ELEC 501: Analog Integrated Circuit Design

Introduction and Background

Shahriar Mirabbasi

Department of Electrical and Computer Engineering
University of British Columbia
shahriar@ece.ubc.ca

Technical contributions of Pedram Lajevardi in revising these slides is greatly acknowledged.

Overview

- Marking Scheme
- References
- Motivation
- Background

Marking

Assignments	50%
Project (Presentation)	25%
Project (Report)	25%

References

- Lecture notes
- Recommended reading material:

- T. Chan Carusone, D. Johns and K. Martin, *Analog Integrated Circuit Design*, 2nd Edition, John Wiley, 2011
- W. Sansen, *Analog Design Essentials*, Springer, 2006
- Behzad Razavi, Design of Analog CMOS Integrated Circuits, McGraw-Hill, 1st or 2nd edition, 2001 or 2016
- P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th Edition, John Wiley, 2009
- D. Holberg and P. Allen, *CMOS Analog Circuit Design*, 3rd Edition, Oxford University Press, 2012
- A. Sedra and K.C. Smith (and T. Chan Carusone and V. Gaudet), Microelectronic Circuits, 5th, 6th, 7th, or (8th) Edition, Oxford University Press, 2004, 2009, 2014, 2020
- R. Sarpeshkar, Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-Inspired Systems, Cambridge University Press, 2010
- Journal and conference articles including including *IEEE Journal of Solid-State Circuits* and *International Solid-State Circuits Conference*

Fun to Watch and Check

Razavi Electronics 1 - YouTube (45 Lectures)

Razavi Electronics 2 – YouTube (46 Lectures)

William F. Brinkman, Douglas E. Haggan, and William W. Troutman, "A History of the Invention of the Transistor and Where It Will Lead Us," *IEEE Journal of Solid-State Circuits*, volume 32, no. 12, December 1997, pp. 1858-1865

Over 6 Decades of Continued Transistor Shrinkage, Innovation (intel.com)

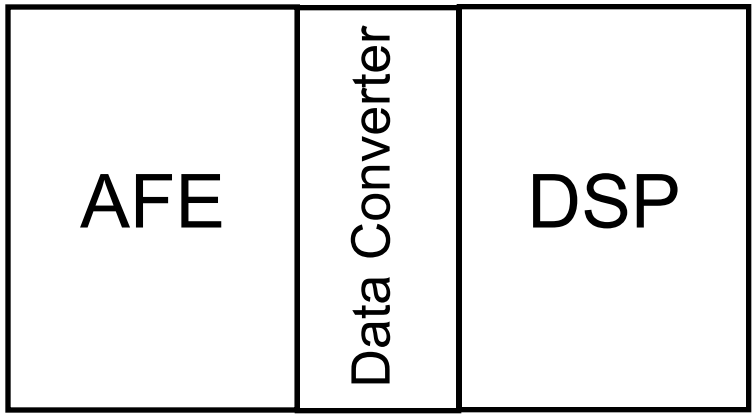
Mark Bohr 2014 IDF Session Presentation (intel.com)

Boris Murmann, "Digitally Assisted Analog Circuits," IEEE Micro, vol. 26, no. 2, pp. 38-47, Mar. 2006.

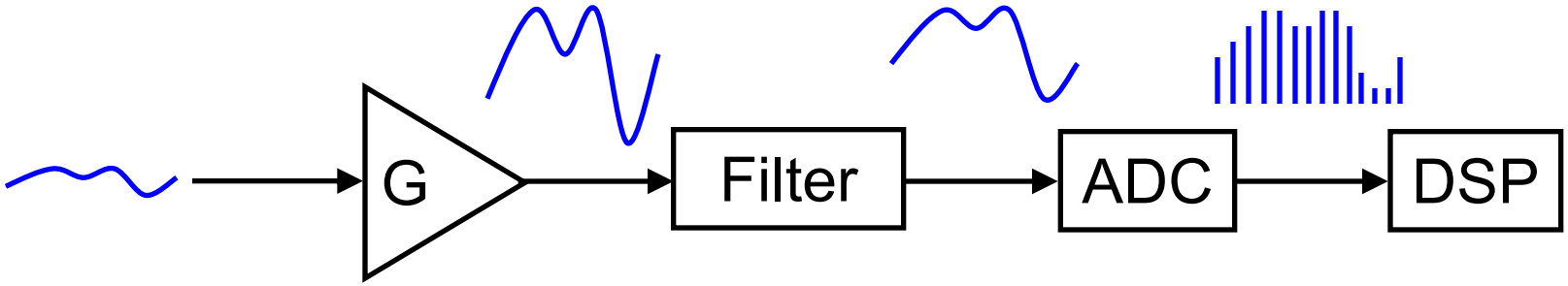
Why Analog?

- Most of the physical signals are analog in nature!
- Although digital is great we need an analog interface to convert physical signals from analog to digital
- Also, in some application after processing the signals in digital domain, we need to convert them back to analog.
- Thus in many applications analog and mixed-signal circuits are the performance bottlenecks.
- Also with constant process improvements the boundary of between high-speed digital and analog circuits becomes more and more fuzzy!
- That is why analog and mixed-signal designers are still and hopefully will be in demand for the foreseeable future.

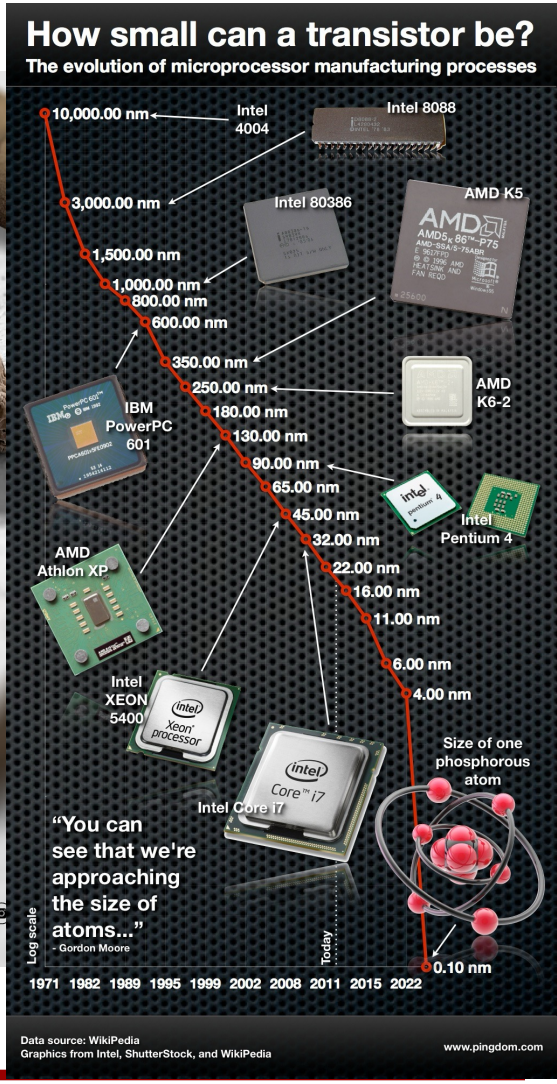
Typical Real World System



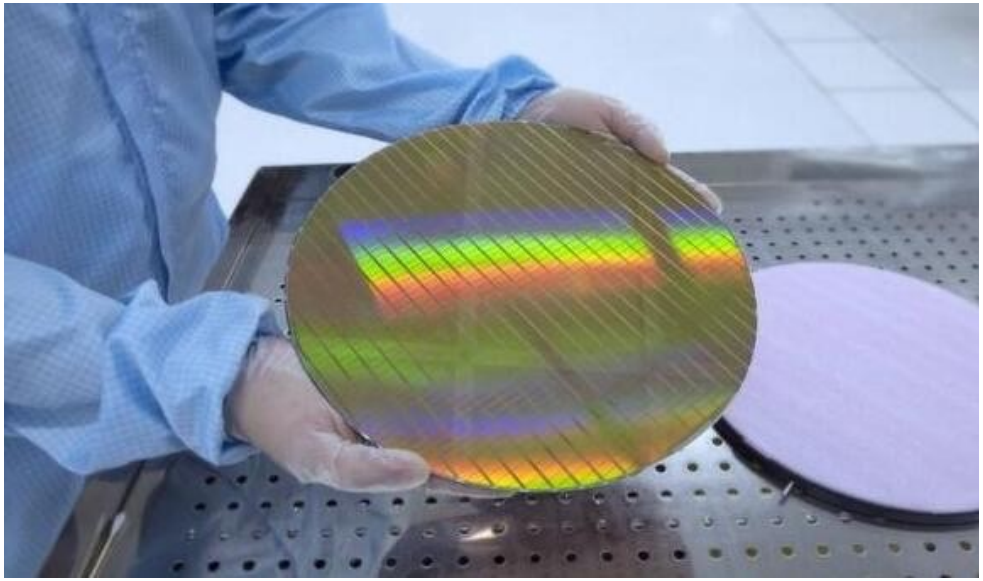
- Example:



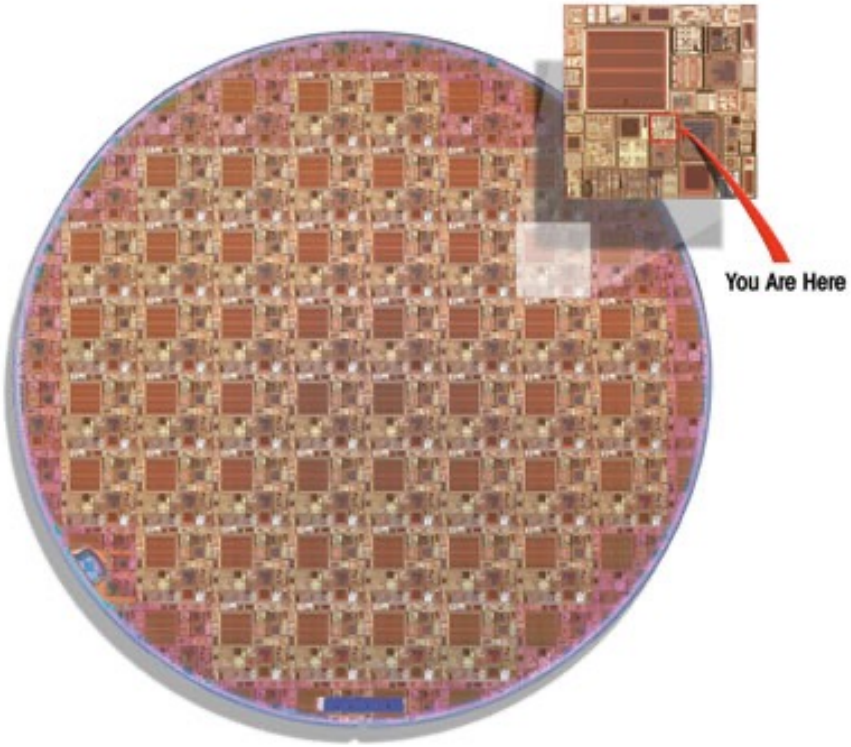

THE RELATIVE SIZE
OF PARTICLES
From the COVID-19 pandemic to the U.S. West Coast wildfires, some of the biggest threats now are also the most microscopic.
A particle needs to be 10 microns (μm) or less before it can be inhaled into your respiratory tract. But just how small are these specks?
Here's a look at the relative sizes of some familiar particles
HUMAN HAIR 50-180μm
FOR SCALE
FINE BEACH SAND 90μm
GRAIN OF SALT 60μm
WHITE BLOOD CELL 25μm
GRAIN OF POLLEN 15μm
DUST PARTICLE (PM10) <10μm
RED BLOOD CELL 7-8μm
RESPIRATORY DROPLETS 5-10μm
DUST PARTICLE (PM2.5) 2.5μm
BACTERIUM 1-3μm
WILDFIRE SMOKE 0.4-0.7μm
CORONAVIRUS 0.1-0.5μm
T4 BACTERIOPHAGE 0.225μm
ZIKA VIRUS 0.045μm
Pollen can trigger allergic reactions and hay fever—which 1 in 5 Americans experience every year.
Source: Harvard Health
The visibility limits for what the naked eye can see hovers around 10-40μm.
Respiratory droplets have the potential to carry smaller particles within them, such as dust or coronavirus.
Wildfire smoke can persist in the air for several days, and even months.
SOURCES Clearstream, Daniel Loverbey, EPA, Financial Times, News Medical, Science Direct, SCMP, Susan Sokolowski, Petroclear, U.S. Dept. of Energy
COLLABORATORS RESEARCH + WRITING Carmen Ang, Iman Ghosh | DESIGN + ART DIRECTION Harrison Schell
VISUAL CAPITALIST
/visualcapitalist


How small can a transistor be?
The evolution of microprocessor manufacturing processes
10,000.00 nm
Intel 4004
Intel 8088
3,000.00 nm
Intel 80386
AMD K5
1,500.00 nm
1,000.00 nm
800.00 nm
600.00 nm
350.00 nm
250.00 nm
AMD K6-2
IBM PowerPC 601
180.00 nm
130.00 nm
90.00 nm
65.00 nm
45.00 nm
32.00 nm
22.00 nm
16.00 nm
11.00 nm
6.00 nm
4.00 nm
Intel Pentium 4
AMD Athlon XP
Intel XEON 5400
Intel Core i7
Size of one phosphorous atom
0.10 nm
"You can see that we're approaching the size of atoms..."
- Gordon Moore
Log scale
Today
1971 1982 1989 1995 1999 2002 2008 2011 2015 2022
Data source: WikiPedia
Graphics from Intel, ShutterStock, and WikiPedia
www.pingdom.com

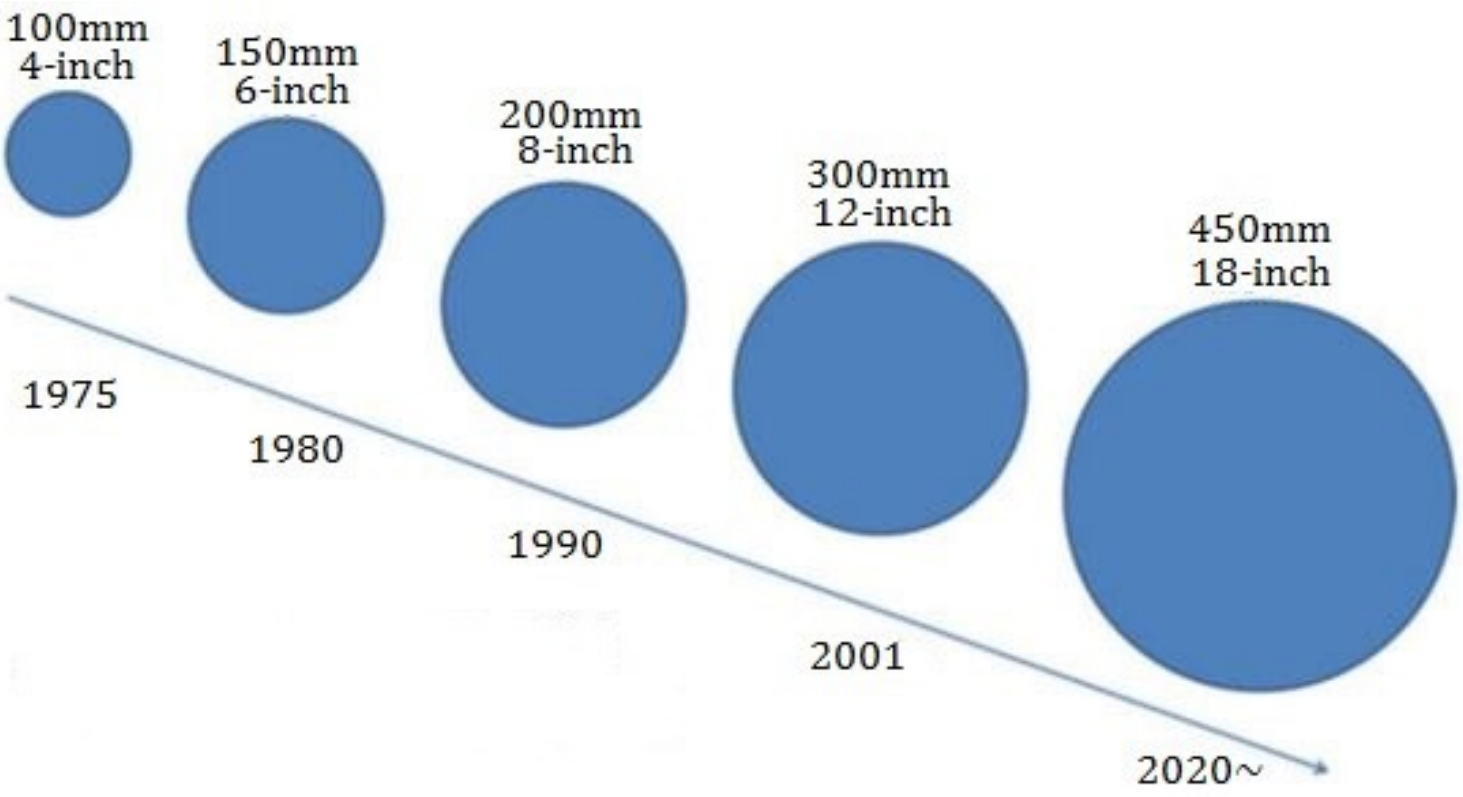
CMOS Wafer



(Image credit Kuke Electronics Limited)

