THz & nm Transistor Electronics: It's All About The Interfaces.

M. Rodwell, Art Gossard
University of California, Santa Barbara

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

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 can the materials growth community help ?

work on interfaces (contacts and gate dielectrics) !

Guidance of utility of other device structures / features

nanowire pillar devices

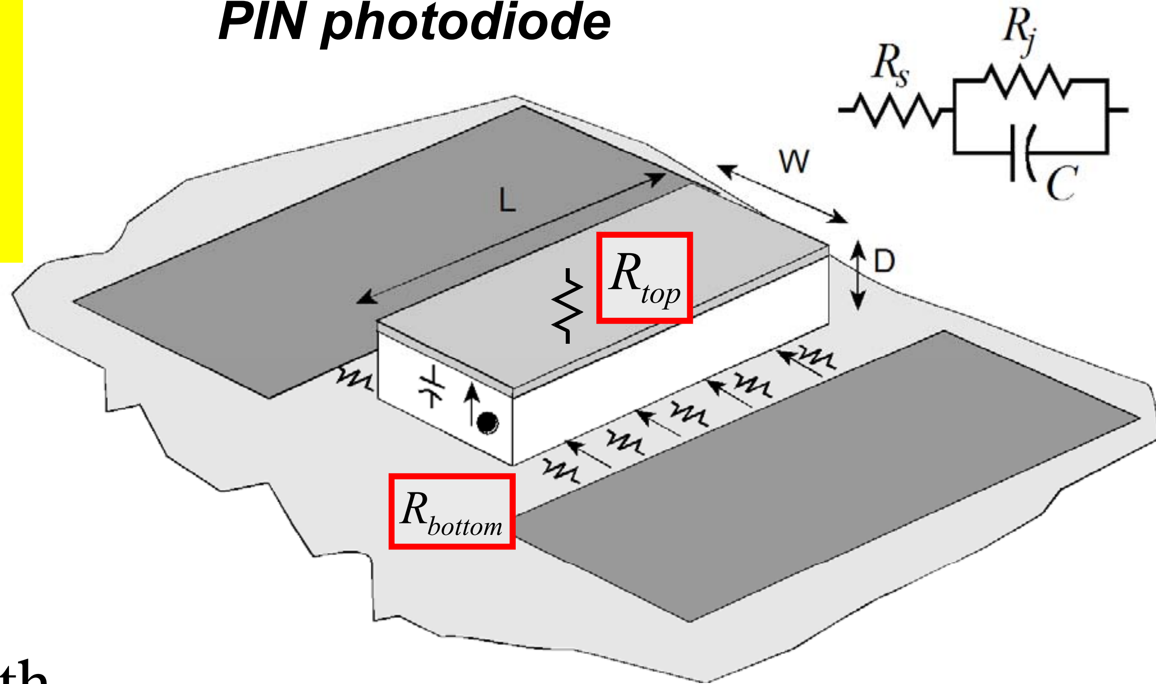
access resistances & capacitances

relevance and irrelevance of mobility

THz & nm Semiconductor Device Design...

... is scaling

Frequency Limits and Scaling Laws of (most) Electron Devices



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

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

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

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

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

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

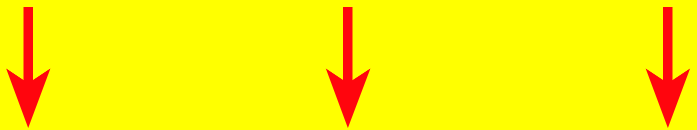
To double bandwidth,

reduce thicknesses 2:1

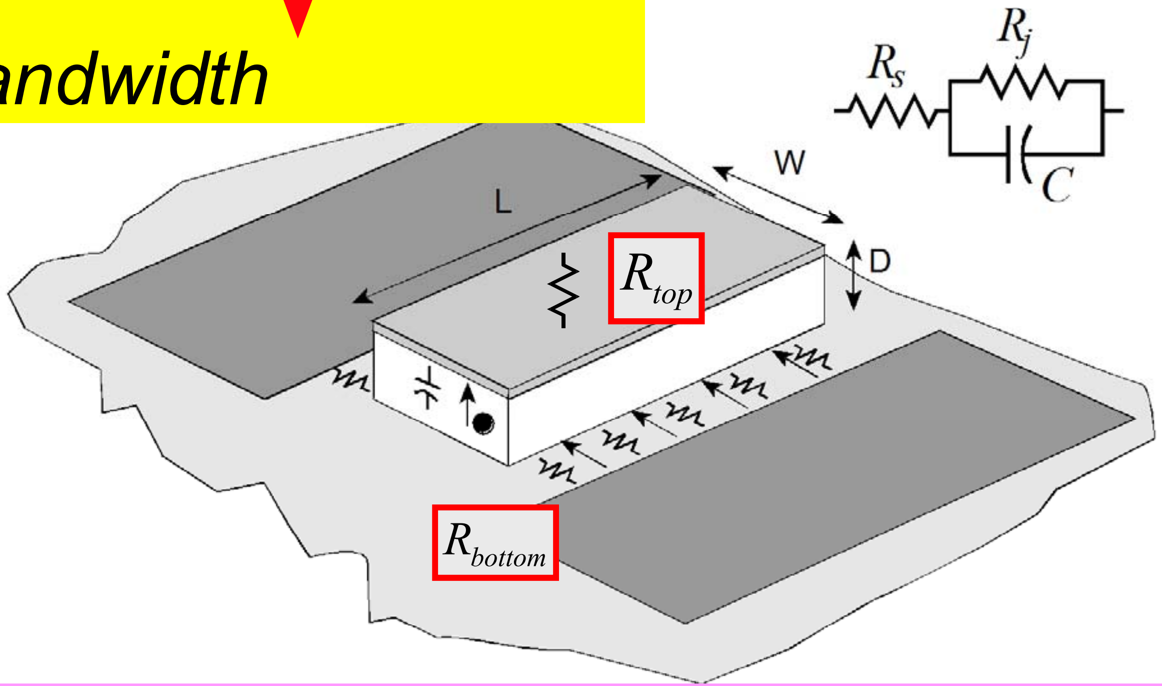
reduce width 4:1, keep constant length

current density has increased 4:1

resistance capacitance transit time



device bandwidth



applies to almost all semiconductor devices:

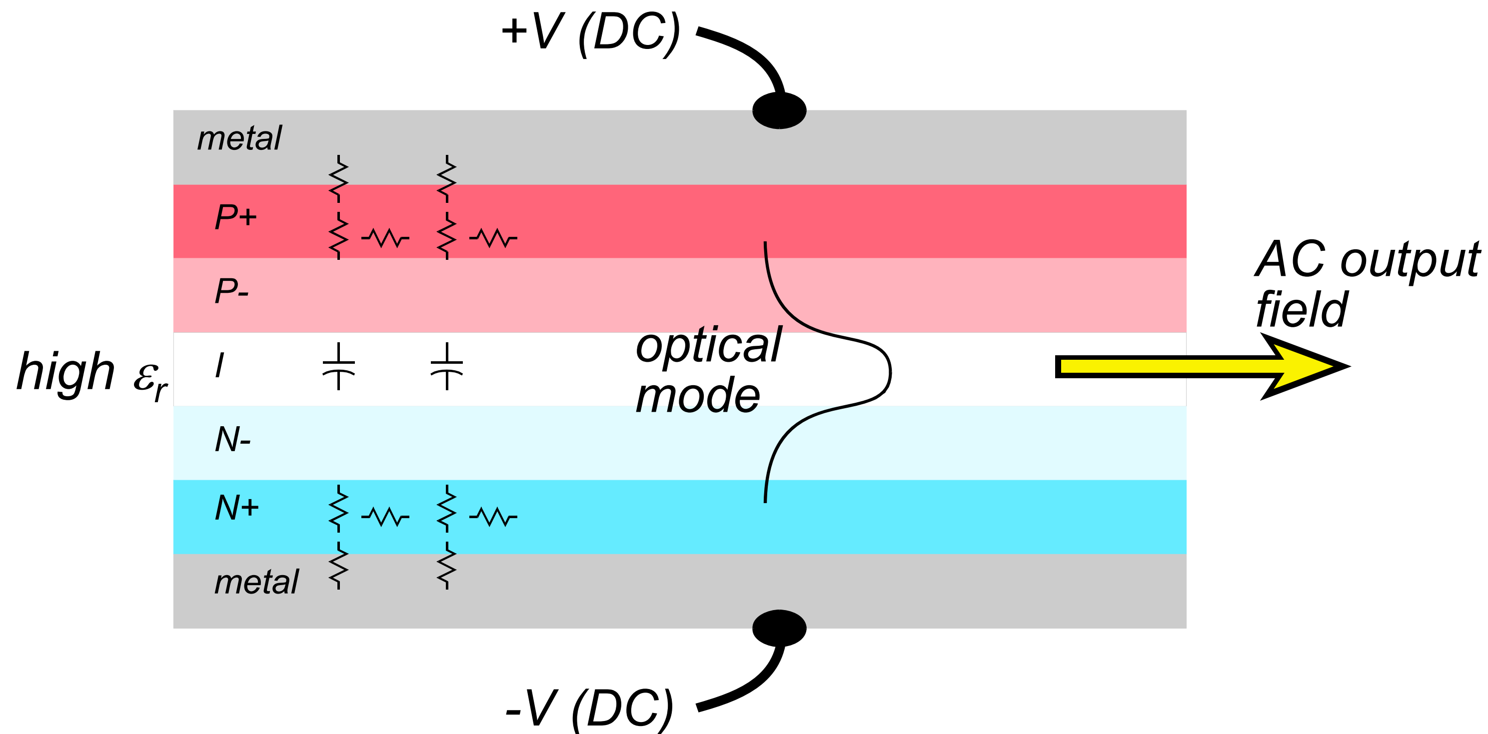
transistors: BJTs & HBTs, MOSFETS & HEMTs,
Schottky diodes, photodiodes, photo mixers, RTDs, ...

