On the Feasibility of Few-THz Bipolar Transistors

University of California, Santa Barbara

Sponsors
J. Zolper, S. Pappert, M. Rosker
DARPA (TFAST, SWIFT, FLARE)
D. Purdy, I. Mack
Office of Naval Research
Kwok Ng, Jim Hutchby
Semiconductor Research Corporation

Collaborators (HBT)
M. Urteaga, R. Pierson, P. Rowell, M-J Choe, B. Brar
Teledyne Scientific Company
X. M. Fang, D. Lubyshev, Y. Wu, J. M. Fastenau, W.K. Liu
International Quantum Epitaxy, Inc.
S. Mohney
Penn State University

*Now at University of Texas, Austin

rodwell@ece.ucsb.edu  805-893-3244, 805-893-5705 fax
Specific Acknowledgements

(Prof.) Erik Lind
125 nm HBTs
process technology
theory / epi design

Dr. Zach Griffith
500 & 250 nm HBTs
150 GHz Logic
100 GHz op-amps
THz Transistors are coming soon; both InP & Silicon

InP Bipolars: 250 nm generation: → 780 GHz $f_{\text{max}}$, 424 GHz $f_\tau$, 4-5 V $BV_{\text{CEO}}$

125 nm & 62 nm nodes → ~THz devices

IBM IEDM '06: 65 nm SOI CMOS → 450 GHz $f_{\text{max}}$, ~1 V operation

Intel Jan '07: 45 nm / high-K / metal gate
continued rapid progress
→ continued pressure on III-V technologies

If you can't beat them, join them!
unclear if Si MOSFETs will work well at sub-22-nm gate length
InGaAs/InAs/InP channels under serious investigation for CMOS VLSI.

Datta, DelAlamo, Sadana, ...
THz InP vs. near-THz CMOS: different opportunities

65 / 45 / 33 / 22 ... nm CMOS
vast #s of very fast transistors
... having low breakdown, sloppy DC parameters
what NEW mm-wave applications will this enable?

massive monolithic mm-wave arrays → 1 Gb/s over ~1 km

DC parameters limit analog precision...
THz InP vs. near-THz CMOS: different opportunities

**InP HBT: THz bandwidths, good breakdown, analog precision**

- $f_{max} = 780$ GHz
- $f_t = 424$ GHz

340 GHz, 70 mW amplifiers (design)
In future: 700 or 1000 GHz amplifiers?

200 GHz digital logic (design)
In future: 450 GHz clock rate?

30-50 GHz gain-bandwidth op-amps $\rightarrow$ low IM3 @ 2 GHz
In future: 200 GHz op-amps for low-IM3 10 GHz amplifiers?
Transistor Benchmarks

\( f_{\text{max}} \) matters

- \( f_t = 660 \text{ GHz} \)
- \( f_{\text{max}} = 218 \text{ GHz} \)

Tuned amplifiers: \( f_{\text{max}} \) sets bandwidth

Mixed-signal:
- \( C_{cb} \Delta V / I_c, C_{je} \Delta V / I_c \),
- \( R_{ex} I_c / \Delta V, R_{bb} I_c / \Delta V \),
- \( \tau_f \)

Goal is >1 GHz \( f_t \) and \( f_{\text{max}} \)
<50 fs \( C \Delta V / I \) charging delays

BVCEO is not the only voltage limit

Need Safe Operating Area
...at least \( BV_{ceo}/2 \) at \( J_{\text{max}}/2 \)

thermal resistance,
high-current breakdown
high-temperature operation (~75°C)?

→ emphasize InP-collector DHBTs
HBT Scaling Laws
**Frequency Limits and Scaling Laws of (most) Electron Devices**

\[ \tau \propto \text{thickness} \]

\[ C \propto \text{area/thickness} \]

\[ R_{\text{top}} \propto \rho_{\text{contact}} / \text{area} \]

\[ R_{\text{bottom}} \propto 1 / \text{stripe length} \]

Applies to almost all semiconductor devices:

transistors: BJTs & HBTs, MOSFETs & HEMTs, Schottky diodes, photodiodes, photo mixers, RTDs, ...

