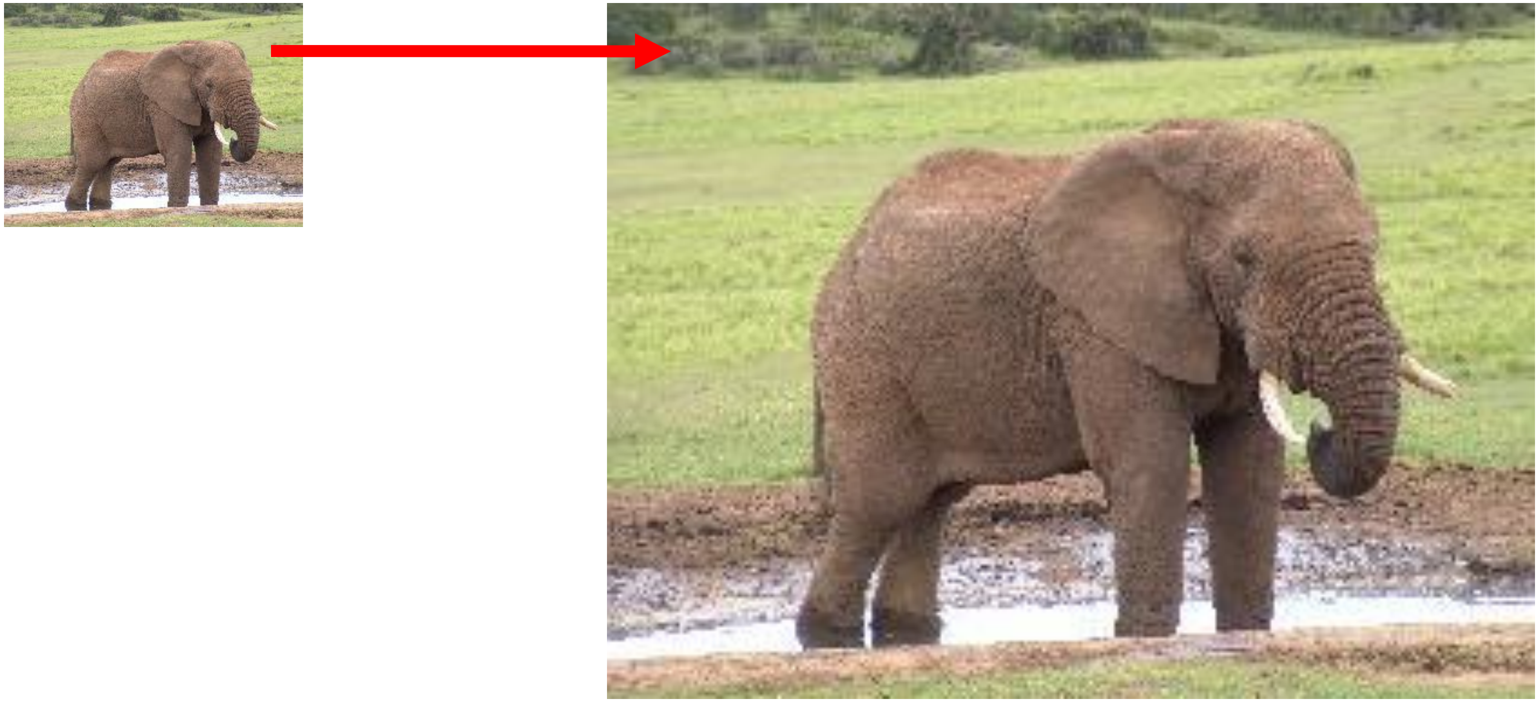
THz Transistors:
It's All About The Interfaces.

Mark Rodwell
University of California, Santa Barbara

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

Forget Ballistic Transport:

Consider Nanotechnology... & Elephants



10:1 (taller /wider/ deeper)

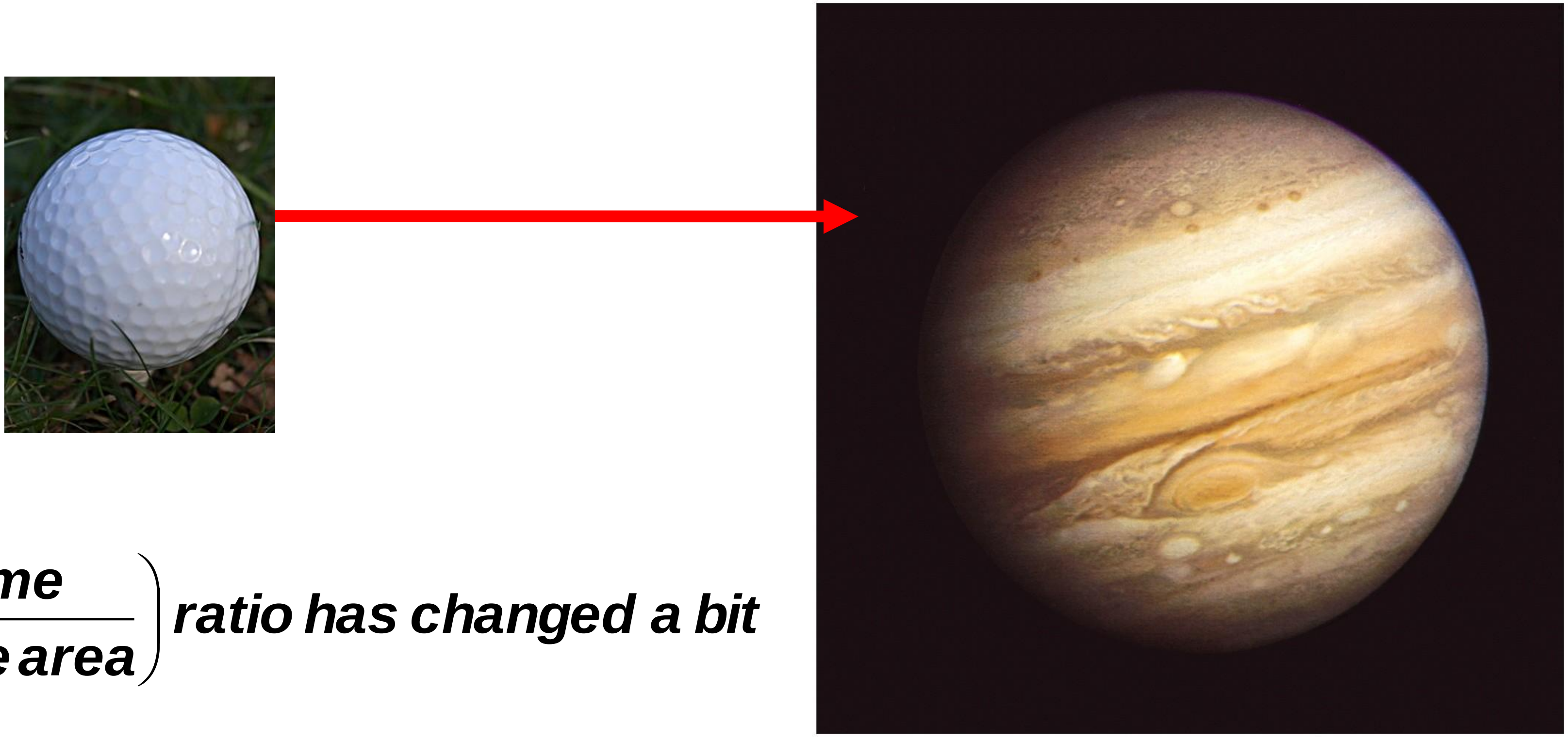
1000: 1 more metabolism, 100:1 larger skin area surface→ overheats

