VLSI Devices
Lecture 22

Sung-Min Hong (smhong@gist.ac.kr)
Semiconductor Device Simulation Laboratory
Department of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)

Coverage

- Two YouTube lectures reserved for advanced topics
  - L14: ~~Substrate bias, channel mobility~~
  - L15: ~~3.2.1~~
  - L16: ~~3.2.1 (Continued)~~
  - L17: ~~Velocity saturation (3.2.2)~~
  - L18: ~~Channel length modulation and so on (3.2.3, 3.2.4, 3.2.5)~~
  - L19: MOSFET scaling
  - L20: MOSFET scaling (Continued)
  - L21: Quantum effect (4.2.4)
  - L22: Double-gate MOSFETs (10.3)
  - L23: FinFETs
  - L24: CFETs

MOSFET scaling

- First of all, we must understand the history. (~ 2011)
  - Comtemporary MOSFETs are not planar.
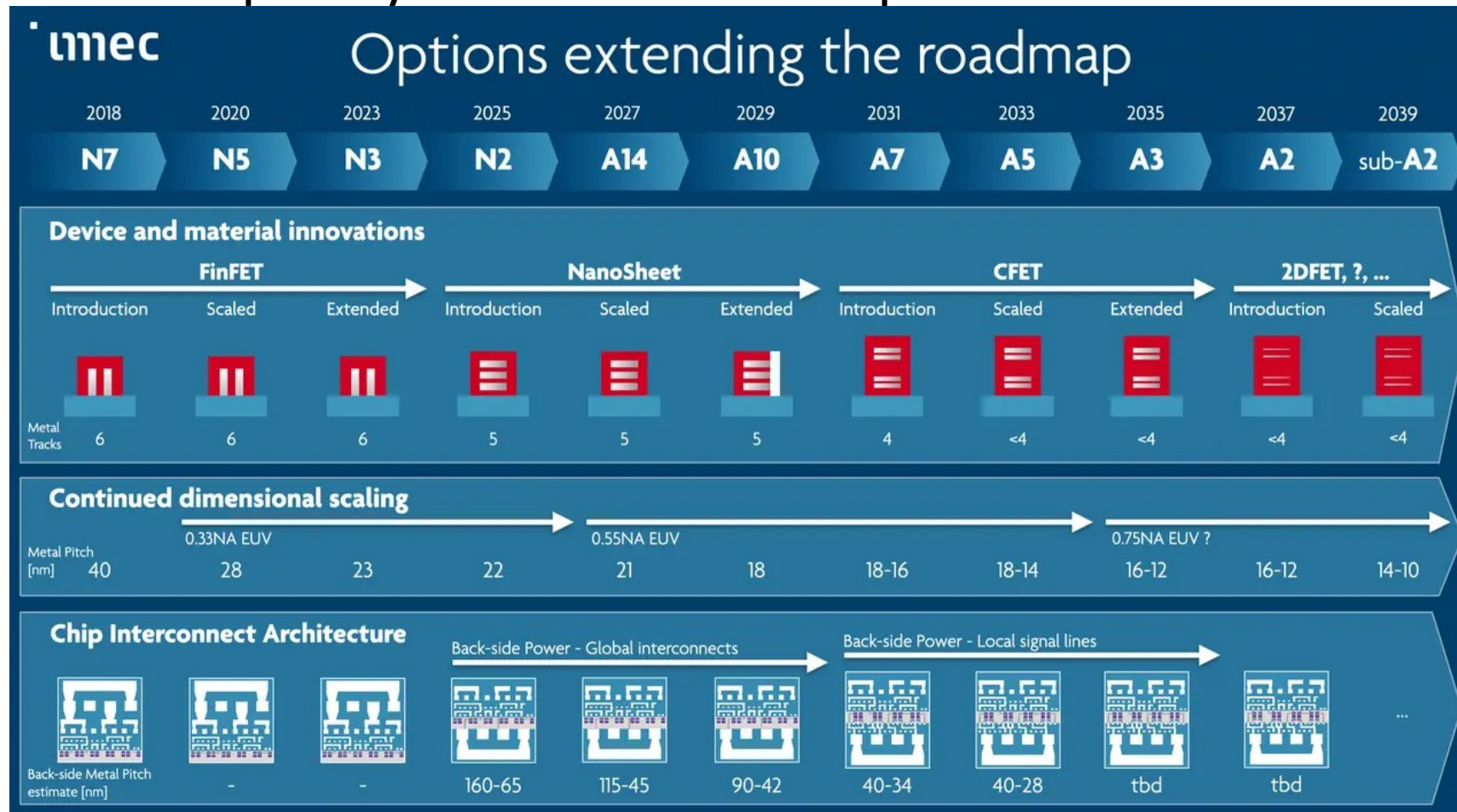
IMEC roadmap

Basic assumptions

- High-resolution lithographic techniques (Minimum L)
- Technological advancement in ion implantation (Shallow junction)

An architect and a construction worker (Image generated by ChatGPT)

Constant-field scaling (Dennard scaling)

