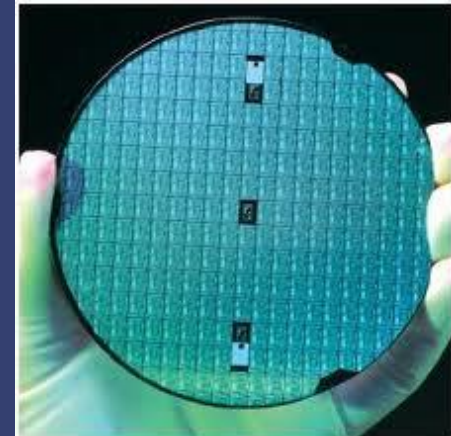
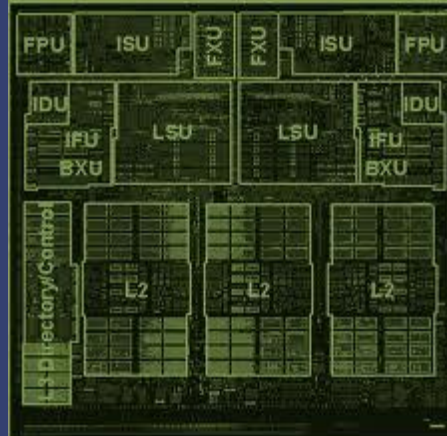
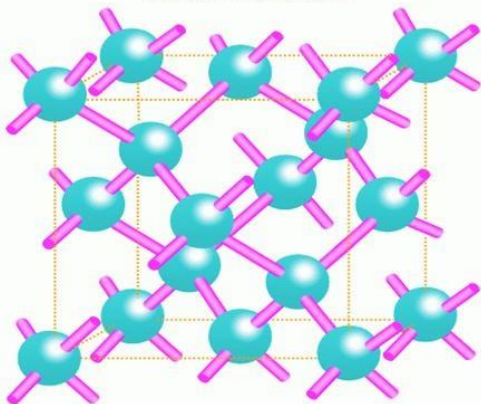


Structure of silicon crystal



ECE 225

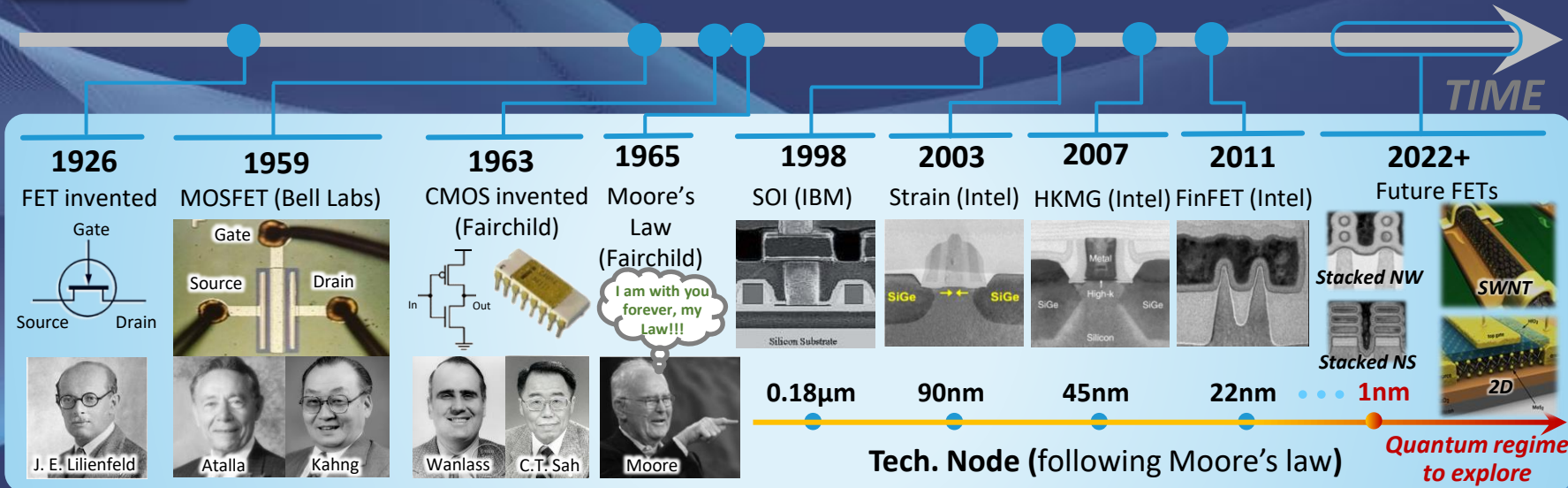
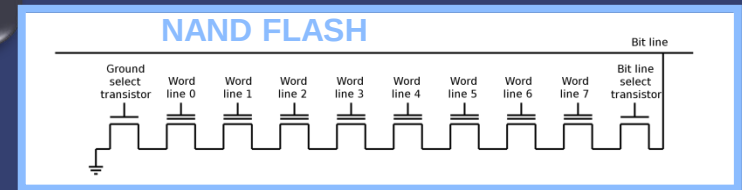
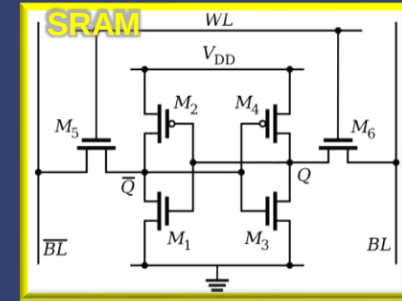
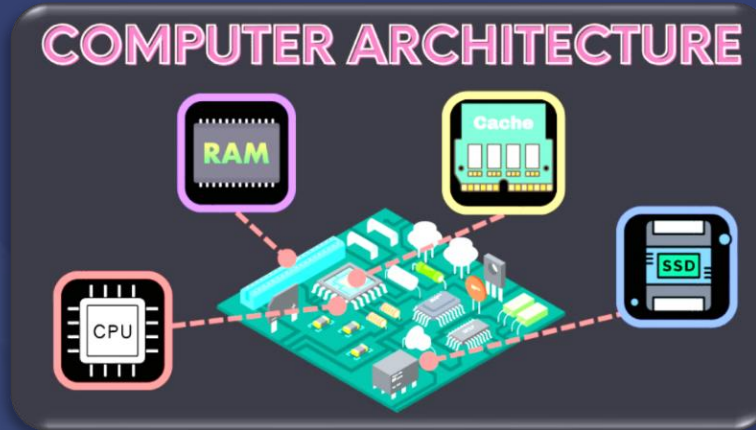
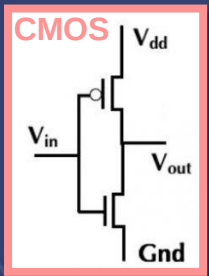
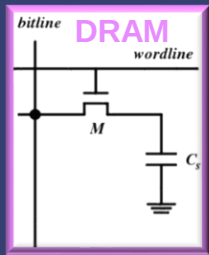
Lecture 5

MOSFET Scaling

Prof. Kaustav Banerjee
Electrical and Computer Engineering
University of California, Santa Barbara

