On the Feasibility of low-THz InP HBTs

Mark Rodwell, Z. Griffith, E. Lind, U. Singisetti,*
M. Wistey, A.C. Gossard
University of California, Santa Barbara

i*Now at Lund University

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

Collaborators (III-V MOS)

A. Gossard, S. Stemmer, C. Van de Walle
University of California Santa Barbara

P. Asbeck, A. Kummel, Y. Taur,
University of California San Diego

J. Harris, P. McIntyre,
Stanford University

C. Palmstrøm,
University of Minnesota

M. Fischetti
University of Massachusetts Amherst

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

Specific Acknowledgements

(Prof.) Erik Lind

125 nm HBTs
process
theory / epi design

Dr. Zach Griffith

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

Dr. Mark Wistey

InGaAs MOSFET
process technology
theory / epi design

TeraHertz and nanoMeter Electron Devices

How do we make very fast electron devices ?

...by scaling

What are the limits to scaling ?

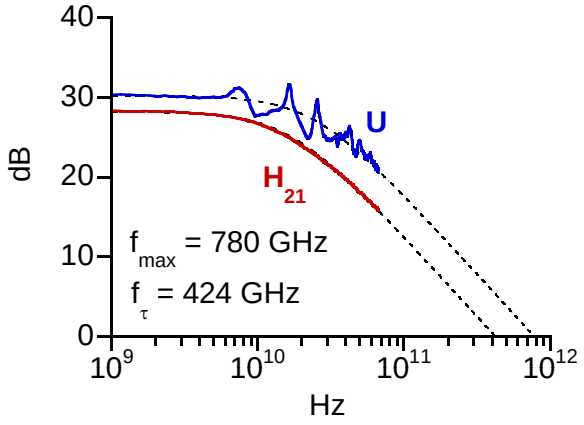
attainable contact resistivities,
attainable thermal resistivities
attainable contact stabilities
and for FETs, attainable capacitance densities

How do we make long-term progress ?

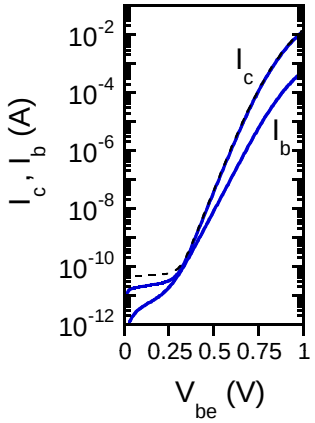
work on interfaces (contacts and gate dielectrics) !

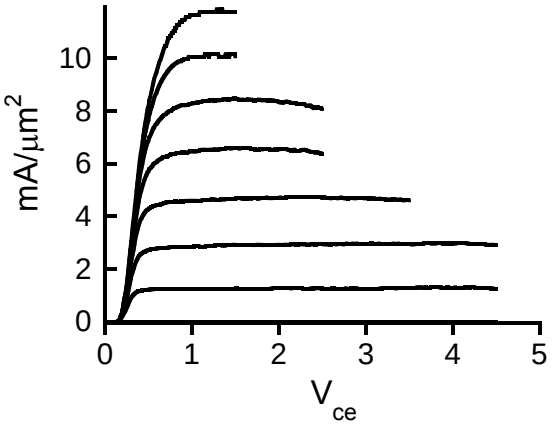
THz InP Transistors: Opportunities

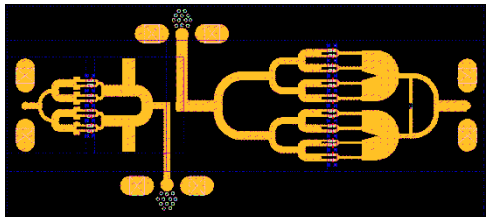
InP HBT: THz bandwidths, good breakdown, analog precision



&



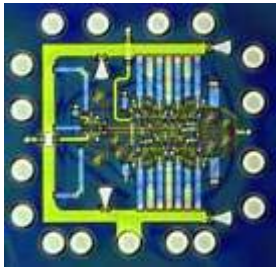




M. Jones

340 GHz, 70 mW amplifiers ***(design)***

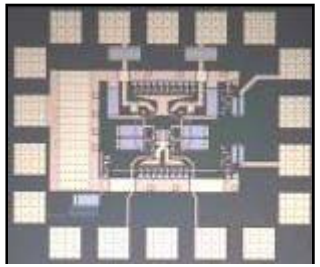
In future: 700 or 1000 GHz amplifiers ?



Z. Griffith

200 GHz digital logic ***(design)***

In future: 450 GHz clock rate ?



M. Urteaga (Teledyne)

Z. Griffith

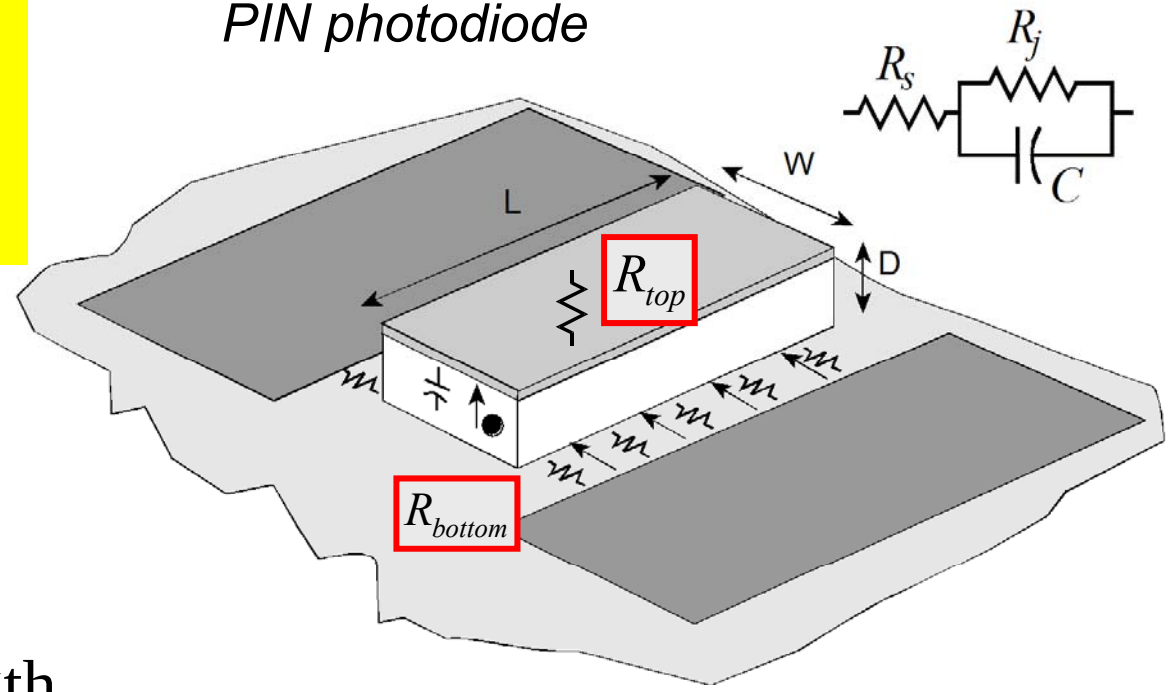
30-50 GHz gain-bandwidth op-amps⟶ low IM3 @ 2 GHz ***In future: 200 GHz op-amps for low-IM3 10 GHz amplifiers?***

We will make THz transistors ...

... by scaling

Frequency Limits and Scaling Laws of (most) Electron Devices

PIN photodiode



$$\tau \propto \text{thickness}$$

$$C \propto \text{area} / \text{thickness}$$

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

$$R_{bottom} \propto 1/\text{stripe length}$$

$$I_{\text{max, space-charge-limit}} \propto \text{area} / (\text{thickness})^2$$

$$\Delta T \propto \frac{\text{power}}{\text{length}} \times \log\left(\frac{\text{length}}{\text{width}}\right)$$

resistance *capacitance* *transit time*

↓ ↓ ↓

device bandwidth

To double bandwidth,

reduce thicknesses 2:1

reduce width 4:1, keep constant length

current density has increased 4:1

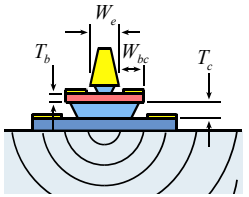
InP Bipolar Transistors

Bipolar Transistor Scaling Laws

Bipolar Transistor Design

$$\tau_b \approx T_b^2 / 2D_n$$

$$\tau_c = T_c / 2v_{sat}$$



(emitter length L_E)

$$R_{ex} = \rho_{\text{contact}} / A_e$$

$$R_{bb} = \rho_{\text{sheet}} \left(\frac{W_e}{12L_e} + \frac{W_{bc}}{6L_e} \right) + \frac{\rho_{\text{contact}}}{A_{\text{contacts}}}$$

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

$$I_{c,Kirk} \propto v_{sat} A_e (V_{\text{ce,operating}} + V_{\text{ce,punch-through}}) / T_c^2$$

