Hardware Fundamentals [CESE4005]

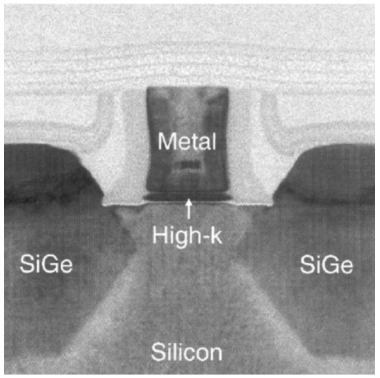
Lecture 1.2: MOS FET transistors fabrication, basic properties and operation

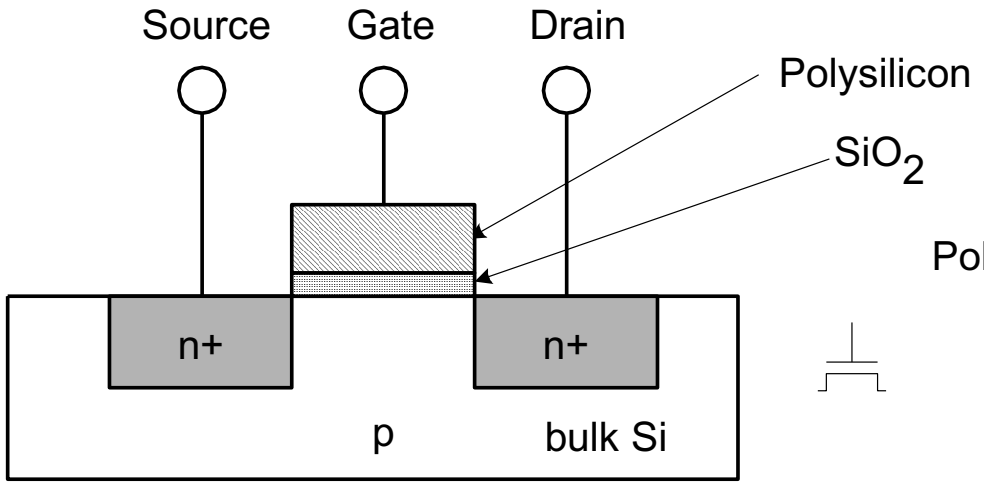
September 2024

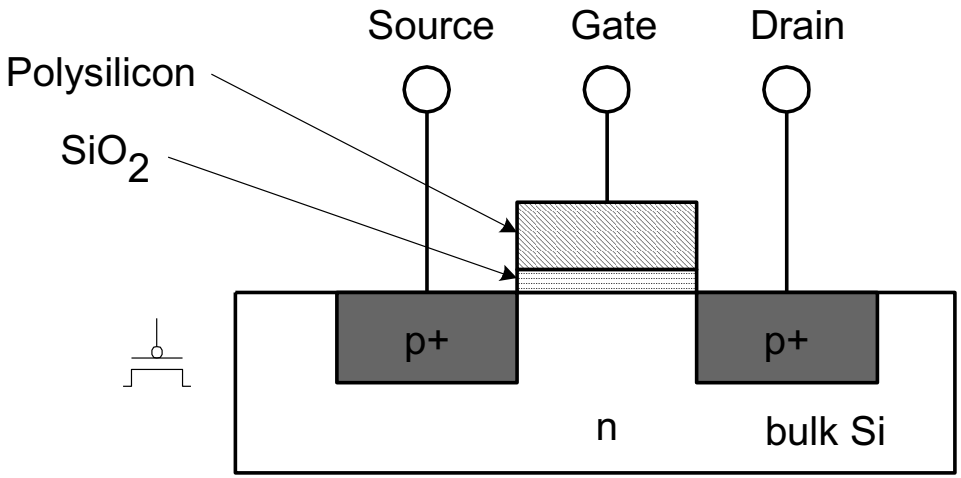
TU Delft

The MOS FET transistor

- MOS FET (**M**etal-**O**xide **S**emiconductor **F**ield-**E**ffect **T**ransistor): a sandwich structure of p-type Si, n-type Si and insulator (SiO_2) materials
- It comes in two flavors: **nMOS** and **pMOS**:

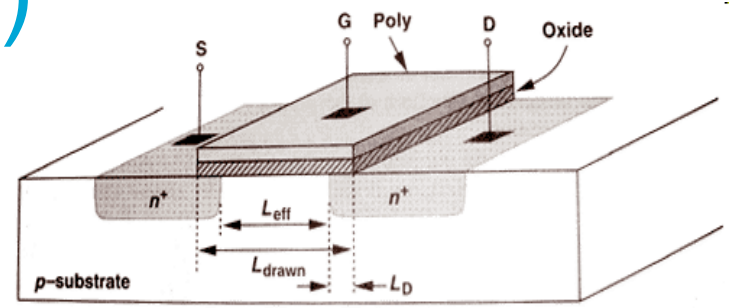


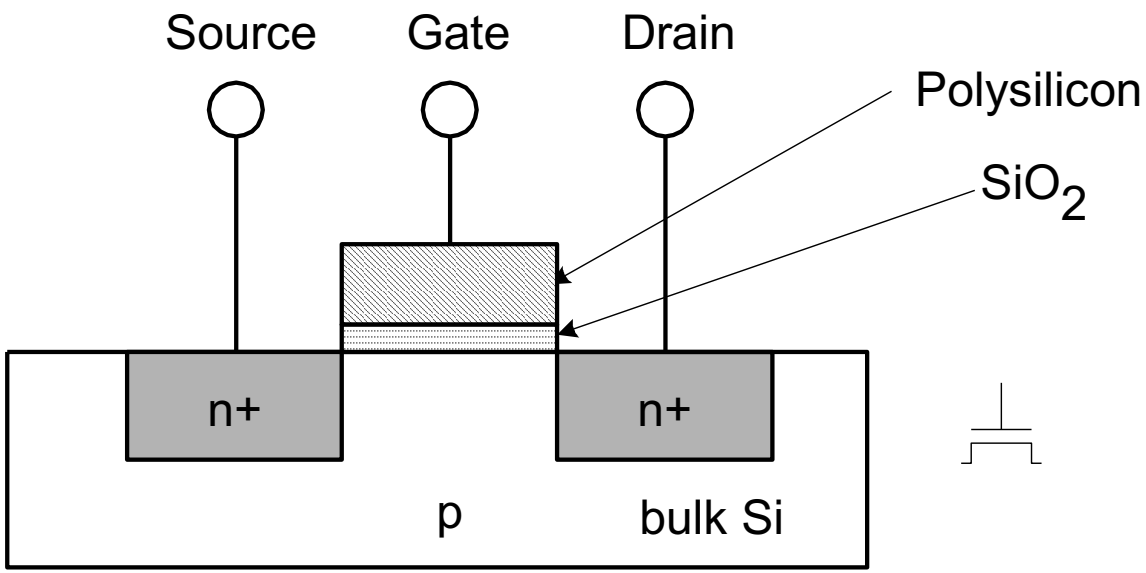


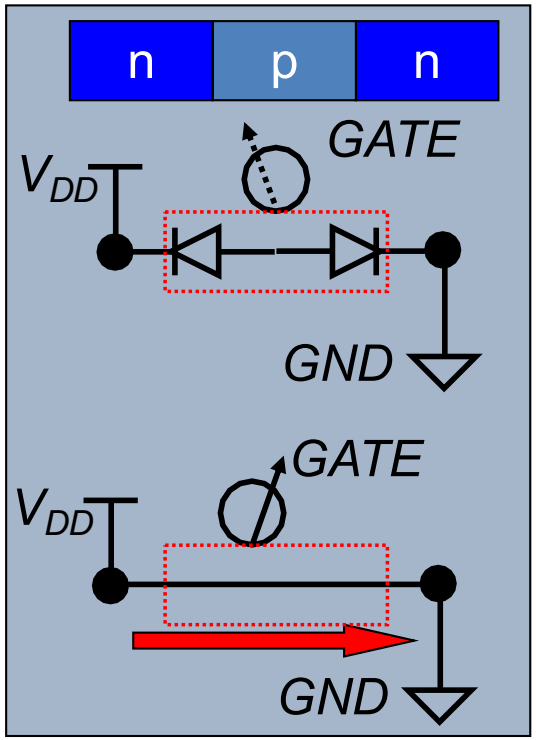


The MOS transistor (cont)

- The *Gate* is a control input:
  - $V_G < V_{th}$ Volt: nMOS transistor is **OFF**
  - $V_G \geq V_{th}$ Volt: nMOS transistor is **ON**
- The MOS transistor is a voltage-controlled **switch**!





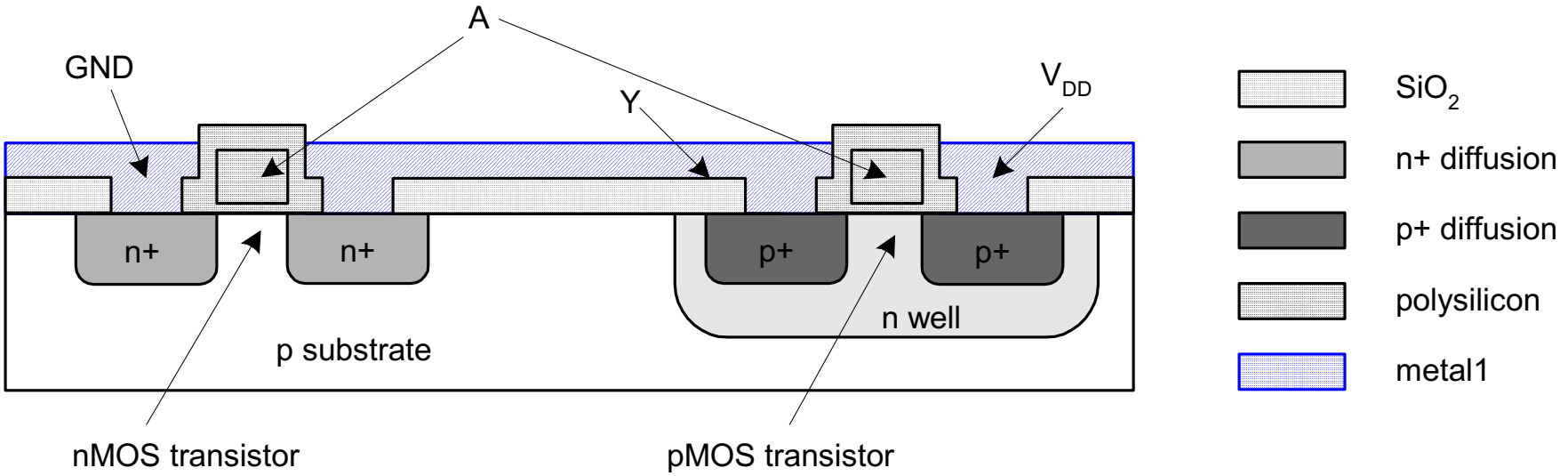


CMOS Fabrication

- CMOS transistors are fabricated on silicon wafer
- Lithography process similar to printing press
- On each step, different materials are deposited or etched
- Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process

Inverter Cross-section

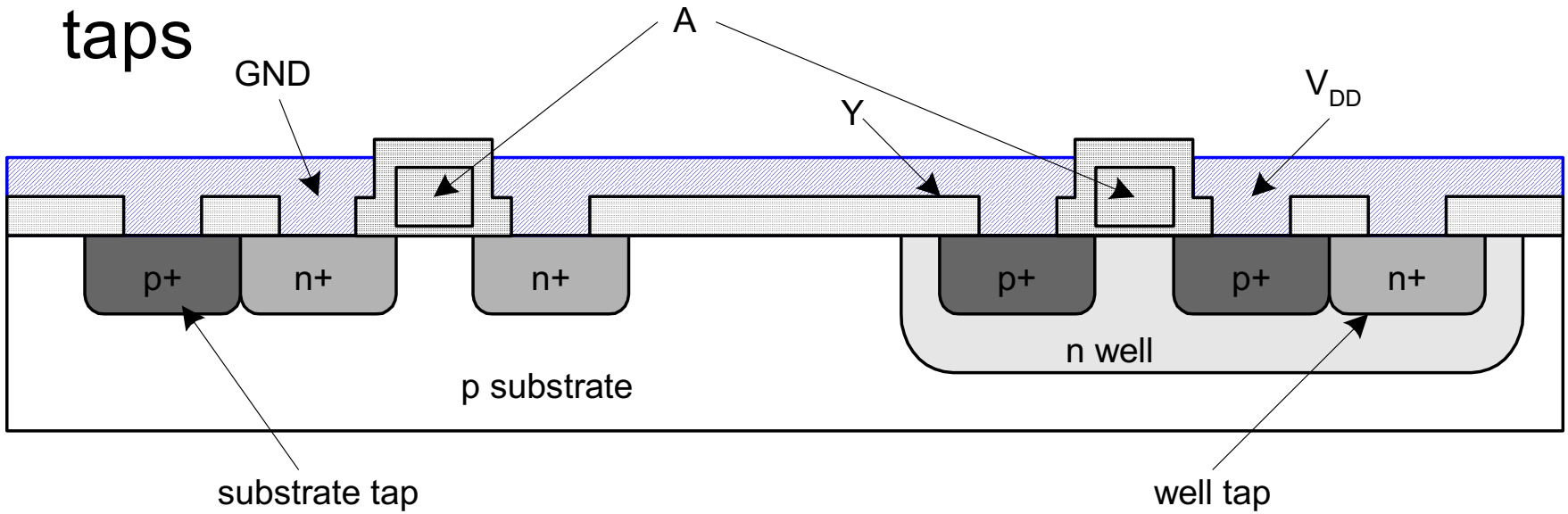
- Typically use p-type substrate for nMOS transistor
  - Requires n-well for body of pMOS transistors
  - Several alternatives: SOI, twin-tub, etc.



