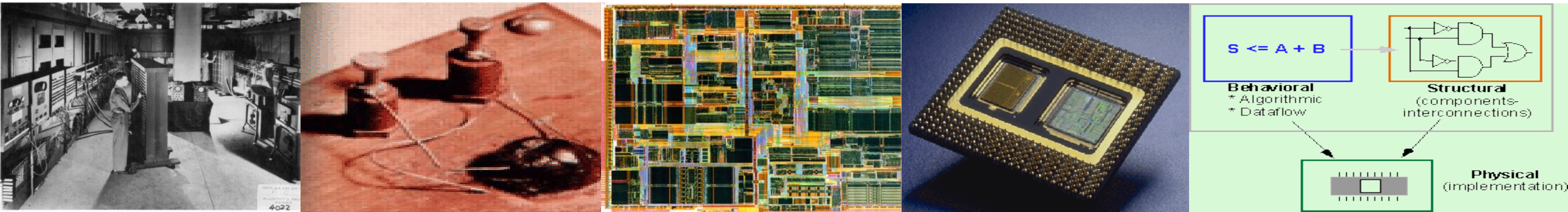
Lecture 6
CMOS Fabrication and Layout

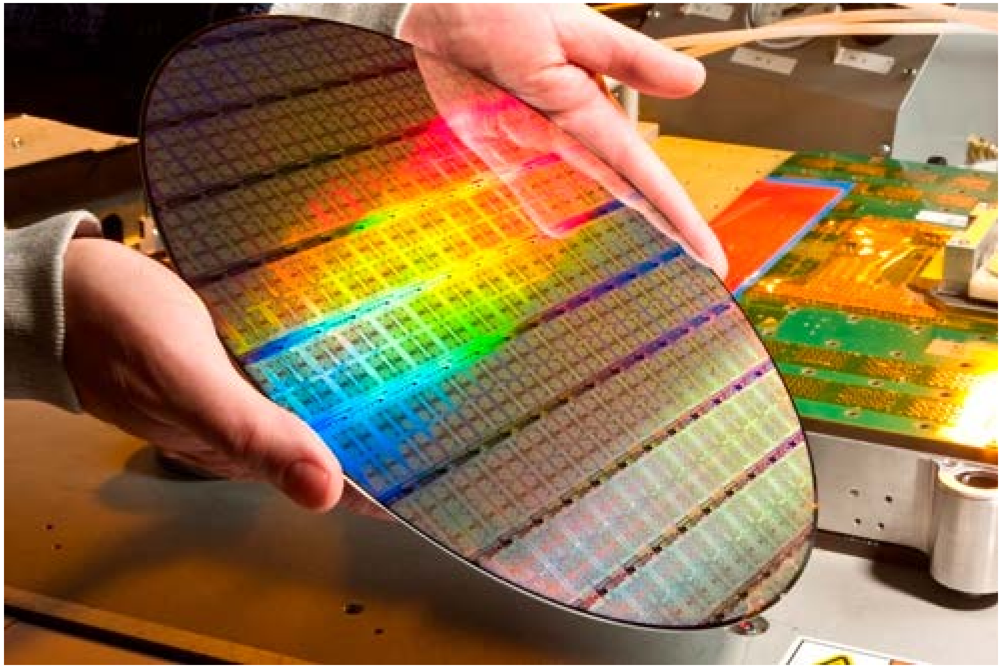
Bryan Ackland
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07030

Adapted from Lecture Notes, David Mahoney Harris CMOS VLSI Design



CMOS Wafers

- CMOS transistors are fabricated on silicon wafer
  - mechanical support
  - electrical ground plane
  - epitaxial layer: “single crystal” substrate (< 0.2 defects/cm^2)



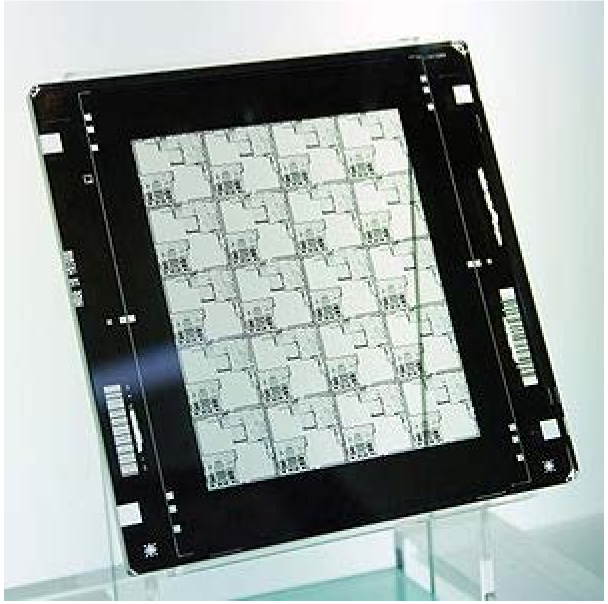
Courtesy IBM

CMOS Fabrication

- Lithography process similar to printing press
  - glass masks and UV light

- On each step, different materials are deposited or etched according to one of these masks

- As process line width shrinks:
  - smaller transistors & wires
  - faster transistors
  - lower power transistors

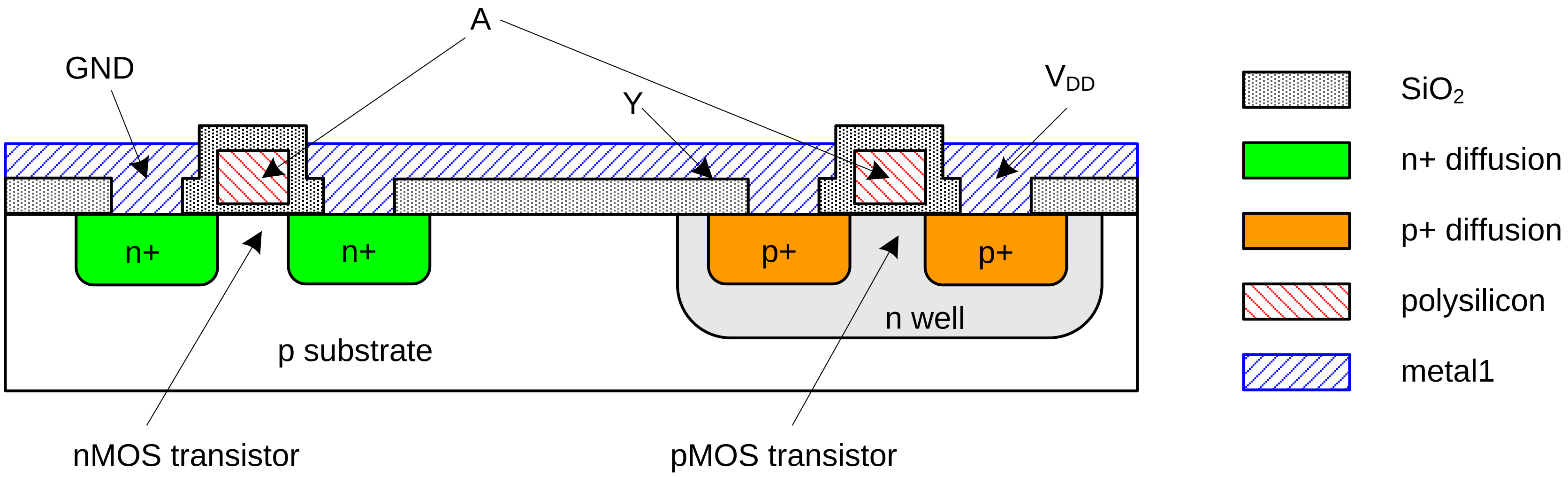


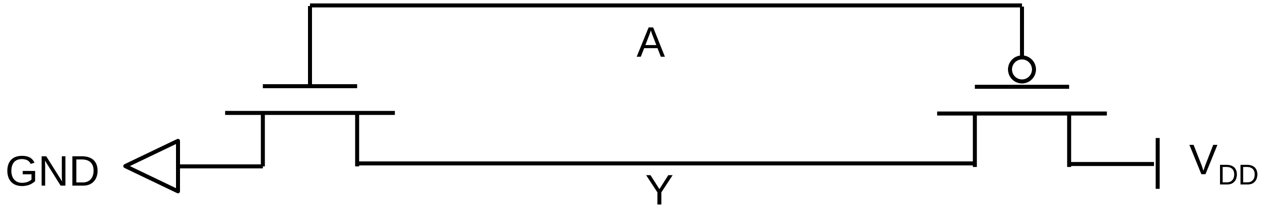
Wikipedia

- Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process

CMOS Fabrication

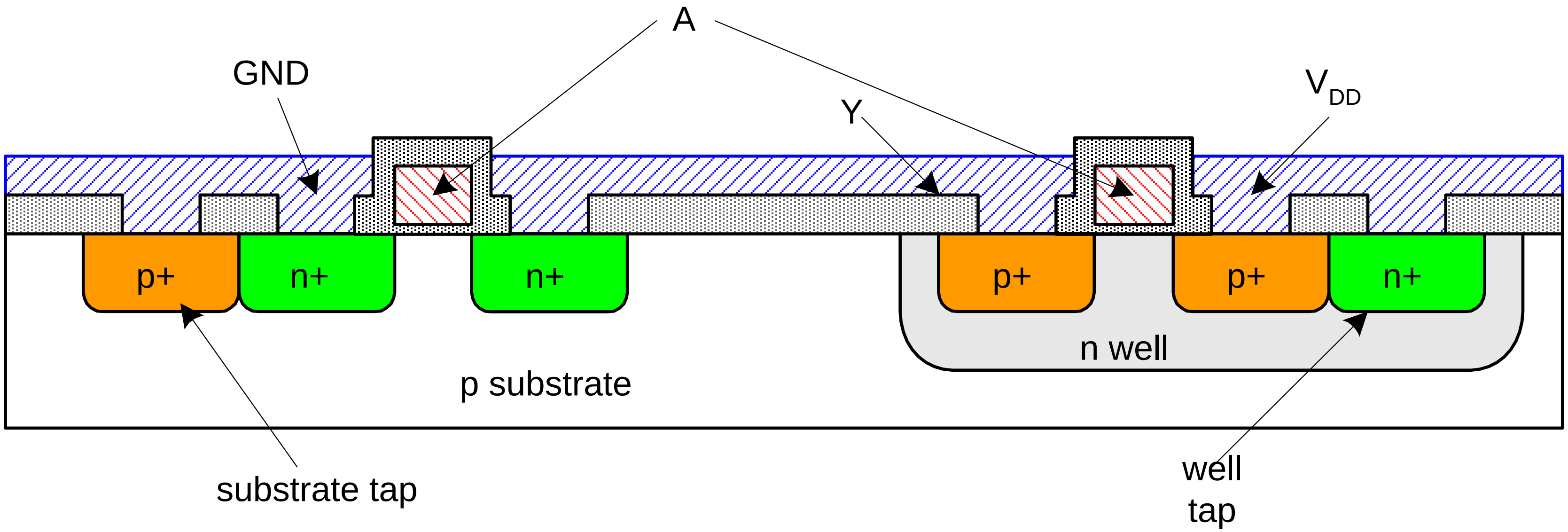
- Typically use p-type substrate for nMOS transistors
- Requires n-well for body of pMOS transistors

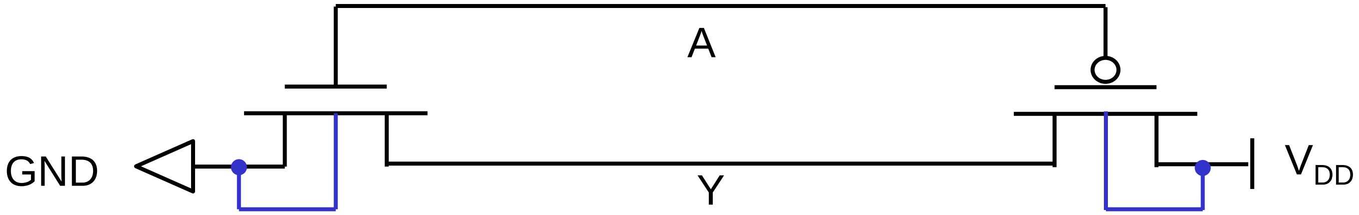




Well and Substrate Taps

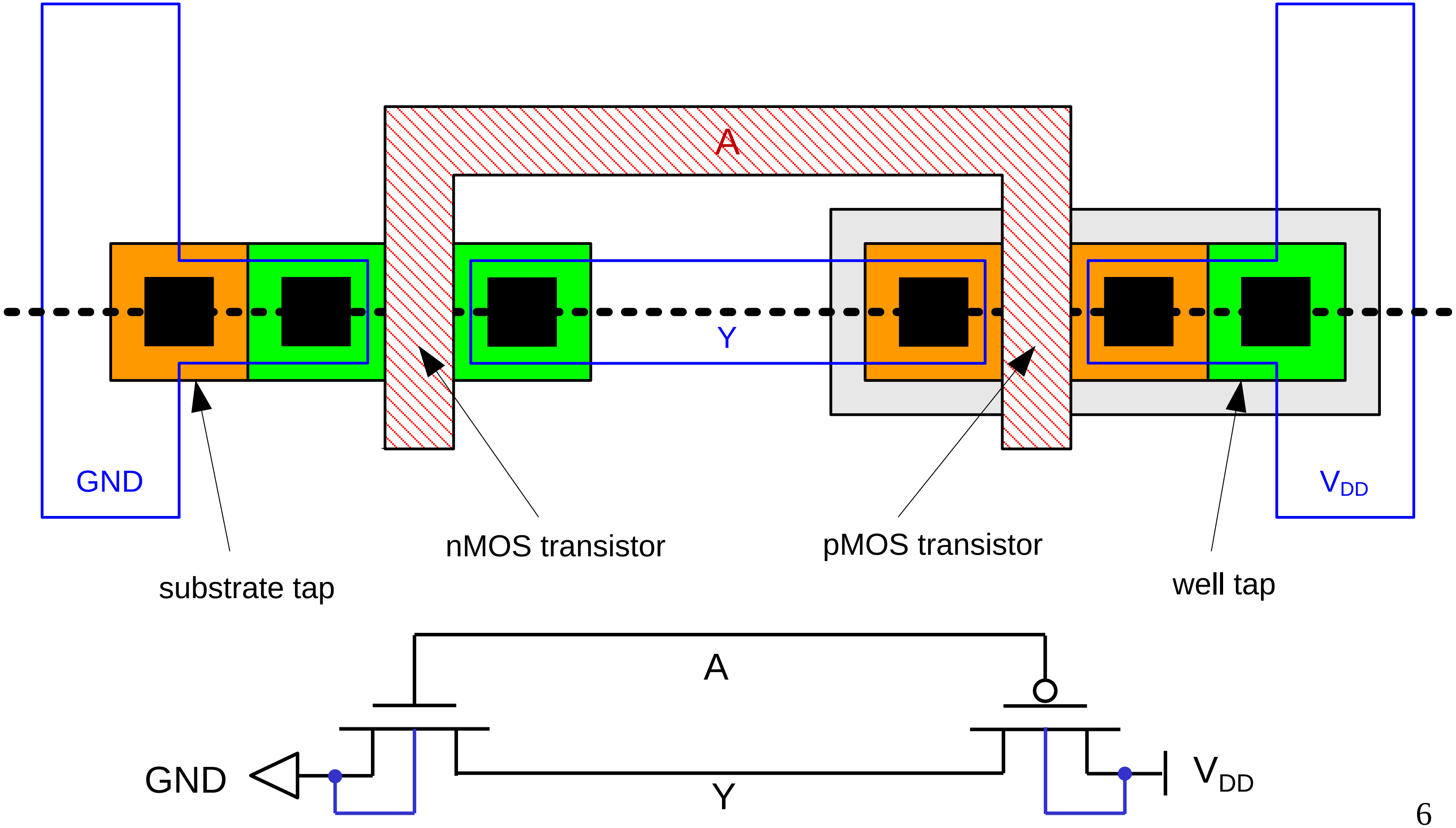
- Substrate must be tied to GND and n-well to VDD
- Metal to lightly-doped semiconductor forms poor connection called Shottky Diode
- Use heavily doped well & substrate contacts / taps / ties