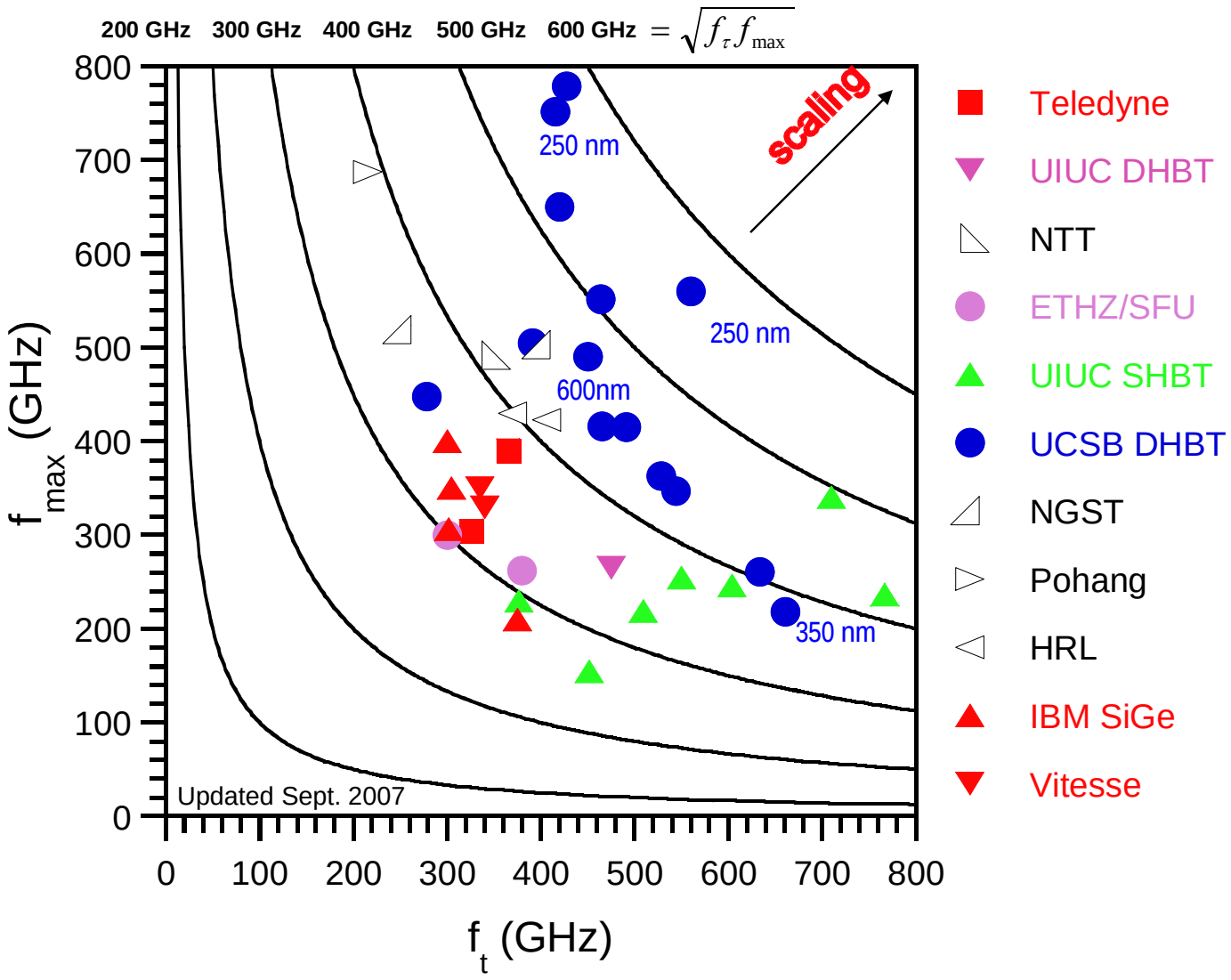
$$\Delta T \propto \frac{P}{L_E} \left[1 + \ln \left(\frac{L_e}{W_e} \right) \right]$$

Changes required to double transistor bandwidth:

parameter	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 contact resistance	decrease 4:1
current density	increase 4:1
base contact resistivity	decrease 4:1

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

Status of Bipolar Transistors : September 2007



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 :

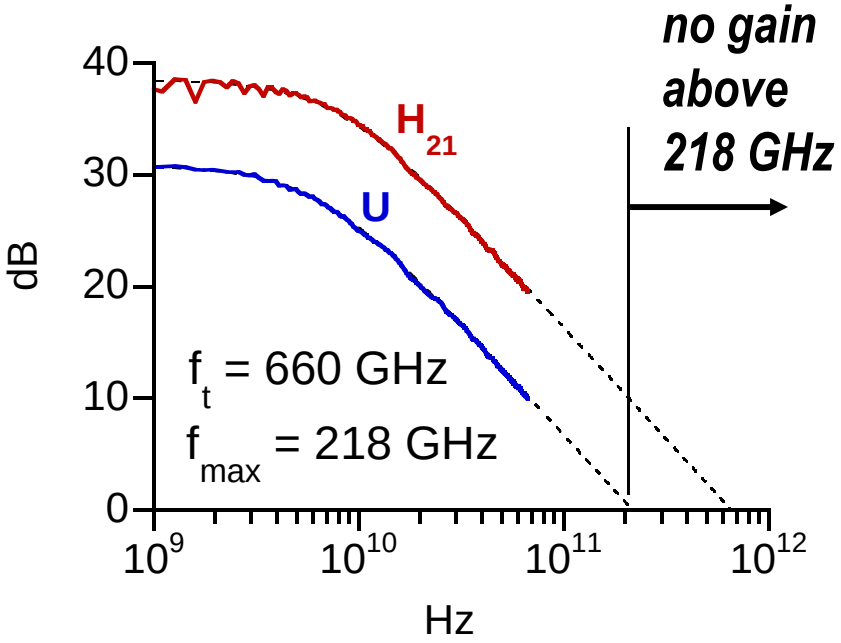
F_{min}, associated gain,

digital :

f_{clock}, hence $(C_{cb}\Delta V / I_c)$, $(R_{ex} I_c / \Delta V)$, $(R_{bb} I_c / \Delta V)$, $(\tau_b + \tau_c)$

What Matters Regarding Transistor Performance ?

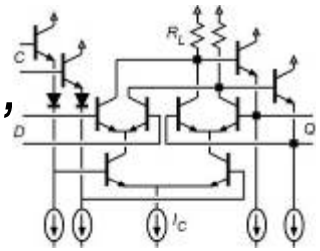
f_{max} matters



Tuned amplifiers: f_{max} sets bandwidth

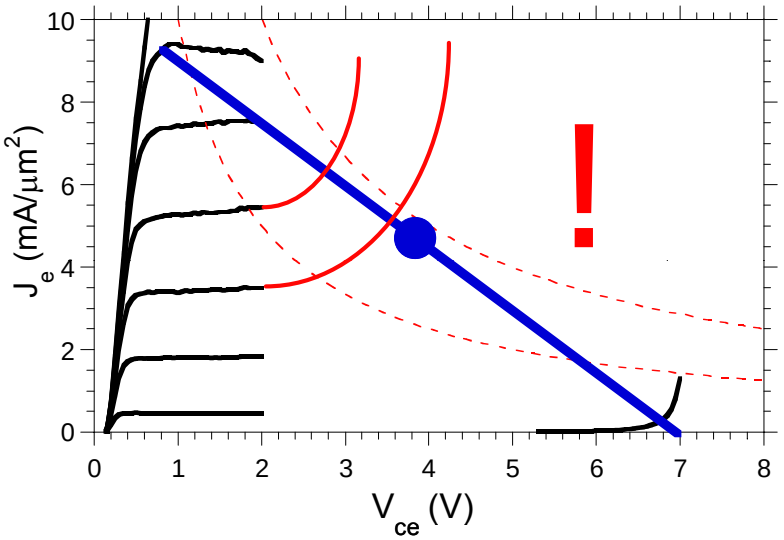
Mixed-signal:

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



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

breakdown is not the only voltage limit



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

thermal resistance,
high-current breakdown
reduced electron velocity at high voltages
high-temperature operation (~75 C) ?

Unilateral Power Gain

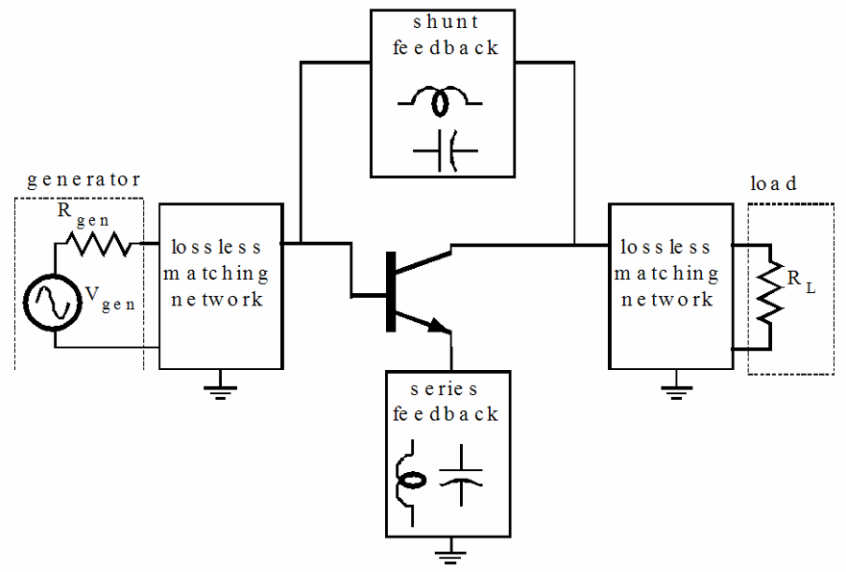
1) Cancel device feedback with external lossless feedback