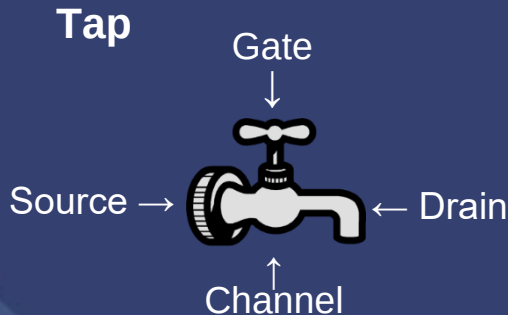
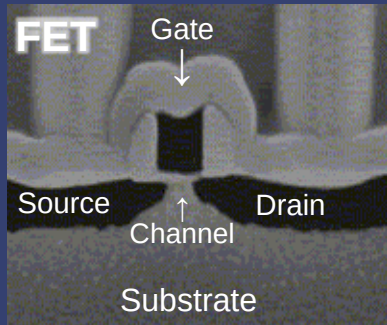
Metal-Oxide-Field-Effect Transistor (MOSFET): Core Element for Computing and Storage

(W. Cao, /K. Banerjee et al., *Nature*, vol. 620, pp. 501-515, 2023)

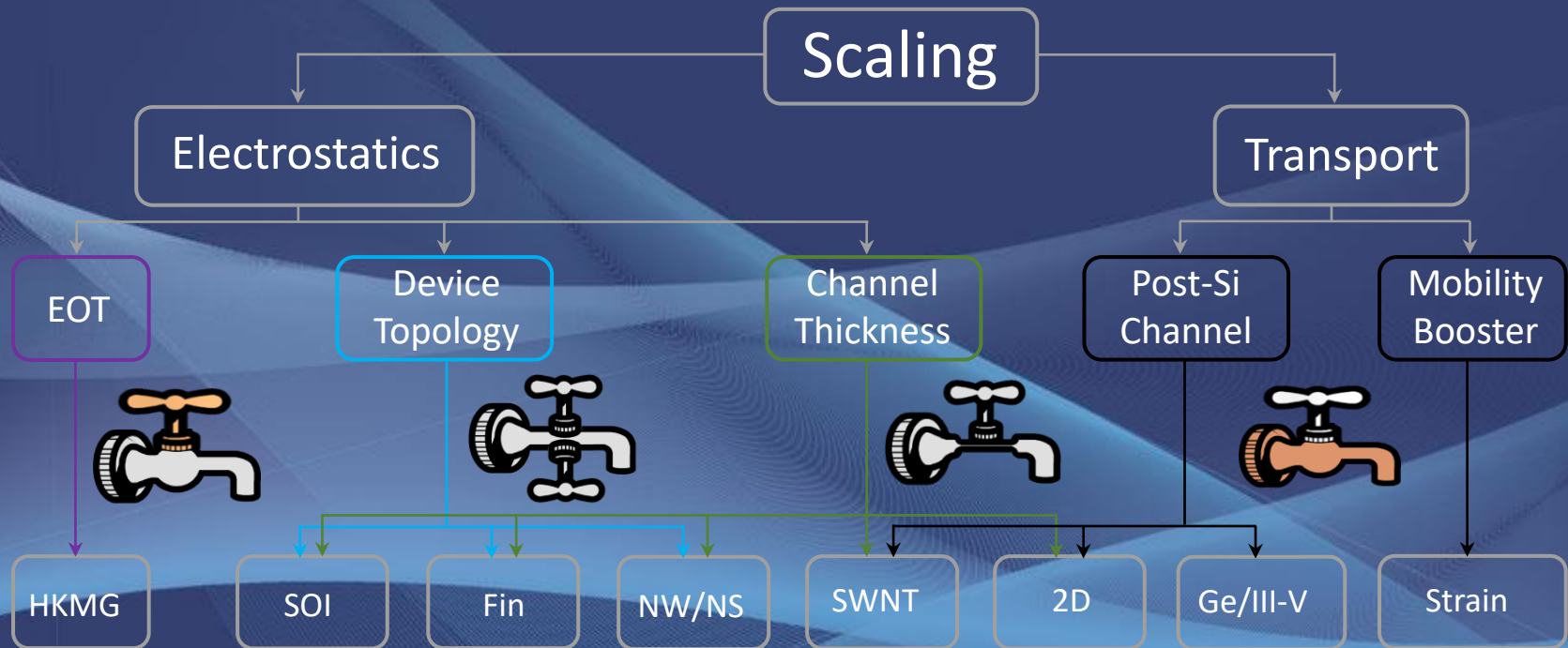


FET Design and Scaling Framework

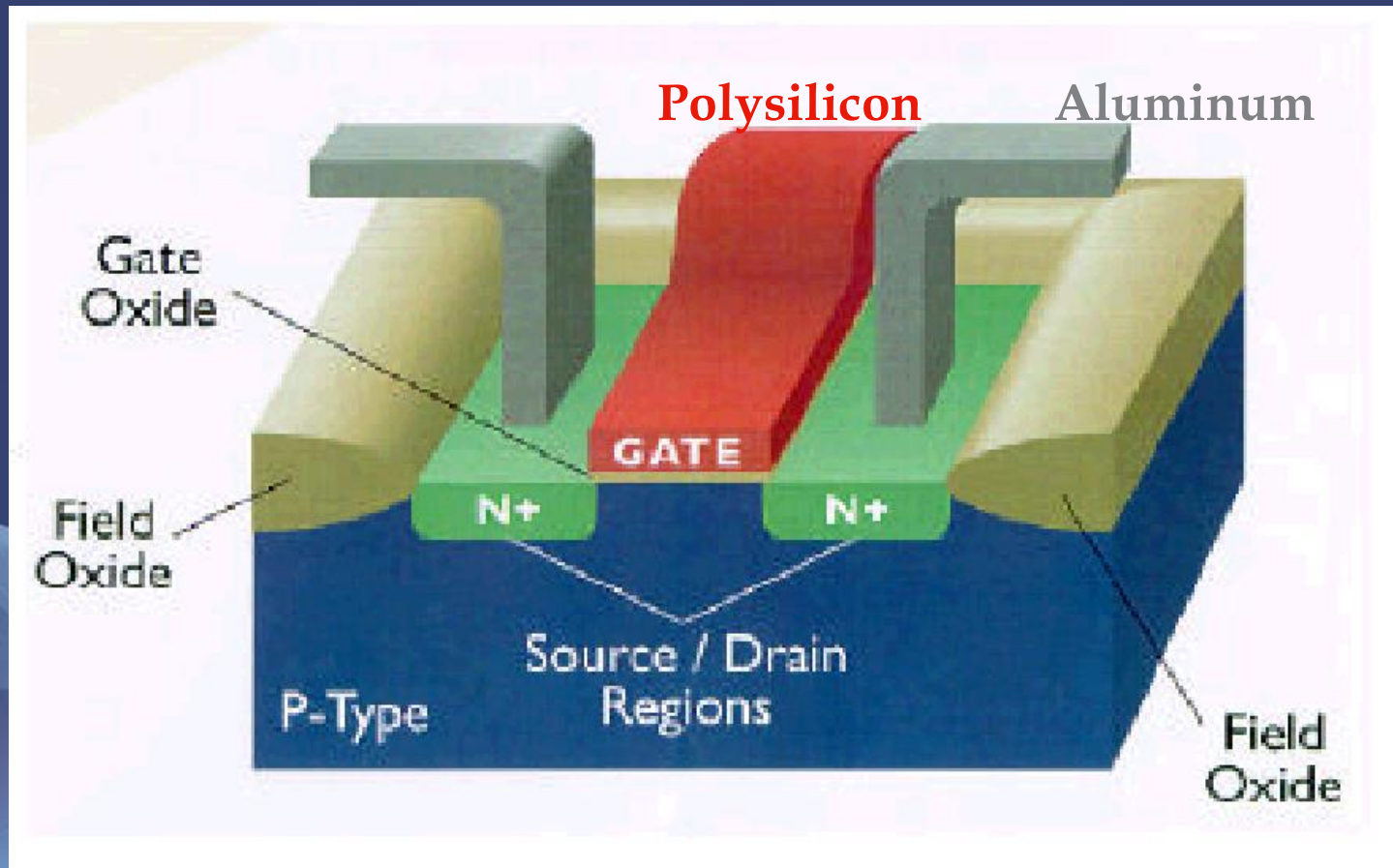
(W. Cao, K. Banerjee et al., *Nature*, vol. 620, pp. 501-515, 2023)



