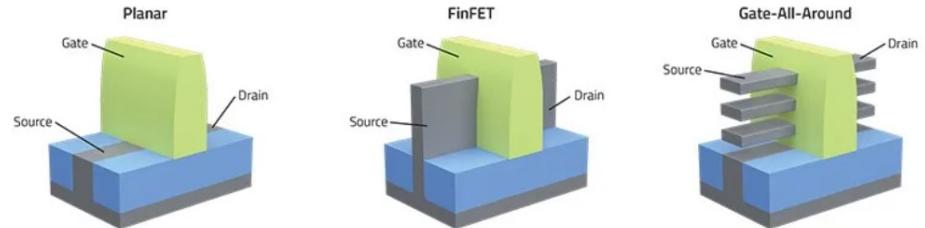


Lecture 10

Microelectronic Devices



- MOSFETs in basic digital circuit elements: inverter
- Performance requirements for MOSFETs in digital electronics and SCALING
- Modern MOSFETs architectures and technologies (FinFETs, FD SOI FETs, strained-channel MOSFETs...)
- Beyond established CMOS technologies: alternative concepts of FETs

Results of the test and feedback: on 21.05 (next week), during the session of exercises, everyone can check the results and consult the corrected copy

$$I_{Dsat} = \frac{W}{2L} \mu_n C_{ox} (V_{GS} - V_T)^2$$

MOSFETs vs BJT

$$I_C \propto \exp \frac{qV_{BE}}{kT}$$

- Output current I_D controlled by voltage
- Main current flow Source to Drain
- Very high input impedance (ideal gate draws zero current)
- Power consumption is low (ideal gate draws zero current)
- Better thermal stability (positive temp. coefficient)
- Linear gain is lower
- unipolar device (single type majority carriers)
- main applications: Switching, digital logic
- Output current I_C controlled by base current
- Main current flow Collector to Emitter
- Moderate input impedance (base draws electric current)
- Power consumption is higher (base draws electric current)
- Worse thermal stability (thermal runaway risk)
- Linear gain is higher
- Bipolar (both electrons and holes involved)
- analog amplification (gain + linearity)

CMOS performance factors

- Performance of CMOS VLSI is measured by:
 - **integration density** (scaling)
 - **switching speed**
 - **power dissipation**

$$P = \frac{CV_{dd}^2}{T} = CV_{dd}^2 f$$

Classic digital switch (before nanoscale era)
only power consumption comes from switching:

$$P = AC_{\text{total}} (V_{dd})^2 f + V_{dd} I_{\text{leak}}$$

In nanoscale CMOS not anymore true:
comparable leakage power!

- Most important logic blocks: CMOS static gates
 - due to their simplicity and noise immunity
- Simplest logic block: **CMOS inverter**

CMOS = Complementary MOS: inverter – key element of logic gates

- For digital electronics MOSFET = switch!

Inverter:

An inverter performs the NOT function: it flips a digital signal ($0 \rightarrow 1$ and $1 \rightarrow 0$)

Inverter is a building block for logic gates

Basic logic circuit using

- 1 x n- and 1 x p-MOSFET

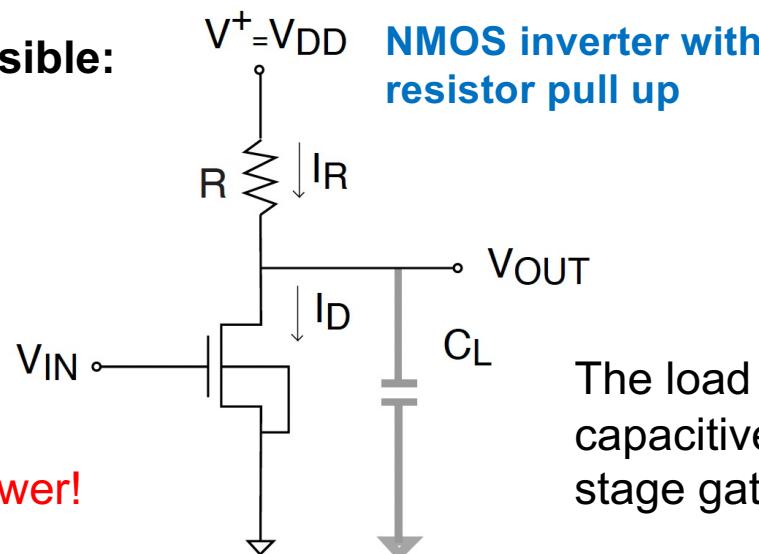
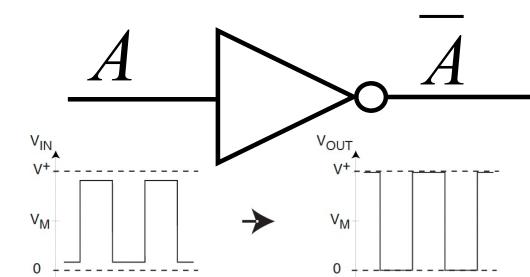
In principle, a simple FET inverter is possible:

if $V_{IN} < V_T$, MOSFET OFF $\Rightarrow V_{OUT} = V_{DD}$

if $V_{IN} > V_T$, MOSFET ON $\Rightarrow V_{OUT}$ small
(value set by resistor/nMOS divider)

The problem of static consumption of power!

Symbol:



NMOS inverter with
resistor pull up

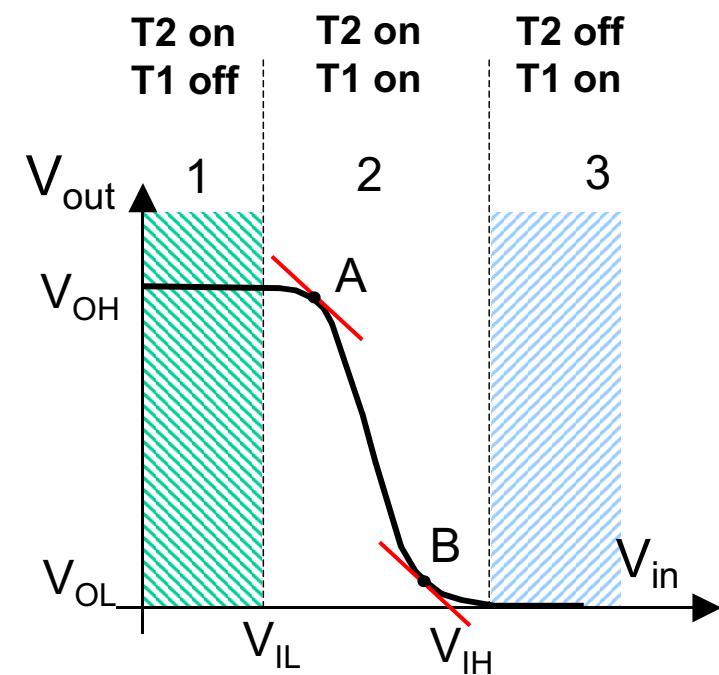
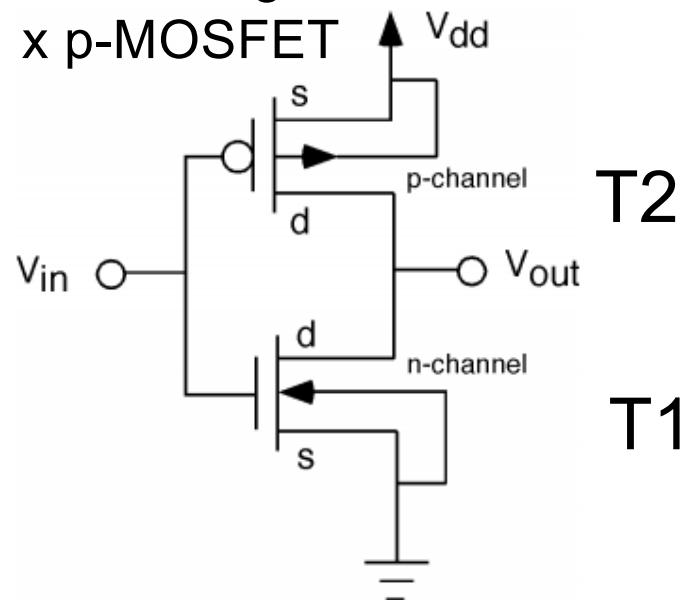
The load C_L is predominantly
capacitive (from following
stage gates)

CMOS = Complementary MOS: inverter

Inverter:

Basic logic circuit using

1 x n- and 1 x p-MOSFET



- CMOS inverter: Consumes power only during switching (dynamic power). In steady-state (logic 0 or 1), no direct path from V_{dd} to ground - very low static power
- CMOS inverters have sharp voltage transfer characteristics with strong logic level separation and high gain around the switching point - better noise performance
- CMOS inverters can drive output fully to V_{dd} or GND (ideal logic levels)
- CMOS offer better speed, better scalability

MOSFET Scaling: speed, efficiency, costs

- More transistors per unit area allow for more complex and powerful integrated circuits (ICs) without increasing chip size
- Shorter channel lengths reduce carrier transit time, increasing switching speed
- Smaller gate capacitance reduces delay and improves dynamic response
- Reduces energy per operation, which is crucial for mobile and battery-powered devices, smaller devices operate at lower supply voltages V_{DD} , reduced dynamic power $P_{dynamic} = \alpha CV_{DD}^2 f$
- Lower costs per function
- Lower global cost

SCALING BENEFITS: *performance, power efficiency, costs*

MOSFET Scaling: challenges

- Short channel effects (threshold voltage roll-off, drain-induced barrier lowering)
- Leakage currents (e.g., subthreshold and gate leakage)
- Quantum mechanical effects (e.g., tunneling through thin gate oxides) Reduces energy per operation, which is crucial for mobile and battery-powered devices, smaller devices operate at lower supply voltages V_{DD} , reduced dynamic power
- Increased variability due to atomic-scale dimensions and process fluctuations
- Overall complexity of technology, equipment and associated cost increase

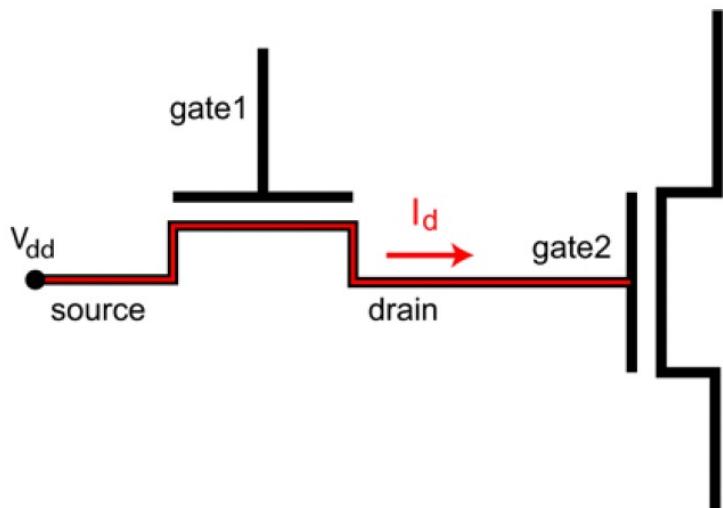
Scaling and speed

On-state

$$\tau = \frac{C_g V_{dd}}{I_d} \propto \frac{L^2}{\mu} \frac{V_{dd}}{(V_{dd} - V_{th})^2}$$

as small as possible

$$I_d \propto \mu \frac{W}{L} (V_{dd} - V_{th})^2 \quad C_g \propto \frac{WL}{t_{ox}}$$



This formula describes the **delay time τ** of a CMOS inverter or logic gate, and how it relates to key MOSFET parameters. It provides a measure of how fast a transistor can switch, which is crucial for digital circuit performance

Delay increases quadratically with channel length L .