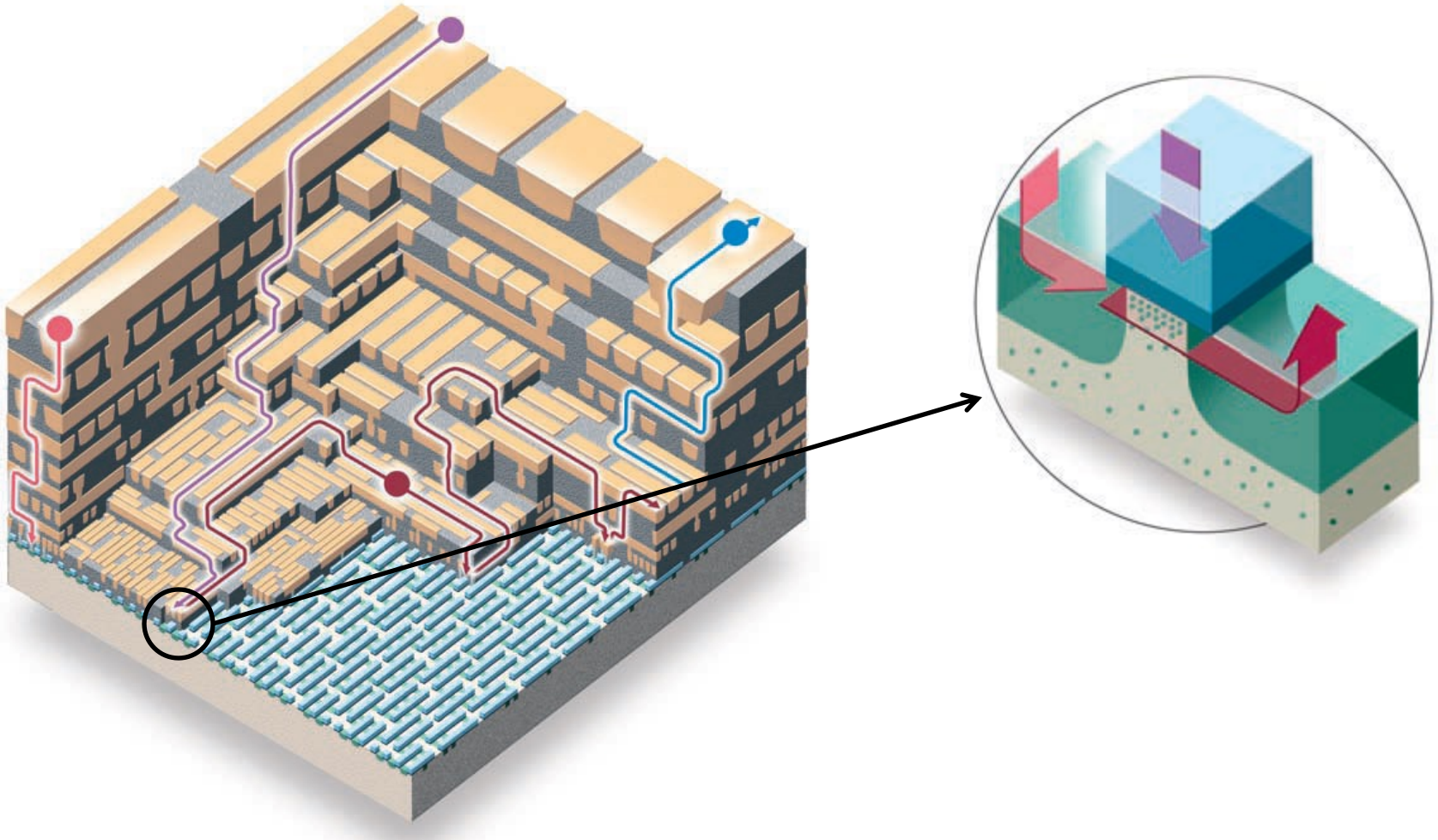


A typical multiproject wafer from Mosis accommodates the needs of integrated device manufacturers and fabless semiconductor companies.



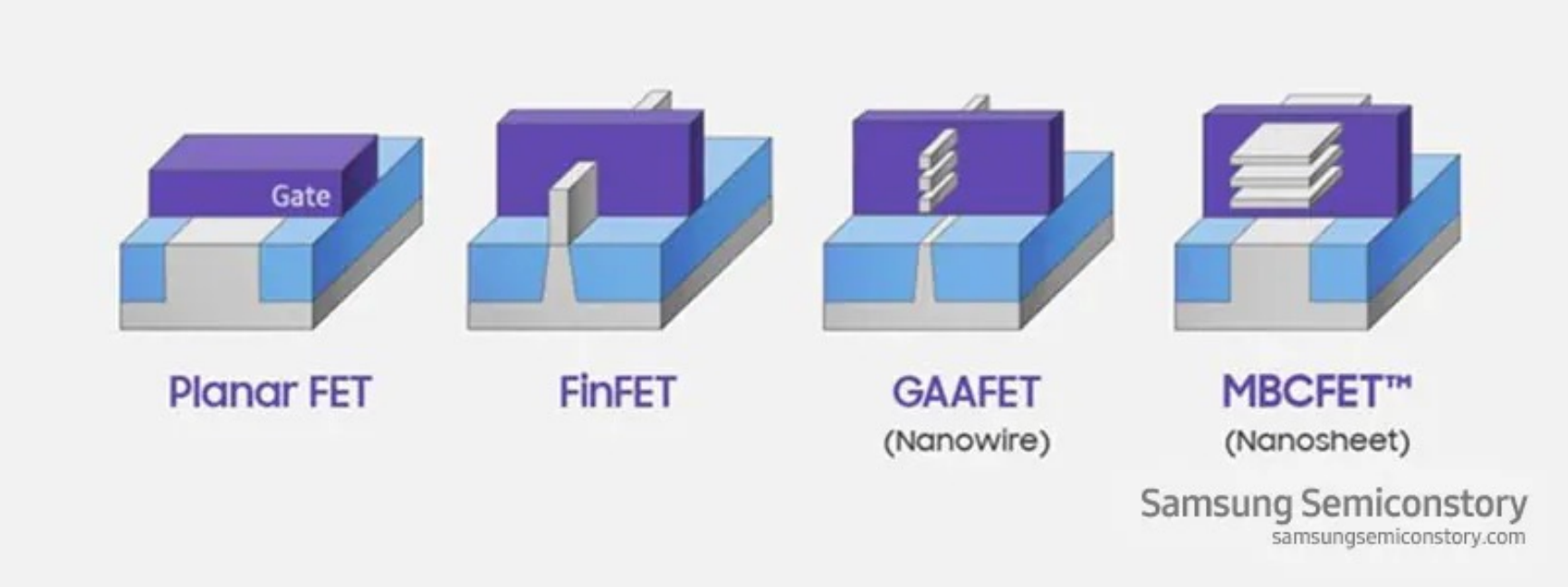
Source: Kuke Electronics Limited

Intel 45 nm Process



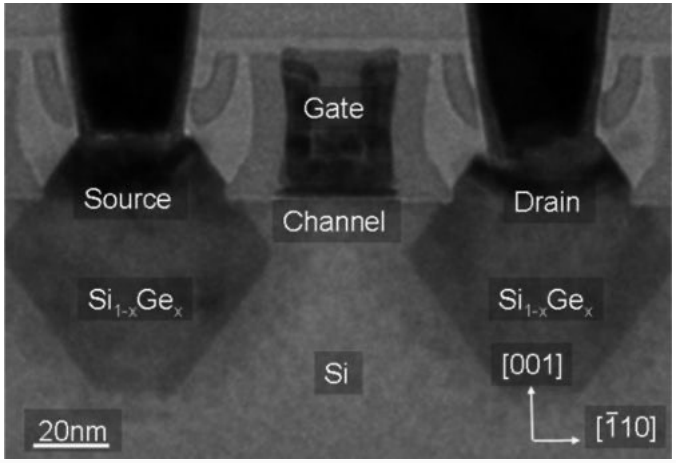
http://blog.oregonlive.com/siliconforest/2007/11/intel11.pdf

Planar to 3D Structures

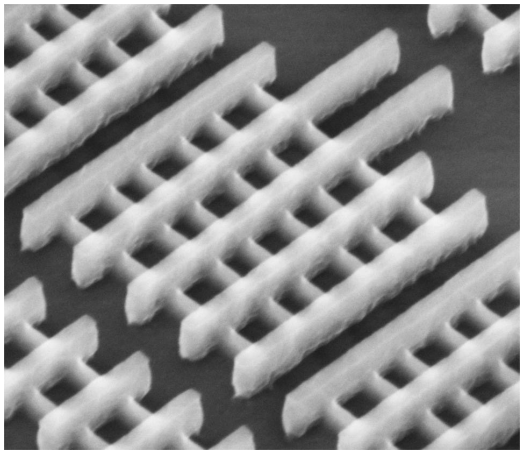


https://semiconductor.samsung.com/
MBCFET (Multi-Bridge Channel Field Effect Transistor)

