2-1(Invited)

Simulation of Quantum Transport in Small Semiconductor Devices

M. V. Fischetti[1‡], S. E. Laux[2], and A. Kumar[2]
[1]Department of Electrical and Computer Engineering, University of Massachusetts
Amherst, Massachusetts 01003, USA
[2]IBM Semiconductor Research and Development Center (SRDC), IBM Research Division, T. J. Watson Research Center
P.O. Box 218, Yorktown Heights, New York 10598, USA
E-mail: fischett@ecs.umass.edu

***Abstract* – We describe a formulation of quantum electron transport in small devices based on the Master equation. We sketch its derivation from the Liouville-von Neumann equation, especially alluding to the subtle issues related to irreversibility. We compare this approach to alternative formulations of quantum transport and present results regarding ballistic and dissipative transport in double gate Si FETs.**

I. INTRODUCTION

The term "quantum transport" of charge-carriers in semiconductor devices addresses several different non-classical issues: Coherence and finite-collision-duration effects at short times[1], quasi-ballistic transport at low temperatures[2], high-field effects[3], such as intracollisional-field and collisional broadening corrections, or size-quantization (quantum confinement) effects[4]. Arguably, the latter corrections represent the effects most important to the study of devices at the nanometer length scale, since they imply large corrections to the 'semiclassical' electrostatics and to the conventional picture of carrier transport based on the Boltzmann Transport Equation (BTE). Within the framework of the BTE the scattering dephasing length, λ_ϕ, is considered to be much smaller than the size, L, of the 'active' region (*e.g.* channel, barrier regions) of the device. Moreover, the scattering-induced energetic broadening, $\delta E \sim \hbar\upsilon/\lambda_\phi$ (where υ is some group velocity and $\hbar$ the reduced Planck constant), is assumed to be much smaller than all other energy scales of the system. The end result of these assumptions is the possibility of characterizing the electronic wave-packets by specifying only their central position and momentum, as in a classical picture.

Here we review an approach to quantum transport which is both parallel and complementary to the BTE picture. We start from the density-matrix (ρ) formulation of transport and make use of Van Hove's observation[5] that most interactions of interest are 'self-averaging' because of the large number of degrees of freedom (phonons, electron-electron) or configurations (ionized impurities)[6] involved in the process. Thus, whenever carriers are injected from reservoirs (contacts) characterized by a dephasing length larger than the size of the device, $\lambda_\phi >> L$, a suitable set of basis-functions – namely, the propagating eigenmodes, $\{\psi_\mu\}$, of the device, instead of the 'usual' plane waves – exists for which the off-diagonal elements of the density matrix ($\rho_{\mu\nu}$ with $\mu \neq \nu$, representing coherent superpositions of the eigenmodes μ and ν) remain much smaller than the diagonal elements ($\rho_{\mu\mu}$, representing the occupation of the eigenmode μ), for times longer than the transit time in the device. This is because of the decoherence caused by the random action of the bath degrees of freedom on the eigenstates of the system. As we shall see below, we cannot overstress the importance of this choice of basis eigenfunctions to represent the density matrix: Any other representation would result in formulations dominated by the off-diagonal elements of ρ and, so, in transport equations significantly more complicated, as they would not make use of what we consider to be the 'natural' basis-states of the system. Thus, we are led to describe transport with the Pauli Master equation (PME):[7]

$$\frac{\partial \rho_{\mu\mu}}{\partial t} = \sum_\nu [W_{\mu\nu}\rho_{\nu\nu} - W_{\nu\mu}\rho_{\mu\mu}] + \left(\frac{\partial \rho_{\mu\mu}}{\partial t}\right)_{res}, \quad (1)$$

where $W_{\nu\mu}$ is the Golden-rule transition rate from a state μ to a state ν and the last term represents the injection/extraction from the contacts (a most troublesome term).

Note that the PME above is quite similar to the BTE, but the 'driving field' term is missing, since the field has been 'diagonalized' via the proper choice of basis states. This is exactly the reason why the PME represents a limit diametrically opposite to the BTE: The basis states for which the PME is valid are fully delocalized states and scattering processes are represented as fully nonlocal interactions.

