Developing Bipolar Transistors for Sub-mm-Wave Amplifiers and Next-Generation (300 GHz) Digital Circuits

Mark Rodwell, Z. Griffith, N. Parthasarathy, E. Lind, C. Sheldon, S. R. Bank, U. Singisetti
ECE Department, University of California, Santa Barbara, 93106, USA.
phone: 805-893-3244, fax 805-893-3262, rodwell@ece.ucsb.edu

M. Urteaga, K. Shinohara, R. Pierson, P. Rowell
Rockwell Scientific Company, Thousand Oaks, CA 91360, USA

Despite formable progress with CMOS VLSI, bipolar transistors continue to find application in high frequency circuits due to the higher bandwidths obtained for a device of a given feature size and the higher breakdown voltages obtained for a device of a given bandwidth. While SiGe heterojunction bipolar transistors (HBTs) offer higher IC device counts and CMOS integration, InP HBTs offer higher bandwidth and higher breakdown voltage. During 2002-2005, under DARPA support, InP HBT transistor cutoff frequencies have been greatly increased -almost doubled- such that balanced devices have ~500 GHz cutoff frequencies, low-power digital ICs have been demonstrated at over 150 GHz clock rate, device junction areas have been reduced for reduced power consumption, and manufacturable high-yield IC fabrication processes have been introduced [1,2,3,4,5].

Here we consider the prospects for continued improvement in InP HBTs, specifically the challenges faced in a further doubling of transistor and IC bandwidth. Our objective is an IC technology supporting 300 GHz digital clock rates, ~600 GHz reactively-tuned amplifiers, and balanced ($f_{max} \geq f_\tau$) cutoff frequencies in the 700-1000 GHz range. Such ICs would permit monolithic transceivers for 300 GHz and 600 GHz imaging systems, ~250 GHz high-rate communications radios, chip sets for 300 Gb/s optical data transmission, and very high-resolution microwave mixed-signal ICs.

HBT bandwidth is increased by scaling[6]; the requirements are summarized in Table 1. A γ :1 increase in bandwidth is accomplished by a γ :1 reduction in the collector and a $\gamma^{1/2}$:1 reduction in the base layer thicknesses, a $\sim\gamma^2$:1 reduction in the emitter and collector junction widths, a $\sim\gamma^2$:1 increase in the current density, and a $\sim\gamma^2$:1 reduction in the emitter contact resistivity. Base contact resistivity scaling requirements depend somewhat upon the collector-base junction geometry.

Table 1 shows illustrative HBT designs for 150 GHz through 440 GHz digital circuits -- and 250 GHz through 750 GHz RF amplifiers -- with transistors designed for highest toggle frequency of an ECL master-slave D-flip-flop (MS-DFF), a representative digital building-block. The table shows relevant digital delay terms and coefficients. We emphasize concurrent reduction of all transistor parasitics; at a given scaling generation, high f_τ can be obtained at the expense of reduced f_{max} and reduced mm-wave and digital IC bandwidth by thinning epitaxial layers. The table shows the maximum operating frequency, conservatively estimated as $\sim f_{max}/2$, of a reactively-tuned amplifier having reasonable gain.

Base $\rho_{v,b}$ and emitter ρ_{ex} contact resistivities, key barriers to scaling, must decrease in proportion to the *square* of circuit bandwidth. A 125-nm scaling generation HBT suitable for ~300 GHz digital clock rate requires ~4-5 $\Omega-\mu\text{m}^2$ contact resistivities. Using Pd solid-phase-reaction contacts[7], we measure 5 $\Omega-\mu\text{m}^2$ base contact resistivity. Emitter contact resistivity is presently 5-10 $\Omega-\mu\text{m}^2$, and thus presents a more serious challenge. Resistivity is strongly influenced by surface oxides. Lower emitter resistance may be obtained through better cleaning, increased doping, flared contacts, or contacts covering both the edge and sidewalls of a tall semiconductor ridge. To avoid surface contaminants and interfacial defects, we are investigating *in-situ* MBE growth of a semiconductor-semimetal (ErAs) contact [8].

At 18 mA/ μm^2 , the 125 nm scaling generation HBT has a nearly degenerate electron concentration at the peak of the InP/InGaAs energy barrier in the base-emitter junction. Approximating to first order in I_c the resulting deviation from the $I_c \propto \exp[qV_{be}/kT]$ Boltzmann characteristics, the effect can be modeled as an increase[9] in R_{ex} , proportional to $1/m^*_{InP}$, of 1-2 $\Omega-\mu\text{m}^2$.

