

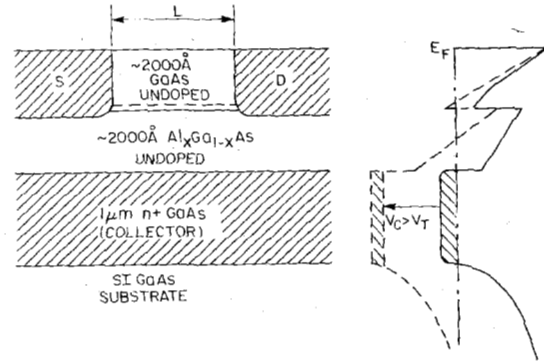
Theory of Hot-Electron Injection in CHINT/NERFET Devices

ANATOLY A. GRINBERG, ALEXANDER KASTALSKY, AND SERGE LURYI, SENIOR MEMBER, IEEE

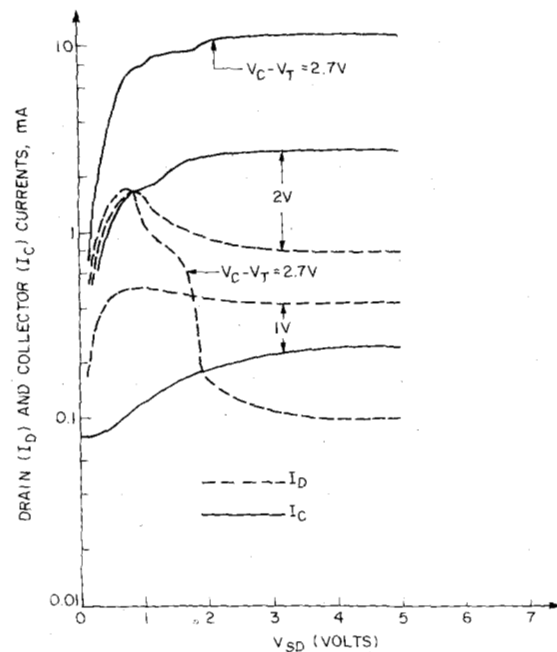
Abstract—We have considered theoretically the basic physical processes underlying the operation of the charge injection transistor (CHINT) and the negative resistance field-effect transistor (NERFET). Our treatment is based on the electron temperature (T_e) approximation for the energy distribution of hot electrons in the two-dimensional electron gas (2DEG) channel. The T_e is determined from an energy-balance equation which includes the following processes: 1) electron heating by the source-drain electric field (assumed uniform), 2) energy losses due to the interaction with phonons, 3) energy losses due to the emission of hot electrons from the channel into the second conducting layer (collector) and the attendant nonconservation of the channel current. Our theory gives a semi-quantitative analytical description of the current-voltage characteristics of CHINT/NERFET devices. Most—but not all—of the important experimental features of the device operation are adequately described. Further improvement of the theory should include a realistic description of the field nonuniformity along the channel.

I. INTRODUCTION

THE EFFECT of real-space transfer (RST) of hot electrons in multilayer semiconductor structures was first suggested by Hess *et al.* [1]. Recently, a novel class of three-terminal devices was proposed [2] that employs the RST between semiconducting layers, separated by a potential barrier and contacted individually. These devices, the charge injection transistor (CHINT), the negative resistance transistor (NERFET), as well as their logic and memory circuits, were subsequently implemented by MBE in GaAs/AlGaAs heterostructures and extensively studied (see [3], [4] and references therein). More recently, a significant improvement was achieved in OMCVD grown structures [5], [6]. A current gain in CHINT was demonstrated at frequencies up to 32 GHz, with a transconductance of more than 1000 mS/mm, and the negative differential resistance (NDR) in NERFET had a peak-to-valley ratio of 160 at room temperature. The typical configuration of these devices, and their basic characteristics are shown in Fig. 1. The operation of CHINT is based on the control of charge injection of hot electrons from a source-drain (SD) channel into a second conducting layer by applying a heating voltage V_{SD} , whereas the NERFET is based on a pronounced negative differential resistance in the SD circuit, controlled by voltage on the second con-



(a)



(b)

Fig. 1. (a) Schematic illustration and energy-band diagram of the CHINT/NERFET device structure. A conducting channel (2D-EG) appears at the undoped heterointerface when the collector voltage exceeds a threshold value, $V_C > V_T$. (b) Typical experimental [5] current-voltage characteristics, I_C - V_{SD} and I_D - V_{SD} at room temperature and different collector voltages $V_C - V_T$. The measured value of the threshold voltage was $V_T = 5$ V. With increasing heating voltage V_{SD} , one clearly sees a rapid rise and subsequent saturation of the collector current I_C , accompanied by an increase and subsequent sharp drop in the drain current I_D .

Manuscript received March 31, 1986; revised July 22, 1986.

A. A. Grinberg and S. Luryi are with AT&T Bell Laboratories, Murray Hill, NJ 07974.

A. Kastalsky is with Bell Communications Research, Murray Hill, NJ 07974.

IEEE Log Number 8611085.

ducting layer (the latter will be referred to as the hot-electron collecting gate or "collector").¹

So far, much of the understanding of the operation of CHINT and NERFET has remained on a qualitative level [3]. Quantitative description of these devices is substantially more complicated than that of an ordinary heterojunction FET, despite an apparent similarity in the device structures. Unlike the FET case, where hot-electron phenomena appear "merely" through their effect on the carrier transport in the channel, these phenomena dominate the CHINT/NERFET operation through RST. Moreover, the RST leads to a *nonconservation* of the channel current $I(x)$, which may have a strong influence on the field distribution within the channel. Another fundamental complication is associated with the energy flux carried away by hot electrons leaving the channel.

The purpose of this work is to present a quantitative description of CHINT/NERFET characteristics. A satisfactory theory should describe the following experimental facts evident from Fig. 1(b):

- 1) The NDR in the SD circuit accompanying the increase in the collector current.
- 2) Saturation in both the drain current I_D and the collecting-gate current I_C with increasing heating voltage V_{SD} .
- 3) Initial growth and subsequent decay in the values of the saturated current $I_{D_{sat}}$ with increasing V_C (a negative transconductance).

The semi-analytical theory presented in this paper does not aim at a complete description of all aspects of the device characteristics. Fitting of the experimental data is complicated by the fact that the device technology has not yet reached a sufficient maturity—the characteristics are not well reproducible from one wafer to another, being influenced by a number of inessential (in principle) effects, such as imperfect contacts, etc.

II. PHYSICAL MODEL

The sheet carrier concentration $n(x)$ in the NERFET channel is essentially governed by the process of hot electrons leaving the channel. This means that $n(x)$ must be determined from the current continuity equation. Capacitive coupling enters only as a boundary condition at the source end of the channel

$$en(0) = \frac{\epsilon}{4\pi d} (V_C - V_T) \quad (1)$$

where d is the thickness of the collector barrier (see Fig. 1(a)) and V_T is a threshold voltage for the appearance of a two-dimensional electron gas (2-DEG) in the channel

¹In first implementations the second conducting layer was realized as a heavily doped n-GaAs substrate and hence designated by SUB. Electrically, the SUB electrode in CHINT plays the role analogous to the collector of a bipolar transistor, whereas in NERFET it is more like the gate of an FET. Physically, in both devices it plays the dual role of a sink for hot electrons and a field source inducing the carrier concentration in the hot-electron channel.

induced by V_C . In what follows we shall be counting V_C from the threshold level and set $V_T = 0$.

