Physica 134B (1985) 453-465
North-Holland, Amsterdam



# HOT-ELECTRON TRANSPORT IN HETEROSTRUCTURE DEVICES

Serge Luryi
AT&T Bell Laboratories, Murray Hill, NJ 07974

Alexander Kastalsky
Bell Communications Research, Inc., Murray Hill, NJ 07974

A number of hot-electron device concepts are reviewed with the emphasis on their potential for applications and the limitations. The discussion is restricted to charge-injection devices, i.e., those in which hot electrons are transported either ballistically or thermionically between adjacent layers. Recent developments related to the metal-base transistor and its all-semiconductor analogs, as well as those related to the real-space-transfer effect in heterolayers, are critically reviewed.

## 1. INTRODUCTION

As the dimensions of semiconductor devices shrink and the internal fields rise, a large fraction of carriers in the active regions of the device during its operation are in states of high kinetic energy. At a given point in space and time the velocity distribution of carriers may be narrowly peaked, in which case one speaks about "ballistic" electron packets. At other times and locations, the non-equilibrium electron ensemble can have a broad velocity distribution — usually taken to be Maxwellian and parameterized by an effective electron temperature $T_e > T$, where T is the lattice temperature. Hot-electron phenomena have become important for the understanding of all modern semiconductor devices.[1] Moreover, a number of devices have been proposed whose very principle is based on such effects. This group of devices will be reviewed in the present work.

Commercial utilization of hot-electron phenomena began with the Gunn effect (ref. 1, chap. 11), based on the Hilsum - Ridley - Watkins mechanism for a negative differential resistance (NDR). The Gunn diode is a bulk device in which the NDR arises due to the transfer of hot electrons from the high-mobility central valley in a direct-gap III-V compound semiconductor to its higher-lying low mobility satellite valleys. This is undoubtedly the best-known hot electron device, for which a mature technology has developed.

Another successful application of a hot-carrier effect has been made in a nonvolatile memory device invented by D. Frohman-Bentchkowsky and called FAMOS (ref. 1, p. 500). It represents a p-channel MOSFET structure with a floating gate electrode. In the process of "writing" the memory, carriers, heated by the drain field, avalanche near the drain junction with hot electrons from the avalanche plasma injected into the floating gate. As the gate is charged up, its potential is lowered and the p-channel conductance increases. The FAMOS bears a conceptual similarity to some of the real-space-transfer devices discussed below.

We shall be concerned only with the hot-electron *injection* devices, i.e. such devices in which hot carries are physically transferred between adjacent semiconductor layers. Two distinct classes of such devices can be identified — depending on which of the two hot-electron regimes is essentially employed (the ballistic or the $T_e$ regime). In the *electron-temperature* devices the heating electric field is applied parallel to the semiconductor layers with hot electrons then spilling over to the adjacent layers over an energy barrier. This process is quite similar to the usual thermionic emission — but at an elevated effective temperature $T_e$ — and the carrier flux over a barrier of height $\Phi$ can be assumed proportional

0378-4363/85/$03.30 

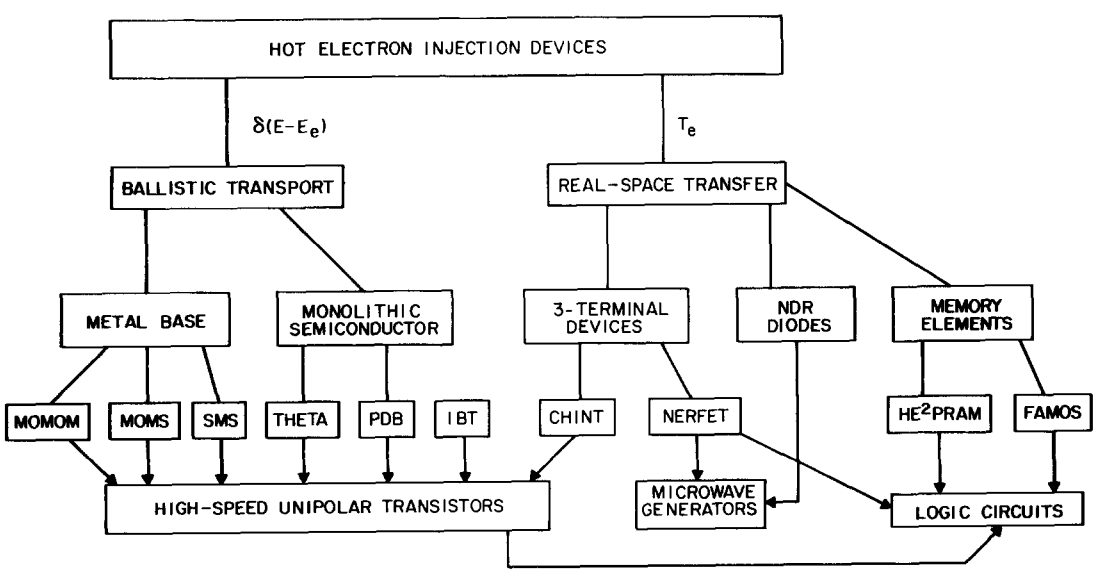
to exp $(-\Phi/kT_e)$. Even though a small fraction of electrons — those in the high-energy tail of the hot-carrier distribution function — can participate in this flux, their number is replenished at a fast rate determined by the energy relaxation time, so that the injection can be very efficient. In the *ballistic* devices, electrons are injected into a narrow base layer at a high initial energy in the direction normal to the plane of the layer. The typical ballistic hot-electron device is illustrated in the logotype of our Conference. Its performance is limited by various energy-loss mechanisms in the base and by the finite probability of a reflection at the base-collector barrier.

Even though the first hot-electron injection devices were proposed a quarter century ago, their full potential has become realizable only with the advent of such heteroepitaxial techniques as the molecular beam epitaxy (MBE) and metalorganic chemical vapor deposition (MOCVD). These techniques can now provide abrupt heterointerfaces and the modulation of doping on the scale of a nanometer — which is essential for the implementation of both classes of hot-electron devices discussed in this work. Latest developments in the heteroepitaxy of III-V compounds have been reviewed by Cho;[2] for general device applications of the "band-gap engineering" see a paper by Capasso.[3] In the next section we shall review the most important hot-electron-injection device structures. Our own work on a novel class of three-terminal electron temperature devices will be discussed in Sect. 3.

