Analytic solution of the Boltzmann equation with applications to electron transport in inhomogeneous semiconductors

S. A. Trugman
Department of Physics, Princeton University, Princeton, New Jersey 08544

A. J. Taylor
AT&T Bell Laboratories, Holmdel, New Jersey 07733
(Received 18 September 1985)

An analytic solution has been obtained to the one-dimensional Boltzmann transport equation in the relaxation-time approximation that is exact to first order in an arbitrary position- and time-dependent potential. This is the first practical method for studying time-dependent transport in inhomogeneous systems. The solution is correct for general relaxation times and arbitrarily large external electric fields, including the ballistic limit. This solution is used to study steady-state transport in two submicron structures: a doping superlattice, where the doping density is a sinusoidal function of position, and an $N^+N^-N^+$ GaAs junction. The distribution function for the $N^+N^-N^+$ device exhibits a ballistic peak by a new mechanism, in a regime where the potential has no local maximum to "skim" electrons. At zero temperature, the ballistic peak is a true singularity, rather than just a qualitative feature. Evidence is presented that an inhomogeneous potential causes a coherent excitation of plasmons. Analytic solutions have also been obtained, within these approximations, for initially nonequilibrium carrier distributions in the presence of potentials with arbitrary time and position dependence.

I. INTRODUCTION

There has been an increasing interest in understanding electron transport in submicron structures, which has been motivated in part by the development of new artificial materials such as epitaxially grown superlattices and high-density integrated circuits.[1] Transport in these systems differs from that in larger structures primarily in that carriers in submicron devices can have mean free paths that are large compared to the spatial extent of the device and can also be subject to large electric fields. Approximate solutions for the hot electron distribution derived from the drift-diffusion equation[2] or from an assumption of the drifted Maxwellian distribution[3] have been shown to be inaccurate in this regime. Reliable solutions have been obtained thus far by Monte-Carlo simulations[4] and by numerical solution of the Boltzmann equation.[5] These methods require large amounts of computer time, and numerical methods require algorithms that avoid the inherent instabilities present in seeking self-consistent solutions to the Boltzmann equation.

We have obtained an exact analytic solution to the one-dimensional Boltzmann transport equation in the relaxation time approximation.[6] The solution is valid to first order in an arbitrary position-dependent potential (which may be caused, for example, by inhomogeneous doping) and is given in terms of special functions. It applies to large external electric fields and general relaxation times, including the ballistic limit of long relaxation times and strong fields. It is fully self-consistent in that it takes into account the modification of the potential due to electron rearrangement. The steady-state solution is derived in Sec. II. In Sec. III, we have used the solutions to model transport for a doping superlattice and for an $N^+N^-N^+$ GaAs junction. The particular $N^+N^-N^+$ structure is chosen so that a qualitative comparison can be made with other theoretical[4,5] and experimental[7(a)] studies. We observe a ballistic peak (and an "antipeak") in this structure in spite of the restriction of a weak doping potential (since our solution is only valid to first order). This is the first observation of a ballistic peak for a potential that does not possess a local maximum. We also discuss evidence for the coherent excitation of plasmons.

Using the same approximations described above, we also discuss time-dependent solutions to the Boltzmann transport equation. In Sec. IV we describe the linear response to the Boltzmann equation for a general weak potential varying in both time and space, where the initial electron distribution is specified in the distant past. This solution is important when considering the response of devices to ac electric fields. In Sec. V we describe to first order the relaxation of an arbitrary initial distribution, that may be specified at any time, in the presence of inhomogeneous, time-varying potentials. This could be applied to describe the transient behavior of submicron devices, as well as the behavior of such devices as fast photodetectors, where free carriers are initially created through the absorption process. These solutions provide the first practical method for studying the time-dependent response of inhomogeneous systems.

II. STEADY-STATE THEORY

The method used for this calculation is to first solve for the exact electron distribution in a uniform electric field,

and then to treat a general position-dependent potential as a perturbation. The starting point is the Boltzmann equation for the electron distribution $n(x,p)$ in a one-dimensional static potential $V_{\rm tot}(x)$:

$$\frac{\partial n}{\partial t}=\frac{\partial n}{\partial p}\frac{\partial V_{\rm tot}}{\partial x}-\frac{\partial n}{\partial x}\frac{p}{m}+\Gamma_{\rm coll}\ . \tag{1}$$

Recent work has shown that the Boltzmann equation accurately describes transport even in the presence of large electric fields.[7(b)] We use the relaxation time approximation for the collision term $\Gamma_{\rm coll}$, where it is assumed that $n(x,p)$ relaxes to its equilibrium value determined by the local density and lattice temperature T, and that this process can be characterized by an energy-independent relaxation time τ.[7(c)] The relaxation time approximation employed is identical to that used by Baranger and Wilkins [Ref. (5)]. The approximation, while not technically realistic, emphasizes the effects of inhomogeneities by being itself structureless. (Note, however, that the optical phonon scattering rates are only weakly energy dependent at the energies where structure appears in the distribution function.) The approximation overestimates the effectiveness of a collision in bringing a particle to equilibrium, so that it may underestimate the nonequilibrium effects. We therefore expect the qualitative features described below to persist in the presence of more realistic scattering processes.

In this section, we consider for simplicity only steady-state solutions, $\partial n/\partial t=0$. Time-dependent solutions are discussed in Secs. IV and V. By further assuming a parabolic band structure described by an effective mass m, Eq. (1) becomes

$$0=\frac{\partial n}{\partial p}\frac{\partial V_{\rm tot}}{\partial x}-\frac{\partial n}{\partial x}\frac{p}{m}-\frac{1}{\tau}\left[n-\left[\frac{\beta}{2\pi m}\right]^{1/2}\exp\left[\frac{-\beta p^2}{2m}\right]\int_{-\infty}^{+\infty}n(x,p)dp\right], \tag{2}$$

where $\beta=(k_BT)^{-1}$. Note that the expression for $\Gamma_{\rm coll}$ ensures particle conservation (the current continuity equation, $\partial J/\partial x=0$, is satisfied).