- Keep short-channel effects under control,
  - By scaling down the vertical dimensions along with the horizontal dimensions.
  - Decrease the applied voltage.
  - Increase the substrate doping concentration.
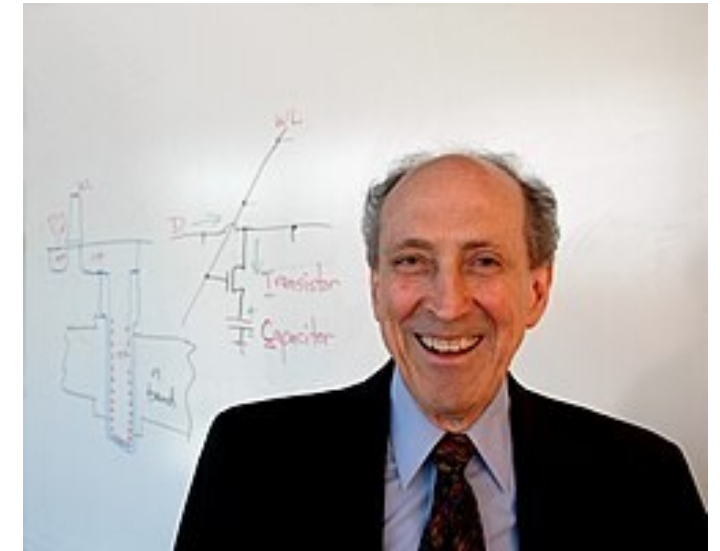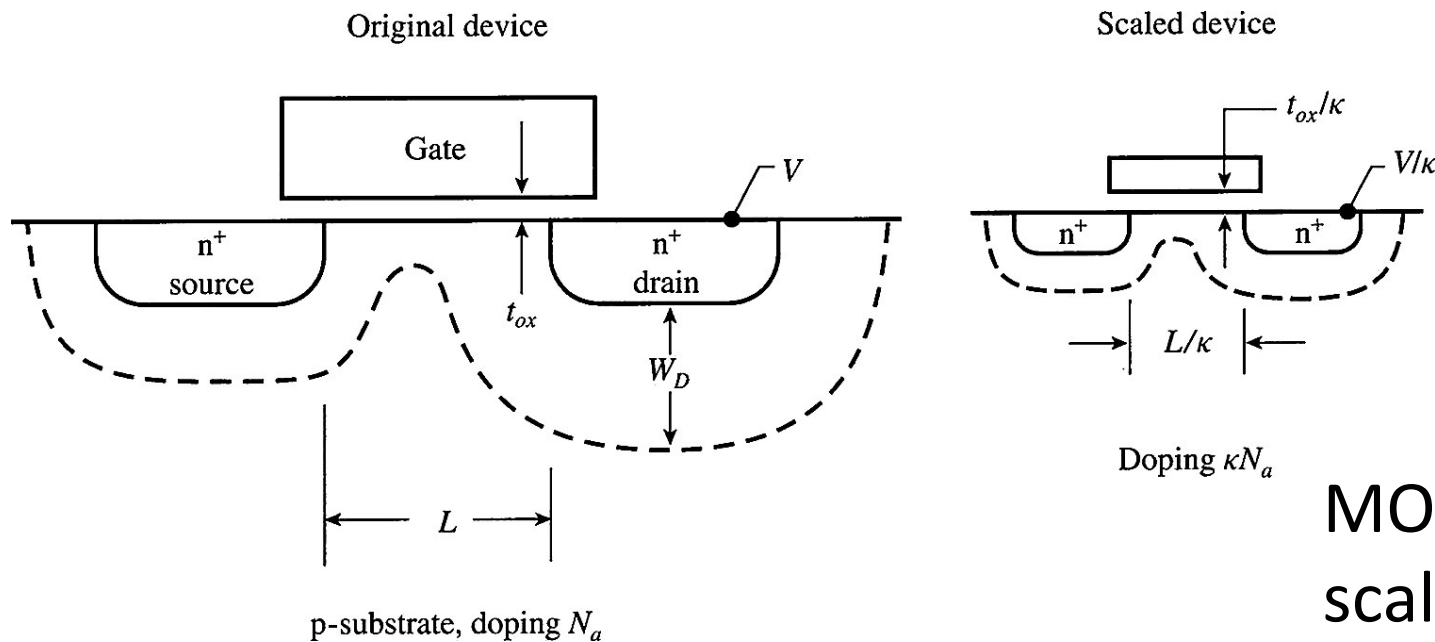
R. H. Dennard
(Inventor of DRAM)
MOSFET constant-electric-field
scaling (Taur, Fig. 4.1)

Rules for constant-field scaling (1)

- Scaling assumption ($\kappa > 1$)
  - Device dimensions (t_{ox}, L, W, and x_j): $1/\kappa$
  - Doping concentration (N_a and N_d): κ
  - Voltage (V): $1/\kappa$

- Maximum drain depletion width

$$W_D = \sqrt{\frac{2\epsilon_{si}(\phi_{bi} + V_{dd})}{qN_a}} \rightarrow \sqrt{\frac{2\epsilon_{si}\left(\phi_{bi} + \frac{1}{\kappa}V_{dd}\right)}{q\kappa N_a}}$$

Taur, Eq. (4.1)

$$\approx \frac{1}{\kappa}\sqrt{\frac{2\epsilon_{si}(\phi_{bi} + V_{dd})}{qN_a}} = \frac{1}{\kappa}W_D$$

Rules for constant-field scaling (2)

- Capacitances
  - They scale down by κ.

- Charge per device ($\sim C \times V$)
  - It scaled down by κ^2.

- Drain current
  - The original one

$$I_d = \mu_n C_{ox} \frac{W}{L}\left[(V_{gs} - V_t)V_{ds} - \frac{1}{2}V_{ds}^2\right]$$

$$\mu_n \kappa C_{ox} \frac{\frac{1}{\kappa}W}{\frac{1}{\kappa}L}\left[\left(\frac{1}{\kappa}V_{gs} - V_{t,scaled}\right)\frac{1}{\kappa}V_{ds} - \frac{1}{2}\frac{1}{\kappa^2}V_{ds}^2\right] \approx \frac{1}{\kappa}I_d$$

Effect of scaling on circuit parameters

- Important colcusion of constant-field scaling:
  - Once the device dimensions and the power-supply voltage are scaled down, the circuit speeds up by the same factor.

$$\tau \sim \frac{CV}{I}$$

with $\frac{1}{\kappa^2}$ pointing to C, $\frac{1}{\kappa}$ pointing to V, and $\frac{1}{\kappa}$ pointing to I.