## 2. CLASSIFICATION OF HOT-ELECTRON INJECTION DEVICES

The family tree of the hot-electron injection devices is displayed in the table below. The family is large and its members often go under different names. In the attempt to represent only distinct ideas, we may well have overlooked some important relatives!

As discussed in the Introduction, our main conceptual classification is made according to the kind of a hot-electron ensemble employed in the device operation. Although all injection devices involve a real-space transfer (RST) of hot electrons, we shall, following the established terminology, reserve this term for devices operating in the electron-temperature regime. As far as we know, the first proposal of a hot-electron injection device was made by Mead[4]. His device, MOMOM (metal-oxide-metal-oxide-metal), belongs to the category of ballistic-transport transistors. Accordingly, we begin by reviewing this group of devices.



### 2.1 Ballistic injection devices

Devices of this group represent a unipolar analog of the bipolar junction transistor.[5,6] Among themselves they differ by the materials employed and by the physical mechanism of hot-electron injection into the base. The original MOMOM proposal by Mead (Fig. 1a) was based on electron tunneling from a metal emitter through a thin oxide barrier into a high energy state in a metal base. Another insulating barrier separated the base from a metal collector electrode. Later versions of this device[7] had the second MOM replaced by a metal-semiconductor junction, resulting in a transistor structure called the MOMS (Fig. 1b). Attempts have also been made to employ a vacuum collector barrier (MOMVM).[6]

Theoretical estimates of the frequency performance of tunnel-emitter transistors have led several authors to conclude[5] that these devices are inherently inferior to the bipolar transistor. Those conclusions were disputed by Heiblum[6] who, using another set of parameters for evaluation, suggested that certain tunnel-emitter configurations may have an edge. Experimentally, this question is open, although the general consensus is probably reflected in the fact that the tunnel-emitter metal-base transistor concepts have not gained much development in recent years.

Metal-base transistors (MBT), which employ thermionic rather then tunneling injection of hot carriers into the base, were first proposed by Atalla and Kahng[8] and Geppert[9] in the form of a metal-semiconductor-metal (SMS) structure. The basic SMS transistor is illustrated in Fig. 1c. Experimental studies of the SMS device are actively pursued to this day: recent advances[10,11] have been associated with the development of epitaxial techniques for the growth of monolithic single-crystal silicon-metal silicide-silicon structures. This continued interest is explained not only by the scientific usefulness of the SMS structure (it is an excellent tool for studying fundamental properties of hot-electron transport through thin films), but also by lingering hopes to produce a transistor which is faster than the bipolar or FET devices.

The potential merits of the SMS transistor had been appraised long ago by Sze and Gummel.[12] They predicted that despite its possibly superior frequency performance this device would hardly ever replace the bipolar junction transistor. The problem which has plagued the SMS (and all other metal-base) transistors is their poor transfer ratio $\alpha$ (the common-base current gain). Even assuming an ideal monocrystalline SMS structure and extrapolating the base thickness to zero, the typical calculated values of $\alpha$ are unacceptably low — mainly due to the quantum-mechanical (QM) reflection of electrons at the base-collector interface. In our view, these conclusions of the 1966 paper[12] remain valid today. The origin of the QM reflection problem can be traced to the large Fermi energy of electrons in a metal base.[21] Indeed, consider an (over) simplified model of a metal-semiconductor barrier, Fig. 2, and assume parabolic energy-momentum relationships in both materials. The well-known solution of this QM problem gives for the above-barrier reflection coefficient R the following expression

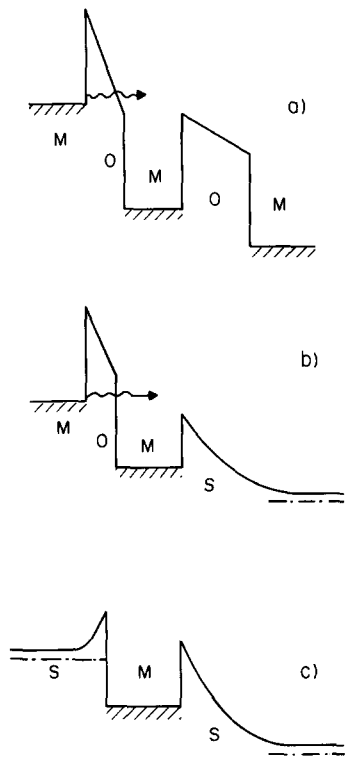



FIGURE 1: Metal Base Transistors. a) MOMOM. b) MOMS. c) SMS

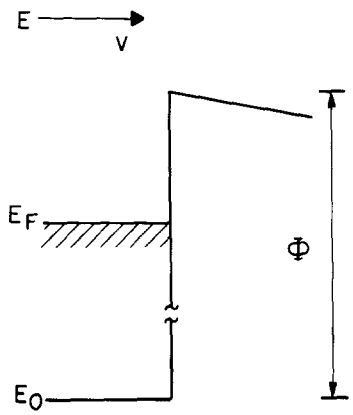



FIGURE 2: Simplified model for estimating the above-barrier reflection of ballistic electrons.

$$R = \left( \frac{1-\nu}{1+\nu} \right)^2, \qquad (1)$$

where

$$\nu = \left( 1 - \frac{\Phi}{E} \right)^{1/2} \qquad (1a)$$

E is the hot-electron energy in the base, and Φ is the barrier height. Note that it is not the clearance $E-\Phi$ but the ratio $E/\Phi$ which enters the expression for R — and hence one must correctly choose the zero energy level, including a large Fermi energy $E_F$. Typically, $\Phi/E$ is close to unity and the reflection is large. For a ballistic electron in Al incident on the interface with GaAs at 0.4 eV above the Schottky barrier ($\Phi \approx 12$ eV), the probability of reflection predicted by (1) is ~ 50%.

This estimate based on the simplest free-electron model is certainly invalid for metals with a complicated band structure. There have been recent reports[11,13] of a transistor action in monocrystalline $Si/CoSi_2/Si$ structures with $\alpha$ as high as 0.6. One cannot rule out some "accidental" resonance which aids the QM transmission of hot electrons in these devices. Such an interpretation, however, appears to us unlikely. A more probable explanation[11] is related to the existence of pinholes in the base metal film, i.e. continuous silicon "pipes" between the emitter and the collector. Such a device would be analogous to the permeable base transistor (PBT),[14] which in our classification is not a hot-electron device. If the area of each pinhole is small, then it is difficult to tell whether the PBT or the MBT mechanism has contributed to the observed IV characteristics. We remark, however, that the thermionic emission through a permeable base has, in our opinion, a greater device potential than the hot-electron transport through a metal base, and thin silicide films may offer an attractive way of fabricating the PBT — if one learns how to control the statistics of pinhole sizes.

