CG2027: Transistor-level Digital Design
Sem I, AY2021

Introduction:
Challenges in Digital Circuit Design

Tuesdays, 15:00 – 18:00, E-Learning on Zoom

Lecturer: Prof. Kelvin Fong Xuanyao

Office: E4-05-23

Email: kelvin.xy.fong@nus.edu.sg

NUS National University of Singapore

CG2027 Course Flow – I

- Lectures
  - Tue. (15:00 – 18:00), Zoom
    - Check LumiNUS for Meeting ID and Password
  - Videos and accompanying slides on LumiNUS for self-paced learning
    - Quizzes are embedded within videos to enhance learning
  - In-meeting discussions to reinforce learning
- Office hours:
  - Mon. & Thu. (14:00 – 15:00), or by appointments
- Grading
  - Participation (2%): punctual completion of embedded quizzes AND assignments
  - Lesson Video Quizzes (18%),
  - Tutorials & Assignments (30%),
  - Final Quiz (50%): Week 7 (format TBD)
- Assignments (released at the same time as lecture)
  - LumiNUS submission (due at 18:00 on Sunday at the same weekend)
  - 10% deduction per day for late turn-in (maximum 2 days)
  - No credit after 2 days or when the solutions are uploaded to LumiNUS

CG2027 Course Flow – II

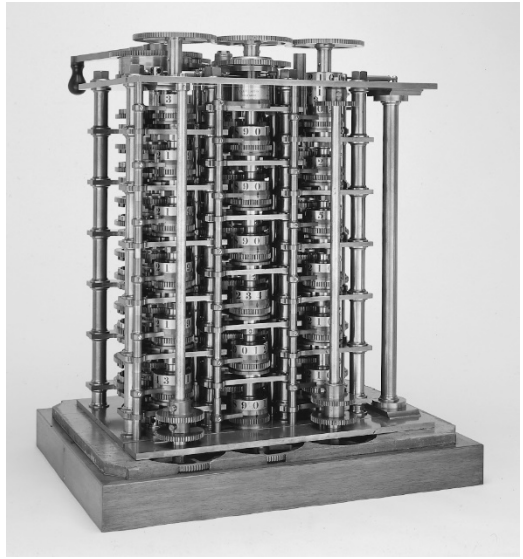
- Tutorials (F2F) are for discussing Assignment solutions
  - T01
    - Mon. (13:00 – 14:00), E1-06-01, Prof. Jerald Yoo
  - T02
    - Mon. (14:00 – 15:00), E1-06-01, Prof. Hong Minghui
  - T03
    - Thu. (11:00 – 12:00), E1-06-01, Prof. Jerald Yoo
  - T04
    - Thu. (13:00 – 14:00), E1-06-01, Prof. Hong Minghui

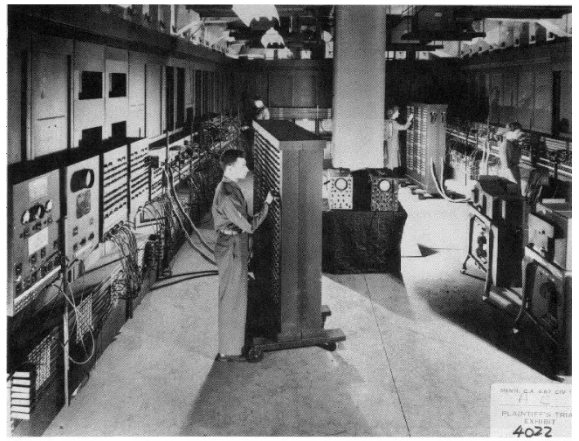
Course Overview

- Basic operating principles of a transistor
  - Types of carriers, diode and MOSFET
  - Device characteristics
  - Basic fabrication process
- Transistors for switches and logic applications
  - Single transistor switching circuit, NMOS, PMOS
  - Transmission gate, CMOS logic, inverter
  - Logic functions into CMOS gates
- Design parameters and issues
  - Power and energy dissipation, sizing, delay, rise/fall time
  - Designing CMOS logic gates, fan-in/fan-out

Early Computer History

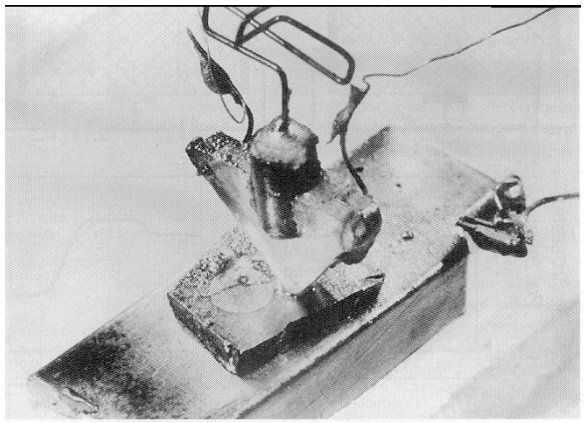
The Babbage Difference Engine (1832)
25,000 parts
Mechanical
Punched cards
Cost: £17,470

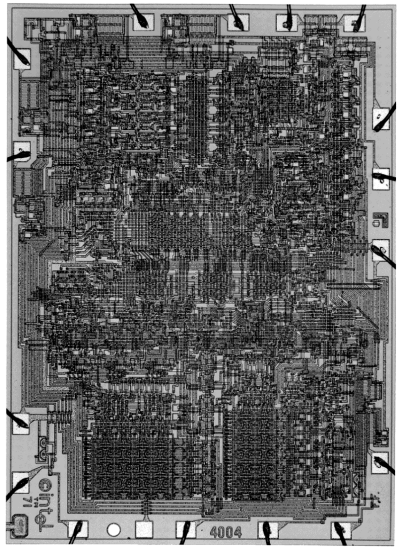




ENIAC (1946)
First electronic computer
Reliability and power dissipation problems

First transistor
Bell Labs, 1948



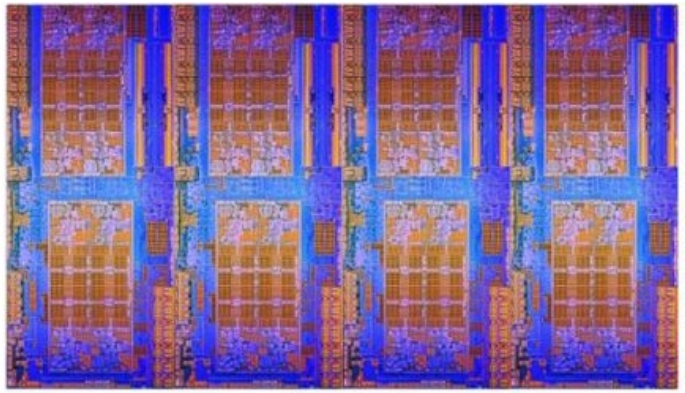


1971
2,300 transistors
0.1 MHz operation
4b buses
10u process

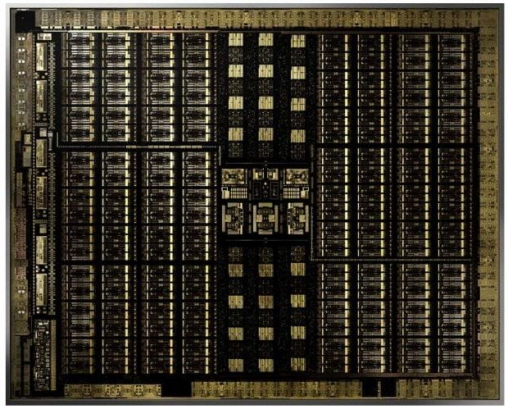
Modern Processors

AMD Epyc (32 Cores)

7nm FinFET
32 billion FET



Nvidia Turing (TU102)

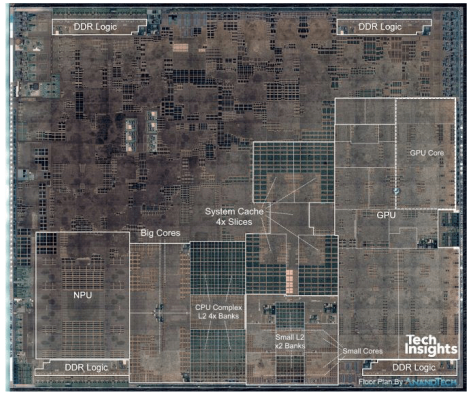


12nm FinFET
18.6 billion FET

Intel Skylake-X (28 Cores)

14nm FinFET
8 billion FET

Apple A12 Bionic



7nm FinFET
6.9 billion FET

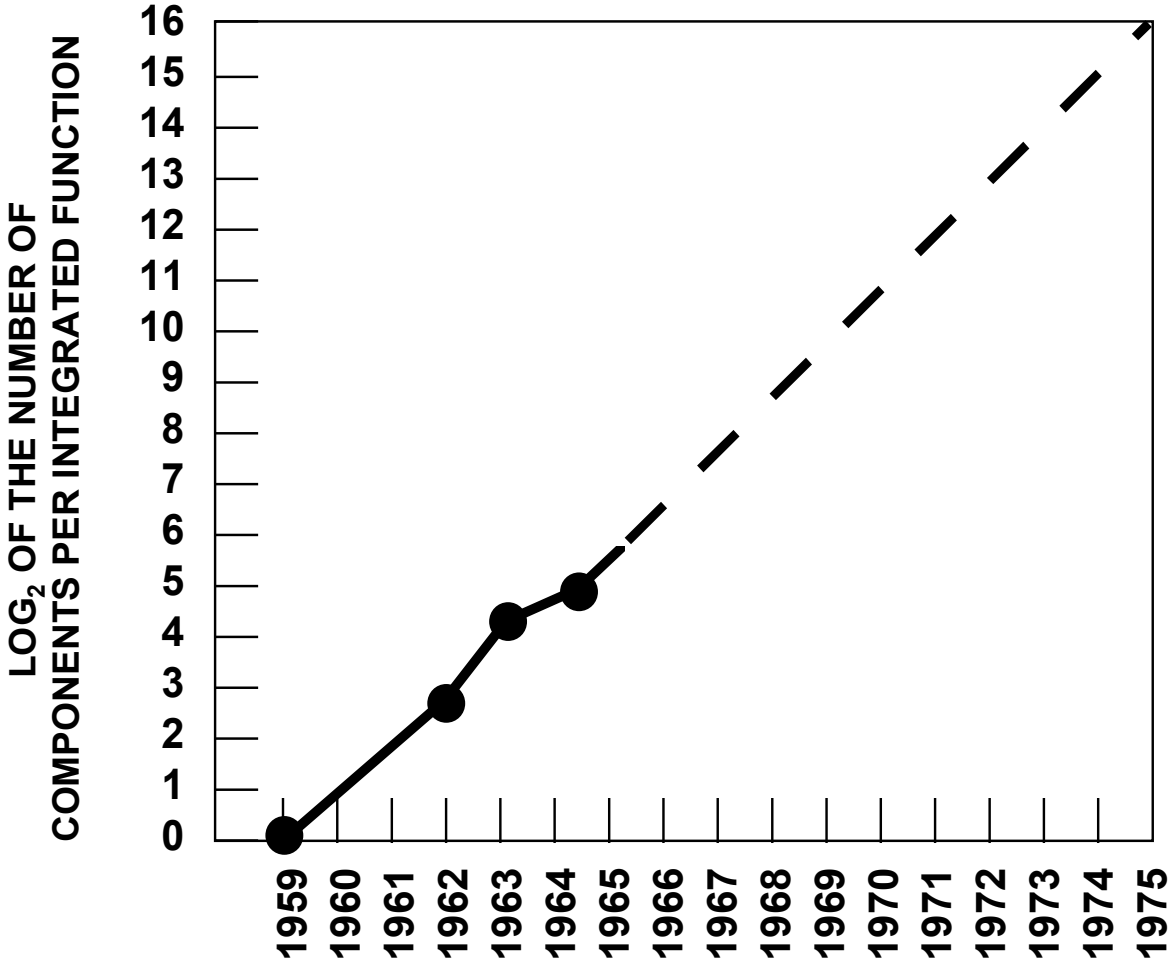
Question:
How do we build modern processors?

Thought Question:
Why do we need to know about transistor-level digital circuits?

Moore's Law

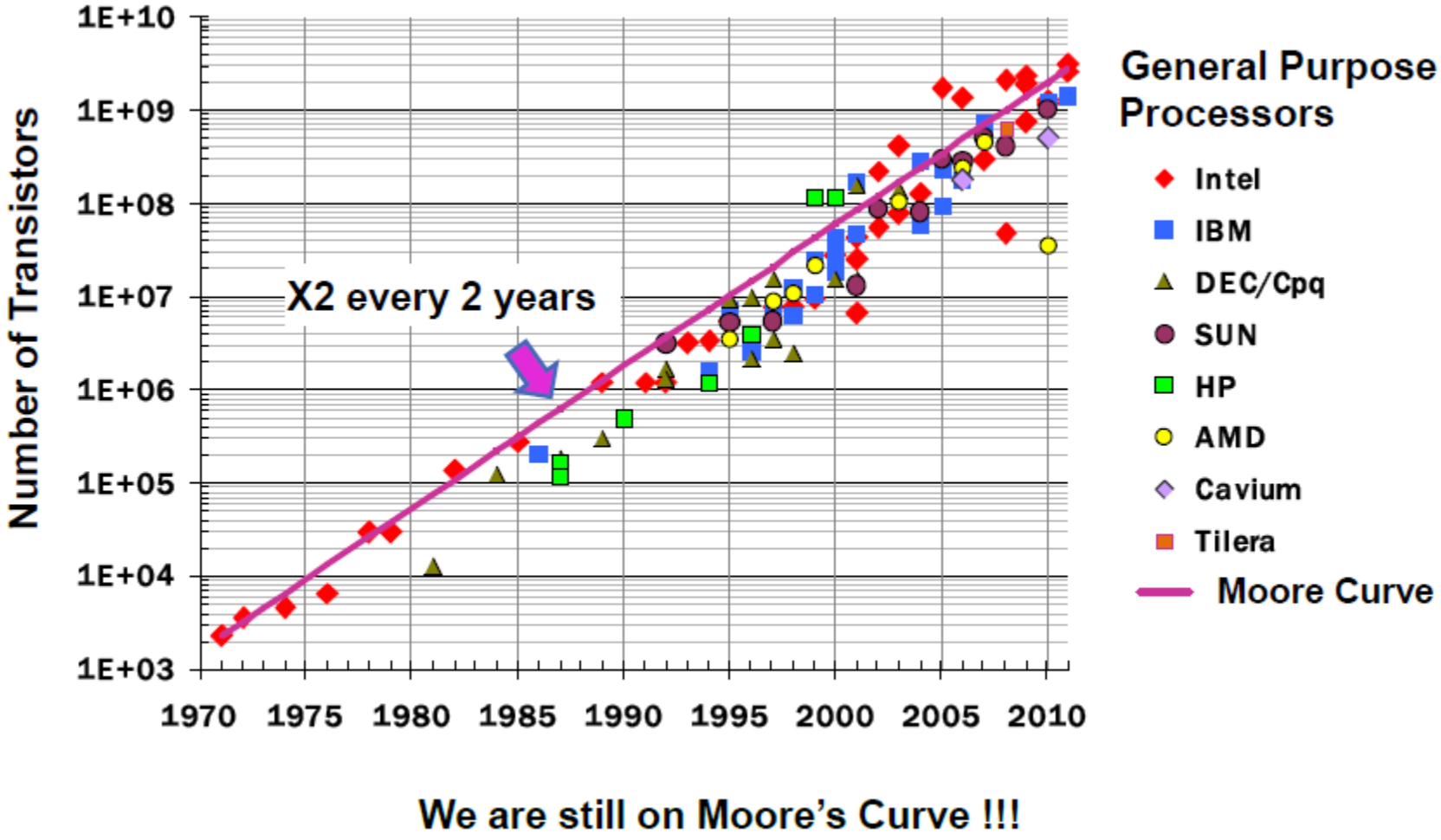
- In 1965, Gordon Moore noted that the number of transistors on a chip doubled every 18 to 24 months.
- Semiconductor technology will double its effectiveness every 18 months

Electronics, April 19, 1965.



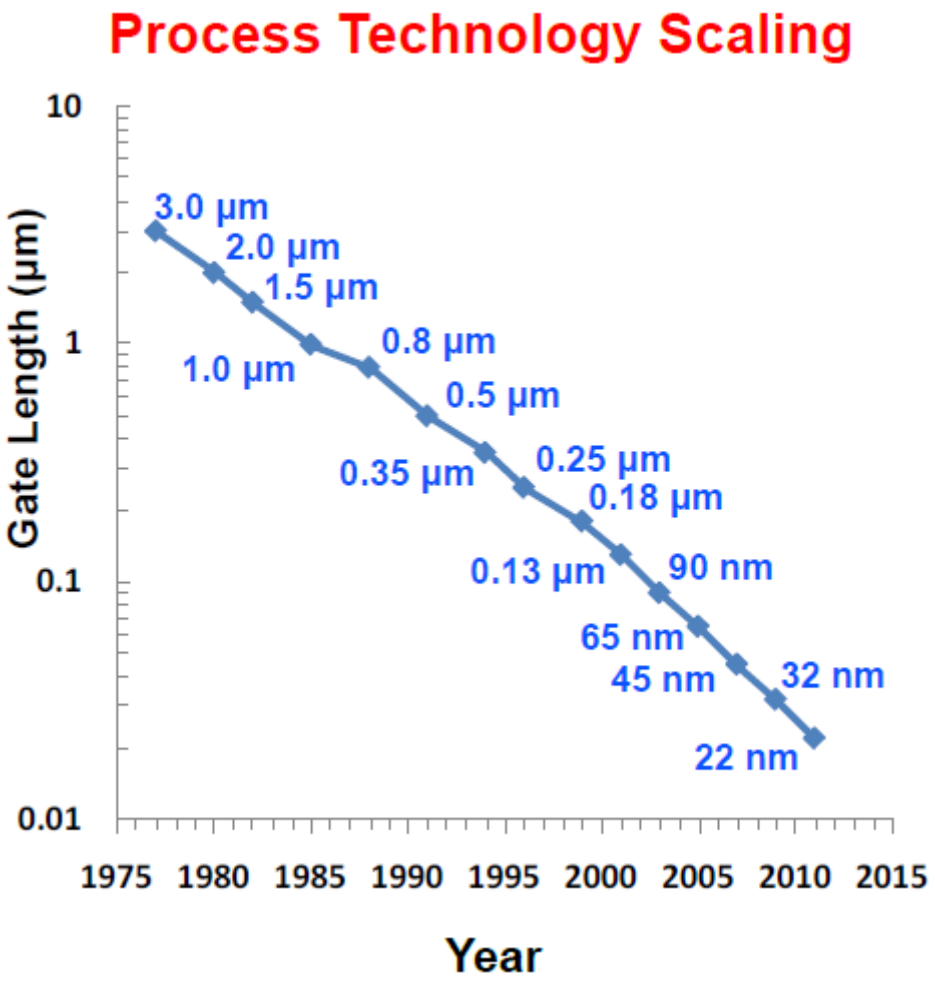
NUS National University of Singapore

Moore's Law: Process Integration



Courtesy, Intel

Process Scaling



Traditionally, every new process node provides:

1.5-2x more TRs
1.5-2x higher speed
1.5-2x less power
1.5-2x lower cost (Si area)

Problems along the way:

• Lithography
Solution: Shorter λ, Immersion, EUV

•Transistor performance does not scale
Solution: Stress techniques (65nm)

•Gate Leakage (t_{ox} does not scale)
Solution: High-K Metal Gate (45nm)

•Source-Drain leakage increases
Solution: FinFETs, FD-SOI (22nm)

Challenges in Digital Design

"Microscopic Problems"

- Ultra-high speed design
- Interconnect
- Noise, Crosstalk
- Reliability, Manufacturability
- Power Dissipation
- Clock distribution.

