AD-A280 181

Scientific Research Associates, inc.

50 Nye Road, P.O. Box 1058 Glastonbury, Connecticut 06033
Tel: (203) 659-0333 Fax: (203) 633-0676

FINAL REPORT R-9231F

THE PHYSICS AND OPERATION OF ULTRA-SUBMICRON LENGTH SEMICONDUCTOR DEVICES

Submitted to
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

May 1994

Approved for Public Release;
Distribution Unlimited

THE PHYSICS AND OPERATION OF ULTRA-SUBMICRON LENGTH SEMICONDUCTOR DEVICES

TABLE OF CONTENTS

THE PHYSICS AND OPERATION OF
ULTRA-SUBMICRON LENGTH SEMICONDUCTOR DEVICES

ABSTRACT

This document summarizes activities under ONR Contract: N00014-86-C-0780, under which equilibrium and nonequilibrium electron and hole transport in micron and submicron structures were studied via a wide range of numerical procedures. These included Monte Carlo methods, moments of the Boltzmann transport equation, Schrodinger's equation and the quantum Liouville equation in the coordinate representation. While all of the studies have resulted in a large collection of publications, the basic theme of the studies was the determination of the physics of device operation and the influence of small structure size on this operation. The most recent activities have involved the quantum Liouville equation with emphasis on dissipation and the calculation of current. This document includes a description of quantum transport via the quantum Liouville equation, as we now understand it, as well as a brief summary of the previous activities involving larger submicron devices. While the principle goal of this study was elucidating the physics and operation of nanoscale devices, a continuing requirement was that all algorithms be menu driven and accessible to device scientists and engineers. The quantum transport algorithm is accessible on UNIX workstations and in a PC Windows format.

THE PHYSICS AND OPERATION OF ULTRA-SUBMICRON LENGTH SEMICONDUCTOR DEVICES

1. PREFACE

From it's inception the study discussed below, performed under ONR Contract N00014-86-C-0780 has concerned itself with equilibrium and nonequilibrium electron and hole transport in micron and submicron structures. All relevant equations and procedures were invoked and included Monte Carlo methods, moments of the Boltzmann transport equation, Schrodinger's equation and the quantum Liouville equation in the coordinate representation.

The more classical problems emphasized hot carrier phenomena and transients, while the quantum transport was concerned with specific quantum phenomena and the best means of studying it. Quantum transport has occupied most of our activities in the past few years, and the major success in the program was the recent ability to compute current self-consistently within the framework of a dissipation model. Two examples serve to illustrate. This model when coupled to earlier models now permits us to deal with transients in a sensible manner in that the relaxation to an intermediate state is better defined.

The approach we have taken is different from those of others because our goals were very general and included the requirement that any and all algorithms include tools that device scientists and engineers could utilize as part of routine device design tasks. In other words one goal was to include algorithms that would be as accessible as the standard drift and diffusion equations.

The quantum transport equation we deal with is the quantum Liouville equation in the coordinate representation. Recall that Schrodinger's equation is a coordinate representation description. In dealing with the quantum Liouville equation in the coordinate representation we broke new ground, particularly with respect to devices. For example boundary conditions that workers typically employ in solving the drift and diffusion equations were discarded. In it's place it was necessary to incorporate quasi-Fermi level conditions at the boundaries to assure flat band contact conditions. The issues of Fermi statistics was not treated within the framework of the differential equations, which would formally require the introduction of the Dirac Hamiltonian into the quantum Liouville equation. Instead statistics were accounted for through boundary conditions.

The calculation of current was introduced self-consistently and coupled to the quasi-Fermi levels. The quantum Liouville equations were also used as a basis for justifying earlier and more recent work on the quantum potential.

This document summarizes these studies.

Many papers were either published or submitted for publication during this study and one huge review article was initiated. A copy of each of these is included with this report.

2. INTRODUCTION

Since the pioneering work of Tsu and Esaki[1], the experimental studies of Sollner et al.[2] on double-barrier resonant tunneling devices, and the superlattice detector work of Levine et al.[3], there has been growing interest in barrier/well devices and in the fundamental underpinnings of quantum device operation. Further, following the work Datta et al.[4], there has also been rising interest in the basic physics accompanying the Aharonov-Bohm[5] effect in heterostructures. Indeed, major advances in material technology has enabled device scientists to conjecture about new device structures that both test and illustrate basic fundamental quantum physics issues of few and many particle systems. For example the issue of *nonlocality* now finds its way into discussions of transport in quantum devices. Nonlocality in classical physics is illustrated by the coulomb interaction that decreases as the square of the distance between particles. In quantum mechanics there are additional interactions that do not necessarily drop off with distance and these are discussed below.

Another issue involves *dissipation*. Schrodinger's equation as traditionally used is dissipationless, and if all transport in subsystems were governed by Schrodinger's equation without interactions between the subsystems, all transport would be ballistic. Dissipation in quantum mechanics is treated by introducing additional systems, e.g., phonons, and allowing the additional system to cause a transition between states of the original system.

A third issue, specific to the treatment of electronic devices, is the *reservoir*. Traditionally, the examination of classical devices involves the specification of densities on the bounding surfaces, regarded as reservoirs. Such specification, which is assumed to remain valid under bias, often involves the

---

[1] R. Tsu and L. Esaki: "Tunneling in a Finite Superlattice," *Appl. Phys. Lett.*, 22, 562 (1973)

[2] T.C.L.G. Sollner, W.D. Goodhue, P.E. Tannewald, C.D. Parker and D.D. Peck: "Resonant Tunneling Through Quantum Wells at Frequencies up to 2.5 THz," *Appl. Phys. Letts.*, 43, 588 (1983).

[3] B.F. Levine, K.K. Choi, C.G. Bethea, J. Walker and R.J. Malik: "New 10 micron Infrared Detector Using Interband Absorption in Resonant Tunneling GaAlAs Superlattices," *Appl. Phys. Letts.*, 50, 1092 (1987).

[4] S. Datta, M.R. Mellcoh, S. Bandyopadhyay and M.S. Lundstrom: *Appl. Phys. Letts.*, 48, 487 (1986).

[5] Y. Aharanov and D. Bohm: *Phys. Rev.*, 115, 485 (1959).

concept of a quasi-Fermi level, in which the energy separation between the bottom of the conduction band and the Fermi level at the boundary remains unchanged.

Presently, our ability to incorporate these quantum mechanical issues to describe physical phenomena in ultra small devices and to propose quantum phase based devices has been evolutionary. Through a coupling of experiment, theory and numerical simulation we have been better able to understand how basic quantum mechanical processes affect device physics. But the 'goodness' of a description of quantum transport lies in the ability of the theory to explain the detailed experimental results obtained from such complex devices as, e.g., two terminal resonant tunneling diodes (RTD), quantum well superlattice detectors, and the more common heterostructure FETs. However, the complexity of the RTD and the puzzle associated with understanding its detailed operational principles has led Ferry[6] to describe it as the *fruit fly* of quantum transport device theory. How good is the fruit fly analog.

Traditionally, transport in RTDs and other barrier structures has been analyzed through implementation of the formula[1] :

$$(1) \qquad J = \left[2e/(2\pi)^3\right] \int d\mathbf{k}\, v(\mathbf{k}) \left[f_{FD}(E) - f_{FD}(E + e\phi)\right] |T(E,\phi)|^2$$

It is the approximations associated with this formulae that provide the bounds of our understanding of transport in quantum structures. In equation (1) f_{FD} is the equilibrium Fermi-Dirac distribution function, $T(E,\phi)$ is the transmission coefficient obtained from solutions to the time independent Schrodinger equation, E is the energy of the tunneling particle and ϕ the applied potential. As discussed by Kluksdahl et al.[7] a major criticism of this approach is that it requires knowledge of the distribution function at each side of the tunneling interface, rather that the bulk like distribution far from the tunneling interface. Additionally the form of equation (1) also implies: (1) the use of equilibrium distribution functions to describe a biased state, when the biased resonant tunneling diode is in a non-equilibrium state; (2) the neglect of scattering, although scattering would be

---

[6] D. K. Ferry, *Theory of Resonant Tunneling and Surface Superlattices*, a chapter in *Physics of Quantum Electron Devices*, F. Capasso (ed) Springer-Verlag, Berlin pp77-106 (1989)

[7] N.C. Kluhsdahl, A.M. Kriman and D.K. Ferry: "Self-Consistent Study of the Resonant Tunneling Diode," *Phys. Rev.*, B39, 7720 (1989).

required to force a system to a state of equilibrium; and (3) the concept of a Fermi level, which clearly implies the presence of strong carrier-carrier interactions, particularly in the quantum well.

While the use of equation (1) has been successful in predicting negative conductance in RTD its inadequacies in explaining experiment have been well documented. These include: *First*: the dc studies do not account for the peak-to-valley ratio of resonant tunneling devices. *Second*: the dc studies do not adequately treat dissipation. *Third*: the dc studies do not treat hysteresis in the current voltage characteristics, observed experimentally. *Fourth*: the dc treatment cannot predict how the devices will be used in applications. *Fifth*: the dc treatment cannot treat the time dependent nature of the boundary conditions that represent physical contacts.

The above studies suffer from lack of incorporating the feature basic to quantum mechanical phenomena: *all quantum mechanical devices are time dependent*. Apart from dissipation, there are always reflections off boundaries, barriers, wells, imperfections and contacts. *What is needed is a time dependent large signal numerical studies of quantum feature size devices*. This need has been discussed by Ravaioli et al.[8] and Frensley[9] and more recently Ferry and Grubin[10]. This approach emphasizes the details of transient behavior, the numbers of particles involved in device operation, the temporal duration under which the effective mass approximation is valid, the significance of the Fermi-golden rule, and other short time phenomena. As currently practiced, when scattering is present, or when time dependent fields are present and treated as perturbations, it is supposed that the perturbation does not modify the states of an unperturbed system, rather the perturbed system instead of remaining permanently in one of the unperturbed states is assumed to be continually changing from one to another, i.e., undergoing transitions from one state to another state. This approach is at the

---

[8] U. Ravaioli, M.A. Osman, W. Potz, N. Kluksdahl and D.K. Ferry: "Investigation of Ballistic Transport Through Resonant-Tunneling Quantum Wells Using Wigner Function Approach," *Physica*, 134B, 36 (1985).

[9] W. Frensley: "Boundary Conditions for Open Quantum Systems Driven Far From Equilibrium," *Reviews of Modern Physics*, 62, 745 (1990).

[10] D. K. Ferry and H. L. Grubin, "Modeling of Quantum Transport in Semiconductor Devices" Chap. In *Solid State Physics* (H. Ehrenreich, ed) Academic Press (1994)

heart of those calculations employing the density matrix[11], those employing the Wigner distribution function[12], and those employing Green's function techniques[13].

In addition to these fundamental approaches there are also *derivative* procedures that enjoy wide spread use, both for the intuitive nature of the equations and because of the ease with which classical concepts emerge. These discussions include the quantum moment equations, see e.g., Iafrate et al.[14], Stroscio[15], and Grubin and Kreskovsky[16].

In the discussion that follows the density matrix and quantum moment equations were implemented in the study of quantum feature size devices. Further we have found insight for multiparticle transport based upon concepts obtained through a recasting of the single particle Schrodinger equation. Adopting the approach of Bohm[17] the single particle wave function is written in the form:

$$(2) \qquad \psi = R\exp\left[\frac{iS}{\hbar}\right]$$

subject to the condition that increasing the phase by 2π, does not change the wave-function. This wave function when inserted into Schrodinger's equation results in two equations:

---

[11] See, e.g., H. Ehrehreich and M. H. Cohen, *Phys. Rev.*, 115, 786 (1959), J. Goldstone and K. Gottfried, *Il Nuovo Cimento*, 13, 849 (1959), and more recently, W.A. Frensley, *Rev. Mod. Phys.*, 62, 745, (1990), which include a discussion of the density matrix in the coordinate representation. Most recently a discussion by J. B. Krieger and G. J. Iafrate, *Phys. Rev.* B35, 9644 (1987) and G. J. Iafrate and J. B. Krieger, *Phys. Rev.* B40, 6144 (1989) for a discussion of the density matrix in the momentum representation.

[12] E. P. Wigner, *Phys. Rev.*, 40, 749 (1932)

[13] R. Lake and S. Datta, *Phys. Rev.* B45, 6670 (1992)

[14] G.J. Iafrate, H.L. Grubin and D.K. Ferry: "Utilization of Quantum Distribution Functions for Ultra-Submicron Device Transport," *J. De Physique*, 10, C7-307 (1981).

[15] M. A. Stroscio: "Moment-Equation Representation of the Dissipative Quantum Liouville Equation," *Superlattices and Microstructures*, 2, 83 (1986).

[16] H.L. Grubin and J.P. Kreskovsky: "Quantum Moment Balance Equations and Resonant Tunneling Structures," *Solid State Electronics*, 32, 1071 (1989).

[17] See, e.g., C. Philippidis, D. Bohm and R.D. Kaye: *Il Nuovo Cimento*, 71B, 75 (1982). More recently see D. Bohm and B. J. Hiley, *The Undivided Universe*, Routledge, London (1993)

(3) $$\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V+Q=0$$

and

(4) $$\frac{\partial R^2}{\partial t}+\nabla\cdot\left(R^2\frac{\nabla S}{m}\right)=0$$

where:

(5) $$Q\equiv-\left(\frac{\hbar^2}{2m}\right)\frac{\nabla^2 R}{R}$$

Equations (3)-(5) indicate that the Schrodinger wave represents a particle with a well defined position whose value is causally determined. The particle is never separate from the quantum force, $-\nabla Q$, that fundamentally affects it. The particle has an *equation of motion:*

(6) $$m\frac{dv}{dt}=-\nabla(V+Q)$$

which means that the forces acting on the particle consist of the classical force, $-\nabla V$, and the quantum force, $-\nabla Q$. It is important to note that the quantum potential is dependent on the shape of the real part of the wave function rather than on its intensity; and does not necessarily fall off with distance. The quantum force is dependent on the momentum of the carrier through the continuity equation, but does not require a source term

The quantum potential is defined in terms of a single particle wavefunction. And if $S(\mathbf{r},t)=s(\mathbf{r},t)-Et$, where E, is a constant independent of position, then under zero current conditions, equation (3) is the real part of Schrodingers equation whose solutions subject to a particular set of conditions leads to a set of bound state eigenvalues. We will come back to this point over and over again, in the discussion that follows.

While the above discussion is for single particle wave functions we are interested in quantal and classical distribution functions, both representing an ensemble of particles. Our experience, has developed from approximate representations of the Wigner distribution function[12], indicates that the incorporation of the quantum potential for an ensemble of particles, where the *amplitude R is replaced by the square root of the self-consistent density, ρ(x), is a significant aid in interpreting much of the salient features of quantum transport in devices*. The use of the quantum potential provides an alternative explanation for the peaking of the charge density at positions away from the interface of wide and narrowband gap structures; for real space transfer, for the potential distribution associated with a Schottky barrier, for density variations associated with variations in effective mass, and a host of additional features. To get to these points we must get through some mathematics, part of which is exact, and part approximate. We begin with the development of the single particle density matrix.

3. THE SINGLE PARTICLE DENSITY MATRIX

While the density matrix approach discussed below and the Wigner approach are mathematically equivalent, we have made the choice of the density matrix because the equation of motion readily submits to algorithms developed by the authors; the use of which are extremely short computational times for steady state solutions. These algorithms are discussed below. There are limitations to our treatment. The most important is that the equation of motion discussed below does not include anti-symmetric components and the density matrix has not been subject to anti-symmetrization[18]. We note that the application of the Wigner formulation to devices suffers from the same limitation. In some of the studies below, the inclusion of Fermi statistics is through the boundary conditions, as in the Wigner studies.

The structures that we discuss fall under the category of *open structures*[9], which can exchange particles with its surrounding, and which mathematically expresses this interaction in terms of boundary conditions. The phenomena we are interested in will be with systems that are far from equilibrium.

---

[18] A brief discussion of anti-symmetrization is included in the monograph: "Foundations of Electrodynamics," S.R. De Groot and L.G. Suttorp, North-Holland Publishing Company, Amsterdam (1972). See also O'Connell (get reference).

The density matrix is obtained from the density operator $\rho_{op}(t)$, which following Dirac notation[19] is:

$$(7) \qquad \rho_{op}(t) = \sum_i |i(t)> P(i) < i(t)|$$

where $|i(t)>$ represents an eigenstate[20]. The time evolution of the density operator is obtained from the time evolution operator[21] $U(t,t_0)$ which has the property $U(t,t_0)|i(t_0)> = |i(t)>$. The time evolution operator is unitary and the dependence of the density operator on previous times is given by:

$$(8) \qquad \rho_{op}(t) = U(t,t_0)\rho_{op}(t_0)U(t,t_0)^\dagger$$

where the symbol '$\dagger$' represents the adjoint. The time dependence of the density operator is governed by the time dependence of the time evolution operator, which is[22]:

$$(9) \qquad i\hbar\frac{dU(t,t_0)}{dt} = H(t)U(t,t_0)$$

Assuming that the Hamiltonian *H(t)* is Hermetian, the time dependence of the density operator is:

---

[19] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, London (1958). Particular attention should be paid to Section 33, where we note that if $P(i)$ is the probability of the system being in the *i*th state, it can never be negative. If ρ' is an eigenvalue of ρ_{op} and $|\rho'>$ is an eigenket belonging to this eigenvalue, then $\rho_{op}|\rho'> = \rho'|\rho'>$. As discussed in section 33, ρ' cannot be negative.

[20] Note, later we will be expressing our results in the coordinate representation. As discussed by P. R. Holland, *The Quantum Theory of Motion*, Cambridge University Press, Cambridge (1993), page 104, a mixed state may be decomposed in an infinite number of ways, and so we cannot uniquely deduce from it the set of eigenstates in the ensemble and their respective weights. The same will apply to the Wigner distribution function, which is obtained from it through a transformation, and has the same content.

[21] Reference [19], section 27.

[22] See, e.g., A. Messiah *Quantum Mechanics*, Volume II, John Wiley & Sons, NY, (1961), particularly Chapter XVII.

(10) $$i\hbar \frac{d\rho_{op}(t)}{dt} = \left[H(t), \rho_{op}(t)\right]$$

The density matrix in the coordinate representation is given by:

(11) $$i\hbar \frac{\partial < \mathbf{x}|\rho(t)|\mathbf{x}'>}{\partial t} = \left\{ -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial \mathbf{x}^2} - \frac{\partial^2}{\partial \mathbf{x}'^2} \right) + V(\mathbf{x}) - V(\mathbf{x}') \right\} < \mathbf{x}|\rho(t)|\mathbf{x}'> + i\hbar \left\{ \frac{\partial < \mathbf{x}|\rho|\mathbf{x}'>}{\partial t} \right\}_{scattering}$$

Notice that we are ignoring any spatial variation in the effective mass, although we will deal with this later[23]. The last term on the right hand side of equation (11) is a generic representation of scattering, which we treat below in a semiclassical manner. All of the quantum features associated with the devices below will arise from the streaming terms.

The density matrix is Hermetian, and $\rho(\mathbf{x},\mathbf{x}') \equiv < \mathbf{x}|\rho_{op}|\mathbf{x}'> = \rho(\mathbf{x}',\mathbf{x})^*$. Additional quantities relevant for transport include the current density matrix:

(12) $$\mathbf{j}(\mathbf{x},\mathbf{x}') = \frac{\hbar}{2mi}(\nabla_{\mathbf{x}} - \nabla_{\mathbf{x}'})\rho(\mathbf{x},\mathbf{x}')$$

and the energy density matrix:

(13) $$E(\mathbf{x},\mathbf{x}') = -\frac{\hbar^2}{8m}(\nabla_{\mathbf{x}} - \nabla_{\mathbf{x}'})^2 \rho(\mathbf{x},\mathbf{x}')$$

The diagonal components of each represent the observables.

Equation (11) when coupled to Poisson's equation :

(14) $$\nabla \cdot (\varepsilon \nabla) V = -e^2(\rho(\mathbf{x}) - \rho_0(\mathbf{x}))$$

[23] A key study is referenced here G. T. Einevoll and L. J. Sham, *Boundary conditions for envelope functions at interfaces between dissimilar materials*, *Phys. Rev.* **B49**, 10533 (1994).

and *equations describing scattering* are the relevant equations for device transport. Note while the above equations are for electrons, we will also discuss hole transport; the relevant modifications to the equation will be indicated.

The Liouville equation in the coordinate representation is a function of six variables plus time. The six variables represent a coordinate phase space whose relation to the standard phase space involving position and momenta may be assessed through application of the Weyl transformation[24], which has been modified to include spin

$$(15) \quad \rho(\mathbf{x},\mathbf{x}') = \left[\frac{2}{(2\pi)^3}\right]\int_{-\infty}^{+\infty} d\mathbf{k}\, f_W\left[\mathbf{k},\frac{(\mathbf{x}+\mathbf{x}')}{2}\right]\exp[i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')]$$

To date the description of transport in devices via the density matrix has been confined to cases where the particles are free in two directions, which for specificity we take as the 'y' and 'z' directions. Further in the discussion below we will deal with diagonal components along the free directions, and treat the density matrix $\rho(x,x',y=y',z=z')\equiv\rho(x,x')$.

To determine the form of the density matrix, we can picture a situation in the absence of dissipation in which boundary conditions permit the separation of equation in two Schrodinger type equations, with a solution that is the product of two wave functions. More generally we seek solutions of the type:

$$(16) \quad \rho(x,x',t)=\sum_i f_i\,\Psi_i^*(x',t)\Psi_i(x,t)$$

for which equation (15) is a special case. We now consider several examples.

4. EXAMPLES OF THE EQUILIBRIUM DENSITY MATRIX

For a Fermi-Dirac distribution function:

$$(17) \quad f_W(\mathbf{k},\mathbf{x})=\frac{1}{1+\exp[(E-E_F)/k_B T]}$$

and for parabolic bands the density matrix is:

---

[24] H. Weyl, *Z. Physik*, 46, 1 (1927).

(18) $$\rho(x,x') = \left[\frac{N_C}{\Gamma(3/2)}\right]\left[\frac{\lambda}{(x-x')}\right]\int_0^{+\infty} d\mu \frac{\sin\left[\mu^{1/2}(x-x')/\lambda\right]}{1+\exp\left[\mu-\mu_n\right]}$$

Here, $Lim_{x\to x'}\rho(x,x') = N_c F_{1/2}(\mu_n)$, $F_{1/2}(\mu_n) = [\Gamma(3/2)]^{-1}\int_0^{+\infty}\frac{\mu^{1/2}d\mu}{1+\exp[\mu-\mu_n]}$, $\mu = (E-E_C)/k_BT$, $\mu_n = (E_F-E_C)/k_BT$, $N_C = \Gamma(3/2)/(2\pi^2\lambda^3)$ is the density of states and $\lambda^2 = \hbar^2/(2mk_BT)$ is *the square of the thermal deBroglie wavelength.*

There are two limiting cases that submit to analytical expression. In the high temperature limit, where Boltzmann statistics apply (the Boltzmann distribution arises when $\mu_n < -4$)

(19) $$\rho(x,x') = N_c \exp\left[\mu_n - (x-x')^2/4\lambda^2\right]$$

This distribution is Gaussian. For a material such as gallium arsenide, the thermal deBroglie wavelength at room temperature is 4.7nm and N_c = 4.4x10^{23}/m^3. For a nominal density of 10^{23}/m^3, $\mu_n = -1.48$. In the low temperature limit, e.g., T=0K [25]:

(20) $$\rho(x,x') = \left[\frac{k_F^{\ 3}}{\pi^2}\right]\frac{j_1[k_F(x-x')]}{k_F(x-x')}$$

where $j_1(z)$ is a spherical Bessel function, $E_F = \hbar^2 k_F^2/2m$, and $k_F = \left[3\pi^2 N\right]^{1/3}$. In the limit as $z \Rightarrow 0$, $j_1(z) \Rightarrow z/3$. One of the earliest applications involving equation (20) was in a discussion by Bardeen [26] where it was demonstrated that the electron density profile a distance 'z' from an infinite barrier was:

[25] See also equation A 5.1.7 in N.H. March, *Solids: Defective and Perfect,* appearing in *The Single-Particle Density in Physics and Chemistry,* N. H. March and B. M. Deb, editors, Academic Press, London (1987)

[26] J. Bardeen, *Phys. Rev.* 49, 653 (1936)

$$(21) \qquad \rho(z) = N\begin{bmatrix} 1-\dfrac{3}{2}\dfrac{j_1[2k_F z]}{k_F z} & for\ z>0 \\ 0 & otherwise \end{bmatrix}$$

Figure 1 displays the density matrix corresponding to equation (20) for a density of $10^{23}/m^3$.

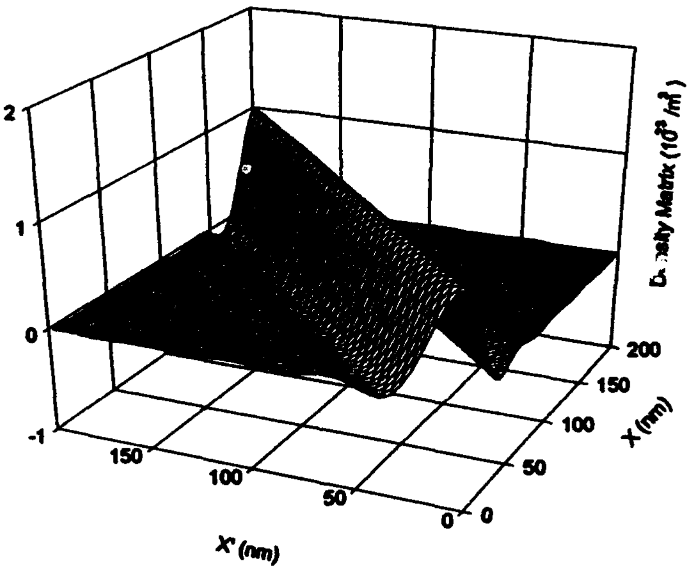


Figure 1. Density matrix for free particles weighted by a Fermi distribution for GaAs at T=0K. The density is $10^{23}/m^3$.

The oscillation in the density matrix along the direction (correlation direction) normal to the diagonal is determined by the argument of the spherical Bessel function. The periodicity depends on density as expressed by the Fermi wave number, and suggests the possibility of a wavenumber dependent resonance. The oscillation disappears at room temperature where the distribution approached as Gaussian as described by equation (19). The progressive decrease in the numbers of oscillations as the temperature increases is displayed in figure 2, which displays a cut of the density matrix in a plane normal to the diagonal of the density matrix. The effects of Fermi statistics are also more pronounced as the density is increased (e.g., k_F is increased) and we expect this to manifest itself in the oscillatory character of the density matrix.

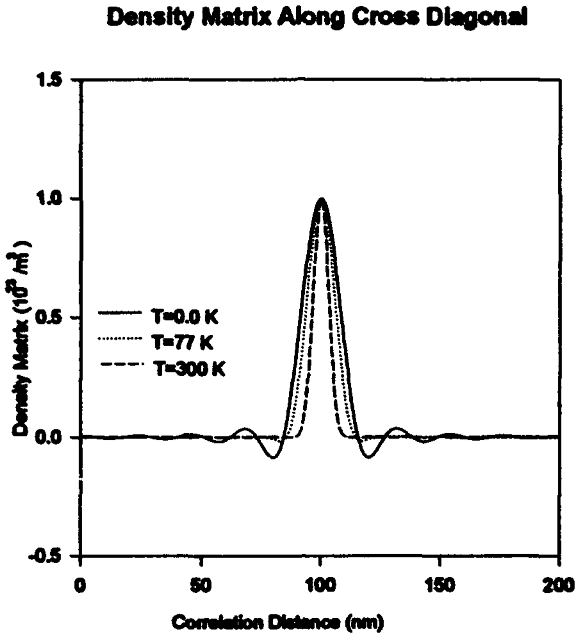


Figure 2. Density matrix versus correlation distance for free particles weighted by a Fermi distribution for GaAs at T=0K, 77K and 300K.. The density is 10^{23} /m^3

The density matrix $\rho(x,x')$ shown in figure 1 is plotted for a range of values of x and x', $(0<x<200nm, 0<x'<200nm)$. The density 'observable' $\rho(x) \equiv \rho(x,x)$ is the value of the density matrix along the diagonal and is *plotted* along the physical coordinate x. Pictorially, the density is a *projection* of the diagonal component of the density matrix onto the x-axis. The density matrix along the cross diagonal is defined as $\rho_{cross}(x) \equiv \rho(L-x,x)$, where L is the length of the structure; it is shown as a projection onto the x-axis.

The above discussion provides an indication of what the density matrix coordinate representation profiles corresponding to standard classical equilibrium distribution functions look like. It is expected on physical grounds that a *classical* problem studied with the classical distribution function in momentum space would yield the same physical results as that obtained with coordinate representation density matrix. For example, classically, with the Boltzmann distribution, the probability distribution is proportional to $\exp -V(x)/k_B T$. Thus, when a potential energy change equal to $k_B T \ell n 10$ (0.059521 ev at room temperature) is considered, classical theory teaches that the density will be reduced by an order of magnitude. Solving the equation of motion of the density matrix for this case provides the same result. If we go to the other extreme at T=0K, and recognize that the Fermi energy relative to the bottom of the conduction band, $E_F - E_C$, corresponding to a density of $10^{24}/m^3$ is 54.4 mev, while that corresponding to a density of $10^{23}\ /m^3$ is 11.7 mev, then introducing a barrier of 42.7 mev will reduce the density by an order of magnitude. This is shown in figure *3*.

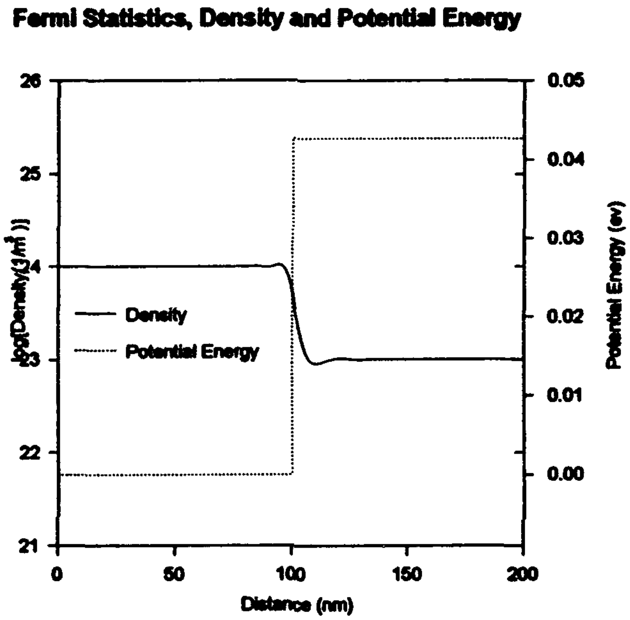


Figure 3. For GaAs at T=0K, Fermi statistics, with a step change in potential energy from 0.0 ev to 0.0427 ev (dotted line), the non-self consistent spatial variation in density (solid line).

Apart from the asymptotic (classical) values of density far from the interface we point to the local oscillation in density on either side of the interface, and make note of the position of the peak and minimum values of density. Classical studies indicate that the peak value of density occurs at the interface; while all quantum mechanical studies indicate that the peak is shifted away from the interface. In a recent density matrix study[27], devoted to Boltzmann statistics, it was analytically demonstrated that the density could be represented in equilibrium as being equal to

[27] H. L. Grubin, T. R. Govindan, J. P. Keskovsky and M. A. Stroscio, *Sol. St. Electron*, **36**, 1697 (1993)

(22) $\rho(x) = N_c \exp[\mu_n - (V(x) + Q(x)/3)]$

In the absence of the quantum potential the density is determined solely by the potential energy, and so the density for the potential energy distribution of figure 3 would be equal to its left hand value right up to the potential barrier, and a second (lower) value within the potential barrier. The finite value of the quantum potential and its spatial variation is responsible for the minimum and maximum values of the density occurring away from the interface. This will be discussed in more detail below where we will also illustrate the value of the quantum potential. The factor '3' that appears in equation (22) is discussed in detail in reference 27.

The potential variation in figure 3 is imposed and abrupt. Alternatively we can envision a structure in which the *density* changes abruptly at the same point (100nm). Then a solution to the Liouville equation and Poisson's equation yield a potential distribution whose values asymptotically approach those of figure 3. The potential distribution at the interface is no longer abrupt, and the local peak seen in figure 3 is absent. Rather, there is a more gradual decrease in density across the interface, with values that cannot be described by the classical distribution, but require the presence of the quantum potential. The two dimensional density matrix for the calculations of figure 3 are shown in figure 4.

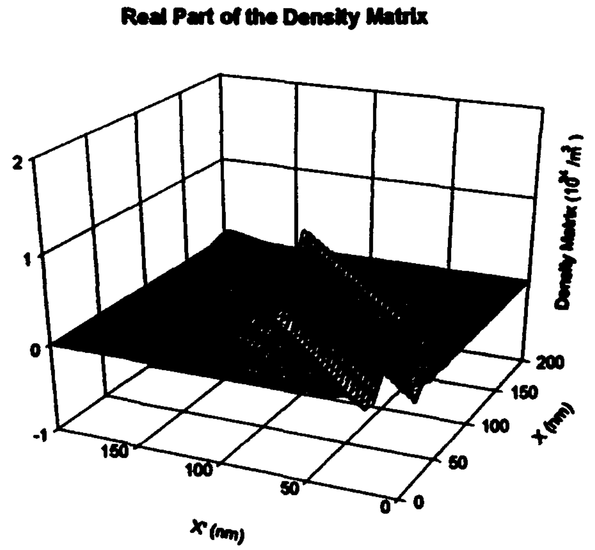


Figure 4. Two dimensional density matrix from which the results of figure 3 are obtained.

The origin of the scales in figure 4 is closest to the reader where the density matrix has its highest values. Notice the ripples in the density matrix closest to the highest density regions. Ripples are also present at the lower density regions but their period and magnitude are weaker. Generally the effects of Femi statistics are more pronounced at higher densities, where from equation (20) it is seen that the amplitude of the oscillation increases and the period decreases, with increasing density.

5. EQUILIBRIUM DISTRIBUTIONS AND THE QUANTUM POTENTIAL

As indicated in the earlier discussion the classical distribution function accounts incorrectly for the charge distribution in the vicinity discontinuities in potential energy and cannot be used if the goal is a description of the operational physics of devices; the quantum potential must be included. Additionally, we have also used the quantum potential as an aid in interpretation. Several cases are treated below which illustrate the significance of the quantum potential. The situation of the resonant tunneling diode will be treated separately where the significance of the quantum potential is most apparent.

The first case of interest is that of a single barrier of modest height, 42.7 mev. This value of barrier height is the same value as that of the step potential of figures (3) and (4) where the asymptotic values of density differed by an order of magnitude. For the case illustrated in figure 5, we again consider a non self-consistent calculation, with a reference density of 10^{24} /m^3, T=0K, Fermi statistics and a device length of 200nm. For the situation when a *very wide* 42.7 barrier, 100nm width and centrally placed, is considered it is found that the asymptotic value of density within a central 80 nm region is equal to 10^{23} /m^3, a result expected from the earlier discussion. There was additionally the structure in density at the potential discontinuity that was seen in figure 3.

When a *narrow* 10nm wide barrier is considered the results are quantitatively different. There is a local peak away from the barrier, but the minimum value of density exceeds that associated with the wider barrier. Of interest, however, is the structure of the quantum potential, shown in figure 5. First we note that the magnitudes of $Q(x)$ and $V(x)$ are approximately the same within the barrier region. The quantum potential is negative within the barrier, a consequence of a positive value of curvature for the density within the barrier (the density reaches a minimum at x=100nm). The quantum potential is positive in the regions immediately upstream and downstream of the barrier, where the curvature of the density is negative. The signs of the quantum potential are consistent with a density that is below its classical value immediately outside the barrier, and above its classical value within the barrier region.

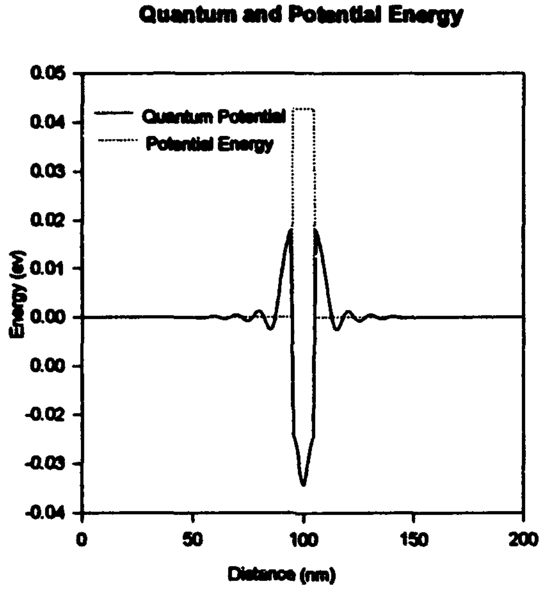


Figure 5 Quantum potential (solid curve) and V(x) (dotted) for a single barrier 10nm wide.

The next case of interest, which again offers the quantum potential as a vehicle for interpretation is the familiar self-consistent charge distribution associated with a wide bandgap/narrow bandgap structure. Figures 6 through 8 illustrates results using the density matrix for a room temperature self-consistent calculation. Here the device length is 200nm where for 0<x<100nm, $N_D=10^{23}/m^3$, for 100nm<x<200nm, $N_D=10^{24}/m^3$. A barrier of 300 mev is imposed. While the non self-consistent calculations of figure 3 show a reduction in charge density within and near the edge of the barrier, there is nothing in figure 3 resembling the extent of the charge reduction seen in figure 6. The contributions to this change are several-fold. *First* the barrier of figure 6 is an order of magnitude higher than that of figure 3. *Second*, the applied potential energy difference across the structure is chosen to yield *flat band conditions*, and thus equal to the height of the barrier plus the built-in potential. *Third*, the self-consistent potential displays structure. What is the origin of this structure?