The problem of QM reflections is largely avoided in the all-semiconductor ballistic hot-electron transistors. A number of such devices have been manufactured recently, using MBE grown planar-doped barriers (PDB),[15] heterostructure barriers[16] or ion-implanted "camel" barriers.[17] Even when the base is degenerately doped, the Fermi energy is typically less than 0.1 eV. It is possible, therefore, to arrange an injection energy so that, say, $\Phi/E \lesssim 1/2$. In this case, Eqs. (1) give $R \lesssim 0.03$. Both tunnel-emitter and thermionic-emitter versions of the ballistic hot-electron transistor have been implemented. Figure 3 shows the schematic energy-band diagrams of these devices. Using the GaAs/GaAlAs heterostructure technology by MBE, Yokoyama et al.[16] manufactured a tunneling device THETA (tunneling hot-electron transfer amplifier, the name coined by Heiblum[6]) with a 500Å-thick $Al_{0.3}Ga_{0.7}A_s$ emitter barrier and a 1000Å-thick n-GaAs base. Evidently, in such a structure a significant tunneling current occurs via the Fowler-Nordheim mechanism, Fig. 3a. A respectable (for hot-electron transistors) transfer ratio $\alpha = 0.28$ was observed in this device. Even more impressive are the recent results achieved with thermionic injection. A device of this type is illustrated in Fig. 3b. Shannon and coworkers obtained $\alpha = 0.95$ in a GaAs PDB transistor[18] which had a nominal base thickness d (distance between the planes of p-type doping) of 700Å. Similar results ($\alpha \approx 0.7$ for $d \approx 550$Å) have been obtained by Hayes and coworkers.[19] When these results are confirmed and made

reproducible, the PDB (camel) transistor may become a serious contender for ultra-high speed applications.

One should understand the trade-off involved in the design of all hot-electron transistors with a doped base: cooling of hot-electrons by phonon emission and other inelastic processes (minimized by thin base layers) against the increasing base resistance for thinner layers. It is easy to estimate the RC delay associated with charging the working base-emitter capacitance and the parasitic base-collector capacitance through the lateral base resistance:

$$RC = \tau_b = \frac{\epsilon L^2}{\ell \mu \sigma}, \qquad (2)$$

where $\ell$ is the thickness of the emitter or the collector barriers, $\ell \sim 10^{-5}$ cm, L the characteristic lateral base dimension (shortest distance to the base contact from the geometric center of the base), $L \sim 10^{-4}$ cm, $\mu$ the mobility in the base, $\sigma$ the mobile charge density per unit base area, and $\epsilon$ the dielectric permittivity. For a hot-electron transistor to be competitive, one must have $\tau_b \approx 1$ psec, which means that the sheet resistance in the base must be $(\mu\sigma)^{-1} \lesssim 1\ k\Omega/\square$. For n–GaAs the active dopant concentration is typically $2\times10^{18}$ cm$^{-3}$ and the mobility $\mu \lesssim 300$ cm$^2$/V·sec. Thus, one needs $\sigma/e \gtrsim 2\times10^{13}$ cm$^{-2}$ and base thickness $d \gtrsim 1000$Å. At such thicknesses one should expect a degradation in $\alpha$ due to various energy-loss mechanisms (for example, hot electrons in GaAs lose energy at the rate of about 0.16 eV/psec due to the emission of optic phonons[20]). The limitation (2) is rather severe. The minimum value of L is governed by the lithographic resolution. One cannot really make the barrier thicknesses $\ell$ much larger than 1000Å, since this would introduce the emitter and the collector delays of more than 1 psec.

An attempt to circumvent (2) was made in a recent proposal[21] of an induced-base transistor (IBT). In this device, illustrated in Fig. 3c, the base conductivity is provided by a 2-dimensional (2-D) electron gas induced by the collector field at an undoped heterointerface. The density of the induced charge is limited by a dielectric breakdown in the collector barrier. For a GaAs/AlGaAs system this means $\sigma/e \lesssim 10^{12}$ cm$^{-2}$. The IBT benefits from the enhanced mobility effect.[22,23] At room temperature $\mu$ is limited by phonon scattering, $\mu \lesssim 9000$ cm$^2$/V·sec, giving $(\mu\sigma)^{-1} \approx 600\ \Omega/\square$. The base conductivity is virtually independent of its thickness down to $d \lesssim 100$Å. At such short distances the loss of

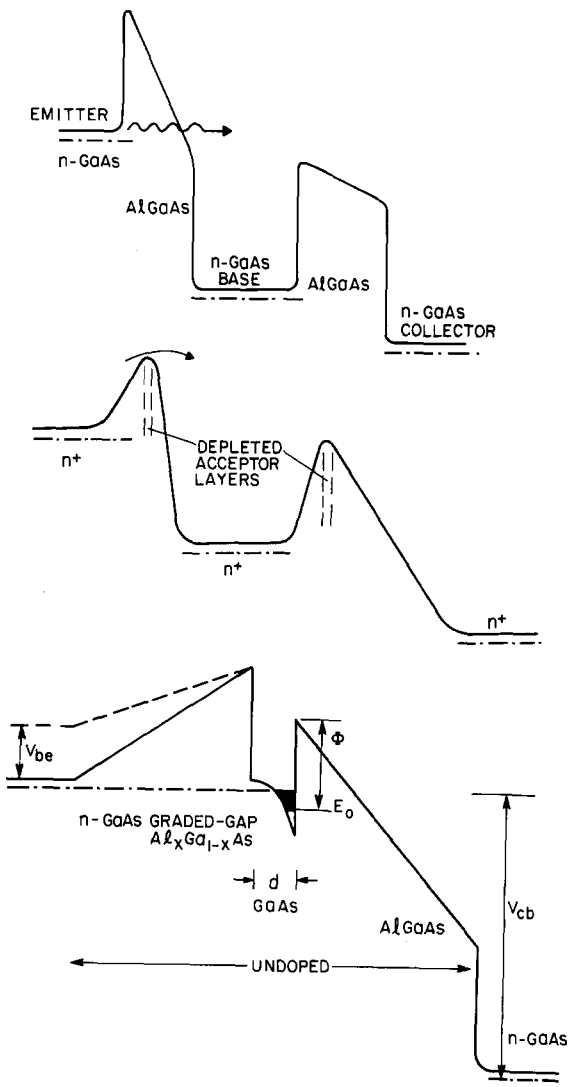



FIGURE 3: Ballistic hot-electron transistors with a monolithic semiconductor structure. a) Tunnel-emitter (THETA) transistor[16]. b) Planar-doped-barrier (PDB) transistor.[15,17–19] c) Induced-base transistor (IBT).[21]

hot electrons in the base due to phonon emission is negligible. The IBT concept is further discussed in ref. 24.

