Bipolar tunneling field-effect transistor: A three-terminal negative differential resistance device for high-speed applications

J. P. Leburton, J. Kolodzey, and S. Briggs
Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois, Urbana, Illinois 61801

(Received 5 August 1987; accepted for publication 3 March 1988)

Carrier injection across a tunnel homojunction is suggested as a new mechanism for a high-speed three-terminal device. The novel feature is the two-dimensional homojunction tunneling within a bipolar modulation doping structure. Negative differential resistance characterized by large peak-to-valley current ratios and high transconductance is anticipated. Estimates of the relevant time constants of the tunnel structure suggest the possibility of very high frequency operation.

The first observation of resonant tunneling (RT) in unipolar heterostructures by Chang *et al.*[1] has stimulated intensive work in the field. Negative differential resistance (NDR) characteristics produced by tunnel injection into confined-particle states of a quantum well structure embedded in a *p-n* junction have been subsequently demonstrated by Holonyak *et al.*[2,3] Meanwhile, interest in RT heterostructures has been motivated by the advances in fabrication techniques and the expectation of extremely high frequency responses.[4] State-of-the-art RT diodes have demonstrated operation up to 200 GHz.[5] However, many desired functions and widespread digital and microwave applications of these structures will be achieved only with three-terminal (3T) tunneling devices. Various attempts have been made to conceive RT structures where the onset of NDR can be externally controlled by a third terminal.[6-8] However, performances of the RT 3T heterobarrier devices seem to be limited by the background current, thermionic emission and high-energy tunneling, which smears out the NDR and requires low-temperature operation.[6-9] Recently, promising approaches using bipolar technology for room-temperature operation have been suggested by Capasso *et al.*[10] and other groups.[11]

As an alternative to existing RT heterojunction devices, we propose a new tunnel device which combines a modulation-doped (MD) structure with a single tunnel homojunction,[12,13] and has the possibility of operating as an ultrafast NDR three-terminal device. Novel features of the device we describe are that carrier injection in the tunnel junction and the onset of NDR are externally controlled by field-effect modulation in the MD structure. Because both electrons and holes are involved in the conduction mechanism of the *p-n* junction, the tunnel field-effect transistor (FET) is bipolar.

There are several advantages of using a tunnel homojunction with the MD structure: First, as shown below, the NDR transitions are expected to be abrupt (small NDR values) and characterized by large peak-to-valley current ratios. This is due to the two-dimensional (2D) density of states in the confining channels, which causes current variations to be sharper than in 3D tunnel junctions. Small NDR values suggest high-frequency response and high-power operation. Second, it seems possible to reduce the background diffusion current more effectively than in unipolar RT heterostructures. This is because the tunnel process in homojunctions permits electron and hole injection only within the energy range corresponding to the overlap between conduction and valence bands. In addition, because of the adjacent heterojunctions with large-gap MD materials, the conventional diffusion (thermionic) current can be strongly reduced over the range of NDR operation. These features offer the possibility of high-performance room-temperature operation. Finally, fabricating the 3T NDR device appears feasible with conventional lithography and ion implantation.

A schematic representation of the bipolar tunneling FET (BiTFET) structure is shown in Fig. 1. The configuration is very similar to the heterojunction field-effect transistor photodetector (HFETPD) recently fabricated by Taylor and Simmons.[14] The lateral doping profile can be obtained using ion implantation which has been shown to have a resolution below 500 Å.[15] A key feature is the tunnel junction of a small-gap semiconductor (GaAs or InGaAs) which consists of two heavily doped layers of opposite conductivity types sandwiched vertically between two lightly doped regions of a large-gap material (AlGaAs or InP). Thin undoped layers of GaAs (InGaAs) are intercalated between the tunnel junction and the two heterojunctions. At equilibrium the two heavily doped layers are totally depleted by forming the space-charge region of the junction, but also by backward diffusion of carriers toward the undoped layer. This mechanism substantially increases the channel concentration with respect to the normal value resulting from the modulation doping mechanism. With this extra modulation doping mechanism, the two (electron and hole) channels are degenerate and form extended lateral 2D electrodes for the tunnel junction. Selective contact electrodes connect with the channels of identical conductivity types while forming a blocking contact for opposite-type carriers. The lightly doped vertical n^- and p^- layers result from the grading of the implantation profile and improve the isolation of each channel from the unlike electrode and prevent leakage at these junctions. Under forward bias, the current flows from the selective contacts to the channels with low resistance. Owing to the carrier degeneracy in the two channels, electrons tunnel through the junction at low bias by a process similar to the tunnel mechanism leading to the NDR in conventional tunnel diodes. An important feature of the device configuration is the 2D nature of the quantum tunneling with step-function initial and final carrier density of states. This causes abrupt changes in the tunnel current when 2D electron and hole levels cease to overlap [Fig. 1(b)]. Consequently, the device will exhibit multiple NDR's with arbi-

