Lecture 2
CMOS Devices

Xuan 'Silvia' Zhang
Washington University in St. Louis

http://classes.engineering.wustl.edu/ese566/

Outline

Transistor Model

Wire Model

Delay Model

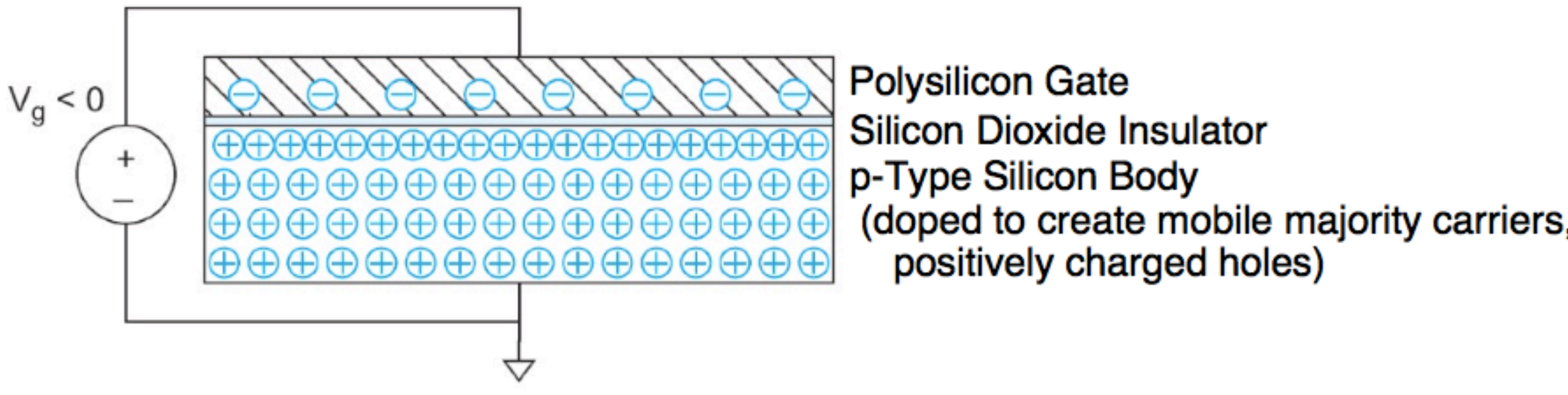
CMOS Fabrication

Fundamental Building Block: MOSFET

Core
Core
Core
Core
L2 Cache
L2 Cache
L2 Cache
L2 Cache
Local SMP Links
Mem Ctrl
L3 Cache and Chip Interconnect
Mem Ctrl
Remote SMP + I/O Links
L2 Cache
L2 Cache
L2 Cache
L2 Cache
Core
Core
Core
Core

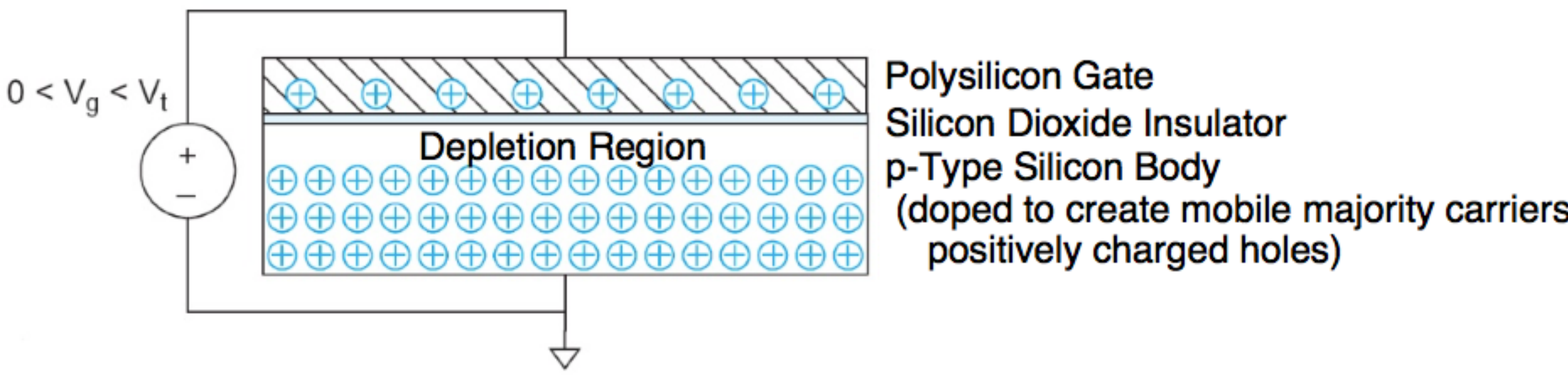
IBM Power 7
1.2 Billion
Transistors

Metal-Oxide-Semiconductor Structure



Accumulation:
Voltage source puts negative charge on gate, attracts positively-charged majority carriers in p-type silicon body

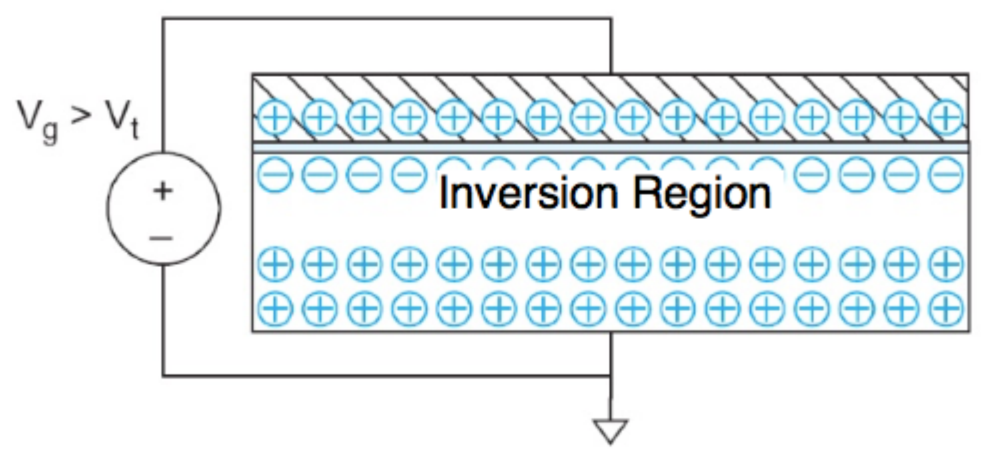
Metal-Oxide-Semiconductor Structure



Depletion:

Voltage source puts positive charge on gate, pushes positively-charged carriers away from surface, uncovers some negatively-charged dopant atoms in substrate

