The Abdus Salam International Centre for Theoretical Physics

2499-4

International Training Workshop on FPGA Design for Scientific Instrumentation and Computing

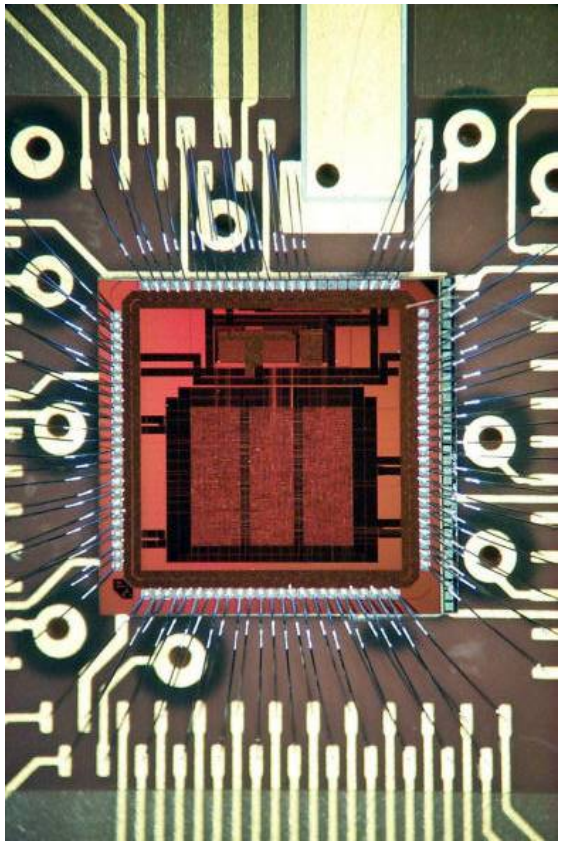
11 - 22 November 2013

Introduction to VLSI Degital Design
Transistors

Sandro BONACINI
CERN, Geneva
Switzerland

Outline

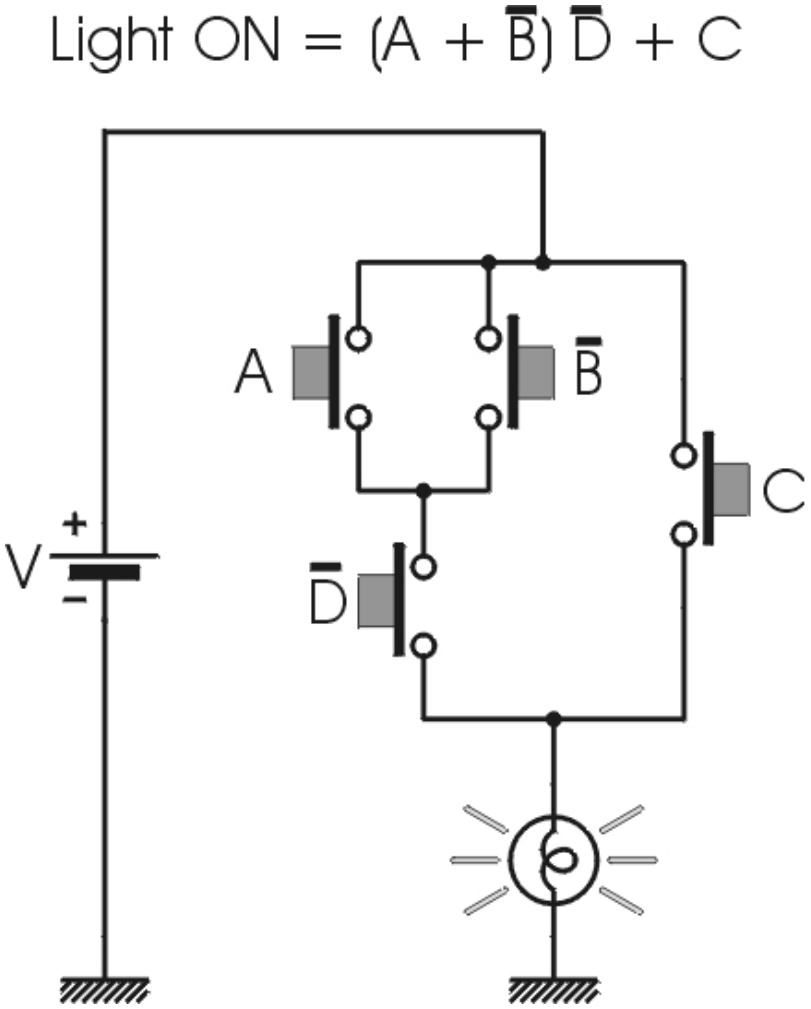
- Introduction
- Transistors
  - *DC behaviour*
  - *MOSFET capacitances*
- The CMOS inverter
- Technology
- Scaling
- Gates
- Sequential circuits
- Storage elements



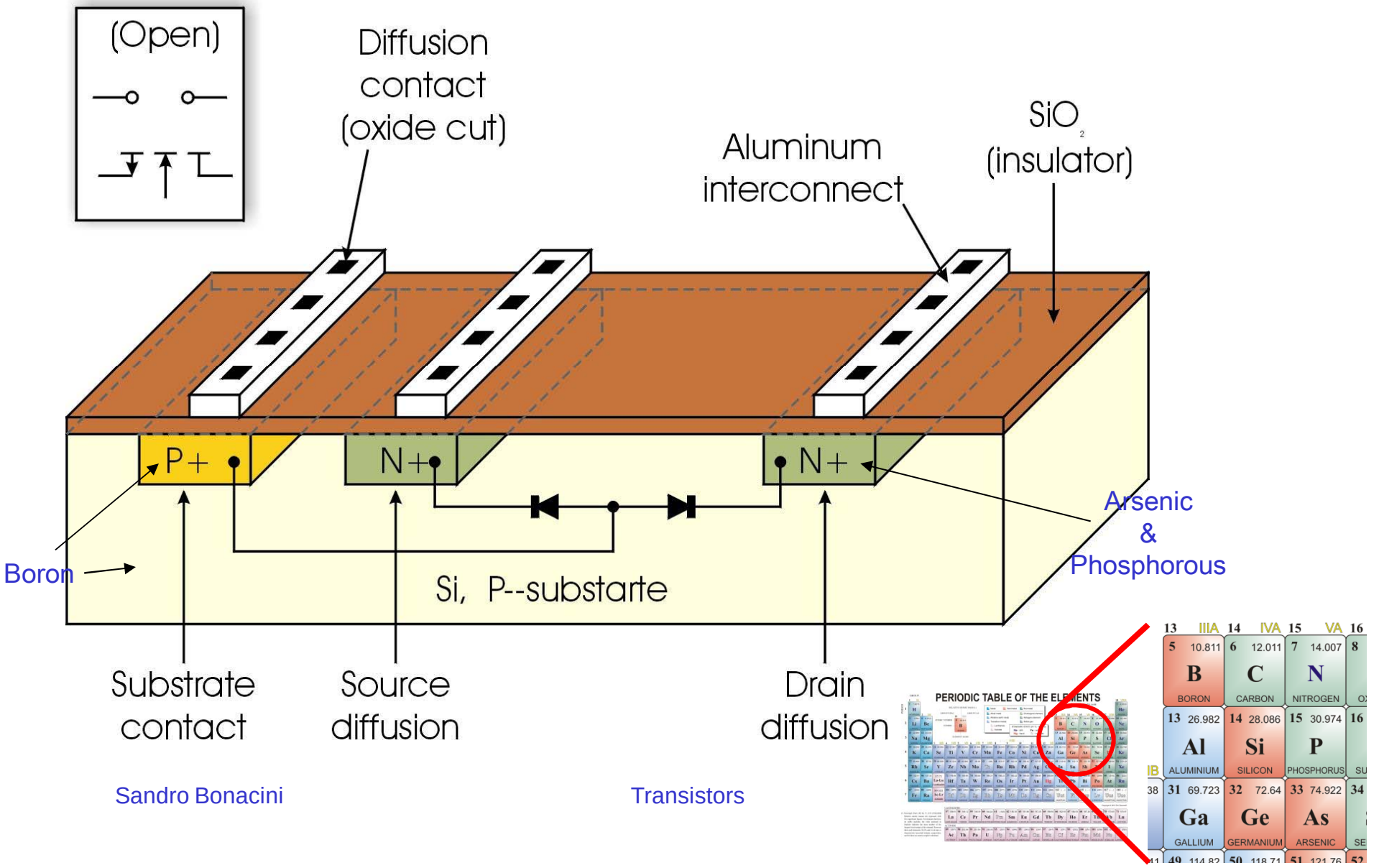
Many slides are a courtesy of Paulo Moreira