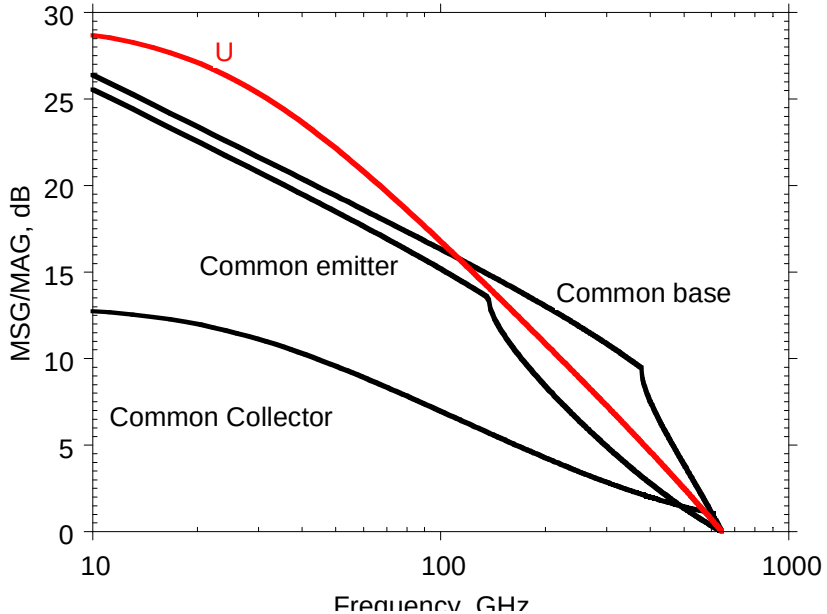
$$\rightarrow Y_{12} = S_{12} = 0$$

2) Match input and output

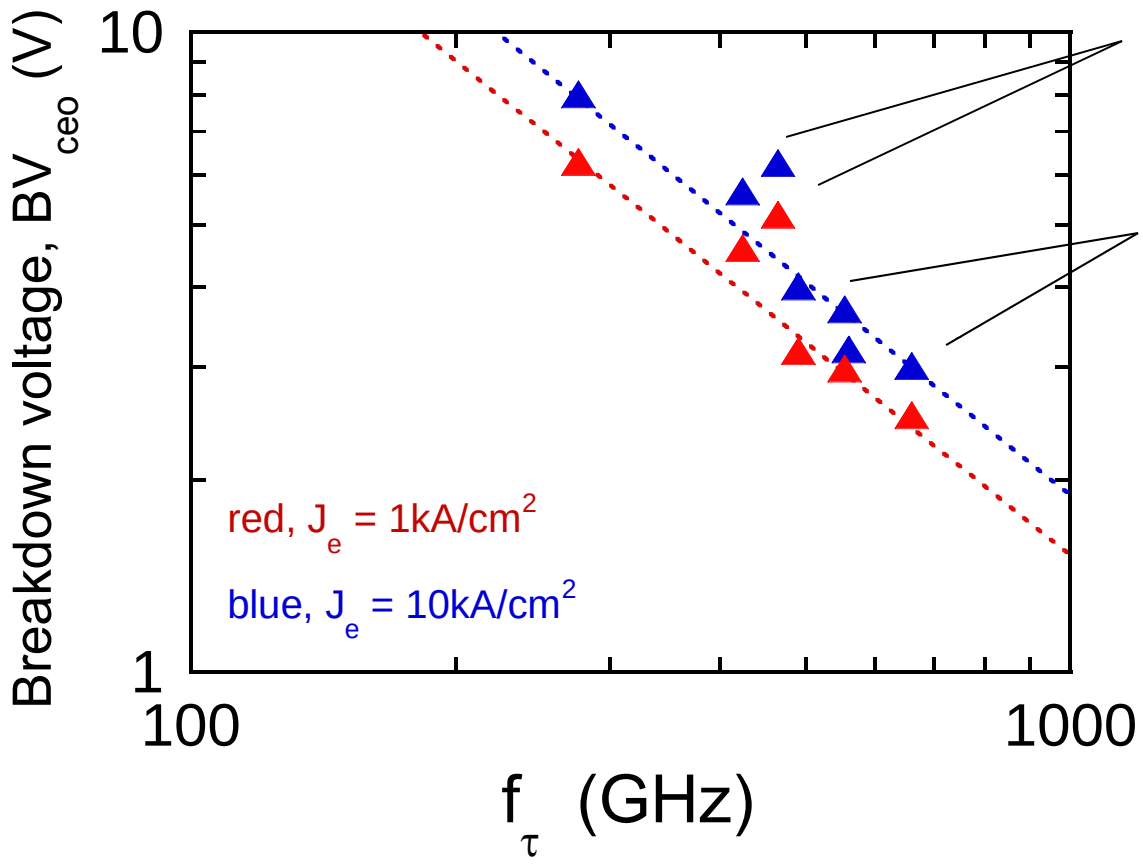
Resulting power gain is Mason's Unilateral Gain

$$U = \frac{\left|Y_{21} - Y_{12}\right|^2}{4\left(G_{11}G_{22} - G_{21}G_{12}\right)} \text{ NOT } \frac{\left|S_{21}\right|^2}{\left(1 - \left\|S_{11}\right\|^2\right)\left(1 - \left\|S_{22}\right\|^2\right)}$$





Breakdown Voltage Scaling: Expect 2.4 V @ 1 THz f_τ



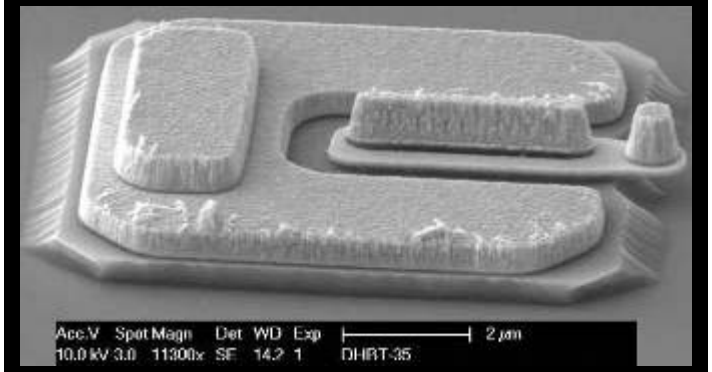
mature devices--
full potential f_τ achieved

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

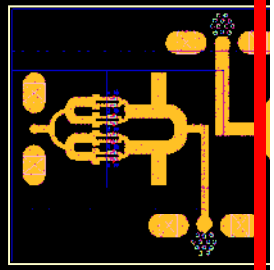
For mature, well-scaled InP DHBTs, f_τ x BVCEO = 2.4 THz-Volts.

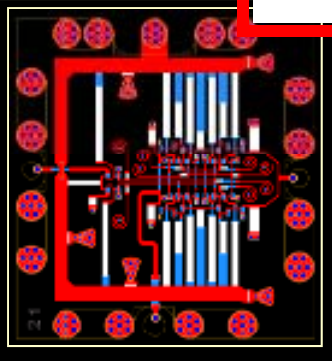
InP/InGaAs/InP & InP/GaAsSb/InP DHBTs have equal breakdown.

256 nm Generation InP DHBT

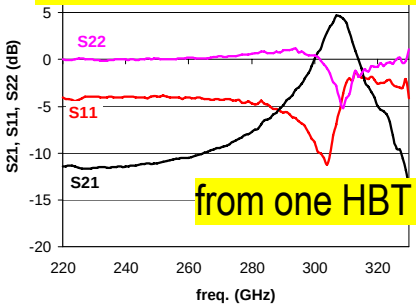


340 GHz, 70 mW





4.7 dB Gain at 306 GHz.

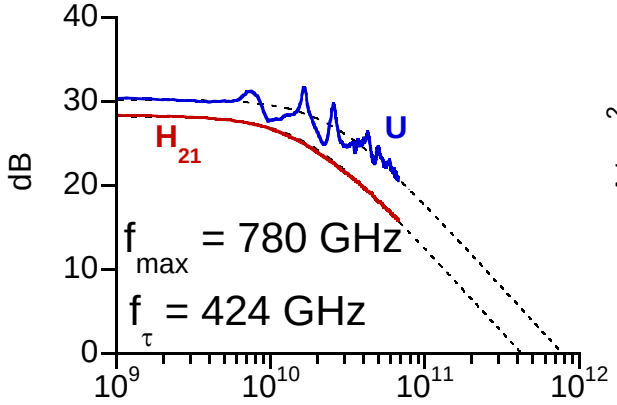


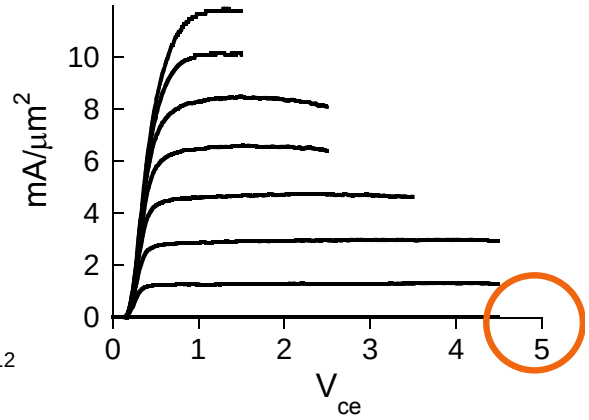
from one HBT

200 GHz master-slave latch design

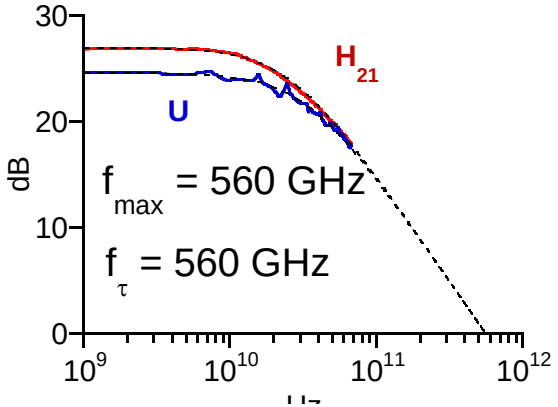
Z. Griffith, E. Lind, J. Hacker, M. Jones