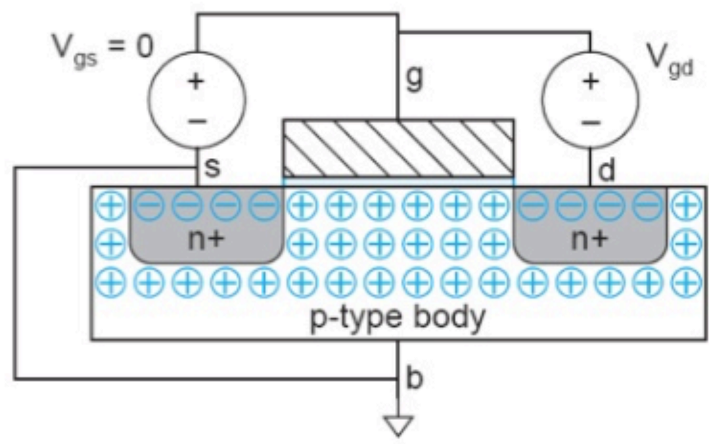
Metal-Oxide-Semiconductor Structure



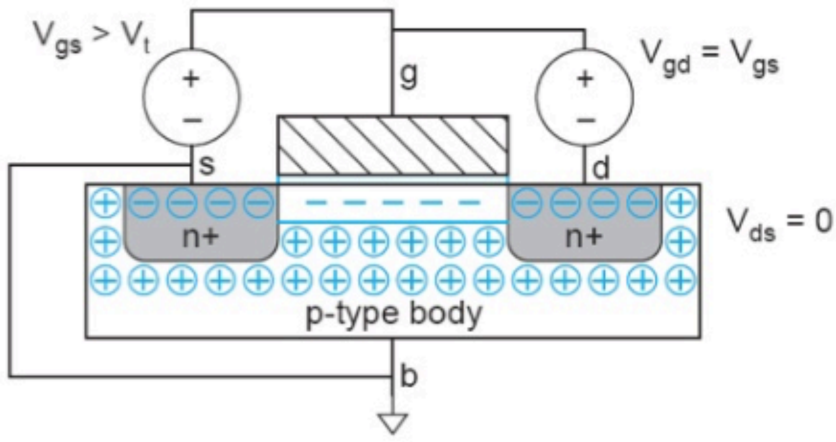
Polysilicon Gate
Silicon Dioxide Insulator
p-Type Silicon Body
(doped to create mobile majority carriers, positively charged holes)

Inversion:
Voltage source puts more positive charge on gate, instead of pushing holes even further away, draws free electrons to surface.

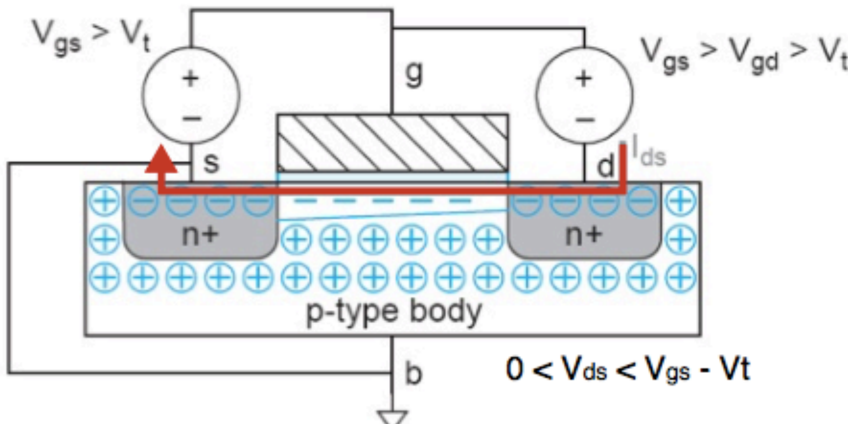
NMOS Transistor



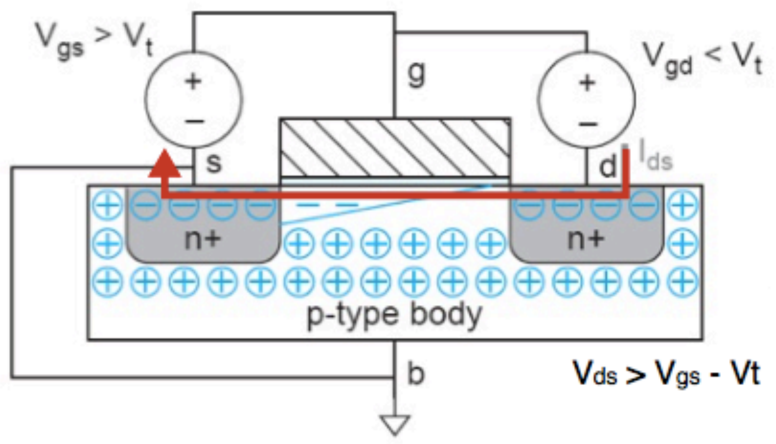
Cutoff: V_{gs} = 0V, Vds can be 0V or Vdd
No Channel, I_{ds} = 0