  - Moreover, power dissipation per circuit, which is proportional to VI, is reduced by κ^2.

Chronicles

- Intel and TSMC. IEDM and VLSI papers
  - 130nm: 2000 (Intel, IEDM)
  - 90nm: 2003 (Intel, IEDM)
  - 65nm: 2004 (Intel, IEDM)
  - 45nm: 2007 (Intel, IEDM)
  - 32nm: 2008 (Intel, IEDM)
  - 22nm: 2012 (Intel, VLSI)
  - 16nm: 2013 (TSMC, IEDM), 14nm: 2014 (Intel, IEDM)
  - 10nm: 2016 (TSMC, IEDM)
  - 7nm: 2016 (TSMC, IEDM)
  - 5nm: 2019 (TSMC, IEDM)
  - 3nm: 2022 (TSMC, IEDM)
  - 2nm: 2024 (TSMC, IEDM)

Prehistoric(?) MOSFET

- 130-nm MOSFET
  - Major issue: Cu interconnection (Previously, Al)
  - Operation voltage: 1.3 V
  - Oxide thickness: 1.5 nm
  - Poly-silicon gate
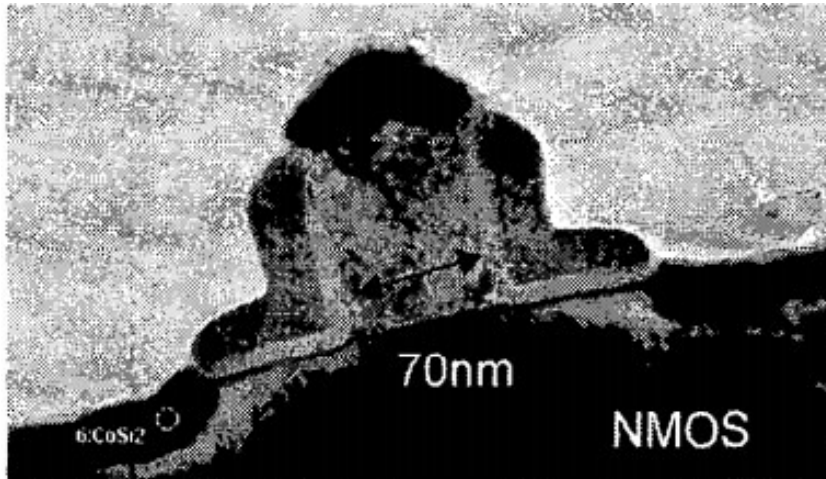
SUMMARY OF TRANSISTOR CHARACTERISTICS

| Parameter		180 nm Generation [1]	This Work
V_{DD}	[V]	1.5	1.3
L_{GATE}	[nm]	130	70
T_{OX}	[nm]	2.0	1.5
I_{OFF}	[nA/μm]	3	10
$I_{DSAT}(n)$	[mA/μm]	1.04	1.02
$I_{DSAT}(p)$	[mA/μm]	0.46	0.5
Low Vt I_{OFF}	[nA/μm]	-	100
Low Vt $I_{DSAT}(n)$	[mA/μm]	-	1.17
Low Vt $I_{DSAT}(p)$	[mA/μm]	-	0.6

Running out of steam

- In the early 2000s. Why?
  - Due to leakage limitations!
  - The SiO_2 gate oxide had scaled to ~1.2 nm at the 90 nm generation. ➔ The gate oxide leakage was increasing exponentially and had become a noticeable percentage of total chip power.
  - Deceasing supply voltage ➔ decreasing threshold voltage ➔ ever-higher subthreshold leakage current
- Increasing transistor leakage
  - Was against the market preferences.
  - The 1980s and 1990s were the era of the home PC.
  - The 2000s was the "mobile" era.

90-nm node

- Oxide thickness: 1.2nm ($\times$ 0.8 scaling, not $\times$ 0.7)
  - SiGe was selectively deposited in PMOS source-drain regions to provide compressive channel strain.
  - A tensile SiN cap layer was deposited over NMOS transistors to provide tensile channel strain.

90nm uniaxial strained silicon transistors
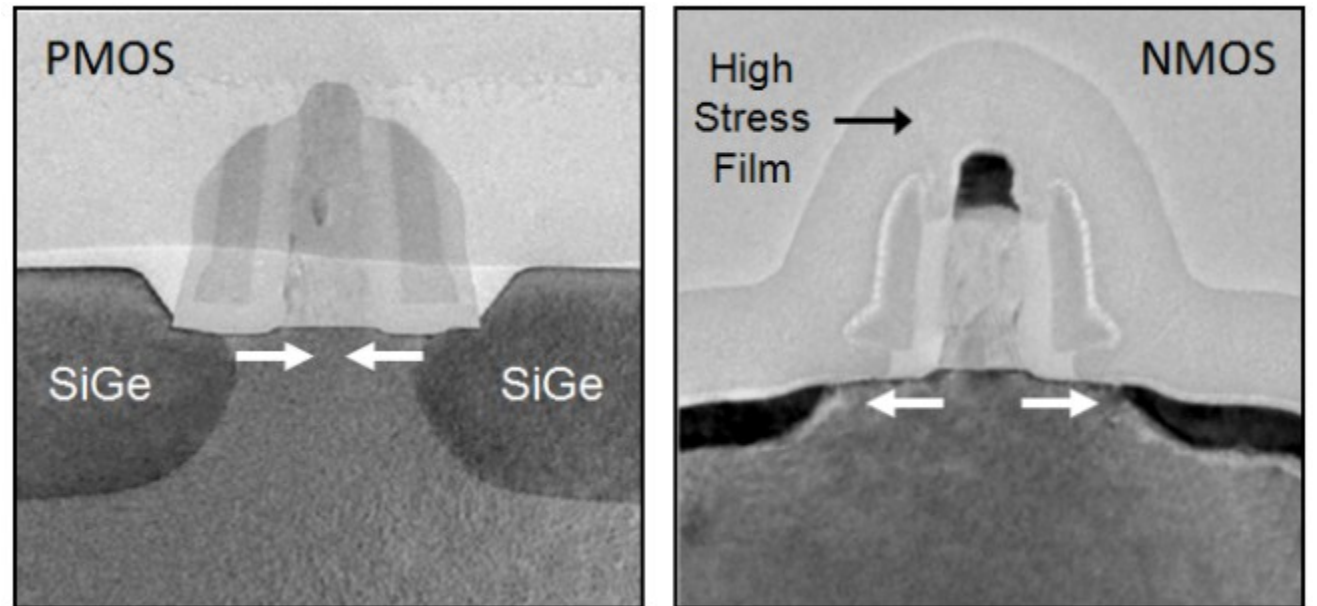
Its impact