1000: 1 larger weight, 100:1 larger bone cross-section→ legs break

1000: 1 more flesh, 100:1 larger lung surface→ suffocates

(plagiarized from Galileo)

Scaling... a golf ball



$\left(\frac{\text{volume}}{\text{surface area}}\right)$ ratio has changed a bit

Scaling: little things change more quickly than big things

Scaling:

the surface matters most in little things,
the bulk matters most in big things

Everything We Need to Know, We Learned in ~~Kindergarten~~

Sophomore Circuits

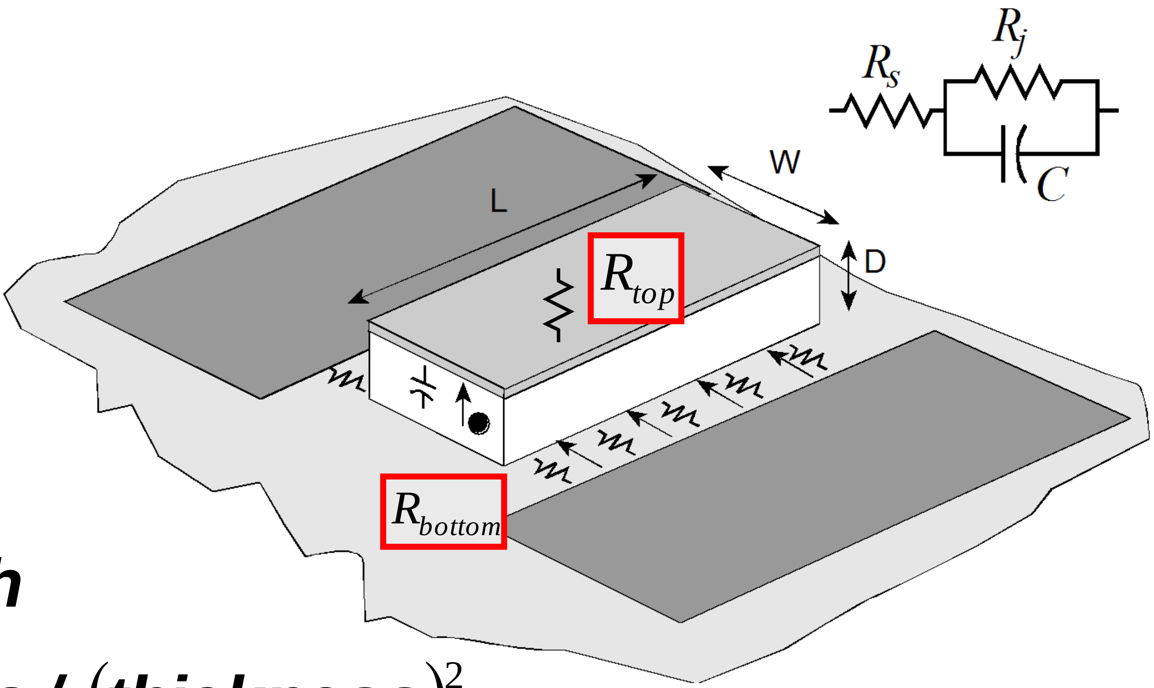
$\tau \propto \textbf{\textit{thickness}}$

$C \propto \textbf{\textit{area}}/\textbf{\textit{thickness}}$

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

$R_{bottom} \propto 1/\textbf{\textit{stripe length}}$

$I_{\max,\,\textbf{\textit{space-charge-limit}}} \propto \textbf{\textit{area}}/\left(\textbf{\textit{thickness}}\right)^2$



photodiode

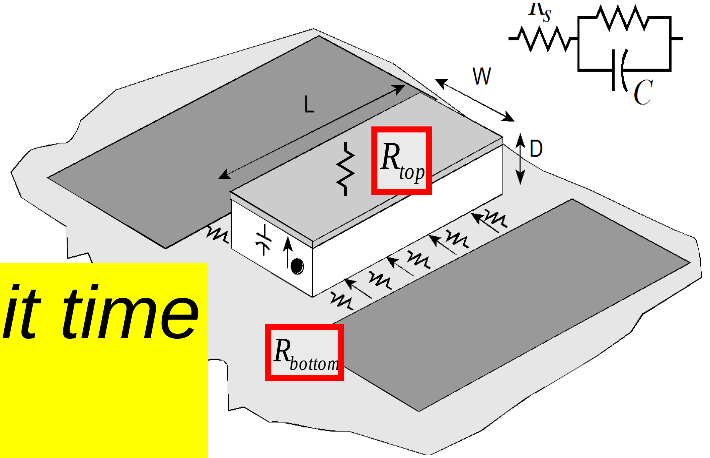
To double bandwidth,

reduce thicknesses 2:1

reduce width 4:1, keep constant length

current density has increased 4:1

It's all very simple, really...



resistance capacitance transit time

device bandwidth

applies to almost all (lumped) semiconductor devices

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

→

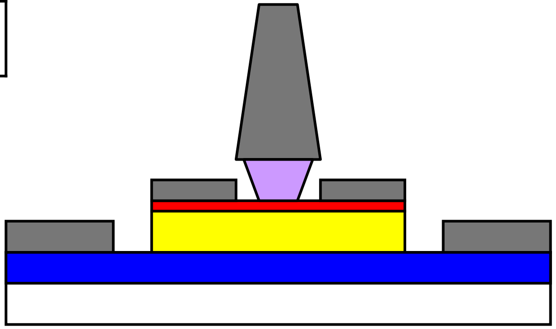
THz
semiconductor
devices

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

Scaling Bipolar Transistors

to double the bandwidth:

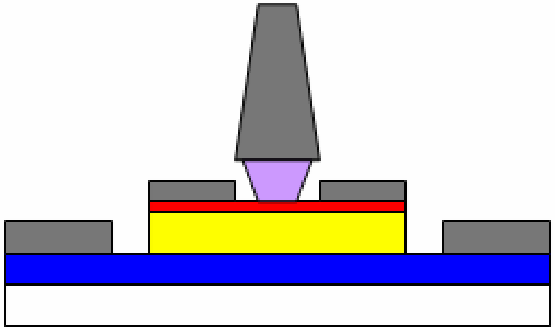
parameter	change
collector depletion layer thickness	decrease 2:1
base thickness	decrease 1.414:1
emitter junction width	decrease 4:1
collector junction width	decrease 4:1
emitter contact resistance	decrease 4:1
current density	increase 4:1
base contact resistivity	decrease 4:1



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

InP HBT Scaling... Is All About Contacts & Current Density

	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,
	16	8	4	2	**1** $\Omega\cdot\mu m^2$ contact ρ
collector	150	106	75	53	**38** 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



Device Designers Don't Matter... Material Scientists Do

To build a 5-THz bipolar Transistor...

