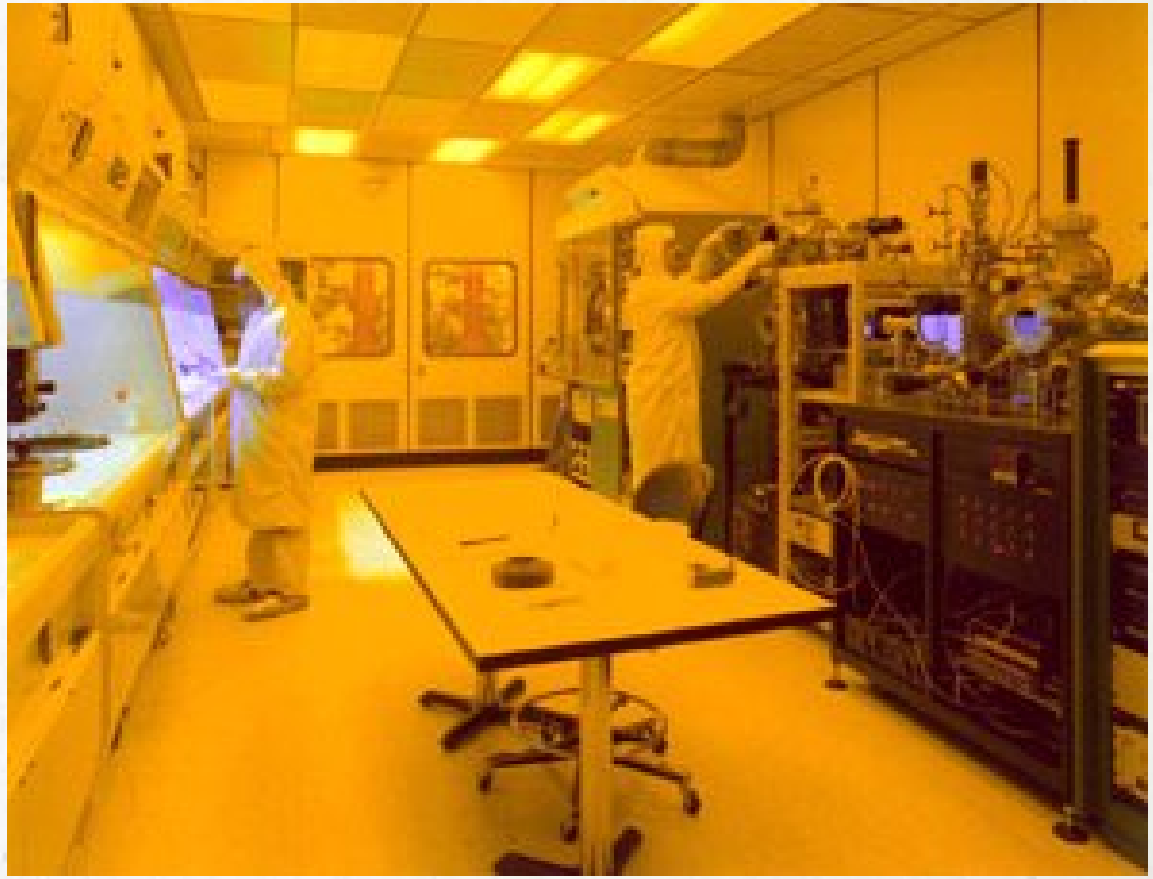
FABRICATION

ALAN HASTINGS

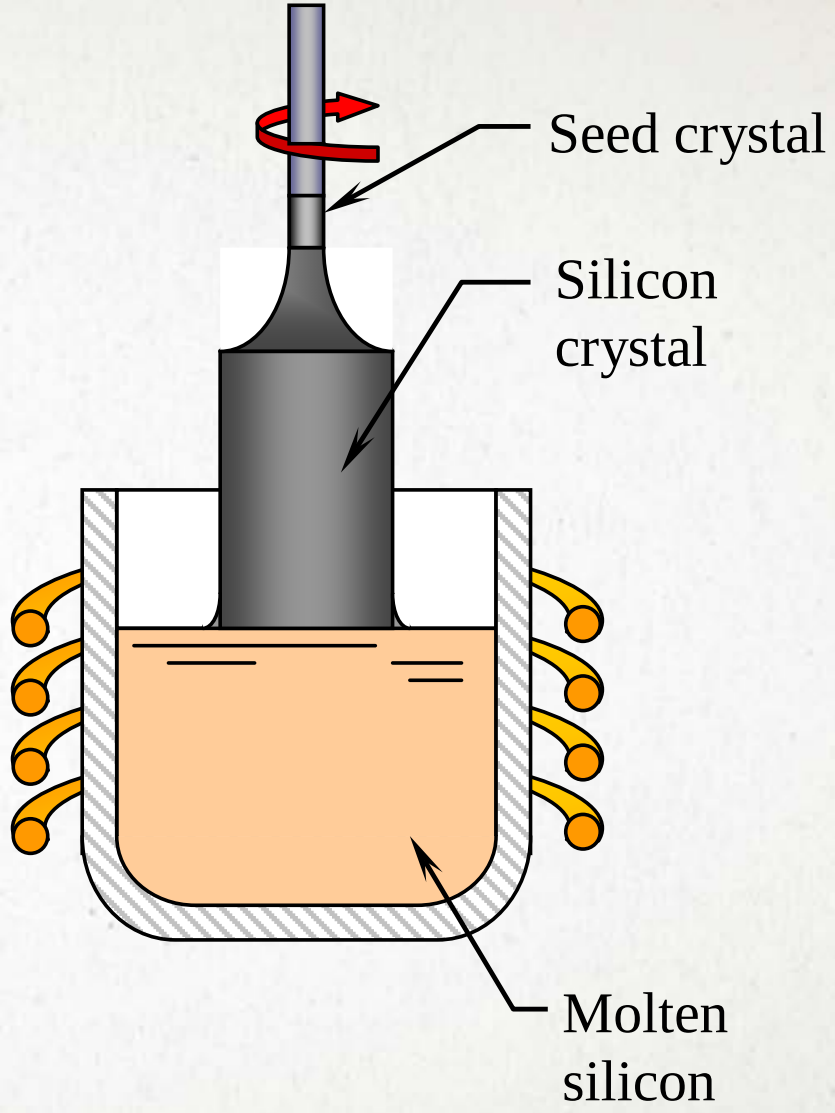
INTRODUCTION

- ***Layout* creates the geometric patterns needed to fabricate an integrated circuit.**
  - We can't understand layout unless we first understand fabrication.
  - Wafer fabs are some of the most sophisticated manufacturing facilities ever created.
  - Luckily, we only need a broad overview of the technologies involved.



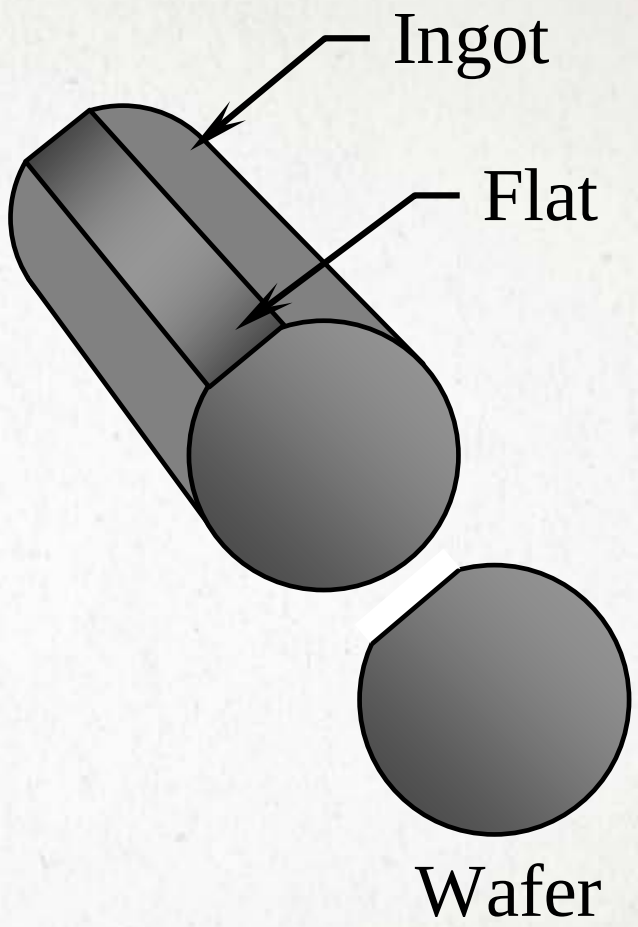
CRYSTAL GROWING

- **The *Czochralski process* transforms purified polycrystalline silicon into monocrystalline silicon.**
  - A silica crucible charged with polysilicon (*poly*) is heated until the silicon melts.
  - A *seed crystal* lowered into the melt provides a surface upon which silicon atoms crystalize.
  - Rotating the silicon crystal and slowly withdrawing it from the crucible creates a cylindrical silicon crystal, or *ingot*.



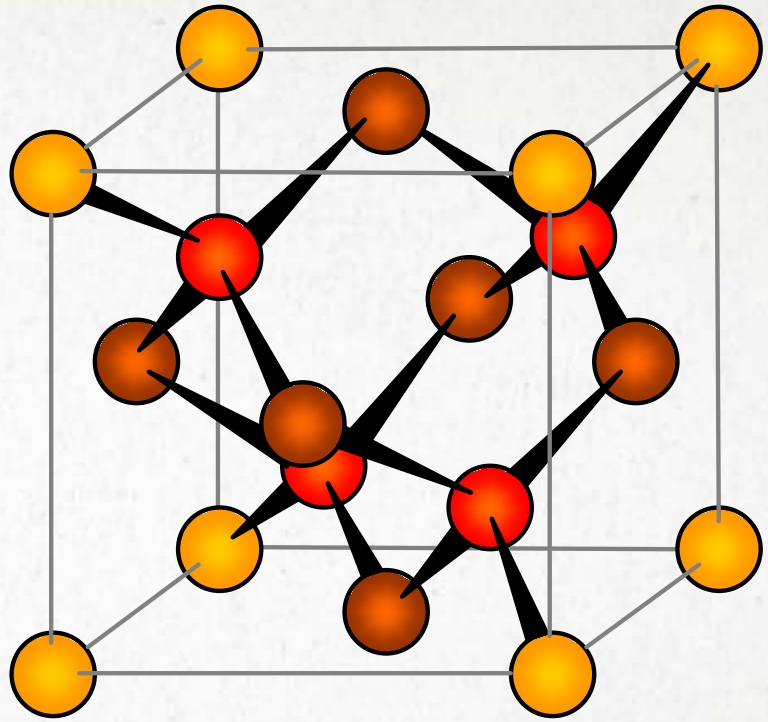
WAFERS

- **The ingot is ground into a cylinder.**
  - A *flat* ground down one side of the cylinder denotes crystal orientation.
- **The ingot is sliced into wafers.**
  - A diamond saw cuts slices called *wafers* from the ingot.
  - Mechanical and chemical polishing produce a nearly perfect surface on each wafer.
  - Each wafer bears a flat indicating the orientation of the crystal axes.



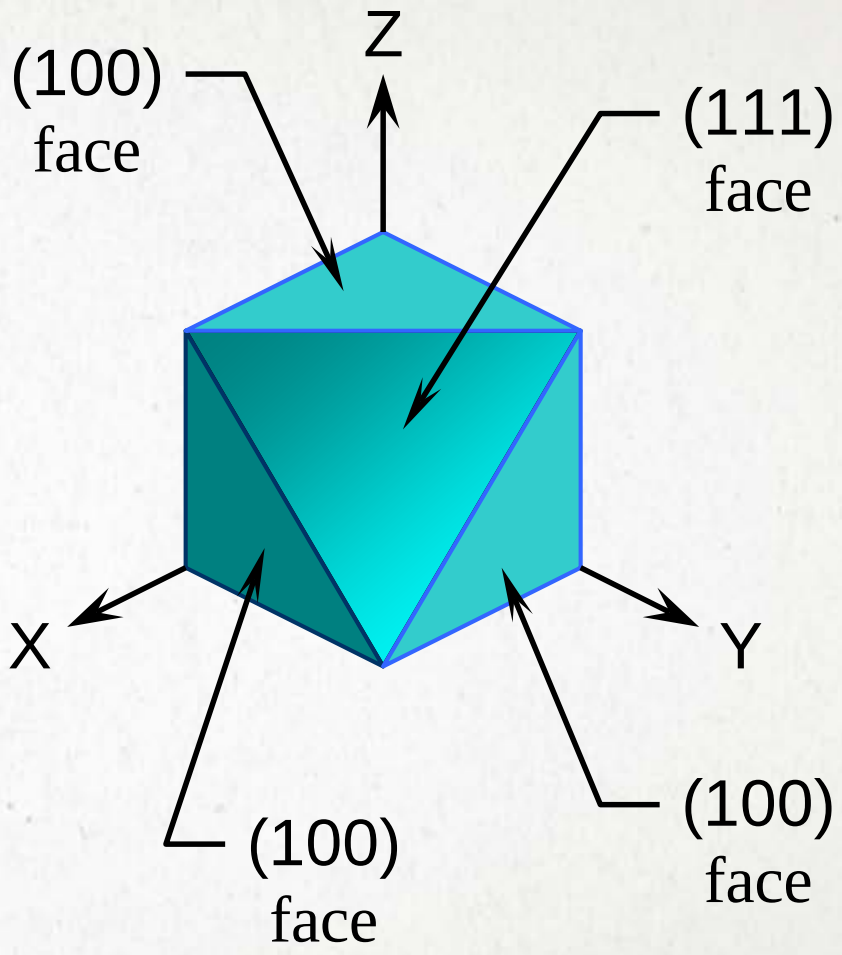
THE CRYSTAL STRUCTURE OF SILICON

- **Silicon crystals display a *face-centered-cubic* (FCC) crystal structure.**
  - Diamond also exhibits a FCC structure.
  - Silicon crystals consist of repetitions of the *unit cell* shown at the right.
  - Planes cut through the crystal in different directions display different properties.



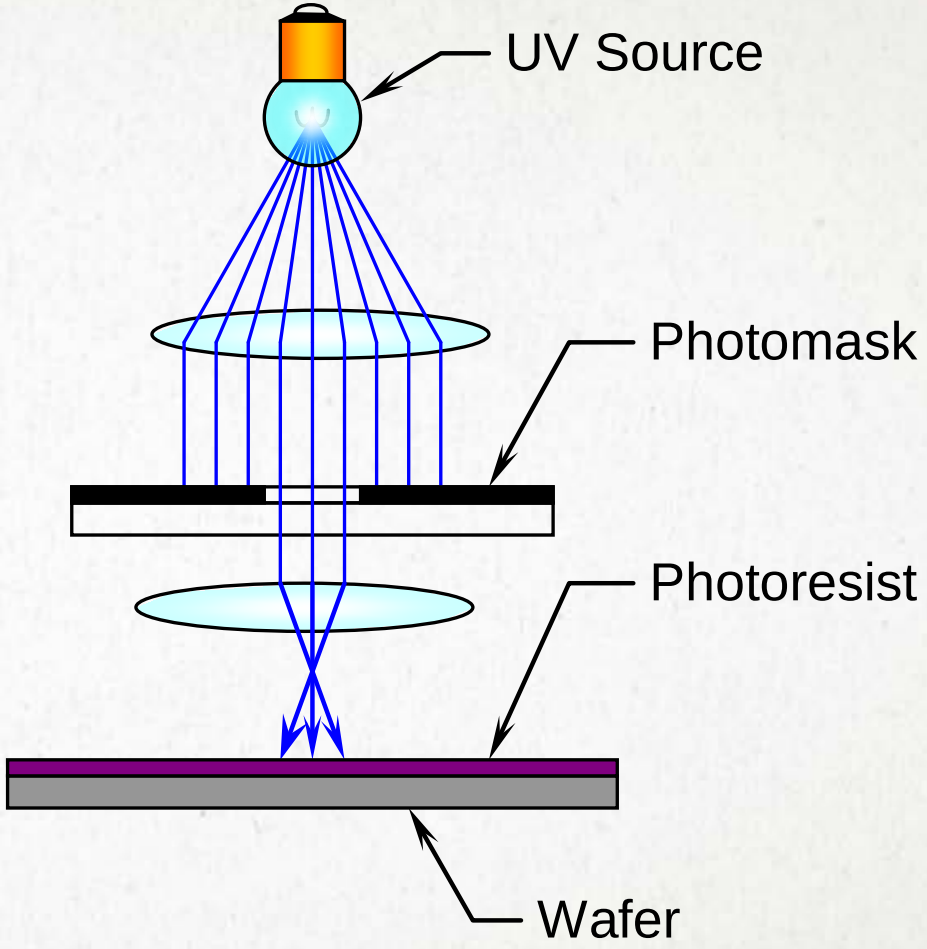
MILLER INDICES

- **The properties of a silicon surface depend upon its orientation with respect to the unit cell.**
  - Each plane surface intersecting the cubic unit cell can be described by a trio of numbers called *Miller indices*.
  - CMOS and BiCMOS processes usually employ wafers whose surfaces are (100) planes.
  - Older processes, such as standard bipolar, employed wafers whose surfaces were (111) planes.



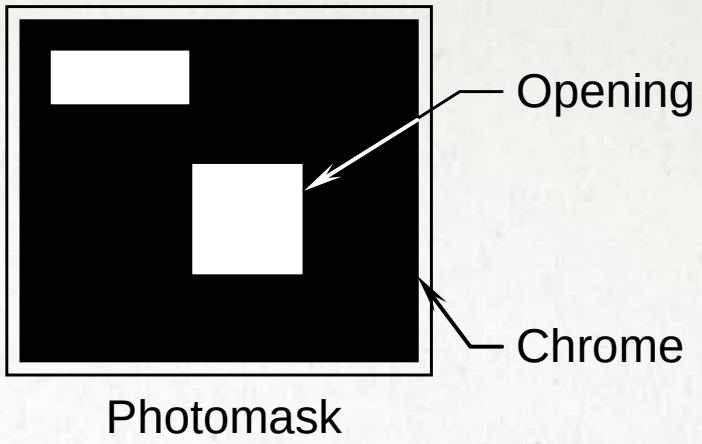
PHOTOLITHOGRAPHY

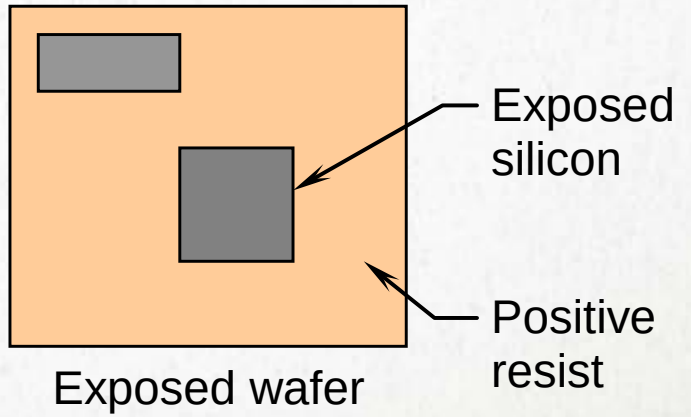
- ***Photolithography* allows selective deposition or removal of materials from the wafer surface.**
  - A thin film of *photoresist* (*resist*) is applied to the wafer.
  - Light shining through openings in a *photomask* (*mask*) exposes portions of the photoresist.
  - Flooding the wafer surface with *developer* creates a pattern of openings in the photoresist.