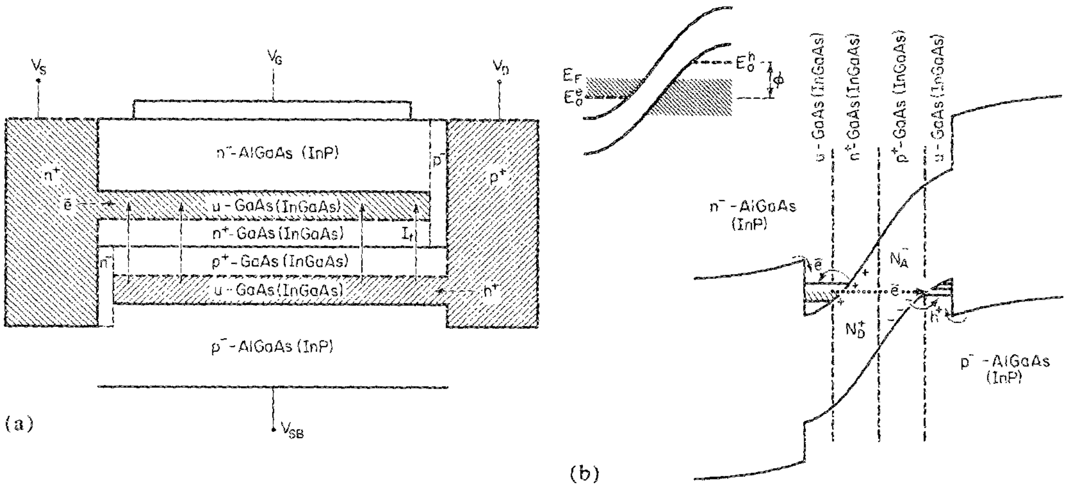

FIG. 1. (a) Schematic representation of a BiTFET structure. Electrons tunnel from the 2D channel connected to the n^+ electrode (source) toward the hole channel connected to the p^+ drain across a thin tunnel diode. Vertical n^{-} and p^{-} isolation layers are preimplanted. (b) Cross-sectional energy diagram of the tunnel active region. Electron and hole 2D degenerate channels are caused by normal modulation doping and diffusion of carriers from residual p^+ and n^+ doped junction layers. The dotted arrow represents the tunneling from quantized levels. Inset shows the position of the electron and hole fundamental levels.

trarily small resistance values. In practice, the latter is determined by the peak current density flowing through the junction [see Eq. (3)]. With this device configuration, a gate-substrate reverse bias may be used to enhance the current by increasing the separation ϕ between the electron and hole ground levels, $\phi = E_0^h - E_0^e$, since there is more overlap between electron and hole energies for tunneling across the junction. This situation is new and unique because in conventional two-terminal tunnel junctions the characteristics are determined by the level and profile of the doping which are process dependent and consequently inalterable. Light doping in the large-gap materials reduces the space-charge field in the regions adjacent to the *p-n* junction and prevents Fowler–Nordheim tunneling across the heterojunctions. There is thus advantage in using heterojunctions with large band-gap offsets such as the new pseudomorphic AlGaAs/InGaAs structures recently proposed to enhance the channel carrier density.[16]

