Ultrafast Operation of Heterostructure Bipolar Transistors Resulting from Coherent Base Transport of Minority Carriers

Serge Luryi

AT&T Bell Laboratories, Murray Hill, NJ 07974

Introduction

The possibility of using transit-time phase shifts in a bipolar junction transistor to extend the active transistor operation to higher frequencies has been discussed by a number of researchers. The discussion has been centered on the negative dynamic output impedance that can be achieved due to the phase delay in carrier drift across the collector space-charge region. This idea, proposed originally by Wright [1], is attractive for millimeter and submillimeter wave applications. It has gained a lot of attention in the context of a conventional junction transistor implementation [2,3], heterostructure bipolar transistors (HBT) [4-6], unipolar ballistic transistors [7], and field-effect transistors [8]. Unfortunately, in transistors with diffusive base transport, the practical possibility of extending the active transistor behavior beyond f_T is severely limited [6] by parasitic effects. Utilization of transit-time resonances, arising from carrier drift across the collector junction in a conventional GaAs/AlGaAs HBT, requires a reduction of the base and collector resistances, as well as the collector capacitance, by factors of 10 from estimates, based on the state-of-the-art technology.

This problem was recently analyzed by Grinberg and myself [9]. Like in transit-time diodes, the negative dynamic output impedance in transistors arises from a combination of the injection phase delay $\varphi = \varphi_E + \varphi_B$ (the total transit angle of the emitter and the base) and the drift delay θ in the base-collector junction. Transit-time effects are swamped by the device parasitics because in diffusive transistors the available phase φ is too small. Exponential decay of the magnitude of the base transport factor α_B with the phase φ_B, acquired in a diffusive propagation across the base, implies that the overall phase of the collector current $(\varphi_B + \varphi_E + \theta)$ must be acquired almost entirely at the expense of the collector phase delay θ. However, with increasing θ the magnitude of the negative dynamic resistance also decreases, and for $\theta \approx \pi$ it becomes too small to overcome the influence of parasitic elements.

Transit time resonances can withstand the parasitics only if they can be achieved while keeping the collector delay $\theta \lesssim 1$, i.e. if a sufficient phase $\omega\tau_B \equiv \varphi_B \gtrsim \pi$ is acquired in the base transport alone – without a significant penalty in the magnitude of the base transport factor α_B. In general, with increasing frequency ω, the base transport factor $\alpha_B(\omega) = |\alpha_B| \exp(-\omega\tau_B)$ describes an inbound spiral $|\alpha_B| \to 0$ in the complex plane [10]. We shall apply the term ***coherent***, to any base transport mechanism that provides a sufficiently ***slow*** spiraling in of $\alpha_B(\omega)$, such that $\alpha_B \gtrsim 0.5$ for $\varphi_B \gtrsim \pi$.

One possibility to achieve coherence arises from ballistic transport at cryogenic temperatures [9]. In the ballistic case the modulated signal injected at the base-emitter interface is washed out because of the thermal spread in the normal velocities of injected carriers, leading to a variance $\delta\tau$ in their base propagation time. The latter process is analogous to the Landau damping of density waves in collisionless plasmas and has an effect similar to diffusion. Collisionless base propagation by itself does not imply coherence. Coherent regime in a ballistic transistor arises when $\delta\tau \ll \tau_B$, i.e., when the injected electrons form a collimated and monoenergetic beam. A good approximation to such a beam results from the passage of electrons across an abrupt heterointerface at low temperatures. For electrons traveling with a velocity v perpendicular to the base layer, a periodic injection modulation at the emitter interface with a frequency f sets up an electron density wave of wavelength $\lambda = v/f$. For $\delta\tau \ll \tau_B$, this wave does not appreciably decay over the entire base, resulting in a transport factor

$$\alpha_B = e^{-(\omega\delta\tau)^2/2}\, e^{-i\omega\tau_B} \quad . \qquad (1)$$

As discussed in detail in Ref. [9], a coherent transistor can have both the common-emitter current gain h_{21} and the unilateral power gain U exceeding unity at frequencies far above the usual f_T. The transistor speed is limited not by the base propagation time τ_B but rather by its variance $\delta\tau$.