POSITIVE RESISTS

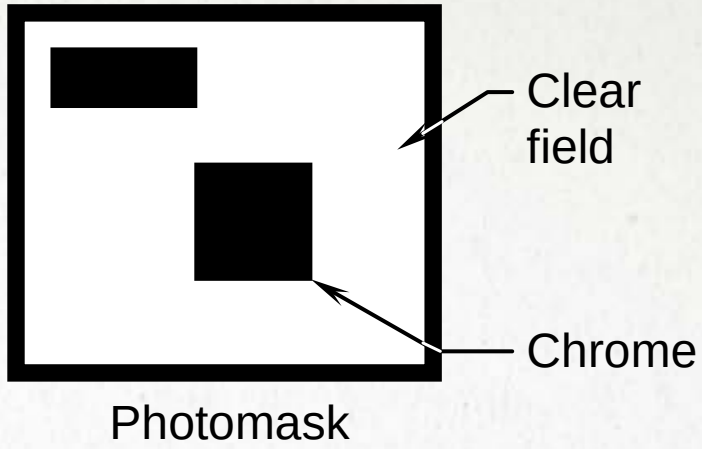
- **A *positive resist* chemically decomposes under UV light.**
  - Exposed areas wash away in the developer.
  - Openings in the mask therefore produce corresponding openings in the resist.
  - Masks for positive resists have a dark field and clear openings; these are called *dark masks*.

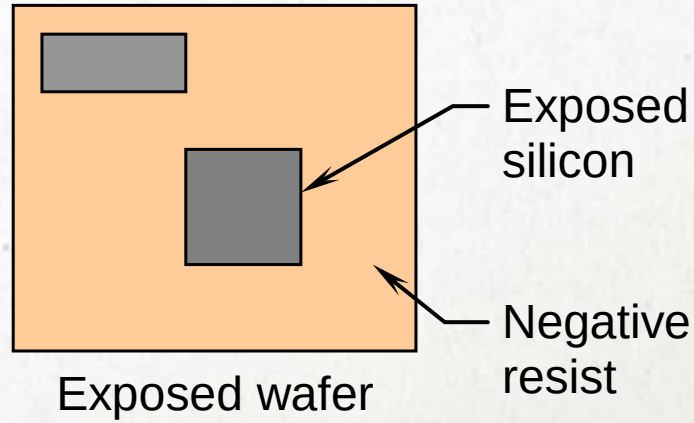




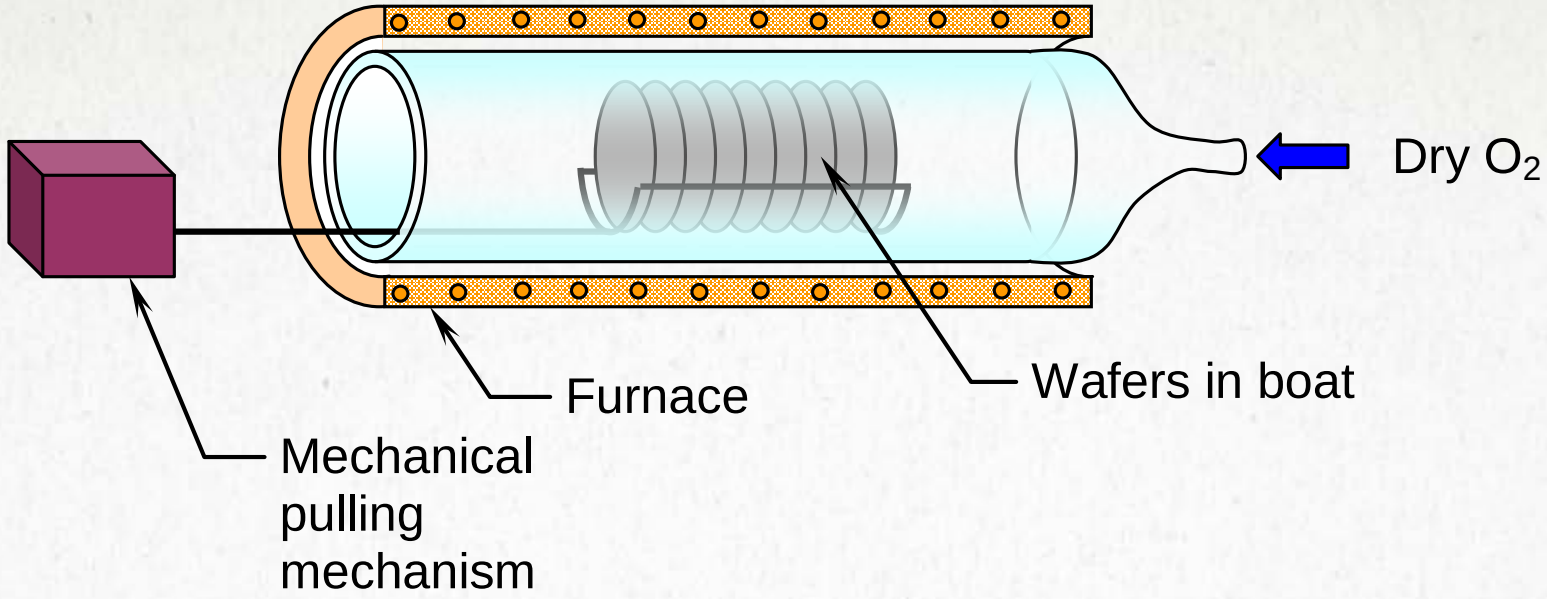
NEGATIVE RESISTS

- **A *negative resist* polymerizes under UV light.**
  - Unexposed areas wash away in the developer.
  - Opaque areas on the photomask therefore produce openings in the photoresist.
  - Masks for negative resists have a clear field and dark figures; these are called *clear masks*.
  - Negative resists tend to swell in development and thus are now seldom used.





OXIDATION



- **Silicon dioxide (*oxide*) is grown by heating the wafers in an oxidizing atmosphere.**
  - Pure dry oxygen produces a slow-growing oxide with minimal defects (a *dry oxide*).
  - Oxygen containing wafer vapor produces a fast-growing oxide with more defects (a *wet oxide*).

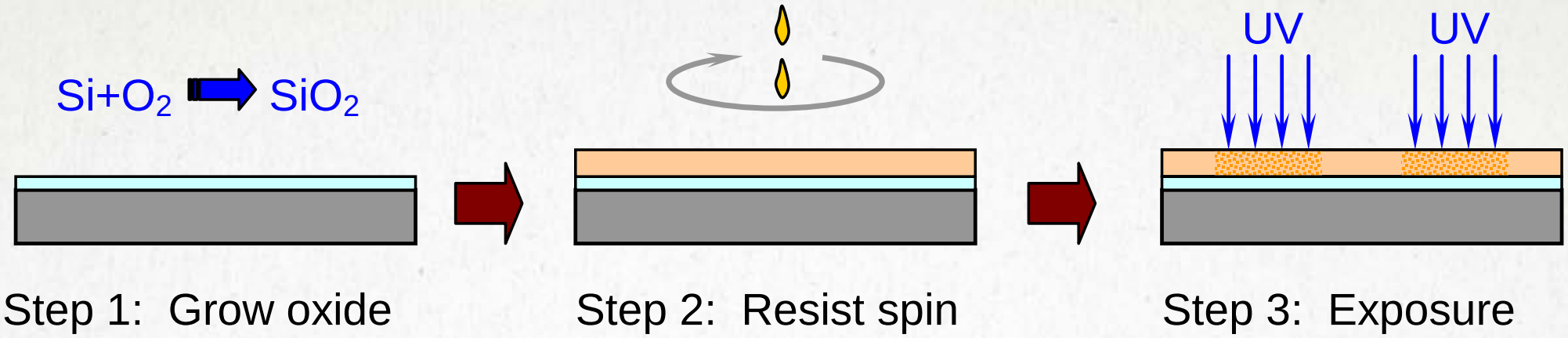
PROPERTIES OF OXIDE

Times Required to Grow 0.1μm (1000Å) of Oxide on (111) Silicon.

Ambient	800°C	900°C	1000°C	1100°C	1200°C
Dry O_2	30 hr	6 hr	1.7 hr	40 min	15 min
Wet O_2	1.7 hr	20 min	6 min		

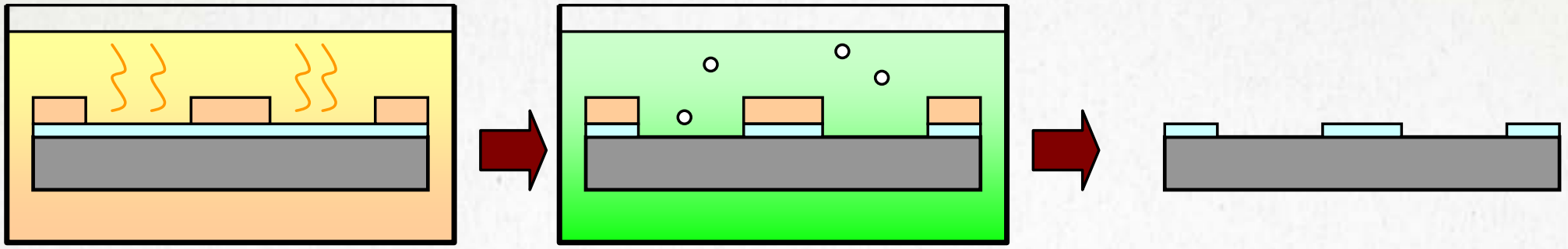
- **Properties of silicon dioxide (oxide):**
  - Readily formed in films as thin as 10 Å (1 nm).
  - Adheres tenaciously to silicon.
  - Resists attack by most chemicals.
  - Readily dissolves by *hydrofluoric acid* (HF).
  - Excellent dielectric for capacitors and MOS transistors.
  - Extremely low defect density when grown on (100) silicon surfaces.

PATTERNING OXIDE



- **Since photoresist cannot withstand high temperatures, a *patterned oxide* film is often used to selectively block dopants during high-temperature diffusion.**
- **A patterned oxide is formed as follows:**
  - **Step1:** Grow a uniform thin film of oxide across the wafer.
  - **Step 2:** Spin photoresist onto the wafer.
  - **Step 3:** photoresist using a suitable photomask.

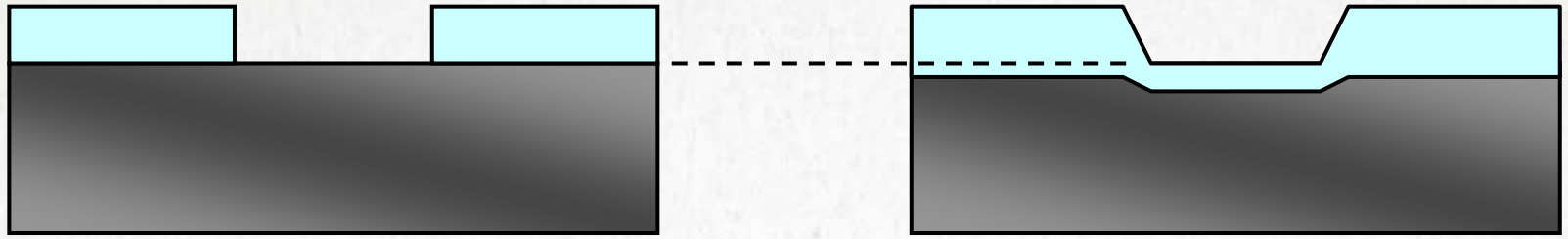
PATTERNING OXIDE (CONTINUED)



Step 4: Develop Step 5: Etch Step 6: Strip resist

- **Step 4:** Develop the photoresist to expose the oxide in the bottom of the oxide removal (OR) windows.
- **Step 5:** Etch the wafer using buffered hydrofluoric acid.
- **Step 6:** Strip the photoresist, leaving the patterned oxide.

EFFECTS OF OXIDATION



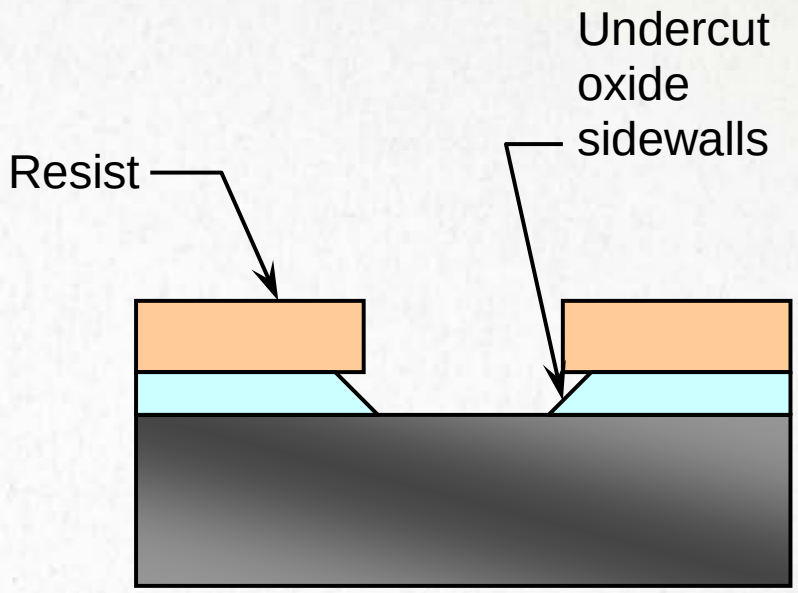
Before oxidation

After oxidation

- **Subsequent oxidation of a previously patterned oxide produces a characteristic *topography*.**
  - Thinner oxides etch more quickly than thick ones.
  - The oxide openings therefore oxidize more quickly than surrounding field oxide.
  - Thermal oxidation consumes silicon, so the silicon surface erodes by about 45% of the oxide thickness grown.

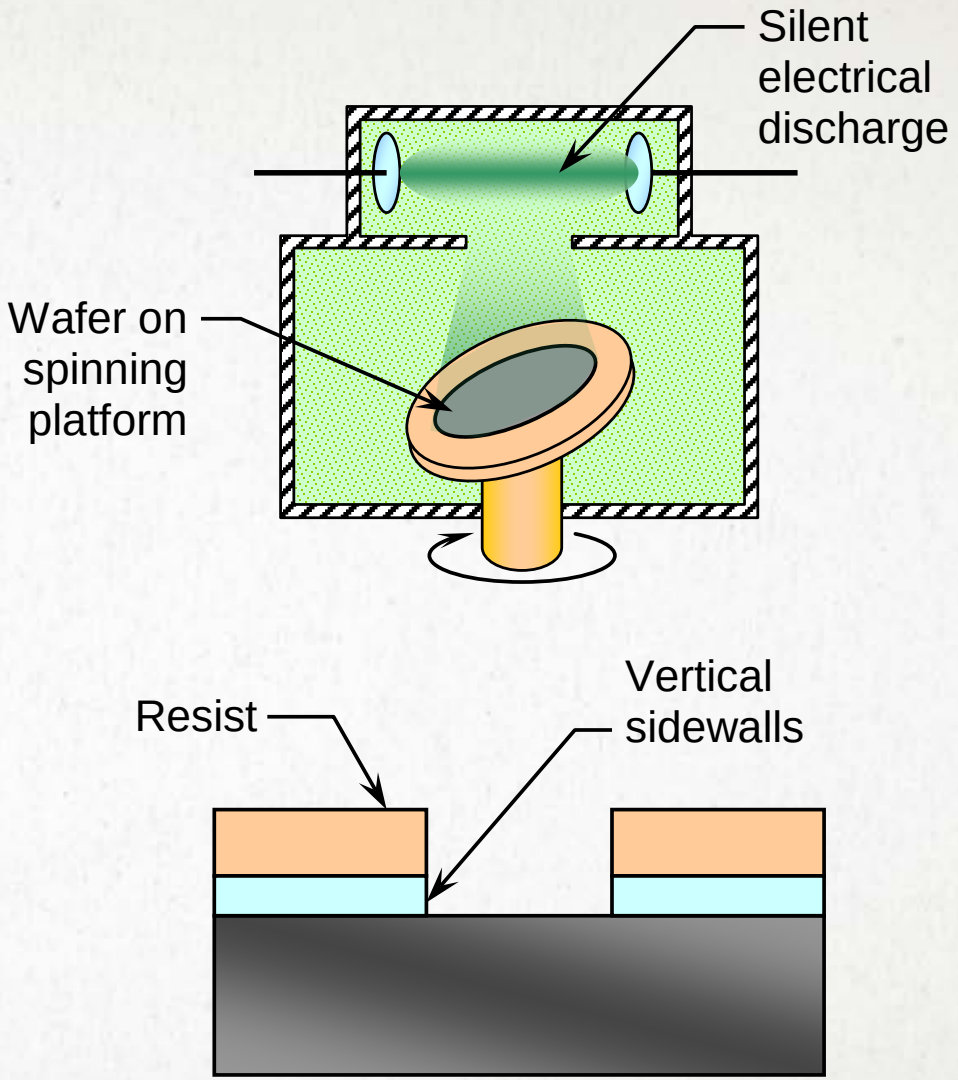
WET ETCHING

- **A *wet etch* uses buffered hydrofluoric acid to dissolve oxide.**
  - The acid attacks all exposed oxide surfaces at the same rate: it is an *isotropic* etchant.
  - The sidewall surfaces exposed by etching erode laterally underneath the photoresist.
  - This *undercutting* interferes with production of fine linewidth features.

