2009 Topical Workshop on Heterostructure Microelectronics, August 25-28, Nagano, Japan,

Process Technologies For Sub-100-nm InP HBTs & InGaAs MOSFETs

Mark. Rodwell,
University of California, Santa Barbara

M. A. Wistey*, U. Singisetti, G. J. Burek, B. J. Thibeault, A. Baraskar, E. Lobisser, V. Jain, J. Cagnon, S. Stemmer, A. C. Gossard
University of California, Santa Barbara
*Now at Notre Dame

E. Kim, P. C. McIntyre
Stanford University

Y.-J. Lee
Intel

B. Yue, L. Wang, P. Asbeck, Y. Taur
University of California, San Diego

III-V transistors: the goal is scaling

2-3 THz InP HBTs: 32 nm / 64 nm scaling generations

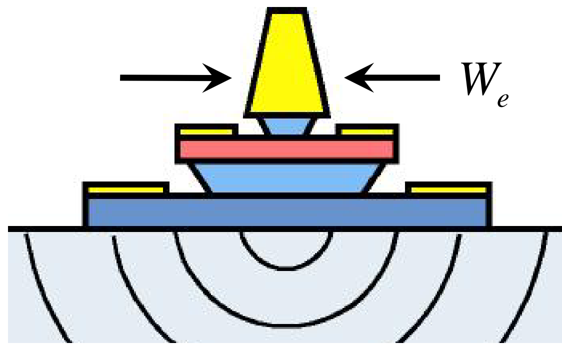
2-3 THz HEMTs: 10-15 nm, balanced / fully scaled devices

15 nm InGaAs MOSFETs for VLSI

implication:

we need new fabrication processes

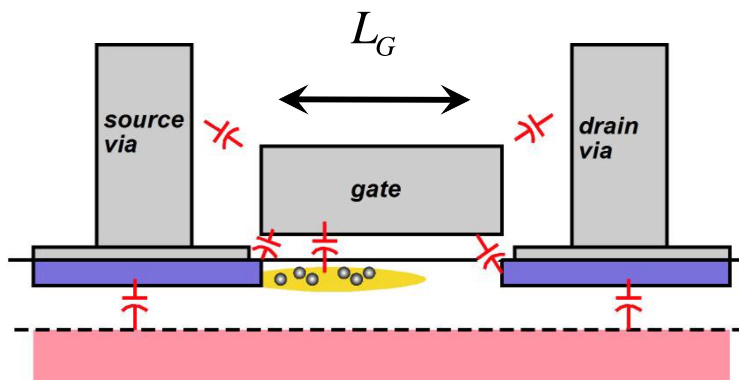
Changes required to double transistor bandwidth



$(\text{emitter length } L_E)$

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

nearly constant junction temperature → linewidths vary as (1 / bandwidth)²



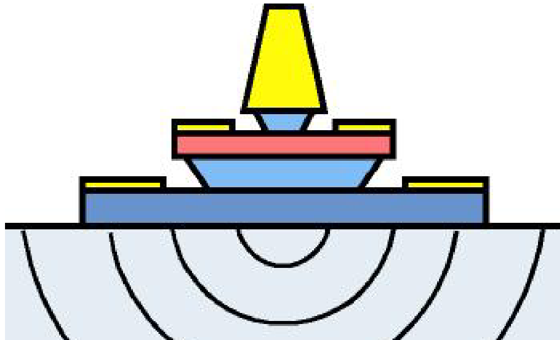
$(\text{gate width } W_G)$

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

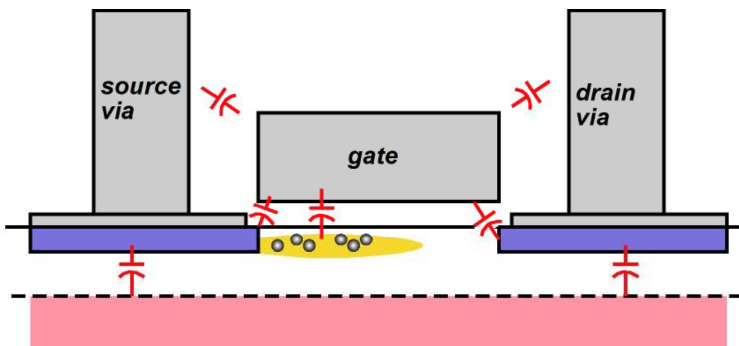
constant voltage, constant velocity scaling

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

III-V Fabrication Processes Must Change... Greatly



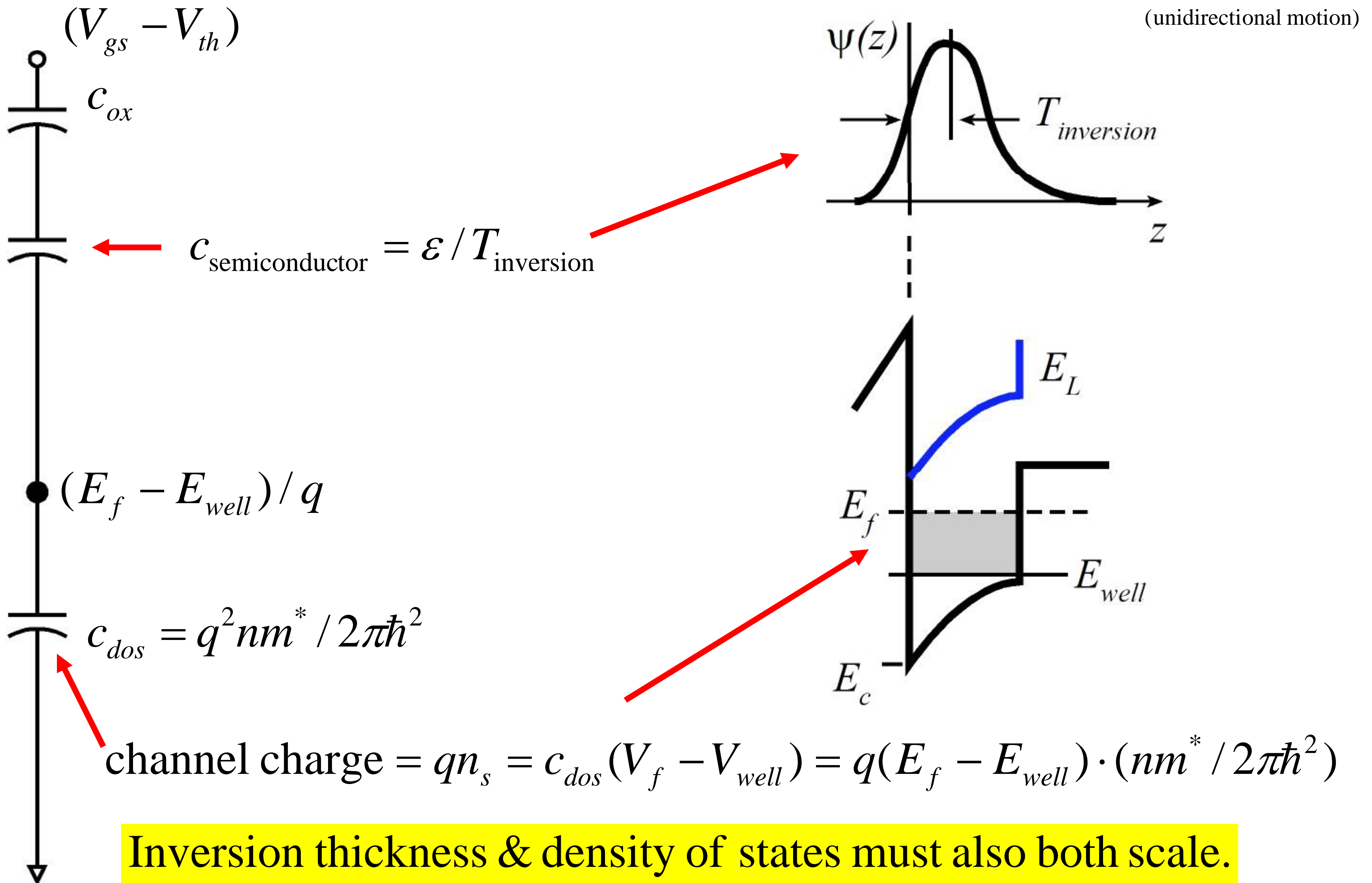
32 nm base & emitter contacts...self-aligned
32 nm emitter junctions
1 $\Omega-\mu m^2$ contact resistivities
70 mA/μm^2 $\rightarrow$ refractory contacts



15 nm gate length
15 nm source / drain contacts...self-aligned
< 10 nm source / drain spacers (sidewalls)
1/2 $\Omega-\mu m^2$ contact resistivities
3 mA/μm $\rightarrow$ 200 mA/μm^2
contacts above ~ 5 nm N+ layer
$\rightarrow$ refractory contacts !

FETs

Semiconductor Capacitances Must Also Scale



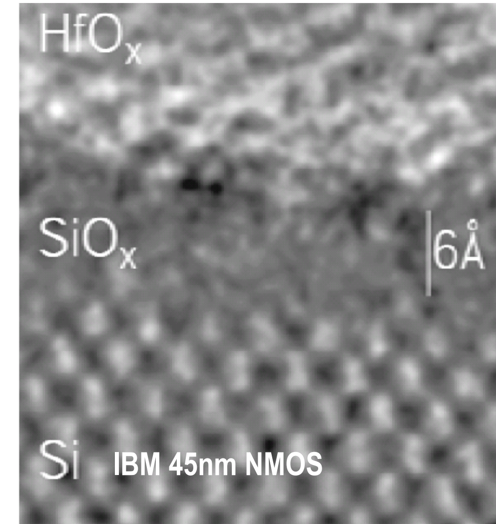
Highly Scaled FET Process Flows

Why III-V MOSFETs

Silicon MOSFETs:

Gate oxide may limit <16 nm scaling

$I_d / W_g \sim c_{ox}(V_g - V_{th}) v_{inj}$



Narayan *et al*, VLSI 2006

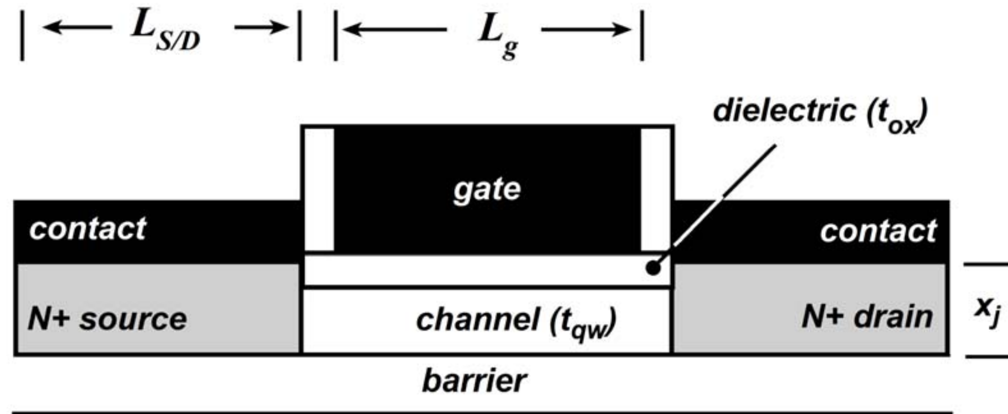
Alternative: $In_{0.53}Ga_{0.47}As$ channel MOSFETs

low m (0.041 m_o) $\rightarrow$ high injection velocity, v_{inj} (~ 2-3×10^7 cm/s)**

→increase drive current, decreased CV/I

* Enoki *et al* , EDL 1990

MOSFET scaling*: lateral and vertical



Goal :

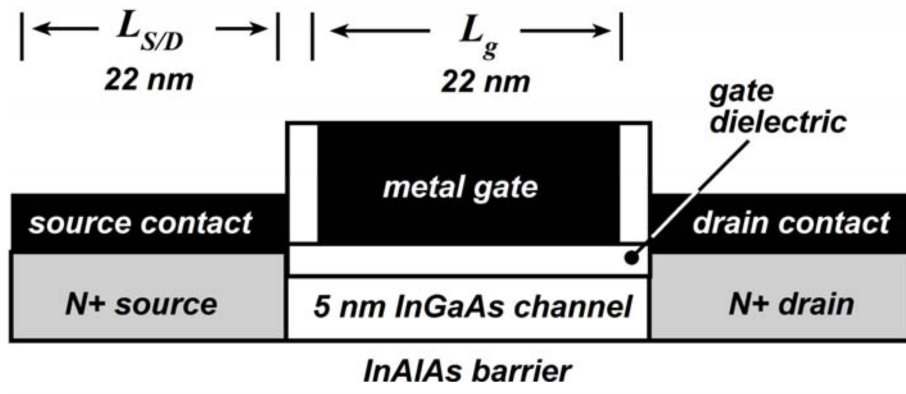
double package density → *lateral scaling* L_g, W_g, $L_{s/d}$

