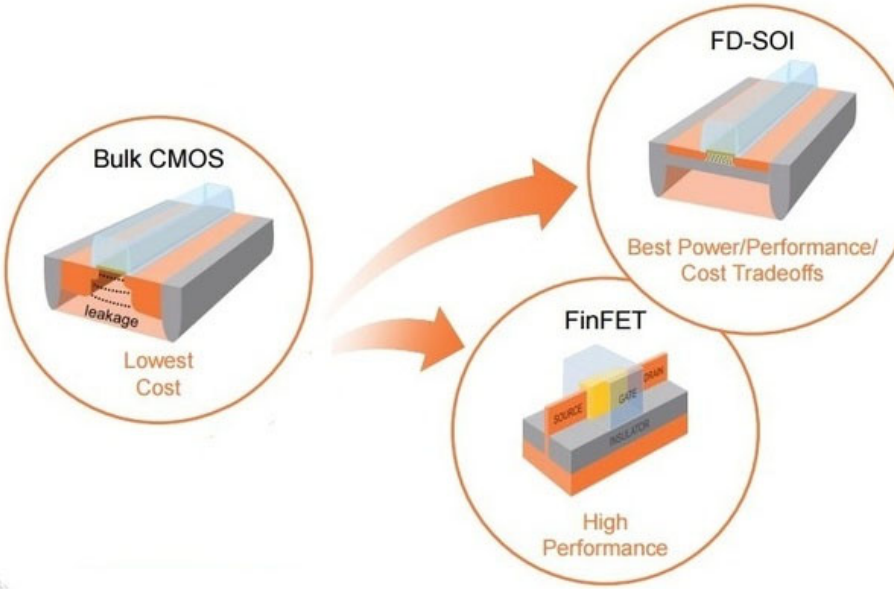
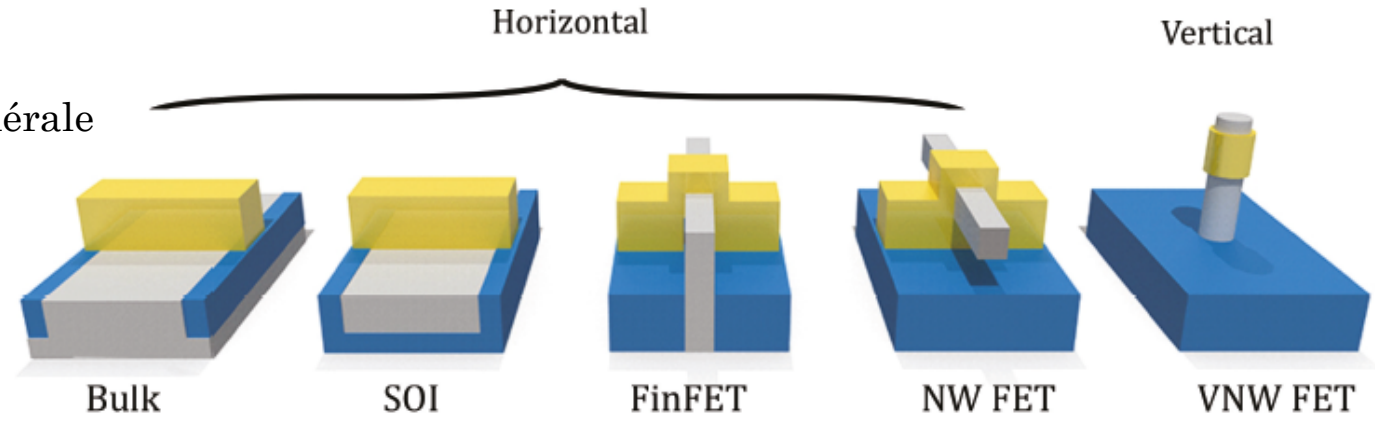
Nanoelectronics

Lecture #3:
End of scaling and
Non-classical CMOS



Adrian M. Ionescu
Ecole Polytechnique Fédérale
de Lausanne



Summary

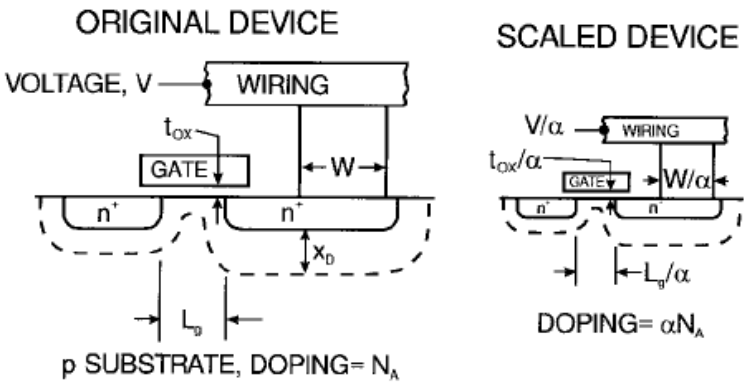
- **Post-Dennardian nanometer scaling challenges:**
  - Short channel effects and DIBL
  - Leakage
  - Ballistic MOSFET
- **Non-classical CMOS device architectures**
  - Fin FET
  - Silicon-On-Insulator (SOI) substrates
  - Ultra Thin Body (UTB) Fully Depleted (FD) SOI MOSFET
  - 1D FETs: GAA FETs and vertical FETs
  - Junctionless MOSFETs

CMOS: Dennard happy scaling (1)

Guide for MOSFET scale-down (happy scaling): *Denard et al. 1974*

Idea: change the device parameters and operating voltage according to:

	Before scaling:	After scaling:
Channel length	$= L_g$	L_g / k
Oxide thickness	$= t_{ox}$	t_{ox} / k
Channel doping	$= N_A$	$N_A \times k$
Operating voltage	$= V_{DD}$	V_{DD} / k



1. Maintain the **SAME constant field in the device** (limitation: ~5MV/cm in the oxide)
2. Consequence: **drain current UNCHANGED** before and after scaling !

CMOS: Dennard happy scaling (2)

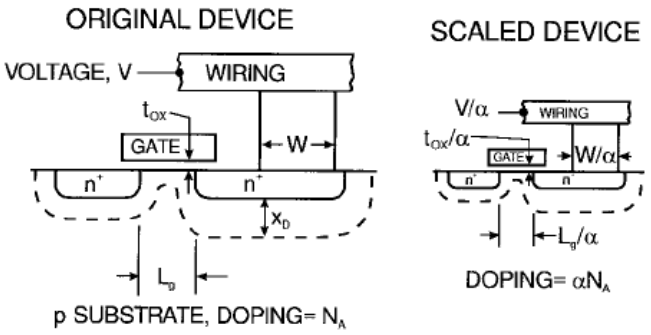


Table 1
Technology Scaling Rules for Three Cases (α Is the Dimensional Scaling Parameter, ϵ Is the Electric Field Scaling Parameter, and α_d and α_w Are Separate Dimensional Scaling Parameters for the Selective Scaling Case; α_d Is Applied to the Device Vertical Dimensions and Gate Length While α_w Applies to the Device Width and the Wiring)

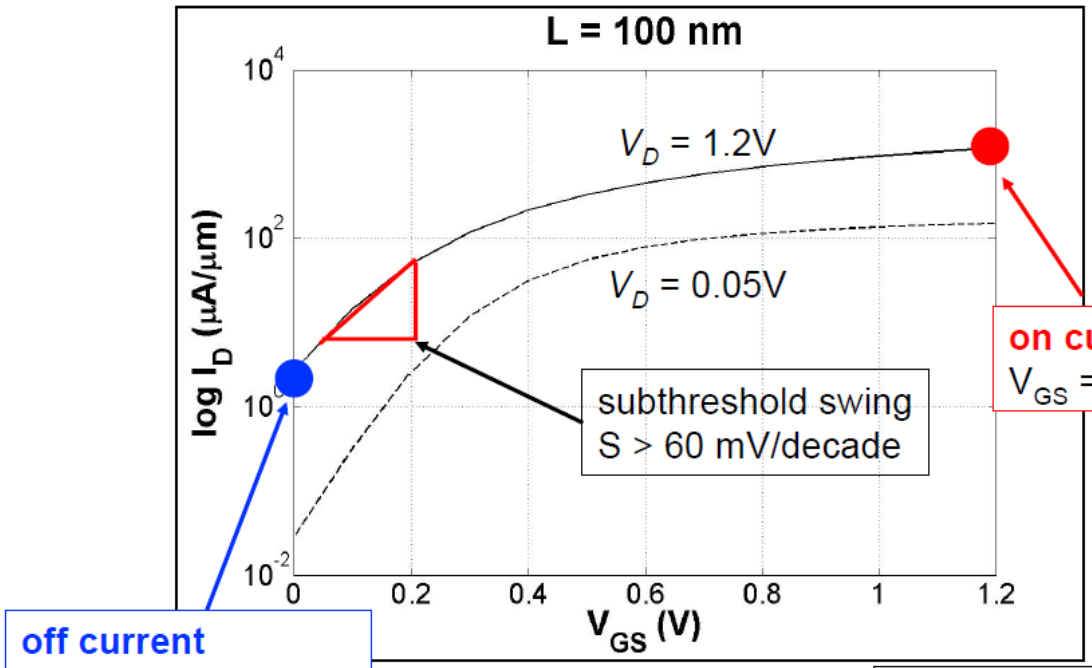
Physical parameter	Constant-Electric Field Scaling Factor	Generalized Scaling Factor	Generalized Selective Scaling Factor
Channel length, Insulator thickness	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Wiring width, channel width	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Electric field in device	1	ϵ	ϵ
Voltage	$1/\alpha$	ϵ/α	ϵ/α_d
Doping	α	$\epsilon\alpha$	$\epsilon\alpha_d$
Area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha_w^2$
Capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Gate delay	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Power dissipation	$1/\alpha^2$	ϵ^2/α^2	$\epsilon^2/\alpha_w\alpha_d$
Power density	1	ϵ^2	$\epsilon^2\alpha_w/\alpha_d$

Post-Dennardian nanometer scaling challenges

Important performance metrics and physical effects for nanometer scale MOSFET:

- **Short channel effects**
- **Drain Induced Barrier Lowering (DIBL)**
- **(Inverse) subthreshold slope**
- **Leakage Power**
- **Low Field Mobility**
- **Carrier Velocity Saturation**
- **Ballistic Transport**

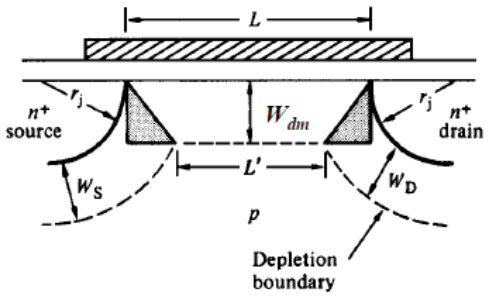
Reminder: MOSFET static performance: needs semi-log plot



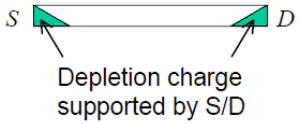
S is the number of millivolts required to increase V_G to produce a factor of 10 increase in I_D (in the sub-threshold region).

