AN ACCURATE, EFFICIENT ALGORITHM FOR CALCULATION OF QUANTUM TRANSPORT IN EXTENDED STRUCTURES

T. J. Godin[a] and Roger Haydock[b]
[a]*MSRC, Pacific Northwest Laboratory,* P.O. Box 999, Richland WA 99352*
[b]*Department of Physics and Materials Science Institute, University of Oregon, Eugene OR 97403*

Abstract

In device structures with dimensions comparable to carrier inelastic scattering lengths, the quantum nature of carriers will cause interference effects that cannot be modeled by conventional techniques. The basic equations that govern these "quantum" circuit elements present significant numerical challenges. We describe the block recursion method, an accurate, efficient method for solving the quantum circuit problem. We demonstrate this method by modeling dirty inversion layers.

I. INTRODUCTION

Numerical "pre-testing" of proposed integrated circuit elements can vastly reduce development time and expense [1]. "Conventional," semiclassical device modeling methods fail, however, for structures with dimensions comparable to the distance between carrier inelastic (*e.g.*, electron-phonon) collisions. On this submicron scale, without inelastic collisions to destroy electron wave coherence, the quantum wavelike nature of a carrier will produce interference effects as the carrier scatters from device walls, individual dopants, and other deviations from crystallinity [2]. Modeling such a "quantum circuit element" presents significant numerical challenges.

In this paper we describe the block recursion method, an efficient, accurate algorithm for solving the quantum circuit problem. We have presented thorough derivations and analyses of this method, and some applications, in specialized papers, [3,4,5] and block recursion software is publicly available [5]. Our purpose here is to present the salient points in a manner accessible to a broader audience. Readers interested in more detail should consult the references or authors.

II. THE BASIC EQUATIONS OF QUANTUM CIRCUIT THEORY

Conventional device modeling starts with a basic set of coupled equations (diffusion-drift, continuity, Poisson, etc.) that are discretized (*e.g.*, by finite difference or finite element schemes) and solved. Quantum effects are only considered implicitly (through effective masses, carrier/band energetics, etc.) [1]. At each node, Ohm's Law relates the local current to the local electric field using *conductivity*, an intensive, local property.

In quantum circuits, however, resistance cannot be locally defined. When electron waves scatter and interfere between two barriers (device walls, dopant atoms, etc.) total resistance depends not only on properties of each scatterer, but on electronic DeBroglie wavelengths and distances between scatterers. It is more useful, then, to calculate a device's total *conductance*, an extensive property. Without inelastic scattering, the conductance Σ of a device with carriers of energy E is given by the Landauer formula [6],

$$\Sigma(E) = (2e^2/\hbar)\, T(E)\,. \tag{1}$$

e is the carrier charge; $\hbar$ is Planck's constant divided by 2π. The transmittance T is the probability that an incident electron will be pass through a device. When transport occurs only near the Fermi Energy, we take this as E. T is calculated by solving the time independent Schrödinger Equation,

$$\{(-\hbar^2/2m^*)\nabla^2 + V(x,y,z)\}\,\psi = E\,\psi, \tag{2}$$

for an unknown quantum state ψ. $(-\hbar^2/2m^*)\nabla^2 + V(x,y,z)$ is H, the Hamiltonian operator. V, the electric potential within the device, arises from fields caused by device walls, dopants, gates, etc. m^* is the effective carrier mass. In a crystalline region, ψ is a traveling wave. If a device is connected to crystalline "leads," the transmittance is found by applying boundary conditions of incident and reflected waves in one lead and a transmitted wave in the other. The transmittance is the ratio of transmitted to incident wave amplitudes.

*Operated for the U.S. Dept. of Energy by Battelle Memorial Institute under contract DE-AC06-76RLO 1830.

transmitted wave in the other. The transmittance is the ratio of transmitted to incident wave amplitudes.

V must be specified on an atomic level, since electrons scatter from individual atoms. Moreover, Eq. (2) must be solved over an inelastic scattering volume (≈ 0.01 μ wide at room temperature [7]), even if this region extends beyond a single device (*i.e.*, an "extended structure"). Electrons can scatter and interfere at interfaces with leads and with other devices. Any model must preserve carrier phase relationships between these boundaries. (Larger volumes require a hybrid quantum/conventional model; this is beyond the scope of this paper.)