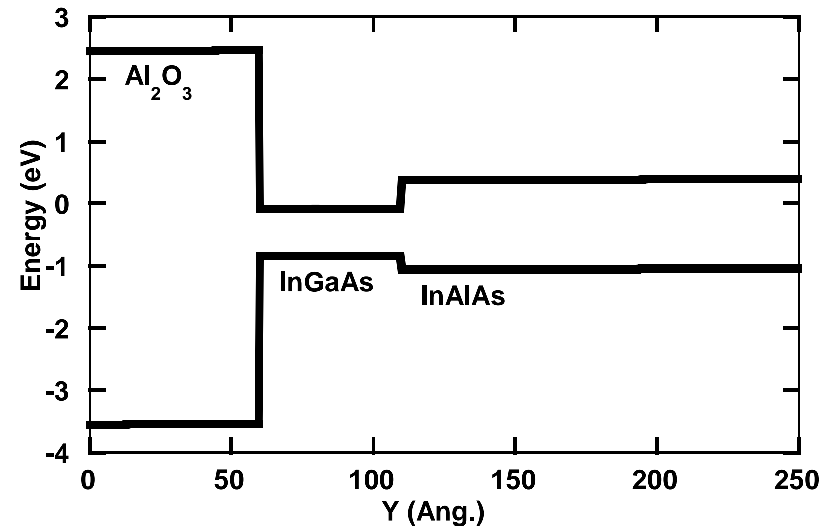
double the MOSFET speed

keep constant gate control → *vertical scaling* t_{ox}, t_{qw}, x_j

*Rodwell, IPRM 2008

Target device structure



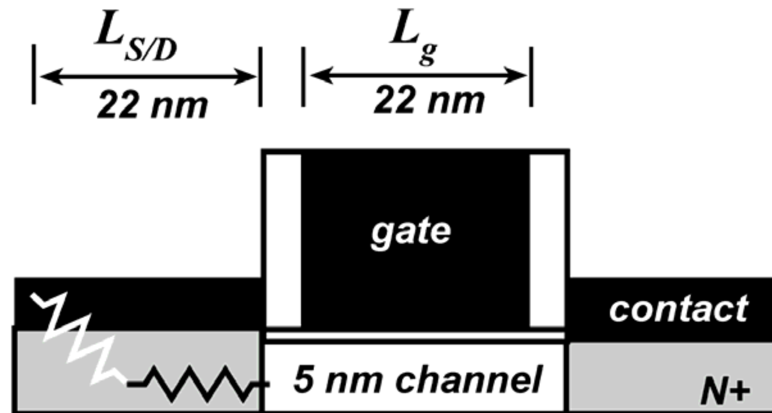


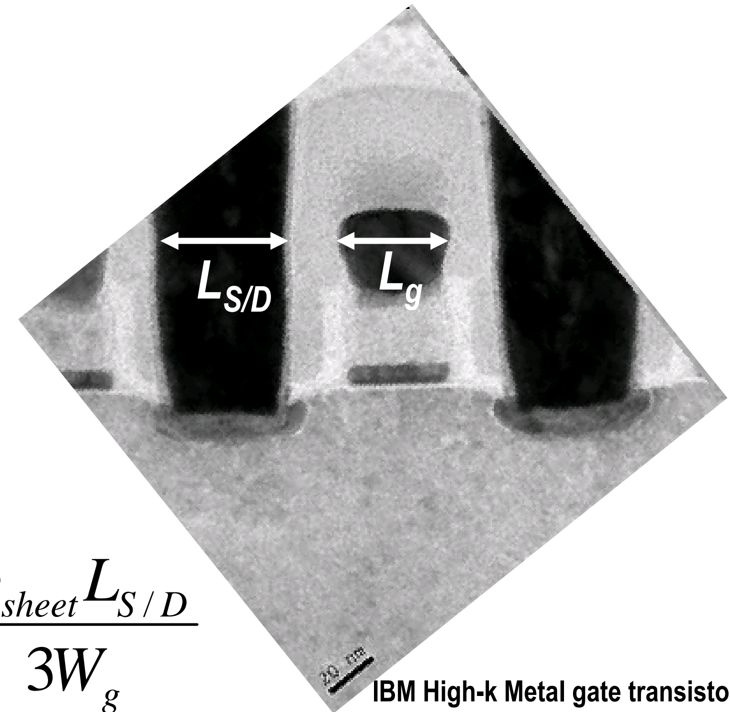
Target 22 nm gate length

Control of short-channel effects → vertical scaling
1 nm EOT: thin gate dielectric, surface-channel device
5 nm quantum well thickness
<5 nm deep source / drain regions

~3 mA/μm target drive current → low access resistance
self-aligned, low resistivity source / drain contacts
self-aligned N+ source / drain regions with high doping

22 nm InGaAs MOSFET: Source Resistance





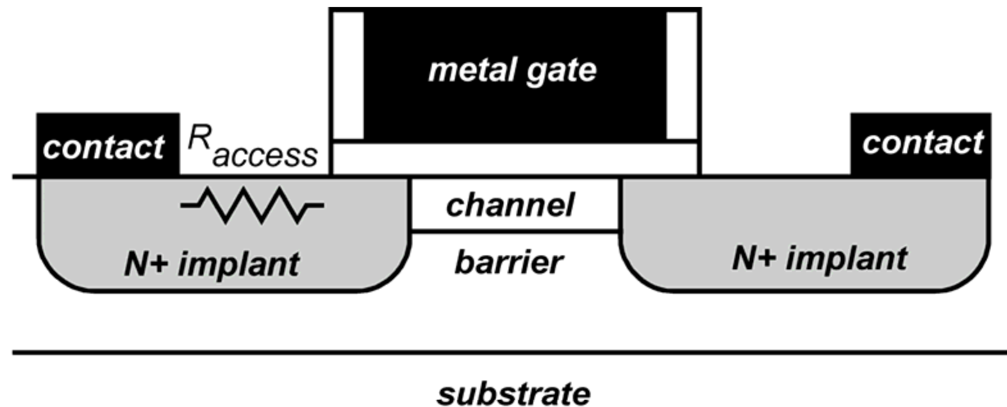
IBM High-k Metal gate transistor

Image Source: EE Times

$$I_d = \frac{I_{di}}{1 + g_{mi} \cdot R_s} \qquad R_s = \frac{\rho_c}{W_g L_{S/D}} + \frac{\rho_{sheet} L_{S/D}}{3W_g}$$

- Source access resistance degrades I_d and g_m
- IC Package density : $L_{S/D} \sim L_g$ =22 nm $\rightarrow$ ρ_c *must be low*
- Need low sheet resistance in thin ~5 nm N+ layer
- Design targets: ρ_c ~1 Ω–μm^2, ρ_{sheet} ~ 400 Ω

22nm ion implanted InGaAs MOSFET



Key Technological Challenges

- Shallow junctions (~ 5 nm), high (~5×10^{19} cm^{-3}) doping
- Doping abruptness (~ 1 nm/decade)
- Lateral Straggle (~ 5 nm)
- Deep junctions would lead to degraded short channel effects

Why HEMTs are Hard to Improve

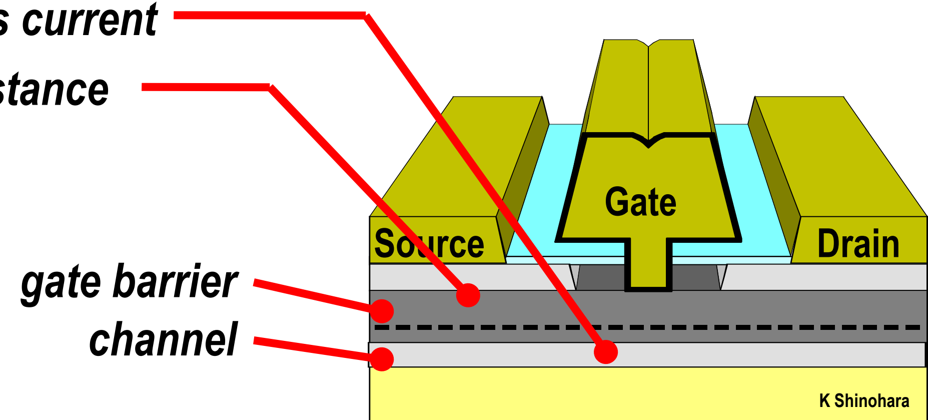
1st challenge with HEMTs: reducing access resistance

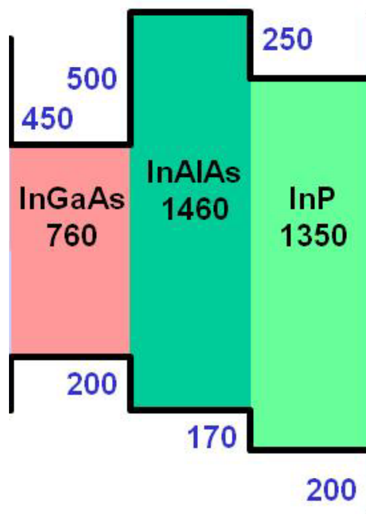
low electron density under gate recess→ limits current

gate barrier lies under S/D contacts → resistance

gate barrier

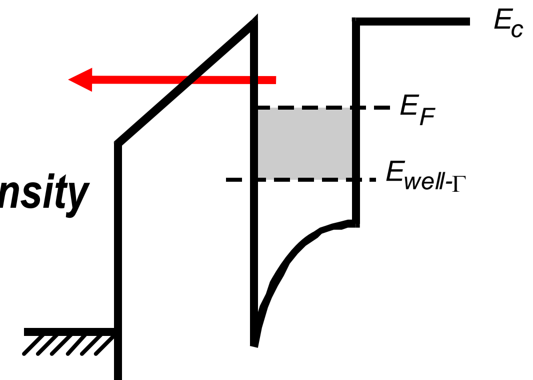
channel





2nd challenge with HEMTs: low gate barrier

high tunneling currents with thin barrier
high emission currents with high electron density

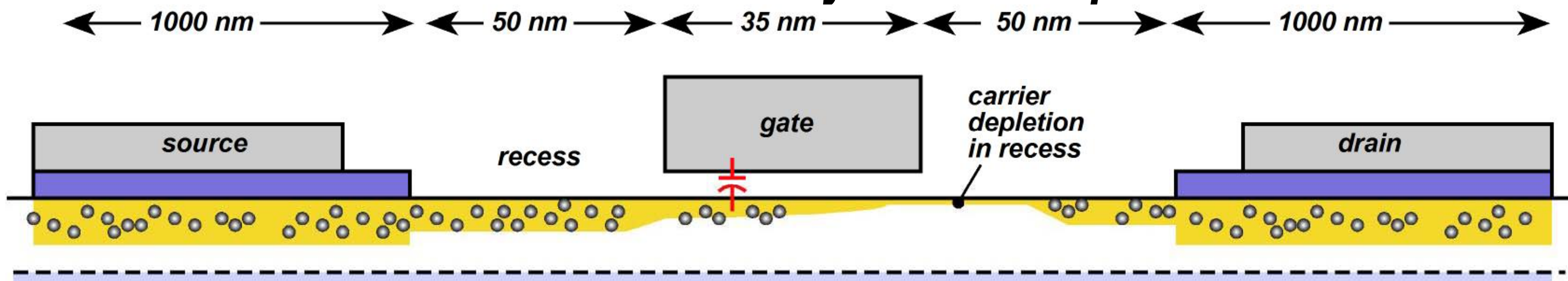


III-V MOSFETs do not face these scaling challenges

HEMTs Differ in Access Resistance, Electrostatics

HEMTs: *short gate lengths, wide spacing / recess, wide contacts*

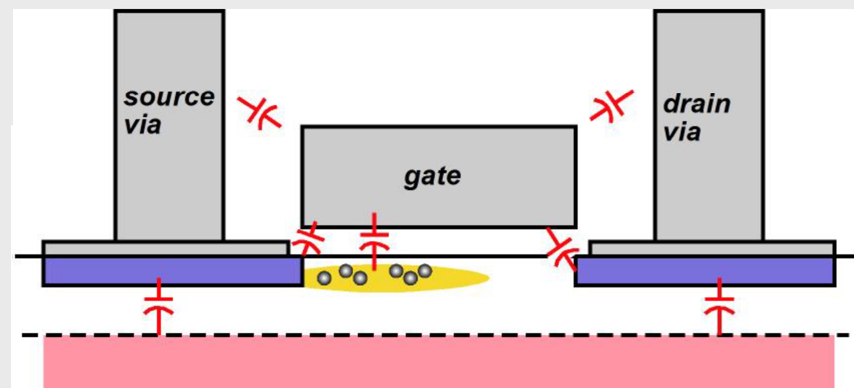
wide recess→ improved DIBL, improved subthreshold slope,
wide contacts→ OK access resistivity even with poor contacts



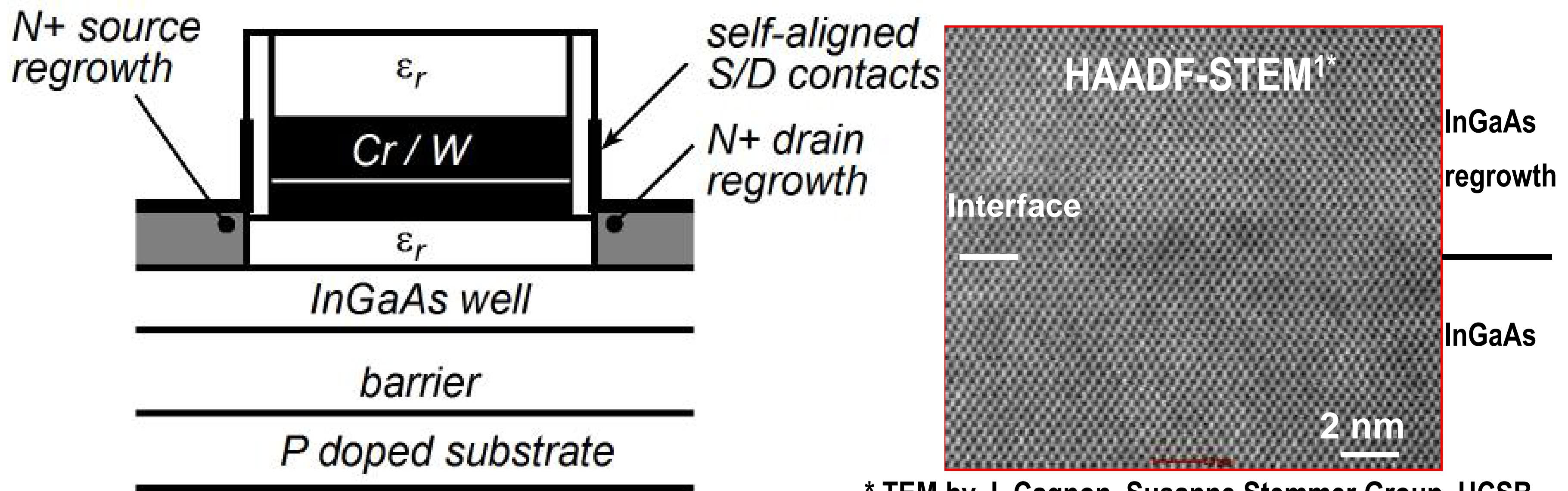
VLSI MOSFETs *: short gate lengths, narrow contacts, no spacing/recess*

Need good DIBL even with zero drain/gate offset.

Need low S/D resistance even with 22 nm width contacts.