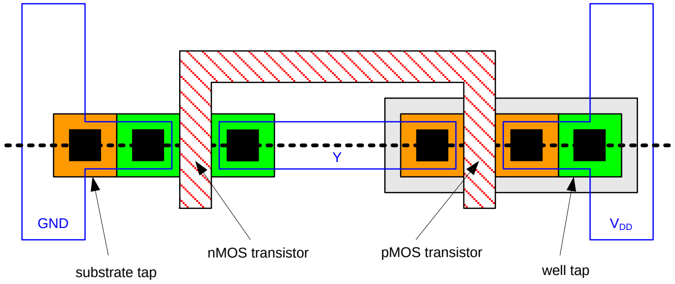
Inverter Layout

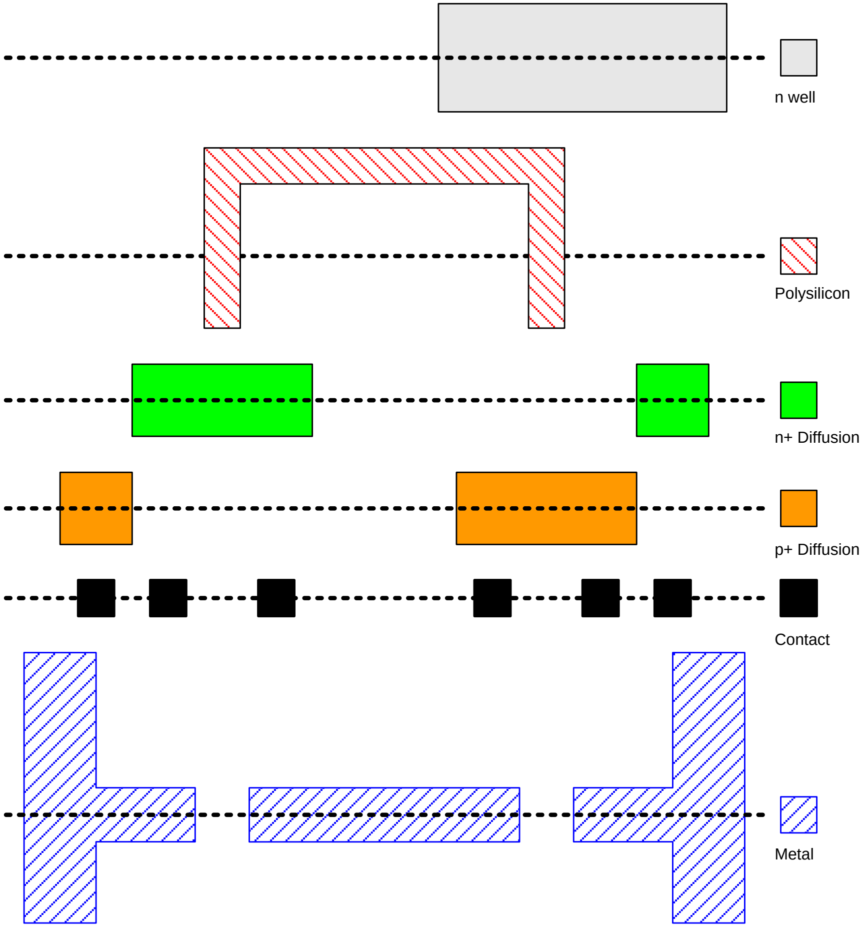
- Transistors and wires are defined by masks
- Cross-section taken along dashed line



Detailed Mask Views

- 6 masks
  - n-well
  - polysilicon
  - n+ diffusion
  - p+ diffusion
  - contact
  - metal

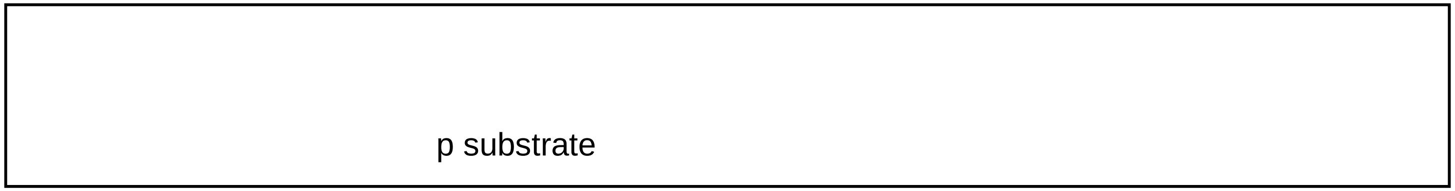




Fabrication Steps

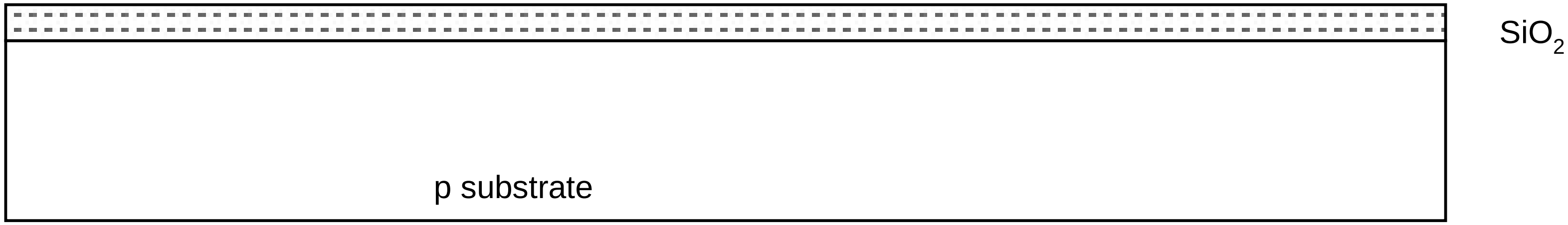
- Start with blank wafer
- Build inverter from the bottom up
- First step will be to form the n-well:

- 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



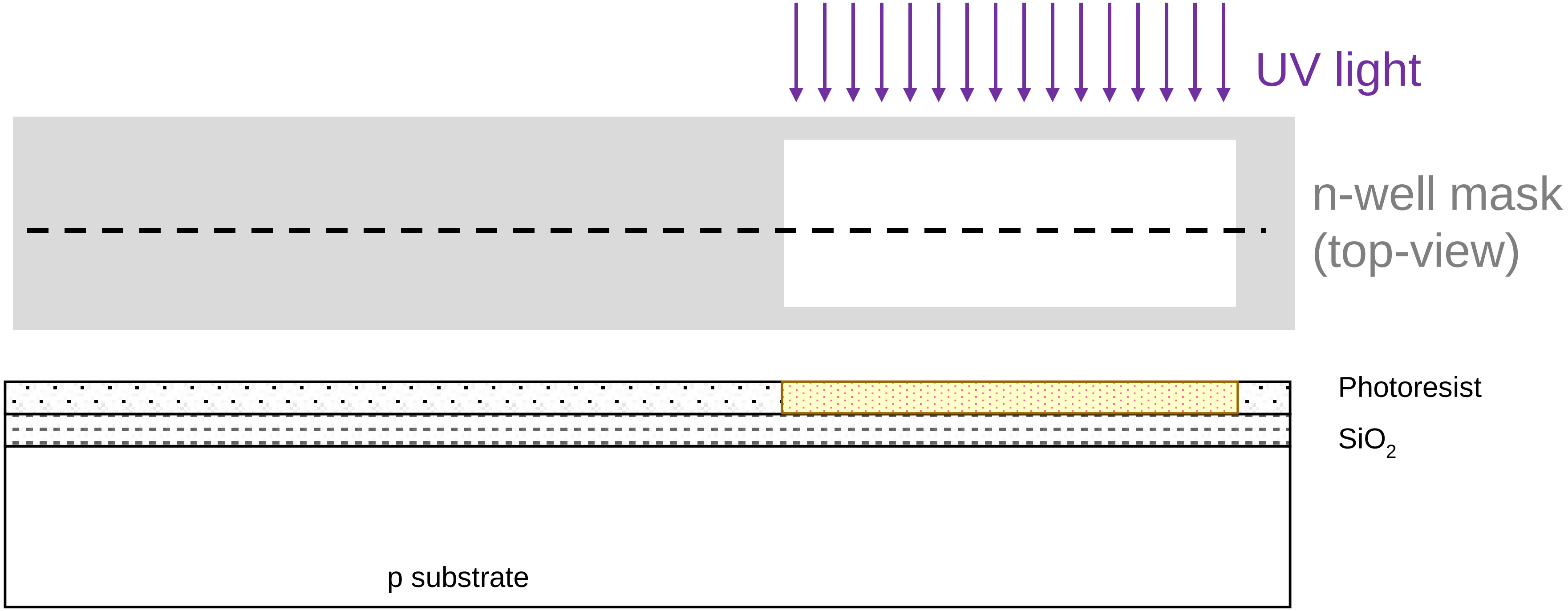
Oxidation

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



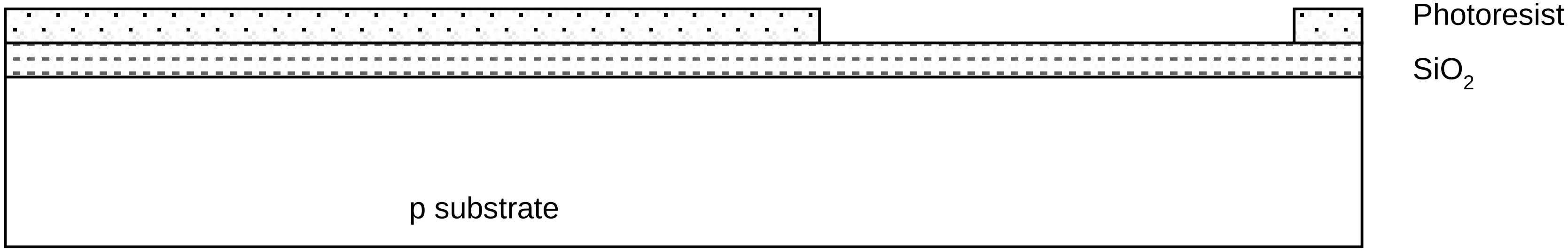
Photoresist

- Spin on photoresist
- Photoresist is a light-sensitive organic polymer
- Softens where exposed to UV light
- Expose photoresist through n-well mask

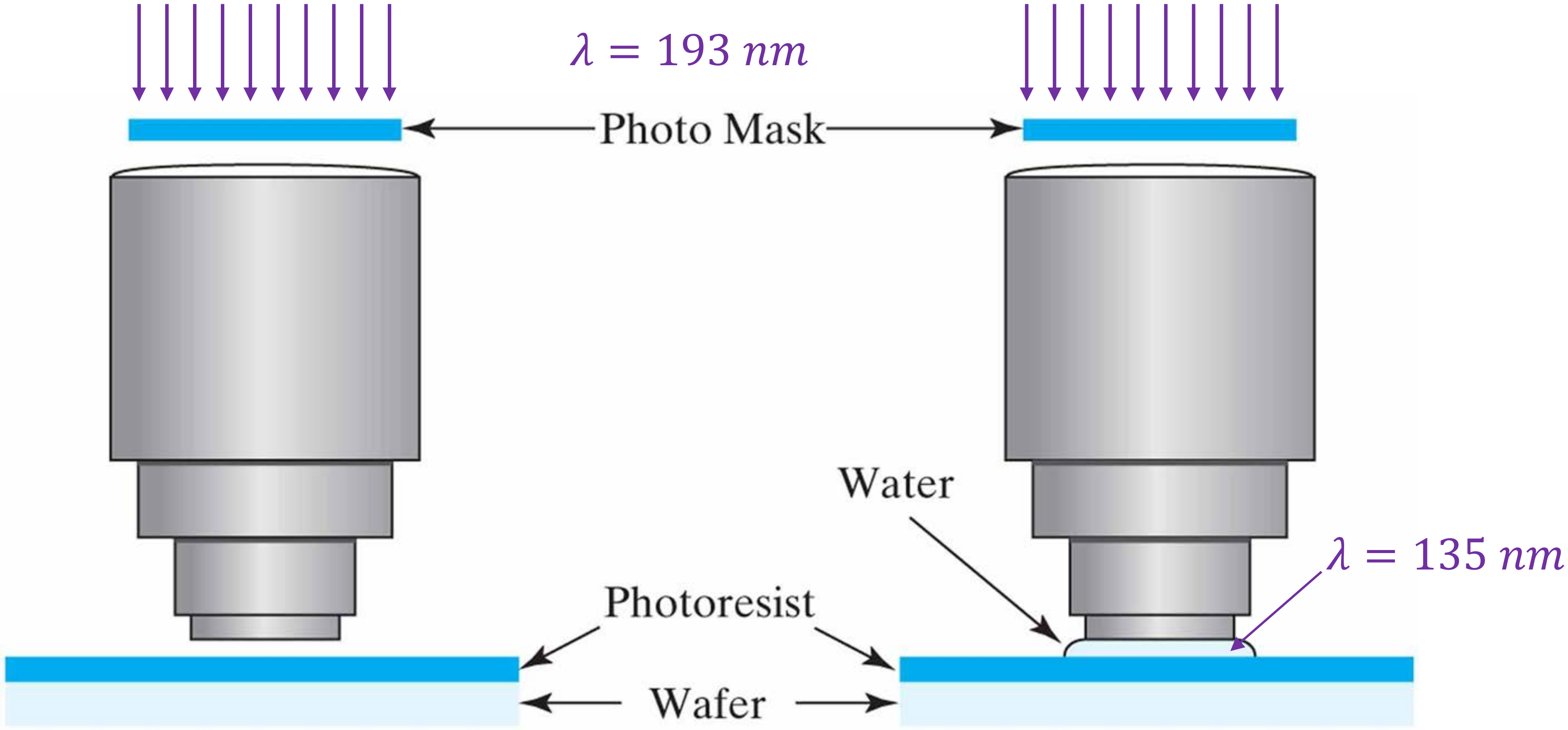


Lithography

- Strip off exposed photoresist with developer
  - organic solvent
- Leaves exposed SiO_2 in pattern determined by n-well mask
- How do we make 65 nm patterns with UV-light where $\lambda = 193\ nm$?



Wet Lithography

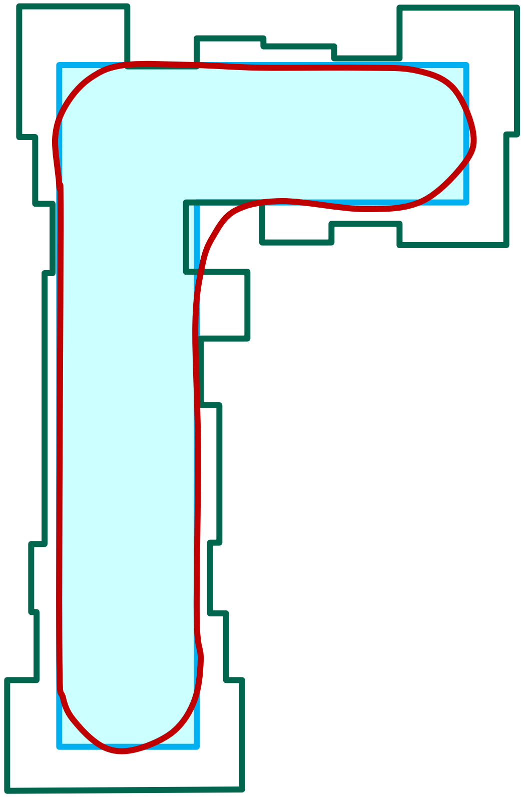


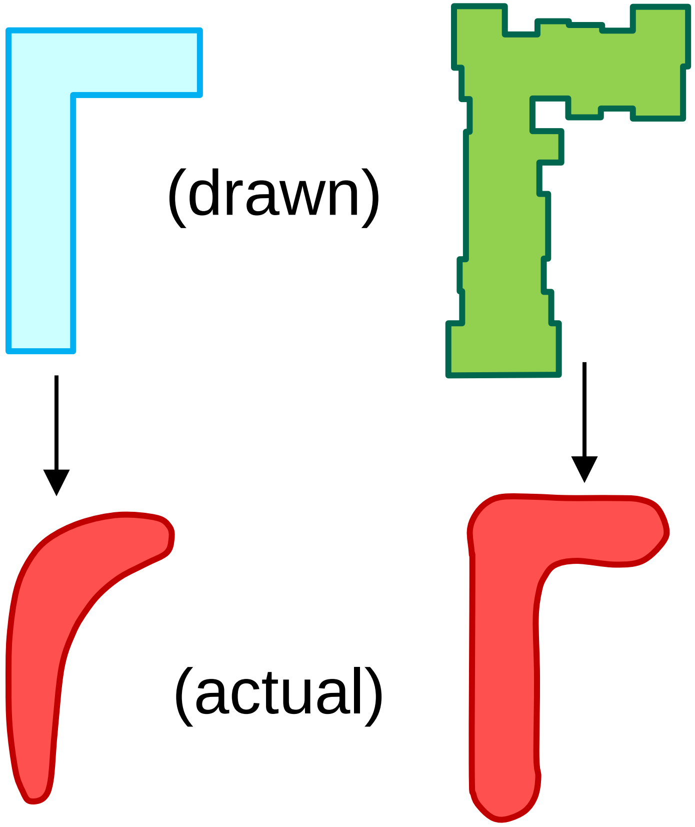
Dry Lithography

Wet Lithography

- Wavelength reduced by refractive index of water

Optical Proximity Correction





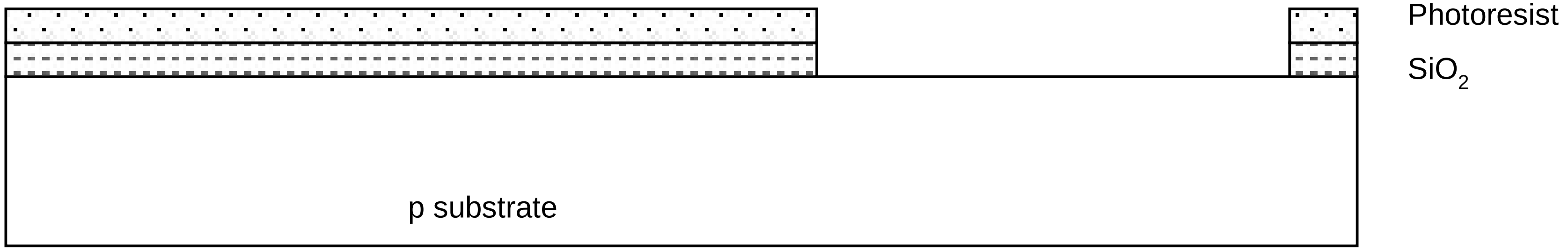
(drawn)

(actual)

- Mask pattern is modified to compensate for diffraction effects
- CAD tools have software to generate these patterns
  - typically based on library of pre-computed shapes

