SEMICONDUCTOR BOLTZMANN TRANSPORT EQUATION IN MACROSCOPIC AND QUANTUM-CONFINED SYSTEMS

In this article we describe the carrier dynamics within a crystal based on the behavior of the distribution function f. The main tool used for determining the characteristics of the distribution function f is the Boltzmann transport equation. A model for the conductivity (and the mobility) of carriers within a crystal is discussed in Transport in Semiconductors, Dynamics of Carriers in Macroscopic and Mesoscopic Systems. This model is based on the transport behavior of one electron having an "average" velocity. We know that in an ensemble of electrons the energy, momentum, velocity, and the spatial position of individual electrons vary quite substantially. To describe this ensemble we define a distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$ in six-dimensional coordinates, called "phase space."

In the first section we define the probability distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$. Then we show that the *equilibrium* value of this distribution function $f_0(\boldsymbol{p}, \boldsymbol{r}, t)$ is the well-known *Fermi–Dirac distribution* function. In the next section we calculate the average concentration and energy of carriers under equilibrium. In the following we derive the Boltzmann transport equation (*BTE*) by equating the change in the distribution function caused by accelerating forces to the change in the distribution function resulting from scattering forces. An approximate solution to the BTE is presented in the subsequent section. This method, which is called relaxation time approximation (*RTA*) provides the basis for describing the conductivity and mobility of crystals. In the next section the conductivity of semiconductors with elliptical bands (such as silicon), multiscattering mobility, Hall mobility, and temperature-dependence of mobility of carriers in semiconductors are described.

A more general solution to Boltzmann transport equation, known as "method of moments" is presented in the following section. In particular, three widely used balance equations are derived from BTE which collectively describe the conservation of carriers, their momentum, and their energy. These balance equations, which form an infinite chain of balance equations, are truncated subject to several simplifying approximations, which are also discussed in this section.

Carrier dynamics within the crystals are analyzed in many other ways. Monte Carlo technique, (*MC*) and the hydrodynamic model (*HD*) are two of the most widely used techniques in recent literature. In the following section we describe the Monte Carlo simulation technique, and in the last section we present a hydrodynamic model based on the solution of the two higher moments of the Boltzmann transport equation. In particular, we show how the Boltzmann equation is solved self-consistently with the Poisson and Schrödinger equations for a high-electron-mobility transistor.

THE DISTRIBUTION FUNCTION *f(r, p, t)*

To describe the ensemble of carriers in a crystal, we must define a distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$, which is a probability distribution function. This function is defined in the phase space which is a space consisting of position and momentum coordinates. The distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$ has various definitions, such as those that follows.

1. The distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$ is the probability of finding an electron in a box volume of size $\Delta\boldsymbol{r}$, centered at $\boldsymbol{r}$, and $\Delta\boldsymbol{p}$, centered at $\boldsymbol{p}$, at time t where $\boldsymbol{p}$ is the momentum and $\boldsymbol{r}$ is the position. The quantities $\boldsymbol{p}$ and $\boldsymbol{r}$ do not refer to a particular carrier but are merely the momentum and position in the phase space. With this definition, the integral of $f(\boldsymbol{p}, \boldsymbol{r}, t)$ over the entire phase space must add up to unity:
$$\iint f(\boldsymbol{r}, \boldsymbol{p}, t) d\boldsymbol{p}\, d\boldsymbol{r} = 1 \tag{1}$$
2. The distribution function $f(\boldsymbol{p}, \boldsymbol{r}, t)$ is the average number of electrons in a phase-space box of size $\Delta\boldsymbol{p}\,\Delta\boldsymbol{r}$ located at the phase space point $(\boldsymbol{r}, \boldsymbol{p})$. With this definition, the integral of $f(\boldsymbol{p}, \boldsymbol{r}, t))$ must add up to
$$\iint f(\boldsymbol{r}, \boldsymbol{p}, t) d\boldsymbol{p}\, d\boldsymbol{r} = N(t) \tag{2}$$
3. where $N(t)$ is total number of electrons at time t.
4. The distribution $f(\boldsymbol{v}, \boldsymbol{r}, t)$ defines the probability of finding one electron in a velocity range $\boldsymbol{v}$ and $\boldsymbol{v} + d\boldsymbol{v}$ having space coordinate between $\boldsymbol{r}$ and $\boldsymbol{r} + d\boldsymbol{r}$. As in the second definition,
$$\iint f(\boldsymbol{r}, \boldsymbol{v}, t) d\boldsymbol{v}\, d\boldsymbol{r} = N(t) \tag{3}$$

Note that in all three definitions, the *Heisenberg uncertainty principle* is violated, for the very existence of such distribution function $f(\boldsymbol{r}, \boldsymbol{p}, t)$ inherently implies the possibility of simultaneous measurement of the two quantities, $\boldsymbol{r}$ and $\boldsymbol{p}$, which according to Heisenberg uncertainty principle is forbidden.

Therefore, the Boltzmann transport equation (BTE), which is based on a time derivative of the distribution function $f(\boldsymbol{r}, \boldsymbol{p}, t)$, is a *nonquantum classical* approach to the transport in semiconductors. In fact BTE was developed far earlier (BTE is now 170 years old) than quantum mechanics. In spite of this violation of the uncertainty principle, the BTE is one of the most powerful tools in studying the transport phenomenon in macroscopic semiconductors (semiconductor devices with dimensions of the order of micrometer). However, recent advances in semiconductor technologies have pushed the dimension of electron devices down to 0.13 micron (130 nanometers), and 0.08 micron (80 nanometers). At the research and development level, devices with dimensions of 40 nanometers are being developed. It is anticipated that "micro-electronics" will soon give way to "nano-electronics", in which the effective dimensions of electron devices will be pushed down to less than 10 nanometers. For a brief introduction to these new development, see Section 10 of the article: Transport in Semiconductors, Dynamics of Carriers in Macroscopic and Mesoscopic Systems. There are other limitations to the validity of BTE which are discussed later in this article.

The function $f(\boldsymbol{r}, \boldsymbol{p}, t)$ is a distribution function that generates numbers between 0 and 1 which are the probability of finding a carrier at location $\boldsymbol{r}$ with crystal momentum $\boldsymbol{p}$ at time t. The value of $f(\boldsymbol{r}, \boldsymbol{p}, t)$ is obtained by solving the Boltzmann transport equation for one (or more) particular scattering source(s). The *equilibrium distribution function*, however, is simply the *Fermi–Dirac distribution* given as (1)

$$f_0(\boldsymbol{r}, \boldsymbol{p}) = \frac{1}{1 + e^{\frac{[E_C(\boldsymbol{r}, \boldsymbol{p}) - E_F]}{k_B T_L}}} \tag{4}$$

where k_B is the Boltzmann constant, E_F is the *Fermi energy*, T_L is the lattice temperature (which is equal to electron temperature only under low applied fields), and $f_0(\boldsymbol{r}, \boldsymbol{p}, t)$ is the equilibrium distribution function.

The Fermi energy E_F is an electrochemical energy and has several implications:

1. The amount of energy that is added to the crystal at 0 K when one electron is added to the crystal.
2. The energy level at which the probability of occupancy is exactly one-half.
3. The highest occupied energy level in metals.
4. The energy level which remains constant throughout the crystal when there is no applied electric field (and no current transport).

The term $E_C(\boldsymbol{r}, \boldsymbol{p})$ is the total energy of carriers consisting of carrier potential energy, $E_{C0}(\boldsymbol{r}, t)$ and kinetic energy $E(\boldsymbol{p})$:

$$E_C(\boldsymbol{r}, t) = E_{C0}(\boldsymbol{r}, t) + E(\boldsymbol{p}) \tag{5}$$

For a parabolic spherical band structure $E(\boldsymbol{p}) = p^2/2m$. Assuming a *nondegenerate* semiconductor (low to moderately doped under low injection levels), the distribution function is given by

$$f_0(\boldsymbol{r}, \boldsymbol{p}) = e^{\frac{-[E_{C0}(\boldsymbol{r}) - E_F]}{k_B T_L}} e^{\frac{p^2}{2m^* k_B T_L}} \tag{6}$$

where $E_{C0}(\boldsymbol{r})$ is the energy level corresponding to the bottom of the conduction band. In a uniformly doped isotropic semiconductor with no composition grading, E_{C0} is independent of position. The equilibrium distribution function $f_0(\boldsymbol{r}, \boldsymbol{p}, t)$ can be graphed as a function of one of the momentum coordinates, as depicted in Fig. 1, where $f_0(p_z)$ is graphed as a function of the z-coordinate with momentum p_z. Note that the distribution function is symmetrical in p-space, implying that the probabilities of carriers having momentum p_z and $-p_z$ are the same. Since there are equal number of carriers with oppositely directed momenta, the net current is zero.

When the distribution function is disturbed from its equilibrium value by an applied field (the solid line in Fig. 1), the Fermi level, E_F is shifted to a higher energy level. This energy level is called a *quasi-Fermi level* which is also called an *imref level* (Fermi spelled backward). The momentum distribution of carriers is altered, as shown by the dashed line in Fig. 1. This distribution function is called the *Maxwell–Boltzmann distribution* which is indeed an approximation to the Fermi–Dirac distribution in nondegenerate semiconductors. Also note that in both distributions, the greatest number of carriers have zero or a very small momentum, which is indeed the equilibrium condition.

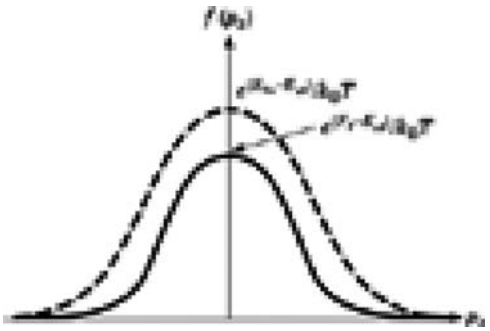

Figure 1. Illustration of Maxwellian distribution functions. The solid line is the equilibrium distribution function, and the dashed line is the nonequilibrium distribution with the Fermi energy shifted to higher energies. After Ref. 1, reprinted with permission.

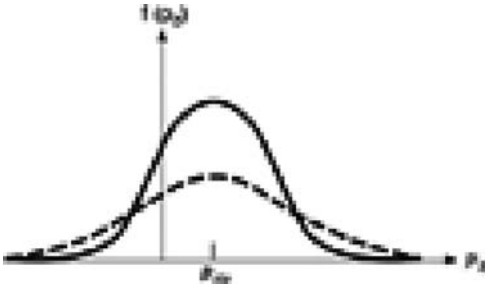

Figure 2. Illustration of displaced Maxwellian distribution functions. The solid line assumes that the carrier temperature and lattice temperature are equal, and the dashed line shows the hot carrier phenomenon where the carrier temperature exceeds that of the lattice. After Ref. 1, reprinted with permission.

When an applied field causes the kinetic energy of carriers (and therefore their momentum) to increase, the distribution function $f(\boldsymbol{p})$ is shifted toward higher momenta, although it preserves its Maxwellian shape. This is shown by the solid line in Fig. 2. Sometimes, the applied field is so intense that the carriers get more kinetic energy than that which keeps them in equilibrium with the crystal lattice. The distribution is shifted toward positive momenta p_z, and there is a net transport of carriers in the positive direction. Under these conditions, the carriers have more kinetic energy than they can lose to the scattering agents, and because of that they attain a temperature T_C, higher than the lattice temperature T_L. Under such a high electric field, the carriers are said to be "*hot carriers*." Their distribution is shown by a dashed line in Fig. 2. These two nonequilibrium distributions are *displaced Maxwellian* distributions. It should be emphasized here that the use of the Fermi–Dirac distribution for $f(\boldsymbol{r}, \boldsymbol{p}, t)$ is limited only to homogeneous semiconductors (2).