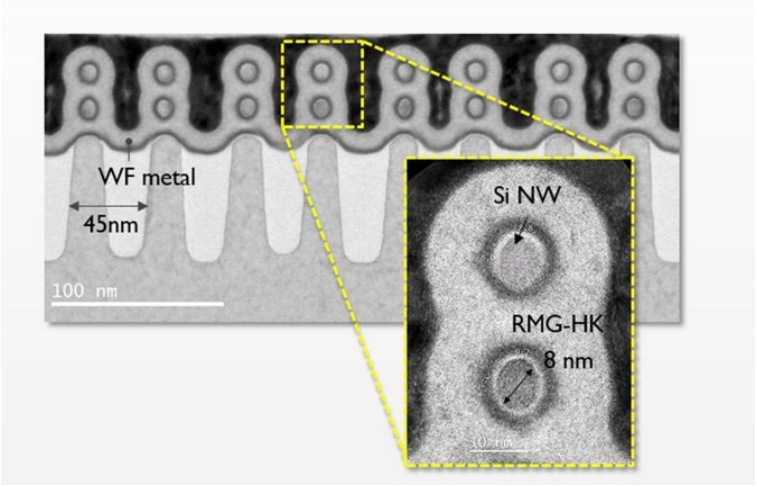
FinFET



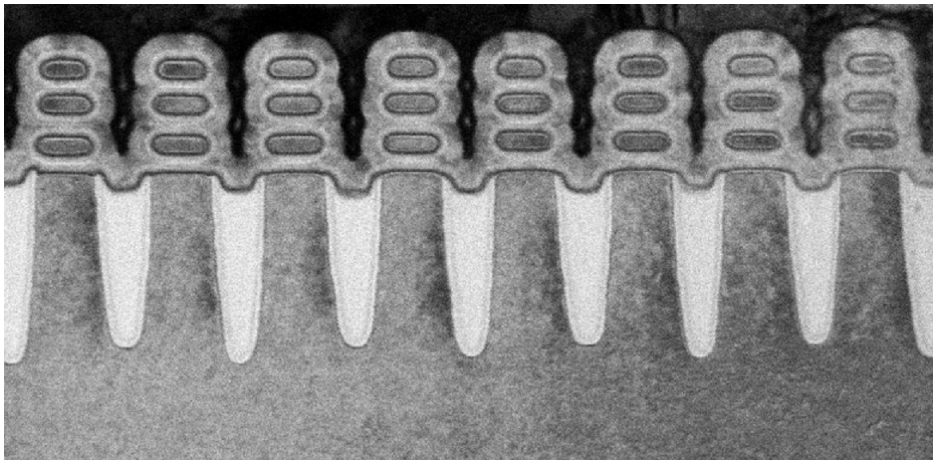
Source: Wiley, Analytical Science



Source: Intel

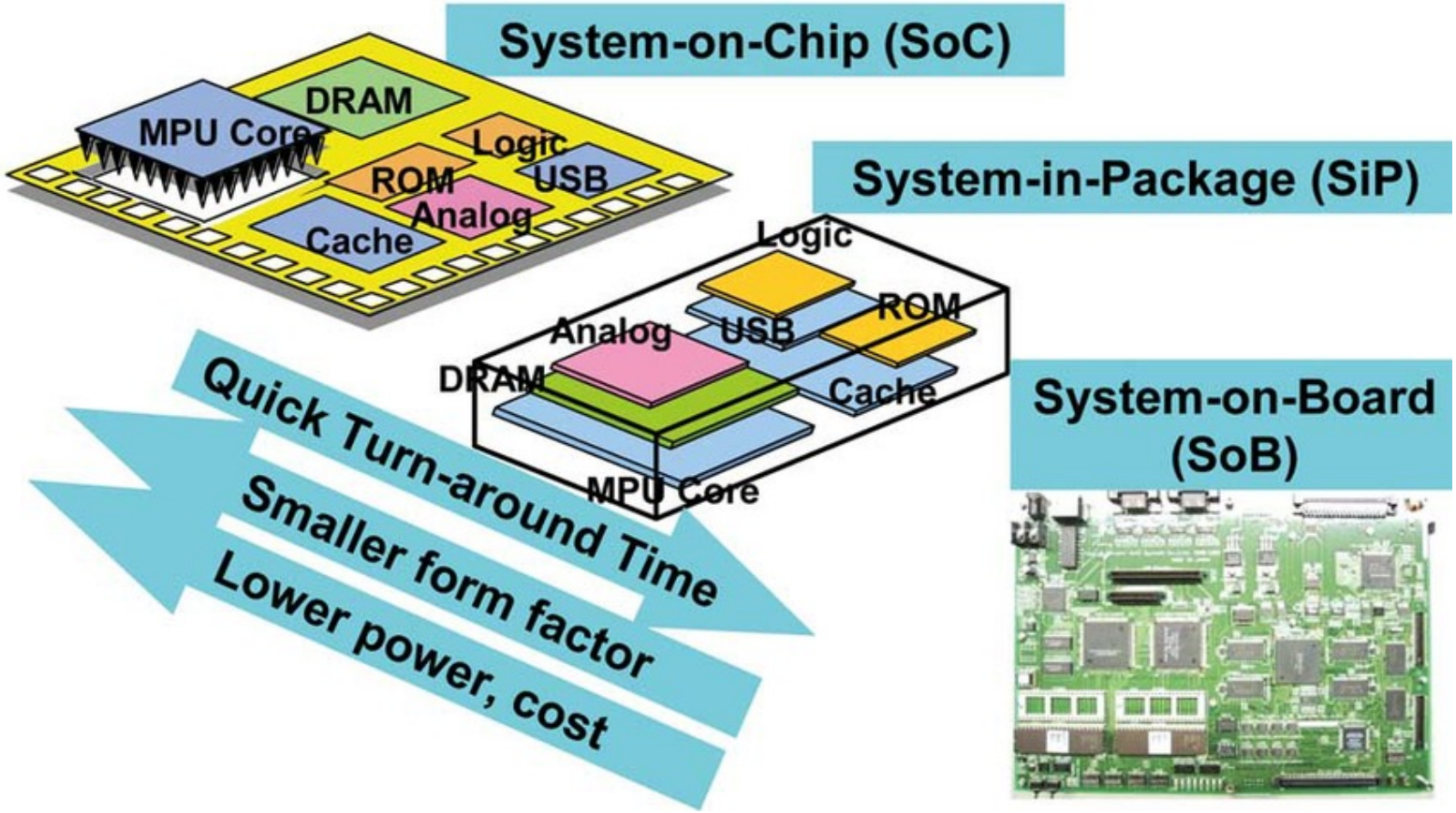


Source: Imec



Source: IBM

System

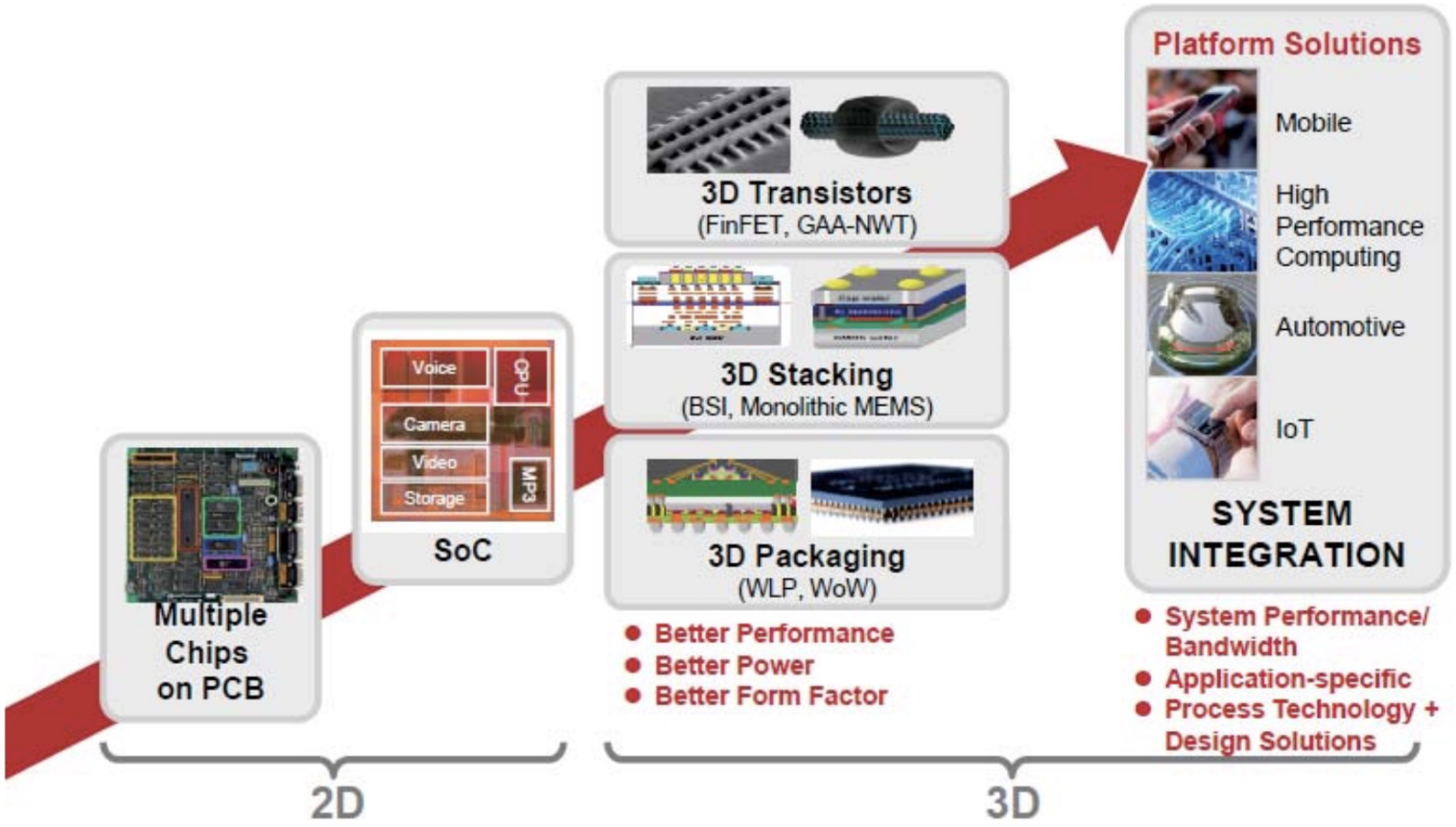


Source: R. Saleh *et al*., "System-on-Chip: Reuse and Integration," *Proceedings of the IEEE*, vol. 94, no. 6, pp. 1050-1069, June 2006, doi: 10.1109/JPROC.2006.873611.

Technology Scaling and Roadmap (Intel)

Source: “Moore’s Law – Now and in the Future,” www.intel.com

Chip and System Integration



Source: Cliff Hou (TSMC), ISSCC 2017 (Plenary)

Background

- Most of the material in this set are based on

  Chapters 2, 16, and 17 of the Razavi's book: Design of Analog CMOS Integrated Circuits

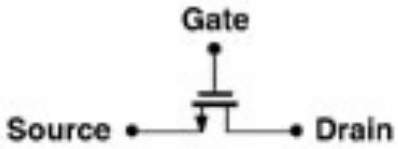
Many of the figures in this set are from © *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, 2001, unless otherwise noted.

Transistor

- Transistor stands for …
- Transistor are semiconductor devices that can be classified as
  - Bipolar Junction Transistors (BJTs)
  - Field Effect Transistors (FETs)
    - Depletion-Mode FETs or (e.g., JFETs)
    - Enhancement-Mode FETs (e.g., MOSFETs)

Simplistic (Digital) Model

- MOS transistors have three terminals: Gate, Source, and Drain

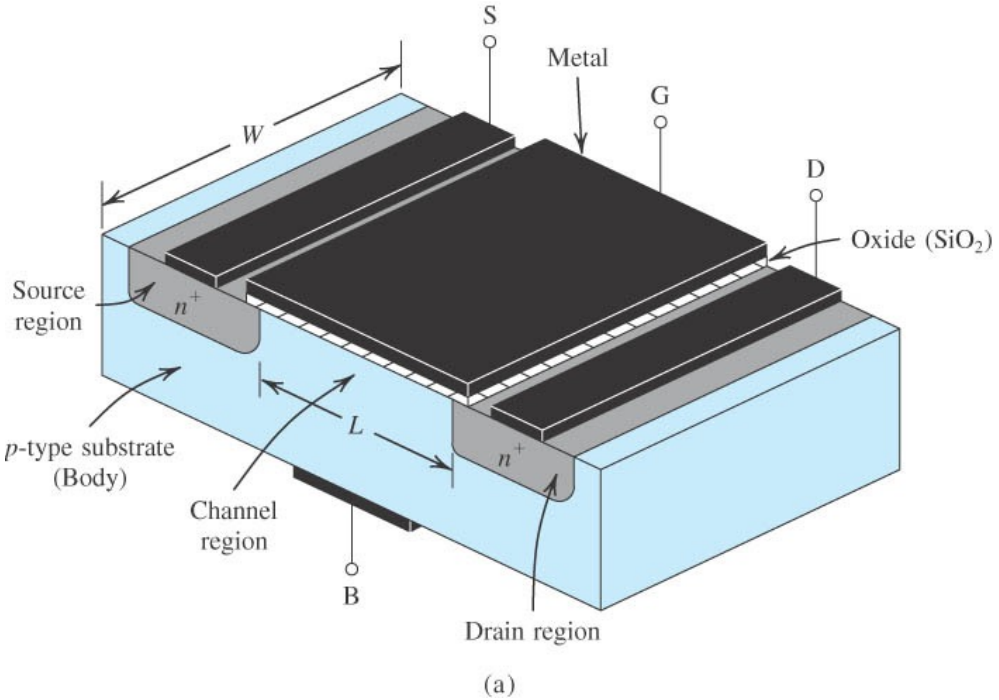


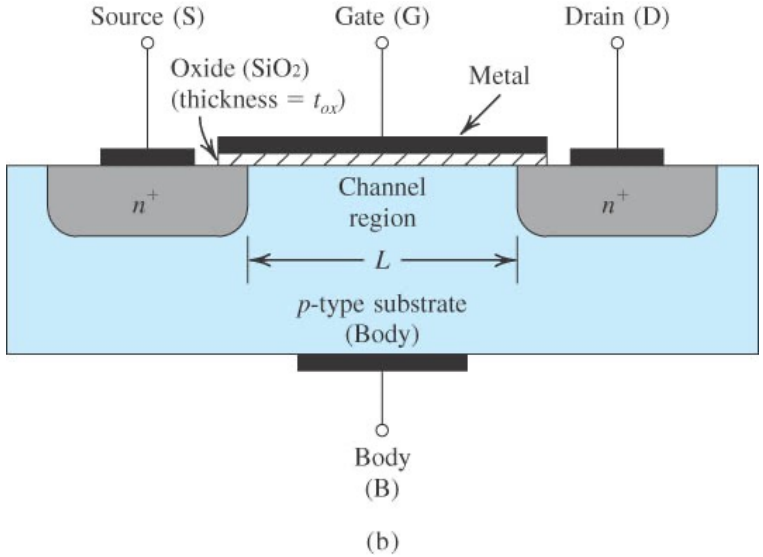
- The voltage of the Gate terminal determines the type of connection between Source and Drain (Short or Open).
- Thus, MOS devices behave like a switch

	NMOS	PMOS
V_G high	Device is ON D is shorted to S	Device is OFF D & S are disconnected
V_G low	Device is OFF D & S are disconnected	Device is ON D is shorted to S

Physical Structure - 1

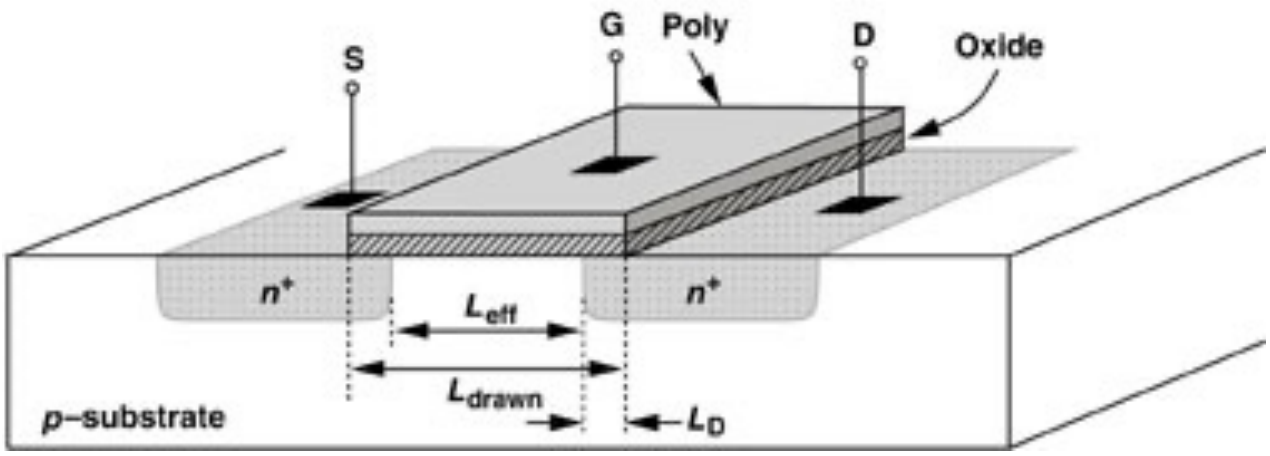
- Source and Drain terminals are identical except that Source provides charge carriers and Drain receives them.





Physical Structure - 2

- Charge Carriers are electrons in NMOS devices, and holes in PMOS devices.
- Electrons have a higher mobility than holes
- So, NMOS devices are faster than PMOS devices
- We rather to have a p-type substrate?!



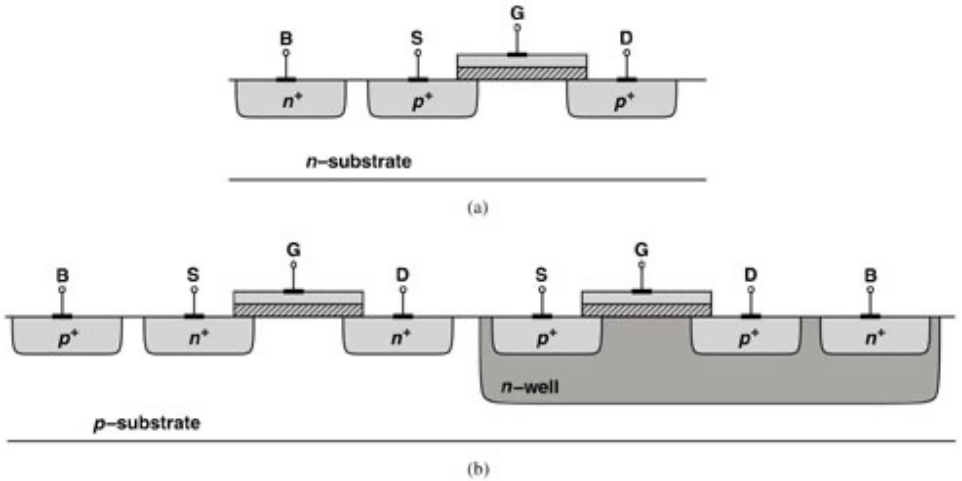
L_D: Due to Side Diffusion

Poly-silicon used instead of Metal for fabrication reasons

- Actual length of the channel (L_{eff}) is less than the width of poly

Physical Structure - 3

- N-wells allow both NMOS and PMOS devices to reside on the same piece of die.



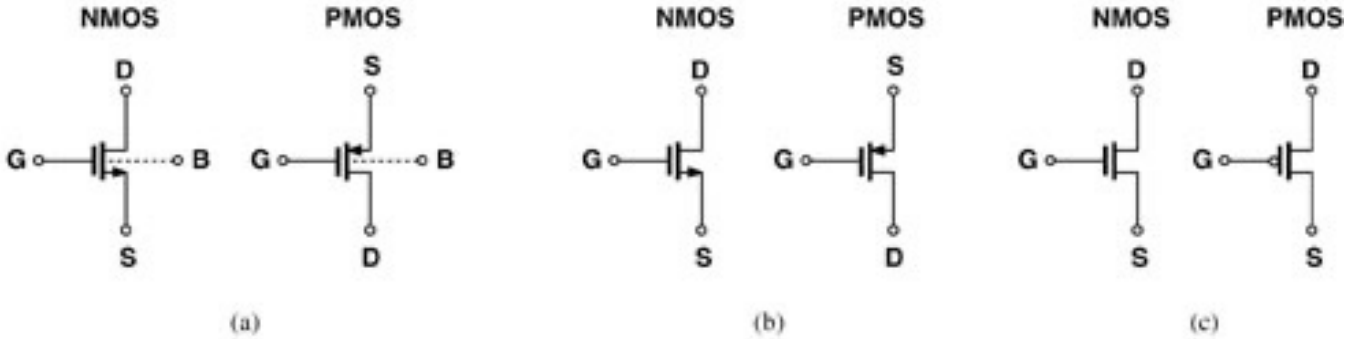
- NMOS and PMOS devices have in fact 4 terminals:

  Source, Drain, Gate, Substrate (bulk)

- In order to have all PN junctions reverse-biased, substrate of NMOS is connected to the most negative voltage, and substrate of PMOS is connected to the most positive voltage.

Physical Structure - 4

- MOS transistor Symbols:



- In NMOS Devices: $Source \xrightarrow{electron} Drain$

  Current flows from Drain to Source

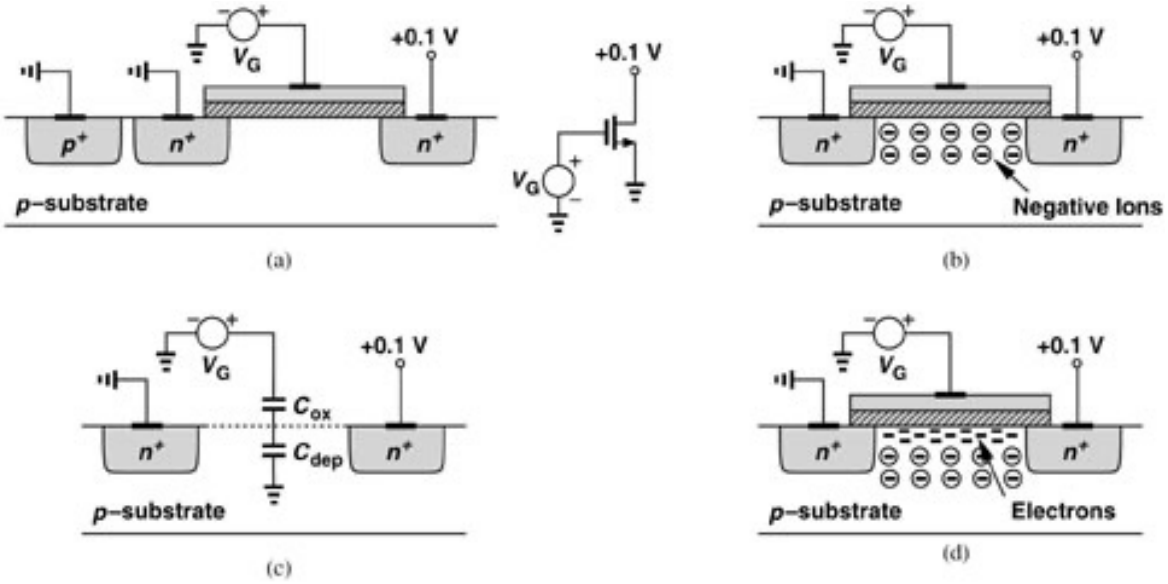
- In PMOS Devices: $Source \xrightarrow{hole} Drain$

  Current flows from Source to Drain

- Current flow determines which terminal is Source and which one is Drain. Equivalently, source and drain can be determined based on their relative voltages.