high current density,
low resistivity contacts,
epitaxial & lithographic scaling

THz semiconductor devices

FETs only: high $\varepsilon_r \varepsilon_o$/D dielectrics

Why aren't semiconductor lasers R/C/τ limited ?



dielectric waveguide mode confines AC field away from resistive bulk and contact regions.

AC signal is not coupled through electrical contacts

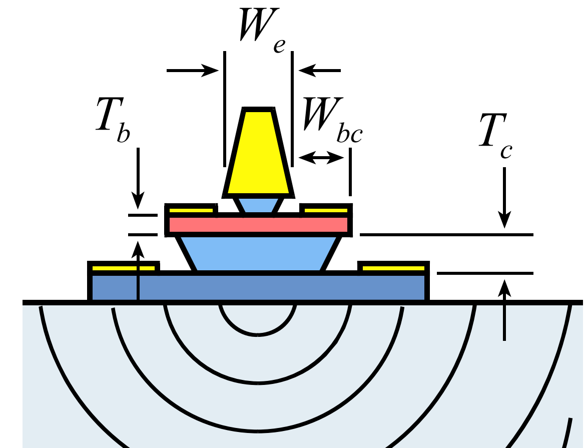
dielectric mode confinement is harder at lower frequencies

Bipolar Transistor Design

Bipolar Transistor Design is Simple

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

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



(emitter length L_E)

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

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

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

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

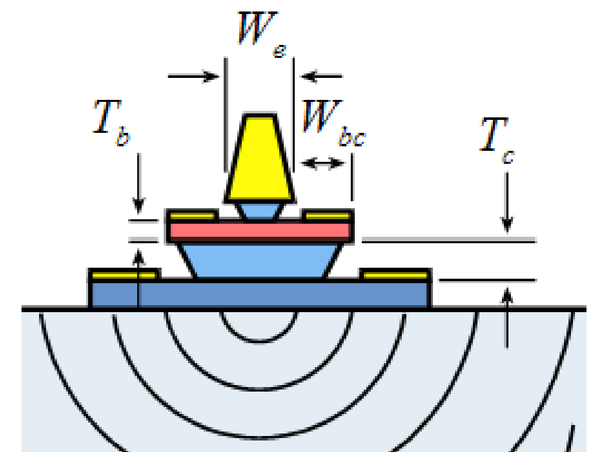
$$\Delta T \propto \frac{P}{L_E} \left[1 + \ln\left(\frac{L_e}{W_e} \right) \right]$$

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



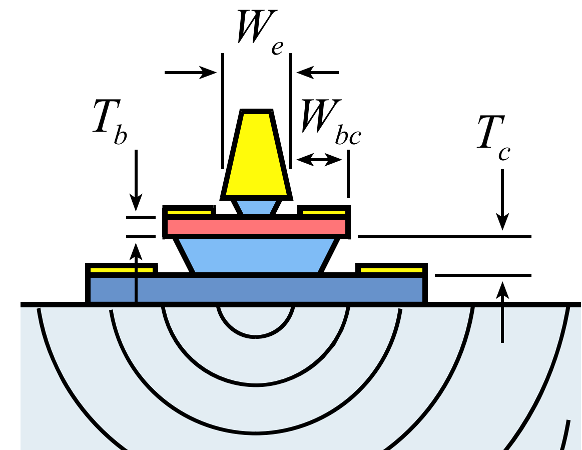
(emitter length L_E)

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*



(emitter length L_E)

$C_{cb} \propto A_c / T_c$ → *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)*

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

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 ✓

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

→ *reduce base contact resistivity 4:1*

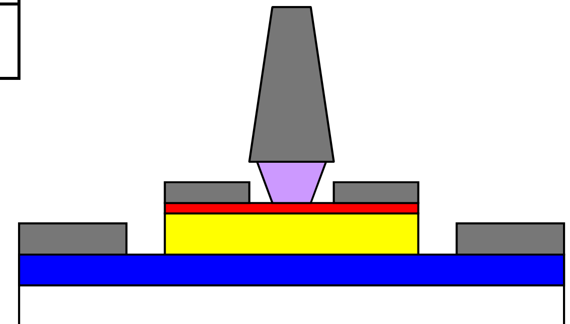
reduce widths 2:1 & reduce length 2:1 → constant R_{bb} ✓
reducing widths 4:1, keep constant length → reduced R_{bb} ✓✓

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

Bipolar Transistor Scaling Laws

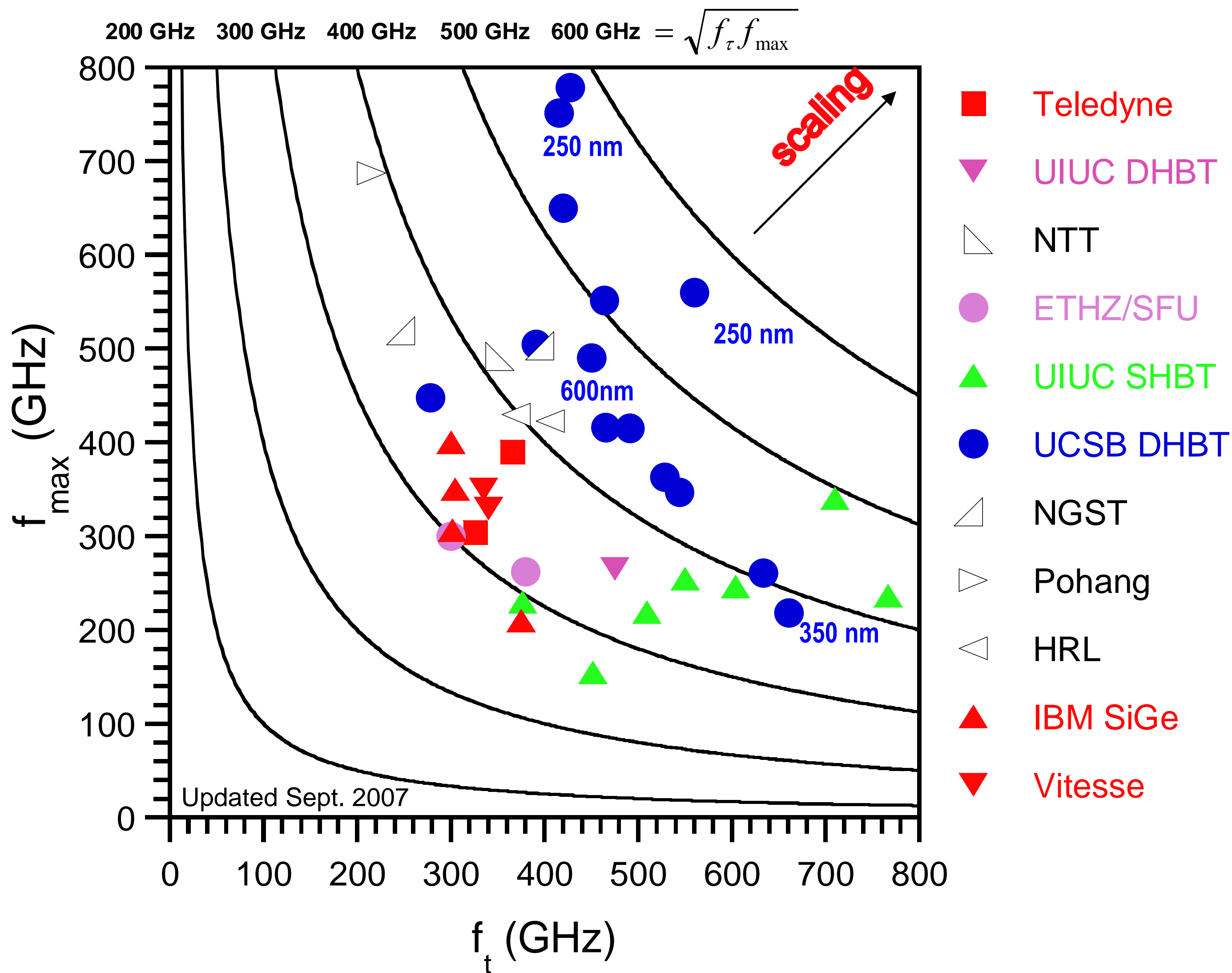
Changes required to double transistor bandwidth:

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 :

<u>power amplifiers</u>:

PAE, associated gain, mW/μm

<u>low noise amplifiers</u>:

F_{min}, associated gain,

<u>digital</u>:

f_{clock}, hence

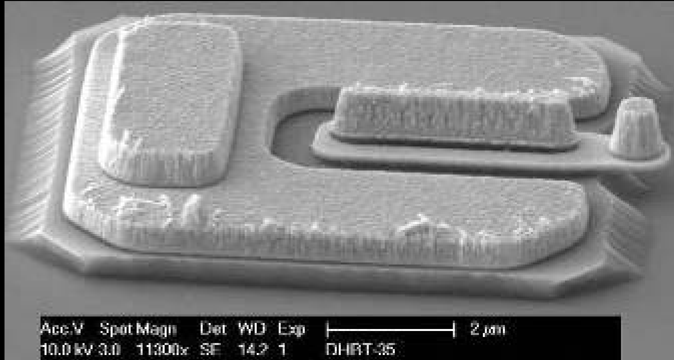
$(C_{cb}\Delta V / I_c)$,

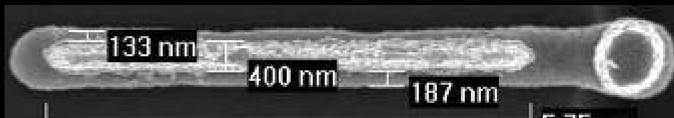
$(R_{ex} I_c / \Delta V)$,

$(R_{bb} I_c / \Delta V)$,

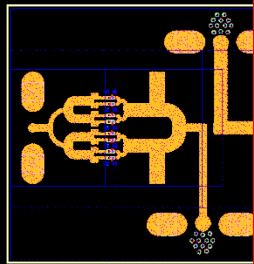
$(\tau_b + \tau_c)$

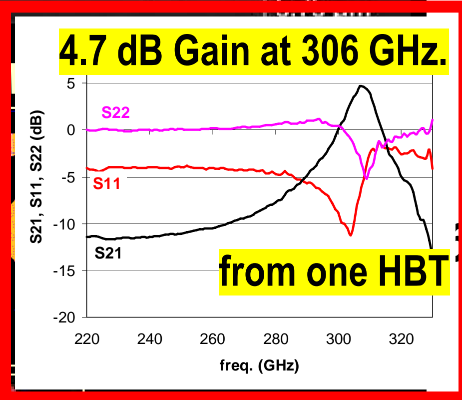
256 nm Generation InP DHBT

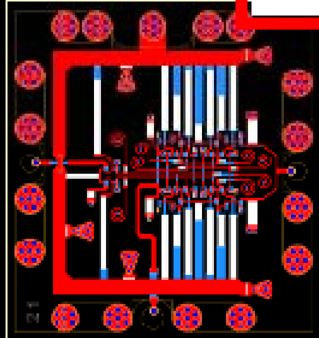




340 GHz, 70 m



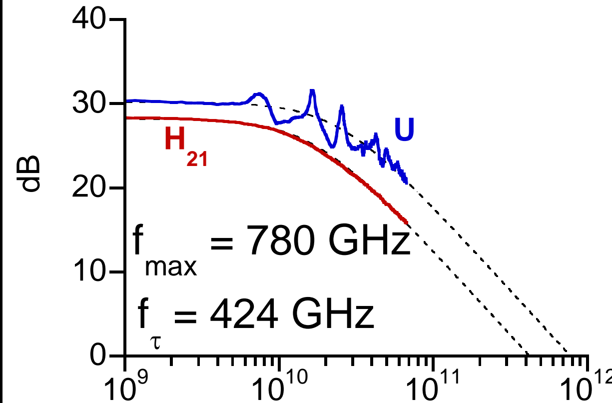


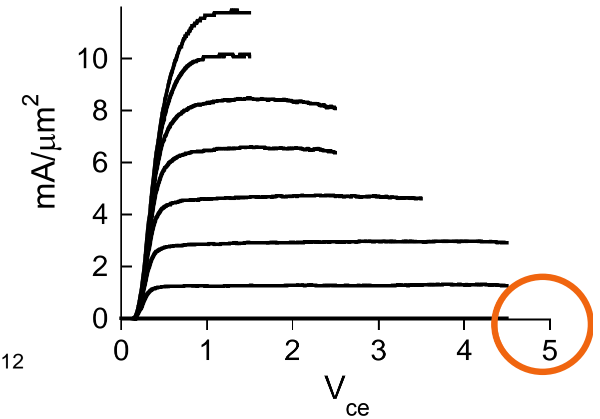


200 GHz master-slave latch design

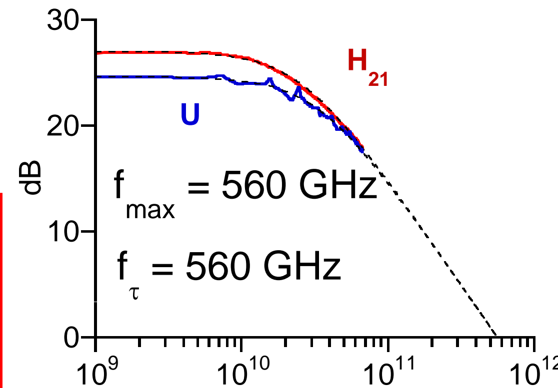
Z. Griffith, E. Lind, J. Hacker, M. Jones

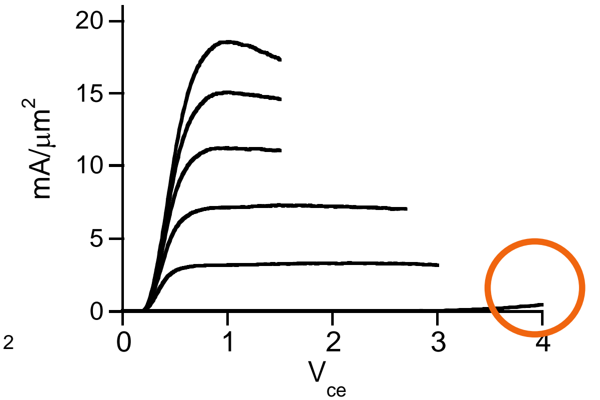
150 nm thick collector



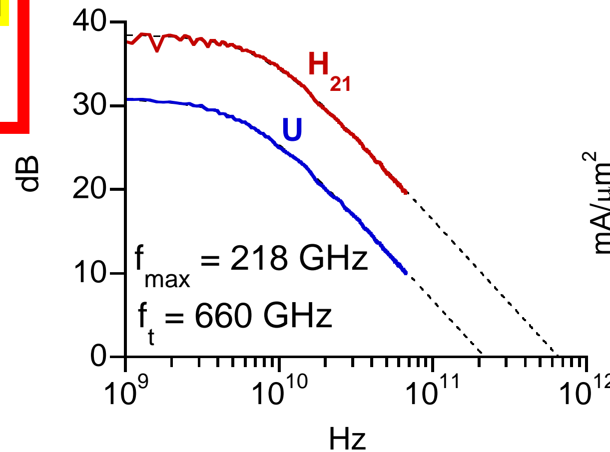


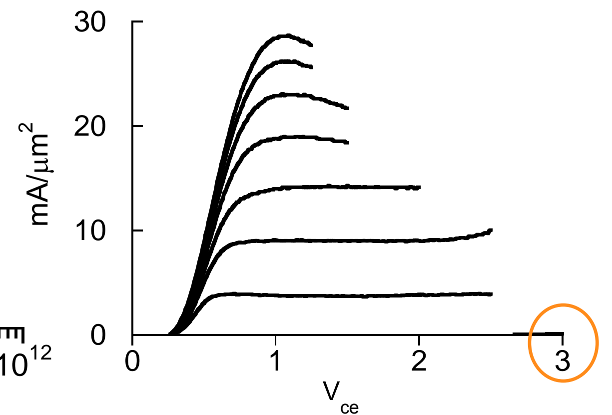
70 nm thick collector





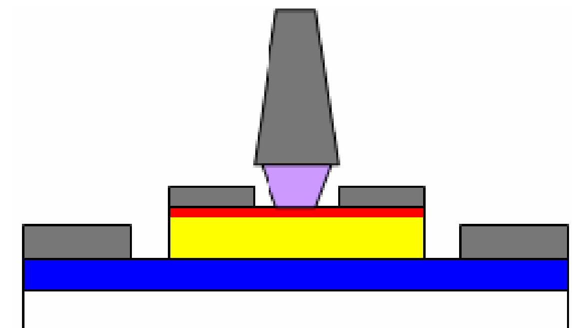
0 nm thick collector





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** $\Omega \cdot \mu m^2$ 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** $mA/\mu m^2$ 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 2:1 divider	150	240	330	480	**660** GHz



How Can Material Scientists Help ?

