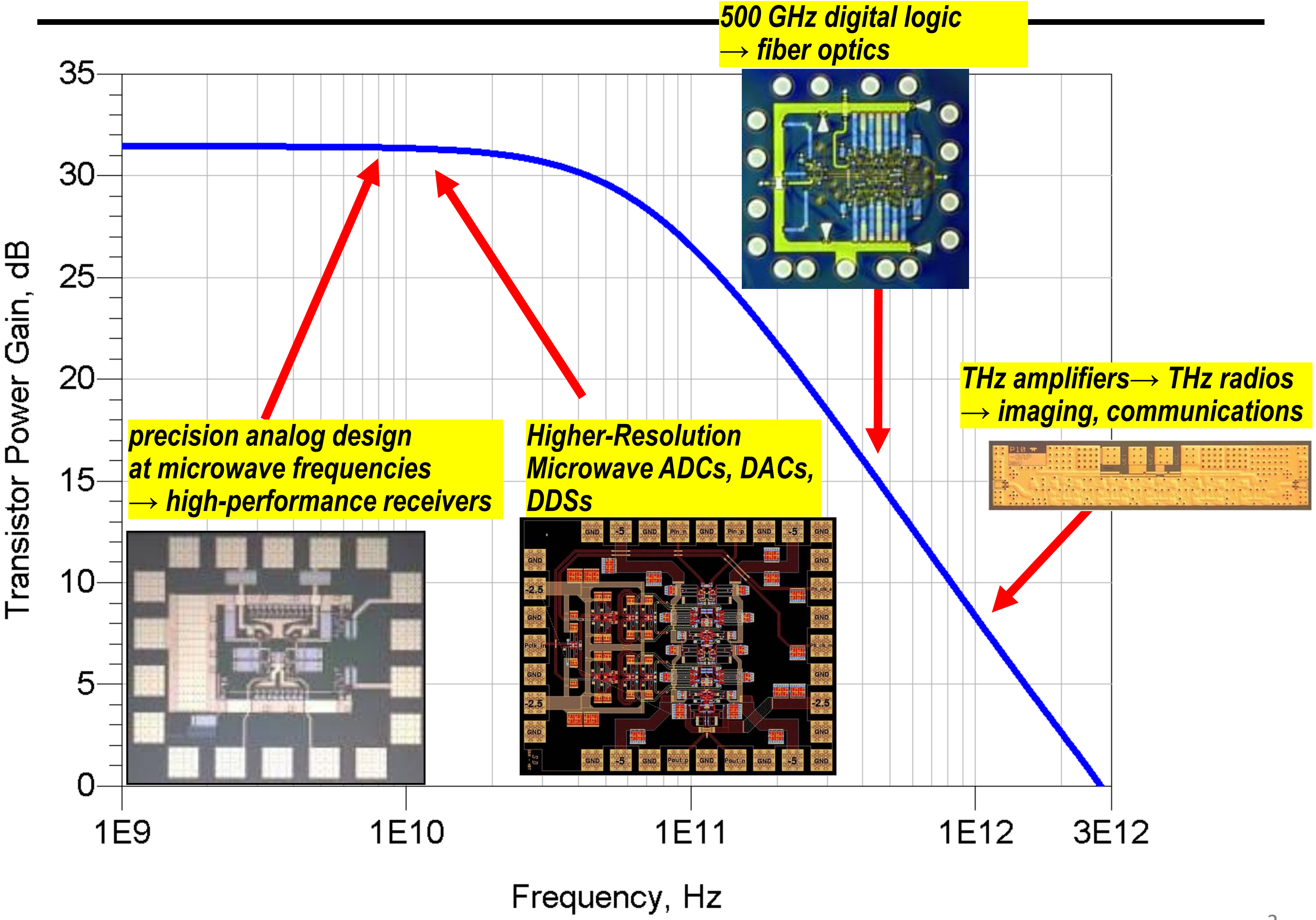
WSG Workshop: Performance Metrics for mm-Wave Devices and Circuits from the Perspective of the International Technology Roadmap for Semiconductors (ITRS), IEEE IMS Symposium, May 17, 2015, Phoenix

III-V HBT and (MOS) HEMT scaling

Mark Rodwell,
University of California, Santa Barbara

THz Transistors: Systems Benefit from 5-500 GHz



500 GHz digital logic
→ fiber optics

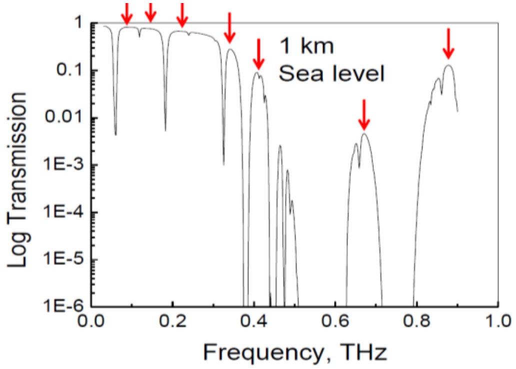
THz amplifiers→ THz radios
→ imaging, communications

precision analog design
at microwave frequencies
→ high-performance receivers

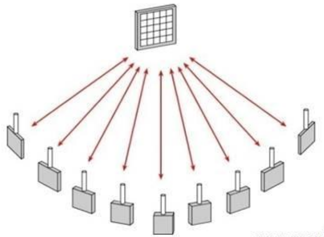
Higher-Resolution
Microwave ADCs, DACs,
DDSs

mm-Wave wireless: attributes & challenges

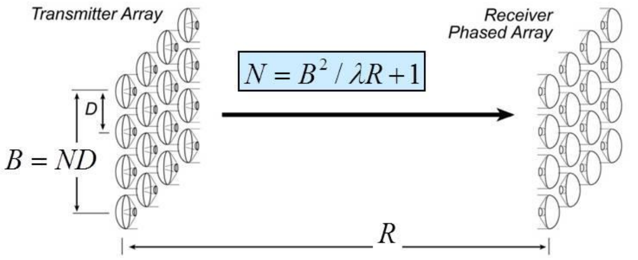
wide bandwidths available



short wavelengths→ many parallel channels

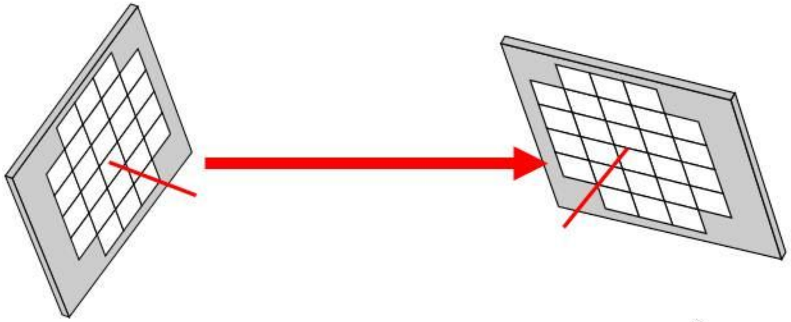


$$\text{angular resolution} \approx \frac{\text{wavelength}}{\text{array width}}$$



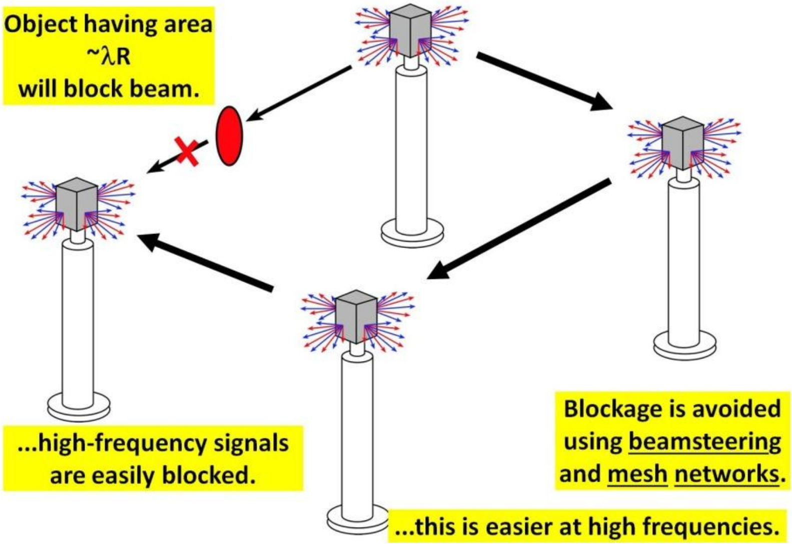
$$\#\text{channels} \propto (\text{aperture area})^2 / (\text{wavelength} \cdot \text{distance})^2$$

Need phased arrays



$$\frac{P_{received}}{P_{transmit}} \propto N_{receive} N_{transmit} \frac{\lambda^2}{R^2} e^{-\alpha R}$$

Need mesh networks



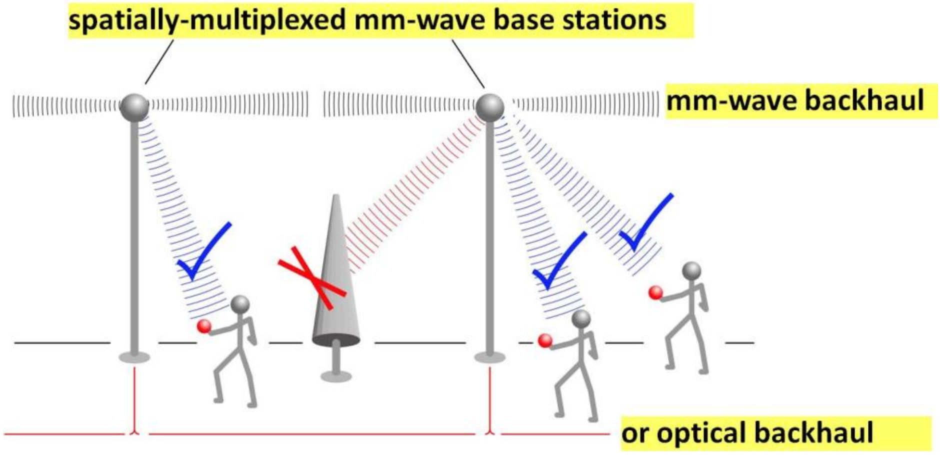
Object having area ~λR will block beam.

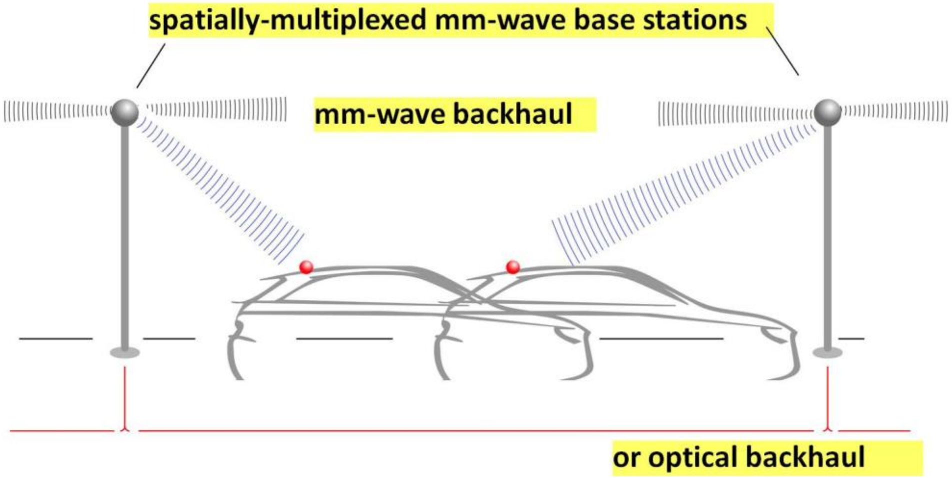
...high-frequency signals are easily blocked.

Blockage is avoided using beamsteering and mesh networks.

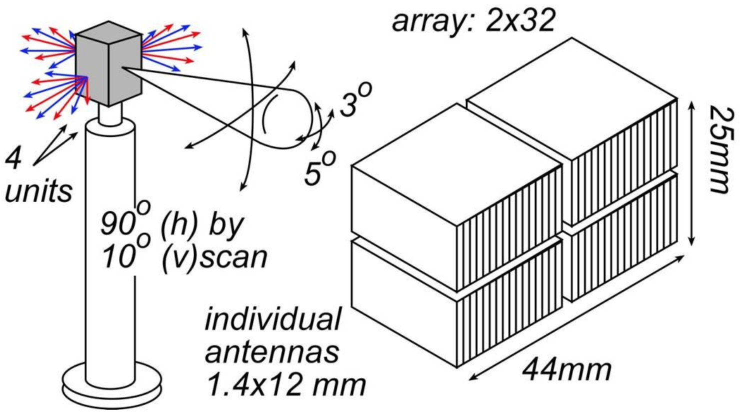
...this is easier at high frequencies.

mm-Waves: high-capacity mobile communications

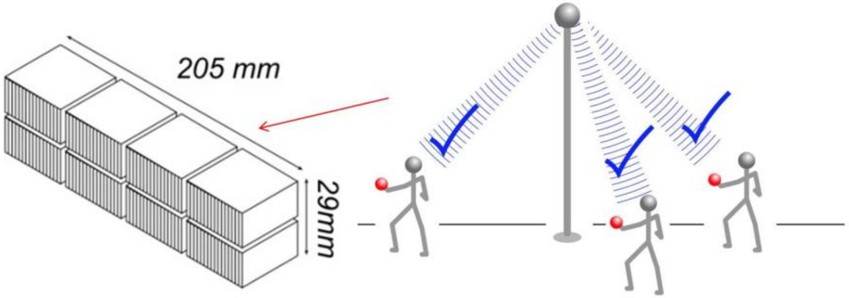




140 GHz, 10 Gb/s Adaptive Picocell Backhaul



60 GHz, 1 Tb/s Spatially-Multiplexed Base Station



128 users/face, 512 total users, each beam 2Gb/s

Needed: phased arrays, 50-500mW power amplifiers, low-noise-figure LNAs

mm-wave imaging radar: TV-like resolution

mm-waves → high resolution from small apertures

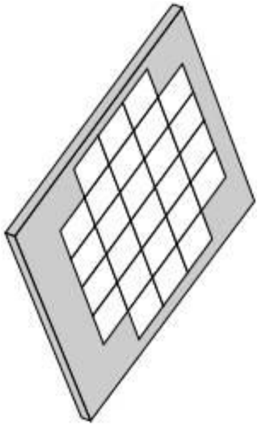
What you see in fog

What 10GHz radar shows

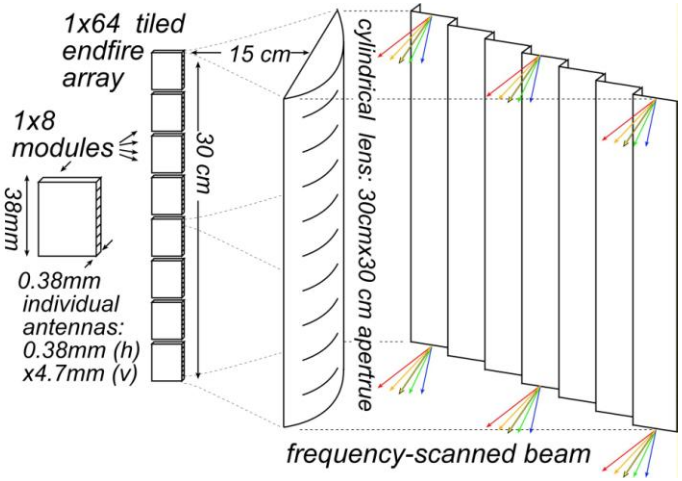
What you want to see

needs: ~0.2^o resolution, 10^3-10^6 pixels

Large NxN phased array



Frequency-scanned 1xN array



InP HBTs and HEMTs for PAs and LNAs

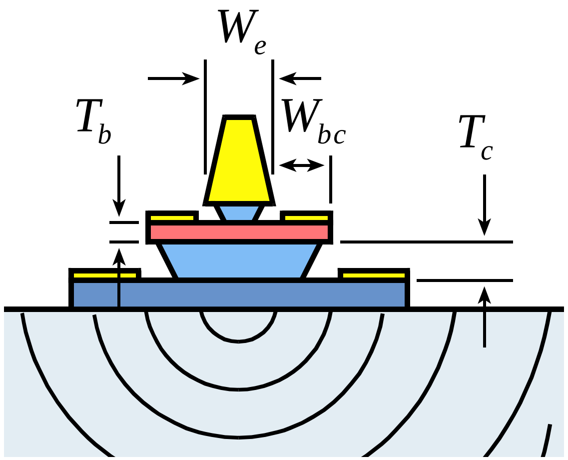
Cell phones and Higher-Performance WiFi sets:
    GaAs HBT power amplifiers
    GaAs PHEMT LNAs

29-34GHz: emerging bands for 5G
    InP HBT PAs, InP HEMT LNAs ?

Later: 60, 71-76, 81-86, 140 GHz

Heterojunction Bipolar Transistors

Bipolar Transistor Design



$(\text{emitter length } L_E)$

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

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

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

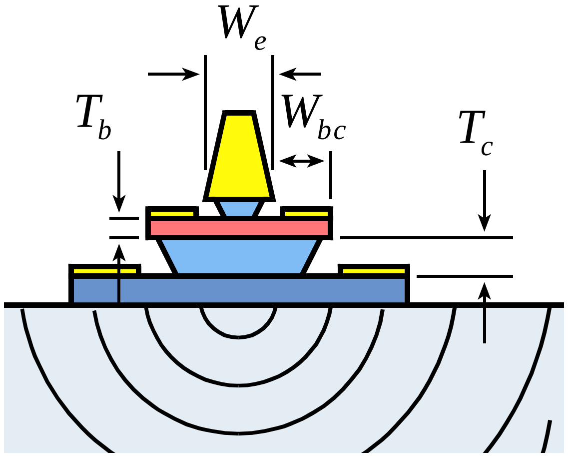
$$I_{c,\max} \propto v_{sat} A_e \left(V_{\text{ce,operating}} + V_{\text{ce,punch-through}}\right) / T_c^2$$

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

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

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

Bipolar Transistor Design: Scaling



$(\text{emitter length } L_E)$

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

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

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

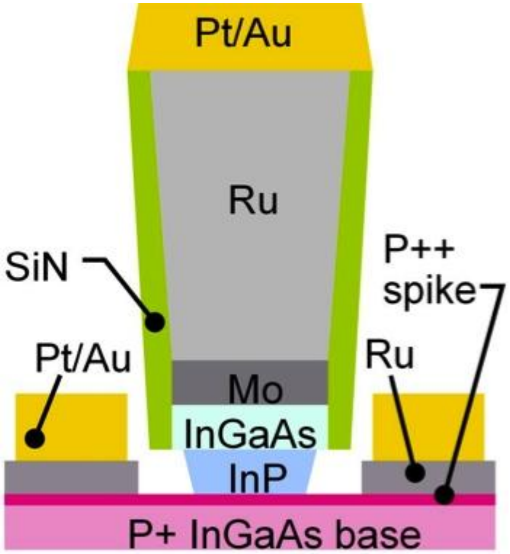
$$I_{c,\max} \propto v_{sat} A_e (V_{\text{ce,operating}} + V_{\text{ce,punch-through}}) / T_c^2$$

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

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

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

Scaling Laws, Scaling Roadmap



HBT parameter	change
emitter & collector junction widths	decrease 4:1
current density (mA/μm²)	increase 4:1
current density (mA/μm)	constant
collector depletion thickness	decrease 2:1
base thickness	decrease 1.4:1
emitter & base contact resistivities	decrease 4:1

Narrow junctions.