Everything Looks a Little Different

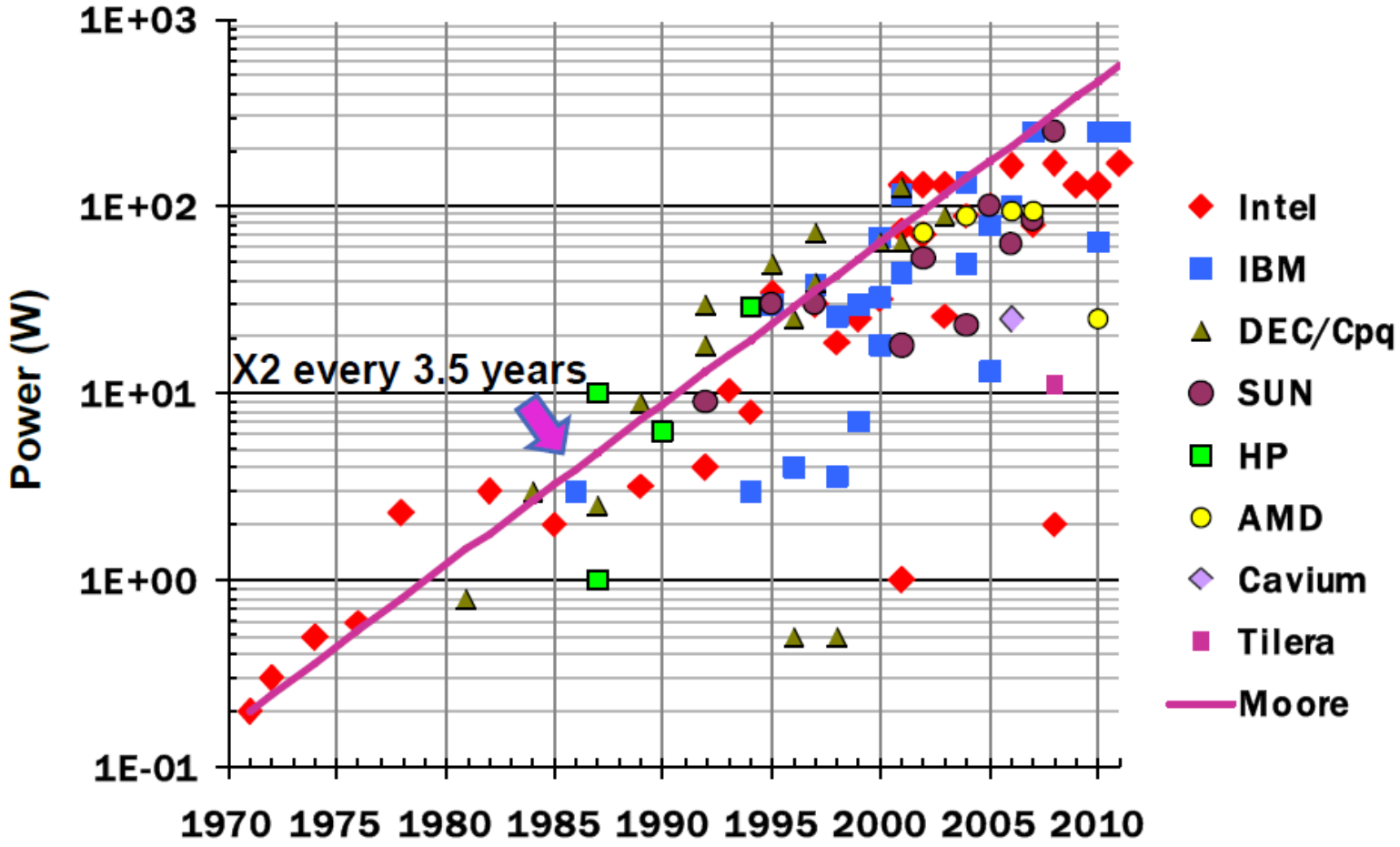
"Macroscopic Issues"

- Time-to-Market
- Billions of Gates
- High-Level Abstractions
- Reuse & IP: Portability
- Predictability
- etc.

…and There's a Lot of Them!

?

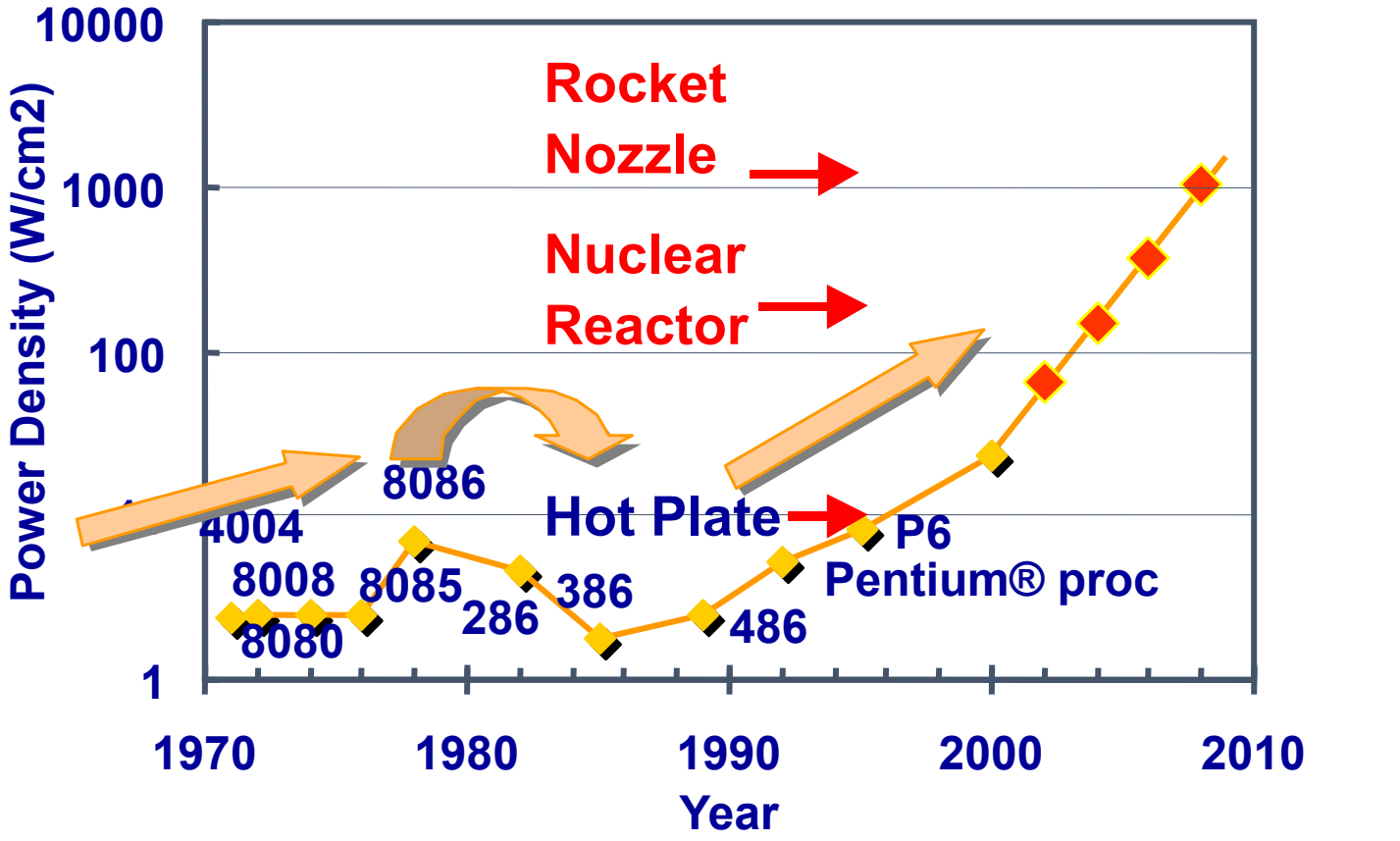
Challenge #1: Power Dissipation



We stopped following Moore's law (in P) since 2007

Courtesy, Intel

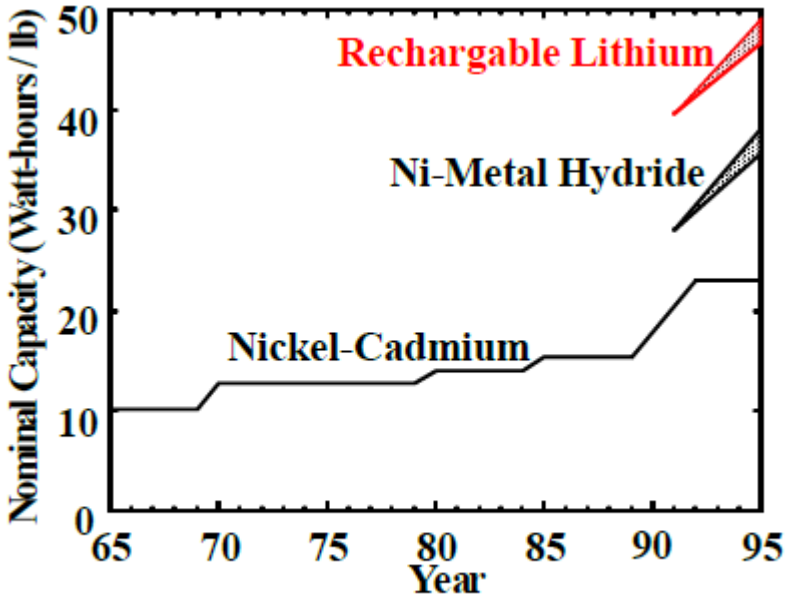
Challenge #1: Power Density



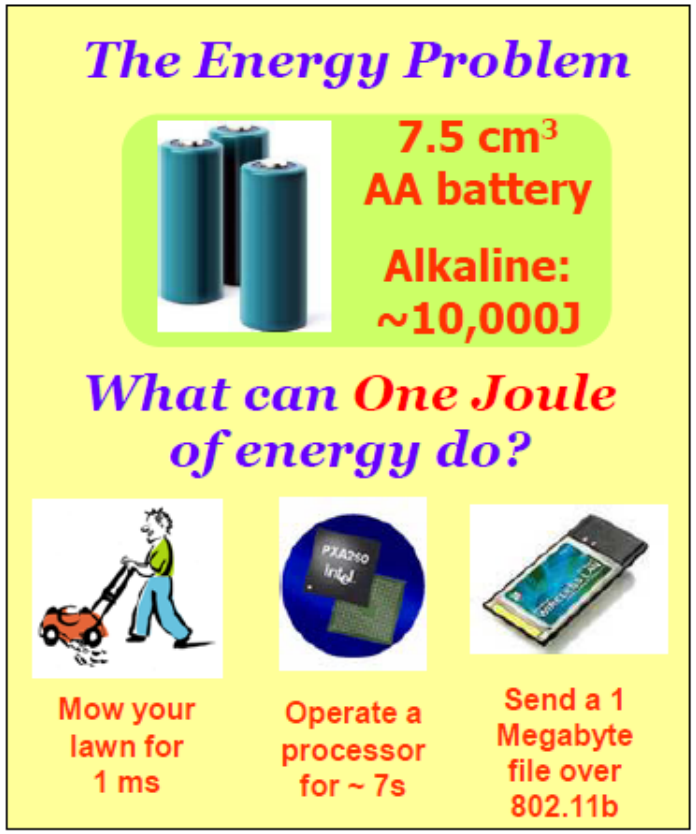
Power density too high to keep junctions at low temp

Courtesy, Intel (S. Borkar)

Challenge #1: Energy Efficiency



(from Jon Eager, Gates Inc. , S. Watanabe, Sony Inc.)

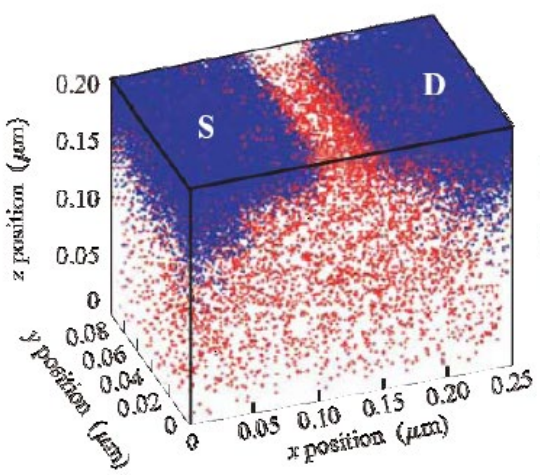


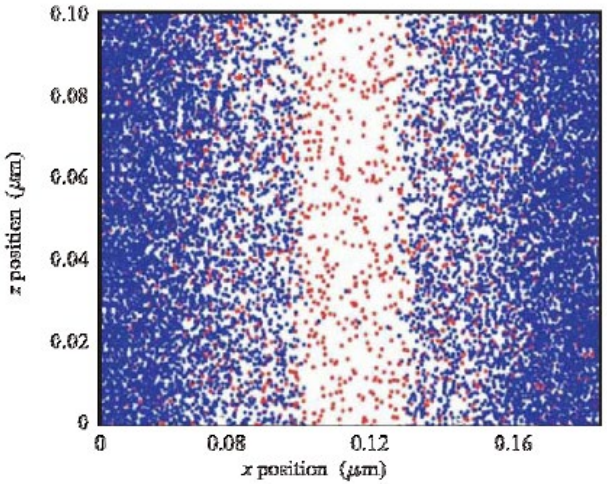
No Moore's law for batteries!

CEG 2027: Understand where power goes and ways to manage it
Differentiate between energy and power

NUS National University of Singapore

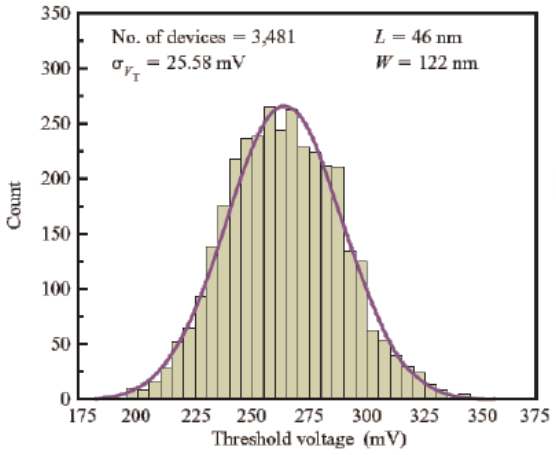
Challenge #2: Variation

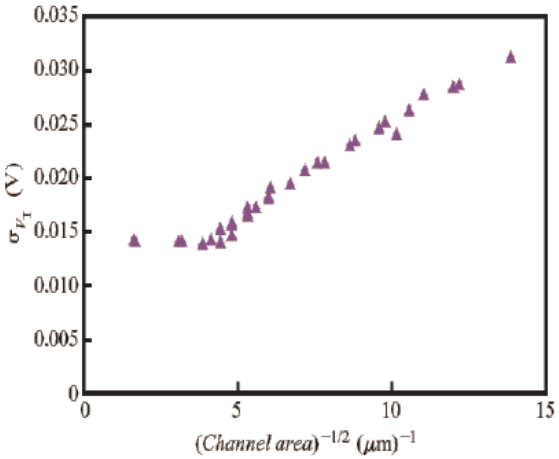




← Device dimension

Threshold voltage →





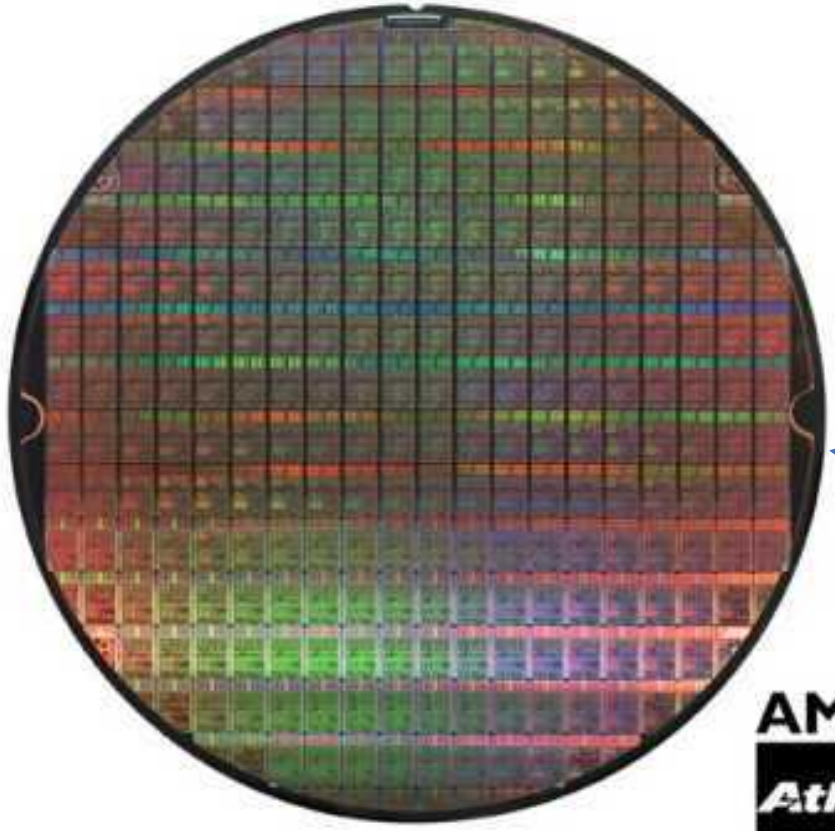
Source: K. Bernstein et al., High-Performance CMOS Variability in the 65nm Regime and Beyond, IBM Journal of Research and Development, Vol50, No 4/5, 2006

Challenge #3: Cost of Integrated Circuits

- NRE (non-recurrent engineering) costs
  - Design time and effort, mask generation
  - One-time cost factor
- Recurrent costs
  - Silicon processing, packaging, test
  - Proportional to volume
  - Proportional to chip area

