Semiclassical description of electron transport in semiconductor quantum-well devices

G. A. Baraff
Lucent Technologies, Bell Laboratories, Murray Hill, New Jersey 07974
(Received 6 December 1996)

Carrier drift, diffusion, and thermionic emission for classical semiconductor devices (p-n junctions, heterostructures, etc.) is most easily described using expressions derived from a Boltzmann transport equation point of view. This point of view is not particularly applicable to quantum-well transport. It is shown here that by postulating a region of phase space that is forbidden to the mobile carriers and then altering the scattering probability so that no particles are scattered to the forbidden region, a Boltzmann-equation-based formalism emerges that can describe the mobile-carrier component of quantum-well transport. [S0163-1829(97)01916-4]

I. INTRODUCTION

In trying to model the dynamics of single- or multiple-quantum-well semiconductor devices, it is usual to regard the carriers as belonging to (at least) two distinct populations, namely, those carriers whose motion is confined to the vicinity of a single quantum well and those that are able to travel more or less freely from one well to the next.[1–3] The practical question that then arises is how to calculate the current and the spatial distribution of carriers in each of the populations. This paper is concerned with answering that question for the mobile carriers.

The problem posed here is one of quantum transport. This paper will not deal with effects that arise because of short-time scales and energy nonconservation or short length scales and wave-packet localization. Still, because the simplicity of a Boltzmann approach makes it a practical one for many aspects of device behavior, it is useful to explore how some simple concepts from quantum mechanics may be incorporated into the Boltzmann equation in a way that makes it applicable to the transport of mobile carriers in quantum-well devices.

In contrast to the situation for mobile carriers, the description of how the *bound* carriers are distributed is conceptually simple: Neglecting tunneling, a self-consistent potential confines the bound carriers to a single well. Assuming that this potential depends only on z and that some single-band envelope-function description is adequate, the wave functions for the bound carriers will have the form

$$\psi = \frac{1}{\sqrt{A}} e^{ik_{\|}\rho}\chi_i(z), \tag{1.1a}$$

where $\rho \equiv (x,y)$ and A is the normalization area. This wave function and its energy

$$E = E_i + \hbar^2 k_{\|}^2/2m^* \tag{1.1b}$$

are determined by the solution to a Schrödinger equation. For bound states, $\chi_i(z)$ tends to zero as z leaves the vicinity of the well and the energies E_i are discrete. The particle density is then evaluated by calculating or assuming some occupation probability and summing it over the densities associated with states of the form (1.1). Such a description is based on the point of view that the bound particles do not interact with each other or, equivalently, that any energy of their mutual interaction is small compared to the separation ΔE between adjacent energies E_i.

This simple picture is harder to justify (and is likely incorrect) when applied to the mobile population of carriers. For these, the functions $\chi_i(z)$ remain finite even for z far from the well and the energies E_i are more closely spaced, becoming continuous in the limit of a large system. As the carrier density rises, one expects the energy uncertainty associated with carrier-carrier scattering also to rise, eventually becoming greater than the energy spacing between states. This will result in a reduced phase coherence length. When the phase coherence length becomes smaller than the well width, the whole picture based on the literal use of Eq. (1.1) becomes suspect: Literal use of Eq. (1.1) predicts strong resonances in the amplitude of the mobile-carrier wave functions in the neighborhood of the well when the shape and depth of the well are such that a quasi-bound-state exists at low energies.[4–7] These resonances, in turn, lead to strong variation of the capture time because of enhanced overlap between the mobile states and the bound state into which the carrier is captured. Experimentally, these oscillations have been hard to observe.[8,9] Experiments cited as supporting their existence[10–13] have been carried out only in a low-density regime or at low temperature or both, conditions that act to maximize the phase coherence needed to observe these resonant phenomena. This challenges the wisdom of basing a theory of high-density room-temperature quantum-well transport on Eq. (1.1) as it stands.

A completely different point of view, appropriate to electron transport dominated by scattering, is that quantum effects are ignored entirely and that the flow of particles is described by a Boltzmann transport equation. This leads to a drift-diffusion description of carrier transport[14] in which particle fluxes arise in response to electric fields (drift current), to concentration gradients (diffusion current), and to concentration discontinuities (thermionic emission current).[15] This point of view does not lend itself readily to describing the specific features of the quantum well and so some synthesis of the classical and quantum descriptions is necessary at this point.

This problem is not new, and progress has been made by Grupen and co-workers.[16] There is a sharp contrast between

their approach and ours. Their approach treats the states of the mobile carriers in the well region the same way it treats the bound states, except for a flow of carriers across the edges of the well into and out of the mobile states. They postulate a separate (position-independent) Fermi energy for both the mobile and bound populations. Assuming this for the *mobile* states is equivalent to taking the mobile-state wave functions, like those of the bound states, to be fully coherent across the well. It is this assumption that we wish to avoid: it is hard to accept that, in the presence of the large capture rate that must occur in lasers under operating conditions, full coherence of the mobile-state wave functions can be maintained. In this paper, we explore the opposite limit, namely, where there is no coherence at all in the mobile states. Clearly, a better theory is needed to describe the real situation where the coherence is likely to be partial and the results intermediate to those that would be obtained using their approach and ours.

To develop a theory of transport of the mobile carriers with no coherence, we return solidly to the Boltzmann-equation point of view. However, certain modifications will be made to account for the possible presence of bound carriers. These modifications can be described by the following postulates about the phase space in which that equation operates.

(i) There is a well-defined region of (classical) phase space that is forbidden to the mobile carriers.

(ii) The probability for mobile carriers to scatter to the forbidden region of phase space is zero.

(iii) The region of phase space allowed to the mobile carriers falls into two disjoint parts, one for the right-traveling carriers, the other for left-traveling carriers.

(iv) The probability to scatter from one allowed region of phase space to the other is far less than the probability to scatter within the same allowed region.

These postulates will be explained, and to some extent rationalized, in Sec. II. In Sec. III the transport equation that results is formulated. In Sec. IV the boundary conditions are presented and in Sec. V the equations are generalized to include carrier capture and emission.