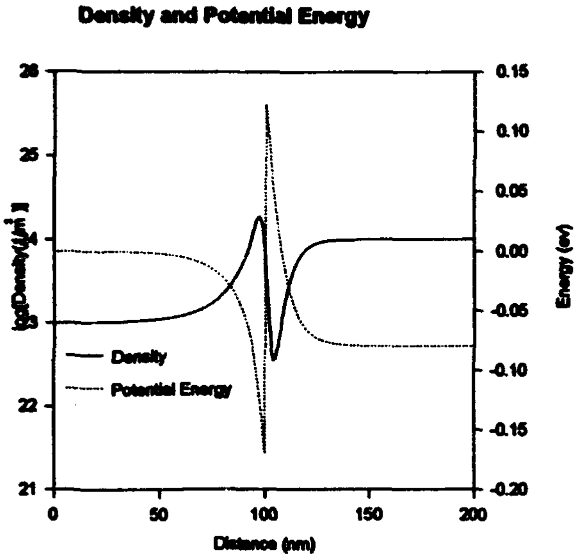


Figure 6. Self-consistent calculation of the density and potential energy for a 300 mev heterostructure diode at T=300K, with Fermi statistics and flat band conditions

In all of the calculations with a heterostructure barrier, once we pass the peak density, there is a progressive decrease in density until a minimum value of density is reached within the interior of the heterobarrier. The simple explanation based upon the quantum potential indicates, from equation (6) that the net force, under zero current conditions is zero. But *the quantum mechanical self-force, generated by variations in the single particle density (from the quantum potential as seen in figure 7) is always nonzero.* Here as we move into the wide band gap region where the density is decreasing and approaching a minimum value, the curvature of the density is positive, resulting in a negative value for the quantum potential. Since there is a minimum value of the density within the wide band gap region, there is structure to the quantum potential leading to a spatially dependent driving force. This force must be balanced by variations in the self consistent potential as seen in figure 7. The self-consistent potential which is driven by Poisson's equation is now subject to the additional constraint imposed by the quantum potential The details are not governed by equation (6), rather they are governed by the Liouville equation; but the qualitative features are represented by equation (6). When examining the classical situation we note that the potential energy, is also constrained by a *diffusive contribution.* Diffusive contributions are also present when quantum transport is considered. The quantum potential contribution is an additional contribution that is not dependent upon the presence of diffusion.

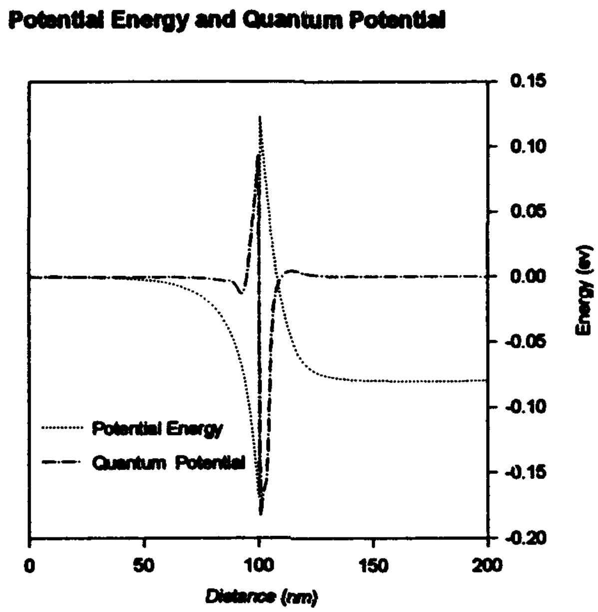


Figure 7. Self-consistent calculation of the quantum potential and potential energy for a 300 mev heterostructure diode at T=300K, with Fermi statistics and flat band conditions

There are several interesting additional points concerning the structure of the charge distribution associated with the calculations of figures (6) and (7). A good approximation to the curvature of the potential energy within the wide band gap region and near the interface, is to assume that the region is free of mobile carriers, whereby $\nabla^2 V(x) = \left(e^2 / \varepsilon\right)\rho_0(x)$. As a consequence, the higher the heterobarrier, the larger the width of the depletion zone on the wide band gap side of the structure. Under flat band conditions where the net charge distribution is zero there is a corresponding increase in charge on the narrow band gap side, and this accumulated charge will increase with increasing barrier height. Thus unlike the non-self consistent calculation of figure (3) there is significant charge accumulation on the narrow band side of the structure. The quantum potential which is negative on the wide band gap side and therefore yields a larger than classical result for the particle density, also has the effect of yielding a lower than classical result for the density just outside of the barrier. The small region of negative quantum potential to the left of the barrier is a consequence of the quantum potential defined in terms of the square root in density. An expansion of the quantum potential leads to contributions from the square of the first derivative of density as well as the second derivative.

What is the situation with multiple barrier structures; the simplest being the double barrier resonant tunneling structure. The characteristic feature of the multiple barrier structures is the existence of quasi-bound states within between the barriers. The density between the barriers depends upon the barrier height, barrier configuration, doping, etc. As discussed earlier[28], the value of the quantum potential within the quantum well of a double barrier structure is approximately equal to the energy of the lowest quasi bound state, relative to the bottom of the conduction band. We note that in terms of the definition of the quantum potential, under steady state, zero current conditions, it is direct to show from Schrodinger's equation that $Q(x)+V(x)=E$, where E is the energy of the quasi-bound state (see also reference 20). We illustrate the quantum potential for a 200nm structure double barrier structure. There are two barriers 5nm wide, each 300 mev high, separated by 5 nm, placed in the center of the structure. The background doping is $10^{24}/m^3$ and uniform, except in the interior 40nm region where it is reduced to $10^{22}/m^3$. Figures 8 and 9 show, respectively the density and donor distribution, the quantum potential and the self consistent potential energy.

With respect to figures 8 and 9, we note that carriers in excess of $4x10^{23}/m^3$ reside within the quantum well. The quantum potential is negative within the barriers of the structure corresponding to the curvature of the density, and is positive within the quantum well. But the remarkable feature is that the quantum potential is approximately constant within the quantum well. We have found that for the 300mev barrier, the quantum potential within the well is approximately 84mev (for a 200mev barrier the quantum potential within the well is approximately 70mev). A key feature in utilizing the density matrix in the coordinate representation. is that the quantum potential behaves like a quasi-bound state.

[28] H. L. Grubin, J. P. Kreskovsky, T. R. Govindan and D. K. Ferry, *Semi. Sci Technolog.* (1994)

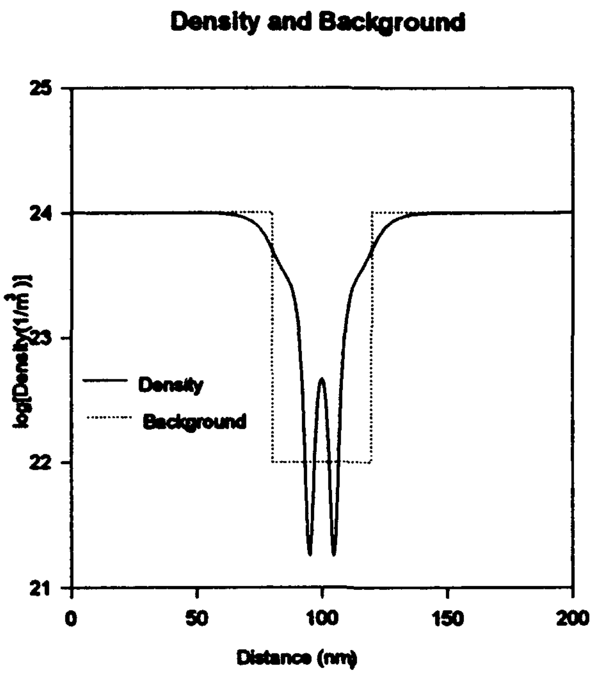


Figure 8. Self-consistent T=300K calculation with Fermi statistics showing the density and donor distribution for a symmetric double barrier structure.

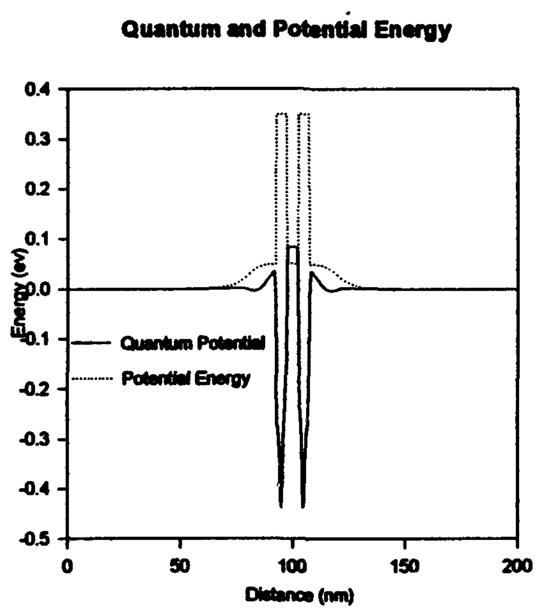


Figure 9. For figure 8, the quantum and potential energy distribution.

Further evidence for use of the quantum potential within the well as a measure of the energy of the quasi-bound state was provided by supplemental with calculations in which the double barrier structure was placed within a 40nm wide quantum well. The depth of the quantum well was varied. As the depth increased the quantum potential between the barriers remained independent of position, but increased slightly in value. The situation when the quantum well was 150mev deep, resulted in a value of the quantum potential between the barriers that increased to 94mev. The detailed results are different than that of figures 8 and 9 in that the density between the barriers has increased[29].

6. DISSIPATION AND CALCULATION OF CURRENT

The calculations of the density and potential profiles for the barrier structures in both non self consistent and self consistent studies indicate that for distances sufficiently far from the interface the results are the same as that expected using the dissipationless Boltzmann (or Vaslov) equation. When current flows, classical device transport studies usually proceed via the drift and diffusion or hydrodynamic equations, or through *solutions* to the Boltzmann transport equation and Monte Carlo procedures. Here, for cases where the ends of the device are heavily doped N^+ regions, boundary conditions on the numerical procedures are invoked to assure that the numbers of particles leaving and entering the structure are the same. An alternative approach that should yield the same results with respect to charge and potential energy distributions at the boundaries, is to implement procedures recognizing that dissipation at the beginning and ends of the structure may be represented by carriers that thermalize to a local equilibrium. The issue then is how is to deal with this situation. To date, very approximate methods have been introduced, and a rational for this approach is discussed below, but it is emphasized that *some procedure for dissipation must be invoked if transport in devices is to be discussed sensibly.*

One of the most succinct way to express the problem of dissipation follows that of Caldeira and Leggett[30]. We consider a system A (the device) interacting

[29] This increase in density has at least two origins: (i) the increased density on either side of the barriers, and (ii) the lowering of the quasi-bound state relative to the Fermi energy of the entering carriers.

[30] A. O. Caldeira and A. J. Leggett, *Physica*, **121A**, 587 (1983).

with a second system B (the reservoir) described by the Hamiltonian $H_T = H_{Device} + H_{Reservoir} + H_{Interaction}$ incorporating the reservoir and the Hamiltonian describing the interaction between the two systems. The breakup between the device and reservoir is problem dependent. If upper case letters denote the coordinates of the reservoir and lower case letters the coordinates of the system of interest (e.g., the electron system) then the quantity interest is the density matrix $<\mathbf{xR}|e^{-iHt/\hbar}\rho_{op}(0)e^{-iHt/\hbar}|\mathbf{x'R'}>$. This quantity describes the behavior of the entire system. We do not need detailed information about the reservoir; rather we need to determine its influence on, in our case, the electron system, which implies: $\rho(\mathbf{x},\mathbf{x}',t) = \int d\mathbf{R}d\mathbf{R}' <\mathbf{xR}|e^{-iHt/\hbar}\rho_{op}(0)e^{-iHt/\hbar}|\mathbf{x'R'}>$.

One method that has been invoked to deal with dissipation and boundaries and current flow in devices, has been guided by perturbation theory on the density matrix [31]. First the equation of motion of the density matrix has been rewritten to include a scattering contribution, as shown by equation (11). Below we concentrate on the modifications of the Liouville equation through the incorporation of scattering and deal only with *the Liouville equivalent of classical scattering.*

In the Boltzmann picture, ignoring Fermi statistics, the scattering rate is:

$$(21)\qquad \left[\frac{\partial f_w(\mathbf{k},\mathbf{x})}{\partial t}\right]_{dis} \equiv -\frac{2}{(8\pi^3)}\int d\mathbf{k}'\{f_w(\mathbf{k},\mathbf{x})W(\mathbf{x},\mathbf{k}',\mathbf{k}) - f_w(\mathbf{k}',\mathbf{x})W(\mathbf{x},\mathbf{k},\mathbf{k}')\}$$

where the subscript 'w' denotes a Wigner function and $W(\mathbf{x},\mathbf{k},\mathbf{k}')$ represents the standard transition probability per unit time. Utilizing the Weyl transformation:

$$(22)\qquad \rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s}) = \frac{2}{8\pi^3}\int d\mathbf{k}\, f_w(\mathbf{k},\mathbf{r})\exp i2\mathbf{s}\cdot\mathbf{k}$$

with the following change in coordinates: $\mathbf{x}+\mathbf{x}' = 2\mathbf{r}$, $\mathbf{x}-\mathbf{x}' = 2\mathbf{s}$, the scattering rate of density matrix (after manipulation of the variables of integration) is given by:

[31] H. L. Grubin, T. R. Govindan and M. A. Stroscio, *Semi. Sci. Technolog.* (1994).

$$(23)\quad \left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv$$

$$-\left[\frac{2}{8\pi^3}\right]^2 \int d\mathbf{k}d\mathbf{k}'\{f_w(\mathbf{k},\mathbf{r})[\exp i2\mathbf{s}\cdot\mathbf{k}]W(\mathbf{r},\mathbf{k}',\mathbf{k})\}\{1-\exp i2\mathbf{s}\cdot(\mathbf{k}'-\mathbf{k})\}$$

The structure of the scattering term within the framework of classical Boltzmann scattering expressed within the coordinate representation is obtained from equation (23). For example the second exponential term in equation (23) can be expressed as an infinite series, in which case the scattering term is expressed as an infinite series in powers of $\mathbf{s}$. The lead term is given by:

$$(24)\quad \left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv$$

$$i2\mathbf{s}\cdot\left[\left[\frac{2}{8\pi^3}\right]^2 \int d\mathbf{k}\, f_w(\mathbf{k},\mathbf{r})[\exp i2\mathbf{s}\cdot\mathbf{k}]\int d\mathbf{k}'(\mathbf{k}'-\mathbf{k})W(\mathbf{r},\mathbf{k}',\mathbf{k})\right]$$

Standard classical theory[32] teaches that:

$$(25)\quad \frac{2}{(8\pi^3)}\int d\mathbf{k}'(\mathbf{k}'-\mathbf{k})W(\mathbf{r},\mathbf{k}',\mathbf{k}) \equiv -\mathbf{k}\Gamma(\mathbf{r},|\mathbf{k}|)$$

where $\Gamma(\mathbf{r},|\mathbf{k}|)$ represents a scattering rate. Thus:

$$(26)\quad \left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv -i2\mathbf{s}\cdot\left[\frac{2}{8\pi^3}\int d\mathbf{k}\, f_w(\mathbf{k},\mathbf{r})[\exp i2\mathbf{s}\cdot\mathbf{k}]\mathbf{k}\Gamma(\mathbf{r},|\mathbf{k}|)\right]$$

which, using the inverse of the Weyl transformation:

$$(27)\quad 2f_w(\mathbf{k},\mathbf{r}) = 2^3\int d\mathbf{s}\,\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})\exp -i2\mathbf{s}\cdot\mathbf{k}$$

can be rearranged as

[32] D. K. Ferry, *Semiconductors*, Macmillan Publishing Company, NY, (1991).

$$\left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv \tag{28}$$
$$-i2\mathbf{s}\cdot\left[\frac{2^3}{8\pi^3}\int d\mathbf{k}\,d\mathbf{s}'\,\rho(\mathbf{r}+\mathbf{s}',\mathbf{r}-\mathbf{s}')[\exp i2(\mathbf{s}-\mathbf{s}')\cdot\mathbf{k}]\mathbf{k}\Gamma(\mathbf{r},|\mathbf{k}|)\right]$$

A significant simplification arises when the crystal momentum in equation (28) is replaced by a divergence of the correlation vector:

$$\left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv \tag{29}$$
$$-i\nabla_{\mathbf{s}}\cdot\left[\frac{2^3}{8\pi^3}\int d\mathbf{k}\,d\mathbf{s}'\,\rho(\mathbf{r}+\mathbf{s}',\mathbf{r}-\mathbf{s}')[\exp i2(\mathbf{s}-\mathbf{s}')\cdot\mathbf{k}]\Gamma(\mathbf{r},|\mathbf{k}|)\right]$$

For the case when the scattering rate is independent of momentum, the dissipation term reduces to:

$$\left[\frac{\partial\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s})}{\partial t}\right]_{scat} \equiv -\Gamma\mathbf{s}\cdot\nabla_{\mathbf{s}}\rho(\mathbf{r}+\mathbf{s},\mathbf{r}-\mathbf{s}) \tag{30}$$

and the Liouville equation in the coordinate representation is modified to read:

$$i\hbar\frac{\partial\rho(\mathbf{x},\mathbf{x}',t)}{\partial t} = \tag{31}$$
$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial\mathbf{x}^2}-\frac{\partial^2}{\partial\mathbf{x}'^2}\right)\rho(\mathbf{x},\mathbf{x}',t)+\left(V(\mathbf{x})-V(\mathbf{x}')\right)\rho(\mathbf{x},\mathbf{x}',t)$$
$$-i\hbar\frac{\Gamma}{2}(\mathbf{x}-\mathbf{x}')\cdot(\nabla_{\mathbf{x}}-\nabla_{\mathbf{x}'})\rho(\mathbf{x},\mathbf{x}')$$

The additional contribution due to dissipation was discussed in reference [27] and in a study by Dekker[33]. Density matrix algorithms incorporating the dissipation contributions of equation (31) have been implemented with some results reported[32]. But because of numerical difficulties at higher bias levels modifications to the scattering were introduced whose consequences go beyond

[33] H. Dekker, *Phys. Rev.* A16, 2126 (1977)

the approximations associated with the expansion of equation (24). It is worthwhile dwelling on these modifications.

In modifying the scattering term in equation (31) it was first recognized that the dissipation term could be reexpressed in terms of a velocity density matrix:

$$\text{(32)} \quad \mathbf{j}(\mathbf{x},\mathbf{x}') = \left(\frac{\hbar}{2mi}\right)(\nabla_{\mathbf{x}} - \nabla_{\mathbf{x}'})\rho(\mathbf{x},\mathbf{x}')$$

The diagonal elements of equation (32) yield the velocity flux density. In terms of $\mathbf{j}(\mathbf{x},\mathbf{x}')$ equation (30) becomes:

$$\text{(33)} \quad \begin{aligned} & i\hbar\frac{\partial\rho(\mathbf{x},\mathbf{x}',t)}{\partial t} = \\ & -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial\mathbf{x}^2} - \frac{\partial^2}{\partial\mathbf{x}'^2}\right)\rho(\mathbf{x},\mathbf{x}',t) + (V(\mathbf{x}) - V(\mathbf{x}'))\rho(\mathbf{x},\mathbf{x}',t) + m\Gamma(\mathbf{x}-\mathbf{x}')\cdot\mathbf{j}(\mathbf{x},\mathbf{x}') \end{aligned}$$

The scattering term in the above equation was then written in the form of a scattering potential. The procedures for this were as follow. *First,* the term $\mathbf{j}(\mathbf{x},\mathbf{x}')$ was rewritten as $\mathbf{j}(\mathbf{x},\mathbf{x}') \equiv \mathbf{v}(\mathbf{x},\mathbf{x}')\rho(\mathbf{x},\mathbf{x}')$, where $\mathbf{v}(\mathbf{x}) \equiv \mathbf{v}(\mathbf{x},\mathbf{x})$ represents the expectation value of the velocity. *Second,* $\mathbf{j}(\mathbf{x},\mathbf{x}')$ was approximated as $\mathbf{j}(\mathbf{x},\mathbf{x}') \approx \mathbf{v}(\mathbf{x})\rho(\mathbf{x},\mathbf{x}')$. Higher order terms are at least second order in (x-x'), and retaining them would be inconsistent with the approximation leading to equation (24). *Third,* quasi-Fermi levels were introduced through the definition:

$$\text{(34)} \quad E_F(\mathbf{x}) - E_F(\mathbf{x}') = -\int_{\mathbf{x}'}^{\mathbf{x}} d\mathbf{x}''\cdot\mathbf{v}(\mathbf{x}'')m\Gamma(\mathbf{x}''),$$

For small values of x-x' about x, equation (34) is approximately represented by: $E_F(\mathbf{x}) - E_F(\mathbf{x}') = -(\mathbf{x}-\mathbf{x}')\cdot\mathbf{v}(\mathbf{x})m\Gamma(\mathbf{x})$. Under this approximation equation (33) becomes:

$$\text{(35)} \quad \begin{aligned} & i\hbar\frac{\partial\rho(\mathbf{x},\mathbf{x}',t)}{\partial t} = \\ & -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial\mathbf{x}^2} - \frac{\partial^2}{\partial\mathbf{x}'^2}\right)\rho(\mathbf{x},\mathbf{x}',t) + \left[(V(\mathbf{x}) - V(\mathbf{x}')) - (E_F(\mathbf{x}) - E_F(\mathbf{x}'))\right]\rho(\mathbf{x},\mathbf{x}',t) \end{aligned}$$

Thus we have taken the differential equation (31) whose right hand side is complex and replaced it by one whose right hand side is real, when the density matrix is real. Side by side calculations at low values of bias yield identical results.

While the above discussion leading to equation (35) appears to be model dependent, the results implied by this equation have greater generality than the means used to arrive at it.

The implementation of equation (35) permits us to calculate current in a direct manner. How is this done? In all of the calculations with current the assumption is that the carriers at the upstream boundary are in local equilibrium and that the distributions are either a displaced Maxwellian or a displaced Fermi-Dirac distribution. As discussed in reference [27] this implies that at the upstream boundary, the zero current quantum distribution function $\rho(\mathbf{x},\mathbf{x}')$ is replaced by $\rho(\mathbf{x},\mathbf{x}')\exp[im\mathbf{v}(boundary)\cdot(\mathbf{x}-\mathbf{x}')/\hbar]$. *Since current is introduced as a boundary condition to the problem as formulated by equation (35) a prescription is necessary for finding its value. An auxiliary condition was constructed.*

To compute a value of current for use in the Liouville equation, a *criteria* was introduced through moments of equation (35)[34]. Under time independent conditions, the momentum balance equation yields the condition:

$$(36) \qquad 2\nabla_{\mathbf{x}}E+[\nabla_{\mathbf{x}}V]\rho(\mathbf{x})-[\nabla_{\mathbf{x}}E_F]\rho(\mathbf{x})=0$$

where E is the kinetic energy and is given by the equation (13). Under the assumptions of current continuity, i.e., $\rho(\mathbf{x})\mathbf{v}(\mathbf{x})$ is independent of distance (satisfied for the Liouville equation), *and the condition that the energy of the entering and exiting carriers are* equal, equation (34) becomes:

$$(37) \qquad E_F(x)-E_F(x')=-j\int_{x'}^{x}dx''m\Gamma(x'')/\rho(x''),$$

where we have restricted the considerations to one space dimension. *The current is chosen so that $E_F(L)-E_F(0)$, is equal to the change in applied energy across the structure.*

[34] These moment equations are discussed in reference (27) and are incorporated into a later section.

We now illustrate some of the above considerations. The simplest type of calculation to deal with is that of a free particle. For this case and with current introduced as a boundary condition the density matrix is complex. The real part is symmetric and the imaginary part (from which current is obtained) is asymmetric about the diagonal. The calculation displayed in figure (10) shows the real part and figure (11) the imaginary part for a 200nm with a doping of $10^{23}/m^3$, subject to a bias of 10mev. For this calculation and parameters appropriate to GaAs, a scattering rate of 10^{-13}sec, yields a mobility of $0.258 m^2/v$-sec. The mean carrier velocity for this calculation is approximately $1.3x10^4$m/sec.

Real Part of the Density Matrix

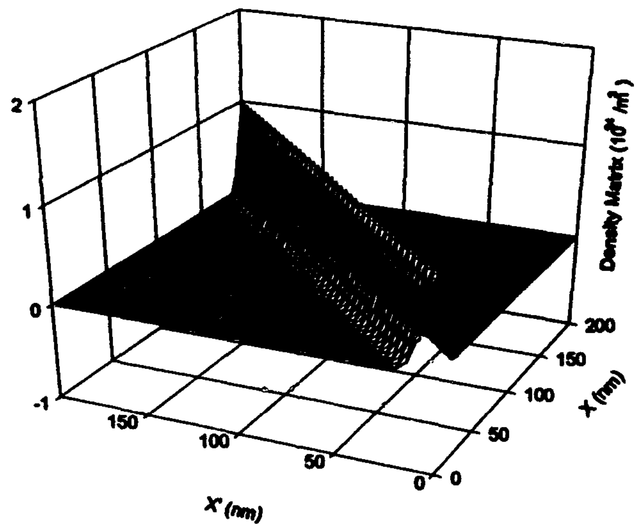


Figure 10. Real part of the density matrix for a free particle subject to a constant force.

Increasing the applied bias results in an increase in the carrier velocity and an increase in the kinetic energy of the carriers. This increase affects the curvature of the density matrix in the correlation direction and is displayed in figure 12.

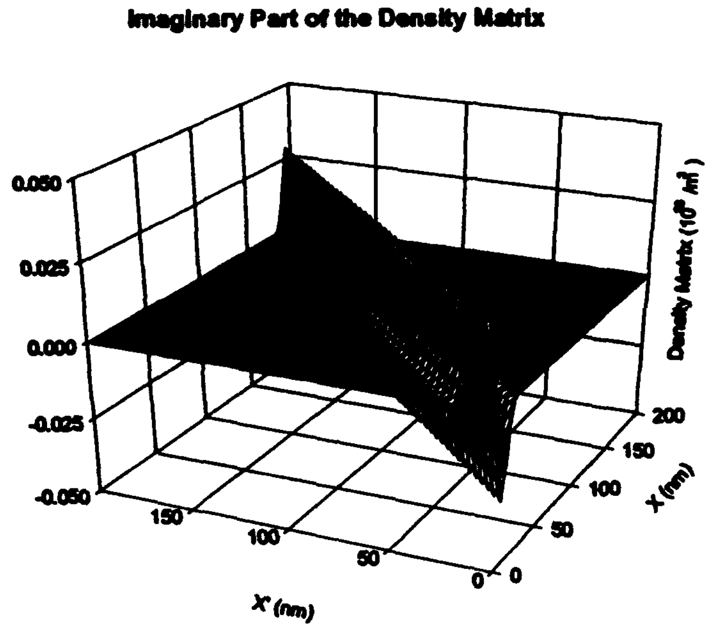


Figure 11. Imaginary part of the density matrix corresponding to figure 10.

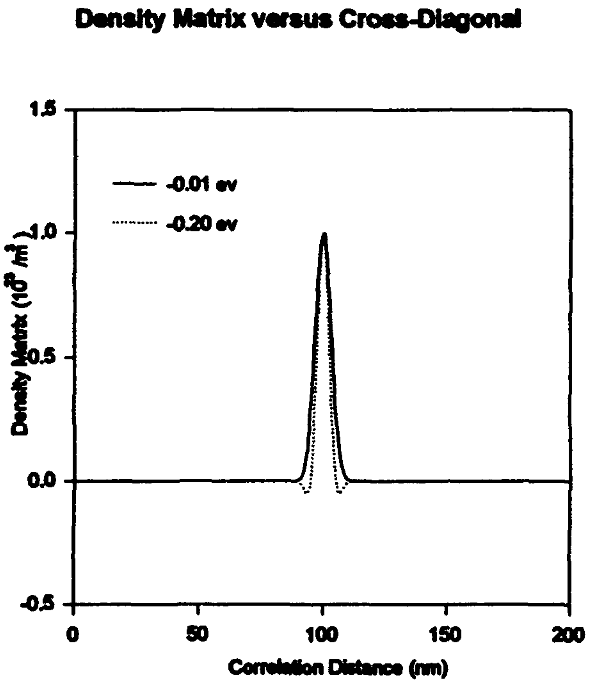


Figure 12. Density matrix versus correlation distance when current is flowing. Dashed line is for a bias of 10 mev and a mean velocity of $1.3x1^{04}$ m/sec; solid line is for a bias 200 mev and a mean velocity of $2.6x10^5$ m/sec.

All semiconductor devices sustain energy dependent scattering, implying that the scattering rate within one region of the structure will be different than at a different region of the structure. To understand how this is implemented in the density matrix algorithm several illustrative examples of nonuniform scattering were performed. These examples deal with the generation of nonuniform fields from variation in the *mobility* (vis scattering). We will treat an element with material parameters nominally the same as those associated with figures 10-12. However, here we vary the scattering rate within the central 5nm of the structure. On the basis of the definition of the quasi-Fermi energy, a decrease in the scattering time, which results in a decrease in mobility, will yield a sharp drop in the quasi-Fermi level. The density cannot change as rapidly, but is constrained by the Debye length and so results in a more gradual change in the self consistent potential energy. The quasi Fermi energy and potential energy as well as the density are displayed in figure 13 for a bias of 10 mev, where the scattering time within the central 5nm was 10^{-14}sec, while that at the boundaries are respectively 10^{-12}sec. There are several points to emphasize. For the calculation of figure 13 the quasi-Fermi energy varies in an approximately linear manner in three separate regions. In particular within the exterior cladding regions the quasi-Fermi level is equal to the potential energy distribution where it assures the presence of local charge neutrality. The departure of the potential energy from the quasi-Fermi energy for this calculation is in large part a consequence of Debye length considerations. The quasi-Fermi energy which is an integral expression follows the same slope, to the interior region, where the precipitous change in value is consequence of the reduction in the scattering time.

Figure 14 displays the scattering rate used in the calculations and the self-consistent density distribution. Of extreme significance here is the formation of a local dipole layer within the interior of the structure.

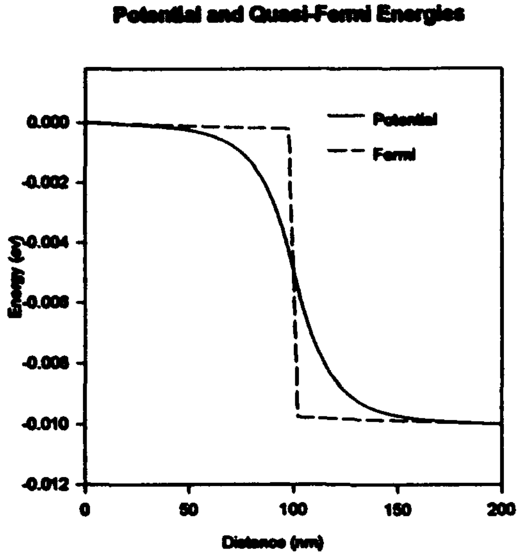


Figure 13. Self-consistent calculations of the potential energy and quasi-Fermi energy for a uniform doped structure with a variable scattering rate within the center of the structure.

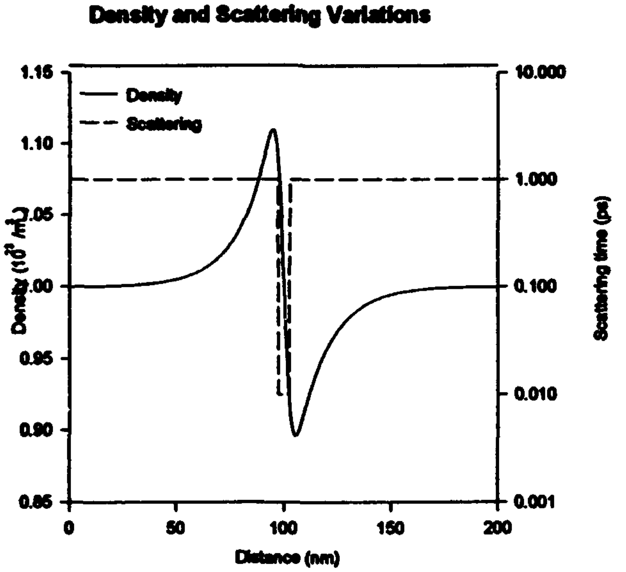


Figure 14. Self-consistent calculation of the density for a uniform doped structure with the displayed variable scattering rate..

7. SINGLE BARRIER DIODE: CONSTANT SCATTERING RATE

The quasi-Fermi scattering model has been applied to a variety of structures including single and multiple barrier diodes as well as electron-hole transport. We illustrate single barrier calculations in figures 15 through 18 for a structure with a constant scattering rate. Preliminary results for this type of structure were presented earlier[35]. The calculations are for a 200nm structure containing a single 300 mev high, 20nm wide barrier embedded within a 30nm N^- region, surrounded by uniformly doped $10^{24}/m^3$ material. The scattering time τ is constant and equal to 10^{-13} sec. The calculations are self-consistent and assume Fermi-Dirac boundary conditions. The first three figures, 15 through 17 show potential energy, density, and quasi-Fermi energy distributions, respectively, for different bias levels.

From figure 15 as the collector boundary is made more negative with respect to the emitter, a local 'notch' potential well forms on the emitter side of the barrier. The potential energy decreases linearly across the barrier, signifying negligible charge within the barrier, followed by a broad region where the potential energy decreases to its value at the collector boundary.

The charge distribution, figure 16, displays a buildup of charge on the emitter side of the barrier, and a compensatory region of charge depletion on the collector side of the barrier. At a bias of 400 mev significant charge accumulation has formed on the emitter side of the barrier, followed by a broad region of charge depletion on the collector side. Note that as the bias increases there is a progressive increase in charge within the interior of the barrier. Both results are consistent with the low temperature experimental findings of Eaves et al[36].

Figure 17 displays the quasi-Fermi energy (relative to the equilibrium Fermi energy). Because of the low values of current E_F, is approximately zero from the emitter to within the first half of the barrier[37] and then drops to a value approximately equal to the bias through the remaining part of the structure.

---

[35] D. K. Ferry and H. L. Grubin, *Proceedings of the International Workshop on Computational Electronics*, Univ. of Leeds 247, (1993)

[36] L. Eaves. F. W. Sheard, and G. A. Toombs, *Physics of Quantum Electron Devices* (ed. F. Capasso), 107 (1990) Springer -Verlag , Berlin.

[37] In the emitter region the variation in E_F matches that of $V(x)$, and insures that ρx is constant in the vicinity of the emitter boundary.

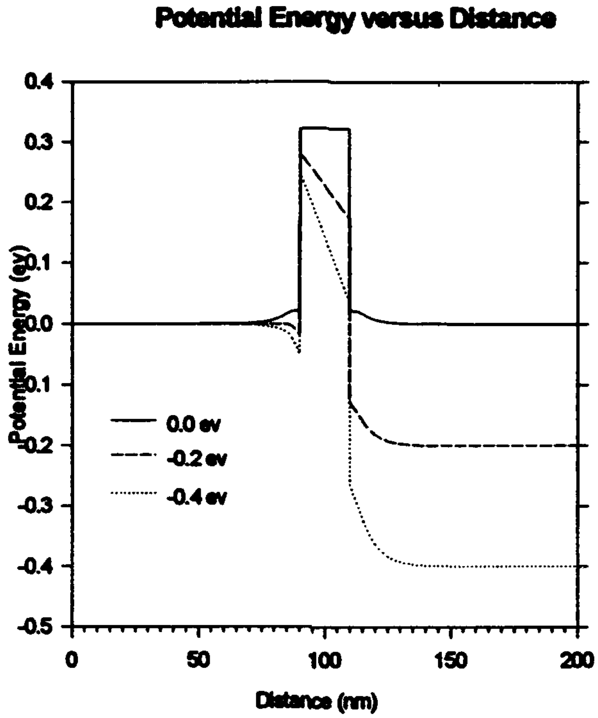


Figure 15. Self consistent room temperature potential energy calculations assuming Fermi-Dirac boundary conditions for a single barrier structure under varying bias conditions.

I versus V for the 20nm barrier is shown in figure 18. Note that for a broad range of voltage the current depends exponentially on voltage; but there is distinct sublinearity to the curve. In words, the sublinearity indicates that at a given value of voltage the current is lower than expected on the basis of a pure exponential relation. In seeking an origin of this sublinearity we note from the accompanying voltage distributions that not all of the voltage falls across the tunnel barrier; indeed a substantial contribution falls across the region immediately adjacent to the collector side of the barrier.

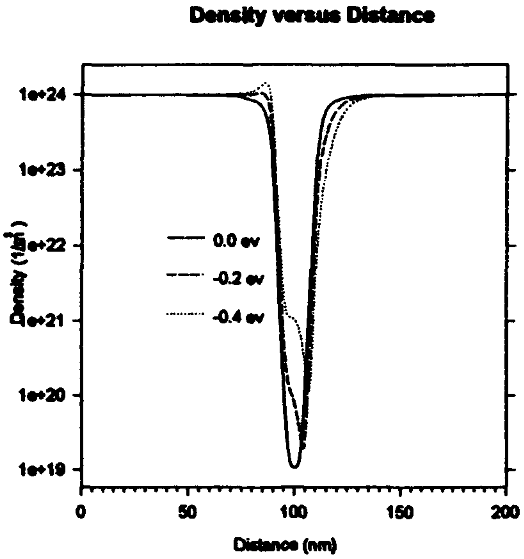


Figure 16. Self consistent density calculations for figure 15.

