Short Course, 2009 Conference on InP and Related Materials, Newport Beach, CA, May 10-14

Scaling of High Frequency III-V Transistors

Mark Rodwell

University of California, Santa Barbara

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

THz Transistors

Transistor bandwidths are increasing rapidly.

Si MOSFETs will soon reach 500+ GHz cutoff frequencies.

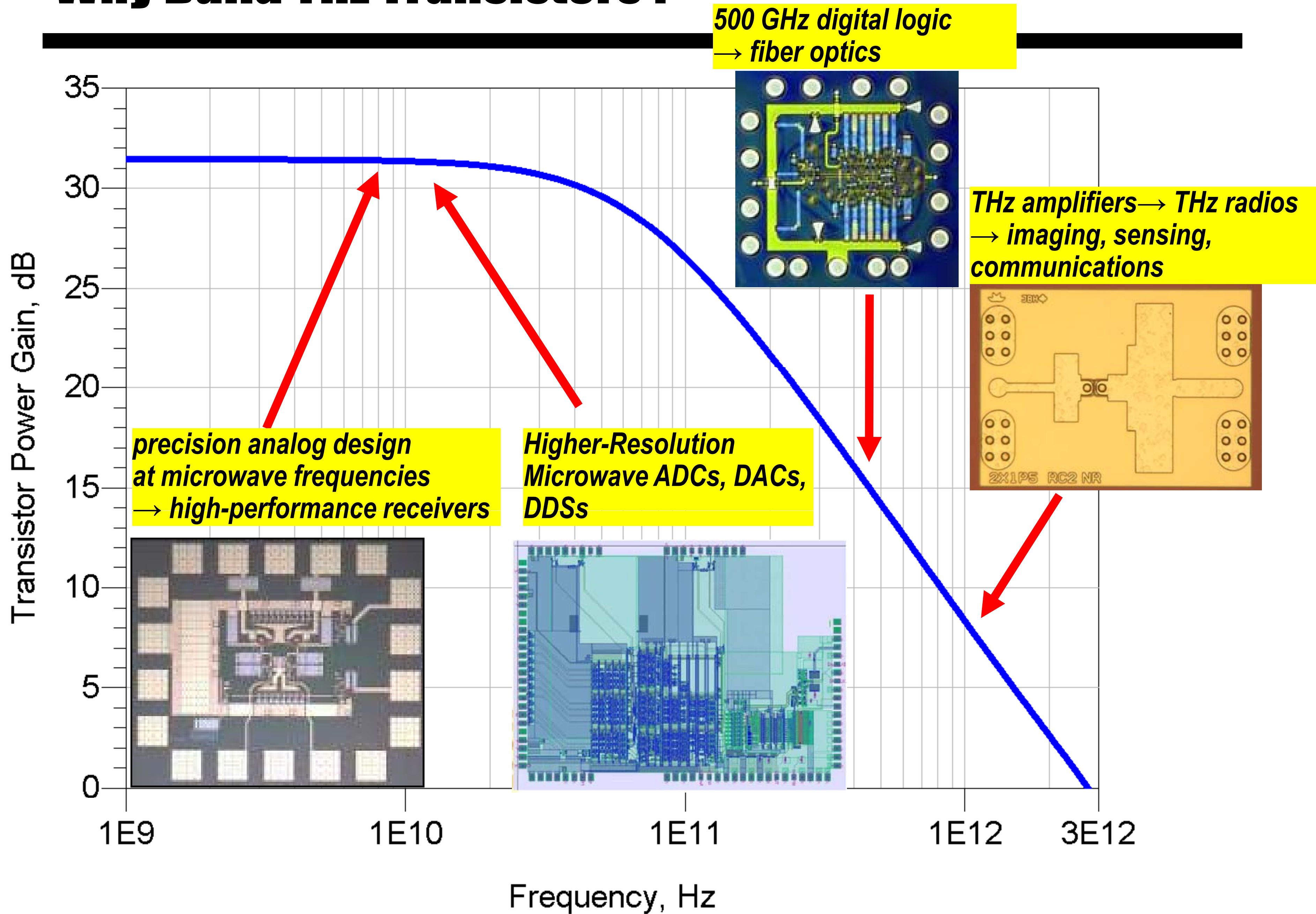
It is now clear III-V bipolar transistors can reach ~2-3 THz cutoff frequencies.

III-V FETs have comparable potential, but the prospects and analysis are less clear.

The limits to transistor bandwidth are:
- *contact resistivities*
- *gate dielectric capacitance densities.*
- *device and IC power density & thermal resistance.*
- *challenges in reliably fabricating small devices.*

Why
THz Transistors ?

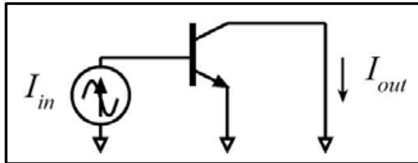
Why Build THz Transistors ?

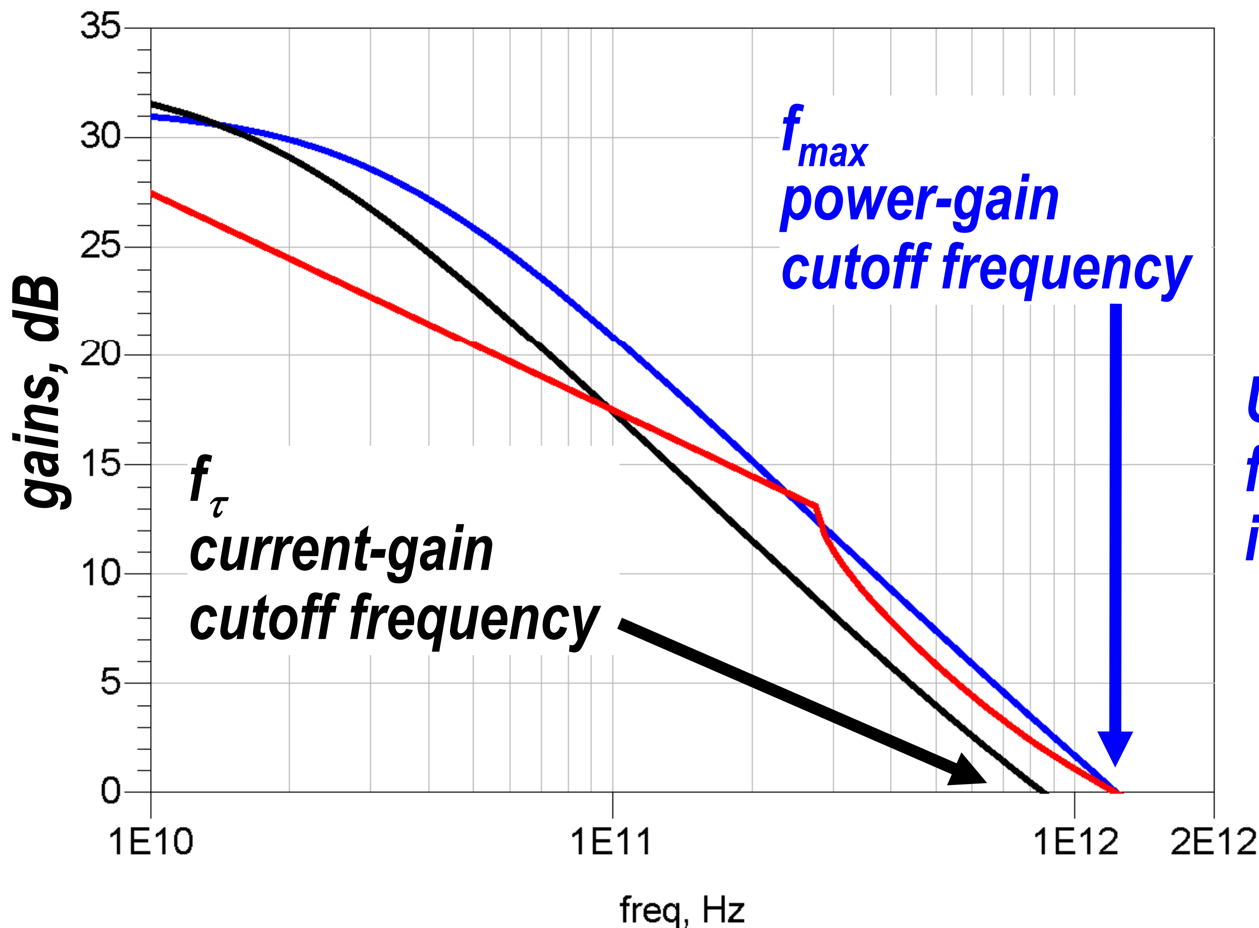


Performance Figures of Merit

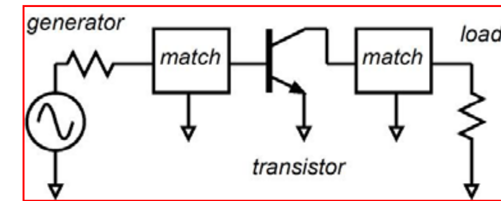
Transistor figures of Merit / Cutoff Frequencies

H_{21}=short-circuit current gain

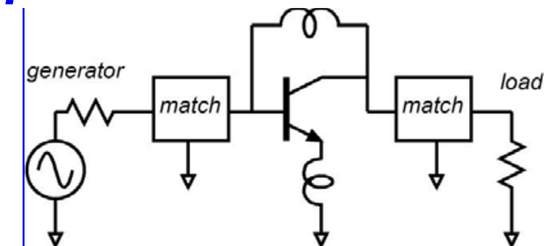




MAG = maximum available power gain: impedance-matched



U= unilateral power gain: feedback nulled, impedance-matched



What Determines Gate Delay ?

Gate Delay Determined by :

Depletion capacitance charging through the logic swing

$$\left(\frac{\Delta V_{LOGIC}}{I_C}\right)\left(C_{cb} + C_{be,\text{depletion}}\right)$$

Depletion capacitance charging through the base resistance

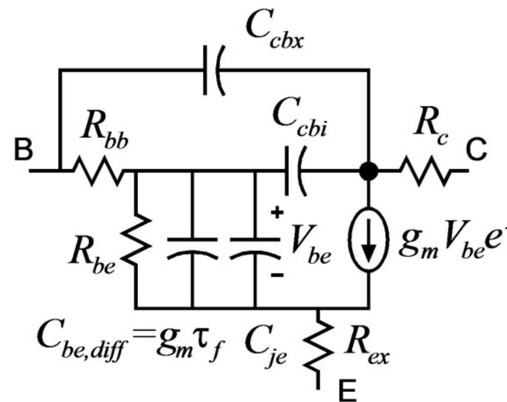
$$R_{\text{bb}}\left(C_{cbi} + C_{be,\text{depletion}}\right)$$

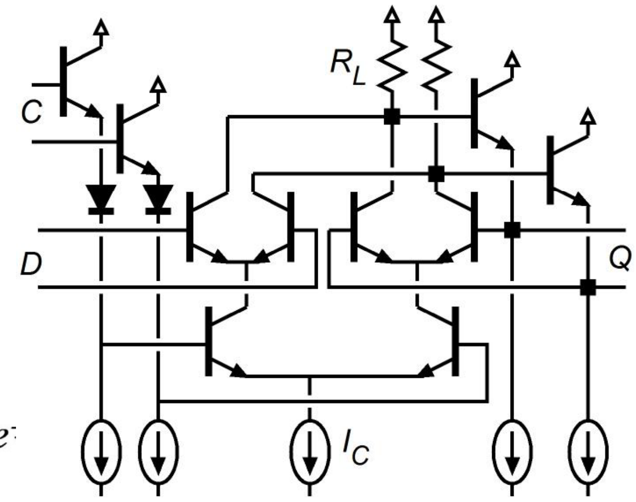
Supplying base + collector stored charge through the base resistance

$$R_{\text{bb}}\left(\tau_b + \tau_c\right)\left(\frac{I_C}{\Delta V_{LOGIC}}\right)$$

The logic swing must be at least

$$\Delta V_{LOGIC} > 4 \cdot \left(\frac{kT}{q} + R_{ex} I_c\right)$$





$(\tau_b + \tau_c)$ typically 10 - 25% of total delay;
Delay not well correlated with f_τ

$(\Delta V_{LOGIC} / I_C)(C_{cb} + C_{be,\text{depl}})$ is 55% - 80% of total.

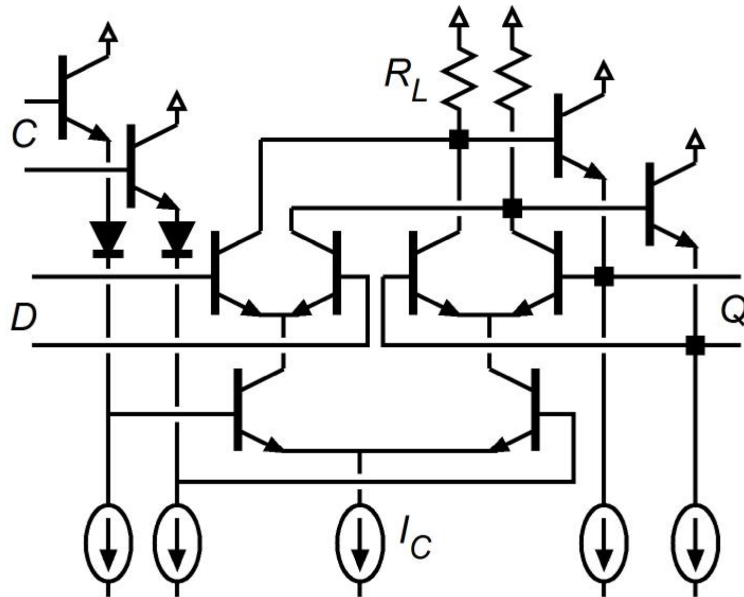
High (I_C / C_{cb}) is a key HBT design objective.

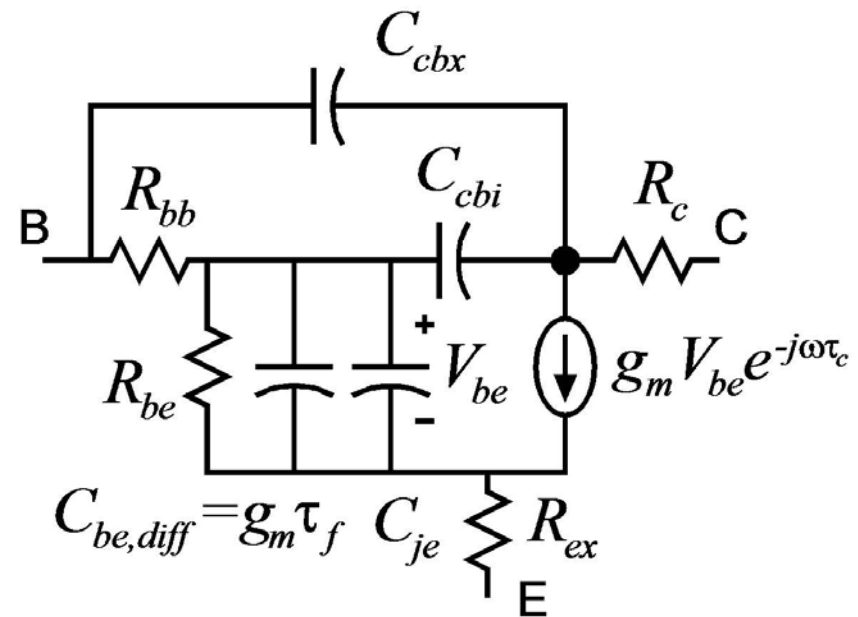
$$J_{\max, Kirk} = 2\varepsilon \bar{v}_{electron}(V_{ce,\text{ operating}} + V_{ce,\text{full depletion}}) / T_c^2$$

$$\Rightarrow \frac{C_{cb} \Delta V_{LOGIC}}{I_C} = \frac{\Delta V_{LOGIC}}{2 V_{CE,\min}} \left(\frac{A_{\text{collector}}}{A_{\text{emitter}}}\right)\left(\frac{T_C}{2\bar{v}_{electron}}\right)$$

R_{ex} must be very low for low ΔV_{logic} at high J

HBT Design For Digital & Mixed-Signal Performance





from charge-control analysis:

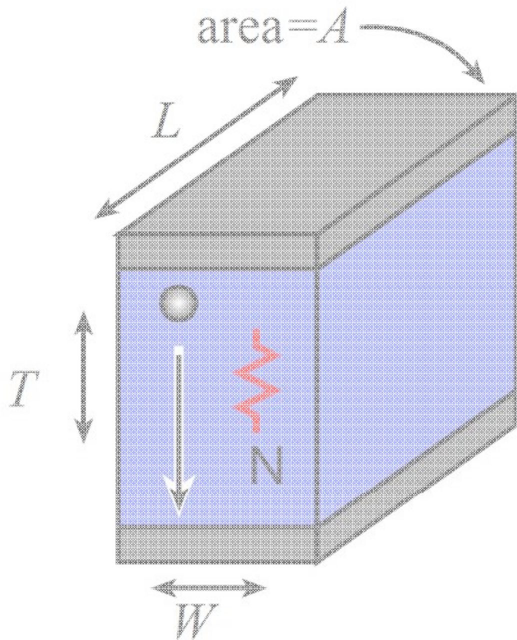
$$
\begin{aligned}
T_{gate} \approx\ & (\Delta V_L / I_C)(C_{je} + 6C_{cbx} + 6C_{cbi}) + \tau_f \\
& + (kT / qI_C)(0.5C_{je} + C_{cbx} + C_{cbi} + 0.5\tau_f I_C / \Delta V_L) \\
& + R_{ex}(0.5C_{cbx} + 0.5C_{cbi} + 0.5\tau_f I_C / \Delta V_L) \\
& + R_{bb}(0.5C_{je} + C_{cbi} + 0.5\tau_f I_C / \Delta V_L).
\end{aligned}
$$

analog ICs have similar bandwidth constraints...

High-Frequency Electron Device Design

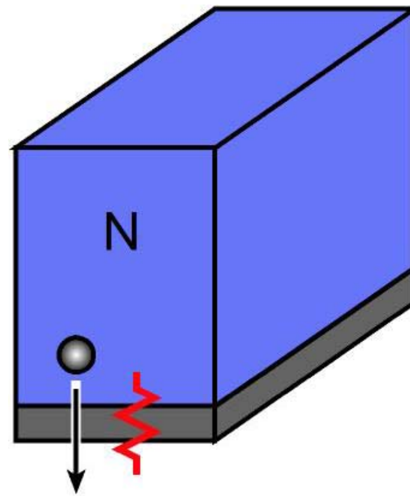
Simple Device Physics: Resistance

bulk resistance



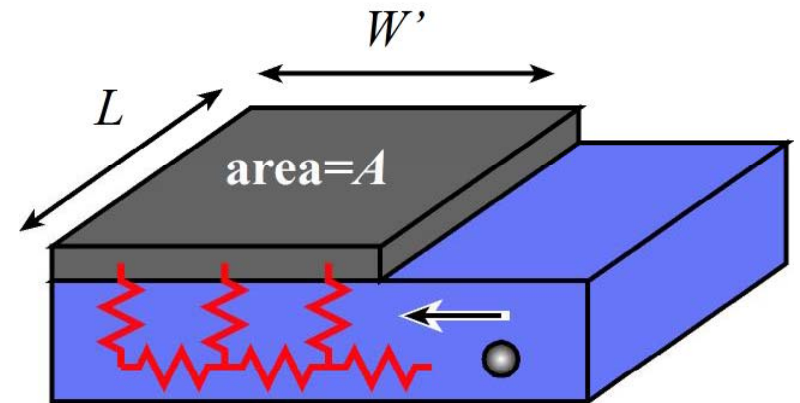
$$R = \frac{\rho_{bulk} \cdot T}{A}$$

contact resistance -perpendicular



$$R = \frac{\rho_{contact}}{A}$$

contact resistance - parallel

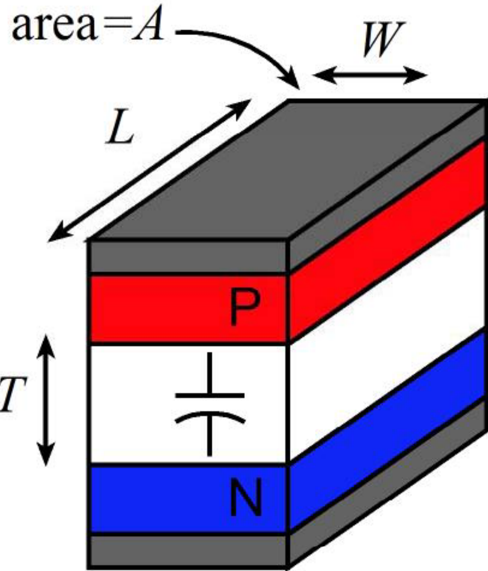


$$R = \frac{\rho_{contact}}{A} + \rho_{sheet} \cdot \frac{W'}{3L}$$

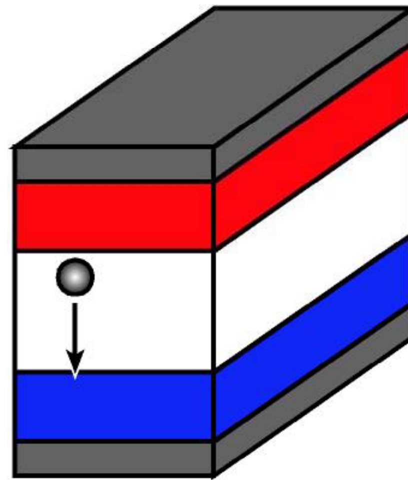
Good approximation for contact widths less than 2 transfer lengths.

Simple Device Physics: Depletion Layers

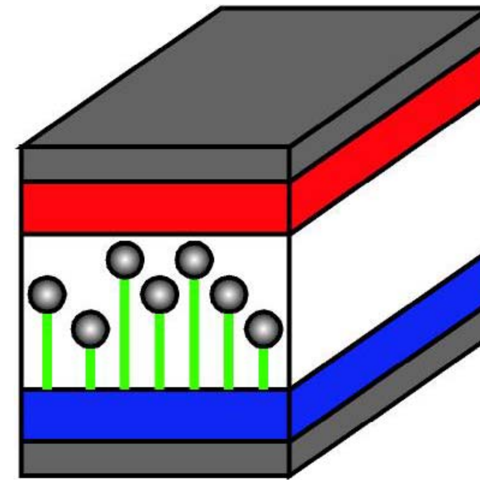
capacitance



transit time



space-charge limited current



$$C = \varepsilon \cdot \frac{A}{T}$$

$$\tau = \frac{T}{2v}$$

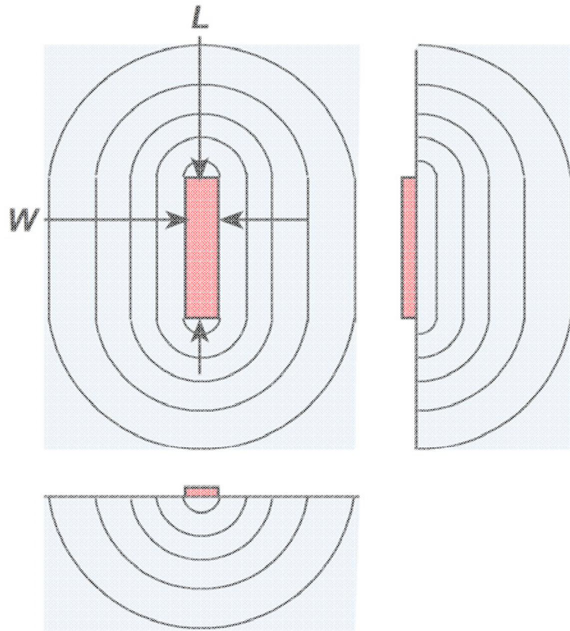
$$\frac{I_{\max}}{A} = \frac{2\varepsilon v}{T^2}\left(V_{applied} + V_{depletion} + 2\phi\right)$$

$$I = C\frac{\Delta V}{\Delta T} \text{ where } \frac{C}{I_{\max}} = \frac{\tau}{V_{applied} + V_{depletion} + 2\phi}$$

Simple Device Physics: Thermal Resistance

Exact

Carslaw & Jaeger 1959



$$R_{th} = \frac{1}{\pi K_{th} L} \sinh^{-1}\left(\frac{L}{W}\right) + \frac{1}{\pi K_{th} W} \sinh^{-1}\left(\frac{W}{L}\right)$$

Long, Narrow Stripe

HBT Emitter, FET Gate

$$R_{th} \cong \underbrace{\frac{1}{\pi K_{th} L} \ln\left(\frac{L}{W}\right)}_{\text{cylindrical heat flow near junction}} + \underbrace{\frac{1}{\pi K_{th} L}}_{\text{spherical heat flow far from junction}}$$

cylindrical heat flow near junction

spherical heat flow far from junction

Square (L by L)

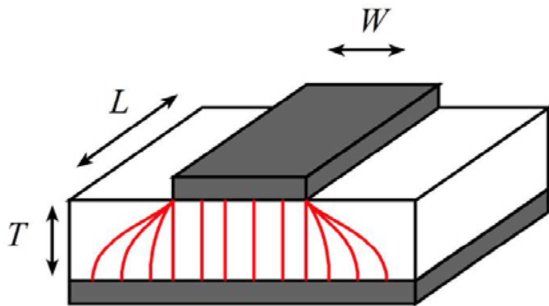
IC on heat sink

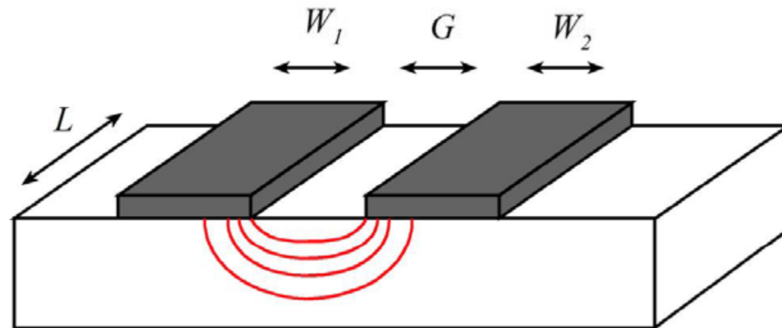
$$R_{th} \cong \underbrace{\frac{1}{4 K_{th} L}}_{\text{planar heat flow near surface}} + \underbrace{\frac{1}{\pi K_{th} L}}_{\text{spherical heat flow far from surface}}$$

planar heat flow near surface

spherical heat flow far from surface

Simple Device Physics: Fringing Capacitance



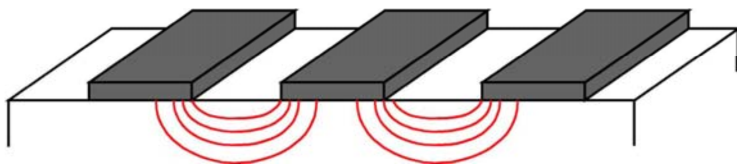


$$\frac{C}{L} \cong \underbrace{\varepsilon \cdot \frac{W}{T}}_{\text{parallel - plate}} + \underbrace{1.5 \cdot \varepsilon}_{\text{fringing}}$$

$$\frac{C}{L} \cong \varepsilon \cdot \left[\begin{array}{l} \text{slowly - varying function} \\ \text{of } W_1 / G \text{ and } W_2 / G \end{array} \right]$$