Threshold Voltage - 1

- As the gate voltage is increased, the surface under the gate is depleted. If the gate voltage increases more, free electrons appear under the gate and a conductive channel is formed.



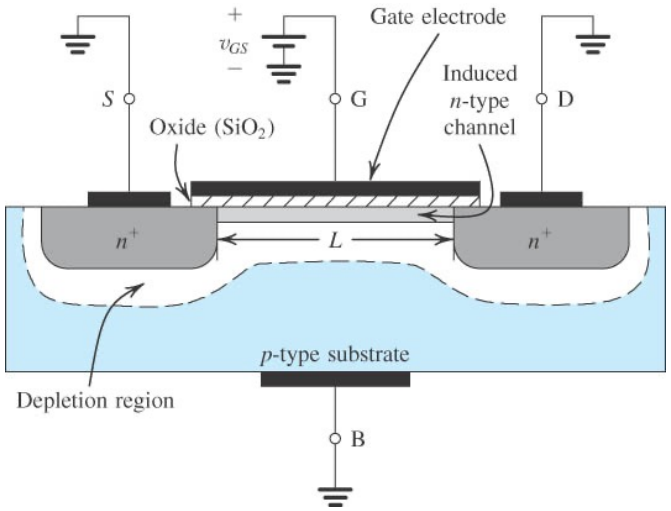
(a) An NMOS driven by a gate voltage, (b) formation of depletion region, (c) onset of inversion, and (d) channel formation

- As mentioned before, in NMOS devices charge carriers in the channel under the gate are electrons.

Threshold Voltage - 2

- Intuitively, the threshold voltage is the gate voltage that forces the interface (surface under the gate) to be completely depleted of charge (in NMOS the interface is as much n-type as the substrate is p-type)

- Increasing gate voltage above this threshold (denoted by V_{TH} or V_t) induces an inversion layer (conductive channel) under the gate.



Threshold Voltage - 3

Analytically:

$$V_{TH} = \Phi_{MS} + 2 \cdot \left|\Phi_F\right| + \frac{\left|Q_{dep}\right|}{C_{ox}}$$

Where:

$$\Phi_{MS} = \text{Built-in Potential} = \Phi_{gate} - \Phi_{Silicon}$$

$$= \text{the difference between the work functions of the polysilicon gate and the silicon substrate}$$

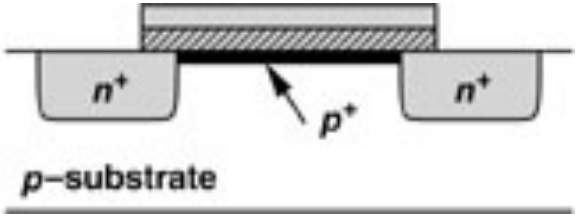
$$\Phi_F = \text{Work Function (electrostatic potential)} = \frac{K \cdot T}{q} \cdot \ln\left(\frac{N_{sub}}{n_i}\right)$$

$$Q_{dep} = \text{Charge in the depletion region} = \sqrt{4 \cdot q \cdot \varepsilon_{si} \cdot \left|\Phi_F\right| \cdot N_{sub}}$$

Threshold Voltage - 4

- In practice, the "native" threshold value may not be suited for circuit design, e.g., V_{TH} may be zero and the device may be on for any positive gate voltage.

- Typically threshold voltage is adjusted by ion implantation into the channel surface (doping P-type material will increase V_{TH} of NMOS devices).



- When V_{DS} is zero, there is no horizontal electric-field present in the channel, and therefore no current between the source to the drain.

- When V_{DS} is more than zero, there is some horizontal electric-field which causes a flow of electrons from source to drain.

Long Channel Current Equations - NMOS

- Current Equation for NMOS:

$$I_D = I_{DS} = \begin{cases} 0 \quad ; if\ V_{GS} < V_{TH}\ (Cut-off) \\ \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \left(V_{GS} - V_{TH}\right) \cdot V_{DS} \quad ; if\ V_{GS} > V_{TH}\ , V_{DS} << 2(V_{GS} - V_{TH})\ (Deep\ Triode) \\ \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[\left(V_{GS} - V_{TH}\right) \cdot V_{DS} - \frac{1}{2} \cdot V_{DS}^2\right] \quad ; if\ V_{GS} > V_{TH}\ , V_{DS} < V_{GS} - V_{TH}\ (Triode) \\ \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH})^2 \quad ; if\ V_{GS} > V_{TH}\ , V_{DS} > V_{GS} - V_{TH}\ (Saturation\ ) \end{cases}$$

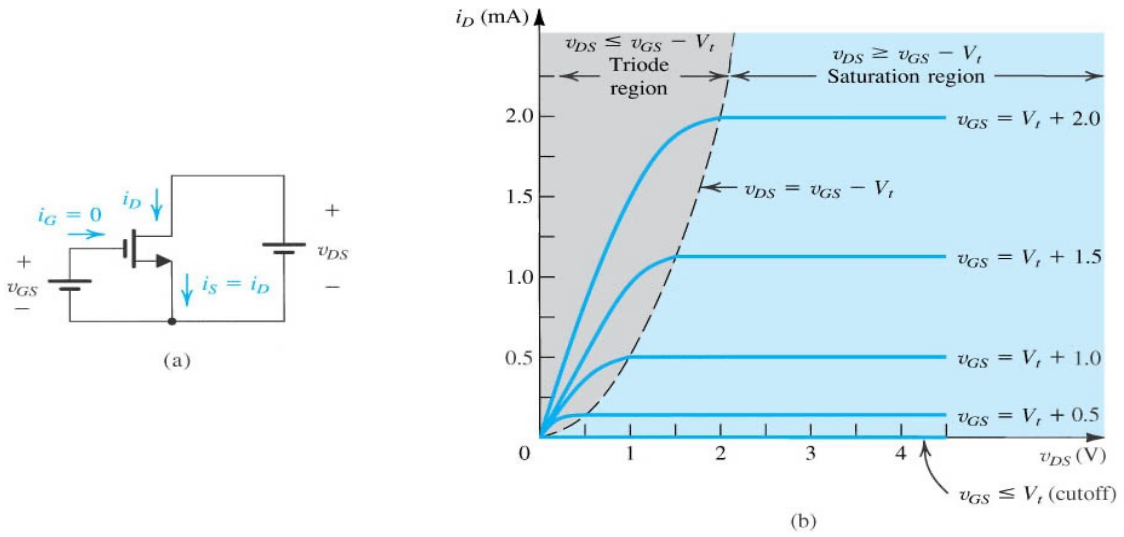
Long Channel Current Equations - PMOS

- Current Equation for PMOS:

$$I_D = I_{SD} = \begin{cases} 0 \quad ; if\ V_{SG} < |V_{TH}|\ (Cut-off) \\ \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot \left(V_{SG} - |V_{TH}|\right) \cdot V_{SD} \quad ; if\ V_{SG} > |V_{TH}|, V_{SD} << 2(V_{SG} - |V_{TH}|)\ (Deep\ Triode) \\ \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[\left(V_{SG} - |V_{TH}|\right) \cdot V_{SD} - \frac{1}{2} \cdot V_{SD}^2\right] \quad ; if\ V_{SG} > |V_{TH}|, V_{SD} < V_{SG} - |V_{TH}|\ (Triode) \\ \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - |V_{TH}|)^2 \quad ; if\ V_{SG} > |V_{TH}|, V_{SD} > V_{SG} - |V_{TH}|\ (Saturation) \end{cases}$$

Regions of Operation - 1

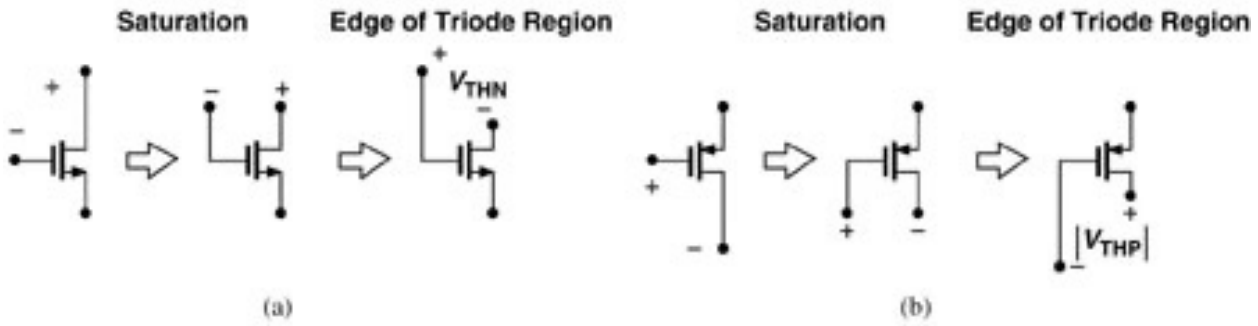
- Regions of Operation:
  Cut-off, triode (linear), and saturation (active or pinch-off)



- Once the channel is pinched off, the current through the channel is almost constant. As a result, the I-V curves have a very small slope in the pinch-off (saturation) region, indicating the large channel resistance.

Regions of Operation - 2

- The following illustrates the transition from pinch-off to triode region for NMOS and PMOS devices.



- For NMOS devices:

  If V_D increases (V_G Const.), the device will go from Triode to Pinch-off.

  If V_G increases (V_D Const.), the device will go from Pinch-off to Triode.

** In NMOS, as V_{DG} increases the device will go from Triode to Pinch-off.

- For PMOS devices:

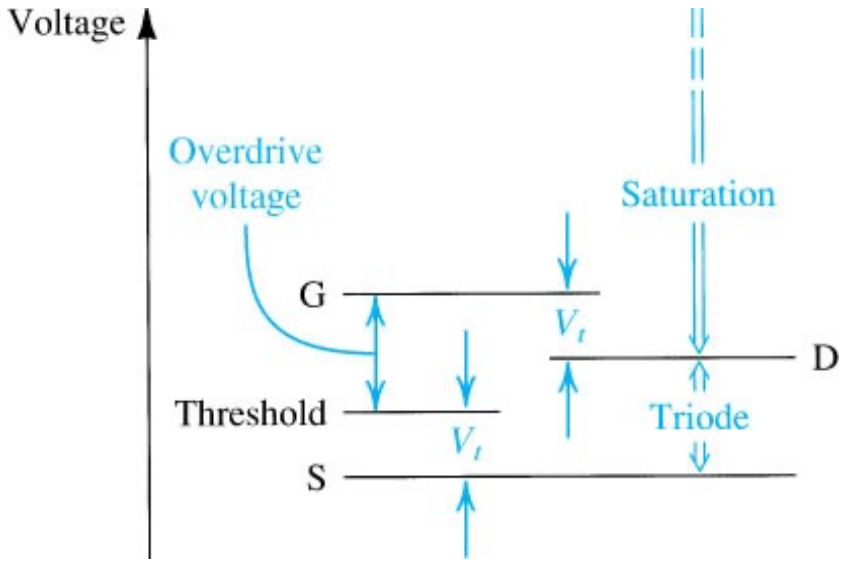
  If V_D decreases (V_G Const.), the device will go from Triode to Pinch-off.

  If V_G decreases (V_D Const.), the device will go from Pinch-off to Triode.

** In PMOS, as V_{GD} increases the device will go from Pinch-off to Triode.

Regions of Operation - 3

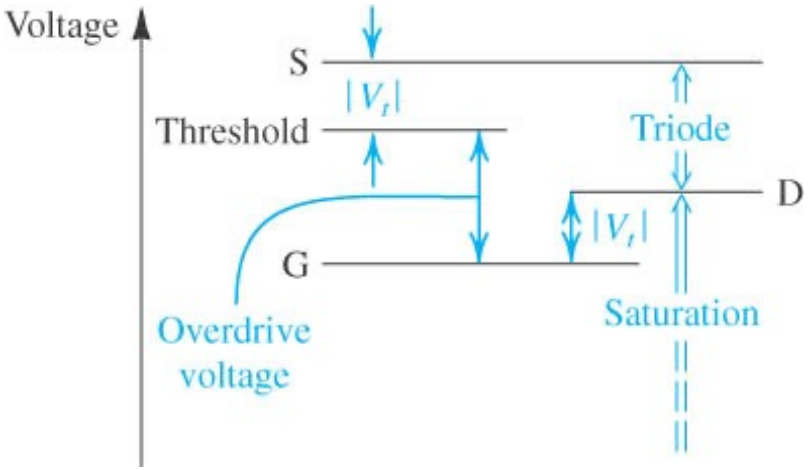
- NMOS Regions of Operation:



- Relative levels of the terminal voltages of the enhancement-type NMOS transistor for different regions of operation.

Regions of Operation - 4

- PMOS Regions of Operation:



- The relative levels of the terminal voltages of the enhancement-type PMOS transistor for different regions of operation.

Transconductance - 1

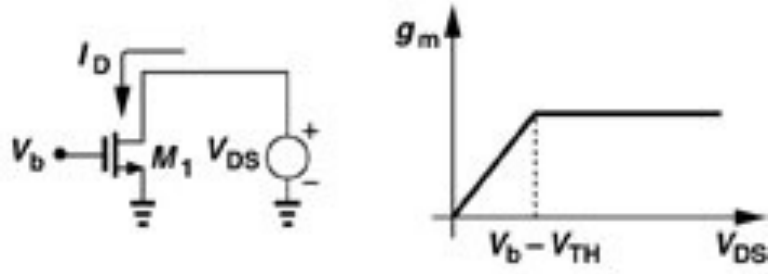
- The drain current of the MOSFET in saturation region is ideally a function of gate overdrive voltage (effective voltage). In fact, it is also a function of V_{DS}.
- It makes sense to define a figure of merit that indicates how well the device converts a voltage to current.
- Which current are we talking about?
- What voltage is in the designer's control?
- What is this figure of merit?

$$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}=Const.}$$

Transconductance - 2

Example:

Plot the transconductance of the following circuit as a function of V_{DS} (assume V_b is a constant voltage).



- Transconductance in triode:

$$g_m = \frac{\partial}{\partial V_{GS}}\left(\mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[\left(V_{GS} - V_{TH}\right) \cdot V_{DS} - \frac{1}{2} \cdot V_{DS}^2\right]\right)\Bigg|_{V_{DS} = Const.}$$

$$= \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot V_{DS}$$

- Transconductance in saturation:

$$g_m = \frac{\partial}{\partial V_{GS}}\left(\frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH})^2\right)\Bigg|_{V_{DS} = Const.}$$

$$= \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH})$$

- Moral: Transconductance drops if the device enters the triode region.

Transconductance - 3

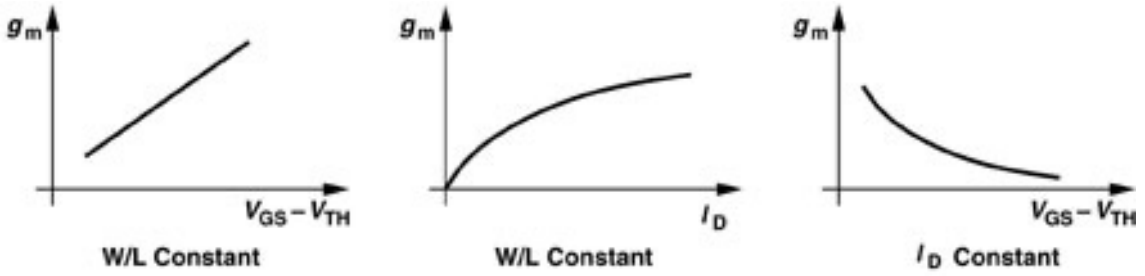
- Transconductance, g_m, in saturation:

$$g_m = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH}) = \sqrt{2\mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D} = \frac{2 \cdot I_D}{V_{GS} - V_{TH}}$$

- If the aspect ratio is constant: g_m depends linearly on (V_{GS} - V_{TH}).
  Also, g_m depends on square root of I_D.

