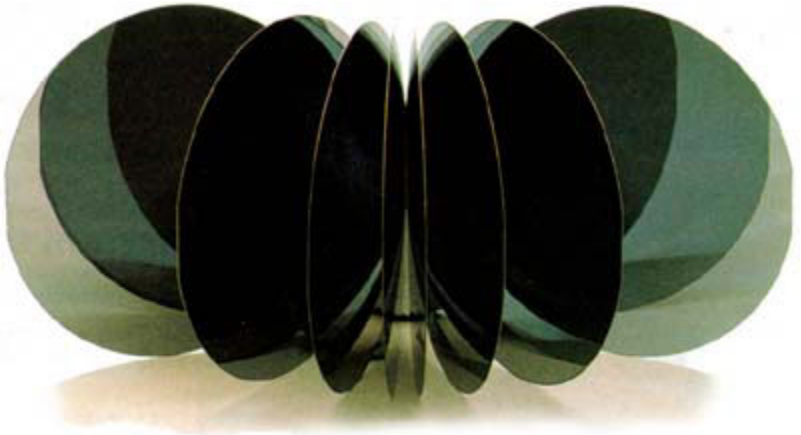
Lecture 2: Fabrication and Layout

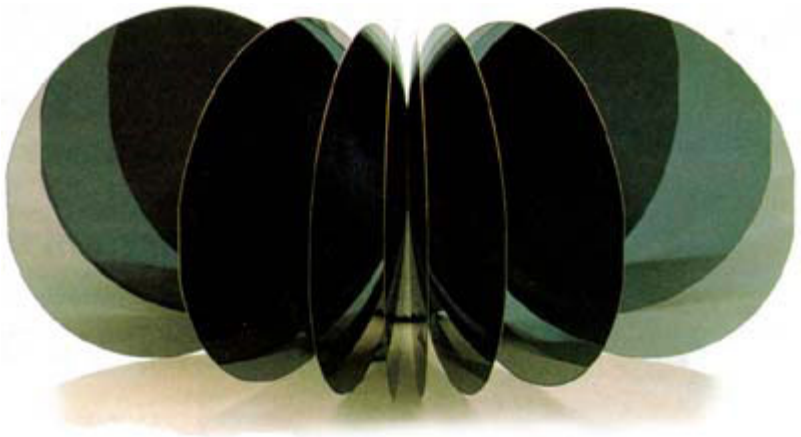
CMOS Fabrication

- 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



Silicon Ingot & Wafer

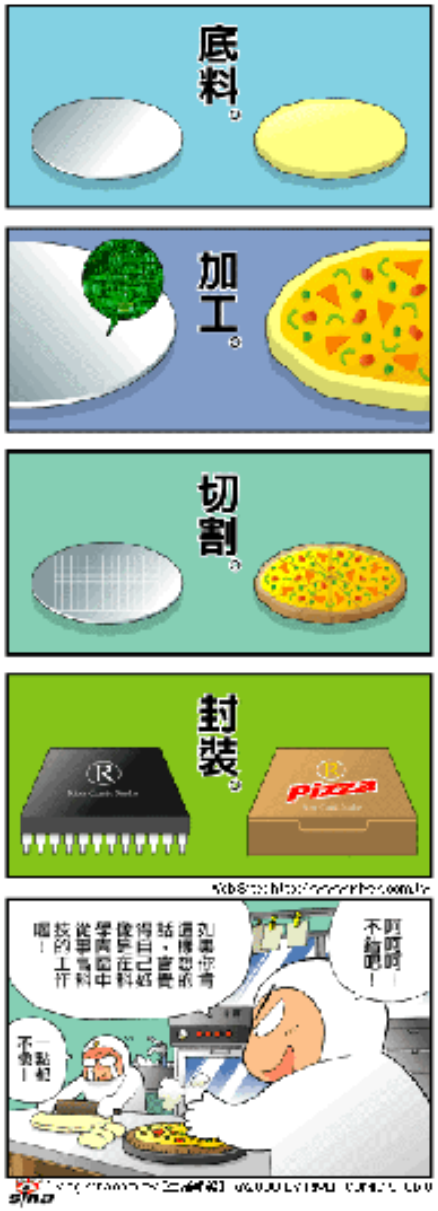
Silicon Ingot



Silicon
Wafer

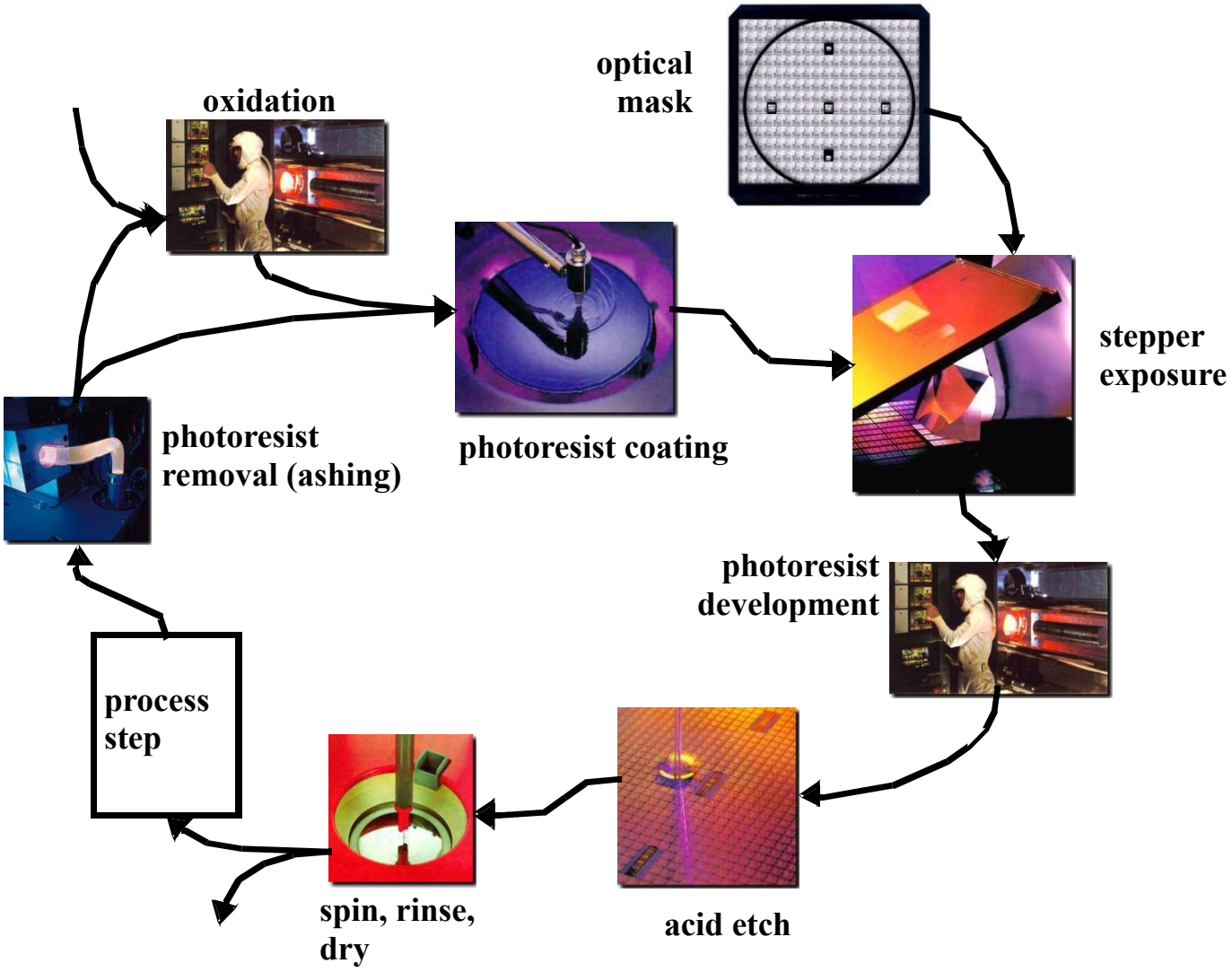
From *Smithsonian*, 2000

Similarity between pizza and IC manufacturing



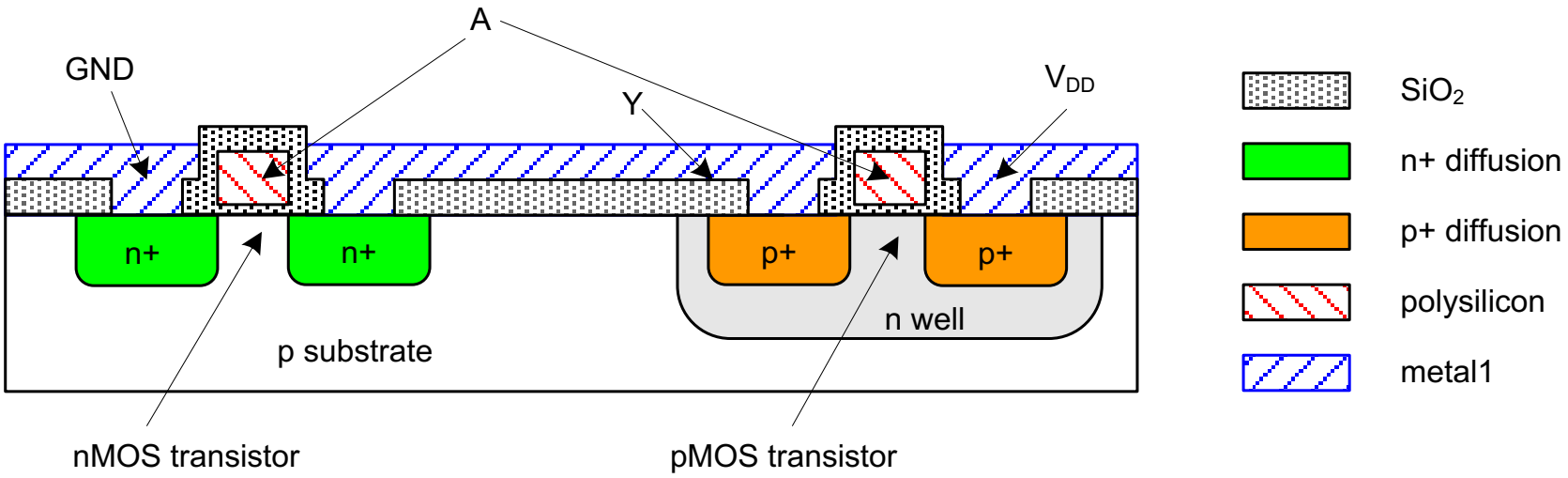
Photolithographic Process

(A method to accomplish selective masking)



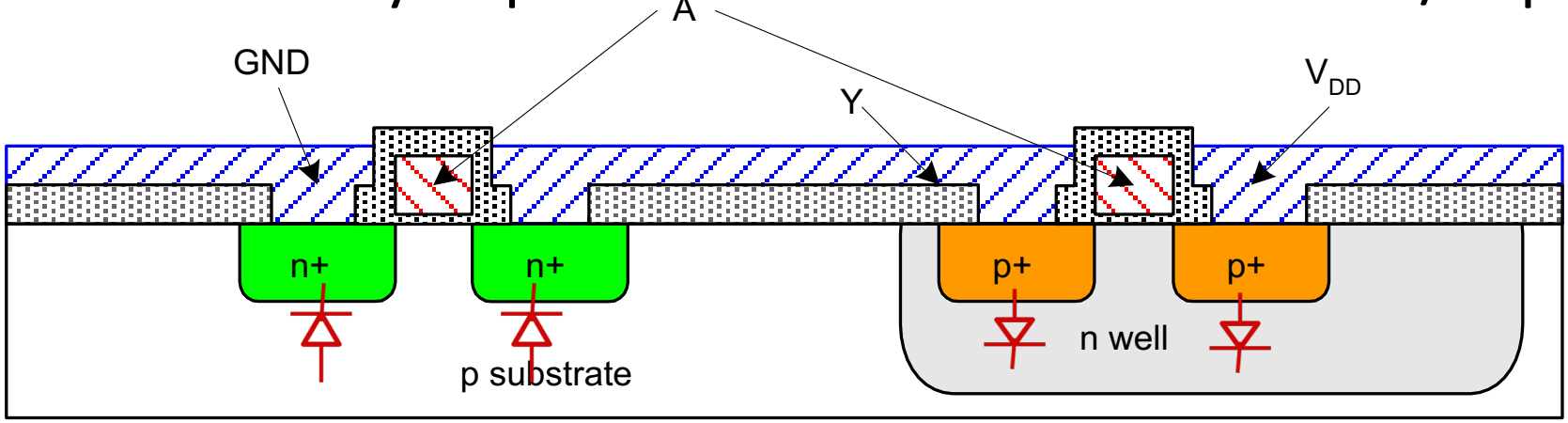
Inverter Cross-section

- Typically use p-type substrate for nMOS transistors
- Requires n-well for body of pMOS transistors



Why Well and Substrate Taps are needed?