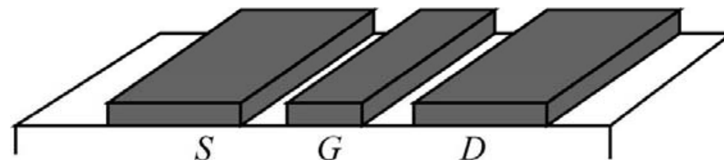
$$\approx (1 \text{ to } 3) \cdot \varepsilon$$

wiring capacitance



$$C / L > \varepsilon$$

FET parasitic capacitances

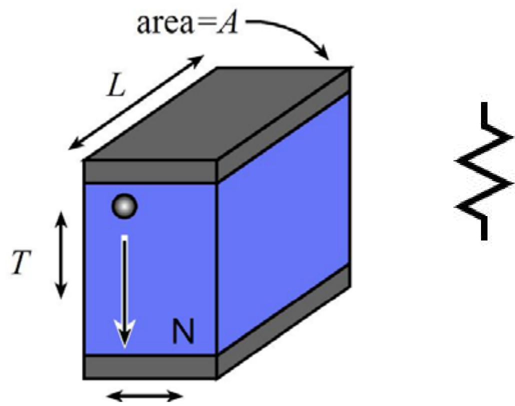


$$C_{parasitic} / L \sim \varepsilon$$

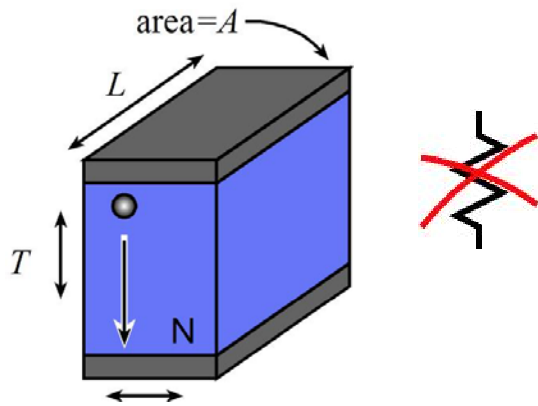
VLSI power-delay limits

FET scaling constraints

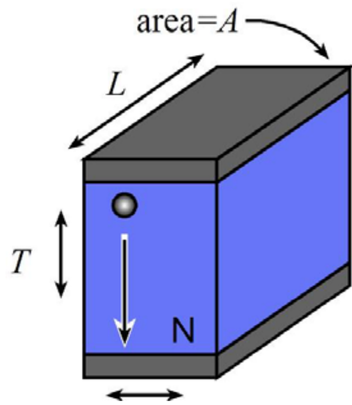
Electron Plasma Resonance: Not a Dominant Limit

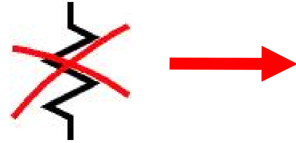


Electron Plasma Resonance: Not a Dominant Limit



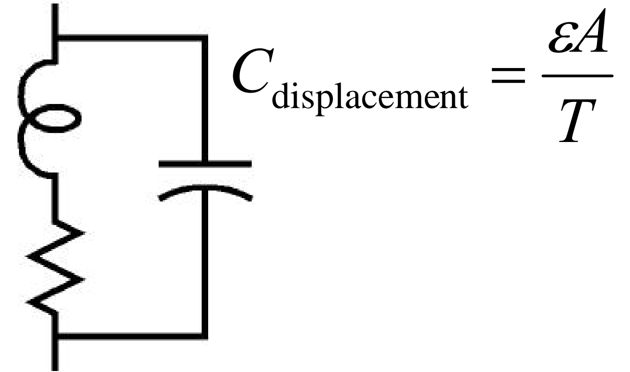
Electron Plasma Resonance: Not a Dominant Limit





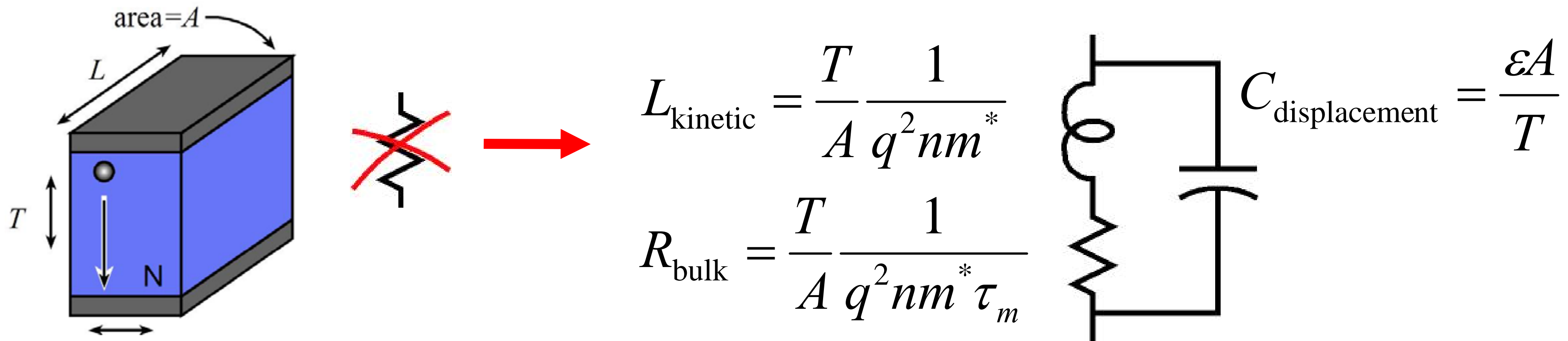
$$L_{\text{kinetic}} = \frac{T}{A}\frac{1}{q^2 n m^*}$$

$$R_{\text{bulk}} = \frac{T}{A}\frac{1}{q^2 n m^* \tau_m}$$



$$C_{\text{displacement}} = \frac{\varepsilon A}{T}$$

Electron Plasma Resonance: Not a Dominant Limit



$$L_{\text{kinetic}} = \frac{T}{A}\frac{1}{q^2 n m^*}$$

$$R_{\text{bulk}} = \frac{T}{A}\frac{1}{q^2 n m^* \tau_m}$$

$$C_{\text{displacement}} = \frac{\varepsilon A}{T}$$

	dielectric relaxation frequency	scattering frequency	plasma frequency
	$f_{dielecic} = \frac{1/2\pi}{C_{\text{displacement}} R_{\text{bulk}}} = \frac{1}{2\pi}\frac{\sigma}{\varepsilon}$	$f_{dielecic} = \frac{1}{2\pi}\frac{R_{\text{bulk}}}{L_{\text{kinetic}}} = \frac{1}{2\pi\tau_m}$	$f_{plasma} = \frac{1/2\pi}{\sqrt{L_{\text{kinetic}} C_{\text{displacement}}}}$
n - InGaAs $3.5 \cdot 10^{19}/\text{cm}^3$	800 THz	7 THz	74 THz
p - InGaAs $7 \cdot 10^{19}/\text{cm}^3$	80 THz	12 THz	31 THz

Frequency Limits and Scaling Laws of (most) Electron Devices

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

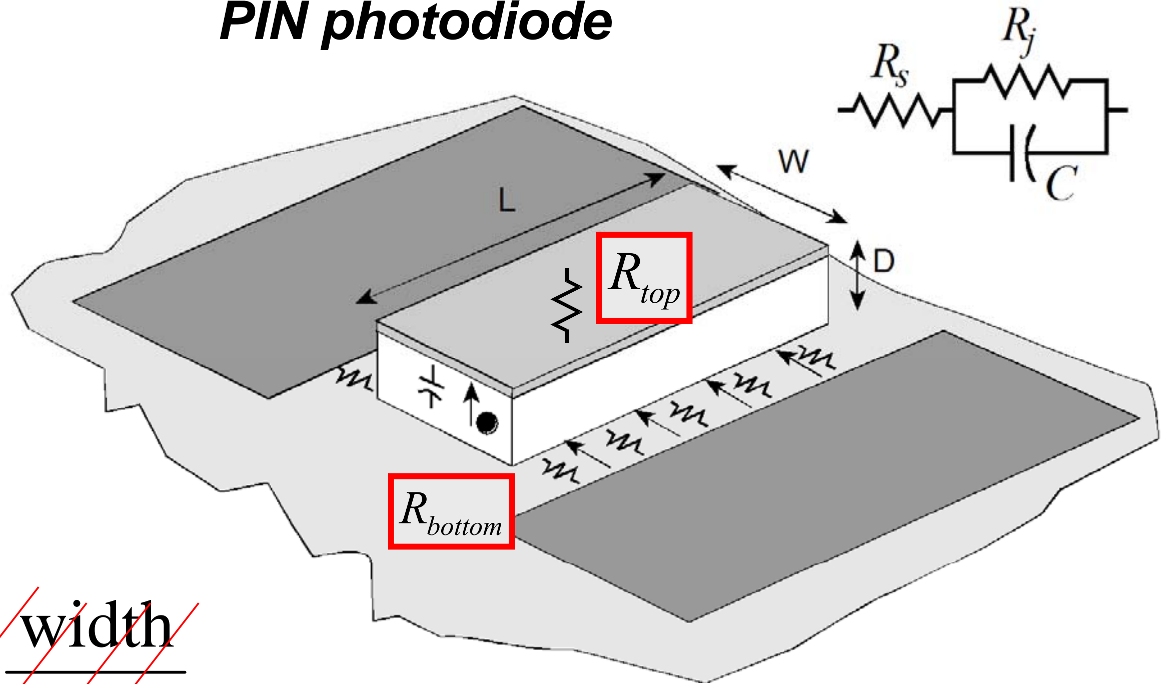
$$C \propto \text{area} / \text{thickness}$$

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

$$R_{bottom} \propto \frac{\rho_{contact}}{\text{area}} + \frac{\rho_{sheet}}{4} \cdot \frac{\text{width}}{\text{length}}$$

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

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



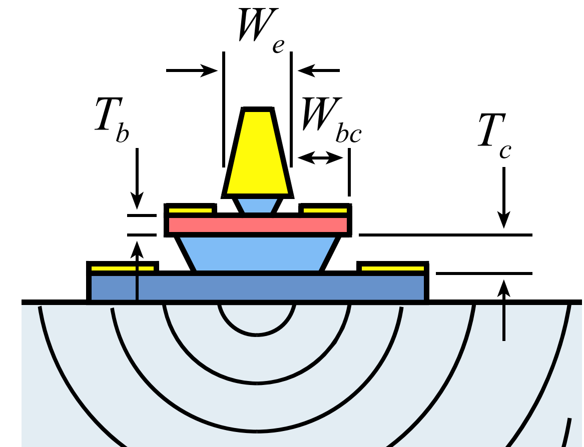
To double bandwidth,

reduce thicknesses 2:1 Improve contacts 4:1

reduce width 4:1, keep constant length

increase current density 4:1

Bipolar Transistor Scaling Laws



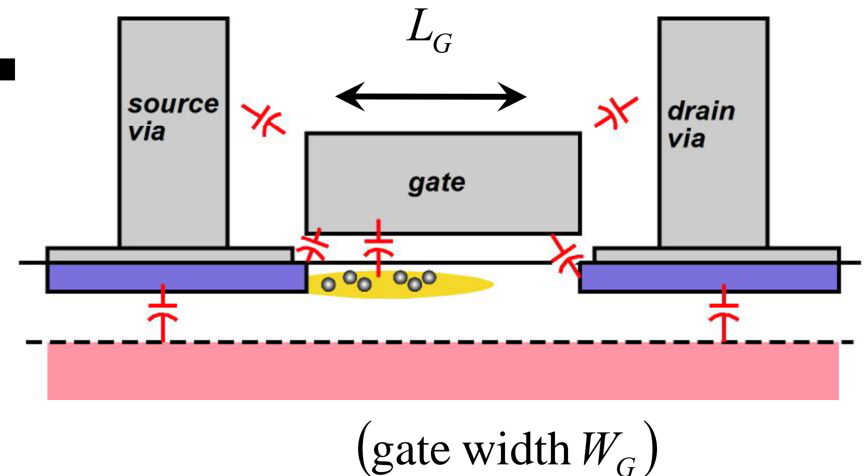
Changes required to double transistor bandwidth:

(emitter length L_E)

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.

FET Scaling Laws



$\left(\text{gate width } W_G\right)$

Changes required to double transistor bandwidth:

parameter	change
gate length	decrease 2:1
gate dielectric capacitance density	increase 2:1
gate dielectric equivalent thickness	decrease 2:1
channel electron density	increase 2:1
source & drain contact resistance	decrease 4:1
current density (mA/μm)	increase 2:1

Linewidths scale as the inverse of bandwidth because fringing capacitance does not scale.

THz & nm Transistors: it's all about the interfaces

Metal-semiconductor interfaces (Ohmic contacts):
very low resistivity

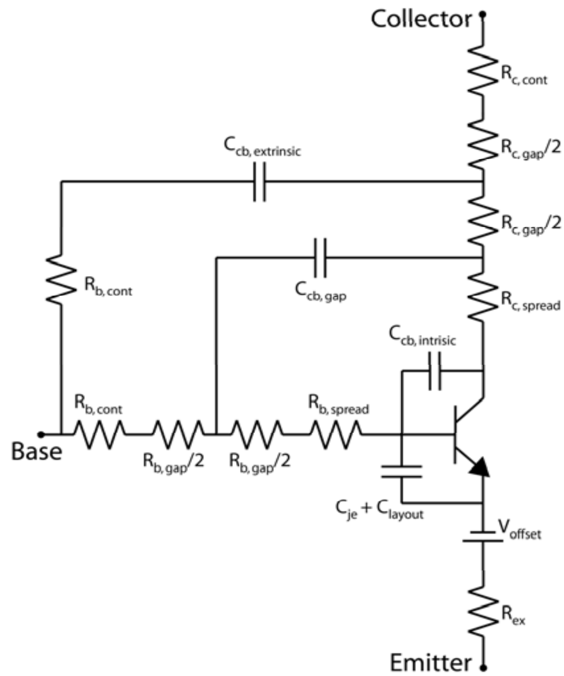
Dielectric-semiconductor interfaces (Gate dielectrics):
very high capacitance density

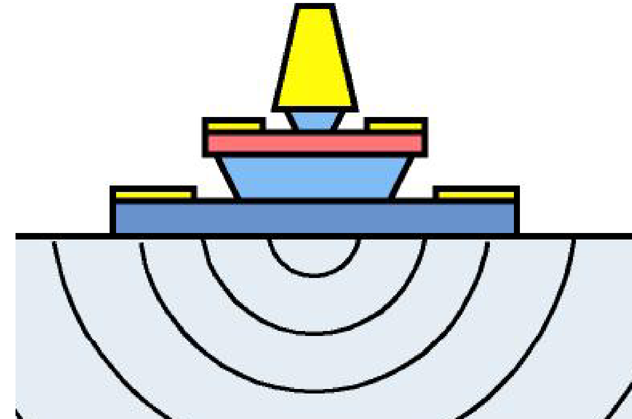
Transistor & IC thermal resistivity.

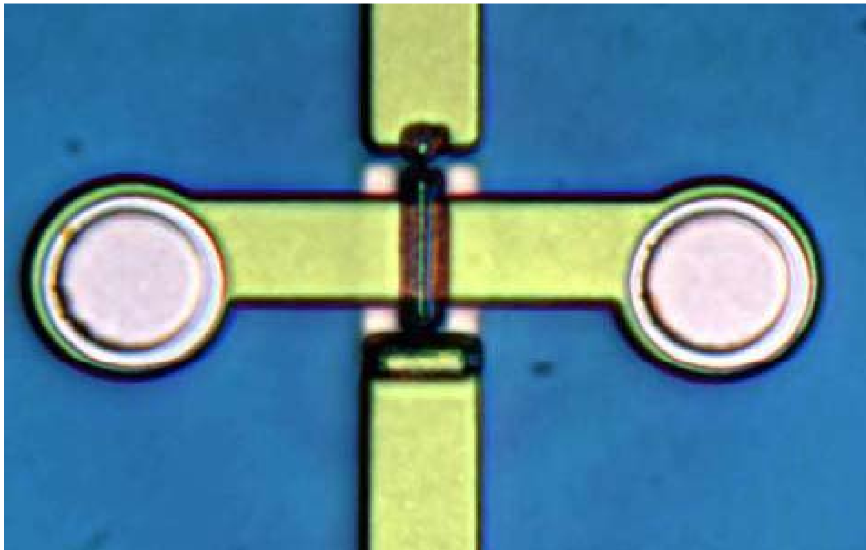
Bipolar Transistors

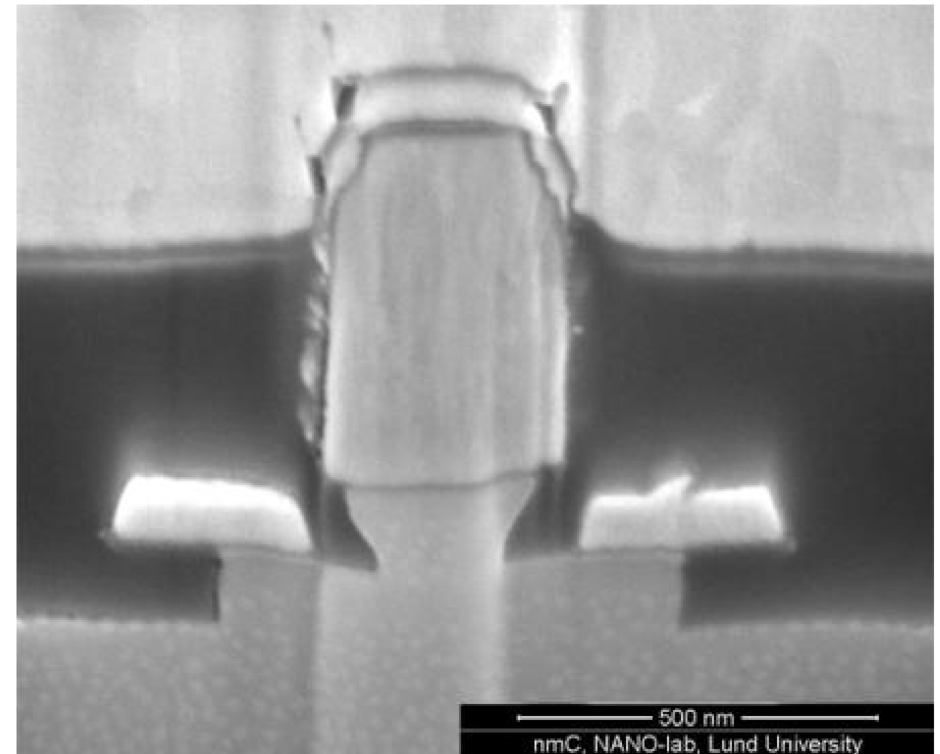
Indium Phosphide Heterojunction Bipolar Transistors

Z. Griffith
E. Lind

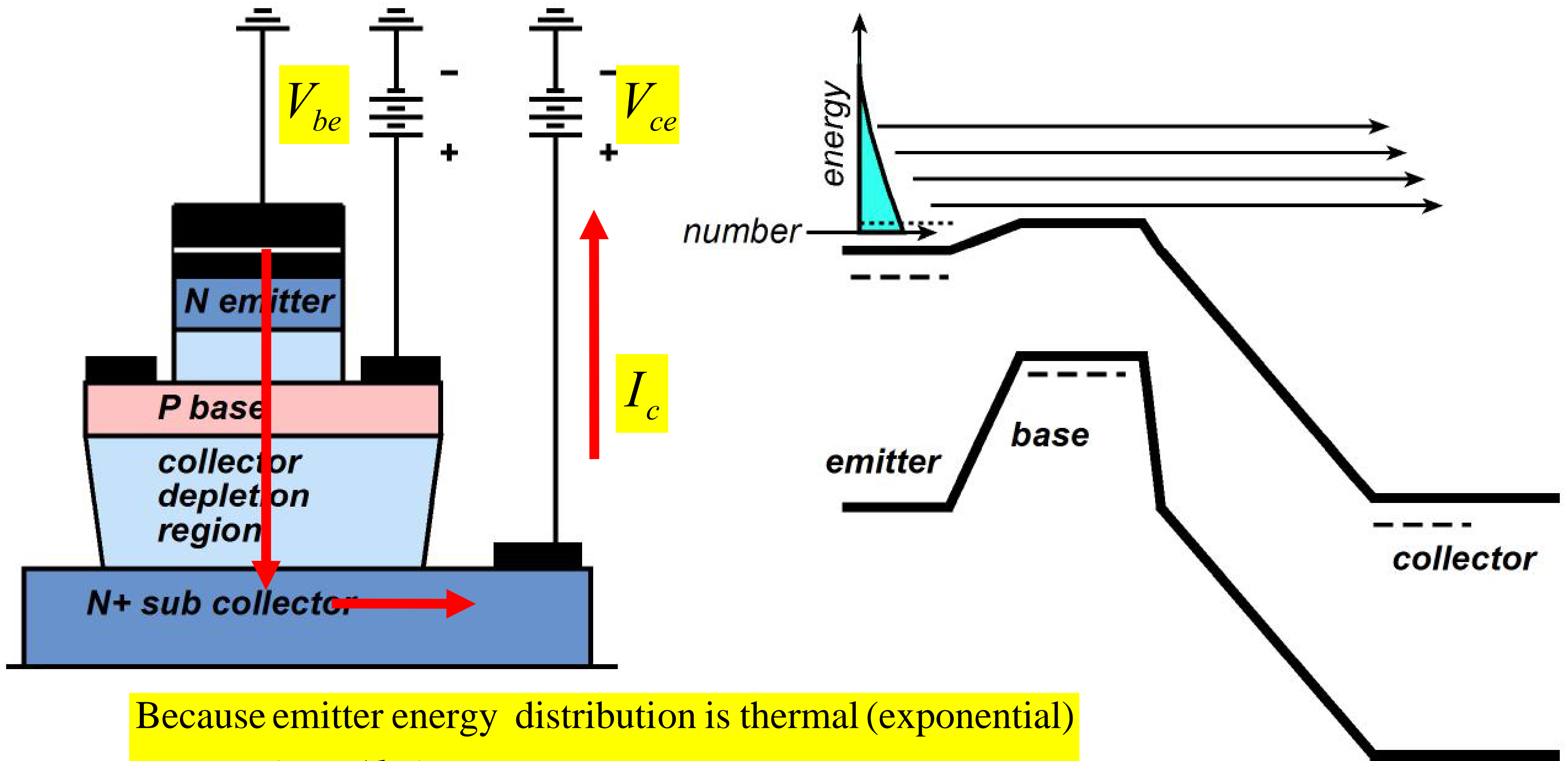








Bipolar Transistor Operation



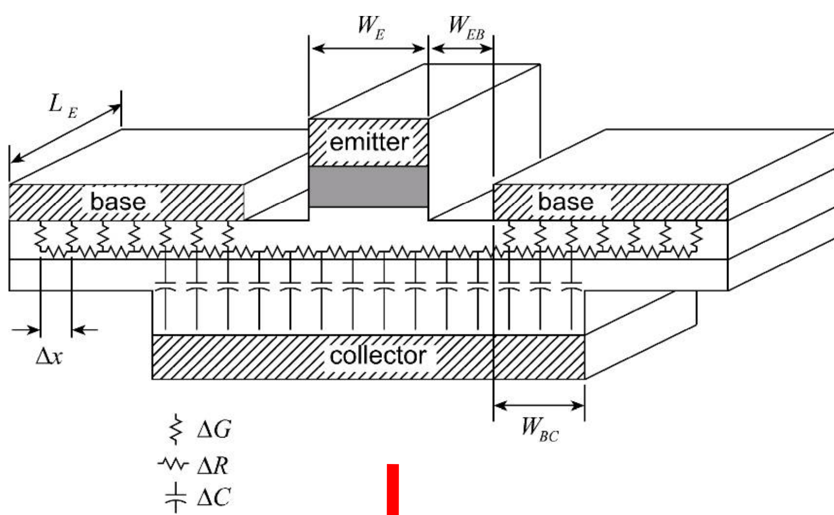
Because emitter energy distribution is thermal (exponential)

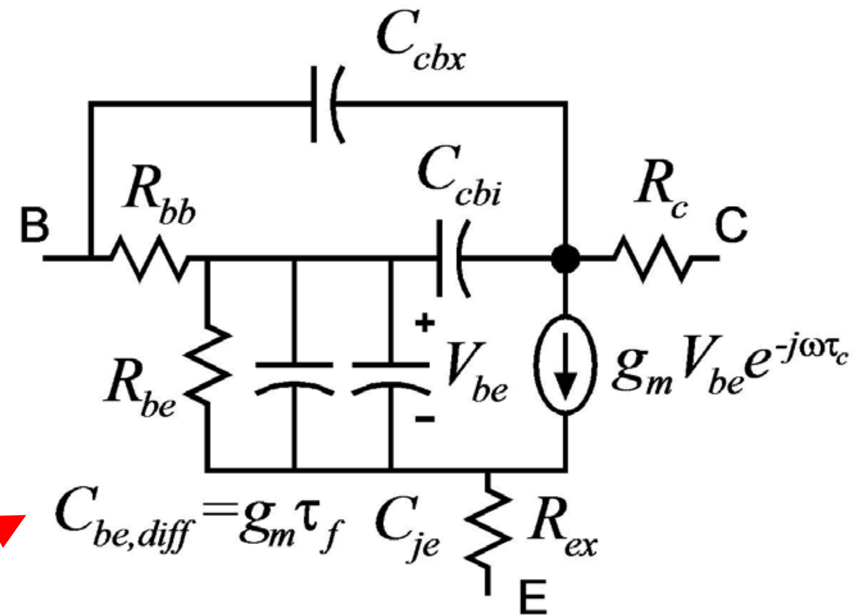
$I_c \propto \exp(qV_{be}/kT)$

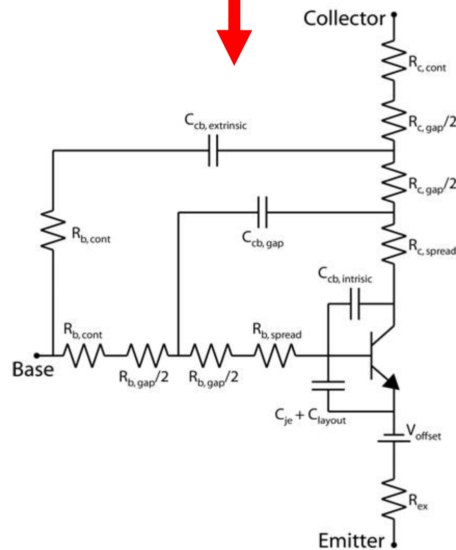
Almost all electrons reaching base pass through it

$\rightarrow I_c$ varies little with collector voltage

Transistor Hybrid-Pi equivalent circuit model





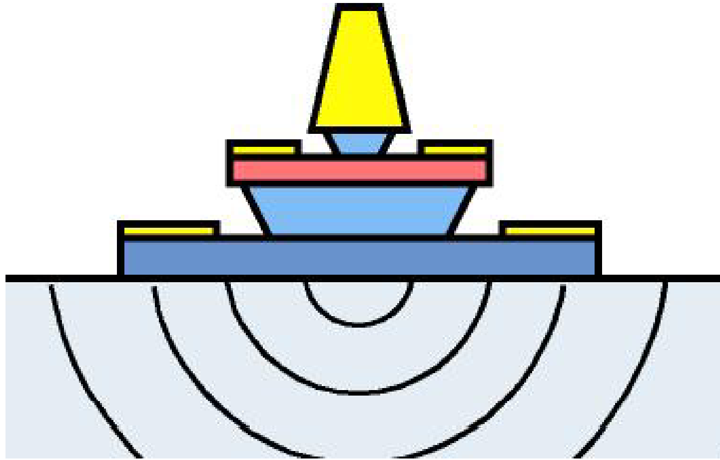


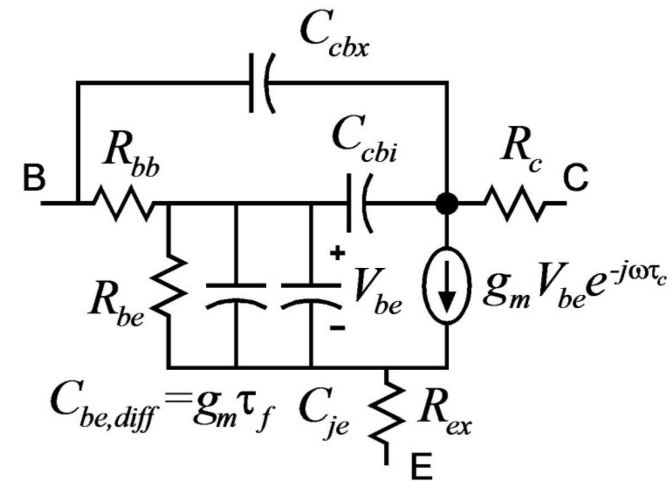
$$R_{be} = \beta / g_m$$

$$g_m = qI_E / nkT$$

$$C_{be} = C_{je} + g_m(\tau_b + \tau_c)$$

Cutoff frequencies in HBTs





$$\frac{1}{2\pi f_\tau} = \tau_{base} + \tau_{collector} + C_{je}\frac{kT}{qI_E} + C_{bc}\left(\frac{kT}{qI_E} + R_{ex} + R_{coll}\right)$$

$$\tau_{base} \approx T_b^2 / 2D_n \qquad \tau_{collector} \approx T_c / 2v_{eff}$$

$$f_{\max} \cong \sqrt{\frac{f_\tau}{8\pi R_{bb} C_{cbi}}}$$

Epitaxial Layer Structure

Epitaxy: InP Emitter, InGaAs Base, InP Collector, Both Junctions Graded

Key Features:

N++ InGaAs emitter contact layer

InP emitter

InGaAs/InAlAs superlattice e/b grade

InGaAs graded base
bandgap or doping grade

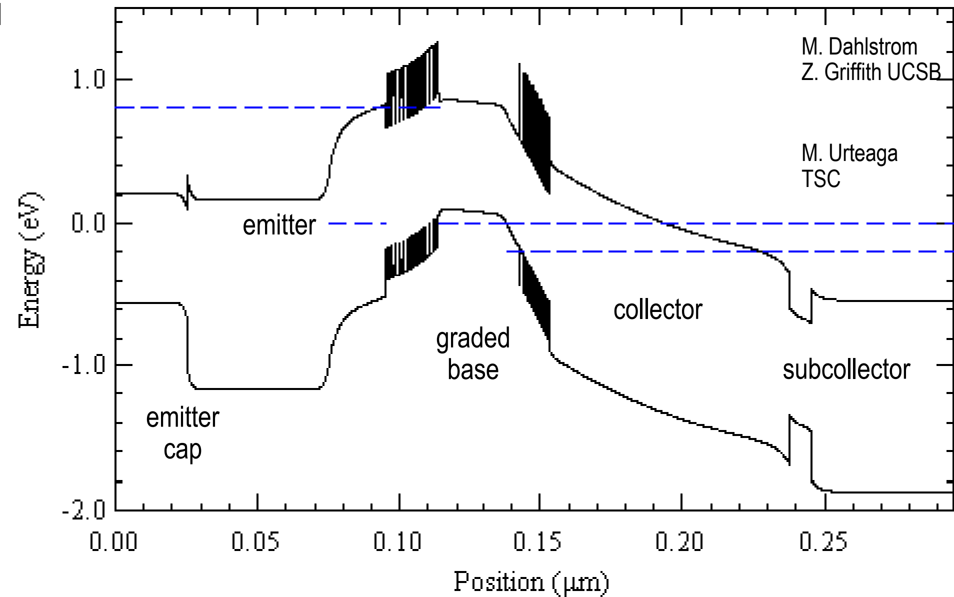
BC setback layer

InGaAs/InAlAs superlattice b/c grade

InP collector

InGaAs etch-stop layer
thin for heat conduction

InP subcollector



Layer	Material	Doping	Thickness (nm)
Emitter cap	$In_{0.53}Ga_{0.47}As$	8×10^{19} cm^{-3}: Si	25
N^+ emitter	InP	8×10^{19} cm^{-3}: Si	50
N^- emitter	InP	1×10^{18} cm^{-3}: Si	20
Emitter-base grade	$In_{0.53}Ga_{0.47}As$ / $In_{0.52}Al_{0.48}As$, 1.6 nm period	N: 1×10^{17} cm^{-3}: Si	24
Emitter-base grade	$In_{0.53}Ga_{0.47}As$ / $In_{0.52}Al_{0.48}As$, 1.5 nm period	P: 8×10^{17} cm^{-3}: C	1
Base	$In_{0.53}Ga_{0.47}As$	P: 7-4 $\times 10^{19}$ cm^{-3}: C	24.5
Setback	$In_{0.53}Ga_{0.47}As$	N: 9×10^{16} cm^{-3}: Si	4.5
Base-collector grade	$In_{0.53}Ga_{0.47}As$ / $In_{0.52}Al_{0.48}As$, 1.6 nm period	N: 9×10^{16} cm^{-3}: Si	15
Pulse doping	InP	6×10^{18} cm^{-3}: Si	3
Collector	InP	N: 2×10^{16} cm^{-3}: Si	82
etch-stop	$In_{0.53}Ga_{0.47}As$	N: 8×10^{19} cm^{-3}: Si	5
Subcollector	InP	N: 8×10^{19} cm^{-3}: Si	~200

Epitaxy with Abrupt BE Junction

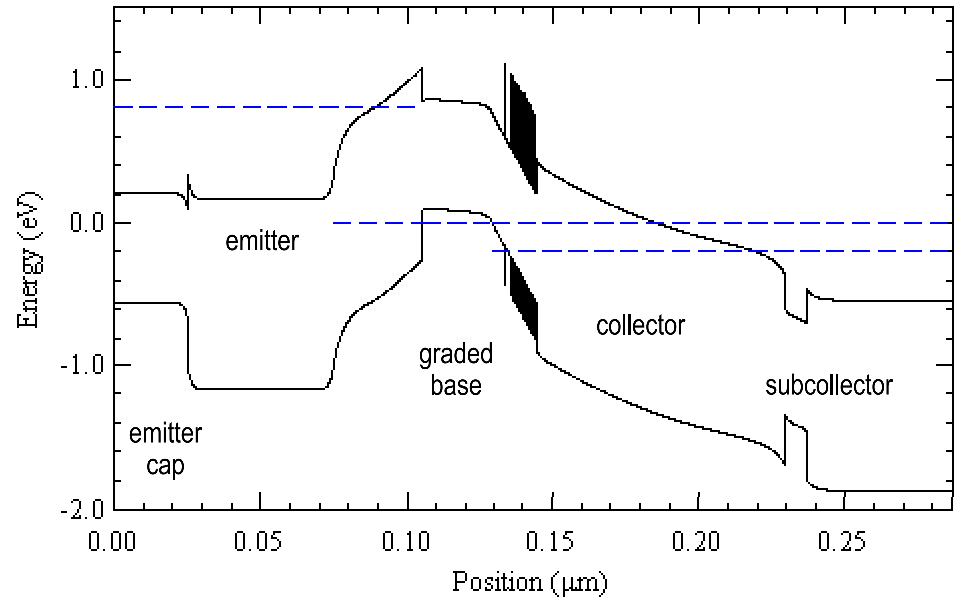
M. Dahlstrom
Z. Griffith
E. Lind

Similar design

Abrupt E/B junction (no e/b grade)

Advantages:
ease of stopping emitter etch on base
→ good base contacts

Disadvantages:
Increased V_{be} .
Cannot make e/b ledge.



Layer	Material	Doping	Thickness (nm)
Emitter cap	$In_{0.53}Ga_{0.47}As$	8×10^{19} cm^{-3}: Si	25
N^+ emitter	InP	8×10^{19} cm^{-3}: Si	50
N^- emitter	InP	1×10^{18} cm^{-3}: Si	20
N- emitter	InP	N: 1×10^{17} cm^{-3}: Si	24
Emitter-base grade	$In_{0.53}Ga_{0.47}As$ / $In_{0.52}Al_{0.48}As$, 1.5 nm period	P: 8×10^{17} cm^{-3}: C	1
Base	$In_{0.53}Ga_{0.47}As$	P: 7-4 $\times 10^{19}$ cm^{-3}: C	24.5
Setback	$In_{0.53}Ga_{0.47}As$	N: 9×10^{16} cm^{-3}: Si	4.5
Base-collector grade	$In_{0.53}Ga_{0.47}As$ / $In_{0.52}Al_{0.48}As$, 1.6 nm period	N: 9×10^{16} cm^{-3}: Si	15
Pulse doping	InP	6×10^{18} cm^{-3}: Si	3
Collector	InP	N: 2×10^{16} cm^{-3}: Si	82
etch-stop	$In_{0.53}Ga_{0.47}As$	N: 8×10^{19} cm^{-3}: Si	5
Subcollector	InP	N: 8×10^{19} cm^{-3}: Si	~200

E . Lind
Z. Griffith

Alternative Grades for Thinner Epitaxy

Common Grade in Literature
InGaAs/InAlAs
18 nm thick, 1.5 nm period

DHBT 42			
Thickness (nm)	***Material***	***Doping (cm^{-3})***	***Description***
10	$In_{0.53}Ga_{0.47}As$	2×10^{17} : Si	Setback
18	InGaAs/InAlAs	2×10^{17} : Si	Superlattice grade
3	InP	3.4×10^{18} : Si	Delta Doping

Sub-monolayer Grade
0.15 nm InAlAs,
(0.15 to 0.165 nm InGaAs)
10.8 nm thick

DHBT 41			
Thickness (nm)	***Material***	***Doping (cm^{-3})***	***Description***
5	$In_{0.53}Ga_{0.47}As$	2×10^{17} : Si	Setback
10.8	InGaAs/InAlAs	2×10^{17} : Si	Sub-monolayer grade
3	InP	6.2×10^{18} : Si	Delta doping

Strained $In_xGa_{1-x}As$ Grade

InGaAs/GaAs 6 nm

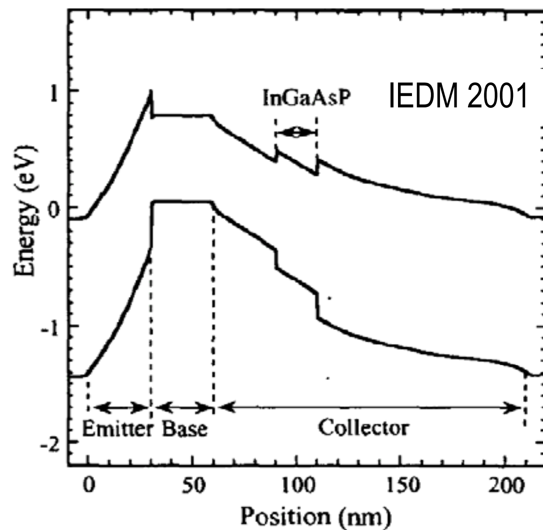
DHBT 40			
Thickness (nm)	***Material***	***Doping (cm^{-3})***	***Description***
1	$In_{0.53}Ga_{0.47}As$	3×10^{17} : Si	Setback
6	InGaAs→GaAs	3×10^{17} : Si	Strained grade
4	InP	6.2×10^{18} : Si	Delta doping

Other Methods of Grading the Junctions

InGaAs/InGaAsP/InP grade

InP/InGaAs DHBTs with 341-GHz f_T at high current density of over 800 kA/cm^2

Minoru Ida, Kenji Kurishima, Noriyuki Watanabe, and Takatomo Enoki



-suitable for MOCVD growth

- excellent results

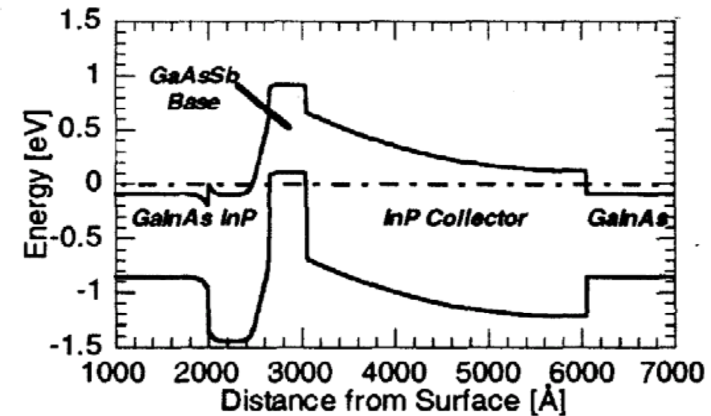
InP/GaAsSb/InP DHBT

11th International Conference on Indium Phosphide and Related Materials
16-20 May 1999 Davos, Switzerland

TuA1-3

InP/GaAsSb/InP DOUBLE HETEROJUNCTION BIPOLAR TRANSISTORS WITH HIGH CUT-OFF FREQUENCIES AND BREAKDOWN VOLTAGES

N. Matine, M. W. Dvorak, X. G. Xu, S. P. Watkins, and C. R. Bolognesi



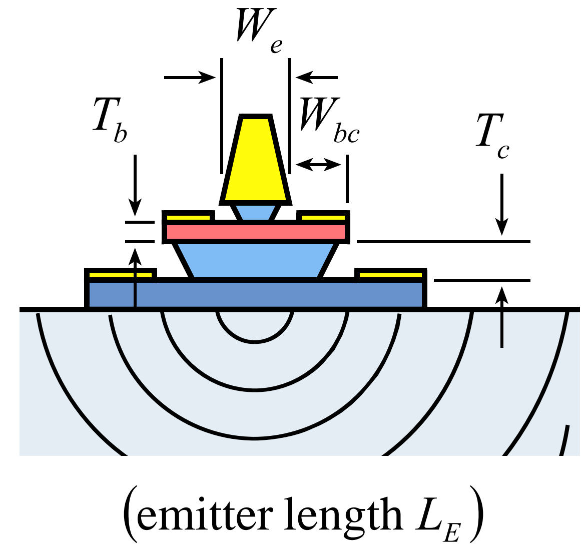
- does not need B/C grading

- E/B band alignment through
  GaAsSb alloy ratio (strain)

  or InAlAs emitter

Transport Analysis

Approximate Carrier Transit Times



(emitter length L_E)

Base Transit Time

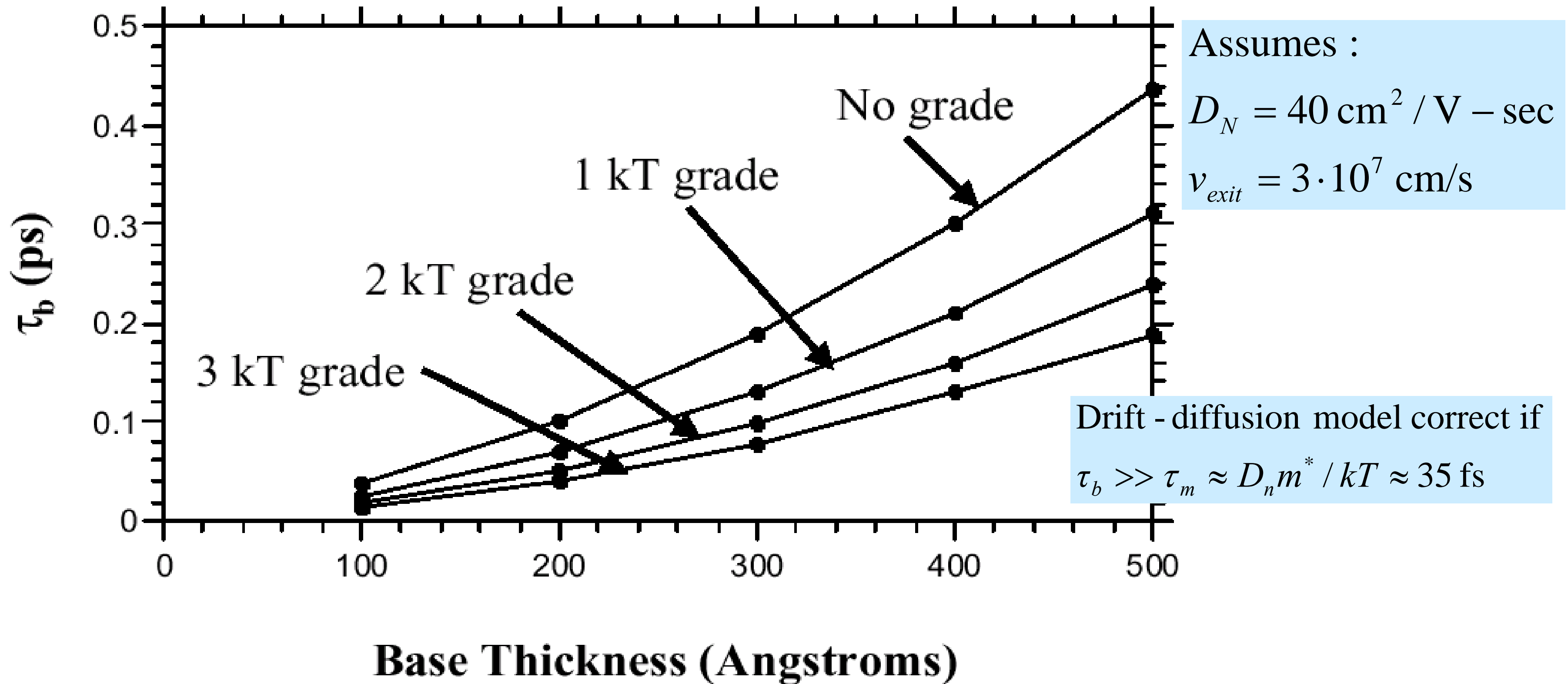
$$\tau_b \approx T_b^2 / 2D_n + T_b / v_{exit}$$

Collector Transit Time

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

Base Transit Time with Graded Base

Dino Mensa



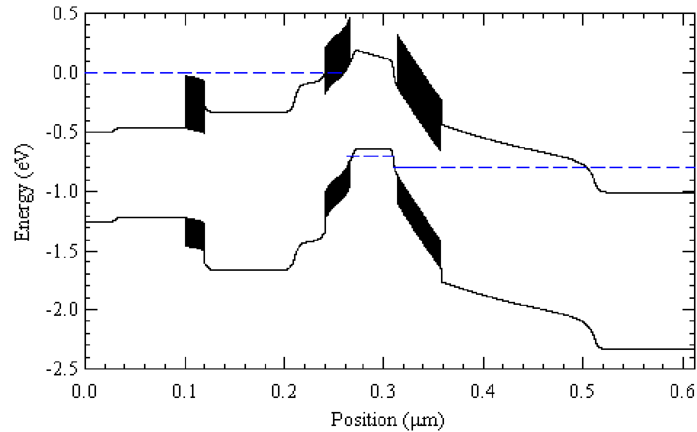
$$\tau_b = W_b L_g / D_n - \left(L_g^2 / D_n - L_g / v_{sat}\right)\left(1 - e^{-W_b / L_g}\right)$$

where L_g is the grading length :

$$L_g = W_b\left(kT / \Delta E_g\right)$$

Base Transit Time: Grading Approaches

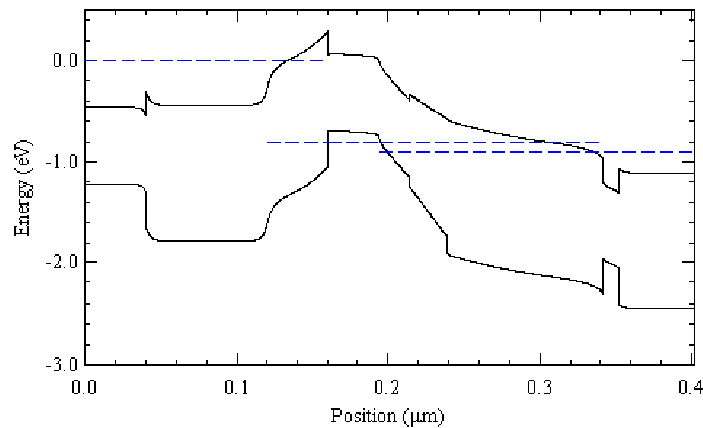
Dino Mensa
Miguel Urteaga
Mattias Dahlström



Compositional grading:
strained graded InGaAs base

52 meV potential drop :

$In_{0.455}Ga_{0.545}As \leftrightarrow In_{0.53}Ga_{0.47}As$ (strained)

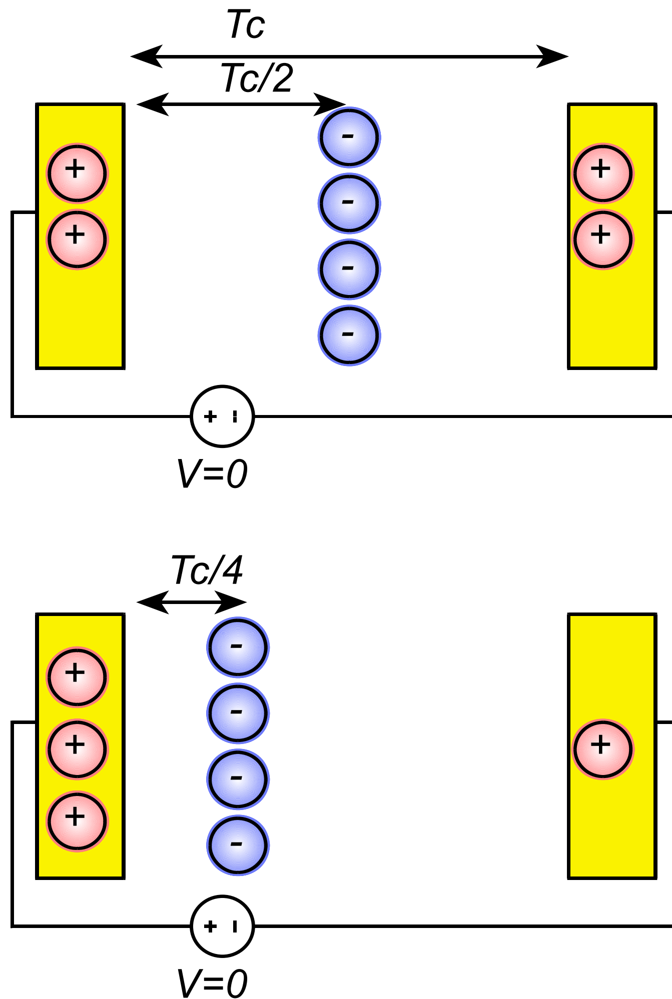


Doping grading
unstrained $In_{0.53}Ga_{0.47}As$ base

Carbon doping varied from :

$\sim 8 \cdot 10^{19}/cm^3$ to $\sim 5 \cdot 10^{19}/cm^3$

Collector Transit Time

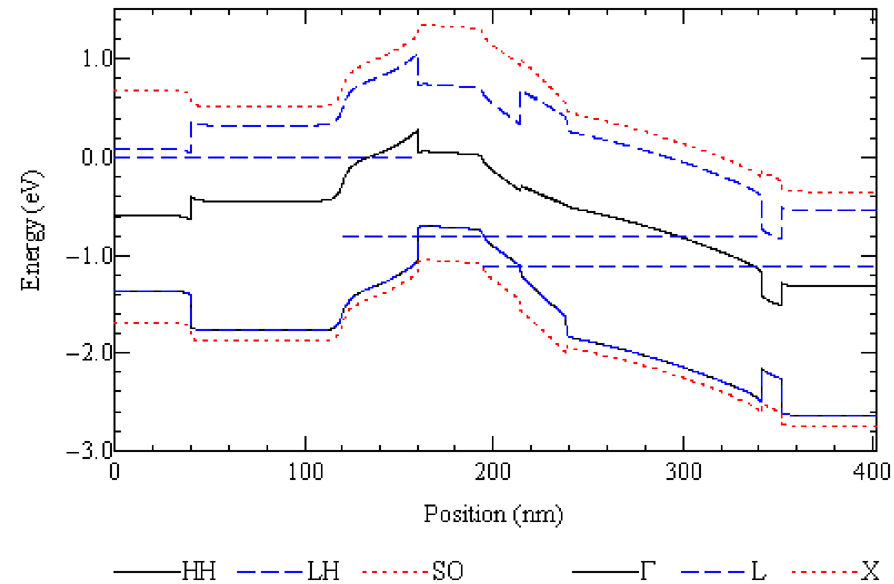


From elementary electrostatics (refer to sketch)

$$\tau_c = \int_0^{T_C} \frac{(1 - x/T_c)}{v(x)} dx \equiv \frac{T_c}{2v_{eff}}$$

τ_c is more sensitive to velocity near base.

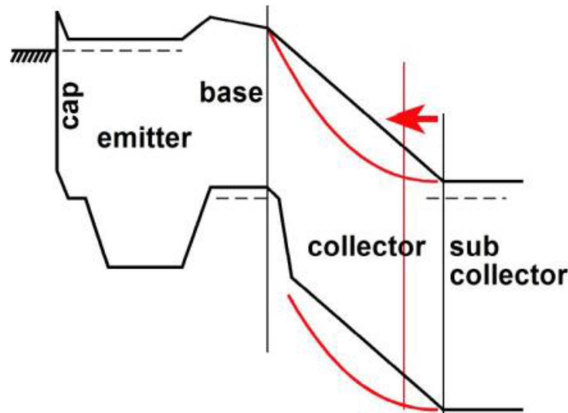
Fortuitous, as initial velocity is high, then decreases due to Γ - L scattering.



From best fit to RF data, or from Kirk current density vs. collector voltage:

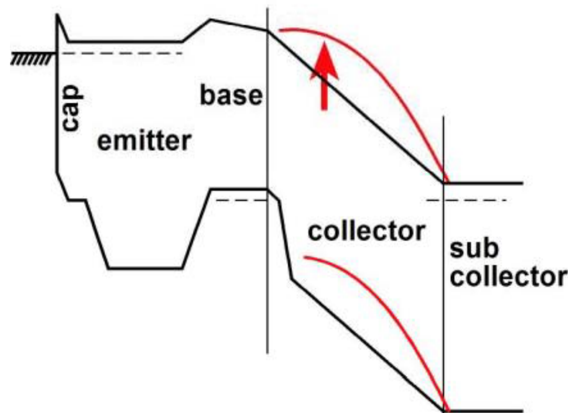
InP: $\approx 3.5 \cdot 10^7$ cm/s for ~ 70 - 200 nm layers

Space-Charge Limited Current Density → C_{cb} charging time



Collector Depletion Layer Collapse

$$V_{cb,\min} + \phi > +(qN_d)(T_c^2 / 2\varepsilon)$$



Collector Field Collapse (Kirk Effect)

$$V_{cb} + \phi > +(J / v_{sat} - qN_d)(T_c^2 / 2\varepsilon)$$

$$\Rightarrow \quad J_{\max} = 2\varepsilon v_{eff} (V_{cb} + V_{cb,\min} + 2\phi) / T_c^2$$

Note that $V_{be} \cong \phi$, hence $(V_{cb} + \phi) \cong V_{ce}$

$$C_{cb} \Delta V_{LOGIC} / I_C = (\varepsilon A_{\text{collector}} / T_c)(\Delta V_{LOGIC} / I_C) = \frac{\Delta V_{LOGIC}}{(V_{CE} + V_{CE,\min})} \left(\frac{A_{\text{collector}}}{A_{\text{emitter}}} \right) \left(\frac{T_C}{2 v_{eff}} \right)$$

Collector capacitance charging time scales linearly with collector thickness if $J = J_{max}$

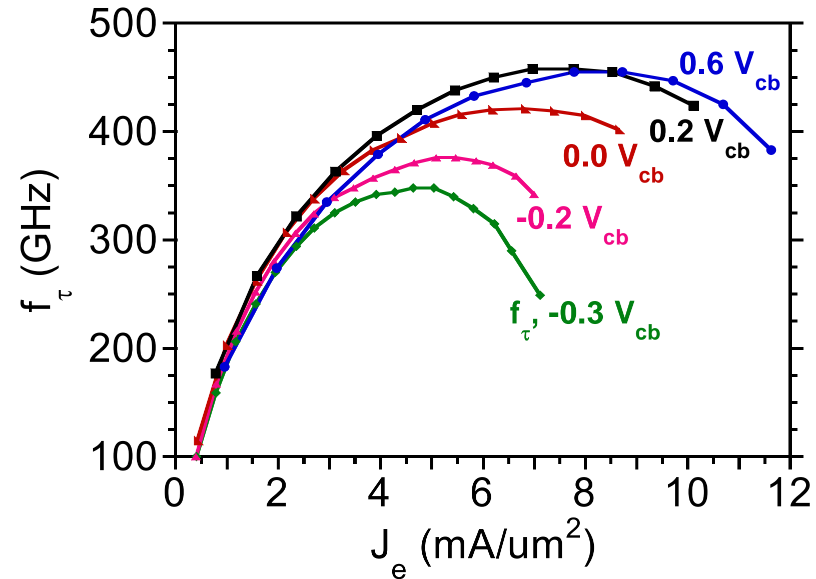
Space-Charge-Limited Current (Kirk effect) in DHBTs

Decrease in f_τ and f_{max} at high J

Kirk - effect threshold increases

with increased V_{ce}

$$J_{max} = 2\varepsilon v_{sat}(V_{cb} + V_{cb,min} + 2\phi)/T_c^2$$
$$\cong 2\varepsilon v_{sat}(V_{ce} + V_{ce,min})/T_c^2$$