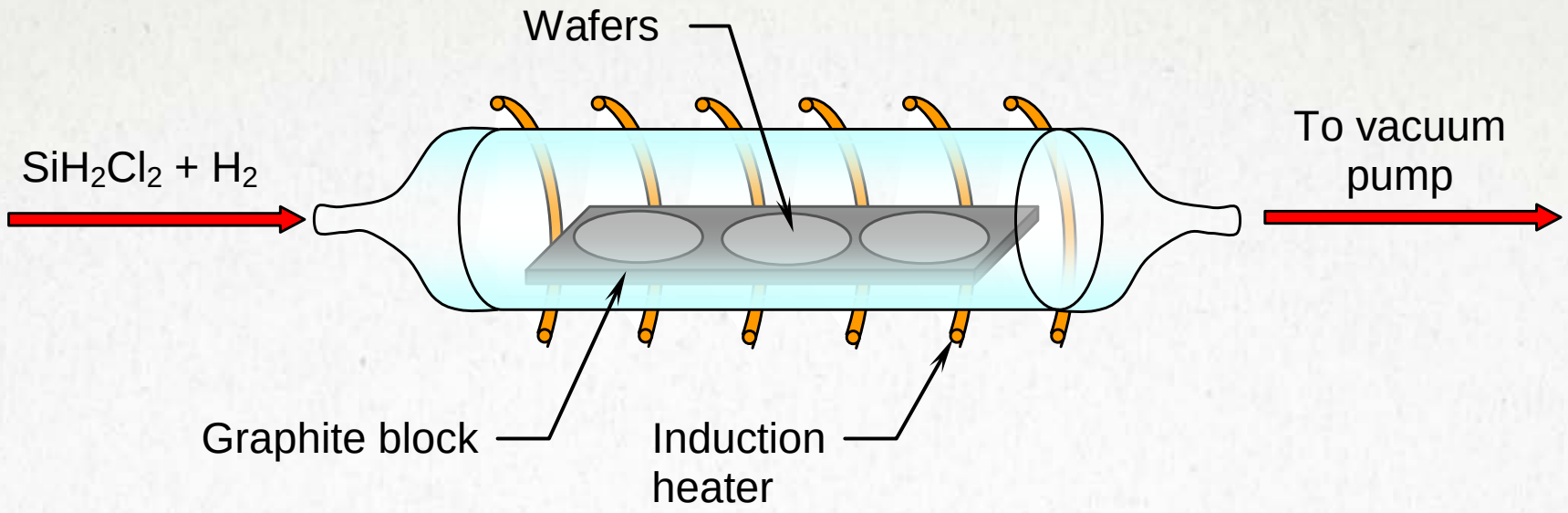


DRY ETCHING

- ***Dry etching* uses a plasma to selectively remove oxide.**
  - Passing a silent electrical discharge through a fluorocarbon gas generates reactive fluorine ions.
  - These ions spray downwards on the wafer, *anisotropically* etching the exposed oxide.
  - Dry etching is now almost universally used for etching fine linewidth features.

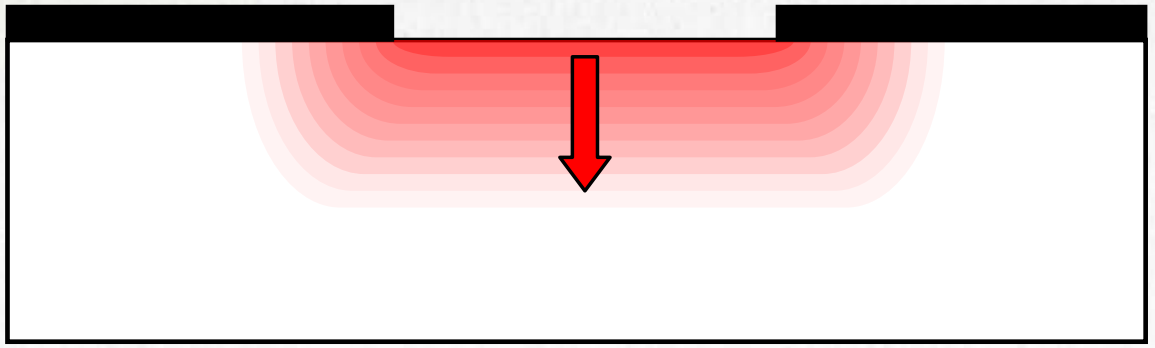


EPITAXY



- **A process called *epitaxial deposition* (*epitaxy*) can deposit monocrystalline silicon on a wafer.**
  - A mixture of dichlorosilane and hydrogen passed over a heated silicon wafer decomposes, depositing silicon.
  - The crystal structure of the *epitaxial layer* (*epi*) exactly duplicates that of the underlying silicon.
  - The same process can also deposit polycrystalline silicon (*poly*) on oxide.

DIFFUSION



- **Dopant atoms become mobile and can *diffuse* through silicon at sufficiently high temperatures.**
  - Diffusion of dopants from an external source through oxide openings allows selective doping of specific regions.
  - Diffusing dopants move downwards and spread outwards from the oxide opening.

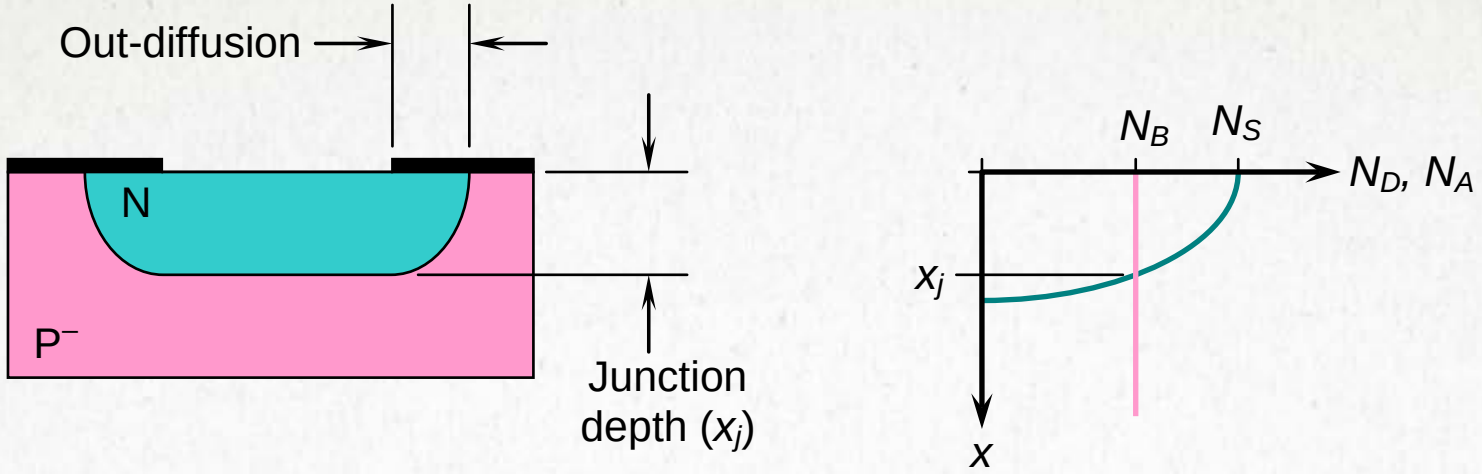
DIFFUSION (CONTINUED)

Representative Junction Depths, in Microns
(10^{20} cm^{-3} source, 10^{16} cm^{-3} background, 15min deposition, 1hr drive)

Dopant	950°C	1000°C	1100°C	1200°C
Boron	0.9μm	1.5μm	3.6μm	7.3μm
Phosphorus		0.5	1.6	4.6
Antimony			0.8	2.1
Arsenic			0.7	2.0

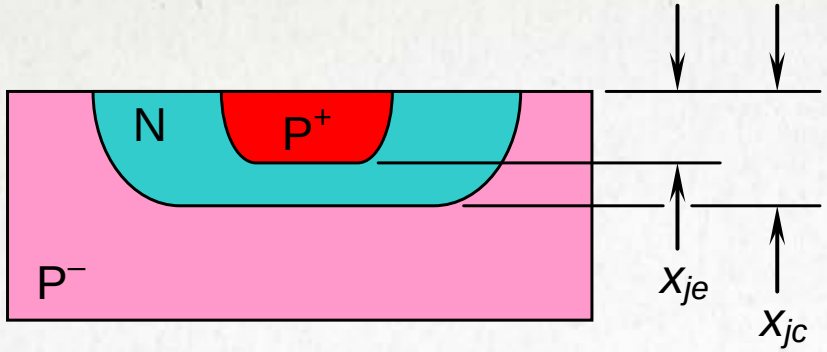
- **Dopants diffuse at different rates.**
  - *Boron* and *phosphorus* diffuse relatively quickly.
  - *Arsenic* and *antimony* diffuse much more slowly.
  - All dopants diffuse more rapidly at higher temperatures.

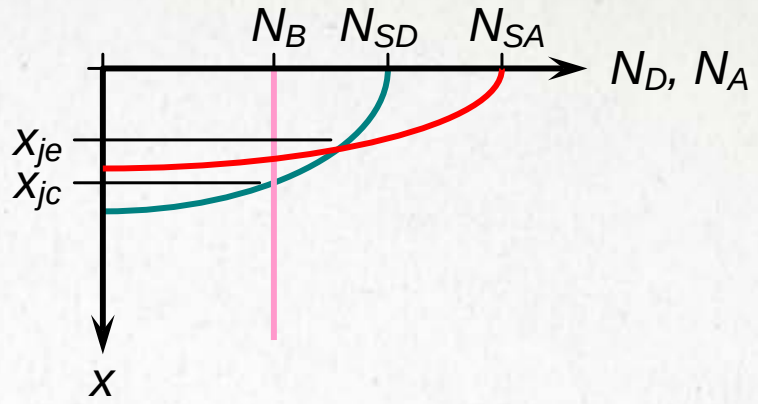
PLANAR DIFFUSIONS



- **A diffusion driven in from the surface of the wafer is called a *planar diffusion*.**
  - Planar diffusions are usually patterned by oxide windows.
  - Dopants diffuse laterally (*outdiffusion*) underneath the edges of the oxide window to about 80% of the junction depth.

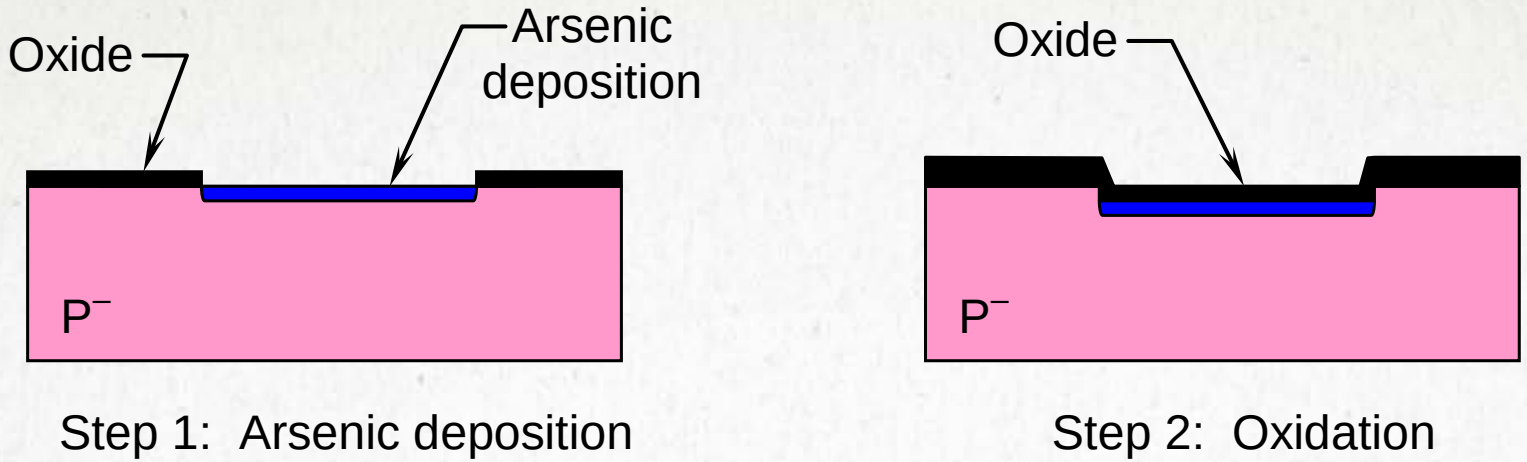
COUNTERDOPING





- **Multiple diffusions can be driven into one another to create complicated structures.**
  - Each diffusion typically has a higher surface concentration than the previous one.
  - Each successive diffusion *counterdopes* the silicon to form a succession of regions of alternating doping polarity.

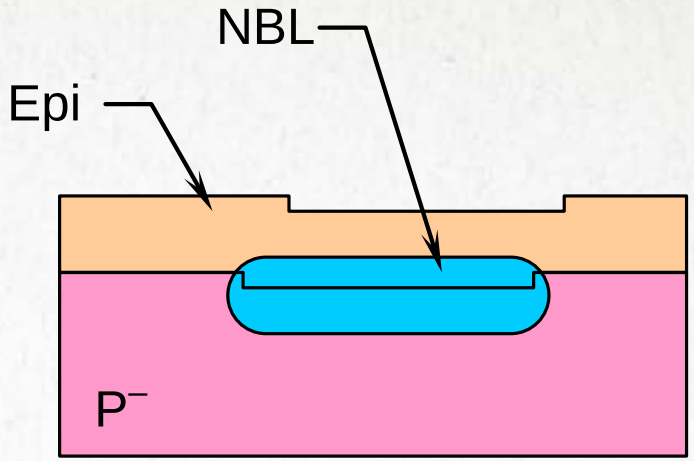
BURIED LAYERS



- **A *buried layer* is a diffusion surrounded on all sides by silicon.**
- **The traditional process for making a buried layer proceeds as follows:**
  - **Step 1:** Open oxide windows were the buried layer will form. Deposit a suitable dopant in these windows.
  - **Step 2:** Drive the dopant into the silicon while oxidizing the silicon surface.

BURIED LAYERS (CONTINUED)

Step 3: Oxide removal

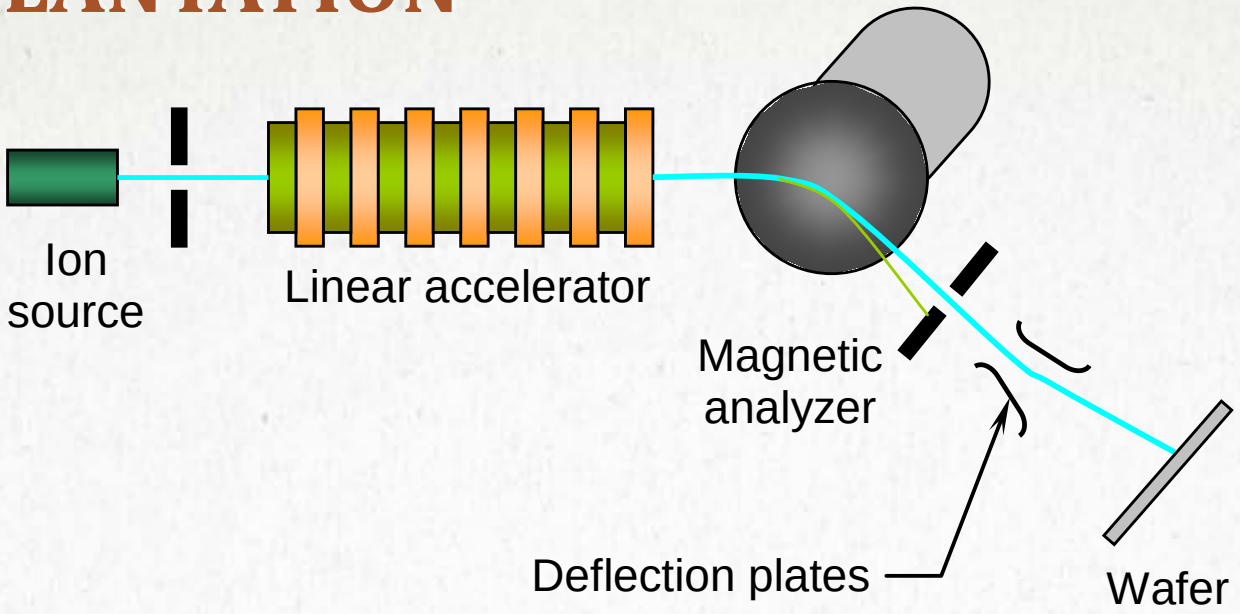


Step 4: Epitaxial deposition

- **Step 3:** Stripping the oxide leaves a surface discontinuity used for subsequent mask alignment.
- **Step 4:** Epitaxially deposit silicon above the buried layer.

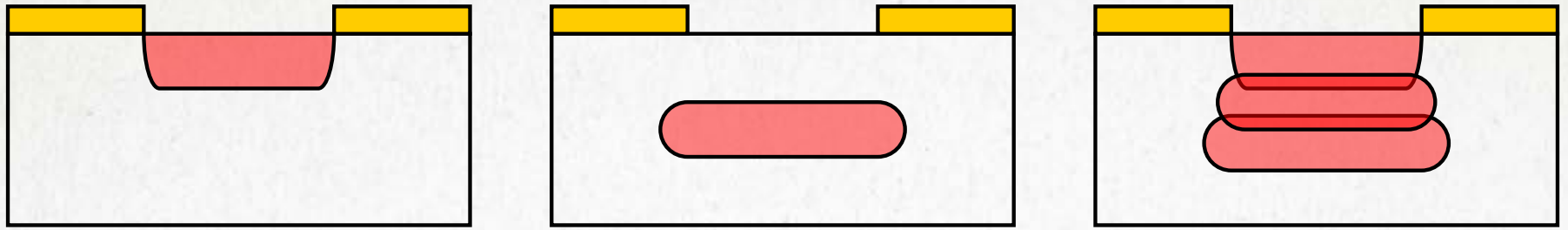
- **The surface discontinuity propagates upwards at an angle during epitaxy, displacing the continuity (*pattern shift*).**

ION IMPLANTATION



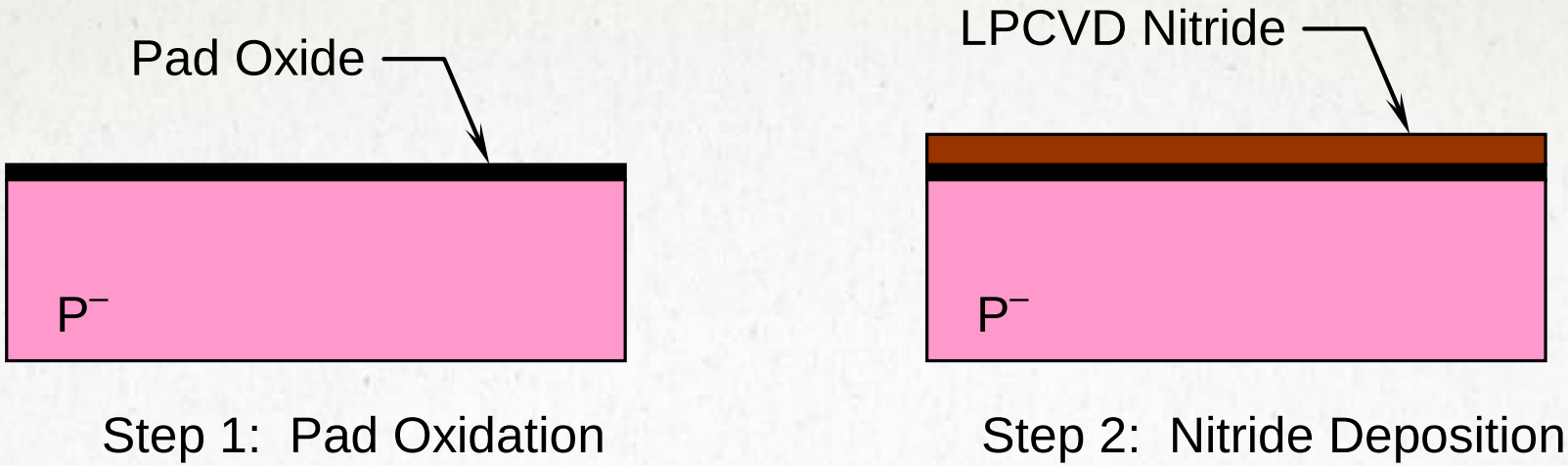
- **An alternative method of doping silicon involves using a particle accelerator to drive dopant atoms into the lattice.**
  - The particle accelerators are low-energy (seldom more than a few MeV) and high beam current (mA) machines called *ion implanters.*
  - *Ion implantation* can deposit extremely precise dosages of dopants.