Q: what is shown on this figure

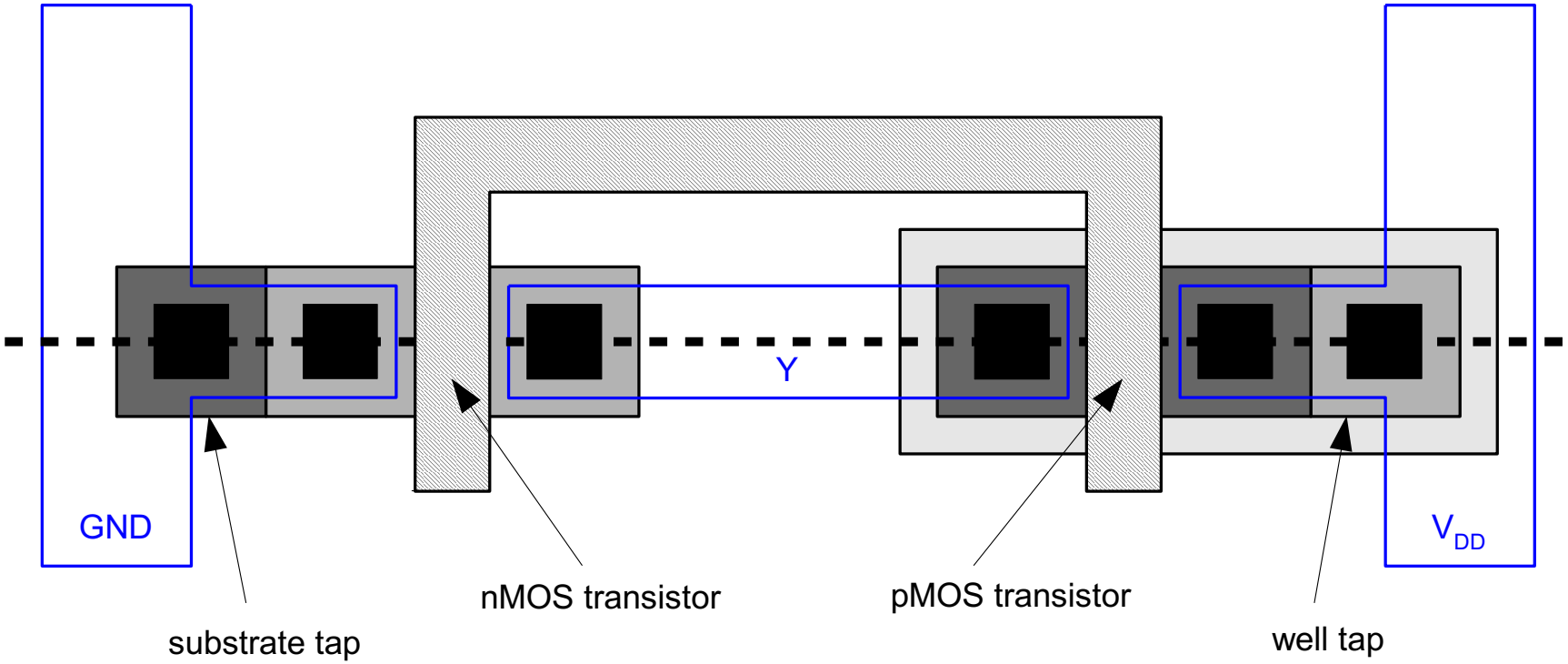
Well and Substrate Taps

- Substrate must be tied to GND and n-well to V_{DD}
- Metal to lightly-doped semiconductor forms poor connection called Shottky Diode
- Use heavily doped well and substrate contacts / taps



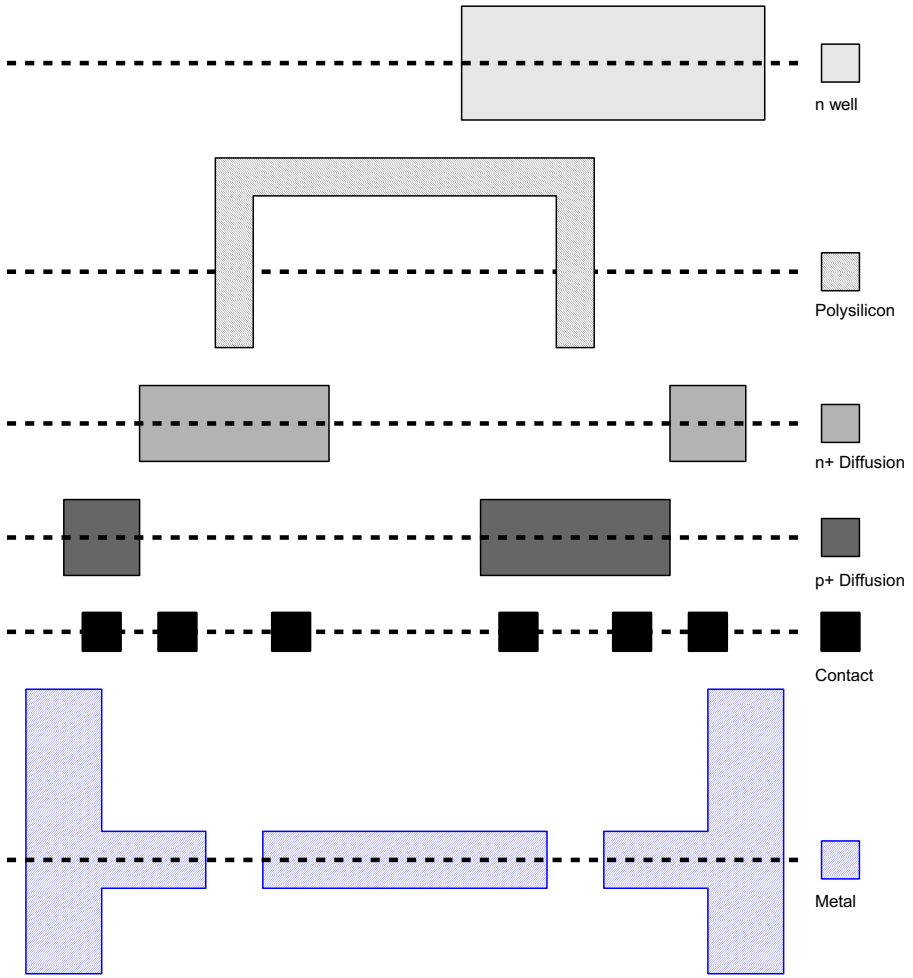
Inverter Mask Set

- Transistors and wires are defined by *masks*
- Cross-section taken along dashed line



Detailed Mask Views

- Six masks
  - n-well
  - Polysilicon
  - n+ diffusion
  - p+ diffusion
  - Contact
  - Metal

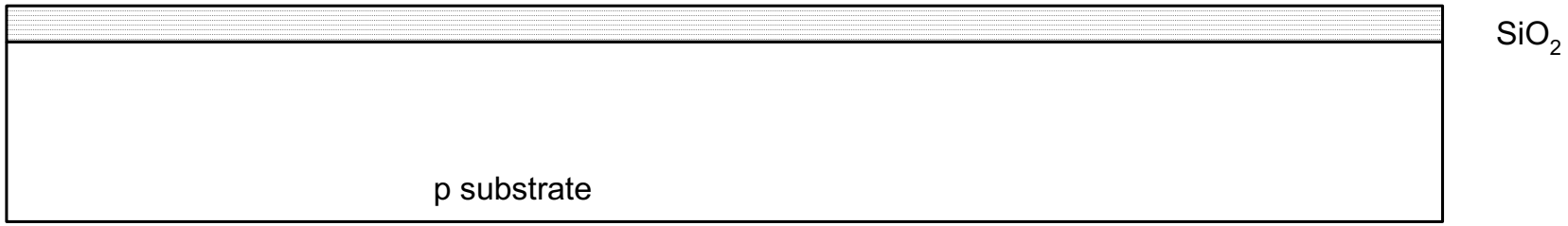


Fabrication Steps

- Start with blank wafer
- Build inverter from the bottom up
- First step will be to form the n-well
  - Cover wafer with protective layer of SiO_2 (oxide)
  - Remove layer where n-well should be built
  - Implant or diffuse n dopants into exposed wafer
  - Strip off SiO_2

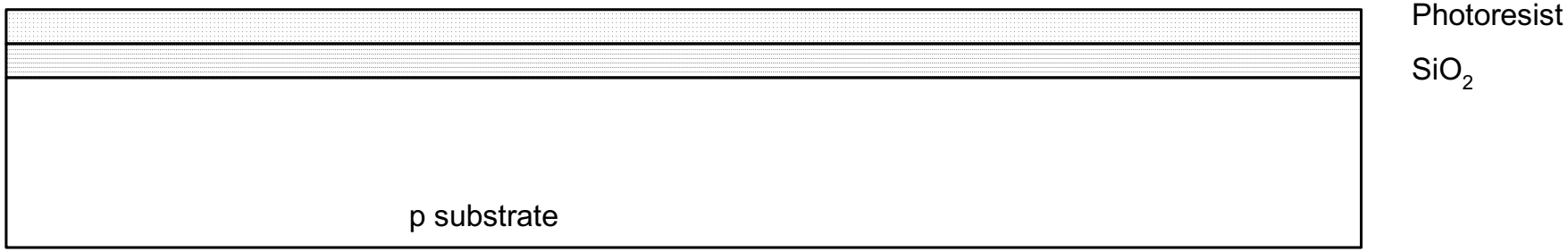
Oxidation

- Grow SiO_2 on top of Si wafer
  - 900 – 1,200 ^oC with H_2O or O_2 in oxidation furnace



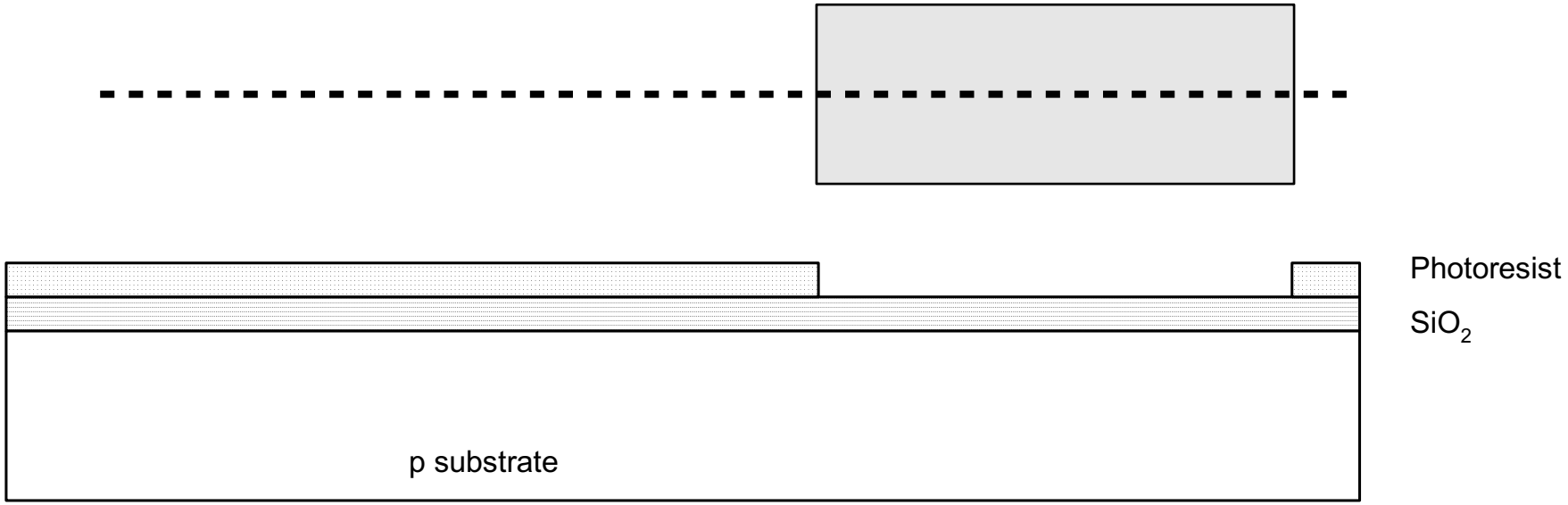
Photoresist

- Spin on photoresist
  - Photoresist is a light-sensitive organic polymer
  - Softens where exposed to light