The gradient of the potential $\partial V_{\rm tot}/\partial x$ is of the form

$$\frac{\partial V_{\rm tot}}{\partial x}=\alpha\left[-1+\frac{1}{\alpha}\frac{\partial V_D}{\partial x}\right]+\frac{\partial \Delta V}{\partial x}\ . \tag{3}$$

The parameter α is the force due to the applied uniform electric field in the device, $V_D(x)$ is the potential due to inhomogeneous doping, and $\Delta V(x)$ is the potential caused by electron rearrangement, assuring that Eq. (2) is self-consistent. We define $V_{\rm ext}\equiv V_D+\Delta V$. Our major assumption in this work is that $\partial V_{\rm ext}/\partial x$ due to the doping potential and electron rearrangement is weak compared to the applied electric field α, thereby allowing a first-order solution. Under this assumption, Eq. (2) is simplified if $V_D(x)$ and $n(x,p)$ are expanded in terms of Fourier components such that

$$V_D(x)=\sum_{k\neq 0}V_k e^{ikx}\ , \tag{4}$$

resulting in

$$\frac{\partial V_{\rm tot}}{\partial x}=\alpha\left[-1+\sum_{k\neq 0}\lambda_k e^{ikx}+\sum_{k\neq 0}\lambda_k\Delta_k e^{ikx}\right], \tag{5}$$

and

$$n(x,p)=n_0(p)+\delta n(x,p)=n_0(p)+\sum_{k\neq 0}\lambda_k g(k,p)e^{ikx}\ . \tag{6}$$

We have defined $\lambda_k=ikV_k/\alpha$. With no loss of generality, we have chosen the zero of V_D such that $V_{k=0}=0$. The variable Δ_k is chosen so that the last sum in Eq. (5) is $\partial(\Delta V)/\partial x$; Δ_k is given by Eq. (11) below. Equation (6) defines $g(k,p)$. The function $n_0(p)$ is the solution of Eq. (2) in the absence of the inhomogeneous potential $V_{\rm ext}$:

$$n_0(p)=\frac{N_0}{2\alpha\tau}\exp\left[-y+\frac{1}{4B}\right]\left\{\Phi\left[\sqrt{B}\left[y-\frac{1}{2B}\right]\right]+1\right\}, \tag{7}$$

where the dimensionless parameters are

$$y=\frac{p}{\alpha\tau}\ , \tag{8a}$$

$$a^2=\frac{m}{2\,|\,k\,|\,\alpha\tau^2}\ , \tag{8b}$$

$$B=\frac{\beta(\alpha\tau)^2}{2m}\ . \tag{8c}$$

The variable y is the momentum measured in terms of the momentum gained during acceleration by the field between scattering events and a^2 is proportional to the ratio of the wavelength of the potential to the mean free path under the uniform field. B is the inverse temperature measured in units of the inverse of the kinetic energy gained from the uniform electric field during the time between scattering events. $\Phi(x)$ is the error function (defined in Ref. 8, p. 930) and N_0 is the average electron density. Note that even to zeroth order in λ, the distribution is not, in general, a drifted Maxwellian. The function $\delta n(x,p)$ is the part of the solution of interest and can be written as

$$\delta n(x,p)=2\sum_{k>0}\mathrm{Re}[\lambda_k g(k,p)e^{ikx}]\ . \tag{9}$$

Equation (9) follows from Eq. (6) by noting that since V_D is real, $\lambda_{-k}=\lambda_k^*$ and $g(-k,p)=g^*(k,p)$. We solve for δn to first order in λ. The Fourier components decouple to this order, so that each Fourier component may be considered separately.

It is convenient here to introduce the dimensionless function

$$g_1(k,y)\equiv(\alpha\tau/N_0)g(k,p=\alpha\tau y)\ , \tag{10a}$$

and the parameter

$$\gamma_k\equiv\int_{-\infty}^{+\infty}g_1(k,y)dy\ . \tag{10b}$$

$\mathrm{Re}(\gamma_k)[-\mathrm{Im}(\gamma_k)]$ is the total carrier density change out of phase (in phase) with the potential of the kth Fourier component. The remaining equations in this section (and their analogues in Secs. IV and V) apply for $k>0$. For $k<0$, use the relations $g_1(-k,y)=g_1^*(k,y)$ and $\gamma_{-k}=\gamma_k^*$. Δ_k can then be found using Poisson's equation

$$\Delta_k=\frac{4\pi i e^2\gamma_k N_0}{\epsilon_0 k\alpha}=ia^2\delta\gamma_k\ , \tag{11a}$$

where

$$\delta=\frac{8\pi N_0 e^2\tau^2}{m\epsilon_0} \tag{11b}$$

is twice the square of the plasma frequency times scattering time and ϵ_0 is the static dielectric constant. By collecting the $O(\lambda)$ terms from Eq. (2), the following ordinary differential equation for $g_1(k,y)$ results in

$$\left[\frac{d}{dy}+1+\frac{iy}{2a^2}\right]g_1=\left[\frac{B}{\pi}\right]^{1/2}[1+\gamma_k(1+ia^2\delta)]\exp(-By^2)$$
$$-\tfrac{1}{2}(1+ia^2\delta\gamma_k)\exp\left[-y+\frac{1}{4B}\right]\left\{\Phi\left[\sqrt{B}\left[y-\frac{1}{2B}\right]\right]+1\right\}. \tag{12}$$

The solution can be written in terms of one-dimensional integrals as

$$g_1(k,y)=\exp\left[-y-\frac{iy^2}{4a^2}\right]\int_{-\infty}^{y}dx\left\{\left[\frac{B}{\pi}\right]^{1/2}[1+\gamma_k(1+ia^2\delta)]\exp(-Bx^2)\right.$$
$$\left.-\tfrac{1}{2}(1+ia^2\delta\gamma_k)\exp\left[-x+\frac{1}{4B}\right]\left\{\Phi\left[\sqrt{B}\left[x-\frac{1}{2B}\right]\right]+1\right\}\right\}\exp\left[x+\frac{ix^2}{4a^2}\right]. \tag{13}$$

Note that the value of the complex number γ_k must be determined by substitution of g_1 into Eq. (10b).[9]

At finite temperatures, a one-dimensional integral must be done numerically. However, in the zero-temperature limit ($B=\infty$), the integrals for both $g_1(k,y)$ and γ_k can be expressed in closed form in terms of special functions:

$$n_0(p)=\frac{N_0}{\alpha\tau}\exp(-y)\Theta(y)\ , \tag{14}$$

$$g_1(k,y)=\exp\left[-y-\frac{iy^2}{4a^2}\right]\Theta(y)\{1+\gamma_k(1+ia^2\delta)-a\sqrt{2\pi}(1+ia^2\delta\gamma_k)[C(y/2a)+iS(y/2a)]\}\ , \tag{15}$$

$$\gamma_k=\left[\left[\frac{\pi}{2}\right]^{1/2}\{(1-i)+2i[C(a)+iS(a)]\}\right.$$
$$\left.-ia[\mathrm{ci}(a^2)+i\,\mathrm{si}(a^2)]\right]\Big/\left[\frac{e^{ia^2}}{a}-\left[\frac{\pi}{2}\right]^{1/2}(1+i\delta a^2)\{(1-i)+2i[C(a)+iS(a)]\}-\delta a^3[\mathrm{ci}(a^2)+i\,\mathrm{si}(a^2)]\right]. \tag{16}$$

$C(x)$ and $S(x)$ are Fresnel integrals, while $\mathrm{ci}(x)$ and $\mathrm{si}(x)$ are the cosine and sine integrals,[8] and $\Theta(x)$ is the unit step function. Equation (14) is the zero-temperature limit of Eq. (7). Even at $T=0$, the strong electric field spreads the electrons over so many states that it is legitimate to use Boltzmann statistics, as we have done, except at very high densities. The unperturbed distribution function described in Eq. (14) contains no left-moving carriers because, after suffering a collision, the electrons emerge at rest (since $T=0$) and are then accelerated to the right by the electric field. We emphasize that our analytic solution [Eq. (13)] is not limited to situations in which all carriers move in one direction. Carriers in fact move in both directions for nonzero temperatures.

