Invited Paper, 2007 IEEE Bipolar Circuits and Technology Meeting, September 30 - October 3, 2007

On the Feasibility of Few-THz Bipolar Transistors

M. Rodwell, E. Lind, Z. Griffith, S. R. Bank, A. M. Crook*, U. Singisetti, M. Wistey, G. Burek, and A. C. Gossard*
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

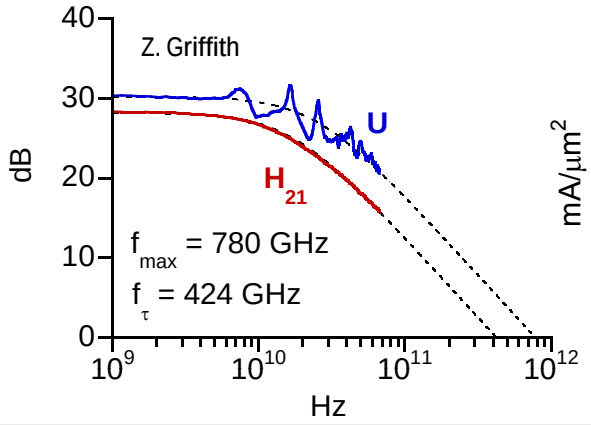
Dr. Zach Griffith

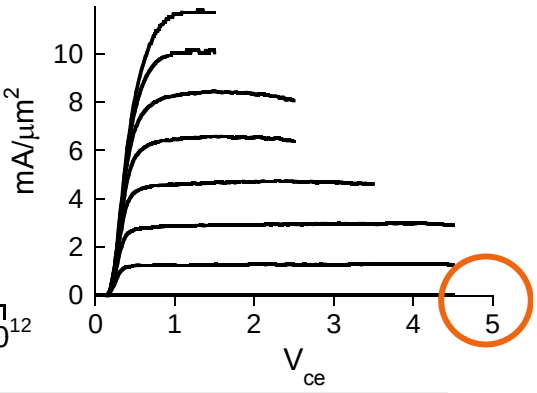
500 & 250 nm HBTs
150 GHz Logic
100 GHz op-amps

THz Transistors are coming soon; both InP & Silicon

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

125 nm & 62 nm nodes
→ ~THz devices

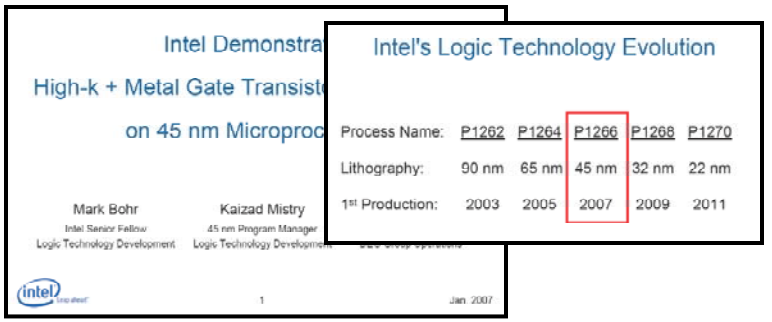




IBM IEDM '06: 65 nm SOI CMOS → 450 GHz f_{max}, ~1 V operation

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, ...

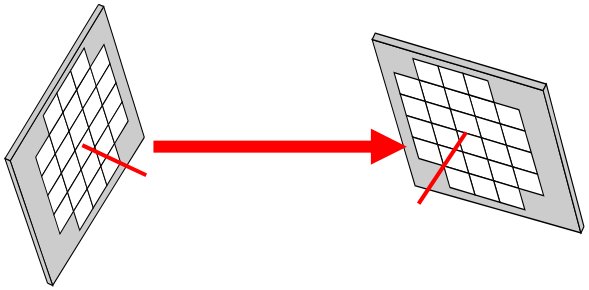
THz InP vs. near-THz CMOS: different opportunities

65 / 45 / 33 / 22 ... nm CMOS

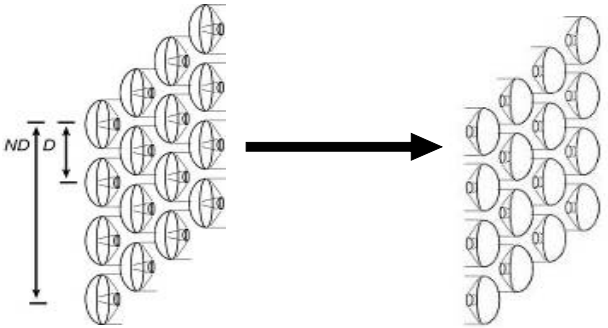
vast #s of very fast transistors

... having low breakdown, sloppy DC parameters

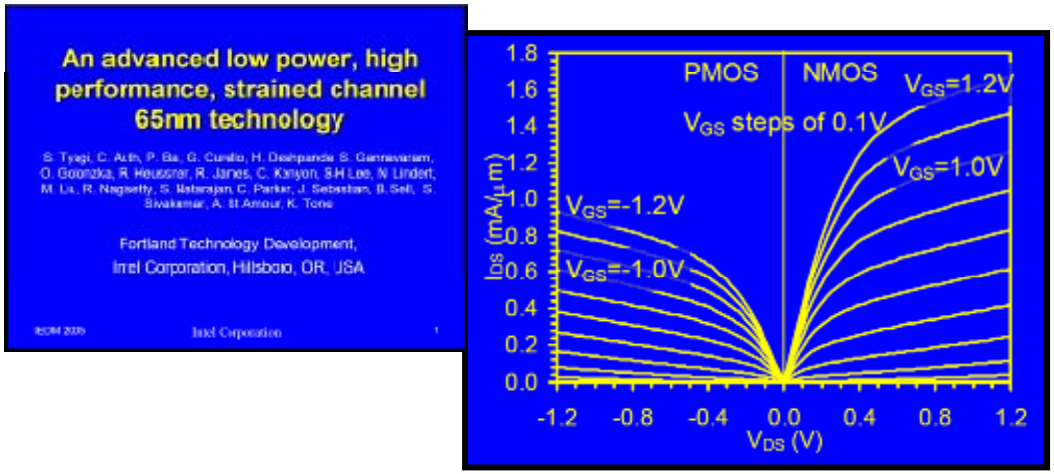
what NEW mm-wave applications will this enable ?



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



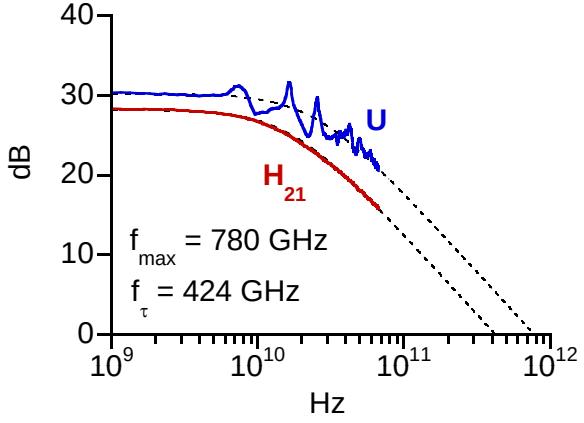
mm-wave MIMO



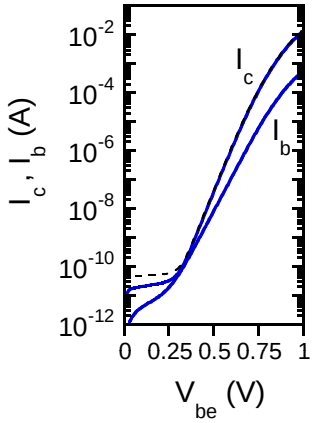
DC parameters
limit analog precision...

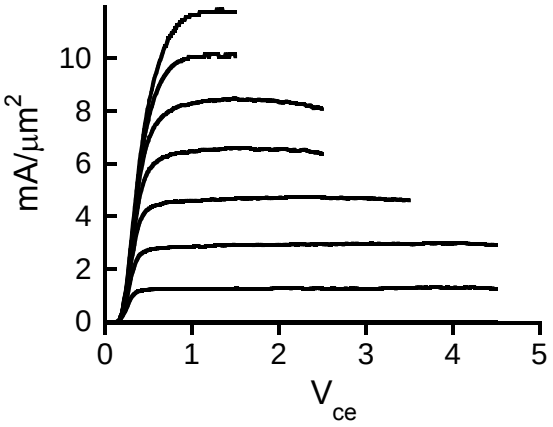
THz InP vs. near-THz CMOS: different opportunities

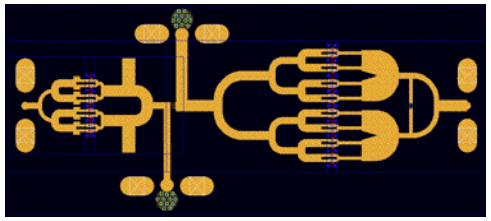
InP HBT: THz bandwidths, good breakdown, analog precision



&



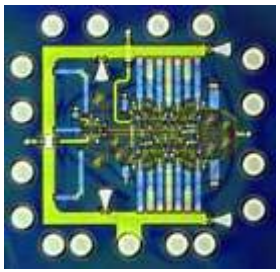




M. Jones

340 GHz, 70 mW amplifiers (design)

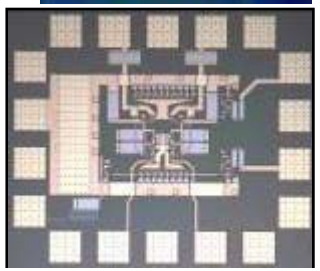
In future: 700 or 1000 GHz amplifiers ?



Z. Griffith

200 GHz digital logic (design)

In future: 450 GHz clock rate ?



M. Urteaga (Teledyne)

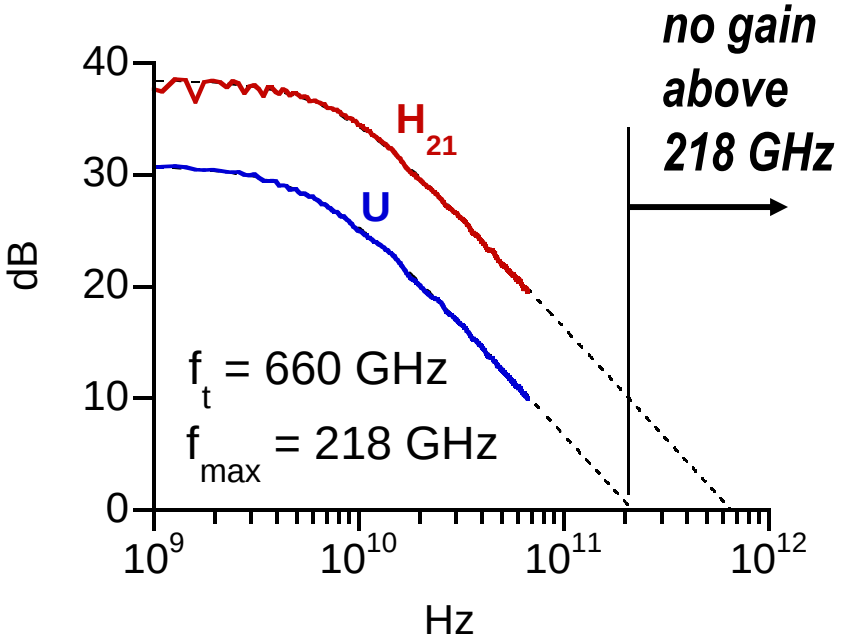
Z. Griffith

30-50 GHz gain-bandwidth op-amps→ low IM3 @ 2 GHz

In future: 200 GHz op-amps for low-IM3 10 GHz amplifiers?