AVERAGE EQUILIBRIUM CARRIER CONCENTRATION AND ENERGY

When the equilibrium distribution function $f(\boldsymbol{r}, \boldsymbol{p}, t)$ is known, the average carrier concentration is obtained by

summing up the carriers in each momentum state and dividing the sum by the volume Ω:

$$n(\boldsymbol{r}, t) = \frac{1}{\Omega} \sum_{\boldsymbol{p}} f_0(\boldsymbol{r}, \boldsymbol{p}, t) \tag{7}$$

where Σ_p is over all states in the first Brillouin zone. Converting the summation over $\boldsymbol{p}$ to integration as

$$\sum_{\boldsymbol{k}} = \frac{\Omega}{4\pi^3} \int_{\boldsymbol{k}} g(\boldsymbol{k}) d\boldsymbol{k} \tag{8}$$

and transforming from wave vector k to momentum space results in

$$n_0(\boldsymbol{r}) = \frac{1}{4\pi^3\hbar^3} \int_p e^{\frac{-[E_{C0}(\boldsymbol{r}) - E_F + \frac{p^2}{2m^*}]}{k_B T_L}} dp \tag{9}$$

which, after performing the integration, yields

$$n_0(\boldsymbol{r}) = \frac{1}{4} \left[\frac{2m^* k_B T_L}{\pi\hbar^2} \right]^{3/2} e^{\frac{-[E_{C0}(\boldsymbol{r}) - E_F]}{k_B T_L}} \tag{10}$$

The coefficient in front of the exponential is the *effective density of states* in the conduction band and is denoted as N_C. Then the equilibrium carrier density is given by

$$n_0(\boldsymbol{r}) = N_C e^{\frac{-[E_{C0}(\boldsymbol{r}) - E_F]}{k_B T_L}} \tag{11}$$

which is valid for a nondegenerate semiconductor in equilibrium. It can be shown (2) that the equilibrium carriers density for degenerate semiconductors (when the concentration of carriers is high compared to the effective density of the states) is given by

$$n_0(\boldsymbol{r}) = N_C F_{\frac{1}{2}} \left(\frac{E_{C0}(\boldsymbol{r}) - E_F}{k_B T_L} \right) \tag{12}$$

where

$$F_n(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{y^n dy}{1 + e^{(y-x)}} \tag{13}$$

is the *Fermi–Dirac integral*, with $n = 1/2$ for $F_{1/2}(x)$. The equilibrium average energy of carriers can be calculated similarly as (1)

$$W_0(\boldsymbol{r}, t) = \frac{1}{\Omega} \sum_{\boldsymbol{p}} E(\boldsymbol{p})\, f_0(\boldsymbol{p}, \boldsymbol{r}, t) \tag{14}$$

which for a parabolic spherical energy band is expressed by

$$W_0(\boldsymbol{r}) = \frac{1}{8\pi^3\hbar^3 m^*} \int_p p^2 e^{\frac{-[E_{C0}(\boldsymbol{r}) - E_F + \frac{p^2}{2m^*}]}{k_B T_L}} dp \tag{15}$$

Performing the integration yields

$$W_0(\boldsymbol{r}) = \frac{3}{2} n_0(\boldsymbol{r}) k_B T_L \tag{16}$$

which gives the well-known average thermal energy per carrier of

$$\begin{aligned} u_0(\boldsymbol{r}) &= \frac{W_0(\boldsymbol{r})}{n_0(\boldsymbol{r})} \\ &= \frac{3}{2} k_B T_L \end{aligned} \tag{17}$$

The average kinetic energy component associated with each direction x, y, or z is similarly calculated as

$$\begin{aligned} u_{x0} &= u_{y0} = u_{z0} \\ &= \frac{1}{2} k_B T_L \end{aligned} \tag{18}$$

Equation (18) establishes the fact that the kinetic energy of carriers is equally distributed among the three directions, a result that is called the *equipartition of energy*.

BOLTZMANN TRANSPORT EQUATION

Now we have established the probability function f. The Boltzmann transport equation simply relates the time rate of change of f caused by forces to the time rate of change f caused by scattering sources:

$$\left. \frac{df(\boldsymbol{r}, \boldsymbol{p}, t)}{dt} \right|_{\text{forces}} = \left. \frac{df}{dt} \right|_{\text{scattering}} \tag{19}$$

Using the chain rule of differentiation and using wave vector $\boldsymbol{k}$ instead of $\boldsymbol{p}$ in Eq. (19), the $df(\boldsymbol{r}, \boldsymbol{p}, \text{t})/dt\,|_{\text{forces}}$ term can be expanded as

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \boldsymbol{r}} \frac{d\boldsymbol{r}}{dt} + \frac{\partial f}{\partial \boldsymbol{k}} \frac{d\boldsymbol{k}}{dt} = \left. \frac{df}{dt} \right|_{\text{scattering}} \tag{20}$$

which is usually written as

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{r}} f + \frac{d\boldsymbol{k}}{dt} \cdot \nabla_{\boldsymbol{k}} f = \left. \frac{df}{dt} \right|_{\text{scattering}} \tag{21}$$

where we have used the fact that $d\boldsymbol{r}/dt$ is the velocity $\boldsymbol{v}$. Equation (21) is the Boltzmann transport equation.

The validity of the BTE is restricted to the following conditions (2):

1. The distribution function f has small variation with position. This allows us to use the band structure model we developed in the third section of the article Transport in Semiconductors, Dynamics of Carriers.
2. The forces acting on the electron ensemble are small, so that the response of the system can be considered semiclassically within the effective mass approximation (see the fourth section of the previous article).
3. The time variation of f is assumed to be slow within the time frame of consecutive collisions. In other words, the variations in f are assumed to be small within the timescale of hydrodynamic time τ_H. For discussions of hydrodynamic time and other timescales in the transport of carriers within crystals, refer to the second section of the previous article.

The left-hand side of the BTE has three terms: $\partial f/\partial t$ which is set to zero for steady-state transport, $\boldsymbol{v} \cdot \nabla_r f$ the velocity term, and $(d\boldsymbol{k}/dt) \cdot \nabla_k$ the acceleration term which is usually related to the electric field (or magnetic field if applied).

All of the scattering processes are lumped together in the scattering term $\partial f/\partial t\,|_{\text{scattering}}$.

Two processes contribute to a change in distribution: (1) carriers from states with wave vector k' could be scattered into state k (in scattering); and (2) carriers at state k could

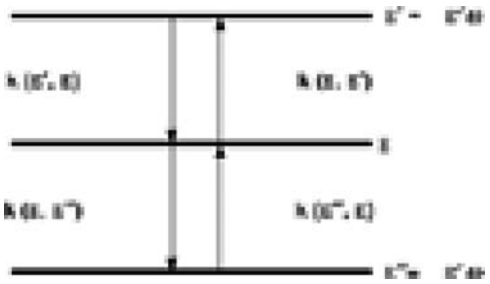

Figure 3. Illustration of four possible processes that occur when the scattering involves phonons. These processes involve either absorption of a phonon or emission of a phonon.

be scattered out to states k' (out scattering). The net scattering is simply the difference between these two processes, which gives rise to a net change in f as a function of time:

$$\begin{aligned}\frac{\partial f}{\partial t}\Big|_{\text{scattering}} &= \sum_{k'} f(k')[1-f(k)]S(k,k') \\ &\quad -\sum_{k} f(k)[1-f(k')]S(k,k')\end{aligned} \tag{22}$$

where $f(k)$ and $f(k')$ are the probabilities that states k and k' are occupied and $[1-f(k)]$ and $f(k')]$ are the probability that the two states are empty. $S(k,k')$ and $S(k,k')$ are the probability of scattering from state k to state k' and back. Note that the summation $\Sigma_{k'}$ is over all possible states k', whereas $S(k,k')$ and $S(k,k')$ contain summations over all possible scattering mechanisms by which electrons can move from state k to k' and back.

In general, the scattering probabilities $S(k,k')$ and $S(k,k')$ are energy-dependent. In particular, when the scattering involves phonons there are four possible processes, as depicted in Fig. 3. Depending on whether a phonon is absorbed or emitted, one of the four possible scatterings occurs. In a nondegenerate semiconductor where $f(E) <$‘1, Eq. (22) achieves detailed balance as (2)

$$\frac{f(E')}{f(E)} = \frac{S(E,E')}{S(E',E)} \tag{23}$$

which converts a sum over k into a sum over energy E. The right-hand side of Eq. (23) is the ratio of absorption of phonons to emission of photons which is given by

$$\begin{aligned}\frac{S(E,E')}{S(E',E)} &= \frac{N_q}{N_q+1} \\ &= e^{\dfrac{\hbar\omega_Q}{k_B T}}\end{aligned} \tag{24}$$

where N_q is the phonon number and f is a Maxwell–Boltzmann distribution. In a degenerate case, the distribution function f is a Fermi–Dirac distribution as discussed previously.

SOLUTION OF THE BTE: RELAXATION TIME APPROXIMATION

In a homogeneous semiconductor without any applied field and with an elastic (or isotropic and inelastic) scattering source, an approximate solution for the BTE can be found by writing a linear relationship for the scattering term in BTE as

$$\frac{\partial f}{\partial t}\bigg|_{\text{scattering}} = \frac{f-f_0}{\tau} \tag{25}$$

where f_0 is the equilibrium distribution function and τ is a characteristic time called *relaxation time*. Equation 25 leads to an exponentially decaying distribution function given by

$$f(t) = f_0 + (f-f_0)e^{\frac{-t}{\tau}} \tag{26}$$

which states simply that the perturbations in the distribution function f decay exponentially and the system relaxes to its equilibrium with a time constant τ. This is known as the *relaxation time approximation* (*RTA*) and is valid so long as the perturbations have not driven the system *too far from equilibrium*. It can be shown (1) that in such a system in equilibrium, even with a position-dependent band structure, the Fermi-energy and temperature throughout the system are constant:

$$\nabla_r E_F = 0 \tag{27}$$

$$\nabla_r T_L = 0 \tag{28}$$

Now we return to the BTE, Eq. (21), use the RTA, Eq. (25), and try to solve the BTE for a semiconductor that has a weak, uniform, applied electric field. Considering only the acceleration term in the BTE, and assuming that $f - f_0 \approx f_0$ (not far from equilibrium) (2),

$$\begin{aligned}f(\boldsymbol{k}) &= f_0(\boldsymbol{k}) - \frac{\tau}{h}\boldsymbol{F}\cdot\frac{\partial f(\boldsymbol{k})}{\partial k} \\ &= f_0(\boldsymbol{k}) - \tau\boldsymbol{F}\cdot\boldsymbol{v}\frac{\partial f(\epsilon)}{\partial\epsilon}\end{aligned} \tag{29}$$

Using $\boldsymbol{F} = -q\boldsymbol{E}$ for a uniform electric field, the solution to the BTE is given by

$$f(\epsilon) = f_0(\epsilon) + q\tau\boldsymbol{E}\cdot\boldsymbol{v}\frac{\partial f_0(\epsilon)}{\partial\epsilon} \tag{30}$$

Now the current density J produced by such an applied field $\boldsymbol{E}$ can be found by summing over all of the electron states as

$$J = \int_0^\infty N(\epsilon)vf(\epsilon)d\epsilon \tag{31}$$

where $N(\epsilon)$ is the density of states. Substituting for $f(\epsilon)$ from Eq. (30) in Eq. (31) gives

$$J = \int_0^\infty N(\epsilon)vf_0(\epsilon)d\epsilon - q^2\boldsymbol{E}\int_0^\infty N(\epsilon)v_E^2\tau\frac{\partial f_0(\epsilon)}{\partial\epsilon}d\epsilon \tag{32}$$

Realizing that the first term in Eq. (32) can be written as

$$J = -q^2 n\boldsymbol{E}\int_0^\infty N(\epsilon)v_E^2\tau\frac{\partial f_0(\epsilon)}{\partial\epsilon}d\epsilon \tag{33}$$

where v_E is the drift velocity caused by an applied electric field $\boldsymbol{E}$.