Because current density increases in proportion to the *square* of bandwidth, thermal resistance is a serious challenge. Further, transistor spacing D must scale as $1/\gamma$ to scale wiring delays. Approximating heat flow by cylindrical, spherical, and planar regions, the substrate temperature rise is $\Delta T_{InP} \cong (P/\pi K_{InP} L_E)\ln(L_e/W_e) + (P/\pi K_{InP})(1/L_E - 1/D) + (P/K_{InP})(T_{sub} - D/2)/D^2$ (Figure 3), where P is the HBT dissipation, L_E the emitter length, and K_{InP} the thermal conductivity. Scaling causes a logarithmic temperature increase arising from narrow emitters and a quadratic temperature increase arising from the decreased transistor spacings. By extreme thinning of the substrate in inverse proportion to the substrate thickness, low junction temperature can be maintained at a target 300 GHz f_{clock}. This might be realized with a thermal via process with Au backfilling. Package heat flow is approximated by cylindrical and planar regions (Figure 4), giving $\Delta T_{package} \cong ((2+\pi)/2\pi)(P_{chip}/K_{Cu}W_{chip}) \propto \gamma$. Heat rise increases as γ^1 because the HBT spacing $D \propto 1/\gamma$, hence die size $W_{chip} \propto 1/\gamma$. At 300 GHz clock rate, package heat rise in a 1000-HBT IC is acceptable given 3 mA switched per transistor (100 Ω loading).

As the collector is thinned, breakdown voltage is reduced. Figure 7 shows I_{ceo} vs. V_{ce} for InP/ InGaAs/ InGaAlAs/ InP DHBTs . If their thicknesses are an excessive fraction of the total collector depletion layer thickness, the narrow-bandgap InGaAs setback and InGaAs/InAlAs grade layers (Figure 2) may reduce V_{br} through either increased Zener tunneling or increased impact ionization. Comparison with InGaAsSb-base DHBTs[10], which do not require the setback layer or grade, has been confused by inconsistent definition of the current density at which $V_{br,ceo}$ is measured. Present data at 75 nm collector thickness[11] suggests that 10 nm setback and 24 nm grade layers do not measurably decrease $V_{br,ceo}$. When necessary, these layers can be thinned (Figure 8). Further, $V_{br,ceo}$ is often not the dominant limit to applied voltage; many published InP HBTs will fail thermally if biased at $J_{\max}/2$ simultaneously with $V_{br,ceo}/2$, or will exhibit thermal instability $(K_{thermal} = [-\partial V_{be}/\partial T][kT/qI + R_{ex}]^{-1} V_{ce}\theta_{ja} < 1)$ in the current distribution of a multi-finger power transistor under such a bias condition. To address both these limitations, normalized thermal resistance $\theta_{ja}A_e$ must be further reduced.

Figure 9 and Figure 10 show representative InP mesa DHBT results at 75 nm collector thickness and 600 nm emitter width. Other DHBTs[1] exhibit 450 GHz f_τ and 490 GHz $f_{\max}$. We are presently developing scaled 250 nm DHBTs with target performance as in Table 1.

To fabricate large ICs, high transistor yield is necessary. For acceptable IC power dissipation, HBT junction areas must be small. In the absence of these two key parameters, a wideband transistor technology has no relevance to ICs of significant integration scales. Because the mesa structure with self-aligned emitter-base contacts defined by liftoff does not permit the emitter semiconductor to be thinned as the junction width is reduced, its future viability in scaled processes is doubtful. Dielectric sidewall emitter-base processes (Figure 11) are showing high yield and high bandwidth[2,3]. Similarly, although laboratory processes reduce pad capacitance through cantilevered base contacts and reduce extrinsic C_{cb} through etch undercuts, it remains to be established whether such processes can provide high IC yield in highly scaled processes. Addressing parasitic reduction within a planar process, collector pedestal processes[12] (Figure 12) eliminate the base pad capacitance and reduce the extrinsic C_{cb} (Figure 13), increasing circuit speed[13] particularly for low-power logic (Figure 14) .