Etch

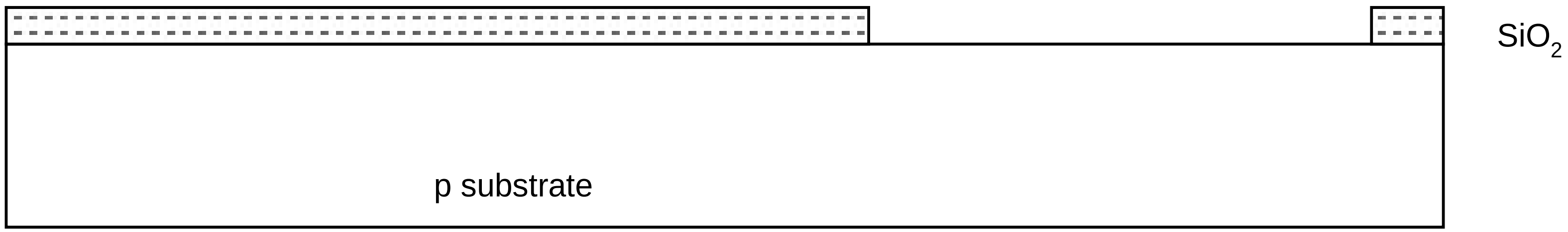
- Etch oxide with hydrofluoric acid (HF)
- Seeps through skin and eats bone; nasty stuff!!!
- Only attacks oxide where resist has been exposed



Strip Photoresist

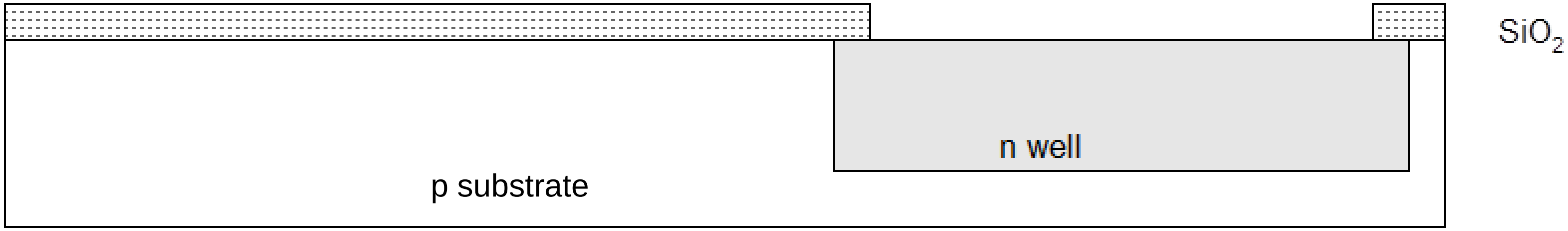
- Strip off remaining photoresist
- Use mixture of acids called *piranha etch*
  - mixture of H_2SO_4 and H_2O_2
- Necessary so resist doesn't melt in next step

食人鱼



Form n-well

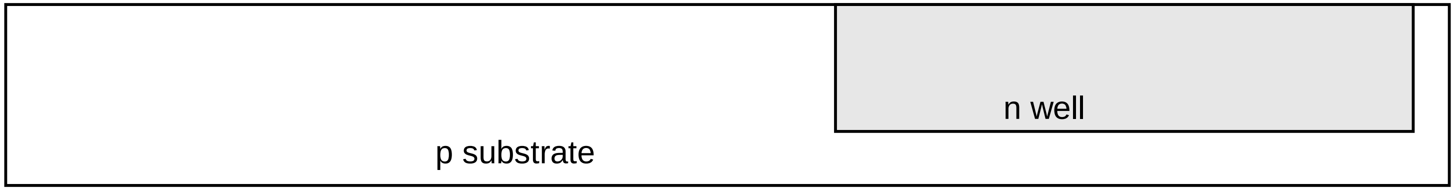
- n-well is formed by counter-doping with arsenic (donor impurity) using diffusion or ion implantation
- Diffusion
  - Place wafer in furnace with arsine
    - AsH_3 – lethal at a few ppm – really nasty stuff!
  - Heat until As atoms diffuse into exposed Si
- Ion Implanatation
  - Blast wafer with beam 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

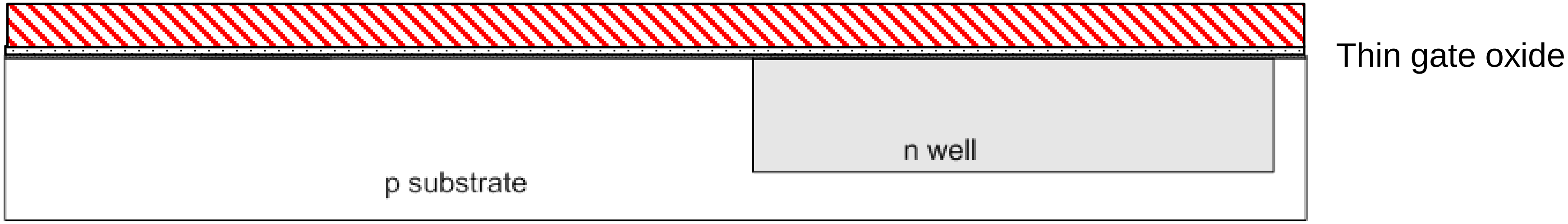
- Subsequent masks involve similar series of steps



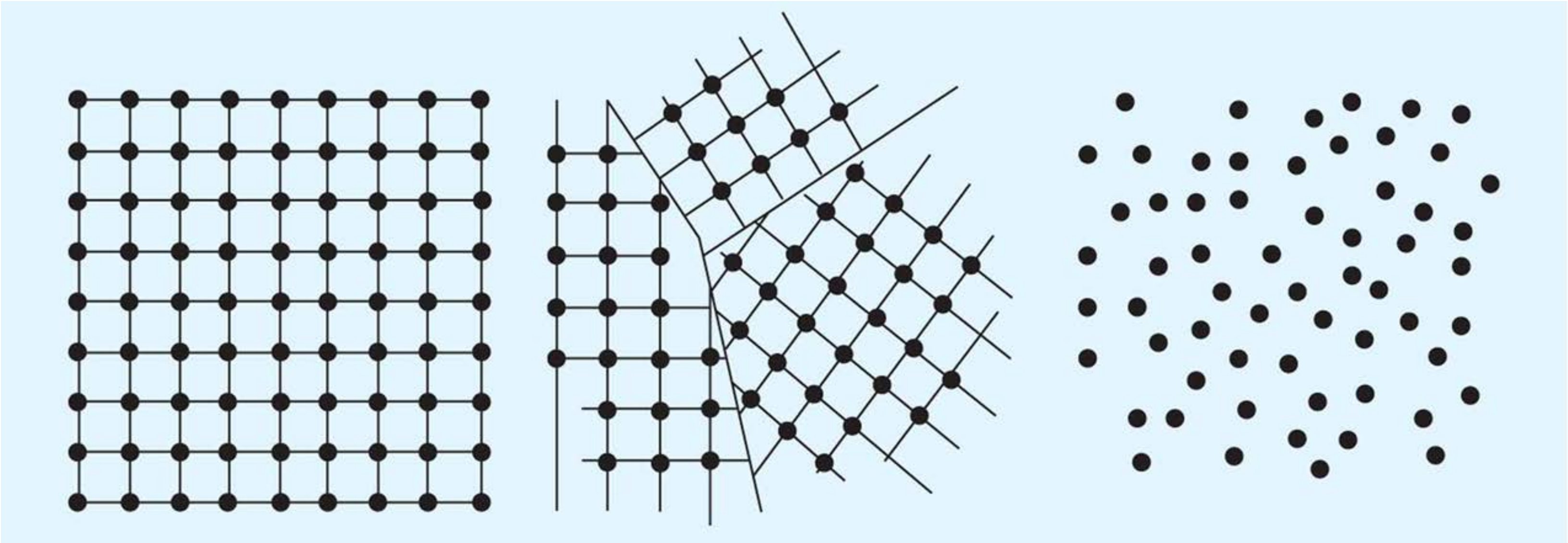
Gate Oxide and Polysilicon

- Deposit very thin layer of gate oxide
  - 40 Å (~13 atomic layers) at 180nm node
  - 20 Å (6-7 atomic layers) at 130nm node
  - 12 Å (4-5 atomic layers) at 65nm node

- Chemical Vapor Deposition (CVD) of silicon layer
  - Place wafer in furnace with Silane gas (SiH_4) - pyrophoric
  - Forms many small crystals called polysilicon
  - Heavily doped to be good conductor



Polysilicon Structure



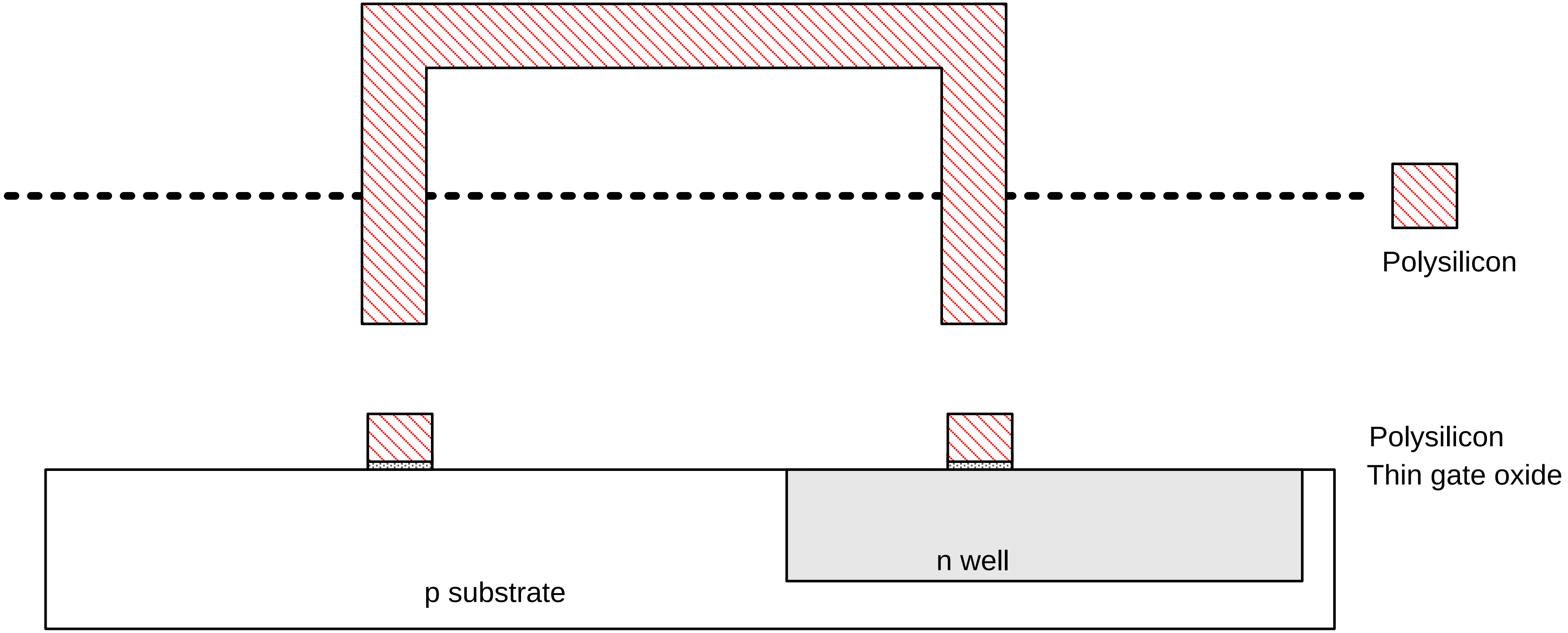
Crystalline silicon

Polycrystalline silicon

Amorphous silicon

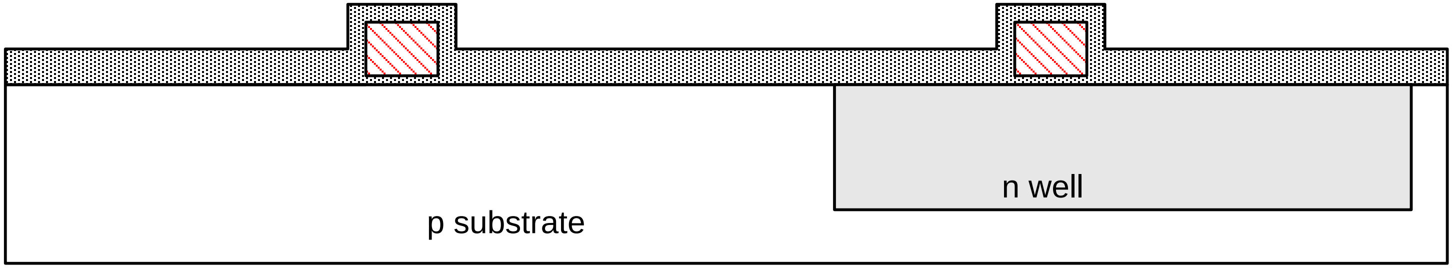
Polysilicon

- Use same lithography process to pattern polysilicon



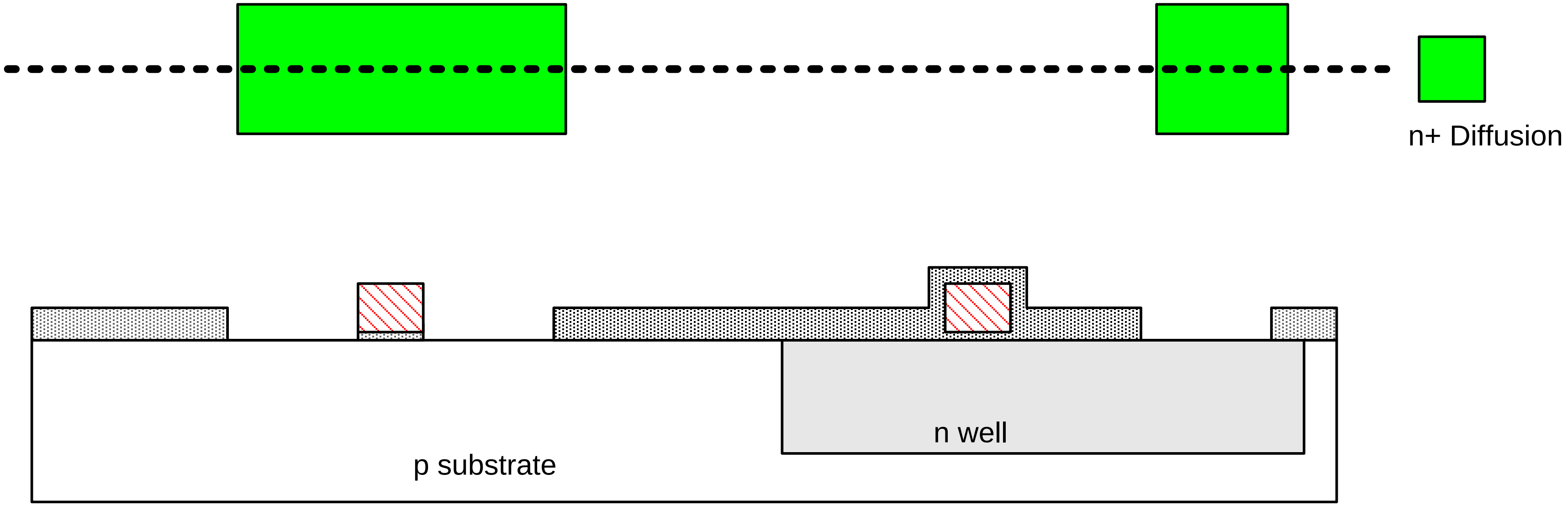
N+ Diffusion / Implantation

- Grow another layer of SiO_2
- Will use oxide and masking to expose where n+ dopants should be diffused or implanted
- N-diffusion forms nMOS source/drain, and n-well contact



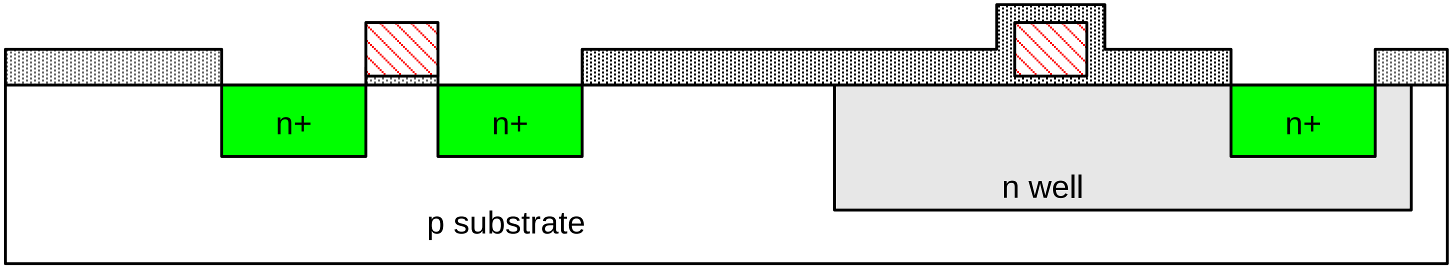
Self-Aligned Process

- 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

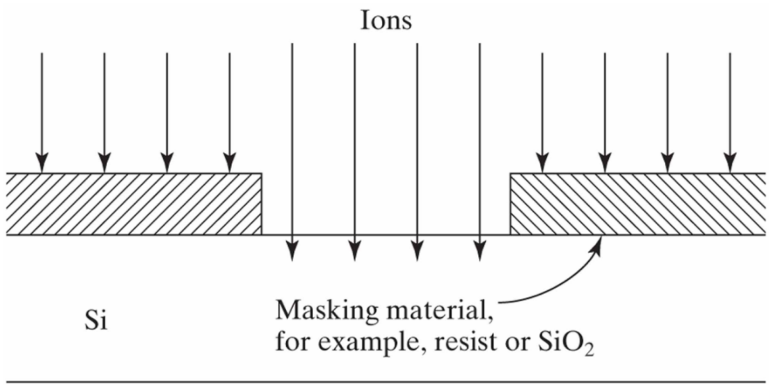


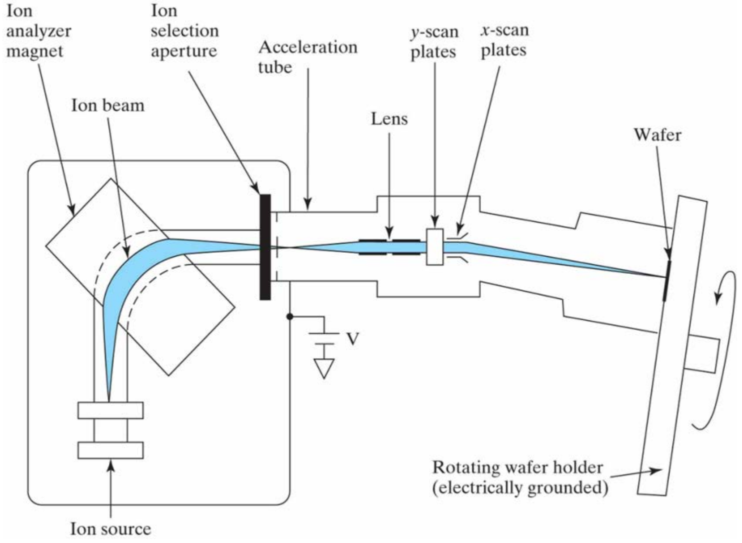
N+ Diffusion (cont.)

