Solid-State Electronics Vol. 33, No. 7, pp. 831–836, 1990
Printed in Great Britain. All rights reserved

0038-1101/90 \$3.00 + 0.00
Copyright © 1990 Pergamon Press plc

A UNIPOLAR TRANSISTOR WITH NEGATIVE OUTPUT RESISTANCE

NADIR DAGLI
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, U.S.A.

(*Received* 20 *July* 1989; *in revised form* 5 *November* 1989)

Abstract—In this work a three-terminal active solid state device, that operates well into submillimeter-wave frequencies is proposed. The activity of this device is due to negative output resistance. The origin of the negative output resistance is shown to be injection and transit-time delays that any finite sized device possesses. Generating proper injection delays utilizing specially designed tunneling junctions and combining these injection delays with transit-time delays a novel hot electron transit time transistor (HETT) is introduced and its small signal analysis is carried out. Having three terminals in an active device provides isolation between input and output and completely integrated monolithic negative resistance amplifiers can be fabricated, without the need for non-reciprocal circuit components, like circulators. It is shown that negative output resistance should also be observed in transistors due to inherent intrinsic delays.

1. INTRODUCTION

Solid state devices that operate at millimeter and submillimeter-wave frequencies are highly desirable and have many important applications such as space based millimeter-wave imaging arrays, all weather radar, weapons detection, space based communications, nondestructive testing and plasma diagnostics. To extend the upper frequency of activity of solid state devices into this frequency spectrum is a very active research area. The technological improvements in device processing and material growth over the past decade accelerated the research towards this goal as well as opening new possibilities to create new and totally novel device structures. In this work a new hot electron transit-time transistor (HETT) with potential to operate at these frequencies is introduced. The principle of operation of this device is to utilize intrinsic time delays (injection and transit-time delays) that the charged particles (electrons or holes) experience in any finite sized structure. In this work such intrinsic delays are taken advantage of, rather than being seen as limitations that need to be minimized. In the next section the existence of the negative output resistance due to these time delays will be explained. Next the generation of proper injection delays utilizing specially designed tunneling junctions and creation of a novel hot electron transit-time transistor (HETT) combining these injection delays with transit-time delays will be explained. Then a specific device structure will be proposed and its small signal analysis will be carried out. Finally implications of negative output resistance and effects of intrinsic time delays on transistors will be discussed.

2. NEGATIVE RESISTANCE DUE TO INTRINSIC TIME DELAYS

Until recently, in active solid state devices the limiting factor for the upper frequency of operation has always been the external parasitics associated with the device rather than the intrinsic limitations. At present, however, technological developments and improved device structures result in devices for which external parasitics are reduced to comparable levels to intrinsic limitations[1, 2]. In the near future external parastics will be reduced even further to insignificant levels compared to intrinsic limitations, hence intrinsic limitations will become the bottleneck for the speed of operation. The main source of intrinsic limitations is the inevitable time delays in any finite sized device. But, when external parasitics are sufficiently reduced so that intrinsic time delays become the dominant limitations to high frequency performance, the resistances representing the drift zones can be negative over certain bands of frequencies. In order to demonstrate the origin of this negative resistance phenomena consider the generic model for an active device shown in Fig. 1(a). It consists of an injector, which injects charge into a drift zone of length L and cross sectional area A. Assume that the voltage across the device is kept sinusoidal by attaching a resonator across the device. A small fraction of this voltage drops across the injector, which injects particles into the drift zone. The small signal, or a.c., particle current, I_{p0}, entering this drift zone is delayed with respect to the total small signal current, I_t, i.e.,

$$I_{p0} = K I_t \exp(-j\phi), \qquad (1)$$

where ϕ is the injection angle which is given as $\phi = \omega\tau$, τ being the injection delay. The origin of this

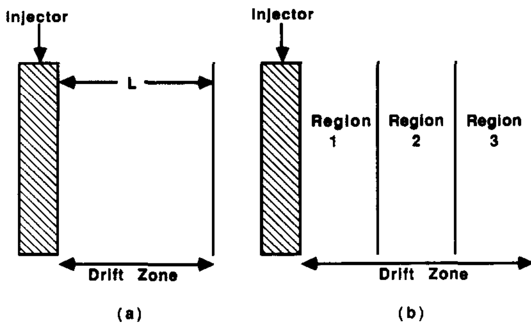

Fig. 1. (a) Schematic of a generic active device. (b) A negative resistance device in which the drift zone is partioned.