Die Cost

Single die

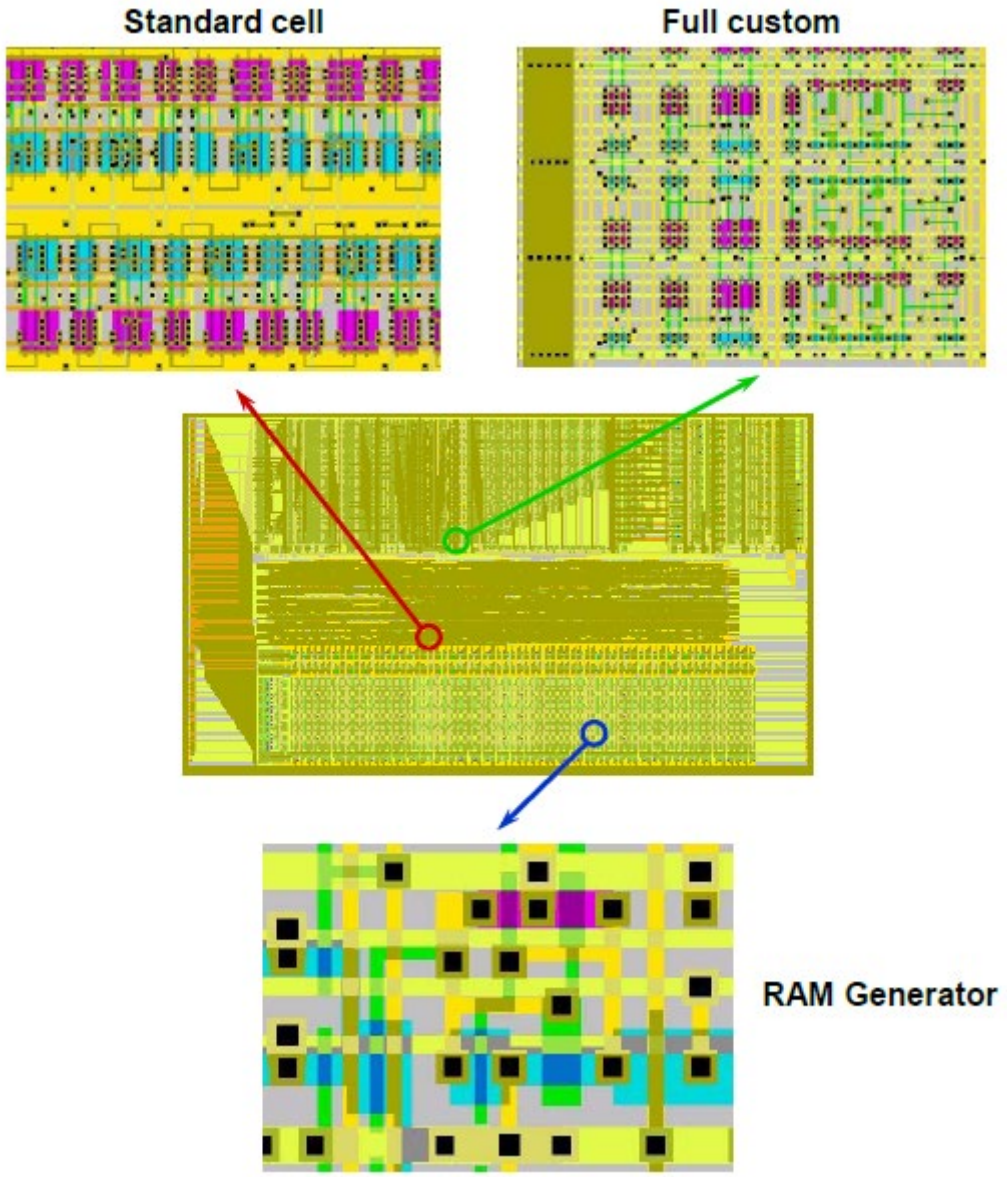


Wafer

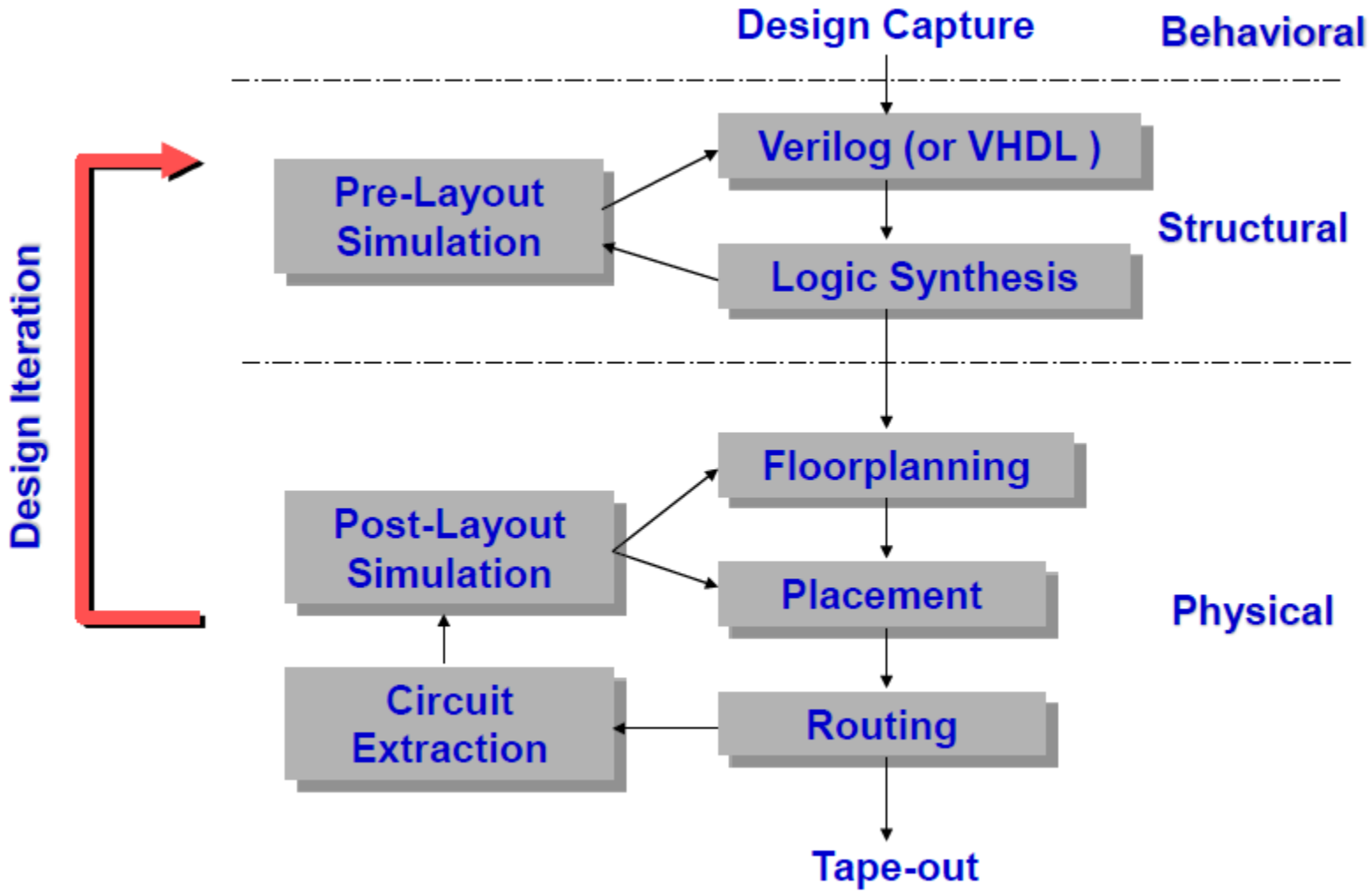
12” (30cm) wafer these days

From http://www.amd.com

Layout Methodologies

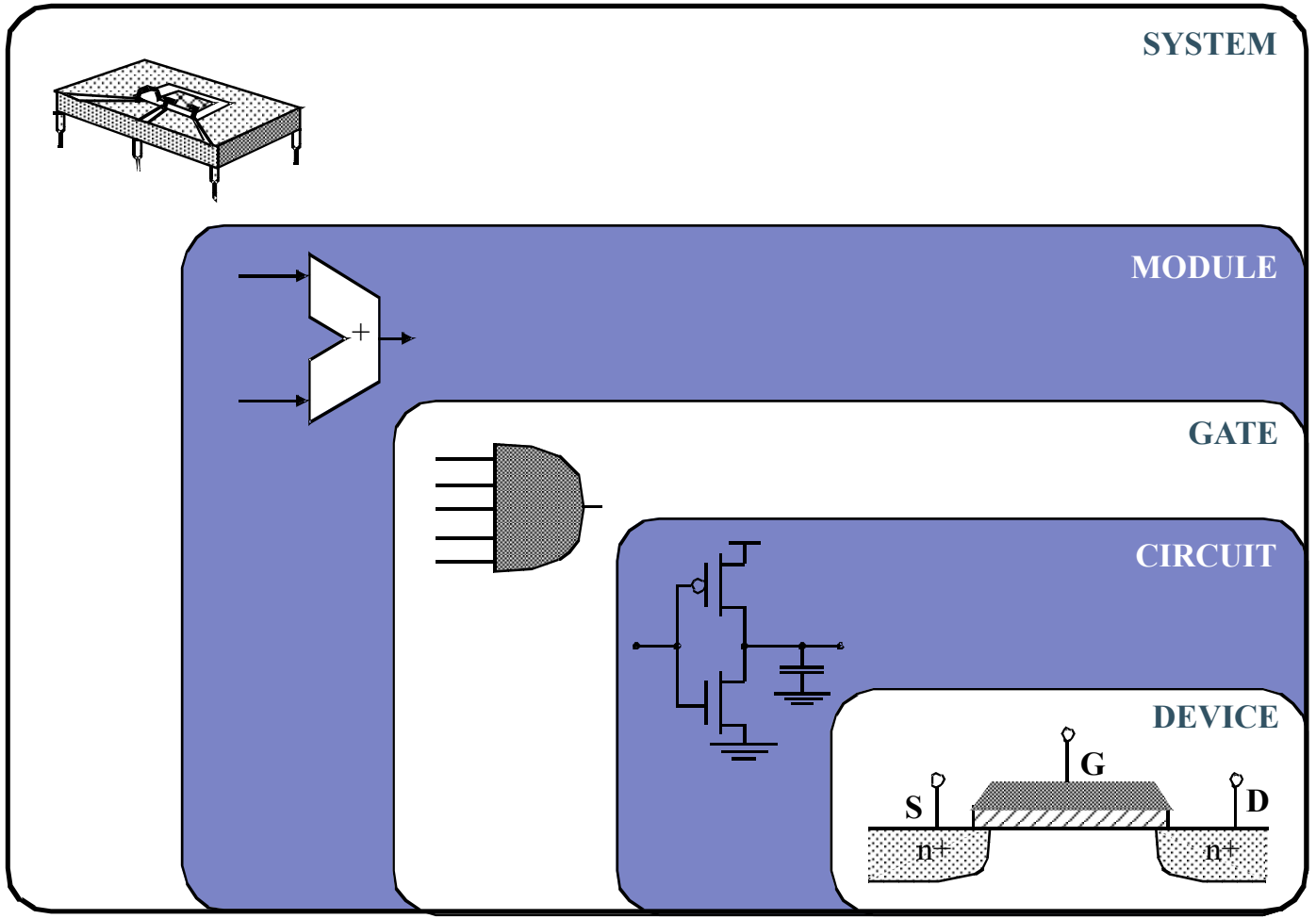


The ASIC Approach



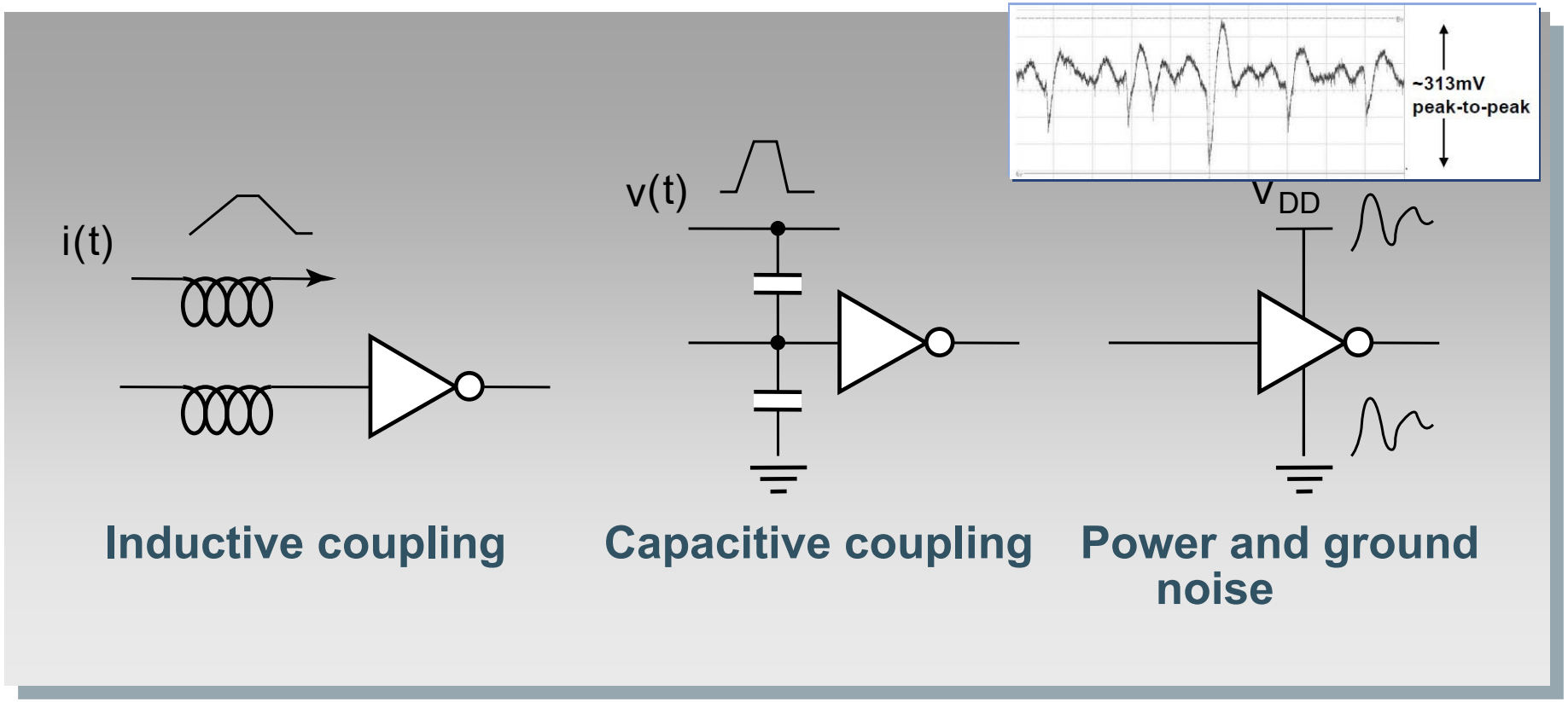
Most Common Design Approach for Designs < 1GHz Clock Rates

