VLSI Design

Introduction

Outline

- Introduction
- Silicon, pn-junctions and transistors
- A Brief History
- Operation of MOS Transistors
- CMOS circuits
- Fabrication steps for CMOS circuits

Introduction

- Integrated circuits: many transistors on one chip.

- *Very Large Scale Integration* (VLSI)

- *Complementary Metal Oxide Semiconductor* (CMOS)

  - Fast, cheap, "low-power" transistors circuits

# WHY VLSI DESIGN?

Money, technology, civilization

Annual Sales

- 10^{18} transistors manufactured in 2003
  - 100 million for every human on the planet
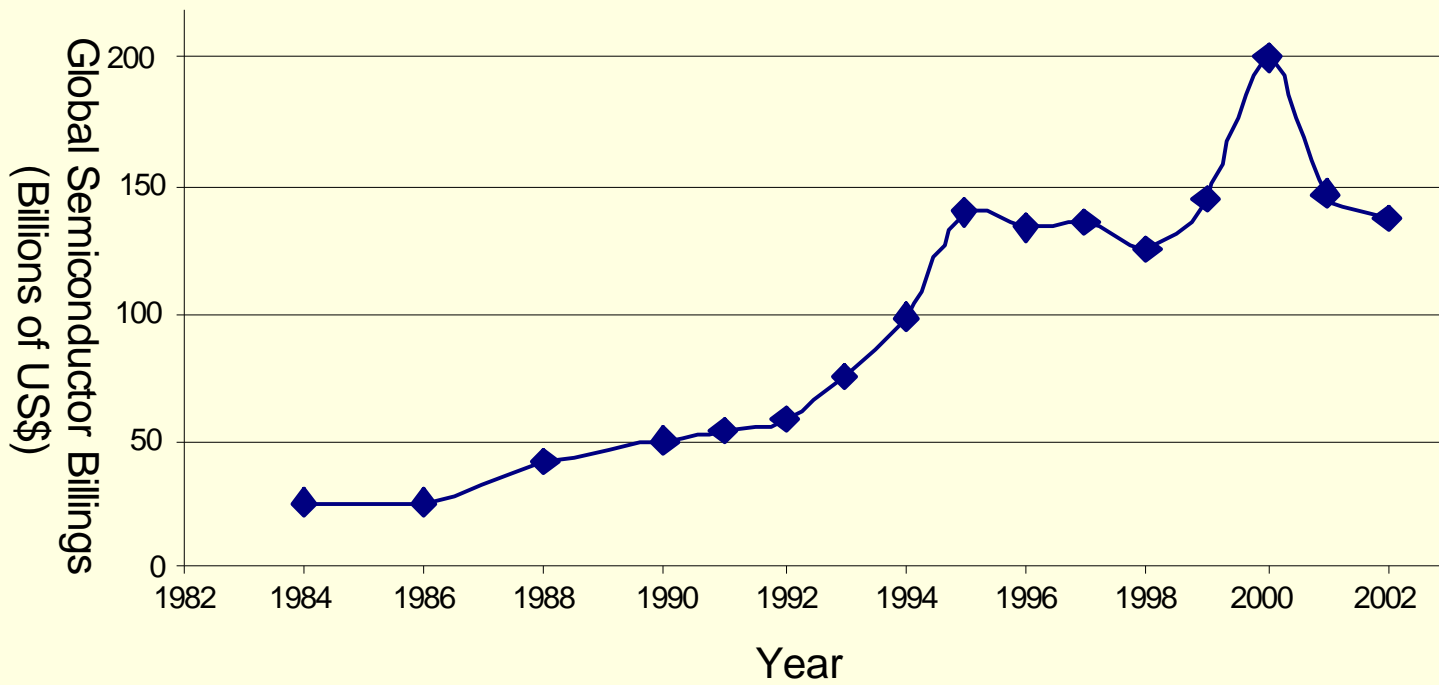
Digression: Silicon Semiconductors

- Modern electronic chips are built mostly on silicon substrates
- Silicon is a Group IV semiconducting material
- crystal lattice: covalent bonds hold each atom to four neighbors

http://onlineheavytheory.net/silicon.html
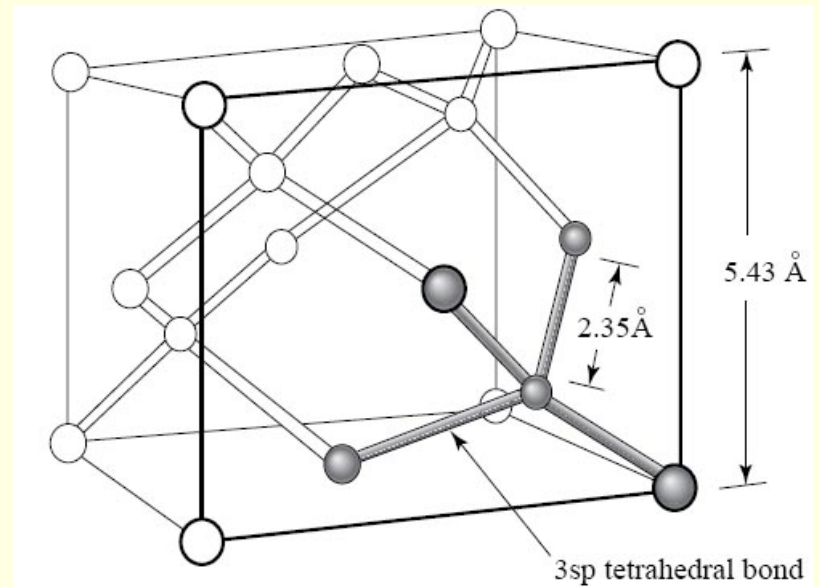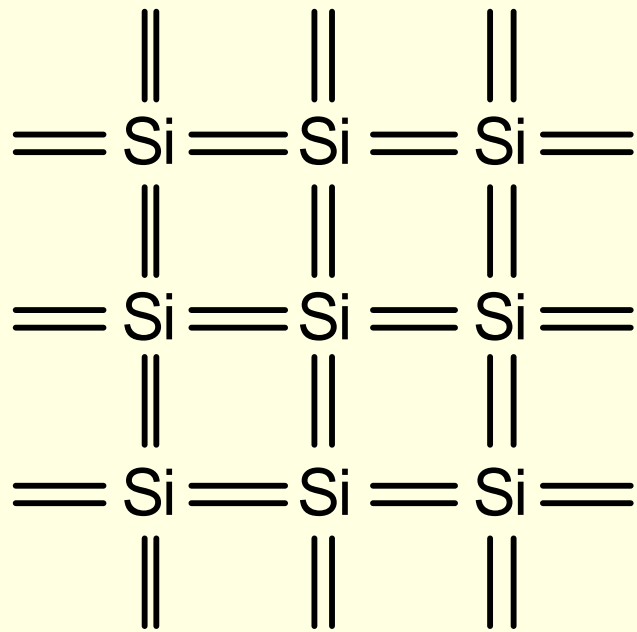
Dopants

- Silicon is a semiconductor at room temperature
- Pure silicon has few free carriers and conducts poorly
- Adding dopants increases the conductivity drastically
- Dopant from Group V (e.g. As, P): extra electron (n-type)
- Dopant from Group III (e.g. B, Al): missing electron, called hole (p-type)
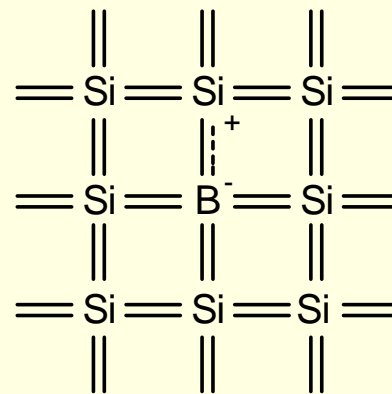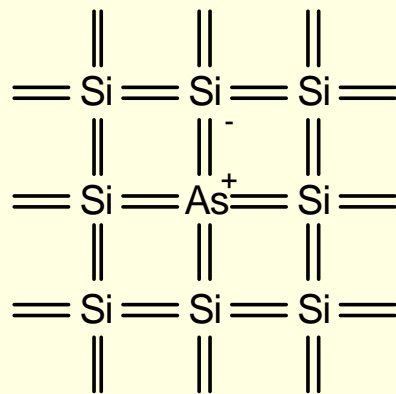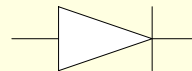
p-n Junctions

- First semiconductor (two terminal) devices
- A junction between p-type and n-type semiconductor forms a diode.
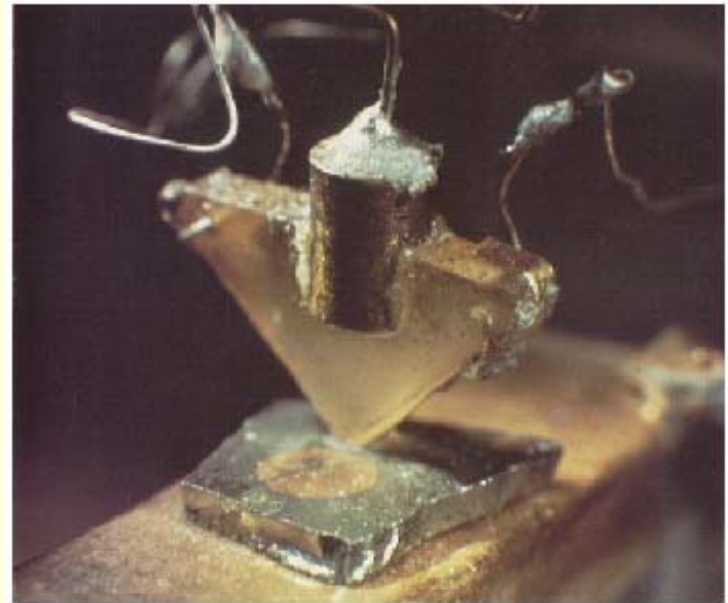- Current flows only in one direction

p-type	n-type
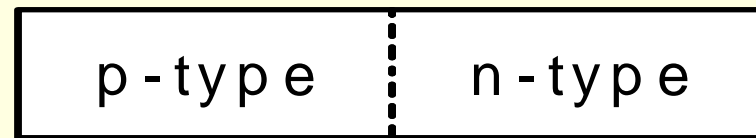
anode   cathode

A Brief History
Invention of the Transistor

- Vacuum tubes ruled in first half of 20th century Large, expensive, power-hungry, unreliable

- 1947: first point contact transistor (3 terminal devices)

  - Shockley, Bardeen and Brattain at Bell Labs

A Brief History, contd..

❑ 1958: First integrated circuit

   ▪ Flip-flop using two transistors
   ▪ Built by Jack Kilby (Nobel Laureate) at Texas Instruments
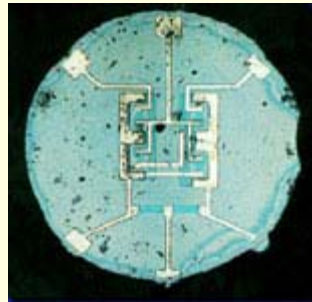   ▪ Robert Noyce (Fairchild) is also considered as a co-inventor
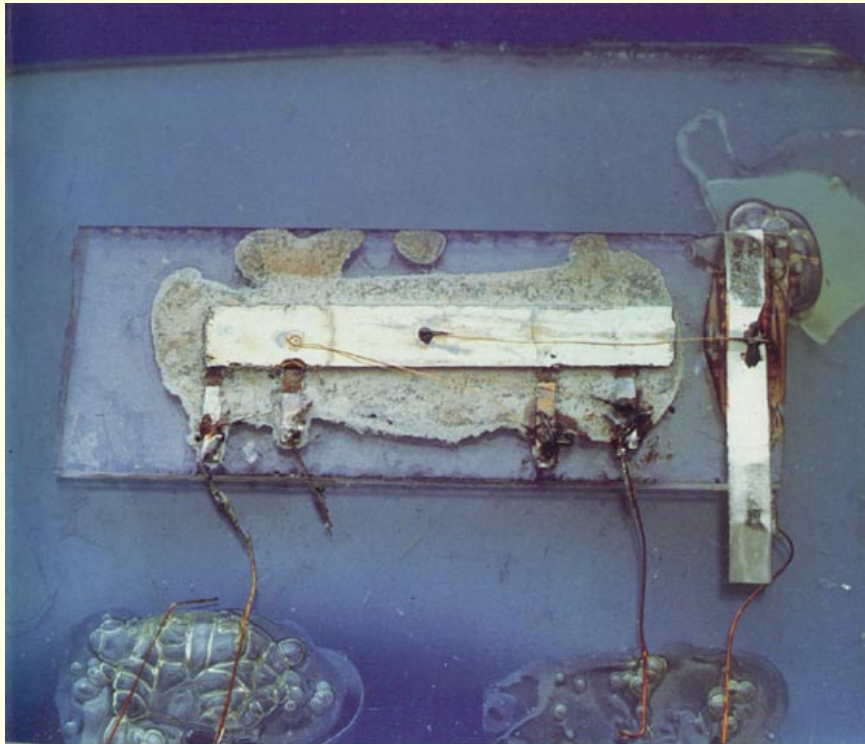
Kilby's IC

smithsonianchips.si.edu/ augarten/

# A Brief History, contd.

- First Planer IC built in 1961

- 2003
  - Intel Pentium 4 μprocessor (55 million transistors)
  - 512 Mbit DRAM (> 0.5 billion transistors)
- 53% compound annual growth rate over 45 years
  - No other technology has grown so fast so long
- Driven by miniaturization of transistors
  - Smaller is cheaper, faster, lower in power!
  - Revolutionary effects on society