II. FORCING THE QUANTUM MECHANICS OF THE WELL INTO THE BOLTZMANN EQUATION

In order to use the Boltzmann equation for quantum wells, a fundamental restriction must be imposed on the scattering probability in that equation. To motivate the need for that change, consider E_z, the z component of the energy. Classically, E_z is given by

$$E_z = H_z(z,p_z) \equiv p_z^2/2m + V(z). \tag{2.1a}$$

Quantum mechanically, E_z is given by

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + V(z) - E_z \right) \psi(z) = 0. \tag{2.1b}$$

Consider a typical quantum well of depth V_0 and width w such as the one whose potential is sketched in Fig. 1: The potential is taken to be

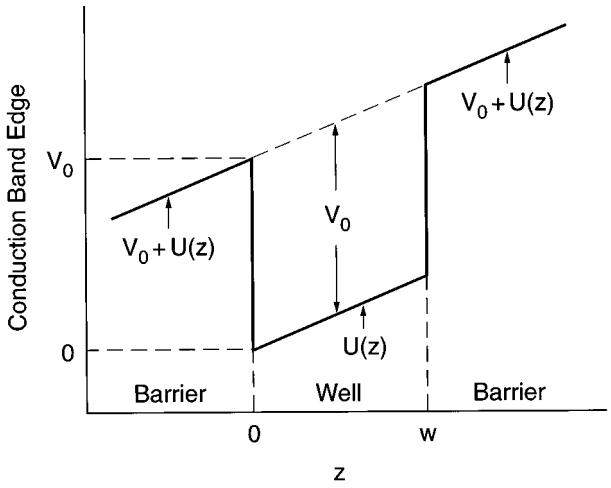

FIG. 1. Potential used for a quantum well of depth V_0 located in a region having a nonzero field.

$$V(z) = \begin{cases} U(z) + V_0, & z < 0, \quad z > w \qquad (2.2a) \\ U(z), & 0 < z < w. \qquad (2.2b) \end{cases}$$

It is convenient to choose the zero of potential at the lowest point in the well and to choose the sense of the z axis so that $U(w) \geq U(0)$. Neglecting the possibility of tunneling through the barrier on the left of the well [say by setting $U(z) = 0$ for $z < 0$] the energy spectrum of Eq. (2.1b) divides into two regions. For $E_z < V_0$, the spectrum is discrete, while for $E_z > V_0$, the spectrum is continuous. The discrete spectrum describes bound states [the two-dimensional (2D) population]. The continuous spectrum describes mobile carriers (the 3D population). We assume that the whole region that might support bound states is unavailable to the mobile carriers. This restriction, expressed in terms of Eq. (2.1a), is $E_z > V_0$, or

$$p_z^2 > 2m[V_0 - U(z)] \equiv p_{\min}^2(z), \quad 0 < z < w. \tag{2.3}$$

This divides the allowed phase space in the well into two disjoint regions, one for right-traveling particles and one for left-traveling particles.

It is easy to see that under the sole influence of the classical Hamiltonian

$$H = \frac{p_x^2 + p_y^2}{2m} + H_z(z,p_z) \tag{2.4}$$

a particle in an allowed region of phase space will never enter the excluded region. Collisions, however, are a different matter: A particle whose collision-free trajectory is in the allowed region might, after a *classical* scattering event, be found in the excluded region. If the region is excluded, scattering to it must be prohibited. Although this prohibition may seem *ad hoc*, a quantum-mechanical calculation of scattering matrix using only final states whose momentum components are allowed in the well would give rise to a corresponding restriction.

The effect of this restriction may be seen as follows. Consider the scattering probability W used in the Boltzmann equation:

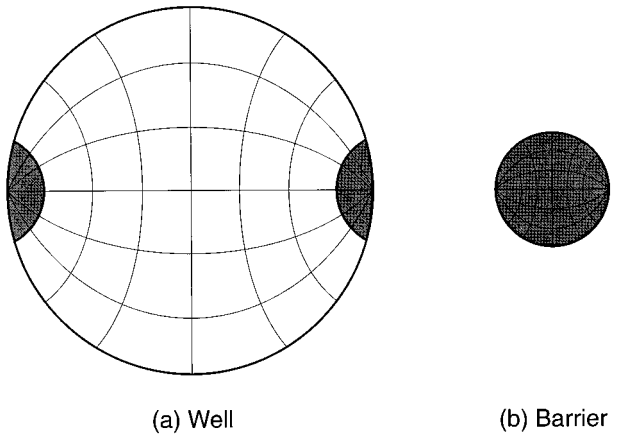

FIG. 2. Surfaces of constant energy in momentum space available to moble carriers of total energy $E=V_0+k_BT$ (a) in the quantum well and (b) in the adjacent barriers.

$$W=W_z(E,\Omega,\rightarrow E',\Omega'). \tag{2.5}$$

The pre- and post-collision states of the particle have been labeled by a total energy given by Eq. (2.4) and an angle Ω specifying the direction of the momentum. The z momentum of a particle in the well must satisfy Eq. (2.3). Its total momentum is

$$p^2=2m[E-U(z)]. \tag{2.6}$$

After collision, the direction cosine μ will therefore be restricted by

$$\mu^2\equiv(p_z/p)^2>[V_0-U(z)]/[E-U(z)]\equiv\mu_{\min}^2. \tag{2.7}$$

Surfaces of constant energy E in momentum space to which carriers may scatter are sketched in Fig. 2. They are small polar caps at opposite ends of a sphere whose radius depends on the energy E and on z. As E decreases towards V_0, the solid angle $\Delta\Omega$ of each cap shrinks towards zero. $\Delta\Omega$ is given by

$$\Delta\Omega=2\pi(1-\mu_{\min})=2\pi\{1-\sqrt{[V_0-U(z)]/[E-U(z)]}\}. \tag{2.8}$$

These are also the surfaces of constant energy that the mobile carriers in the well can occupy. The small area of the polar caps means that, within the well, carriers are strongly collimated to travel along the z axis. However, outside the well, allowed surfaces of constant energy are small complete spheres and particles travel more or less uniformly in all directions.

Within the well, the population of each polar cap will be considered to be a distinct species. The reason for doing so is that for a carrier to scatter from one polar cap to the other, a momentum of order $2\sqrt{2mV_0}$ must be supplied. There are few scattering mechanisms that can supply this much momentum. Significant scattering between one cap and the other is far less likely than scattering within a single cap.