Eq. (2) may be discretized by a tight-binding method [8]. This is like a simple finite element model, with one node per atom. For example, consider a simple quantum circuit (Fig. 1): an (n-2) atom device connected to one-dimensional leads. First, we convert Eq. (2) to a matrix eigenvalue equation by defining a set of n×1 vectors $\{\phi_m: m= 1,2,...n\}$. The mth element of ϕ_m equals one, and all other elements equal zero. ϕ_m represents an electron localized at node m. In this basis, H is a matrix whose diagonal elements $H_{mm} = V_m$ equal the potential V at each node m. We set the off-diagonal elements of H equal to energy v for nearest neighbor atoms, and zero for all other atom pairs. v can be fit to the device band width. This model is now a discrete network (shown by the solid lines in Fig. 1) through which carriers move. Using this matrix form of H, Eq. (2) becomes a matrix eigenvalue equation whose solution is an n×1 vector Ψ. The n components of vector Ψ equal the function values of state ψ at each of the n nodes. The boundary condition[the values of ψ at nodes 0, 1, n-1 and n (Fig. 1)] must be satisfied; the solution algorithm presented in Sec. III does this automatically.)

This model is appropriate for simple potentials, and carriers confined to a single band. More complicated potentials can be modeled by varying the off-diagonal elements of H, and multiple bands by adding more basis vectors per node.

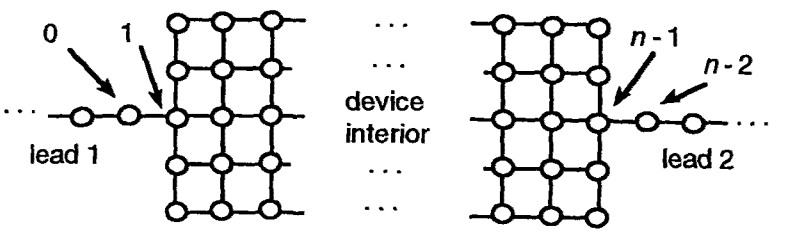

Figure 1. Schematic representation of a model of a simple two-lead quantum device with n atoms. Nodes (circles) are located at atoms. Lines connect nearest neighbors. Nodes (atoms) at the lead-device interfaces are numbered.

III. THE BLOCK RECURSION ALGORITHM

Eq. (2) presents significant numerical challenges. First, solids have on the order of 100 atoms per cubic nanometer. Models of even small device models may require millions of nodes. Moreover, this problem requires calculation of extremely sensitive coupling between leads. Many previous algorithms have worked only for small atomic clusters. For realistically sized devices, unstable matrix inversions or divergent recursions made these earlier algorithms impractical (for a review, see [5]).

We developed the Block Recursion Method for this problem. To illustrate this approach, consider the device in Fig. 1. Suppose we define an n×2 matrix, u_1, whose columns are the vectors representing ϕ_1 and ϕ_{n-2} respectively. Similarly, let the columns of u_0 be ϕ_0 and ϕ_{n-1}. We then use these two matrices to generate a set of n×2 matrices $u_2, u_3, \ldots u_{n/2}$. The (m + 1)th member of this set is generated from the mth and (m -1)th elements using the following recurrence:

$$u_{m+1} B_{m+1} = H u_m - u_m A_m - u_{m-1} B_m^{\dagger}$$
$$\text{where } A_{m-1} = u_m^{\dagger} H u_m \quad \text{and} \quad B_m^{\dagger} = u_{m-1}^{\dagger} H u_m \,. \tag{3}$$

The A_m and B_m are 2×2 matrices. $u_{m+1}B_{m+1}$ is factored by requiring the columns of u_{m+1} to be mutually orthonormal. When this procedure is finished, the columns of the u_m form a new basis for the matrix eigenvalue problem of Eq. (2). The new basis has two important features. First, in this basis H is block tridiagonal: if H is divided into 2×2 blocks, the A_m form the diagonal blocks, the B_m and $B_m^{\dagger}$ form, respectively, the super- and sub-diagonal blocks, and all elements of all other blocks of H equal zero. Second, although most new basis elements no longer represent electrons localized on a single atom, the basis states in u_0 and u_1 still

represent electrons localized on atoms at the lead-device interface. This means that the boundary conditions can still be imposed on these components of the solution Ψ.