Calculations carried out for exemplary heterostructures, implemented at the state-of-the-art rules of technology, indicate that active transistor behavior can be extended to about 1 THz.

It is clear, however, that the ballistic coherent operation requires cryogenic temperatures. For a collisionless transport to hold over the entire base width, the electron kinetic energy Δ (the injection energy, corresponding to a conduction-band discontinuity in the base-emitter junction [11]) should not exceed the optical phonon emission threshold, $\Delta \lesssim \hbar\omega_{\rm opt}$. On the other hand, to achieve coherence, the injection energy Δ must substantially exceed the thermal spread, whence we need $\hbar\omega_{\rm opt} \gtrsim \Delta \gg kT$. This means that the implementation of base transport coherence in a ballistic HBT requires at least liquid N_2 temperatures. Moreover, the ballistic coherent operation is limited to ultra-high frequencies. The concept cannot even be tested at lower frequencies, because the first resonance in U appears only at $f_\pi \approx \pi f_{\rm T} = 1/2\tau_{\rm B}$, which must evidently be higher than the collision rate $1/\tau_{\rm coll}$ governing the momentum relaxation of ballistic electrons at energy Δ.

Another idea for achieving base transport coherence (which is not limited to ultra-high frequencies and also works at room temperature) can be traced back to the famous paper by Shockley [12]. In that paper he introduced the concept of transit-time diodes and suggested that the delay in minority-carrier transit across a transistor base can lead to an active device at extended frequencies. A necessary condition for this to occur is that the directed transport across the base be much faster than the diffusive transport, that tends to wash out a modulated structure of the injected distribution. Shockley suggested [12] that this condition can be met in a minority-carrier delay diode with a variable doping in the base. However, because of a limited range of the potential variation available with an exponentially graded doping, the feasibility of this approach is marginal and it has never been realized. Recently, Shockley's argument was reconsidered [13] in the context of HBT with a graded alloy base composition. It was shown that coherence of the base transport in such devices is feasible and may lead to useful applications. In addition to transistors with a linearly graded base, we discussed a new a structure in which the base bandgap narrows down toward the collector in N discontinuous steps of minimum height larger than $\hbar\omega_{\rm opt}$. In this structure the minority transport occurs by a ***strongly accelerated diffusive process***, adequate for achieving the transit time resonance both at ultra-high and conventional frequencies. Our discussion below is based on Ref. [13].

Step Base Transistor

Consider a HBT whose base bandgap narrows down toward the collector in a stepwise fashion, Fig. 1. Let the band edge for minority carriers in the base consist of N steps W_j which are high enough, $\Delta_j \gg kT$, that carriers are effectively forbidden to return once they have fallen off a particular step. Moreover, let us assume that all the memory of the previous journey is lost in each step. The latter condition is reasonable if the steps are higher than the threshold for a rapid inelastic process, e.g., optical phonon emission: $\Delta_j \gtrsim \hbar\omega_{\rm opt}$. With this condition fulfilled, we can treat the transport in each step individually and characterize it by a step transport factor α_j,

$$\alpha_j(\omega) = \frac{1}{\cosh[(2i\varphi_j)^{1/2}]}, \tag{2a}$$

where $\varphi_j \equiv \omega\tau_j$ is the phase acquired on step j and $\tau_j = W_j^2/2D$ is the step propagation time by diffusion. We assume that the steps are narrow enough that $\varphi_j \ll 1$. In this case, to within quadratic in φ_j terms, Eq. (2a) reduces to the form

$$\alpha_j(\omega) \approx e^{-\varphi_j^2/3}\, e^{-i\varphi_j}. \tag{2b}$$

We see that at small values of φ_j, the magnitude $|\alpha_j|$ deviates from unity ***quadratically*** in φ_j. This observation enables us to construct a coherent $\alpha_{\rm B}$ with a large phase and little decay in magnitude. Indeed, the total base transport factor is the product of α_j's:

$$\alpha_{\rm B}(\varphi) = \prod_{j=1}^{N} \alpha_j \approx e^{-\varphi^2/3N}\, e^{-i\varphi}, \tag{3}$$

where $\varphi = \sum \varphi_j$ is the overall phase acquired in the base transport; we have assumed for simplicity that all steps are equal and $\varphi = N\varphi_j$.

