HOT ELECTRON INJECTION DEVICES

Serge Luryi[†] and Alexander Kastalsky[§]

[†]AT&T Bell Laboratories, Murray Hill, N.J. 07974

[§]Bell Communications Research, Murray Hill, N.J. 07974

(Received 27 November, 1984)

Recently we proposed a novel three-terminal structure which employs hot-electron transfer between two conducting layers separated by a potential barrier and contacted individually. Several new device concepts based on this principle have been demonstrated experimentally. In the present work we review these devices — a charge injection transistor, a negative-resistance microwave oscillator controlled by a third terminal, and a memory element — emphasizing the physics of their operation.

1. *Introduction*

The effect of real-space hot-electron transfer in multilayer semiconductor structures was first suggested by Hess et al.[1] Guided by an analogy with the momentum-space intervalley transfer (Gunn effect) they proposed that in layered heterostructures high-energy (hot) electrons, heated by an applied electric field, can move between adjacent layers causing enhancement of the mobile charge concentration in one set of layers and depletion in the other. When the layers had different mobilities, the real-space transfer was predicted to result in a negative differential resistance (NDR) in the two-terminal circuit.

Recently, we proposed[2] a novel three-terminal structure in which the effect of real-space transfer gives rise to charge injection between two conducting layers separated by a potential barrier and contacted individually. The idea of our structure can be illustrated by the glow-cathode analogy, displayed in Fig. 1. In a vacuum diode the anode current as a function of the anode voltage saturates at a value determined by the cathode work function and the temperature. One can think of a hypothetical amplifier in which an input circuit controls the cathode temperature and thus the output current, but that would be a slow device. In our structure the input circuit controls *the electron temperature* T_e which, unlike the temperature of a material, can be rapidly varied in one of the conducting layers ("the channel"), resulting in an efficient charge injection into the other layer. Based on this principle, we suggested several new device concepts, most of which by now have been

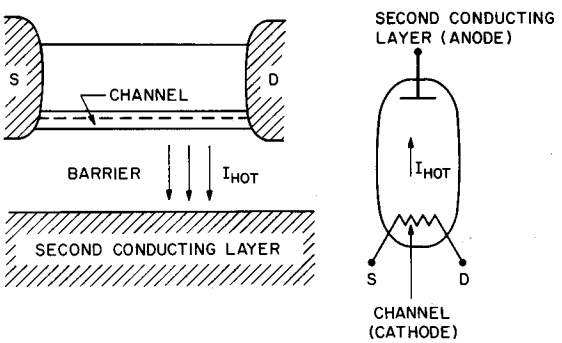


Fig. 1 Illustration of the principle of our charge injection structure. The channel serves as a cathode whose effective electron temperature is controlled by the source-to-drain field. The second conducting layer separated by a potential barrier serves as an anode and is biased positively.

demonstrated experimentally.[3–5] In the present work we shall review these devices: the charge injection transistor (CHINT),[3] the negative resistance field-effect transistor (NERFET),[4] and the hot-electron erasable programmable random access memory element[5] (HE^{2}PRAM) — emphasizing the physics of their operation, as we see it now, after one year of experimentation. It should be said that much of our understanding remains on a qualitative level.

0749-6036/85/050389+12 $02.00/0

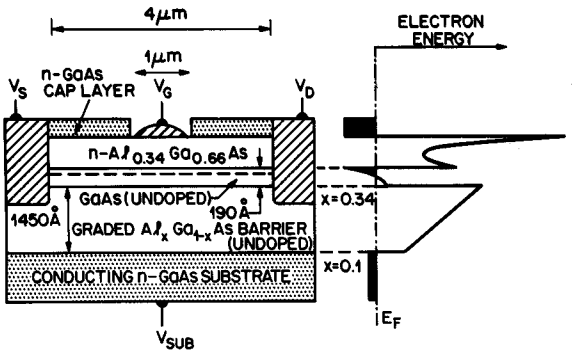


Fig. 2 Device structure and energy diagram.

2. *Device Structure*

Cross-section of the typical device used in most of our experiments is shown in Fig. 2. The device has been implemented in an AlGaAs/GaAs heterojunction structure. Details of its MBE growth and processing can be found in ref. 3. One of the conducting layers represents an FET channel, the other a heavily doped GaAs substrate. The device is essentially three-terminal with the electrodes labeled source (S), drain (D), and substrate (SUB). Application of a source-to-drain voltage V_{SD} leads to a heating of channel electrodes and charge injection into the substrate over the AlGaAs potential barrier. The aluminum content in the barrier was graded from x=0.34 near the channel to x=0.1 near the substrate. This triangular or, more precisely, trapezoidal form of the barrier is not essential for most applications. We have done experiments (to be reported elsewhere) with a similar structure having a rectangular-shaped barrier.

The fourth (gate) electrode plays an auxilliary role in our devices. Its purpose is to concentrate the source-drain electric field in a 1μm region under the gate notch. The gate bias can be used to fine-tune the operation of hot-electron injection devices by modifying the equilibrium carrier concentration in the channel. Devices can also be operated without gate metallization, the field concentration being still provided by the notch etched into the AlGaAs layer enough to produce a slight channel depletion.

A critical step in manufacturing the CHINT/NERFET structure is to provide ohmic contacts to the 2-dimensional electron gas in the channel, while keeping them electrically insulated from the conducting substrate. This is no easy task since the two conducting layers are separated by less than 0.15μm, cf. Fig. 2. Our solution was to use Au/Ge-Ag-Au contacts which appear to give an abrupt edge of alloy penetration into the semiconductor material. The resultant diode characteristics between the two conducting layers are shown in Fig. 3. The curves are similar to those one would expect for an asymmetric triangular-barrier diode[6] of an approximate barrier height 0.3 eV. Quality of the contact is demonstrated by the absence of any ohmic leakage in the reverse-bias characteristic, Fig. 3a, down to picoampere currents. Large ideality factor $n_1 \approx 13$ for thermionic emission in the reverse direction (i.e., when the substrate is biased positively) ensures that the barrier can be regarded as blocking for $V_{SUB} \lesssim 3$ V (at T=77 K). On the other hand, at 300 K one has to worry about the reverse-bias leakage already at $V_{SUB} \sim 1$ V.[3]

Analysis of the forward-bias characteristic, Fig. 3b, suggests the existence of acceptor-like traps in the nominally undoped AlGaAs barrier layer. Indeed, for an ideal triangular-barrier of the form shown in Fig. 2 one would expect an ideality factor n~1 in the forward direction (neglecting the channel thickness), while we have $n_2=1.5$ at low currents. Moreover, the inverse ideality factors $\ell_1 \equiv 1/n_1$ and $\ell_2 \equiv 1/n_2$ do not add up to unity, as would be the case for an ideal asymmetric triangular barrier.[7] We believe that after an initial forward biasing the barrier traps are charged and the shape of the barrier changes as shown in the insert to Fig. 3b. The measured barrier height $\Phi=0.32$ eV is somewhat larger than what one would expect for a GaAs/$Al_{0.34}Ga_{0.66}As$ heterojunction (assuming the recently proposed[8] conduction-band discontinuity rule $\Delta E_c \approx 0.6\,\Delta E_g$, the expected $\Phi=0.25$ eV). The difference $\Delta\Phi \approx 0.07$ eV can be attributed to the space-charge potential of electrons trapped in the barrier layer, and the curve in Fig. 3b can be fitted by assuming a uniform trap density $\sim 1.5\times10^{16}$ cm^{-3}. It should be said that this interpretation is only tentative. We are not even certain at this time whether most of the thermionic current flows in the channel-barrier region or under the alloyed contacts. Moreover, the locus of this current may actually be different for the forward and reverse bias conditions.

