Frequency Limits of Bipolar Integrated Circuits

Mark Rodwell, Z. Griffith, N. Parthasarathy, U. Singisetti, V. Paidi,
M. Urteaga* , R. Pierson*, P. Rowell*, B. Brar*
ECE Department, University of California, Santa Barbara, 93106, USA,
*Rockwell Scientific Company, Thousand Oaks, CA 91360, USA

***Abstract* — InP HBTs at the 500 nm scaling generation have attained 450 GHz balanced cutoff frequencies and ~4 V breakdown. With such devices, 150 GHz digital circuits (static dividers) have been demonstrated. 175 GHz amplifiers have been demonstrated with slower HBTs. We discuss transistor scaling laws and scaling limits for realizing digital and analog/RF circuits at sub-mm-wave frequencies; the most critical limitations are metal/semiconductor contact resistivities and dissipated power densities. Given present contact performance and thermal design, 230 GHz digital ICs and 300 GHz power amplifiers are feasible today.**

***Index Terms* — Heterojunction bipolar transistors, Millimeter wave integrated circuits, Bipolar digital integrated circuits**

I. Introduction

Radios, imaging systems, and radars today employ transistor circuits only at low mm-wave frequencies, with higher frequencies served by Schottky diodes, vacuum electron devices, and gas lasers. Traveling-wave vacuum devices offer much higher output power than transistors, lasers offer a much wider frequency range, and both lasers and bolometers offer lower noise. Despite these disadvantages, transistors offer key attributes which motivate us to extend their operation to sub-mm-wave frequencies. Transistors are very small, are efficient power amplifiers, and can be very reliable. Unlike diodes, THz tubes and lasers, transistors serve many functions, including amplifiers, oscillators, mixers, harmonic multipliers, switches, and logic gates. With transistors, one can compactly construct a wealth of important communications circuits, including PLLs and frequency synthesizers, monolithic frequency multiplier chains, QAM (de)modulators, superheterodyne receivers, and monolithic phased arrays. Thus, with transistors of sufficient bandwidth, we will be able to construct compact and low-power *integrated circuits* which perform sophisticated functions at sub-mm-wave frequencies.

Today (Figure 1) InP transistors, both HEMTs and HBTs, obtain balanced ($f_\tau \approx f_{\max}$) cutoff frequencies slightly below 500 GHz. HEMTs and HBTs are complementary: HEMTs exhibit much lower noise for receivers while HBTs provide faster digital circuits and have higher breakdown hence serve better in power amplifiers and in ADCs and DACs. Breakdown is such that a 450-GHz- f_τ HBT can produce a 4-V signal (Figure 1).

Further increases in transistor bandwidth will enable amplifiers and digital logic circuits operating at sub-mm-wave frequencies. How can this bandwidth be further improved ? In the early development of III-V transistors, motivated by the superior transport properties compared to those of Silicon, much early research focused on identifying materials with the highest electron and hole mobilities, and determining the role of transient transport effects in providing very high peak electron velocities [1]. With these effects now well-understood, further improvement in transistor bandwidth are now obtained by *scaling* [2,3], e.g. reducing device epitaxial and lithographic dimension. Scaling, of course, faces challenges and limits, key among these being nonzero metal-semiconductor contact resistivities, high dissipated junction power densities, decreasing breakdown voltages, and decreasing yield in the fabrication of progressively smaller device features.