- Mobility enhancement
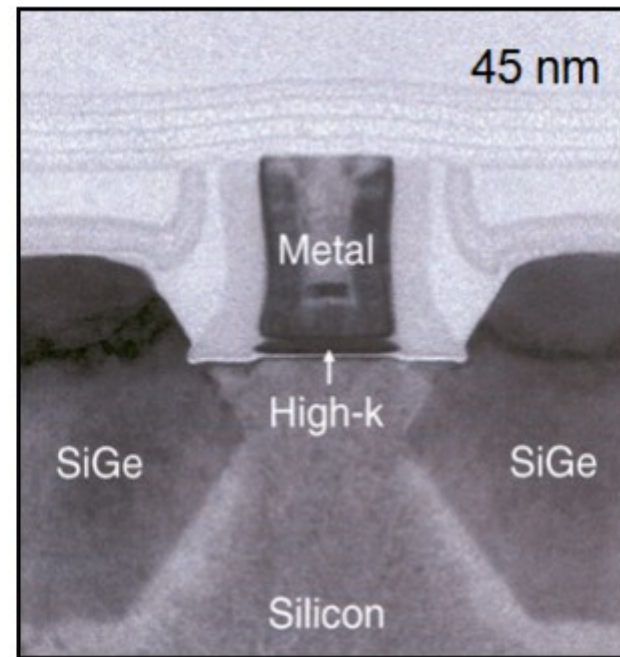 - For electrons, tensile strain. For holes, compressive strain
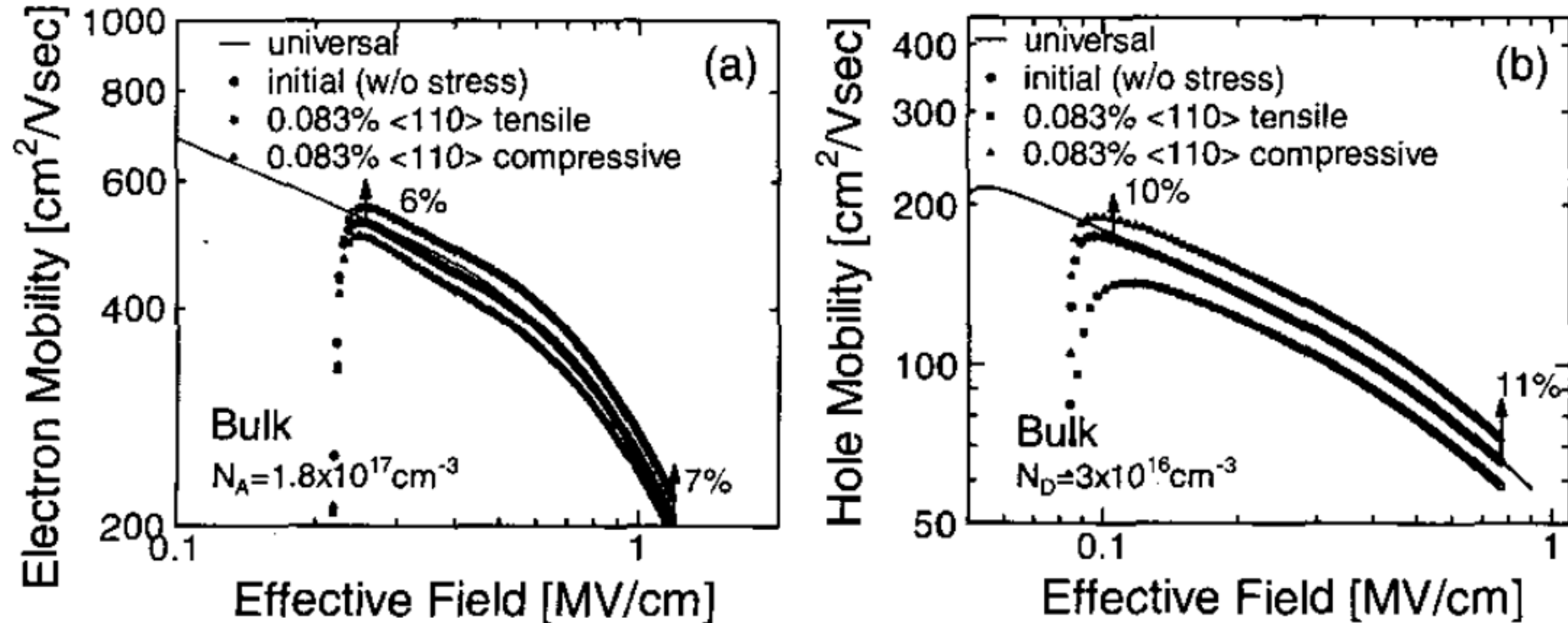
Mobility characteristics under <110> uniaxial strain (K. Uchida, IEDM 2004)