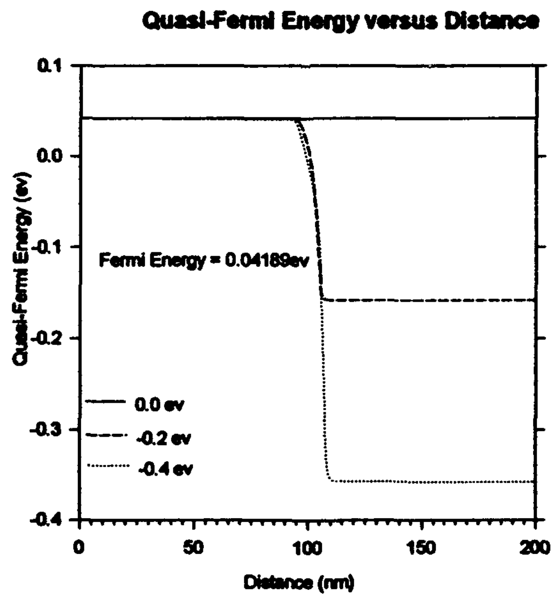


Figure 17 Quasi-Fermi energy distribution for the calculations of figure 15.

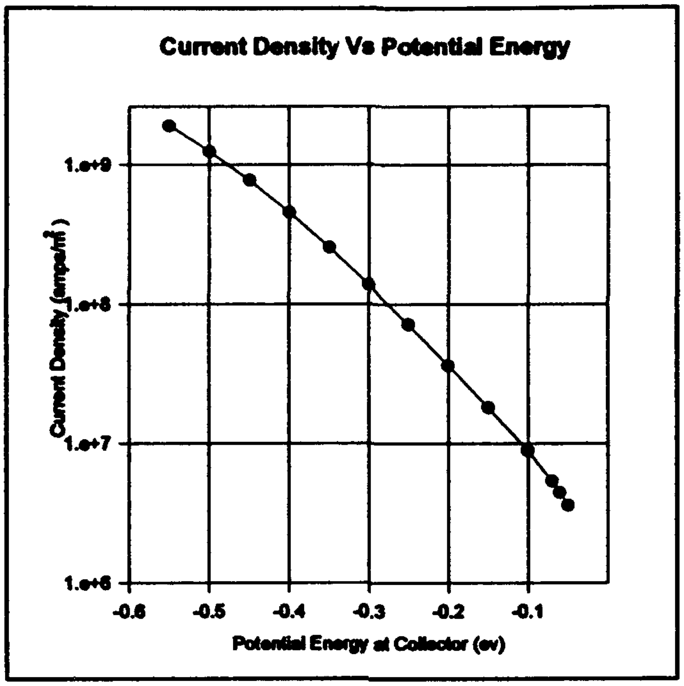


Figure 18. Current-voltage relation for the calculations of figures 15 through 17.

8. RESONANT TUNNEL DIODE; VARIABLE SCATTERING RATE.

To illustrate the calculation for resonant tunneling structures we treat a 200nm structure, with two 5 nm - 300 mev barriers separated by a 5nm well. The structure has a nominal doping of $10^{24}/m^3$ except for a central 50nm wide region where the doping is reduced to $10^{22}/m^3$. The effective mass is constant and equal to that of GaAs ($0.067m_0$); Fermi statistics are imposed; the ambient is 77K; and current is imposed through the density matrix equivalent of a displaced distribution at the boundaries. In these computations only one set of scattering rates was used, *although scattering was increased in the vicinity of the double barriers.*

The signature of the RTD is it's current-voltage relation with the region of negative differential conductivity; for the structure considered this is displayed in figure 19. The current is numerically negligible until a bias of approximately 50 mev, with the peak current occurring at 260 mev, followed by a sharp but modest drop in current at 270 mev. The interpretation of these results is assisted by figures (20) and (21) and the Bohm quantum potential. As indicated earlier, we have found, through an extensive number of numerical simulations, that the value of *V(x)+Q(x)*, between the barriers of an RTD is a measure of the position of the quasi-bound state.

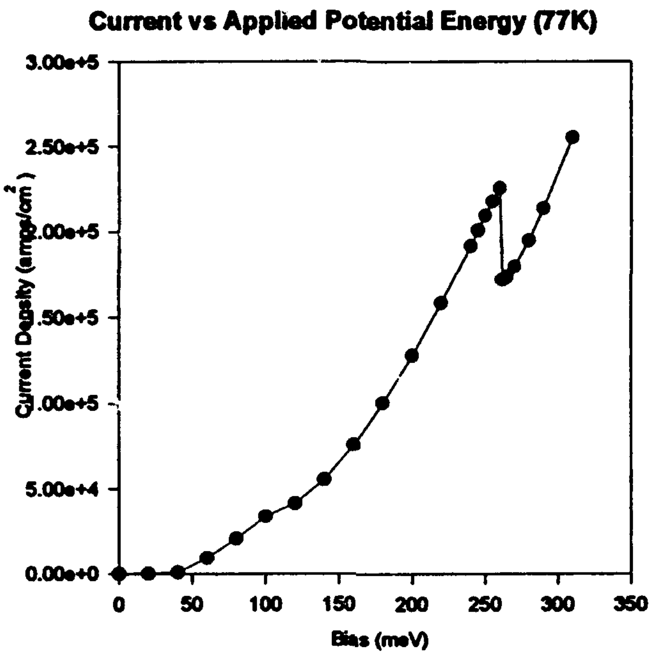


Figure 19. Current versus (magnitude) voltage for the resonant tunneling structure.

Consider figure 20 which displays the equilibrium self-consistent potential for the RTD. Also shown is the value of the equilibrium Fermi energy (approximately 54 mev) and the values, at five different values of applied potential energy, of *V(x)+Q(x)* within the quantum well. At 100 mev the quasi-bound state is approximately equal to the equilibrium Fermi energy and significant current begins to flow. The current continues to increase until the bias equals 260 mev, where there is a sudden drop in current.

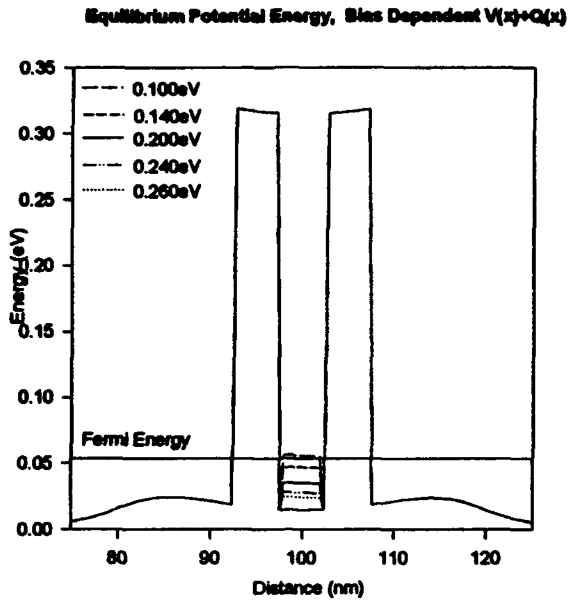


Figure 20. Equilibrium potential energy and the bias dependence of V(x)+Q(x) within the quantum well. Legend denotes collector bias.

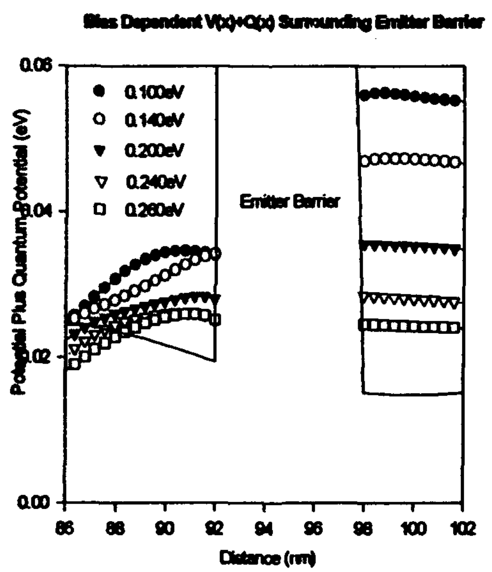


Figure 21. Blow up of figure 20.

To see what is happening we blow up the region on either side of the emitter barrier, where we display values of *V(x)+Q(x)* before the emitter barrier and within the quantum well (figure 21). Within the quantum well we see the quasi bound state decreasing as the bias on the collector is increasing. In the region prior to the emitter barrier where a 'notch' potential forms signifying charge accumulation, we see the formation with increased bias of a region where $V(x)+Q(x)$ is relatively flat. Of significance here is that for values of bias associated with the initial current increase the value of $V(x)+Q(x)$ within the quantum well is greater then its value in the emitter region. The current reaches a maximum at the cross-over where $V(x)+Q(x)$ in the emitter region and in the quantum well are approximately equal. (Implementation of an earlier algorithm, generally resulted in solutions oscillating between high and low values of current when this condition was reached). While it is tempting to associate $V(x)+Q(x)$ within the emitter region with a quasi-bound state, this may be premature.

The distribution of potential energy $V(x)$ as a function of bias is displayed in figure 22, where the notch potential is deepened with increasing bias, signifying increased charge accumulation. This is accompanied by a smaller share of the potential drop across the emitter barrier, relative to the collector barrier region. Comparing the slopes of the voltage drop across the emitter and collector barriers, we see larger fractions of potential energy fall across the collector barrier.

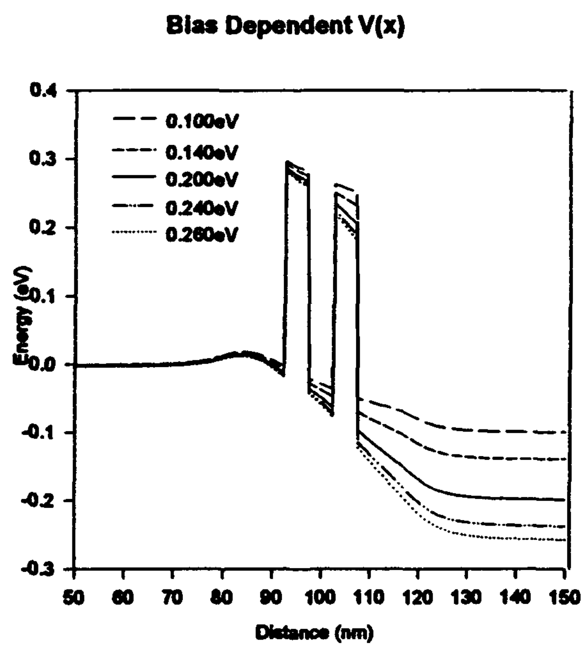


Figure 22. Distribution of potential energy as a function of applied bias.

Explicit in this calculation is dissipation which is incorporated through the quasi-Fermi level. Within the vicinity of the boundaries the quasi-Fermi level is parallel to the conduction band edge. Indeed, for this calculation the quasi-Fermi level departs from the conduction band edge only within the vicinity of the barriers. The quasi-Fermi level is displayed in figure 23 at a bias of 260 mev, where we see that the quasi-Fermi level is relatively flat until the middle of the first barrier at which point there is a small drop in value followed by a flat region within the quantum well. There is a strong drop of the quasi Fermi level within the second barrier.

The charge distribution accompanying these variations in bias shows accumulation on the emitter side of the barrier along with charge accumulation within the quantum well. The increase in charge within the quantum well and adjacent to the emitter region is accompanying by charge depletion downstream of the second barrier, with the result that the net charge distribution throughout the structure is zero.

Variations in the quasi Fermi level were accompanied by variations in density and current which were all obtained in a self-consistent manner. Supplemental computations were performed in which the quasi-Fermi level was varied by altering the scattering rates. The calculations were applied to the post threshold case with values for the scattering rate chosen so to provide a large drop in current. Indeed a current drop by greater than a factor of three was obtained followed by a shallow current increase with increasing bias. The significant difference leading to these changes was the manner in which the quasi-Fermi level changed. Rather than the shallow change depicted in figure 19, there was a larger change in the quasi-Fermi level across the first barrier, a result similar to that obtained for single barriers.

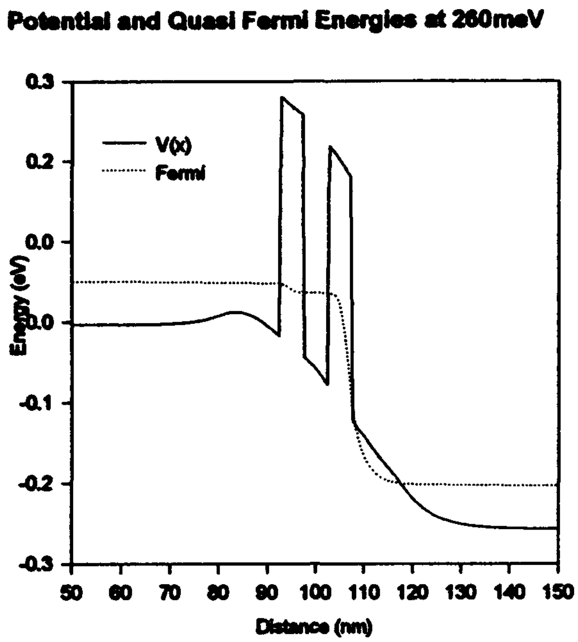


Figure 23 Potential and quasi-Fermi energy at a bias of 260 mev.

The calculations obtained for figures 19 through 23 were obtained from a new solution algorithm that was constructed for the quantum Liouville equation that permits a more convenient specification of boundary conditions, in particular when the device is under bias. The algorithm is based on a reformulation of the governing equations in which a higher order differential equation in the local direction $[(x+x')/2]$ is constructed from the quantum Liouville equation. The reformulated equation behaves like an elliptical equation in the local direction rather than the hyperbolic behavior of the quantum Liouville equation. With appropriate boundary conditions, solutions to the two forms of the quantum Liouville equations are equivalent. However the reformulated equation allows construction of a more robust algorithm that provides desired solution behavior at the contacts by boundary condition specification at both contacts.

9 THE QUANTUM HYDRODYNAMIC EQUATIONS

A detailed description of the density matrix under equilibrium and nonequilibrium conditions was given in the previous sections. Much of the work reported there was a consequence of considerable effort at understanding the nature of the quantum Liouville equation in the coordinate representation. As such many of these results were obtained at the end of the program; with some results particularly the RTD results obtained after the program was completed.

Simultaneous with this effort was attempts at determining an understanding of the quantum hydrodynamic equations, as it was felt that these equations being approximate in nature would find greater acceptance in the engineering community as a vehicle for the design of multidimensional devices. A discussion of these equations is not given here as an extensive paper along with several smaller supplement al studies have either been published or will be published. These papers form part of this final report and constitute part of this section.

Keyword	PAGE PROOFS	Please note that these first proofs have been output by a Laserjet printer. At revise, we will output through a Monotype Lasercomp, effecting a far superior image.
SST		

PLEASE RETURN THIS PROOF WITHIN 3 DAYS OF RECEIPT

(124) SST/51569/SYM

NOT YET READ IN PRODUCTION OFFICE

Semicond. Sci. Technol. ? (1994) ?–?. Printed in the UK

Tanner—SSTP.128

Uses of the quantum potential in modelling hot-carrier semiconductor devices

H L Grubin†, J P Kreskovsky†, T R Govindan† and D K Ferry‡

† Scientific Research Associates Inc., Glastonbury, CT 06033-6058, USA

‡ Center of Solid State Electronics Research, Arizona State University, Tempe, AZ 85287-6206, USA

d) ~~Received ?~~

Abstract. Through the use of examples, various forms of the quantum potential are examined for modelling and interpreting the operation of semiconductor devices.

1. Introduction

Quantum effects occur in device structures when the lateral confinement dimensions (the distance over which significant changes in density occur) are comparable to the thermal de Broglie wavelength. It is possible to model quantum structures with a quantum density from the Schrödinger equation, the Liouville equation or the Wigner equation. However, modelling of these structures has also evolved from the use of a set of classical hydrodynamic equations, which include corrections in the form of the quantum potential. The most familiar forms used involve the Bohm potential [1]

$$Q_B = -\left(\frac{\hbar^2}{2m}\right)\left(\frac{d^2\sqrt{\rho}}{\sqrt{\rho}\,dx^2}\right)$$

and the Wigner potential [2]

$$Q_W = -\left(\frac{\hbar^2}{8m}\right)\left(\frac{d^2\ln\rho}{dx^2}\right).$$

For example, using the density matrix in the coordinate representation, in which dissipation is modelled by a Fokker–Planck contribution, the hydrodynamic equations are of the form [3]

$$\frac{\partial\rho}{\partial t}+\left(\frac{1}{m}\right)\frac{\partial\rho p_d}{\partial x}=0 \qquad (1a)$$

$$\frac{\partial\rho p_d}{\partial t}+\frac{2\partial E(x)}{\partial x}+\frac{\rho\partial V(x)}{\partial x}+\frac{\rho p_d}{\tau}=0 \qquad (1b)$$

$$\frac{\partial E}{\partial t}+\left(\frac{1}{2m^2}\right)\frac{\partial P^{(3)}(x)}{\partial x}+\left(\frac{\rho p_d}{m}\right)\frac{\partial V(x)}{\partial x}+\frac{2E}{\tau}-\frac{\rho\Xi}{m}=0 \qquad (1c)$$

where p_d is the mean momentum, $E(x)$ is the mean kinetic energy and $P^{(3)}$ represents the energy flux [3] (which in the classical case represents the transport of energy). The Wigner potential is usually interpreted as a quantum correction to $E(x)$. The term Ξ/m represents an equilibrium energy term to which the system relaxes [3]. To place equations (1) in a familiar form requires additional manipulation, and one common form is [4]

$$\frac{\partial(\rho p_d)}{\partial t}+\left(\frac{1}{m}\right)\frac{\partial(\rho p_d^2)}{\partial x}+\frac{\partial(\rho k_B T)}{\partial x}+\frac{\rho\partial(Q_B(x)/3)}{\partial x}+\frac{\rho\partial V(x)}{\partial x}+\frac{\rho p_d}{\tau}=0 \qquad (2a)$$

$$\frac{\partial E}{\partial t}+\left(\frac{1}{m}\right)\frac{\partial(p_d(E(x)+\rho k_B T))}{\partial x}+\left(\frac{\rho p_d}{m}\right)\frac{\partial((Q_B/3)+V(x))}{\partial x}-\left(\frac{\rho\lambda^2 k_B T}{6m}\right)\left(\frac{\partial^2(\ln\rho)}{\partial x^2}\right)\frac{\partial p_d}{\partial x}+\frac{2(E-E_0)}{\tau}=0 \qquad (2b)$$

lines are.

where λ is the thermal de Broglie wavelength and E_0 is an equilibrium energy. In the above equations, the quantum potential has been reduced by a factor of three, although this reduction is a subject of some debate. We have taken this view, with others [5], that this is an adjustable parameter, calculations below illustrate the effect of varying this parameter.

It is possible to determine the validity of the above equations for quantum devices by performing comparable calculations with one of the full quantum transport equations. In addition, it is useful to ask whether the quantum potential aids in the interpretation of results. The full quantum treatment used here involves the density matrix (in the coordinate representation) [6]. The latter calculations include the appropriate Fermi or Boltzmann statistics, although the moment equations discussed below involve only Boltzmann statistics. When a comparison between the quantum hydrodynamic equations and the density matrix is made, Boltzmann statistics is assumed.

Keyword	PAGE PROOFS	*Please note that these first proofs have been output by a Laserjet printer. At revise, we will output through a Monotype Lasercomp, effecting a far superior image.*
SST		

2. Resonant tunnel diode under equilibrium

The first example is that of a GaAs/AlGaAs double-barrier (5 nm barriers and 5 nm well) resonant tunnelling structure, with a nominal doping of 10^{18} cm^{-3}. This structure has a quasi-bound state of 85 meV. It is anticipated that within the quantum well Q_B would determine the quasi-bound state. Figure 1 displays a blow-up of the region surrounding the double-barrier structure. In figure 1(*a*), the self-consistent potential energy, $V(x)$, the Bohm and Wigner quantum potentials are compared. It is seen that the Bohm potential is constant within the well at the quasi-bound state value. There is a small accumulation of carriers whose value peaks in the centre of the well. Figure 1(*b*) replaces $V(x)$ with the mean kinetic energy per particle for this comparison. It is seen that the best representation of this energy is with the Wigner potential. The closeness of the results suggests that the Wigner potential is important for the structure of the energy within the well.

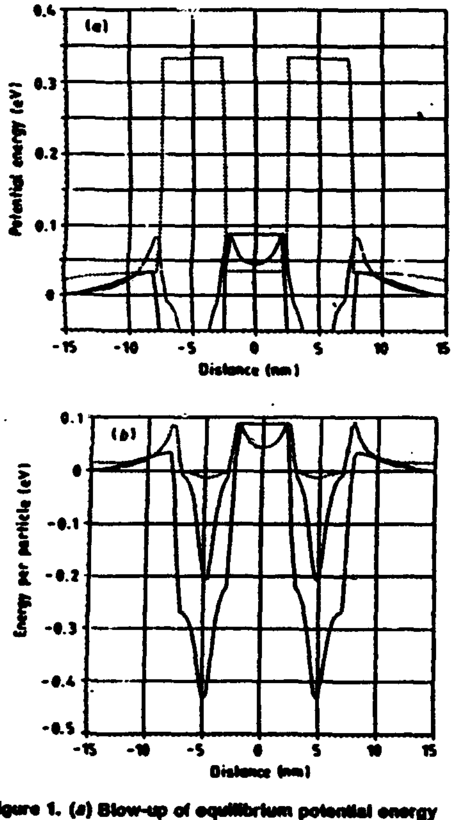


Figure 1. (*a*) Blow-up of equilibrium potential energy (broken curve), Bohm potential (full curve) and Wigner potential (dotted curve) within and surrounding a resonant tunnelling structure. (*b*) Blow-up of equilibrium energy per particle (broken curve), Bohm potential (full curve) and Wigner potential (dotted curve) within and surrounding a resonant tunnelling structure.

3. MODFET calculation

The next example is a comparative calculation along a line perpendicular to the conduction channel of a MODFET. For this device (figure 2) the GaAs region (left-hand side of the structure) is 100 nm long (doped to 10^{14} cm^{-3}) and is adjacent to a wide-bandgap (300 meV) region doped to 10^{18} cm^{-3}. The first 10 nm of the wide-bandgap region is undoped. The hydrodynamic equations were solved using the Bohm potential with three different values for the multiplicative constants: 3 (as in equations (2)), 1 and 9. Figure 2(*a*) shows the density obtained from the density matrix, while figure 2(*b*) shows density obtained from the hydrodynamic equations. Several points are worth emphasizing. The structure of the density is the same for the three quantum hydrodynamic equations and for the density matrix calculation. The value of the density for the hydrodynamic equations is closest to that of the density matrix when the constant is between 1 and 3. It should be mentioned that while the value of the multiplicative constant was varied by almost an order of magnitude in these calculations, the quantum potential also underwent changes, and was different for each of the three hydrodynamic equation

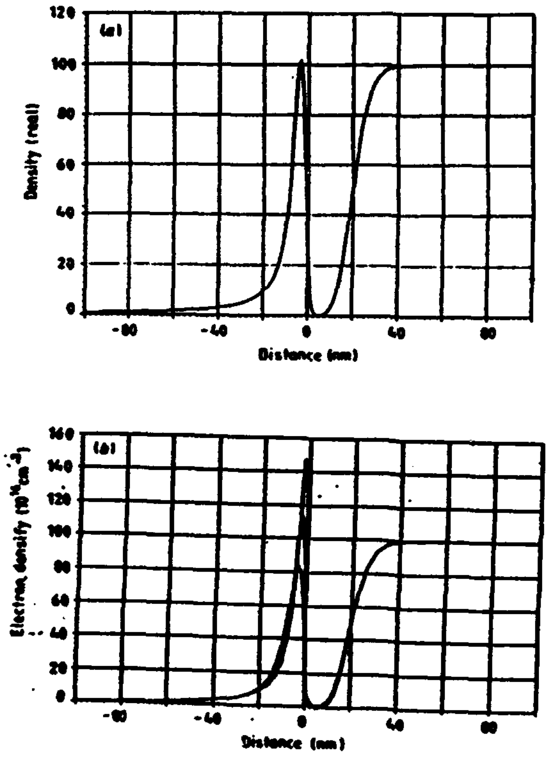


Figure 2. (*a*) Density matrix calculation of the density within a MODFET configuration. (*b*) Quantum hydrodynamic calculation of the density for three different values of the multiplicative factor: $q/9$ (broken curve), $q/3$ (full curve) and $q/\sqrt{}$ (dotted curve). Should read q/1

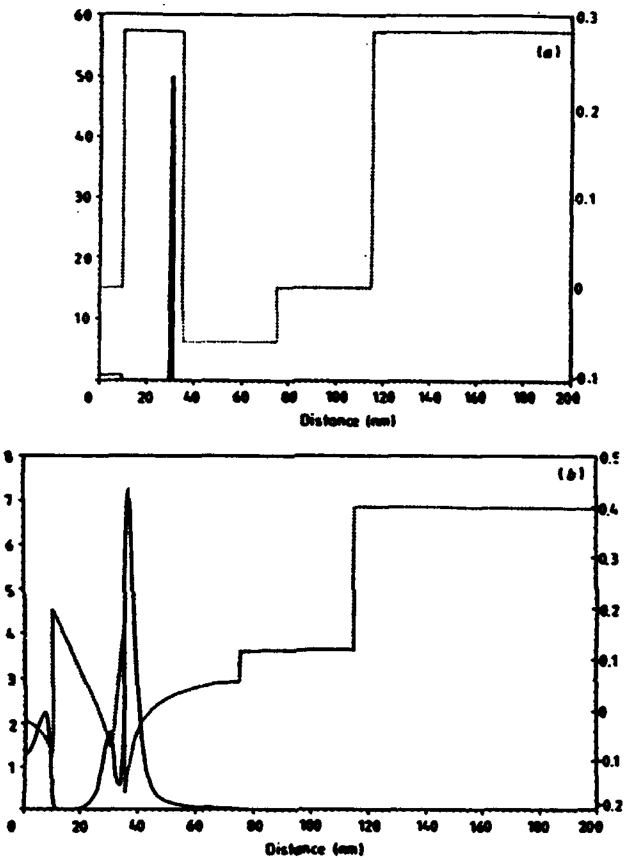


Figure 3. (*a*) Doping (full curve) and barrier configuration (broken curve) for a quantum hydrodynamic calculation. (*b*) Equilibrium electron density (full curve) and potential energy (broken curve) for the structure of (*a*).

Au: curves OK?

calculations, with the net result that all three results are satisfactory in some sense of the word.

4. Quantum well FET calculations

The final example of the use of the quantum potential in modelling devices is that of a quantum well FET calculation. For this calculation, a one-dimensional slice through the source region showing the planar doped region in the wide-bandgap material is shown in figure 3(*a*). In the two-dimensional structure, the gate sits 20 nm into the structure, which is otherwise nominally doped. There is a large concentration of carriers adjacent to the wide-bandgap region, due to the placement of the delta-doped region. The potential energy displays a linear variation with charge depletion in the wide-bandgap region (figure 3(*b*)). All of the density within the quantum well is a consequence of the planar-doped region and the barrier is depleted of carriers. This latter calculation shows the significant advantages found in using the quantum potential to model semiconductor FET structures.

5. Conclusions

In the above discussion, we have illustrated the use of the quantum potential as both an adjunct for interpretation and as a tool for examining transport with the hydrodynamic equations. While calculations using the quantum potential have been performed [7], and illustrate the significance of its contribution to the density distribution within small devices, significant advances will occur when one can model two-dimensional flow in MODFET structures. Preliminary studies indicate that realistic simulations can be performed [8].

Keyword	PAGE PROOFS	*Please note that these first proofs have been output by a Laserjet printer. At revise, we will output through a Monotype Lasercomp, effecting a far superior image.*
SST		

Acknowledgments

This study was supported by ARO and ONR and AFOSR.

References

[1] Bohm D 1952 *Phys. Rev.* 85 166
[2] See, for example, Iafrate G J, Grubin H L and Ferry D K 1981 *J. Physique* C7 307
Au: no?
[3] Grubin H L, Govindan T R, Kreskovsky J P and Stroscio M A *Solid-State Electron.* at press
[4] Wollard D L, Stroscio M A, Littlejohn M A, Trew R J and Grubin H L 1991 *Proc. Workshop on Computational Electronics* (Boston: Kluwer) p 59
[5] Ancona M A and Iafrate G J 1989 *Phys. Rev.* B 39 9536
[6] Grubin H L, Govindan T R and Stroscio M A 1994 *Semicond. Sci. Technol.* 9 xx
[7] Zhou J R and Ferry D K 1992 *IEEE Trans. Electron Devices* 39 1793
[8] Kreskovsky J P and Grubin H L to be published Au: details?

Solid-State Electronics Vol. 36, No. 12, pp. 1697–1709, 1993
Printed in Great Britain.
0038-1101/93 $6.00 + 0.00

TRANSPORT VIA THE LIOUVILLE EQUATION AND MOMENTS OF QUANTUM DISTRIBUTION FUNCTIONS

H. L. Grubin†, T. R. Govindan† and J. P. Kreskovsky†
Scientific Research Associates, P.O. Box 1058, Glastonbury, CT 06033, U.S.A.

M. A. Stroscio
Army Research Office, Research Triangle Park, NC 27709-2211, U.S.A.

(Received 7 February 1993; in revised form 6 May 1993)

Abstract—This paper (i) examines through numerical solutions of the coupled *coordinate representation* Liouville and Poisson equations, the use of the Bohm quantum potential to represent the equilibrium distribution of density and energy in quantum feature size structures; (ii) discusses the development of the nonequilibrium quantum hydrodynamic (QHD) equations with dissipation through the truncation of the quantum distribution function; and (iii) compares select results of the QHD equations incorporating the Bohm potential to the exact Liouville equation solutions. The broad conclusion of the study is that for structures of current interest such as HEMTs, only quantum mechanical solutions, or the incorporation of the quantum potential as a modification of the classical equations will permit representative solutions of such critical features as the sheet charge density.

INTRODUCTION

Advances in crystal growth and processing techniques have assured the construction of nanoscale devices with sharp interfaces. Concomitantly, new device concepts have emerged, including resonant tunneling structures, quantum wires, quantum dots; and varients of classical structures with quantum features, e.g. HEMTs and HBTs. While all devices are governed by quantum mechanics, many devices including HEMTs and HBTs do not require quantum transport for a description of their basic operation. Nevertheless, quantum mechanics is required to provide key electrical features. For example, HBTs sustain low levels of current at low bias levels; these currents are dominantly tunneling currents. Thermionic contributions to current occur at high bias levels. Recently, device formulations utilizing the drift and diffusion equations and the moments of the Boltzmann transport equation were generalized to include a description of tunneling currents (Ancona and Iafrate[1], Grubin and Kreskovsky[2]). These newer studies indicated that quantum contributions of the type first considered by Wigner[3], could be incorporated as modifications to the more traditional approaches to studying transport of carriers through devices. Such an approach was taken by Zhou and Ferry[4,5] in a study of quantum contributions to transport in MESFETs. How well do the quantum modifications of classical transport represent actual transport? This question is addressed for a limited number of cases through comparison of (i) quantum "corrected" solutions with (ii) exact *coordinate representation* solutions to the quantum Liouville equation for the density operator ρ_{op}, whose time dependence is governed by the Hamiltonian H:

$$i\hbar\partial\rho_{op}/\partial t = [H, \rho_{op}]. \quad (1)$$

The relevant quantity in the Liouville simulations is the density matrix $\rho(x, x', t) = \langle x|\rho_{op}|x'\rangle$ whose role is similar to that of the distribution function in classical physics.

The procedure for assessing the quantum contributions has two parts: *First*, approximate and exact equilibrium solutions to the dissipationless quantum Liouville equation for a variety of structures, including a barrier, are compared. The approximate solutions which arise from a new procedure, with results similar to that of Wigner[3], are also expressed in terms of the Bohm quantum potential[6]:

$$Q_B \equiv -(\hbar^2/2m)[d^2(\rho)^{1/2}/dx^2]/(\rho)^{1/2}, \quad (2)$$

whose physical significance is addressed. *Second*, the quantum Liouville equation with Fokker–Planck dissipation mechanisms is introduced[7]; from which a new derivation of the quantum hydrodynamic (QHD) equations are obtained. Nonequilibrium zero current QHD and exact Liouville solutions are compared for a simple heterostructure diode configuration relevant to HEMT structures. We confirm that the simplest version of the QHD equations, the drift and diffusion current density equation, and its zero current solution are modified as follows[1,2]:

$$J(x, t) = \rho\mu k_b T\partial[(V + aQ_B)/k_b T + \ln(\rho)]/\partial x \quad (3)$$

$$\rho = \rho_0 \exp - [V(x) + aQ_B(x)]/k_b T, \quad (4)$$

†Supported by AFOSR, ARO and ONR.

where a is a constant, determined analytically below and in [1] to be $a = 1/3$. More often a is chosen to provide the best fit to exact results, and is thus determined from numerical simulations as discussed below. Any value of a other than $a = 1$ is of concern, in that arguments associated with the single particle Schrodinger equation, suggest a value of unity, see e.g. [2]. Nevertheless, we show for conditions appropriate to Boltzmann statistics, that the exact and approximate solutions for $a \neq 1$ are remarkably similar; and that solutions without quantum contributions will not represent the local charge distribution in barrier dominated structures. Finally, the results are related to earlier work on the Wigner function for mixed and pure states[8]. These latter issues are addressed in the appendices, which also include a discussion of the numerical algorithm.

THE EXACT EQUATION OF MOTION FOR THE DENSITY MATRIX

The Liouville equation in the coordinate representation without dissipation is:

$$\partial\rho/\partial t + (\hbar/2mi)(\nabla_x^2 - \nabla_{x'}^2)\rho - (1/i\hbar)[V(X,t) - V(X',t)]\rho = 0. \quad (5)$$

Solutions yield the time dependent density matrix $\rho(X, X', t)$, whose diagonal components provide the density, and whose values are constrained by the integral: $\int d^3X\rho(X, X) = N_0$, where N_0 is the total *number of electrons*. Assuming free particle conditions along the Y and Z directions, the density matrix, with $\lambda^2 = \hbar^2/2mk_bT$, separates and we seek $\rho(X, X', t)$:

$$\rho(X, X', t) = \rho(X, X', t) \times \exp - [\{(Y - Y')^2 + (Z - Z')^2\}/4\lambda^2]. \quad (6)$$

Here λ, is the thermal de Broglie wavelength. Equation (5) separates and the X, X' portion is rewritten in terms of center of mass and nonlocal coordinates:

$$\text{center of mass coordinates: } (X + X')/2 = x;$$
$$\text{nonlocal coordinates: } (X - X')/2 = \zeta. \quad (7)$$

Note: the transformation is consistent with [3], but is not area preserving (the Jacobian is not unity). In terms of these variables and for free particle conditions along the other directions, the governing equation for $\rho(x + \zeta, x - \zeta, t)$ is:

$$\partial\rho/\partial t + (\hbar/2mi)\partial^2\rho/\partial x\partial\zeta - (1/i\hbar)[V(x + \zeta, t) - V(x - \zeta, t)]\rho = 0. \quad (8)$$

All results arise from $\rho(x + \zeta, x - \zeta)$; nevertheless, we require expressions for current and energy, which are obtained from the diagonal components of the following matrices:

$$\text{density: } \rho(x + \zeta, x - \zeta); \quad (9a)$$

$$\text{current density: } j(x + \zeta, x - \zeta) = [\hbar/(2mi)]\partial\rho/\partial\zeta; \quad (9b)$$

$$\text{energy density: } E(x + \zeta, x - \zeta) = -(\hbar^2/8m)\partial^2\rho/\partial\zeta^2. \quad (9c)$$

The above definitions are discussed in [7], and in Appendix B.

THE APPROXIMATE DENSITY MATRIX EQUATION AND EQUILIBRIUM SOLUTION

Numerical solutions are obtained from eqn (8). For interpretive purposes and for developing the QHD equations, we approximate eqn (8) in two steps. *First*, we assume an infinitely differentiable potential, in which case eqn (8) becomes:

$$\partial\rho/\partial t + (\hbar/2mi)\partial^2\rho/\partial x\partial\zeta - (2/i\hbar) \times \sum[1/(2l+1)!]\zeta^{(2l+1)}\partial^{(2l+1)}V/\partial x^{(2l+1)}\rho = 0, \quad (10)$$

where the sum is over $0 \leq l < \infty$. *Second*, we retain only the first two terms in the expansion:

$$\partial\rho/\partial t + (\hbar/2mi)\partial^2\rho/\partial x\partial\zeta - (1/i\hbar)[2\zeta\partial V/\partial x + (\zeta^3/3)\partial^3V/\partial x^3]\rho = 0. \quad (11)$$

Note: retaining only the term linear in ζ, yields an equation equivalent to the time dependent collisionless Boltzmann equation; demonstrating that quantum effects arise from higher order terms in the expansion of $[V(x + \zeta, t) - V(x - \zeta, t)]$. For the density matrix equivalent to the collisionless Boltzmann equation, and for $\partial\rho/\partial t = 0$:

$$\rho(x + \zeta, x - \zeta, t) = \rho_0 \exp - [\zeta^2/\lambda^2 + \beta V(x)] \quad (12)$$

is an exact solution for free particles (no collisions) in a potential energy distribution $V(x)$. More generally: $\rho(X, X', t) = \rho_0 \exp - [\zeta^2/\lambda^2 + \beta V(x)]\exp - [\{(Y - Y')^2 + (Z - Z')^2\}/4\lambda^2]$. For a reference density ρ_0, *the Fermi energy* $E_F = [1/\beta]\ln[\rho_0/N]$, where $N = 2(m/2\pi\beta\hbar^2)^{3/2}$. Equation (12) is equivalent to $\exp[-\beta\{(p^2/2m) + V(x)\}]$ (see Appendix B).