Before leaving the subject of ballistic transistors, let us briefly discuss their potential frequency performance. It is sometimes stated[16] that hot-electron transistors are capable of subpicosecond operation because such is the time of flight of ballistic electrons across the base. This is a much too often repeated fallacy: that time of flight has nothing to do with the intrinsic device speed. Like the bipolar, the FET, and most other transistors, hot-electron transistors have a regime in which their output current I rises exponentially with the input (base-emitter) voltage. In this regime, the maximum speed of operation is proportional to I. However, like every exponent in nature (except, possibly, the inflation), this dependence eventually saturates and goes over into a linear law. One gains no further advantage in speed by increasing I, since the charge stored in all input capacitances will rise proportionally. Ultimately, the speed of a transistor is determined by the current level at which one has a crossover between the exponential and the linear regimes.[25] In transistors with a thermionic emitter this crossover occurs because of the accumulation of the mobile charge diffusing up the emitter barrier and drifting down the collector barrier. A rigorous $g_m/C$ analysis[21] leads to the characteristic delays $\tau_e = \ell_e/v_{th}$ and $\tau_c = \ell_c/v_s$, where $\ell_e$ and $\ell_c$ are the thicknesses of the emitter and the collector barriers, respectively, $v_{th}$ is the thermal velocity of carriers, and $v_s$ their saturated drift velocity. Of course, neither of the $\ell$'s can be shrunk below, say, 1000Å — because of the complementary limitation (2). We conclude that an ideally optimized ballistic transistor will be a roughly 3 picosecond device.

## 2.2 Real-Space-Transfer Devices

The term "real-space transfer" (RST) was coined by Hess et al.[26] to describe a new mechanism for NDR they proposed and subsequently discovered[27] in layered heterostructures. The original RST structure is shown in Fig. 4. In equilibrium the mobile electrons reside in undoped GaAs quantum wells and are spatially separated from their parent donors in AlGaAs layers. Guided by an analogy with the momentum-space intervalley transfer, Hess et al.[26] suggested that carriers, heated by an electric field applied parallel to the layers, will move to the adjacent layers by thermionic emission, causing an enhancement of the mobile charge concentration in one set of layers and depletion in the other. Since the layers had different mobilities, the RST process was predicted to result in an NDR in the two-terminal circuit. This effect was discovered experimentally[27] and used for microwave generation.[28] Several RST diode configurations have been reviewed by Hess.[29] If the device is used as an oscillator, electrons must cycle back and forth between the high and low mobility layers. The maximum oscillation frequency is, in our view, limited by the delay due to "cold" electrons returning from the potential "pockets" in the wide-gap layers, cf. Fig. 4. This process can be viewed as a thermionic emission over the potential barrier due to the space-charge of ionized donors. For a modulation-doped AlGaAs/GaAs heterostructure at room temperature one can estimate the return time to be at least $10^{-11}$ sec and still longer at lower temperatures. On the other hand, the time constants involved in the initial transfer of hot electrons are considerably shorter.[29]

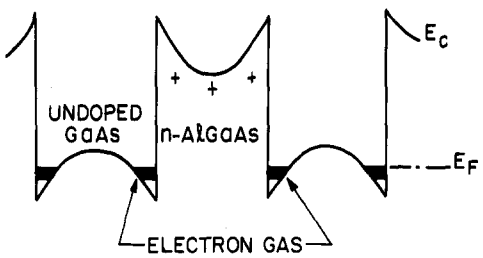



FIGURE 4: The real-space-transfer diode.[26]

The important idea of real-space transfer was taken up in our recent proposal[30] of a *three-terminal* hot-electron device structure. In this structure the RST effect gives rise to charge injection between two conducting layers isolated by a potential barrier and contacted separately. Our idea can be best illustrated

by the glow-cathode analogy, displayed in Fig. 5. In a vacuum

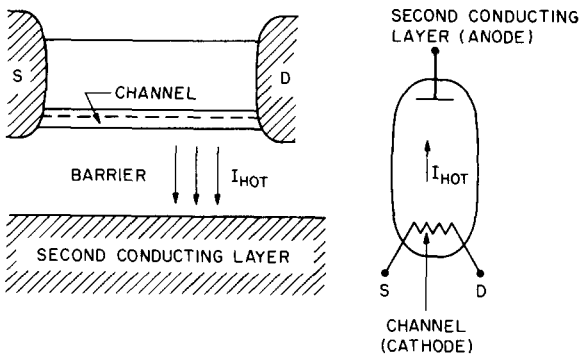



FIGURE 5: Illustration of the principle of three-terminal RST devices.[31]

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 $T_e$ 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 demonstrated experimentally. This work has been thoroughly discussed in our recent review.[31] In the next section we shall briefly review the key ideas, trying not to repeat the review[31] but rather concentrate on the most recent developments and the future perspectives.

## 3. THREE-TERMINAL RST DEVICES

### 3.1 The device structure

Two basic structures used in our work are illustrated in Fig. 6. In the original structure[30] the second conducting layer was implemented as a conducting GaAs substrate separated by a graded-gap AlGaAs barrier from the channel of a modulation-doped FET with source (S) and drain (D) contacts, Fig. 6a. Details of the MBE growth and processing can be found in ref. 32. This device had an auxilliary fourth electrode (gate) which concentrated the lateral electric field under a 1 $\mu$m wide notch. In the most recent work[33] both the gate electrode and the modulation-doping were eliminated, Fig. 6b, and the channel was induced at the undoped heterointerface by a back-gate action of the second conducting layer. (Even though the latter is now implemented as a heavily doped n-GaAs layer on a semi-insulating substrate, we shall keep the designation SUB for this electrode.) Also in the new structure the rectangular potential barrier provides a better insulation between the two conducting layers. A similar structure has been implemented using InGaAs/InAlAs heterojunctions.[34]