InGaAs MOSFET with N+ Source/Drain by MEE Regrowth[1]



* TEM by J. Cagnon, Susanne Stemmer Group, UCSB

Self-aligned source/drain defined by MBE regrowth[2]

Self-aligned in-situ Mo contacts[3]

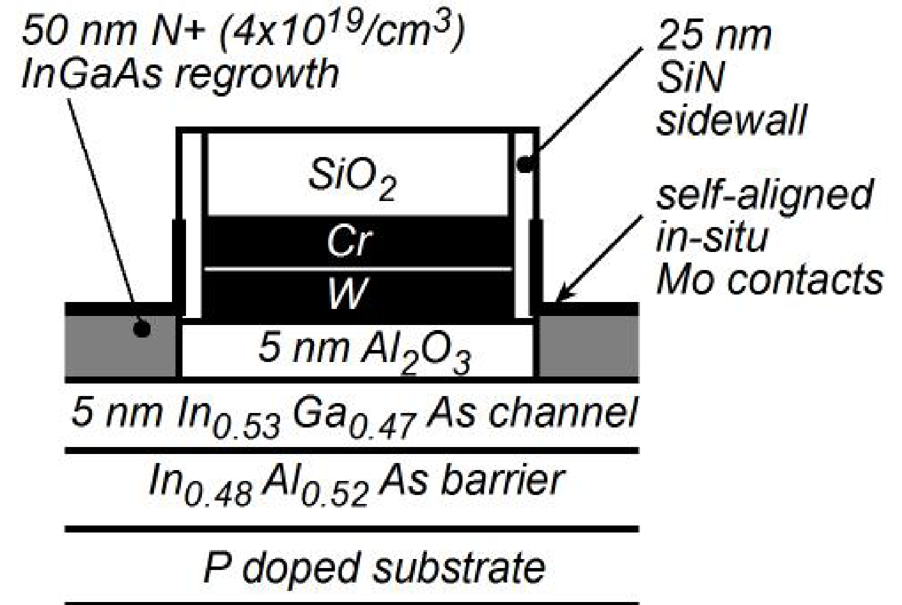
Process flow & dimensions selected for 22 nm L_g design; present devices @ 200 nm gate length

[1]Singisetti, ISCS 2008

[2]Wistey, EMC 2008

[3]Baraskar, EMC 2009

Regrown S/D process: key features



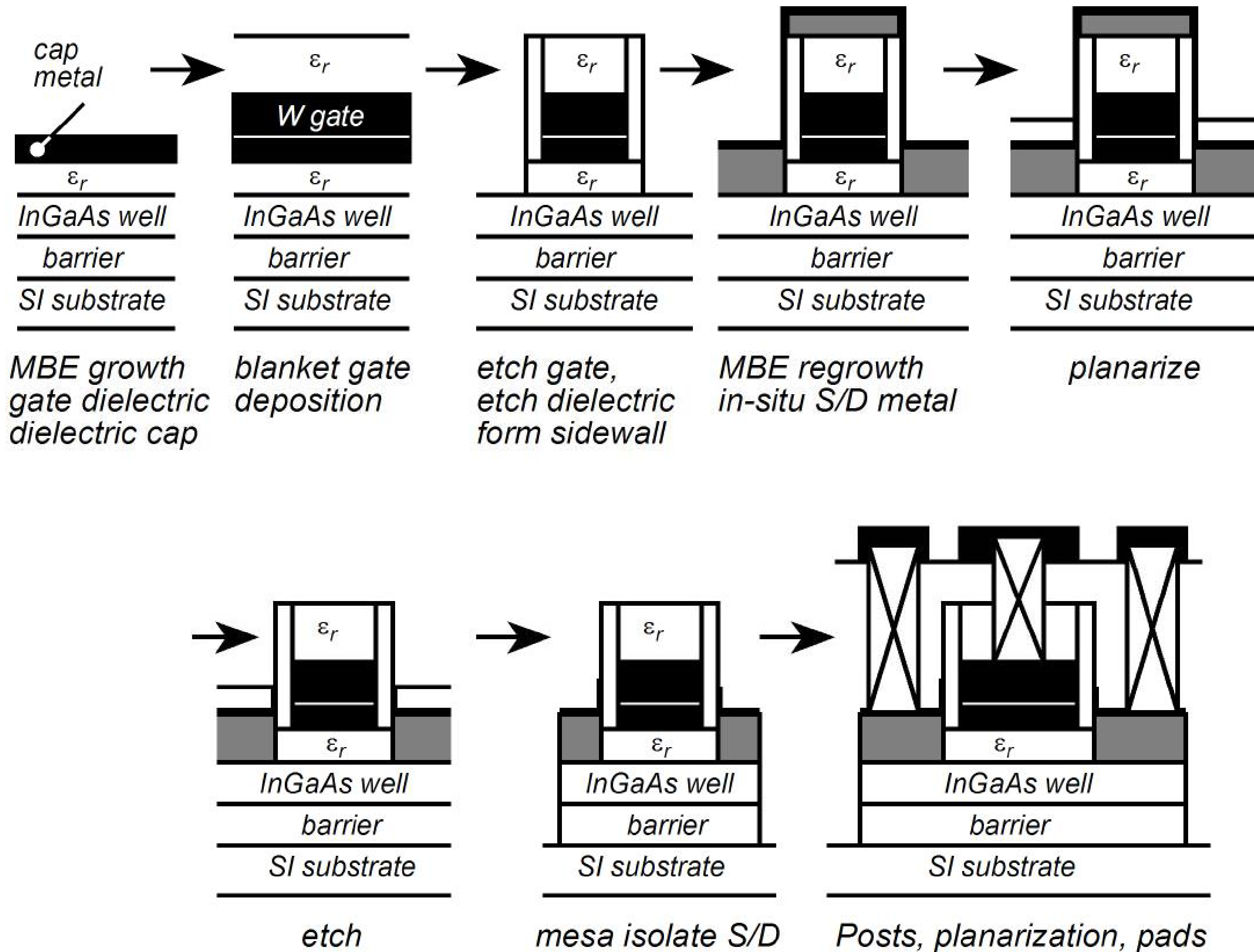
Self-aligned <u>& low resistivity</u>
...source / drain N+ regions
...source / drain metal contacts

Vertical S/D doping profile set by MBE
no n+ junction extension below channel
abrupt on few-nm scale

Gate-first
gate dielectric formed after MBE growth
uncontaminated / undamaged surface

Process flow*

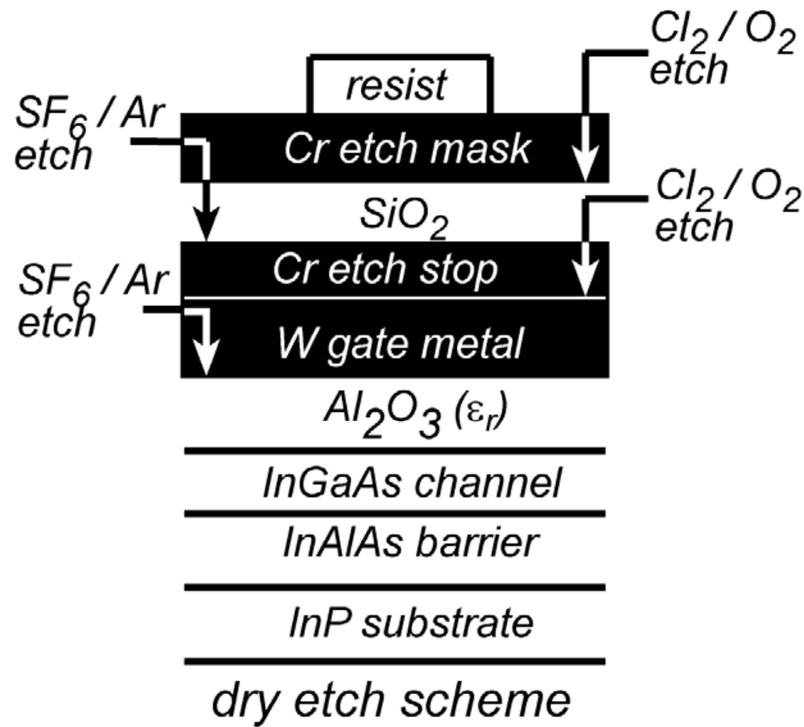
* Singisetti et al, 2008 ISCS, September, Frieburg
Singisetti *et al*; Physica Status Solidi C, vol. 6, pp. 1394,2009

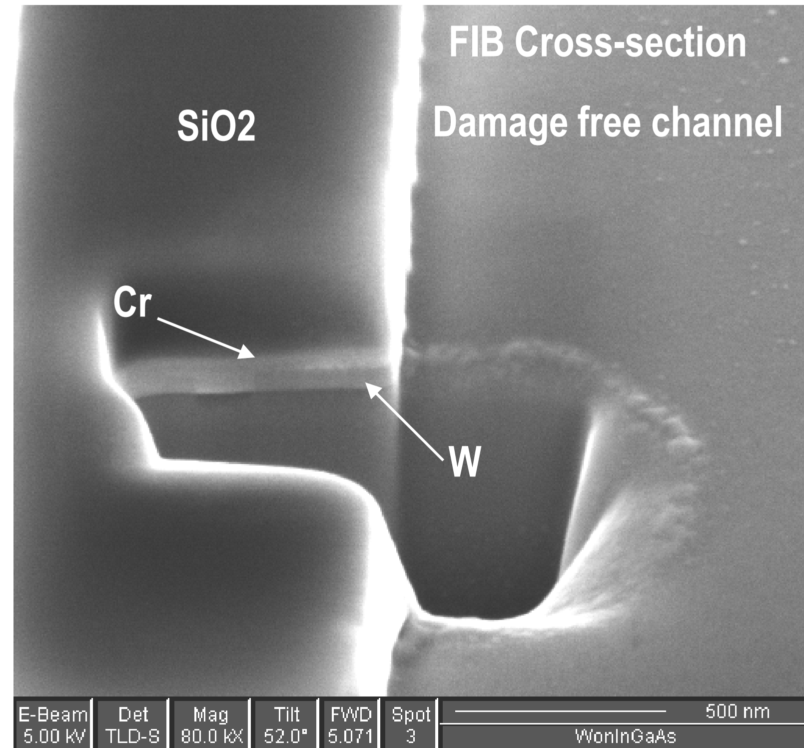


Key challenge in S/D process: gate stack etch

Requirement: avoid damaging semiconductor surface:

Approach: Gate stack with multiple selective etches*

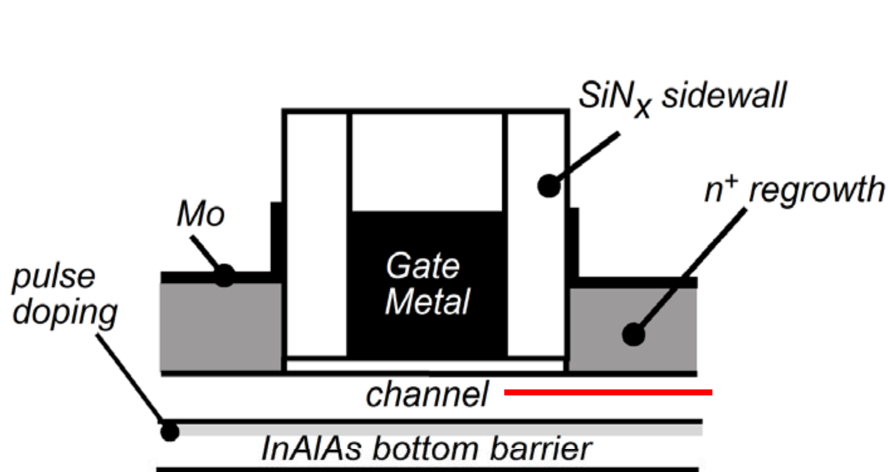


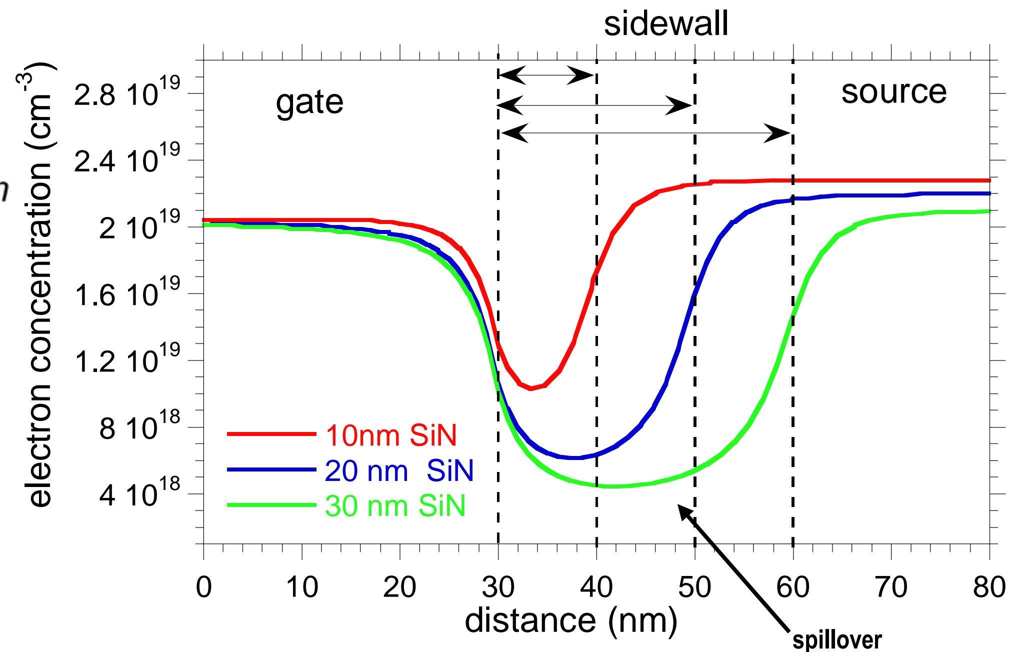


Process scalable to sub-100 nm gate lengths

*** Singisetti *et al*; Physica Status Solidi C, vol. 6, pp. 1394,2009**

Key challenge in S/D process: dielectric sidewall





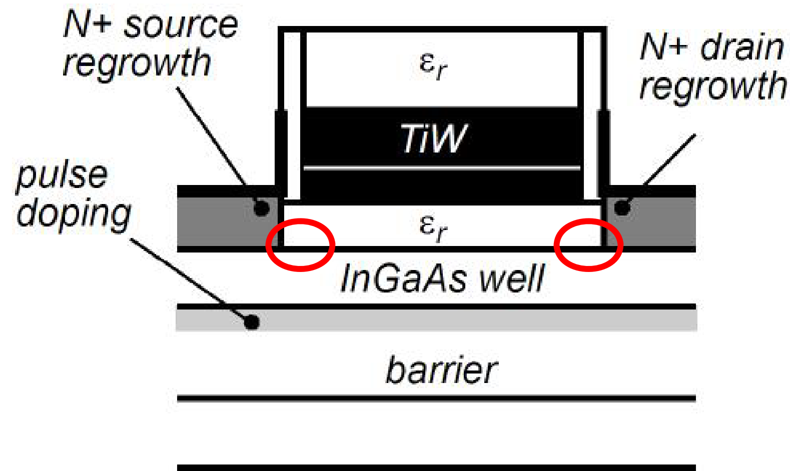
n_s under sidewall:
electrostatic spillover from source, gate