Equation (33) can be written as

$$J = -q^2\boldsymbol{E}\frac{\int_0^\infty N(\epsilon)v_E^2\tau\frac{\partial f_0(\epsilon)}{\partial\epsilon}d\epsilon}{n} \tag{34}$$

and realizing that the carrier concentration n is given by

$$n = \int_0^\infty N(\epsilon)f_0\epsilon d\epsilon \tag{35}$$

the current density becomes

$$J = -nq^2 \boldsymbol{E} \frac{\int_0^\infty N(\epsilon) v_E^2 \tau \frac{\partial f_0(\epsilon)}{\partial \epsilon} d\epsilon}{\int_0^\infty N(\epsilon) f_0 \epsilon d\epsilon} \tag{36}$$

Now we can introduce the energy-dependency of the relaxation time τ by noting that

$$\begin{aligned} v_E^2 &= \frac{v^2}{3} \\ &= \frac{2E}{m^*} \end{aligned} \tag{37}$$

which assumes that the kinetic energy of carriers is mainly thermal energy. Then the current density is given by

$$J = -\frac{nq^2}{m^*} \langle \tau \rangle E \tag{38}$$

where $\langle \tau \rangle$ is the average relaxation time of the ensemble given by

$$\langle \tau \rangle = -\frac{2}{3} \frac{\int_0^\infty N(\epsilon) v_E^2 \tau \frac{\partial f_0(\epsilon)}{\partial \epsilon} d\epsilon}{\int_0^\infty N(\epsilon) f_0 \epsilon d\epsilon} \tag{39}$$

Using the fact that the density of state varies as the square root of the energy, the denominator of Eq. (39) is proportional to

$$\int_0^\infty \epsilon^{1/2} f_0(\epsilon) d\epsilon = -\frac{2}{3} \int_0^\infty \epsilon^{3/2} \frac{\partial f_0(\epsilon)}{\partial \epsilon} d\epsilon \tag{40}$$

which makes Eq. (39) become

$$\langle \tau \rangle = \frac{\int_0^\infty \epsilon^{3/2} \tau(\epsilon) \tau \frac{\partial f_0(\epsilon)}{\partial \epsilon} d\epsilon}{\int_0^\infty \epsilon^{3/2} \tau \frac{\partial f_0(\epsilon)}{\partial \epsilon} d\epsilon} \tag{41}$$

$\langle \tau \rangle$, the *average scattering relaxation time*, is calculated from Eq. (41), once the energy-dependency of the particular scattering source is incorporated. Now Eq. (38) is used to define the mobility of electrons:

$$\boldsymbol{J} = nq\mu_n \boldsymbol{E} \tag{42}$$

where

$$\mu_n = \frac{q\langle \tau \rangle}{m^*} \tag{43}$$

is the mobility. Equation (43) leads to the conclusion that the carriers can attain high drift velocity in semiconductors with smaller effective masses and longer relaxation times. A slightly different derivation of average scattering relaxation time can be worked out by expressing the power-law form of energy-dependent relaxation time:

$$\tau = \tau_0 \left[\frac{E(\boldsymbol{p})}{k_B T_L} \right]^s \tag{44}$$

where $\tau 0$ is a constant relaxation time. Then by definition the average scattering time is given by

$$\langle\langle \tau \rangle\rangle = \frac{\langle E_\tau(E) \rangle}{\langle E \rangle} \tag{45}$$

where $\langle A \rangle$ is given by

$$\langle A \rangle = \frac{\int_0^\infty A f \, dp}{\int_0^\infty f \, dp} \tag{46}$$

Then the carrier mobility is defined by

$$\mu_n = \frac{q\langle\langle \tau \rangle\rangle}{m^*} \tag{47}$$

Substituting Eq. (44) in Eq. (45) and using the Γ-function

$$\Gamma(p) = \int_0^\infty y^{p-1} e^{-y} dy \tag{48}$$

the average relaxation time becomes

$$\langle\langle \tau \rangle\rangle = \tau_0 \frac{\Gamma(s + \frac{5}{2})}{\Gamma(\frac{5}{2})} \tag{49}$$

Once the power-law dependency of relaxation time on energy is known, the average relaxation time can be calculated from Eq. (49). For example, with a scattering source of s = 1/2, the average scattering time relaxation time is given by (1)

$$\langle\langle \tau \rangle\rangle = \tau_0 \frac{4}{3\sqrt{\pi}} \tag{50}$$

The Γ-function has the following three properties which enable us to evaluate $\langle\langle \tau \rangle\rangle$ for any exponent s:

$$\Gamma(n) = (n-1)!, \; n = \text{integer} \tag{51}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{52}$$

$$\Gamma(p+1) = p\Gamma(p) \tag{53}$$

CONDUCTIVITY AND MOBILITY IN SEMICONDUCTORS

Spherical Versus Ellipsoid Bands

The average relaxation time and the mobility given by Eqs. (41) and (43) are derived for a semiconductor with spherical energy bands, such as GaAs. In a semiconductor with elliptical constant-energy surfaces (such as Si) the derivation of average relaxation time (and mobility) is somewhat more complicated. Unlike the spherical energy bands in which the energy is related to a single effective mass, the energy in elliptical energy bands is related to momentum by

$$E(p) = E_{c0} + \frac{(p_x - p_{x0})^2}{2m_{xx}^*} + \frac{(p_y - p_{y0})^2}{2m_{yy}^*} + \frac{(p_z - p_{z0})^2}{2m_{zz}^*} \tag{54}$$

where m_{xx}^*, m_{yy}^*, and m_{zz}^*, are the effective masses along the three principal axes in the constant-energy surfaces, and $p_0 = (p_{x0}, p_{y0}, p_{z0})$ specifies the location of the ellipsoid's center (see Fig. 6b in the article Transport in Semiconductors, Dynamics of Carriers).

Assuming that the applied field is along the x-axis and allowing m_{xx}^*, to coincide with the longitudinal effective mass m_l^*, the contribution to conductivity from two ellipsoids along the x-axis is given by

$$\sigma_{1,2} = 2nq \left(\frac{q\langle\langle \tau \rangle\rangle}{m_l^*} \right) \tag{55}$$

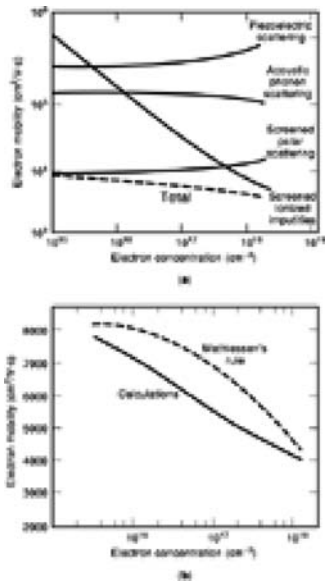

Figure 4. (a) Individual and combined mobility of GaAs as a function of electron concentration at 300 K. (b) Comparison of the room temperature mobility of electrons in GaAs with the estimate based on Mathussen's rule. After Ref. 1 originally after Ref. 29, reprinted with permission.

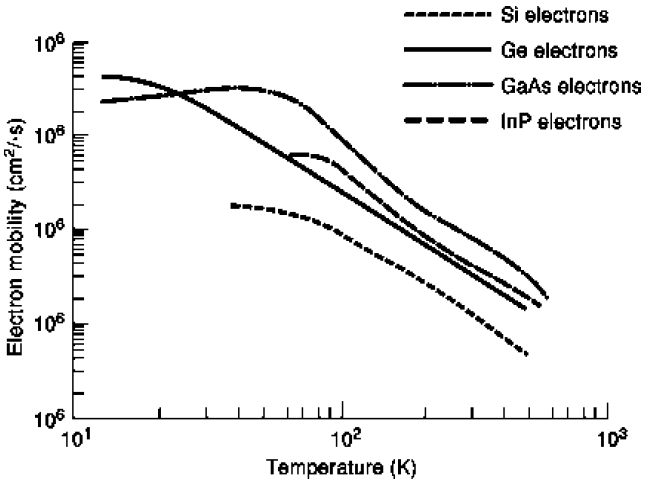

Figure 5. Temperature dependence of electron mobility in a number of semiconductors. After Ref. 2, reprinted with permission.

The other four ellipsoids respond with the transverse effective mass m_t^*:

$$\sigma_{3,4,5,6} = \frac{4n}{6} q \left(\frac{q\langle\langle\tau\rangle\rangle}{m_t^*} \right) \tag{56}$$

Adding the contributions of all six ellipsoids result in

$$\sigma = nq \left(\frac{q\langle\langle\tau\rangle\rangle}{m_c^*} \right) \tag{57}$$

with

$$\frac{1}{m_c^*} = \frac{1}{3m_l^*} + \frac{2}{3m_t^*} \tag{58}$$

where m_c^* is the *conductivity effective mass*. Although the conductivity of each ellipsoid depends on the direction of the applied field (x-direction in the above derivation), the *total* conductivity is independent of the direction because of the high degree of symmetry.

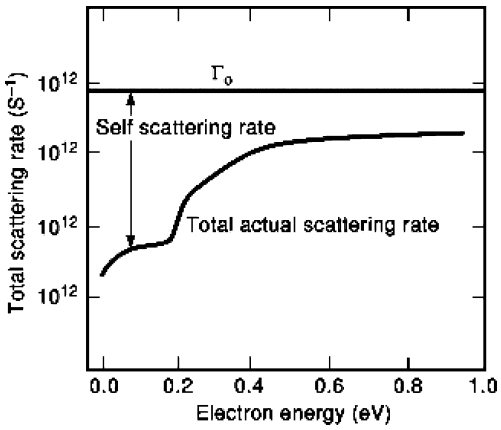

Figure 6. A schematic picture of total scattering rate versus electron energy and a fictitious scattering rate Γ_0.

Multiscattering Mobility

When several independent scattering mechanisms are present, the relaxation time approximation leads to a rule of combination of mobilities:

$$\frac{1}{\mu} = \sum_{i=1}^{n} \frac{1}{\mu_i} \tag{59}$$

where n is the number of scattering mechanisms involved, which may involve impurities, acoustic phonons, and a variety of optical phonons in a typical semiconductor. A plot of individual mobilities and total combined mobility in GaAs is shown in Fig. 4(a).

The rule of combination expressed in Eq. (59) (known as Mathiessen's rule) is valid only when the s-parameters (s is the power-law form of energy dependency of the scattering time) are equal. Because independent scattering mechanisms usually have unequal s-parameters, the use of Mathiessens rule introduces error (usually overestimation) in overall mobility. This is shown in Fig. 4(b), where the overall mobility calculated by more exact methods is compared to that estimated by Mathiessen's rule. Although the discrepancy is obvious, the ease by which Mathiessen's rule estimates the combined mobility makes it rather attractive. This is especially true considering the level of difficulty involved in simultaneously treating several independent scattering sources.

Hall Mobility

The solution of the BTE is rather complicated when a magnetic field is applied, even with the relaxation time approximation. The addition of an applied magnetic field causes a pile-up of carriers in the direction perpendicular to that of the magnetic field, a phenomenon called the *Hall effect*.The resulting conductivity (called *magnetoconductivity*) and the associated mobility are altered from the

electric mobility μ by a factor of

$$r_{\mathrm{H}} = \frac{\mu_{\mathrm{H}}}{\mu} \Rightarrow \mu_{\mathrm{H}} = r_{\mathrm{H}}\mu \tag{60}$$

where μ_{H} is the *Hall mobility* and r_{H} is the *Hall factor*. Although similar to that of mobility, the mathematics of calculating, the Hall mobility are, significantly more involved. The final result is given by (1, 2)

$$\begin{aligned} r_{\mathrm{H}} &= \frac{\langle\langle \tau^2 \rangle\rangle}{\langle\langle \tau \rangle\rangle^2} \\ &= \frac{\Gamma\left(2s+\frac{5}{2}\right)\Gamma\left(\frac{5}{2}\right)}{\left[\Gamma\left(s+\frac{5}{2}\right)\right]^2} \end{aligned} \tag{61}$$

where τ is the relaxation time and s is the exponential parameter of the power-law energy-dependence of the relaxation time. For acoustic phonons with an s-exponent of 1/2 the Hall factor is 1.18, whereas for neutral impurities the Hall factor is unity, making the magnetic conductivity and electric conductivity identical. The s-exponent and Hall factor of these and other scattering sources can be found in Ref. 1.

Temperature-Dependence of Mobility

The energy relaxation time of all scattering sources are temperature-dependent, albeit to different degrees. Equation (44), although an imperial relaxation, clearly expresses the general temperature dependence of the scattering source. This equation however, does not provide any information on the range of temperatures in which one scattering source becomes dominant compared to others. We know, however, that the deformation potential, for example, dominates at temperatures below 80 K, whereas optical polar scattering is the dominant source at room temperature and higher. Figure 5 displays the temperature dependence of electron mobility of a variety of semiconductors.

SOLUTION OF THE BTE: METHODS OF MOMENTS: BALANCE EQUATIONS

From the Boltzmann transport equation we can obtain the carrier density J and other quantities we may need to characterize a semiconductor device. Several complications arise, however, when we attempt to solve the BTE for a real device with finite dimensions subject to applied electric or magnetic fields. These complications which make finding solutions for BTE almost formidable include the following:

1. Incorporation of boundary conditions for carriers concentrations and current densities.
2. Quantum mechanical effects, especially in devices with extremely narrow regions where carriers face reduced dimensionality, such as a *two-dimensional electron gas* in a high-electron-mobility transistor. The quantum mechanical effects can be incorporated by the Schrödinger equation.
3. The electric field and the resulting potential profiles depend on the carrier density (intrinsically generated or externally injected). This dependence is usually accounted for by self-consistently solving the Poisson equation which relates the electric field and the potential profiles to the charge distribution.

These complications force us to look for simpler ways of characterizing semiconductor devices. One such way is the *method of moments* in which BTE is used to derive several *balance equations*. These balance equations describe the continuity (or conservation) of particles, momentum, energy, and other physical quantities related to the transport phenomenon. To obtain the moments of BTE we proceed as follows:

1. First, we multiply each and every term in BTE by 1/4 τ^3 and then by $\phi = \phi(\hbar k)$, which is a function of momentum. Depending on which physical quantities we want to "balance," we must choose an appropriate function. For example, $\phi(\hbar k) = 1$ results in conservation of particles, $\phi(\hbar k) = \hbar k$ conserves momentum, and energy balance requires $\phi(\hbar k) = E(k)$, where $E = \hbar^2 k^2/2m^*$.
2. Then we integrate the resulting terms in BTE over all possible momenta. To work out the integrations necessary, we may have to employ certain simplifying assumptions, such as parabolic energy bands, simplified energy-dependent scattering rates, and low and uniform field profiles.
3. Then we write the terms for any integral that has a nonzero contribution.