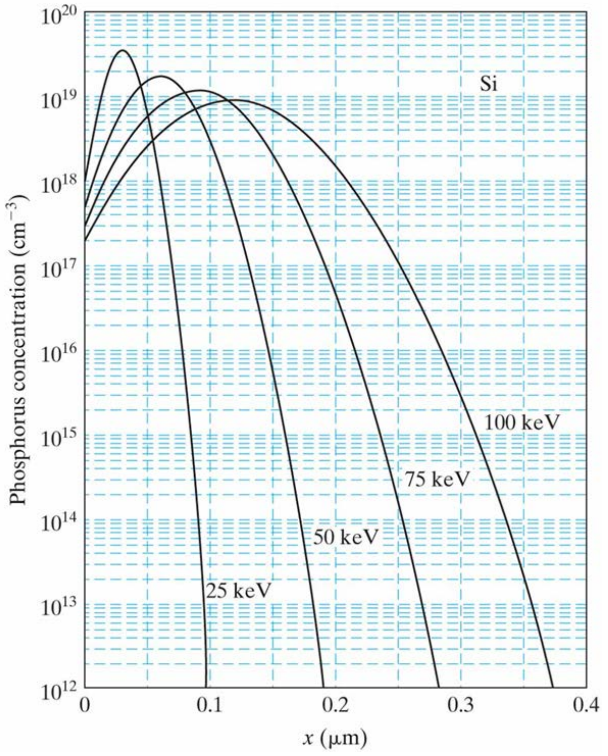
- Historically dopants were diffused
- Usually ion implantation today
- But these n+ regions are still called diffusion



Ion Implantation



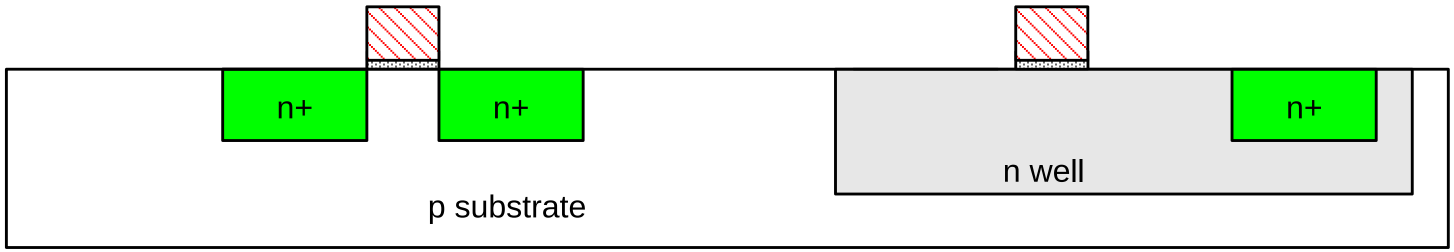




- The dominant doping method
- Excellent control of dose (ions/cm^2)
- Good control of implant depth with ion energy (KeV to MeV)
- Repairing crystal damage and dopant activation requires annealing, which can cause dopant diffusion and loss of depth control.

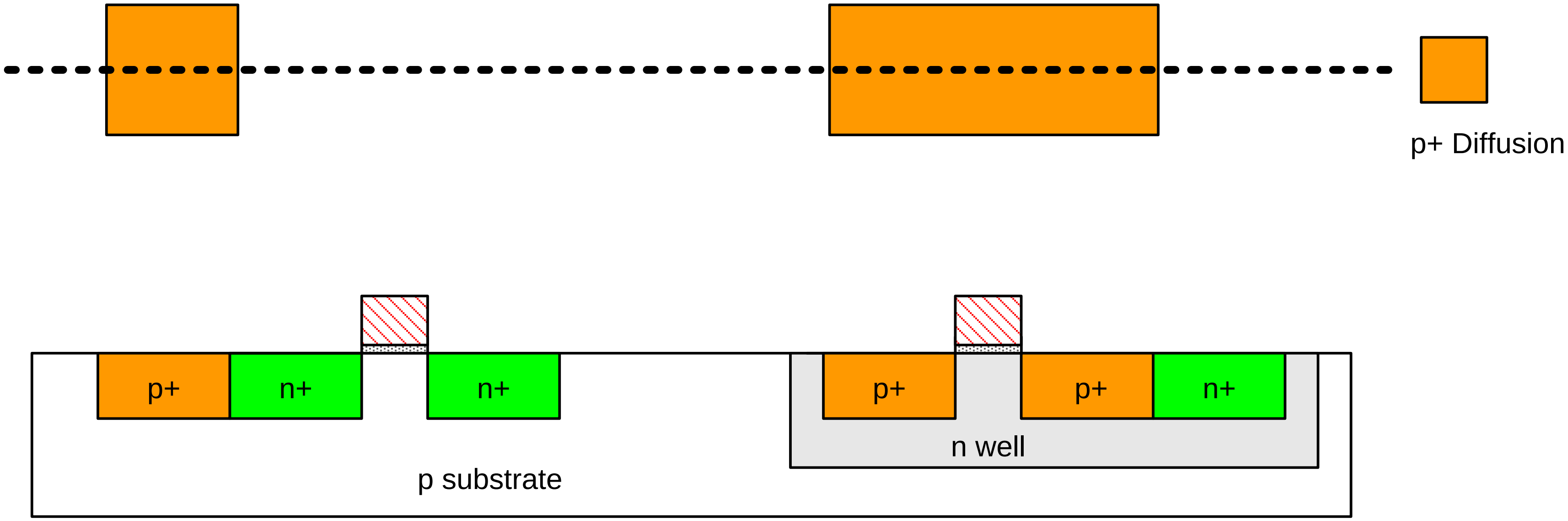
N+ Diffusion (cont.)

- Strip off oxide to complete patterning step



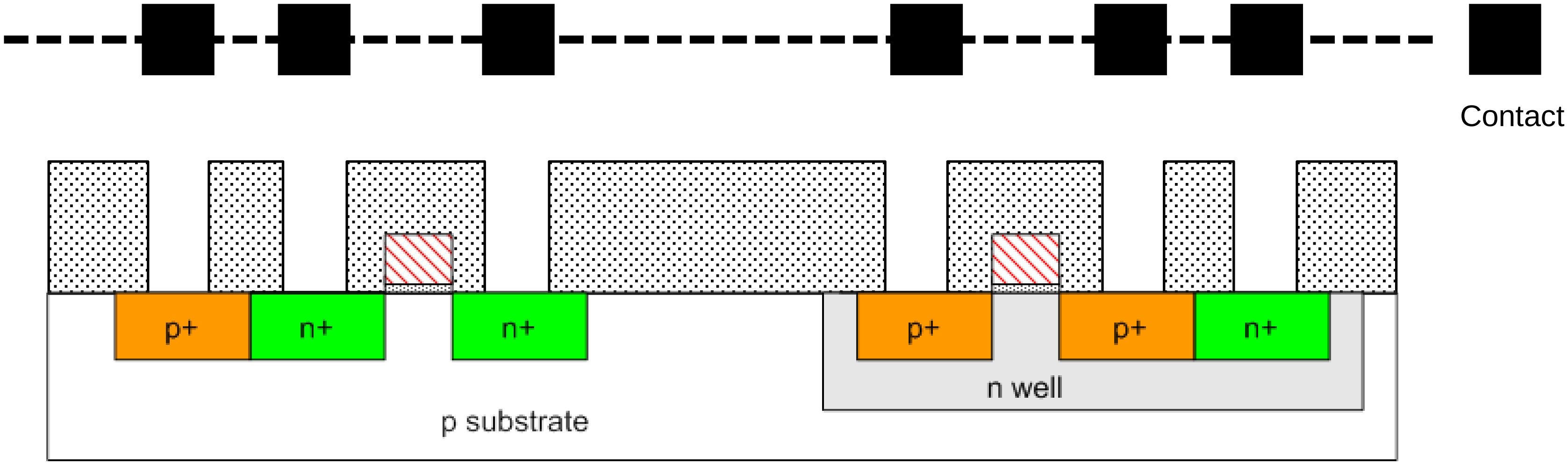
P+ Diffusion / Implant

- Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact
- Boron atoms are implanted in the unmasked silicon



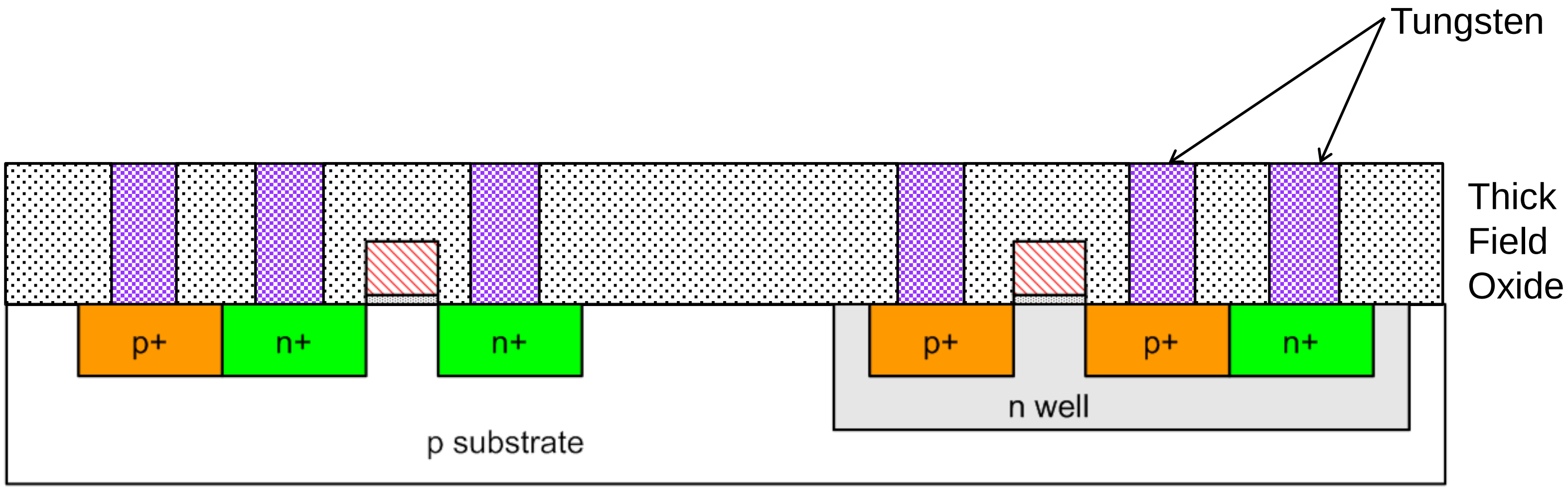
Contacts

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



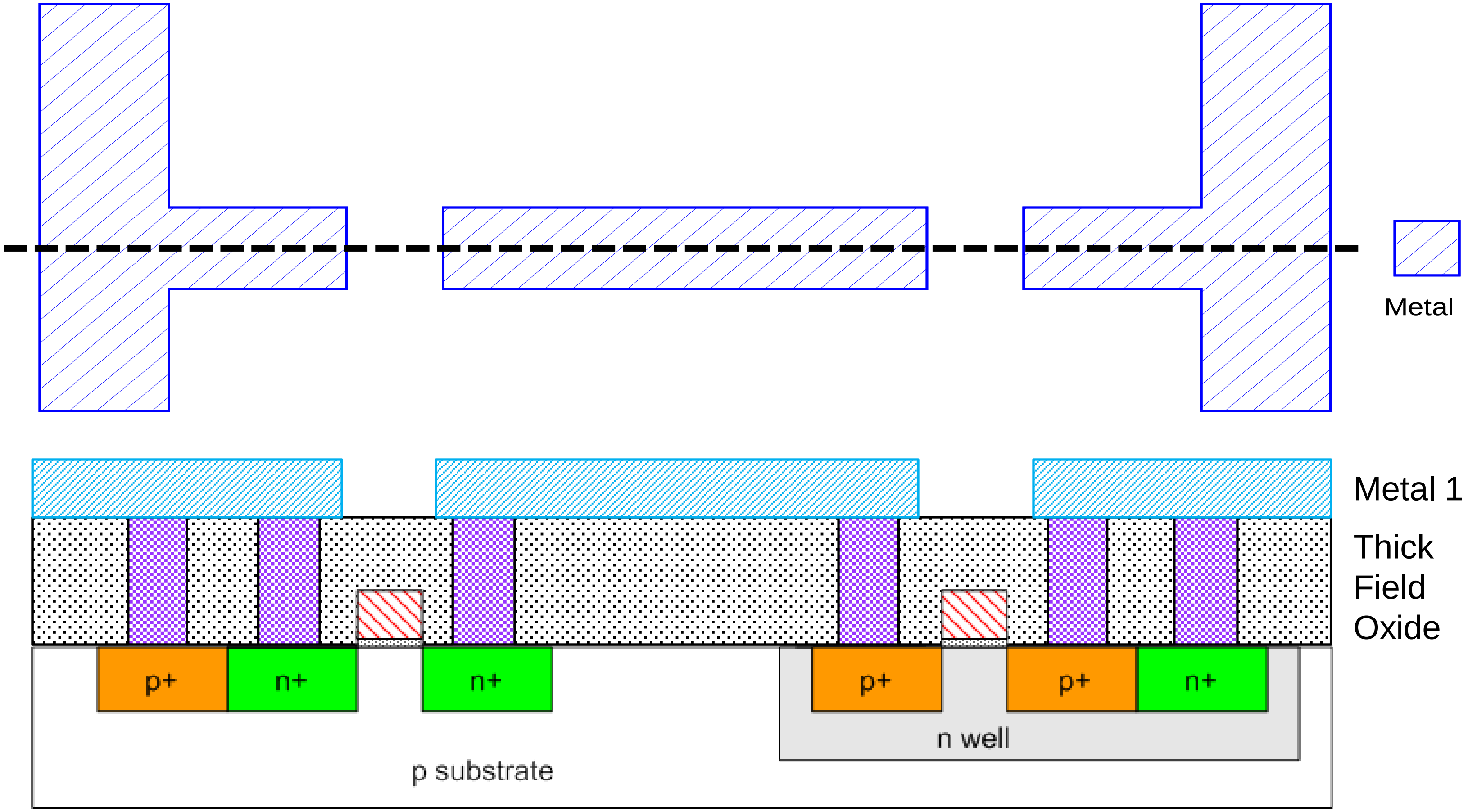
Tungsten Plugs

- A layer of tungsten is grown over surface
- Etched away to leave only contact holes filled with tungsten
- Tungsten conforms better (than Al) to geometry of small holes