- Any FET can be simply understood as an electric version of a tap. They share very similar design and scaling rules

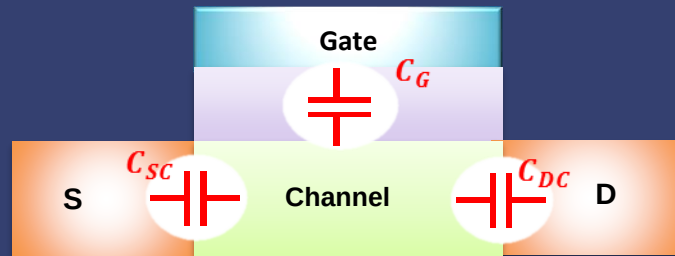


MOSFET Structure



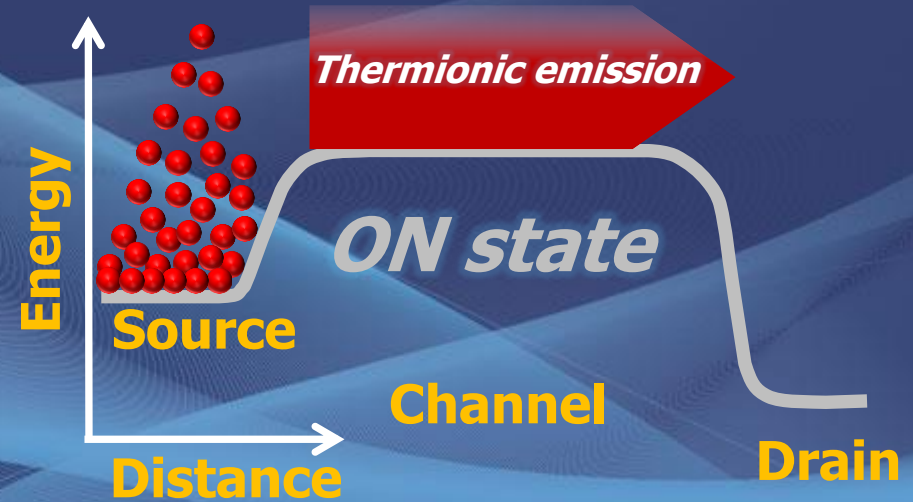
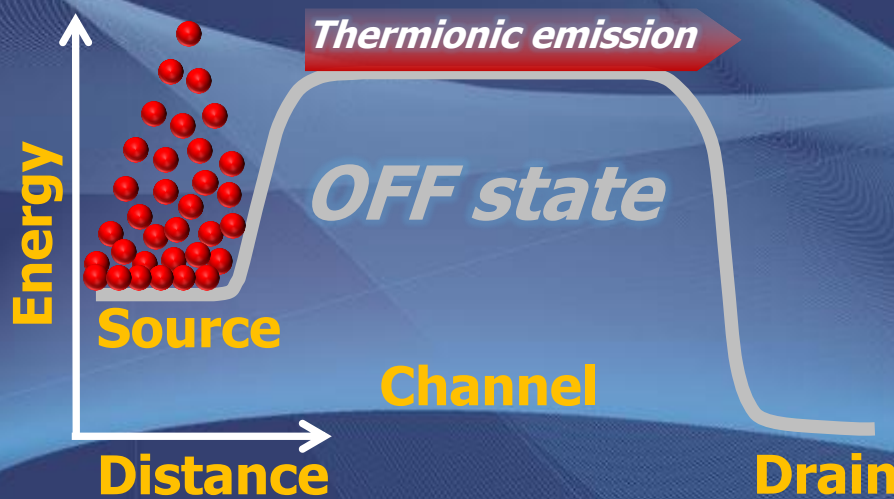
3D Schematic of a n-type planar MOSFET

The Essential Physics of MOSFET



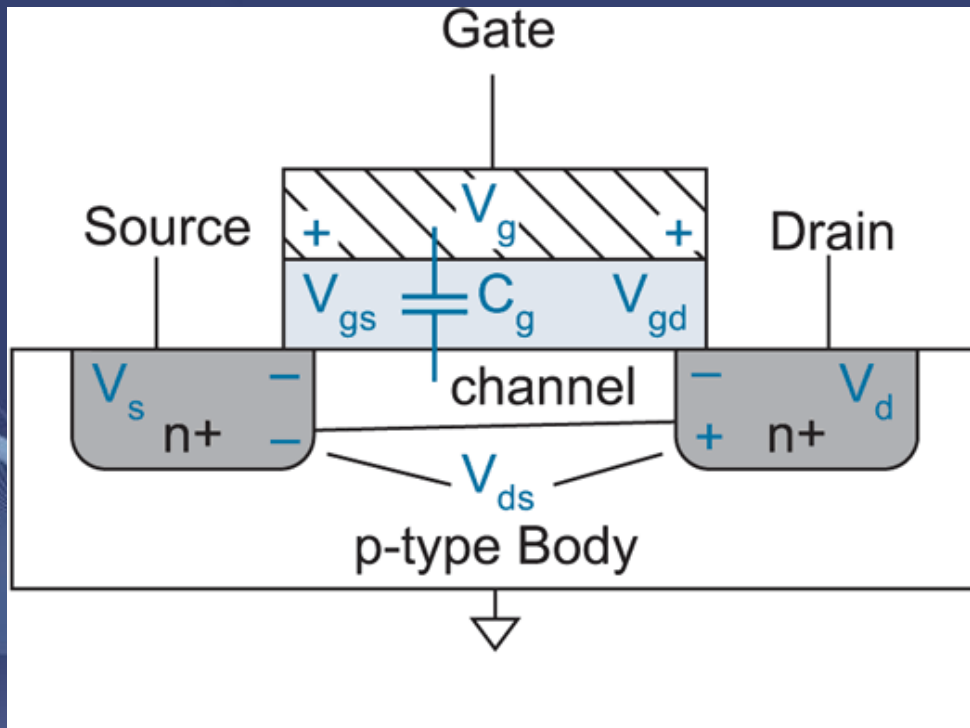
$$V_{Ch} = \frac{C_G}{C_G + C_{SC} + C_{DC}} V_G$$

