

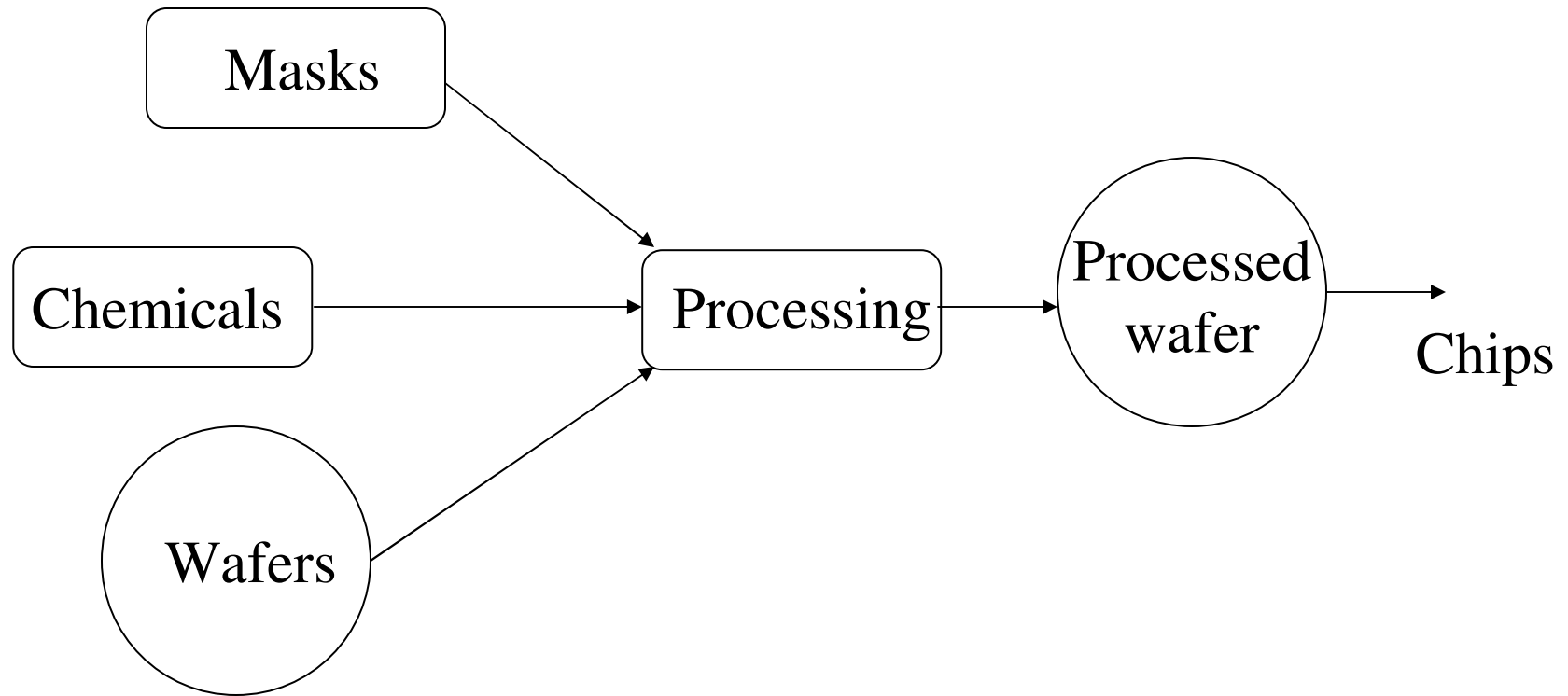
# Fabrication and Manufacturing (Basics)

- Batch processes
  - Fabrication time independent of design complexity
- Standard process
  - Customization by masks
  - Each mask defines geometry on one layer
  - Lower-level masks define transistors
  - Higher-level masks define wiring
- Silicon is neat stuff
  - Oxide protects things from impurities
  - Can be etched selectively on silicon or metal
- Can be doped
  - Add P or As impurities

# CMOS Fabrication

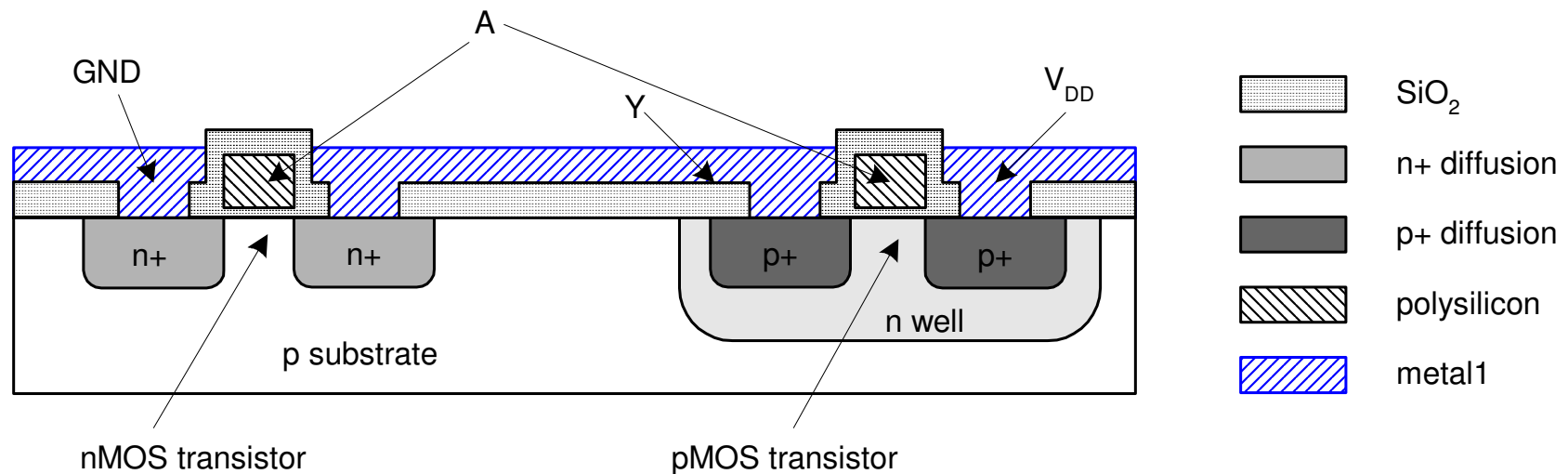
- CMOS transistors are fabricated on silicon wafer
- Lithography process similar to printing press
- On each step, different materials are deposited or etched
- Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process

