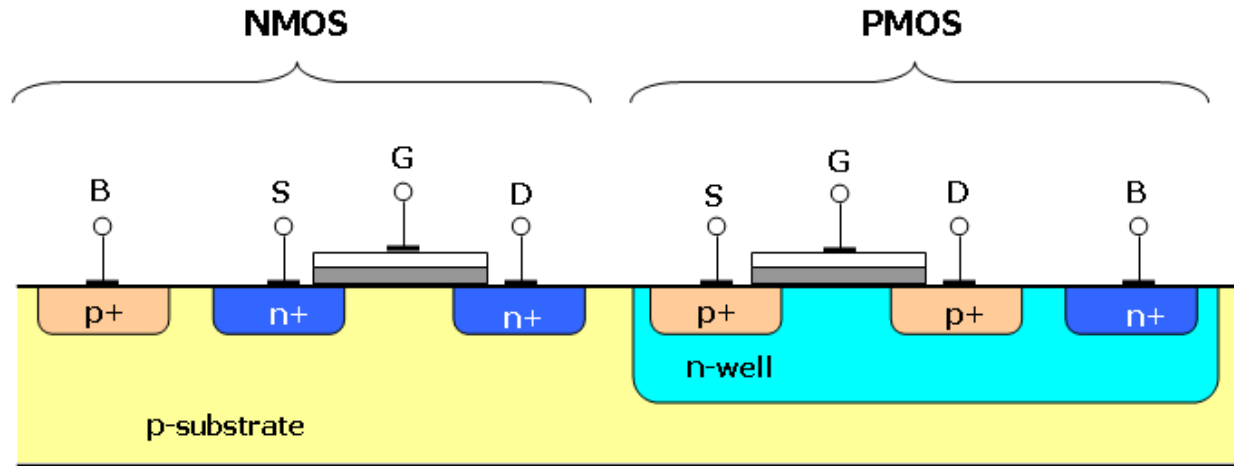


Devices III Outline

- Complementary Metal Oxide Semiconductor (CMOS)
- Applying MOSFETs as switches
 - > CMOS logic gate
 - > Static power dissipation
 - > Dynamic power dissipation
- Downscaling theory and Moore's law
 - > Capacitance
 - > Intrinsic delay
 - > Power dissipation

Complementary MOS (CMOS)



Fabrication TECHNOLOGY where complementary pairs of p-channel and n-channel MOSFETs (PMOS, NMOS) are fabricated together on the same substrate to construct integrated circuits (IC).

The substrate is p-type based on which n+ regions are doped for the source and drain of the NMOS.

For the PMOS, an n-well is doped to form the required n-type body of the PMOS. Within the n-well, p+ regions are doped for the source and drain of the PMOS.

Applications of CMOS

- Digital circuits:
 - > Microprocessors
 - > CPUs
 - > Microcontrollers
 - > Static RAMs
- Analogue circuits:
 - > Image sensors
 - > Transceivers in communications (Transceiver: Combined functions of Transmitter & Receiver)
 - > Application specific integrated circuits (ASICs)

Analogue versus Digital

Analogue (UK) / Analog (US):

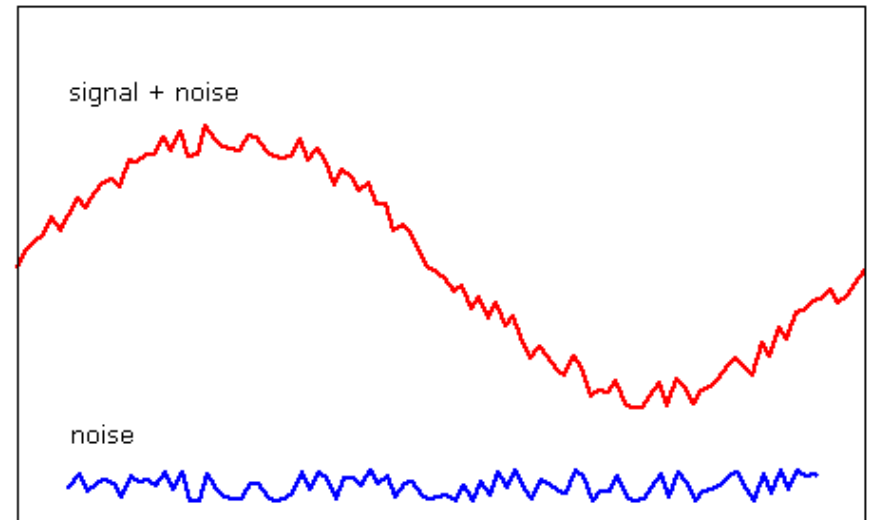
An analogue signal is any continuous electrical signal that varies with time.

Any information may be conveyed by an analogue signal

Often a measured response to changes in physical phenomena (e.g. sound, light, temperature, position).



Noise degrades quality of the signal

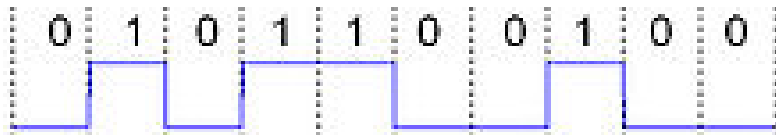


Analogue versus Digital

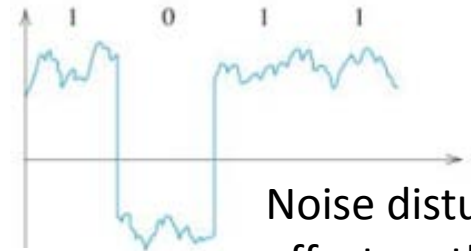
Digital:

A digital signal is a signal that represents a sequence of discrete values. A logic signal is a digital signal with *only two possible values*, and describes an arbitrary bit stream.

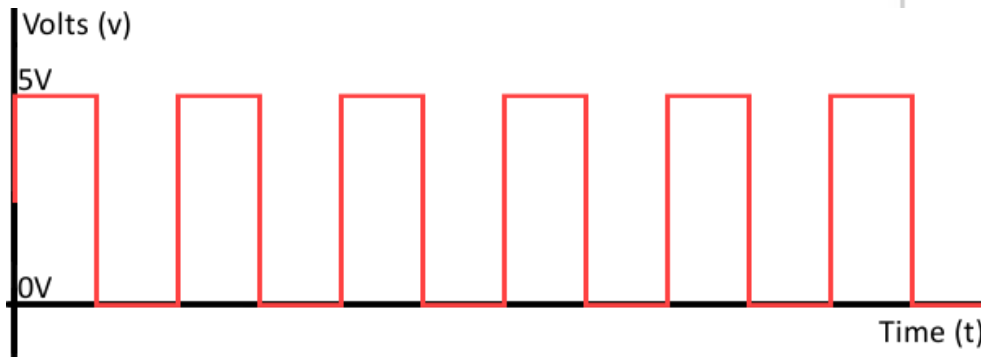
With digital signals, system noise, provided it is not too great, will not change the quantification whereas with signals handled using analogue processing, noise always degrades the operation to some degree.