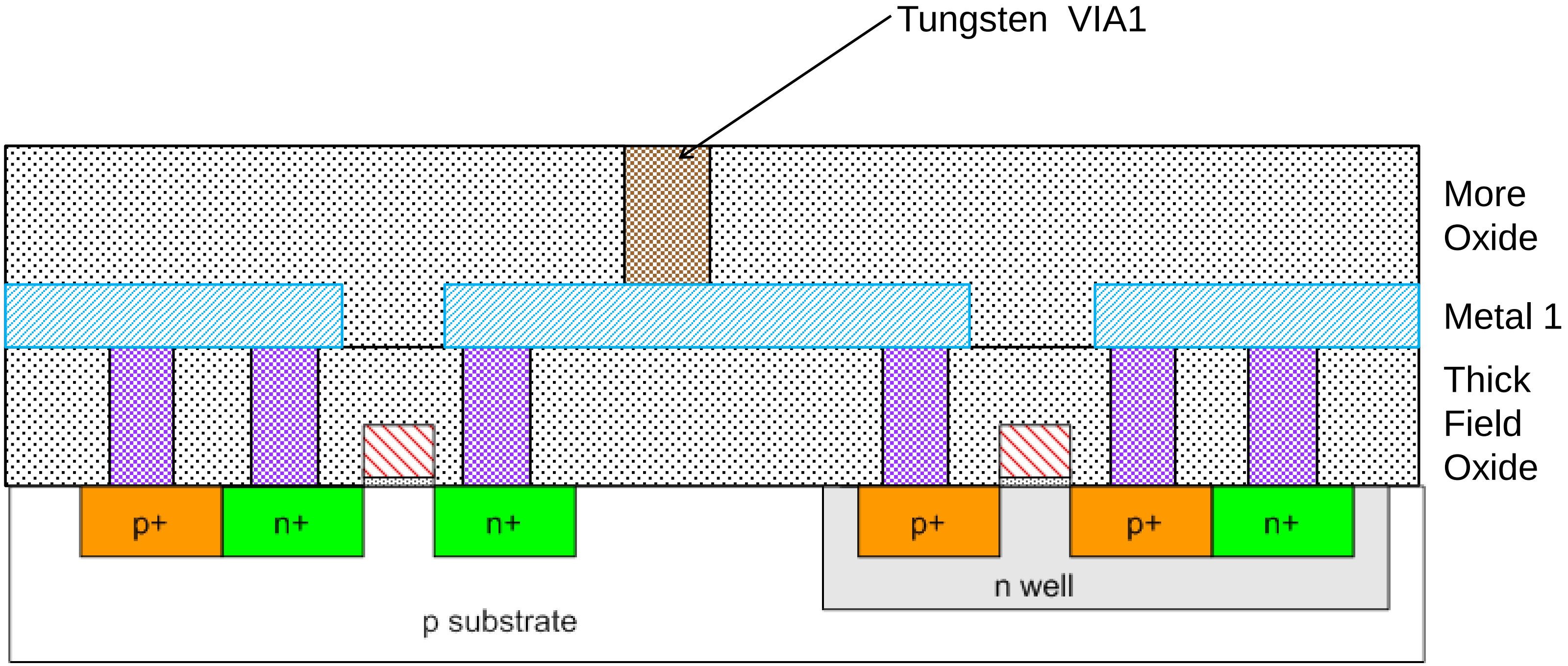
Metallization – Metal 1

- Sputter on aluminum over whole wafer
  - Patterned and plasma etched to remove excess metal, leaving wires
  - Aluminum (metal 1) wires connect (via plugs) to source/drain regions
  - M1 also connects to poly (not shown in this example)



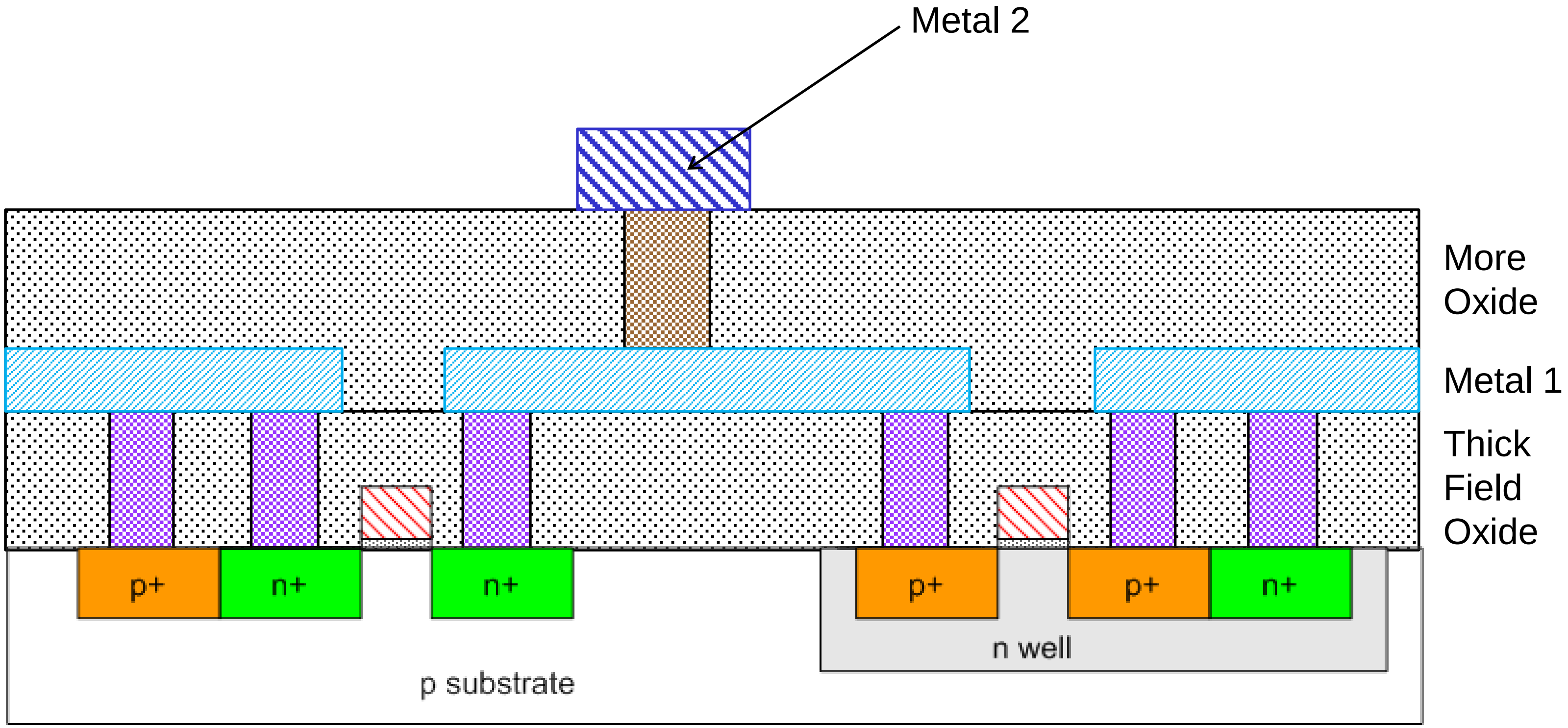
More Tungsten Plugs

- Suppose we want to connect our first layer metal (M1) to a higher metal routing layer (M2)
- Grow another layer of SiO_2 as an insulating dielectric
- Etch VIA holes (VIA1) to connect M2 to M1
- Fill with Tungsten



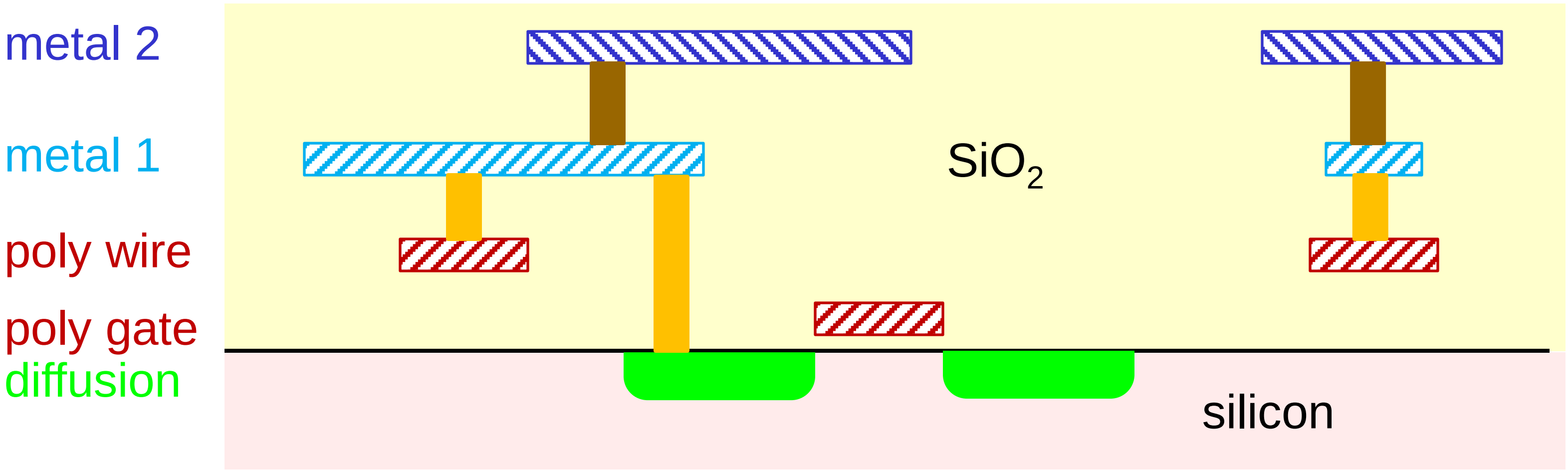
Second Layer of Metallization – M2

- Pattern and plasma etch second layer of metal (M2)
- M2 connects to M1 through VIA1
- If there is a third layer of metallization, M2 connects to M3 through VIA2 (not shown)
- M2 <u>cannot</u> connect (directly) to poly or diffusion



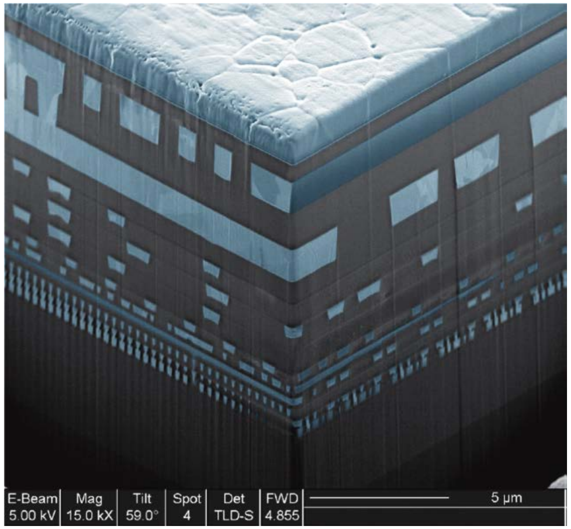
Contacts & Vias

	Diffusion	Poly Gate	Poly Wire	Metal1	Metal2
Diffusion	✓	✗	✗	contact	✗
Poly Gate	✗	✓	✓	✗	✗
Poly Wire	✗	✓	✓	contact	✗
Metal1	contact	✗	contact	✓	VIA1
Metal2	✗	✗	✗	VIA1	✓



Higher Metallization Layers

- Need at least 3-4 layers of metal to support dense custom (hand-drawn) layout.
- Automatic place & route tools rely on multiple metal layers to create dense designs with good power & clock distribution and minimum parasitics.
- Modern processes have 5-10 layers of metal
  - upper layers often Cu (rather than Al_)
  - each layer requires *via* and a *metal pattern* mask

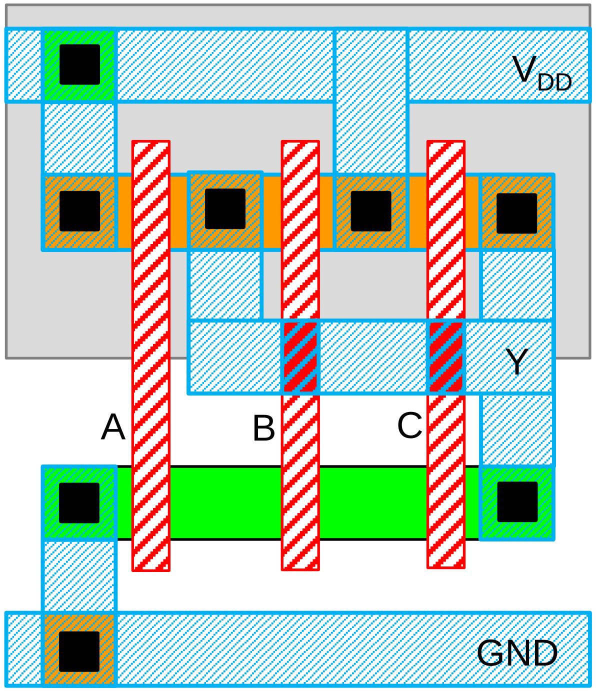


cross-section showing 11 metallization layers (Courtesy IBM)

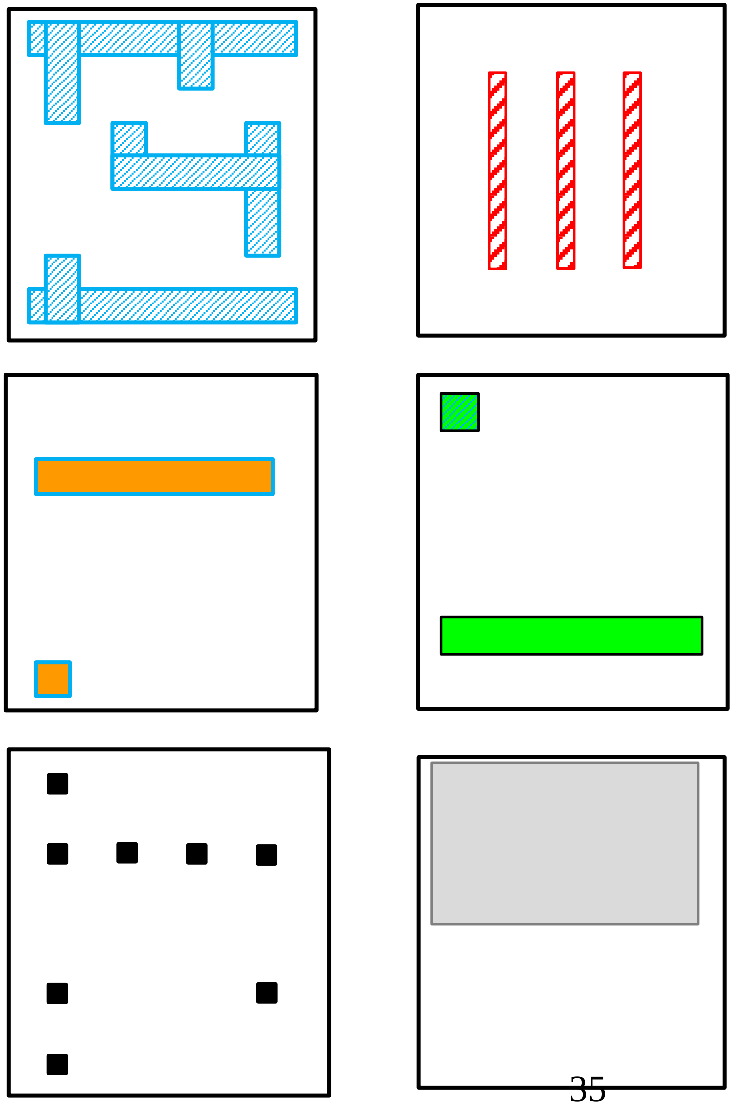
Mask vs. Layout

- Chips are built with set of masks
- Layout designers job is to define patterns for each mask
- Layout is specified using a number of “layers”
  - Layout layers are mapped to mask levels
- Some layers correspond directly to specific masks
  - e.g. poly, metal1, contact
- Other layers might be combined to create a mask
  - e.g. $(\text{diff}_{\text{layout}}$ AND $\text{nplus}_{\text{layout}}) \Rightarrow \text{NDIFF}_{\text{mask}}$
  - $(\text{diff}_{\text{layout}}$ AND $\text{pplus}_{\text{layout}}) \Rightarrow \text{PDIFF}_{\text{mask}}$
- Other layers may be added to assist CAD tools
  - e.g. designating ndiff wire as diffusion resistor

Mask to Layout



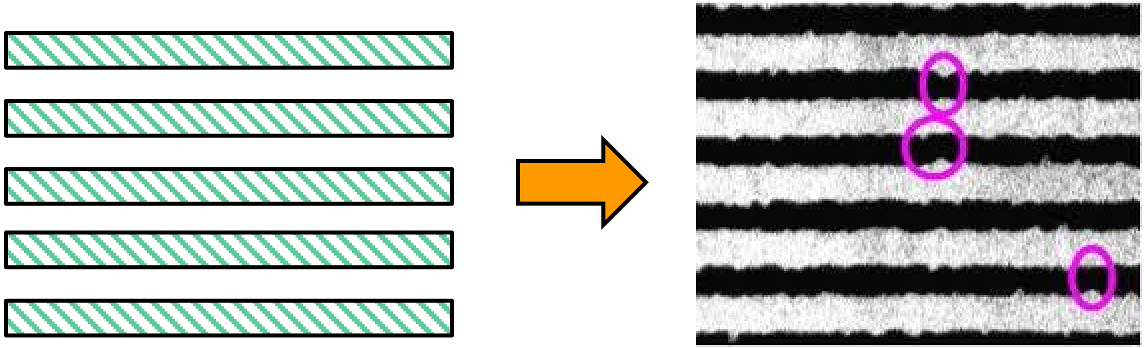
(layout)



(masks)

Process Limitations and Design Rules

- Would like to make objects (transistors, wires etc.) as small as possible
  - to increase speed, decrease cost & power
- Object size and spacing is limited by precision of photolithography & manufacturing process



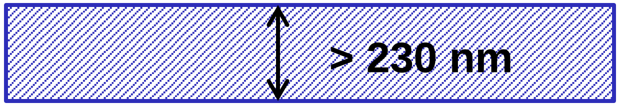
layout

on-chip wiring

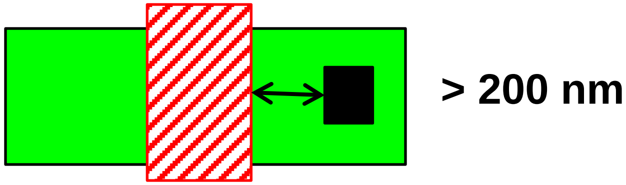
- Need “Design Rules” to constrain layout engineer
  - ensure design is manufacturable

Layout & Design Rules

- Design Rules set minimum size and spacing for each layer to give acceptable yield (e.g. M1 min width = 230nm)