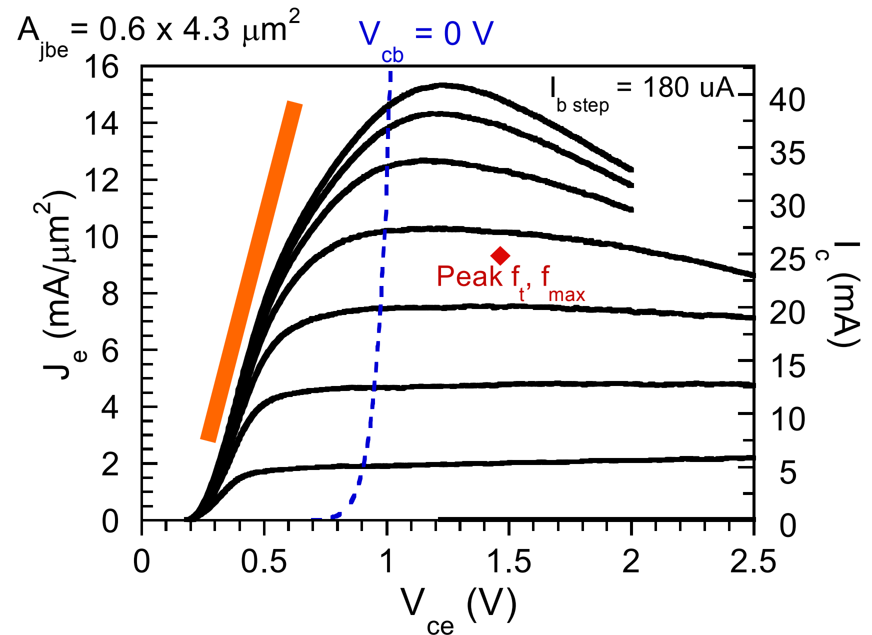


Increase in $V_{ce,sat}$ with increased J

$$\frac{dV_{ce}}{dI_c} = R_{space-charge} = \frac{T_c^2}{2\varepsilon v_{sat} A_{effective}}$$

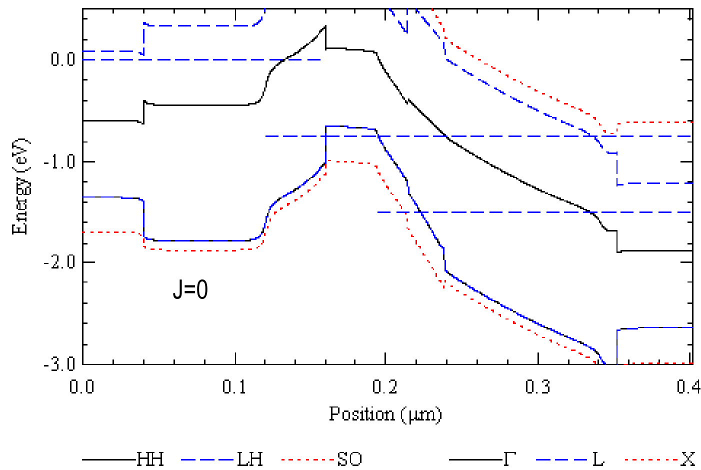
where the effective collector current flux area is

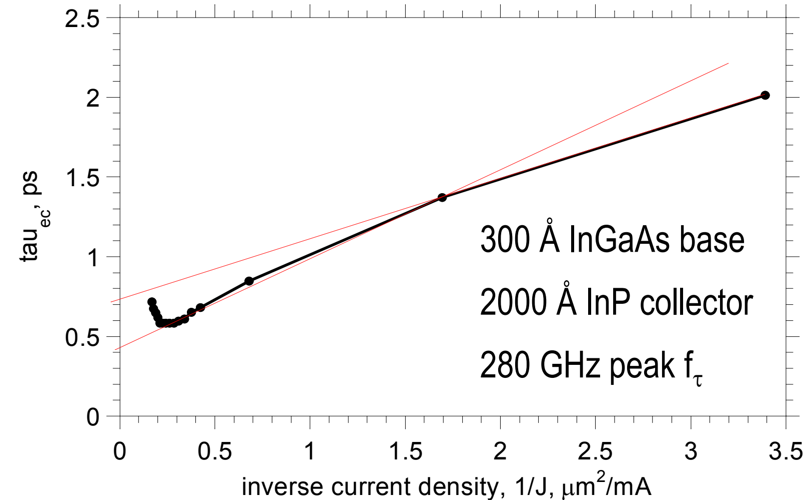
$$A_{effective} \approx L_E(W_E + 2T_C)$$

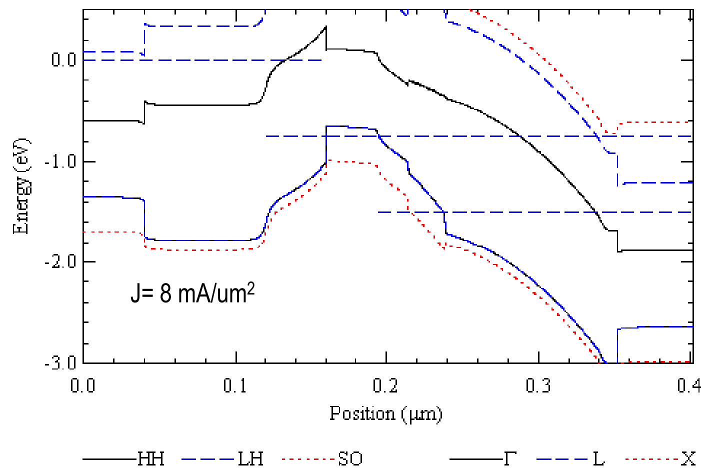


T. Ishibashi

Current-induced Collector Velocity Overshoot







Increased current reduces Γ - L scattering,

increases $v(x)$ in early part of collector

$\Rightarrow$ reduced collector transit time

$$Q_{base} = I_c \cdot \int_0^{T_C} \frac{(1 - x/T_c)}{v(x)} dx \text{ is not exactly proportional to } I_c$$

correct definition of collector transit time is

$$\tau_c = \frac{\partial Q_{base}}{\partial I_c} \quad \text{not} \quad \tau_c = \frac{Q_{base}}{I_c}$$

Nakajima, H. "A generalized expression for collector transit time of HBTs taking account of electron velocity modulation," Japanese Journal of Applied Physics, vo. 36, Feb. 1997, pp. 667-668

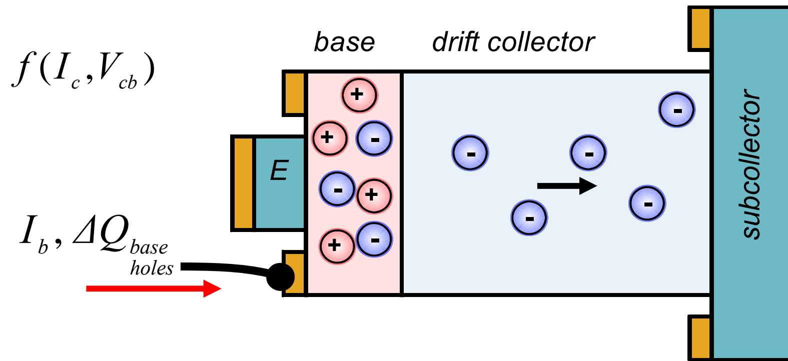
CAUTION : observed nonlinear τ_{ec} variation is also in part due to modulation in emitter ideality factor with bias current ($1/g_m$ often does not vary as $R_{ex} + nkT/qI_E$), and due to variation of C_{je} with bias.

Transit time Modulation Causes C_{cb} Modulation

$$Q_{\substack{base\\holes}} = \text{constant} + Q_{\substack{base\\electrons}} + \int_0^{T_c} qn(x)A(1 - x/T_c)dx + V_{bc}\varepsilon A/T_c = f(I_c, V_{cb})$$

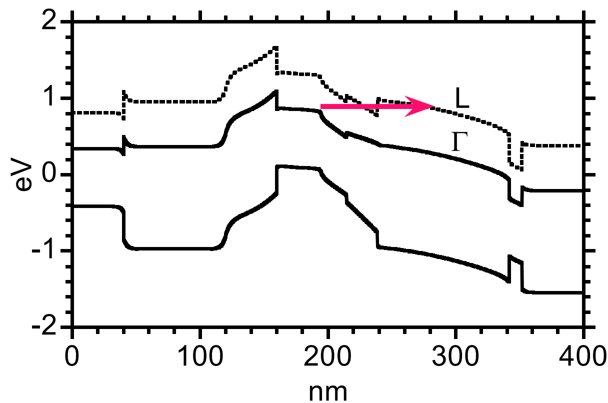
$$C_{cb} \equiv -\frac{\partial Q_{base\ holes}}{\partial V_{cb}} \qquad \tau_f \equiv \frac{\partial Q_{base\ holes}}{\partial I_c} \Rightarrow \frac{\partial C_{cb}}{\partial I_c} = -\frac{\partial \tau_f}{\partial V_{cb}}$$

Camnitz and Moll, Betser & Ritter, *D. Root*



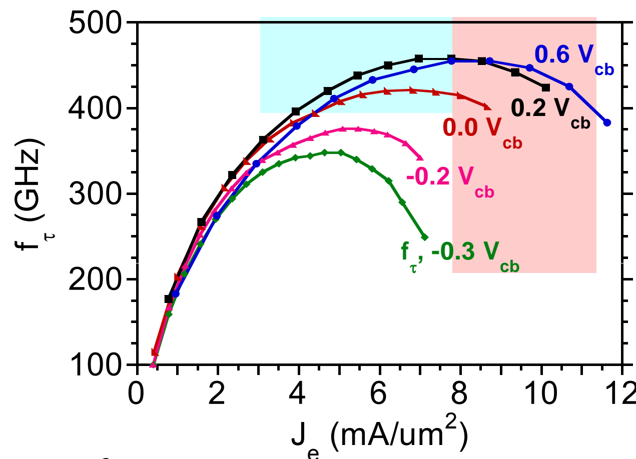
Collector Velocity Modulation :

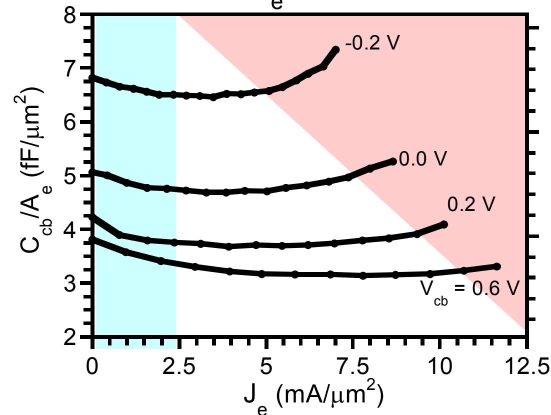
$$\partial \tau_f / \partial V_{cb} > 0 \Rightarrow \partial C_{cb} / \partial I_c < 0$$



Increase in τ_c with $V_{cb} \rightarrow$ reduced C_{cb}

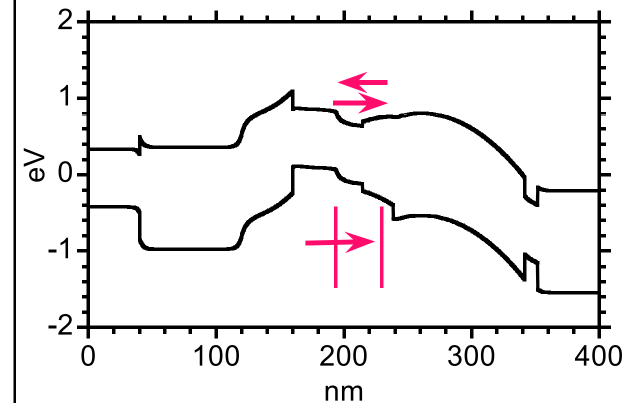
- ***strong effect in InGaAs SHBTs***
- ***weak effect in InP DHBTs***





Kirk Effect :

$$\partial \tau_f / \partial V_{cb} < 0 \Rightarrow \partial C_{cb} / \partial I_c > 0$$

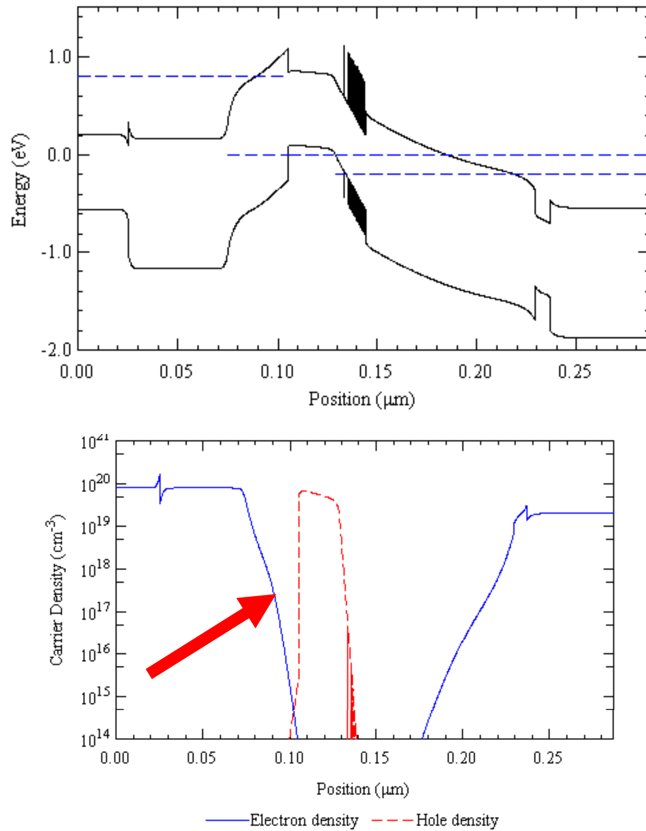


Increase in C_{cb} is due to both

- ***base pushout into collector***
- ***and modulation of τ_b by V_{cb}***

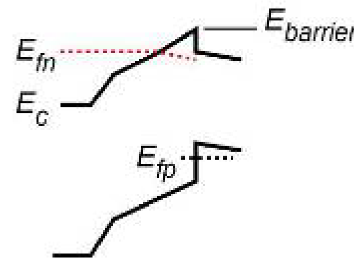
Emitter-Base Junction Effects

Space-charge storage



need thin layer to avoid substantial charge storage delays

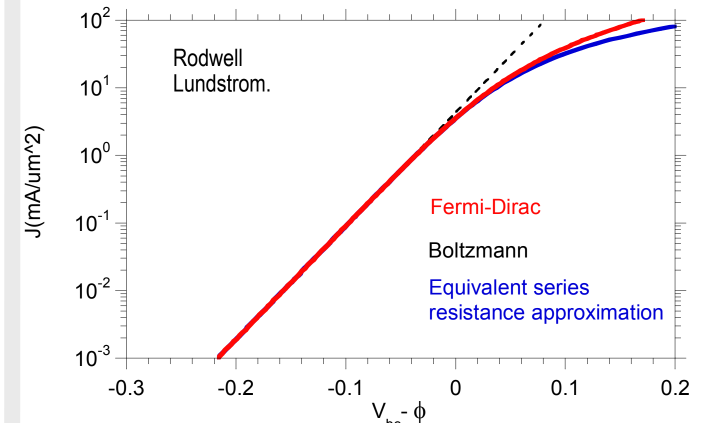
Voltage drops in depletion region



$$\frac{\partial E_{fn}(x)}{\partial x} = \frac{-J}{q\mu_n n(x)}$$

need thin layer & high electron density

Electron degeneracy contributes 1 Ω - μm^2 equivalent series resistance



Highly degenerate limit

$$J = \frac{qm^*(E_f - E_c)^2}{2\pi^2\hbar^3}$$

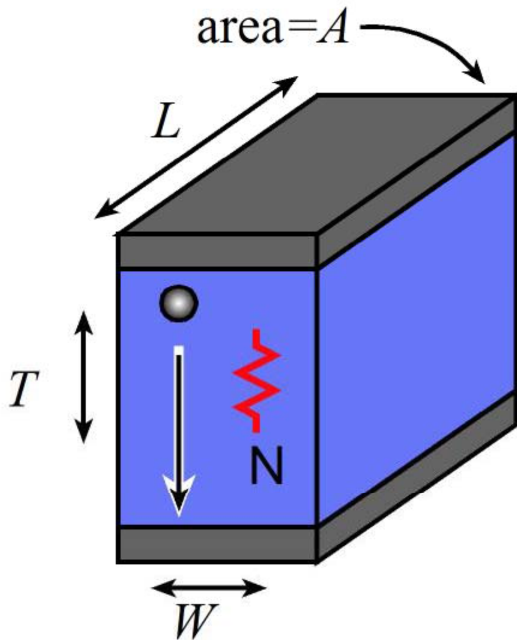
$$= \left(130\frac{\text{mA}}{\mu\text{m}^2}\right)\left(\frac{E_f - E_c}{0.1\,\text{eV}}\right)^2$$

for InP emitter ($m^*/m_0 = 0.08$).

RC parasitics

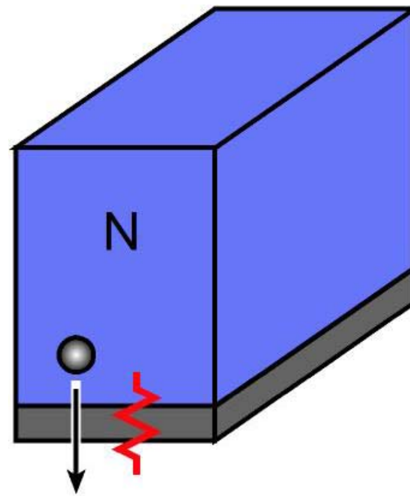
Simple Device Physics: Resistance

bulk resistance



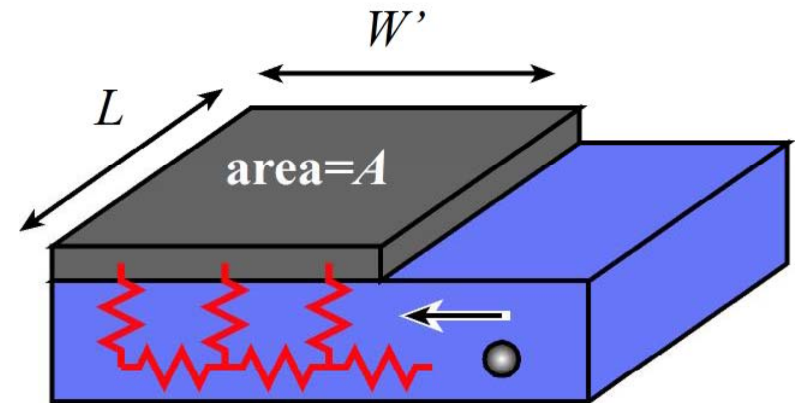
$$R = \frac{\rho_{bulk} \cdot T}{A}$$

contact resistance -perpendicular



$$R = \frac{\rho_{contact}}{A}$$

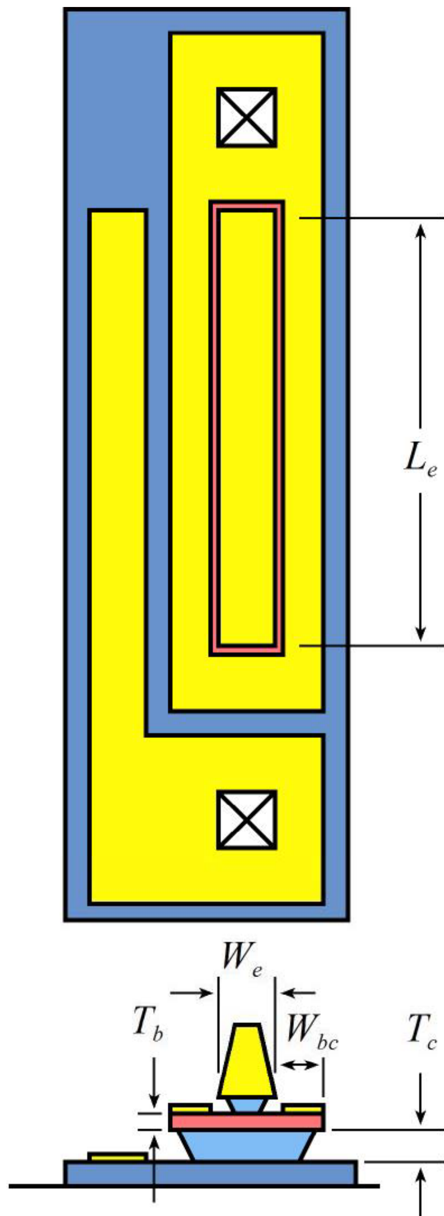
contact resistance - parallel

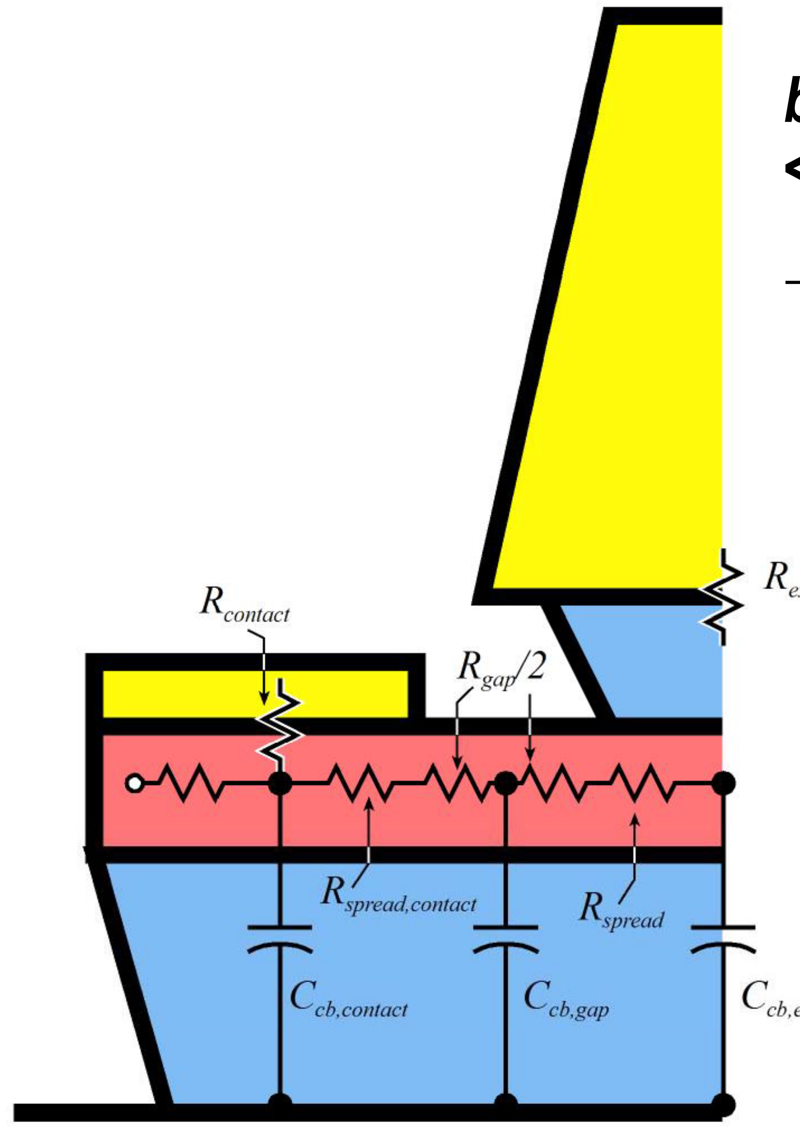


$$R = \frac{\rho_{contact}}{A} + \rho_{sheet} \cdot \frac{W'}{3L}$$

Good approximation for contact widths less than 2 transfer lengths.

HBT RC Parasitics



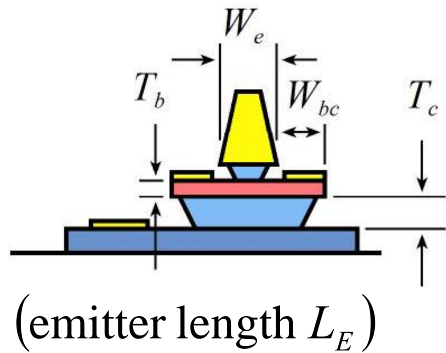


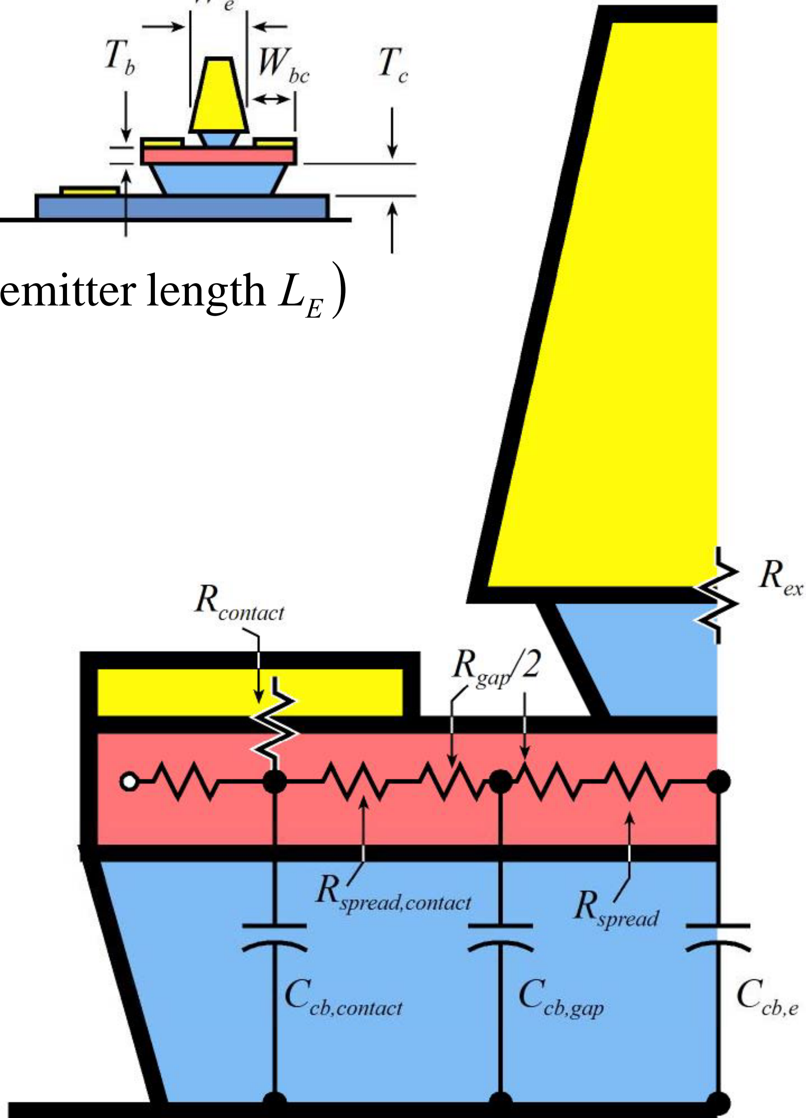
base contact width < 2 transfer lengths

→ simple analysis

Limiting case of Pulfrey / Vaidyanathan f_{max} model.