Short channel effect (1)

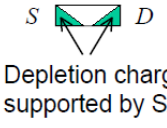
- The gate supports the depletion charge in the trapezoidal region. This is smaller than the rectangular depletion region underneath the gate, by the factor $1-\frac{L+L'}{2L}$
- This is the factor by which the depletion charge Q_{dep} is reduced from the ideal
- One can deduce from simple geometric analysis that $L' = L - 2r_j\left[\sqrt{1+\frac{2W_{dm}}{r_j}}-1\right]$



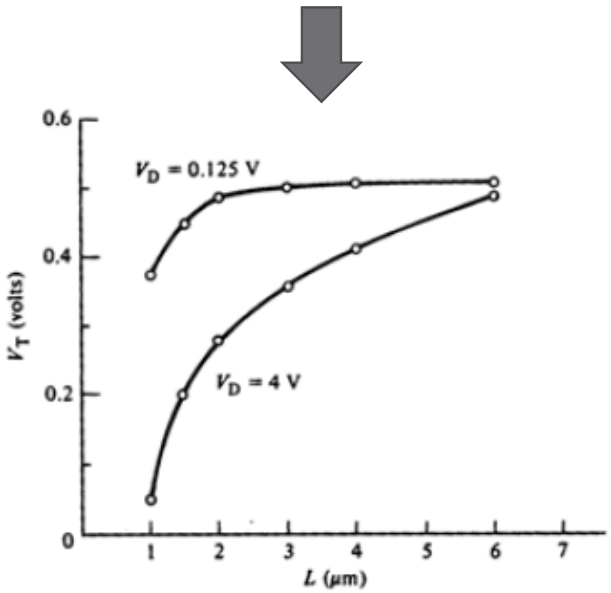
Large L:



Small L:



(Bad) consequence: V_T Roll-Off



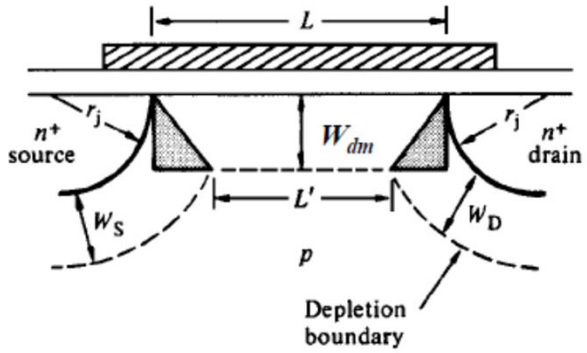
⇒ less gate charge is required to reach inversion (*i.e.* $|V_T|$ decreases)

Short channel effect (2)

$$|V_T| - |V_{T(long-channel)}| \equiv \Delta V_T = \frac{-qN_A W_{dm}}{C_{oxe}} \frac{r_j}{L}\left(\sqrt{1+\frac{2W_{dm}}{r_j}} - 1\right)$$

Minimize ΔV_T by

- reducing T_{oxe}
- reducing r_j
- increasing N_A
  (trade-offs: degraded m, μ)



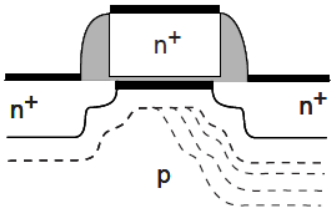
⇒ **MOSFET vertical dimensions should be scaled along with horizontal dimensions!**

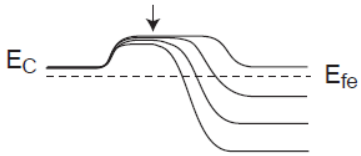
Solutions:

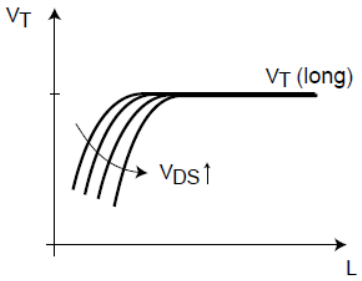
- Thinner oxide
- Shallow S/ regions
- Retrograde channel doping
- Make transistors on thin semiconducting films

Drain Induced Barrier Lowering (DIBL)

Depletion region associated with drain junction expands as VDS ↑
⇒ additional VT shift.







Quantification of DIBL:

$$DIBL = |\frac{V_T(V_{DS} = V_{DD}) - V_T(V_{DS} = 0.1\ V)}{V_{DD} - 0.1}|\ \ mV/V$$

for a certain L device.

Goal: suppress or minimize DIBL below 100mV/V

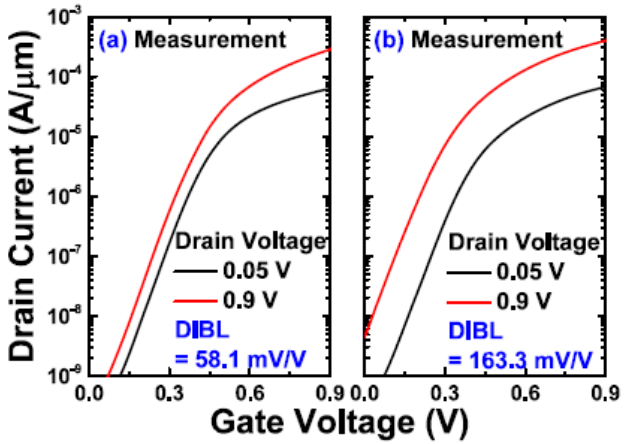
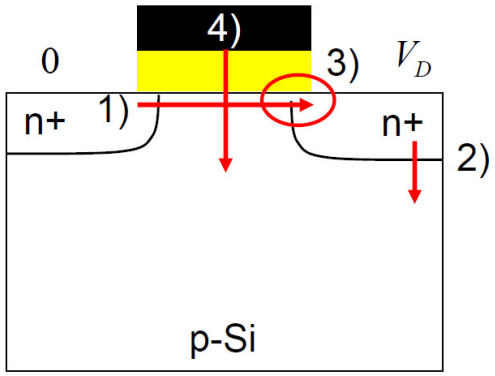


Fig. 1. (Color online) Measured *I–V* characteristics of two conventional bulk nFETs with (a) small and (b) large DIBL values.

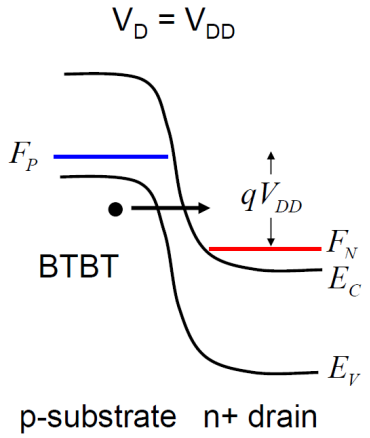
Leakage sources in nanometer scale MOSFET



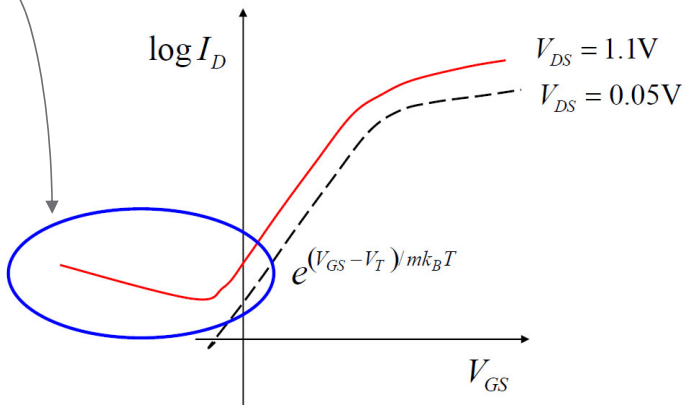
- Subthreshold current
- Junction leakage
- Gate-induced drain leakage
- Gate-leakage

Source: Mark Lundstrom.

Band-To-Band-Tunneling (BTBT)

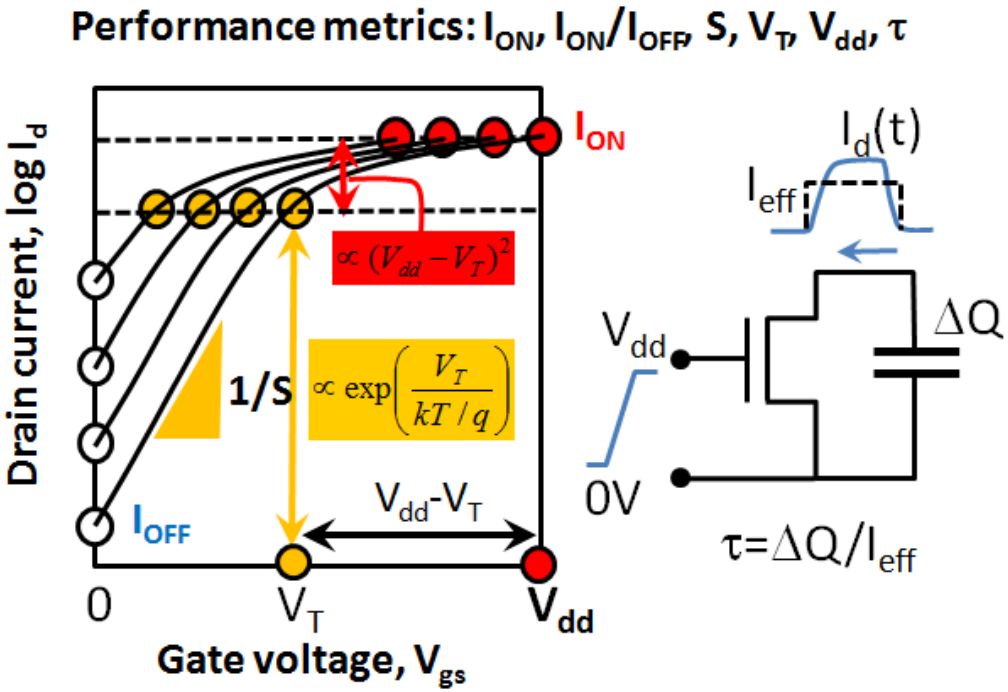


Gate Induced BTBT



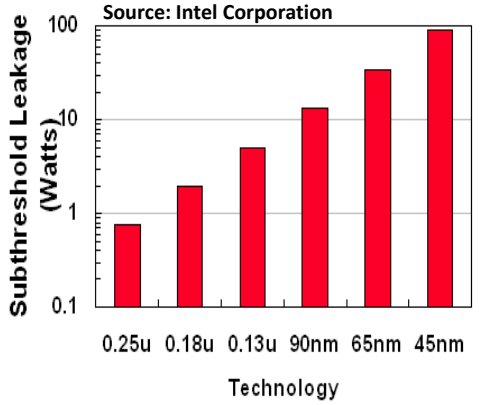
Inverse subthreshold slope, S, and leakage, Ioff

Reducing threshold voltage
by 60mV
increases the leakage current (power)
by ~10 times

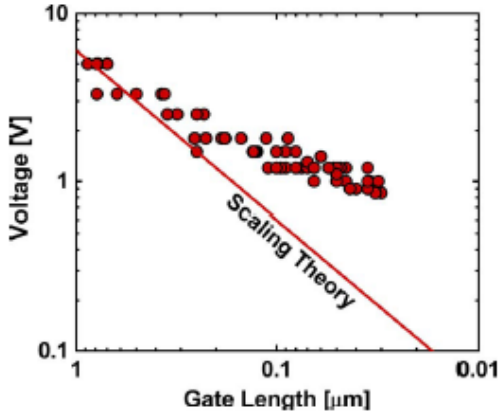


A.M. Ionescu and H. Riel, Nature 2011

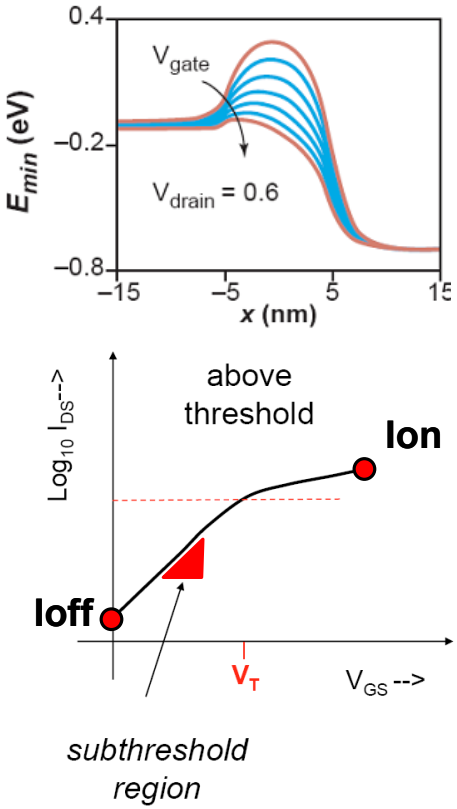
→ Leakage power dominates in advanced technology nodes