Transistor Benchmarks

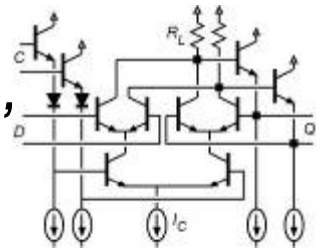
f_{max} matters



Tuned amplifiers: f_{max} sets bandwidth

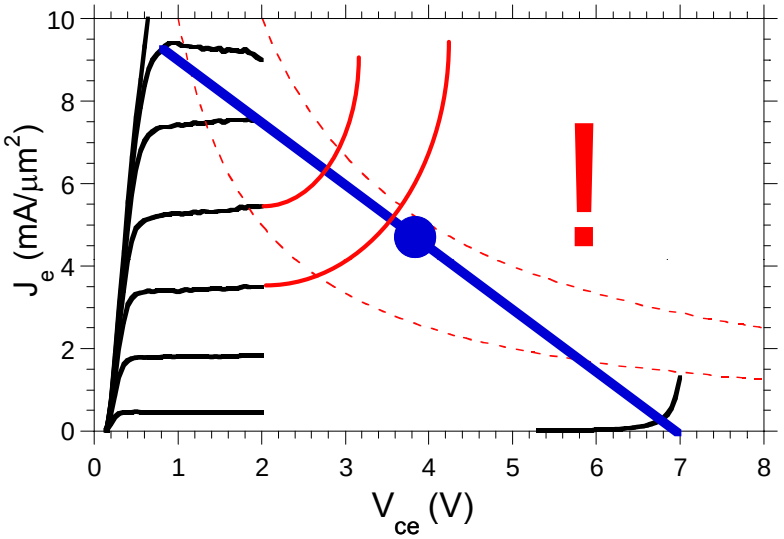
Mixed-signal:

$C_{cb}\Delta V/I_c$, $C_{je}\Delta V/I_c$,
$R_{ex}I_c/\Delta V$, $R_{bb}I_c/\Delta V$,
τ_f



Goal is >1 THz f_τ and f_{max}
<50 fs $C\Delta V / I$ charging delays

BVCEO is not the only voltage limit



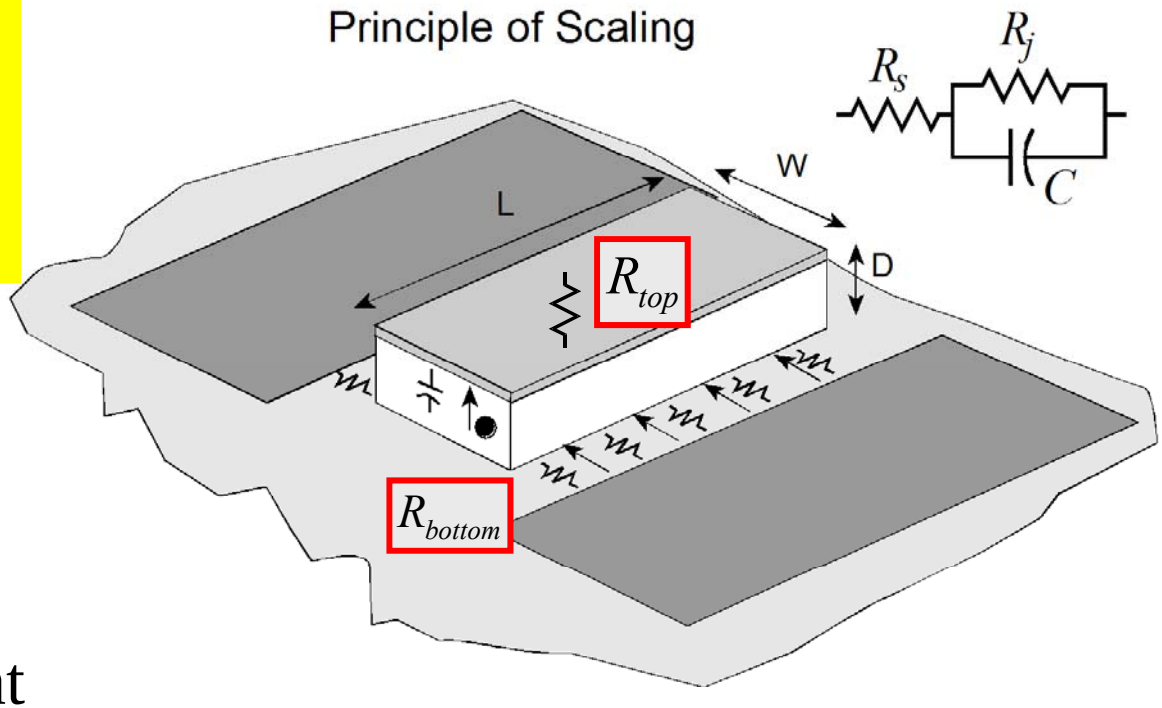
Need Safe Operating Area
...at least $BV_{ceo}/2$ at $J_{max}/2$

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



$\tau \propto$ thickness

$C \propto$ area / thickness

$R_{top} \propto \rho_{contact}$ / area

$R_{bottom} \propto 1/$ stripe lenght

resistance *capacitance* *transit time*

device bandwidth

applies to almost all semiconductor devices:

transistors: BJTs & HBTs, MOSFETS & HEMTs,
Schottky diodes, photodiodes, photo mixers, RTDs, …

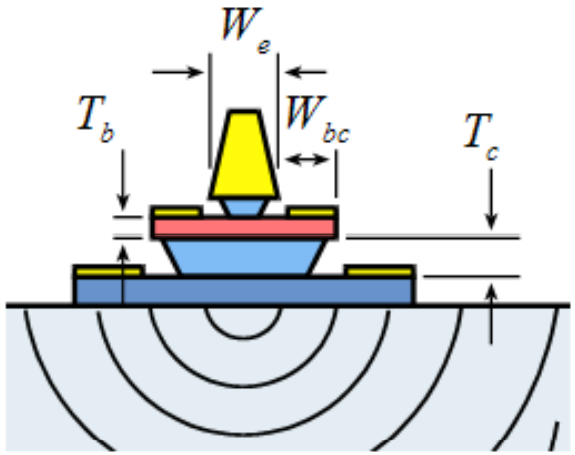
Applies whenever AC signals are removed though Ohmic contacts
Diode lasers avoid R/C/τ limits by radiating through end facets

HBT scaling laws

*Goal: double transistor bandwidth when used in **any** circuit*

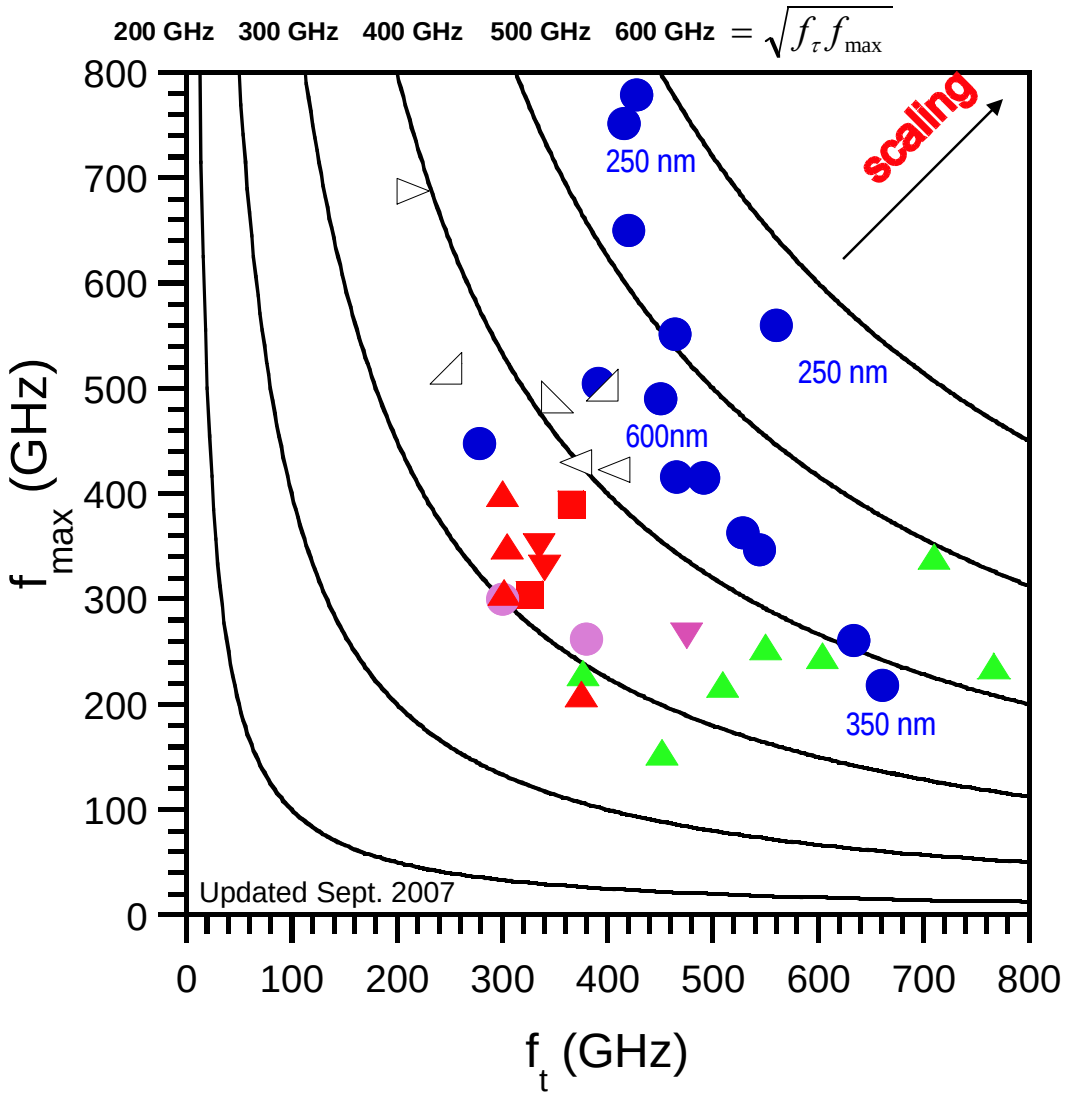
→ keep constant all resistances, voltages, currents

→ reduce 2:1 all capacitances and all transport delays



(emitter length L_E)

InP DHBTs: September 2007



- Teledyne
- UIUC DHBT
- NTT
- ETHZ/SFU
- UIUC SHBT
- UCSB DHBT
- NGST
- Pohang
- HRL
- IBM SiGe
- Vitesse

popular metrics :

f_τ or f_{max} alone

$(f_\tau + f_{max})/2$

$\sqrt{f_\tau\, f_{max}}$

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

much better metrics :

power amplifiers:

PAE, associated gain, mW/μm

low noise amplifiers:

F_{min}, associated gain,

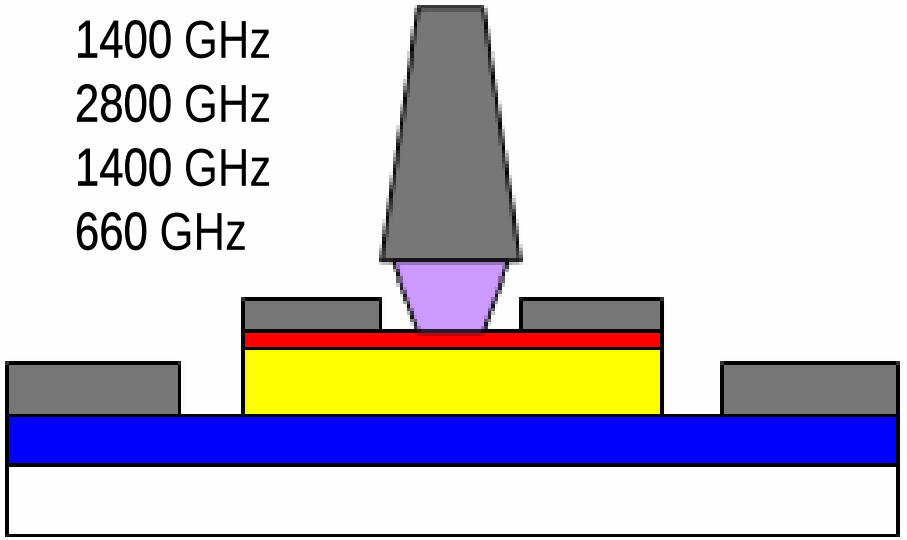
digital:

f_{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

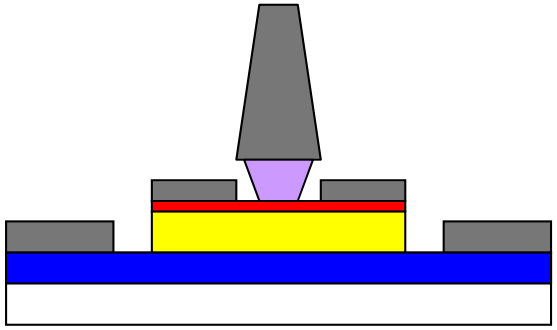
multi-THz InP HBT Scaling Roadmap

emitter	512	256	128	64	32 nm width
	16	8	4	2	1 $\Omega\cdot\mu m^2$ access ρ
base	300	175	120	60	30 nm contact width,
	20	10	5	2.5	1.25 $\Omega\cdot\mu m^2$ contact ρ
collector	150	106	75	53	37.5 nm thick,
	4.5	9	18	36	72 mA/μm^2 current density
	4.9	4	3.3	2.75	2-2.5 V, breakdown
f_τ	370	520	730	1000	1400 GHz
f_{max}	490	850	1300	2000	2800 GHz
power amplifiers	245	430	660	1000	1400 GHz
digital clock rate (static dividers)	150	240	330	480	660 GHz



HBT Scaling Challenges

Scaling challenges: What looks easy, what looks hard ?



key device parameter	required change
collector depletion layer thickness	decrease 2:1
base thickness	decrease 1.414:1
emitter junction width	decrease 4:1
collector junction width	decrease 4:1
emitter resistance per unit emitter area	decrease 4:1
current density	increase 4:1
base contact resistivity (if contacts lie above collector junction)	decrease 4:1

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

cylindrical heat flow near junction

spherical flow for $r > L_e$

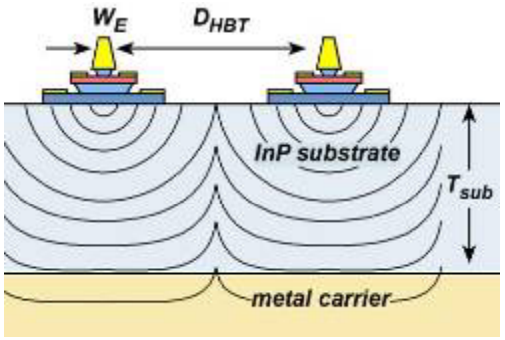
planar flow for $r > D_{HBT}/2$

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

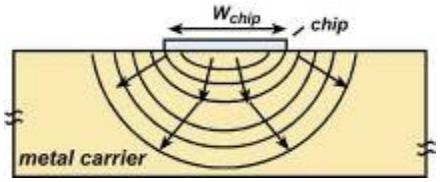
increases logarithmically

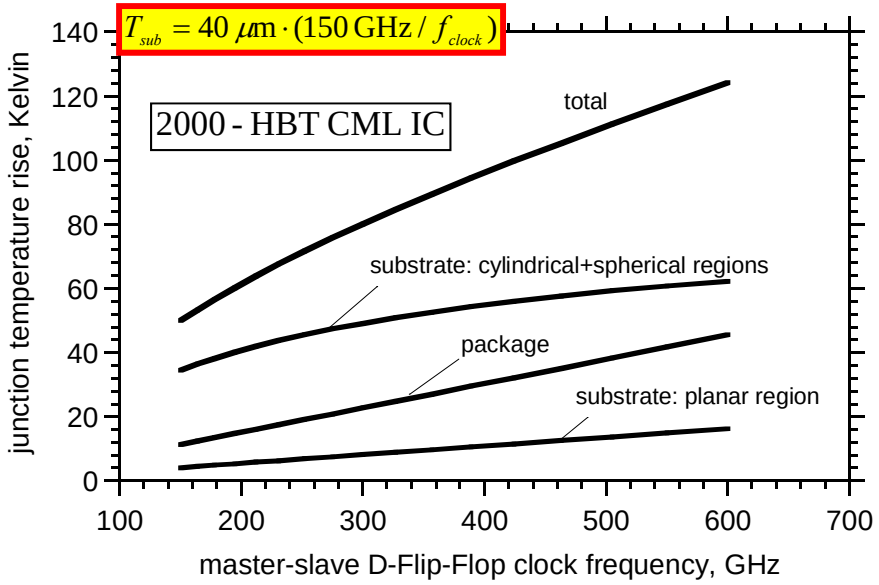
insignificant variation

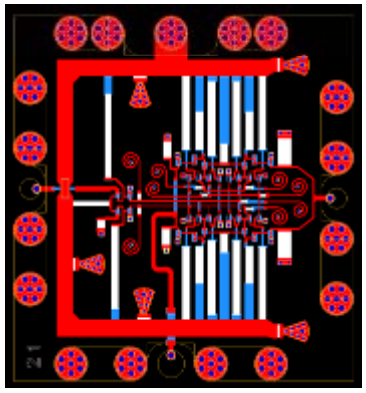
increases quadratically if T_{sub} is constant



$$\Delta T_{package} \cong \left(\frac{1}{4} + \frac{1}{\pi}\right) \frac{P_{chip}}{K_{Cu} W_{chip}}$$

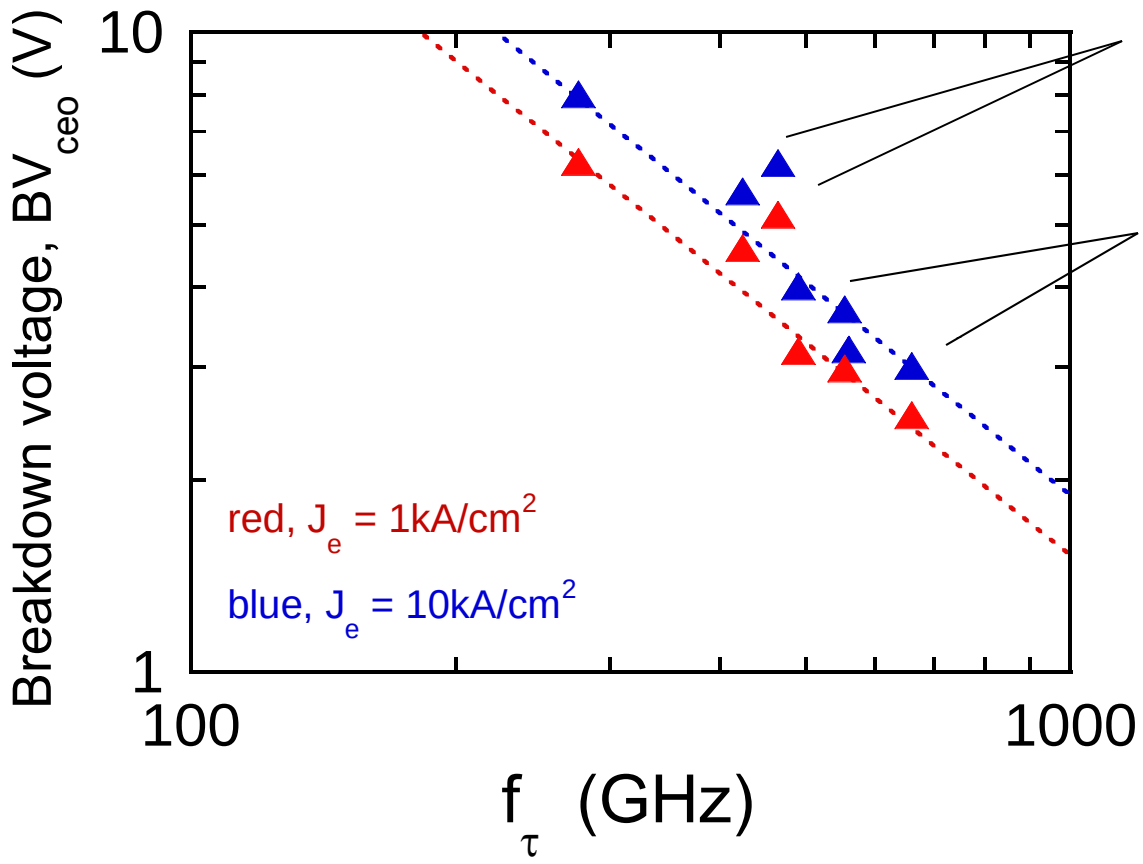






Wiring lenghts, clock rates, power densities, etc. scaled from demonstrated 150 GHz digital ICs

Breakdown Voltage Scaling: Expect 2.4 V @ 1 THz f_τ



mature devices--
full potential f_τ achieved

newer, immature devices-
reduced f_τ from:
high contact parasitics
insufficient lateral scaling

For mature, well-scaled InP DHBTs, f_τ x BVCEO = 2.4 THz-Volts.

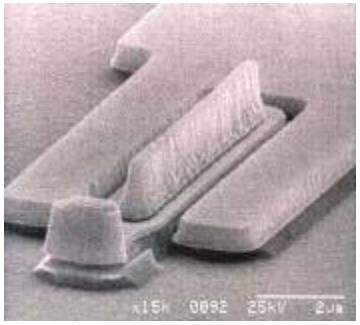
HBTs:

500 nm Generation

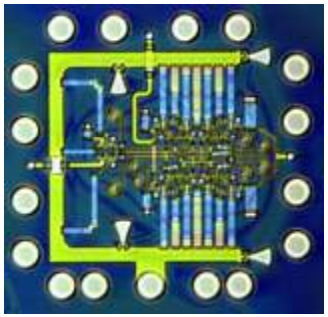
512 nm InP DHBT

Laboratory Technology

500 nm mesa HBT

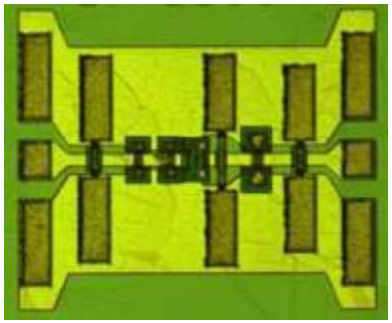


150 GHz M/S latches



UCSB / Teledyne / GCS

175 GHz amplifiers

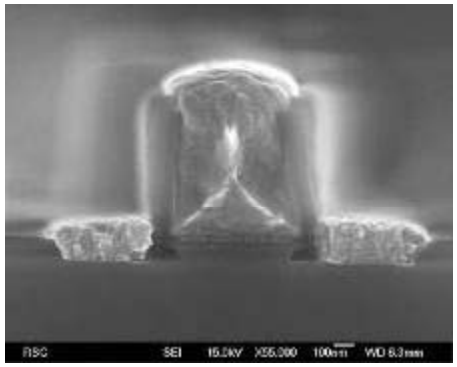


UCSB

Production

(Teledyne)