delay is related to the speed of response of the injector. K, which is usually a frequency dependent coefficient between 0 and 1, indicates what fraction of the total current entering the drift zone is particle current. Assuming that these injected carriers drift with a velocity, $v(x)$, in the drift region, the particle current at any position in this region is

$$I_p(x) = K I_t \exp(-j\phi)\exp\left[-j\omega \int_0^x dx'/v(x')\right] \quad (2)$$

The total small signal current through the drift zone is the summation of a displacement current, I_d, and the particle current. Then

$$I_t = I_p(x) + I_d(x) = K I_t \exp(-j\phi)\exp \times \left[-j\omega \int_0^x dx'/v(x')\right] + j\omega\epsilon A\, E(x) \quad (3)$$

where $E(x)$ is the small electric field in the drift zone. One can solve for $E(x)$ from eqn (3) and integrate $E(x)$ along the drift zone to find the small signal voltage across the drift zone. The ratio of this voltage to the total current is the small signal impedance of the drift zone. Carrying out the algebra and assuming a constant velocity v in the drift zone, one obtains that the real part of this impedance, R, which is the resistance across the drift zone, is

$$R = K\frac{\cos(\phi) - \cos(\phi + \theta)}{\omega\theta C A}, \quad (4)$$

where $C = \epsilon/L$ is the capacitance per unit area and $\theta = \omega L/v$ is the transit angle of the drift zone respectively. Since a lage fraction of the total voltage across the device drops across the dift zone, the impedance of the device will be dominated by the impedance of the drift zone. Hence R can be regarded as the resistance of the device. From eqn (4) one observes that if $\phi > 0$, i.e. if there is a finite injection delay, by adjusting the length or the transit angle of the drift zone one can achieve negative resistance. Utilizing this principle a class of two terminal transit-time devices were successfully generated. The most well known examples of such devices are IMPATT[3], BARITT[4] and TUNNETT[5] diodes. But the fact that negative resistance develops across a drift zone if the carriers entering the drift zone are delayed has broad implications. This can happen in any finite sized device due to inevitable intrinsic delays, since all active solid state devices have drift zones in which charge is injected utilizing various mechanisms and all these injection mechanisms have inevitable injection delays. Since negative resistance is a means of activity, its existence implies that intrinsic time delays can be taken advantage of rather than being regarded as limitations. The approach proposed in this work is to take advantage of this inevitable phenomena to generate a class of three-terminal devices with negative output resistance. Then, at least in principle, finding active devices that show negative output resistance is reduced to finding methods of delaying the particle current injected into a drift zone with respect to the total current in a controllable way. From eqn (4) it is obvious that the most desirable value for both ϕ and θ, which makes the magnitude of the negative resistance maximum, is π. One way of achieving this is to partition the drift zone. This way one can obtain a three-terminal device as well as introducing control on the injection delay as explained in the next section.

3. A HOT ELECTRON TRANSIT-TIME TRANSISTOR (HETT) BASED ON INJECTION AND TRANSIT-TIME DELAYS

Consider the device structure shown in Fig. 1(b). In this structure the drift zone after the injector consists of three separate regions. Carriers injected from the injector drift in regions 1 and 2 first and suffer a delay before entering to region 3. Even if the injection process in the injector is very fast, it is possible to create the desired injection delay into region 3 by properly designing the length of the regions 1 and 2. Due to this injection delay the resistance across the third region will become negative over certain bands of frequencies. If one can achieve the required injection delay keeping the total thickness of regions 1 and 2 smaller than thickness of region 3, the majority of the voltage across the device will drop across region 3. Hence the overall resistance of the device will be dominated by the resistance of region 3 and will be negative. Yet there is another significant advantage in dividing the drift zone into separate regions. If a barrier to the d.c. current flow between regions 2 and 3 is created, current flow between these two regions will be inhibited when a d.c. bias is applied. Then one can contact regions 2 and 3 separately and create a three-terminal device. Schematic of one possible device that satisfy all the requirements that the proposed structure should satisfy and its conduction band diagram in equilibrium are shown in Figs 2(a) and (b) respectively. In this example injector, which injects electrons, is an asymmetrical tunneling barrier, whereas an isotype heterojunction, a planar doped barrier or a resonant tunneling injector are other possibilities. Let's call the three terminals of this