Bit stream: sequence of two possible values



Noise disturbance has little effect on the digits



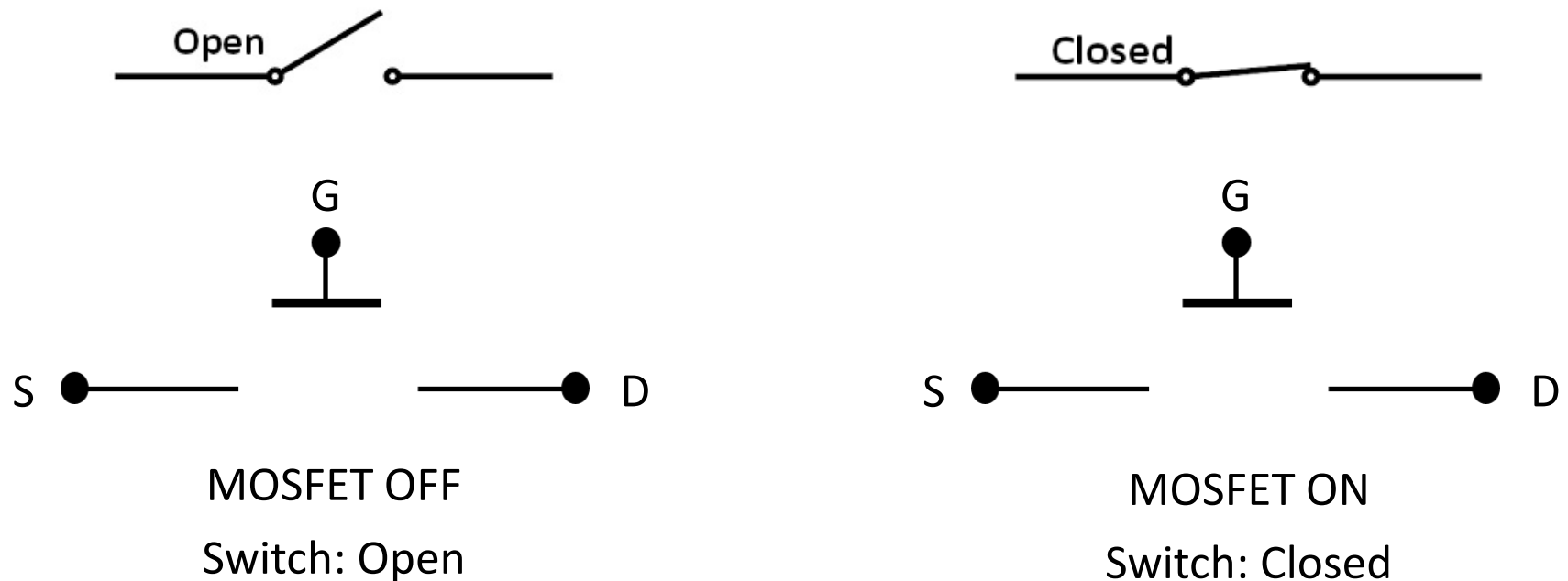
Hi: 5V (typically the supply voltage)

Lo: 0V

Applying MOSFETs as switches

MOSFETs are used as the building blocks for logic gates (subject of EE1001)

To build logic gates, the MOSFETs function as switches (turn ON/OFF)



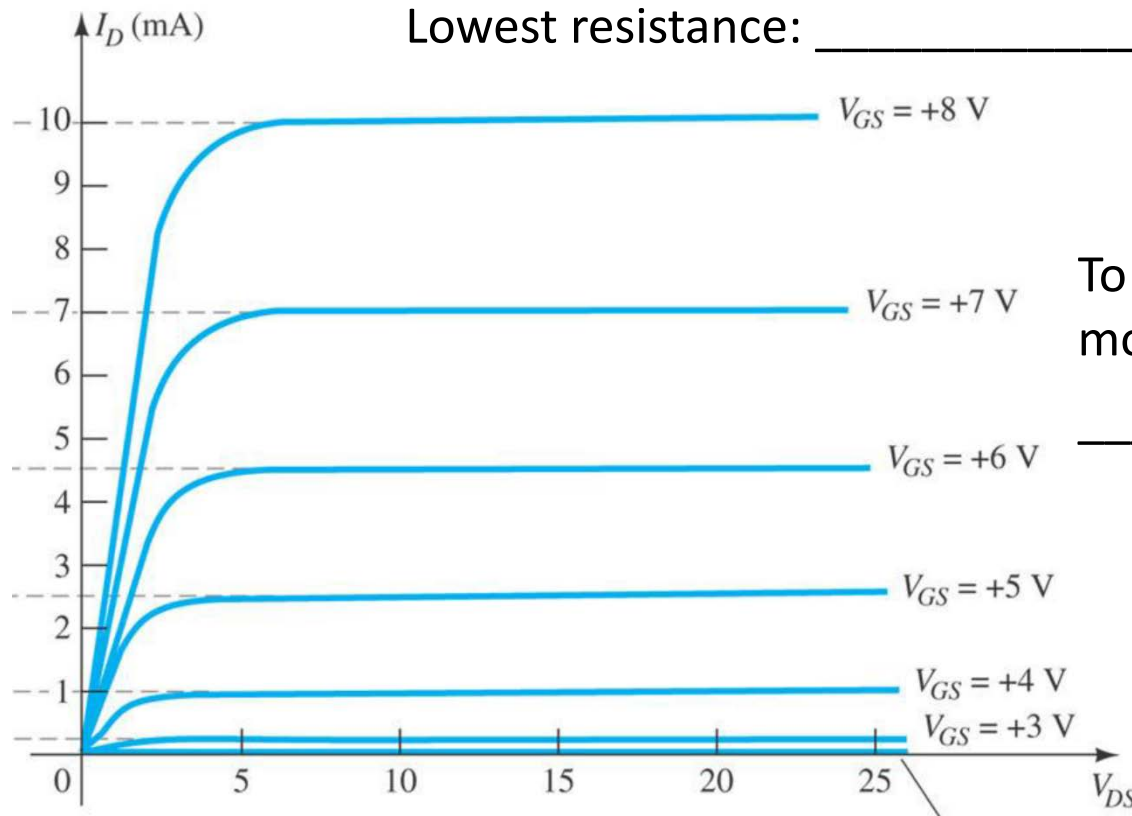
To be an effective closed switch, R_{DS} should be _____
and V_{DS} should be _____

Operating modes for switches

To realize a closed switch, which operating mode would give the lowest possible channel resistance? _____

Note: Resistance given by the inverse of the slope of I_D vs V_{DS}

Lowest resistance: _____ slope



To realize an open switch, which mode would give an open circuit?