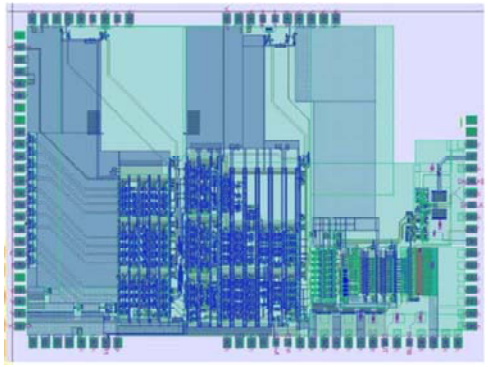
500 nm sidewall HBT



Teledyne

f_τ = 405 GHz
f_{max} = 392 GHz
$V_{br,\ ceo}$ = 4 V

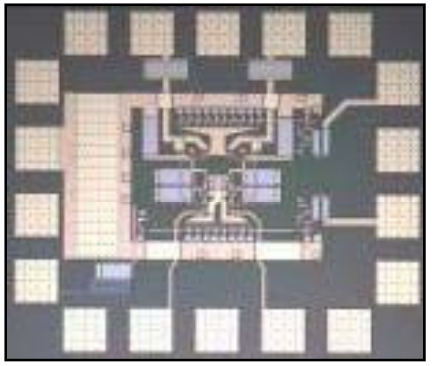
DDS IC: 4500 HBTs



Teledyne / BAE

20 GHz clock

38 GHz op-amps



Teledyne / UCSB

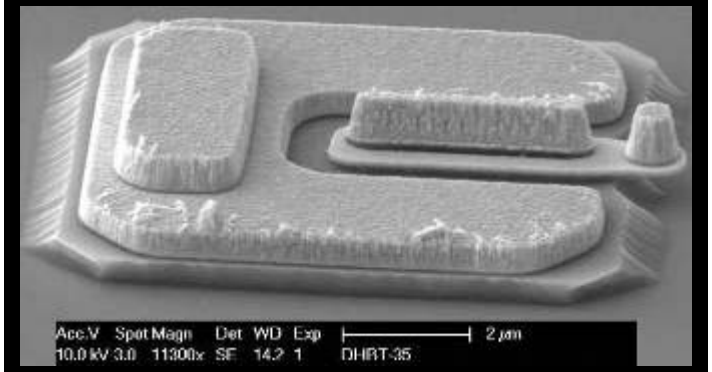
46 dBm OIP3 @ 2 GHz with 1 W dissipation

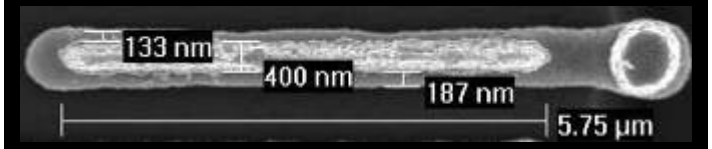
Z. Griffith
M. Urteaga
P. Rowell
D. Pierson
B. Brar
V. Paidi

HBTs:

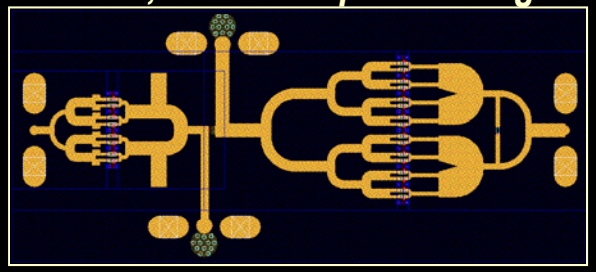
250 nm Generation

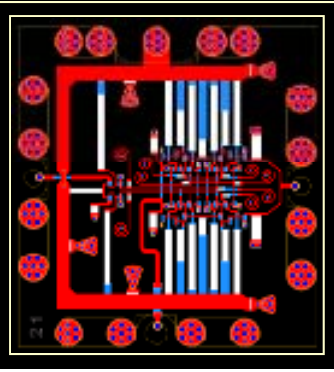
256 nm Generation InP DHBT





340 GHz, 70 mW amplifier design

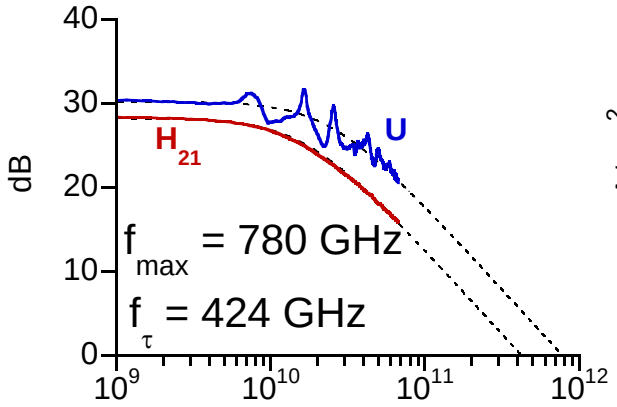


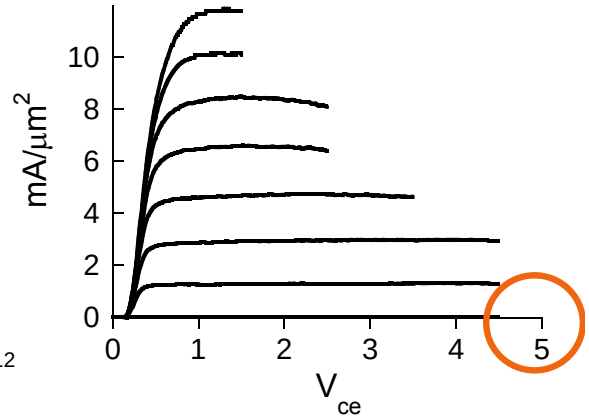


200 GHz master-slave latch design

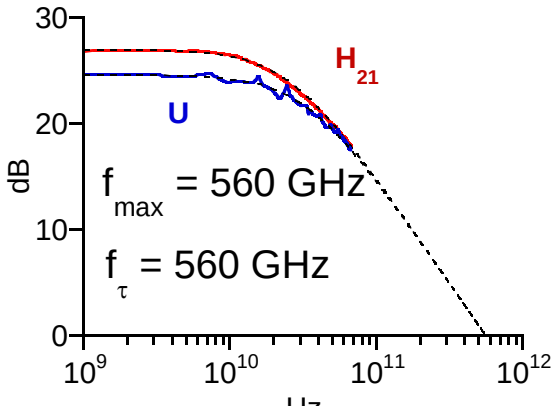
Z. Griffith, E. Lind, J. Hacker, M. Jones