"High-k + metal" gate

- We need to scale the "effective" oxide thickness.
  - Keeping the physical thickness & increasing the oxide capacitance
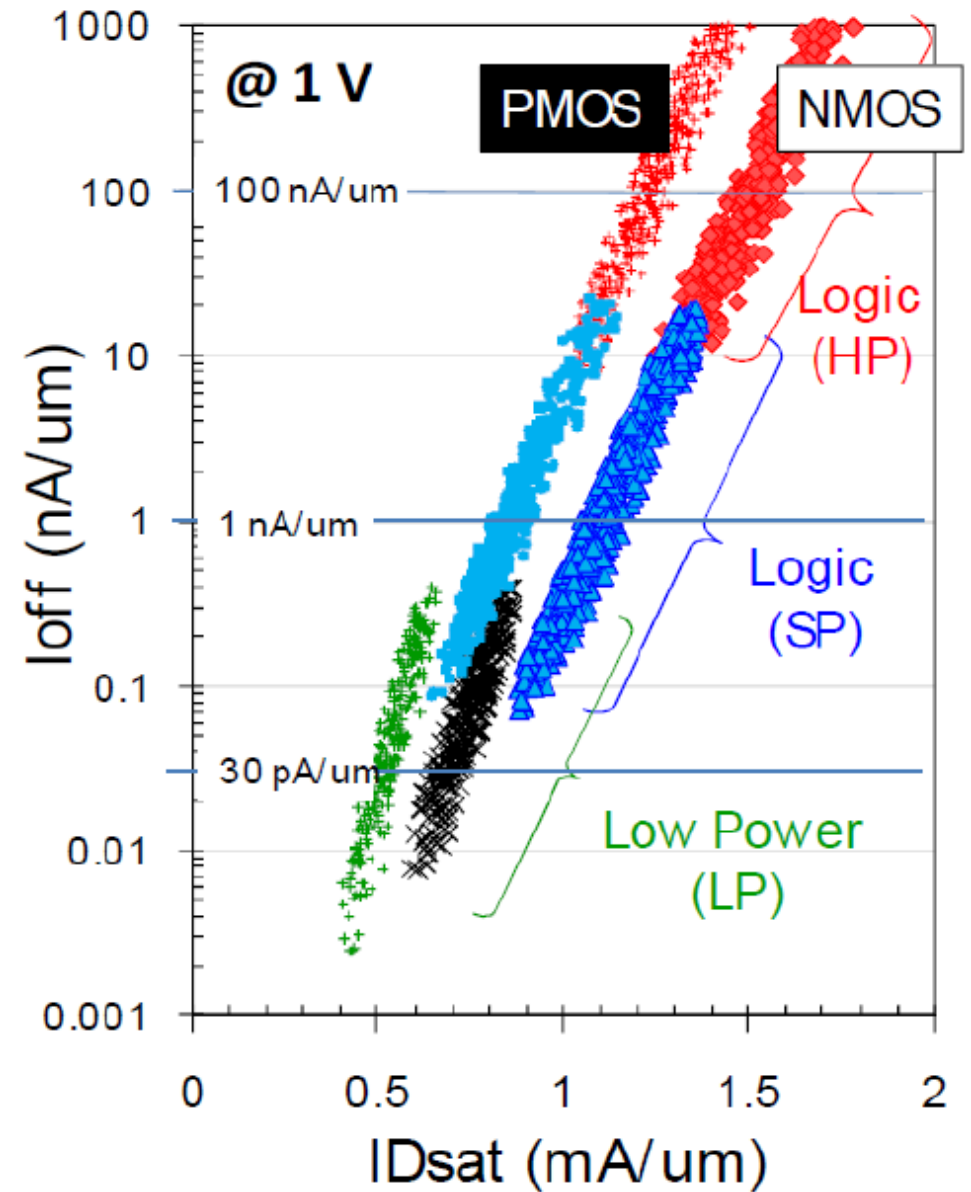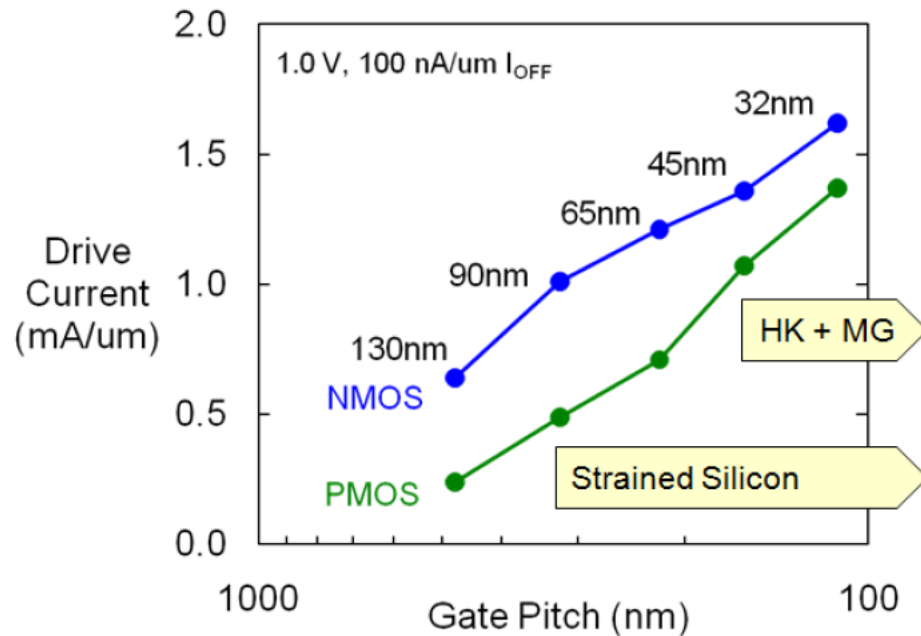- Poly depletion effect is now removed.
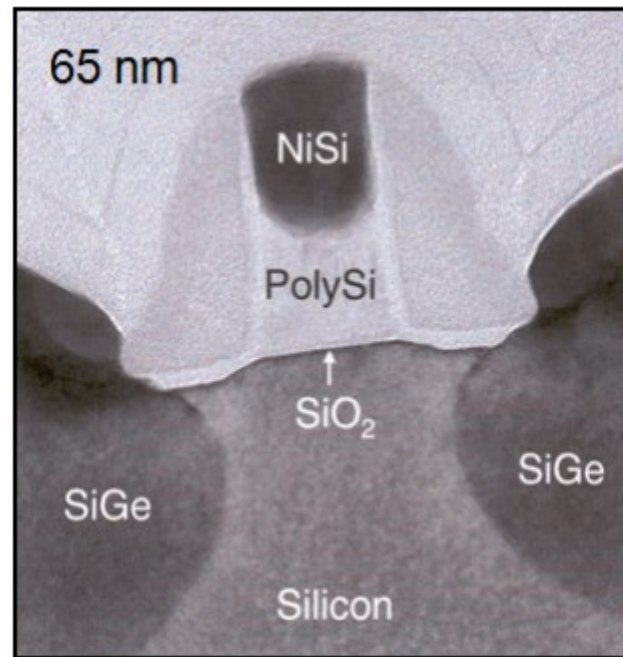
65nm   45nm (high-k metal gate)
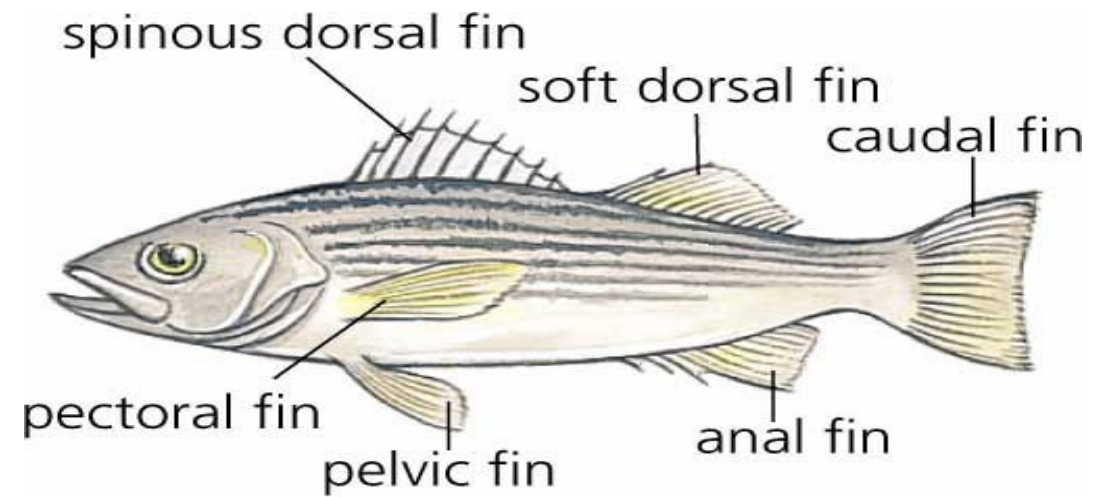
Performance