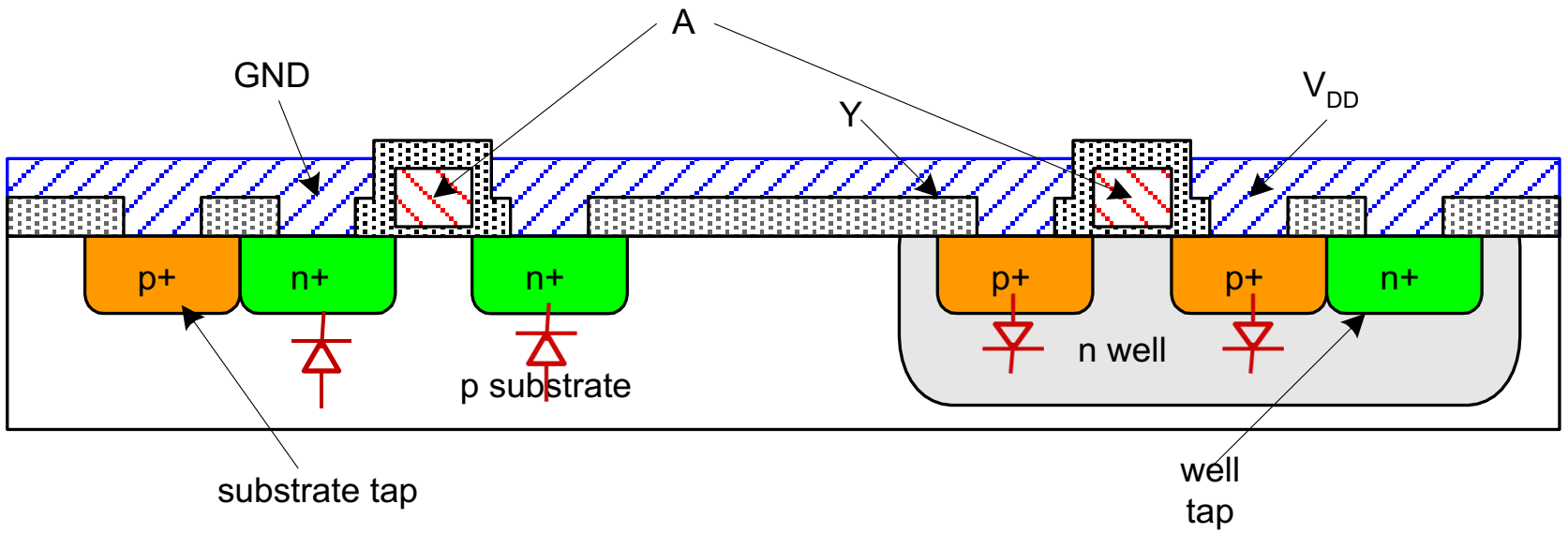
- P-N junctions exists between p substrate and n+
  - Forward-biasing p-n junctions may occurs between p substrate and nmos source and drain if $V_{Substrate} > 0$
- To avoid forward-biasing the p-n junctions
  - Substrate must be tied to GND and n-well to V_{DD}
- Use heavily doped well and substrate contacts / taps



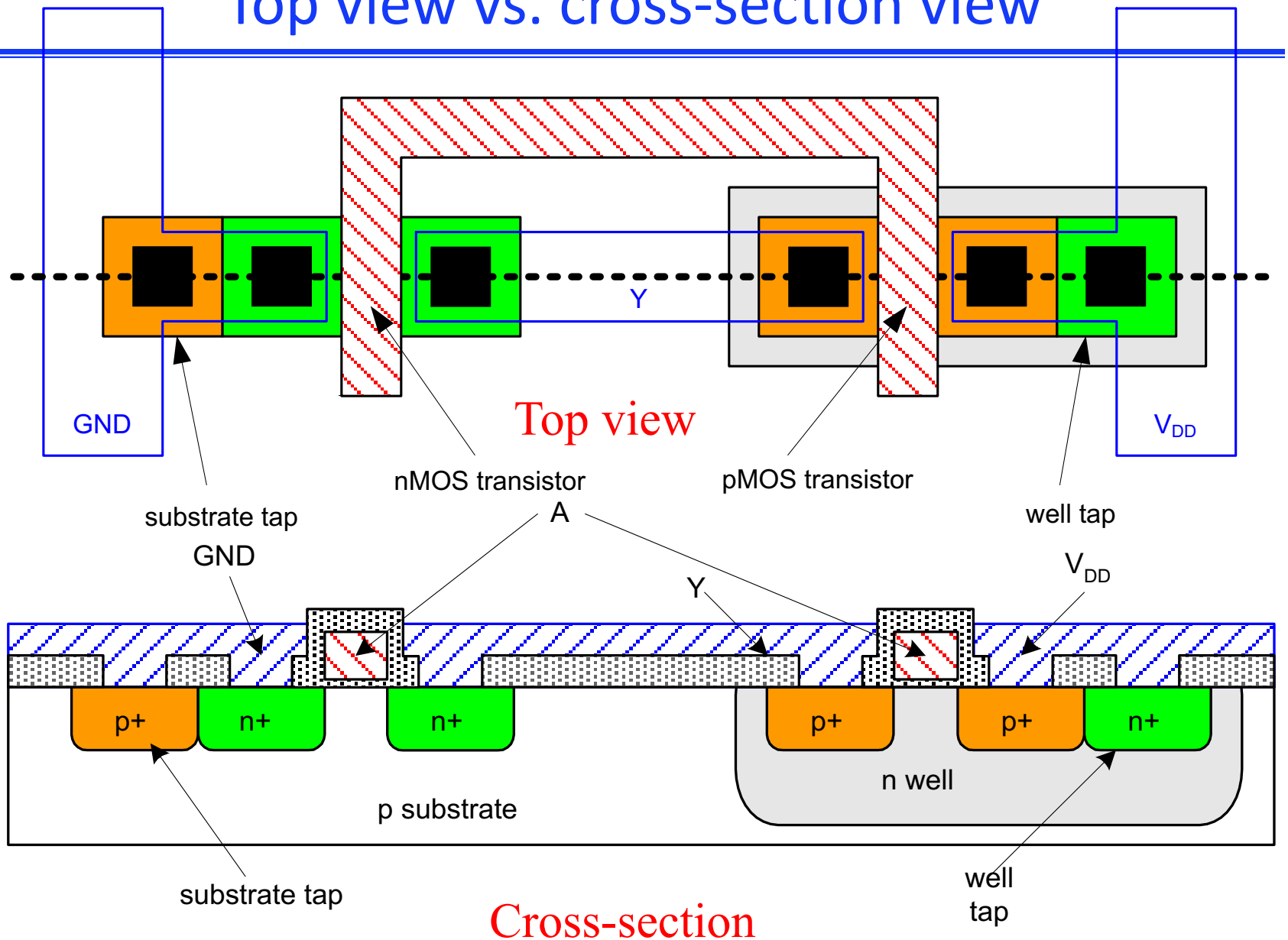
Well and Substrate Taps

- Use heavily doped well and substrate contacts / taps

To avoid forward-biasing the p-n junctions

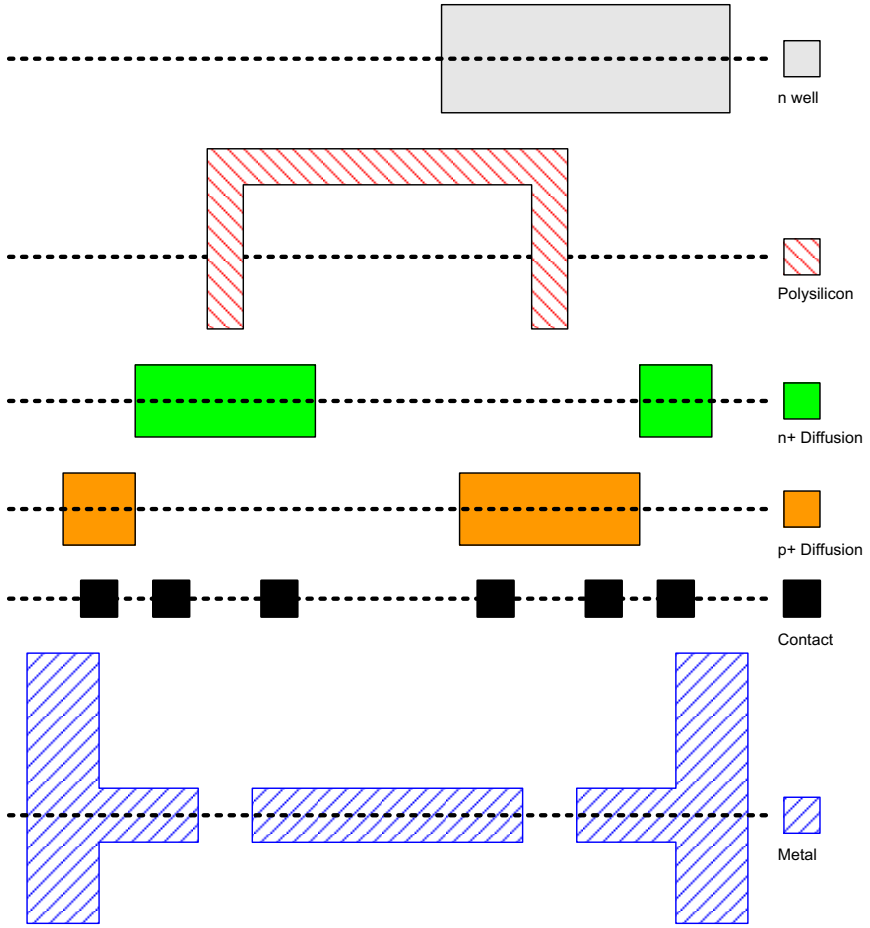


Top view vs. cross-section view



Inverter Mask Views

- Transistors and wires are defined by *masks*
- Six masks
  - n-well
  - Polysilicon
  - n+ diffusion
  - p+ diffusion
  - Contact
  - Metal



Fabrication

- Chips are built in huge factories called fabs
- Contain clean rooms as large as football fields

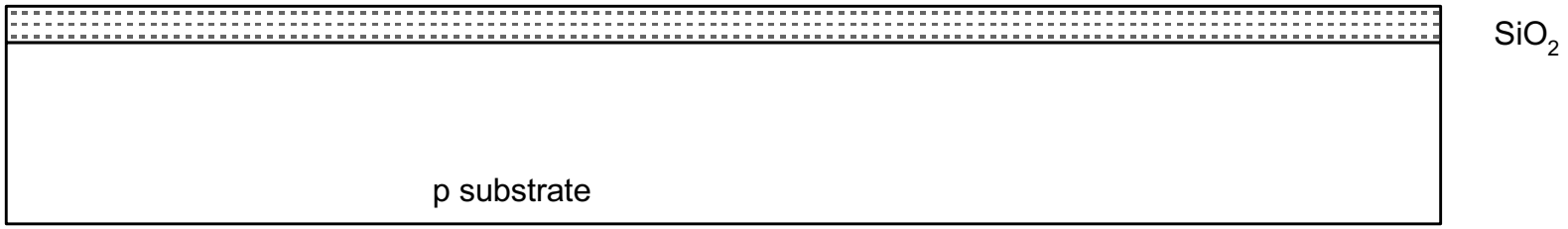
Courtesy of International Business Machines Corporation. Unauthorized use not permitted.