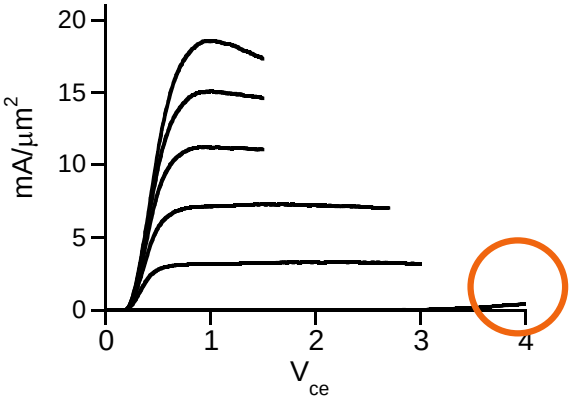
150 nm thick collector



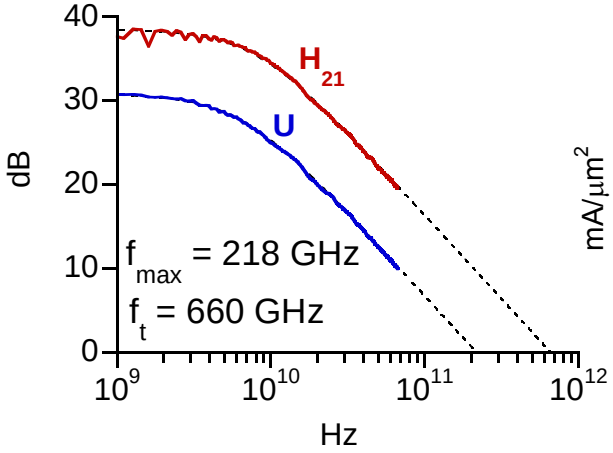


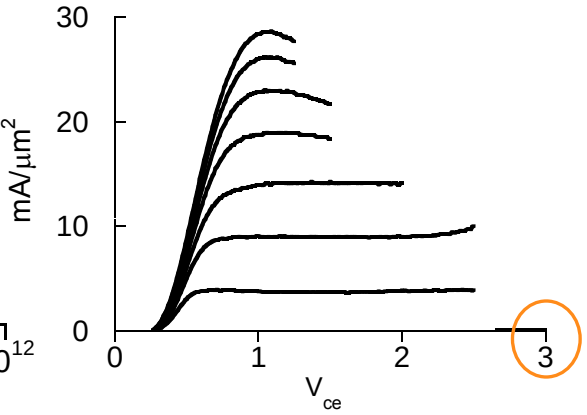
70 nm thick collector





60 nm thick collector



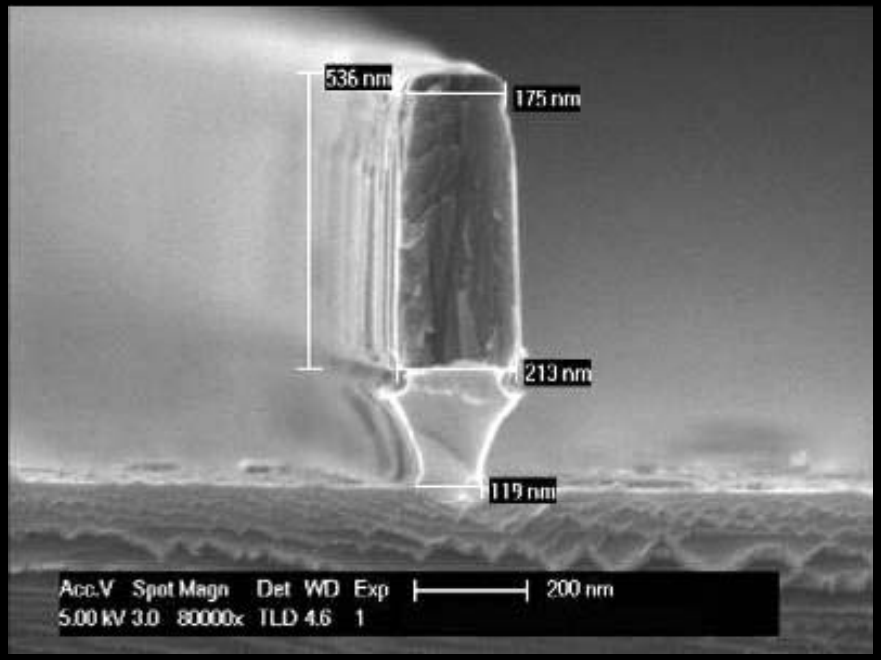


125 nm InP HBT development

E. Lind

125 nm Technology Development

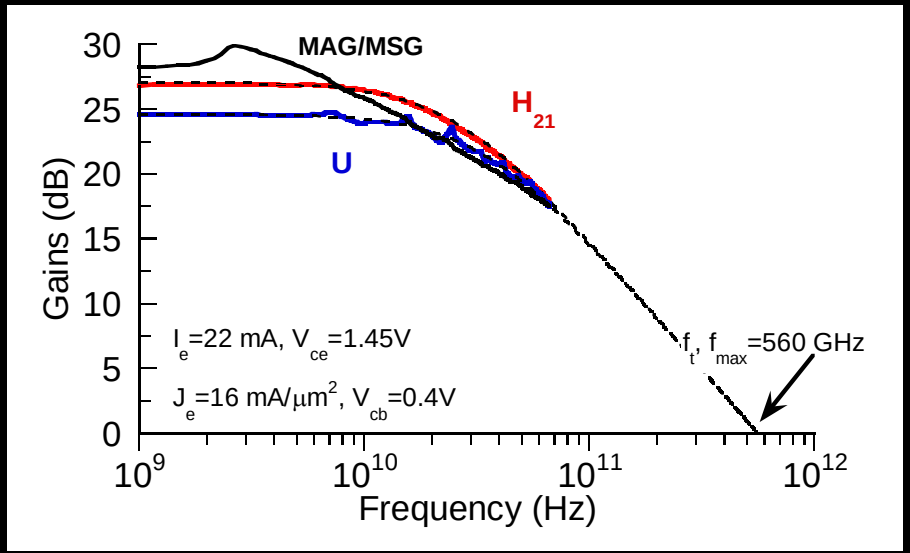
New Emitter Process



Scalable below 128 nm width

First results

are at 250 nm emitter width



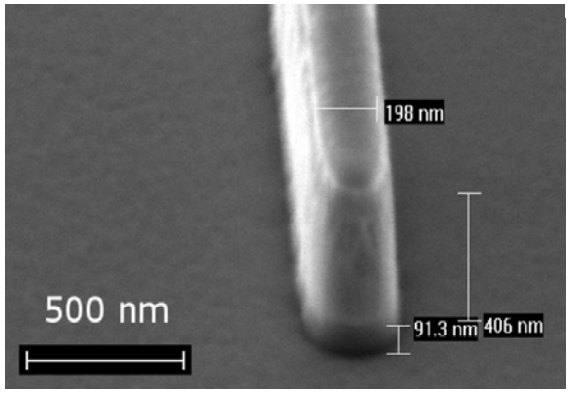
Simultaneously 560 GHz f_t & f_{max}

BVceo = 3.3V

...can do much better...

128 nm InP HBT: Technology Development

E. Lind



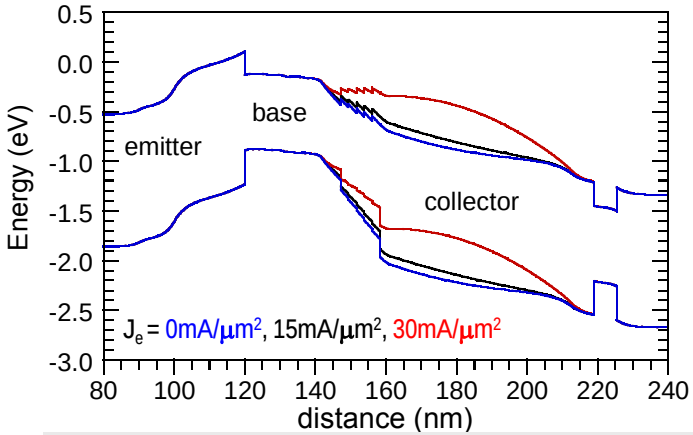
New Emitter Process for 128 and 64 nm junctions

dry etched metal

dry etched junction

refractory W or Mo contact→ stable at very high J_e

< 0.8 Ω-μm² contact resistivity



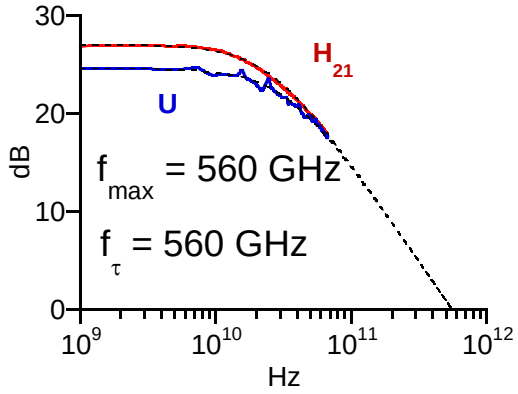
New, thin --12 nm -- base-collector grade:

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

U. Singisetti
A. Crook
S. Bank
E. Lind

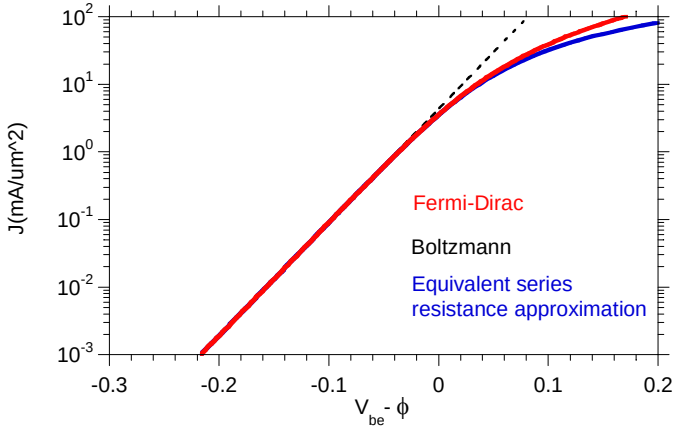
125 nm generation requires 5 Ω - μm² emitter resistivities

65 nm generation requires 1-2 Ω - μm²

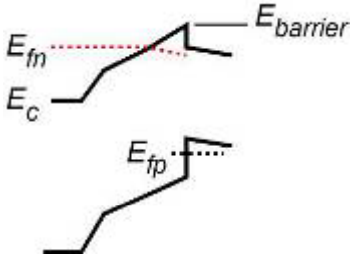
Recent Results (ONR contacts program)

ErAs/Mb	MBE in-situ	1.5 Ω - μm²
Mb	MBE in-situ	0.6 Ω - μm²
TiPdAu	ex-situ	0.5 Ω - μm²
TiW	ex-situ	0.7 Ω - μm²

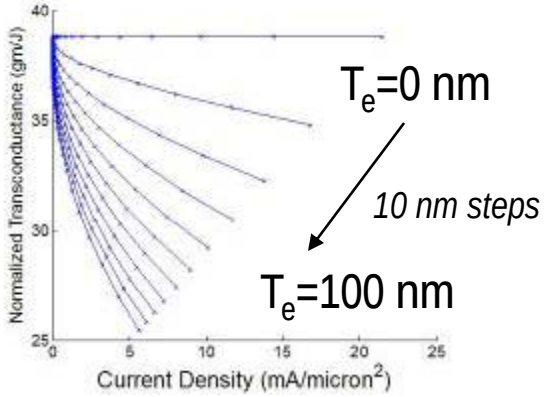
Degeneracy contributes 1 Ω - μm²



20 nm emitter-base depletion layer contributes 1 Ω - μm² resistance

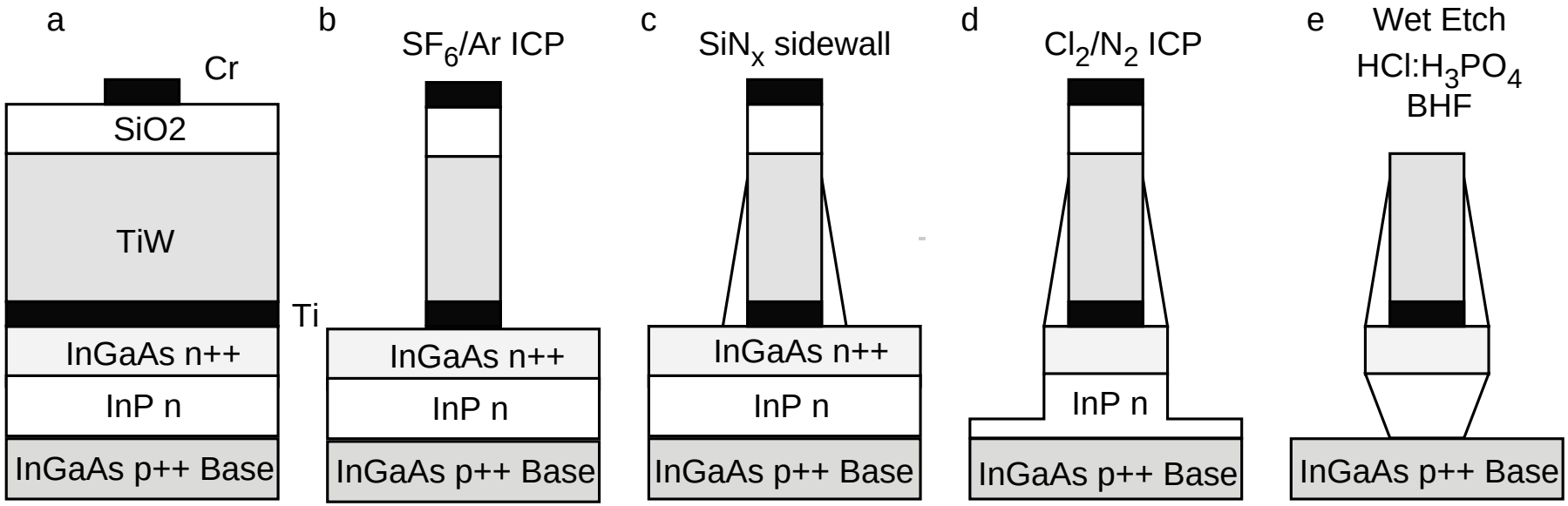


$$\frac{\partial E_{fn}(x)}{\partial x} = \frac{-J}{qn(x)}$$

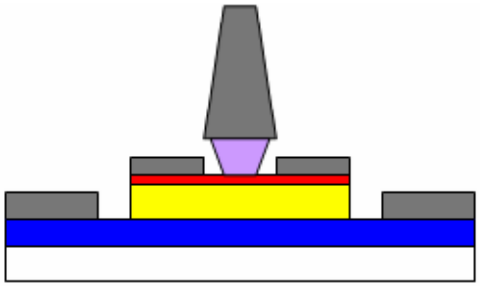


E. Lind

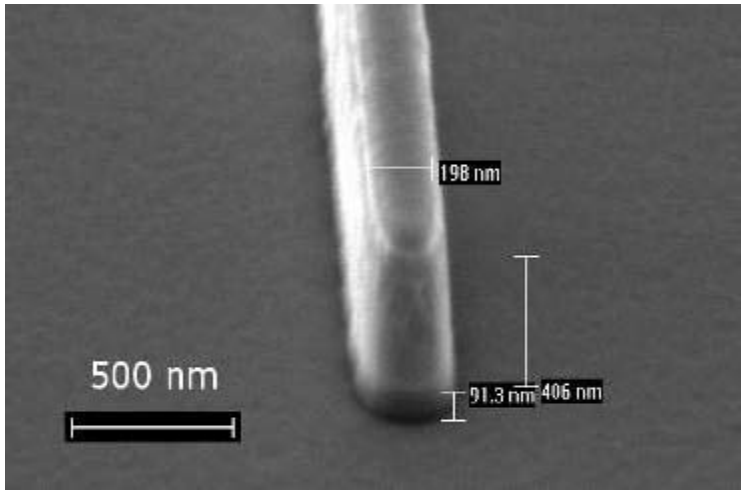
Current UCSB TiW emitter process



- 5 nm Ti layer for improved adhesion
- 25 nm SiN_x sidewalls protects Ti/TiW during Cl_2 and BHF etch, improves adhesion
- Standard triple mesa
- BCB passivation