Figure 2 shows a calculation of the tunnel current as a function of the drain voltage for different values of ϕ. A 2D transfer Hamiltonian method[17] has been used to compute the tunnel current and reproduces the essential features of Zener tunneling. For the sake of simplicity, the 2D junction has been assumed to be symmetrical with parabolic confining potential in the channel. This is in contrast with quasi-triangular potentials in FET channels, but is justified by the strong carrier degeneracy combined with low surface fields resulting from low MD in AlGaAs. J_{t0} is a constant which depends on the characteristics of the junction. The main features here are the multiple abrupt NDR transitions which occur when 2D electron and hole tunneling states cease to overlap. Similar abrupt transitions have been obtained by Tabatabaie and Tamargo.[18] Calculations show that because of channel degeneracy, the tunnel current is a sensitive function of V_{DS} and ϕ, and varies by several orders of magnitude with voltage. This results in large peak-to-valley ratios which decrease with ϕ.

The speed performances of the BiTFET are determined by the time constants associated with various processes controlling the transport in the structure.[1,19] It is well known that the tunneling time is essentially a function of the level broadening ΔE resulting from the interaction between electronic states across the junction,[18,20] i.e.,

$$\tau_{tun} = \pi\hbar/\Delta E. \qquad (1)$$

In strong junction fields and low-gap materials such as InGaAs, ΔE resulting from the overlap of electron and hole wave functions can reach 2–3 meV so that τ_{tun} is about 1–0.5 ps. However, the charging time of each 2D layer is also of importance and can introduce some limitations in the device high-frequency operation. For 2D degenerate channels and a quarter μm source-drain spacing, layer specific resistances ($R_t \times$ device width) as low as $10^{-2}\ \Omega$ cm can be achieved at room temperature where the mobility difference between electrons and holes can be offset by doping the *p* layer more heavily than the *n* layer in the tunnel junction. This is fortunately possible with beryllium doping in GaAs. Among the various capacitances associated with the tunnel structure, the junction capacitance C_j is certainly the most critical for this time constant. With a distance $d = 150$ Å between the two heterojunctions, C_j is about 17 pF/cm; the gate capacitance C_G is three times less with a top AlGaAs epilayer of 500 Å, so that the total charging time is below 1 ps.

At constant reverse gate bias, the 2D junction can be driven into the NDR regime by applying the appropriate source-drain voltage and operating essentially as a 2D tunnel diode. The cutoff frequency is given by[21]

$$f_{co} = (2\pi R_{NDR} C_j)^{-1}\sqrt{R_{NDR}/R_s - 1}, \qquad (2)$$

where R_{NDR} is the NDR value and R_s is the series resistance

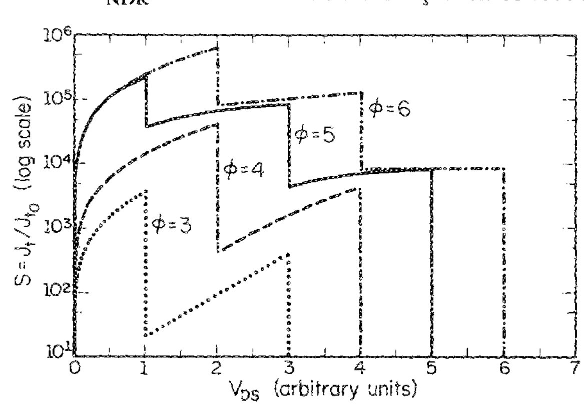

FIG. 2. Normalized current J_t/J_{t0} voltage characteristics of the 2D tunnel junction for various values of $\phi = E_0^h - E_0^e$ dotted line, $\phi = 3$; dashed line, $\phi = 4$; solid line, $\phi = 5$; and dot-dashed line, $\phi = 6$. The voltage and energy are expressed in units of the energy quantum of the channel parabolic potential $\hbar\omega$. $T = 300$ K.