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 (Oxidation)
  - Remove layer where n-well should be built
  - Implant or diffuse n dopants into exposed wafer
  - Strip off SiO_2

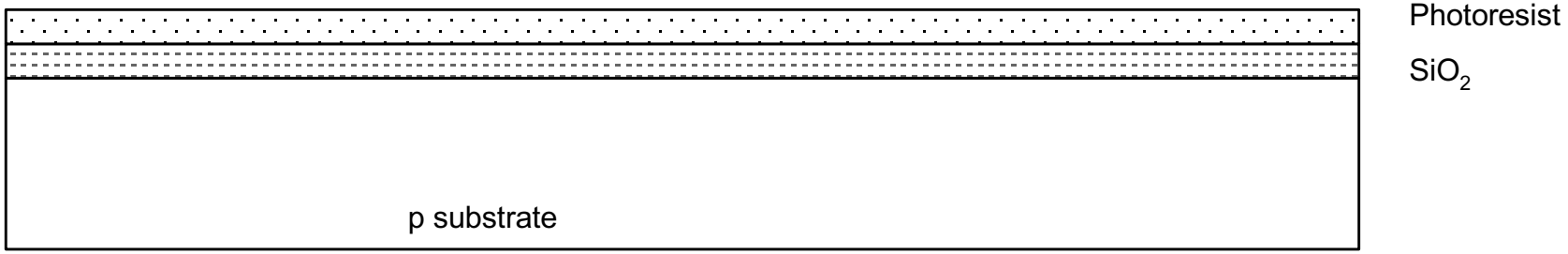
Oxidation

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



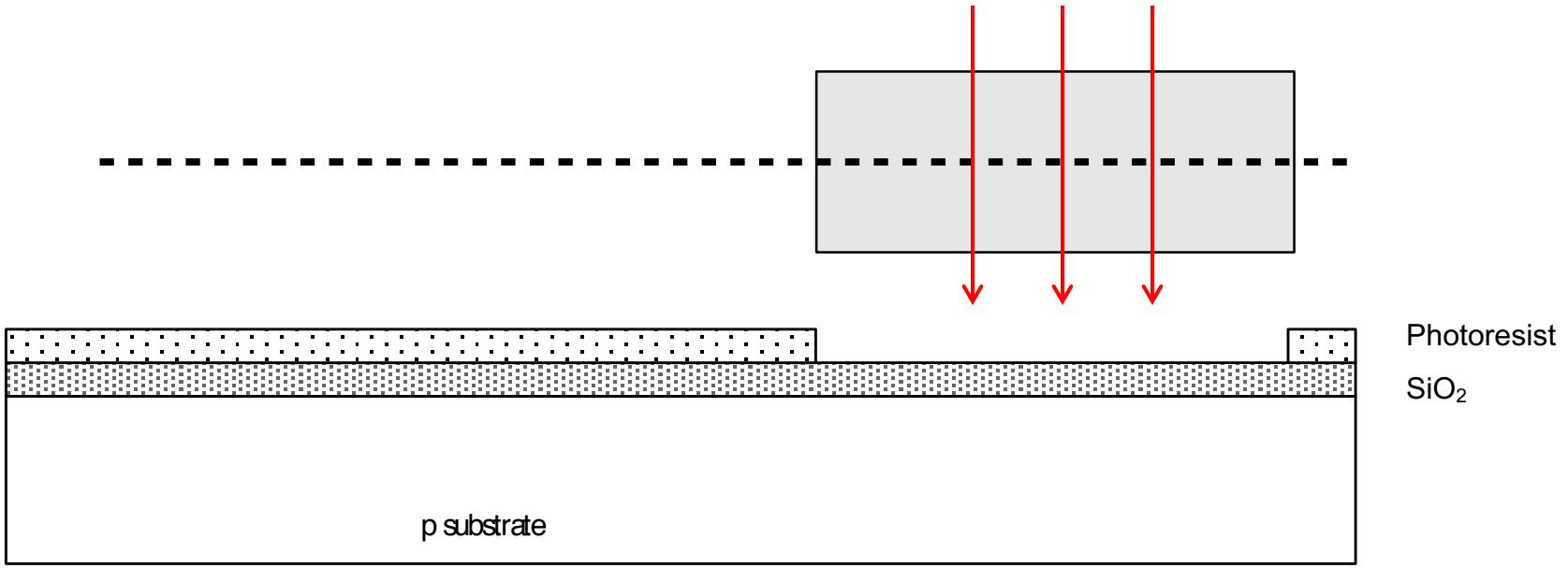
Photoresist

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



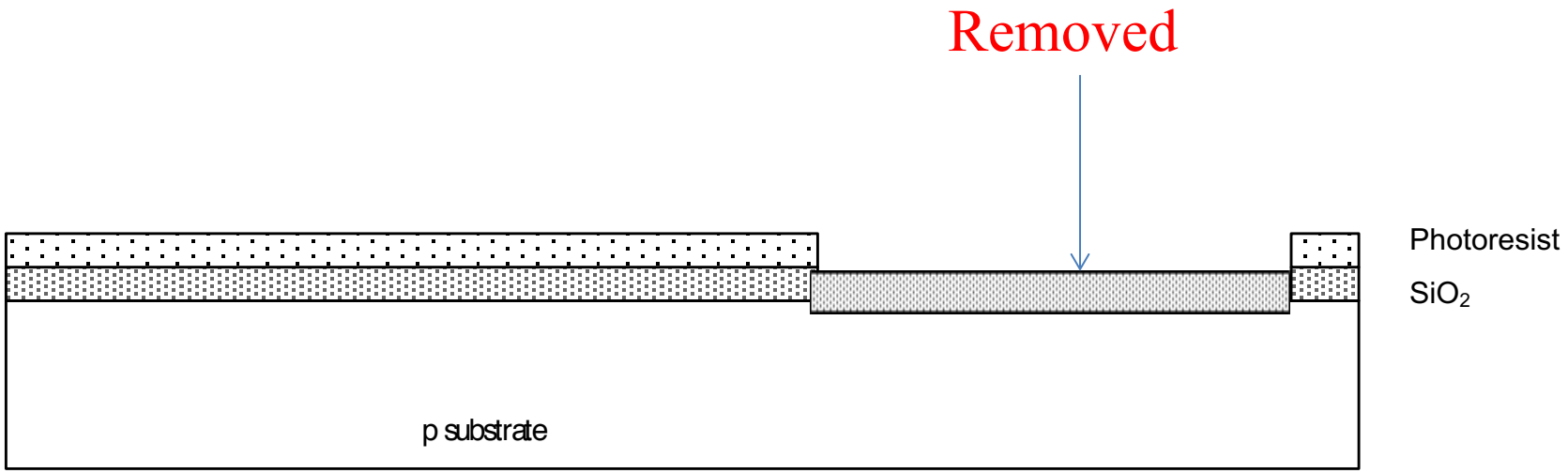
Lithography

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



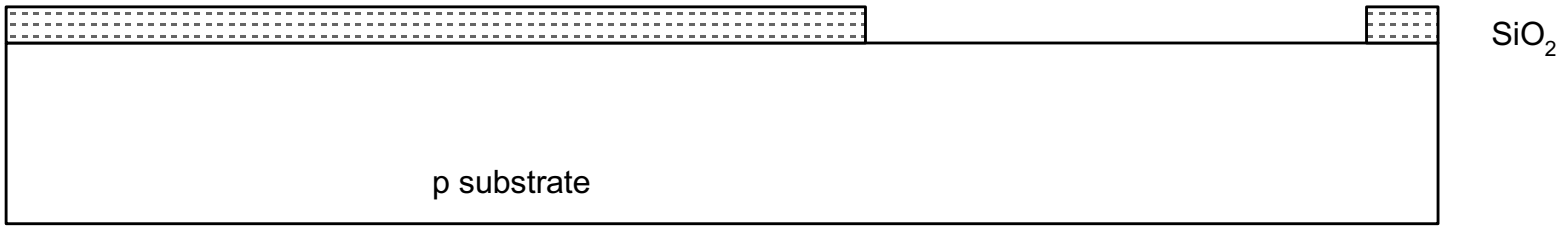
Etch

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



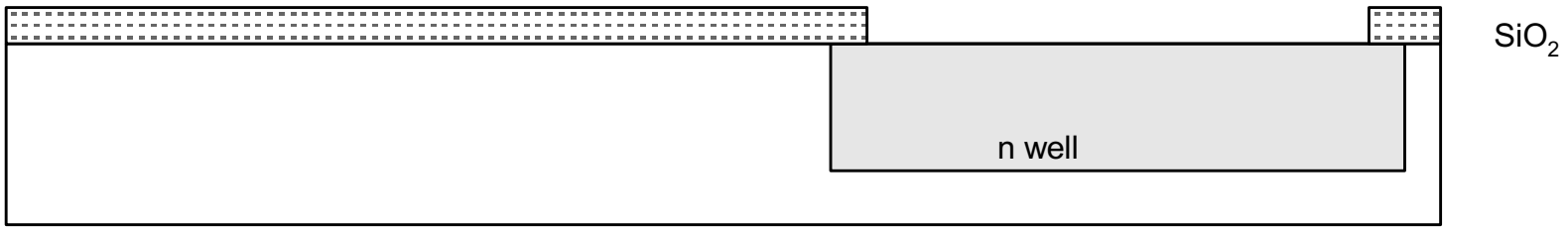
Strip Photoresist

- Strip off remaining photoresist
  - Use mixture of acids called piranha 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