To obtain the quantum modifications, we recognize that the classical carrier density and mean kinetic energy density under zero current conditions are respectively: $\rho(x, x) = \rho_0 \exp - [\beta V(x)]$, and $E(x, x) \equiv \epsilon(x, x)\rho(x, x) = \rho(x, x)k_bT/2$, where $\epsilon(x, x)$ is the *mean kinetic energy per particle*, and that eqn (12) can be recast as:

$$\rho(x + \zeta, x - \zeta) = \rho(x, x)\exp - [2\zeta^2\beta\epsilon(x, x)/\lambda^2]. \quad (13)$$

Equations (13) and (12) have the same content for classical transport. For quantum transport the mean kinetic energy per particle includes modifications to the classical value[3]. The numerical studies below suggest that the effect of the quantum correction is to either *pinch* or *stretch* the density matrix along the nonlocal direction. Equation (13) represents both contributions. To obtain these corrections eqn (13) is substituted into eqn (11) with the resulting equation ordered in powers of ζ:

$$\zeta\{4\partial[\epsilon(x,x)\rho(x,x)]/\partial x + 2(\partial V/\partial x)\rho(x,x)\} - (8\zeta^3/\lambda^2)\{(\beta\epsilon(x,x)\partial\epsilon(x,x)/\partial x - (\lambda^2/24)\partial^3 V/\partial x^3)\rho(x,x)\} = 0. \quad (14)$$

Thus separately:

$$2\partial[\epsilon(x,x)\rho(x,x)]/\partial x + (\partial V/\partial x)\rho(x,x) = 0 \quad (15)$$

$$\epsilon(x,x)\partial\epsilon(x,x)/\partial x - (\lambda^2/24\beta)\partial^3 V/\partial x^3 = 0. \quad (16)$$

Equation (16) submits to an immediate solution: $\epsilon(x,x) = \epsilon_0[1 + (\lambda/\epsilon_0)^2(1/12\beta)\partial^2 V/\partial x^2]^{1/2}$, where $\epsilon_0 = k_b T/2$ is independent of position and the integration constant is chosen to retrieve the classical result under uniform field and density conditions. If the quantum corrections are small compared to ϵ_0, the quantum corrected energy is:

$$E(x,x) \equiv \epsilon(x,x)\rho(x,x) = [k_b T/2 + (\lambda^2/12)\partial^2 V/\partial x^2]\rho(x,x), \quad (17)$$

which corresponds to Wigner's result ([3], eqn (30)). For the quantum corrected density, we solve a rearranged eqn (15), using eqn (17) for energy:

$$[(\lambda^2/6)(\partial^3 V/\partial x^3) + \partial V/\partial x] /[(\lambda^2/6)(\partial^2 V/\partial x^2) + 1/\beta] + \partial \ln \rho/\partial x = 0. \quad (18)$$

For small quantum modifications the above integrates (approximately) to:

$$\rho(x,x) = \rho_0[\exp - \beta(V + Q_W/3)], \quad (19)$$

where:

$$Q_W = (h^2/4m)\beta[\partial^2 V/\partial x^2 - \beta(\partial V/\partial x)^2/2] \quad (20)$$

defines a Wigner quantum potential. *For small* modifications eqn (19) becomes: $\rho(x,x) \approx \rho_0[\exp - \beta V(x)]\{1 - \beta Q_W/3\}$, which corresponds to Wigner's equation (28).

Equations (19) and (4) have the same form although the modification is in terms of potential energy rather than density. To the extent that the above approximations are of order h^2, we replace the potential energy in eqn (20) with its classical density equivalent: $\beta V(x) = -\ln[\rho(x,x)/\rho_0]$. In this case $Q_B = Q_W$, and eqn (4) is retrieved with $a = \frac{1}{3}$. In terms of density, the energy [eqn (17)] is reexpressed as:

$$E(x,x) = [k_b T/2 - (h^2/24m)\partial^2(\ln \rho)/\partial^2 x]\rho(x,x). \quad (21)$$

The quantum corrected form of the density matrix using eqns (13), (17) and (19) is:

$$\rho(x+\zeta, x-\zeta) = \rho_0 \exp - [\beta\{V(x) + Q_W/3\} + (\zeta/\lambda)^2\{1 + (\lambda^2/6)\beta\partial^2 V/\partial x^2\}]. \quad (22)$$

For small corrections, $\rho(x+\zeta, x-\zeta) \simeq \rho_0 \exp - [(\zeta/\lambda)^2 + \beta V(x)]\{1 - \beta Q_W/3 - (\zeta^2/6)\beta\partial^2 V/\partial x^2\}$, which as discussed in Appendix B, yields upon application of the Weyl integral, Wigner's form of the quantum corrections (eqn (25) of [3]).

Equation (19) highlights the quantum modifications. For example, in the case of a symmetric barrier, classical theory teaches that density is determined solely by the value of the potential. Quantum theory is predicated upon continuity of the wave functions, permits tunneling, and teaches that the density within a barrier can be higher than that determined classically. At the peak of the barrier, $V_x = 0, V_{xx} < 0, Q_W < 0$ and the density exceeds its classical value. Within a symmetric quantum well, at the center of symmetry, $V_x = 0, V_{xx} > 0, Q_W > 0$ and the density can be less than that obtained classically.

COMPARISON OF EXACT AND APPROXIMATE EQUILIBRIUM DISTRIBUTION FUNCTIONS

The extent to which quantum modification represent quantum transport in structures under equilibrium was addressed in two steps. *First*, solutions were obtained for the coupled Liouville equation (8) and Poisson's equation:

$$\partial[\epsilon(x)\partial V/\partial x]/\partial x = -e^2[\rho(x,x) - \rho_D(x)]. \quad (23)$$

(In the discussion below, the permittivity and effective mass are constant, with values are those appropriate to GaAs.) *Second*, the *exact* density computed from the Liouville equation was inserted into eqn (2), Q_B was computed, and an *approximate density* and energy per particle was obtained. The results of part *one* and part *two* were compared. In all calculations global charge neutrality occured: $\int dx[\rho(x,x) - \rho_D(x)] = 0$. Two examples were considered, each at 300 K, where for GaAs the thermal de Broglie wavelength is $\lambda = 45$ Å. In both calculations the nominal density was $10^{18}/\mathrm{cm}^3$. (At these densities gallium arsenide calculations necessitate the use of Fermi statistics. These have been performed by the authors[10], and demonstrate *two* density dependent contributions to energy, one classical and a second quantum mechanical in origin. At lower densities where Fermi statistics are not an issue calculations demonstrate that the effects of the quantum potential Q_B are qualitatively similar to the results discussed below, but the magnitudes of the density derivatives are smaller (longer Debye length) and the quantum corrections are reduced.)

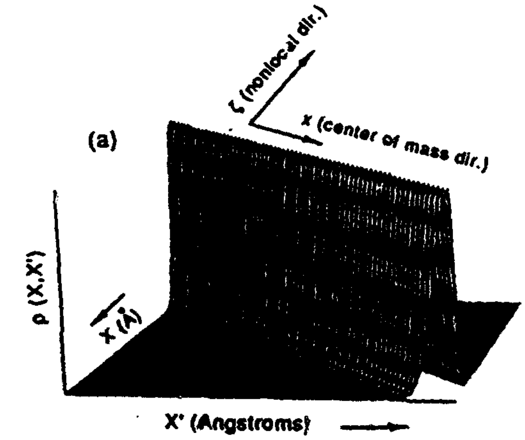
Classical $N^+N^-N^+$ structures

The structure is 1600 Å long with a nominal doping of $10^{18}/\mathrm{cm}^3$ and a centrally placed 500 Å, $10^{15}/\mathrm{cm}^3$ region. The variation in background doping was over one grid point or 4 Å. Solutions yield $\rho(X, X')$, which in equilibrium is real and symmetric, $\rho(X, X') = \rho(X', X)$, as displayed in Fig. 1(a). The inset to Fig. 1(a) is the free particle density matrix. *Free particle Boltzmann boundary conditions* are assumed; i.e. $\rho(X, X') = \rho_0 \exp - (\zeta/\lambda)^2$. All numerical calculations are compared to eqn (4) for the approximate density and the following for the approximate energy per particle [where a is has the same significance as that in eqn (4)]:

$$\varepsilon(x, x) = k_b T/2[1 - (a\lambda^2/2)\partial^2(\ln \rho)/\partial x^2]. \quad (24)$$

The density $\rho(x) \equiv \rho(x, x)$, is represented by the line plot in Fig. 1(b); as is the density obtained from the eqn (4), where $a = 1/3$. The density is ostensibly classical; the quantum corrected density, represented by the dashed lines, closely follows the exact solution. The self consistent potential energy $V(x)$ is displayed in Fig. 1(c), and shows an increase across the N^- region, which accompanies a decrease in charge density across this same region. We are also dealing with a tunneling problem, particularly with those carriers whose energy is below the potential barrier, which in this case has a height of approx. 75 meV. Note: the mean energy of the entering carriers is $k_b T/2 < 75$ meV.

The quantum potential is shown in Fig. 1(d). At the boundary regions $Q_B = 0$; at or near the interface regions Q_B is alternatively positive and negative, and reflects changes in the curvature of the potential. The magnitude of Q_B is approx. 3–5 meV and is nearly 30% of the mean energy of the entering carriers, as given by $k_b T/2$. The signs of Q_B are consistent with continuity of the wavefunction and its derivative through the potential energy barrier region, and weakly prevent the density from approaching its classical value outside (within) the barrier, which instead assumes a smaller (larger) value. This decreased value of density outside the barrier has been described as arising from quantum "repulsion"[11]. The mean energy per particle is displayed in Fig. 1(e), where the solid line represents the exact solution to the Liouville equation [obtained from the ratio of the diagonal components of eqns (9a) and (9c)]. The dashed lines represent eqn (24) for energy, with different values of the coefficient a. The results closest



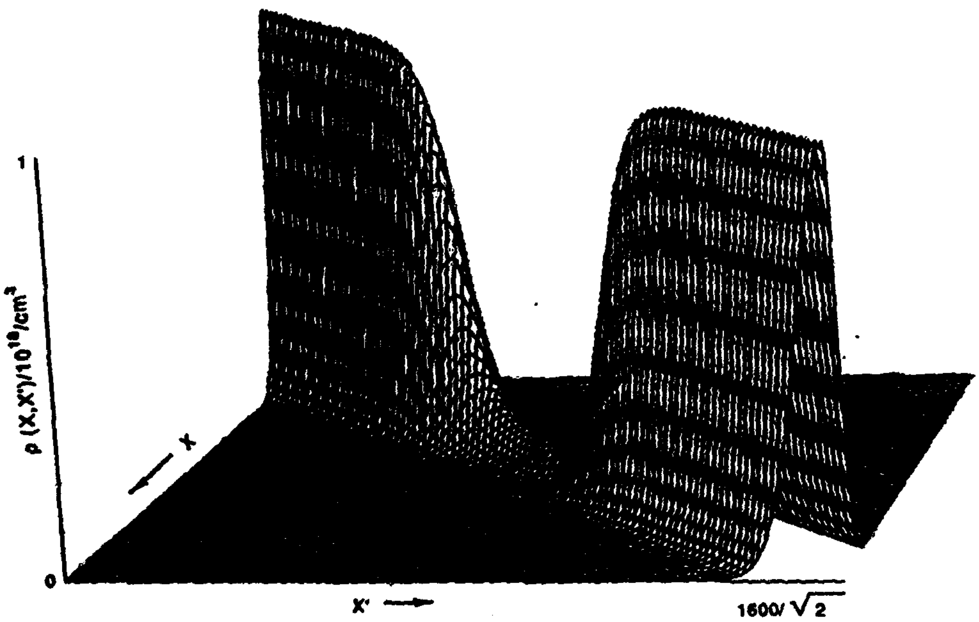


Fig. 1(a). *Caption on facing page.*

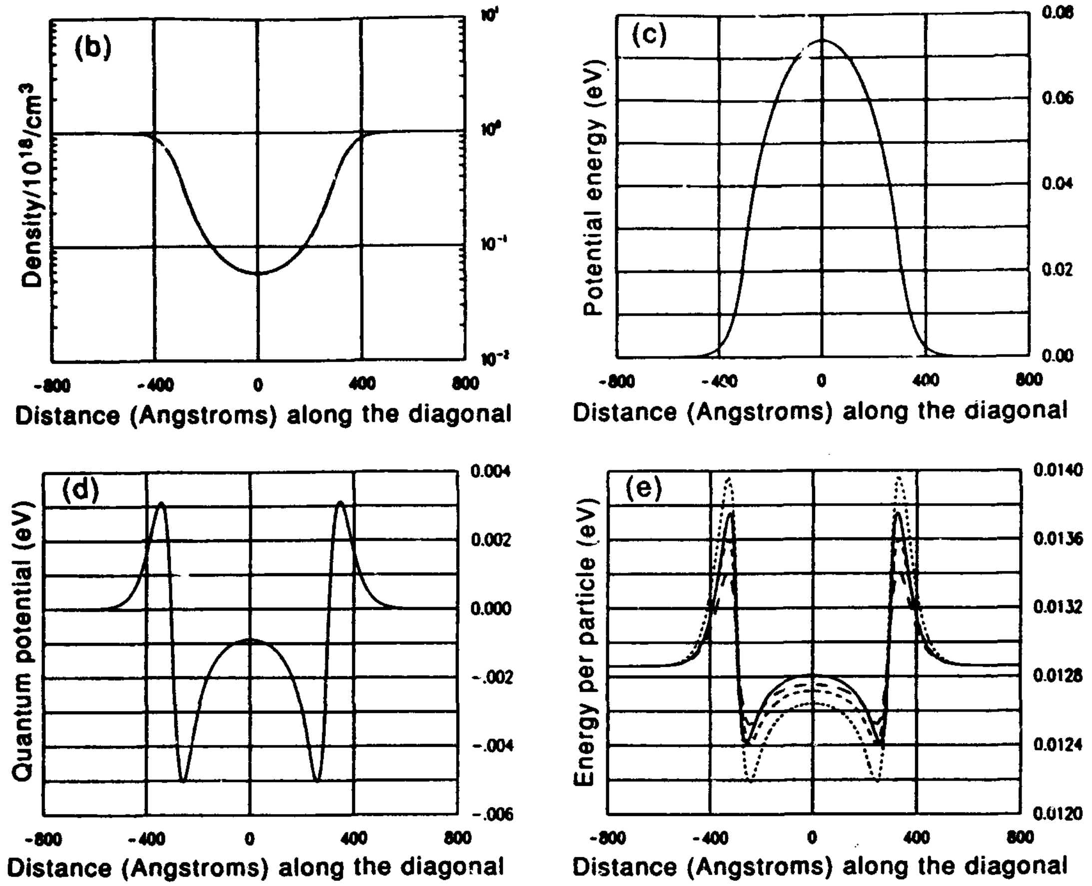


Fig. 1. (a) Density matrix for the $N^+N^-N^+$ structure with free particle boundary conditions, as obtained from the Liouville equation (8). The physical dimension of the structure is 1600 Å, requiring that the density matrix, which is calculated over a square matrix, is of side 1600 Å/$\sqrt{2}$. The center of mass and nonlocal coordinates are indicated. The inset is the free particle density matrix. (b) (——) Diagonal component of the density matrix from (a). (---) Density as obtained from eqn (4) with $a = 0$; (– – –) with $a = 1/3$. (c) Self consistent potential energy for the calculation of (a). (d) Quantum potential for the calculation of (a). (e) (——) Energy per particle as obtained from the diagonal component of eqn (9c) for the calculation of (a). (— —) Energy per particle as obtained from eqn (24) with $a = 1/3$; (– – –) with $a = 1/4$; (---) with $a = 1/5$.

to the exact solution occur for a between 1/3 and 1/4. The significance of these results is that the quantum contributions are solely responsible for the spatial variation in the energy per particle, and demonstrates the presence of quantum contributions with classical structures.

Single barrier diodes

For this calculation the background density is flat and equal to 10^{18}/cm^3; the structure is 2000 Å long and the grid spacing is uniform and equal to 3.33 Å. Figure 2 displays the results for a 500 meV barrier represented analytically by:

$$V_{\text{barrier}}(x) = 500\ \text{meV}[\{1 + \tanh[(x - a_1)/b]\}/2 + \{1 - \tanh[(x - a_2)/b]\}/2 - 1], \quad (25)$$

where $a_1 = -150$ Å, $a_2 = 150$ Å, $b = 50$ Å/3.80. Figure 2(a) displays eqn (25), where V_{barrier} increases continuously (from near zero) to 500 meV, over approx. 75 Å.

Figure 2(b) displays $\rho(X, X')$. As in the Fig. 1 calculation, free particle boundary conditions are assumed. The dramatic "hole" is a consequence of the barrier. Figure 2(c) is a line plot of density. The solid line is obtained from the Liouville and Poisson equations; the dashed lines are from eqn (4) with $a = 0$ (long dashed line) and 1/3 (short dashed line). Common to each calculation is a significant reduction of charge within the barrier, as well as charge accumulation on either side of the barrier. At the edges of the barrier the solutions closest to the Liouville equation are those for $a = 1/3$. The reduction of charge within the barrier is a consequence of the barrier, while the presence of charge adjacent to but outside of the barrier is a consequence of self-consistency in the calculation. Its magnitude is dependent on the condition of global charge neutrality. Figure 2(d) displays the potential energy distribution. The lowering of the potential adjacent to but outside of the barrier (approx. 65 meV) is a consequence of the excess charge and self-consistency. It is important to recognize that the difference between the peak voltage and the minimum voltage is approx. 380 meV which is less than the 500 meV maximum value of the barrier. This lower value is a consequence of the gradual increase in potential from its minimum value. For an abrupt heterostructure, later calculations demonstrate that all of the offset voltage falls at the interface.

Q_B is displayed in Fig. 2(e). Of significance here is the positive (negative) value of Q_B outside (inside), but adjacent to the barrier edge. The positive (negative) value of Q_B is qualitatively similar to that associated with the $N^+N^-N^+$ structure; and there is a corresponding reduction (increase) in charge outside (inside) but adjacent to the barrier from that expected on the basis of classical calculations. Note: (i) the value of Q_B is approximately equal to the value of the self-consistent potential at the edges of the barrier; (ii) the density in the region of highest potential energy, is greater than the density in the center of the structure. (This variation in density is qualitatively accounted for by the value and sign of Q_B.) The inset to Fig. 2(e) is a combination of Q_B and potential energy plots. Figure 2(f) displays the energy per particle. The solid line is obtained from the Liouville equation and eqns (9a) and (9c). The dashed line is obtained from eqn (24) with the coefficient $a = 1/3$. The results are qualitatively in agreement, although the results adjacent to and outside of the barrier are in poor quantitative agreement. The energy per particle is dominated, as in the $N^+N^-N^+$ calculation by Q_B, and emphasizes the role of *gradients in the charge density* to the quantum contributions.

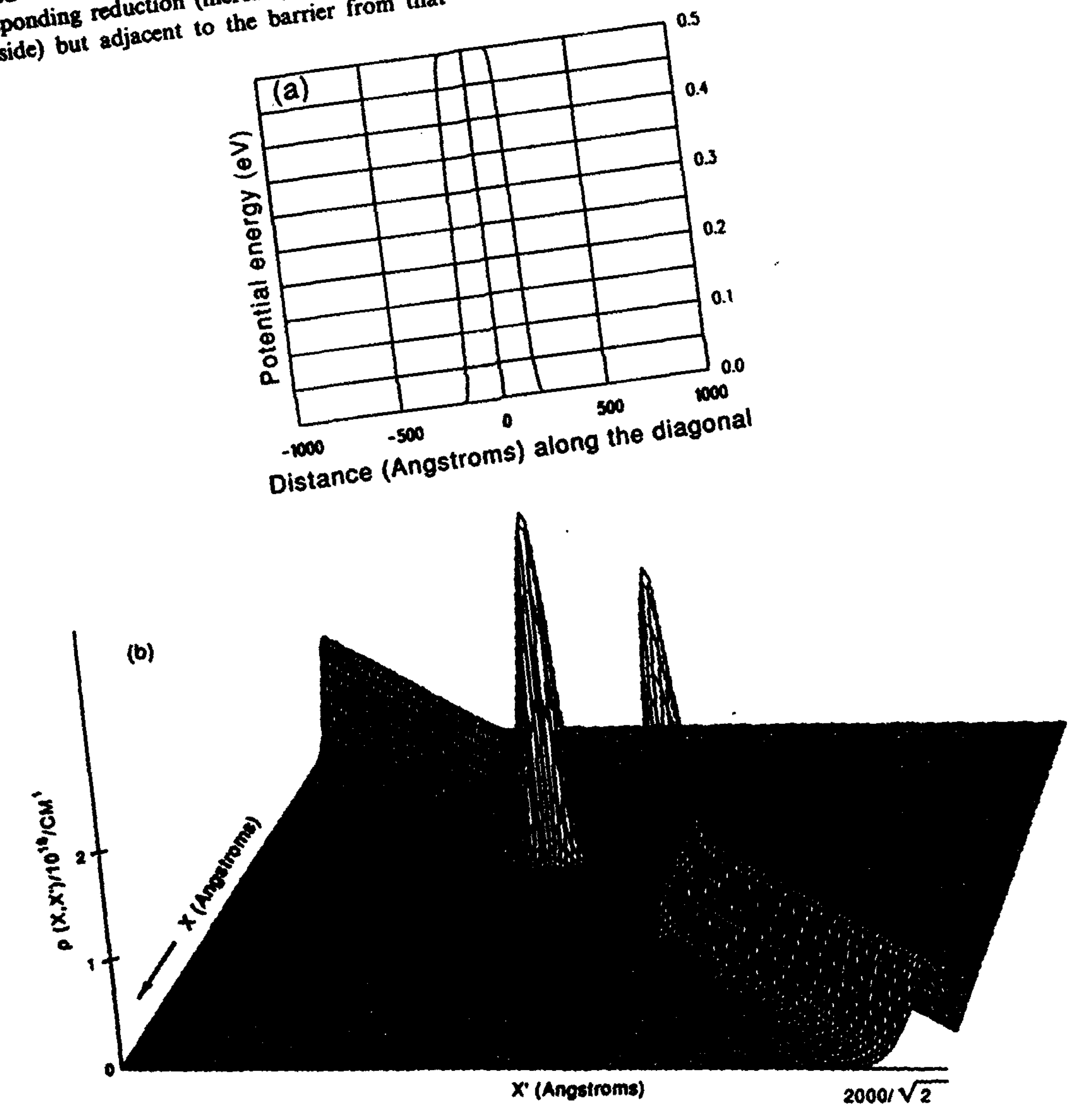


Fig. 2(a, b). *Caption on facing page.*

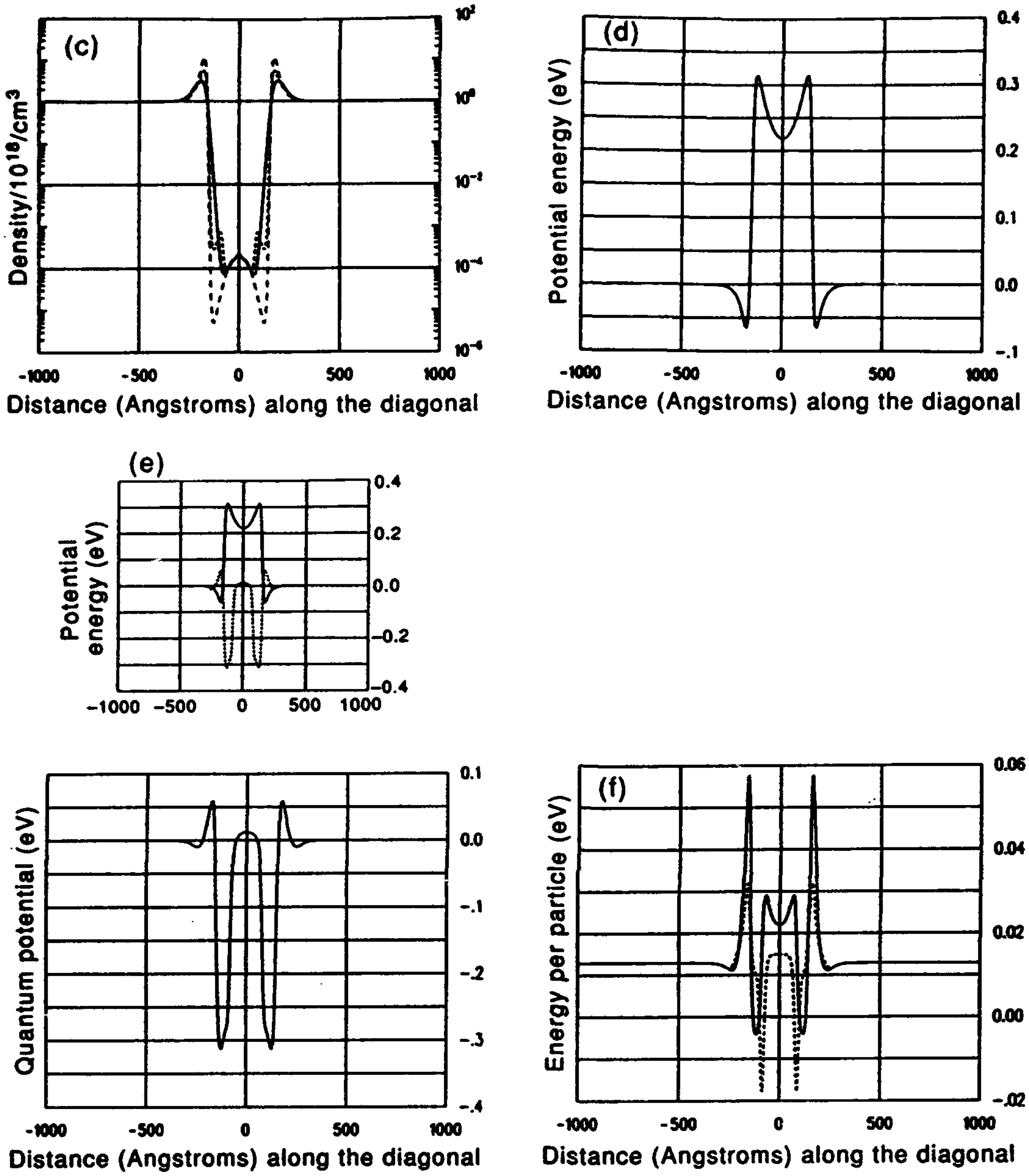


Fig. 2. (a) Sketch of the hyperbolic tangent barrier centrally placed within the 2000 Å structure. (b) Density matrix for the single barrier structure with free particle boundary conditions, as obtained from the Liouville equation (8). The physical dimension of the structure is 2000 Å, requiring that the density matrix, which is calculated over a square matrix, is of side 2000 Å/$\sqrt{2}$. The center of mass and nonlocal coordinates are indicated. (c) (——) Diagonal component of the density matrix from (a). (---) Density as obtained from eqn (4) with $a = 1/3$; (— —) with $a = 0$. (d) Self consistent potential energy for the calculation of (a). (e) Quantum potential for the calculation of (a). Inset includes (d). (f) (——) Energy per particle as obtained from eqn (9c). (---) Energy per particle as obtained from eqn (24) with $a = 1/3$.

THE APPROXIMATE NONEQUILIBRIUM DENSITY MATRIX; THE CONSTRUCTION OF THE QHD EQUATIONS

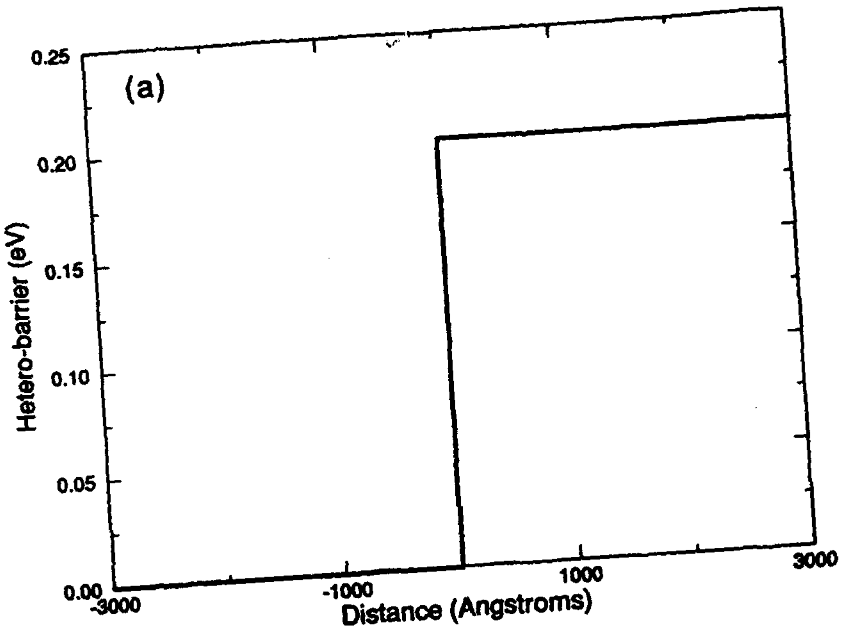
The nonequilibrium situation, is considered within the framework of the QHD equations, which incorporate quantum contributions as *modifications*. The QHD equations incude dissipation within the context of Fokker–Planck (FP) scattering. The motivation for FP dissipation is simplicity. When scattering is treated as in the Boltzmann transport equation, utilization of the Weyl transformation results in an equivalent scattering integral, that is approximately dependent upon powers of ζ, and derivatives with respect to ζ[10]. Under special circumstances these take the form of FP dissipation. The equation of motion of the density matrix with FP dissipation[12] is:

$$\begin{aligned}\partial\rho/\partial t &+ (\hbar/2mi)(\nabla_x^2 - \nabla_{x'}^2)\rho \\ &- (1/i\hbar)[V(\mathbf{X},t) - V(\mathbf{X}',t)]\rho \\ &+ (1/2\tau)(\mathbf{X}-\mathbf{X}')\cdot(\nabla_x - \nabla_{x'})\rho \\ &+ [(\Xi/\hbar^2)(X-X')\cdot(X-X')]\rho = 0, \qquad (26)\end{aligned}$$

where τ represents a scattering time, and Ξ represents a diffusive term in the momentum representation (see Appendix B). Using procedures leading to eqn (11), and assuming that the Ξ is directionally dependent, i.e. along the Y and Z, $\Xi = mk_bT/\tau$ (see also [7]), the equation from which the QHD equations are obtained is:

$$\begin{aligned}\partial\rho/\partial t &+ (\hbar/2mi)\partial^2\rho/\partial x\partial\zeta - (1/i\hbar)[2\zeta\partial V/\partial x + (\zeta^3/3) \\ &\times \partial^3 V/\partial x^3]\rho + (1/\tau)\zeta\partial\rho/\partial\zeta + (4\Xi/\hbar^2)\zeta^2\rho = 0. \qquad (27)\end{aligned}$$

We note that a general set of moment equations is obtained by taking successive nonlocal direction derivatives of the Liouville equation. Truncating the moment equations requires assumptions on the form of the density matrix; and that used below evolved from the approximate equilibrium case. The motivation for such a form is the semi-classical situation where moment equations are often trun-



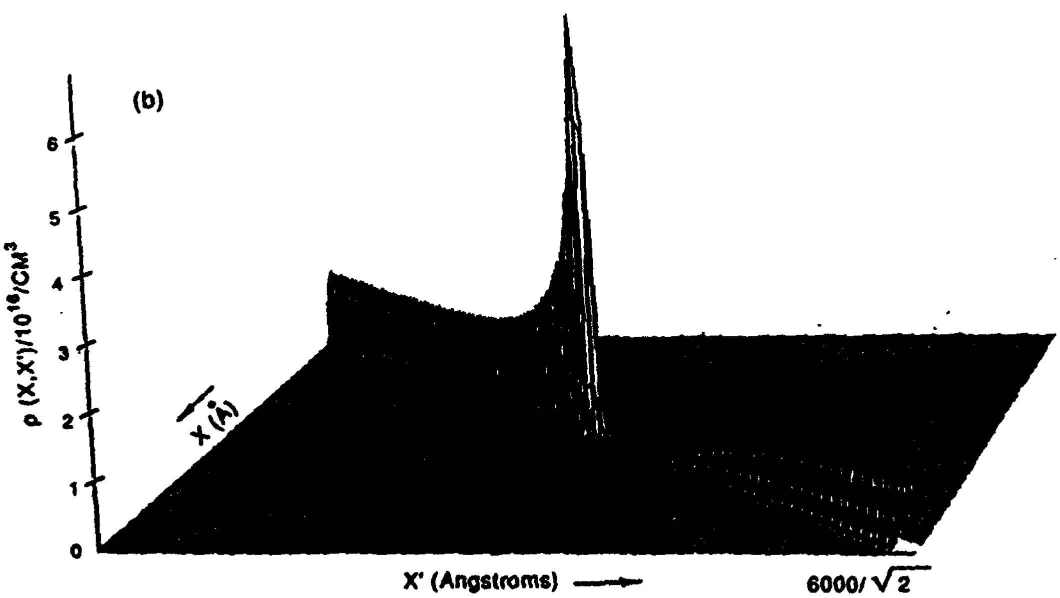


Fig. 3(a, b). *Caption on facing page.*

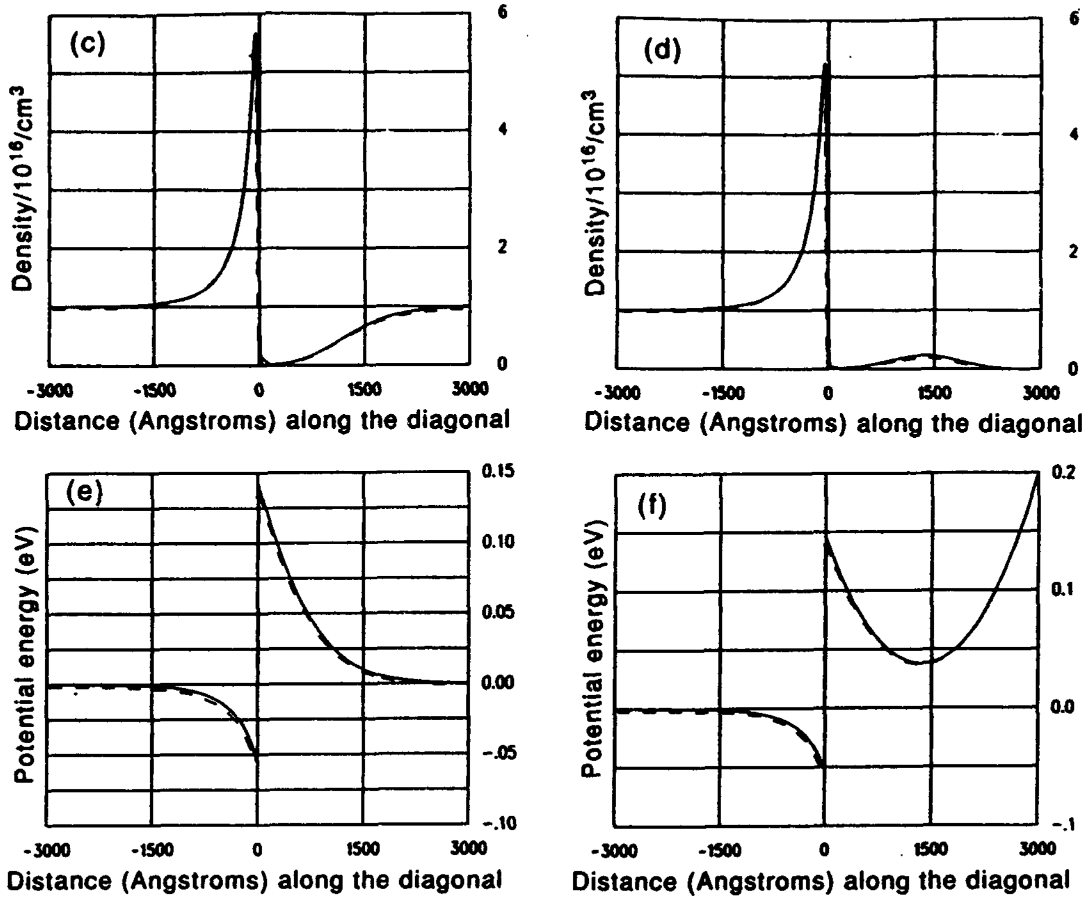


Fig. 3. (a) 200 meV barrier within the 6000 Å structure studied with the simulation. (b) Density matrix for the single barrier structure with free particle boundary conditions, as obtained from the Liouville equation (8) for a bias of −0.2 meV. The physical dimension of the structure is 6000 Å, requiring that the density matrix, which is calculated over a square matrix, is of side 6000 Å/$\sqrt{2}$. The center of mass and nonlocal coordinates are indicated. (c) (——) Self consistent diagonal component of the density matrix for a bias of −0.2 meV. (- - -) Quantum corrected solution. (d) (——) Self consistent diagonal component of the density matrix for a bias of 0.0 meV. (- - -) Quantum corrected solution. (e) (——) Self consistent potential energy for a bias of −0.2 meV. (- - -) Quantum corrected solution. (f) (——) Self consistent potential energy for a bias of 0.0 meV. (- - -) Quantum corrected solution.

cated by representing nonequilibrium by a displaced Maxwellian, $\exp - [\beta\{(p - p_d)^2/2m + V\}]$, and where the mean momentum, p_d, the density, and a particle temperature, are to be determined. The argument for a displaced Maxwellian is the assumption of rapid thermalization. While there is experimental evidence that some quantum feature size devices sustain strong relaxation effects, such phenomena is not likely to be universal. Nevertheless, as a first step in developing a set of nonequilibrium QHD equations we examine the consequences of modifying the quantum equilibrium distributions to describe nonequilibrium conditions. Within the context of the coordinate representation, the Weyl transformation as discussed in the Appendix B, dictates that the displaced equilibrium density matrix (generally non-Maxwellian) used below, is obtained through the following modification of a zero current density matrix:

$$\rho(x+\zeta, x-\zeta) \Rightarrow \rho(x+\zeta, x-\zeta)\exp + [2ip_d\zeta/\hbar], \quad (28)$$

where p_d is at most a function of x. With the current incorporated as in eqn (28) the *construction* of the QHD equations proceeds in three parts: *First*, the truncated density matrix is identified as eqn (28) with the form of the equilibrium contribution given by eqn (22) (the potential is replaced by the classical density equivalent); *second*, the relevant transport quantities are identified as carrier density, $\rho(x)$, mean momentum, $p_d(x)$ and electron temperature, $T_E(x) = 1/(\beta_E k_b)$; *third*, the moment equations, are obtained from a succession of derivatives, followed by the limit as $\zeta \Rightarrow 0$. In taking moments we note that much information contained in the off-diagonal elements of the density matrix is lost.