In full equilibrium, carriers in the well occupy a Fermi-Dirac distribution. In partial equilibrium, the two populations can have different Fermi energies because of inhibited scattering from one population to the other. Even farther from equilibrium, the two Fermi energies can depend on position, becoming ''quasi-Fermi energies,'' such as those used for describing the populations of electrons and holes in bulk semiconductors. The two quasiequilibrium distribution functions are

$$f_0^+=\theta(p_z)\,\theta[H_z-V_0]\,\frac{2}{(2\pi\hbar)^3}\times\frac{1}{1+\exp\{[H(z,p)-E_F^+(z)]/k_BT\}}, \tag{2.9a}$$

$$f_0^-=\theta(-p_z)\,\theta[H_z-V_0]\,\frac{2}{(2\pi\hbar)^3}\times\frac{1}{1+\exp\{[H(z,p)-E_F^-(z)]/k_BT\}}, \tag{2.9b}$$

where θ is the unit step function. Each of these gives rise to a large flux of particles

$$J_0^+(z)=\frac{2}{(2\pi\hbar)^3}\int dp_xdp_y\times\int_{p_{\min}(z)}^{\infty}\frac{V_zdp_z}{1+\exp[H(z,p)-E_F^+(z)]/k_BT}, \tag{2.10a}$$

$$J_0^-(z)=\frac{2}{(2\pi\hbar)^3}\int dp_xdp_y\times\int_{-\infty}^{-p_{\min}(z)}\frac{V_zdp_z}{1+\exp[H(z,p)-E_F^-(z)]/k_BT}. \tag{2.10b}$$

In true equilibrium, the two quasi-Fermi energies are equal and independent of z. The two fluxes are then equal and opposite and there is no net current in the system of carriers.

In spite of the appearance of $U(z)$ in both the integrand and the limits of Eqs. (2.10), the flux depends on z only if $E_F^{\pm}$ is z dependent. This can be seen by using E_z as the variable of integration instead of p_z. The integrals are then

$$J_0^{\pm}(z)=\pm\frac{2}{(2\pi\hbar)^3}\int dp_xdp_y\times\int_{V_0}^{\infty}\frac{dE_z}{1+\exp[(p_x^2+p_y^2)/2m+E_z-E_F^{\pm}(z)]/k_BT}. \tag{2.11}$$

In contrast, the densities $n^{\pm}(z)$ depend explicitly on $U(z)$,

$$n_0^{\pm}(z)=\frac{2}{(2\pi\hbar)^3}\int dp_xdp_y\times\int_{p_{\min}(z)}^{\infty}\frac{dp_z}{1+\exp[H(z,p)-E_F^{\pm}(z)]/k_BT}. \tag{2.12}$$

When the densities are low enough for the Fermi-Dirac distribution to be replaced by the Boltzmann distribution, the current and density become

$$J_0^{\pm}(z) = \pm \frac{m k_B^2 T^2}{2\pi^2 \hbar^3} e^{[E_F^{\pm}(z) - V_0]/k_B T}, \tag{2.13}$$

$$n_0^{\pm} = \left(\frac{m k_B T}{2\pi\hbar^2} \right)^{3/2} \mathrm{erfc}[p_{\min}(z)^2 / 2 m k_B T]^{1/2} e^{[E_F^{\pm}(z) - U(z)]/k_B T}, \tag{2.14a}$$

where

$$\mathrm{erfc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt. \tag{2.14b}$$

The first factor in Eq. (2.13) is the usual Richardson factor for thermionic emission. The first factor in Eq. (2.14a) is half the usual thermal density of states of a bulk semiconductor. If no phase space is excluded [so that $p_{\min}(z)=0$ and $\mathrm{erfc}(x)=1$], then n_0^+ and n_0^- each are half the density of carriers in a band whose edge is at $U(z)$ and whose Fermi energy is $E_F^{\pm}$. The exclusion makes a profound difference, however. To show this, we work to first order in $k_B T/V_0$ (which is always small for quantum wells of practical interest) and retain the first two terms in the large argument expansion of the complementary error function[17]

$$\mathrm{erfc}(x) \approx \frac{1}{x\sqrt{\pi}} e^{-x^2} \left(1 - \frac{1}{2x^2} \right).$$

Using the definition of $p_{\min}(z)$, this gives

$$n_0^{\pm} = \frac{m k_B^2 T^2}{2\pi^2 \hbar^3} \sqrt{\frac{m}{2[V_0 + k_B T - U(z)]}}\, e^{[E_F^{\pm}(z) - V_0]/k_B T}. \tag{2.15}$$

The average z velocity is

$$\langle V_z \rangle \equiv \frac{J_0^{\pm}(z)}{n_0^{\pm}(z)} = \sqrt{\frac{2[V_0 + k_B T - U(z)]}{m}}, \tag{2.16}$$

which is the velocity of a particle with $E_z = V_0 + k_B T$. The proportionality of $n_0^{\pm}(z)$ to the inverse velocity is expected classically for a beam of particles whose speed varies. Equation (2.13) is exactly the usual form for thermionic emission current.[15,18] Equations (2.13), (2.15), and (2.16) demonstrate that the assumption of a quasiequilibrium distribution function in a restricted phase space gives a reasonable description of particles of low-z kinetic energy ($\approx k_B T$) accelerated by a strong potential drop V_0 at the edge of the well and then acted upon by a potential $U(z)$ in the well itself.

III. BOLTZMANN EQUATION

A fundamental aspect of transport of mobile carriers (the 3D population) across the wells is that carriers injected at $z=0$, traveling towards the right, have come out of a reservoir whose Fermi energy may differ from that of the other reservoir, which, at $z=w$, injects the carriers traveling towards the left. Equilibration of these two populations of traveling carriers will occur only via events that can transfer carriers from one population to the other. If the mechanisms for doing this are so weak that equilibration does not occur during the traversal time of the well, some measure of the Fermi-level difference between the two reservoirs will persist into the well.