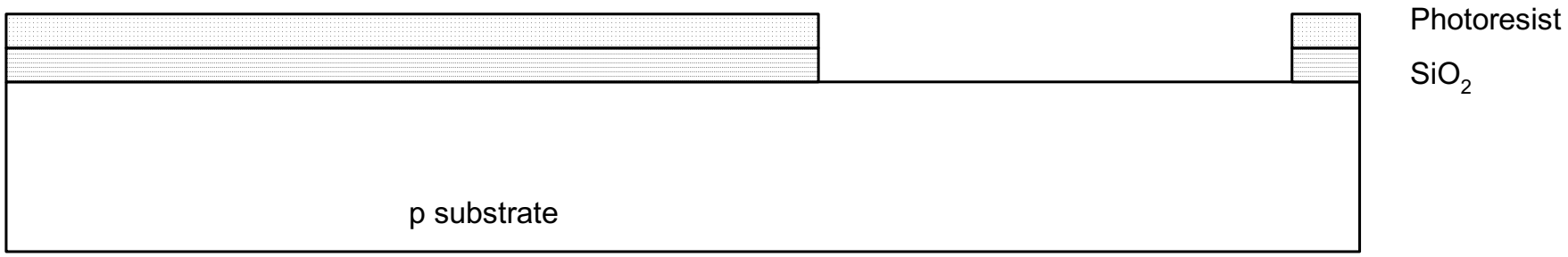
Lithography

- Expose photoresist through n-well mask
- Strip off exposed photoresist



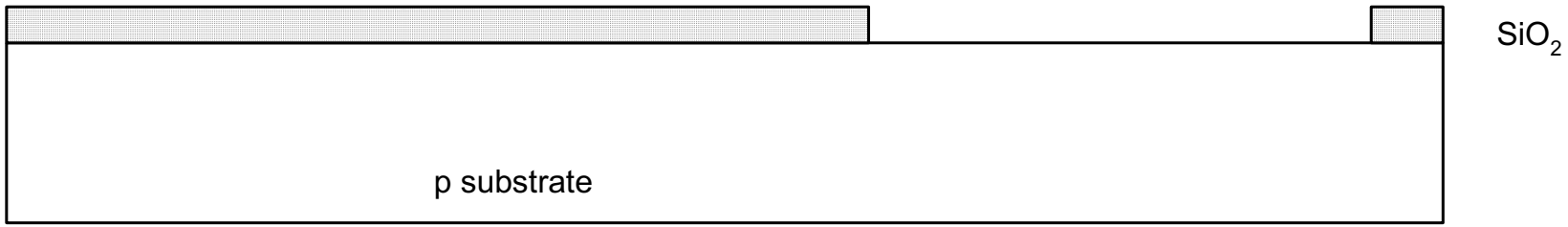
Etch

- Etch oxide with hydrofluoric acid (HF)
  - Seeps through skin and eats bone; nasty stuff!!!
- Only attacks oxide where resist has been exposed



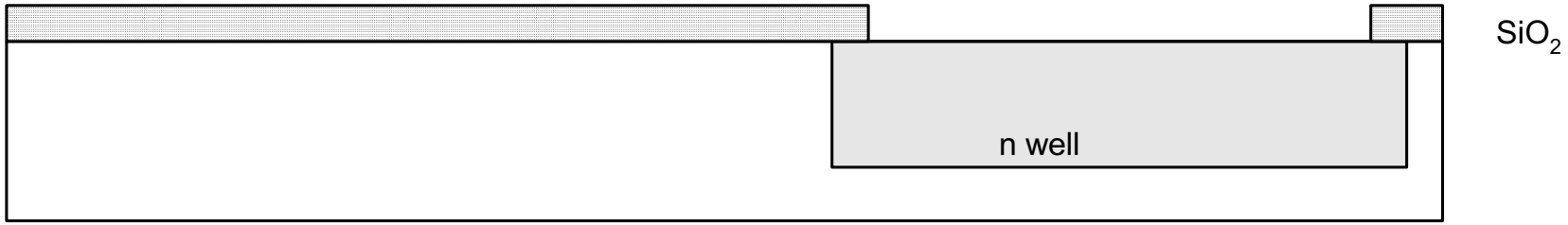
Strip Photoresist

- Strip off remaining photoresist
  - Use mixture of acids called piranah etch
- Necessary so resist doesn't melt in next step



n-well

- n-well is formed with diffusion or ion implantation
- Diffusion
  - Place wafer in furnace with arsenic gas
  - Heat until As atoms diffuse into exposed Si
- Ion Implanatation
  - Blast wafer with beam of As ions
  - Ions blocked by SiO_2, only enter exposed Si

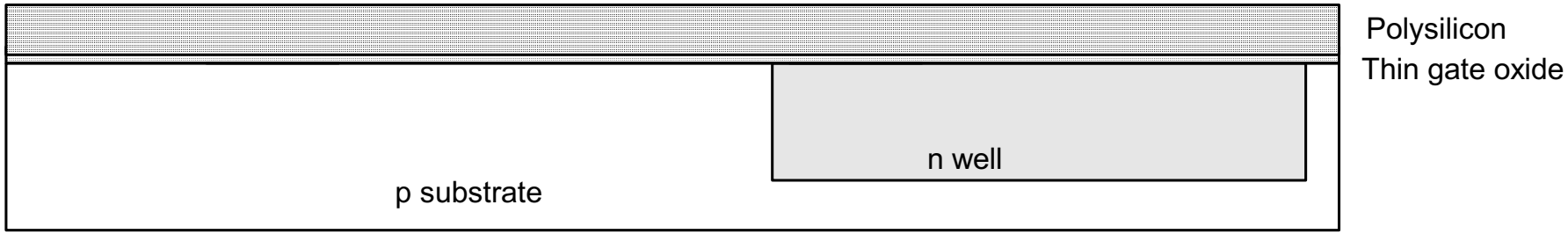


Strip Oxide

- Strip off the remaining oxide using HF
- Back to bare wafer with n-well
- Subsequent steps involve similar series of steps

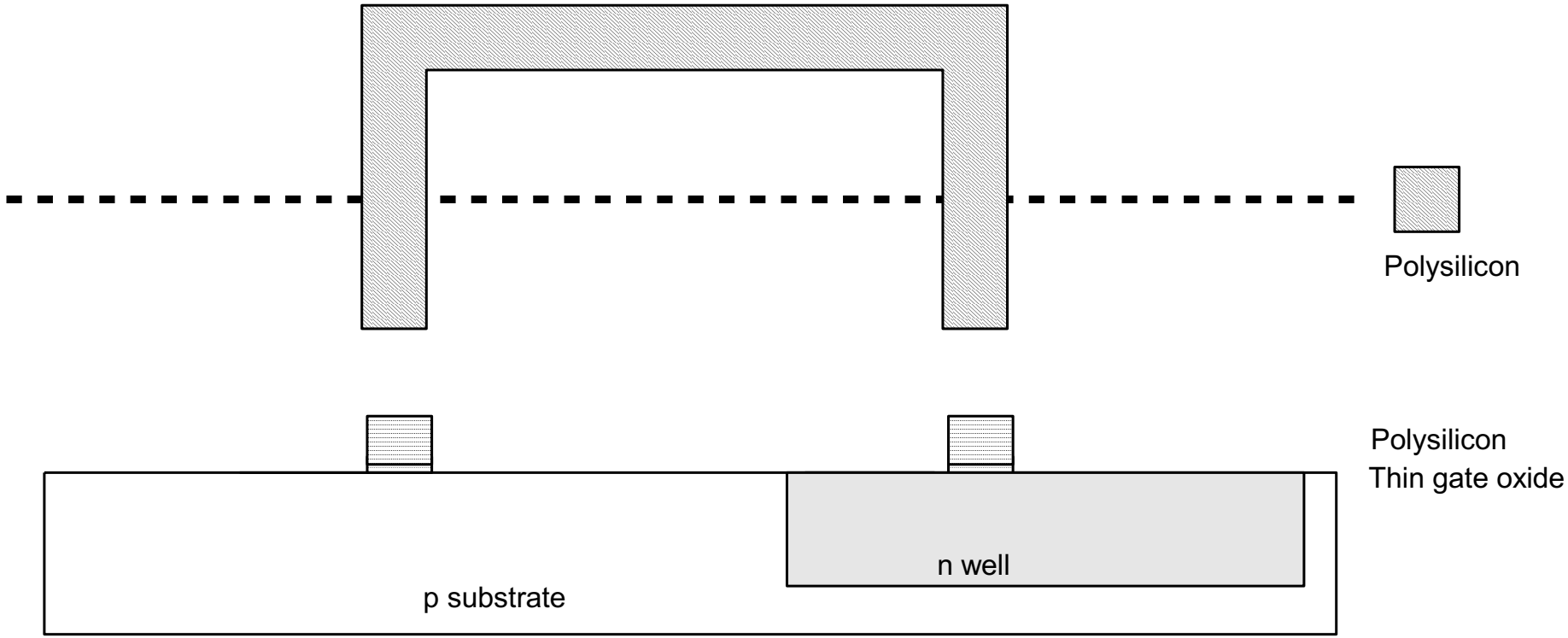
Polysilicon

- Deposit very thin layer of gate oxide
  - < 20 Å (6-7 atomic layers)
- Chemical Vapor Deposition (CVD) of silicon layer
  - Place wafer in furnace with Silane gas (SiH_4)
  - Forms many small crystals called polysilicon
  - Heavily doped to be good conductor



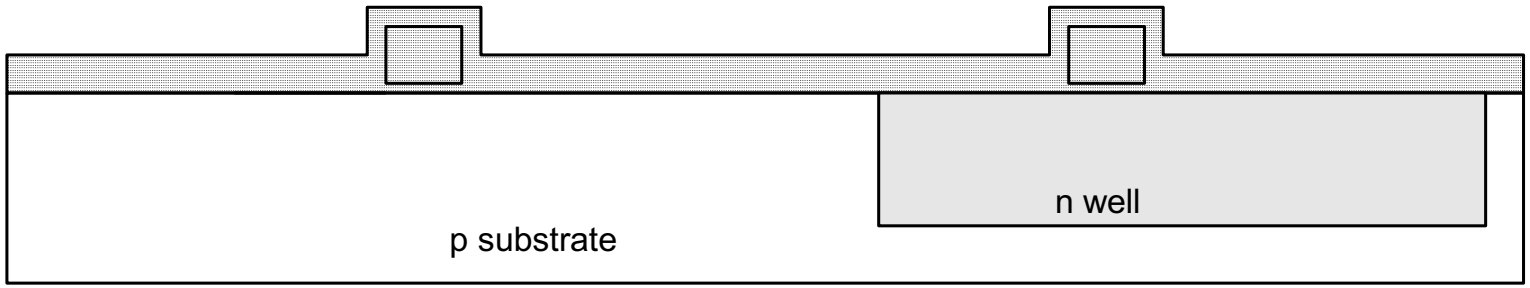
Polysilicon Patterning

- Use same lithography process to pattern polysilicon



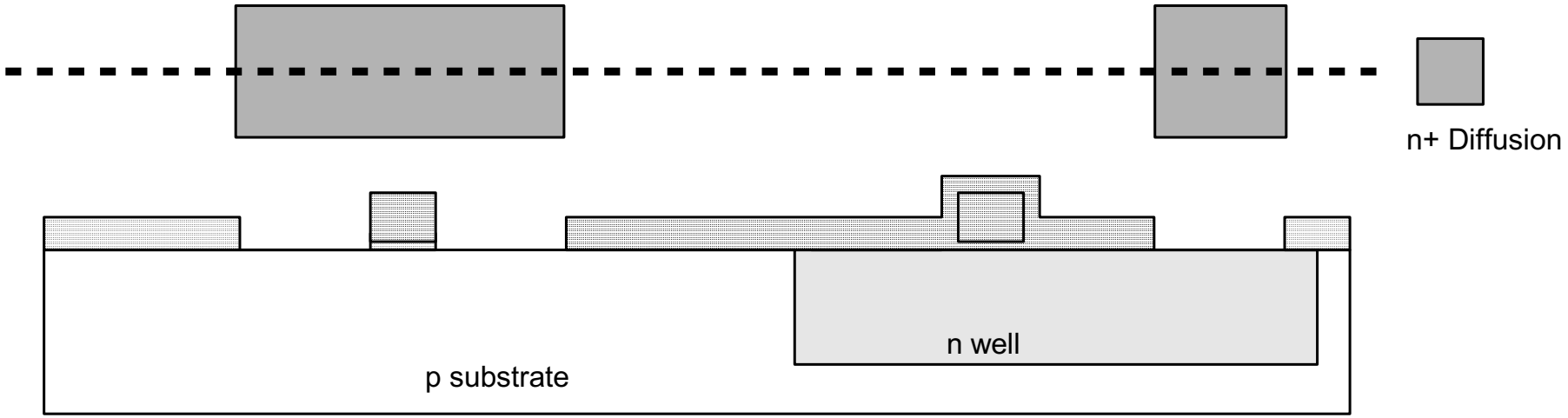
Self-Aligned Process

- Use oxide and masking to expose where n+ dopants should be diffused or implanted
- N-diffusion forms nMOS source, drain, and n-well contact