It is straightforward to show that a position-dependent doping field, $\partial V_D/\partial x$, which is $O(\lambda)$ results in a change in current that is $O(\lambda^2)$. (Changing λ to $-\lambda$ is equivalent to changing the origin of a cosine, which does not affect the current.) Thus, our $O(\lambda)$ calculation cannot determine the leading modification in the dc I-V characteristic. The fact that the change in current is $O(\lambda^2)$ may however help to explain why the I-V characteristic is not a sensitive probe of position-dependent doping.[5,10,11] The dc result is to be contrasted with the finite frequency result for a time-dependent potential, which is treated in Secs. IV and V. The ac current is $O(\lambda)$ and *can* be determined by our methods.

III. APPLICATIONS

We now consider two applications of the solutions that were derived in Sec. II. The first is the solution for a sine

potential in an external electric field. This potential would arise in a doping superlattice where the doping density is a sinusoidal function of position. This example is also of interest because the solution for a general potential is obtained by a superposition of the separate Fourier components. The second application is to an $N^+N^-N^+$ structure, where the doping density in a narrow slab is lower than that in the wider contacts. We qualitatively compare our solution with that of Baranger and Wilkins.[5] Both applications are done at zero temperature, because qualitative features are most clearly evident and are not smeared by finite temperature, and because our formalism is simpler at $T=0$.

A. Sine potential

For a sinusoidal potential V_D is given by $\mathrm{Re}(\lambda\alpha e^{ikx}/ik)$. Since this doping potential contains only one Fourier component, the distribution function is given by

$$n(x,p)=n_0(p)+\lambda e^{ikx}g(k,p)\ , \tag{17}$$

where n_0 and g are given by Eqs. (14)—(15). We focus our attention on the first-order response, g. The real part of g is the response in phase with the gradient of the doping potential, and the imaginary part is the response out of phase. Some limiting cases of the dimensionless distribution function $g_1(y)$ are easily understood. The limit $a\rightarrow\infty$ is the high drag, small electric field limit in which the mean free path is short compared with the wavelength of the external potential. This is essentially a local problem, which could have been treated more simply by other means had it been our only aim. When the electron density approaches zero ($\delta\ll a^{-2}$), each electron moves independently under a force

$$F=\alpha-\lambda\alpha e^{ikx}\ . \tag{18}$$

At each position, the average electron velocity is proportional to the local force; by the continuity equation, the density is proportional to the inverse of the force. It is straightforward to show that in this limit $g_1(y)$ is real and given by $g_1(y)=(2-y)\exp(-y)\Theta(y)$ [see Fig. 1(a)]. The resulting distribution g_1 at every point is the one appropriate to the local electric field. As δ becomes large, there is a large additional contribution to the force given by Eq. (18) from the potential $\Delta V(x)$ due to the other electrons. For δ sufficiently large ($\delta\gg a^{-2}$), the electrons move so as to screen the potential V_D due to inhomogeneous doping.

The low-drag, high electric field limit of $a\rightarrow 0$ is more relevant to submicron structures. In this limit, the mean free path is large compared to the scale on which the doping potential changes, and $g_1(y)$ shows a great deal more structure [see Fig. 1(b)]. Local maxima in $\mathrm{Re}[g_1(y)]$ occur at a sequence $y_1,y_2,\ldots$. For large j, the y_j are given by

$$y_j\approx(8\pi j)^{1/2}a\ , \tag{19}$$

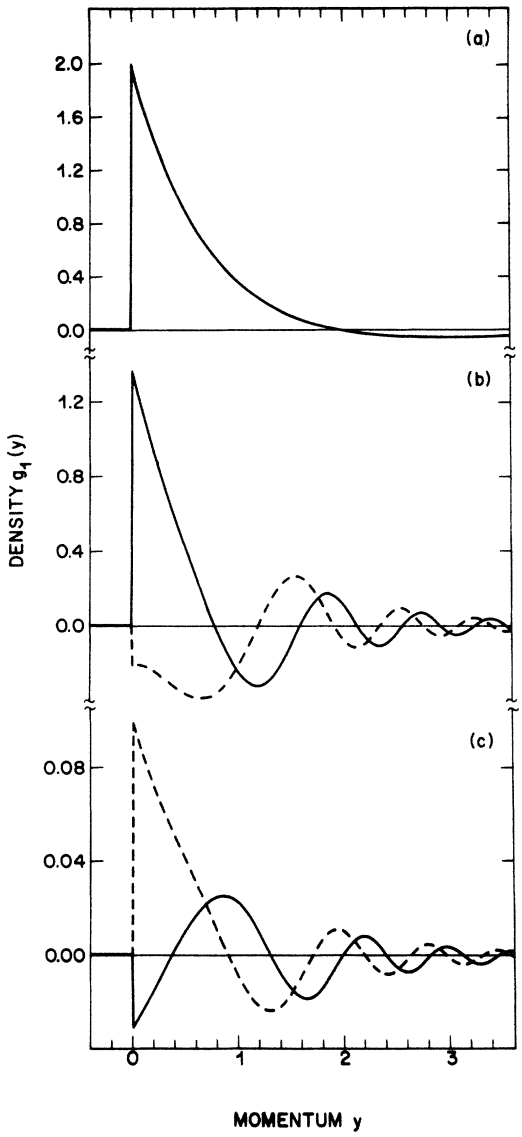

FIG. 1. First-order response of the electron distribution function $g_1(y)$ is plotted as a function of the dimensionless momentum y for a sinusoidal potential. The real part of $g_1(y)$ (solid line) is in phase with the gradient of the doping potential and the imaginary part (dashed line) is out of phase. (a) The parameters $a=20$, $\delta=0$ describe a high-drag superlattice with a very low carrier density. The imaginary part of g_1 vanishes on this scale. (b) The parameters $a=0.4$, $\delta=0$ describe a low-drag superlattice with a mean free path comparable to the lattice spacing and a very low carrier density. The qualitatively periodic structure for larger momentum y is caused by electrons that have traveled several lattice spacings since their last collision. The period decreases as y increases. (c) The parameters $a=0.4$, $\delta=200$ describe the same low drag as (b), but with a high carrier density (3.2×10^{17} cm^{-3} for the GaAs parameters of Sec. III B). In contrast with (b), the area under $\mathrm{Re}(g_1)$ is very small compared with that under $\mathrm{Im}(g_1)$, indicating that the doping potential is being screened. The smaller vertical scale of Fig. 1(c) results from the fact that the electron density is sufficiently high that a small fraction of them can screen the doping potential.