τ = delay time

C_g = gate capacitance

V_{dd} = supply voltage

I_d = drain current

L = channel length

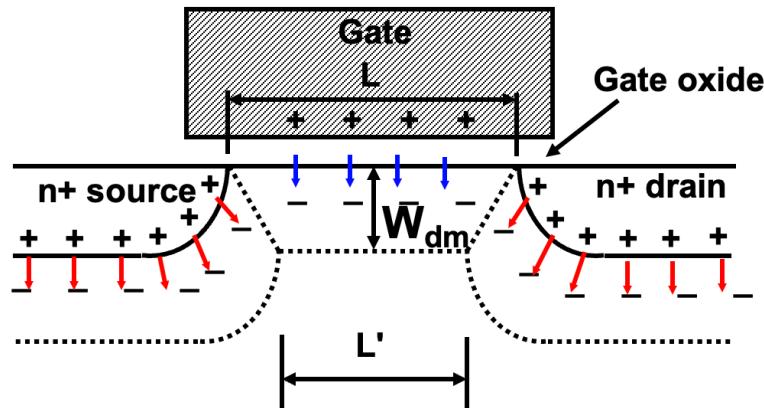
μ = carrier mobility

V_{th} = threshold voltage

Implications:

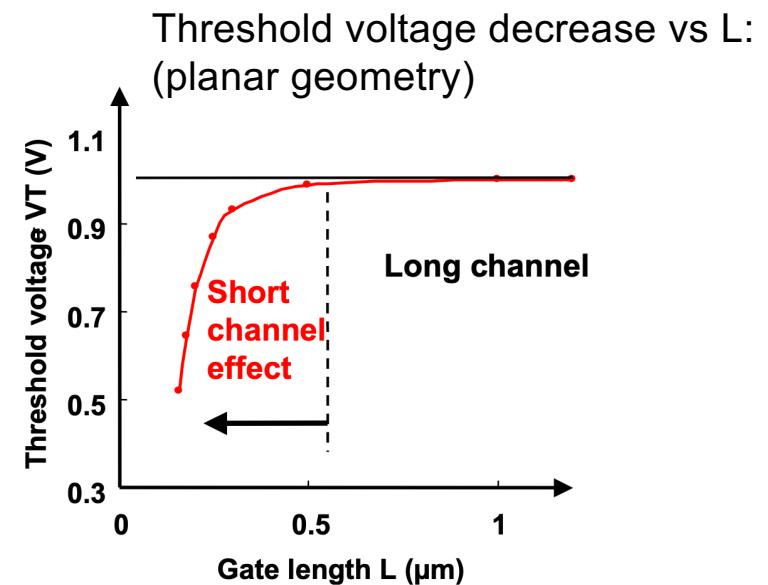
- Smaller devices (shorter L) and higher μ result in faster circuits.
- Lowering V_{dd} to save power increases τ if V_{dd} approaches V_{th} too closely.
- There is a trade-off between power and speed.

Scaling, short channel effects and leakage



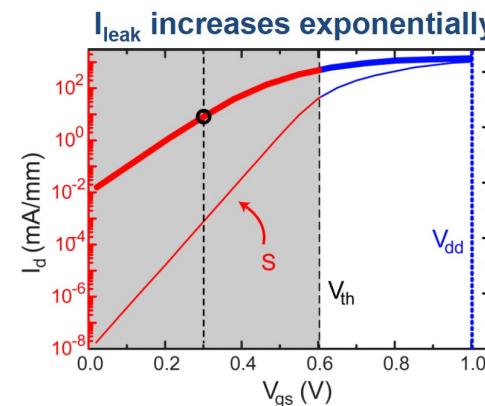
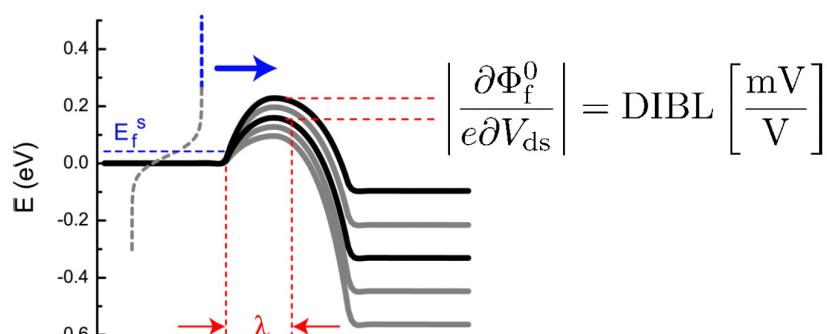
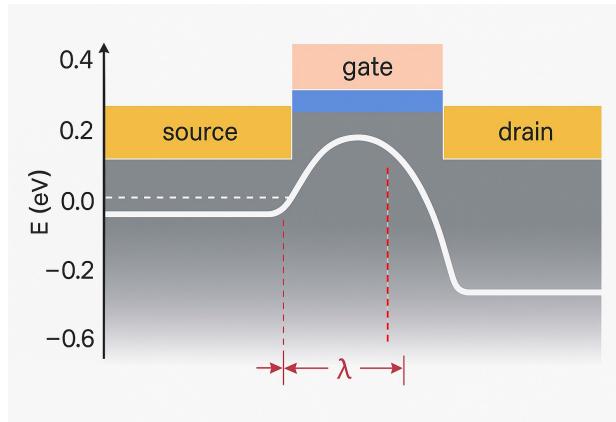
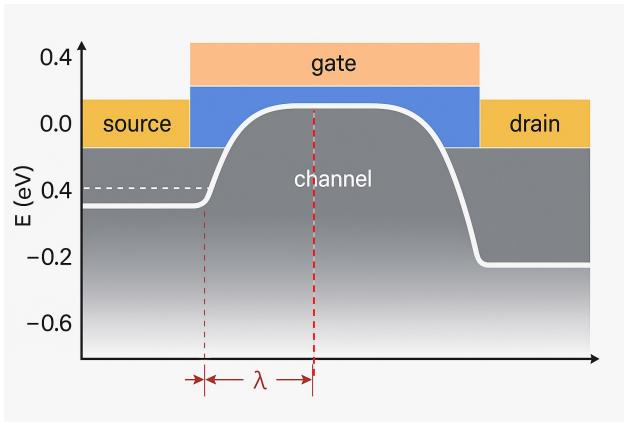
- **V_T roll-off and its impact:**
 - threshold voltage reduces with L reduction
 - I_{off} increases (subthreshold slope is not scalable: limit is 60mV/decade!)
 - Reduced control of the gate on the channel

To some extent, the problem can be addressed by using the so-called retrograde doping (retrograde well) – a special doping profile of the channel with a concentration gradient from the surface to the bottom



Scaling and short channel effects

Conduction band profile of a conventional MOSFET



DIBL = „Drain – Induced – Barrier - Lowering“

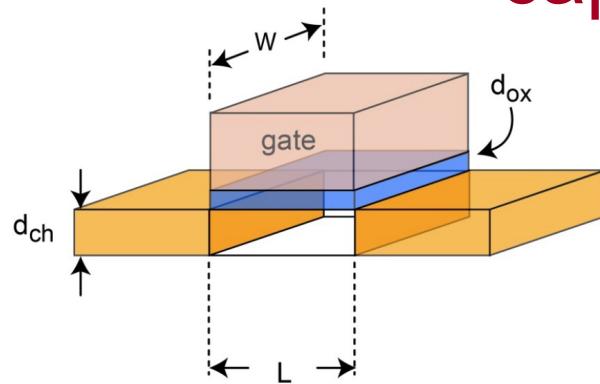
Aggressive scaling:

- barriers are overlapping
- barrier lowering
- off current increases exponentially

The exponential increase of leakage current:
Power consumption increase

Geometrical factors are important

Scaling: geometrical factors to be considered, capacitance ratio



$$C_{\text{ox}} \approx 10 \times C_d$$

$$C_{\text{ox}} = \epsilon_0 \epsilon_{\text{ox}} \frac{W \cdot L}{d_{\text{ox}}}$$

$$C_d = \epsilon_0 \epsilon_{\text{ch}} \frac{W \cdot d_{\text{ch}}}{L}$$

$$\epsilon_0 \epsilon_{\text{ox}} \frac{W \cdot L}{d_{\text{ox}}} \approx 10 \times \epsilon_0 \epsilon_{\text{ch}} \frac{W \cdot d_{\text{ch}}}{L}$$

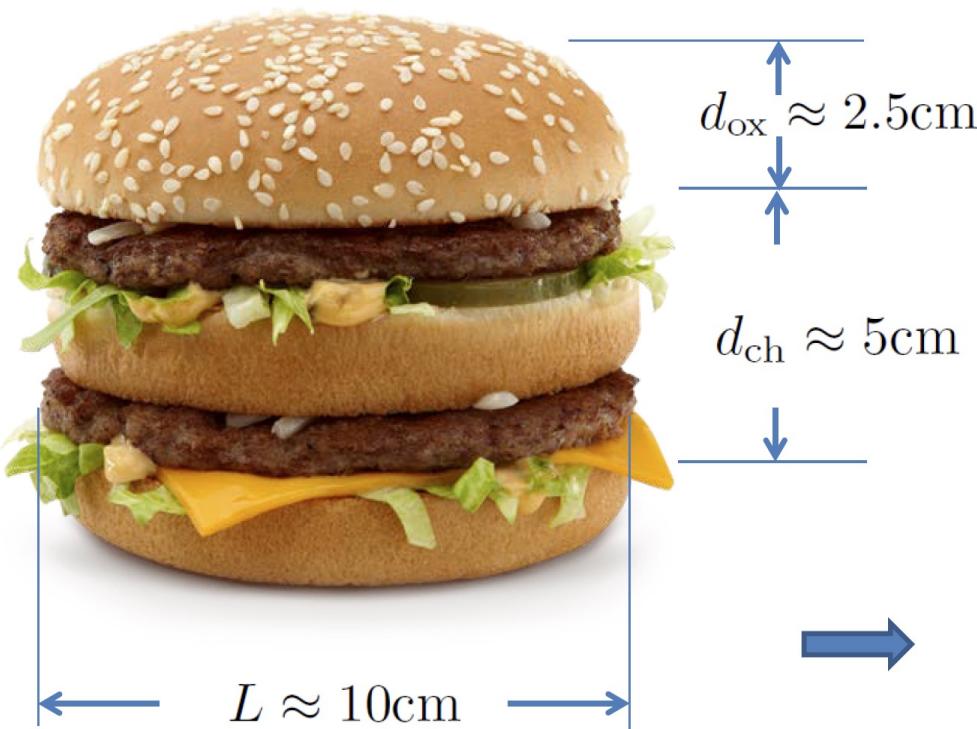
Estimation:

$$L^2 \approx 10 \times \frac{\epsilon_{\text{ch}}}{\epsilon_{\text{ox}}} d_{\text{ox}} d_{\text{ch}} = 10 \times \lambda^2$$

$$L > 3 \times \lambda$$

$$\lambda = \sqrt{\frac{\epsilon_{\text{ch}}}{\epsilon_{\text{ox}}} d_{\text{ox}} d_{\text{ch}}}$$