MOS Integrated Circuits

❑ 1970's processes usually had only nMOS transistors

Inexpensive, but consume power while idle
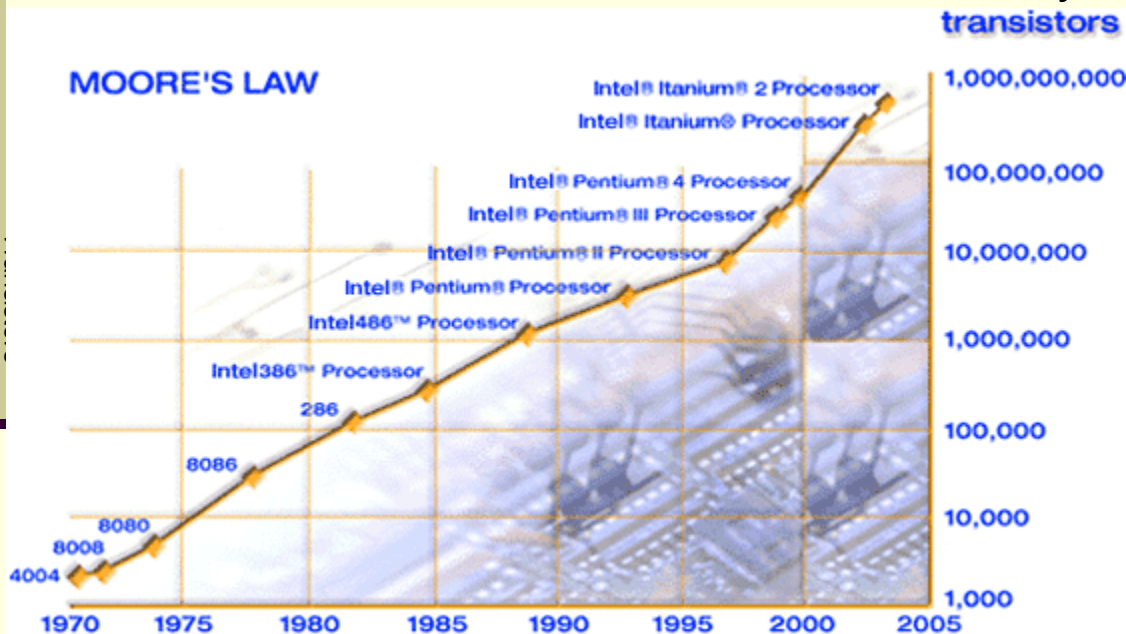
❑ 1980s-present: CMOS processes for low idle power
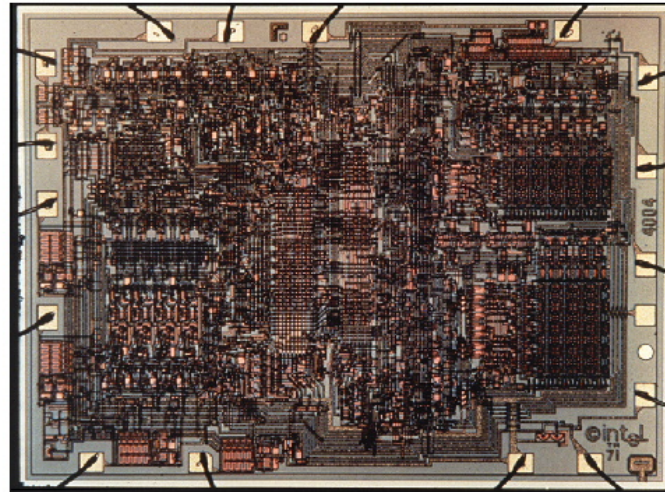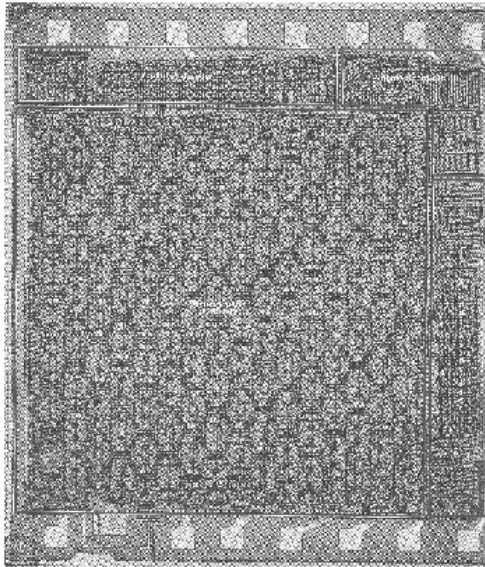
Intel 1101 256-bit SRAM   Intel 4004 4-bit μProc

Moore's Law

- 1965: Gordon Moore plotted transistor on each chip
  - Fit straight line on semilog scale
  - Transistor counts have doubled every 26 months

Integration Levels
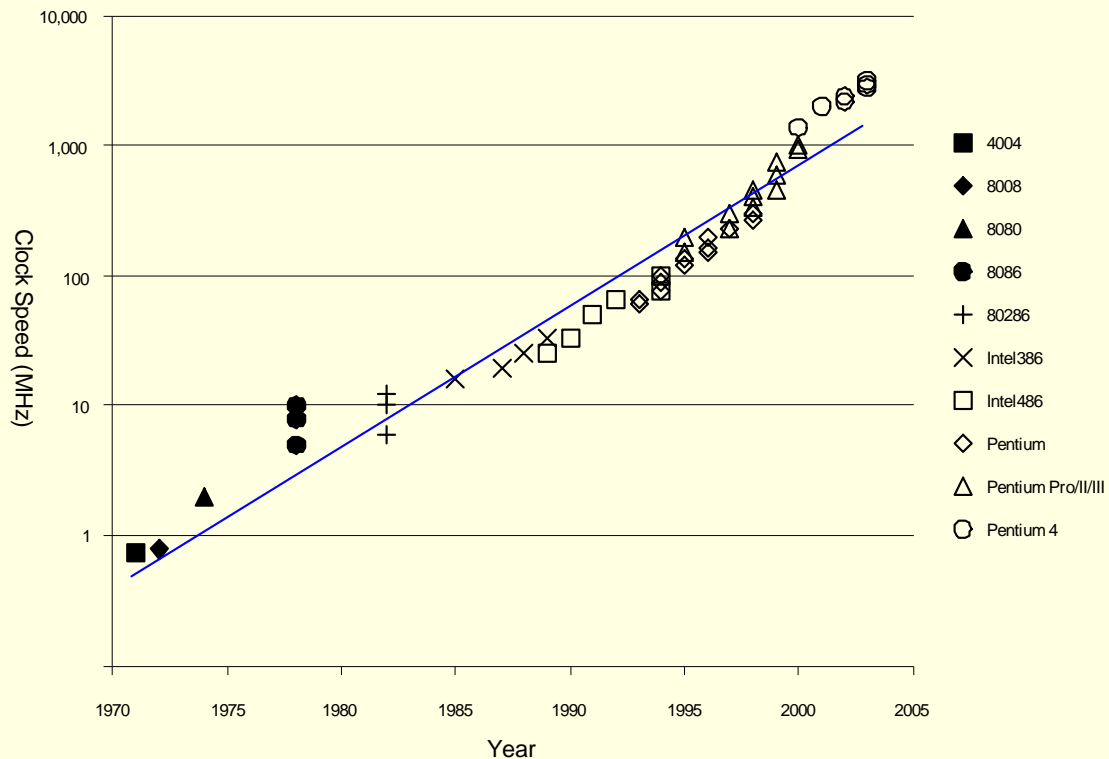
SSI: 10 gates

MSI: 1000 gates

LSI: 10,000 gates

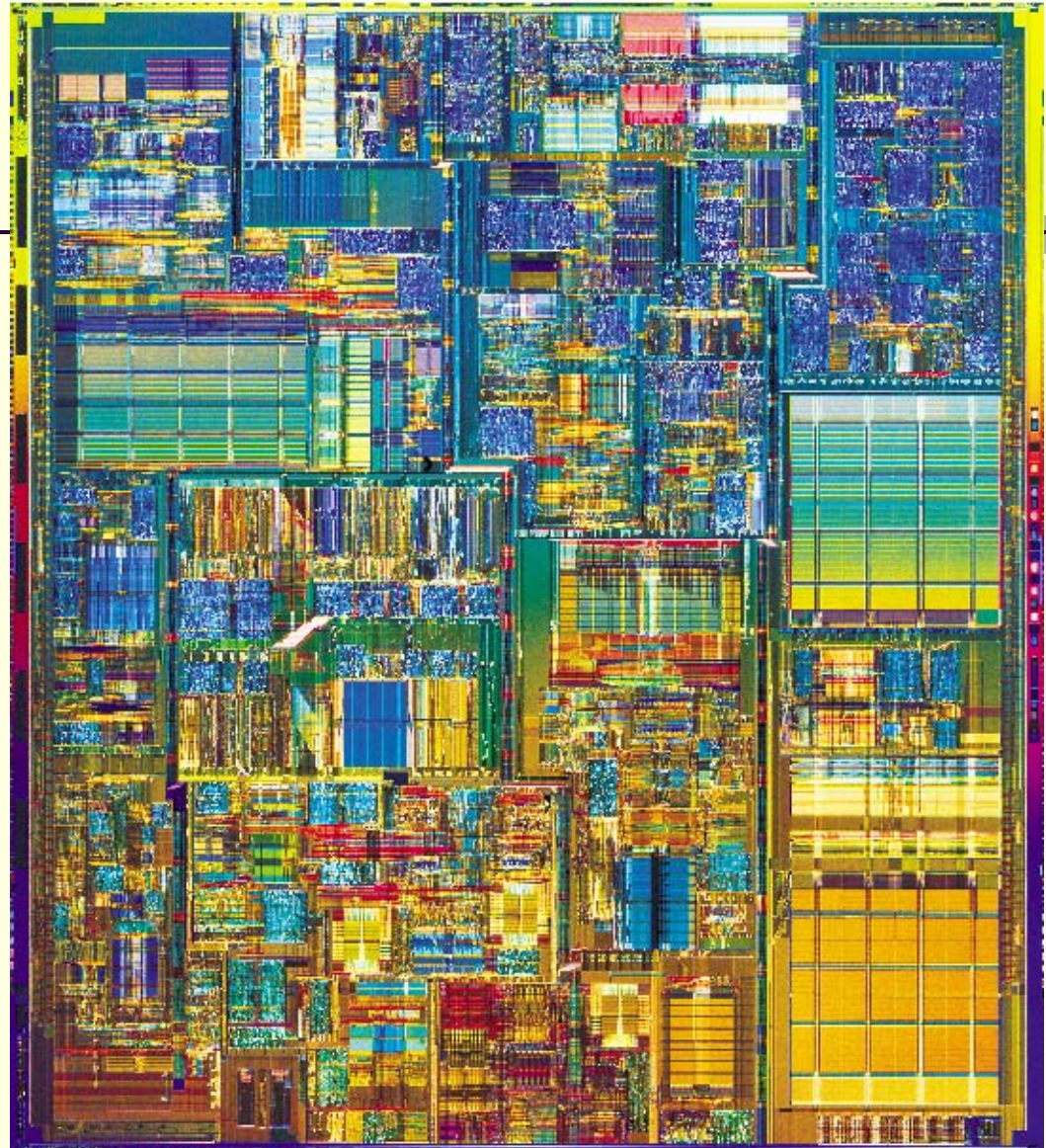**VLSI**: > 10k gates

http://www.intel.com/technology/silicon/mooreslaw/

Corollaries

- **Many other factors grow exponentially**
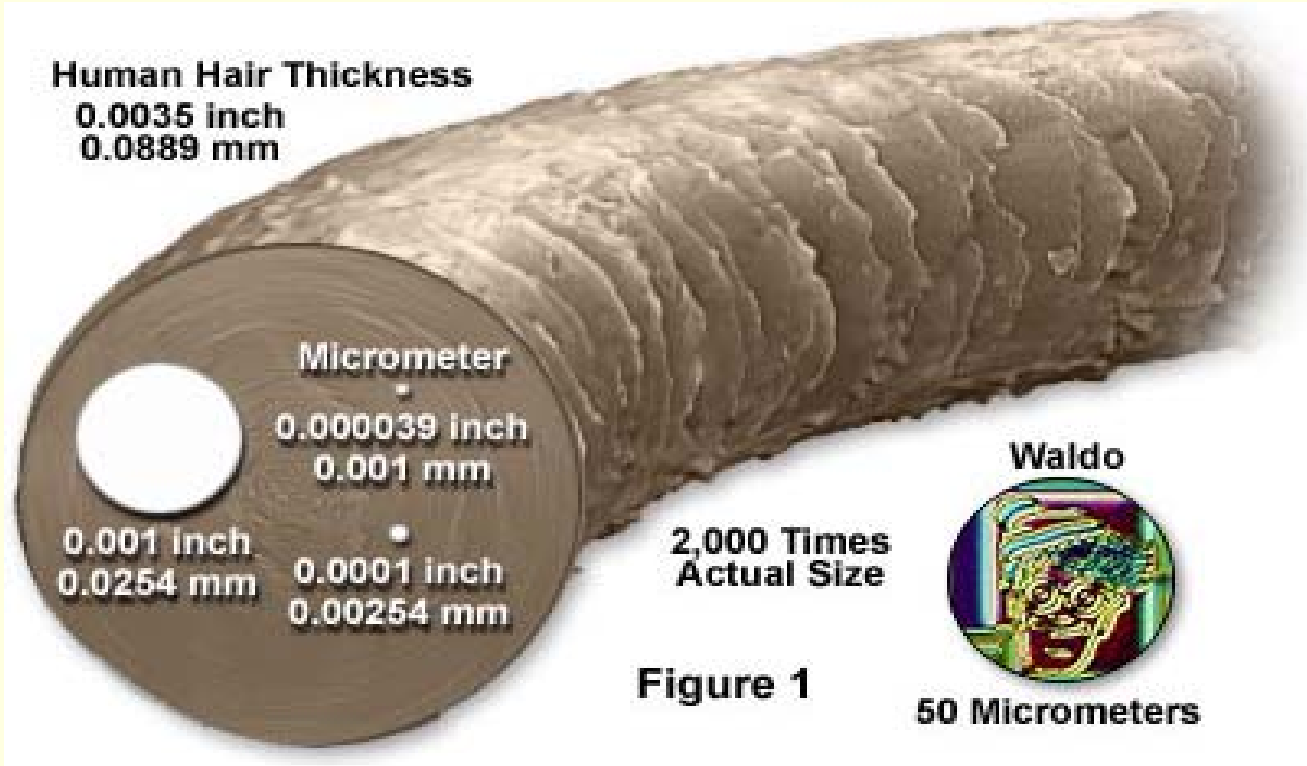 - Ex: clock frequency, processor performance

Pentium 4 Processor

http://www.intel.com/intel/intelis/museum/online/hist_micro/hof/index.htm

- Modern transistors are few microns wide and approximately 0.1 micron or less in length
- Human hair is 80-90 microns in diameter

Human Hair Thickness
0.0035 inch
0.0889 mm

Micrometer
0.000039 inch
0.001 mm

0.001 inch
0.0254 mm

0.0001 inch
0.00254 mm

2,000 Times Actual Size

Waldo
50 Micrometers

Figure 1

Transistor Types

- Bipolar transistors
  - npn or pnp silicon structure
  - Small current into very thin base layer controls large currents between emitter and collector
  - Base currents limit integration density
- Metal Oxide Semiconductor Field Effect Transistors
  - nMOS and pMOS MOSFETS
  - Voltage applied to insulated gate controls current between source and drain
  - Low power allows very high integration
  - First patent in the '20s in USA and Germany
  - Not widely used until the '60s or '70s

MOS Transistors

- Four terminal device: gate, source, drain, body
- Gate – oxide – body stack looks like a capacitor
  - Gate and body are conductors (body is also called the substrate)
  - SiO_2 (oxide) is a "good" insulator (separates the gate from the body
  - Called metal–oxide–semiconductor (MOS) capacitor, even though gate is mostly made of poly-crystalline silicon (polysilicon)
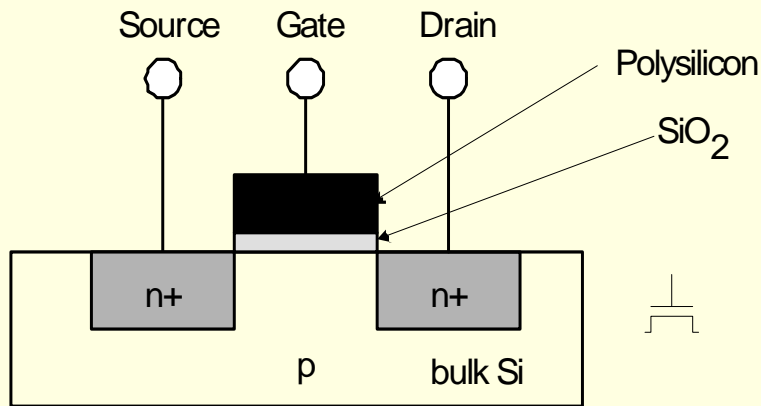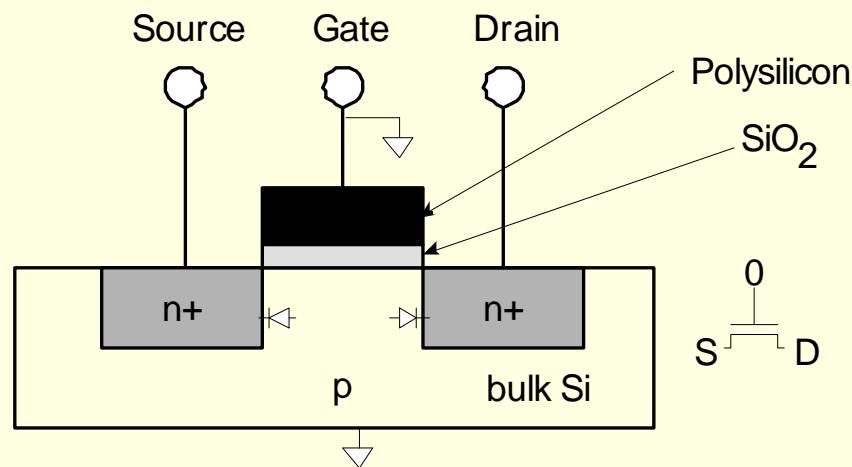

