor (MOS) structure.

Our main objective in this theoretical work has been to provide guidelines for the fabrication of devices that may exhibit the proposed quantum transistor effect. Even for relatively simple geometries, remarkable structure is found for the scattering amplitudes as a function of device dimensions and incident electron quantum numbers. Since one can only perform a limited exploration of the space of experimental parameters, it is important to emphasize that our results can be made more general by noting some scaling rules which we describe below.

The scattering by geometrical structures in electron waveguides is essentially a classical wave matching problem. Therefore, it depends only on the relative value of the incident electron wavelengths with respect to the dimensions of the structure itself. It follows that the scattering amplitudes are invariant under a global scaling of the lengths in the problem. We have performed the numerical calculations for a typical effective mass $M = 0.05\ m_0$ and a fixed width of the main wire $W = 100$ Å. This is a very small width for the present state of the art of nanofabrication technology, although probably within reach in a not too distant future. However, the numerical results presented here can be applied to larger structures, since the scattering amplitudes in our model are invariant under any of the following scaling transformations:

(i) $\lambda' = \eta\lambda$ and $L' = \eta L$, where λ stands for the wavelengths involved in the problem (those defining the transverse and longitudinal motion in the various modes of propagation) while L stands for the lengths defining the geometrical structure. Note that, for a fixed electron effective mass M, the scaling of the wavelengths requires also the scaling of all the energies in the problem, according to the rule $E' = E/\eta^2$.

(ii) $M' = \eta M$ and $E' = E/\eta$. This transformation leaves the wavelengths unchanged.

II. GREEN'S FUNCTION METHOD FOR SCATTERING BY BOUNDARIES

In this section we formulate the scattering problem related to geometrical modifications of quantum wires and describe the method of calculation of the scattering matrix (S-matrix) elements (transmission and reflection coefficients). We idealize our system by assuming that the geometry of the two-dimensional electron waveguide is defined by hard-wall confining potentials and that the potential is otherwise flat. For a device that is based on quantum interference, the electron phase coherence throughout the region defining the device is essential. Therefore, we assume the absence of inelastic scattering, and focus on the one-electron problem of scattering by boundaries. We neglect in this work the effects of electron-electron interaction except for the use of an effective Fermi liquid picture. This is probably a reasonable assumption, and is consistent with the experimental observation of nonlocal phenomena such as the Aharonov–Bohm effect in real metals and semiconductors.[4,5]

Far from the sidearms, the electron propagates in a perfect narrow wire where eigenstates can be factorized into a longitudinal part $\exp(ikx)$ (free-plane-wave motion) and a transverse part $\chi_n(y)$ (motion between two infinite potential walls). The scattering eigenstates are therefore of the form

$$\psi_{E,m}(x,y) = \begin{cases} k_m^{-1/2} e^{-ik_m x}\chi_m(y) \\ \qquad + \sum_n r_{nm} k_n^{-1/2} e^{-ik_n x}\chi_n(y) \\ \qquad \text{for } x \to -\infty, \\ \sum_n t_{nm}\, k_n^{-1/2} e^{ik_n x}\chi_n(y) \quad \text{for } x \to +\infty, \end{cases} \tag{1}$$

where the index n runs over all possible transverse modes (including the incident mode m) with energy $E_n < E$, so that the energy conservation requirement

$$E = E_n + (\hbar^2 k_n^2/2M) \tag{2}$$

can be satisfied. Here M denotes the effective mass of the electron. We will consider only the case where the wires have constant and identical width for $x \to \pm\infty$ and the corresponding sets of transverse modes are therefore identical (generalization to the case of asymmetric wires is straightforward). Thus, the wave function (1) describes an electron which comes from the left in the transverse mode m with total energy E and is transmitted (reflected) into mode n with probability amplitude t_{nm} (r_{nm}). Current conservation is expressed through the unitarity of the S matrix by the relation.

$$\sum_n (|t_{nm}|^2 + |r_{nm}|^2) = 1. \tag{3}$$

In order to calculate the scattering amplitudes t_{nm} and r_{nm}, we take advantage of the fact that plane wave motion can be

emulated by a nearest-neighbors tight-binding (TB) Hamiltonian in a lattice with periodicity a that is much smaller than all the other length scales of the problem (particularly the electron wavelength and device dimensions). We therefore fill the geometry of Fig. 1(a) by a sufficiently dense square lattice, as is schematically shown in Fig. 2(a), and we study the resulting scattering problem of a TB stripe with a sidearm. Consequently, we search for the scattering eigenstates of the Hamiltonian

$$H = \epsilon_a \sum_{\mathbf{R}} |\mathbf{R}\rangle\langle\mathbf{R}| + V \sum_{\mathbf{R},\,\delta_{\mathbf{R}}} |\mathbf{R}\rangle\langle\mathbf{R}+\delta_{\mathbf{R}}|, \tag{4}$$

where each site $\mathbf{R}$ of the lattice is connected to its nearest neighbors $\delta_{\mathbf{R}}$ within the structure depicted in Fig. 2(a). It is assumed that $\langle\mathbf{R}|\mathbf{R}'\rangle = \delta_{\mathbf{R},\,\mathbf{R}'}$. We set

$$V = -(\hbar^2/2Ma^2) \tag{5}$$

so that in the continuum limit we describe the motion within the wire boundaries of an otherwise free particle with positive effective mass M. Solving the TB Hamiltonian (4) is equivalent to solving the Schrödinger equation for a free particle within the hard-wall boundaries by means of a finite difference method. Therefore, it is natural to write

$$\epsilon_a = -4V \tag{6}$$

to obtain the same zero of energies in the continuous as well as in the discretized problem (the difference between the energy at the bottom of the band and the diagonal site energies is $2|V|d$, where d is the physical dimension).

The use of a square TB lattice permits fast numerical calculations in geometries defined by boundaries of straight lines and right-angle corners. The method can be used for arbitrary geometries, although a higher density of sites is needed to achieve a comparable degree of accuracy. More importantly, numerical calculation of transmisssion coefficients in arbitrary geometries may require brute force procedures such as the recursive Green's function method,[29] while, as shown below, the computational effort is greatly simplified in our problem because the wires can be partitioned into regions within which the propagation is known exactly.

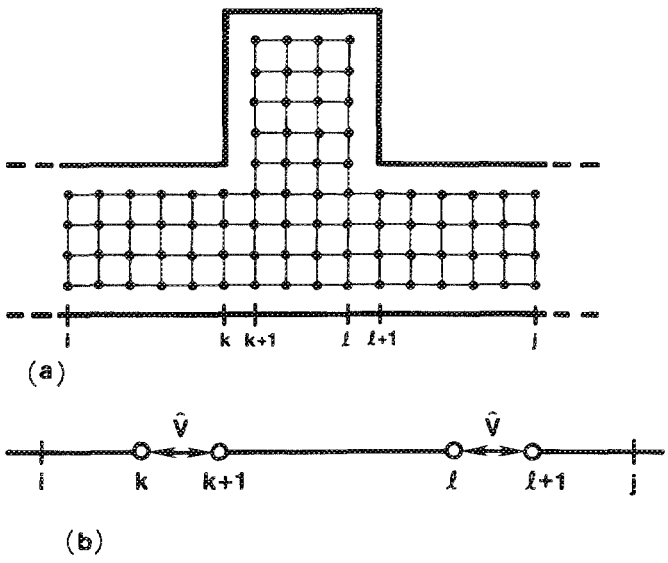

FIG. 2. (a) Tight-binding lattice used in the calculations. In this particular example the number of chains is $N = 4$. (b) Representation of the tight-binding lattice as a sequence of uniform stripes connected through discontinuities.

In order to calculate the scattering amplitude, we also make use of a Green's function method. Formally, the problem is similar to that of a TB chain where each "site" represents a transverse chain (slice) of the TB stripe, which contains several real lattice sites. To understand how an electron coming from the left in mode m is transmitted (reflected) into mode n, we have to calculate the matrix G_{ji} (G_{ii}), where i labels a slice on the left of the region under consideration and j labels one on the right (see Fig. 2). The matrix G_{ji} is obtained by taking the matrix elements of the Green's function operator G between states of the parts j and i. Since we wish to determine scattering amplitudes, we use a representation of transverse modes on each slice, i.e., we calculate the matrix elements $\langle n|G_{ji}|m\rangle$ and $\langle n|G_{ii}|m\rangle$ where

$$\langle n|G_{ji}|m\rangle = (G_{ji})_{nm} = \langle n, j|G|m, i\rangle, \tag{7}$$

and similarly for G_{ii}. In the notation of (7) G is the Green's operator [see (10)] and $|m, i\rangle$ is the ket for the transverse mode m located on the transverse chain i. This Green's function approach has been a standard procedure in the literature to calculate the transmission through disordered samples and a recursive method[29] has generally been used for the numerical calculation of the matrix elements. This method requires the inversion of an $N \times N$ matrix for every disordered slice, where N is the number of sites in the slice itself. In our problem, the geometry (in particular, the change of width in a portion of the electron waveguide) rather than the presence of impurities is the cause of scattering. As discussed, we can then take advantage of the relatively simple geometrical structure of Fig. 2(a) and decompose the lattice into left and right semi-infinite stripes and a central finite rectangular lattice formed by the sidearm plus the adjacent portion of the main wire. The eigenmodes and progagators within each of these three isolated regions can be calculated exactly and this fact simplifies the calculations considerably, since the total Hamiltonian H can be partitioned into two parts:

$$H = H_0 + \hat{V}. \tag{8}$$

H_0 describes the isolated regions and $\hat{V}$ corresponds to the connection by hopping between them. The propagators G^0 in the separate regions and the propagators G in the connected system are related by Dyson's equation at a given energy E:

$$G = G^0 + G^0\hat{V}G, \tag{9}$$

where

$$G(E)(E-H) \equiv 1, \quad G^0(E)(E-H_0) \equiv 1. \tag{10}$$

In the remainder of this section, we assume that the Green's function operator $G^0(E)$ is known. (A summary of expressions for the energy eigenvalues, wave functions, and propagators in finite and semi-infinite chains is given in the Appendix.)