The resulting equation is the zeroth moment of BTE if $\phi(\hbar k) = 1$. The function $\phi(\hbar k) = \hbar k$ results in the first moment, and the function $\phi(\hbar k) = E(k) = \hbar^2 k^2/2m^*$ produces the second moment. There are, of course, higher moments of BTE. In fact there are infinitely many balance equations one can generate. Here, we discuss the first three balance equations. Complete derivation of these equations can be found in Refs. 1–3.

Carrier Density Balance Equation

Multiplying BTE by 1/4 π^3 and then by $\phi(\hbar k) = 1$ and integrating over all k space yields (1)

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot J_n + G_n - R_n \tag{62}$$

where $\partial n/\partial t$ is the net rate of increase (or decrease) of *average* carrier density $n(r, t)$ at a specified location and time. The term $1/q\nabla\cdot J_n$ is the divergence of electron flux (electron current density). G_n is the generation rate of electrons, which includes contributions from optical excitation of carriers, avalanche breakdown, or other generation processes. The last term R_n is the rate at which the electrons are lost via recombination which may include recombination through traps and Auger and radiative recombination sources. Although generation and recombination may involve interaction with scattering agents, such as phonons, the integral of the $\partial n^{\mathrm{phonons}}/\partial t_{\mathrm{scattering}}$ over all k-space vanishes. Equation (62) is the well-known *carrier continuity*

equation and can also be written for holes:

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot J_p + G_p - R_p \tag{63}$$

where p stands for holes. Usually $G_n = G_p$ because carriers are generated in pairs of electrons and holes. In many cases, especially under low injection levels the recombination rate can be written as

$$\begin{aligned} R_n &= \frac{n - n_0}{\tau_n} \\ R_p &= \frac{p - p_0}{\tau_p} \end{aligned} \tag{64}$$

where τ_n and τ_p are recombination lifetimes for electrons and holes, respectively, and are totally different from the scattering times or relaxation times discussed earlier. ∇n_0 and p_0 are equilibrium concentration of electrons and holes, respectively. The recombination rate generally can be written in terms of excess carrier concentration $n = n - n_0$ as

$$R = A\nabla n + B(\nabla n)^2 + C(\nabla n)^3 \tag{65}$$

where A is the radiative recombination coefficient, B is the coefficient for recombination through traps, (also referred to as the Schokley–Read–Hall recombination) and C is the Auger recombination coefficient.

The Momentum Balance Equation

Using $\phi(\hbar k) = \hbar k$, multiplying BTE by 1/4 π^3 and then by hk, and integrating over all k-space results in (1)

$$\frac{\partial J_n}{\partial t} = \frac{2q\nabla \cdot \boldsymbol{w}}{m^*} + \frac{q^2 n\mathscr{E}}{m^*} - \langle\langle\frac{1}{\tau_m}\rangle\rangle J_n \tag{66}$$

where $\partial J_n/\partial t$ is the rate of change of carrier current, $\nabla\cdot\boldsymbol{w}$ is the momentum flux and $\boldsymbol{w}$ is a tensor given by

$$w_{ij} = \frac{\hbar^2}{8\pi^3}\int_k v_i k_j f\, dk \tag{67}$$

The product of the velocity v_i and momentum $\hbar k_j$ is related to kinetic energy. The term $q^2 n\mathscr{E}/m^*$ is the increase in momentum caused by the accelerative force of the electric field. The term $\langle\langle 1/\tau_m\rangle\rangle J_n$ is the loss of momentum caused by scattering, and the quantity $\langle\langle 1/\tau_m\rangle\rangle$ is the *average momentum relaxation time* given by (1)

$$\left\langle\left\langle \frac{1}{\tau_m} \right\rangle\right\rangle = \frac{1}{\tau_0}\frac{\Gamma(\frac{5}{2} - s)}{\Gamma(\frac{5}{2})} \tag{68}$$

A simpler version of the momentum balance equation, known as the *drift-diffusion* equation, can be derived as follows. First, we write the momentum balance equation, Eq. 66, as

$$J_n + \frac{1}{\frac{1}{\langle\langle\tau_m\rangle\rangle}}\frac{\partial J_n}{\partial t} = \frac{q^2 n}{m^*\frac{1}{\langle\langle\tau_m\rangle\rangle}}\mathscr{E} + \frac{2q}{m^*\frac{1}{\langle\langle\tau_m\rangle\rangle}}\nabla\square \tag{69}$$

Assuming that $\partial J_n/\partial t$ is negligible during the time $1/(1/\langle\langle\tau_m\rangle\rangle)$ and defining mobility as

$$\mu_n = \frac{q}{m^*\frac{1}{\langle\langle\tau_m\rangle\rangle}} \tag{70}$$

Eq. (69) becomes

$$J_n = nq\mu_n\mathscr{E} + 2\mu_n\nabla\cdot\mathcal{W} \tag{71}$$

Now ignoring the drift velocity component of w assuming that the kinetic energy results entirely from thermal velocity, and using the equipartition of energy ($w_{ij} = w/3\delta_{ij}$), where $w = 3/2nk_B T_c$, Eq. (71) becomes (1)

$$J_n = nq\mu_n\mathscr{E} + qD_n\nabla n + qS_n\nabla T_c \tag{72}$$

where the first term is the *drift* component, the second term is the *diffusion*, and the third term is the contribution to current caused by the *thermal gradient* of the carriers. The two parameters $D_n = k_B T_c/q\ \mu_n$ and $S_n = \mu_n k_B/qn$ are the *diffusion coefficient* and the *Soret coefficient*, respectively. The contribution of $qS_n\nabla T_c$ to the current is extremely small in many isothermal transport processes.

Energy Balance Equation

Similarly, the energy balance equation is obtained as (1)

$$\frac{\partial w}{\partial t} = -\nabla\cdot\boldsymbol{F}_w + \boldsymbol{J}_n\cdot\mathscr{E} - \frac{1}{\langle\langle\tau_E\rangle\rangle}(w - w^0) \tag{73}$$

where w is the carrier energy. The first term is the energy flux, the second term is the energy gained from the field, and the last term is the energy lost to scattering agents. $\langle\langle\tau_E\rangle\rangle$ is the average energy relaxation time, and w^0 is the equilibrium carrier energy. We can go on and generate more balance equations. In fact using $\phi(\hbar k) = vE$ generates yet higher moments of BTE. Careful comparison of the three balance equations already generated reveals a very interesting fact. The continuity equation has two unknowns, n and $\boldsymbol{J}_n$, and it contains the current density contains flux $\nabla\cdot\boldsymbol{J}_n$. Needing an additional equation for the two unknowns n and $\boldsymbol{J}_n$, we generated the momentum balance equation which contains the momentum flux $\nabla\cdot\boldsymbol{w}$. But the tensor $\boldsymbol{w}$. is related to kinetic energy, and when we generate the energy balance equation, it contains the energy flux $\nabla\cdot\boldsymbol{F}_w$. Now we need another balance equation for $\boldsymbol{F}_w$, which in turn generates yet another flux. Therefore, we must truncate this chain of balance equations. Indeed, we did truncate the chain of balance equations when we wrote the drift-diffusion equation for isothermal conditions as

$$\boldsymbol{J}_n = nq\mu_n\mathscr{E} + qD_n\nabla n \tag{74}$$

Equations (62) and (74), together with auxiliary equations for generation and recombination, and with all scattering effects lumped in a mobility model of some kind, such as $\mu = \mu(T, \mathscr{E}, N_D$, have been the basis of modeling and characterizing many devices ranging from diodes to transistors and to solar cells. These models are called *drift-diffusion models*. In the last section we describe the highlights of a model developed by the authors, in which the two higher moments of BTE along with the Poisson and Schrödinger equations are self-consistently solved for a high-electron-mobility transistor; see Refs. 4 through 10.

MONTE CARLO SIMULATION TECHNIQUE

Carrier dynamics within a device structure can be obtained by the Boltzmann transport approach (*BTE*), the hydrody-

namic model (*HD*), the Monte Carlo (*MC*) technique, and the molecular dynamics (*MD*) technique. In this section, the fundamentals of the MC techniques and its salient features and limitations are described.

In the MC technique, individual trajectories of carriers are simulated by using Newton's second law allowing for intermittent random scattering of the particle during its travel. A detailed description of the dynamics of the carriers can be obtained from ensemble averages of various dynamic quantities: position, velocity, and kinetic and potential energies.

Typically, the initial conditions of various particles are obtained from the temperature and the Fermi level at the initial contact using either the Maxwell–Boltzmann or Fermi–Dirac distributions, depending on the dopant concentration. The particle starting at the initial contact has an initial momentum and is subjected to position-dependent electric fields. The particle motion in between two consecutive collisions (free flight) is simulated by using Newton's second law. Assuming an electric field $\boldsymbol{E}$, the force $\boldsymbol{F}$, which is the rate of change of momentum $\boldsymbol{p}$, is given by

$$\begin{aligned} \boldsymbol{F} &= \frac{d\boldsymbol{p}}{dt} \\ &= (-q)\boldsymbol{E} \end{aligned} \tag{75}$$

for an electron. Based on Eq. (75), the position and velocity of one electron is computed after its travel for the intercollisional time t_c. The time t_c, between collisions is determined on the basic of a conditional probability given by:

$$t_c = -\frac{1}{\Gamma_0} ln(r_1) \tag{76}$$

where r_1 is a pseudorandom number uniformly generated between 0 and 1 and Γ_0 is the maximum scattering rate of the carrier which is usually chosen to be larger than largest scattering rate of the carriers. A schematic diagram showing the total scattering rate versus energy of the electron is depicted in Fig. 6. The reason for the choice of Γ_0 is to avoid using an energy-dependent scattering rate in Eq. (76). The difference between Γ_0 and the actual scattering rate is called the *fictitious scattering rate*, and a carrier undergoing such an event just continues its travel without any change in its momentum. Use of Γ_0 has no influence on the physics of the problem, but makes the MC code extremely efficient. Knowing t_c, one can compute the kinetic energy of the carrier at the time of collision. Knowing the initial kinetic energy and momentum, one can identify the rates of scattering for various mechanisms. For a discussion of scattering rates see the previous article entitled "Transport in Semiconductors: Dynamics of Carriers." Using a pseudorandom number uniformly distributed between 0 and 1 and the relative rates of various scattering mechanisms including the fictitious scattering, one can pick the scattering event that occurred. Choice of a uniform distribution guarantees that the fastest scattering event is chosen most often and the slowest event, least often. Based on the nature of the scattering event, such as elastic or inelastic and isotropic or anisotropic, the final momentum and energy of the particle after the collision are determined from energy and momentum conservation principles.

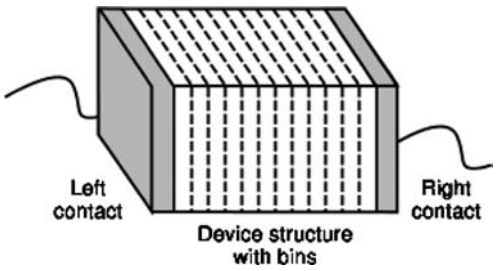

Figure 7. A schematic picture of a device structure with partitioned bins.

Two different Monte Carlo approaches are employed for studying the carrier dynamics within a device; the *many-particle approach* and the *incident-flux approach*. In the many particle Monte Carlo method, the device is divided into many slabs, and carriers are distributed within them based on the Maxwell–Boltzmann distribution. A schematic diagram in which the device is partitioned into *bins* is shown in Fig. 7. The number of carriers in each slab is initially chosen to satisfy the charge neutrality condition based on the local dopant concentrations. Typically, the contacts are assumed to be perfect absorbers. The simulation is started with the initial conditions described. After every few *femtoseconds*, Δt, the carrier concentration in every slab is checked and updated to satisfy charge neutrality. The Poisson equation is solved to recompute the modified electrostatic potentials. By continuing the simulation for many time intervals, one can achieve a steady-state condition. At any time during the simulation, one can collect data, such as carrier concentration, average velocity, average energy, and other required dynamic properties as functions of position. A flow chart for the simulation is shown in Fig. 8.

In the incident-flux MC approach, carriers are introduced into the device from one of the contacts that have an initial velocity, and the carriers are followed through the device allowing them to scatter and respond to the local electrostatic fields until they reach the other contact. The initial momentum, (p_x, p_y, p_z) of the particle is chosen on the basis of three random numbers, r_x, r_y, and r_z, respectively, chosen from a uniform distribution, depending on the type of contact and the direction of carrier injection. The dynamics of many particles with varying initial momentum are simulated, and average dynamic properties are computed from the individual statistics. To eliminate statistical fluctuations, typically, 10,000 carriers are simulated. A flow chart describing the algorithm is shown in Fig. 9.