3. *Electron Temperature in the Channel*

We are certainly not yet at that stage of understanding where we could present a quantitative analysis of what happens in our structure when voltage is applied between the surface electrodes S and D. A substantial amount of theoretical *and* experimental work is required before such an analysis becomes possible. Electron heating in the channel involves an interplay between the real-space and the momentum-space transfers, formation of domains of high electric field, redistribution of electrons both laterally and vertically within the channel, transient "ballistic" transport, and various quantum effects due to the two-dimensional confinement. Our treatment of most of these "complications" will not go beyond naming them.

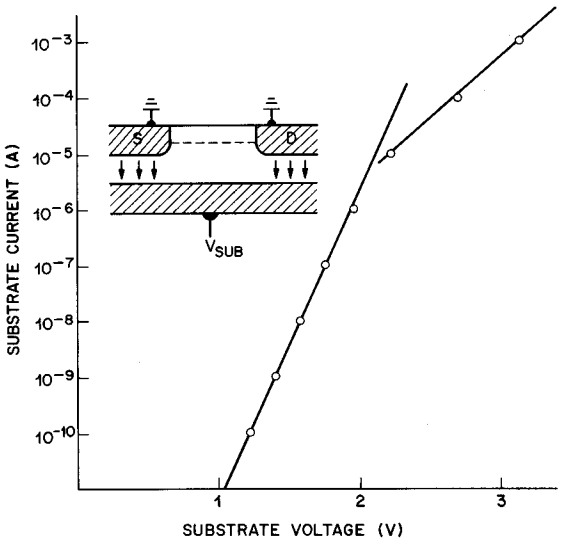
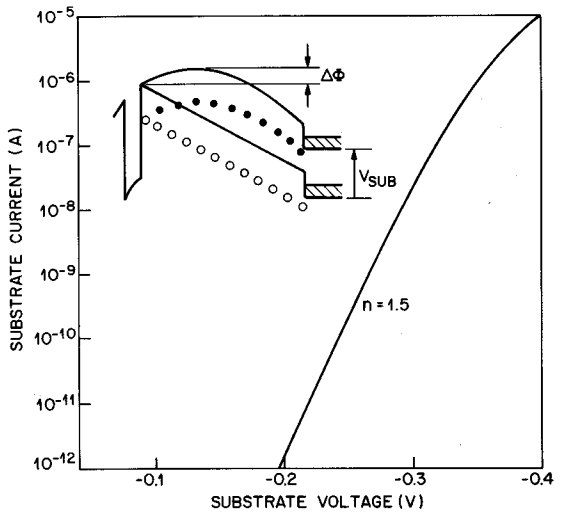


Fig. 3 Diode characteristics between the S-D (tied together) and the substrate terminals at 77 K. a) Reverse-bias characteristic ($V_{SUB} > 0$); b) forward characteristic.



Our qualitative picture of hot-electron injection is based on the assumption that the relevant non-equilibrium properties of the ensemble of channel electrons can be described by postulating a local electron temperature $T_e(x)$ which may be a function of the channel position x between S and D. The areal density J(x) of the hot-electron current is then given by the Richardson-like expression

$$J = \frac{\sigma}{\Delta} v(T_e)\, e^{-\Phi/kT_e}, \qquad (1)$$

where $\sigma(x)$ is the charge density per unit area, Δ the channel thickness, and

$$v(T_e) = (kT_e/2\pi m^*)^{1/2} \qquad (2)$$

is the mean velocity of hot electrons in a given direction. Equation (1) assumes the validity of the electron-temperature approximation in the entire range of the electron energy distribution. This approximation is reasonable even in the high-energy tail of the distribution because of the high ($N \gtrsim 10^{17}$ cm^{-3}) electron concentration in the channel. Indeed, it is well known[9] from studies of bulk hot-electron transport that at such concentrations establishment of the electron temperature is provided by electron-electron (e-e) collisions.

Typical hot-electron injection characteristics taken at 77 K are shown in Fig. 4. One of the surface electrodes (labeled D) was grounded. The heating voltage V_{SD} of both polarities was applied to the other electrode. The substrate was kept at a fixed positive voltage V_{SUB}. We see that the substrate current I_{SUB} exhibits a sharp minimum when $V_{SD} \to 0$. Consider

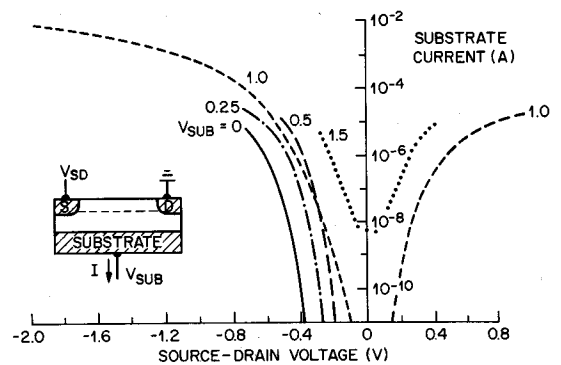


Fig. 4 Hot-electron injection at 77 K for different substrate biases.

the curve corresponding to $V_{SUB} = 1$ V. As the heating voltage increases *in either polarity* the current rises by many orders of magnitude and its polarity corresponds to electrons injected into the substrate. This is a direct evidence of the hot-electron nature of I_{SUB}. An analysis of the characteristics in Fig. 4 allowed us to determine[3] the electron temperature T_e as a function of the heating voltage V_{SD}. For low V_{SUB} we found (Fig. 5) that $T_e \propto (V_{SD})^2$ and at a fixed V_{SD}, the T_e increases with V_{SUB}. It should be born in mind that this determination was based on rather crude assumptions [neglect of the variations in the pre-exponential factor and in the barrier height Φ in eq. (1)], so that the accuracy of our estimates for T_e is not better than 20%.