The essence of the problem that we want to solve is

shown schematically in Fig. 2(b). We know the propagator G^0 in each of the three isolated regions, represented by straight lines, as well as the coupling between them. From this information we calculate the propagation in the total system and determine the submatrices G_{ji} and G_{li} defined in (7). By taking matrix elements between suitable pairs of transverse chains in the general Dyson's equation (9), one obtains the following set of coupled matrix equations (two per discontinuity):

$$\begin{aligned} G_{ki} &= G^0_{ki} + G^0_{kk}\, V_{k,k+1} G_{k+1,i}, \\ G_{k+1,i} &= G^0_{k+1,k+1}\, V_{k+1,k} G_{ki} \\ &\quad + G^0_{k+1,l}\, V_{l,l+1} G_{l+1,i}, \\ G_{li} &= G^0_{l,k+1}\, V_{k+1,k} G_{ki} + G^0_{ll}\, V_{l,l+1} G_{l+1,i}, \\ G_{l+1,i} &= G^0_{l+1,l+1}\, V_{l+1,l} G_{li}. \end{aligned} \tag{11}$$

The matrix $V_{l,l+1}$ is obtained by taking matrix elements of $\hat{V}$ between sites or modes of slices l and $l+1$:

$$\begin{aligned} (V_{l,l+1})_{nm} &= \langle n|V_{l,l+1}|m\rangle = \langle n,l|\hat{V}|m,l+1\rangle \\ &= V\sum_j \psi^*_n(j)\psi_m(j). \end{aligned} \tag{12}$$

In Eq. (12) $\psi_n(j)$ and $\psi_m(j)$ are the wave functions evaluated on nearest-neighbor sites on the slices l and $l+1$, respectively [see Eq. (A2) of the Appendix for an explicit expression]. The index j spans the possible pairs of nearest-neighbor sites which connect the two chains, while n and m indicate the transverse mode. From the orthonormality of the transverse modes one can see that, if the slices l and $l+1$ are part of a perfect stripe of uniform width, then $(V_{l,l+1})_{nm} = V\delta_{nm}$. This verifies that transverse modes do not mix in a wire of uniform width.

Once G_{li} and $G_{k+1,i}$ are known, the desired propagators G_{ji} and G_{ii} are readily calculated from

$$\begin{aligned} G_{ji} &= G^0_{j,l+1}\, V_{l+1,l}\, G_{li}, \\ G_{ii} &= G^0_{ii} + G^0_{ik}\, V_{k,k+1} G_{k+1,i}. \end{aligned} \tag{13}$$

The set of matrix operations (11) and (13) can be solved analytically and the resulting expressions can be evaluated numerically. Only two matrices have to be inverted, one for each discontinuity. However, we are also interested in calculating the transmission across multiple stubs. As the number of discontinuities in the stripe increases, the explicit solution of the resulting set of equations equivalent to (11) becomes too cumbersome. Therefore, it is useful to employ a method in which the effect of additional discontinuities can be included by recursion. This constitutes a generalization of the commonly used recursive Green's function method[29] that is particularly suitable to our present needs.

In the standard recursive Green's function method, the effect of one additional transverse chain is included at each step. Such a method is necessary in the study of disordered samples, for instance, where the distribution of impurities is random. If perfect semi-infinite leads are attached to both sides of the sample, the first iteration will consist in connecting the right-most disordered slice to the right semi-infinite lead. In an intermediate iteration, an additional slice is added to the outcome of previous iterations. In the final iteration the semi-infinite perfect lead on the left is added to the composite system of the whole disordered sample plus the perfect lead on the right.

In our generalized Green's function method, the exact knowledge of the propagation within a finite stripe is exploited and a portion of wire of uniform width is integrated at every step. This technique can be reduced to the standard recursive Green's function method by simply reducing the finite perfect stripe to a single slice.

For computer calculations, it is convenient to include the discontinuities in such a way that the result from a particular discontinuity is used as the input for the calculation of the following discontinuity. A generic recursion process of this kind is shown schematically in Fig. 3, where j indicates a slice on the right of the structure, within the open perfect stripe on the right; q and $q+1$ indicate neighboring transverse chains at the boundary of portions with different width in the TB stripe; and finally, p is either the index of the last slice of the region of uniform width A or, in the final iteration, the index of a slice on the left of the structure ($p = i$). The propagation in the region B extending from $q+1$ to $+\infty$ is assumed to be known, either because it is known eactly (as in the first iteration) or because discontinuities to the right of $q+1$ have already been included. A is a region within which the propagators are exactly known and which we want to connect to the region B through the hopping interaction $\hat{V}$. The goal in a recursion step is to obtain the propagators G^{A+B} in the composite region $A+B$ from the knowledge of the propagation in the separate regions A and B. In particular, we want to obtain G^{A+B}_{jp} and G^{A+B}_{pp} in terms of $G^{B}_{j,q+1}$ and $G^{B}_{q+1,q+1}$. In the following iteration, the region $A+B$ becomes the new region B and the transverse section p becomes the new $q+1$.

In order to derive the equations for the recursive process, we make use of the general Dyson equation (8). The propagators G^A and G^B in the separate regions A and B now play the role of G^0 in (8). Since we want $\hat{V}$ to represent the connection by hopping between regions A and B, we label here the full propagators in the connected blocks as G^{A+B} [equivalent to G in (8)]. By taking matrix elements of (8) between appropriate pairs of slices (note that a "matrix element" between two slices is a matrix), one obtains the following set of coupled matrix equations:

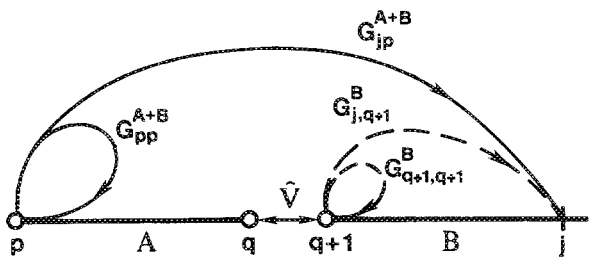

FIG. 3. Representation of the recursion steps for the calculation of the recursion Green's functions.

$$G^{A+B}_{jp} = G^{B}_{j,q+1}\, V_{q+1,q}\, G^{A+B}_{qp},$$
$$G^{A+B}_{pp} = G^{A}_{pp} + G^{A}_{pq}\, V_{q,q+1}\, G^{A+B}_{q+1,p},$$
$$G^{A+B}_{qp} = G^{A}_{qp} + G^{A}_{qq}\, V_{q,q+1}\, G^{A+B}_{q+1,p},$$
$$G^{A+B}_{q+1,p} = G^{B}_{q+1,q+1}\, V_{q+1,q}\, G^{A+B}_{qp}. \quad (14)$$

which can be recast into

$$G^{A+B}_{jp} = G^{B}_{jq+1}\, V_{q+1,q}(1 - G^{A}_{qq}\, S^{B}_{q})^{-1} G^{A}_{qp},$$
$$G^{A+B}_{pp} = G^{A}_{pp} + G^{A}_{pq}\, S^{B}_{q}(1 - G^{A}_{qq}\, S^{B}_{q})^{-1} G^{A}_{qp}, \quad (15)$$

where

$$S^{B}_{q} = V_{q,q+1}\, G^{B}_{q+1,q+1}\, V_{q+1,q} \quad (16)$$

is the self-energy matrix for the transverse chain q, due to the presence of B. In the first iteration, B is a perfect semi-infinite stripe for which G^B is known exactly, and A is a perfect finite stripe. In an intermediate iteration, the propagation in B has been computed numerically while G^A is known exactly. In the final iteration, A is a perfect semi-infinite stripe that extends from q to $-\infty$, in which the propagation is also known exactly.

The recursion procedure described above is a powerful method for calculating the transmission across an arbitrary set of width discontinuities in an electron waveguide. Other types of discontinuities like elbows, bends, or open sidearms will require other calculational methods which can nevertheless be included in a generalized recursive scheme like the one we have described.

The relation between the Green's functions and the S matrix for scattering in leads has been given by Fisher and Lee[30] and has been generalized to an arbitrary scattering structure by Stone and Szafer.[31] It is straightforward to adapt their results to the case of scattering in a TB stripe. We obtain for the transmission and reflection coefficients:

$$t_{nm} = -i2V(\sin\theta_n \sin\theta_m)^{1/2} e^{i(\theta_m l - \theta_m j)} \langle n|G_{jl}|m\rangle, \quad (17a)$$

$$r_{nm} = -(\sin\theta_n/\sin\theta_m)^{1/2} e^{i2(\theta_n + \theta_m)l} \times (i2V \sin\theta_m \langle n|G_{ll}|m\rangle + \delta_{nm}). \quad (17b)$$

where l and j are indices of slices on the left and on the right of the device, respectively. Also, $\theta_n \equiv k_n\, a$, where k_n is the wave vector available for longitudinal motion, as indicated in (2). We note that $\sin\theta_n$ is related to the velocity of propagation in the transverse mode n:

$$v_n = -(2Va/\hbar)\sin\theta_n. \quad (18)$$

It is important to remark that in (17) the retarded Green's function must be used, which means that, in the definitions (10), $(E-H)$ must be understood as $(E-H+i0^+)$.

