BIRZEIT UNIVERSITY

DEPARTMENT OF COMPUTER SYSTEM ENGINEERING

Digital Integrated Circuits - ENCS333

Dr. Khader Mohammad

Lecture #4 -

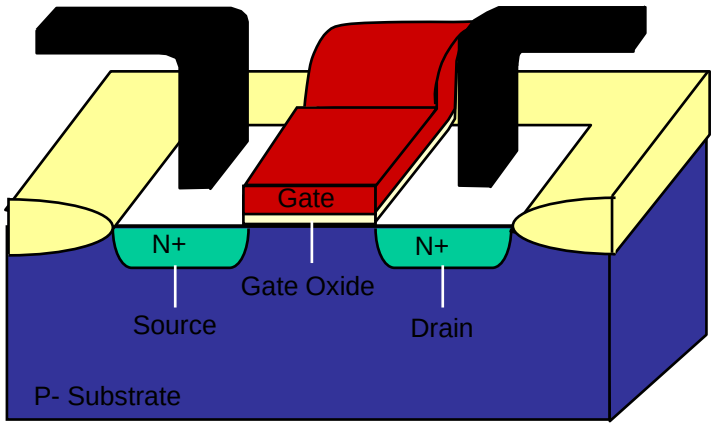
Semiconductor material: PN-junction, NMOS, PMOS

Transistor Theory

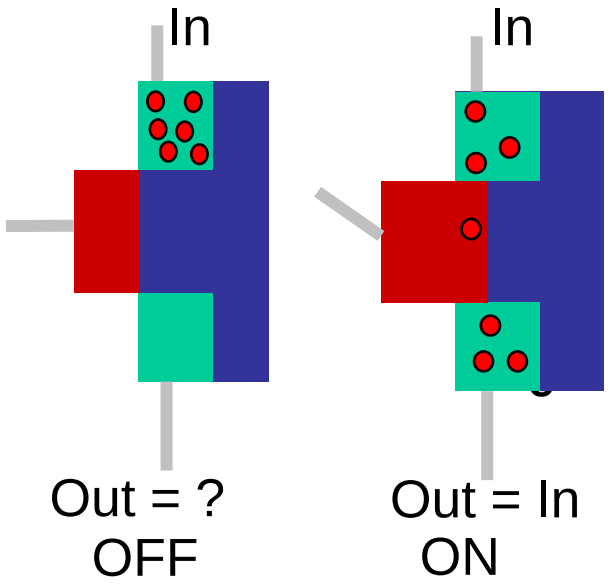
Integrated-Circuit Devices and Modeling

Basic Element

- CMOS Transistor is a switch



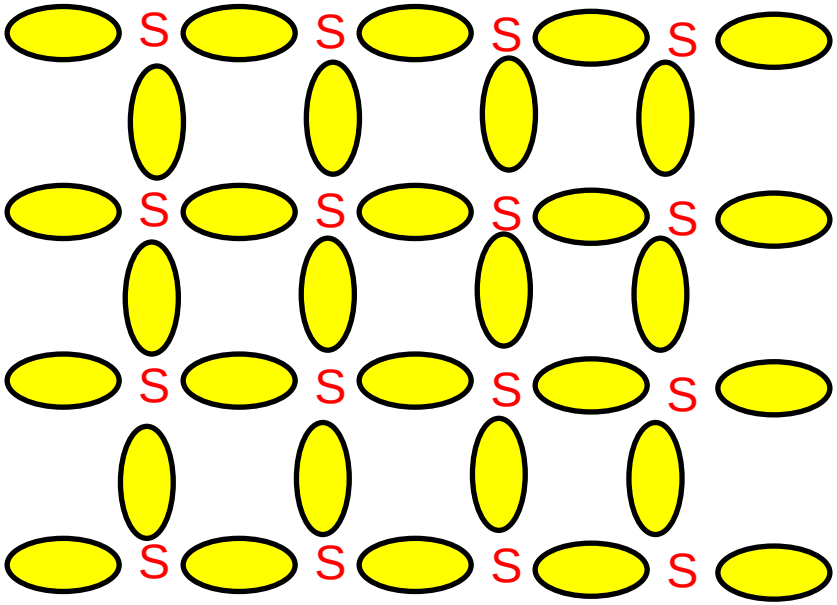
Cross section



Symbol view

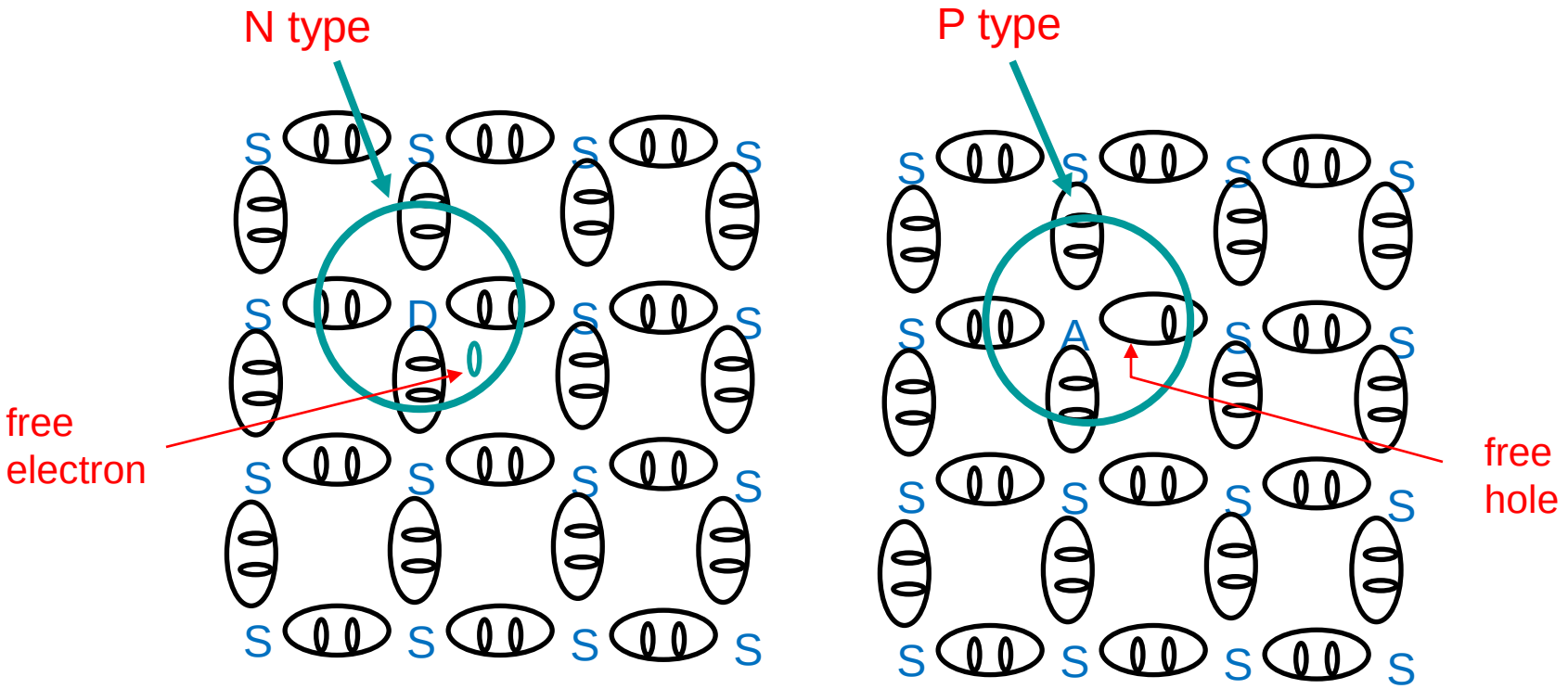
The Semiconductor

- Silicon crystals is composed by atoms with 4 valence electrons
  - Diamond lattice
  - No free charge
  - Very high resistance
  - Thermal excitation: energy gap ~ 1.1eV, KT ~ 25 meV

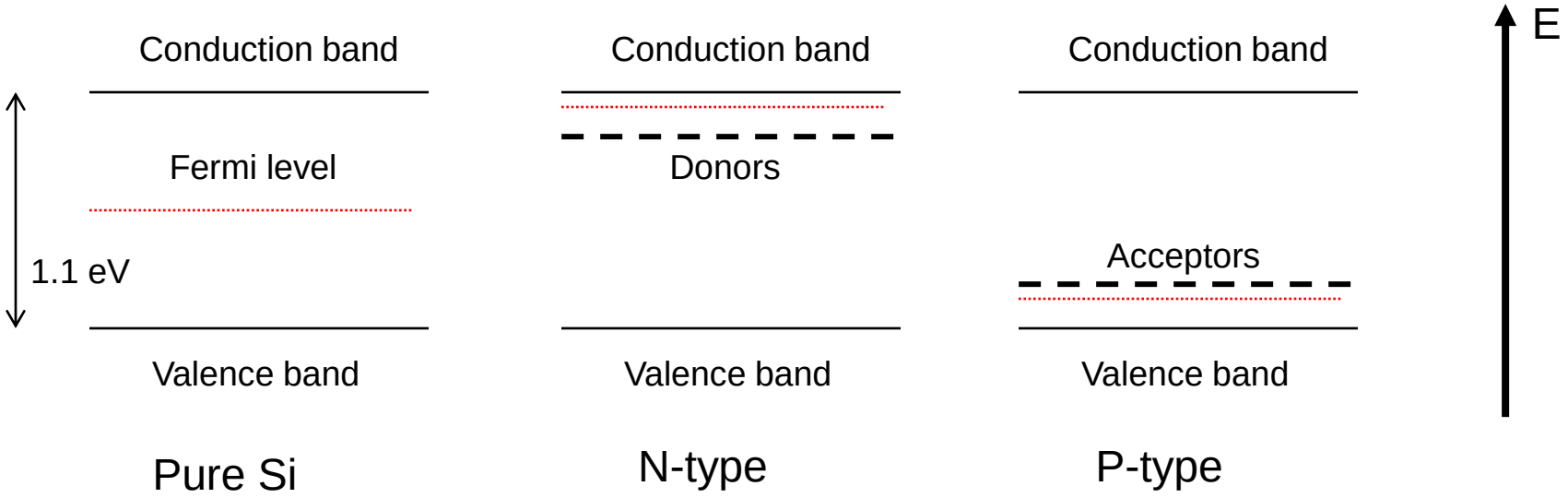


Doping the Semiconductor

- Creates free electrical charges by adding impurities
  - Donors: 5 valence electrons creates free electrons
    - Phosphorus (P), Arsenic (As)
  - Acceptors: 3 valence electrons creates free “holes”
    - Boron (B), Gallium (Ga)
  - Ionization energy ~ KT: all donors/acceptors become free carriers

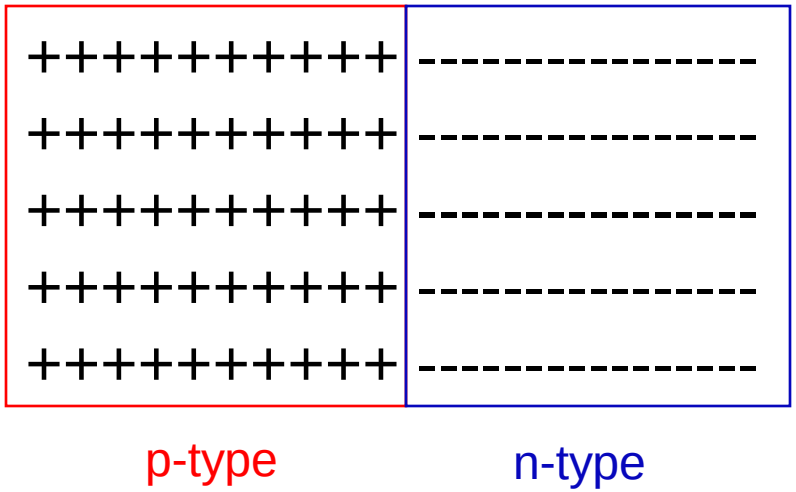


Energy Gap and Doping



Metals: half filled band
Insulators: very large energy gap, valence band is full
Semiconductors: moderate energy gap, valence band is full

The Semiconductor Diode

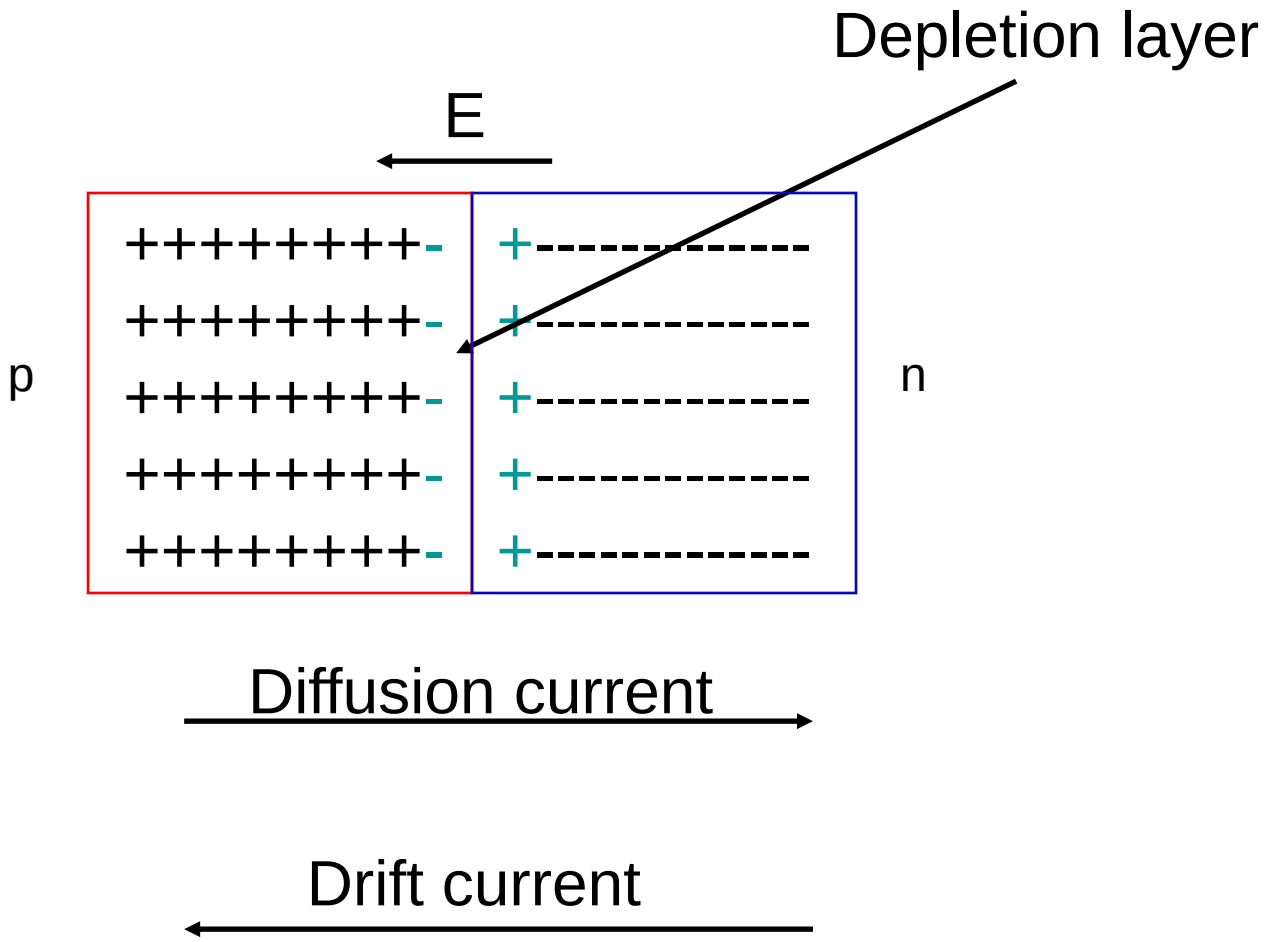


Two materials in contact ==>

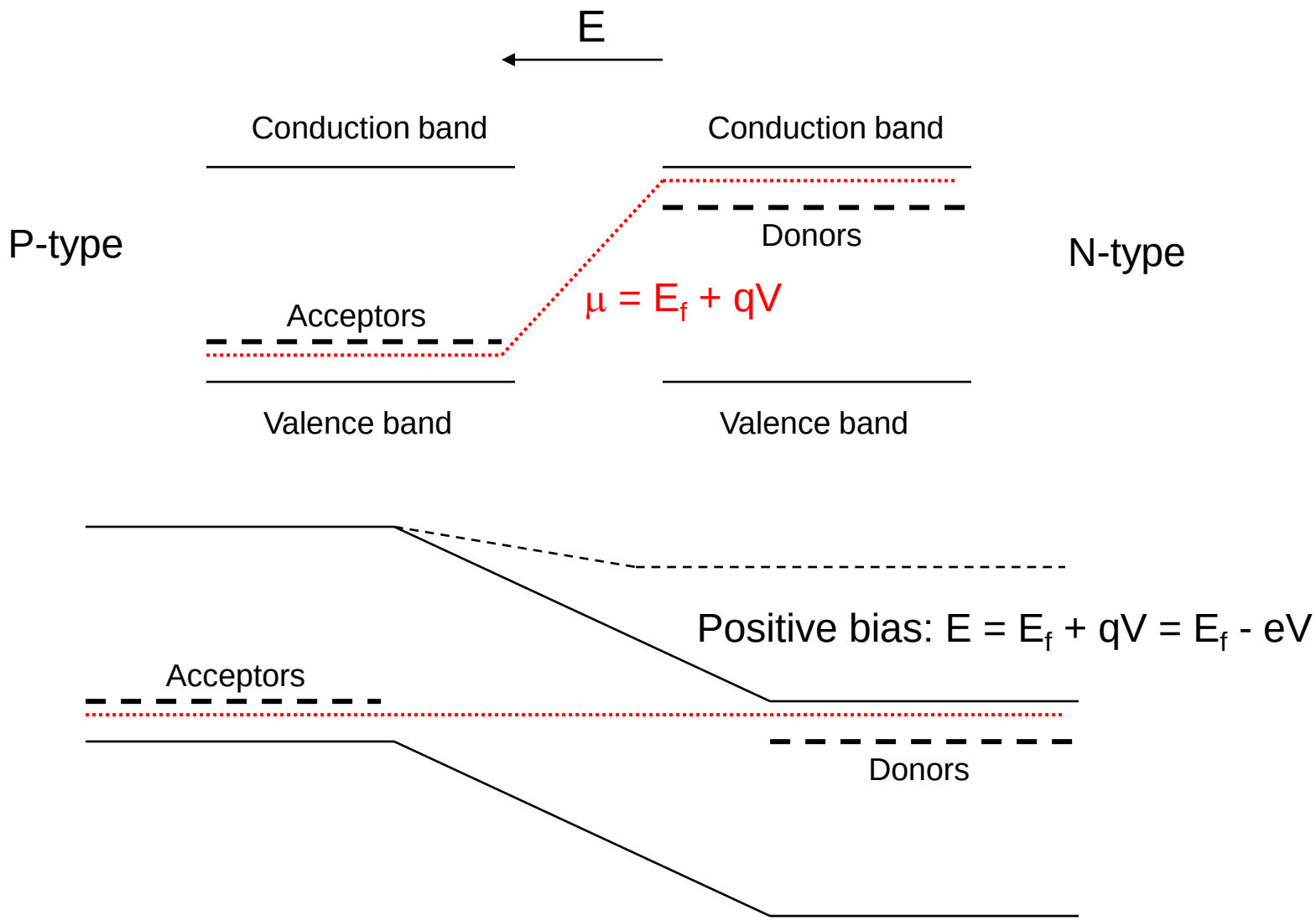
particles are free to move from one material to the other one

Equilibrium: equal electro-chemical potential (μ)

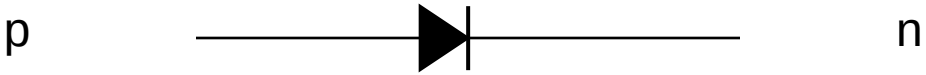
The Semiconductor Diode



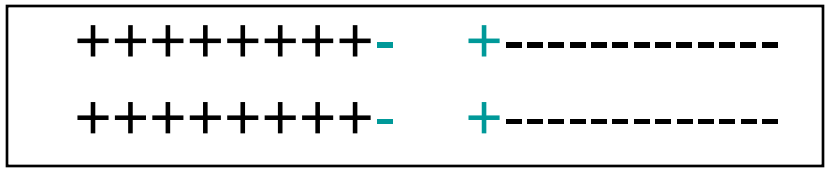
PN junction and energy bands



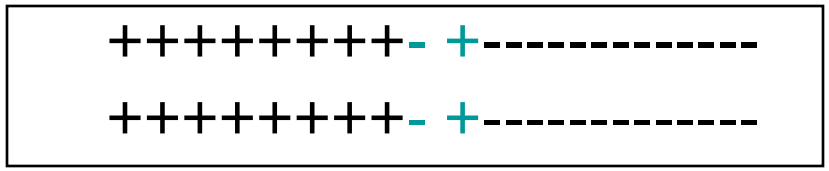
Diode Capacitance



C_{diode}



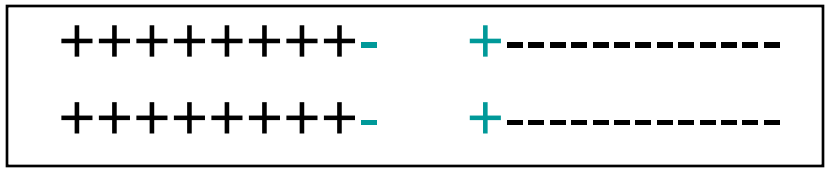
Forward bias



+ -

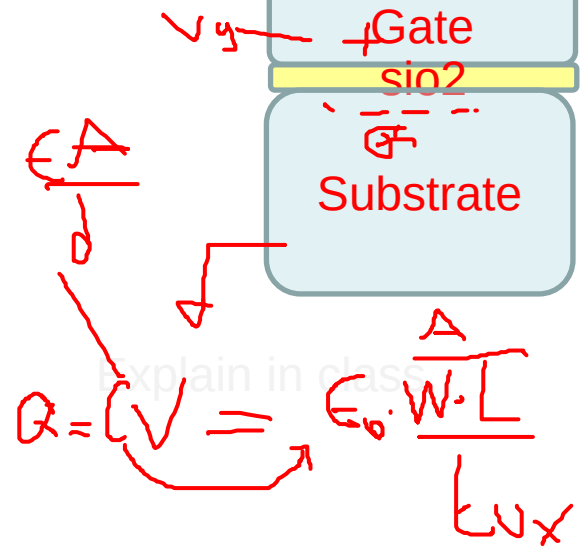
C increases

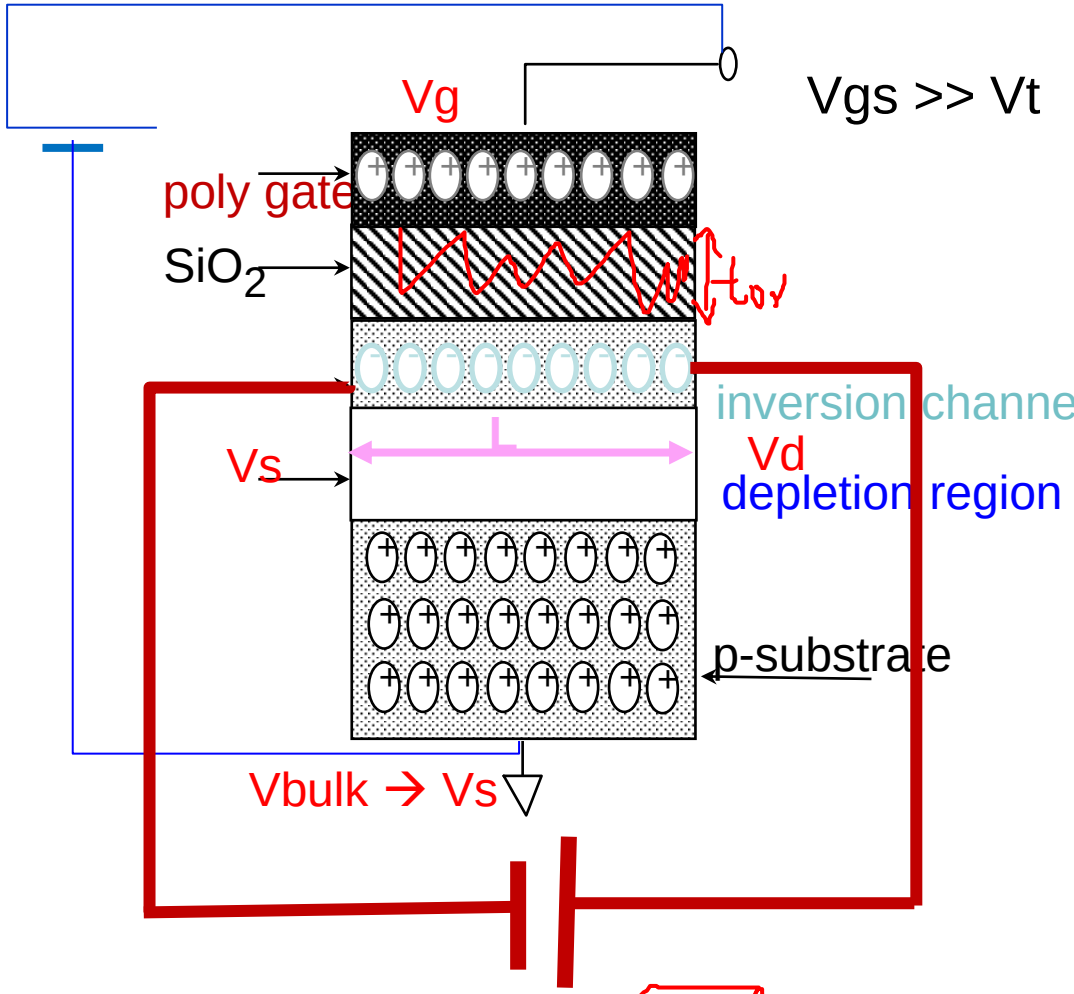
Reverse bias

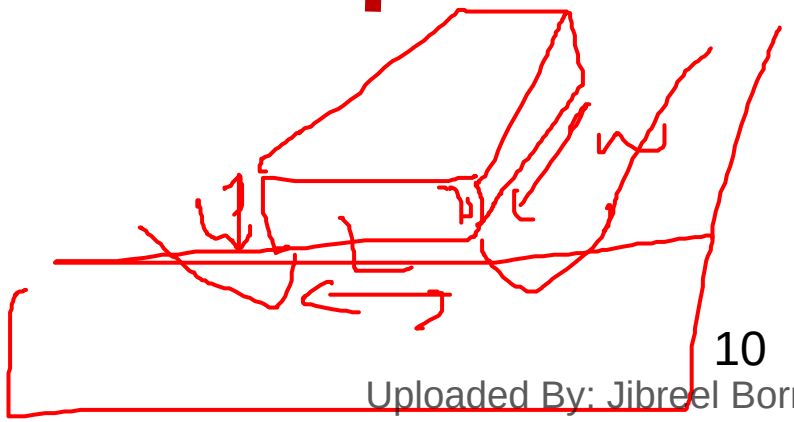


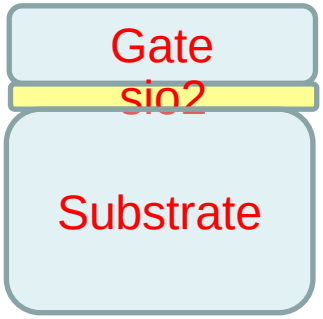
- +

C decreases









Q (channel) = C V

V INC ➔ Q INCREASE

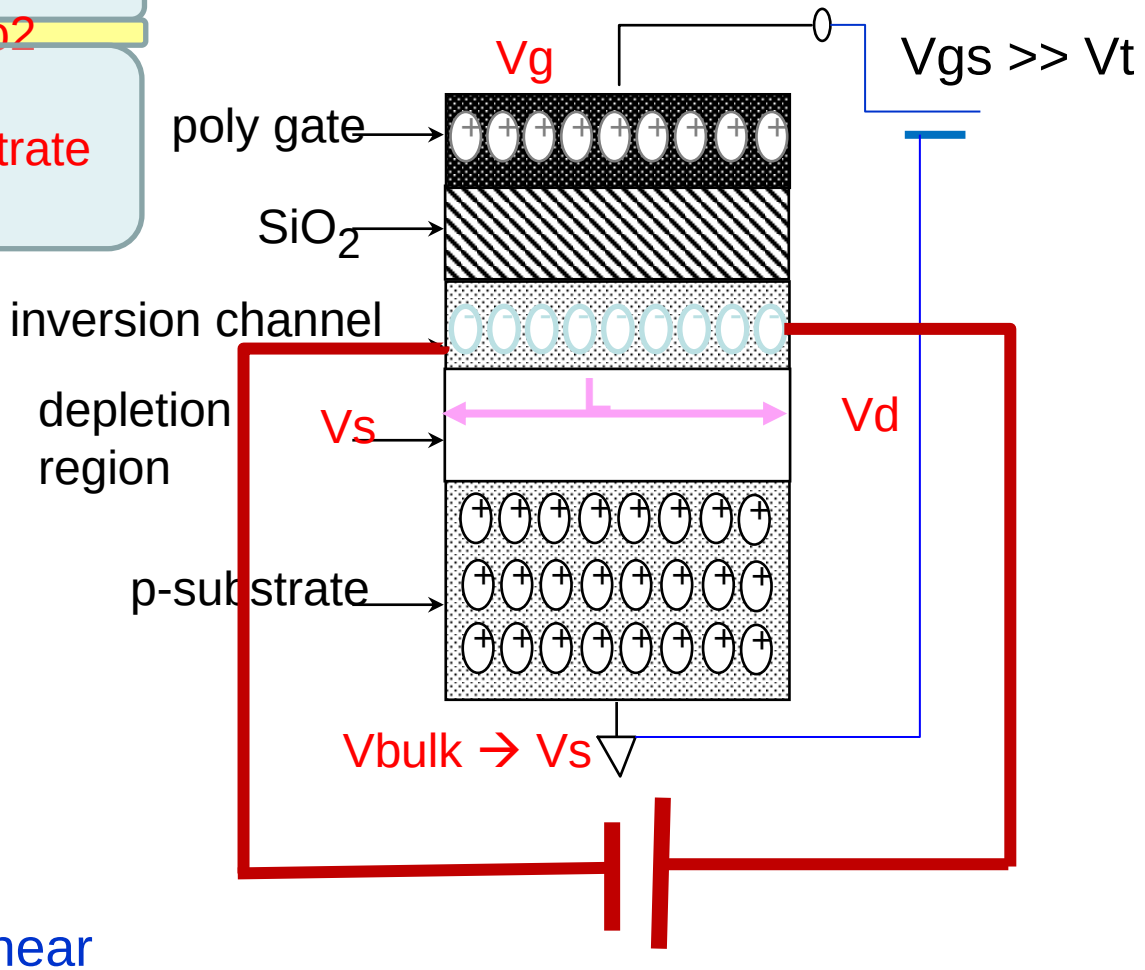
C= W.L. Cox (cap per unit length)

V called over drive voltage =Vgs – Vt

So

Q= W.L. Cox * (Vgs – Vt)

Cox=£eps/tox



i=dQ/dt

Assuming VGS>>Vt, Vds >0 linear

t=L(distance)/v(speed)

=L/(μ Mobility of electron * E electric field

= L/ μ * (Vds/L) = L^2/ (μ *Vds)

i= μ * (1/L) *(£ox/tox) * (w/L) (Vgs-vt) *Vds

In sat , Vds>Vgs-vt

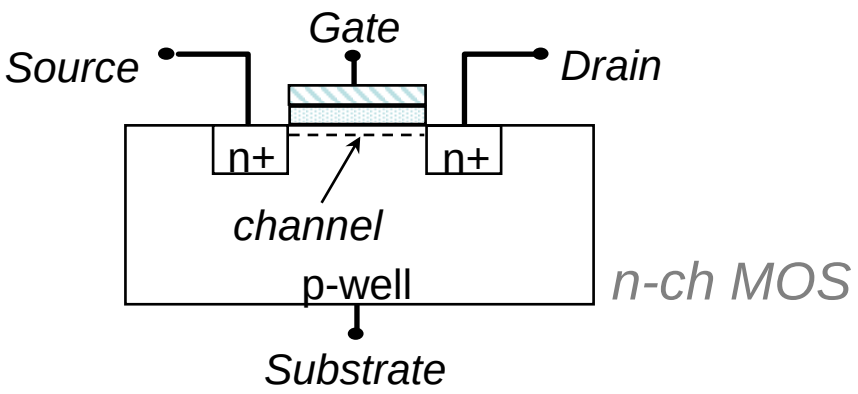
V=Vgs-vt- (1/2 Vds)

Q=CV = (£ox/tox) * (w/L) (Vgs-vt –Vds/2)

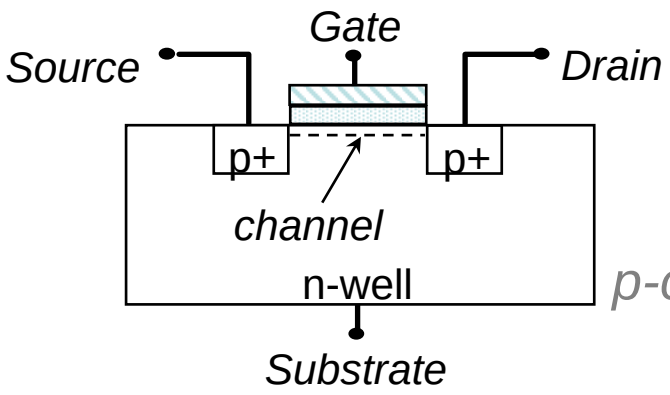
i= μ/2 * (1/L) *(£ox/tox) * (w/L) (Vgs-vt)^2

The MOS transistor

- The MOS transistor is a four terminal device (Gate, Source, Drain and Substrate).
- The n+ (p+) diffusions provide the Source / Drain terminals of n-ch (p-ch) MOS transistors.
- Ids (drain-source current) flows through a region that is controlled by the MOS gate - “the channel”.
- Ids is modulated by the applied voltage at the gate (trans-conductance).
- Note that the source and drain are circuit definitions, not device definitions (the process forms symmetric devices).