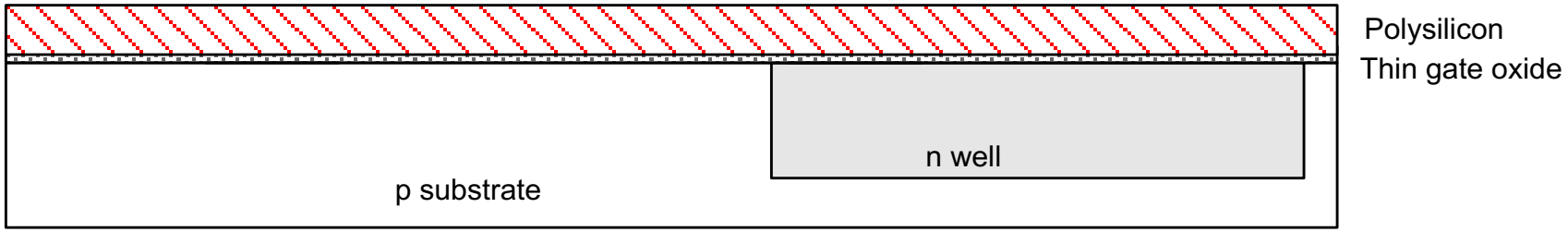


Strip Oxide

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

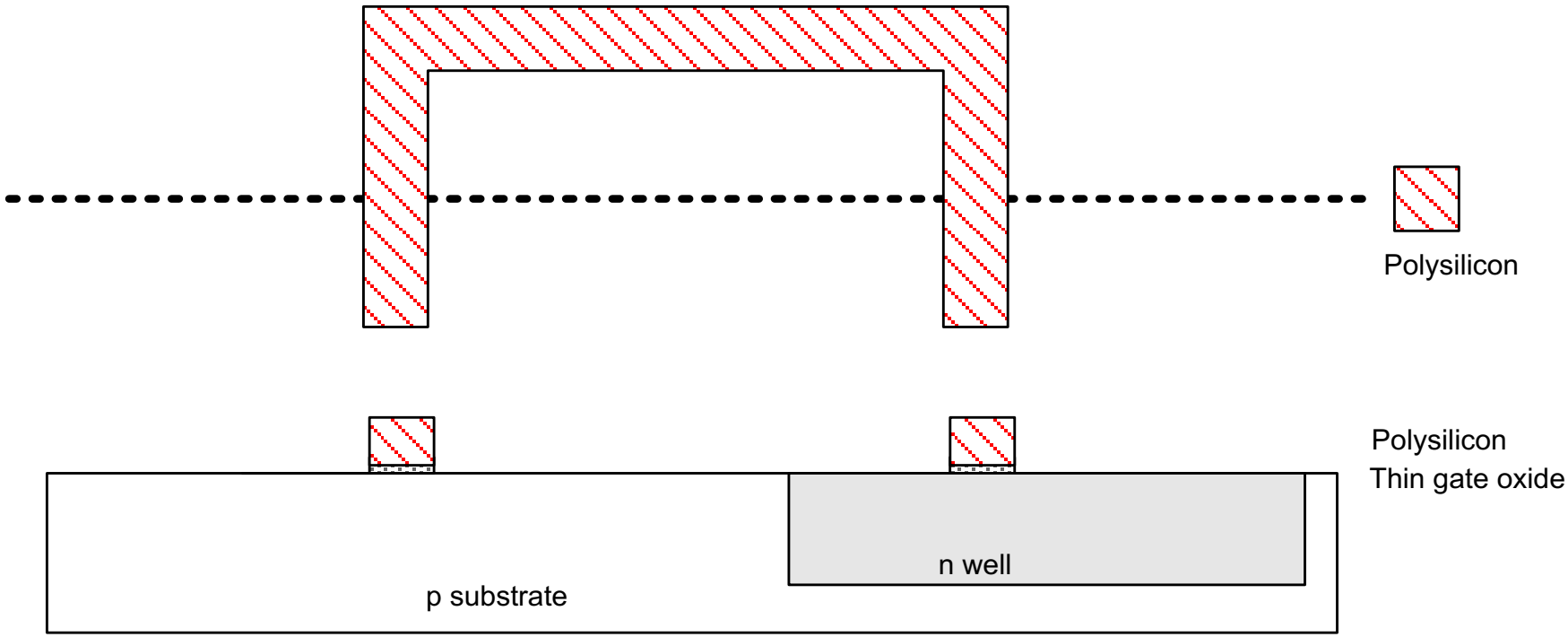
Polysilicon

- 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



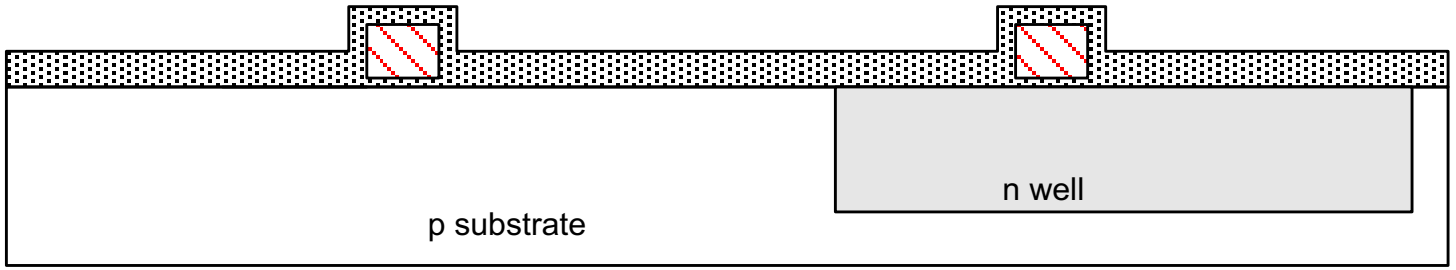
Polysilicon Patterning

- Use same lithography process to pattern polysilicon



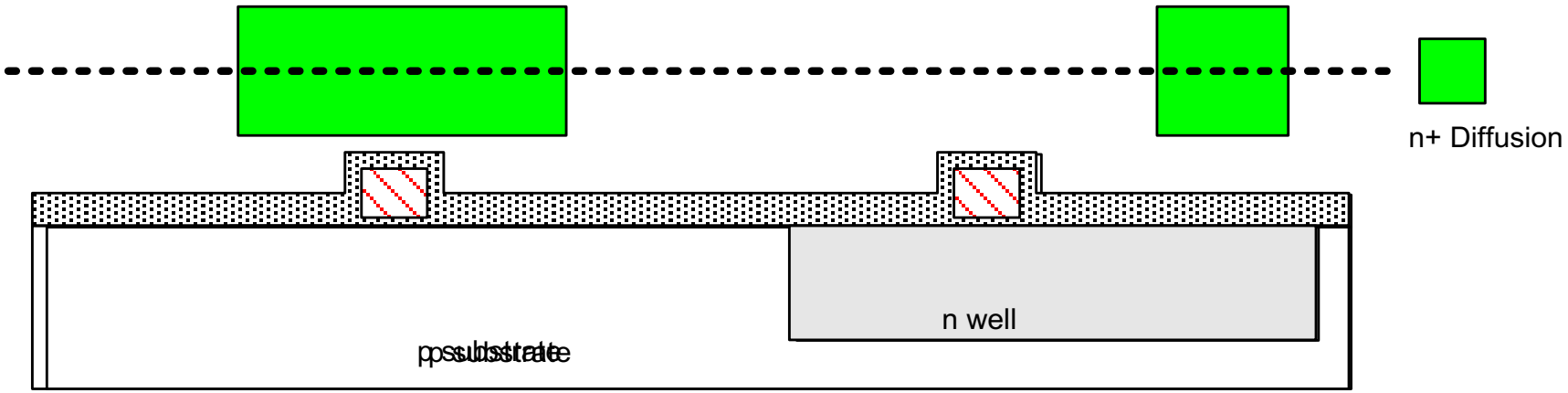
Self-Aligned Process

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



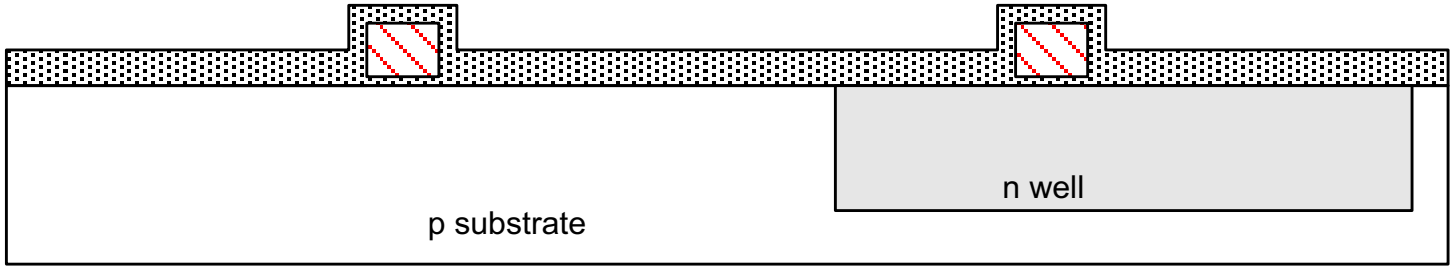
N-diffusion

- 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

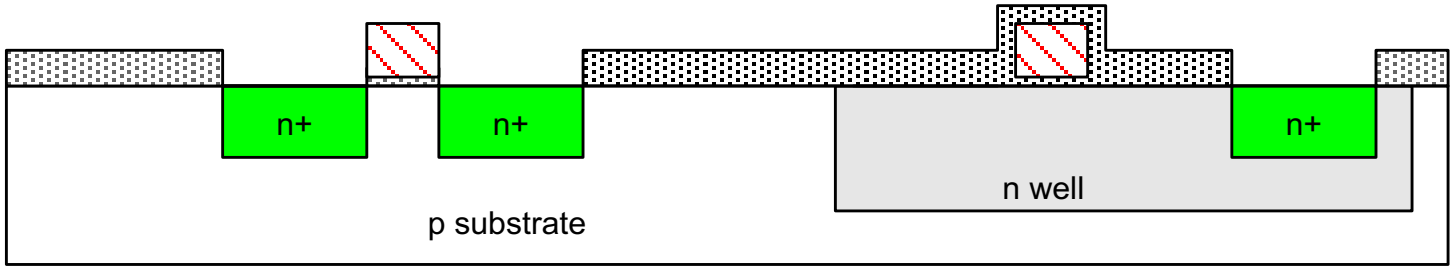


N-diffusion cont.

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



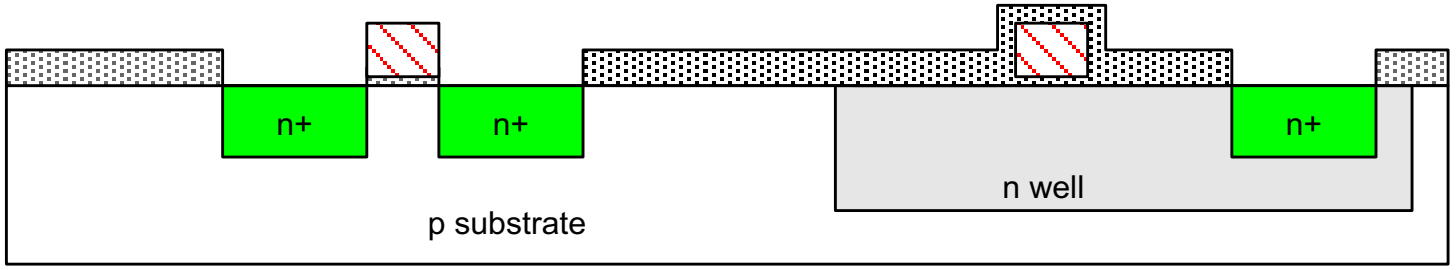
Before

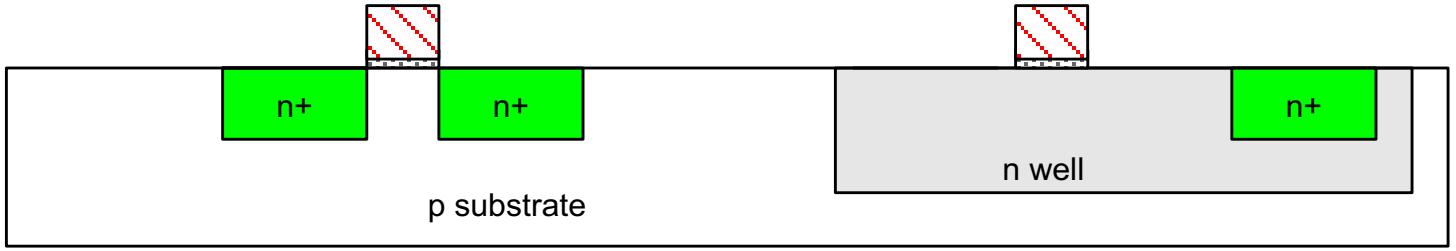


After

N-diffusion cont.