Applies whenever AC signals are removed through Ohmic contacts

Diode lasers avoid \( R/C/\tau \) limits by radiating through end facets
HBT scaling laws

Goal: double transistor bandwidth when used in any circuit
→ keep constant all resistances, voltages, currents
→ reduce 2:1 all capacitances and all transport delays
InP DHBTs: September 2007

Equation: $f_t$ or $f_{\text{max}}$ alone

$(f_t + f_{\text{max}})/2$

$\sqrt{f_t f_{\text{max}}}$

$(1/f_t + 1/f_{\text{max}})^{-1}$

**Popular metrics:**

- $f_t$ or $f_{\text{max}}$ alone
- $(f_t + f_{\text{max}})/2$
- $\sqrt{f_t f_{\text{max}}}$
- $(1/f_t + 1/f_{\text{max}})^{-1}$

**Much better metrics:**

- Power amplifiers: PAE, associated gain, mW/$\mu$m
- Low noise amplifiers: $F_{\text{in}}$, associated gain, digital

- $f_{\text{clock}}$, hence
  - $(C_{cb} \Delta V / I_c)$
  - $(R_{ex} I_c / \Delta V)$
  - $(R_{bb} I_c / \Delta V)$
  - $(\tau_b + \tau_c)$
HBT Scaling Roadmaps
**multi-THz InP HBT Scaling Roadmap**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>512</th>
<th>256</th>
<th>128</th>
<th>64</th>
<th>32 nm width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emitter Width</td>
<td>16</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1 Ω⋅μm² access ρ</td>
</tr>
<tr>
<td>Base Width</td>
<td>300</td>
<td>175</td>
<td>120</td>
<td>60</td>
<td>30 nm contact width,</td>
</tr>
<tr>
<td>Collector Width</td>
<td>150</td>
<td>106</td>
<td>75</td>
<td>53</td>
<td>37.5 nm thick,</td>
</tr>
<tr>
<td>Collector Thickness</td>
<td>4.5</td>
<td>9</td>
<td>18</td>
<td>36</td>
<td>72 mA/μm² current density</td>
</tr>
<tr>
<td>fₜ</td>
<td>370</td>
<td>520</td>
<td>730</td>
<td>1000</td>
<td>1400 GHz</td>
</tr>
<tr>
<td>f_max</td>
<td>490</td>
<td>850</td>
<td>1300</td>
<td>2000</td>
<td>2800 GHz</td>
</tr>
<tr>
<td>Power Amplifiers</td>
<td>245</td>
<td>430</td>
<td>660</td>
<td>1000</td>
<td>1400 GHz</td>
</tr>
<tr>
<td>Digital Clock Rate</td>
<td>150</td>
<td>240</td>
<td>330</td>
<td>480</td>
<td>660 GHz</td>
</tr>
<tr>
<td>Static Dividers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HBT Scaling
Challenges
Scaling challenges: What looks easy, what looks hard?

<table>
<thead>
<tr>
<th>key device parameter</th>
<th>required change</th>
</tr>
</thead>
<tbody>
<tr>
<td>collector depletion layer thickness</td>
<td>decrease 2:1</td>
</tr>
<tr>
<td>base thickness</td>
<td>decrease 1.414:1</td>
</tr>
<tr>
<td>emitter junction width</td>
<td>decrease 4:1</td>
</tr>
<tr>
<td>collector junction width</td>
<td>decrease 4:1</td>
</tr>
<tr>
<td>emitter resistance per unit emitter area</td>
<td>decrease 4:1</td>
</tr>
<tr>
<td>current density</td>
<td>increase 4:1</td>
</tr>
<tr>
<td>base contact resistivity</td>
<td>decrease 4:1</td>
</tr>
<tr>
<td>(if contacts lie above collector junction)</td>
<td></td>
</tr>
</tbody>
</table>

Hard:

- Thermal resistance (particularly IC-level)
- Emitter contact + access resistance
- Base contact resistance
- Contact electromigration
- Yield in deep submicron processes