→V_T & V_{dd} scaling saturated



Inverse subthreshold slope and diffusion current



- carrier injection by lowering the barrier
- subthreshold current is a diffusion current

$$I_D = q\frac{W}{L}\left(\frac{n_i^2}{N_A}\right)\frac{(k_BT/q)^2}{E_S}\mu_{eff}e^{qV_{GS}/mk_BT}\left(1-e^{-qV_{DS}/k_BT}\right)\text{A}$$

$$m = 1 + C_D / C_{OX}$$

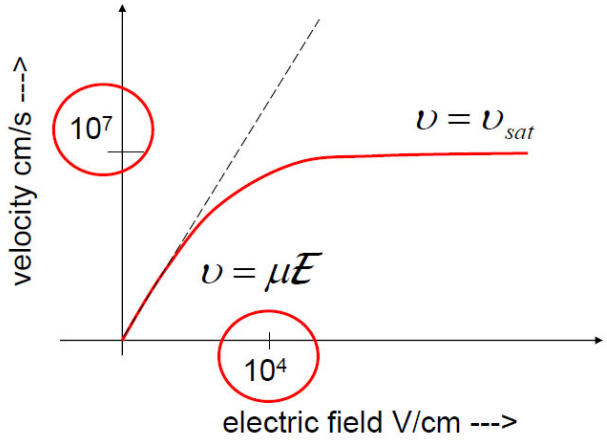
$$S = \frac{dV_g}{d(\log_{10} I_d)} = \ln 10 \frac{kT}{q}\left(1+\frac{C_{dep}}{C_{ox}}+\frac{C_{ss}}{C_{ox}}\right)$$

$$\rightarrow \frac{kT}{q}\ln 10 = 60mV/decade \quad @\,RT$$

R. Van Overstraeten,G.J. Declerck, P.A. Muls,
IEEE Transactions on Electron Devices, Volume 22, May 1975.

Carrier Velocity Saturation & Saturation Drain Voltage in nanometer gate length MOSFETs

Micrometer to submicrometer FETs



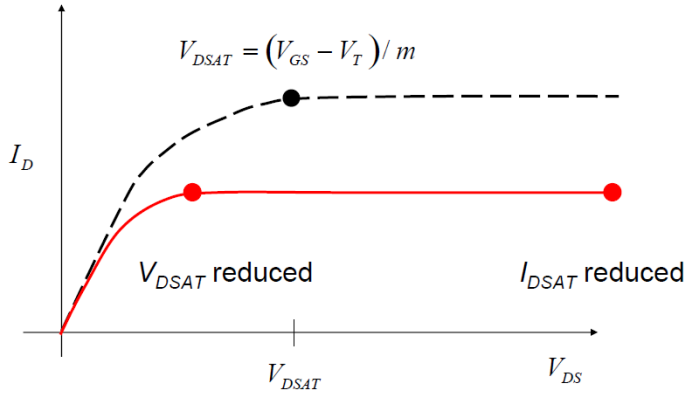
$$I_D = W\,Q_I(y)\,\upsilon_y(y)$$

$$\upsilon_y(y) = \mu_{eff} E_y(y)\,?$$

$$E_y \sim \frac{V_{DD}}{L} << 10^4 \text{ V/cm}$$

OK for *L* >> 1 micrometer

Nanometer scaled FETs



$$V_{DSAT} \rightarrow \sqrt{2\upsilon_{sat} L (V_{GS} - V_T)/m\mu_{eff}}$$

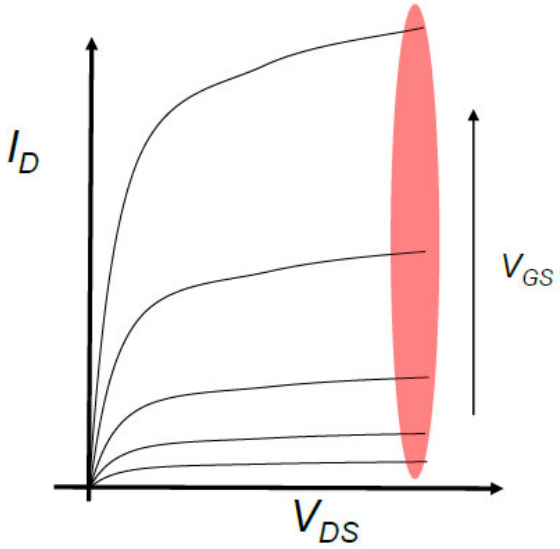
$$I_{DSAT} = W\,C_{ox}\,\upsilon_{sat}\,(V_{GS} - V_T)$$

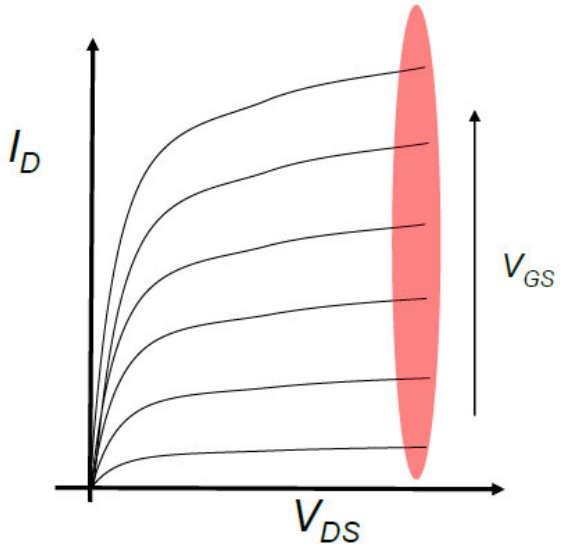
"complete velocity saturation"

current independent of *L*

for L < 100nm

Signature of carrier velocity saturation: Dependence of I_{Dsat} on gate voltage





$$I_D(V_{DS} = V_{DD}) \sim (V_{GS} - V_T)^{\alpha}$$

$$1 < \alpha < 2$$

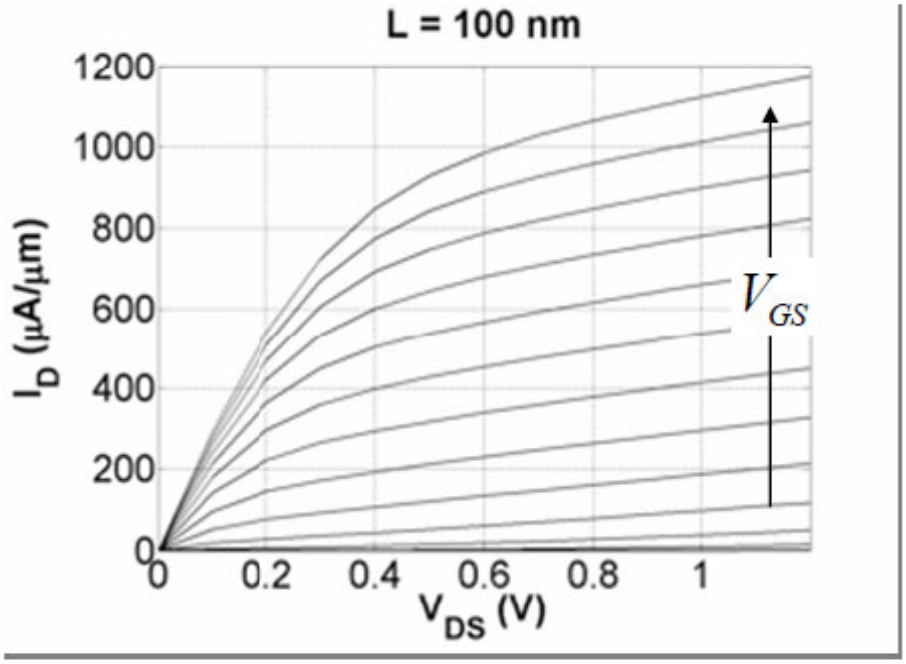
complete velocity saturation

long channel

$$I_D = \frac{W}{2L}\mu_{eff} C_{ox} \frac{(V_{GS} - V_T)^2}{m}$$

$$I_D = W \upsilon_{sat} C_{ox} (V_{GS} - V_T)$$

Experimental data @ 100nm



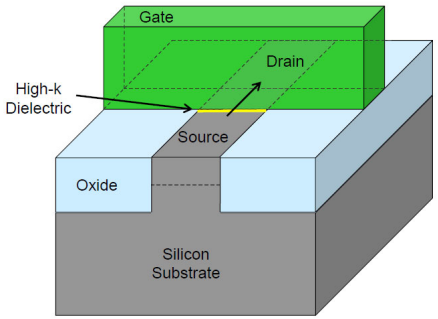
$$I_D \propto W\left(V_{GS} - V_T\right)^{\alpha}$$

$$\alpha \approx 1$$

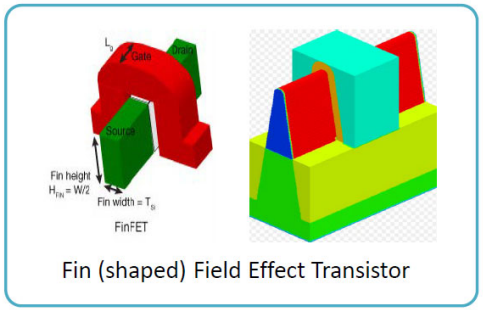
(Courtesy, Shuji Ikeda, ATDF, Dec. 2007)

Limitations of bulk planar transistors

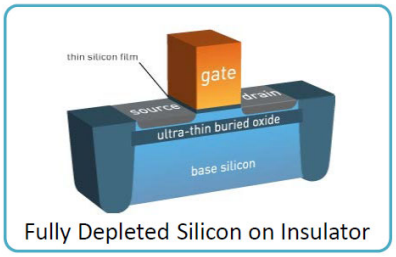
- Channel area underneath the gate is too deep and far away from the gate to be well-controlled
- Consequences:
  - **Short channel effects** and **DIBL**
  - **Higher leakage power** (static/stand-by power)
  - **Deteriorated subthreshold slope**
  - Gate is never truly turned off
- Solution: make the **channel thinner** so that it is well controlled by the gate