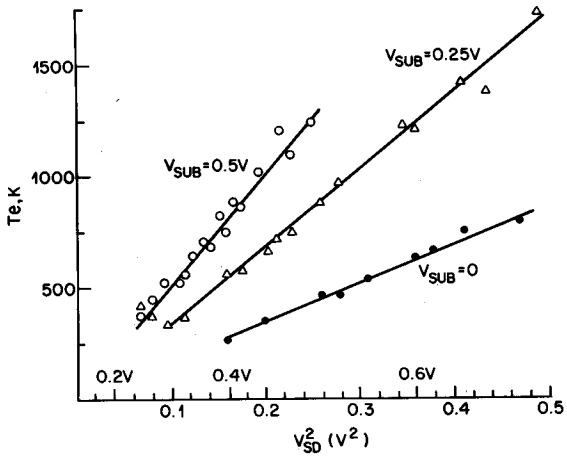


Fig. 5 Electron temperature T_e as a function of the heating voltage V_{SD}, for several *low* values of V_{SUB}. For higher V_{SUB} these simple dependences no longer hold.

At low substrate biases, the hot-electron injection is observed only when $V_{SD}<0$. At positive V_{SD} the effect is masked by a "cold" electron current of opposite polarity thermionically emitted from the substrate into the S electrode (which actually is a drain now). In this case the thermionic emission in the forward-biased triangular barrier diode occurs before the onset of an efficient hot-electron injection. At a sufficiently large $V_{SD}>0$ the cold thermiomic current always wins and the total substrate current changes polarity.[3] It would be very interesting to observe a competition between these currents when $V_{SUB}=0$. In that case, even a small range of $V_{SD}>0$ accompanied by charge injection into the substrate would imply a spectacular effect of "absolute negative resistance". Despite considerable efforts, we have failed (so far) to detect the hot-electron injection at $V_{SD}>0$ and $V_{SUB}=0$. Assuming that most of the hot-electron injection occurs near the drain and using (1), it is easy to estimate the conditions for this effect to occur:

$$(T_e/T)\ (1-V_{SD}/\Phi) > 1 \qquad (3)$$

Clearly, this condition favors lower lattice temperatures and higher heating efficiencies. In our experiments we were not able to achieve a sufficiently high T_e at low heating voltages ($V_{SD}<\Phi$).

The curves in Fig. 4 were measured in the configuration with unbiased gate electrode (so that we could register the injection current under heating voltages of opposite polarities). Figure 6 shows the analogous curves for the case when the gate is biased relative to the source. The hot-electron current I_{SUB} (solid lines) is measured simultaneously with the drain

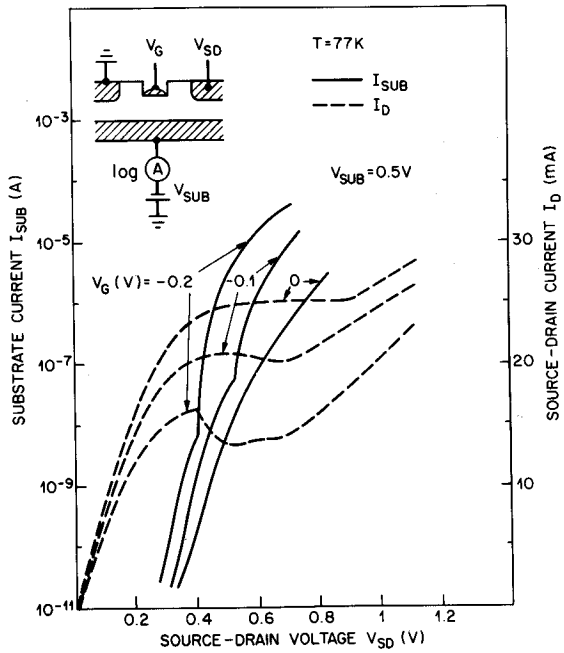


Fig. 6 Injection curves at 77 K with a biased gate electrode. The kinks represent an evidence for the high-field domain formation in the channel. Dashed lines correspond to the drain current I_D measured simultaneously with the injection current I_{SUB}.

current I_{SD} (broken lines). We see that for the gate biases $V_G \lesssim -0.1$ V the injection curves have a kink (indicating a discontinuous process), which occurs exactly at the onset of an NDR in the drain channel. This feature reveals an abrupt transition of the channel into a new state in which the electric field under the gate is highly nonuniform. This phase transition (formation of a high-field domain) accounts for the observed discontinuous rise in T_c. The role of the negative V_G is to provide the initial field concentration within the 1μm region under the gate. Note that the drop in I_{SD} is several orders of magnitude higher than the accompanying increase in I_{SUB}. Therefore, the observed NDR in this case cannot be explained by a real-space transfer. It is probable that the responsible mechanism here is the momentum-space transfer (Gunn effect) in the channel.

Figure 7 shows our results for hot electron injection at 4.2 K in the presence of a magnetic field **H** perpendicular to the sample's surface. Even a moderate H~1T shifts the threshold for hot-electron injection by about 0.3 V. It is clear that the magnetic field inhibits electron heating. A theory of such an effect in the bulk[10] gives qualitatively similar

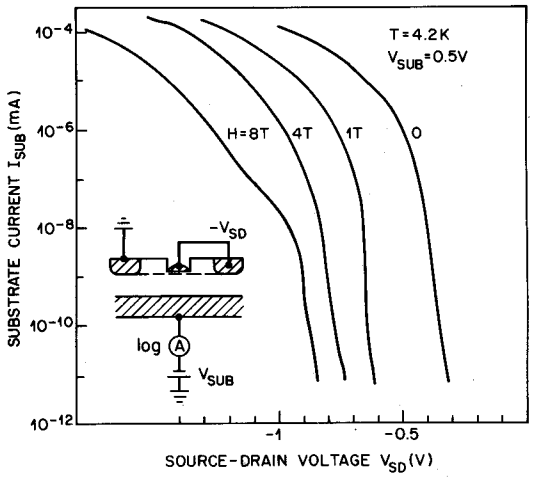


Fig. 7 Hot-electron injection in the presence of a magnetic field transverse to the plane of the channel.