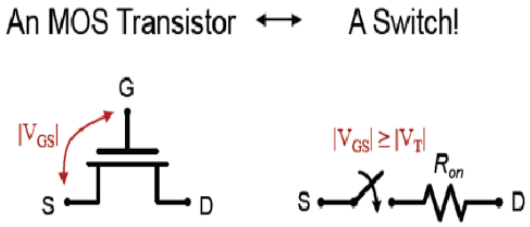
n-ch MOS

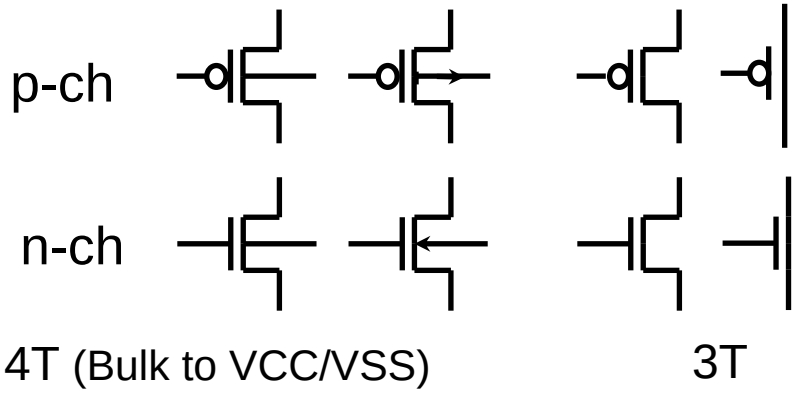


p-ch MOS

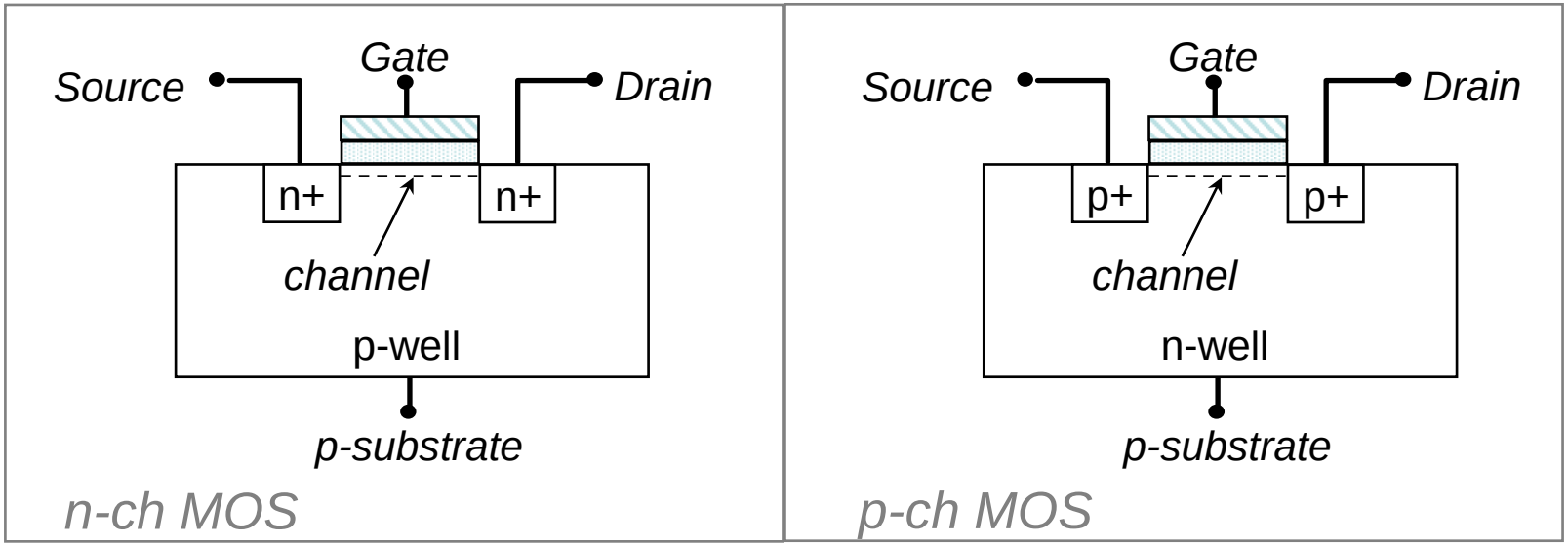
The MOS transistor

- Symbols:





- Physical structure:



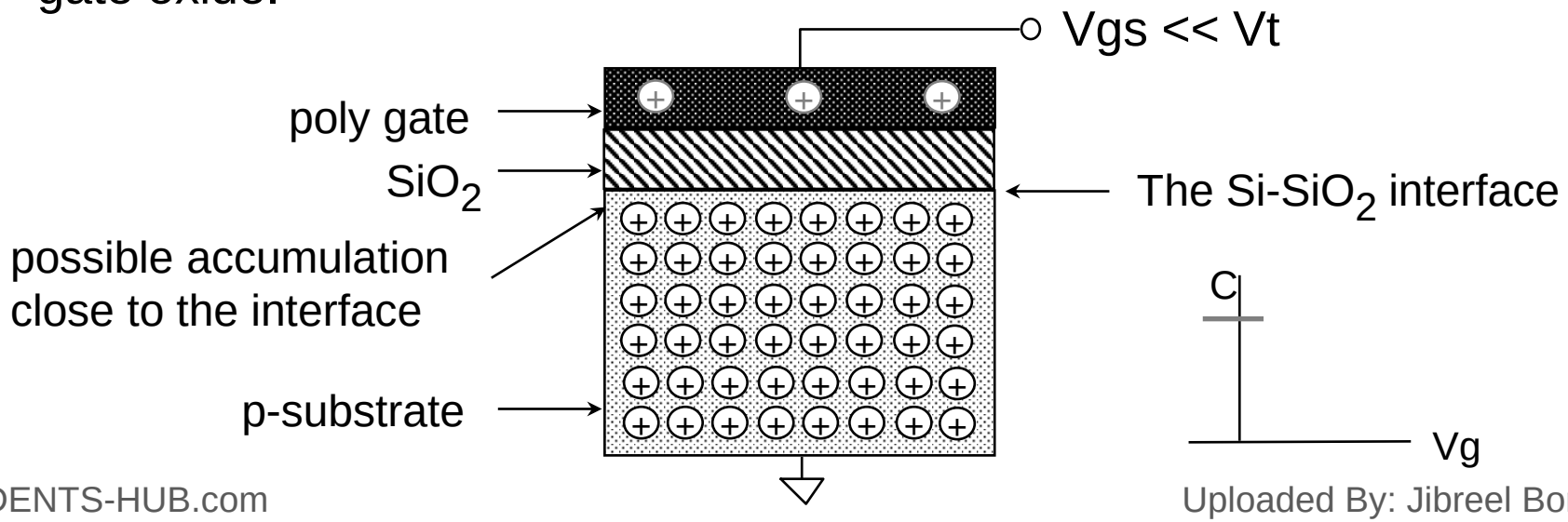
The surface state in a MOS structure

➡ *Accumulation close to the Si-SiO_2 interface*

- Refer to a MOS structure on a p-type substrate.
- Assume that a negative or small positive voltage is applied to the gate of the MOS.

THEN,

- The free majority carriers at the substrate (holes) will be attracted to the Si-SiO_2 interface and will form there an accumulation layer .
- While in this region, the gate capacitance is solely determined by the gate oxide.

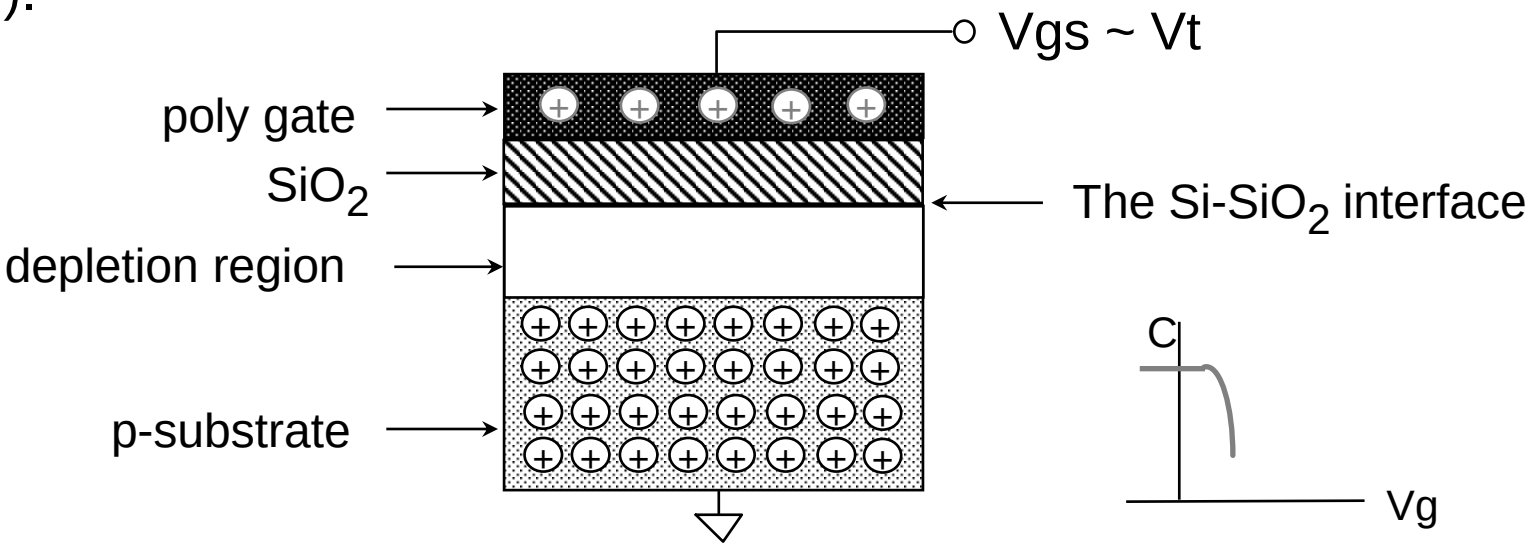


The surface state in a MOS structure (cont'd)

➡ *Depletion close to the Si-SiO_2 interface*

- If the gate voltage is now increased,

  THEN,
- The free majority carriers at the substrate (holes in this case) will be repelled from the Si-SiO_2 interface and the region close to the interface will become depleted of free carriers.
- The gate capacitance in this region is reduced - distance between charges is increased (effective dielectric thickness = oxide + depletion layer).



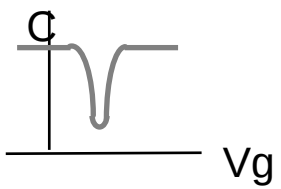
The surface state in a MOS structure (cont'd)

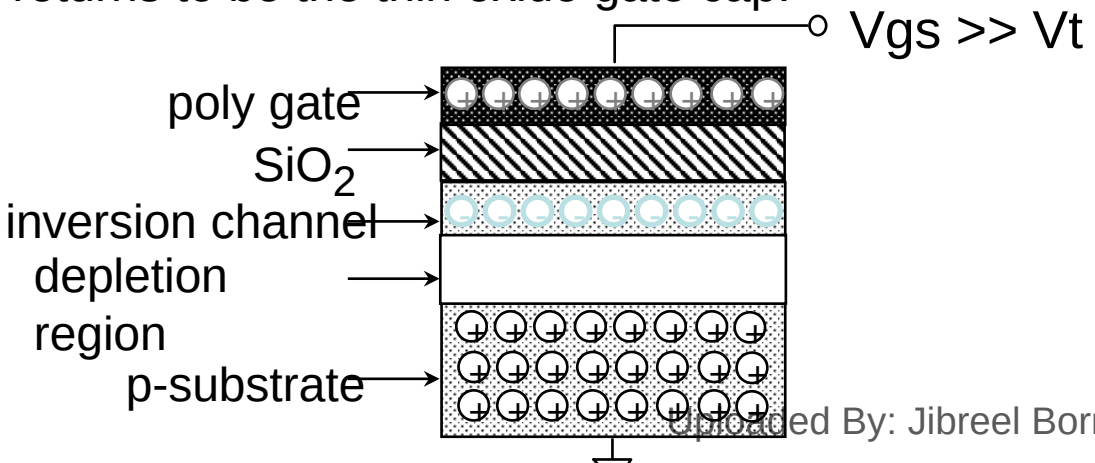
➡ *Inversion close to the Si-SiO_2 interface*

- If the gate voltage is higher than a point that is called the "threshold voltage" (Vt)

THEN:

- Minority carriers in the substrate (electrons in this case) will be attracted towards the Si-SiO_2 interface and this region will become an 'n-type' region (i.e. the Si type was inverted). Note: usually the carriers are injected from the source.
- From now on the depletion region stops to grow. The charge change at the gate is balanced by a change in the inversion charge (the depletion charge is kept constant). This assumes the availability of minority carriers (which is the case in a MOS trans.).
- The total gate capacitance returns to be the thin oxide gate cap.





The threshold voltage - V_T

- The existence of trapped charge in the oxide and the different work functions between the gate material and the Si substrate, cause that even at $V_{GS} = 0$ there is some charge built up at the Si side of the MOS structure. The 'flat band' voltage, V_{FB}, is the V_{GS} for which the band structure of the semiconductor is flat, meaning: no net charge at the Si side of the interface.
- If we now want to create an n-type channel below the Si-SiO_2 surface we need to increase V_{GS}. As V_{GS} is increased the Si close to the surface first becomes depleted of holes and only then (at higher V_{GS}) electrons start to build the channel.

  Vt = V_{FB} + … (details are quite involved)
- Doping the Si, the poly, or both, impact the flat-band V_{GS} and therefore Vt.
  - High Si doping ==> harder to deplete and inverse the channel ==> higher Vt
- We can define the 'threshold voltage' as the V_{GS} that below it the transistor's current (I_{DS}) effectively drops to zero.
- Negative bulk bias and or positive source bias ==> source-bulk diode in reverse bias ==> larger Vt (also known as the "body effect")

The threshold voltage (cont'd)

Watch : https://www.youtube.com/watch?v=S6SagsuMcxM

https://www.youtube.com/watch?v=FRRZV5eHhUk

- The threshold voltage equation (the simplest one):

$$V_T = V_{FB} + |2\phi_F| + \frac{|Q_B|}{C_{ox}}$$

$$V_{FB} = \phi_{MS} - \frac{Q_{ss}}{C_{ox}} - \frac{1}{C_{ox}} \int_0^{t_{ox}} \frac{x}{t_{ox}} P(x)\, dx$$

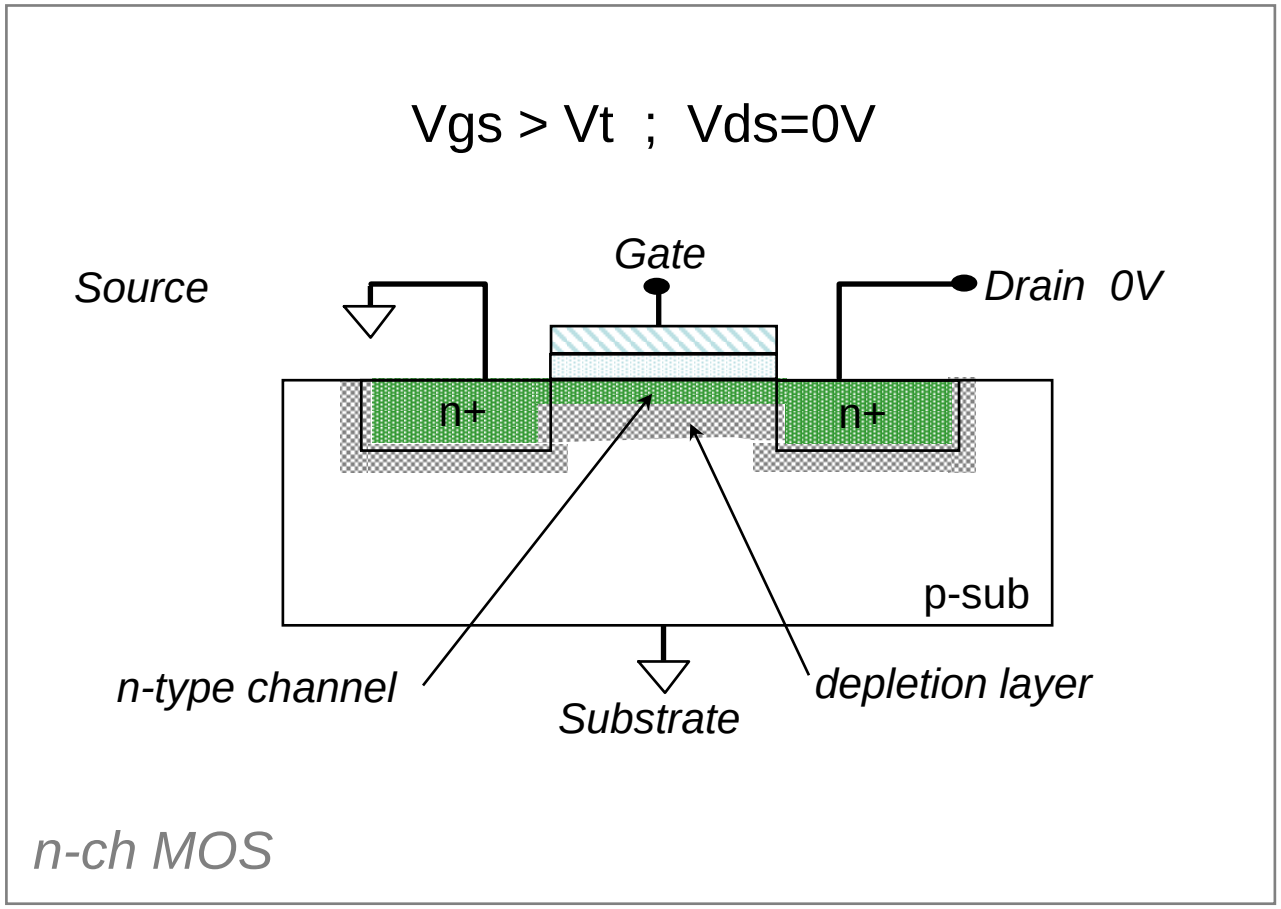
where the "flat-band" voltage is:

$$|Q_B| = \sqrt{2\varepsilon q N_B |2\phi_F|}$$

and the "bulk charge" term is given by:

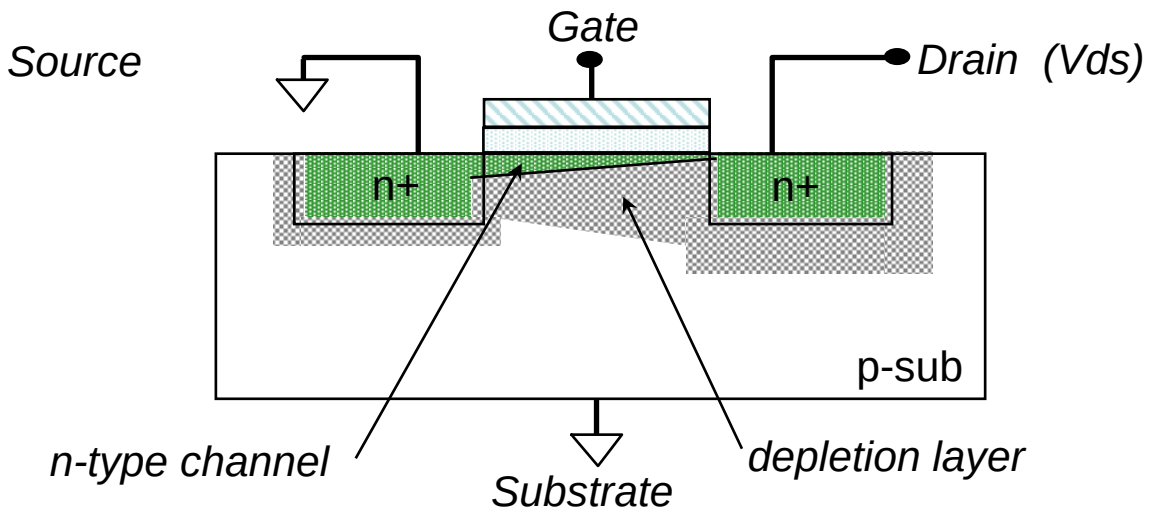
Q_B: bulk charge, N_B- bulk doping, ϕ_F - energy diff: donor/acceptor to Fermi level, Q_{ss} - surface charges (Si-SiO_2), ϕ_{ms} - difference in work function between metal (poly) and Silicon, P(x) - charges within the oxide

The MOS transistor - the different modes of operation



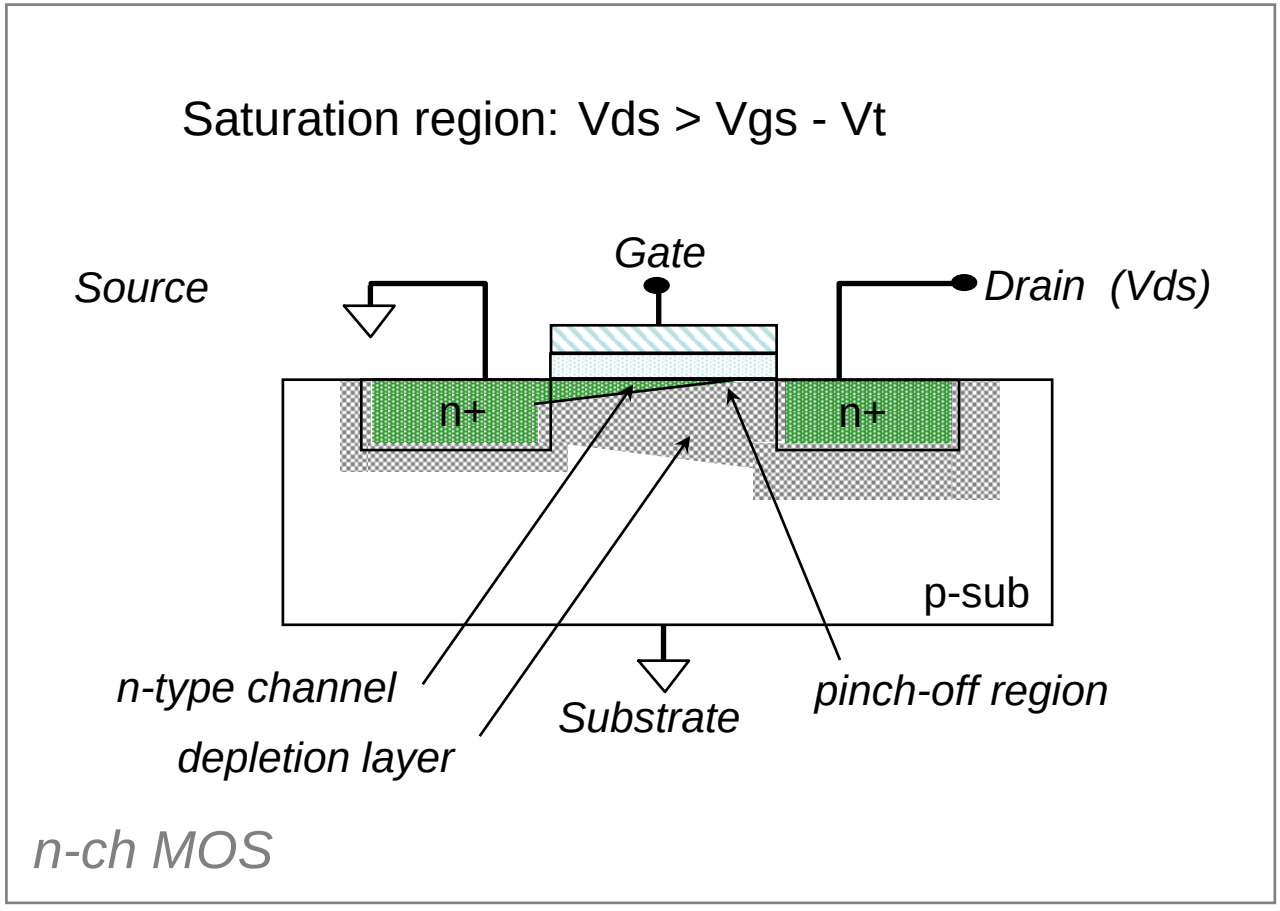
The MOS transistor - the different modes of operation

Linear region: Vgs - Vt > Vds > 0V



n-ch MOS

The MOS transistor - the different modes of operation



NMOSFET Transistors

- cut off region

$V_{gs} < V_t \Rightarrow \boxed{I_D = 0}$

- Linear region

when $V_{ds} < V_{gs} - V_t \Rightarrow i = \frac{dq}{dt} = \frac{\Delta Q}{\Delta t}$

$Q = C \cdot V = \varepsilon_{ox} \cdot \frac{w \cdot L}{t_{ox}} \cdot V$

* but $V = V_{GS} - V_T$

$\Rightarrow Q = \varepsilon_{ox} \cdot \frac{w \cdot L}{t_{ox}} (V_{GS} - V_T)$

$t = \frac{L}{V} = \frac{L}{\mu_n \cdot E} = \frac{L}{\mu_n \cdot \frac{V_{DS}}{L}} = \frac{L^2}{\mu_n \cdot V_{DS}}$

velocity $\rightarrow V$

$\Rightarrow \boxed{I_D = \mu_n \cdot \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{w}{L} \cdot (V_{GS} - V_t) \cdot V_{DS}}$

- Saturation region

$Q = C \cdot V = \varepsilon_{ox} \cdot \frac{w \cdot L}{t_{ox}} \left(V_{GS} - V_t - \frac{1}{2} V_{DS}\right)$

$t = \frac{L^2}{\mu_n \cdot V_{DS}}$ $\Rightarrow$ At pinch off $V_{DS} = V_{GS} - V_t$