In modeling the properties of 2-DEG, we shall take into consideration only the two lowest subbands. Positions of the bottom edges, E_1 and E_2 , of these subbands will be assumed given in terms of the local carrier concentration in the channel [7]

$$\begin{aligned} E_1 &= \beta_1 n^{2/3}(x) \\ E_2 &= \beta_2 n^{2/3}(x) \end{aligned} \quad (2)$$

where β_1 and β_2 are numerical coefficients

$$\begin{aligned} \beta_1 &= 7.24 \cdot 10^{-10} \text{ eV} \cdot \text{cm}^{4/3} \\ \beta_2 &= 11.37 \cdot 10^{-10} \text{ eV} \cdot \text{cm}^{4/3}. \end{aligned} \quad (3)$$

(These values follow from the numerical calculations of Stern and Das Sarma [8], which give $E_1 = 45.6$ meV and $E_2 = 71.6$ meV for $n = 5 \cdot 10^{11} \text{ cm}^{-2}$.) Exchange-correlation corrections to the subband energies ($\propto n^{4/9}$ [8], [9]) will be neglected for the sake of simplicity.

We shall assume that the nonequilibrium properties of the hot-electron ensemble can be described by a pseudo-Maxwellian distribution characterized by an electron temperature T_e , which in turn will be determined from an energy-balance equation. In the electron-temperature approximation, the local sheet carrier concentration $n(x)$ in the 2-DEG is related to the quasi-Fermi level $\tilde{E}_F(x)$, counted from the classical edge of conduction band, as follows [10]:

$$n = D \ln ([1 + \exp(\tilde{E}_F - \tilde{E}_1)][1 + \exp(\tilde{E}_F - \tilde{E}_2)]) \quad (4)$$

where $D = mkT_e/\pi\hbar^2$ is the areal density of states in the energy interval kT_e , and \tilde{E}_F and \tilde{E}_1 are the corresponding energies in units of kT_e . Below, we shall be using the tilde notation always to describe energies in units kT_e and voltages in units kT_e/e .

The main simplifying approximation made in this work is the assumption of a uniform lateral electric field in the channel, $F_{SD} = V_{SD}/L$. As will be discussed in Section III, this approximation is not entirely satisfactory and its use will cost us an inability to give a complete quantitative account of the device behavior, especially in the range of current saturation. Another approximation made is that we shall neglect electron scattering into the subsidiary minima (L and X) of the conduction band. We feel this approximation may be justified by the fact that for $x \approx 0.4$, the satellite valleys are nearly continuous at the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ interface, cf. the discussion at the end of Section II-B.

A. Charge Injection Current

Hot electrons are injected into the collector either by a thermionic emission over the barrier or by a "thermally assisted" tunneling under the top of the barrier. As will be seen below, the tunneling component cannot be neglected in the calculation of device characteristics.

1) *Thermionic Emission*: Although, for $kT_e \lesssim (E_2 - E_1)$, most of the channel electrons are two-dimensional (i.e., located within the lowest subband), those electrons which participate in the charge injection over the barrier (of height $\Phi \approx 0.3$ eV equal to the conduction-band discontinuity between $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and GaAs at $x \approx 0.4$) are located in the high-energy tail of the hot-electron distribution function and, therefore, must be treated as three-dimensional. Their flux over the barrier Φ is hence given by a Richardson-like equation

$$J = A^* T_e^2 \exp(\tilde{E}_F - \tilde{\Phi}) \quad (5)$$

where J is the current density of hot-electron injection, and A^* is the effective Richardson constant containing the electron effective mass in the channel [11] (because the effective electron mass in the channel is lower than that in the AlGaAs barrier). We have assumed in (5) that the shape of the barrier is such that the reverse flux of electrons from the collector into the channel can be neglected (because the collector temperature is low) even when $V_{SD} > V_C$.

2) *Thermally Assisted Tunneling*: Inclusion of tunneling gives rise to an additional factor γ in the current equation (5) (see, for instance, [12]). Although the shape of the potential barrier is trapezoidal, we can regard it as approximately triangular near the top of the barrier—where most of the tunneling occurs. The usual quasi-classical theory for tunneling under a barrier gives in this case (see, for instance, [13])

$$\gamma = 1 + \int_0^{\tilde{\Phi}} \exp(\xi - (\xi/\tilde{E}_{00})^{3/2}) d\xi \quad (6)$$

where

$$\tilde{E}_{00} = \frac{1}{kT_e} \left(\frac{3\hbar e F(0)}{4(2m)^{1/2}} \right)^{2/3} \quad (7)$$

and $F(0)$ is the electric field near the top of the barrier. We shall assume that this field is given by the Gauss law $F(0) = 4\pi en/\epsilon$. The total current density across the barrier is of the form

$$J = \gamma A^* T_e^2 \exp(\tilde{E}_F - \tilde{\Phi}) \quad (8)$$

Fig. 2 shows the dependence of the injection current density J on the electron temperature for different values of the carrier concentration, $n = 10^{11}$, $5 \cdot 10^{11}$, and 10^{12} cm^{-2} . For comparison, the values of J calculated from (5), i.e., without the tunneling effect, also are shown (by the dashed lines). When tunneling is not included, then the concentration dependence (at a given T_e) arises solely due to the variation of the Fermi level. As expected, the importance of tunneling diminishes with increasing T_e . Tunneling also becomes less important at lower concentrations, since those correspond to lower electric fields $F(0)$ and hence less transparent barriers. We see that in a wide range of T_e and n the tunneling component cannot be neglected. As will be shown below, taking account of tunneling also improves our description of the experimental data.

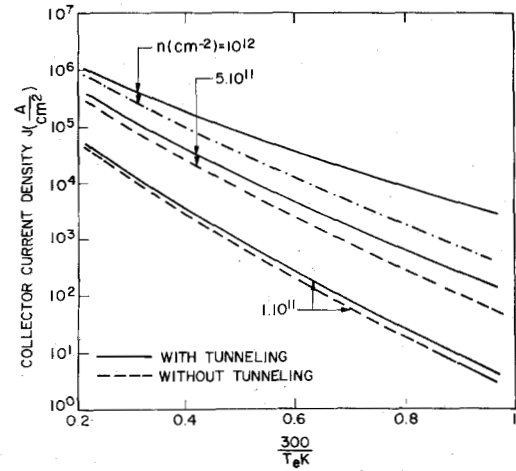


Fig. 2. Calculated injection current density J as a function of the electron temperature T_e for different electron densities n in the channel. Note that the tunneling component becomes quite important at higher n (corresponding to higher electric fields and, therefore, more transparent barriers).

B. Energy Balance Equation

The energy balance equation for the electron gas in the channel besides the usual terms corresponding to the energy gain in the external field ($I \cdot F$) and the energy loss to the lattice, must contain two additional terms, associated with the electron emission over the barrier. On the one hand, electrons leaving the channel carry away their energy, and on the other hand, the nonconservation of channel current results in a peculiar energy transport along the channel. These effects, proportional to the injection current density J , are entirely novel and specific to the CHINT/NERFET device. As will be shown below, at high electron temperatures, the energy carried away by hot electrons injected across the barrier becomes dominant in the balance equation. It gives rise to a feedback mechanism which stabilizes the T_e and thus limits the injection current.