Thin layers

High current density

Ultra low resistivity contacts

Scaling Node	64	32	16	nm
Emitter Width	64	32	16	nm
Resistivity	2	1	0.5	Ω-μm^2
Base Thickness	18	15	13	nm
Contact width	60	30	15	nm
Contact ρ	2.5	1.25	0.63	Ω-μm^2
Collector Width	180	90	45	nm
Thickness	53	37.5	26	nm
Current Density	36	72	140	mA/μm^2
f_τ	1.0	1.4	2.0	THz
$f_{\max}$	2.0	2.8	4.0	THz

Can we make a 2 THz SiGe Bipolar Transistor ?

Simple physics clearly drives scaling

transit times, C_{cb}/I_c

→ thinner layers, higher current density

high power density → narrow junctions

small junctions→ low resistance contacts

Key challenge: Breakdown

15 nm collector → very low breakdown

Also required:

low resistivity Ohmic contacts to Si

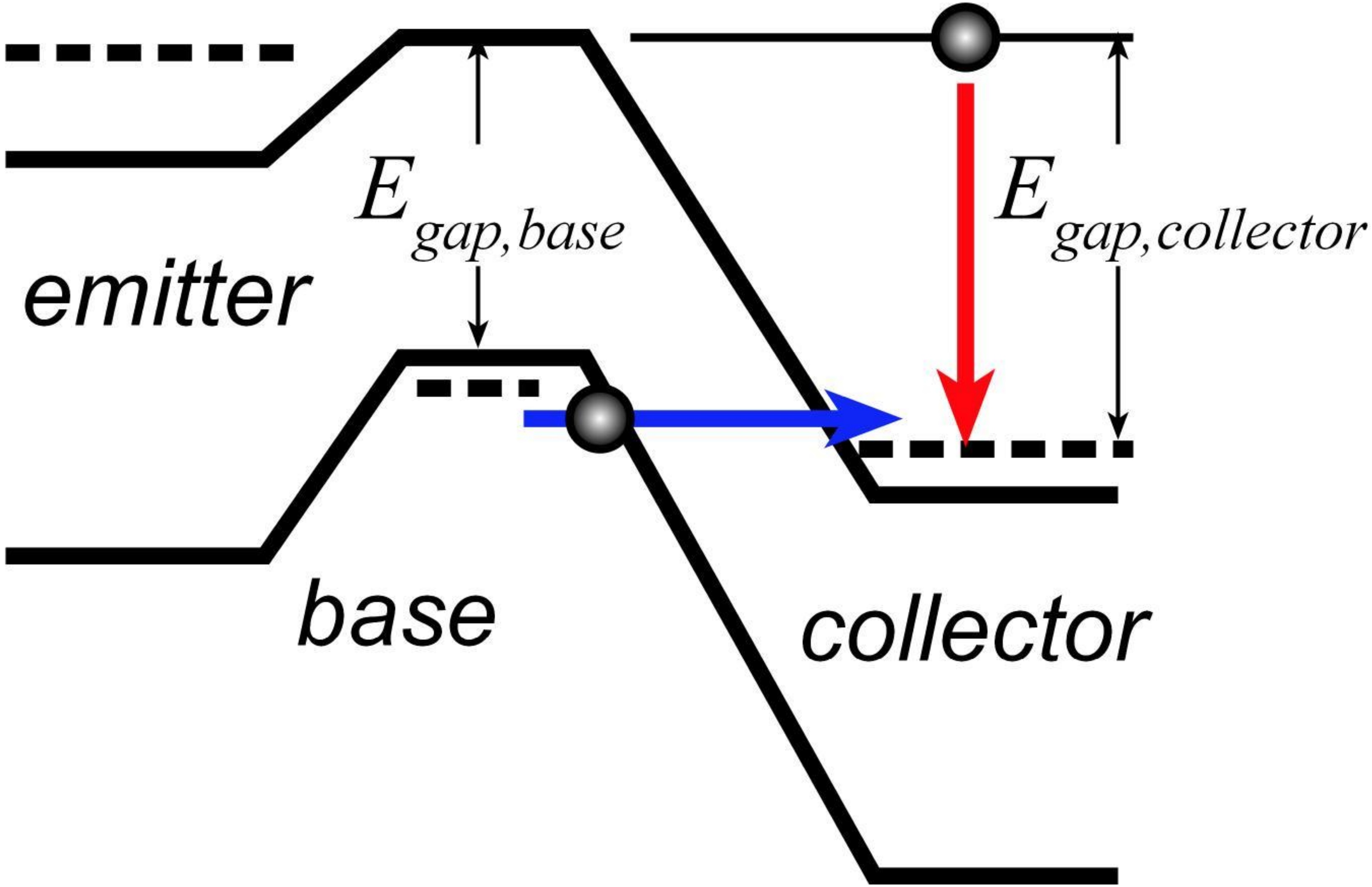
very high current densities: heat

	InP	SiGe	
emitter	64	18	nm width
	2	**0.6**	$\Omega\cdot\mu m^2$ access ρ
base	64	18	nm contact width,
	2.5	**0.7**	$\Omega\cdot\mu m^2$ contact ρ
collector	53	**15**	nm thick
	36	125	$mA/\mu m^2$
	2.75	**1.3?**	V, breakdown
f_τ	**1000**	**1000**	GHz
f_{max}	**2000**	**2000**	GHz
PAs	1000	1000	GHz
digital	480	480	GHz
(2:1 static divider metric)			

Assumes collector junction 3:1 wider than emitter.

Assumes SiGe contacts no wider than junctions

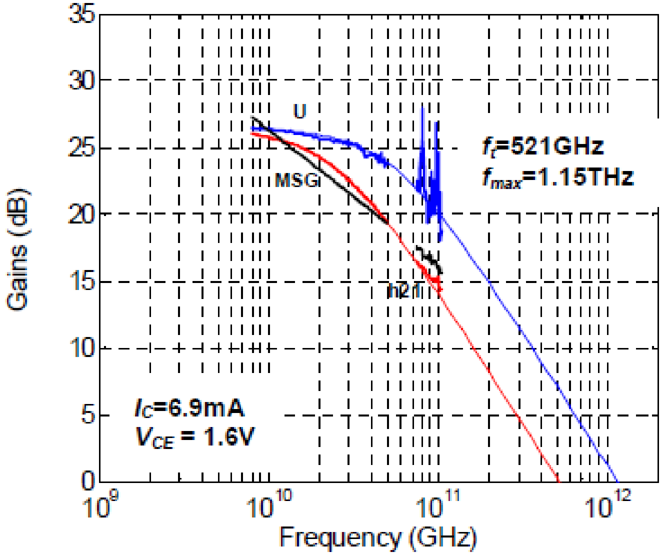
Energy-limited vs. field-limited breakdown



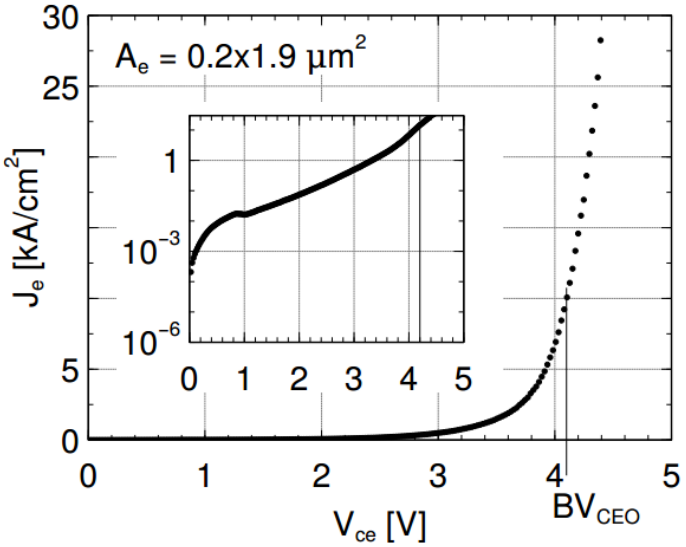
band-band tunneling: base bandgap
impact ionization: collector bandgap

THz InP HBTs: Performance @ 130 nm Node

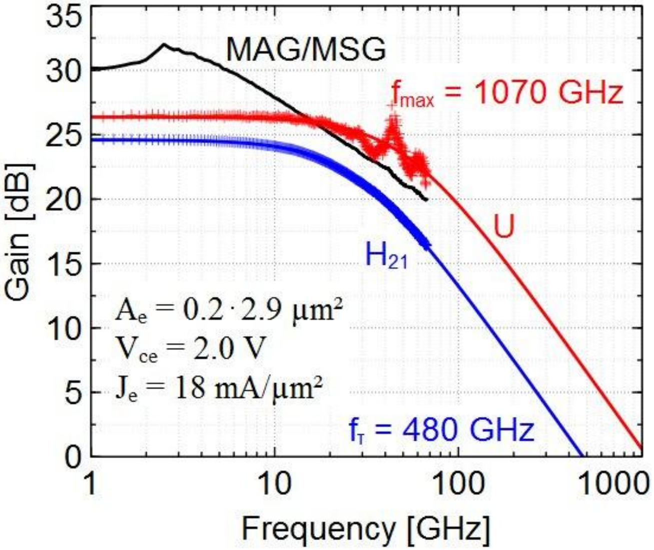
Teledyne: M. Urteaga *et al*: 2011 DRC



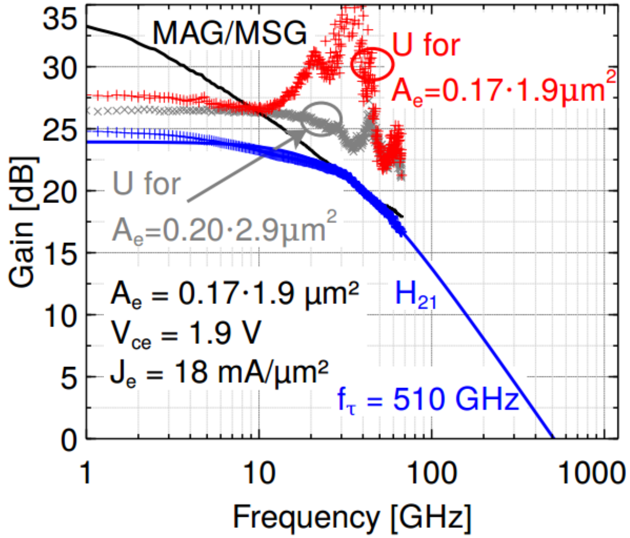
UCSB: J. Rode *et al*: in review



UCSB: J. Rode *et al*: in review

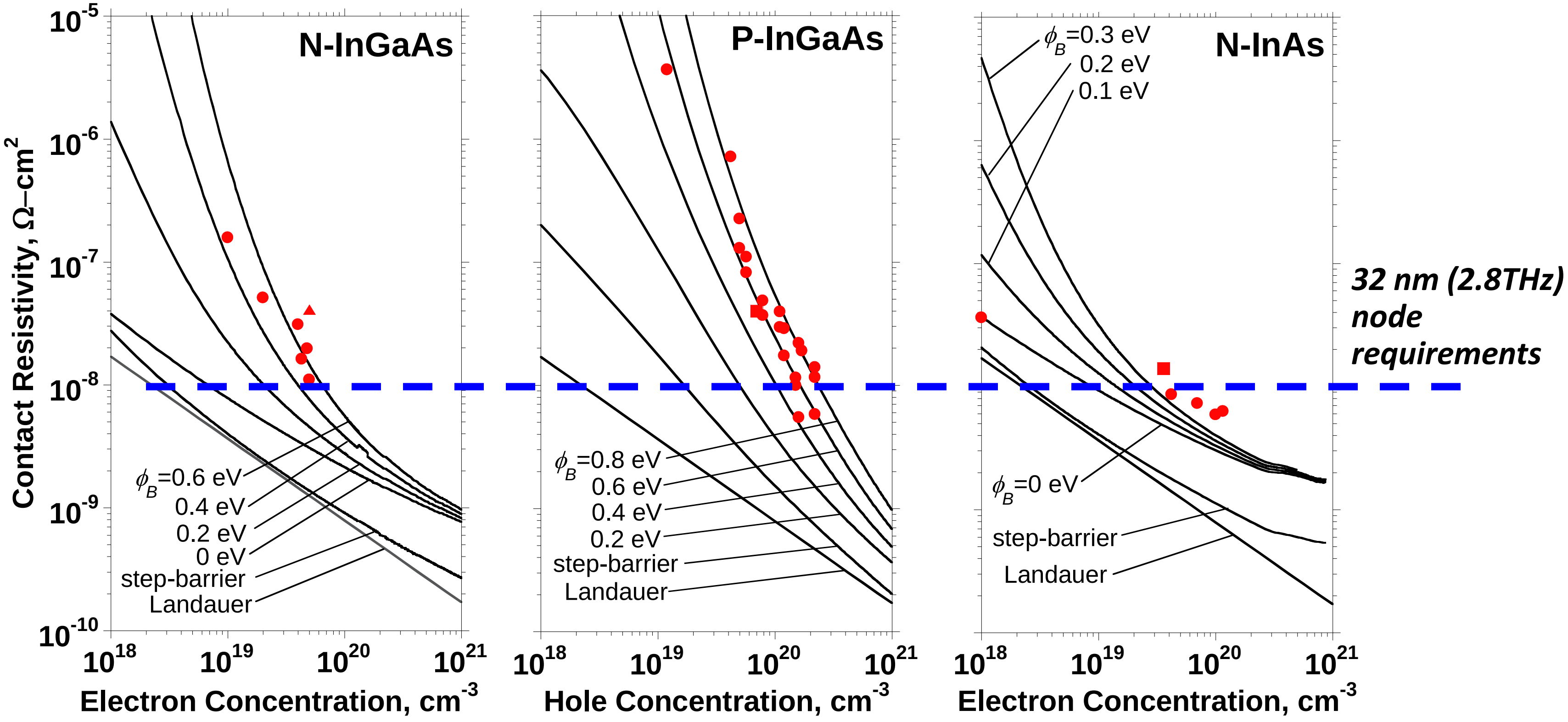


UCSB: J. Rode *et al*: in review



Refractory Contacts to In(Ga)As

Baraskar *et al*, Journal of Applied Physics, 2013

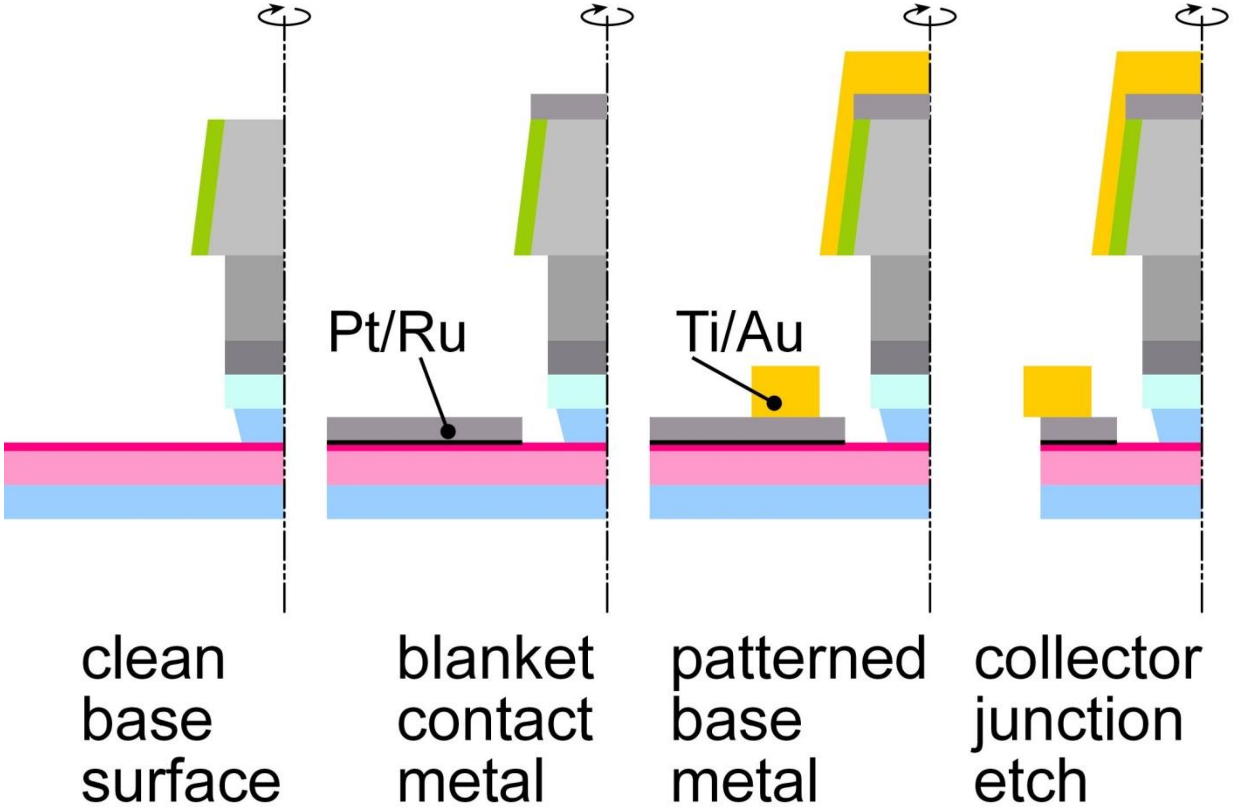


Refractory: robust under high-current operation / Low penetration depth: ~ 1 nm / Performance sufficient for 32 nm /2.8 THz node.