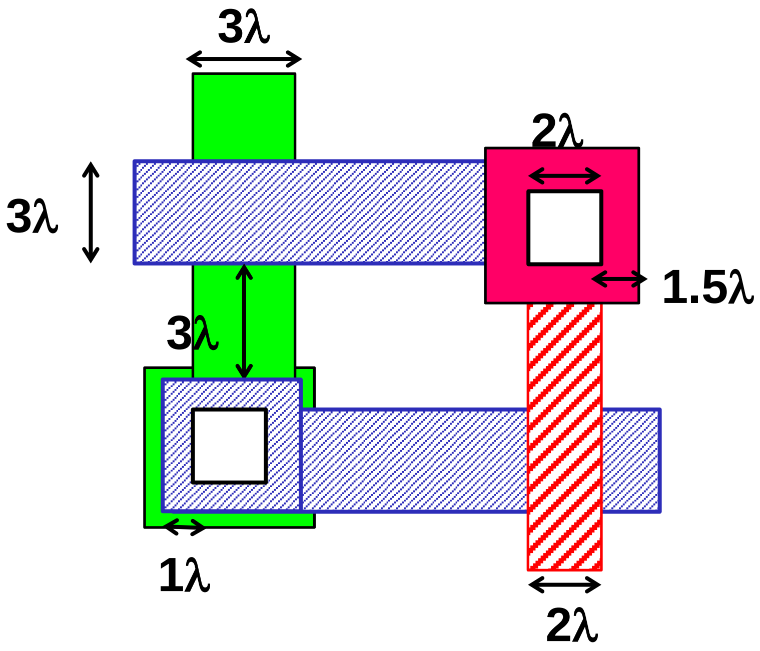
- Design Rules also specify spacing between objects on different layers (e.g. min distance of contact from gate = 200nm)



- Design Rules typically expressed in μm or nm.
- Each CMOS process typically characterized by feature size f = minimum distance between source and drain
- Set by minimum width of polysilicon (e.g. 180nm)
- Feature size improves 30% every 3 years or so

λ based Design Rules

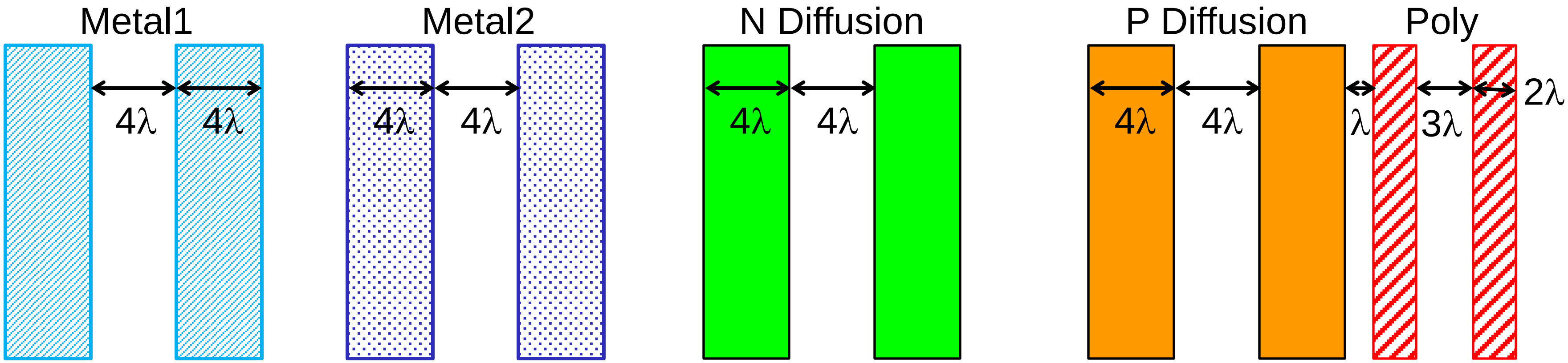
- We can simplify design by adopting a conservative set of rules normalized to minimum feature size *f*
- Express rules in terms of λ = *f*/2
  - e.g. for 180nm process, λ = 90nm
- For example MOSIS SCMOS rules:



- Layout can be scaled to new process by simply changing value of λ

Simplified Design Rules: Conductors

- A simpler (more conservative) set to get you started:

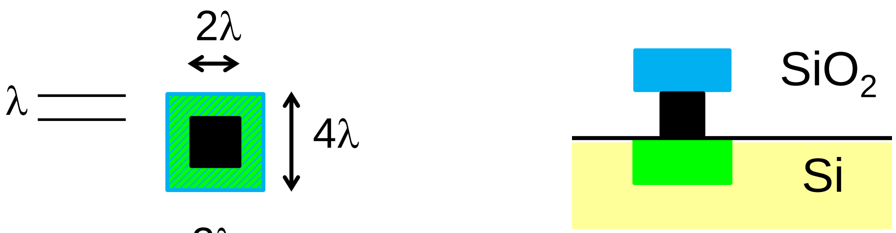


Contact Design Rules

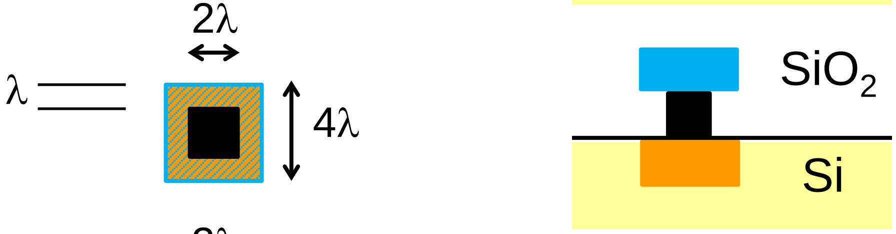
Layout

Cross-section

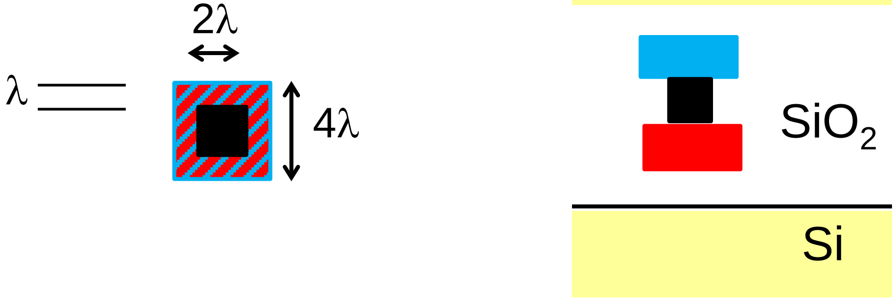
Metal1-Ndiff Contact



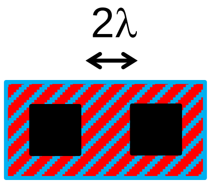
Metal1-Pdiff Contact



Metal1 – Poly Contact



Contact Spacing Rule
(for all 3 types)

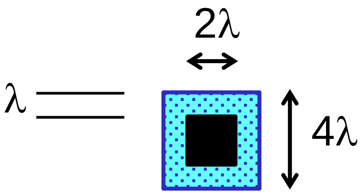


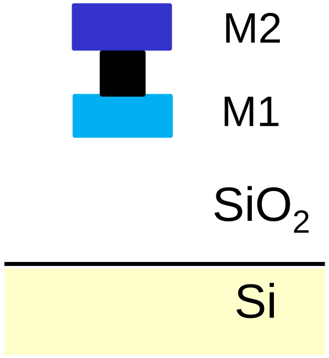
Via Design Rules

Layout

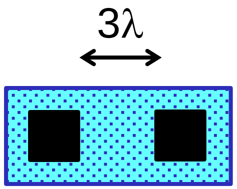
Cross-section

Metal1-Metal2 Via



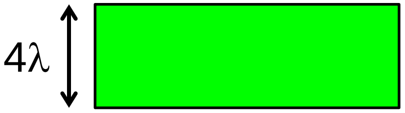


Via Spacing Rule

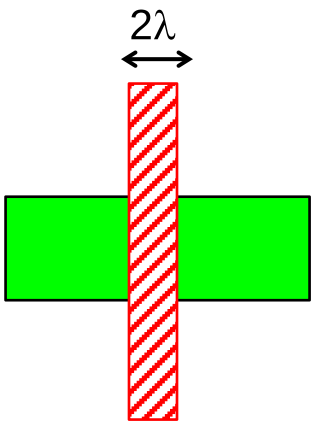


Simplified Design Rules (cont.)

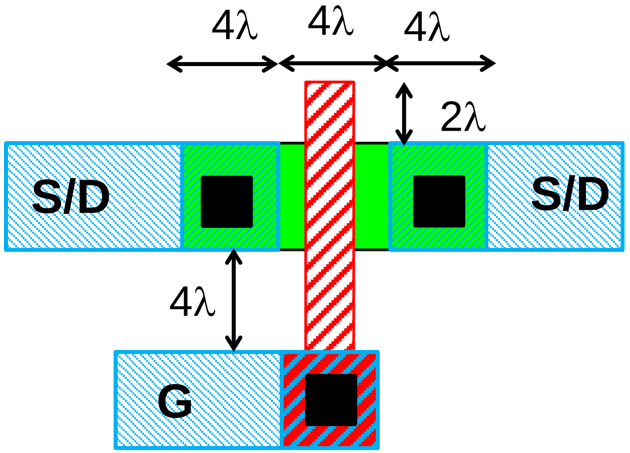
- Creating NMOS devices:



diffusion region becomes S/D regions



poly gate crosses diffusion region, separating S/D



contacts connect M1 to gate and S/D regions

- This produces minimum size device: $W = 4\lambda, \quad L=2\lambda$

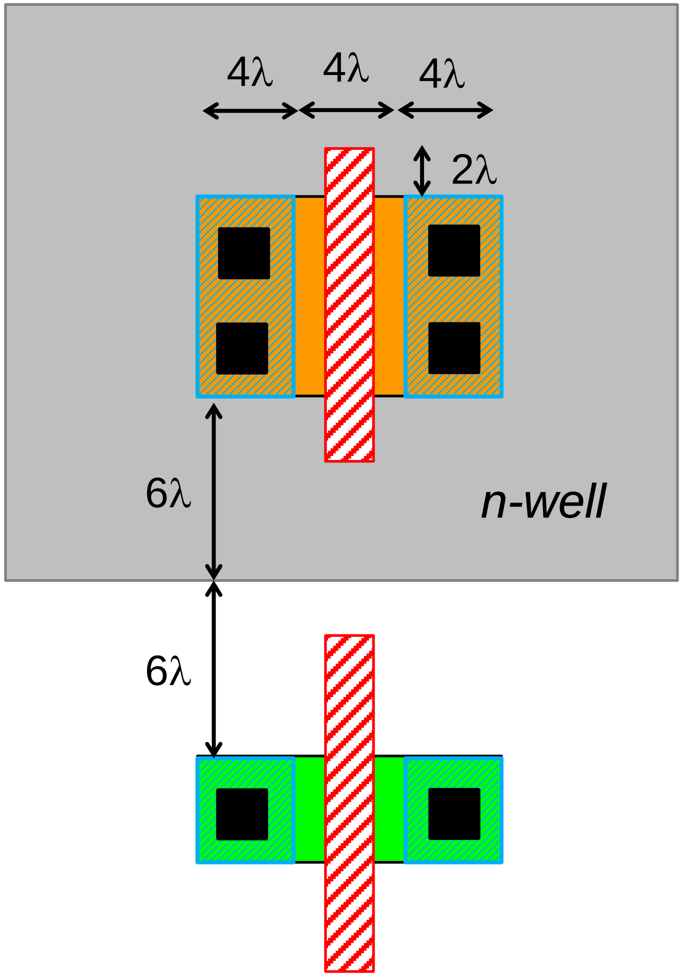
Simplified Design Rules (cont.)

- PMOS device created in same way:

- PMOS device often wider to match drive strength of NMOS: (W = 8λ, L = 2λ)

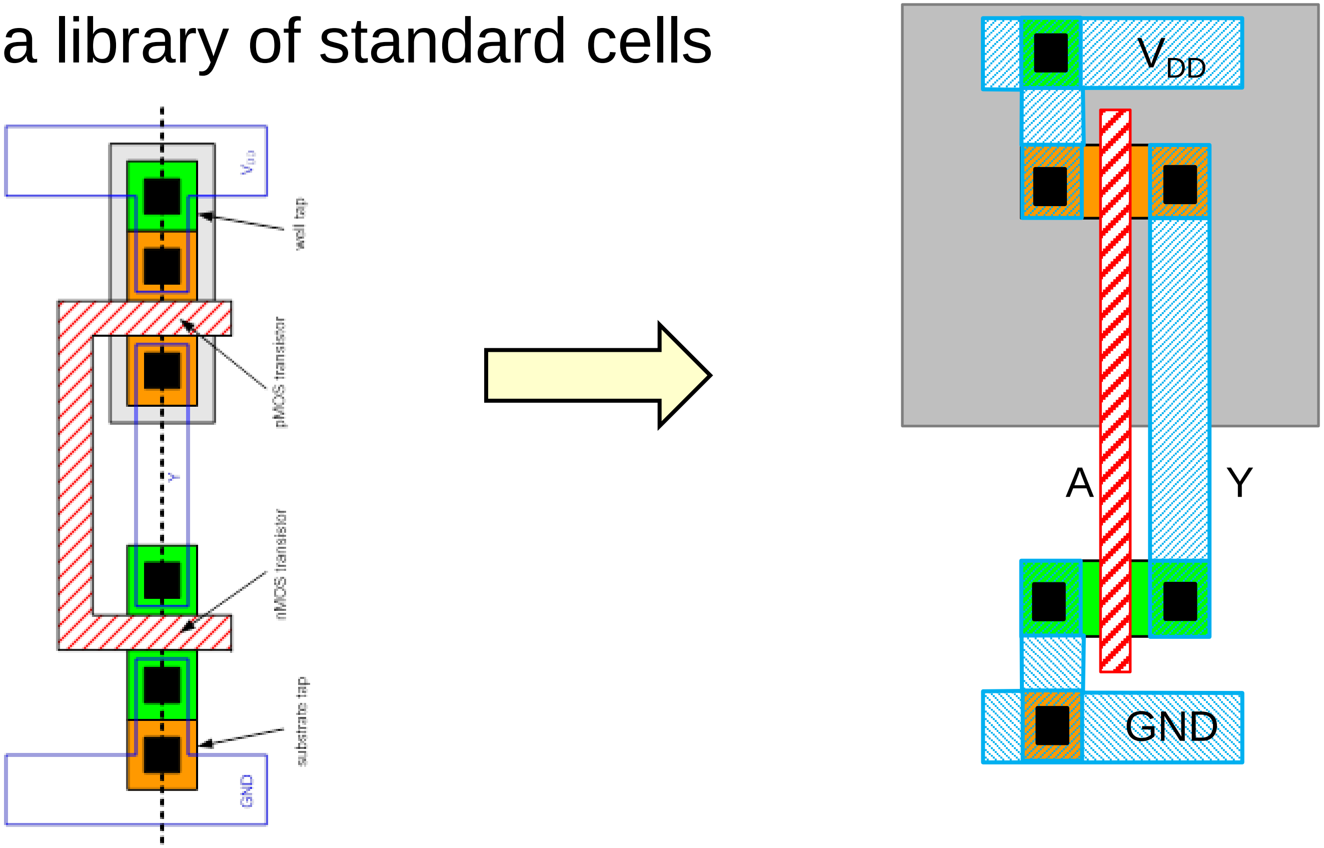
- PMOS device surrounded by nwell (at least 6 λ)

- NMOS device must be separated from nwell by at least 6λ



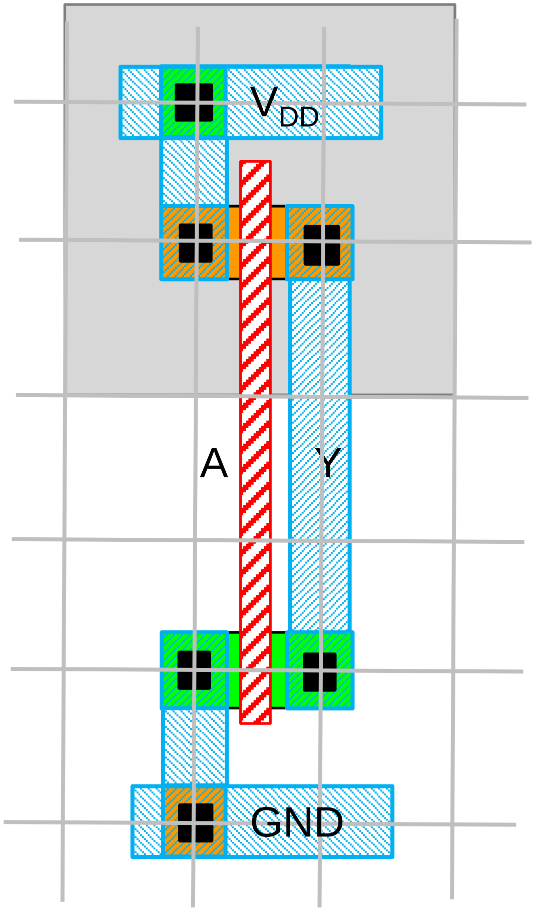
Gate Layout

- Layout can be very time consuming
  - can waste a lot of time trying to squeeze last micron out
- Layout more efficient if we design gates to fit together nicely
- Build a library of standard cells



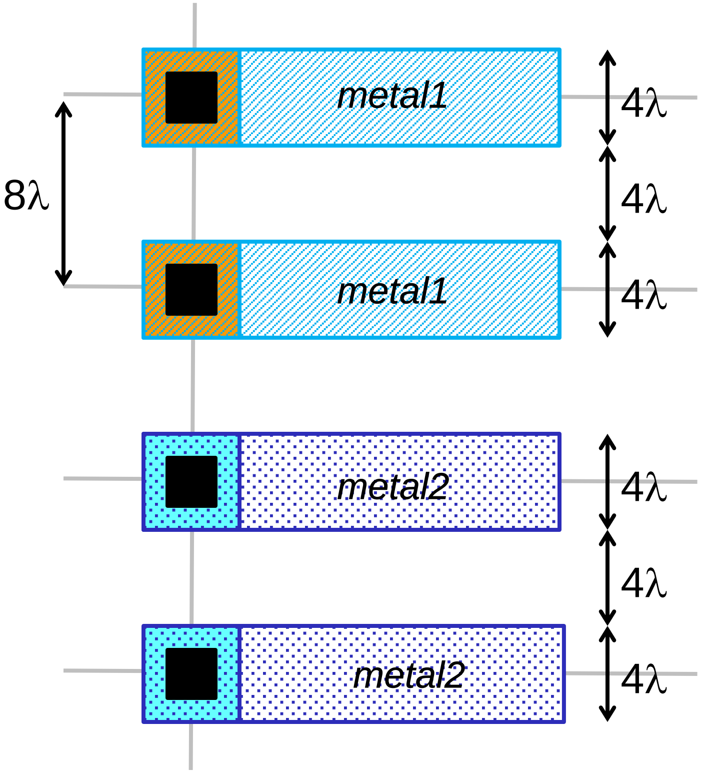
Standard Cell Design Methodology

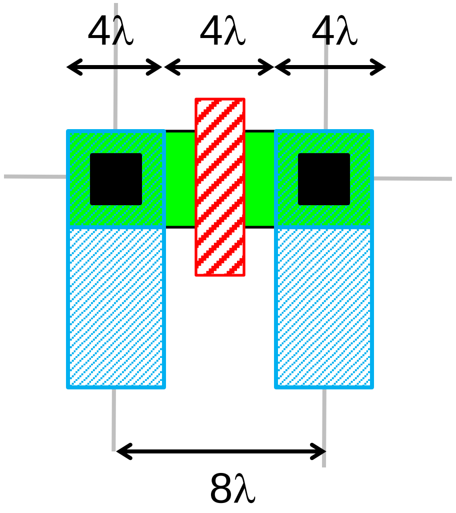
- VDD and GND run horizontally & should abut
  - standard height cell
- nMOS horizontally at bottom and pMOS at top
- Polysilicon runs vertically to connect transistor gates
- All gates include well and substrate contacts
- Adjacent gates should satisfy design rules
  - extend VDD and GND rails by 2λ
- Layout can be built on 8λ x 8λ grid with metal1 wiring tracks between nMOS and pMOS devices.