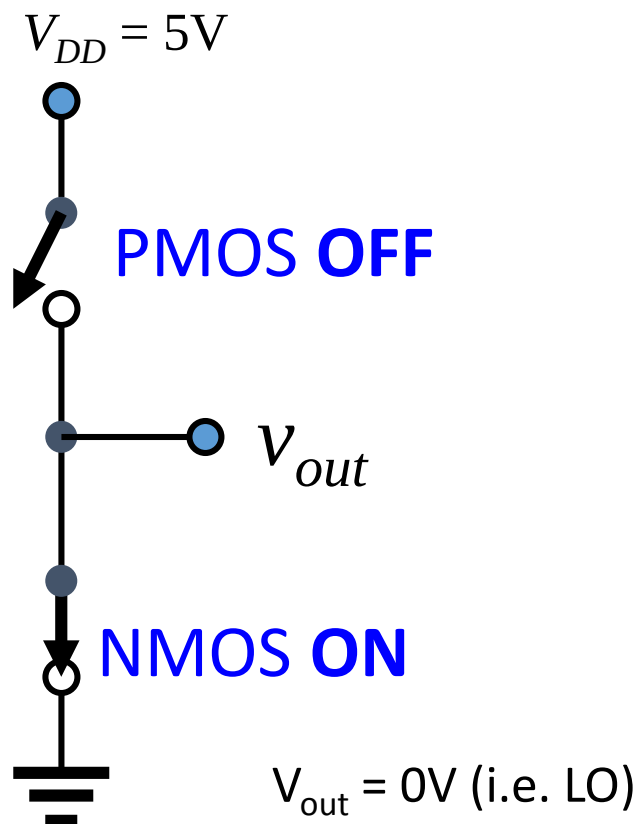
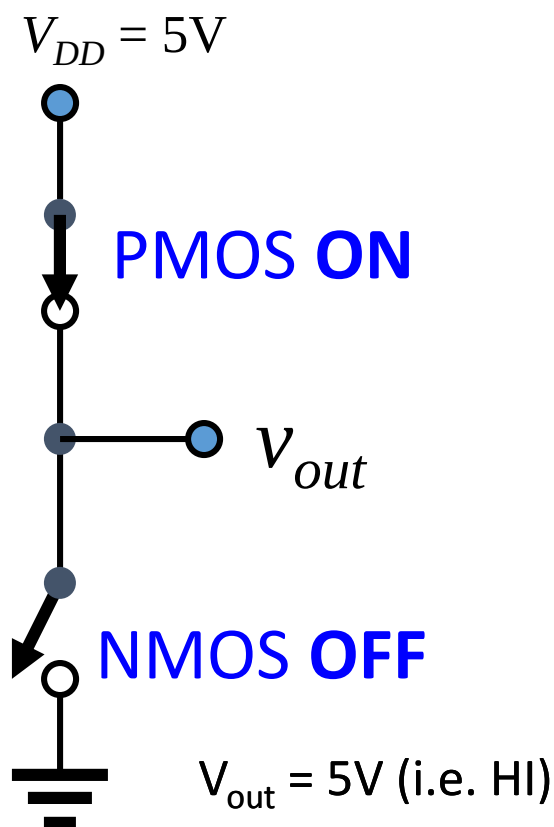
Complementary action

When the NMOS is OFF, the PMOS is ON

When the NMOS is ON, the PMOS is OFF

Never both off nor both on

ON: MOSFET is in **triode/ohmic** mode (R_o is very low)

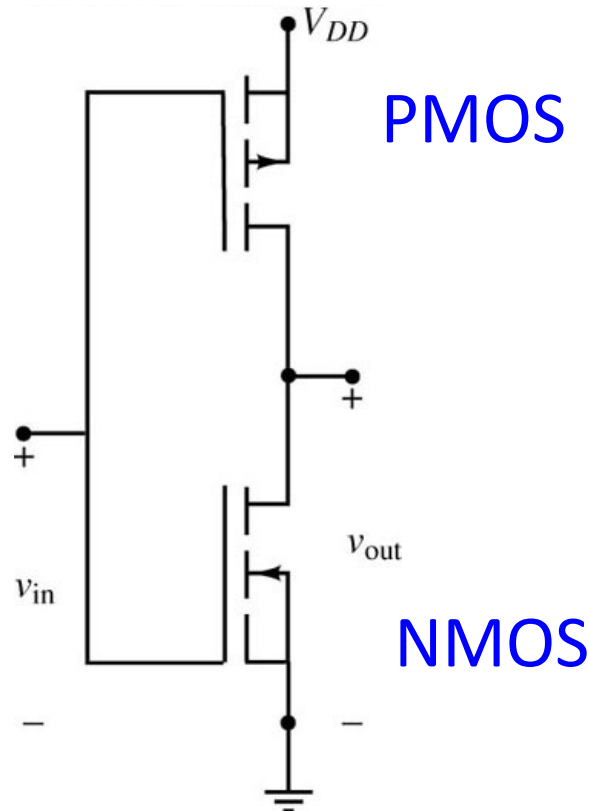


CMOS logic

CMOS: Complementary MOS (PMOS and NMOS pairs)

The simplest logic gate is the NOT gate (requires 1 PMOS and 1 NMOS)

NOT gate: also called the Inverter



Gates (GP, GN): connected together
Forms the input

Drains (DP, DN): connected together
Forms the output

PMOS source (SP): Power supply (V_{DD})

NMOS source (SN): Ground

Logic function and MOS state

Assume $|V_T| = 1V$

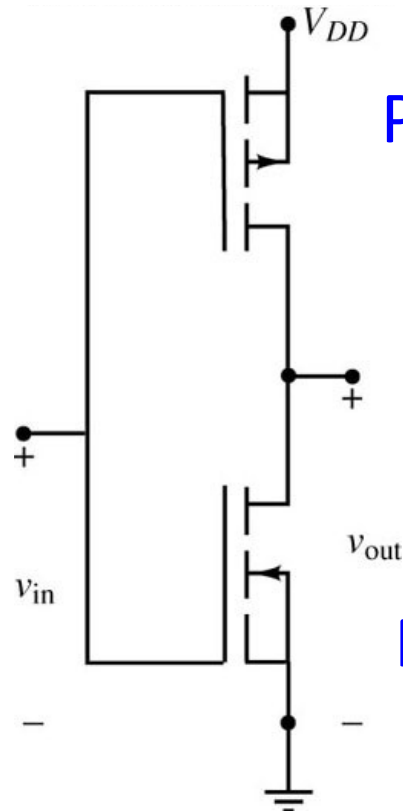
$V_{TN} = 1V$

$V_{TP} = -1V$

When $V_{in} = 0V$ (i.e. LO)

$V_{GSN} = \underline{\hspace{2cm}}$ ($< V_T$): NMOS is OFF (Cut-off)

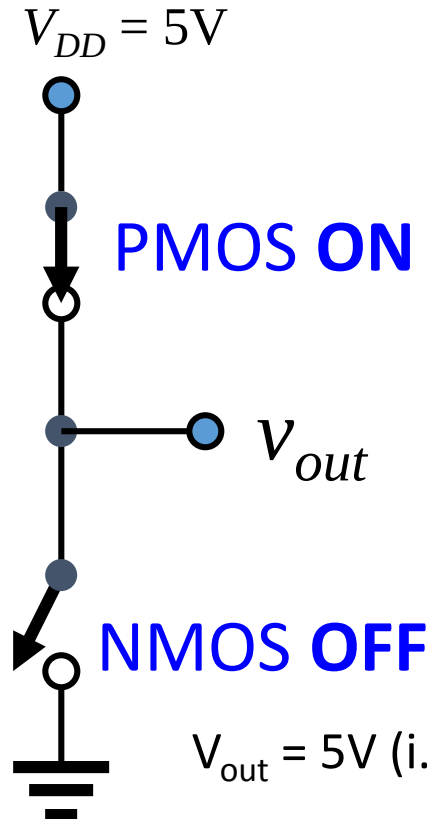
$V_{SGP} = \underline{\hspace{2cm}}$ ($> V_T$): PMOS is ON (Triode)



PMOS

NMOS

When V_{in} is LO,
 V_{out} is HI



$V_{out} = 5V$ (i.e. HI)