# Making Chips



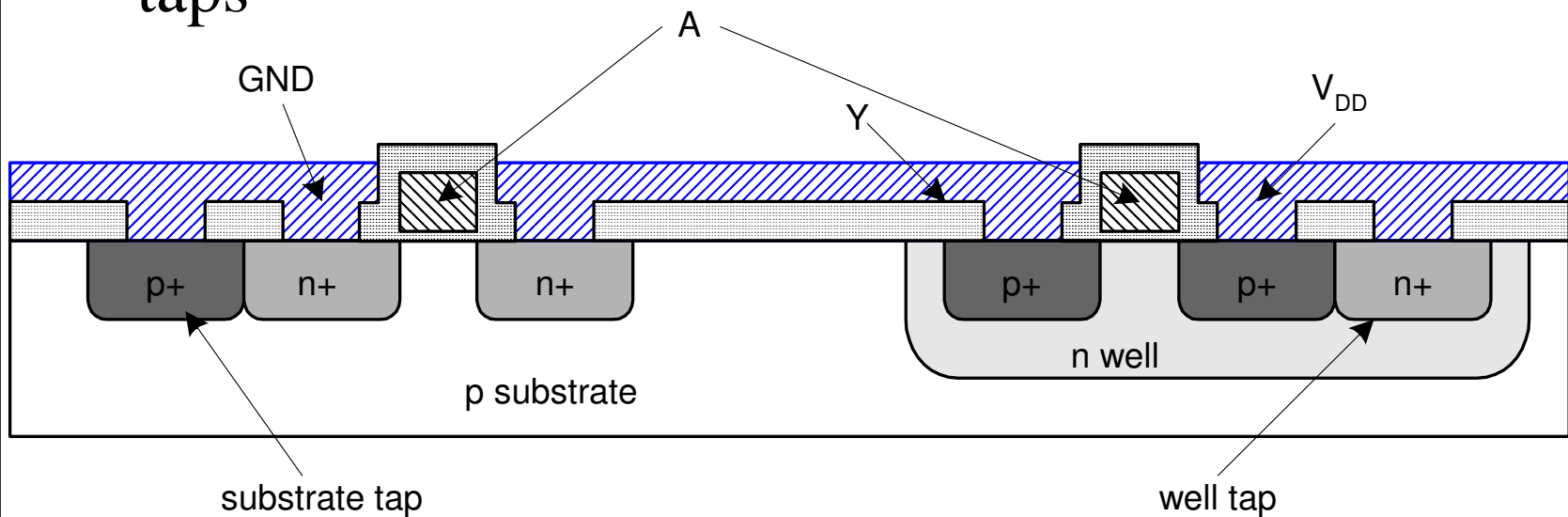
# Inverter Cross-section

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



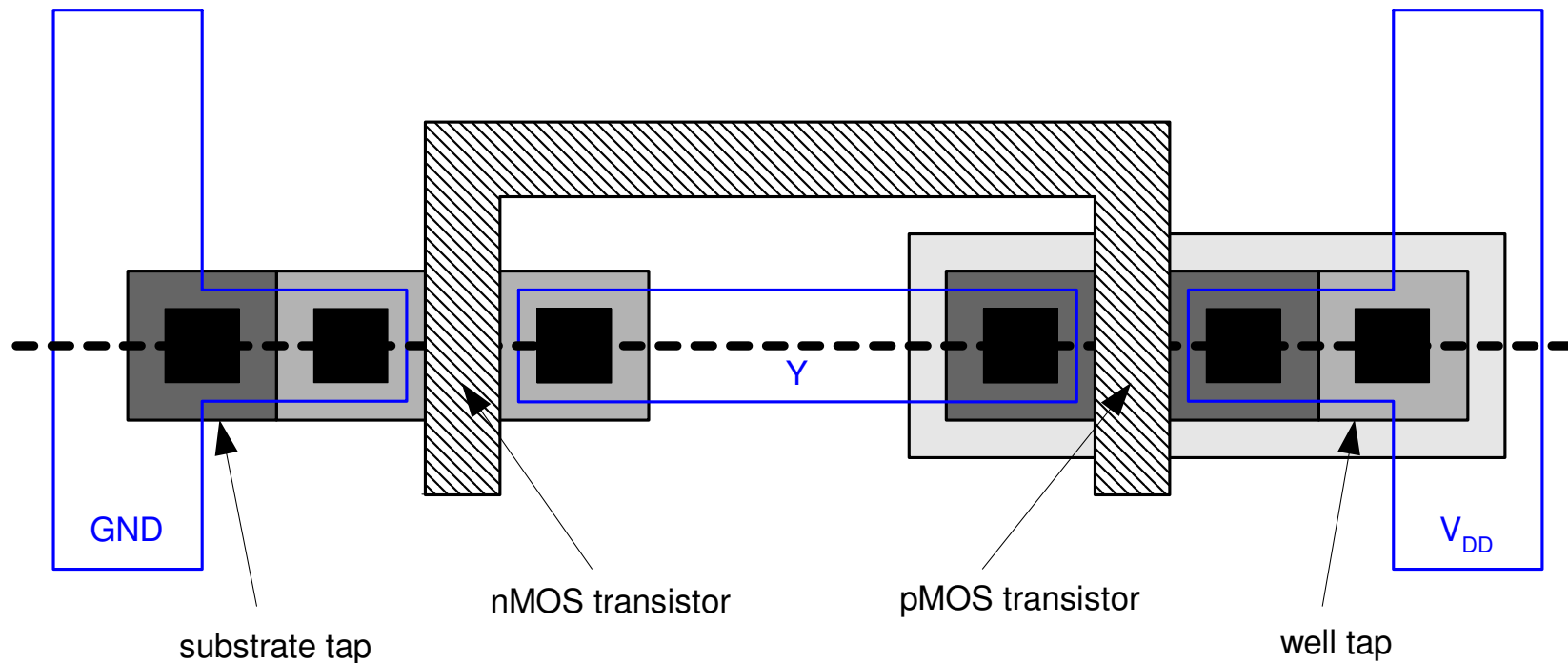
# Well and Substrate Taps

- 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



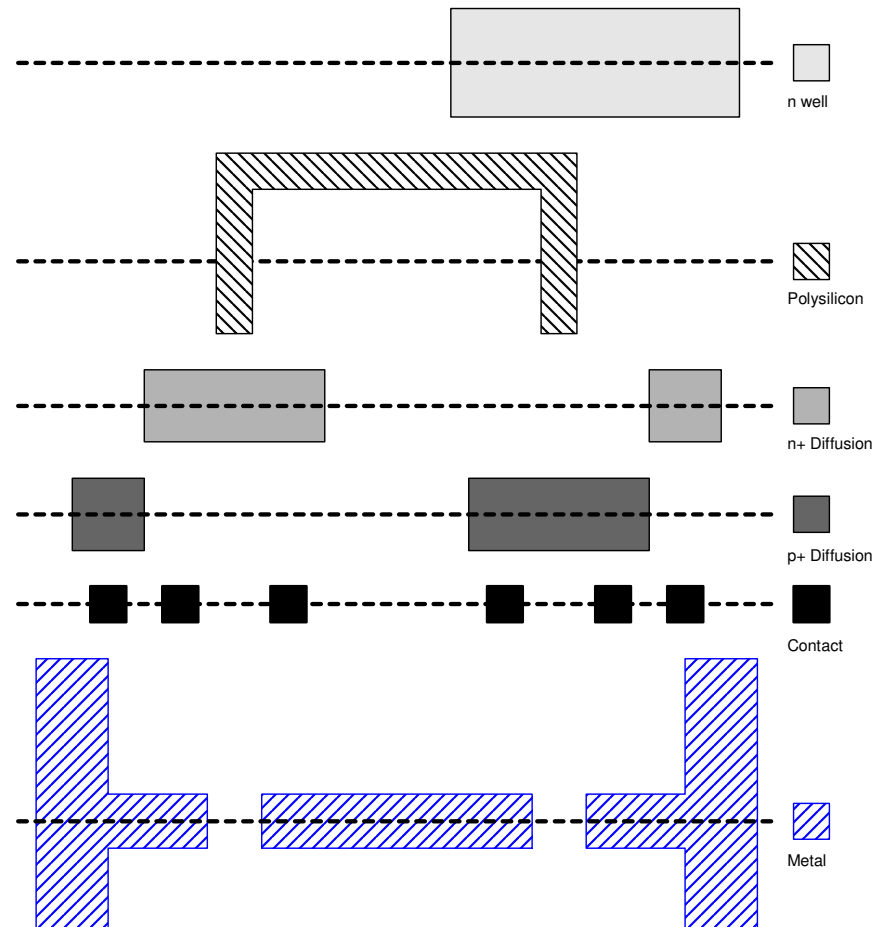
# Inverter Mask Set

- 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



# Basic Processing Steps

- N-diffusion created by doping regions of the substrate
- Poly and metal are laid over the substrate, with oxide to insulate them from substrate and each other
- Wires are added in layers, alternating with oxide
- Vias are cut in the oxide



# Fabrication Steps

- Features are patterned on a wafer by a photolithographic process
  - *Photo-light lithography, n.* process of printing from a plane surface on which image to be printed is ink-receptive and the blank area is ink-repellant
- Cover the wafer with a light-sensitive, organic material called *photoresist*
- Expose to light with the proper pattern (mask)
- Patterns left by photoresist can be used to control where oxide is grown or materials are placed on surface of wafer

# Fabrication Steps

- Layout contains information on what patterns have to be made on the wafer
- Masks are created using the layout information provided by the designer
- Procedure involves selective removal of the oxide
  - Coat the oxide with photoresist, polymerized by UV light (applied through mask)
  - Polymerized photoresist dissolves in acid
  - Photoresist itself is acid-resistant

# 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  $\text{SiO}_2$  (oxide)
  - Remove layer where n-well should be built
  - Implant or diffuse n dopants into exposed wafer
  - Strip off  $\text{SiO}_2$

p substrate

# Oxidation

- Grow  $\text{SiO}_2$  on top of Si wafer
  - 900 – 1200 C with  $\text{H}_2\text{O}$  or  $\text{O}_2$  in oxidation furnace



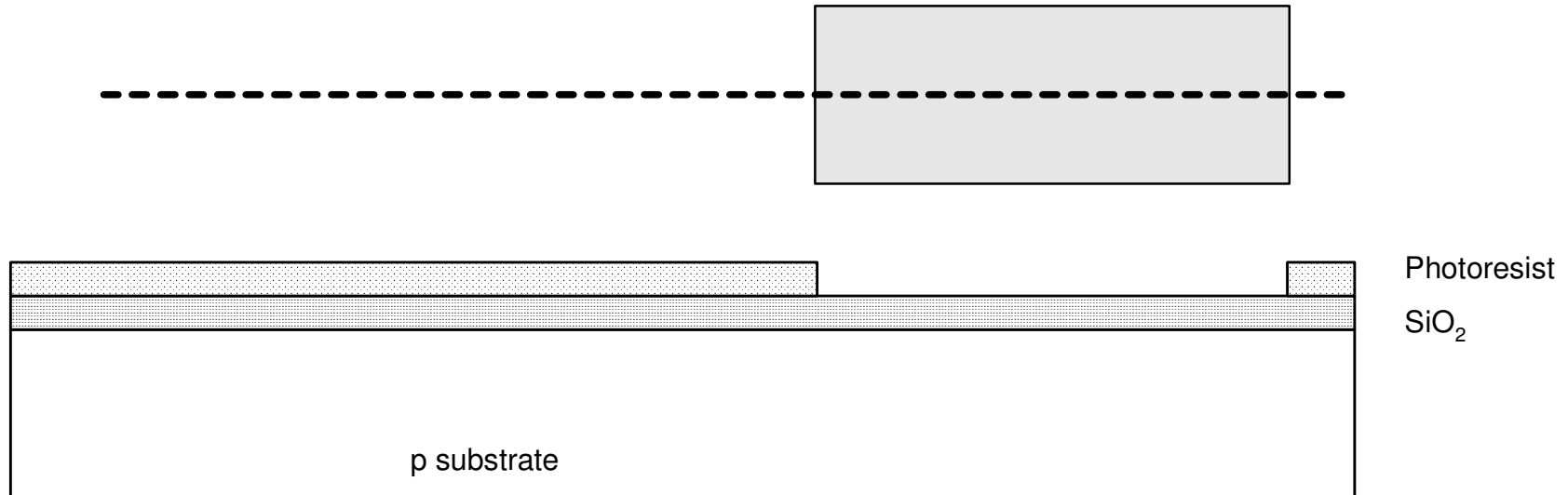
# Photoresist

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



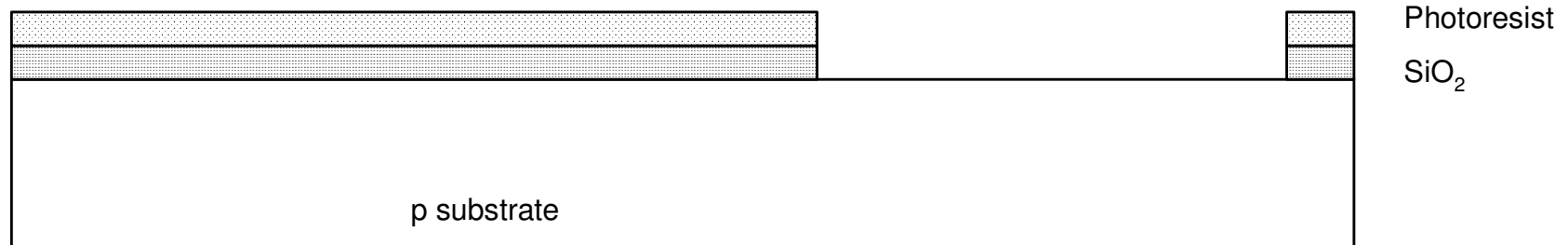
# Lithography

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



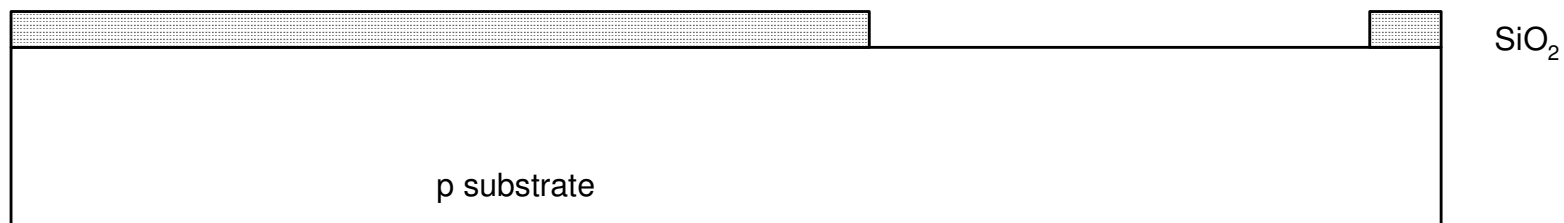
# Etch

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



# Strip Photoresist

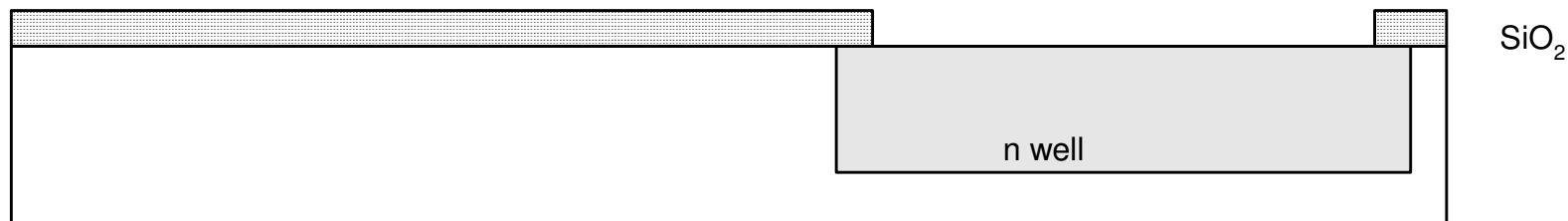
- Strip off remaining photoresist
  - Use mixture of acids called piranha etch
- Necessary so resist doesn't melt in next step





# 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 Implantation
  - Blast wafer with beam of As ions
  - Ions blocked by  $\text{SiO}_2$ , only enter exposed Si