The intrinsic current gain $\beta_B = \alpha_B (1-\alpha_B)^{-1}$ will have peaks $|\beta_B| > 1$ at $\varphi = 2\pi m$, provided

$$|\alpha_B(2\pi m)| > 0.5 \quad . \tag{4}$$

The first peak $(m=1)$ occurs at frequency $f = 2\pi f_T$. According to Eq. (3), this requires $N > N_{2\pi} = (2\pi)^2/3\ln 2 \approx 19$.

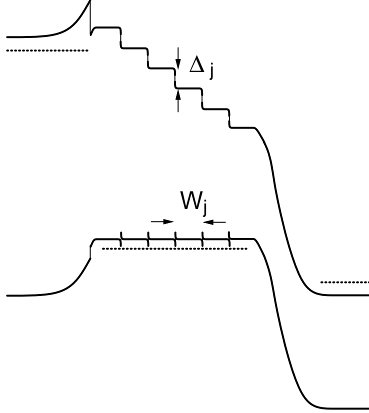

Figure 1. Schematic band diagram of a step-base heterostructure bipolar transistor.

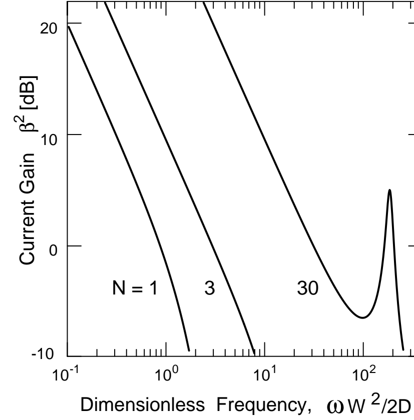

Figure 2. Common emitter current gain β for different number N of steps in the base.

Physically, the described effect originates from an enhancement of the ***forward*** diffusion transport. The essential condition is carrier thermalization at every step, which provides the independence of α_j's and restricts particles from returning to a preceding step. The resultant "stepped up" diffusion is a rather peculiar, at first glance even counter-intuitive, process. Thus, in a ***static*** regime, the minority concentration in the base is a ***periodic*** (rather than decreasing) function of the distance. Indeed, the steady-state current in each step is

$$J = \frac{e\,[\,n(0_j) - n(W_j)\,]}{D\,W_j} \approx \frac{N\,e\,n(0)}{D\,W} \ , \tag{5}$$

where $W = NW_j$ is the total base thickness and $n(0_j)$ is the concentration at the beginning of step j. From the current continuity in the absence of recombination, we have $n(0_j) = n(0)$, independent of j. Equation (5) shows that the diffusion flux is enhanced by a factor of N and so is the effective diffusion velocity, which becomes $2D/W_j$. If we increase the number of steps keeping W constant, the conventional cutoff frequency f_T will increase in proportion to N, cf. Fig. 2. Clearly, the recombination limited static current gain will also scale in proportion to N. This effect has in fact been observed by Ohishi et al. [14] in a InGaAsP/InP HBT with a double-layer ($N=2$) base.

For high enough N, we would see a peaked structure in the current gain $|h_{21}|$ and for $N > N_{2\pi}$ the peak value is greater than unity, cf. the curve in Fig. 2 for $N = 30$. However, this situation is difficult to realize in practice, because each step must be high enough ($\Delta_j \gtrsim \hbar\omega_{opt}$) to ensure the no-return condition, while the overall potential drop is limited by the bandgap difference.

Power gain

Conditions for the transistor oscillation activity are more relaxed. Consider the intrinsic case first. An extended-frequency peak in the unilateral power gain U appears when the real part of the common-emitter output impedance $r_{22} \equiv Re(z^e_{22})$ changes sign [5,9]. In an intrinsic transistor we have $r_{22} = R_\varphi + R_E$, where

$$R_\varphi = \frac{\cos(\varphi) - \cos(\varphi+\theta)}{\omega\,C_C\,\theta}\;|\alpha_B| \approx \frac{|\alpha_B|\;\sin(\varphi+\theta')}{\omega\,C_C} \ , \tag{6}$$

R_E is the emitter resistance, and $\theta \equiv \omega\tau_C = 2\theta'$ is the collector transit angle. The phase φ includes the (usually small) delay $\varphi_E = \omega R_E C_E$, due to emitter capacitance C_E. The approximate relation in the right-hand side of Eq. (6) corresponds to $\theta \lesssim 1$. In this case a power gain peak occurs at $\varphi \approx \pi$, provided $|\alpha_B| > \omega C_C R_E$.