We conclude this section by noting that, through a careful dimensional analysis of the equations involved in the calculation of the amplitudes t_{nm} and r_{nm} (including those in the Appendix), it is possible to show that these have a well-defined continuum limit ($a \to 0$), as required, and that the scaling rules given in Sec. I are rigorously correct.

III. NUMERICAL RESULTS

In this section we present numerical results for the transmission probability of an electron approaching one or several sidearms in a quantum wire. These results have been obtained following the Green's function method outlined in the previous section. We have studied the case of the single stub in detail, extending our previous work.[22] In particular, we explore here the dependence of the transmission on the electron energy and the device dimensions. We present also results for the case of structures with two and three stubs. We show that the scattering properties of multiple stubs are promising from an application point of view.

Throughout this paper, we limit ourselves to the single-channel (also referred to as mode or subband) regime, in which the electron can only propagate along the main wire in the fundamental transverse mode ($E_1 < E < E_2$). The reason for this restriction is that this is the regime that provides the greatest hope of achieving a sharp and controllable modulation of the electron transmission probability.[22] At higher energies, propagation in various modes is possible and, due to the presence of the stub, an incident electron in mode m may be transmitted into mode n with probability $T_{nm}(E)$. The result of this mode mixing is that the structure of the transmission modulation tends to decrease. Even when it is feasible, operation at high energies with a single incident mode does not solve the problem entirely, since the sum over various possible emerging modes tends by itself to average out the oscillations in the total transmission probability.[22] We further show that the presence of a stub gives rise to a bound state in the quantum wire, and the corresponding binding energy is calculated. The nature of this bound state is very similar to that of the bound state recently found by Schult *et al.*[32] at the intersection of two infinite quantum wires. This provides another example of a quantum bound state in a classically unbound system.

A. Single stub

We show in Fig. 4 the transmission probability T for an electron impinging on a single-stub structure as a function of the electron energy E and the effective length of the stub L^* for various stub widths L_x. In all the calculations the width of the main wire is nominally $W = 100$ Å. In order to apply our results to a structure with a different W, the energies and all the other lengths in the problem must be changed according to the scaling rules given in Sec. I. As expected, the transmission probability approaches unity for vanishing stub length and tends to zero as the wavelength of the incident electron becomes infinite. In Fig. 4(b) ($L_x = W = 100$ Å), it can be seen that, for a given energy, the transmission exhibits a periodic pattern as a function of the effective length L^*, the period being one half of the wavelength which the electron assumes along the stub (in the case of $L_x = W$ it coincides with the incident wavelength). At high energies, close to threshold for propagation in the first excited transverse mode ($E \lesssim E_2$), the presence of this slowly decaying mode slightly distorts the periodicity of the signal. The structure of $T(L^*)$ for a fixed energy $E = 0.08$ eV is shown in Fig. 7(a) [see also Fig. 2(b) of Ref. 22]. A relatively small change in the effective length L^* can induce dramatic changes in the transmission. This remarkable fact constitutes the basis for our conjecture that transistor action by quantum interference should be possible.

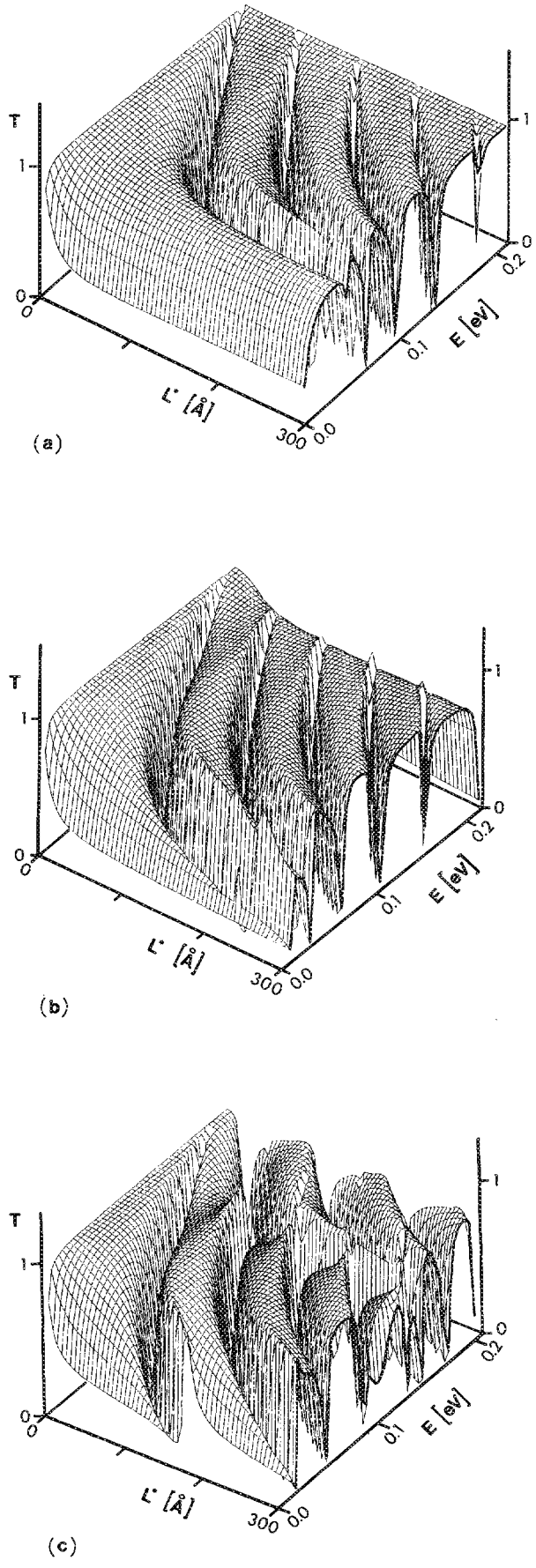

FIG. 4. Transmission probability T as a function of the effective length of the stub L^* and of the electron energy E for a single-stub structure with $W = 100$ Å and (a) $L_x = 80$ Å, (b) $L_x = 100$ Å, (c) $L_x = 120$ Å.

In Fig. 4(c) we show the transmission probability $T(E, L^*)$ for $L_x = 120$ Å $> W$ (wide stub). At high energies, $T(L^*)$ exhibits a quasiperiodic pattern due to the fact that the electron can propagate along the stub with two different wavelengths, corresponding to the two available transverse modes (an increase of the stub width lowers the energy of the transverse modes, making more of them available for propagation at a given energy). It can generally be said that the more channels in the stub resonate with the incident energy, the more bizarre (and effectively random) the behavior of $T(L^*)$ becomes. For narrow stubs [see Fig. 4(a), where $L_x = 80$ Å], electrons with a relatively low energy cannot penetrate the stub due to the higher zero-point energy marking the threshold for propagation in the stub. The result is that the low-energy maximum in $T(E)$ becomes broad and independent of the effective length L^*, as Fig. 4(a) clearly shows, since low-energy electrons penetrate very little in the sidearm and are thus insensitive to changes in the effective stub length. This geometry is therefore incapable of transistor action at low electron energies.

The interplay between stub length and stub width can be seen in Fig. 5, where a contour plot of $T(L_x, L^*)$ is shown for a typical energy $E = 0.08$ eV. It is clear that for small values of the width L_x, $T(L^*)$ is approximately constant due to the relatively high maximum energy required to propagate along the stub. For a wider stub, propagation in the fundamental ($n = 1$) transverse mode of the stub becomes possible and $T(L^*)$ reveals a periodic pattern. As the stub width increases further, more transverse modes become

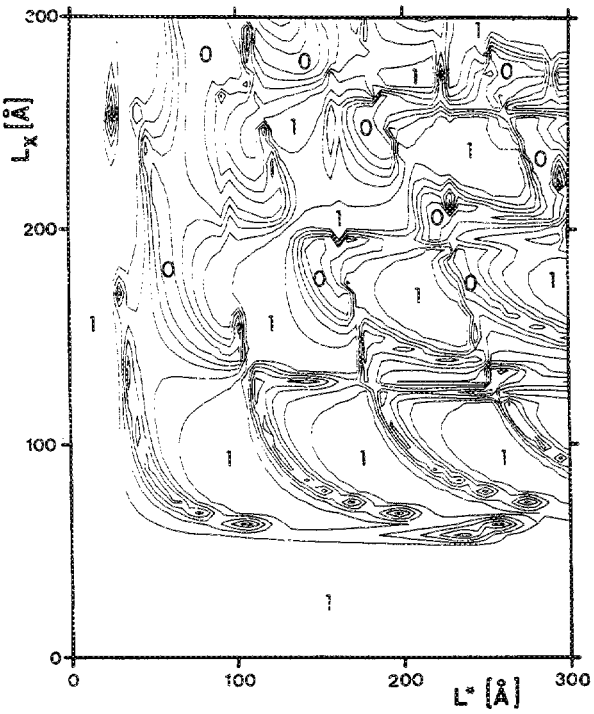

FIG. 5. Contour plot of the transmission probability $T(L_x, L^*)$ for a single-stub structure with $W = 100$ Å. The electron energy considered is $E = 0.08$ eV. The contour lines between probability 0 and 1 correspond to the values $T = 0.1$, 0.3, 0.5, 0.7, and 0.9.

available for propagation in the stub. At a given energy, each of these modes has a different value of the kinetic energy which is available for "longitudinal" motion. The result is that as L_x increases, more periodicities become involved in $T(L^*)$. The thresholds for the onset of propagation in a new mode are evenly spaced as a function of L_x and correspond to integer multiples of $\lambda/2$, where λ is the wavelength of the incident electron.