NMOS
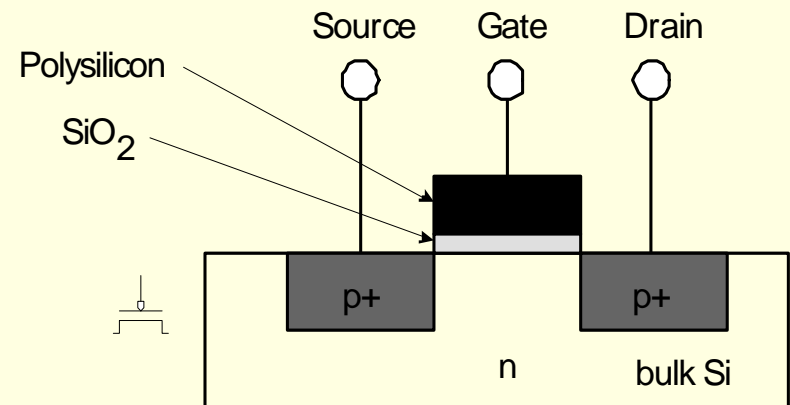
PMOS

NMOS Operation

- Body is commonly tied to ground (0 V)
- Drain is at a higher voltage than Source
- When the gate is at a low voltage:
    - P-type body is at low voltage
    - Source-body and drain-body "diodes" are OFF
   - No current flows, transistor is OFF

NMOS Operation Cont.

- When the gate is at a high voltage: Positive charge on gate of MOS capacitor
  - Negative charge is attracted to body under the gate
  - Inverts a channel under gate to "n-type" (N-channel, hence called the NMOS) if the gate voltage is above a threshold voltage (VT)
  - Now current can flow through "n-type" silicon from source through channel to drain, transistor is ON
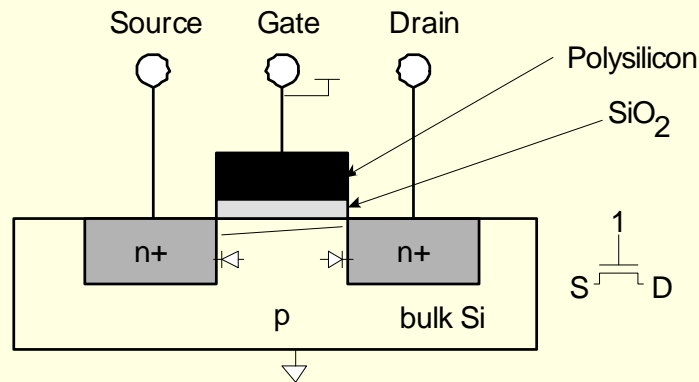
PMOS Transistor

- Similar, but doping and voltages reversed
  - Body tied to high voltage (V_{DD})
  - Drain is at a lower voltage than the Source
  - Gate low: transistor ON
  - Gate high: transistor OFF
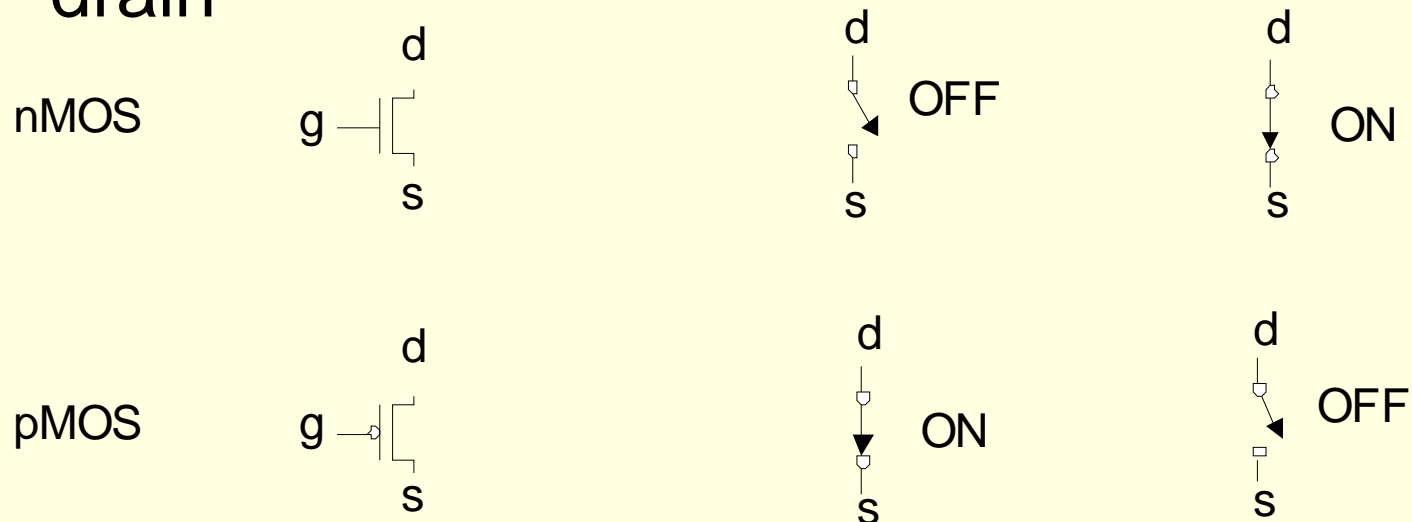 - Bubble indicates inverted behavior

Source   Gate   Drain

Polysilicon

SiO_2

p+   p+

n   bulk Si
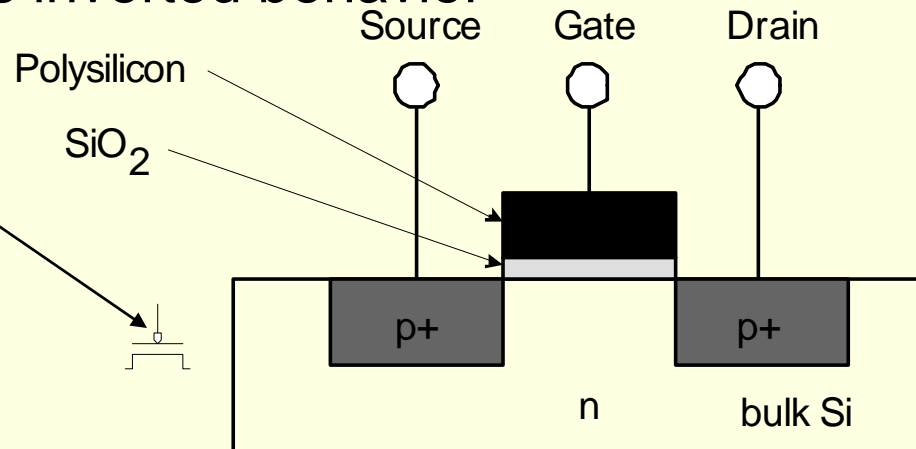
Power Supply Voltage

- GND = 0 V
- In 1980's, V_{DD} = 5V
- V_{DD} has decreased in modern processes
  - High V_{DD} would damage modern tiny transistors
  - Lower V_{DD} saves power
- V_{DD} = 3.3, 2.5, 1.8, 1.5, 1.2, 1.0,
- Effective power supply voltage can be lower due

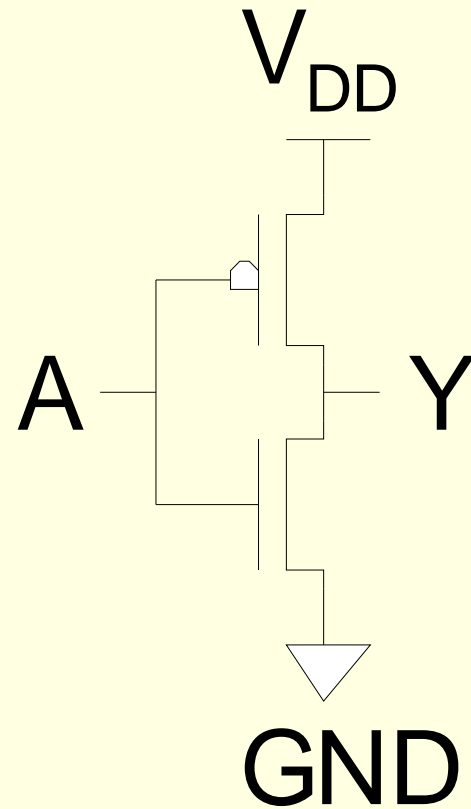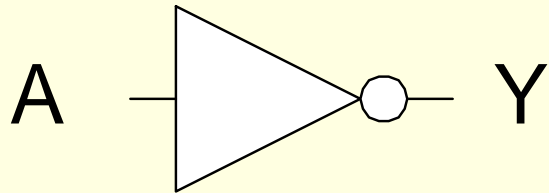 to IR drop across the power grid.

Transistors as Switches

- In Digital circuits, MOS transistors are electrically controlled switches
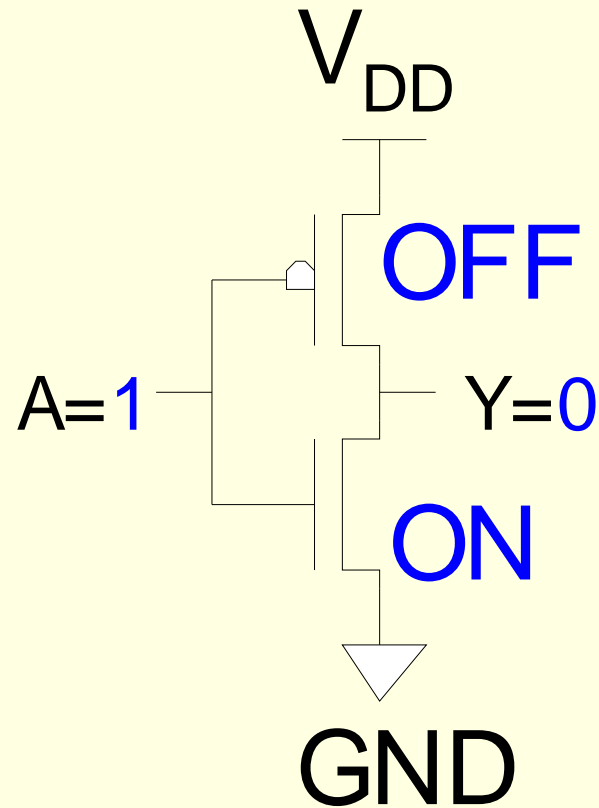
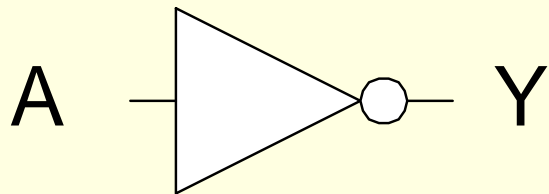- Voltage at gate controls path from source to drain

$g = 0$ $g = 1$

nMOS

g —| [
d
s

d
s
OFF

d
s
ON

pMOS

g —o| [
d
s

d
s
ON

d
s
OFF

CMOS Inverter

A	Y
0	
1	

V_{DD}

A ———○——— Y

A ———▷○——— Y

GND

CMOS Inverter

A	Y
0	
1	**0**

V_{DD}

OFF

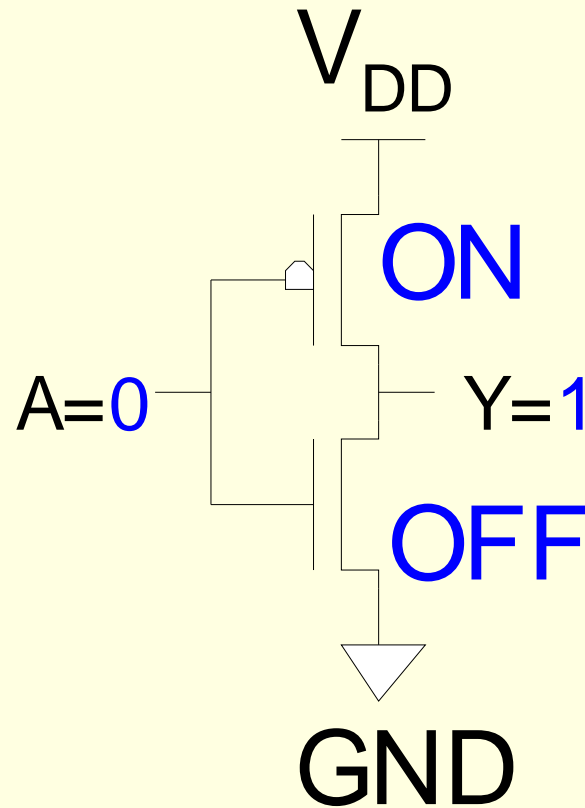A=1   Y=0

ON

Y is pulled low by the turned on NMOS Device. Hence NMOS is the pull-down device.