- Drive current improvement
  - Strain engineering
  - High-k + metal gate
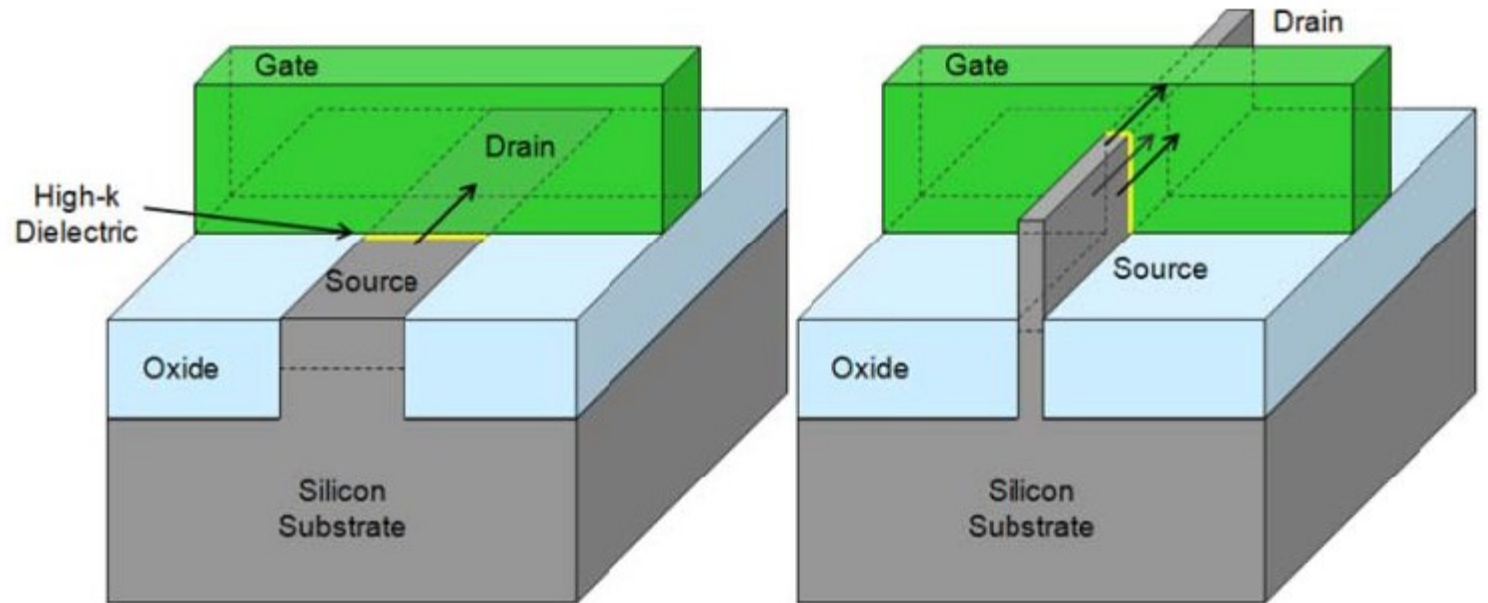