The latter condition is relatively easy to accommodate. However, as pointed out by Tiwari [6], considerations of the phaseshift effects on unilateral gain are rather meaningless without including extrinsic resistances and capacitances. A careful analysis [9] shows that when these "parasitics" are included, the condition for a peak in U near $\varphi = \pi$ is

$$|\alpha_B(\pi)| > \omega\,\tau_X\,, \tag{7}$$

where τ_X is a parasitics-limited time constant. For example, in a model which includes extrinsic emitter, base, and collector resistances (R_{Ex}, R_{Bx}, and R_{Cx}, respectively, cf. the equivalent circuit [9] of an abrupt-junction HBT) but neglects an extrinsic collector capacitance C_{Cx}, the expression for U is

$$U = \frac{|\alpha_B\alpha_C|^2}{4\omega^2 C_C^2 (R_B + R_{Bx})}\;\frac{1}{R_\varphi + R_X}\,, \tag{8}$$

where $R_X \equiv R_E + R_{Ex} + R_{Cx} + R_{Cx}(R_E + R_{Ex})/(R_B + R_{Bx})$ and $\alpha_C \equiv (\sin\theta'/\theta')\,e^{-i\theta'}$ is the collector transport factor. The corresponding τ_X in Eq. (7) is $\tau_X = C_C R_X$. Besides the "low" frequency regime, where $U > 1$, the transistor will be active in the range of frequencies, where

$$|\alpha_B|\,\sin(\varphi + \theta') + \omega\tau_X \le 0\,. \tag{9}$$

In this range $U < 0$ and hence one can obtain $U \gg 1$ by adding a series resistance. Obviously, inequality (9) can only be obtained if $|\alpha_B| > \omega\tau_X$; from Eq. (3) this means that the number of steps in the base must exceed the value $N_\pi = (\pi^2/3)\,|\ln(\omega\tau_X)|^{-1}$. For a transistor not overdamped by the parasitics, say $\omega\tau_X \lesssim 0.5$, we need $N \gtrsim 5$.

The solid line in Fig. 3 displays the gain $|U|$, calculated for a model HBT with 5 steps in the base. The two peaks correspond to a vanishing denominator in (8); between the peaks U is negative.

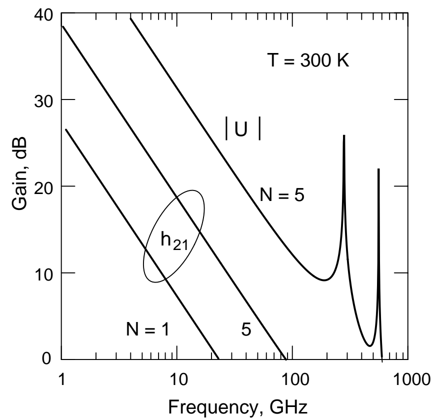

Figure 3. Common-emitter current gain $|h_{21}|$ and the unilateral gain $|U|$ of HBT's with the same base width $W = 0.25$ µm, the same diffusivity $D = 50\ \mathrm{cm^2/s}$, the same collector transit time $\tau_C = 1$ ps, and different number N of steps in the base.

Assumed parameters: $R_B = R_{Bx} = 50\ \Omega\cdot\mu\mathrm{m}$, $R_E = 10\ \Omega\cdot\mu\mathrm{m}$, $R_{Ex} = R_{Cx} = 20\ \Omega\cdot\mu\mathrm{m}$, $C_C = 0.25$ fF/µm, and $C_C = 10$ fF/µm.