"Making Logic"

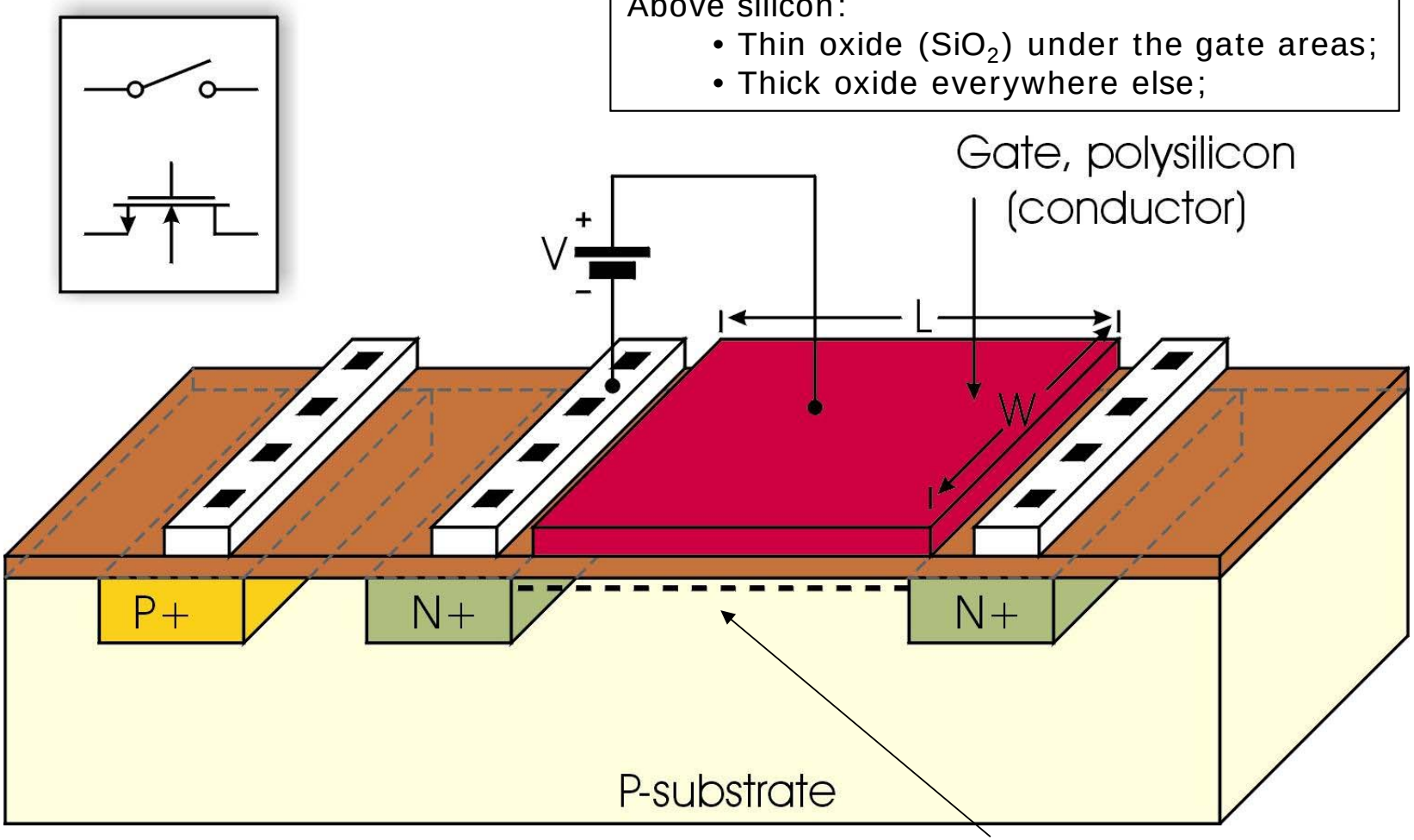
- Logic circuit "ingredients":
  - Power source
  - Switches
  - Inversion
  - Power gain (for multiple stages)
- Power always comes from some form of external generator.
- NMOS and PMOS transistors:
  - Can perform the last three functions
  - They are the building blocks of CMOS technologies!