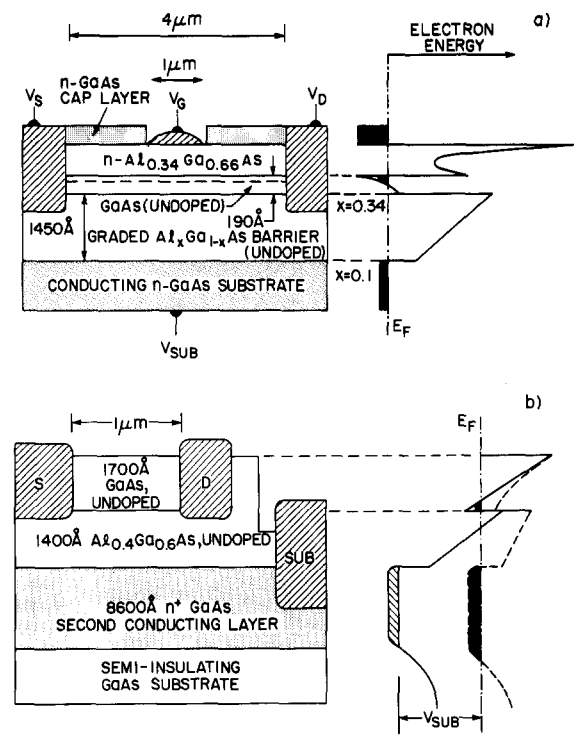



FIGURE 6: Cross-section and the energy-band diagram of three-terminal RST devices. a) The original structure[30–32], type-1. b) The new structure[33], type-2, with its band diagram shown for equilibrium (dashed line) and for a positive voltage $V_{SUB}$ applied to the second conducting layer.

A critical step in manufacturing the three-terminal RST devices is to provide ohmic contacts to the 2-D electron gas in the channel, while preserving the insulation from the SUB layer. Figure 7a shows the characteristics of a diode formed between the SUB electrode and S,D terminals tied together and grounded. A better insulation at 300K in the type-2 device is evident: for $V_{SUB} \lesssim 1.5$ V the barrier is blocking in both directions. For the graded-gap structure the diode characteristic is asymmetric, as expected. The observed current at $V_{SUB} > 0$ is probably due to a combination of barrier lowering and (thermally assisted) tunneling, especially at lower temperatures. Recently, Grinberg[35] calculated the expected $I_{SUB}$ ($V_{SUB}$) characteristics for a type-1 structure. His results, Fig. 7b, show that at lower electron temperatures the tunneling component is dominant. On the other hand, for charge injection of hot electrons tunneling can be neglected.

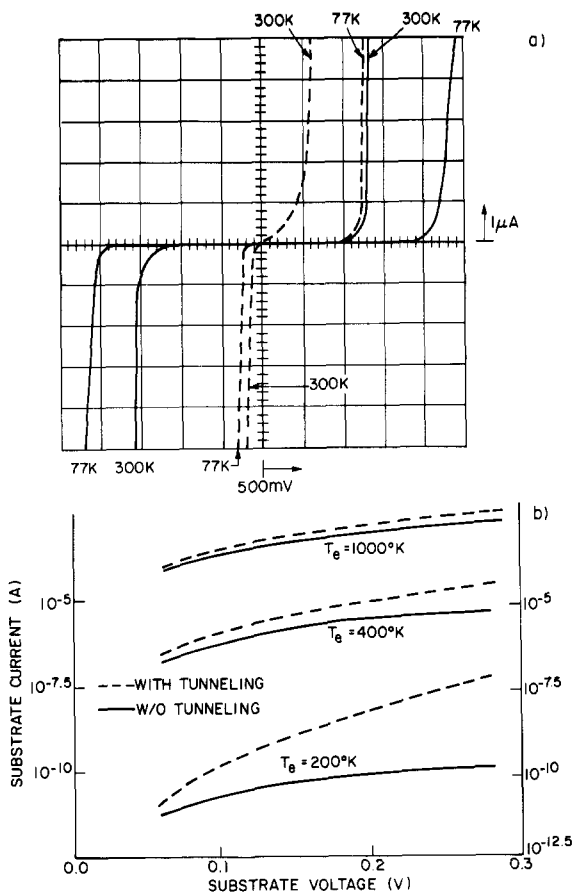



FIGURE 7: Barrier leakage between the two conducting layers. a) Observed $I_{SUB}$ ($V_{SUB}$) dependence for $V_{SD} = 0$ in both types of structures at 300K and 77K. Dashed line corresponds to type-1 and solid line to type-2 structures. b) Calculated $I_{SUB}$ versus $V_{SUB}$ characteristics[35] for different electron temperatures in the channel, assuming a type-1 barrier of equilibrium height 0.3 eV and thickness 1500Å.

## 3.2 CHINT and NERFET

The charge injection transistor (CHINT) is a solid-state analog of the hypothetical vacuum diode with controlled cathode temperature, discussed above in connection with Fig. 5. Application of a voltage $V_{SD}$ produces a lateral electric field which heats the channel electrons to temperatures $T_e = 1500$K and higher[32] and leads to an exponential enhancement of charge injection over the barrier. Figure 8 displays the anode characteristics in CHINT with the heating voltage $V_{SD}$ as a parameter. These characteristics demonstrate the existence of a power gain and a transconductance of 240 mS/mm at 77 K. They were obtained, however, by subtracting a parasitic diode leakage due to the imperfect contact insulation in our type-1 structure.

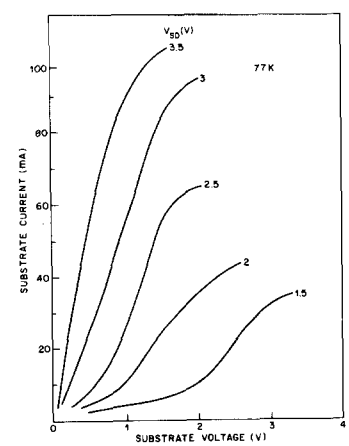



FIGURE 8: Transistor characteristics of CHINT in the common-drain configuration with the heating voltage $V_{SD}$ as a parameter.[32]

The hot-electron injection in CHINT is accomplished by a strong NDR in the channel circuit. This permits the implementation[36,33] of a related device called NERFET (negative-resistance FET). The typical NERFET characteristics are shown in Fig. 9 for both types of structure. We note that the NDR appears for $V_{SUB} \gtrsim 2V$ and it is strongly affected by $V_{SUB}$. It is clear that higher $V_{SUB}$ enhances the electron concentration in the channel (backgate action). The physical mechanism of the NDR consists in the dynamical screening effect predicted earlier.[30] As discussed in detail in ref. 31, hot electrons emitted over the barrier constitute a negative space charge dynamically stored in the AlGaAs barrier layer. This charge screens the backgate field and thus depletes the channel. The associated space-charge potential can be regarded as a threshold shift in a field-effect transistor in which $V_{SUB}$ plays the role of a gate bias. The dynamical screening mechanism of NDR is extremely fast — intrinsically limited by the time of flight of injected electrons toward the second conducting layer. The same limitation applies to the operation of CHINT. Microwave generation by NERFET in the gigahertz range has been observed.[37] Although in principle NERFET is a picosecond device, its real speed limit at present arises from the RC delay due to large contact pads (the parasitic D—SUB capacitance). The main advantage of NERFET over two-terminal microwave generators lies in the possibility of controlling the oscillations by a third electrode. This advantage can also be used in logic applications, as discussed below.