$i_{DS} = \mu_n \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{w}{L} \left[V_{GS} - V_t - \frac{1}{2}(V_{GS} - V_t)\right](V_{GS} - V_t)$

$\Rightarrow \boxed{i_{DS} = \frac{1}{2} \mu_n \cdot \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{w}{L} \cdot (V_{GS} - V_t)^2}$

Working Region of PMOS is the same but with negative Voltages & Current

- Cut off region

$V_{gs} < |V_t| \Rightarrow \boxed{I_{SD} = 0}$

- Linear region

when $V_{SD} < V_{SG} - |V_t|$

$I_{SD} = \frac{dq}{dt}$, $q = C.V = \frac{\varepsilon_o \cdot W.L}{t_{ox}} (V_{SG} - V_t)$

$t = \frac{L}{V} = \frac{L^2}{\mu_{holes} \cdot V_{SD}}$

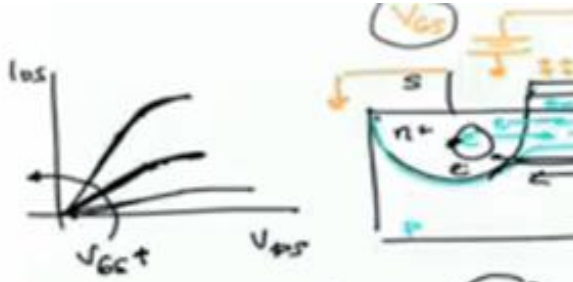
$\Rightarrow \boxed{I_{SD} = \mu_p \cdot \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{W}{L} \cdot (V_{SG} - V_T) \cdot V_{SD}}$

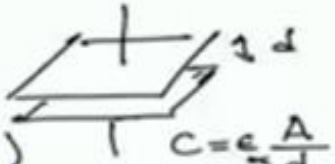
- Saturation Region

$Q = C.V = \frac{\varepsilon_{ox} \cdot W.L}{t_{ox}} (V_{SG} - V_t - \frac{1}{2} V_{SD})$

$t = \frac{L^2}{\mu_p \cdot V_{SD}}$

$\Rightarrow \boxed{I_{SD} = \frac{1}{2} \mu_p \cdot \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{W}{L} (V_{SG} - V_t)^2}$





- $i = \frac{dq}{dt} = \frac{\Delta Q}{\Delta t}$
- $Q = C \cdot V = \epsilon_{ox} \cdot \frac{W \cdot L}{t_{ox}} \cdot (V_{GS} - V_t)$

$$Q = \frac{\epsilon_{ox}}{t_{ox}} \cdot W \cdot L \, (V_{GS} - V_t) \}$$

$$t = \frac{L}{v} = \frac{L}{\mu_n \cdot E} = \frac{L}{\mu_n \cdot \frac{V_{DS}}{L}} = \frac{L^2}{\mu_n \cdot V_{DS}} \}$$

- $\boxed{I_D = \mu_n \cdot \frac{\epsilon_{ox}}{t_{ox}} \cdot \frac{W}{L} (V_{GS} - V_t) \cdot V_{DS}}$

small V_{DS}
$V_{GS} > V_t$

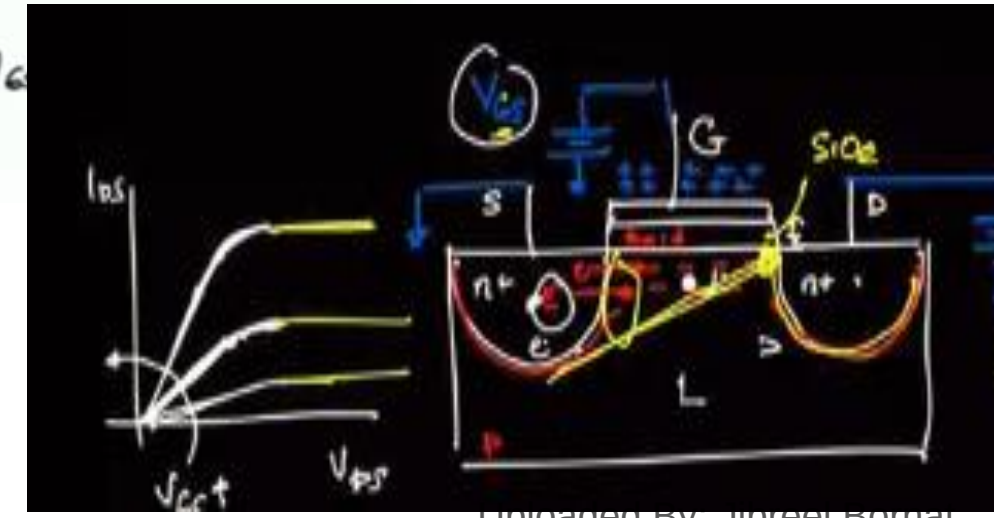
$$i_D = g \cdot V_{DS}$$

$$Q = C \cdot V = \epsilon_{ox} \cdot \frac{W \cdot L}{t_{ox}} \cdot (V_G$$

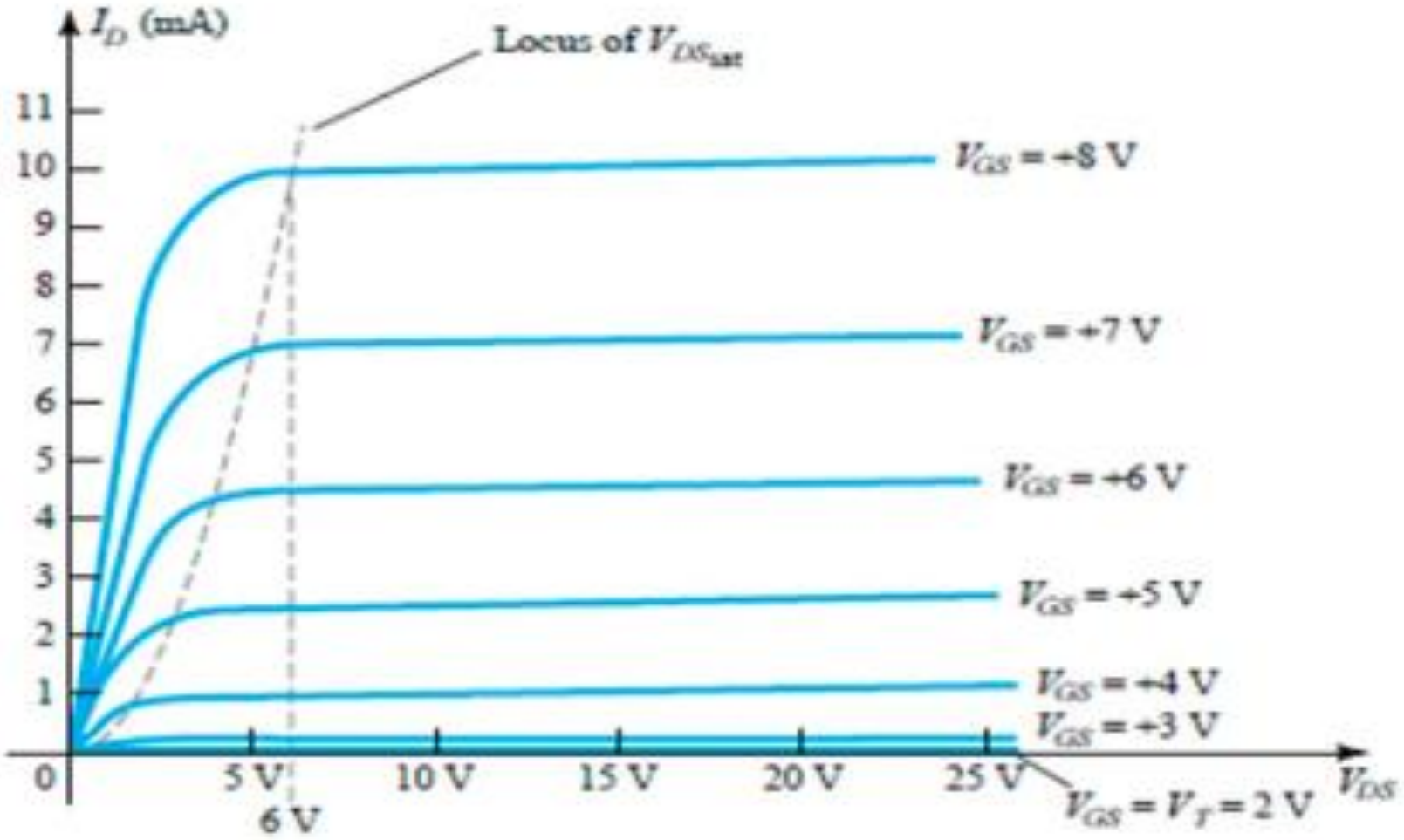
$$t = L^2 / \mu_n \cdot V_{DS}$$

$$i_{DS} = \mu_n \frac{\epsilon_{ox}}{t_{ox}} \cdot \frac{W}{L} \left[V_{GS} - V_t - \frac{1}{2} (V_{GS} - V_t) \right] (V_{GS} - V_t)$$

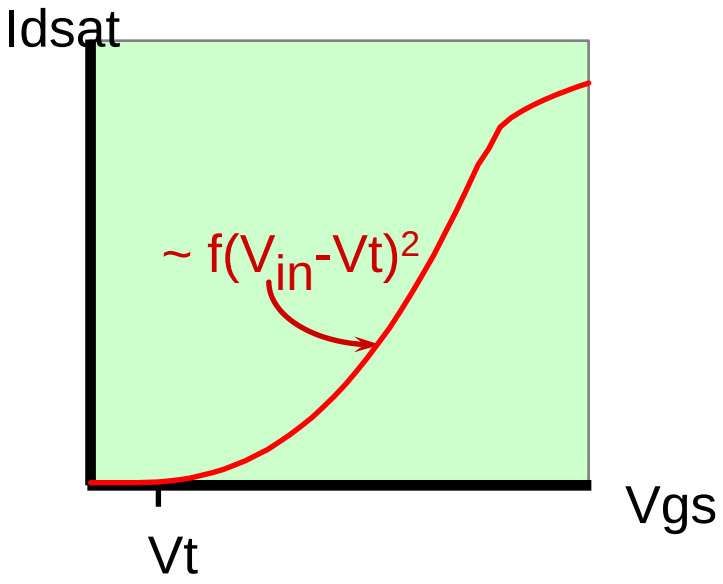
$$\boxed{i_{DS} = \frac{1}{2} \mu_n \cdot \frac{\epsilon_{ox}}{t_{ox}} \cdot \frac{W}{L} (V_{GS} - V_t)^2}$$

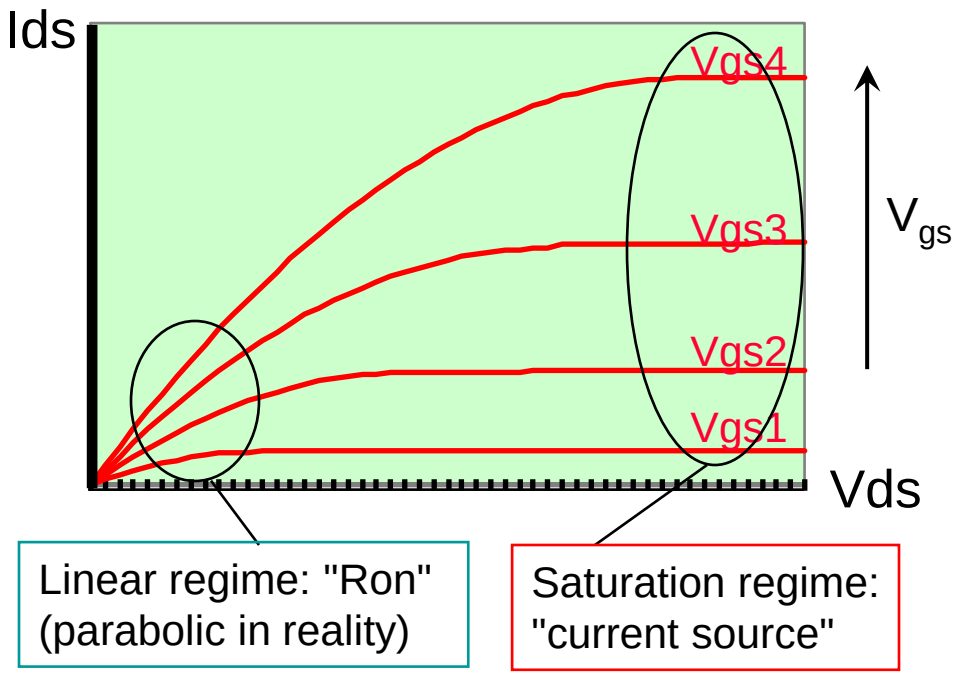


Drain Characteristics of an n-channel enhancement-type MOSFET



The MOS I-V characteristics

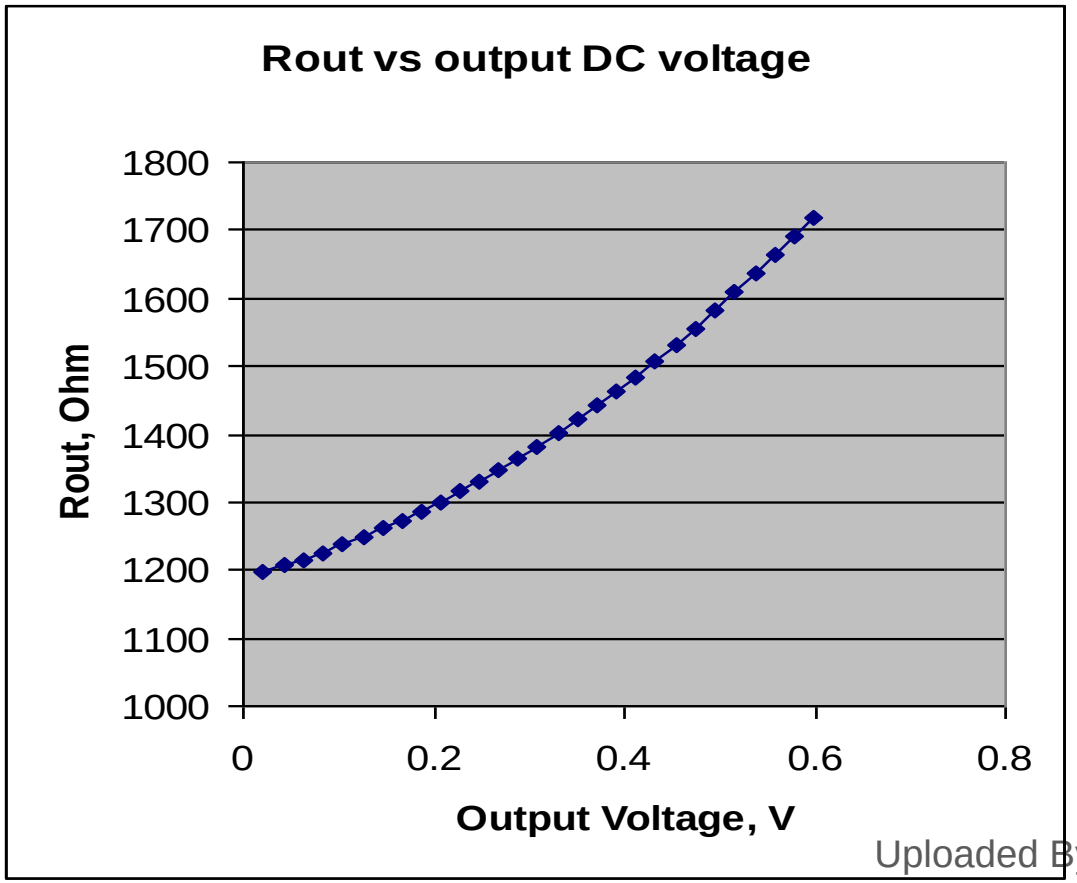




1. The transistor behaves as a non-linear resistor.
2. It's effective resistance depends on the operating point.
3. The transistor's current saturates at high Vds values.
4. Changing Vt (low Vt transistor) has high impact on Ids and hence on transistors' speed

R-on example

- The example below indicates that the linear regime is not really linear: Ron (or Rout) has non negligible voltage dependence
- This is significant for noise rejection: the larger is the noise on output pin (the drain), the weaker is the pull-up or pull-down



The MOS transistor models

- The simplest model:
  - Cut-off: Ids = 0 for |Vgs| < |Vt|
  - Linear: Ids = K/2 *[2*(Vgs-Vt)*Vds - Vds^2] for |Vgs-Vt| > |Vds|
  - Saturation: Ids = K/2 *$(Vgs-Vt)^2$ for |Vds| > |Vgs-Vt|
  - * where K = μ Cox (W/L)
- These equations can only model giant devices (long channel).

- Note that K is larger if C_{ox} is larger, fast transistors need large C_{ox}
- μ is the electron / hole mobility
  - e: 1300 ; h: 500 cm^2/V sec , respectively, in bulk Silicon

Sub-threshold current

- When the gate voltage is reduced to Vt, the current does not abruptly drop to zero.
- Around Vt the current drops off in an exponential manner with a slope of ~1decade/100mV.
- The current in the sub-threshold region can be approximated as:

$$Ids \sim Io * exp[q(Vgs-Vt)/nkT]$$

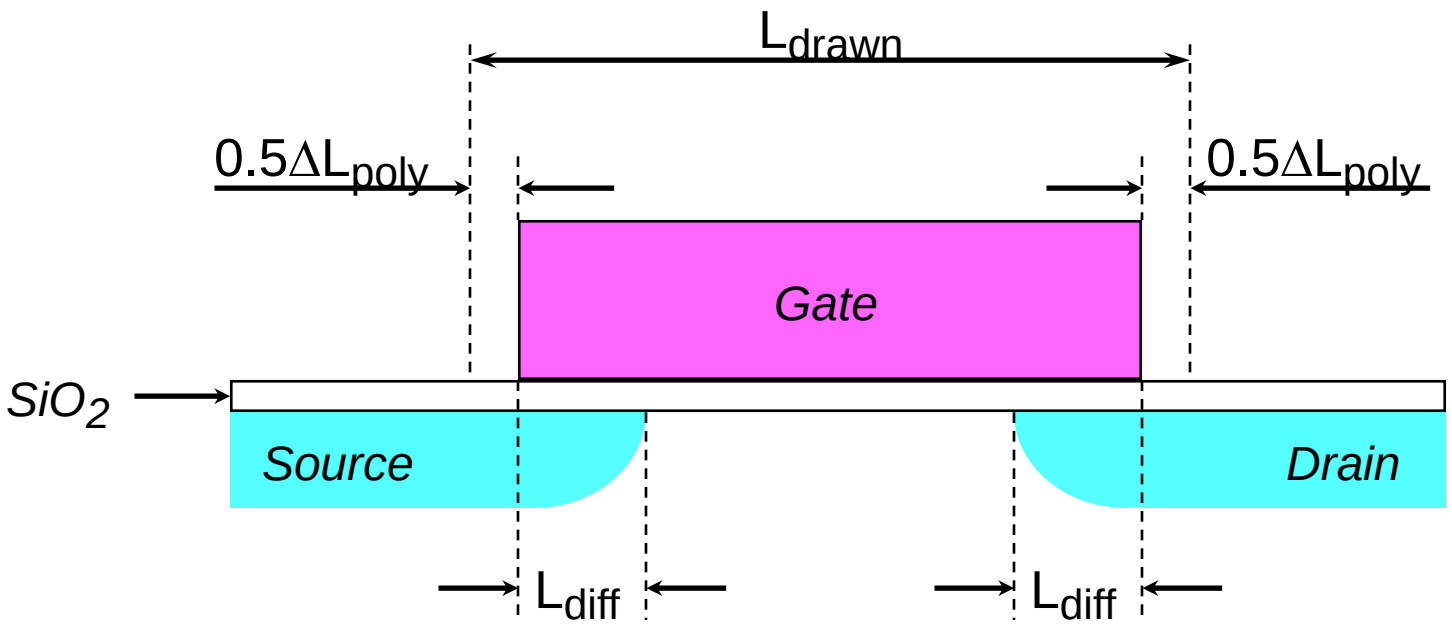
- This exponential behaviour is exploited in many analog circuits operating in weak or moderate inversion.
- In deep sub-micron technologies the sub-threshold current (or off current or "source-drain leakage") is a major power consumption concern, especially for low activity nodes like caches
- Note that leakage Ids grows exponentially at elevated temperature

è *The sub-threshold conduction places a lower limit on the Vt reduction for low voltage operation!*

Gate leakage

- As oxides become thinner, gate leakage is showing up
- **Tunnelling effect** enables gate-bulk current through the "isolating" oxide layer
- Gate current is weakly temperature dependent
  - Significant to the overall leakage at low temperatures
    - for example standby mode in laptops - die reduces to ambient temperature, say 20-30C
  - Negligible overall impact at high temperatures - normal operating conditions (typical Tj is 80-110C)
- P1262, 1264: gate leakage is comparable to source-drain leakage at room temperature
  - Still 4-5X smaller at operating conditions
- Process trend
  - Oxide cannot be thinner (higher Cox should come from high K oxide)
  - P1264 is probably the worst process in terms of gate leakage

Le - The effective channel length

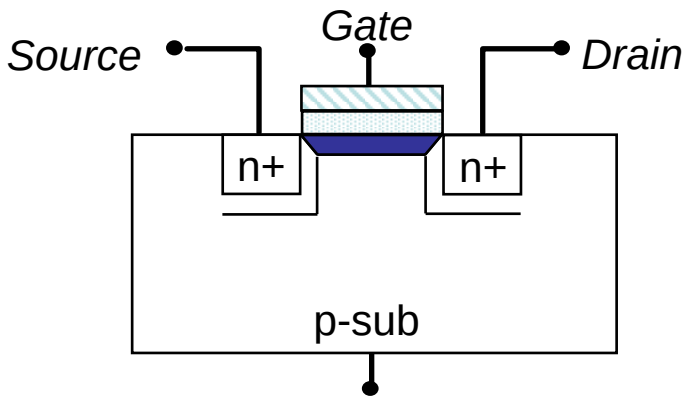


$$L_e = L_{drawn} - \Delta L_{poly} - 2\,L_{diff}$$

- In terms of the process file: ΔL_{poly} = -dell and L_{diff} = dfl

‘Short’ (… i.e. normal) channel effects

- Vt(L) - As the channel length decreases, the depletion region below the gate can no longer be approximated as a rectangular region. So, as L decreases so does Qb and therefore Vt will also decrease.



- Vt(Vd) - As Vd is higher the drain depletion region increases, causing a decrease in Vt.

Channel length modulation

- When the drain voltage is beyond onset of saturation, the pinch-off point starts to move towards the source.
- The voltage at the pinch-off point is still Vgs-Vt. The remaining Vds-(Vgs-Vt) is dropped on the depletion layer between the p-o point and the drain.
- Since the same voltage is now dropped on a shorter channel length, the drain current increases.
- This effect is seen as a non-zero slope in the Ids-Vds curves.
- The non-zero slope represents a finite output impedance.
- The simplest way to model this effects is:

$$Ids = K/2 *(Vgs-Vt)^2(1+\lambda Vds)$$

The ‘body effect’

- The ‘bulk charge’ term in the Vt equation changes with the reverse bulk-source bias.
- As the reverse |Vsb| bias increases so does |Vt|.

$$Vt = Vto + \gamma\left[\sqrt{|2\phi_f| + |Vsb|} - \sqrt{|2\phi_f|}\right]$$

where $\gamma = \frac{1}{Cox}\sqrt{2q\varepsilon N_B}$ **(γ ~ *0.3-1.4 V $^{0.5}$*)**

- A higher Vt leads to lower device currents (i.e. slower circuits).
- Transistor matching is also degraded by the ‘body effect’
- Low Vt devices can be obtained by forward body bias (0.3-0.5 Vcc)
- Note: transistors in stack have "body effect" since Vs > Vbulk

The temperature dependence of I_{DS}

- The transistor **current changes** with the operating temperature mainly through the mobility and Vt temperature dependences.
- Mobility: at normal operating temperatures

$$\mu(T) = \mu_0 \left(\frac{T}{T_0}\right)^{-\frac{3}{2}}$$

i.e. as T increases μ decreases

- Vt(T):

$$V_T = V_{T0} + \alpha(T - T_0)$$

where α ~2 mV/^{o}C

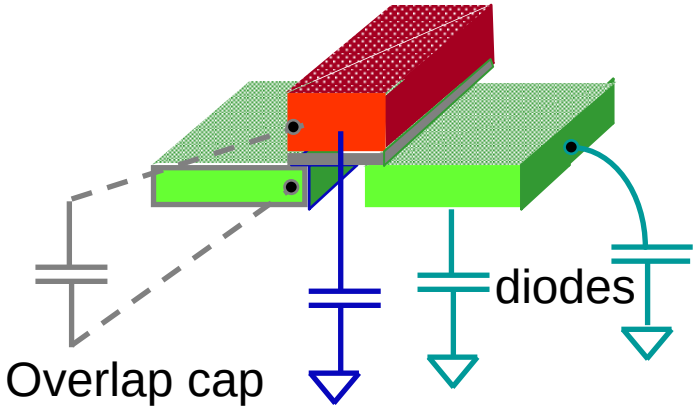
i.e. as T increases Vt increases

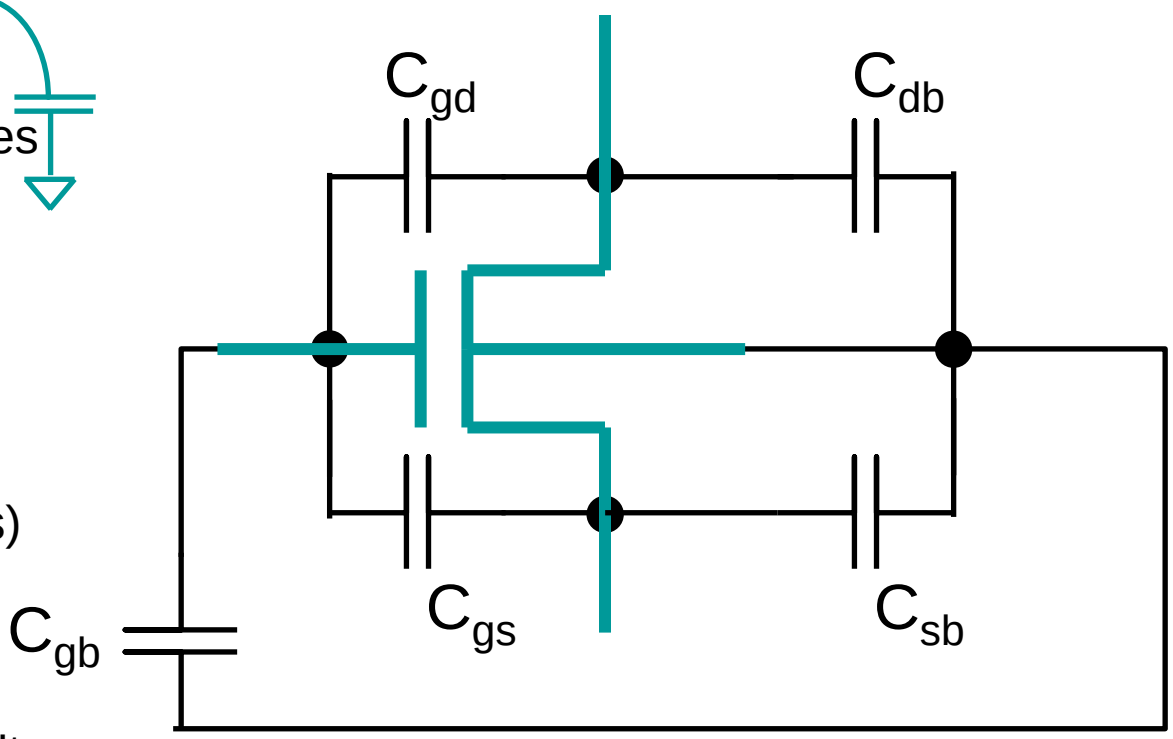
The transistor's current temperature dependence is more dominated by the mobility. So, *As T increases Ids decreases.*

*Vt dependency means that low an high Vt **scale differently** with temperature*

Shutting down units (power saving) introduce dynamic thermal variation

Capacitance of the MOS Transistor





Parameters:
Diffusions area (diodes)
Diffusions perimeter (diodes)
Gate W (overlap cap)
Gate W*L (gate cap)
Doping profiles
Note: all capacitors have voltage dependence (not simple caps)

Reliability Issues

- **Hot-e degradation (mainly n-ch concern)**
  - **Vt and Ids degrade**
- **P-ch BT (Bias Temp)**
  - **Ids instability: degrade under bias (accelerated with temperature), recovers under zero bias**
- **Vt stability**
  - **Affects AC performance and races**
  - **After long periods of inactivity**
  - **Reversible**
- **Gate stress (gate oxide wearout)**
  - **High voltage at the gate may ruin the transistor**
  - **Lifetime strongly depends on voltage, spikes, and temperature**

Hot-e degradation

- When a MOS transistor is in saturation, the electric field across the pinch-off region may be high enough that carriers gain there enough energy to excite electron-hole pairs.