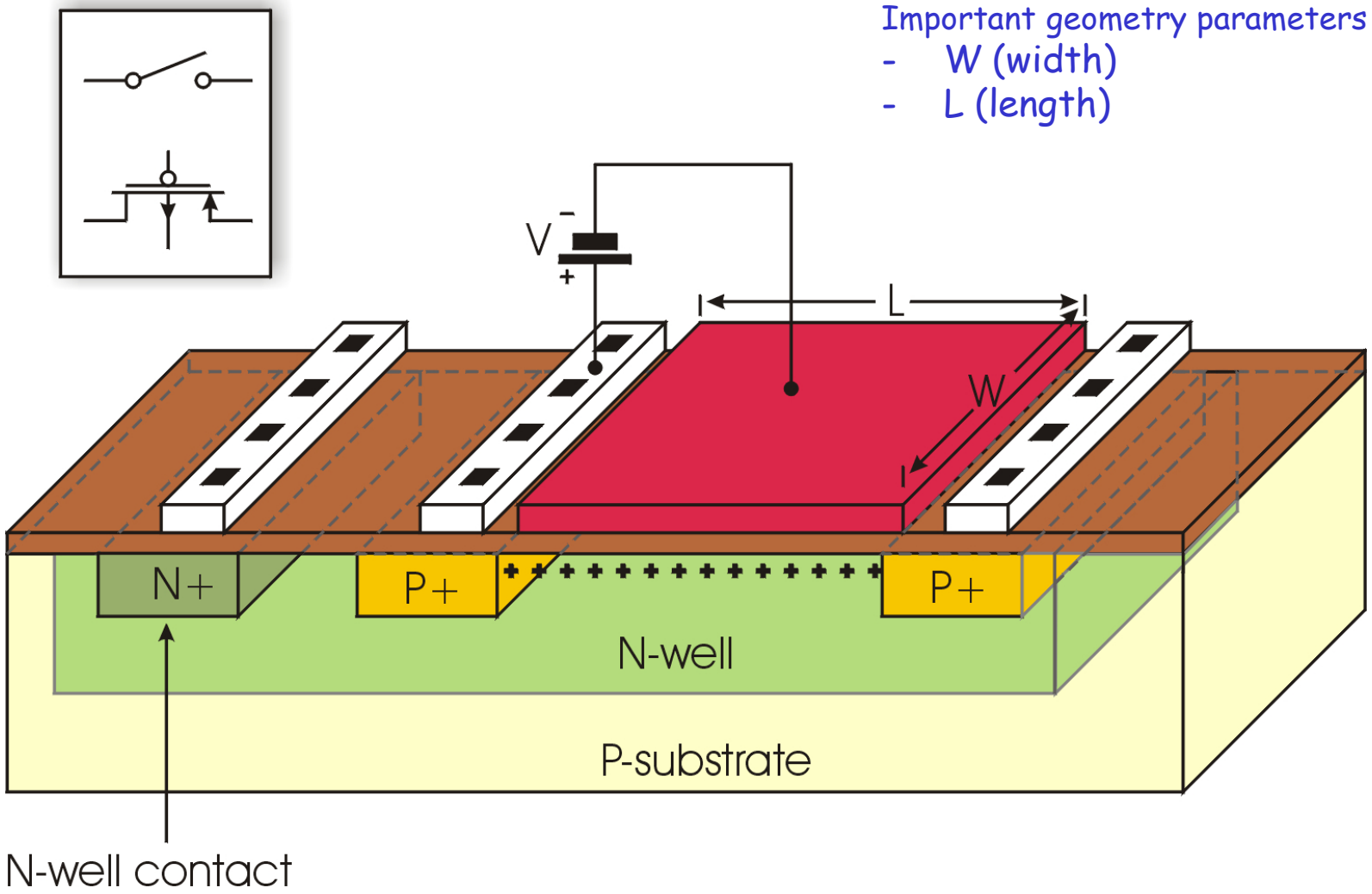
Silicon switches: the NMOS



Silicon switches: the NMOS



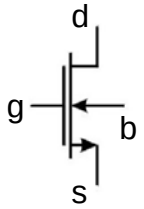
Silicon switches: the PMOS



MOSFET equations

- Cut-off region:

$$I_{ds} = 0 \quad \text{for} \quad V_{gs} - V_T < 0 \qquad (1)$$



- Linear region:

$$I_{ds} = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[\left(V_{gs} - V_T \right) \cdot V_{ds} - \frac{V_{ds}^2}{2} \right] \cdot \left(1 + \lambda \cdot V_{ds} \right) \text{ for } 0 < V_{ds} < V_{gs} - V_T \qquad (2)$$

- Saturation:

$$I_{ds} = \frac{\mu \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot \left(V_{gs} - V_T \right)^2 \cdot \left(1 + \lambda \cdot V_{ds} \right) \text{ for } V_{ds} > V_{gs} - V_T \qquad (3)$$

- Oxide capacitance

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \quad \left(\mathrm{F} / \mathrm{m}^2 \right)$$

- Process "transconductance"

$$\mu \cdot C_{ox} = \frac{\mu \cdot \varepsilon_{ox}}{t_{ox}} \quad \left(\mathrm{A} / \mathrm{V}^2 \right)$$

0.25μm process

t_{ox} = 5 nm (~10 atomic layers)

C_{ox} = 5.6 fF/μm^2

65 nm process

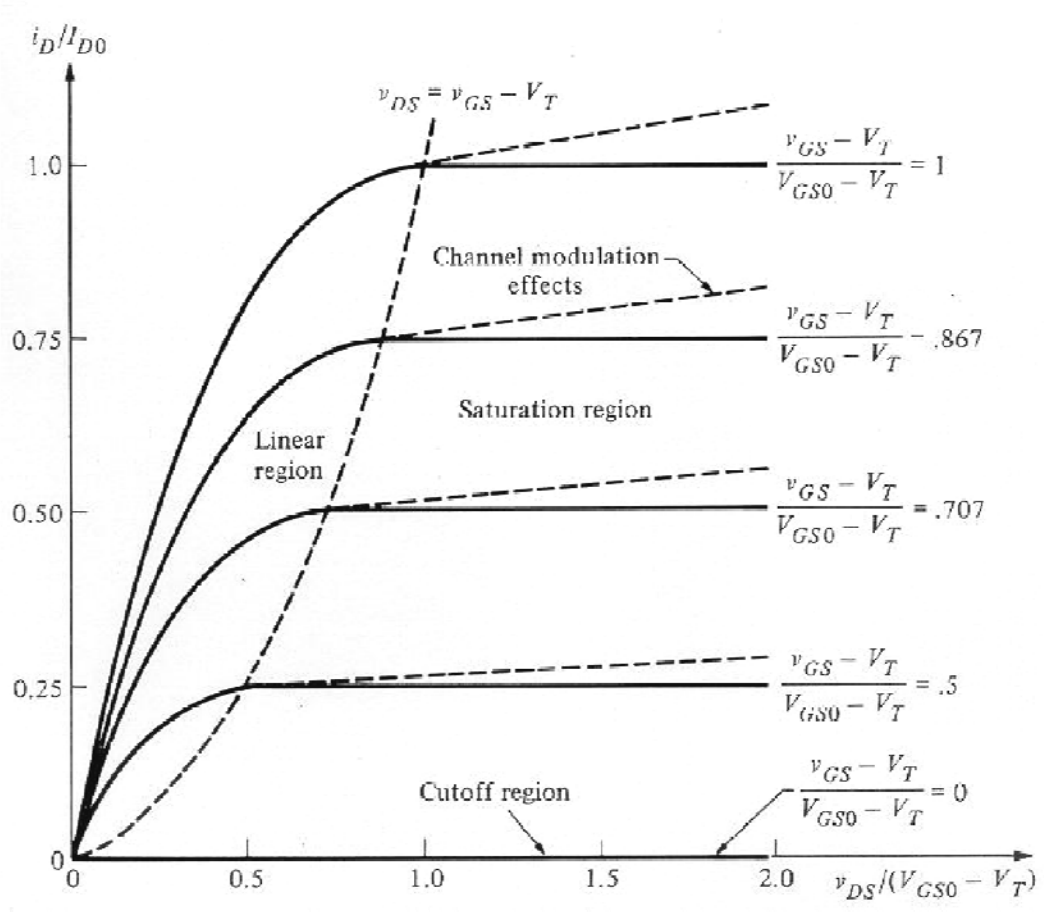
t_{ox} = 2.4 nm (~5 atomic layers)

C_{ox} = 15 fF/μm^2

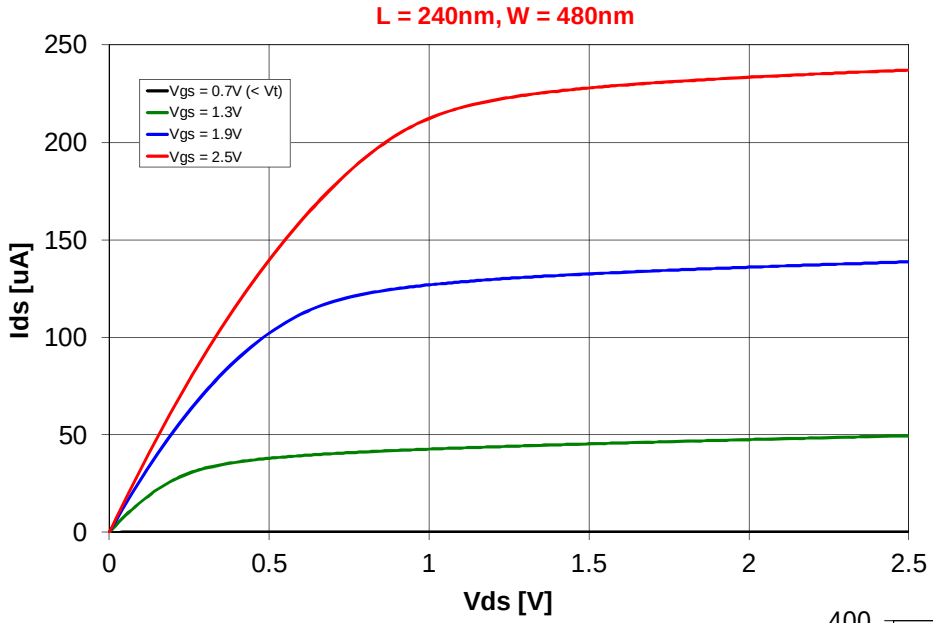
MOS output characteristics

- **Linear region**:
  $V_{ds}<V_{gs}-V_T$
  - Voltage controlled resistor
- **Saturation region**:
  $V_{ds}>V_{gs}-V_T$
  - Voltage controlled current source
- Curves deviate from the ideal current source behavior due to:
  - Channel modulation effects