...we need $0.25\ \Omega\text{-}\mu m^2$ Ohmic contacts,

& these must be stable at $300\ mA/\mu m^2$.

For this, we need help

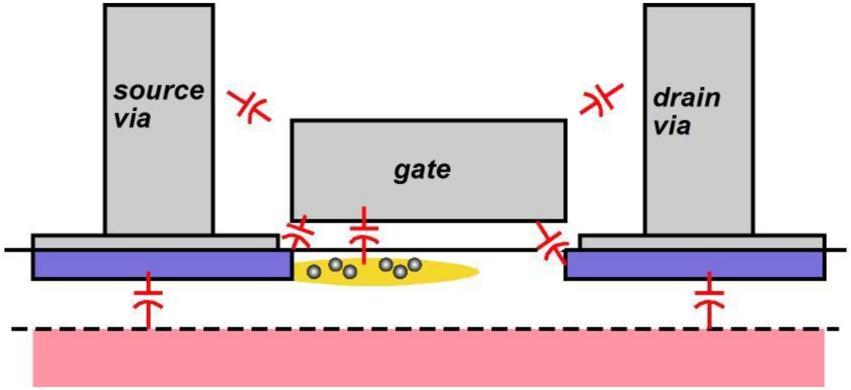
**** finite density-of-states may cause problems at the 2 THz generation***

FETs

FETs: Think About Fringing Capacitances

Reducing gate length improves $\tau = C_{gs} / g_m$

...but transconductance must increase to improve $C_{parasitic} / g_m$



source resistance limits g_m----- so contacts & access must improve

gate dielectric thickness limits g_m --- so must make gate barrier thinner

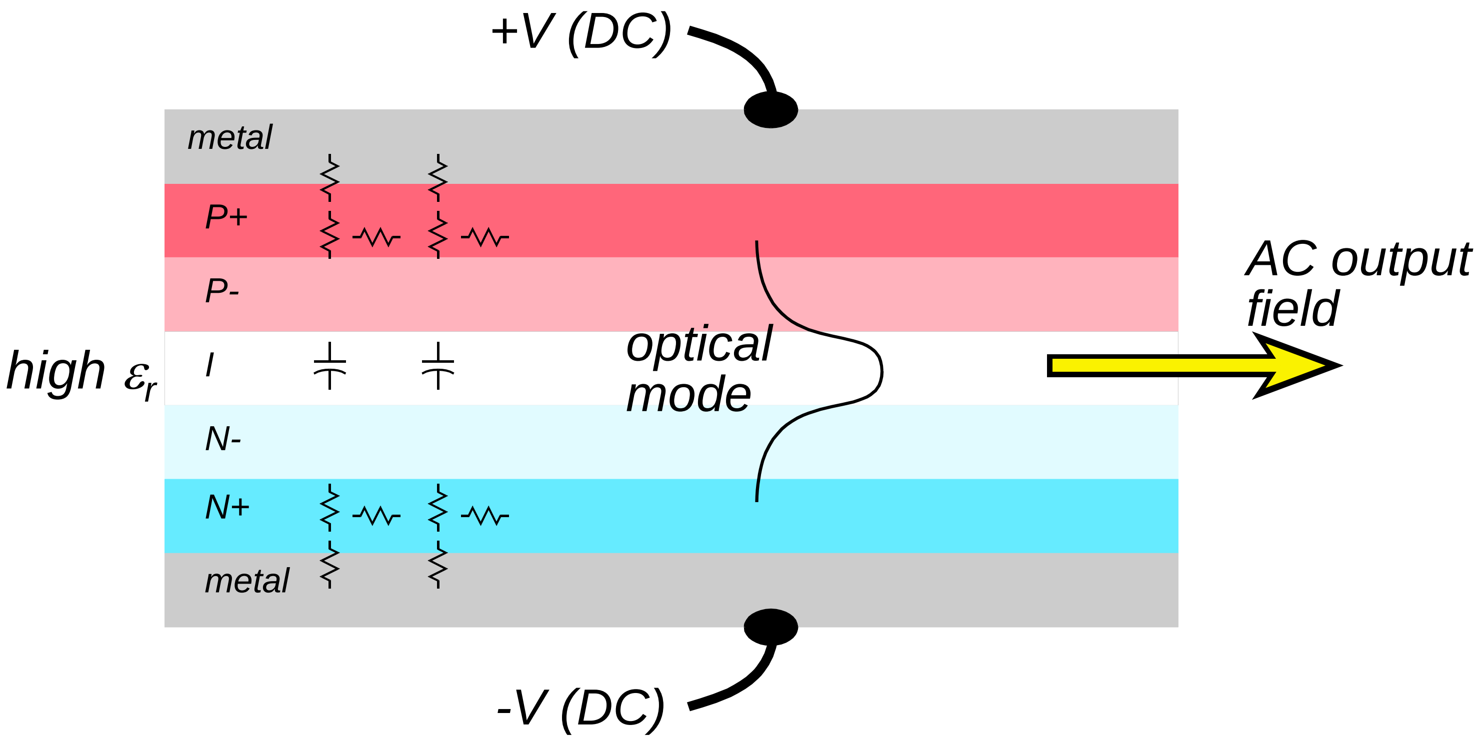
density of states limits gm ---- so must make gate barrier thinner