ION IMPLANTATION (CONTINUED)



- **Ion implantation can form implants at different depths.**
  - *Low-energy implants* deposit dopants near the surface.
  - *High-energy implants* deposit dopants up to a micron or two beneath the surface.
- **Ion implantation disrupts the silicon lattice. Heating repairs (*anneals*) the damage.**

LOCAL OXIDATION OF SILICON

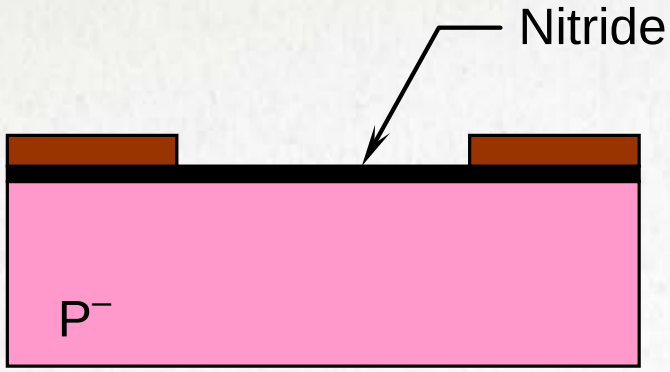


Step 1: Pad Oxidation

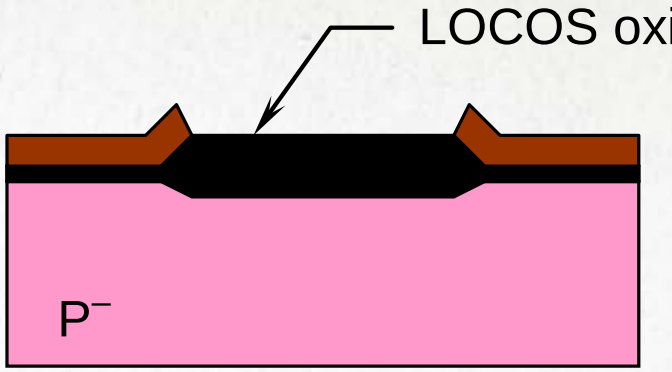
Step 2: Nitride Deposition

- **CMOS processes need a thick *field oxide* to cover the inactive regions of the die.**
- **A technique called *local oxidation of silicon* (LOCOS) can form thick patterned oxide layers as follows:**
  - **Step 1:** A thin *pad oxide* is grown across the silicon to protect it from mechanical stress.
  - Step 2: *Silicon nitride* (*nitride*) is deposited atop the pad oxide.

LOCAL OXIDATION OF SILICON (CONTINUED)

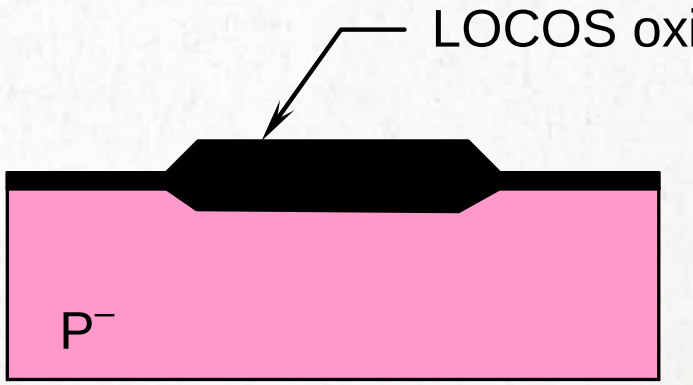


Step 3: Nitride Deposition



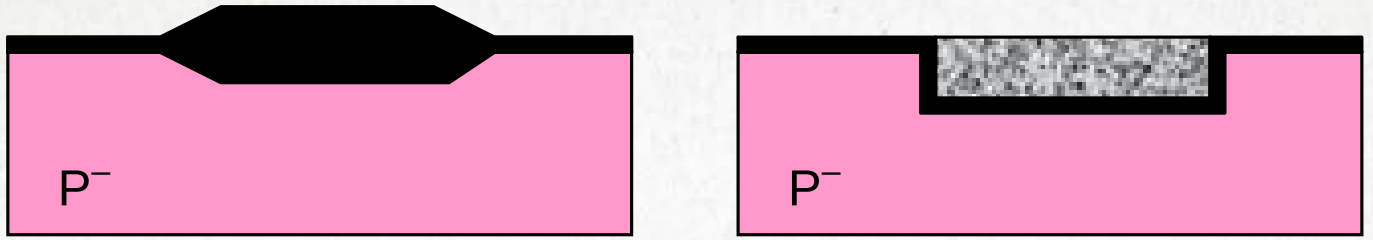
Step 4: LOCOS oxidation

- **Step 3:** A patterned etch opens windows in the nitride.
- **Step 4:** A wet oxidation grows oxide in the nitride windows.
- **Step 5:** Stripping the nitride away leaves the patterned oxide.



Step 5: Nitride Strip

SHALLOW TRENCH ISOLATION

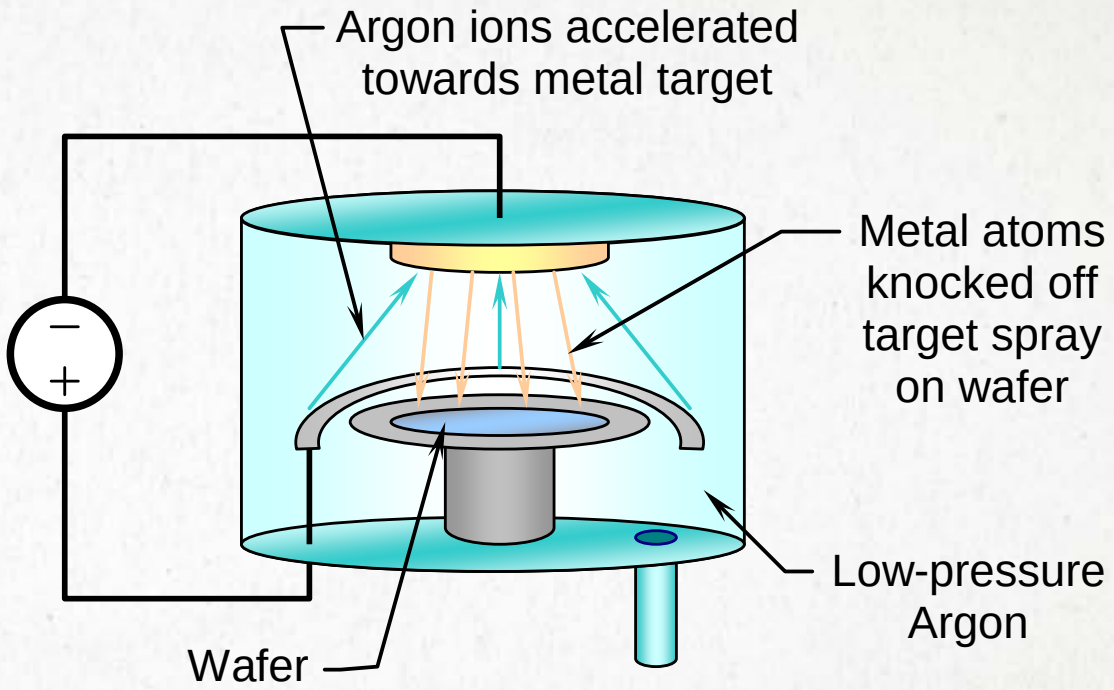


LOCOS STI

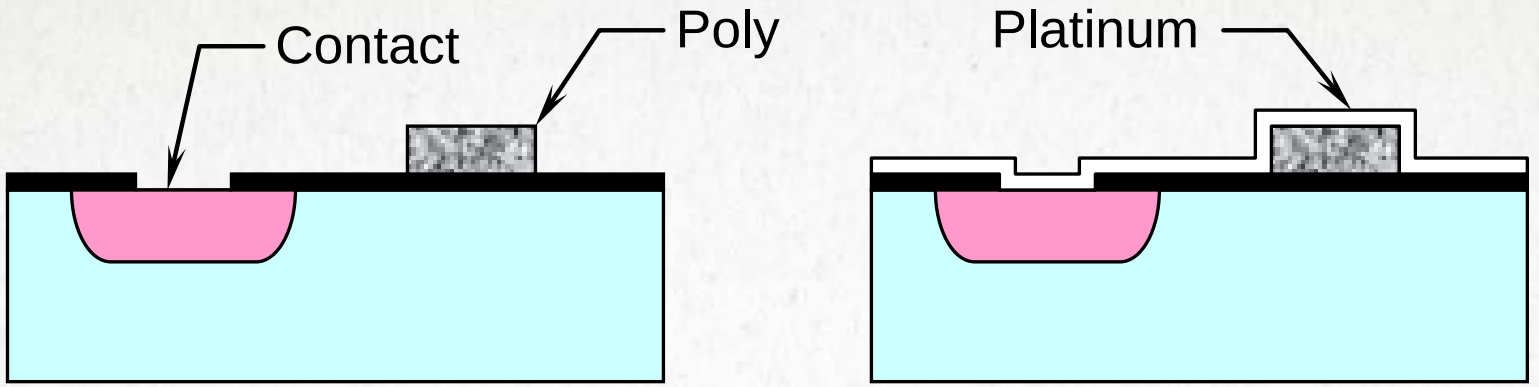
- **LOCOS cannot support deep-submicron processes because:**
  - Most forms of LOCOS generate significant topography.
  - The transition region between LOCOS and thin oxide (*bird's beak*) increases minimum spacings.
- ***Shallow trench isolation* (STI) eliminates these problems:**
  - The anisotropically etched trench has nearly vertical sidewalls.
  - The trench is filled with oxide and then polished back to the surface.

SPUTTERING ALUMINUM

- **Aluminum is often chosen to form metallic interconnects.**
  - Aluminum is highly conductive, adheres well to oxide, and (unlike copper) can be dry etched.
  - Addition of about 0.5% copper improves *electromigration* resistance.
- **Aluminum alloys are deposited by *sputtering*.**
- This process uses inert ions to knock metal atoms off a target.
- The sputtered atoms redeposit on nearby surfaces.



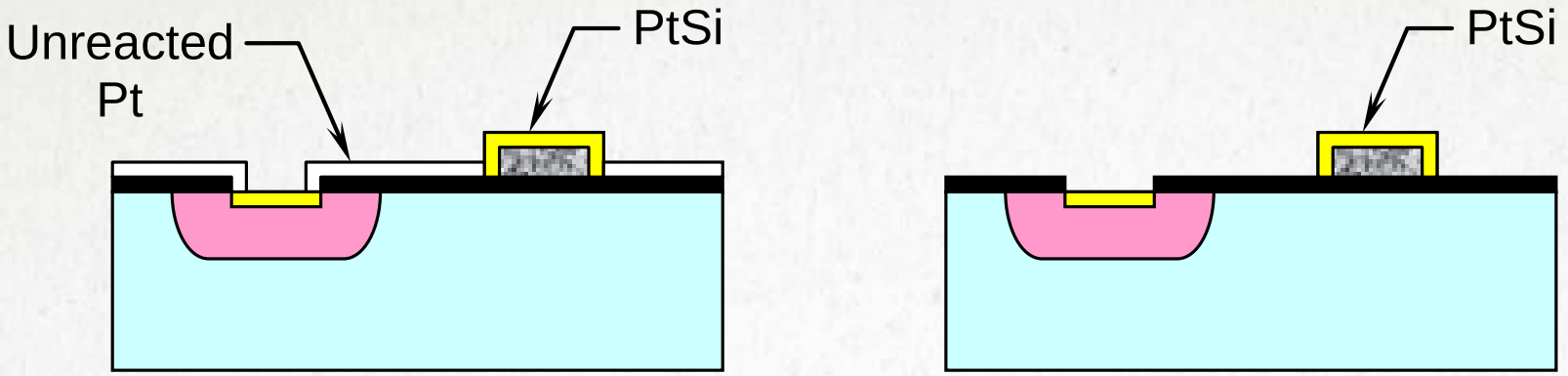
SILICIDATION



Step 1: Contact OR Step 2: Platinum Deposition

- ***Silicides*** **are definite compounds of silicon and various metals.**
  - Silicides are used to reduce contact resistance to silicon and to reduce the resistance of polysilicon traces.
- **To deposit a silicide layer,**
  - **Step 1:** Expose silicon or poly surfaces to be silicided.
  - **Step 2:** Sputter an appropriate metal (such as platinum) onto the wafer.

SILICIDATION (CONTINUED)



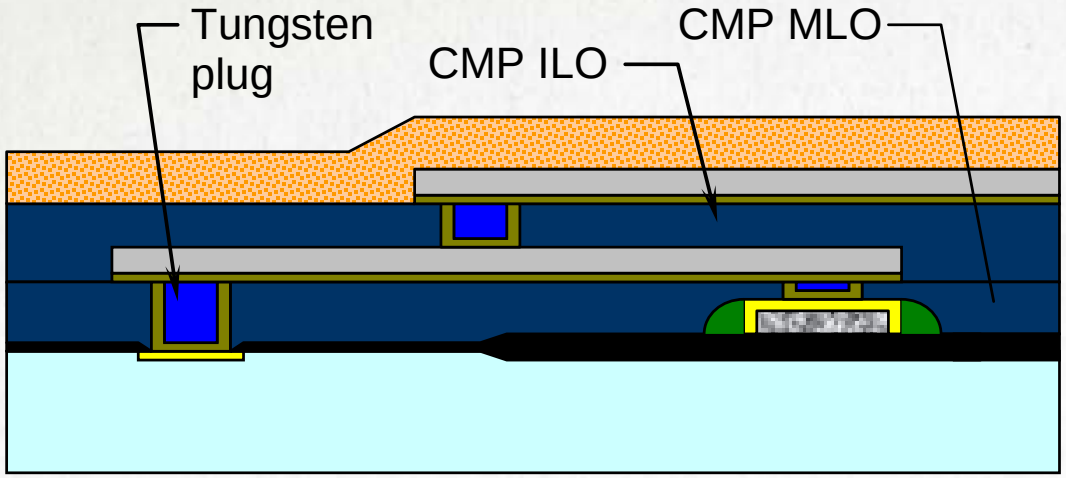
Step 3: Sinter

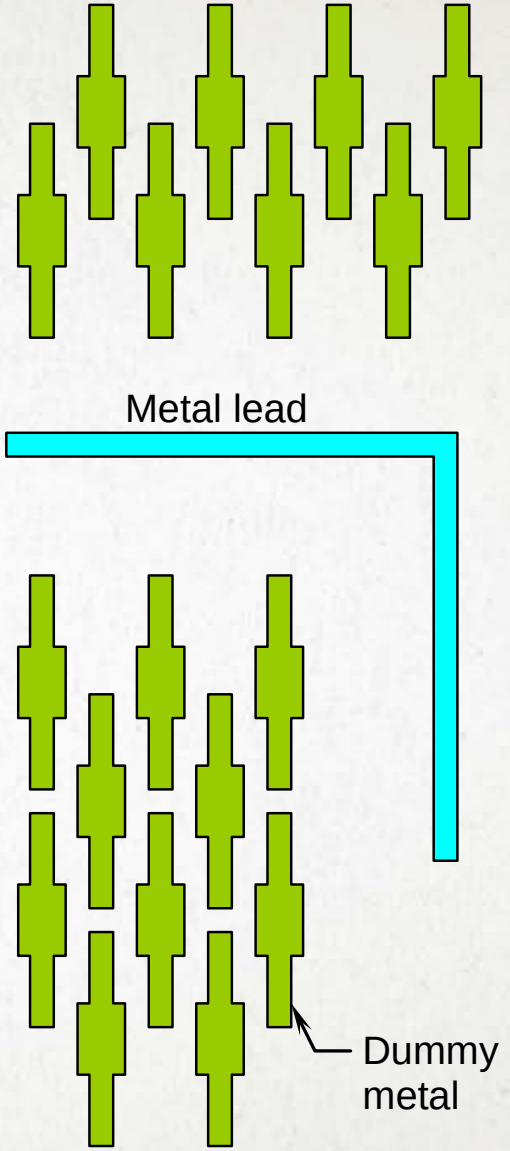
Step 4: Aqua Regia Etch

- **Step 3:** Heating the wafer causes silicon to react with the metal.
- **Step 4:** Remove unreacted metal with an appropriate etchant.

- **Popular choices for silicides include:**
  - *Platinum and palladium silicides:* Form Schottky diodes.
  - *Titanium silicide:* Forms *self-aligned silicide* (*salicide*) on poly.
  - *Nickel and cobalt silicides:* Suitable for fine-linewidth salicidation.

CHEMICAL MECHANICAL POLISHING





- ***Chemical-mechanical polishing* (CMP) is used to *planarize* the wafer to accommodate the narrow depth of field of modern photolithography.**
  - CMP requires approximately the same density of metal everywhere to prevent *dishing*.
  - Areas with insufficient density require addition of *dummy metal,* a pattern of metal figures that do not serve any electrical function.