- Strip off oxide to complete patterning step

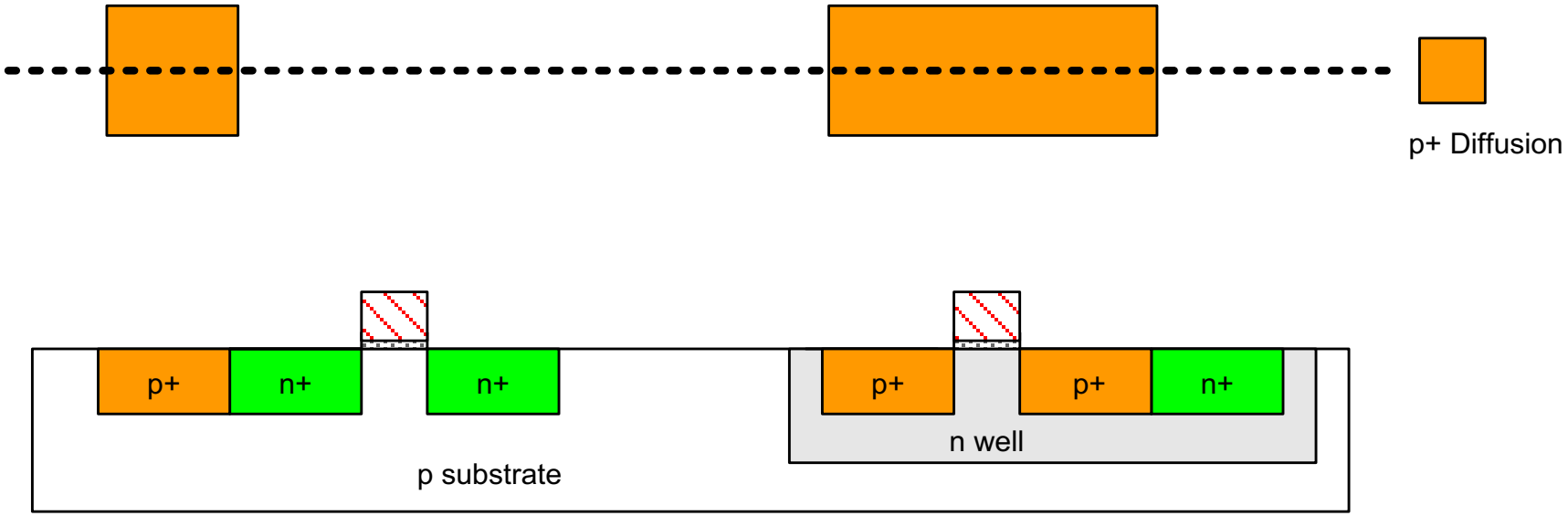




After

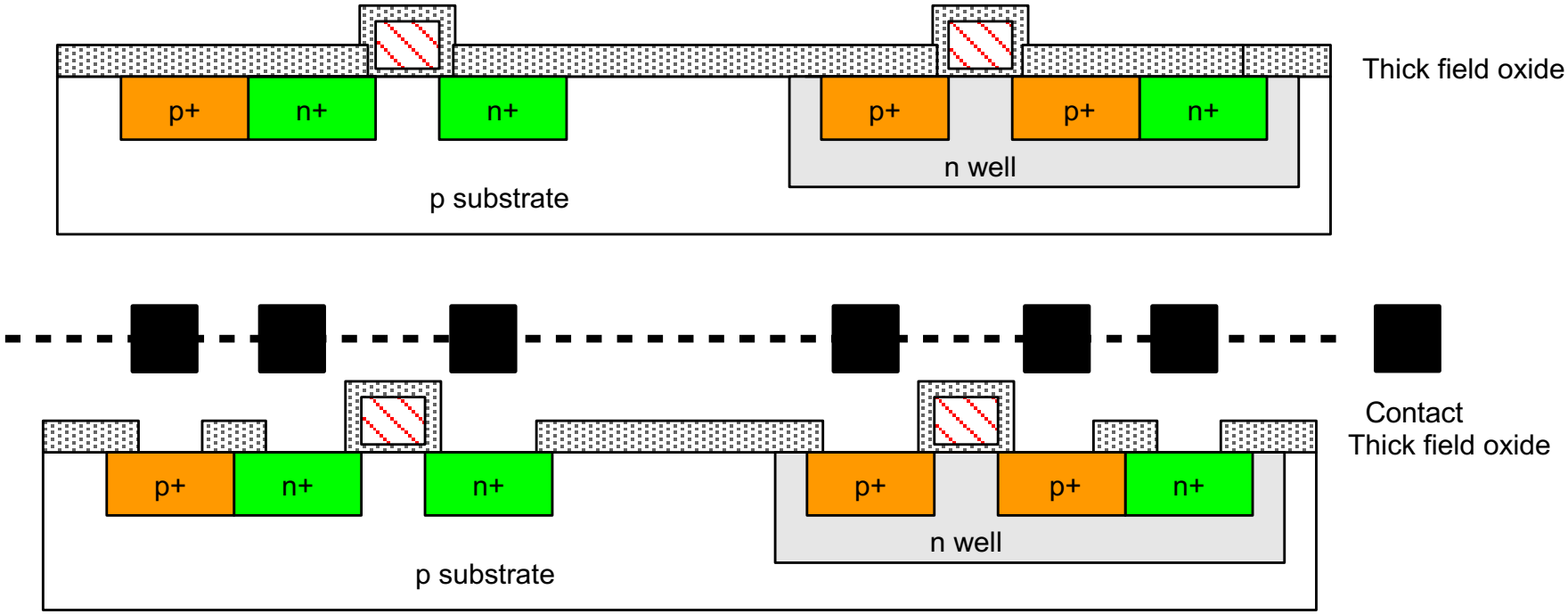
P-Diffusion

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



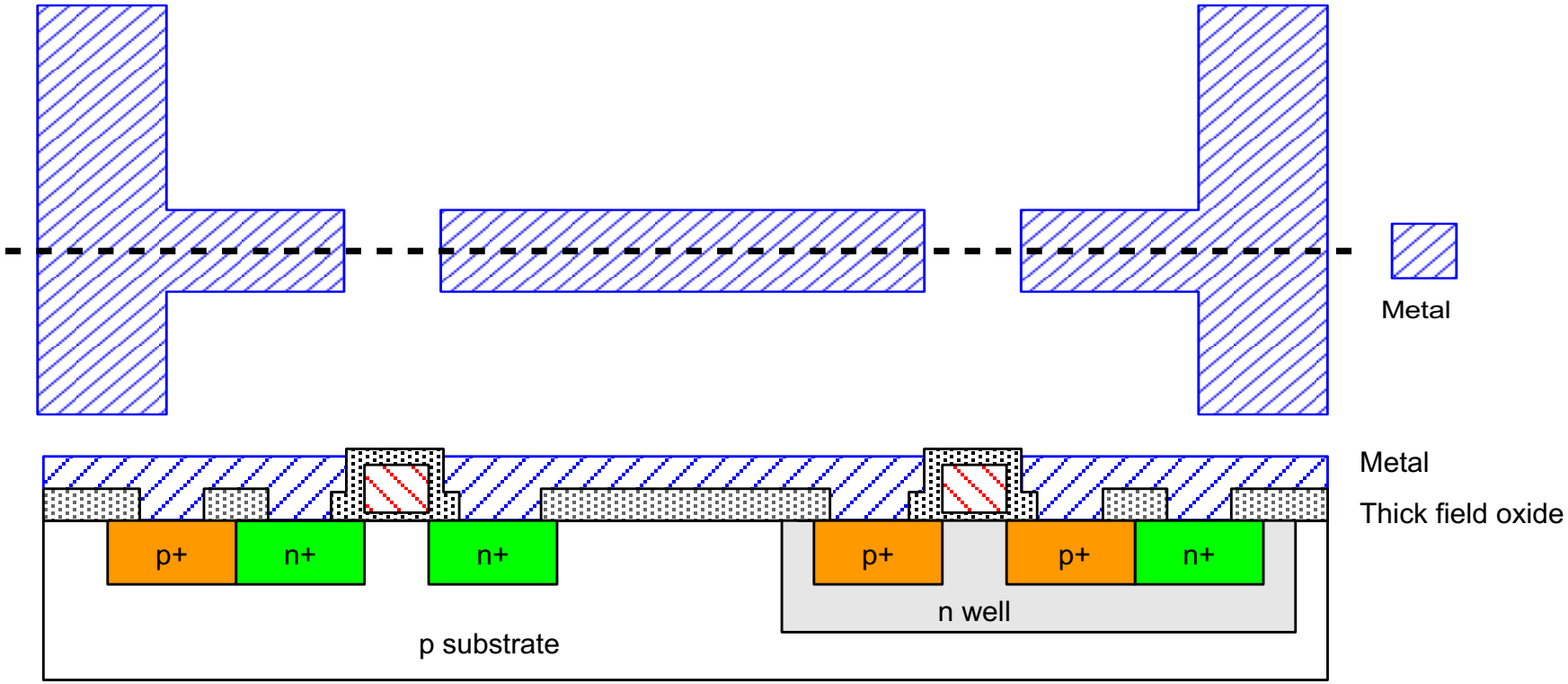
Contacts

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



Metalization

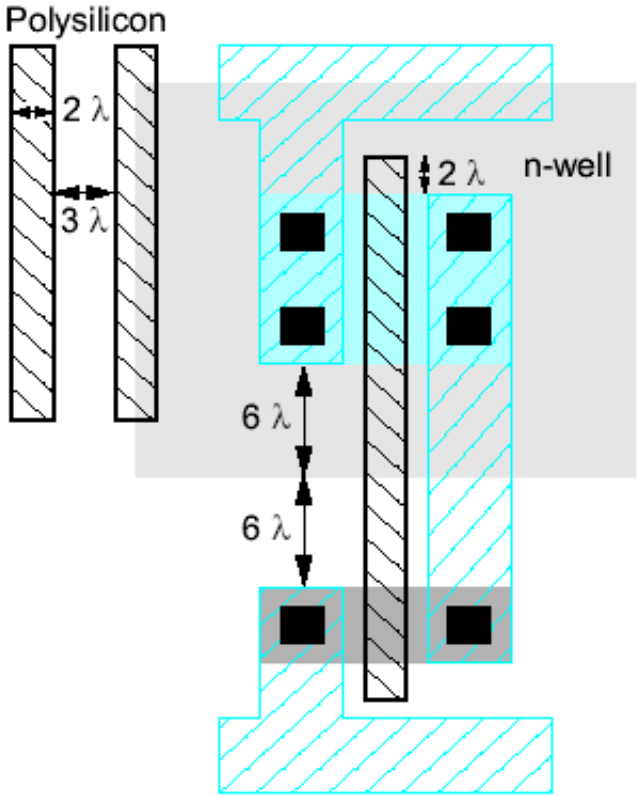
- Sputter on aluminum over whole wafer
- Pattern to remove excess metal, leaving wires



Layout Design Rules

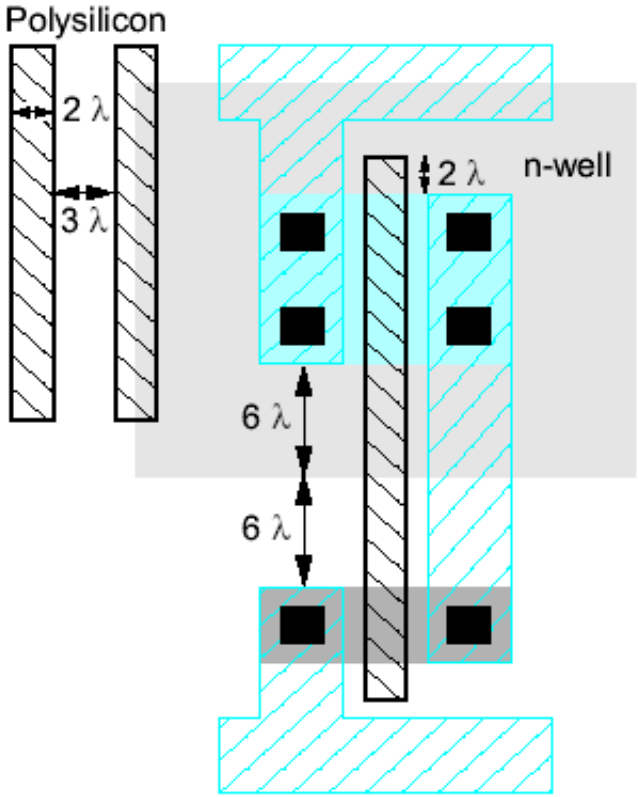
Layout

- 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



Layout and Design Rules

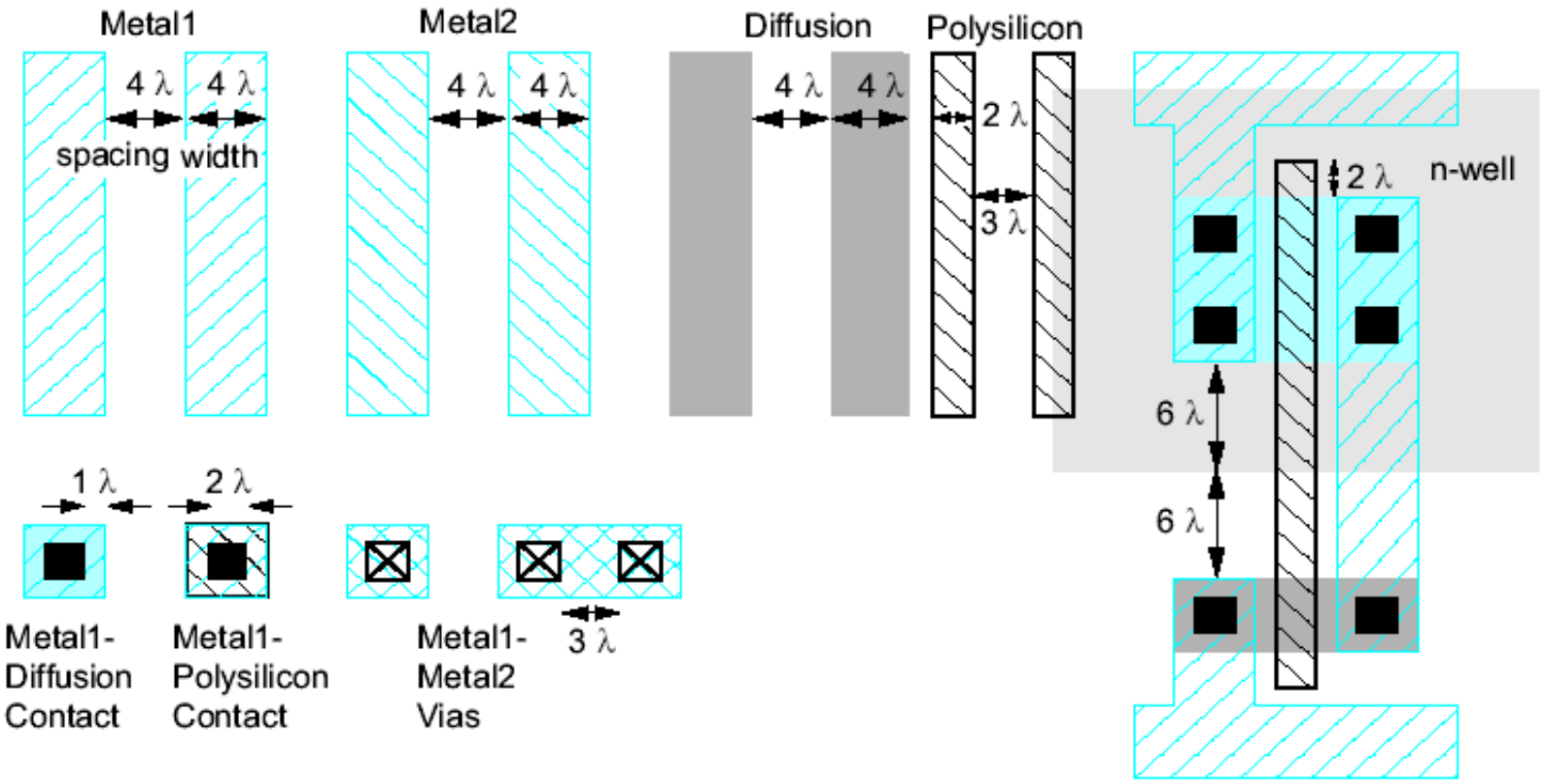
- Feature size *f* improves 30% every 3 years or so
- Design Rules to specify width to avoid breaks in a line, minimum spacing to avoid shorts.
- Express rules in terms of $\lambda = f/2$
  - E.g. $\lambda = 0.09\ \mu m$ in $0.18\ \mu m$ process



Simplified Design Rules

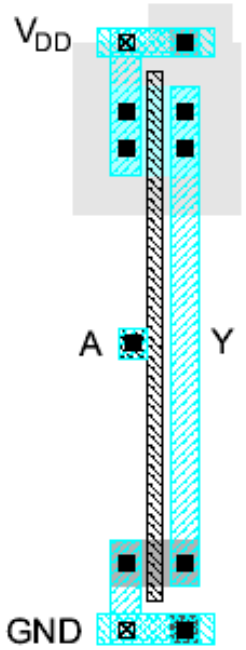
- Metal and diffusion: 4 λ
- Contact is 2 λ x2 λ , polysilicon: 2 λ

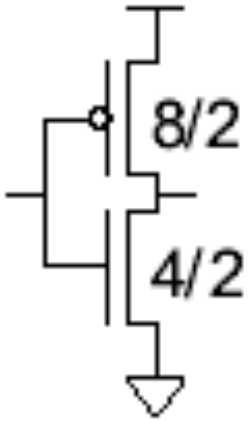
Many more….