Design Abstraction Levels

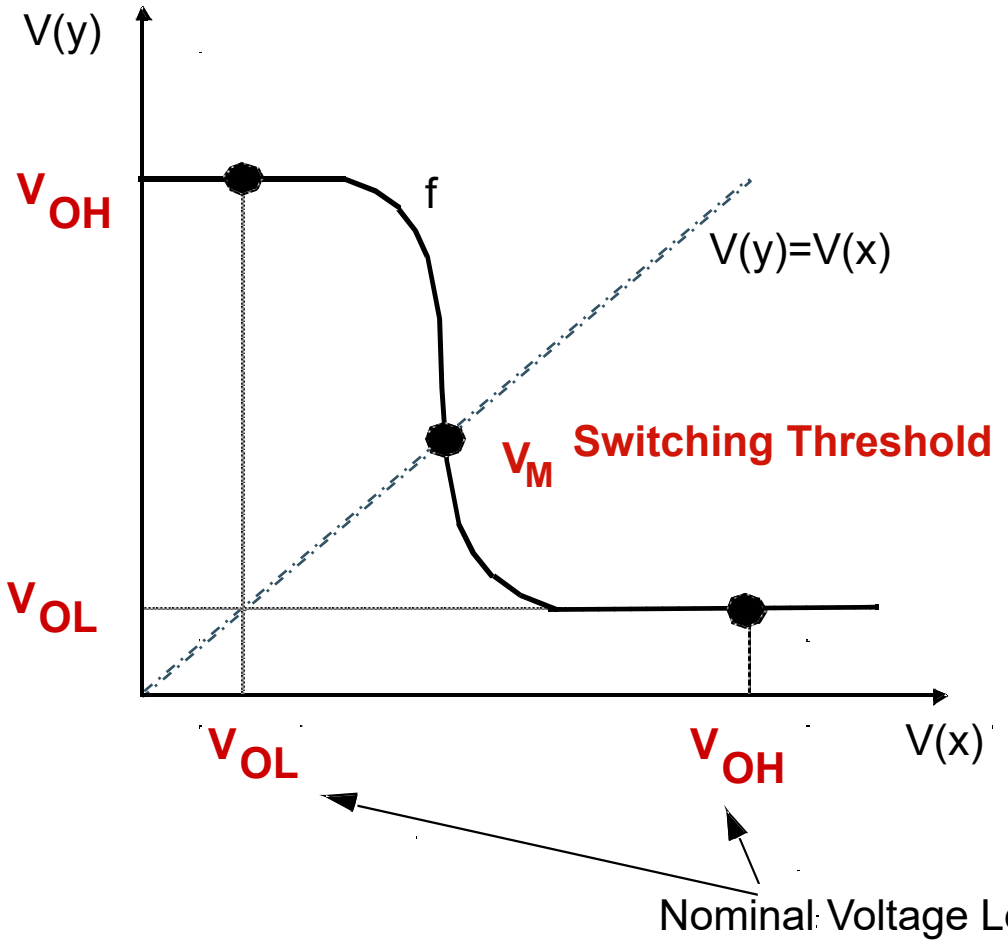


Reliability

Noise in Digital Integrated Circuits

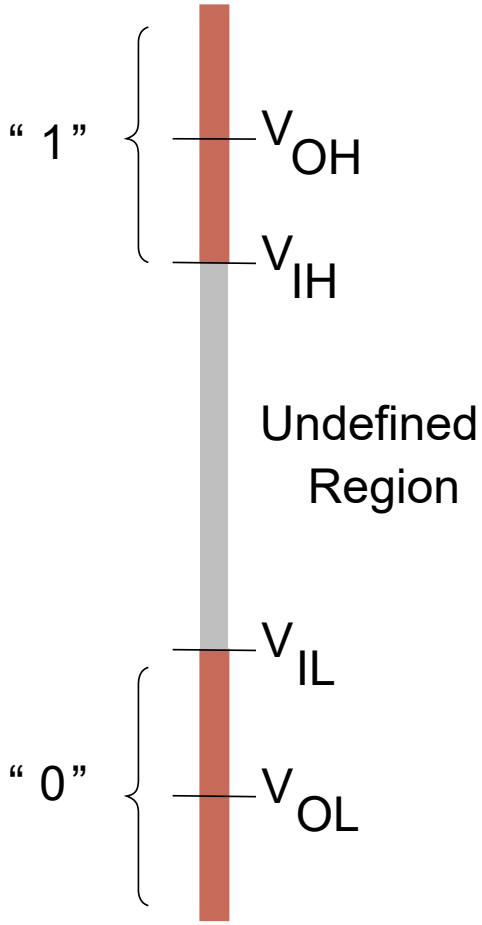


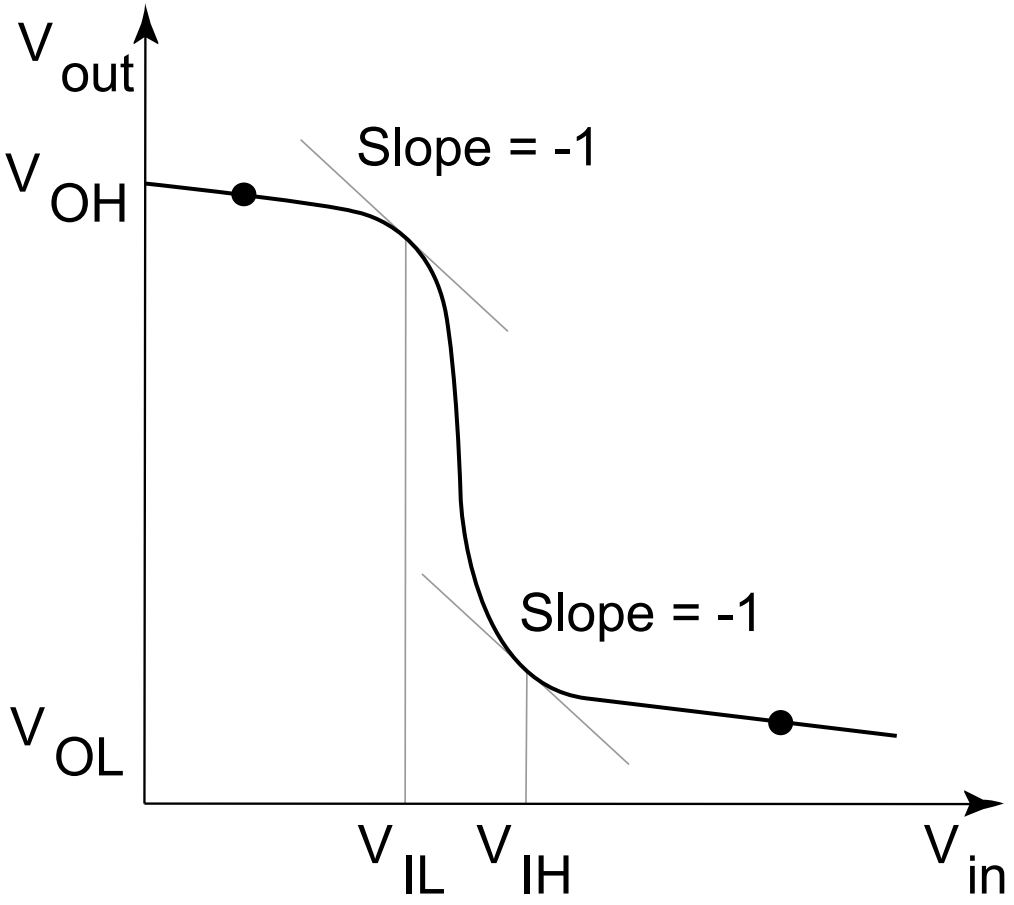
The Inverter: Voltage Transfer Characteristic



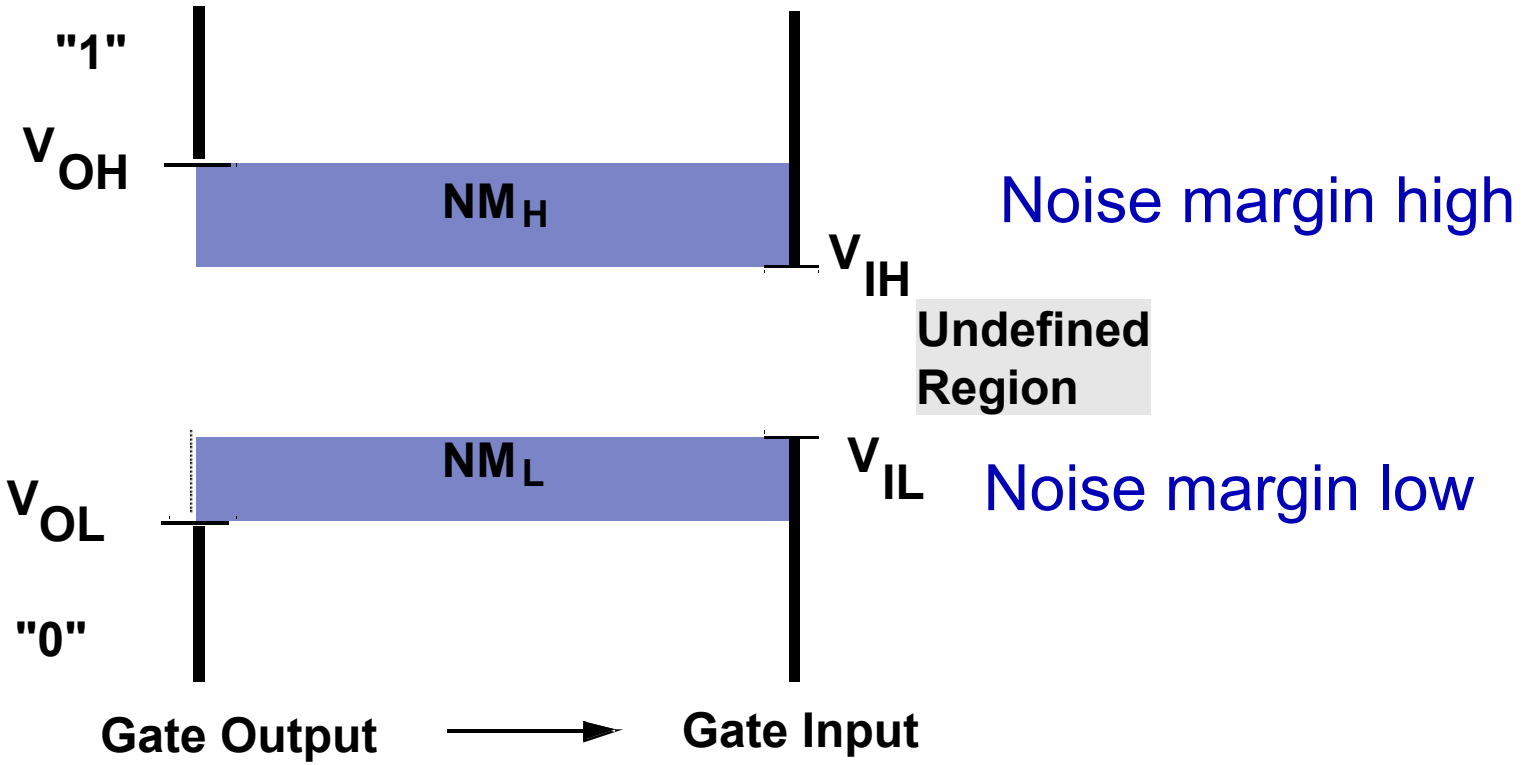
$$V_{OH} = f(V_{OL})$$
$$V_{OL} = f(V_{OH})$$
$$V_M = f(V_M)$$

Mapping between Analog and Digital Signals





Definition of Noise Margins



Noise margin high

Noise margin low

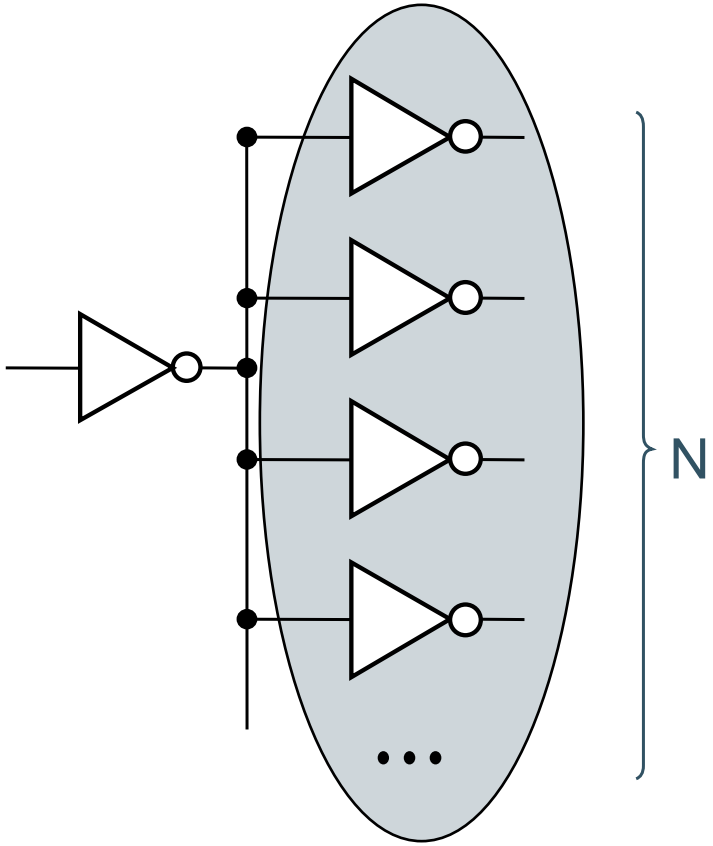
Noise Budget

- Allocates gross noise margin to expected sources of noise
- Sources:
  - Supply noise
  - Cross talk
  - Interference
  - Offset
- Differentiate between fixed and proportional noise sources

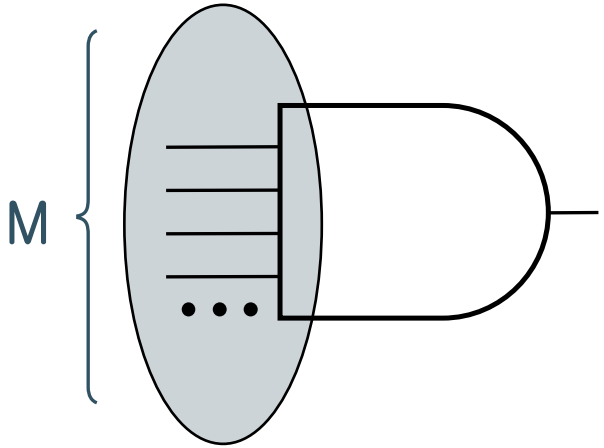
Key Reliability Properties

- Absolute noise margin values are deceptive
  - A floating node is more easily disturbed than a node driven by a low impedance (in terms of voltage)
- Noise immunity is the more important metric
  - Describes the capability to suppress noise sources
- Key metrics:
  - Noise transfer functions
  - Output impedance of the driver
  - Input impedance of the receiver

