On the Feasibility of low-THz InP HBTs

Mark Rodwell, Z. Griffith, E. Lind, U. Singisetti, M. Wistey, A.C. Gossard

***Abstract*—We present an InP HBT scaling roadmap through 32 nm. Target device bandwidths increase as the inverse square root of lithographic feature size. The required improvements of emitter and base contact resistivities, and the increase in IC thermal resistance, are the major limits of scaling. The present state-of-art in P- and N- contact resistivities are similar to those required of the 64 nm generation, a device for which ~1 THz f_τ and 2 THz f_{max} would be expected, and for which 1 THz tuned amplifiers would be feasible. Present UCSB devices at 250 nm emitter size obtain record 560 GHz *simultaneous* f_τ and f_{max}, while record 755 GHz f_{max} has been attained for a device with a thicker collector.**

BANDWITHS of InP bipolar transistors are rapidly increasing, and THz bandwidths should soon be realized. 250 nm HBTs have demonstrated 416 GHz f_τ and 755 GHz $f_{\max}$ [1] and 560 GHz simultaneous f_τ and $f_{\max}$ [2]. We here address both recent progress, and prospects for further significant improvements .

Table 1 summarizes scaling laws [3] and provides a roadmap through 32 nm. To improve by 2:1 the bandwidth of *any* circuit employing the HBT, we must reduce by 2:1 all transit delays and capacitances while maintaining constant all resistances and all bias and signal voltages and currents. Thinning the collector depletion layer by 2:1 reduces the collector transit time 2:1 but doubles junction capacitances per unit area; the collector and emitter junction areas are therefore reduced 4:1 so as to obtain the required 2:1 capacitance reduction. The emitter junction area has decreased 4:1; yet the emitter current and the emitter access resistance must both remain constant. The emitter current density and specific access resistivity must thus both be reduced 4:1. The 4:1 required increase in current density is consistent with the Kirk-effect limit as this varies as the inverse square of depletion layer thickness. Base resistance is *dominated* by the Ohmic contact, and is approximately the contact resistivity divided by the contact area. With a 4:1 reduction in junction areas, the base contact resistivity must thus be reduced 4:1.

HBT lithographic scaling laws are driven by thermal constraints. The junction temperature rise of an isolated HBT on a thick substrate is $\Delta T \approx \left(P/\pi K_{InP} L_E\right) + \left(P/\pi K_{InP} L_E\right)\ln\left(L_E/W_E\right)$. Here K_{InP} is the substrate thermal conductivity and P the dissipated power. We must reduce the emitter junction area $L_E W_E$ by 4:1; maintaining constant L_E while reducing W_E by 4:1 results in only a logarithmic increase in junction temperature with scaling.

Table 1 shows both HBT cutoff frequencies and key time constants (e.g. $C_{cb}\Delta V_{\text{logic}}/I_c$) associated with digital gate delay [3]. HBTs for mm-wave amplification are designed for highest $f_{\max}$ and for $f_\tau > f_{\max}/2$, while HBTs for mixed-signal ICs are designed for minimum ECL gate delay [3]. Transistors having $f_{\max} < f_\tau$ are of limited utility.

Table 1: Bipolar transistor scaling laws and InP HBT scaling roadmap.

Parameter	scaling law	Gen. 3 (256 nm)	Gen. 4 (128 nm)	Gen 5 (64 nm)	Gen 5 (32 nm)
MS-DFF speed	γ^1	240 GHz	330 GHz	480 GHz	660 GHz
Amplifier center frequency	γ^1	430 GHz	660 GHz	1.0 THz	1.4 THz
Emitter Width	$1/\gamma^2$	256 nm	128 nm	64 nm	32 nm
Resistivity	$1/\gamma^2$	8 Ω-μm^2	4 Ω-μm^2	2 Ω-μm^2	1 Ω-μm^2
Base Thickness	$1/\gamma^{1/2}$	250 Å	212 Å	180 Å	180 Å
Contact width	$1/\gamma^2$	175 nm	120 nm	60 nm	30 nm
Doping	γ^0	7 10^{19} /cm^2	7 10^{19} /cm^2	7 10^{19} /cm^2	7 10^{19} /cm^2
Sheet resistance	$\gamma^{1/2}$	600 Ω	708 Ω	830 Ω	990 Ω
Contact ρ	$1/\gamma^2$	10 Ω-μm^2	5 Ω-μm^2	2.5 Ω-μm^2	1.25 Ω-μm^2
Collector Width	$1/\gamma^2$	600 nm	360 nm	180 nm	90 nm
Thickness	$1/\gamma$	106 nm	75 nm	53 nm	37.5 nm
Current Density	γ^2	9 mA/μm^2	18 mA/μm^2	36 mA/μm^2	72 mA/μm^2
$A_{collector}/A_{emitter}$	γ^0	2.4	2.9	2.8	2.8
f_τ	γ^1	520 GHz	730 GHz	1.0 THz	1.4 THz
$f_{\max}$	γ^1	850 GHz	1.30 THz	2.0 THz	2.8 THz
$V_{BR,CEO}$		4.0 V	3.3 V	2.75 V	?
I_E/L_E	γ^0	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm
τ_f	$1/\gamma$	240 fs	180 fs	130 fs	95 fs
C_{cb}/I_c	$1/\gamma$	280 fs/V	240 fs/V	170 fs/V	120 fs/V
$C_{cb}\Delta V_{\text{logic}}/I_c$	$1/\gamma$	85 fs	74 fs	52 fs	36 fs
$R_{bb}/(\Delta V_{\text{logic}}/I_c)$	γ^0	0.47	0.34	0.26	0.23
$C_{je}(\Delta V_{\text{logic}}/I_C)$	$1/\gamma^{3/2}$	180 fs	94 fs	50 fs	33 fs
$R_{ex}/(\Delta V_{\text{logic}}/I_c)$	γ^0	0.24	0.24	0.24	0.24