Why no ~2THz HBTs today ?

Problem: reproducing these base contacts in full HBT process flow

Refractory Blanket Base Metal Process (1)

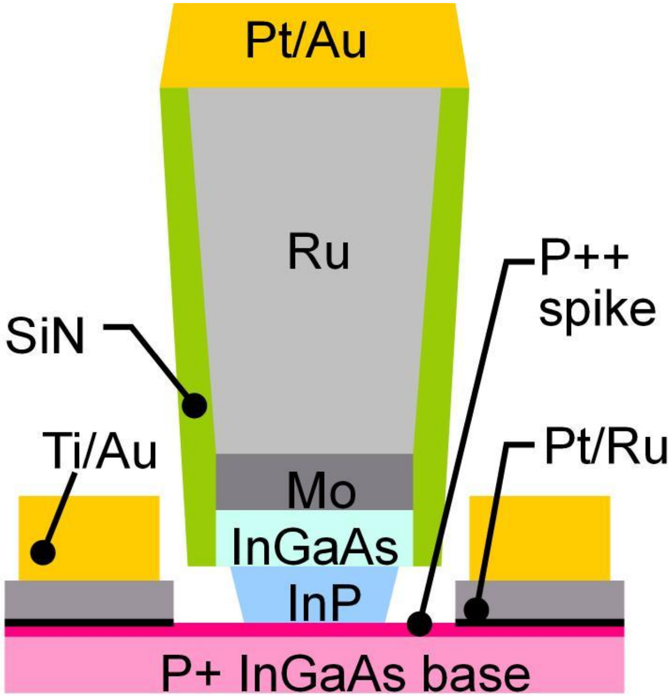


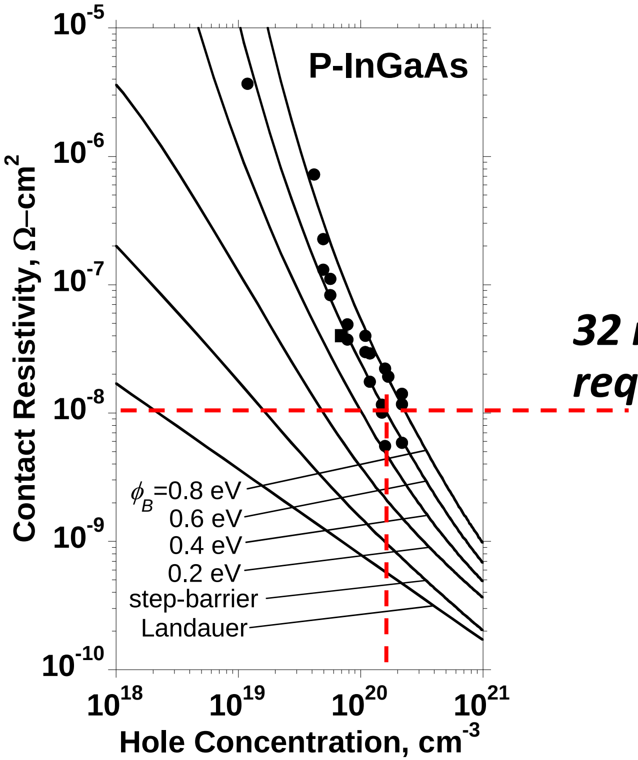
Metal deposited on clean surface; no resist residue

Refractory Ru contact layer→ low penetration depth

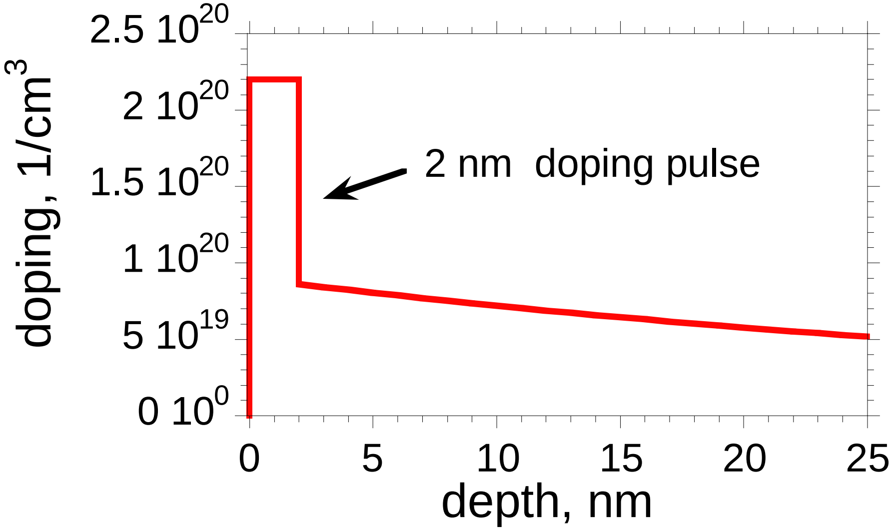
2nm Pt reaction layer→ penetrate surface contaminants

Refractory Blanket Base Metal Process (2)





32 nm node requirement

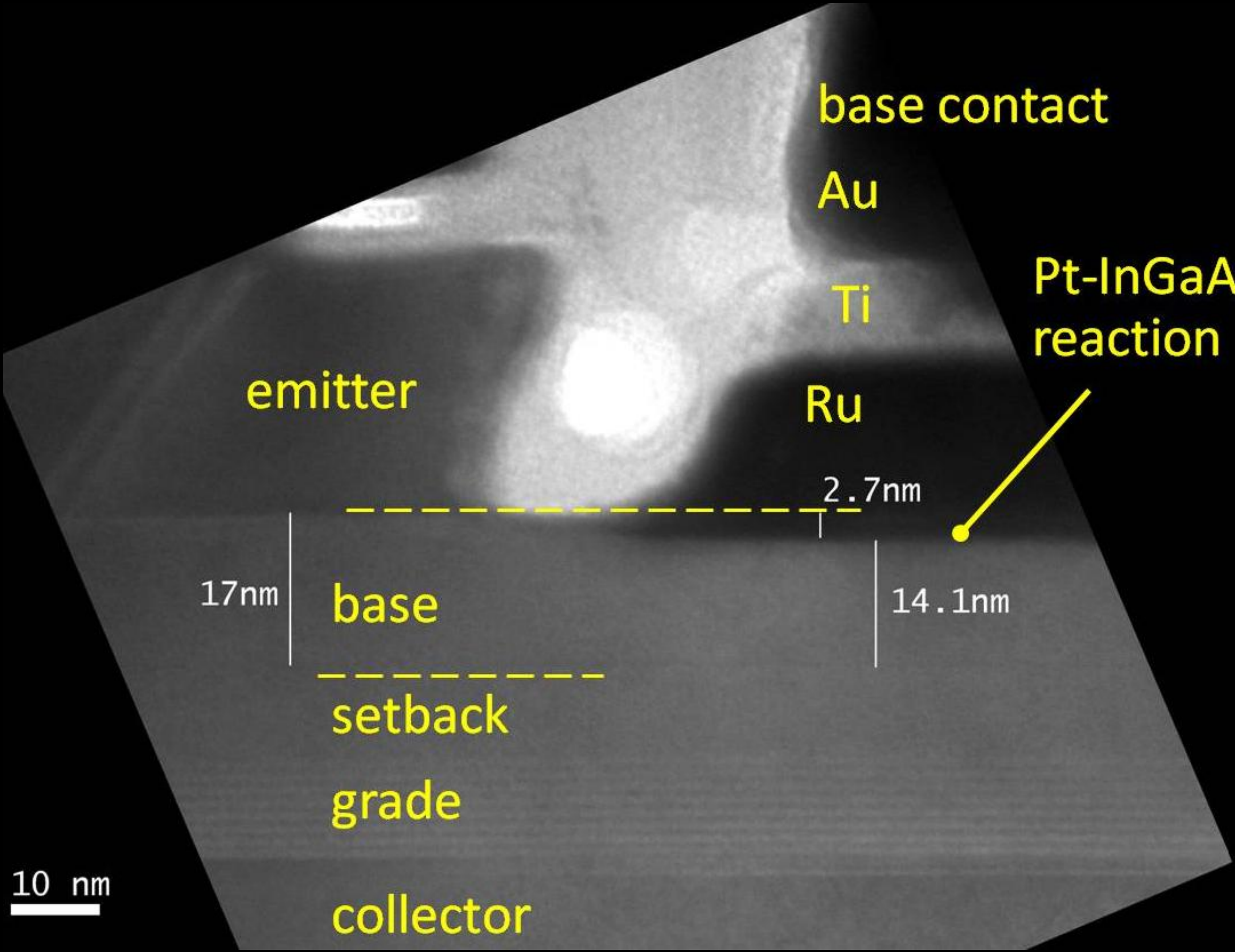


Increased surface doping:
reduced contact resistivity,
but increased Auger recombination.

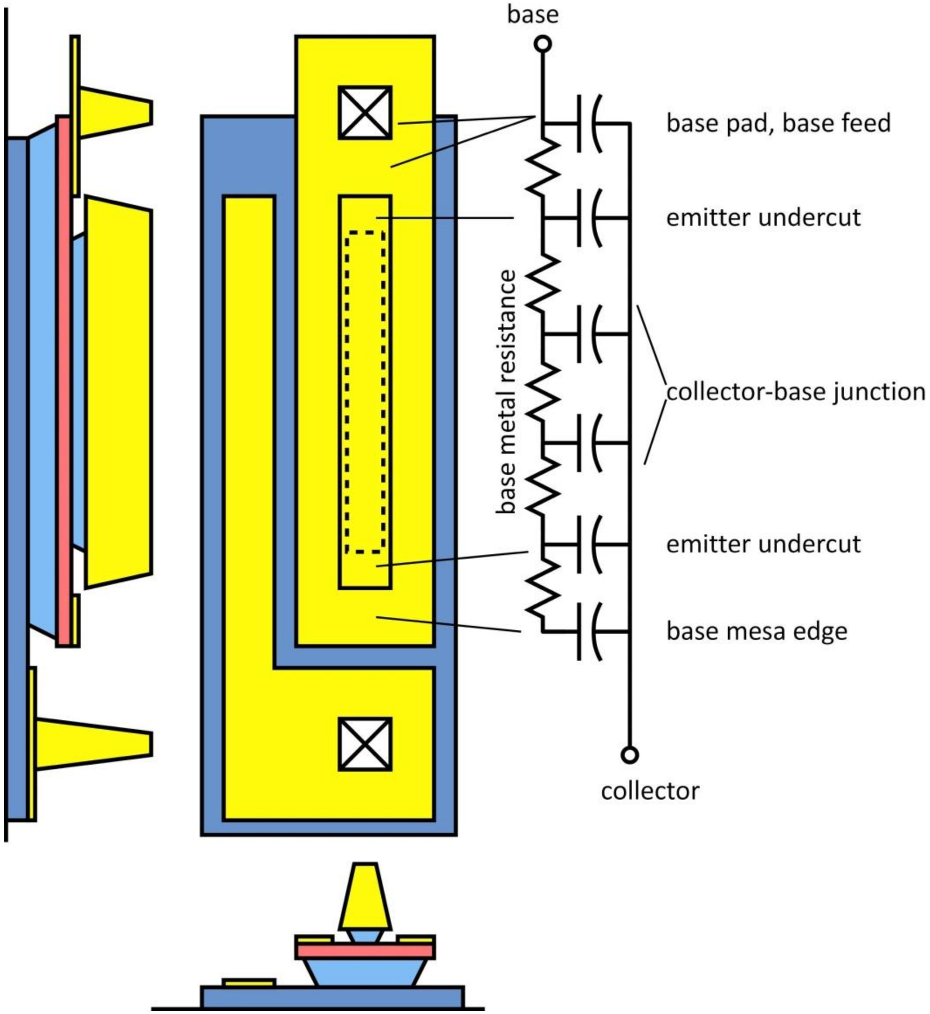
→ Surface doping spike at most 2-5 thick.

Refractory contacts do not penetrate; compatible with pulse doping.

Blanket Base Metal Process



Parasitics along length of HBT emitter



Base pad & feed
increases C_{cb}

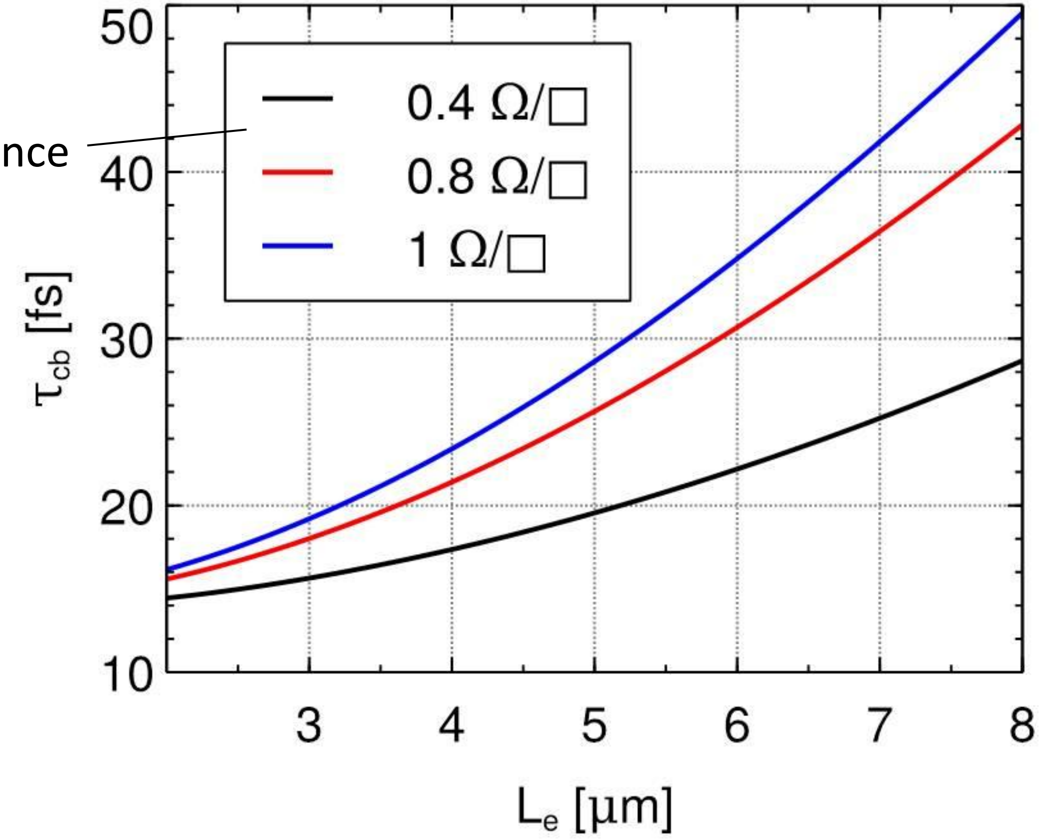
Emitter undercut
actual junction shorter than drawn.
→ excess C_{cb} , excess base metal resistance

Base metal resistance
adds to R_{bb}

all these factors decrease f_{max}

Emitter Length Effects: Decreased f_{max}

Results from finite-element modeling

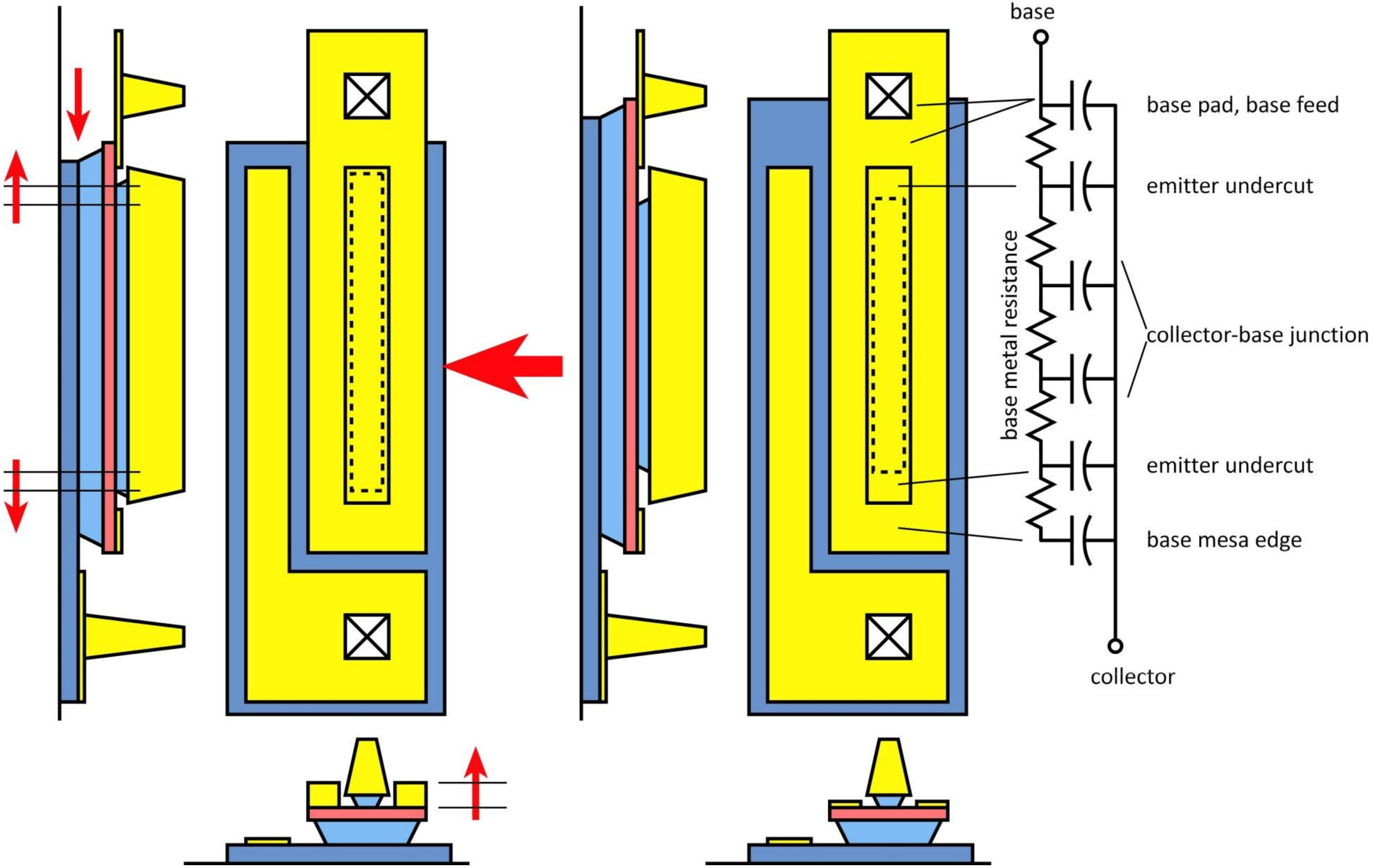


base metal
sheet resistance

$$f_{max} = \sqrt{\frac{f_\tau}{8\pi\,\tau_{cb}}}$$

On a 2 μm emitter finger, effect of base metal resistance can be comparable to adding 3 Ω-μm^2 to the base contact resistivity !