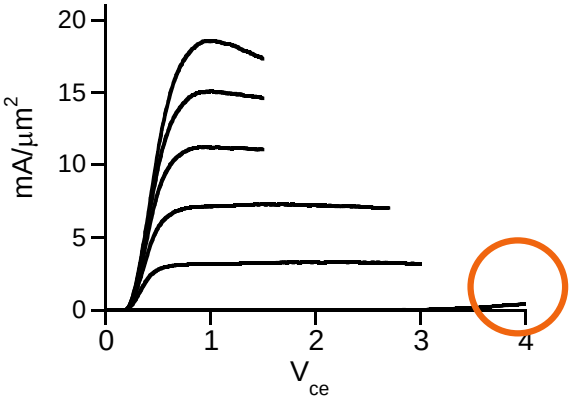
150 nm thick collector



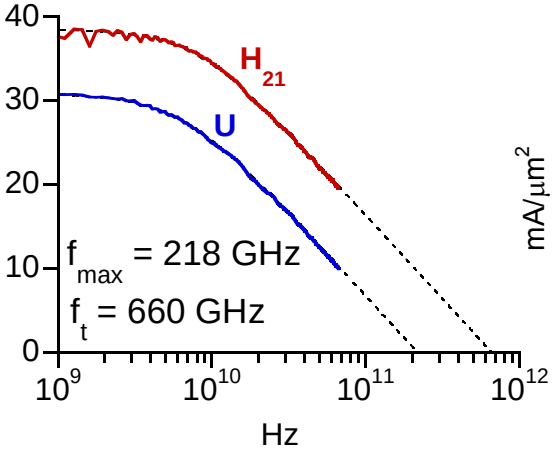


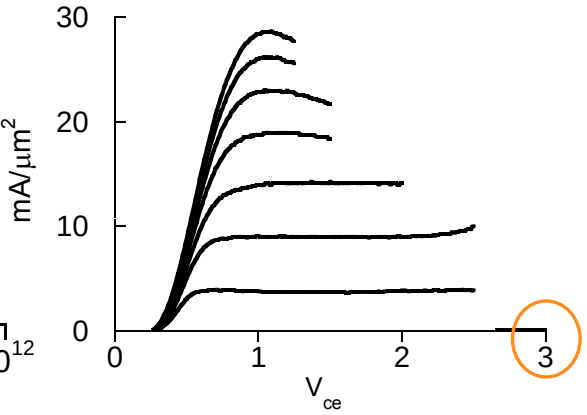
70 nm thick collector





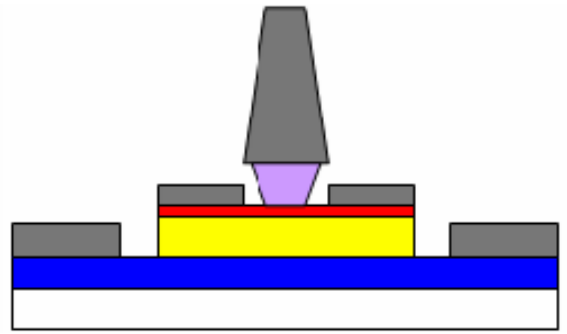
30 nm thick collector





InP Bipolar Transistor Scaling Roadmap

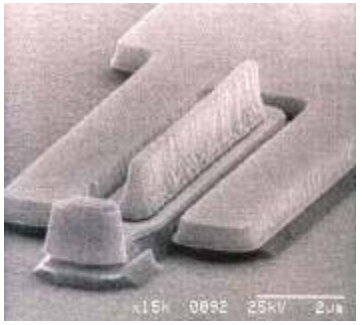
	industry	university →industry	university 2007-8	appears feasible	maybe
emitter	512	256	128	64	32 nm width
	16	8	4	2	1 Ω·μm² access ρ
base	300	175	120	60	30 nm contact width,
	20	10	5	2.5	1.25 Ω·μm² contact ρ
collector	150	106	75	53	37.5 nm thick,
	4.5	9	18	36	72 mA/μm² current density
	4.9	4	3.3	2.75	2-2.5 V, breakdown
f_τ	370	520	730	1000	1400 GHz
f_{max}	490	850	1300	2000	2800 GHz
power amplifiers	245	430	660	1000	1400 GHz
digital 2:1 divider	150	240	330	480	660 GHz



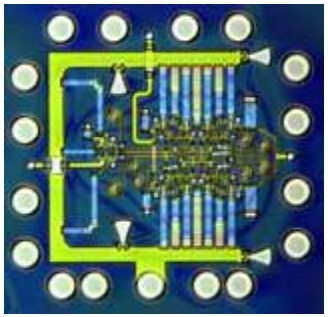
512 nm InP DHBT

Laboratory Technology

500 nm mesa HBT

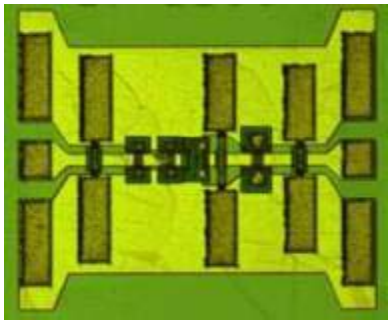


150 GHz M/S latches



UCSB / Teledyne / GCS

175 GHz amplifiers

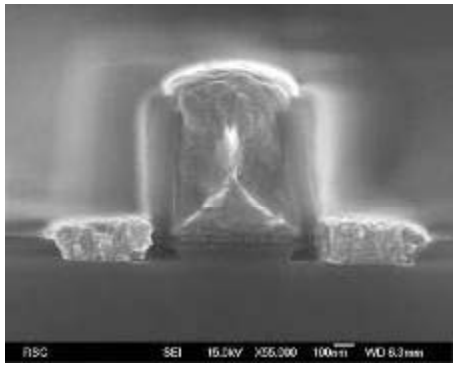


UCSB

Production

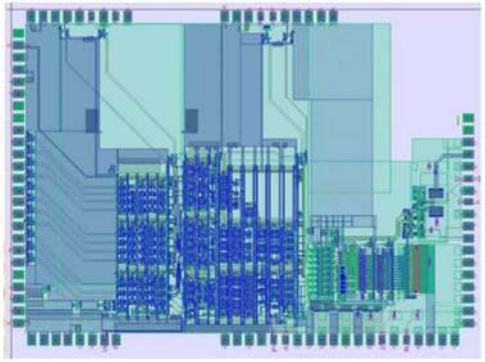
(Teledyne)

500 nm sidewall HBT



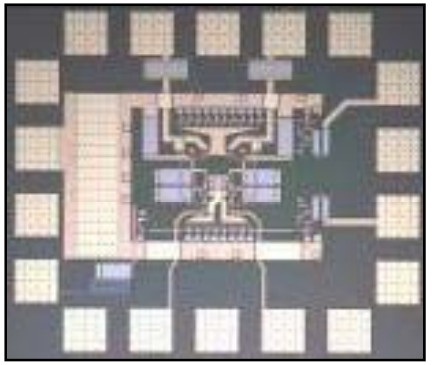
Teledyne

DDS IC: 4500 HBTs



Teledyne / BAE

40 GHz op-amps



Teledyne / UCSB

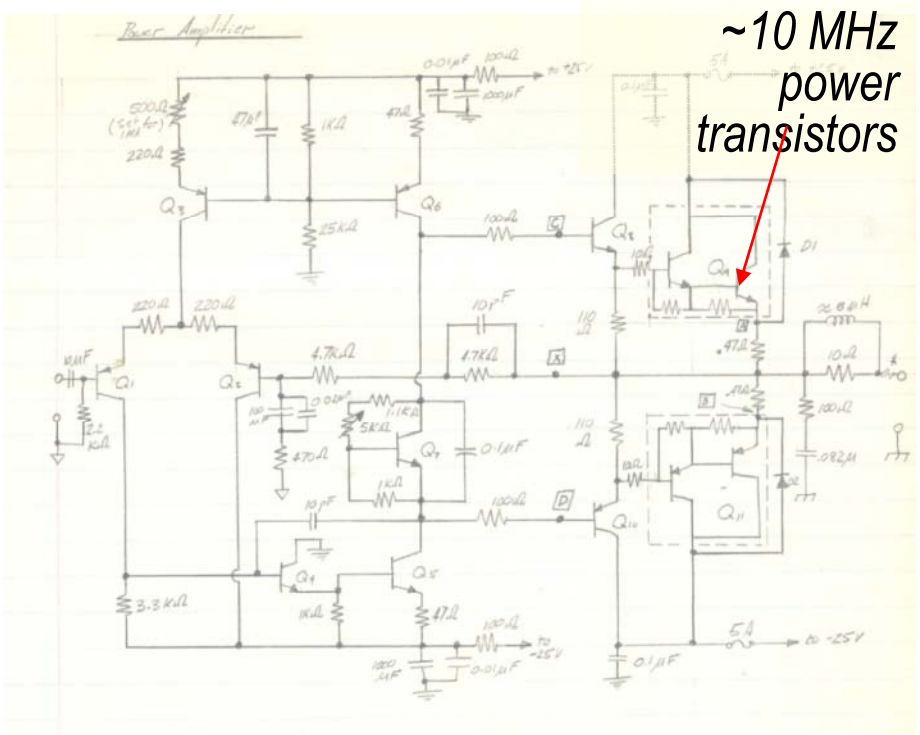
Z. Griffith
M. Urteaga
P. Rowell
D. Pierson
B. Brar
V. Paidi

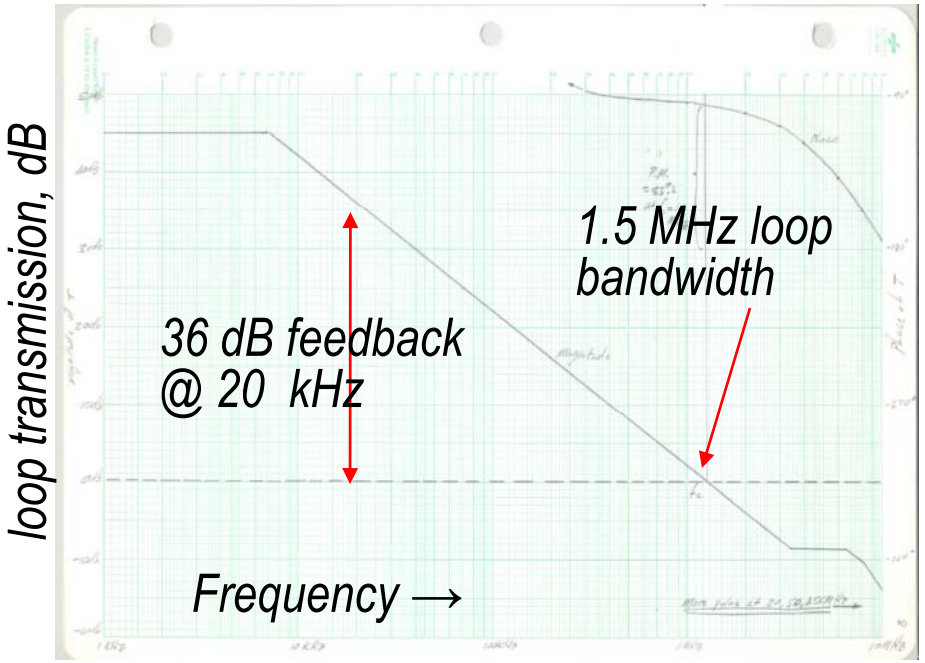
f_τ = 405 GHz
f_{max} = 392 GHz
$V_{br,\ ceo}$ = 4 V

20 GHz clock

50 dBm OIP3 @ 2 GHz with 1 W dissipation

Let's make Audio Power Amplifiers . . . in the GHz !