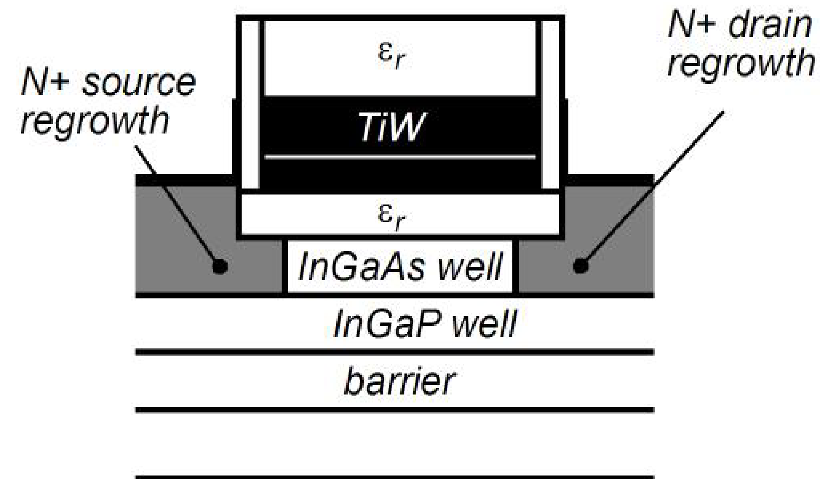
Sidewall must be kept thin:
avoid carrier depletion,
avoid source starvation

t_{sw}	n (cm^{-3})	R_s (Ω–μm)
10 nm	> 1×10^{19}	6
20 nm	> 5×10^{18}	20
30 nm	~ 4×10^{18}	60

2-DSimulation of an artificially on state device in Atlas, Silvaco. Source doping 6e19 cm^{-3}

Raised vs. Recessed S/D Regrowth:





planar regrowth

need thin sidewalls (now ~25nm)

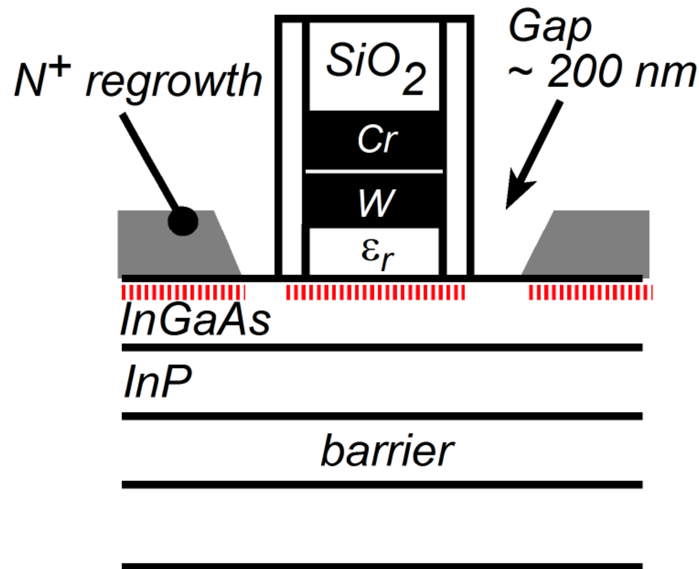
High D_{it} ? → severe carrier depletion

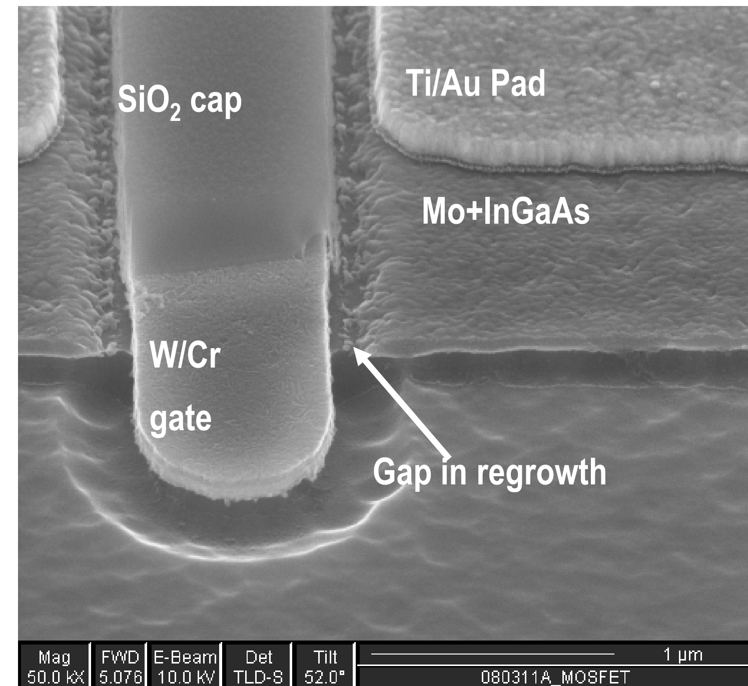
regrowth under sidewalls

more difficult growth...

...tolerate of high D_{it} in access region

MBE Regrowth→ Gap Near Gate→ Source Resistance





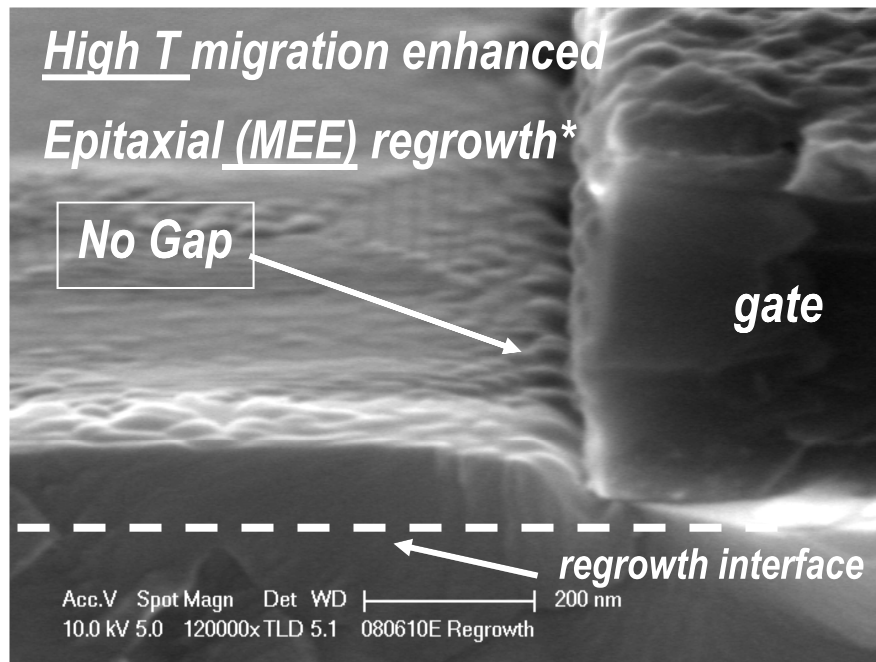
- ***Shadowing by gate: No regrowth next to gate***

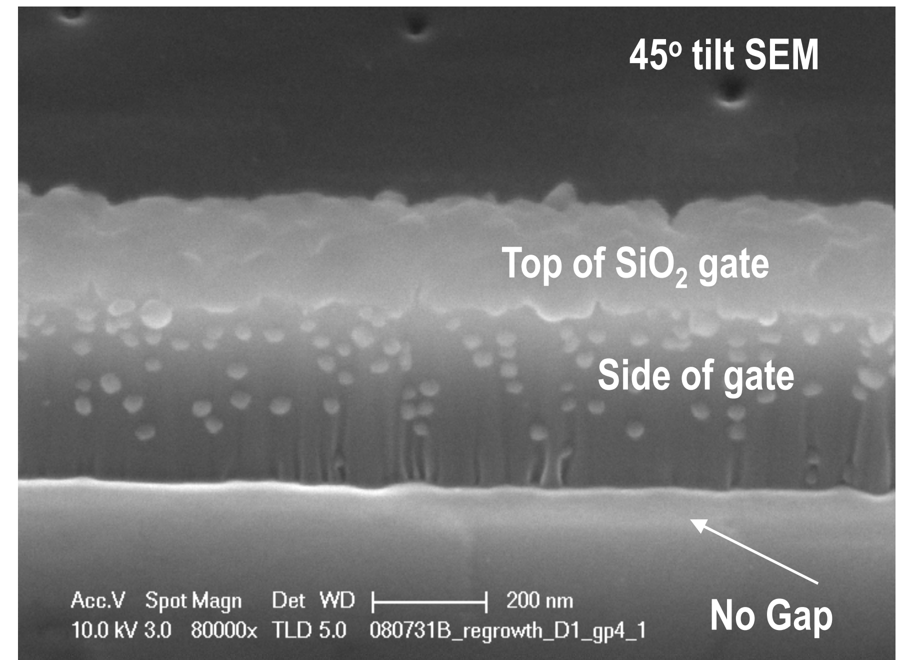
- ***Gap region is depleted of electrons***

High source resistance because of electron depletion in the gap

MBE growth by Dr. Mark Wistey, device fabrication and characterization by U. Singisetti

Migration Enhanced Epitaxial (MEE) S/D Regrowth*





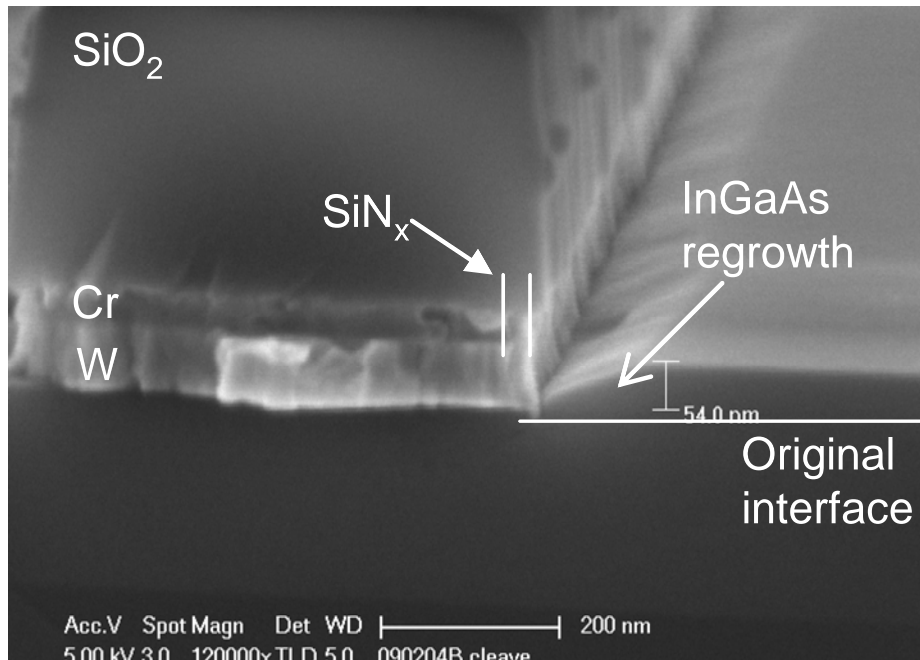
High temperature migration enhanced epitaxial regrowth

***Wistey, EMC 2008**

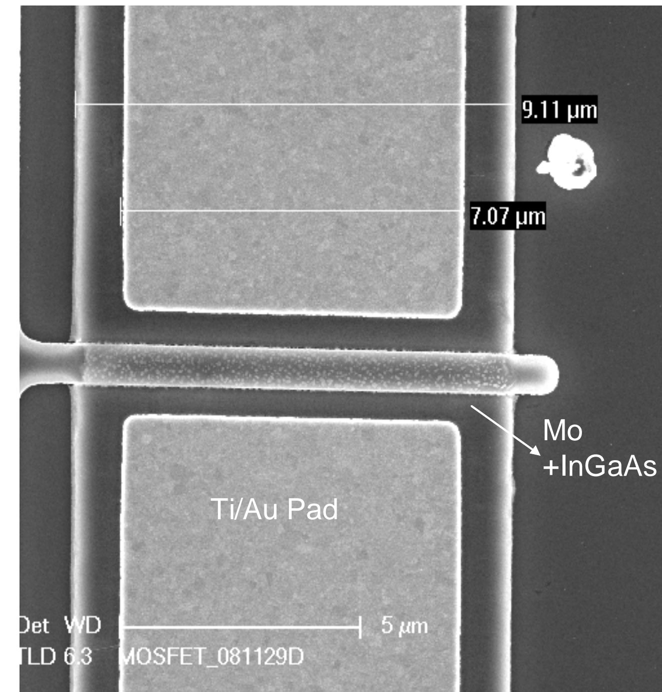
Wistey, ICMBE 2008

MBE growth by Dr. Mark Wistey, device fabrication and characterization by U. Singisetti

Regrown S/D III-V MOSFET: Images

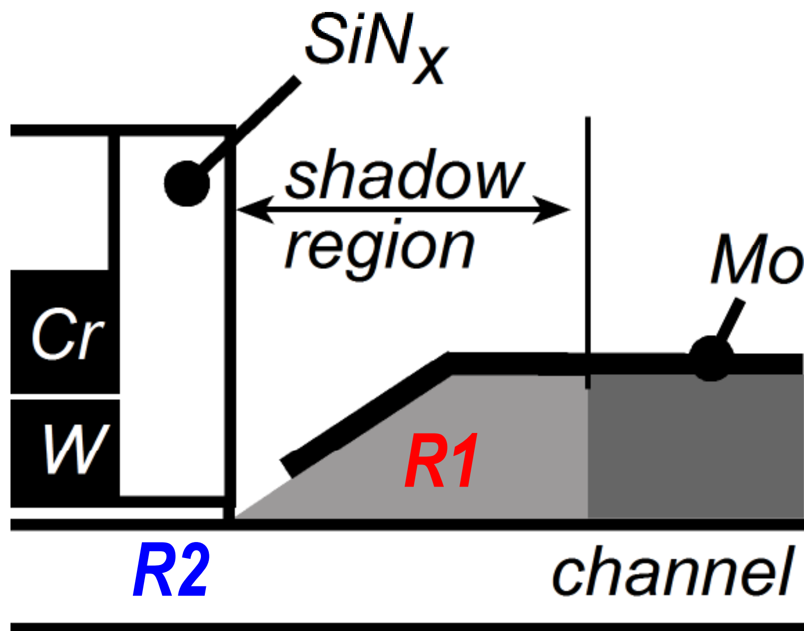


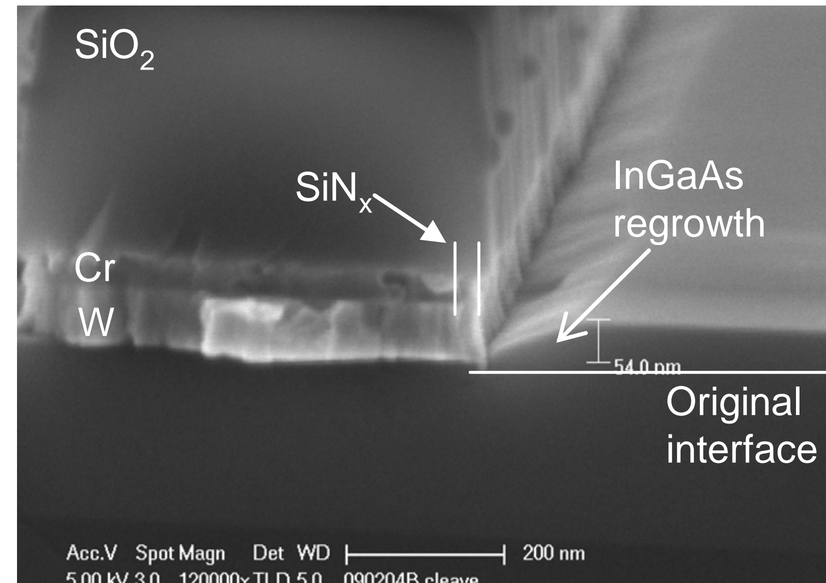
Cross-section after regrowth, but before Mo deposition