Linear: V_{gs} = Vdd, V_{ds} = 0V
Channel Formed, I_{ds} increases with V_{ds}

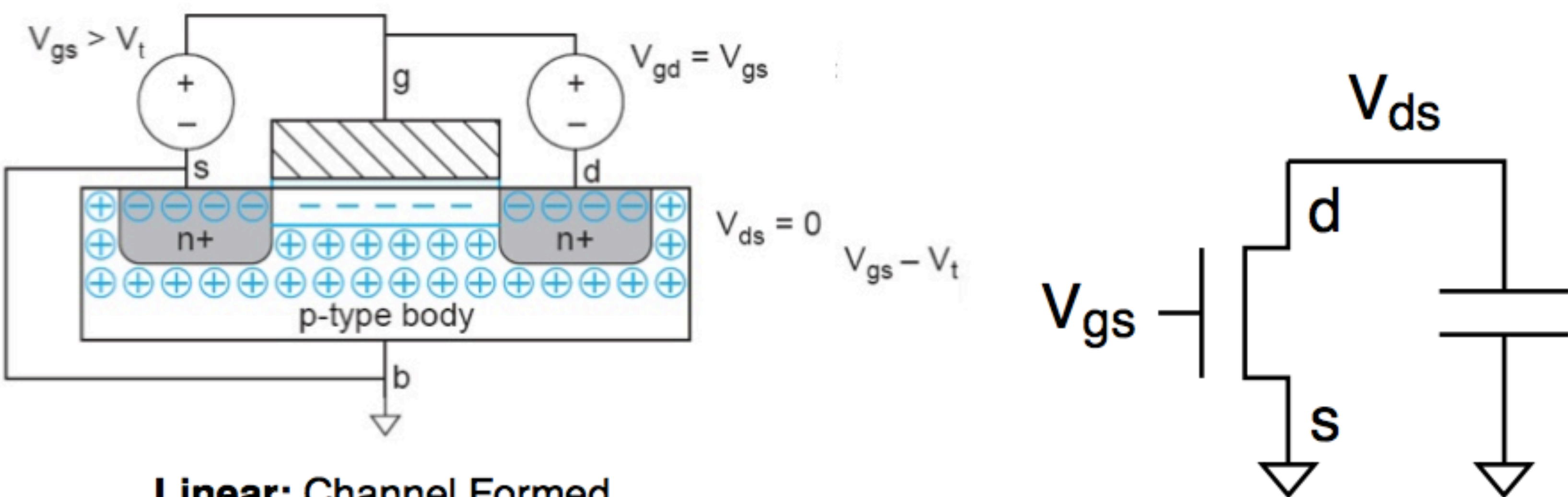


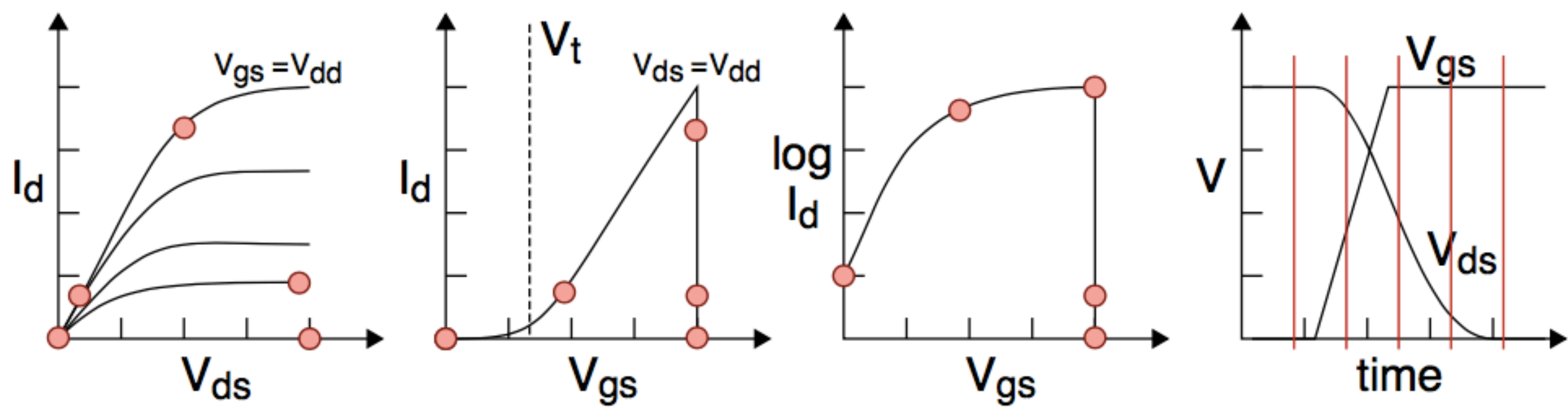
Linear: Vgs = Vdd, Vds = Vdd
Channel Formed, I_{ds} increases with V_{ds}



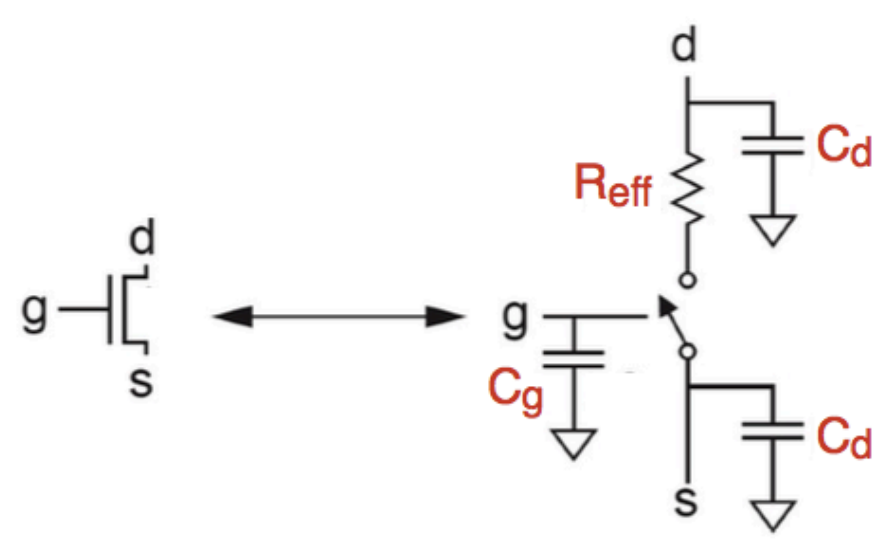
Saturation: Channel Pinched Off,
I_{ds} independent of V_{ds}

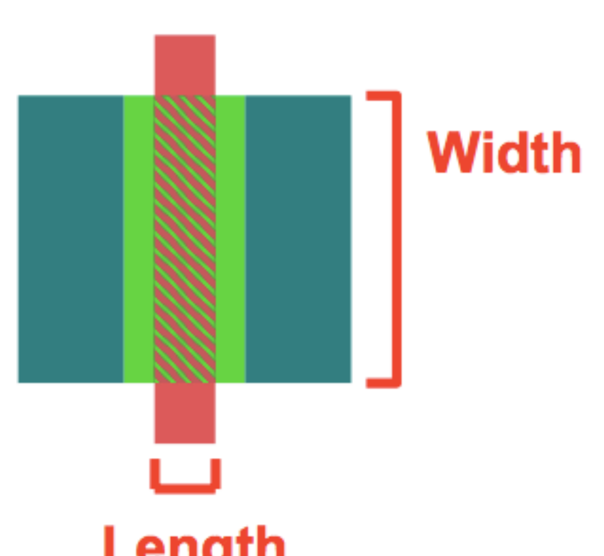
Simple NMOS Circuit





Key Qualitative Characteristics of MOSFET





- V_t sets when transistor turns on, impacts leakage current
- $I_d \propto \mu \times (W/L)$
- $\mu_n > \mu_p \implies R_{N,eff} < R_{P,eff}$
- $C_g \propto (W \times L)$
- $C_d \propto W$
- $\uparrow W = \downarrow R_{eff} = \uparrow I_d = \uparrow C_d, C_g$
- $\uparrow L = \uparrow R_{eff} = \downarrow I_d = \uparrow C_g$

Quantitative CMOS Model

- Threshold Voltage

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

$$\phi_0 = \frac{kT}{q}\ln\frac{N_A N_D}{n_i^2}$$

- I-V Curve
  - linear/triode $I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left(V_{gs} - V_{th}\right)V_{ds} - \frac{V_{ds}^2}{2}$
  - saturation $I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left(V_{gs} - V_{th}\right)^2\left(1 + \lambda V_{ds}\right)$
- Parasitic Capacitance
  - gate cap
  - junction cap $C_{j0} = \sqrt{\frac{\varepsilon_{si} q}{2}\left(\frac{N_A N_D}{N_A + N_D}\right)\frac{1}{\sqrt{\phi_0}}}$