II. MASTER EQUATION AND IRREVERSIBILITY

A rigorous derivation of the PME, Eq. (1), remains an elusive goal. Indeed, starting from the Liouville-von Neumann equation,

$$\frac{\partial \rho(t)}{\partial t} = -\frac{i}{\hbar}[\mathbf{H}_0(t), \rho(t)] - \alpha\frac{i}{\hbar}[\mathbf{H}'(t), \rho(t)], \quad (2)$$

(where $\mathbf{H}_0$ is the unperturbed Hamiltonian and $\alpha\mathbf{H}'$ the perturbing interaction term of dimensionless strength α), we note that this is a time-reversible equation, while Eq. (1) is irreversible (dissipative). Along the path from Eq. (2) to Eq. (1) we must necessarily meet the problem of understanding where irreversibility (and the associated problem of decoherence) enters the picture. Here we shall sketch the formal path we

‡A similar version of this work – performed at the IBM T. J. Watson Research Center, Yorktown Heights, New York – was presented at the International Electron Device Meeting, December 2003.

have followed[7], stressing many open problems.

Starting from Eq. (2) in the 'weak coupling' limit, $\alpha << 1$, we expand its formal solution by retaining only terms of order α^2. We are left with sums over matrix elements of the form:

$$\sum_{\lambda s} <\mu r|\mathbf{H}'(t)|\lambda s><\lambda s|\mathbf{H}'(t')|\nu p> , \qquad (3)$$

where λ, μ, and ν label the electronic states defined above and the indices r, s, and p label the internal states of the bath (*e.g.*, phonon occupation and momenta) or the various configurations (*e.g.*, various distributions of impurities). Van Hove observed[5] that most interactions of interest are such that when $\mu \neq \nu$ the sum over the infinite degrees of freedom, s, involves random phases. Thus, the contribution of these 'off diagonal' terms is of the order of the bath degrees of freedom (or of the number of configurations of the impurities[6]), N_b. If, however, $\mu = \nu$, then the sums yield an in-phase contribution $\sim N_b^2$. Thus, in the infinite-volume limit ($N_b \to \infty$), only the diagonal terms survive. Based on this observation, Van Hove has shown that a proper selection of 'initial' (or 'injected') states (namely, wavefunctions fully delocalized throughout our small device) guarantees that off-diagonal elements, initially absent, will not grow in time (ignoring virtual non-energy-conserving states at short times during a collision). We can now disregard off-diagonal terms, $\rho_{\mu\nu}$ with $\mu \neq \nu$, of the density matrix and consider only diagonal terms $\rho_{\mu\mu}$ (simply written as ρ_μ). Next, we invoke the Markov approximation for the phonon bath (or the random, uncorrelated distribution of impurities), the Hartree-Fock approximation for the many-electron Green's functions, and reach the equation (called 'preMaster' by Zwanzig[8]):

$$\frac{\partial \rho_\mu}{\partial t} \;=\; \sum_\lambda \int_0^t \; dt' [\; K_{\mu\lambda}(t,t')\; \rho_\lambda(t') \;-\; K_{\lambda\mu}(t,t')\; \rho_\mu(t') \;] \;, \qquad (4)$$

having denoted by $K_{\mu\lambda}(t,t')$ the kernel of the scattering processes, $\mathrm{Re}\sum_{r,s} <\mu r|\mathbf{H}'(t)|\lambda s><\lambda s|\mathbf{H}'(t')|\mu r>$. Surprisingly, Eq. (4) is still reversible[9], despite having used the Markov approximation and having gone to the infinite-volume limit. The final step required to convert Eq. (4) to the PME form, Eq. (1), requires taking the 'Van Hove limit' $\alpha^2 t \to$ constant as $\alpha^2 \to 0$ while $t \to \infty$, while also converting the kernel $K_{\mu\lambda}(t,t')$ to the Golden-Rule transition rate $W_{\mu\lambda}$. This last step requires the regularization of the (diverging) time integral in Eq. (4) by multiplying the integrand by $\exp(-\eta t)$. *This is the only term which breaks the time-reversal symmetry in the entire derivation*. Finally, taking the limit $\eta \to 0$, we obtain the energy-conserving delta-function characterizing completed collisions.