The problem of electron-gas heating and determination of the distribution function of hot electrons is quite involved in general. (In the 2-D case it is further complicated by the subband structure, which requires the consideration of not only intra- but also inter-subband transitions, as discussed in the extensive literature on the subject; cf. the review [7]). In a phenomenological treatment of the device characteristics attempted here, it is unreasonable to include all the details of the electron heating problem—which would be appropriate in a study of the heating effect itself. We shall be using instead the simplest and most common approximation in which the non-equilibrium distribution will be assumed to differ from the equilibrium case only by the temperature T_e , and the latter will be evaluated from the energy balance equation. As is well known, the establishment of such a quasi-equilibrium distribution can be ensured by a sufficiently strong electron-electron interaction, which in turn is realized at high enough electron concentrations. Although, the electron concentration will be shown to decrease substantially

along the channel due to the injection effect, nevertheless the main portion of the collector current—and it is for the determination of this current that we need to determine T_e anyway—flows from the channel region where the concentration is still high enough ($\geq 10^{11} \text{ cm}^{-2}$). We can expect, therefore, that the electron temperature approximation will be satisfactory for our purposes.

Consider the basic components of the energy balance equation which governs the temperature of channel electrons heated by an applied electric field.

1) *The Energy Transport Along the Channel:* Under stationary conditions the net number of electrons arriving at a point in the channel from the adjacent regions must equal their number leaving the channel at that point. The electron flux per unit area of the channel is given by J/e . Denote the mean electron energy of a channel electron by $\langle E \rangle$. (Both J and $\langle E \rangle$ are position dependent, in general.) Then the rate of energy change per channel electron at a given point is of the form

$$w_1 = \langle E \rangle \frac{J}{en} \quad (9)$$

The physical meaning of (9) is as follows. In the absence of electron transfer across the barrier, when the number of electrons entering an elementary channel volume equals their number leaving that volume, the divergence of the electron energy-density flux must equal the gradient of the average energy density (cf. the Appendix). Neglecting processes of thermal conductivity, the total variation of the energy of a given elementary volume would be owing only to the difference in the average energy of electrons entering at point x and leaving at point $x + \delta x$. If we neglect that difference (the Thomson heat), then the total variation of energy along δx vanishes. In the presence of a real-space transfer, however, the incoming flux of electrons entering an elementary volume exceeds their outgoing flux after the distance δx by the amount $(J/e)\delta x$ and the taken-away energy equals $\langle E \rangle (J/e)\delta x$, which leads to (9) per unit length of the channel.

The value of $\langle E \rangle$ can be expressed in terms of the local quasi-Fermi level E_F as follows:

$$\begin{aligned} \langle E \rangle &= \frac{2kT_e}{(2\pi)^2 n} \int \frac{\tilde{E}_k d^2k}{1 + \exp(\tilde{E}_k - \tilde{E}_F + \tilde{E}_1)} \\ &= \frac{DkT_e}{n} \int_0^\Delta \frac{\ln(1 + \xi)}{\xi} d\xi \end{aligned} \quad (10)$$

where

$$E_k = \frac{\hbar^2 k^2}{2m}$$

and

$$\Delta = \exp(\tilde{E}_F - \tilde{E}_1) = \exp(n/D) - 1. \quad (11)$$

For the sake of simplicity, in the above equations and below we restrict our consideration to one 2-D subband (E_1).

2) *Energy Losses in the Emission of Channel Electrons into the Collector:* Electrons leaving the channel by going over the barrier of height Φ take away the energy (per channel electron) at the rate

$$w_2 = \frac{2}{(2\pi)^3 n} \int f(E_k) E v_\perp d^3k D^*(E_\perp) \quad (12)$$

where $v_\perp = \hbar k_\perp / m$ is the component of electron velocity normal to the interface, $f(E)$ the electron Fermi distribution function corresponding to the temperature T_e , and $D^*(E_\perp)$ is a factor describing the barrier transparency, which we shall take equal to unity when the transverse kinetic energy component exceeds the barrier height, $E_\perp > \Phi$ (i.e., we shall neglect the probability of a quantum-mechanical above-barrier reflection). Counting all energies from the classical bottom of the conduction band in the channel (the bottom of the quantum well for 2DEG), we can express the energy near the top of the barrier in the form

$$E_k = \frac{\hbar^2 k_\parallel^2}{2m} + \frac{\hbar^2 k_\perp^2}{2m}$$

where k_\parallel and k_\perp are, respectively, the parallel and the perpendicular (to the plane of the channel) components of an electron wave vector. At low energies this dispersion law goes over into $E_k = E_1 + (\hbar^2 k_\parallel^2 / 2m)$, so that the actual kinetic energy carried away by an electron equals

$$\frac{\hbar^2 k_\parallel^2}{2m} + \frac{\hbar^2 k_\perp^2}{2m} - E_1.$$

Therefore, (12) assumes the form

$$\begin{aligned} w_2 &= \frac{2}{(2\pi)^3 n} \int f(E_k) (E_k - E_1) v_\perp D^*(E_\perp) d^3k \\ &= \frac{m \exp(\tilde{E}_F / (kT_e))^3}{2\pi^2 \hbar^3 n} \int_0^\infty e^{-\tilde{E}_\parallel} d\tilde{E}_\parallel \int_{\tilde{E}_1}^\infty \\ &\quad \cdot e^{-\tilde{E}_\perp - \chi(\tilde{E}_\perp)} (\tilde{E}_\parallel + \tilde{E}_\perp - \tilde{E}_1) d\tilde{E}_\perp \end{aligned} \quad (13)$$

where, for a triangular barrier

$$\chi(\tilde{E}_\perp) = [(\tilde{\Phi} - \tilde{E}_\perp) / \tilde{E}_{00}]^{3/2}.$$

In the integration over \tilde{E}_\perp we can extend the lower limit to zero and obtain

$$\begin{aligned} w_2 &= A^* T_e^2 (kT_e / ne) [2 + \tilde{\Phi} - \tilde{E}_1 \\ &\quad + (1 + \tilde{\Phi} - \tilde{E}_1) I_1 - I_2] e^{\tilde{E}_F - \tilde{\Phi}} \end{aligned} \quad (14)$$

where

$$I_\alpha = \int_0^{\tilde{\Phi}} \exp \left[\xi - \left(\frac{\xi}{\tilde{E}_{00}} \right)^{3/2} \right] \xi^{\alpha-1} d\xi, \quad \alpha = 1, 2. \quad (15)$$

The first three terms in the square brackets in (14) correspond to the loss of energy in the purely thermionic emis-

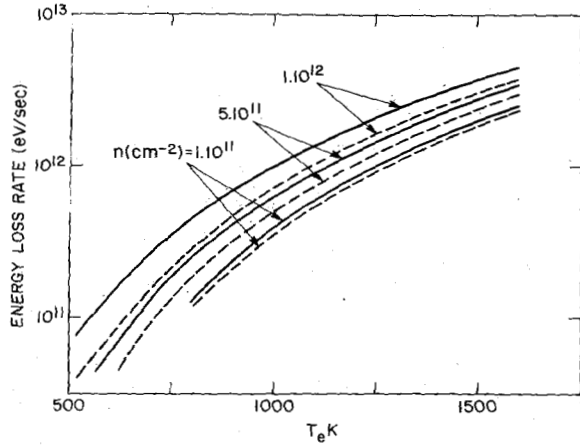


Fig. 3. Energy loss rate w_2 per one electron as a function of the electron temperature T_e at different channel concentrations n . These losses correspond to the energy carried away from the channel by hot electrons emitted over the barrier.

sion process, while the other terms describe the contributions due to tunneling.