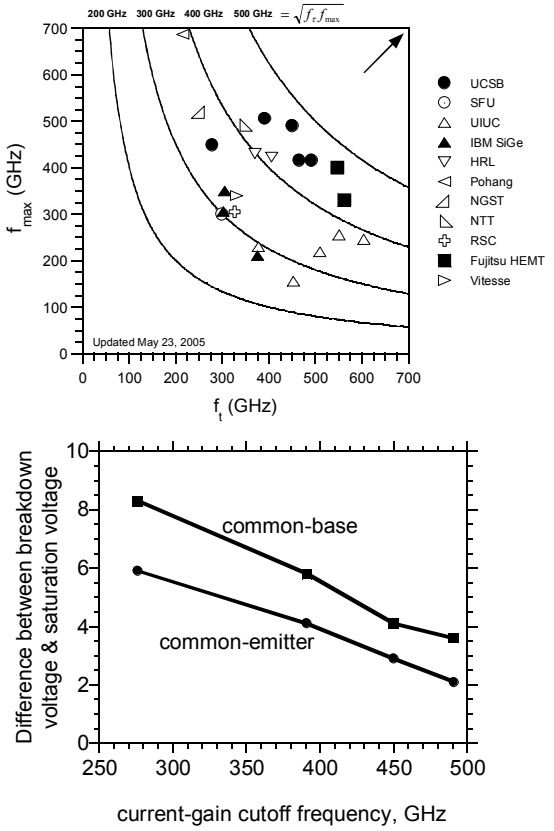

Figure 1: InP DHBTs. (a) Recent results (b) DHBT maximum voltage swing as a function of device f_τ. Higher f_τ will be obtained at a given breakdown (hence collector thickness) as emitter and collector contacts are improved and the base layer thinned.

II. Semiconductor Device Scaling

Scaling laws are central to high-frequency semiconductor device design. Consider the design of a Schottky or PIN diode (Figure 2). The Schottky diode has capacitance and transit time arising from its depletion layer, bulk and contact resistance from its bottom N+ contact, and small-signal junction resistance $R_j = kT/qI$. A PIN diode also has a top Ohmic contact resistance $R_{s,top}$.

To double the diode bandwidth, we must reduce all transit times and capacitances 2:1 while maintaining constant all resistances and bias and signal currents. To reduce the space-charge transit time 2:1, the depletion layer thickness D must be reduced 2:1. With depletion capacitance $C \propto WL/D$, we have inadvertently doubled C; we had desired a 2:1 reduction. Clearly, WL must be reduced 4:1. If W is small, the resistance associated with the N+ buried layer and its contacts varies approximately as $R_{s,bottom} \propto 1/L$, and is nearly independent of W. Thus, to maintain constant $R_{s,bottom}$ while reducing C by 2:1, L is held constant while W is reduced 4:1.

$R_j = kT/qI$ must be held constant, hence I is unchanged. This requires a 4:1 increase in current density $J = I/LW$. This 4:1 increase in current density is feasible within the limits imposed by the Kirk effect (space charge screening) because $J_{Kirk} \propto 1/D^2$. Note that the power density has increased 4:1

For a PIN diode, there is also a top contact resistance $R_{s,top} = \rho_{s,top}/LW$. To maintain constant $R_{s,top}$ given the 4:1 junction area reduction, the contact resistivity $\rho_{s,top}$ must be reduced 4:1. Schottky diodes do not face this limitation.

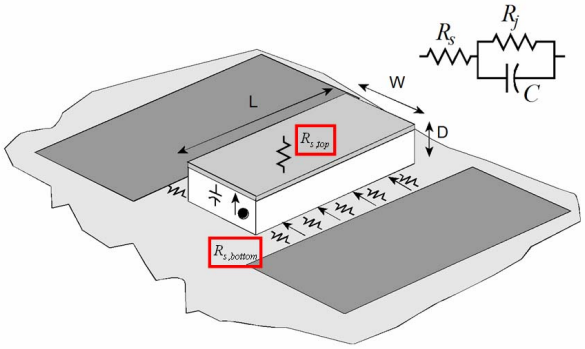

Figure 2: Schottky diode as an example of semiconductor device scaling.