High current density, low resistivity contacts, epitaxial & lithographic scaling ➔ THz semiconductor devices
Thermal Resistance Scaling: Transistor, Substrate, Package

\[ \Delta T_{\text{substrate}} \approx \frac{P}{\pi K_{\text{InP}} L_e} \ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{\text{InP}}} \left( \frac{1}{L_e} - \frac{1}{D} \right) + \frac{P}{K_{\text{InP}}} \left( \frac{T_{\text{sub}} - D/2}{D^2} \right) \]

- cylindrical heat flow near junction increases logarithmically
- spherical flow for \( r > L_e \) insignificant variation
- planar flow for \( r > D_{\text{HBT}} / 2 \) increases quadratically if \( T_{\text{sub}} \) is constant

\[ \Delta T_{\text{package}} \approx \left( \frac{1}{4} + \frac{1}{\pi} \right) \frac{P_{\text{chip}}}{K_{\text{Cu}} W_{\text{chip}}} \]

Wiring lengths, clock rates, power densities, etc. scaled from demonstrated 150 GHz digital ICs
Breakdown Voltage Scaling: Expect 2.4 V \( @ 1 \text{ THz} f_\tau \)

For mature, well-scaled InP DHBTs, \( f_\tau \times BVCEO = 2.4 \text{ THz-Volts} \).
HBTs:

500 nm Generation
512 nm InP DHBT

Laboratory Technology

500 nm mesa HBT

150 GHz M/S latches

175 GHz amplifiers

Production

( Teledyne )

500 nm sidewall HBT

DDS IC: 4500 HBTs

38 GHz op-amps

Z. Griffith
M. Urteaga
P. Rowell
D. Pierson
B. Brar
V. Paidi

\[ f_r = 405 \text{ GHz} \]
\[ f_{\text{max}} = 392 \text{ GHz} \]
\[ V_{br, ceo} = 4 \text{ V} \]
HBTs:

250 nm Generation
256 nm Generation InP DHBT

150 nm thick collector

$H_{21}$

$f_{\text{max}} = 780 \text{ GHz}$

$f = 424 \text{ GHz}$

70 nm thick collector

$H_{21}$

$f_{\text{max}} = 560 \text{ GHz}$

$f = 560 \text{ GHz}$

60 nm thick collector

$H_{21}$

$f_{\text{max}} = 218 \text{ GHz}$

$f = 660 \text{ GHz}$

Z. Griffith, E. Lind, J. Hacker, M. Jones
125 nm InP HBT development
**125 nm Technology Development**

**New Emitter Process**

*First results are at 250 nm emitter width*

*Scalable below 128 nm width*

Simultaneously 560 GHz $f_t$ & $f_{\text{max}}$

$BV_{ceo} = 3.3V$

...can do much better...
**128 nm InP HBT: Technology Development**

**New Emitter Process for 128 and 64 nm junctions**
- **dry etched metal**
- **dry etched junction**
- **refractory W or Mo contact** → stable at very high $J_e < 0.8 \Omega \cdot \mu m^2$ contact resistivity

**New, thin --12 nm -- base-collector grade:**
- most of collector is high-$E_g$ InP
  → does not degrade $V_{brceo}$
- grade sufficiently thin even for 64 nm HBTs

Alternative epi layer designs (InP/GaAsSb/InP) are not necessary

**First results:** close but not perfect
- slip-ups: wide 250 nm emitters, poor base contacts
- only a 560GHz / 560GHz / 3 V device
- target was 700 / 700 / 3 ... try again soon...
Improvements in Emitter Access Resistance

125 nm generation requires 5 $\Omega \cdot \mu m^2$ emitter resistivities

65 nm generation requires 1-2 $\Omega \cdot \mu m^2$

**Recent Results (ONR contacts program)**

<table>
<thead>
<tr>
<th>Material</th>
<th>Process</th>
<th>Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ErAs/Mb</td>
<td>MBE in-situ</td>
<td>1.5 $\Omega \cdot \mu m^2$</td>
</tr>
<tr>
<td>Mb</td>
<td>MBE in-situ</td>
<td>0.6 $\Omega \cdot \mu m^2$</td>
</tr>
<tr>
<td>TiPdAu</td>
<td>ex-situ</td>
<td>0.5 $\Omega \cdot \mu m^2$</td>
</tr>
<tr>
<td>TiW</td>
<td>ex-situ</td>
<td>0.7 $\Omega \cdot \mu m^2$</td>
</tr>
</tbody>
</table>