- If I_D is constant: g_m is inversely proportional to (V_{GS} - V_{TH}).
  Also, g_m depends on square root of the aspect ratio.

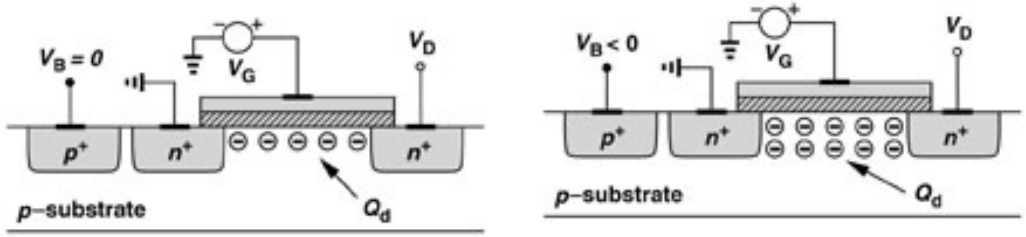
- If the overdrive voltage is constant: g_m depends linearly on I_D.
  Also, g_m depends linearly on the aspect ratio.



Second-Order Effects (Body Effect)

Substrate Voltage:

- So far, we assumed that the bulk and source of the transistor are at the same voltage (V_B=V_S).
- If $V_B > V_s$, then the bulk-source PN junction will be forward-biased, and the device will not operate properly.
- If $V_B < V_s$,
  - the bulk-source PN junction will be reverse-biased.
  - the depletion region widens, and Q_{dep} increases.
  - V_{TH} will be increased (Body effect or Backgate effect).



- It can be shown that (what is the unit for γ ?):

$$V_{TH} = V_{TH0} + \gamma \cdot \left(\sqrt{\left|2 \cdot \Phi_F + V_{SB}\right|} - \sqrt{\left|2 \cdot \Phi_F\right|} \right) \quad \text{where } \gamma = \frac{\sqrt{2 \cdot q \cdot \varepsilon_{si} \cdot N_{sub}}}{C_{ox}}$$

Body Effect - 2

Example:

Consider the circuit below:

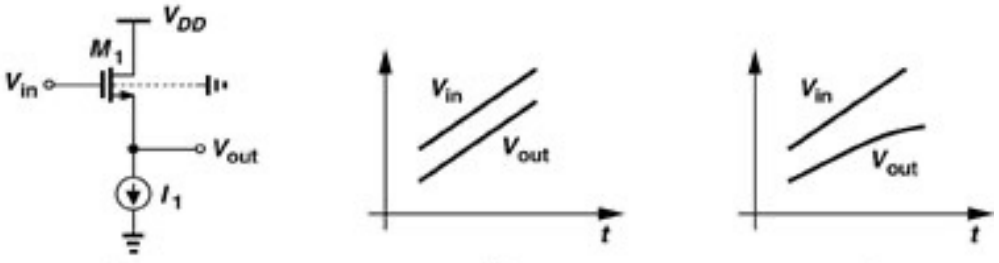
- If body-effect is ignored, V_{TH} will be constant, and I_1 will only depend on $V_{GS1}=V_{in}-V_{out}$. Since I_1 is constant, $V_{in}-V_{out}$ remains constant.

$$V_{in} - V_{out} - V_{TH} = C = Const. \rightarrow V_{in} - V_{out} = V_{TH} + C = D = Conts.$$

- In general I_1 depends on $V_{GS1}- V_{TH} =V_{in}-V_{out}-V_{TH}$ (and with body effect V_{TH} is not constant). Since I_1 is constant, $V_{in}-V_{out}-V_{TH}$ remains constant:

$$V_{in} - V_{out} - V_{TH} = C = Const. \rightarrow V_{in} - V_{out} = V_{TH} + C$$

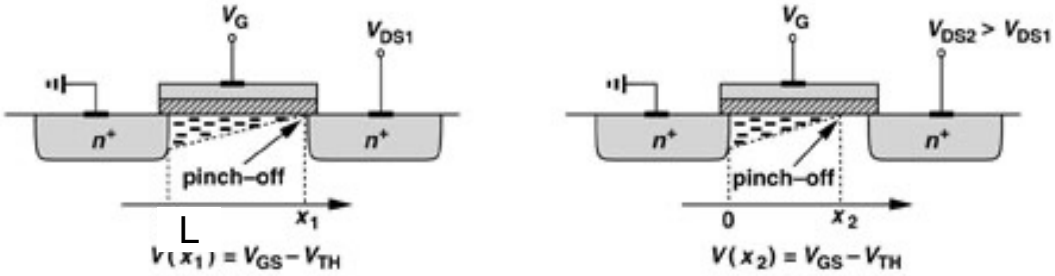
- As V_{out} increases, V_{SB1} increases, and as a result V_{TH} increases. Therefore, $V_{in}-V_{out}$ Increases.



No Body Effect With Body Effect

Channel Length Modulation - 1

- When the transistor is in the saturation region ($V_{DS} > V_{GS} - V_{TH}$), the channel is pinched off.



- The drain current is $I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2$ where $L' = L - \Delta L$

$$\frac{1}{L'} = \frac{1}{L - \Delta L} = \frac{1}{L}\cdot\frac{1}{1 - \Delta L/L} \approx \frac{1}{L}\cdot\left(1 + \Delta L/L\right)$$

- Assuming $\Delta L/L = \lambda \cdot V_{DS}$ we get: $\frac{1}{L'} \approx \frac{1}{L}\cdot\left(1 + \Delta L/L\right) = \frac{1}{L}\cdot\left(1 + \lambda \cdot V_{DS}\right)$

- The drain current is $I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2 \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2 \cdot (1 + \lambda \cdot V_{DS})$

- As I_D actually depends on both V_{GS} and V_{DS}, MOS transistors are not ideal current sources (why?).

Channel Length Modulation - 2

- λ represents the relative variation in effective length of the channel for a given increment in V_{DS}.
- For longer channels λ is smaller, i.e., $\lambda \propto 1/L$
- Transconductance: $$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}=Const.}$$

  In Triode:
  $$g_m = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot V_{DS}$$

  In Saturation (ignoring channel length modulation):
  $$g_m = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH}) = \sqrt{2\mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D} = \frac{2 \cdot I_D}{V_{GS} - V_{TH}}$$

  In saturation with channel length modulation:
  $$g_m = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH}) \cdot (1 + \lambda \cdot V_{DS}) = \sqrt{2\mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D \cdot (1 + \lambda \cdot V_{DS})} = \frac{2 \cdot I_D}{V_{GS} - V_{TH}}$$
- The dependence of I_D on V_{DS} is much weaker than its dependence on V_{GS}.

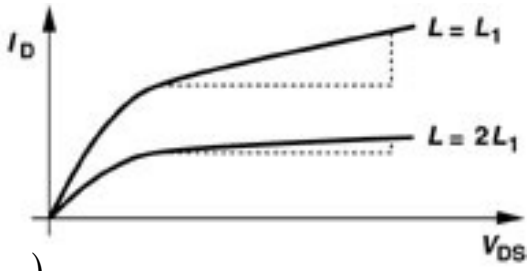
Channel Length Modulation - 3

Example:

Given all other parameters constant, plot I_D-V_{DS} characteristic of an NMOS for $L=L_1$ and $L=2L_1$

- In Triode Region: $I_D \approx \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[\left(V_{GS} - V_{TH} \right) \cdot V_{DS} - \frac{1}{2} \cdot V_{DS}^2 \right]$

  Therefore : $\frac{\partial I_D}{\partial V_{DS}} \propto \frac{W}{L}$



- In Saturation Region: $I_D \approx \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left(V_{GS} - V_{TH} \right)^2 \cdot \left(1 + \lambda \cdot V_{DS} \right)$

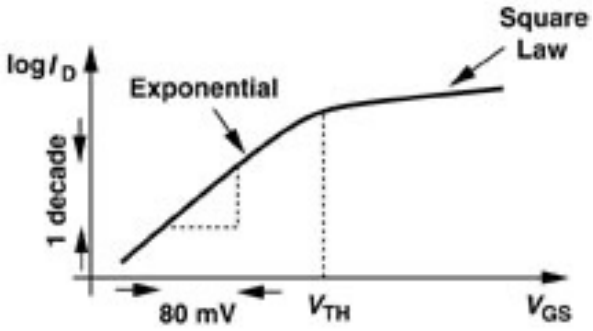
  So we get : $\frac{\partial I_D}{\partial V_{DS}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left(V_{GS} - V_{TH} \right)^2 \cdot \lambda$

  Therefore : $\frac{\partial I_D}{\partial V_{DS}} \propto \frac{W \cdot \lambda}{L} \propto \frac{W}{L^2}$

- Changing the length of the device from L_1 to $2L_1$ will flatten the I_D-V_{DS} curves (slope will be divided by two in triode and by four in saturation).
- Increasing L will make a transistor a better current source, while degrading its current capability.
- Increasing W will improve the current capability.

Sub-threshold Conduction

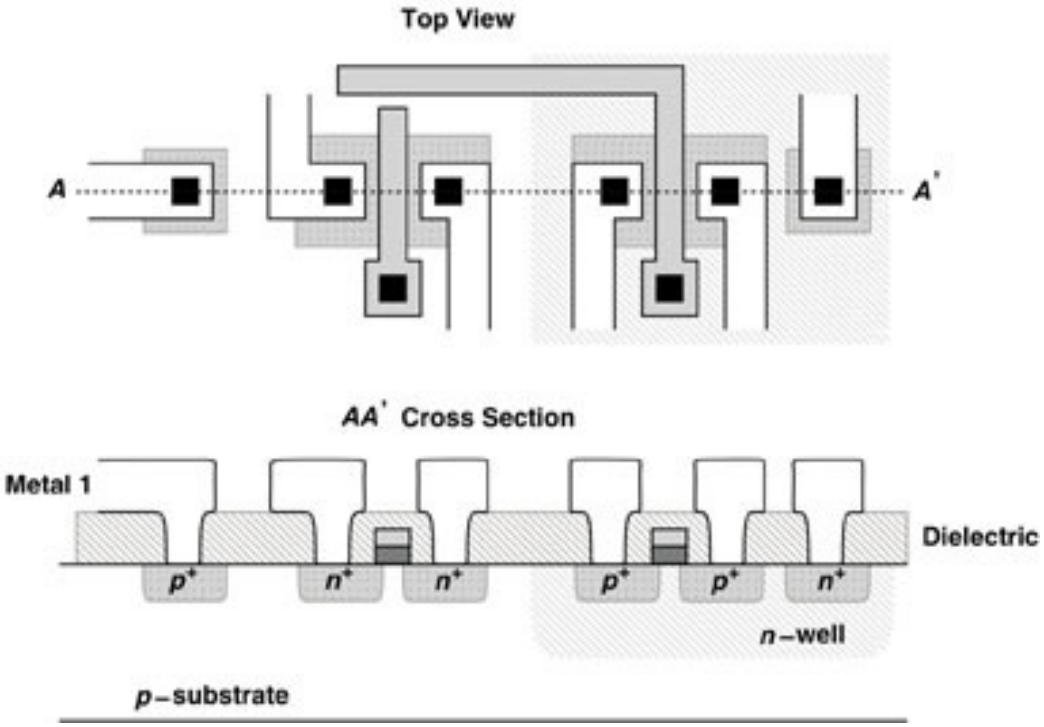
- If $V_{GS} < V_{TH}$, the drain current is not zero.
- The MOS transistors behave similar to BJTs.
- In BJT: $I_C = I_S \cdot e^{\frac{V_{BE}}{V_T}}$



- In MOS: $I_D = I_0 \cdot e^{\frac{V_{GS}}{\zeta \cdot V_T}}$
- In the example shown here, the drain current drops by one decade for approximately each 80mV of drop in V_{GS}.
- This current is known as Sub-threshold or Weak-inversion current.
- In BJT devices the current drops faster (one decade for approximately each 60mV of drop in V_{BE}).