It is interesting to understand the connections between the PME and alternative formulations of quantum transport. The Keldysh-Kadanoff-Baym (KB) approach[10], especially in its excellent implementation by the NEMO group[11], considers as central quantity the nonequilibrium Green's function (NEGF) $G^<(\mathbf{r},\mathbf{r}',t,t')$, (see, for example, Datta's 'tutorial'[12]). Together with the function $G^<$, the advanced and retarded Green's functions, G^A and G^R, and the corresponding self energies, Σ, provide a complete description of nonequilibrium quantum transport in terms of the KB and Dyson equations. Use of the Wigner coordinates $T = (t+t')/2$, $\tau = t - t'$, $\mathbf{R} = (\mathbf{r}+\mathbf{r}')/2$ and $\rho = \mathbf{r} - \mathbf{r}'$ emphasizes the 'slowly varying' temporal and spatial dependences via the 'central' coordinates T and $\mathbf{R}$, amenable to an interpretation analogous to the classical picture, and the dependence on internal 'intracollisional' variables τ and ρ of a pure quantum nature. With a few exceptions – most notably the NEMO group[11], the problem is usually tackled by Fourier-transforming $G^<$ with respect to the 'internal' coordinates $\tau = t - t'$ and $\rho = \mathbf{r} - \mathbf{r}'$, and eliminating the dependence on fast (intracollisional) times by integrating over energy – as is done, for example, to obtain the Wigner function, $f(\mathbf{R},T,\mathbf{k})$:

$$f(\mathbf{R},T,\mathbf{k}) \;=\; -\frac{1}{2\pi} \int dE\; G^<(\mathbf{R},T,\mathbf{k},E) \;. \qquad (5)$$

This amounts to considering only the *equal-time* two-particle Green's function. The same step connects the equal-time $G^<$ to the density matrix: Abandoning the plane-wave (**k**) representation in favor of our basis-states $|\mu>$, one recognizes that $(\hbar/i)G^<_{\mu\nu}(t,t)$ is just the density matrix $\rho_{\mu\nu}(t)$. Indeed, one can follow Zubarev *et al.*[13] and derive the PME from the (generalized) KB equations by using Van Hove's argument to retain only the diagonal elements of the self-energies, and several *Ansätzen*, most notably, embracing the quasi-particle approximation by collapsing the spectral density function into a delta-function, thus losing information about the intracollisional dynamics and intercollisional correlations. As noted above, the Wigner function approach is identical to our Master equation approach, the latter being represented in $|\mu>$-space, the former in real space. Finally, the Semiconductor Bloch Equation (SBE) approach followed by Rossi and co-workers[14] accounts for the possible presence of off-diagonal terms of the density matrix, but is otherwise equivalent to the PME approach presented here.

Thus, the main strengths of the NEGF method originate from the use of the two-time Green's function and of basis 'atomic' basis function (often tight-binding localized orbitals). The latter feature permits the accurate simulation of structures at the atomic scale. On the down-side, since localized basis functions do not carry current, *the current operator is dominated by off-diagonal terms, unlike the PME*. It is this characteristic which renders scattering such a difficult proposition in the NEGF framework. The use of the two-time Green's function, however, permits, *at least in principle*, the inclusion of effects related to a 'broad' spectral-density function, (although, in practice, only second-order collisional broadening is often considered, scattering processes often being handled in a golden-rule fashion, all because of the mind-boggling complication of the Dyson equation), of off-diagonal terms of the density matrix resulting from intracollisional short-time effects and from injection from the reservoirs. In practice, though, reservoirs usually damp off-diagonal terms and a 'diagonal' injection has been recently found to be a necessary ingredient to avoid physical artifacts in the context of the Wigner and SBE approaches[15]. Finally, the real-space formulation usually employed to formulate the NEGF equations often causes scattering to be treated simplistically as a local process and requires contacts to be treated as 'decoupled' regions,

a treatment which is satisfactory only in the case of weak device-contact coupling, as in resonant-tunnel-diodes.