To double diode bandwidth, the epitaxial layer thicknesses were scaled 2:1, the lithographic dimensions scaled 4:1, current and power densities increased 4:1, and resistivities of contacts normal to the current flow were reduced 4:1. This concurrently reduced the devices' resistance-capacitance (RC) and transit ($\tau_{transit}$) time constants. Such $R/C/\tau_{transit}$ bandwidth limits are relevant to almost all high-frequency semiconductor devices, including bipolar and field-effect transistors, Schottky and PIN diodes, photodiodes, photo mixers, and resonant-tunnel-diodes [4], and scaling is the primary tool to improve the bandwidth of such devices.

The interaction of RC and transit time effects arises whenever AC currents generated in a depletion region must subsequently pass though Ohmic contacts, and effective THz devices must minimize or circumvent contact resistance, depletion capacitance, and depletion layer transit times. Such effects must be critically examined in any candidate sub-mm-wave semiconductor device. Semiconductor lasers, operating at mid-IR to optical (~5-600 THz) frequencies, avoid RC losses because the laser dielectric waveguide confines the AC signal field away from resistive N+ and P+ bulk and contact regions, and the AC field is removed from the laser end facets rather than via displacement currents accessed through Ohmic contacts to doped semiconductor layers. In semiconductors, such mode confinement is hard to realize below ~ 1 THz.

Table 1: Device scaling laws, and illustrative HBT designs for 160 GHz through 330 GHz digital clock rates

Parameter	scaling law	Gen. 2	Gen. 3	Gen. 4
MS-DFF speed	γ^1	158 GHz	230 GHz	330 GHz
Emitter Width	$1/\gamma^2$	500 nm	250 nm	125 nm
Resistivity	$1/\gamma^2$	15 Ω-μm^2	7.5 Ω-μm^2	5 Ω-μm^2
Base Thickness	$1/\gamma^{1/2}$	300Å	250 Å	212 Å
Doping	γ^0	7 10^{19}/cm^2	7 10^{19}/cm^2	7 10^{19}/cm^2
Sheet resistance	$\gamma^{1/2}$	500 Ω	600 Ω	707 Ω
Contact ρ	$1/\gamma^{1/2}$	20 Ω-μm^2	10 Ω-μm^2	5 Ω-μm^2
Collector Width	$1/\gamma^2$	1.1 μm	0.54 μm	0.27 μm
Thickness	$1/\gamma$	1500 Å	1060 Å	750 Å
Current Density	γ^2	5 mA/μm^2	10 mA/μm^2	20 mA/μm^2
$A_{collector}/A_{emitter}$	γ^0	2.8	2.8	2.8
f_τ	γ^1	371 GHz	517 GHz	720 GHz
f_{max}	γ^1	483 GHz	724 GHz	1.06 THz
I_E/L_E	γ^0	2.4 mA/μm	2.4 mA/μm	2.4 mA/μm
τ_f	$1/\gamma$	340 fs	250 fs	170 fs
C_{cb}/I_c	$1/\gamma$	440 fs/V	310 fs/V	220 fs/V
$C_{cb}\Delta V_{logic}/I_c$	$1/\gamma$	130 fs	94 fs	66 fs
$R_{bb}/(\Delta V_{logic}/I_c)$	γ^0	0.66	0.51	0.41
$C_{je}(\Delta V_{logic}/I_C)$	$1/\gamma^{3/2}$	350 fs	250 fs	180 fs
$R_{ex}/(\Delta V_{logic}/I_c)$	γ^0	0.24	0.24	0.24

III. Bipolar Transistor Scaling Laws

Following the principles outlined above, HBT bandwidth is also increased by scaling; the requirements are summarized in Table 1. A γ:1 increase in HBT 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 upon the collector-base junction geometry. Table 1

shows illustrative HBT designs for 160 GHz through 330 GHz digital clock rates. In this table, the primary circuit speed metric is the maximum toggle frequency of an ECL master-slave D-flip-flop (MS-DFF). Such ICs find broad application in small-scale high-frequency mixed-signal and communications ICs, and their maximum clock frequency is a generally-reported performance metric for a mixed-signal IC technology. Time constants at the bottom of Table 1 reflect delay terms relevant to the MS-DFF maximum toggle frequency . Note the emphasis in Table 1 on 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 the epitaxial layers without concurrent lateral scaling. Such devices have little relevance to ICs.