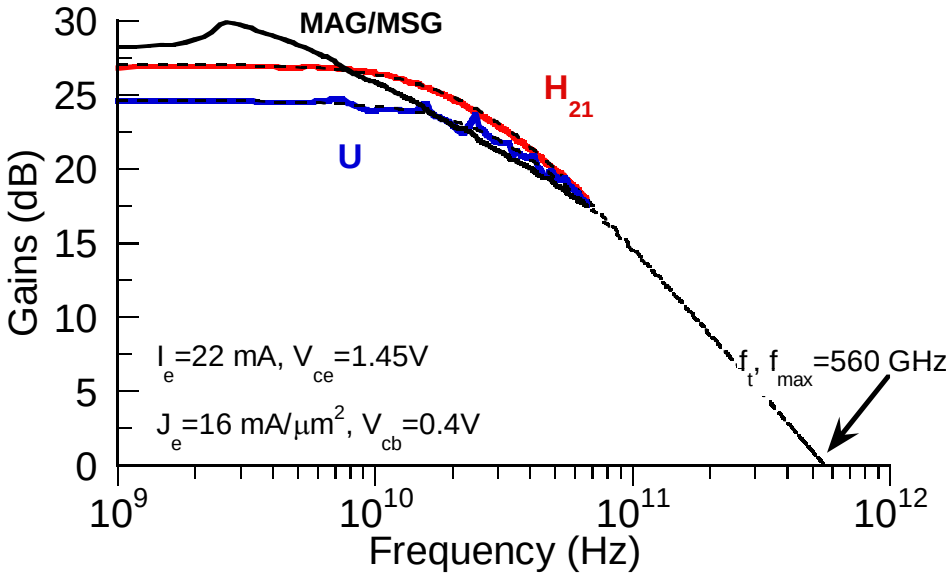


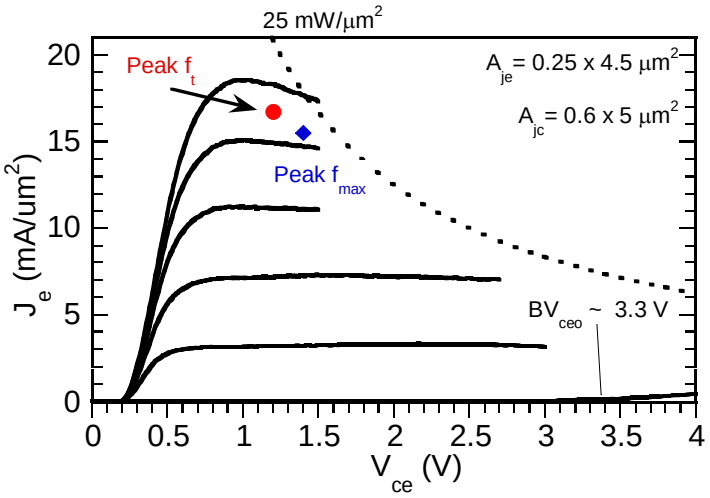
Emitter prior to InP wet etch

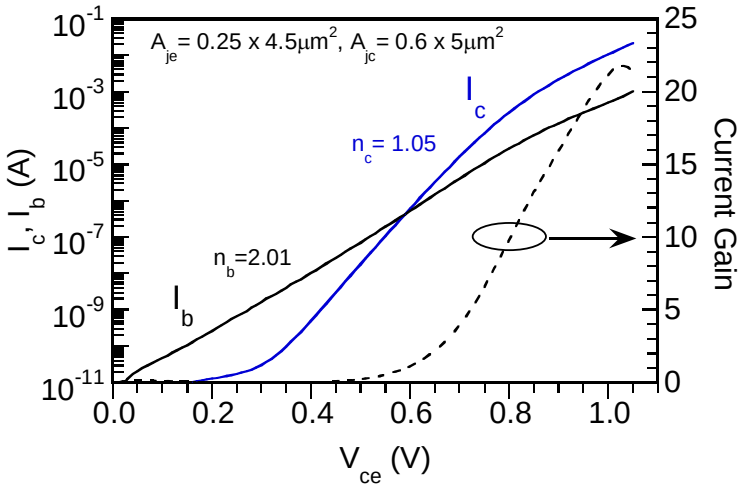


RF & DC data –70 nm collector, 22 nm base InP Type-I DHBT

E. Lind







Emitter width ~ 250 nm

First reported device with f_t, f_{max} > 500 GHz

BV_{CEO} ~ 3.3 V, BV_{CBO} = 3.9 V ($J_{e,c}$ = 15kA/cm^2)

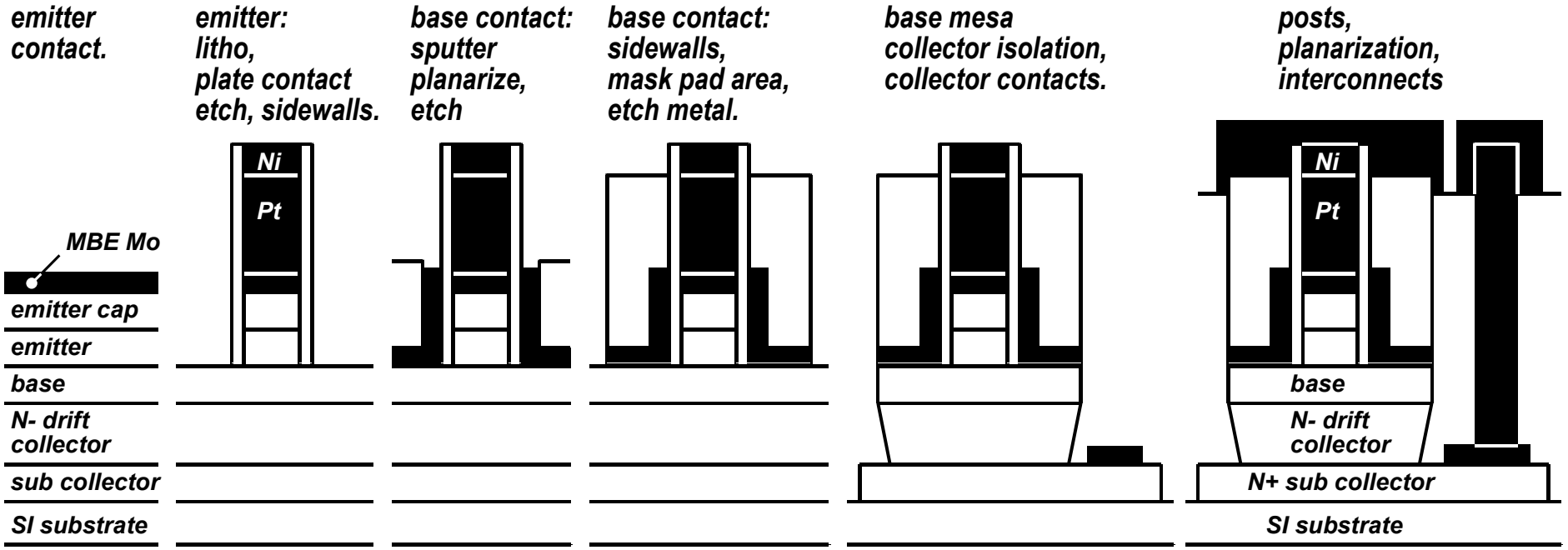
Emitter contact (from RF extraction), R_{cont} < 5 Ω·μm^2

Base: R_{sheet} = 780 Ω/sq, R_{cont} ~ 15 Ω·μm^2

Collector: R_{sheet} = 11.1 Ω/sq, R_{cont} ~ 10.1 Ω·μm^2

64 nm & 32 nm (THz) InP HBT

64 & 32 nm Generations: Example Process Flow



One critical lithographic step, no critical alignments

Refractory base and emitter contacts

Base-emitter ledge for leakage current control

Reliability…

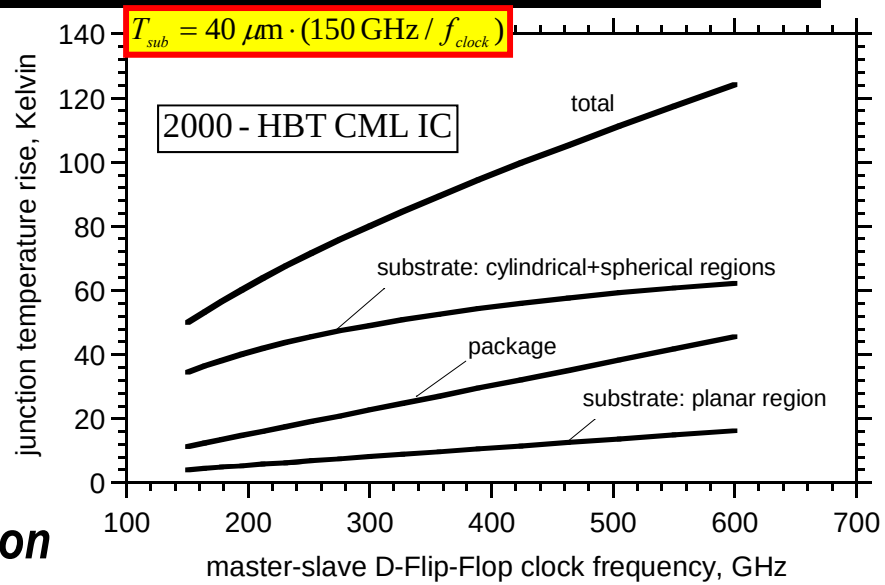
…depends upon stress
and upon device structure.

high current density→ heating
thermal design is critical

high current density→ contact electromigration
need refractory (W, TiW, Mo,…) contacts

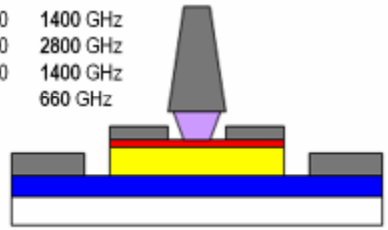
must investigate failure mechanisms
driven by high current density in semiconductor
dark-line defects

Given the need for large I/C charging rates, high current density is unavoidable.



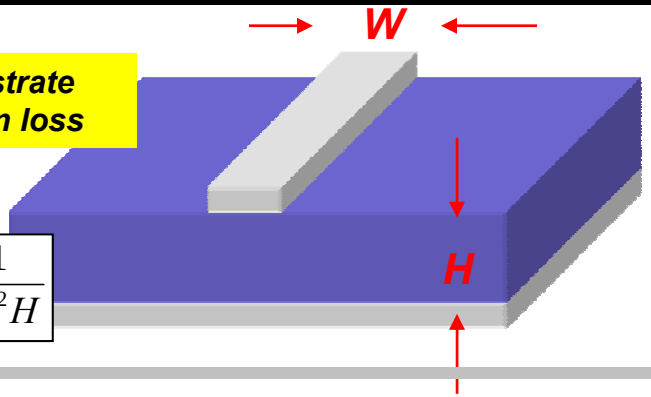
multi-THz InP HBT Scaling Roadmap

emitter	512	256	128	64	**32** nm width
	16	8	4	2	**1** Ω·μm² access ρ
base	300	175	120	60	**30** nm width,
	20	10	5	2.5	**1.25** Ω·μm² contact ρ
collector	150	106	75	53	**37.5** nm thick,
	4.5	9	18	36	**72** mA/μm² current density
	4.9	4	3.3	2.75	**2-2.5** V, breakdown
f_t	370	520	730	1000	**1400** GHz
f_{max}	490	850	1300	2000	**2800** GHz
power amplifiers	245	430	660	1000	**1400** GHz
digital clock rate (static dividers)	150	240	330	480	**660** GHz