A ▷○ Y

GND

CMOS Inverter

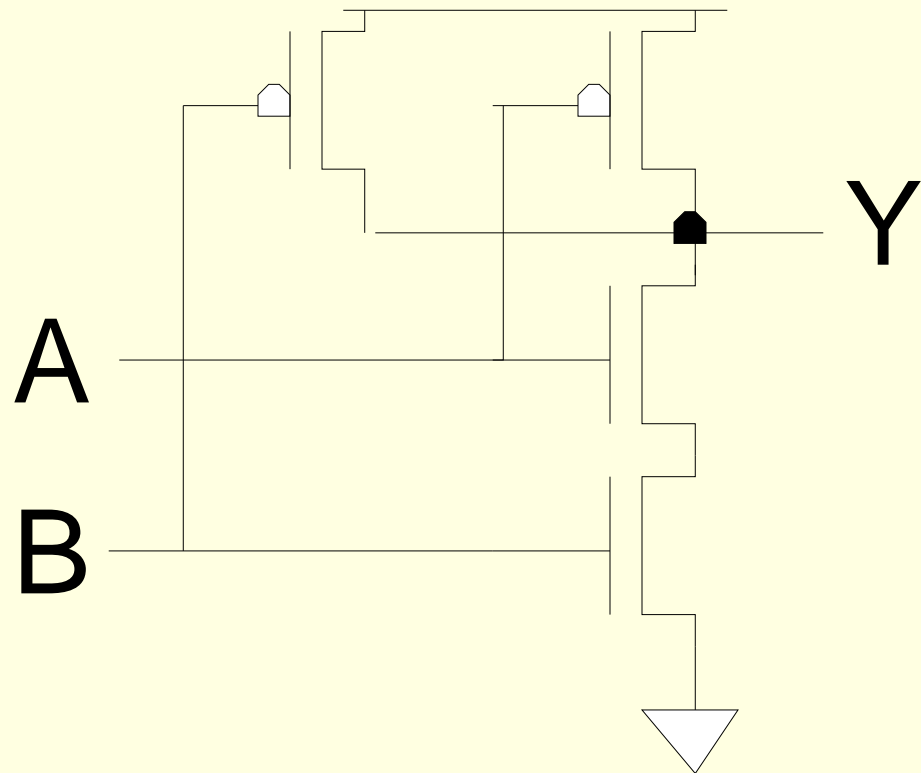| A	Y
0	**1**
1	0

V_{DD}

A=0 — ▷○ — Y

$A=0$

ON

$Y=1$

OFF

GND
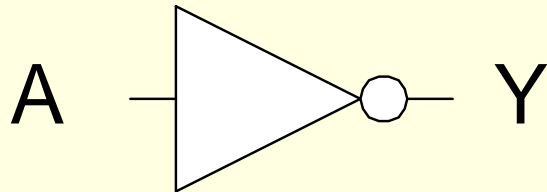
Y is pulled high by the turned on PMOS Device. Hence PMOS is the pull-up device.

CMOS NAND Gate

A	B	Y
0	0	
0	1	
1	0	
1	1	
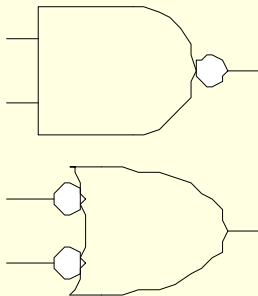
CMOS NAND Gate

A	B	Y
0	**0**	**1**
0	1	
1	0	
1	1	

ON   ON

Y=1

A=0

OFF

B=0

OFF
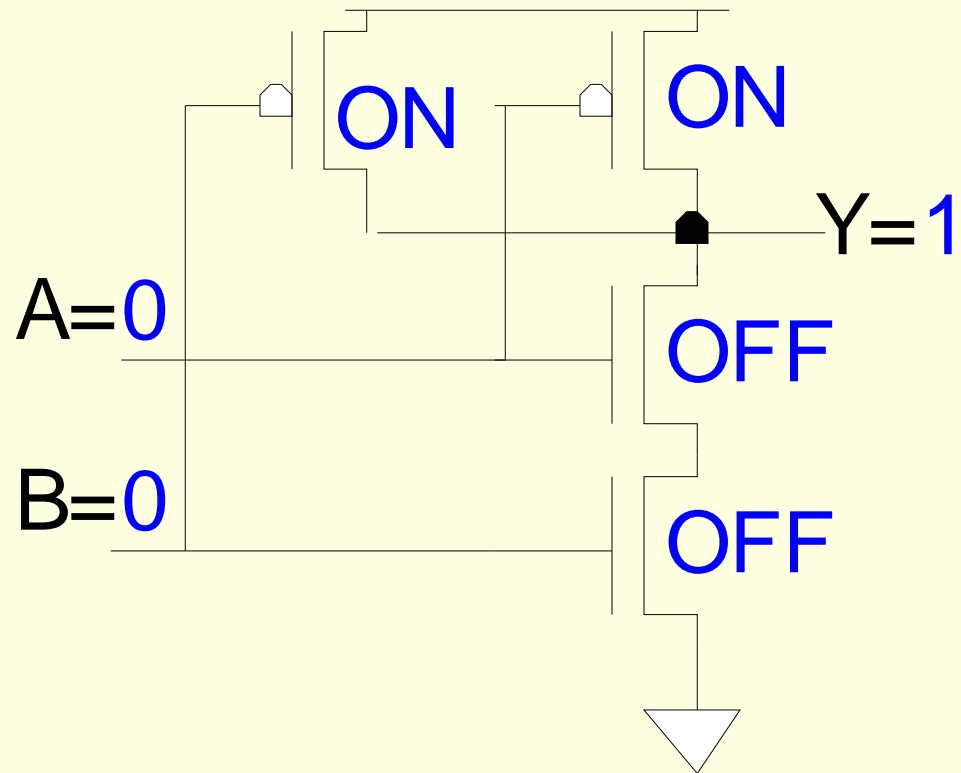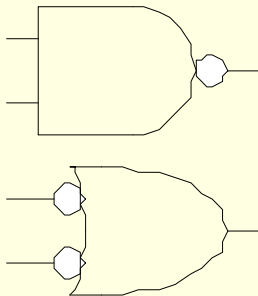
CMOS NAND Gate

A	B	Y
0	0	1
0	**1**	**1**
1	0	
1	1	

CMOS NAND Gate

A	B	Y
0	0	1
0	1	1
1	**0**	**1**
1	1	

ON   OFF

A=1   ON

B=0   OFF

Y=1
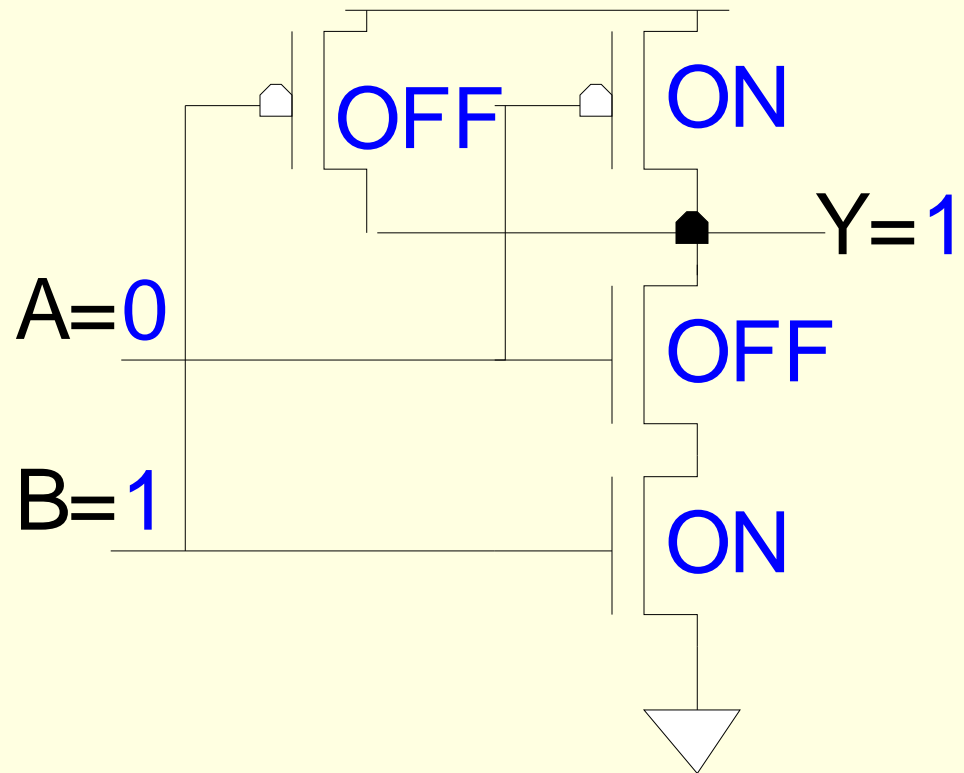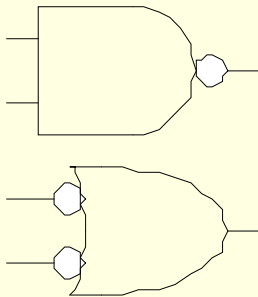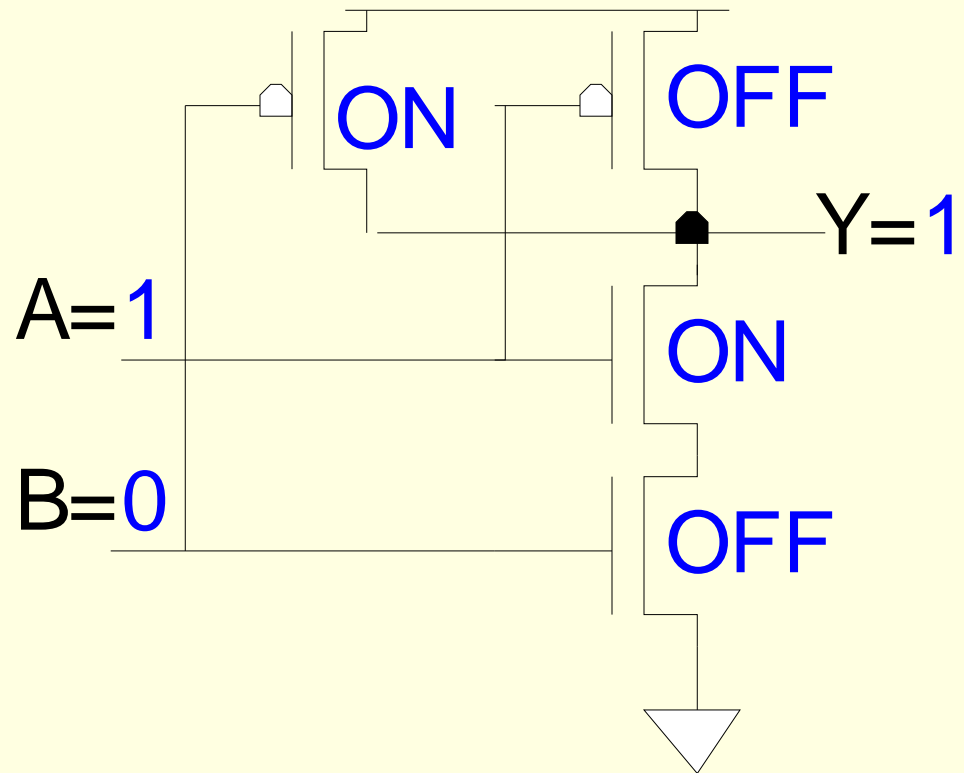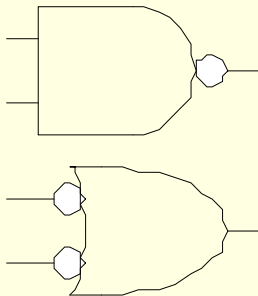
CMOS NAND Gate

A	B	Y
0	0	1
0	1	1
1	0	1
1	**1**	**0**
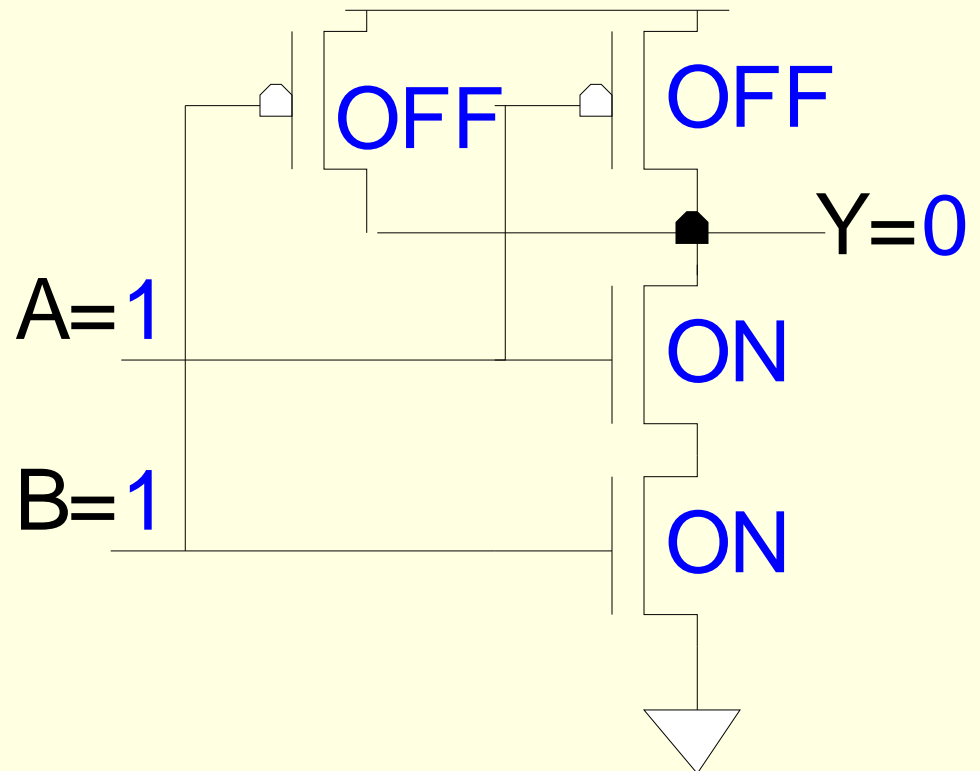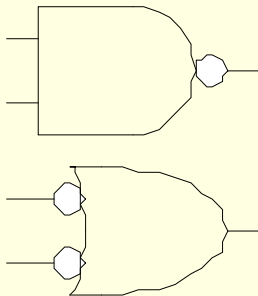
CMOS NOR Gate

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0
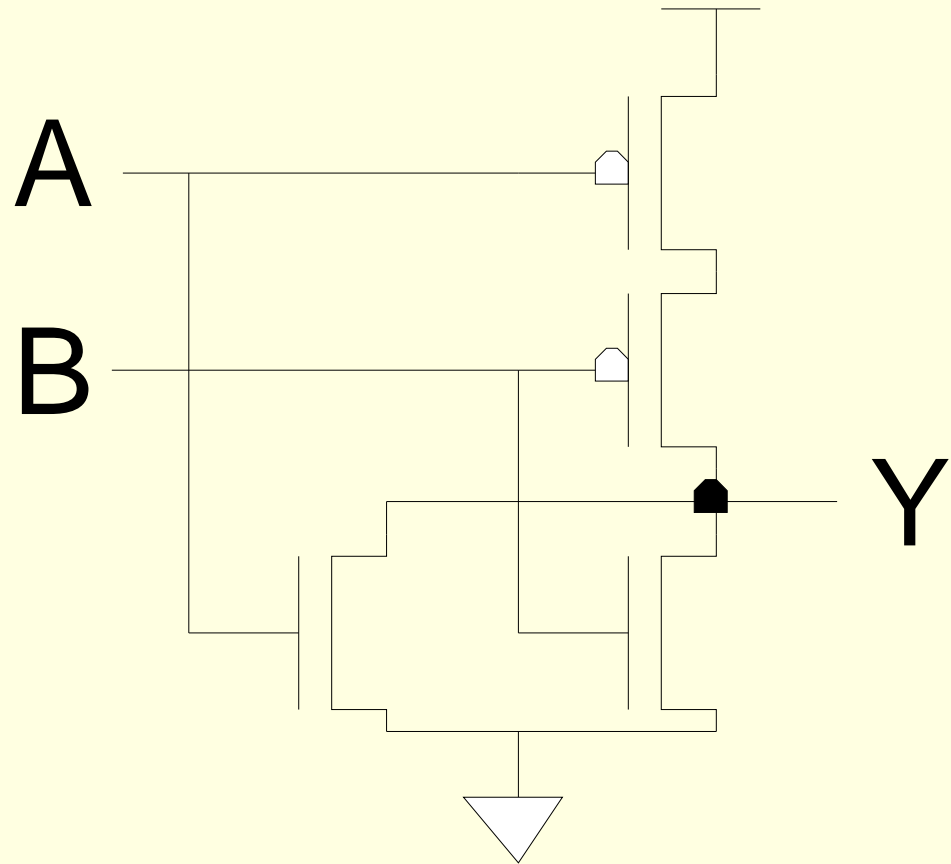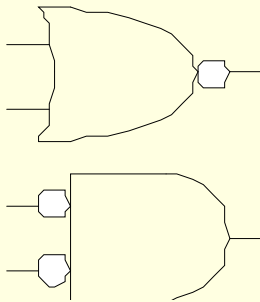
3-input NAND Gate