32nm transistor performance and leakage options

FinFET

- FinFET
  - Following its shape
 - Initially proposed as a SOI FinFET
  - Later, a bulk FinFET

spinous dorsal fin
soft dorsal fin
caudal fin
pectoral fin
pelvic fin
anal fin

Elizabeth Morales

Fins (Google images)

Gate
Drain
High-k Dielectric
Source
Oxide
Silicon Substrate

Gate
Drain
Source
Oxide
Silicon Substrate

Planar transistor structure (left) and FinFET structure (right)
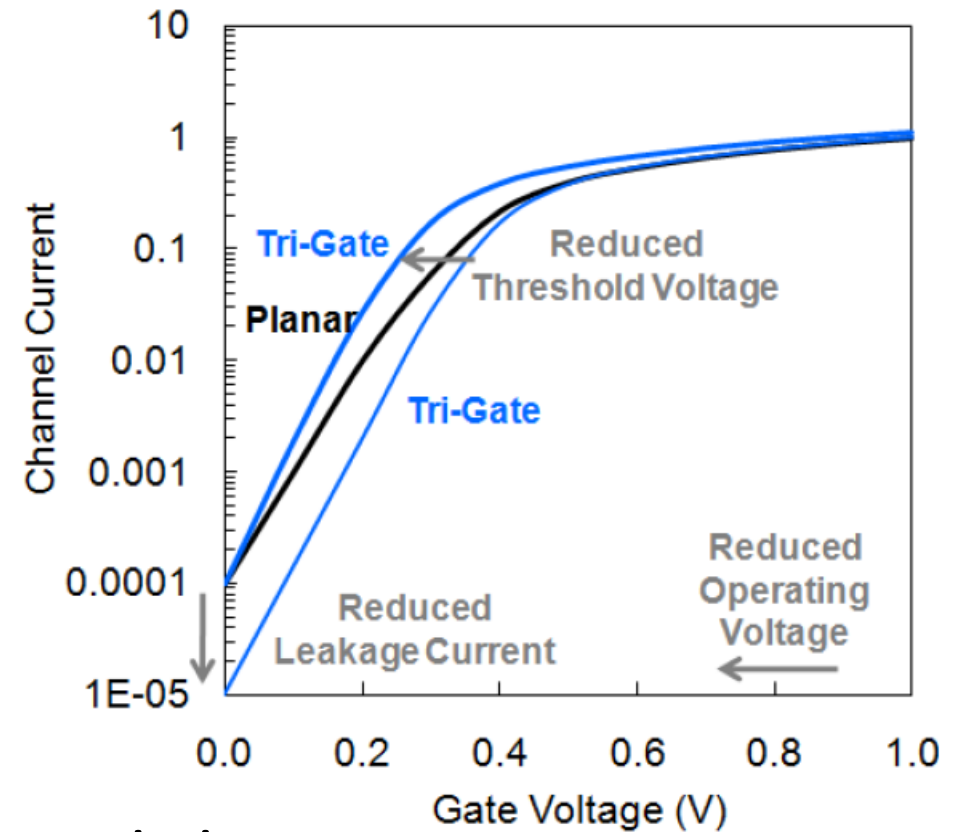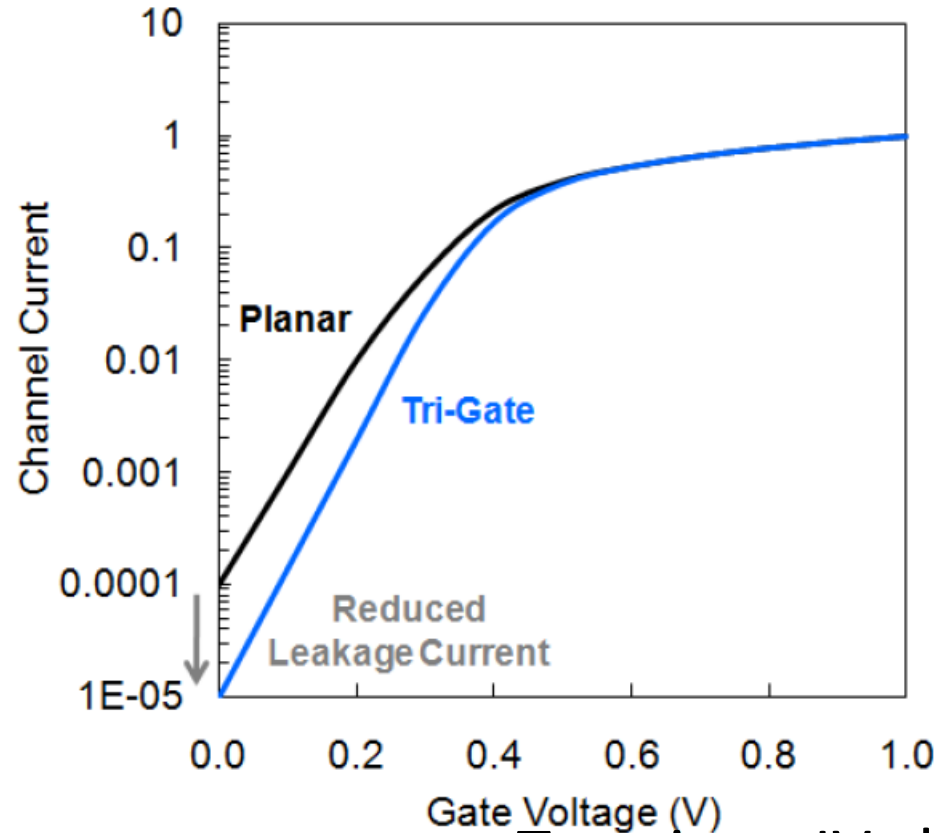
SEM image