Scaling: geometrical factors to be considered, capacitance ratio, example from everyday life



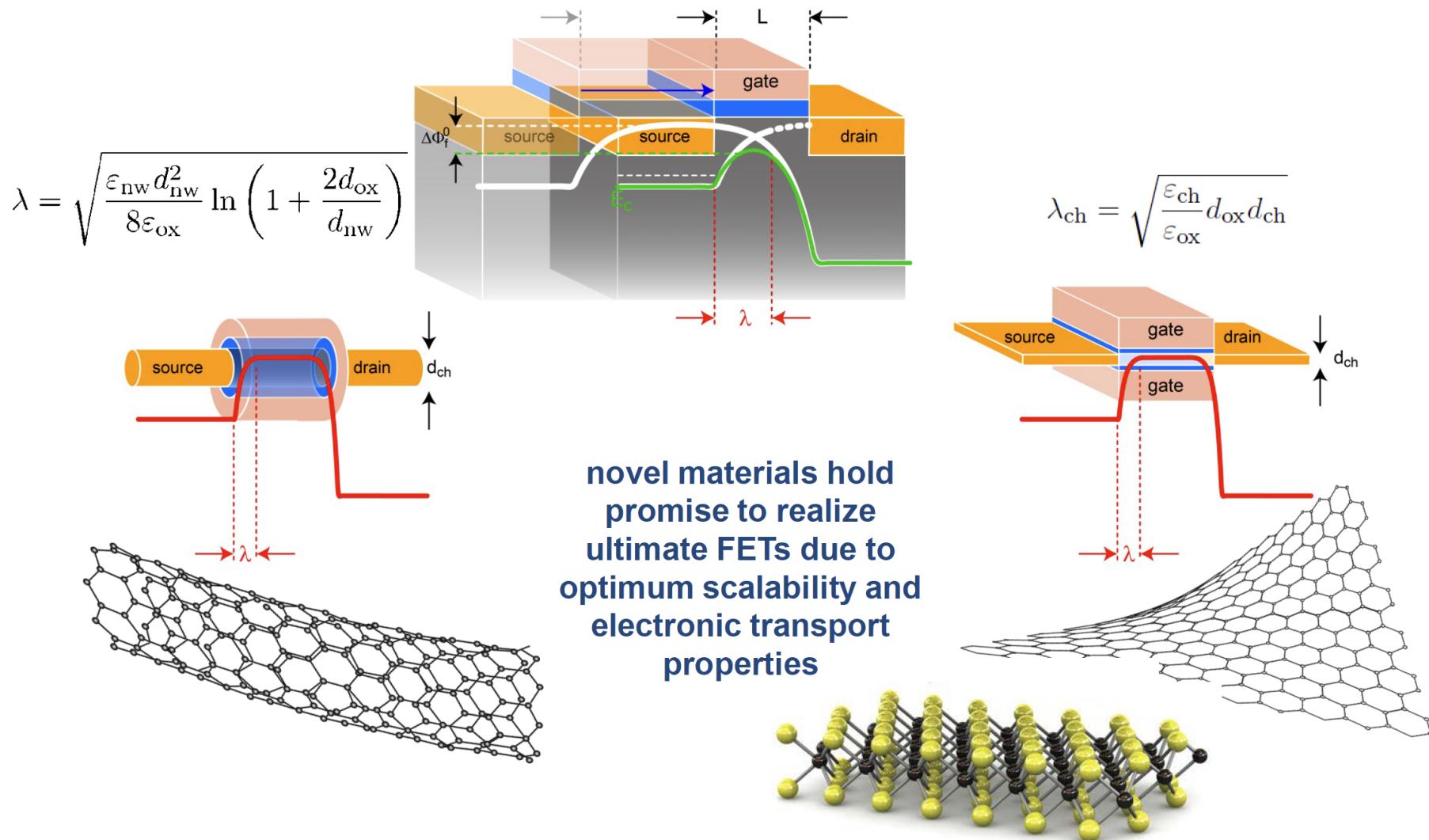
$$\lambda = \sqrt{\frac{\epsilon_{ch}}{\epsilon_{ox}} d_{ox} d_{ch}} \approx 7.5\text{cm}$$

Leaky device!

$$L \approx \lambda$$

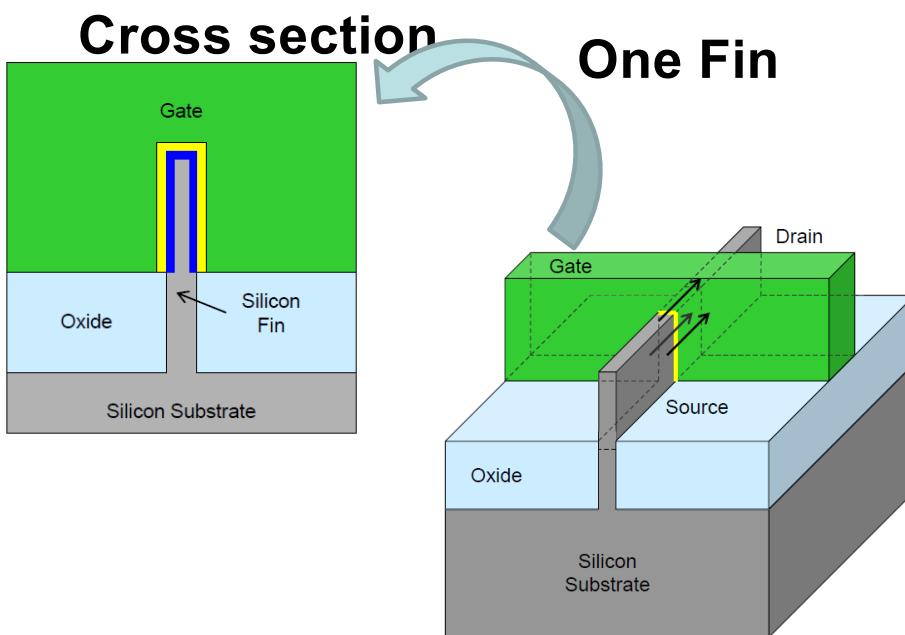
a leaky device

Scaling: the geometrical factor: quest for an ultra-thin channel

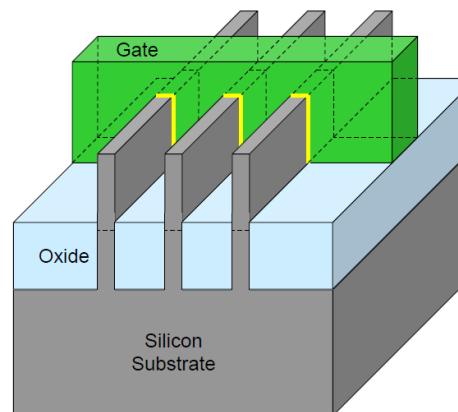


Fin FETs

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



Multiple Fins
for higher drive and
performance



Fin: Thin vertical silicon body where current flows (channel)

Gate: Surrounds the fin on three sides for better control

Source/Drain: Regions at both ends of the fin

Oxide layer: Thin insulator separating gate from channel

FinFETs

- Just like planar MOSFETs, FinFETs operate by applying a voltage to the gate to invert the channel and allow current to flow between source and drain.

However, due to the 3D structure FinFET offers:

- Better electrostatic control of the channel
- Reduced subthreshold leakage
- Higher on/off ratio
- In particular, 3D gate control significantly reduces short channel effects; high drive current results in a better performance (speed) per area; lower leakage helps keeping static power dissipation to minimum
- **Scalable below 10nm**
- https://en.wikipedia.org/wiki/3_nm_process

Fin FETs process challenges

Spacer

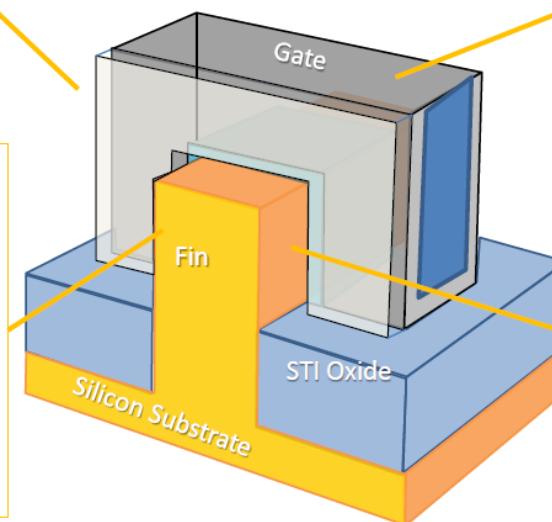
- Complete spacer removal from fin area

Gate Stack (high-k & metal gate)

- Material selectivity
- Material deposition thickness uniformity on vertical walls
- Metal gate composition uniformity/stability

Fin Formation:

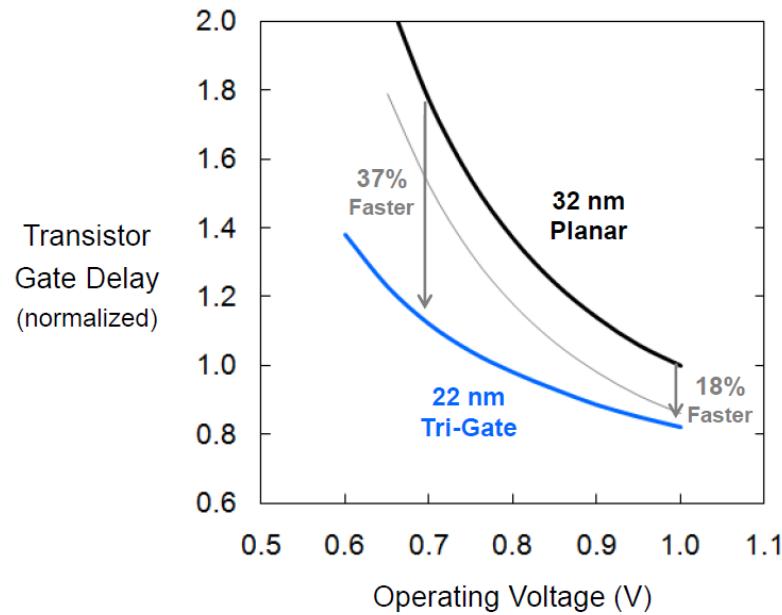
- Precision etch
- Structural integrity (collapse, erosion, thermal shock)
- Precise Recess to control fin height
- Channel materials to increase mobility



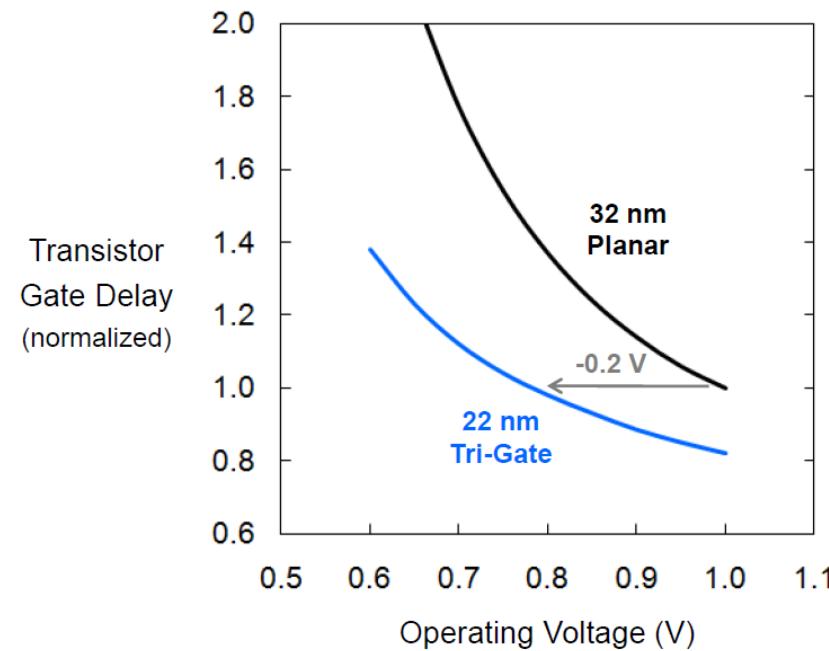
Fin Junctions:

- Conformal doping on sidewalls

FinFET performance benefits



Improved speed performance.