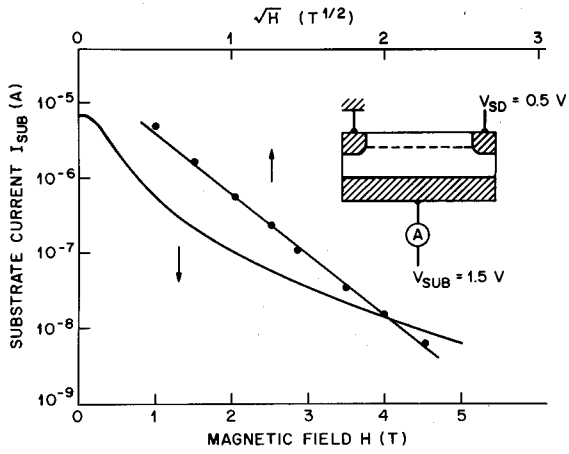


Fig. 8 Magnetic field dependence of the hot-electron injection at a fixed bias configuration (V_{SD}, V_{SUB}). In the region $H < 0.25$ T the injection current I_{SUB} varies little. The straight line refers to the upper horizontal axis and corresponds to log I_{SUB} plotted without change of scale against square root of the magnetic field in the region $H > 0.25$ T.

predictions. Figure 8 shows the magnetic field dependence of charge injection at a fixed voltage configuration $V_{SD}=0.5$ V, $V_{SUB}=1.5$ V. We see that as H is increased to 5T, the injection current drops by three orders of magnitude. Most of this drop occurs in the region $H \gtrsim 0.25$ T; at lower H we see almost no

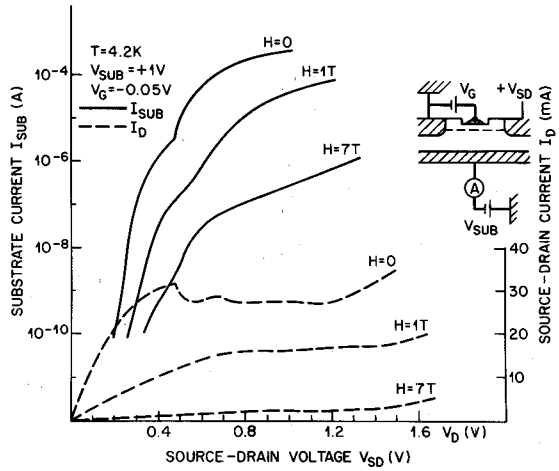


Fig. 9 Injection curves similar to those in Fig. 6 but in the presence of a magnetic field transverse to the plane of the channel.

variation in I_{SUB}. If we plot log I_{SUB} versus $\sqrt{H}$ for $H > 0.25$ T, we find a nearly perfect linear dependence over 3 decades in current. It is tempting to interpret this variation as arising mainly from the dependence $T_e(H)$ in the exponent of (1), and thus conclude that $T_e \propto H^{-1/2}$. It should be born in mind, however, that application of the magnetic field can also affect the electric field distribution in the channel — especially its concentration in a high-field domain. Consider our data in Fig. 9, obtained in a configuration similar to that of Fig. 6, but at the temperature of liquid helium rather than nitrogen. We see that application of a magnetic field suppresses the kink on the injection curve which we had attributed to a Gunn-effect transition. At the same time we see a disappearance of the NDR in the channel circuit and a substantial depression of the channel current by the magnetic field. On the other hand, as we shall see below (Sect. 5), the NDR characteristics in the regime of NERFET operation (high substrate currents) is *little* affected even by strong magnetic fields (up to 8T).

4. *The CHINT*

The charge injection transistor or CHINT is a direct solid-state analog of the hypothetical vacuum diode with controlled cathode temperature discussed above in connection with Fig. 1. It is a three-terminal device whose output (substrate) current I_{SUB} is a function of the input voltage V_{SD}. Increasing V_{SD} results in a higher T_e and an enhanced injection into the substrate - just like the anode current in a vacuum diode responds to increasing temperature of the cathode filament. By the physical principle involved, CHINT is

different from all previous three-terminal semiconductor devices — all of which could be classified in either of the two groups: *potential*-effect and *field*-effect transistors. In the first group (which includes the bipolar transistor and the so-called analog transistors, such as the static induction transistor,[11] the permeable base transistor,[12] and the thermionic emission transistor[13]) the transistor action results from modulating the height of a potential barrier by an input electrode. The second group, containing a great variety of FET's, employs the modulation of charge in a resistive channel due to the screening of an input field. This classification, of course, is quite relative (for example, MOSFET in its subthreshold regime should be regarded as a potential-effect device), since both the field-screening and the potential-modulation effects are at work in every transistor. Neither of these effects, however, constitutes the physical mechanism of CHINT which is based on charge emission over a barrier of fixed height due to a modulation of the electron temperature in the emitter.

Transistor characteristics of the CHINT have been studied in ref. 3 at 300 K and 77 K. A family of curves I_{SUB} versus V_{SUB} with the heating voltage V_{SD} as a parameter (drain grounded, negative source bias varied) is similar to the collector characteristics of a bipolar transistor in its common-base configuration.[15] Electrically, one can draw an analogy between the two devices: *source* ≡ emitter, *drain* ≡ base, and *substrate* ≡ collector, although the CHINT is entirely unipolar. The existence of power gain in CHINT at 77 K has been inferred from the static measurements.[3] The highest value of the mutual conductance,

$$g_m \equiv (\partial I_{SUB}/\partial V_{SD})_{V_{SUB}=\text{const}}, \qquad (4)$$

achieved to-date is about 300 mS/mm at 77 K. A peculiar feature of CHINT is the value of its common-base current gain α,

$$\alpha \equiv (\partial I_{SUB}/\partial I_S)_{V_{SUB}=\text{const}}, \qquad (5)$$

which can substantially exceed unity. Since the source, drain and substrate currents are related by Kirchhoff's law

$$I_{SUB} = I_S - I_D \qquad (6)$$

we can have $\alpha > 1$ only in the region of a negative differential resistance in the drain circuit. Interestingly, the early bipolar transistors also had an alpha greater than unity and their static emitter base characteristic exhibited an NDR.[16] However, this effect resulted from a parasitic positive feedback mechanism (the so-called p−n hook effect) and was not an intrinsic property of the ideal transistor. As will be discussed in the next Section, the strong NDR in the drain circuit is an inherent property of the CHINT-NERFET structure. A simultaneous plot of the drain and the substrate currents versus the heating voltage

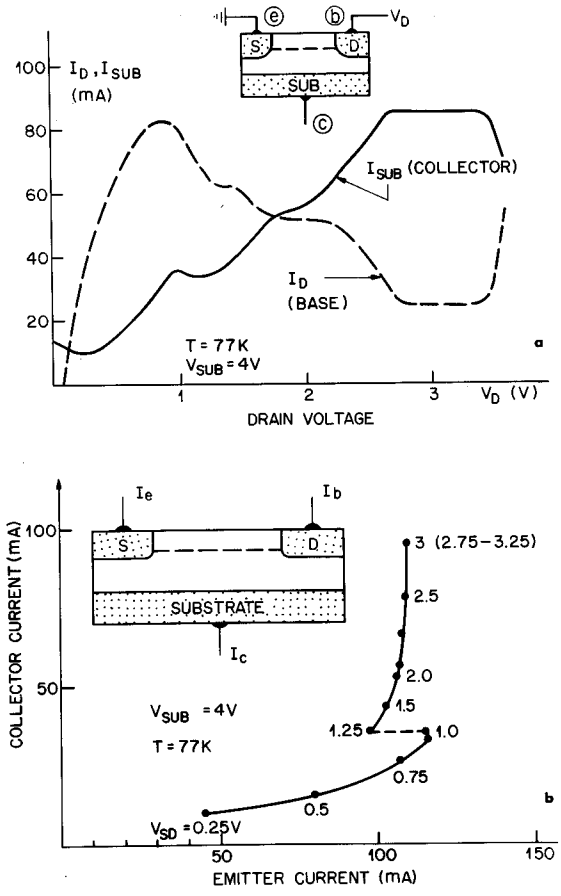


Fig. 10 Variation of currents in the common-drain configuration of CHINT at 77 K.
a) The substrate (I_{SUB}) and the drain (I_D) currents versus the heating voltage V_{SD} at fixed V_{SUB} = 4 V.
b) The dependence $I_{SUB}(I_S)$ at a fixed V_{SUB}, obtained from a) by eliminating V_{SD} and using (6).