Wiring tracks

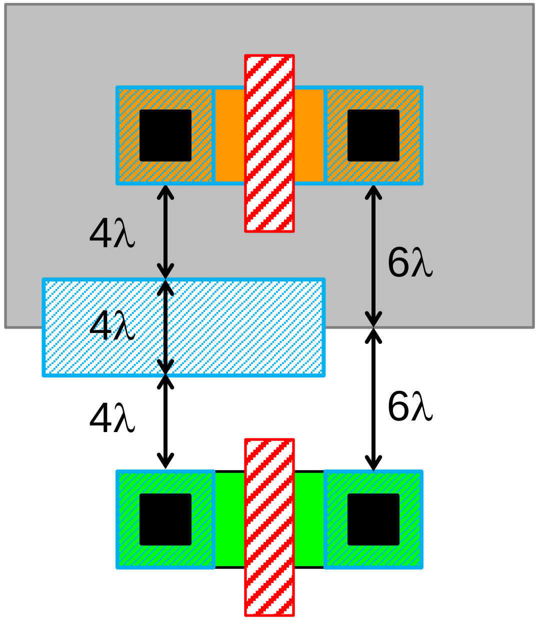
- A wiring track is the space required for a wire
- 4 λ width, 4 λ spacing from neighbor = 8 λ pitch
- Transistors also consume one wiring track





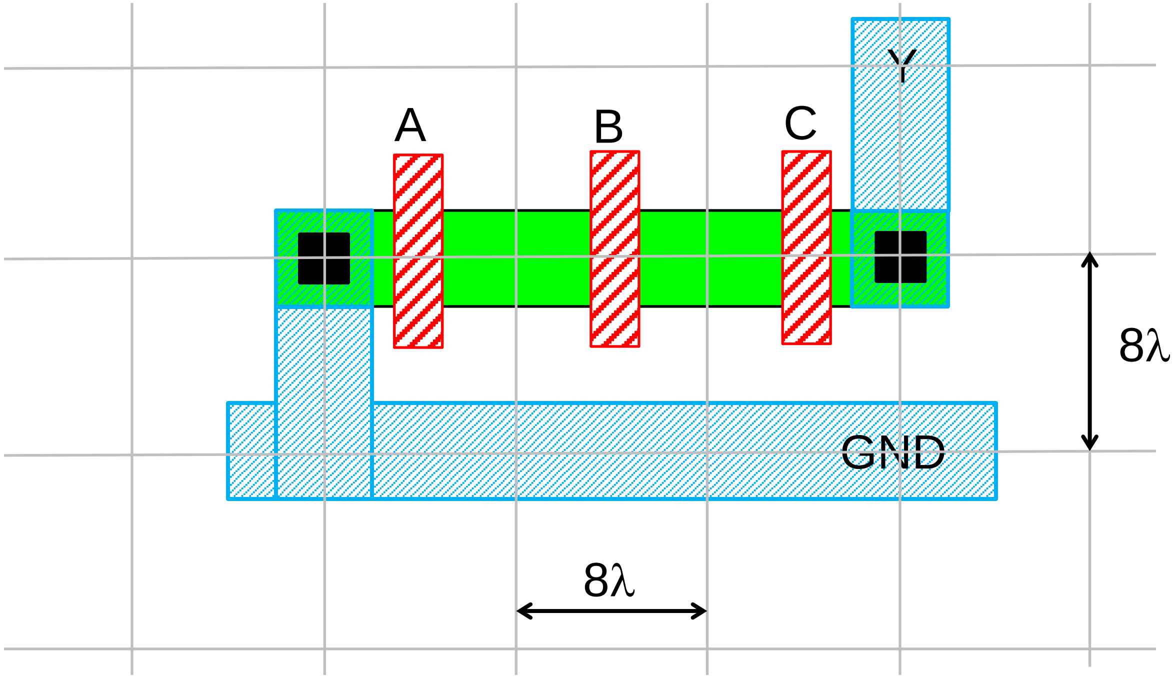
Well Spacing

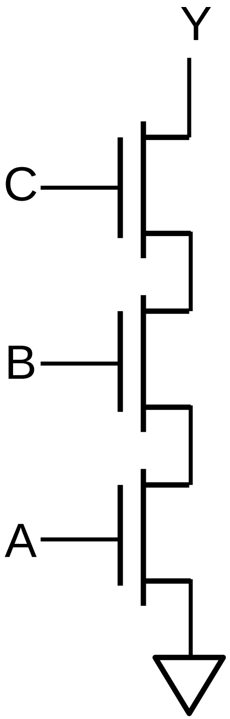
- Wells must surround transistors by 6 λ
- Implies 12 λ between opposite transistor flavors
- Leaves room for one wire track



Building Gates: Transistors in Series

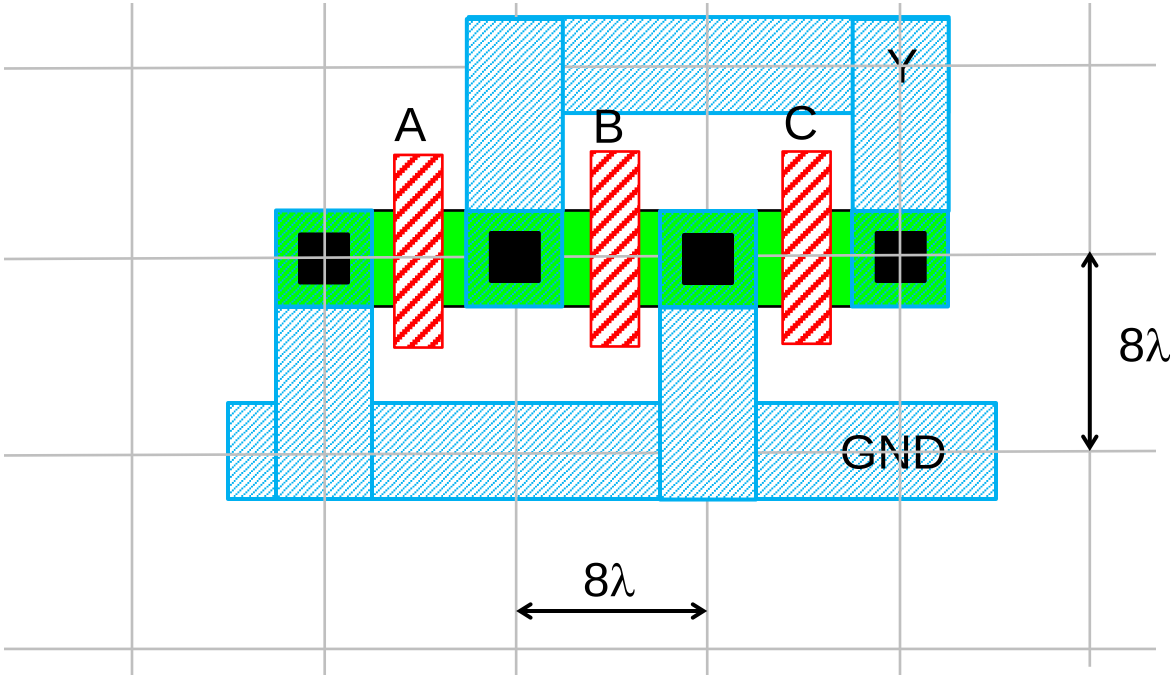
- Transistors can be placed in series by simply overlaying their common source/drain region

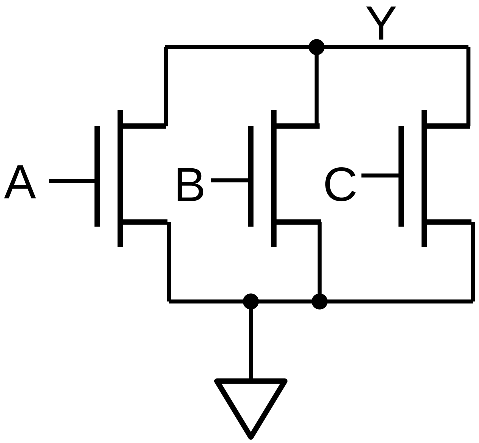




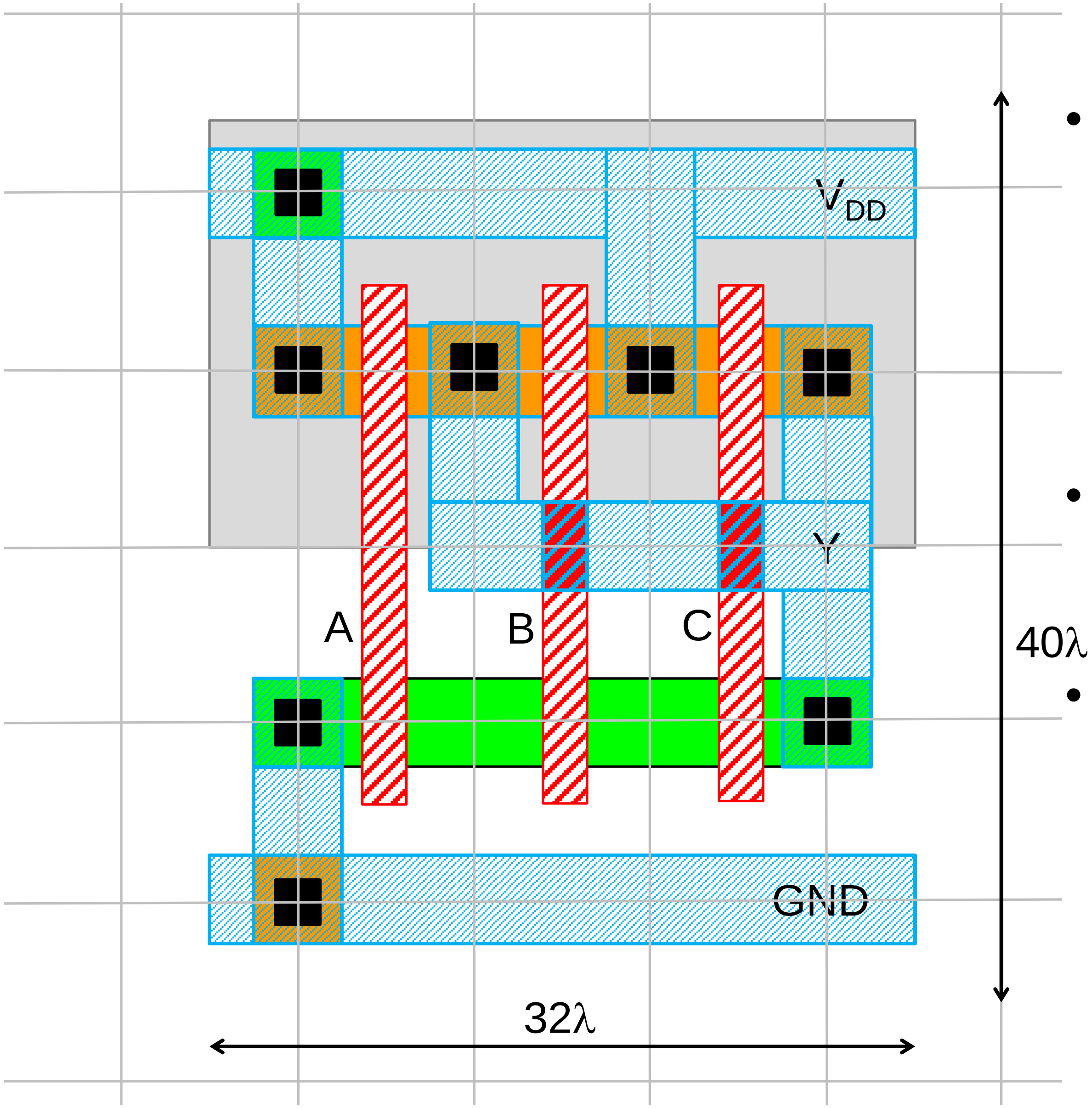
Building Gates: Transistors in Parallel

- Transistors can be placed in parallel by using a combination of source/drain overlap and metal interconnect





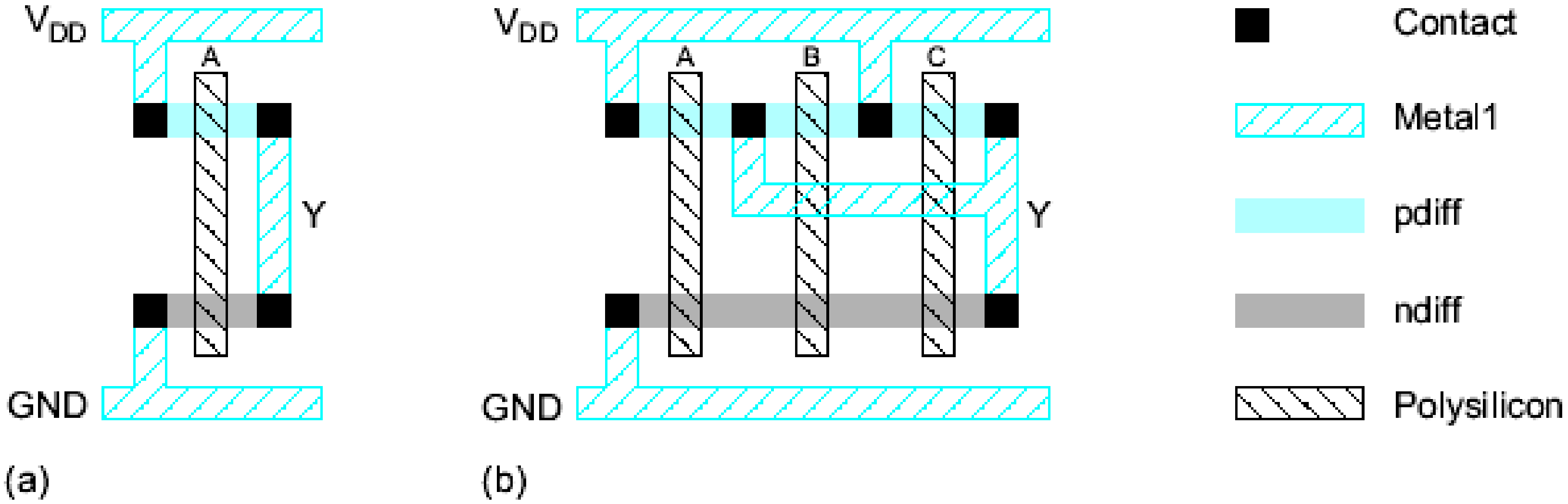
Example: NAND3



- Try to find placement of transistors that maximizes use of common vertical polysilicon and common source/drain overlap
- Estimate area by counting wiring tracks
- Area of this Nand3 is:
  $40\lambda \times 32\lambda = 1280\ \lambda^2$