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

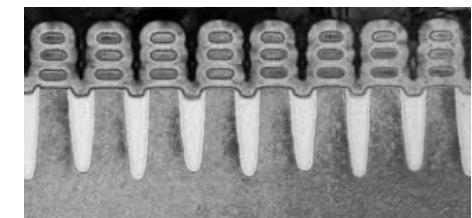
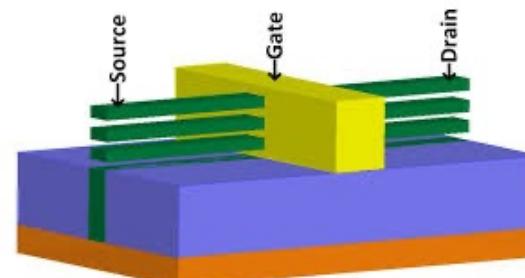
FinFETs: a mainstream technology

FinFETs are used in all major commercial chips at 22 nm and below:

- TSMC, Intel Samsung, GlobalFoundries: Use FinFETs at 16/14 nm, 10 nm, 7 nm, 5 nm and below
- FinFET is standard in smartphones, CPUs, GPUs, FPGAs, etc.
- **Gate-all-around (GAA) FETs are the next step beyond FinFET, offering even better control**

The channel is formed by **nanosheets** (wider) or **nanowires** (narrower), suspended between source and drain

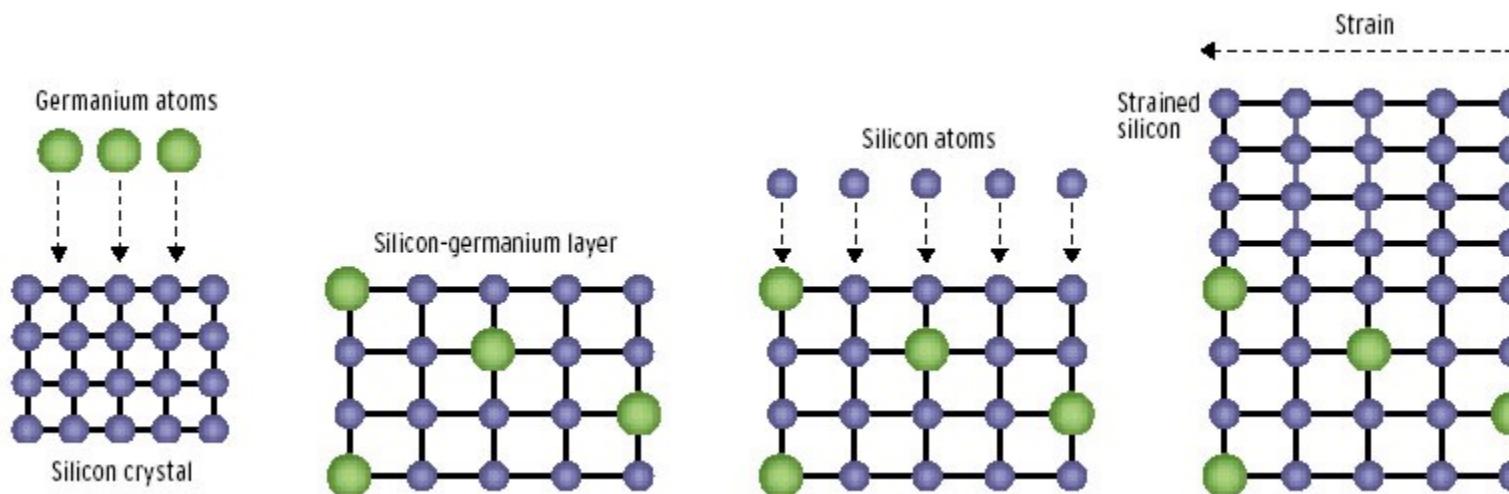
The **gate wraps around all 4 sides**, hence "Gate-All-Around" widely accepted by key players for 3nm node and beyond



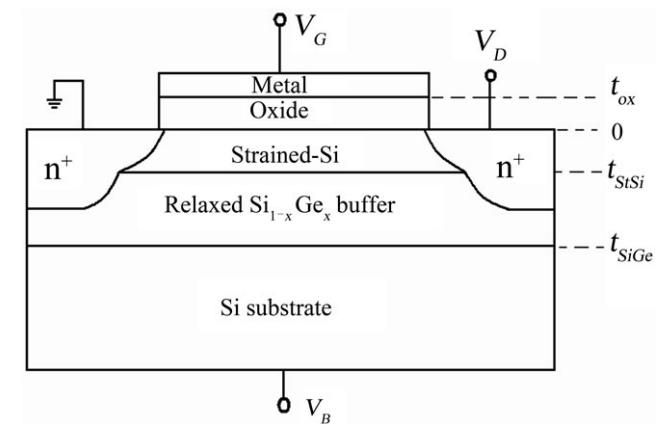
Strained Si Channels (via SiGe integration)

- Tensile strain in silicon (usually achieved by growing Si on a relaxed SiGe buffer) stretches the Si lattice- tensile strain
- Increases electron mobility by reducing effective mass and scattering
- Enhances hole mobility when compressive strain is introduced, often via SiGe in PMOS channels
- Mobility enhancements: electrons: up to $2 \times$ higher mobility; holes: up to $3 \times$ with compressively strained SiGe channels (vs. unstrained Si)
- Improved mobility translates to higher current for the same gate voltage - **speed**
- Higher mobility allows operation at lower supply voltages – V_{dd} – **lower power**

Strained Si Channels (via SiGe integration)

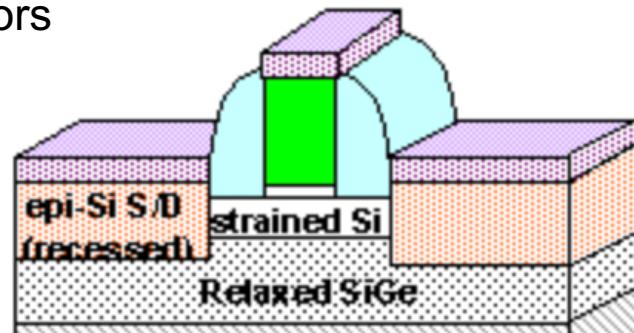


- (A) Ge is introduced into the lattice
 - $a_{\text{SiGe}} > a_{\text{Si}}$
- (B) Si is deposited on top of the SiGe.
 - Atoms align causing a strain in the lattice.

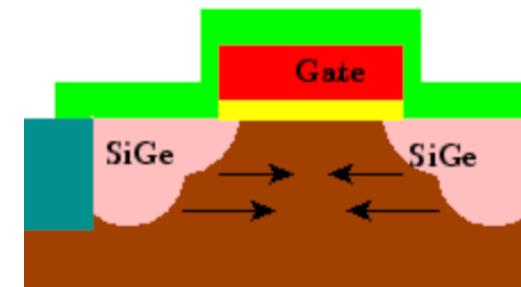


Strained Si Channels

- **Tensile Strain in Silicon Channel (for NMOS)**
- Si layer grown on a relaxed SiGe substrate or buffer
- When grown on relaxed SiGe, the Si layer stretches laterally to match the lattice of the underlying SiGe
- This introduces biaxial tensile strain in the Si film
- Enhanced electron mobility → used for NMOS transistors

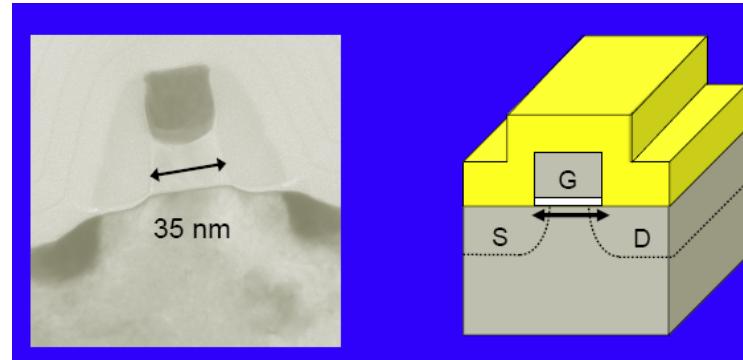


- **Compressive Strain in Channel (for PMOS)**
- Standard silicon substrate and silicon channel.
- Etch cavities in the source/drain regions of PMOS devices.
- Epitaxially grow SiGe into these cavities (called embedded SiGe)
- During cooling, lattice mismatch and thermal expansion differences generate compressive strain in the adjacent silicon channel
- The presence of SiGe in the source/drain regions pulls laterally on the silicon channel, inducing compressive uniaxial strain along the channel length

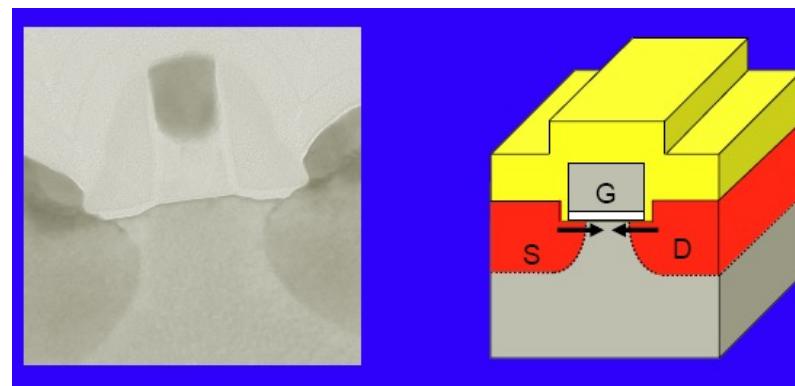


There are also other ways of strain engineering like cap TiN layers

Different strain technology: NMOS & PMOS strain technology



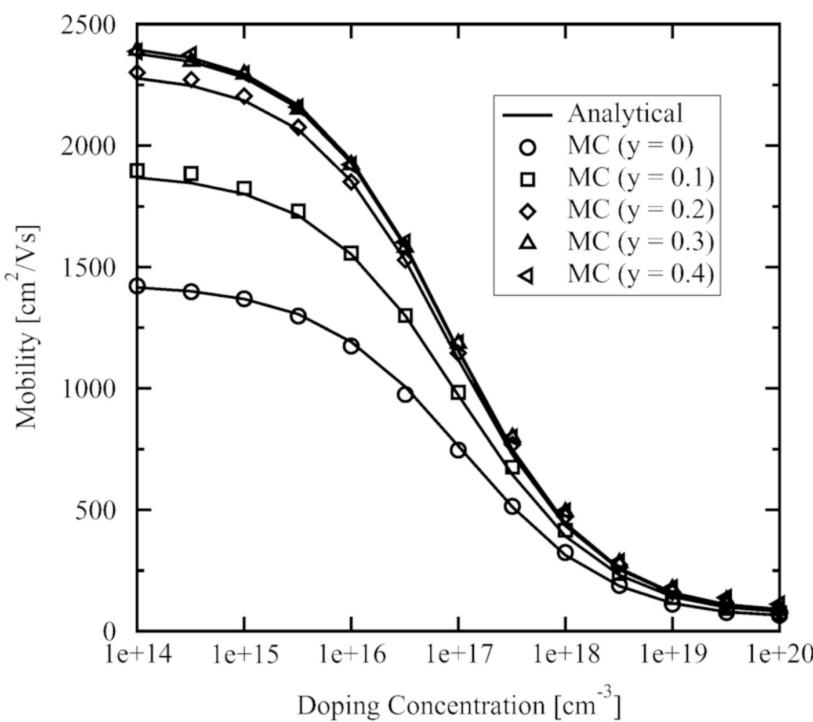
NMOS: Tensile Tensile SiN capping films for enhanced for enhanced uniaxial strain