V_{SD} is shown in Fig. 10a. We see that rise in I_{SUB} is accompanied by a drop in I_D so that the source current varies little in the NDR region. The curves in Fig. 10a give the dependence $I_{SUB}(I_S)$ at a fixed V_{SUB} in a parametric form, $I_{SUB}(V_{SD})$ and $I_S(V_{SD})$. This dependence is plotted explicitly in Fig. 10b by eliminating V_{SD} and using (6). Its slope gives the value of α which, as we see, can be very large in the NDR region. The $\alpha > 1$ effect is highly reproducible and being an inherent property of the device it does not impede its speed of operation (the fundamental limitations on the speed of CHINT will be discussed in Sec. 6).

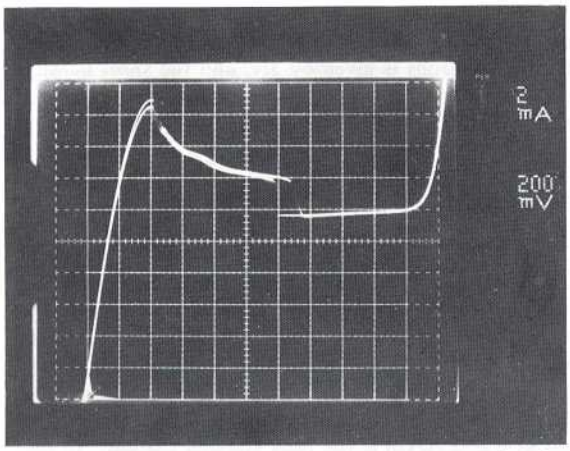


Fig. 11 NERFET at room temperature; the I_D versus V_{SD} characteristic at $V_{SUB}=2.4$ V.

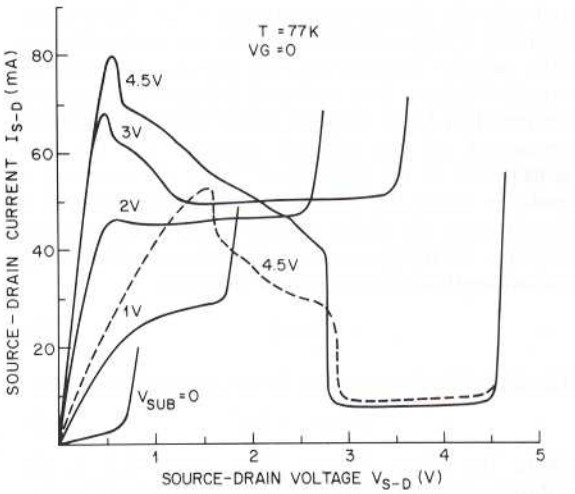


Fig. 12 Typical IV characteristics in the channel circuit of the NERFET at 77 K and different substrate biases. Dashed curve was obtained at 4.2 K and in a transverse magnetic field H = 5 T.

The dependence $I_{SUB}(I_S)$ in Fig. 10b clearly shows a discontinuity which is indicative of an instability corresponding to the formation of a high-field domain in the channel. In contrast to the regime discussed above in connection with Fig. 6, the domain is now formed at high densities of the injection current. We believe that in this case the instability is caused by the real-space transfer and the responsible mechanism is the dynamical screening of the substrate field by the injected charge of hot electrons. This mechanism, first predicted in ref. 2, will be discussed in the next Section.

5. *The NERFET*

As we have seen in Fig. 10a, the hot electron injection in our device is accompanied by a pronounced NDR in the source-drain circuit. This effect is observed not only at low temperatures but at 300 K as well, see Fig. 11. The NDR device can work as an efficient generator and amplifier of electromagnetic oscillations at any frequency below a fundamental cut-off (discussed in Sec. 6). Microwave generation in NERFET has been experimentally demonstrated[17] with high dc to ac conversion efficiency in the gigahertz range. An important advantage of our device over other NDR oscillators is the possibility of controlling the oscillations by a third electrode.

The NDR effect in NERFET arises due to the hot-electron injection into the substrate. One can say that the substrate circuit "steals" current from the drain circuit. On the other hand, one should realize that the injected charge does not reach the substrate instantly but drifts toward it with a constant (saturated) velocity in the conduction band of the AlGaAs barrier. The space-charge thus dynamically stored (i.e., stored while in transit) in the barrier layer depletes the channel by screening it from the positive substrate potential and therefore lowers the drain current. These two effects — the current stealing and the dynamical screening (charge storage) — are both at work in our structure in an inseparable way. Both aspects of the NDR mechanism in NERFET should be clearly understood in order to optimize the device performance.

Below we shall discuss these aspects in conjunction with Fig. 12 which shows a typical family of I-V characteristics in the drain circuit of NERFET at 77 K. These characteristics were taken with different substrate biases as a parameter. Firstly, we note that substrate bias exerts a backgate action: higher V_{SUB} enhances the electron concentration in the channel. In the absence of a real-space transfer this would lead to a higher drain saturation of the transistor. Next, we see that an NDR region appears only for $V_{SUB} \gtrsim 2$ V and that higher V_{SUB} leads to a deeper NDR. The NDR region is typically followed by a flat plateau reminiscent of an FET saturation and then by a sharp rise in the drain current (the latter is obviously due to the "ordinary" thermionic emission SUB $\rightarrow$D).