Logic function and MOS state

Assume $|V_T| = 1V$

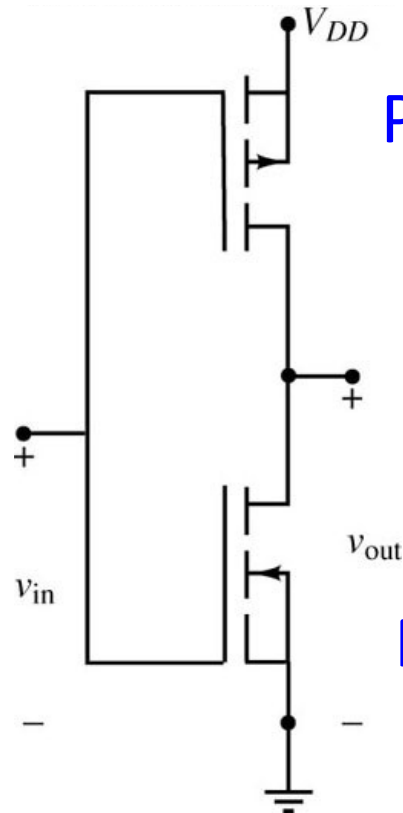
$V_{TN} = 1V$

$V_{TP} = -1V$

When $V_{in} = 5V$ (i.e. HI)

$V_{GSN} = \underline{\hspace{2cm}}$ ($> V_T$): NMOS is ON (Triode)

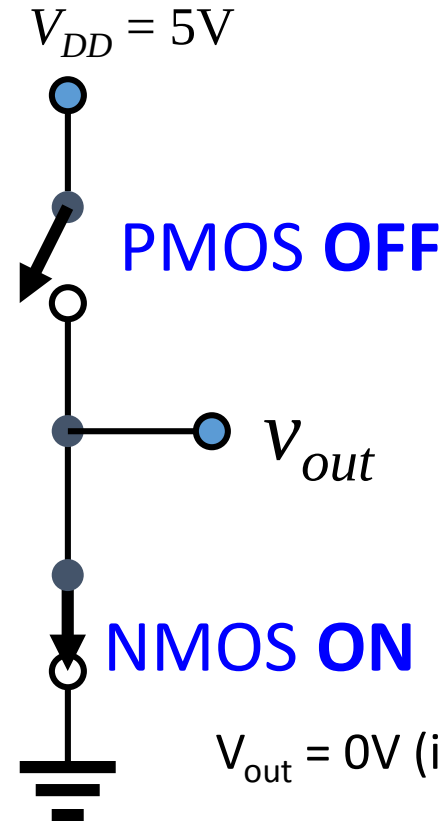
$V_{SGP} = \underline{\hspace{2cm}}$ ($< V_T$): PMOS is OFF (Cut-off)



PMOS

NMOS

When V_{in} is HI,
 V_{out} is LO



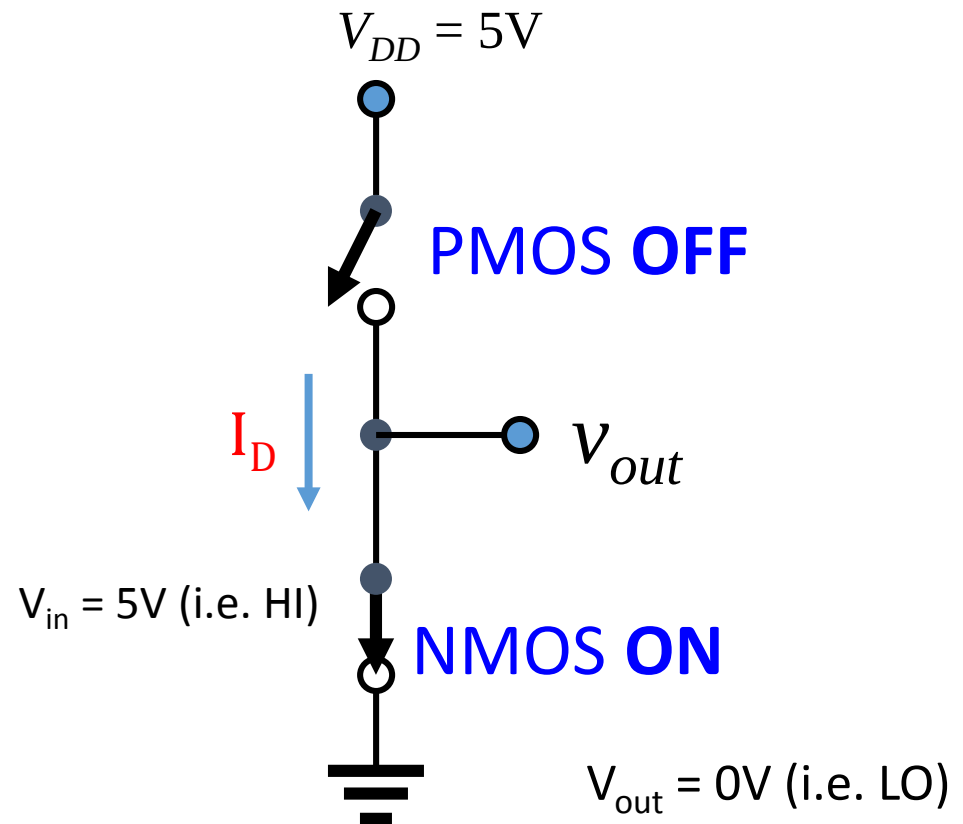
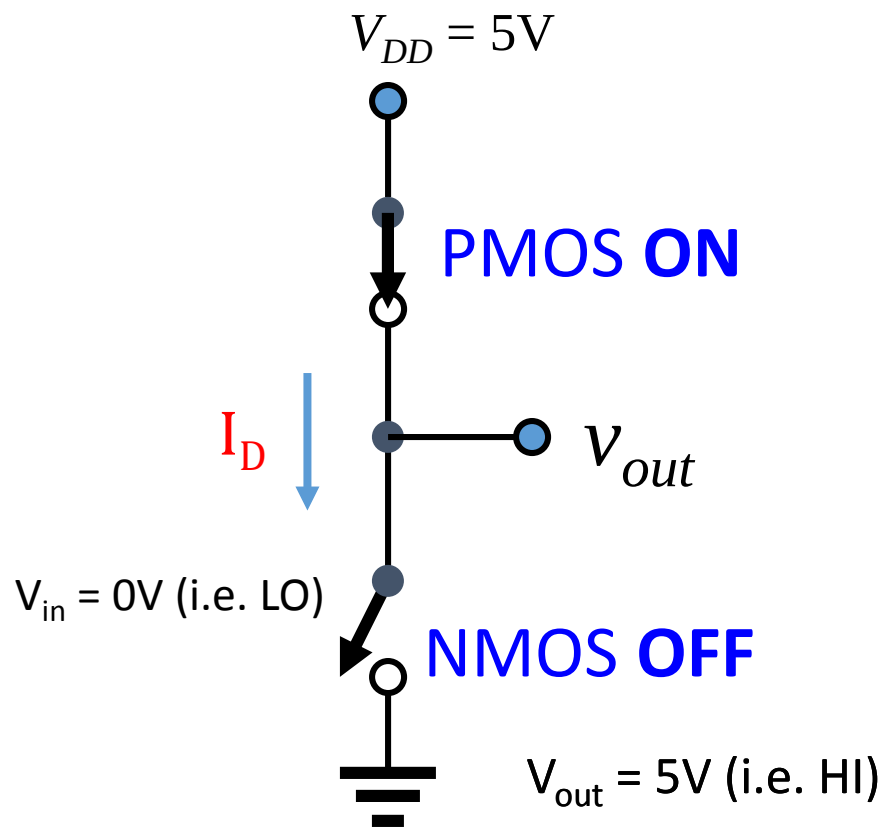
$V_{out} = 0V$ (i.e. LO)