To build a 5-THz bipolar Transistor...

...we need 0.25 Ω-μm^2 Ohmic contacts,

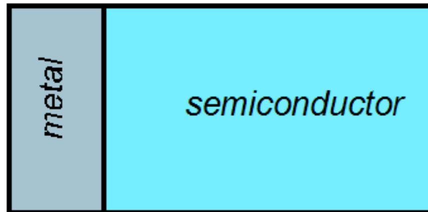
& these must be stable at 300 mA/μm^2.

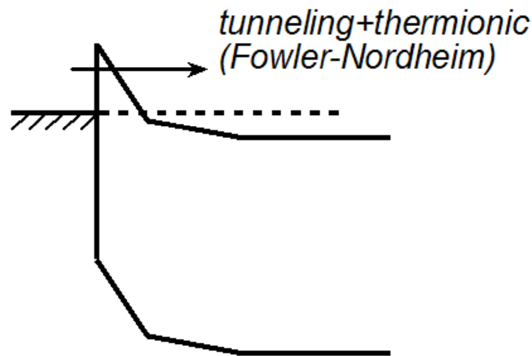
...Can you help ?

Ohmic Contacts

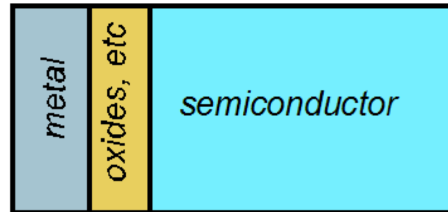
Ex-Situ Ohmic Contacts are a Mess

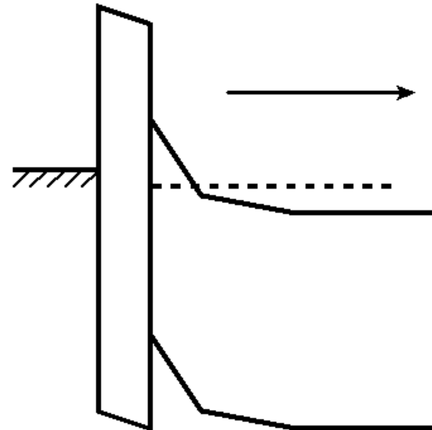
textbook contact



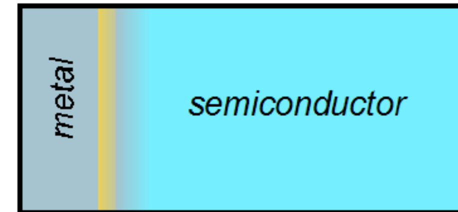


with surface oxide





with metal diffusion



Surface contaminated by semiconductor oxides

On InGaAs surface: Indium and Gallium Oxides, elemental As

Metals Interdiffuse with Semiconductor

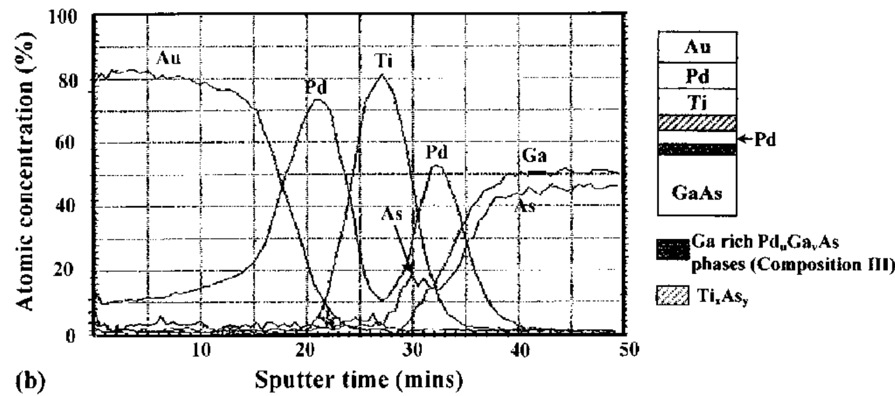
TiPtAu contacts: Ti diffusion. Pt contacts: reaction. Pd contacts: reaction

Interface is degraded → poor conductivity

Interface is badly-controlled→ hard to understand→ hard to improve

Our HBT Base Contacts Today Use Pd or Pt to Penetrate Oxides

TEM : Lysczek, Robinson, & Mohney, Penn State
Sample: Urteaga, RSC



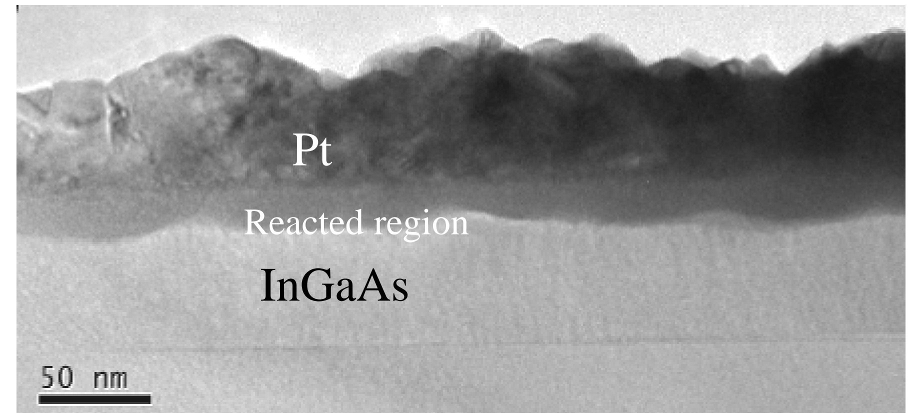
Wafer first cleaned in reducing

Pd & Pt react with III-V semiconductor

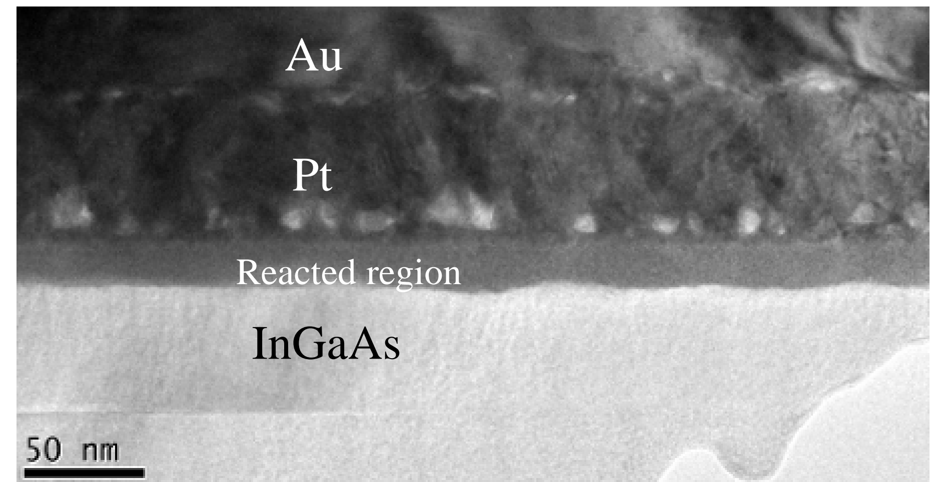
Penetrate surface oxide

Provide ~5 Ω-μm^2 resistivity (InGaAs base, $8*10^{19}/cm^3$)

reaction depth is a problem for HBT base



Pt Contact after 4hr 260C Anneal



Pt/Au Contact after 4hr 260C Anneal

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

Improvements in HBT Emitter Access Resistance

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

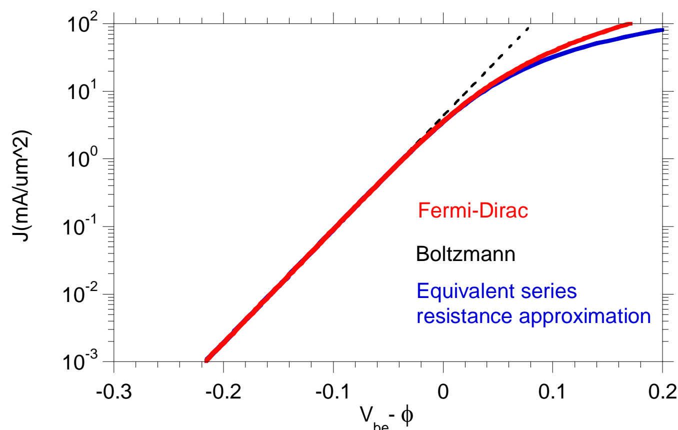
125 nm generation requires 5 Ω - µm^2 emitter resistivities