of the junction. Owing to the 2D level broadening ΔE, R_{NDR} is not infinitely small, but of the order of

$$R_{NDR} \sim \Delta E / q \, \Delta J_{pv} \, A \, , \tag{3}$$

where ΔJ_{pv} is the peak-to-valley current density difference and A is the junction area. Optimum values of $\Delta J_{pv} \gtrsim 10^3$ A/cm^2 can be achieved with use of low-gap materials[21] (e.g., InGaAs) for the junction and external control of the tunneling injection by the gate field through the channel density. This latter advantage is a unique feature of the 3T device where the control of the input current is separated from the tunnel characteristics. With a gate width of 100 μm, we estimate $R_{NDR} \sim 8 \Omega$. R_s has contribution from contacts and spreading resistance, and is typically 0.7 Ω, but can be reduced to half this value.[22] Therefore, we obtain $f_{co} \sim 300$–500 GHz.

Because of the predicted high magnitude of the NDR peak-to-valley ratio, the BiTFET can also operate as a switch between high current level (pre-NDR transition, ON state) and low current level (post-NDR transition, OFF state). In this mode, the source-drain voltage is biased in the NDR region, and the gate voltage switches the device between ON and OFF states. Owing to the 2D nature of the tunneling states, only tiny variations of the gate voltage are required to shift the quantized levels with respect to one another during tunneling and give rise to the NDR. This results in large current gain and extremely fast switching. The current response time is given by[21]

$$\tau_{resp} = C_G / g_m (V_{NDR}), \tag{4}$$

where $g_m(V_{NDR})$ is the transconductance at $V_{DS} = V_{NDR}$ which for this process is controlled by the current in the n channel. τ_{resp} is obtained from the cutoff frequency at which the input gate current equals the output drain current. This represents a condition of unity incremental current gain. One can write

$$g_m (V_{NDR}) = \frac{dI_{DS}(V_{NDR})}{dV_G} = \frac{dI_{DS}}{d\phi} \frac{d\phi}{dV_G} \, . \tag{5}$$

For a current variation corresponding to ΔJ_{pv}, the first term is equal to $1/qR_{NDR}$, since $d\phi = dV_{DS}$ in the limit of vanishing series resistance. On the other hand, $\Delta\phi = -\Delta E_0^e$, if E_0^h is kept constant. By assuming that under strong degeneracy most of the electrons are in the lowest level, the expression of the 2D density of states yields

$$E_f - E_0^e = \pi \hbar^2 n_s / m^*, \tag{6}$$

where E_f is the Fermi level and n_s is the 2D electron density. Combining all these equations, one obtains

$$g_m (V_{NDR}) = \frac{\pi \hbar^2}{q m^* R_{NDR}} \frac{dn_s}{dV_G} = \frac{\pi \hbar^2}{q^2 m^* R_{NDR}} \frac{C_G}{A} \, , \tag{7}$$

with $C_G / A = q dn_s / dV_G$ and $q^2 m^* / \pi \hbar^2$ is an elementary unit of 2D layer capacitance.[23] Using Eqs. (1) and (3), the final expression for the current response time is given by

$$\tau_{resp} = \frac{q m^*}{\hbar \tau_{tun} \, \Delta J_{pv}} \text{ or } \tau_{resp} \tau_{tun} = \frac{q m^*}{\hbar \, \Delta J_{pv}} < 10^{-23} \text{ s}, \tag{8}$$

for $m^* = 0.05 m_0$ in InGaAs junctions. Ultimately, the current response time is determined by the magnitude of the tunneling injection, which causes large values of ΔJ_{pv} as shown in Fig. 2. Extremely short time is expected if ΔJ_{pv} exceeds 10^3 A/cm^2 and would correspond to unusual large transconductance $[g_m(V_{NDR}) \sim 10^3$ mS/mm$]$ at the NDR transition. This is due to the slight V_G variation which is required to move the relative position of the n and p tunneling states, thereby causing large current variation without the need to redistribute the channel charge.