HBT RC Parasitics





$$R_{ex} = \rho_{contact,emitter} / A_{emitter}$$

$$R_{spread} = \rho_s W_e / 12 L_E$$

$$R_{gap} = \rho_s W_{gap} / 4 L_E$$

$$R_{spread,contact} = \rho_s W_{bc} / 6 L_E$$

$$R_{contact} = \rho_{contact,base} W_{bc} / A_{base_contacts}$$

$$C_{cb,e} = \varepsilon A_{emitter} / T_c$$

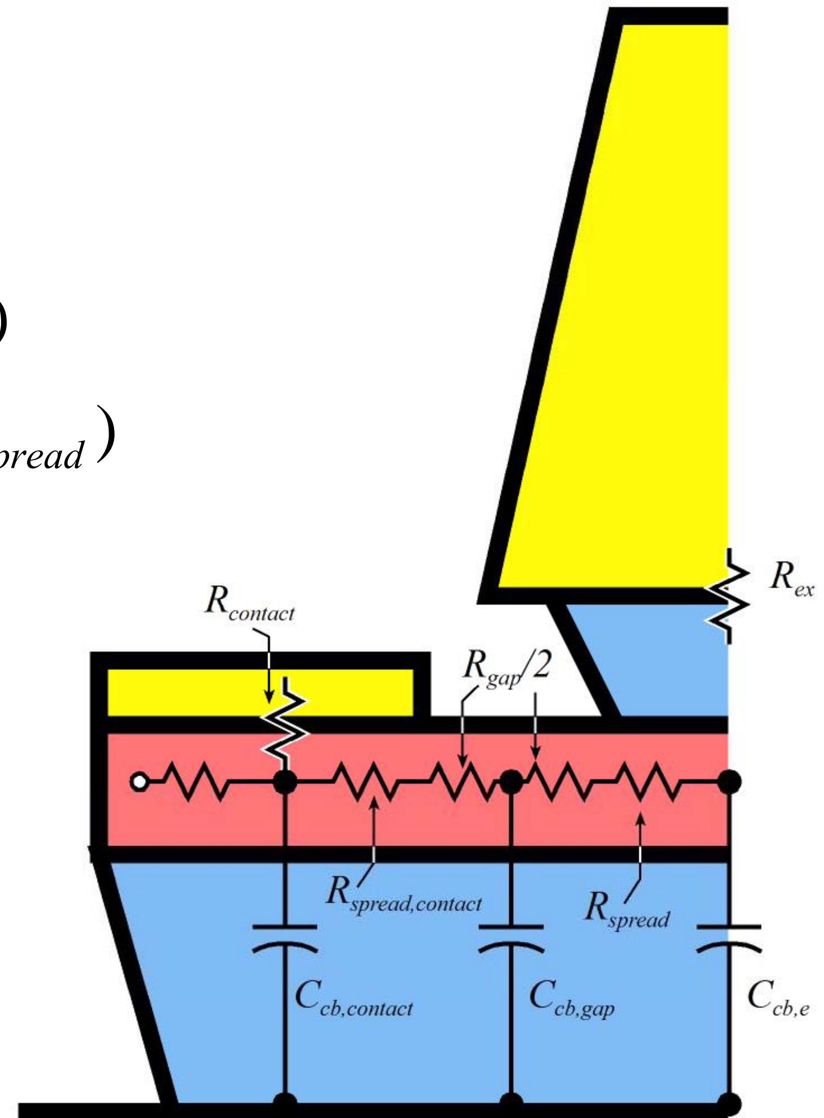
$$C_{cb,gap} = \varepsilon A_{gap} / T_c$$

$$C_{cb,contact} = \varepsilon A_{base_contacts} / T_c$$

Base-Collector Time Constant & Fmax.

$$f_{\max} \cong \sqrt{\frac{f_\tau}{8\pi R_{bb} C_{cbi}}} \text{ where}$$

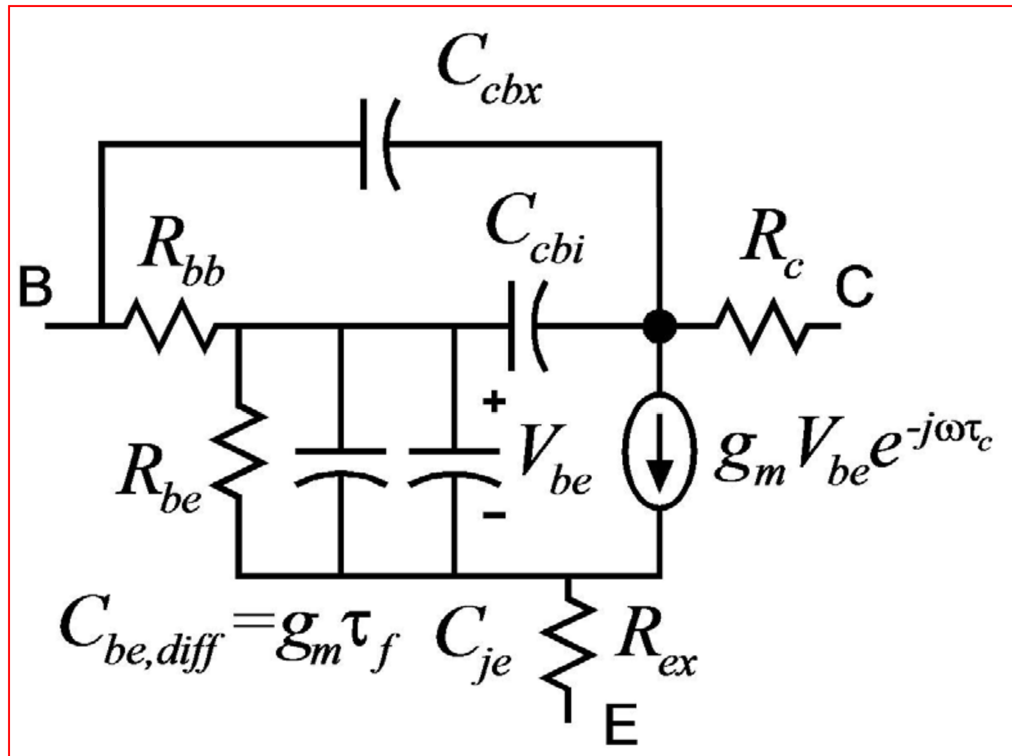
$$\tau_{cb} = R_{bb} C_{cbi} = C_{cb,contact} R_{contact}$$
$$+ C_{cb,gap}(R_{contact} + R_{spread,contact} + R_{gap}/2)$$
$$+ C_{cb,e}(R_{contact} + R_{spread,contact} + R_{gap} + R_{spread})$$

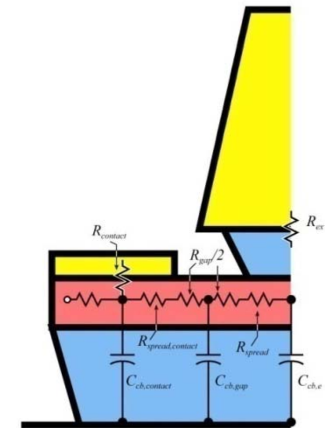


Relationship to Equivalent Circuit Model

$$C_{cbx} + C_{cbi} = C_{cb,e} + C_{cb,gap} + C_{cb,contact}$$

$$R_{bb} = R_{spread} + R_{gap} + R_{contact,spread} + R_{contact}$$





$$R_{bb}C_{cbi} = C_{cb,contact}R_{contact} + C_{cb,gap}(R_{contact} + R_{spread,contact} + R_{gap}/2)$$
$$+ C_{cb,e}(R_{contact} + R_{spread,contact} + R_{gap} + R_{spread})$$

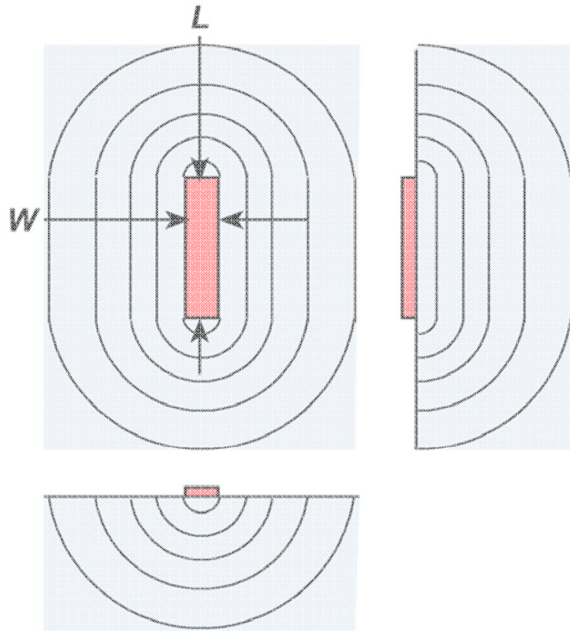
Device Design

Device Scaling

Simple Device Physics: Thermal Resistance

Exact

Carslaw & Jaeger 1959



$$R_{th} = \frac{1}{\pi K_{th} L} \sinh^{-1}\left(\frac{L}{W}\right) + \frac{1}{\pi K_{th} W} \sinh^{-1}\left(\frac{W}{L}\right)$$

Long, Narrow Stripe

HBT Emitter, FET Gate

$$R_{th} \cong \underbrace{\frac{1}{\pi K_{th} L} \ln\left(\frac{L}{W}\right)}_{\text{cylindrical heat flow near junction}} + \underbrace{\frac{1}{\pi K_{th} L}}_{\text{spherical heat flow far from junction}}$$

Square (L by L)

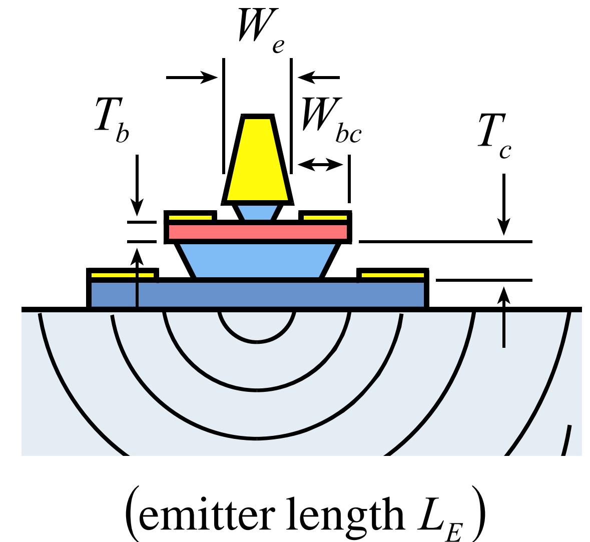
IC on heat sink

$$R_{th} \cong \underbrace{\frac{1}{4 K_{th} L}}_{\text{planar heat flow near surface}} + \underbrace{\frac{1}{\pi K_{th} L}}_{\text{spherical heat flow far from surface}}$$

Bipolar Transistor Design

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

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



(emitter length L_E)

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

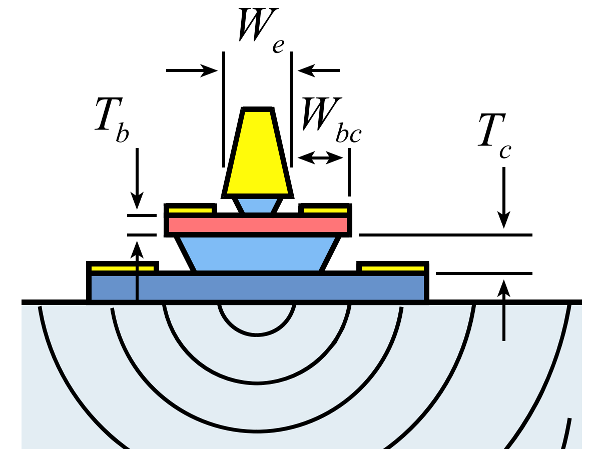
$$I_{c,\max} \propto v_{sat} A_e (V_{\text{ce,operating}} + V_{\text{ce,punch-through}}) / T_c^2$$

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

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

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

Bipolar Transistor Design: Scaling



(emitter length L_E)

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

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

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

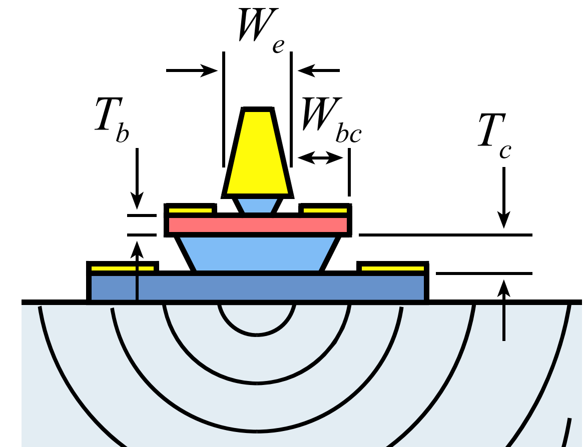
$$I_{c,\max} \propto v_{sat} A_e (V_{\text{ce,operating}} + V_{\text{ce,punch-through}}) / T_c^2$$

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

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

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

Bipolar Transistor Scaling Laws



Changes required to double transistor bandwidth:

(emitter length L_E)

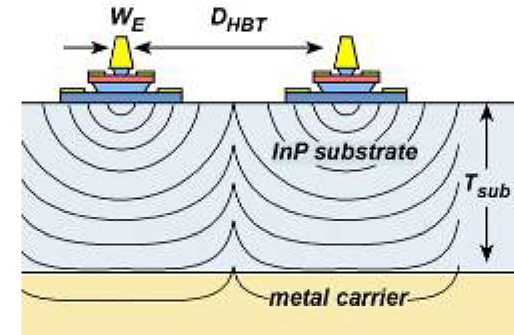
parameter	change
collector depletion layer thickness	decrease 2:1
base thickness	decrease 1.414:1
emitter junction width	decrease 4:1
collector junction width	decrease 4:1
emitter contact resistance	decrease 4:1
current density	increase 4:1
base contact resistivity	decrease 4:1

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

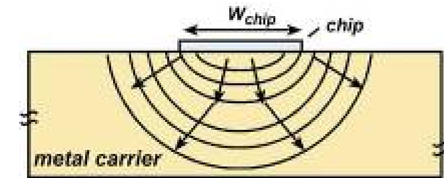
Thermal Resistance Scaling : Transistor, Substrate, Package

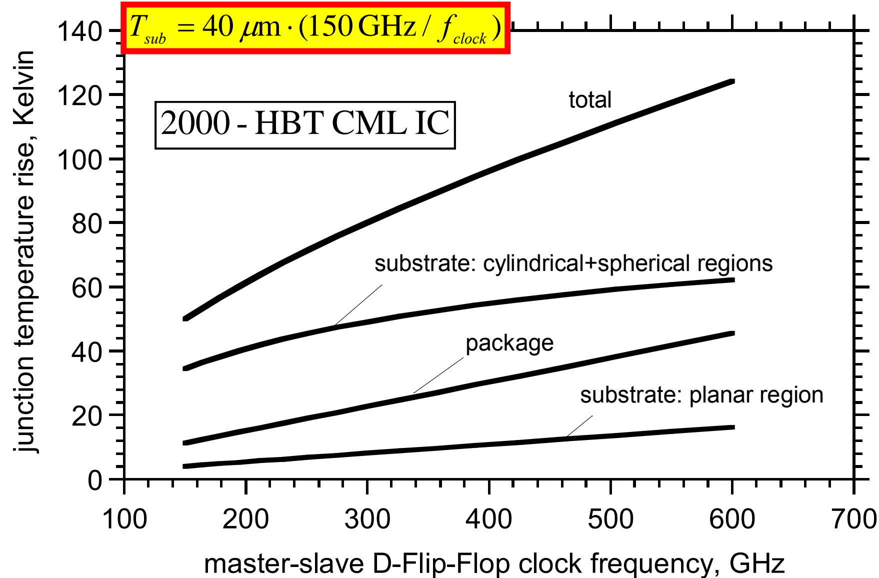
$$\Delta T_{\text{substrate}} \cong \underbrace{\frac{P}{\pi K_{InP} L_E} \ln\left(\frac{L_e}{W_e}\right)}_{\text{cylindrical heat flow near junction}} + \underbrace{\frac{P}{\pi K_{InP}}\left(\frac{1}{L_E} - \frac{1}{D}\right)}_{\text{spherical flow for } r > L_e} + \underbrace{\frac{P}{K_{InP}} \cdot \left(\frac{T_{sub} - D/2}{D^2}\right)}_{\text{planar flow for } r > D_{HBT}/2}$$

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



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





Wiring lenghts scale as 1/bandwidth. Power density, scales as $(\text{bandwidth})^2$.

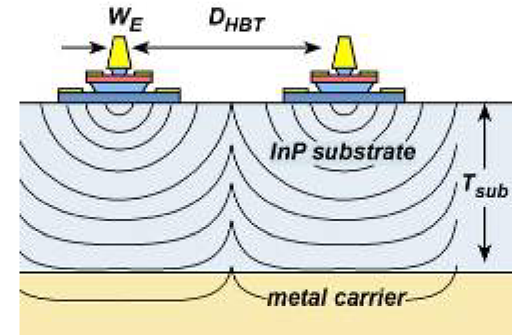
Thermal Resistance Scaling : Transistor, Substrate, Package

cylindrical heat flow near junction — increases logarithmically

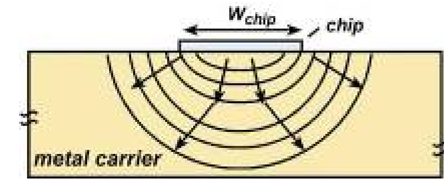
spherical flow for $r > L_e$ — insignificant variation

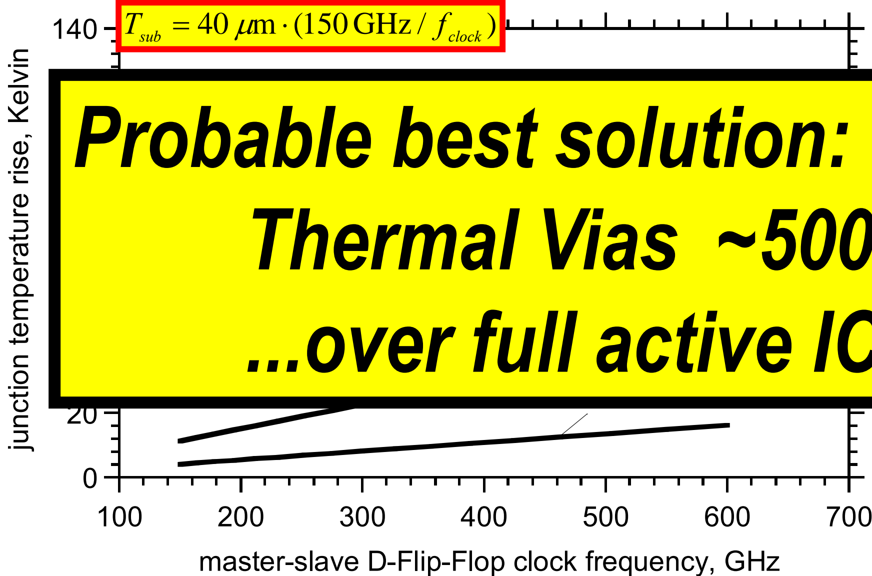
planar flow for $r > D_{HBT}/2$ — increases quadratically if T_{sub} is constant

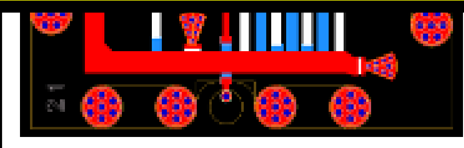
$$\Delta T_{\text{substrate}} \cong \frac{P}{\pi K_{InP} L_E} \ln\left(\frac{L_e}{W_e}\right) + \frac{P}{\pi K_{InP}}\left(\frac{1}{L_E} - \frac{1}{D}\right) + \frac{P}{K_{InP}} \cdot \left(\frac{T_{sub} - D/2}{D^2}\right)$$



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







(bandwidth)^2.

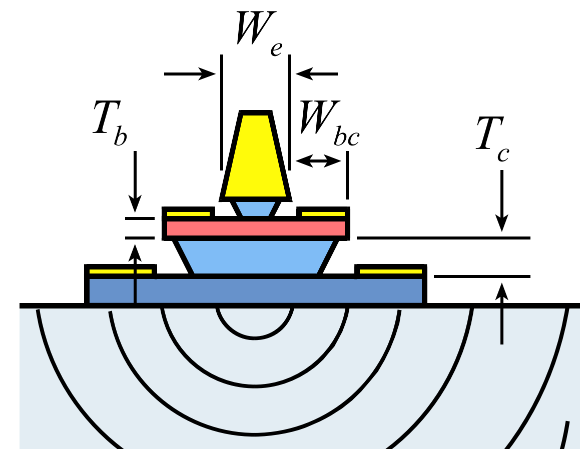
Probable best solution:

Thermal Vias ~500 nm below InP subcollector

...over full active IC area.

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



Can we make a 1 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

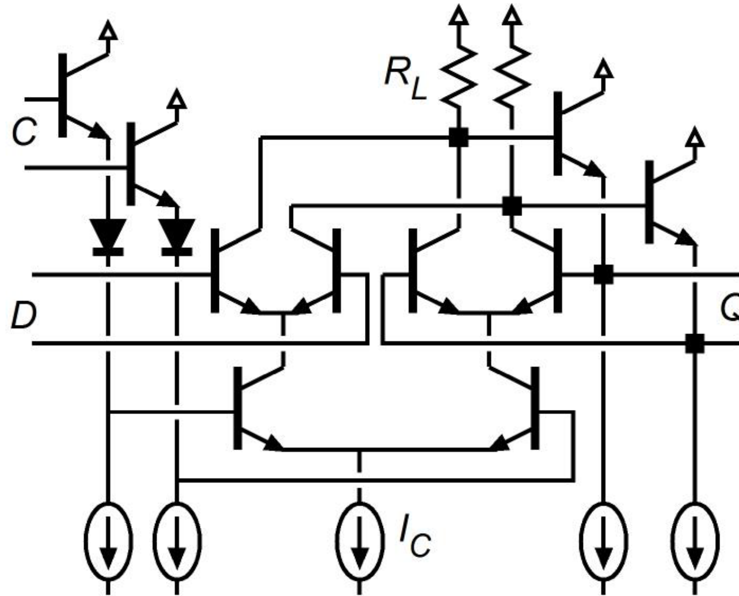
(also need better Ohmic contacts)

	InP	SiGe	
emitter	64	18	nm width
	2	**1.2**	$\Omega\cdot\mu m^2$ access ρ
base	64	56	nm contact width,
	2.5	**1.4**	$\Omega\cdot\mu m^2$ contact ρ
collector	53	**15**	nm thick
	36	125	$mA/\mu m^2$
	2.75	**???**	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 2:1 wider than junctions

HBT Design For Digital & Mixed-Signal Performance



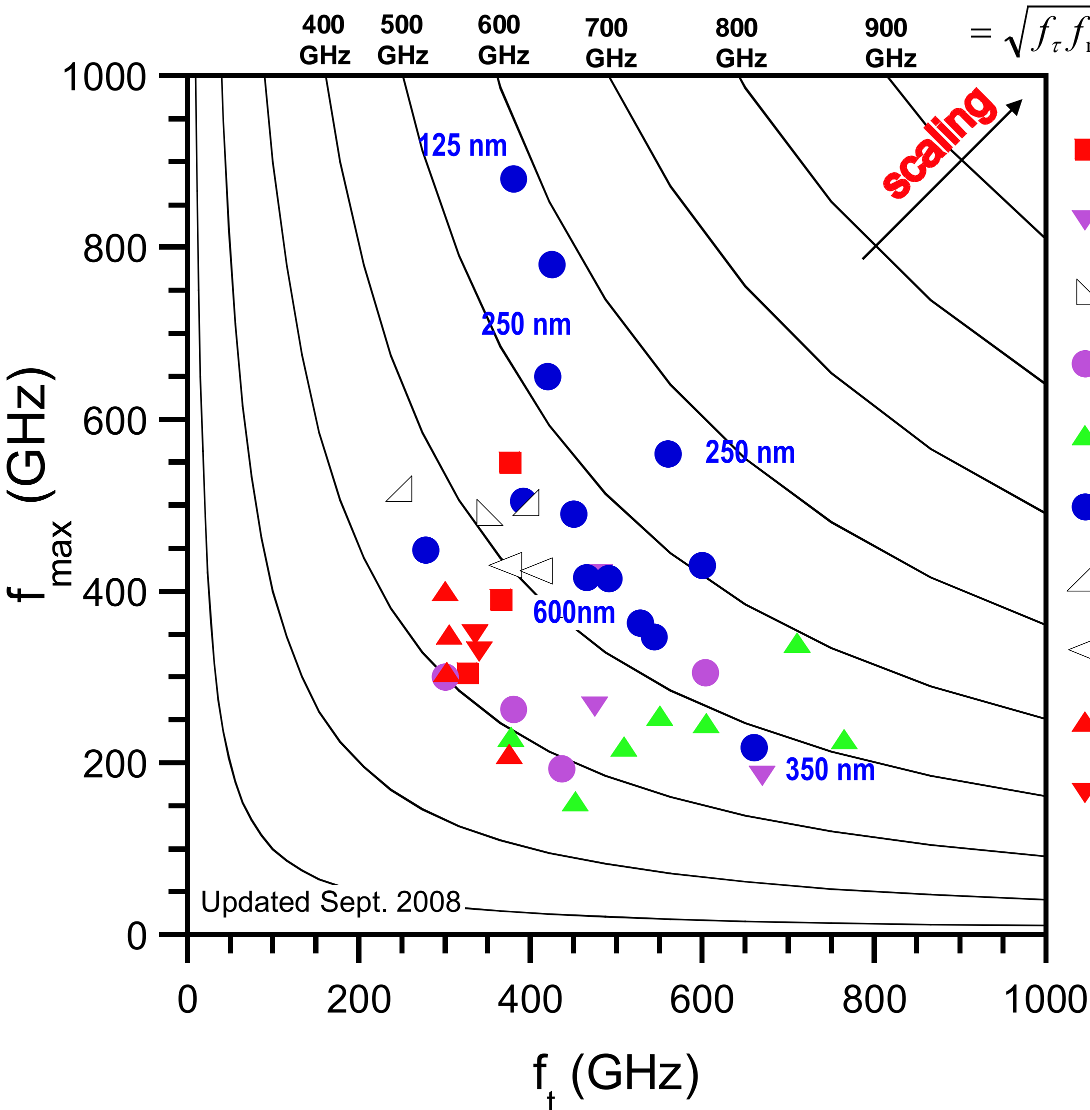
Parameter	scaling law	Gen. 3 (256 nm)	Gen. 4 (128 nm)	Gen 5 (64 nm)	Gen 5 (32 nm)
MS-DFF speed	γ^1	240 GHz	330 GHz	480 GHz	660 GHz
Amplifier center frequency	γ^1	430 GHz	660 GHz	1.0 THz	1.4 THz
Emitter Width	$1/\gamma^2$	256 nm	128 nm	64 nm	32 nm
Resistivity	$1/\gamma^2$	8 Ω-μm²	4 Ω-μm²	2 Ω-μm²	1 Ω-μm²
Base Thickness	$1/\gamma^{1/2}$	250 Å	212 Å	180 Å	180 Å
Contact width	$1/\gamma^2$	175 nm	120 nm	60 nm	30 nm
Doping	γ^0	$7\ 10^{19}$ /cm²	$7\ 10^{19}$ /cm²	$7\ 10^{19}$ /cm²	$7\ 10^{19}$ /cm²
Sheet resistance	$\gamma^{1/2}$	600 Ω	708 Ω	830 Ω	990 Ω
Contact ρ	$1/\gamma^2$	10 Ω-μm²	5 Ω-μm²	2.5 Ω-μm²	1.25 Ω-μm²
Collector Width	$1/\gamma^2$	600 nm	360 nm	180 nm	90 nm
Thickness	$1/\gamma$	106 nm	75 nm	53 nm	37.5 nm
Current Density	γ^2	9 mA/μm²	18 mA/μm²	36 mA/μm²	72 mA/μm²
$A_{collector}/A_{emitter}$	γ^0	2.4	2.9	2.8	2.8
f_τ	γ^1	520 GHz	730 GHz	1.0 THz	1.4 THz
f_{max}	γ^1	850 GHz	1.30 THz	2.0 THz	2.8 THz
$V_{BR,CEO}$		4.0 V	3.3 V	2.75 V	?
ΔT		50 K	61 K	72K	83 K
I_E/L_E	γ^0	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm	2.3 mA/μm
τ_f	$1/\gamma$	240 fs	180 fs	130 fs	95 fs
C_{cb}/I_c	$1/\gamma$	280 fs/V	240 fs/V	170 fs/V	120 fs/V
$C_{cb}\Delta V_{logic}/I_c$	$1/\gamma$	85 fs	74 fs	52 fs	36 fs
$R_{bb}/(\Delta V_{logic}/I_c)$	γ^0	0.47	0.34	0.26	0.23
$C_{je}(\Delta V_{logic}/I_C)$	$1/\gamma^{3/2}$	180 fs	94 fs	50 fs	33 fs
$R_{ex}/(\Delta V_{logic}/I_c)$	γ^0	0.24	0.24	0.24	0.24
670 GHz gain	--	--	4.3 dB	8.7 dB	12.8 dB
670 GHz Fmin	--	--	7.4 dB	5 dB	3.8 dB
1030 GHz gain	--	--	--	4.9 dB	7.9 dB
1030 GHz Fmin	--	--	--	7.3 dB	5.0 dB

from charge-control analysis:

$$
\begin{aligned}
T_{gate} \approx\ & (\Delta V_L / I_C)(C_{je} + 6C_{cbx} + 6C_{cbi}) + \tau_f \\
& + (kT/qI_C)(0.5C_{je} + C_{cbx} + C_{cbi} + 0.5\tau_f I_C/\Delta V_L) \\
& + R_{ex}(0.5C_{cbx} + 0.5C_{cbi} + 0.5\tau_f I_C/\Delta V_L) \\
& + R_{bb}(0.5C_{je} + C_{cbi} + 0.5\tau_f I_C/\Delta V_L).
\end{aligned}
$$