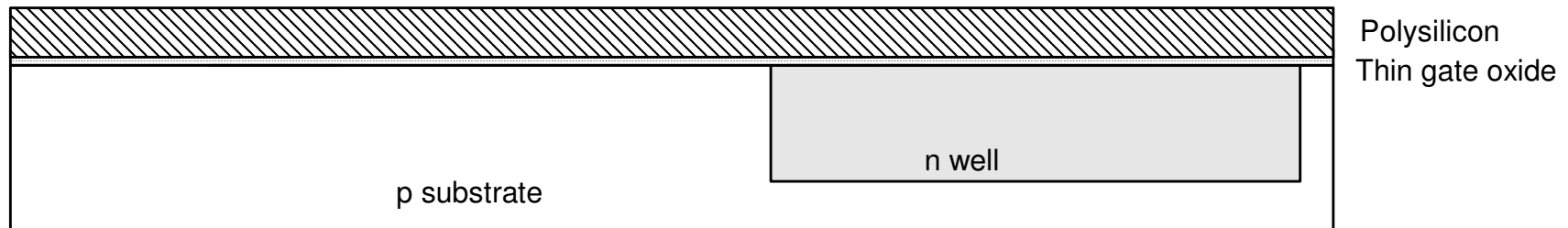
# Strip Oxide

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



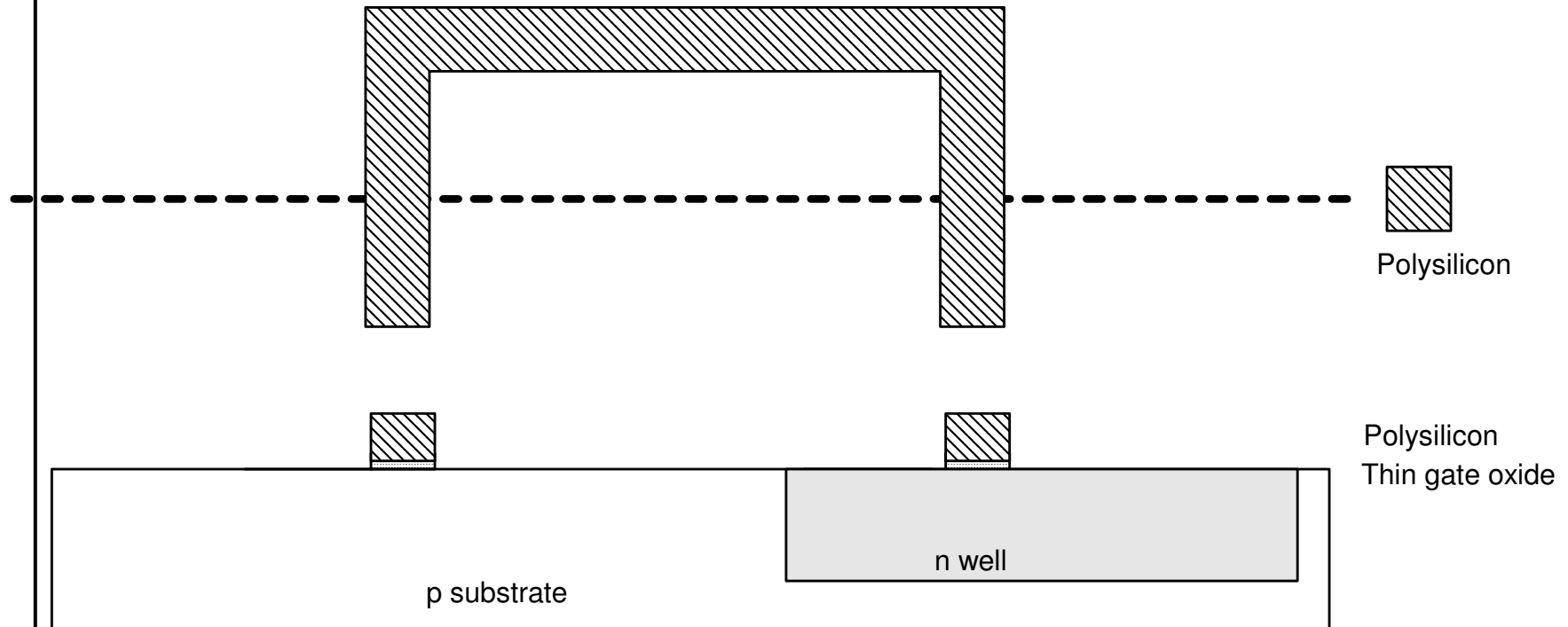
# Polysilicon

- Deposit very thin layer of gate oxide
  - $< 20 \text{ \AA}$  (6-7 atomic layers)
- Chemical Vapor Deposition (CVD) of silicon layer
  - Place wafer in furnace with Silane gas ( $\text{SiH}_4$ )
  - Forms many small crystals called polysilicon
  - Heavily doped to be good conductor



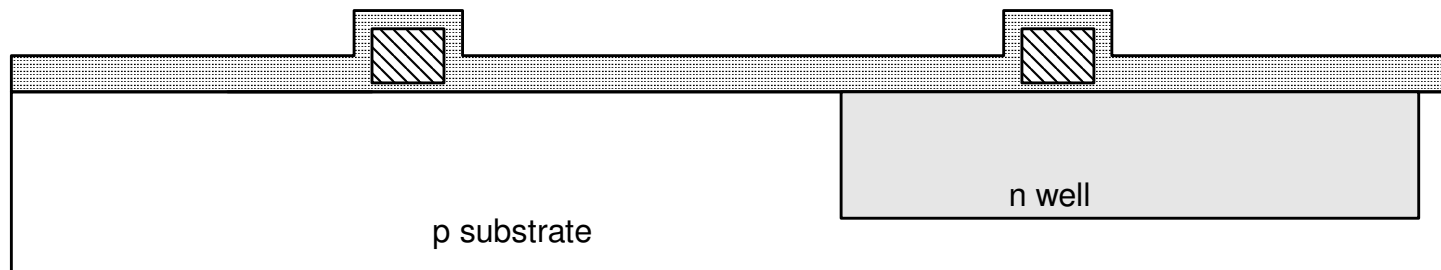
# Polysilicon Patterning

- Use same lithography process to pattern polysilicon



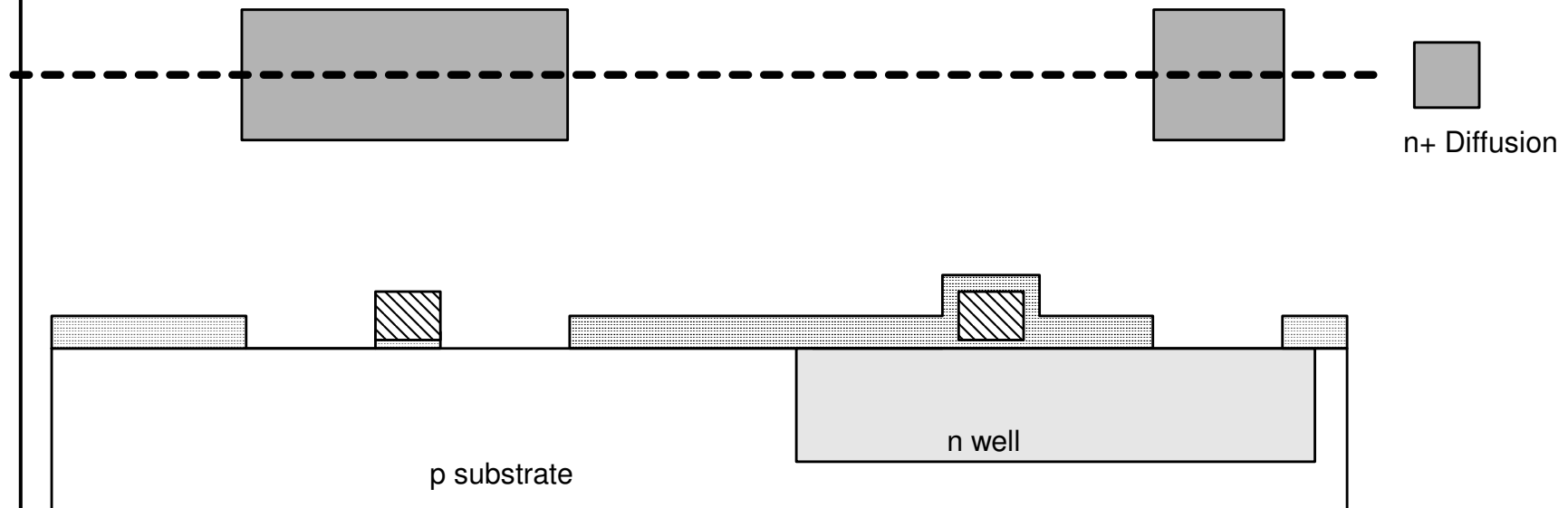
# Self-Aligned Process

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



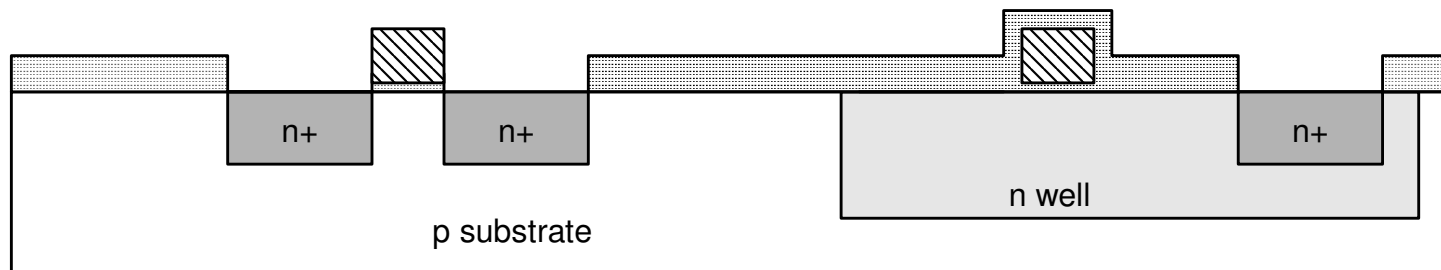
# N-diffusion

- 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



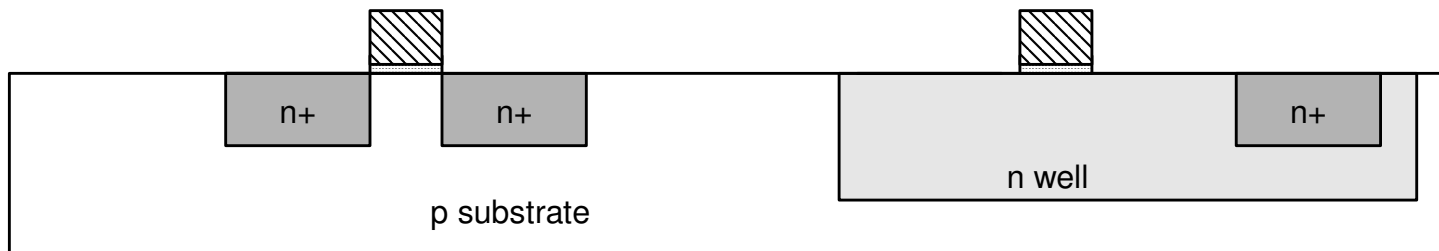
# N-diffusion cont.