- Y is pulled low if ALL inputs are 1
- Y is pulled high if ANY input is 0
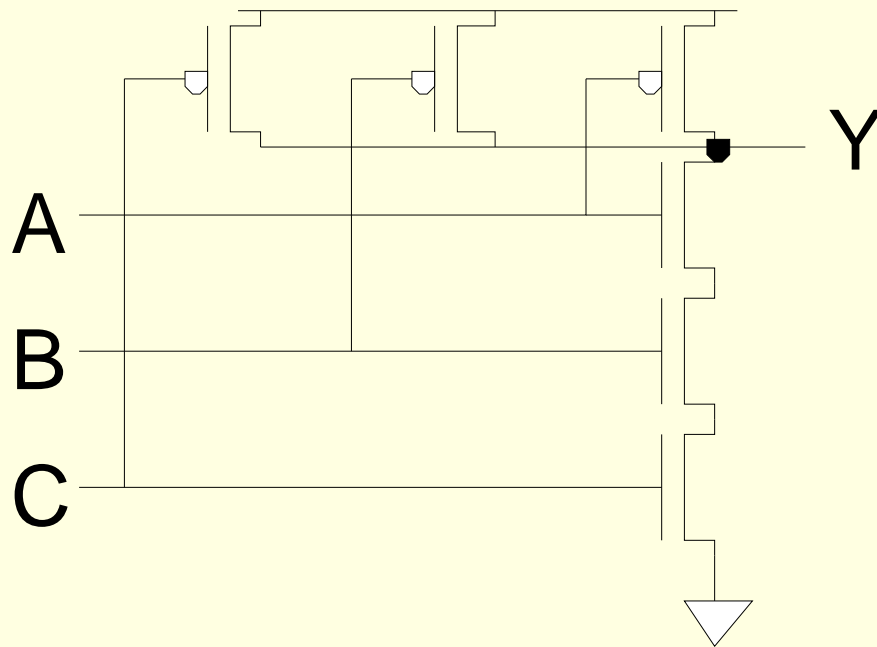
CMOS Fabrication

- CMOS transistors are fabricated on silicon wafer

- Wafers diameters (200-300 mm)

- Lithography process similar to printing press

- On each step, different materials are deposited, or patterned 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 transistors
- Requires to make an n-well for body of pMOS transistors
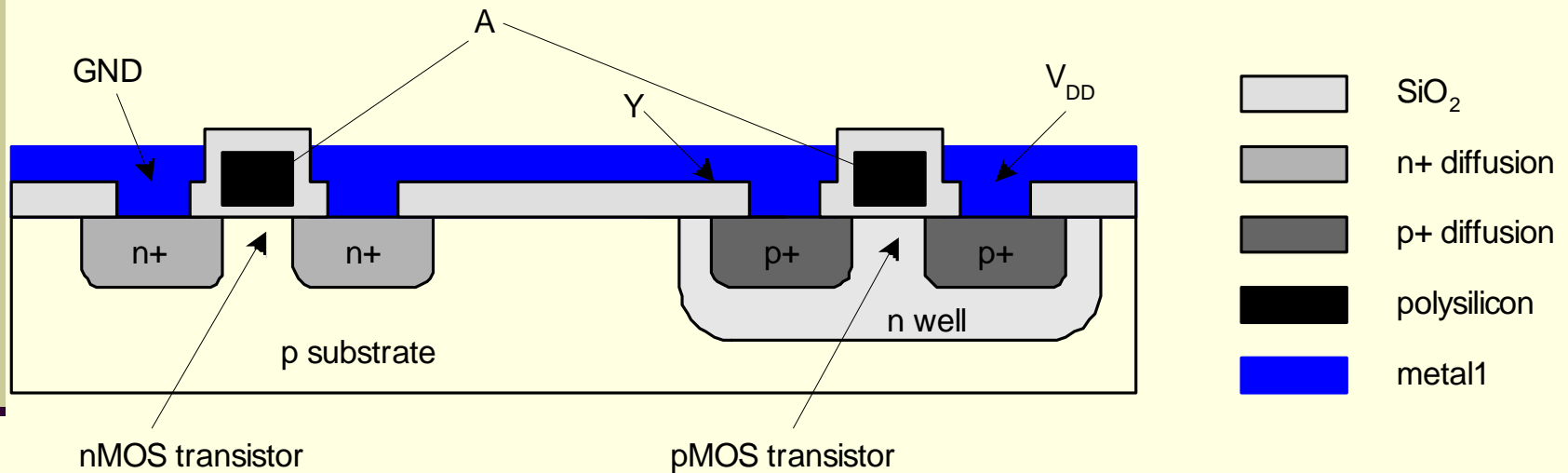
Well and Substrate Taps

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

GND

A

Y

V_{DD}

p+   n+   n+   p+   p+   n+

n well

p substrate

substrate tap

well tap
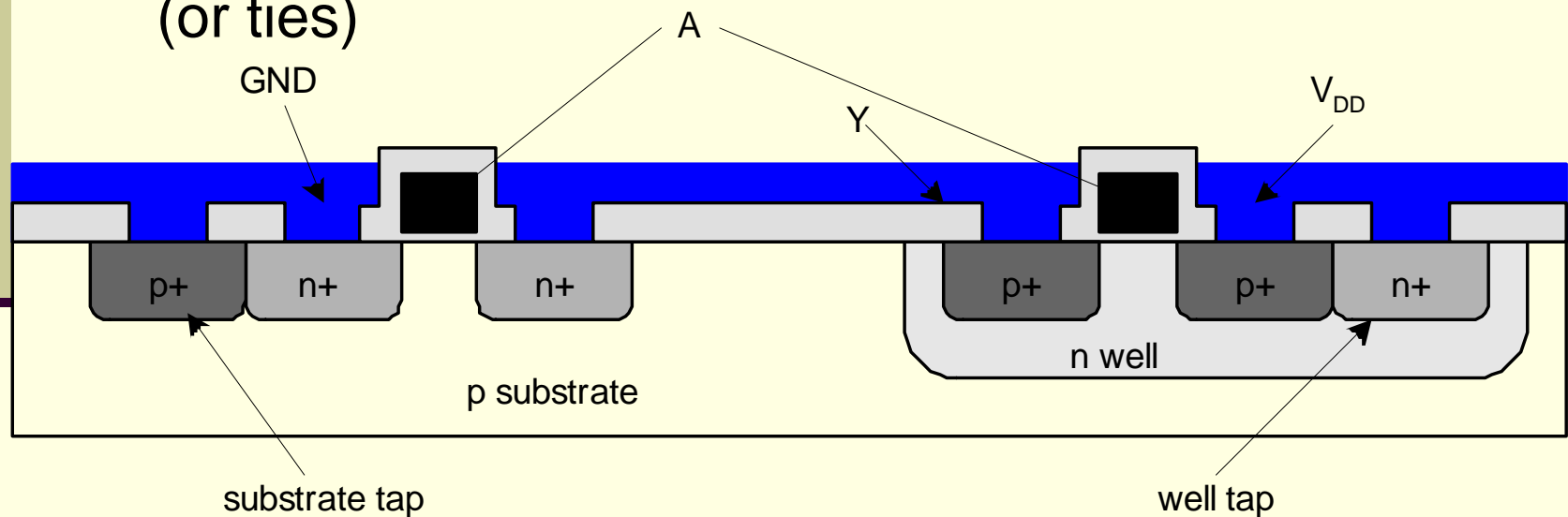
Inverter Mask Set

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

GND

Y

V$_{DD}$

substrate tap

nMOS transistor

pMOS transistor

well tap
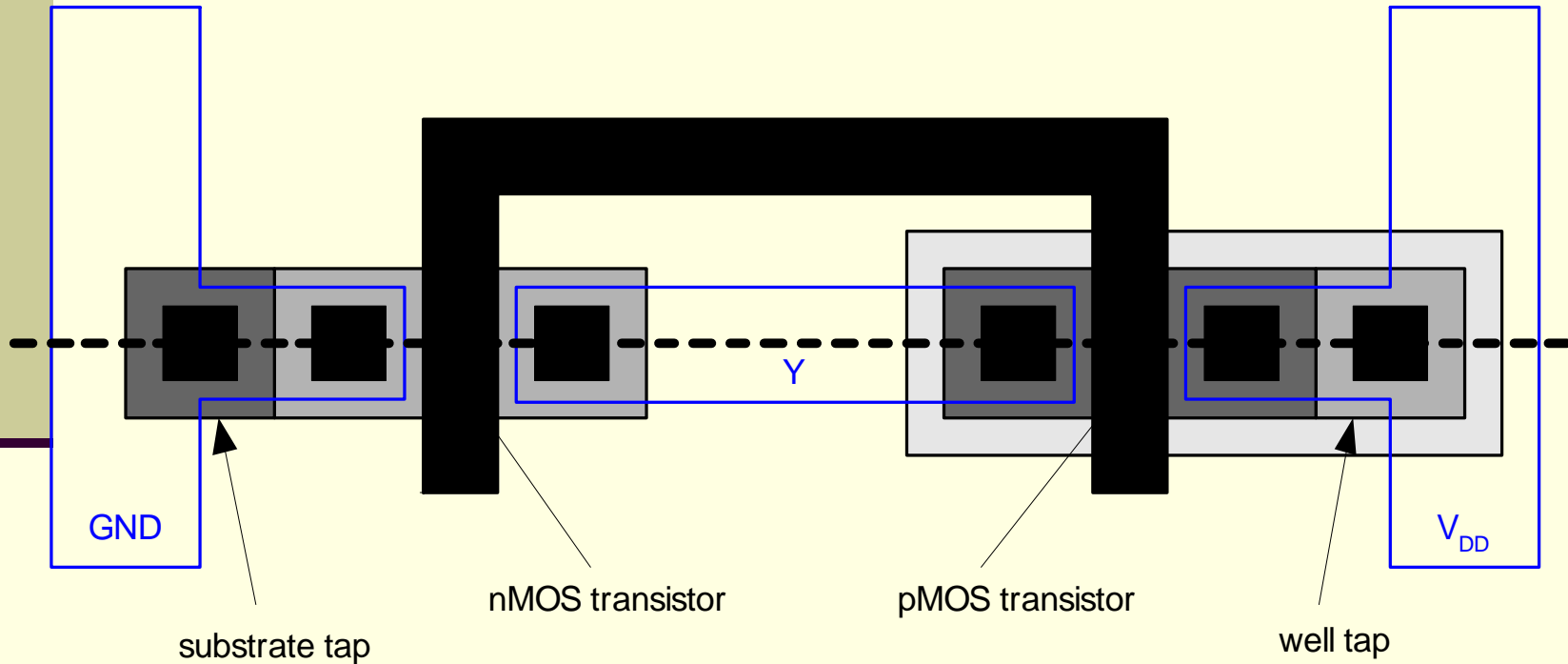
Detailed Mask Views

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

In reality >40 masks may be needed

n well

Polysilicon

n+ Diffusion

p+ Diffusion

Contact

Metal
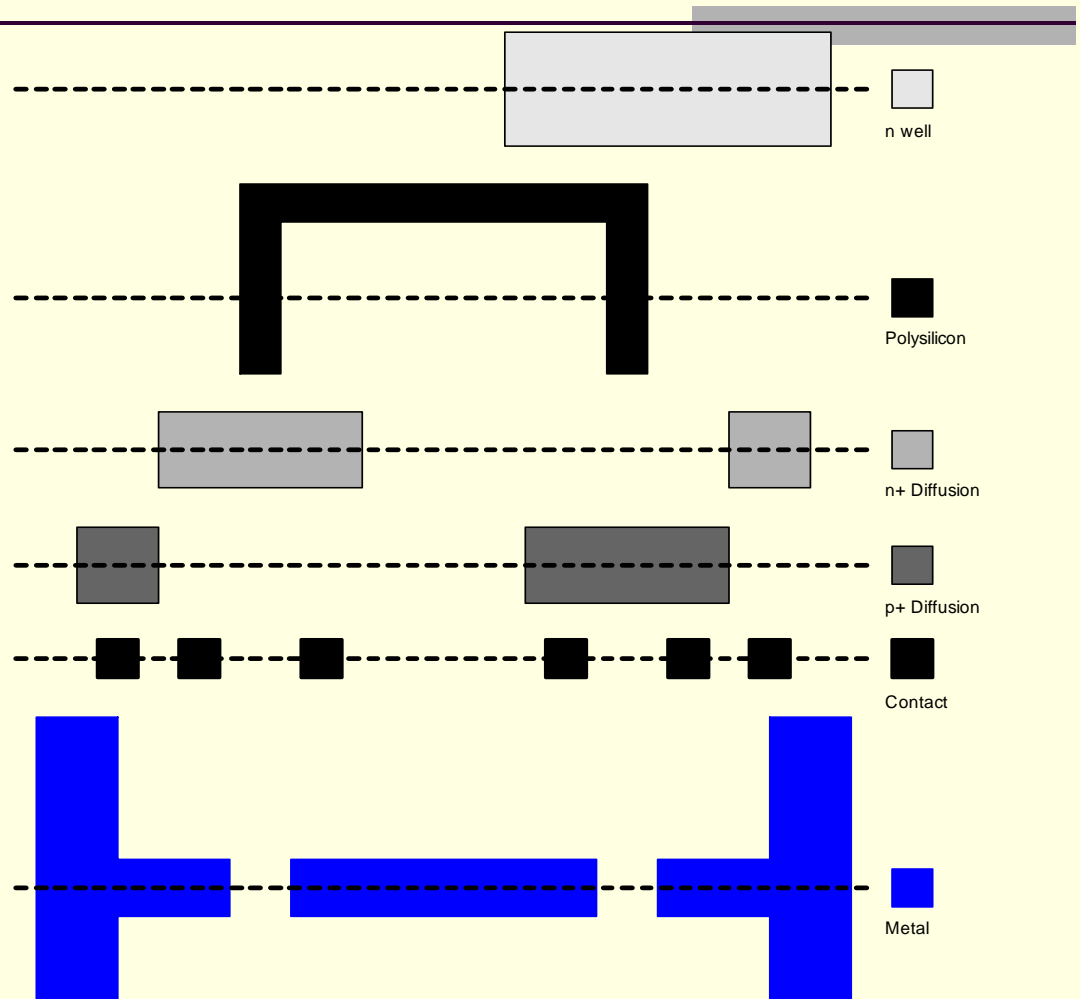
Fabrication Steps

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

p substrate

Oxidation

- Grow SiO_2 on top of Si wafer
  - 900 – 1200 C with H_2O or O_2 in oxidation furnace