65 nm generation requires 1-2 Ω - µm^2

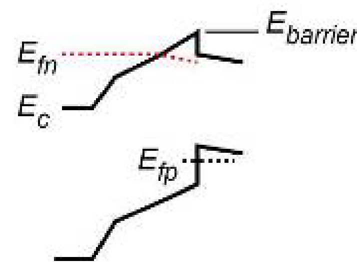
<u>Recent Results</u>

ErAs/Mb	*MBE in-situ*	*1.5 Ω - µm^2*
Mo	*MBE in-situ*	*0.6 Ω - µm^2*
TiPdAu	*ex-situ*	*0.5 Ω - µm^2*
TiW	*ex-situ*	*0.7 Ω - µm^2*

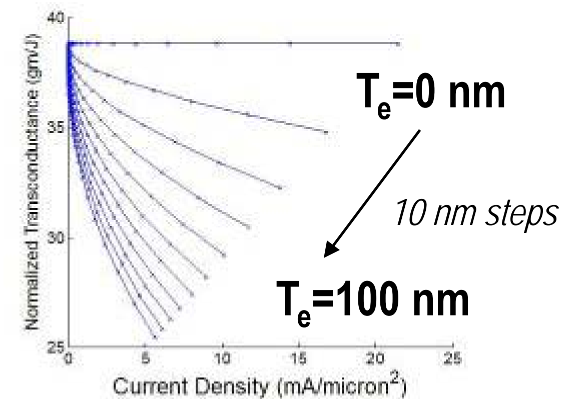
Degeneracy contributes 1 Ω - µm^2



20 nm emitter-base depletion layer contributes 1 Ω - µm^2 resistance

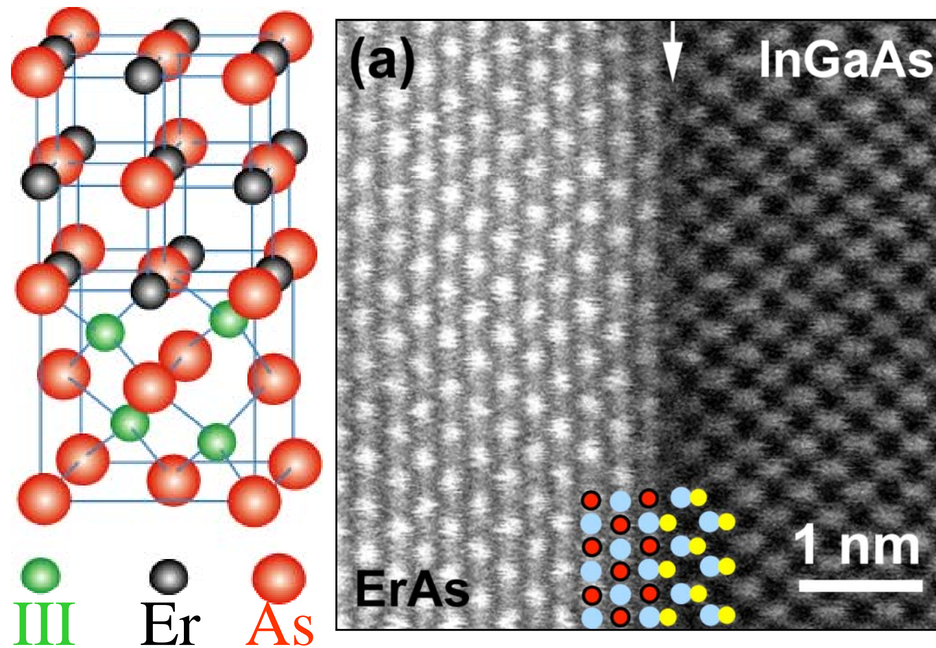


$$\frac{\partial E_{fn}(x)}{\partial x} = \frac{-J}{qn(x)}$$

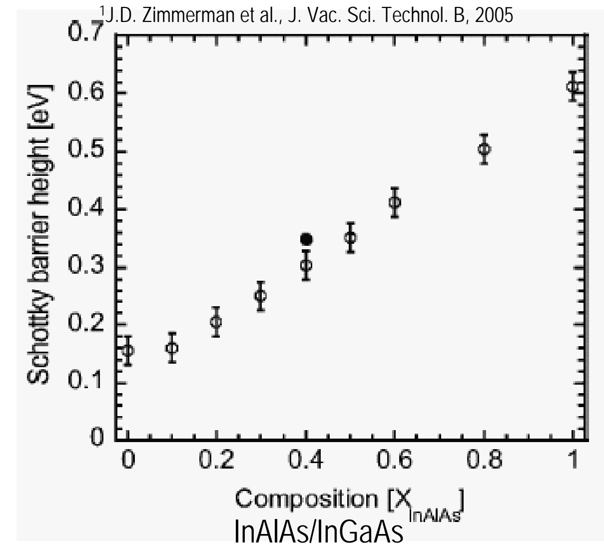


In-situ ErAs-InGaAs Contacts

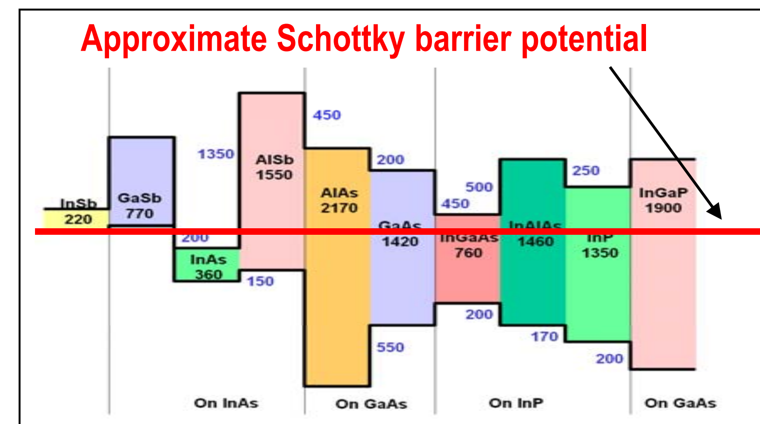
Epitaxially formed, no surface defects, no Fermi level pinning (?)
In-situ, no surface oxides, coherent interface, continuous As sublattice
Thermodynamically stable
ErAs/InAs Fermi level should be above conduction band



D. O. Klenov, Appl. Phys. Lett., 2005



[1]J.D. Zimmerman et al., J. Vac. Sci. Technol. B, 2005



S.R. Bank, NAMBE , 2006

Results nevertheless disappointing:
1.5 Ω - μm^2

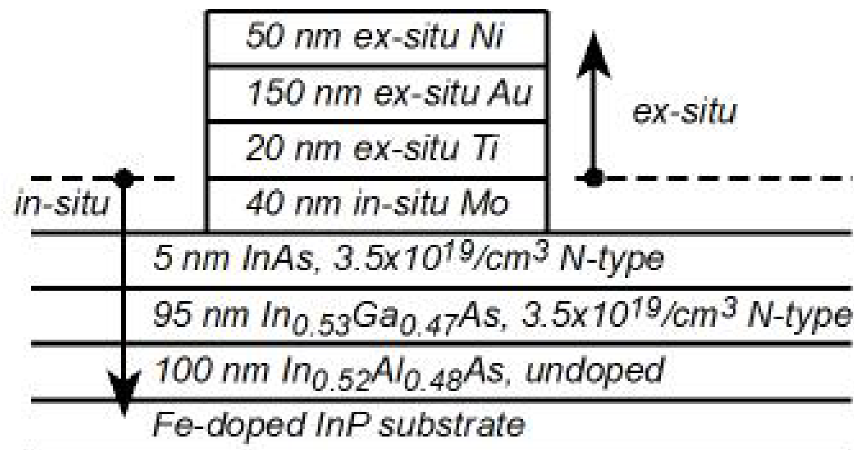
Low-Resistance Refractory Contacts to N-InGaAs