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



# N-diffusion cont.

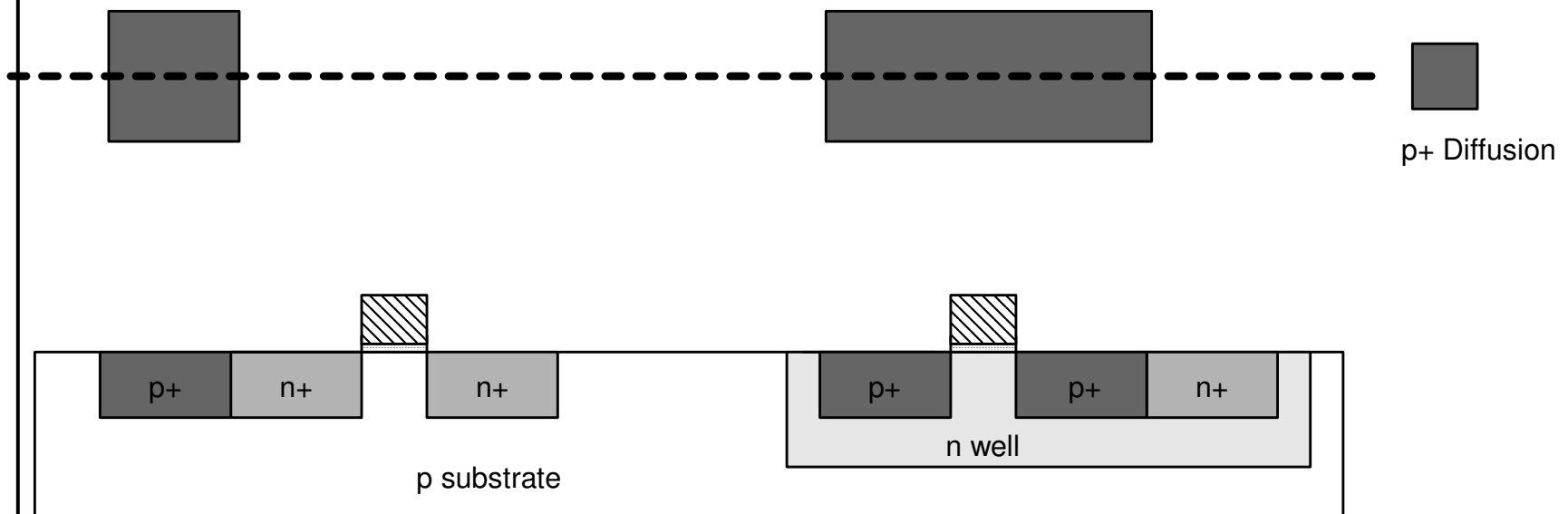
- Strip off oxide to complete patterning step





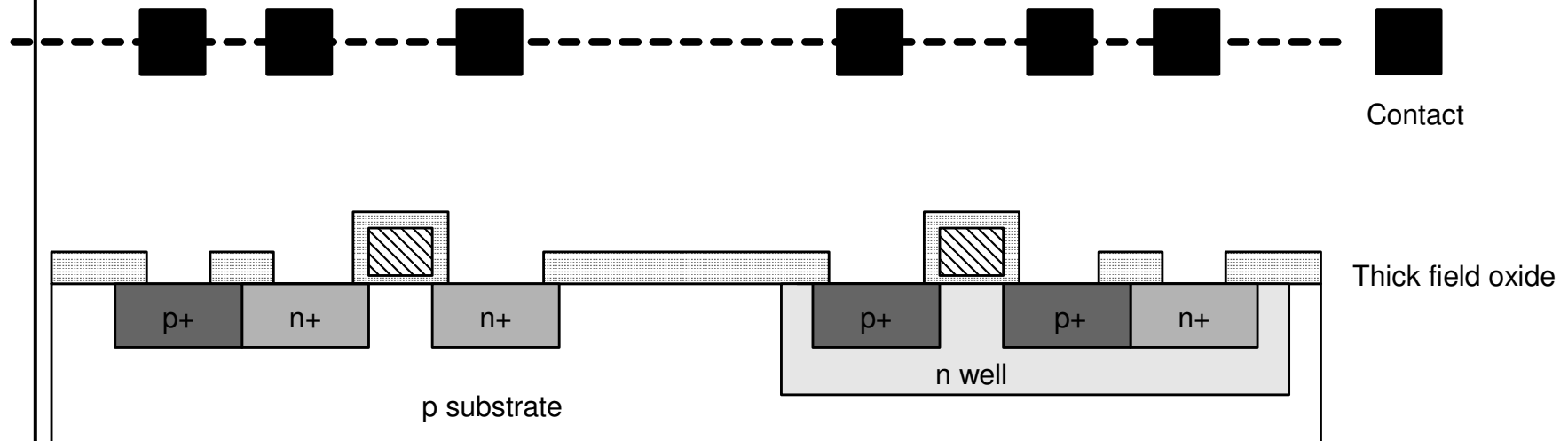
# P-Diffusion

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



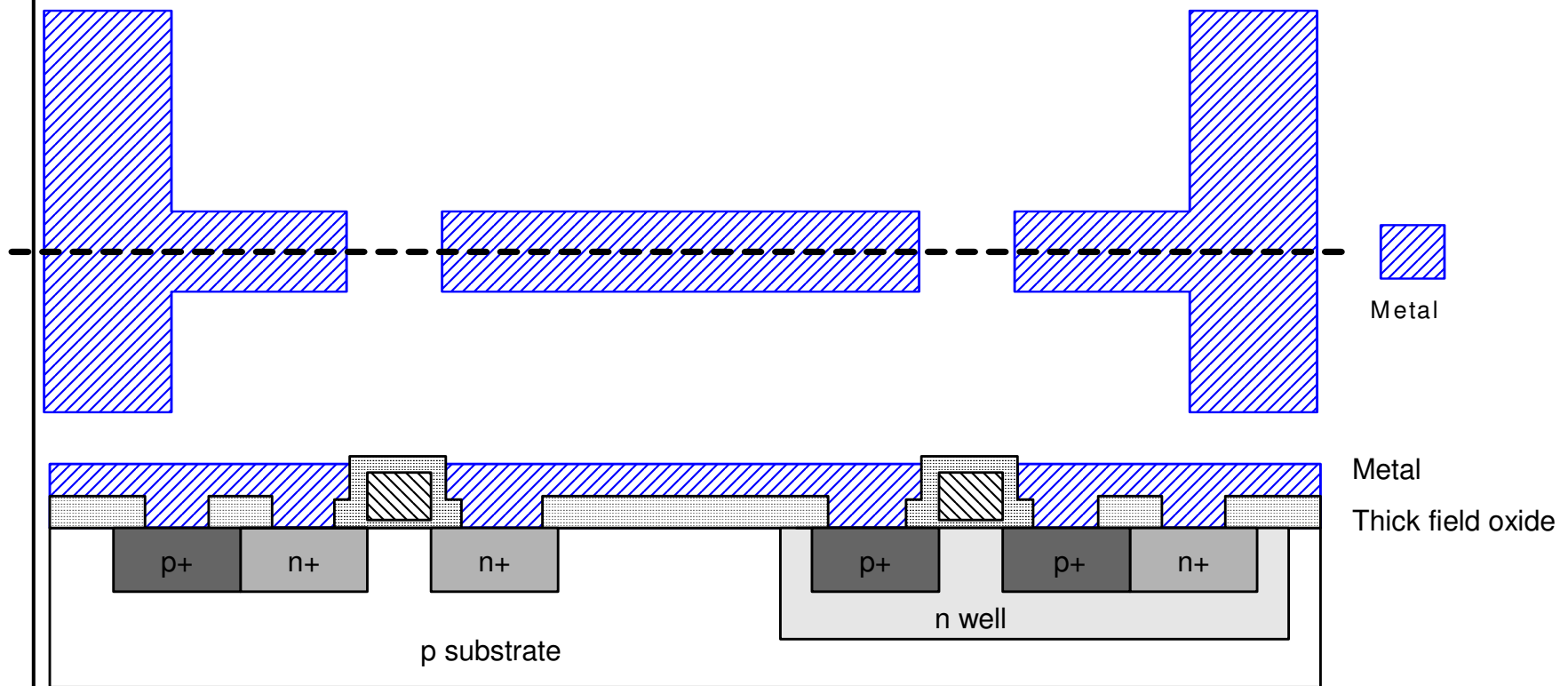
# Contacts

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



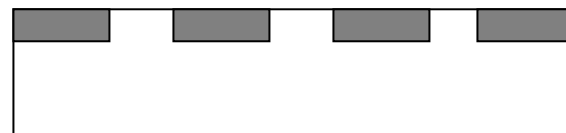
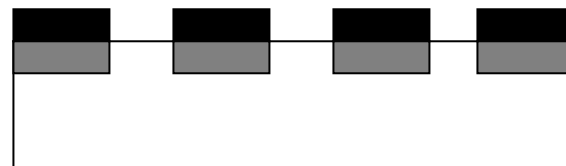
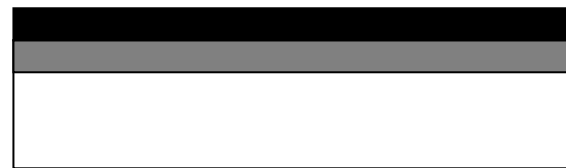
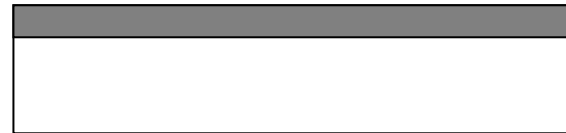
# Metalization

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



# Basic Processing Steps (Summary)

- Start with wafer at current step
- Add photoresist
- Pattern photoresist with mask
- Step-specific etch, implant, etc.
- Wash off resist



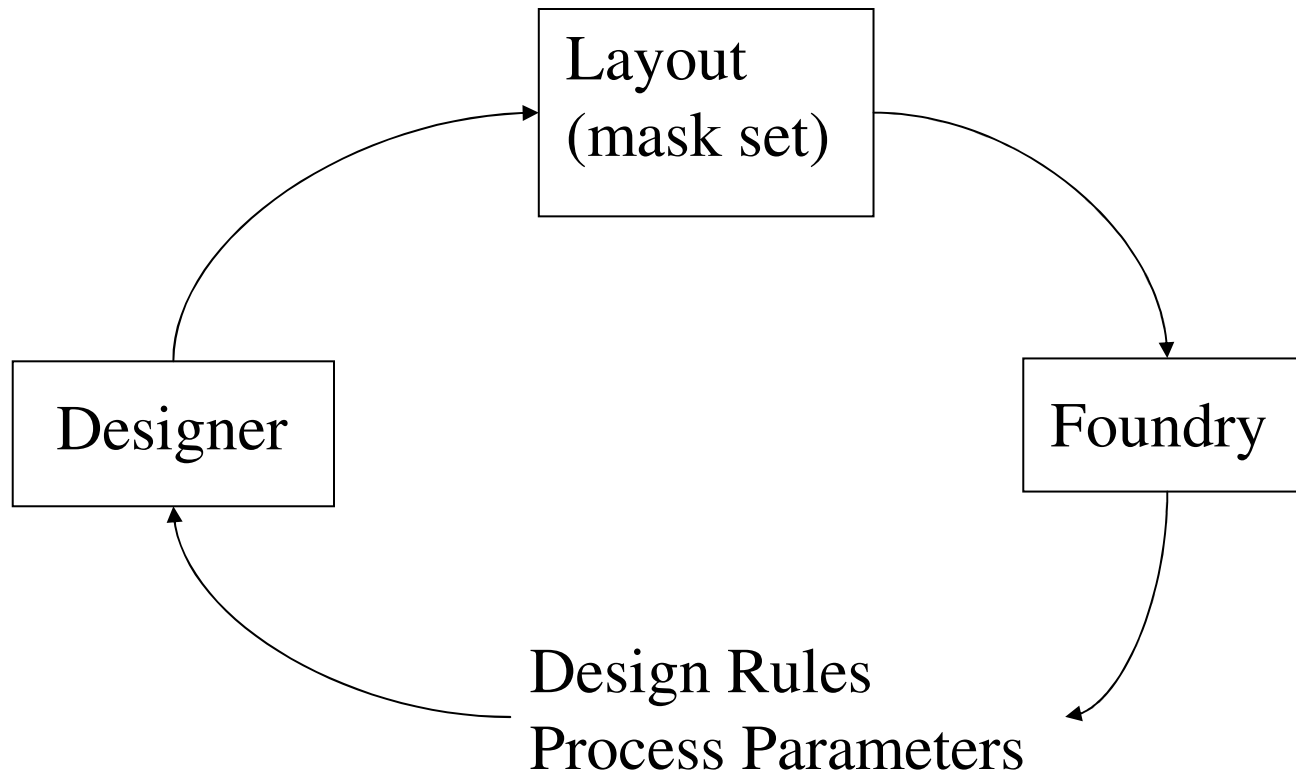
# Layout

- 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.  $\lambda = 0.3 \mu\text{m}$  in  $0.6 \mu\text{m}$  process