Thermal resistance is a severe scaling constraint. Transistor spacings D must scale in inverse proportion to circuit bandwidth to scale wiring delays. The substrate temperature rise is approximately [4]

$$\Delta T_{InP} \sim \Delta T_{\text{cylindrical}} + \Delta T_{\text{hemisperical}} + \Delta T_{\text{planar}}$$
$$= \left(\frac{P}{\pi K_{InP} L_E}\right)\ln\left(\frac{L_e}{W_e}\right) + \left(\frac{P}{\pi K_{InP}}\right)\left(\frac{1}{L_E} - \frac{1}{D}\right)$$

M. Rodwell, Z. Griffith, A.M. Crook, U. Singisetti, M. Wistey, G. Burek, and A.C. Gossard are with the ECE and Materials Departments, University of California, Santa Barbara, CA 93106. E. Lind, is now with Lund University, Lund, Sweden. Work supported by the DARPA TFAST and SWIFT programs, and by the ONR ultra-low resistance epitaxial contacts program.

$$+\left(\frac{P}{K_{InP}}\right)\frac{T_{sub}-D/2}{D^2} \qquad (1)$$

In addition to the logarithmic temperature increase arising from narrow emitters, at fixed substrate thickness T_{sub}, scaling causes ΔT_{planar} to scale in proportion to the square of circuit bandwidth. The third (planar) term can be reduced by extreme substrate thinning using wafer lapping or thermal vias [5].

Given a square IC of linear dimensions W_{chip} on a large copper heat sink, package thermal resistance is $\Delta T_{package} \cong (1/4+1/\pi)(P_{chip}/K_{Cu}W_{chip})$. Because D varies as the inverse of IC bandwidth, W_{chip} varies by the same law, and $\Delta T_{package}$ increases in proportion to bandwidth.

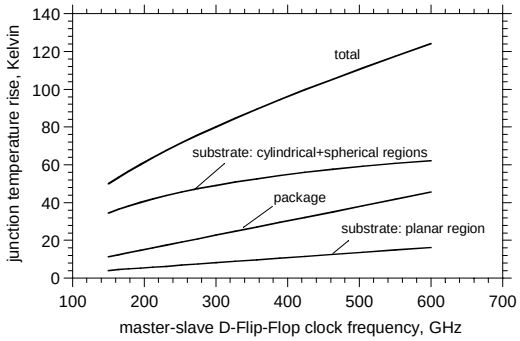


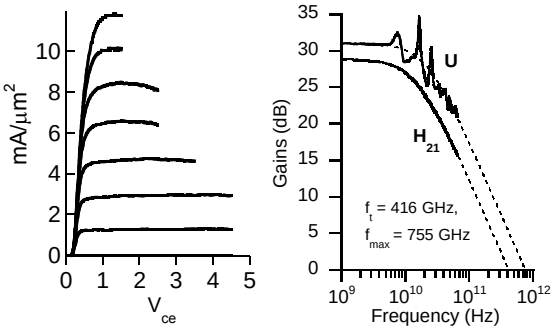
Figure 1: Calculated package and substrate temperatures rise, as a function of digital clock rate, for a 2048-HBT CML integrated circuit**.**

Using these relationships, and scaling from an existing 150 GHz IC design the dissipation, wire lengths, transistor parameters, and circuit bandwidths, we project (Figure 1) the junction temperature rise of a 2048-HBT CML digital IC vs. clock rate. The substrate is thinned aggressively with the assumed use of thermal vias: $T_{sub} = 40\,\mu\text{m}\cdot(150\,\text{GHz}/f_{clock})$ to obtain these results. It appears that 450 GHz f_{clock} is thermally feasible in a 2000-transistor IC.