- FinFET
  - Improved electrostatic control of the channel region
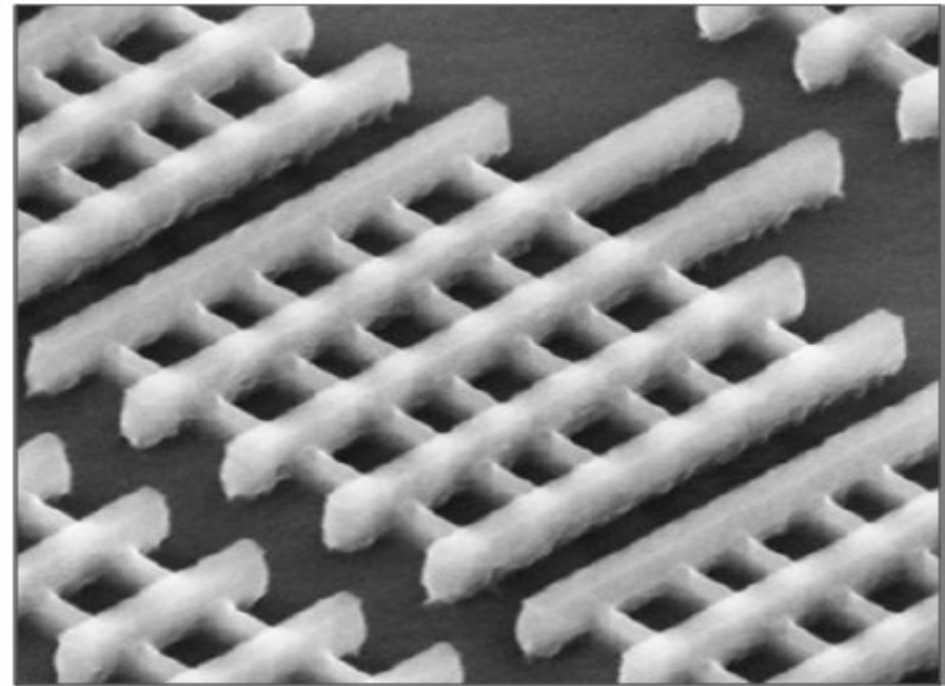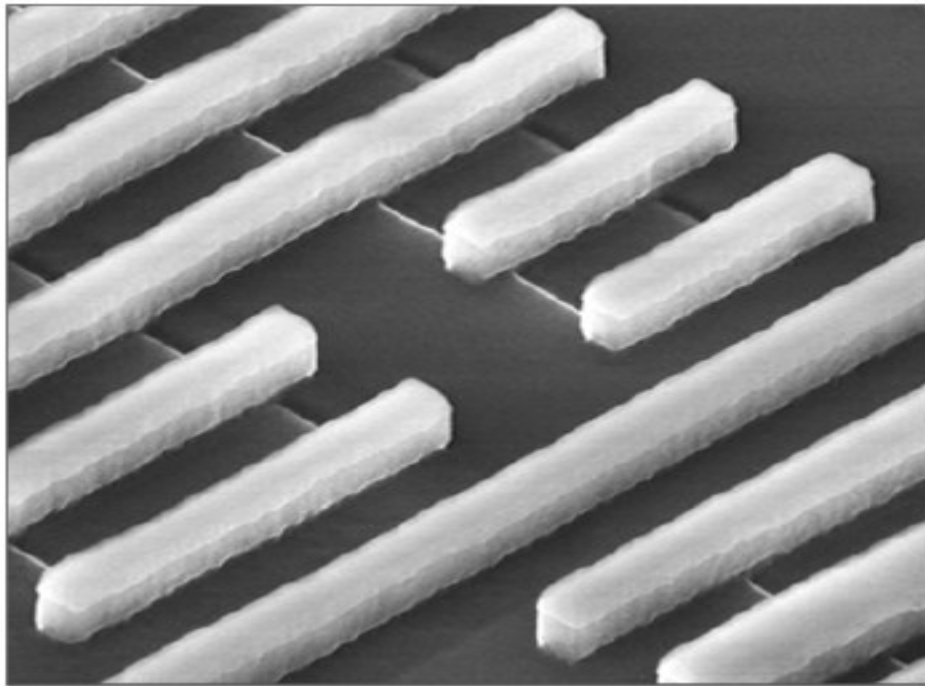
32nm planar transistors (left) and 22nm FinFETs (right)

Performance

- Steeper sub-threshold slope
  - × 10 off-state leakage reduction

Transistor IV characteristics

Thank you!