Interconnects: Substrate Microstrip Has Problems

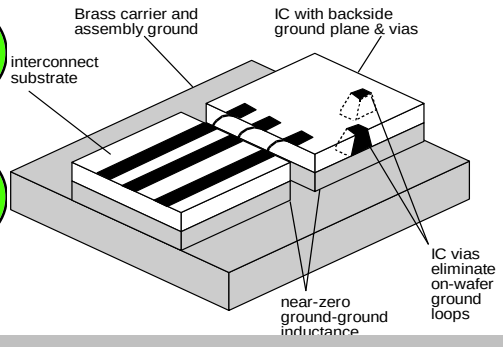
Thick Substrate → low skin loss



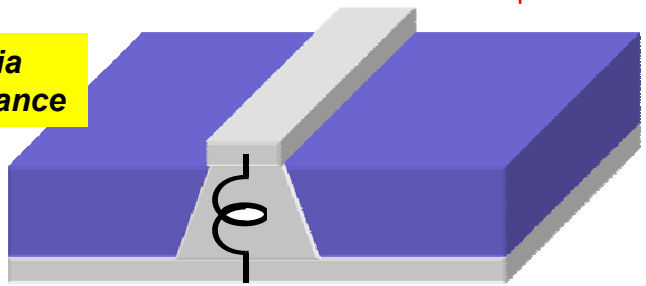
$$\alpha_{skin} \propto \frac{1}{\varepsilon_r^{1/2} H}$$

Zero ground inductance in package

No ground plane breaks in IC

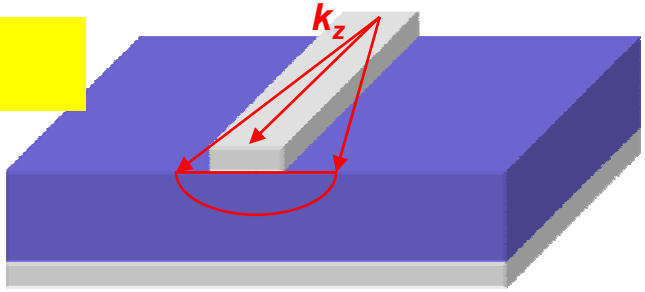


High via inductance



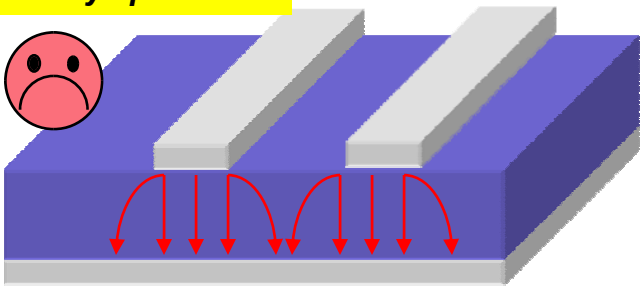
12 pH for 100 μm substrate -- 7.5 Ω @ 100 GHz

TM substrate mode coupling



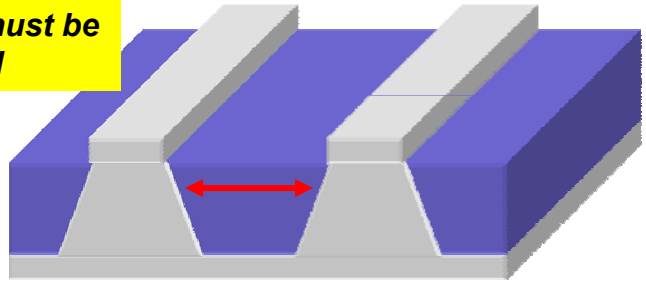
Strong coupling when substrate approaches ~λ_d / 4 thickness

lines must be widely spaced



Line spacings must be ~3(substrate thickness)*

ground vias must be widely spaced

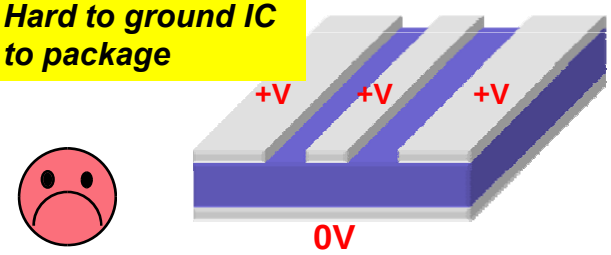


all factors require very thin substrates for >100 GHz ICs
→ lapping to ~50 μm substrate thickness typical for 100+ GHz

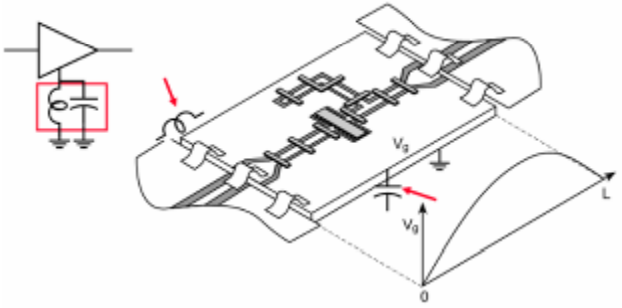
Interconnect: Coplanar Waveguide Has Many Problems !

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

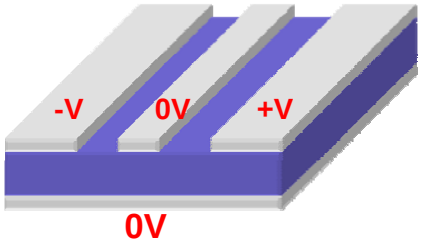
Hard to ground IC to package



Parasitic microstrip mode

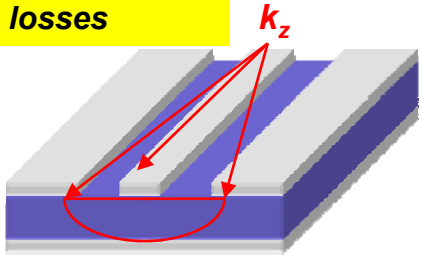


ground plane breaks → loss of ground integrity



Parasitic slot mode

substrate mode coupling or substrate losses

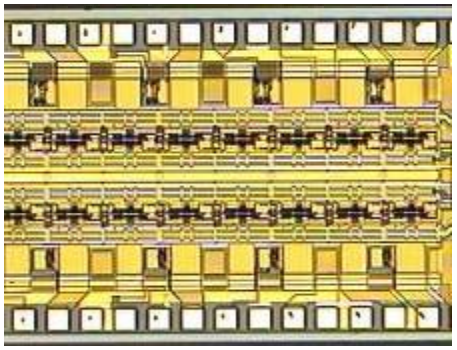


→ substrate *must* be thinned !

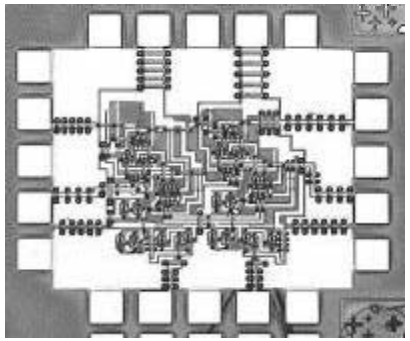
III-V:
semi-insulating substrate→ substrate mode coupling → must thin wafer to $\lambda_d/2$, must have vias to kill microstrip mode.

Silicon conducting substrate → substrate conductivity losses

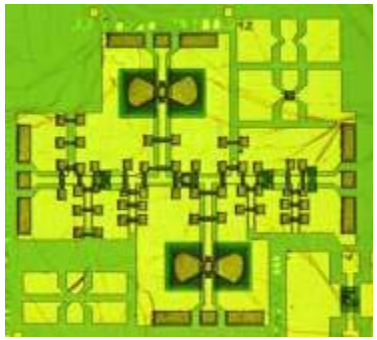
Repairing ground plane with ground straps is effective only in simple ICs
In more complex CPW ICs, ground plane rapidly vanishes
→ common-lead inductance → strong circuit-circuit coupling