The dependence of w_2 on the electron temperature is displayed in Fig. 3 for three different values of the concentration. Dashed lines show the same curves evaluated without the inclusion of tunneling. We see that the tunneling corrections, although important, do not give rise to any qualitative change in the temperature dependence of w_2 .

3) Energy Losses Due to Interaction with Polar Optic Phonons: In order to describe the energy losses associated with the phonon emission by hot electrons, one has to assume a specific form of the envelope wave function of 2-D electrons. Considering only the lowest subband electrons, we shall assume that their motion transverse to the channel is described by a wave function of the form

$$\psi(z) = \frac{\alpha^{3/2}}{\sqrt{2}} z \exp\left(-\frac{\alpha z}{2}\right) \quad (16)$$

where the parameter α is determined variationally [10], [14], and in terms of the bottom energy of the lowest subband E_1 it is given by

$$\alpha = \left(\frac{8mE_1}{5\hbar^2}\right)^{1/2} \quad (17)$$

At high temperatures ($T_e \geq 300$ K) the dominant mechanism of energy losses in phonon emission is known to be the interaction with polar optic phonons (see, for example, [15], [16]). The energy loss rate in these processes was calculated by Price [17] for electronic wave functions corresponding to a rectangular quantum well. Our calculation below is similar to [17] except for the assumed form (16) of the wave functions.

To calculate the average rate of electron energy losses due to their interaction with polar optic phonons, we shall

use the master equation describing the rate of change of the number N_q^* of phonons of wave vector q (per unit area of the 2-DEG)

$$\begin{aligned} \frac{\partial N_q^*}{\partial t} = & \frac{4\pi}{\hbar S_0} \sum_k W(q_{\parallel}, q_{\perp}) \delta(E_{k+q_{\parallel}} - E_k - \hbar\omega_0) \\ & \times \{(N_q + 1) f(E_{k+q_{\parallel}}) [1 - f(E_k)] \\ & - N_q f(E_k) [1 - f(E_{k+q_{\parallel}})]\} \end{aligned} \quad (18)$$

where k is the 2-D electron wave vector, $E_k = \hbar^2 k^2 / 2m$, $\hbar\omega_0$ is the optical phonon energy, S_0 the normalizing area, q_{\parallel} and q_{\perp} are, respectively, the components of the phonon wave vector that are parallel and perpendicular to the plane of the 2-DEG, N_q is the Planck function for phonons, which we shall assume to correspond to an equilibrium at the lattice temperature T , i.e.,

$$N_q = [\exp(\hbar\omega_0/kT) - 1]^{-1}$$

and $W(q_{\parallel}, q_{\perp})$ is the square of the matrix element for electron-phonon interaction, given by [18]

$$W(q_{\parallel}, q_{\perp}) = \frac{2\pi\hbar^2 e E_0}{m V_0 (q_{\perp}^2 + q_{\parallel}^2)} |I(q_{\perp})|^2 \quad (19)$$

where V_0 is a normalizing volume

$$e E_0 = \frac{m e^2 \omega_0}{\hbar} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_0} \right) \quad (20)$$

ϵ_0 and ϵ_{∞} are, respectively, the static and the high-frequency permittivities of the sample, and

$$|I(q_{\perp})|^2 = \left| \int_{-\infty}^{\infty} \psi(z) e^{iq_{\perp}z} dz \right|^2 = \frac{\alpha^6}{(\alpha^2 + q_{\perp}^2)^3} \quad (21)$$

The total rate of change of the phonon energy per unit area of the 2-DEG is given by

$$\begin{aligned} \frac{\partial \langle E_{ph} \rangle}{\partial t} = & \frac{V_0}{(2\pi)^3} \int \hbar\omega_0 \frac{\partial N_q^*}{\partial t} d^3 q \\ = & \frac{2\hbar e E_0 (\hbar\omega_0)}{(2\pi)^3 m} N_q \left(\exp \left[\frac{\hbar\omega_0}{k} \left(\frac{1}{T} - \frac{1}{T_e} \right) \right] - 1 \right) \\ & \times \int \frac{|I(q_{\perp})|^2}{(q_{\parallel}^2 + q_{\perp}^2)} f(E_k) [1 - f(E_k + \hbar\omega_0)] \\ & \cdot \delta(E_{k+q_{\parallel}} - E_k - \hbar\omega_0) dq_{\perp} d^2 q_{\parallel} d^2 k \end{aligned} \quad (22)$$

where we have taken into account that the electron distribution function $f(E_k)$ is assumed to differ from the equilibrium distribution only by the electron temperature parameter T_e .

Integration over q_{\perp} and the directions of q_{\parallel} and k brings

(22) in the form

$$\begin{aligned} \frac{\partial \langle E_{\text{ph}} \rangle}{\partial t} &= \frac{eE_0\omega_0\alpha(2m)^{1/2}}{2^4\pi\hbar} N_q \left(\exp \left[\frac{\hbar\omega_0}{k} \left(\frac{1}{T} - \frac{1}{T_e} \right) \right] - 1 \right) \\ &\times \int_0^\infty \frac{(3q_{\parallel}^2 + 9q_{\parallel}\alpha + 8\alpha^2)}{q_{\parallel}(q_{\parallel} - \alpha)^3} \\ &\cdot \frac{f[E_k + E^*(q_{\parallel})] \{1 - f[E_k + E^*(q_{\parallel}) + \hbar\omega_0]\}}{\sqrt{E_k}} dE_k dq_{\parallel} \end{aligned} \quad (23)$$

where

$$E^*(q_{\parallel}) = \frac{\hbar^2}{2m} \left(\frac{q_{\parallel}}{2} - \frac{m\omega_0}{\hbar q_{\parallel}} \right)^2.$$

Introducing dimensionless variables $\zeta = q_{\parallel}/\alpha$ and $\eta^2 = \tilde{E}_k$, we have

$$\begin{aligned} \frac{\partial \langle E_{\text{ph}} \rangle}{\partial t} &= \frac{eE_0\omega_0}{2^3\pi\hbar} (2m)^{1/2} N_q (kT_e)^{1/2} \\ &\cdot \left(\exp \left[\frac{\hbar\omega_0}{k} \left(\frac{1}{T} - \frac{1}{T_e} \right) \right] - 1 \right) \\ &\times \int_0^\infty \frac{3\zeta^2 + 9\zeta + 8}{\zeta(1 + \zeta)^3} f[\eta^2 + \tilde{E}^*(\zeta)] \\ &\cdot \{1 - f[\eta^2 + \tilde{E}^*(\zeta) + \hbar\tilde{\omega}_0]\} d\eta d\zeta \end{aligned} \quad (24)$$

where

$$\tilde{E}^*(\zeta) = \frac{1}{5} \tilde{E}_1 \left(\zeta - \frac{5}{4\zeta} \frac{\hbar\tilde{\omega}_0}{\tilde{E}_1} \right)^2 \quad (25a)$$

$$\begin{aligned} f[\eta^2 + \tilde{E}^*(\zeta)] &= (e^{\eta^2 + \tilde{E}^*(\zeta) + \tilde{E}_1 - \tilde{E}_F} + 1)^{-1} \\ &= \left(\frac{\exp[\eta^2 + \tilde{E}^*(\zeta)]}{[\exp(n/D) - 1]} + 1 \right)^{-1} \end{aligned} \quad (25b)$$

$$\begin{aligned} f[\eta^2 + \tilde{E}^*(\zeta) + \hbar\tilde{\omega}_0] &= (e^{\eta^2 + \tilde{E}^*(\zeta) + \hbar\tilde{\omega}_0 + \tilde{E}_1 - \tilde{E}_F} + 1)^{-1} \\ &= \left(\frac{\exp[\eta^2 + \tilde{E}^*(\zeta) + \hbar\tilde{\omega}_0]}{[\exp(n/D) - 1]} + 1 \right)^{-1} \end{aligned} \quad (25c)$$