Consider first the effect of current diversion. In what follows we shall measure the channel current I per unit width of the gate. In the presence of hot-electron transfer the current I(x) is position dependent

and so is the electron concentration $\sigma(x)$. Even in the region of a high source-drain field, where the electron drift velocity is saturated, the current $I(x)$ can be expected to have both drift and diffusion components. On the other hand, we shall neglect the thermoelectric component of the channel current, i.e. the one proportional to the gradient of $T_e(x)$, by assuming a model in which the electric field is concentrated in a domain near the drain, within which T_e is assumed uniform. In the presence of a hot-electron flux J the current-continuity equation is of the form

$$\frac{dI}{dx} = J \,. \tag{7}$$

The drift-diffusion current I in the channel is given by

$$I = \sigma v_s - D\sigma' \tag{8}$$

where the prime denotes d/dx, $D = D(T_e)$ is the diffusion coefficient in a hot-electron ensemble,[18] and v_s is the saturation velocity. The injection current density $J(x)$ is assumed given by eq. (1). Substituting (1) and (8) in (7) we arrive at a differential equation

$$\sigma'' - L_D^{-1}\sigma' - (\lambda L_D)^{-1}\sigma = 0 \,, \tag{9}$$

in which $L_D \equiv D/v_s$ and

$$\lambda = \Delta \exp(\Phi/kT_e)\, v_s/v(T_e) \tag{10}$$

with $v(T_e)$ given by (2). In our model domain of uniform T_e equation (9) predicts that the density of charge in the channel decays exponentially from the beginning of the domain and toward the drain with a characteristic length

$$\Lambda = \frac{\lambda}{2}\left[1 + (1+4L_D/\lambda)^{1/2}\right] . \tag{11}$$

In the limit $\lambda \gg L_D$ which is realized when $\Phi \gg kT_e$ we have $\Lambda \approx \lambda$. This limit is usually realized (except for highest electron temperatures) in our device where $\Phi \approx 0.3$ eV. At low T_e the length $\Lambda \approx \lambda$ is effectively infinite but for $T_e \sim 1500$ K we have $\lambda \sim 10^{-5}$cm and the diffusion component is not negligible. However, at high electric fields the hot-electron diffusivity D drops sharply in polar semiconductors like GaAs, so that we can estimate[18] that in this regime L_D is also quite short ($\lesssim 10^{-5}$cm). Therefore, the limit $\lambda \ll L_D$ (in which $\Lambda \approx \sqrt{\lambda L_D}$) is, probably, never realized in our structure.

The described above effect of current diversion by itself does not explain, even qualitatively, some of the salient features of the NDR characteristics (Figs. 10-12), notably their strong dependence on V_{SUB} and the existence of flat plateaus. To explain these features one must bring into consideration the dynamical screening effect which we shall now discuss.

Once emitted over the top of the barrier, the injected charge travels downhill toward the substrate with a saturated velocity v_s (for simplicity we shall assume that v_s has the same value in the channel and in the AlGaAs barrier). The volume charge density of injected electrons is given by J/v_s and the sheet density of charge stored in transit in the barrier of thickness ℓ is therefore

$$\Delta\sigma = J\ell/v_s \tag{12}$$

The corresponding space-charge potential is given by

$$\Delta\psi = J\ell^2/2\epsilon v_s \tag{13}$$

Let us make simple estimates. In the above model for current diversion we can relate J to the total substrate current (per unit gate width): $I_{SUB} = J\Lambda$. For an efficient channel depletion we must have $\Delta\sigma \approx \sigma_0$ where $\sigma_0 = I_0/v_s$ is the electron sheet concentration in the channel just prior to the onset of hot-electron transfer and $I_0 \sim 2$–3 A/cm is the peak channel current. In the case of a strong NDR, when $I_{SUB} \approx I_0$ (see Fig. 10a), we have therefore $\Lambda \approx \ell = 1500$ Å which agrees with the above estimates for $T_e \gtrsim 10^3$ K, and have $J \sim 2\times10^5$ A/cm^2. >From eqs. (12) and (13) we then find $\Delta\sigma \sim 2\times10^{12}$ e/cm^2 and $\Delta\psi \sim 2$ V. The space-charge potential $\Delta\psi$ opposes charge injection; to overcome this additional barrier the substrate bias V_{SUB} must exceed $\Delta\psi$ and hence for the indicated current densities the NDR is obtained only when $V_{SUB} \gtrsim 2$ V, see Fig. 12.

We can regard $\Delta\psi$ as a threshold shift in a field-effect transistor in which V_{SUB} plays the role of a gate bias. Due to the dynamically stored charge the FET drain current saturates at a lower value corresponding to $V_{SUB} - \Delta\psi$. As the heating field increases, the injection current density goes up and so does $\Delta\psi$ by eq. (13). For a given V_{SUB} the highest possible threshold shift corresponds either to the situation when most of the channel charge is effectively transferred away before reaching the drain (in which case ideally one would have zero valley drain current) or to the situation where the initial slope of the barrier near the channel interface and in the vicinity of the drain (where most of the injection occurs) becomes flat due to the space-charge accumulation in the barrier region. It is this latter situation which is normally realized in our devices. In this case further enhancement of the charge injection is effectively quenched. Indeed, if the initial slope would become *negative* (as in the insert to Fig. 3b) then the injected hot electrons would cool down before reaching the barrier top. In this case one can expect no further reduction of the drain current with increasing V_{SD}. However, for a higher V_{SUB} the similar balance is established at a higher value of V_{SD} and a lower I_D. Therefore, increasing V_{SUB} not only gives rise to an enhancement of the *peak* current (backgate action) but it can also lead to a depression of the saturated *valley* current. This effect, clearly seen in Fig. 12, is an evidence in favor of our peculiar charge-storage mechanism. Of course, the dynamically

induced threshold shift in NERFET is highly volatile: when the charge injection ceases it disappears as soon as the injected electrons reach the substrate. The speed with which this happens determines the fundamental frequency limit discussed in the next section.

The dynamical screening effect itself is independent of the ambient temperature and it gives rise to a qualitatively similar NDR at 300, 77, and 4.2K. Of course, the NDR *is* affected indirectly by changing conditions for the electron heating: the different low-field mobility μ, the domain formation, etc. The higher μ typically leads to an earlier onset of the NDR. However, the difference between the NERFET curves at 77 K and 4.2 K is almost intangible since the advantage of higher mobility at 4.2 K is lost already at low electric fields — well before the NDR region. It is instructive to consider the NERFET characteristic in a high magnetic field, dashed line in Fig. 12. Note that most of the change introduced by H occurs before the onset of NDR and manifests itself in a degraded slope of the IV characteristic and a lower peak current. The NDR portion of the curve is affected very little by the magnetic field up to 8 T.

6. *Speed of CHINT and NERFET*

Fundamental limitations on the intrinsic speed of our hot-electron devices arise[3] due to the time-of-flight delays characteristic of a space-charge-limited current and because of a finite time required for the establishment of an electron temperature.