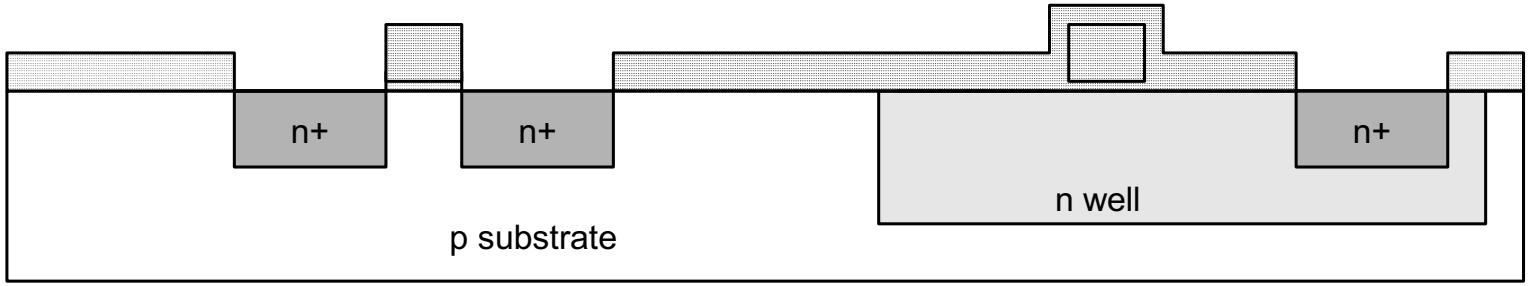
N-diffusion

- Pattern oxide and form n+ regions
- *Self-aligned process* where gate blocks diffusion
- Polysilicon is better than metal for self-aligned gates because it doesn't melt during later processing



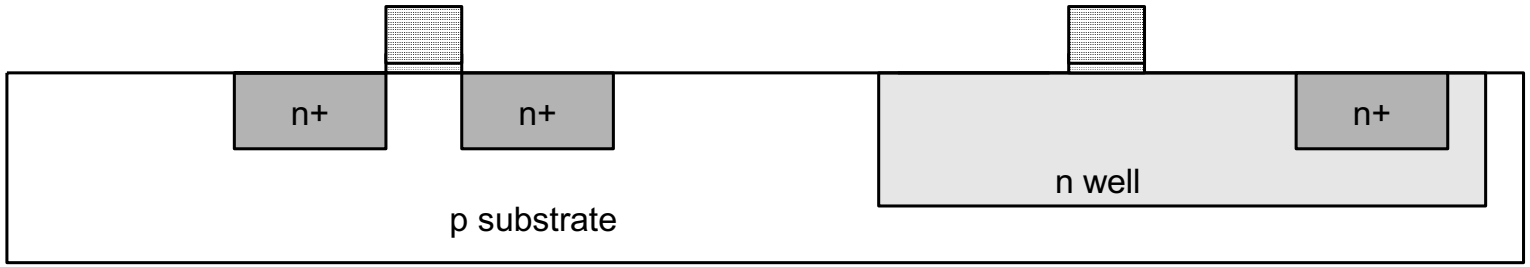
N-diffusion

- Historically dopants were diffused
- Usually ion implantation today
- But regions are still called diffusion



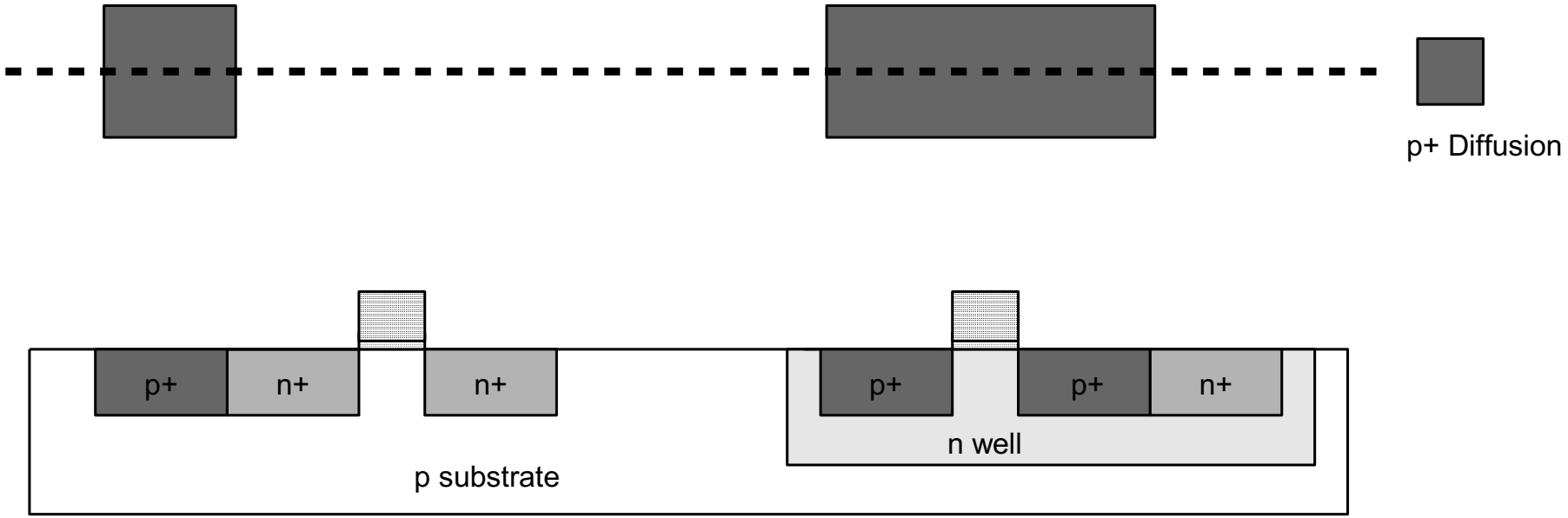
N-diffusion

- Strip off oxide to complete patterning step



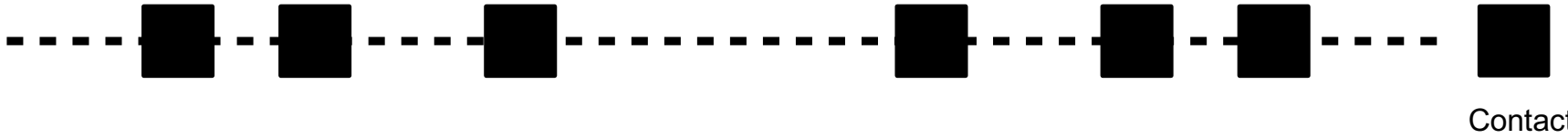
P-Diffusion

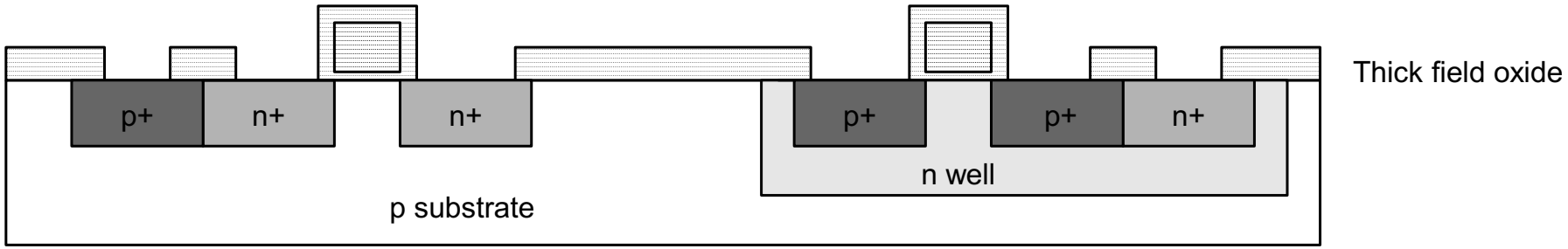
- Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact



Contacts

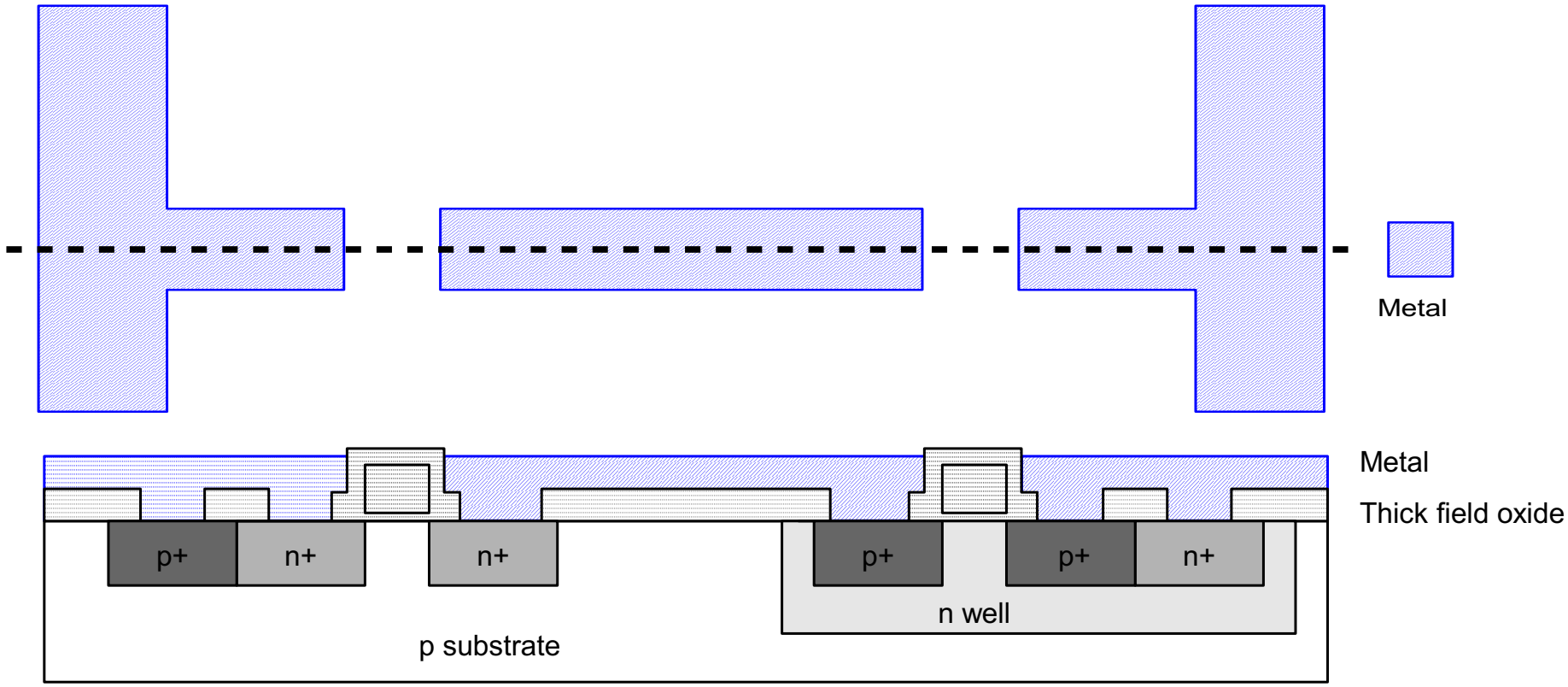
- Now we need to wire together the devices
- Cover chip with thick field oxide
- Etch oxide where contact cuts are needed