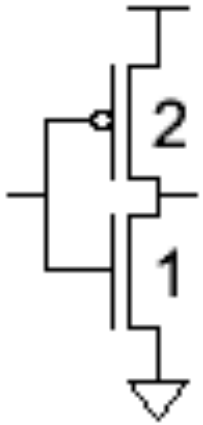
Inverter Layout

- Transistor dimensions specified as Width / Length
  - Minimum size is 4λ / 2λ, sometimes called 1 unit
  - In *f* = 0.6 μm process, this is 1.2 μm(4λ) wide, 0.6 μm(2λ) long





Inverter annotated with width and length

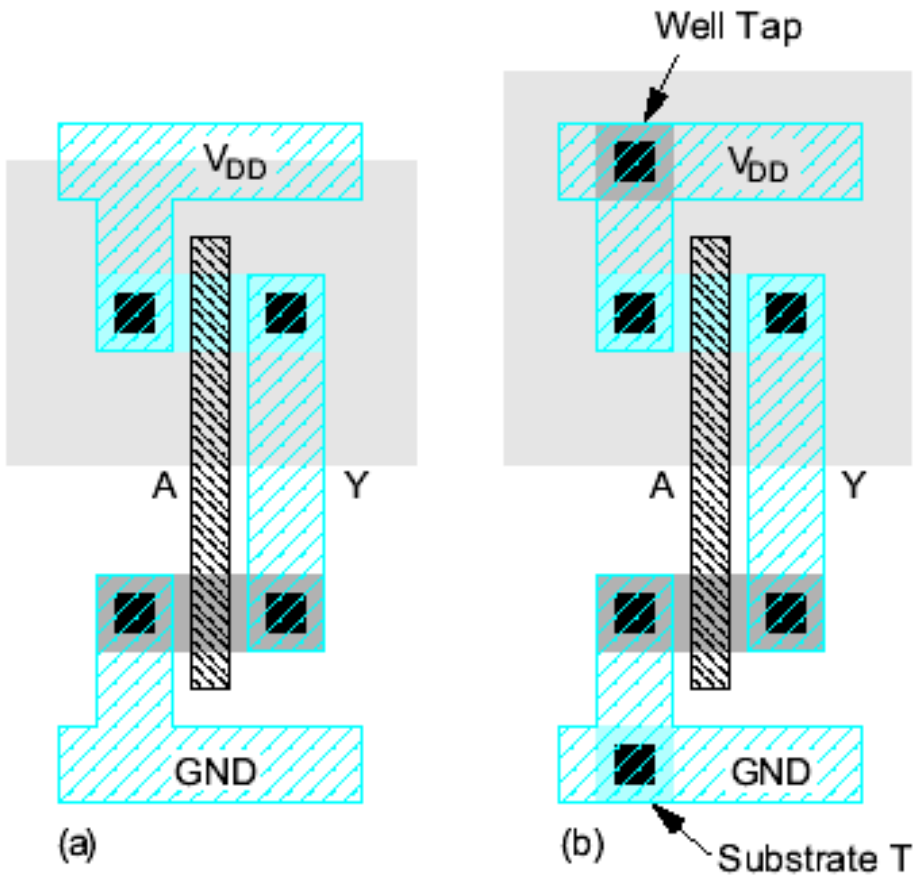


Simplified

Only width is specified . Length is assumed minimum.

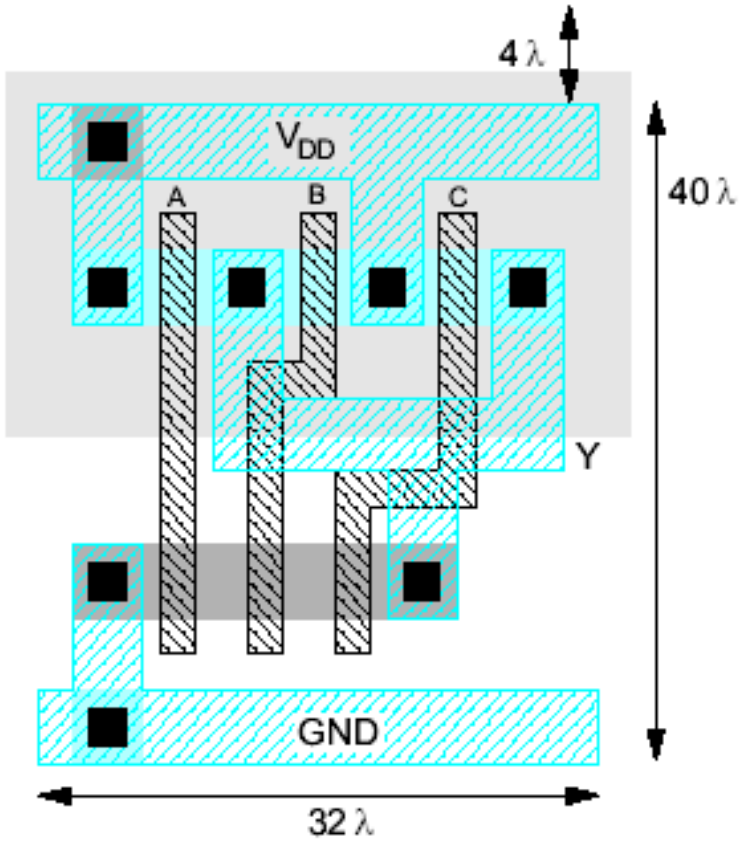
Standard cell design methodology

- Standard cell design methodology
  - V_{DD} and GND should abut (standard height)
  - Adjacent gates should satisfy design rules
  - nMOS at bottom and pMOS at top
  - All gates include well and substrate contacts



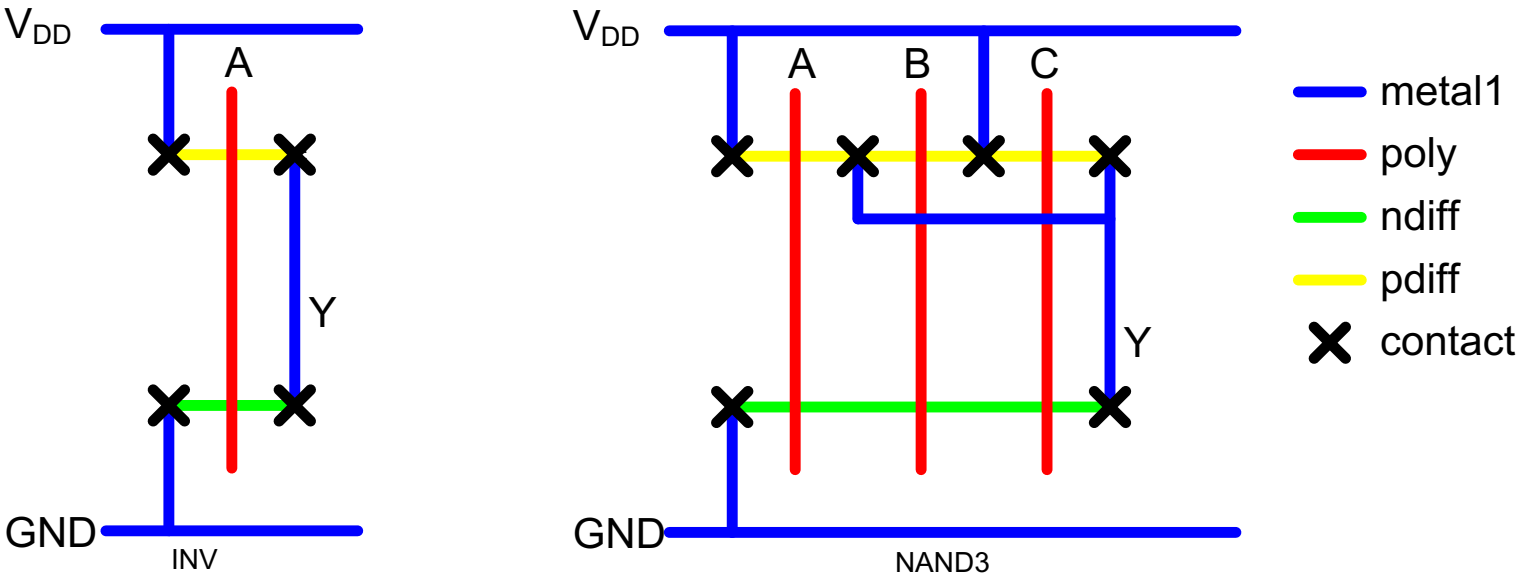
Example: NAND3

- Horizontal N-diffusion and p-diffusion strips
- Vertical polysilicon gates
- Metal1 V_{DD} rail at top
- Metal1 GND rail at bottom
- 32 λ by 40 λ



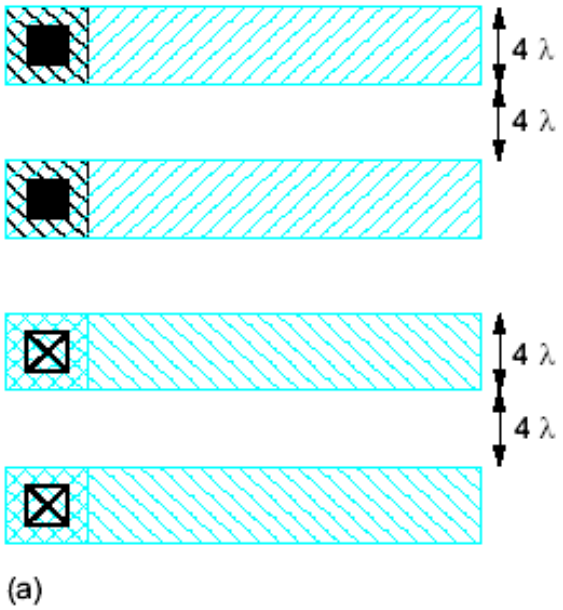
Stick Diagrams

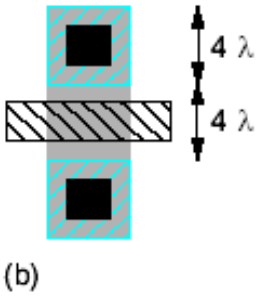
- *Stick diagrams: faster ways to plan cells and estimate area*
- *Stick diagrams* help plan layout quickly
  - Need not be to scale
  - Draw with color pencils or dry-erase markers



Wiring Tracks

- A *wiring track* is the space required for a wire
  - 4 λ width, 4 λ spacing from neighbor = 8 λ pitch
- Transistors also consume one wiring track