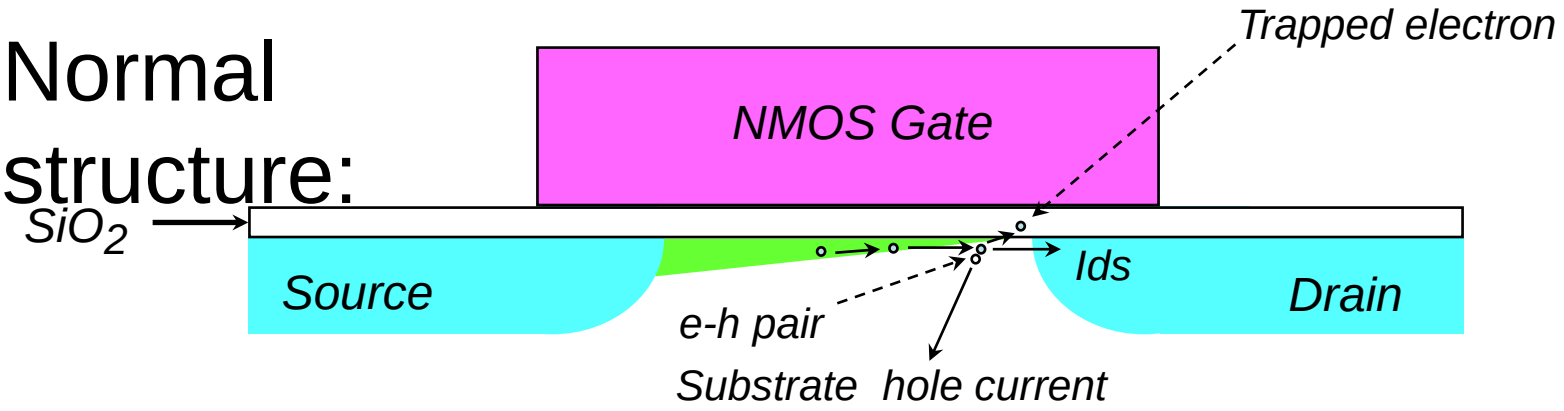
  1V / 0.01um = 10^6 V/cm, huge electrons (holes) acceleration
- The e-h pairs become components of the drain and substrate currents.
  - The holes (*electrons*) usually flow towards the p-substrate (*n-well*) in an n-ch (*p-ch*) device, increasing the substrate currents (bad for latch-up!).
  - The excited electrons (holes) that reach the drain cause an increase in Ids (weak avalanche).
- However, the reliability concern is that part of the hot-e can penetrate the gate oxide !!!
- P-ch transistors are usually less susceptible to hot-e degradation due to the holes lower mobility (higher effective mass).

Hot-e degradation (cont'd)

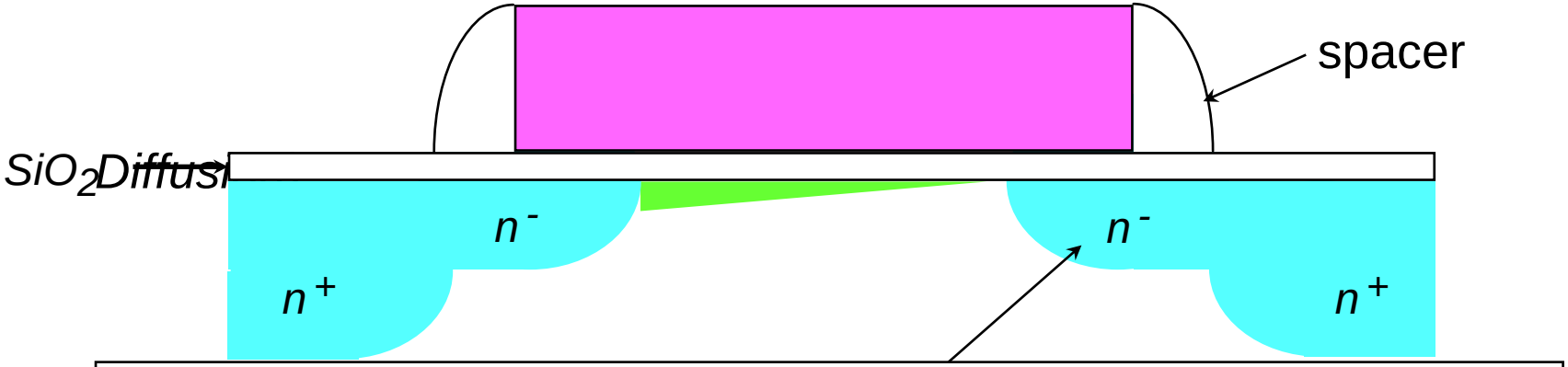
- Electrons that penetrated the gate oxide remain trapped there (in normal operating conditions). ***The hot-e effect is accumulative*** *!*
- The negative trapped charge in the oxide (near the drain) of an NMOS cause an increase in Vt.
- *The hot-e effect result is a degradation in the NMOS Ids* (due to the higher effective Vt). Meaning: *circuits slow down!*
- *Note!* the hot-e degradation has a negative feedback behaviour.
- Hot-e degradation is a long term reliability concern. The device life time that Intel guarantees is 7 years of constant operation at the worst case conditions (within the spec). Meaning: When a device frequency is tested after fabrication, the hot-e degradation with time must be taken into account.
- Therefore, if a device frequency should be F it will be tested at F + 'the hot-e guard-band'. For example, in the P55C the hot-e GB was ~4.4MHz (2.2%) for 200MHz devices.

Hot-e degradation (cont'd)

- Normal structure:

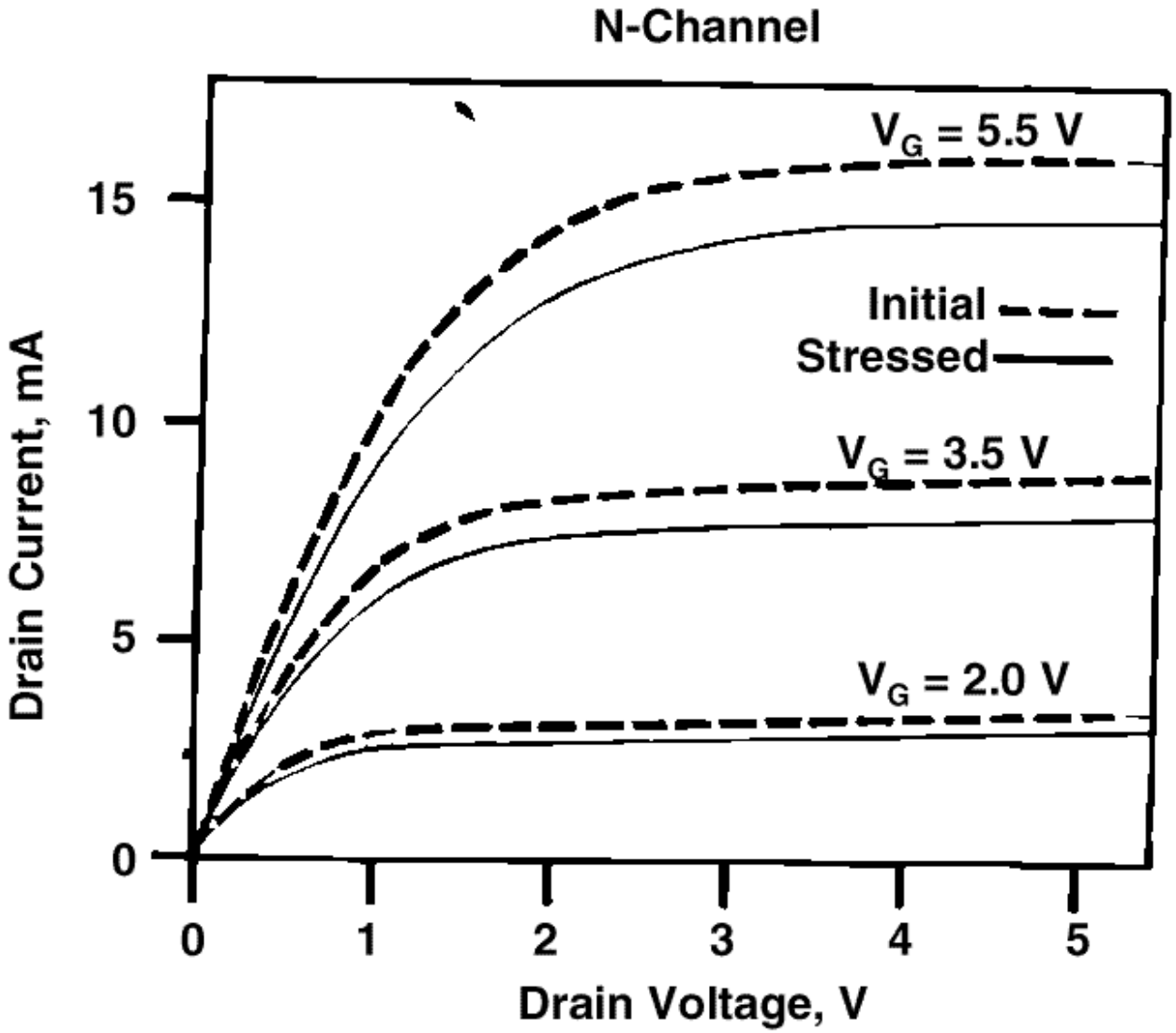


LDD structure:

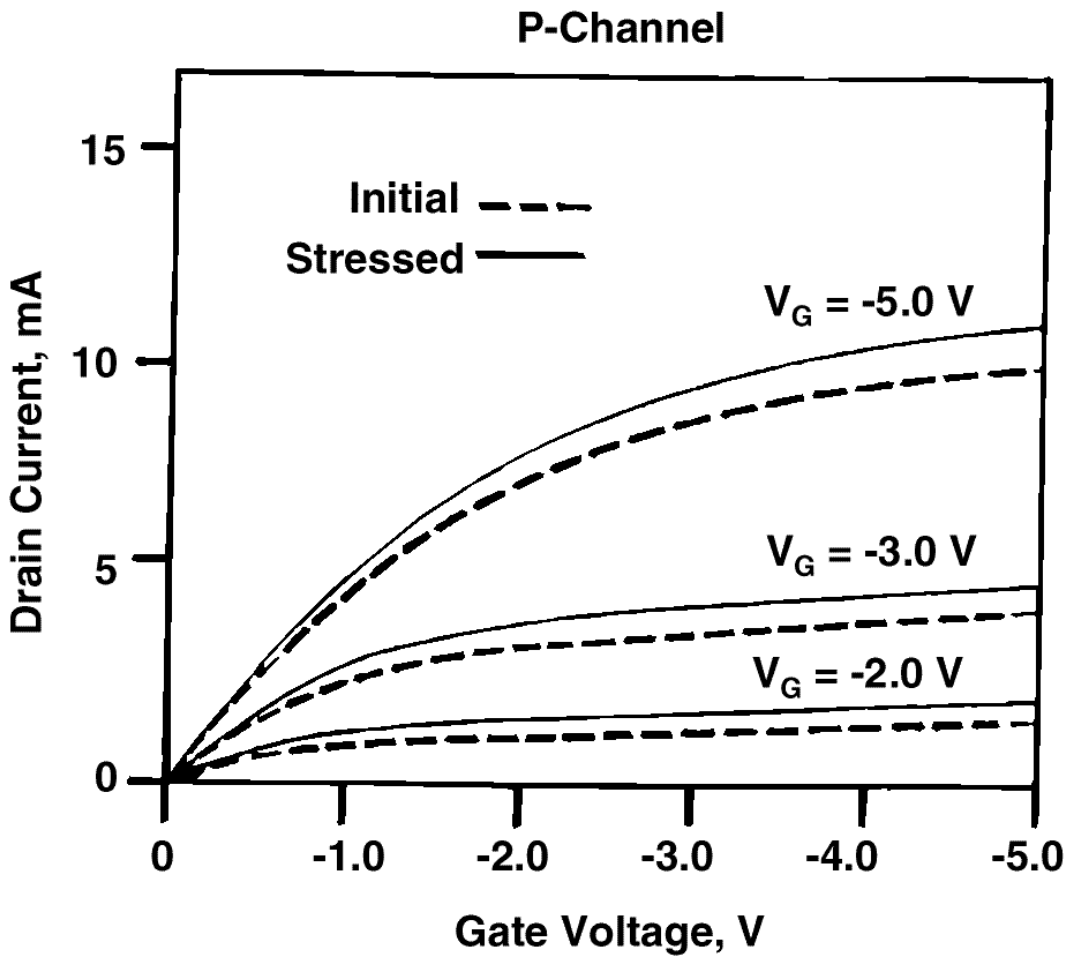


The n^- region acts as a series resistor therefore the transistor's Vds is reduced. This lightly doped region also reduces the electric field across the pinch-off region. (Side effect - the transistor is slower!)

N-channel hot-e degradation

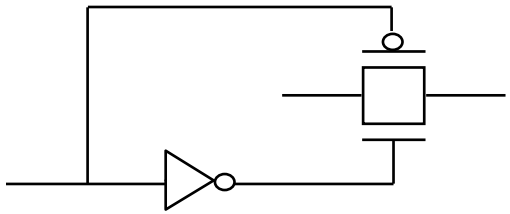


P-channel hot-e degradation



Hot-e degradation (cont'd)

- Pass-gates and bi-directional n-ch MOS.
  - Transistors in pass-gates can operate in both directions. Therefore, the degradation occurs at both ends of the transistor.
  - When operating in bi-directional mode, %ΔIDL is usually greater than in forward operation only (~1.4x in P854 & P856).
- Special design guidelines for pass-gates:
  - Turn-on the p-ch one inversion before the n-ch (usually not practical)
  - Increase the NMOS channel length (some area and power penalty)



  - The above guidelines will maintain the pass-gate degradation at similar levels as for other uni-directional transistors.

Will discuss this later

Hot-e degradation (cont'd)

- *How to reduce the hot-e degradation ?*
  - Decrease Cload (reduce fanout).
  - Speed up the input edge rate.
  - Avoid slowly varying output signals where possible.
  - Increase Weff of NMOS.
  - Use NAND gates if possible (or gates with NMOS in series).
  - Increase NMOS channel length.
  - Avoid false transitions (glitches) - reduce the real AF.
  - Avoid capacitive coupling above Vcc (overshoots).

Vt Stability

- The ac performance of a transistor may be affected after applying a dc voltage to the transistor's gate for a long time. This is not related to hot-e, it occurs even with Ids=0 (as long as there is an E-field across the oxide).
- This effect is called 'Vt stability'. It is primarily caused by the charging/discharging of slow interface states and by mobile ions drift in the oxide.
- This effect is reversible!

- What can be done in the design ?
  - Avoid long periods of high E-field on the transistors gates (e.g. at stop clk).

Will discuss this later

Gate Stress (Oxide Wearout)

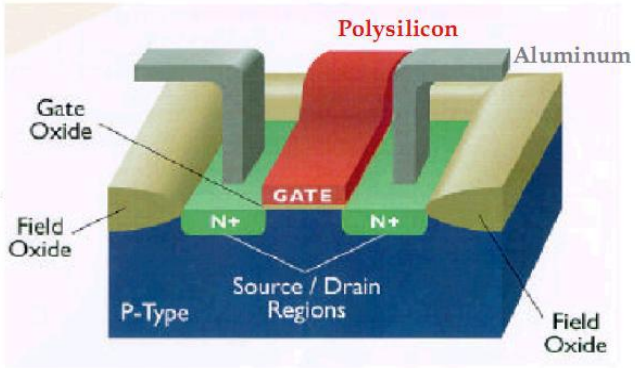
- Every process has a maximum allowed voltage stress on the transistors. This may be a dc and/or an ac value.
- The thin gate oxide has a certain density of defects that may ruin the transistor if a high voltage is applied to the gate. Note that this max voltage is far lower than the Breakdown Voltage of the oxide.
- Gate oxide life-time is exponential in both voltage and temperature
  - Max Vcc (and hence frequency) is limited by Oxide wearout
  - Voltage overshoots / undershoots significantly limit oxide life time
  - Noise-like analysis is needed to predict oxide wearout
    - Aided by AF since it is a cumulative (ageing) effect
- What can be done in the design ?
  - Avoid or limit the coupling noise.
  - Provide a 'gate clamping' mechanism at all the gates.

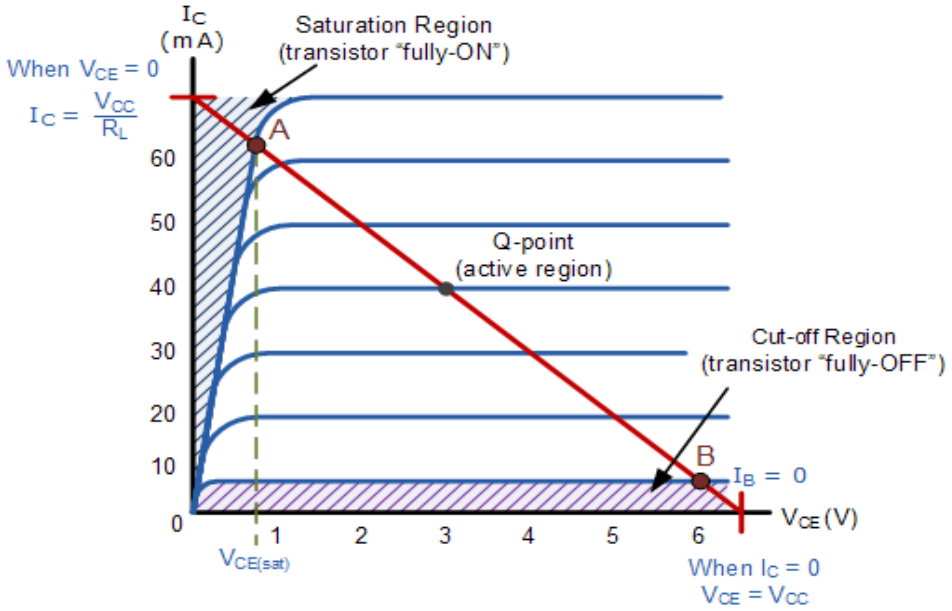
Will discuss this later

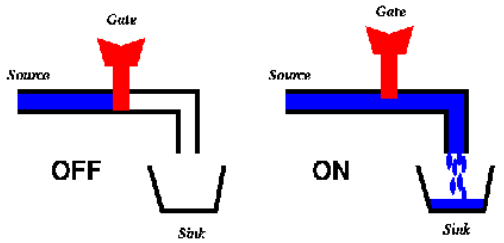
Relax and watch

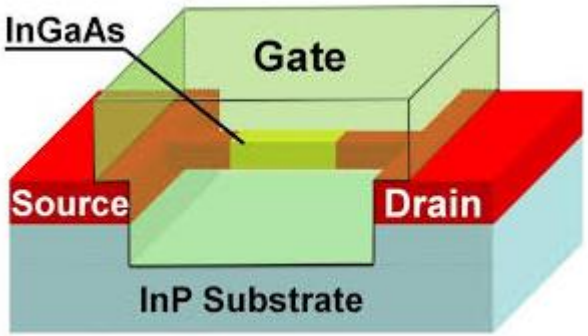
- https://www.youtube.com/watch?v=pvhWzOE2CYk&list=PLRqxkeNbuY7206obBKw_IGdhkqHXcDzVO
- https://www.youtube.com/watch?v=6-tKOHICqrI
- *https://www.youtube.com/watch?v=M3a19mMwhto*
- https://www.youtube.com/watch?v=o31uckqYchE
- Transistor :https://www.youtube.com/watch?v=U2wWpLQACo8&t=2s
- MOSFET BASIC 2: https://www.youtube.com/watch?v=GrvvkYTW_0k
- Chip flow-synopsys
- https://www.youtube.com/watch?v=sTfGzfyMrVc
- https://www.youtube.com/watch?v=o31uckqYchE

The MOS Transistor

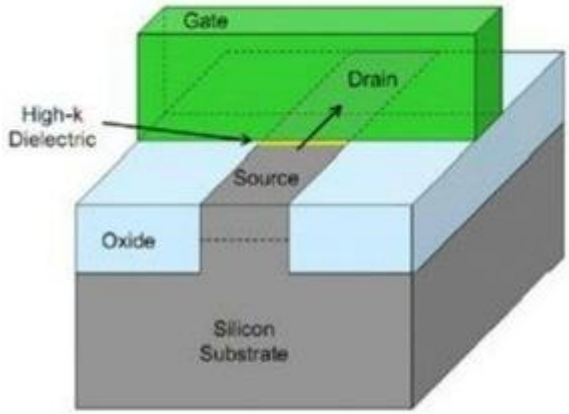






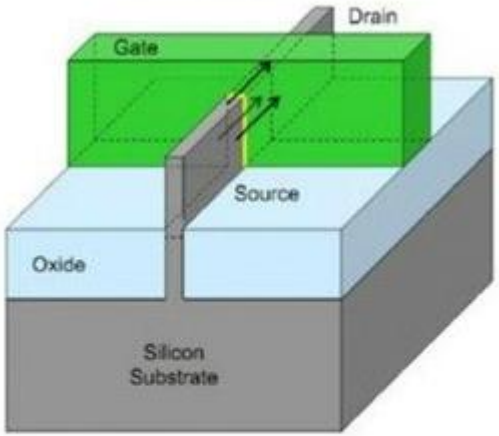


Traditional Planar



Traditional 2-D planar transistor form a conducting channel in the silicon region under the gate electrode when in the "on" state

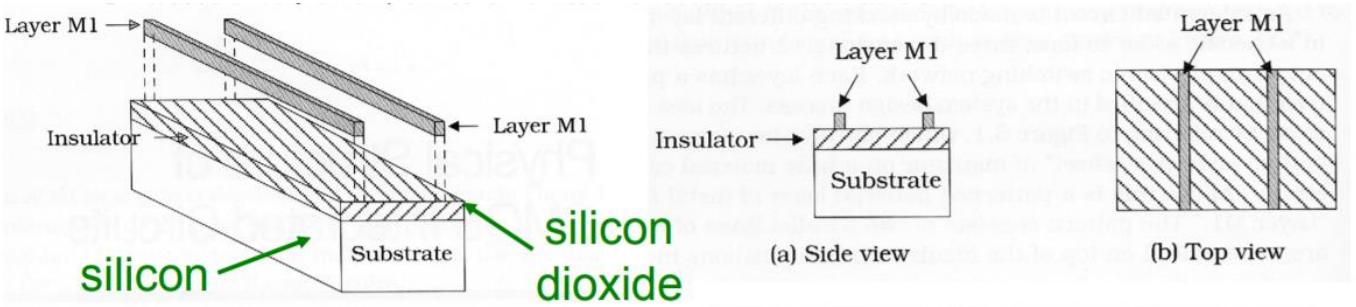
3D FinFET



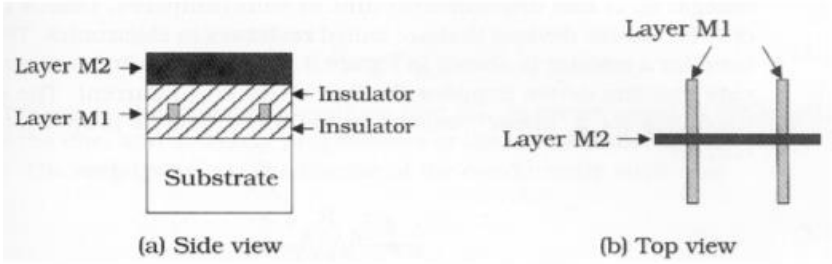
3-D Tri-Gate transistor form conducting channels on three sides of a vertical fin structure, providing "fully depleted" operation

Integrated Circuit Layers

- Integrated circuits are a stack of patterned layers
  - – metals, good conduction, used for interconnects (was Al, now copper)
  - – insulators (silicon dioxide, High K-dielectric), blocks conduction
  - – semiconductors (silicon), conducts under certain conditions
  - – Polysilicon, resistive element, forms tx gates.
- Stacked layers form 3-dimensional structures



- Multi-layer metals :
  - background assumed to be silicon covered by silicon dioxide



Intrinsic Silicon Properties

- Read textbook, section 3.2.1, 3.2.2, 3.2.3 – Won't have time to cover this in detail in lecture, but it's important to understanding how transistors work
- Intrinsic Semiconductors – undoped (i.e., not n+ or p+) silicon has intrinsic charge carriers
  - electron-hole pairs are created by thermal energy – intrinsic carrier concentration ≡ ni = 1.45x10^10 cm^-3, at room temp.
  - function of temperature: increase or decrease with temp?
  - – n = p = ni, in intrinsic (undoped) material
  - n ≡ number of electrons, p ≡ number of holes
  - mass-action law, np = ni2 applies to undoped and doped material

Book:
J. Uyemura, Introduction to VLSI Circuits and Systems, Wiley, 2002. ISBN 0-471-12704-3
Rabaey, Jan, Anantha Chandrakasan, and Bora Nikolic. Digital Integrated Circuits: A Design Perspective. 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002. ISBN: 0130909963.

Extrinsic Silicon Properties

- doping, adding **dopants** to modify material properties
  - n-type = n+, add elements with extra electrons
    - (arsenic, As, or phosphorus, P), Group V elements
    - $n_n \equiv$ concentration of electrons in n-type material
    - $n_n = N_d$ cm^{-3}, $N_d \equiv$ concentration of donor atoms
    - $p_n \equiv$ concentration of holes in n-type material
    - $N_d\, p_n = n_i^2$, using mass-action law
      - always a lot more n than p in n-type material
  - p-type = p+, add elements with an extra hole
    - (boron, B)
    - $p_p \equiv$ concentration of holes in p-type material
    - $p_p = N_a$ cm^{-3}, $N_a \equiv$ concentration of acceptor atoms
    - $n_p \equiv$ concentration of electrons in p-type material
    - $N_a\, n_p = n_i^2$, using mass-action law
      - always a lot more p than n in p-type material
  - if both N_d and N_a present, $n_n = N_d - N_a$, $p_p = N_a - N_d$

n+/p+ defines region as heavily doped, typically ≈ 10^{16}-10^{18} cm^{-3} less highly doped regions generally labeled n/p (without the +)

Example

$n_i^2 = 2.1 \times 10^{20}$

Conduction in Silicon Devices

- doping provides free charge carriers, alters conductivity
- conductivity in semic. w/ carrier densities *n* and *p*
  - $\sigma = q(\mu_n n + \mu_p p)$
    - q ≡ electron charge, $q = 1.6 \times 10^{-19}$ [Coulombs]
    - μ ≡ mobility [$cm^2/V\text{-}sec$], $\mu_n \cong 1360$, $\mu_p \cong 480$ (typical values in ***bulk*** Si)
- in n-type region, $n_n \gg p_n$
  - $\sigma \approx q\mu_n n_n$
- in p-type region, $p_p \gg n_p$
  - $\sigma \approx q\mu_p n_p$
- resistivity, $\rho = 1/\sigma$
- Can now calculate the resistance of an n+ or p+ region

$\mu_n > \mu_p$

electrons more mobile than holes

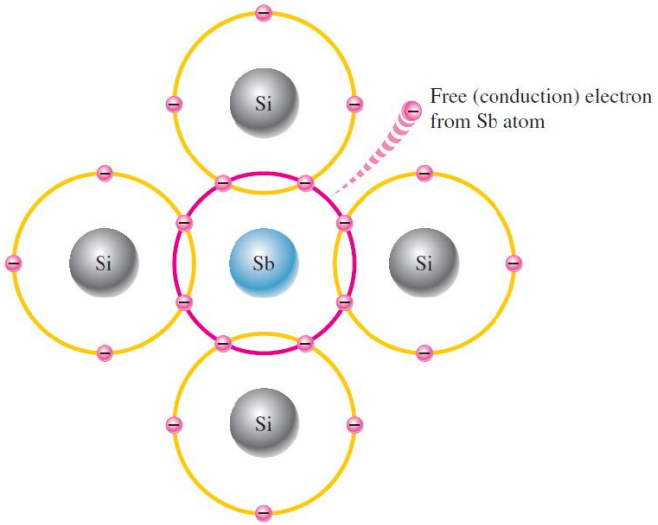
conductivity of n+ > p+

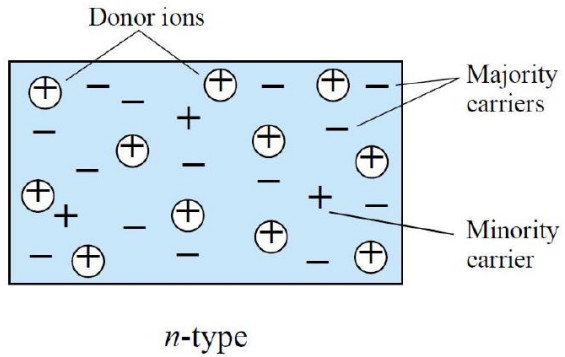
Mobility often assumed constant

but is a function of Temperature and Doping Concentration

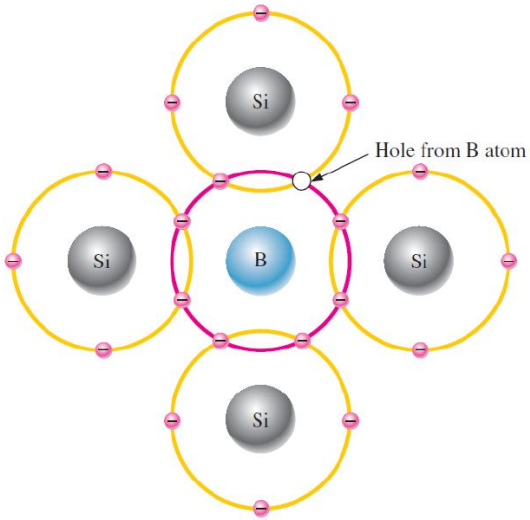
Semiconductor

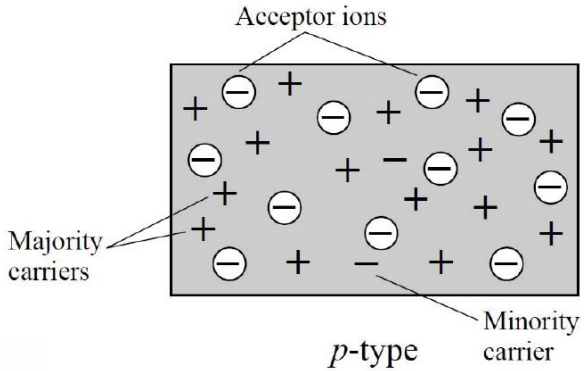
N - type semiconductor

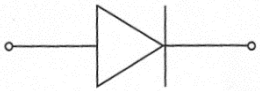


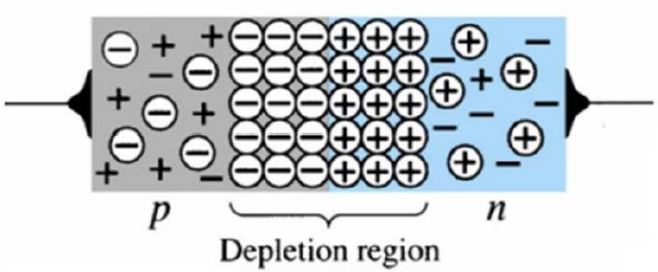


P - type semiconductor





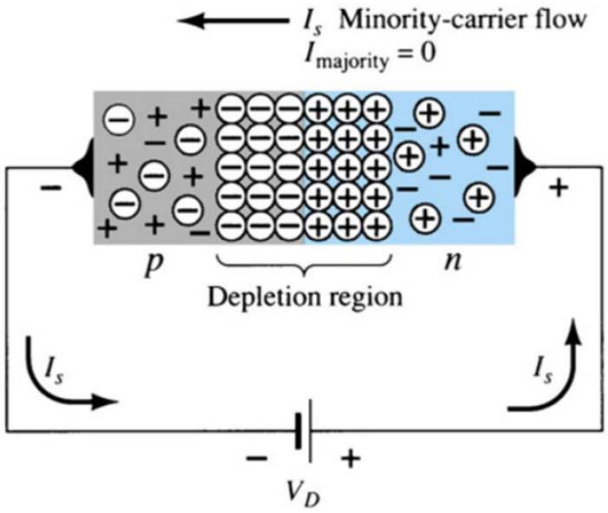




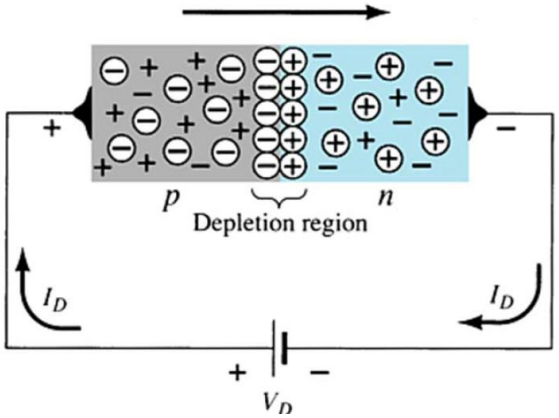
- The p-n junction is the basis for diodes, certain transistors ,and other devices.