valid FET scaling analyses must consider well quantization

& finite density of states→ density of states capacitance

For Later

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

For Later: BJTs

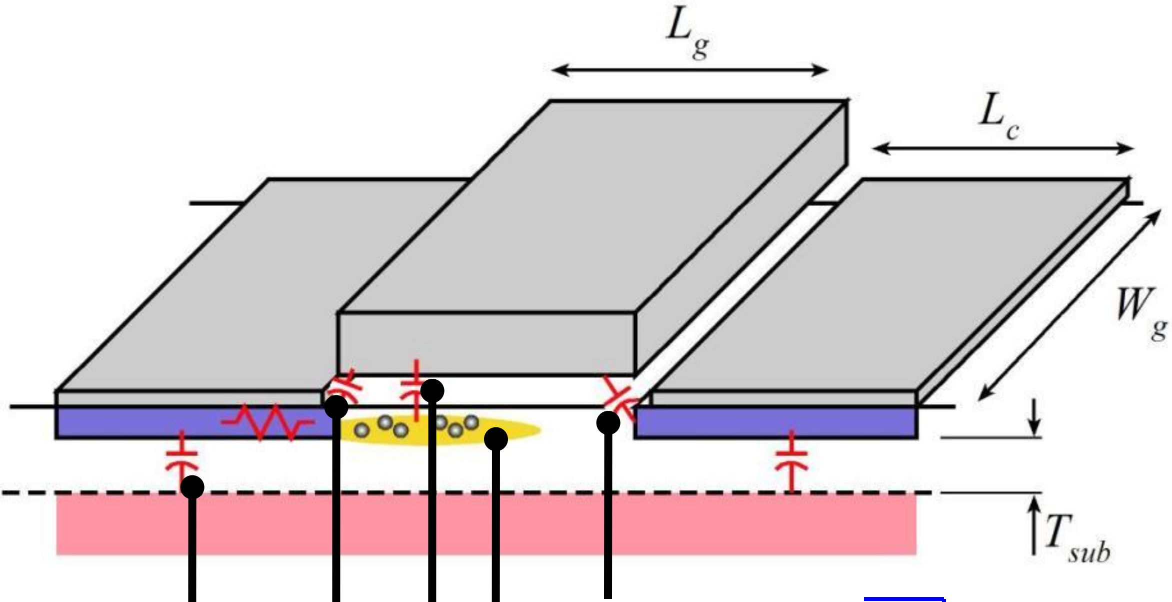
HBT Scaling

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-μm^2	4 Ω-μm^2	2 Ω-μm^2	1 Ω-μm^2
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} /cm^2	7 10^{19} /cm^2	7 10^{19} /cm^2	7 10^{19} /cm^2
Sheet resistance	$\gamma^{1/2}$	600 Ω	708 Ω	830 Ω	990 Ω
Contact ρ	$1/\gamma^2$	10 Ω-μm^2	5 Ω-μm^2	2.5 Ω-μm^2	1.25 Ω-μm^2
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/μm^2	18 mA/μm^2	36 mA/μm^2	72 mA/μm^2
$A_{\text{collector}}/A_{\text{emitter}}$	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	?
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
$C_{cb}\Delta V_{\text{logic}} / I_c$	$1/\gamma$	85 fs	74 fs	52 fs	36 fs
$R_{bb} / (\Delta V_{\text{logic}} / I_c)$	γ^0	0.47	0.34	0.26	0.23
$C_{je}(\Delta V_{\text{logic}} / I_C)$	$1/\gamma^{3/2}$	180 fs	94 fs	50 fs	33 fs
$R_{ex} / (\Delta V_{\text{logic}} / I_c)$	γ^0	0.24	0.24	0.24	0.24

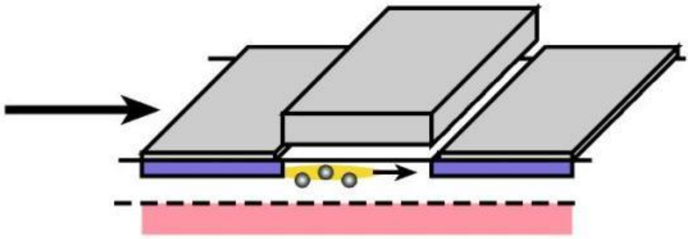
For Later: FETs

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

$$C_{gd}/W_g \sim \varepsilon$$

$$g_m/W_g \sim v\varepsilon/T_{ox}$$

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

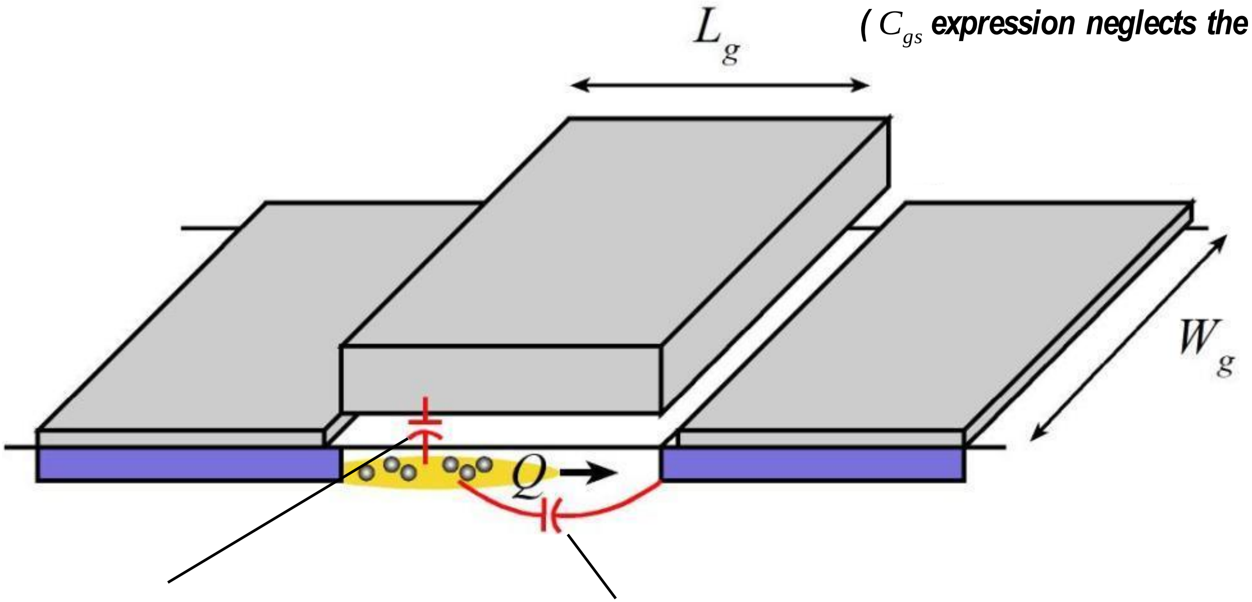
$$C_{gs,f}/W_g \sim \varepsilon$$

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

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 ✗

FET scaling: Output Conductance & DIBL

(C_{gs} expression neglects the effect of a finite density of states)