Fan-in and Fan-out

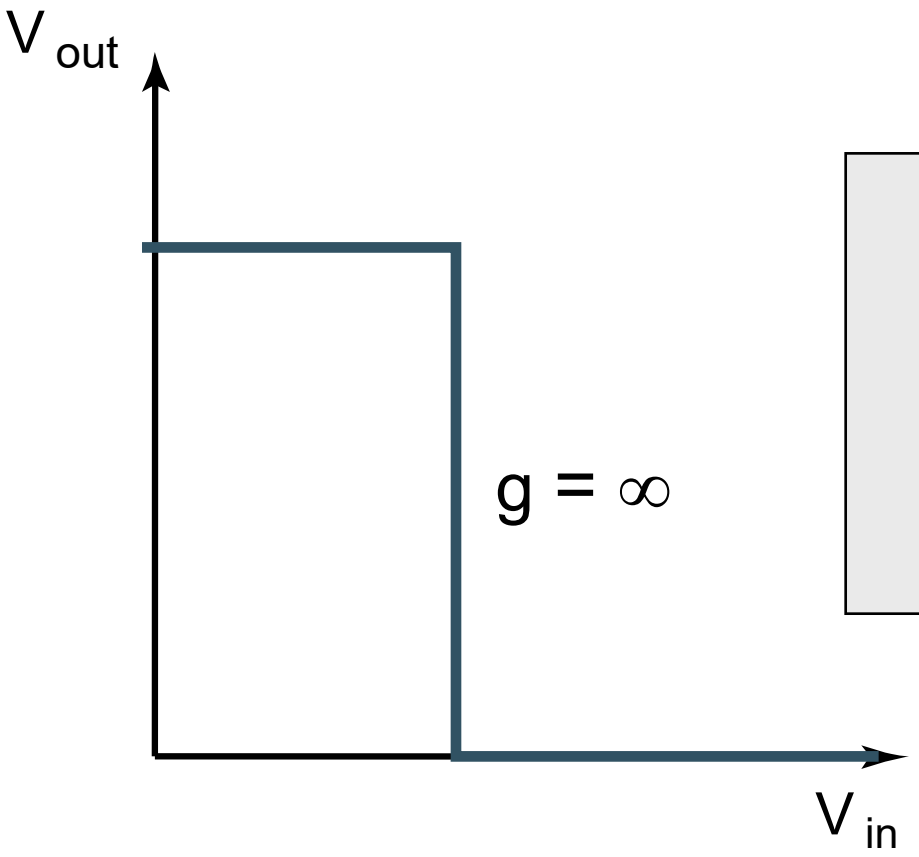


Fan-out N



Fan-in M

The Ideal Gate



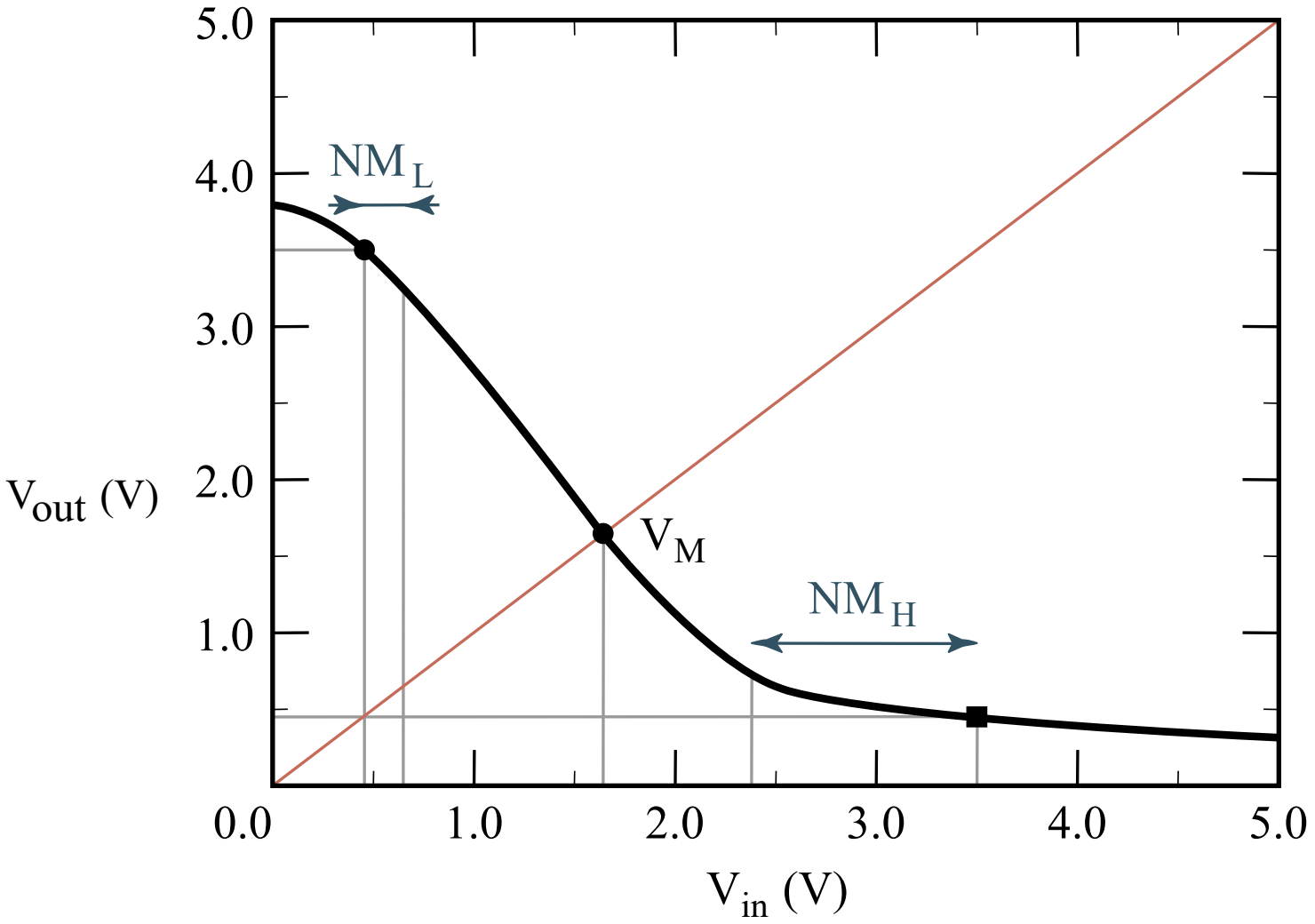
$R_i = \infty$

$R_o = 0$

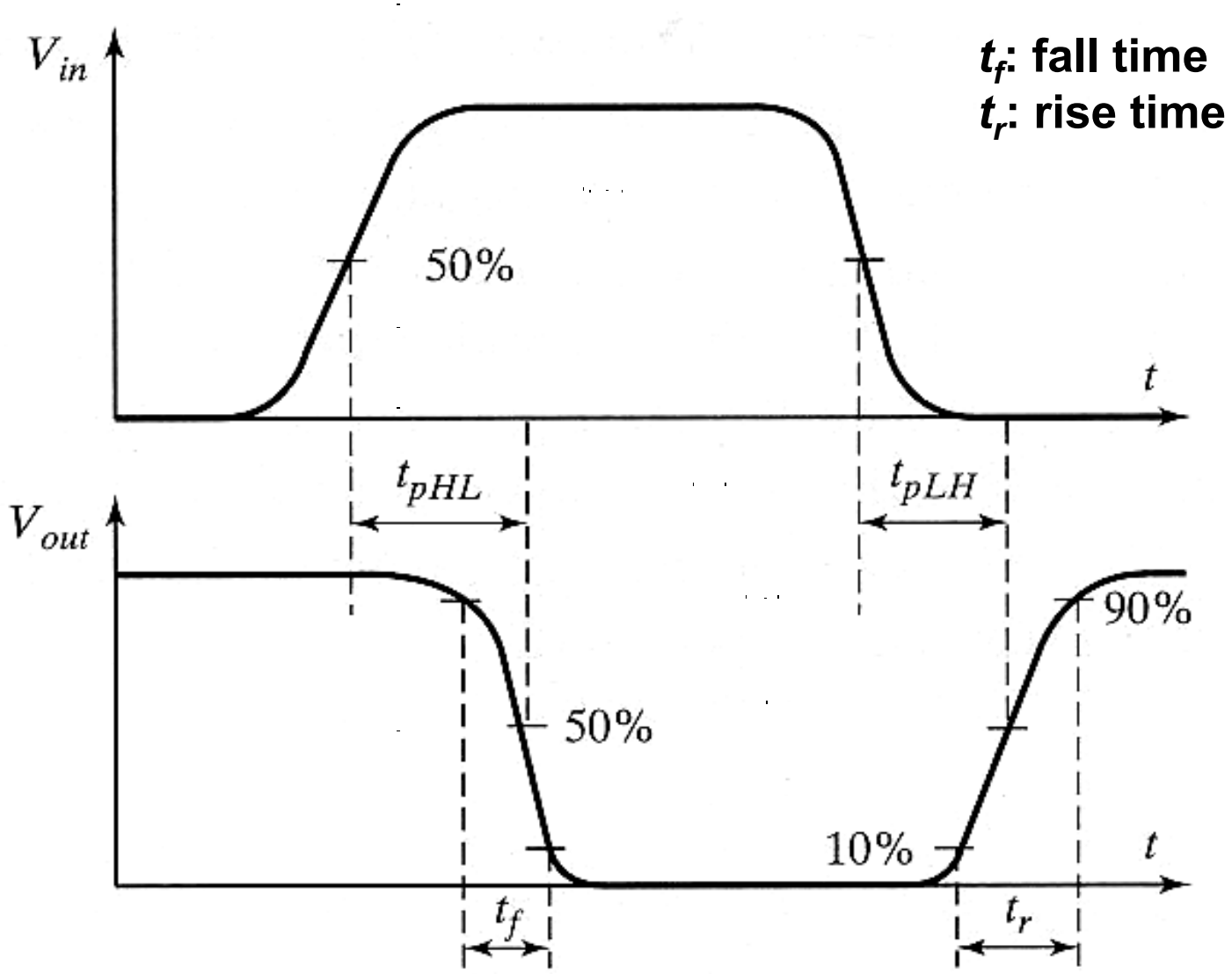
Fanout = ∞

$NM_H = NM_L = V_{DD}/2$

An Old-time Inverter



Delay Definitions



t_f: fall time
t_r: rise time

Summary

- Digital integrated circuits have come a long way and still have quite some potential left for the coming decades
- Some interesting challenges ahead
  - Obtaining basics of digital circuits in ***transistor level*** to system perspective
  - Getting a clear perspective on the challenges and potential solutions is the purpose of this module
- Understanding the design metrics and trade-offs that govern digital design is crucial
  - Speed, sizing, power and energy dissipation

CG2027 Transistor-Level Digital Circuits

Handout #1: The Device

National University of Singapore

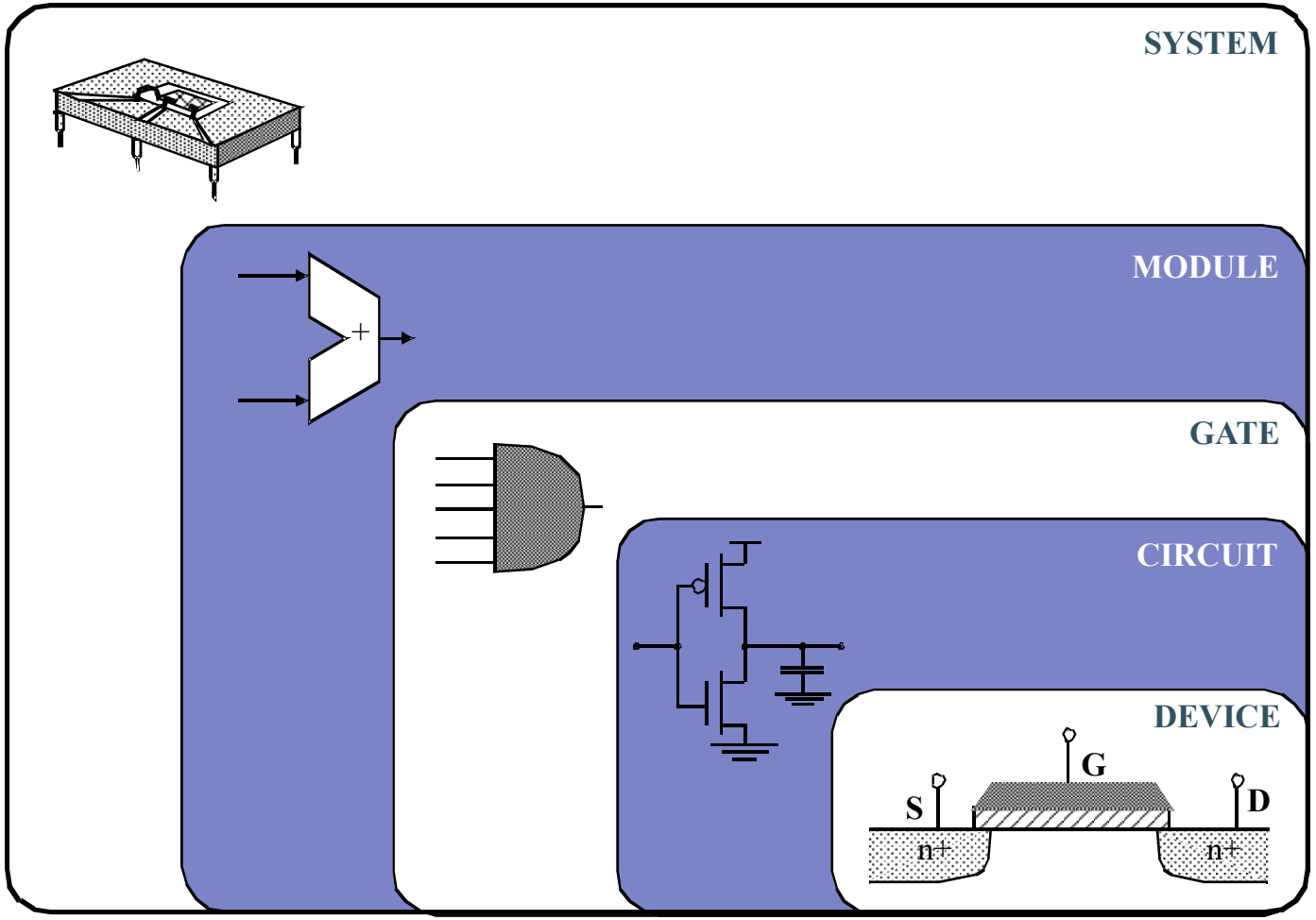
Kelvin Fong

Lecture Overview

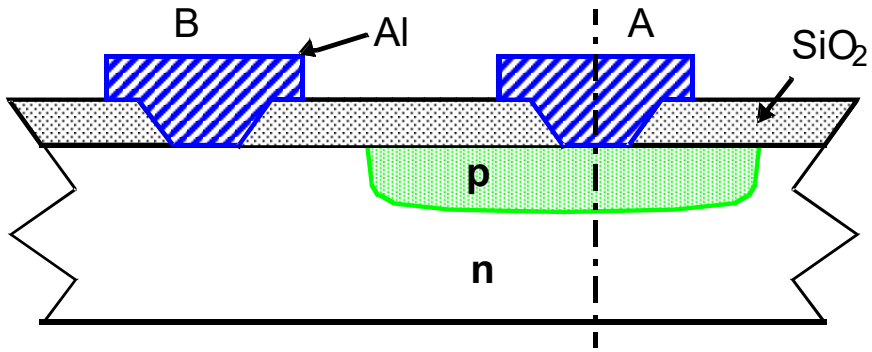
In this lecture, you will learn about

- Basic devices (diode, BJT and MOS transistor).
- Brief overview of fabrication process (lithography, etching and deposition).
- Brief usage of doping/thermal steps like oxidation, diffusion and ion implantation.
- Evolution of mask layout and design rules.
- Circuit implementation which will continue in further chapters.

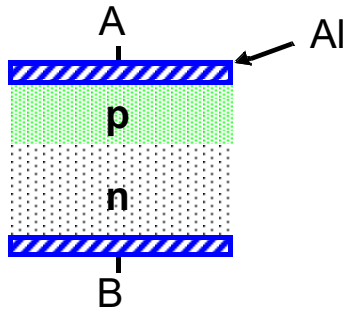
Design Abstraction Levels



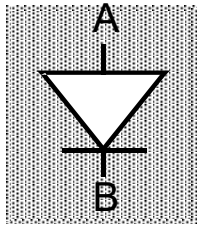
The Diode



Cross-section of pn-junction in an IC process



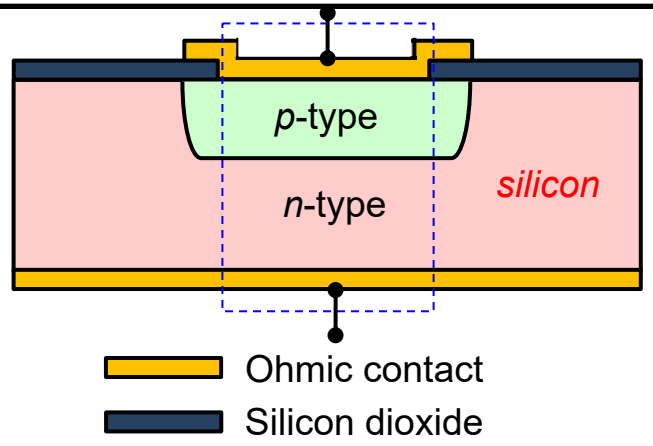
One-dimensional representation



diode symbol

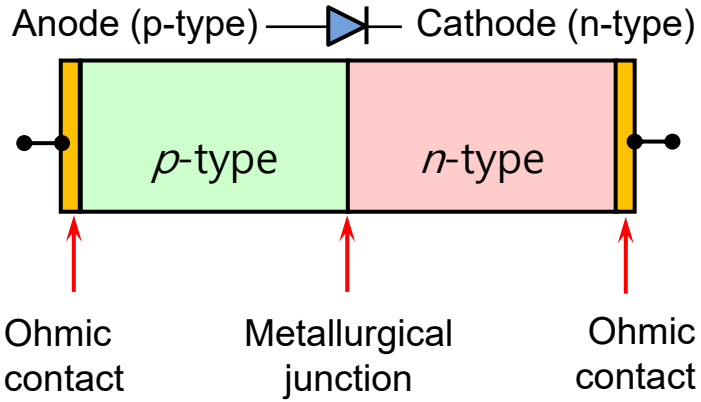
Mostly occurring as **parasitic** element in Digital ICs

Diode - Introduction



Realistic cross-section of fabricated silicon *pn* junction diode

Circuit symbol of diode:



Simplified structure and circuit symbol of a semiconductor *pn* junction.

The diode to be discussed is a semiconductor *pn*-junction.

- Made using a single crystal semiconductor (typically silicon), with impurities added to one side to contain negative charge carriers (electrons), called an *n*-type semiconductor; and to the other side to contain positive charge carriers (holes), called a *p*-type semiconductor.
- In a semiconductor, there are two types of charge carriers: electrons (with charge of -1.602×10^{-19} C) and holes (with charge of $+1.602\times10^{-19}$ C).

Diode – Semiconductor

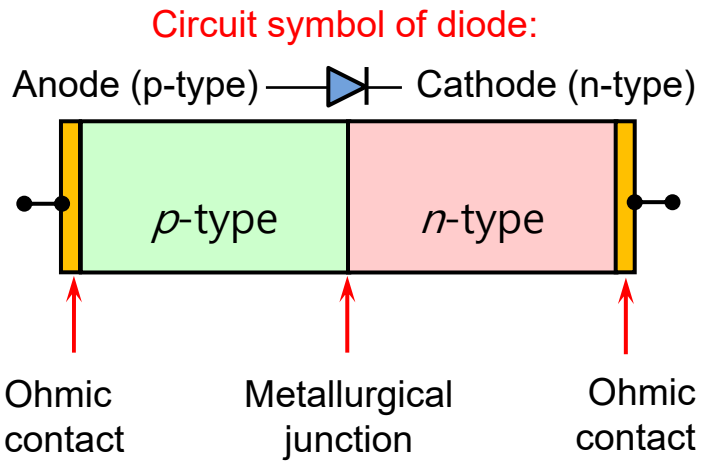
- Semiconductors have electrical conductivities between that of metals (conductors) and insulators.
- Unlike metal and insulator, a unique property of semiconductor is that impurities can be added (in a controlled manner) into it –
  - to make it *n*-type or *p*-type, and
  - to change its conductivity (or resistivity).

	Material	Typical Resistivity* (Ω-cm)	Typical Carrier Concentration (cm^{-3})
Metal	Copper Gold Aluminum Stainless Steel	1.69×10^{-6} 2.20×10^{-6} 2.67×10^{-6} $70\text{-}78\times10^{-6}$	$\sim10^{23}$
Semiconductor	Germanium Silicon Gallium Arsenide	46 2.3×10^{5} 10^{8}	Wide range up $\sim10^{18\text{-}19}$ (with doping)
Insulator	Silicon Nitride Silicon Dioxide Polyimide	10^{14} $10^{14}\text{-}10^{16}$ 10^{18}	Negligible

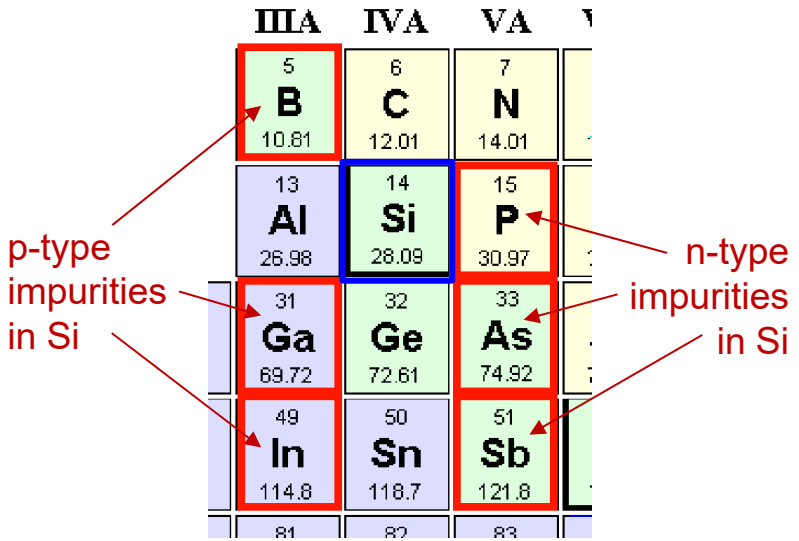
* Resistivity is reciprocal of conductivity. Temperature ~ 300 K (room temperature).

NUS National University of Singapore

Diode – Semidonductor



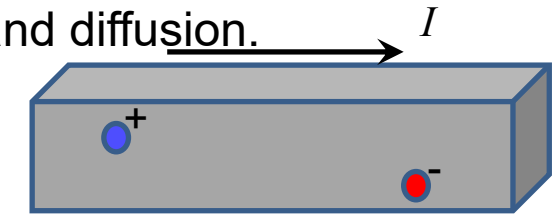
Simplified structure and circuit symbol of a semiconductor pn junction.



- The process of adding impurities to semiconductor is known as doping.
- Impurities added to semiconductor to make it *n*-type and *p*-type are different. For silicon (a group IV element) –
  - *p*-type impurity is a group III element (Boron, Aluminium and Gallium).
  - *n*-type impurity is a group V element (Phosphorus, Arsenic and Antimony)
- The process of doping can also change a *p*-type semiconductor to *n*-type semiconductor, and vice versa. For example, by adding more *n*-type impurities to an originally *p*-type semiconductor, it can be changed to *n*-type. This allows the making of *pn*-junction, and transistors (BJT and MOSFET).

Diode – Origin of Current

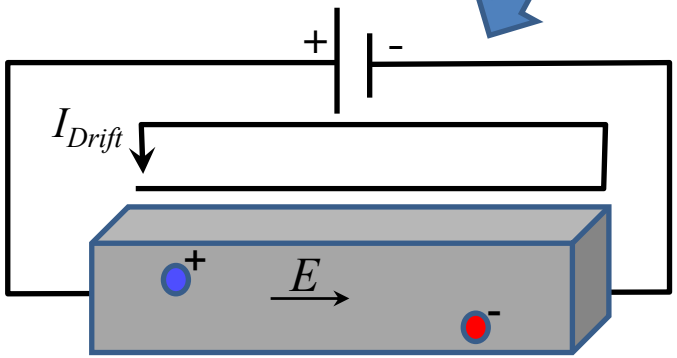
- Take note there are two types of charged carrier movement, leading to two types of current: drift and diffusion.



Movement of charged carriers (electrons or holes) gives rise to current

Two types of movements

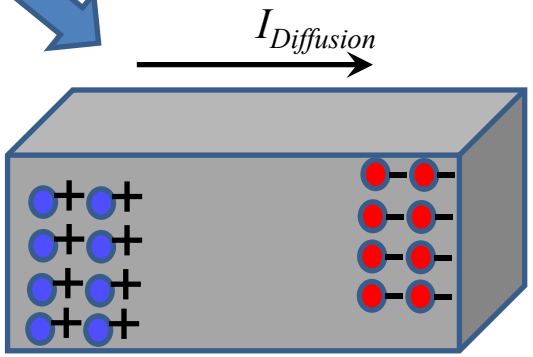
Drift Current, I_{Drift}



Charged carriers move/drift in the presence electric field. Carriers drift at a velocity proportional to the electric field.

The current in all electronic devices originates from either of the two mechanisms

Diffusion Current, $I_{Diffusion}$



Charged carriers diffuse owing to the difference in carrier concentration. The resulting current is proportional to the concentration gradient.