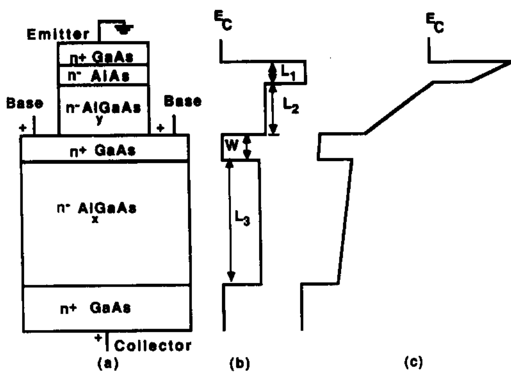

Fig. 2. A hot electron transit-time transistor (HETT) and its conduction band diagrams under equilibrium and operating bias conditions.

device emitter, base and collector respectively as shown in Fig. 2(a). A barrier in the conduction band exists between base and collector regions due to an isotype heterojunction. When the device is biased as shown in Fig. 2(a), the conduction band minimum will be as shown in Fig. 2(c). Now one can control the injection of electrons using base and emitter terminals. The presence of the barrier between base and collector regions hinders the flow of electrons from base to collector under d.c. bias conditions. This barrier will also be present for the electrons injected from the emitter. But, since the injected electrons are hot electrons, if the base region is made thin enough so that the transport through the base is mostly ballistic, a sizable fraction of the injected electrons can make it over the base–collector potential barrier into the base–collector drift zone[6]. In such structures collection efficiencies up to 82% with minimal collector leakage have experimentally been observed [7]. Even though such collection efficiencies result in poor β in the sense of conventional transistor operation, they are very satisfactory as far as injecting carriers into a drift zone after they suffer an injection delay. In this case the injection delay is the combination of the delay through the tunneling barrier and the delays due to the transit-time of the carriers in the first Al_y GaAs drift region and the time of flight of ballistic electrons in the base. Due to this injection delay resistance across the base collector drift region will become negative over certain frequency bands as explained in Section 2. With existing technology it is possible to control the thickness of drift regions precisely. The control over the thickness of the first drift region (L_2) is equivalent to engineering the injection delay. Similarly one can control the thickness of the second drift region (L_3) precisely, hence transit-time delay can also be engineered. Therefore, this structure permits the independent adjustment of the injection and transit-time delays, hence the value of the negative resistance and the efficiency of the device can be optimized. One can determine the frequency bands over which the negative resistance will exist using the results of the analysis in the previous section. Integrating eqn 3 along the drift zone one obtains

$$I_t = \left\{ \frac{1}{L} \int_0^L \exp\left[-j\omega \int_0^x \mathrm{d}x'/v(x') \right] \mathrm{d}x \right\} I_{p0} + j\omega \frac{\epsilon A}{L} V = \alpha(\omega) I_{p0} + j\omega CAV \quad (5)$$

where V is the small signal voltage across the drift zone. Equation (5) indicates that an equivalent circuit model for a drift zone is a parallel combination of a capacitor, whose value is the geometrical capacitance of the drift zone, and a current controlled current source. The controlling current for the controlled source is the particle current which is injected into the drift zone. The tunneling barrier can be modeled as a parallel RC combination, where the RC time constant is equal to the tunneling time across the barrier[8]. Using these models the small signal equivalent circuit representing the device becomes as shown in Fig. 3. C_1 is the geometrical capacitance of the tunneling barrier. The associated dynamic resistance value can be estimated using the RC product associated with the tunneling barrier[9]. For a 20 Å thick AlAs barrier a very conservative estimate of RC product is about 5×10^{-12} s. Current through the resistor, I_1, is the particle current that is injected into the first drift zone. τ_b and D are the time of flight and collection efficiency of the ballistic electrons through the base respectively. Particle current through the first drift zone is transported ballistically across the base and injected into the second drift zone. Hence only part of the particle current through the first drift zone enters the second drift zone after a certain delay. The factor $D\ \exp(-j\omega\tau_b)$ scaling the second controlled source accounts for this effect. R_b is the base resistance. α_2 and α_3 are current controlled current source coefficients for the first and second drift zones respectively, i.e.