Metallization

- Sputter on aluminum over whole wafer
- Pattern to remove excess metal, leaving wires

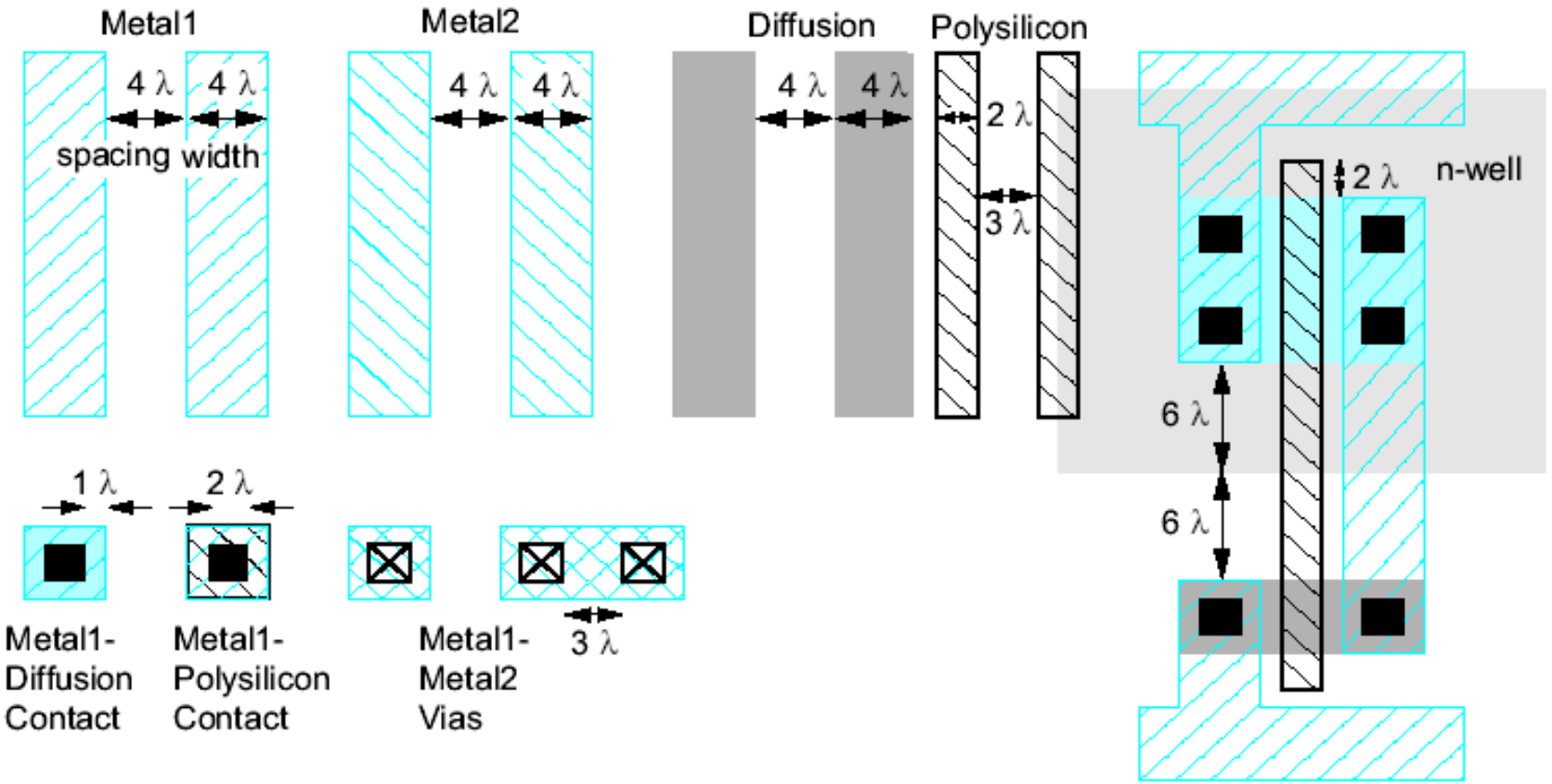


Layout

- Chips are specified with set of masks
- Minimum dimensions of masks determine transistor size (and hence speed, cost, and power)
- Feature size *f* = distance between source and drain
  - Set by minimum width of polysilicon
- Feature size improves 30% every 3 years or so
- Normalize for feature size when describing design rules
- Express rules in terms of $\lambda = f/2$
  - E.g., λ = 0.3 μm in 0.6 μm process

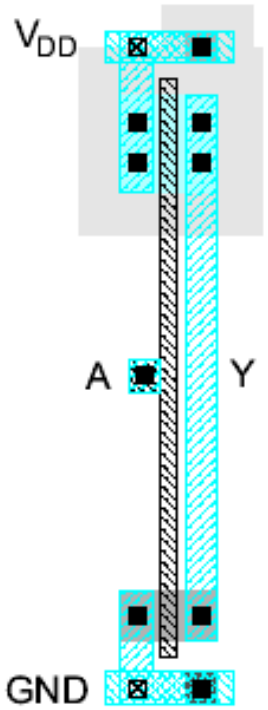
Simplified Design Rules

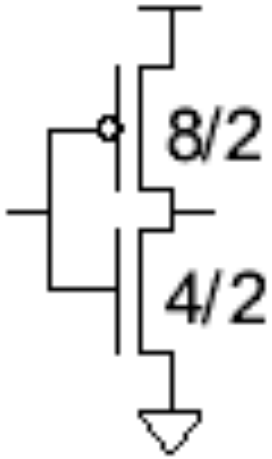
- Conservative rules to get you started

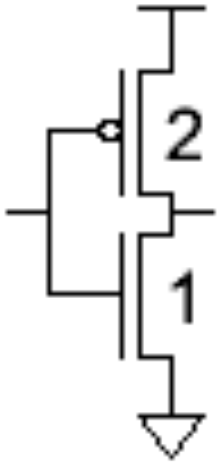


Inverter Layout

- Transistor dimensions specified as Width / Length
  - Minimum size is 4λ / 2λ, sometimes called 1 unit
  - For 0.6 μm process, W=1.2 μm, L=0.6 μm

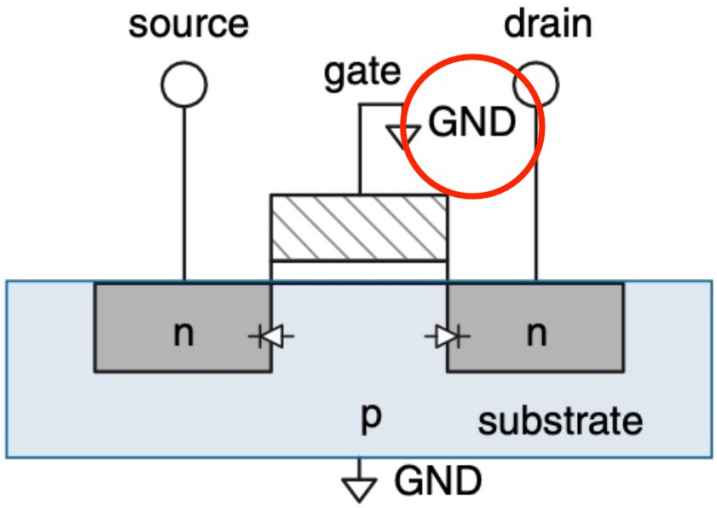


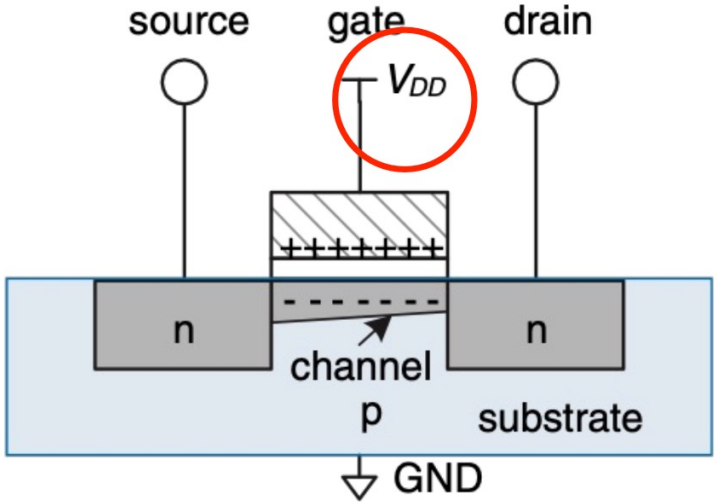




This is a bit outdated

nMOS transistor operation

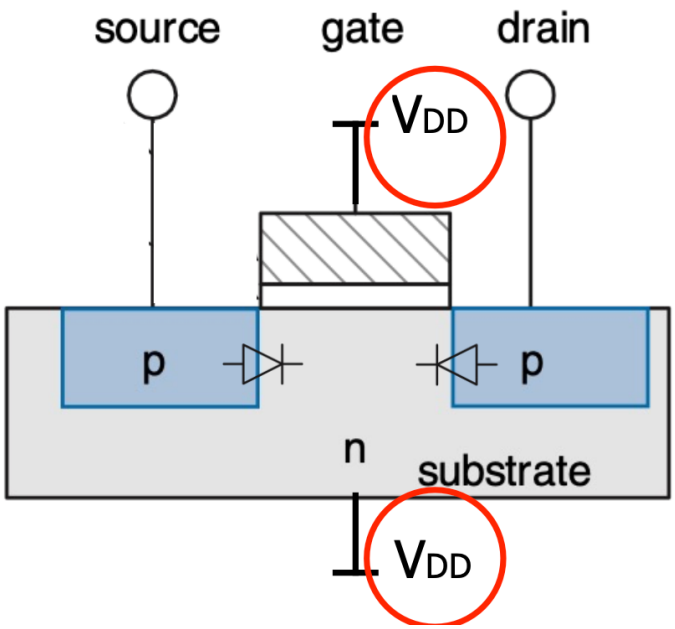


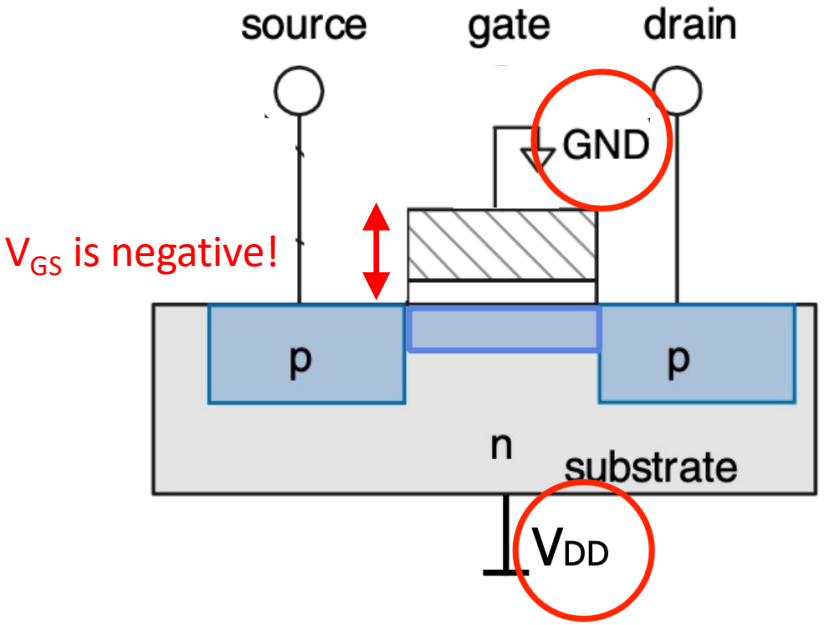


- The two diodes between source (S) and drain (D) are reverse biased as S/D voltages =0 or >0
- There is no current path between S and D: **Transistor is OFF**.