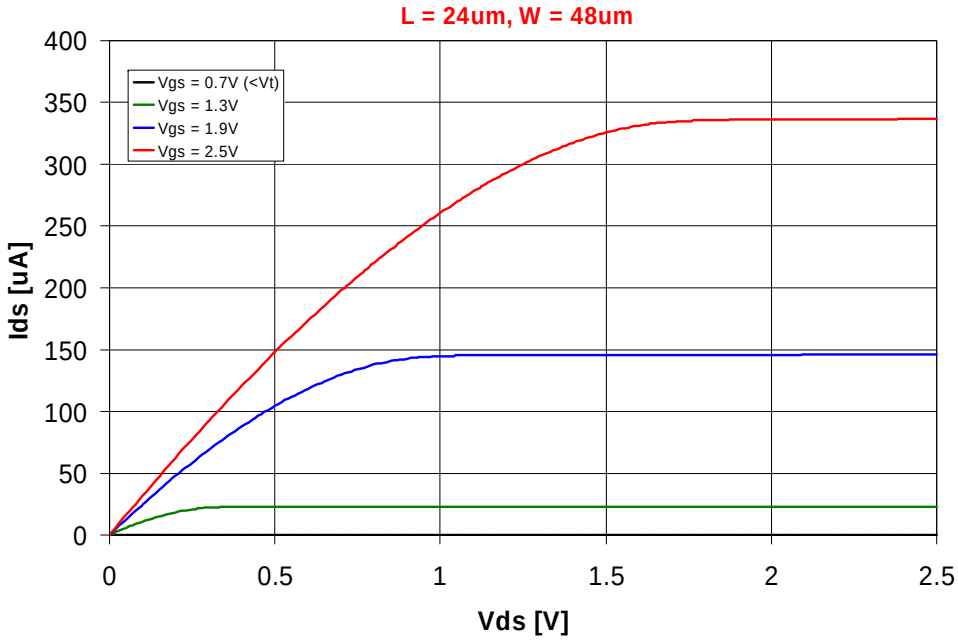
- (V_{gs0} is the maximum V_{gs}, I_{ds0} is the maximum I_{ds})