With eqn (28) the following quantities [from eqn (9)] are relevant to the moment equations (without the equipartition contributions of the Y and Z directions):

$$j(x, x) = \rho(x, x)p_d/m; \quad (29a)$$

$$E(x, x) = [p_d^2/2m + k_b T_E/2 - (\hbar^2/24m)\partial^2(\ln \rho)/\partial x^2]\rho; \quad (29b)$$

$$P^{(3)}(x, x) = [p_d^3 + 3p_d m k_b T_E - p_d(\hbar^2/4)\partial^2(\ln \rho)/\partial x^2]\rho. \quad (29c)$$

Equation (29c) is the diagonal component of $P^{(3)}(x+\zeta, x-\zeta) = (\hbar/2i)^3\partial^3\rho/\partial\zeta^3$, and represents the energy flux, (as typically appears, e.g. in the third moment of the Boltzmann transport equation). Equations (29) and their dependence on derivatives of density are valid only in the limits discussed in the above sections, and represent *modifications* to classical situations. *In this sense it is important to note that the derivation of the quantum potential in terms of Q_W* explicitly involved the carrier temperature. The Bohm potential Q_B, is independent of electron temperature. The consequences of using Q_B rather than Q_W, in the QHD equations should be examined.

The QHD equations are obtained by taking successive derivatives with respect to ζ, as defined by eqn (9) and taking the limit $\zeta \to 0$. The *QHD particle, momentum and energy balance equations*, are respectively:

$$\partial\rho/\partial t + \partial[\rho p_d/m]/\partial x = 0; \quad (30)$$

$$\partial(\rho p_d)\partial t + 2\partial E(x, x)/\partial x + (\partial V/\partial x)\rho + \rho p_d/\tau = 0; \quad (31)$$

$$\partial E/\partial t + 1/(2m^2)\partial P^{(3)}/\partial x + (\rho p_d/m)\partial V/\partial x + 2E/\tau - (\Xi/m)\rho = 0. \quad (32)$$

We rearrange eqns (31) and (32), noting that the quantum correction driving force is *implicit* in $E(x, x)$ and $P^{(3)}(x, x)$. Using eqn (29b) for $E(x, x)$ and noting that $\partial[\rho\partial^2(\ln \rho)/\partial x^2]/\partial x = -(4m/\hbar^2)\rho\partial Q_B/\partial x$, the QHD momentum equation is[2]:

$$\partial(\rho p_d)/\partial t + \partial(\rho p_d^2/m)\partial x + \partial(\rho kT)/\partial x + \rho\partial(Q_B/3)/\partial x + \rho\partial V/\partial x + \rho p_d/\tau = 0, \quad (33)$$

which differs from its classical analog through the presence of Q_B[2]. When the first two terms are zero, and the electron temperature is spatially independent, the drift momentum density reduces to: $\rho p_d = -\tau k_b T\rho\partial[(V + Q_B/3)/k_b T + \ln(\rho)]/\partial x$. Then for $a = 1/3$: and $J = -e\rho p_d/m$, eqn (3) is retrieved; for $pd = 0$, the density, as given by eqn (4) is a solution to eqn (33). Note: the form of the scattering term in eqn (33) identifies the first part of the FP scattering as a frictional term (see [7]).

For the energy balance equation, using eqn (29), eqn (32) becomes:

$$\partial E/\partial t + \partial\{(p_d/m)[E + (\rho/\beta)(1 - [\lambda^2/6]\partial^2(\ln \rho)/ \times \partial x^2)]\}/\partial x + (\rho p_d/m)V_x + 2E/\tau - (\Xi/m)\rho = 0. \quad (34)$$

To determine Ξ, we note that it generally depends upon x, as does τ. In the context of eqn (34) we require that E relax to E_0 which is the $p_d = 0$ value given by eqn (21). This is guaranteed with $\Xi = 2mE_0/\tau$. Thus eqn (34) becomes:

$$\partial E/\partial t + \partial[(p_d/m)(E + \rho k_b T)]\partial x + (\rho p_d/m)\partial[Q_B/3 + V]/\partial x - \rho(\lambda^2 k_b T/6)[\partial^2(\ln \rho)/\partial x^2]\partial \times (p_d/m)/\partial x + (2/\tau)[E - E_0] = 0. \quad (35)$$

The second part of the FP dissipation involves a relaxation to a non-zero thermal energy. E_0 above is the same as used by Woolard *et al.*[13].

The consequences of the above approximations is the appearance of the quantum potential with the factor "1/3". (The situation for Fermi statistics is not addressed here.)

QUANTUM MOMENT EQUATION COMPUTATIONS; COMPARISON TO THE EXACT SOLUTIONS

The development of the QHD equations, is predicted on future use in the design and understanding of multi-dimensional quantum feature size devices. The degree to which this is useful remains to be determined for nonequilibrium phenomena, and the work of [4,5] represents an important beginning. Another relevant case is the evaluation of density across an abrupt heterostructure region, as occurs in either a heterostructure diode or in modulation doped FETs. While the sheet charge density can be obtained from solutions to Schrodinger's equation, the incorporation of such a calculation in a quantum corrected standard set of device simulation equations has only recently been addressed. We consider this in assessing solutions of the QHD equations against the Liouville equation in the zero current limit. It is noted that the use of an abrupt interface violates the following conditions regarded as the basis for the development of the quantum modifications: *the potential is continuous, and the value of Q is small enough to be regarded as a "correction"*. It may be conjectured that the use of quantum potential has more generality than that uncovered in the above derivations; at this time there is no justification for this claim.

The computation is for a 6000 Å structure with constant 10^{16}/cm^3 doping. The grid spacing for the Liouville equation was constant and equal to 7.5 Å; the grid was nonuniformly spaced for the QHD calculation. A 200 meV abrupt barrier is placed across the right half of the structure, as shown in Fig. 3(a). The self-consistent space charge profiles were computed for two values of applied bias: $V_{applied} = 0.0$ eV and -0.2 eV. In both computations the quantum potential was finite within the vicinity of the interface, with structure similar to that of the barrier problem discussed in Fig. 2; it was zero within the vicinity of the boundaries. The two dimensional zero current density matrix for $V_{applied} = -0.2$ eV is shown in Fig. 3(b).

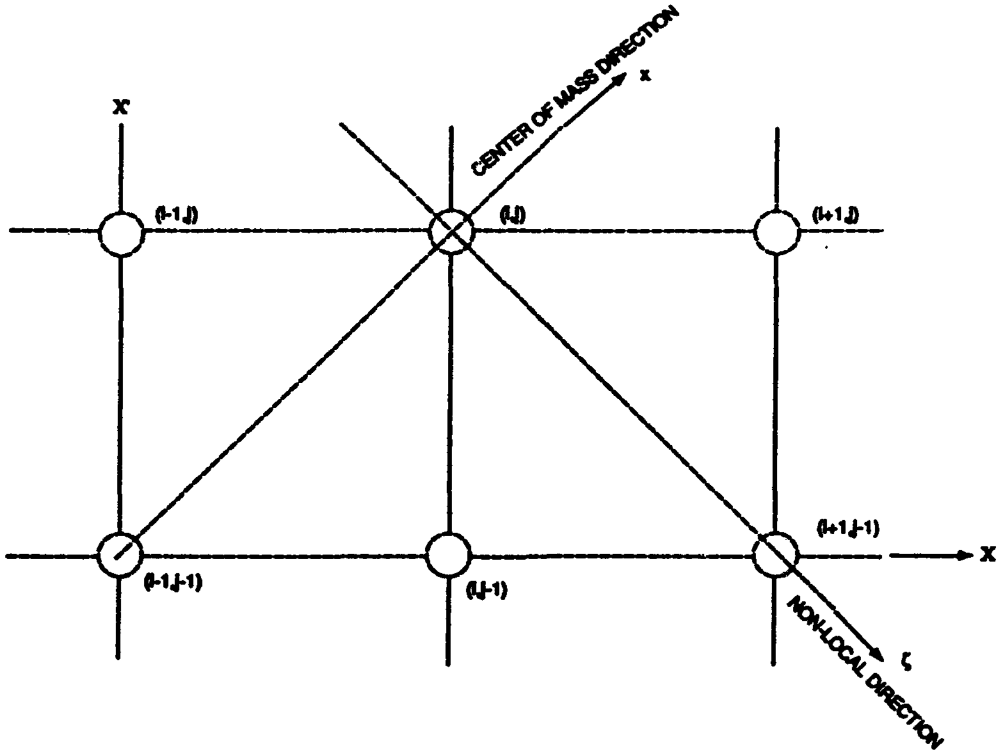


Fig. 4. Schematic representation of the characteristic directions with respect to the grid points.

Line plots of the density matrix for both the exact (solid line) and quantum corrected solution (dashed line) are displayed in Fig. 3(c,d). Note the accumulation of charge on the narrow bandgap side of the structure followed by depletion (with non-negligible values) of charge on the wide bandgap portion of the structure. Under bias the edge of the structure is depleted of charge. The potential energy distribution for the two values of bias are displayed in Fig. 3(e,f) where we see that the discontinuity in potential is equal to the full 200 meV associated with the barrier. The character of this solution is similar to that at the edge of the hyperbolic tangent barrier shown in Fig. 2. In particular the quantum potential is positive (negative) to the left (right) of the metallurgical boundary. The comparative density and potential profiles are extremely close and attest to the confidence of the approximation, for this type of structure. But caution is in order! The excellent agreement for the calculations of Fig. 3, *but* the less certain agreement of Fig. 2, indicate that a careful case-by-case assessement may be necessary. Nevertheless it appears that obtaining representative charge densities necessitates the incorporation of quantum effects through such additions as the Bohm quantum potential. Alternatively, realistic device simulations must resort to a full multidimensional quantum transport calculation.

SUMMARY

This study assessed the introduction of quantum modifications of classical transport, with the results indicating that quantum corrective transport is useful under certain circumstances, and that many simple device studies, such as those for HEMTs would benefit from its incorporation. It is likely that such corrective transport considerations would also be valuable under nonequilibrium conditions particularly in evaluating transport across heterostructure regions. It is important to note that introduction of the quantum potential in a generic form of the QHD equations is not new: it has been linked to density functional theory, as discussed by Deb and Ghosh[14] who also identify the force as being of quantum origin. Bohm and Hiley[15], point out that an essential new feature of the quantum potential is that for single particle Schrodinger fields, only the *form* of the Schrodinger field counts, not the intensity. The force arising from this potential is not like a mechanical force of a wave pushing on a particle with a pressure proportional to the wave intensity; rather the force arises from information content, e.g. structure, rather than value, of the wave[15]. Bohm and Hiley[15] distinguish this force from the Madelung[16] hydrodynamic model in which the particle is pushed mechanically by the fluid.

Acknowledgements—The authors are grateful for conversations with F. de Jong, G. J. Iafrate, W. Frensley, A. Kriman, M. Cahey, and especially D. K. Ferry. The authors are grateful for the assistance of B. Morrison. This work was supported by AFOSR, ARO and ONR.

REFERENCES

1. M. A. Ancona and G. J. Iafrate, *Phys. Rev. B* 39, 9536 (1989).
2. H. L. Grubin and J. P. Kreskovsky, *Solid St. Electron.* 32, 1071 (1989).
3. E. Wigner, *Phys. Rev.* 40, 749 (1932).
4. J.-R. Zhou and D. K. Ferry, *IEEE Trans. Electron Devices* 39, 1793 (1992).
5. J.-R. Zhou and D. K. Ferry, *Semicond. Sci. Technol* 7, B546 (1992).
6. D. Bohm, *Phys. Rev.* 85, 166, 180 (1952).
7. W. A. Frensley, *Rev. Mod. Phys.* 62, 745 (1990).
8. G. J. Iafrate, H. L. Grubin and D. K. Ferry, *J. Physique* C7, 307 (1981).
9. H. L. Grubin, T. R. Govindan, B. J. Morrison, D. K. Ferry and M. A. Stroscio, *Semicond. Sci. Technol.* 7, B360 (1992).
10. To be published.
11. A. M. Kriman, N. C. Kluksdahl and D. K. Ferry, *Phys. Rev. B* 36, 5953 (1987).
12. A. O. Calderia and A. J. Leggett, *Physica* A121, 587 (1983).
13. D. L. Woolard, M. A. Stroscio, M. A. Littlejohn, R. J. Trew and H. L. Grubin, *Proc. Workshop on Computational Electronics*, p. 59. Kluwer Academic Publishers, Boston (1991).
14. B. M. Deb and S. K. Ghosh, *The Single-Particle Density in Physics and Chemistry*, p. 219. Academic Press, New York (1987).
15. D. Bohm and B. J. Hiley, *Foundations Phys.* 14, 255 (1984).
16. E. Madelung, *Z. Phys* 40, 332 (1926).
17. T. R. Govindan, H. L. Grubin and F. J. DeJong, *Proc. 1991 NASCODE.*
18. M. A. Stroscio, *Superlattices Microstruct.* 2, 83 (1986).

APPENDIX A

Solution Procedure

For convenience of solution and determining suitable forms of boundary conditions, eqn (8), is rewritten as a coupled first order system of equations[17]:

$$u(X,X')+[i\hbar/2m][\partial\rho/\partial X+\partial\rho/\partial X']=0 \qquad (A1)$$

$$\partial\rho/\partial t+[\partial u/\partial X-\partial u/\partial X'] + [i/\hbar][V(X,t)-V(X',t)]\rho=0. \qquad (A2)$$

Equation (A1) defines $u(X, X', t)$; eqn (A2) is an alternative form of eqn (8) after accounting for free particle conditions along the Y and Z directions; rewritten in terms of u and ρ. The characteristic directions for eqns (A1) and (A2) are:

$$x=(X+X')/2=\text{constant} \qquad (A3)$$

$$\zeta=(X-X')/2=\text{constant}. \qquad (A4)$$

In terms of the characteristic directions x and ζ, eqns (A1) and (A2) can be written as:

$$u(X,X')+[i\hbar/2m]\partial\rho/\partial x=0 \qquad (A5)$$

$$\partial\rho/\partial t+\partial u/\partial\zeta+[i/\hbar][V(X,t)-V(X',t)]\rho=0. \qquad (A6)$$

Suitable boundary conditions for eqns (A5) and (A6) are the specification of ρ and u along the boundary $X'=0$ and the specification of u along the boundary $X=L/\sqrt{2}$, where L is the length of the device. Along the boundary $X=0$, ρ is specified as the complex conjugate of $\rho(X,0)$, since ρ is hermitian, and u is computed form the outgoing characteristic eqn (A6). Along the boundary $X=L/\sqrt{2}$, ρ is computed from the outgoing characteristic eqn (A5).

An alternative system can be formulated in terms of the current matrix:

$$j(X,X')+[i\hbar/2m]\partial\rho/\partial\zeta=0 \qquad (A7)$$

$$\partial\rho/\partial t+\partial j/\partial x+[i/\hbar][V(X,t)-V(X',t)]\rho=0. \qquad (A8)$$

Equations (A7) and (A8) have the same characteristic directions x and ζ as equations (A1) and (A2). Suitable boundary conditions for eqns (A7) and (A8) are the specification of ρ and j along the boundary $X'=0$ and the specification of ρ along the boundary $X=L$. Along the boundary $X=0$, j is specified as the complex conjugate of $j(x,0)$ and ρ is computed from the outgoing characteristic eqn (A7). Along the boundary $X=L$, j is computed from the outgoing characteristic eqn (A8). Both sets of the first order system of equations, eqns (A5) and (A6) and eqns (A7) and (A8), are useful in applications since they allow different forms of boundary conditions. Both sets of equations can be solved by the same numerical procedure.

The solution procedure consists of solving the first order system of equations as an initial boundary-value problem starting from conditions along the line $X'=0$ and marching to the line $X'=L$ using the method of characteristics. A characteristic net for the equation of motion of the density matrix can be constructed *a priori* from grid points of a uniform square grid. A discrete form of eqns (A5) and (A6) on this grid is [Fig. 4]:

$$[u]_{av}+[i\hbar/2m][\rho(i,j)-\rho(i-1,j-1)]/\Delta x=0 \qquad (A9)$$

$$[\partial\rho/\partial t]_{av}+[u(i+1,j-1)-u(i,j)]/\Delta\zeta + [i/\hbar][V(X,t)-V(X',t)]_{av}[\rho]_{av}=0, \qquad (A10)$$

where $[\cdot]_{av}$ represents an average over the grid cell. Depending upon the form of averaging chosen, eqns (A9) and (A10) form a system of 2×2 block tridiagonal or block diagonal algebraic equations that can be solved at $X'=j$ from known values at $X'=j-1$. Thus, the solution procedure can be marched from boundary conditions at $X'=0$, in steps along X', to $X'=L$. Similar procedures can be utilized for eqns (A7) and (A8).

Self-consistency is included in the analysis by iterating the solution of the density matrix equation with the solution of Poisson's equation to convergence, by successive substitution. For this purpose, Poisson's equation is written in the form:

$$\{\partial[\epsilon(x)\partial(\Delta V)/\partial x]/\partial x\}^n+e^2(\partial\rho/\partial V)\Delta V^n = -\{\partial[\epsilon(x)\partial V/\partial x]/\partial x\}^n-e^2[\rho(x,x)-\rho_0(x)]^n, \qquad (A11)$$

with $V^{n+1}=V^n+\Delta V^n$, where n is the iteration number. The second term on the left hand side of equation (A11) serves to accelerate convergence of the iteration, wherein $\partial\rho/\partial V$ is evaluated at x either numerically from previous iterations or analytically as $\partial\rho/\partial V=-\rho(x,x)/k_bT$, for Boltzmann statistics. A 3-point centered finite difference approximation to (A11) results in a tridiagonal system of algebraic equations that can be solved easily and efficiently for ΔV, which is the increment in V between iterations.

The first step of the iteration procedure consists of assuming a distribution for the self consistent potential (typically, zero everywhere) and solving the density matrix equation to obtain the density distribution. Based on the computed density, Poisson's equation (A11) is solved to update the self consistent potential distribution. For the computations of this paper, the analytical expression for $\partial\rho/\partial V$ was utilized. For cases where $\partial\rho/\partial V$ is computed numerically, several iterations (typically, four or five) are required before a reliable estimate for the gradient can be computed. During these initial iterations, the second term in equation (A11) is replaced by a term of the form $-[\epsilon(x)\Delta V/(\Delta\tau\{\Delta x\}^2)]$ (Δx is the mesh spacing, $\Delta\tau\approx 50$) for convergence. This term could be utilized for all iterations, but convergence is not rapid. Solution to Poisson's equation is used to update the self consistent potential based upon which the density matrix equation is solved again. The iterations are repeated until the density and potential distributions converge to the self consistent solution. For the computations presented here, six orders of residual reduction was obtained in less than 10 iterations.

APPENDIX B

Relation of Results to the Wigner Formulation

The connection between the density matrix in the coordinate representation and the Wigner function is through the Weyl-type transformations with normalizations peculiar to the problem of interest. For the density matrix:

$$\rho(\mathbf{x}+\zeta, \mathbf{x}-\zeta) = 2[1/(2\pi\hbar)]^3 \times \int_{-\infty}^{\infty} \mathrm{d}^3 \mathbf{p} f_w(\mathbf{p}, \mathbf{x}) \exp[2i\mathbf{p}\cdot\zeta/\hbar], \quad \text{(B1)}$$

where the factor of 2 accounts for the fact that each momentum state can hold two electrons. The inverse transformation is:

$$f_w(\mathbf{p}, \mathbf{w}) = 2^3/2 \int_{-\infty}^{\infty} \mathrm{d}^3 \zeta \rho(\mathbf{x}+\zeta, \mathbf{x}-\zeta) \exp[-2i\mathbf{p}\cdot\zeta/\hbar], \quad \text{(B2)}$$

where the factor 2^3 is a consequence of the definition of the nonlocal coordinate [see eqn (7)]. In this transformation it is asserted that the Wigner function and all necessary derivatives with respect to momentum vanish as $p \to \pm\infty$. Note: (a) $\rho(\mathbf{x}, \mathbf{x}) = [1/(2\pi\hbar)]^3 \int_{-\infty}^{\infty} \mathrm{d}^3 \mathbf{p} f_w(\mathbf{p}, \mathbf{w})$; (b) substitution of eqn (12) into eqn (B2) yields the results: $f_w = \exp[-\beta\{(p^2/2m) + V(x) - E_F\}]$.

The Wigner equation including FP scattering, as discussed by Stroscio[18] is:

$$\partial f/\partial t + (p/m)\partial f/\partial x + (1/i\hbar)(1/2\pi\hbar) \int_{-\infty}^{+\infty} \mathrm{d}p' \int_{-\infty}^{+\infty} \mathrm{d}x' f(p', x) \times [V(x,t) - V(x',t)] \exp[i(p-p')x'/\hbar] = (1/\tau)\mathrm{div}_p[\mathbf{p}f] + \Xi \nabla_p^2 f, \quad \text{(B3)}$$

where, as in the main text, all spatial variations are along the x direction, Boltzmann statistics apply, and momentum variations in all three dimensions are allowed. The coefficients τ and Ξ are chosen as in the density matrix studies. For transport in one space dimension it is direct to demonstrate that the integral in eqn (B3) reduces in the classical case to $(\partial V/\partial x)(\partial f/\partial p)$. To second order in $\hbar$, the Wigner equation:

$$\partial f/\partial t + (p/m)\partial f/\partial x - \partial V/\partial x \partial f/\partial p + (\hbar^2/24)(\partial^3 V/\partial x^3)\partial^3 f/\partial p^3 = (1/\tau)\mathrm{div}_p[\mathbf{p}f] + \Xi\nabla_p^2 f. \quad \text{(B4)}$$

The left hand side of eqn (B2) has been discussed in [1,2]. Application of the transformation, eqn (B1), yields eqn (11).

In the absence of dissipation the approximate Wigner distribution function to second order in $\hbar$ is [Wigner[3], see, e.g. eqn (25)]:

$$f_w = \exp -\beta[p^2/2m + V(x)]\{1 - (\lambda^2\beta/4) \times [(\partial^2 V/\partial x^2 - \beta(\partial V/\partial x)^2/3) - \beta(p^2/3m)\partial^2 V/\partial x^2]\} \quad \text{(B5)}$$

which upon application of eqn (B1) yields the equation following eqn (22).

The relation between the density matrix and the Wigner function extends to observables, permitting a concise definition of the associated matrices. Defining *current density, energy density and third moment matricies* respectively, as:

$$\mathbf{j}(\mathbf{x}+\zeta, \mathbf{x}-\zeta) = 2[1/(2\pi\hbar)]^3 \times \int_{-\infty}^{\infty} \mathrm{d}^3 p (\mathbf{p}/m) f_w(\mathbf{p}, x) \exp[2i\mathbf{p}\cdot\zeta/\hbar] \quad \text{(B6)}$$

$$E(\mathbf{x}+\zeta, \mathbf{x}-\zeta) = 2[1/(2\pi\hbar)]^3 \times \int_{-\infty}^{\infty} \mathrm{d}^3 p (\mathbf{p}\cdot\mathbf{p}/2m) f_w(\mathbf{p}, \mathbf{x}) \exp[2i\mathbf{p}\cdot\zeta/\hbar] \quad \text{(B7)}$$

$$\mathbf{P}^{(3)}(\mathbf{x}+\zeta, \mathbf{x}-\zeta) = 2[1/(2\pi\hbar)]^3 \times \int_{-\infty}^{\infty} \mathrm{d}^3 p [\mathbf{p}(\mathbf{p}\cdot\mathbf{p})] f_w(\mathbf{p}, x) \exp[2i\mathbf{p}\cdot\zeta/\hbar], \quad \text{(B8)}$$

it is direct to demonstrate the validity of eqns (9b) and (9c). The derivative definition of the third moment [see discussion following eqn (29c)] follow directly from the above. Note: If the distribution function in eqn (B1) is for zero current, and a finite current function is obtained from $f_w(p - p_D, x)$, then the zero current and finite current density matrix are related as eqn (28).

APPENDIX C

Pure State Results and a Comparison to Iafrate, Grubin and Ferry[8]

The pure state results for current, energy density and third moment may be obtained as follows. Express the wave function as $\Psi(x,t) = \rho(x,t)^{1/2} \exp i\Theta(x,t)$, with $p(x,t) = \hbar\partial\Theta/\partial x$. Then Schrodinger's equation, $i\hbar\partial\Psi/\partial t = -(\hbar^2/2m)\partial^2\Psi/\partial x^2 + V(x,t)\Psi$, which is complex is rewritten as two real partial differential equations:

$$\partial\rho/\partial t + \partial j_{(\mathrm{Schrodinger})}/\partial x = 0 \quad \text{(C1)}$$

$$\partial(\rho p)/\partial t + 2\partial E_{(\mathrm{Schrodinger})}/\partial x + \rho\partial V/\partial x = 0, \quad \text{(C2)}$$

where

$$j_{(\mathrm{Schrodinger})}(x,t) \equiv \rho(x,t)p/m \quad \text{(C3)}$$

$$E(x,t)_{(\mathrm{Schrodinger})} \equiv [p^2/2m - (\hbar^2/8m)\partial^2(\ln\rho)/\partial x^2]\rho. \quad \text{(C4)}$$

While the content of Schrodinger's equation is contained in eqns (C1) and (C2), an expression for the time dependence of the energy may be obtained through the time derivative of eqn (C4) and judicious use of eqns (C1) and (C2). We find, with:

$$P^{(3)}_{(\mathrm{Schrodinger})}(x,t) = [p^3 - (\hbar^2/4)\{3p\partial^2(\ln\rho)/\partial x^2 + \partial^2 p/\partial x^2\}]\rho \quad \text{(C5)}$$

$$\partial E_{\mathrm{Schrodinger}}/\partial t + (1/2m^2)\partial P^{(3)}{}_{(\mathrm{Schrodinger})}/\partial x + (\rho p/m)\partial V/\partial x = 0. \quad \text{(C6)}$$

Note the differences between the pure state definitions [eqns (C3), (C4), and (C5)], and that of eqns (29). For the pure state: *there is the absence of a temperature dependence, the factor of 3 is absent, and there is a velocity correction.*

Solid-State Electronics Vol. 32, No. 12, pp. 1071–1075, 1989
Printed in Great Britain

0038-1101/89 $3.00 + 0.00
Pergamon Press plc

QUANTUM MOMENT BALANCE EQUATIONS AND RESONANT TUNNELLING STRUCTURES*

H.L.Grubin and J.P.Kreskovsky
Scientific Research Associates, Inc., P.O.Box 1058
Glastonbury, Connecticut, 06033-6058

Abstract

This study describes the evolution and implementation of a set of quantum balance equations for examining transport im mescoscopic structures.

Key Word:

Wigner functions, quantum potentials, quantum balance equations.

Introduction

This study describes the evolution and implementation of a set of quantum balance equations for examining transport im mescoscopic structures. The study is motivated by a perceived need for an intuitively accessible set of multi-dimensional quantum transport equations, that permit the self-consistent calculation of particle current and current density. The goal is the development of a set of quantum hydrodynamic equations that reduce to the single particle equations [1] for a pure state, and the classical hydrodynamic equations [2] as $\hbar \rightarrow 0$. As discussed below, these goals have been partially met.

Pure State and Classical Moment Equations

The hydrodynamic equatons for a pure state, for single particle transport, spatial variations in one dimension, and a classical potential U(x,t), are, with $\Psi(x,t) = \sqrt{\rho}\exp[iS(x,t)/\hbar]$, and $p_d = \partial S/\partial x$:

$$\partial_t\rho + \partial_x(p_d\rho/m) = 0 \tag{1}$$

$$\partial_t(\rho p_d) + \partial_x(\rho p_d^2/m) + \rho\partial_x(U+Q) = 0 \tag{2}$$

$$Q = -(\hbar^2/2m\sqrt{\rho})\partial_x^2\sqrt{\rho} \tag{3}$$

In the above, ρ and p_d/m represent probability density and probability current flux; quantum mechanics is represented by the quantum potential, Q [3]. (Note: because of the one dimensional nature of the transport, 'Bohr-Sommerfeld' constraints are automatically satisfied [3].) The classical moment equations for single band transport and spatially independent effective mass are [2]:

$$\partial_t\rho + \partial_x(p_d\rho/m) = 0 \tag{4}$$

$$\partial_t(\rho p_d) + \partial_x(\rho p_d^2/m) + \rho\partial_x U + \partial_x(\rho kT) = \partial\rho p_{d,coll} \tag{5}$$

$$\partial_t W + \partial_x(p_d W/m) + \partial_x(p_d\rho kT/m) + (\rho p_d/m)\partial_x U = \partial W_{,coll} \tag{6}$$

$$W = 3\rho kT/2 + \rho p_d^2/2m \tag{7}$$

It is worthwhile emphasizing that the above equations involve three dimensional momentum space integration, with spatial variations in only one direction, and that density and momentum now represent particle density. The derivative notation in the above equations is $\partial_x = \partial/\partial_x$, etc.

Structure of the Quantum Mechanical Equations

If the quantum transport equations for a pure state are given by equations (1) through (3), and the classical equations are given by (4) through (7), at the very least it may be anticipated that quantum contributions will arise by replacing the classical potential U(x,t) by U(x,t) + Q(x,t). How good is this statement? To examine this we turn to an approximate non-equilibrium Wigner function, discussed in [4].

The non-equilibrium distribution function is constructed [4] from the zero current equilibrium distribution function obtained by Wigner [5], and discussed more recently by Ancona and Iafrate [6]:

$$f_W = \exp{-\beta}[p^2/2m + U]\{1 - 2\alpha(\partial_x^2 U - \beta(\partial_x U)^2/2)/3 - \alpha(1-\beta p_x^2/m)\partial_x^2 U/3\} \tag{8}$$

In equation (8), $\beta = 1/kT$, $\alpha = \hbar^2\beta^2/2m$, and $p^2 = p_x^2 + p_y^2 + p_z^2$.

The construction of the nonequilibrium distribution function involves replacing the potential and its derivatives in equation (8) by corresponding density expressions. The carrier density is

$$\rho = 2(1/2\pi\hbar)^3\int f_W(x,p)d^3p = N\exp{-\beta}U[1 - 2\alpha(\partial_x^2 U - \beta(\partial_x U^2)/2)/3] \tag{9}$$

where $N = 2(m/2\pi\beta\hbar^2)^{3/2}$. After demonstrating that $\partial_x U = -(\partial_x\rho)/(\beta\rho) + O(\alpha)$, and $\partial_x^2 U = -\partial_x\{(\partial_x\rho)/\rho)\}/\beta + O(\alpha)$, it is a direct matter to show:

SSE 32/12—C

$$f_0 = (\rho/N)\exp[-\beta p^2/2m][1+(\alpha/3\beta)(1-\beta p_x^2/m)(\partial_x\{(\partial_x\rho)/\rho\})+O(\alpha^2)] \quad (10)$$

Note: when equation (9) is substituted into equation (4), with U representing the equilibrium potential, equation (8), to order α, is retrieved.

To see what equation (10) offers, consider the steady state small signal Wigner function within a relaxation time approximation, and to second order in $\hbar$ [6],

$$f_w = f_0-\tau_w[(p_x/m)\partial_x f_0-(\partial_x U)\partial_p f_0+(\hbar^2/24)(\partial_x^3 U)\partial_p^3 f_0] \quad (11)$$

Inserting equation (10) into (11), the following key results emerge: $\int f_w d^3p=\int f_0 d^3p$, and for $j_x = -(2e/(2\pi\hbar)^3)\int f_w(p_x/m)d^3p$:

$$j_x = \mu\{\rho\partial_x(U+Q/3)+\partial_x(kT\rho)\} \quad (12)$$

which was first obtained by Ancona and Iafrate [6]. Here, $\mu = e\tau/m$. While this result is consistent with the general philosophy of the introductory paragraph of this section the factor of 3 on the quantum potential needs to be explored. For a displaced version of equation (10), the factor of 3 is retained for the moments of the Wigner-Boltzmann (WB) equation, as considered next.

Moments of the Wigner-Boltzman Equation

The quantum moment equations (see also [7]) have been obtained for the WB equation with quantum contributions to order $\hbar^2$, and for a displaced distribution function in which p in equation (10) is replaced by $p-p_d$. The WB equation of motion is:

$$\partial_t f_w+(p_x/m)\partial_x f_w-(\partial_x U)\partial_p f_w+(\hbar^2/24)(\partial_x^3 U)\partial_p^3 f_w] = f_{w,coll} \quad (13)$$

and the first three moment equaton corresponding to that of equations (4) through (7) are:

$$\partial_t\rho+\partial_x(p_d\rho/m) = 0 \quad (14)$$

$$\partial_t(\rho p_d)+\partial_x(\rho p_d^2/m)+\rho\partial_x(U+Q/3)+\partial_x(\rho kT) = \partial\rho p_{d,coll} \quad (15)$$

$$\partial_t W+\partial_x(p_d W/m)+\partial_x(p_d\rho kT/m)+(\rho p_d/m)\partial_x(U+Q/3)$$
$$-(\rho\hbar^2/12m)\partial_x\{(\partial_x\rho)/\rho\}\partial_x(p_d/m) = \partial W_{,coll} \quad (16)$$

$$W = 3\rho kT/2+\rho p_d^2/2m-(\rho\hbar^2/24m)\partial_x\{(\partial_x\rho)/\rho\} \quad (17)$$

Equation (17) and equation (9) for density has strict quantum mechanical meaning (see also the discussion following equation (25) in [5]). Before attending to the above results it is important to establish a confidence level in the quantum balance equatins. To this end the general moment equation formulation, including dissipation, of Stroscio [8] is recalled. Stroscio's results while specific to a phase space that includes one space dimension and one momentum direction, overlap those of this study, as demonstrated below. In this case, borrowing the notation of [8] within the framework of the displaced Wigner function used herein, it is straightforward to first show that:

$$\rho<(p-p_d)^2> = m\rho[1-(2\alpha/3\beta)(\partial_x\{(\partial_x\rho)/\rho\}/\beta] \quad (18)$$

$$<\rho(p-p_d)^3> = 0 \quad (19)$$

from which reference [8] equations (10a), and (10b) when combined with (10a) yield equations (14) and (15) of this study, where the collision integrals are treated generically. The energy balance equation is treated similarly. Here, reference [7] equation (10b), when multipled by $\rho p_d/m$ is added to (10c), which is multipled by 1/2m; the continuity equation is included in this procedure. The result of this manipulation is a one dimensional phase space version of the energy balance equation of this study, W is replaced by:

$$W' = \rho kT/2+\rho p_d^2/2m-(\rho\hbar^2/24m)\partial_x\{(\partial_x\rho)/\rho\} \quad (20)$$

From the point of view of device modeling, it is pointed out a quantum corrected quasi-Fermi energy can be defined. Writing $E=U+Q/3+kT\ell n[\rho/\rho_r]$, where ρ_r is a reference density, the current density in equation (12) can be written as $j=\mu\rho\nabla E$ [6], and equations (15) and (16) can be reexpressed as:

$$\partial_t(\rho p_d)+\partial_x(\rho p_d^2/m)+\rho\partial_x E = \partial\rho p_{d,coll} \quad (21)$$

$$\partial_t W+\partial_x(p_d W/m)+\partial_x(p_d\rho kT/m)+(\rho p_d/m)\partial_x E$$
$$-(\rho\hbar^2/12m)\partial_x\{(\partial_x\rho)/\rho\}\partial_x(p_d/m) = \partial W_{,coll} \quad (22)$$

In the above form it appears that the dynamics of the transport are governed by an energy E. However, E is introduced as a transformation of variables.