We conclude from the analysis above that operation in the single-channel regime is desirable and that the width of the sidearm should be about the same as that of the main wire.

B. Multiple stubs

For purpose of modulation by quantum interference, it is desirable that the curve for $T(L^*)$ display broad regions of high and low transmission at the Fermi energy ($E = E_F$). These regions should be separated by short intervals in which $T(L^*)$ varies rapidly. In addition, this structure should be as insensitive to energy variations as possible.

We focus now on the first requirement. In Fig. 6(a) we show a contour plot of $T(E, L^*)$ for a stub of width $L_x = W = 100$ Å. One can quickly recognize that the most promising range of energies lies roughly in the middle of the single-channel regime. However, the transmission valleys are still very narrow. This shortcoming may be eliminated or at least reduced by introducing one or more additional stubs, which is standard procedure in the design of microwave filters. One may expect this addition to broaden the transmission valleys on the grounds that it brings the structure closer to a periodic array of stubs where transmission would be forbidden in the energy gap regions. We show in Fig. 6(b) a contour plot of $T(E, L^*)$ for two stubs of width $L = 100$ Å separated by $S = 95$ Å. The broadening of the transmission valleys can be clearly seen. However, the presence of extra resonances due to multiple internal reflection in the region between the stubs creates new undesirable structure that reduces the range of energies suitable for modulation (which in our model stays between 0.08 and 0.11 eV, although the low-energy region around 0.04 eV could also be considered).

The broadening of the transmission valleys can be further increased by slightly mismatching the two stubs, a concept once again borrowed from microwave design. A high-reflection probability by any of the stubs suffices to make the total transmission small (in the absence of resonances). In principle, such a mismatch can be created by making the stubs different in length or width. We will consider in the following only stubs with same width and different lengths, since this appears to be slightly more advantageous for the situation under study. We show in Figs. 6(c) and 6(d) a contour plot of $T(E, L^*)$ for structures with two stubs of width $L_x = 100$ Å separated by $S = 95$ Å. The stub length difference is $\Delta = 5$ and 10 Å, respectively (L^* is the effective length of the shortest stub). As compared with the case of two identical stubs [Fig. 6(b)], the difference in length gives rise to a more effective broadening, particularly for $\Delta = 10$ Å. A feature which emerges from the comparison of Figs. 6(c) and 6(d) is that the actual location of the energy region of optimum modulation is not very sensitive to the choice of Δ. This property is particularly useful for real implementation, since an accurate control of the sidearm lengths may prove difficult. It is not desirable to increase the length difference between stubs much beyond 10 Å since the transmission in the middle of the valley would increase significantly. This trend can be observed in the high-energy region where the single-stub transmission valleys are extremely narrow and relatively high peaks develop for $\Delta = 10$ Å, in contrast to the case of $\Delta = 5$ Å, where a much weaker splitting of the valleys is observed. That these spurious peaks appear more pronounced at high energies is not surprising, since the shorter wavelengths increase the relative effect of a given difference in stub lenghts. The excessive structure in both the single- and double-stub case makes the high-energy region unsuitable for practical modulation purposes.

The problem of the presence of undesired resonances also appears in the case of two dissimilar stubs. If the separation between the stubs is further increased, more resonances develop in a given energy range. This can be appreciated by comparing Figs. 6(d) and 6(e) for double-stub systems with a separation of $S = 95$ and 195 Å, respectively. A larger separation between sidearms causes an unwanted reduction of the middle-energy region that seems ideal for operation. This trend may be expected to continue at larger separations. We conclude that the stubs should not be separated by distances much larger than their width for quantum modulation to be effective.

We have finally studied the case of three stubs in series separated by 95 Å, with a relative length difference of 10 Å between consecutive ones. There are actually three possible arrangements for this choice of the three stub lengths, depending on which stub is positioned in the center. The scattering properties in the single-channel regime are invariant under the exchange of the two lateral stubs due to time reversal symmetry. We have considered the case in which the stub of intermediate length lies in the middle and do not expect important qualitative changes for the other possible arrangements. The results shown in Fig. 6(f) display a triple splitting of the narrow valleys in the high-energy region, as expected. The transmission valleys in the middle-energy region acquire a considerable width, which is comparable to the width of the high transmission plateaus. This makes the structure very attractive for possible switching applications. It also adds flexibility for the design of more complicated logic functions, which will be discussed later.

The broadening of the transmission valleys by insertion of additional stubs can be more clearly appreciated in Fig. 7, where, for a fixed energy $E = 0.08$ eV, $T(L^*)$ is plotted for the cases of (a) one single stub, (b) two identical stubs, (c) two dissimilar stubs, and (d) three dissimilar stubs. The separation between stubs in 95 Å and the relative length difference is $\Delta = 10$ Å. These results suggest that the modulation of the transmission by changes in the effective length of the stub is significant enough to be observable, particularly if sets of two or more dissimilar stubs are employed.

We point out that for the results shown in Figs. 4, 5, and 8 a tight-binding lattice with lattice spacing $a = 5$ Å has been used, which corresponds to a representation of the main wire ($W = 100$ Å) by $N = 19$ parallel coupled chains. For the

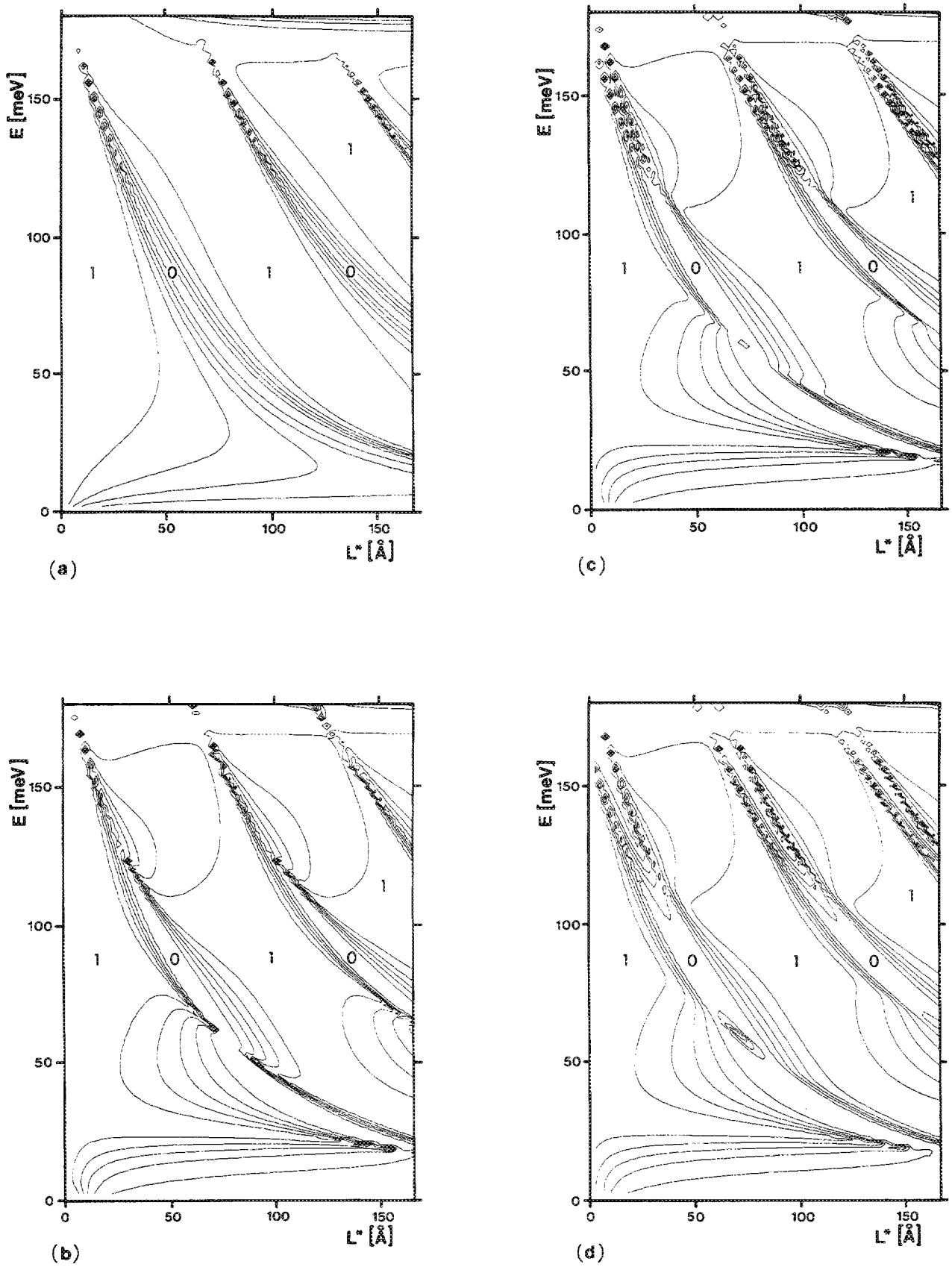

FIG. 6. Contour plots of the transmission probability $T(E, L^*)$ for several stub structures. The contour lines between probability 0 and 1 correspond to the values $T = 0.1, 0.3, 0.5, 0.7$, and 0.9. In all cases, the width of the main wire is $W = 100$ Å, and the stub width is $L_x = 100$ Å. The structures considered are (a) single stub, (b) two identical stubs with separation $S = 95$ Å, (c) two stubs with separation $S = 95$ Å and length difference $\Delta = 5$ Å, (d) two stubs with separation $S = 95$ Å and length difference $\Delta = 10$ Å, (e) two stubs with separation $S = 195$ Å and length difference $\Delta = 10$ Å, (f) three stubs with separation $S = 95$ Å and length difference $\Delta = 10$ Å between consecutive ones. In all cases with dissimilar stubs, L^* refers to the length of the shortest one.