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

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

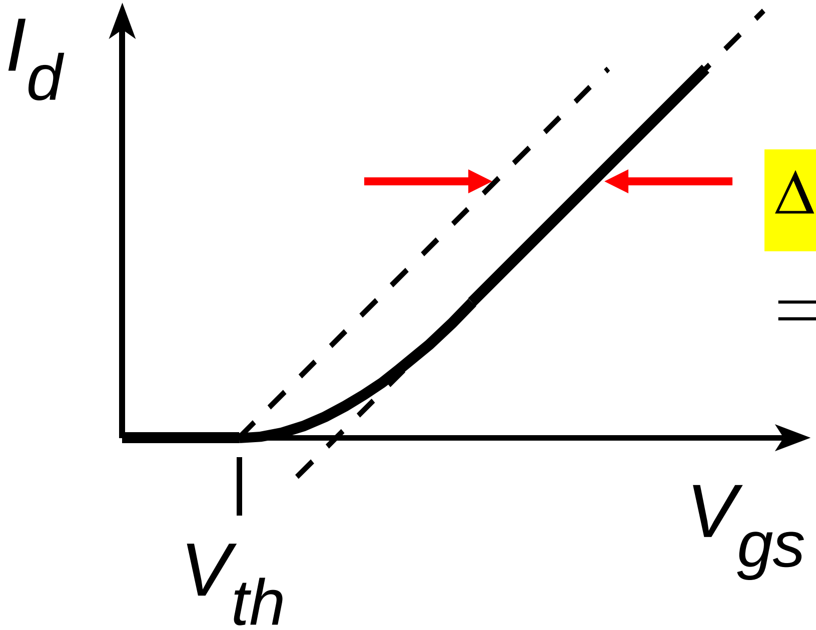
transconductance ↑ ($C_{gs}\delta V_{gs}$) output conductance ↑ ($C_{d-ch}\delta V_{ds}$)

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

FETs: Think about Mass, Not Mobility

Simple drift-diffusion theory, nondegenerate, far above threshold:

$$I_D \approx c_{ox} W_g v_{injection} (V_{gs} - V_{th} - \Delta V) \quad \text{where } v_{injection} \sim v_{thermal} = (kT / m^*)^{1/2}$$



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

$$\Rightarrow \text{Ensure that } \Delta V << (V_{gs} - V_{th})$$

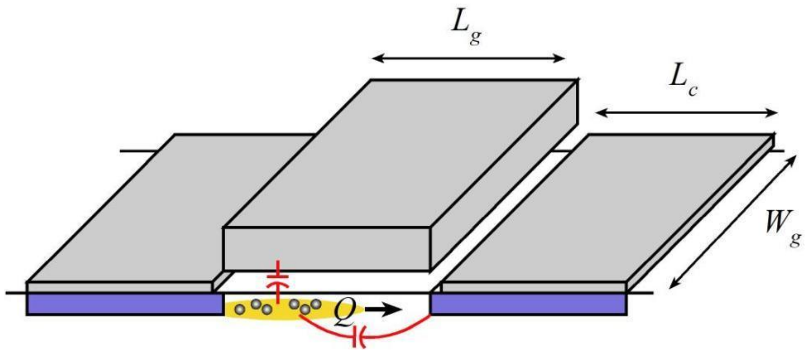
$$\sim 700 \text{ mV}$$

low effective mass → high currents

mobilities above ~ 1000 cm^2/V-s of little benefit at 22 nm Lg

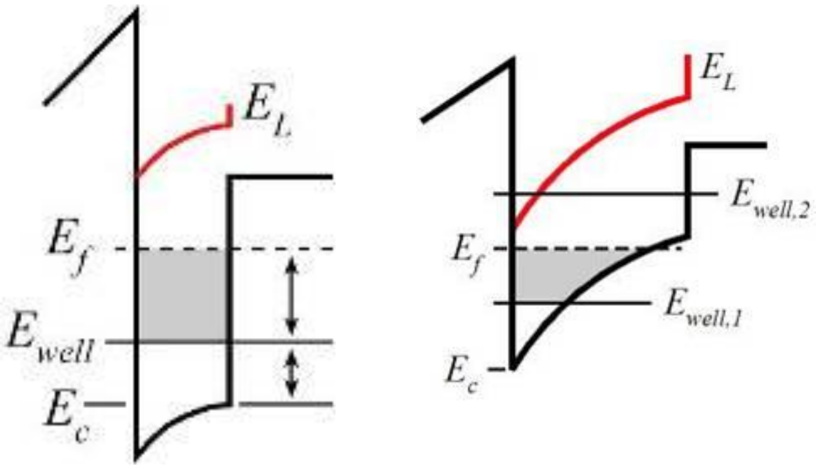
Low Effective Mass Impairs Vertical Scaling

Shallow electron distribution needed for high g_m / G_{ds} ratio, low drain-induced barrier lowering.



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

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.
→ Hard to scale below 22 nm L_g.

Density-Of-States Capacitance

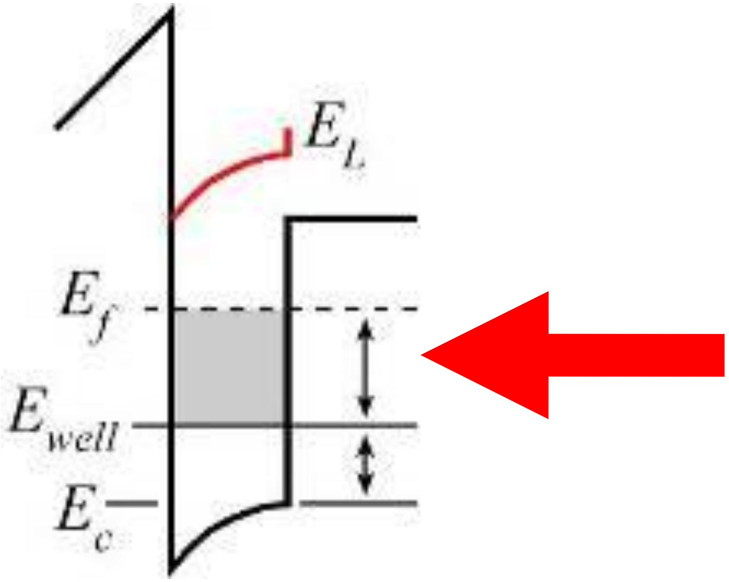
$$E_f - E_{well} = n_s / (nm^* / 2\pi\hbar^2)$$

$$\downarrow$$

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

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

and n is the # of band minima



Two implications:

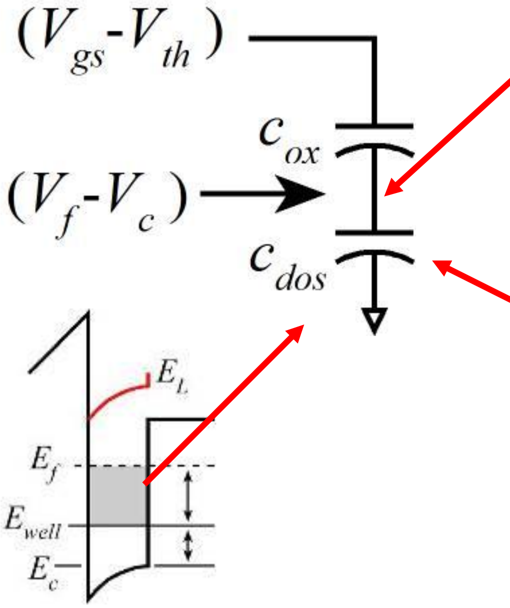
- With N_s >10^{13}/cm^2, electrons populate satellite valleys