A 1980 hobby-project Audio power amp:
10 MHz transistors→ 1.5 MHz loop → 36 dB feedback @ 20 kHz → 0.02% distortion

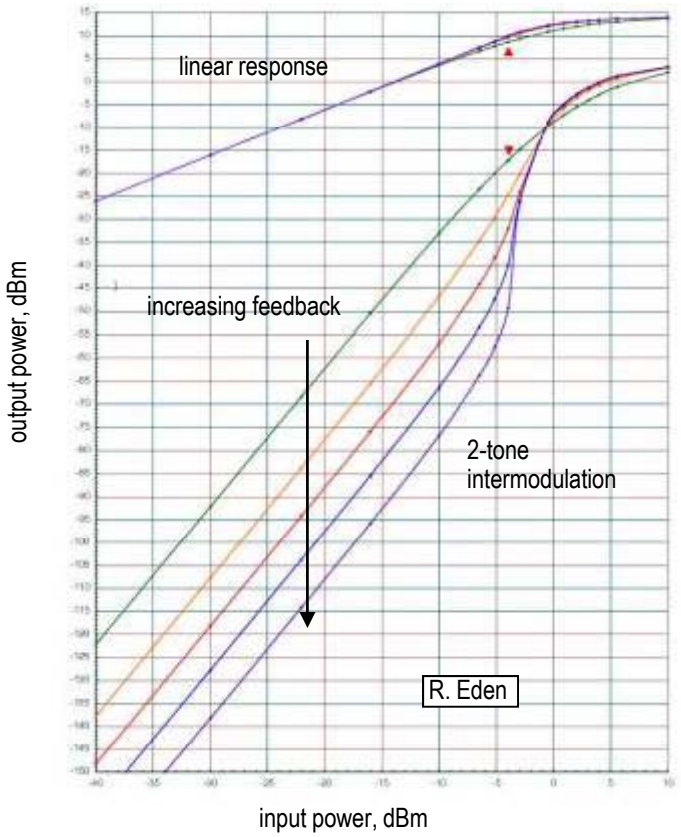
What if we used modern InP transistors ?
350 GHz transistors→ 50 GHz loop → 26 dB feedback @ 2.5 GHz
→ very low distortion for 2 GHz (cell phone band etc) amplifiers ?

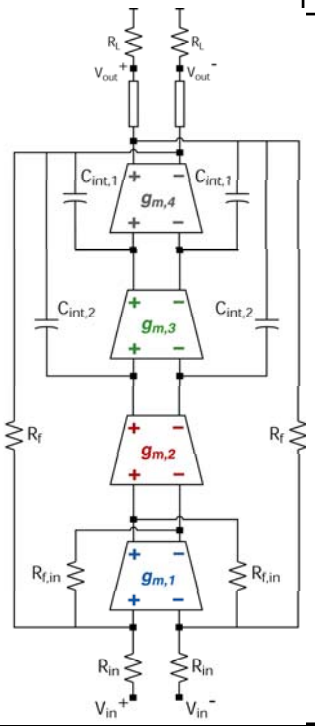
THz transistors → precision analog design at RF & microwave

mm-wave Op-Amps for Linear Microwave Amplification

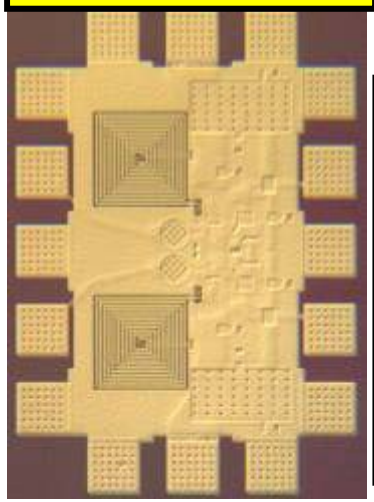
DARPA / UCSB / Teledyne FLARE: Griffith & Urteaga

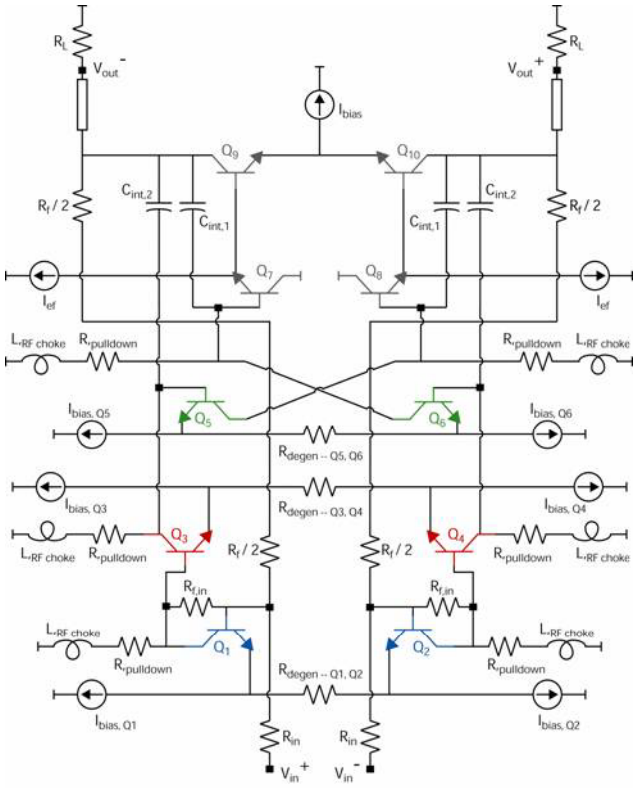
Reduce distortion with strong negative feedback

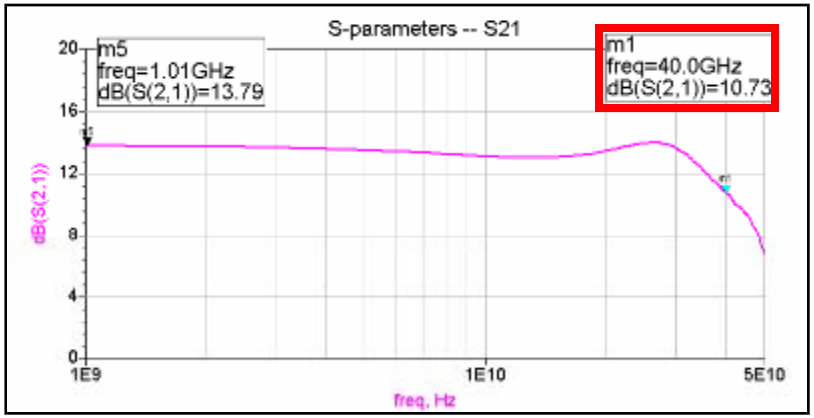




300 GHz / 4 V InP HBT







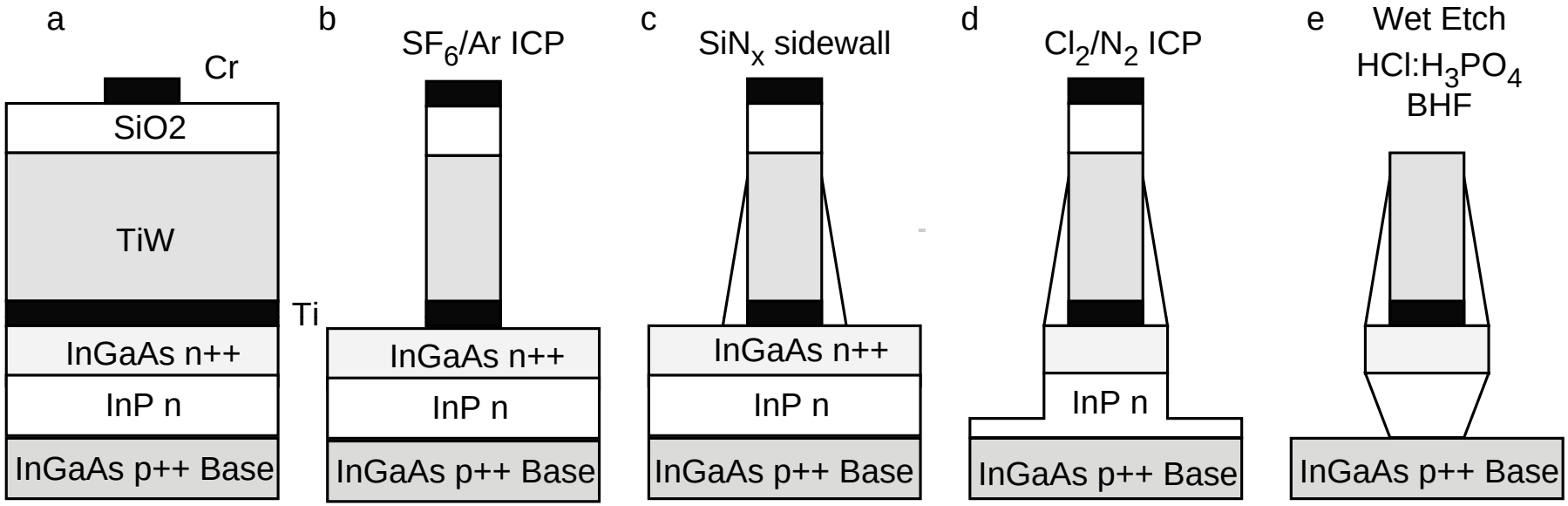
measured 40 GHz bandwidth
measured 50 dBm OIP3 @ 2 GHz