# Design Rules

- Design rules govern the layout of individual components: transistors, wires, contacts, vias
  - How small can the gates be, and how small can the wires be made?
- Conflicting Demands:
  - component packing: more functionality, higher speed
  - Chip yield: smaller sizes can reduce yield (fraction of good chips)
- Conservative vs aggressive design rules

# Foundry Interface



# Geometric Design Rules

- Resolution
  - Width and spacing of lines on one layer
- Alignment
  - make sure interacting layers overlap (or don't)
  - Contact surround
  - Poly overlap of diffusion
  - Well surround of diffusion



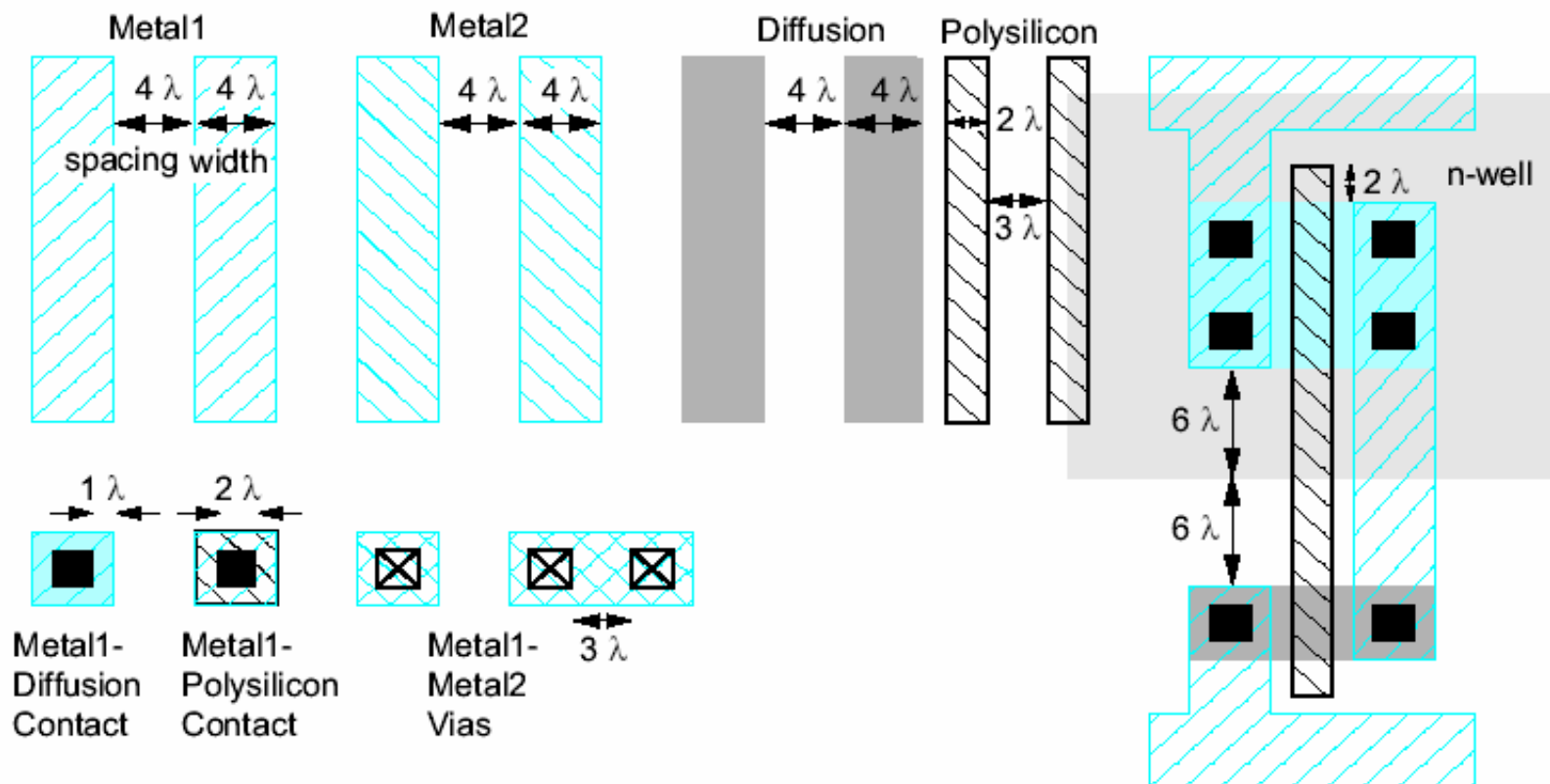
# SCMOS Design Rules

- Scalable CMOS design rules
- Feature size  $\lambda$  = half the drawn gate length (poly width)
- Mentor Graphics IC tool has built-in design rule checker (DRC)

Example design rules:

Layer	Minimum Width	Separation
Metal 1	$3 \lambda$	$3 \lambda$
Metal 2	$3 \lambda$	$4 \lambda$
Poly	$2 \lambda$	poly-poly: $2 \lambda$ poly-diff: $1 \lambda$

# Simplified Design Rules

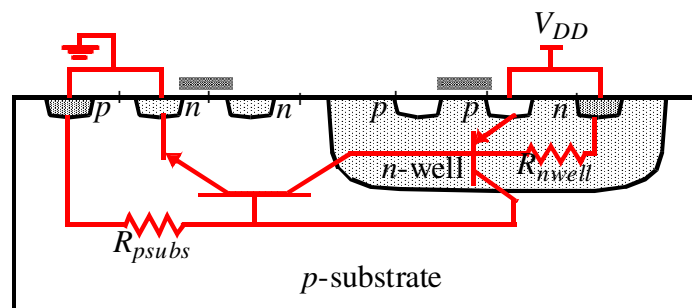


# Tub Ties and Latchup

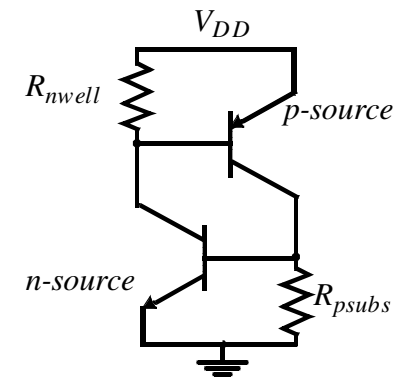
- Substrate must be connected to power supply
- p-tub for nMOS to  $V_{SS}$  (Gnd)
- N-tub for pMOS to  $V_{DD}$
- Connections made by special vias called tub ties
- Conservative design rule: place tub ties for every one or two transistors
- Why not place one tie in each tub that has 50 transistors?

# Latchup

- Too few ties: high resistance between tub and power supply, leads to parasitic bipolar transistors inhibiting normal chip operation
- Parasitic silicon-controlled rectifier (SCR)
- When both bipolar transistors are off, SCR conducts no current
- SCR turns on: high current short-circuit between  $V_{DD}$  and Gnd.



(a) Origin of latchup



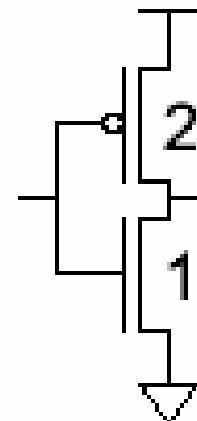
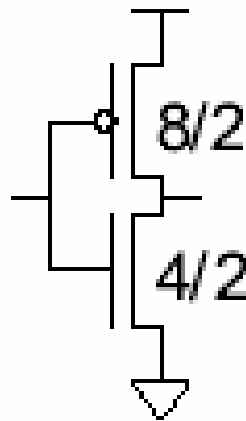
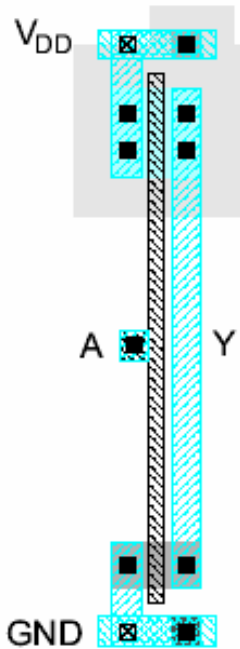
(b) Equivalent circuit

# Gate Layout

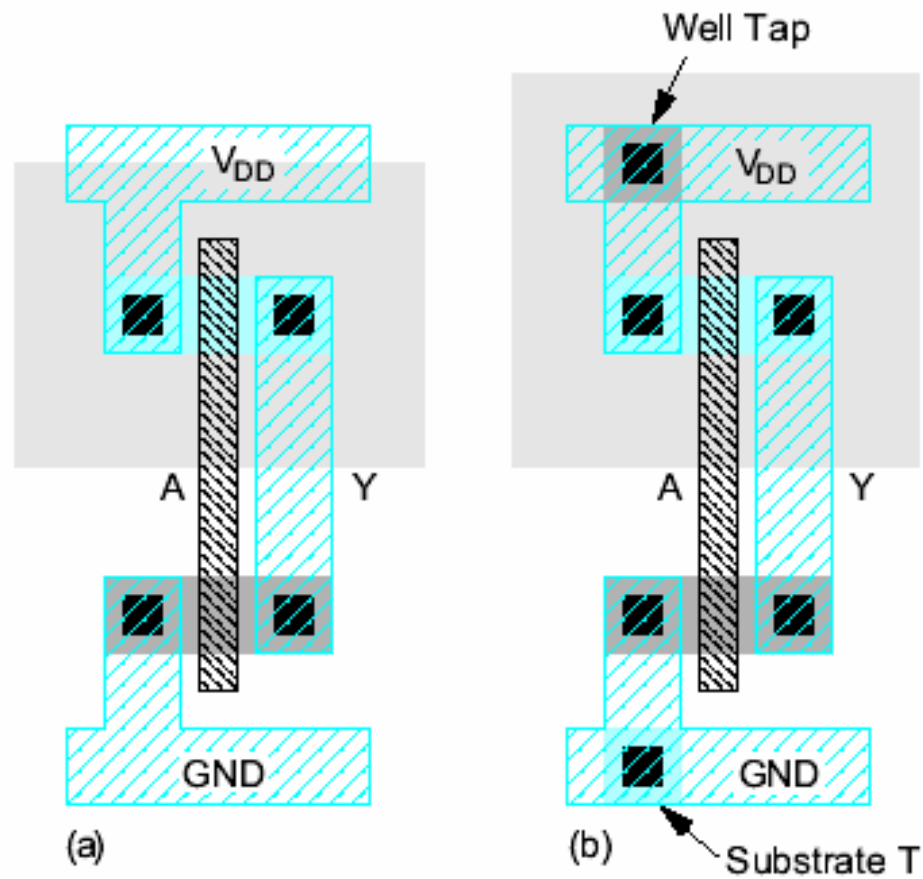
- Layout can be very time consuming
  - Design gates to fit together nicely
  - Build a library of standard cells
- Standard cell design methodology
  - $V_{DD}$  and GND should abut (standard height)
  - Adjacent gates should satisfy design rules
  - nMOS at bottom and pMOS at top
  - All gates include well and substrate contacts

# Inverter Layout

- Transistor dimensions specified as Width / Length
  - Minimum size is  $4\lambda / 2\lambda$ , sometimes called 1 unit
  - In  $f = 0.6 \mu\text{m}$  process, this is  $1.2 \mu\text{m}$  wide,  $0.6 \mu\text{m}$  long