Contact resistivities must decrease in proportion to the inverse *square* of circuit bandwidth. Resistivity of *ex-situ* deposited contacts is strongly influenced by surface oxides and by cleaning procedures. For the base contacts, we observe the lowest resistivity with Pd solid-phase-reaction contacts [6], which penetrate oxides. We have fabricated emitter contacts [7] by depositing (refractory) Molybdenum contact metal *in-situ* by MBE, avoiding surface oxidation. We observe 0.5 $\Omega-\mu\text{m}^2$ contact resistivity. Such contacts are readily integrated into the HBT process flow. With *ex-situ* refractory TiW contacts, low 0.7 $\Omega-\mu\text{m}^2$ contact resistivity is obtained if, immediately prior to TiW contact metal deposition, the wafer surface is first exposed to UV-generated ozone and subsequently rinsed in concentrated ammonia.

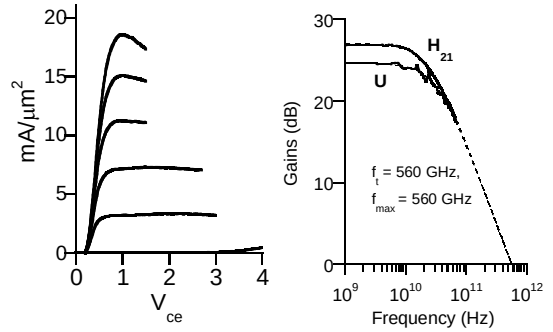
The target emitter access resistivity is the sum of contact resistance, bulk semiconductor resistance, and transport effects. Electron degeneracy in the emitter-base depletion layer contributes an effective increase in the aggregate emitter resistance, proportional to $1/m^*_{InP}$, of ~1.0 $\Omega-\mu\text{m}^2$.

Including the effects of degeneracy, present contact resistivities to meet the 64 nm generation requirements. HBTs with simultaneous 1.0 THz f_τ and 2 THz $f_{\max}$ appear feasible; these would enable 450 GHz digital clock rates and 750 GHz amplifiers. Feasibility of the 32 nm device is unclear, as substantial progress in contact resistivity is required. Figure 2, Figure 3, and Figure 4 show recent results [1,2,8].



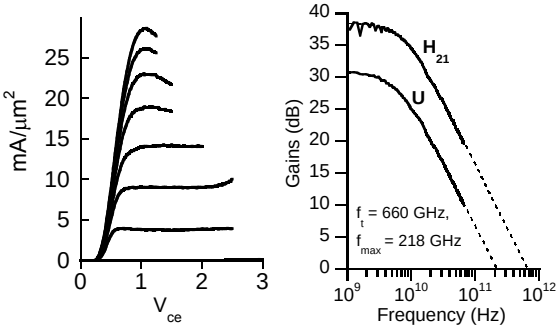
(a) (b)

Figure 2: Measured common-emitter characteristics (a) and RF gains (b) of a DHBT having T_c=150 nm, T_b=30 nm, and W_e=250 nm, biased at J_e=12 mA/μm^2. The DHBT exhibits $V_{br,ceo}$ =5.6V at I_c/A_e=10 kA/cm^2.



(a) (b)

Figure 3: Measured common-emitter characteristics (a) and RF gains (b) of a DHBT having T_c=70 nm, T_b=22 nm, and W_e=250 nm biased at J_e=13 mA/μm^2. The DHBT exhibits $V_{br,ceo}$ =3.3V at I_c/A_e=15 kA/cm^2.



(a) (b)

Figure 4: Measured common-emitter characteristics (a) and RF gains (b) of a DHBT having T_c=60 nm, T_b=14 nm, and W_e=400 nm biased at J_e=13 mA/μm^2. The DHBT exhibits $V_{br,ceo}$ =3.0V at I_c/A_e=10 kA/cm^2.

[1] Z. Griffith *et al* ,IEEE/LEOS Int. Conf. Indium Phosphide and Related Materials, Matsue, Japan, May 14-18, 2007

[2] E. Lind *et al*, post deadline paper, 2007 IEEE Device Research Conference, June 18-20, Univ. of Notre Dame, South Bend, Illinois.

[3] M. Rodwell *et al*, IEEE TED , vol. 48, no. 11, pp. 2606–2624, Nov. 2001.

[4] M. Rodwell *et al* , IEEE/LEOS Int. Conf. Indium Phosphide and Related Materials, Matsue, Japan, May 14-18, 2007

[5] J. S. Kofol *et al*, 1992 GaAs IC Symposium, 1992, 4-7 Oct, pp. 267-270

[6] E. F. Chor *et al* , Journal of Applied Physics, vol.87, (no.5), AIP, 1 March 2000. pp .2437-44.

[7] U. Singisetti, *et al*, 2007 IEEE Device Research Conference, June 18-20, Univ. of Notre Dame, South Bend, Illinois

[8] Z. Griffith et al, 2006 IEEE CSIC Symposium, Nov. 12-15, San Antonio