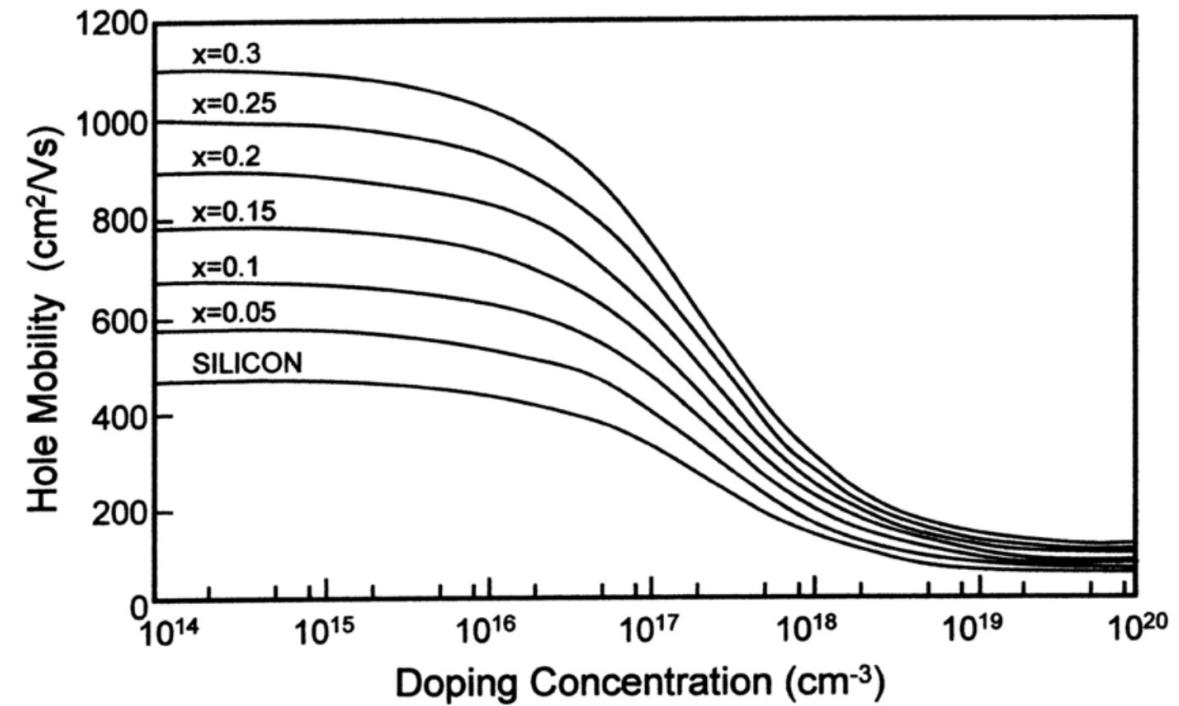
PMOS: Epitaxial SiGe film embedded into S/D

Strained Si Channels: electron and hole mobility

Electron mobility



Hole mobility

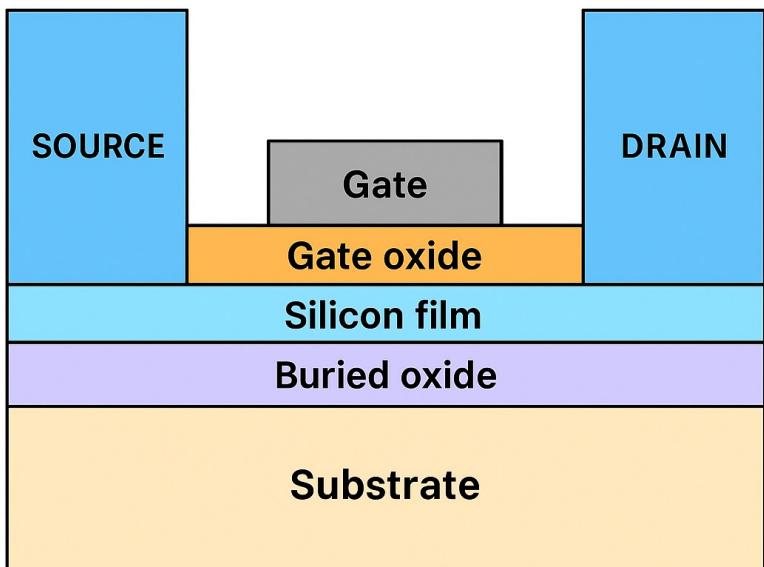


The use of strained-channel technology

- Strained-channel MOSFETs have been used in commercial semiconductor manufacturing since the 90 nm node
- Intel introduced strained silicon in its 90 nm "Prescott" Pentium 4 processors (2003)
- Later generations (45 nm, 32 nm, 22 nm, 14 nm, etc.) used sophisticated strain techniques
- Strained channels are used in high-performance logic to boost electron (nMOS) and hole (pMOS) mobility
- Smartphones, tablets: Qualcomm, Apple, and Samsung use strained-channel MOSFETs in their SoCs (System-on-Chip) to improve performance/power efficiency
- Though FinFETs (introduced at ~22 nm) reduce short-channel effects without requiring strain, strain engineering is still used to enhance carrier mobility within the fins
- Techniques like selective epitaxial growth of SiGe (for p-FinFET) and Si:C or tensile nitride liners (for n-FinFET) are commonly applied
- **Strained-channel MOSFETs have been mainstream in commercial semiconductor manufacturing since the 90 nm node. They are still used today in advanced logic and mobile chips — often combined with FinFET or SOI architectures to further improve performance**

Silicon-On-Insulator (SOI) CMOS:

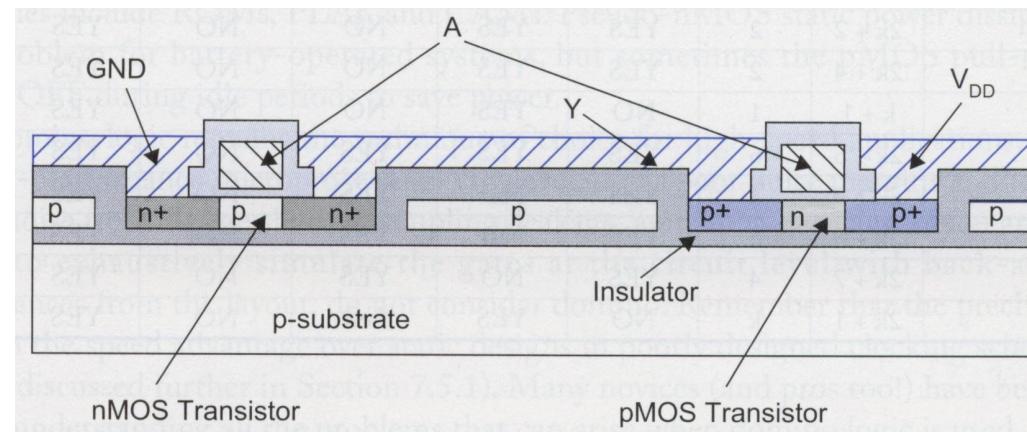
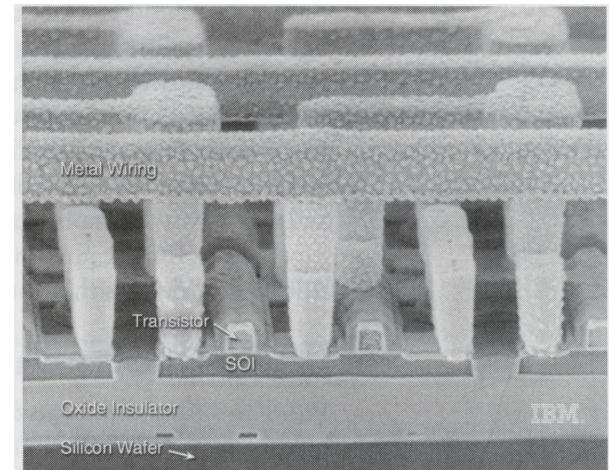
SOI MOSFET



- a thin layer of silicon is isolated from the substrate by a buried oxide (BOX) layer
- significantly reduces junction capacitance, improving switching speed and lowering dynamic power consumption
- SOI devices are less affected by short-channel effects like drain-induced barrier lowering (DIBL)
- easier scaling for smaller nodes
- Due to the reduced junction area and full isolation from the substrate, SOI MOSFETs exhibit lower off-state leakage currents
- Especially in fully-depleted SOI (FD-SOI) devices, the subthreshold slope is steeper, allowing for better threshold voltage control and low-voltage operation
- The reduced capacitance and improved electrostatics enable higher-speed digital and RF circuits

Silicon-On-Insulator (SOI) CMOS

- Adopted for IBM PowerPC microprocessors in 1998
 - Higher performance and lower power than CMOS
 - Higher cost and complicated circuit design
- Differences from bulk CMOS:
 - Transistor source, drain & body surrounded by insulating SiO_2
 - Eliminates most diffusion parasitic C
 - Body no longer tied to GND or V_D