InP HBT: Status

InP DHBTs: September 2008



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

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

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

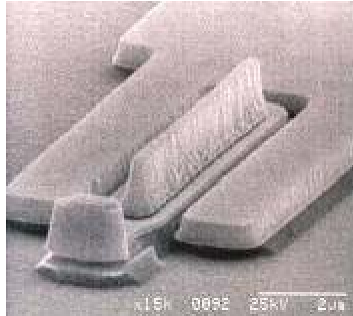
$(R_{bb} I_c / \Delta V)$,

$(\tau_b + \tau_c)$

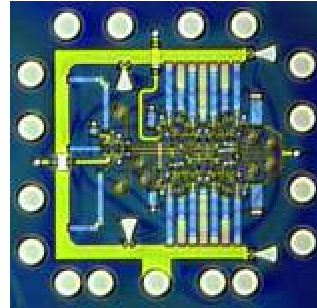
512 nm InP DHBT

Laboratory Technology

500 nm mesa HBT

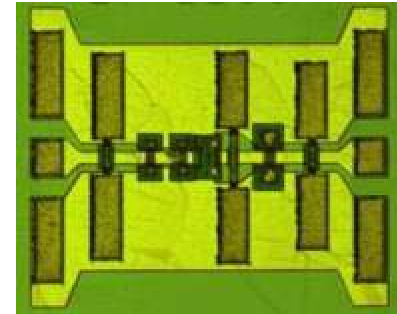


150 GHz M/S latches



UCSB / Teledyne / GCS

175 GHz amplifiers

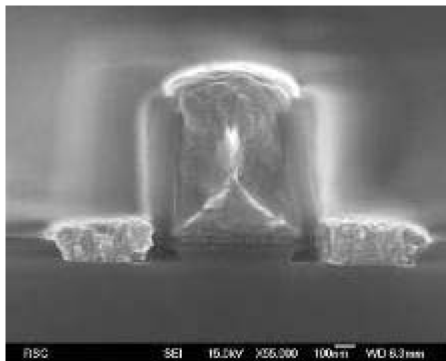


UCSB

Production

(Teledyne)

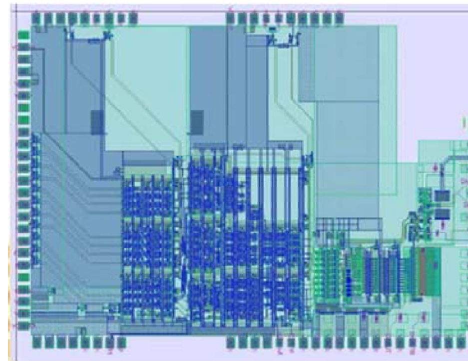
500 nm sidewall HBT



Teledyne

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

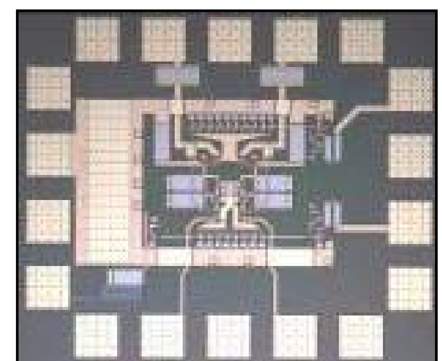
DDS IC: 4500 HBTs



Teledyne / BAE

20 GHz clock

20-40 GHz op-amps

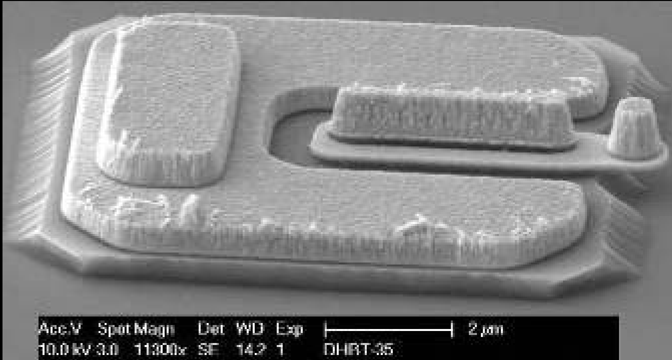


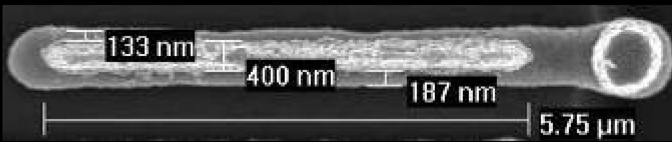
Teledyne / UCSB

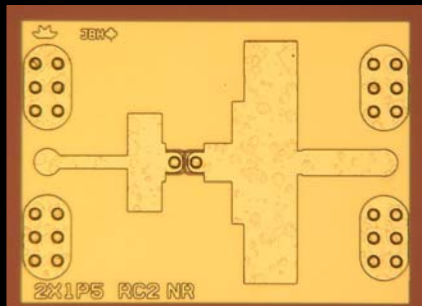
53-56 dBm OIP3 @ 2 GHz with 1 W dissipation

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

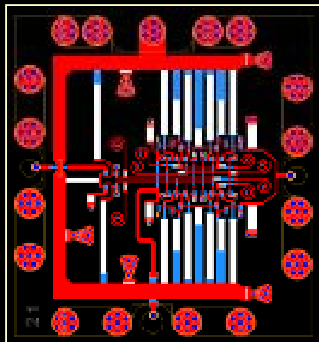
256 nm Generation InP DHBT







324 GHz Amplifier

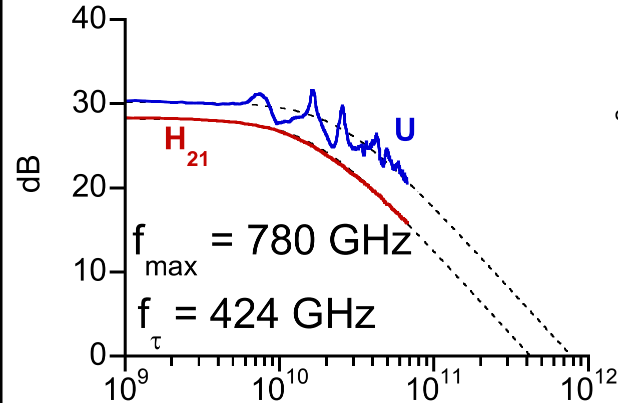


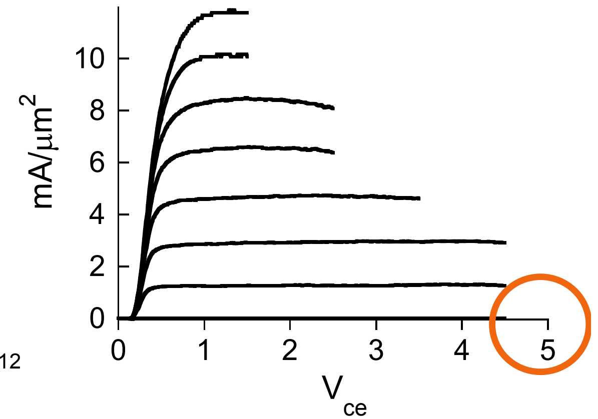
200 GHz master-slave latch design

Z. Griffith, E. Lind

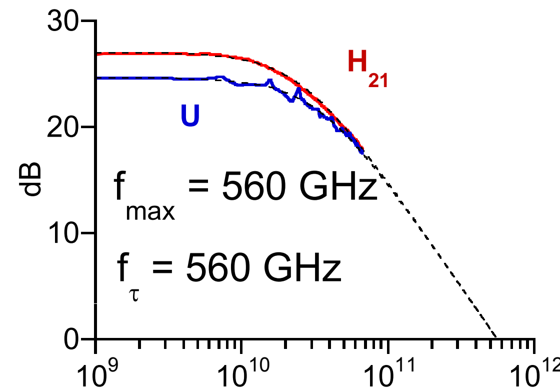
J. Hacker, M. Jones

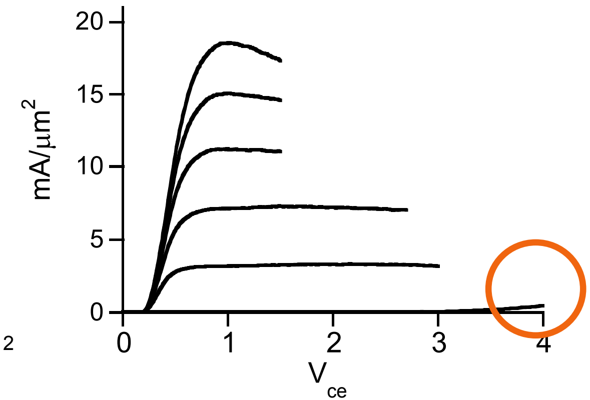
150 nm thick collector



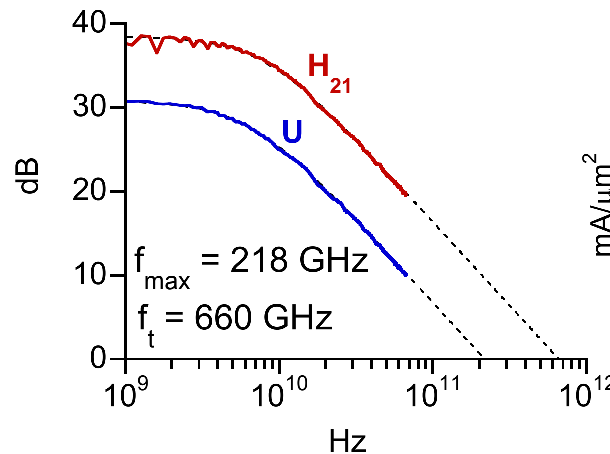


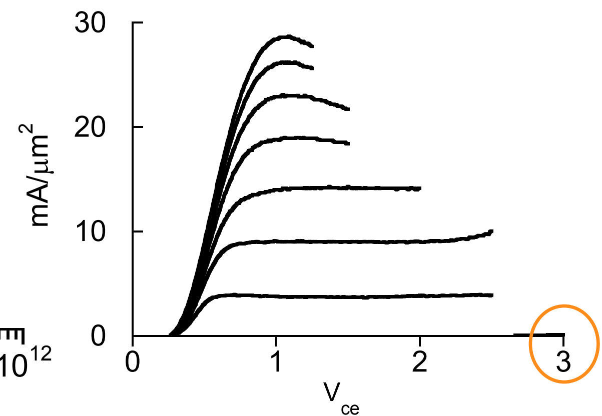
70 nm thick collector





60 nm thick collector





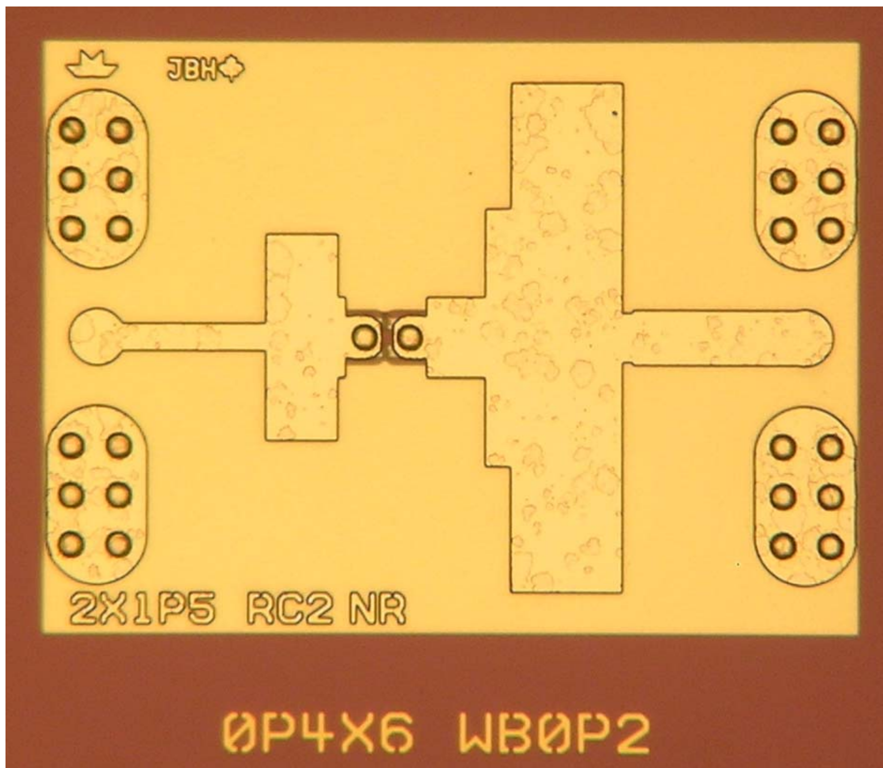
324 GHz Medium Power Amplifiers in 256 nm HBT

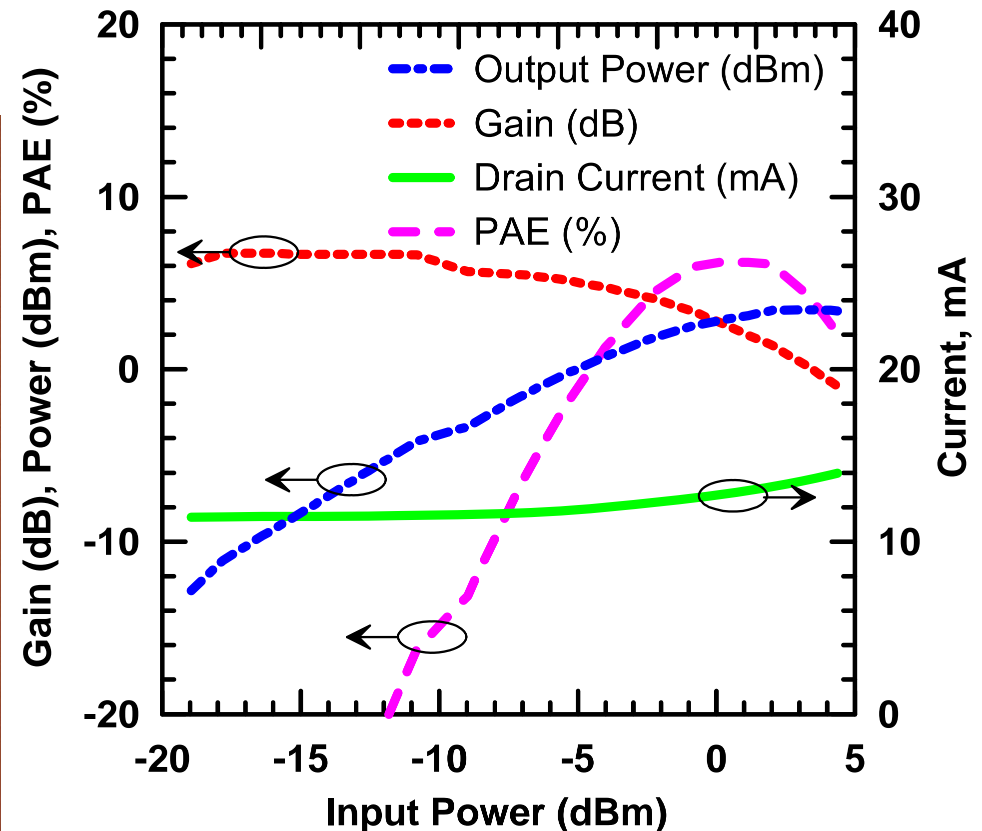
ICs designed by Jon Hacker / Teledyne

Teledyne 256 nm process flow-

Hacker et al, 2008 IEEE MTT-S

~2 mW saturated output power



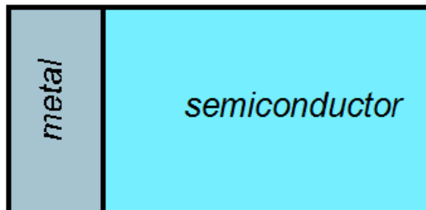


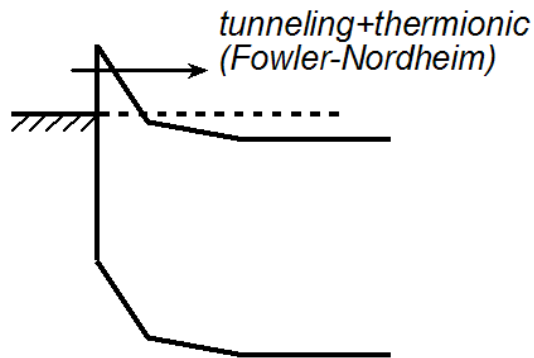
128 / 64 / 32 nm
HBT Technologies

Conventional ex-situ contacts are a mess

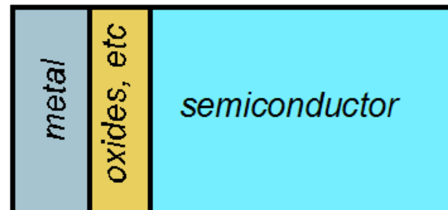
THz transistor bandwidths: very low-resistivity contacts are required

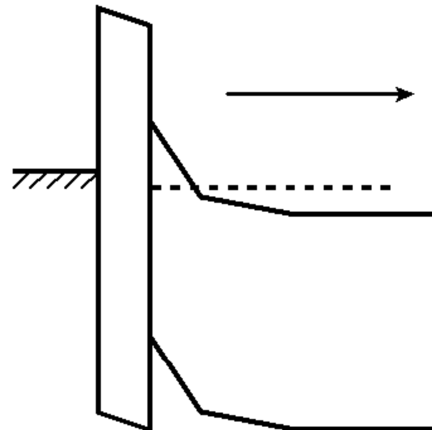
textbook contact



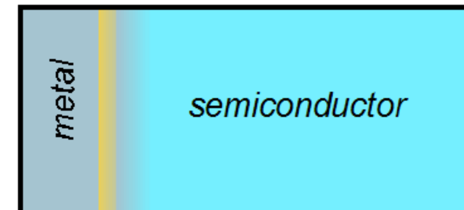


with surface oxide





with metal penetration



Interface barrier → resistance

Further intermixing during high-current operation → degradation

Improvements in Ohmic Contacts

A.. Crook
V. Jain
A. Barakshar
M. Wistey
U. Singisetti
S. Bank

128 nm generation requires ~ 4 Ω - μm^2 emitter & base resistivities

64 nm generation requires ~ 2 Ω - μm^2

Contacts to N-InGaAs*:

Mo	*MBE in-situ*	*2.2 (+/- 0.5) Ω - μm^2*
TiW	*ex-situ / NH4 pre-clean*	*~2.2 Ω - μm^2*
	variable between process runs	

Contacts to P-InGaAs:

Mo	*MBE in-situ*	*below 2.5 Ω - μm^2*
Pd/Ti...	*ex-situ*	*~4 Ω - μm^2*

...far better contacts coming...

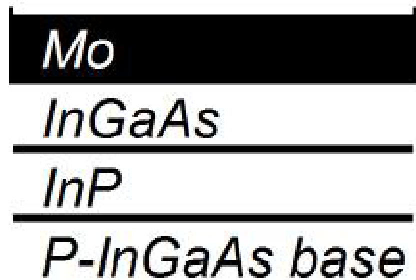
**measured emitter resistance remains higher than that of contacts.*

Mo Emitter Contacts: Robust Integration into Process Flow

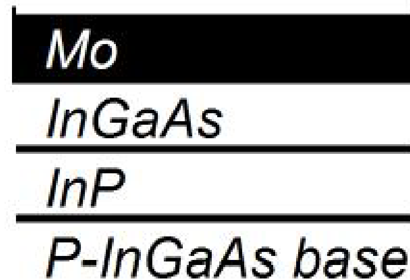
M. Wistey
A. Barakshar
U. Singisettti
V. Jain

Proposed Process Integration:

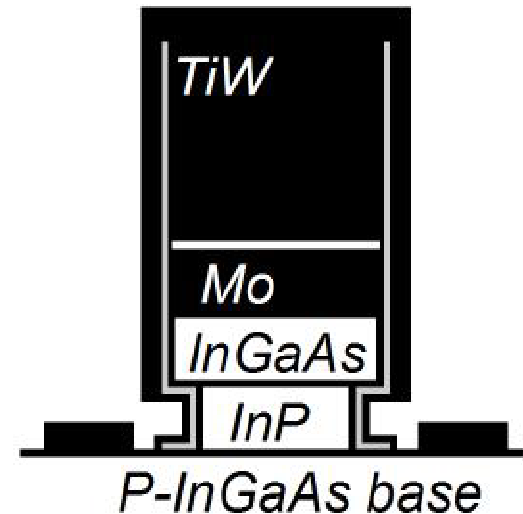
MBE growth and Mo deposition



store wafer as long as desired



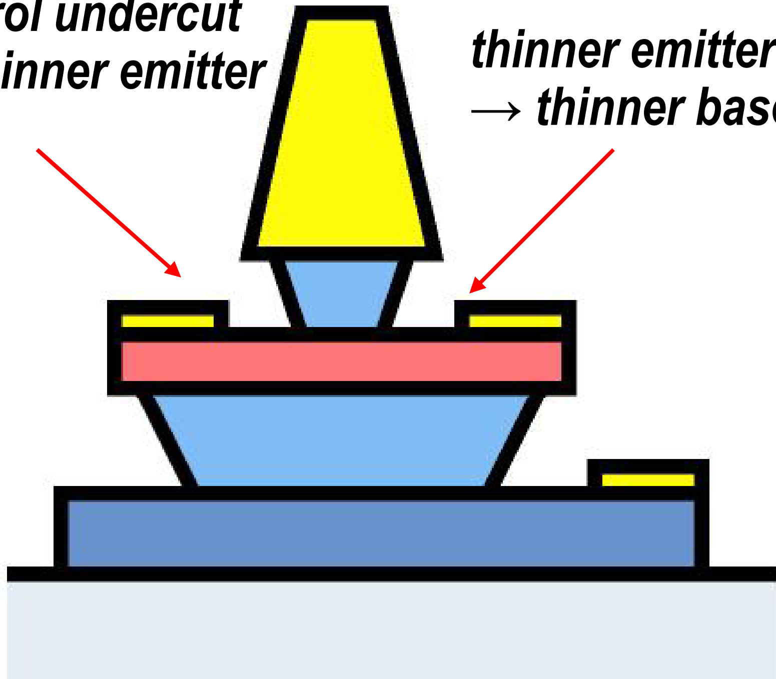
build HBTs



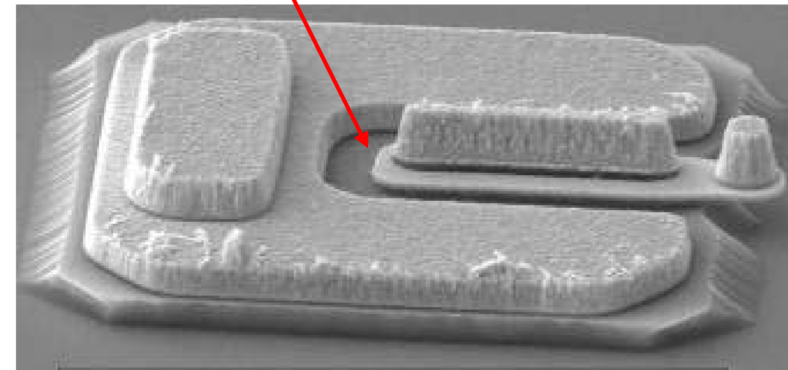
Process Must Change Greatly for 128 / 64 / 32 nm Nodes