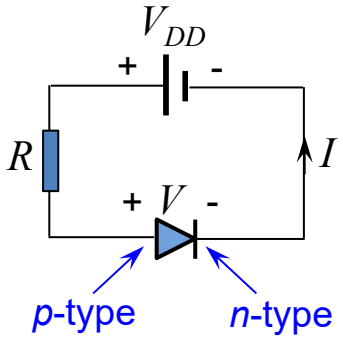
Diode – Carrier Movement in Devices

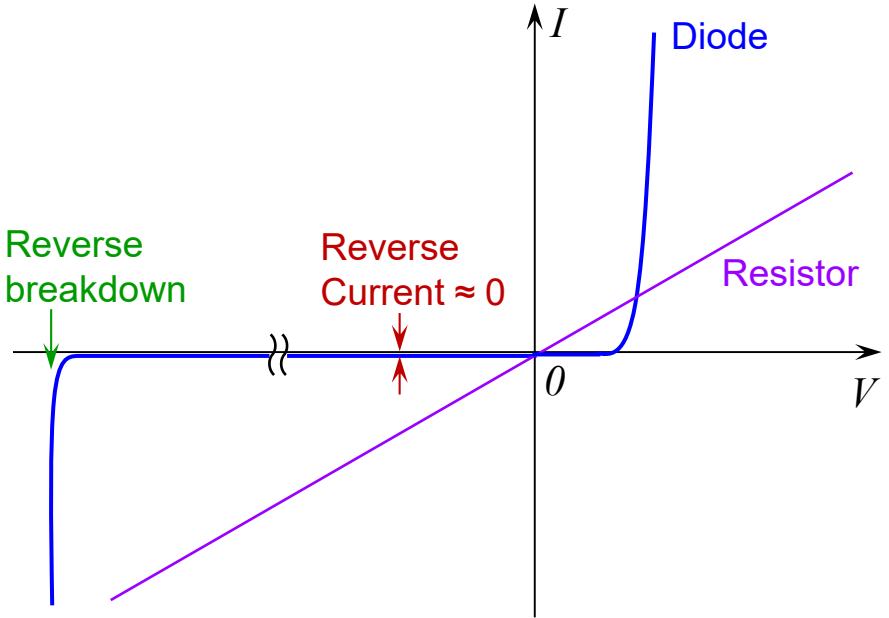
Devices	Movement Mechanism in on-state	Type of Carriers
Resistor	Drift	▪ Electrons (Metal) ▪ Electrons and holes (Semiconductor)
Diode	Diffusion	▪ Electrons and holes
Bipolar Junction Transistor	Diffusion	▪ Electrons and holes
MOSFET	Drift	▪ Electrons (NMOS) ▪ Holes (PMOS)

Diode – Operation

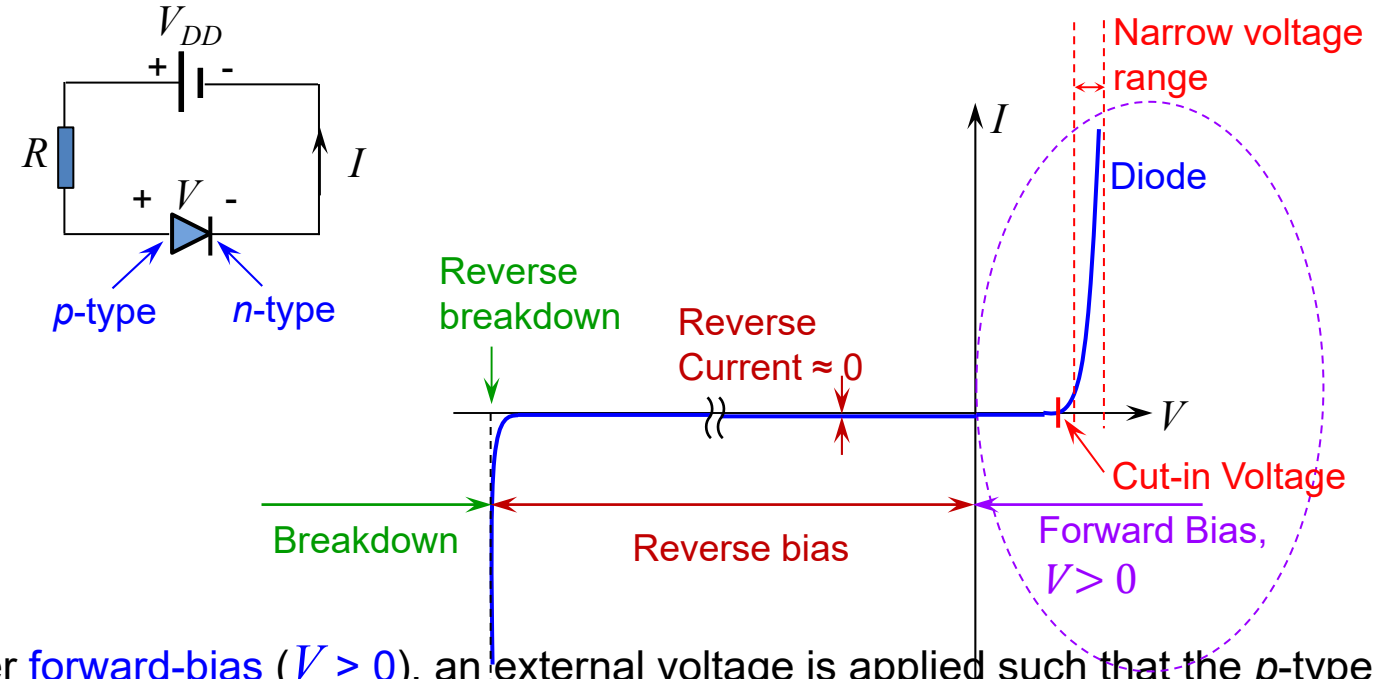
Diode (semiconductor *pn*-junction) is the simplest (2-terminal) and most fundamental *nonlinear* circuit element.

- It allows a current flow through it easily in one direction (known as the forward direction, $V > 0$), but not in the opposite direction (known as the reverse direction, $V < 0$), except for the reverse breakdown region. This is unlike a resistor, which is a linear element that has a linear current-voltage relation.
- Diode can be used as a switch and in a rectifier circuit to convert ac into dc.





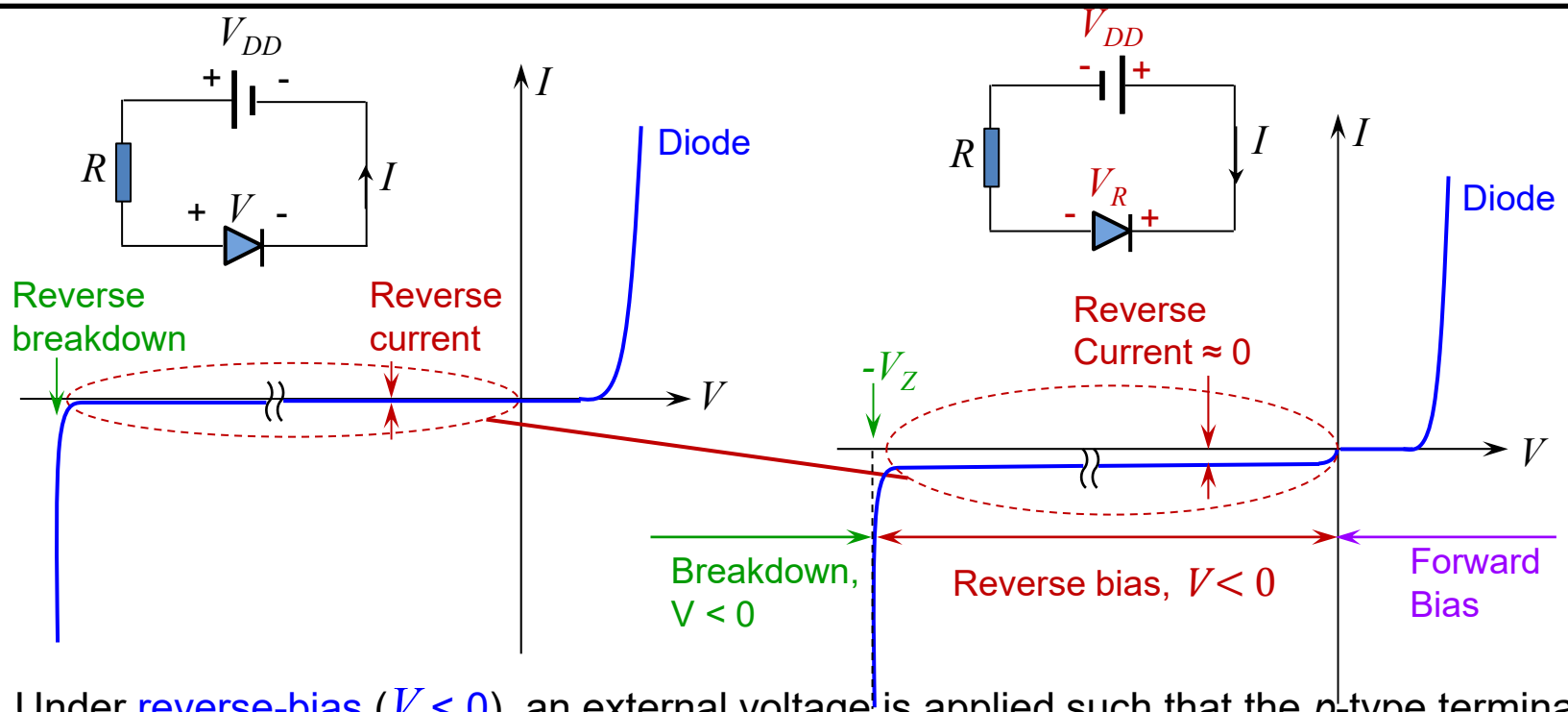
Diode – Forward Bias



- Under forward-bias ($V > 0$), an external voltage is applied such that the *p*-type terminal is at a higher (positive) voltage with respect to the *n*-type terminal.
- The forward current flows through the diode from the *p*-type side to the *n*-type side
- The forward current remains small (≈ 0 practically) until the cut-in voltage, and increases quickly with a small increase in V thereafter.
- With a substantial forward current, the voltage drop across the diode lies in a narrow range.

Typically avoided in Digital IC
Diffusion current becomes dominant

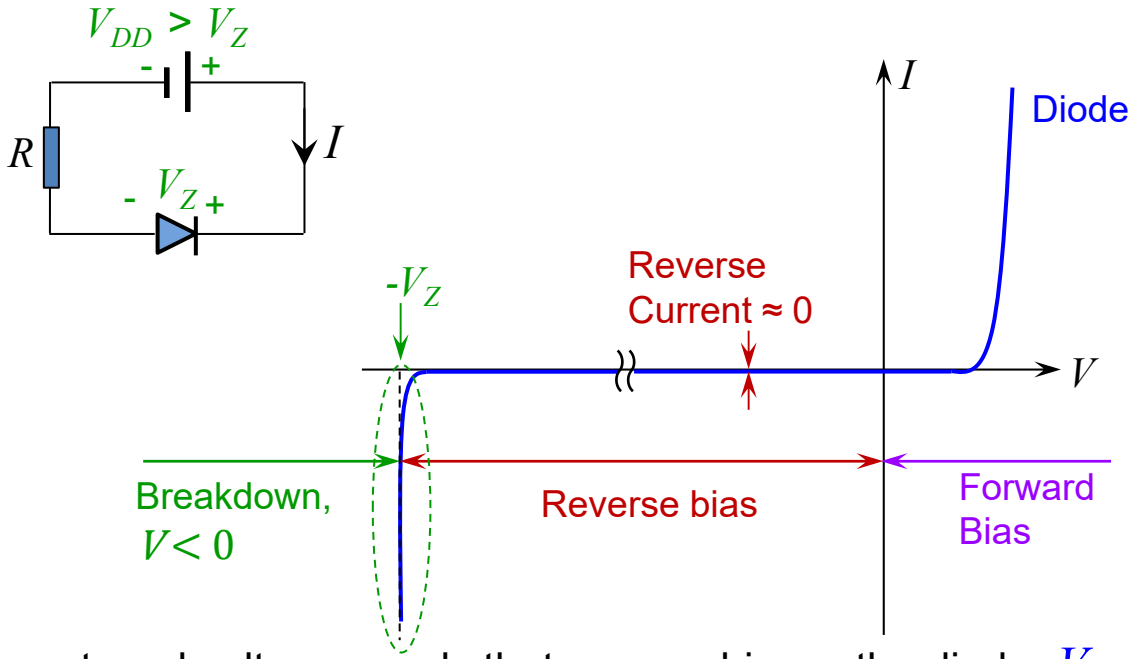
Diode – Reverse Bias



- Under reverse-bias ($V < 0$), an external voltage is applied such that the *p*-type terminal is at a lower (or negative) voltage with respect to the *n*-type terminal.
- The reverse current flows through the diode from the *n*-type side to the *p*-type side.
- For reverse bias voltage magnitude, $|V| = V_R < V_Z$, the breakdown voltage, the reverse current is very small and can be treated practically as zero, meaning the diode is equivalent to an open circuit.

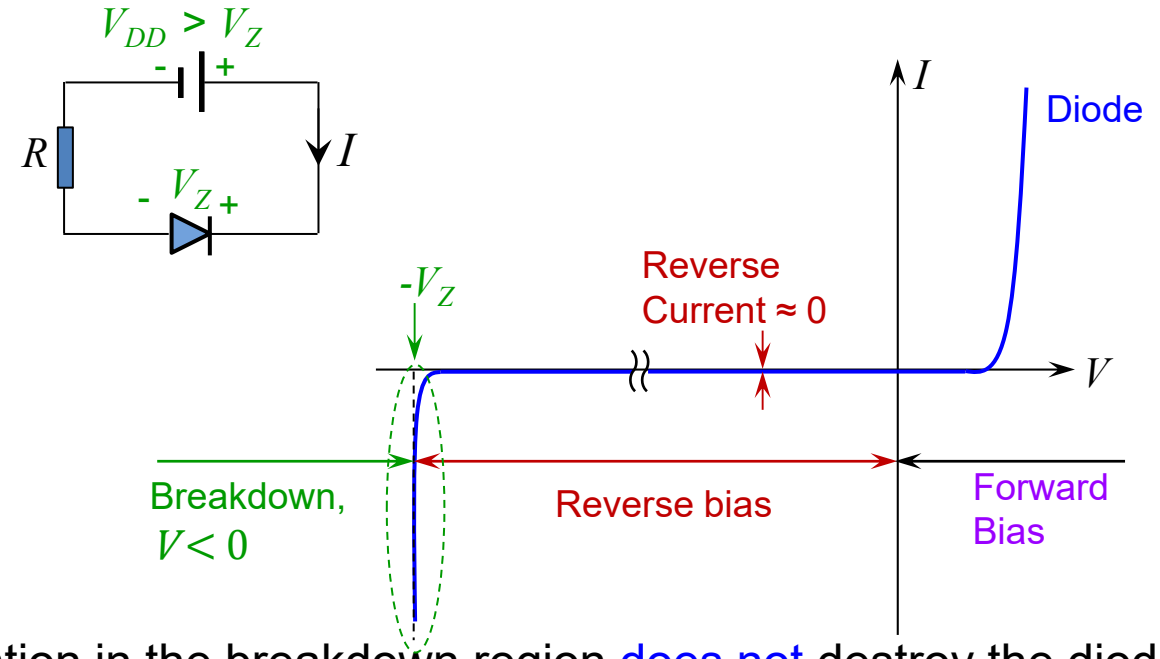
The dominant operation mode in digital IC
Drift current becomes dominant

Diode – Breakdown Region



- With an external voltage supply that reverse biases the diode, $V_{DD} > V_Z$ (breakdown voltage), the reverse current is no longer ≈ 0 but increases rapidly with practically no increase in the voltage across the diode. This condition is known as breakdown.
- Under breakdown condition, the voltage across the pn junction diode stays practically constant at $-V_Z$. Minus sign highlights that breakdown is a reverse biased condition.

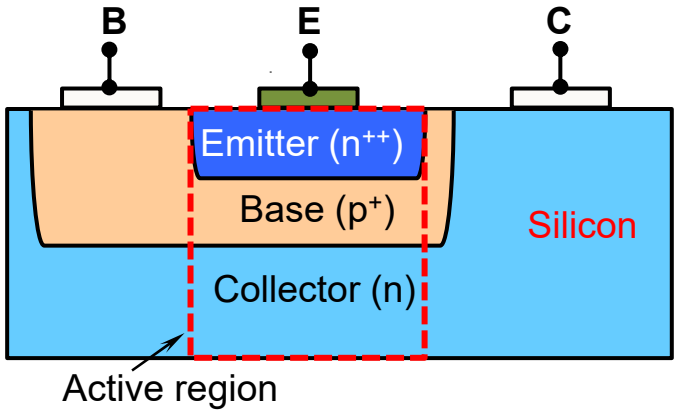
Diode – Breakdown Region

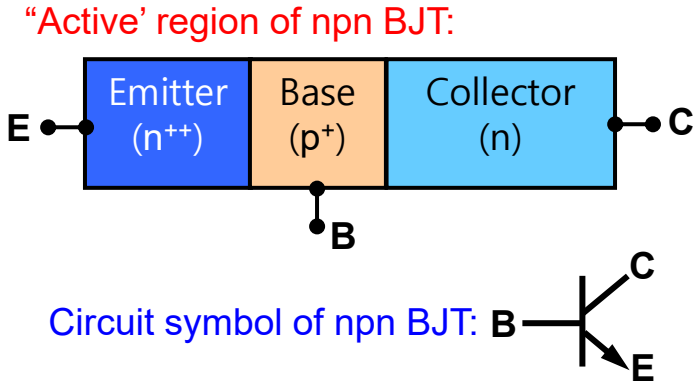


- Operation in the breakdown region does not destroy the diode, provided the current through it is kept below a certain level, such that the power dissipation ($V \times I$) is below what the diode can handle.
- Current, while operating in the breakdown region, can be limited by connecting a resistor, R, of suitable value in series with the *pn* junction diode -

$$I = \frac{V_{DD} - V_Z}{R}$$

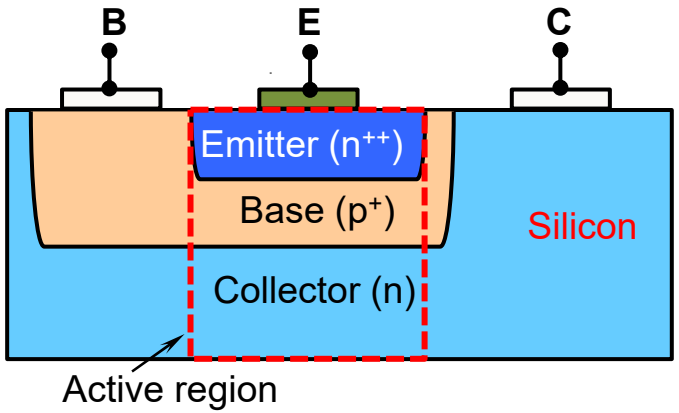
BJT - Introduction

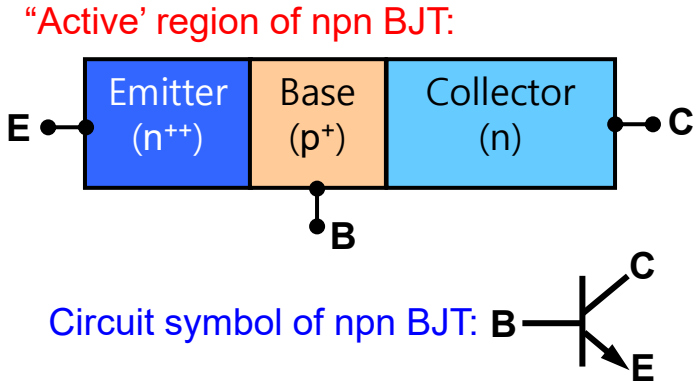




- Bipolar junction transistor (BJT) is a 3-terminal device made using a single crystal semiconductor (typically silicon), just like the *pn*-junction diode.
- BJT is made with 3 doped semiconductor regions, namely emitter, base and collector, corresponding to the 3 terminals.
- Left figure above shows a schematic cross-sectional view of an npn BJT, where the emitter is n-type, base is p-type and collector is n-type.
- The "active" region of the BJT is the region under (and including) the emitter. This is the part of the device that provides, for example, the amplifying function of the BJT. The rest of the structure is to facilitate the movement of the currents into or out of the transistor. We will focus on the "active' region of BJT.