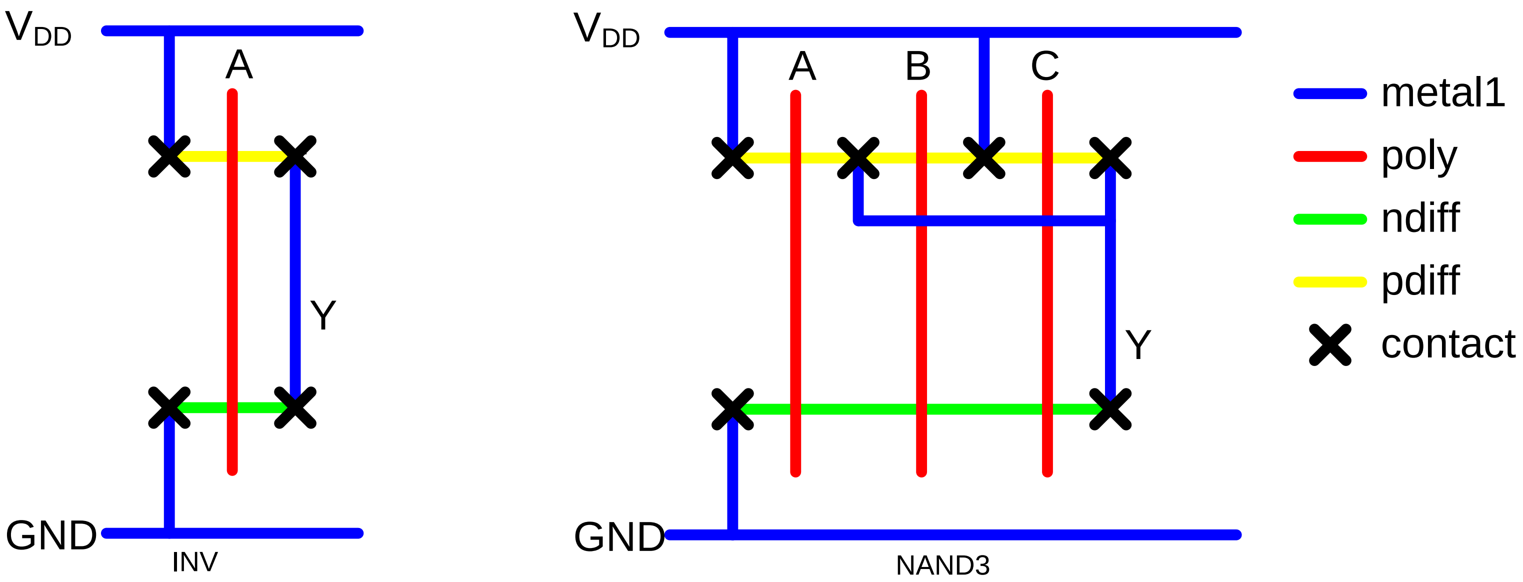
Stick Diagrams

- Stick diagrams help plan layout quickly
- Need not be to scale
- As drawn in textbook (Weste & Harris)



Stick Diagrams

- Or with color pencils or markers:



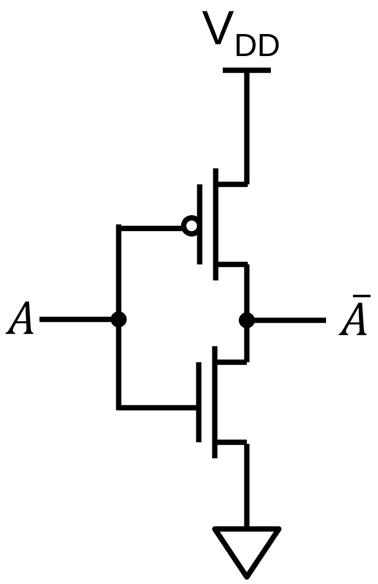
Example: O3AI

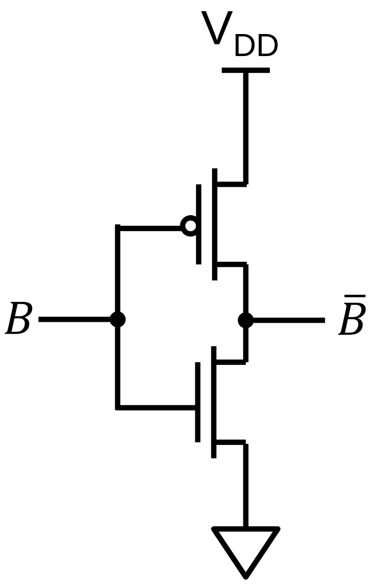
- Sketch a stick diagram for O3AI and estimate area.

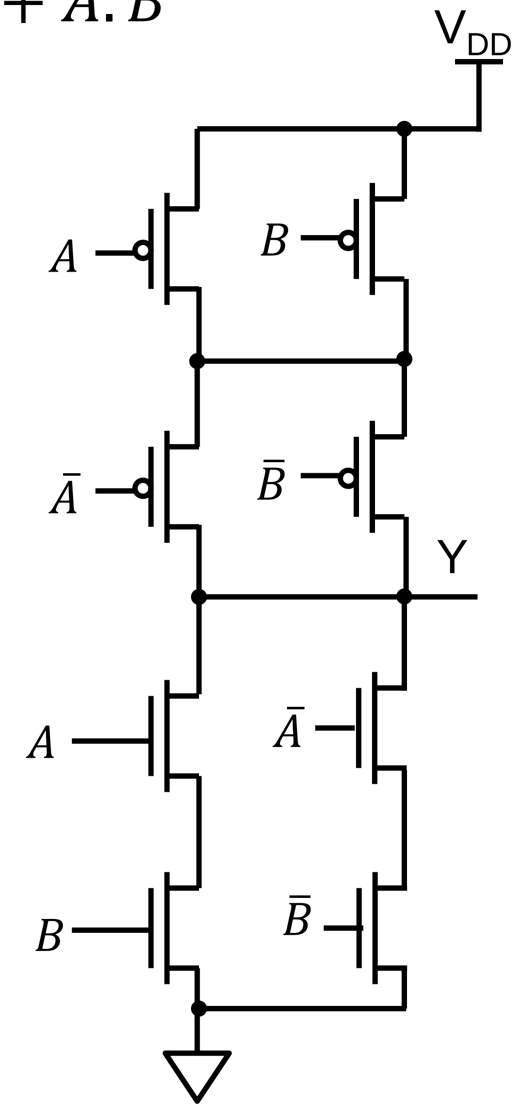
$$Y = \overline{(A + B + C) \bullet \mathit{D}}$$

Example: XOR gate

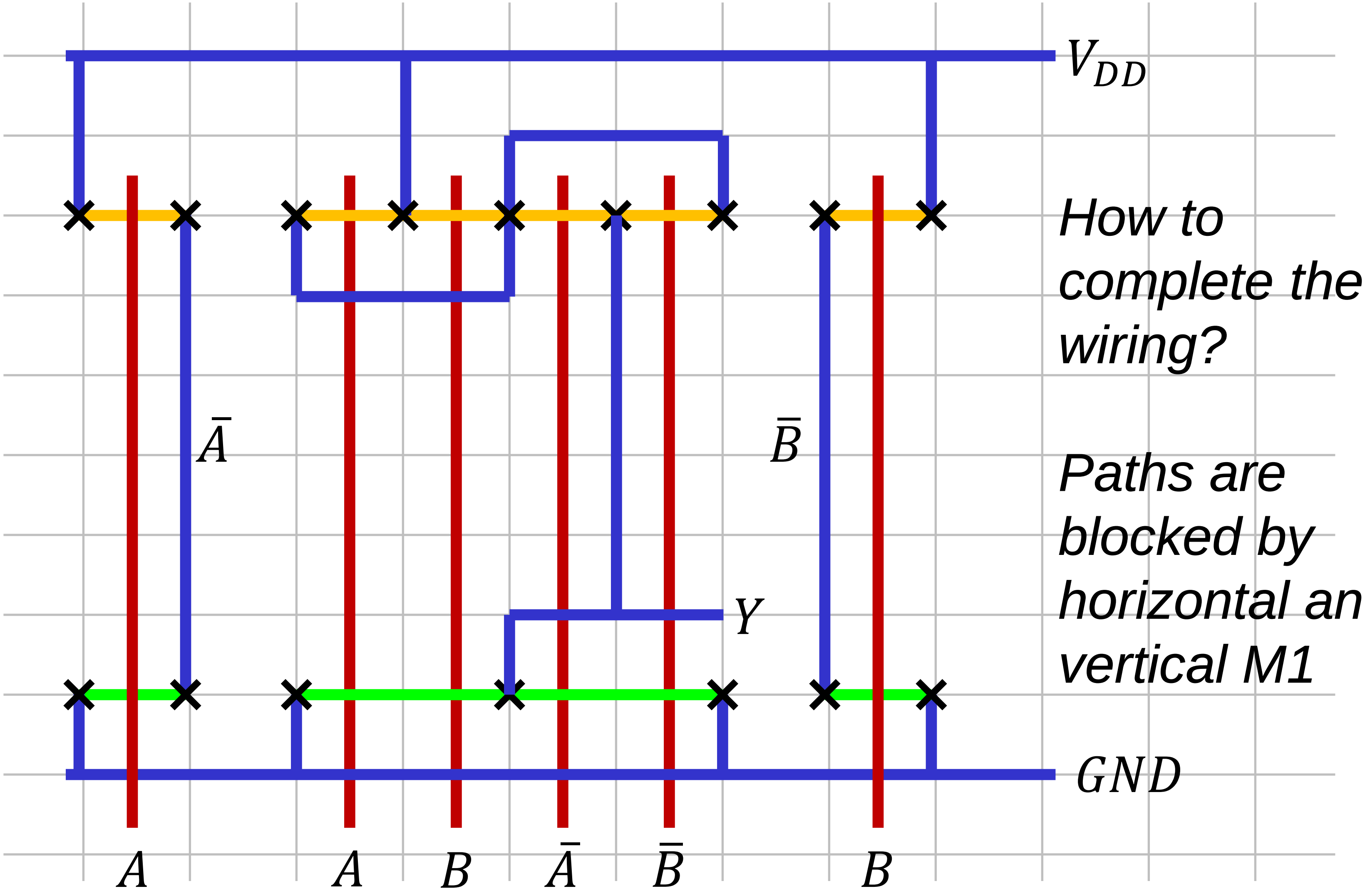
- $Y = A\ xor\ B = A.\bar{B} + \bar{A}.B = \overline{A.B + \bar{A}.\bar{B}}$







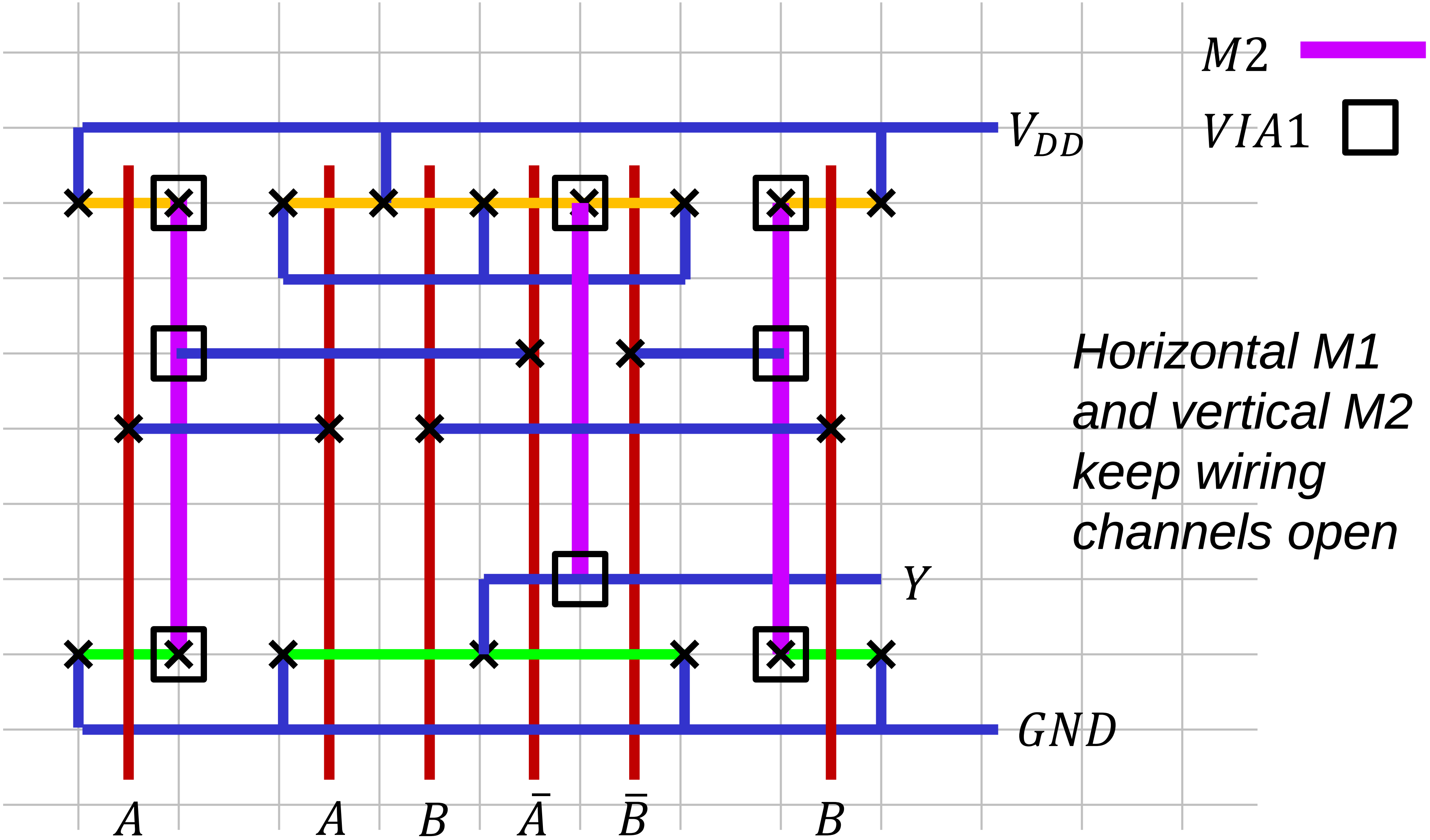
XOR gate layout with M1



How to complete the wiring?

Paths are blocked by horizontal and vertical M1

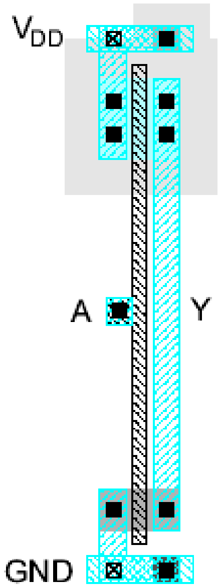
XOR gate layout with M1 & M2

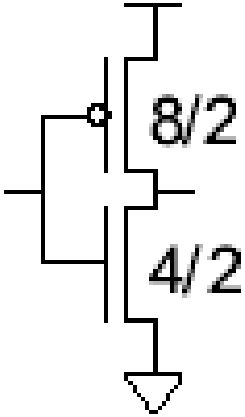


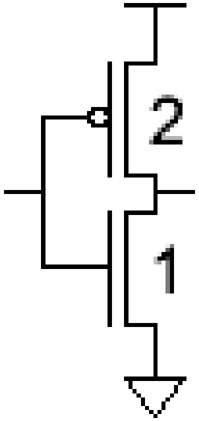
Horizontal M1 and vertical M2 keep wiring channels open

Transistor Sizing

- In most layouts, transistors are not all of same size
  - pMOS has about ½ drive of same size nMOS
  - series/parallel combinations lead to different drive strength
- Transistor dimensions specified as Width / Length
- Minimum size is 4λ / 2λ, sometimes called 1 unit
  - e.g. in f = 0.5 μm process, this is 1.0 μm wide, 0.5 μm long

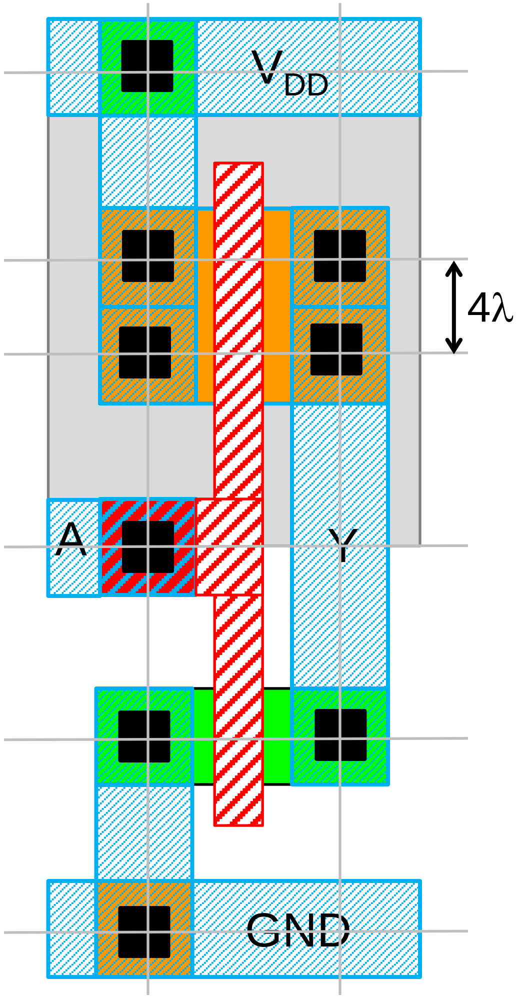






Impact of Sizing on Layout

- Adding sized transistors complicates simple 8λ x 8λ grid
- Still useful for draft layout and approximate area calculations
- When estimating area, add (w-1).4λ in height to accommodate a transistor of width w.
- Add extra contacts when possible
  - Improved contact resistance
  - Improves yield
  - Many designers will use two-contact transistor as “minimum width” device



V_{DD}
4λ
A
Y
GND