This transformation, in effect, replaces the network of Fig. 1 with a different network (Fig. 2), without altering the description of the leads. We call this new network an "effective quantum circuit," analogous to the effective circuit theory used to calculate the total impedance of a network of conventional resistors. This result is general—any device geometry can be transformed to a model like Fig. 2. More importantly, the transmittance of such an effective network is given by the following:

$$T(E) = 4 \sin^2 k \, | [e^{ik} I - v\, G(E)]^{-1}{}_{1,2} |^2, \quad \text{where} \tag{4a}$$

$$G(E) = [EI - A_1 - B_2^{\dagger}(EI - A_2 - \ldots)^{-1} B_2]^{-1} . \tag{4b}$$

k is the crystal momentum given by $E = \hbar^2 k^2/(2m^*)$. I is the 2×2 identity matrix. The subscript "1,2" signifies the off-diagonal element of the symmetric 2×2 matrix calculated in Eq. (4a).

This algorithm is descended from the Lanczos method of matrix diagonalization. The high stability and efficiency of this class of methods is well understood. [9] Briefly, a large rounding error is introduced in the transformation described by Eq. (3). This causes a rapid loss of orthogonality of the new basis vectors as the recursion progresses. These errors do not affect the calculated transmittance, however, because the new elements of H contained in the A_m and B_m are still accurate, to within a simple rounding error, for the new, nonorthogonal basis vectors.

In practice, the orthogonality loss prevents the recurrence relation Eq. (3) from terminating after n new vectors have been produced. That is, new vectors will continue to be generated, eventually leading to an overspecified basis. This presents no real problem; the matrix continued fraction (4b), which is highly stable, need only be continued until the transmittance converges. While this may take more than n levels; in practice it usually takes far less.

This method has impressive numerical performance. First, the number of operations required is proportional to n (most conventional schemes require n^3). Second, using Eqs. (3) and (4) provides transmittances correct to machine precision (that is, on random rounding architectures, the number of mantissa digits lost to rounding error will be approximately $\log_{16} \sqrt{n}$). Fig. 3 demonstrates the accuracy of this method by showing the calculated transmittance as a function of the number of levels used in the continued fraction (Eq. (4b)), for a model system whose transmittance is exactly known (dashed line). The calculation converges sharply to the correct value, once enough basis vectors are included.

We have tested the accuracy of this method for elastic scattering volumes containing up to 10^5 atoms (nodes). This exhausted our available computer memory. However, algorithmic advances and the availability of massively parallel architectures make much larger calculations possible. Multiple lead geometries can be accommodated by tridiagonalizing in larger blocks.

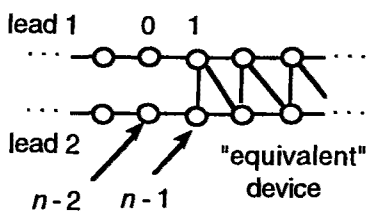

Figure 2. The block recursion method transforms the device model into an equivalent network without altering lead nodes.

IV. AN EXAMPLE: DIRTY INVERSION LAYERS

We have modeled inversion layers as clusters of atoms arranged in a two dimensional square lattice [3,4,5]. Fig. 4a shows conductance as a function of Fermi Energy (controlled by gate voltage) for a "clean" (pure crystalline) device. Fig. 4b represents a "dirty" layer (having a high concentration of defects, dopants, etc.) Both layers are 100 atoms square ($n = 10^5$), roughly equivalent to an elastic scattering area of a metallic layer at room temperature [7]. Here, we will only highlight some important results of this study.

In the clean sample, carriers with energies within the band conduct well because the electronic states of a crystal obey Bloch's theorem [7], and are thus extended through the layer and have high mobility. There is noticeable resistance, however; this comes from electron reflection and interference at the device walls and device-lead interfaces. At the band edge, conductance falls off sharply, since carriers above this energy must tunnel through the band gap.