BJT - Introduction

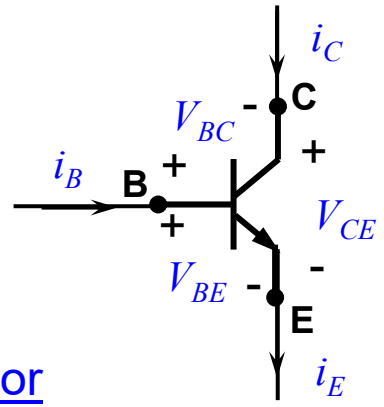




- BJT is not a symmetrical device, in particular, impurities added to the emitter is at a much higher concentration than that added to collector. Hence, emitter is indicated as n^{++}-type (very heavily doped and has many more electrons) and collector as n-type.
- Concentration of impurities (of a different type from that for emitter/collector) added to base is in between those of emitter and collector. Base is thus indicated as p^{+}-type.
- BJT comprises two back-to-back *pn*-junctions: emitter-base junction and base-collector junction. The two *pn*-junctions must be close enough to each other to interact, and this requires the base to be thin.

BJT – Modes of Operation

- BJT has 2 *pn*-junctions (emitter-base junction and collector-base junction), and each *pn*-junction can be either forward biased or reverse biased, hence there are 4 possible permutations of the biasing as shown in the table below for an npn BJT. We will not focus on BJT as it is not used in CMOS logic.

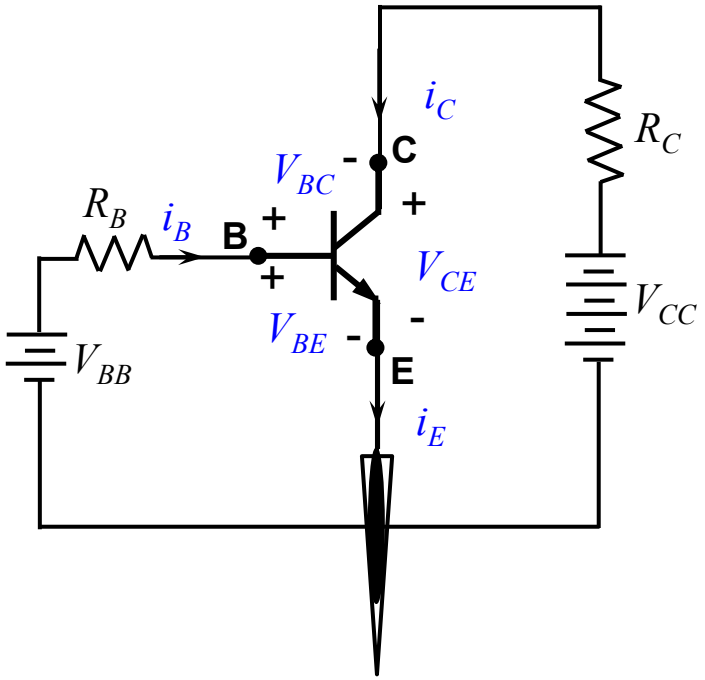


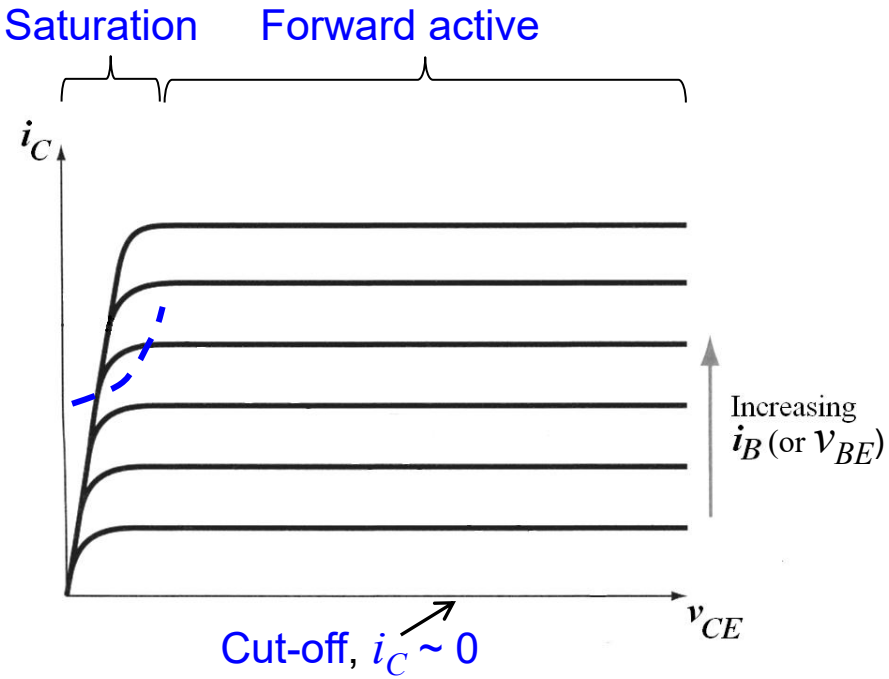
Modes of operation of the **npn** bipolar junction transistor

Mode of Operation	Emitter-Base Junction	Collector-Base Junction	Applications
Cut-off	Reverse biased (V_{BE} < 0 for npn)	Reverse biased (V_{BC} < 0 for npn)	Logic - OFF State
Forward Active	Forward biased (V_{BE} > 0 for npn)	Reverse biased (V_{BC} < 0 for npn)	Amplifier
Saturation	Forward biased (V_{BE} > 0 for npn)	Forward biased (V_{BC} >0 for npn)	Logic - ON State
Reverse Active	Reverse Biased (V_{BE} < 0 for npn)	Forward Biased (V_{BC} > 0 for npn)	Not used

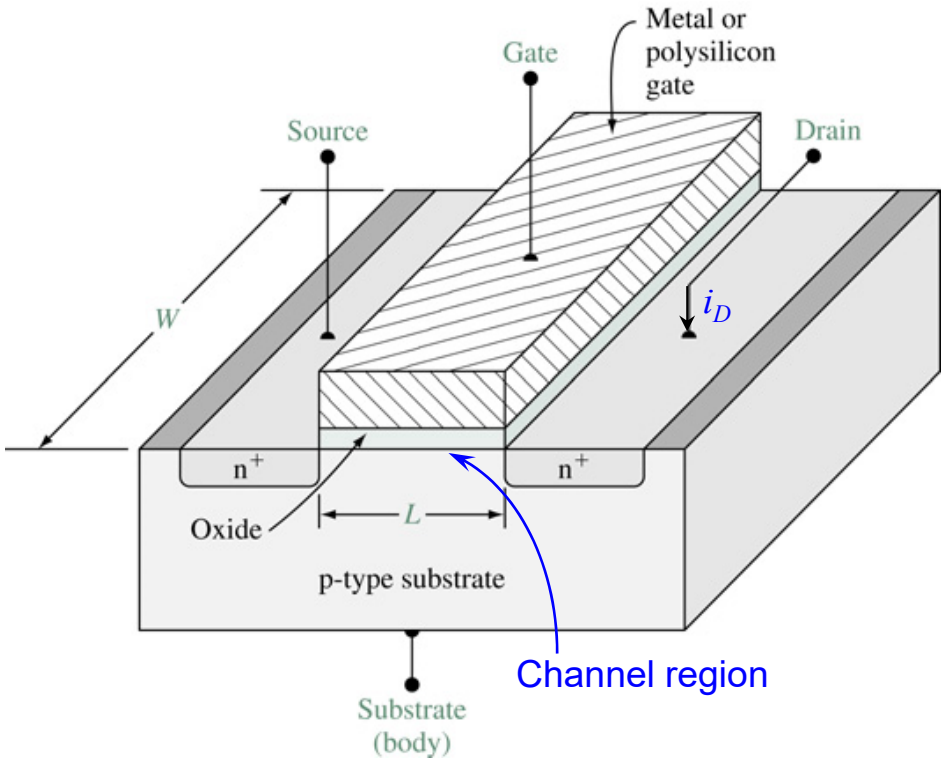
BJT – Modes of Operation

- The regions corresponding to the forward active, saturation and cut-off modes of operation are as indicated in the plot below.
- IV characteristics show the relationships between the collector current, i_C, and the collector-emitter voltage, v_{CE}, for different base currents, i_B, (or equivalently different v_{BE}). The current gain i_C / i_B is high around 100 in the forward active region which is used in amplifiers.





MOSFET – Introduction

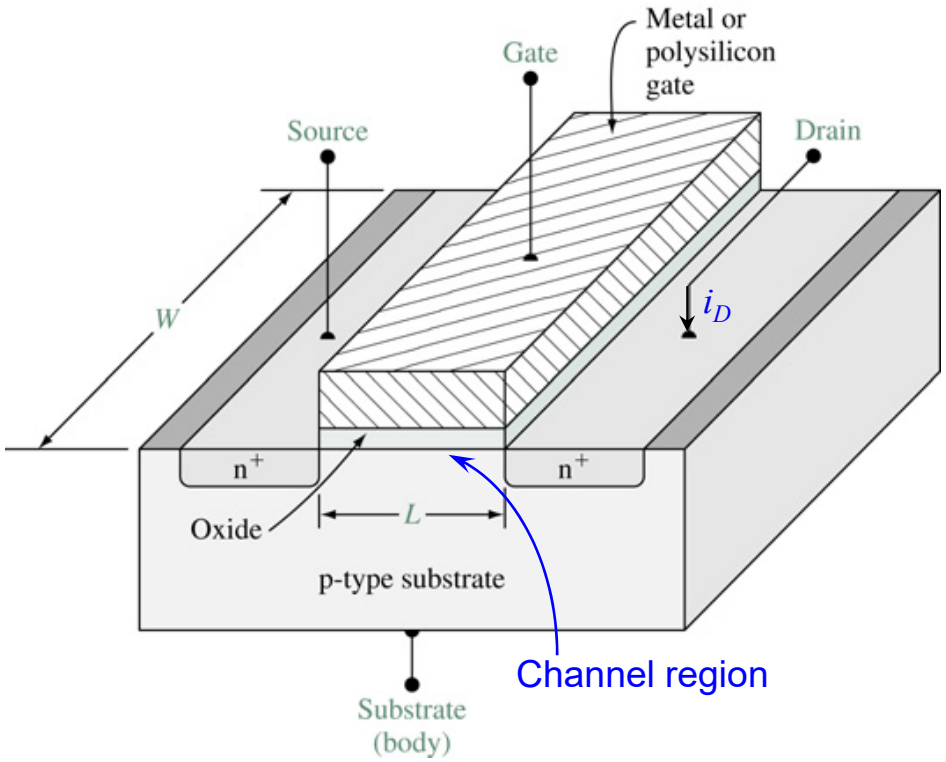


Basic structure of an n-channel MOSFET.

- Schematic on the left shows the basic structure of an *n*-channel MOSFET, also called an NMOS transistor.
- An n-channel MOSFET is made using a *p*-type single-crystal silicon substrate.
- Heavily doped n^+-type regions, created in the substrate, form the source and drain regions.
- The metal or polysilicon electrode on top of the thin oxide (dielectric) layer, between the source and drain regions, is called the gate.

Note: A11 processor has 4.3 billion TRs (10nm FinFET)

MOSFET – Introduction



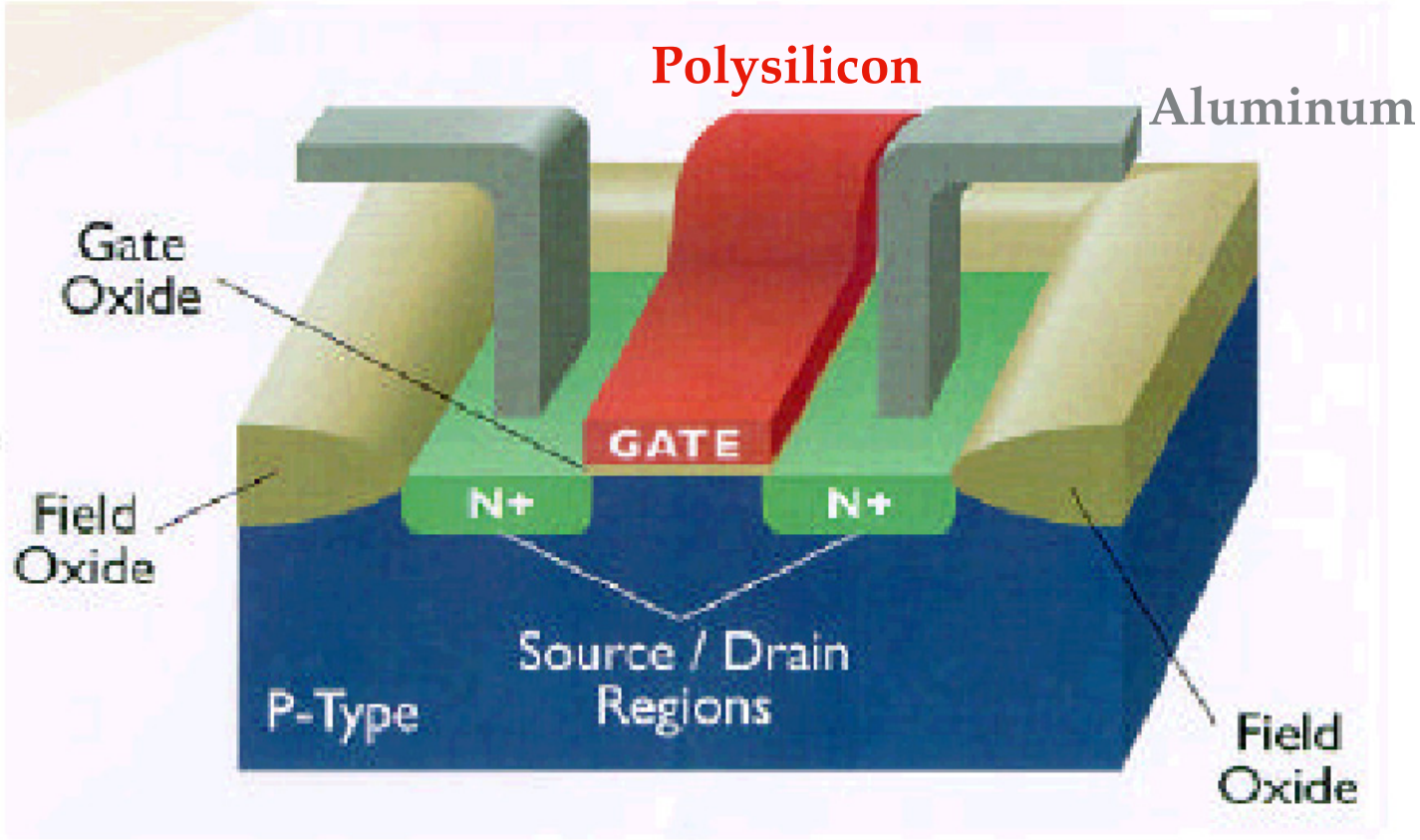
Basic structure of an n-channel MOSFET.

- Source terminal is the source of the carriers that will flow through the channel to the drain terminal.
- In an *n*-channel MOSFET, electrons (negatively charged) flow from the source to the drain.
- Conventional current therefore enters from the drain terminal and flows to the source terminal.
- Note that MOSFET has a fourth terminal - the substrate or body.
- There are two types of MOSFET: n-channel and p-channel MOSFETs.

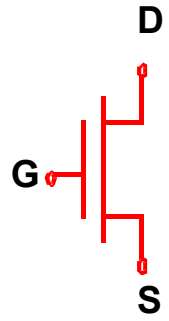
NUS National University of Singapore

As n and p-channel MOSFETs are building blocks of Complementary Metal Oxide Semiconductor (CMOS) Technology, we will discuss details in the subsequent lectures.