new designs in fabrication
simulated 56 dBm OIP3 @ 2 GHz

HBTs:
128 nm Generation

Sputter / Dry-Etch Emitter Process for 128 nm

E. Lind



Contact metal is sputtered & dry-etched

Contact metal is refractory

Emitter semiconductor is (mostly) dry-etched

Improvements in Emitter Access Resistance

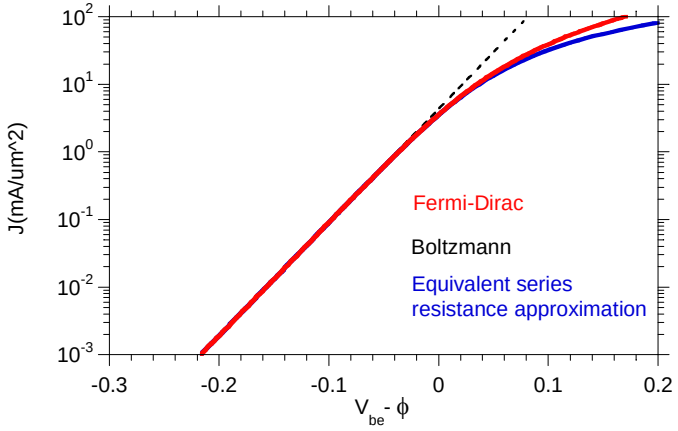
U. Singisetti
A. Crook
S. Bank
E. Lind

125 nm generation requires 5 Ω - μm² emitter resistivities

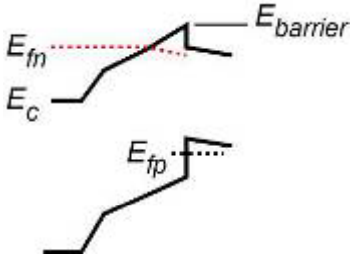
65 nm generation requires 1-2 Ω - μm²

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***

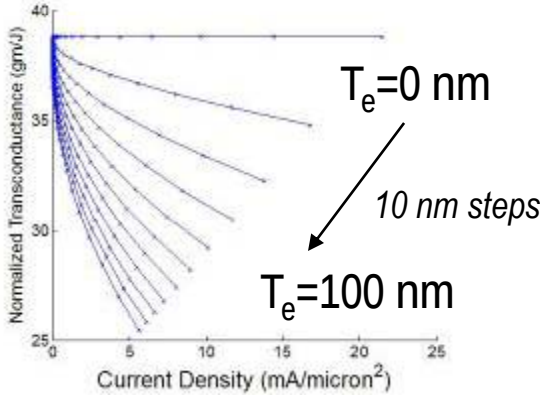
Degeneracy contributes 1 Ω - μm²



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

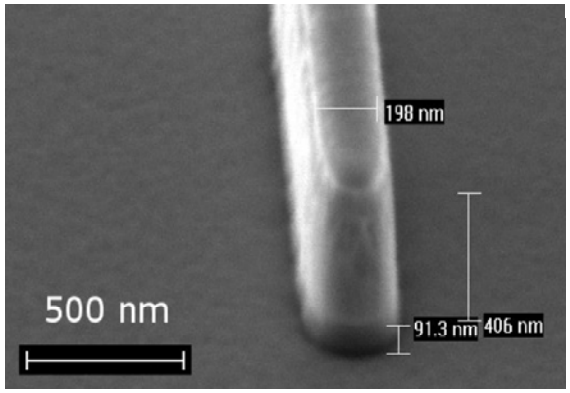


$$\frac{\partial E_{fn}(x)}{\partial x} = \frac{-J}{qn(x)}$$



128 nm InP HBT: Technology Development

E. Lind



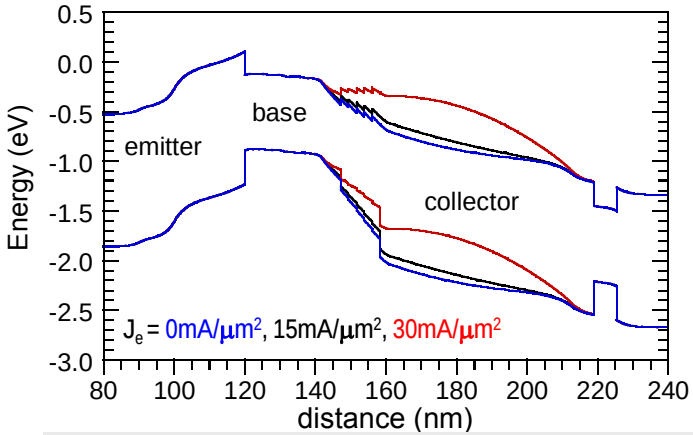
New Emitter Process for 128 and 64 nm junctions

dry etched metal

dry etched junction

refractory W or Mo contact→ stable at very high J_e

< 0.8 Ω-μm² contact resistivity



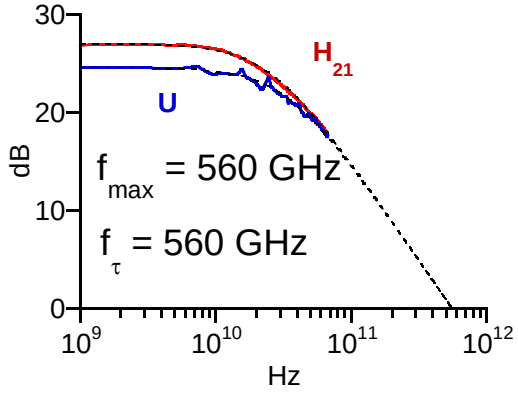
New, thin --12 nm -- base-collector grade:

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



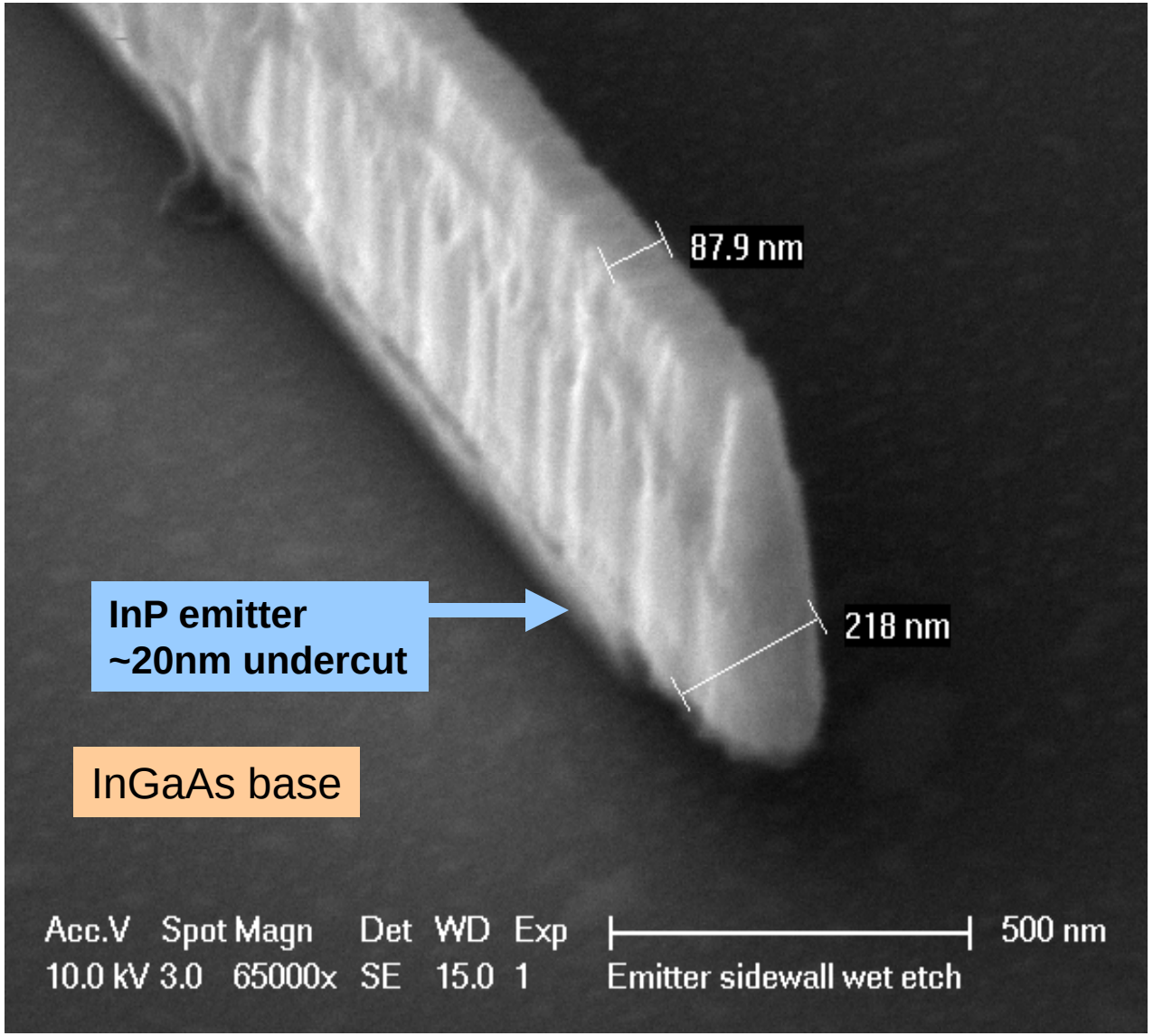
first results: close but not perfect

slip-ups: wide 250 nm emitters, poor base contacts

only a 560GHz / 560GHz / 3 V device

target was 700 / 700 / 3 ... try again soon...

128 nm HBTs in development



True scaled technology:

Thinned epi

reduced access resistance

Should we add a launcher in the B-C Junction ???

No !