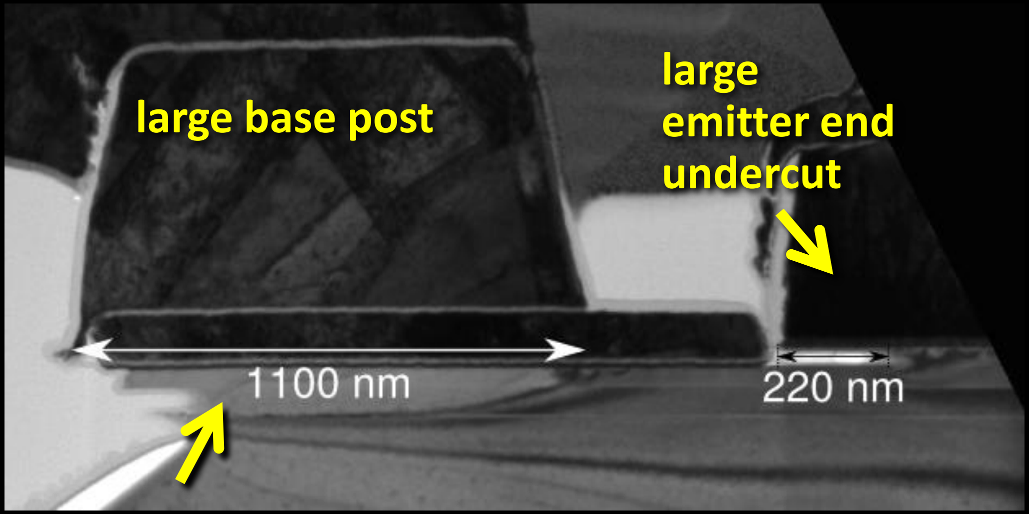
Reducing Emitter Length Effects



Reducing Emitter Length Effects

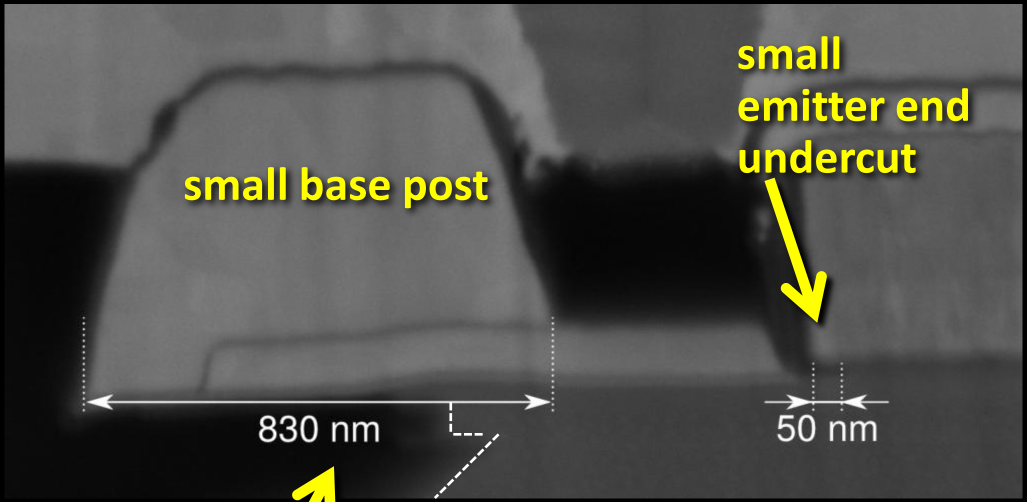
J. Rode
in review

before



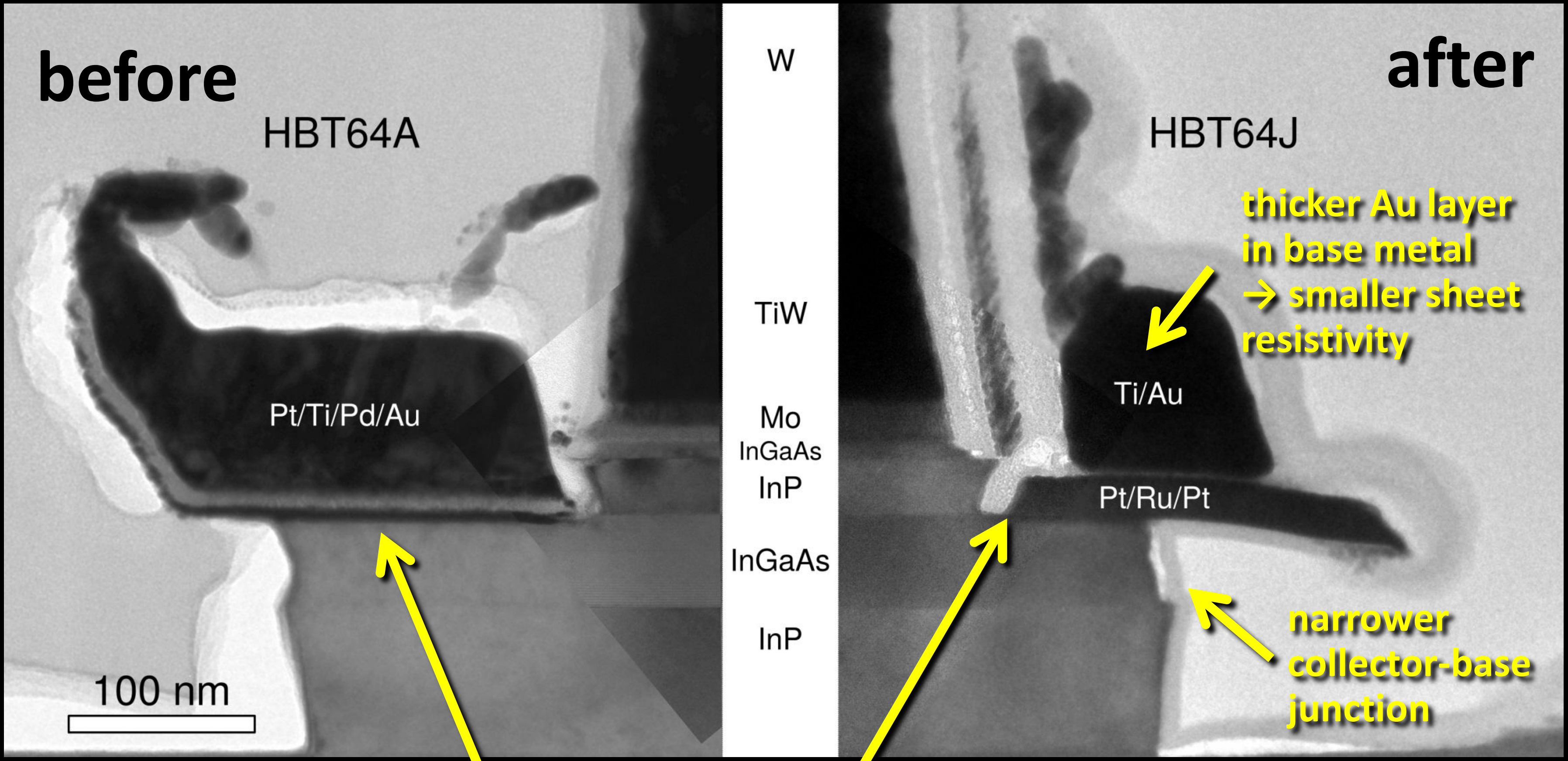
Small Base Post Undercut

after



Large Base Post Undercut

Reducing Emitter Length Effects



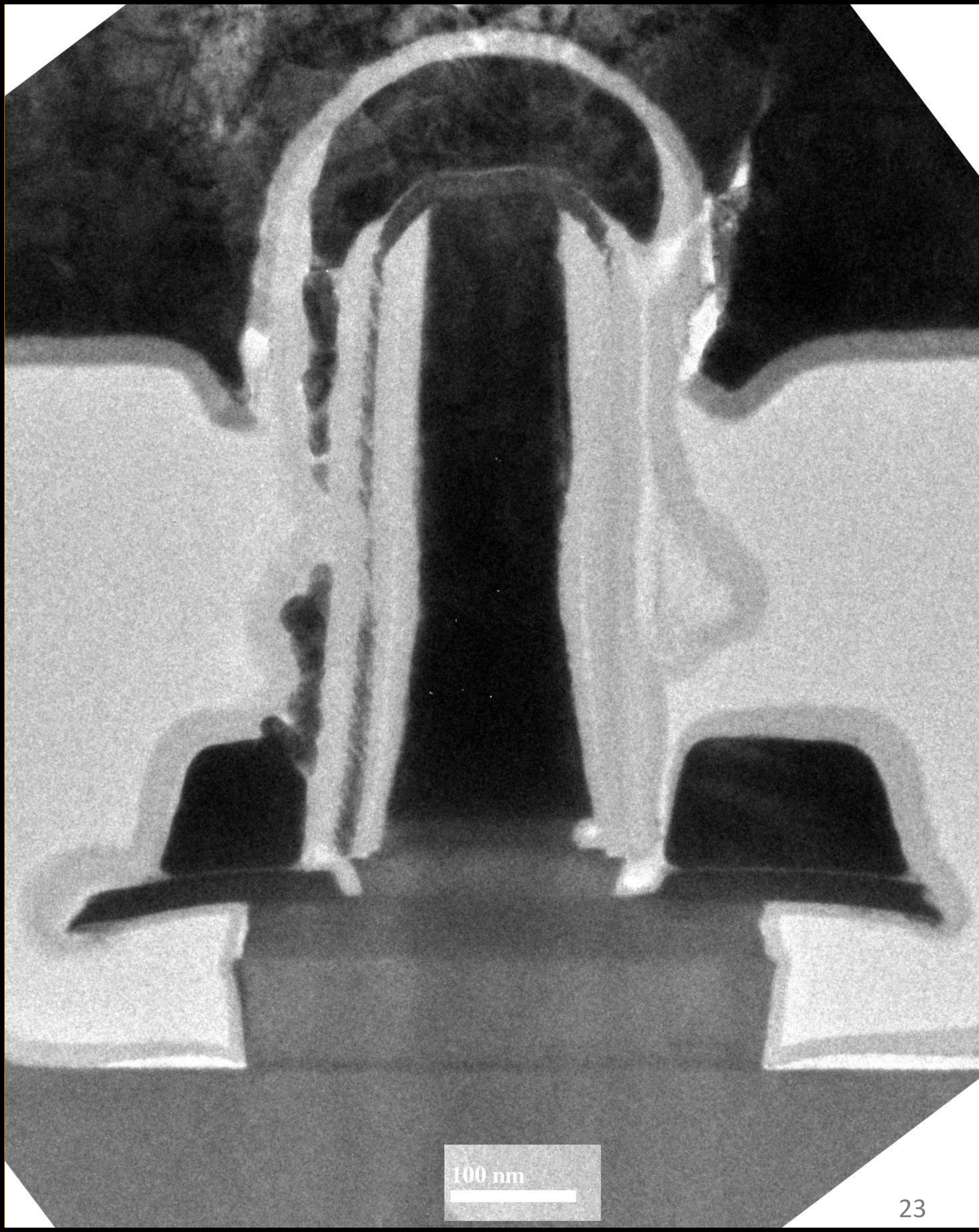
thicker Au layer in base metal → smaller sheet resistivity

narrower collector-base junction

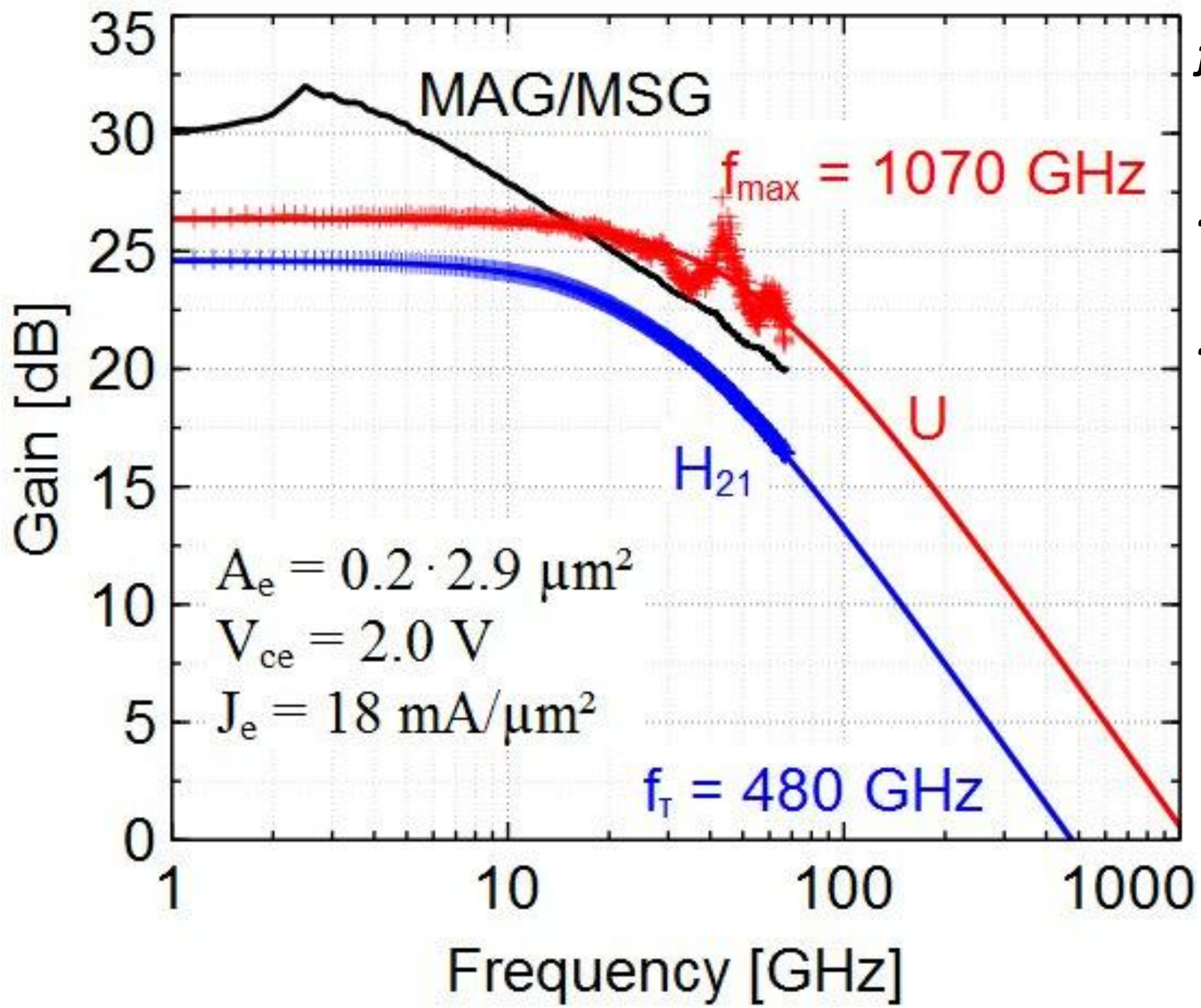
smaller contact penetration into base

J. Rode in review

200nm emitter InP HBT



200nm emitter width: High Fmax



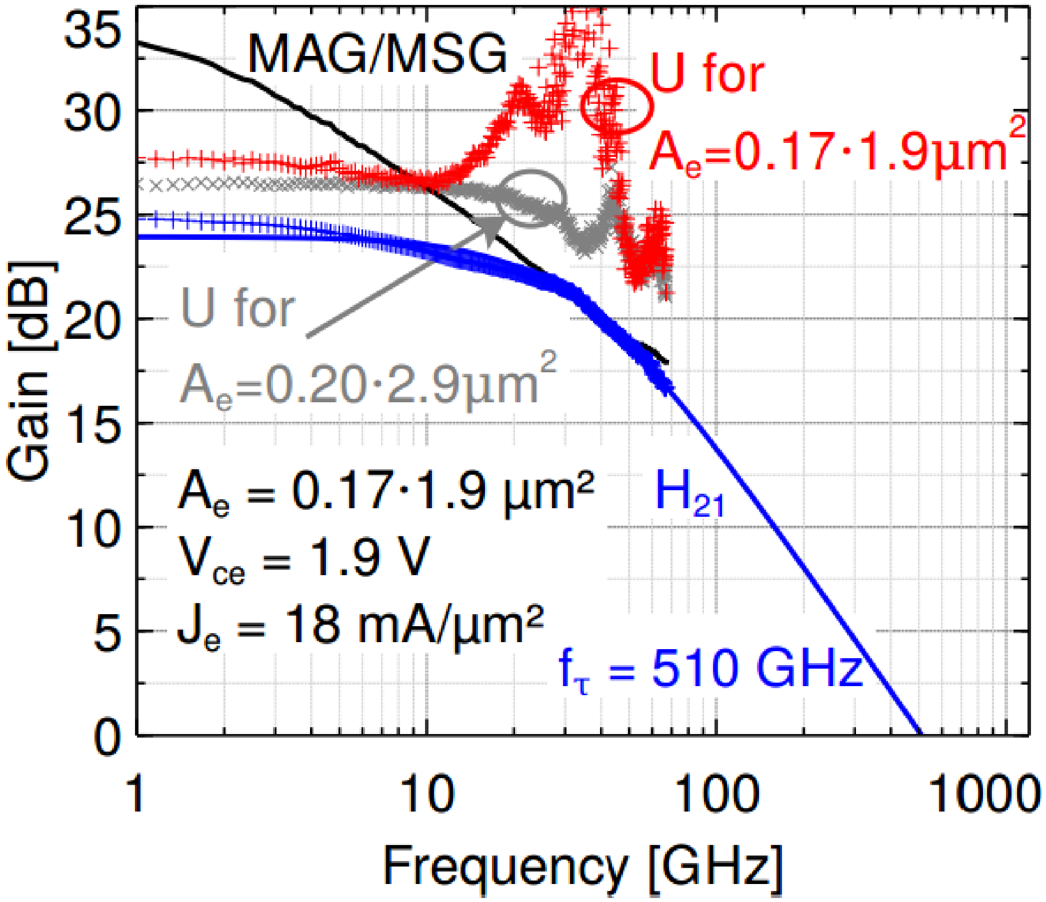
f_{max} is high:

...even at 2.9 μm emitter length

...even at 200nm emitter width

J. Rode
in review

160nm emitter width: Unmeasurable Fmax



J. Rode
in review

on HBTs with

...shorter 1.9 μm emitter length
...narrower 170nm emitter width

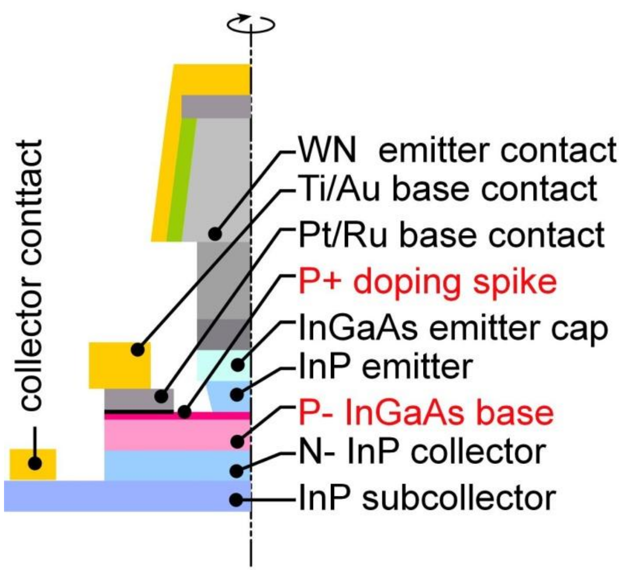
f_{max} cannot be measured because of calibration difficulties (small Y_{12})

f_{max} probably above 1.1THz,
but we cannot prove this.

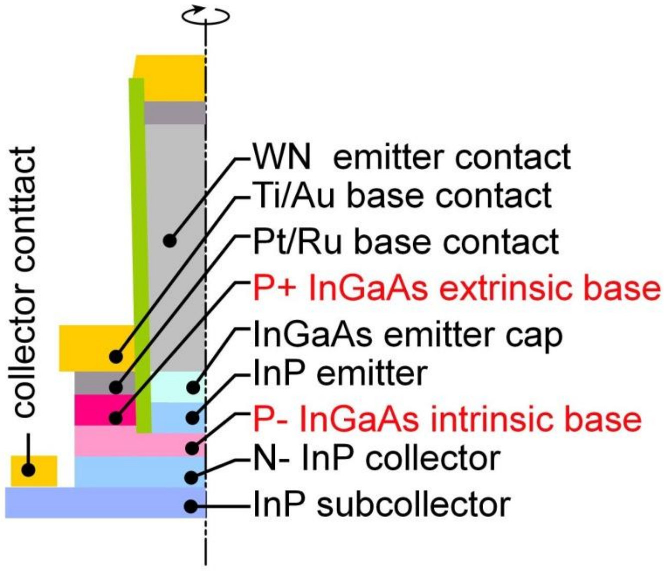
Better fmax measurement would require on-wafer LRL standards.

We no do not at present
have the resources to pursue this.

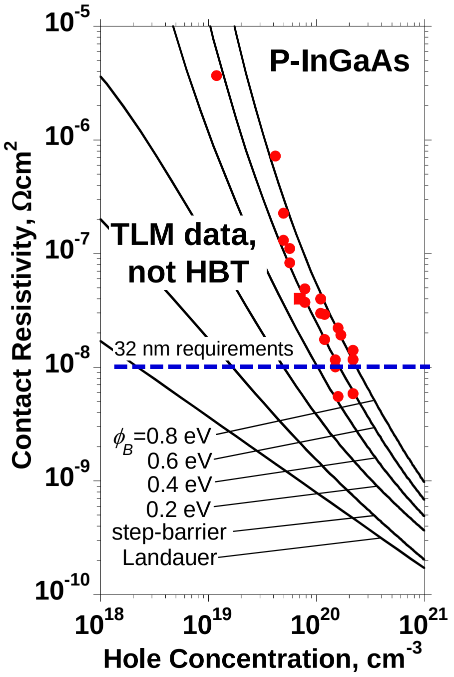
Regrowth for high β in THz HBTs ?



present



regrown



2-3 THz f_{max} HBTs need ~$1.5*10^{20}$ cm^{-3} doping under base contacts
→ high Auger recombination→ low β.

Desire: high doping under contacts, lower doping elsewhere.

Regrowth processes enable this.

THz InP HBT Scaling Roadmap

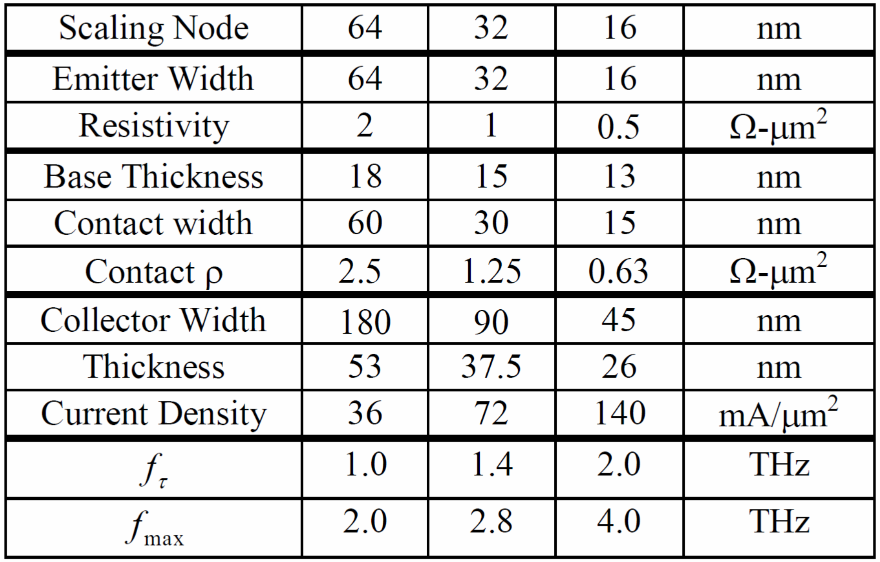
130nm node: 550GHz f_τ, 1100 GHz f_{max}

Are the 64 nm and 32nm nodes feasible ?

Key challenge: base contacts

Recent demonstration of <2 Ω–μm^2 contacts ***in HBT process flow***.

Longer term challenge :
decoupling doping under contacts vs. under base



Scaling Node	64	32	16	nm
Emitter Width	64	32	16	nm
Resistivity	2	1	0.5	Ω-μm^2
Base Thickness	18	15	13	nm
Contact width	60	30	15	nm
Contact ρ	2.5	1.25	0.63	Ω-μm^2
Collector Width	180	90	45	nm
Thickness	53	37.5	26	nm
Current Density	36	72	140	mA/μm^2
f_τ	1.0	1.4	2.0	THz
f_{max}	2.0	2.8	4.0	THz

86 GHz InP HBT Power Amplifier

UCSB/Teledyne

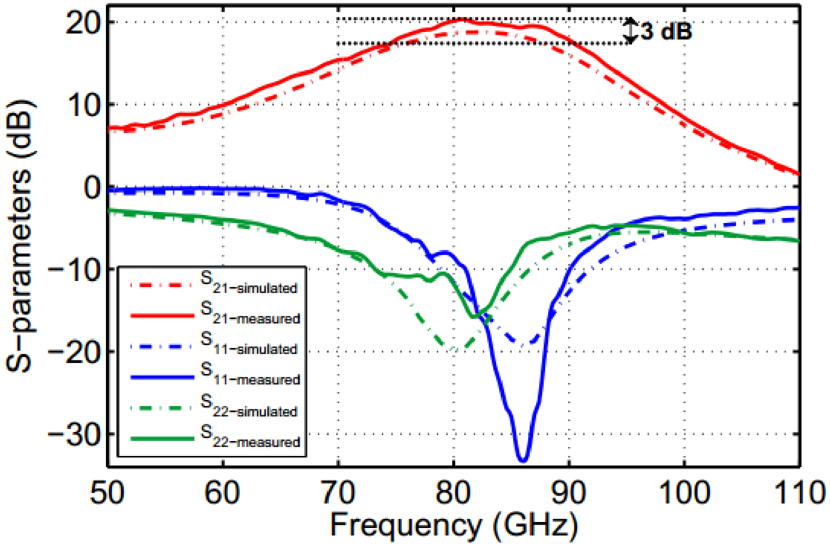
Gain: 20.4dB S21 Gain at 86GHz

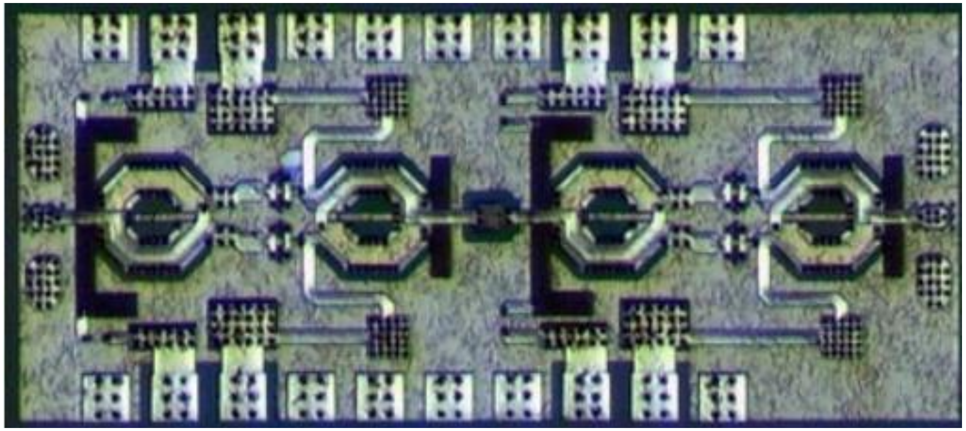
Saturated output power: 188mW at 86GHz

Output Power Density: 1.96 W/mm

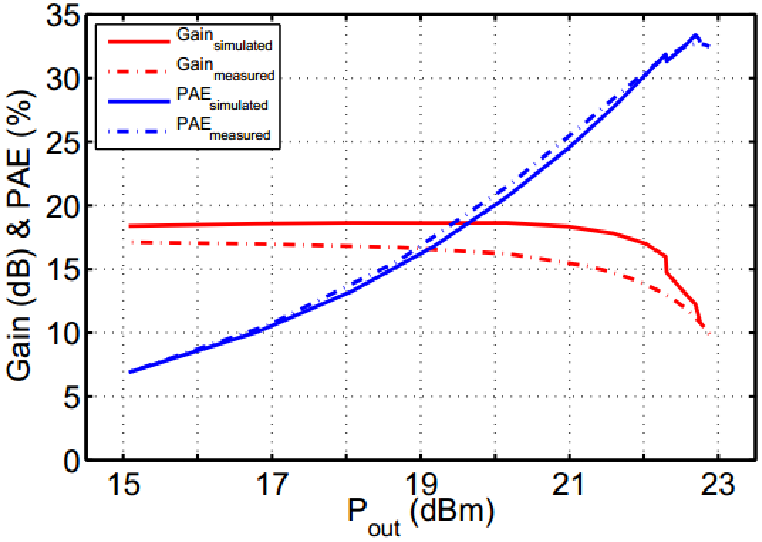
PAE: 32.8%

Technology: 250 nm InP HBT





1.4 mm x 0.60 mm



High W/mm, very small die

Park et al, JSSC, Oct. 2014 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6847236&tag=1

81 GHz InP HBT Power Amplifier

UCSB/Teledyne

Gain: 17.4dB S21 Gain at 81GHz

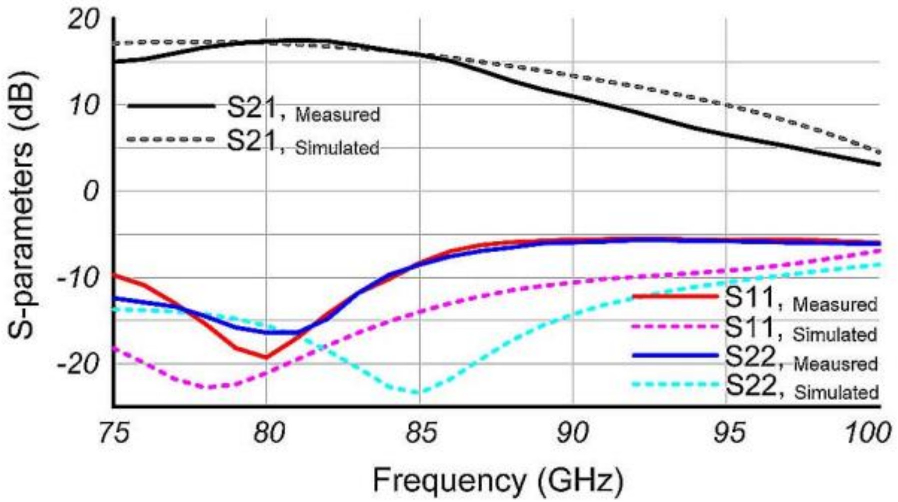
Saturated output power: 470mW at 81GHz

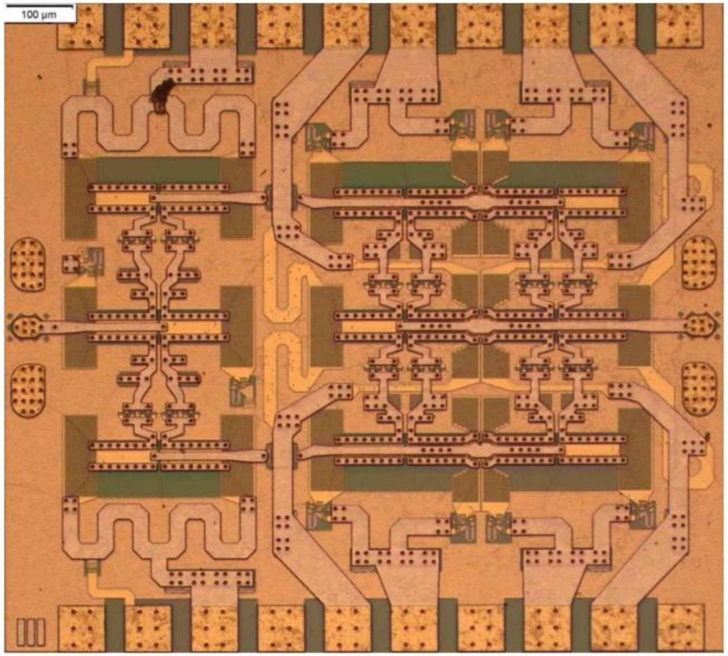
Output Power Density: 1.22 W/mm*

PAE: 23.4%

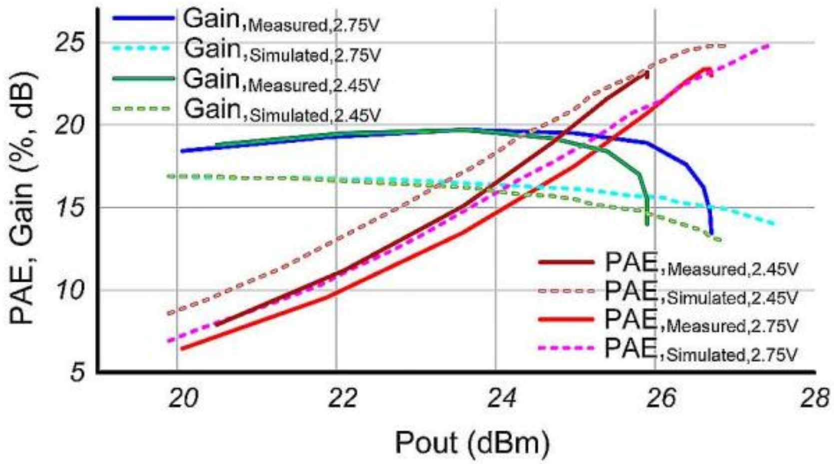
Power/(core die area): 1020W/mm²

Technology: 250 nm InP HBT





0.82mm x 0.82 mm



***design error: IC should have produced P_{sat}=700mW, ~2 W/mm**

High Power, very small die

Park et al, JSSC, Oct. 2014 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6847236&tag=1