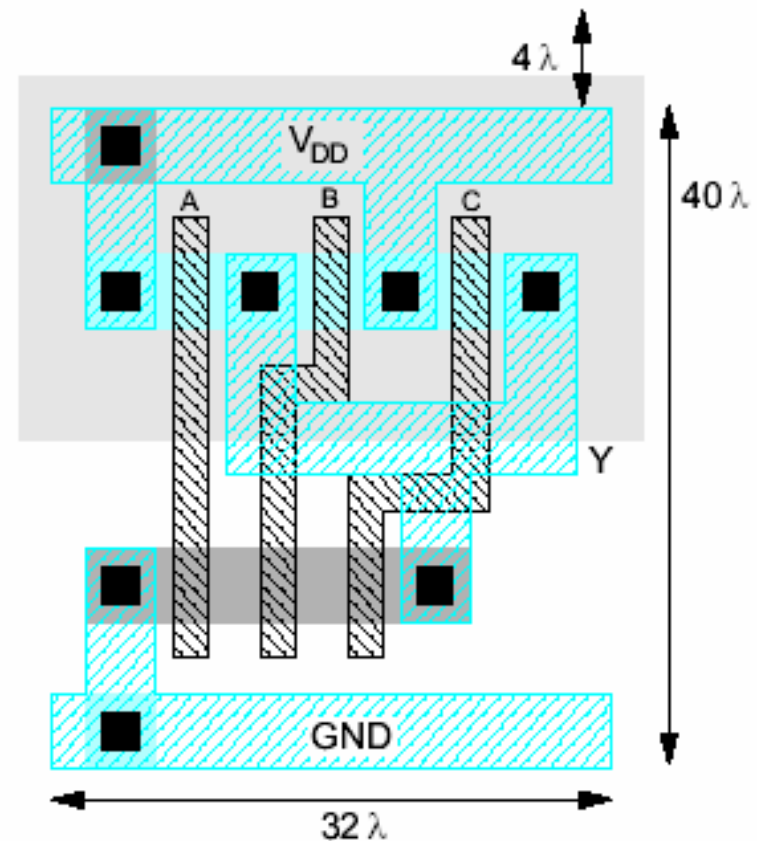


# Example: Inverter



# Example: NAND3

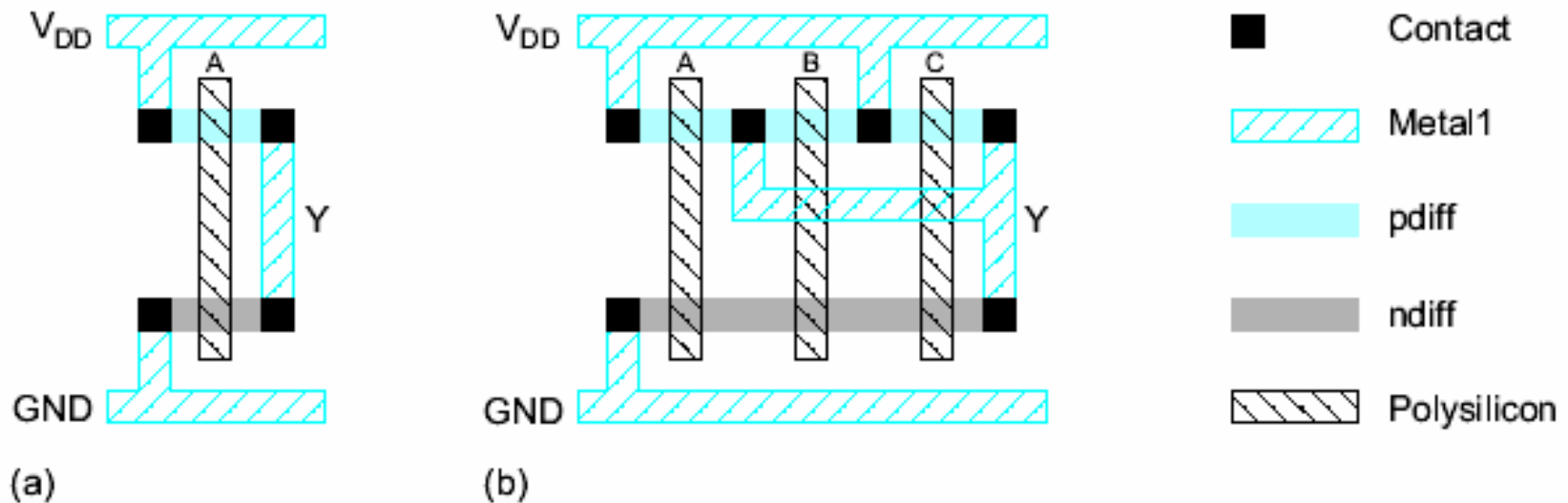
- Horizontal N-diffusion and p-diffusion strips
- Vertical polysilicon gates
- Metal1  $V_{DD}$  rail at top
- Metal1 GND rail at bottom
- $32 \lambda$  by  $40 \lambda$





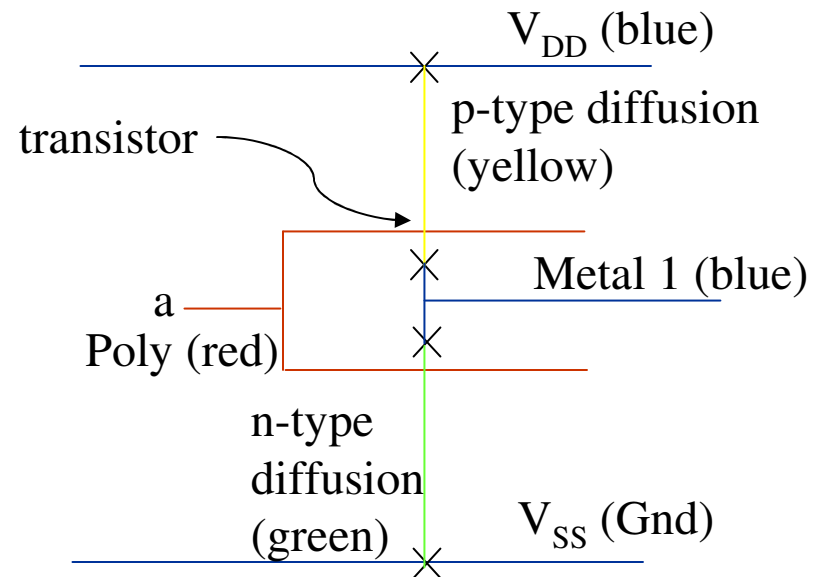
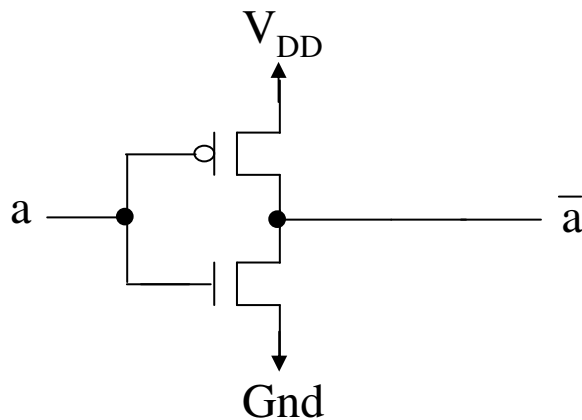
# Stick Diagrams

- *Stick diagrams* help plan layout quickly
  - Need not be to scale

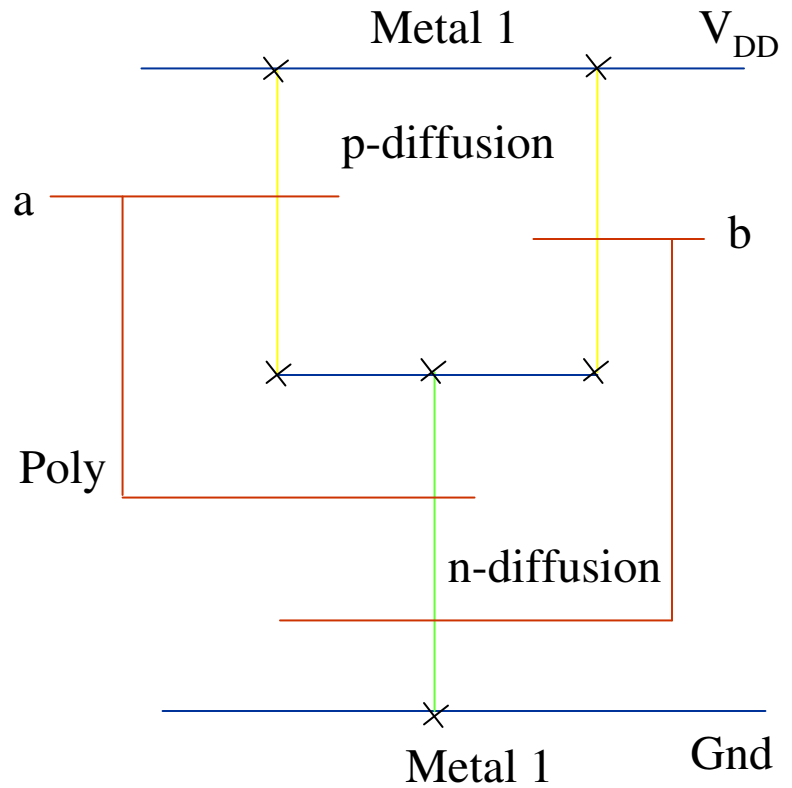
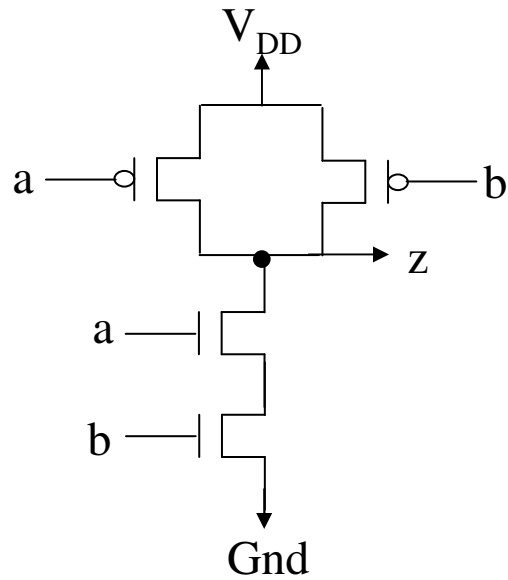


# Stick Diagrams

- Designing complete layout in terms of rectangles can be overwhelming
- Stick diagram: abstraction between transistor schematic and layout
  - Cartoon of a chip layout
- Replace rectangles by lines