Bulk Si MOSFET (Dennard scaling):
Works down to ~50 – 100 nm

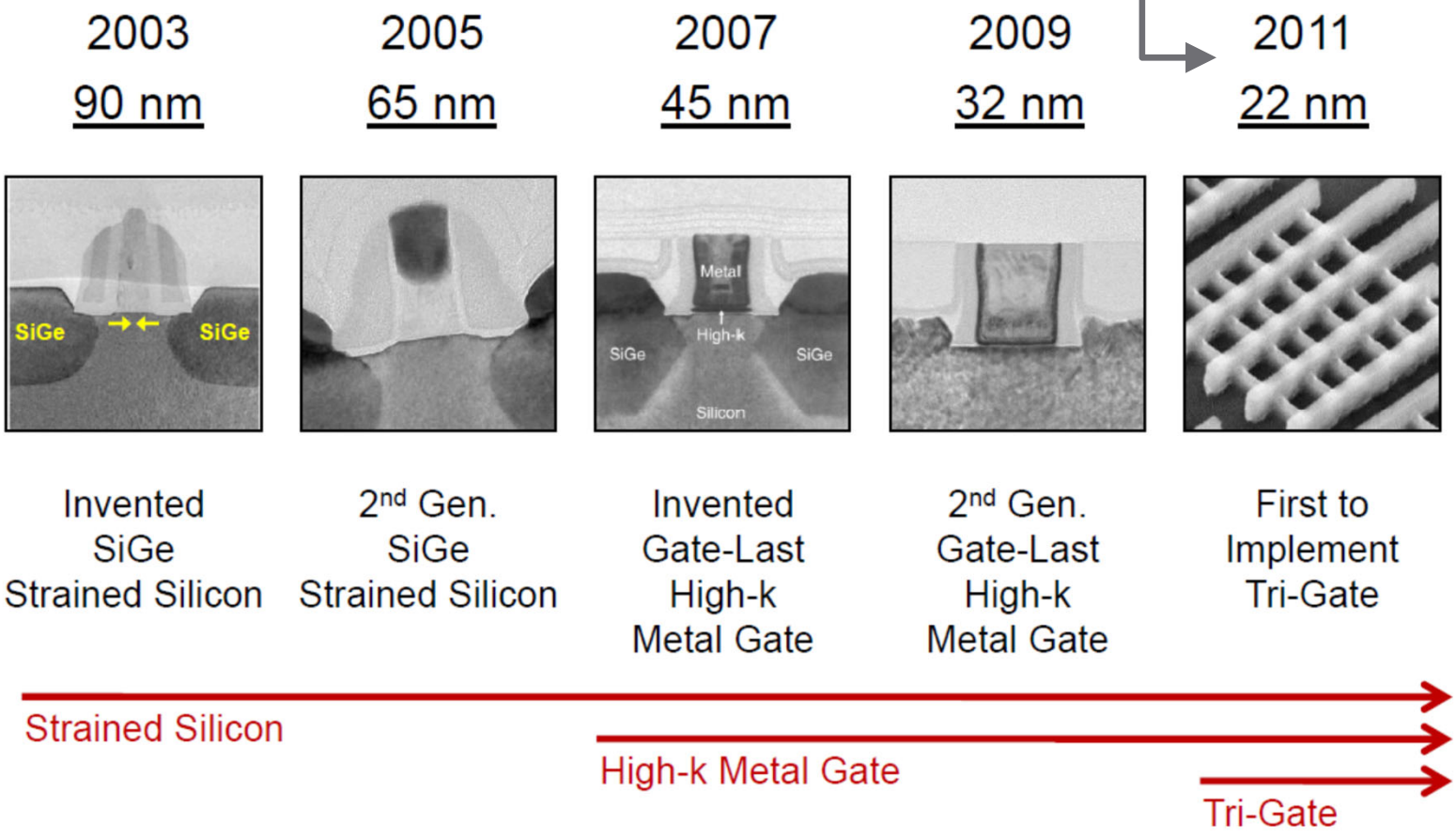


Fin (shaped) Field Effect Transistor



Fully Depleted Silicon on Insulator

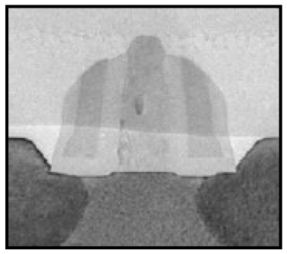
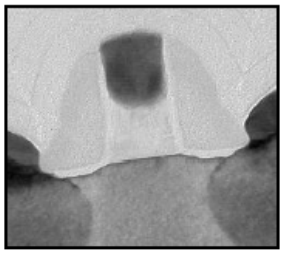
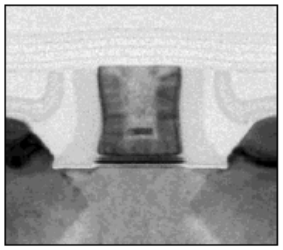
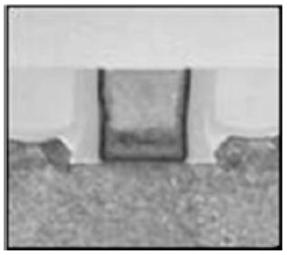
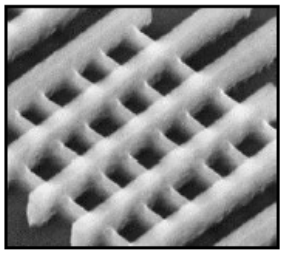
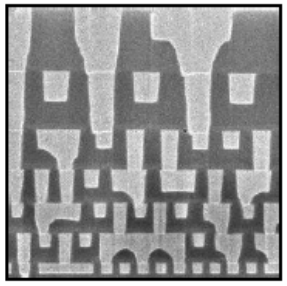
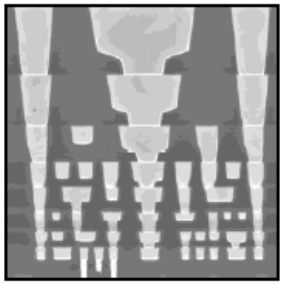
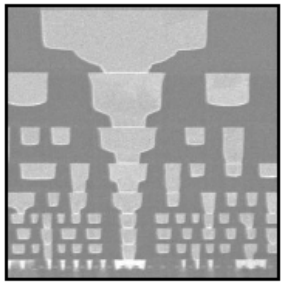
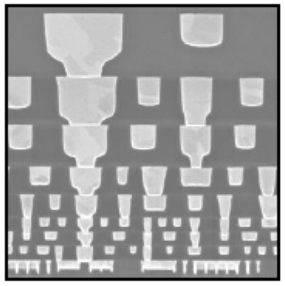
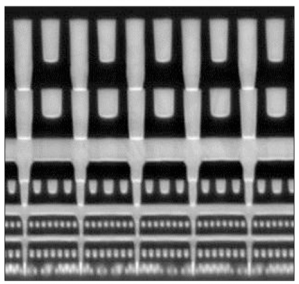
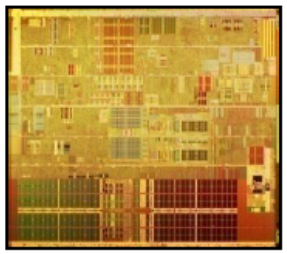
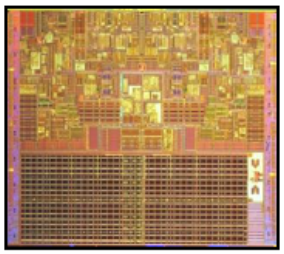
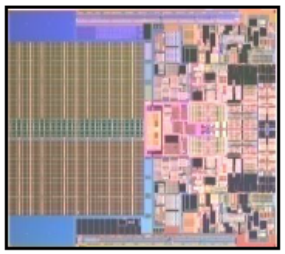
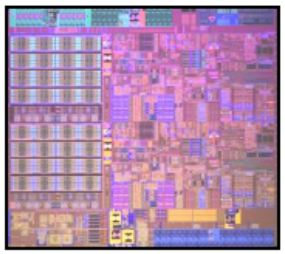
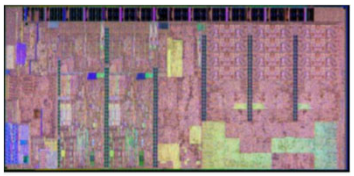
From bulk planar transistor to multigate (Multiple Gate = MuG) FET