214 GHz InP HBT Power Amplifier

UCSB/Teledyne

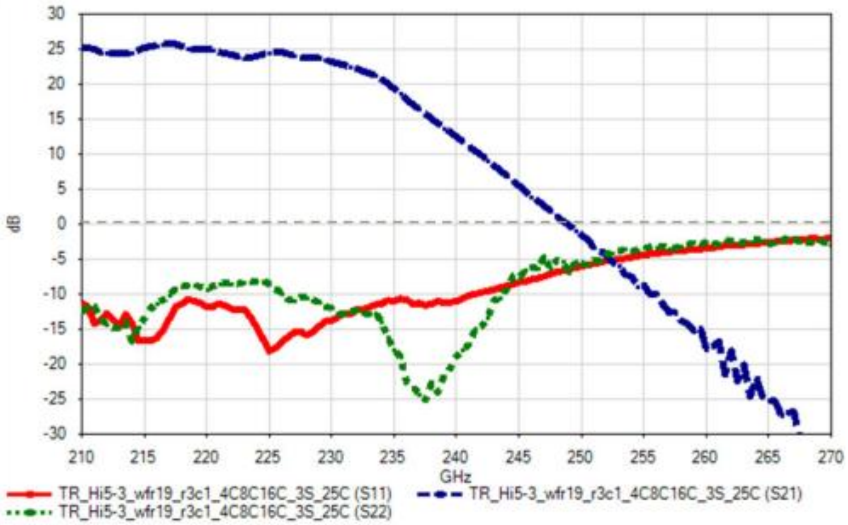
Gain: 25dB S21 Gain at 220GHz

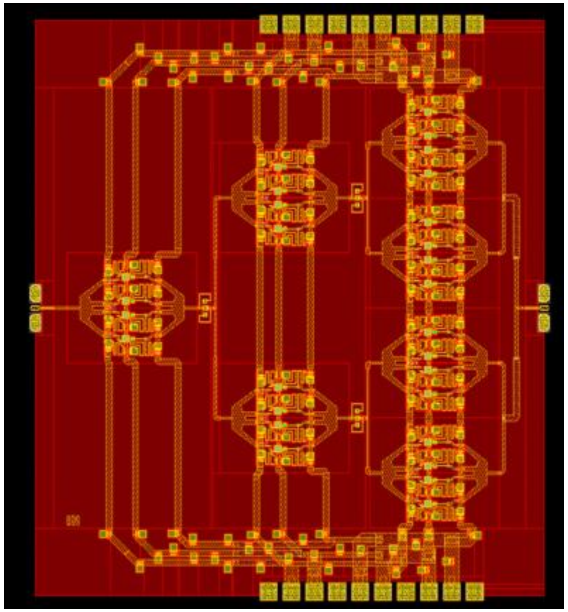
Saturated output power: 164mW at 214GHz

Output Power Density: 0.43 W/mm

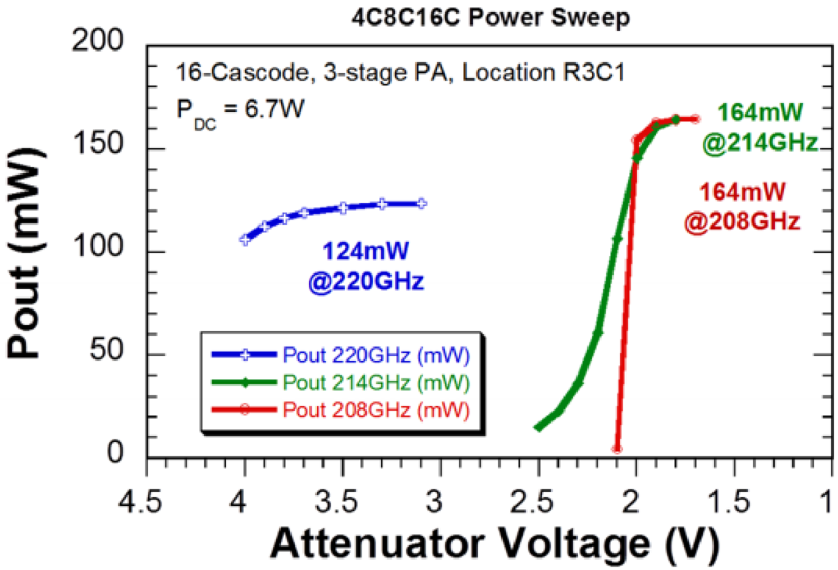
PAE: 2.4%

Technology: 250 nm InP HBT





(no die photo) 2.5mm x 2.1 mm

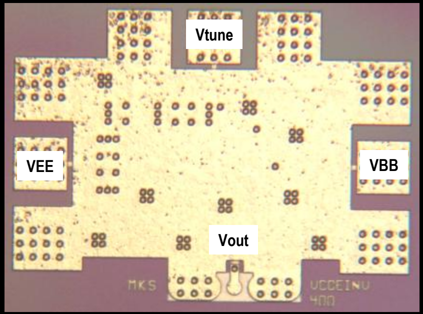


Reed et al, 2014 CSICS http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6659187&tag=1

InP HBT Integrated Circuits: 600 GHz & Beyond

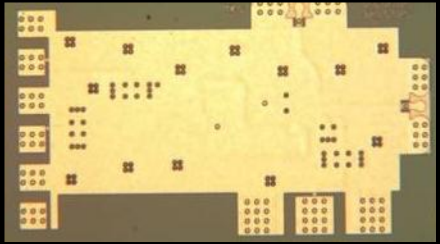
614 GHz fundamental VCO

M. Seo, TSC / UCSB



340 GHz dynamic frequency divider

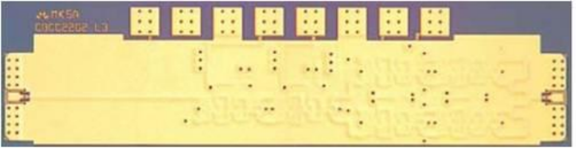
M. Seo, UCSB/TSC
IMS 2010



620 GHz, 20 dB gain amplifier

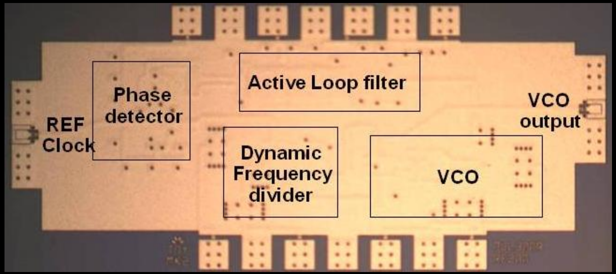
M Seo, TSC
IMS 2013

Not shown: 670 GHz amplifier:
J. Hacker, TSC
IMS 2013



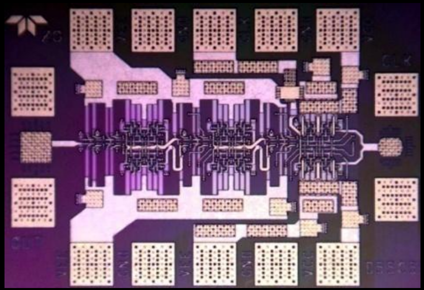
300 GHz fundamental PLL

M. Seo, TSC
IMS 2011



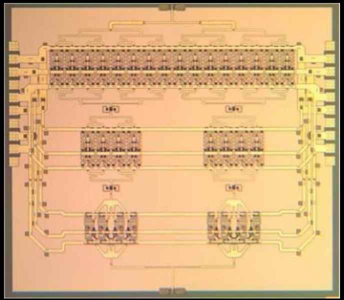
204 GHz static frequency divider (ECL master-slave latch)

Z. Griffith, TSC
CSIC 2010



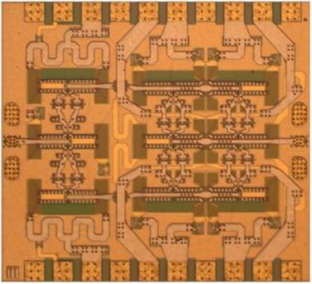
220 GHz 180 mW power amplifier

T. Reed, UCSB
CSICS 2013



81 GHz 470 mW power amplifier

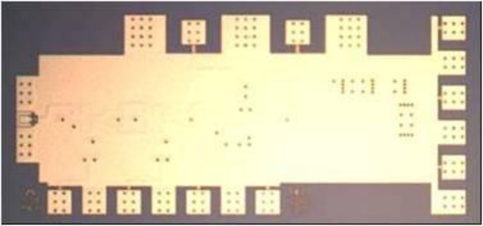
H-C Park UCSB
IMS 2014



Integrated 300/350GHz Receivers:

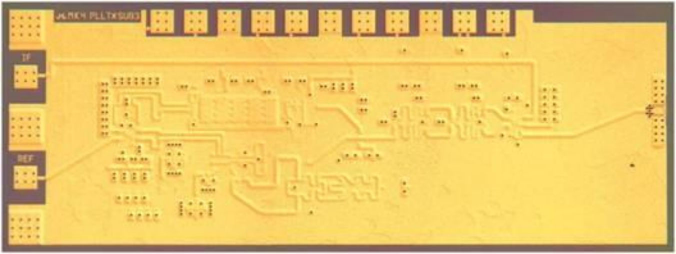
LNA/Mixer/VCO

M. Seo TSC



600 GHz Integrated Transmitter

PLL + Mixer
M. Seo TSC



Field-Effect Transistors

State of the Art (IMS 2014)

Recent Progress in Scaling InP HEMT TMIC Technology to 850 GHz

W.R. Deal, K. Leong, A. Zamora, V. Radisic and X.B. Mei
Northrop Grumman Corporation

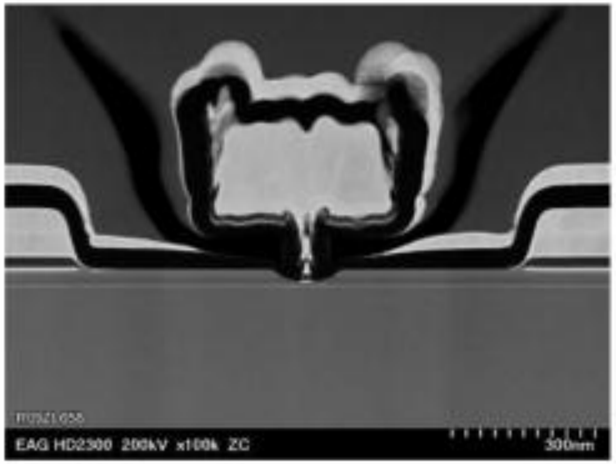
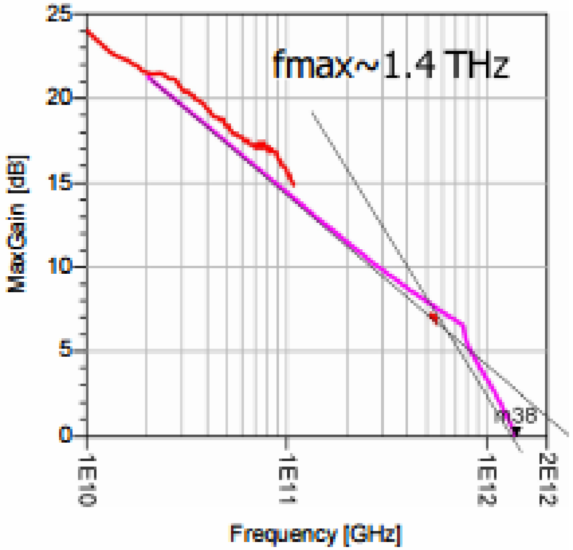


Fig. 1. A STEM image of a 30 nm InP HEMT.



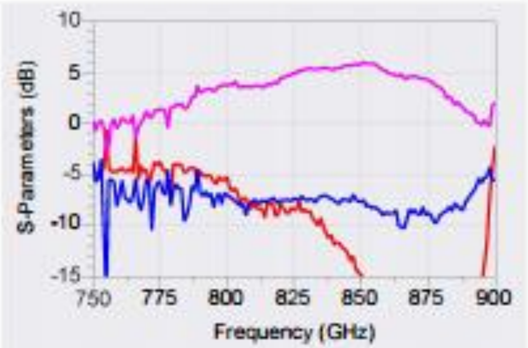


Fig. 7. Measured performance of 850 GHz amplifier. Magenta is s21,
s11 and blue is s22.

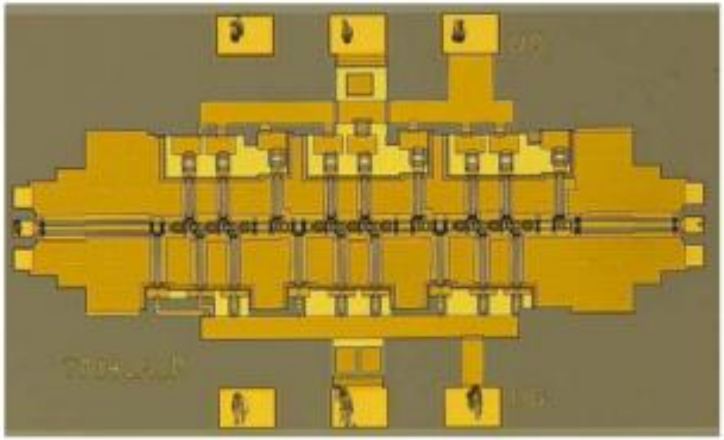
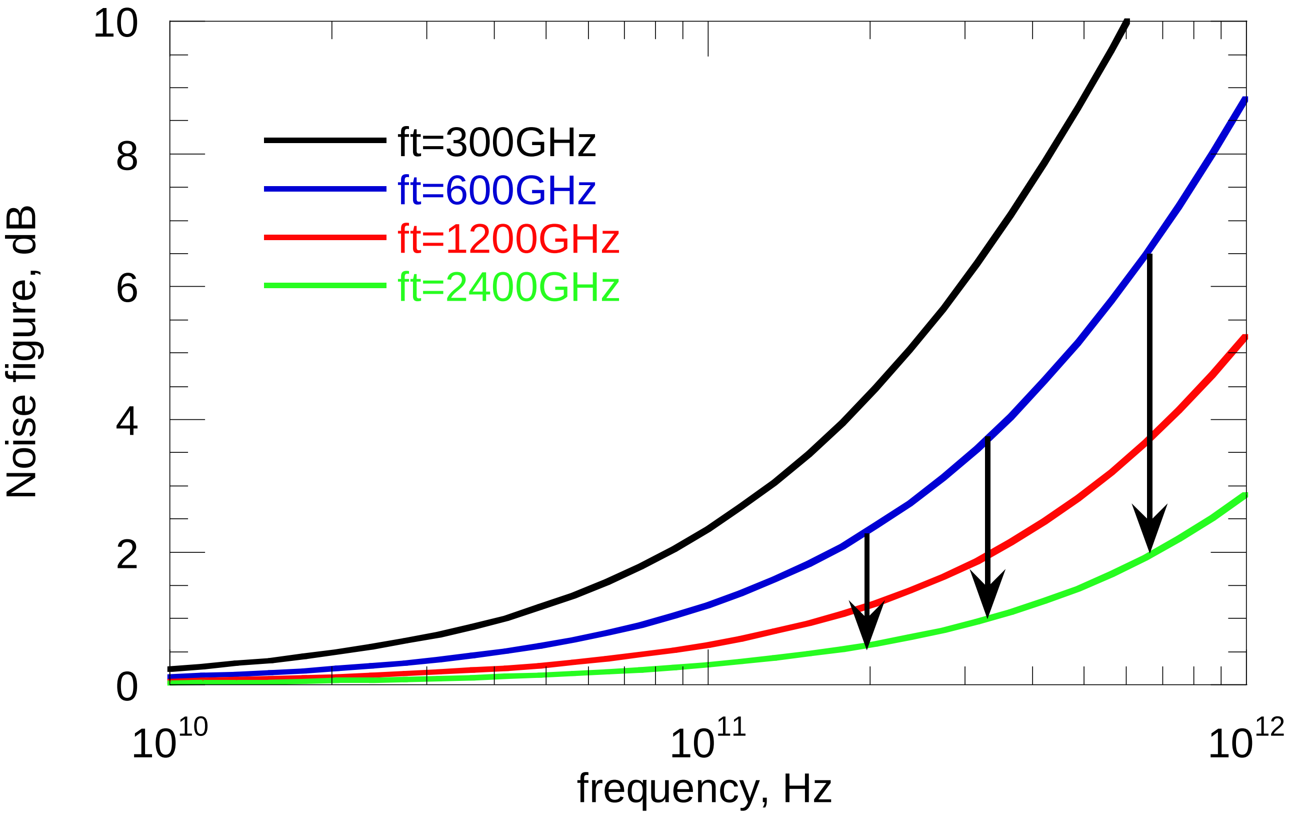


Fig. 6. Microphotograph of 850 GHz TMIC amplifier.

HEMTs: Key Device for Low Noise Figure



$$F_{\min} \approx 1 + 2\sqrt{g_m (R_s + R_g + R_i)\Gamma} \cdot \left(\frac{f}{f_\tau} \right) + 2 g_m (R_s + R_g + R_i)\Gamma \cdot \left(\frac{f}{f_\tau} \right)^2$$

$$\Gamma \approx 1$$

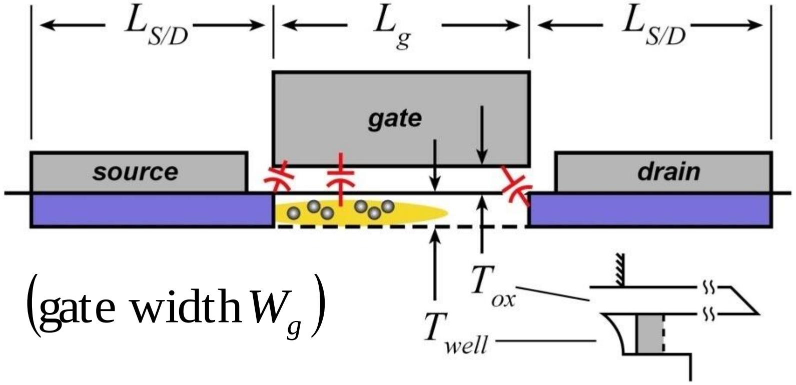
Hand-derived modified Fukui Expression, fits CAD simulation extremely well.

2:1 to 4:1 increase in f_τ → greatly improved noise @ 200-670 GHz.

Better range in sub-mm-wave systems; or use smaller power amps.