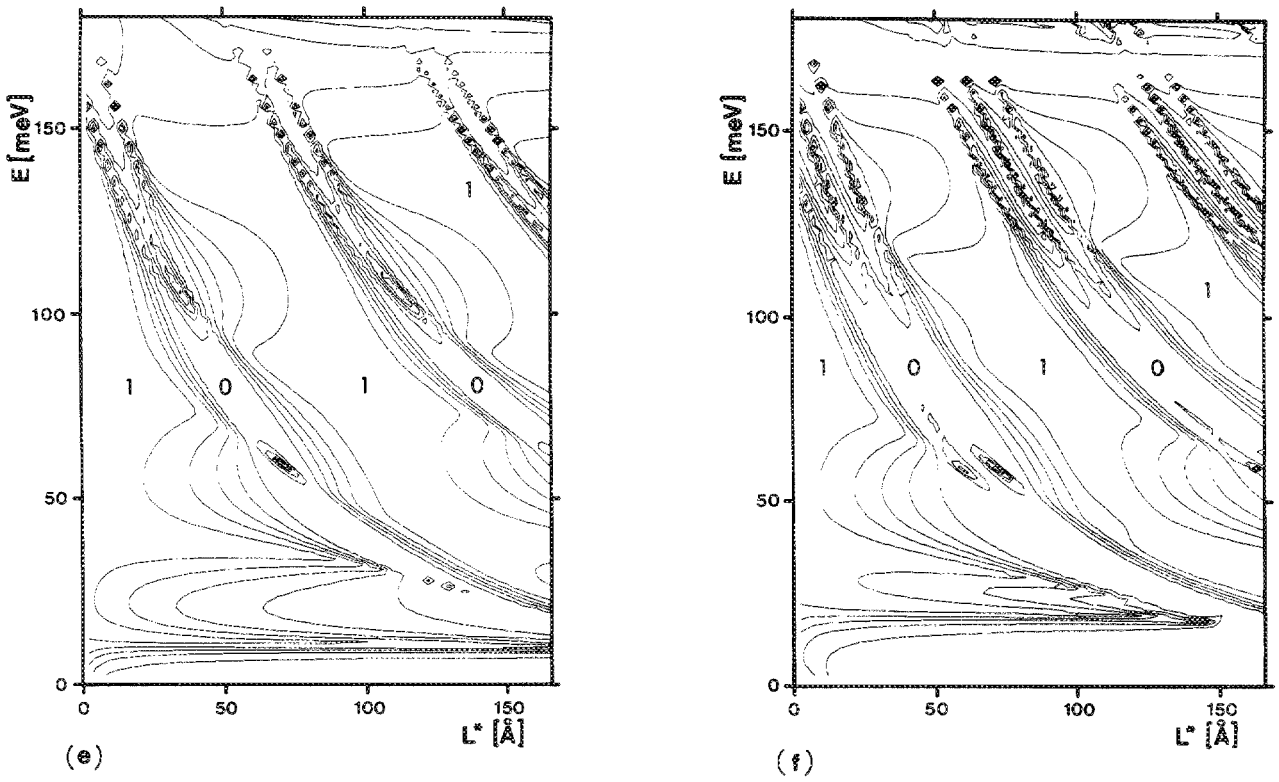

FIG. 6 (*continued*).

calculations in Figs. 6 and 7, the value $a = 2.5$ Å has been employed ($N = 39$ for the main wire) to increase the resolution. No significant discrepancy was found between results from these two different choices of TB lattice, as should be expected from a proper description of the continuum limit ($\lambda \ll a$). Note that calculations of the scattering at higher energies (shorter wavelengths) may require a finer tight-binding grid.

C. Bound states

It is possible to extract from the Green's function approach information on the existence of bound states localized around the intersection of the sidearm with the main wire. One has simply to look for poles in the Green's function at energies below the minimum threshold for propagation ($E < E_1$). The position of the pole is the energy level of the bound state. In Fig. 8 we show the resulting bound state energy E_{bs} as a function of the length of a single stub of width $L_x = W = 100$ Å. The results are easy to understand: as the stub length vanishes the effective strength of the scattering decreases and, consequently, the binding energy tends to zero. Although for small values of L^* our method looses accuracy (due to the specifics of our discretization method), the results of Fig. 8 indicate that $E_{bs} \propto L^*$. As the stub length increases, the position of the energy level approaches a constant value [here, $E_{bs}(L^* \to \infty) \simeq -14.2$ meV, which is in excellent agreement with the result obtained by Schult *et al.*[32] for a bound state in an open T structure]. When the length of the stub becomes sufficiently large, the wave function reaches its maximum penetration into the stub compatible with the geometry of the intersection. For stub lengths much larger than this intrinsic decay length (which in our problem is of the order of 50 Å), the form of the bound state wave function becomes insensitivie to changes in the effective stub length, hence the saturation of the energy level. An important basic fact, which is clear from Fig. 8, is that a bound state exists for all values of the stub length. This bound state originates from the local lowering of the zero-point energy that the presence of the stub induces in the electron waveguide. The result that a bound state nucleates at all finite values of the effective scattering strength is probably a universal property of geometrical modifications that induce an effective broadening of the quantum wire. An interesting question in this context is whether or not the electron-electron interaction can screen out the bound state. These issues deserve further study.

IV. PHYSICS OF THE QUANTUM MODULATED TRANSISTOR

A. Optimum conditions for modulation by quantum interference

We provide in this section estimates of the physical parameters that determine the conditions which seem most promising for an observation of the proposed transistor effect. Ideally, one would like to meet the following requirements: (i) Large inelastic mean free path, (ii) ballistic electron motion, (iii) operation on the lowest transverse subband, and (iv) a sharp Fermi surface ($T \ll T_F$).

Condition (i) is necessary for the preservation of phase coherence throughout the structure. We will not elaborate on this topic since it is common to all mesoscopic systems and is phenomenologically well investigated.[12] We note that,