[1] Z. Griffith *et al*., *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 2061-2069, 2005
[2] M. Urteaga *et al*., Proc. Device Research Conf., Notre Dame, IN, June 21-23 2004, pp. 239-240
[3] G. He *et al*., IEEE Electron Device Letters, vol. 25, no. 8, pp. 520-522, 2004.
[4] T. Hussain *et al*., *Proc. IEEE International Electron Devices Meeting*, San Francisco, CA, 13-15 December, 2004
[5] D. Sawdai *et al*., IEEE Int. Conf. on Indium Phosphide and Related Materials, Glasgow, Scotland, 8-13 May, 2005
[6] M. Rodwell *et al*., IEEE Trans. Electron Devices, vol. 48, no. 11, pp. 2606-2624, 2001.
[7] E.F. Chor *et al,* Journal of Applied Physics, vol.87, (no.5), AIP, 1 March 2000. p.2437-44.
[8] H. Kazemi, *et al* SPIE Defense and Security Symposium, 28 March-1 April 2005. Orlando Fl.
[9] We are indebted to G. Liang and M. Lundstrom (Purdue) for their unpublished analyses.
[10] W. Snodgrass *et al*, IEEE Electron Device Letters, Jan. 2006.
[11] Z. Griffith *et al,* IEEE Int. Conf. on Indium Phosphide and Related Materials, Princeton, May 2006.
[12] N. Parthasarathy *et al*, IEEE Electron Device Letters, to be published.
[13] Z. Griffith *et al*, IEEE MTT-S International Microwave Symposium, San Francisco, June 11-16, 2006

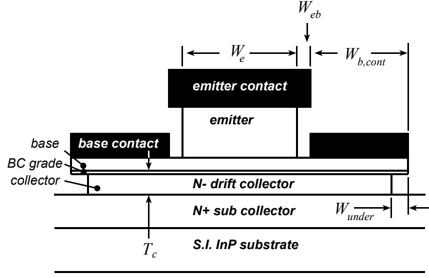

Figure 1: HBT critical dimensions.

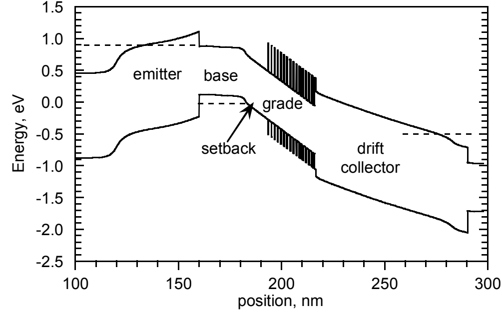

Figure 2: HBT band diagram & layer structure.

Parameter	scaling law	Gen. 2 (500 nm)	Gen. 3 (250 nm)	Gen. 4 (125 nm)	Gen 5 (62.5nm)
MS-DFF speed	γ^1	150 GHz	235 GHz	330 GHz	440 GHz
Amplifier center frequency	γ^1	245 GHz	400 GHz	650 GHz	750 GHz
Emitter Width	$1/\gamma^2$	500 nm	250 nm	125 nm	62.5 nm
Resistivity	$1/\gamma^2$	16 Ω-μm²	9 Ω-μm²	4 Ω-μm²	2 Ω-μm²
Base Thickness	$1/\gamma^{1/2}$	300Å	250 Å	212 Å	180 Å
Contact width	$\sim 1/\gamma^2$	300 nm	175 nm	120 nm	70 nm
Doping	γ^0	$7\ 10^{19}$/cm²	$7\ 10^{19}$/cm²	$7\ 10^{19}$/cm²	$7\ 10^{19}$/cm²
Sheet resistance	$\gamma^{1/2}$	500 Ω	600 Ω	707 Ω	830 Ω
Contact ρ	$1/\gamma^{1/2}$	20 Ω-μm²	10 Ω-μm²	5 Ω-μm²	5 Ω-μm²
Collector Width	$1/\gamma^2$	1.2 μm	0.60 μm	0.37 μm	0.20 μm
Thickness	$1/\gamma$	1500 Å	1060 Å	750 Å	530 Å
Current Density	γ^2	4.5 mA/μm²	9 mA/μm²	18 mA/μm²	36 mA/μm²
$A_{collector}/A_{emitter}$	γ^0	2.4	2.4	2.9	2.8
f_τ	γ^1	370 GHz	530 GHz	730 GHz	1.0 THz
f_{max}	γ^1	490 GHz	801 GHz	1.30 THz	1.5 THz
I_E / L_E	γ^0	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm
τ_f	$1/\gamma$	340 fs	240 fs	180 fs	130 fs
C_{cb}/I_c	$1/\gamma$	400 fs/V	280 fs/V	250 fs/V	190 fs/V
$C_{cb}\Delta V_{logic}/I_c$	$1/\gamma$	120 fs	85 fs	74 fs	57 fs
$R_{bb}/(\Delta V_{logic}/I_c)$	γ^0	0.76	0.54	0.34	0.39
$C_{je}(\Delta V_{logic}/I_C)$	$1/\gamma^{3/2}$	380 fs	180 fs	94 fs	50 fs
$R_{ex}/(\Delta V_{logic}/I_c)$	γ^0	0.24	0.24	0.24	0.24