It should be emphasized that we are not considering any "ballistic" boosts in the particle speed at the step edges. Such effects may in fact be beneficial; they are not expected to qualitatively modify the result (3). If we increase N at the expense of making the step height smaller, $\Delta_j < \hbar\omega_{opt}$, then the necessary energy relaxation will not occur at each step and hot electrons will diffuse backward as well as forward. In order to treat this regime quantitatively, a Boltzmann transport model for minority carriers has been developed [15]. It includes the effects of a finite optical phonon scattering time τ_{op} and finite exit velocity at each step. Preliminary results obtained in this model indicate that for $\Delta_j > \hbar\omega_{opt}$ the present simplified approach is valid, provided that τ_{op} is shorter than the elastic collision time of hot carriers exiting a step.

Comparison with Linear Grading

In the limit of $N \to \infty$ and Δ_j, $W_j \to 0$ the base structure becomes equivalent to that in a linear-graded HBT [16-18] with $\nabla E_G = \lim \Delta_j / W_j$. In such a device, the phase of α_B is acquired with the drift velocity $v \equiv -\mu \nabla E_G / e$, while the magnitude $|\alpha_B|$ is attenuated due to spreading by diffusion. An equation analogous to (3) can be derived by solving the continuity equation $\nabla \cdot J = e(\partial n / \partial t)$ for the minority current taken in the drift-diffusion form $J = n\mu \nabla E_G + eD\nabla n$. Using the phasor notation

$$n(x, t) \equiv \bar{n}(x) + \hat{n}(x) e^{i\omega t} ,$$

the continuity equation for the dynamic component $\hat{n}$ reduces to

$$i\omega \hat{n} = -v\hat{n}' + D\hat{n}'' . \tag{10}$$

Normalizing the coordinate x to the base width W, we find that the solution of Eq. (10), satisfying $\hat{n}(1) = 0$, is of the form

$$\hat{n} = A e^{rx} \sinh[\lambda(x-1)] , \tag{11}$$

where $\lambda \equiv \sqrt{r^2 + 2i\omega\tau_D}$ and r is the ratio of the characteristic diffusion time $\tau_D = W^2/2D$ to the drift time $\tau_B = W/v$, viz.

$$r \equiv \frac{\tau_D}{\tau_B} = \frac{Wv}{2D} \approx \frac{\Delta E_G}{2kT} , \tag{12}$$

where ΔE_G is the total bandgap variation. The last equation in the right-hand side of (12) results form Einstein's relation $eD = \mu kT$ (valid for not too high bandgap gradients).

From Eq. (11) we find the base transport factor $\alpha_B \equiv J(1)/J(0)$ in the form

$$\alpha_B(\omega) = \frac{\exp(r)}{\cosh(\lambda) + (1 + 2i\omega\tau_B/r)^{-1/2} \sinh(\lambda)} , \tag{13}$$

In the absence of a grading, $v \to 0$, Eq. (13) reduces to $\alpha_B = \cosh^{-1}[(2i\omega\tau_D)^{1/2}]$, which corresponds to Eq. (2a) extended to the entire base. For a large grading, $r \gg 1$, one has, asymptotically, $\lambda \approx r + i\omega\tau_B + (\omega\tau_B)^2/2r$ and Eq. (13) reduces to

$$\alpha_B(\varphi) = e^{-\varphi^2/2r} e^{-i\varphi} , \tag{14}$$

where $\varphi \equiv \omega\tau_B(1 - 1/2r)$. The minority carrier drift effects are qualitatively similar to those resulting from the enhanced forward diffusion. Parameter $2r$ in Eq. (14) plays the same role as $3N$ in Eq. (3). [In the ballistic case, Eq. (1), the parameter equivalent to r is $r_B \equiv (\tau_B/\delta\tau)^2$.] For our example of 5 steps (Fig. 3) we need $\Delta E_G = 5\Delta_j \gtrsim 5\hbar\omega_{opt} \approx 180\,\text{meV}$ (in GaAs/AlGaAs). Precisely the same effect will be achieved with a graded-gap base with $r = 7.5$, requiring $\Delta E_G = 15\,kT \approx 380\,\text{meV}$ at room temperature. As means for achieving the extended frequency operation at $\varphi \gtrsim \pi$, both approaches appear equally feasible.