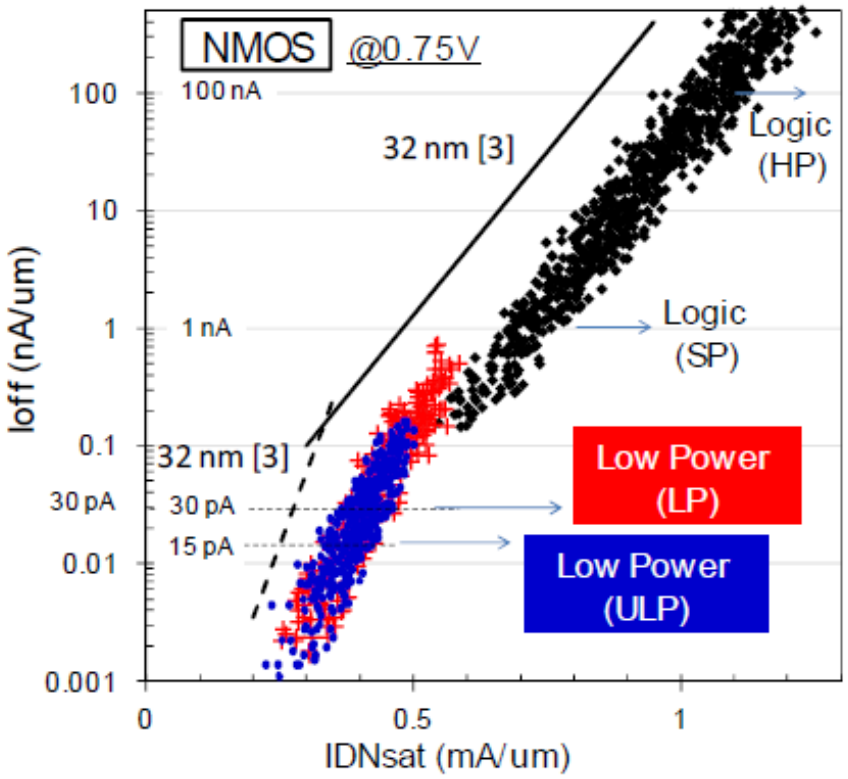
Intel's technology

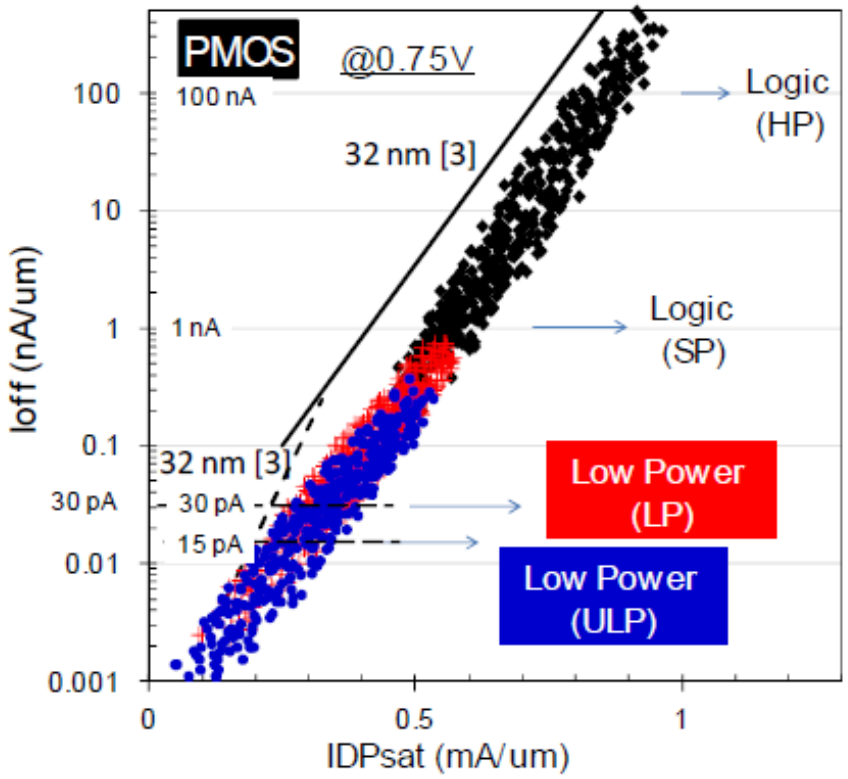
Trigate/FinFET

90 nm	65 nm	45 nm	32 nm	22 nm
2003	2005	2007	2009	2011

State of the art technology performance

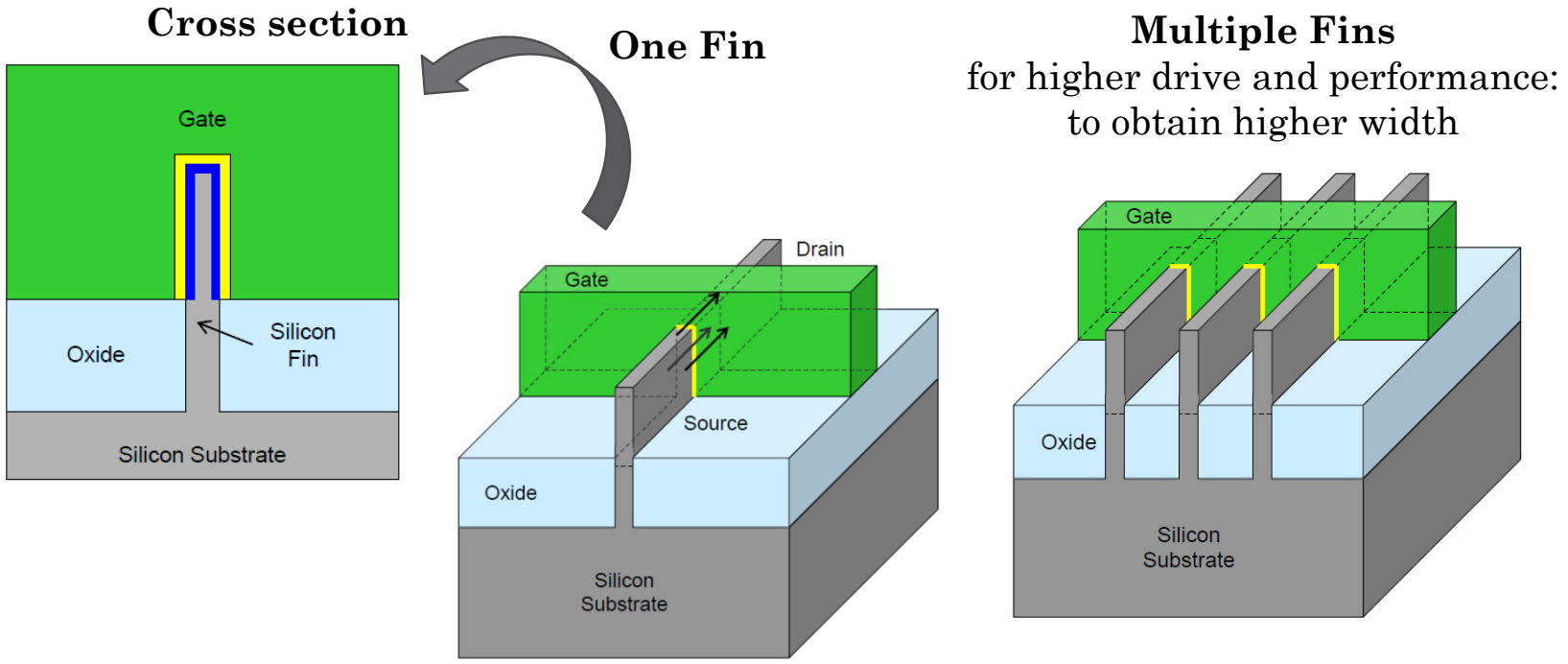




Intel @ IEDM 2012

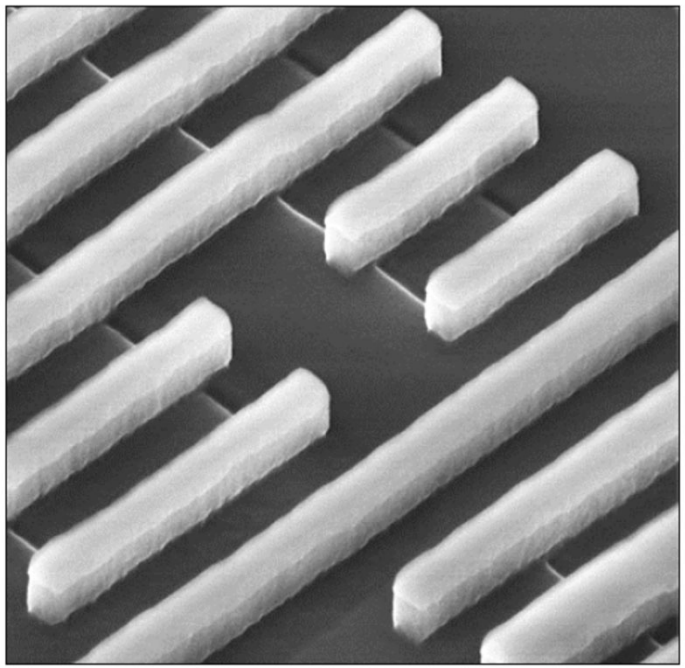
Fin FETs (Intel)

- 3D tri-gate transistors form conducting channels on the three sides of a vertical fin structure, being a fully depleted transistor with excellent electrostatic control,



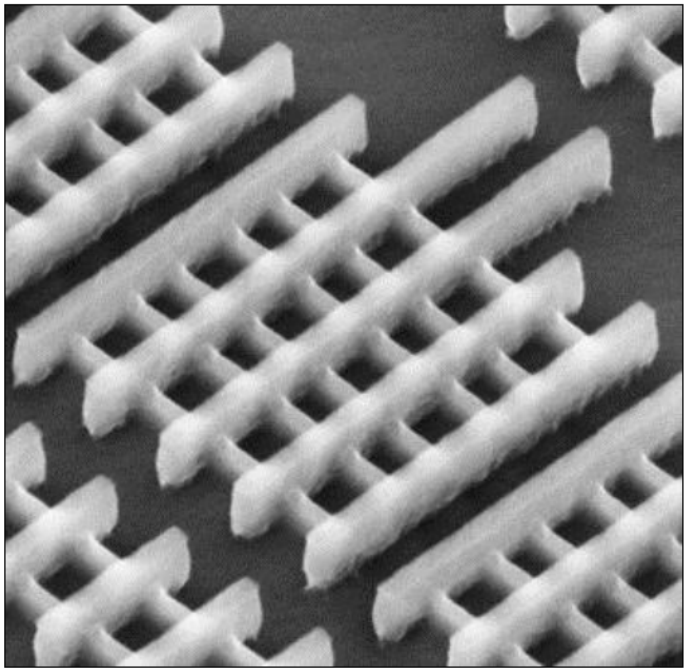
Scanning Electron Microscopy (SEM) Pictures of FinFETs

32 nm Planar Transistors



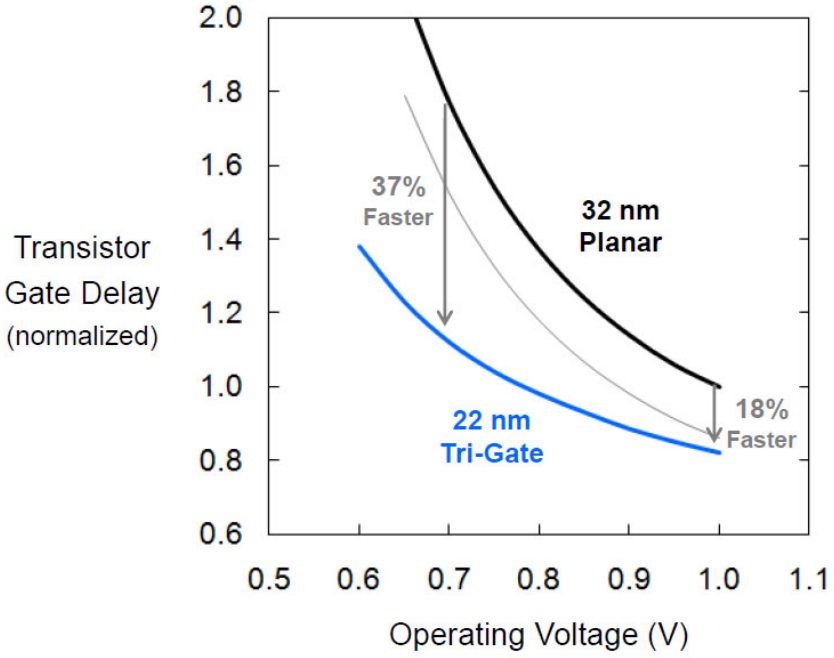
Bulk

22 nm Tri-Gate Transistors

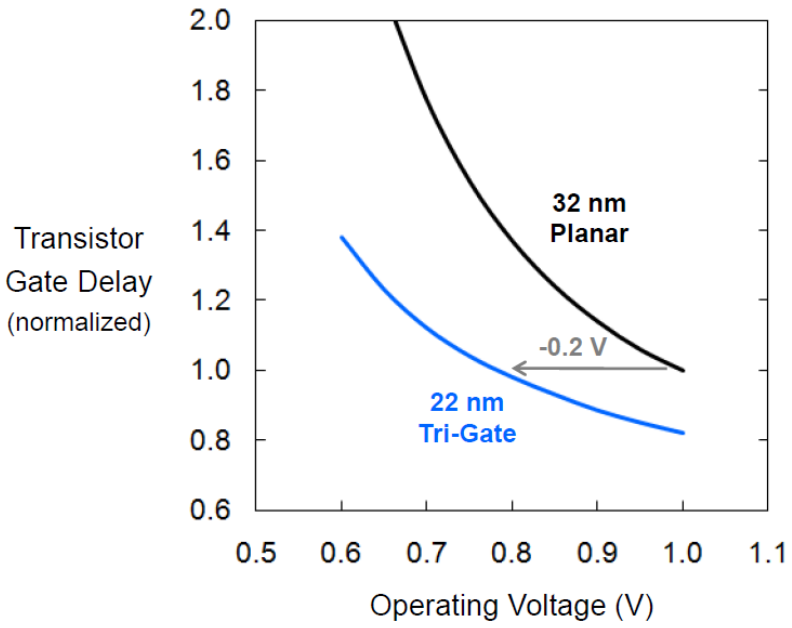


Tri-Gate/FinFET

Tri-gate/FinFET performance benefits



Improved performance at high gate voltage.

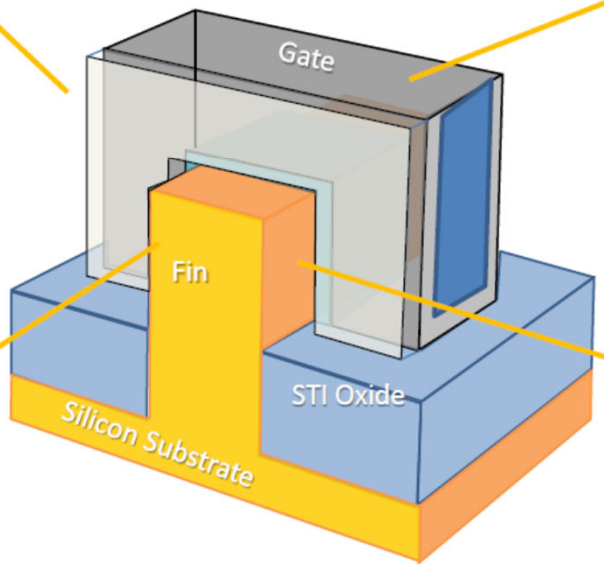


22 nm Tri-Gate transistors can operate at lower voltage with good performance, reducing active power by >50%

Fin FETs process challenges

Spacer
- Complete spacer removal from fin area

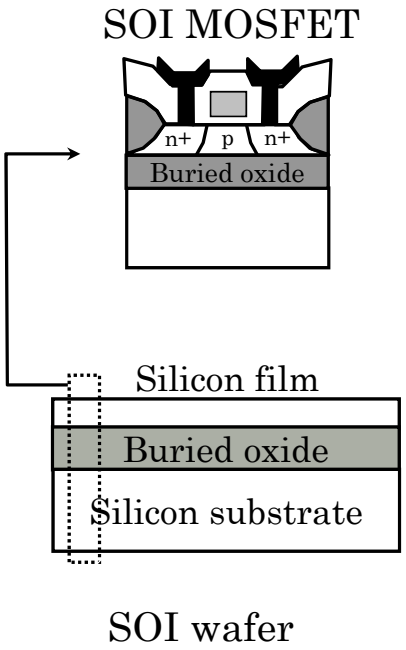
Gate Stack (high-k & metal gate)
- Material selectivity
- Material deposition thickness uniformity on vertical walls
- Metal gate composition uniformity/stability



Fin Formation:
- Precision etch
- Structural integrity (collapse, erosion, thermal shock)
- Precise Recess to control fin height
- Channel materials to increase mobility

Fin Junctions:
- Conformal doping on sidewalls

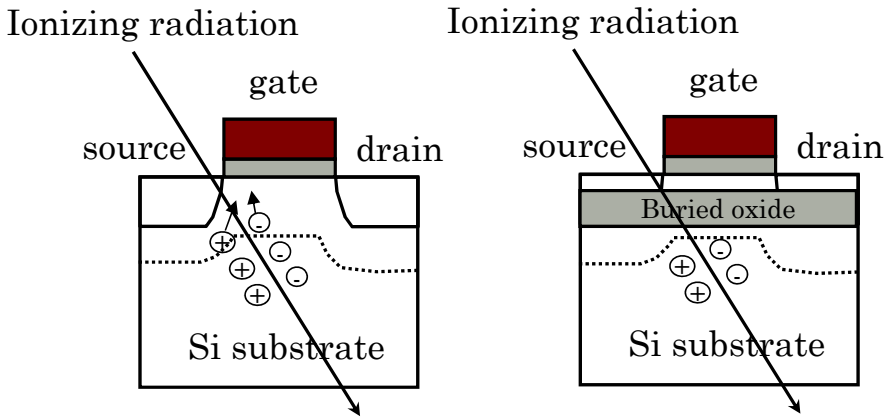
Silicon-On-Insulator (SOI) Substrates



Why ?