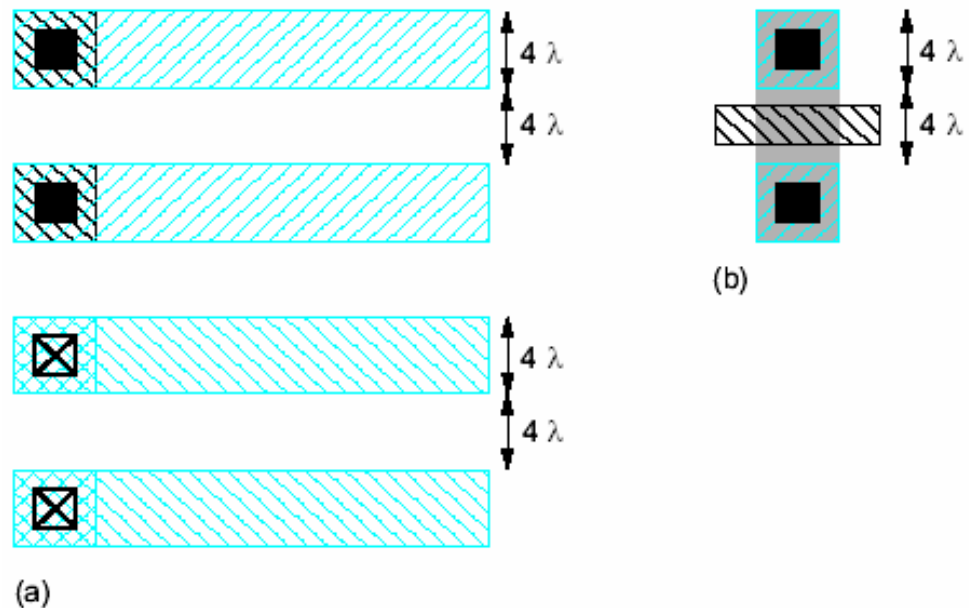


# Stick Diagram



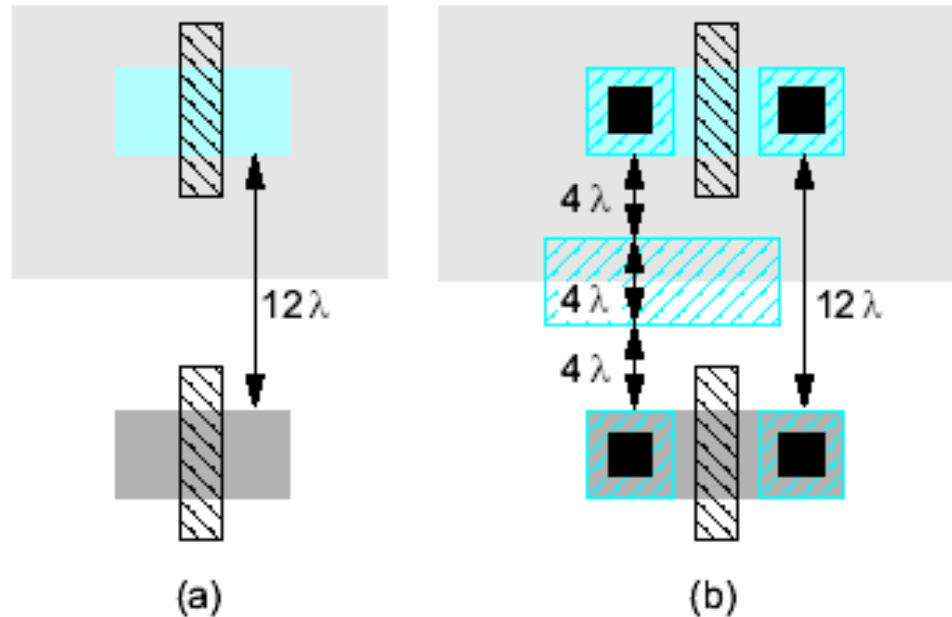
# Wiring Tracks

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



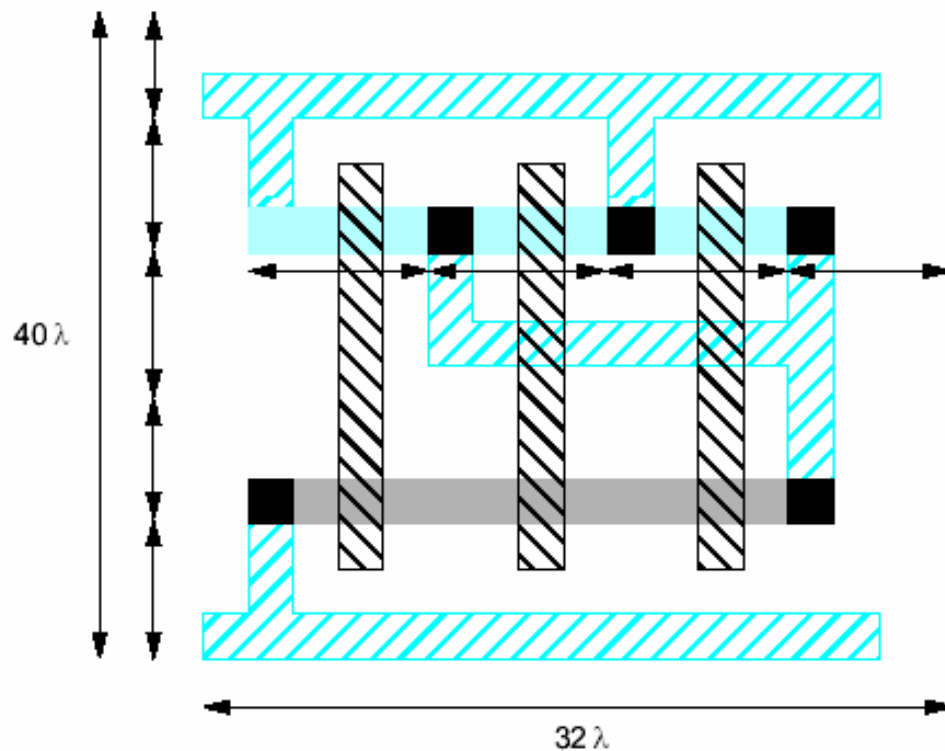
# Well spacing

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



# Area Estimation

- Estimate area by counting wiring tracks
  - Multiply by 8 to express in  $\lambda$



# Example: O3AI

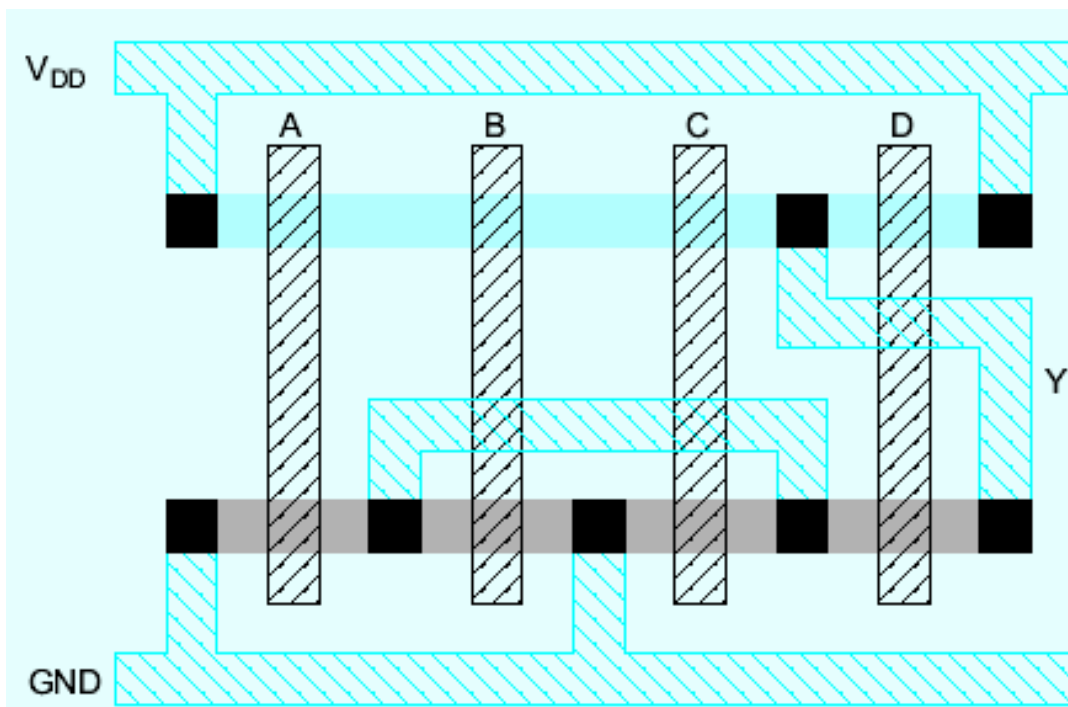
- Sketch a stick diagram for O3AI and estimate area

$$- \quad Y = \overline{(A + B + C)} \square D$$

# Example: O3AI

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

- Sketch a stick diagram for O3AI and estimate area

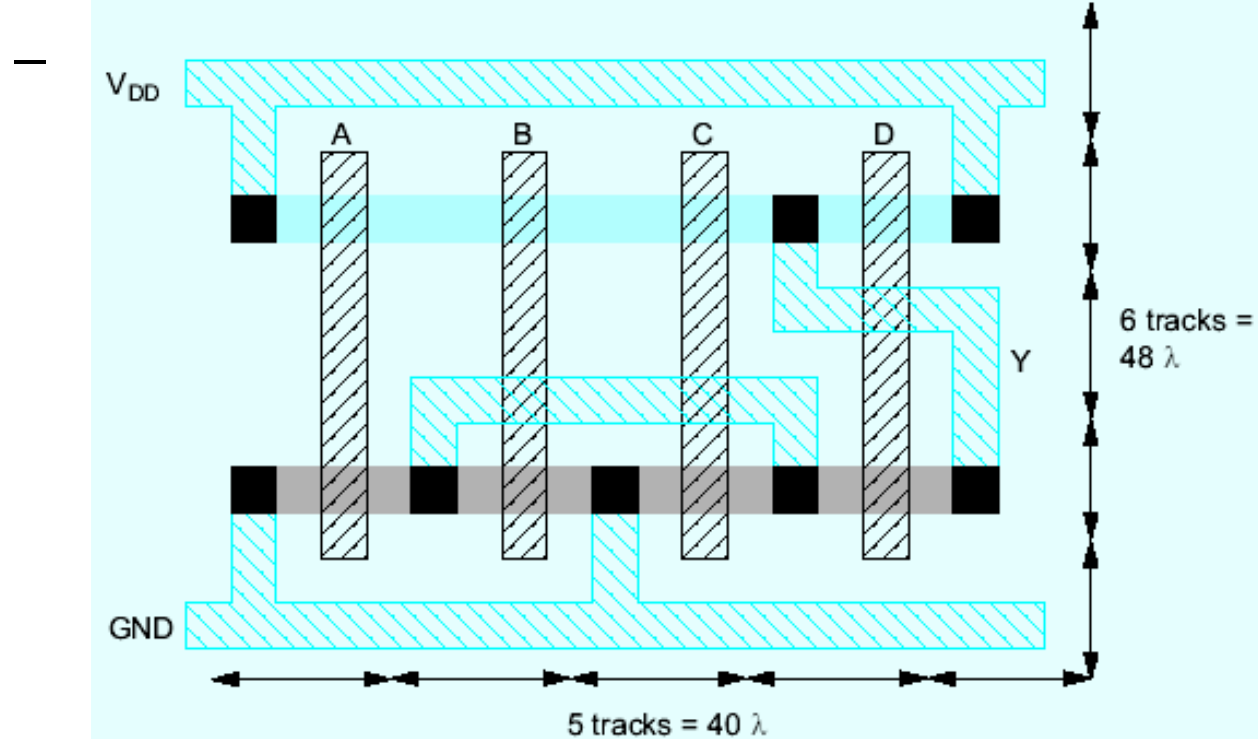




# Example: O3AI

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

- Sketch a stick diagram for O3AI and estimate area



# Some Layout Hints

- Plan the global structure (“big picture”), then design cells
  - Floorplan
  - Wiring strategy
  - Power and ground distribution
  - Systematic placement
  - Keep all pMOS/nMOS together
  - Place transistors in rows: share source/drain diffusion
- Wiring on orthogonal metal layers
  - Assign preferred directions to M1 and M2
  - Use diffusion only for devices, not for interconnect
  - Use poly only for very local interconnect