in the subthreshold region



Mobile Charge

NMOS Transistor:



$$V_{ds} = V_{gs} - V_{gd}$$

Threshold Voltage V_t :
the minimum V_g for
inversion charge layer
to appear

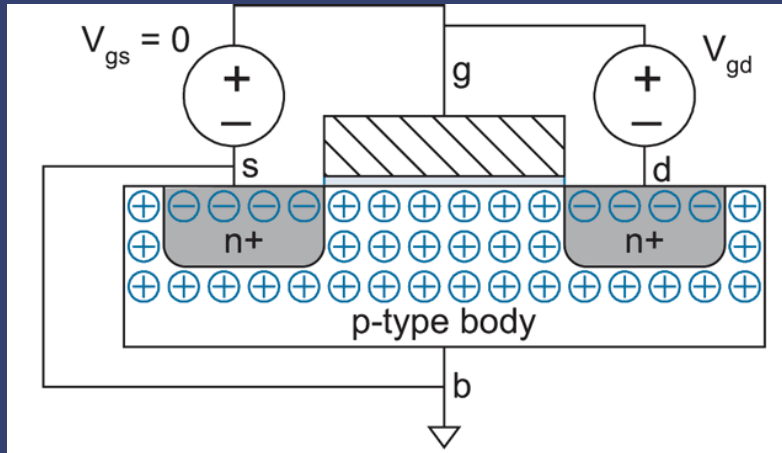
$$Q_c(x) = C_g [V_{gc}(x) - V_t]$$

Note:

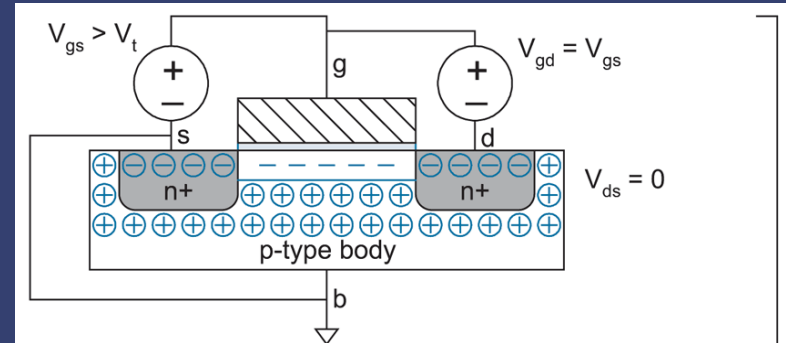
The inversion layer thickness is
assumed to be zero: all
charges are assumed to be
located at the Si surface....like
a sheet of charge....

MOSFET Operation

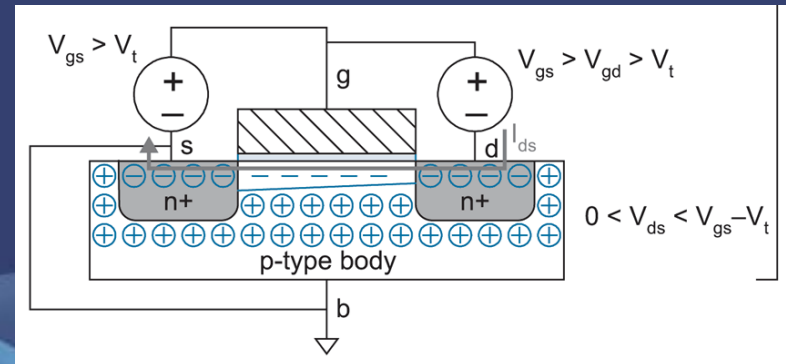
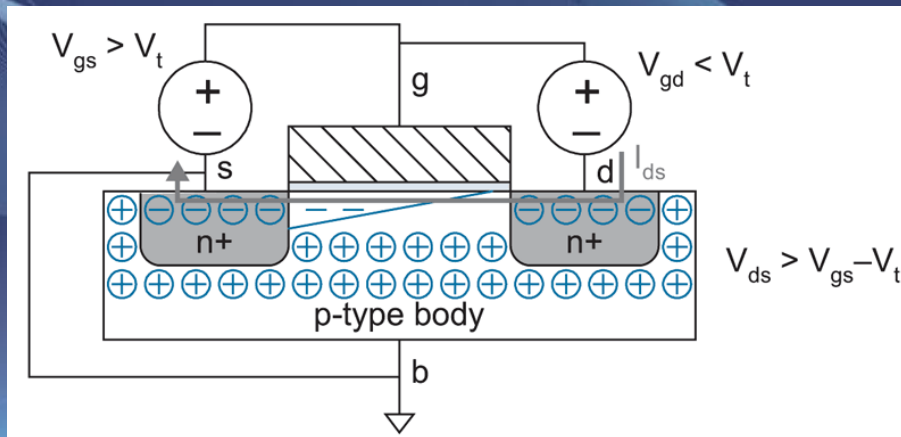
Cut-off



Turn-on



Pinch-off



$$I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t - V_{ds}/2) V_{ds}$$

$$I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) / 2$$

Goals of MOSFET Scaling

- Making things cheaper
 - Integrating more transistors per chip for the same money
 - Price of fabrication of a transistor to be reduced
- But also want the transistor to be faster and low power

$$I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) / 2$$

Technology Scaling Models

- Fixed Voltage Scaling
 - most common model in the early years
 - only dimensions scale, voltages remain constant
- Full Scaling (Constant Electrical Field)
 - ideal model, dimensions and voltage scale together by the same factor S $Q_c(x) = C_g [V_{gc}(x) - V_t]$
- General Scaling
 - most realistic for today's situation
 - voltages and dimensions scale with different factors