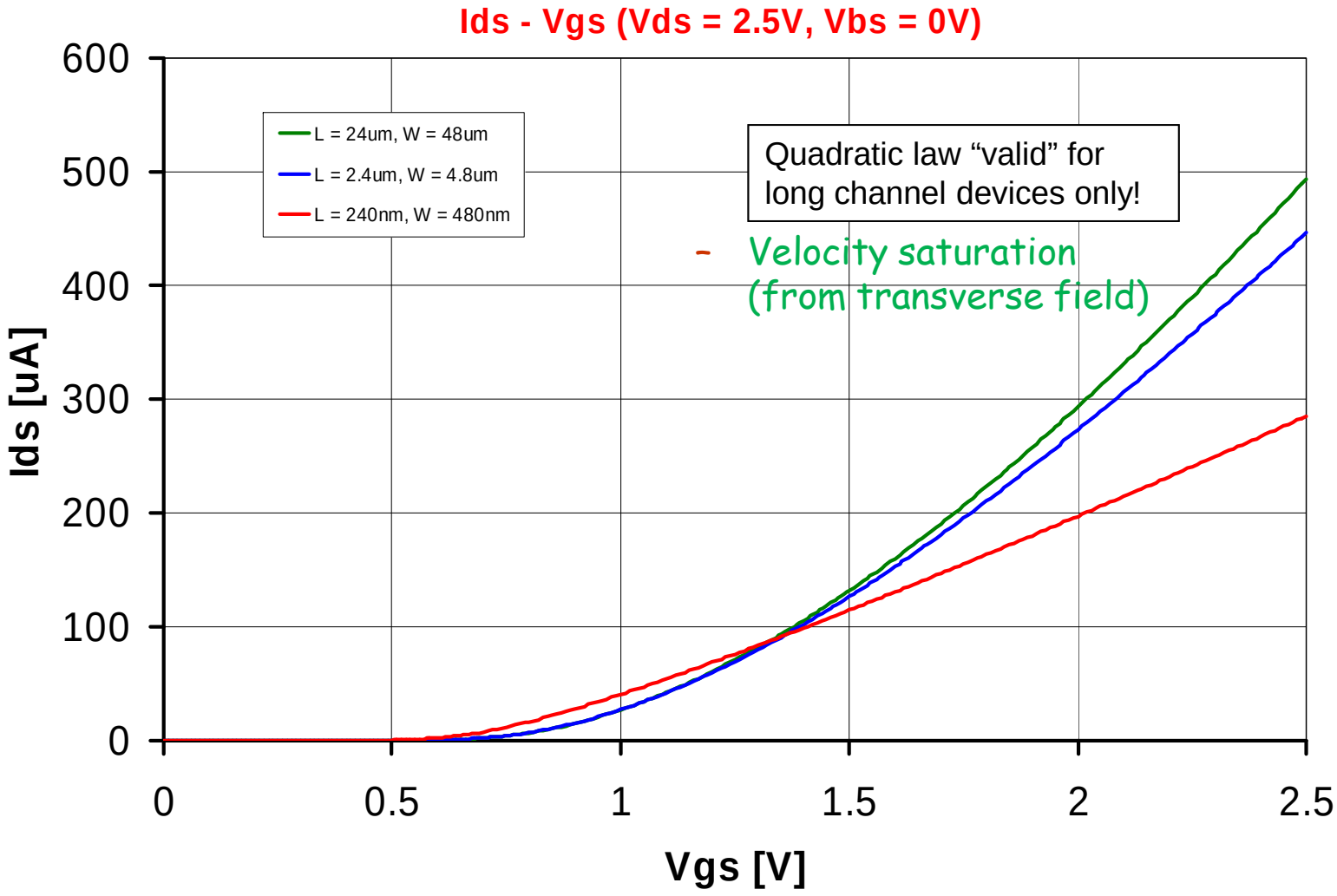
MOS output characteristics



- Channel modulation effects
- Visible in small devices

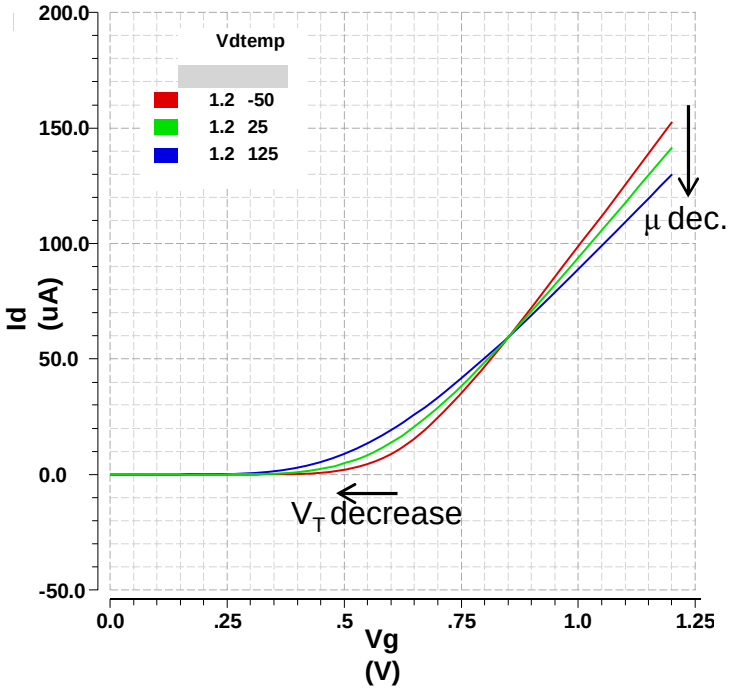


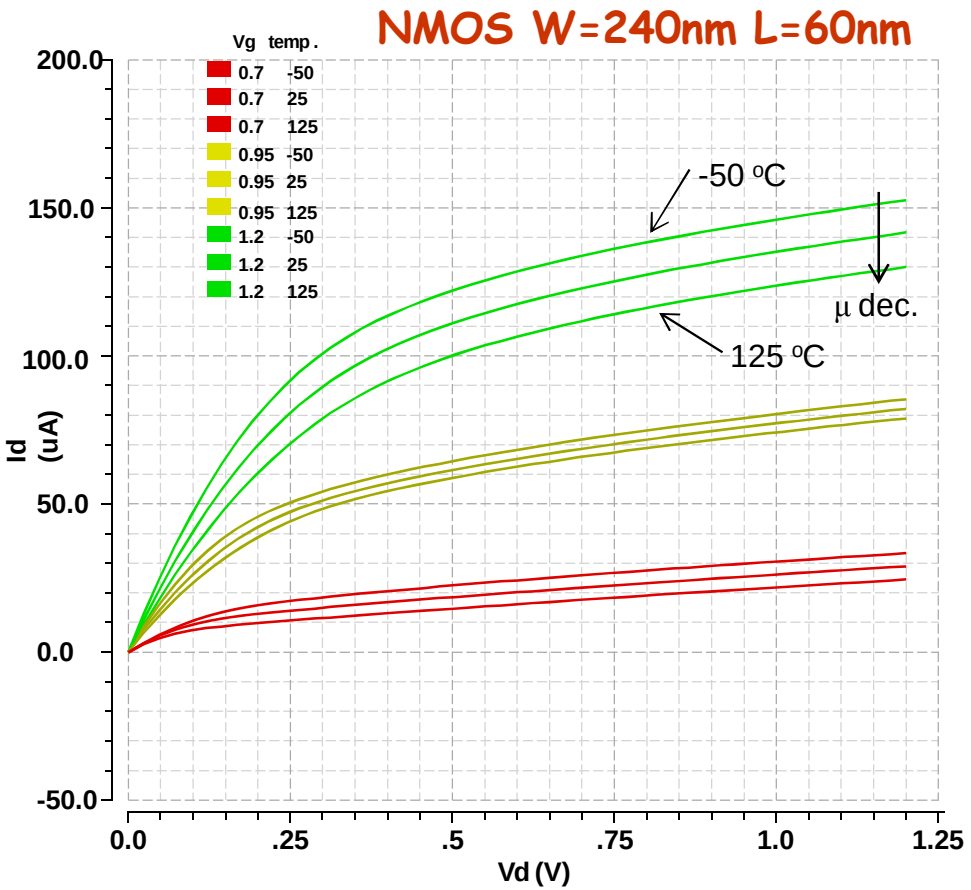
Is the quadratic law valid?



Temperature dependence

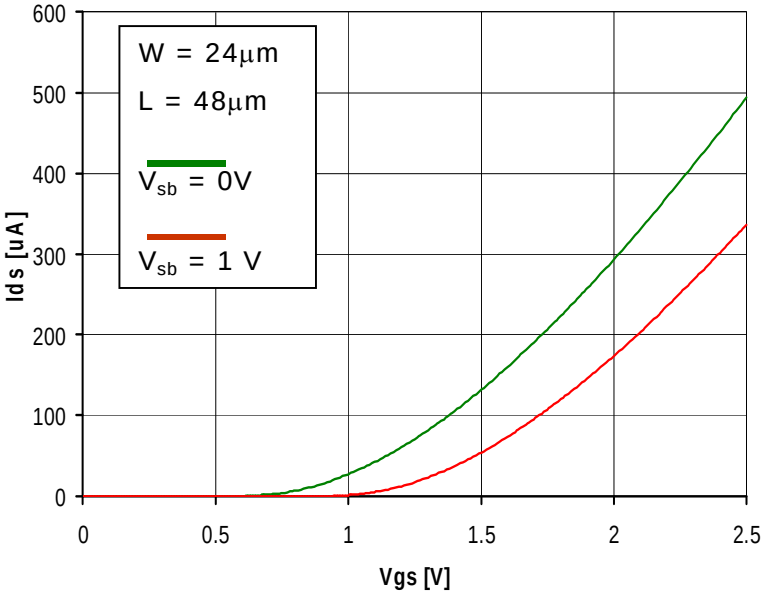
- **Transistor characteristics are influenced by temperature (T)**
  - V_T decreases linearly with T
  - μ decreases with T
  - $I_{leakage}$ increases with T
- Circuit performances are worst at high temperature

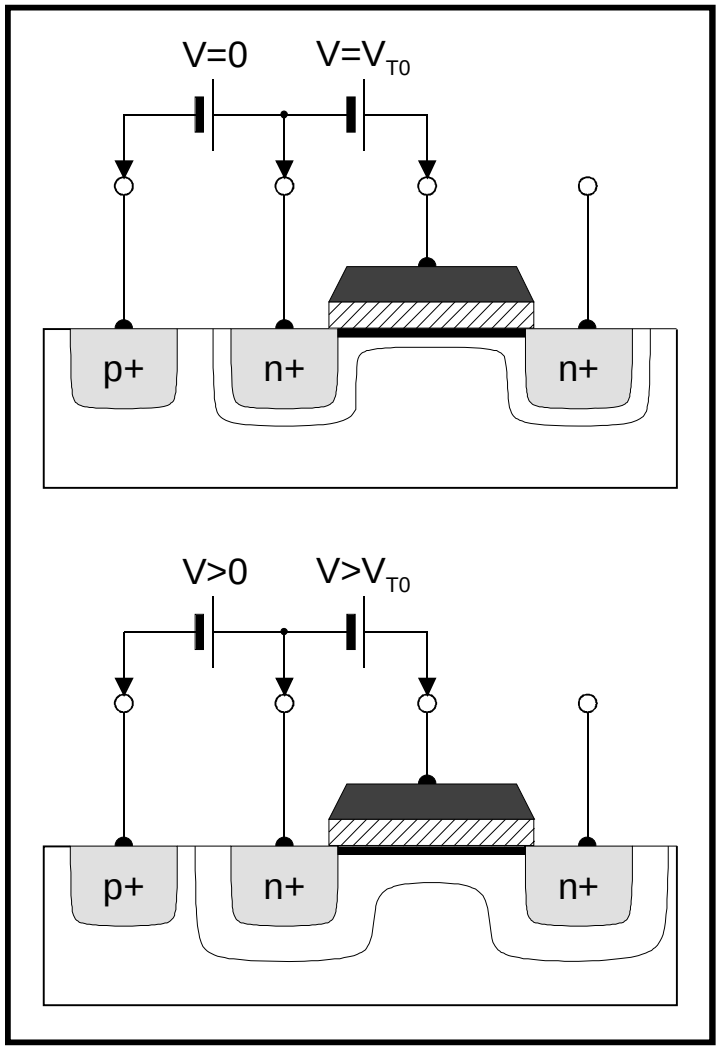




Bulk effect

- The threshold depends on:
  - Gate oxide thickness
  - Doping levels
  - Source-to-bulk voltage
- If $V_{sb} > 0$ the threshold voltage is higher than $V_{sb} = 0$