Degeneracy contributes 1 $\Omega \cdot \mu m^2$  
20 nm emitter-base depletion layer contributes 1 $\Omega \cdot \mu m^2$ resistance
Current UCSB TiW emitter process

- 5 nm Ti layer for improved adhesion
- 25 nm SiN\textsubscript{x} sidewalls protects Ti/TiW during Cl\textsubscript{2} and BHF etch, improves adhesion
- Standard triple mesa
- BCB passivation

Emitter prior to InP wet etch
RF & DC data – 70 nm collector, 22 nm base InP Type-I DHBT

Emitter width ~ 250 nm

First reported device with \( f_t, f_{\text{max}} > 500 \) GHz

\( \text{BV}_{CEO} \sim 3.3 \) V, \( \text{BV}_{CBO} = 3.9 \) V (\( J_{e,c} = 15\text{kA/cm}^2 \))

Emitter contact (from RF extraction), \( R_{\text{cont}} < 5 \Omega \cdot \mu \text{m}^2 \)

Base: \( R_{\text{sheet}} = 780 \Omega/\text{sq}, R_{\text{cont}} \sim 15 \Omega \cdot \mu \text{m}^2 \)

Collector: \( R_{\text{sheet}} = 11.1 \Omega/\text{sq}, R_{\text{cont}} \sim 10.1 \Omega \cdot \mu \text{m}^2 \)
64 nm & 32 nm (THz) InP HBT
### 64 & 32 nm Generations: Example Process Flow

<table>
<thead>
<tr>
<th>Step</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emitter Contact</strong></td>
<td>![Emitter Contact Diagram]</td>
</tr>
<tr>
<td><strong>Emitter Litho</strong></td>
<td>![Litho Diagram]</td>
</tr>
<tr>
<td><strong>Plate Contact</strong></td>
<td>![Contact Diagram]</td>
</tr>
<tr>
<td><strong>Etch, Sidewalls</strong></td>
<td>![Etch Diagram]</td>
</tr>
<tr>
<td><strong>Base Contact</strong></td>
<td>![Base Contact Diagram]</td>
</tr>
<tr>
<td><strong>Sputter Planarize</strong></td>
<td>![Planarize Diagram]</td>
</tr>
<tr>
<td><strong>Etch</strong></td>
<td>![Etch Metal Diagram]</td>
</tr>
<tr>
<td><strong>Base Mesa</strong></td>
<td>![Mesa Diagram]</td>
</tr>
<tr>
<td><strong>Collector Isolation</strong></td>
<td>![Isolation Diagram]</td>
</tr>
<tr>
<td><strong>Collector Contacts</strong></td>
<td>![Collector Contacts Diagram]</td>
</tr>
<tr>
<td><strong>Posts</strong></td>
<td>![Posts Diagram]</td>
</tr>
<tr>
<td><strong>Planarization</strong></td>
<td>![Planarization Diagram]</td>
</tr>
<tr>
<td><strong>Interconnects</strong></td>
<td>![Interconnects Diagram]</td>
</tr>
</tbody>
</table>

**Key Elements**

- **MBE Mo**
- **Emitter Cap**
- **Emitter**
- **Base**
- **N-drift Collector**
- **Sub Collector**
- **Substrate**

**Remarks**

- One critical lithographic step, no critical alignments
- Refractory base and emitter contacts
- Base-emitter ledge for leakage current control
Reliability...

...depends upon stress and upon device structure.

high current density $\rightarrow$ heating
thermal design is critical

high current density $\rightarrow$ contact electromigration
need refractory (W, TiW, Mo,...) contacts

must investigate failure mechanisms
驱动 by high current density in semiconductor
dark-line defects

Given the need for large I/C charging rates, high
current density is unavoidable.
Interconnects: Substrate Microstrip Has Problems

Thick Substrate → low skin loss

\[ \alpha_{\text{skin}} \propto \frac{1}{\varepsilon_r^{1/2} H} \]

Zero ground inductance in package

No ground plane breaks in IC

Brass carrier and assembly ground

IK with backside ground plane & vias

High via inductance

Strong coupling when substrate approaches \( \lambda/4 \) thickness

TM substrate mode coupling

12 \( \varepsilon \)H for 100 \( \mu \)m substrate → 7.5 \( \Omega \) @ 100 GHz

Ground vias must be widely spaced

lines must be widely spaced

Line spacings must be \( \sim 3\) (substrate thickness)

all factors require very thin substrates for >100 GHz ICs

→ lapping to \( \sim 50 \) \( \mu \)m substrate thickness typical for 100+ GHz
Interconnect: Coplanar Waveguide Has Many Problems!