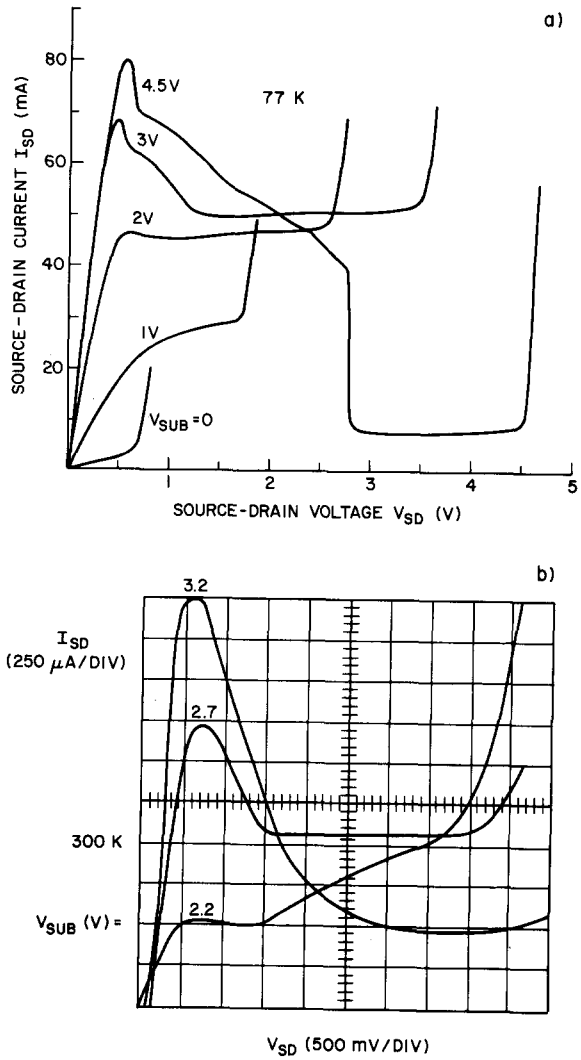



FIGURE 9: Typical NERFET characteristics $I_{SD}$ versus $V_{SD}$ with $V_{SUB}$ as a parameter. a) Type-1 structure at 77K.[36] b) Type-2 structure at 300K.[33]

### 3.3 Logic applications

When two negative-resistance devices (like tunnel diodes or Gunn diodes) are connected in series and the total applied voltage $V_{DD}$ exceeds roughly twice the critical voltage for the onset of NDR in the single device, then an instability occurs in which one of the devices takes most of the applied voltage, that is to say, contains a high-field domain, while the other is in the low-field mode. This is illustrated by the usual load-line graphical construct, Fig. 10. As is well-known[38], the operating points A and C are stable, while B is unstable. Which of the devices contains the domain is determined by an accidental fluctuation or, if the system is prepared in one of the stable states, by the history. Various schemes have been proposed to utilize this bistability. Due to the existence of a controlling electrode, NERFET offers new possibilities for logic.

Room-temperature operation of a simplest NERFET logic circuit[33] is illustrated in Fig. 11. Two type-2 NERFETs with nearly identical characteristics (displayed in Fig. 9b) were connected in series, as shown in Fig. 11a. One of the controlling voltages was fixed, $V_{SUB2} = 2.5V$, and the output voltage $V_{OUT}$ was measured as

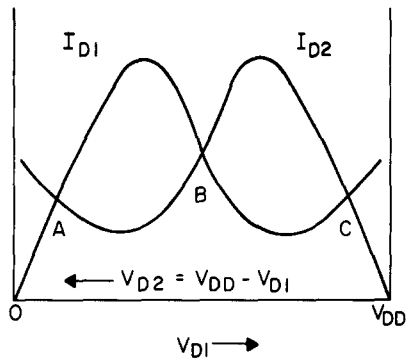



FIGURE 10: Graphical construct for determining the operating points of a circuit formed by two identical NDR elements in series. Points A and C are stable, point B unstable.

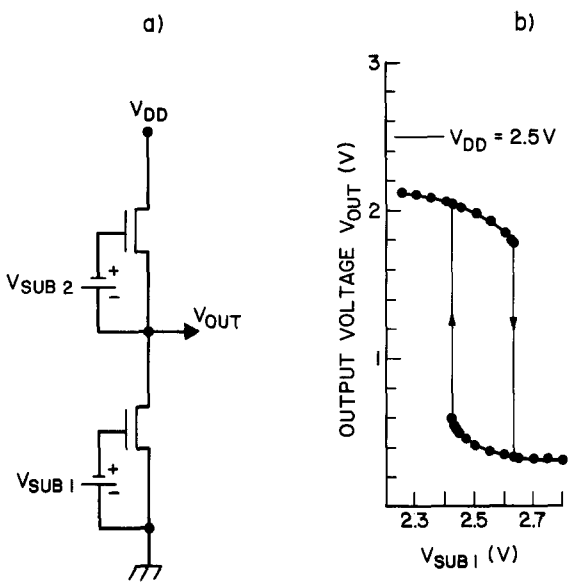



FIGURE 11: NERFET logic circuit.[33] a) Schematic diagram. Conventional FET circuit symbols are used with the understanding that the SUB electrode plays the role of a gate. b) Logic transitions at room temperature. The output voltage was measured at fixed $V_{SUB2} = 2.5V$ as a function of $V_{SUB1}$ slowly varied in the direction shown by the arrows.