Outline

Transistor Model

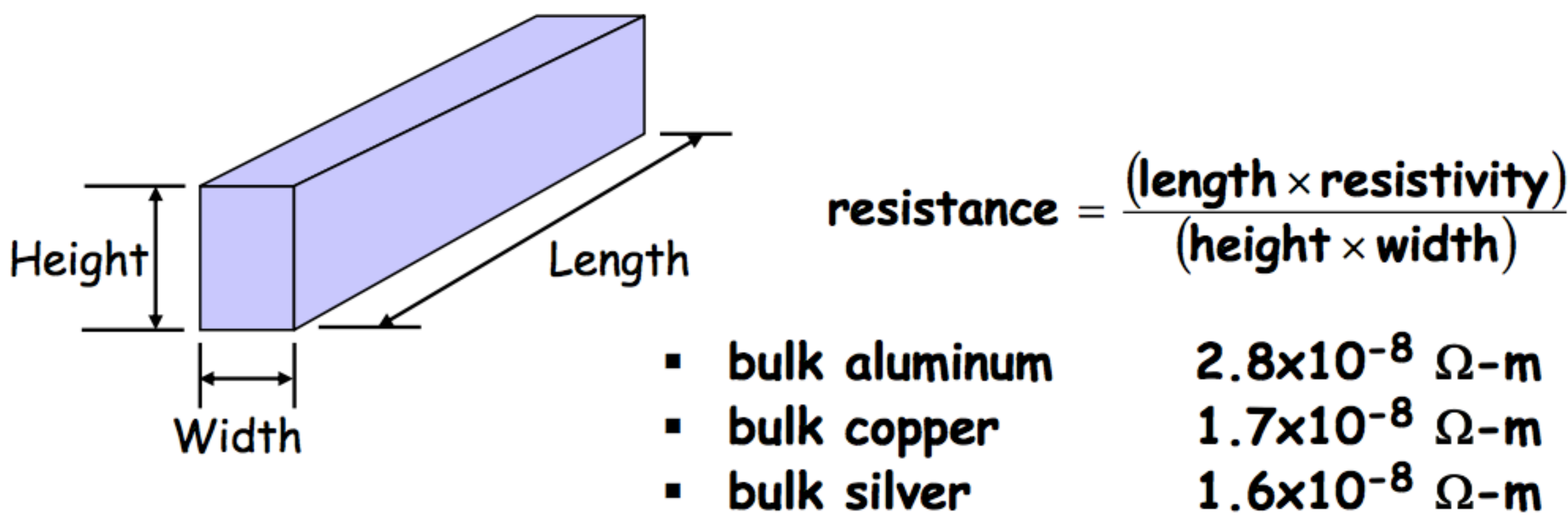
Wire Model

Delay Model

CMOS Fabrication

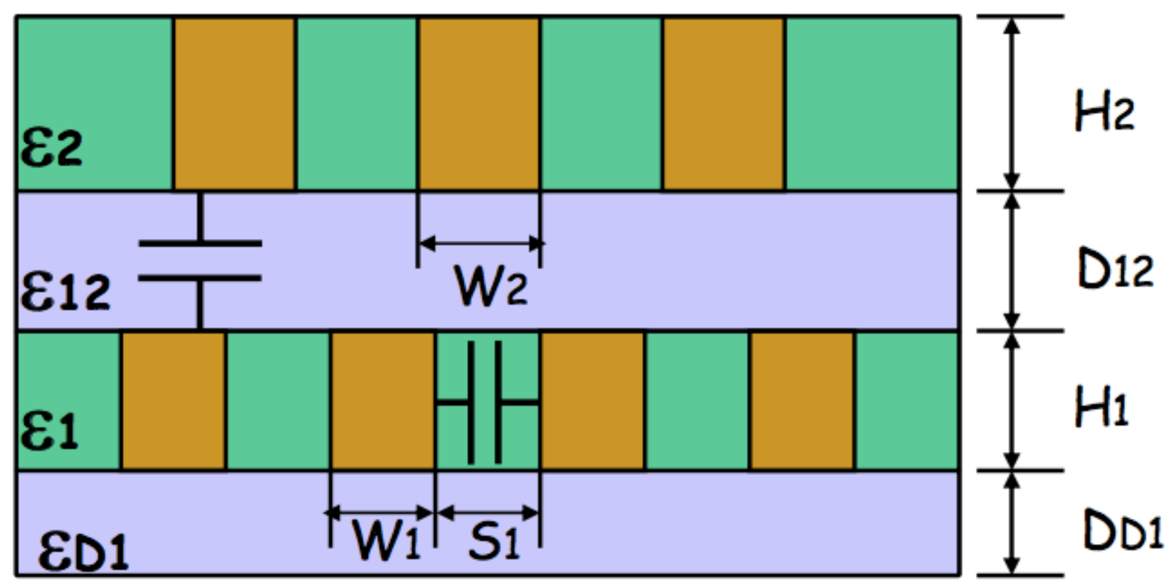
Wire Resistance

- Thickness fixed in given manufacturing process
- Resistance quoted as ohm/square
- TSMC 180nm 6 Aluminum metal layers
  - M1-M5: 0.08 ohm/square (0.5um x 1mm wire = 160 ohm)
  - M6: 0.03 ohm/square (0.5um x 1mm wire = 60 ohm)



Wire Capacitance

- Capacitance depends on geometry of surrounding wires and relative permittivity, ε_r, of dielectric
  - Silicon dioxide, SiO_2, $\varepsilon_r = 3.9$
  - Silicon flouride, SiOF, $\varepsilon_r = 3.1$
  - SiLK polymer, $\varepsilon_r = 2.6$



Outline

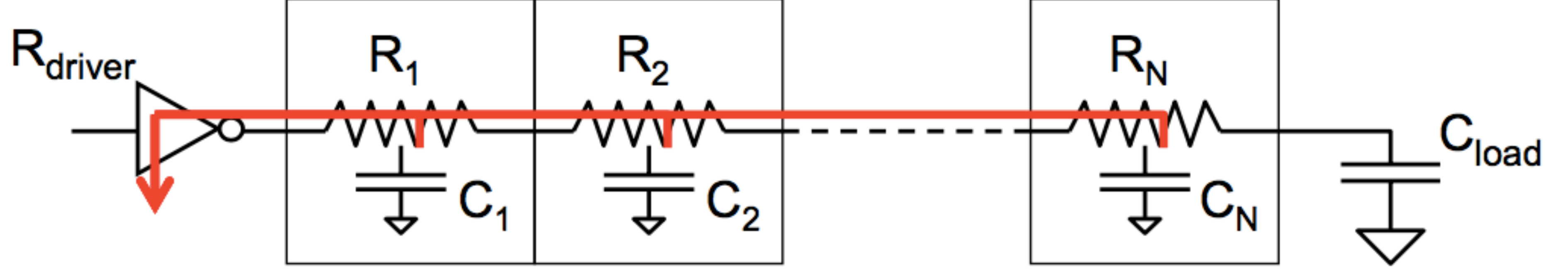
Transistor Model

Wire Model

Delay Model

CMOS Fabrication

Qualitative Characteristics of Wire Delay

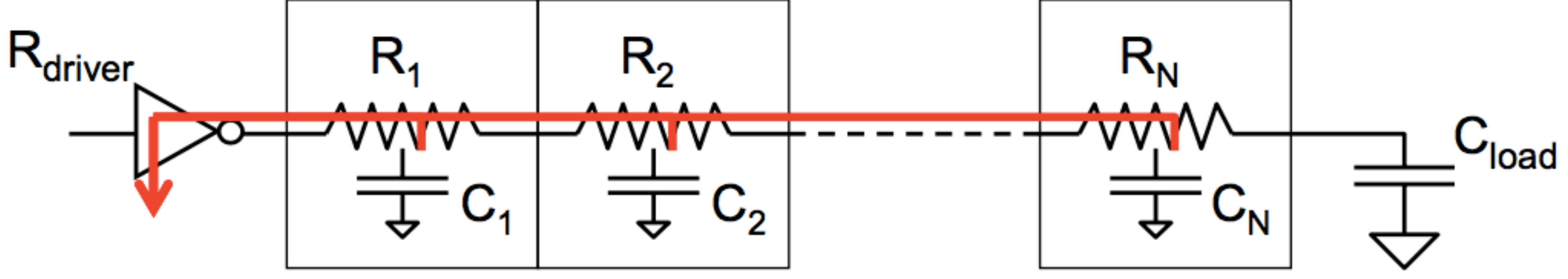


Wire resistance ∝ length
Wire capacitance ∝ length
➔
Wire delay ∝ length^2

Rdriver
Rw
Cw/2
Cw/2
Cload

Quantitative Delay Model

- RC Ladder
  - Elmore delay model



$$t_{pd} = \sum_{nodes\, i} R_{i-to-source} C_i$$

$$= R_1C_1 + (R_1 + R_2)C_2 + \cdots + (R_1 + R_2 + \cdots + R_N)C_N$$

Outline

Transistor Model

Wire Model

Delay Model

CMOS Fabrication

Mask Set for NMOS (circa 1986)

Vd = 1V

Vg = 0V

Vs = 0V

I ≈ nA

dielectric

n+

n+

p-

Top-down view:

Masks

- #1: n+ diffusion
- #2: poly (gate)
- #3: diff contact
- #4: metal

Layers to do p-Fet not shown. Modern processes have 6 to 10 metal layers (or more) (in 1986: 2).