Consider the latter limitation first. Energy relaxation of hot carriers in bulk semiconductors has been a subject of considerable number of studies (see e.g., the reviews[18,19] and references therein). The dominant mechanisms for the maxwellization of the hot-electron energy distribution function are the polar optic phonon scattering and the electron-electron interaction. The phonon mechanism is expected to be not too different in the CHINT/NERFET structure compared to the bulk. Monte-Carlo studies[20] indicate that the energy loss rate due to polar optic phonon emission by electrons in GaAs is nearly constant for electron energies above 0.1 eV and equals 1.6×10^{11} eV/sec. This translates into about 1 psec equilibration time for $T_e \sim 1500$ K. The influence of e-e scattering (which, as discussed in Sect. 3, is of primary importance for the establishment of the quasi-equilibrium in high-energy tails of electron distribution at relatively low heating voltages) is much more difficult to take into consideration. As far as we know, there is no satisfactory treatment of this process in a two-dimensional problem. The results of Inoue et al.[21] are hardly applicable to our case since they are based on a Debye-like screening, expressed by a length $(\epsilon kT_e/ne^2)^{1/2}$, where n is the electron concentration. Such a screening presupposes the existence of a positive neutralizing background — a condition not fulfilled in regions of a space charge limited current flow. We would expect a much stronger n dependence of the e-e scattering rate, than that calculated in ref. 21. We believe it is probably a safe bet to assume that at the operating voltages of CHINT and NERFET (for which $n = \sigma/e\Delta \gtrsim 10^{18}\text{cm}^{-3}$) the relevant energy relaxation times are less than 1 psec. Further work is clearly required in this area — especially to include the interplay between the real-space and the momentum-space transfers. Since the latter may considerably slow down the device operation, we proposed[3] to employ such heterostructure materials as InGaAs/InAlAs — in which the satellite valleys have a higher energy separation. A further bonus in using these materials is the lower electron effective mass, which enhances the heating effects.

The second fundamental limitation of the speed discussed previously[3] in connection with CHINT arises due to the space-charge capacitance associated with the mobile charge drifting in the high-field regions of the device. It reduces to the time of flight of electrons over these regions — the high field domain in the channel and the downhill slope of the potential barrier. Both regions are of order 10^{-5}cm and the corresponding delay is about 1 picosecond.

It should be emphasized that our time-of-flight limitation is different from the time-of-flight-under-the-gate limitation characteristic of an FET. The latter results from charging the channel by the gate field through the output resistance of a previous identical device which necessarily gives $\tau = L/v$ with L being the gate length. In the CHINT the controlling electrode is the drain and L is the total length of the space-charge-limited current regions — which can be substantially shorter than the source-to-drain distance. The same limitation governs the frequency cut-off in NERFET.

7. *The HE^{2}PRAM*

A memory effect due to the real-space hot-electron transfer results when the second conducting layer (substrate) is floating. In this case the hot-electron injection induces a negative charge in the substrate, which raises the substrate potential and depletes the channel. We have studied this effect[5] by using the same structure with a graded potential barrier, Fig. 2, although it is certainly not an ideal structure for charge retention. Indeed, in our structure the rising substrate potential (which can be interpreted as a thermoelectric force of hot electrons) is applied to the substrate-to-channel triangular-barrier diode in the forward direction. A steady-state situation results when the "cold" thermionic emission exactly balances the hot-electron injection.[2] (Evidently, one can optimize

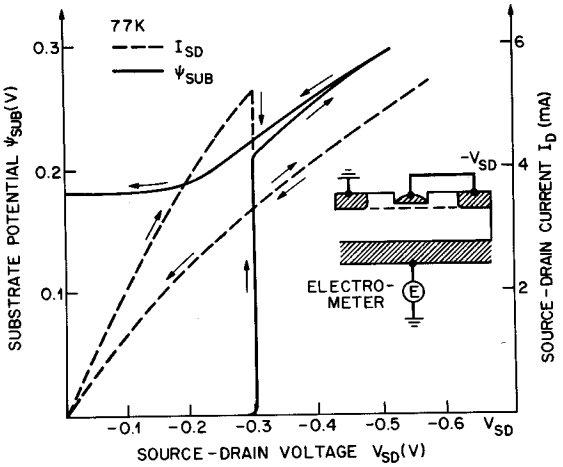


Fig. 13 Substrate potential Ψ_{SUB} and the channel current I_{SD} as function of the heating voltage V_{SD}. Arrows indicate the direction of slow (10 mV/sec) voltage ramping.

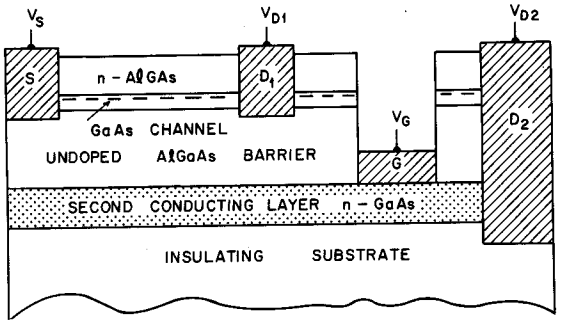


Fig. 14 Structure and energy diagram of the HE^{2}PRAM logic element. Thickness ($\sim 10^{-5}$ cm) and doping level in the second conducting layer are chosen so that this layer can be depleted by the gate field.

the structure emission by using a narrow rectangular barrier instead of the graded triangular one, as we have done with successful results to be reported elsewhere). If one now removes the heating voltage, one can expect a retention of charge in the substrate for a period of time controlled by the cold thermionic emission.

Experimental demonstration of this effect is presented in Fig. 13. The solid line shows the development of the substrate potential Ψ_{SUB} (relative to the grounded drain) as a function of V_{SD}. As the heating voltage is slowly increased, we reach a point $V_{SD} \approx 0.3$ V at which Ψ_{SUB} rises sharply and saturates at $\Psi_{SUB} \approx 0.2$ V within several seconds. At the same time the drain current drops by about 30%, as indicated by the broken line in Fig. 13. This shows the existence of NDR in the channel circuit. Such an effect was predicted by Price[22]. In his words, the disconnected conducting layer "will act as a giant trap causing a negative differential mobility to occur." We should note, however, that this is hysteretic NDR, not capable of generating oscillations — in contrast to the NDR discussed in Sect. 5. As seen from Fig. 13, the current does not rise again when V_{SD} in ramped back and neither does the induced potential Ψ_{SUB} go back to zero. Instead, Ψ_{SUB} settles at a residual value $\Psi_{SUB}(0) \approx 0.18$ V which holds for a long time controlled by a discharge through the high impedance electrometer in the substrate circuit. If the latter is disconnected, then at 77 K the residual potential does not change appreciably for several hours.

As mentioned above, at $V_{SD} = 0.2$ V the rise time of Ψ_{SUB} is several seconds. However, at higher heating voltages the charge-up kinetics by applying V_{SD} in the form of a rectangular pulse of variable amplitude and duration. We found that for $V_{SD} = 0.6$ V it takes about 100 μsec to complete the charge transfer (i.e., reach the state with $\Psi_{SUB}(0) = 0.18$ V). At higher pulse amplitudes this time is still shorter and at $V_{SD} \gtrsim 1.5$ V the ideal memory charge-up time is less than 10^{-9} sec. After each charge-up, the device can be rapidly discharged by grounding the substrate.