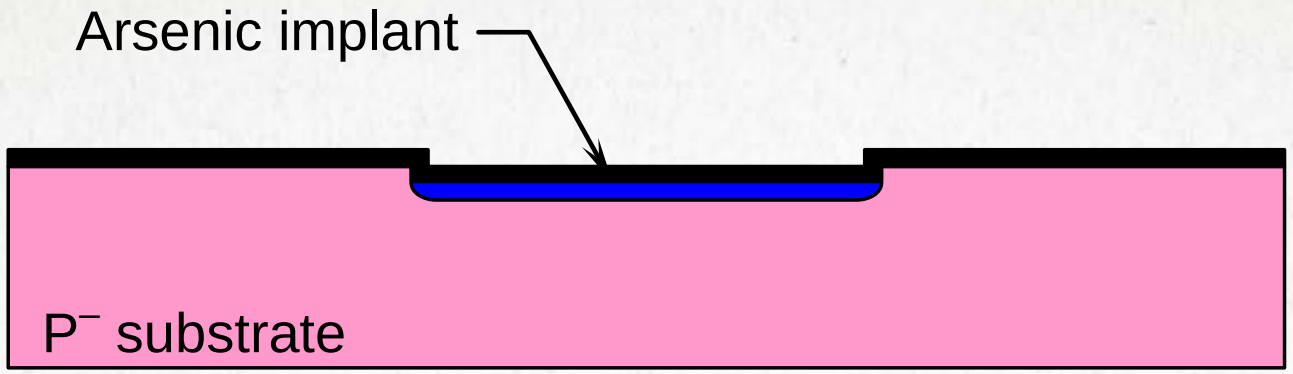
PROCESSES

ALAN HASTINGS

STANDARD BIPOLAR

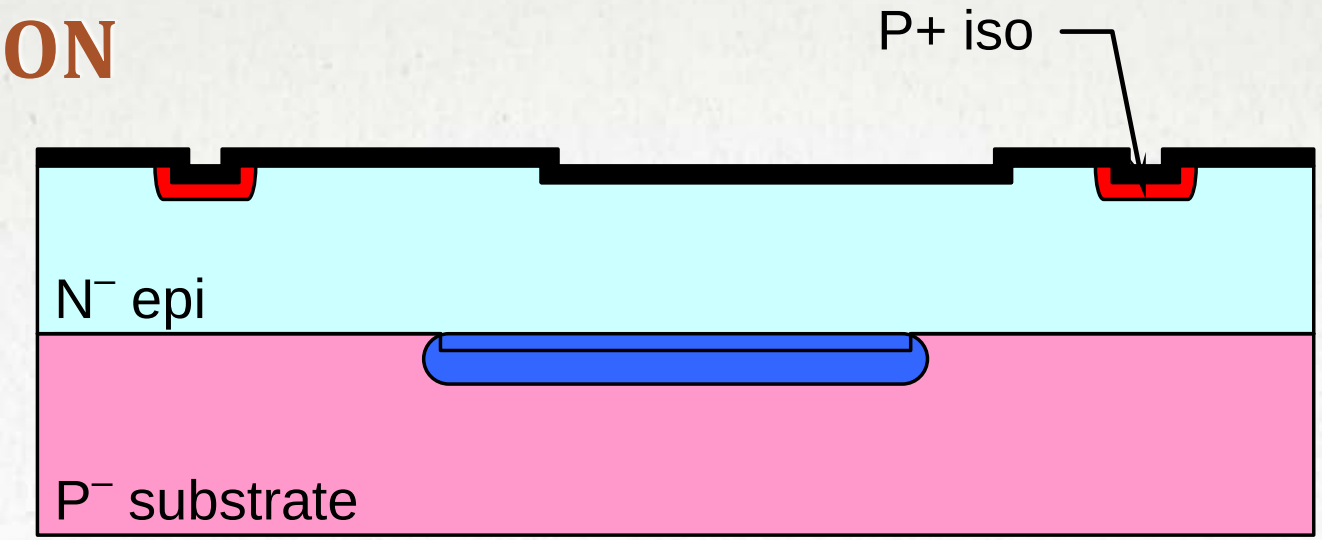
- **Standard bipolar was the first widely used analog integrated circuit process.**
  - Although new designs are seldom implemented in this process, many standard bipolar devices are still manufactured today.
- **Standard bipolar uses many features that are now common components of other processes:**
  - Epitaxy
  - Buried layers
  - Isolation tanks
  - Vertical bipolar transistors

NBL



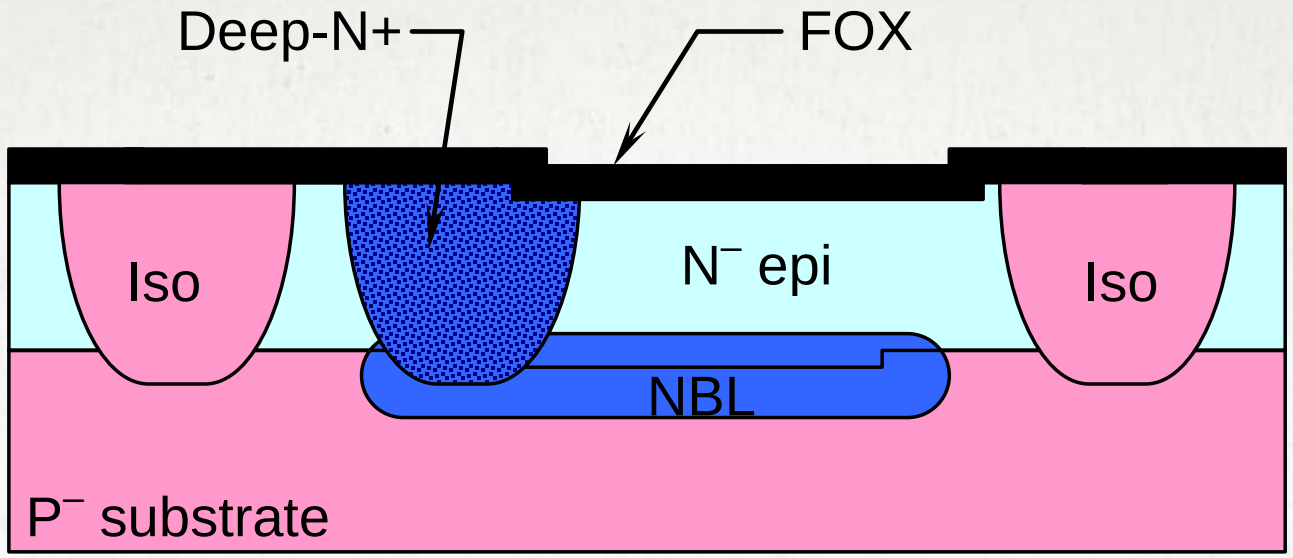
- **Deposit N-type buried layer (NBL):**
  - Thermal oxidation of P– substrate.
  - Pattern using NBL (mask #1).
  - Remove oxide over NBL regions.
  - Implant antimony to form NBL.
  - Anneal NBL implant while growing thermal oxide.

ISOLATION



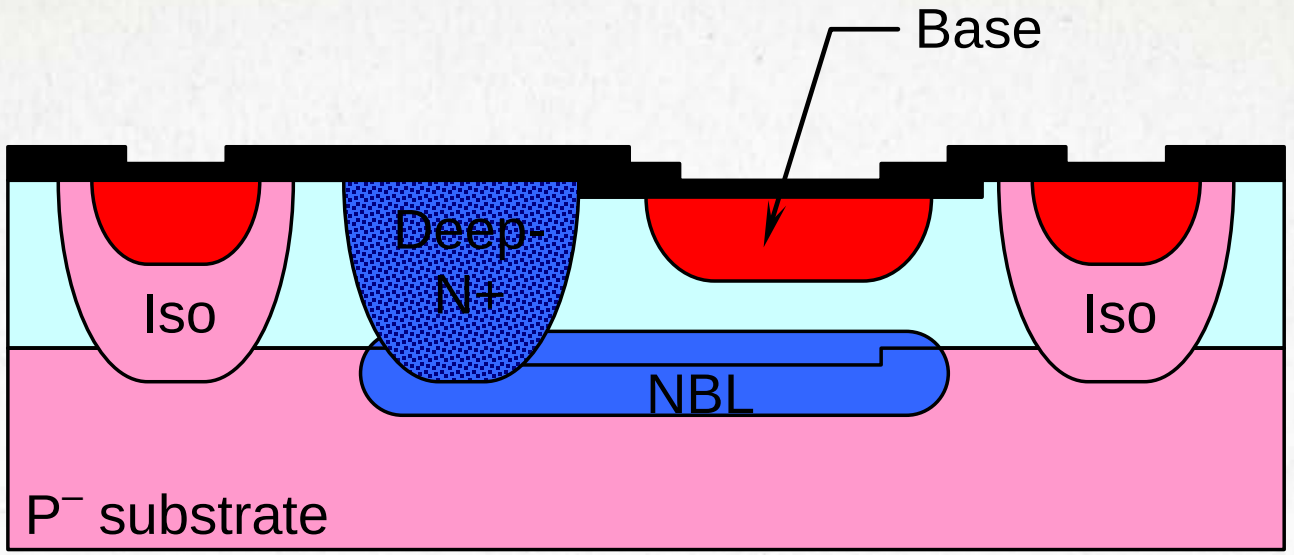
- **Grow epitaxial layer (epi) and deposit isolation:**
  - Grow N– epitaxial layer.
  - NBL shadow propagates upwards diagonally to surface, forming alignment mark.
  - Grow thermal oxide.
  - Pattern using ISO (mask #2).
  - Remove oxide over isolation regions.
  - Deposit or implant boron into isolation regions.
  - Drive isolation partway down while growing thermal oxide.

DEEP-N+



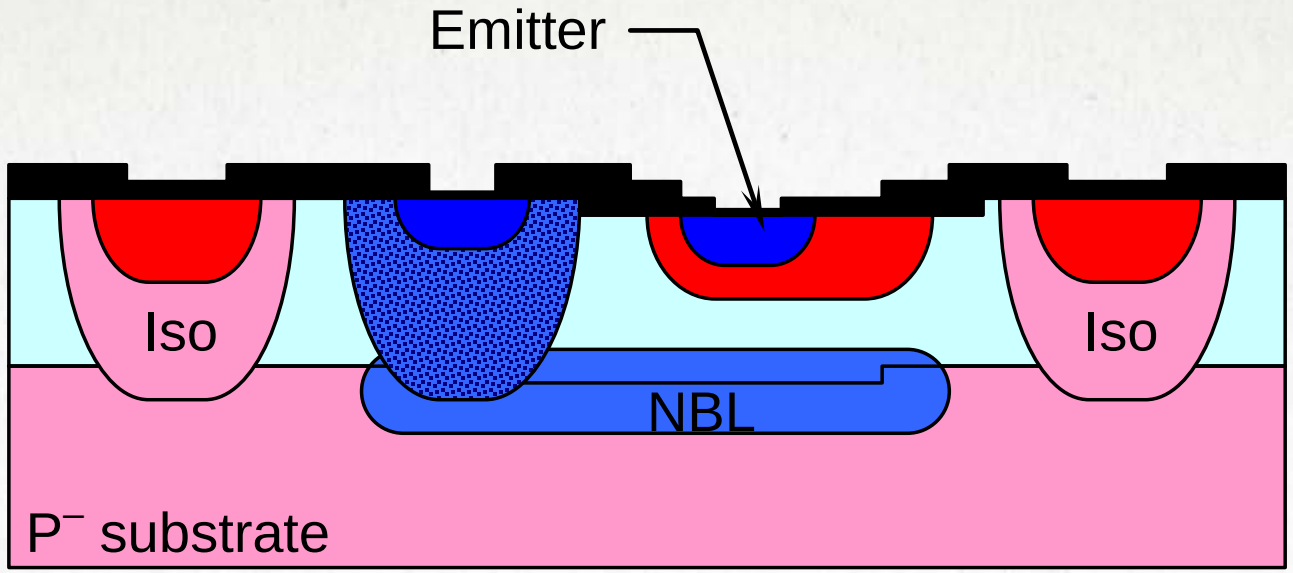
- **Form deep-N+ sinkers:**
  - Pattern using DEEPN (mask #3).
  - Open oxide over DEEPN regions.
  - Deposit phosphorus.
  - Drive deep-N+ down to meet NBL.
  - This step also completes the isolation drive.
  - Wet oxide grown during the deep-N+ drive forms the field oxide (FOX).

BASE



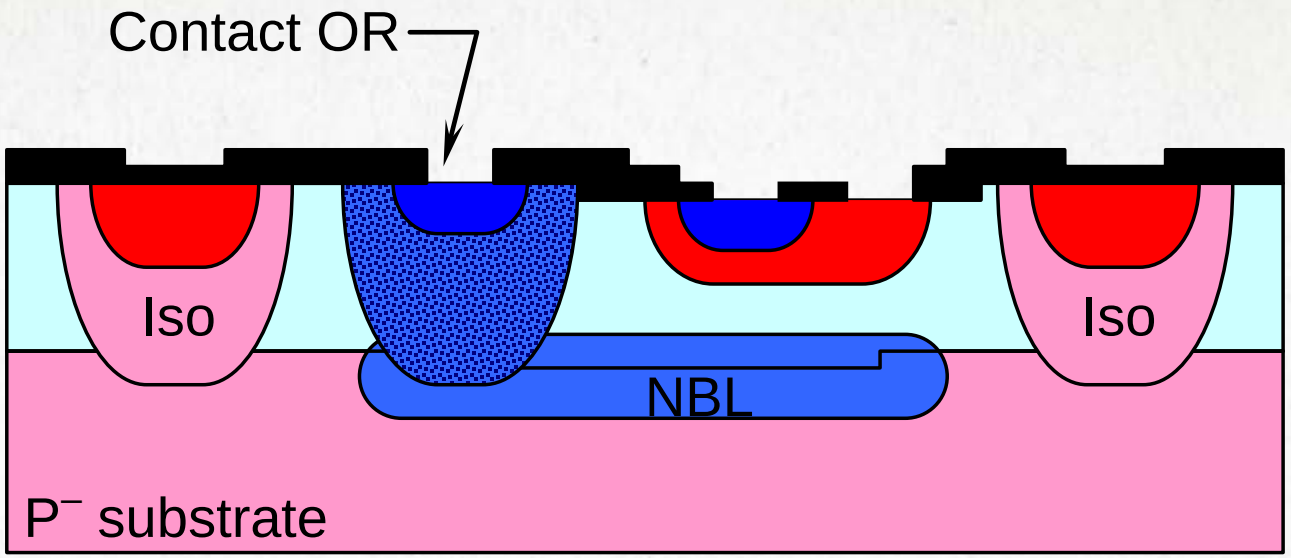
- **Implant base:**
  - Pattern using BASE (mask #4).
  - Remove oxide over base regions.
  - Implant boron.
  - Drive base down while simultaneously growing thermal oxide.

EMITTER



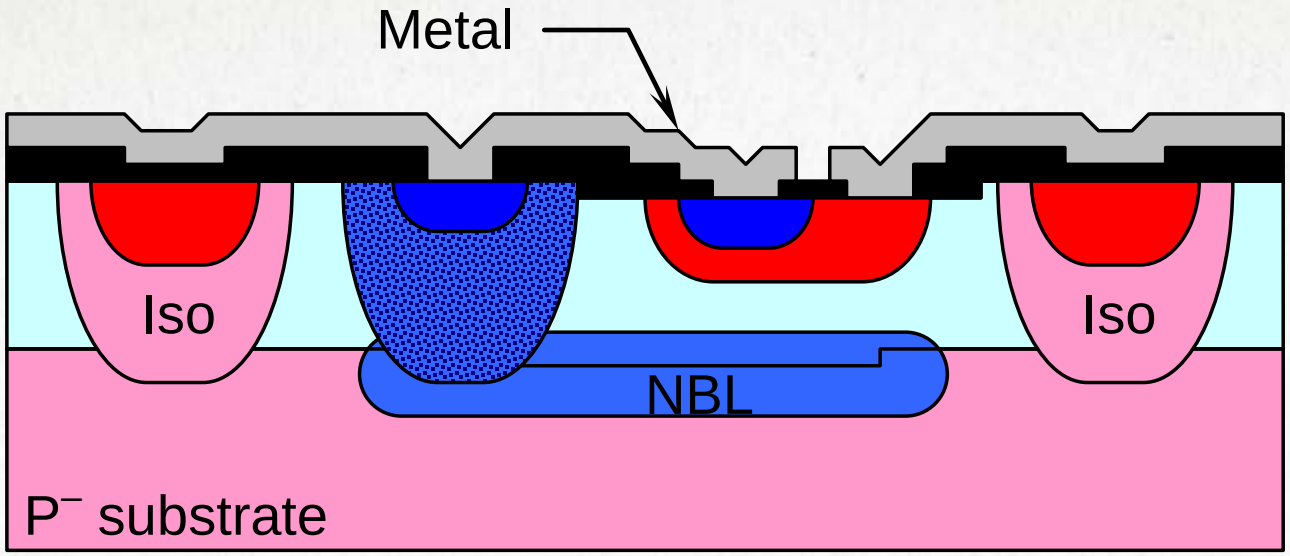
- **Implant emitter:**
  - Pattern using EMIT (mask #5).
  - Deposit or implant phosphorus to form emitter.
  - Drive emitter down while simultaneously growing thermal oxide.

CONTACT



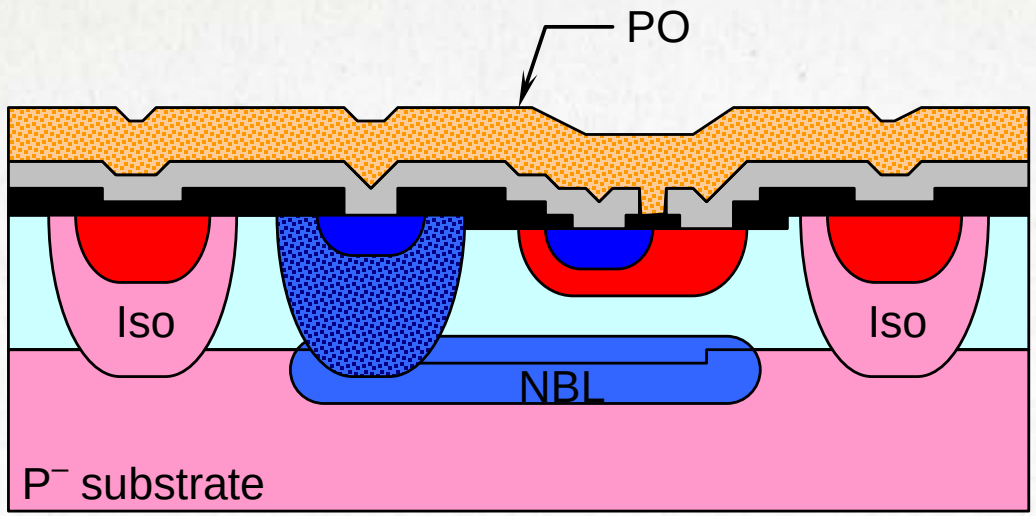
- **Create contact openings.**
  - Pattern using CONT (mask #6).
  - Remove oxide to open contacts.