Weak inversion

- Is I_d=0 when V_{gs}<V_T?
- For V_{gs}<V_T the drain current depends exponentially on V_{gs}
- Weak inversion is when:

$$I_d < 2n\mu C_{ox}\frac{W}{L}\left(\frac{kT}{q}\right)^2 = I_{ST}$$

- In weak inversion and saturation (V_{ds} > ~150mV):

$$I_d \cong \frac{W}{L} I_{d0} e^{\frac{qV_{gs}}{nkT}}$$

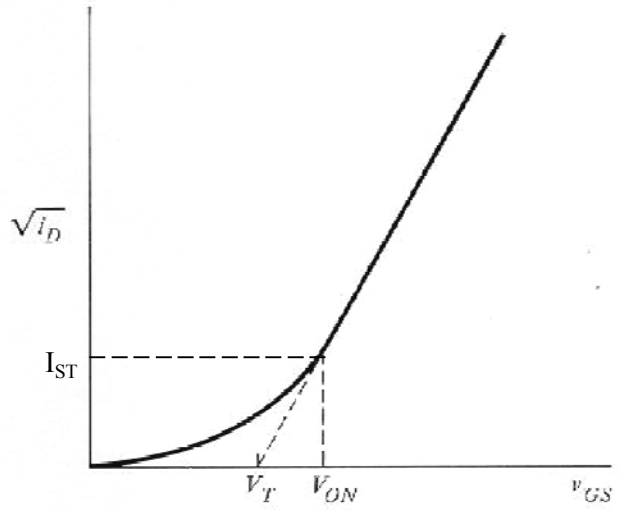
where

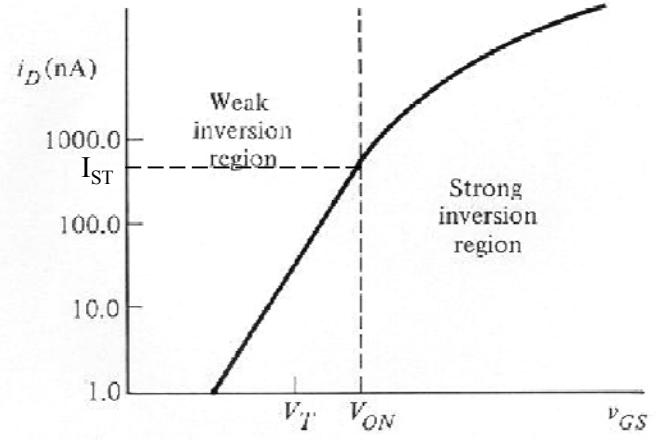
$$I_{d0} = 2n\left(\frac{kT}{q}\right)^2 e^{-\frac{qV_T}{nkT}}$$

- Used in very low power designs
- Highest g_m/I_d
- Slow operation

(n is the slope factor)

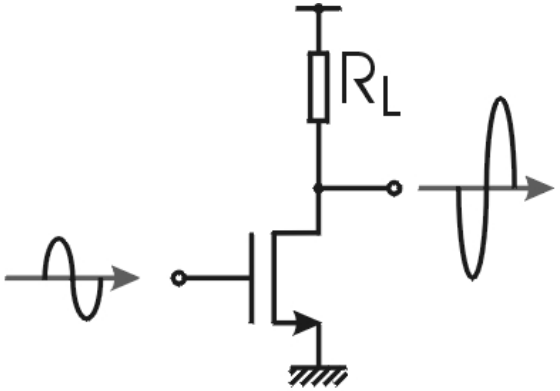
$$\frac{1}{n} = 1 - \frac{\gamma}{2\sqrt{V_g - V_{T0} + \left(\frac{\gamma}{2} + \sqrt{\Psi_0}\right)^2}}$$





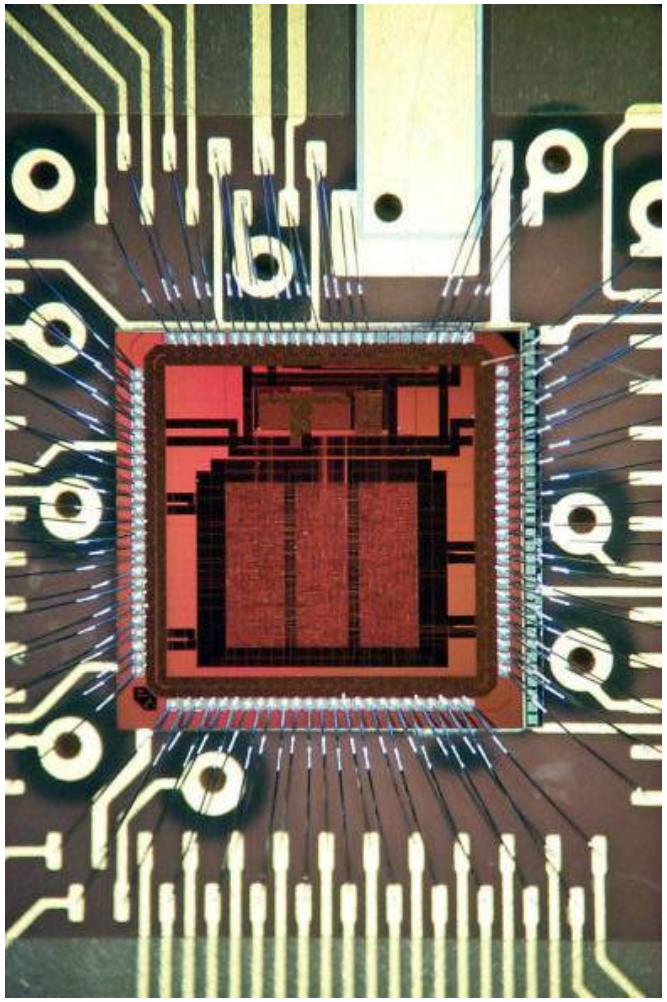
Gain & Inversion

- Gain:
  - Signal regeneration at every logic operation
  - "Static" flip-flops
    - -> clock can be turned off
  - "Static" RW memory cells
- Inversion:
  - Intrinsic to the common-source configuration
- The gain cell load can be:
  - Resistor
  - Current source
  - Another gain device (PMOS)



Outline

- Introduction
- Transistors
  - *DC behaviour*
  - *MOSFET capacitances*
  - *MOSFET model*
- The CMOS Inverter
- Technology Scaling
- Gates
- Sequential circuits
- Storage elements



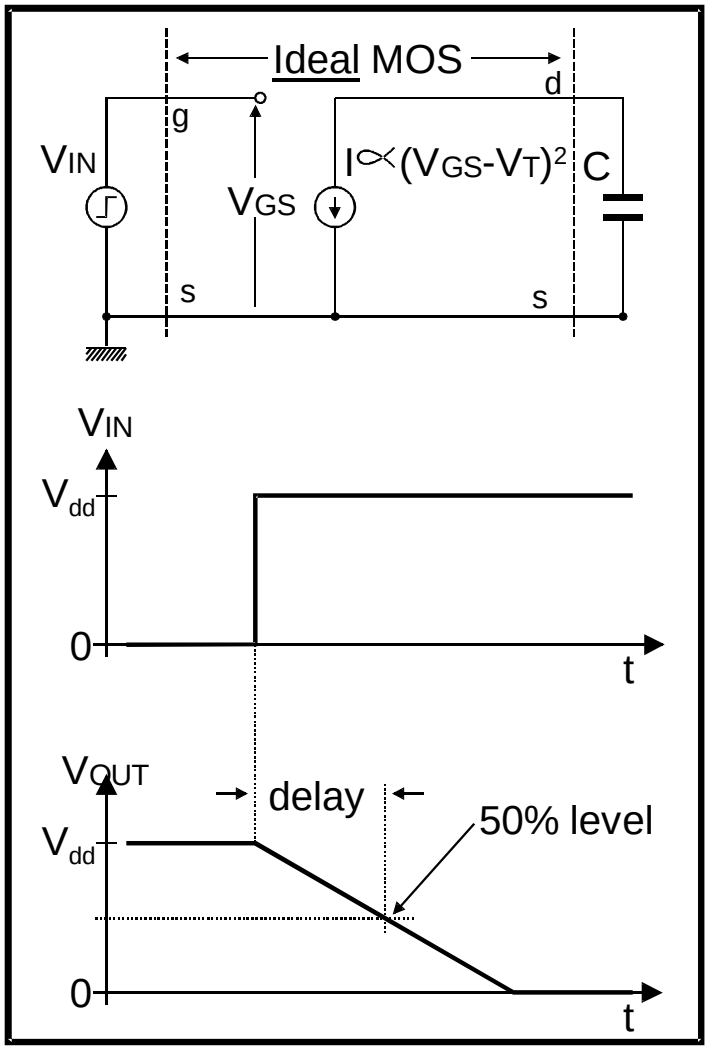
What causes delay?