In conclusion, the proposed BiTFET structure is a novel device configuration. Detailed calculations indicate NDR operation and the possibility of high-speed performance. The predicted features are large peak-to-valley current ratios and small NDR values due to the 2D nature of the tunnel states. Ultimately the device speed is limited by the magnitude of the tunnel current between the two channels. In actual devices, the tunnel current will critically depend on the junction band gap, the band offsets at the heterojunctions, and the doping level achieved in the junction layers.

We would like to thank K. Hess for encouragement. We are indebted to N. Holonyak, Jr., A. Y. Cho, R. J. Fischer, D. Sivco, and B. Schwartz for critical comments and fruitful discussions. This work was supported by the Joint Services Electronics program and the National Science Foundation Engineering Research Center for Compound Semiconductor Microelectronics.

[1] L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
[2] E. A. Rezek, N. Holonyak, Jr., B. A. Vojak, and H. Shichijo, Appl. Phys. Lett. **31**, 703 (1977).
[3] B. A. Vojak, S. W. Kirchoefer, N. Holonyak, Jr., R. Chin, R. D. Dupuis, and P. D. Dapkus, J. Appl. Phys. **50**, 5830 (1979).
[4] F. Capasso, K. Mohammed, and A. Y. Cho, IEEE J. Quantum Electron. **QE-22**, 1853 (1986).
[5] E. R. Brown, T. C. L. G. Sollner, W. D. Goodhue, and C. D. Parker, IEEE Trans. Electron Devices **ED-34**, 2381 (1987).
[6] N. Yokoyoma, K. Imamura, S. Muto, S. Hiyamizu, and H. Nishi, Jpn. J. Appl. Phys. **24**, L 853 (1985).
[7] S. Luryi and F. Capasso, Appl. Phys. Lett. **47**, 1337 (1985); **48**, 1693 (1986).
[8] B. Vinter and A. Tardella, Appl. Phys. Lett. **50**, 410 (1987).
[9] F. Capasso, S. Sen, F. Beltram, and A. Y. Cho, Electron. Lett. **23**, 225 (1987).
[10] F. Capasso, S. Sen, A. C. Gossard, A. L. Hutchinson, and J. H. English, IEEE Electron Device Lett. **EDL-7**, 573 (1986).
[11] T. Futatsugi, Y. Yamagushi, K. Imamura, S. Muto, N. Yokoyama, and A. Shibatomi, Jpn. J. Appl. Phys. **26**, L131 (1987).
[12] L. Esaki and Y. Miyahara, Solid-State Electron. **1**, 13 (1960).
[13] N. Holonyak, Jr. and I. A. Lesk, Proc. IRE **48**, 1405 (1960).
[14] G. W. Taylor and J. G. Simmons, Appl. Phys. Lett. **50**, 1754 (1987).
[15] D. B. Rensch, D. F. Matthews, M. W. Utlaut, M. D. Courtney, and W. M. Clark, Jr., IEEE Trans. Electron Devices **ED-34**, 2232 (1987).
[16] T. E. Zipperian, L. R. Dawson, C. G. Osbourn, and I. J. Fritz, in International Electron Devices Meeting, Technical Digest, Washington DC, 1983, p. 696.
[17] See, e.g., C. B. Duke, in *Tunneling in Solids*, Solid State Physics, Suppl. 10, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), p. 207.
[18] N. Tabatabaie and M. C. Tamargo, in International Electron Devices Meeting, Technical Digest, Los Angeles, 1986, p. 80.
[19] T. C. L. G. Sollner, E. R. Brown, W. D. Goodhue, and H. Q. Le, Appl. Phys. Lett. **50**, 332 (1987).
[20] E. O. Kane, in *Tunneling Phenomena in Solids*, edited by E. Burstein and S. Lindqvist (Plenum, New York, 1969).
[21] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981), pp. 534 and 343.
[22] J. J. Coleman (private communication).
[23] In the course of writing this letter, Luryi has reported a similar expression: S. Luryi, Appl. Phys. Lett. **52**, 501 (1988).