METAL



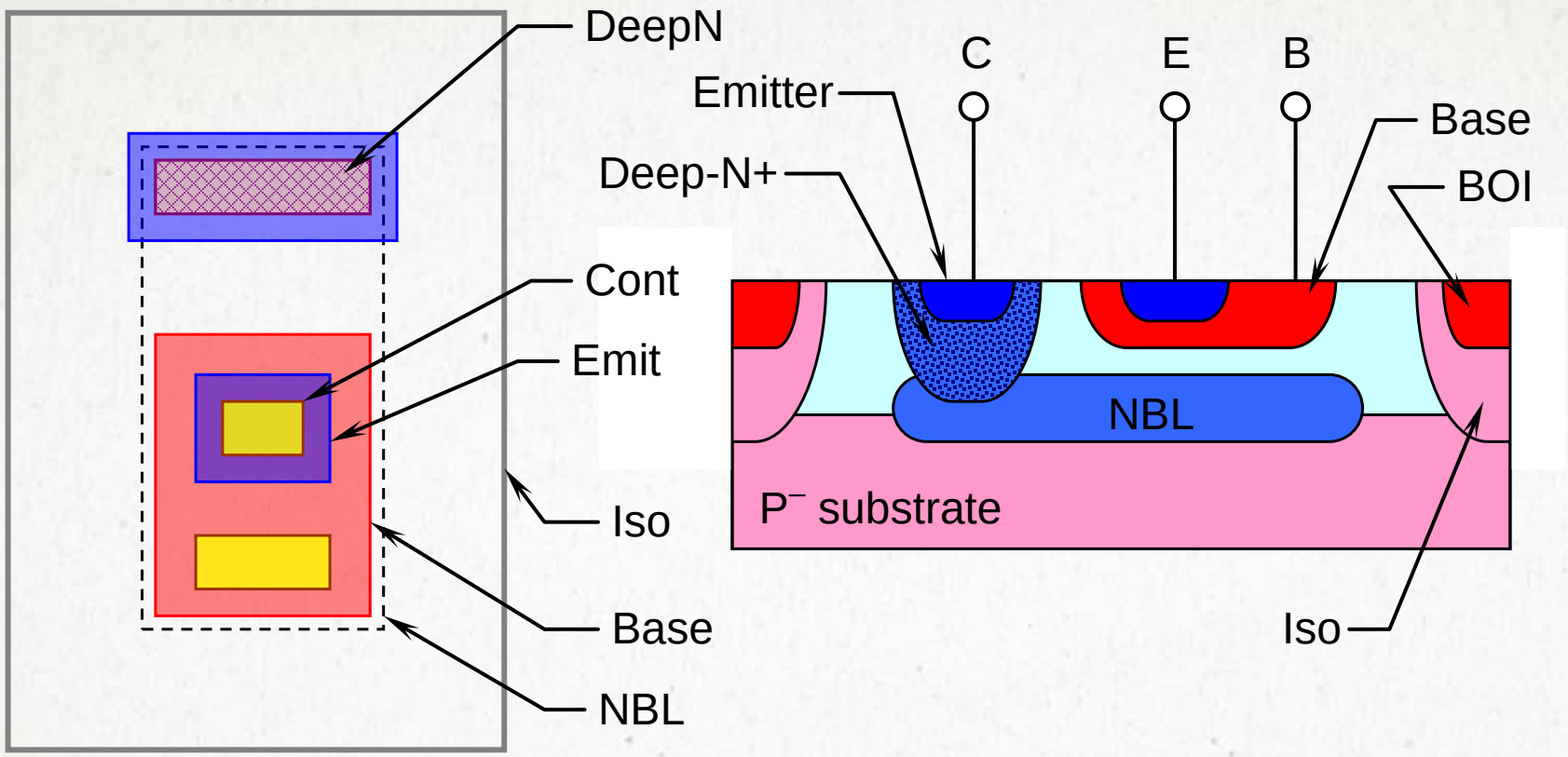
- **Form metal pattern (assume single-level metallization):**
  - Deposit aluminum-copper-silicon alloy.
  - Pattern using MET (mask #7).
  - Etch metal pattern.

POR (PROTECTIVE OVERCOAT REMOVAL)



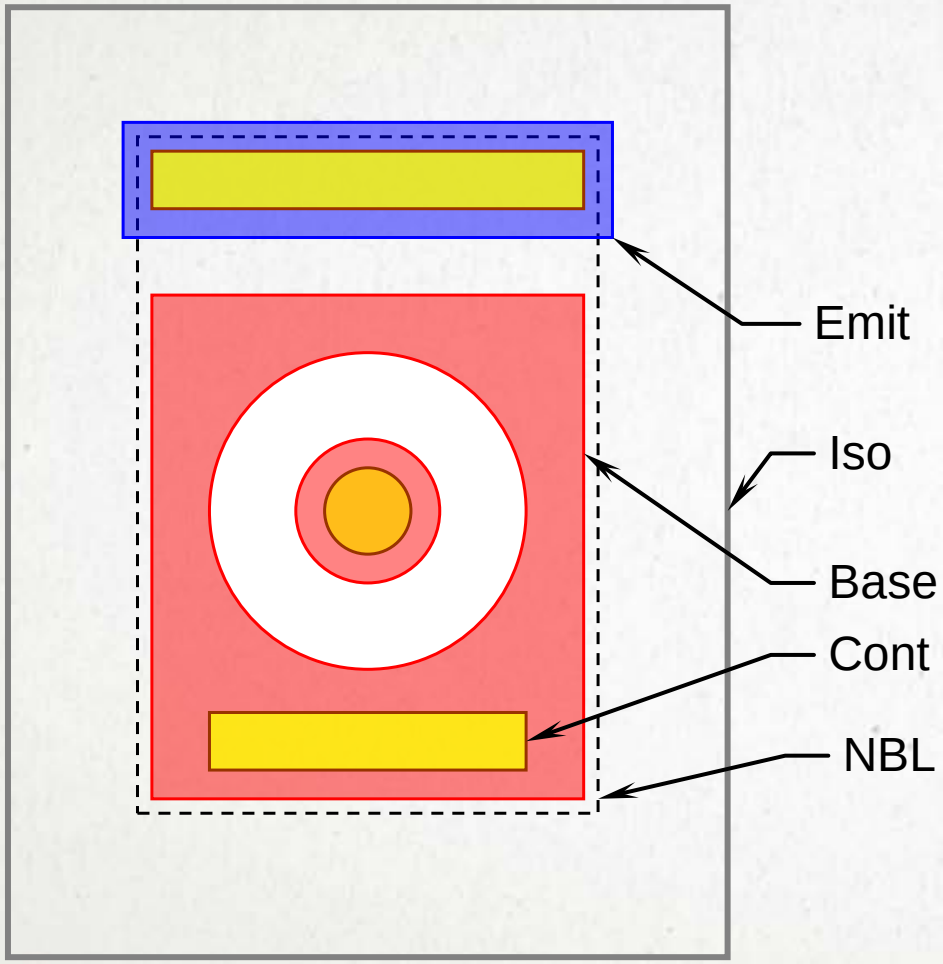
- **Deposit protective overcoat (PO):**
  - Deposit compressive nitride.
  - Pattern using POR (mask #8).
  - Etch POR openings for bondpads (none exist in the illustrated cross section).

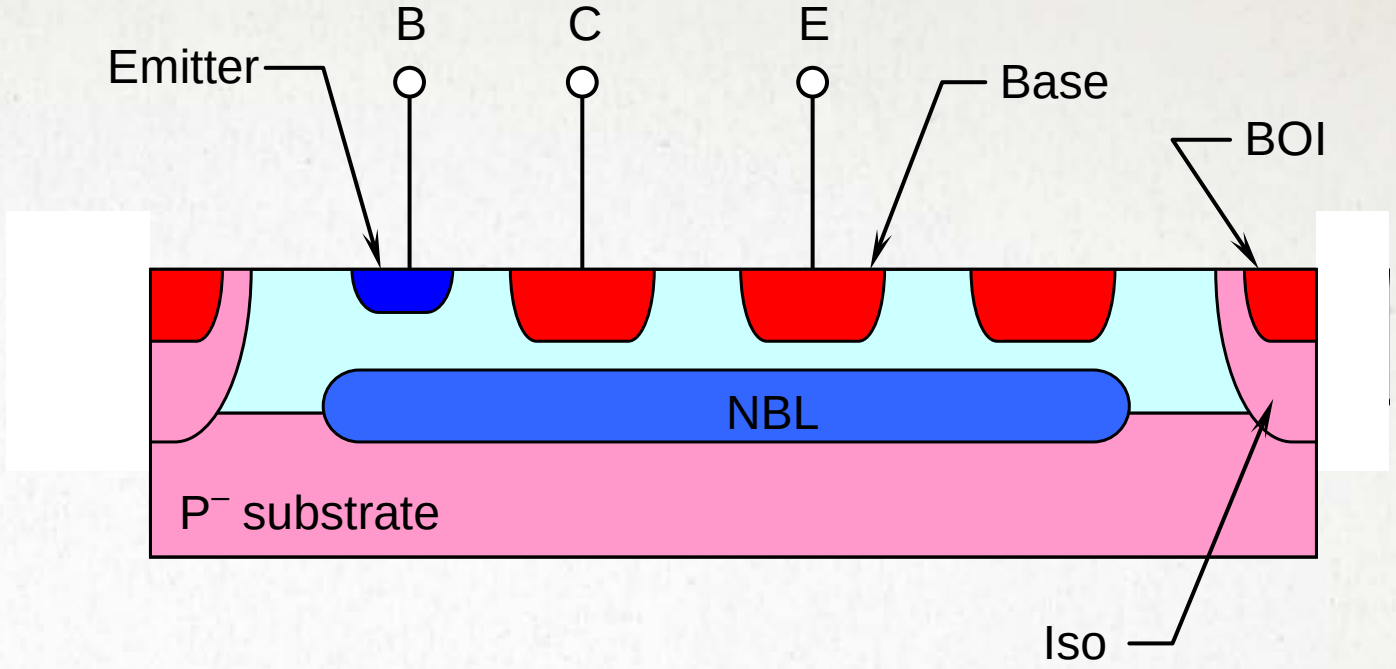
NPN



- **Features and limitations:**
  - Fully isolated with low-resistance collector.
  - High beta (100–300) and moderate V_{CEO} (typically 30 V).

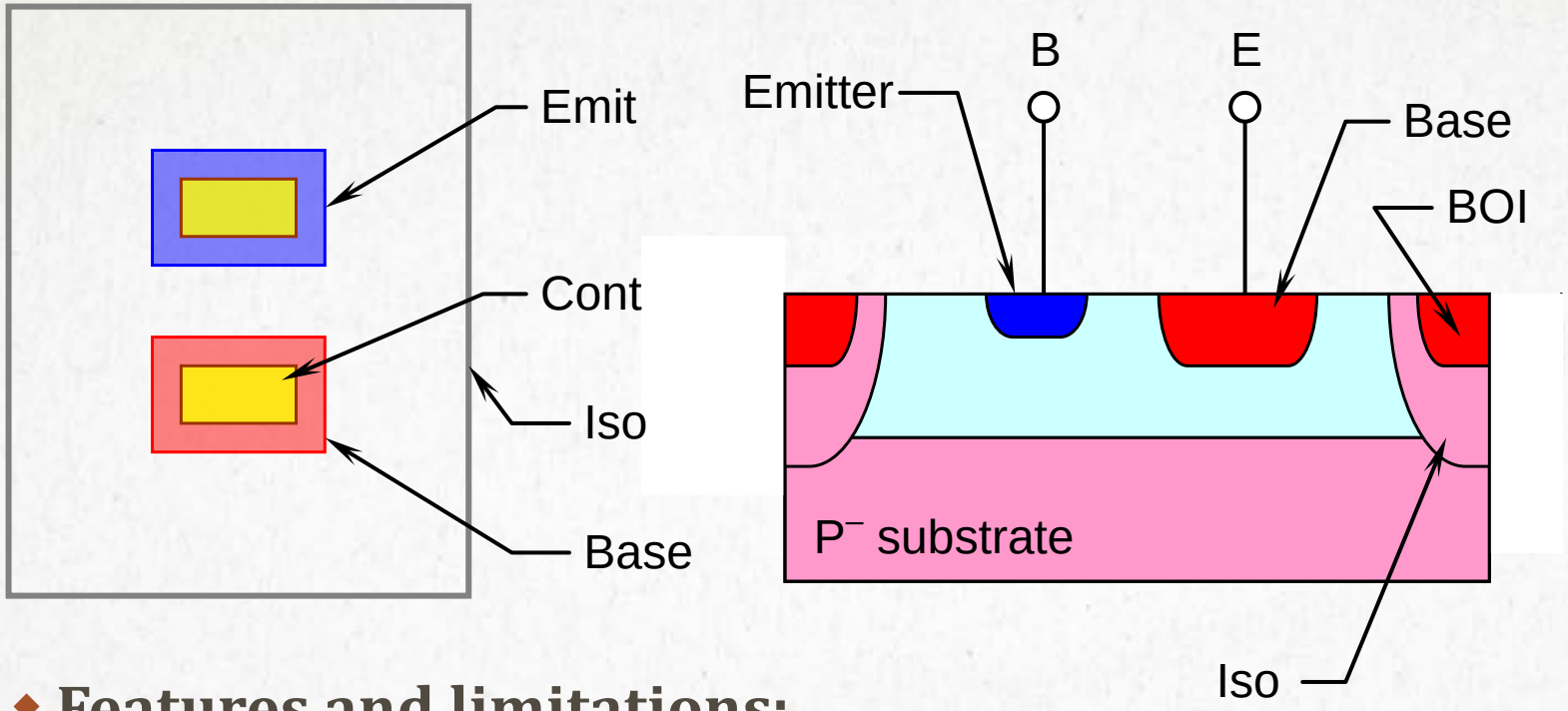
LATERAL PNP





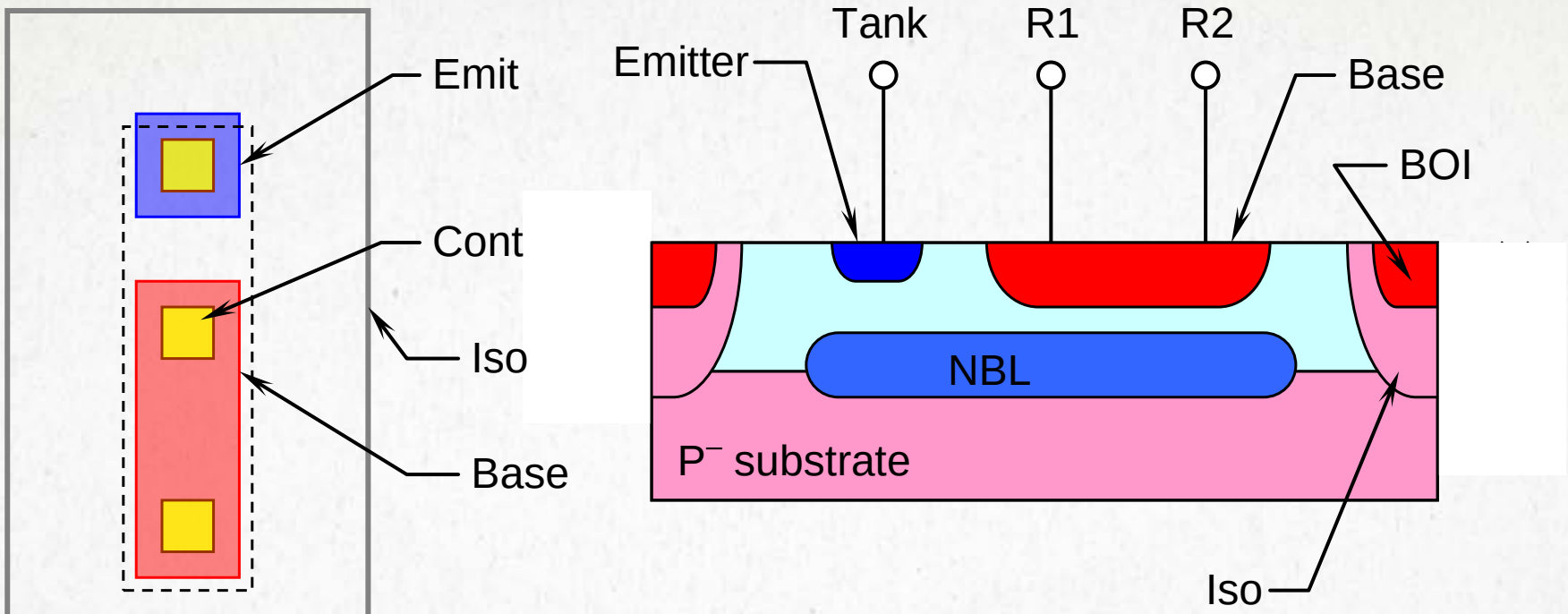
- **Features and limitations:**
  - Fully isolated; NBL prevents excessive substrate injection.
  - Moderate beta (typically 30–100).
  - Low frequency (f_T < 5 MHz).

SUBSTRATE PNP



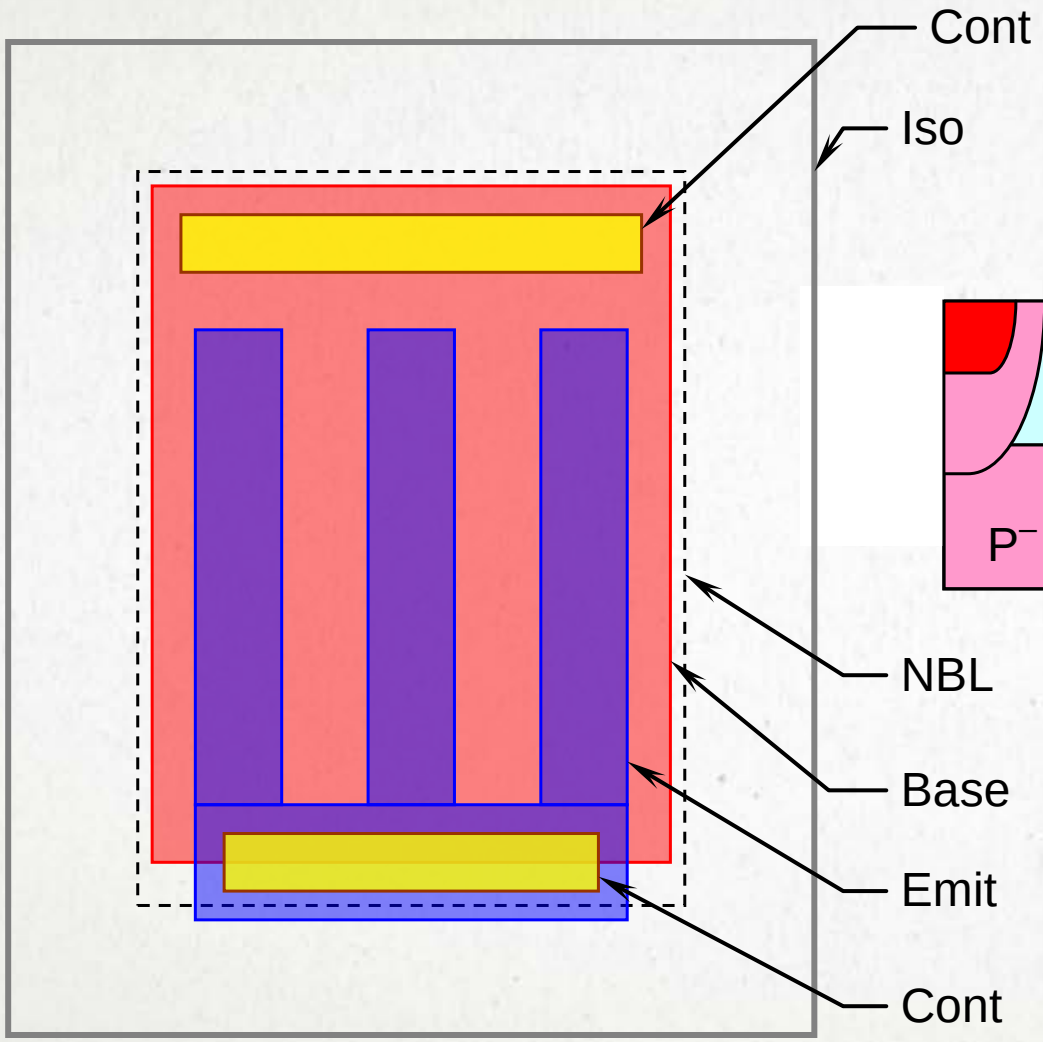
- **Features and limitations:**
  - Not fully isolated; collector is common to substrate.
  - High beta (50–200).