No ground vias
No need (???) to thin substrate

Hard to ground IC to package

Parasitic microstrip mode

Parasitic slot mode

ground plane breaks → loss of ground integrity

substrate mode coupling or substrate losses

→ substrate *must* be thinned!

Repairing ground plane with ground straps is effective only in simple ICs
In more complex CPW ICs, ground plane rapidly vanishes
→ common-lead inductance → strong circuit-circuit coupling

III-V: semi-insulating substrate → substrate mode coupling → must thin wafer to λ/2, must have vias to kill microstrip mode.

Silicon conducting substrate → substrate conductivity losses

poor ground integrity

loss of impedance control

ground bounce

coupling, EMI, oscillation

40 Gb/s differential TWA modulator driver note CPW lines, fragmented ground plane

35 GHz master-slave latch in CPW note fragmented ground plane

175 GHz tuned amplifier in CPW note fragmented ground plane
THz IC Interconnects will be Silicon-Like

Microstrip wiring; metal 4 ground plane

P-doped substrate, junction isolation → kill substrate modes, negligible substrate capacitance

Through-wafer thermal vias (etch stop) → address device & IC thermal scaling

Through-wafer electrical vias (no etch stop) → low-ground-bounce IC-package connection
Comparison to SiGe
SiGe Today: Parasitic Reduction for Increased Bandwidth

wide emitter contact: low resistance
narrow emitter junction: scaling (low $R_{bb}/A_e$)

thick extrinsic base: low resistance
thin intrinsic base: low transit time

wide base contacts: low resistance
narrow collector junction: low capacitance

These are planar approximations to radial contacts:

- $R_{bulk} = \frac{2\rho_{bulk}}{\pi L} \ln\left(\frac{\sqrt{2} \cdot r}{W}\right)$
- $R_{contact} = \frac{2\rho_c}{\pi L r}$
- $R_{\text{total,min}} = \frac{2\rho_{bulk}}{\pi L} \left[1.34 + \ln\left(\frac{\rho_{contact}}{W \rho_{bulk}}\right)\right]$
Parasitic Reduction Could Help Less with Small Devices

Scaling for increased HBT bandwidth:

$Lateral$ $dimensions$ $vary$ $as$ $\sim$ $(bandwidth)^{-1}$, $vertical$ $as$ $\sim$ $(bandwidth)^{-2}$.

$\rightarrow$ Proportionally larger parasitic capacitances from extrinsic contact regions.

Let us compare intrinsic device structures...
**InP vs. SiGe: Comparison with Intrinsic Device**

**Assumption:** Mesa structure, change from SiGe to InP, keep the same bandwidth

\[ \tau_c = \frac{T_c}{2v} \frac{C_{cb}}{I_c} \approx \frac{A_c}{A_e} \left( \frac{\tau_c}{2} + \frac{V_{cb,app} + \phi}{V_{cb,app}} \right) \]

\( v \) increased 3.5:1 \( \rightarrow \) increase collector thickness 3.5:1

\[ J_{Kirk} = \frac{2e\nu(V_{cb,applied} + V_{cb,depletion} + 2\phi)}{T_c^2} \]

\( J \) decreases 3.5:1 \( \rightarrow \) junction areas increase 3.5:1

\[ \Delta T \approx \frac{P}{\pi K_{Th} L_e} \ln \left( \frac{L_e}{W_e} \right) + \frac{P}{\pi K_{Th} L_e} \]

For equal \( \Delta T \), \( L_e \) remains constant, \( W_e \) increases 3.5:1

\( (K_{th} \) is higher in Si, but Si needs SiO_2 trenches)