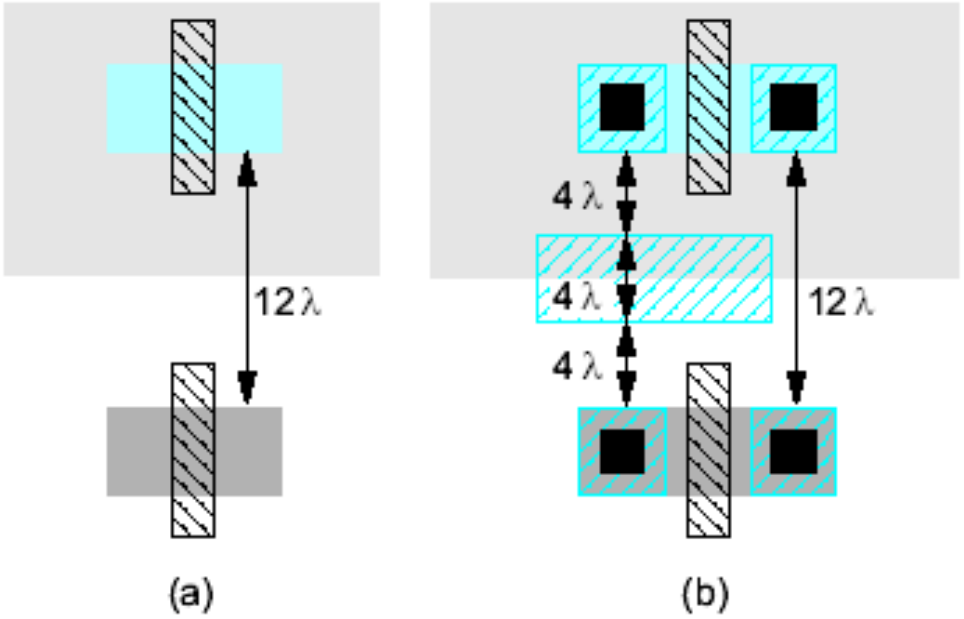




A transistor can be inserted in a wire track

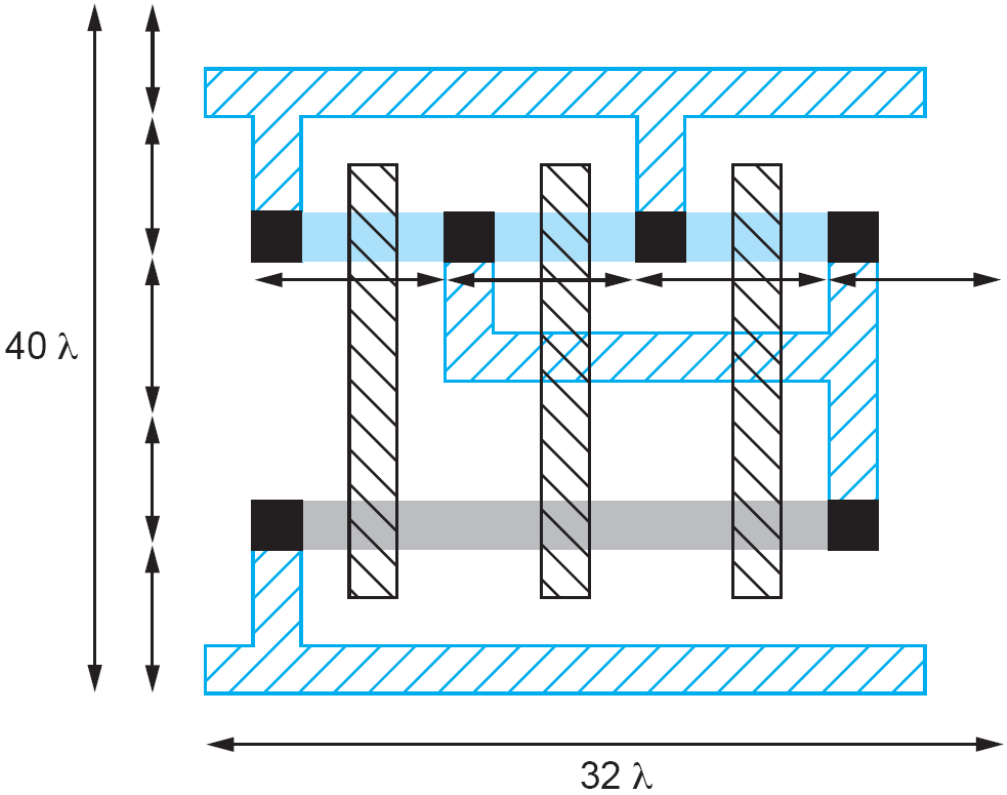
Well spacing

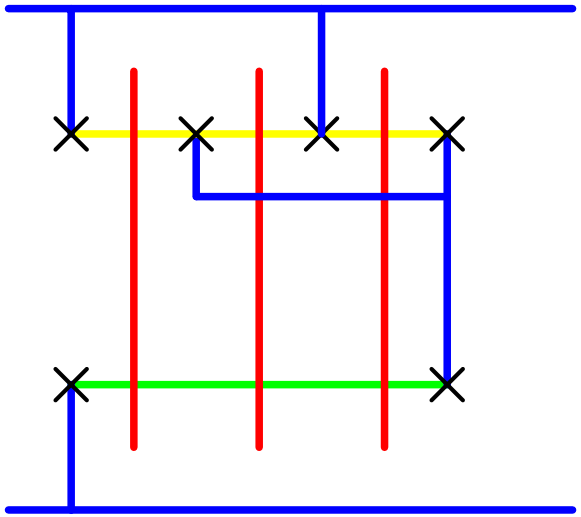
- Wells must surround transistors by 6 λ
  - Implies 12 λ between opposite transistor flavors
  - Leaves room for one wire track



Area Estimation for 3-input NAND

- Area Estimation for 3-input NAND
- 4 vertical metal track: 32 λ
- 5 vertical metal track: 40 λ





Hidden

Area Estimation

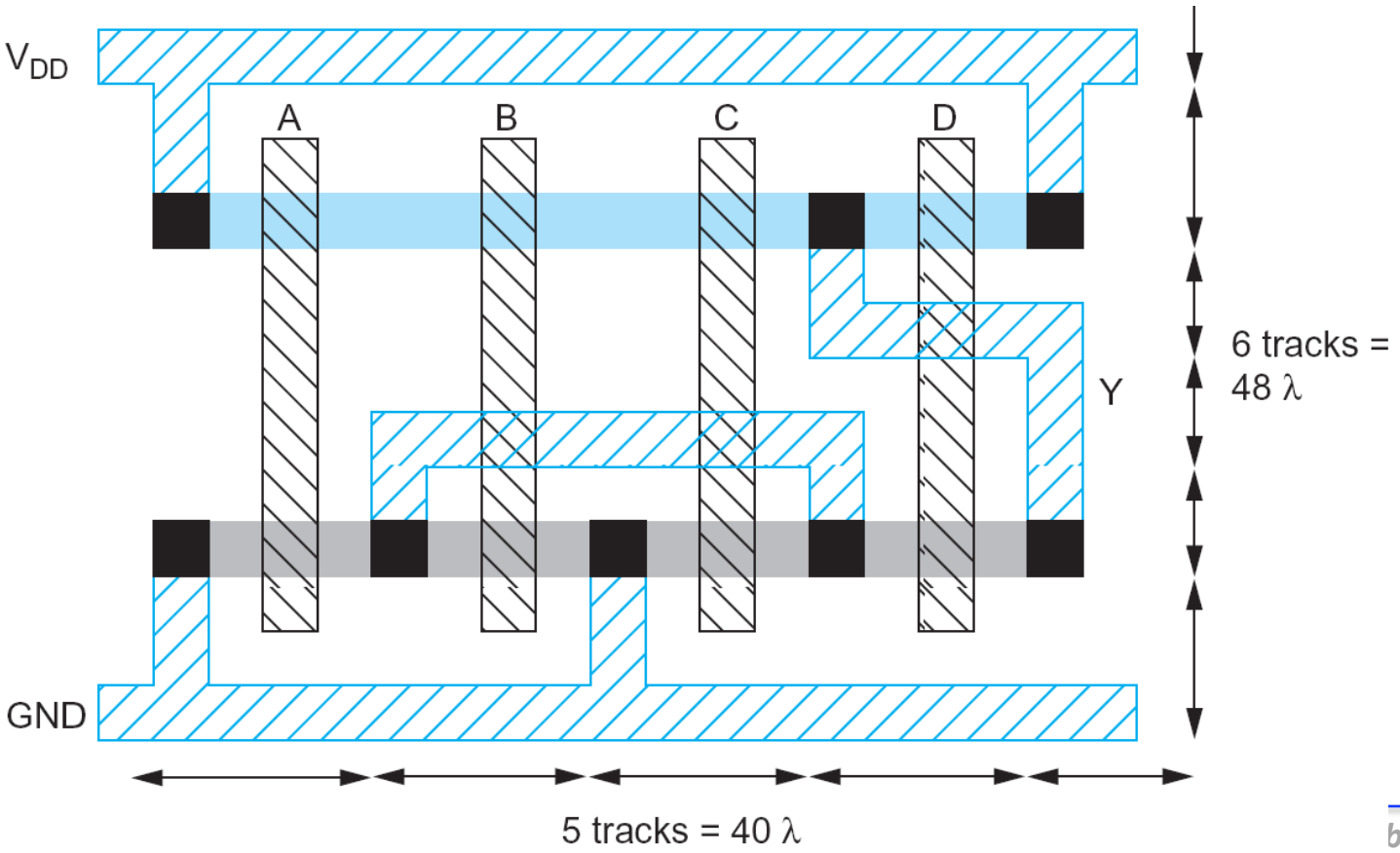
- Sketch a stick diagram for O3AI and estimate area

$$Y = \overline{(A + B + C) \cdot D}$$

Area Estimation

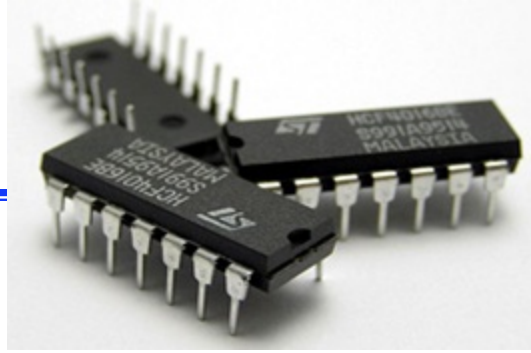
- Sketch a stick diagram for O3AI and estimate area

$$Y = \overline{(A + B + C) \cdot D}$$



Packaging

Package



- Package Functionality
  - Bring signal and supply wires in and out of die
  - Remove the heat generated by the circuit
  - Protect the die against environmental conditions such as humidity
- Major impact on the performance and power of the chip
  - 50% delay of a high-performance chip caused by package
  - Cost vs. power dissipation tradeoff

http://www.analog.com/en/technical-library/packages/sip-single-inline-package/index.html

Package (cont.)

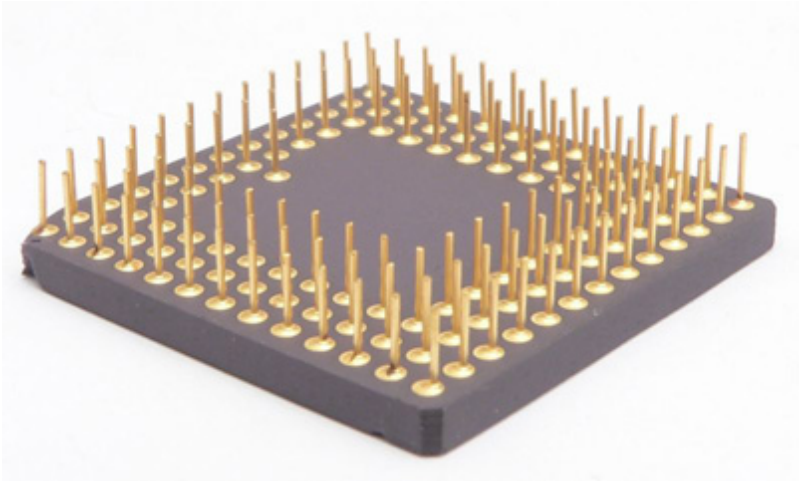
- Package requirement:
  - Electrical: Low parasitics
  - Mechanical: Reliable and robust
  - Thermal: Efficient heat removal
  - Economical: Cheap
- Package materials
  - Plastics/polymers:
    - Cheaper but inferior thermal properties
  - Ceramic: more expensive
- Two Levels
  - Die to package substrate
    - Wire bonding or tap-automated bonding
  - Package substrate to Board
    - Through hole or surface mount

Rent's rule

- The number of connections going off-chip (I/O pins) tends to be roughly proportional to the complexity of the circuitry on the chip

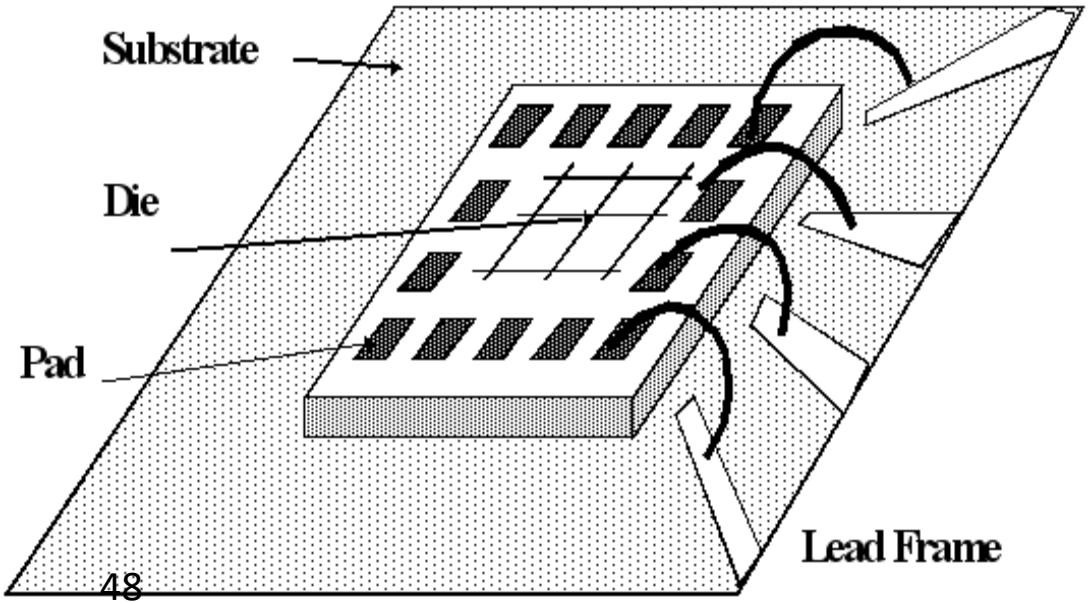
$$P = K \times G^{\beta}$$

K: average number of I/O per gates
G: the number of gates
β : the Rent exponent, 0.1 ~ 0.7
P : the number of I/O pin