- When the positive voltage is applied to the gate, an electric field attracts holes on the gate and electrons in the p-type substrate: **"inversion"**.
- The current flow through the channel: **Transistor is ON.**

pMOS transistor operation





- The two diodes between S and D are reverse biased as S/D voltages =0 or <VDD
- There is no current path between S and D: **Transistor is OFF.**

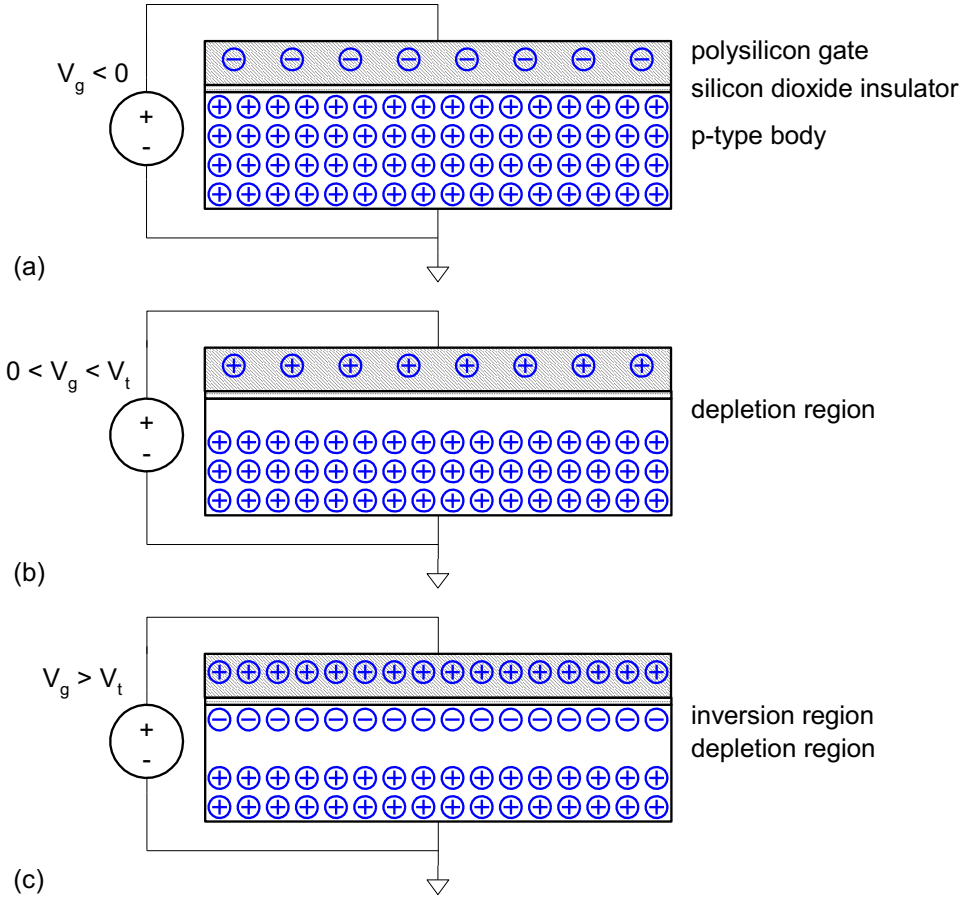
- When the gate is grounded, an electric field attracts electrons on the gate and holes in the n-type substrate: "**inversion**".
- The current flow through the channel: **Transistor is ON.**

What happens in reality (or closer to)

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - $I = C (\Delta V/\Delta t) \rightarrow \Delta t = (C/I) \Delta V$
  - Capacitance and current determine speed
- Also explore what a “degraded level” really means

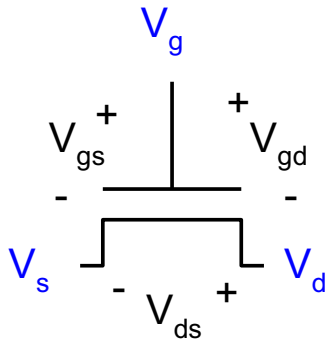
MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
  - Accumulation
  - Depletion
  - Inversion



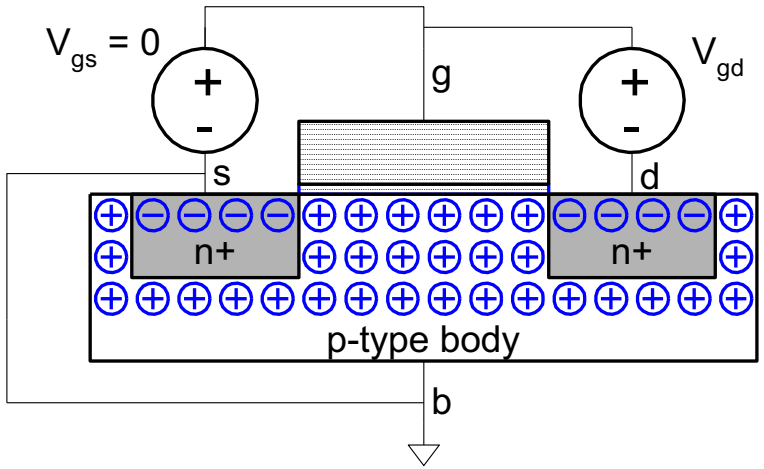
Terminal Voltages

- Mode of operation depends on V_g, V_d, V_s
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$
- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} \geq 0$
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - *Cutoff*
  - *Linear*
  - *Saturation*



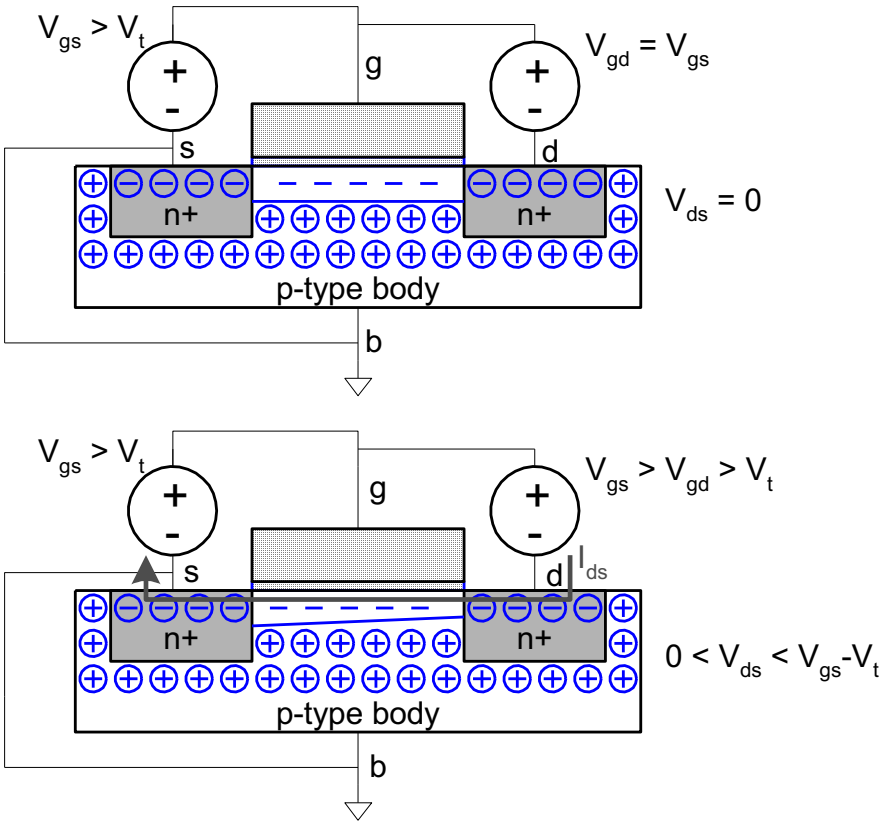
nMOS Cutoff

- No channel
- $I_{ds} = 0$



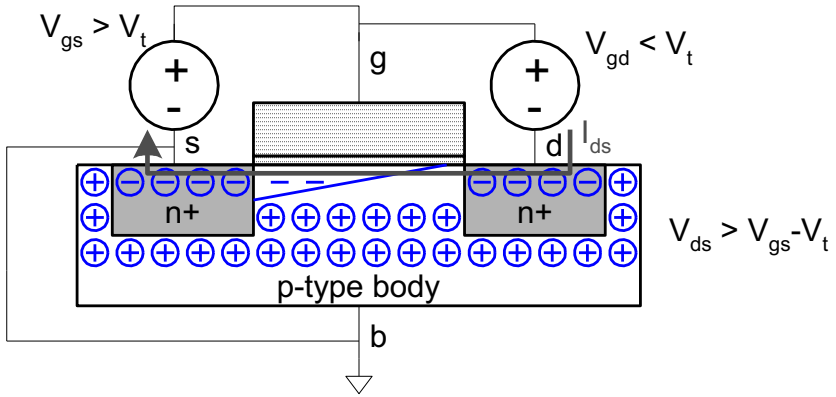
nMOS Linear

- Channel forms
- Current flows from d to s
  - e^- from s to d
- I_{ds} increases with V_{ds}
- Similar to linear resistor



nMOS Saturation

- Channel pinches off
- I_{ds} independent of V_{ds}
- We say current saturates
- Similar to current source

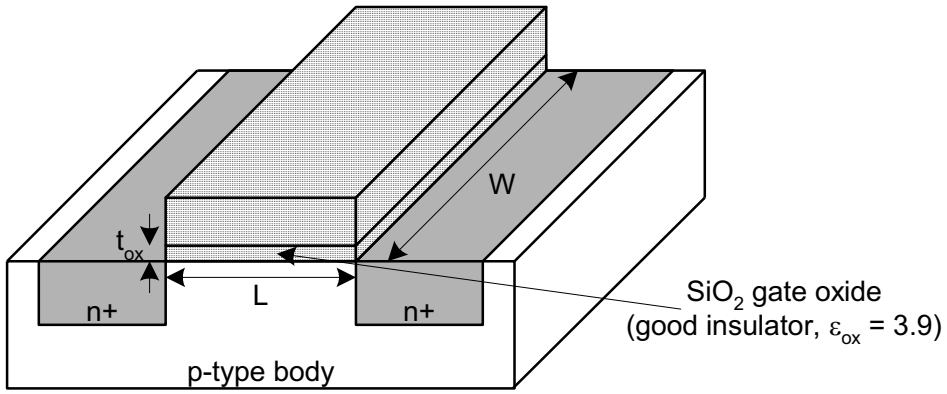


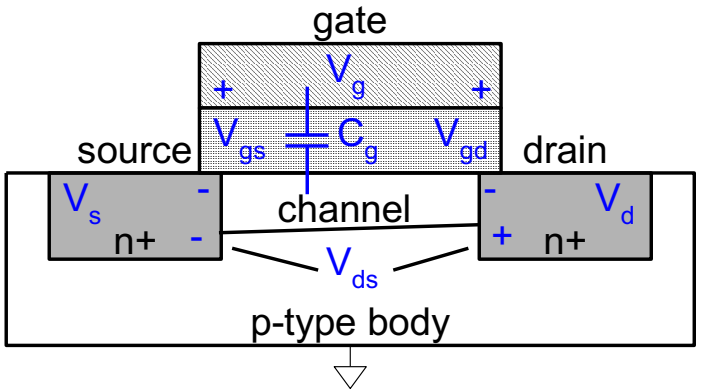
I-V Characteristics

- In Linear region, I_{ds} depends on
  - How much charge is in the channel?
  - How fast is the charge moving?

Channel Charge

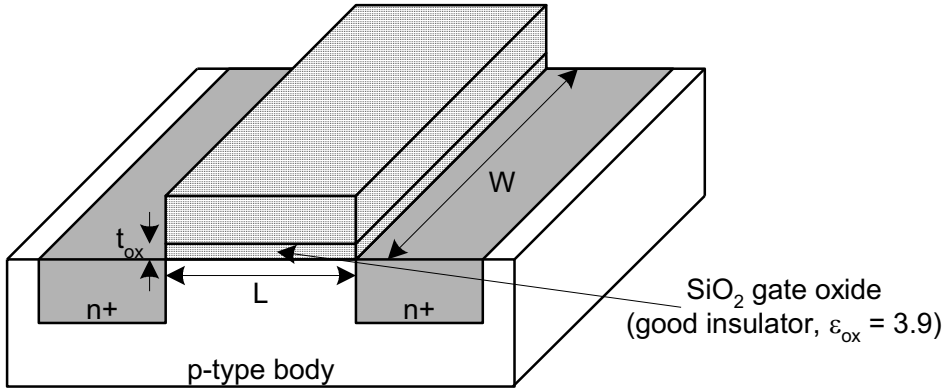
- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel}$ =