Base $\rho_{v,b}$ and emitter ρ_{ex} contact resistivity, thermal resistance θ_{ja} [5], and fabrication yield of submicron features are the key barriers to further scaling. Current density must increase, and emitter and base contact resistivities must decrease in proportion to the *square* of circuit bandwidth. Thermal resistance normalized to the emitter junction area $\theta_{ja} A_E$, must also be reduced in proportion to the square of circuit bandwidth.

A 125-nm scaling generation (Table 1) HBT suitable for ~300 GHz digital clock rate and ~600 GHz power amplifiers ($f_{signal} / f_{\max}$ ~2/3) requires ~5 $\Omega - \mu\text{m}^2$ contact resistivities and must operate at ~40 mW/μm^2 power density, requiring < 3-4 K-μm^2/mW thermal resistance. Advanced fabrication processes (pedestal implants, emitter regrowth) may ease these requirements [3].

IV. Scaling Limits: Contact & Thermal Resistivities

Contact and thermal resistivities are presently the most serious barriers to further HBT scaling. By using Pd solid-phase-reaction base contacts [6], base contact resistivity is presently below 10 $\Omega - \mu\text{m}^2$. Emitter contact resistivity is also presently 10 $\Omega - \mu\text{m}^2$. Resistivity of *ex-situ* deposited contacts is strongly influenced by process history and the presence of surface oxides, and Pd contacts provide the lowest contact resistivities if efforts are made to minimize surface oxidation. Flared extrinsic emitter contacts formed by emitter regrowth [3] can reduce emitter resistance through a large ratio of contact to junction areas. A radical alternative, presently under investigation, is to avoid surface contaminants through *in-situ* formation of a semiconductor-semimetal (ErAs) contact during MBE growth [7].

After emitter resistivity, thermal resistance is the most serious scaling challenge. Emitter length L_e is constant with scaling, but emitter width $W_e \propto 1/\gamma^2$. Transistor spacing D must also scale as $1/\gamma$ to scale wiring delays. Approximating heat flow by cylindrical, spherical, and planar regions (Figure 3), the temperature rise between junction and package is

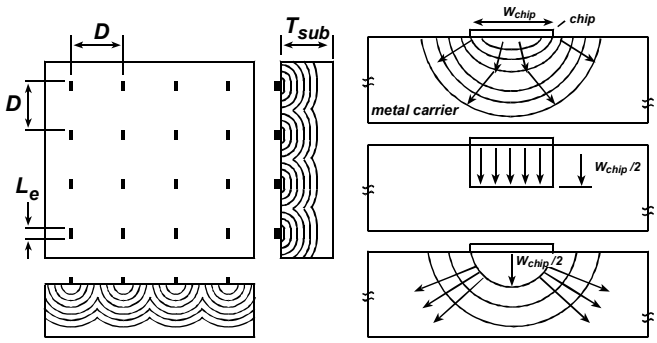

Figure 3: Thermal scaling analysis: heat flow geometry in the IC substrate (left) and in package (right)