The above properties are exploited in the memory device, shown schematically in Fig. 14. The structure must be grown on an insulating substrate, followed by a thin conducting GaAs layer. The key new element is the guard-gate MESFET-like structure G and the second, "deep", drain D_2, contacting both the channel and the second conducting layer. Electrically, D_2 is connected to the source S. When the guard-gate voltage is negative, the substrate conducting layer is isolated from D_2. Applying voltage to D_1 (*write*), we charge up this layer by the hot electron transfer and thus deplete the main channel. Information is *read* by probing the channel resistance. Applying a positive voltage to G, we can rapidly *erase* this information (with the characteristic MESFET time). An important advantage of this device, called the HE^{2}PRAM (hot-electron erasable programmable random-access memory) is the high speed of all logic operations.

8. *Conclusions*

We have reviewed the physical principles of several novel devices which employ hot-electron transfer between two conducting layers separated by a potential barrier. Their operation is based on controlling charge injection over the barrier by modulating the electron temperature in one of the

layers. The devices have been experimentally demonstrated with the help of an AlGaAs/GaAs heterostructure, in which one of the conducting layers is realized as the channel of a modulation-doped transistor and the other as a heavily doped GaAs substrate.

First we discussed the CHINT (charge injection transistor), which is a general-purpose three-terminal amplifying element. Its work can be best illustrated by a comparison to a hypothetical vacuum diode whose cathode temperature is controlled without inertia by an input electrode. This is different from all previous three-terminal devices — which were based either on the potential effect, i.e., the modulation of a potential barrier by an applied voltage (vacuum triode, bipolar transistor, various analog transistors), or on the field effect, which is the screening of an applied field by a variation of charge in a resistive channel. In CHINT the control of output current is effected by a modulation of the electron temperature resulting in charge injection over a barrier of fixed height. The substrate serves as an anode and the channel as a hot-electron cathode, whose effective temperature is controlled by the source-to-drain field. The existence of power gain in this device has been demonstrated experimentally. The value of the mutual conductance obtained in first CHINT devices is comparable to the best bipolar or field-effect transistors.

Next, the hot-electron injection is accompanied by a strong negative differential resistance in the channel circuit. This allows the implementation of a related device, called the NERFET (negative resistance FET), which is essentially a two-terminal microwave oscillator, controllable by the gate and substrate voltages. The NDR in NERFET arises due to a peculiar dynamical screening effect. Injected electrons moving in the AlGaAs barrier region constitute a space-charge limited current. The associated space-charge potential screens the applied positive substrate voltage and thus depletes the channel. This dynamical screening mechanism of NDR is extremely fast — intrinsically limited by the time of flight of electrons across the high-field regions of the device, which is in the picosecond range. (The same speed limitation applied to the operation of CHINT). Microwave generation in NERFET has been observed.

Finally, we discussed a memory effect, which obtains when the substrate is left unbiased. Here one also sees an NDR in the channel circuit — but this is a hysteretic NDR, not capable of generating oscillations. It indicates a charge accumulation due to the hot-electron injection into the floating substrate (which remains charged even after the heating voltage is removed). The thermoelectric force developed between the two conducting layers has a characteristic decay time determined by the ambient temperature and the barrier height. The proposed device based on this effect, the HE^2PRAM (hot electron erasable programmable random access memory) has the advantage of a high speed of all logic operations: write, read, and erase of information.

Throughout most of our discussion of the underlying physics of hot electron injection in double-layered heterostructures, we emphasized the (many) points where our present understanding is incomplete and where further work, both theoretical and experimental, is required.

Acknowledgement — We wish to thank Art Gossard and Rudi Hendel whose collaboration was indispensable for the experimental realization of our devices.

REFERENCES

1. K. Hess, H. Morkoc, H. Shichijo, and B. G. Streetman, *Appl. Phys. Lett.* **35**, 469 (1979); see also K. Hess, *Physica* **117B**, 723 (1983) and references therein.
2. A. Kastalsky and S. Luryi, *IEEE Electron Device Lett.* **EDL-4**, 334 (1983).
3. S. Luryi, A. Kastalsky, A. C. Gossard, and R. Hendel, *IEEE Trans. Electron Devices* **ED-31**, 832 (1984).
4. A. Kastalsky, S. Luryi, A. C. Gossard, and R. Hendel, *IEEE Electron Device Lett.* **EDL-5**, 57 (1984).
5. S. Luryi, A. Kastalsky, A. C. Gossard, and R. Hendel, *Appl. Phys. Lett.* **45**, 1294 (1984).
6. C. L. Allyn, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.* **36**, 373 (1980).
7. R. F. Kazarinov and S. Luryi, *Appl. Phys. Lett.* **38**, 810 (1981); a misprint in eq. (6) of that work is corrected in ref. 14, eq. (7).
8. R. C. Miller, D. A. Kleinman, and A. C. Gossard, *Phys. Rev.* **29**, 7085 (1984); W. I. Wang, E. E. Mendez, and F. Stern, *Appl. Phys. Lett.* **45**, 639 (1984).
9. E. M. Conwell, *High Field Transport in Semiconductors*, New York: Academic Press (1967).
10. R. F. Kazarinov and V. G. Skobov, *Zh. Eskp.Teor. Fiz.*, **42**, 1047 (1962) [Eng. Transl. *Sov. Phys. JETP* **15**, 726 (1962)].
11. J.-I. Nishizawa, T. Terasaki, and J. Shibata, *IEEE Trans. Electron Devices* **ED-22**, 185 (1975)
12. C. O. Bozler and G. D. Alley, *IEEE Trans. Electron Devices* **ED-27**, 619 (1980).

13. S. Luryi and R. F. Kazarinov, *Solid State Electron.* **25**, 933 (1982).

14. R. F. Kazarinov and S. Luryi, *Appl. Phys.* **A 28**, 151 (1982).

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

16. W. Shockley, *Electrons and Holes in Semiconductors*, Van Nostrand: Princeton (1950), p. 110.

17. A. Kastalsky, R. A. Kiehl, S. Luryi, A. C. Gossard, and R. Hendel, *IEEE Electron Device Letters* **EDL-5**, 321 (1984).

18. C. Jacoboni and L. Reggiani, *Adv. Phys.* **28**, 493 (1979).

19. D. K. Ferry, in *Handbook on Semiconductors*, Vol. 1, ed. by T. S. Moss, North-Holland (1982), pp. 563-597.

20. T. J. Maloney and J. Frey, *J. Appl. Phys.*, **48**, 781 (1977); T. J. Maloney, *IEEE Electron Device Lett.* **EDL-1**, 54 (1980).

21. M. Inoue and J. Frey, *J. Appl. Phys.* **51**, 4234 (1980); N. Takenaka, M. Inoue, and Y. Inuishi, *J. Phys. Soc. Jpn.* **47**, 861 (1979).

22. P. J. Price, *IEEE Trans. Electron Devices* **ED-28**, 911 (1981).