Top view of completed device

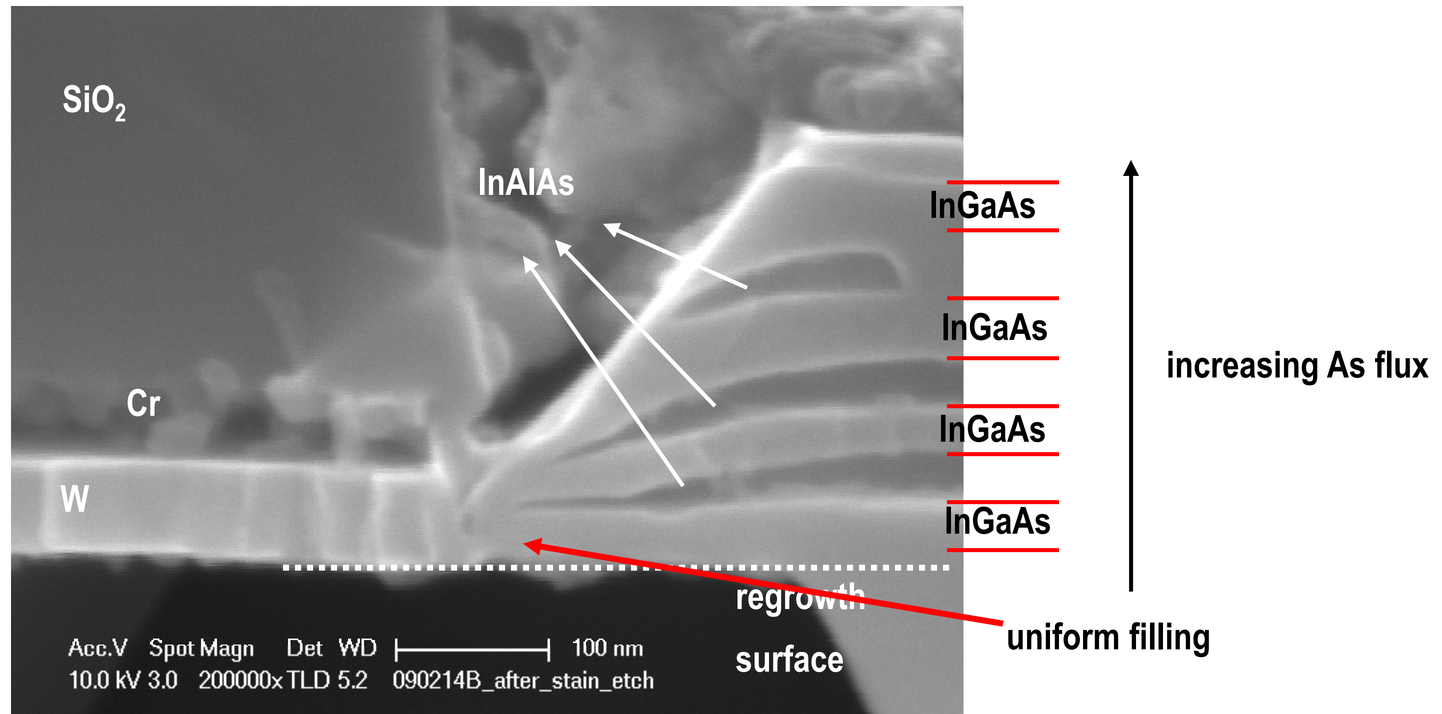
Source Resistance: electron depletion near gate





- *Electron depletion in regrowth shadow region (R_1)*
- *Electron depletion in the channel under SiN_x sidewalls (R_2)*

Regrowth profile dependence on As flux*



multiple InGaAs regrowths with InAlAs marker layers

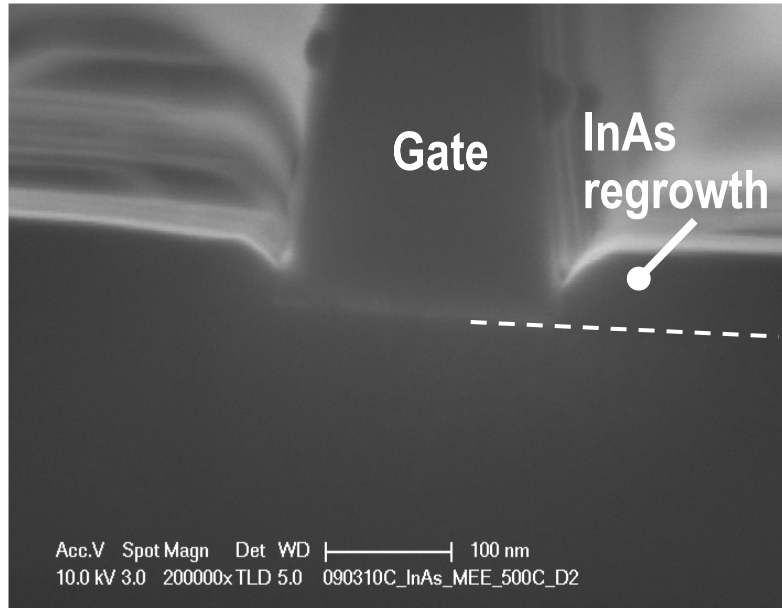
Uniform filling with lower As flux

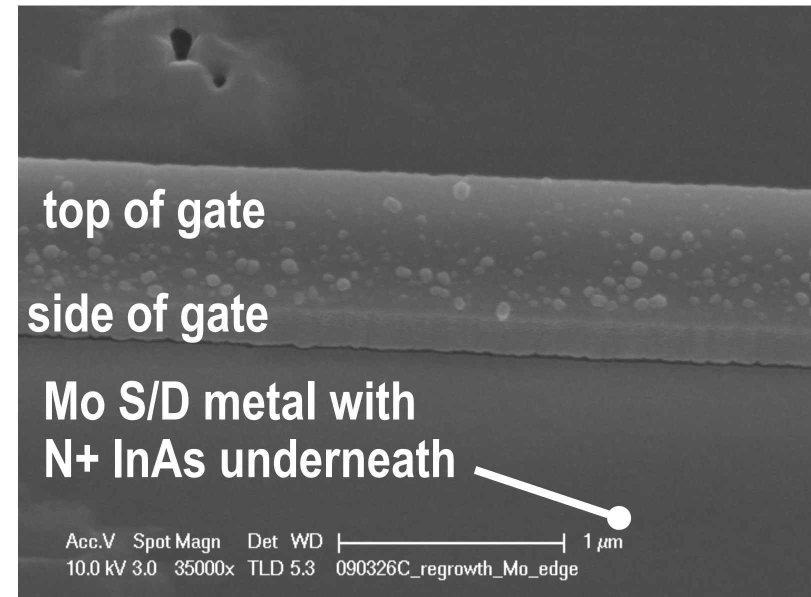
* Wistey *et al*, EMC 2009

Wistey *et al* NAMBE 2009

MBE growth by Dr. Mark Wistey, device fabrication and characterization by U. Singisetti

InAs source/drain regrowth





***Improved InAs regrowth with low As flux for uniform filling*[1]**

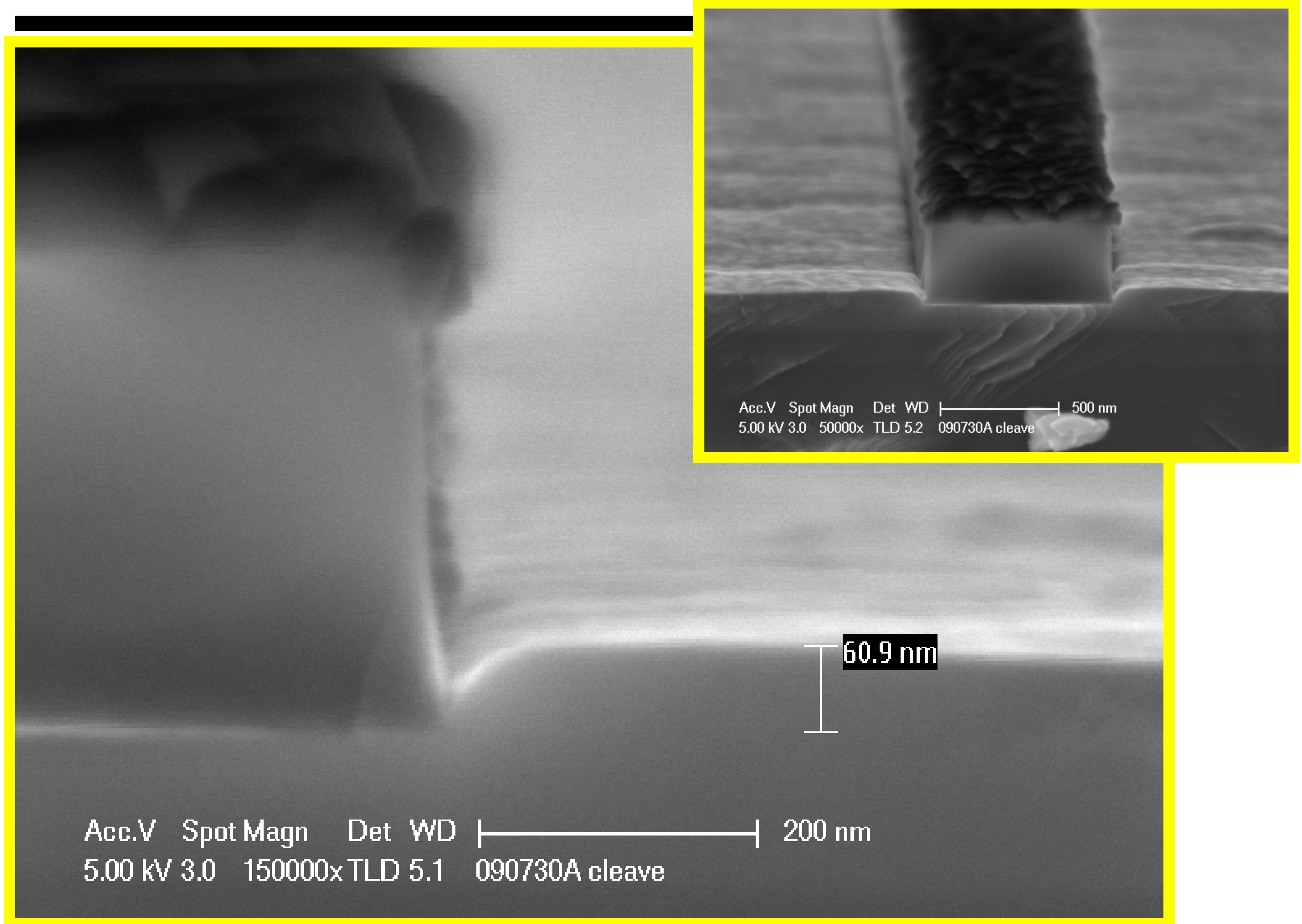
***InAs less susceptible to electron depletion: Fermi pinning above E_c*[2]**

[1] Wistey *et al*, EMC 2009

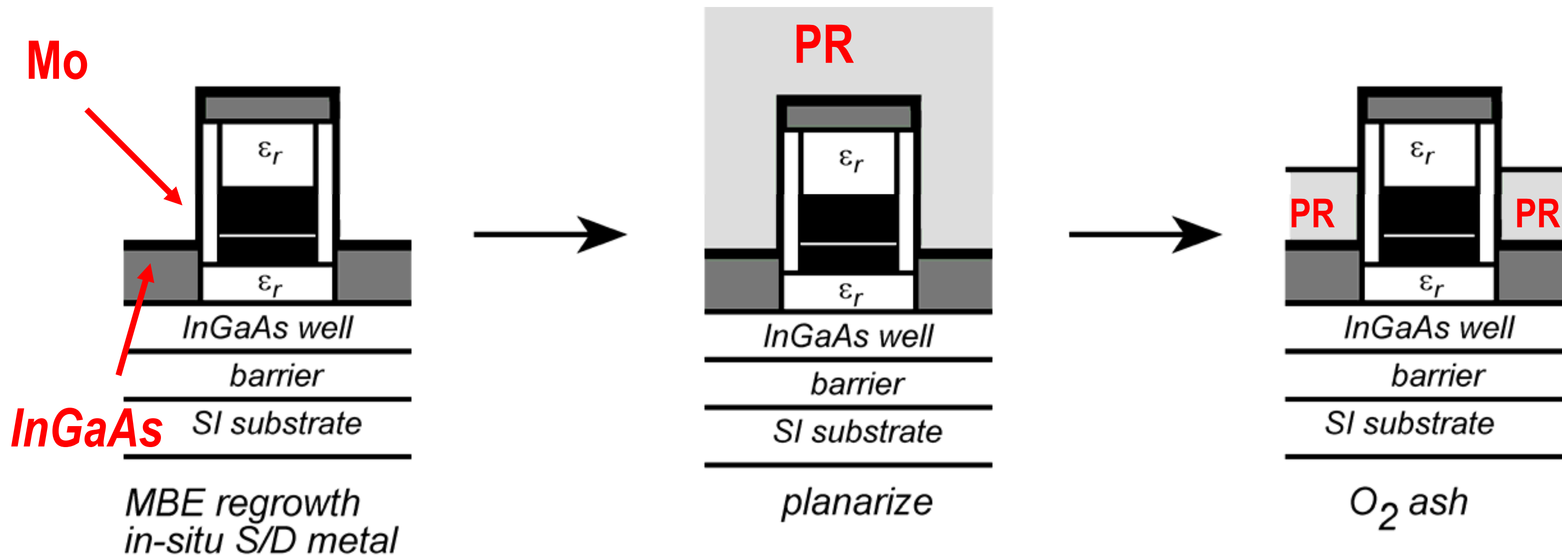
Wistey *et al* NAMBE 2009.