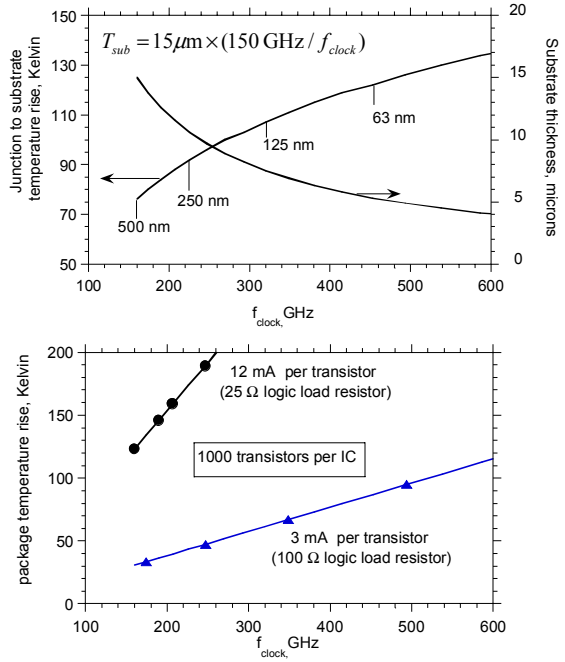

Figure 4: Computed temperature rise in IC substrate (above) and in package (below). Plots assume proportional scaling from our 150 GHz ICs, e.g., $L_e = 5\ \mu m$, I_c =5mA, V_{ce} =2 V, $W_e = 500\text{ nm}\,(160\text{ GHz} / f_{clock})^2$, $D = 20\ \mu\text{m}\,(160\text{ GHz} / f_{clock})$, and $T_{sub} = 15\ \mu\text{m}\,(160\text{ GHz} / f_{clock})$. HBT scaling generations (W_e from table 1) are indicated in the upper plot.

$$\Delta T_{InP} \cong \frac{P}{\pi K_{InP} L_E} \ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{InP}}\left(\frac{1}{L_E} - \frac{1}{D}\right) + \frac{P}{K_{InP}} \cdot \frac{T_{sub} - D/2}{D^2} \quad . \qquad (1)$$

P is the HBT dissipation, L_E the emitter length, and K_{InP} the thermal conductivity. The first term in (1) varies as $k_1 + k_2 \ln(\gamma)$, the second is small, and the third varies as γ^1. Only by extremely aggressively thinning the substrate in inverse proportion to the substrate thickness (Figure 4) can low junction temperature be maintained at a target 300 GHz

f_{clock}. Such thinning might be best realized with a thermal via process with Au via backfilling.

Heat flow in the package is similarly approximated by cylindrical and planar regions (Figure 3), giving

$$\Delta T_{package} \cong \left(\frac{2+\pi}{2\pi} \right) \frac{P_{chip}}{K_{Cu} W_{chip}} \propto \gamma . \qquad (2)$$

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 rises in a 1000-HBT IC are made acceptable only by restricting HBT switched logic currents to 3 mA per transistor (e.g. 100 Ω loading, vs. the 25 Ω loading used in [3])

V. Results

Present results [3] with 500-nm scaling generation HBTs include 450 GHz f_τ and 490 GHz f_{max} (Figure 5), and 142 GHz (Figure 6) and 150 GHz digital benchmark circuits (static frequency dividers). 172 GHz medium-power amplifiers have been realized [3] with 300-GHz-f_{max} HBTs (Figure 7).

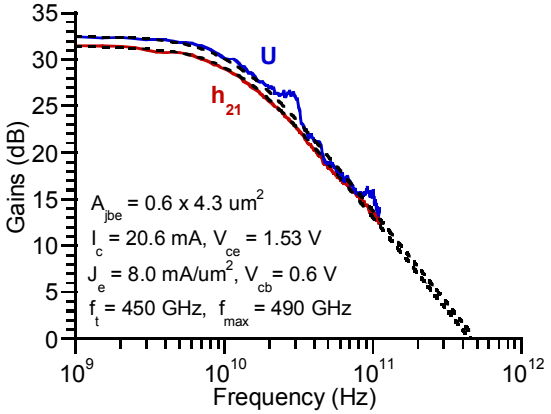

Figure 5: RF characteristics of and InP DHBT 120 nm collector thickness.-common-emitter breakdown is 3.9 V

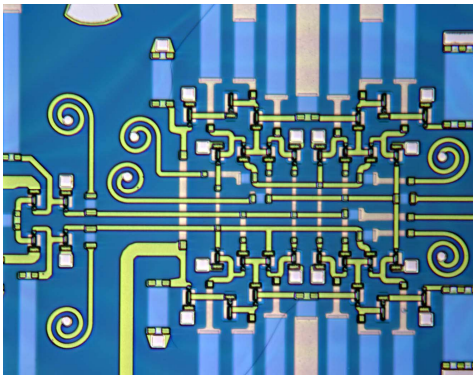

Figure 6: 142 GHz static frequency divider IC.