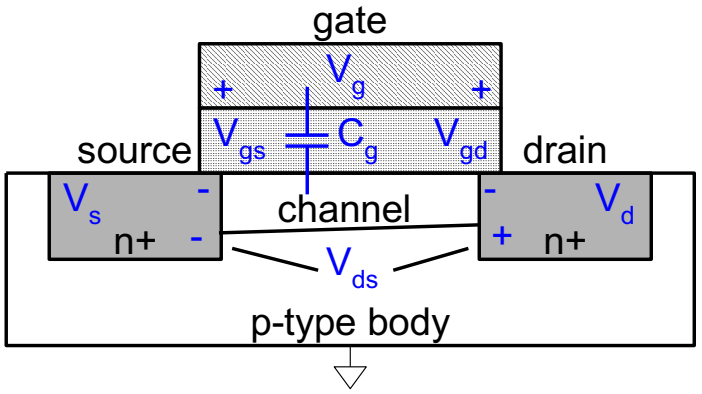




Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- C =

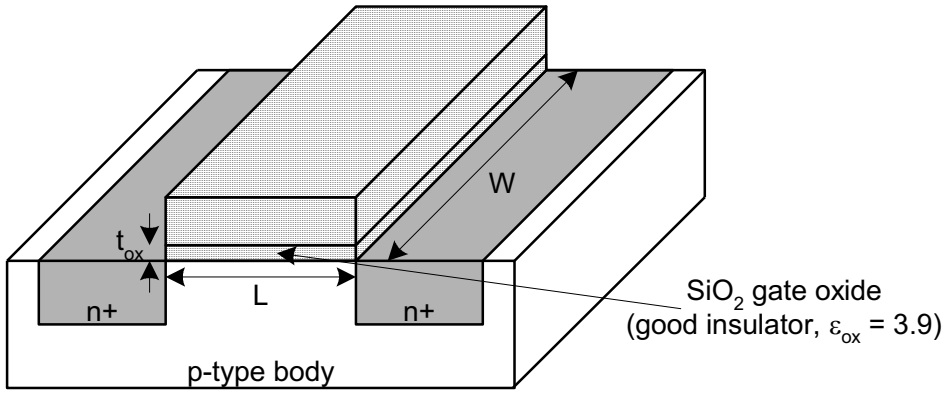


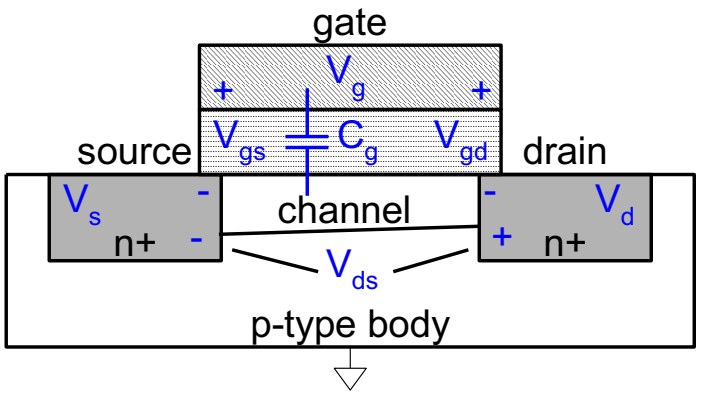


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL$
- $V =$

$C_{ox} = \varepsilon_{ox} / t_{ox}$

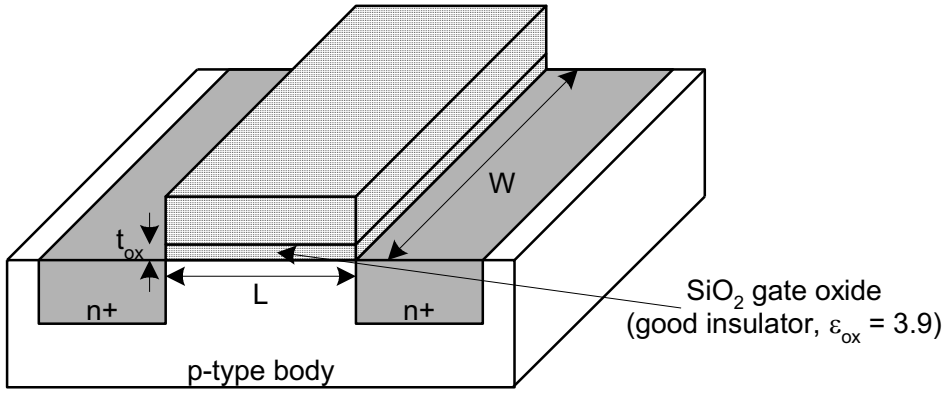


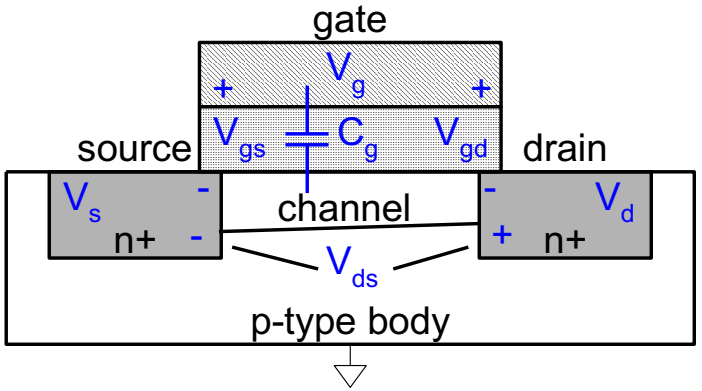


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL$
- $V = V_{gc} - V_t = (V_{gs} - V_{ds}/2) - V_t$

$C_{ox} = \varepsilon_{ox} / t_{ox}$





Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- *v* =

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- E =

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t =$

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t = L / v$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} =$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$=$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$= \mu C_{\text{ox}} \frac{W}{L}\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds}$$

$$= \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds} \qquad \beta = \mu C_{\text{ox}} \frac{W}{L}$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} =$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta\left(V_{gs} - V_t - V_{dsat}/2\right)V_{dsat}$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta\left(V_{gs} - V_t - V_{dsat}/2\right)V_{dsat}$$

$$= \frac{\beta}{2}\left(V_{gs} - V_t\right)^2$$

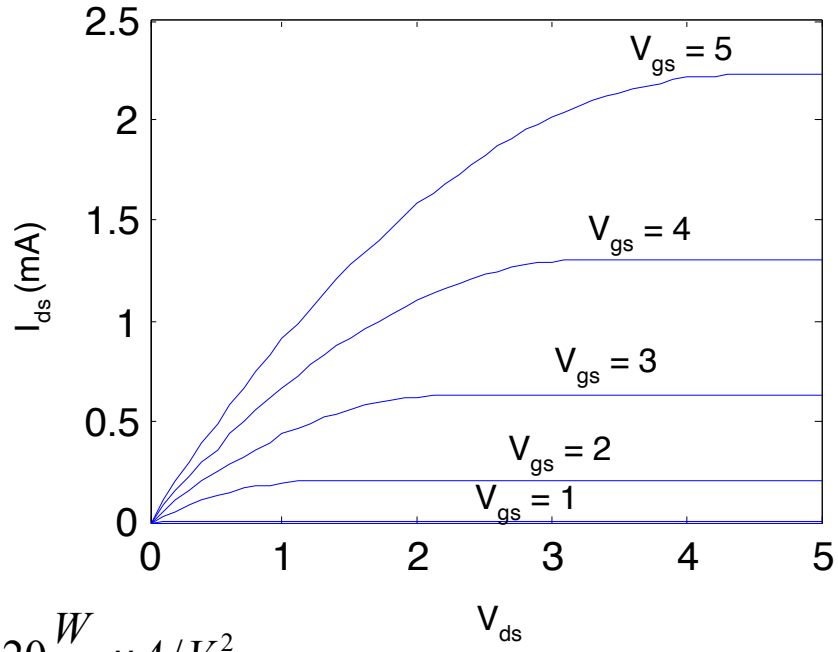
nMOS I-V Summary

- *Shockley* 1st order transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

Example

- Example: a 0.6 μm process from AMI semiconductor
  - t_{ox} = 100 Å
  - μ = 350 $cm^2/V*s$
  - V_t = 0.7 V
- Plot I_{ds} vs. V_{ds}
  - V_{gs} = 0, 1, 2, 3, 4, 5
  - Use W/L = 4/2 λ



$$\beta = \mu C_{ox}\frac{W}{L} = (350)\left(\frac{3.9 \bullet 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}}\right)\left(\frac{W}{L}\right) = 120\frac{W}{L}\mu A/V^2$$

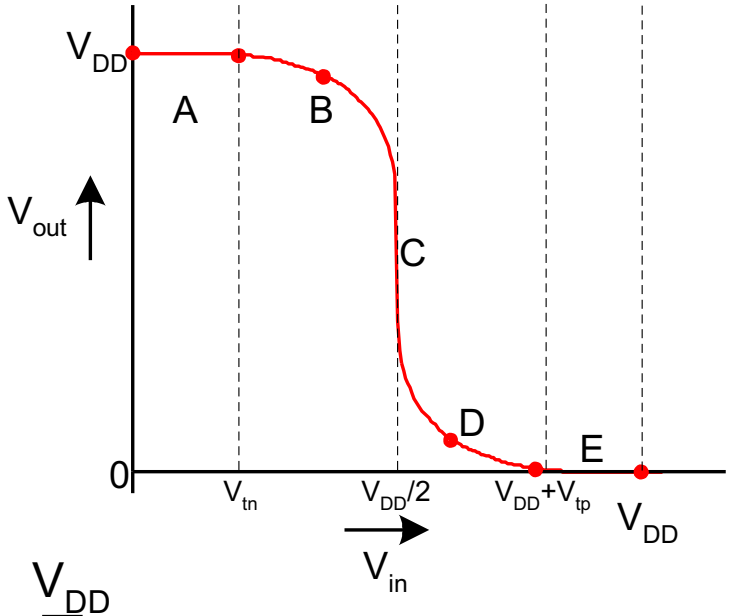
pMOS I-V

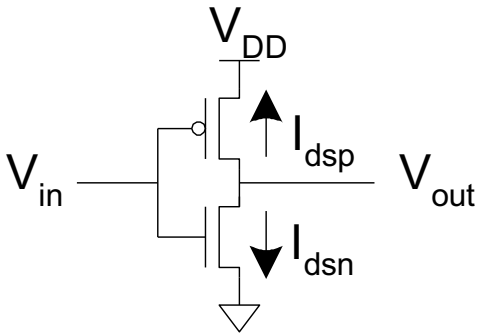
- All dopings and voltages are inverted for pMOS
- Mobility μ_p is determined by holes
  - Typically 2-3x lower than that of electrons μ_n
  - 120 $cm^2/V*s$ in AMI 0.6 μm process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$

Operating Regions of Inverter

- Revisit transistor operating regions

Region	nMOS	pMOS
A	Cutoff	Linear
B	Saturation	Linear
C	Saturation	Saturation
D	Linear	Saturation
E	Linear	Cutoff



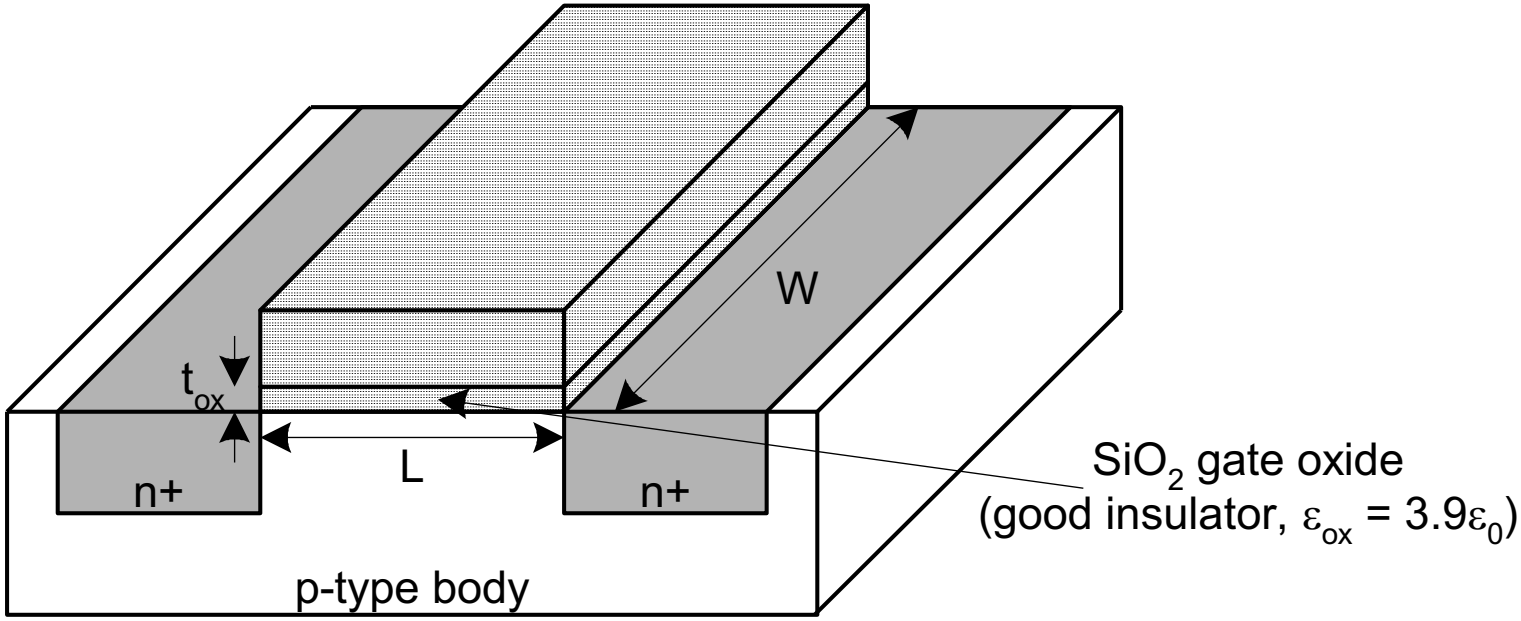


Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
  - Creates channel charge necessary for operation
- Source and drain have capacitance to body
  - Across reverse-biased diodes
  - Called diffusion capacitance because it is associated with source/drain diffusion