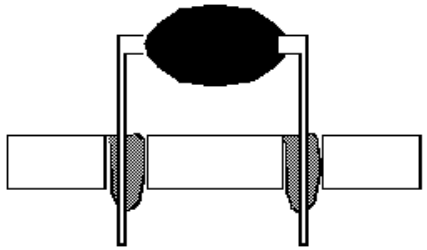
Die to package substrate - Wire bonding

- Pros: Cheap, and easier
- Cons:
  - Wire must be attached serially.
  - Difficult to find bonding pattern for larger pin counts
  - Hard to predict parasitics with irregular outlay

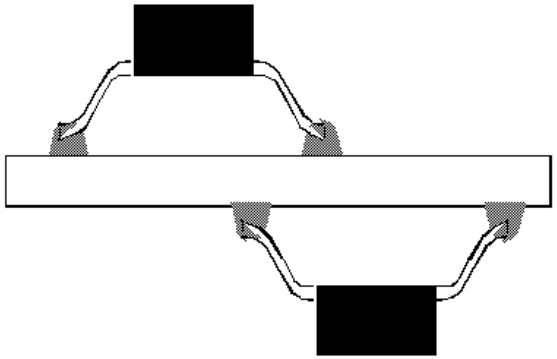


Package-to-Board Interconnect

- Connection between the package to PC board
  - Through-hole mounting
  - Surface mounting

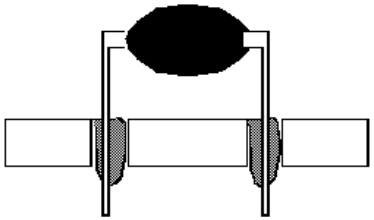


(a) Through-Hole Mounting



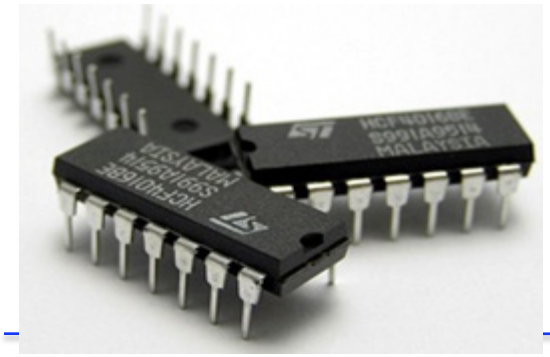
(b) Surface Mount

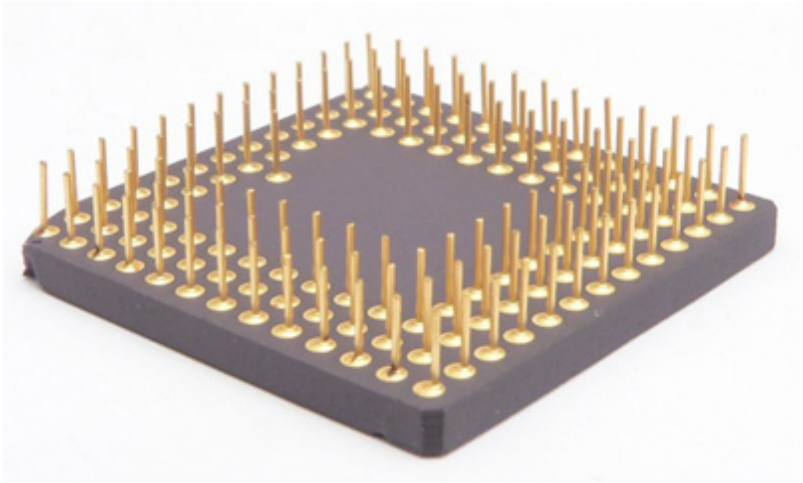
Through-hole mounting



(a) Through-Hole Mounting

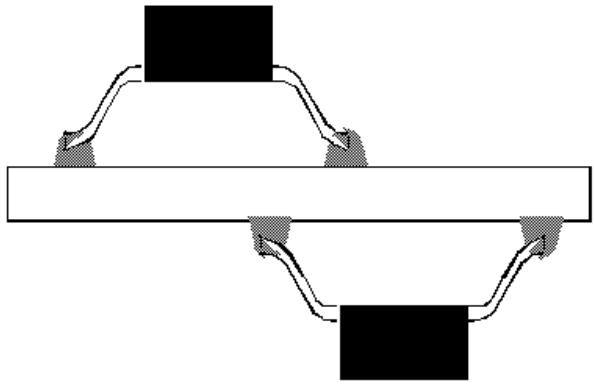
- Cheap, Mechanically reliable and sturdy
- Lower packaging density
  - A min pitch of 2.54mm between holes is required
- PC board is weaker when pin count is larger
- Additional routing layer may be needed
  - Hole is blocking routing path
- Through-hole mounting
  - DIP (Dual-in-line package)
  - PGA (Pin-grid-array package)





Surface mount

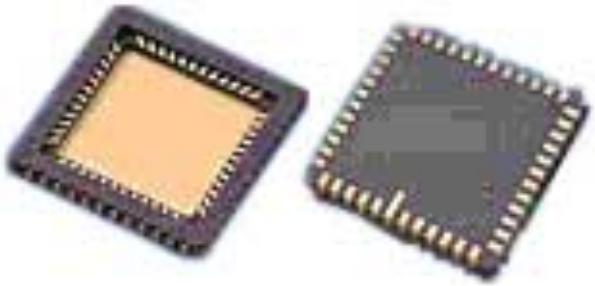
- Surface mount avoids many shortcomings of the through hole mounting, packing density is increased because
- Advantage
  - No through hole, more wiring space
  - Lead pitch is reduced
  - Both side of board can be mounted
- Disadv.
  - Chip and board connection is weaker
  - Test of the board is more complex
- Types of Surface mounting
  - Small-outline package
  - Plastic leaded Chip Carrier
  - Leadless chip carrier
  - Ball grid array (BGA)



(b) Surface Mount

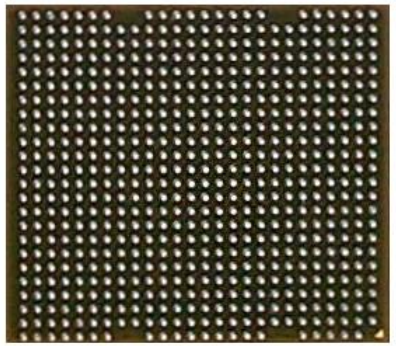
Small outline package

Leadless Chip Carrier



Plastic leaded Chip Carrier

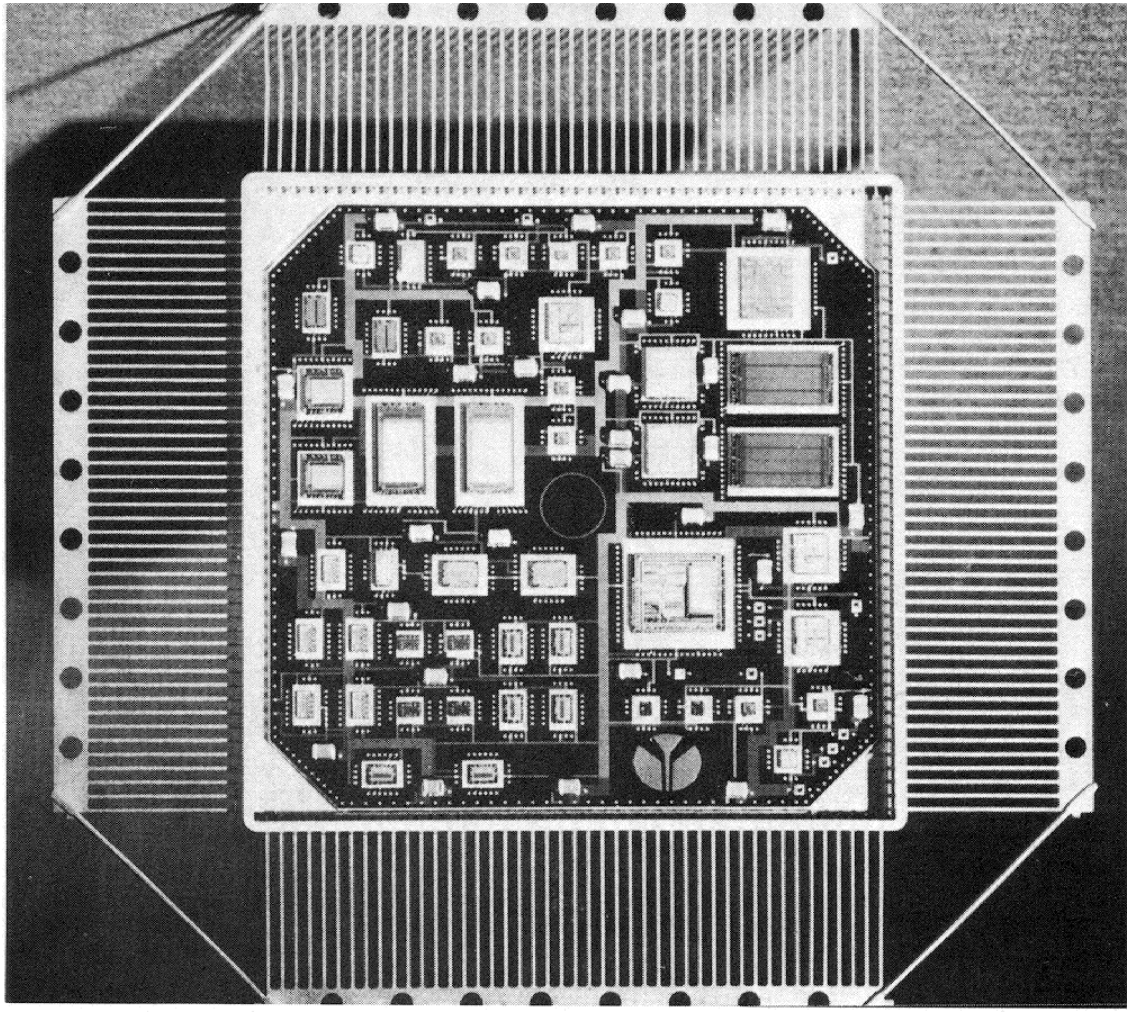
Top



Bottom

BGA

Multi-Chip Modules



Thermal Considerations in packaging

- As the power of consumption of integrated circuits, it becomes increasingly important to efficiently remove the heat
- To prevent the failure of the chip, the temperature of the die must be kept within certain ranges
  - Range for commercial devices: 0^oC to 70^oC
  - Range for military parts: -55^oC to 125^oC
- Improving heat removal efficiency
  - Better material
  - Finned metal heat sinks
  - Force air, liquid…etc.

Heat flow equation

- The temperature difference between chip and environment is

$$\Delta T = T_{chip} - T_{env} = \theta Q$$

- T_{chip}: chip temperature, T_{env}: env temperature
- Θ: thermal resistance, expressed in $^{o}C/W$
- Q is the heat flow (in Watt) (c.f. V= R* I)
- Example: A DIP has a thermal resistance of $25^{o}C/W$ when natural air is used. How many watts of heat can it remove, when the temperature between the die and environment is $75^{o}C$

Heat flow equation

- The temperature difference between chip and environment is

$$\Delta T = T_{chip} - T_{env} = \theta Q$$

- T_{chip}: chip temperature, T_{env}: env temperature
- Θ: thermal resistance, expressed in $^{o}C/W$
- Q is the heat flow (in Watt) (c.f. V= R* I)
- Example: A DIP has a thermal resistance of $25^{o}C/W$ when natural air is used. How many watts of heat can it remove, when the temperature between the die and environment is $75^{o}C$

Ans: 3W. Because 75/25 =3

Coping with Complexity

- How to design System-on-Chip?
  - Many millions (even billions!) of transistors
  - Tens to hundreds of engineers
- Structured Design
- Design Partitioning

Structured Design

- **Hierarchy**: Divide and Conquer
  - Recursively system into modules
- **Regularity**
  - Reuse modules wherever possible
  - Ex: Standard cell library
- **Modularity**: well-formed interfaces
  - Allows modules to be treated as black boxes
- **Locality**
  - Physical and temporal

Backup slides