Anticipated Solutions of the Quantum Balance Equations

We focus attention on the single particle pure state equations, whHere of a zero time derivative of the momentum balance equation implies that $S(x,t) = S_1(x) + S_2(t)$, and $\rho(x,t) = \rho(x)$. For zero time derivatives, one space integration yields (ignoring spatial derivatives of the effective mass) energy conservation, $p_d^2/2m + (U+Q) = E_0$, and constant probability current, $\rho p_d/m = J$, where E_0 and J are integration constants. Using the definition of the quantum potential, energy conservation can be rewritten as:

$$\partial_x^2 \sqrt{\rho} + (2m/\hbar^2)[E_0 - U - (mJ^2/2\rho^2)]\sqrt{\rho} = 0 \tag{23}$$

For bound states, $J = 0$, and equation (23) is an eigenvalue problem, one that in the case of a resonant tunnel structure leads to quasi-bound states. Further under zero current conditions, with E_0 representing the eigenvalues, $Q = E_0 - U$, and the values of Q are spatially dependent and, in some cases, are approximately equal to the bound states. This result will be prominantly displayed in the discussion below.

The single particle Schrodinger picture is limited, in that being dissipationless it does not permit a direct transition transition to a multiparticle problem when contacts are considered. For example, in the case of multiparticle transport with electrons moving ballistically within the N^- region of an $N^+N^-N^+$ structure, the mean carrier energy and velocity increase from the cathode to the anode. Conservation of multiparticle current requires that increases in velocity are accompanied by decreases in particle density. Thus in the absence of dissipation there will necessarily be charge depletion at the downstream anode, unless dissipation is present in the interior of the device. If the assumption is made, that the physical contact are boundaries where the numbers of carriers at the cathode and anode are equal, then scattering within the interior of the structure is conceptually necessary. For the hydrodynamic formulation of the single particle Schrodinger's equation, there is no meaning to introducing N^+ cathode and anode regions, since we are dealing with a single particle.

In that dissipation is an essential feature of transport in devices, the quantum balance equations represented by equations (13) through (17) form a staring point for the simulations to be discussed below. To date, our simulations include the first two moment equations, and Poisson's equation. These have been solved for a spatially dependent effect mass, and for Fermi statistics. Here, since we have neither generated a set of WB moment equations for a spatially dependent effective mass nor have we obtained a displaced Wigner function that satisfied Fermi statistics, we have instead patched on these contributions. Further, we have treated the factor of '3' associated with equations (14) and (15) as an adjustable parameter that reflects the statistical distribution used as a basis for the calculation, as discussed in [6], and have replaced it by unity. In this case with $v = p_d/m$, the continuity equation is unchanged, while the momentum balance equation reads: :

$$\partial_t \rho v + \partial_x(\rho v^2) + (\rho/m)\{[\partial_x(U+Q)] + [(\rho v^2/2) - (NkTF_{3/2}/\rho)]\partial_x \ell nm\} + (2/3)\partial_x NkTF_{3/2} + \rho v\Gamma = 0 \tag{24}$$

where $F_\ell(x_f) = (2/\sqrt{\pi})\int[x^\ell/(1+\exp(x-x_f))]dx$, where the integration range is $0 < x < \infty$. 'x_f' is defined implicitly as a change of variables through the relation $F_{1/2} = \rho/N$; where $x_f = \ell n(\rho/N) + \rho/(N\sqrt{8})$, for $x_f < 4.4426$, and $x_f = (9\pi/16)^{1/3}(\rho/N)^{2/3}$, for $x_f > 4.4426$. Using the identities associated with the Fermi integral, namely $F'_{3/2} = (2/3)F_{1/2}$, and introducing the term $E = U + Q + kTx_f$, which is a generalization of the variable transofrmation discussed above the momentum balance equation is:

$$\partial_t \rho v + \partial_x(\rho v^2) + (\rho/m)\{[\partial_x E] + [(\rho v^2/2) - (NkTF_{3/2}/\rho)]\partial_x \ell nm\} + \rho v\Gamma = 0 \tag{25}$$

Equation (25) is coupled to the equation of continuity and Poisson's equation, with U(x) representing the conduction band energy. The heterostructure is represented by the Anderson rule: $U = \Sigma - \chi(x)$, where $\chi(x)$ is a position dependent electron affinity. Σ is obtained from Poisson's equation: $\nabla\epsilon\nabla\Sigma = e^2[\rho - \rho_e]$, where $\epsilon(x)$ is a position dependent permitivity, and ρ_e is a position dependent doping level. For conduction band variations between GaAs and $Al_xGa_{1-x}As$, the following relationships were used: $m^* = 0.067 + 0.083x$, $\Delta E_c = 0.697x$.

Calculations

The calculations discussed below are for the structure shown in figure 1, with resonant tunnelling barriers located symmetrically at the center of the structure. The structure and dimensions of the barriers are displayed in figure 2, which shows the current voltage characteristics of this device at 77K. There is a weak region of negative conductance. The conduction band profile at different bias levels, figure 3, shows the expected tilt as the bias is increased. The distribution of energy is such that at 0.1v, approximatley 20% of the voltage drop falls across the upstream accumulation layer, 30% within the confines of the barrier, and 50% across downstream from the second barrier. The charge distribution, figure 4, shows a region of charge accumulation upstream of the barrier that increases with increasing bias, as does the charge in the well. While different boundary conditions have not been studied these results should be extremely sensitive to the boundary conditions at the cathode; as should the effects of incorporating the energy balance equation. It is not clear that including the latter will reproduce the charge depletion at resonance seen by several other studies.

In all our simulations, we have noticed the formation of a depletion layer downstream of the second barrier once we pass the valley current of the I-V characteristic. This depletion layer is a specific single particle quantum effect. The depletion layer keeps on extending for biases greater than the voltage at the valley of the I-V curve until the depletion layer touches the heavily doped region ($210^{18}cm^{-3}$). Then, the electron density downstream of the second barrier gradually increases and the depletion region disappears.

The quantum potential is displayed in figure 5. If we concentrate on its value in the well, the most dramatic point to note is that as the bias is raised the value of the quantum potential tends to cluster around a narrow range, increasing in magnitude from the upstream barrier to the downstream barrier. Within the barrier the

10. ELECTRON AND HOLE TRANSPORT

Multispecies transport is part of the density matrix algorithm with applications to heterostructure barriers and diodes. While extensive work on electron and hole transport in underway as part of the ULTRA program, some of this type of activity has been published and forms part of this document, and is incorporated into this section.

7.5 High Field Transport and the Gunn Effect in AlGaAs/GaAs Structures

by H.L.Grubin

September 1992

A. INTRODUCTION

When a device physicist studies high field transport there are several issues to address. First, what is the momentum space kinetics, and second what are the high field dynamics in devices? For much of the early phases of device studies in which length scales were of the order of tens and hundreds of microns, a satisfying picture often emerged in which it was assumed that the space charge distribution did not affect, in any significant way, the kinetics of transport. In this picture the semiconductor was represented as a material with a specified field dependent velocity, which in the case of GaAs sustained a region of negative differential mobility associated with k-space transfer. As the structure length scales decreased, it was clear that the separation of the kinetics from the space charge contributions was no longer possible, and a more complicated picture emerged that required the use of advanced numerical algorithms for the study of device physics. These algorithms, however, depended in a detailed way on uncertain parameters that were used in the k-space calculation. This was the best we could do; and today when high field transport is considered it is most often examined without respect to device configuration. In the discussion below, we take this viewpoint.

With high field transport considered within a device context, the ternary alloy AlGaAs can be regarded as an enabling material. For while AlGaAs for a range of aluminium mole fraction possesses a region of negative differential mobility (NDM), other samples possess superior NDM regions. A major interest in AlGaAs lies in the fact that within the Anderson rule, its electron affinity is significantly different from such materials as gallium arsenide and indium gallium arsenide. This, in its simplest version, was responsible for the presence of barrier structure devices, and the earliest AlGaAs/GaAs devices were among the first band structure engineered devices. While a host of AlGaAs/GaAs devices have emerged as a result of barrier engineering we will focus attention on only several versions of this enabling technology. The structures we will focus on are Gunn diodes, AlGaAs/GaAs HBTs, AlGaAs/GaAs MODFETs and AlGaAs/GaAs BICFETs. The emphasis will be on transport. The following will briefly review the high field k-space transport properties of AlGaAs, and then turn to high field transport in the devices mentioned above.

B. HIGH FIELD BULK PROPERTIES

Electron transport on the AlGaAs alloy depends in a detailed manner on the numbers of carriers in the Γ, L and X portions of the conduction band. For moderate values of the mole composition of aluminium, and at low values of electric field the electrons are dominantly in the Γ valley. As the field increases and LO phonon scattering no longer effectively removes the excess carrier energy, a certain fraction of electrons, with the assistance of phonons, transfers to the subsidiary L valley, of which there are four equivalent valleys. The rate at which these electrons are transferred determines whether negative differential mobility will occur. At further increases in field the electrons can also transfer to the next higher valley, the X valley, of which there are three equivalent valleys. Transfer between any two valleys including equivalent valleys occurs. The specific properties needed to determine these transfer rates are listed in TABLE 1, where the density of states effective mass in the table is $m_d \equiv (m_l^2 m_{perpendicular})^{1/3}$.

from ref 1

TABLE 1 Properties of $Al_xGa_{1-x}As$

Parameter	GaAs	AlAs	$Al_xGa_{1-x}As$
Static dielectric constant ε_s	13.18	10.06	13.18 - 3.12x
High frequency dielectric constant ε_∞	10.89	8.16	10.89 - 2.73x
Band gap energy E_g (eV)	1.424	2.168	1.424 + 1.247x ($0<x<0.45$)
			4.07 - 1.1x ($0<x<0.45$)
Indirect band gap (L)	1.708	2.35	1.708 - 0.642x
Indirect band gap (X)	1.900	2.168	$1.900 + 0.125x + 0.143x^2$
Electron affinity χ_e (eV)	4.07	3.5	3.64 - 0.14x ($0<x<0.45$)
			$1.900 + 0.125x + 0.143x^2$ ($0.45<x$)
Density of states electron mass m_e			
Γ valley	0.067	0.150	0.067 + 0.083x
X valley	0.49	0.34	0.49 - 0.15x
L valley	0.22	0.26	0.22 + 0.04x
Acoustic deformation potential Ξ (eV)	6.7	5.5	6.7 - 1.2x
Phenomenological def. potential E_{ac} (eV)	3.6	2.9	3.6 - 0.7x
Intervalley def. potential D(ij) (eV/cm)			
D(Γ,X)	$(0.5 - 1.1) \times 10^9$		
D(Γ,L)	$(0.15 - 1.0) \times 10^9$		
D(X,L)	$(0.34 - 1.1) \times 10^9$		
D(X,X)	$(0.27 - 1.1) \times 10^9$	1.47×10^9	
D(L,L)	1×10^9		
LO phonon energy (eV)	0.033	0.050	0.033 + 0.017x
Intervalley phonon energy (eV)			
($\Gamma \Rightarrow$X)	0.0299		0.0299 + 0.0175x
($\Gamma \Rightarrow$L)	0.03		0.03 + 0.0134x
(X$\Rightarrow$L)	0.0293		0.0293 + 0.0181x
(X$\Rightarrow$X)	0.0299		0.0299 + 0.0175x
(L$\Rightarrow$L)	0.029		0.029 + 0.0150x

A number of studies have been performed which show the degree to which the percentage of aluminium affects the region of NDM. The parameters used in these studies sustain a certain degree of uncertainty in that such scattering contributions as the deformation potential contribution, the relevant optical phonon frequencies, etc., are all dependent upon the aluminium mole fraction. A representative calculation was performed in 1988 [1]. This was a Monte Carlo calculation of the field dependent velocity of $Al_{0.32}Ga_{0.68}As$. Comparison was not made to experiment, as no experimental drift velocity measurements were available. However, the calculation is of sufficiently general nature to reproduce its results here. In ref [1] a set of parameters was chosen from the literature, about which a parameter variation was made. The parameters are approximately represented by inserting 32% into the expressions of TABLE 1, from which the velocity field curve of FIGURE 1 was obtained.

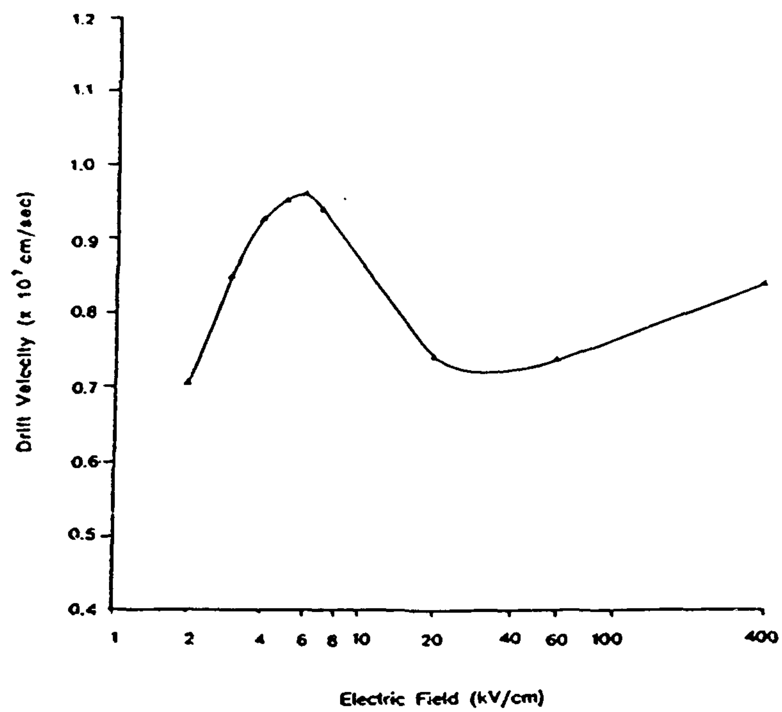


FIGURE 1 Field dependent velocity for $Al_{0.32}Ga_{0.68}As$, from [1].

Several points are relevant: first the peak velocity is slightly in excess of 9 x 10^6 cm/s, which is considerably below that of gallium arsenide. The saturated drift velocity is however approximately equal to that of gallium arsenide. The study in [1] undertook a number of parameter variations. For example when the Γ-L separation was reduced there was as expected a larger fraction of carriers in the L valley in equilibrium, with a consequent reduction in the mobility. Indeed for an energy separation of 91 meV the NDM region is absent. For an increase in the Γ-L separation to 140 meV there is a reduced population in the L valley, the mobility is higher and a region of negative differential mobility is present with an increase in the peak velocity to nearly 1.2 x 10^7 cm/s. Increasing the effective masses of all the valleys tends to reduce the peak velocity as well as the saturated drift velocity. The LO phonon scattering rate increases as the optical phonon frequency increases. Increase in the scattering rate results in fewer electrons heated to sufficient values for intervalley transfer to occur. It was found that for an optical phonon energy increase from 38 to 45 meV the peak velocity increased by approximately 15%. A decrease in the optical phonon energy to 30 meV also resulted in a near 15% decrease in peak velocity. It is important to note that for percentages of aluminium greater than 50%, the region of NDM disappears (see the general discussion in [2]). Additionally, in the vicinity of 40% aluminium the band structure energy minimum order changes from the Γ-L-X ordering to the X-L-Γ ordering.

The phrase heated, in the above paragraphs, implies nonequilibrium electrons, and an electron temperature model is the one most often invoked to deal with this description. There are a variety of means by which the electron temperature model is invoked, the most common being usually predicated on a displaced Maxwellian for a distribution function and solving three sets of equations: a carrier balance equation, a momentum balance equation and an energy balance equation. This model is briefly considered as the language associated with it is invoked in the device discussion.

Under uniform field conditions these equations for two levels of transfer (e.g. Γ and L) represent particle, momentum and energy conservation. For particle conservation:

$$\partial a_1n_1/\partial t = -a_1n_1\Gamma_1 + a_2n_2\Gamma_2 \quad (1a)$$

$$\partial a_2n_2/\partial t = a_1n_1\Gamma_1 - a_2n_2\Gamma_2 \quad (1b)$$

For Eqn (1a) there are n_1 electrons in each lower energy valley and n_2 electrons in each higher energy valley. Eqn (1a) indicates that there are a_1n_1 electrons scattered out, distributed equally to the a_2 higher energy valleys; and there are a_2n_2 electrons scattered from the higher energy valleys into the lower energy valleys. The respective carrier scattering rates are designated Γ_1 and Γ_2. For uniform fields and steady state, a condition under which the velocity field curve is generated, the following condition holds: $a_1n_1\Gamma_1 = a_2n_2\Gamma_2$.

For momentum conservation, the second set of equations describes the rate of change of momentum (or velocity) in the individual valleys under application of an applied field and scattering events. Under uniform fields and for the low energy carriers this equation is:

$$\partial(n_1\mathbf{p}_1)/\partial t = -n_1e\mathbf{F}_n - n_1\mathbf{p}_1\Gamma_3 \quad (2)$$

where the momentum is designated $\mathbf{p}_1, \equiv m_1v_1$, and Γ_3 is the momentum scattering rate for the low energy carriers. Under steady state conditions: $n_1m_1v_1 = -(1/\Gamma_3)n_1e\mathbf{F}_n$. Similar equations can be written for the high energy valley carriers and for holes.

For energy conservation, there are various forms in which the lower and higher energy valley energy equations can be described. We cast the energy equations in terms of the species '1' and species '2' electron temperatures:

$$\partial(a_1n_1T_1)/\partial t = [m_1v_1^2/(3k_B)]\,\{a_1n_1(2\Gamma_3-\Gamma_1) + a_1n_2\Gamma_2\} - a_1n_1T_1\Gamma_5 + a_2n_2T_2\Gamma_6 \quad (3)$$

In Eqn (3) k_B denotes the Boltzmann constant; Γ_5 denotes energy relaxation within the species '1' valley plus energy exchange with the species '2' valley; Γ_6 denotes return energy between species '2' and species '1' valleys. The above analysis requires calculations of the scattering rates. These are taken from scattering integrals. For a review see [2], in which the significance of the above description is dwelled upon. In particular, with the electric field as a driving force the three parameters density, momentum and temperature are determined, as a function of field. It is generally assumed that under equilibrium conditions the distribution of carriers in each of the valleys is determined by the density of states of each valley and the energy separation of each valley.

Adachi [3] has performed some of the above electron temperature calculations. In particular, an estimate of the increase of electron temperature as a function of electric field for polar phonon scattering is shown in FIGURE 2. It is clear from FIGURE 2 that as the alloy composition is increased carrier heating is less severe, at any given value of

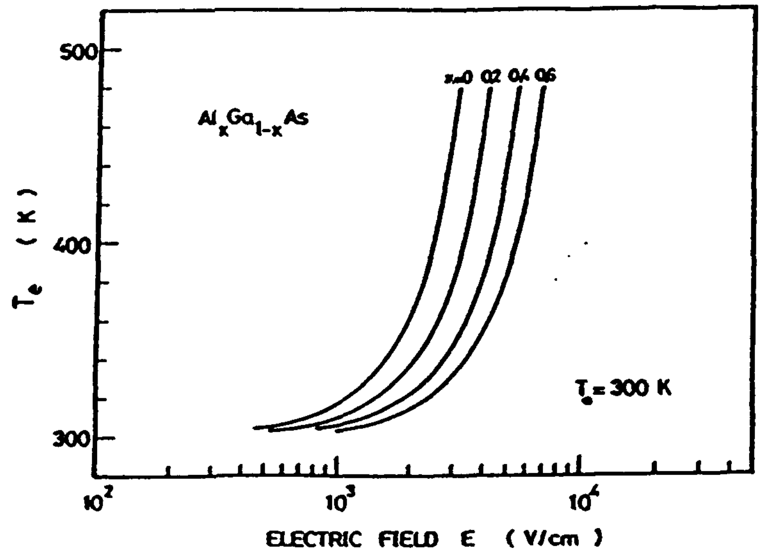


FIGURE 2 Electron temperature as a function of electric field for various alloy compositions, under the condition of polar optical scattering. From [3].

field. When all of the scattering mechanisms are included the electron temperature increases much more rapidly with field, indicating that at high values of electric field there is a considerable amount of transfer from the low energy regions to the high energy regions [2].

Polar optical scattering also dominates the low field mobility of many of these alloys, and in FIGURE 3 the dependence of the polar optical scattering limited mobility for GaAs and indicated alloys is shown, with the apparent degradation as the alloy composition is increased.

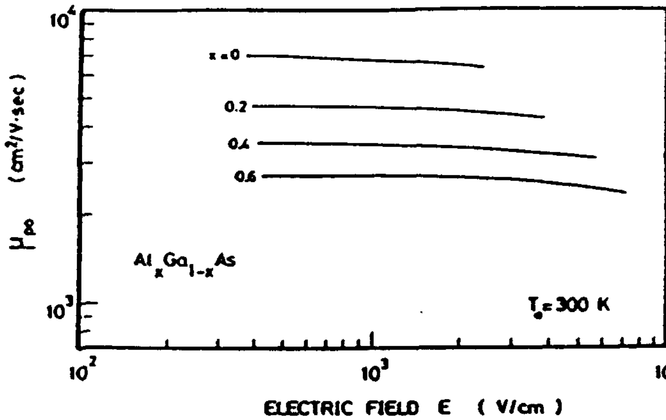


FIGURE 3 Polar optical scattering limited mobility as a function of alloy composition using the electron temperature model. From [3].

C. PERPENDICULAR TRANSPORT - HOT ELECTRON INJECTOR CATHODES

While most of the discussion of high field transport in AlGaAs will be concerned with transport parallel to the interface, the following discussion on perpendicular transport will provide an indication of how high field transport in devices is affected by such materials as AlGaAs.

Consider FIGURE 4, which within the context of the Anderson model is a generic conduction band diagram of a three region structure: HEA-LEA-HEA (H: high, L: low, EA: electron affinity). In the case of hot electron injector cathodes [4], a recent application choice for Gunn diodes, the HEA material was GaAs while the LEA material was AlGaAs. As configured the structure consisted of an n^+ heavily doped cathode, a linearly graded AlGaAs region extending over 500 Å with a height of 300 meV, an n^+ spiked doping layer, approximately 50 Å long, followed by a drift region and an n^+ anode. The design parameters of the structure are consistent with the requirement that for Gunn oscillations to occur electrons must transfer from a low energy region to a higher energy with a consequent local region of negative differential mobility. Early theories recognised that satisfactory control of the oscillations required the presence of local regions to force at least one parameter, e.g. a bounding electric field, to be insensitive to bias and temperature conditions such that electrons would enter the drift region with a distribution of energies consistent with a sufficient number of carriers in the L valley, in the case of e.g. GaAs.

The equilibrium band structure for this case (ignoring subsidiary valleys and invoking Boltzmann statistics) is represented in FIGURE 4 for an undoped AlGaAs launcher. One concept behind this design is that carriers enter the drift region with a non-zero velocity whose value is estimated from the conversion of carrier potential energy to kinetic energy. It is assumed that a large fraction of the electrons that enter the device are Γ valley carriers, as are those that are in the AlGaAs region. Those carriers that pass through the spiked region are Γ valley carriers, but when they get to the boundary of the drift region there is enough energy to place a considerable number of carriers in the L valleys. The design appears to provide improved performance of the transferred electron oscillator, but the presence of the wide band gap AlGaAs next to the GaAs introduces, as in all structures of this type, a triangular potential well and a force that tends to confine carriers. Indeed the presence of such a force on carriers in the continuum is of the order of 10 kV/cm and in the wrong direction! Under bias of course there may, depending on conditions, be a net force acting on the carriers pulling them into the transit region, but then the operating conditions would display a significant dependence on bias. The origin of this dilemma lies in the incomplete manner in which transport problems are addressed. This incomplete feature is highlighted here because the contribution we are about to focus on is present in all structures where there are large carrier density gradients, often associated with heterojunction interfaces.

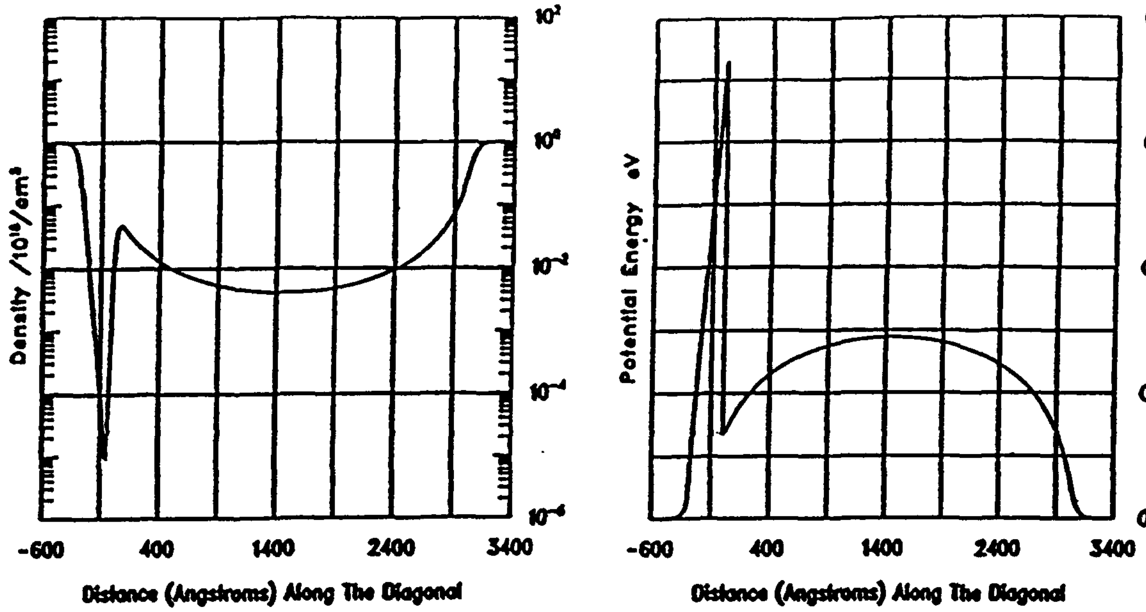


FIGURE 4 Equilibrium density and potential energy for a hot electron injection launcher with an n^+ spike of $10^{17}/cm^3$. From [5].

The calculation displayed in FIGURE 4 was obtained through a solution to the Liouville equation in the coordinate representation [5], and contains the relevant quantum features. However, through an analytical expansion of the Liouville equation, in which classical transport dominates and quantum contributions are treated as corrections, and Boltzmann statistics prevail, it has been demonstrated that the net driving force on carriers is given by:

$$F = -\nabla(V + Q/3) \tag{4}$$

$$Q = -(\hbar^2/2m)(\rho^{-1/2})(\partial^2 \rho^{1/2}/\partial^2 x) \tag{5}$$

We have calculated Q, often referred to as the quantum potential, from the structure of FIGURE 1, and find that in equilibrium there is a net force acting on the carriers that is of the order of 10 keV/cm, acting in such a direction as to move carriers into the drift region. It is perhaps important to emphasise that the configuration of FIGURE 4 is representative of perpendicular transport; and for the specific situation of the hot electron launcher it is expected that the details of such things as the field dependent velocity may be of secondary importance to the feature of providing a heterostructure offset region.

D. PERPENDICULAR TRANSPORT - AlGaAs/GaAs HETEROSTRUCTURE BIPOLAR TRANSISTORS

The second example of perpendicular transport is that of the heterostructure bipolar transistor (HBT). In this case the band structure of the device falls into the same generic category as that of the hot electron injector. Typically the HEA region is a heavily doped n^+ gallium arsenide region, followed by an LEA region which is a heavily doped n^+ AlGaAs region (whose properties are often application specific), followed by HEA material. The HEA material is often a standard p^+ GaAs base followed by a low doped GaAs collector region. Most of the attention associated with this structure involves compositional grading of the AlGaAs emitter. Here there are several aspects to consider. First, the electrons must get from the heavily doped GaAs region to the wide band gap AlGaAs. If the first n^+(HEA)n^+(LEA) interface is not obliterated in the device processing steps, then the interface is expected to look like that shown in FIGURE 5, where the wide band gap material is at the left. Note that in equilibrium the density in the wide band gap material approaches background, which for this case is $10^{18}/cm^3$, within 300 Å, and is relatively insensitive to the density of the adjacent material. Similar remarks apply to the potential energy, which is also shown.

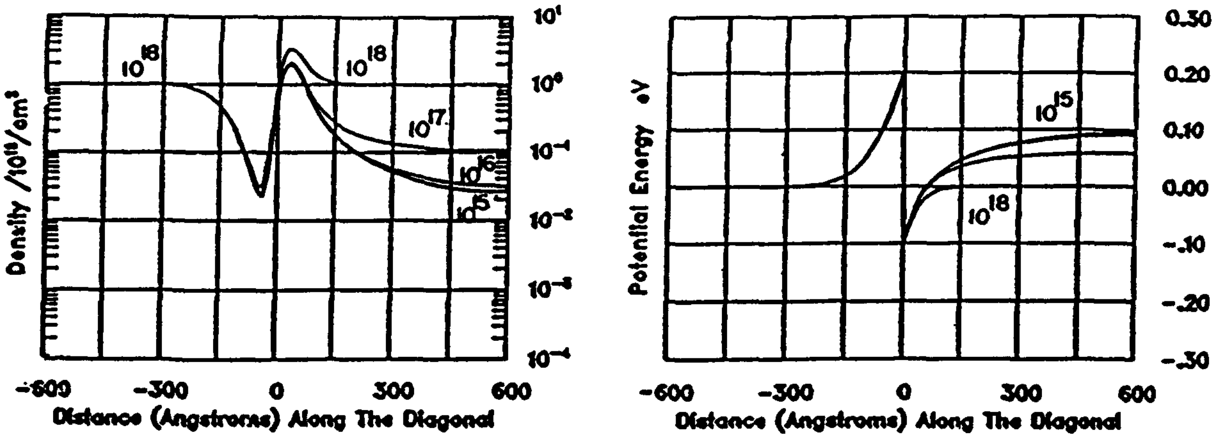
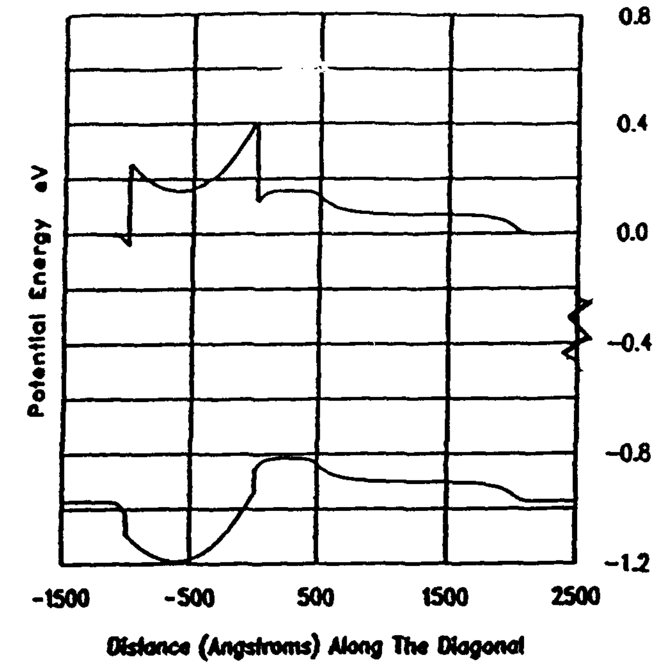


FIGURE 5 Quantum mechanical calculation of the distribution of charge and potential energy for a wide band gap/narrow band gap structure with varying doping distributions. From [5].

In the design of the HBT, if the upstream emitter interface is graded, as in the hot electron injector, the necessity of requiring tunnelling mechanisms to move the electrons from the gallium arsenide region to the aluminium gallium arsenide region is minimised. But most of the effort in designing the HBT is concerned with the compositional grading of the AlGaAs in the vicinity of the base.

The quantum mechanical calculation of the equilibrium potential energy profile and electron and hole density, for an emitter that incorporates an abrupt 1000 Å long, 300 meV barrier adjacent to the base, is displayed schematically in FIGURES 6 and 7, respectively. For this calculation an acceptor doping of $10^{18}/cm^3$, and a wide band gap emitter doping of $10^{17}/cm^3$, was assumed. This latter is generally an order of magnitude below that of the usual design of the HBT, but is sufficient to illustrate the features of the role of the heterostructure in the device - it prevents the diffusion of holes from the base to the emitter while enhancing the injection of electrons into the base. The latter is represented by the dip in the potential at the n-hetero-p interface. There is a diffusion of mobile holes from the base to the collector region, that follows the standard results when recombination is ignored. Higher doping in the emitter will pull down the potential within the centre of the heterostructure region to a near zero value, which is consistent with the higher emitter background doping. While quantum effects associated with electron injection into the base are apparent, on the emitter side the hole density goes from its peak value within the base to a negligible value at a distance of approximately 70 Å into the emitter base; quantum effects are likely here. The details indicate that the band structure of the base is dominated by near charge neutrality within the base away from the heterointerface. Notice that the decrease of the conduction band energy on the emitter side of the barrier corresponds to the increase in the space charge on the emitter side of the barrier. This is similar to the space charge distribution in FIGURE 5.



Relative position within each band are correct. Band separation is incorrect

FIGURE 6 Quantum mechanical calculation of the equilibrium conduction and valence band energies for an HBT with an abrupt electron barrier in the emitter. From [5].

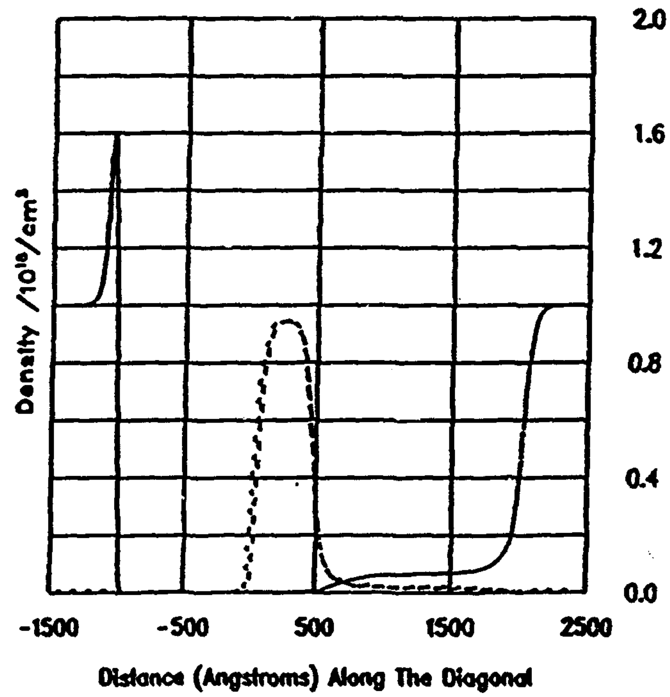


FIGURE 7 For the calculation of FIGURE 6, the equi[illegible]um electron (solid) and hole (dashed) distribution for th

E. PERPENDICULAR TRANSPORT - THE BIPOLAR INVERSION FIELD EFFECT TRANSISTOR

The bipolar inversion field effect transistor [6] (BICFET) has the same generic ordering as that of the two previous devices, but is of interest because in some of its present configurations it involves the strategic placement of a planar doped (or delta doped) layer of acceptors or donors. In particular, the configuration of the HBT shown in the previous figures can be altered to that of the BICFET by replacing the heavily doped p base with a planar doped p region, approximately 30 Å long. Unlike ordinary inversion regions which arise at suitable values of the local potential energy, the local region of holes generated by planar doping is thought to contain many of the bias dependent characteristics of the more common inversion layer (hence the term inversion in the name of the device).

For the BICFET the operating voltages are different because the critical device lengths are of a nanostructure scale. For the configuration of an HBT, and on the basis of earlier studies [7], it is anticipated that the presence of the planar doped layer would lead to potential contours that would lie parallel to the interface everywhere except at the vicinity of the metallisation or contact regions. In the vicinity of the contacts they would spread from a small region at the emitter planar doped edge into the contact regions. For the case in which the planar doped barrier is introduced as a replacement for the base in the HBT, it is anticipated that such terms as the base transit time would improve, simply because of a reduction in the base dimensions. The presence of the wide band gap material such as aluminium gallium arsenide is crucial for the operation of the device because it eliminates the possibility of a remote migration of holes toward the emitter and confines them to the collector region.

F. PARALLEL TRANSPORT - GaAs/AlGaAs MODFETS

While much recent activity has concentrated on the structures discussed above particularly with respect to the analog and digital properties of the devices, the vertical devices, as the above are referred to, do not reflect the transport properties of the ternary AlGaAs. Rather they reflect primarily the band structure of the material in concert with the lower band gap material. To examine the role of high field transport in devices and the Gunn effect, we need to examine the MODFET as a generic device (see ref [8] for a review).

FIGURE 8 is a sketch of a MODFET, and while these structures can be very complicated, the essential feature of the device is that the wide band gap material is doped, and a two dimensional electron gas forms within the narrow band gap material, near the heterostructure interface. The structure shown incorporates AlGaAs/GaAs. Other structures incorporate AlGaAs/InGaAs/GaAs, which includes the possibility of transport in a quantum well.

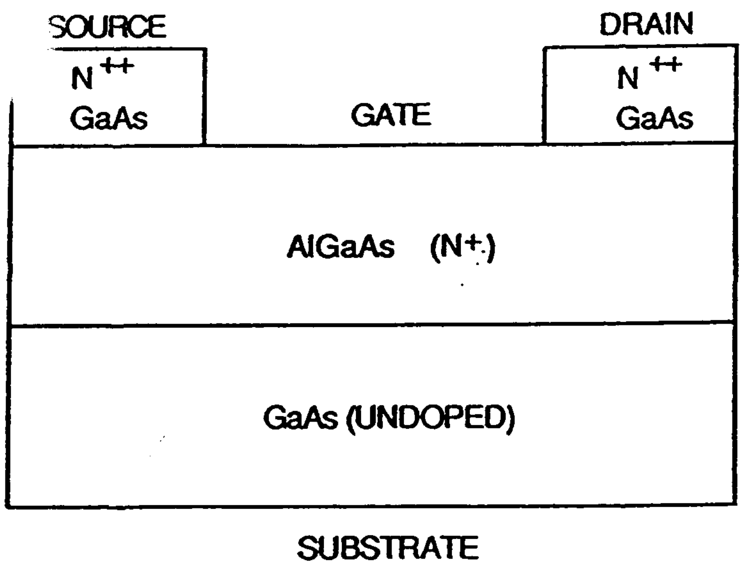


FIGURE 8 Schematic of an n-channel MODFET.