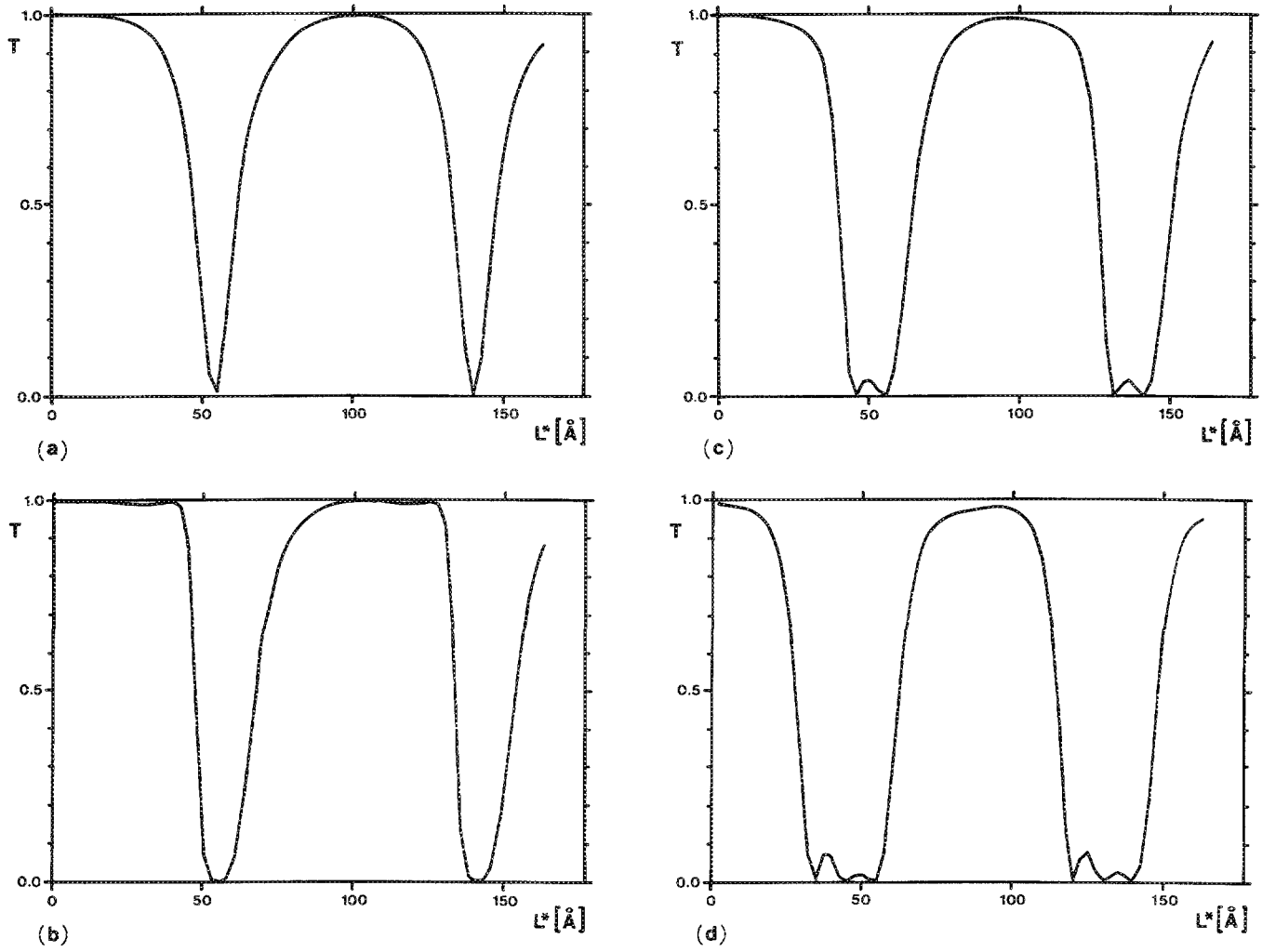

FIG. 7. Plot of the transmission probability $T(L^*)$ at a given electron energy $E = 0.08$ eV for several stub structures. In all cases, the width of the main wire is $W = 100$ Å, and the stub width is $L_x = 100$ Å. The structures considered are (a) single stub, (b) two identical stubs with separation $S = 95$ Å, (c) two stubs with separation $S = 95$ Å and length difference $\Delta = 10$ Å, (d) three stubs with separation $S = 95$ Å and length difference $\Delta = 10$ Å between consecutive ones. In all cases with dissimilar stubs, L^* refers to the length of the shortest one.

for length scales of thousands of angstroms, cooling down to liquid-helium temperature and even far below will be required, while operation at room temperature would require structures of 100-Å length scale. A detailed treatment of the role of dissipation in the crossover from coherent to incoherent electron motion in submicron devices is still not available.

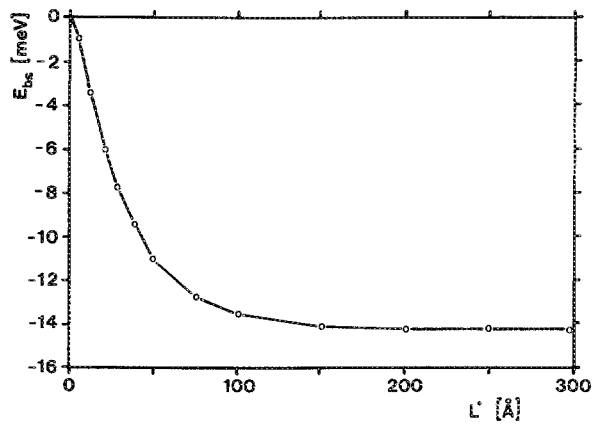

FIG. 8. Calculated energy of the bound state, at the intersection between the wires in a single-stub structure with $W = L_x = 100$ Å, as a function of the effective stub length L^*.

Requirement (ii) is met if the elastic mean free path due to scattering by impurities is larger than the scale of the device. Under such conditions, the main source of scattering is the geometry of the structure, as is the case for electromagnetic waveguides. Of course, the ideal requirement of ballistic electron motion is not strictly necessary for a single device to exhibit quantum interference effects as long as inelastic scattering is avoided, but it is desirable in order to have strong oscillations in the electron transmission probability. The problem of reproducibility is common to all mesoscopic systems. It is currently believed that the change in the position of a single impurity may change the whole interference pattern. This is certainly true for metallic samples and may be less important in semiconductors structures where the electron wavelength is much larger than the lattice spacing. However, reproducibility may still be a problem if distances involved in the sample-to-sample fluctuations are comparable with the electron wavelength. If low impurity concentration is desirable, modulation-doped structures are in princi-

ple the most promising for semiconductor applications, since it is well known that in the conduction channel the density of impurities can be controlled to a level better than 10^{15} cm^{-3}, corresponding to a mean distance between impurities of 1000 Å. In the Introduction we propose a structure consisting of many modulation-doped layers to achieve larger output current levels. If layer-to-layer fluctuations are sufficiently small, this arrangement can also yield some degree of self-averaging and therefore higher reproducibility, and this could be seen as a way of achieving a more uniform device response. We want to strongly emphasize at this point that to realize this kind of structures one has to carefully tradeoff between the ideal or desired properties and the actual technological limitations.

The necessity of operating in the fundamental mode regime [condition (iii)] is a clear conclusion from the numerical analysis of Sec. III. This was also pointed out in Ref. 22. In particular, operation in the regime of optimum performance in double- and triple-stub devices requires that the Fermi energy lies roughly in the middle of the single-channel energy range. This can be stated as the requirement that $\alpha \simeq 1.5$, where

$$E_F = \alpha E_0, \tag{19}$$

and $E_0 = \hbar^2\pi^2/2MW^2$ is the zero-point energy of the lowest transverse mode, and sets the energy scale of the problem [note that in this convention the thresholds for the different subbands lie at $E_n = (n^2 - 1)E_0$, i.e., the zero of energies is taken at the bottom of the lowest subband]. In Table I, the value of E_0 is given for various combinations of wire width and electron effective mass. In the single-channel regime, the carrier density that corresponds to a particular value of the parameter α is independent of the effective mass M and depends only on the width of the main wire through the simple relation:

$$n_s = \alpha/W^2. \tag{20}$$

In Table I the carrier density is given for $\alpha = 1.5$ and for several widths of the wire. We conclude that, for operation in the fundamental channel, the use of narrow wires ($W \lesssim 500$ Å) is desirable.

The operating temperature of the device should be much lower than the Fermi temperature of the carriers [condition (iv)], so that the electrons responsible for the electric current occupy a relatively narrow energy range, thus responding approximately in the same way to changes in the effective stub depth. From the results of Sec. III, this requirement can be formulated numerically as $T \sim T_0$, where $T_0 \equiv 0.05\ T_F$. In Table I, some values of T_0 are given for several combinations of wire width and carrier effective mass (assuming $\alpha = 1.5$). It is clear that the requirement of a sharp Fermi surface is approximately as restrictive as condition (i) of a large inelastic mean free path, which is quintessential to the mesoscopic regime.

B. General considerations concerning device applications

Although quantum interference effects open new horizons for electronic conduction and eventual device applications, the technological problems that need to be overcome to bring some of these effects to fruition are enormous. One of the major objections to practical usefulness arises from the large value of the quantum of resistance $h/e^2 \simeq 25\ \mathrm{k}\Omega$ which seems to preclude reasonably short switching times because of practical resistance-capacitance limitations.[12] We address this and other questions below, and show that some problems are less restrictive than they appear at a first glance.

The calculation of switching times is in general a difficult problem, since account has to be taken of dynamical screening in a complicated geometry. It is nevertheless possible to make simple estimates of some fundamental limits to the speed of the transistor. A more elaborate theoretical study might require the identification of other time scales involved in the operation of the device. We analyze below the transit time and cutoff frequency of a single-stub structure. The transit time τ_t can be defined as the time necessary for an electron to travel from the main wire to the end of the lateral sidearm and return to the wire. This is the minimum time required for the current to adapt itself to the new value of the effective stub length. Short transit times can be obtained with short stubs and high carrier densities, since $\tau_t \simeq 2L^*/v_F$, where $v_F = \hbar\pi n_s W/M$ is the carrier Fermi velocity. The order of magnitude of the transit time can be made more explicit by writing

$$\tau_t(\mathrm{ps}) = 5.5\times 10^{-5}[(M/m_0)/\alpha]L^*(\text{Å})W(\text{Å}), \tag{21}$$

where the notation (20) has been introduced. The estimated values of the transit time in picoseconds are given in Table I

TABLE I. Estimated values of some physical quantities related to a QMT for various values of the wire width W and the electron effective mass M. See the text for the definitions and the assumptions introduced.

W (Å)	M/m_0	E_0 (meV)	n_s (cm^{-2}) $\times 10^{10}$	T_0 (K)	v_F (cm s^{-1}) $\times 10^7$	τ_t (ps)	f_T (THz)
100	0.01	375.6	150.0	327.0	54.6 [a]	0.0183	5.77
100	0.05	75.1	150.0	65.0	10.9 [a]	0.0915	1.15
500	0.01	15.02	6.0	13.0	10.9 [a]	0.457	0.225
500	0.05	3.0	6.0	2.6	2.2	2.29	0.045
1000	0.01	3.75	1.5	3.2	5.46	18.3	0.057
1000	0.05	0.75	1.5	0.65	1.09	91.5	0.011

[a] Note that in III-V semiconductors the electron velocity is limited to about 10^8 cm s^{-1} by Bragg refraction.

for several combinations of wire width and electron effective mass, assuming $\alpha = 1.5$ and for explicitness $L^* = 5$ W. The corresponding Fermi velocities are also given in Table I. The values of v_F involved make the effective-mass approximation acceptable, at least for order of magnitude estimates.