Design Rules for Masks (circa 1986)

Poly overhang. So that if masks are misaligned, channel doesn't short out.

Minimum gate length. So that the source and drain depletion regions do not meet!

length

Metal rules: Contact separation from channel, one fixed contact size, overlap rules with metal, etc ...

#1: n+ diffusion
#2: poly (gate)
#3: diff contact
#4: metal

Start with an Un-Doped Wafer

UV hardens exposed resist. A wafer wash leaves only hard resist.

oxide

p-

Steps

#1: dope wafer p-

#2: grow gate oxide

#3: deposit undoped polysilicon

#4: spin on photoresist

#5: place positive poly mask and expose with UV.

Wet Etch to Remove Unmasked Regions

HF acid etches through poly and oxide, but not hardened resist.

oxide

p-

oxide

p-

After etch and resist removal

Use Diffusion Mask to Implant N-Type

accelerated donor atoms

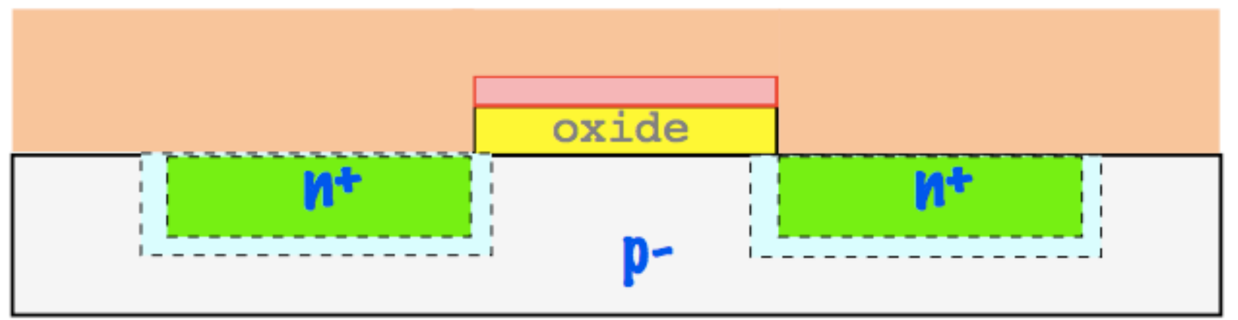
Arsenic
As
74.922

oxide
n+
n+
p-

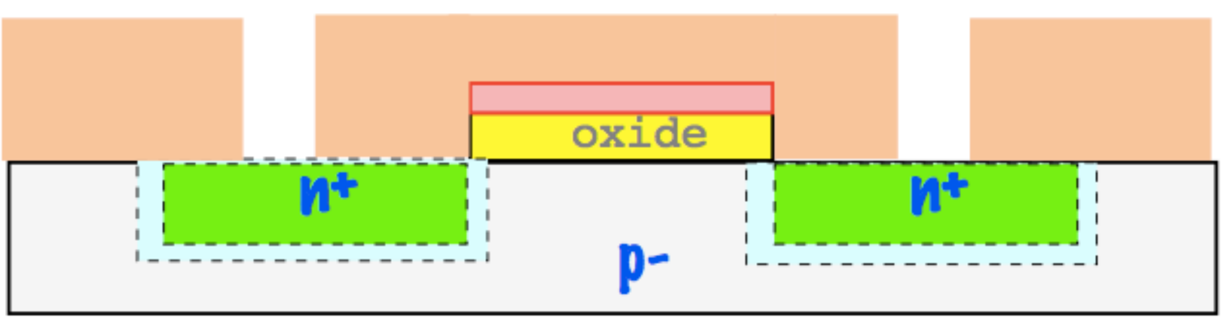
Notice how donor atoms are blocked by gate and do not enter channel.

Thus, the channel is "self-aligned", precise mask alignment is not needed!

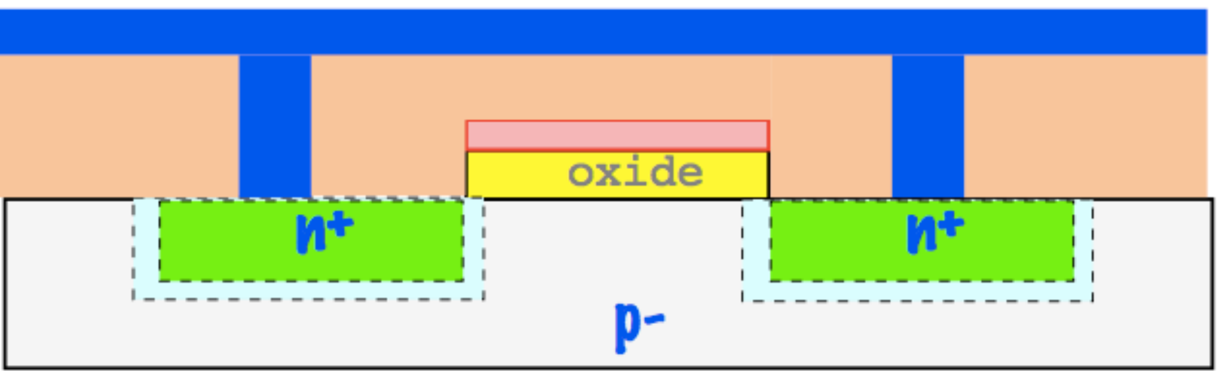
Metallization Completes Device



Grow a thick oxide on top of the wafer.