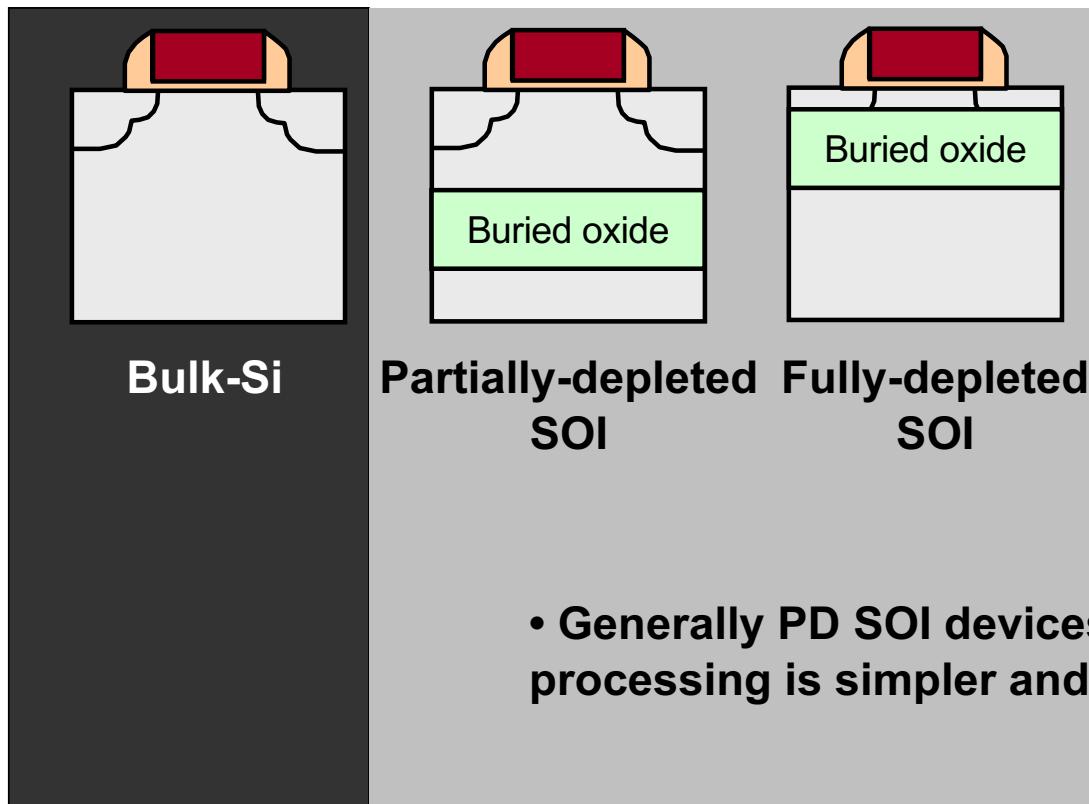


Two types of SOI MOSFET

- **Partially depleted (PD)**
 - Body thicker than channel depletion width
 - Body voltage can change
 - Depending on charge injected into bulk
 - Causes *history effect*, which changes V_t
- **Fully depleted (FD)**
 - Body thinner than channel depletion width
 - Fixed body charge
 - Body voltage does not change
 - Thin body makes this very hard to manufacture
 - Therefore, the use is limited

SOI MOSFET: partially depleted vs full-depleted

The difference between Partially Depleted (PD) SOI and Fully Depleted (FD) SOI MOSFETs lies in the thickness of the silicon layer above the buried oxide and how completely it can be depleted



PD SOI

- Top Si $> 40\text{nm}$
- Only part is depleted
- Floating body effects
- some short channel effects (better than bulk)

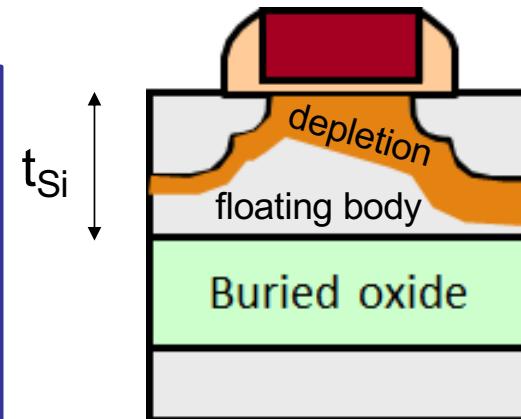
FD SOI

- Top Si $< 20\text{nm}$
- fully depleted
- Floating body effects negligible
- weaker short channel effects (better than bulk)

Partially depleted (PD) SOI MOSFET

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

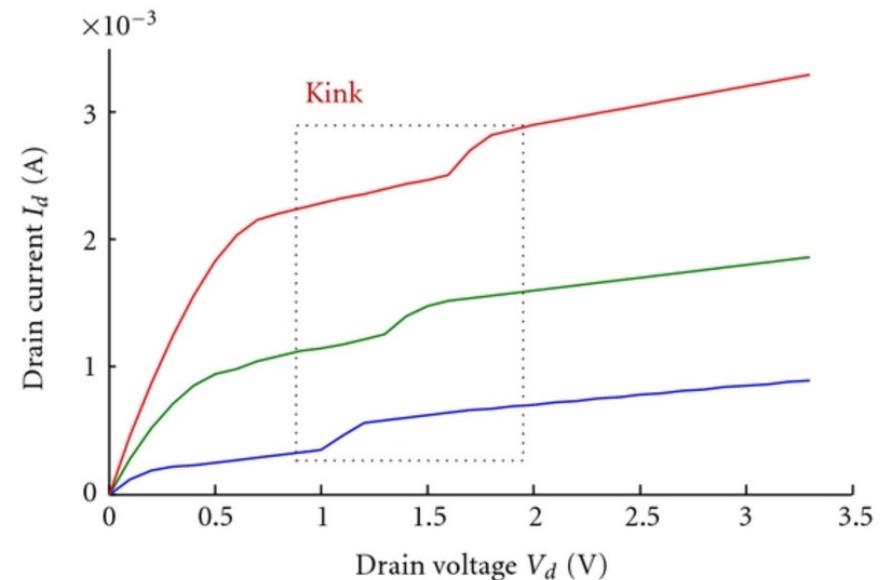
- + similar to bulk-Si
- + no coupling
- Short channel effects (SCE): as in bulk
- floating body
- Kink effect
- dynamic over- and under-shoots
- self-heating effects



- 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**

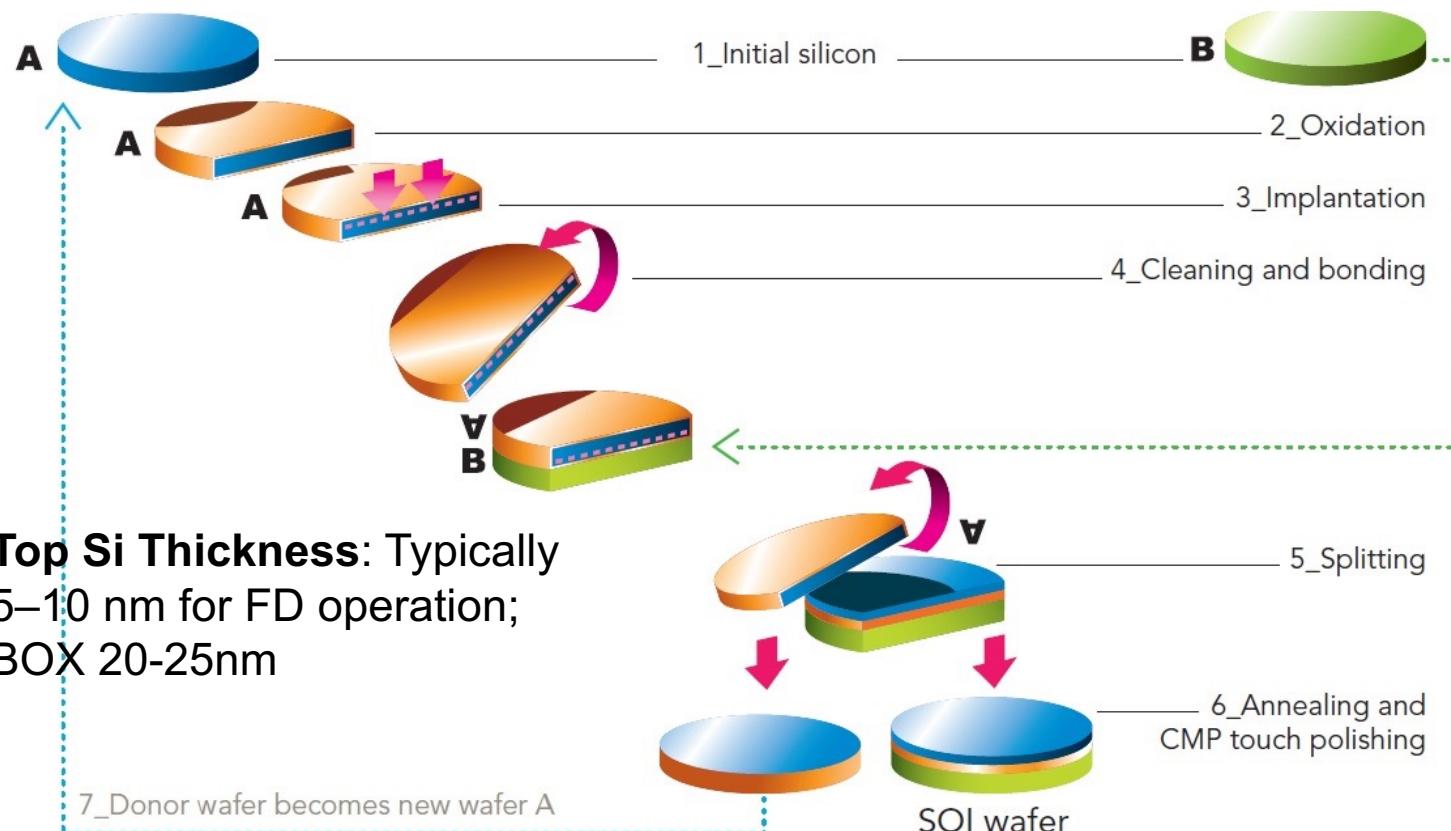
Partially Depleted (PD) SOI MOSFET: floating body effect

- The kink effect is a phenomenon observed in PD-SOI MOSFETs, particularly at high drain voltages, and is a direct consequence of the floating body in these devices
- In PD-SOI devices, the body (substrate region under the channel) is not grounded and is electrically floating (not tied to a fixed potential)
- High electric field near the drain - electrons gain energy and ionize silicon atoms
- This **impact ionization** generates electron-hole pairs
- Electrons are swept into the drain
- Holes accumulate in the body because they cannot escape easily (the body is floating)
- The body potential rises, forward-biasing the body-source junction
- This effectively lowers the threshold voltage, then I_D jumps



In FD SOI the ultra-thin body is fully depleted by the gate field during operation
There is no neutral region to accumulate charge carriers, no body potential build-up, therefore there is no kink

Smartcut process for Silicon On Insulator substrates © SOITEC



Invented by CEA-LETI & SOITEC, Bernin, France

- A donor wafer (with high-quality silicon) is implanted with hydrogen ions at a precise depth
- The donor wafer is bonded to a handle wafer (also silicon), which already has a thermal oxide grown on its surface to serve as the buried oxide (BOX) layer
- Upon thermal annealing, the implanted hydrogen forms microbubbles, causing the donor wafer to split at the implantation depth, transferring the thin silicon layer to the handle wafer
- The transferred silicon layer is often rough and is polished (using chemical mechanical polishing, CMP) to improve surface quality
- The remaining part of the donor wafer can be reused after polishing

SOI Wafers: state of the art in thin film SOI

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.