# Cell Minimization

- Chip area (cell size) must be minimized carefully

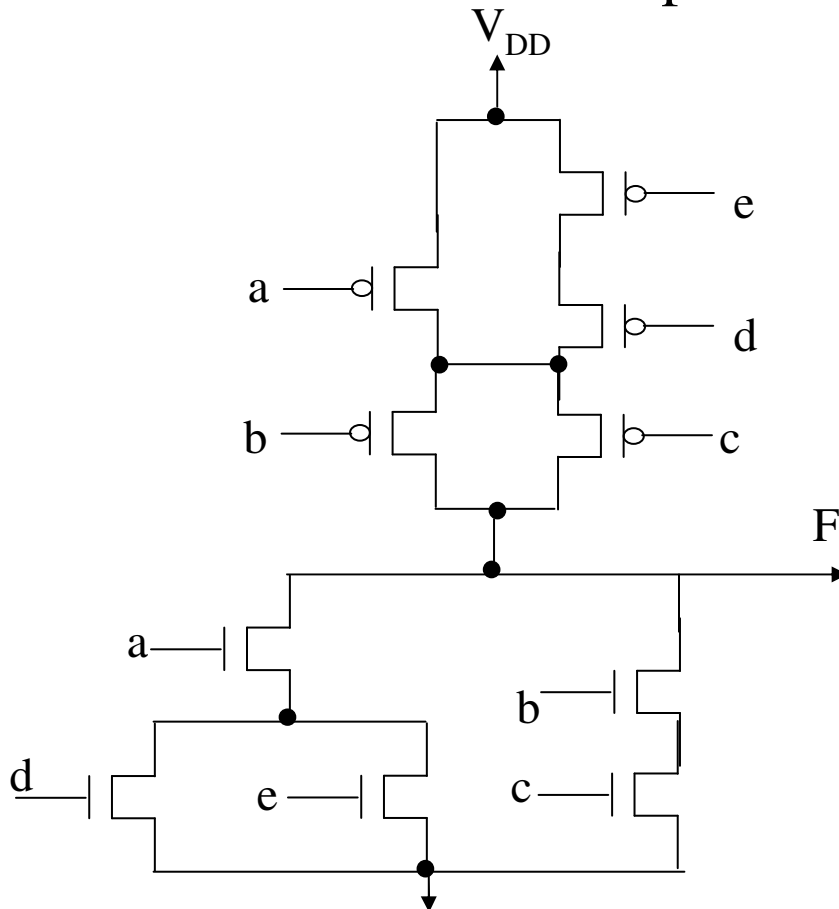
Impact of die size/chip area on cost (unpackaged dies)

	Nominal Pentium die	1% increase in die size	15% increase in die size
Wafer cost	\$1,460	\$1,460	\$1,460
Die size	160.2 mm <sup>2</sup>	161.8 mm <sup>2</sup>	184.2 mm <sup>2</sup>
Die cost	\$84.06	\$85.33	\$102.55
Chips fabricated per week	498.1 K	482.9 K	337.5 K
Added annual cost	—	\$63.5 M	\$961 M

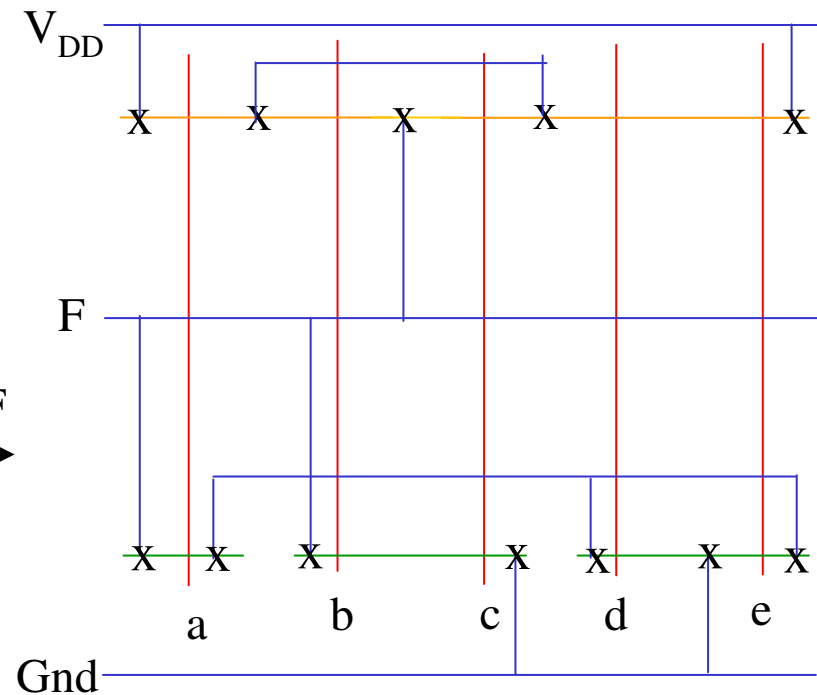
**1% increase in die size leads to 3% decrease in stock price for Intel!**

# Minimize number of diffusion strips

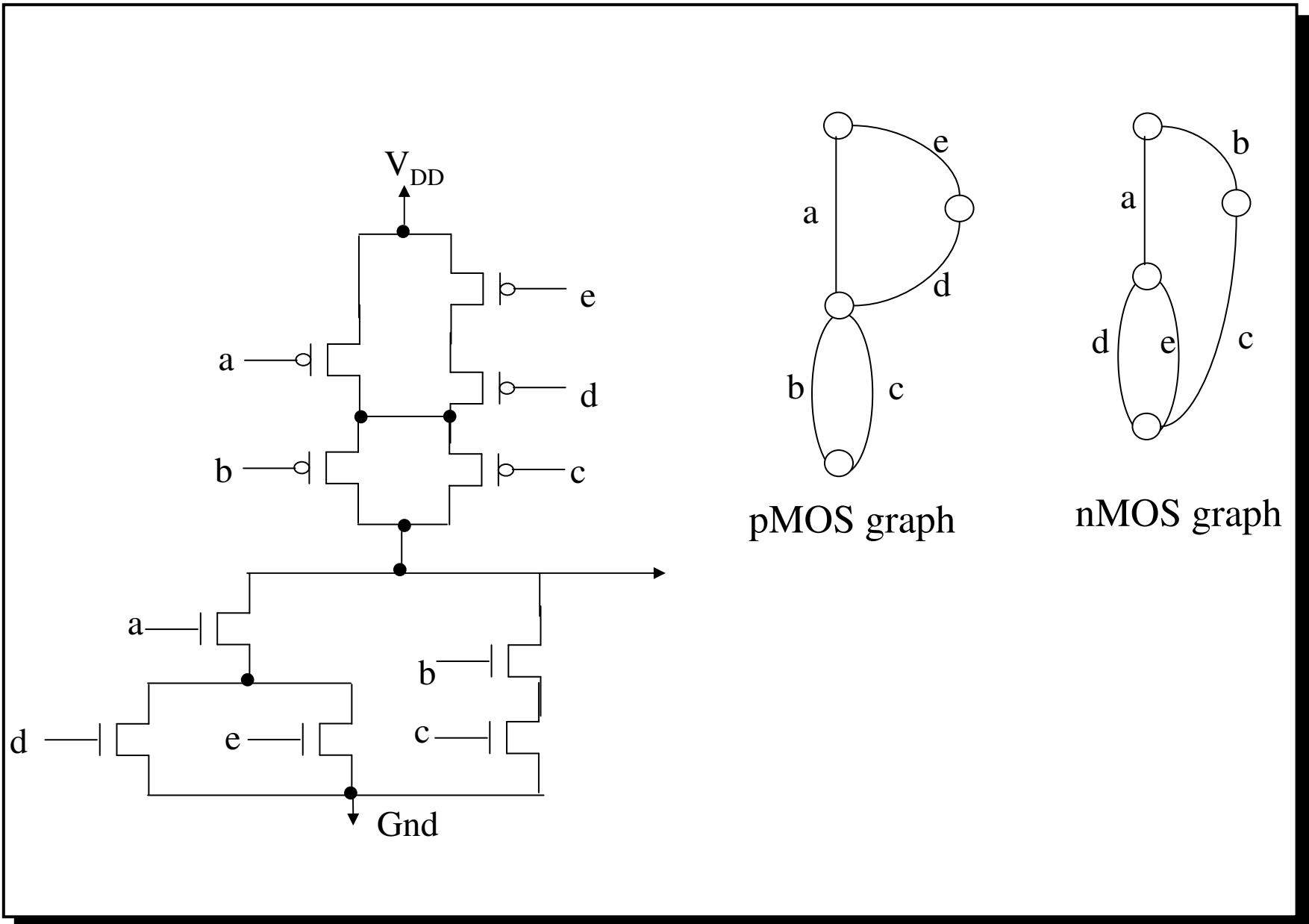
- How do we order the gate inputs (poly)?
- More diffusion strips  $\Rightarrow$  more spacing, more area



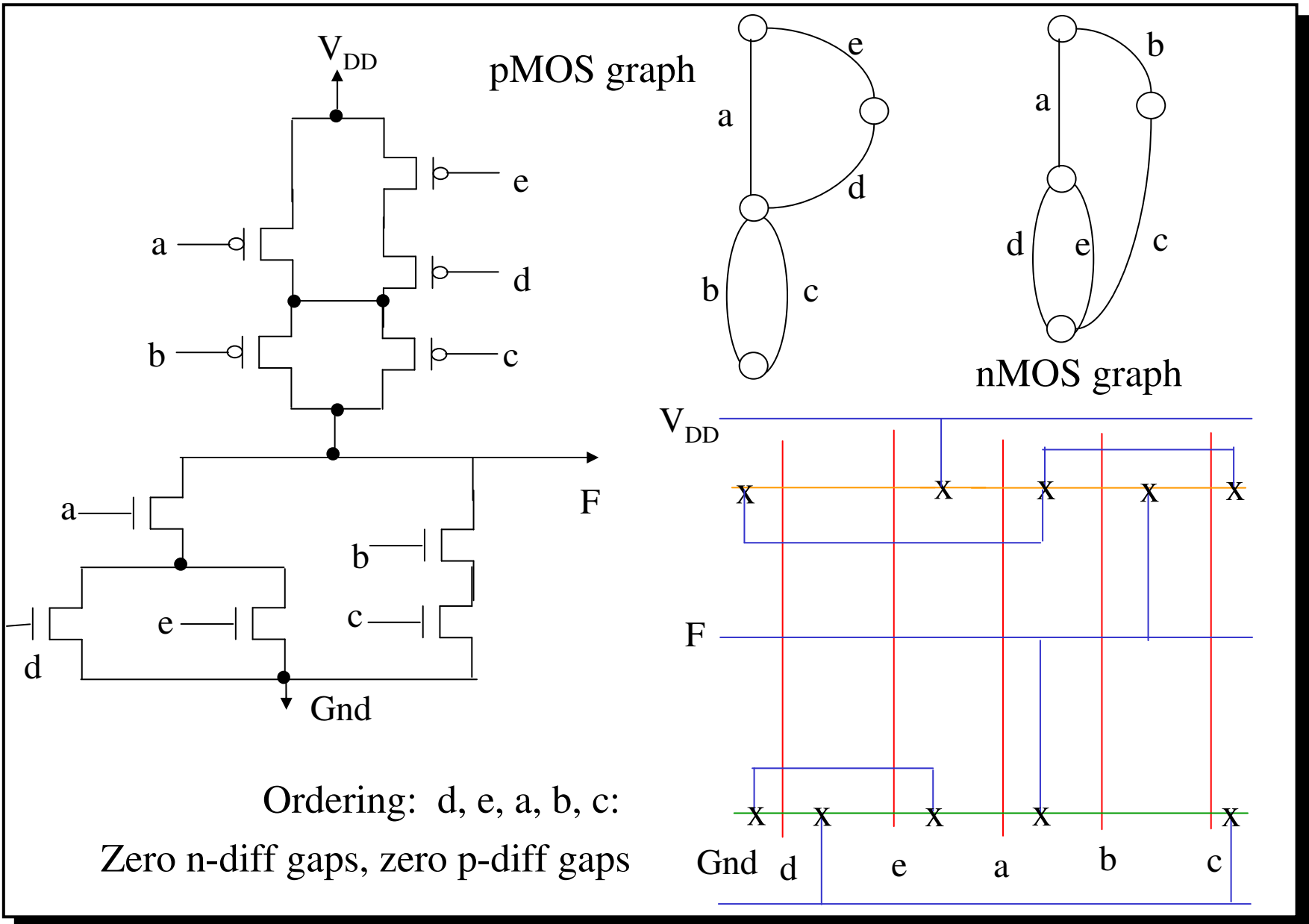
Try a, b, c, d, e:



Two n-diff gaps, zero p-diff gaps



- Euler path: Visit every edge *exactly* once
  - Find all Euler paths for nMOS and pMOS graphs
  - Find p- and n-path that have identical labeling
- 
- For example: d, e, a, b, c
  - If no such path exists, then break diffusion into strips



# Summary

- MOS Transistors are stack of gate, oxide, silicon
- 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!