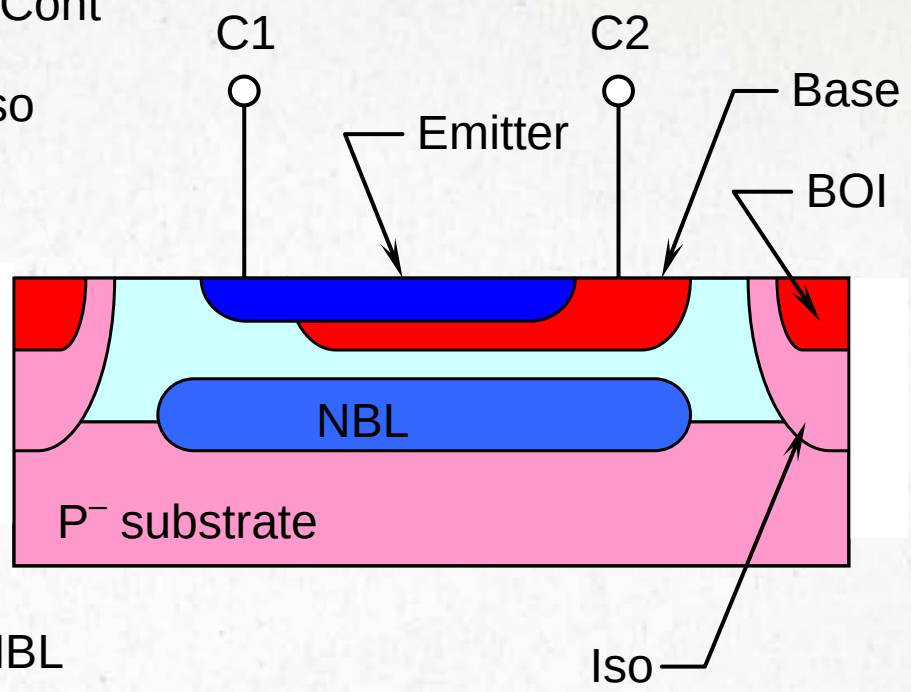
BASE RESISTOR



- **Features and limitations:**
  - Moderate sheet resistance (typically 160 Ω/□).
  - Good tolerance (typically ±25%).

JUNCTION CAPACITOR



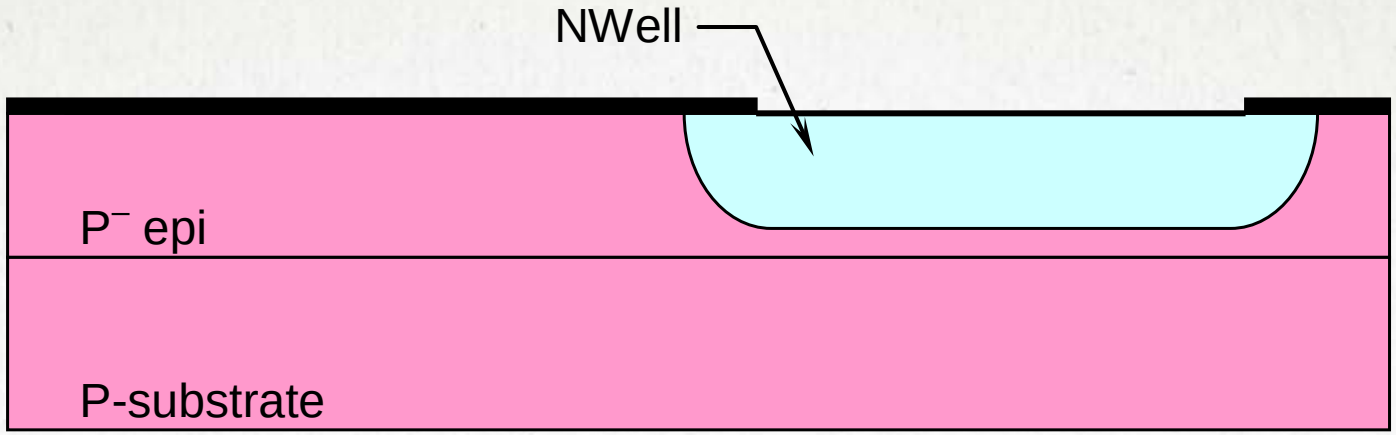


- **Features and limitations:**
  - High capacitance (typically 1.6 fF/μm^2).
  - Extremely variable.
  - Can only handle about 4–5 V.

POLY-GATE CMOS

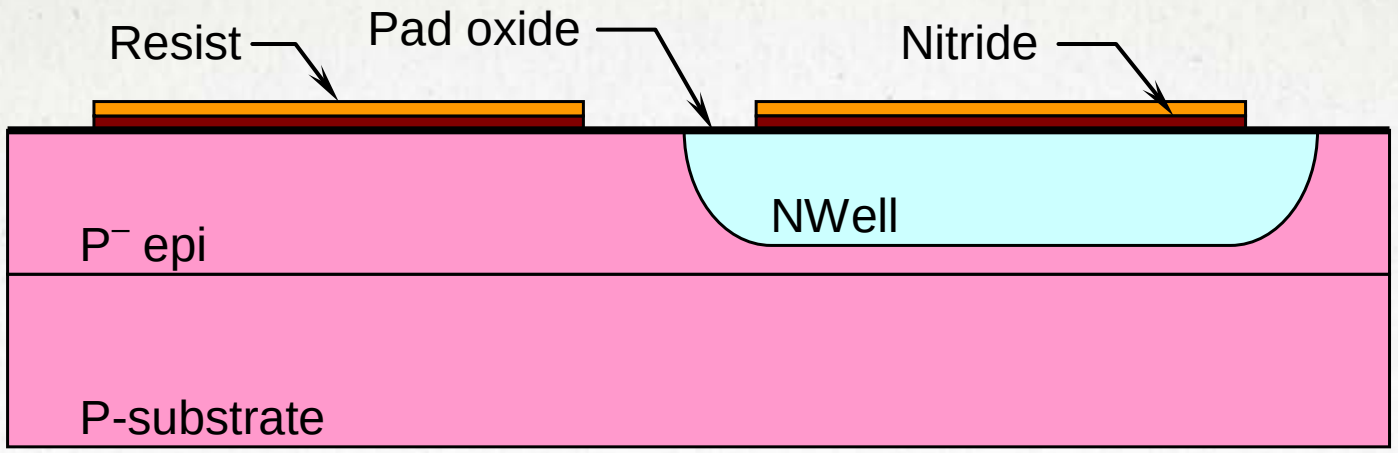
- **The self-aligned polysilicon-gate CMOS process is the basis for most modern analog CMOS and BiCMOS processes.**
- **Many features of this process are still widely used:**
  - P+ substrates
  - Self-aligned polysilicon gates
  - LOCOS field oxidation
  - Channel stop and threshold adjust implants
  - Silicidation and aluminum metallization

NWELL



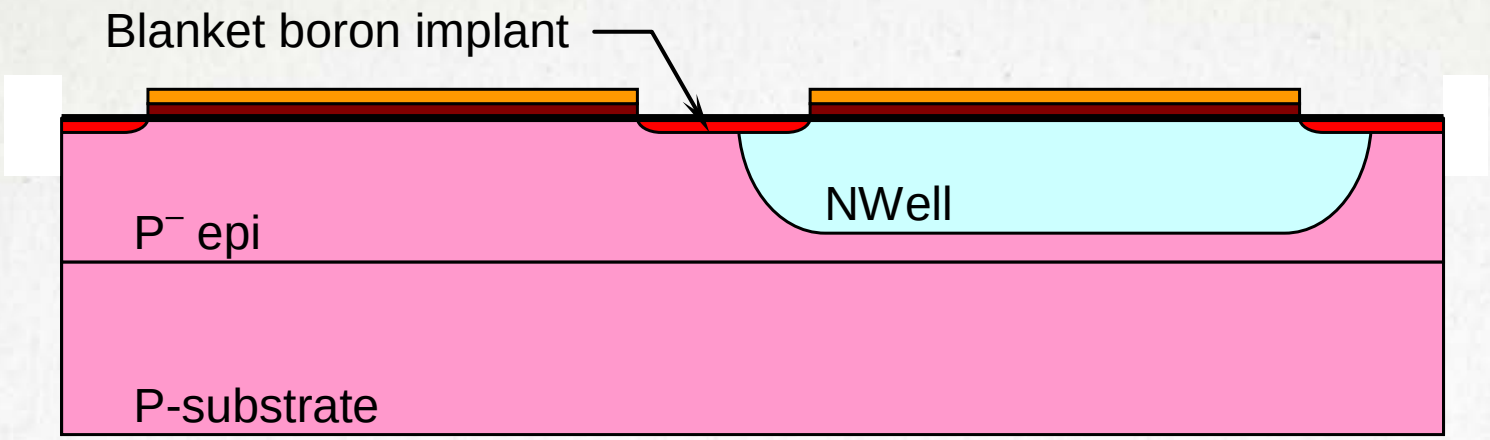
- **Form N-type well regions:**
  - Deposit P– epitaxial layer on P+ substrate.
  - Grow thermal oxide.
  - Pattern with NWell mask (mask #1).
  - Remove oxide over NWell regions.
  - Implant phosphorus into oxide openings.
  - Drive phosphorus down to form NWells.

MOAT



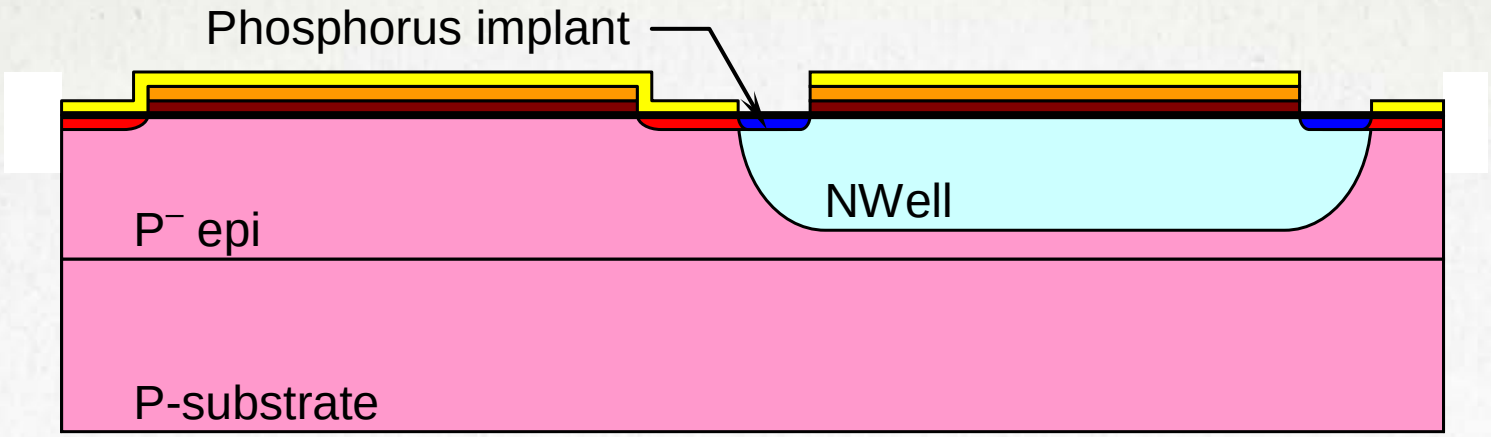
- **Deposit and pattern nitride for LOCOS oxidation:**
  - Strip thermal oxide.
  - Grow pad oxide.
  - Deposit nitride.
  - Pattern with moat mask (mask #2).
  - Etch nitride in inverse moat (field) oxide openings.

PCHST



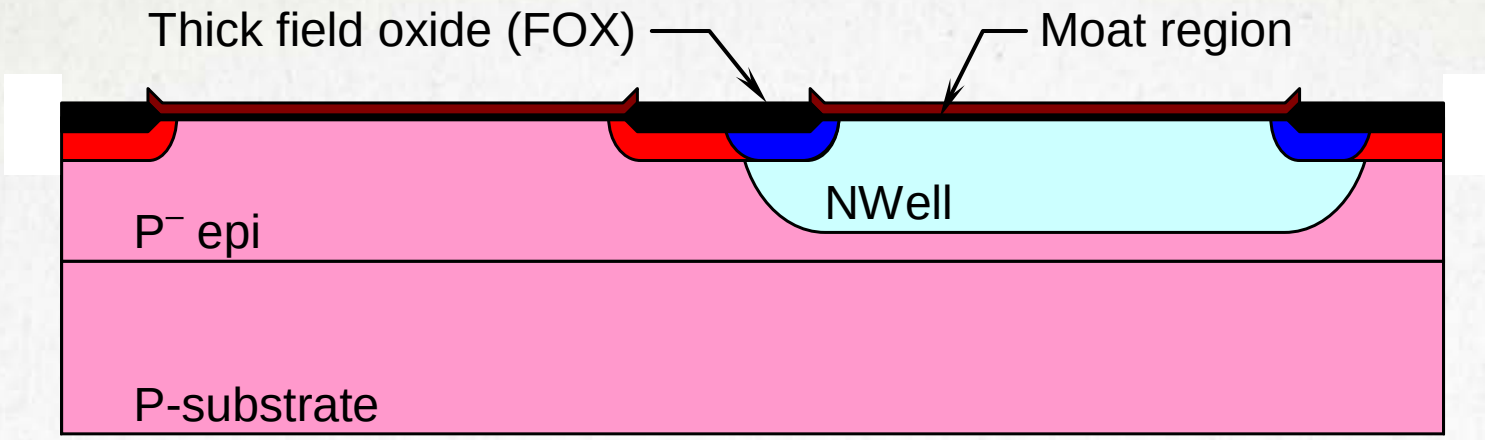
- **Following nitride etch, form P-type channel stop regions:**
  - Implant boron through nitride openings.
  - This P-type channel stop implant prevents parasitic channels from forming across P-type field regions.

NCHST



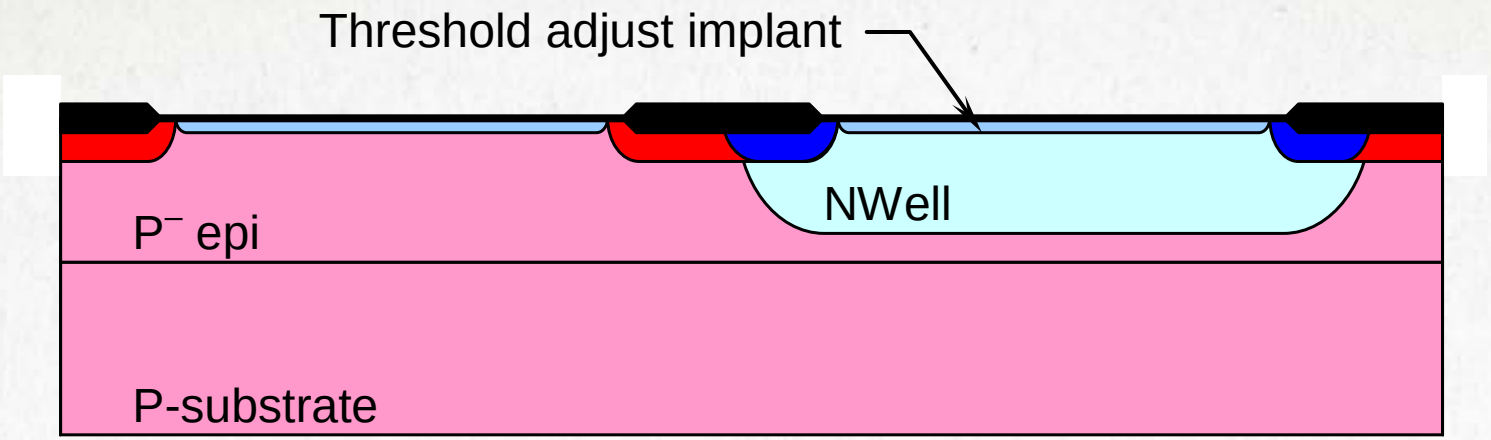
- **Next form patterned N-type channel stop implants:**
  - Spin another layer of photoresist on the wafer.
  - Pattern the photoresist using the channel stop mask (mask #3).
  - Implant phosphorus to form N-type channel stop (NChst) regions.
  - The N-type channel stop implants prevent parasitic channels from forming across NWell regions.

LOCOS



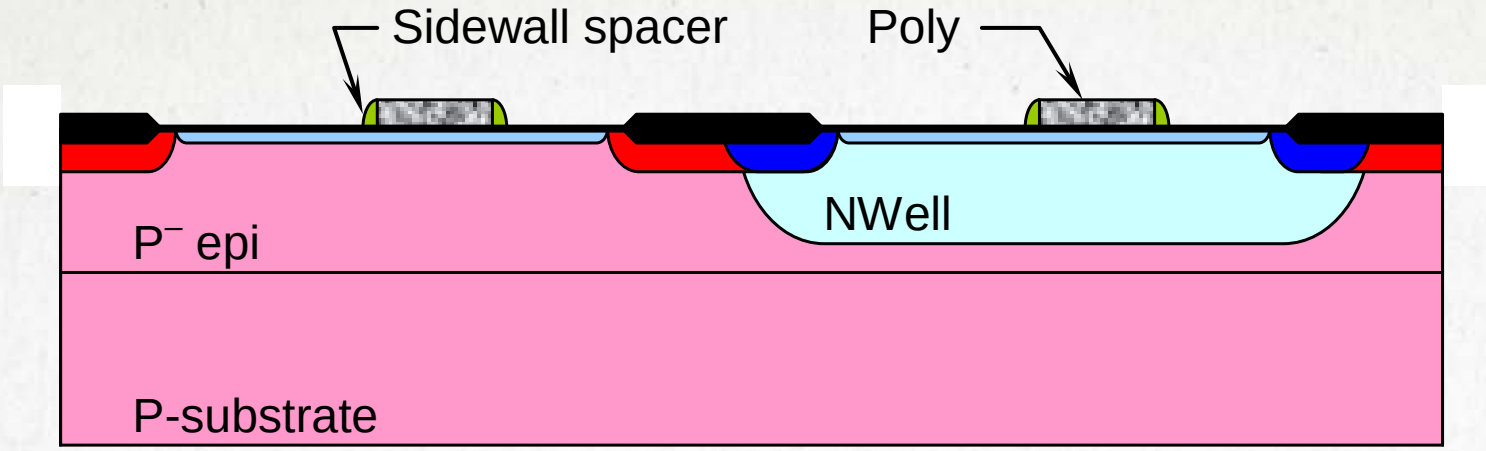
- **Grow field oxide (FOX):**
  - Remove photoresist.
  - Perform local oxidation of silicon (LOCOS).
  - Strip nitride.
  - Strip pad oxide.