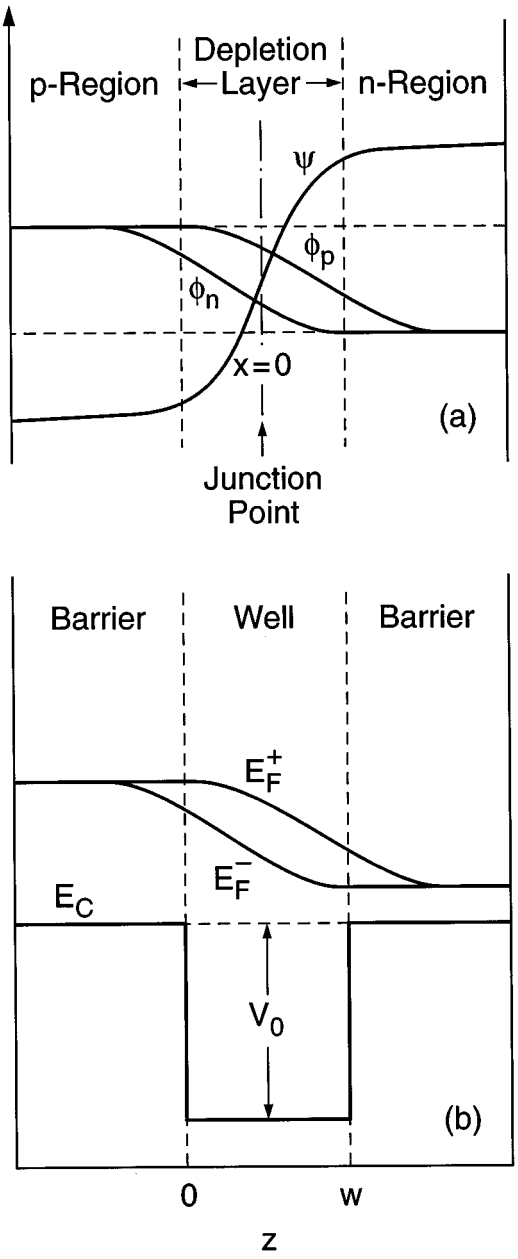

FIG. 3. Quasi-Fermi energies versus position in (a) a semiconductor p-n junction, where the two quasi-Fermi energies describe electrons and holes, and (b) a quantum well where the two quasi-Fermi energies describe right-going and left-going electrons.

The situation is analogous to that of electrons and holes in a p-n semiconductor junction. In a p-n junction under bias, the two types of carriers may have different quasi-Fermi energies until carrier generation and recombination equilibrate the two populations. A sketch of the quasi-Fermi energies for electrons and holes in a p-n junction under bias and a sketch of the quasi-Fermi energies for the left-going and right-going electrons in a quantum well under bias have the same form, as illustrated in Fig. 3. In the junction itself, there is insuffi-

cient time for the electron and hole populations to equilibrate and the two quasi-Fermi energies separate there. In the regions adjacent to the junction, generation and recombination equilibrate the two populations and the two quasi-Fermi energies coalesce. In the quantum well, the particles in the well have been accelerated to high speed and scattering is ineffective in equilibrating the two populations. In the regions adjacent to the wells, the particles are traveling more slowly, and scattering from the forward to the back direction can take place, equilibrating the two populations.

To construct a Boltzmann formalism for this situation, two particle distribution functions f^+ and f^- are used to describe the carrier distributions in the two separate allowed regions of phase space. Each of these distributions separately satisfies a Boltzmann equation

$$\frac{\partial f^{\pm}}{\partial t}+\left(\frac{\partial H}{\partial p_z}\frac{\partial}{\partial z}-\frac{\partial H}{\partial z}\frac{\partial}{\partial p_z}\right)f^{\pm}=\left(\frac{\partial f^{\pm}}{\partial t}\right)_{\text{coll}}. \tag{3.1}$$

The collisional derivative is the sum of a scattering out term and a scattering in term

$$\left(\frac{\partial f^{\pm}}{\partial t}\right)_{\text{coll}}=\left(\frac{\partial f^{\pm}}{\partial t}\right)_{\text{in}}-\left(\frac{\partial f^{\pm}}{\partial t}\right)_{\text{out}}. \tag{3.2}$$

We assume that the scattering probability $W_z(\mathbf{p}\rightarrow\mathbf{p}')$ to be used in the Boltzmann equation is of the relaxational form, which means that after scattering, the distribution of particles is a quasiequilibrium one at the local density. Let $p_{\pm}$ and $p'_{\pm}$ be the notation used to indicate in which of the two allowed regions of phase space $\mathbf{p}$ and $\mathbf{p}'$ are located. On the assumption that scattering from one region to the other is much less frequent than scattering within the same region, the scattering probability to be used in the Boltzmann equation is

$$W_z(p_+\rightarrow p'_+)=\frac{1}{\tau^+}\frac{f_0^+(p')}{n^+(z)}, \tag{3.3a}$$

$$W_z(p_+\rightarrow p'_-)=\frac{1}{\tau^-}\frac{f_0^-(p')}{n^-(z)}, \tag{3.3b}$$

$$W_z(p_-\rightarrow p'_+)=\frac{1}{\tau^-}\frac{f_0^+(p')}{n^+(z)}, \tag{3.3c}$$

$$W_z(p_-\rightarrow p'_-)=\frac{1}{\tau^+}\frac{f_0^-(p')}{n^-(z)}, \tag{3.3d}$$

where τ^+ and τ^- are the scattering times for forward (intraregional) and backward (interregional) scattering, respectively. These two are vastly different, with $\tau^+/\tau^-\ll 1$.

The scattering terms in Eq. (3.2) can be evaluated using the probabilities $W_z(p\rightarrow p')$,

$$\left(\frac{\partial f^{\pm}}{\partial t}\right)_{\text{out}}=\int dp' f^{\pm}(p)W_z(p\rightarrow p')=\left(\frac{1}{\tau^+}+\frac{1}{\tau^-}\right)f^{\pm}(p), \tag{3.4a}$$

$$\left(\frac{\partial f^{\alpha}}{\partial t}\right)_{\text{in}}=\sum_{\beta}\int dp' f^{\beta}(p')W_z(p'\rightarrow p), \tag{3.4b}$$

$$\left(\frac{\partial f^{+}}{\partial t}\right)_{\text{in}}=\left(\frac{n^+}{\tau^+}+\frac{n^-}{\tau^-}\right)\frac{f_0^+(p)}{n^+}, \tag{3.4c}$$

$$\left(\frac{\partial f^{-}}{\partial t}\right)_{\text{in}}=\left(\frac{n^-}{\tau^+}+\frac{n^+}{\tau^-}\right)\frac{f_0^-(p)}{n^-}. \tag{3.4d}$$

Integrating the transport equation (3.1) over all momenta in the appropriate allowed region of phase space and using the scattering expressions (3.4) gives

$$\frac{\partial n^+}{\partial t}+\frac{\partial}{\partial z}J^+=\frac{n^--n^+}{\tau^-}, \tag{3.5a}$$

$$\frac{\partial n^-}{\partial t}+\frac{\partial}{\partial z}J^-=\frac{n^+-n^-}{\tau^-}, \tag{3.5b}$$

where the fluxes J^+ and J^- are given by

$$J^+(z,t)\equiv\int_{p_z>p_{\min}(z)}d^3p\ V_z f^+(z,p,t), \tag{3.6a}$$

$$J^-(z,t)\equiv\int_{p_z<-p_{\min}(z)}d^3p\ V_z f^-(z,p,t). \tag{3.6b}$$

The technique for evaluating the fluxes $J^{\pm}$ is similar to that used to evaluate the fluxes in bulk semiconductors.[14,19] The transport equation for $f^{\pm}$ is solved neglecting both the time dependence and scattering into or out of the distribution because these possibilities are already included in the continuity equations (3.5). The transport equation (3.1) then becomes

$$\left(V_z\frac{\partial}{\partial z}-\frac{\partial U}{\partial z}\frac{\partial}{\partial p_z}+\frac{1}{\tau^+}\right)f^{\pm}=\frac{f_0^{\pm}}{\tau^+}. \tag{3.7}$$