Required lithographic feature size is \( \sim 3.5:1 \) larger

\[ R_{ex} = \frac{\rho_c}{A_e} \]

Required contact resistivities are similarly relaxed

\[ R_{bb} \approx \frac{\rho_s W_e}{12L_e} + \frac{\rho_s W_{bc}}{6L_e} + \frac{\rho_c}{A_{contacts}} \]

\( \rightarrow \) about 3.5:1 larger dimensions, 3.5:1 larger breakdown for a given device bandwidth

\( \rightarrow \) higher bandwidths achieved due to high velocities and low resistance contacts
On the Feasibility of Few-THz Bipolar Transistors

InP Bipolar Transistors
Scaling limits: contact resistivities, device and IC thermal resistances.
62 nm (1 THz \( f_{\tau} \), 1.5 THz \( f_{\text{max}} \)) scaling generation is feasible.

700 GHz amplifiers, 450 GHz digital logic

Is the 32 nm (1 THz amplifiers) generation feasible?

SiGe Bipolar Transistors
Sophisticated device structure \( \rightarrow \) harder to project further progress

Contact + access resistivities & thermal resistivities are key scaling limits
non-animated versions of the three key scaling slides
HBT scaling laws

Goal: double transistor bandwidth when used in any circuit
→ keep constant all resistances, voltages, currents
→ reduce 2:1 all capacitances and all transport delays

\[ \tau_b = \frac{T_b^2}{2D_n + T_b/v} \rightarrow \text{thin base } \sim 1.414:1 \]
\[ \tau_c = \frac{T_c}{2v} \rightarrow \text{thin collector } 2:1 \]

\[ C_{cb} \propto \frac{A_c}{T_c} \rightarrow \text{reduce junction areas } 4:1 \]
\[ R_{ex} = \rho_e/A_e \rightarrow \text{reduce emitter contact resistivity } 4:1 \]
\[ I_{c,Kirk} \propto \frac{A_e}{T_c^2} \text{ (current remains constant, as desired) } \]

\[ \Delta T \approx \frac{P}{\pi K_{InP} L_E} \ln \left( \frac{L_e}{W_e} \right) + \frac{P}{\pi K_{InP} L_E} \]
\[ \text{need to reduce junction areas } 4:1 \]
\[ \text{reduce widths } 2:1 \text{ & reduce length } 2:1 \rightarrow \text{doubles } \Delta T \times \]
\[ \text{reducing widths } 4:1, \text{ keep constant length} \rightarrow \text{small } \Delta T \text{ increase } \checkmark \]

\[ R_{bb} \approx \frac{\rho_s W_e}{12L_e} + \frac{\rho_s W_{bc}}{6L_e} + \frac{\rho_c}{A_{contacts}} \rightarrow \text{reduce base contact resistivity } 4:1 \]
\[ \text{reduce widths } 2:1 \text{ & reduce length } 2:1 \rightarrow \text{constant } R_{bb} \checkmark \]
\[ \text{reducing widths } 4:1, \text{ keep constant length} \rightarrow \text{reduced } R_{bb} \checkmark \checkmark \]

Linewidths scale as the inverse square of bandwidth because thermal constraints dominate.
First-Order HBT Design

\[ \tau_b = T_b^2 / 2D_n + T_b / v \]

\[ \tau_c = T_c / 2v \]

\[ R_{ex} = \rho_c / A_e \]

\[ R_{bb} \approx \frac{\rho_s W_e}{12L_e} + \frac{\rho_s W_{bc}}{6L_e} + \frac{\rho_c}{A_{contacts}} \]

\[ C_{cb} = \varepsilon A_c / T_c \]

\[ I_{c,Kirk} = 2\varepsilon v_{eff} L_E (W_E + 2T_c)(V_{cb} + V_{cb,depletion} + 2\phi) / T_c^2 \]

\[ C_{cb} / I_c = (A_c / A_e) \cdot \tau_c \cdot (V_{cb,app} + V_{cb,depl} + 2\phi)^{-1} \]

\[ \Delta T \approx \frac{P}{\pi K_{InP} L_E} \ln \left( \frac{L_e}{W_e} \right) + \frac{P}{\pi K_{InP} L_E} \]