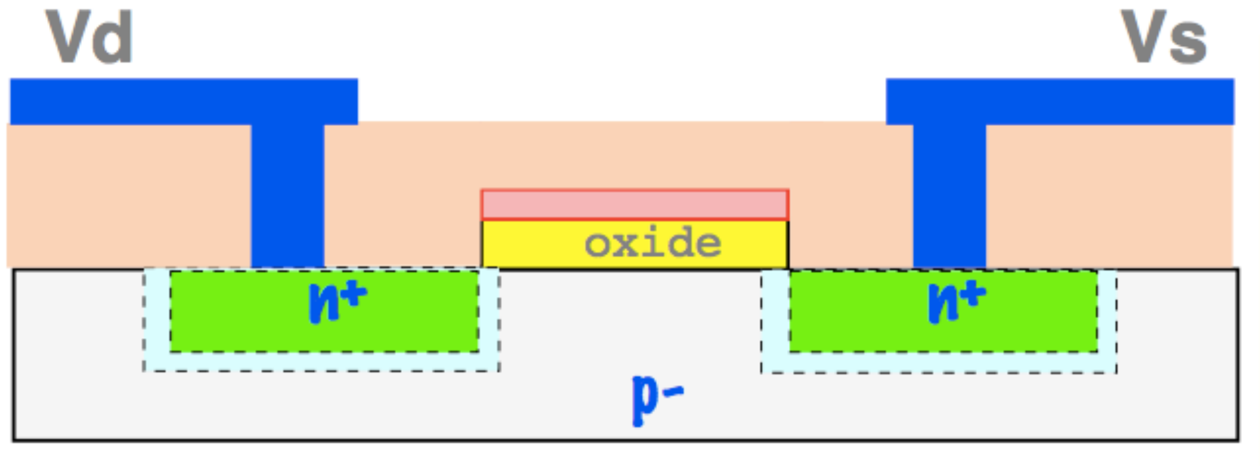


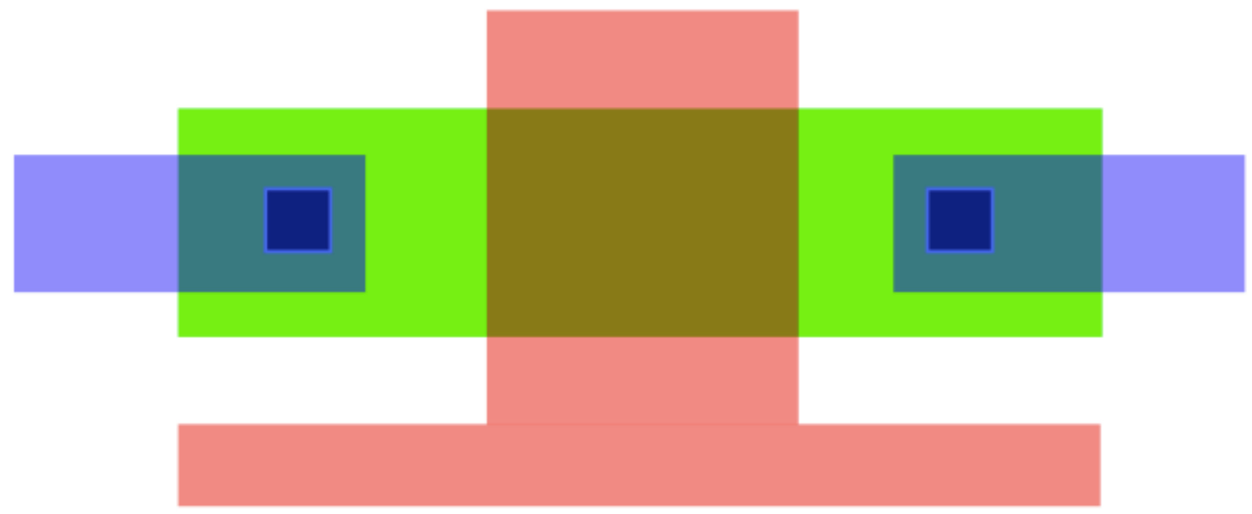
Mask and etch to make contact holes



Put a layer of metal on chip. Be sure to fill in the holes!

Final NMOS Transistor





"The planar process"

Jean Hoerni, Fairchild Semiconductor 1958

PMOS—Dual of NMOS

V_{well} = Vs = 1V

Vg = 0V

Vd = 0V

↓I ≈ μA

dielectric

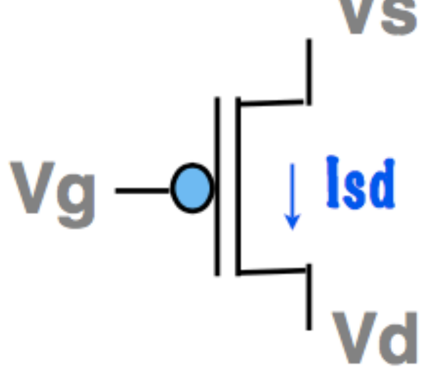
p+

p+

n-well

p-

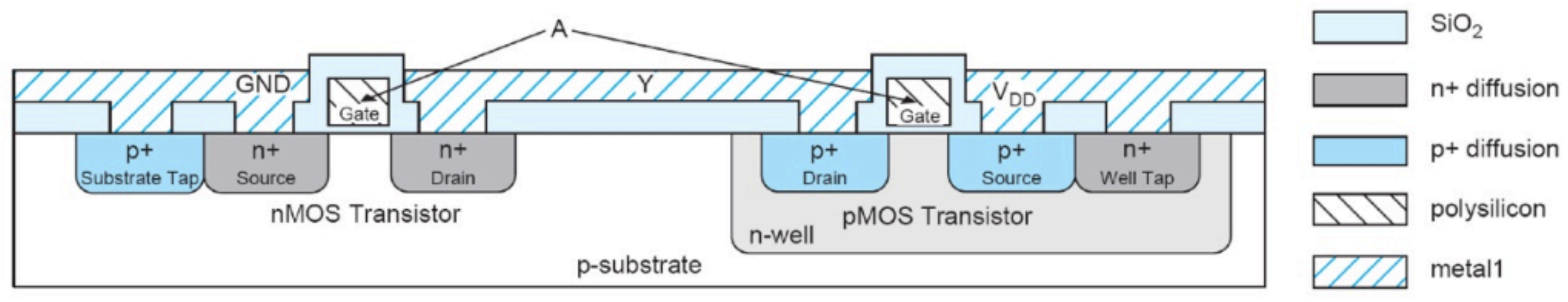
New "n-well" mask



"Mobility" of holes is slower than electrons.

p-Fets drive less current than n-Fets, all else being equal

Single- and Triple-Well Process



nMOS Transistor

pMOS Transistor

p+ n+ n+

p+ p+ n+

p-well

n-well

Deep n-well

p-substrate

Local Interconnect

metal1
V_{DD}
Unit Cell
Contact Stud
Polysilicon
Wordline

Local interconnect
between n and p
diffusion
n-diffusion

IBM 6-Transistor SRAM Cell

Intel Metal Stacks: 90nm and 45nm

M6
M5
M4
M3
M2
M1
Transistors

(a)

1 μm

M8
M7
M6
M5
M4
M3
M2
M1
Transistors

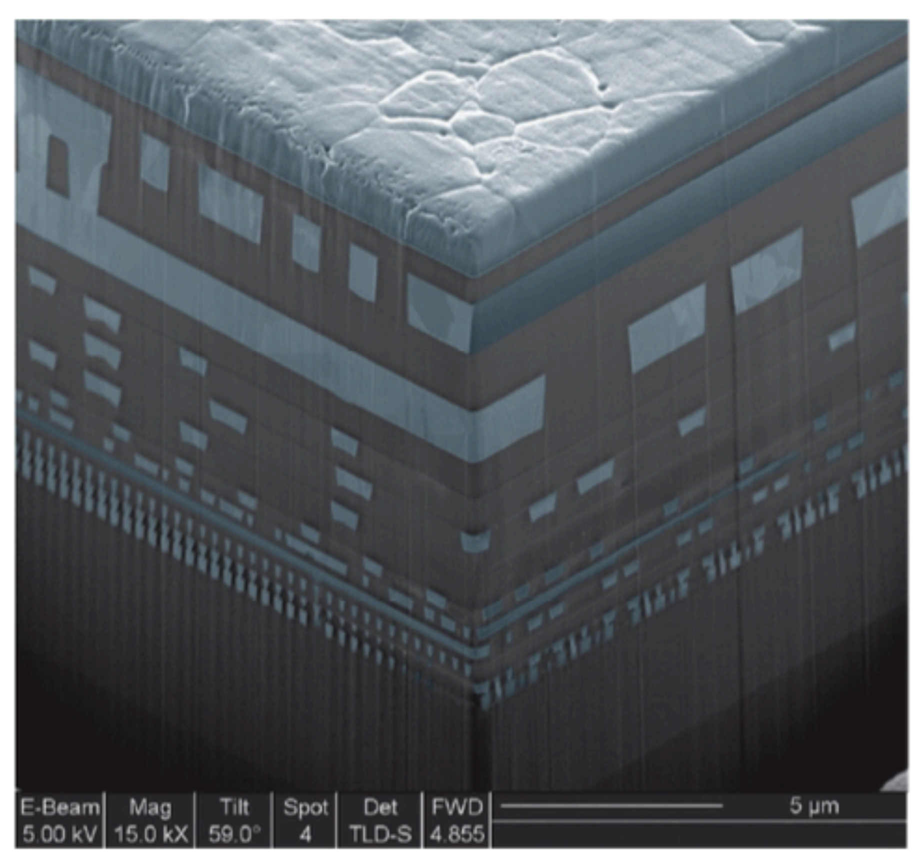
(b)