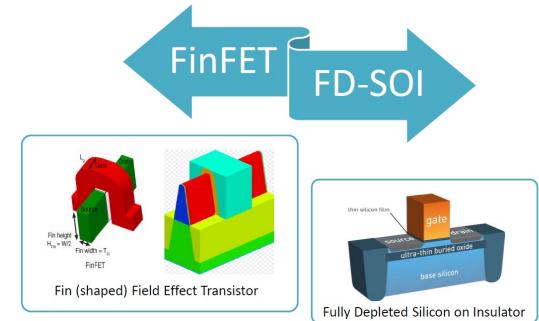


<http://www.soitec.com/en/products-and-services/microelectronics/fd-2d/>

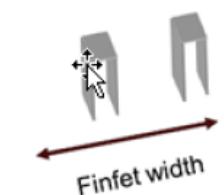
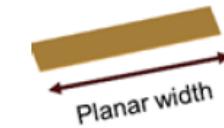
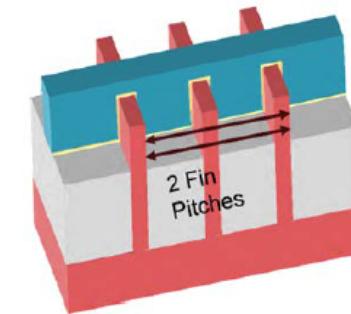
FD-SOI is **actively used and commercially viable**, especially where **power efficiency, analog/RF performance, and cost-effectiveness** are critical. It is not replacing FinFETs at leading-edge nodes but complements them in different segments.

Fin FET versus FD SOI MOSFET

The only two mainstream industrial technology options for sub-10 nm scaling are FinFETs and FD SOI MOSFETs, followed by 1D/2D transistors (these last ones are under research).

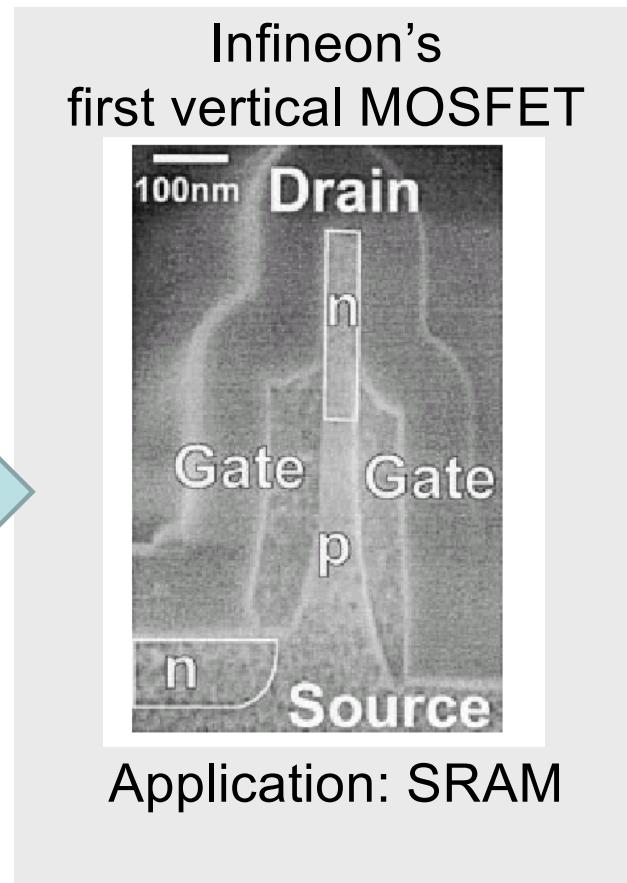
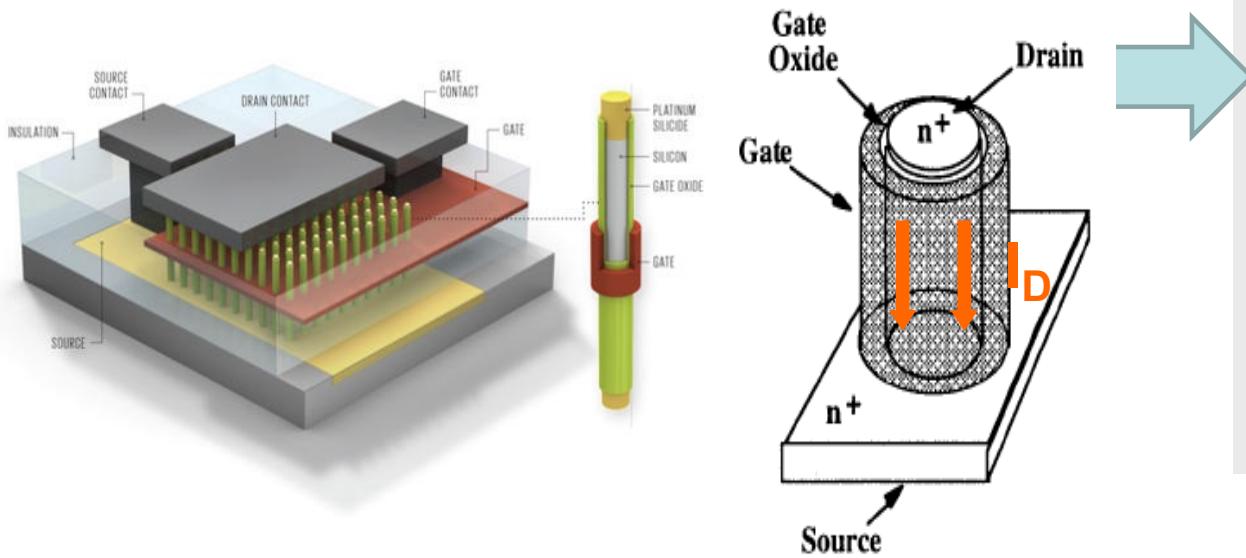


Category	FinFet	FD-SOI
Base Wafer Cost		x4
Process Complexity	↑	↓
Overall Wafer Cost	→ to ↑	→ To ↓
Die Yields	??????	??????
Unit Cost	??????	??????
Process control & metrology challenges	↑	↓
Active transistor area density	↑	↓
Performance (Ion vs. Ioff)	Similar	Similar



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 (very suitable for memory)



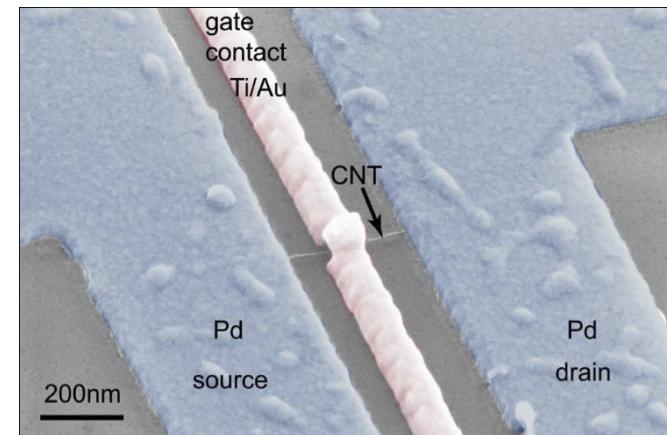
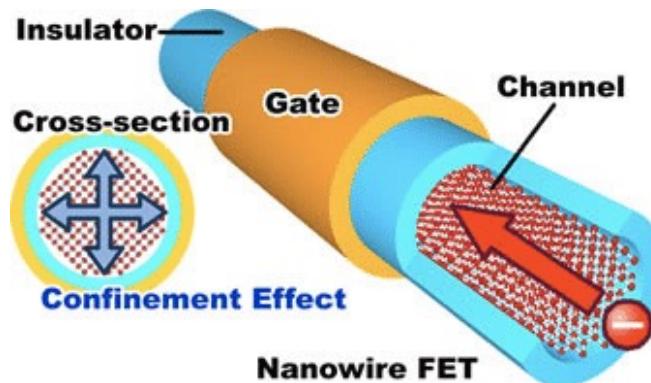
1D (nanowire) Gate-All-Around FETs

Why **semiconducting nanowires (NWs)** and **carbon nanotubes (CNTs)**:

- due to their **potential to test fundamental concepts** about dimensionality
- can serve as basic **building blocks of emerging technology platforms**
- enable **new integrated functionality** in highly dense low cost integrated circuits
- use advanced existing fabrication schemes: **Si-toolset** with 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

Nanowires are used by the key players (3nm node)

CNT – only R&R



Semiconductor manufacturing facilities (fab): Estimated Costs and Construction Times by Node

Technology Node	Estimated Cost (USD)	Estimated Time to Build	Notes
22 nm	\$5–7 billion	~2–3 years	Mature node, many foundries already operational.
10 nm	\$7–9 billion	~2.5–3.5 years	Requires advanced lithography (193i with multiple patterning).
7 nm	\$10–12 billion	~3–4 years	EUV may begin to be used; very tight tolerances.
5 nm	\$15–20 billion	~4–5 years	Full EUV integration; extreme control required.
3 nm	\$18–25+ billion	~5 years or more	GAA FETs, ultra-complex infrastructure, cutting-edge EUV.

Hi-k dielectric gates in MOSFETs

Problems with ultra-thin SiO_2 :

- Tunneling leakage currents (quantum mechanical tunneling through thin oxide)
- Increased power consumption
- Reduced device reliability

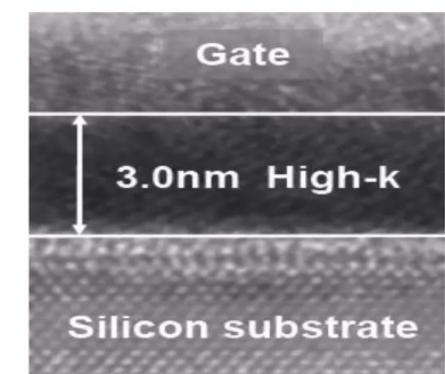
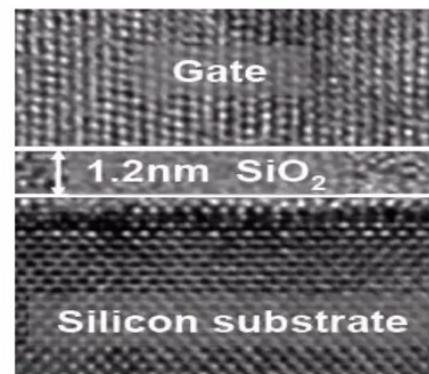
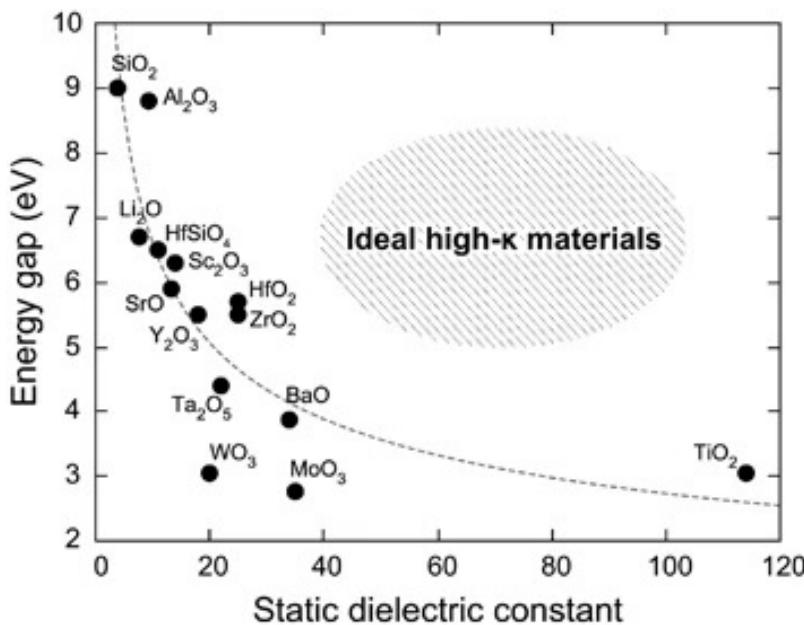
To solve this, high-k materials such as HfO_2 (hafnium dioxide) replaced SiO_2