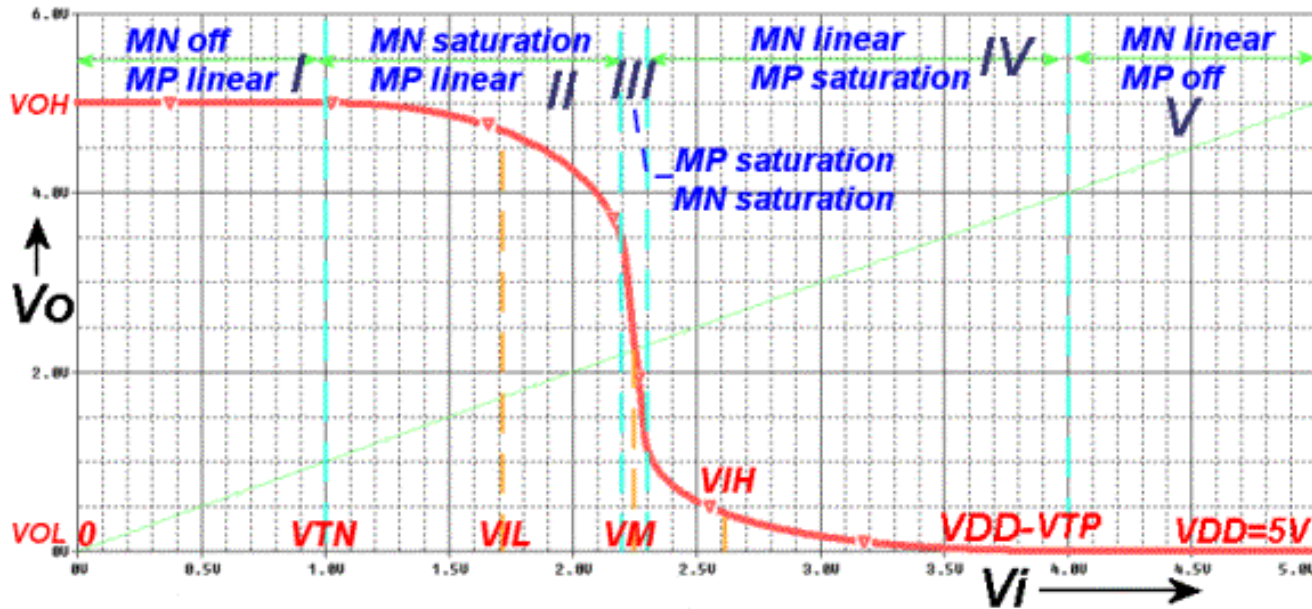
Static power dissipation

In either case, you can see that I_D is always zero (in the ideal case)

Hence the power dissipation in the static case is zero

This is a huge attractive feature of CMOS logic





MN: NMOS

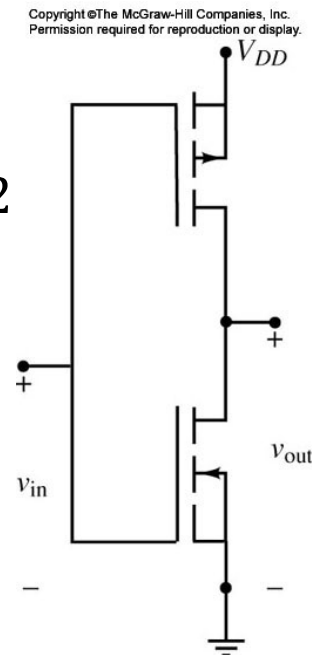
MP: PMOS

At V_M , $V_{in} = V_{out}$

NMOS: $I_D = K_N (V_{GS} - V_T)^2$ If $K_N = K_P$ and $V_{TN} = -V_{TP}$: $V_M = V_{DD}/2$

PMOS: $I_D = K_P (V_{GS} - V_T)^2$ When $V_{in} = V_{out} = V_{DD}/2 = 2.5V$:

	V_{GSN}, V_{SGP}	V_{GDN}, V_{DGP}	Mode
NMOS			Saturation
PMOS			Saturation



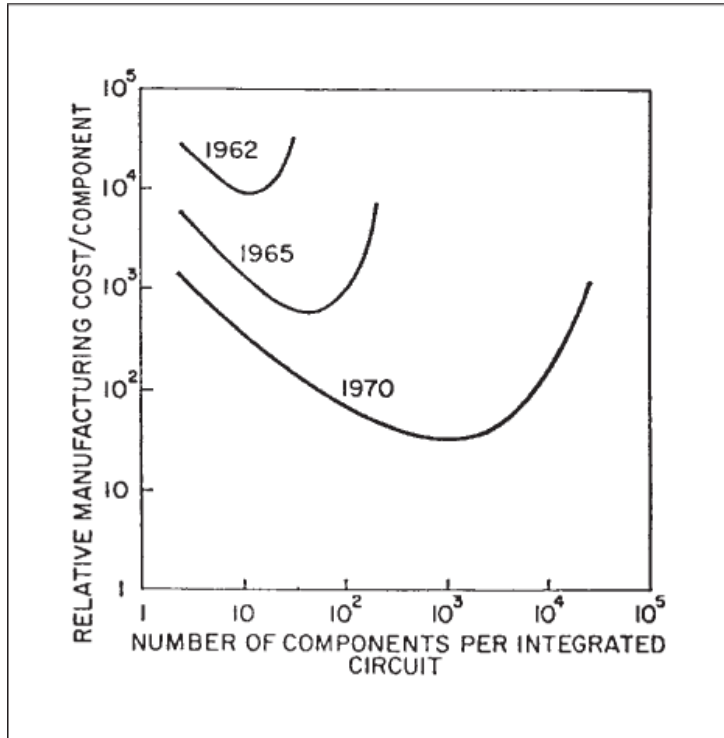
Notes on power consumption

- Due to the complementary action, no current flows through the transistors when the logic gate output is in logic 1 or 0 state
 - > No power consumed in MOSFETs when $V_{in} = 0V$ or V_{DD}
 - > This the ideal case; there will be some small leakage current
 - > i.e. static power dissipation is ideally zero
- In the ideal case, power is only consumed when in the transition state (when the MOSFET switch between on/off); i.e. dynamic power dissipation
- Smaller supply voltage (V_{DD}) helps to reduce power consumption
 - > Smaller threshold voltage allows for smaller supply voltage

Scaling of transistor size

Why do we cram as many transistors as possible on a chip?

What advantage do we get from making transistors smaller?



Graph published in 1965 by
Gordon Moore (founder of Intel)

Prediction made in 1965 based on data from recent years:

Observation: Number of components on a chip manufactured at minimum cost doubled every year

Manufacturing cost per component was falling from year to year

Prediction: Trend would continue into the future

Moore's law: Number of transistors on a chip produced at minimum cost doubles every 18 months

Why make transistors smaller?