Intel Metal Stacks: 45nm with M9 and I/O Bump

Intel Metal Layer Dimensions in 45nm

Layer	t (nm)	w (nm)	s (nm)	pitch (nm)
M9	7µm	17.5µm	13µm	30.5µm
M8	720	400	410	810
M7	504	280	280	560
M6	324	180	180	360
M5	252	140	140	280
M4	216	120	120	240
M3	144	80	80	160
M2	144	80	80	160
M1	144	80	80	160

IBM Metal Stack

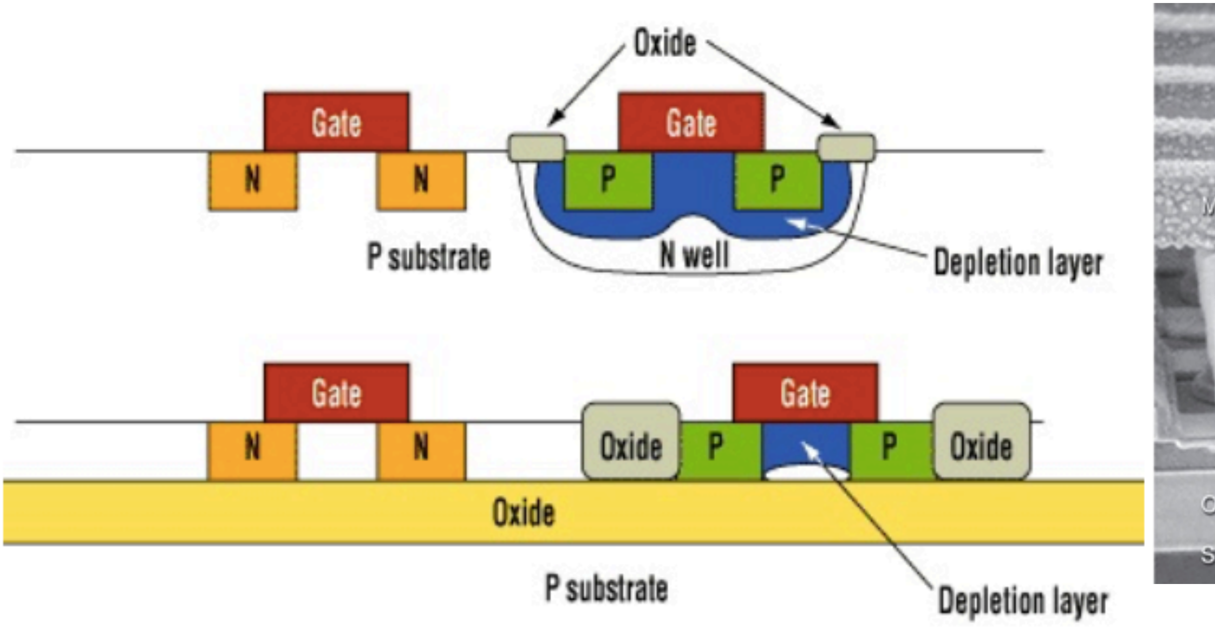


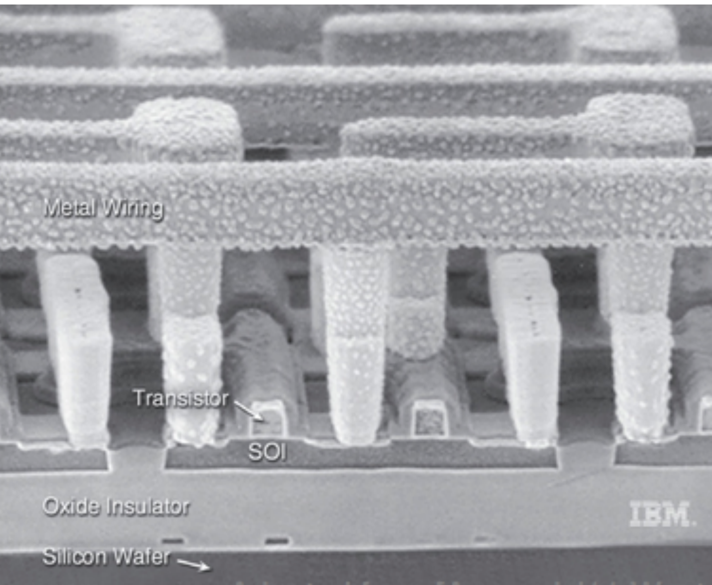
IBM 11-layer Copper Metal Stack

IBM 6-layer Copper Metal Stack

Bulk vs Silicon-on-Insulator (SOI) Processing

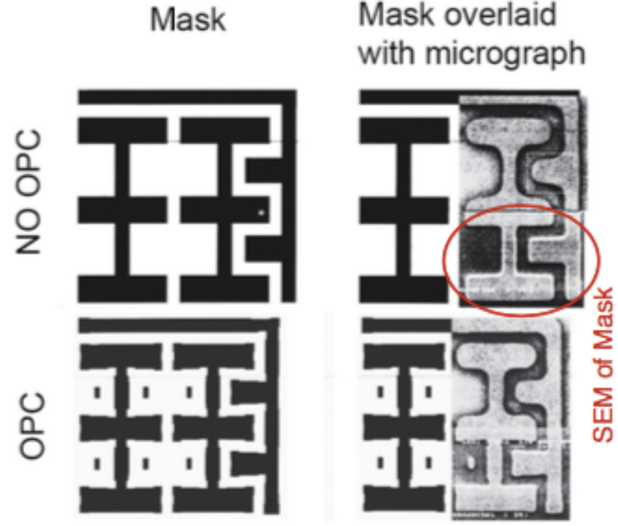
- Eliminate parasitic cap between S/D and body
  - lower energy, higher performance
- Lower sub-Vth leakage, but Vth varies over time
- 10-15% increase in total manufacturing cost
  - due to substrate cost