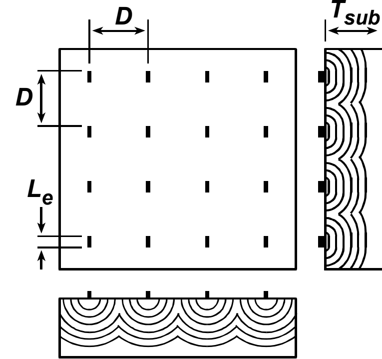

Figure 3: Thermal scaling analysis: heat flow in the IC substrate.

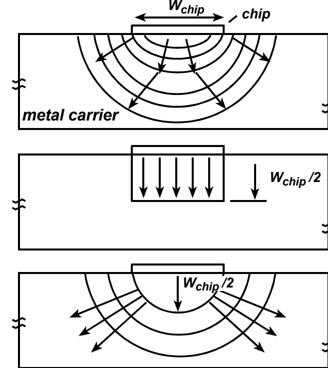

Figure 4: Thermal scaling analysis: heat flow in the IC package.

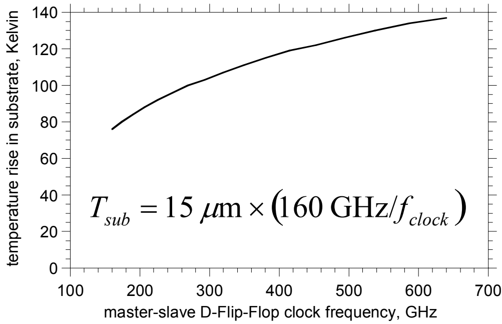

Figure 5: Approximate substrate temperatures rise as a function of digital clock rate.

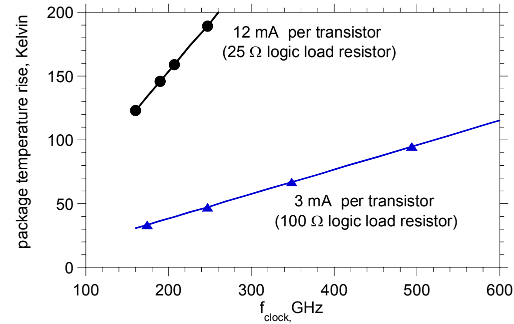

Figure 6: Approximate package temperature rise as a function of digital clock rate.

Plots assume proportional scaling from our 150 GHz ICs, e.g., $L_e = 5\ \mu m$, V_{ce}=2 V, $W_e = 500\ \text{nm}\,(160\ \text{GHz}/f_{clock})^2$, $D = 20\ \mu\text{m}\,(160\ \text{GHz}/f_{clock})$, and $T_{sub} = 15\ \mu\text{m}\,(160\ \text{GHz}/f_{clock})$.

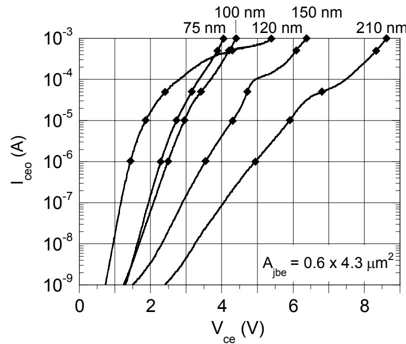

Figure 7: Breakdown characteristics: I_{ceo} vs. V_{ce}.

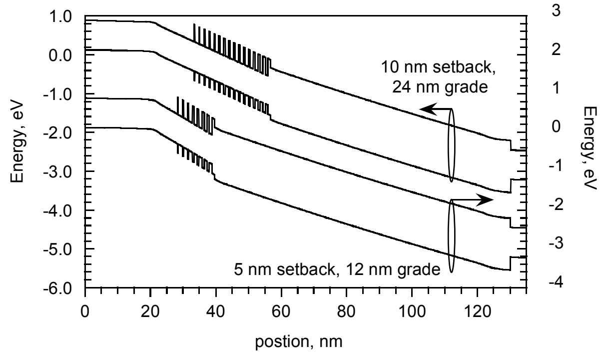

Figure 8: Thinned grade for improved $V_{br,ceo}$.