Scaling Models

$$I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) / 2$$

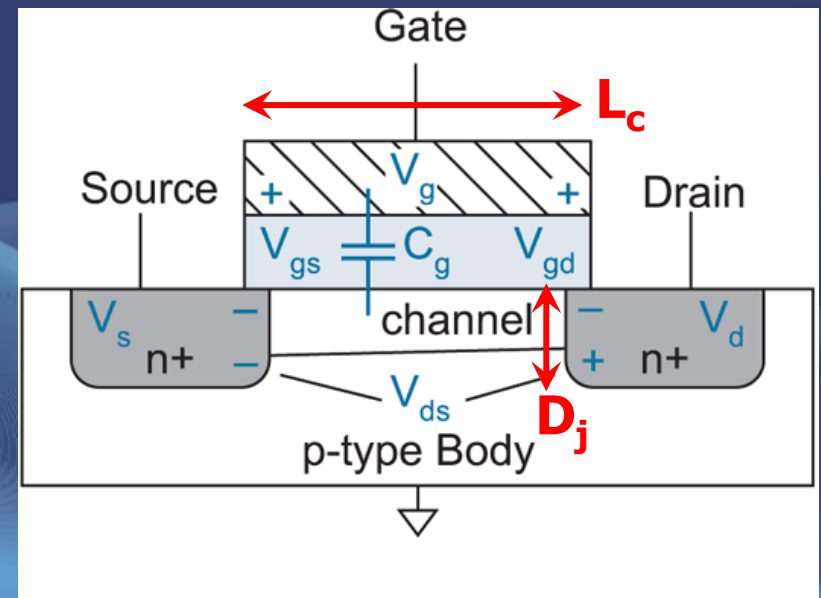
Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD} , V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	$C_{ox}WL$	1/S	1/S	1/S
k_n , k_p	$C_{ox}W/L$	S	S	S
I_{av}	$k_{n,p} V^2$	1/S	S/U^2	S
t_p (intrinsic)	$C_L V / I_{av}$	1/S	U/S^2	$1/S^2$
P_{av}	$C_L V^2 / t_p$	$1/S^2$	S/U^3	S
PDP	$C_L V^2$	$1/S^3$	$1/SU^2$	1/S

Scaling Problems Faced

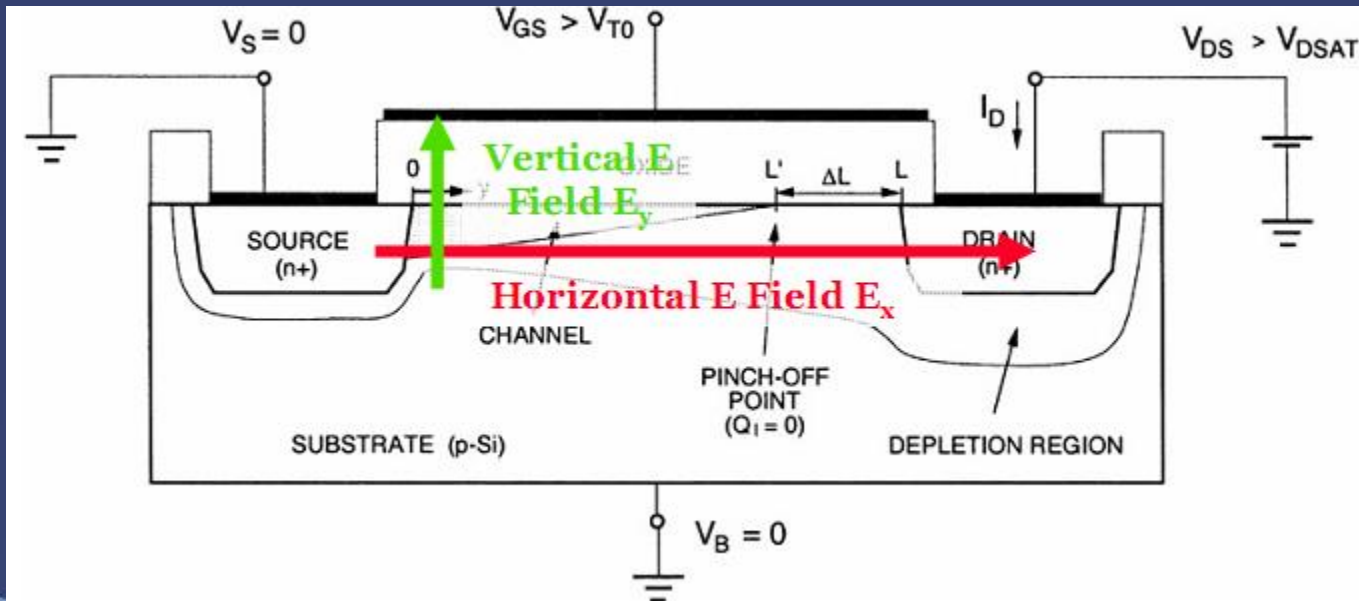
- Short channel effects – gate efficiency degradation, parasitic R/C, leakage currents, variability, reliability
- Interconnect delay becomes significant
- Temperature on the die increases → higher resistance

Short Channel Effects

- Short-channel device: channel length is comparable to depth of drain and source junctions and depletion width
 - In general, visible when $L_c \sim 1\mu\text{m}$ and below
- Short-channel effects
 - Carrier velocity saturation
 - Mobility degradation
 - Threshold voltage variation
 - V_T Roll-Off
 - Drain-Induced Barrier Lowering
 - Oxide breakdown



Increasing Electric Field in MOSFET



Vertical Electric Field

1980

$E_y = 5V/1000\text{\AA} = 50 \times 10^4 \text{ V/cm}$

1995

$E_y = 3.3V/75\text{\AA} = 4.4 \times 10^6 \text{ V/cm}$

2001

$E_y = 1.2V/22\text{\AA} = 5.5 \times 10^6 \text{ V/cm}$

Horizontal Electric Field

1980

$E_x = 5V/5\mu\text{m} = 10^4 \text{ V/cm}$

1995

$E_x = 3.3V/0.35\mu\text{m} = 9.4 \times 10^4 \text{ V/cm}$

2001

$E_x = 1.2V/0.1\mu\text{m} = 1.2 \times 10^5 \text{ V/cm}$

E_x induced Velocity Saturation

- As the horizontal E_x field grows, so does the velocity of the electrons $\langle v_x \rangle = \mu_n E_x$
- As the gate shrinks, the velocity reaches the point where it cannot increase as much with increasing E field
- This is due to velocity saturation
 - Electrons become so energetic that they scatter (crash into the crystal lattice)

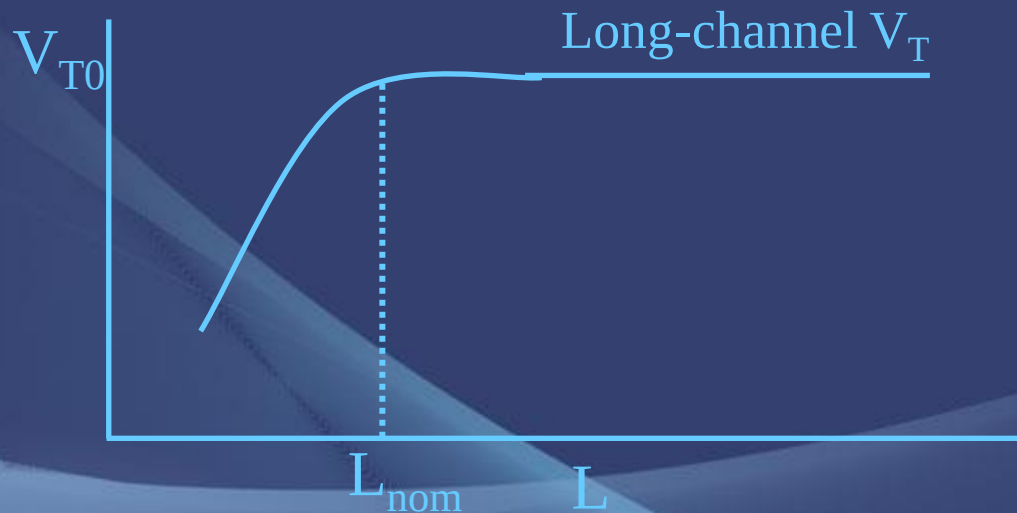
E_y induced Mobility Degradation

- MOS I/V equations depend on surface mobility μ_n (or μ_p)
 $I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) / 2$
 - Surface mobility is the mobility of the minority charge carriers at the surface of the channel
- In short-channel devices, μ_n and μ_p are not constant
 - As vertical electric field E_y increases, surface mobility decreases because the electrons' wave function extends increasingly into the extremely low-mobility oxide
 - μ_0 = low-field mobility, η is empirical constant
 - As V_{GS} increases, effective surface mobility decreases