Note that we used relation (11) in (25b) and (25c). Also, at high electron temperatures when the degeneracy is lifted, one can replace the Fermi functions by the corresponding Boltzmann factors and (24) reduces to

$$\begin{aligned} \frac{\partial \langle E_{\text{ph}} \rangle}{\partial t} &= \frac{eE_0\omega_0}{2^4\pi^{1/2}\hbar} (2m)^{1/2} N_q (kT_e)^{1/2} \\ &\cdot \left(\exp \left[\frac{\hbar\omega_0}{k} \left(\frac{1}{T} - \frac{1}{T_e} \right) \right] - 1 \right) [e^{(n/D)} - 1] \\ &\times \int_0^\infty \frac{3\zeta^2 + 9\zeta + 8}{\zeta(1 + \zeta)^3} e^{-\tilde{E}^*(\zeta)} d\zeta. \end{aligned} \quad (26)$$

Fig. 4 shows the effective energy relaxation time τ_E defined by

$$\tau_E = k(T_e - T) \left(\frac{1}{n} \frac{\partial \langle E_{\text{ph}} \rangle}{\partial t} \right)^{-1} \quad (27)$$

and calculated from (24) as a function of the electron temperature at a fixed lattice temperature $T = 300$ K. Three curves shown correspond to the electron concentrations $n = 10^{11}$, $5 \cdot 10^{11}$, and 10^{12} cm^{-2} . The relaxation time decreases with increasing concentration for the following reason:

Higher concentrations correspond to stronger transverse electric fields and hence entail higher values of the parameter α describing the compression of the 2DEG wave function (16). As evident from (21), this leads to an increase in the magnitude of the form factor $|I|$. Physically, the more compressed the wave function is the shorter is the wavelength of phonons which can efficiently interact with the 2DEG. At sufficiently high concentrations, however, τ_E ceases to decrease with increasing n —because all phonons with $\hbar q \leq (2mkT_e)^{1/2}$ (i.e., those which can interact with electrons conserving both energy and momentum) are already fully involved in the interaction. At lower T_e this saturation of the $\tau_E(n)$ dependence occurs earlier, e.g., for $T_e \sim 300$ K this dependence is practically saturated already at $n \geq 10^{11} \text{ cm}^{-2}$. (In this discussion we have ignored the Fermi degeneracy, which is permissible to do since in most cases we have $n \leq D$. In the opposite limit ($n \gg D$) the situation is more complicated, since for $kT_e > \hbar\omega_0$ only those electrons whose energies lie within a band of order kT_e from the Fermi surface can both absorb and emit phonons. Electrons below this band do not participate in the scattering processes.)

Our numerical calculations were carried out on the basis of (24), which is applicable for arbitrary degrees of degeneracy. In order to facilitate the comparison, the loss rate per electron

$$w_3 = \frac{1}{n} \frac{\partial \langle E_{\text{ph}} \rangle}{\partial t}$$

due to the interaction with optical phonons, is presented in Fig. 5 along with the rates w_1 and w_2 . The figure clearly shows that for $n = 10^{11} \text{ cm}^{-2}$ and $T_e \geq 500$ K the dominant contribution in the energy balance equation is due to the energy flux carried away by hot electrons emitted

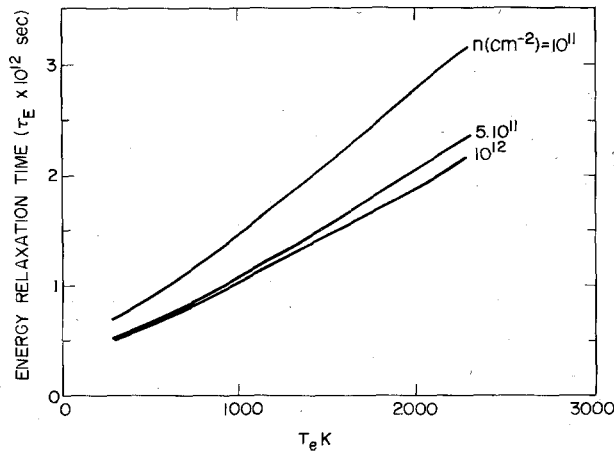


Fig. 4. Energy relaxation time τ_E (due to the interaction with polar optical phonons) defined by (27) and calculated as a function of the electron temperature T_e for different channel concentrations n .

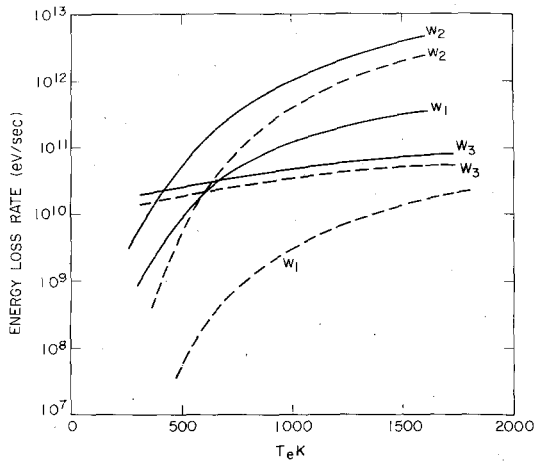


Fig. 5. Summary of the energy loss rates per one channel electron for different processes discussed in this work. Solid lines correspond to $n = 10^{12} \text{ cm}^{-2}$ and dashed lines to $n = 10^{11} \text{ cm}^{-2}$.

over the barrier into the collector. At higher concentrations this contribution becomes dominant even at lower T_e .

It should be noted that in the absence of real-space transfer (say, if the barrier had "infinite" height, so that both w_1 and w_2 terms would vanish) the optical-phonon term w_3 alone can compensate the input power $I \cdot F$ only for fields $F \lesssim 4 \text{ kV/cm}$. At higher fields our approximation in this situation would lead to a runaway breakdown process. Clearly, the reason for such a behavior is associated with the fact that we had neglected the transfer into heavy-mass valleys, which would reduce the rate at which the $I \cdot F$ term increases with the field. We believe, however, that in a $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ CHINT/NERFET structure under consideration the neglect of momentum-space transfer processes may be justified for the following reason. In this structure, for $x \approx 0.4$, the barrier height Φ is approximately equal to the intervalley Γ -L energy separation. The transferred L electrons see virtually no

barrier for real-space transfer and their collection efficiency is near unity (the situation with those electrons is quite analogous to that with minority carriers in a bipolar transistor). The energy carried away by an L electron is of the same order as that for Γ electrons. Therefore, this process (transfer into a satellite valley with subsequent diffusion across the barrier) is just another real-space transfer channel, and its contribution to the balance equation is of the same form as w_2 . Taking account of satellite valleys would, therefore, lead to a slight increase of the w_2 rate compared to that calculated above.