Results initially by luck: control samples for ErAs experiments

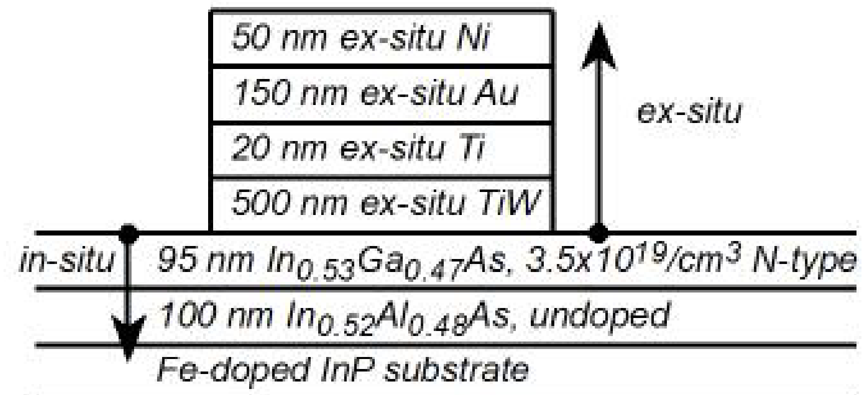
Mo contacts: deposition by MBE immediately after InGaAs growth

TiW contacts: sputter deposition
after UV-Ozone & 14.8-normality ammonia soak

Both give ~ 1 Ω-μm^2 resistitivity



in-situ **Mo contact**



ex-situ **TiW contact**

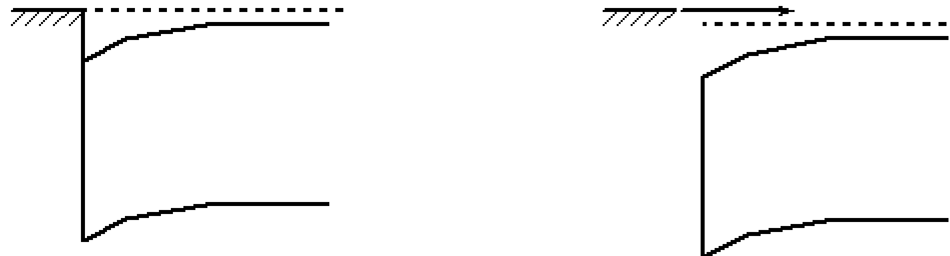
Coherent Epitaxial Metal Semiconductor Contacts ?

Chris Palmstrom suggests materials such as
Fe_3Ga, CoGa, NiAl

It might be possible to grow these with low interfacial densities on InGaAs or InAs.

Key question: what resistivity would we expect for a zero-defect, zero-barrier metal-semiconductor interface ?

If we introduce a small difference in Fermi Level between metal and semiconductor, what current do we compute from integration of N(E) v(E)F(E)T(E) ?

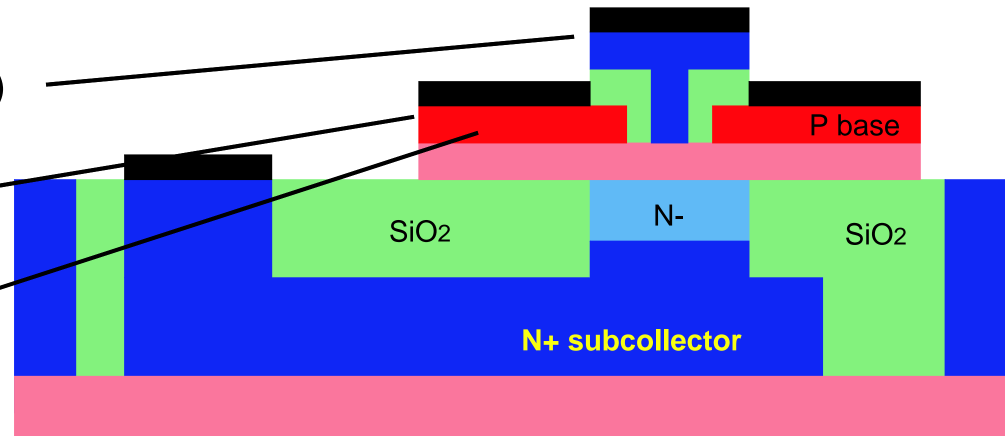


Shape as Substitute for Low-Resistance Contacts: SiGe HBTs

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

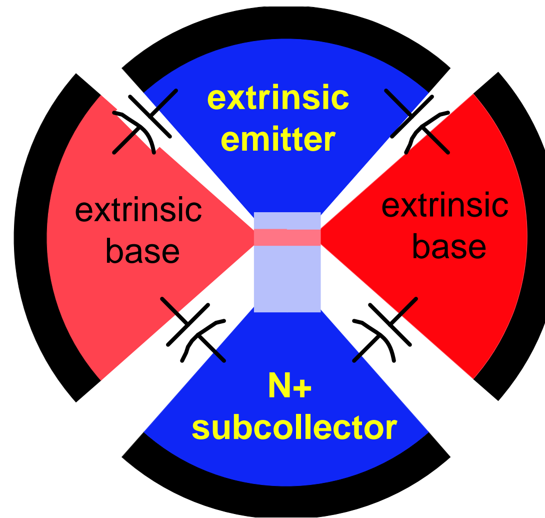
thick extrinsic base : low resistance
thin intrinsic base: low transit time

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



These are planar approximations to radial contacts:

→ ***reduced access resistance***



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

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

should help less with small devices:

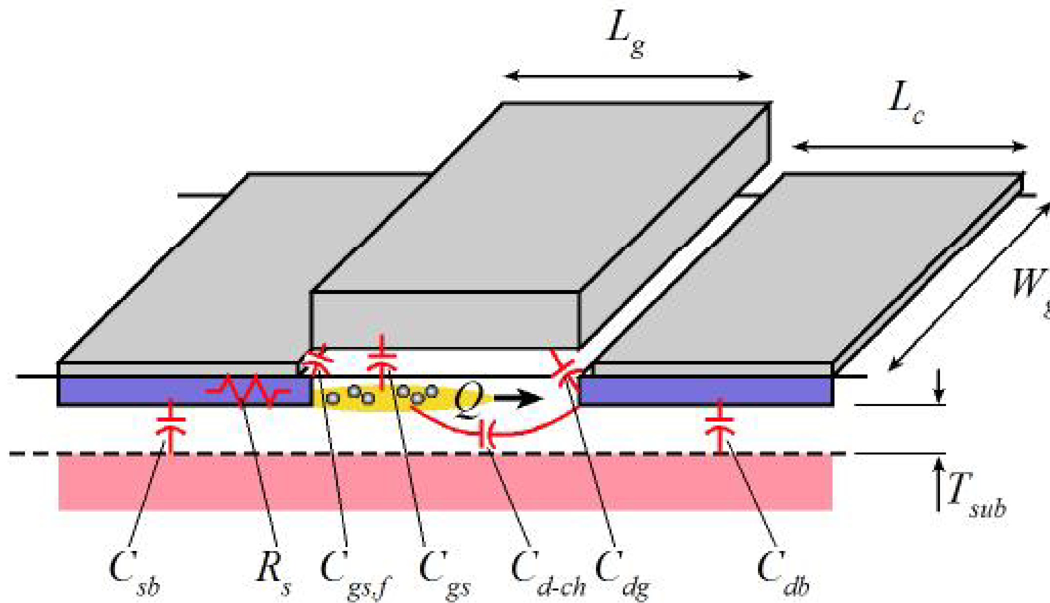
...widths scale faster than thicknesses→ trench fringing capacitance

dielectric trench conducts heat badly

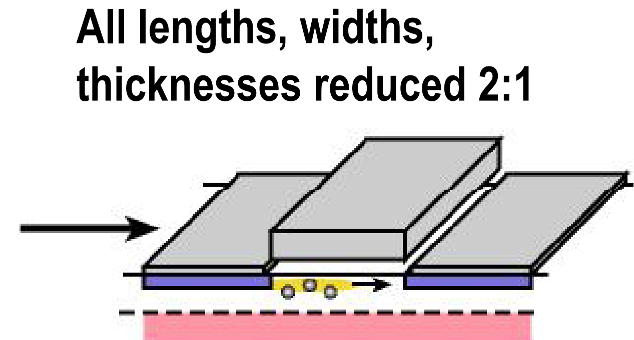
Field-Effect Transistors

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

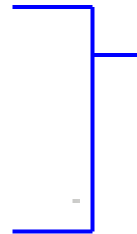


S/D contact resistivity reduced 4:1

$$C_{gs} \sim \varepsilon W_g L_g / T_{ox}$$

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

$$C_{sb} \sim C_{db} \sim \varepsilon W_g L_c / T_{sub}$$

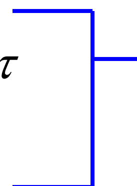


If T_{ox} cannot scale with gate length,
$C_{parasitic}$ / C_{gs} increases, ✗
g_m / W_g does not increase
hence $C_{parasitic}$ /g_m does not scale ✗

$$\tau \sim L_g / v$$

$$g_m \sim C_{gs} / \tau \sim (\varepsilon L_g W_g / T_{ox}) / \tau$$

$$G_{ds} \sim C_{d-ch} / \tau \sim \varepsilon W_g / \tau$$



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
σV_{th} matching (mV·µm) [5]	6	6	6	6	5	5	5	5	5
I_{ds} (µA/µm) [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: often significant SiO2 interlayer, can limit EOT scaling

S/D access resistance also a challenge: about 1 Ω-μm^2 required for 20 nm

Because gate equivalent thickness is not scaling, present devices scale badly
- output conductance is degrading with scaling
- other capacitances are not scaling in proportion to Cgs
  - hence are starting to dominate high frequency performance

How Can Materials Scientists Help ?