These local maxima are "echoes" caused by electrons that last equilibrated a time $t_j=\tau y_j$ earlier near the jth previous local maximum of the potential.

One feature of Fig. 1(b) is that the number of low-momentum electrons is largest near the point in space where the total force is the smallest. The reason for this can be seen most clearly in the context of a simple example: Suppose that a spatially uniform system in a constant electric field suddenly has the electric field reduced by a factor of 2 at a time $t=0$. The distribution function $g_1(y)$ is $\exp(-y)\Theta(y)$ for $t<0$ and will become $2\exp(-2y)\Theta(y)$ for $t\to\infty$. For positive finite t, the distribution is $2\exp(-2y)\Theta(y)$ for small momenta ($y<t/2\tau$), and drops discontinuously at $y=t/2\tau$. [Equation (37) describes this quantitatively.] The density for small y has doubled. (The distribution function has changed only for those electrons which have equilibrated after the electric field was changed.) A weaker field causes a particle's momentum to increase more slowly, so that the density (per unit momentum) for small y is proportional to the inverse of the instantaneous electric field. This observation explains the small y features of Fig. 1(b), where the electric field now varies as a function of position rather than time.

Just as in the high drag case, it is true for low drag that a sufficiently large density N_0 of electrons will arrange themselves so as to completely screen the potential caused by inhomogeneous doping [see Fig. 1(c)]. The electron density must, however, be higher to accomplish this screening when the drag is low than when it is high. The low drag limit will be discussed in more detail below in the context of $N^+N^-N^+$ structures.

B. $N^+N^-N^+$ structures

In this section, we consider $N^+N^-N^+$ structures, which are prototypical submicron structures consisting of a slab of low dopant density N^- between "contacts" of higher density N^+. Because our solution applies only to systems where the gradient of $V_{\rm ext}$ is small compared to the force α due to the applied field, we can only consider structures where N^+ is not too different from N^-. We call these structures weak $N^+N^-N^+$ junctions. The purpose of this section is to explain where qualitative features appear in the density function $\delta n(x,p)$, and what form they take. One particular issue of interest is whether the prominent ballistic peak that was seen by Baranger and Wilkins[5] (BW) in an $N^+N^-N^+$ structure that was *not* weak is also present in weak structures, or whether on the contrary it is an effect that is fundamentally nonlinear in $V_{\rm ext}$.

The potential V_0 due to inhomogeneous doping in $N^+N^-N^+$ junctions (assuming that the electrons are uniformly distributed) is piecewise parabolic with a continuous slope. The function $\delta n(x,p)$ is obtained by resolving V_D into its Fourier components, and adding together the response to each component [see Eq. (9)]. The Fourier integral is, in practice, approximated numerically as a Fourier sum so the structure is, in effect, repeated periodically in space. To study an isolated slab the repeat distance can be made arbitrarily large. The qualitative remarks made above concerning the high drag limit apply equally to $N^+N^-N^+$ structures. We will instead focus on parameters describing the low drag regime in structures of physical interest.

We choose parameters to represent a device that is similar in some respects to an $N^+N^-N^+$ GaAs structure studied experimentally by Hollis *et al.*[7] and theoretically by Baranger and Wilkins.[5] The width w of the N^- region is chosen to be a fraction of the length L of the entire device, $b\equiv w/L=0.15$. The dimensionless parameter a_1 is a evaluated at the wave number $k=2\pi/L$, $a_1\equiv(mL/4\pi\alpha\tau^2)^{1/2}$, which is determined so that the voltage drop across the N^- region is 0.47 volts, as in Ref. 5. The other parameters required to set a_1 are the width $w=0.4\ \mu m$, the effective mass $m=0.069m_e$, and the relaxation time $\tau=2.9\times10^{-13}$ sec,[5] which result in $a_1=0.917$.

There is some freedom in choosing the remaining parameters δ [Eq. (11b)] and ϵ [Eq. (20), below] to resemble the conditions of Ref. 5. (The parameter ϵ is proportional to the strength of the position-dependent potential.) Together, δ and ϵ determine N^- and N^+. We cannot choose the densities $N_0^-=2\times10^{15}$ cm^{-3} and $N_0^+=10^{18}$ cm^{-3} of Ref. 5, which differ by almost 3 orders of magnitude, and still be in the linear regime. We have instead simulated two sets of densities: (i) $N^-=2.5\times10^{16}$ cm^{-3}; $N^+=(1.48)N^-=3.70\times10^{16}$ cm^{-3} and (ii) $N^-=2\times10^{15}$ cm^{-3}; $N^+=(2.44)N^-=4.88\times10^{15}$ cm^{-3}. The densities were chosen so that N^- is (i) approximately an order of magnitude larger than N_0^- and (ii) equal to N_0^-. Given N^-, N^+ was chosen so that the "small" parameter η is equal to 0.7, where η is the maximum magnitude of $\alpha^{-1}(\partial V_{\rm ext}/\partial x)$. One suspects that for larger η, higher-order effects are becoming important. The choices for N^+ and N^- above, together with $\epsilon_0=12.5$ for GaAs, result in (i) $\delta=21.84$; $\epsilon=0.342$ and (ii) $\delta=2.77$; $\epsilon=0.648$.

Because we are simulating a device with a far smaller $\Delta N\equiv N^+-N^-$, we only expect a qualitative resemblance to the numerical results of BW. The electron density as a function of position is plotted in Fig. 2. For case (i) (solid line, Fig. 2), the electron density is high enough that it can follow the dopant density rather closely, resulting in a nearly complete screening of the dopant potential V_D. The main consequence of incomplete screening is that the electron density has been swept downstream (to the right) of the dopant density by the strong external electric field. These results are in qualitative agreement with those of Baranger and Wilkins. They do not, however, see the overscreening bump in the electron density (arrow in Fig. 2) that is present for weak $N^+N^-N^+$ junctions at low temperatures. If the electron density had been somewhat lower [case (ii)], the nearly complete screening would not have occurred (see Fig. 2, dashed line).