The MC simulation technique has several limitations. First, one needs to run large simulations to eliminate statistical fluctuations. Secondly, the simulation is CPU time-intensive. Thirdly, treating electron–electron scattering in the simulation, which is important for degenerate materials, is difficult. Finally, including the band structure information, which is important for accurate results in realistic semiconductor systems, makes the simulations computationally intensive. There are several excellent reviews on applications of MC simulation to semiconductors (11–15) and to devices (16, 17).

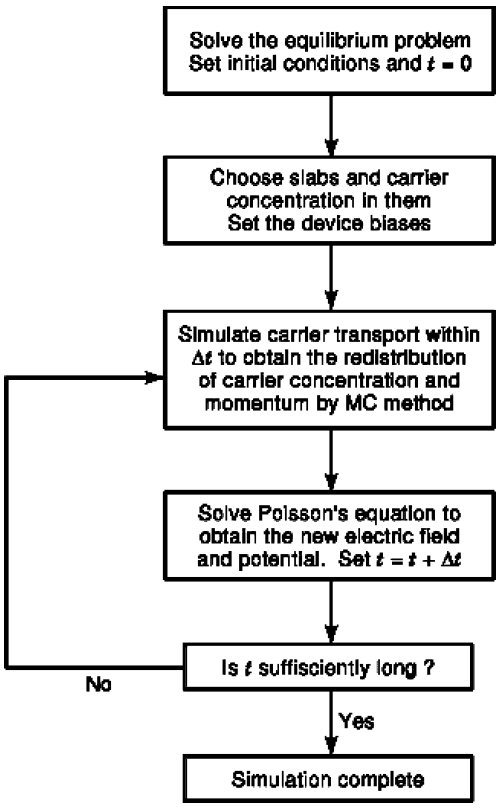

Figure 8. A flow chart for the many particle Monte Carlo method.

SELF-CONSISTENT HYDRODYNAMIC NUMERICAL MODELS

Hydrodynamic models consisting of simultaneous solution of the moments of the Boltzmann transport equation have been the basis of many analytical and numerical models in the recent literature. In this section we present the highlights of a self-consistent Boltzmann–Poisson–Schrödinger solver in which the two higher moments of the Boltzmann transport equation are numerically solved along with the Poisson and Schrödinger equations. This model, which is developed for a *high-electron-mobility transistor* (*HEMT*), takes into account the effects of quantization of the *two-dimensional electron gas* in the channel of the device. The two higher moments of the BTE in the form of a current continuity equation and an energy balance equation are solved to obtain the transient and steady-state behavior. Using this Boltzmann–Poisson–Schrödinger solver the $(i_{ds} - v_{ds})$ characteristics, transconductance, and unity-gain frequency of a single quantum well HEMT can be simulated. For details of this model and its further extension to a "full quantum two-subband model," see Refs. 4–10.

The HEMT Structure

The HEMT structure illustrated in Fig. 10 is considered for simulation. The gate length is 0.5 μm, and two 0.5 μm regions on the two sides of the gate separate the gate from the source and drain. Beneath the gate is a highly doped $Al_{0.3}Ga_{0.7}AS$ layer of 50 nm wide and that has a doping level of $5{\cdot}10^{17}$ cm^{-3}. A 10 nm undoped AlGaAs spacer lies

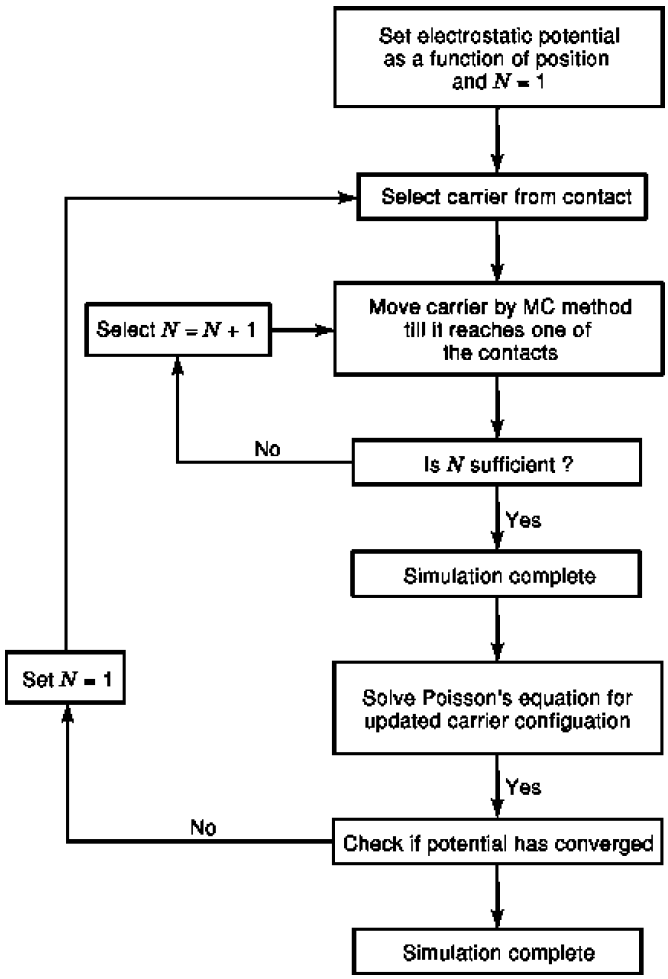

Figure 9. A flow chart for the incident-flux Monte Carlo method. N is the number of carriers simulated.

between the highly doped AlGaAs layer and the GaAs layer. Such a spacer is included to separate the free electrons in the GaAs channel from their donor impurities in AlGaAs, thereby reducing the remote scattering of electrons by ionized impurities (18). The GaAs region consists of a quantum well 100 nm wide and a bulk layer of 300 nm wide and the two layers overlapping each other, as shown in Fig. 10. The doping level of GaAs is 10^{14} cm^{-3}. On the two sides are boundaries to two highly-doped GaAs regions that are ohmic contacts to the source and drain.

Moments of the Boltzmann Transport Equation

To start, we assume that the two-dimensional electron gas in the channel consists mainly of the electrons in the lowest subband and that electrons in the second and higher subbands are three-dimensional bulk electrons without any reduced dimensionality. The current continuity equation that describes the transport of electrons and an energy balance equation that describes the spatial and temporal variation of the average electron energy in the *bulk* of the device are written as:

$$\frac{\partial n_b}{\partial t} = \nabla_b \cdot [-\mu_b n_b \nabla_b V + \nabla_b (D_b n_b)] + G_b \qquad (77)$$

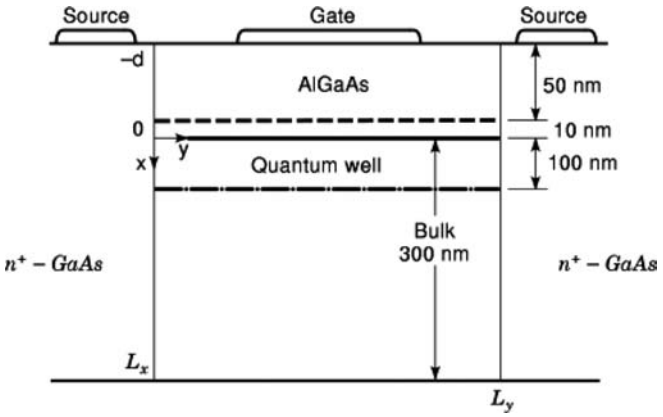

Figure 10. The HEMT structure used in the simulation. The quantum well and the bulk regions overlap with a total width of 300 nm. After Ref. 11, reprinted with permission.

and

$$\frac{\partial n_b E_b}{\partial t} = -j_b \cdot \nabla_b V - n_b B_b + \nabla_b \cdot \alpha[-\mu_b n_b E_b \nabla_b V + \nabla_b (D_b n_b E_b)] + F_b \tag{78}$$

and for the *quantum well* these equations are:

$$\frac{\partial n_{qw}}{\partial t} = \nabla_{qw} \cdot [-\mu_{qw} n_{qw} \nabla_{qw} V + \nabla_{qw} (D_{qw} n_{qw})] + G_{qw} \tag{79}$$

and

$$\frac{\partial n_{qw} \boldsymbol{E}_{qw}}{\partial t} = -\boldsymbol{j_{qw}} \cdot \nabla_{qw} V - n_{qw} B_{qw} + \nabla_{qw} \cdot \alpha[-\mu_{qw} n_{qw} E_{qw} \nabla_{qw} V + \nabla_{qw} (D_{qw} n_{qw} E_{qw})] + F_{qw} \tag{80}$$

where subscripts b and qw denote the bulk and the quantum well, respectively. The terms n and $\boldsymbol{j}$ are the electron concentration and current density, V is the potential, $\boldsymbol{E}$ is the average electron density, μ is the mobility, D is the diffusion constant, and B is the energy dissipation factor. The term G is a generation-like term used to redistribute the electrons between the bulk and the quantum system, and the term F accounts for energy transfer between the two systems, as explained in the following section. The coefficient α is the ratio of flux mobility to carrier mobility given by:

$$\alpha = \frac{\mu_E}{\mu} = \frac{\langle \tau E_i^2 \rangle}{\langle \tau E_i \rangle \langle E_i \rangle} \tag{81}$$

where τ is the momentum relaxation time E_i is the electronic energy, and the brackets in Eq. (81) refer to statistical averages over the entire sample. For a power-law scattering $\alpha = 2/3\ (p + 5/2)$, where the power-law scattering is defined as $\tau \mu E_i^p$. For polar optical phonons in GaAs, p has a value of 0.5. In deriving these equations, it is assumed that the high-frequency terms in the Boltzmann moment equations are negligible because the high frequency parameter, τ_{HF}, is on the order of 0.1 ps (19), whereas the typical calculated transient time is on the order of 3 ps.

Poisson Equation

The previous transport equations are solved along with Poisson equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = -\frac{q}{\epsilon}[N_D(x, y) - n(x, y)] \tag{82}$$

where V is the electrostatic potential, ϵ is the dielectric constant, n is the total electron concentration in the channel, and N_{D} is the impurity doping level.

Schrödinger Equation

The quantization of electron energy levels and the spatial spread of electrons in the quantum well are taken into consideration by a self-consistent solution of the Schrödinger and Poisson equations. The Schrödinger equation that describes the electrons in a quantum well is given by:

$$-\frac{\hbar^2}{2m_x}\frac{d^2\psi_i(x)}{dx^2} - qV(x, y)\psi_i(x) = E_i \psi_i(x) \tag{83}$$

where m_x is the electron effective mass in the x-direction, ψ_i is the wave function corresponding to the eigenvalue E_i for the ith subband, and $V(x, y)$ is the electrostatic potential. The boundary conditions require that the wave functions vanish at both infinities. The Schrödinger equation is solved by using a Rayleigh–Ritz variational method (20).

To model the quantum well, one possible approach is to define an artificial boundary across the GaAs region that separates the bulk system from the quantized system. Electrons confined by such artificial boundary and the heterojunction are considered quantized and their motion is restricted to the y-direction, whereas electrons lying below the artificial boundary are considered bulk carriers that have restricted motion. Such an approach has a number of shortcomings. First, no definite rule defines the quantum well/bulk boundary. Because the wave function spans a relatively wide region in the quantum well, much of the wave function outside the boundary is truncated and the quantum effects are greatly distorted if the well width is taken too small. On the other hand, if the well width is taken sufficiently large to include a significant portion of the wave function, the bulk electronic behavior is neglected. Neither case is desirable for simulating a device. Secondly, the electron concentration over the quantum well/bulk boundary is generally discontinuous, which gives rise to a large diffusion current across the boundary. This can cause erroneous results in the simulation. Thirdly, at points where the electric field at the heterojunction is weak, the quantum well is too shallow to confine the electrons, and the electrons at the heterojunction behave essentially as bulk carriers. Therefore, it is important that both the bulk and the quantum characters of the electrons are considered, particularly at the heterojunction where the concentration of electrons is highest.

To avoid setting an artificial boundary separating the bulk GaAs from the quantum well, we assume that the two systems overlap. The quantum well and the bulk systems both start at $x = 0$. The quantum well (electrons residing in the lowest subband, E_0) spreads a distance d into the GaAs layer as shown in Fig. 10. The quantum well's width is determined from the wave function for the lowest sub-

band. The bulk system (electrons residing in the second subband, E_1, and higher subbands) extend 300 nm into the GaAs layer. Because of this overlap of the two systems, the electrons at any point across the channel can be in the quantum well or in the bulk, depending on their eigenenergies and the quantum well width at that point. Electrons in the bulk undergo transport in both the x- and y-directions, whereas electrons in the quantum well are restricted to transport only in the y-direction.