This equation is linearized by setting $f^{\pm}=f_0^{\pm}+\delta f^{\pm}$ and assuming that the scattering term is larger than the convective flow term. This gives

$$\delta f^{\pm}=-\tau^+V_z\frac{\partial E_F^{\pm}}{\partial z}\frac{\partial f_0^{\pm}}{\partial E_F^{\pm}}. \tag{3.8}$$

The change in the distribution function must produce no change in the particle density. This happens automatically for bulk semiconductors where the integration of Eq. (3.8) over all momentum space causes the integral to vanish by symmetry. It does not happen here because the integrand for each of the functions $\delta f^{\pm}$ extends only over a part of momentum space in which the integrand does not change sign. The integral of $\delta f^{\pm}$ can vanish only if $E_F^{\pm}$ is independent of z. In such a case, $\delta f^{\pm}$ vanishes identically and $J^{\pm}$ and $n^{\pm}$ are the explicit functions of $E_F^{\pm}$ given in Eqs. (2.11) and (2.12). We can therefore rewrite Eqs. (3.5) in the form

$$\frac{\partial n^+}{\partial t}+\frac{\partial E_F^+}{\partial z}\frac{\partial J^+}{\partial E_F^+}=\frac{n^--n^+}{\tau^-}, \tag{3.9a}$$

$$\frac{\partial n^-}{\partial t}+\frac{\partial E_F^-}{\partial z}\frac{\partial J^-}{\partial E_F^-}=\frac{n^+-n^-}{\tau^-}, \tag{3.9b}$$

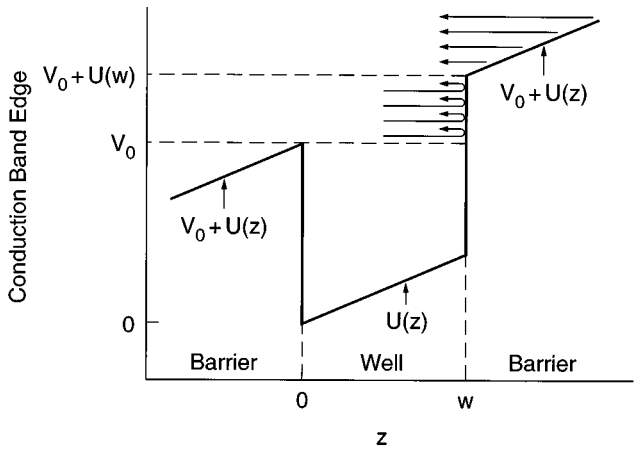

FIG. 4. Division of the right-going and left-going carriers into two groups, according to the boundary conditions they must satisfy at the edges of the well.

in which $E_F^{\pm}(z,t)$ are the unknowns. To make the solution unique, the values $E_F^+(0)$ and $E_F^-(w)$ must be specified. Thermionic emission theory provides the way to specify these boundary conditions.

IV. BOUNDARY CONDITIONS: THERMIONIC EMISSION

Thermionic emission theory is based on the idea that the carriers *leaving* the half spaces on both sides of a potential step are in equilibrium distributions.[15,18] We assume that the carriers leaving the barrier region and entering the well at $z=0$ are in an equilibrium distribution having a Fermi energy $E_F(0)$. Using the Boltzmann equation (3.1), it can be shown that in the well, the incoming distribution at the left-hand side has the same Fermi energy as the outgoing carriers in the barrier. This gives the boundary condition

$$E_F^+(0)=E_F(0). \tag{4.1}$$

The situation at the other side of the well is somewhat more complex. We assume that the carriers leaving the barrier region have a Fermi energy $E_F(w)$. In Fig. 4 it is clear that the incoming (left-going) distribution at $z=w$ consists of two separate parts. For E_z lying between V_0 and $V_0+U(w)$, the incoming distribution contains particles that had been traveling towards the right and have since been reflected by the edge of the well. Since at $z=w$ there can be no current in this energy range, the Fermi energy of this part of the left-going distribution must equal the Fermi energy of the right-going particles, namely, $E_F^+(w)$. For $E_z>V_0+U(w)$, the incoming distribution is injected from the barrier. This part of the distribution has the Fermi energy of the particles in the barrier, namely, $E_F(w)$. As the two parts of the left-going distribution move into the well, rapid forward scattering blends them into a single equilibrium distribution with a single Fermi energy. In the limit that τ^+ is very small, this blending will be complete very close to $z=w$ and so $E_F^-(w)$ is determined by the requirement that the density of particles in f_0^- is equal to the sum of the densities in the two parts of the left-going distributions just described.

The three left-going distributions, corresponding to reflected, injected, and equilibrated particles as described above at $z=w$, are

$$f_{\text{refl}}^-=\theta(-p_z)\,\theta[H_z(w,p)-V_0]\,\theta[V_0+U(w)-H_z(w,p)]\times\frac{2}{(2\pi\hbar)^3}\,\frac{1}{1+\exp\{[H(w,p)-E_F^+(w)]/k_BT\}}, \tag{4.2a}$$

$$f_{\text{inj}}^-=\theta(-p_z)\,\theta[H_z(w,p)-V_0-U(w)]\times\frac{2}{(2\pi\hbar)^3}\,\frac{1}{1+\exp\{[H(w,p)-E_F(w)]/k_BT\}}, \tag{4.2b}$$

$$f_0^-=\theta(-p_z)\,\theta[H_z(w,p)-V_0]\times\frac{2}{(2\pi\hbar)^3}\,\frac{1}{1+\exp\{[H(w,p)-E_F^-(w)]/k_BT\}}. \tag{4.2c}$$

Equating the equilibrated density to the two partial densities,

$$\int d^3p\; f_{\text{refl}}^- + \int d^3p\; f_{\text{inj}}^- = \int d^3p\; f_0^-, \tag{4.3}$$

which determines the Fermi energy $E_F^-(w)$.