- Historically: → rad-hard devices
- Better lateral isolation, compact solution
- Less parasitic capacitances
- Reduced junction leakage
- Better short channel effects
- 3D integration possible

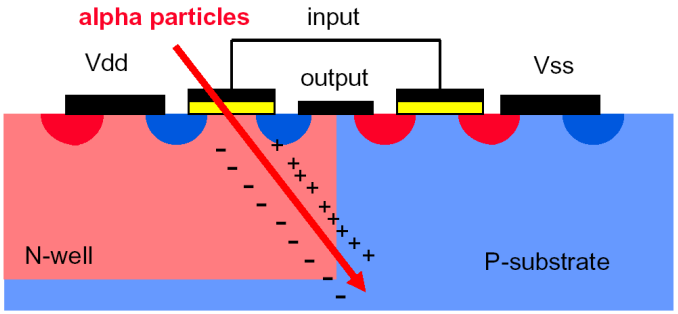
Today SOI: → low-voltage, high-speed, RF



Naturally less sensitive to ionizing radiation

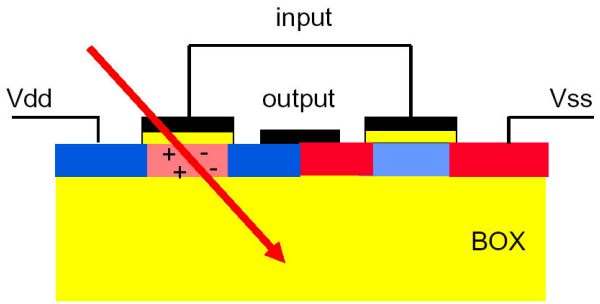
Less soft errors in SOI than in bulk CMOS

soft errors in bulk CMOS



Alpha particles → "soft" errors

soft errors in SOI



Low soft error rate

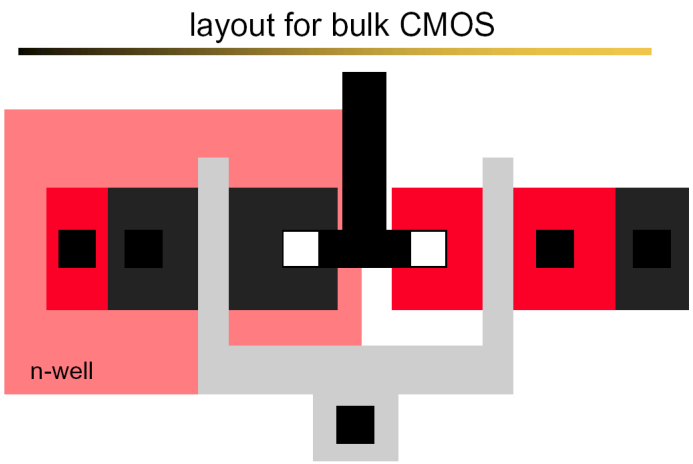
Sources of alpha particles

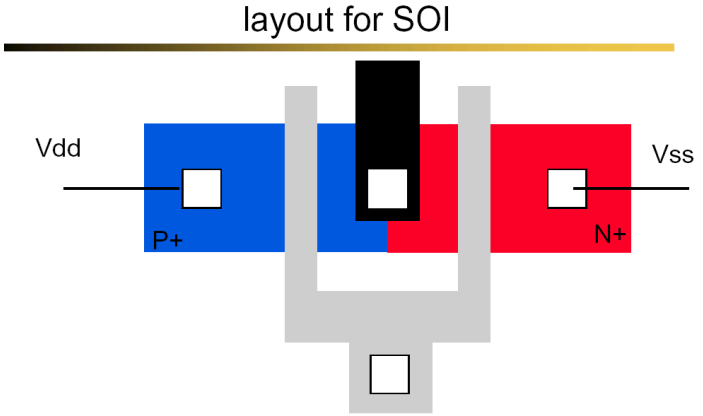
- Cosmic Rays (aircraft electronics vulnerable)
- Decaying uranium and thorium impurities in integrated circuit interconnect

Generates electron-hole pairs in substrate

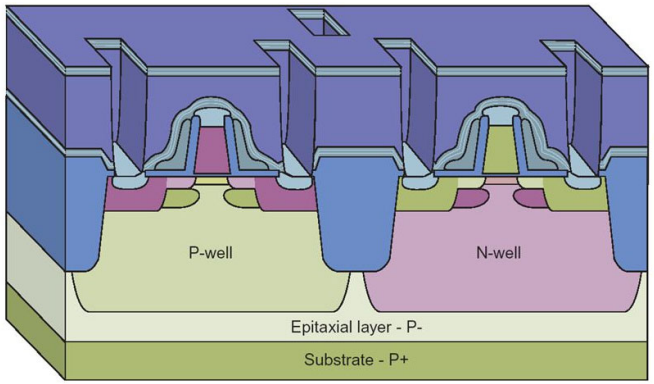
- Excess carriers collected by diffusion terminals of transistors
- Can cause upset of state nodes – floating nodes, DRAM cells most vulnerable

SOI versus bulk CMOS inverter design



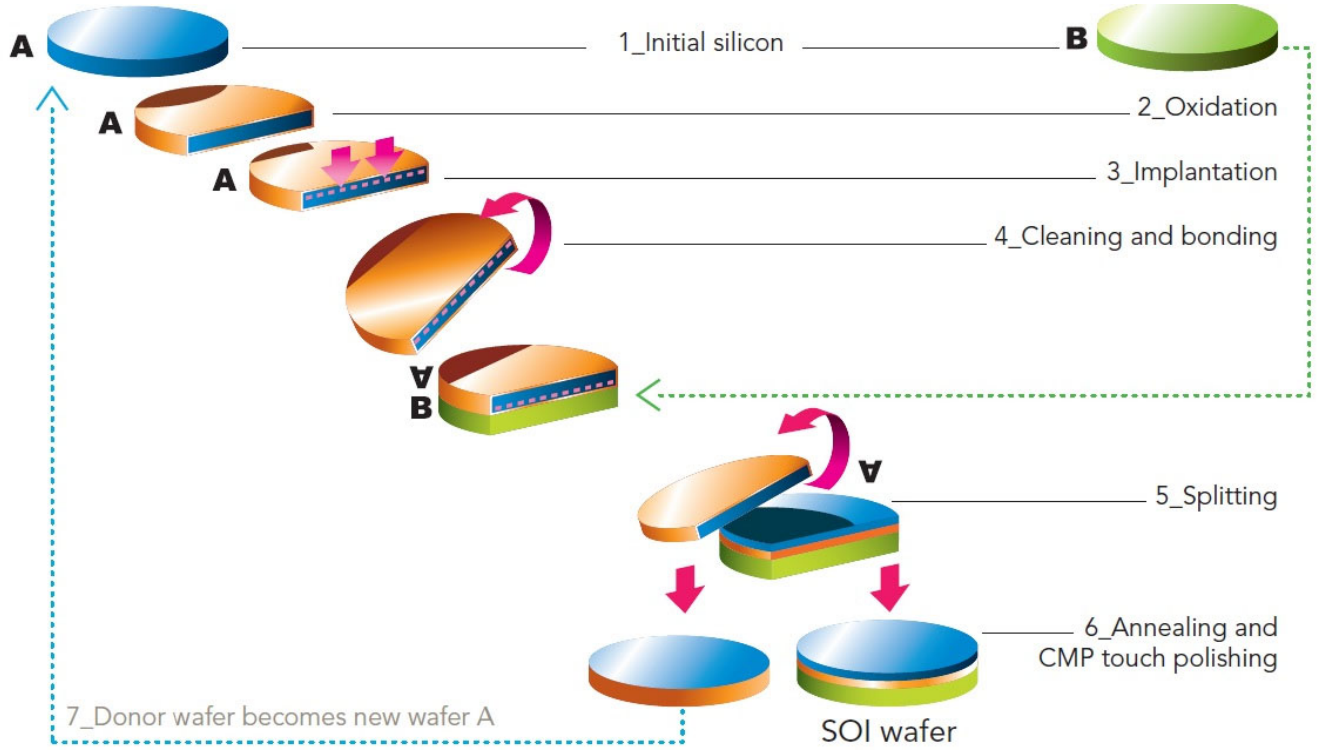


Simpler isolation → simpler process → smaller layout



MUCH COMPACT, SMALLER AREA

Smartcut process for Silicon On Insulator substrates



Invented by CEA-LETI & SOITEC, Bernin, France

SOI Wafers: state of the art

Fully-depleted Silicon-on-Insulator (FD-SOI) – substrates

- The prevalent method for FD-SOI substrate manufacture uses the "SmartCut" process, licensed by SOITEC to wafer suppliers.

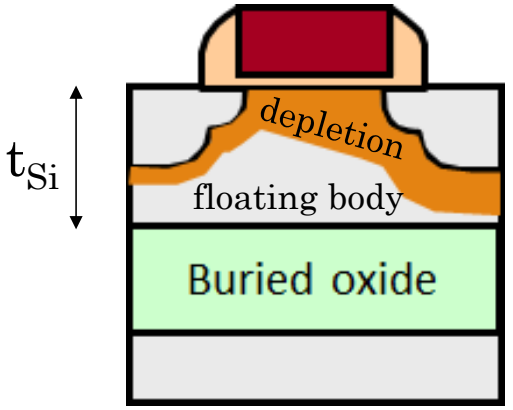
For the 28nm node, the uniformity of the thin Silicon layer is +/- 5A over the 300mm wafer, equivalent to 0.2" between Chicago and San Francisco.

http://www.soitec.com/en/products-and-services/microelectronics/fd-2d/

Partially Depleted (PD) SOI MOSFET

- **Si thickness is higher than the depletion depth → PD, neutral body**

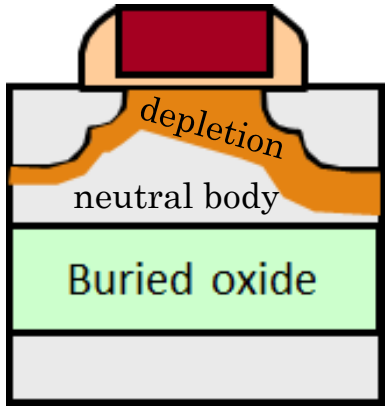
+ similar to bulk-Si
+ no coupling
- Short channel effects (SCE): as in bulk
- floating body
- Kink effect
- dynamic over- and under-shoots
- self-heating effects



- CMOS easier to manufacture in PD SOI
- no significant advantage for nanometer scaling
- unless the body is tied electrically to the source → **floating body effects**

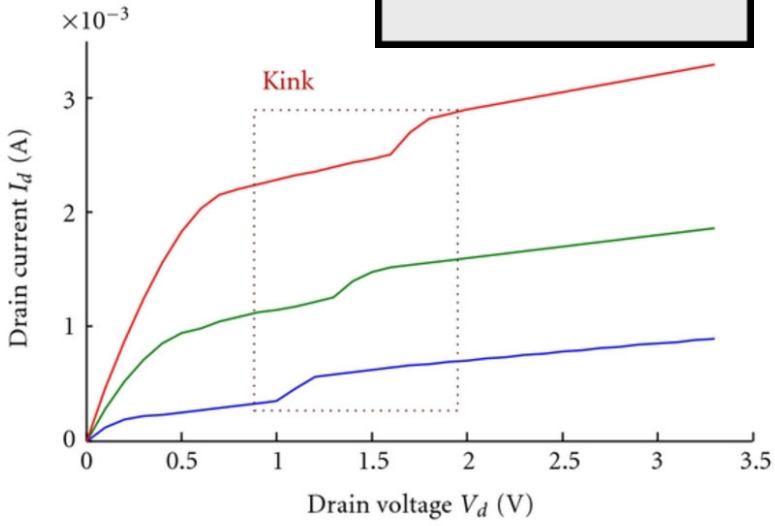
Partially Depleted (PD) SOI MOSFET: What is floating body effect

$$T_{SOI} > W_T, \text{ where } W_T = \sqrt{\frac{2\varepsilon_s(2\phi_F)}{qN_{body}}}$$



Floating body effect (*history dependent*):

- When a PD-SOI NMOSFET is in the ON state, at moderate-to-high V_{DS}, holes are generated via impact ionization near the drain
- Holes are swept into the neutral body, collecting at the source junction
- The body-source pn junction is forward biased: → V_T is lowered → I_{Dsat} increases → **"kink" in output I_D *vs.* V_{DS} curve**



Fully Depleted (PD) SOI MOSFET

$$T_{SOI} < W_T, \text{ where } W_T = \sqrt{\frac{2\varepsilon_s(2\phi_F)}{qN_{body}}}$$

+ Excellent control of short channel effects

+ No floating body effect

+ Less capacitance parasitics

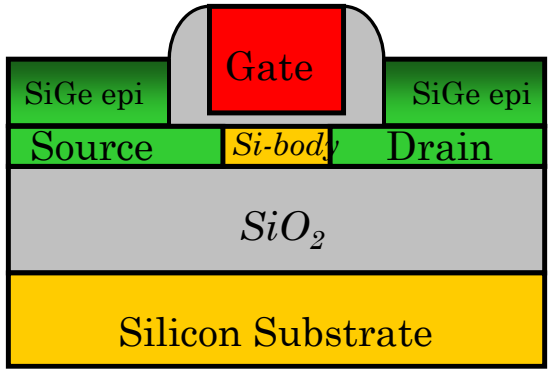
+ Less junction leakage

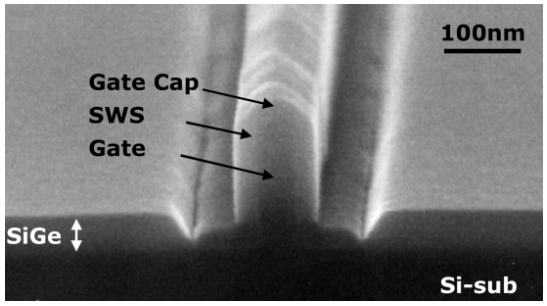
+ Quasi-ideal subthreshold swing

- V_T is sensitive to SOI film thickness

- Elevated (or recessed) S/D contact structure needed to reduce series resistances, R_S, R_D