High K-dielectrics for Si CMOS are still extremely important

Self-aligned (Salicide-like) contacts
of very low resistivity are needed

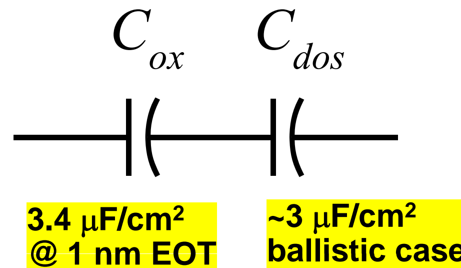
...for 2 mA/micron operation at 700 mV gate overdrive,
we want ~300 Ohm-micron lateral access resistivity
→ about 0.7 Ohm-micron^2 resistivity in a 25 nm wide contact

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

Low access resistance: 1 Ω-μm^2 , 10 Ω-μm
Light electron→ high electron velocity (thermal or Fermi injection)
→ 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

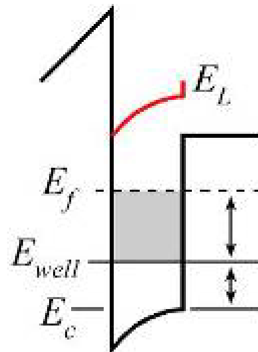
$$C_{dos} = \frac{q^2 m^*}{\pi\hbar^2}$$



limits n_s to ~ 6*10^{12} /cm^2
limits I_d/ W_g
limits g_m/W_g

Challenge:
filling of low-mobility satellite valleys

limits n_s to ~ 8*10^{12} /cm^2
limits I_d/ W_g

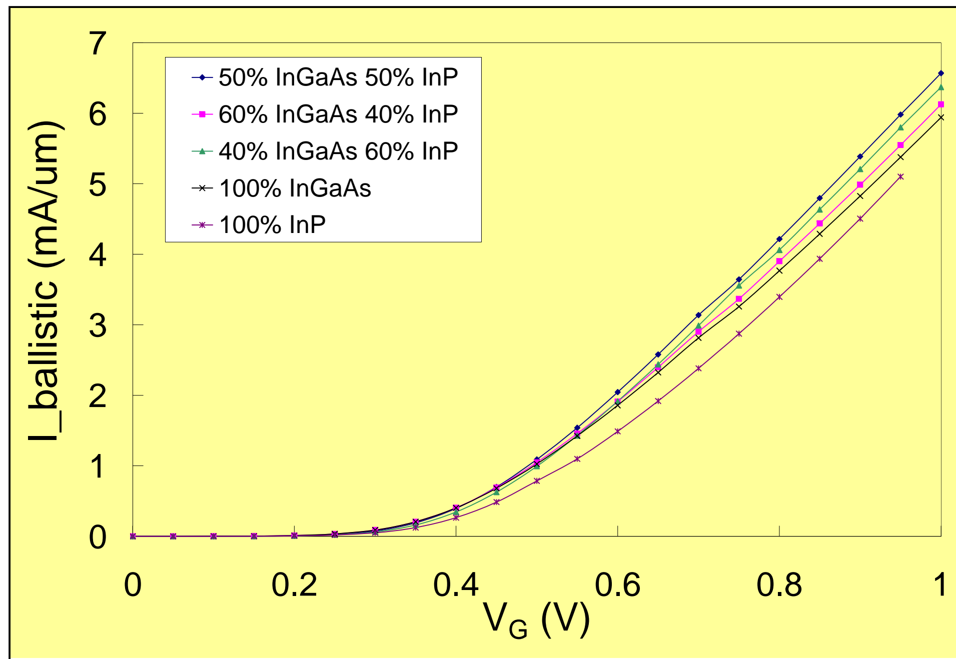


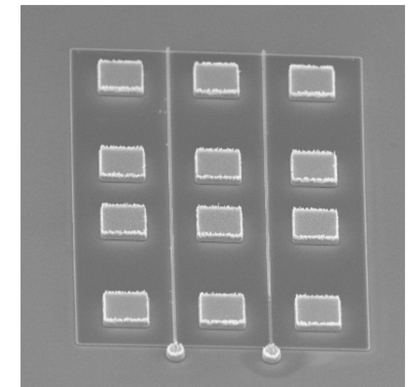
Challenge:
light electron limits vertical scaling
~1.5-2.5 nm minimum mean electron depth