The transport of electrons in the device is governed by six coupled, nonlinear, partial differential, equations, Eqs. (77), (78), (79), (80), (82), and the Schrödinger equation, Eq. (83) that have the unknown variables n_{qw}, E_{qw}, n_b, E_b, V, $y_i(x)$, and E_i. The system of equations consisting of these six equations along with several auxiliary equations is solved iteratively by using a two-dimensional finite-difference scheme; see Refs. (4) through (10). In each iterative cycle the values of the variables $n_{qw}, E_{qw}, n_b, E_b, V$, $y_i(x)$, and E_i are updated until the correction terms to these variables are within a tolerance range. The Schrödinger equation is solved at each (x, y) point, and the values of the wave functions and the eigenenergies are updated in each iterative cycle. The discretization of the current continuity equation the numerical instability problems associated with the discretized equation, and the time step and spatial mesh sizes used in these simulations are described in a later section.

Full Quantum Two-Subband Transport

At 77 K, 98% of the electrons in the quantum well of a HEMT device reside in the lowest subband, whereas at 300 K, the population of electrons residing in the lowest subband is reduced to 68% (21). In fact, at 300 K close to 20% of electrons in the quantum well reside in the second subband, and their motion is restricted to two dimensions.

The Boltzmann–Schrödinger–Poisson simulator described previously is a one-subband model in which it is assumed that only the electrons in the first subband are quantized and that the electrons in the second and higher subbands behave as bulk electrons unrestricted in their motion. This assumption becomes questionable particularly at 300 K because only 68% of electrons reside in the first subband.

Now we extend the one-subband Boltzmann–Poisson–Schrödinger (*OS-BPS*) simulator to a full quantum two-subband Boltzmann–Poisson–Schrödinger (*FQ-BPS*) model. This is done by considering that electrons in the lowest *two subbands* are in the quantum well form the two-dimensional electron gas and the electrons in the third and higher subbands behave as bulk electrons unrestricted in their motion. In the FQ-BPS model, we solve the Boltzmann, Poisson, and Schrödinger equations for the first subband, second subband, and the bulk system. An additional self-consistency is added by calculating the intersubband and intrasubband scattering due to polar optical phonon and ionized impurity scattering mechanisms (7). The rates of transfer of electrons and their energies to and from each subband are calculated from the intersubband and intrasubband scattering rates and are used self-consistently in the Boltzmann–Poisson–Schrödinger simulator. In the FQ-BPS model, the two higher moments of Boltzmann transport equation are given by

$$\begin{aligned}\frac{\partial[n_i(x,\,t)]}{\partial t} &= \nabla\cdot\{-\mu_i n_i(x,t)\nabla V(x)+\nabla[D_i n_i(x,t)]\}\\ &+ \sum_{j\neq i}\left(\frac{n_j-n_{oj}}{\tau_{ji}}\right)-\sum_{j\neq i}\left(\frac{n_i-n_{oj}}{\tau_{ij}}\right)\quad i=1,2\end{aligned}\tag{84}$$

$$\begin{aligned}\frac{\partial[n_i(x,t)E_i(x,t)]}{\partial t} &= -\,J\cdot\nabla V(x)+\nabla\cdot\{-\mu_{E,i}n_i(x,t)E_i(x,t)\nabla V(x)\\ &+ \nabla[D_{E,i}n_i(x,t)E_i(x,t)]\}\\ &+ \sum_{j\neq 1}\left(\frac{n_jE_j-n_{j0}E_{0j}}{\tau_{E,ji}}\right)\\ &- \sum_{j\neq i}\left(\frac{n_iE_i-n_{i0}E_{i0}}{\tau_{E,ij}}\right)\\ &- \sum_{j\neq i}\left(\frac{n_j-n_{0j}}{\tau_{ji}}\right)\hbar\omega_0\quad i=1,2\end{aligned}\tag{85}$$

In these equations i = 1, 2 refers to the first and second subbands, respectively. n is the electron concentration, J is the electron current density, q is the electronic charge, μ_i is the mobility, D_i is diffusivity, $\mu_{E,i}$ is flux mobility, $D_{E,i}$ is the flux diffusivity, E is the average electron energy, τ_{ij} is the particle relaxation time, $\tau_{E,ij}$ is the energy relaxation time for particles moving from subband i to subband j, and $\hbar\omega_0$ is the polar optical phonon energy. ∇ represents $\partial/\partial y$ in the one-dimensional quantum well. The first summation term $(\sum_{j\neq i}(n_i - n_{0i}/\tau_{ji}))$ in Eq. (84) accounts for the particles moving from subband j to subband i, and the second summation term $[-\sum_{j\neq i}(n_i - n_{0i}/\tau_{ij})]$ in Eq. (84), accounts for the particles moving from subband i to subband j. The first summation term $[\sum_{j\neq i}(n_jE_j - n_{j0}E_{0j}/\tau_{E,ji})]$ in Eq. (85) accounts for the energy gained from the movement of particles from subband j to subband i, and the second summation term $[-\sum_{j\neq i}(n_iE_i - n_{i0}E_{i0}/\tau_{E,ij})]$ in Eq. (85), accounts for the energy loss from the movement of particles from subband i to subband j. Finally, the last summation term $[-\sum_{j\neq i}(n_j - n_{0j}/\tau_{ji})\hbar\omega_0]$ in Eq. (85) includes the effect of the loss of energy by the electrons to the polar optical phonons.

In the bulk (electrons in the third and higher subbands), current continuity and energy balance equations are

$$\frac{\partial n}{\partial t} = \nabla[-\mu n\nabla V+\nabla(Dn)]+G_i\quad i=1,2\tag{86}$$

$$\begin{aligned}\frac{\partial(nE)}{\partial t} &= -J\cdot\nabla V - nB+\nabla\alpha[-\mu nE\nabla V+\nabla(DnE)]+F_i\\ i &= 1,2\end{aligned}\tag{87}$$

where B is the energy dissipation factor, α is a constant relating μ to μ_E and D to D_E. ∇ represents $(\partial/\partial x, \partial/\partial y)$ in the two-dimensional bulk. The term G_i is the generation-like term that accounts for the transfer of electrons between the bulk and the first (i = 1) and second (i = 2) subbands. F_i is a similar term that account for the rate of energy transfer between the bulk and the two subbands (8).

Self-Consistent Scattering Rates

The eigenfunctions obtained from the Boltzmann–Poisson–Schrödinger simulator are used to calculate the ionized impurity scattering and the polar optical

phonon scattering rates for the two lowest subbands in the quantum well. The rates of transfer of electrons and their energies to and from each subband are calculated from these intersubband and intrasubband scattering rates. The polar optical phonon scattering rate is given by (21)

$$S_{mn}^{pop} = \frac{e^2\omega_0}{8\pi\epsilon_0}\left[\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon_s}\right]\left(N_q+\frac{1}{2}\pm\frac{1}{2}\right) \times \int \frac{H_{mn}(Q)}{Q}\,\delta[E(\mathbf{k}_2)-E(\mathbf{k}_1)\pm\hbar\omega_0]d\mathbf{k}_2 \tag{88}$$

where ϵ_∞ and ϵ_s are the optical and static dielectric constant, $\hbar\mu_0$ is the polar optical phonon energy, Q is the phonon wave-vector component parallel to the interface, $\mathbf{k}_1$ and $\mathbf{k}_2$ denote the initial and final state wave vectors. N_q is the phonon occupation number, and $E(\mathbf{k}_1)$ and $E(\mathbf{k}_2)$ are the initial and final state energies. $H_{mn}(Q)$s are the multisubband coupling coefficients. The ionized impurity scattering rate is given by

$$S_{mn}^{\text{imp}} = \frac{1}{2\pi\hbar}\int |M_{mn}(Q)|^2\delta(E(\mathbf{k}_2)-E(\mathbf{k}_1))d\mathbf{k}_2 \tag{89}$$

The matrix elements $M_{mn}(Q)$ account for the electron–impurity interactions. For more details of the modeling of scattering rates, see Ref. 7.

Numerical Methods

As shown in the flow chart of Fig. 11, we begin the numerical simulation with an initial guess for the electron densities in the bulk and in the quantum well. Then, we start the iterative process by solving the Poisson and Schrödinger equations self-consistently, from which we obtain the electrostatic potential V, eigenenergies E_i, and eigenfunctions ψ_i. The scattering rates are calculated next. The mobilities and the coupling terms in the moments equations are derived from these scattering rates. Finally we solve the two moments of the BTE and obtain the $(i_{ds}-V_{ds})$ characteristics of the device.

The numerical solution of the partial differential equations outlined previously requires iterative computation both in time and space. Numerical instability problems often occur with such an iterative technique. An explicit approach is commonly employed to solve the continuity equation,

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla\cdot\mathbf{J}+G \tag{90}$$

where

$$\mathbf{J} = [-\mu n\,\nabla V+\nabla(Dn)] \tag{91}$$

Equation (90) is discretized into the form

$$n^{k+1} = n^k+\nabla t\left[\left(\frac{1}{q}\nabla\cdot\mathbf{J}\right)+G\right]^k \tag{92}$$

The values of n, V, and G at time k are inserted into the right-hand side of Eq. (92) and yield the value of n at the next time step $(k+1)$. Although this method is extremely straightforward, it requires excessively small time steps to guarantee numerical stability. Specifically, it has been

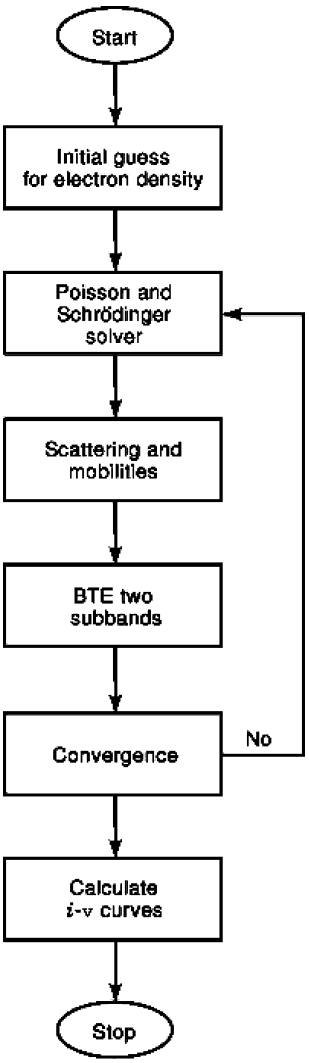

Figure 11. Flow chart of the numerical iteration in the FQ-BPS model. After Ref. 13, reprinted with permission.

shown (22) that the maximum time step without any instability is given by

$$\Delta t < \min\left[\frac{\Delta x^2\Delta y^2}{2D(\Delta x^2+\Delta y^2)}, \frac{2D}{v_\infty^2}\right] \tag{93}$$

where Δx and Δy are the mesh spacings, D the diffusivity, and v_∞ the saturation velocity. When the maximum allowable time step is exceeded, a minor perturbation in the values of $n_{i,j}$ at mesh point (i,j) results in a diverging solution. The smallest mesh dimensions Δx and Δy used in these simulations are $2{\cdot}10^{-7}$ cm and $5{\cdot}10^{-6}$ cm, respectively. The diffusivity at low field is about 300 cm^2/s. Assuming that the mesh spacing is the factor limiting the speed of the iterative process, the maximum time step without causing numerical instability problem is (for all values of v_∞)

$$\Delta t < 6.6\cdot 10^{-17}\,\text{s}$$

which is on the order of 10^6 times smaller than the typical transient time of HEMT. This poses a serious problem for the convergence of the program. To increase the time step

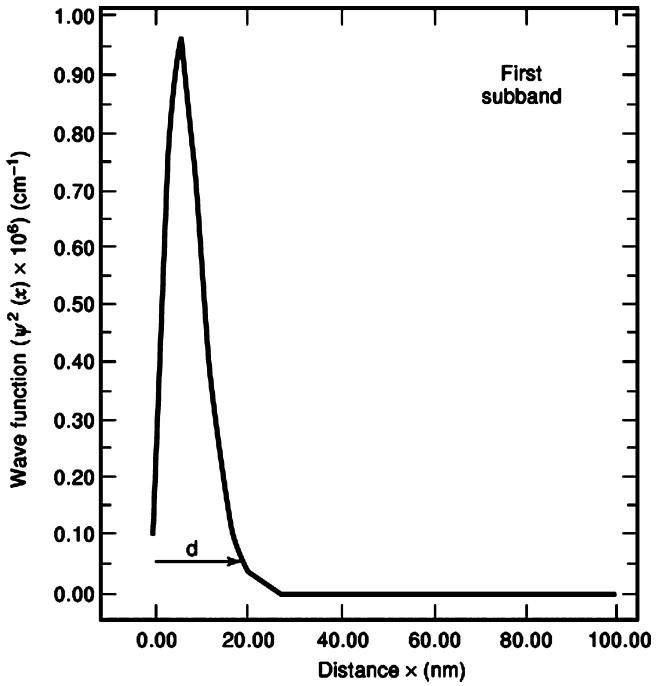

Figure 12. Calculated wave function $\psi_1(x)^2$ of the lowest subband at 300 K with an applied gate voltage of 0.5 V and a drain bias of 1.5 V. The wave function peaks at 7 nm from the interface and spreads about 20 nm in the GaAs layer. After Ref. 11, reprinted with permission.

and to speed up the program, one has to increase the mesh spacing which in turn sacrifices the accuracy of the result. Another approach to the solution of the continuity equation is to write the equation in an implicit form:

$$\frac{\partial n}{\partial t} = \frac{1}{2q}[\nabla \cdot \mathbf{J}^k + \nabla \cdot \mathbf{J}^{k+1}] + G^k \tag{94}$$

where the superscript k represents time. The price to pay is complicated discretization and a tedious solution. Using the *implicit discretization scheme* given by Eq. (95), the time step size used in these simulations is 10^{-15} s. The spatial mesh employed has nonuniform mesh spacing in the x-direction and mesh sizes Δx ranging from $2{\cdot}10^{-7}$ cm to $5{\cdot}10^{-6}$ cm. The mesh sizes are smaller in the proximity of the heterojunction. The mesh spacing Δy in the y-direction is uniform and has a mesh size of $5{\cdot}10^{-6}$ cm.