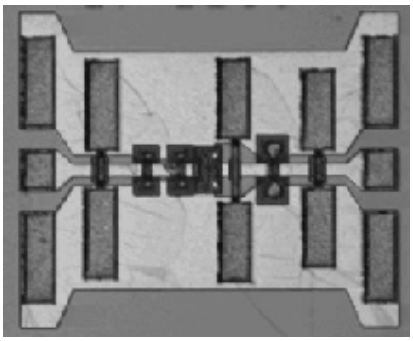

Figure 7: Common-base DHBT amplifier at 172 GHz. The IC has 8.3 dBm Psat and 6 dB small signal gain

Acknowledgement

Work supported by the Defense Advanced Research Projects Agency (DARPA) under the TFAST program N66001-02-C-8080, by the Office of Naval Research under N0014-04-1-0071, N00014-01-1-0024, and N00014-99-1-0041, and by the Jet Propulsion Laboratory (JPL) President's Fund

References

[1] T. Ishibashi, "Nonequilibrium Electron Transport in HBTs", IEEE Trans. Electron Devices, vol. 48, no. 11, pp. 2595-2605, 2001.

[2] M. Rodwell, M. Urteaga, T. Mathew, D. Scott, D. Mensa, Q. Lee, J. Guthrie, Y. Betser, S. C. Martin, R.P. Smith, S. Jaganathan, S. Krishnan, S. I. Long, R. Pullela, B. Agarwal, U. Bhattacharya, L. Samoska, and M. Dahlström, "Submicron Scaling of HBTs" IEEE Trans. Electron Devices, Vol. 48, pp. 2606-2624, November 2001.

[3] Z.Griffith, Y. Dong, D. Scott, Y. Wei, N. Parthasarathy, M. Dahlström, C. Kadow, V. Paidi, M. J. W. Rodwell, M. Urteaga, R. Pierson, P. Rowell, B. Brar, S. Lee, N. X. Nguyen, C. Nguyen , "Transistor and Circuit Design for 100–200-GHz ICs", IEEE Journal Solid-State Circuits, Vol. 40, No. 10, pp. 2061-2068, October 2005,

[4] M. Reddy, S.C. Martin, A.C. Molnar, R.E. Muller, R.P. Smith, P.H. Siegel, M.J. Mondry, M.J.W. Rodwell, and S.J. Allen, Jr., "Monolithic Schottky-Collector Resonant Tunnel Diode Oscillator Arrays to 650 GHz." IEEE Electron Device Letters, Volume 18, No. 5, May 1997, pp. 218-221

[5] I. Harrison, M. Dahlström, S. Krishnan, Z. Griffith, Y.-M. Kim, and M. J. W. Rodwell, "Thermal limitations of InP HBTs in 80- and 160-Gb ICs," IEEE Trans. Electron Devices, vol. 51, no. 4, pp. 529–534, Apr. 2004.

[6] E.F. Chor, D. Zhang, H. Gong, W.K. Chong, S. Y. Ong, "Electrical characterization, metallurgical investigation, and thermal stability studies of (Pd, Ti, Au)-based ohmic contacts". Journal of Applied Physics, vol.87, (no.5), AIP, 1 March 2000. p.2437-44.

[7] H. Kazemi, , E. R. Brown, A. C. Gossard, J. Zimmerman, G. D. Boreman, J. B. Hacker, "First MMW characterization of ErAs/InAlGaAs/InP semimetal-semiconductor-Schottky diode (S3) detectors for passive millimeter wave and infrared imaging", SPIE Defense and Security Symposium, 28 March-1 April 2005. Orlando Fl.