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

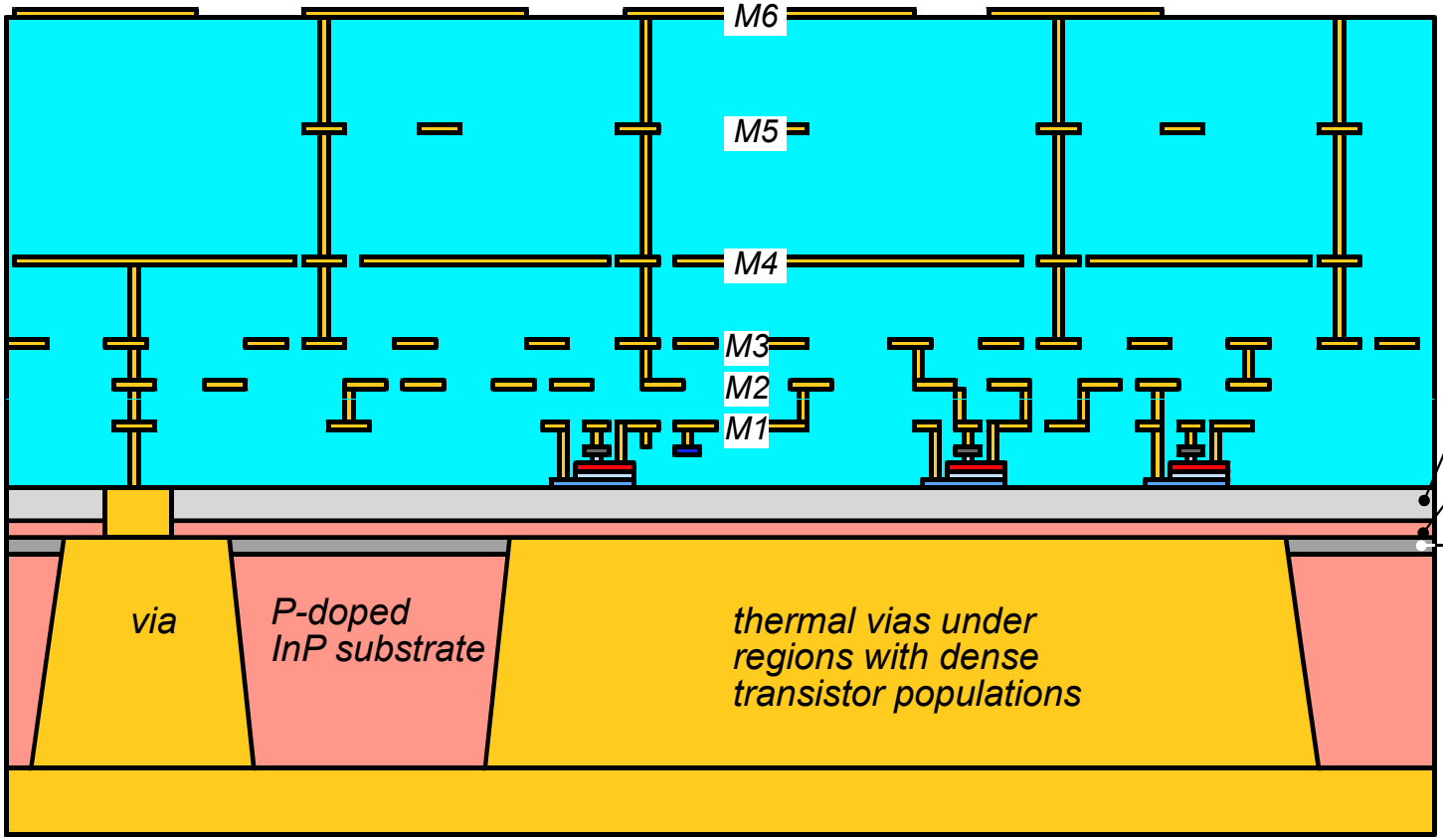
poor ground integrity

loss of impedance control

ground bounce

coupling, EMI, oscillation

THz IC Interconnects will be Silicon-Like



metal 6:
microstrip lines

metal 5:
microstrip crossovers

metal 4: ground plane
for M5 & M6 microstrips

undoped 1 um
InP spacer

P-InP

P-InGaAs etch stop

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

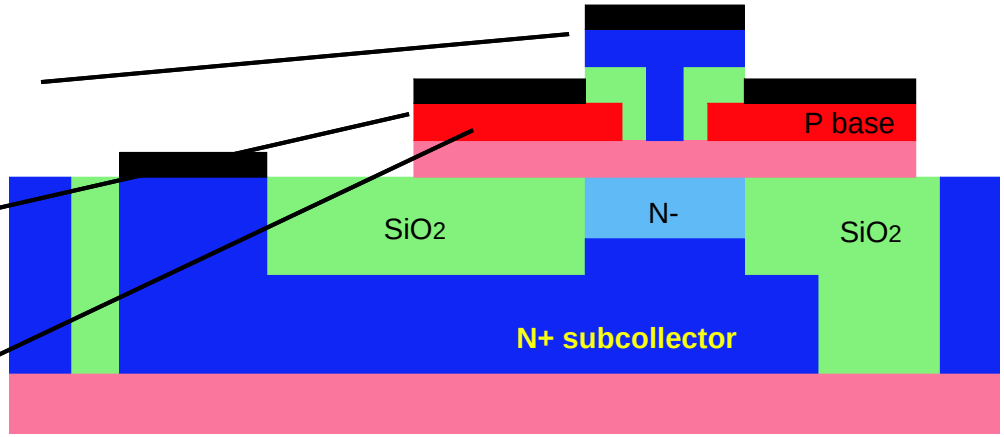
Comparison to SiGe

SiGe Today : Parasitic Reduction for Increased Bandwidth

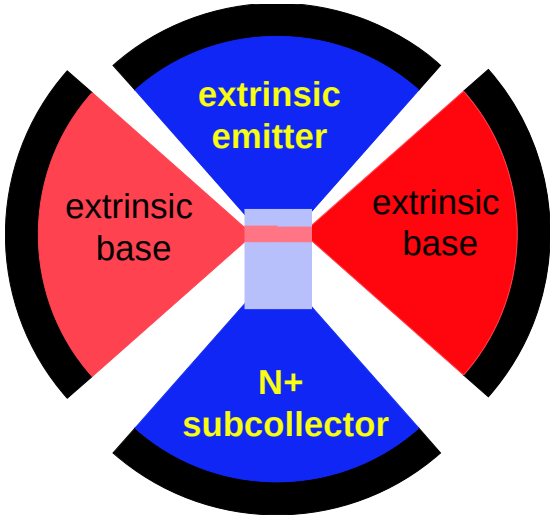
wide emitter contact: low resistance
narrow emitter junction: scaling (low R_{bb}/A_e)

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:



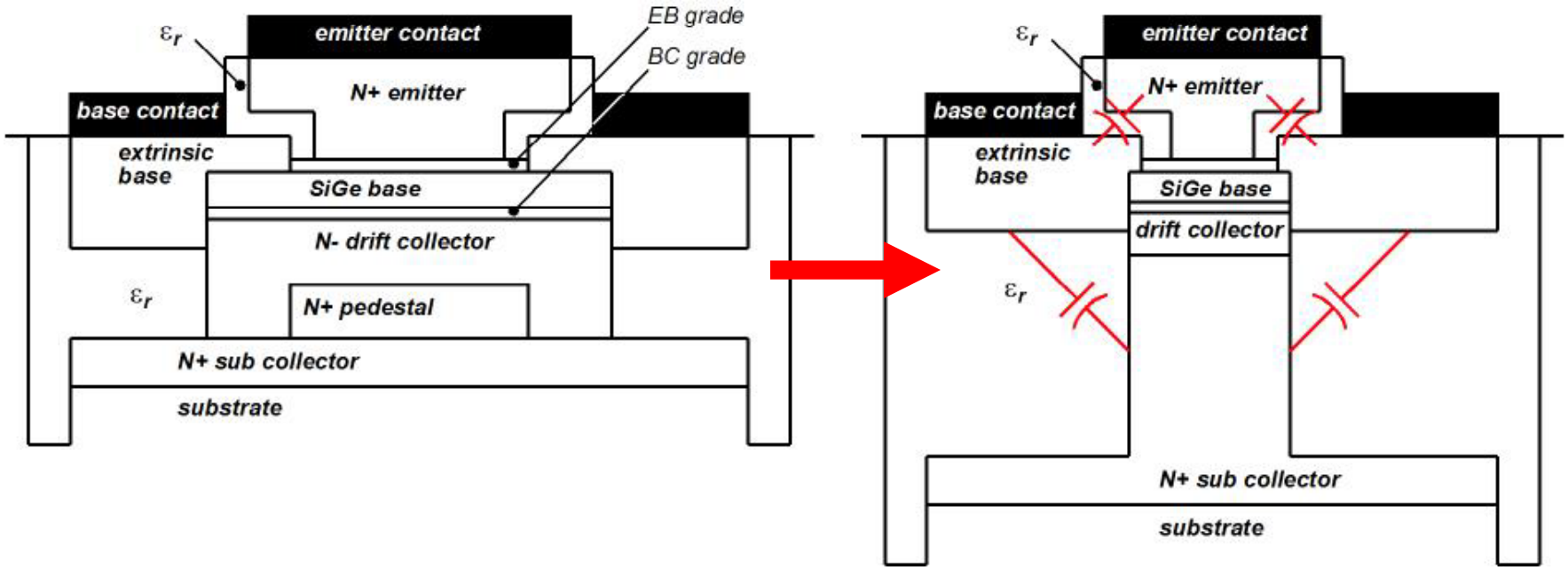
$$R_{bulk} = \frac{2\rho_{bulk}}{\pi L}\ln\left(\frac{\sqrt{2}\cdot r}{W}\right)$$

$$R_{contact} = \frac{2\rho_c}{\pi L r}$$

$$R_{\text{total,min}} = \frac{2\rho_{bulk}}{\pi L}\left[1.34 + \ln\left(\frac{\rho_{contact}}{W\rho_{bulk}}\right)\right]$$

→ ***reduced access resistance***

Parasitic Reduction Could Help Less with Small Devices

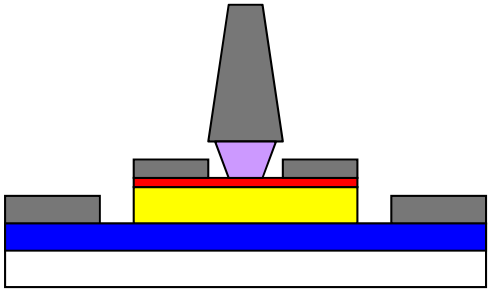


Scaling for increased HBT bandwidth:

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

→ 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**

$$\tau_c = T_c/2v \qquad \frac{C_{cb}}{I_c} \sim \frac{A_c}{A_e}\frac{\tau_c/2}{(V_{cb,app}+\phi)}$$

v increased 3.5 : 1 → increase collector thickness 3.5 : 1

breakdown increased in proportion to $(E_{max}v)$

$$J_{Kirk} = \frac{2\varepsilon v(V_{cb,\text{applied}} + V_{cb,\text{depletion}} + 2\phi)}{T_c^2}$$

J decreases 3.5 : 1 → junction areas increase 3.5 : 1

$$\Delta T \cong \frac{P}{\pi K_{Th} L_E}\ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{Th} L_E}$$

For equal ΔT, 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 ~ 3.5 : 1 larger

$$R_{ex} = \rho_c/A_e$$

$$R_{bb} \cong \frac{\rho_s W_e}{12L_e} + \frac{\rho_s W_{bc}}{6L_e} + \frac{\rho_c}{A_{contacts}}$$

Required contact resistivities are similarly relaxed

→ about 3.5:1 larger dimensions, 3.5:1 larger breakdown for a given device bandwidth

→ 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_τ, 1.5 THz f_{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 → harder to project further progress

Contact + access resistivies & thermal resistivities are key scaling limits

(end)

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

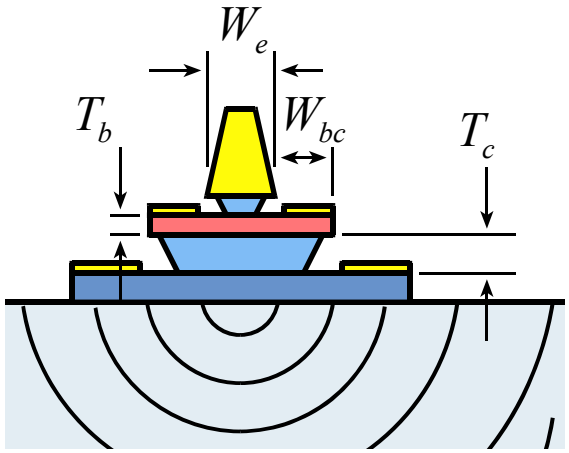
$\tau_b = T_b^2 / 2D_n + T_b / v$ → *thin base ~1.414:1*

$\tau_c = T_c / 2v$ → *thin collector 2:1*

$C_{cb} \propto A_c / T_c$ → *reduce junction areas 4:1*

$R_{ex} = \rho_c / A_e$ → *reduce emitter contact resistivity 4:1*

$I_{c,Kirk} \propto A_e / T_c^2$ *(current remains constant, as desired)*



(emitter length L_E)

$$\Delta T \cong \frac{P}{\pi K_{InP} L_E} \ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{InP} L_E}$$

need to reduce junction areas 4:1

reduce widths 2:1 & reduce length 2:1 → doubles ΔT ✗

reducing widths 4:1, keep constant length→ small ΔT increase ✓

$$R_{bb} \cong \underbrace{\frac{\rho_s W_e}{12 L_e} + \frac{\rho_s W_{bc}}{6 L_e}} + \frac{\rho_c}{A_{contacts}}$$

→ *reduce base contact resistivity 4:1*

reduce widths 2:1 & reduce length 2:1 → constant R_{bb} ✓

reducing widths 4:1, keep constant length → reduced R_{bb} ✓✓

Linewidths scale as the inverse square of bandwidth because thermal constraints dominate.

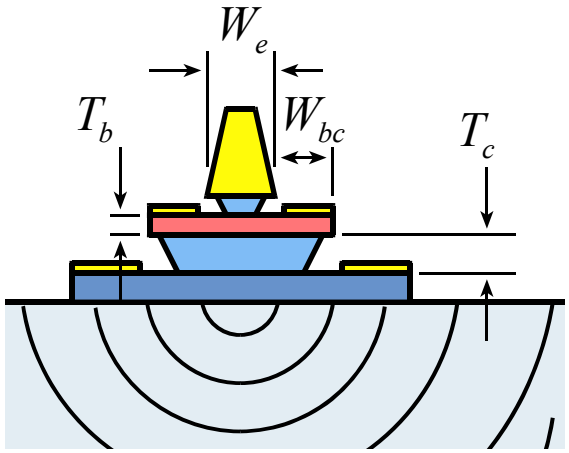
First-Order HBT Design

$$\tau_b = T_b^2 / 2D_n + T_b / v$$

$$\tau_c = T_c / 2v$$

$$R_{ex} = \rho_c / A_e$$

$$R_{bb} \cong \frac{\rho_s W_e}{12 L_e} + \frac{\rho_s W_{bc}}{6 L_e} + \frac{\rho_c}{A_{contacts}}$$



(emitter length L_E)

$$C_{cb} = \varepsilon A_c / T_c$$

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

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

$$\Delta T \cong \frac{P}{\pi K_{InP} L_E} \ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{InP} L_E}$$