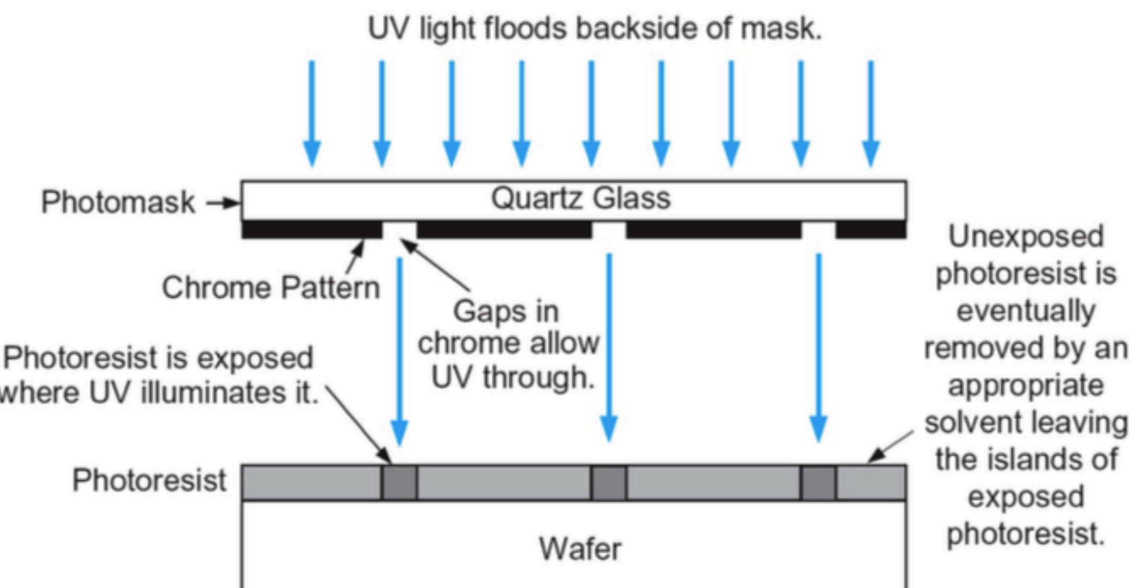


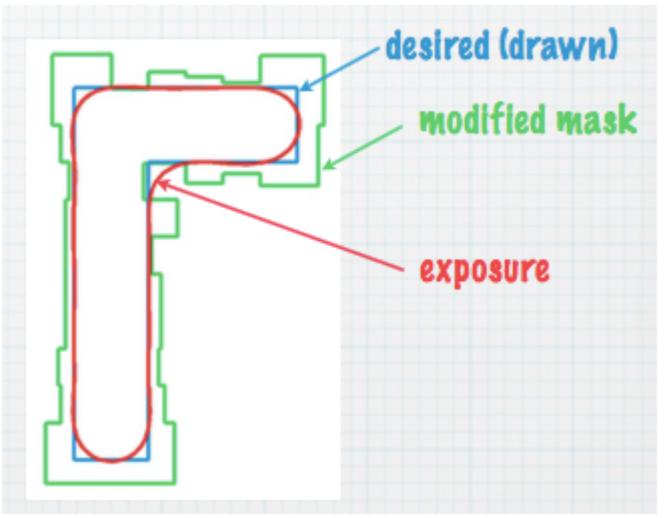


Lithography

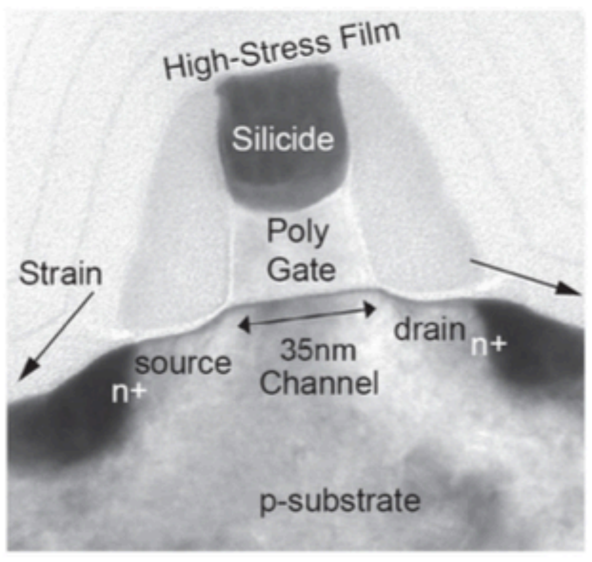
- Pattern resolution exceeds wavelength of light
  - 193nm from argon fluoride laser
- Sophisticated patterning tricks
  - immersion lithography
  - optical proximity correction (OPC)
  - double patterning

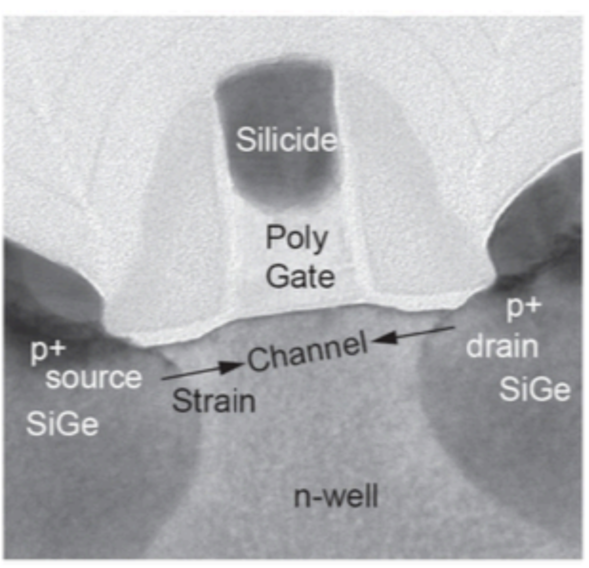






Processing Enhancements

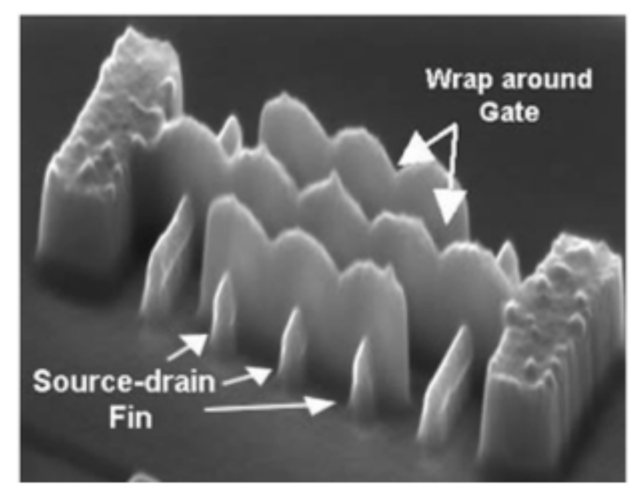




- **High-K Dieletrics and Metal Gates –** Replacing silicon dioxide gate dielectric with a high-K material allows increased vertical electric field without increasing gate leakage
- **Strained Silicon** – Layer of silicon in which silicon atoms are stretched beyond their normal interatomic distance leading to better mobility
- **Gate Engineering** – Multiple transistor designs with different threshold voltages to allow optimization of delay or power

FinFET Transistors

- Small footprint, good gate control
  - use the vertical dimension
  - Intel starts using FinFET in 20nm



Gate Oxide
Drain
Gate
Source
W

Cross-Section
Gate
Gate Oxide
Source Drain
BOX

Top View
Source
Gate
L
Drain

Questions?

Comments?

Discussion?

Acknowledgement

N. Weste and D. Harris, “CMOS VLSI Design”, 2011

Cornell University, ECE 5745

UC Berkeley, CS 230

MIT, 6.371