C. Summary of the Model

Within the assumed approximation of a uniform electric-field F along the channel, the energy balance equation is of the form

$$w_1 - w_2 - w_3 + I \cdot F/n = 0 \quad (28)$$

where $I = I(x)$ is the channel current per unit width of the channel

$$I(x) = env_d. \quad (29)$$

The drift velocity of electrons in the channel is assumed in the form

$$v_d = \begin{cases} \mu \frac{V_{SD}}{L} & \text{if } \mu \frac{V_{SD}}{L} \leq v_{sat} \\ v_{sat} & \text{if } \mu \frac{V_{SD}}{L} \geq v_{sat} \end{cases} \quad (30)$$

where μ is the mobility and v_{sat} the saturation velocity of channel electrons. Inasmuch as all terms in (28) depend only on the local concentration $n(x)$ and the electron temperature $T_e(x)$, the latter can be expressed as an implicit function of n . Therefore the density of the collector current J also can be regarded as a function of the concentration.

The current continuity equation is of the form

$$\frac{\partial I(x)}{\partial x} = -J(n). \quad (31)$$

Combined with (29), this gives a differential equation for the variation of the channel concentration

$$\frac{dn}{dx} = -\frac{J(n)}{ev_d}. \quad (32)$$

Integration of (32) determines $n(x)$

$$ev_d \int_{n(x)}^{n(0)} \frac{dn}{J(n)} = x. \quad (33)$$

For any value of the concentration near the source ($x = 0$), given by (1), the channel concentration at the drain $n(L)$ is determined by the equation

$$ev_d \int_{n(L)}^{n(0)} \frac{dn}{J(n)} = L \quad (34)$$

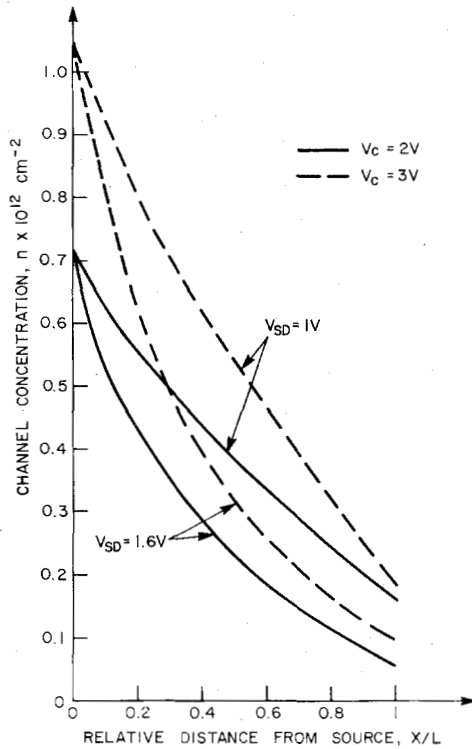


Fig. 6. Variation of the channel concentration $n(x)$ for different drain (V_{SD}) and collector (V_C) voltages. Note that at a given heating voltage V_{SD} the concentration decay along the channel is faster for higher V_C .

where L is the channel length. Thus, the drain current is given by

$$I_D = ev_d n(L) \quad (35)$$

and the collector current by

$$I_C = ev_d [n(0) - n(L)]. \quad (36)$$

The above procedure can, of course, be implemented only numerically. To determine the dependence $T_e(n)$, we solve (28) for T_e at a fixed n . Thus determined electron temperature is used to calculate $J(n)$ and the latter is substituted in the integrand of (33). Integration is then performed with ever decreasing lower limit—until the integral becomes equal to $Liev_d$. At this point, (34) is satisfied, thus determining the concentration $n(L)$.

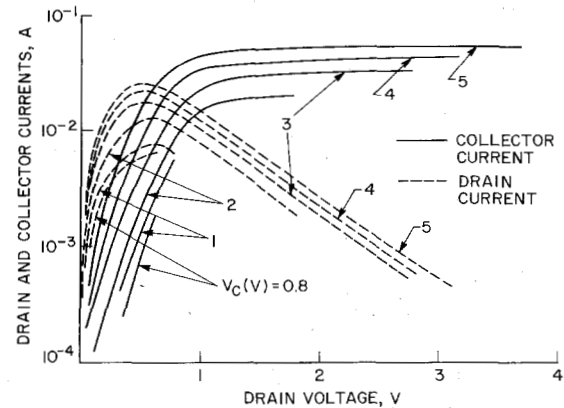
III. RESULTS

Physical parameters assumed in our calculation are presented in Table I.

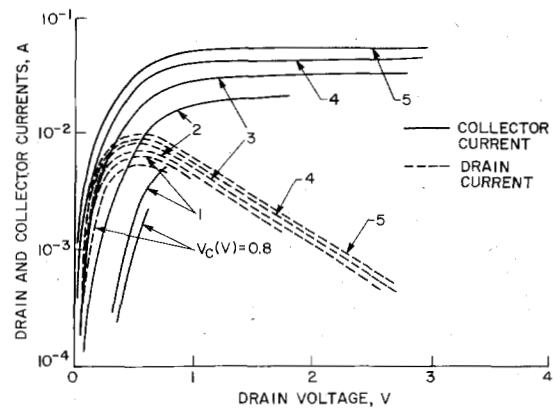
Fig. 6 shows the variation of electron concentration along the channel at two exemplary bias configurations. At high collector voltages ($V_C \geq 4$ V) the concentration near drain becomes so low ($\leq 10^{11}$ cm $^{-2}$) that one would have to consider diffusion processes in the channel and our uniform-field approximation would certainly break down. Moreover, at low concentrations our basic assumption of an electron temperature parameter T_e describing the entire hot-electron distribution function is questionable because the electron-electron scattering becomes

TABLE I

| | |
|-------------------------------|-------------------------------------|
| $T = 300$ K | lattice temperature |
| $d = 2 \cdot 10^{-5}$ cm | barrier thickness |
| $L = 2 \cdot 10^{-4}$ cm | channel length |
| $W = 10^{-2}$ cm | channel width |
| $\epsilon_0 = 12.85$ | static permittivity of GaAs |
| $\epsilon_\infty = 10.9$ | high-frequency permittivity of GaAs |
| $\Phi = 0.3$ eV | barrier height |
| $\hbar\omega_0 = 35.2$ meV | optical phonon energy |
| $m = 0.067 m_0$ | electron effective mass |
| $v_{sat} = 2 \cdot 10^7$ cm/s | saturation velocity |
| $E_0 = 5.95$ kV/cm | optical-phonon coupling constant |



(a)



(b)

Fig. 7. Current-voltage characteristics of a model CHINT/NERFET device with parameters assumed as in Table I, calculated (a) without and (b) with inclusion of the tunneling component of the injection current.

weak. Accordingly, we have left this range out of consideration. Also, in order not to run into a channel pinch-off effect, which would ruin our uniform-field assumption, we have continued the calculations only until $V_{SD} \leq V_C - 0.2$ V.