The major advantage of the PME approach remains the intuitive picture it provides, the exact treatment of the field, the fact that it leads to a diagonal-dominated formulation, thanks to the basis functions employed, and the realistic (namely, non-local) model one can afford for scattering processes. In addition, from a purely philosophical perspective, the PME approach introduces irreversibility in a transparent way, namely, when taking the Van Hove limit of completed collisions. The NEGF and the SBE approaches, instead, struggle with the formalization of scattering, leaving it as a "yet to be defined" concept: the NEGF introduces formally the self-energies $\Sigma^{<}$, $\Sigma^{>}$, Σ^{A}, and Σ^{R}, but they are defined as solutions of tremendously complicated Dyson equations. In practice, the golden rule (or, worse yet, local scattering processes) are used. The Wigner function also must employ a "Boltzmann-like collisional integral" of dubious origin, while, finally, scattering in the SBE approach is often treated in a 'golden rule' fashion. All of these methods find the source of irreversibility in the treatment of the contacts (defined as equilibrium, and so irreversible, reservoirs), but make no attempts to introduce or even clarify the origin of dissipation within their framework. Finally, the objections raised by Frensley against the PME concerning the violation of current continuity[16] obviously do not affect steady-state solutions[7] and, even during time transients, recent work by Gebauer and Car[17] identifies fluctuations in the bath during collisions as the component of the current, previously neglected, required to satisfy current continuity *exactly* within the PME context. Which method constitutes the best compromise is probably a function of the applications one has in mind (as well as "philosophical" personal preferences). NEGF is clearly well suited to resonant tunneling devices – for which scattering is not a major player and coupling to the reservoirs is relatively weak – and to "molecular" devices well described by a basis of localized orbitals. The SBE approach has been used extensively to investigate fast coherent effects, when extended to include also the so-called phonon-assisted density matrix[18]. The PME is probably the best choice for MOSFETs, in which the correct electrostatics and the role of quantitatively correct scattering rates are of paramount importance.

III. NUMERICAL IMPLEMENTATION

The problem we face consists of solving simultaneously the two-dimensional Schrödinger equation with open boundary conditions, the PME, Eq. (1), and the two-dimensional Poisson equation using the charge-density resulting from the occupations ρ_μ and appropriate boundary conditions. We start by solving the associated ballistic problem, that is, by ignoring the scattering terms in Eq. (1), so that only the contacts determine the occupations ρ_μ. The Schrödinger equation is formulated using a six-valley, parabolic band approximation describing approximately the conduction bands of Si. Full details have been given elsewhere[20, 19]. Here we shall simply emphasize three major elements of our procedure: 1. The discretization of the continuous spectrum of the electronic propagating states is performed by first solving the associated Schrödinger equation with *closed* boundary conditions[7]. By definition, the resulting spectrum of standing waves samples the local density of states of the device. Alternative, simpler discretization schemes (equi-energy or equi-wavevector discretization) have been found to be ineffective. 2. The contacts are treated as 'active' elements, the Fermi-Dirac distribution describing the reservoirs being 'drifted' in **k**-space so as to maintain charge neutrality near the contacts and satisfy continuity of flux across the device/reservoir boundaries. Alternative schemes have been proposed before (see the discussion in [16] and [7]). The use of any of these schemes is absolutely required in order to reach physically meaningful solutions. 3. Convergence is sought using a novel Broyden-Newton method[19] which allows fast convergence even in high-bias, high-current, strongly off-equilibrium situations.