$$\mu_{eff} = \frac{\mu_0}{1 + \eta(V_{GS} - V_T)}$$

Threshold voltage Roll-off...

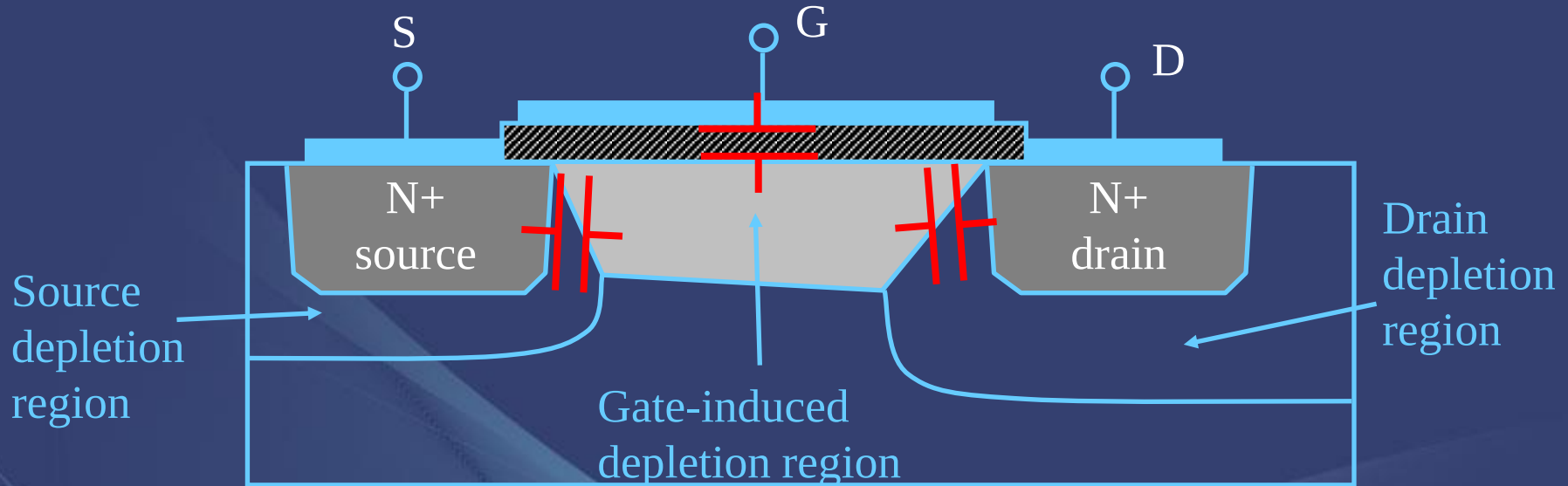
Graphically: V_{T0} versus channel length L



V_T Roll-off:

V_T decreases rapidly with channel length

Reason for V_T Roll-off....



- Even with $V_{GS}=0$, part of the channel is already depleted
- Bulk depletion charge is smaller in short-channel device $\rightarrow V_T$ is smaller

Estimation of V_T Roll-off....

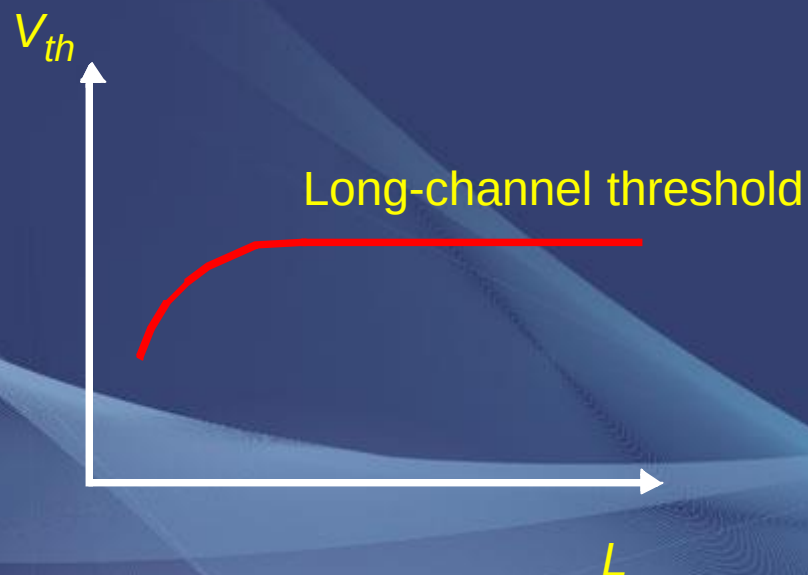
- Change in V_{t0} :
 - x_{dS} , x_{dD} : depth of depletion regions at S, D
 - x_j : junction depth
- ΔV_{t0} is proportional to (x_j/L)
 - For short channel lengths, ΔV_{t0} is large
 - For large channel lengths, term approaches 0

DIBL

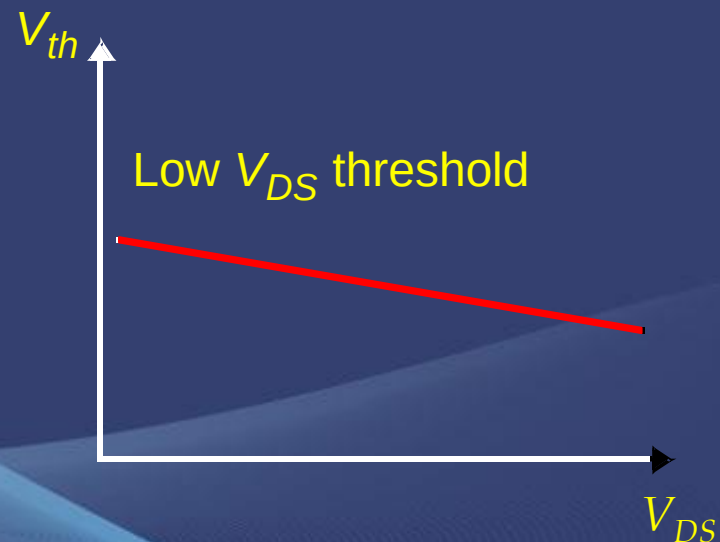
- Drain-induced barrier lowering (DIBL)
 - Drain voltage V_{DS} causes change in threshold voltage
 - As V_{DS} is increased, threshold voltage decreases
- Cause: depletion region around drain
 - Depletion region depth around drain depends on drain voltage
 - As V_{DS} is increased, drain depletion region gets deeper and extends further into channel
 - For very large V_{DS} , source and drain depletion regions can meet → *punch-through!*
- Issue: results in uncertainty in circuit design

Effect of SCE and DIBL on V_{th}

- $V_{th} = V_{th_long-channel} - SCE - DIBL$



Threshold as a function of the length (for low V_{DS})



Drain-induced barrier lowering (DIBL) (for low L)

Easier to envision V_{th} drop in short-channel devices with a band diagram!