There is an indication from Fig. 2 that the inhomogeneous potential causes a *coherent excitation of plasmons,* or equivalently, a macroscopic excitation of a plasma mode. We consider the overscreening bump (arrow in Fig. 2), and ask why it appears where it does, and in particular why it is shifted so far downstream when the electron density is lower (near $s=0.5$ in Fig. 2, dashed line). The answer is that the bump appears at a position that is

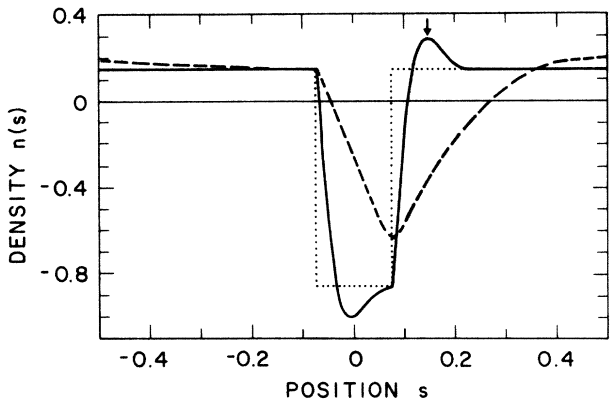

FIG. 2. Dimensionless change in the number density (integrated over momentum) from the average density is plotted as a function of position s for an $N^+N^-N^+$ junction, which repeats periodically in space with period $\Delta s=1$. The solid line is $n(s)$ for case (i). The overscreening bump is indicated by an arrow. The dashed line is $n(s)$ for the lower carrier density case (ii), and the dotted line is the dopant density minus the average dopant density. In case (ii) the electrons are less effective in screening.

reached by electrons that last collided at the downstream N^-N^+ junction, at a time equal to about half of a plasma oscillation cycle. [Both the solid and dashed curve maxima correspond to a time t_m since the last collision such that $\omega_p t_m=0.6(2\pi)$.] The implication is that a plasma oscillation is excited at the junction and manifests itself downstream, half of a plasma oscillation later. We believe that this is a dynamic rather than a static effect. (Note that the usual Debye screening length $[=(\epsilon_0 k_B T/4\pi e^2 N)^{1/2}]$ does not provide a static length scale, since it is zero for $T=0$.) It should be possible to design specialized structures to either enhance or suppress this effect.

We now consider the electron density as a function of momentum, at a given position x. Singularities in the density appear at particular values of the momentum. We believe that at $T=0$ these features that correspond to ballistic peaks are true mathematical singularities (points of nonanalyticity) and not just smooth bumps; this is discussed later in the section. It is of interest to understand first the momentum at which the singularities appear, and second, the form that the singularities take. The first-order change in the electron density is $g_1(y)$. To aid in comparison with previously published data, we also plot the *total density* $n(y)$, given by

$$n(y)=\tilde{n}_0(y)+\epsilon g_1(y)\ . \tag{20}$$

In Sec. III B, $n,\tilde{n}_0$, and g_1 denote real dimensionless densities, which are the dimensionful densities multiplied by $(\alpha\tau/N_0)$ [e.g., $\tilde{n}_0(y)=\exp(-y)\Theta(y)$]. They are functions of the dimensionless momentum y at a given dimensionless position $s\equiv x/L$.

Figure 3(a) is a plot of $g_1(y)$ at a position somewhat into the downstream N^+ contact, and Fig. 3(b) is the corresponding total density.[12] The position is $s=0.10$, where the N^- region extends from $s=-0.075$ to $s=0.075$. The singularity marked S_2 is the ballistic peak that was noted by Baranger and Wilkins. The ballistic peak occurs at precisely the momentum attained by electrons that last suffered a collision at the upstream edge of the N^- region at $s_0=-0.075$. Because at $T=0$ all particles emerge at rest, simple kinematics give a momentum

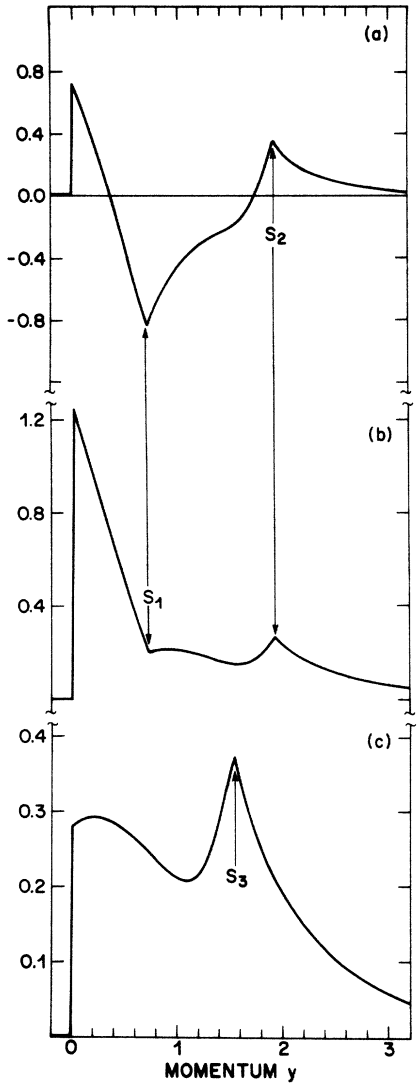

FIG. 3. Plots are for a weak $N^+N^-N^+$ junction with case (i) parameters. (a) The response $g_1(y)$ is plotted at $s=0.1$, corresponding to a position in the N^+ contact, 0.067 μm downstream of the 0.4 μm-wide junction. The momentum $y=1$ corresponds to a velocity of 8.69×10^7 cm/sec. (b) The total electron density, defined in Eq. (20), is plotted at $s=0.1$ as in (a). Singular points are labeled S_1 and S_2. (c) The total electron density is plotted at $s=0.0375$, corresponding to a position within the N^- region, 0.3 μm from the upstream contact. The position is identical to position x_3 of Ref. (5). The ballistic peak singularity is labeled S_3. For weak potentials ($\epsilon\to0$), the average electron kinetic energy would correspond to a "temperature" of 3400 K, via $\frac{1}{2}k_B T=\frac{1}{2}m\langle(v-\langle v\rangle)^2\rangle$.

$$y_0 = a_1[8\pi(s - s_0)]^{1/2} \tag{21}$$

for particles at position s that last collided at position s_0. Note also the singularity marked S_1, which could be called a "ballistic antipeak." It occurs at the momentum y given by Eq. (21) for those electrons that last collided at the *downstream* edge of the N^- region ($s_0 = +0.075$). (The singularity at $s = 0$ is trivial, and simply reflects the fact that there are no left-moving electrons at $T = 0$.)