The transconductance of the device is defined as

$$g_m = \frac{\partial I_D}{\partial V_G}. \tag{22}$$

Assuming that the linear response and the modified Landauer formula for the conductance[12,30,31,33] hold, the drain current I_D (single-channel regime) is

$$I_D = (2e^2/h)TV_D, \tag{23}$$

where the factor of 2 accounts for spin degeneracy. Introducing (23) into (22) leads to

$$g_m = \frac{2e^2}{h} V_D \frac{\partial T}{\partial V_G} = \frac{2e^2}{h} V_D \frac{\partial T}{\partial L^*} \frac{\partial L^*}{\partial V_G}. \tag{24}$$

We have assumed in (24) that the transmission probability depends on the gate voltage V_G through L^* only. The quantity $\partial T/\partial L^*$ is directly available from the results discussed in Sec. III.

In the small-signal regime, the gate capacitance can be defined as[34]

$$C_G = \left|\frac{\partial Q}{\partial V_G}\right| = en_s L_x \left|\frac{\partial L_d}{\partial V_G}\right|, \tag{25}$$

where L_d is the depletion length [see Fig. 1(a)]. In a multiple-stub structure L_x should be understood as the sum of the stub widths. We note at this point that $L_d = L_y - L^*$, where L_y is a constant equal to the geometrical length of the stub, and therefore one can write

$$\frac{\partial L^*}{\partial V_G} = -\frac{\partial L_d}{\partial V_G}. \tag{26}$$

The transconductance then becomes

$$g_m = \frac{2e^2}{h} V_D \frac{\partial T}{\partial L^*} \frac{C_G}{en_s L_x} = I_{D0} \frac{C_G}{en_s L_x} \frac{\partial T}{\partial L^*}, \tag{27}$$

where $I_{D0} = 2e^2 V_D/h$ is the drain current for $T = 1$.

The cutoff frequency f_T can be defined as the frequency at which the current gain is unity, i.e., the incremental currents through the input capacitance C_G and through the drain, are equal in amplitude. The gate and drain incremental currents are simply

$$i_G = C_G \frac{\partial V_G}{\partial t} = C_G i\omega v_G, \tag{28}$$

$$i_D = g_m v_G, \tag{29}$$

where t indicates time, ω is the circular frequency, and v_G is the incremental gate voltage. When $\omega = \omega_T = 2\pi f_T$, from (28) and (29) one obtains the cutoff frequency as

$$f_T = |g_m|/2\pi C_G, \tag{30}$$

which is formally identical to the expression used to FETs. From (25), (26), and (27), Eq. (30) becomes

$$f_T = I_{D0} \frac{1}{2\pi en_s L_x} \left|\frac{\partial T}{\partial L^*}\right|, \tag{31}$$

which can be made more explicit by writing

$$f_T(\text{GHz}) = 76.9 \frac{V_D(\text{mV})}{\alpha\beta} W \left|\frac{\partial T}{\partial L^*}\right|, \tag{32}$$

where $\beta \equiv L_x/W$. Some numerical estimates of f_T are shown in Table I for $\alpha = 1.5$ and $\beta = 1$. The calculations of Sec. III suggest that, in the energy region of optimum performance, $W\partial T/\partial L^* \cong 4$. The drain voltage has been taken as $V_D = 0.05E_F/e$ which is a rather conservative assumption that should guarantee an acceptable monochromaticity of the electrons responsible for the current. One can see from Table I that the estimated values of f_T are attractively high. We also remark that, unlike in FETs, the cutoff frequency f_T is not simply related to the transit time τ_t defined above due to the different principle of operation.

We give now an estimate for the quantity $\partial L^*/\partial V_G$ which appears in the expressions for g_m and C_G. Assume that the stub is terminated by a Schottky barrier. Then the depletion length L_D is determined by the impurities which are located at position y_i in the stub. If we exclude at this point difficult questions of screening and use the depletion approximation we have[34]

$$V_G + V_{\text{bi}} = -\frac{e}{\epsilon_r \epsilon_0 A} \sum_{i=1}^{N} y\,\delta(y - y_i) = -\frac{e}{\epsilon_r \epsilon_0 A} \sum_{i=1}^{N} y_i, \tag{33}$$

where ϵ_r is the relative dielectric constant and the reference $y = 0$ has been set at the Schottky barrier interface. A is the cross-sectional area of the stub and V_{bi} is the equilibrium built-in voltage. By using (26), the finite average changes of L^* can be estimated. Of course, a large impurity density in the conducting channel is not desirable. If a modulation-doped structure is used, the impurities will reside for instance in an AlGaAs layer, with the mobile carriers in a neighboring low-doped GaAs layer. In the continuum approximation one has

$$\sum_{i=1}^{N} y_i = N_D^+ A \frac{L_d^2}{2}, \tag{34}$$

where N_D^+ is the density of positively charged donors in the doped layer. From Eq. (26), (33), and (34) one obtains

$$\frac{\partial L^*}{\partial V_G} = -\frac{1}{2}\left(\frac{2\epsilon_r \epsilon_0}{N_D^+ e}\right)^{1/2} \frac{1}{\sqrt{V_G + V_{\text{bi}}}}. \tag{35}$$

For ϵ_r and a typical doping of $N_D^+ = 10^{18}$ cm^{-3}, we obtain

$$\frac{\partial L^*}{\partial V_G}(\text{Å/V}) \cong 190 \frac{1}{\sqrt{V_G + V_{\text{bi}}}}, \tag{36}$$

where V_G and V_{bi} are given in volts. From the above estimate, we may conclude that in principle a practical control of the effective stub length is possible, since a modulation of L^* by 20–30 Å requires then a gate voltage swing of a few tenths of a volt. However, applications in integrated circuits will require comparable gate and drain voltages.[35] Unfortunately, as mentioned above, the drain voltages must be much smaller than the Fermi energy (~ 10 mV) and gate voltages of the same order are desirable. This can be achieved by using a gate on top of and parallel to the layers of the stub. The top AlGaAs layer [Fig. 1(b)] can then be chosen of an appropriate thickness which permits depletion almost precisely down to the undoped GaAs. In such an arrangement,

any infinitesimal increase of gate voltage will deplete the GaAs layer and therefore change the effective stub length. We would like to emphasize that such operation is possible only because of the use of the quantum interference effect which allows to put the gate at a remote location. If one uses the standard field effect and fringing fields to pinch off the channel, one either needs large gate voltages or a gate in close proximity, which leads to large capacitances.

Another important area to be investigated involves the possibility of creating logic works based on the single devices discussed. Of primary importance is to assess how quantum interference devices can be interconnected without destroying the interference effects. This may require the design of appropriate filtering and matching passive elements to minimize unwanted reflections. Due to the low current levels, which constitute an intrinsic limitation for this type of devices, it is also important to determine configurations which do not require large currents for logic operation. The best example of low-power devices in conventional integrated circuits is probably given by the complementary metal-oxide-semiconductor (CMOS) family. The basic CMOS inverter is realized by connecting two complementary p- and n-channel MOS transistors, so that one of the devices is turned off for either logic input (0 or 1). No drain current flows, since the devices are connected in series, except for a small charging current during the switching transients. It is conceivable to realize a similar elementary inverter using the proposed QMT. The three-stub structure is probably the most attractive one, since as already discussed, it is possible to obtain plateaus and valleys of transmission of comparable width [Fig. 7(d)]. The idea is to use two devices with slightly different stub length, so that for the same voltage applied to the stub termination one device is on and the other one is off. As long as the connections are realized in a way that does not modify the essential quantum interference effects, and inverter logic element is obtained with a configuration similar to the CMOS one. This could be envisioned as the basic building block of an entire logic family.

V. CONCLUSIONS

We have presented numerical results for the transport properties of waveguide like T-shaped semiconductor structures and have illustrated various aspects of a new transistor principle which is based on quantum interference phenomena. Our results have been presented for particular dimensions. However, we have also given scaling laws which make the results more general and applicable to mesoscopic systems of any size, as determined mainly by the inelastic mean free path of electrons.

From the point of view of physics, these structures offer interesting and unexpected behavior, such as the complicated form of the transmission and reflection coefficients and the existence of a bound state.

From the point of view of applications, mesoscopic systems offer some very interesting properties as well. We have shown that the frequency limitations of such structures used as transistors are determined to a good approximation by rather universal equations which involve only basic material parameters and derivatives of the transmission coefficient T which have been computed in this paper. The cutoff frequency may of the order of terahertz (inverse picoseconds) and thus attractively high in spite of the large value of the quantum of resistance h/e^2. Besides the potential for high-speed operation, the appeal of this class of quantum devices is in the very small dimensions and the possibility of operation at very-low-power levels, which could allow an extremely high level of integration.

We conclude therefore that the inherently new mechanisms of transport in mesoscopic systems may open new horizons in semiconductor electronics.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and the Army Research Office. F. Sols wishes to acknowledge the support from the Joint Program between the Fulbright Commission and Spain's Ministerio de Educación y Ciencia. U. Ravaioli wishes to acknowledge the support of the National Science Foundation. Computations were performed on the CRAY X-MP/48 of the National Center for Supercomputing Applications of the University of Illinois.