control undercut
→ thinner emitter

thinner emitter
→ thinner base metal

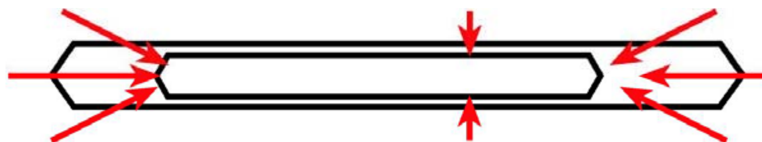


thinner base metal
→ excess base metal resistance

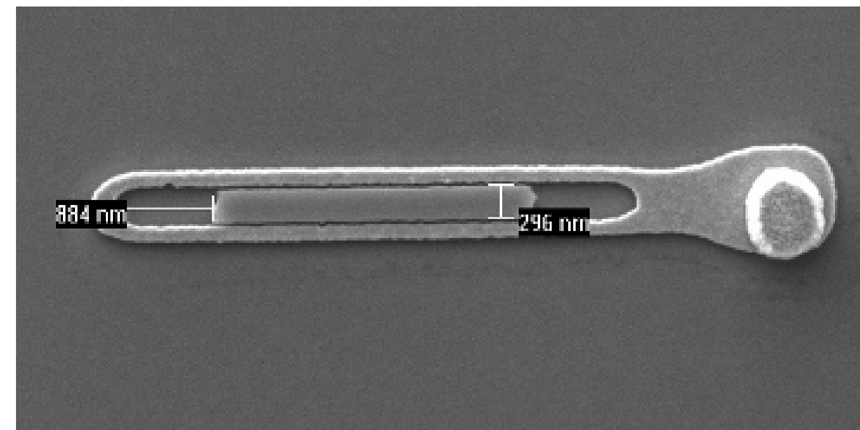


Undercutting of emitter ends

{101}A planes: fast

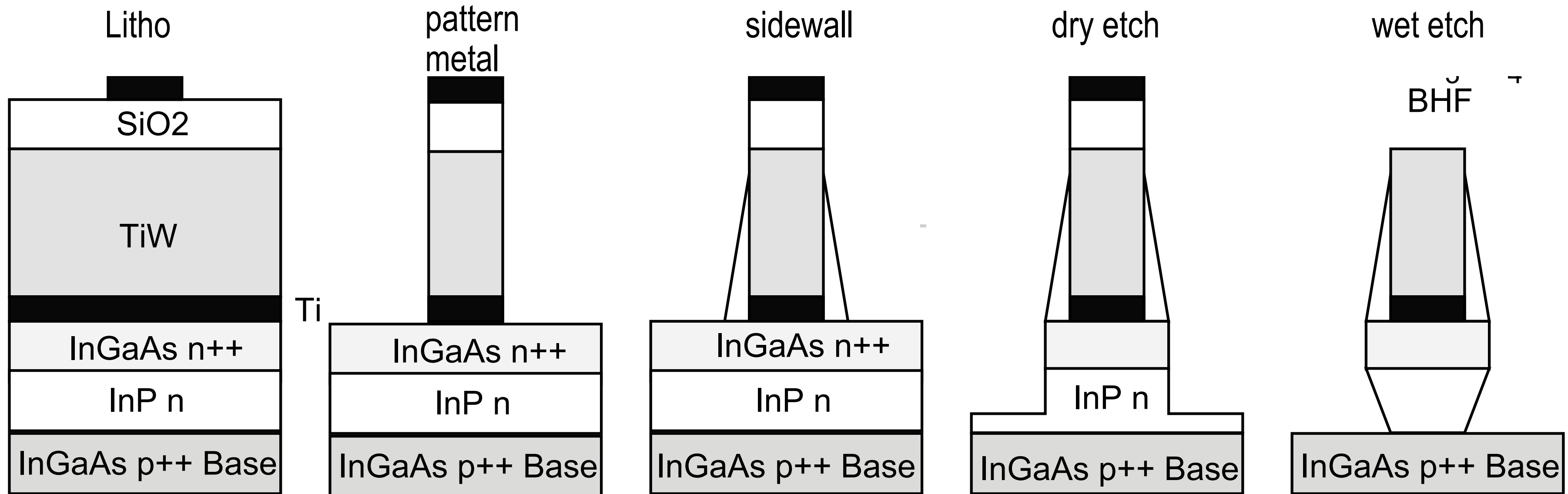


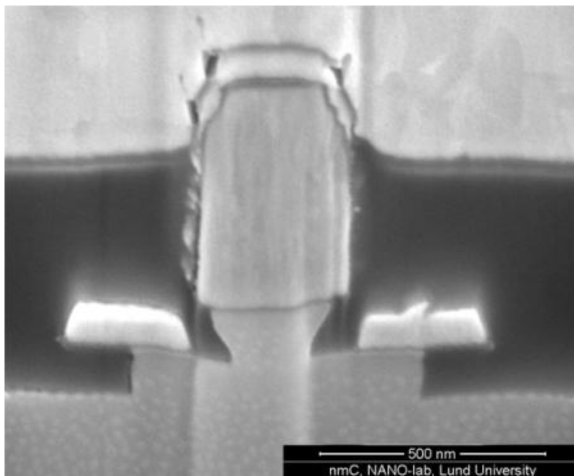
{111}A planes: slow

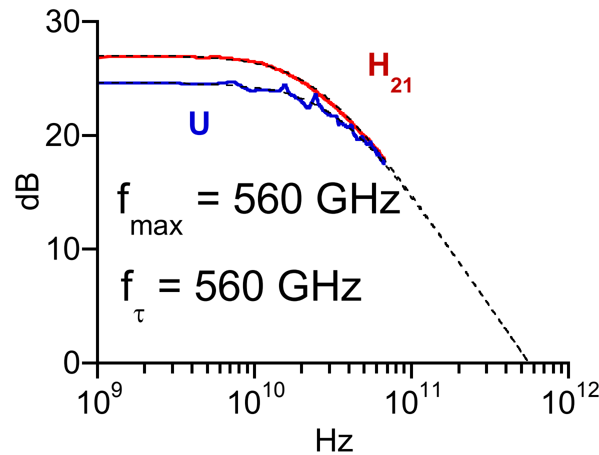


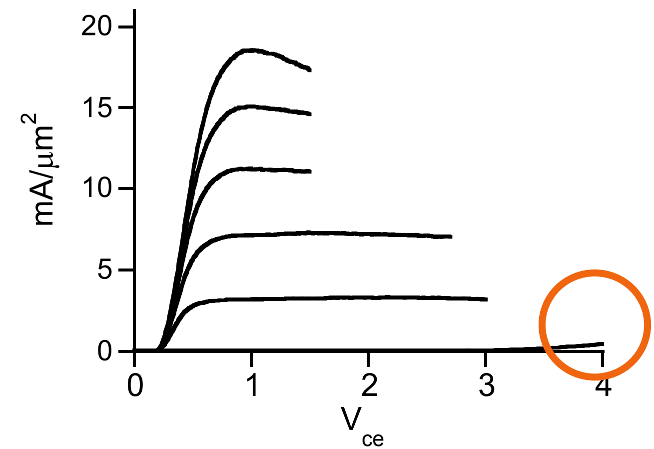
E. Lind

128 nm Emitter Process: Dry Etched Metal & Semiconductor





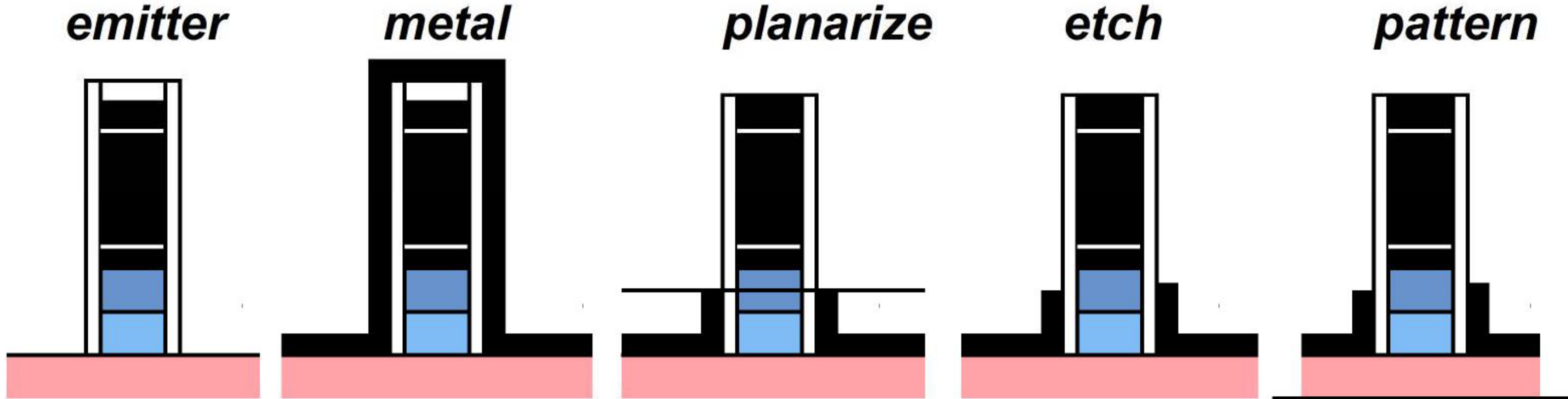


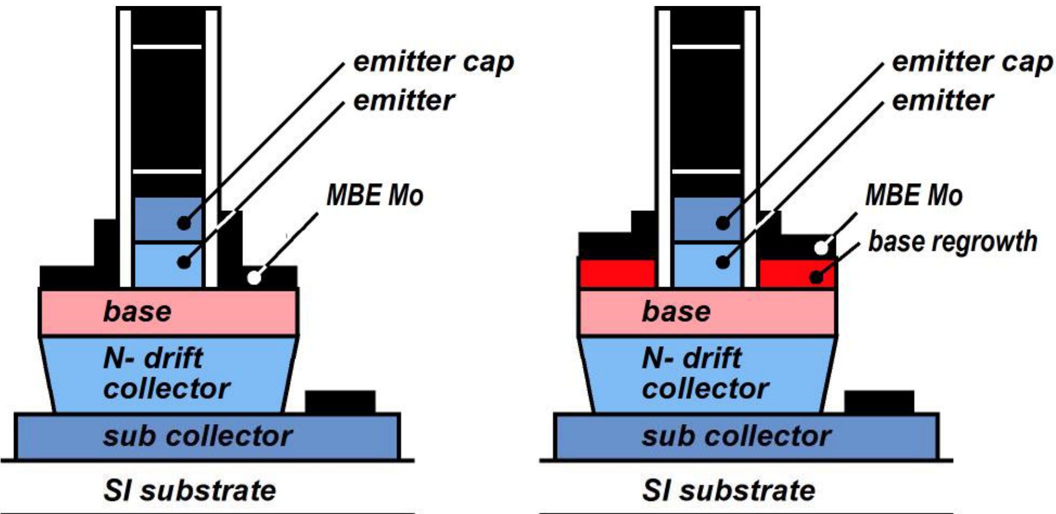


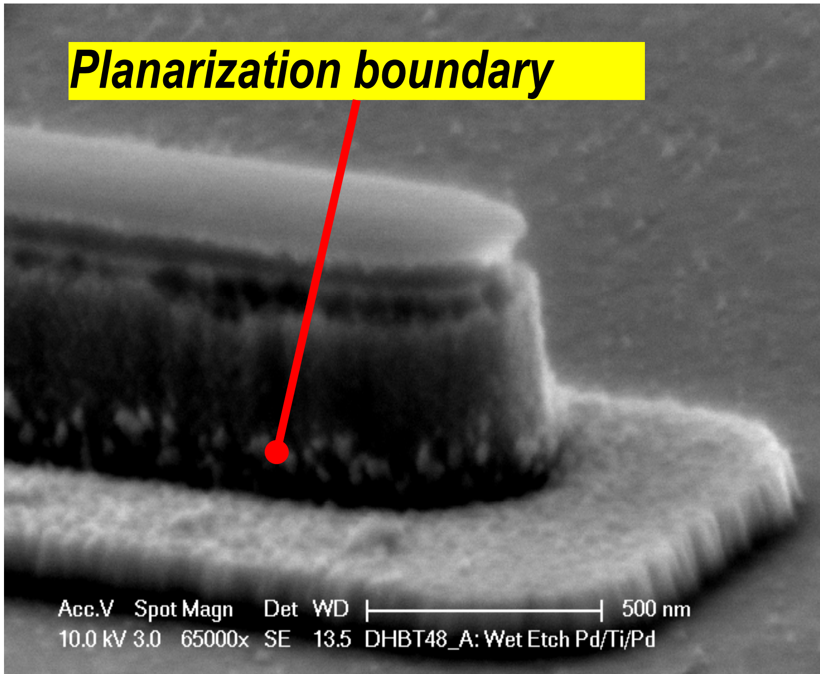
results @ c.a. 200 nm emitter metal width

Planarization E/B Processes for 64 & 32 nm

E; Lobisser
V. Jain
G. Burek



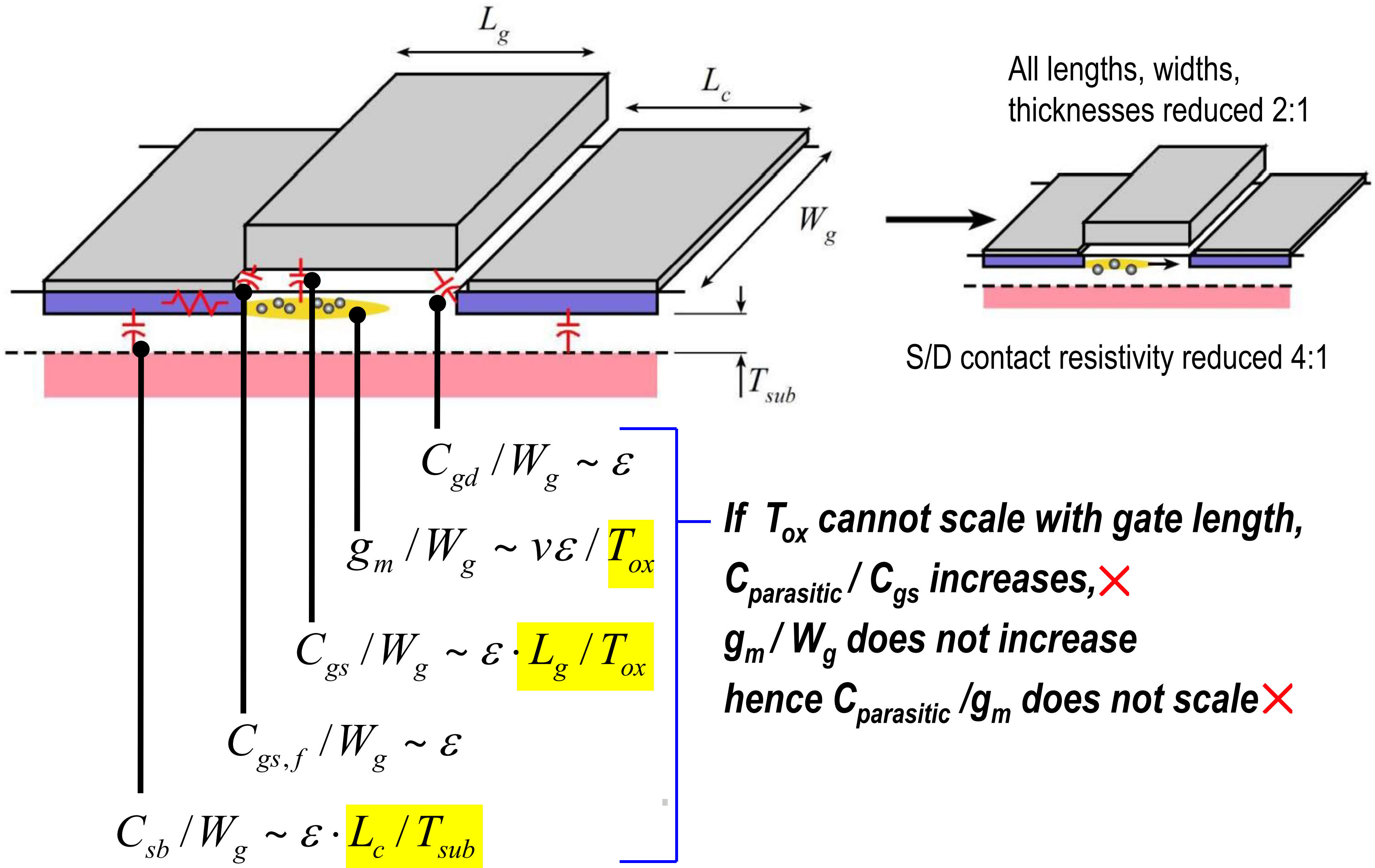




III-V FET Scaling

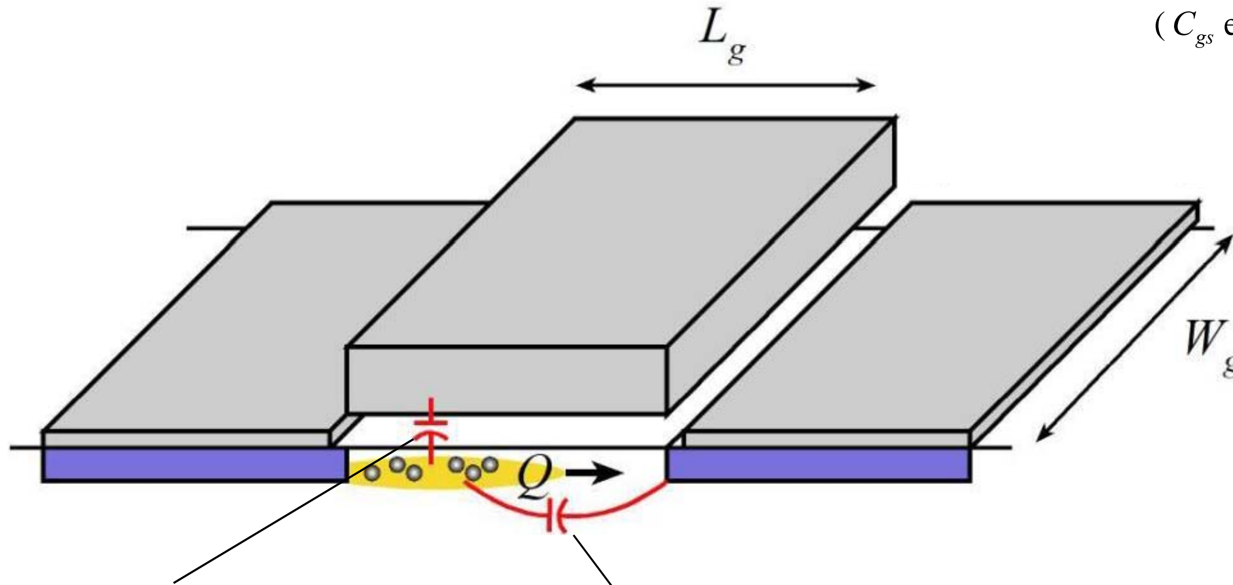
Simple FET Scaling

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



FET scaling: Output Conductance & DIBL

(C_{gs} expression neglects D.O.S. effects)



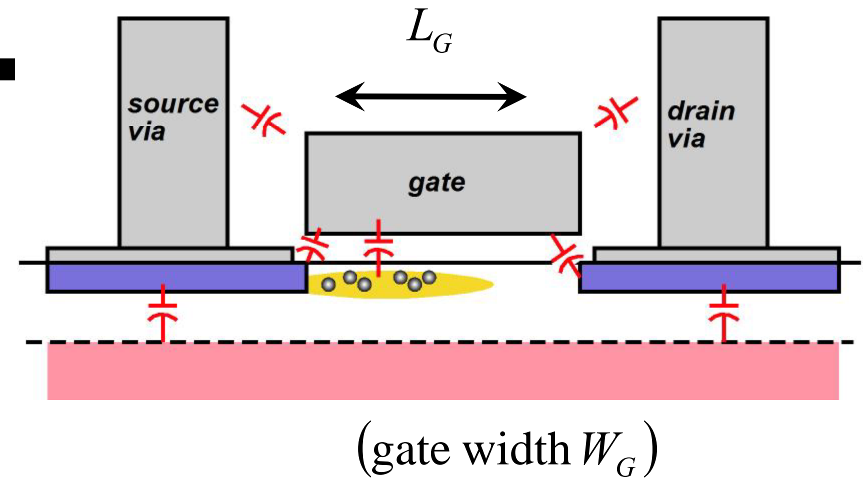
$$C_{gs} \sim \varepsilon W_g L_g / T_{ox} \qquad C_{d-ch} \sim \varepsilon W_g$$

$$I_d = Q/\tau \quad \text{where} \quad \delta Q = C_{gs}\delta V_{gs} + C_{d-ch}\delta V_{ds}$$

transconductance ($C_{gs}\delta V_{gs}$) — *output conductance* ($C_{d-ch}\delta V_{ds}$)

→ ***Keep L_g / T_{ox} constant as we scale L_g***

FET Scaling Laws

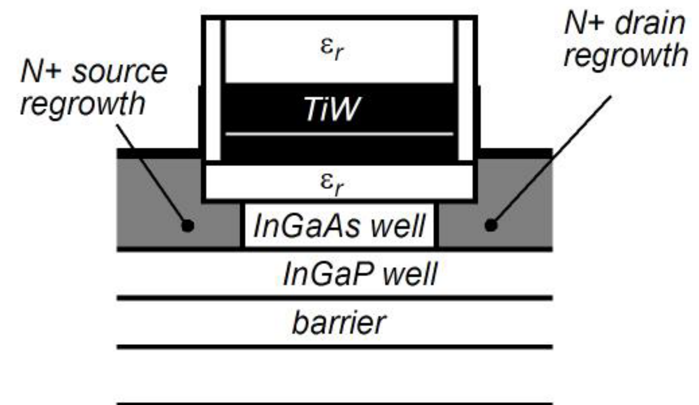


$(\text{gate width } W_G)$

Changes required to double transistor bandwidth:

parameter	change
gate length	decrease 2:1
gate dielectric capacitance density	increase 2:1
gate dielectric equivalent thickness	decrease 2:1
channel electron density	increase 2:1
source & drain contact resistance	decrease 4:1
current density (mA/μm)	increase 2:1

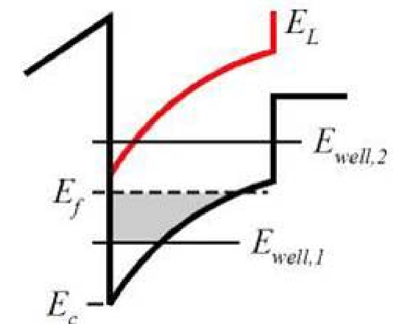
III-V MOSFETs for VLSI



What is it ?
MOSFET with an InGaAs channel

Why do it ?
low electron effective mass→ higher electron velocity
more current, less charge at a given insulator thickness & gate length
very low access resistance

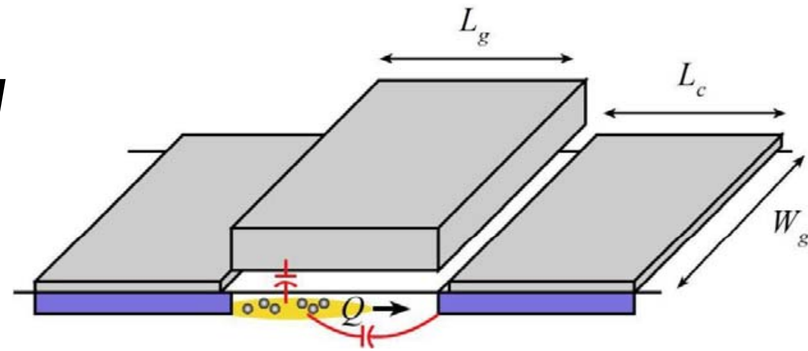
What are the problems ?
low electron effective mass→ constraints on scaling !
must grow high-K on InGaAs, must grow InGaAs on Si



Our focus today is III-V FET scaling generally

Low Effective Mass Impairs Vertical Scaling

Shallow electron distribution needed for high g_m / G_{ds} ratio.



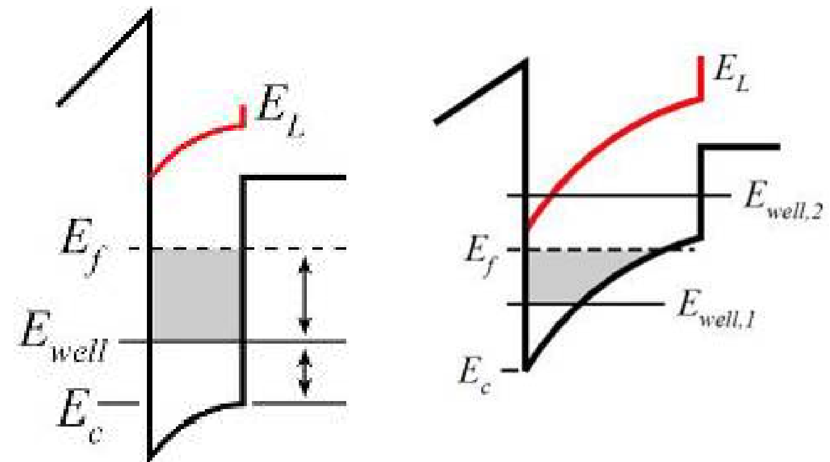
Energy of Lth well state $\propto L^2 / m^* T_{well}^2$.

For thin wells,

only 1st state can be populated.

For very thin wells,

1st state approaches L-valley.



Only one vertical state in well.

Minimum ~ 5 nm well thickness.

→ constrains gate length scaling.

Density-Of-States Capacitance

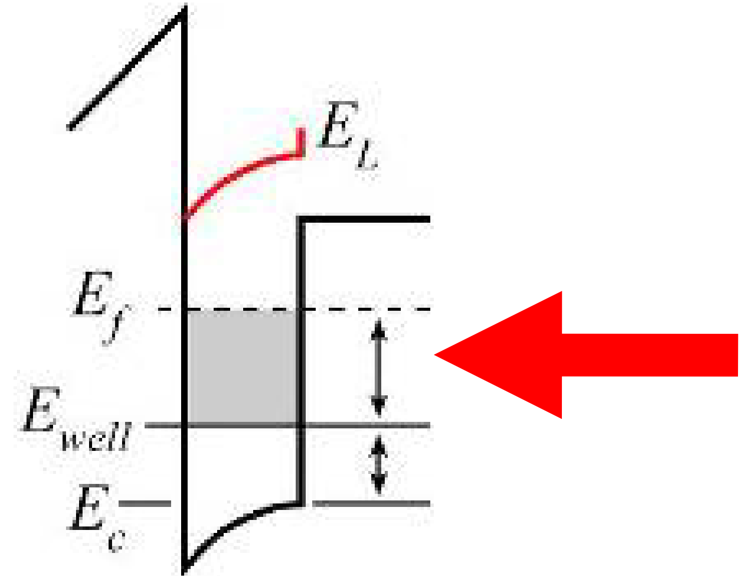
$$E_f - E_{well} = n_s /(nm^* / \pi\hbar^2)$$

(bidirectional motion)

$$V_f - V_{well} = \rho_s / c_{dos}$$

where $c_{dos} = q^2 nm^* / \pi\hbar^2$

and n is the # of band minima



Two implications:

- With N_s >10^{13}/cm^2, electrons populate satellite valleys

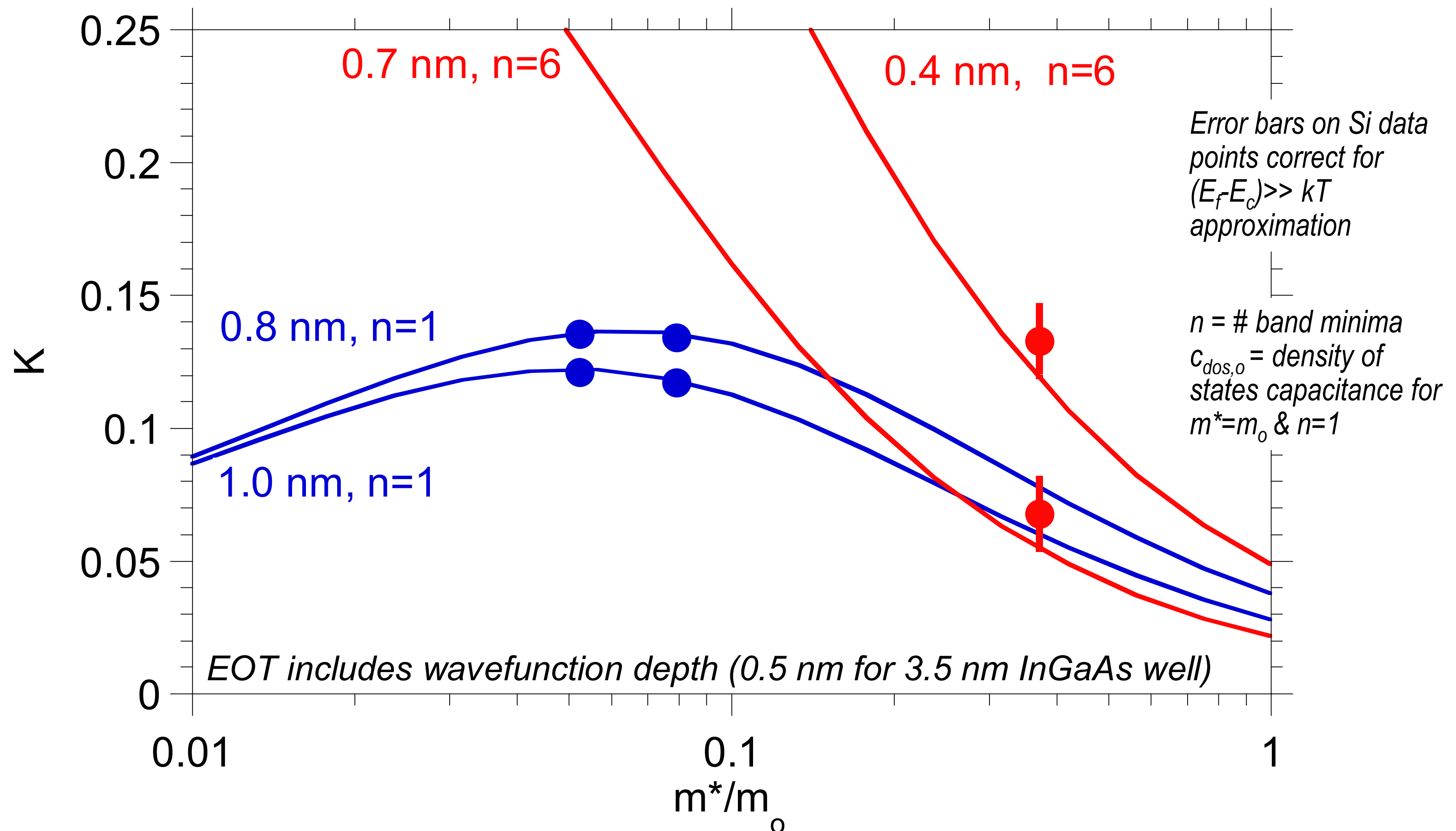
Fischetti et al, IEDM2007

- Transconductance limited by finite state density

Solomon & Laux , IEDM2001

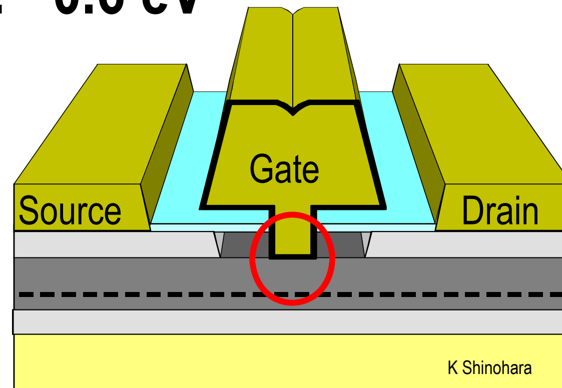
Drive Current in the Ballistic & Degenerate Limits

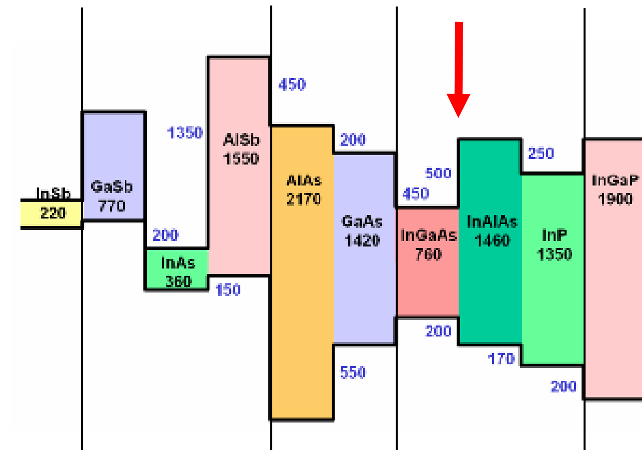
$$J = \underline{K} \cdot \left(84\frac{\text{mA}}{\mu\text{m}}\right) \cdot \left(\frac{V_{gs} - V_{th}}{1\,\text{V}}\right)^{3/2}, \quad \text{where } \underline{K} = \frac{n \cdot \left(m^*/m_o\right)^{1/2}}{\left(1 + (c_{dos,o}/c_{ox}) \cdot n \cdot (m^*/m_o)\right)^{3/2}}$$