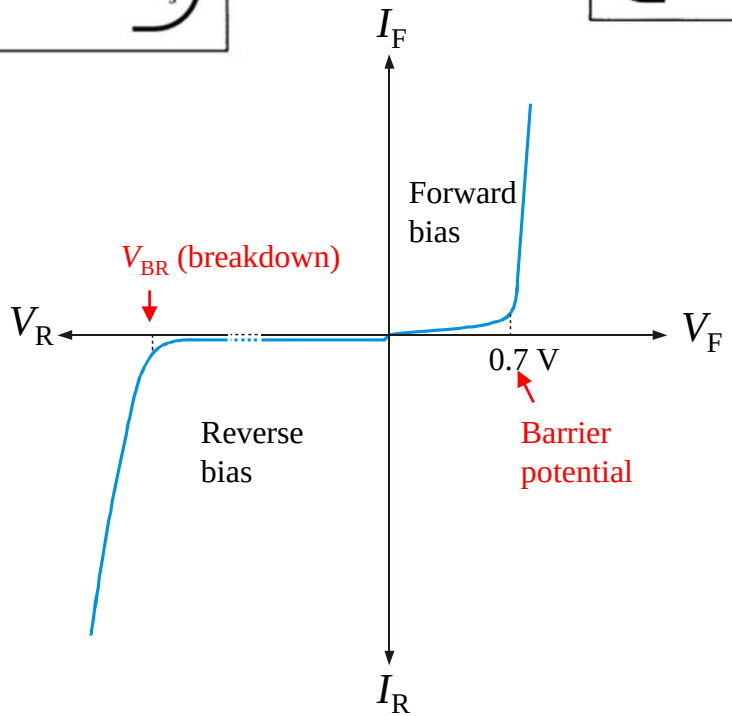
Semiconductor- Diods

Reverse bias of a pn junction



Forward bias of a pn junction



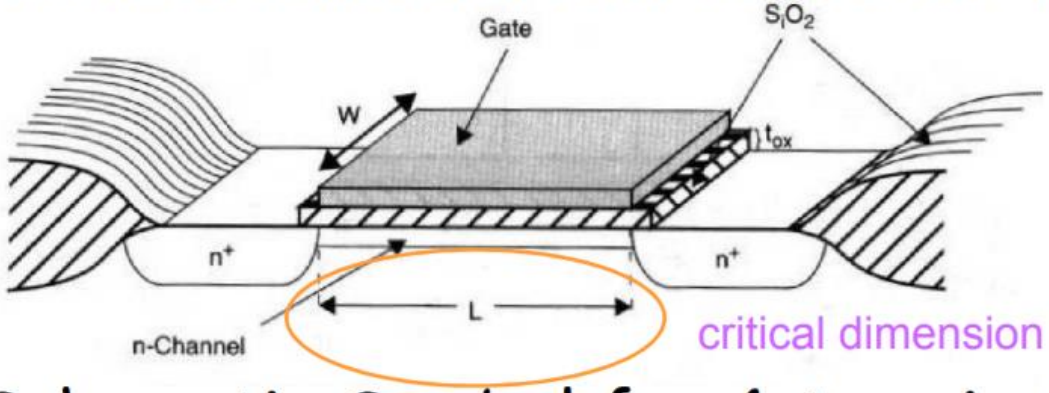


What are P-type and N-type ?

- Semiconductors are classified in to P-type and N-type semiconductor
- P-type: A P-type material is one in which holes are majority carriers i.e. they are positively charged materials (++++)
- N-type: A N-type material is one in which electrons are majority charge carriers i.e. they are negatively charged materials (-----)

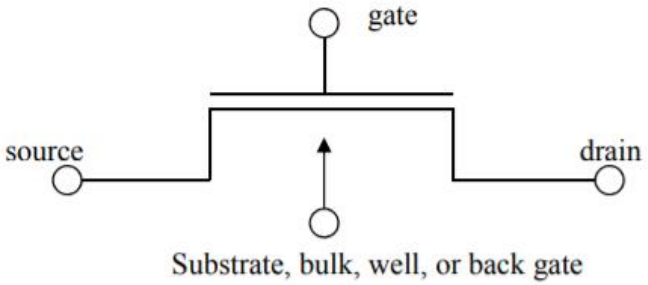
MOSFET Physical View

- Physical Structure of a MOSFET Device

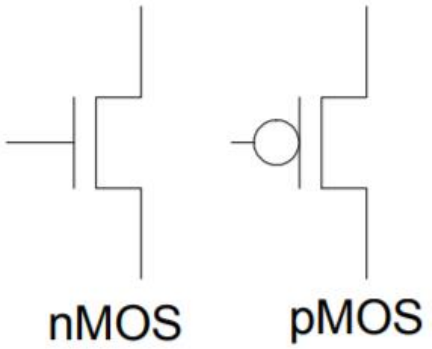


critical dimension = "feature size"

- Schematic Symbol for 4-terminal MOSFET



- Simplified Symbols



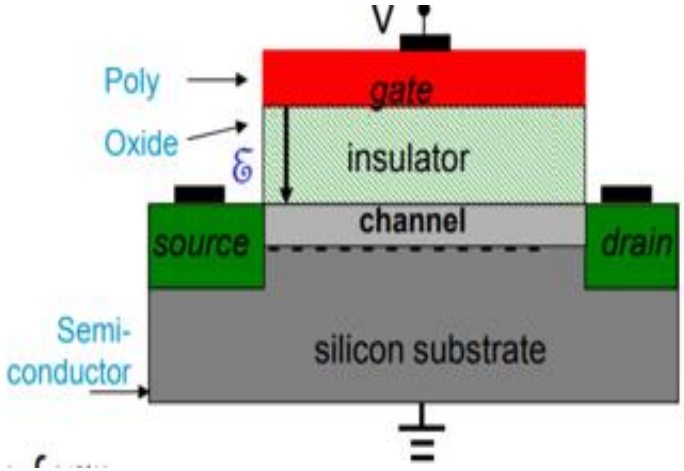
“Electronics” Building block(s)

- MOSFET Device-- 1950+ to 2020

- New elements in nano technologies are emerging. These include:
  - Fin-Transistor
  - Memristor: memory resistor- see IEEE Spectrum
  - Nano-tubes
  - Molecular devices
  - Quantum dots
  - Etc.

What is a MOSFET?

- Digital integrated circuits rely on transistor switches
  - – most common device for digital and mixed signal: MOSFET
- Definitions
  - **MOS = Metal Oxide Semiconductor**
    - physical layers of the device
  - **FET = Field Effect Transistor**
    - What field? What does the field do?
    - Are other fields important? – CMOS =
  - **Complementary MOS**
    - use of both nMOS and pMOS to form a circuit with lowest power consumption



Primary Features

– gate; gate oxide (insulator)

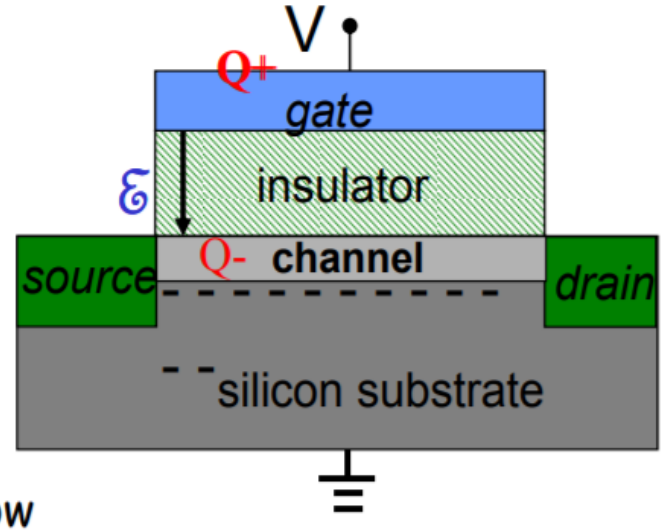
– source and drain

– channel

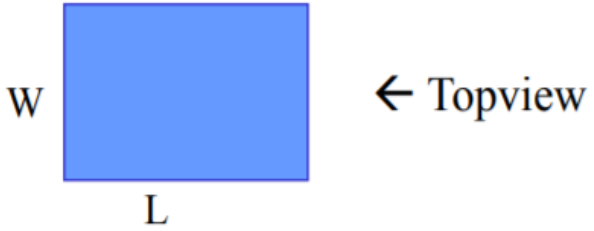
– bulk/substrate

Fundamental Relations in MOSFET

- Electric Fields
  - fundamental equation
    - electric field: E = V/d
  - vertical field through gate oxide
    - determines charge induced in channel
  - horizontal field across channel
    - determines source-to-drain current flow



- Capacitance
  - fundamental equations
    - capacitor charge: Q = CV
    - capacitance: C = ε A/d
  - charge balance on capacitor, Q+ = Q-
    - charge on gate is balanced by charge in channel
    - what is the source of channel charge? where does it come from?



CMOS Circuit Basics

CMOS = complementary MOS

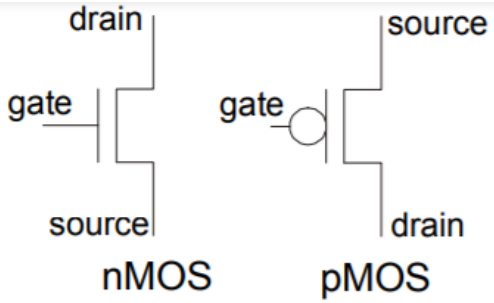
- uses 2 types of MOSFETs to create logic functions
  - nMOS
  - pMOS

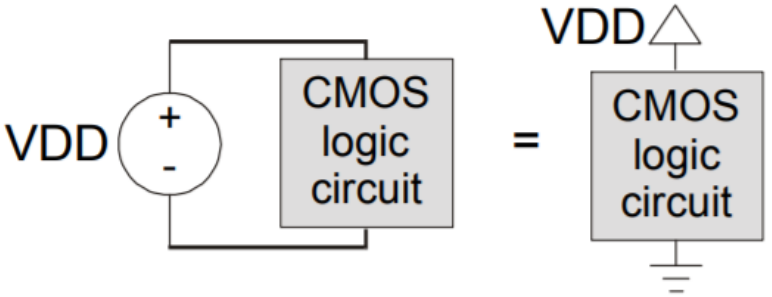
CMOS Power Supply

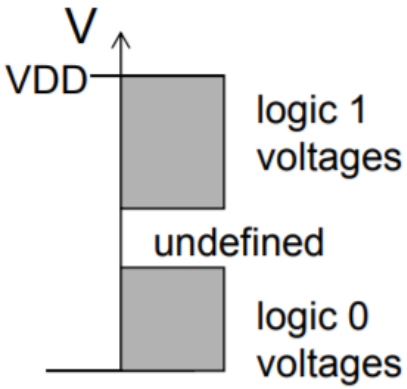
- typically single power supply
- **VDD**, with Ground reference
  - typically uses single power supply
  - VDD ranges from (0.6V) 1V to 5V

Logic Levels (voltage-based)

- all voltages between 0V and VDD
- Logic '1' = VDD
- Logic '0' = ground = 0V







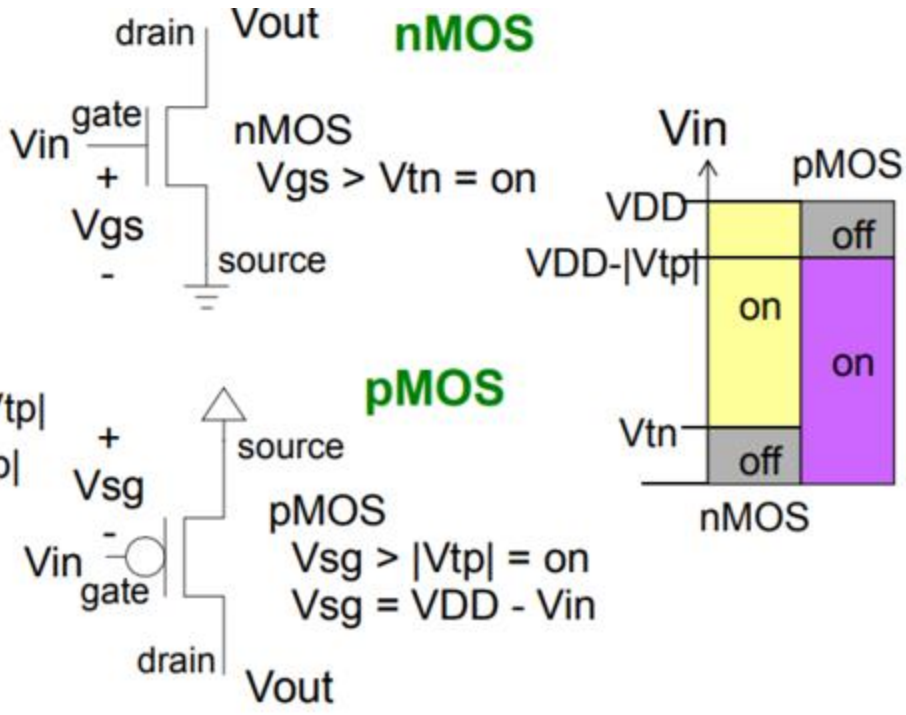
Transistor Switching Characteristics

- nMOS
  - switching behavior
    - on = closed, when Vin > Vtn
    - off = open, when Vin < Vtn
- pMOS
  - switching behavior
    - on = closed, when Vin < VDD - |Vtp|
    - off = open, when Vin > VDD - |Vtp|
- Digital Behavior
  - nMOS

Vin	Vout (drain)	
1	Vs=0	device is ON
0	?	device is OFF

  - pMOS

Vin	Vout (drain)	
1	?	device is OFF
0	Vs=VDD=1	device is ON



Rule to Remember

'source' is at

- lowest potential for nMOS
- highest potential for pMOS

MOSFET Pass Characteristics

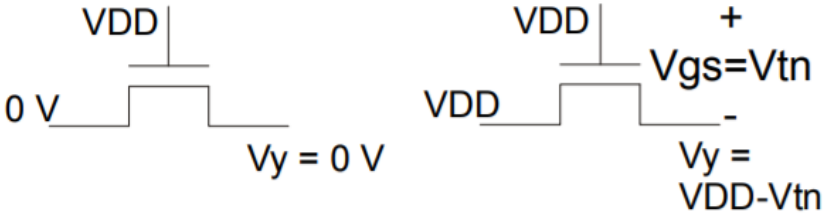
- Each type of transistor is better at passing (to output) one digital voltage than the other
  - nMOS passes a good low (0) but not a good high (1)
  - pMOS passes a good high (1) but not a good low (0)

TABLE 1.1 The Output Logic Levels of N-SWITCHES and P-SWITCHES

LEVEL	SYMBOL	SWITCH CONDITION
Strong 1	1	P-SWITCH gate = 0, source = V_{DD}
Weak 1	1	N-SWITCH gate = 1, source = V_{DD} or P-SWITCH connected to V_{DD}
Strong 0	0	N-SWITCH gate = 1, source = V_{SS}
Weak 0	0	P-SWITCH gate = 0, source = V_{SS} or N-SWITCH connected to V_{SS}
High impedance	Z	N-SWITCH gate = 0 or P-SWITCH gate = 1

nMOS

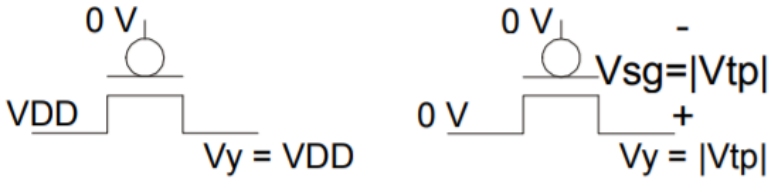
on when gate is 'high'



Passes a good low
Max high is VDD-Vtn

pMOS

on when gate is 'low'



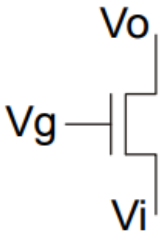
Passes a good high
Min low is |Vtp|

Rule to Remember

'source' is at lowest potential (nMOS) and highest potential (pMOS)

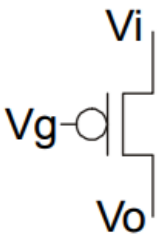
MOSFET Terminal Voltages

- How do you determine one terminal voltage if other 2 are known?
  - nMOS
    - case 1) if Vg > Vi + Vtn, then Vo = Vi (Vg-Vi > Vtn)
      - here Vi is the "source" so the nMOS will pass Vi to Vo
    - case 2) if Vg < Vi + Vtn, then Vo = Vg-Vtn (Vg-Vi < Vtn)
      - here Vo is the "source" so the nMOS output is limited
    - Example (Vtn=0.5V): Vg=5V, Vi=2V ⇒ Vo = 2V
      Vg=2V, Vi=2V ⇒ Vo = 1.5V



For nMOS,
max(Vo) = Vg-Vtn

  - pMOS
    - case 1) if Vg < Vi - |Vtp|, then Vo = Vi (Vi-Vg > |Vtp|)
      - here Vi is the "source" so the pMOS will pass Vi to Vo
    - case 2) if Vg > Vi - |Vtp|, then Vo = Vg+|Vtp| (Vi-Vg < |Vtp|)
      - here Vo is the "source" so the pMOS output is limited
    - Example (Vtp=-0.5V): Vg=2V, Vi=5V ⇒ Vo = 5V
      Vg=2V, Vi=2V ⇒ Vo = 2.5V



For pMOS,
min(Vo) = Vg+|Vtp|

Switch-Level Boolean Logic

- Logic gates are created by using sets of controlled switches
- Characteristics of an **assert-high** switch

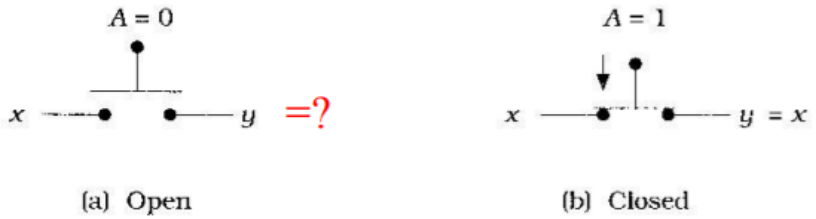


Figure 2.1 Behavior of an assert-high switch

nMOS acts like an **assert-high** switch

– $y = x \cdot A$, i.e. y = x iff A = 1 (iff=if and only if)

Series switches ⇒ AND function

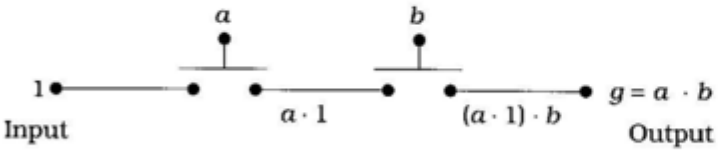


Figure 2.2 Series-connected switches

Parallel switches ⇒ OR function

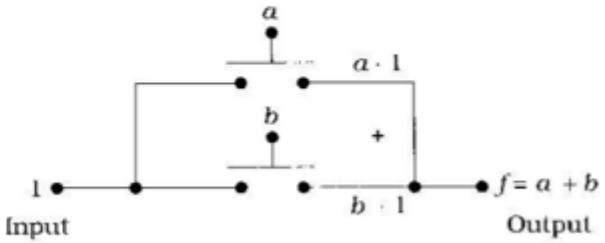
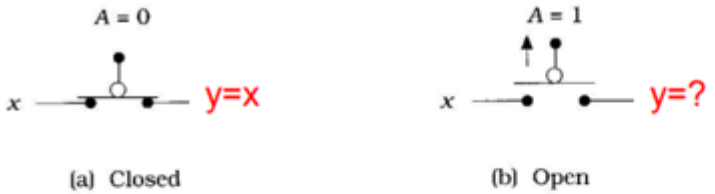


Figure 2.4 Parallel-connected switches

Switch-Level Boolean Logic

- Characteristics of an **assert-low** switch

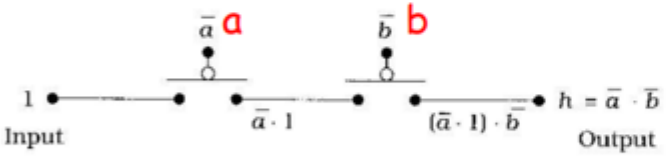


pMOS acts like an **assert-low** switch

– $y = x \cdot \overline{A}$, i.e. y = x if A = 0

error in figure 2.5

Series assert-low switches ⇒ ?



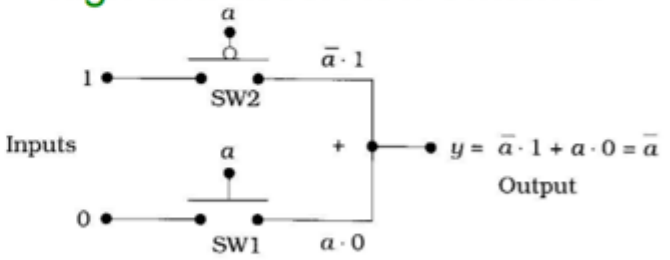
NOR

Remember This??

$\overline{a} \cdot \overline{b} = \overline{a + b}, \quad \overline{a} + \overline{b} = \overline{a \cdot b}$

DeMorgan relations

NOT function, combining assert-high and assert-low switches

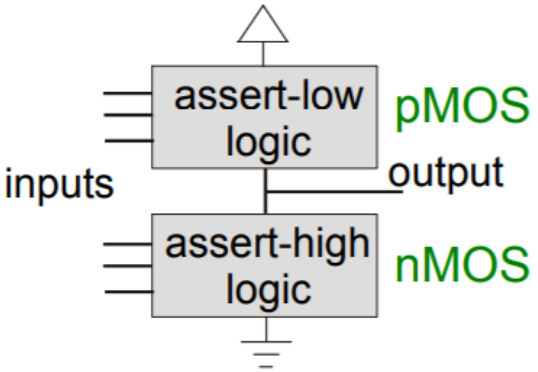


a=1 ⇒ SW1 closed, SW2 open ⇒ y=0 = $\overline{a}$

a=0 ⇒ SW1 open, SW2 closed ⇒ y=1 = a

CMOS "Push-Pull" Logic

- CMOS Push-Pull Networks
  - pMOS
    - "on" when input is low
    - pushes output high
  - nMOS
    - "on" when input is high
    - pulls output low



- only one logic network (p or n) is required to produce (1/2-) the logic function???
- but the complementary set allows the "load" to be turned off for zero **static** power dissipation

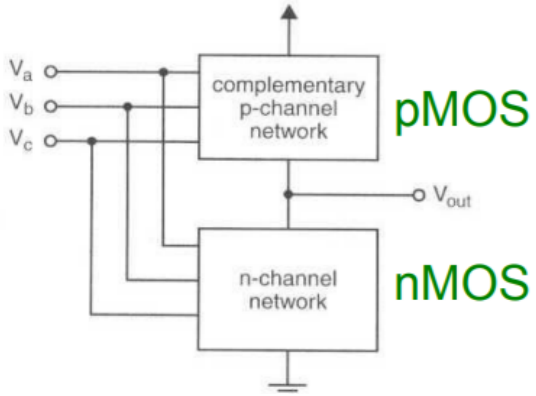
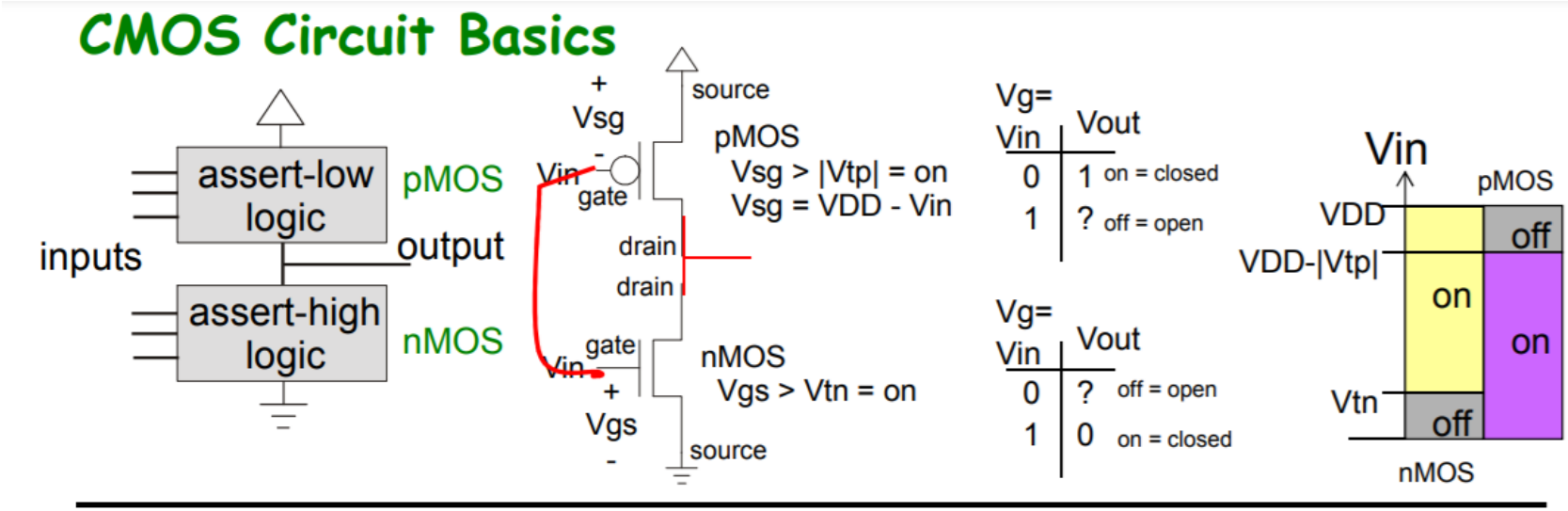


TABLE 1.1 The Output Logic Levels of N-SWITCHES and P-SWITCHES

LEVEL	SYMBOL	SWITCH CONDITION
Strong 1	**1**	P-SWITCH gate = 0, source = V_{DD}
Weak 1	1	N-SWITCH gate = 1, source = V_{DD} or P-SWITCH connected to V_{DD}
Strong 0	**0**	N-SWITCH gate = 1, source = V_{SS}
Weak 0	0	P-SWITCH gate = 0, source = V_{SS} or N-SWITCH connected to V_{SS}
High impedance	Z	N-SWITCH gate = 0 or P-SWITCH gate = 1

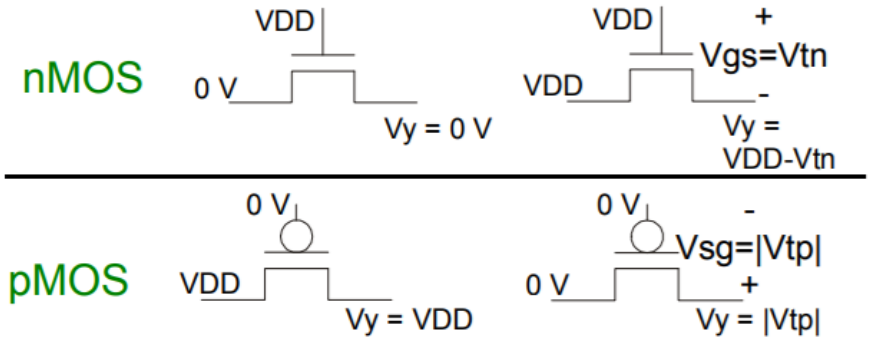
VSS = ground

Review: Basic Transistor Operation



CMOS Pass Characteristics

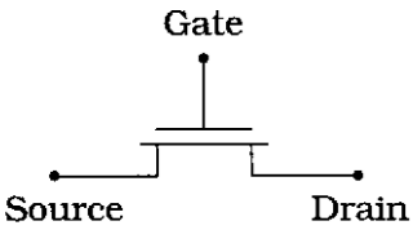
'source' is at lowest potential (nMOS) and highest potential (pMOS)