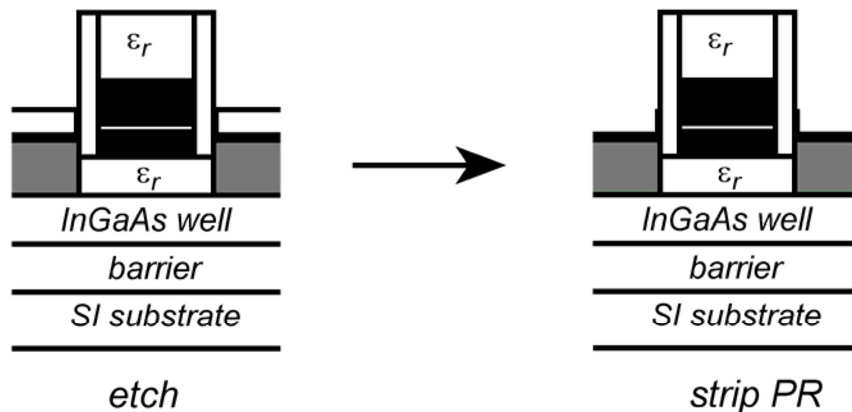
[2]Bhargava *et al* , APL 1997

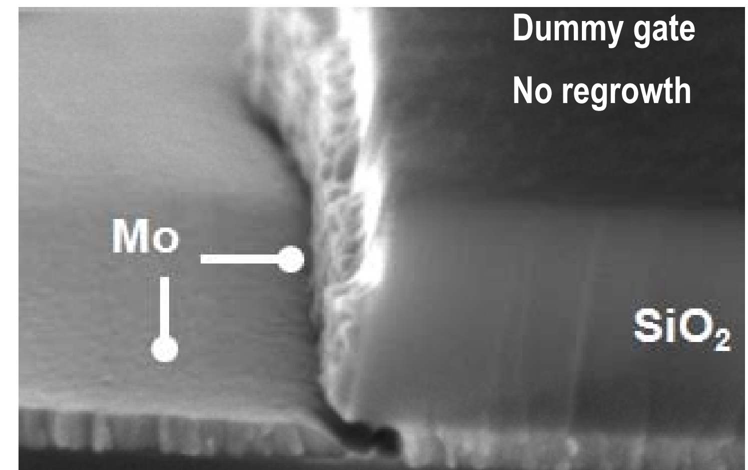
Self-Aligned Source/Drain regrowth



Self-Aligned Contacts: Height Selective Etching*

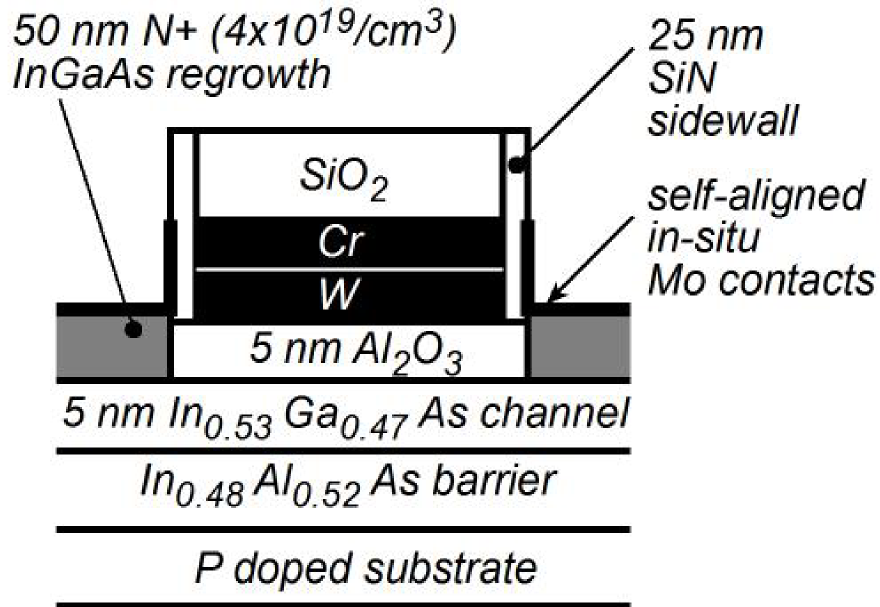


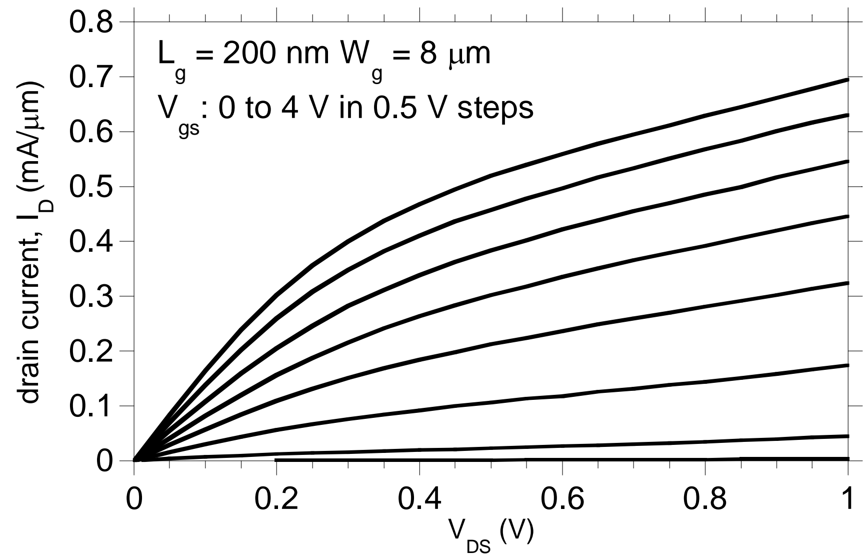


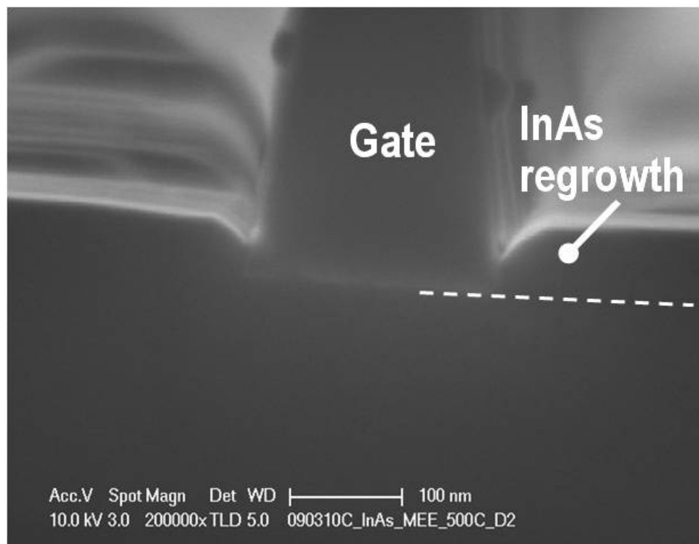


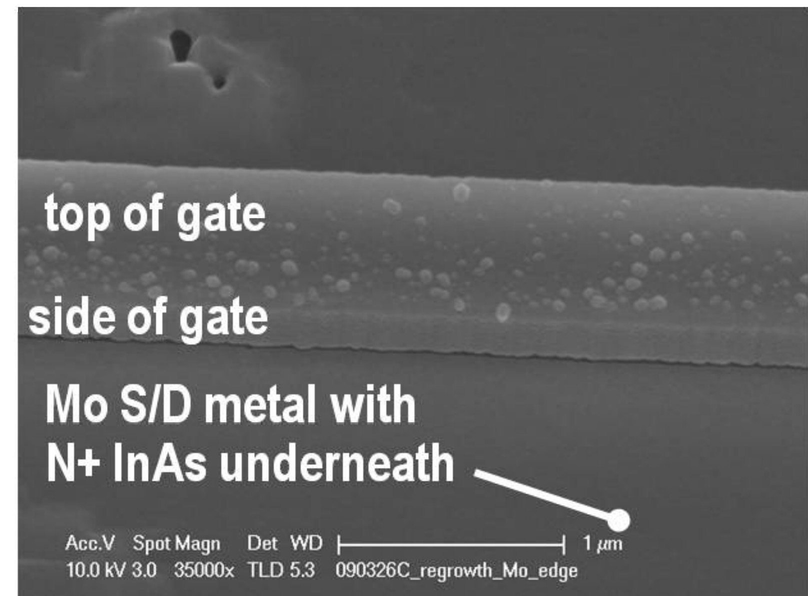
* Burek *et al*, J. Cryst. Growth 2009

Fully Self-Aligned III-V MOSFET Process









Why Is the Device Drive Current Low ? → D_{it}

Devices used Stanford / McIntyre ALD Al_2O_3 gate dielectric

best Stanford results: H passivation for low D_{it}.

FET results: H gets driven away in process

need, but do not yet have, post-process H anneal

→ high Dit on present FETs, c.a. 10^{13} / cm^2/eV.

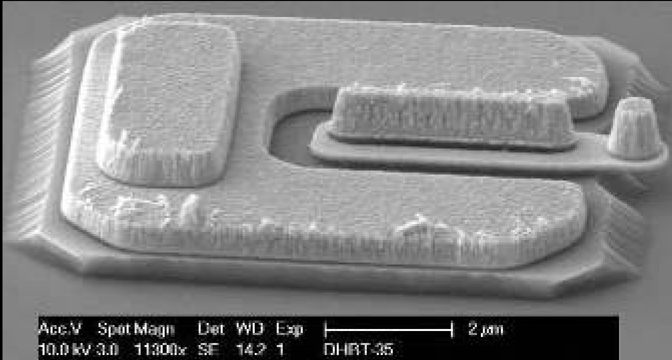
High D_{it} → Carrier depletion under sidewalls

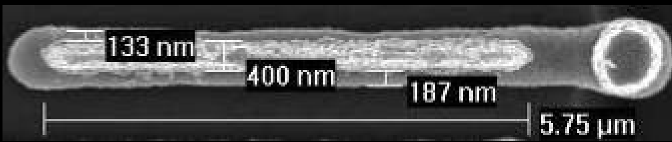
greatly increased access resistance.

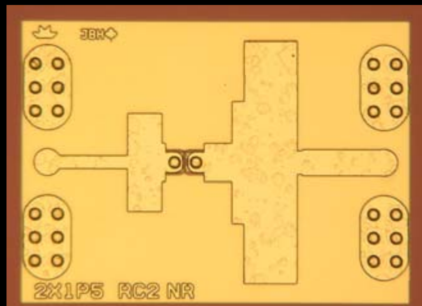
High D_{it}→ inefficient charge modulation→ low g_m.