We now consider the shape of the singularities. Because we are working to first order in V_D, the momentum distribution can be given in terms of a real space Green's function G,

$$g_1(y,s) = \int_{-\infty}^{\infty} ds_0\, G(s - s_0, y) V_D(s_0) \,. \tag{22}$$

An analytic study of the singular structure of G indicates[13] that its most singular part is proportional to the derivative of a δ function, $-\delta'(y - y_0)$, where y_0 is given by Eq. (21). It also contains a term proportional to $\delta(y - y_0)$, and a diffuse term that contributes at all momenta, not just at y_0. These properties are consistent with the Green's function that is obtained numerically for a potential V_D that is a narrow Gaussian (a "fat" δ function); see Fig. 4. By integrating Eq. (22) by parts, one sees that the leading singular behavior of $g_1(y,s)$ is proportional to *the derivative* of V_D,

$$g_1(y,s) \sim (dV_D/ds)\,|_{s_0 = s - y^2/8\pi(a_1)^2} \,. \tag{23}$$

Thus if V_D is n times continuously differentiable but the $(n+1)$st derivative is discontinuous at s_0, then the leading singularity of $g_1(y)$ is a discontinuous nth derivative at $y = y_0$.

We believe that these singularities in $g_1(y,s)$ are present not only to first order in λ, but also in the exact solution at nonzero λ. The strength of the singularity (for example, the magnitude of the discontinuity in the first derivative in the case considered below) is a function of λ whose leading term is linear in λ. As long as λ is smaller than the radius of convergence of the perturbation series, which we assume is nonzero, this function cannot vanish identically except at isolated points.

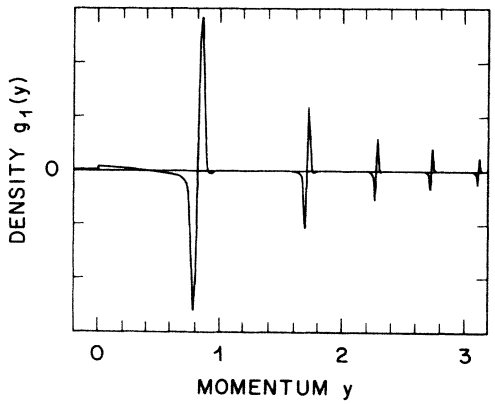

FIG. 4. Response $g_1(y)$ is the numerically determined Green's function. It is plotted at $s = 0.3$ for parameters $a_1 = 0.3$ and $\delta = 21.84$. The potential consists of narrow Gaussians of standard deviation $\sigma = 0.02$ at positions $s = \ldots, -2, -1, 0, 1, \ldots$. The most noticable part of the response to a δ function ($\sigma \to 0$) in the potential is the *derivative* of a δ function in $g_1(y)$.

At nonzero temperature, electrons emerge with a distribution of velocities so that there is no unique relation between the position s_0 of a singularity in V_D and a particular momentum at position s. What was at $T = 0$ a true singularity in $g_1(y)$ becomes a smeared feature that may still, however, be rather sharp at low temperatures.

One can now explain the shape of the singularities in Figs. 3(a) and 3(b). An ideal $N^+N^-N^+$ junction has a potential with a discontinuous second derivative, which is a consequence of the discontinuous dopant density. Equation (23) implies that the singularity marked S_2 is a discontinuous first derivative forming an upward pointing cusp. (Note that s_0 decreases as y increases.) Equation (23) also implies that the singularity marked S_1 is a downward pointing cusp.

The momentum distribution $g_1(y)$ is plotted at $s = 0.0375$ in Fig. 3(c). Since this cut is taken above the downstream contact, only one singularity exists, the upward pointing cusp marked S_3. This cut is taken at the same position in the device as the distribution marked x_3 in BW. There is a qualitative resemblance between these two ballistic peaks, in spite of the fact that Fig. 3(c) of the present work arises in a weak $N^+N^-N^+$ junction and that of BW does not. There is, however, a notable difference in interpretation. Baranger and Wilkins have interpreted their ballistic peak as arising from electrons that are "skimmed", emerging with near-zero velocity from the vicinity of the local maximum in their total potential at a position $x_1 = 0.05$ μm downstream from the upper contact. In the present work, the total potential *does not possess* a *local maximum;* it decreases monotonically for small V_D. At zero temperature the ballistic peak is unambiguously associated with the point $x = 0$ at the upstream contact. The discontinuity in d^2V_D/ds^2 at this point causes a discontinuity in the slope of $g_1(y)$. The fact that the electric field is a minimum, results in a peak in $g_1(y)$ for the reasons given in the discussion of Fig. 1(b). [The locally weak electric field contributes to the formation of a ballistic peak, but other factors also shape g_1 in Fig. 3(c), including a partial screening of the dopant potential by a spatially nonuniform electron density.] The novel formation of a ballistic peak by a monotonically decreasing potential is expected to persist for junctions that are *not* weak. Note that in steady state at zero temperatures, junctions must have monotonically decreasing total potentials. (Otherwise a local minimum would trap electrons that could never escape, until their space charge eliminated the minimum.)

Figure 3(c) of this work also bears a qualitative resemblance to the Monte Carlo simulation of Hesto *et al.*, Fig. 4(a),[4] which was done at a lower temperature (77 K) than that of BW. Our simulations at an even lower density (not shown) more closely resemble the conditions and the results of Hesto *et al.*

IV. TIME-DEPENDENT POTENTIAL

The calculation of Sec. II for an arbitrary position-dependent potential can be generalized to obtain the response for arbitrary position *and time* dependence. The calculation closely follows that of Sec. II. The potential is now given by

$$V_D(x,t)=\sum_{k,\omega} V_{k,\omega}e^{i(kx-\omega t)} . \tag{24}$$

The uniform electric field provides a time-independent force α, which still dominates $\partial V_{\text{ext}}/\partial x$. We consider only one Fourier component here; a solution for a general weak potential (varying in time and space) can be obtained by summing over Fourier components as shown earlier. Following our previous formalism, we look for a solution of the Boltzmann equation of the form

$$n(x,p,t)=n_0(p)+\frac{\lambda N_0}{\alpha\tau}g_1(p)e^{i(kx-\omega t)} , \tag{25}$$

where $n_0(p)$ is again given by Eq. (7). The solution for g_1 is given by the one-dimensional integral,

$$g_1(k,y)=\exp\left[-y(1-i\omega\tau)-\frac{iy^2}{4a^2}\right]\int_{-\infty}^{y}dx\left[\left[\frac{B}{\pi}\right]^{1/2}[1+\gamma_k(1+ia^2\delta)]\exp(-Bx^2)\right.$$
$$\left.-\tfrac{1}{2}(1+ia^2\delta\gamma_k)\exp\left[-x+\frac{1}{4B}\right]\left\{\Phi\left[\sqrt{B}\left[x-\frac{1}{2B}\right]\right]+1\right\}\right]$$
$$\times\exp\left[x(1-i\omega\tau)+\frac{ix^2}{4a^2}\right] . \tag{26}$$

In the zero-temperature limit, g_1 is

$$g_1(y)=\exp\left[-y(1-i\Omega)-\frac{iy^2}{4a^2}\right]\Theta(y)\left\{1+\gamma(1+ia^2\delta)-a\sqrt{2\pi}(1+ia^2\delta\gamma)\right.$$
$$\times\exp(-i\Omega^2a^2)\left[C(\Omega a)+iS(\Omega a)+C\left[\frac{y}{2a}-\Omega a\right]\right.$$
$$\left.\left.+iS\left[\frac{y}{2a}-\Omega a\right]\right]\right\} \tag{27}$$

where y, a^2, and δ have been defined earlier [see Eqs. (8) and (11b)] and $\Omega=\omega\tau$.