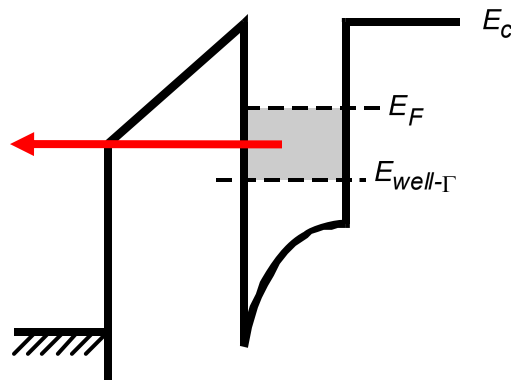
HEMT Scaling Challenge: Low Gate Barrier

Gate barrier is low: ~0.6 eV

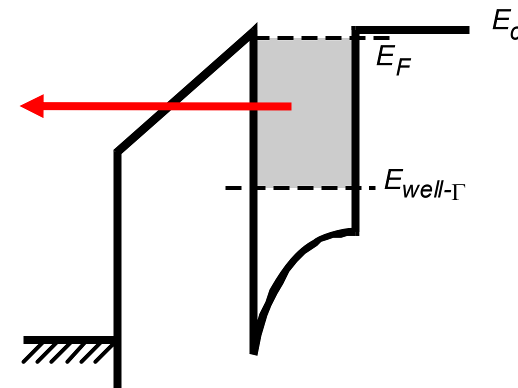




Tunneling through barrier
→ sets minimum thickness



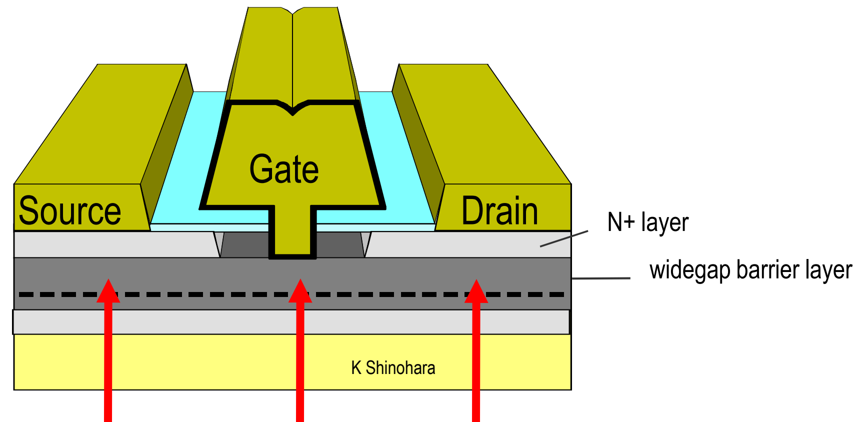
Emission over barrier
→ limits 2D carrier density



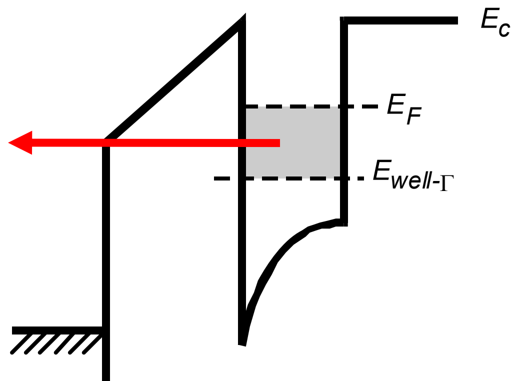
At $N_s = 10^{13}/\text{cm}^2, (E_f - E_c) \sim 0.6\,\text{eV}$

HEMT Scaling Challenge: High Access Resistance

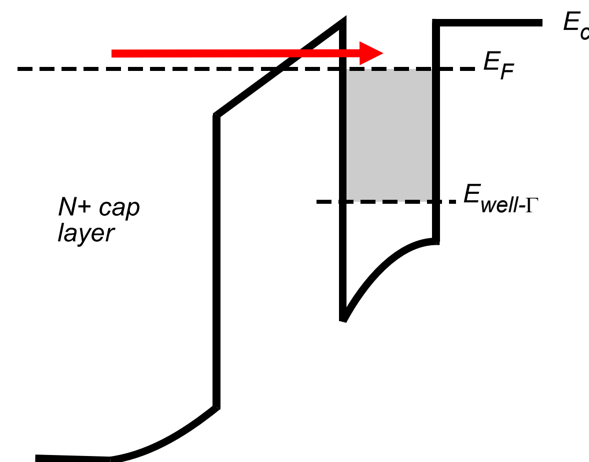
Gate barrier also lies under source / drain contacts



low leakage:
need high barrier under gate



low resistance:
need low barrier under contacts



THz III-V FET Scaling: What Must Be Done

As gate length is reduced...

channel thickness should be reduced...

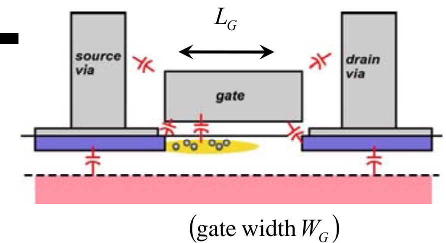
barrier thickness should be reduced...

target g_m/W_g and I_d/W_g should be increased...

source and drain access resistivity should be reduced...

We face serious difficulties in doing these.

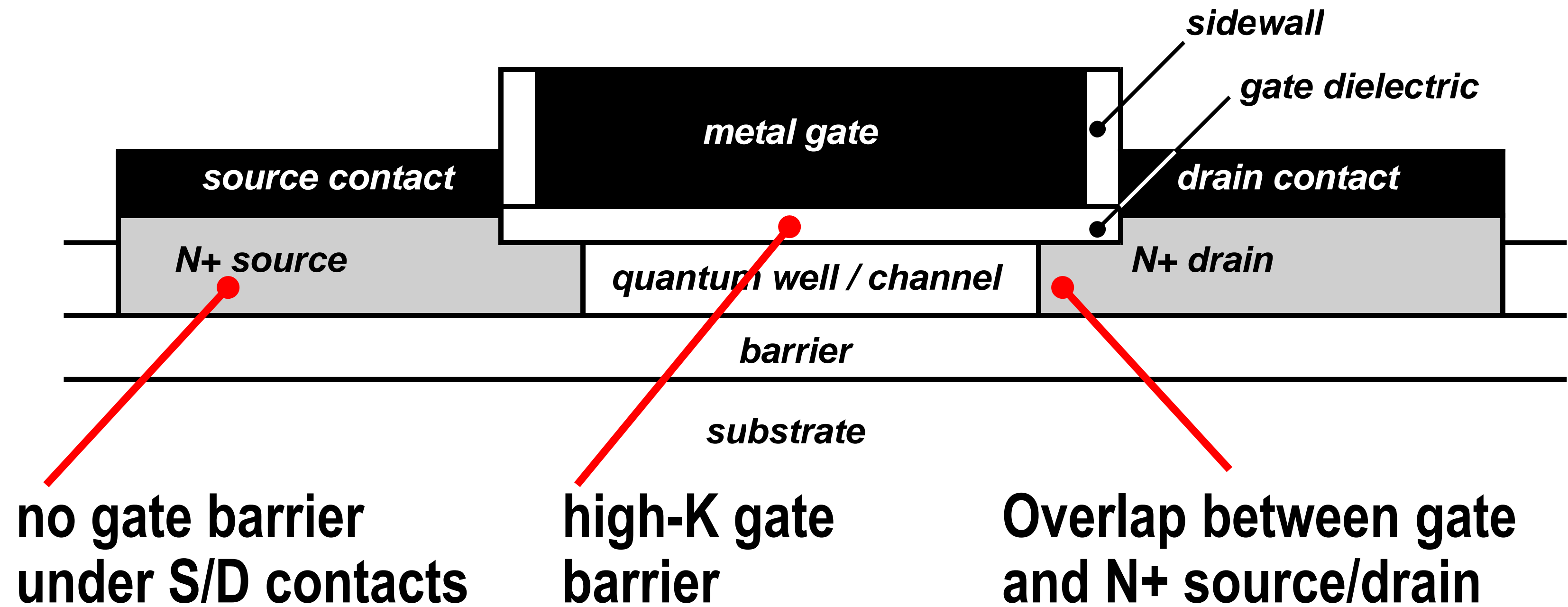
FET Scaling Laws



Changes required to double transistor bandwidth:

parameter	change
gate length	decrease 2:1
gate dielectric capacitance density	increase 2:1
gate dielectric equivalent thickness	decrease 2:1
channel electron density	increase 2:1
source & drain contact resistance	decrease 4:1
current density (mA/µm)	increase 2:1

A MOSFET Might Scale Better than a HEMT



no gate barrier under S/D contacts

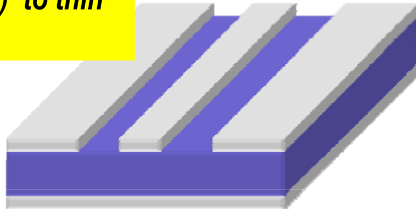
high-K gate barrier

Overlap between gate and N+ source/drain

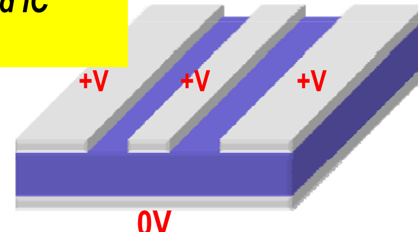
Interconnects

Coplanar Waveguide

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

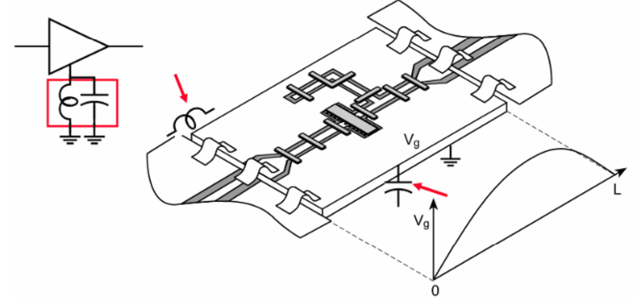


Hard to ground IC to package

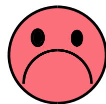


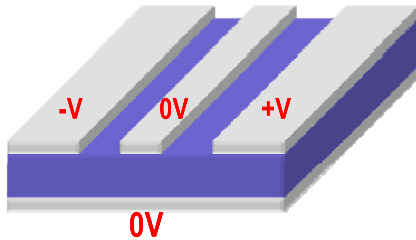
0V

Parasitic microstrip mode



ground plane breaks → loss of ground integrity

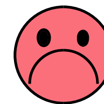




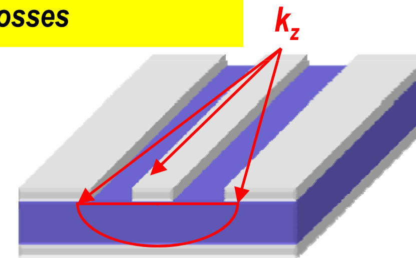
0V

Parasitic slot mode

substrate mode coupling or substrate losses



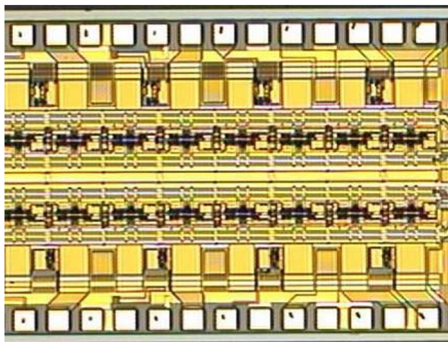
k_z



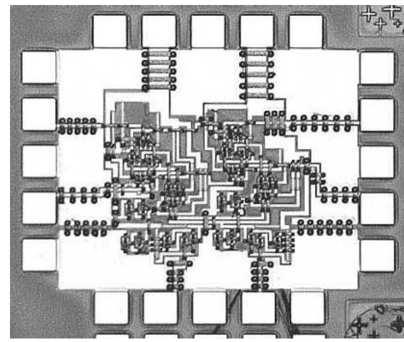
III-V:
semi-insulating substrate→ substrate mode coupling

Silicon
conducting substrate → substrate conductivity losses

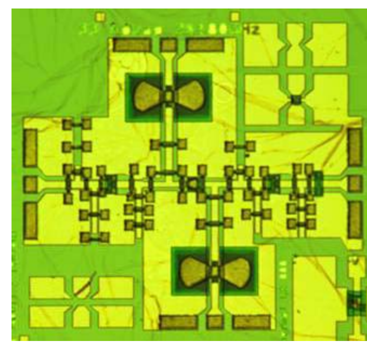
Repairing ground plane with ground straps is effective only in simple ICs
In more complex CPW ICs, ground plane rapidly vanishes
→ common-lead inductance → strong circuit-circuit coupling



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

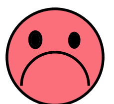


35 GHz master-slave latch in CPW
note fragmented ground plane

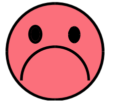


175 GHz tuned amplifier in CPW
note fragmented ground plane

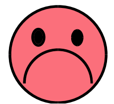
poor ground integrity



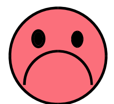
loss of impedance control



ground bounce



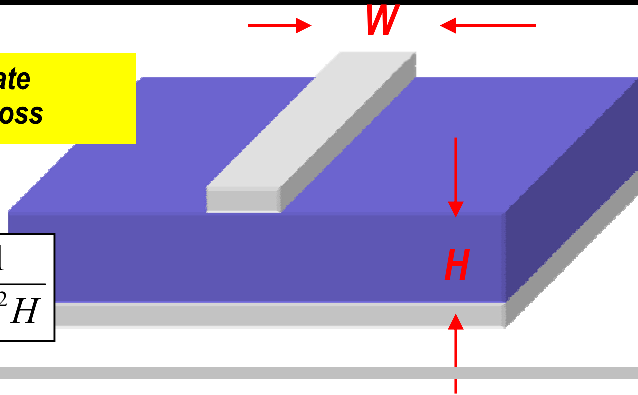
coupling, EMI, oscillation



Classic Substrate Microstrip

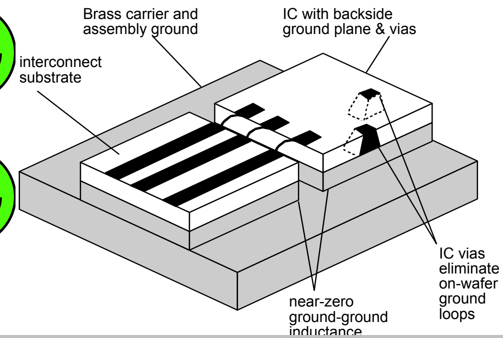
Thick Substrate → low skin loss

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

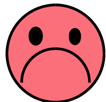


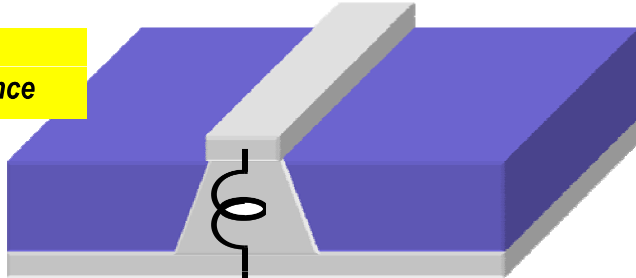
Zero ground inductance in package

No ground plane breaks in IC



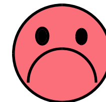
High via inductance

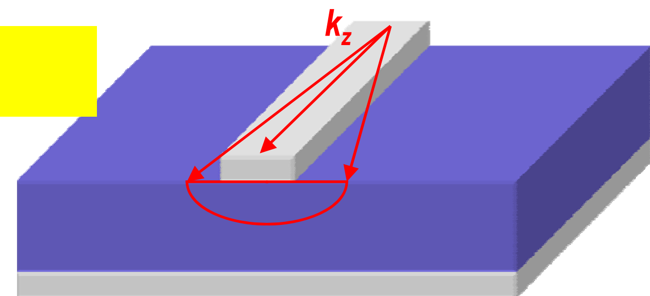




12 pH for 100 μm substrate -- 7.5 Ω@ 100 GHz

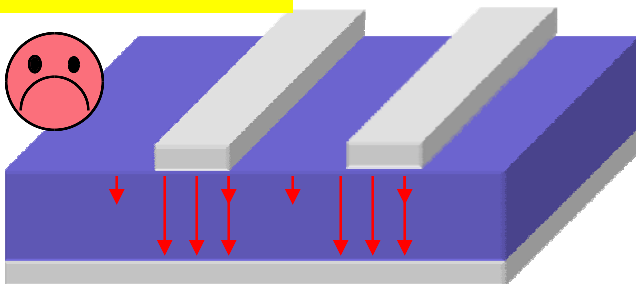
TM substrate mode coupling





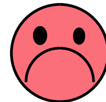
Strong coupling when substrate approaches ~λ_d/ 4 thickness

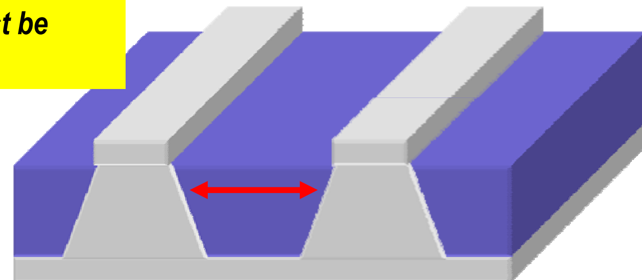
lines must be widely spaced



Line spacings must be ~3*(substrate thickness)

ground vias must be widely spaced

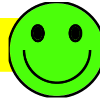




all factors require very thin substrates for >100 GHz ICs
→ lapping to ~50 μm substrate thickness typical for 100+ GHz

III-V MIMIC Interconnects -- Thin-Film Microstrip

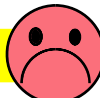
narrow line spacing → IC density



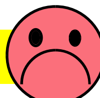
no substrate radiation, no substrate losses

fewer breaks in ground plane than CPW

... but ground breaks at device placements

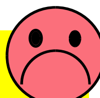


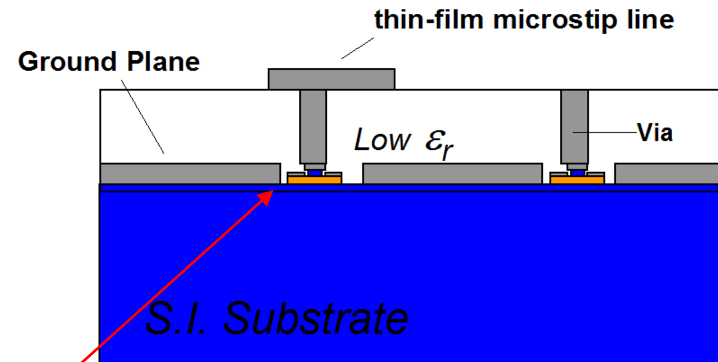
still have problem with package grounding

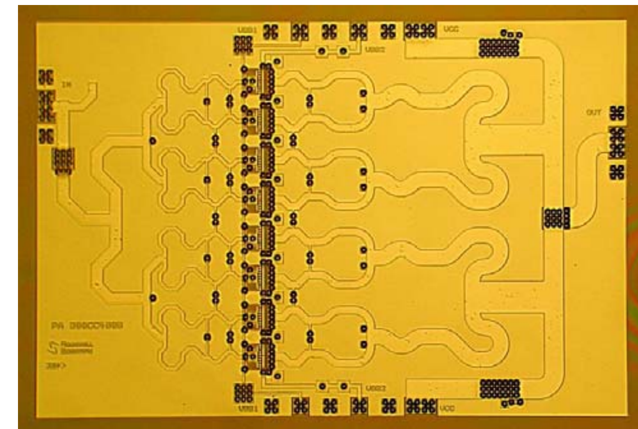


...need to flip-chip bond

thin dielectrics → narrow lines
→ high line losses
→ low current capability
→ no high-Z_o lines

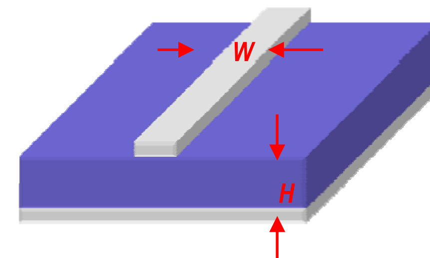






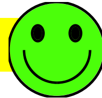
InP mm-wave PA (Rockwell)

$$Z_o \sim \frac{\eta_o}{\varepsilon_r^{1/2}}\left(\frac{H}{W+H}\right)$$

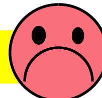


III-V MIMIC Interconnects -- Inverted Thin-Film Microstrip

narrow line spacing → IC density



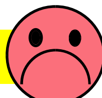
Some substrate radiation / substrate losses



No breaks in ground plane

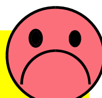
... no ground breaks at device placements

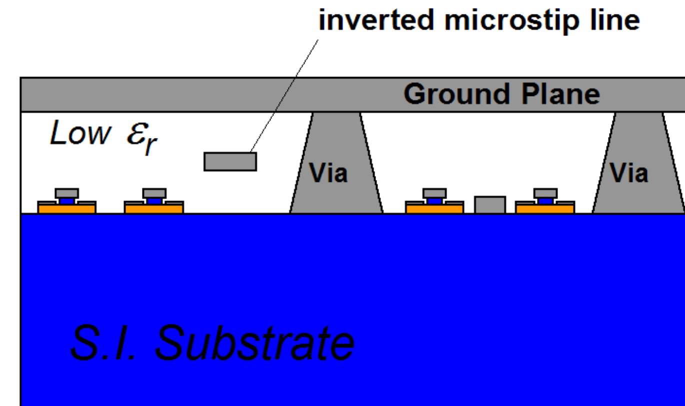
still have problem with package grounding

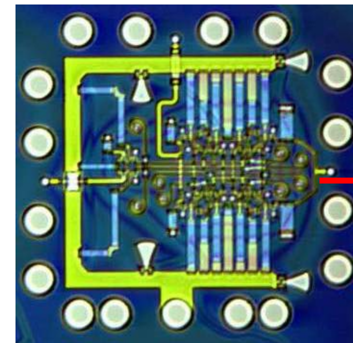


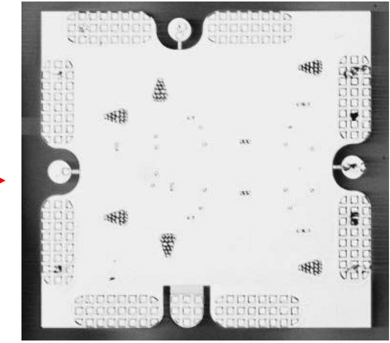
...need to flip-chip bond

thin dielectrics → narrow lines
→ high line losses
→ low current capability
→ no high-Z_o lines

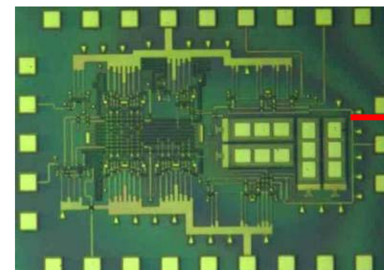


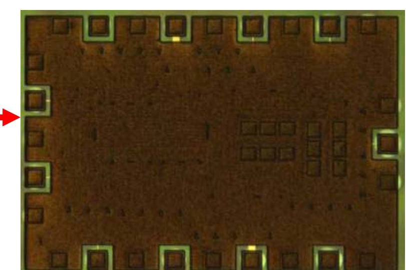






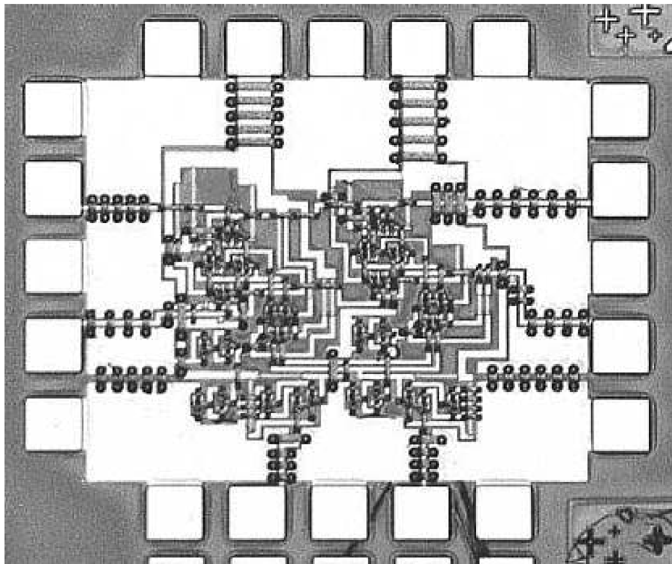
InP 150 GHz master-slave latch



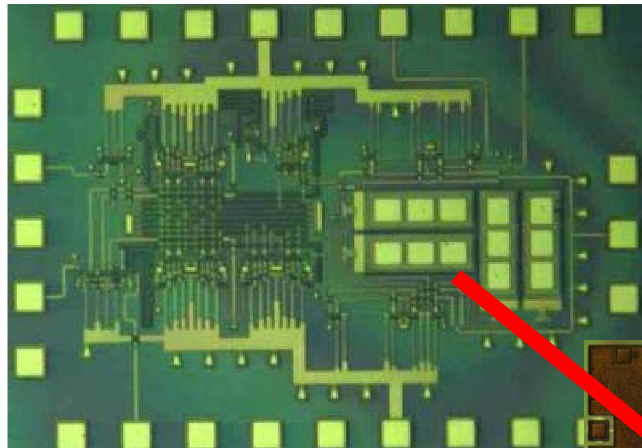


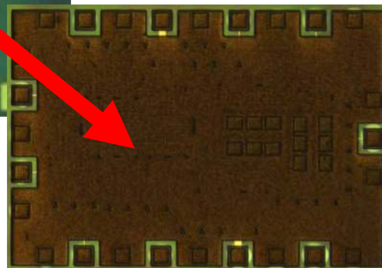
InP 8 GHz clock rate delta-sigma ADC

No clean ground return ? → interconnects can't be modeled !



35 GHz static divider
interconnects have no clear local ground return
interconnect inductance is non-local
interconnect inductance has no compact model





8 GHz clock-rate delta-sigma ADC
thin-film microstrip wiring
every interconnect can be modeled as microstrip
some interconnects are terminated in their Zo
some interconnects are not terminated
...but ALL are precisely modeled

InP 8 GHz clock rate delta-sigma ADC

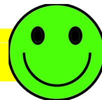
VLSI Interconnects with Ground Integrity & Controlled Z_0

narrow line spacing → IC density

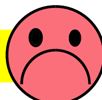
no substrate radiation, no substrate losses

negligible breaks in ground plane

negligible ground breaks @ device placements

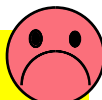


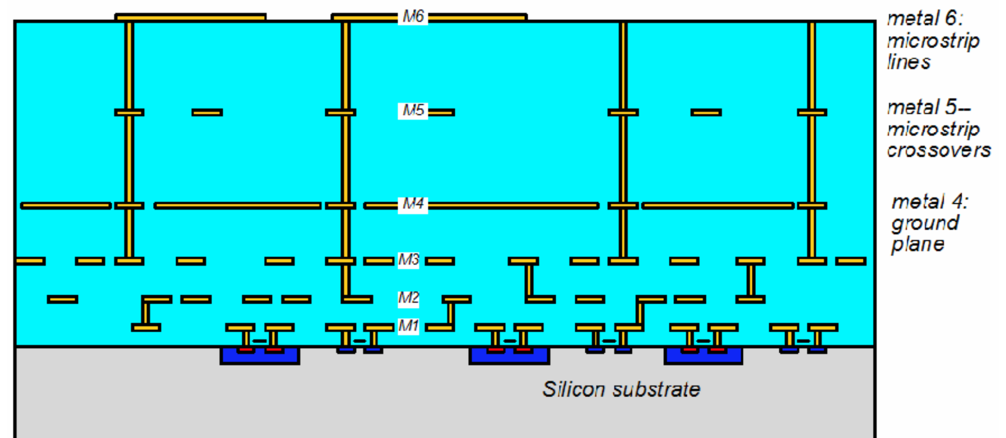
still have problem with package grounding



...need to flip-chip bond

thin dielectrics → narrow lines
→ high line losses
→ low current capability
→ no high-Z_0 lines





Conclusions

Few-THz Transistors

<u>Few-THz InP Bipolar Transistors: can it be done ?</u>

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 ?

<u>Few-THz InP Field-Effect Transistors: can it be done?</u>

challenges are gate barrier, vertical scaling,
source/drain access resistance, increased gm and drive current.

2DEG carrier concentrations must increase.

S/D regrowth offers a path to lower access resistance.

Solutions needed for gate barrier: maybe even MOSFET ?

What Would We Do With Them ?