The step-base approach offers additional flexibility in the design of ultrahigh speed heterostructure transistors. Perhaps its most obvious practical application is to relieve the stringent trade-off between the requirements of low base resistance and short base propagation time. As discussed above, this can be accomplished even with a small number of steps, e.g., $N = 2$. For higher number of steps, $N \gtrsim 5$, one can obtain an active behavior of the transistor at an extended frequency $f \approx \pi f_{max}$, provided (cf. Eq. 7) the device at this frequency is not overdamped by extrinsic elements.

Acknowledgement

This paper reviews the work carried out in collaboration with Anatoly Grinberg and Vera Gorfinkel, to whom I am deeply grateful for all the insights they provided.

References

1. G. T. Wright, "Transistor transit-time oscillator", *Electron. Lett.* **3**, pp. 234-235 (1967).
2. G. T. Wright, "Small-signal theory of the transistor transit-time oscillator (translator)", *Solid-St. Electron.* **22**, pp. 399-403 (1979).
3. Y. Druelle, Y. Crosnier, and G. Salmer "Large-signal behaviour and capability of the transistor transit-time oscillator", *IEE Proc.* **127**, Pt. I, pp. 72-77 (1980).
4. N. Dagli, W. Lee, S. Prasad, and C. G. Fonstad, "High-frequency characteristics of inverted-mode heterojunction bipolar transistors", *IEEE Electron Device Lett.* **EDL-8**, pp. 472-474 (1987).
5. N. Dagli, "Physical origin of the negative output resistance of heterojunction bipolar transistors", *IEEE Electron Device Lett.* **EDL-9**, pp. 113-115 (1988).
6. S. Tiwari, "Frequency dependence of the unilateral gain in bipolar transistors", *IEEE Electron Device Lett.* **EDL-10**, pp. 574-576 (1989).
7. N. Dagli, "A unipolar transistor with negative output resistance", *Solid-St. Electron.* **33**, pp. 831-836 (1990).
8. P. Roblin and S. C. Kang, "Unilateral power gain resonances and roll-off with frequency for the velocity-saturated MOSFET/MODFET wave equation", *IEEE Trans. Electron Devices* **ED-39**, pp. 1490-1495 (1992).
9. A. A. Grinberg and S. Luryi, "Coherent Transistor", *IEEE Transactions Electron Devices* **ED-40**, pp. 1512-1522 (1993).
10. A. A. Grinberg and S. Luryi, "Ballistic versus diffusive base transport in the high-frequency characteristics of bipolar transistors," *Appl. Phys. Lett.* **60** pp. 2770-2772 (1992).
11. H. Kroemer, "Heterostructure bipolar transistors and integrated circuits", *Proc. IEEE* **70**, pp. 13-25 (1982).
12. W. Shockley, "Negative resistance arising from transit time in semiconductor diodes", *Bell. Syst. Tech. J.* **33**, pp. 799-826 (1954).
13. S. Luryi, A. A. Grinberg, and V. B. Gorfinkel, "Heterostructure Bipolar Transistors with Enhanced Diffusion of Minority Carriers", *Appl. Phys. Letters* **63** (September 13, 1993).
14. T. Ohishi, Y. Abe, H. Sugimoto, K. Ohtsuka, and T. Matsui, "Ultra-high current gain InGaAsP/InP heterojunction bipolar transistor", *Electron. Lett.* **26**, pp. 392-393 (1990).
15. M. M. Dignam, A. A. Grinberg and S. Luryi, "Boltzmann equation studies of minority-carrier transport in the base of a heterojunction transistor", unpublished.
16. H. Kroemer, "Quasi-electric and quasi-magnetic fields in nonuniform semiconductors", *RCA Rev.* **18**, pp. 332-342 (1957).
17. D. L. Miller, P. M. Asbeck, R. J. Anderson, and F. H. Eisen, "(GaAl)As/GaAs heterojunction bipolar transistors with graded composition in the base", *Electron. Lett.* **19**, pp. 367-368 (1983).
18. J. R. Hayes, F. Capasso, A. C. Gossard, R. J. Malik, and W. Wiegmann, "Bipolar transistor with graded band-gap base", *Electron. Lett.* **19**, pp. 410-411 (1983).