Critical: Also enables THz systems beyond 820 GHz

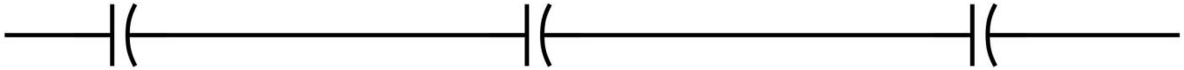
FET Design



$$C_{gd} \cong C_{gs,f} \cong \varepsilon W_g$$

$$g_m = C_{g-ch} \cdot (v / L_g)$$

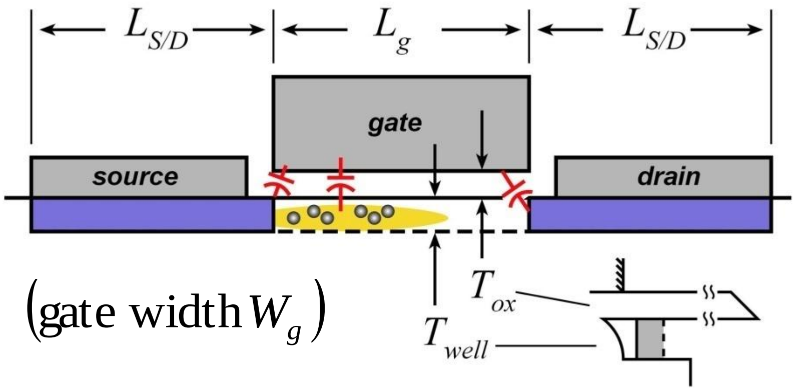
$$C_{g-ch} = \frac{L_g W_g}{T_{ox}/\varepsilon_{ox} + T_{\text{well}}/2\varepsilon_{\text{well}} + (q^2/\text{well state density})}$$



$$v \propto \begin{pmatrix} \text{voltage division ratio between} \\ \text{the above three capacitors} \end{pmatrix}^{-1/2} \cdot \frac{1}{\sqrt{\text{transport mass}}}$$

$$R_{DS} \approx L_g /(W_g v \varepsilon) \qquad R_S = R_D = \frac{\rho_{\text{contact}}}{L_{\text{S/D}} W_g}$$

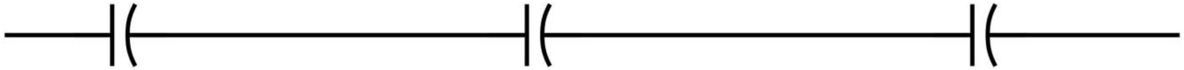
FET Design: Scaling



$$C_{gd} \cong C_{gs,f} \cong \varepsilon W_g$$

$$g_m = C_{g-ch} \cdot (v / L_g)$$

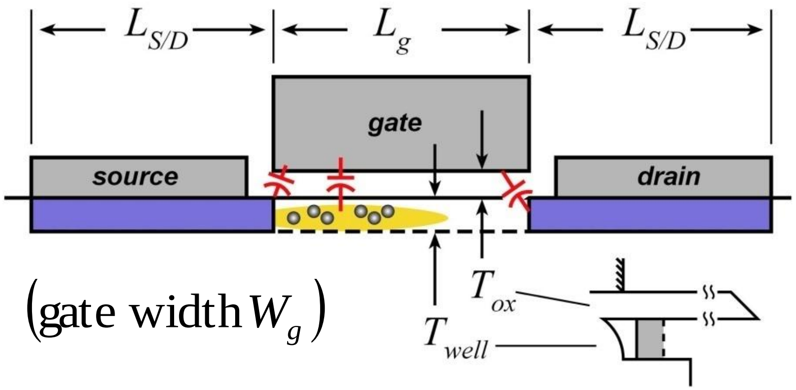
$$C_{g-ch} = \frac{L_g W_g}{T_{ox}/\varepsilon_{ox} + T_{\text{well}}/2\varepsilon_{\text{well}} + (q^2/\text{well state density})}$$



$$v \propto \begin{pmatrix}\text{voltage division ratio between}\\ \text{the above three capacitors}\end{pmatrix}^{-1/2} \cdot \frac{1}{\sqrt{\text{transport mass}}}$$

$$R_{DS} \approx L_g/(W_g v \varepsilon) \qquad R_S = R_D = \frac{\rho_{\text{contact}}}{L_{\text{S/D}} W_g}$$

FET Design: Scaling

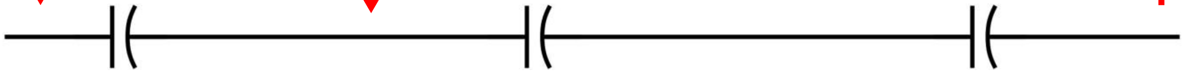


$$\text{2:1}\downarrow C_{gd} \cong C_{gs,f} \cong \varepsilon W_g \quad \text{2:1}\downarrow$$

constant

$$g_m = \overset{\text{2:1}\downarrow}{C_{g-ch}} \cdot (v/L_g) \quad \text{2:1}\downarrow$$

$$\overset{\text{2:1}\downarrow}{C_{g-ch}} = \frac{\text{2:1}\downarrow\, L_g W_g \,\downarrow\text{2:1}}{\underset{\text{2:1}\downarrow}{T_{ox}}/\varepsilon_{ox} + \underset{\text{2:1}\downarrow}{T_{\text{well}}}/2\varepsilon_{\text{well}} + (q^2/\underset{\text{2:1}\uparrow}{\text{well state density}})}$$



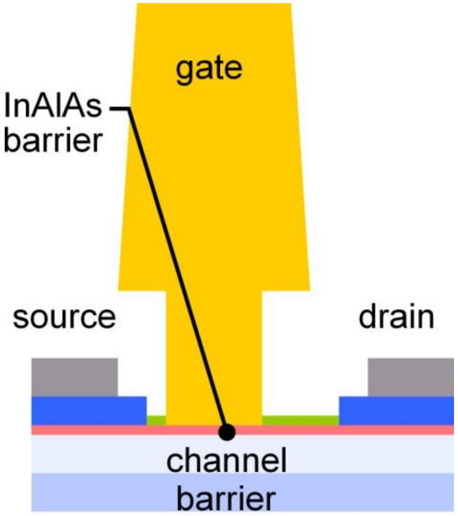
$$\underset{\text{constant}}{v} \propto \underset{\text{constant}}{\begin{pmatrix}\text{voltage division ratio between} \\ \text{the above three capacitors}\end{pmatrix}^{-1/2}} \cdot \frac{1}{\sqrt{\underset{\text{constant}}{\text{transport mass}}}}$$

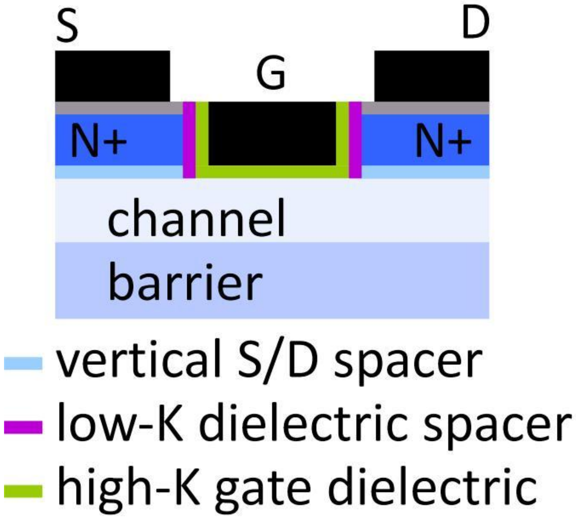
constant

$$R_{DS} \approx \underset{\text{2:1}\downarrow}{L_g}/(\underset{\downarrow\text{2:1}}{W_g} v\varepsilon)$$

$$\underset{\text{constant}}{R_S} = R_D = \frac{\rho_{\text{contact}}}{\text{2:1}\downarrow L_{S/D} W_g \downarrow\text{2:1}} \quad \text{4:1}\downarrow$$

Field-Effect Transistor Scaling Laws

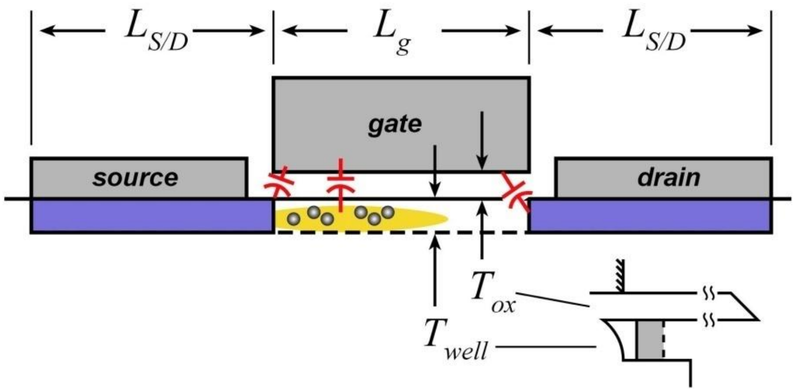




FET parameter	change
gate length	decrease 2:1
current density (mA/μm), g_m (mS/μm)	increase 2:1
transport effective mass	constant
channel 2DEG electron density	increase 2:1
gate-channel capacitance density	increase 2:1
dielectric equivalent thickness	decrease 2:1
channel thickness	decrease 2:1
channel density of states	increase 2:1
source & drain contact resistivities	decrease 4:1

fringing capacitance does not scale → linewidths scale as (1 / bandwidth)

Field-Effect Transistors No Longer Scale Properly



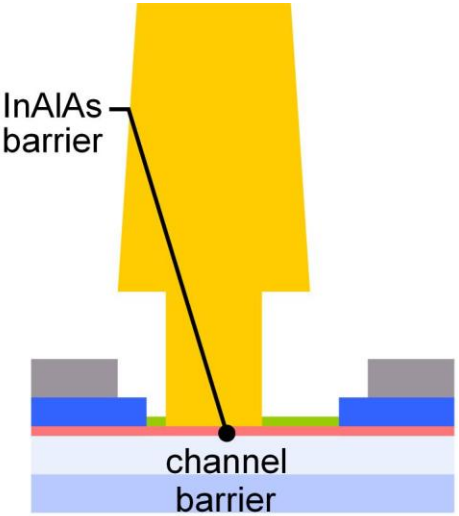
FET parameter	change
gate length	decrease 2:1
current density (mA/μm), g_m (mS/μm)	increase 2:1
transport effective mass	constant
channel 2DEG electron density	increase 2:1
gate-channel capacitance density	increase 2:1
dielectric equivalent thickness	decrease 2:1
channel thickness	decrease 2:1
channel density of states	increase 2:1
source & drain contact resistivities	decrease 4:1

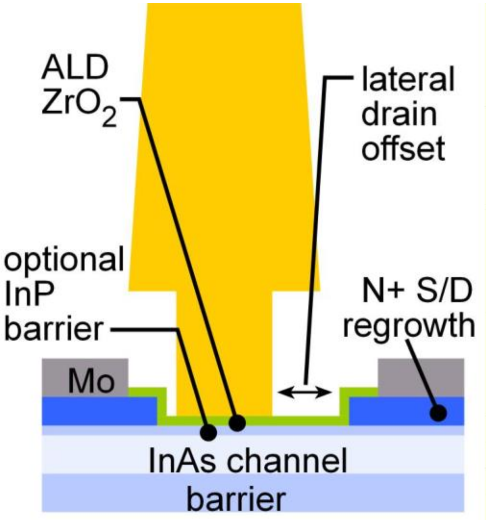
Gate dielectric can't be much further scaled.
Not in CMOS VLSI, not in mm-wave HEMTs

g_m/W_g (mS/μm) hard to increase→ C_{fringe}/g_m prevents f_τ scaling.

Shorter gate lengths degrade electrostatics→ reduced g_m/G_{ds}

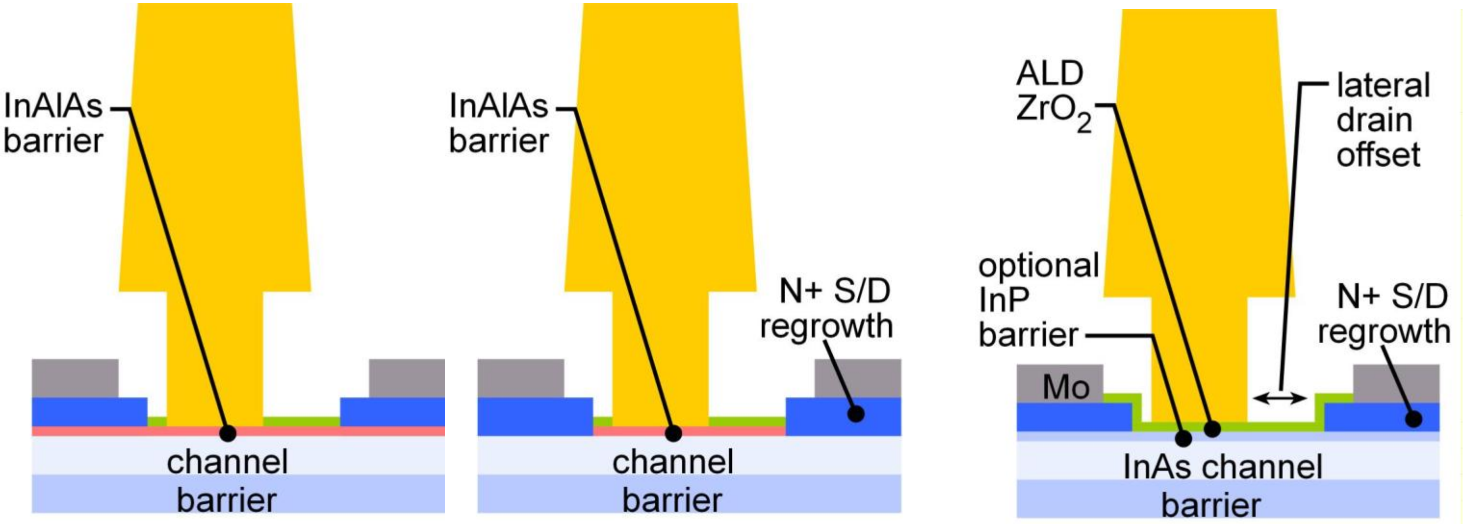
Scaling roadmap for InP HEMTs





gate length	36	18	9	nm
EOT	0.8	0.4	0.2	nm
well thickness	5.6	2.8	1.4	nm
effective mass	0.05	0.08	0.08	times m_0
# bands	1	1	1	--
S/D resistivity	150	74	37	Ω-μm
extrinisic g_m	2.5	4.2	6.4	mS/μm
on-current	0.55	0.8	1.1	mA/μm
f_τ	0.70	1.2	2.0	THz
$f_{\max}$	0.81	1.4	2.7	THz

Why THz HEMTs no longer scale; how to fix this



HEMTs: gate barrier also lies under S/D contacts → high S/D access resistance

S/D regrowth→ no barriers under contacts→ low $R_{S/D}$→ higher f_{max}, lower F_{min}

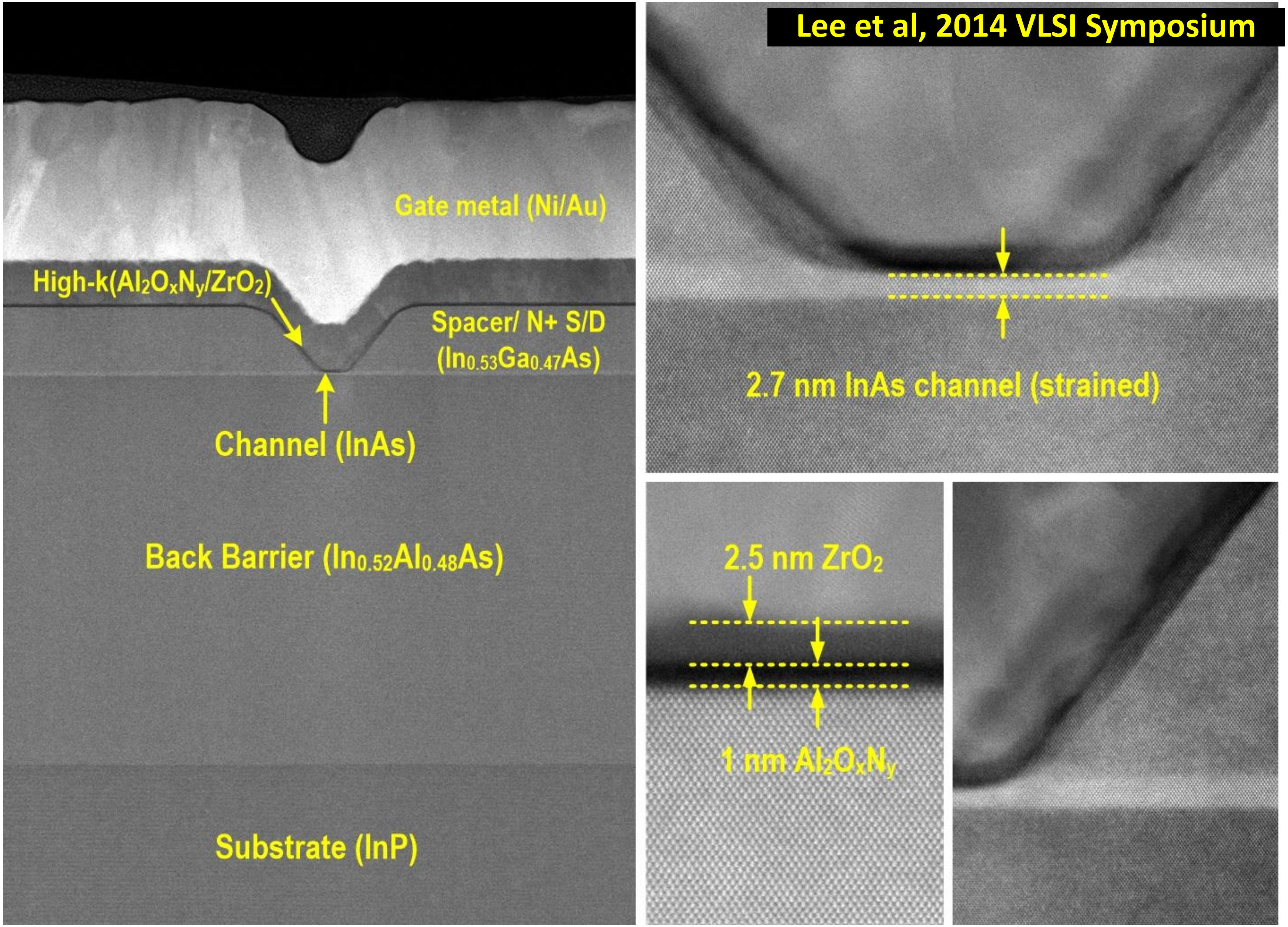
As gate length is scaled, gate barrier must be thinned for high g_m, low G_{ds}

HEMTs: High gate leakage when gate barrier is thinned→ cannot thin barrier

ALD high-K gate dielectrics→ ultra-thin→ improved g_m, G_{ds} , increased (f_τ,f_{max})

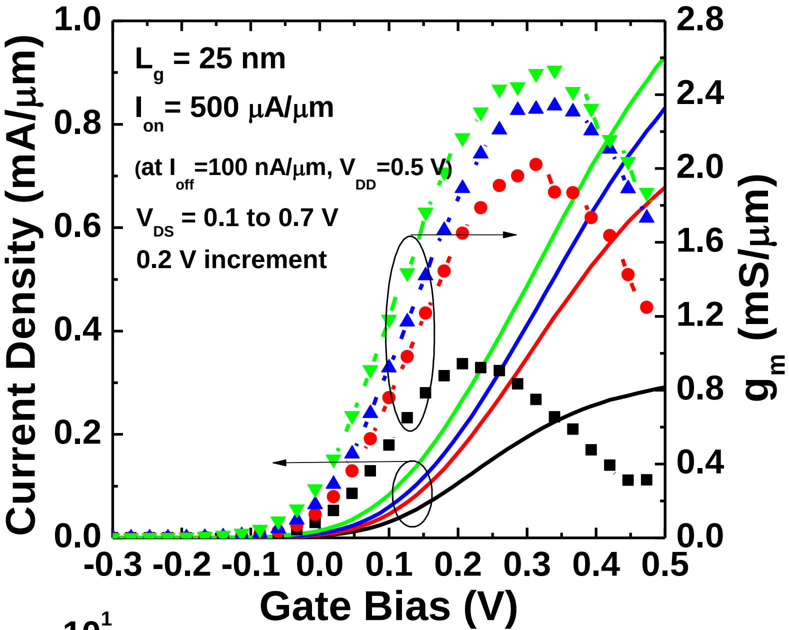
Solutions to key HEMT scaling challenges have been developed during the development of III-V MOS for VLSI.