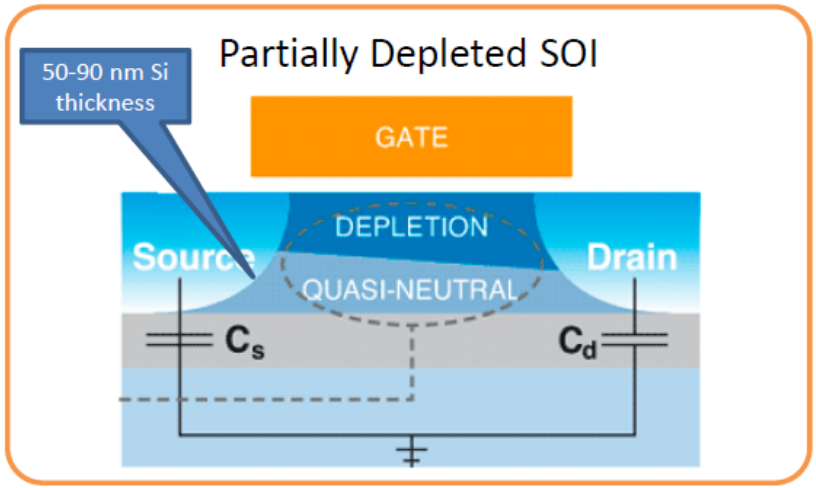
- Self heating effects



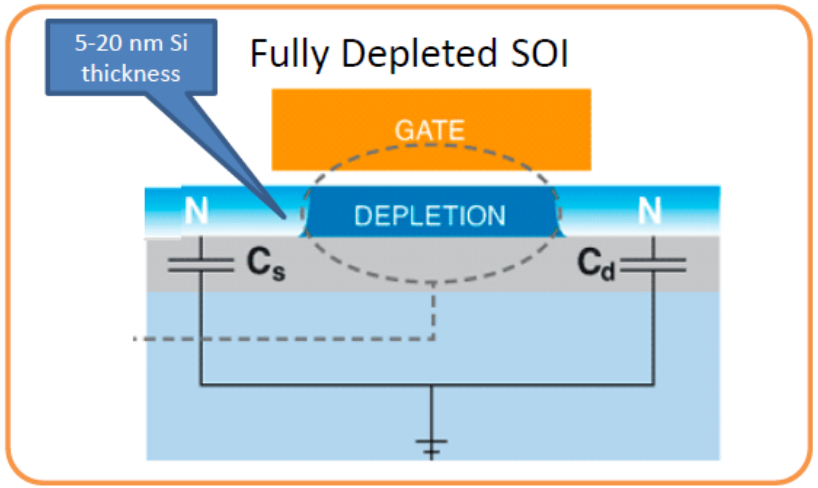


Summary: Fully Depleted (FD) versus Partially Depleted (PD) SOI MOSFET

PD-SOI

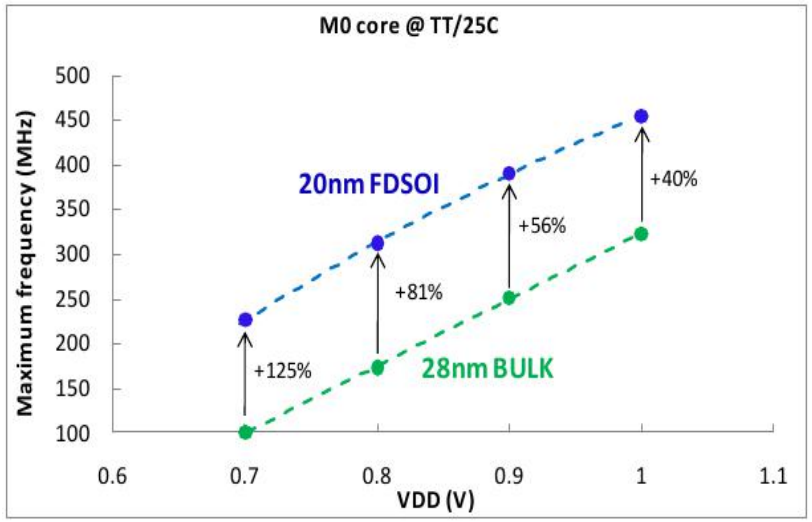


FD-SOI



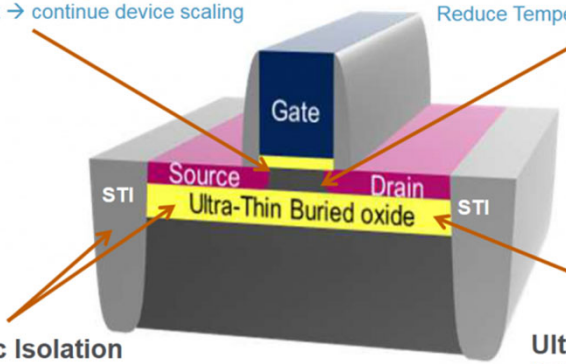
Best solution for nanometer scaling and low power

FD SOI CMOS versus Bulk CMOS: Role of Body Bias (!)



Ultra-Thin BODY
Excellent Electrostatic Control of the channel
Low SCE, DIBL → Low V_t @ High V_D
Gate length shrink → continue device scaling

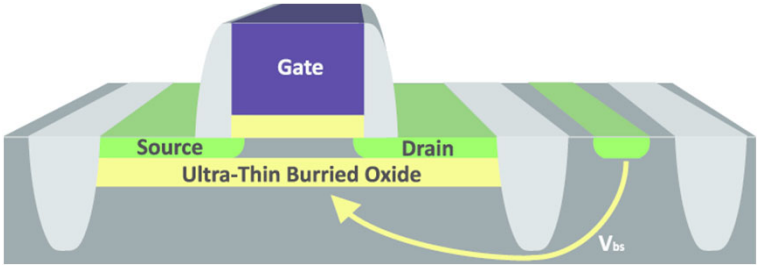
Undoped Channel & no Pocket
Less V_t variability & SRAM V_{min} improved
Lower Power Consumption
Reduce Temperature dependency & no RDF



Total Dielectric Isolation
Lower SD capacitances & Lower SD Leakage
Less Sensitive to Temperature
Higher power efficiency

Ultra Thin BOX
Body Biasing (BB) → FBB & RBB
Speed boost due to BB
GP Implantation → V_t adjustment

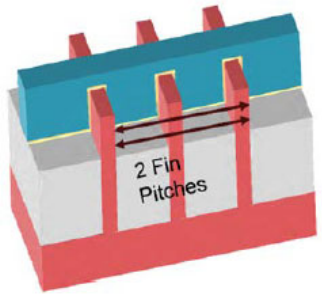
+ body bias, Vbs, design flexibility

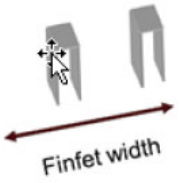


Fin FET versus FD SOI MOSFET: which one is best?

The only two remaining technology options for sub-10nm scaling

Category	FinFet	FD-SOI
Base Wafer Cost		x4
Process Complexity	↑	↓
Overall Wafer Cost	→ to ↗	→ To ↘
Die Yields	??????	??????
Unit Cost	??????	??????
Process control & metrology challenges	↑	↓
Active transistor area density	↑	↓
Performance (Ion vs. Ioff)	Similar	Similar





1D (nanowire) FETs

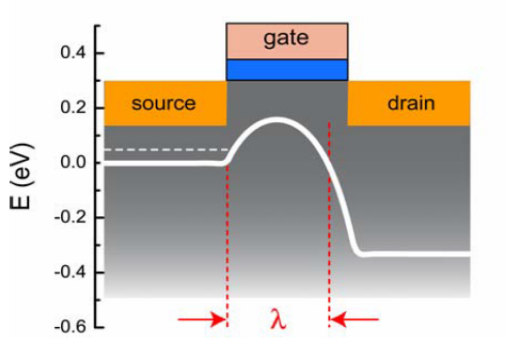
Why **nanowires (NWs)** and nanotubes (NTs):

- due to their **potential to test fundamental concepts** about how dimensionality and size affect physical properties
- can serve as basic **building blocks of emerging technology platforms**
- enable **new integrated functionality** in highly dense low cost integrated circuits
- advanced existing fabrication schemes: **Si-toolset**
- tunable electrical properties by **controlled doping**
- **predictable electron transport**: enhanced engineering of device characteristics
- well-defined surface structural properties: **enhanced interfacial engineering**
- **Vertical and lateral isolation** possible with SOI devices

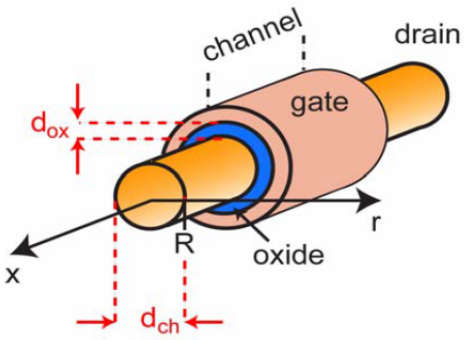
Electrostatics of nanowire FET vs. planar MOSFET

Comparison of Planar vs. Nanowire Architecture

Planar MOSFET



Nanowire MOSFET:



- **Planar gate:**
  limited electrostatic control of the channel

$$\lambda = \sqrt{\frac{\varepsilon_{ch}}{\varepsilon_{ox}} d_{ox} d_{channel}}$$

- **Example:**
  8nm SOI, 1 nm SiO2: $\lambda \approx 5$ nm
  $\Rightarrow$ $L_G > 20$ nm

- **Surround gate (Nanowire):**
  ultimate electrostatic control of channel

$$\lambda = \sqrt{\frac{\varepsilon_{nw} d_{nw}^2 \ln\left(1 + \frac{2d_{ox}}{d_{nw}}\right)}{8\varepsilon_{ox}}}$$

- **Example:**
  8 nm SiNW 1 nm SiO_2: $\lambda \approx 2.3$ nm
  $\Rightarrow$ $L_G > 9$ nm

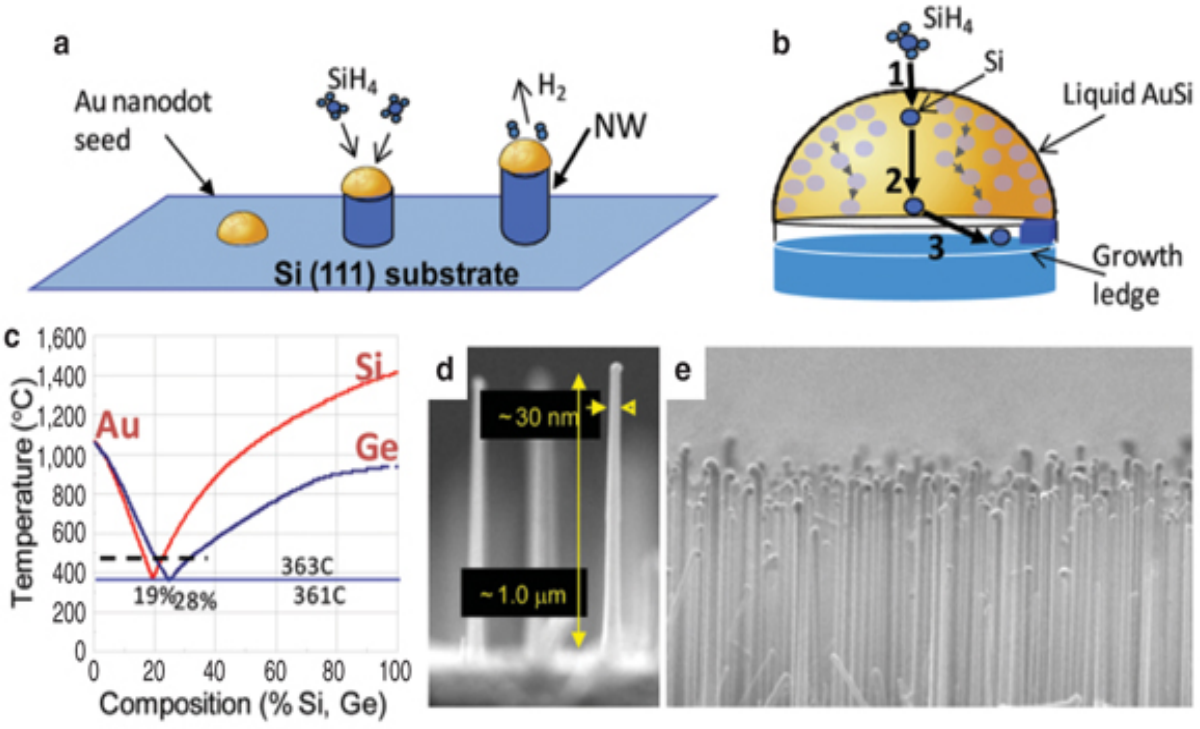
⇨ **NW device geometry yields improved scaling and better inv. subthreshold slope**

Source: Heike Riel, IBM.