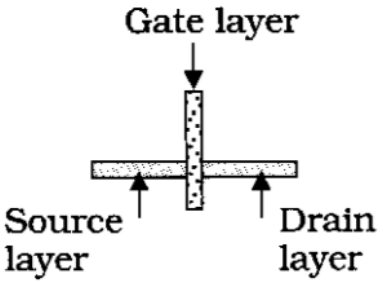
- nMOS
  - 0 in = 0 out
  - VDD in = VDD-Vtn out
  - strong '0', weak '1'
- pMOS
  - VDD in = VDD out
  - 0 in = |Vtp| out
  - strong '1', weak '0'

Physical MOSFET Switch

- nMOS Switch

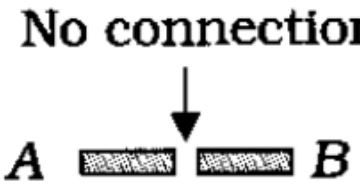


(a) nFET symbol

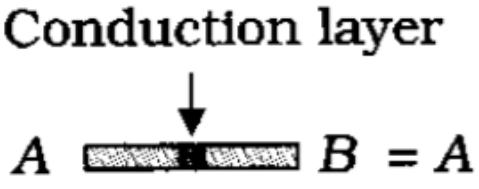


(b) nFET layers

- Physical Switching Operation Layers



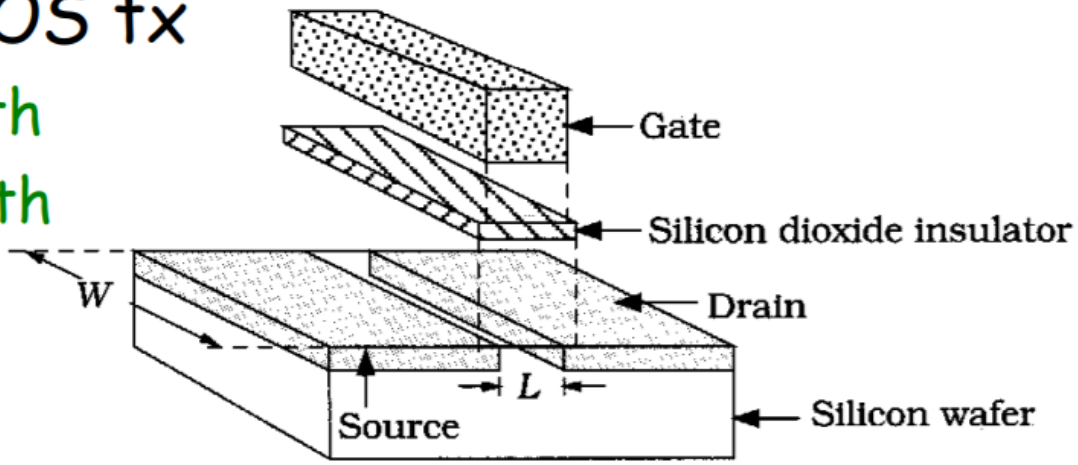
(a) Open switch



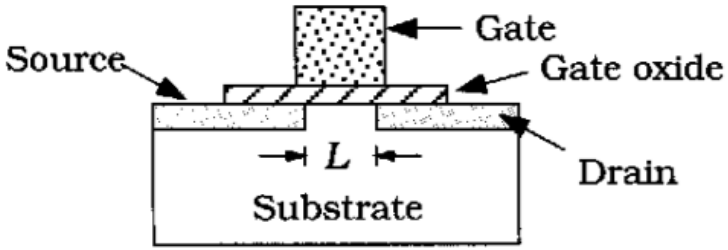
(b) Closed switch

nMOS Layers and Layout

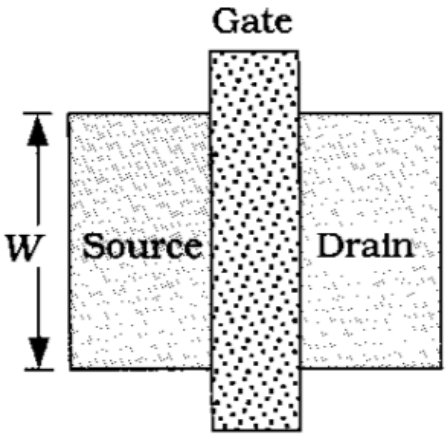
- Layers of an nMOS tx
  - L = channel length
  - W = channel width
  - gate oxide
    - separates gate from substrate



- Side and Top views



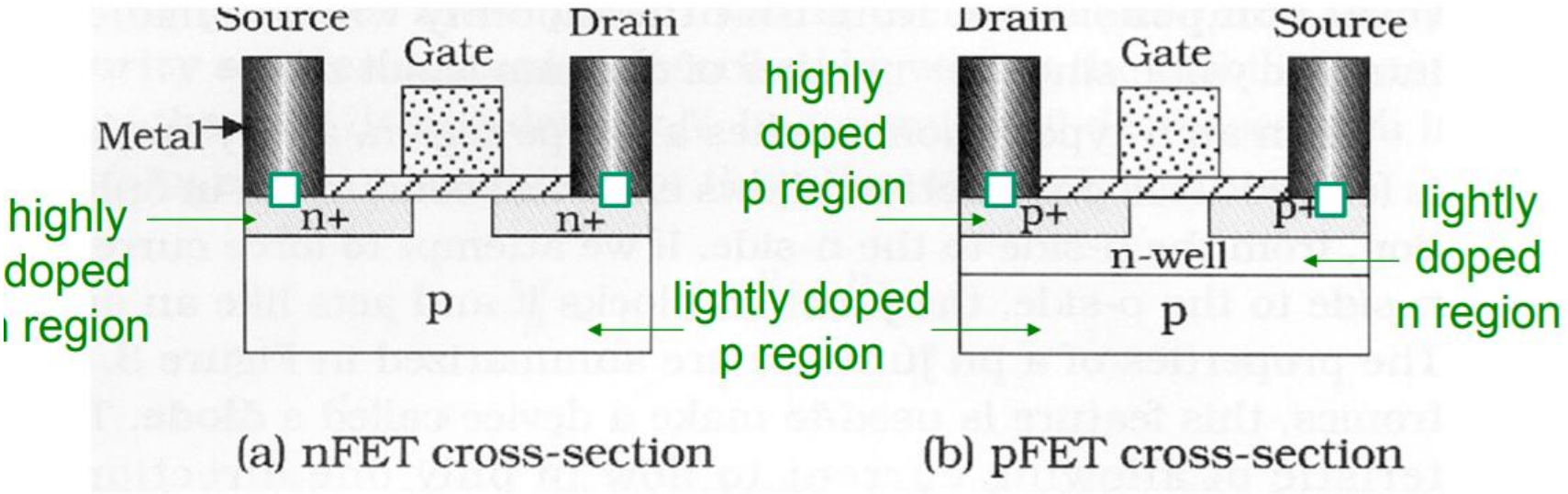
(a) Side view



(b) Top view

Physical n/pMOS Devices

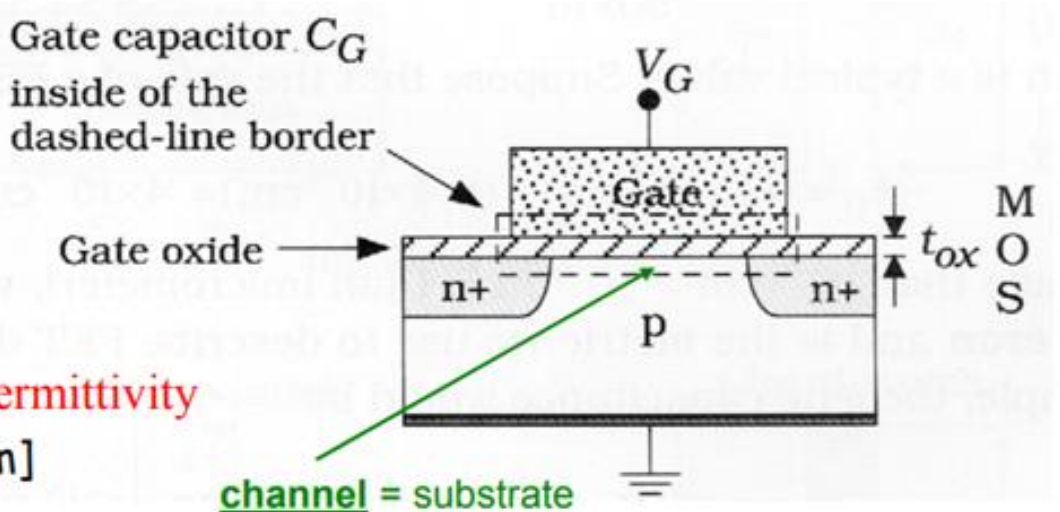
- nMOS and pMOS cross-section

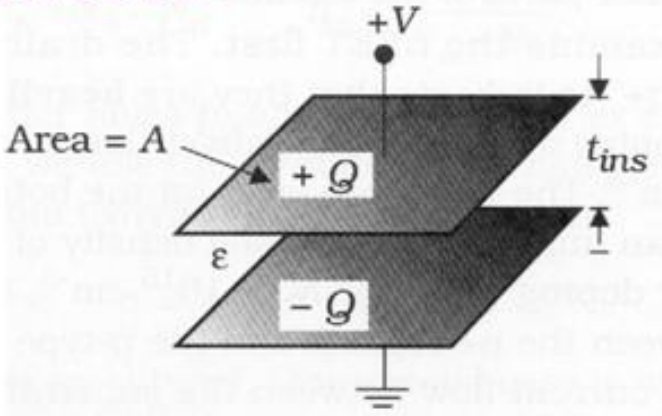


- Layers
  - substrate, n-well, n+/p+ S/D, gate oxide, polysilicon gate, S/D contact, S/D metal
- Can you find all of the diodes (pn junctions)?
  - where? conduct in which direction? what purpose?

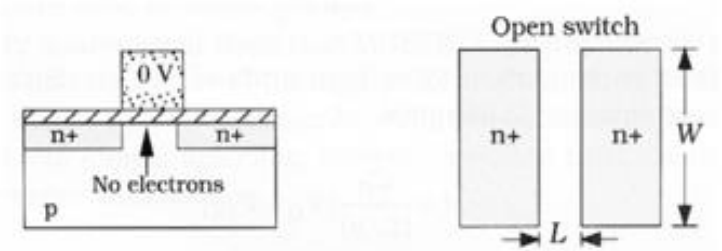
MOSFET Gate Operation

- Gate Capacitance
  - gate-substrate parallel plate capacitor
  - $C_G = \varepsilon_{ox} A / t_{ox}$ [F]
    - $\varepsilon_{ox} = 3.9\ \varepsilon_o$ ← permittivity
    - $\varepsilon_o = 8.85 \times 10^{-14}$ [F/cm]
- Oxide Capacitance
  - $C_{ox} = \varepsilon_{ox} / t_{ox}$ [F/cm²]
  - $C_G = C_{ox} A_G$ [F]
    - A_G=gate area = L·W [cm²]
- Charge on Gate, +Q, induces charge -Q in substrate channel
  - channel charge allows conduction between source and drain*

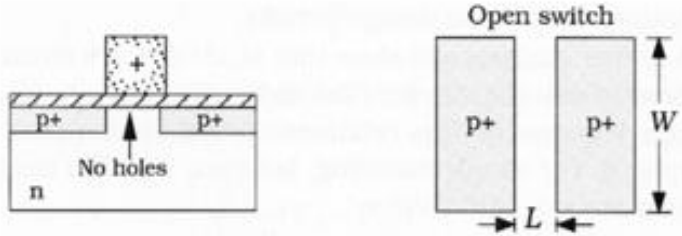




Physical Switching in nMOS and pMOS

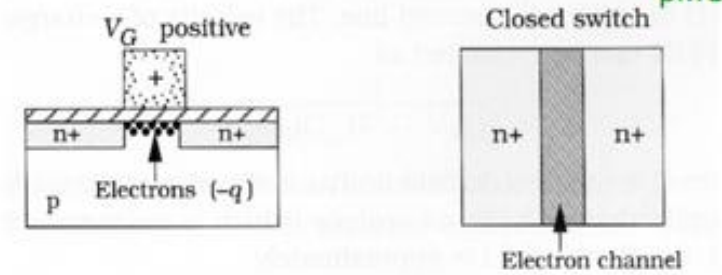


(a) Zero gate voltage

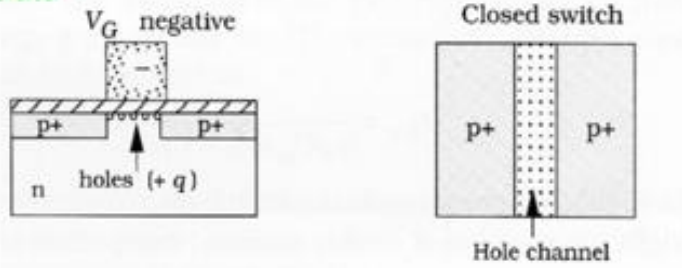


(a) High gate voltage

notice,
nMOS in p-substrate,
pMOS in n-substrate



(b) Positive gate voltage



(b) Negative gate voltage

- nMOS
  - zero or negative Q on gate
    - no charge in channel
  - positive Q on gate
    - negative charge (e^-) in channel
    - conduction path between n+ S/D
- pMOS
  - positive Q on gate
    - no charge in channel
  - negative Q on gate ($V_G < V_S$)
    - positive charge (h^+) in channel
    - conduction path between p+ S/D

Channel Charge and Current

- Threshold Voltage = Vtn, Vtp
  - amount of voltage required on the gate to turn tx on
  - gate voltage > Vtn/p will induce charge in the channel
- nMOS Channel Charge
  - $Qc = -C_G(V_G - Vtn)$, from Q=CV, (-) because channel holds electrons
- nMOS Channel Current (linear model)
  - $I = |Qc| / t_t$, where t_t = *transit time*, average time to cross channel
    - t_t = channel length / (average velocity) = L / v
    - average drift velocity in channel due to electric field $E \rightarrow v = \mu_n E$
    - assuming constant field in channel due to $V_{DS} \rightarrow E = V_{DS} / L$
    - $\rightarrow$ $$I = Qc\frac{\mu_n \frac{V_{DS}}{L}}{L} \qquad C_G = CoxWL \Rightarrow |Qc| = CoxWL(V_G - Vtn)$$
  - $I = \mu_n Cox\,(W/L)\,(V_G - Vtn)\,V_{DS}$ linear model, assumes constant charge in channel

similar analysis applies for pMOS, see textbook

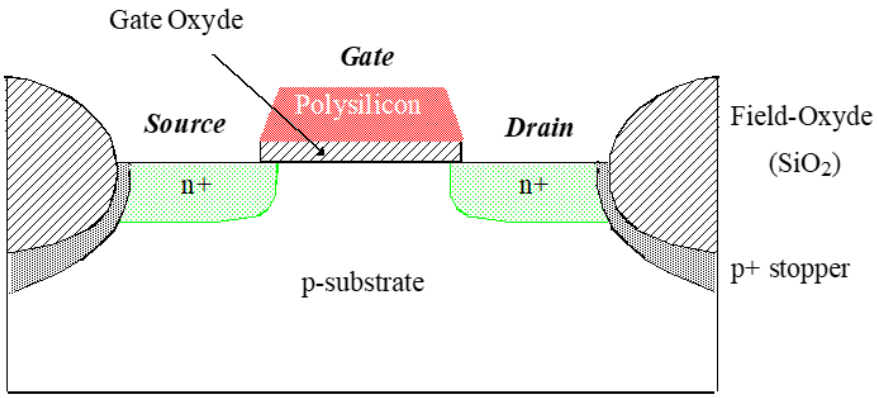
Transconductance and Channel Resistance

- nMOS Channel Charge: $Qc = -C_G(V_G - Vtn)$
- nMOS linear model Channel Current:
  - $I = \mu_n Cox(W/L)(V_G - Vtn)\, V_{DS}$
    - assumes constant charge in channel
    - valid only for very small V_{DS}
- nMOS Transconductance
  - $\beta_n = \mu_n Cox\,(W/L)\ [A/V^2] \Rightarrow I = \beta_n\,(V_G - Vtn)\, V_{DS}$
  - constant for set transistor size and process
- nMOS Channel Resistance
  - channel current flows between Drain and Source
  - channel resistance = V_{DS} / I_{DS}
  - $Rn = 1/(\beta_n\,(V_G - Vtn))$

$$Rn = \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_G - V_{tn})}$$

similar analysis applies for pMOS, see textbook

The MOS Transistor/SPICE Parameters



CROSS-SECTION of NMOS Transistor

Parameter Name	Symbol	SPICE Name	Units	Default Value
Saturation current	I_S	IS	A	1.0 E-14
Emission coefficient	n	N	-	1
Series resistance	R_S	RS	Ω	0
Transit time	τ_T	TT	sec	0
Zero-bias junction capacitance	C_{j0}	CJ0	F	0
Grading coefficient	m	M	-	0.5
Junction potential	ϕ_0	VJ	V	1

First Order SPICE diode model parameters.

Threshold Voltage: Concept

$$V_T = \phi_{ms} - 2\phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{SS}}{C_{ox}} - \frac{Q_I}{C_{ox}}$$

Workfunction Difference

Depletion Layer Charge

Surface Charge

Implants

Body Effect Coefficient

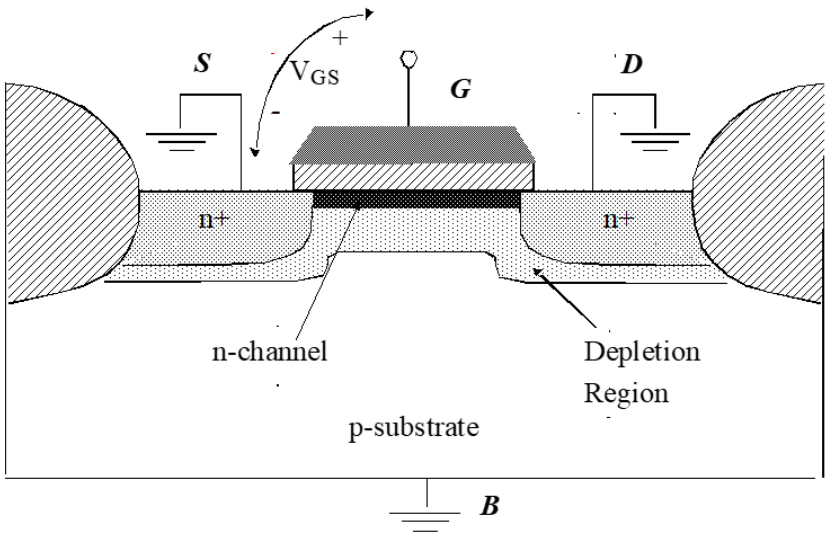
$$V_T = V_{T0} + \gamma\left(\sqrt{\left|-2\phi_F + V_{SB}\right|} - \sqrt{\left|-2\phi_F\right|}\right)$$

with

$$V_{T0} = \phi_{ms} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{SS}}{C_{ox}} - \frac{Q_I}{C_{ox}}$$

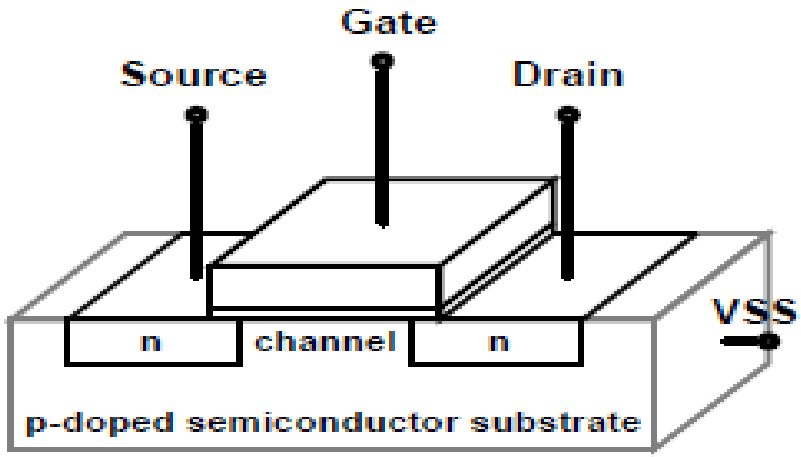
and

$$\gamma = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{ox}}$$

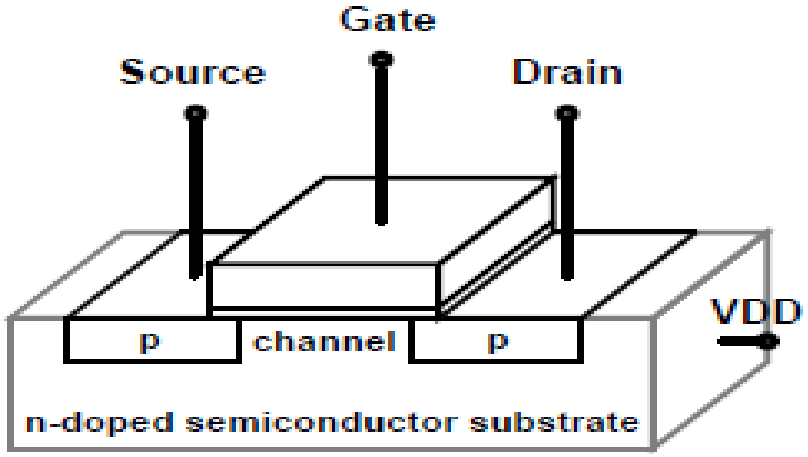


How Is It Done? (devices)

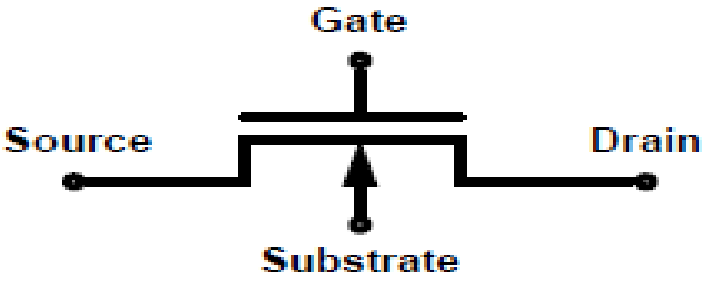
MOS Transistors:

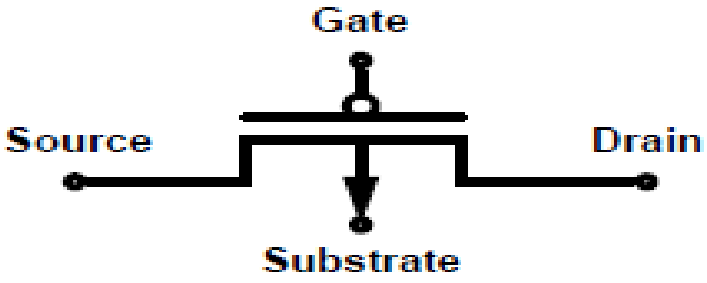


NMOS



PMOS

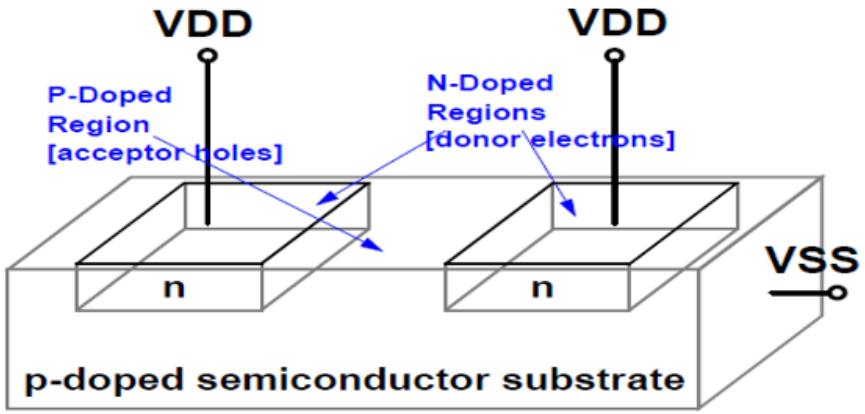


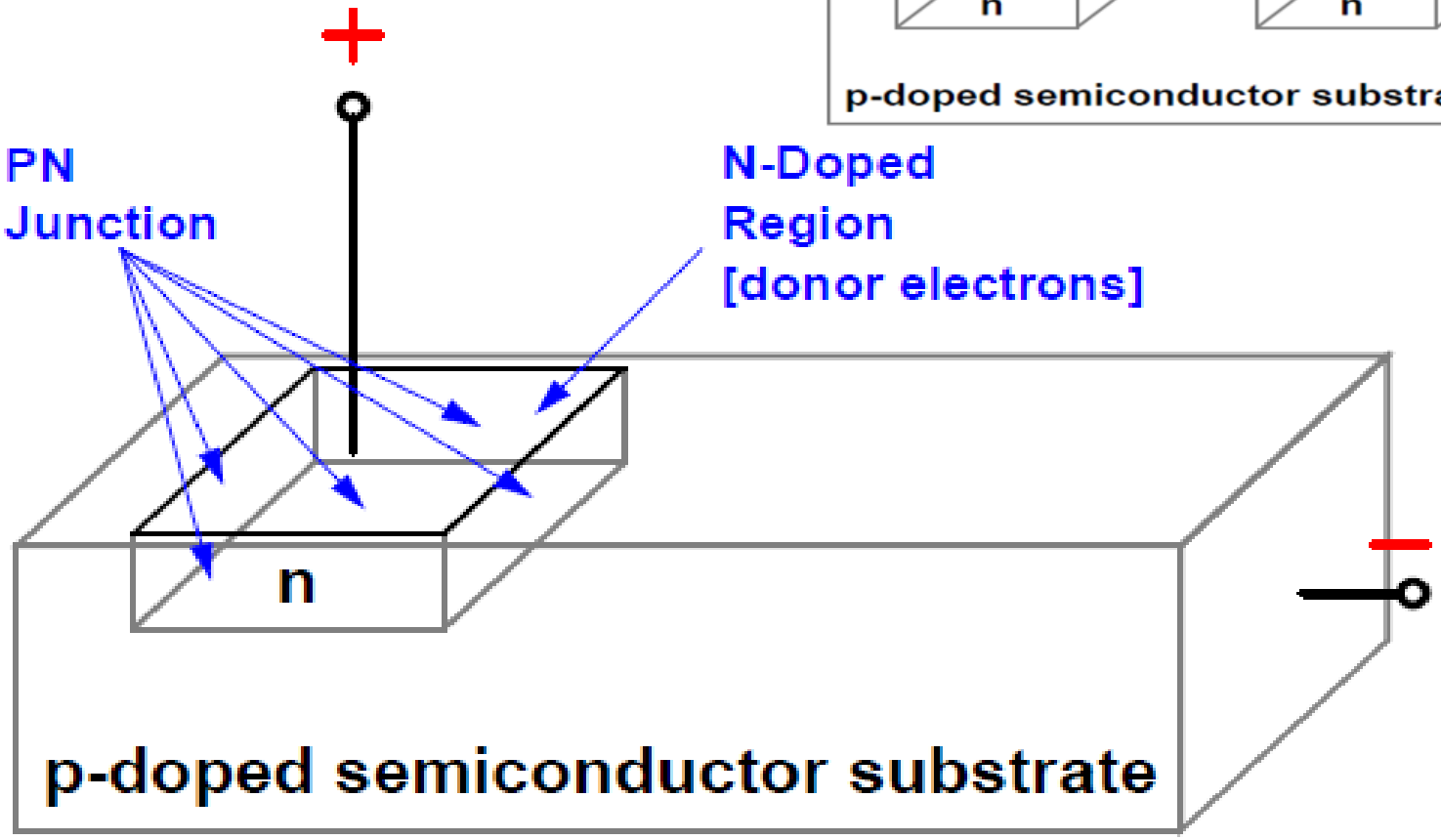


What's a "C" MOS?