- In MOS circuits capacitive loading is the main cause of delay.
- Capacitance loading is due to:
  - Device capacitance
  - Interconnect capacitance

    (RC delay in the interconnects will be addressed later)

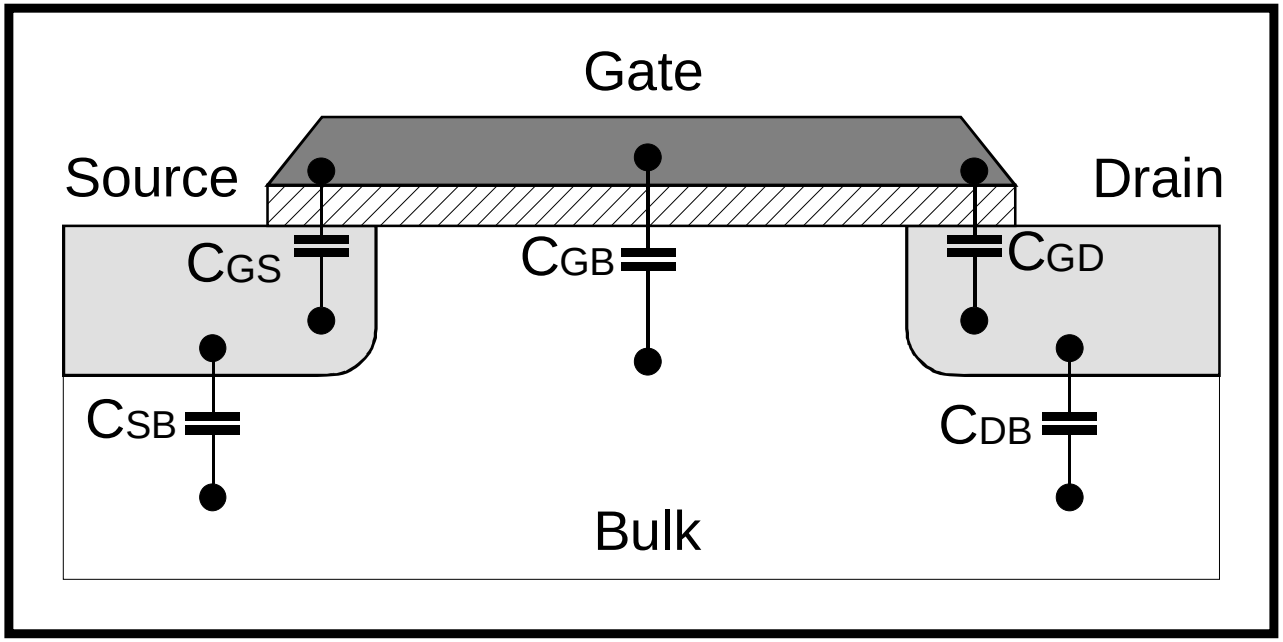
$$\Delta t = C \cdot \frac{\Delta V}{I} \approx \frac{C}{\mu \cdot C_{ox} \cdot V_{dd}} \cdot \frac{L}{W}$$

Assuming $V_T = 0$



MOSFET capacitances

- MOS capacitances have three origins:
  - The basic MOS structure
  - The channel charge
  - The pn-junctions depletion regions



MOS Structure Capacitances

- Source/drain diffusion extend below the gate oxide by:

  x_d - the lateral diffusion

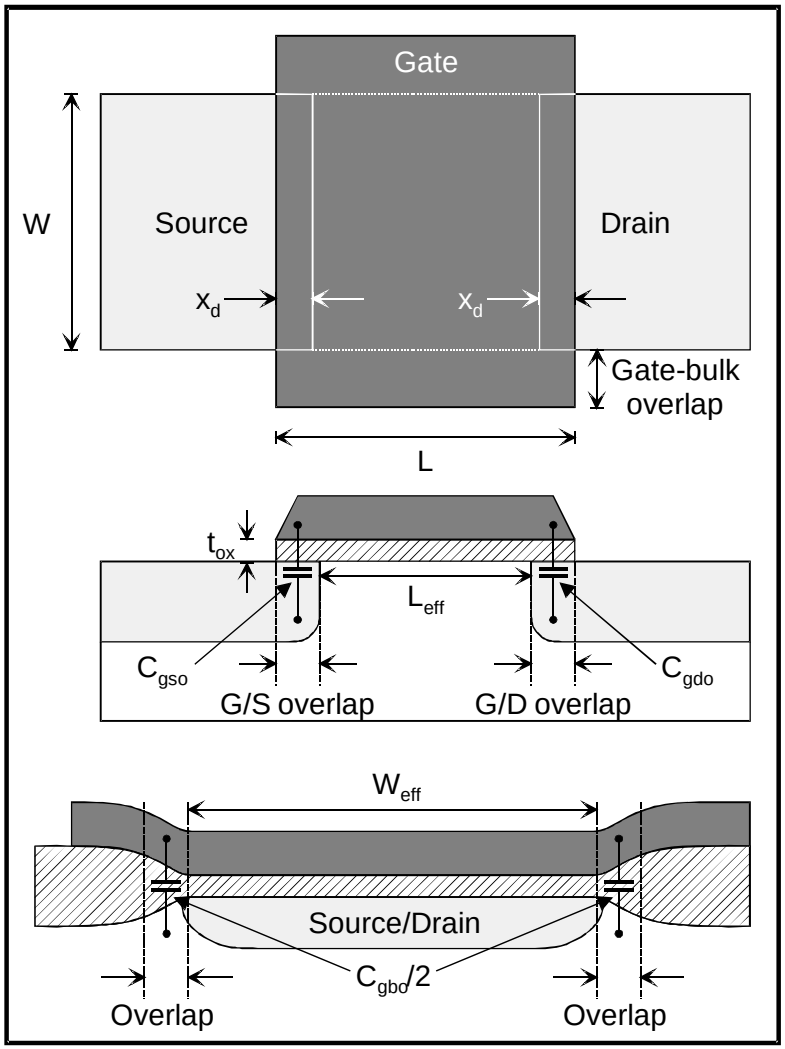
- This gives origin to the source/drain overlap capacitances:

$$C_{gso} = C_{gdo} = C_o \times W$$

$$C_o \ (\mathrm{F/m})$$

- Gate-bulk overlap capacitance:

$$C_{gbo} = C_o' \times L, \quad C_o' \ (\mathrm{F/m})$$



MOS Structure Capacitances

0.24 μm process

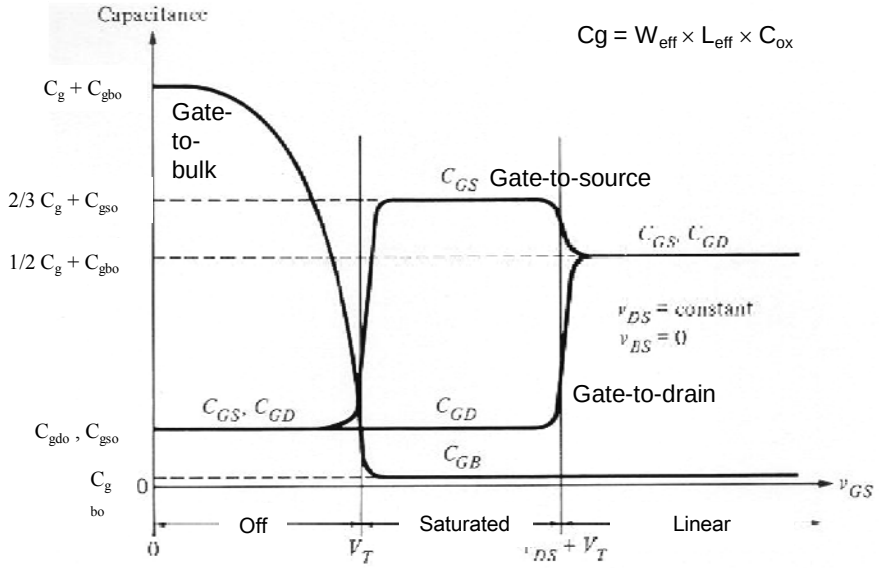
NMOS
L(drawn) = 0.24 μm
L(effective) = 0.18 μm
W(drawn) = 2 μm
C_o (s, d, b) = 0.36 fF/μm
C_{ox} = 5.6 fF/μm^2

$C_{gso} = C_{gdo}$ = 0.72 fF
C_{gbo} = 0.086 fF

C_g = 2.02 fF

Channel Capacitance

- The channel capacitance is nonlinear
- Its value depends on the operation region
- Its formed of three components:
  - C_{gb} - gate-to-bulk capacitance
  - C_{gs} - gate-to-source capacitance
  - C_{gd} - gate-to-drain capacitance



Operation region	C_{gb}	C_{gs}	C_{gd}
Cutoff	C_{ox} W L	0	0
Linear	0	(1/2) C_{ox} W L	(1/2) C_{ox} W L
Saturation	0	(2/3) C_{ox} W L	0

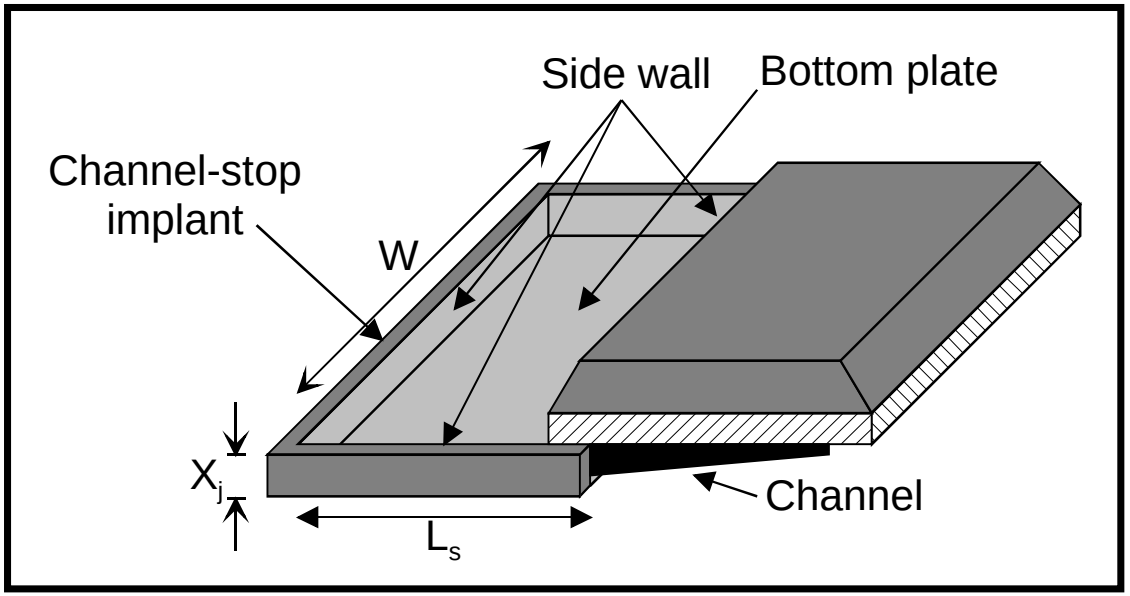
Junction capacitances

- C_{sb} and C_{db} are diffusion capacitances composed of:
  - Bottom-plate capacitance:

$$C_{bottom} = C_j \cdot W \cdot L_s$$

  - Side-wall capacitance:

$$C_{sw} = C_{jsw} \cdot (2\,L_s + W)$$



Junction capacitances

0.24 μm process

NMOS
L(drawn) = 0.24 μm
L(effective) = 0.18 μm
W(drawn) = 2 μm
L_s = 0.8 μm
C_j (s, d) = 1.05 fF/μm^2
C_{jsw} = 0.09 fF/μm

C_{bottom} = 1.68 fF
C_{sw} = 0.32 fF

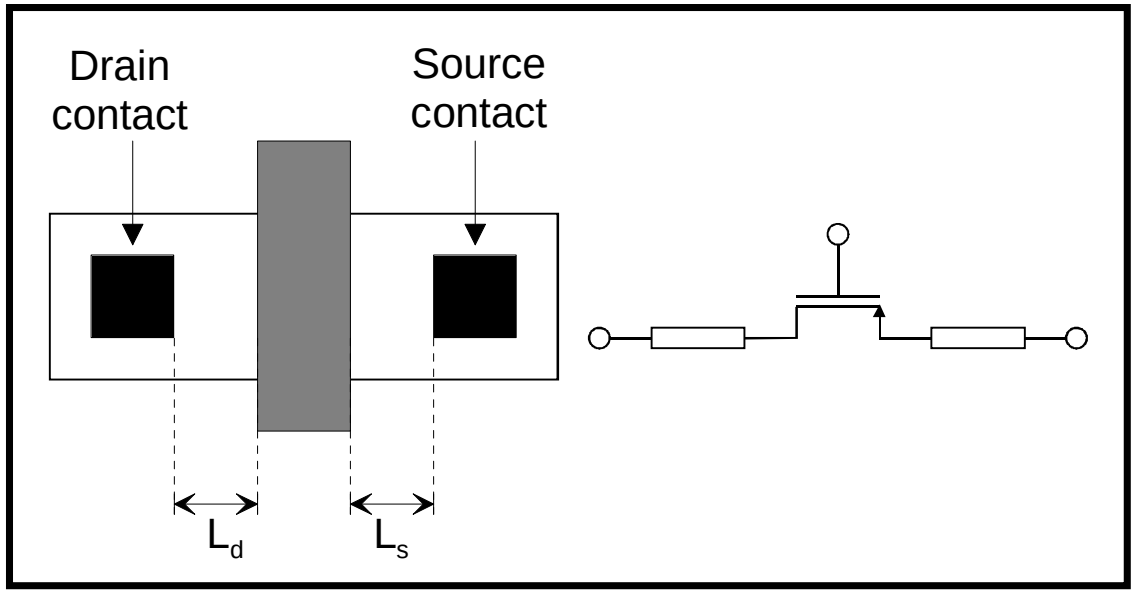
C_g = 2.02 fF

Source/drain resistance

- Scaled down devices $\Rightarrow$ higher source/drain resistance:

$$R_{s,d} = \frac{L_{s,d}}{W} \cdot R_{sq} + R_c$$

- In sub-μ processes *silicidation* is used to reduce the source, drain and gate parasitic resistance



0.24 μm process

R (P+) = 4 Ω/sq
R (N-) = 4 Ω/sq