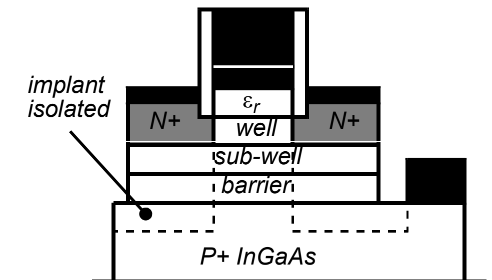
III-V MOS: What might be accomplished

Drive current simulation- ideal (ballistic) assumptions

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





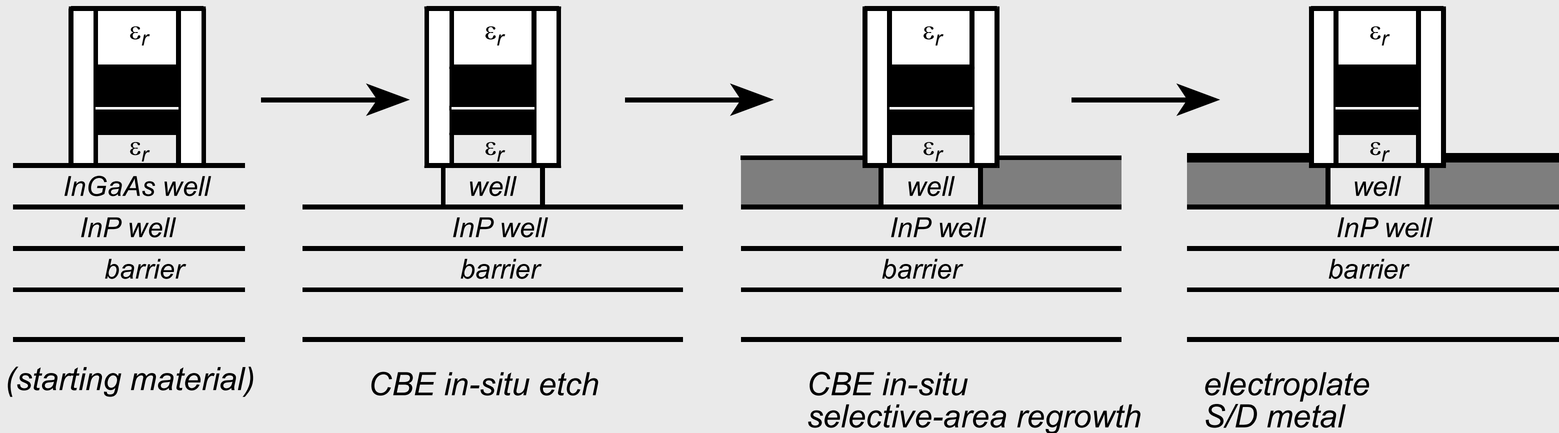


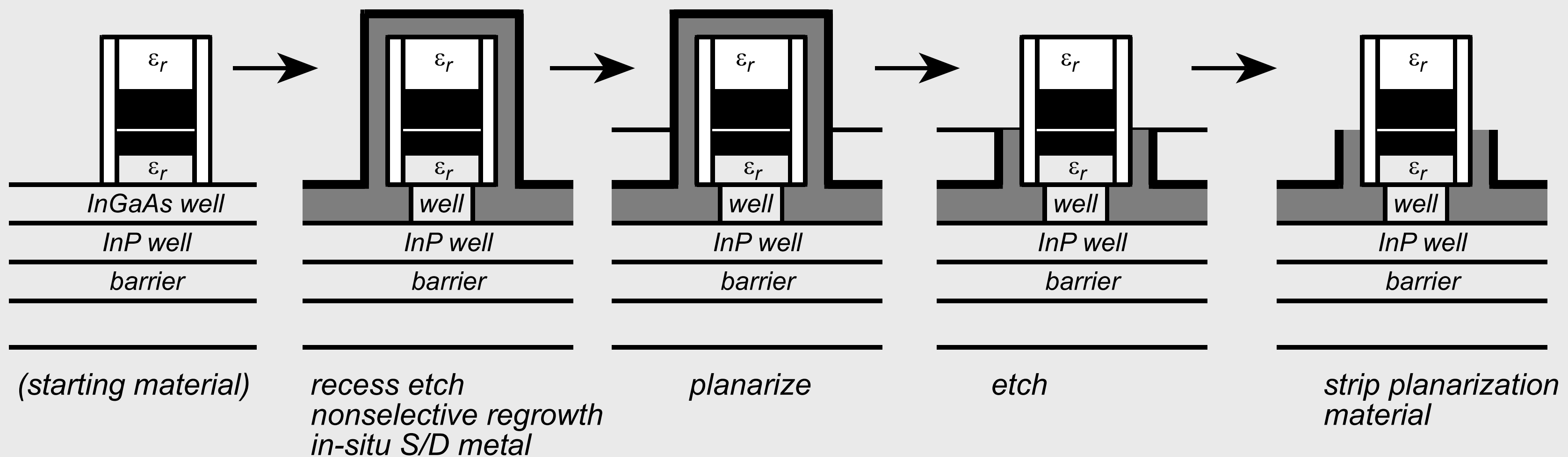
22 nm gate length, 5 nm thick InGaAs / InP channel

under similar assumptions, silicon channels show 3-4 mA /μm

intrinsic C_{gs} ~350 fF/mm --- comparable to fringing and stray capacitances

S/D Contact Process Flow For III-V MOSFETs





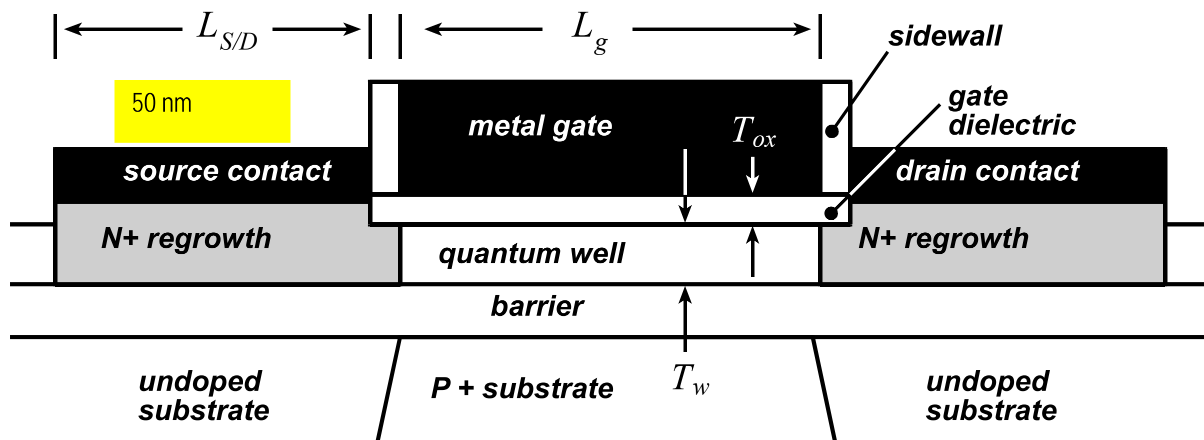
III-V MOSFETs Can Provide Very Low S/D Access Resistance

Objective : $I_d/W_g \sim 5$ mA/μm @ $(V_{gs} - V_{th}) = 0.7$ V

$\Rightarrow$ transconductance $g_m/W_g > 7$ mS/μm

$\Rightarrow$ 14 Ω - μm source resistance will reduce g_m and I_d by 10%.

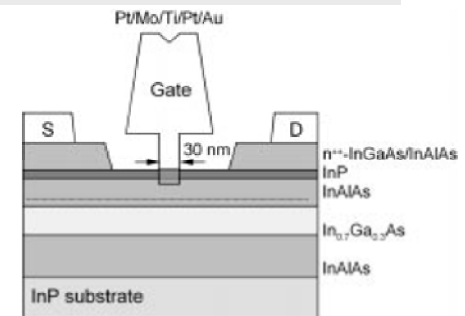
With 50 nm wide contacts, this requires $\rho_c < 0.7\ \Omega\text{-}\mu\text{m}^2$



Modern III - V HEMTs have ~ 10 : 1 larger (~ 100 Ω - μm) source resistance...
because of the poor extrinsic source access region.

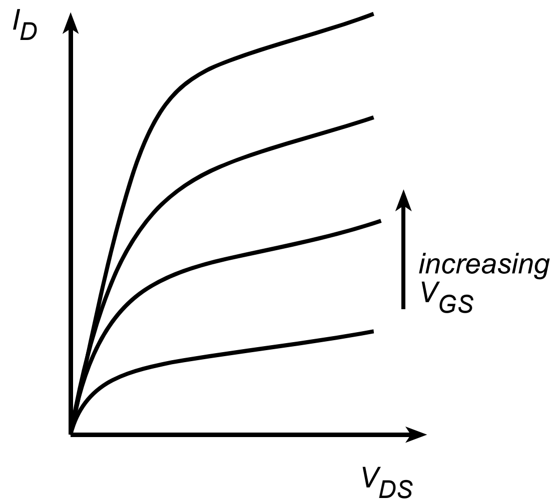
Extremely High g_m > 2.2 S/mm and f_T > 550 GHz in 30-nm Enhancement-Mode InP-HEMTs with Pt/Mo/Ti/Pt/Au Buried Gate

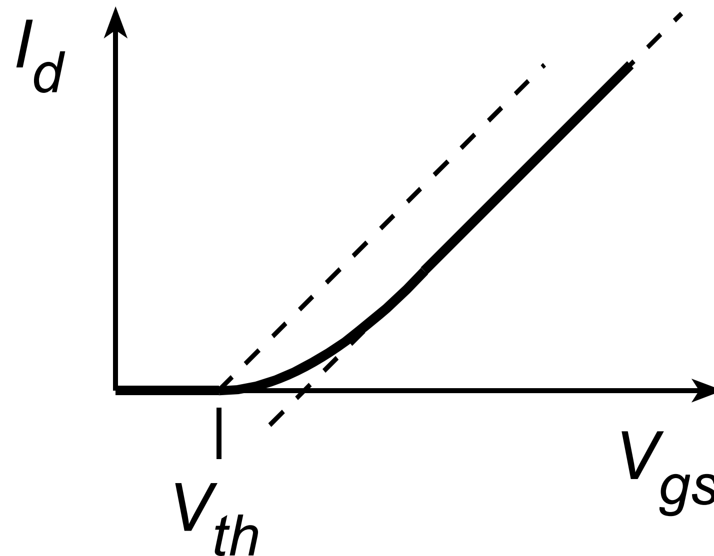
Keisuke Shinohara[1], Wonill Ha[1], Mark J.W. Rodwell[2], and Berinder Brar[1]



Improving FETs by Developing Other Materials

Other materials may offer high mobilities but...





$$I_D \approx c_{ox} W_g v_{injection} (V_{gs} - V_{th} - \Delta V) \text{ for } (V_{gs} - V_{th}) / \Delta V >> 1$$

$$\text{where } \Delta V = v_{injection} L_g / \mu$$

→ ***mobilities above ~ 1000 cm²/V-s of little benefit at 22 nm Lg***

increased injection velocities are of value...

...but not at sacrifice in density of states

Nanopillar and Nanowire Devices

Nanopillar devices might have improved 2-D electrostatics
... but only if wire diameter is ~10 nm or less

Access resistances are serious issue

Capacitances to source-drain pad regions a serious concern

III-V Nanowires FETs still must address defect density dielectric-semiconductor interface

III-V nanopillar devices experience same DOS, confinement challenges as planar III-V devices

Conclusion

THz & nm Transistor Electronics is all about the interfaces

Bipolar Transistors:

P and N ohmic contacts with very low resistivity

stability at high current density

FETs

gate dielectrics

contact resistance

density of states