Having obtained a ballistic solution, we employ a Monte Carlo method to solve the PME[7]. Its implementation parallels the conventional technique used to solve the BTE, since the PME is nothing but the BTE expressed on the basis of the traveling eigenstates $|\mu>$ of the Hamiltonian $\mathbf{H}_0$, and 'stripped' of the term expressing the action of the electrostatic field, since this is already diagonalized exactly when using the basis $|\mu>$ instead of plane waves. Results shown below have been obtained by including only intravalley scattering with acoustic phonons (in the elastic, equipartition approximation) and inelastic intervalley processes. The inclusion of scattering with ionized impurities, currently being studied, allows alternative interesting formulations, either by exploiting the configuration average to account for dissipation, in the spirit of Kohn and Luttinger[6], or by accounting for non-phase-breaking processes when a given configuration is considered. In this case correlations between scattering processes are retained, but a time-consuming configuration average (over many simulations) is required. Carrier-carrier scattering can also be shown to be an interaction of the Van Hove type, as long as the devices are sufficiently large to be populated by many carriers.

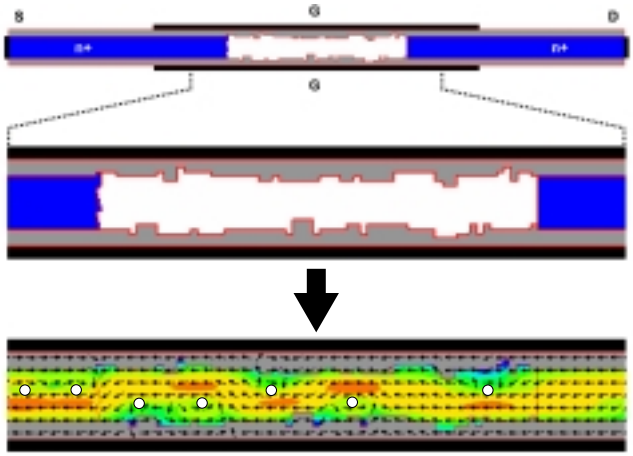

Figure 1: Bottom frame: Current density associated with one particular eigenmode of a 'straight' double-gate Si FET. Note the many vortices, signature of a reduced transmission amplitude. The top two frames show the geometry of the 'roughened' device.

IV. RESULTS AND CONCLUSIONS

We have previously investigated the role of the access geometry in ballistic double-gate (DG) Si FETs, showing that pure quantum effects (quantum reflection and diffraction in the source/channel constriction) can influence the current to an extent comparable to the that of scattering[21]. Another interesting 'ballistic' effect is the emergence of vorticity in 'bent' resonant-tunnel-diodes[22]. But, as illustrated in Fig. 1, the phenomenon is not unique when looking for vorticity in individual eigenstates: The roughened interfaces of a 'straight' DG Si FET induce the transfer of enough angular momentum to generate a large amount of vorticity. The observed reduced transmission amplitude for eigenstates showing vortices suggests the existence of an intriguing general correlation between vorticity, scattering, and resistance. Coherent scattering with ionized impurities is expected to cause similar effects, as already investigated by Barker[23].

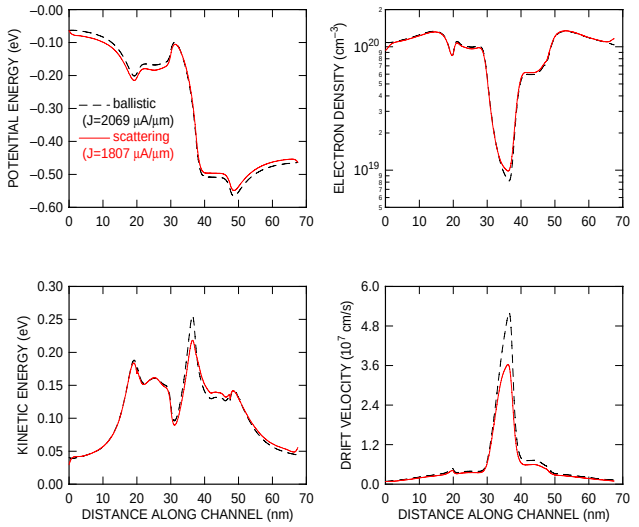

Figure 2: Clockwise from top left: Potential energy, carrier density, drift velocity, and kinetic energy along the channel of a 'taper' DG Si FET at 300 K with V_{GS} = 0.2 V and V_{DS} = 0.4 V obtained from a 2D solution of the coupled PME-Schrödinger-Poisson problem. All quantities are averaged over the width of the channel. The dashed black lines refer to a ballistic simulation, the solid red lines to a simulation performed acounting for intra- and inter-valley electron-phonon processes.