How Is It Done? (devices)

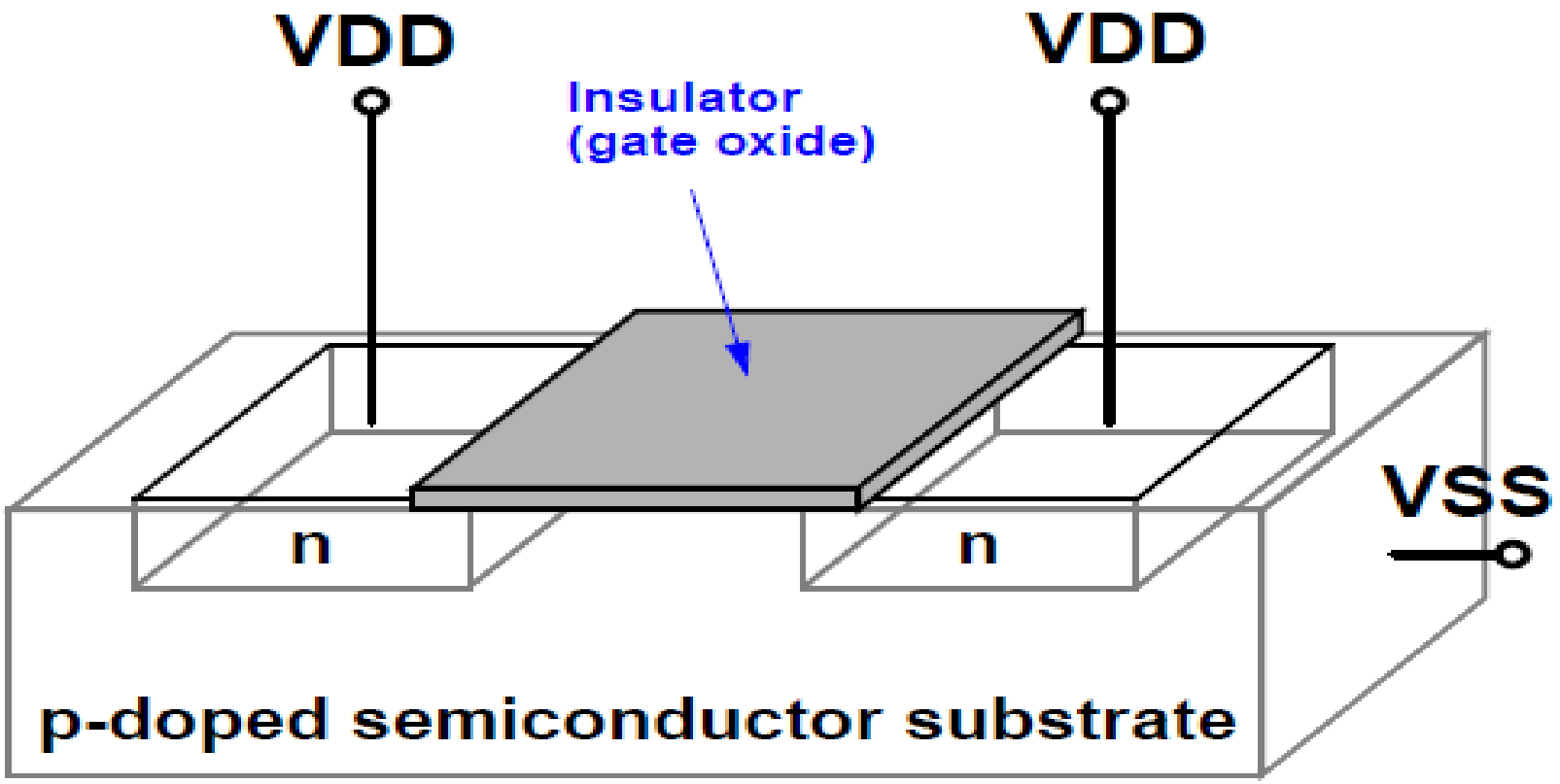
MOS Transistor:





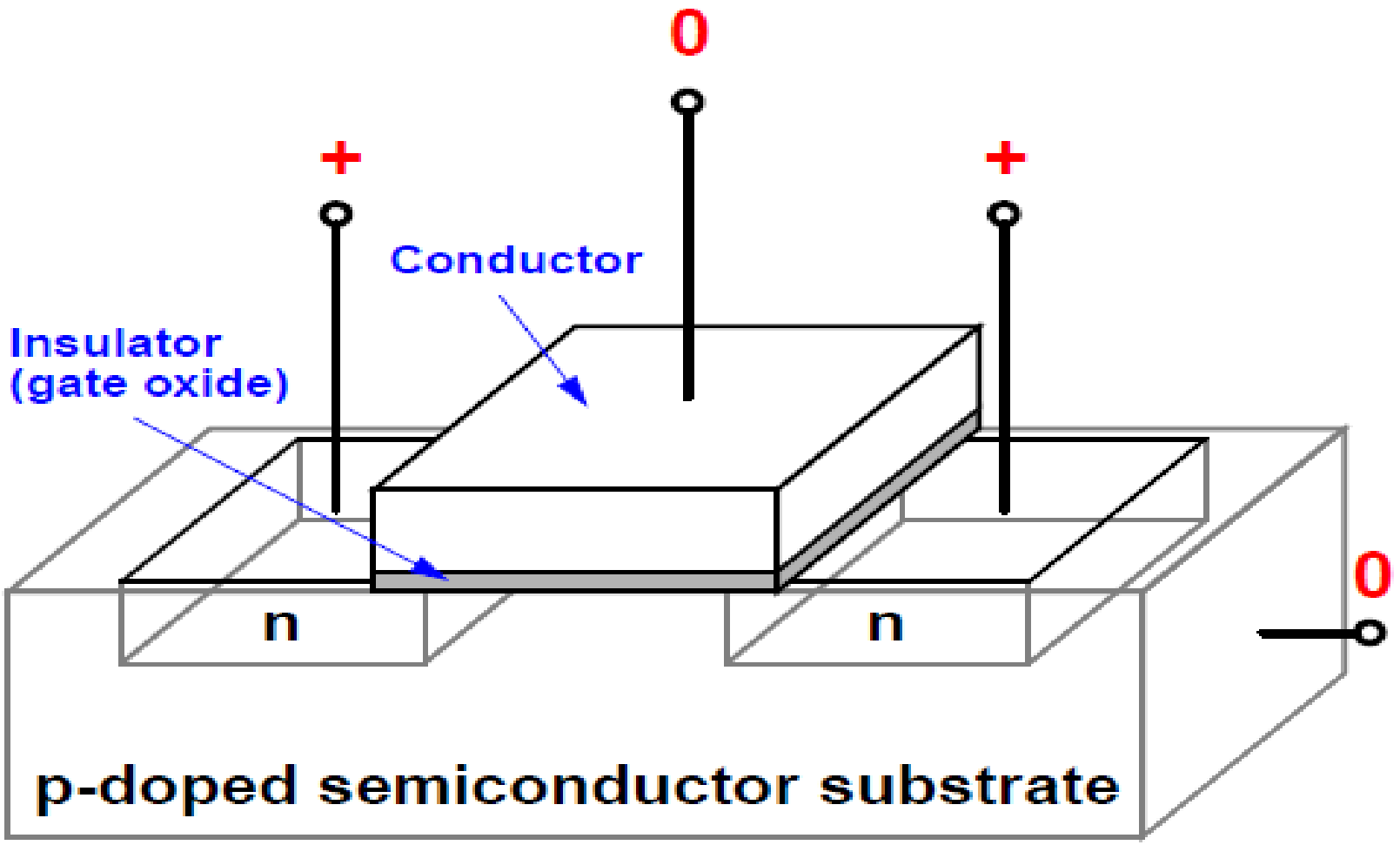
How Is It Done? (devices)

MOS Transistor:



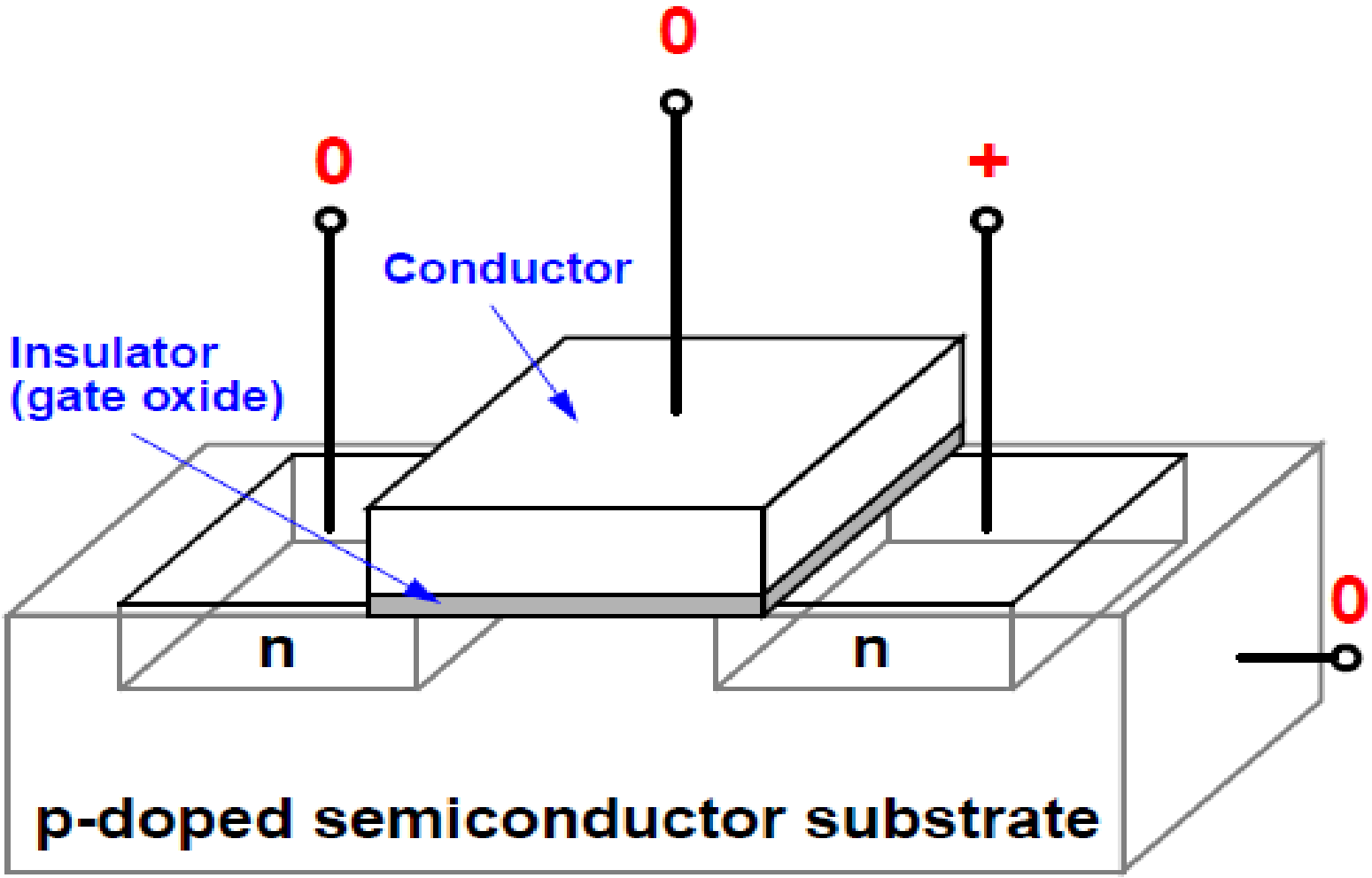
How Is It Done? (devices)

NMOS Transistor with gate:



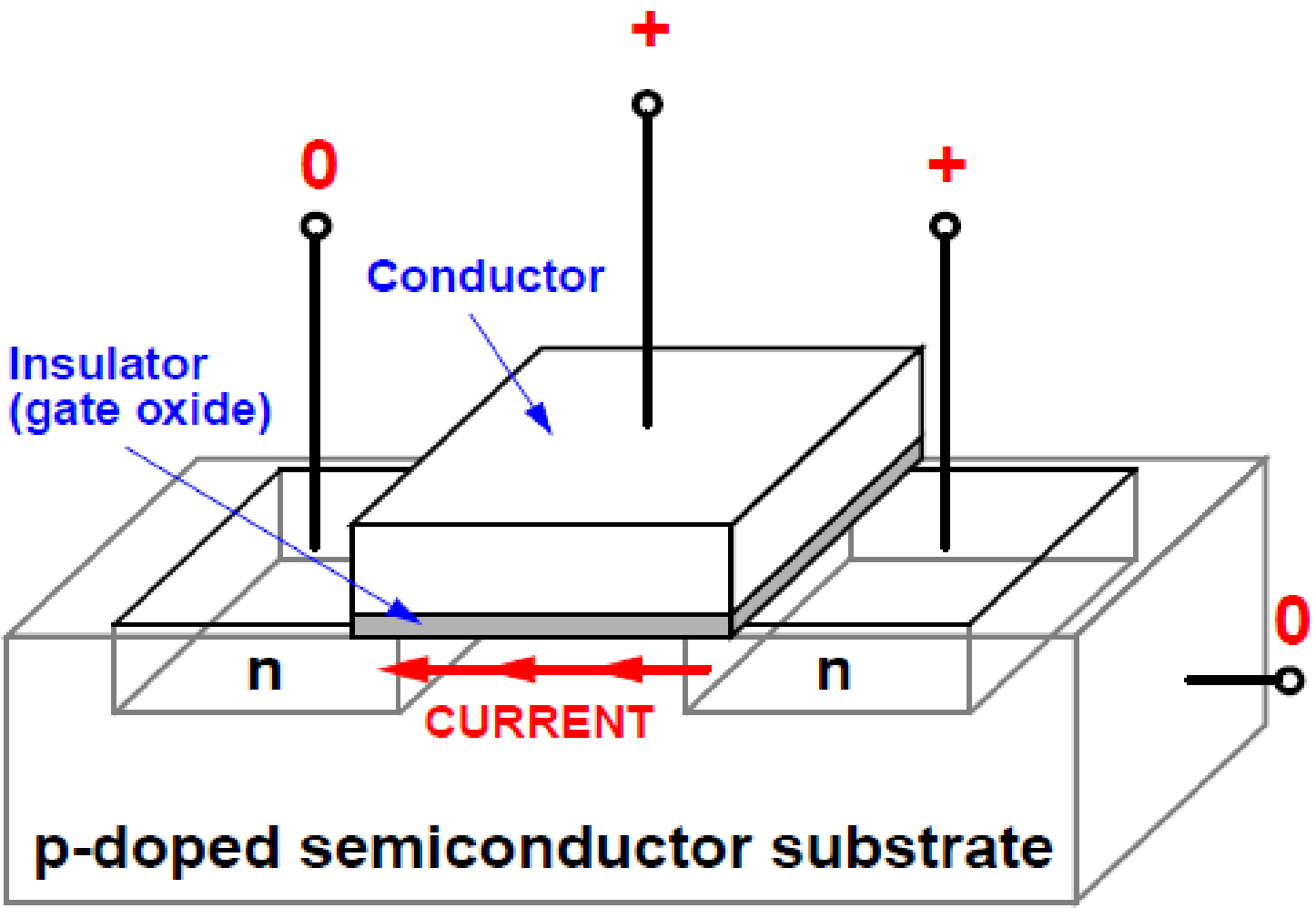
How Is It Done? (devices)

NMOS Transistor with bias voltages:



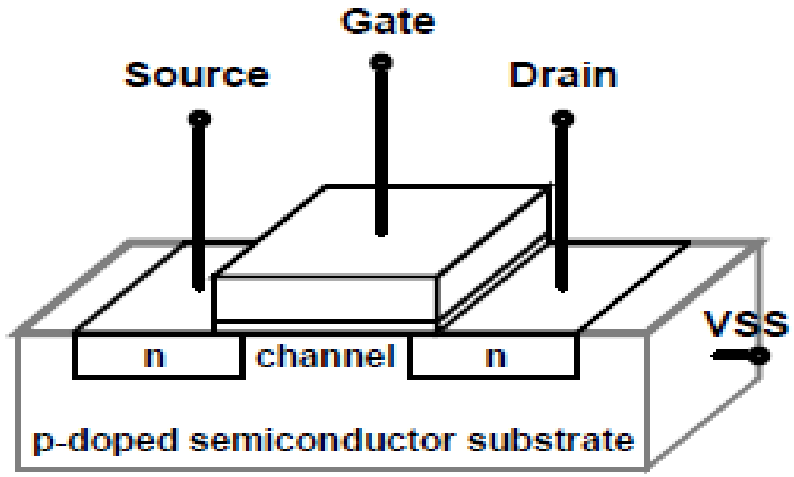
How Is It Done? (devices)

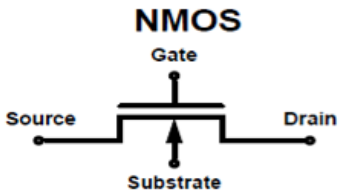
NMOS Transistor with bias voltages:

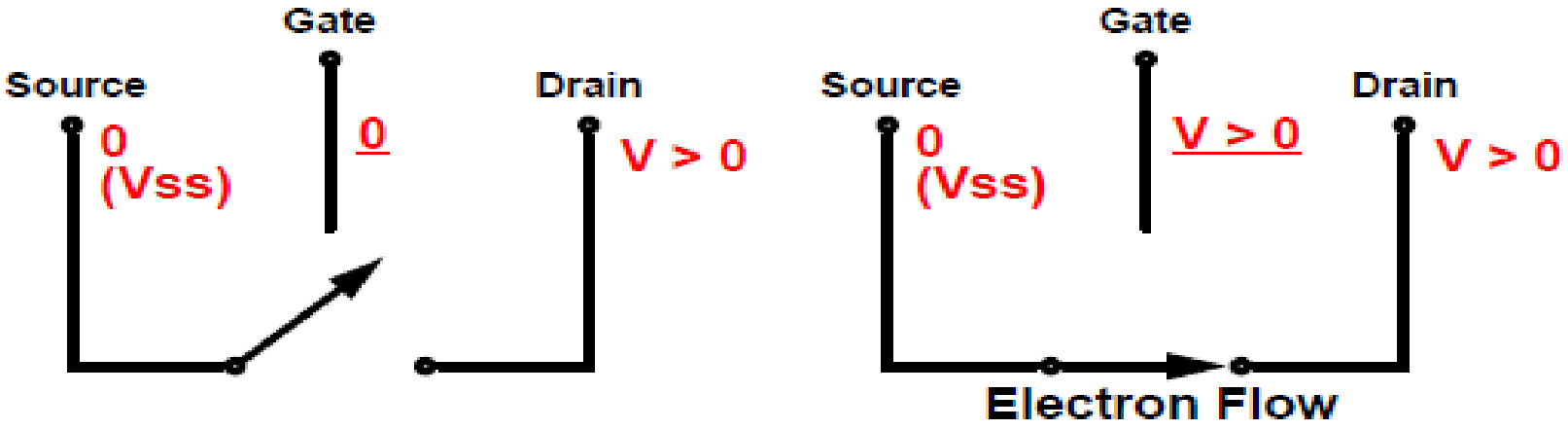


How Is It Done? (devices)

NMOS Transistor with bias voltages:

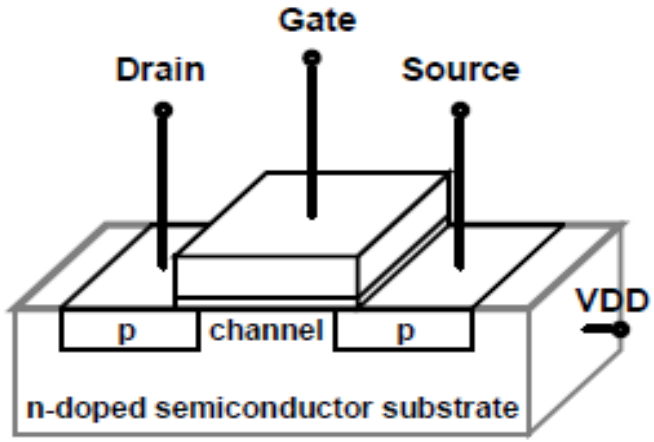


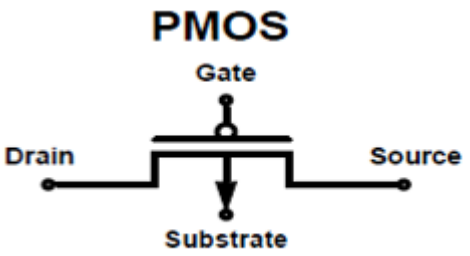


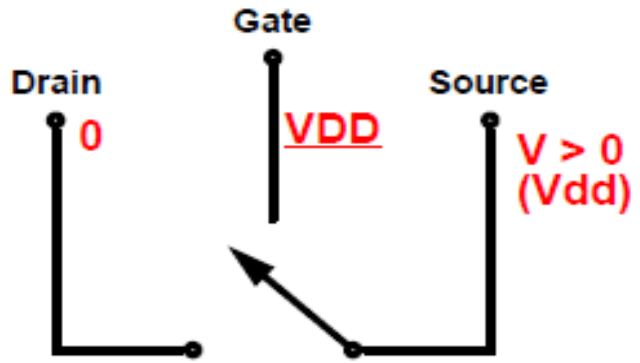


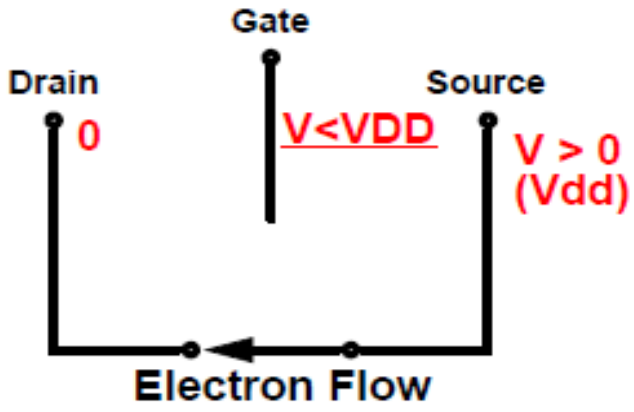
How Is It Done? (devices)

PMOS Transistor with bias voltages:



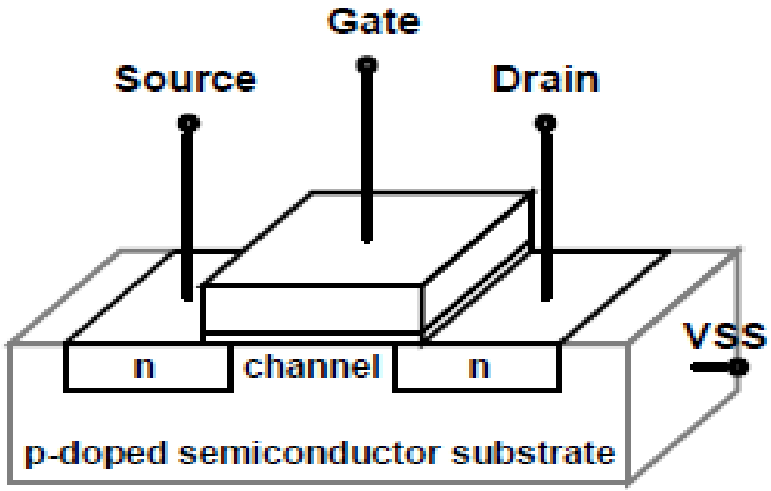




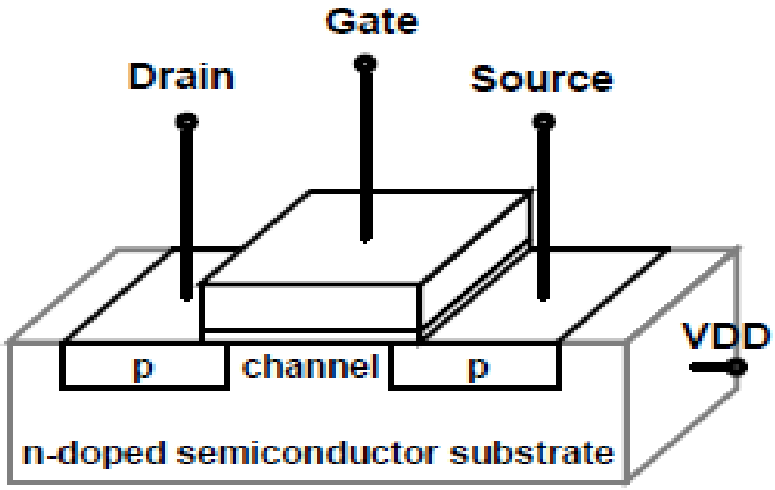


How Is It Done? (devices)

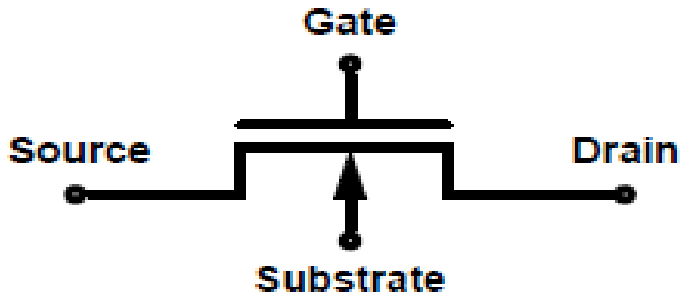
MOS Transistors:

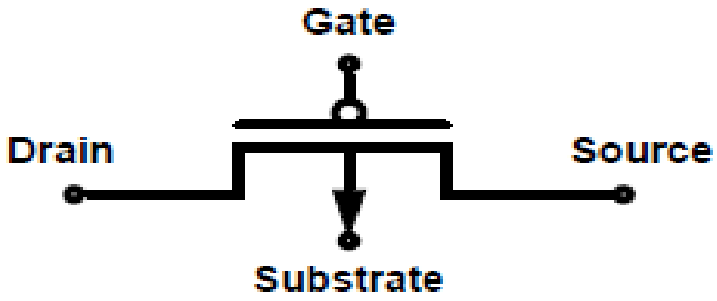


NMOS



PMOS

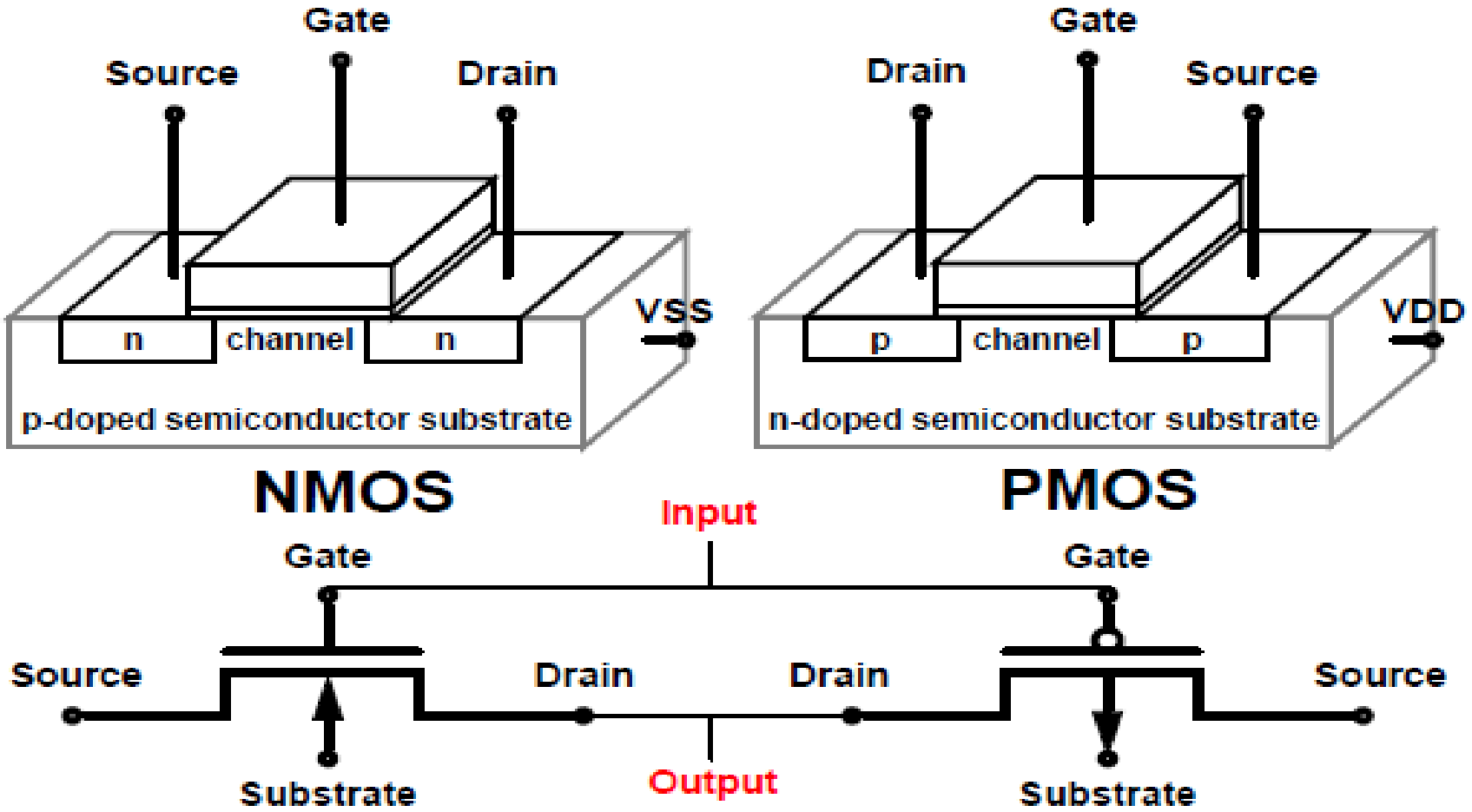




CMOS Inverter = one of each

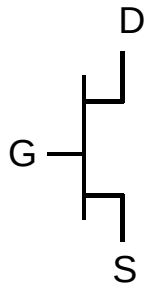
How Is It Done? (devices)

MOS Transistors:

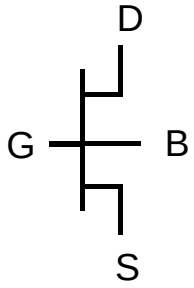


CMOS Inverter = one of each

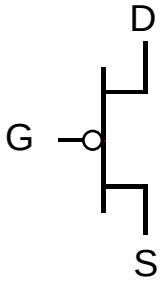
CMOS Transistor - Types and Symbols



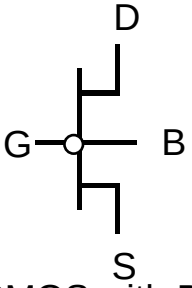
NMOS Enhancement



NMOS with Bulk Contact

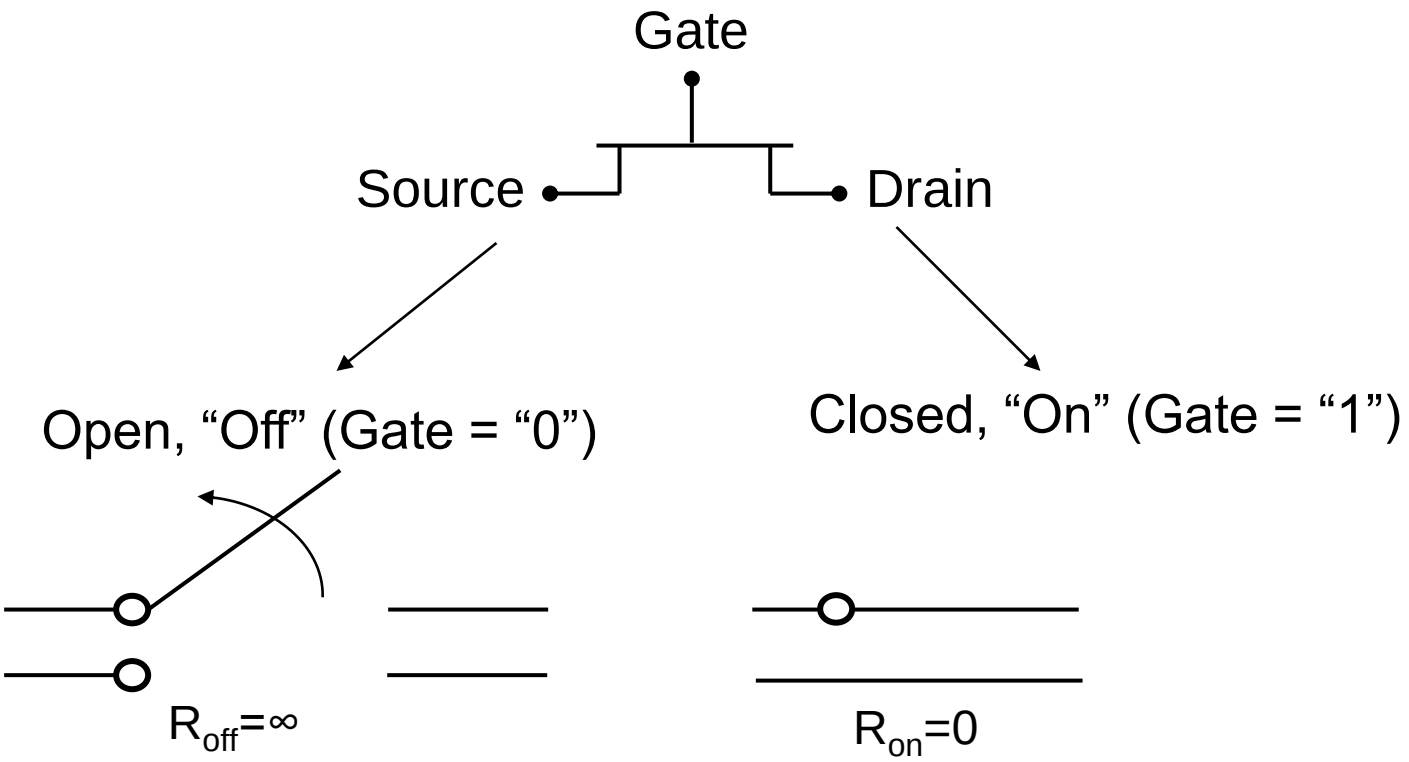


PMOS Enhancement



PMOS with Bulk Contact

Switch Model of NMOS Transistor

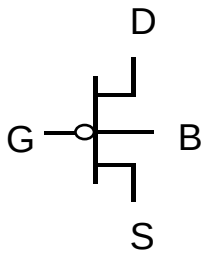


Gate voltage smaller than threshold voltage

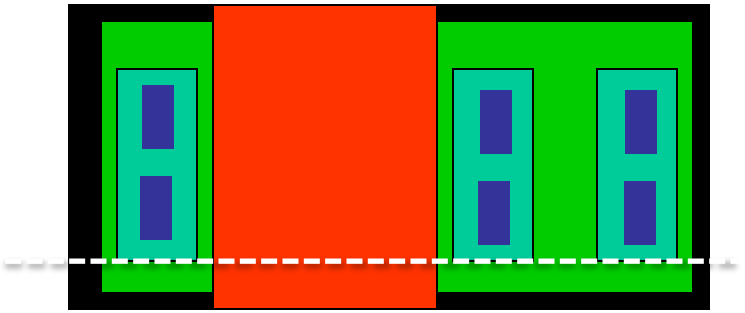
Gate voltage larger than threshold voltage

Concepts of the Circuit and Layout

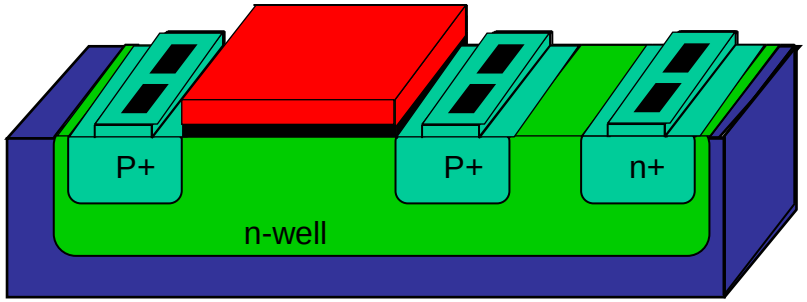
Circuit



Layout

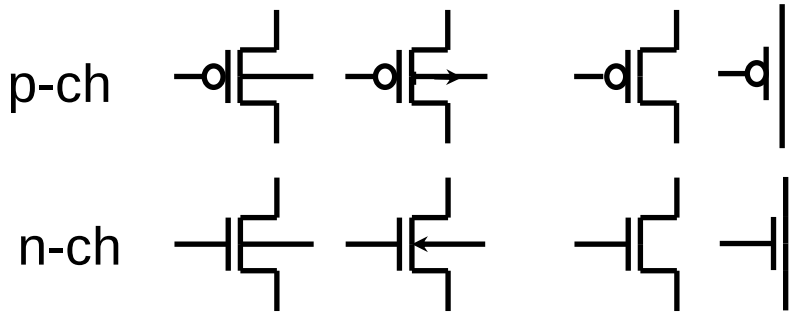


Resulting structure in manufactured IC

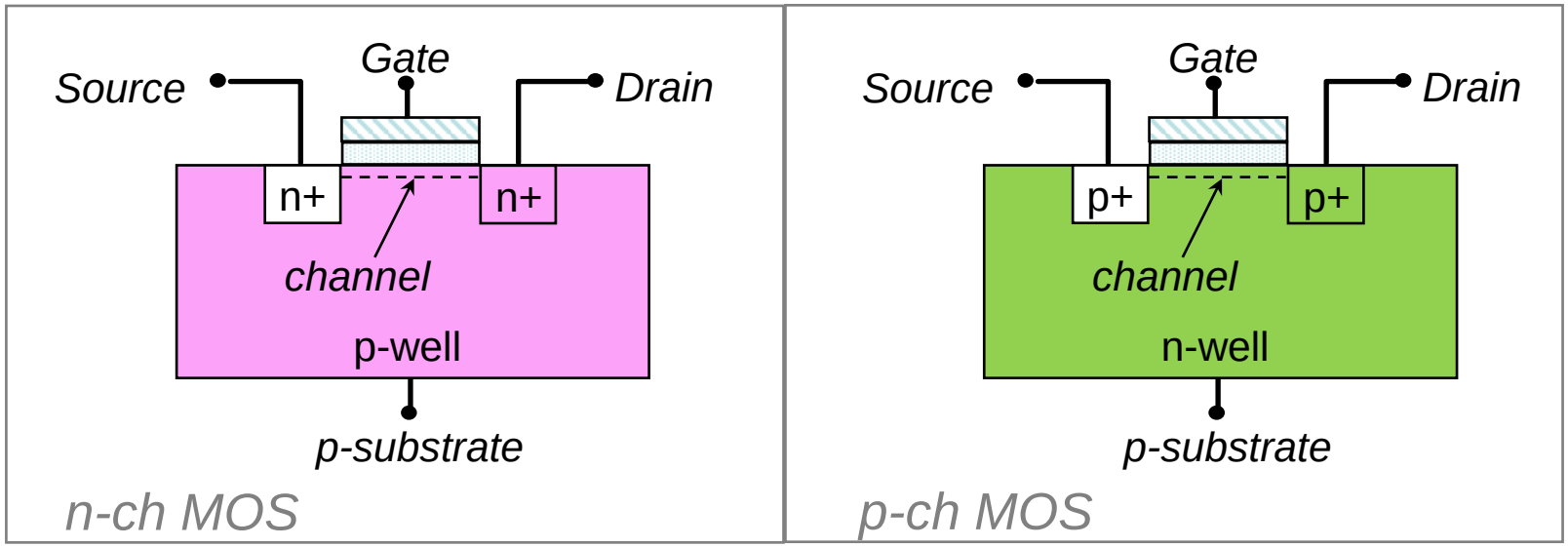


The MOS transistor

- Symbols:

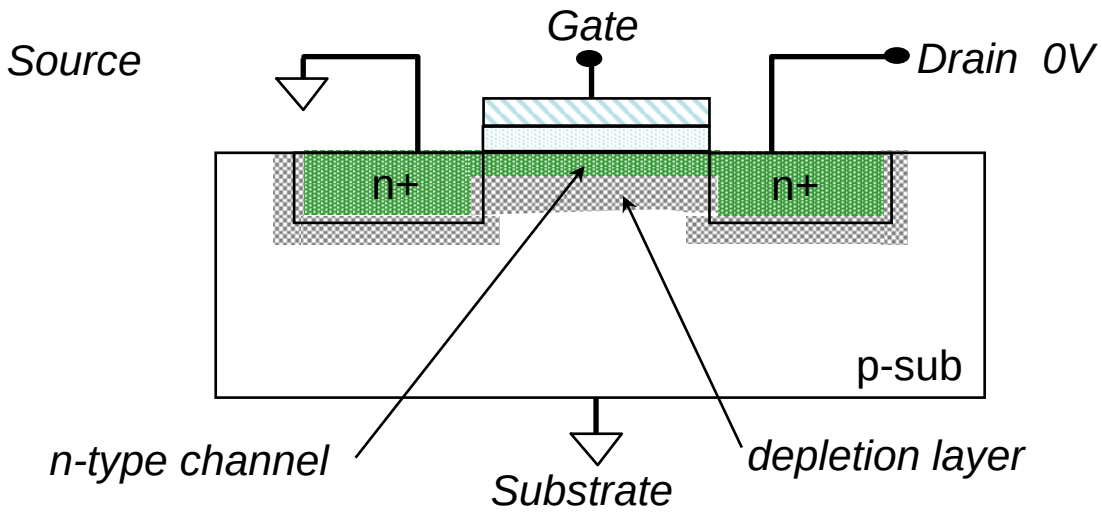


- Physical structure:

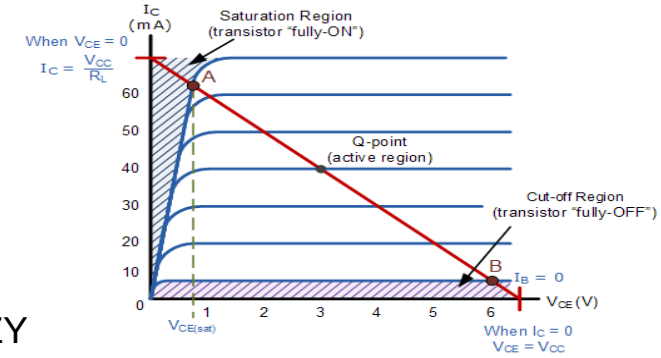


The MOS transistor - the different modes of operation

Vgs > Vt ; Vds=0V

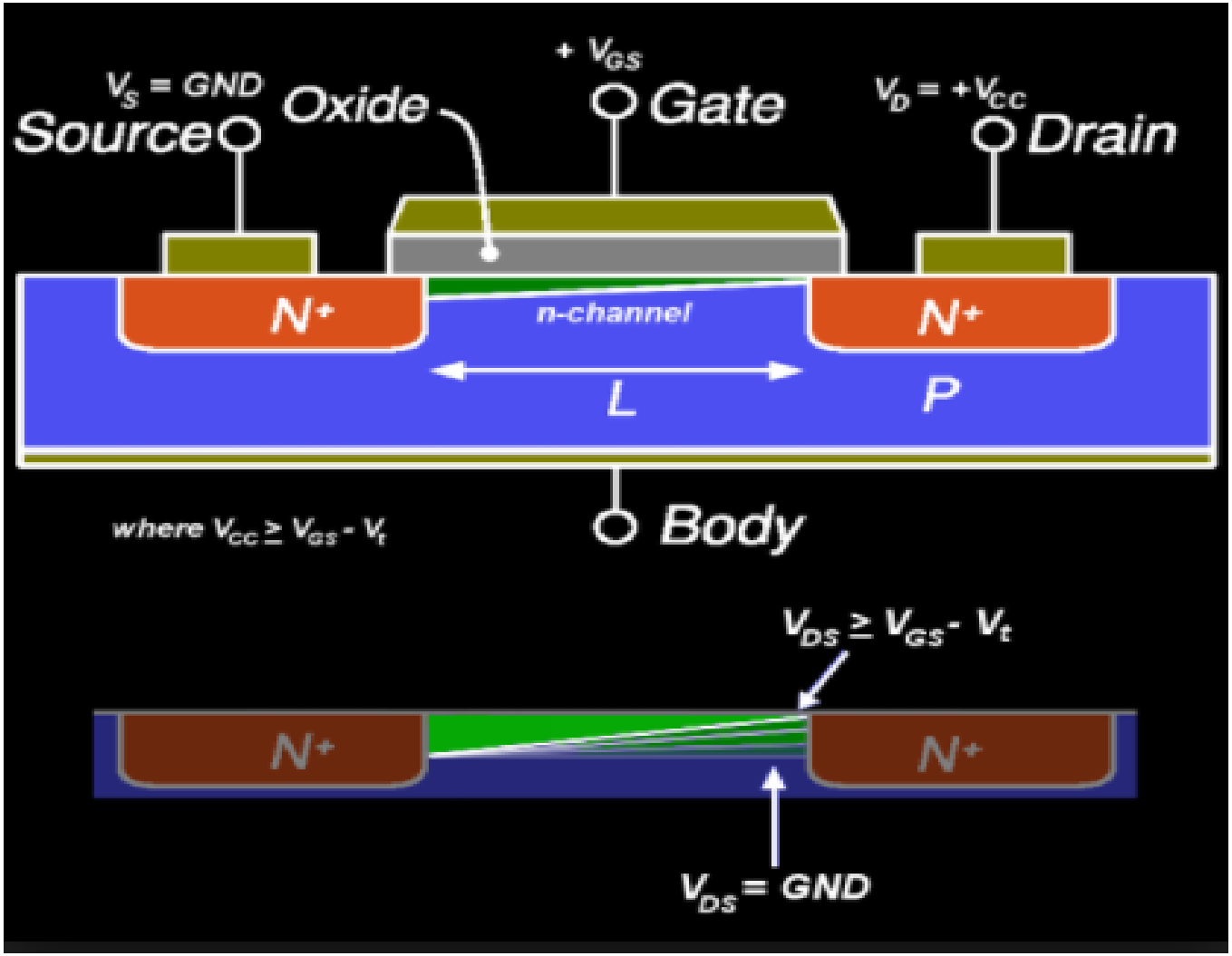


n-ch MOS

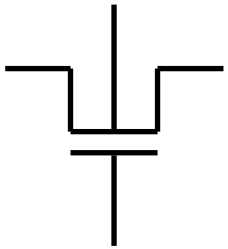


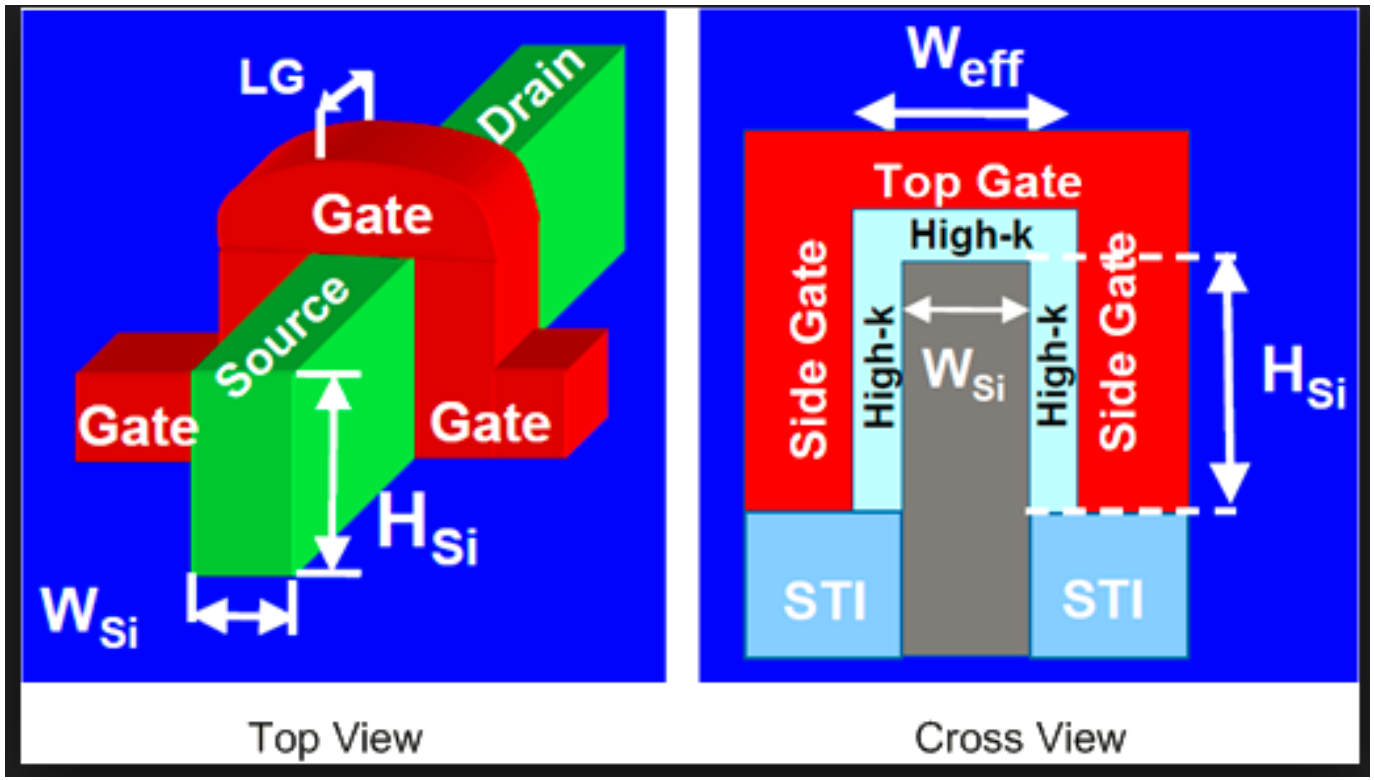
https://www.youtube.com/watch?v=p34w6ISouZY

CMOS

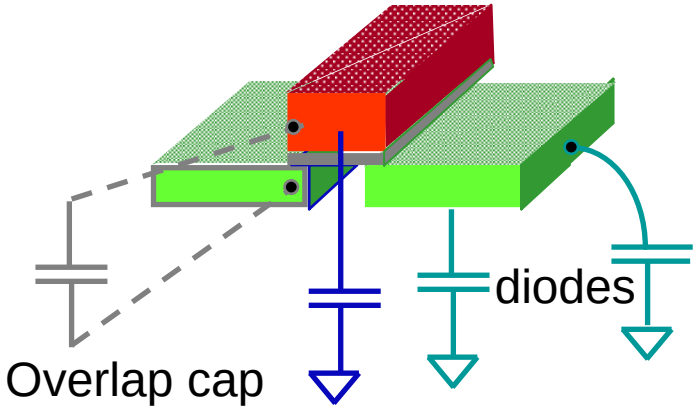


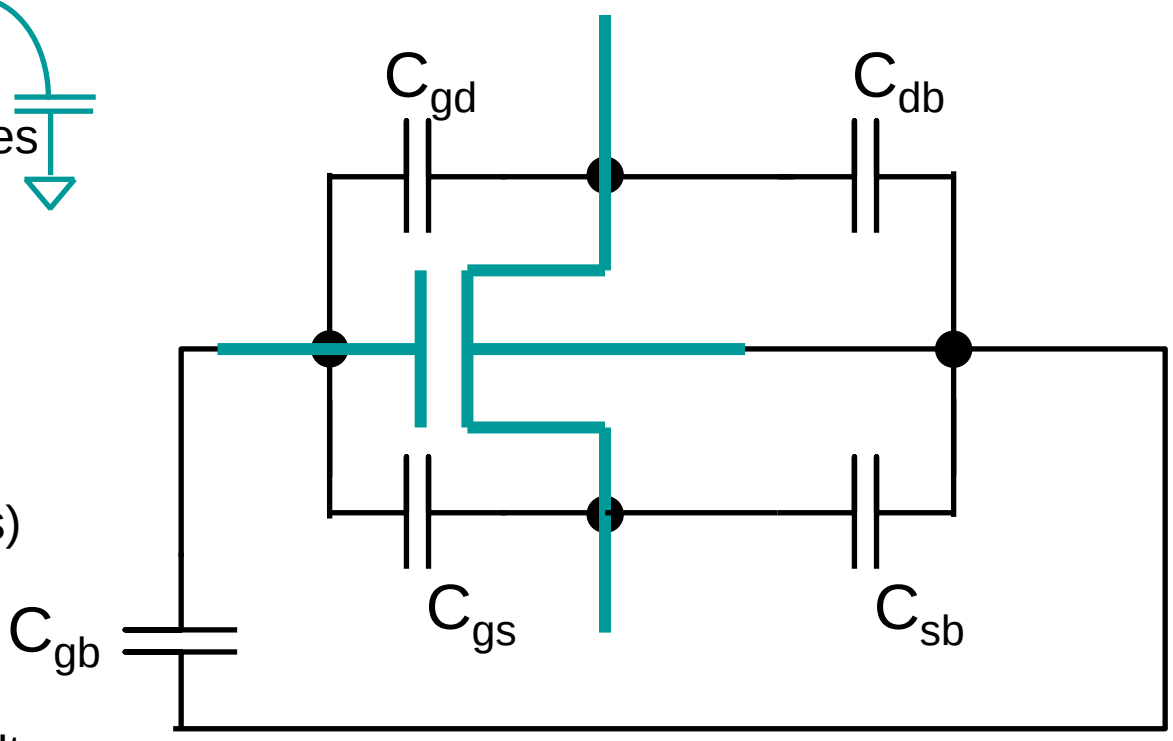
3D Gate





Capacitance of the MOS Transistor

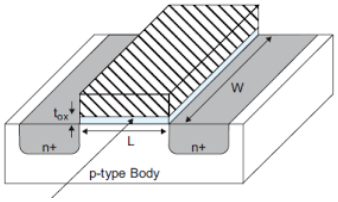




Parameters:
Diffusions area (diodes)
Diffusions perimeter (diodes)
Gate W (overlap cap)
Gate W*L (gate cap)
Doping profiles
Note: all capacitors have voltage dependence (not simple caps)

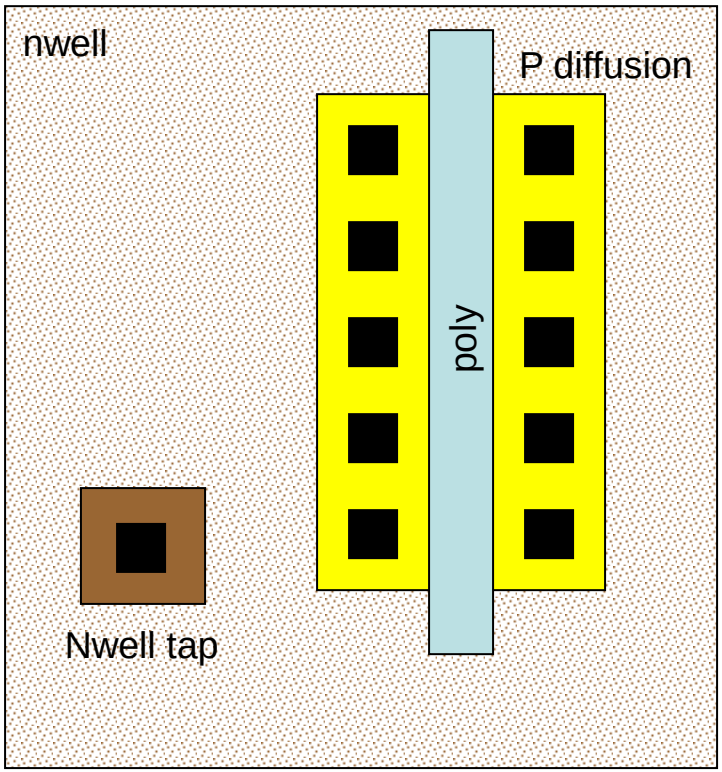
Layout

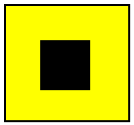
- Transistor defined as poly over diffusion

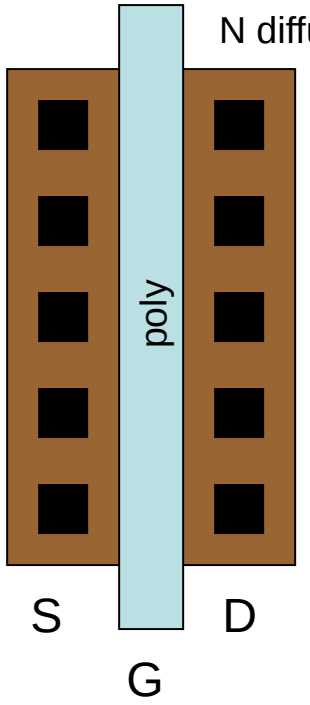


PMOS

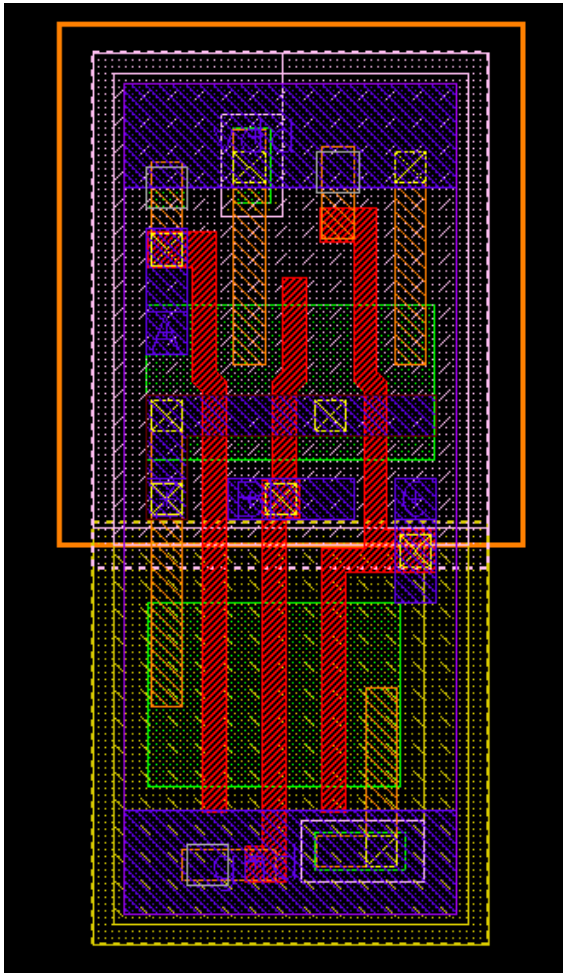
NMOS

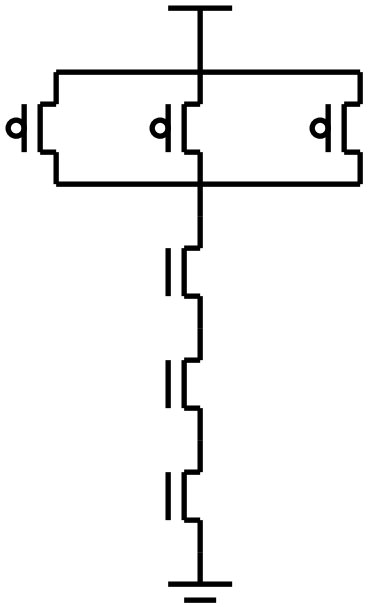






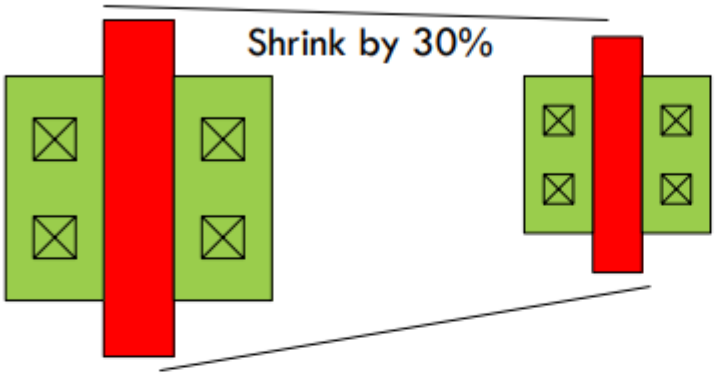
Layout vs. Schematic

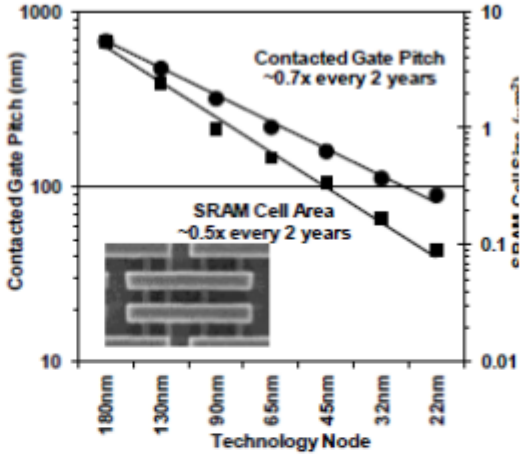




Course Content

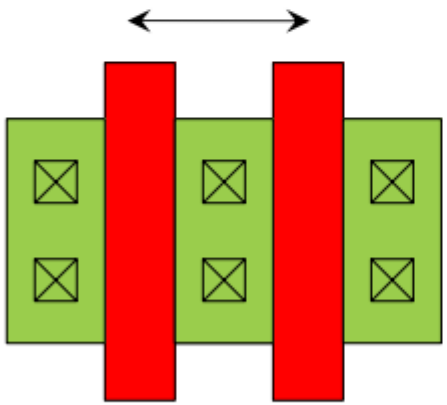
Transistor Scaling



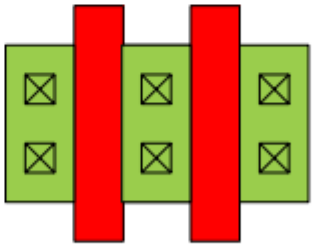


C. Auth, VLSI'12

"Contacted gate pitch"



Shrink by 30%



Gate pitch scales 0.7x every node

Intel	45nm	32nm	22nm	14nm	10nm
Contacted gate pitch	160nm	112.5nm	90nm	70nm	54nm