SiO_2

p substrate
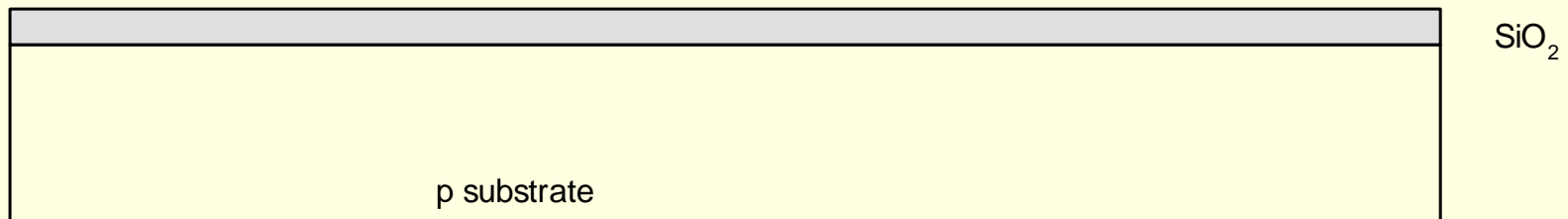
Photoresist

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

- Two types of photoresists (positive or negative)
  - Positive resists can be removed if exposed to UV light
  - Negative resists cannot be removed if exposed to UV light

Photoresist

SiO$_2$

p substrate
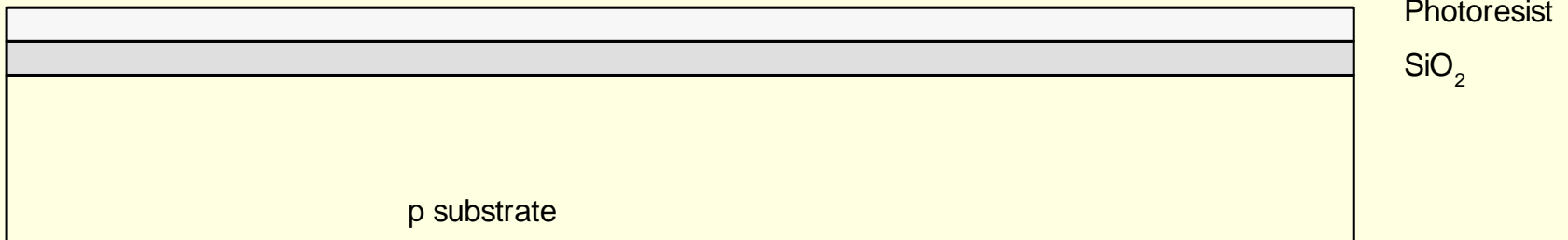
Lithography

- Expose photoresist to Ultra-violate (UV) light through the n-well mask
- Strip off exposed photoresist with chemicals

Photoresist

SiO_2

p substrate
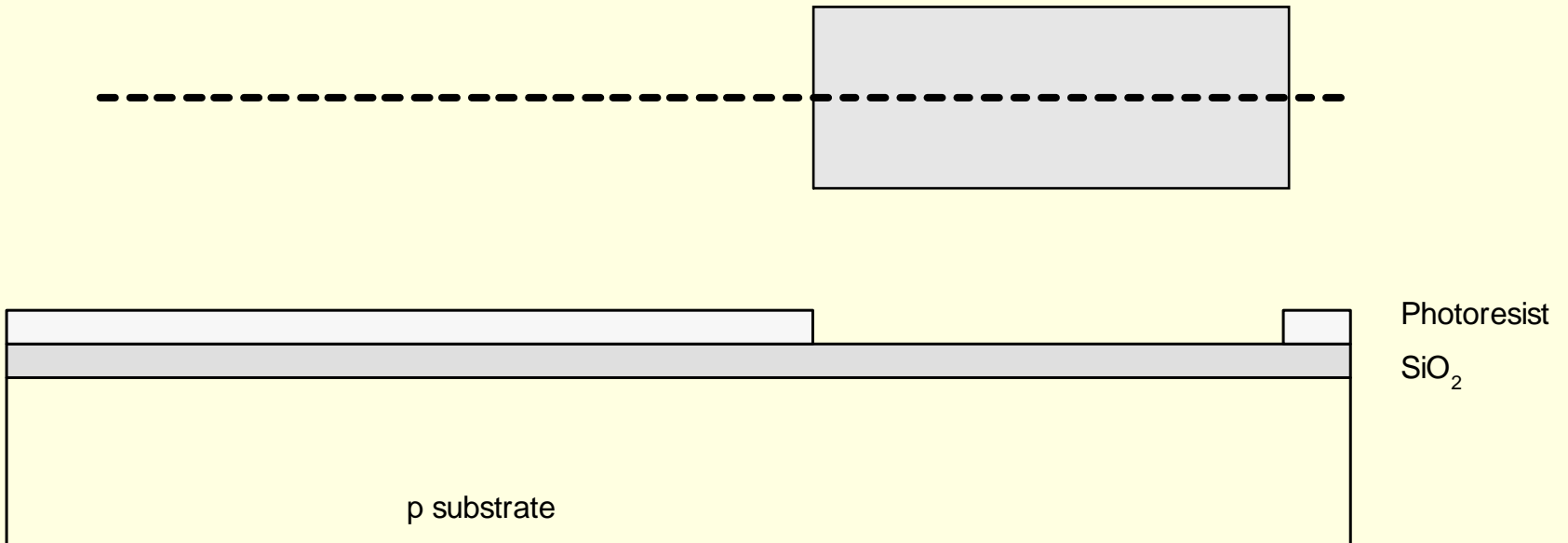
Etch

- Etch oxide with hydrofluoric acid (HF)
  - Seeps through skin and eats bone; nasty stuff!!!
- Only attacks oxide where resist has been exposed
- N-well pattern is transferred from the mask to silicon-di-oxide surface; creates an opening to the silicon surface

Photoresist

SiO_2

p substrate
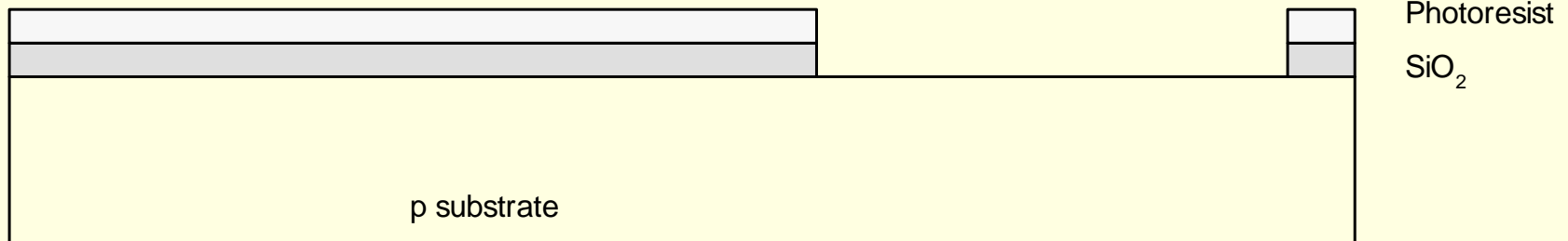
Strip Photoresist

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

SiO_2

p substrate
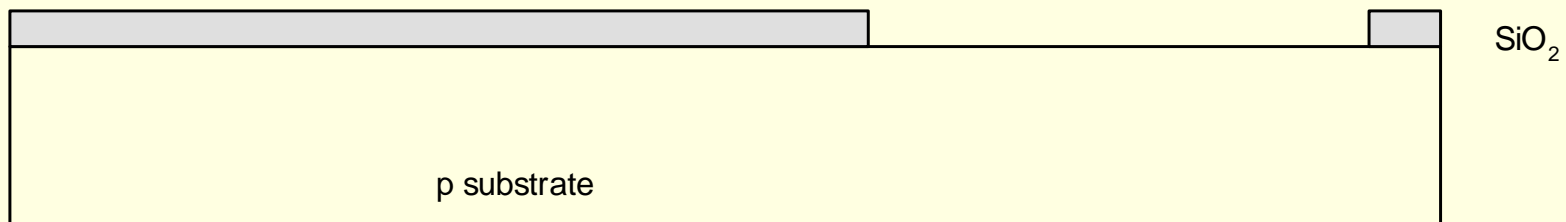
n-well

- n-well is formed with diffusion or ion implantation
- Diffusion
  - Place wafer in furnace with arsenic-rich 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
- SiO_2 shields (or masks) areas which remain p-type

n well

SiO_2
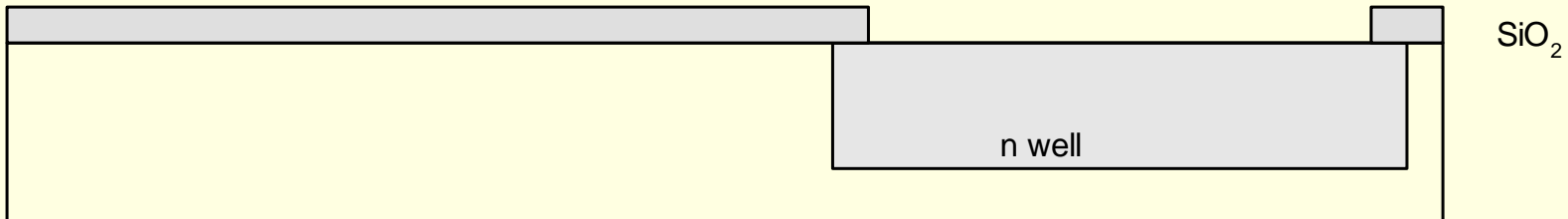
Strip Oxide

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

n well

p substrate
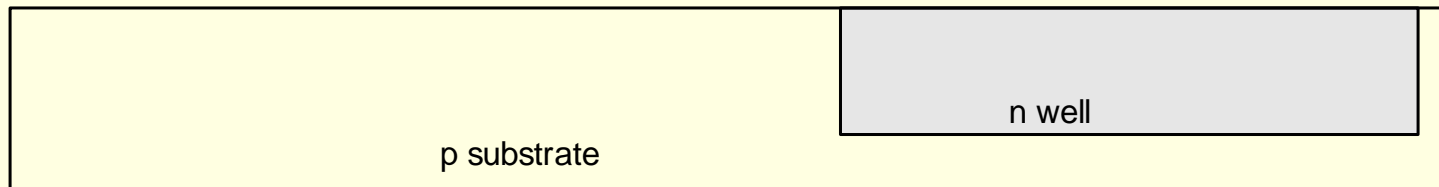
Polysilicon
(self-aligned gate technology)

- 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
Thin gate oxide

n well

p substrate
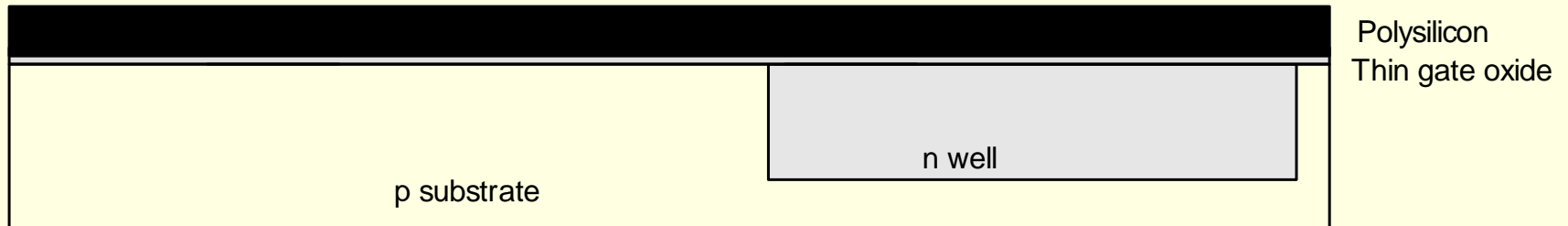
Polysilicon Patterning

- Use same lithography process discussed earlier to pattern polysilicon

Polysilicon

Polysilicon
Thin gate oxide

n well

p substrate
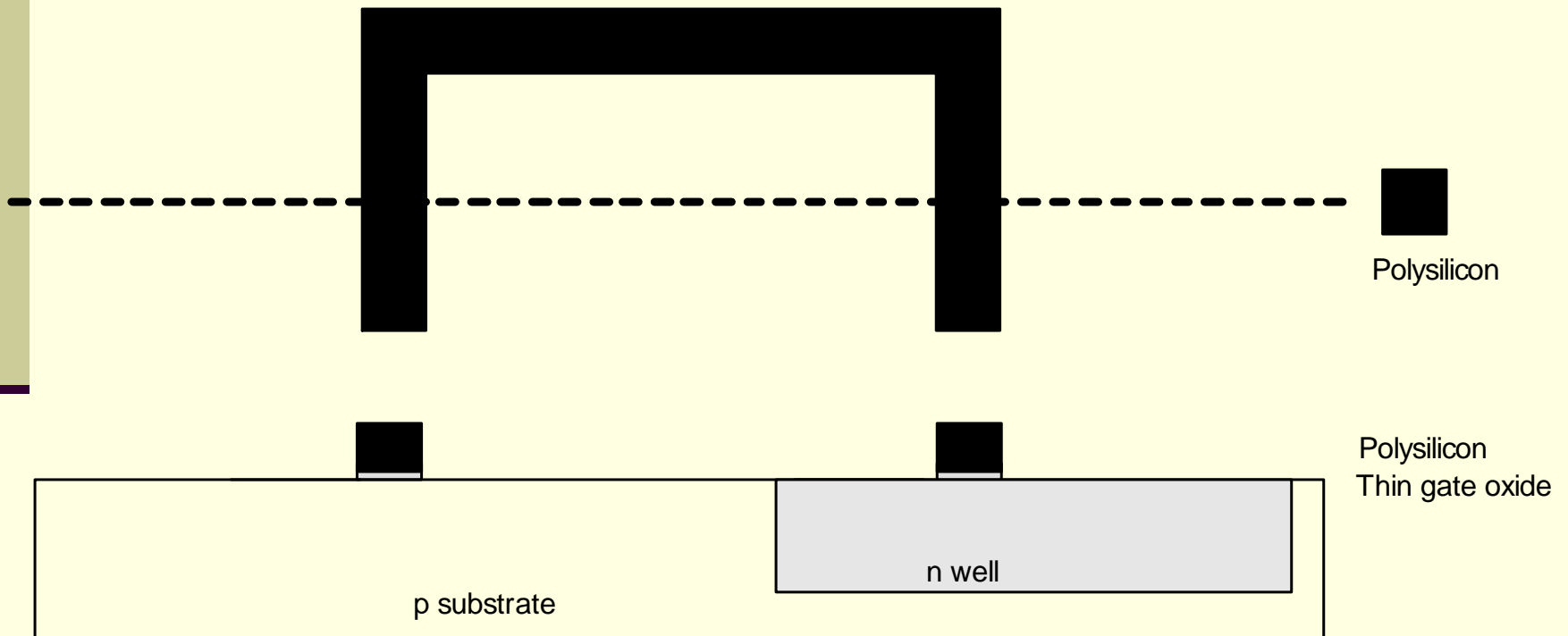
Self-Aligned Process

- Use gate-oxide/polysilicon and masking to expose where n+ dopants should be diffused or implanted