Bottom-up nanowires by VLS Growth

VLS Growth Technique

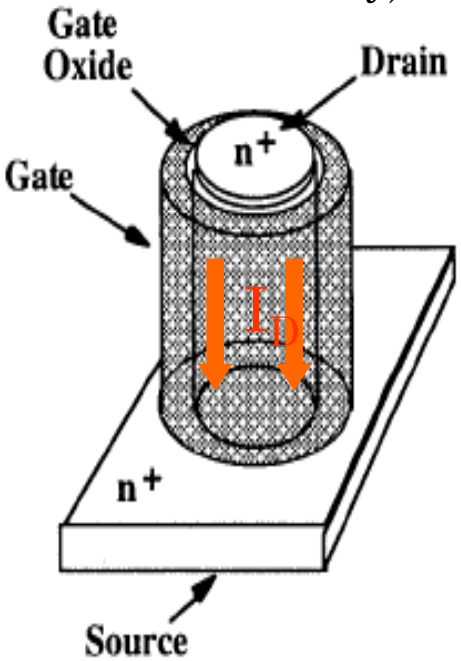
- **Gas precursor molecules** or vapor containing the growth species are introduced into a furnace or chemical vapor deposition (CVD) reactor and liquid metal catalysts on a surface react with the source atoms to grow nanowires.
- **A metal nanodot** catalyzes the growth of the nanowire from a vapor source by forming a liquid metal droplet through which the growth atoms are transported to the crystallizing interface. The catalytic seed floats on top of the nanowire and defines the nanowire diameter while the growth rate kinetics together with the growth time defines the nanowire length.
- **The nanowire 'self-assembles' in a bottom-up synthesis technique.**



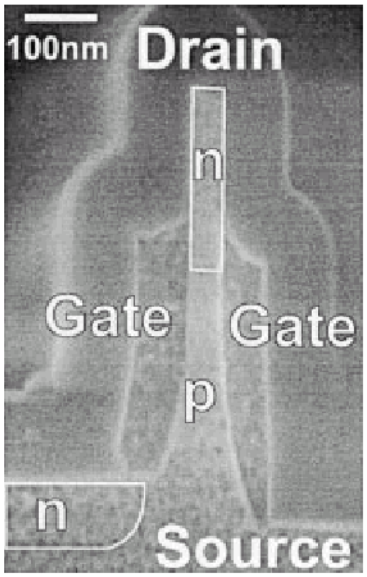
diameters ~5 nm up to ~100 nm

Vertical nanowire MOSFETs

- Idea: **channel length is defined by the thickness of a semiconductor material**
- Design: VERTICAL, contacts at top & bottom (source / drain = bottom / top or vice-versa)
- Performance: similar to 1D NW MOSFET
- Advantage: more dense (suitable for memory)



Infineon's vertical MOSFET



Application: SRAM

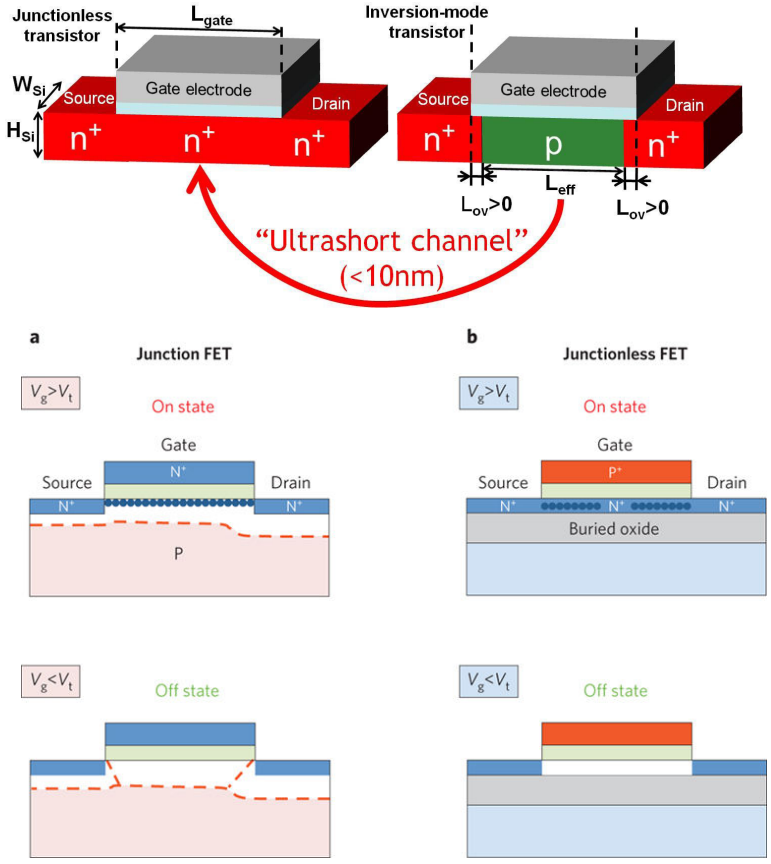
T. Schulz et al., IEEE TED, August 2001.

Junctionless MOSFETs (1)

Operation principle for a junctionless n-type accumulation device:

1. **On-state:** the whole nanowire body behaves as a conducting channel: on current depends on the nanowire's geometry, mobility and doping.
2. **Off-state:** the gate voltage is decreased, the transistor body is gradually depleted, until eventually the transistor turns off.

Merit: much simpler than the junction-based inversion-mode MOSFET and much easier to manufacture.



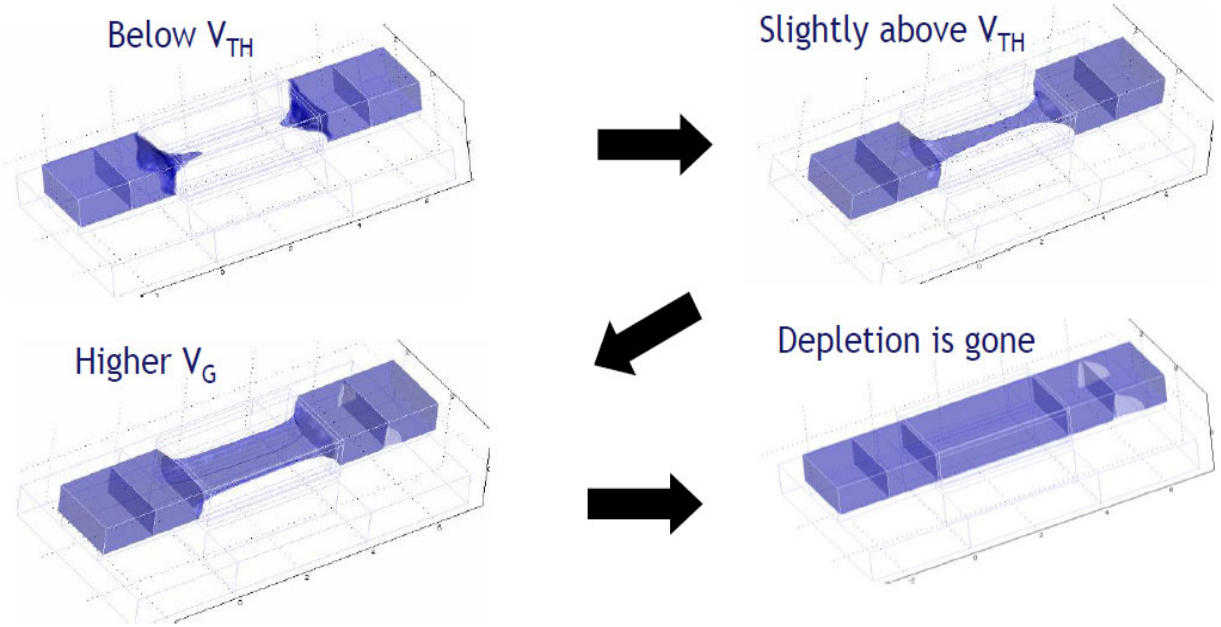
J.P. Colinge et al., Nanowire transistors without junctions, Nature Nanotechnology 5, 225 - 229 (2010)
A.M. Ionescu, Nanowire transistors made easy, Nature Nanotechnology 5, 178 - 179 (2010).

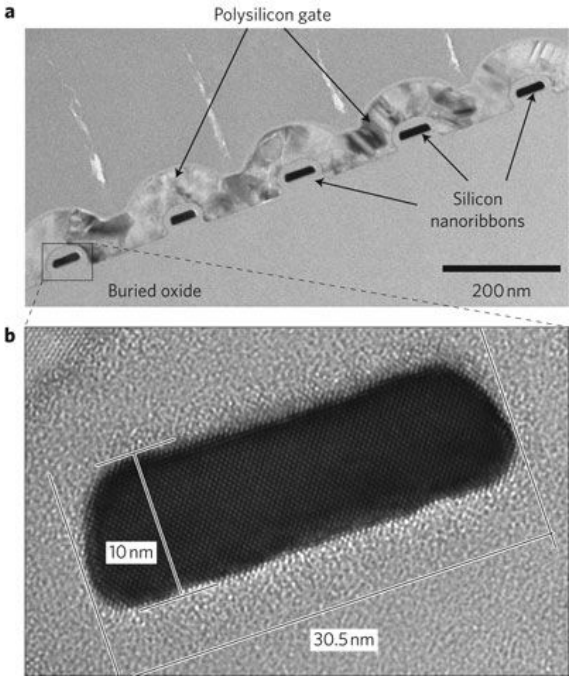
Junctionless MOSFETs (2)

Electrostatic pinch-off:

The cross section is small enough for the channel region to be depleted

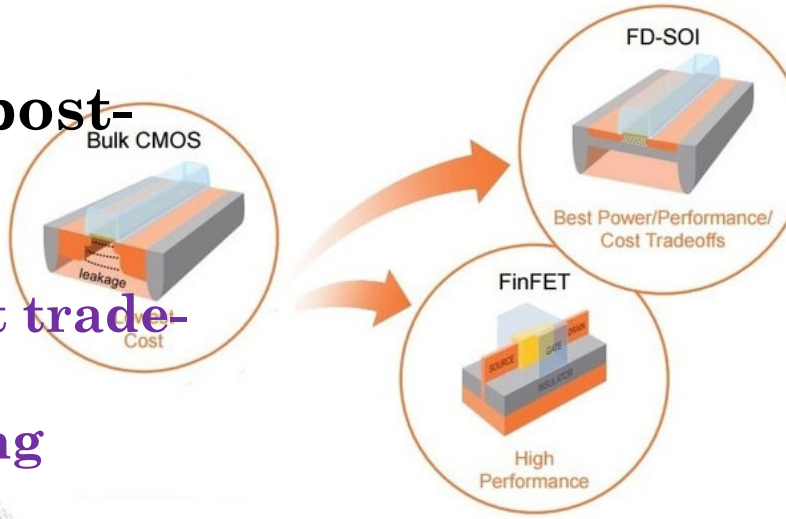
(V_D=50mV, N_d>5e18/cm^3)





Conclusions

- **Main challenges of non-conventional post-Dennardian CMOS:**
  - **3D devices for best electrostatics**
  - **FinFET / UTB FD SOI / NW FETs for best trade-offs**
  - **FInFETs for high performance computing**
  - **Power leakage control**
  - **High performance versus energy efficiency (low power) demands**
  - **Complexity and cost**



~2nm NW MOSFET seems to be the end of the roadmap in channel length scaling.