Equations (4.1)–(4.3), plus perhaps an initial condition, are sufficient boundary conditions to use with the transport equations (3.9) when the Fermi energies in the barriers are known. However, the Fermi energies of the particles in the barriers are usually not known at the outset: They are usually obtained in a self-consistent calculation of currents and densities throughout the device.[16] Such a calculation makes use of the expression for the currents at the edges of the well and so the distribution of particles transmitted from the well into the barrier at $z=w$ is needed. It is

$$f_{\text{trans}}^+=\theta(p_z)\,\theta[H_z(w,p)-V_0-U(w)]\times\frac{2}{(2\pi\hbar)^3}\,\frac{1}{1+\exp\{[H(w,p)-E_F^+(w)]/k_BT\}}. \tag{4.4}$$

The currents across the edges of the well can then be expressed in terms of these distribution functions as

$$J(0)=\int d^3p\; V_z f^+[E_F^+(0)]+\int d^3p\; V_z f^-[E_F^-(0)], \tag{4.5a}$$

$$J(w)=\int d^3p\; V_z f_{\text{trans}}+\int d^3p\; V_z f_{\text{inj}}^+. \tag{4.5b}$$

These expressions appear to be more familiar when they are evaluated for a nondegenerate distribution

$$J(0)=\frac{mk_B^2T^2}{2\pi^2\hbar^3}\,e^{-V_0/k_BT}\left(e^{E_F(0)/k_BT}-e^{E_F^-(0)/k_BT}\right), \tag{4.6a}$$

$$J(w)=\frac{mk_B^2T^2}{2\pi^2\hbar^3}\,e^{-[V_0+U(w)]/k_BT}(e^{E_F^+(w)/k_BT}-e^{E_F(w)/k_BT}). \tag{4.6b}$$

This expression is, except for the procedures by which the Fermi energies of the mobile carriers are fixed, the same as that used by Grupen and Hess in Eq. (1) of Ref. 16(c).

In the nondegenerate limit, the three densities in Eq. (4.3) at $z=w$ are

$$n_{\rm refl}=\frac{mk_B^2T^2}{2\pi^2\hbar^3}\,e^{[E_F^+(w)-V_0]/k_BT}\left(\frac{1}{\langle V_z(w)\rangle}-\frac{e^{-U(w)/k_BT}}{\langle V_z(0)\rangle}\right), \tag{4.7a}$$

$$n_{\rm inj}=\frac{mk_B^2T^2}{2\pi^2\hbar^3}\,\frac{e^{[E_F(w)-V_0-U(w)]/k_BT}}{\langle V_z(0)\rangle}, \tag{4.7b}$$

$$n_0=\frac{mk_B^2T^2}{2\pi^2\hbar^3}\,\frac{e^{[E_F^-(w)-V_0]/k_BT}}{\langle V_z(w)\rangle}. \tag{4.7c}$$

The equation that determines $E_F^-(w)$ is then

$$e^{E_F^-(w)/k_BT}=(1-\alpha)e^{E_F^+(w)/k_BT}+\alpha e^{E_F(w)/k_BT}, \tag{4.8a}$$

$$\alpha\equiv\left(\frac{\langle V_z(w)\rangle}{\langle V_z(0)\rangle}\right)e^{-U(W)/k_BT}. \tag{4.8b}$$

When there is no potential rise across the well, $\alpha=1$ and $E_F^-(w)=E_F(w)$. When there is a large potential rise across the well, $\alpha=0$ and $E_F^-(w)=E_F^+(w)$. These limits also hold in the degenerate case.

V. CAPTURE AND EMISSION

Capture and emission processes may be regarded as inelastic scattering events in which carriers are transferred back and forth between the mobile (3D) population of the well and the bound (2D) population of the wells. From the standpoint of the mobile carrier equations (3.9), they are an additional sink and source of carriers. The transport equations are now

$$\frac{\partial n^+}{\partial t}+\frac{\partial E_F^+(z,t)}{\partial z}\,\frac{\partial J^+}{\partial E_F^+}=\frac{n^--n^+}{\tau^-}-\frac{n^+(z,t)}{\tau_{\rm capt}(z)}+\frac{S(z,t)}{2}, \tag{5.1a}$$

$$\frac{\partial n^-}{\partial t}+\frac{\partial E_F^-(z,t)}{\partial z}\,\frac{\partial J^-}{\partial E_F^-}=\frac{n^+-n^-}{\tau^-}-\frac{n^-(z,t)}{\tau_{\rm capt}(z)}+\frac{S(z,t)}{2}. \tag{5.1b}$$

$\tau_{\rm capt}(z)$ is the capture time for transferring carriers from the mobile population to the bound one and $S(z,t)$ is the rate of emission from the bound state to the mobile population. It is assumed that half the carriers will be emitted to the right and half to the left.

There are now two mechanisms for equilibrating the populations of left-traveling carriers and right-traveling carriers, namely, direct cap-to-cap scattering and capture (from one distribution) followed by remission (to both distributions). In the absence of precise knowledge of the various rates, we ignore direct cap-to-cap scatter, which we have assumed to be very weak. This will have little effect unless capture and emission is an even weaker process. We rewrite Eq. (5.1) in a form that makes use of the relationship between the flux, the density, and the average z velocity,

$$\frac{\partial n^+}{\partial t}+\frac{\partial E_F^+}{\partial z}\,\frac{\partial J^+}{\partial E_F^+}=\frac{S(z,t)}{2}-\frac{J^+[E_F^+(z,t)]}{\tau_{\rm capt}(z)\langle V_z^+(z)\rangle}, \tag{5.2a}$$

$$\frac{\partial n^-}{\partial t}+\frac{\partial E_F^-}{\partial z}\,\frac{\partial J^-}{\partial E_F^-}=\frac{S(z,t)}{2}-\frac{J^-[E_F^-(z,t)]}{\tau_{\rm capt}(z)\langle V_z^-(z)\rangle}. \tag{5.2b}$$

Because emission of mobile carriers comes primarily from the uppermost bound state in the well, we assume that the shape and intensity of the source function are proportional to the density of carriers in that state,

$$S(z,t)=\frac{\psi(z)^2N[E_F^b(t)]}{\tau_e}. \tag{5.3}$$

$\psi(z)$ is the wave function for the uppermost bound state and $N(E_F^b)$ is the occupation of the associated transverse states, assuming that their occupancy can be characterized by a bound-state Fermi energy. τ_e is the emission time. In equilibrium, there will be no time dependence. The two quasi-Fermi energies $E_F^\pm$ will be equal and independent of position. The two fluxes will be equal and opposite. E_F^b will have the same common value E_F^0 as the other two. From Eqs. (5.2) and (5.3), it follows that

$$\frac{J^\pm(E_F^0)}{\langle V_z^\pm(z)\rangle\tau_{\rm capt}(z)}=\frac{\psi(z)^2N[E_F^0]}{2\tau_e}. \tag{5.4}$$