It is evident from Fig. 6 that higher V_C brings about a more rapid decay of the concentration along the channel at a given heating voltage V_{SD} . This effect is due to the enhancement of the charge-injection current J by the barrier lowering. It should be noted, however, that our theory does not describe the effect evident in Fig. 1(a)—the decrease of I_C with higher V_C at a sufficiently high fixed

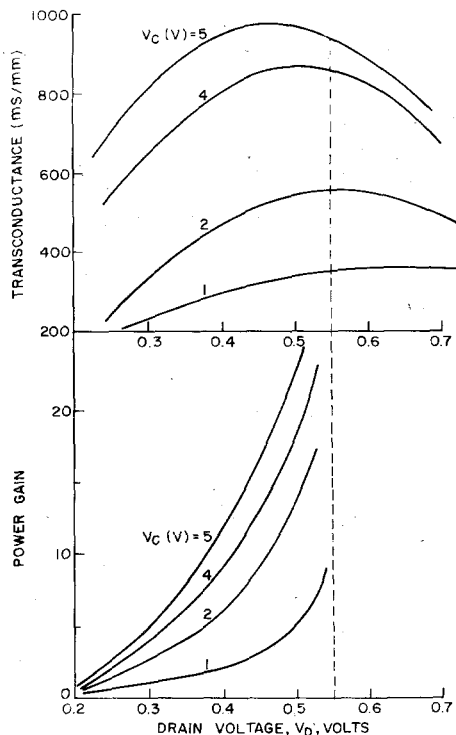


Fig. 8. Calculated transconductance and power gain in a model CHINT structure (parameters, as in Table I) at different bias configurations. The dashed line indicates the onset of NDR in the source-drain circuit (approximately taken to be independent of V_C).

V_{SD} . In the plot of Fig. 6 this effect would manifest itself by an intersection of curves corresponding to the same V_{SD} but different V_C .

The calculated current-voltage characteristics of a CHINT/NERFET device with assumed parameters (Table I) are displayed in Fig. 7. Characteristics calculated with and without the inclusion of tunneling are shown in Fig. 7(b) and 7(a), respectively. It is evident that tunneling improves significantly the efficiency of CHINT operation, since the output current (I_C) becomes a steeper function of the input heating voltage V_{SD} . In other words, tunneling improves the transconductance of CHINT.

Fig. 8 shows the transconductance $g_m \equiv (\partial I_C / \partial V_{SD})|_{V_C}$ and the power gain in CHINT, calculated for different bias configurations. As expected, g_m increases with the collector voltage V_C . As a function of V_{SD} , the transconductance peaks near the onset of the NDR in the drain circuit. At higher voltages g_m decreases—due to the saturation of the collector current. Peak values of the transconductance and the qualitative functional dependences $g_m(V_{SD}, V_C)$ are in agreement with the experimental data [6]. We remark, however, that the experimental curves $g_m(V_{SD})$ exhibit a sharper rise than that seen in the calculated dependences. The calculated power gain in our model CHINT structure is shown in the bottom of Fig. 8. Note that as a function of V_{SD} the gain $\rightarrow \infty$ when the device is biased into the NDR region. Physically, in this region the system goes over from the small-signal amplification regime into the regime of spontaneous oscillations. The calculated behav-

ior of the power gain in the vicinity of the NDR threshold (the latter is shown in Fig. 8 by a dashed line) fits very well the experimental observation [6].

The calculated characteristics also describe a strong NDR in the drain circuit of the device, i.e., the operation of a NERFET. Our theory predicts a current drop by a factor 30 in the bias range considered, which compares well with the typical experimental data at room temperature (Fig. 1(b)). Note that in the NDR region of the NERFET the CHINT collector current saturates, and therefore the operation range of CHINT occurs prior to the onset of the NDR, as discussed above. Fig. 7(b) also shows that even near the peak of I_D the collector current dominates ($I_C \gg I_D$)—which implies a high current gain. These theoretical results are in good agreement with the experimental observations. Comparison of the curves in Figs. 7 and 1(b) shows that inclusion of the tunneling component of the injection current significantly improves the agreement between the theory and experiment.

IV. CONCLUSIONS

Comparing our results with the objectives set out in the Introduction, we can say that some of the major features we sought to explain are adequately described by the theory. Our calculations correctly predict the existence of a strong NDR in the drain circuit accompanied by a rapid increase and subsequent saturation of the collector current. Moreover, the theory describes the experimental fact that at high enough heating voltages V_{SD} the collector current substantially exceeds the drain current (inclusion of the tunneling component of the injection current turned out to be crucial in establishing this fact, which is very important for obtaining a high current gain in the operation of CHINT).

However two important experimental features were not captured by our model. Firstly, the theory does not describe the observed saturation of the drain current from below—after the NDR region. Secondly, although the theory correctly shows the increase in the peak drain current with higher collector voltage V_C (which is simply a back-gate field effect), it does not describe the salient (and quite puzzling) experimental feature of the negative transconductance effect in the saturation region. This effect (drop in the saturated I_D with increasing V_C) is clearly evident in Fig. 1(b). We believe that the major deficiency of the present theory consists in the assumption of a uniform field in the channel and the neglect of the diffusion component of the channel current.

The actual profile of the electric field in the channel is nonuniform and qualitatively this can be expected to affect the picture as follows. At low heating voltages V_{SD} , when $I_C \lesssim I_D$, the situation is close to that in an ordinary FET—the collector playing the role of a gate. In this case the field is highest near the drain and is approximately proportional to the gradient of the channel concentration dn/dx (since, in a long-channel approximation, $en(x) = C[V_C - V(x)]$). However, as V_{SD} rises accompanied by the

hot-electron emission, the electric field in the channel must redistribute so as to support the larger channel current $I(0)$ near the source—since the latter is given exactly by the sum of the drain and collector currents. (Obviously, the channel current must vary from $I(0) = I_C + I_D$ to $I(L) = I_D$.) The required redistribution of the field toward the source will lead to a still sharper rise in I_C and a further drop of I_D . This “run-away” process is stabilized when the concentration in the channel region adjacent to the drain becomes so low that most of the further increase in V_{SD} will drop in that depleted region. At this point the field distribution will again resemble that in a field-effect transistor.

Another process not included in the present theory is the dynamical screening effect, arising due to the accumulation of carriers drifting downhill in the collector barrier [3]. This effect becomes important when the collector current exceeds the drain current, i.e., when the areal density of the injected mobile carriers becomes comparable to that of channel electrons. We believe that inclusion of this effect may account for the experimentally observed negative transconductance.

APPENDIX

ENERGY BALANCE EQUATION IN THE PRESENCE OF REAL SPACE TRANSFER

Our derivation of the balance equation is based on the assumption that the energy flux Q_x along the channel equals the mean electron energy at a given point x multiplied by the electron particle flux $I(x)/e$. This assumption is valid to within a constant factor of order unity, which depends on the form of the energy distribution function, as well as on the energy dependence of the electron momentum relaxation time τ_m . Indeed, the linear current density I in the channel is given by

$$I(x) = F_x \frac{e^2 n}{m} \langle \tau_m \rangle \quad (A1)$$

where in the two-dimensional case

$$\langle \tau_m \rangle = - \frac{\int E_k \tau(E_k) (\partial f_0 / \partial E_k) dE_k}{\int f_0(E_k) dE_k} \quad (A2)$$

and $f_0(E_k)$ is the isotropic part of the distribution function. Similarly, the energy flux along the channel can be written in the form

$$Q_x = \frac{F_x e n}{m} \frac{\int E_k^2 \tau(E_k) (\partial f_0 / \partial E_k) dE_k}{\int f_0(E_k) dE_k} \quad (A3)$$

Throughout this work, the function f_0 is taken in the electron-temperature approximation, $f_0 \propto \exp(-E_k/kT_e)$. Assuming that $\tau_m(E_k) = \tau_0(E_k/kT_e)^{-\nu}$ and comparing (A2) with (A3), we find

$$I(x) = \frac{e^2 \tau_0}{m} \Gamma(2 - \nu) F_x$$

$$Q_x = -kT_e \frac{e \tau_0}{m} \Gamma(3 - \nu) F_x$$

$$\langle E \rangle = kT_e \quad (A4)$$

Thus we have

$$Q_x = -\langle E \rangle \frac{I(x)}{e} (2 - \nu) \quad (A5)$$

On the other hand, for a degenerate distribution $\langle E \rangle = E_F/2$, and from (A2) and (A3) one obtains

$$Q_x = -\langle E \rangle \frac{I(x)}{e} \quad (A6)$$

In this work the difference between (A5) and (A6) was neglected by setting the numerical coefficient in (A5) equal to unity.