The self-consistency condition for the complex constant γ is

$$I(\gamma)\equiv\int_{-\infty}^{\infty}g_1(y)dy=\gamma . \tag{28}$$

In this case we have not been able to solve for γ in terms of common special functions as in Eq. (16). However, by noting that $g_1(y)$, and therefore $I(\gamma)$ are linear in γ, the proper value of γ can be found by numerically evaluating $I(\gamma)$ at two values γ_1 and γ_2

$$\gamma=\frac{I(\gamma_2)\gamma_1-I(\gamma_1)\gamma_2}{I(\gamma_2)+\gamma_1-[I(\gamma_1)+\gamma_2]} . \tag{29}$$

The frequency appears in Eq. (27) primarily through the dimensionless combination $(\alpha\Omega)^2$. This is proportional to the ratio of the wavelength of the potential to the distance traveled in one period $T=2\pi/\omega$ under the uniform electric field. In the low-frequency limit ($\Omega\ll1$ and $\Omega a\ll1$), $g_1(y)$ becomes indistinguishable from the static solution of Sec. II. A detailed investigation of higher frequencies has not yet been performed. In some regimes, we expect finite frequencies to cause significant changes in qualitative behavior. For example, $g_1(y)$ should be strongly modified for particles that are locally traveling with a velocity equal to the phase velocity ω/k of the potential. These particles should be accelerated or decelerated depending on their positions.

V. SOLUTION FOR AN INITIALLY NONEQUILIBRIUM DISTRIBUTION

Here we describe the time-dependent linear response of the distribution function $n(x,p,t)$ under more general conditions than those considered in Sec. IV. The distribution function at time t is, in general, a function of both the potential $V_D(x,t)$ and the initial condition $n(x,p,t_0)$. If the time at which an observation is made is much later than t_0, one expects that $n(x,p,t)$ becomes essentially independent of the initial condition, and is a function of $V_D(x,t)$ alone. (Initial conditions that violate overall charge neutrality are not allowed.) This case was solved in Sec. IV. The solutions of this section allow one to follow the time evolution of an initial nonequilibrium distribution (which may be due, for example, to photoexcited carriers) at time that need not be much later than t_0.

For simplicity we work at zero temperature (generalization to finite T is straightforward). The equation to be solved is

$$\frac{\partial n}{\partial t}=\frac{\partial n}{\partial p}\frac{\partial V_{\text{tot}}}{\partial x}-\frac{\partial n}{\partial x}\frac{p}{m}$$
$$-\frac{1}{\tau}\left[n-\delta(p)\int_{-\infty}^{\infty}n(x,p,t)dp\right] , \tag{30}$$

with an initial nonequilibrium distribution $n(x,p,t=0)$ specified. The potential $\partial V_{\rm tot}/\partial x$ is of the form

$$\frac{\partial V_{\rm tot}}{\partial x}=-\alpha+\frac{\partial V_D}{\partial x}+\frac{\partial \Delta V}{\partial x}\ , \tag{31}$$

where, as before, α is the force due to an applied dc electric field which dominates $\partial V_{\rm tot}/\partial x$ ($\alpha \gg \partial V_D/\partial x + \partial \Delta V/\partial x$). $V_D(x,t)$ is an arbitrary potential due to doping, ac electric fields, etc., whose spatial and temporal variation is given. $\Delta V(x,t)$ is an additional potential that arises through the dynamic electron rearrangement. The solution $n(x,p,t)$ describes the behavior of an initial nonequilibrium distribution $n(x,p,0)$ in the presence of a strong electric field as well as weak inhomogeneous and time varying potentials. The deviation of the initial distribution from steady state is also assumed small.

To solve Eq. (30) we write n in terms of Fourier components:

$$n(x,p,t)=f(0,p,t)+\sum_{k\neq 0} f(k,p,t)e^{ikx}\ . \tag{32}$$

As before, the potential $\partial V_{\rm tot}/\partial x$ is also written in terms of Fourier components:

$$\frac{\partial V_{\rm tot}}{\partial x}=\alpha\left[-1+\sum_{k\neq 0} ikV_k(t)e^{ikx}/\alpha+\sum_{k\neq 0} 4\pi i e^2 e^{ikx}\Gamma(k,t)/k\alpha\epsilon_0\right]\ , \tag{33}$$

where the last sum ΔV is found through Poisson's equation and Γ is defined by

$$\Gamma(k,t)\equiv\int_{-\infty}^{\infty} f(k,p,t)dp\ . \tag{34}$$

We assume $f(k\neq 0,p,t)\ll f(0,p,t)$ and $\lambda_k=ikV_k/\alpha\ll 1$, and solve for the linear response. To first order, the $k\neq 0$ Fourier components do not mix and the equations for each mode separate. The $k=0$ mode satisfies a particularly simple equation,

$$\left[\frac{\partial}{\partial t}+\alpha\frac{\partial}{\partial p}+\frac{1}{\tau}\right]f(0,p,t)=\frac{N_0}{\tau}\delta(p)\ , \tag{35}$$

where we have used $\Gamma(0,t)=N_0$. The $k\neq 0$ mode is driven by $f(0,p,t)$ above in the following manner:

$$\left[\frac{\partial}{\partial t}+\alpha\frac{\partial}{\partial p}+\frac{ik}{m}p+\frac{1}{\tau}\right]f(k,p,t)=\frac{1}{\tau}\delta(p)\Gamma(k,t)+\left[\frac{4\pi i e^2}{k\epsilon_0}\Gamma(k,t)+\alpha\lambda_k(t)\right]\frac{\partial f(0,p,t)}{\partial p}\ . \tag{36}$$

Note that the solution $f(k,p,t)$ must satisfy Eq. (34).