Our first example of a 2D simulation with scattering is the 'taper' DG FET described before[20]. Figure 2 shows the main results of the ballistic simulation (dashed black lines) – averaged over the transverse direction of the device – compared to the results obtained when accounting for electron-phonon scattering (solid red lines): The potential energy and carrier densities are barely altered (only noticeable is a slight increase of the electron concentration associated with the scattering-induced slow-down), while a drop in the maximum carrier energy and an even more significant reduction of the drift velocity are seen in the bottom panels. An interesting observation concerns the effect that inelastic scattering processes have on the current density: As shown in the top-left panel of Fig. (2), the current density drops from its 'ballistic' value of $\sim$ 2.09 mA/μm to a value barely 10% lower, $\sim$ 1.8 mA/μm. On the contrary, under purely ballistic transport the current density under the same bias conditions exhbits a variation of about 30% when moving from a 'straight' to a 'dog-bone' access geometry[21]. This unexpected result shows that purely quantum mechanical effects such as interference/refraction at geometrical contrictions may matter more than phase-breaking scattering processes at length scales comparable to the electron wavelength. Admittedly, the effect of the geometry may be amplified in purely balistic simulations (for example: scattering with impurities may perturb transport more than the geometry, thus damping the effect of geometric constrictions themselves on transport). Yet, these results are suggestive of the importance that non-phase-breaking, non-dissipative 'collisions' may have on transport at the nanometer length scale.

REFERENCES

[1] F. Rossi and T. Kuhn, Rev. Mod. Phys. **74**, p. 895 (2003).

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

[3] J. Bardeen and W. Shockley, Phys. Rev. **81**, p. 69 (1950).

[4] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, p. 437 (1982).

[5] L. Van Hove, Physica **XXI**, p. 517 (1955).

[6] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, p. 590 (1957).

[7] M. V. Fischetti, J. Appl. Phys. **83**, p. 270 (1998); Phys. Rev. B **59**, p. 4901 (1999).

[8] R. W. Zwanzig, in *Quantum Statistical Mechanics*, Paul H. E. Meijer, Ed., (Gordon and Breach, New York, 1966), p. 139.

[9] R. Jancel, *Foundations of Classical and Quantum Statistical Mechanics*, Pergamon, Braunschweig, 1969).

[10] L. Kadanoff and G. Baym, *Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and Nonequilibrium Problems*, (Benjamin, New York, 1962).

[11] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, p. 7845 (1997).

[12] S. Datta, IEDM Tech. Digest, p. 703 (2002).

[13] V. Morozov D. Zubarev and G. Roepke, *Statistical Mechanics of Nonequilibrium Processes*, (Akademie Verlag, Berlin, 1997).

[14] F. Rossi, A. Di Carlo, and P. Lugli, Phys. Rev. Lett. **80**, p. 3348 (1988).

[15] R. Proietti-Zaccaria and F. Rossi, Phys. Rev. B **67**, p. 113311 (2003).

[16] W. R. Frensley, Rev. Mod. Phys. **62**, p. 745 (1990).

[17] R. Gebauer and R. Car, Phys. Rev. B **70**, p. 125324 (2004).

[18] J. Schilp, T. Kuhn and G. Mahler, Phys. Rev. B **50**, p. 5435 (1994).

[19] S. E. Laux, A. Kumar, and M. V. Fischetti, IEDM Tech. Digest, p. 715 (2002).

[20] S. E. Laux, A. Kumar, and M. V. Fischetti, IEEE Trans. Nanotechnol. **1**, p. 255 (2002).

[21] S. E. Laux, A. Kumar, and M. V. Fischetti, J. Appl. Phys. **95**, 5545 (2003).

[22] S. E. Laux, A. Kumar, and M. V. Fischetti, J. Comp. Electron. **2**, p. 105 (2003).

[23] J. R. Barker, J. Comp. Electron. **2**, p. 153 (2003).