$$\alpha_i = \frac{1}{L_i} \int_0^{L_i} \exp\left[-j\omega \int_0^x \mathrm{d}x'/v(x') \right] \mathrm{d}x \quad i = 2, 3 \quad (6)$$

An initial estimate of the device dimensions for very high frequency operation can be performed in the following way. At very high frequencies the tunneling barrier will introduce almost a phase delay of $\pi/2$. Delay due to ballistic transport of electrons through

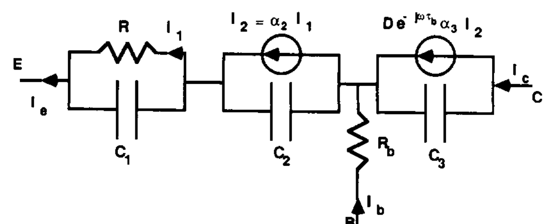

Fig. 3. Small signal equivalent circuit representing the device shown in Fig. 2.

the base is very small even in THz range. So in order to create a π injection delay an additional delay of $\pi/2$ is needed. This is provided by the first Al_yGaAs drift zone. The desired value for the transit time delay in the second drift zone is π. So assuming equal carrier velocities in both drift regions, the second drift region should be twice as long as the first one. Assuming a value for the carrier velocity the absolute values for the lengths of the drift zones can be determined. For lower frequency operation the length of the first drift zone should be slightly longer than the half of the length of the second drift zone to provide a delay larger than $\pi/2$ needed to bring the total injection delay to π. Starting with these initial estimates one can optimize the device dimensions for the frequency range of interest and determine the frequency bands over which the negative output resistance will exist using the equivalent circuit. In defining an output impedance to the device, however, one should be careful, since the device has three terminals and the output impedance depends on the input termination. In this case it is best to define the output impedance in a way which is consistent with the definitions of the two port device parameters, i.e. z, y or h parameters. The output impedance obtained this way can be used in the gain expressions given in terms of two port device parameters, hence the circuit performance of the device can be related to the device physics. Then one immediately observes that for a common emitter configuration the output resistance that needs to be considered is r_{22}, which is the resistance seen between the emitter and collector (output) when the base (input) is open circuited in the small signal sense. Variation of the output resistance and device active area product of a device for which $L_2 = 0.5\ \mu m$ and $L_3 = 1.0\ \mu m$ are shown in Fig. 4. In the calculations for both AlGaAs regions an average drift velocity of 5×10^6 cm/s is used. Basewidth (W) is 500 Å, for

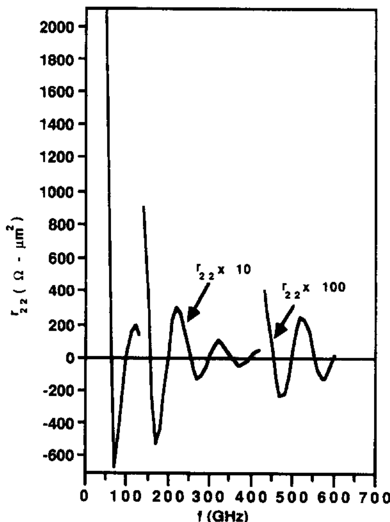

Fig. 4. Variation of the output resistance of a hot electron transit-time transistor (HETT).