These materials:

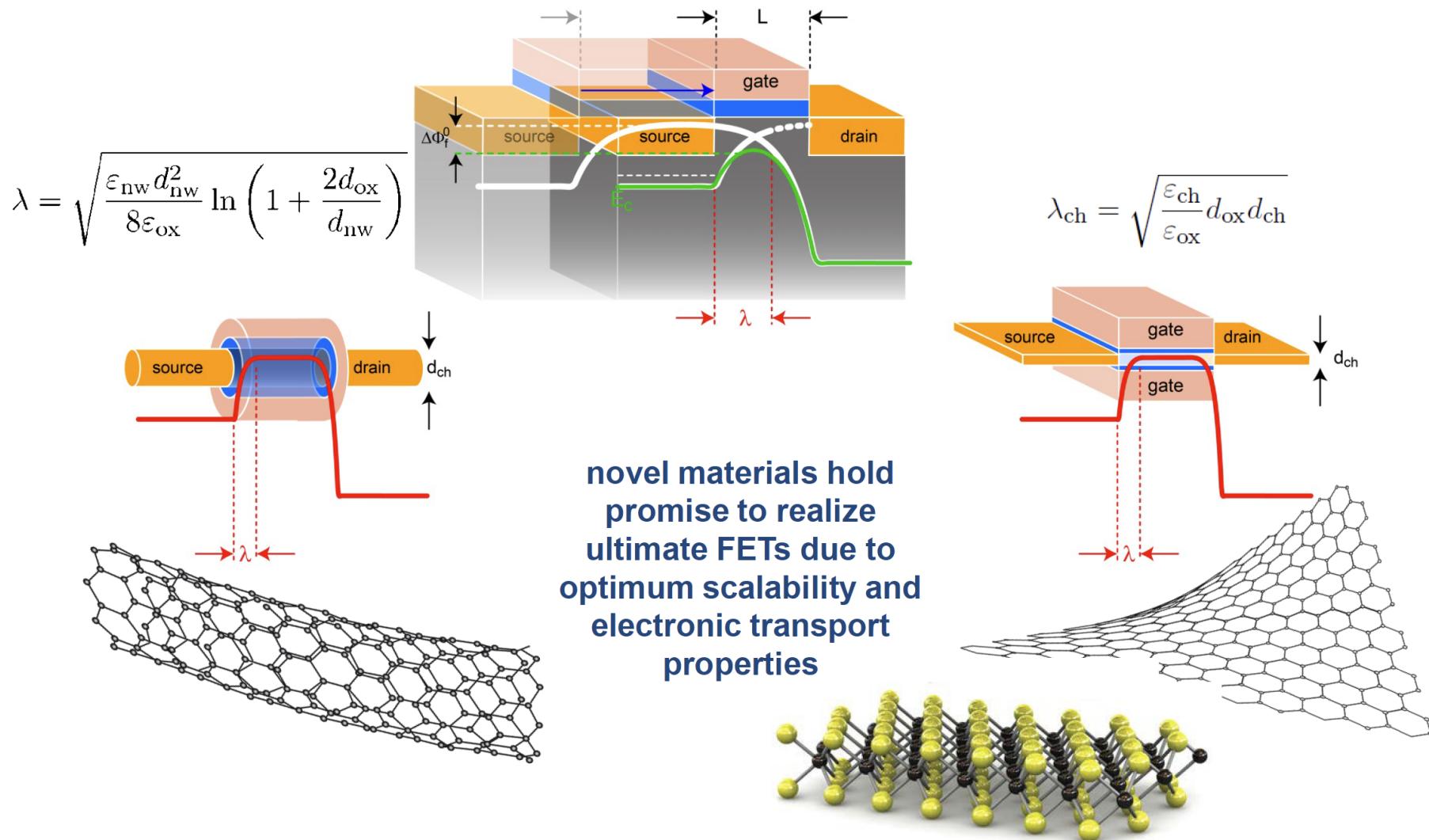
- Have a higher dielectric constant than SiO_2 ($\text{HfO}_2 \approx 20\text{--}25$ vs. $\text{SiO}_2 \approx 3.9$)
- Allow for physically thicker dielectric layers while maintaining same equivalent oxide thickness (EOT):

$$\text{EOT} = \frac{t_{\text{high-k}} \cdot \varepsilon_{\text{SiO}_2}}{\varepsilon_{\text{high-k}}}$$

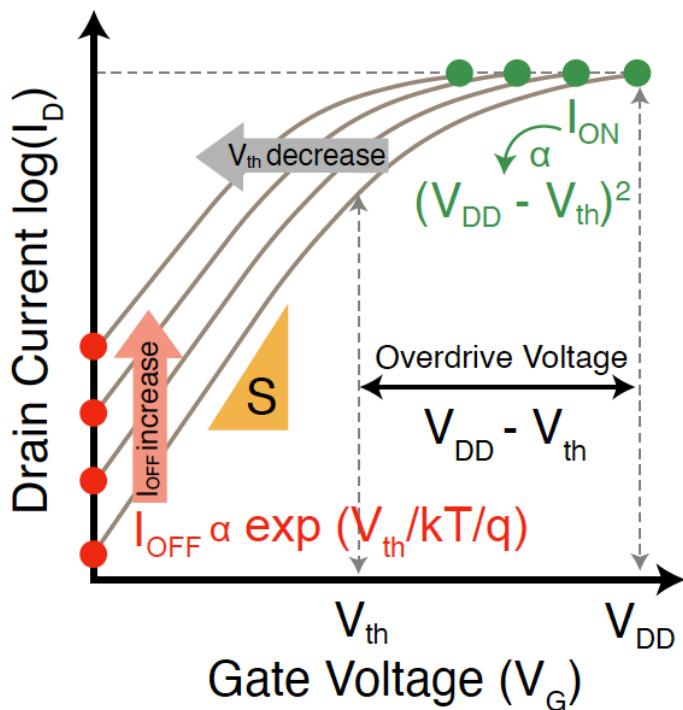
This reduces gate leakage current without sacrificing gate control over the channel.



Scaling: the geometrical factor: quest for an ultra-thin channel



Subthreshold slope: a fundamental limitation of MOSFETs



$$SS = \frac{\partial V_{GS}}{\partial \log I_{DS}} \geq 60 \text{ mV/dec}$$

There is a fundamental limitation for the subthreshold slope, which is determined by Boltzmann's statistics (60= mV/dec is calculated at the room temperature)

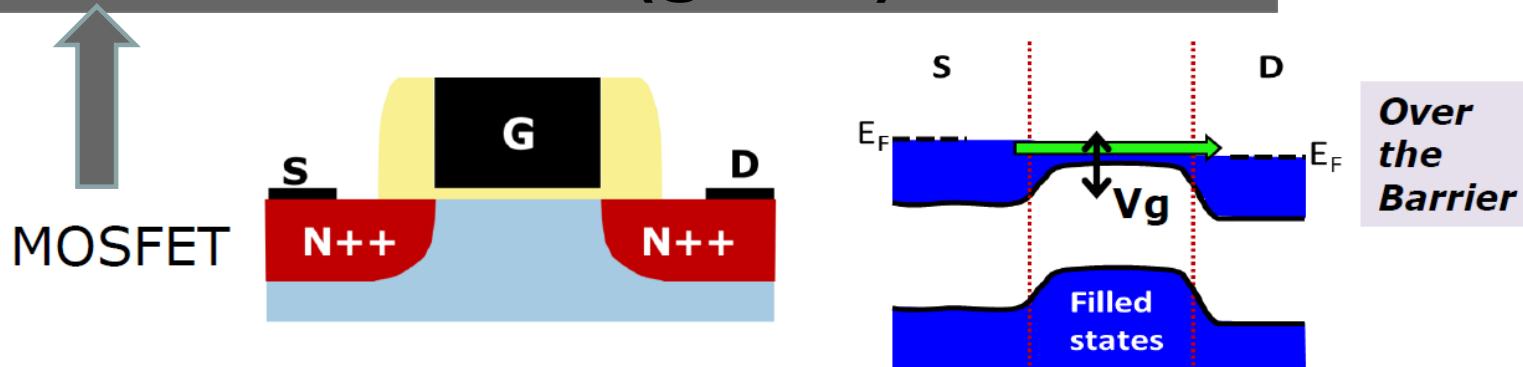
An ideal switch must have a steep slope (i.e. smallest possible voltage per current decade), the "Boltzmann's tyranny" is an important limitation for modern MOSFETs

This limitation cannot be overcome by any technological improvements without changing the physical principles

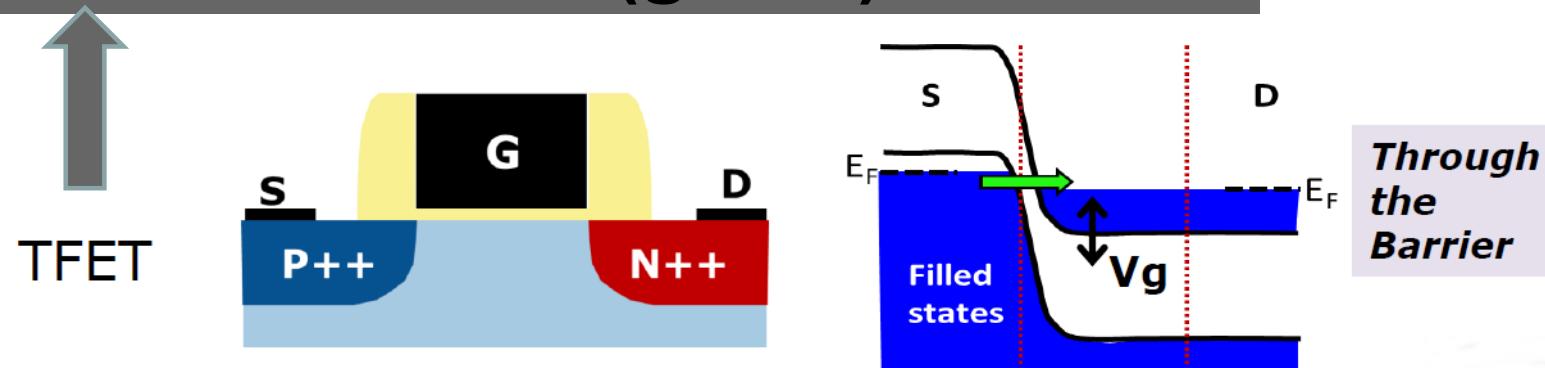
New device physics needed (only research so far...) 39

Towards steeper slope (FETs with different principles): Tunnel FET versus MOSFET

Architecture: N+ / (gated) I / N+



Architecture: P+ / (gated) I / N+



Summary

- Principal MOSFET architectures for scaling towards advanced technological nodes: FinFETs, FD SOI FETs
- Important performance boosters adopted by the industry: hi-k semiconductors, strain engineering
- A trade-off between power dissipation and speed is a key consideration for FET design and choice of parameters
- Alternative materials are under research (1D, 2D)
- Different physical principles are needed to overcome “Boltzmann’s tyranny”