Other V_t Variation Sources

- Hot-carrier effect
 - Threshold voltages drift over time
- Negative-Bias Temperature Instability (NBTI)
 - Issue in PMOS transistors
 - V_t drifts over time
 - Typical stress temperature 100-150 C
 - Typical oxide electric fields of 5-6 MV/cm

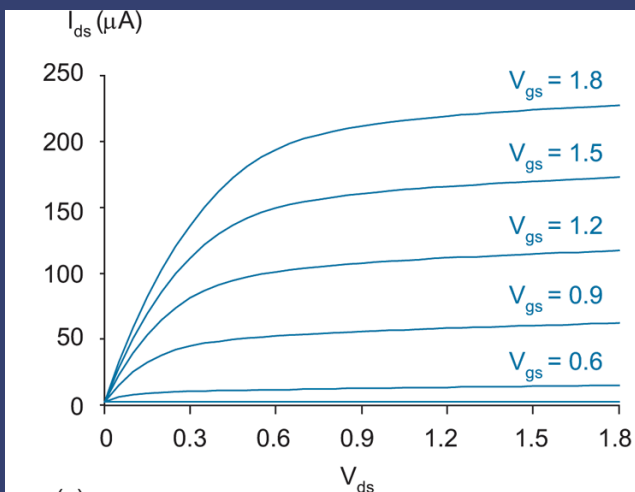
Sub-threshold Conduction

- When $V_{GS} < V_T$, transistor is “off”
 - However, small drain current I_D still flows
 - Called *subthreshold leakage* current
- Model for subthreshold current:

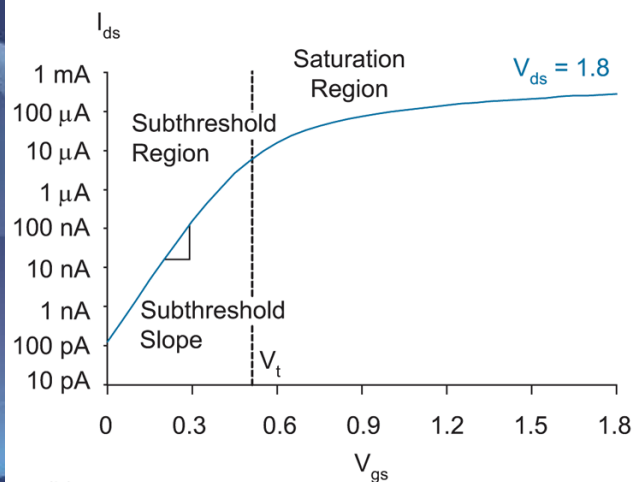
$$I_D(\text{subthreshold}) = I_S W e^{\frac{q}{kT}(AV_{GS} + BV_{DS})}$$

- Increases as V_{GS} increases (potential barrier lowered)
- Increases as V_{DS} increases (DIBL)

Subthreshold Leakage



(a)



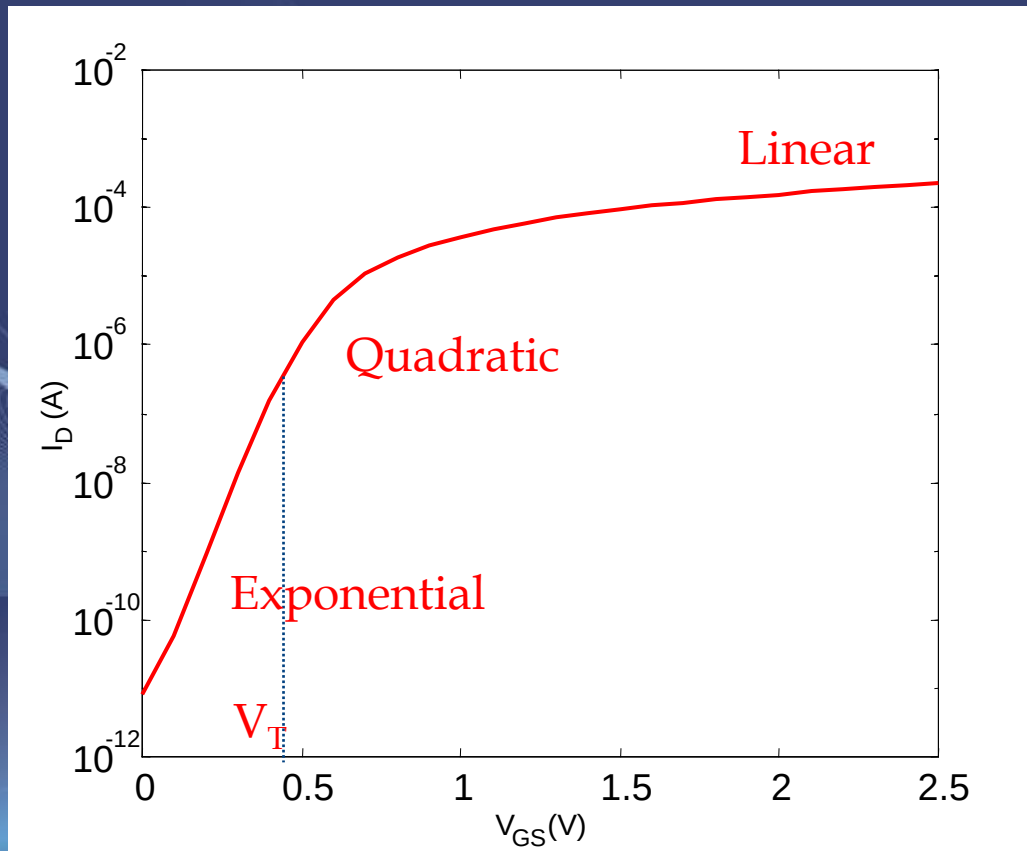
(b)

FIG 2.15 Simulated I-V characteristics

- Dominant leakage mechanism
- Increases exponentially with temperature and V_t reduction

$$S = \left(\frac{d(\log I_d)}{dV_g} \right)^{-1} = \left(\frac{\partial V_g}{\partial \psi_s} \right) \left(\frac{\partial \psi_s}{\partial (\log I_d)} \right) = \left(1 + \frac{C_{dm}}{C_{ox}} \right) \frac{kT}{q} \ln(10)$$