Pack more transistors on the same area

Build more sophisticated systems with more transistors

When transistors are smaller:

Transistors can switch faster

Transistors consume less power

V_T is reduced

Smaller supply voltage

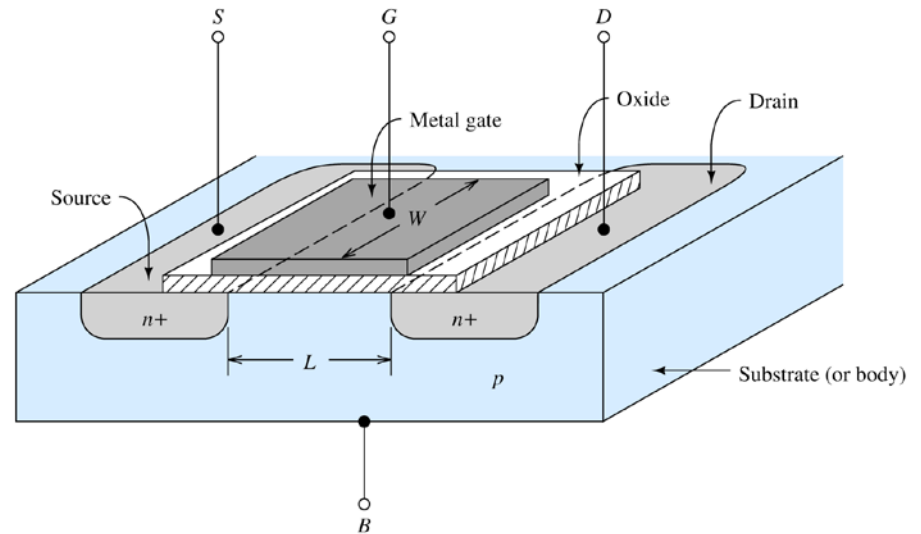


Figure 12.1 n-channel enhancement MOSFET showing channel length L and channel width W .

Manufacturing cost per transistor comes down with time

Faster and more powerful microchips with minimum increase in the cost per chip

Downscaling theory

Constant field scaling

Device parameter	Symbol	Scaling factor
Gate length	L	α
Gate width	W	α
Gate oxide thickness	t_{ox}	α
Supply voltage	V_{DD}	α
Threshold voltage	V_{Th}	α
Doping concentration	N_A, N_D	$1/\alpha$

If $\alpha = 1/2$, then the parameter is half the original value

Example: If the original gate length is L , then scaled down gate length is αL

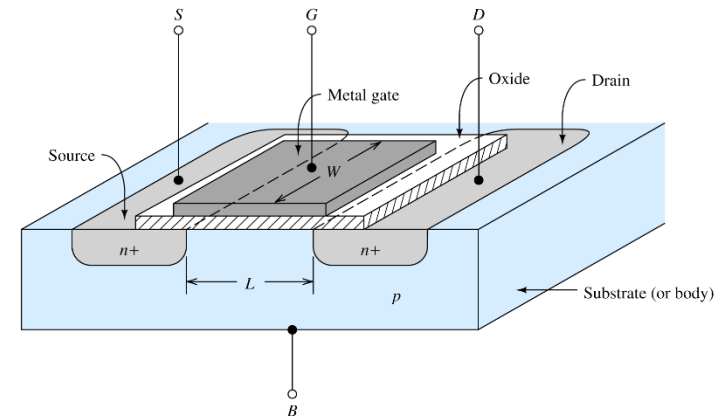
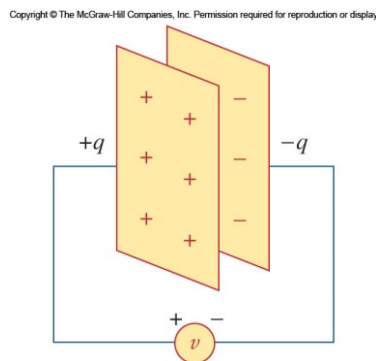
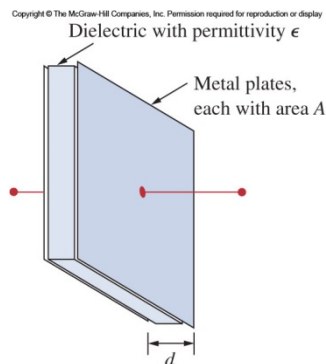


Figure 12.1 n-channel enhancement MOSFET showing channel length L and channel width W .

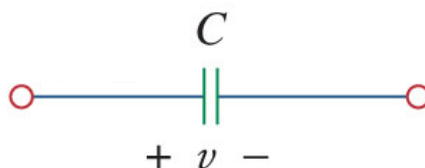
Capacitance

From MOSFETs, we make a small detour to look at the concept of **capacitance**

Capacitor: Formed by two conductor plates separated by an insulator



Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display



Symbol for a capacitor is two parallel lines separated by a gap

Capacitor is usually symbolized by the letter "C"

Unit: Farad (F)

When a voltage is applied across the plates, opposite charge pairs are stored on the plates.

The charges do not flow between the plates due to the insulator (i = 0 when the plates are fully charged up)

Capacitance relates the voltage and charge across a capacitor:

$$Q = CV$$

Capacitance and charging

$$Q = CV$$



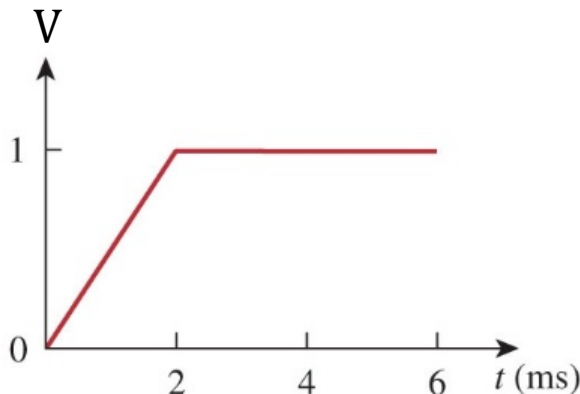
$$i = \frac{dQ}{dt} = C \frac{dV}{dt}$$

If we take the time derivative, we get an expression that relates current to voltage in a capacitor

Recall: Current is defined as the rate of change of charge

So we now see that where C is constant, current depends on the rate of change of voltage (i.e. dV/dt)

The figure below shows the voltage applied across a $0.1\mu\text{F}$ cap. Work out the corresponding current in the cap.



For the first 2ms, when V rises at a rate of $1\text{V}/2\text{ms}$:

$$I = (0.1\mu)(1/2\text{m}) = 50\mu\text{A}$$

Note that charge (Q) rises proportionally with V during this charging stage.

After 2ms, I drops to zero as V is constant and Q reaches $0.1\mu\text{C}$

Capacitance and switch time

What affects the time required to charge a capacitor?

Apply KVL around the circuit loop:

$$V_s = V + IR$$

$$V_s = V + RC \frac{dV}{dt}$$

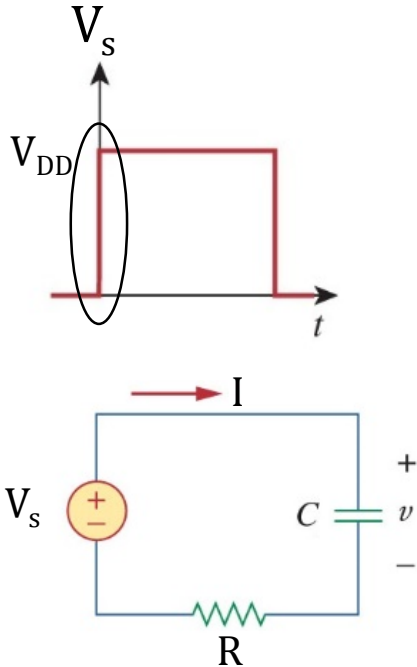
This is a First Order Ordinary Differential Equation, for which the general solution takes the following form:

$$V = Ae^{-\frac{t}{RC}} + B \quad \text{A and B are constants to be determined depending on the specific case}$$

Initially (when $t = 0$), we know that $V = 0$

Finally (after a long time of $t \rightarrow \infty$), we that $V \rightarrow V_{DD}$ (current stops flowing and the capacitor is fully charged)

We make the substitutions into the general solution



Rise time

When $t = 0$, $V = 0$:

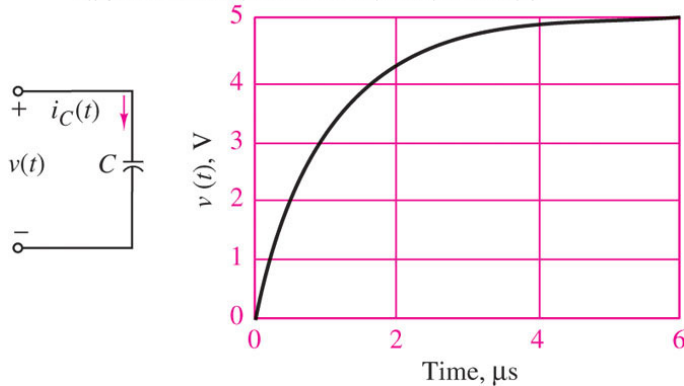
$$\Rightarrow A + B = 0$$

When $t \rightarrow \infty$, $V \rightarrow V_{DD}$:

$$\Rightarrow B = V_{DD} \Rightarrow A = -V_{DD}$$

$$V = V_{DD}(1 - e^{-t/(RC)})$$

Copyright©The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



Make the following substitutions:

$$\text{When } t = RC: V = 0.63V_{DD}$$

$$\text{When } t = 5RC: V = 0.995V_{DD}$$

RC is known as the time constant

We can see from above that RC controls the duration required to charge up a cap

A smaller C and R reduces the time constant

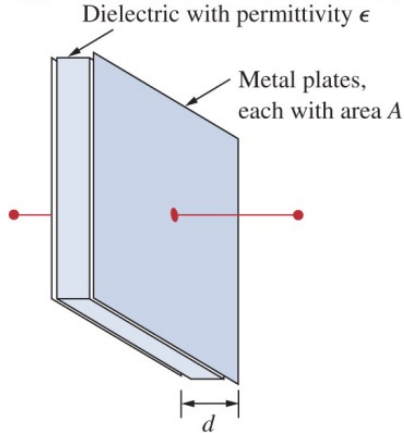
The same can be said about discharging a capacitor

RC time constant controls the delay time required to switch on/off a MOSFET

Capacitance

We can work out the capacitance of a parallel plate for a given insulator:

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display



$$C = \frac{\epsilon_0 \epsilon_r A}{d}$$

Overlapping area of plates: A

Thickness of the insulator: d

Permittivity in vacuum (universal): $\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$

Relative permittivity / dielectric constant of insulator: ϵ_r

(Absolute) permittivity of insulator: $\epsilon_0 \epsilon_r$

Material	ϵ_r
Air	~ 1
Water	~ 80
Silicon Dioxide	3.9

Gate capacitance

$$C_{gate} = \frac{\epsilon_0 \epsilon_r W L}{t_{ox}}$$

Given $L = 200\text{nm}$, $W = 800\text{nm}$, $t_{ox} = 50\text{nm}$, work out the gate capacitance.

$$C_{gate} = (8.854 \times 10^{-12})(3.9)(200\text{n})(800\text{n})/(50\text{n}) = 0.1\text{fF}$$

C_{ox} : oxide capacitance per unit area

$$C_{ox} = \frac{\epsilon_0 \epsilon_r}{t_{ox}} \quad \text{This is required to find the conductance parameter, } K$$

Can you see how a capacitor is formed at the gate of the MOSFET?

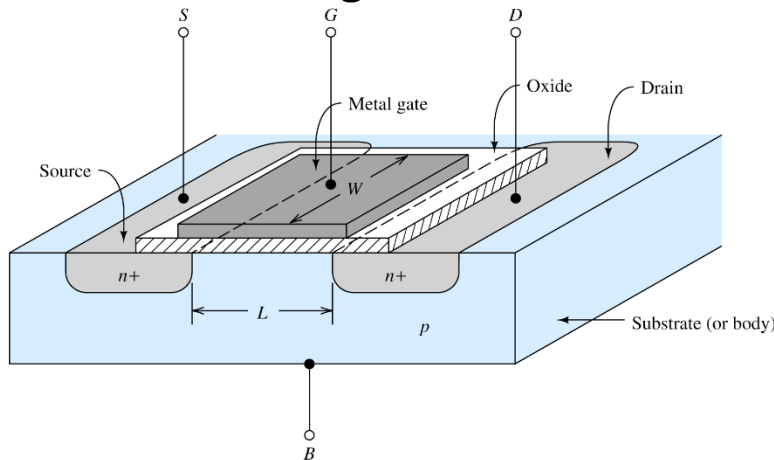
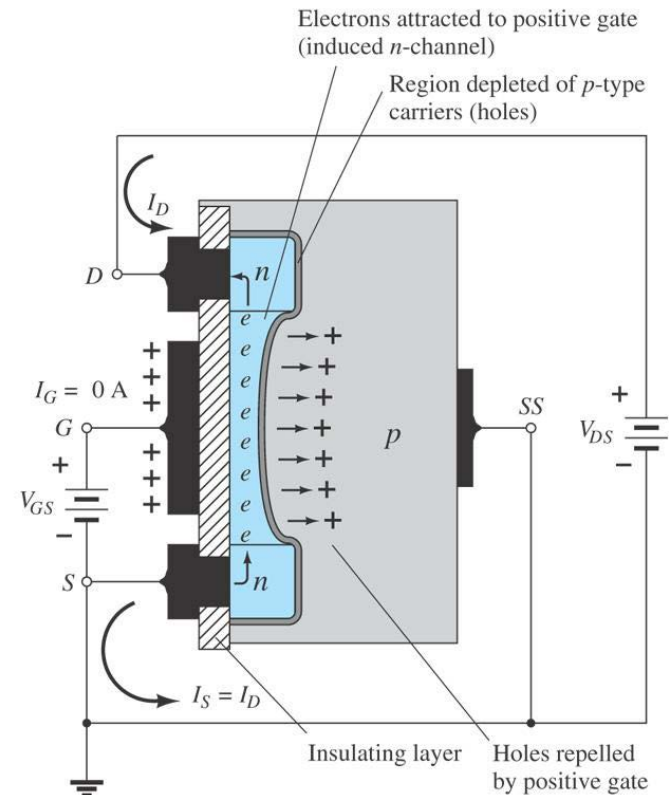


Figure 12.1 n -channel enhancement MOSFET showing channel length L and channel width W .



Results of downscaling

Performance of device	Symbol	Expression	Scaling factor
Number of devices per unit area	N_{dev}		$1/\alpha^2$
Gate oxide capacitance	C_{gate}		α
Drain current	I_D		α
Intrinsic delay per device	τ	$C_{\text{gate}} V_{DD}/I_D$	α
Power dissipation per device	P	$I_D V_{DD}$	α^2
Power density	P_{dens}	$I_D V_{DD} N_{\text{dev}}$	1
Transit frequency (analogue)	f_T		$1/\alpha$

Intrinsic delay per device scales down with device size:

Faster switching times

Power density per unit area constant with device size:

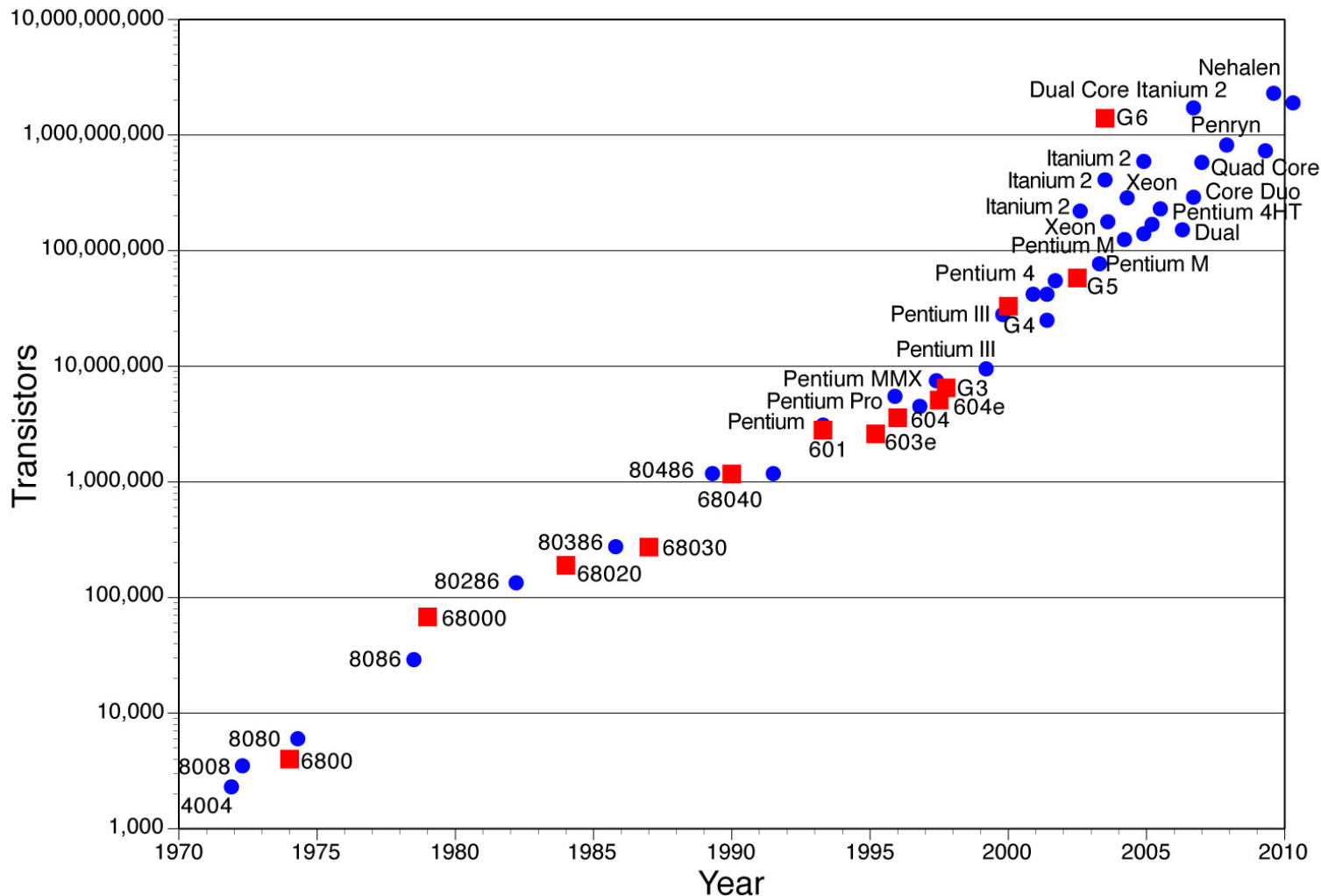
More devices for increased functionality with no cost in power budget

Evolution of transistors over time

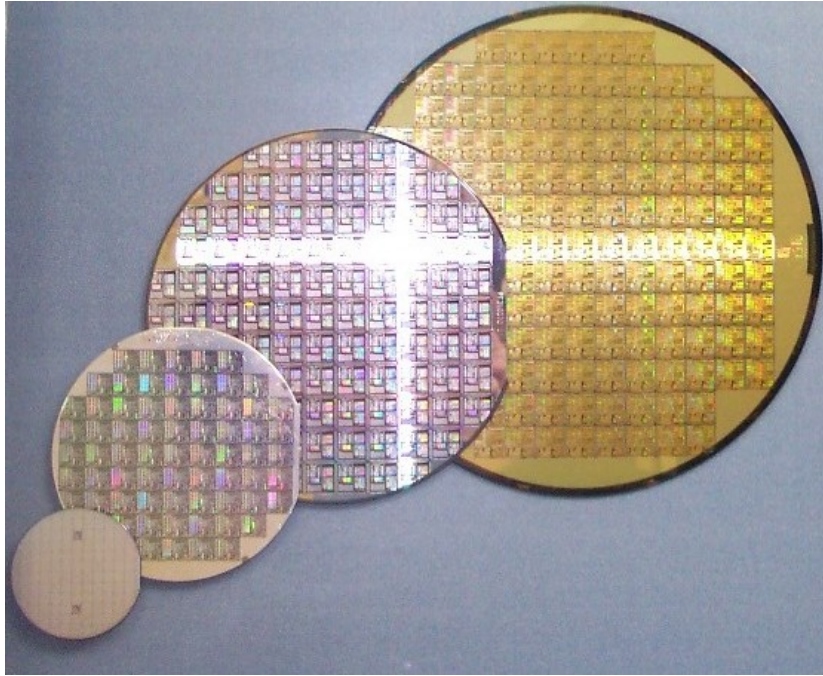
Today: 7 nm

Technology Nodes

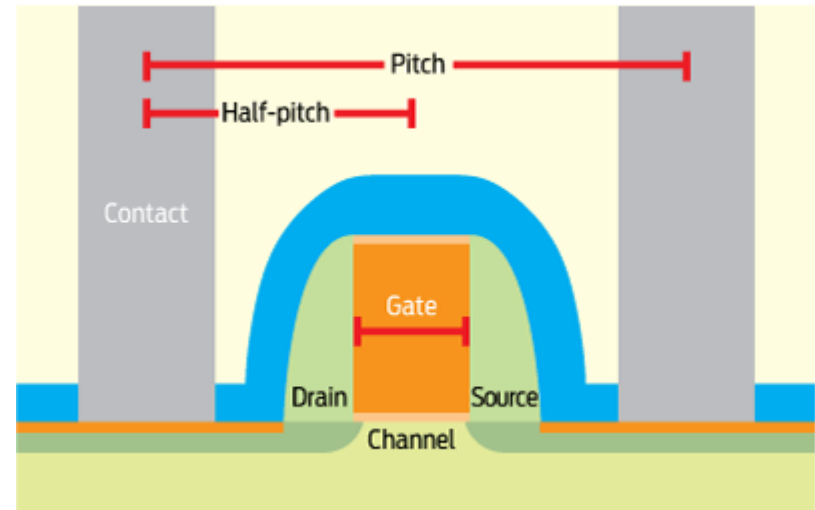
10 μm – 1971
6 μm – 1974
3 μm – 1977
1.5 μm – 1982
1 μm – 1985
800 nm – 1989
600 nm – 1994
350 nm – 1995
250 nm – 1997
180 nm – 1999
130 nm – 2001
90 nm – 2004
65 nm – 2006
45 nm – 2008
32 nm – 2010
22 nm – 2012
14 nm – 2014



Technology nodes and wafers



2-inch (51 mm), 4-inch (100 mm), 6-inch (150 mm), and 8-inch (200 mm) wafers
State-of-art: 12-inch wafers



The technology node (e.g. 22nm) refers to the average half-pitch

Two things happening:

- 1) Scaling down of device size
- 2) Increasing size of wafers for processing devices (with the same guaranteed reliability and yield)