APPENDIX: WAVE FUNCTIONS AND PROPAGATORS IN TIGHT-BINDING CHAINS

The purpose of this Appendix is to provide the mathematical expressions for the energies, wave functions, and Green's functions that are needed for the calculations described in Sec. II. In particular, we give formulas for the energies and wave functions of the normal modes, which are needed for the transverse component of the scattering states, as well as expressions for the propagators in finite and semi-infinite lattices. We only need to study decoupled longitudinal chains since transverse modes do not mix in stripes of uniform width. Of course, for a given total energy E, the energy available for longitudinal motion will not be the same in the various chains (each one corresponding to a transverse subband) and will depend on the transverse kinetic energy E_n. For asymptotic propagation, $E_n < E$ is required. However, in the presence of discontinuities, the local propagation through evanescent modes with $E_n > E$ has to be taken into account since they do play an important role in the wave matching that determines the final scattering amplitudes.

The Hamiltonian for a one-dimensional tight-binding (TB) chain can be written [replacing $\mathbf{R}$ in Eq. (4) by l]:

$$H = \epsilon_a \sum_l |l\rangle\langle l| + V \sum_l (|l\rangle\langle l+1| + |l+1\rangle\langle l|), \quad \text{(A1)}$$

where the hopping energy V is given in (5) and here $\epsilon_a = -2$ V, since (A1) corresponds to a one-dimensional problem.

For a finite chain with N sites, the index l in (A1) runs from 1 to N. There are N eigenmodes with energies

$$E_n = 2V\{\cos[n\pi/(N+1)] - 1\}, \quad n = 1,\ldots, N \quad \text{(A2)}$$

and wave functions

$$\psi_n(l) = A_n \sin[n\pi l/(N+1)], \quad \text{(A2a)}$$

where

$$A_n = \left[\frac{N}{2} + \frac{1}{2}\,\mathrm{Re}\left(\frac{1 - e^{i2\pi nN/(N+1)}}{1 - e^{-i2\pi n/(N+1)}}\right)\right]^{-1/2}, \quad \text{(A2b)}$$

so that $\langle \psi_n | \psi_m \rangle = \delta_{nm}$. The continuum limit is obtained by taking $y = la$ and $L = (N+1)a$ and letting $a \to 0$ and $N \to \infty$ with L remaining constant.

For the calculation involved in Eqs. (11)–(16), we need the propagators between the end points of finite TB chains and between two points in a semi-infinite chain. For a finite chain we have

$$G_{11} = \frac{\sin N\theta}{V \sin[(N+1)\theta]}, \quad G_{1N} = \frac{\sin\theta}{V \sin[(N+1)\theta]} \quad \text{(A3)}$$

where θ is related to the propagation energy through

$$E = 2V(\cos\theta - 1). \quad \text{(A4)}$$

Equations (A3) can be shown by recursion, starting from the simple cases $N = 1$ and 2 and using Dyson's equation to include additional sites.

For a semi-infinite chain extending from $l = 1$ to $+\infty$, the following propagators are needed:

$$G_{11} = \frac{e^{i\theta}}{V}, \quad G_{1l} = \frac{e^{il\theta}}{V}, \quad G_{ll} = \frac{e^{il\theta}}{V}\frac{\sin l\theta}{\sin\theta}. \quad \text{(A5)}$$

For a semi-infinite chain extending to the left from $l = -1$ to $-\infty$, Eq. (A5) applies if we keep $\theta > 0$ and replace l by its absolute value. For the derivation of (A5) it must be known that the retarded propagator between sites l and j in an infinite chain is[36]

$$G_{jl} = \frac{ie^{i|l-j|\theta}}{2V\sin\theta}. \quad \text{(A6)}$$

In the energy convention adopted in (A4), the case of evanescent propagation corresponds to $E < 0$ or $E > -4$ V, the second case being unphysical in the continuum limit. We have evanescent propagation in a piece of chain whenever we consider an energy E that is not allowed in the bulk of the corresponding infinite perfect chain. The Green's functions for the evanescent modes are easily obtained from Eqs. (A3), (A5), and (A6) by substituting $i\alpha$ for θ, where $\alpha \equiv \cosh^{-1}(E/2V + 1)$.

To summarize, we have given expressions for the eigenmodes in a finite TB chain and for the propagators in finite and semi-infinite chains. The wave functions will be used for the "transverse" chains and define the transverse modes, while the propagators will be applied to the "longitudinal" chains which result from mapping a TB stripe into a set of decoupled parallel chains each corresponding to a particular transverse mode. Such modes do not mix as long as the width of the stripe is uniform. For a given total energy, the energy that has to be considered in the propagation within these longitudinal chains is that which is left for longitudinal motion in the particular subband.

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

[2]P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).

[3]P. A. Lee, A. D. Stone, and H. Fukuyama, Phys. Rev. B **35**, 1039 (1987).

[4]R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, Phys. Rev. Lett. **54**, 2696 (1985); S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren, M. Vaziri, M. Miller, and R. Reifenberger, Phys. Rev. Lett. **55**, 2344 (1985).

[5]S. B. Kaplan and A. Hartstein, IBM J. Res. Develop. **32**, 347 (1988).

[6]P. W. Anderson, Phys. Rev. **109**, 1492 (1958); Philos. Mag. B **52**, 505 (1985); C. M. Soukoulis, E. N. Economou, G. S. Grest, and M. H. Cohen, Phys. Rev. Lett. **62**, 575 (1989).

[7]B. L. Al'tshuler, A. G. Aronov, D. E. Khmel'nitskii, and A. I. Larkin, in *Quantum Theory of Solids*, edited by I. M. Lifshitz (Mir, Moscow, 1982).

[8]T. C. L. G. Soliner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983); F. Capasso, in *Semiconductors and Semimetals* (Academic, New York, 1987), Vol. 24, Chap. 6.

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

[10]R. Landauer, IBM J. Res. Develop. **1**, 223 (1957).

[11]R. Landauer, Philos. Mag. **21**, 863 (1970).

[12]R. Landauer, IBM J. Res. Develop. **32**, 306 (1988) (see also other references in the same issue).

[13]M. Büttiker, Phys. Rev. B **35**, 4123 (1987); **38**, 12724 (1988); IBM J. Res. Develop. **32**, 306 (1988).

[14]M. L. Roukes, A. Scherer, H. G. Craighead, S. J. Allen, Jr., R. M. Ruthen, E. D. Beebe, and J. P. Harbison, Surf. Sci. **196**, 79 (1988); A. Scherer, M. L. Roukes, H. G. Craighead, R. M. Ruthen, E. D. Beebe, and J. P. Harbison, Appl. Phys. Lett. **51**, 2133 (1987).

[15]J. H. F. Scott-Thomas, Stuart B. Field, M. A. Kastner, H. I. Smith, and D. A. Antoniadis, Phys. Rev. Lett. **62**, 583 (1989).

[16]G. Timp, H. U. Baranger, P. deVegvar, J. E. Cunningham, R. E. Howard, R. Behringer, and P. M. Mankiewich, Phys. Rev. Lett. **60**, 2081 (1988).

[17]A. C. Warren, D. A. Antoniadis, H. I. Smith, and J. Melngailis, IEEE Electron Dev. Lett. **EDL-6**, 294 (1985).

[18]Y. Tokura and K. Tsubaki, Appl. Phys. Lett. **51**, 1807 (1987); K. Tsubaki and Y. Tokura, Appl. Phys. Lett. **53**, 859 (1988).

[19]G. Bernstein and D. K. Ferry, Superlattices and Microstructures **2**, 373 (1986).

[20]A. B. Fowler, U. S. patent No. 4550330 (1985).

[21]S. Bandyopadhyay, S. Datta, and M. R. Melloch, Superlattices and Microstructures **2**, 539 (1986).

[22]F. Sols, M. Macucci, U. Ravaioli, and K. Hess, Appl. Phys. Lett. **54**, 350 (1989).

[23]A. B. Fowler, Workshop on Quantum Interference, Atlanta, Sept. 1988.

[24]S. Datta, Supperlattices and Microstructures **6**, 83 (1989).

[25]D. C. Miller, R. K. Lake, S. Datta, M. S. Lundstrom and R. Reifenberger, to appear in the Proceedings of the International Symposium on Nanostructure Physics and Fabrication, College Station, Texas, March 13–15, 1989.

[26]B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988); D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ajmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L209 (1988).

[27]Y. Gefen, Y. Imry, and M. Ya. Azbel, Phys. Rev. Lett. **52**, 129 (1984).

[28]F. Guinea and J. A. Vergés, Phys. Rev. B **35**, 979 (1987).

[29]D. J. Thouless and S. Kirkpatrick, J. Phys. C **14**, 235 (1981).

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

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

[32]R. L. Schult, D. G. Ravenhall, and H. W. Wyld, Phys. Rev. B **39**, 5476 (1989).

[33]M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).

[34]K. Hess, *Advanced Theory of Semiconductor Devices* (Prentice-Hall, Englewood Cliffs, NJ, 1988).

[35]We thank R. Landauer and H. van Houten for bringing this point to our attention.

[36]E. N. Economou, in *Green's Functions in Quantum Physics*, Vol. 7 of *Springer Series in Solid State Sciences*, edited by M. Cardona, P. Fulde, and H. J. Queisser (Springer, Berlin, 1983).