- N-diffusion forms nMOS source, drain, and n-well contact

n well

p substrate
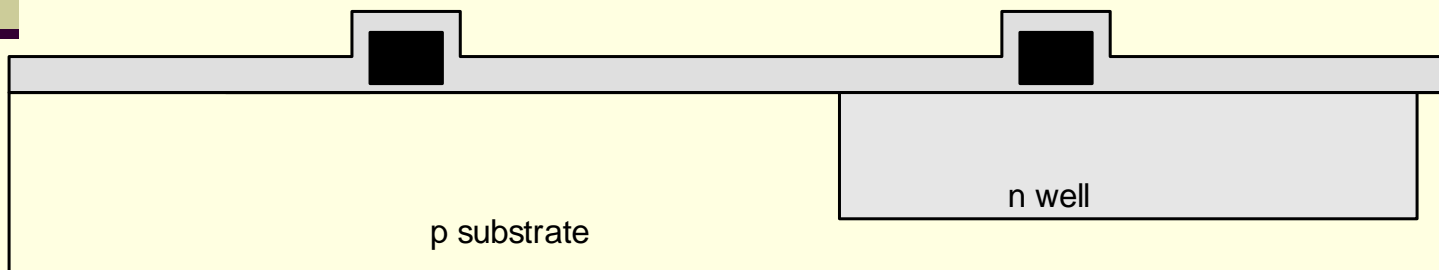
N-diffusion/implantation

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

n+ Diffusion

n well

p substrate
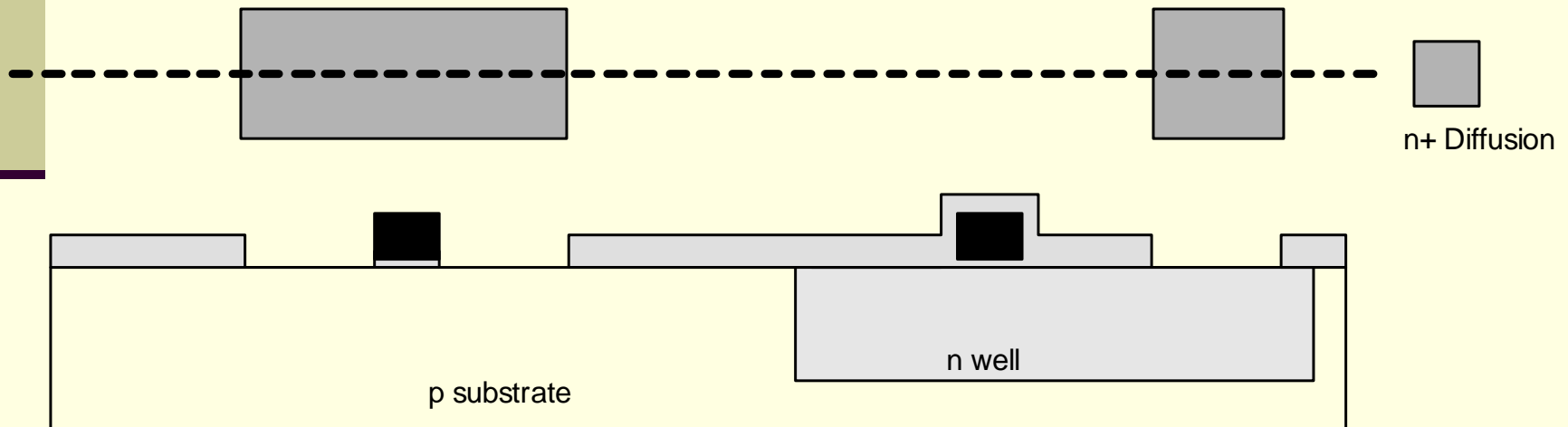
N-diffusion/implantation cont.

- Historically dopants were diffused
- Usually high energy ion-implantation used today
- But n+ regions are still called diffusion
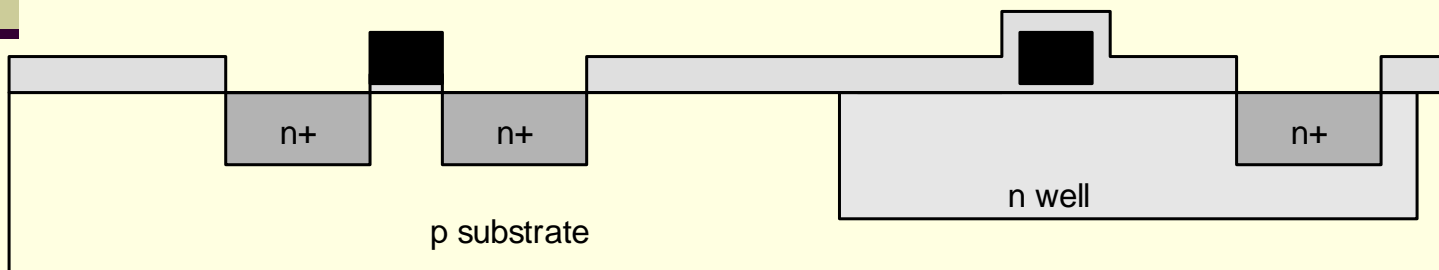
N-diffusion cont.

- Strip off oxide to complete patterning step

P-Diffusion/implantation

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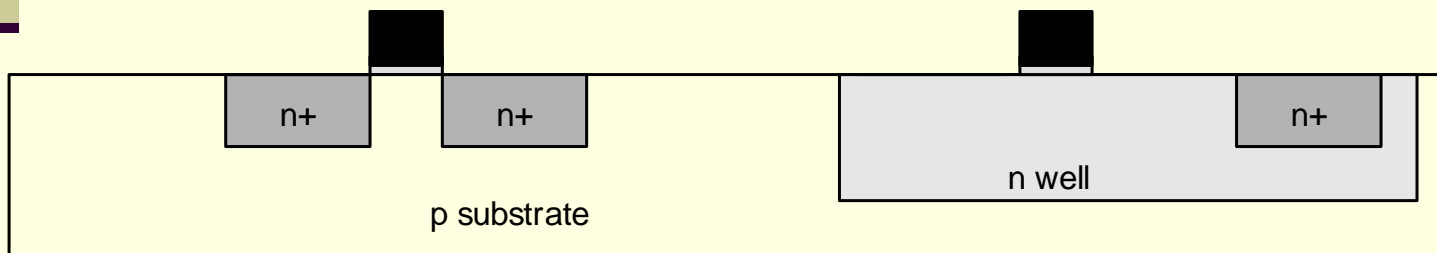
p+ Diffusion

| p+ | n+ | | n+ | | p+ | p+ | n+ |

n well

p substrate
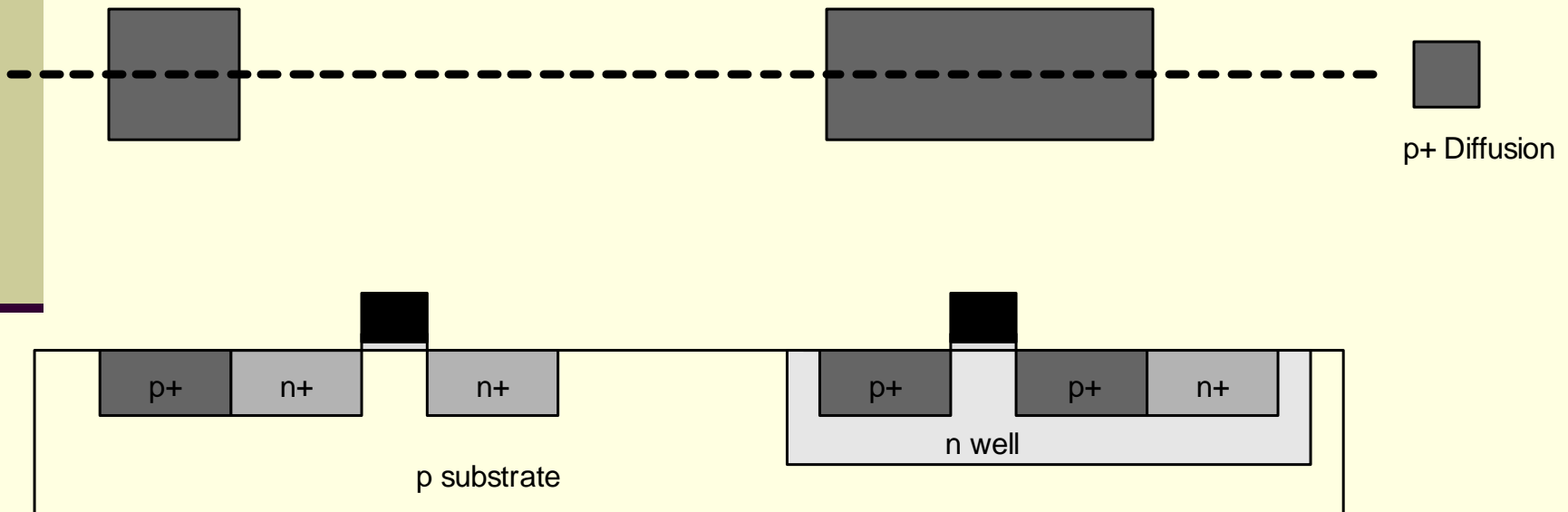
Contacts

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

Contact

Thick field oxide

| p+ | n+ | | n+ | | p+ | | p+ | n+ |

n well

p substrate
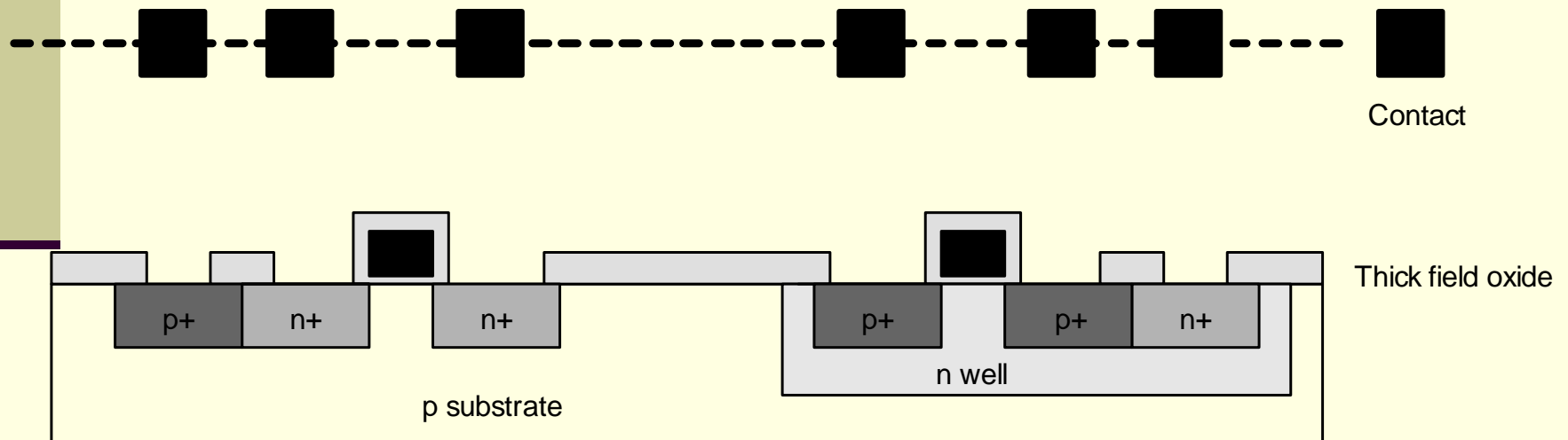
Metalization

- Sputter on aluminum over whole wafer
- Copper is used in newer technology
- Pattern to remove excess metal, leaving wires

Metal

Metal

Thick field oxide

| p+ | n+ | | n+ | | p+ | p+ | n+ |

n well

p substrate
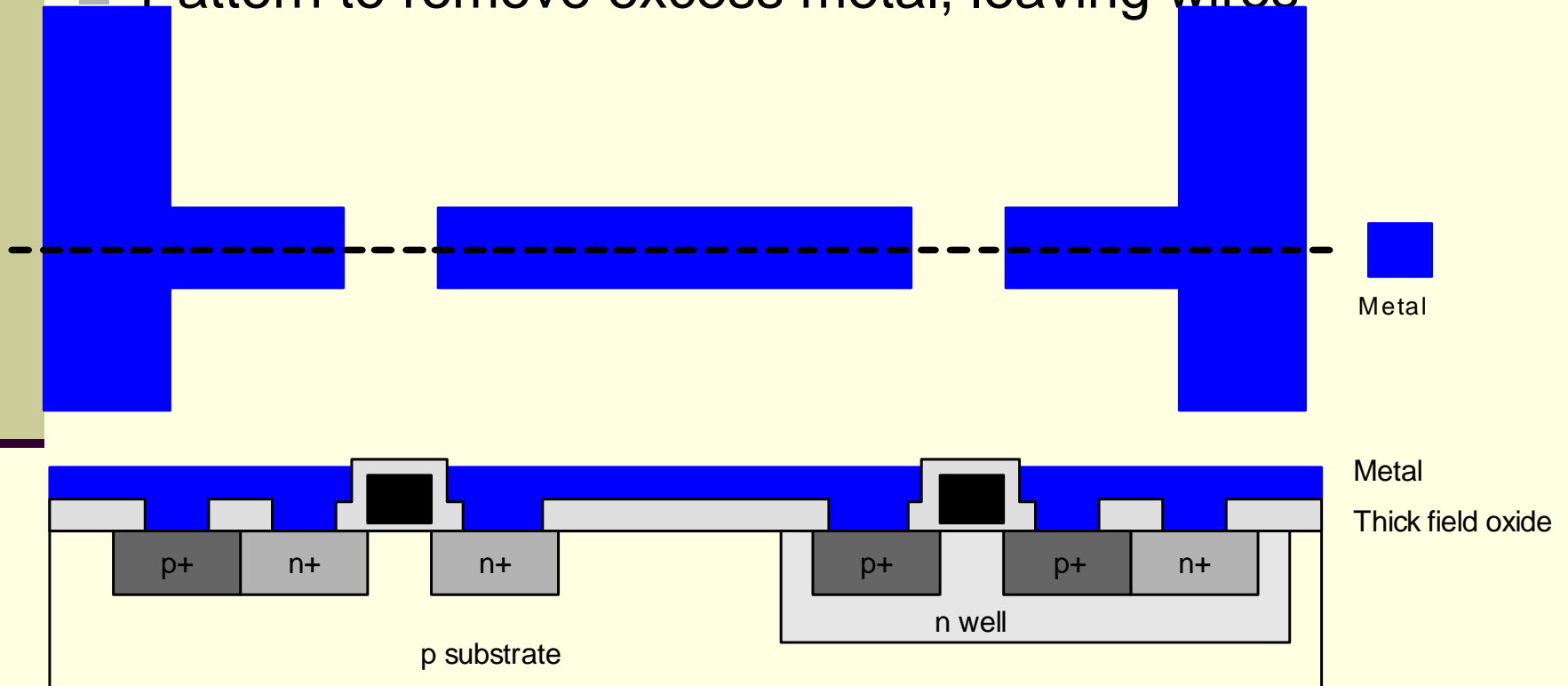
Physical Layout

- 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 $\lambda = f/2$
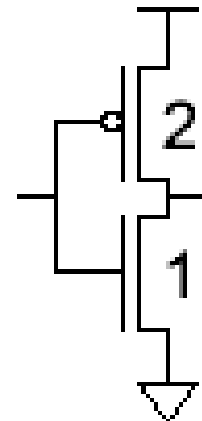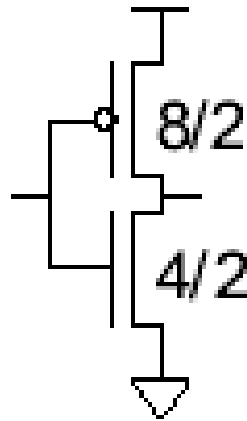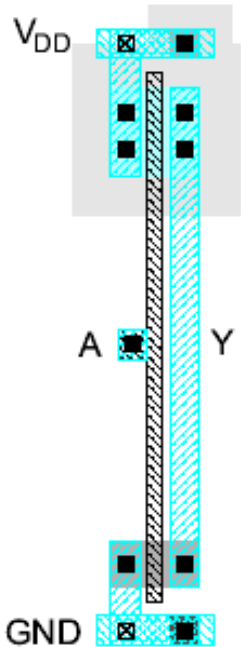 - E.g. $\lambda = 0.3\ \mu m$ in 0.6 μm process