Fischetti et al, IEDM2007

- Transconductance dominated by finite state density

Solomon & Laux , IEDM2001

Drive Current in the Ballistic & Degenerate Limits



Channel Fermi voltage = voltage applied to c_{dos}

determines Fermi velocity v_f through $E_f = qV_f = m^* v_f^2 / 2$

Channel charge : $\rho_s = c_{dos}\left(V_f - V_c\right) = \dfrac{c_{dos}c_{ox}}{c_{ox} + c_{dos}}\left(V_{gs} - V_{th}\right)$

$c_{dos} = q^2 n m^* / 2\pi\hbar^2 = c_{dos,o} \cdot n \cdot (m^* / m_o)$, where n is the # of band minima

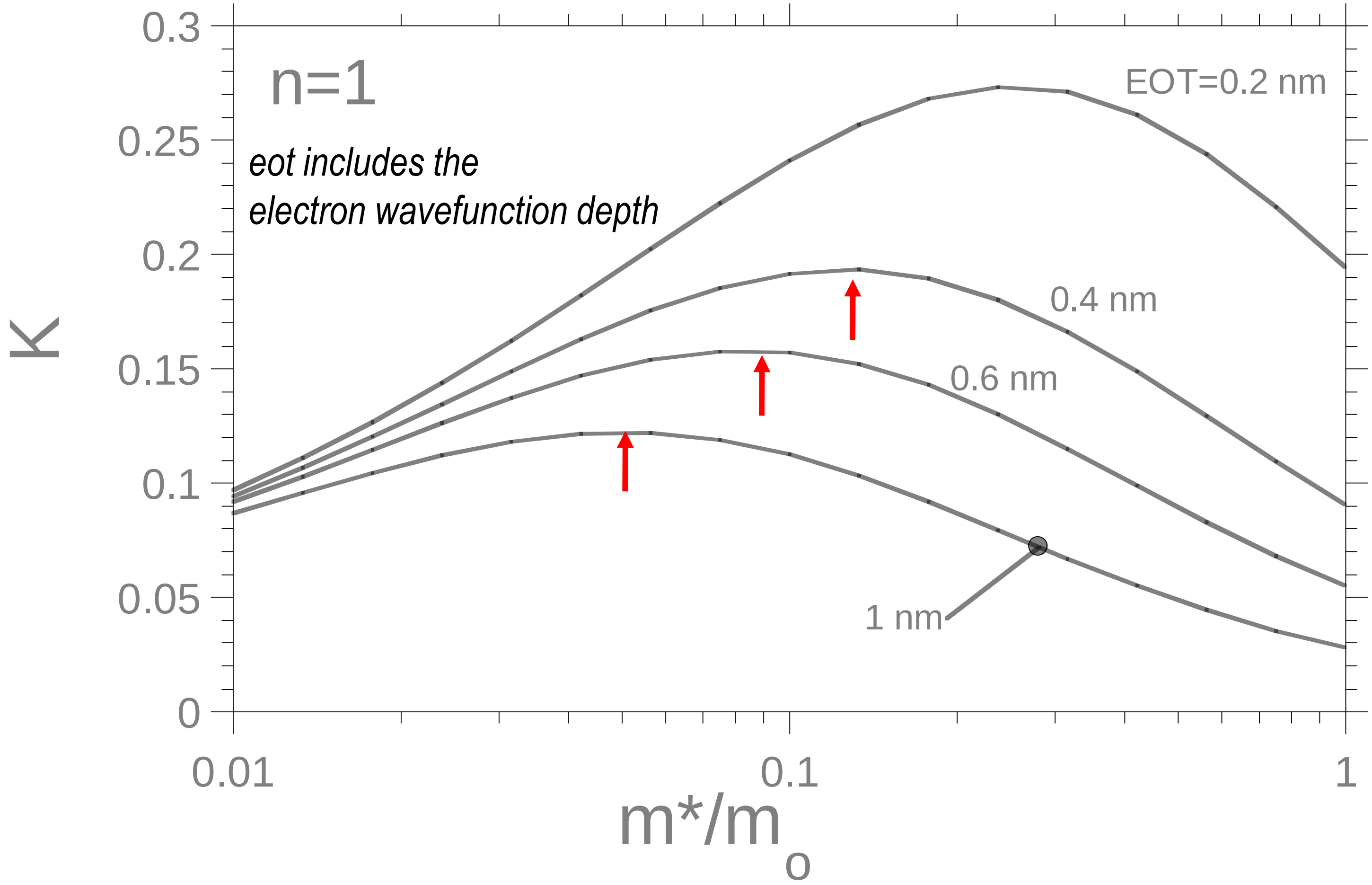
$$\Rightarrow \quad J = \left(84\,\frac{\text{mA}}{\mu\text{m}}\right)\cdot\left(\frac{V_{gs} - V_{th}}{1\,\text{V}}\right)^{3/2} \cdot \frac{n\cdot\left(m^*/m_o\right)^{1/2}}{\left(1 + (c_{dos,o}/c_{ox})\cdot n\cdot(m^*/m_o)\right)^{3/2}}$$

Ballistic but nondegenerate case : $J \approx \left(kT/m^*\right)^{1/2} c_{ox}\left(V_{gs} - V_{th}\right)$

More careful analyses by Taur & Asbeck Groups, UCSD; Fischetti Group: U-Mass: IEDM2007

Drive Current in the Ballistic & Degenerate Limits

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



eot includes the electron wavefunction depth

Inclusive of non-parabolic band effects, which increase c_{dos} ,

InGaAs & InP have near-optimum mass for 0.4-1.0 nm EOT gate dielectrics

Rough Projections From Simple Ballistic Theory

22 nm gate length

0.5-1.0 fF/μm parasitic capacitances

Channel	*EOT*	*drive current* *(700 mV overdrive)*	*intrinsic gate capacitance*
InGaAs	*1 nm*	*6 mA/μm*	*0.2 fF/μm*
InGaAs	*1/2 nm*	*8.5 mA/μm*	*0.25 fF/μm*
Si	*1 nm*	*2.5-3.5 mA/μm*	*0.7 fF/μm*
Si	*1/2 nm*	*5-7 mA/μm*	*1.4 fF/μm*

InGaAs has much less gate capacitance

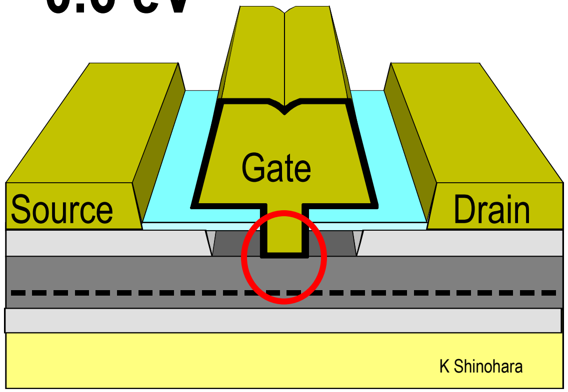
1 nm EOT → InGaAs gives much more drive current