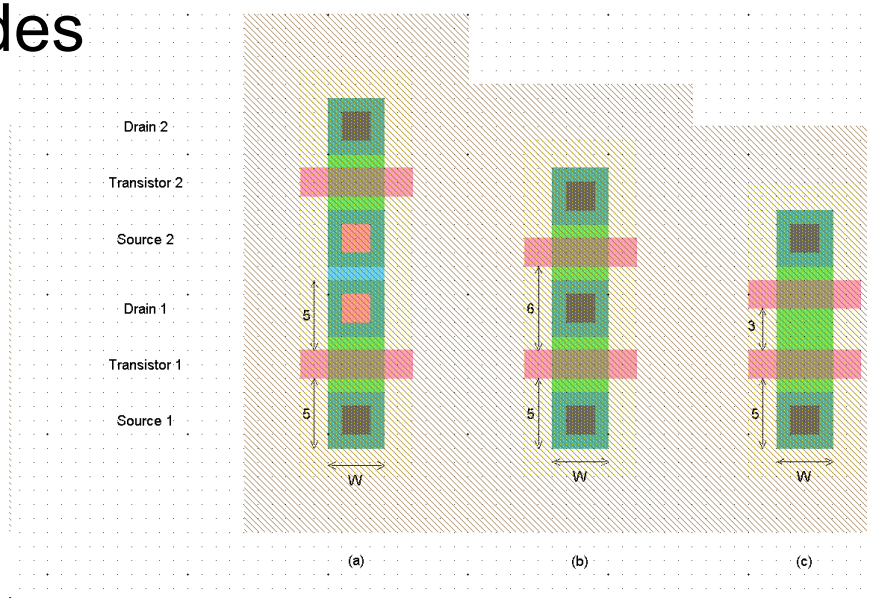
Gate Capacitance

- Approximate channel as connected to source
- $C_{gs} = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL = C_{permicron}W$
- $C_{permicron}$ is typically about 2 fF/μm

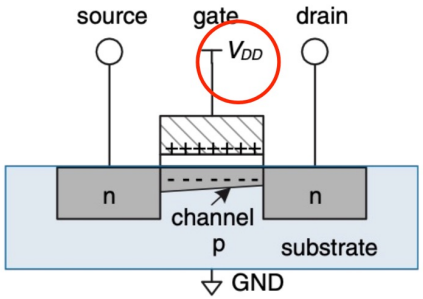


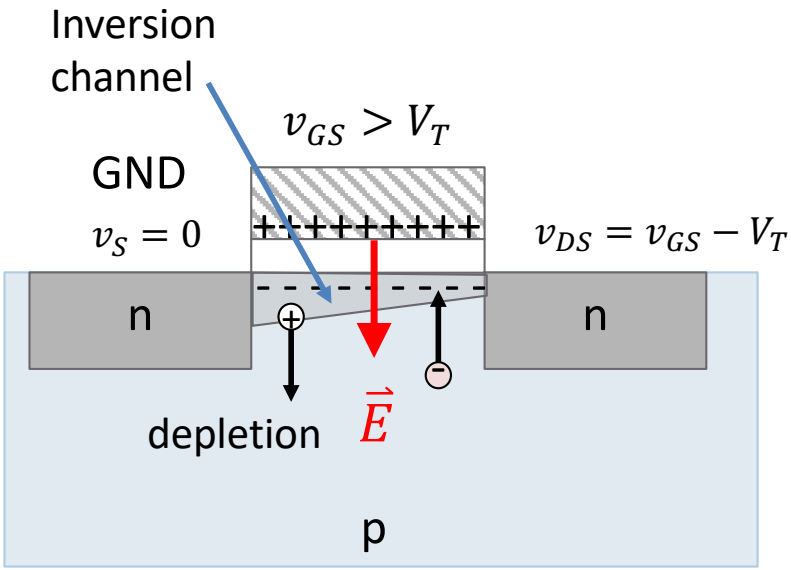
Diffusion Capacitance

- C_{sb}, C_{db}
- Undesirable, called *parasitic* capacitance
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to C_g for contacted diff
  - ½ C_g for uncontacted
  - Varies with process



nMOS transistor (summary)



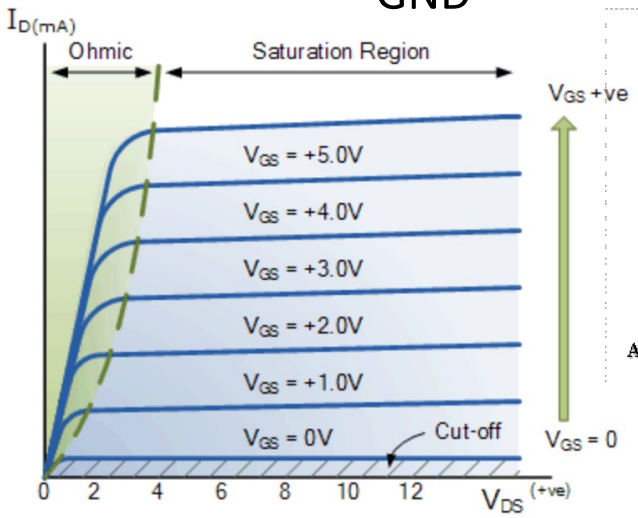


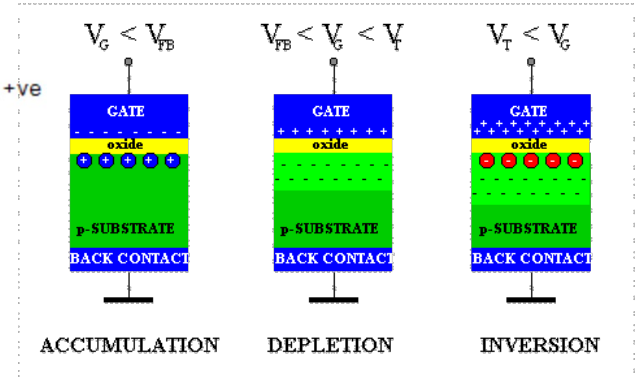
$$I_{DS} = \mu_{eff} C_{ox} \frac{W}{L} [(V_{GS} - V_T)V_{DS} - \tfrac{1}{2} V_{DS}^2]$$

$$R_{ON} = \frac{1}{\mu_n} C_{ox} \frac{W}{L} [(V_{GS} - V_T)V_{DS} - \tfrac{1}{2} V_{DS}^2]$$

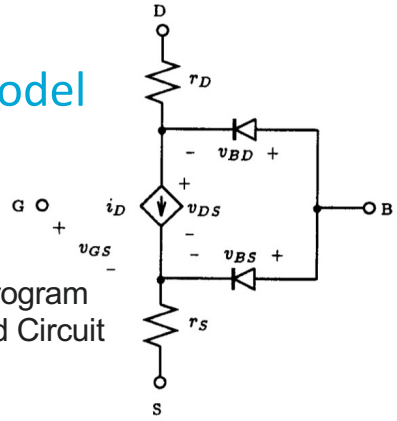
μ_{eff} is the effective mobility of the charge carrier.

'C_{OX}' is the capacitance of gate oxide for each unit area.





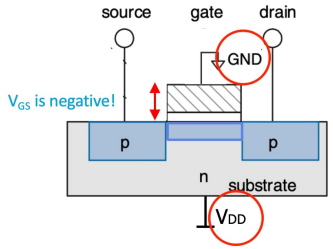
SPICE model

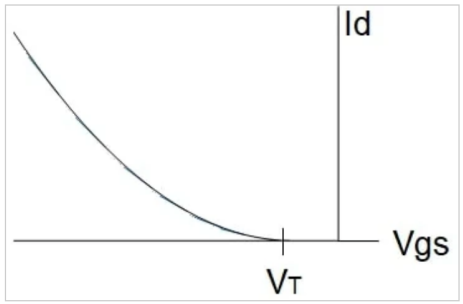


(Simulation Program with Integrated Circuit Emphasis) [1]

[1] http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE/

pMOS transistor (summary)

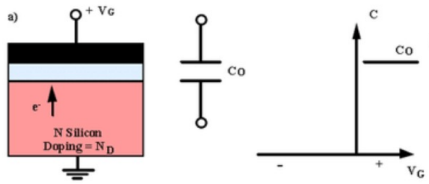




Negative v_{gs} voltage

- **DC bias + small AC high frequency signal applied.**

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

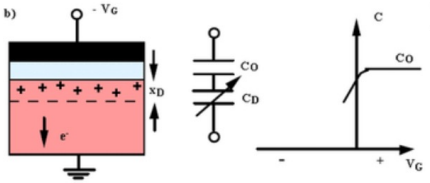


a) Accumulation majority carrier drawn to surface

$$C_{ox} = \frac{\varepsilon_{ox} \cdot A}{x_{ox}}$$

Charge Density

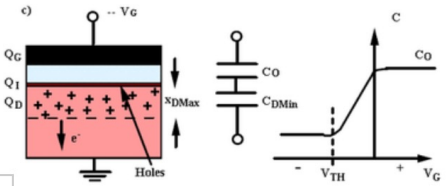
$$|Q_G| = |Q_D| = N_D x_D$$



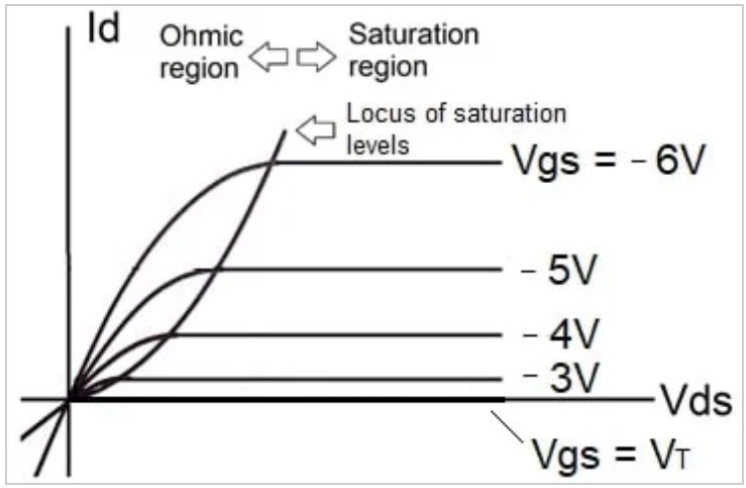
b) Depletion minority carrier drawn to surface majority carriers repelled

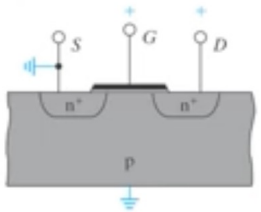
$$C_D = \frac{\varepsilon_{Si} \cdot A}{x_D}$$

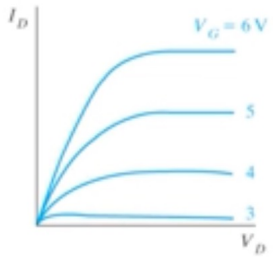
$$|Q_G| = N_D x_D + Q_I$$



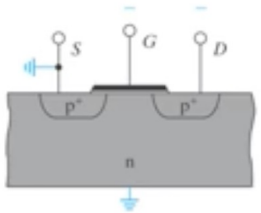
c) Inversion minority carrier layer exists

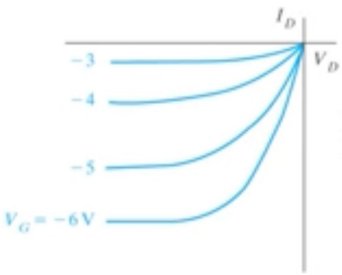






Enhancement Mode NMOS





Enhancement Mode PMOS

Complementary CMOS logic

- Full rail-to-rail swing; high noise margins
- Logic levels not dependent upon the relative device sizes; ratioless
- Always a path to V_{dd} or GND in steady state; low output impedance
- Extremely high input resistance; nearly zero steady-state input current
- No direct path steady state between power and ground; no static power dissipation*
- Propagation delay function of load capacitance and resistance of transistors

* well, this is gone for a long time now

Next Time

We continue with CMOS transistors

62