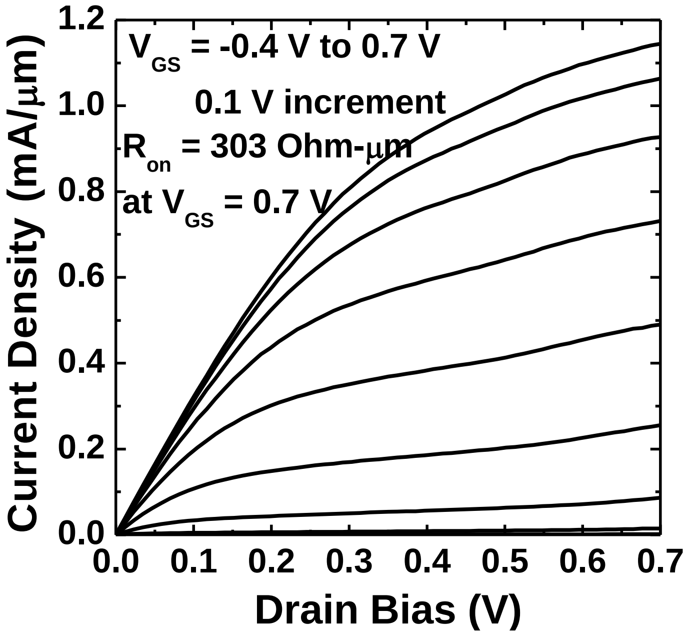
UCSB's Record VLSI-Optimized MOSFET @ 25nm L_g.

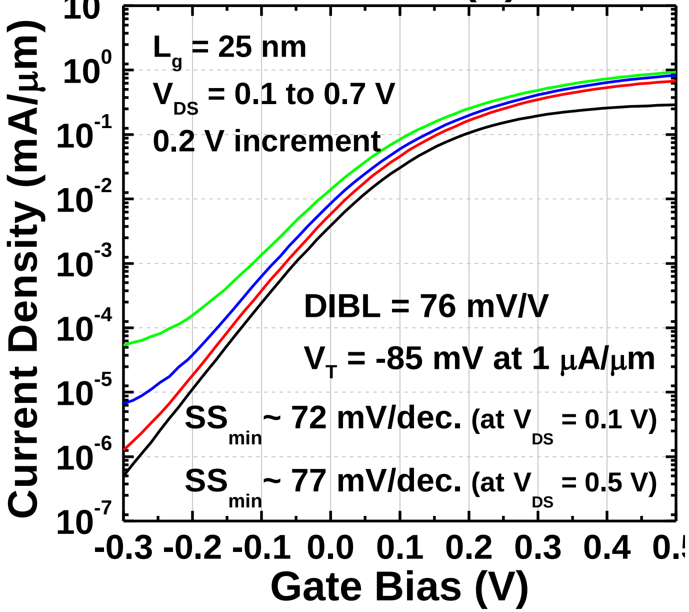


Lee et al, 2014 VLSI Symposium

UCSB's Record VLSI-Optimized MOSFET @ 25nm L_g.





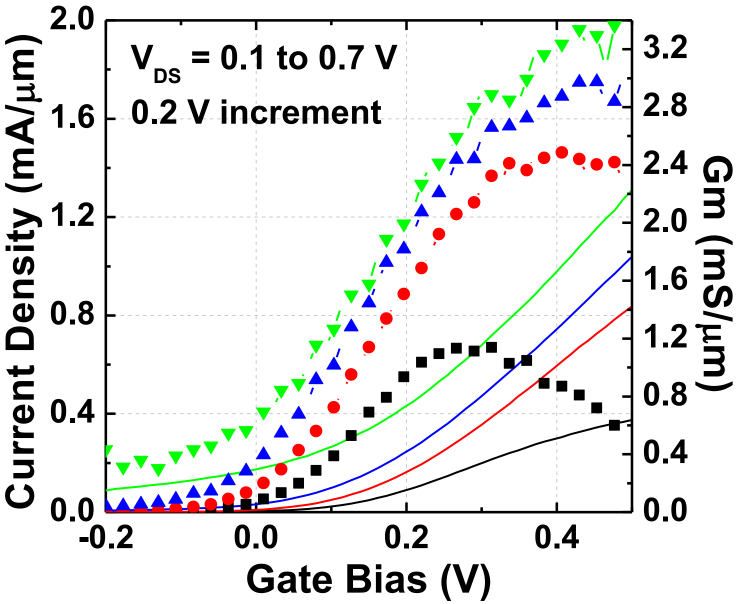


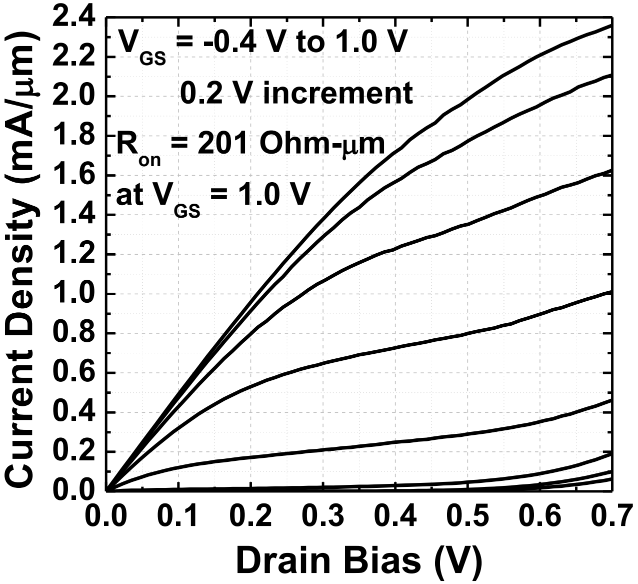
- ~2.4 mS/μm Peak g_m at V_{DS}=0.5 V
- ~300 Ohm-μm on-resistance at V_{GS}=0.7 V
- 77 mV/dec Subthreshold Swing at V_{DS}=0.5 V, 76 mV/V DIBL at 1 μA/μm
- 0.5 mA/μm I_{on} at I_{off}=100 nA/μm and V_{DD}=0.5 V
- 61 mV/dec subthreshold swing @1 μm L_g

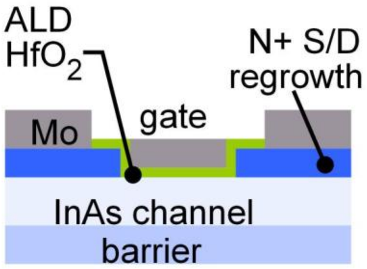
Lee et al, 2014 VLSI Symposium

High Transconductance III-V MOSFETs

Lee et al, EDL, June 2014







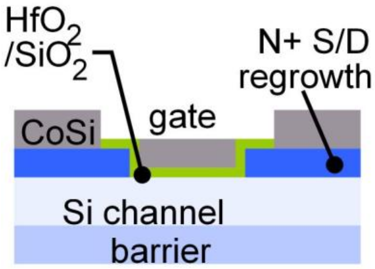
High g_m, with low G_{DS}, is critical for THz FETs

Here:
18nm gate length, 5nm InAs channel → 3mS/µm g_m.

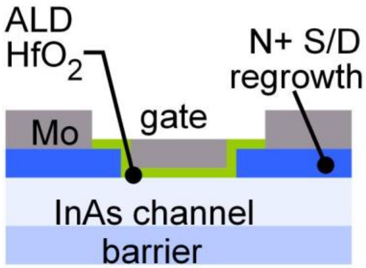
These FETs have large access resistance from non-self-aligned contacts;
so g_m can be readily increased.

Future: shorter gates, thinner channels, better dielectrics better contacts
→ higher g_m.

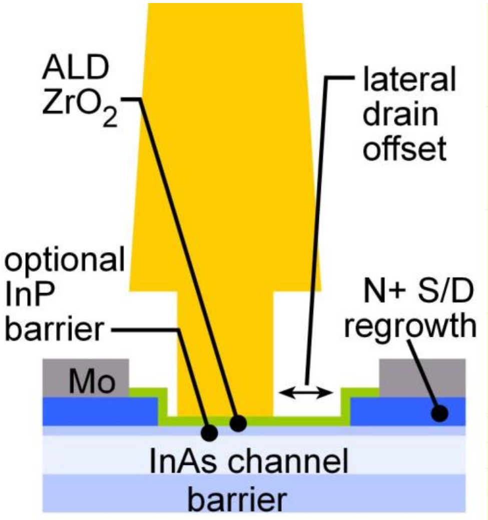
THz III-V MOS: Not the same as VLSI III-V MOS



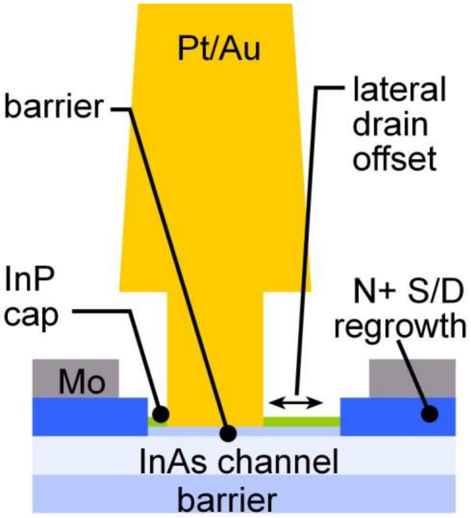
UTB Si MOS



UTB III-V MOS



III-V THz MOS/HEMT



III-V THz HEMT

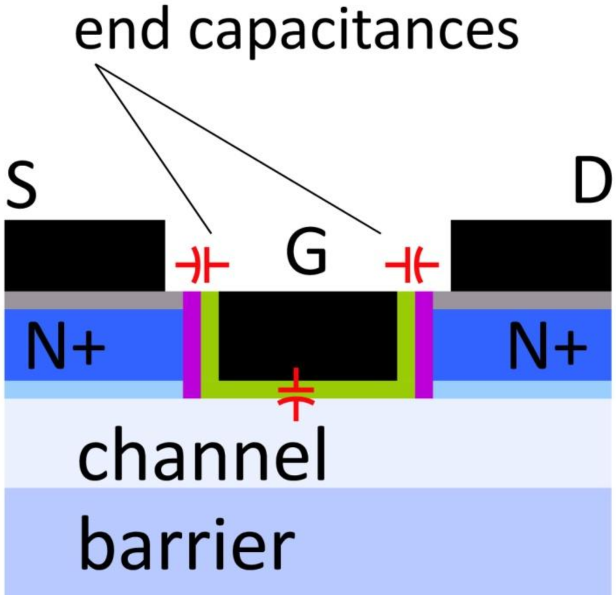
III-V MOS has a reasonable chance of use in VLSI at the 7nm node
These will *not* be THz devices

The real mm-wave / VLSI distinction:
Device geometry optimized for high-frequency gain (THz)
vs. optimized for small footprint & high DC on/off ratio (VLSI).

mm-wave / THz devices:
minimize overlap capacitances, drain offset for low C_{gd} & G_{ds},
thicker channels optimized for g_m, T-gates for low resistance

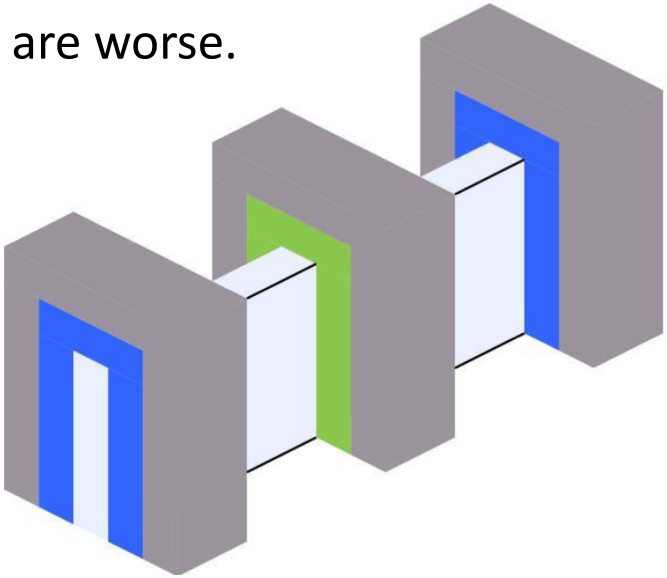
Prospects for Higher-Bandwidth CMOS VLSI

Recall:
Gate-dielectric can't scale much further.
That stops g_m (mS/μm) from increasing.
(end capacitance)/g_m limits achievable f_τ.



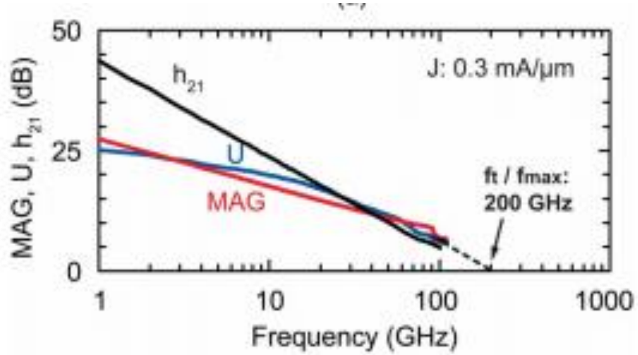
Also:
Given fixed dielectric EOT,
G_{ds} degrades with scaling.

FinFETs have better electrostatics,
hence better g_m/G_{ds}...
But in present technologies the end capacitances are worse.



And W via resistances reduce the gain

Inac et al, CSICS 2011 (45nm SOI CMOS)



InP Field-Effect-Transistor Scaling Roadmap

2-3 THz InP HEMTs are Feasible.

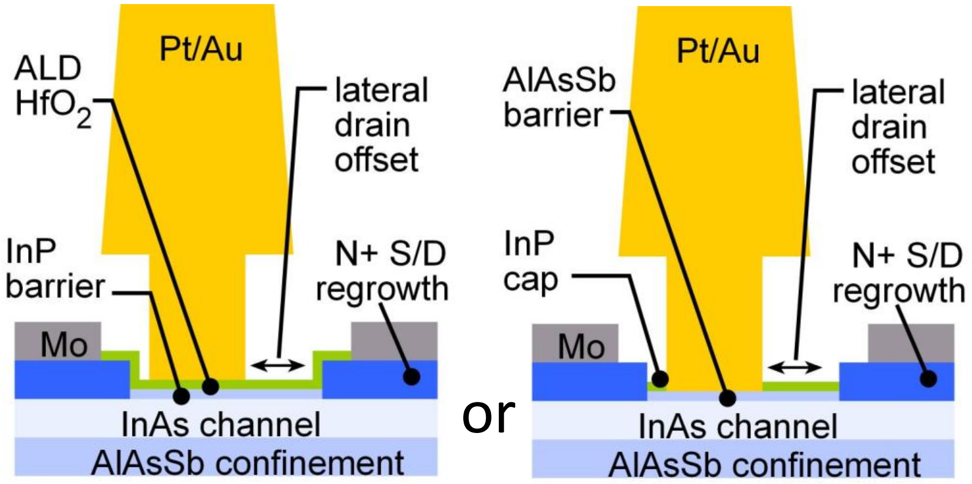
2 THz FETs realized by:

Ultra low resistivity source/drain

High operating current densities

Very thin barriers & dielectrics

Gates scaled to 9 nm junctions



MOSFET

high-barrier HEMT

gate length	36	18	9	nm
EOT	0.8	0.4	0.2	nm
well thickness	5.6	2.8	1.4	nm
effective mass	0.05	0.08	0.08	times m_0
# bands	1	1	1	--
S/D resistivity	150	74	37	Ω-μm
extrinisic g_m	2.5	4.2	6.4	mS/μm
on-current	0.55	0.8	1.1	mA/μm
f_τ	0.70	1.2	2.0	THz
$f_{\max}$	0.81	1.4	2.7	THz

Impact:
Sensitive, low-noise receivers from 100-1000 GHz.

3 dB less noise →
need 3 dB less transmit power.

Conclusions

Roadmap for High-Frequency Transistors

Beware of physics-free roadmaps
- 20% improvement /year extrapolations are meaningless.
- Real transistors are approaching scaling limits.
- VLSI transistors are optimized for density & digital, not RF.
- Lower standby power processes are slower RF processes.

Bandwidths of Si CMOS VLSI have leveled off.

There is market for application-specific high-frequency transistors.
- LNAs, PAs, front-ends generally.
- Just like cell phones today.

InP HBTs & HBTs have perhaps 2-3 scaling generations left.
- Doubling of bandwidth, perhaps a little more.
- Process technology development is getting quite hard.