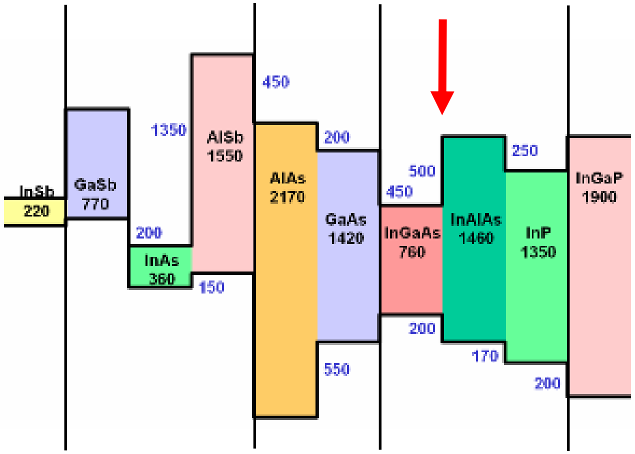
1/2 nm EOT → InGaAs & Si have similar drive current

InGaAs channel→ no benefit for sub-22-nm gate lengths

HEMTs have Low Gate Barriers: Limits Scaling !

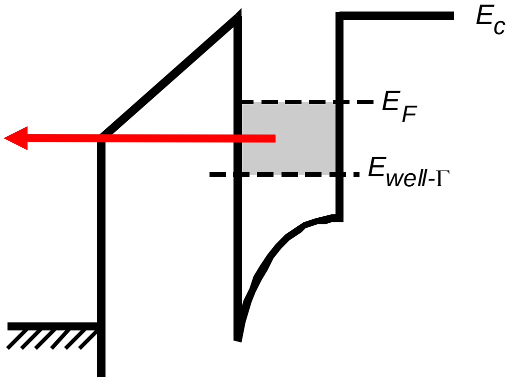
Gate barrier is low: ~0.6 eV

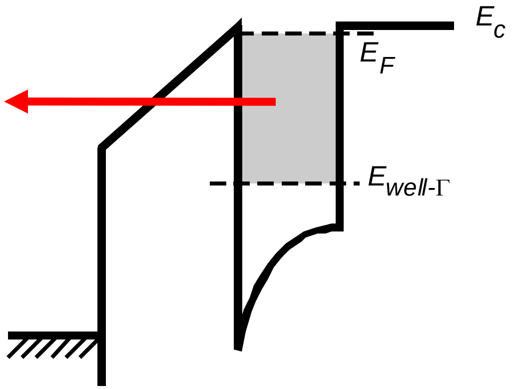




Tunneling through barrier
→ sets minimum thickness

Emission over barrier
→ limits 2D carrier density

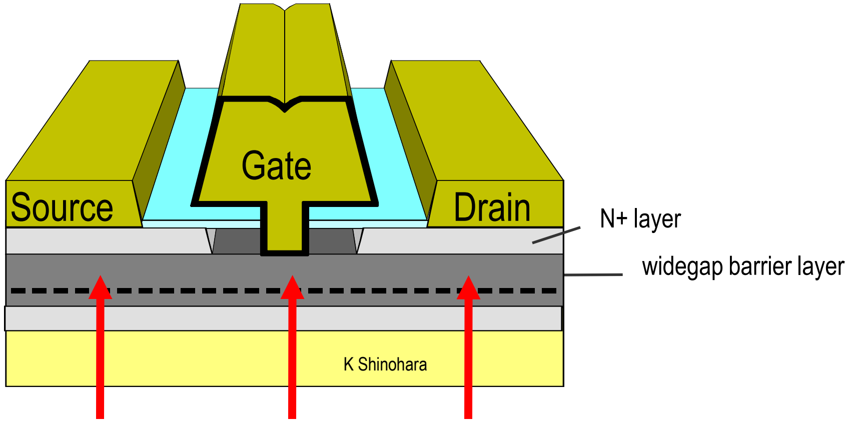




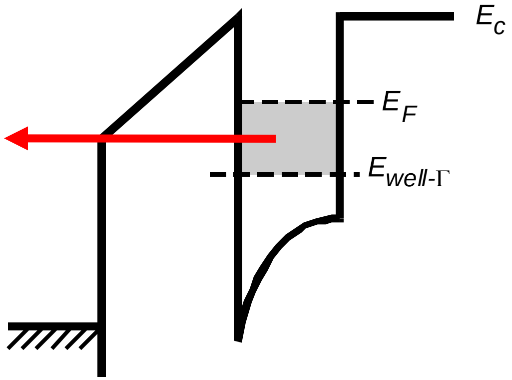
At $N_s = 10^{13}/\text{cm}^2, (E_f - E_c) \sim 0.6\,\text{eV}$

HEMT Gate Barrier Ruins S/D Contacts

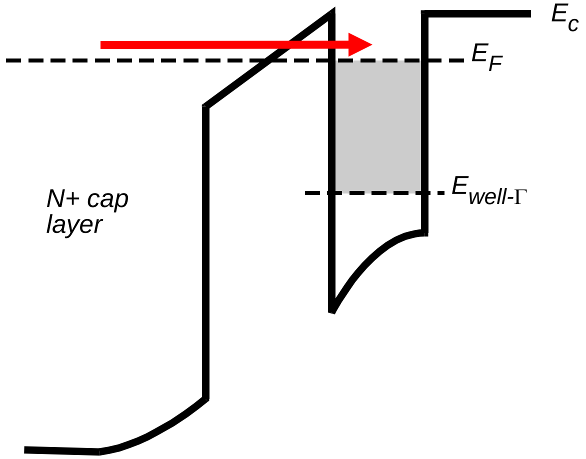
Gate barrier also lies under source / drain contacts



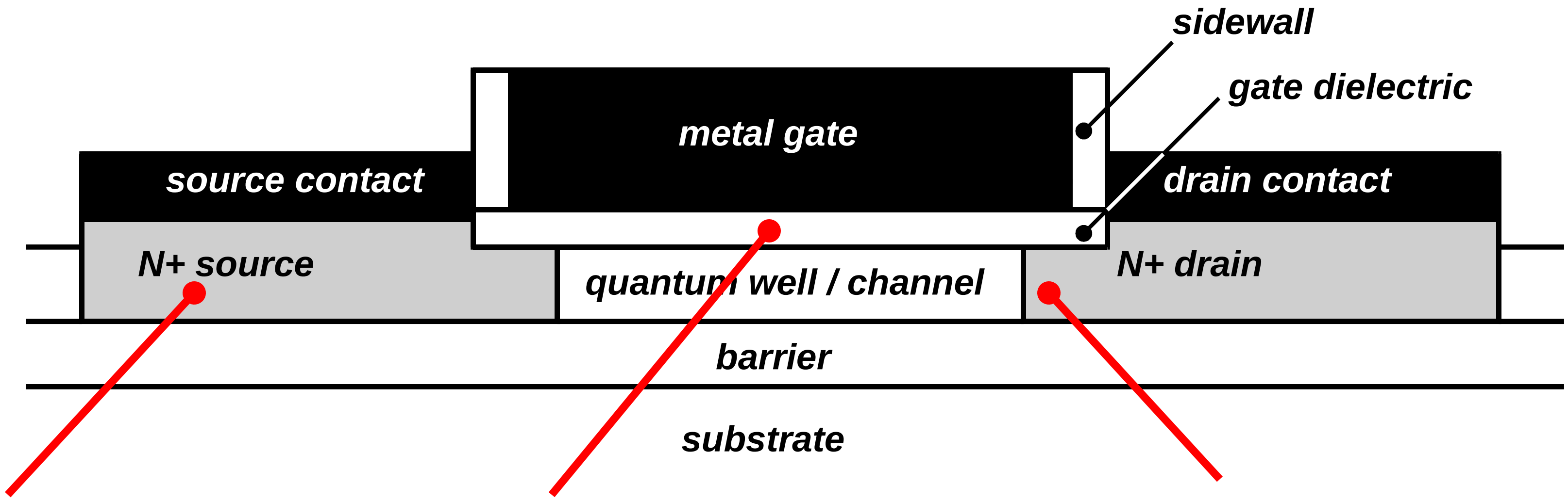
low leakage:
need high barrier under gate