which τ_b and D can be estimated as 0.05 ps and 0.6 respectively [6]. For this device the output resistance changes its sign frequently and there are many frequency bands over which output resistance is negative. In the lower end of the spectrum the magnitude of the negative resistance is very high. For example one can generate a negative output resistance of $-66\ \Omega$ from such a device of $10\ \mu m^2$ active area at 70 GHz. Such high magnitudes of negative resistances are quite desirable to offset the other inevitable device parasitic resistances due to ohmic contacts and substrates and to make external circuit design easier. As frequency increases magnitude of the negative resistance decreases due to the intrinsic capacitances associated with the drift zones and the tunneling barrier. One can offset this disadvantage to a certain degree by decreasing the device active area. For example a device of $1\ \mu m^2$ cross sectional area will have an output resistance of $-5\ \Omega$ at 370 GHz, $-2.4\ \Omega$ at 470 GHz and $-1.3\ \Omega$ at 570 GHz. The frequency at which output resistance becomes negative can be pushed higher and higher by decreasing the lengths of the two drift zones. The lowest frequency at which output resistance becomes negative, f_l, can be estimated in the following way. The injection and transit angles (ϕ and θ) should each be π as discussed previously and naturally should add up to 2π, which is a complete cycle of oscillations. But $\phi = \omega(\tau_e + \tau_b)$ where $\tau_e = \tau_w + \tau_{d1}$ is the addition of delays due to tunneling barrier (τ_w) and the transit-time of the carriers in the first Al_yGaAs drift region (τ_{d1}), and $\theta = \omega\tau_c$, where τ_c is the transit-time of the carriers in the base collector drift region. Hence one obtains

$$f_l = \frac{1}{\tau_e + \tau_b + \tau_c}. \tag{7}$$

Utilizing eqn (7) one can predict that f_l can be pushed well into THz range by decreasing the lengths of the drift zones, since τ_{d1} and τ_c are directly proportional to the lengths of drift zones. Preliminary simulations also indicate that negative output resistance, hence activity, can be obtained well into THz range. The analysis presented so far is valid as far as the intrinsic operation of the device is concerned. The external parasitics associated with the device were neglected altogether assuming that they are reduced sufficiently by proper design, such as utilizing self aligned processes and non-alloyed graded gap ohmic contacts. This is not a fundamental limitation, however, because any kind of an active device that is expected to work at such high frequencies should have these properties. One can also improve the performance of the device utilizing other compound semiconductors like InGaAs/InAlAs or InAs/AlSb system which have higher conduction band offsets[10]. High conduction band offsets will help to eliminate the thermionic emission currents and increase the isolation between input and output.

4. IMPLICATIONS OF THE NEGATIVE OUTPUT RESISTANCE

When the output resistance of a three-terminal device is negative the device is clearly active. The power gain that can be obtained from such a device can be calculated using the gain expressions for two ports. As an example consider the unilateral power gain expression in terms of the generalized two port device parameters, which is[11],

$$U = \frac{|k_{21}|^2}{4\,Re\,[k_{11}]\,Re\,[k_{22}]}, \tag{8}$$

where k parameters stand for any one of the h, y or z parameters. When there is a possibility, however, of negative output resistance one should be very careful in using the various gain expressions that exist in literature, such as the unilateral gain which is being considered presently. In the derivation of U, for example, it is assumed that $Re\,[k_{22}]$, (output resistance, r_{22}, or conductance, g_{22}) is positive and input and output ports are conjugately matched. This requires that load resistance must be equal to output resistance. However, if the output resistance is negative this condition cannot be satisfied, because negative resistive terminations do not exist. Over the frequency ranges where the output resistance is negative the unilateral power gain of the device should be calculated utilizing the formula obtained using a realistic positive load resistance, hence eqn (8) becomes

$$U = \frac{|k_{21}|^2}{2\,Re\,[k_{11}]\,[K_L - |Re\,[k_{22}]|]} \tag{9}$$

where K_L is either a load resistance or conductance depending on the type of two port device parameters utilized. This expression demonstrates that it is possible to fabricate amplifiers with high gain or even oscillators by choosing a suitable load resistance. Another significant advantage is due to the fact that in the proposed structure there are three terminals and the input and the output are naturally isolated. As a result one can design fully integrated circuits without the need to use nonreciprocal circuit components like circulators. This is a significant advantage compared to two terminal negative resistance devices, like IMPATT's or Gunn's, where input and output isolation requires external circuit elements like circulators. Hence it is very difficult to fabricate fully integrated two terminal negative resistance amplifiers for millimeter-wave applications.