128 nm / 64 nm / 32 nm HBT Fabrication

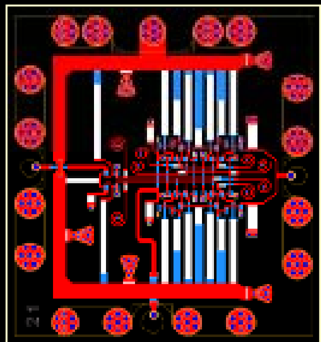
256 nm Generation InP DHBT







324 GHz Amplifier

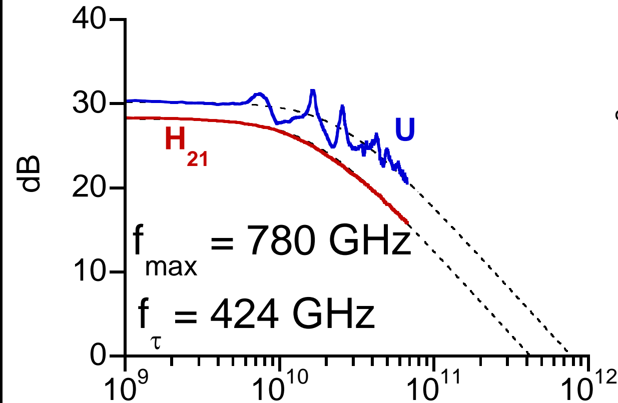


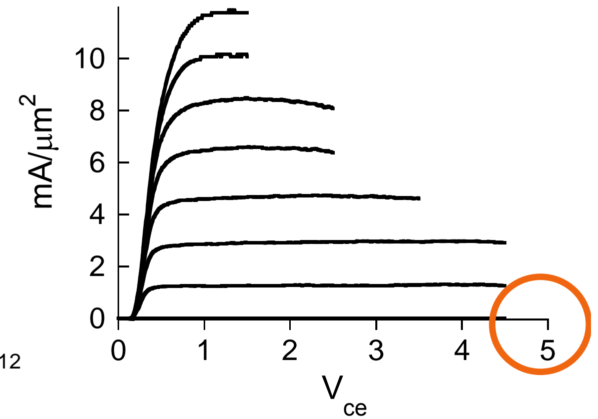
200 GHz master-slave latch design

Z. Griffith, E. Lind

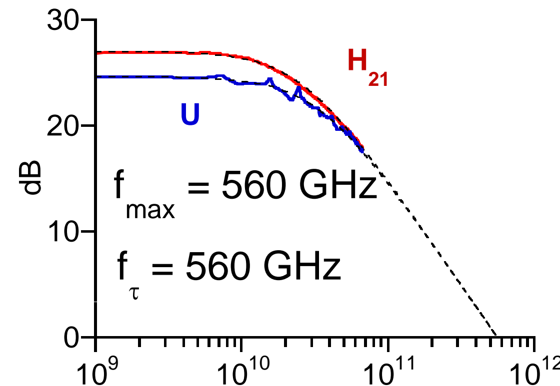
J. Hacker, M. Jones

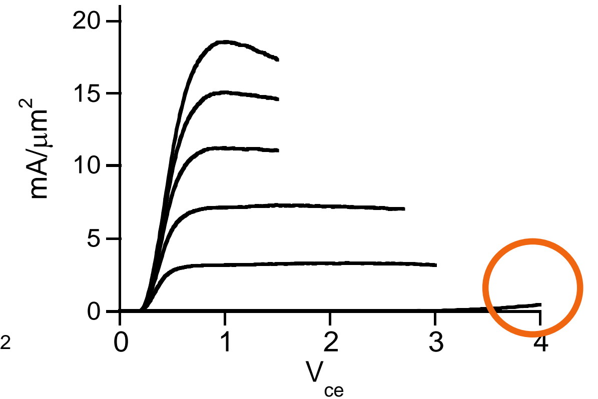
150 nm thick collector



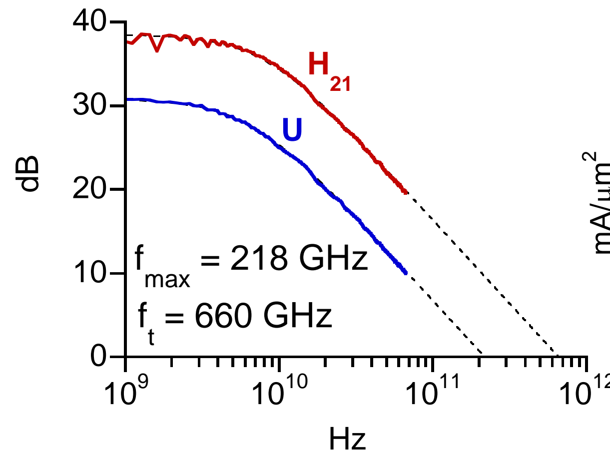


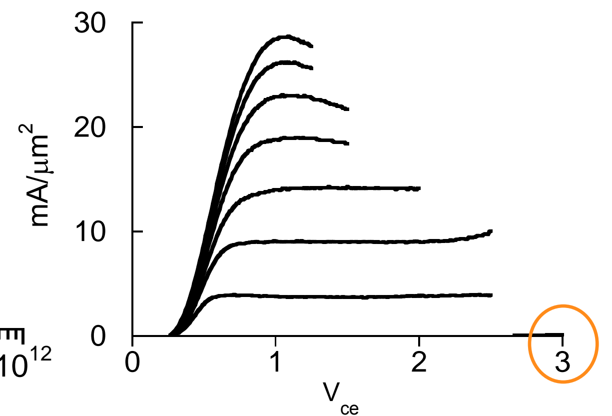
70 nm thick collector





60 nm thick collector

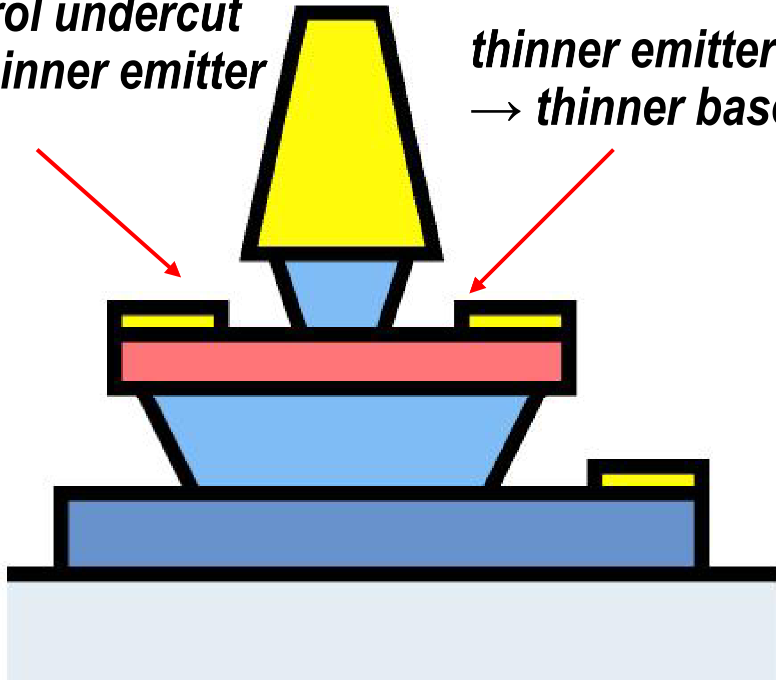




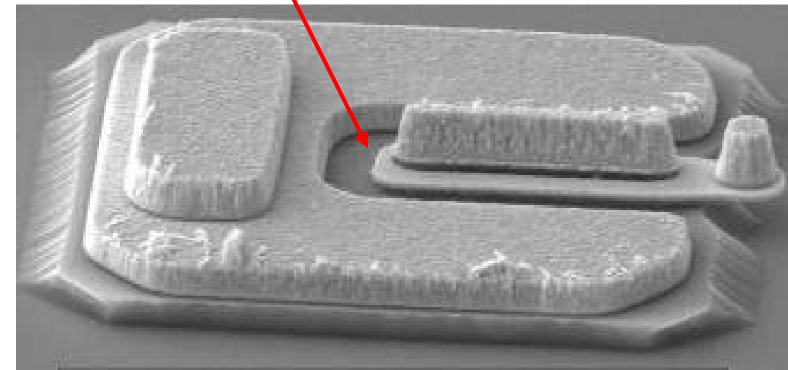
Process Must Change Greatly for 128 / 64 / 32 nm Nodes

control undercut
→ thinner emitter

thinner emitter
→ thinner base metal

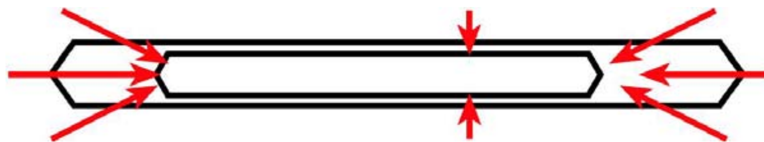


thinner base metal
→ excess base metal resistance

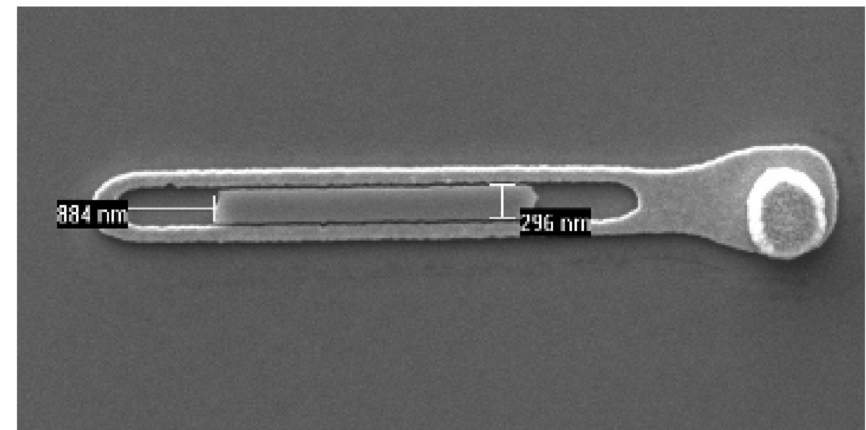


Undercutting of emitter ends...
...and loss of emitter adhesion.

{101}A planes: fast



{111}A planes: slow



128 / 64 nm HBT Process: Where We Are Going

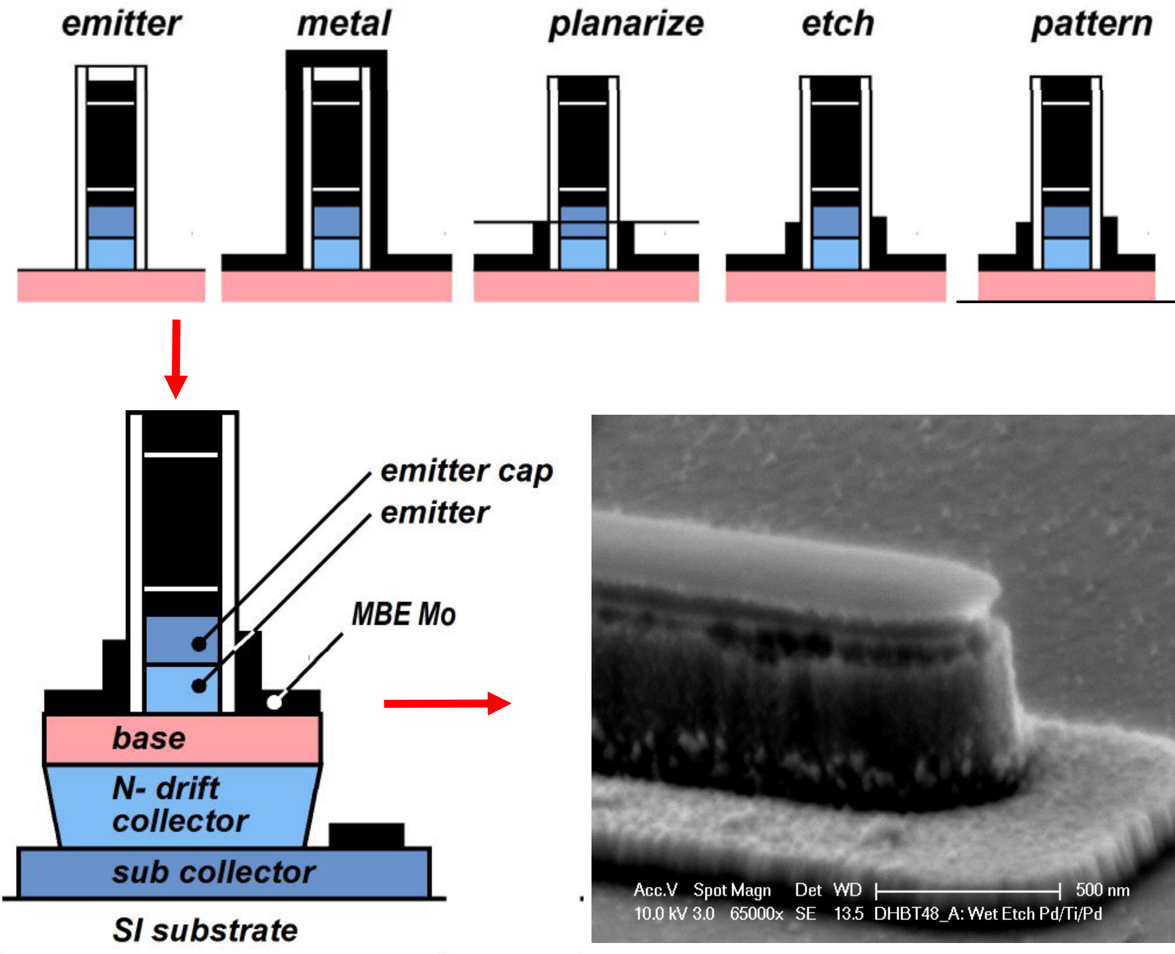
Key Features:

contact metals:
- no liftoff
- sputter deposition
- dry etched

ohmic contacts
- base & emitter
- refractory:
- thermally stable

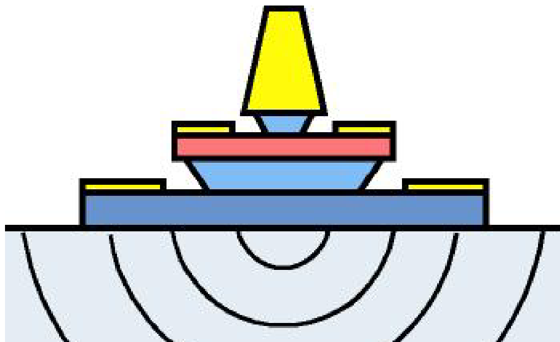
semiconductor junctions
- dry etched
- self-aligned

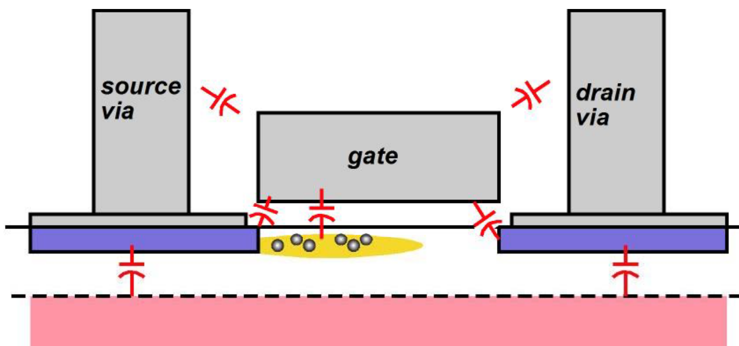
target ~2000 GHz device



Conclusion

Fabrication Processes for nm/THz III-V Transistors





10-30 nm junctions ...

~1 Ω–μm^2 contact resistivities

~100 mA/μm^2 current densities

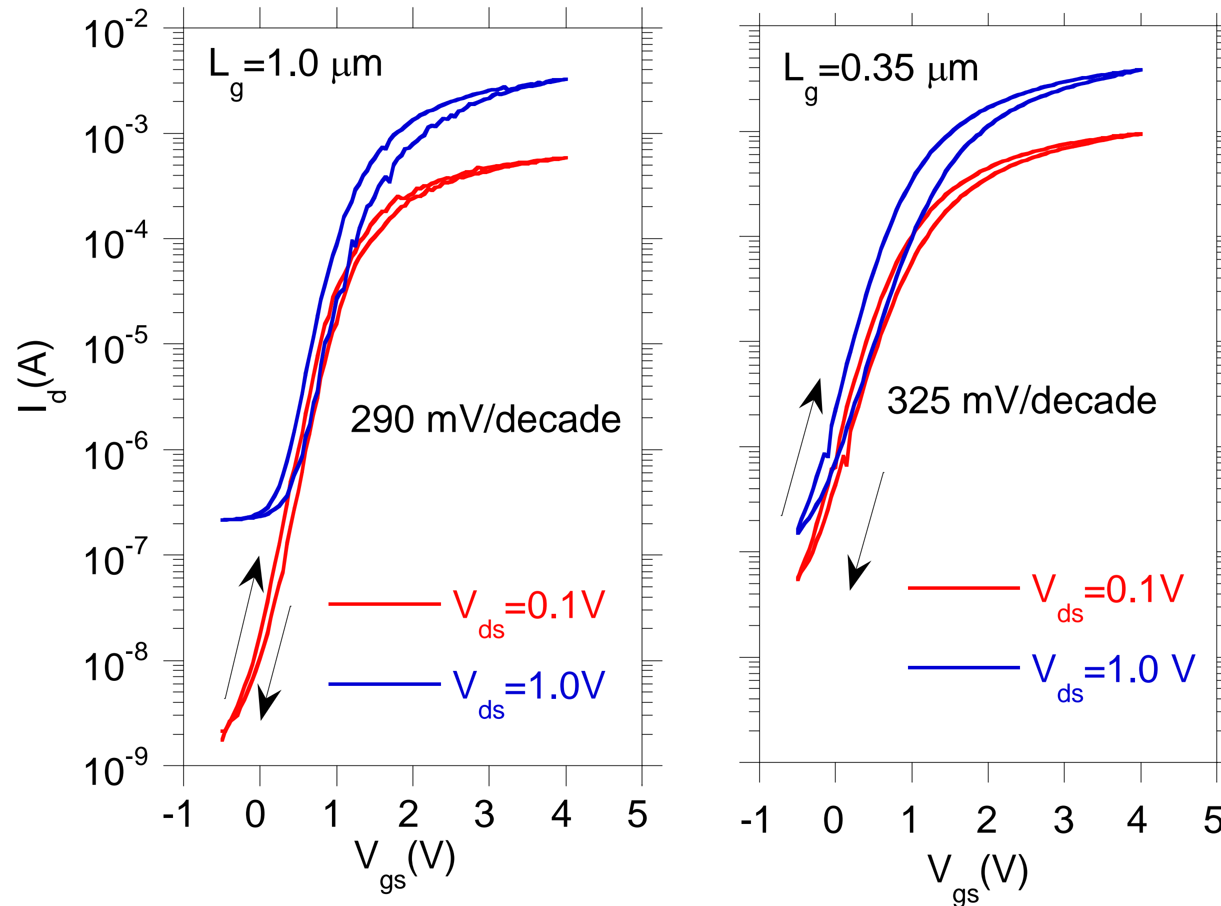
refractory contacts
- *sputter-deposited*
- *dry-etched*