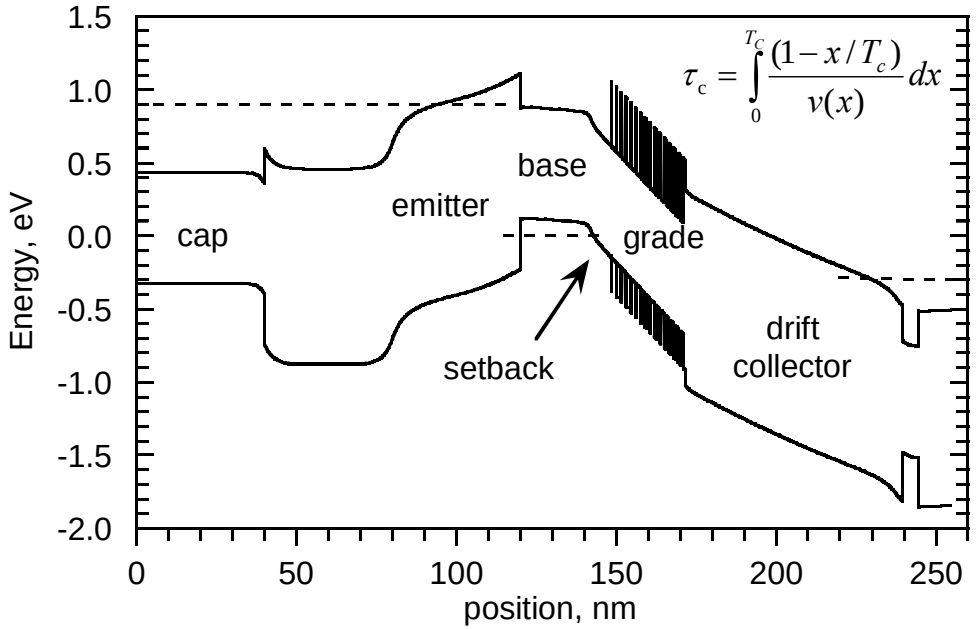
1) Electrons already leave base @

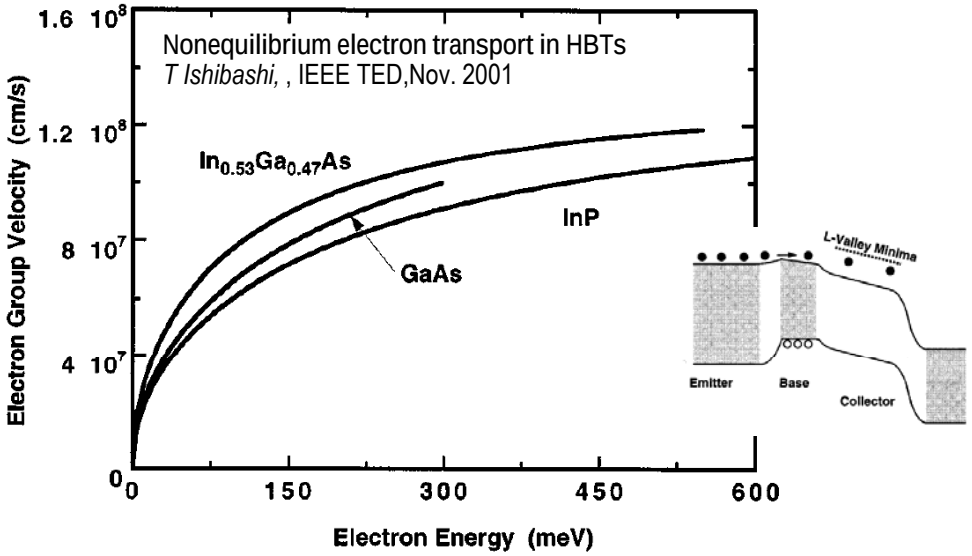
$v_{exit} = v_{thermal} = \sqrt{kT/m^*} \sim 3 \cdot 10^7$ cm/s

2) P + /N - base - collector dipole already gives electrons 0.2 eV, accelerating them to

$v_{ballistic} \sim 8 \cdot 10^7$ cm/s

Large BC band offsets reduce the Kirk - effect - limited drive current, a limitation for GaAsSb/InP DHBTs in Logic ICs.





Comparison to SiGe

1st - Order Design of a 1 THz SiGe Bipolar Transistor

Calculated from
simple BJT model,
& simple scaling laws.

Parameters look ***scary:***
contact resistance is off ITRS roadmap
thin collector→ very high tunneling currents
Is there an alternative ?

Collector thickness determined by:
supposedly: smaller transit time;
actually: smaller C_{cb}/I_c ratio.
current density also determined by C_{cb}/I_c .

Emitter width determined by mA/μm, heating.

Contact resistivity determined by
emitter: low *I*R* drops given high *J*
base: desired f_{max}

Eliminating excess collector area would greatly ease all these considerations

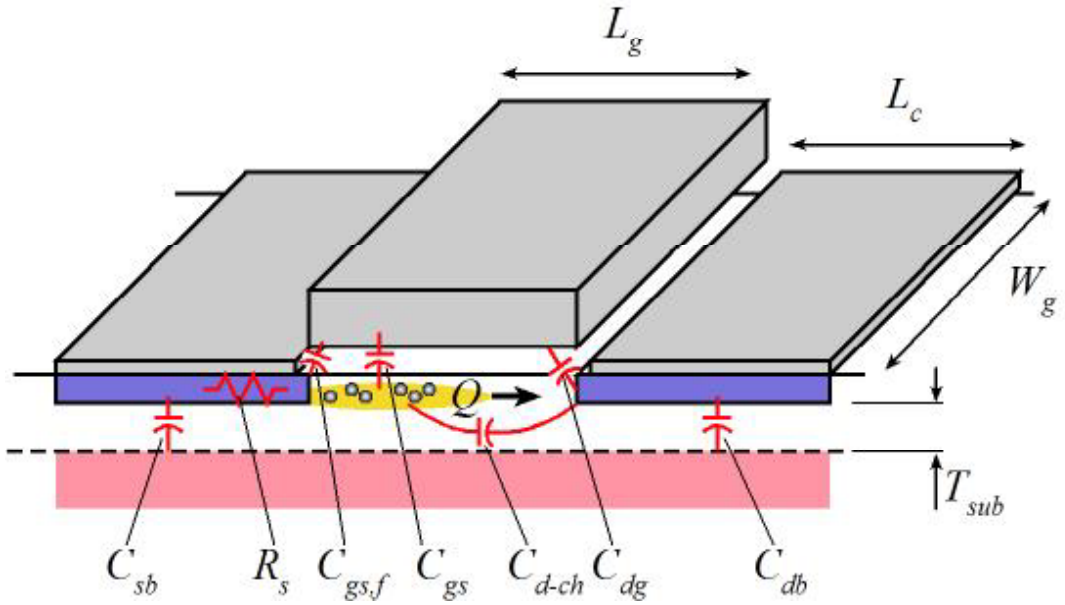
emitter	18	nm width
	1.2	$\Omega\cdot\mu m^2$ access ρ
base	56	nm contact width,
	1.4	$\Omega\cdot\mu m^2$ contact ρ
collector	15	nm thick
	125	$mA/\mu m^2$ current density
	???	V, breakdown
f_τ	1000	GHz
f_{max}	2000	GHz
PAs	1000	GHz
digital	480	GHz
(2:1 static divider metric)		

Assumes collector junction 3:1 wider than emitter.
Assumes contacts 2:1 wider than junctions

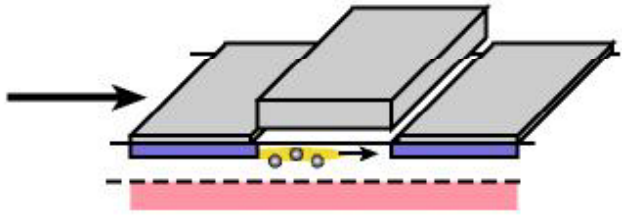
Field-Effect Transistors

Simple FET Scaling

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



All lengths, widths,
thicknesses reduced 2:1



S/D contact resistivity reduced 4:1

$$C_{gs} \sim \varepsilon W_g L_g / T_{ox}$$
$$C_{gs,f} \sim C_{gd} \sim \varepsilon W_g$$
$$C_{sb} \sim C_{db} \sim \varepsilon W_g L_c / T_{sub}$$

If T_{ox} cannot scale with gate length,
$C_{parasitic} / C_{gs}$ increases, ✗
g_m / W_g does not increase
hence $C_{parasitic} / g_m$ does not scale ✗

$$\tau \sim L_g / v$$
$$g_m \sim C_{gs} / \tau \sim (\varepsilon L_g W_g / T_{ox}) / \tau$$
$$G_{ds} \sim C_{d-ch} / \tau \sim \varepsilon W_g / \tau$$

If T_{ox} cannot scale with gate length,
G_{ds}/g_m increases ✗

Well-Known: Si FETs no longer Scale Well

EOT is not scaling as $1/L_g$

T_{ox} (nm) [2]	2.2	2.1	2.0	1.9	1.6	1.5	1.4	1.4	1.3
Gate Length (nm) [2]	75	65	53	45	37	32	28	25	22
g_m/g_{ds} at $5 \cdot L_{min\text{-}digital}$ [3]	47	40	32	30	30	30	30	30	30
1/f-noise ($\mu V^2 \cdot \mu m^2/Hz$) [4]	190	180	160	140	100	90	80	80	70
σ V_{th} matching ($mV \cdot \mu m$) [5]	6	6	6	6	5	5	5	5	5
I_{ds} ($\mu A/\mu m$) [6]	19	15	13	11	9	8	7	6	6
Peak F_t (GHz) [7]	120	140	170	200	240	280	320	360	400
Peak F_{max} (GHz) [8]	200	220	270	310	370	420	480	530	590

(ITRS roadmap copied from Larry Larson's files)

High-K gate dielectrics: significant SiO_2 interlayer→ limits gate capacitance density

It is also hard to reduce access resistance by the amount needed

Because gate oxide scales badly, modern MOSFETs scale badly

output resistance drops, voltage gain drops

gate capacitance does decrease, but other capacitances don't !