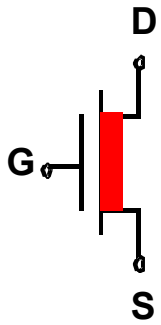
The MOS Transistor



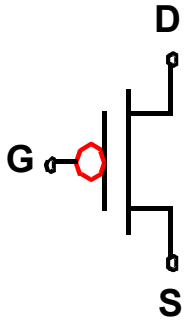
MOS Transistors-Types and Symbols



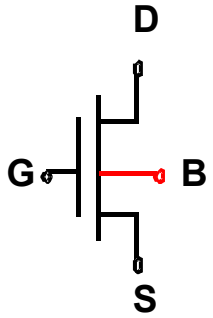
NMOS Enhancement



NMOS Depletion



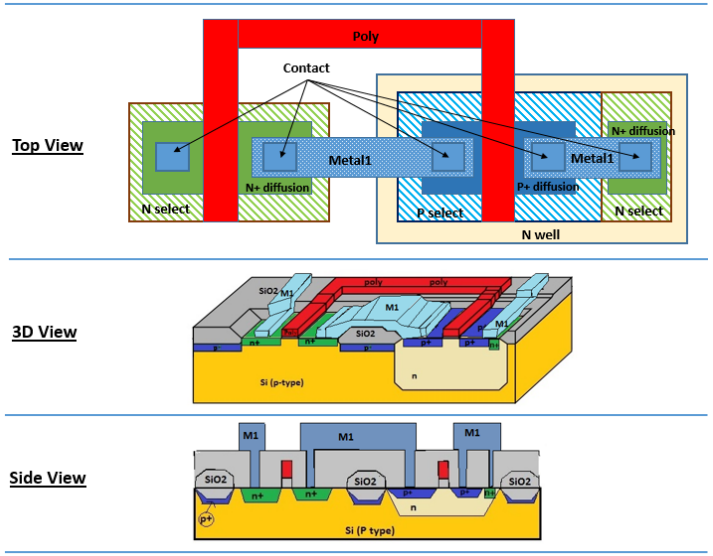
PMOS Enhancement



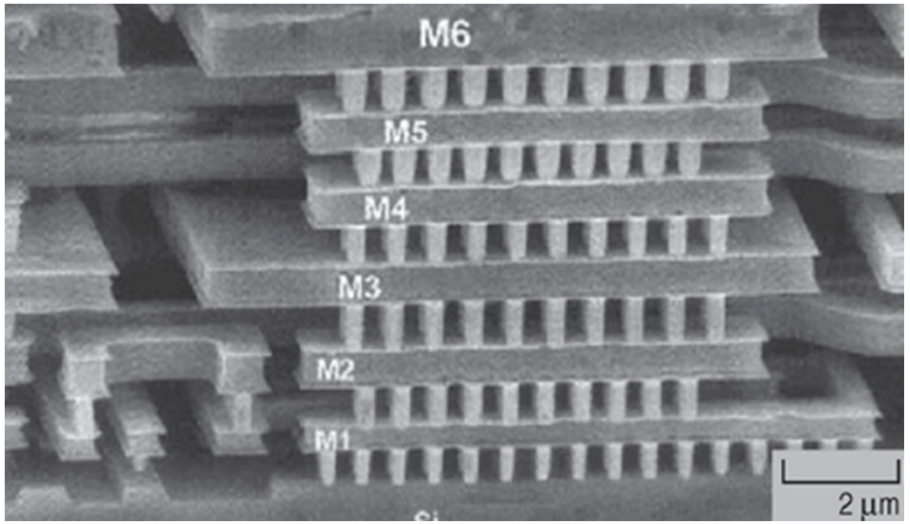
NMOS with Bulk Contact

CMOS Transistor

- Transistors have 3D structures
- We need to fabricate them layer-by-layer



* Image from VLSI-Experts

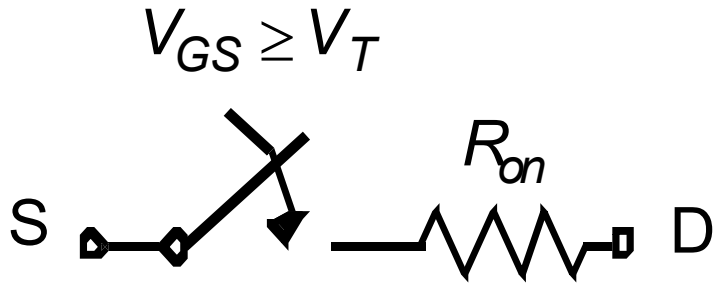


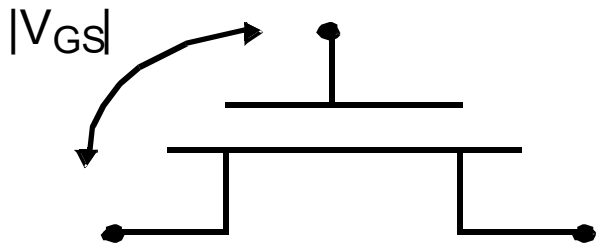
* Image from TMS

NUS National University of Singapore

What is a Transistor?

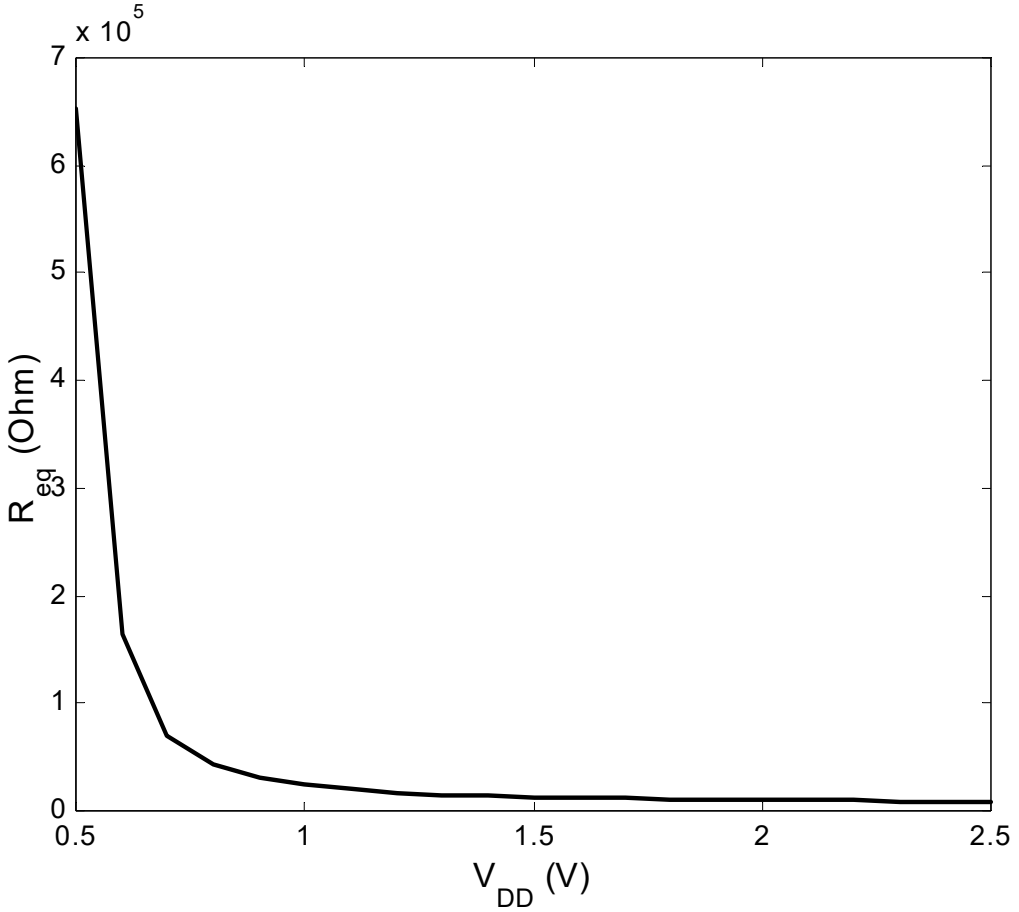
A Switch! ⟷ An MOS Transistor





The Transistor as a Switch

R_{eq} represents the channel resistance when the transistor is ON.



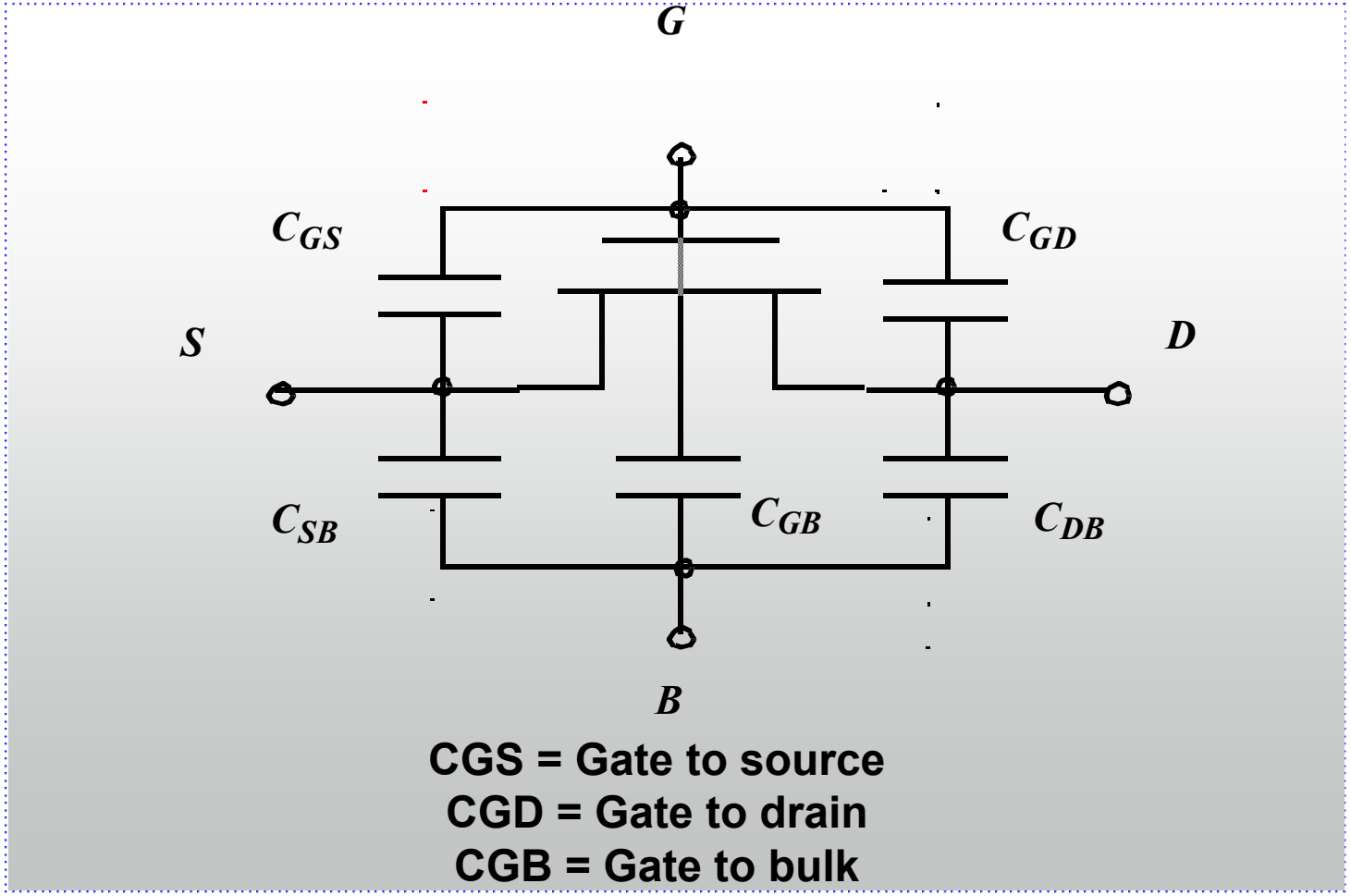
NUS National University of Singapore

The Transistor as a Switch

Table 3.3 Equivalent resistance R_{eq} (W/L= 1) of NMOS and PMOS transistors in 0.25 μm CMOS process (with $L = L_{min}$). For larger devices, divide R_{eq} by W/L.

V_{DD} (V)	1	1.5	2	2.5
NMOS (kΩ)	35	19	15	13
PMOS (kΩ)	115	55	38	31

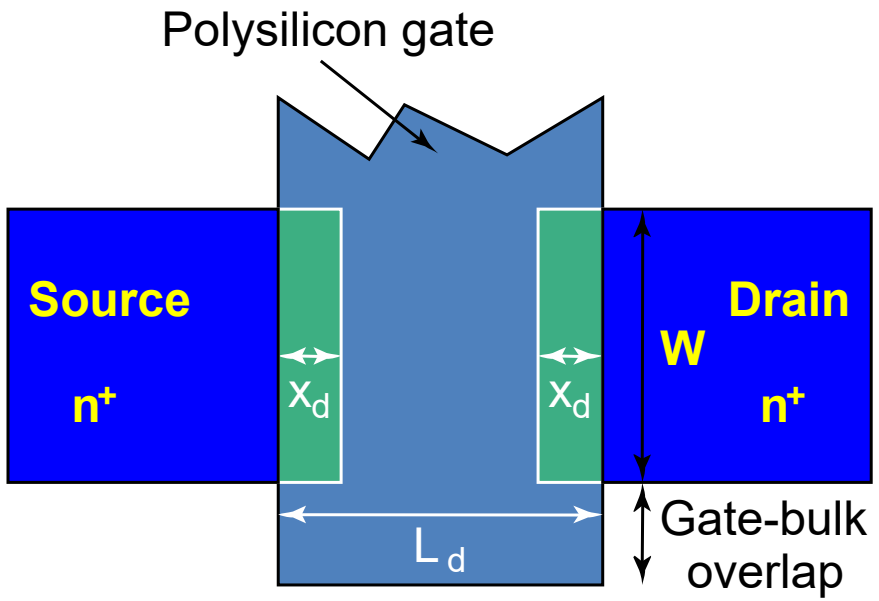
Dynamic Behavior of MOS Transistor



CGS = Gate to source
CGD = Gate to drain
CGB = Gate to bulk

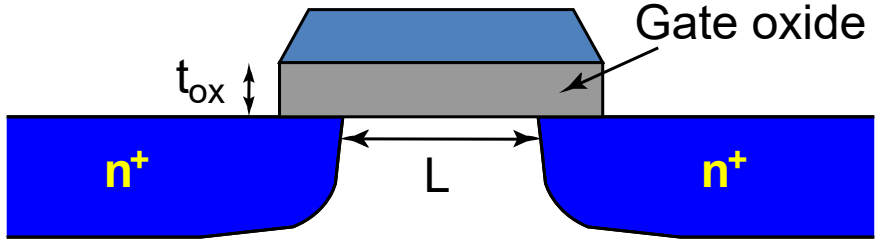
All components are non-linear and values depends on the region of operation

The Gate Capacitance



Top view

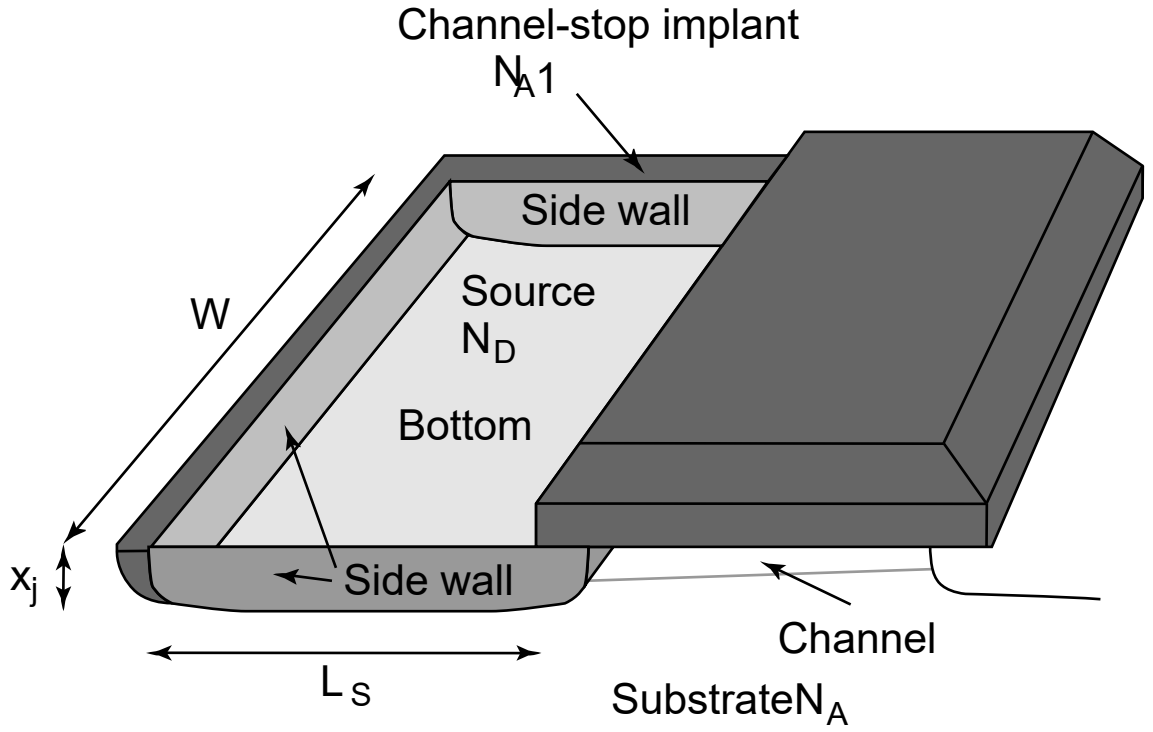
$$C_{gate} = \frac{\varepsilon_{ox}}{t_{ox}} WL$$



Cross section

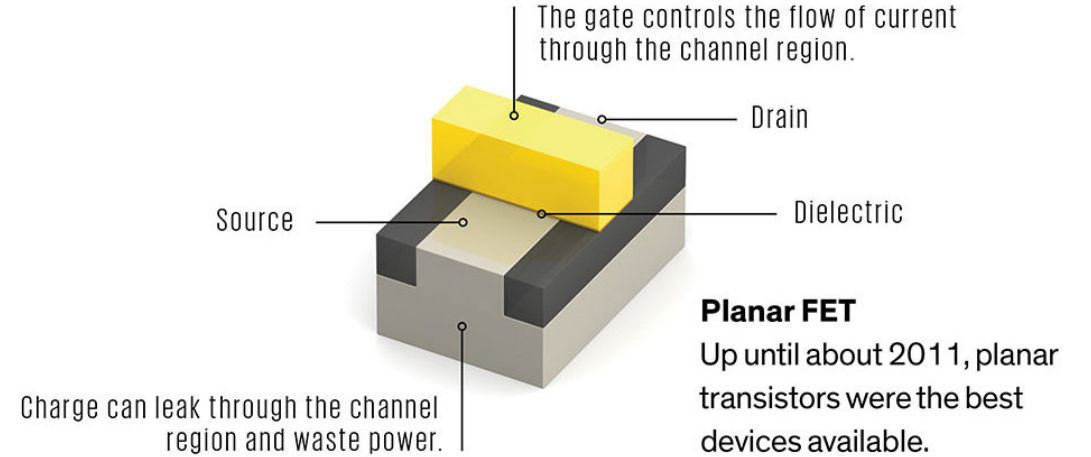
$$L_{effective} = L_{drawn} - 2x_d$$

Junction Capacitance ($C_{diffusion}$)

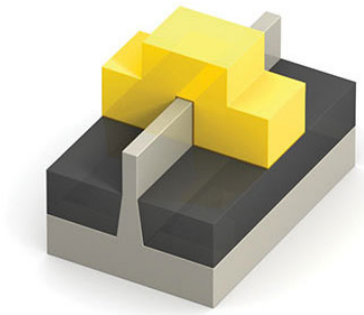


$$C_{diff} = C_{bottom} + C_{sw} = C_j \times AREA + C_{jsw} \times PERIMETER$$
$$= C_j L_S W + C_{jsw}(2L_S + W)$$

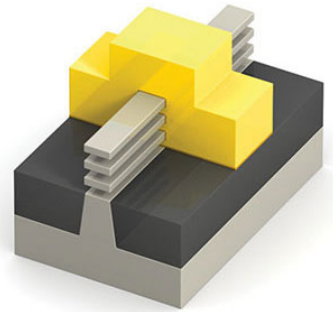
New Transistor Architectures: 3D FET



Planar FET
Up until about 2011, planar transistors were the best devices available.



FinFET
Surrounding the channel region on three sides with the gate gives better control and prevents current leakage.



Stacked nanosheet FET
The gate completely surrounds the channel regions to give even better control than the FinFET.