Inserting Eq. (5.4) into Eq. (5.2) gives

$$\frac{\partial n^+}{\partial t}+\frac{\partial E_F^+}{\partial z}\,\frac{\partial J^+}{\partial E_F^+}=\frac{\psi(z)^2N[E_F^b(t)]}{2\tau_e}\left(1-\frac{J^+[E_F^+(z,t)]N[E_F^0]}{J^+[E_F^0]N[E_F^b(t)]}\right), \tag{5.5a}$$

$$\frac{\partial n^-}{\partial t}+\frac{\partial E_F^-}{\partial z}\,\frac{\partial J^-}{\partial E_F^-}=\frac{\psi(z)^2N[E_F^b(t)]}{2\tau_e}\left(1-\frac{J^-[E_F^-(z,t)]N[E_F^0]}{J^-[E_F^0]N[E_F^b(t)]}\right). \tag{5.5b}$$

In the nondegenerate limit, one can show that

$$\left(1-\frac{J^\pm[E_F^\pm(z,t)]N[E_F^0]}{J^\pm[E_F^0]N[E_F^b(t)]}\right)=1-e^{[E_F^\pm(z,t)-E_F^b(t)]/k_BT}. \tag{5.6}$$

We assert that this form is the correct one to use even in the degenerate limit, although it requires a better derivation than the one presented above to obtain it. The reason for this assertion is given in the Appendix. It makes use of standard expressions for transition rates for Fermi statistics and detailed balance considerations that relate transition rates in the forward and backward directions.

Equations (5.5) and (5.6) are the transport equations that have to be solved for the quantum well in which mobile

carriers can be captured to and emitted from a bound state in the well. The functions $n^{\pm}$ depend explicitly on z and on $E_F^{\pm}$ and are given by Eq. (2.12) (degenerate case) or (2.15) (nondegenerate case). The functions $J^{\pm}$ depend explicitly only on $E_F^{\pm}$ and are given by Eq. (2.11) (degenerate case) or (2.13) (nondegenerate case). Thus Eq. (5.5) is to be regarded as a pair of equations for $E_F^{\pm}(z,t)$. The boundary conditions are Eqs. (4.1) and (4.3) (degenerate case) or (4.8) (nondegenerate case). The coupling between the two equations arises via the bound-state occupancy and via the boundary condition at $z=w$, namely, Eq. (4.3) or (4.8), if there is a potential drop across the well.

The bound-state occupancy (or, equivalently, E_F^b) must be determined by a rate equation that describes the transfer of carriers between the bound state and all populations with which it communicates. This certainly includes the mobile states in the well, but it may include others. For example, carriers in the bound population might recombine directly, either in a spontaneous or stimulated emission process, with carriers of opposite sign in the same well. Alternatively, if the state discussed so far is not the lowest-energy one, then carriers may transfer between it and the lower-energy bound states from which the recombinations actually occur. These considerations, which will differ from case to case, should present no special problems, especially in light of the progress already made by Grupen and co-workers.[16]

VI. SUMMARY

We have presented a pair of coupled equations for describing the distribution and current in a quantum well in which there is a healthy rate of exchange between the mobile (3D) population in the well and the bound (2D) population. To capture the essential physics of this situation, it was necessary to abandon the drift-diffusion formalism and to propose a formalism that is in the same spirit as the flux method used by McKelvey, Longini, and Brody,[20] Shockley,[21] and Alam, Tanaka, and Lundstrom.[22] The quantum-well situation, in which injected particles are accelerated to high speed on entering the well, provides a somewhat clearer physical justification for the method than exists in the original flux method papers.

A key feature of quantum-well transport turned out to be the difficulty of changing the number of particles in the left-going and right-going parts of momentum space, where left and right refer to the direction across the well. There is no such difficulty with respect to transport parallel to the plane of the well. For this reason, ordinary drift-diffusion theory can be used in the parallel direction. The only change needed from drift-diffusion theory as used in the bulk is the value of the mobility and diffusion constants to be used in the well. That change comes about because the scattering time in the well is shorter than in the bulk (because of the increased density of final states at the lowest allowed energy in the restricted momentum space, compared to a lowest allowed energy of zero in the bulk). There may be experimental data relating to this point. Such data would, by its nature, also include a reduction in mobility caused by (or attributed to) additional interface scattering.

Note added in proof. Just before Eq. (3.9), the argument that $E_F^{\pm}$ must be independent of Z is really a consistency condition on the use of Eqs. (3.7) and (3.9). It should be interpreted as stating that these equations are valid to the extent that the change of E_F over the forward scattering mean free path $\tau^+ V_Z$ can be ignored.

ACKNOWLEDGMENT

It is a pleasure to thank M. A. Alam for discussions of early versions of this paper and for pointing out that one of the intermediate assumptions in it has been confirmed by numerical work on quantum tunneling.[23,24]

APPENDIX: CAPTURE AND EMISSION RATES

Consider two types of states $i=b,m$ (bound and mobile) that can hold electrons. Let $g_i(E_i)$ be the density of states of type i at energy E_i, $f_i(E_i)$ be the fraction of states of type i occupied at E_i, $S_{ij}(E_i \rightarrow E_j)$ be the transition probability, $n_i(E_i)=g_i(E_i)f_i(E_i)$, and $\overline{n_i}(E_i)=g_i(E_i)[1-f_i(E_i)]$. The rate equation for the mobile carriers, taking into account the Pauli principle, is

$$\frac{dn_m(E_m)}{dt}=\int n_b(E_b)S_{bm}(E_b\rightarrow E_m)\overline{n_m}(E_m)dE_b - \int n_m(E_m)S_{mb}(E_m\rightarrow E_b)\overline{n_b}(E_b)dE_b\,. \tag{A1}$$

Similarly, the rate equation for the bound carriers has the same form with the subscripts m and b interchanged. The total number of carriers of type i is

$$N_i=\int n_i(E_i)dE_i\,.$$

Thus

$$\frac{dN_m}{dt}=\int\int dE_b dE_m n_b(E_b)S_{bm}(E_b\rightarrow E_m)\overline{n_m}(E_m) \times\left(1-\frac{n_m(E_m)}{\overline{n_m}(E_m)}\frac{S_{mb}(E_m\rightarrow E_b)}{S_{bm}(E_b\rightarrow E_m)}\frac{\overline{n_b}(E_b)}{n_b(E_b)}\right). \tag{A2}$$