self-alignment:
- *dielectric sidewall spacers*
- *height-selective etching*

dry-etched junctions, minimal wet-etching

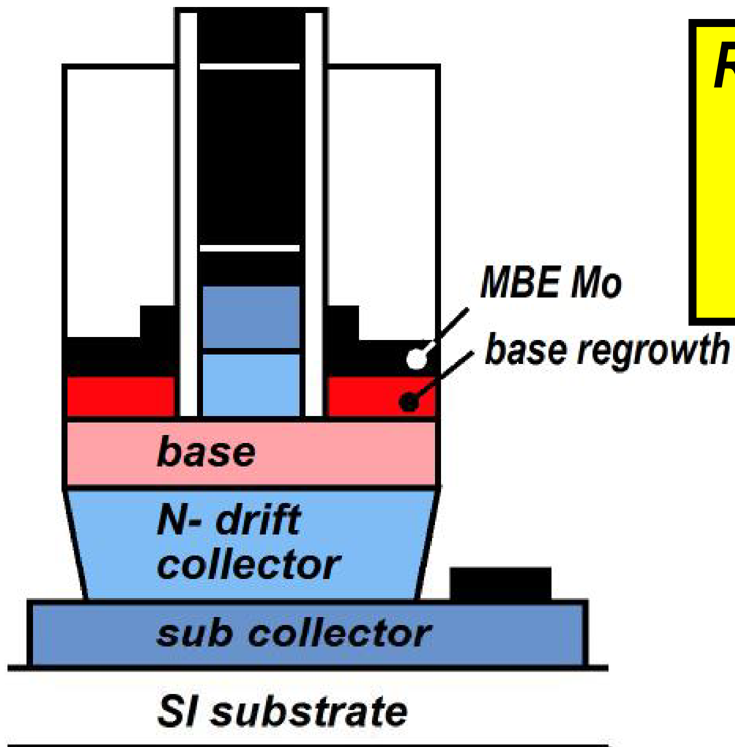
(end)

Subthreshold characteristics



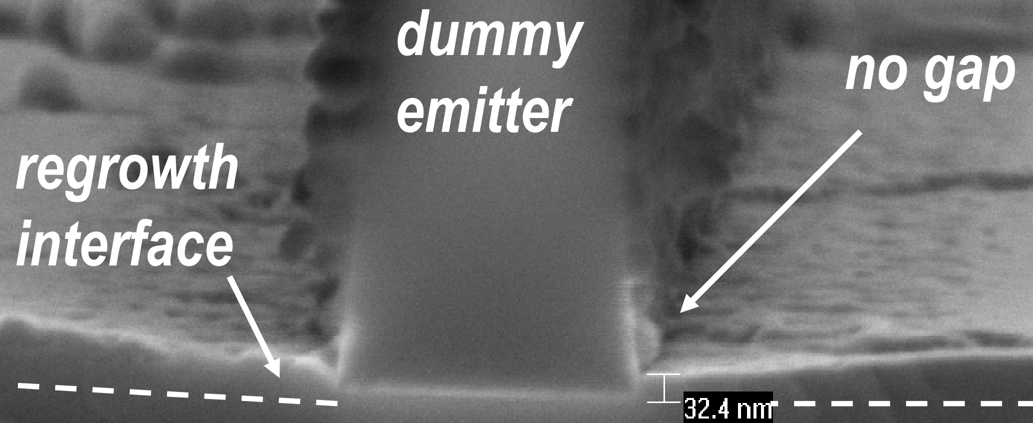
- $I_{on}/I_{off} \sim 10^4:1$

Why do we need base regrowth?

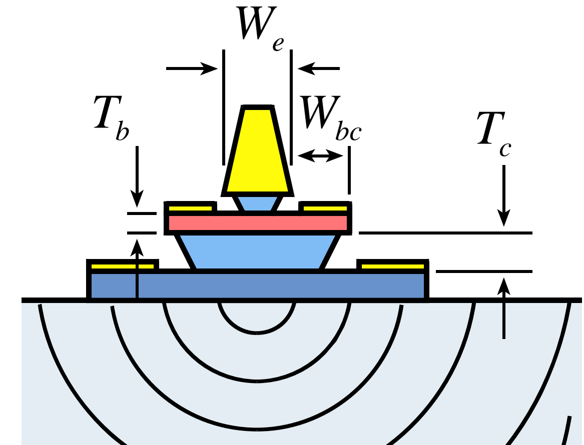


Regrowth for
- *less resistive base contacts*
- *contact moved away from c/b junction*
- *better reliability with thin base layers*

$p = 5x10^{19}$ cm^{-3} ,
$\mu = 15$ cm^2/Vs



Bipolar Transistor Scaling Laws



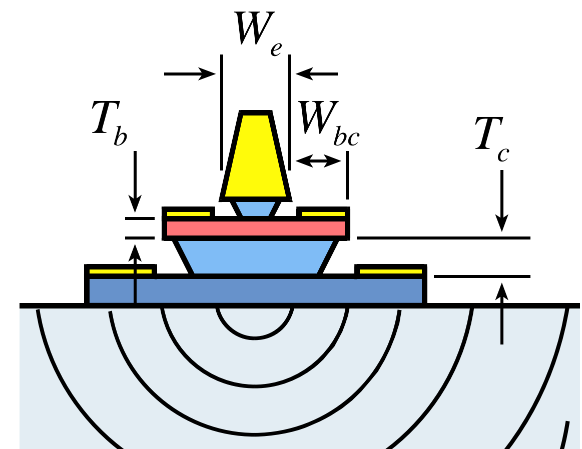
Changes required to double transistor bandwidth: (emitter length L_E)

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.

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/μm^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



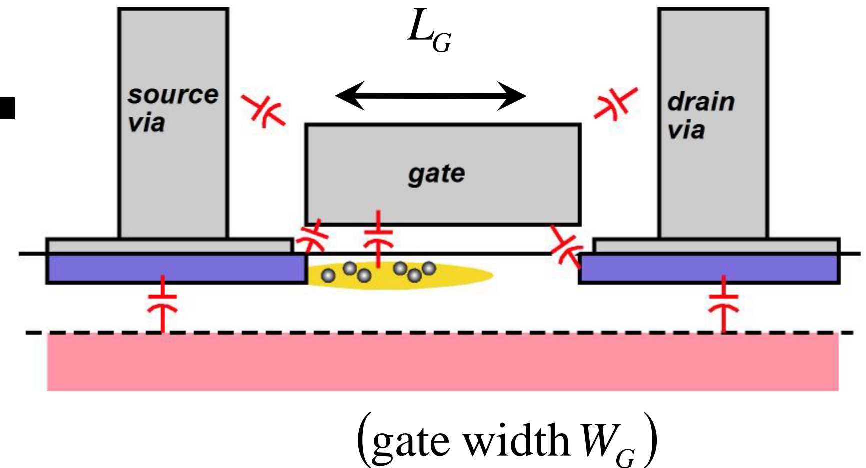
THz / nm Transistors: it's all about the interfaces

Metal-semiconductor interfaces (Ohmic contacts):
<u>very low resistivity</u>

Dielectric-semiconductor interfaces (Gate dielectrics):
<u>very high capacitance density</u>

Transistor & IC thermal resistivity.

FET Scaling Laws

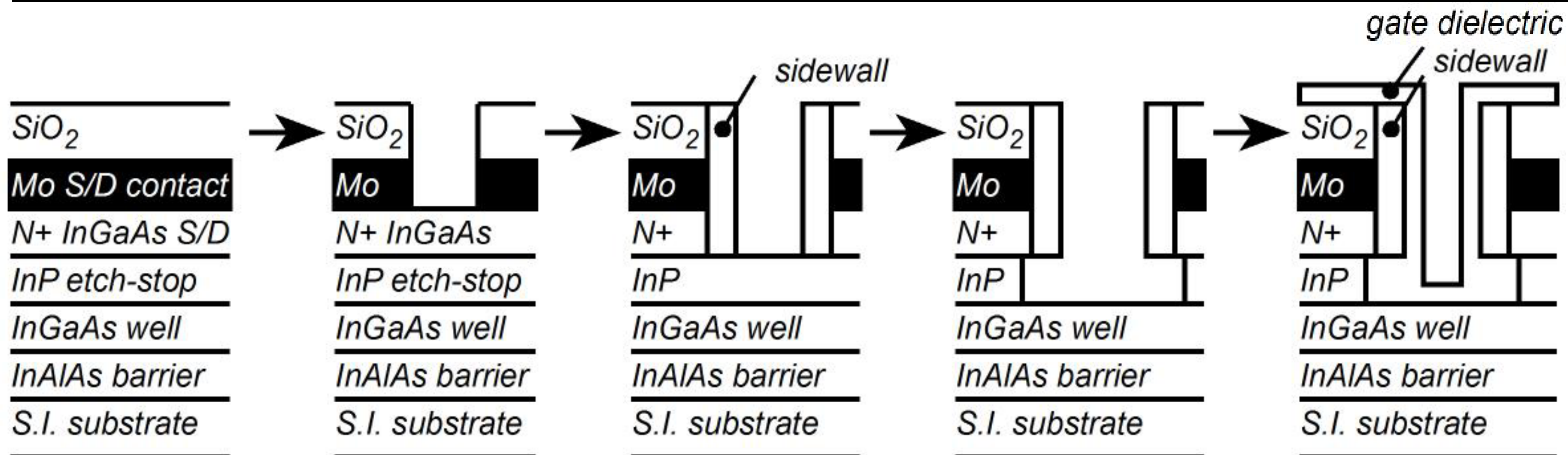


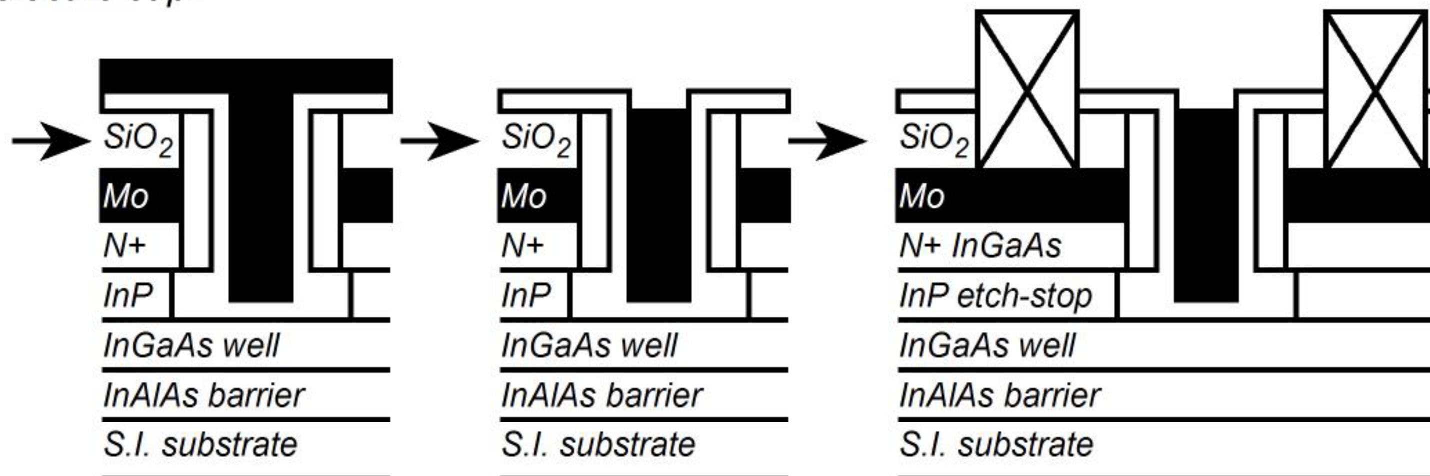
Changes required to double transistor bandwidth:

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

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

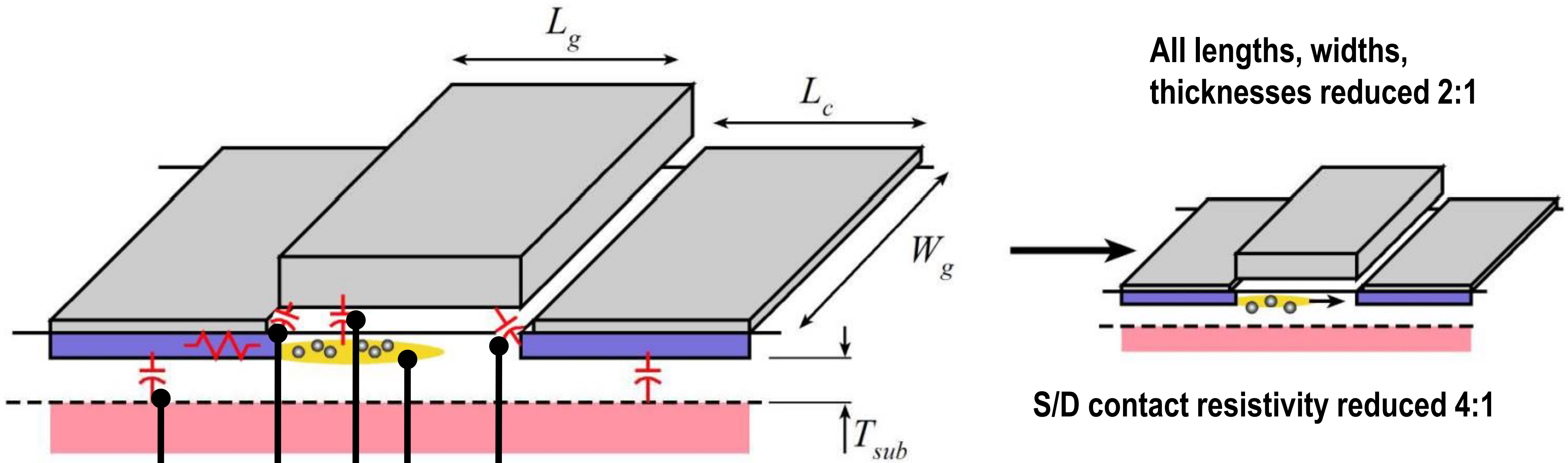
Self-Aligned VLSI Gate-Last Process





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



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

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

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

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

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

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 ✗

$(\text{also}: G_{ds} / W_g \sim \varepsilon v / L_g \rightarrow g_m / G_{ds} \sim L_g / T_{ox})$