Sub-Threshold Conduction (2)



The Slope Factor

$$I_D \sim I_0 e^{\frac{qV_{GS}}{nkT}}, n = 1 + \frac{C_D}{C_{ox}}$$

S is ΔV_{GS} for $I_{D2}/I_{D1} = 10$

$$S = n \left(\frac{kT}{q} \right) \ln(10)$$

Typical values for S :
60 - 100 mV/decade

Gate Leakage

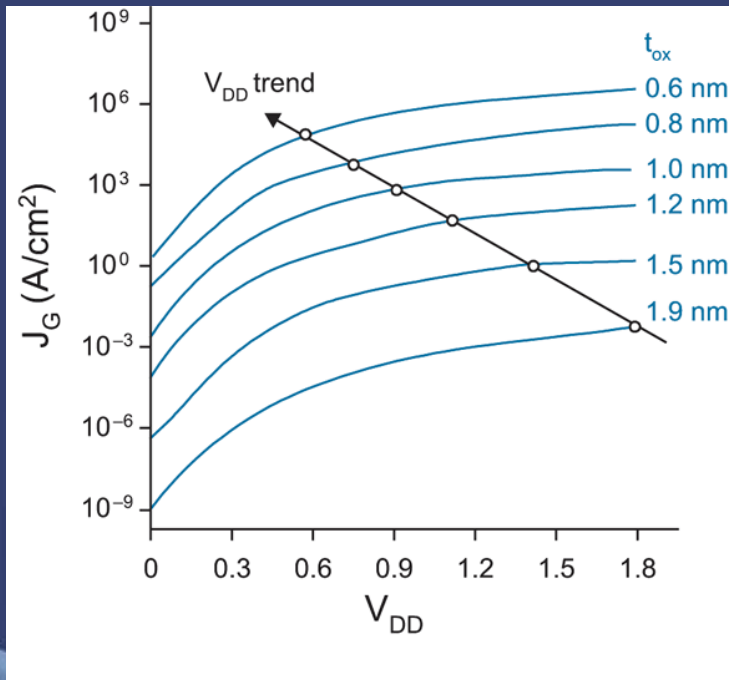
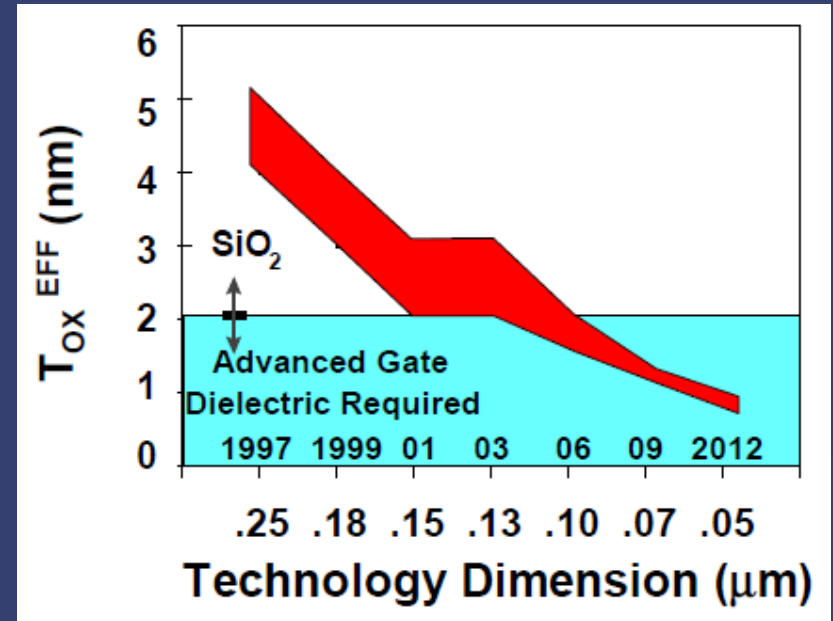


FIG 2.20 Gate leakage current from [Song01]

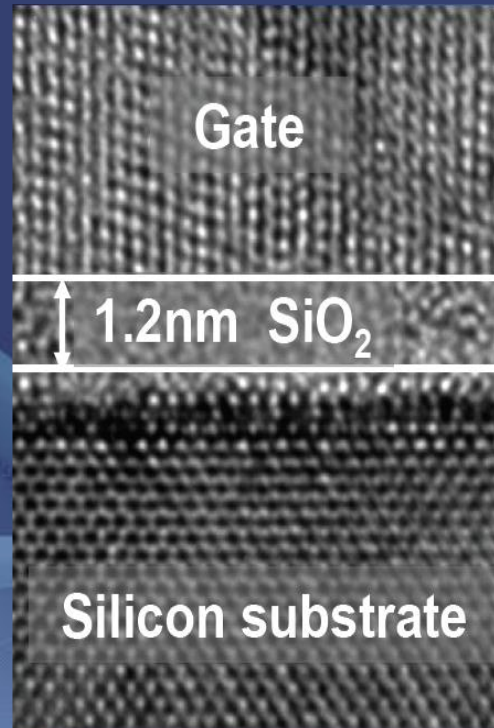


- *Increases with gate oxide (SiO_2) scaling*
- *High-k gate oxides can be used to lower gate leakage*

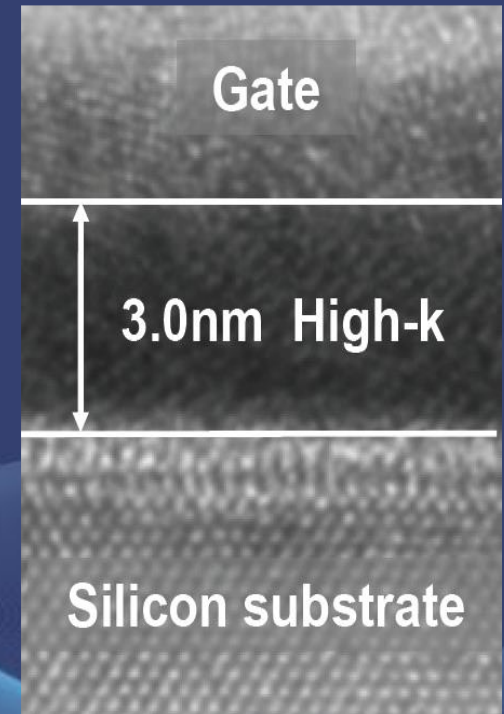
High-k/Metal Gate Technology

A High- k dielectric can deliver an equivalent gate capacitance (per unit area) that SiO_2 can provide at only a very small thickness, called EOT.

Gate leakage $\times 1$
Gate capacitance $\times 1$



Gate leakage $\times 0.01$
Gate capacitance $\times 1.6$



Source: Intel

- High-k/metal gate devices offer lower gate leakage with small increase in gate capacitance

Avalanche Breakdown

- Due to scaling, high-velocity electrons can impact the drain, dislodging holes---Called impact ionization
- Holes are swept towards grounded substrate → cause substrate current
- Also affects long-term reliability
- This is another factor which limits the process scaling → voltage must scale down as length scales