Results of OS-BPS and FQ-BPS Simulations

Figure 12 shows the square of the wave function $\psi_1(x)^2$ for the first lowest subband that has an applied gate voltage of 0.5 V and a drain bias of 1.5 V. The function $\psi_1(x)^2$ peaks at about 7 nm from the interface to a value of about $96{\cdot}10^4$ cm^{-1}. The spatial spread of the well is 20 nm. The calculated eigenenergies for the first and second subbands are 46 and 69 meV, respectively.

Figure 13 shows the electrostatic potentials in the device with a gate voltage of 0.7 V and two drain biases of 0.5 V and 1.35 V. Figure 14 shows the electron concentrations in the quantum well and the bulk GaAs under the same biasing conditions. The electron concentration in the quantum well is calculated by multiplying the electron

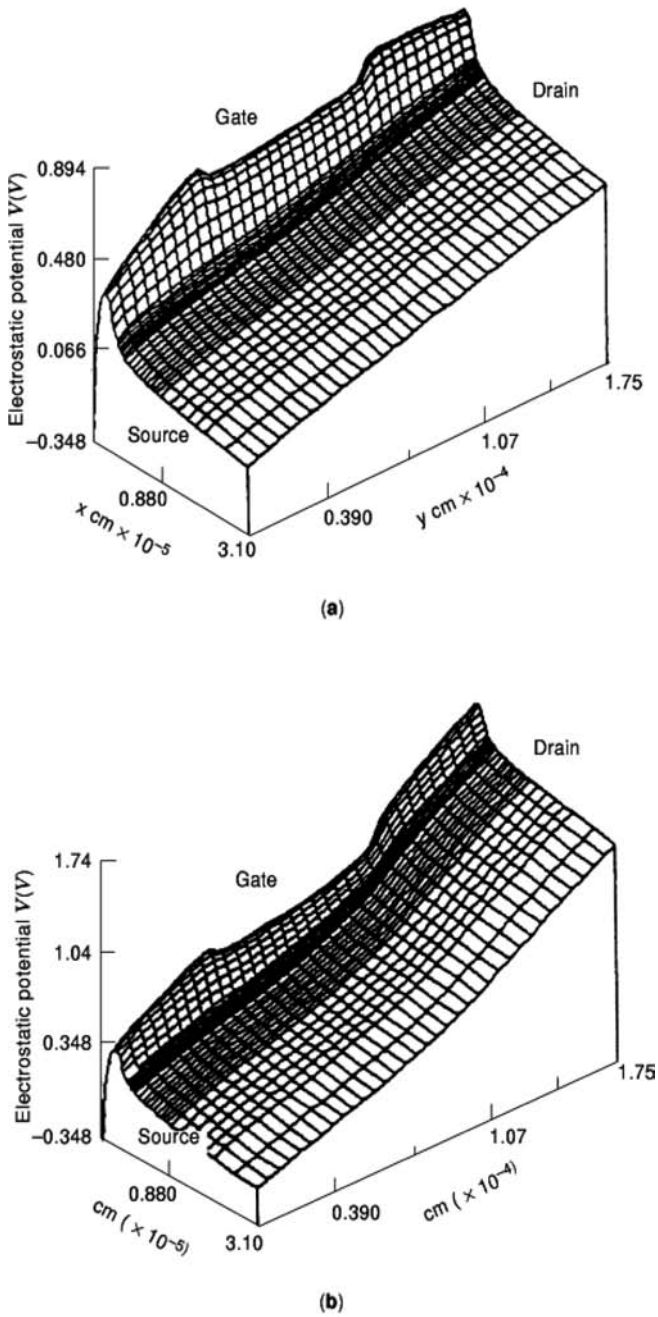

Figure 13. Electrostatic potential profiles with a gate voltage of 0.7 V and drain voltage of (a) 0.5 V, and (b) 1.35 V. These potentials are calculated by the one-subband Boltzmann–Poisson–Schrödinger solver. After Ref. 11, reprinted with permission.

sheet density in the quantum well by the probability density $\psi_1(x)^2$. When $V_D = 0.5$ V, the electron concentrations in the quantum well and the bulk GaAs are relatively uniform from source to drain, whereas when $V_D = 1.35$ V, the electron concentration is sharply reduced in the region underneath the drain end of the gate, which is the *pinch-off* point. The spatial spreading of the quantum well and the bulk systems are also shown in Fig. 14. It should be noted that although the concentration of electrons in the quan-

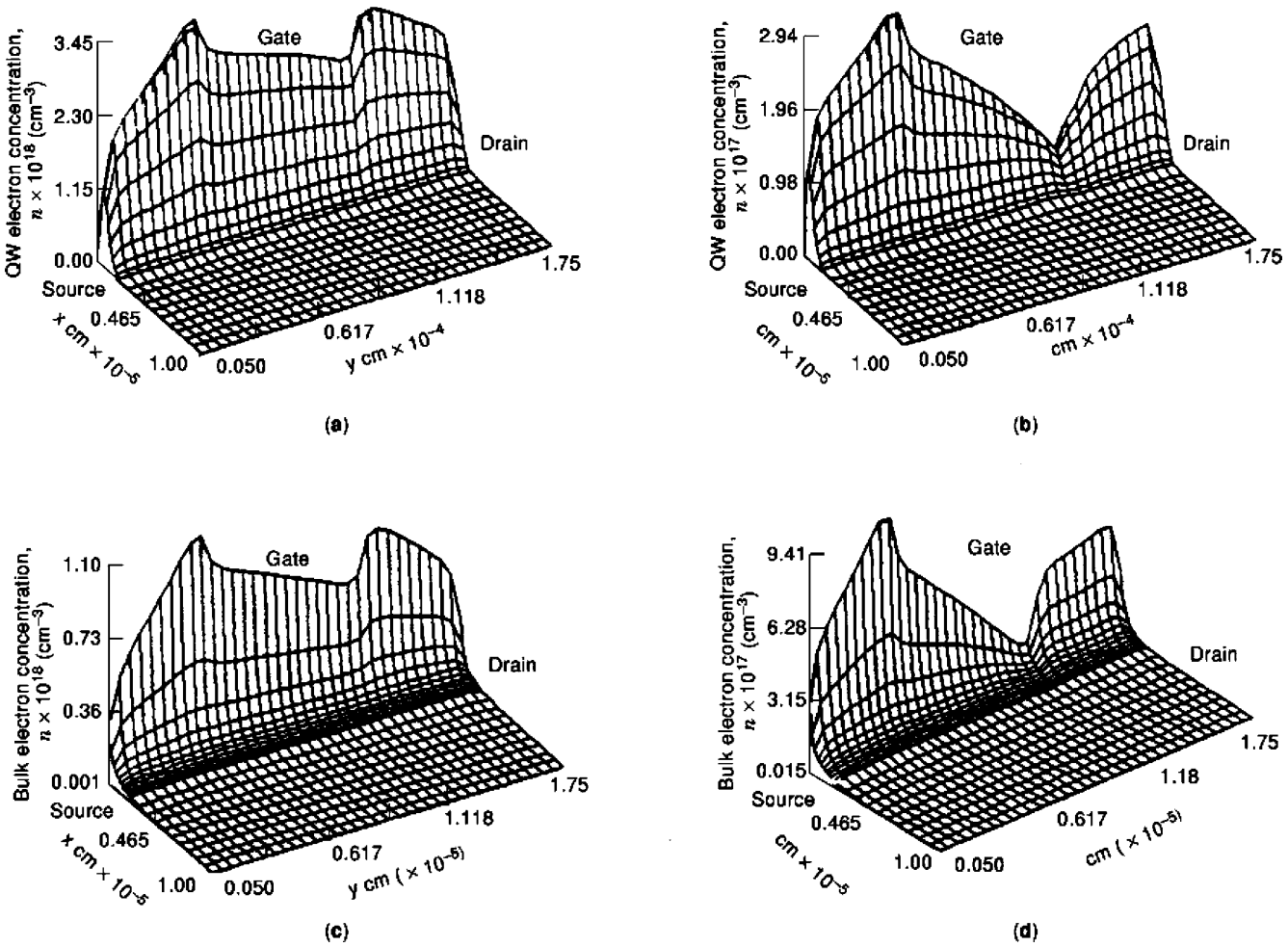

Figure 14. Electron concentrations in (a, b) quantum well and (c, d) bulk GaAs. The applied gate voltage is 0.7 V in all cases. The drain bias is 0.5 V in (a) and (c) and 1.35 V in (b) and (d). These results are calculated by the one-subband Boltzmann–Poisson–Schrödinger solver. After Ref. 11, reprinted with permission.

tum well reduces to zero at a distance of 200 Å from the interface [Fig. 14(a)], the concentration of electrons in the bulk system reduces to 15×10^{14} approaching the doping level of the bulk GaAs at a distance of 1000 Å from the interface. Thus, it is concluded that the quantum well width is about one-fifth of the active bulk layer. The results shown in Figs. 12–14 are calculated by the one-subband Boltzmann–Poisson–Schrödinger solver.

Figure 15 shows the $i_{ds} - v_{ds}$ characteristics of the device calculated by the OS-BPS simulator along with those simulated by the full quantum two-subband Boltzmann–Poisson–Schrödinger solver. In both cases, the slopes of the $i_{ds} - v_{ds}$ curves decrease as drain voltage increases, but the FQ-BPS model predicts lower drain currents. Because the scattering rates are included in the model, the electron density in the quantum well decreases, which results in lower drain currents. Overestimation of the drain current by the OS-BPS model can be seen from the $i_{ds} - v_{ds}$ characteristics of Fig. (15). When a gate bias of 0.7 V is applied in the FQ-BPS model, the slope of the drain current decreases substantially above a drain voltage of 0.8 V, which is the saturation region, whereas the drain current obtained from OS-BPS model still has a sharp slope as the drain voltage is increased, and the onset of saturation occurs at around 1.2 V. Therefore, the FQ-BPS model shows that the device goes into saturation at a lower drain voltage than the value predicted by the OS-BPS model.

The transconductance of the device as a function of gate bias using the two simulators is plotted in Fig. 16. The value of transconductance with the OS-BPS model decreases slightly (and almost linearly) as the gate voltage is varied over the range of 0.6 V to 1.0 V, whereas the transconductance increases at first with the FQ-BPS model and reaches a peak value of about about 470 mS/mm at 0.9 V, and then decreases beyond that point. The pattern of the FQ-BPS data is closer to the experimental results reported (23, 24). The transconductance degradation at high gate voltages has been investigated by many researchers. Among the reasons cited for this degradation are increased gate leakage current (23) reduced inversion charge in the channel (25), dislocations (26), decreased average electron velocity (24), increased population of electrons in the donor AlGaAs layer (27), and neutralized donor effect (28). The

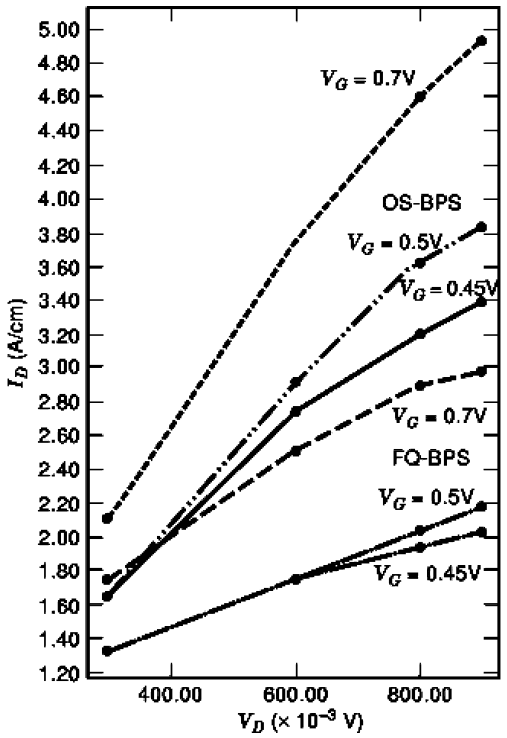

Figure 15. (a) $i_{ds} - v_{ds}$ characteristic: (OS-BPS) are the results from the one-subband Boltzmann–Poisson–Schrödinger solver, and (FQ-BPS) are the results from the full quantum two-subband Boltzmann–Poisson–Schrödinger simulator. After Ref. 13, reprinted with permission.