Conduction in the disordered layer is weaker. Quantum states in a disordered system do not obey Bloch's theorem, and instead tend to destructively interfere, and hence decay with distance. This lowers carrier

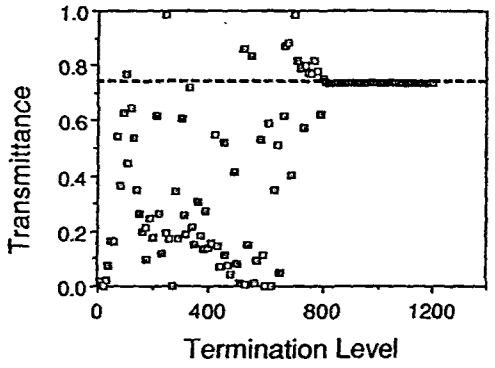

Figure 3. Convergence of the transmittance with dimension of the block-tridiagonal matrix.

mobility compared to a crystal [10]. Fig. 4b shows a "kink" in conductance, similar to the crystal band edge, at a "critical" energy E_C. However, Ec is not at the band edge ($E/v = 6.0$ for the disordered layer). We have shown that there are large numbers of allowed states for carriers with energies above E_C; however, their spatial decay length is much shorter than states with energies at or below E_C. That is, carrier mobilities change drastically when the Fermi Energy passes through E_C.

This result is surprising. Previously, it was widely believed that in disordered layers the strength of carrier state decay (and hence mobility) varied smoothly with carrier energy. Instead, we find a singularity at E_C. This mobility transition explains the sharp switching of even highly disordered FETs. If mobility varied smoothly with energy, current in such a device should switch gradually with gate voltage. Our results suggest, however, that the switching occurs when the Fermi energy sweeps past E_C, not the band edge. In other words, the current shuts off not because there are no carriers, but because carrier mobility has decreased sharply. This conclusion is supported by capacitance measurements in dirty inversion layers, which measure carrier densities as well as conductance [11]. Obtaining this result in a simulation, however, would be difficult without the sensitivity and resolution of the block recursion algorithm.

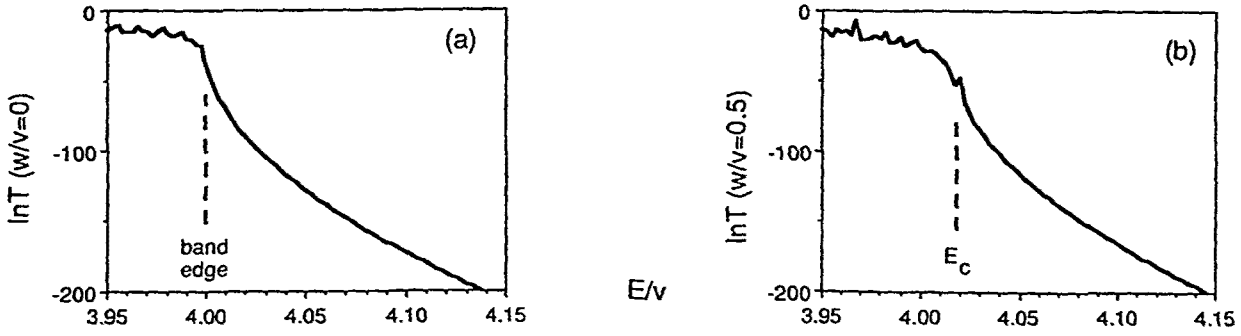

Figure 4. The variation of the logarithm of the transmittance with electron density for an ordered device (a), and a disordered device (b).

REFERENCES

1. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer Verlag, New York, 1984).
2. J.C. Wu, M.N. Wybourne, A. Weisshaar, and S.M. Goodnick, J. Appl. Phys. **74** (7) 4590 (1993).
3. T.J. Godin and R. Haydock, Phys. Rev. B **46**, 1528 (1992).
4. T.J. Godin and R. Haydock, Comput. Phys. Commun. **64**, 123 (1991).
5. T.J. Godin and R. Haydock, Phys. Rev. B **38**, 5237 (1988) and references therein.
6. R. Landauer, Phil. Mag. **21**, 863 (1970).
7. N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Holt Saunders, Philadelphia, 1976).
8. W.A. Harrison, *Electronic Structure and the Properties of Solids* (Dover, New York, 1989).
9. C.C. Paige, J. Inst. Math. Appl. **10**, 372 (1972).
10. N.F. Mott and E.A. Davis, *Electronic Processes in Non-Crystalline Materials*, 2nd ed (Oxford University Press, London, 1979).
11. R.A. Davies and M. Pepper, J. Phys. C **15**, L371.