Simplified Design Rules

■ Conservative rules to get you started
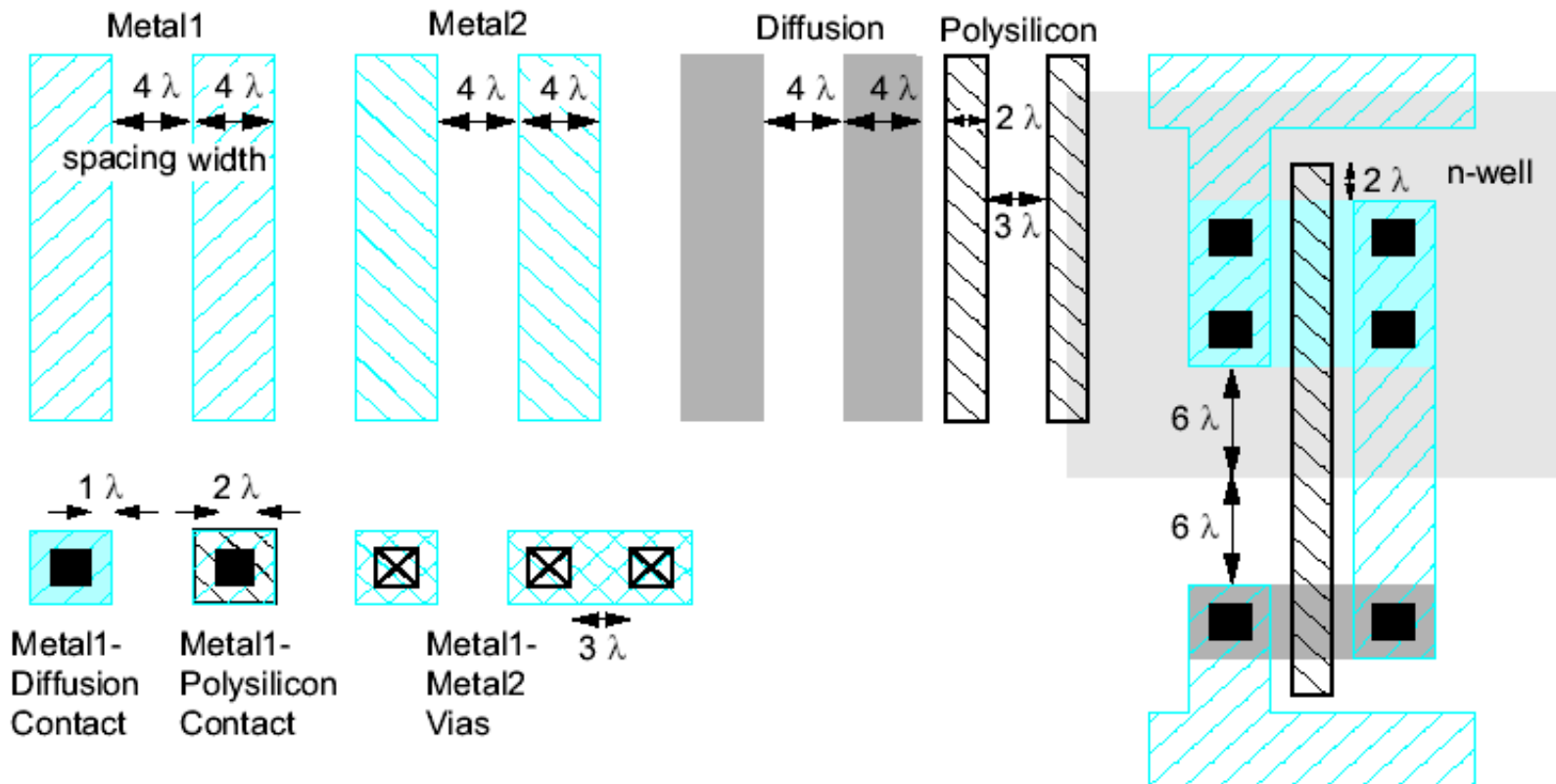
Inverter Layout

- Transistor dimensions specified as Width / Length
  - Minimum size is 4-6λ/ 2λ, sometimes called 1 unit
  - In f = 0.25 μm process, this is 0.5-0.75 μm wide (W), 0.25 μm long (L)
  - Since $\lambda = f/2$, $\lambda = 0.125$ μm.

The Future?
International Technology Roadmap for Semiconductors

Table B ITRS Table Structure—Key Lithography-related Characteristics by Product Type

Near-term Years

YEAR OF PRODUCTION	2003	2004	2005	2006	2007	2008	2009
Technology Node		hp90			hp65		
DRAM ½ Pitch (nm)	100	90	80	70	65	57	50
MPU/ASIC M1 ½ Pitch (nm)	120	107	95	85	75	67	60
MPU/ASIC Poly Si ½ Pitch (nm)	107	90	80	70	65	57	50
MPU Printed Gate Length (nm)	65	53	45	40	35	32	28
MPU Physical Gate Length (nm)	45	37	32	28	25	22	20

Long-term Years

YEAR OF PRODUCTION	2010	2012	2013	2015	2016	2018
Technology Node	hp45		hp32		hp22	
DRAM ½ Pitch (nm)	45	35	32	25	22	18
MPU/ASIC M1 ½ Pitch (nm)	54	42	38	30	27	21
MPU/ASIC Poly Si ½ Pitch (nm)	45	35	32	25	22	18
MPU Printed Gate Length (nm)	25	20	18	14	13	10
MPU Physical Gate Length (nm)	18	14	13	10	9	7

The ORTC and technology requirements tables are intended to indicate current best estimates of introduction timing for specific technology requirements. Please refer to the Glossary for detailed definitions for Year of Introduction and Year of Production.
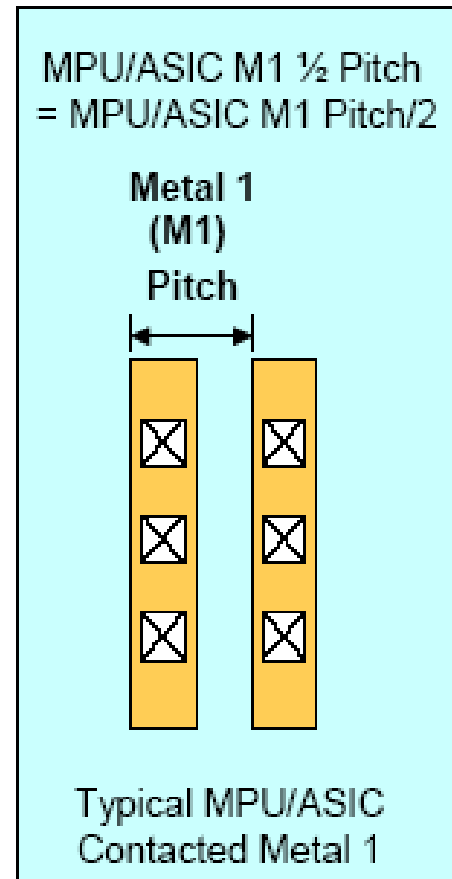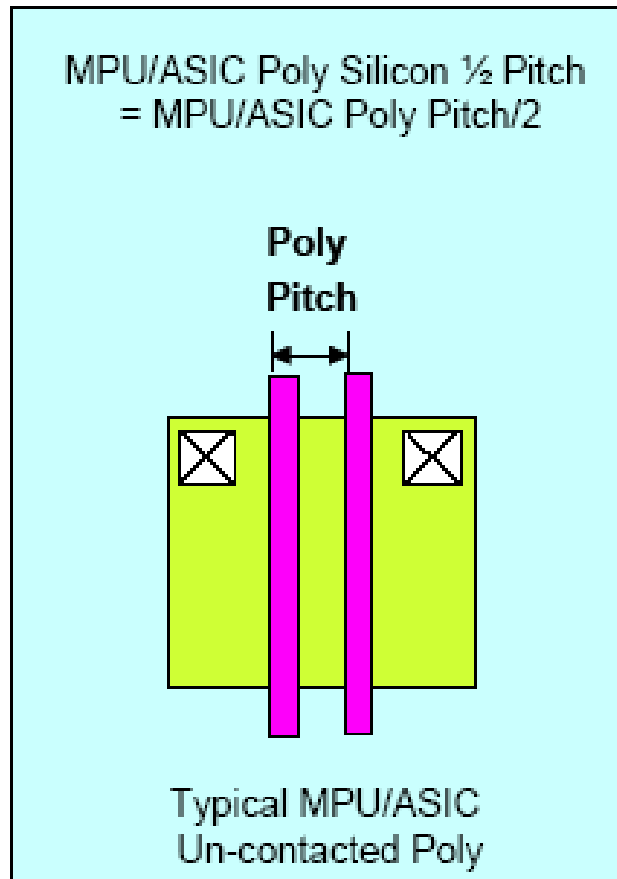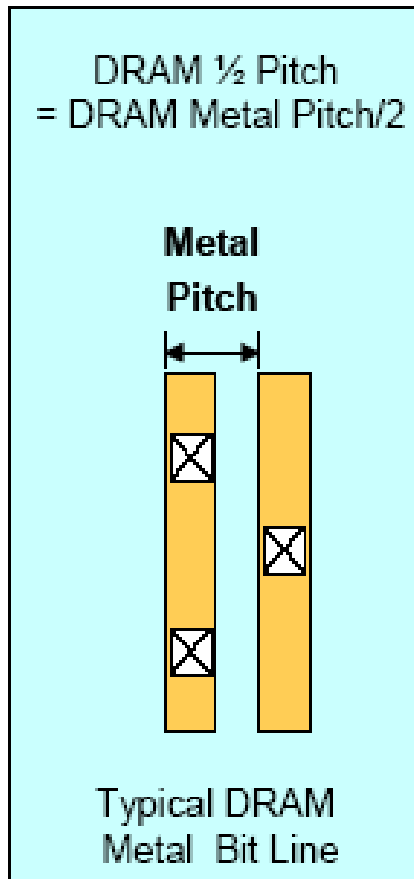
http://public.itrs.net/Files/2003ITRS/Home2003.htm

DRAM ½ Pitch = DRAM Metal Pitch/2

Metal Pitch

Typical DRAM Metal Bit Line

MPU/ASIC Poly Silicon ½ Pitch = MPU/ASIC Poly Pitch/2

Poly Pitch

Typical MPU/ASIC Un-contacted Poly

MPU/ASIC M1 ½ Pitch = MPU/ASIC M1 Pitch/2

Metal 1 (M1) Pitch

Typical MPU/ASIC Contacted Metal 1

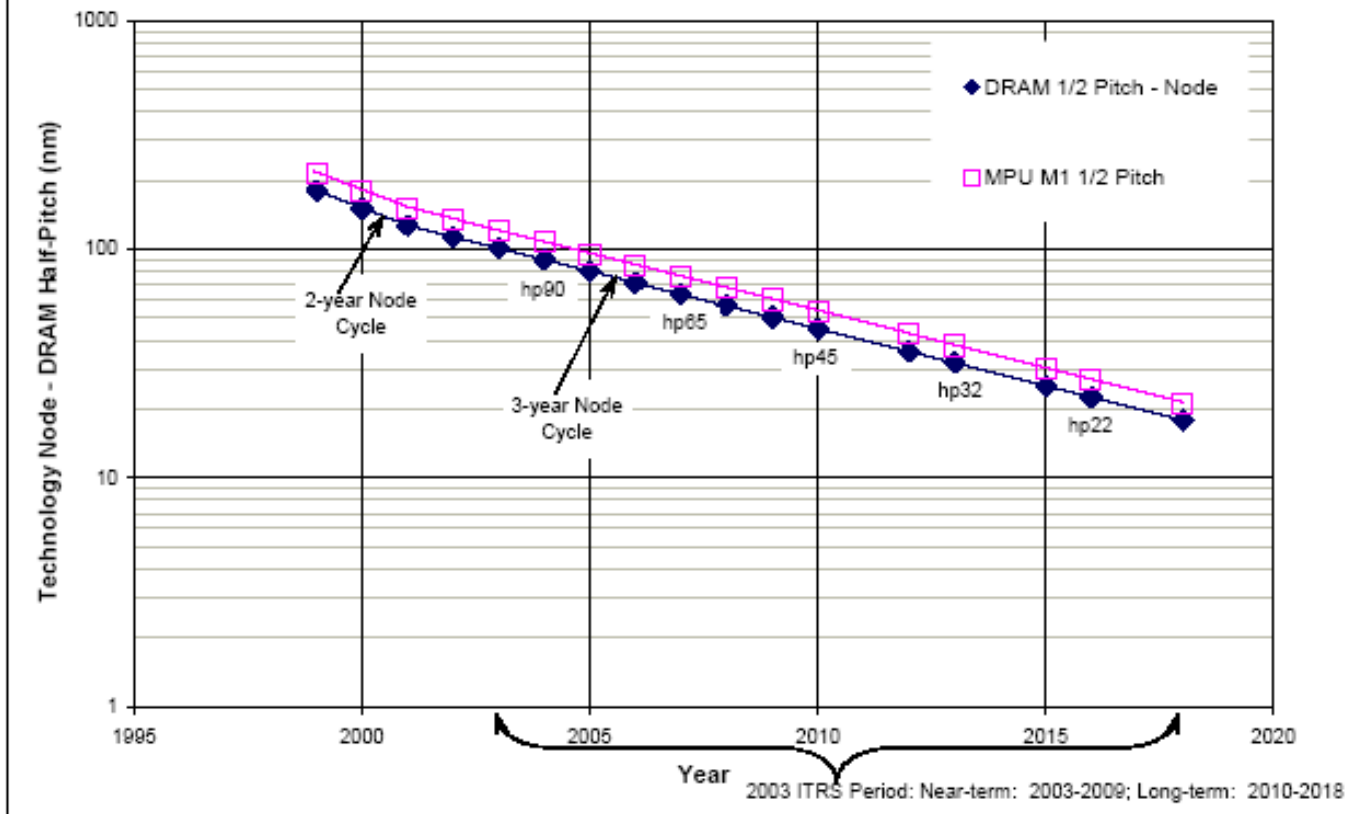*Figure 4   Definition of Metal Half Pitch*

Figure 7   2003 ITRS—Half Pitch Trends

Summary

- MOS Transistors are stack of gate, oxide, silicon

  and p-n junctions

- Can be viewed as electrically controlled switches

- Build logic gates out of switches

- Draw masks to specify layout of transistors

- Now you know everything necessary to start designing schematics and layout for a simple chip!