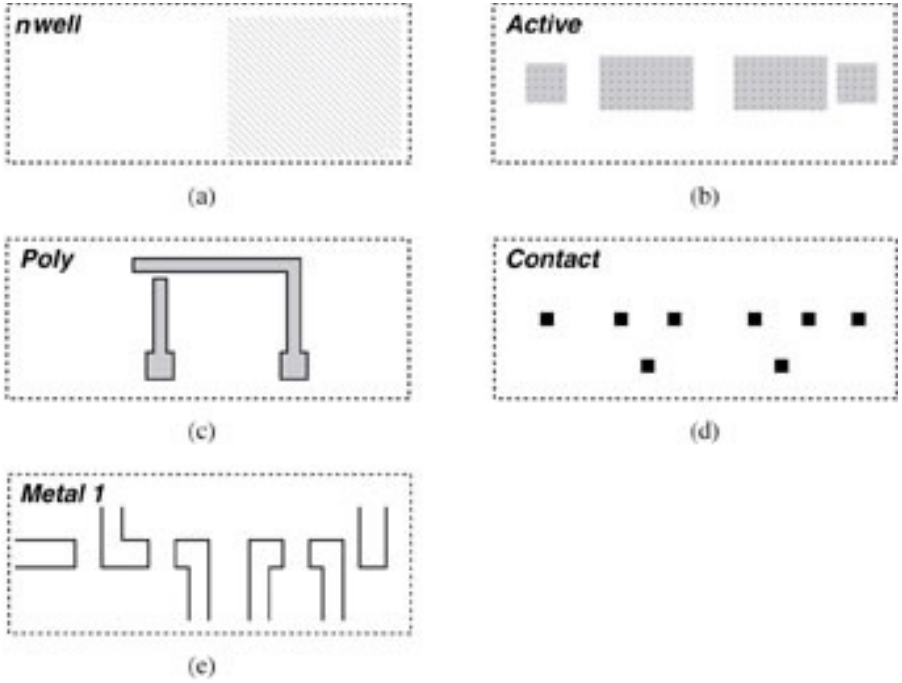
CMOS Processing Technology

- Top and side views of a typical CMOS process



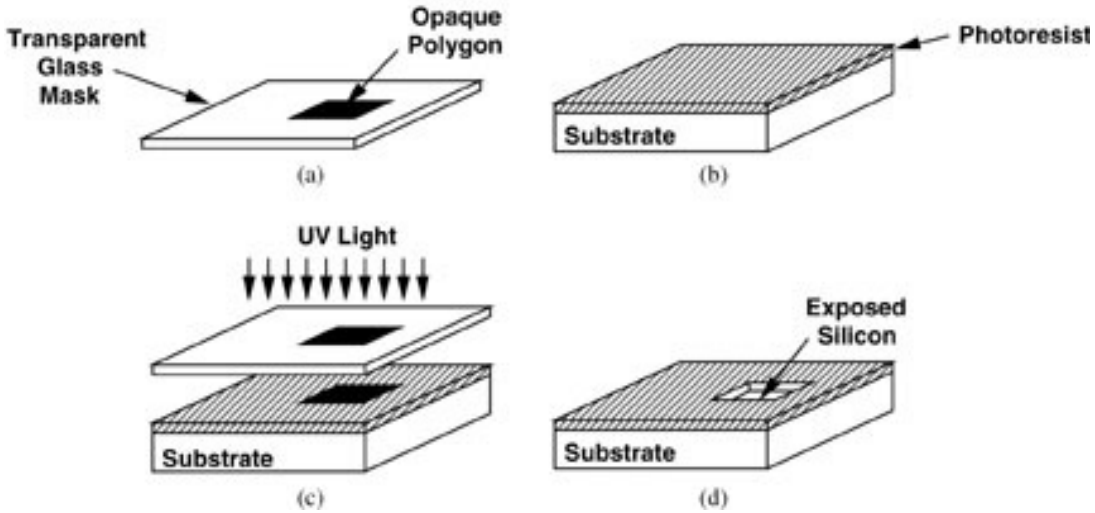
CMOS Processing Technology

- Different layers comprising CMOS transistors

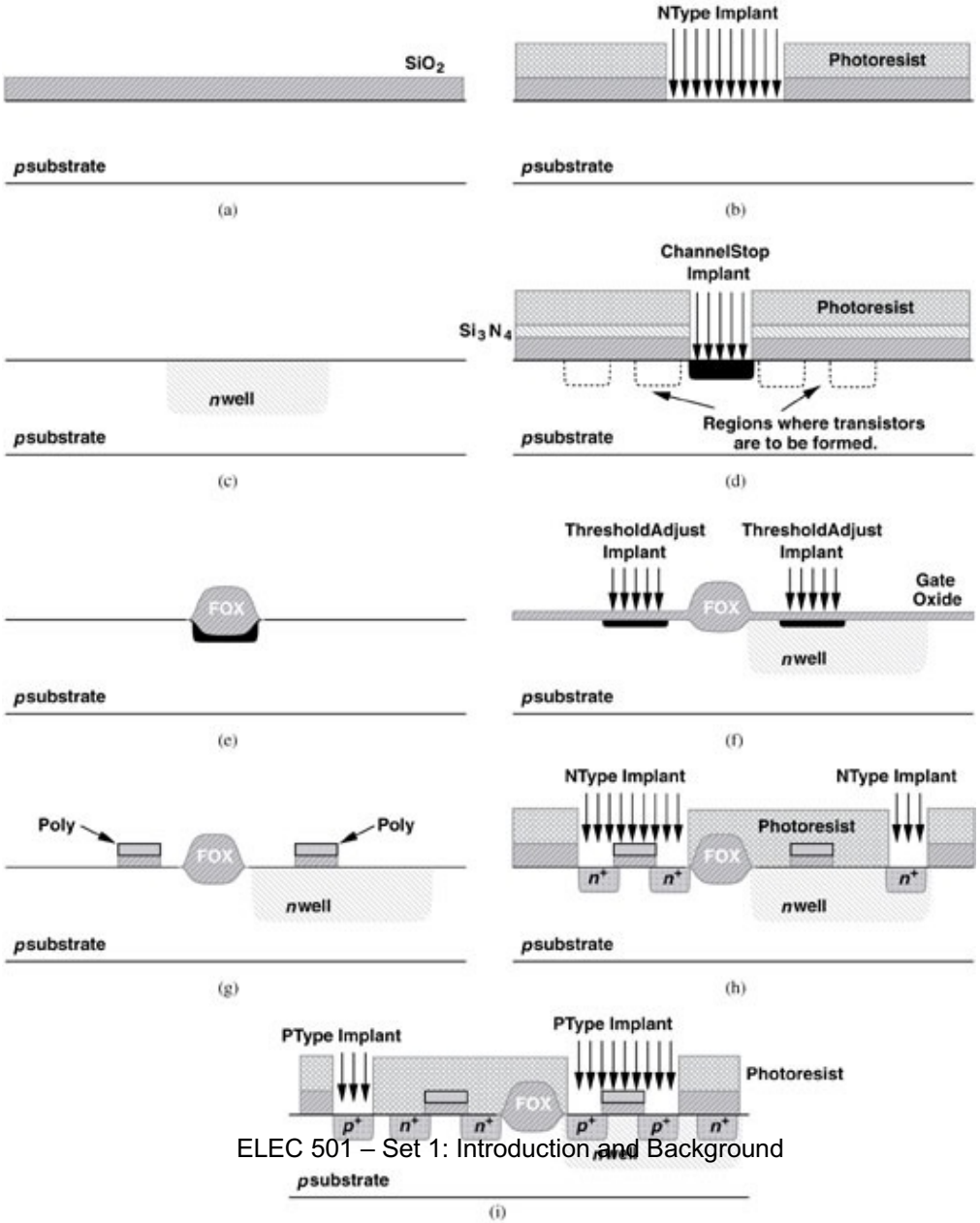


Photolithography (Lithography)

- Used to transfer circuit layout information to the wafer

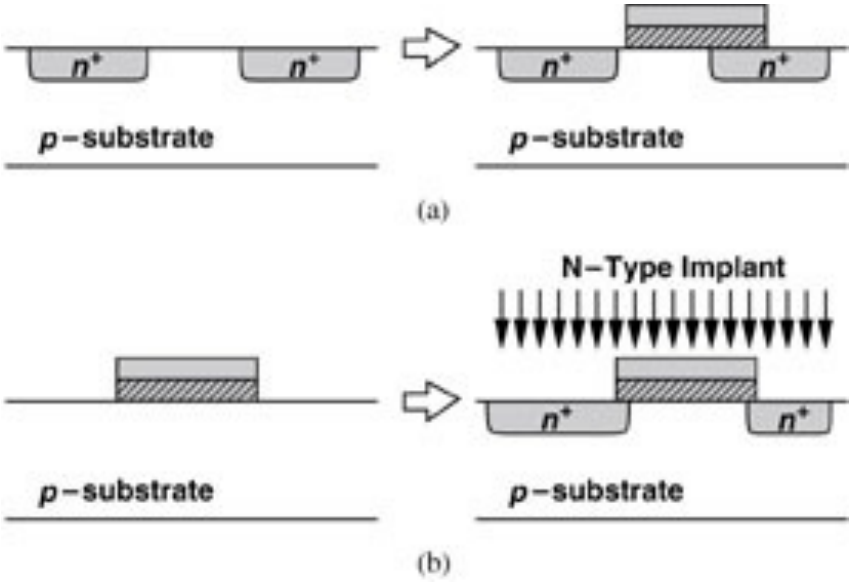


Typical Fabrication Sequence



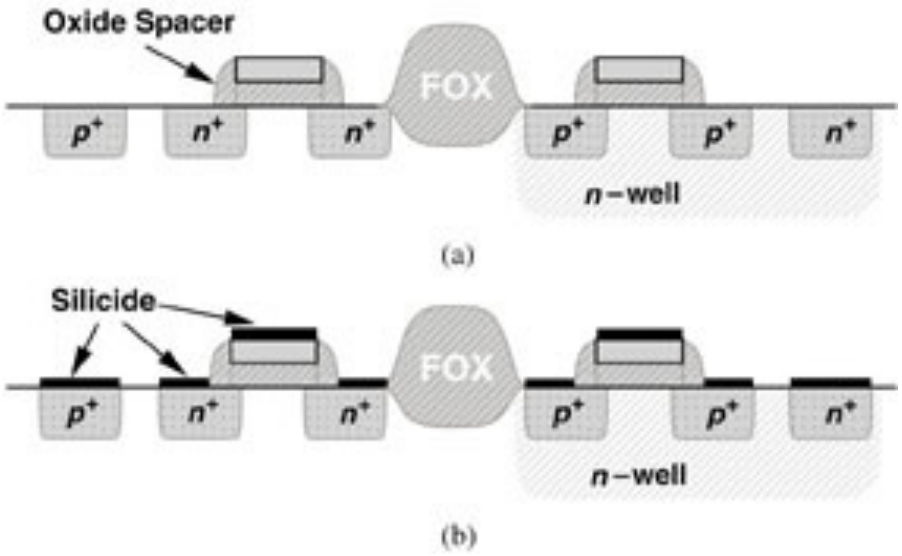
Self-Aligned Process

- Why source and drain junctions are formed after the gate oxide and polysilicon layers are deposited?



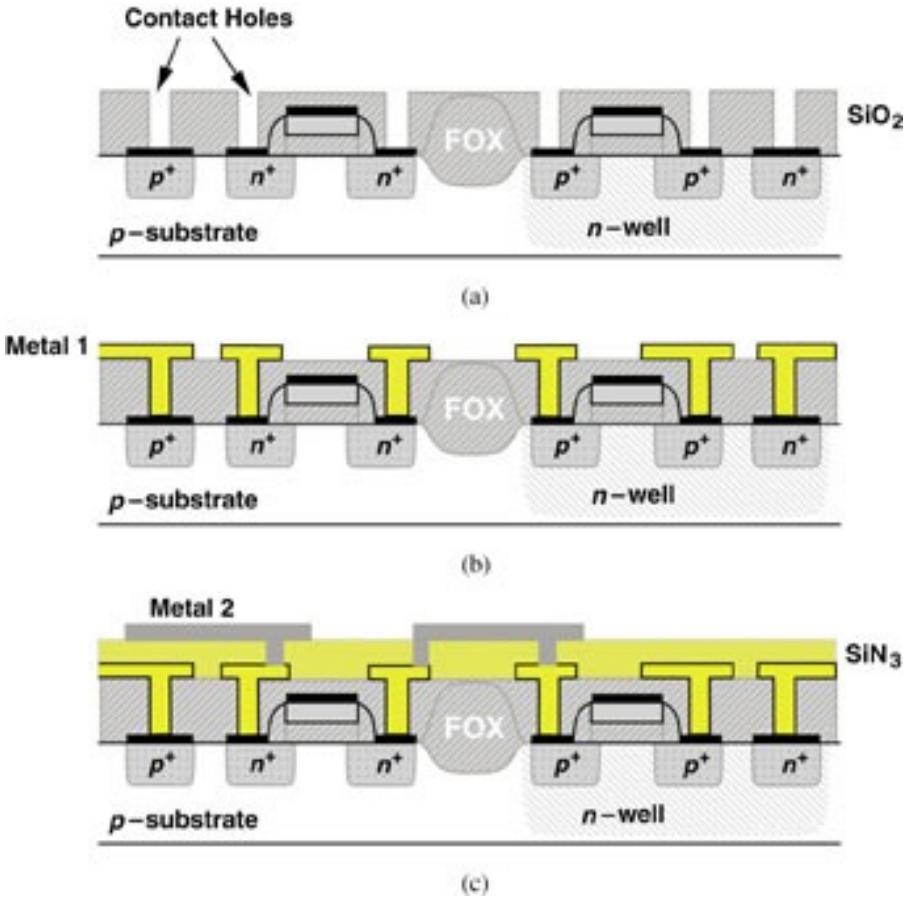
Back-End Processing

- Oxide spacers and silicide



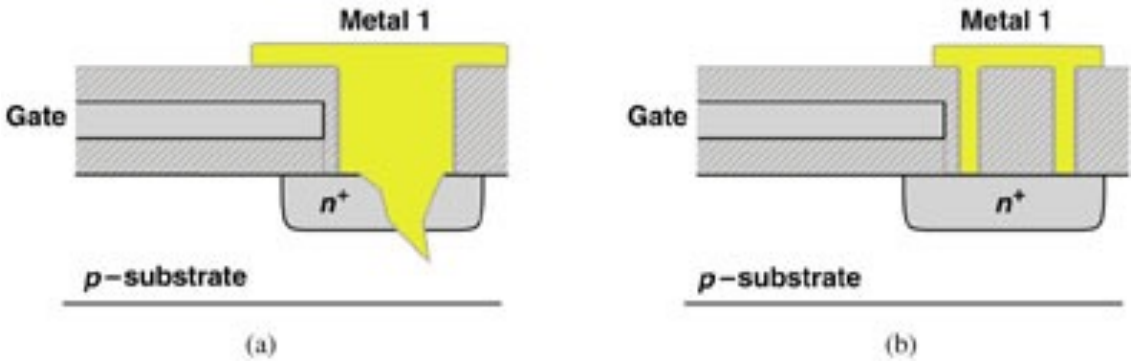
Back-End Processing

- Contact and metal layers fabrication



Back-End Processing

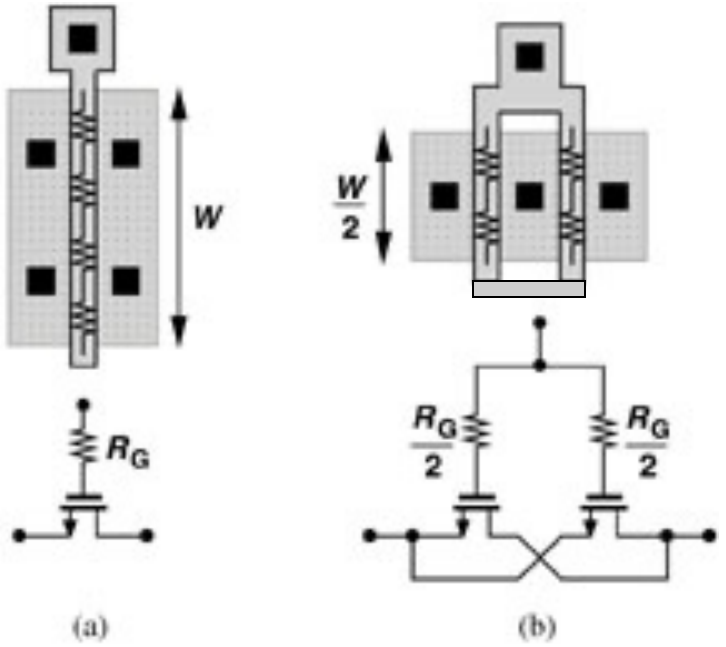
- Large contact areas should be avoided to minimize the possibility of spiking



Importance of Layout

Example (Folded Structure):

Calculate the gate resistance of the circuits shown below.

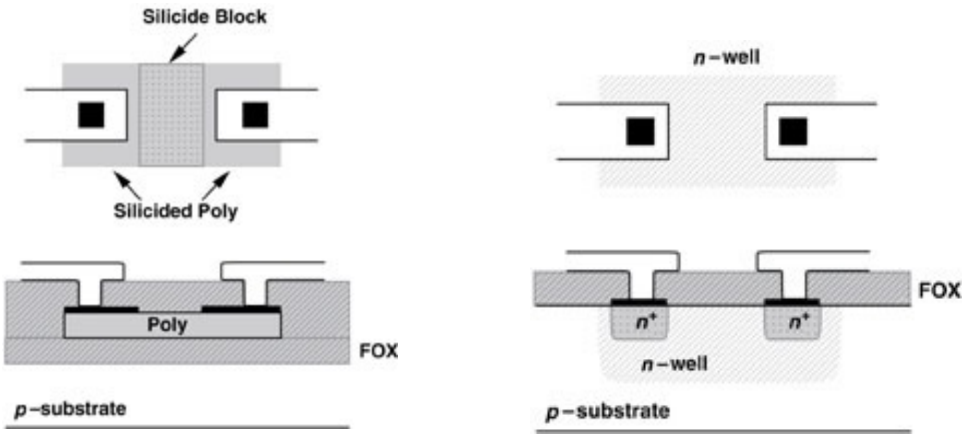


Folded structure:

- Decreases the drain capacitance
- Decreases the gate resistance
- Keeps the aspect ratio the same

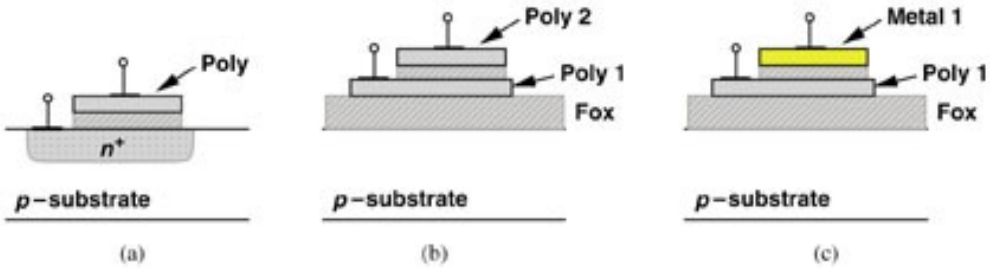
Passive Devices

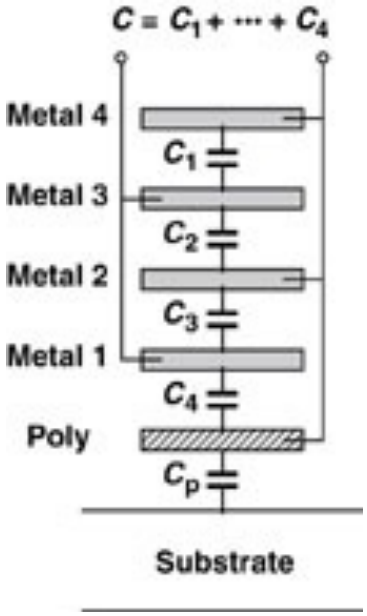
- Resistors



Passive Devices

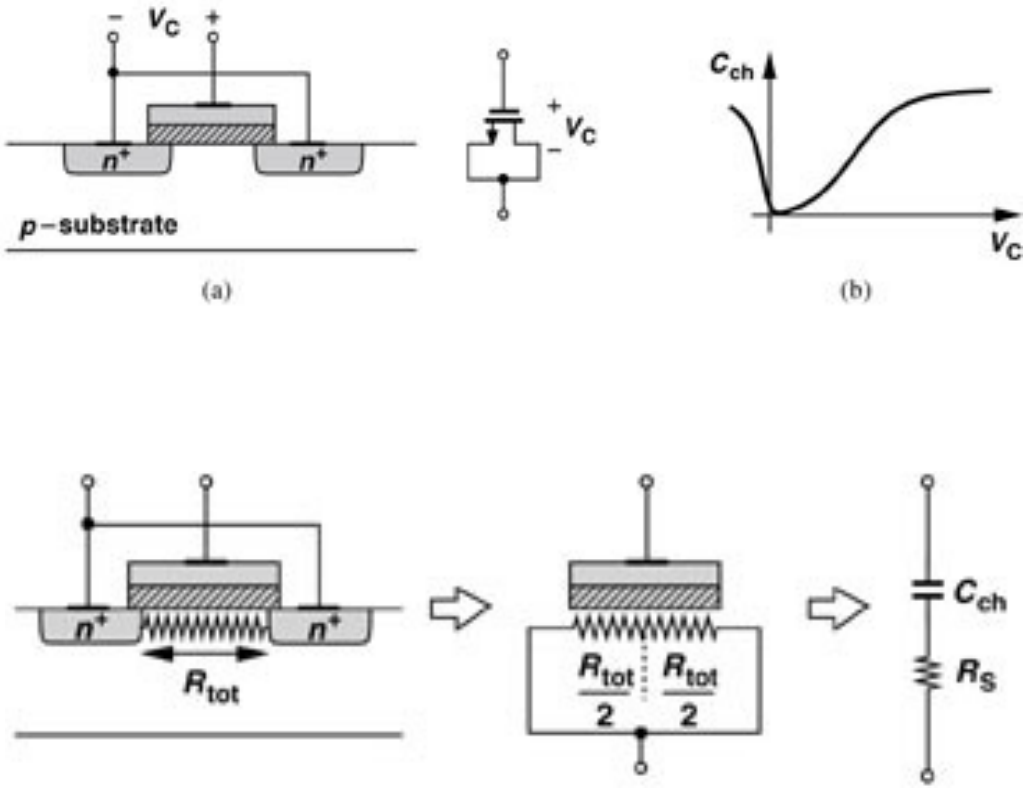
- Capacitors:





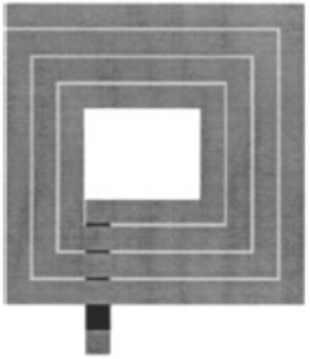
Passive Devices

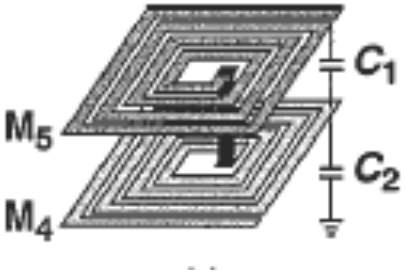
- Capacitors

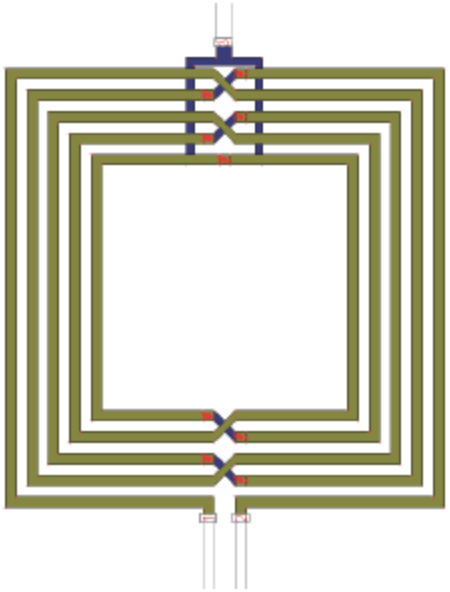


Passive Devices

- Inductors

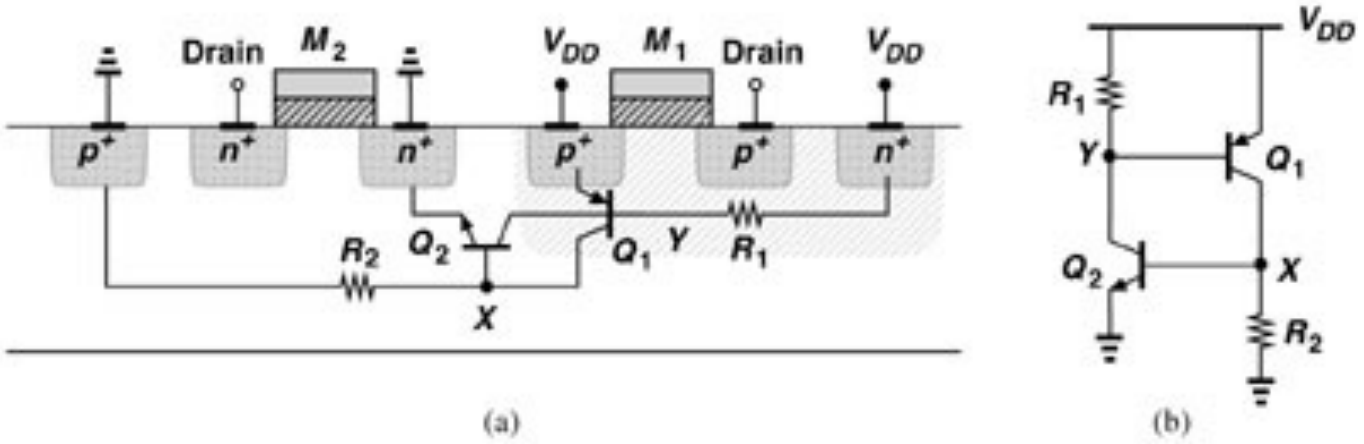






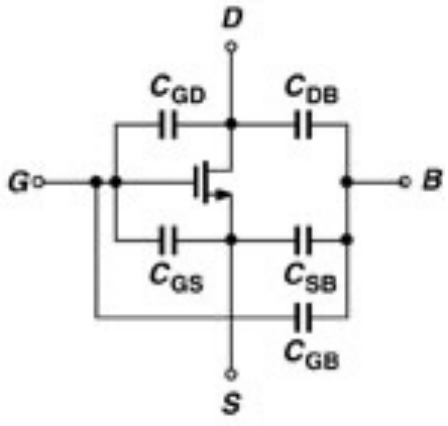
Latch-Up

- Due to parasitic bipolar transistors in a CMOS process



Device Capacitances - 1

- The quadratic model determines the DC behavior of a MOS transistor.
- The capacitances associated with the devices are important when studying the AC behavior of a device.
- There is a capacitance between any two terminals of a MOS transistor. So there are 6 Capacitances in total.
- The Capacitance between Drain and Source is negligible (C_{DS}=0).



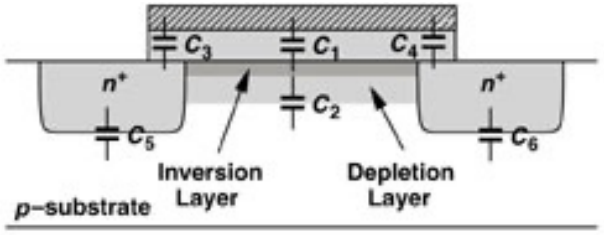
- These capacitances will depend on the region of operation (Bias values).

Device Capacitances - 2

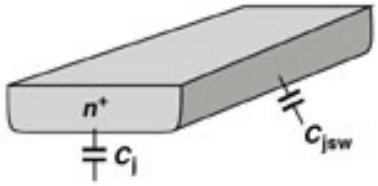
- The following will be used to calculate the capacitances between terminals:
  1. Oxide Capacitance: $C_1 = W \cdot L \cdot C_{ox}$, $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$
  2. Depletion Capacitance: $C_2 = C_{dep} = W \cdot L \cdot \sqrt{\frac{q \cdot \varepsilon_{si} \cdot N_{sub}}{4 \cdot \Phi_F}}$
  3. Overlap Capacitance: $C_3 = C_4 = C_{ov} = W \cdot L_D \cdot C_{ox} + C_{fringe}$
  4. Junction Capacitance:
     - Sidewall Capacitance: C_{jsw}
     - Bottom-plate Capacitance: C_j

$$C_{jun} = \frac{C_{j0}}{\left[1 + \frac{V_R}{\Phi_B}\right]^m}$$

$$C_5 = C_6 = C_j + C_{jsw}$$



(a)



(b)

Device Capacitances - 3

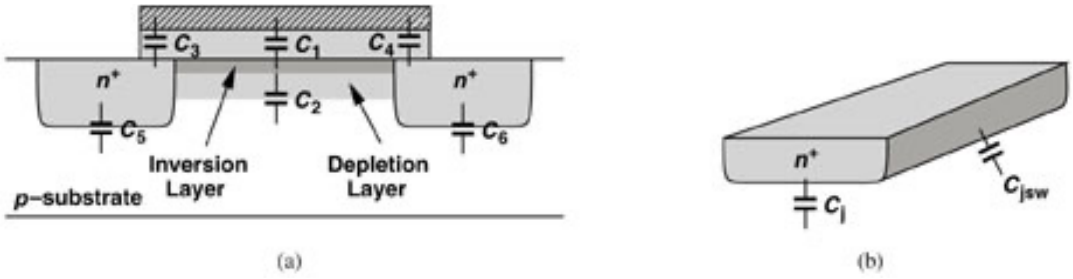
In Cut-off:

1. C_{GS}: is equal to the overlap capacitance. $C_{GS} = C_{ov} = C_3$
2. C_{GD}: is equal to the overlap capacitance. $C_{GD} = C_{ov} = C_4$
3. C_{GB}: is equal to $C_{gate\text{-}channel} = C_1$ in series with $C_{channel\text{-}bulk} = C_2$.

4. C_{SB}: is equal to the junction capacitance between source and bulk.
5. C_{DB}: is equal to the junction capacitance between drain and bulk.

$$C_{SB} = C_5$$

$$C_{DB} = C_6$$

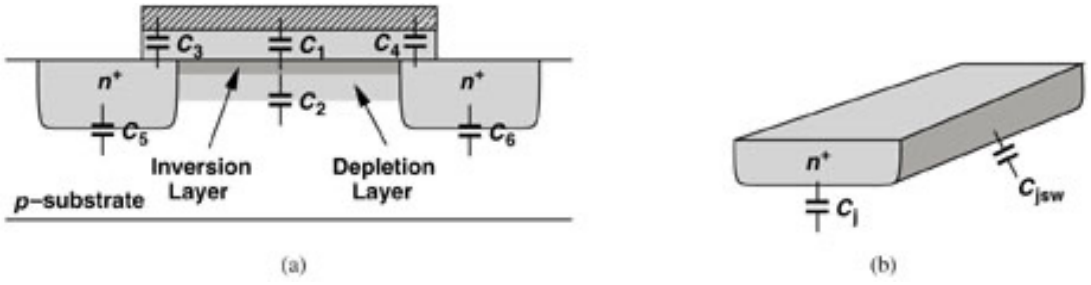


Device Capacitances - 4

In Triode:

- The channel isolates the gate from the substrate. This means that if V_G changes, the charge of the inversion layer are supplied by the drain and source as long as V_{DS} is close to zero. So, C_1 is divided between gate and drain terminals and gate and source terminals.

1. C_{GS}: $C_{GS} = C_{ov} + \frac{C_1}{2}$
2. C_{GD}: $C_{GD} = C_{ov} + \frac{C_1}{2}$
3. C_{GB}: the channel isolates the gate from the substrate. $C_{GB} = 0$
4. C_{SB}: $C_{SB} = C_5 + \frac{C_2}{2}$
5. C_{DB}: $C_{DB} = C_6 + \frac{C_2}{2}$

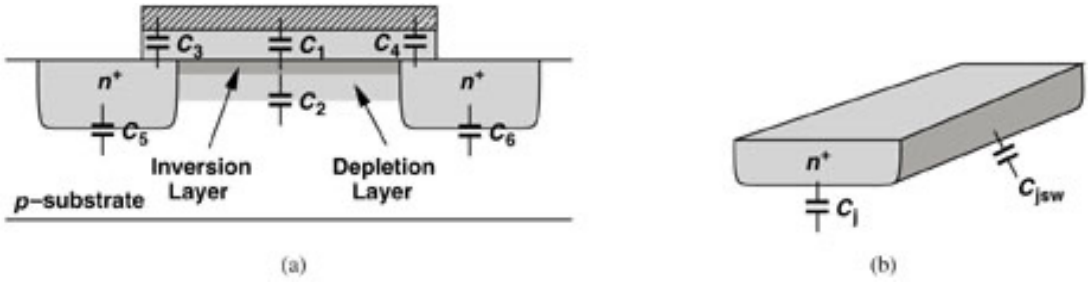


Device Capacitances - 5

In Saturation:

- The channel isolates the gate from the substrate. The voltage across the channel varies which can be accounted for by adding two equivalent capacitances to the source. One is between source and gate, and is equal to two thirds of C_1. The other is between source and bulk, and is equal to two thirds of C_2.

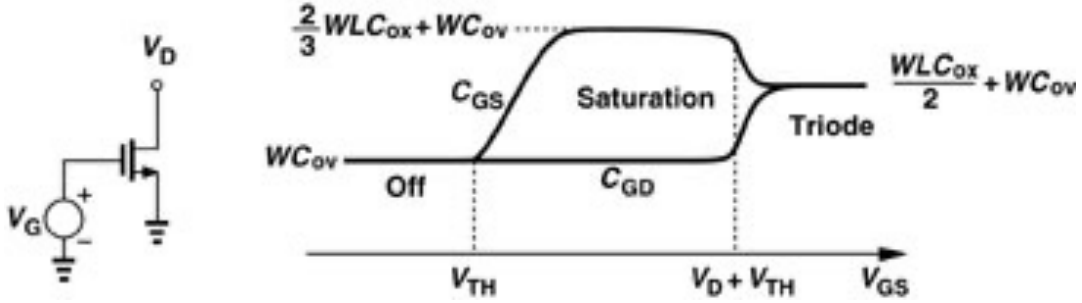
1. C_{GS}: $C_{GS} = C_{ov} + \frac{2}{3}C_1$
2. C_{GD}: $C_{GD} = C_{ov}$
3. C_{GB}: the channel isolates the gate from the substrate. $C_{GB} = 0$
4. C_{SB}: $C_{SB} = C_5 + \frac{2}{3}C_2$
5. C_{DB}: $C_{DB} = C_6$



Device Capacitances - 6

- In summary:

	Cut-off	Triode	Saturation
C_{GS}	C_{ov}	$C_{ov}+\frac{C_1}{2}$	$C_{ov}+\frac{2}{3}C_1$
C_{GD}	C_{ov}	$C_{ov}+\frac{C_1}{2}$	C_{ov}
C_{GB}	$(1/C_1+1/C_2)^{-1}$	0	0
C_{SB}	C_5	$C_5+\frac{C_2}{2}$	$C_5+\frac{2}{3}C_2$
C_{DB}	C_6	$C_6+\frac{C_2}{2}$	C_6

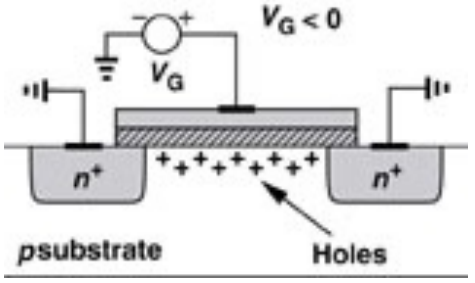


Device Capacitance

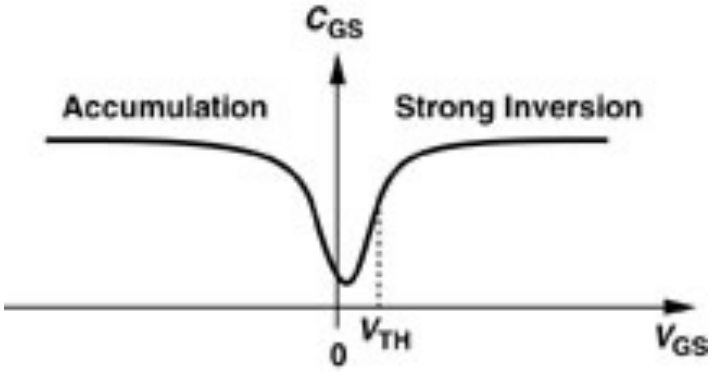
Example:

Draw the gate-source capacitance of the transistor shown below.