a function of $V_{SUB1}$, Fig. 11b. As the controlling voltage $V_{SUB1}$ is varied, the system smoothly approaches the switch points (sharply defined and repetitive within 1 mV), at which $V_{OUT}$ jumps between the low and the high values. Two types of logic operation can be thought of in this configuration. Firstly, we can dc pre-bias our input voltage to a value in the middle of the hysteretic loop, say $V_{SUB1} = 2.5V$. Applying controlling signals $\Delta V_{SUB1}(t)$ in the form of short low-amplitude ($|\Delta V| \gtrsim 0.15V$) pulses of varying polarity, we have a *bistable element*: the system will "remember" the sign of the last pulse, viz. $V_{OUT}$ = high for $\Delta V < 0$ and $V_{OUT}$ = low for $\Delta V > 0$. A second type of logic operation — *inverter* action with amplification — can be obtained by dc prebiasing $V_{SUB1}$ to high enough voltages ($V_{SUB} \gtrsim 2.7V$) to ensure a stable low state. The system will then switch to its high state only during a pulse of negative polarity $|\Delta V_{SUB1}| \gtrsim 0.3V$. Both operations have been demonstrated[33] by pulse-mode experiments.

The width of the hysteresis decreases with increasing $V_{DD}$. At $V_{DD} \gtrsim 3.0V$, the output swing of the inverter is capable of switching a second inverter without any level shifting, so that direct coupling of inverters, such as in a ring oscillator, is possible. At $V_{DD} > 3.5V$ we saw[33] a new feature: the appearance of a third stable state characterized by $V_{OUT}$ being in a range near $V_{DD}/2$. Within that range the dependence $V_{OUT}$ versus $V_{SUB1}$ was strictly linear with a voltage gain of nearly 4 and nonhysteretic. To our knowledge, a tri-stable operation with a stable midpoint has never been observed before with two series-connected voltage-controlled NDR elements. It may become useful for ternary logic.

Another hot-electron logic device can be based on a memory effect[39] which obtains when the second conducting layer in a CHINT/NERFET structure is left unbiased. In this case, the hot electron injection leads to a charge accumulation in the floating layer and a drop in its electrostatic potential $\Psi_{SUB}$, which persists for a long time after the heating voltage $V_{SD}$ is set to zero. The negative $\Psi_{SUB}$ depletes the channel much like it happens in non-volatile memory devices. Based on this effect, we proposed a memory device which allows a fast operation of all functions: *write*, *read*, and *erase*.[31] It received the name HE$^2$PRAM (hot-

electron erasable programmable random access memory).

### 3.4 Perspectives

The CHINT is a multi-purpose high-speed device. Electrically, its operation is analogous to the bipolar transistor and the ballistic hot-electron transistors discussed in Sect. 2, if one identifies the terminals S ≡ emitter, D ≡ base, and SUB ≡ collector. An interesting feature of CHINT is the fact that its differential common-base current gain $\alpha \equiv (\partial I_{SUB}/\partial I_S)$ at $V_{SUB}$ = const can substantially exceed unity (due to the NDR in the S-D circuit). Practical utilization of CHINT may become possible only after one gets rid of the parasitic collector leakage. An important advance in this regard is our type-2 structure. However, we have experienced difficulties with ohmic S and D contacts in this structure, which deviates from the mature MODFET technology. A possible approach is to incorporate some modulation doping in the barrier separating the two conducting layers in a type-2 structure.

The frequency of microwave oscillations in NERFET is presently limited by parasitic RC delays. The main parasitic capacitance — between the D and SUB electrodes — can be reduced by shrinking the contact pads and/or by ion implantation of oxygen underneath the pads. (The latter should also help reduce the leakage in CHINT.) At the same time one should try to minimize the channel resistance (at the peak prior to onset of the NDR), which includes the series contact resistance. Ideally, the intrinsic RC delays in CHINT/NERFET devices are of the order of several picoseconds.[31] Limits on the switching speed in a two-NERFET logic circuit are uncertain at present and require further study, both experimental and theoretical.

The type-1 structure used[39] to study the memory effect due to RST into a floating layer was certainly not an ideal structure for charge retention. Indeed, in that structure the thermoelectric force of hot electrons $\Psi_{SUB}$ is applied to the graded-gap triangular barrier diode in the forward direction. A steady-state situation, which results when the "cold" thermionic emission back into the channel exactly balances hot-electron injection, occurs at a lower level of charge transfer than in structures with a rectangular barrier. The residual $\Psi_{SUB}$ after $V_{SD} \to 0$ was only 0.18V[39], whereas in our later experiments with type-2 structures persistent values of $\Psi_{SUB}$ as high as ~ 0.7V were seen at 77 K.

Another interesting direction of future work involves using materials other than GaAs/AlGaAs. For example, InGaAs/InAlAs heterostructure appears to be a promising candidate due to the lower effective mass in InGaAs and possibly higher efficiency of electron heating. In preliminary experiments[34] we observed a pronounced NDR in the InGaAs/InAlAs NERFET, see Fig. 12. The structure, shown in the insert, combined a type-2 rectangular barrier with the modulation doping. A difficulty (not yet overcome) is to obtain good ohmic contacts to the channel. Our usual Au/Ge-Ag-Au alloy junctions did not provide a satisfactory contact in this structure, as is evidenced in Fig. 12 by a diode-like offset at low currents. On the other hand, Ni based contacts do not give an abrupt edge of alloy penetration, and typically short to the SUB layer. An important technological problem in the implementation of all hot-electron heterojunction devices is to provide an abrupt and shallow ohmic contact to the 2-D electron gas.

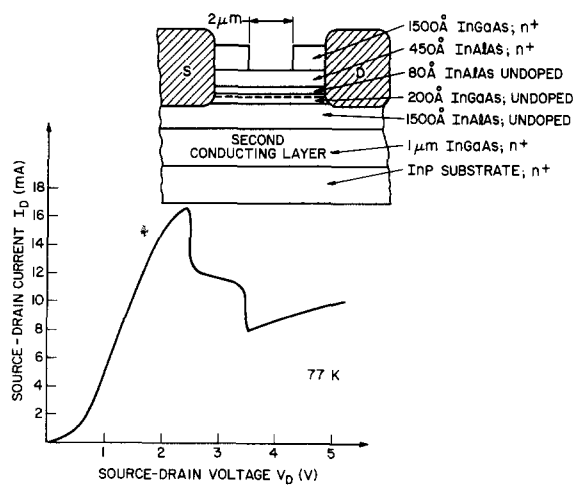



FIGURE 12: First results for $In_{0.53}Ga_{0.47}As/In_{0.48}Al_{0.52}As$ NERFET structure: $I_{SD}$ versus $V_{SD}$ at a fixed $V_{SUB}$ = 4V. Insert shows the device cross-section.