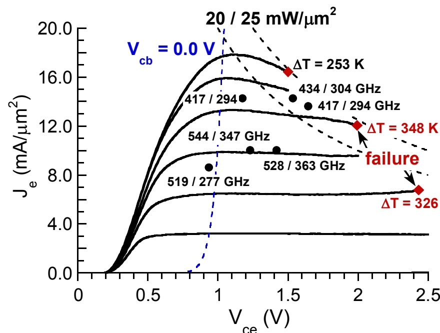

Figure 9: DC parameters for DHBT with 75 nm thick collector.

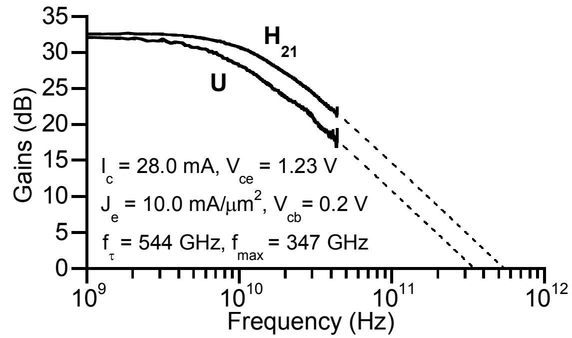

Figure 10: mm-wave gains for DHBT with 75 nm thick collector.

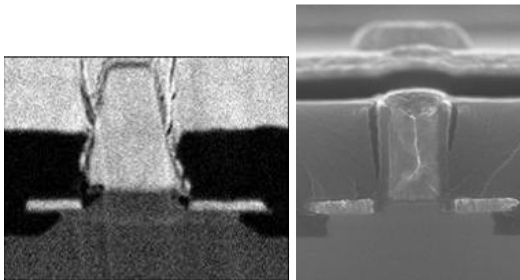

Figure 11: SEMs comparing UCSB mesa and RSC sidewall HBTs[2].

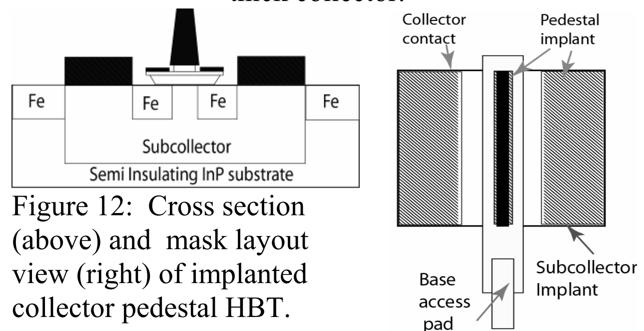

Figure 12: Cross section (above) and mask layout view (right) of implanted collector pedestal HBT.

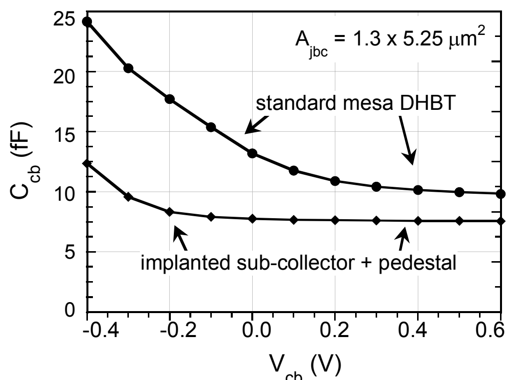

Figure 13: Observed collector capacitance reduction in collector pedestal HBTs.

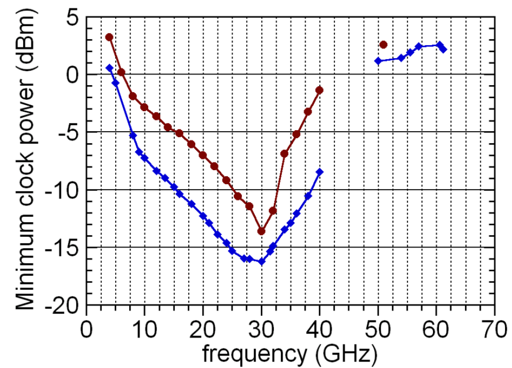

Figure 14: Sensitivity vs. Frequency of 13.6 mW CML static dividers with and without collector pedestal.