The first term in the integral in Eq. (A2) is the rate of emission from the bound states to the mobile population; the second is the rate of capture from the mobile states to the bound population, that is,

$$\frac{dN_m}{dt}=E-C.$$

Assume that the mobile population is in a Fermi distribution with a Fermi energy E_F^m and that the bound population is in a Fermi distribution with a Fermi energy E_F^b. Then

$$\frac{\overline{n_b}(E_b)}{n_b(E_b)}=e^{[E_b-E_F^b]/k_BT}, \tag{A3a}$$

$$\frac{n_m(E_m)}{\overline{n_m}(E_m)}=e^{-[E_m-E_F^m]/k_BT}. \tag{A3b}$$

In equilibrium, the two Fermi energies are equal and, according to the principle of detailed balance, the contents of the large parentheses in Eq. (A2) must vanish. Using Eq. (A3), this condition requires that

$$\frac{S_{mb}(E_m \rightarrow E_b)}{S_{bm}(E_b \rightarrow E_m)} = e^{[E_m - E_b]/k_B T}. \tag{A4}$$

We assume that the system is not so far out of equilibrium as to affect the transition rates. In that case, the large parentheses in Eq. (A2) can be evaluated using Eqs. (A3) and (A4),

$$\left(1 - \frac{n_m(E_m)}{\overline{n_m}(E_m)} \frac{S_{mb}(E_m \rightarrow E_b)}{S_{bm}(E_b \rightarrow E_m)} \frac{\overline{n_b}(E_b)}{n_b(E_b)}\right) = 1 - e^{[E_F^m - E_F^b]/k_B T}. \tag{A5}$$

The spirit of the classical development used in the paper is that everything happens locally with no regard to coherence. Clearly, that is not correct, but if this principle were used as a heuristic guide, we could regard Eq. (A2) as a local equation, in which $n_b(E_b)$, $\overline{n_m}(E_m)$, and E_F^m were all z-dependent quantities. We would then replace $E_F^m \rightarrow E_F^{\pm}(z)$ on the right-hand side of Eq. (A5), providing thereby the heuristic justification for the use of Eq. (5.6) in the text. Furthermore, with the identification

$$\int dE_m S_{bm}(E_b \rightarrow E_m) \overline{n_m}(E_m) = \frac{1}{2\tau_e(E_b)}, \tag{A6}$$

we have

$$\frac{dN_m}{dt} = (1 - e^{[E_F^{\pm}(z) - E_F^b]/k_B T}) \int dE_b \frac{n_b(z, E_b)}{2\tau_e(E_b)}. \tag{A7}$$

The approximation made to go from Eq. (A7) to Eq. (5.5) and (5.6) neglecting the energy dependence of τ_e.

[1] W. Rideout, W. F. Sharfin, E. S. Koteles, M. O. Vassell, and B. Elman, IEEE Photon. Technol. Lett. **3**, 784 (1991).

[2] R. Nagarajan, M. Ishikawa, T. Fukushima, R. S. Geels, and J. E. Bowers, IEEE J. Quantum Electron. **28**, 1990 (1992).

[3] H. Hirayama, J. Yoshida, Y. Miyake, and M. Asada, IEEE J. Quantum Electron. **30**, 54 (1994).

[4] G. Bastard, Phys. Rev. B **30**, 3547 (1984).

[5] J. A. Brum and G. Bastard, Phys. Rev. B **33**, 1420 (1986).

[6] S. V. Kozyrev and A. Ya. Shik, Fiz. Tekh. Poluprovodn. **19**, 1667 (1985) [Sov. Phys. Semicond. **19**, 1024 (1986)].

[7] P. Sotirelis and K. Hess, Phys. Rev. B **49**, 7543 (1994).

[8] S. Weiss, J. M. Weisenfeld, D. S. Chemla, G. Raybon, G. Sucha, M. Wegener, G. Eisenstein, C. A. Burrus, A. G. Dentai, U. Koren, B. I. Miller, H. Temkin, R. A. Logan, and T. Tanbun-Ek, Appl. Phys. Lett. **60**, 9 (1992).

[9] R. Kersting, R. Schwedler, K. Wolter, K. Leo, and H. Kurz, Phys. Rev. B **46**, 1639 (1992).

[10] P. W. M. Blom, C. Smit, J. E. M. Haverkort, and J. H. Wolter, Phys. Rev. B **47**, 2072 (1993).

[11] D. Morris, B. Devaud, A. Regreny, and P. Auvray, Phys. Rev. **47**, 6819 (1993).

[12] M. R. X. Barros, P. C. Becker, D. Morris, B. Devaud, A. Regreny, and F. Beisser, Phys. Rev. B **47**, 10 951 (1993).

[13] B. Devaud, A. Chomette, D. Morris, and A. Regreny, Solid State Commun. **85**, 367 (1993).

[14] R. Stratton, IEEE Trans. Electron Devices **ED-19**, 1288 (1972).

[15] A. A. Grinberg, Phys. Rev. B **33**, 7256 (1986).

[16] (a) M. Grupen and K. Hess (unpublished); (b) M. Grupen, G. Kosinovsky, and K. Hess, *Proceeding of the International Electron Device Meeting, Washington, DC* (IEEE, New York, 1993), p. 23.6.1; (c) M. Grupen and K. Hess (unpublished); (d) M. Grupen, SPIE Proc. **2693**, 374 (1996); (e) M. Grupen and K. Hess, IEEE J. Quantum Electron. (to be published).

[17] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).

[18] K. Hess, *Advanced Theory of Semiconductor Devices* (Prentice-Hall, Englewood Cliffs, NJ, 1988).

[19] A. H. Marshak and K. M. Van Vliet, Solid State Electron. **21**, 417 (1978).

[20] J. P. McKelvey, R. L. Longini, and T. P. Brody, Phys. Rev. **123**, 51 (1961).

[21] W. Shockley, Phys. Rev. **125**, 1570 (1962).

[22] M. A. Alam, S.-I. Tanaka, and M. S. Lundstrom, Solid State Electron. **38**, 177 (1995).

[23] A. N. Khondker and M. A. Alam, Phys. Rev. B **45**, 8516 (1992).

[24] M. A. Alam, R. A. Morrisey, and A. N. Khondker, J. Appl. Phys. **71**, 3077 (1992).