Finally, we remark that semiconductor heterojunctions are not the only way to implement three-terminal RST devices. An interesting possibility lies in using thin semimetal films forming a Schottky barrier with a semiconductor collector underneath. We found experimentally[40] that Bi/Si junctions have a barrier height of 0.63 eV. Because of reduced electron-phonon and electron-electron scattering rates (due to high $\epsilon$ and low carrier concentration), bismuth is known to exhibit strong hot-electron effects. If the electric field is applied laterally to a Bi film on a silicon substrate, one can expect an efficient emission of hot electrons over the Schottky barrier. So far, we have failed to detect this effect.

## 4. CONCLUSIONS

We have reviewed and discussed a number of hot-electron injection devices. Let us quote from the well-known out-of-print book[5]: *"...a useful solid-state device is one which can be used in electronic applications or can be used to study the fundamental physical parameters. All the hot electron transistors, at the present time, belong to the latter category."* Today, after 15 years, they still do... Is there any hope then that these devices will ever leave the fall-back category and move into the world "where the money is"? We believe that the answer is in the affirmative. The new force is coming from the remarkable advancement of the last decade in the techniques of crystal growth (such as MBE and MOCVD) and device processing (submicron lithography, ion implantation, etc.). We hope that the next decade will indeed see a commercial exploitation of hot-electron injection devices — such as some of those reviewed in this paper and those not yet invented.

## REFERENCES


1. S. M. Sze, Physics of Semiconductor Devices, 2nd edition (Wiley, New York, 1981).
2. A. Y. Cho, Proc. 17th Int. Conf. Phys. Semicond., San Francisco 1984, J. D. Chadi and W. A. Harrison, eds. (Springer, New York, 1985) 1515.
3. F. Capasso, Physica 129B (1985) 92.
4. C. A. Mead, Proc. IRE 48 (1960) 359; J. Appl. Phys. 32 (1961) 646.
5. A review of the early developments in the history of the ballistic hot-electron transistor can be found in the 1st edition of ref. 1, Chap. 11, Wiley (1969) and in ref. 6.
6. M. Heiblum, Solid-State Electron. 24 (1981) 343.
7. J. P. Spratt, R. F. Schwarz, and W. M. Kane, Phys. Rev. Lett. 6 (1981) 341.
8. M. M. Atalla and D. Kahng, IRE-AIEE Solid State DRC, University of New Hampshire (1962).
9. D. V. Geppert, Proc. IRE 50 (1962) 1527.
10. E. Rosencher, S. Delage, Y. Campidelli, and F. A. d'Avitaya, Electron. Lett. 20 (1984) 762.
11. J. C. Hensel, A. F. J. Levi, R. T. Tung, and J. M. Gibson, Appl. Phys. Lett. 46 (1985) July 15.
12. S. M. Sze and H. K. Gummel, Solid-State Electron. 9 (1966) 751.
13. E. Rosencher et al., this volume and private communication.
14. C. O. Bozler and G. D. Alley, IEEE Trans. Electron. Devices ED-27 (1980) 619.
15. R. J. Malik, M. P. Hollis, L. F. Eastman, C. E. C. Wood, and T. T. AuCoin, Proc. 8th Biennial Cornell Conf. on Active Microwave Semicond. Devices and Circuits (1981) 87.
16. N. Yokoyama, K. Imamura, T. Ohshima, H. Nishi, S. Muto, K. Kondo, and S. Hiyamizu, Japan. J. Appl. Phys. 23 (1984) L311; also IEDM-84 Tech Dig. (1984) 532.
17. J. M. Shannon, Proc. Inst. Phys. Conf., Ser no. 69, (1984) 45.
18. J. M. Shannon, private communication.
19. J. R. Hayes, private communication.
20. T. J. Maloney and J. Frey, J. Appl. Phys. 48 (1980) 781; also J. Y.-F. Tang and K. Hess, IEEE Trans. Electron Devices ED-29 (1982) 1906. The convenient number quoted is from T. J. Maloney, IEEE Electron Device Lett. EDL-1 (1980) 54.
21. S. Luryi, IEEE Electron Device Lett. EDL-6 (1985) 178.
22. T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. 54 (1982) 437.
23. P. M. Solomon and H. Morkoç, IEEE Trans. Electron Devices ED-31 (1984) 832.

24. S. Luryi, this volume.
25. R. F. Kazarinov and S. Luryi, Appl. Phys. A28 (1982) 151.
26. K. Hess, H. Morkoç, H. Shichijo, and B. G. Streetman, Appl. Phys. Lett. 35 (1979) 469.
27. M. Keever, H. Shichijo, K. Hess, S. Banerjee, L. Witkowski, H. Morkoç, and B. G. Streetman, Appl. Phys. Lett. 35 (1979) 459.
28. P. D. Coleman, J. Freeman, H. Morkoç, K. Hess, B. G. Streetman, and M. Keever, Appl. Phys. Lett. 40 (1982) 493.
29. K. Hess, Physica 117B & 118B (1983) 723.
30. A. Kastalsky and S. Luryi, IEEE Electron Device Lett. EDL-4 (1983) 334.
31. S. Luryi and A. Kastalsky, Superlattices and Microstructures 1 (1985) no. 5.
32. S. Luryi, A. Kastalsky, A. C. Gossard, and R. H. Hendel, IEEE Trans. Electron Devices ED-31 (1984) 832.
33. A. Kastalsky, S. Luryi, A. C. Gossard, and W. K. Chan, IEEE Electron Device Lett. EDL-6 (1985) July.
34. To be published in coll. with A. Y. Cho.
35. A. A. Grinberg, unpublished.
36. A. Kastalsky, S. Luryi, A. C. Gossard, and R. Hendel, IEEE Electron Device Lett. EDL-5 (1984) 57.
37. A. Kastalsky, R. A. Kiehl, S. Luryi, A. C. Gossard, and R. Hendel, IEEE Electron Device Lett. EDL-5 (1984) 321.
38. R. Landauer, Phys. Today 31 (1978) 23 and references therein.
39. S. Luryi, A. Kastalsky, A. C. Gossard, and R. Hendel, Appl. Phys. Lett. 45 (1984) 1294.
40. In collaboration with E. Labate.