Since the offset voltage of the AlGaAs/GaAs system is dependent upon the mole fraction of aluminium, and since the distribution and number density of the two dimensional electron gas are dependent upon the barrier height between the dissimilar materials, as shown in FIGURE 9, the mole fraction becomes an important design feature. For the FIGURE 9 calculations it is noted that the wide band gap material is uniformly doped, and that the entire offset is at the heterostructure interface. The density distribution displays a decrease within the wide band gap material where a minimum is reached. The maximum value of charge density occurs to the left of the interface and within the narrow band gap portion of the structure. For the 100 meV offset calculation the peak density approaches 7 x $10^{17}/cm^3$, for an approximate sheet carrier concentration of 3 x $10^{11}/cm^2$. For the 200 meV offset calculation the sheet carrier concentration increases by approximately 40%.

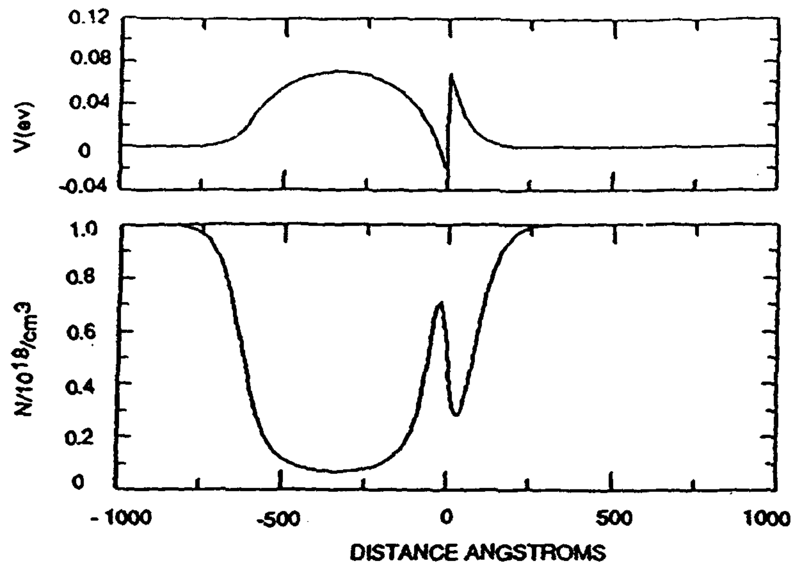


FIGURE 9a Quantum mechanical calculation of the dependence of two dimensional electron gas on the offset voltage. Potential and electron distribution for an n^+n^-(LEA)n^+(HEA) structure with an offset of 100 meV. From [5].

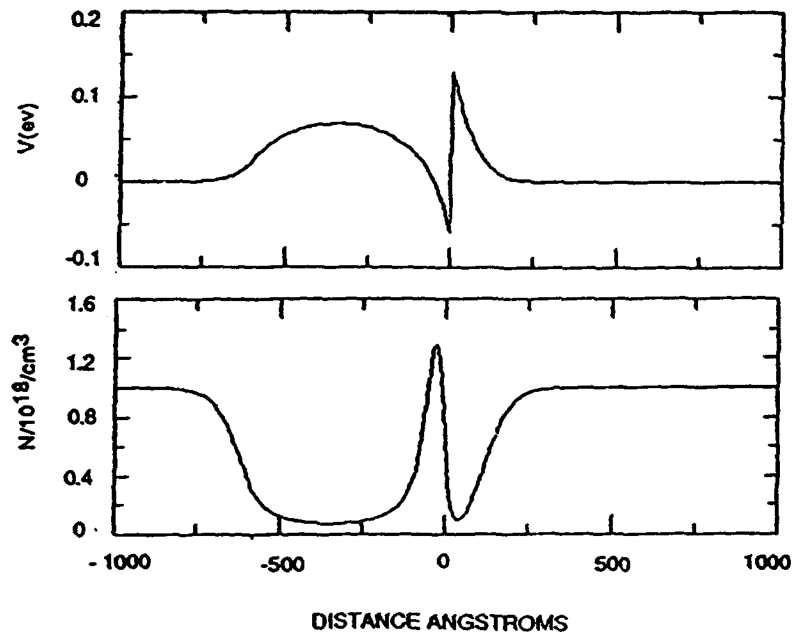


FIGURE 9b As in FIGURE 9a, but for an offset of 200 meV.

An examination of FIGURE 9 provides some important features of transport in the MODFET. First we are interested in transport in the GaAs region. In this region, there is a high concentration of electrons whose origins are in the wide band gap material. Thus ideally, there is no impurity scattering or alloy scattering to worry about. But if we look at the distribution of charge in the aluminium gallium arsenide, it is seen that only a marginal amount of current flows in the structure near the interface, but a substantial amount of current can flow at regions away from the interface. Using the parameters discussed earlier it is known that the alloy and density contribution and the scattering contributions in the aluminium gallium arsenide indicate that any contributions to the current from the wide band gap material are undesirable. This means that design efforts must be introduced to minimise the contributions of transport within the AlGaAs region. One prominent means is to introduce a region of planar doping into the wide band gap region. The situation is displayed in FIGURE 10 for an undoped 400 meV barrier, that incorporates a 30 Å wide planar doped region with a doping of $10^{19}/cm^3$. The important feature to note is that there is very little mobile charge within the wide band gap material, and thus the parasitic current is minimised. The details indicate that the peak in charge falls within the narrow band gap material, and that the presence of exposed donors, associated with the planar doping within the barrier, results in a potential energy minimum in the vicinity of the donors. But the significant reduction in charge density implies a reduction in parasitic current.

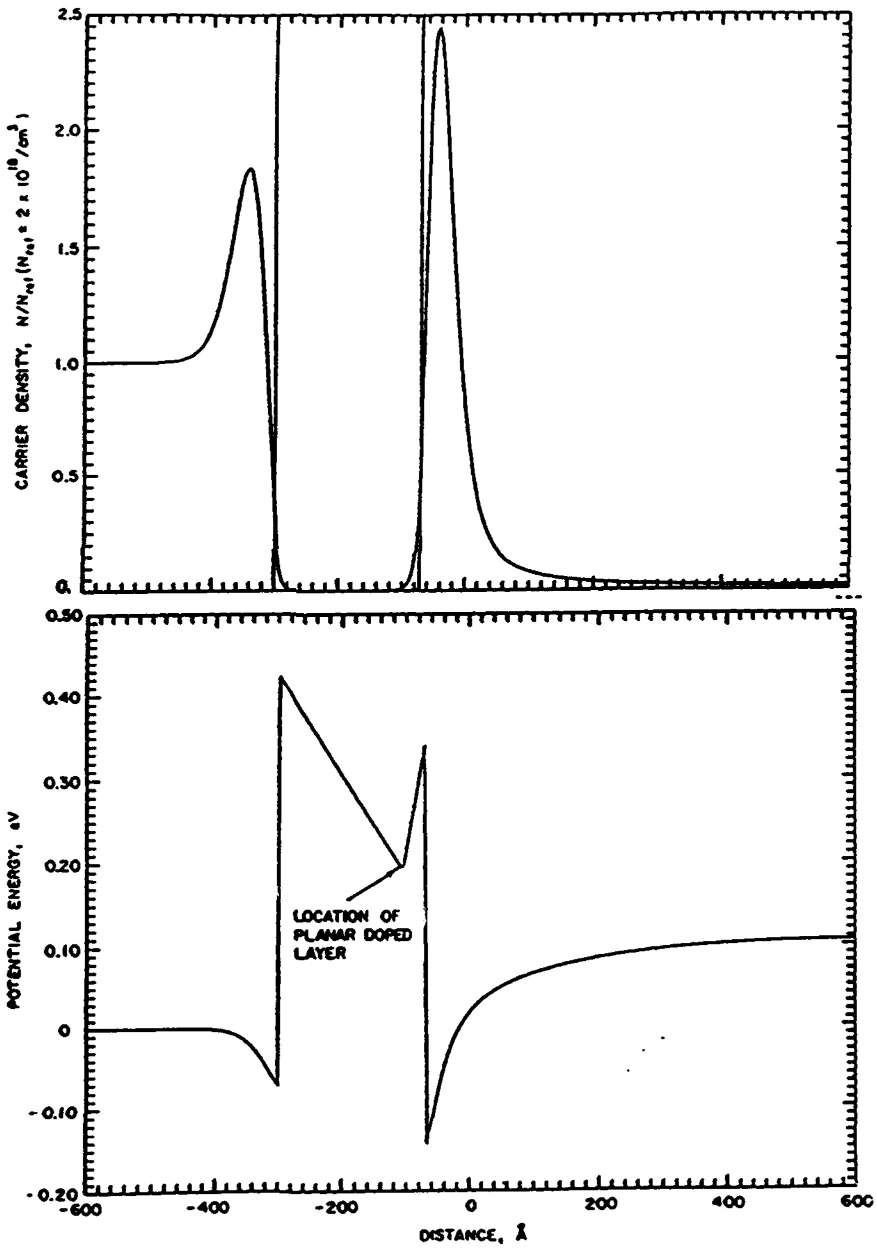


FIGURE 10 Quantum mechanical calculation of the dependence of two dimensional electron gas form in the presence of a planar doped donor region. Potential and electron distribution for an n^+(HEA)n^-(LEA)n^-(HEA) structure with an offset of 400 meV.

In addition to the above issues several others emerge. For the materials of interest there are three bands to consider, the Γ, L and X valleys for each of the materials. Here, generally only the offset voltage between the gamma valleys of the two materials are considered. For concreteness consider $\Delta E(x,i,j)$ where $i,j = (\Gamma,L,X)$ and x denotes the composition of aluminium. Then while the $\Delta E(x,\Gamma,\Gamma)$ increases as x increases from $x = 0$, for sufficiently large x the subsidiary valleys begin to influence the statistics. The commonly used value for x in MODFETs was 30%, where the offset voltage is approximately 260 meV. At this value the commonly known DX centre comes into play and leads to instabilities particularly in digital circuits, as a result of which much effort has been confined to MODFETs with lower aluminium composition.

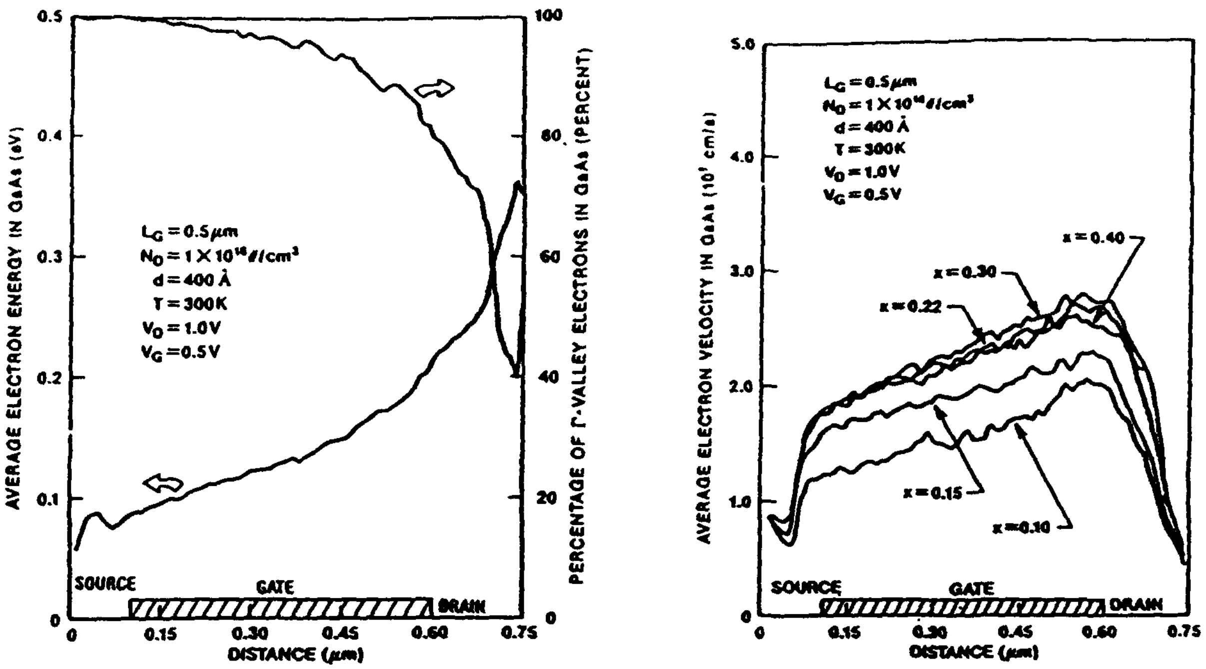
G. PARALLEL NONEQUILIBRIUM TRANSPORT - HIGH FIELD EFFECTS AND MODFETS

We now examine high field transport in these structures. There are several ways in which these problems can be studied. The drift and diffusion equations can be invoked, requiring the presence of field dependent drift velocity for the carriers, or a nonequilibrium formulation transport can be examined. For the case of nonequilibrium transport, the common approach has been to invoke either Monte Carlo or balance equation procedures. Only the Monte Carlo approach will be discussed below.

A particularly interesting set of results has been discussed in [9] and [10]. Both of these calculations include real-space transfer [11], permitting electrons from the narrow band gap material, with a sufficiently large energy, to transfer back into the wide band gap material. Typical material parameters used in these studies are represented in TABLE 1.

The structure studied in [9] consisted of a source drain spacing of 0.75 microns and a gate contact, 0.5 microns long, located 0.1 microns from the source boundary. The thickness of the AlGaAs was 400 Å with a uniform donor concentration of $10^{18}/cm^3$. The composition of aluminium was 22% with a barrier near 190 meV. The gallium arsenide layer was undoped and 0.2 microns thick. While calculations were performed to examine the switching speed in going from one voltage state to a state in which the gate and drain voltages were respectively 0.5 V and 1.0 V, we will concern ourselves only with the steady state.

What is to be expected? For these calculations in which only one volt falls across the source to drain region, the average electric field is approximately 15 kV/cm, and is nonuniformly distributed and determined by Poisson's equation. While the electrons acquire energy from the field as they traverse the structure there are phonon losses within the gallium arsenide layer, and the average energy of the carriers increases to only 350 meV at the end of the channel. Within the channel and for approximately 75% of the channel length most of the carriers in the GaAs are Γ valley carriers, with most of the transfer to the subsidiary valleys occurring near the downstream portion of the structure as shown in FIGURE 11a. The interesting feature of this result is that if most of the carriers in the structure are gamma valley carriers, then the electron velocity within the channel should be dominated by the low field mobility of the material. But a study of the mean carrier velocity (FIGURE 11b) in the structure demonstrates that in regions where there is significant electron transfer, and a reduction of the numbers of gamma valley carriers, the average electron velocity in the gallium arsenide continues to increase. This is reasonable, since the mobility of the subsidiary valleys is considerably smaller than that of the gamma valley and thereby makes a negligible contribution to the total current. Thus, even when electron transfer occurs the current is dominated by the high mobility carriers.



(a) (b)

FIGURE 11 (a) Average electron energy and percentage of Γ-valley electrons along the GaAs channel for y = 0.3. (b) Average electron velocity. [9]

Another feature of importance is the relative concentration of carriers in the subsidiary valleys of AlGaAs. According to the model of [9], with the exception of a composition of 10% aluminium, in which case most of the carriers remain in the lower energy portions of the conduction band or the first 75% of the device length, the movement of the energy bands closer to the gamma valley with increased aluminium composition implies that the percentage of carriers in the subsidiary valleys increases. Thus there are fewer light mass carriers available to conduct current and the mean electron velocity in the aluminium gallium arsenide is expected to be significantly below that of gallium arsenide. FIGURE 12 displays the percentage of carriers in the subsidiary valley and the corresponding average velocity. We point out that the velocity for all but the x = 0.1 mole fraction is approximately an order of magnitude less than that of the mean velocity in gallium arsenide.

In [9], the authors point out that in the low field region of the structure, which is dominated by gamma valley transport, real space transfer from the gallium arsenide to the aluminium gallium arsenide, which is the thermionic emission of electrons from one device layer to another because of heating of the carriers by an electric field, is approximately balanced. In the high field regime, which occurs near the end of the structure, gamma valley electrons transfer to the L valley. The L valley electrons in the GaAs easily undergo real space transfer into the L valley of AlGaAs due to deformation potential scattering and a relatively low offset barrier for the L valley electrons. The electrons in the AlGaAs undergo further transfer to the X and L valleys. The X valley carriers do not transfer efficiently to the X carriers in GaAs due to an unfavourable offset voltage.

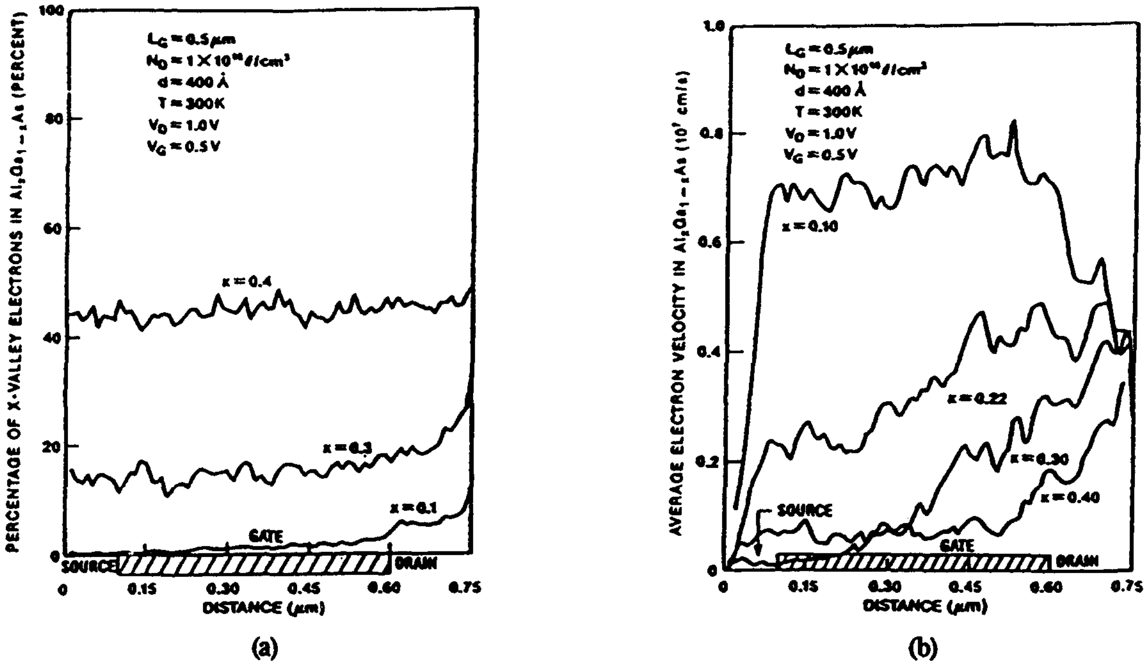
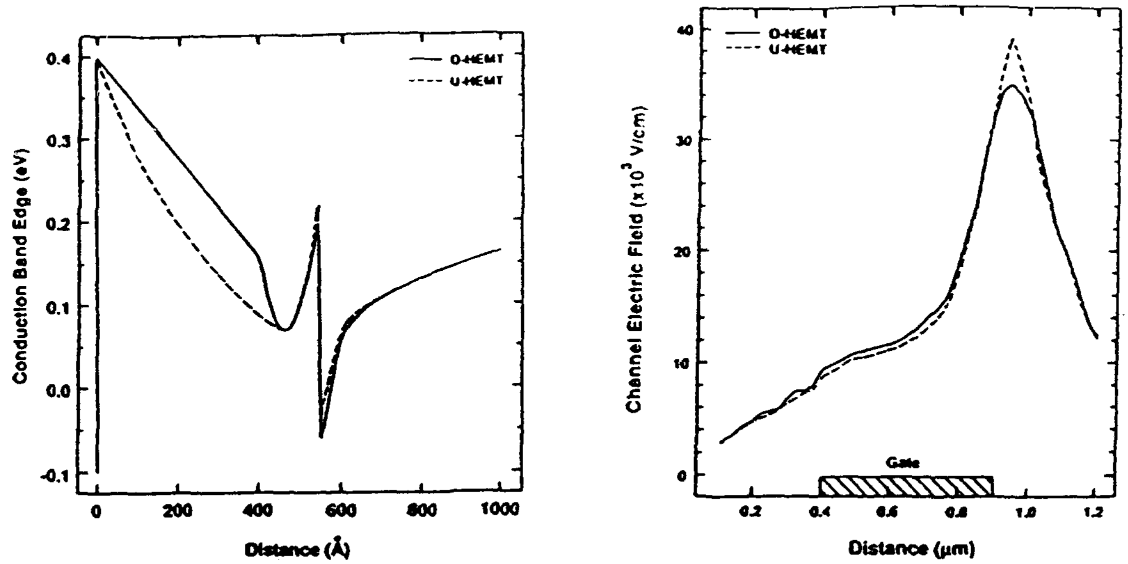


FIGURE 12 (a) Percentage of electrons in AlGaAs which have scattered into the X valley for the indicated aluminium compositions. (b) Average velocity in AlGaAs. [9].

The study of [10] while similar to that discussed above emphasised the role of the transport in the AlGaAs region subject to different doping distributions. In particular, they examined an AlGaAs/GaAs MODFET, with a gate length of 0.5 microns, centrally placed, with a spacing of 0.3 microns from the source and drain metallisation regions. The aluminium percentage was 30%. Two cases were considered. In the first an n^+ AlGaAs region 500 Å deep was doped to $1.8 \times 10^{18}/cm^3$, followed by an undoped AlGaAs spacer layer 50 Å deep, and a 2000 Å undoped gallium arsenide layer. In the second structure, the first 400 Å of the AlGaAs are undoped, the last 100 Å is doped to $5 \times 10^{18}/cm^3$; the remaining parts of the structure were unchanged. The specific doping levels were chosen to assure identical gate capacitance and threshold voltage levels. A self-consistent calculation was then performed, with the electrons subject to the same scattering mechanisms as those discussed for [9]; namely polar optical, intervalley, ionised impurity, and electron-electron scattering. Fermi statistics is also included. The calculations were performed to reveal the differences in the high field transport in the AlGaAs for the two structures, and thereby to reveal the relative merits of one against another. For a composition of 30% aluminium, the offset voltage was taken as 256 meV.

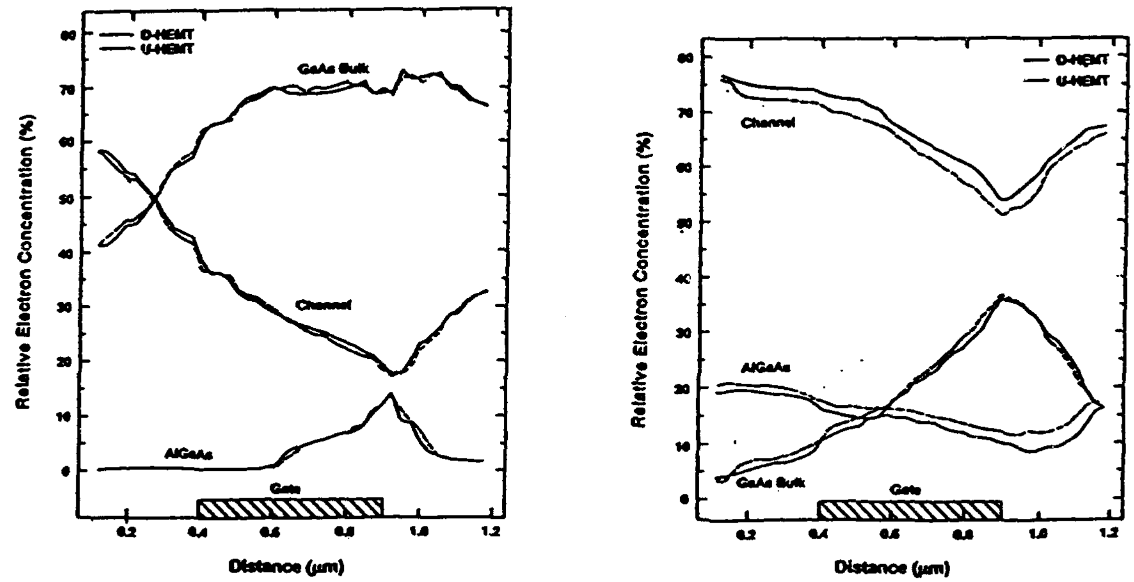
The calculations were performed for a drain bias of 2.0 V and two different values of gate bias, $\pm$0.4 V. A classical calculation of the conduction band profile under the gate contact for a gate bias of 0.4 V is shown in FIGURE 13 (D-HEMT denotes delta doped structure, U-HEMT denotes uniformly doped structure). Note that the higher conduction band levels for the D-HEMT in the vicinity of the gate represent the presence of fewer electrons than for the U-HEMT. The channel electric field profile for this calculation for both structures at a gate bias of 0.4 V is also displayed in FIGURE 13. Note that with the exception of a small region near the drain contact the field profiles are nearly the same, signifying that comparisons of the two structures are relevant.



(a) (b)

FIGURE 13 (a) Conduction band profile along a line perpendicular to the middle of the gate contact for a drain bias of 2.0 V and a gate bias of 0.4 V. The zero coordinate is the Schottky contact. (b) Channel electric field profile along the device. The zero coordinate is at the source. [10].

The observations of FIGURE 14, which are profiles of total electron concentration along: (1) select regions of the AlGaAs (at a distance of 550 Å), (2) the heterointerface (at a distance of 100 Å from physical interface), and (3) the GaAs interior (at a distance of 1900 Å), indicate that under reverse bias there is little distinction between the two structures. However, at a bias of 0.4 V, there is a reduction of electrons in the AlGaAs, and a reduction in the parasitic contributions. A supplementary calculation of [10] shows transconductance levels that are nearly the same under reverse bias, but with substantial improvements under forward bias.



(a) (b)

FIGURE 14 (a) Relative electron concentration within various layers for a drain bias of 2.0 V and a gate bias of -0.4 V. The zero coordinate is at the source. (b) As in (a) but for a bias of 0.4 V. [10].

H. NEGATIVE DIFFERENTIAL RESISTANCE THROUGH REAL SPACE TRANSFER

In the discussion of transport in the MODFETS, we raised the issue of real space transfer. This phenomenon was first discussed in 1979 [11], and occurs as a function of bias in all structures in which transport is parallel to a heterointerface. Consider the superlattice sketch shown in FIGURE 15, and imagine carriers travelling parallel to the interface. If the structure is modulation doped then with the doping only in the wide band gap material, there will be transport in the narrow as well as wide band gap regions, with the mobility of the narrow band gap material being higher because of a reduced number of scattering mechanisms. Under application of an applied bias electrons travelling in the quantum well can acquire a sufficient energy far above their thermal equilibrium value. Electron-electron interactions will help randomise the energy gained in the field direction, and the transfer rate of electrons in the gallium arsenide and the aluminium gallium arsenide will be determined by the thermionic currents from the wide band gap material to the narrow band gap material and vice versa. As in the discussion of the energy gained in the AlGaAs/GaAs MODFETs, electrons in the AlGaAs will not be heated to as high an energy as those electrons in the narrow band gap region, and we may expect that the transfer rate at high fields will not be the same. Indeed more carriers are expected to transfer from the GaAs to the AlGaAs, with the consequent reduction in local mobility.

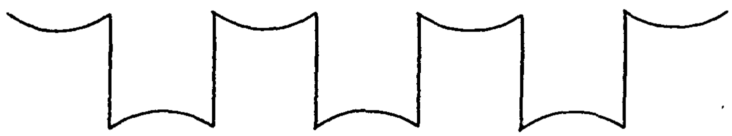


FIGURE 15 Sketch of a superlattice.

Thus we can imagine a device in which the source and drain regions permit transport parallel to the interface, and two components of transport emerge. First: at sufficiently high values of electric field electrons will undergo k-space transfer to subsidiary valleys. For sufficiently long structures this is known to lead to negative differential conductivity and the consequences thereof. Second: real space transfer introduces another component to the negative conductance, where it is possible for gamma valley carriers of sufficient energy in gallium arsenide to transfer to gamma valley carriers in a wide band gap material with a consequent region of negative differential conductivity. It is also possible for L valley carriers in GaAs to transfer to L valley carriers in AlGaAs with a consequent region of negative differential conductivity. Several initial experimental results are summarised in [12]. But perhaps one of the more unusual results associated with real space transfer has been the construction of negative differential resistance transistors NERFETs [13]. These are reviewed in [14].

I. CONCLUSION

The alloy aluminium gallium arsenide is a material whose transport properties under conditions of high fields offer a new dimension in terms of the design of electron devices. Those designs in which transport is perpendicular to the interfaces are based in terms of the energetics of the carriers entering the regions of interest, or, e.g., in the case of HBT upon the reduction of hole injection into the emitter, and the enhancement of electron injection into the base. These devices do not depend upon the specific high field properties of the AlGaAs for their operation. Parallel transport, however, does expose the high field transport properties of AlGaAs. Here while the high field transport properties associated with electron transfer in k-space are present in all devices whose structure size is large enough to sense the NDM region, this is not the feature emphasised.

Rather the effects of negative conductance through real space transfer are emphasised through an understanding of its effect on the operation of such devices as MODFETs, or in the construction of newer devices such as the NERFET.

ACKNOWLEDGEMENT

Portions of this review were sponsored by the US Office of Naval Research and the US Air Force Office of Scientific Research, to whom the author is grateful.

REFERENCES

[1] K.F.Brennan, D.H.Park, K.Hess, M.A.Littlejohn [J. Appl. Phys. (USA) vol.63 (1988) p.5004]

[2] H.L.Grubin ['Transport and material consideration for submicron devices' in 'The Physics of Submicron Semiconductor Devices' Eds H.L.Grubin, D.K.Ferry, C.Jacobnis (Plenum, New York, 1988)]

[3] S.Adachi [J. Appl. Phys. (USA) vol.58 (1985) p.R1]

[4] N.R.Couch, P.H.Beton, M.J.Kelly, T.M.Kerr, D.J.Knight, J.Ondria [Solid-State Electron. (UK) vol.31 (1988) p.613]; N.R.Couch, M.J.Kelly, H.Spooner, T.M.Kerr [Solid-State Electron. (UK) vol.32 (1989) p.1685]

[5] H.L.Grubin, T.R.Govindan [to be published]

[6] G.W.Taylor, J.G.Simmons [IEEE Trans. Electron Devices (USA) vol.32 (1985) p.2345]

[7] M.Meyyappan, J.P.Kreskovsky, H.L.Grubin [Solid-State Electron. (UK) vol.31 (1988) p.1023]

[8] H.Morkoc, H.Unlu [Principles and Technology of MODFETS (J.Wiley and Sons, Chichester, 1991)]

[9] I.C.Kizilyalli, M.Artaki, A.Chandra [IEEE Trans. Electron Devices (USA) vol.38 (1991) p.197]

[10] K.W.Kim, H.Tian, M.A.Littlejohn [IEEE Trans. Electron Devices (USA) vol.38 (1991) p.1737]

[11] K.Hess, H.Morkoc, H.Shichijo, B.G.Streetman [Appl. Phys. Lett. (USA) vol.35 (1979) p.469]

[12] M.Keever et al [Appl. Phys. Lett. (USA) vol.38 (1981) p.36]

[13] A.Kastalsky, R.Bhat, W.K.Chan, M.Koza [Solid-State Electron. (UK) vol.29 (1979) p.1073]

[14] S.Luryi ['Hot-electron injection and resonant tunnelling heterojunction devices' in 'Heterojunction Band Discontinuities, Physics and Device Applications' Eds F.Capasso, G.Mararitondo (North Holland, Amsterdam, 1987)]

11. TRANSIENTS IN QUANTUM WIRES

Dissipation in rectangular quantum wires was studies through Monte Carlo simulations. Optical and acoustic phonons were considered and it was demonstrated that hot-electron cooling is determined by cascade emission of optical phonons followed by a slow second stage of inelastic electron-acoustic phonon interactions (as well as by nonequilibrium hot optical phonons). A copy of a recent paper is enclosed.

Hot-electron relaxation dynamics in quantum wires

R. Gaška, R. Mickevičius, and V. Mitin
Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan 48202

Michael A. Stroscio and Gerald J. Iafrate
U.S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709

H. L. Grubin
Scientific Research Associates Inc., Glastonbury, Connecticut 06033-6058

(Received 16 February 1994; accepted for publication 28 March 1994)

Monte Carlo simulations of hot nonequilibrium electron relaxation in rectangular GaAs quantum wires of different cross sections are carried out. The simulations demonstrate that the initial stage of hot-electron cooling dynamics is determined by cascade emission of optical phonons and exhibits strong dependence on the excitation energy. The second (slow) relaxation stage is controlled by strongly inelastic electron interactions with acoustic phonons as well as by nonequilibrium (hot) optical phonons. The relaxation times obtained in our simulations are in good agreement with the results of recent luminescence experiments. At low electron concentrations where hot phonon effects are negligible the cascade emission of optical phonons may lead to the overcooling of the electron system to temperature below the lattice temperature. These electrons then slowly (during tens of picoseconds) relax to equilibrium due to the interaction with acoustic phonons. At certain excitation energies strong intersubband electron scattering by optical phonons leads to electron redistribution among subbands and intersubband population inversions. If the electron concentration exceeds 10^5 cm^{-1}, hot phonon effects come into play. In contrast to bulk materials and quantum wells, hot phonon effects in quantum wires exhibit strong dependence on the initial broadening of the energy distribution of the electrons. The very initial electron gas relaxation stage in quantum wires is faster in the presence of hot phonons, while for $t>0.5$ ps the hot phonon thermalization time defines the characteristic electron cooling time.

I. INTRODUCTION

Potential possibilities of utilizing unique properties of quasi-one-dimensional (1D) semiconductor structures (peak-like density of states, high packing density, high operation frequencies) for the development of a new generation of electronic and optoelectronic devices have stimulated enhanced interest in investigating the nonequilibrium electron relaxation processes in these structures. Energy dissipation processes of hot electron gases define such characteristics as device operating speed, efficiency, gain, transport characteristics, noise, etc. Although there exist a number of publications dealing with relaxation processes in 1D electron gases,[1–14] the relaxation dynamics of photoexcited carriers under highly nonequilibrium conditions have received less attention. Recent experimental investigations clearly demonstrate different behaviors of photoexcited electron-hole plasmas in 1D and 2D systems.[9–13] Time-resolved luminescence measurements in 1D quantum wires (QWIs)[9–12] indicate that nonequilibrium carrier relaxation to the lowest 1D state is rather slow compared to that measured in quantum wells. Thus, with current interest in developing a new generation of devices based on 1D semiconductor structures it is important to understand the temporal evolution of relaxation of nonequilibrium carriers following initial excitation. This evolution, especially its initial stage (first few picoseconds following pulse excitation), is of great importance for device applications, and, in particular, for high-speed photonic devices. There are several important aspects of electron relaxation in 1D quantum wires to be considered in great detail.

At the operating temperatures of most optoelectronic devices (30–300 K), often the only important energy and momentum relaxation mechanism is the electron-phonon interaction.[4,15] In contrast to 3D or 2D systems, electron-electron pair collisions in QWIs do not affect electron relaxation, whereas electron-phonon interactions remain very strong.[4,16–18] That is why at low electron concentration when the intersubband electron-electron interaction is weak,[4,19] electron scattering by phonons determines the entire electron relaxation dynamics. It appears that phonons in QWIs are far from the same as in bulk materials.[20–22] Along with electron quantization there may exist optical phonon quantization. It has been demonstrated[23] that acoustic phonon scattering in QWIs is essentially inelastic due to the lack of translational symmetry and the resultant uncertainty of momentum conservation. Therefore, in considering the electron dynamic behavior in QWIs it is extremely important to allow for the real phonon spectrum in QWIs.

When electrons are excited well above the bottom of the conduction band they relax via cascade emission of phonons and drive the phonon system out of equilibrium. It is now a commonly accepted notion that nonequilibrium (hot) phonons strongly affect electron transport and relaxation in bulk materials (see, e.g., Ref. 24). Hot phonon effects also explain observations of very slow electron cooling in quantum wells following subpicosecond photoexcitation of hot electrons (see, e.g., Ref. 25). The hot phonon problem in QWIs has been addressed previously in Refs. 2, 6, and 8. However, the kinetic approach used in Ref. 2 does not provide information about relaxation dynamics, while in Ref. 8 the peculiarities related to 1D nature of nonequilibrium electron-phonon system have been overlooked.

In contrast to 3D or 2D systems, the intrasubband electron-electron pair interaction in QWIs leads only to momentum exchange between interaction indistinguishable electrons and does not contribute to the relaxation process. Hence, at least during initial relaxation stage, the electron gas cannot be described by a Maxwellian distribution function and it is necessary to investigate electron relaxation without any *a priori* assumptions about the electron distribution function.[5,6,8] Moreover, the 1D nature of optical phonons in QWIs results in some specific peculiarities of hot phonon buildup that should strongly modify hot phonon effects in QWIs.[6]

In this paper the simulation of the relaxation of hot nonequilibrium electrons has been carried out allowing for all the specific aspects of electron dynamics in QWIs mentioned above.