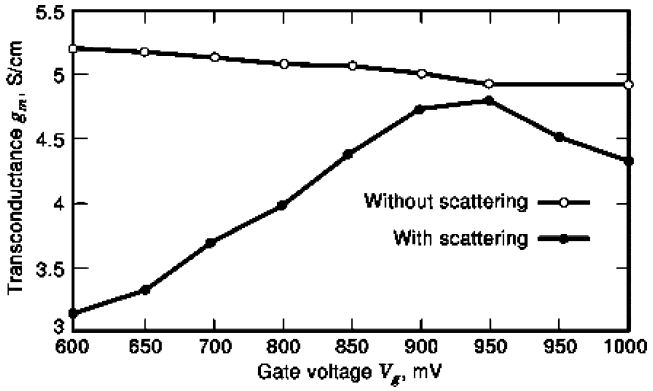

Figure 16. Transconductance as a function of gate bias in a HEMT device calculated by our Boltzmann–Poisson–Schrödinger solver (•) with self-consistent scattering rates (ç) without scattering rates. After Ref. 17, reprinted with permission.

results of our simulations support the reason cited by Nguyen et al., (24) namely, decreased electron velocity caused by an increased rate of scattering of carriers by polar optical phonons and ionized impurities.

MODELING QUANTUM-CONFINED TRANSPORT IN MULTIPLE QUANTUM WELLS

As the semiconductor industry progresses toward nanoelectronics, more and more electron devices utilize single or multiple quantum wells in their design. As was discussed in Section 8, in a quantum well, the carriers behave as two-dimensional entities, since their transport is restricted to only two dimensions. In devices where multiple GaAs quantum (MQW) wells (5 to 30 wells each with a width of the order of 10 to 100 Å) are used in bulk AlGaAs, the transport of carriers will be influenced by two additional processes: *capture* of carriers by the wells, and *escape* of carriers from the wells.

In the AlGaAs bulk regions of the device, where the electrons and holes behave as bulk carriers with no quantum confinement, the current continuity and current density equations are the traditional drift-diffusion equations written for electrons and holes as:

$$\frac{dn_b}{dt} = G_b - U_b + \frac{n_{qw}}{\tau_{en}} - \frac{n_b}{\tau_{cn}} + \frac{1}{q}\frac{dJ_n}{dx} = 0 \tag{95}$$

$$\frac{dp_b}{dt} = G_b - U_b + \frac{p_{qw}}{\tau_{ep}} - \frac{p_b}{\tau_{cp}} - \frac{1}{q}\frac{dJ_p}{dx} = 0 \tag{96}$$

$$J_n = q\mu_n n_b \vec{E} + qD_n \frac{dn_b}{dx} \tag{97}$$

$$J_p = q\mu_p p_b \vec{E} - qD_p \frac{dp_b}{dx} \tag{98}$$

In these equations, $\vec{E}$ is the electric field, n_b and p_b are the bulk electron and hole densities, respectively, n_{qw} and p_{qw} are the quantum well electron and hole densities, respectively, and J_n and J_p are the electron and hole current densities, respectively. The terms $\frac{n_{qw}}{\tau_{en}}$ and $\frac{p_{qw}}{\tau_{ep}}$ are the rates by which the electrons and holes escape the quantum well (and enter the bulk), whereas $\frac{n_b}{\tau_{cn}}$ and $\frac{p_b}{\tau_{cp}}$ are the rates by which electrons and holes are captured by the quantum wells from the bulk. The *escape* and *capture* rates of the bulk are (assumed to be) equal to the *capture* and *escape* rates of the quantum wells. In other words, the carriers that escape the bulk are captured by the quantum wells, and those that escape the quantum wells are captured by the bulk. The symbols τ_{en} and τ_{ep} stand for the electron and hole escape times, respectively, and τ_{cn} and τ_{cp} are the electron and hole capture times, respectively. The terms U_b and G_b are the rates of recombination and generation in the bulk, respectively. The other terms in the current density equations have their usual meanings. The above equations must be solved together with Poisson equation:

$$\frac{\partial^2 V}{\partial x^2} = \frac{q}{\epsilon}[N_A^- - N_D^+ + n_b + n_{qw} - p_b - p_{qw}] \tag{99}$$

The values of n_{qw} and p_{qw} are set to zero for the bulk regions. N_D^+ and N_A^- are the ionized donor and acceptor doping levels, V is the potential, and ϵ is the dielectric constant.

The eigenfunctions and eigenenergies of carriers in the quantum wells must be obtained from a field-dependent Schrödinger equation given by:

$$\left[\frac{\hbar^2}{2}\frac{d}{dx}\frac{1}{m^*(x)}\frac{d}{dx} + V(x)\right]\psi_i(x) = E_i\psi_i(x) \tag{100}$$

where E_i and $\psi_i(x)$ are the energy level and the wavefunction of the subband i, respectively. A non-constant effective mass, $m^*(x)$, must be used to account for different

material systems throughout the balk and the quantum wells.

The current continuity equations for the quantum well system are written as the rate equations in which the net balance of four rates must vanish. These rate equations are given by:

$$\frac{dn_{qw}}{dt} = \frac{n_b}{\tau_{cn}} - \frac{n_{qw}}{\tau_{en}} + G_{qw} - U_{qw} = 0 \tag{101}$$

$$\frac{dp_{qw}}{dt} = \frac{p_b}{\tau_{cp}} - \frac{p_{qw}}{\tau_{ep}} + G_{qw} - U_{qw} = 0 \tag{102}$$

The terms U_{qw} and G_{qw} are the modified Shockley-Read-Hall (SRH) recombination rate, and the generation rate of the quantum wells, respectively. The absence of current density terms in the above equations stems from the fact that carriers in the quantum wells are assumed to be bound, and therefore, to contribute to the current in the device, these carriers must first escape from the quantum wells. For application of the MQW model in solar cells as well as detailed modeling of escape and capture processes, see Ref. (30) and (31).

BIBLIOGRAPHY

1. M. Lundstrom, *Fundamentals of Carrier Transport*, Vol. X, Reading, MA: Addison-Wesley, 1990.
2. D. K. Ferry, *Semiconductors*, New York: Macmillan, 1991.
3. K. Hess, *Advanced Theory of Semiconductor Devices*, Englewood Cliffs, NJ: Prentice-Hall, 1988.
4. Z. H. Ng, R. Khoie, R. Venkat, A two-dimensional self-consistent numerical model for high electron mobility transistor, *IEEE Trans. Electron Devices*, **38**: 852–861, 1991.
5. R. Khoie, A self-consistent numerical method for simulation of quantum transport in high electron mobility transistor. Part I: The Boltzmann-Poisson-Schrödinger solver, *Math. Probl. Eng.*, **2**: 205–218, 1996.
6. R. Khoie, A self-consistent numerical method for simulation of quantum transport in high electron mobility transistor. Part II: The full quantum transport, *Math. Probl. Eng.*, **2**: 219–231, 1996.
7. H. Arman, R. Khoie, A self-consistent multisubband model for calculation of scattering rates in quantum well structures, *Natl. Center Computational Electron., Proc. Int. Workshop Computational Electron.*, K. Hess, U. Ravaioli, and R. Dutton (eds.), Beckman Inst., Univ. of Illinois at Urbana-Champaign, 1992, pp. 175–179.
8. R. Khoie, H. Arman, A two-subband self-consistent model for high electron mobility transistor including intersubband and intrasubband scattering mechanisms, *Natl. Center Computational Electron., Proc. Int. Workshop Computational Electron.*, K. Hess, U. Ravaioli, andR. Dutton (eds.), Beckman Inst., Univ. of Illinois at Urbana-Champaign, 1992, pp. 181–184.
9. Z. H. Ng, R. Khoie, R. Venkat, A self-consistent calculation of spatial spreading of the quantum well in HEMT, inK. Hess, J. P. Leburton, and U. Ravaioli (eds.), *Computational Electronics, Semiconductor Transport and Device Simulation*, Boston: Kluwer, 1991, pp. 55–58.
10. R. Khoie, A study of transconductance degradation in HEMT using a self-consistent Boltzmann-Poisson-Schrödinger solver, *VLSI Design*, **6**: 73–77, 1997.
11. D. H. Park, K. F. Brennan, Monte Carlo simulation of 0.35- mM gate-length GaAs and InGaAs HEMTs, *IEEE Trans. Electron Devices*, **37**: 618–628, 1990.
12. C. Jacoboni, L. Reggiani, The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials, *Rev. Modern Phys.*, **55**: 645–705, 1983.
13. L. Reggiani, *Topics in Applied Physics*, Vol. 58, New York: Springer Verlag, 1985.
14. W. Fawcett, A. D. Boardman, S. Swain, Monte Carlo determination of electron transport properties in gallium arsenide, *J. Phys. Chem. Solids*, **31**: 1963–1990, 1970.
15. P. J. Price, Monte Carlo calculation of electron transport in solids, *Semicond. Semimetals*, **14**: 249–334, 1979.
16. R. W. Hockney, E. J. Eastwood, *Computer Simulation Using Particles*, New York: McGraw-Hill, 1981.
17. K. Tomizawa, N. Hashizume, Ensembe Monte Carlo simulation of AlGaAs/GaAs heterostructure MIS-like FET, *IEEE Trans. Electron Devices*, **35**: 849–856, 1988.
18. K. L. Priddy *et al.*, Design of enhanced Schjottky-barrier AlGaAs/GaAs MODFET's using highly doped p^+ surface layers, *IEEE Trans. Electron Devices*, **ED-34**: 175–179, 1987.
19. D. J. Widiger *et al.*, Two-dimensional transient simulation of an idealized high electron mobility transistor, *IEEE Trans. Electron Devices*, **ED-32**: 1092–1102, 1985.
20. S. H. Gould, *Variational Methods for Eigenvalue Problems*, 2nd ed., Toronto, Canada: Univ. of Toronto Press, 1957.
21. K. Yokoyama, K. Hess, Monte Carlo study of the electronic transport in $Al_{1-x}Ga_xAs$/GaAs single-well heterostructure, *Phys. Rev. B*, **33**: 5595–5606, 1986.
22. C. M. Snowden, *Semiconductor Device Modelling*, IEEE Materials and Device Series 5, 1988.
23. D. Greenberg, J. Del Alamo, R. Bhat, Impact ionization and transport in InAlAs/n^+-InP HFET, *IEEE Trans. Electron Devices*, **42**: 1574–1582, 1995.
24. L. Nguyen *et al.*, 50-nm self-aligned-gate pseudomorphic AlInAs/GaInAs high electron mobility transistors, *IEEE Trans. Electron Devices*, **39**: 2007–2014, 1992.
25. S. Sargood *et al.*, A quantum well inversion channel heterostructure as a multifunctional component for optoelectronic integrated circuits, *IEEE J. Quantum Electron.*, **29**: 136–149, 1993.
26. G. Wang, Y. Chen, L. Eastman, A 0.1 mM gate $Al_{0.5}In_{0.5}As/Ga_{0.5}In_{0.5}As$ MODFET fabricated on GaAs substrate, *IEEE Trans. Electron Devices*, **35**: 818–823, 1988.
27. I. Kizilyalli *et al.*, Scaling properties and short-channel effects in submicrometer AlGaAs/GaAs MODFET's: A Monte Carlo study, *IEEE Trans. Electron Devices*, **40**: 234–249, 1993.
28. Y. Wang *et al.*, Self-consistent simulation of modulation doped field effect transistors, *Solid State Electron.*, **37**: 237–241, 1994.
29. W. Walukiewicz *et al.*, Electron mobility and free carriers absorption in GaAs: Determination of the compensation ratio, *J. Appl. Phys.*, **50**: 899–908, 1979.

30. R. Khoie and S. Ramey, "Self-Consistent Modeling of Escape and Capture of Carriers in Quantum Wells," *Physica E: Low Dimensional Systems and Nanostructures*, **34**, 449–451, 2006.

31. S. Ramey and R. Khoie, "Modeling of Multiple Quantum Well Solar Cells Including Capture, Escape, and Recombination of Photo-excited Carriers in Quantum Wells," *IEEE Transactions on Electron Devices*, **50**, No. 5, 1179–1188, 2003.

R. KHOIE
University of the Pacific,
Stockton, CA
R. VENKAT
University of Nevada, Las
Vegas, Las Vegas, NV