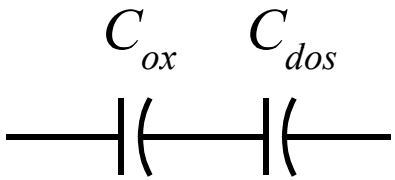
... which hurts high-frequency performance

Why consider III-V (InGaAs/InP) CMOS ?

Low access resistance: 1 Ω-μm², 10 Ω-μm
Light electron→ high electron velocity
→ increased I_d / W_g at a given oxide thickness (?)
→ decreased C_{gs}/g_m at a given gate length

Challenge:
Low density of states

$$C_{dos} = \frac{q^2 m^*}{\pi \hbar^2}$$



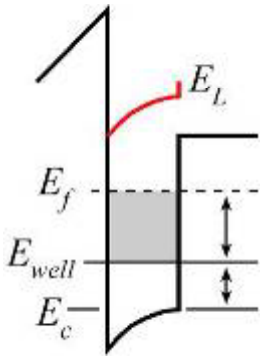
3.4 μF/cm² @ 1 nm EOT

~3 μF/cm² ballistic case

*limits n_s to ~ 6*10¹² /cm²*
limits I_d / W_g
limits g_m / W_g

Challenge:
filling of low-mobility satellite valleys

*limits n_s to ~ 8*10¹² /cm²*
limits I_d / W_g

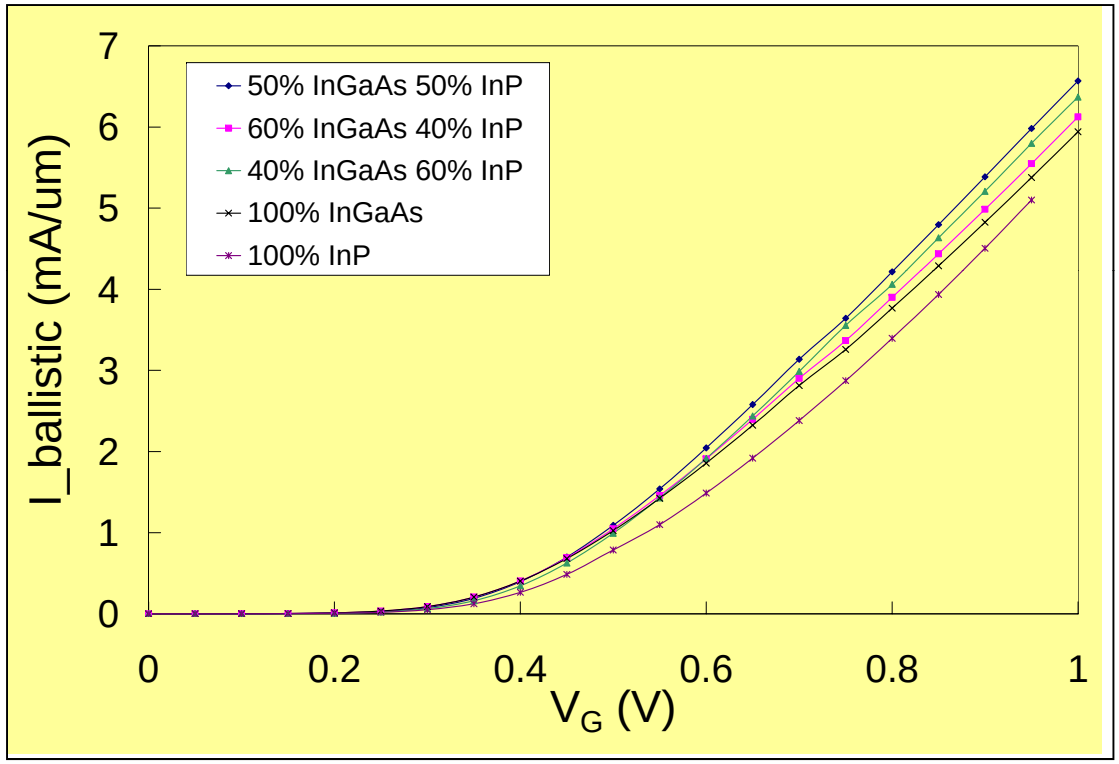


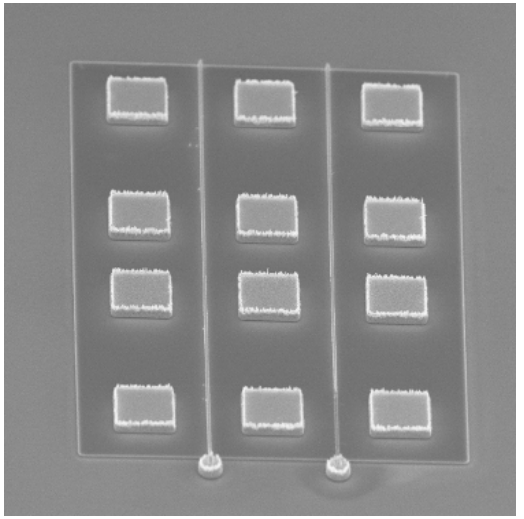
Challenge:
light electron limits vertical scaling
~1.5-2.5 nm minimum mean electron depth

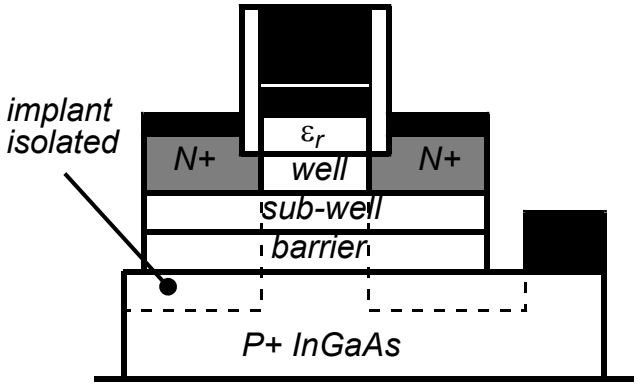
SRC III-V CMOS Center : What we might accomplish

Drive current simulation- ideal (ballistic) assumptions

Taur & Asbeck Groups, UCSD; Fischetti Group: U-Mass: IEDM2007







22 nm gate length, 5 nm thick InGaAs / InP channel

III-V FET Scaling Limits: Implications for HEMTs

Extremely High g_m > 2.2 S/mm and f_T > 550 GHz in 30-nm Enhancement-Mode InP-HEMTs with Pt/Mo/Ti/Pt/Au Buried Gate

Keisuke Shinohara[1], Wonill Ha[1], Mark J.W. Rodwell[2], and Berinder Brar[1]

[1]*Teledyne Scientific Company, 1049 Camino Dos Rios, Thousand Oaks, CA, USA*
[2]*Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA*
Phone: +1-805-373-4098, Fax: +1-805-373-4860, E-mail: kshinohara@teledyne.com

Limited density of states limits sheet charge
→ Schottky barrier leakage
→ limits drive current, places lower limit on $C_{parasitic}\Delta V/I$

Well energy limits vertical scaling
→ poor output conductance for < 35 nm gate length → degraded f_{max}

High access resistance in standard HEMT structure

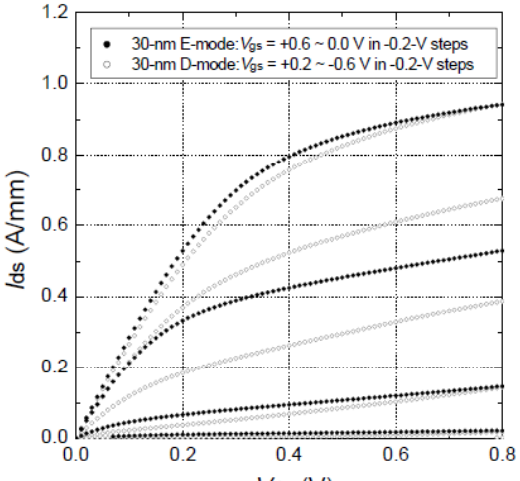
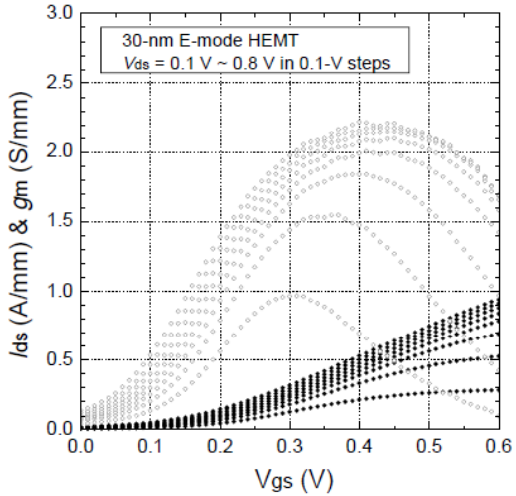
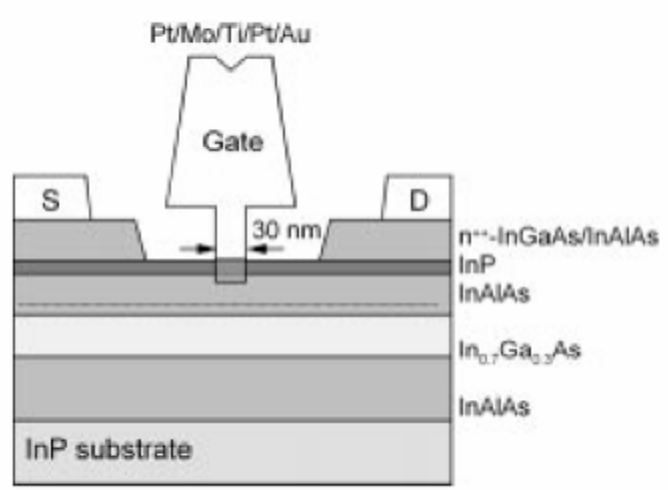


Fig. 3. Output characteristics of 30-nm E/D-mode HEMTs

Ti/Pt/Au gate before/after annealing.





Conclusions...

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)