II. MODEL AND METHOD

Simulation of hot-electron relaxation has been performed in a rectangular GaAs QWIs embedded in AlAs. We employ an ensemble Monte Carlo technique specially suited for 1D electron simulation.[15] A two-dimensional, infinitely deep square potential well confines electrons in the QWI with a multisubband energy structure. The hot-electron energy dissipation model includes electron interactions with confined longitudinal-optical (LO), localized surface (interface) optical (SO) phonons,[17,18] and bulk-like acoustic phonons[23] as well as nonequilibrium optical phonon populations. We have included inelasticity of electron-acoustic phonon scattering in a QWI in full detail using the technique proposed in Ref. 23. The initial distribution of excited electrons among subbands is considered to be defined by the density of states for a given excess energy in each subband.

We start the simulation of electron relaxation after the initial excitation by a short pulse with a duration of 0.1 ps. We have not simulated electron relaxation in coherent regime ($t<100$ fs) which requires a *quantum* mechanical description. Instead, we have focused our attention on the time range $t>0.1$ ps when electrons can be treated semiclassically.[26–29] We are interested primarily in peculiarities of the electron-phonon interaction in QWIs. Therefore, we do not take into account the electron-hole interaction. This situation could take place when electrons are photoexcited from near-monoenergetic impurity levels. The initial state for electron relaxation accounts for the broadening of the electron energy distribution due to two effects: (i) uncertainty in electron initial energy due to the short electron average lifetime at the excited level ($\Delta\epsilon\sim10^{-2}$ eV for $\Delta t\sim10^{-13}$ s); (ii) spectral broadening of the exciting pulse with duration of the order of 10^{-13} s. In accounting for these effects we assume that they both lead to a Gaussian distribution of electron energy at $t=t_0$, which corresponds to the end of the excitation pulse. A Gaussian broadening factor is used instead of a Lorentzian (typically incorporated into idealized theoretical models[26,27]) to prevent the unrealistically large spread of electron energies. This approach has been justified for the electron-phonon interaction in a QWIs demonstrating that the Golden Rule formalism may be retained by convolving a Gaussian broadening function containing a constant broadening factor.[30] We vary the excitation energy ϵ_{ex}, which corresponds to the center of a Gaussian distribution, as well as, $\Delta\epsilon$, the half-width of this distribution.

Nonequilibrium phonons have been included by calculating the phonon occupation number versus phonon wave vector (phonon distribution) within the Monte Carlo procedure. In accordance with the 1D nature of optical phonons in QWIs, the increment of phonon occupation number after each emission (sign +) or absorption (sign −) event is given by the term $\pm(2\pi/\Delta q)(n/N)$, where Δq is the step of the grid in q space used to record the N_q histogram, n is the electron concentration per unit length of a QWI, and N is the actual number of particles in the simulation.

In Monte Carlo simulations of bulk and 2D nonequilibrium electron-optical phonon systems, the mesh interval for the phonon occupation number Δq is not a crucial parameter, given that the interval is much less than the q-space region populated by nonequilibrium phonons which can be easily estimated. This is due to the fact that the phonon reabsorption rate depends on the integrated (average) occupancy over the entire region which is not crucially sensitive to the mesh interval. However, in 1D systems as we will see in a due course, the reabsorption rate depends only on the local value of phonon occupancy N_q at an appropriate q value. Therefore, as the mesh interval becomes smaller, both the local occupancy and the reabsorption rate become larger. This problem is particularly important when considering near-monoenergetic electron excitation. There are, of course, physical limits on the magnitude of Δq. These limits follow from the uncertainty in the phonon longitudinal wave number due to the finite length of the QWI.

We have taken a QWI of length $L_x=10$ μm, so that $\Delta q=2\pi/L_x\sim6\times10^3$ cm^{-1}. Hot phonon thermalization due to the decay of optical phonons into acoustic phonons is taken into account by recalculating N_q for every mesh interval at the end of each time step. It has been demonstrated[31] that the phonon thermalization time τ_{ph} in low-dimensional structures depends weakly on the sizes of the structure and is close to the bulk value. For simulations at $T=30$ and 77 K, we have used the value $\tau_{ph}=7$ ps.[24] The time step in our simulations has been chosen to be smaller than the average time between two events of electron scattering by optical phonons and much less than the phonon thermalization time τ_{ph}. We have not taken into account the increase in the acoustic phonon population as a result of the decay of nonequilibrium optical phonons. There are two reasons for this. First, the buildup of nonequilibrium optical phonons occurs only in a very narrow region of the Brillouin zone (near the zone center), so that over the entire zone the average occupation number increases only negligibly. This is true for systems of any dimensionality since the electrons interacting with phonons populate only the center region of the Brillouin zone. Second, the acoustic phonons in QWIs embedded in surrounding materials with similar elastic properties (GaAs in AlAs in our case) may penetrate through GaAs/AlAs interfaces and escape from the QWI. Therefore, we have ex-

32p

FIG. 1. Time evolution of the mean-electron energy in a QWI with cross section 150×250 Å^2 after excitation at two lattice temperatures: (a) T=77 K and (b) T=300 K. Curve 1 in (a) corresponds to electron excitation energy ϵ_{ex}=20 meV; 2, 42 meV; 3, 67 meV; 4, 100 meV. In (b) 1, 42 meV; 2, 67 meV; 3, 100 meV.

cellent thermal conductivity and the QWI should not be heated much more than the whole GaAs/AlAs structure. Given that the surrounding AlAs is sufficiently massive, the increase in temperature would be negligible even if the QWI strongly radiates acoustic phonons.

III. RESULTS AND DISCUSSION

A. Low electron concentrations

Let us first consider electron concentrations less than 10^5 cm^{-1} where nonequilibrium phonon effects can be neglected.

Calculations with various excitation energies show that in the time scale of 10^{-9} s electron relaxation exhibits one or two distinguishable stages. Figure 1 demonstrates the electron cooling dynamics in a QWI with cross section 150×250 Å^2 for lattice temperatures of T=77 K and T=300 K, as well as for different electron excitation energies ϵ_{ex} counted from the bottom of the lowest conduction subband. The mean electron energy plotted on the vertical axis in Figs. 1(a) and 1(b) is calculated relative to the bottom of the first subband. Thus it consists of two parts: the kinetic energy corresponding to one degree of freedom in a QWI and the intersubband separation energy. The electron excitation energy ϵ_{ex} has been varied from 20 to 100 meV. This implies that for this particular cross section of the QWI, up to the three lowest subbands can be occupied by electrons at the initial time, t=0. One can see from Fig. 1(a) that for electron excitation at 20 meV, the electron gas cooling is slow ("slow" stage). The "fast" stage in the mean electron energy dependence on time is observed when electrons are excited above the optical phonon energy ($\hbar w_{LO}$ or $\hbar w_{SO}$, where $\hbar w_{LO}$ and $\hbar w_{SO}$ are energies of LO and SO phonons, respectively). Electrons initially (in the subpicosecond time scale) cool down losing their energy due to the interaction with optical phonons. Since the optical phonon absorption events at temperature T=77 K are negligibly rare, the electron gas relaxation dynamics is determined by the emission of optical phonons with characteristic times $\tau_{e\text{-}LO}\sim10^{-13}$ s and $\tau_{e\text{-}SO}\sim10^{-12}$ s (for electron-LO and electron-SO phonon interaction, respectively). It is worth to mention that ultrafast nonequilibrium carrier relaxation with characteristic cooling times of the same order of magnitude have been experimentally observed in time-resolved photoluminescence and cathodoluminescence measurements.[9,11,12] At low optical excitation levels the sharp line of band-edge photoluminescence occurs during the laser excitation.[9,12] This implies that hot carriers lose the major portion of their excess energy during the time much shorter than the excitation pulse of 25 ps. The analysis of low-temperature cathodoluminescence spectra suggests that carrier capture and relaxation to the bottom subbands in GaAs QWIs grown on nonplanar substrates occurs in a subpicosecond time scale.[11] At a lattice temperature of T=300 K the electron cooling dynamics is influenced strongly by optical phonon absorption which reduces the electron gas cooling rate [Fig. 1(b)].

The duration of the "fast" relaxation stage as well as the entire electron gas cooling dynamics for $\epsilon_{ex}>\hbar w_{LO}$ exhibits a strong dependence on the excitation energy. As discussed below, when electrons are excited just above the LO phonon energy they cool down to the bottom of the first subband on a subpicosecond time scale [curve 2 in Fig. 1(a) and curve 1 in 1(b)]. Electrons emit optical phonons and occupy states near the subband bottom. Therefore, the mean electron energy drops below that for ϵ_{ex}=20 meV [curve 1 in Fig. 1(a)]. For a lattice temperature of T=300 K we observe anomalous cooling dynamics when electrons occur below the thermal equilibrium energy [curve 1 in Fig. 1(b)]. Overcooling of the electron gas occurs if the electron excitation energy falls into the range $\hbar w_{LO}<\epsilon_{ex}<\hbar w_{LO}+k_BT/2$, where $k_BT/2$ is the electron kinetic energy at a given temperature T corresponding to one degree of freedom in a QWI. At lower temperatures (T=77 K) the transient electron overcooling disappears because the chosen broadening of electron initial energy distribution exceeds the electron thermal equilibrium energy $k_BT/2$.

The "slow" stage of electron relaxation is controlled by the electron interaction with acoustic phonons. Our calculations demonstrate that electron gas thermalization process in a QWI of cross section of 150×250 Å^2 lasts about 1 ns at a

15p

FIG. 2. Time evolution of mean-electron energy after excitation for QWIs with two different cross sections: 40×40 Å^2 (curves 1 and 2); 150×250 Å^2 (curves 3 and 4). Curves 1 and 3 correspond to excitation energy ϵ_{ex}=20 meV; 2 and 4, 100 meV. The lattice temperature is T=77 K.

lattice temperature T=30 K [Fig. 1(a)], and 30 ps at T=300 K [Fig. 1(b)]. This time depends strongly not only on the lattice temperature but also on the cross section of a QWI as does the acoustic phonon scattering rate.[23] The role of acoustic phonon scattering is demonstrated in Fig. 2 for electron cooling dynamics in a QWI with a cross section 40×40 Å^2 compared with the cooling dynamics in a 150×250 Å^2 QWI. The electron energy relaxation due to the interaction with acoustic phonons is much faster in the thin QWI as a result of two factors: (i) the acoustic phonon scattering rate is roughly inversely proportional to the cross section of a QWI, and (ii) the inelasticity of the electron-acoustic phonon interaction also increases with the decrease of the cross section.[23]

The relaxation times of the same order of magnitude have been derived from the time-resolved photoluminescence measurements at low excitation levels.[12] After initial hot-carrier relaxation below optical phonon energy, the further evolution of the band-edge luminescence line shape is characterized by the time of the order of hundreds of picoseconds.[12] Therefore, the electron-acoustic phonon interaction might be responsible for the time evolution of luminescence spectra. Rough estimates yield electron thermalization time due to interaction with acoustic phonons of the order of 500 ps for the structure parameters and temperature (T=5 K) of Ref. 12.

One can see from Fig. 1(a) and Fig. 1(b) that the electron thermal equilibrium energy for T=300 K is larger than could be expected from $k_BT/2$~13 meV, while for T=77 K it practically coincides with $k_BT/2$=3.3 meV. The difference in thermal equilibrium energies comes from the calculation of the electron mean energy in QWIs with multisubband energy structure. Approximately one third of electrons occupy upper subbands in the equilibrium state at a lattice temperature T=300 K due to the Boltzmann distribution. The mean electron energy includes the electron gas kinetic energy ($k_BT/2$) of free motion along the wire and the energy representing the spatial quantization (separation between subbands) in two

31p

FIG. 3. Time evolution of the occupancy of the first subband for the same structure parameters and excitation energies in Fig. 1: (a) T=77 K, (b) T=300 K.

other directions. In the extreme limit of thick QWIs, when a large number of subbands becomes occupied, the electron mean energy tends to $3k_BT/2$ corresponding to the 3D electron gas.

Simulation of hot-electron relaxation dynamics in QWIs demonstrates that intersubband electron scattering primarily by optical phonons leads to a significant carrier redistribution among subbands (Fig. 3). When electrons are excited well above the bottom of the second subband (ϵ_{ex}=100 meV) multiple electron transitions between various subbands due to interaction with optical phonons lead to a nonmonotonous time dependence of the relative occupancy of the first (lowest) subband [curves 3 in Figs. 3(a) and 3(b)]. Acoustic phonon scattering is also responsible for electron intersubband transitions. For the case where the cross section is 150×250 Å^2, the separation between the first and the second subband is less than the optical phonon energy, so that electrons cannot be scattered from the bottom of the second subband by the emission of optical phonons. Accordingly, at low temperatures electron can be "trapped" in the second subband. They are slowly (with characteristic time of tens and hundreds of picoseconds) released from it due to intersubband

15.6p

FIG. 4. Electron distribution in momentum space in a QWI with cross section 150×250 Å^2 3 ps after excitation with $\epsilon_{ex}=67$ meV. Curve 1 represents the electron distribution in the first (lowest) subband; curve 2 represents the second subband. The lattice temperature is $T=77$ K.

electron scattering by acoustic phonons [curves 3 and 4 in Fig. 3(a)]. In this case the second subband serves as a hot-electron reservoir and significantly slows electron cooling. Even in the case when most electrons are excited into the first subband ($\epsilon_{ex}=20$ meV), a small fraction of them (from the high-energy tail of the Gaussian excitation energy distribution) are initially scattered by acoustic phonons to the second subband and then return to the first one [curve 1 in Fig. 3(a)]. The energy of the plateau [curve 3 in Fig. 1(a)] virtually coincides with the position of the second subband with respect to the first subband bottom (27 meV) indicating that electrons are "trapped" there. At high temperatures [$T=300$ K, Fig. 3(b)] electrons "escape" from the upper subbands due to intersubband absorption of optical phonons as well as stronger intersubband acoustic phonon scattering, and reach an equilibrium distribution among subbands in 30 ps.

Under certain excitation conditions intersubband electron scattering by optical phonons may lead to intersubband population inversion. We observe an intersubband population inversion when two conditions are met: (i) separation between two lowest subbands in the QWI is less than minimum optical phonon (LO or SO) energy, so that electrons cannot be scattered from the bottom of the second subband by the emission of optical phonons (we demonstrate results for a QWI with cross section 150×250 Å^2, where this condition is fulfilled); (ii) electrons are excited just above characteristic energy $\epsilon=\epsilon_{12}+\hbar w_{LO}$, where ϵ_{12} is the energy of the bottom of the second subband. Due to a significant difference in the number of final states (peak-like density of states near each subband bottom) electrons from both the first and the second subbands are scattered predominantly into the second subband bottom after the emission of LO phonons. Thus, the number of electrons at the bottom of the second subband exceeds the number of electrons at the bottom of the first subband and a strong intersubband population inversion occurs near the center of the Brillouin zone ($k=0$). Figure 4 presents the distribution of electrons in momentum space for the two lowest subbands 3 ps after excitation. Electrons in the first subband are still hot (wave numbers $k>2\times10^7$ cm^{-1} on curve 1 in Fig. 4) after emission of optical phonons and they relax to the bottom of the subband by interacting with acoustic phonons. Electrons in the second subband occupy states with smaller wave vectors near the subband bottom. This population inversion near the center of the Brillouin zone ($k=0$) lasts about 10 ps at a lattice temperature of $T=77$ K. This time is defined by intrasubband and intersubband electron scattering by acoustic phonons. Intersubband electron scattering by acoustic phonons is responsible for electron release from the second subband at low temperatures where optical phonon absorption is virtually frozen out, while intrasubband acoustic phonon scattering leads to the thermalization of the electron distribution. As one can see from Fig. 4 the population inversion at small electron wave vectors is reduced due to presence of some fraction of electrons near the bottom of the first subband. The number and energy of these electrons depend strongly on the excitation regime. Due to the Gaussian electron excitation energy distribution some electrons from the high-energy tail can emit two optical phonons and cool down to the bottom of the first subband. Thus, as the electron initial energy broadening increases, the occupation of states with small wave vector in the first subband also increases, and the effect of population inversion decreases.

B. Hot phonon effects

Due to optical phonon quantization and the resultant 1D momentum conservation in quantum wires, electrons can emit or absorb optical phonons with wave vectors which are strictly defined by the electron momentum and the phonon energy. In general, the phonon wave number is defined by the energy and momentum conservation equations and is given by

$$q=\sqrt{k^2+k'^2-2kk'\cos\theta}\,, \tag{1}$$

where k is the electron wave number before scattering, $k'=\sqrt{k^2\pm2m^*\omega_0/\hbar}$ is the electron wave number after absorption (sign +) or emission (sign −) of the optical phonon of frequency ω_0, and θ is the angle between electron wave vectors before and after scattering. In 1D structures there are just two final states for scattered electrons: forward scattering with $\cos\theta=1$ or backward scattering with $\cos\theta=-1$. Consequently, there are two possible phonon wave vectors available for emission (and two for absorption) by any single electron:

$$q_1=|k-k'|,\quad q_2=|k+k'|\,. \tag{2}$$

In contrast, in quantum wells (or bulk materials) due to existence of additional degree(s) of freedom, $\cos\theta$ can take any value in the range $(-1,+1)$, so that there is an entire range of a phonon q values from $|k-k'|$ to $k+k'$ available for electron interactions.

Therefore, electrons in QWIs having appreciably different energies generate nonequilibrium phonons in different narrow q-space regions which do not overlap. In turn, these phonons can be reabsorbed only by the electrons that have generated them, unlike in bulk materials and in quantum

15p

FIG. 5. Mean-electron energy as a function of time after initial electron excitation at an energy equal to 4.5 times the LO phonon energy, for two initial widths of the electron energy distribution. Electron concentration is $n=10^5$ cm^{-1} and lattice temperature is $T=30$ K. Solid curves describe the energy evolution for the case of an equilibrium phonon distribution. Results apply to the case of a single-subband QWI neglecting SO phonons.

wells where electrons can reabsorb phonons emitted by other electrons. Consequently, electrons which have different energies cannot interact through the emission and subsequent reabsorption of optical phonons. Thus, electrons and the phonons associated with them (with appropriate wave vectors) are isolated from other electron-phonon pairs if electrons are in different states. Thus, for 1D structures we eliminate one cross-correlation effect which is always present in the nonequilibrium electron-phonon system in bulk materials and in quantum wells and which is important when considering electric noise.[32]

Another consequence of the 1D nature of the electron-phonon interaction in QWIs is that the reabsorption probability for each single electron in the QWI does not depend on the integrated phonon occupation number but only on the local occupancy at a certain q. This reabsorption probability decreases as the phonon distribution spreads over q space (given that the integrated occupancy is defined by the concentration of excited electrons and remains constant). The spread of the nonequilibrium phonon population in q space results from the broadening of the electron energy distribution. As a result, the reabsorption rate and hot phonon effects depend strongly on the energy distribution of excited electrons.

Let us first consider a simplified picture in which there is only one energy subband and SO phonons are neglected; that is, only LO and acoustic phonons are present in the QWI. This simplified picture allows us to refine the pure 1D effect of the broadening of the electron energy distribution on the buildup of hot phonons and, hence, on electron cooling dynamics. Figure 5 illustrates electron cooling dynamics in a 150×250 Å^2 QWI at $T=30$ K following initial electron excitation at an energy 4.5 times the LO phonon energy for two different Gaussian electron distribution half widths: 30 and 4 meV. For comparison, we plot the electron relaxation dynamics without nonequilibrium optical phonons. When hot phonons are neglected the cooling dynamics displays two distinguishable stages: the fast stage (with subpicosecond duration) due to the cascade emission of optical phonons, and the second, slow stage of electron thermalization due to interactions with acoustic phonons. It must be noted, however, that in the given time scale of 10 ps, acoustic phonon scattering does not visibly influence the electron relaxation dynamics in this QWI with a rather large cross section of 150×250 Å^2. As one can see from Fig. 5 the very initial relaxation stage ($t<0.5$ ps) is faster in the presence of hot phonons. The higher nonequilibrium phonon populations are created (4 meV), the faster is the very initial relaxation stage. This effect can be understood if one first considers the temperature dependence of the relaxation rate. At high temperatures both the emission and absorption rates are higher. This leads to fast energy redistribution of excited electrons. The cooling rate of electrons which emit optical phonons increases and that of electrons which absorb phonons decreases because of the $\epsilon^{-1/2}$ energy dependence of 1D density of states and scattering rates. The increase, however, is faster than the decrease due to the same $\epsilon^{-1/2}$ function. Therefore, the total 1D electron gas cooling rate increases when the electron energy redistributes due to emission and absorption of optical phonons. At very high temperatures this happens on a very short time scale while electron distribution at low temperatures still remains unchanged. Hence, the very initial electron cooling rate in QWIs increases when increasing the lattice temperature, provided that electrons are excited well above optical phonon energy and thermal equilibrium energy. (Note that in bulk materials, where the emission and absorption rates increase with energy the relaxation rate is higher at low lattice temperatures. There should be no temperature dependence of the initial relaxation rate in 2D systems.) To observe an appreciable temperature effect on the relaxation rate it is necessary that phonon occupation number be greater than 1. Under phonon equilibrium such occupation numbers could even be unachievable in a solid state. However, due to strong buildup of nonequilibrium phonons at high excited electron concentrations the occupation number for certain phonon modes may be considerably higher than 1. This is why the initial relaxation is faster for higher nonequilibrium phonon occupations and thus, for narrower initial electron energy distributions (Fig. 5).

One can see from Fig. 5 that the onset of hot phonons leads to a substantial reduction of the electron gas cooling rate for $t>0.5$ ps due to strong reabsorption of nonequilibrium phonons. The onset of hot phonons occurs sooner if the electron energy distribution is narrower (4 meV). Hence, electron cooling is slower for narrow electron distributions. The effect of narrowing of the electron energy distribution is similar to that of increasing the electron concentration and, as we have already discussed, it is a purely 1D effect. As has been demonstrated in a previous subsection, in QWIs with small cross sections (40×40 Å^2) the acoustic phonon scattering rate is higher and this scattering is much more inelastic than in QWIs with large cross sections (150×250 Å^2). Therefore, in a QWI with a 40×40 Å^2 cross section, acoustic phonon scattering is a very effective energy dissipation mechanism and it is responsible for fast relaxation and

15p

FIG. 6. Mean electron energy vs time for a realistic multisubband QWI structure with both LO and SO optical modes included as well as acoustic phonons. Electron concentration is equal $n=10^6$ cm^{-1}, other parameters and conditions are the same as in Fig. 5.

15p

FIG. 7. Time evolution of the relative occupation of the first (lowest) subband. Parameters are the same as in Fig. 6.

smearing out the effect of the initial broadening of the electron distribution on the cooling rate.

We have also considered the realistic case where the multisubband structure of the QWI is taken into account along with all possible optical phonon modes (LO and SO). Figure 6 shows the electron cooling dynamics in this realistic structure for 4 and 30 meV electron excitation linewidths. The dependence of hot phonon buildup on the electron distribution broadening is washed out almost completely in this realistic structure due to various intrasubband and intersubband transitions assisted by the LO and the two SO modes. The reason is that nonequilibrium phonon peaks in q space in this case overlap and form a complex broad distribution in q space virtually independent of initial electron distribution. The main effect which comes into play within this realistic model is the dependence of the number of the upper subbands involved in electron cooling on the initial electron energy distribution. Figure 7 demonstrates the time evolution of the subband filling by electrons. In the case of a broad electron initial distribution (30 meV) there are more subbands occupied by electrons scattered from the high-energy tail. Therefore, the return of electrons to the first subband is slower than for a narrow electron distribution (4 meV). Hot phonons lead to stronger intersubband electron redistribution and slower return to the lowest subband. By comparing Figs. 6 and 7 one notices that the different occupation of subbands for 4 and 30 meV excitation linewidths virtually does not affect the mean-electron energies which coincide for both excitation regimes after 3 ps following excitation. This, at the first glance, strange behavior is related to the fact that the electron kinetic energy related to 1D free motion in each subband is higher for 4 meV excitation linewidth due to stronger hot phonon effects. This difference in kinetic energies is compensated by the higher occupation of the upper subbands in the case of a 30 meV linewidth. Consequently, the 4 meV curve in Fig. 6 contains a larger part of kinetic energy and a smaller part of "potential" energy than the 30 meV curve.

IV. SUMMARY

Simulation reveals complex dependence of hot-electron gas cooling dynamics on excitation energy, lattice temperature, and structure parameters of QWI. Electron relaxation in subpicosecond time scale is controlled by their interaction with confined optical phonons, whereas thermalization of the electron distribution is defined by essentially inelastic electron-acoustic phonon scattering. Electron gas thermalization is much faster in a thinner QWI due to higher acoustic-phonon scattering rate and stronger inelasticity of electron-acoustic phonon interaction. The relaxation times obtained in our simulations are in good agreement with the results of optical measurements.[9,11,12]

Variation of the initial electron energy substantially changes the entire picture of hot-electron relaxation due to the interaction of electrons with various phonons in QWIs. Calculations demonstrate potential possibilities of two effects: electron gas overcooling and dynamic intersubband population inversion. Both effects exhibit strong dependence on the lattice temperature.

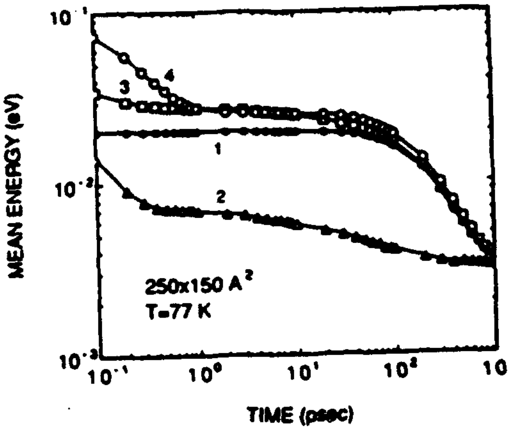
Population inversion is well pronounced at low temperatures, whereas electron gas overcooling benefits from high temperatures. At low temperatures electrons can be "trapped" in the upper subbands below the optical phonon energy and stay there for quite a long time defined by intersubband electron-acoustic phonon interaction.

We have found that hot phonon effects in QWIs are well pronounced for electron concentrations equal to or higher than 10^5 cm^{-1} and depend strongly on the energy distribution of excited electrons. Hot phonon effects become weaker as the broadening of the excited electron energy distribution increases. This result is in complete contrast to the case of bulk materials and quantum wells where the energy distribution of excited electrons virtually does not affect the buildup of nonequilibrium phonons and the reabsorption rate.

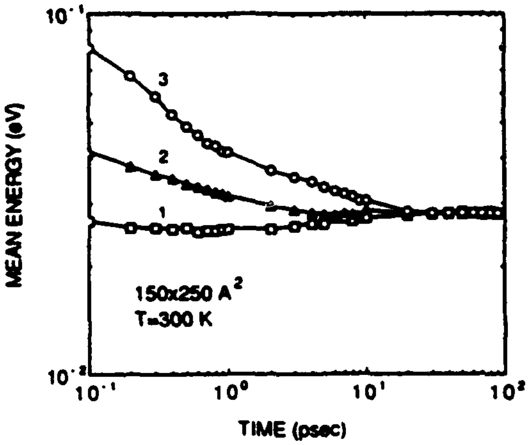
ACKNOWLEDGMENTS

The work was supported by the U.S. Army Research Office, the National Science Foundation, and the Center for Compound Semiconductor Microelectronics of the University of Illinois at Urbana-Champaign.

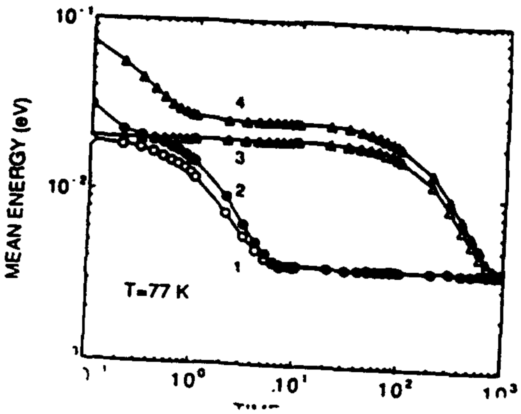
[1] V. B. Campos and S. Das Sarma, Phys. Rev. B 45, 3898 (1992).
[2] V. B. Campos, S. Das Sarma, and M. A. Stroscio, Phys. Rev. B 46, 3849 (1992).
[3] D. Jovanovic, J. P. Leburton, K. Ismail, J. M. Bigelow, and M. H. Degani, Appl. Phys. Lett. 62, 2824; J. P. Leburton, S. Briggs, and D. Jovanovic, Superlattice Microstructure 8, 209 (1990).
[4] I. Vurgaftman and J. Singh, Appl. Phys. Lett. 62, 2251 (1993).
[5] R. Gaška, R. Mickevičius, V. Mitin, and H. L. Grubin, Semicond. Sci. Technol. (1994, to be published).
[6] R. Mickevičius, R. Gaška, V. Mitin, M. A. Stroscio, and G. J. Iafrate, Semicond. Sci. Technol. (1994, to be published).
[7] I. Vurgaftman and J. Singh, Semicond. Sci. Technol. (1994, to be published).
[8] L. Rota, F. Rossi, P. Lugli, and E. Molinari, Semicond. Sci. Technol. (1994, to be published).
[9] R. Cingolani, H. Lage, L. Tapfer, D. Heitmann, and K. Ploog, Semicond. Sci. Technol. 7, B287 (1992).
[10] A. C. Maciel, J. F. Ryan, R. Rinaldi, R. Cingolani, M. Ferrara, U. Marti, D. Martin, F. Morier-Gemoud, and F. K. Reinhart, Semicond. Sci. Technol. (1994, to be published).
[11] J. Christen, M. Grundmann, E. Kapon, E. Colas, D. Hwang, and D. Bimberg, Appl. Phys. Lett. 61, 1611 (1992).
[12] R. Cingolani, R. Rinaldi, M. Ferrara, G. C. La Rocca, H. Lage, D. Heitmann, K. Ploog, and H. Kalt, Phys. Rev. B 48, 14331 (1993).
[13] A. Endoh, H. Arimoto, Y. Sugiyama, H. Kitada, and S. Muto, Appl. Phys. Lett. 64, 449 (1994).
[14] R. Cingolani, R. Rinaldi, M. Ferrara, G. C. La Rocca, H. Lage, D. Heitmann, K. Ploog, and H. Kalt, Semicond. Sci. Technol. (1994, to be published).
[15] R. Mickevičius, V. Mitin, K. W. Kim, and M. A. Stroscio, Semicond. Sci. Technol. 7, B299 (1992).
[16] H. Sakaki, Jpn. J. Appl. Phys. 19, L735 (1980).
[17] K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevičius, and V. V. Mitin, J. Appl. Phys. 70, 319 (1991).
[18] R. Mickevičius, V. Mitin, K. W. Kim, M. A. Stroscio, and G. J. Iafrate, J. Phys. Condens. Matter 4, 4959 (1992).
[19] Yu. M. Sirenko and P. Vasilopoulos, Phys. Rev. B 46, 1611 (1992).
[20] N. Mori and T. Ando, Phys. Rev. B 40, 6175 (1989).
[21] H. Rücker, E. Molinari, and P. Lugli, Phys. Rev. B 44, 3463 (1991).
[22] P. A. Knipp and T. L. Reinecke, Phys. Rev. B 45, 9091 (1992).
[23] R. Mickevičius and V. Mitin, Phys. Rev. B 48, 17194 (1993).
[24] R. Mickevičius and A. Reklaitis, J. Phys. Condens. Matter 1, 9401 (1989).
[25] D. J. Westland, J. F. Ryan, M. D. Scott, J. I. Davies, and J. R. Riffat, Solid-State Electron. 31, 431 (1988).
[26] A. V. Kuznetsov, Phys. Rev. B 44, 8721 (1991).
[27] T. Kuhn and F. Rossi, Phys. Rev. B 46, 7469 (1992).
[28] C. J. Stanton, A. V. Kuznetsov, and C. S. Kim, Semicond. Sci. Technol. (1994, to be published).
[29] J. Schilp, T. Kuhn, and G. Mahler, Semicond. Sci. Technol. (1994, to be published).
[30] J. P. Leburton and D. Jovanovic, Semicond. Sci. Technol. 7, B202 (1992).
[31] G. Mahler, A. M. Kriman, and D. K. Ferry (unpublished).
[32] P. Bordone, L. Varani, L. Reggiani, and T. Kuhn, in *AIP Conference Proceedings on Noise in Physical Systems and 1/f Fluctuations, St. Louis, MO, 1993* (AIP, New York, 1993), p. 657.

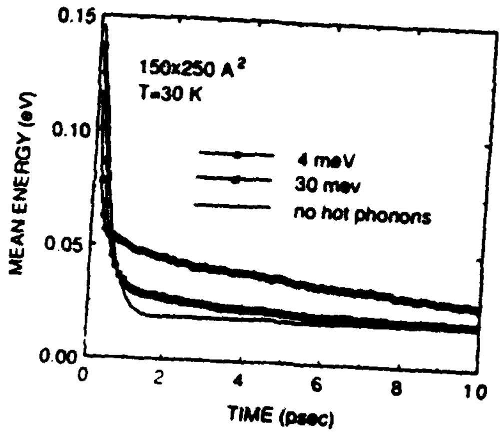


JAP/049414/FIG1A/607

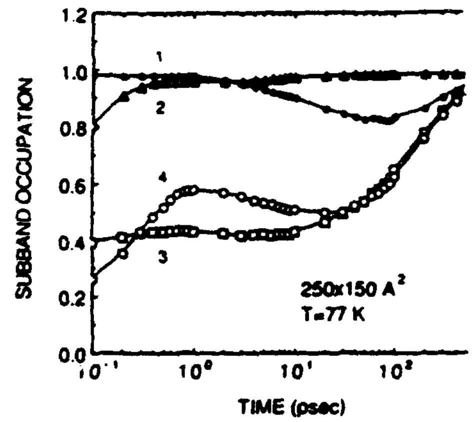


JAP/049414/FIG1B/607

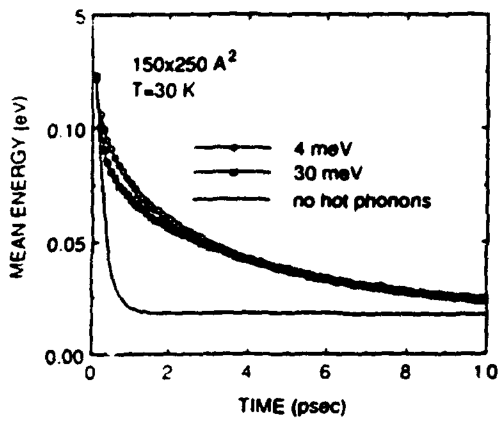




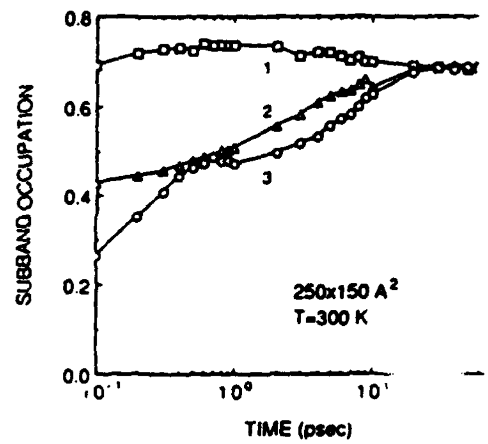
JAP/049414/FIG5/60?



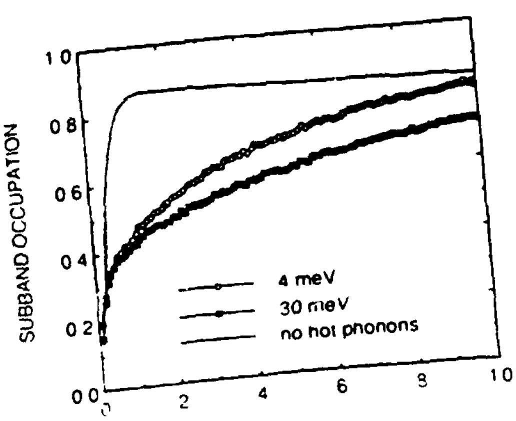
JAP/049414/FIG3A/60?

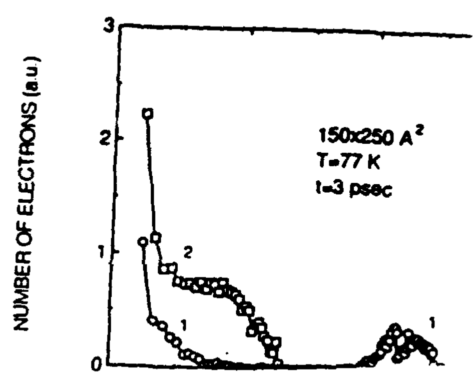


JAP/049414/FIG6/60?



JAP/049414/FIG3B/60?





12. RECOMMENDATIONS

The density matrix in the coordinate representation is an extraordinary means of studying quantum transport in nanoscale devices. While not discussed in this document this approach is being used to study the formation of barriers at metal/semiconductor interfaces. Presently, the density matrix in the coordinate representation, along with Green's function approaches are the only quantum device simulations explicitly accounting for dissipation. The density matrix is also the only quantum transport procedure presently capable of handling both electron and hole transport. Its ability to directly compute current from the nonequilibrium quantum distribution function is a significant advance over those current algorithms that employ the Tsu-Esaki formulation. Further, the presence of menu-driven algorithms for both workstations and PCs suggests that the approach taken at SRA will provide a significant design tool.