The energy conservation law can be written in the form

$$\frac{\partial Q}{\partial t} = -n(w_2 + w_3) - \frac{\partial Q_x}{\partial x} + I \cdot F = 0 \quad (A7)$$

Substituting (A6) into (A7), using (31), and neglecting the Thomson heat transport, i.e., taking

$$\frac{\partial \ln \langle E \rangle}{\partial x} \ll \frac{\partial \ln I}{\partial x}$$

we obtain the balance equation in the form (28) with w_1 given by (9). It should be mentioned that, as seen from Fig. 5, the actual magnitude of the term w_1 is small, $w_1 \ll w_2$ in the range of T_e of interest, so that its introduction is more important for logical consistency than for accuracy of the calculations.

REFERENCES

- [1] K. Hess, H. Morkoç, H. Shichijo, and B. G. Streetman, “Negative differential resistance through real space electron transfer,” *Appl. Phys. Lett.*, vol. 35, pp. 469–471, 1979.
- [2] A. Kastalsky and S. Luryi, “Novel real-space hot-electron transfer devices,” *IEEE Electron Device Lett.*, vol. EDL-4, pp. 334–336, 1983.
- [3] S. Luryi and A. Kastalsky, “Hot electron injection devices,” *Superlatt. Microstructures*, vol. 1, pp. 389–400, 1985.
- [4] —, “Hot-electron transport in heterostructure devices,” *Physica*, vol. 134B, pp. 453–465, 1985.
- [5] A. Kastalsky, R. Bhat, W. K. Chan, and M. Koza, “Negative resistance field-effect transistor,” *Solid-State Electron.*, to be published.
- [6] A. Kastalsky, J. H. Abeles, R. Bhat, W. K. Chan, and M. Koza, “High-frequency amplification and generation in charge injection devices,” *Appl. Phys. Lett.*, vol. 48, pp. 71–73, 1986.
- [7] T. Ando, A. B. Fowler, and F. Stern, “Electronic properties of two-dimensional systems,” *Rev. Mod. Phys.*, vol. 54, pp. 437–672, 1982.
- [8] F. Stern and S. Das Sarma, “Electron energy levels in GaAs-Al_xGa_{1-x}As heterojunctions,” *Phys. Rev. B*, vol. 30, pp. 840–848, 1984.
- [9] A. A. Grinberg, “New variational solution for the lowest level of the two-dimensional electron gas,” *Phys. Rev. B*, vol. 32, pp. 4028–4033, 1985.
- [10] F. Stern and W. E. Howard, “Properties of semiconductor surface inversion layers in the electric quantum limit,” *Phys. Rev.*, vol. 163, pp. 816–835, 1967.

- [11] A. A. Grinberg, "Thermionic emission in heterosystems with different effective electronic masses," *Phys. Rev. B*, 1986, to be published.
- [12] A. A. Grinberg, M. S. Shur, R. J. Fisher, and H. Morkoç, "An investigation of the effect of graded layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1758-1765, 1984.
- [13] R. Stratton, "Theory of field emission from semiconductors," *Phys. Rev.*, vol. 125, pp. 67-82, 1961.
- [14] F. F. Fang and W. E. Howard, "Negative field-effect mobility on (100) Si surfaces," *Phys. Rev. Lett.*, vol. 16, p. 797, 1966.
- [15] K. Hess, "Impurity and phonon scattering in layered structures," *Appl. Phys. Lett.*, vol. 35, p. 484, 1979.
- [16] B. Vinter, "Phonon-limited mobility in GaAlAs/GaAs heterostructures," *Appl. Phys. Lett.*, vol. 45, p. 581, 1984.
- [17] P. J. Price, "Polar optical-mode scattering for an ideal quantum-well heterostructure," *Phys. Rev. B*, vol. 30, pp. 2234-2235, 1984.
- [18] —, "Two-dimensional electron transport in semiconductor layers—I. Phonon scattering," *Ann. Phys. (NY)* vol. 133, p. 217, 1981.

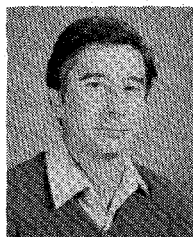
*



Anatoly A. Grinberg received the Ph.D. degree in 1961 from the Lebedev Institute of Physics, Moscow, U.S.S.R., in 1961 and the Dr.Sci. degree in 1969 from the A. F. Ioffe Institute of Physics and Technology, Leningrad, U.S.S.R.

From 1958 to 1980, he worked at the Ioffe Institute as a Senior Staff Scientist. In 1980 he came to the United States and joined the Department of Physics of New York University. Beginning in 1982, he has been associated with the Department of Electrical Engineering of the University of

Minnesota, and since 1985 also with Bell Laboratories. His research has included transport properties as well as linear and nonlinear optics of semiconductors, properties of the two-dimensional electron gas, and high-speed devices. His work has appeared in over 80 publications in physics and engineering journals.



Alexander Kastalsky received an undergraduate degree from the Polytechnical Institute, Leningrad, U.S.S.R., in 1962 and the Ph.D. degree from the Ioffe Physics-Technical Institute, Leningrad, in 1969.

From 1963 until 1980, he worked at the Ioffe Physics-Technical Institute where his field of interest covered the electrical and optical properties of semiconductors. Since 1981, he has been a member of the technical staff at Bell Laboratories, and since 1985 at Bell Communications Research, Inc., Murray Hill, NJ. His work is devoted to the electron properties of two-dimensional systems and the physics of high-speed semiconductor devices.

*



Serge Luryi (M'81-SM'85) received an undergraduate degree in physics in 1971 from the Leningrad State University, U.S.S.R. In 1973, he emigrated to Canada, where he received the M.Sc. and Ph.D. degrees in theoretical physics from the University of Toronto, in 1975 and 1978, respectively. His graduate and postdoctoral research was devoted to intermolecular interactions and lattice vibrations in solid H₂.

Since 1980, he has been a member of the technical staff at AT&T Bell Laboratories, Murray

Hill, NJ, where he is currently a group supervisor in the Advanced VLSI Development Laboratory. He has published on topics ranging from properties of polysilicon gates to the theory of the quantized Hall effect, altogether over 50 papers. His main research interests are in the physics of exploratory semiconductor devices. He has filed 10 patents on novel high-speed and optoelectronic devices as well as new methods of epitaxial crystal growth.

Dr. Luryi is a member of the American Physical Society.