low resistance:
need low barrier under contacts



The Structure We Need -- is Much Like a Si MOSFET



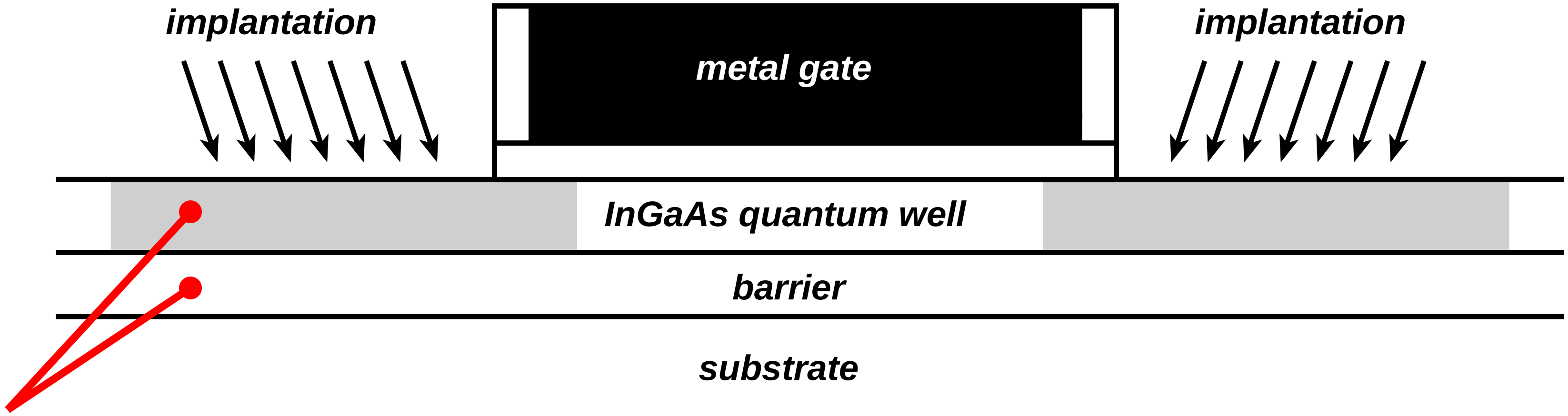
no gate barrier under S/D contacts

high-K gate barrier

Overlap between gate and N+ source/drain

How do we make this device ?

Source/Drain Implantation Does Not Look Easy



Implantation will intermix InGaAs well & InAlAs barrier

Annealing can't fix this.

Incommensurate sublimation of III vs. V elements during anneal

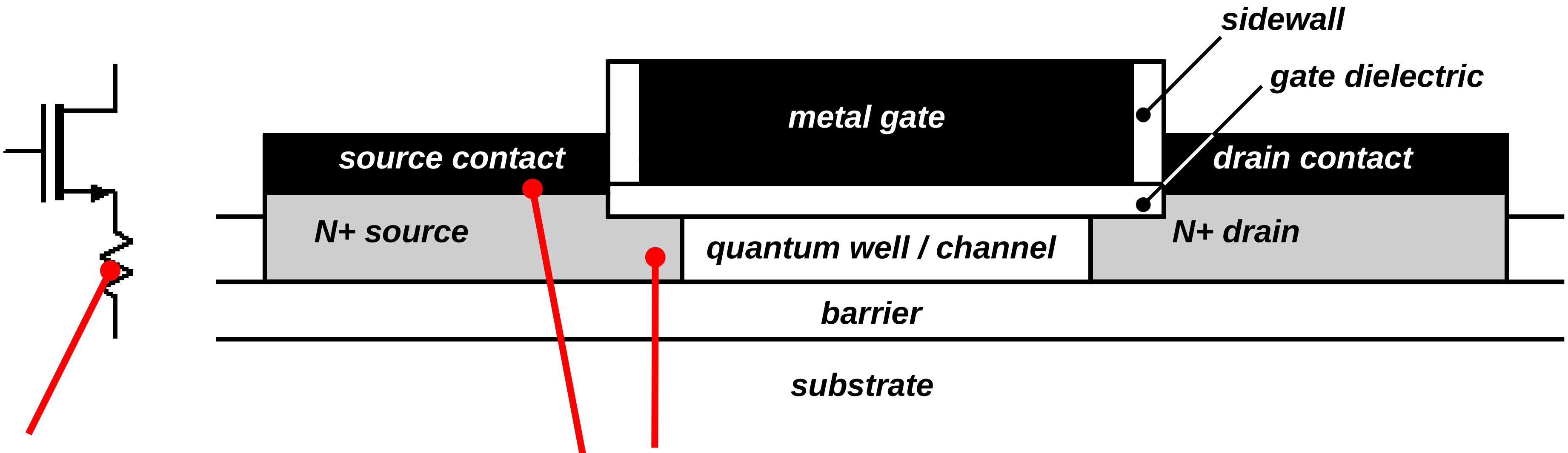
Need ~ 5 nm implant depth & ~ $6*10^{19}$ /cm^3 doping

Implanted structures have not shown the necessary low contact resistivity.

Source Access for a 2 THz FET

~5 nm thick well 1 nm Insulator EOT

Target ~7 mA/µm @ 700 mV gate overdrive



For $< 10\%$ impact on drive current,

$I_D R_S < 70\ \text{mV}.$

$\rightarrow R_s < 10\ \Omega - \mu\text{m}$

$(20\ \text{nm N} + \text{extension}) \times (100\ \Omega/\text{square}) = 2\ \Omega - \mu\text{m}$

$(0.25\ \Omega - \mu\text{m}^2)\ /(25\ \text{nm wide contact}) = 10\ \Omega - \mu\text{m}$