The device cannot be in Saturation (V_{DS}=0).



- For negative V_G, holes will be attracted to the surface of the oxide. The device is said to work in accumulation region. As V_G increases, the device enters weak and then strong inversion.



Small Signal Models - 1

- Small signal model is an approximation of the large-signal model around the operation point.
- In analog circuits most MOS transistors are biased in saturation region.
- In general, I_D is a function of V_{GS}, V_{DS}, and V_{BS}. We can use this Taylor series approximation:

$$\text{Taylor Expansion}: I_D = I_{D0} + \frac{\partial I_D}{\partial V_{GS}} \cdot \Delta V_{GS} + \frac{\partial I_D}{\partial V_{DS}} \cdot \Delta V_{DS} + \frac{\partial I_D}{\partial V_{BS}} \cdot \Delta V_{BS} + \text{second order terms}$$

$$\Delta I_D \approx \frac{\partial I_D}{\partial V_{GS}} \cdot \Delta V_{GS} + \frac{\partial I_D}{\partial V_{DS}} \cdot \Delta V_{DS} + \frac{\partial I_D}{\partial V_{BS}} \cdot \Delta V_{BS} = g_m \cdot \Delta V_{GS} + \frac{\Delta V_{DS}}{r_o} + g_{mb} \cdot \Delta V_{BS}$$

Small Signal Models - 2

- Current in Saturation: $I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS}-V_{TH})^2 \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2\cdot(1+\lambda\cdot V_{DS})$
- Taylor approximation: $\Delta I_D \approx \frac{\partial I_D}{\partial V_{GS}}\cdot\Delta V_{GS} + \frac{\partial I_D}{\partial V_{DS}}\cdot\Delta V_{DS} + \frac{\partial I_D}{\partial V_{BS}}\cdot\Delta V_{BS}$
- Partial Derivatives:

$$\frac{\partial I_D}{\partial V_{GS}} = \mu_n\cdot C_{ox}\cdot\frac{W}{L}\cdot(V_{GS}-V_{TH})\cdot(1+\lambda\cdot V_{DS}) = g_m$$

$$\frac{\partial I_D}{\partial V_{DS}} = \frac{1}{2}\cdot\mu_n\cdot C_{ox}\cdot\frac{W}{L}\cdot(V_{GS}-V_{TH})^2\cdot\lambda \approx I_D\cdot\lambda = \frac{1}{r_o}$$

$$\frac{\partial I_D}{\partial V_{BS}} = \frac{\partial I_D}{\partial V_{TH}}\cdot\frac{\partial V_{TH}}{\partial V_{BS}} = \left[-\mu_n\cdot C_{ox}\cdot\frac{W}{L}\cdot(V_{GS}-V_{TH})\cdot(1+\lambda\cdot V_{DS})\right]\cdot\left[-\frac{\gamma}{2\sqrt{|2\cdot\Phi_F+V_{SB}|}}\right]$$

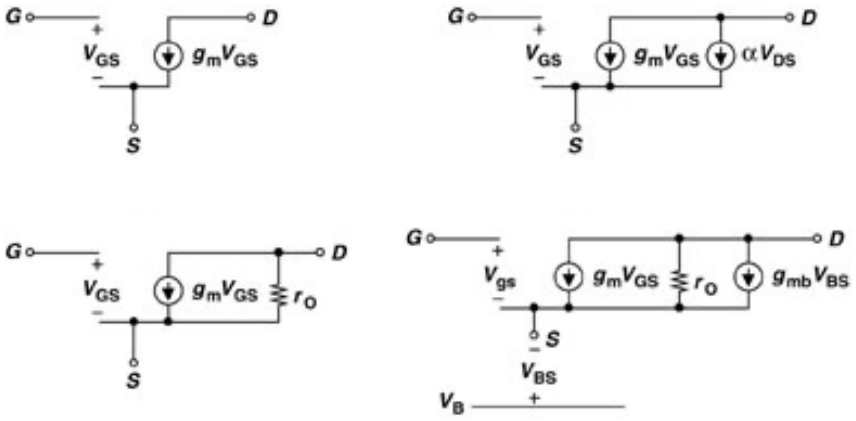
$$= -g_m\cdot\left[-\frac{\gamma}{2\sqrt{|2\cdot\Phi_F+V_{SB}|}}\right] = g_m\cdot\eta = g_{mb}$$

Small Signal Models - 3

- Small-Signal Model:

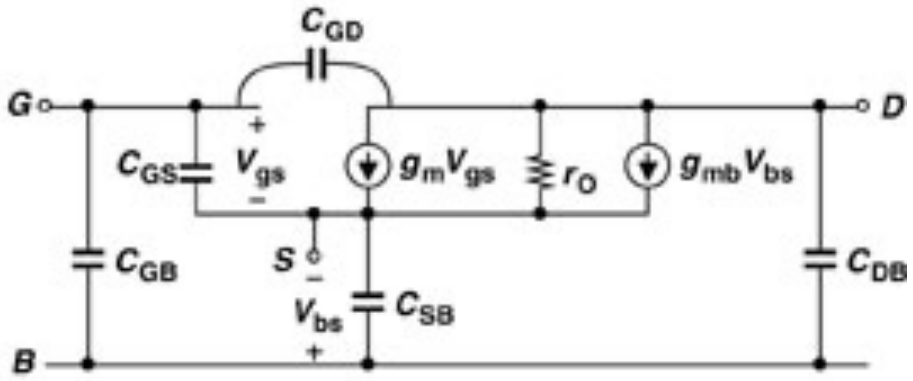
$$i_D = g_m \cdot v_{GS} + \frac{v_{DS}}{r_o} + g_{mb} \cdot v_{BS}$$

- Terms, $g_m v_{GS}$ and $g_{mb} v_{BS}$, can be modeled by dependent sources. These terms have the same polarity: increasing v_G, has the same effect as increasing v_B.
- The term, v_{DS}/r_o can be modeled using a resistor as shown below.



Small Signal Models - 4

- Complete Small-Signal Model with Capacitances:



- Small signal model including all the capacitance makes the intuitive (qualitative) analysis of even a few-transistor circuit difficult!

- Typically, CAD tools are used for accurate circuit analysis

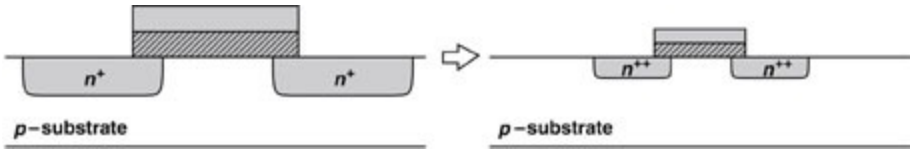
- For intuitive analysis we try to find a simplest model that can represent the role of each transistor with reasonable accuracy.

Short-Channel Effects

- Threshold Reduction
  - Drain-induced barrier lowering (DIBL)
- Mobility degradation
- Velocity saturation
- Hot carrier effects
  - Substrate current
  - Gate current
- Output impedance variation

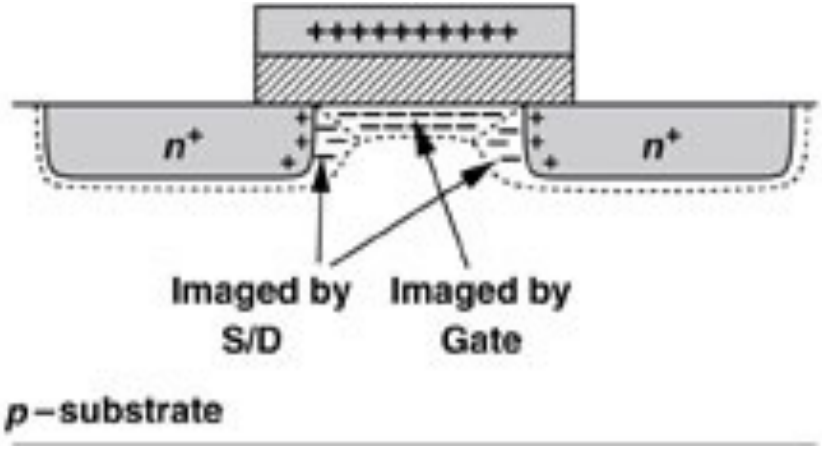
Technology Scaling

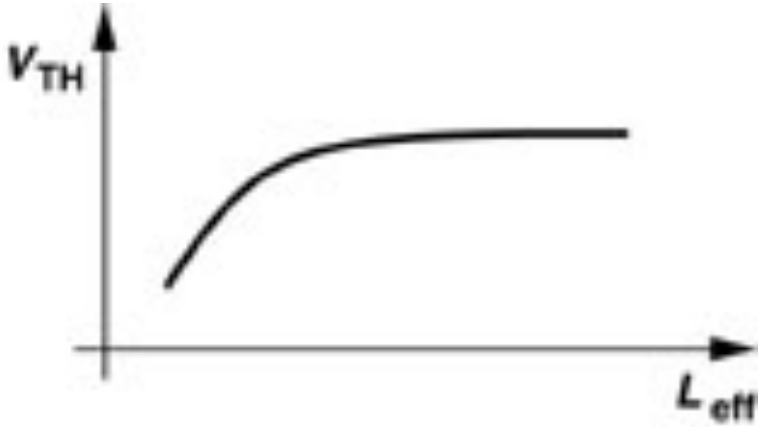
- Motivations for technology scaling:
  - Area (cost)
  - Speed
  - Power consumption in digital circuits $\propto fCV_{DD}^2$



Threshold Voltage Variation in Short Channel Devices

- The Threshold of transistors fabricated on the same chip decreases as the channel length decreases.

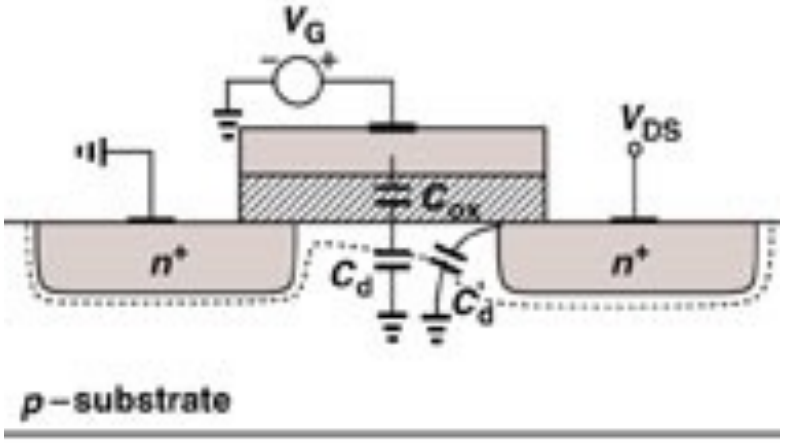




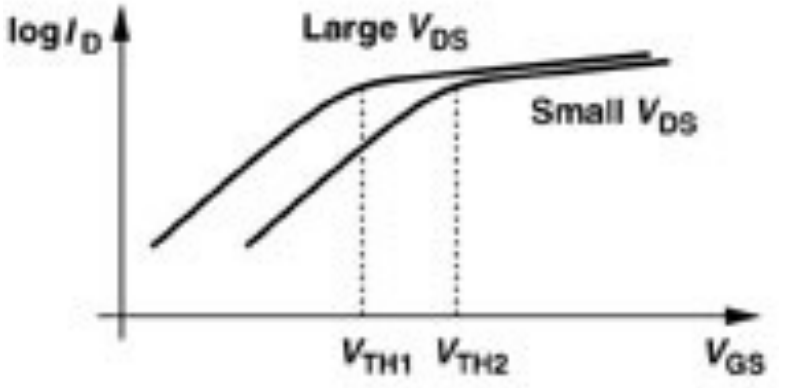
- Intuitively, the extent of depletion regions associated with drain and source in the channel area, reduces the immobile charge that must be imaged by the charge on the gate.

Drain-Induced Barrier Lowering (DIBL)

When the channel is short, the drain voltage increases the channel surface potential, lowering the barrier to flow charge from source (think of increased electric field) and therefore, decreasing the threshold.



(a)



(b)

Mobility Degradation with Vertical Field

- High electric field between gate and channel confines charge carriers to a narrower region in the channel.
- This leads to more carrier scattering and therefore lower mobility.
- Empirical equation modeling this effect:

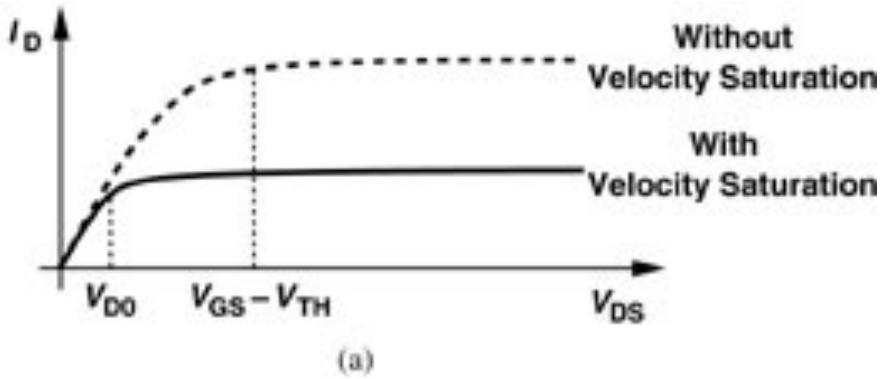
$$\mu_{eff} = \frac{\mu_0}{1 + \theta(V_{GS} - V_{TH})}$$

where θ is a fitting parameter ($\approx 10^{-7}/t_{ox}$)

- Mobility degradation lowers current capability and g_m as well as causing I-V characteristic of the MOS to deviate from square-law.

Effects of Velocity Saturation

- Due to drop in mobility at high electric fields



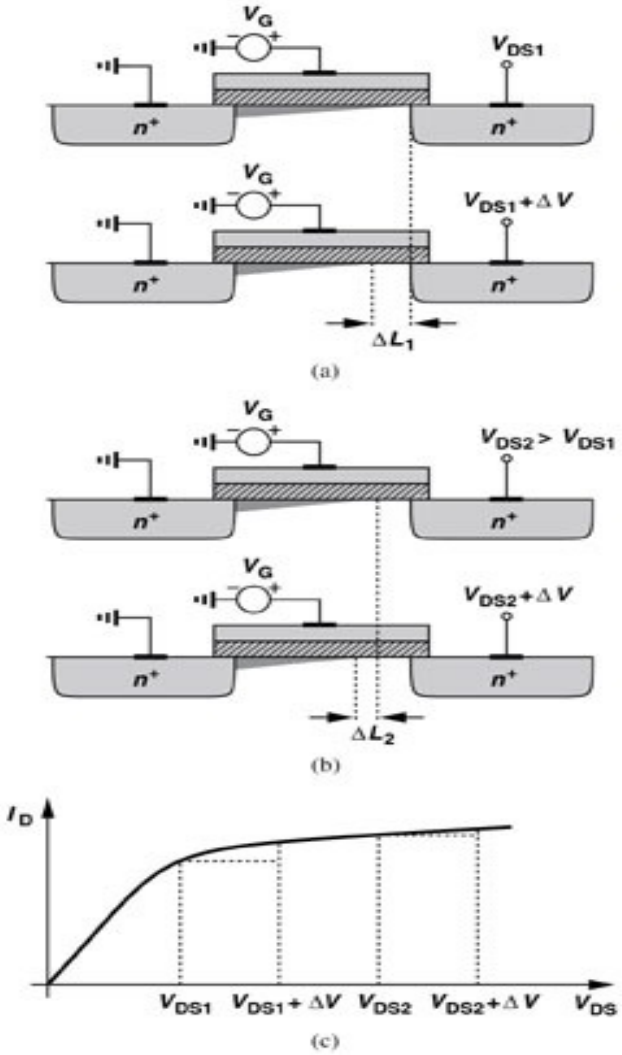
- (a) Premature drain current saturation and (b) reduction in g_m

Hot Carrier Effects

- Short channel devices may experience high lateral drain-source electric field
- Some carriers that make it to drain have high velocity (called "hot" carriers)
- "Hot" carriers may "hit" silicon atoms at high speed and cause impact ionization
- The resulting electron and holes are absorbed by the drain and substrate causing extra drain-substrate current
- Really "hot" carriers may be injected into gate oxide and flow out of gate causing gate current!

Output Impedance Variation

Recall the definition of λ.



Output Impedance Variation in Short-Channel Devices