At this stage one must point out that, in the past there have been other three terminal device ideas that exhibit negative output resistance. The principle behind these device ideas was to add a third terminal to a Gunn or an IMPATT[12]. The first device was called an emitter controlled negative resistance device [13] and the second one is called IMPISTOR (IMPATT Transistor)[12] or CATT (Controlled Avalanche Transit Time Triode)[14]. The function of the third terminal in these devices was to control the current injected into the Gunn or IMPATT to control the domain formation or impact avalanche. The resulting negative resistance was due to the Gunn effect or impact ionization delay rather than the intrinsic delays in the emitter, base and collector regions. Indeed none of these device ideas took advantage of the intrinsic delays that inevitably exist in any finite sized device. As a consequence of this principle negative output resistance should also be seen in transistors (bipolar or field-effect), which have drift zones in which carriers are injected after they suffer an intrinsic delay. Existence of the negative output resistance in a bipolar transistor was predicted both based on device physics[15] and equivalent circuit simulations[16]. However, presence of junction capacitances and minority carrier storage effects limit the upper frequency of operation of a bipolar transistor in this mode of operation to lower values than the ones predicted for the proposed structure.

5. CONCLUSIONS

In this work a new transistor with negative output resistance, called hot electron transit-time transistor (HETT), is introduced and its small signal analysis is carried out. The basic principle of operation of this transistor is to take advantage of the intrinsic time delays that exist in any finite sized device. It is shown that a HETT with vertical dimensions at the order of a micrometer starts to exhibit negative output resistance around 70 GHz. If the device is properly engineered it is possible to create negative output resistance extending all the way to submillimeter-wave frequencies utilizing intrinsic delays. This principle of operation is completely different than the conventional mode of operation where transit-times are seen as limitations. Negative output resistance should also be observed in transistors (bipolar or field effect) and can be utilized to enhance the gain due to conventional transistor action.

Acknowledgements—The author is grateful to Dr Joseph Maserjian of JPL for stimulating discussions and to the UC Regents for a Junior Faculty Fellowship.

REFERENCES

1. P. M. Asbeck, M. F. Chang, K. C. Wang, D. L. Miller, G. J. Sullivan, N. H. Sheng, E. Sovero and J. A. Higgins, *IEEE Trans. Electron Devices* **34,** 2571 (1987).
2. L. D. Hguyen, P. T. Tasker, D. C. Radulescu and L. F. Eastman, *Technical Digest of International Electron Devices Meeting, San Fransisco, CA*, p. 176 (1988).
3. W. T. Read, *Bell Syst. tech. J.* **37,** 401 (1958).
4. W. Harth and M. Claassen, *Nachrichtentechnishe Z.* **2,** 87 (1973).
5. J. Nishizawa, K. Motoya and Y. Okuno, *IEEE Trans. microw. Theory Tech.* **20,** 1029 (1978).
6. M. Heiblum, M. I. Nathan, D. C. Thomas and C. M. Knoedler, *Phys. Rev. Lett.* **55,** 2200 (1985).

7. J. Chen, U. K. Reddy, D. Mui, C. K. Peng and H. Morkoç, *Appl. Phys. Lett.* **51,** 1254 (1987).
8. K. K. Thornber, Thomas C. McGill and C. A. Mead, *J. appl. Phys.* **38,** 2384 (1967).
9. G. Lewicki and C. A. Mead, *Phys. Rev. Lett.* **16,** 939 (1966).
10. A. F. J. Levi and T. H. Chiu, *Appl. Phys. Lett.* **51,** 984 (1987).
11. R. S. Carson, *High-Frequency Amplifiers.* Wiley, New York (1975).
12. J. E. Carroll, *Inst. Phys. Conf. Ser. No. 25, Solid State Devices*, p. 109 (1974).
13. M. M. Atalla and J. L. Moll, *Solid-St. Electron.* **12,** 619 (1969).
14. S. P. Yu, W. Tantraporn and J. R. Eshbach, *IEEE Trans. Electron Devices* **23,** 332 (1976).
15. N. Dagli, *IEEE Electron Device Lett.* **9,** 113 (1988).
16. N. Dagli, W. Lee, S. Prasad and C. G. Fonstad, *IEEE Electron Device Lett.* **8,** 472 (1987).