The solution for $f(0,p,t)$ is

$$f(0,p,t)=f_0(p-\alpha t)e^{-t/\tau}+\frac{N_0}{\alpha\tau}e^{-p/\alpha\tau}\Theta(p)\Theta(\alpha t-p)\ , \tag{37}$$

where $f_0(p)\equiv f(0,p,t=0)$ is the initial electron distribution in the $k=0$ mode. From this expression for $f(0,p,t)$, the Green's function solution for $f(k,p,t)$ can be found. For the Green's function, the initial condition is $f(k,p,t=0)=\delta(p-q)$. (The solution for an arbitrary initial distribution is obtained by superposition.) The solution to Eq. (36) is

$$\begin{aligned}f(k,p,t)=&\delta(p-q-\alpha t)\exp\left[-\frac{t}{\tau}-\frac{ik}{2m}(2qt+\alpha t^2)\right]\\&+\int_0^t ds\left\{\exp\left[-\frac{(t-s)}{\tau}-\frac{ik}{2m}[2p(t-s)-\alpha(t-s)^2]\right]\right.\\&\left.\times\left[\frac{1}{\tau}\delta(p-\alpha(t-s))\Gamma(k,s)+\left[\frac{4\pi i e^2}{k\epsilon_0}\Gamma(k,s)+\alpha\lambda_k(s)\right]\frac{\partial f}{\partial q}(0,p-\alpha(t-s),s)\right]\right\}\ ,\end{aligned} \tag{38}$$

where $\partial f(0,q,s)/\partial q$ is obtained from Eq. (37). To make use of Eq. (38), one must first determine $\Gamma(k,t)$. Integrating Eq. (38) with respect to p, one obtains the following equation for Γ:

$$\begin{aligned}\Gamma(k,t)=&\exp\left[-\frac{t}{\tau}-\frac{ik}{2m}(2qt+\alpha t^2)\right]\\&+\int_0^t ds\left\{\exp\left[-\frac{(t-s)}{\tau}-\frac{ik\alpha}{2m}(t-s)^2\right]\right.\\&\left.\times\left[\frac{\Gamma(k,s)}{\tau}+\frac{ik(t-s)}{m}\left[\frac{4\pi i e^2}{k\epsilon_0}\Gamma(k,s)+\alpha\lambda_k(s)\right]\int_{-\infty}^{\infty}dp\exp\left[-\frac{ik(t-s)p}{m}\right]f(0,p,s)\right]\right\}\ .\end{aligned} \tag{39}$$

This is an integral equation for $\Gamma(k,t)$; however, a direct numerical integration from the initial condition, $\Gamma(k,t=0)$, is possible since one only needs $f(0,p,t)$, which is known from Eq. (37), and $\Gamma(k,s)$ for $s<t$. Knowing $\Gamma(k,t)$, the Green's function $f(k,p,t)$ can be found from Eq. (38).

VI. CONCLUSION AND SUMMARY

The study of transport in submicron structures requires the understanding of inhomogeneous, strongly driven systems far from equilibrium. Our approach has been to incorporate the strongly driven aspect exactly, and to treat the inhomogeneous structure as a perturbation. We have obtained exact analytic solutions to first order in an arbitrary position and time-dependent perturbation, and in an initially nonequilibrium distribution.

Studying the response to weak potentials is a logical first step in acquiring insight into the behavior of submicron structures in general. Once the Green's function is known, either in k space or in real space (Figs. 1 and 4, respectively), any structure can be understood by superposition, rather than by solving the Boltzmann equation anew for each structure. Several new features have emerged from linear response. (1) A ballistic peak is formed in an $N^+N^-N^+$ structure, even when the potential decreases monotonically, so that there is no local maximum to "skim" the electrons. Antipeaks are also present. (2) At zero temperature, the ballistic peak is a true singularity (a discontinuous derivative). (3) Inhomogeneous structures appear to be capable of coherently exciting plasmons. (4) It is now possible to study time-dependent transport in inhomogeneous systems. This new capability will at last permit the study of the time response of submicron devices.

Many topics remain for future study. The finite temperature and finite frequency solutions have not yet been examined in detail. It would be of particular interest to apply the time-dependent solutions to the design of fast photodetectors. Furthermore, the linear response could provide a starting point for algorithms seeking full nonlinear solutions, since these algorithms often fail to converge due to inaccurate starting points. Finally, our methods may be generalizable to study three-dimensional structures, or to calculate higher order in perturbation theory.

ACKNOWLEDGMENTS

We thank H. Baranger and J. Wilkins for valuable discussions and for a careful reading of the manuscript. This work was supported in part by National Science Foundation (NSF) under Grant No. DMR-80-20263.

---

[1]See for example, *Physics of Submicron Devices, Proceedings of the Workshop on the Physics of Submicron Structures, Urbana, Illinois, 1982,* edited by H. L. Grubin, K. Hess, G. J. Iafrate, and D. K. Ferry (Plenum, New York, 1984).

[2]S. M. Sze, *Physics of Semiconductor Devices,* 2nd edition (Wiley, New York, 1981).

[3]M. S. Shur and L. F. Eastman, Solid State Electron. **24**, 11 (1981).

[4]P. Hesto, J. F. Pone, R. Castagne, and J. L. Pelouard in *Physics of Submicron Devices,* Ref. 1, p. 101; Y. Awano *et al.*, Electron. Lett. **18**, 133 (1982); A. Ghis, E. Constant, and B. Boittiaux, J. Appl. Phys. **54**, 214 (1983).

[5]H. U. Baranger and J. W. Wilkins, Phys. Rev. B **30**, 7349 (1984).

[6]N. W. Ashcroft and D. Mermin, *Solid State Physics* (Holt, Rinehardt, and Wilson, New York, 1976).

[7](a) M. Hollis, L. F. Eastman, and C. E. C. Wood, Electron. Lett. **18**, 571 (1982); (b) F. S. Khan, Bull. Am. Phys. Soc. **30**, 520 (1985); S. K. Sarker, J. H. Davies, F. S. Khan, and J. W. Wilkins (unpublished); (c) An analytic solution for a different Boltzmann equation, where energy is absorbed only in discrete units by optical phonons, has recently been obtained by G. D. Mahan (unpublished).

[8]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).

[9]This is a self-consistency requirement that $\gamma=\int g_1(y;\gamma)dy$. In the simplest circumstances, γ can be calculated in closed form [see Eq. (16)]. However, even in the more difficult situations where one-dimensional integrals must be done numerically, one need not conduct an open ended search for the correct γ. Since the self-consistency equation is linear in γ, any two trials at different values of γ allow one to solve immediately for the self-consistent value [see Eq. (29)].

[10]P. Hesto, Surf. Sci. **132**, 623 (1983).

[11]J. R. Hayes, A. F. J. Levi, and W. Wiegmann, Phys. Rev. Lett. **54**, 1570 (1985).

[12]Note that although Fig. 3(a) is a graph of the exact first-order response, Fig. 3(b) is only approximate. It is approximate in the sense that $f(x_2)\approx f(x_1)+(df/dx)|_{x_1}(x_2-x_1)$ when (x_2-x_1) is small, where (df/dx) is the exact derivative.

[13]S. A. Trugman and A. J. Taylor (unpublished).