THRESHOLD ADJUST



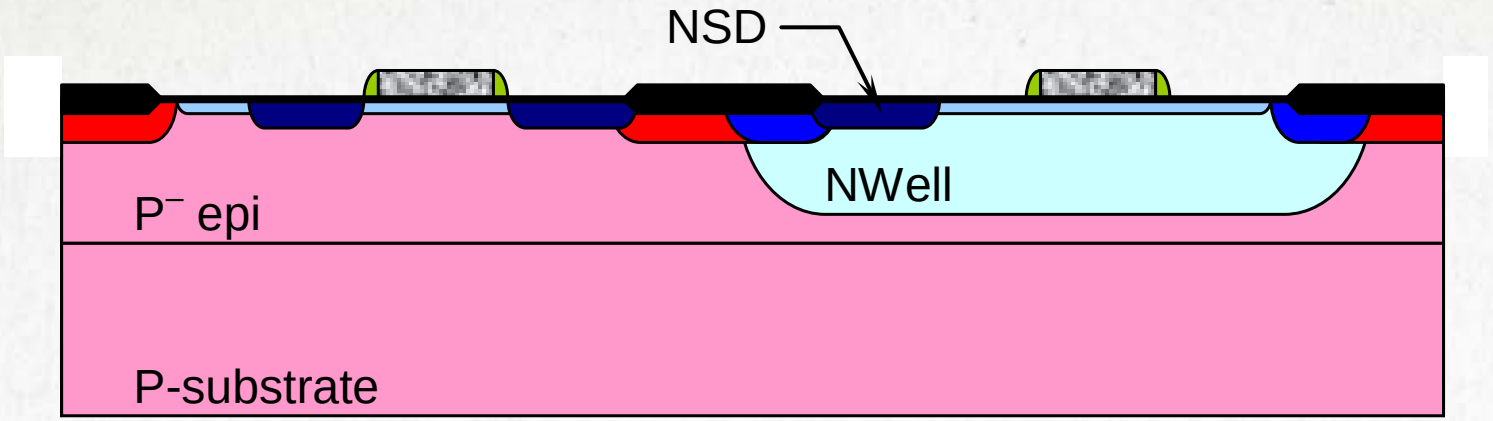
- **Grow gate oxide and perform threshold adjust implant:**
  - Grow dummy gate oxide.
  - Pattern with threshold adjust mask (mask #4).
  - Implant phosphorus to adjust threshold voltages of MOS transistors.
  - Strip dummy gate oxide.
  - Grow true gate oxide.

POLY



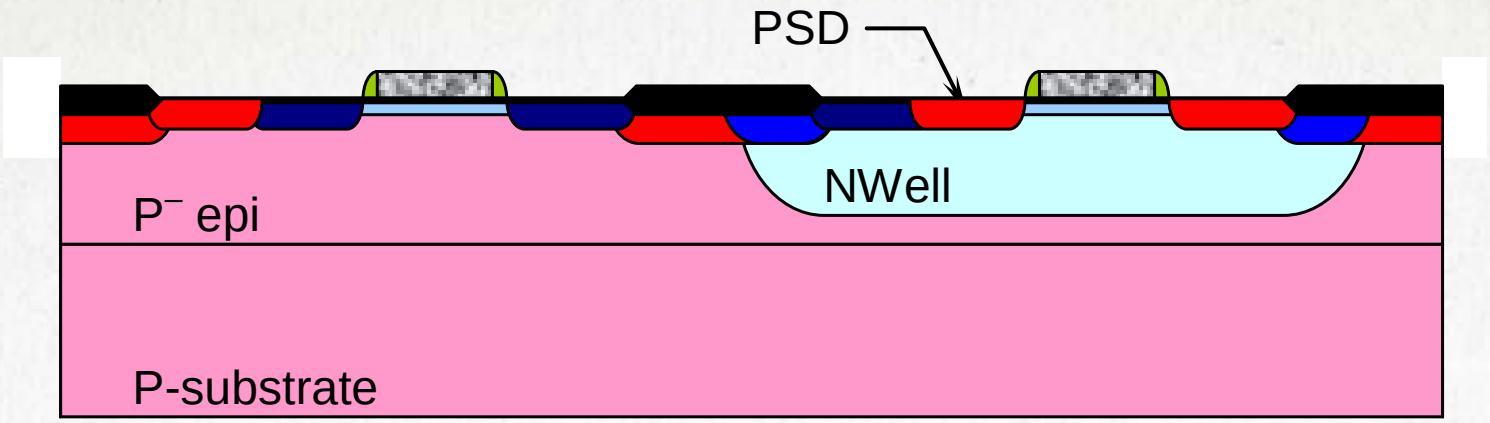
- **Form polysilicon gate regions:**
  - Deposit polysilicon (poly).
  - Dope poly with phosphorus.
  - Pattern polysilicon with poly mask (mask #5).
  - Etch poly.
  - Isotropically deposit oxide using spin-on glass.
  - Anisotropically etch oxide to form sidewall spacers.

NSD



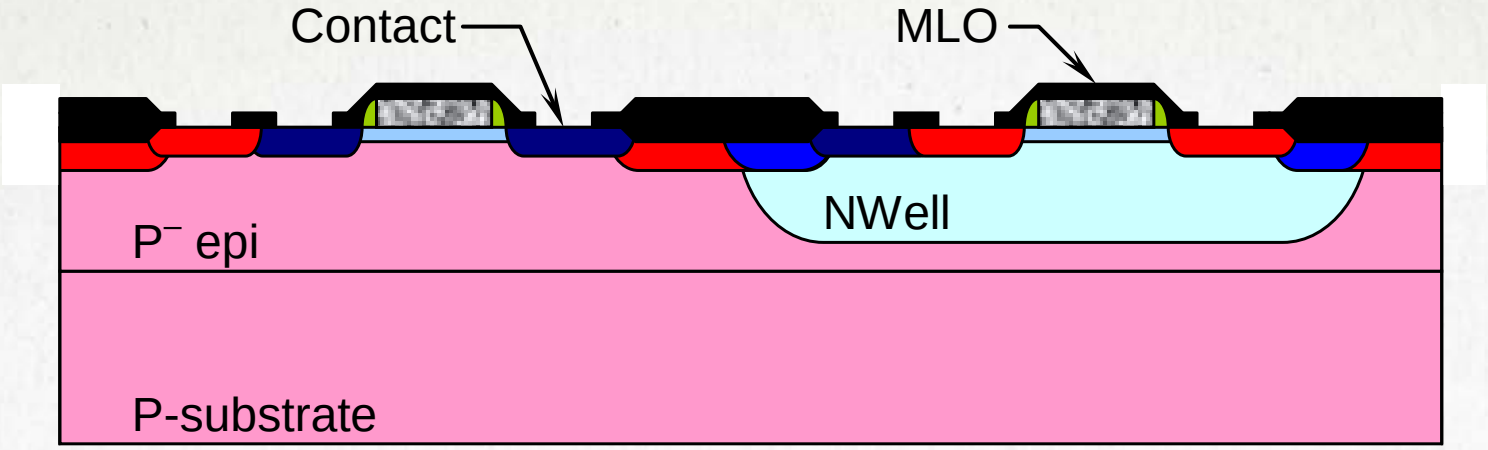
- **Form N-type source/drain (NSD) regions:**
  - Pattern using NSD mask (mask #6).
  - Implant arsenic.

PSD



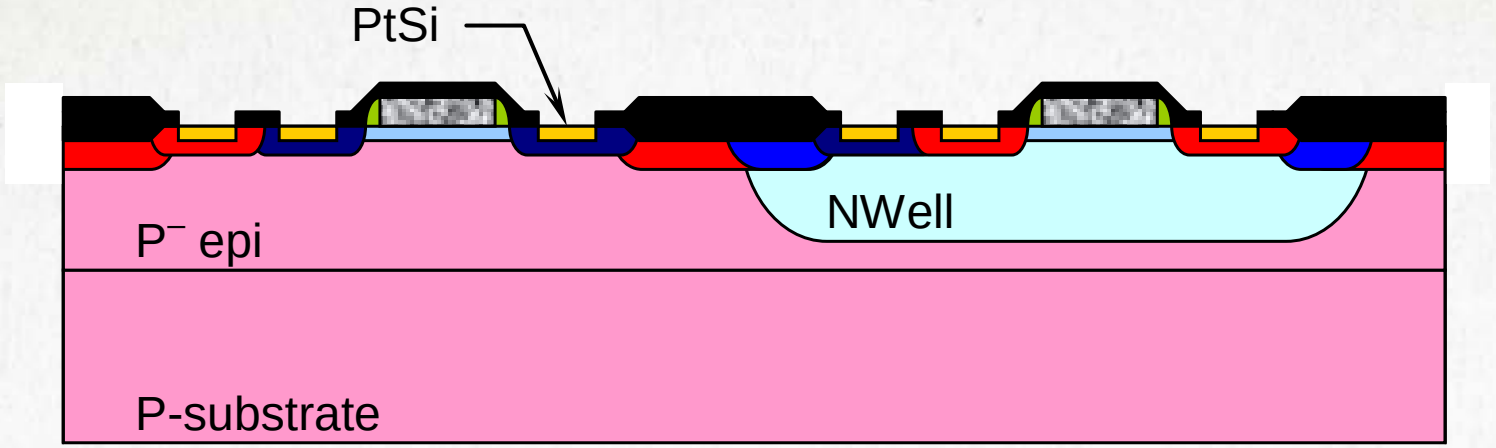
- **Form P-type source/drain (PSD) regions:**
  - Pattern using PSD mask (mask #7).
  - Implant boron.
  - Anneal source/drain implants.

CONTACT



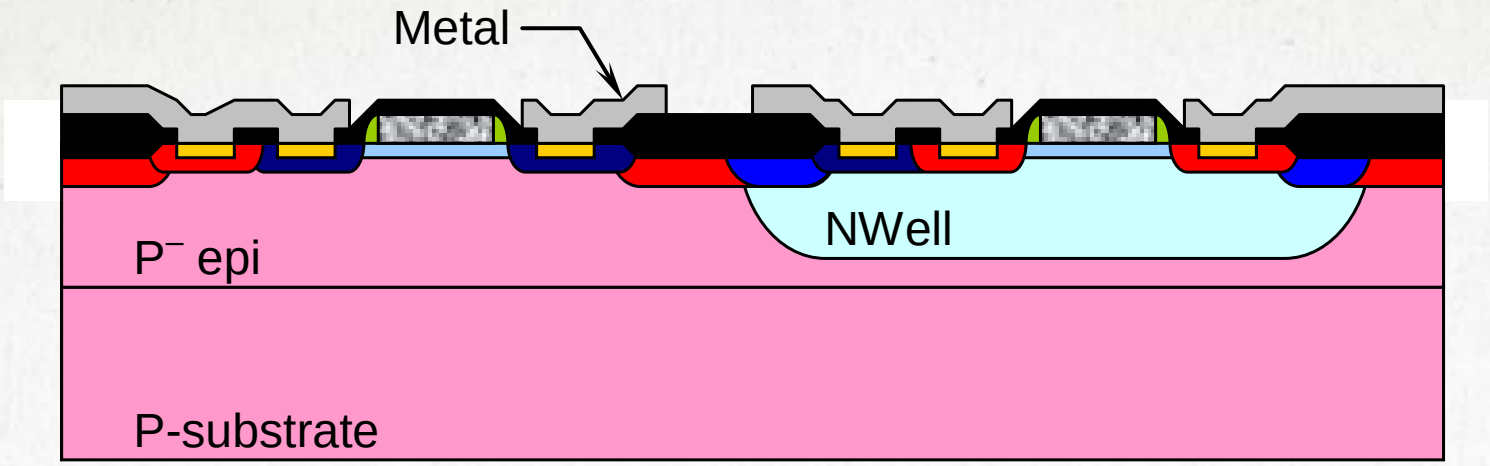
- **Form contact openings:**
  - Spin on multilevel oxide (MLO).
  - Pattern with contact mask (mask #8).
  - Reflow MLO to moderate contact sidewall slope.

SILICIDATION



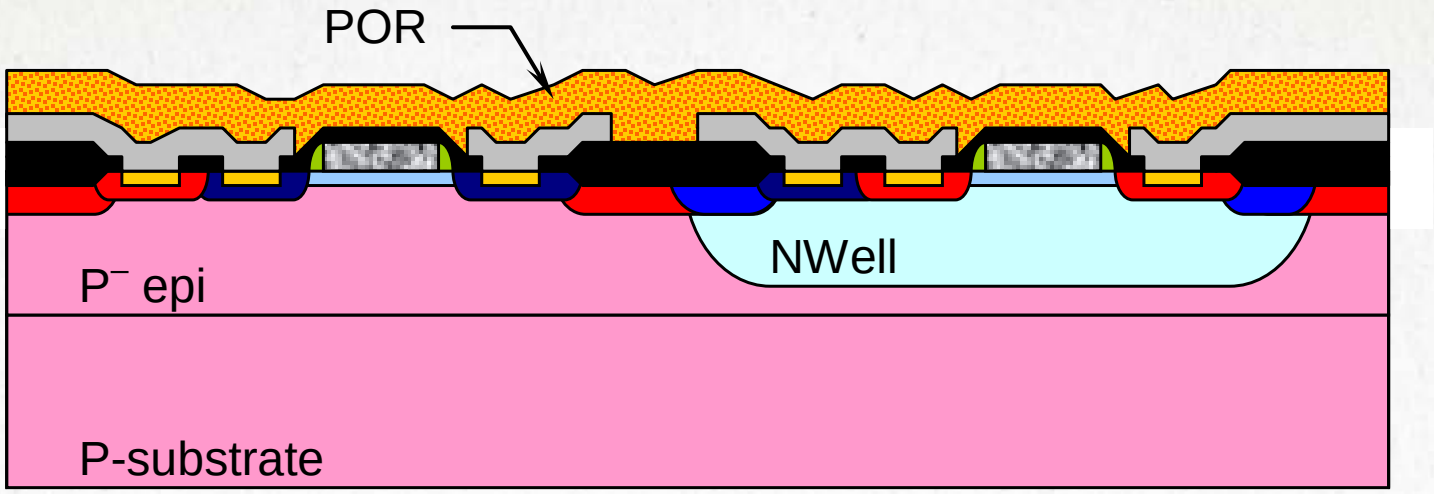
- **Silicide contact openings:**
  - Deposit platinum.
  - Sinter contacts to form platinum silicide.
  - Remove unreacted platinum with aqua regia.

METAL



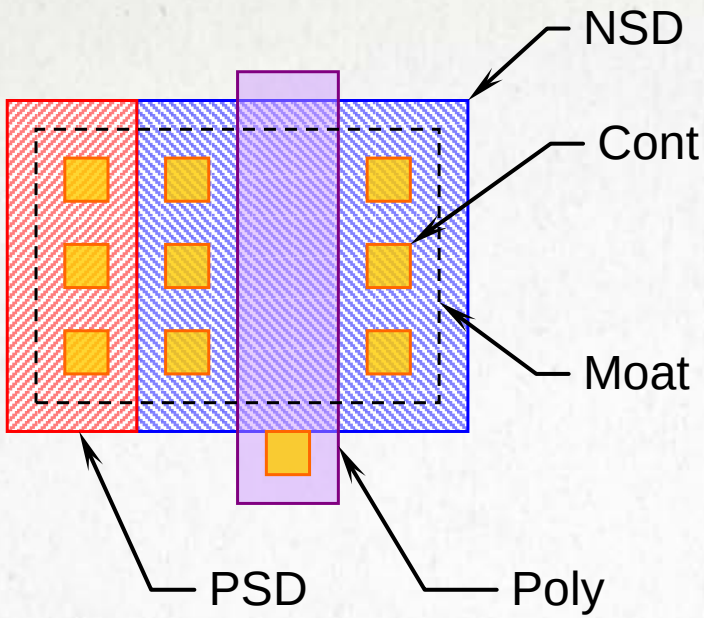
- **Form metal system:**
  - Deposit refractory barrier metal (RBM).
  - Deposit aluminum doped with 0.5% copper.
  - Pattern using metal mask (mask #9).
  - Etch metal.

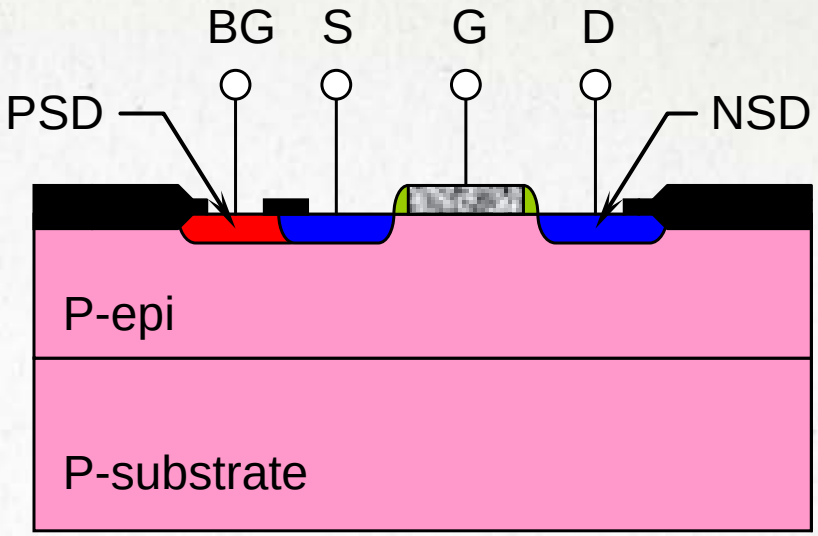
PROTECTIVE OVERCOAT



- **Deposit protective overcoat:**
  - Deposit compressive nitride.
  - Pattern using protective overcoat removal (POR) mask (mask #10).
  - Etch openings in overcoat for bondpads (none shown in cross section).

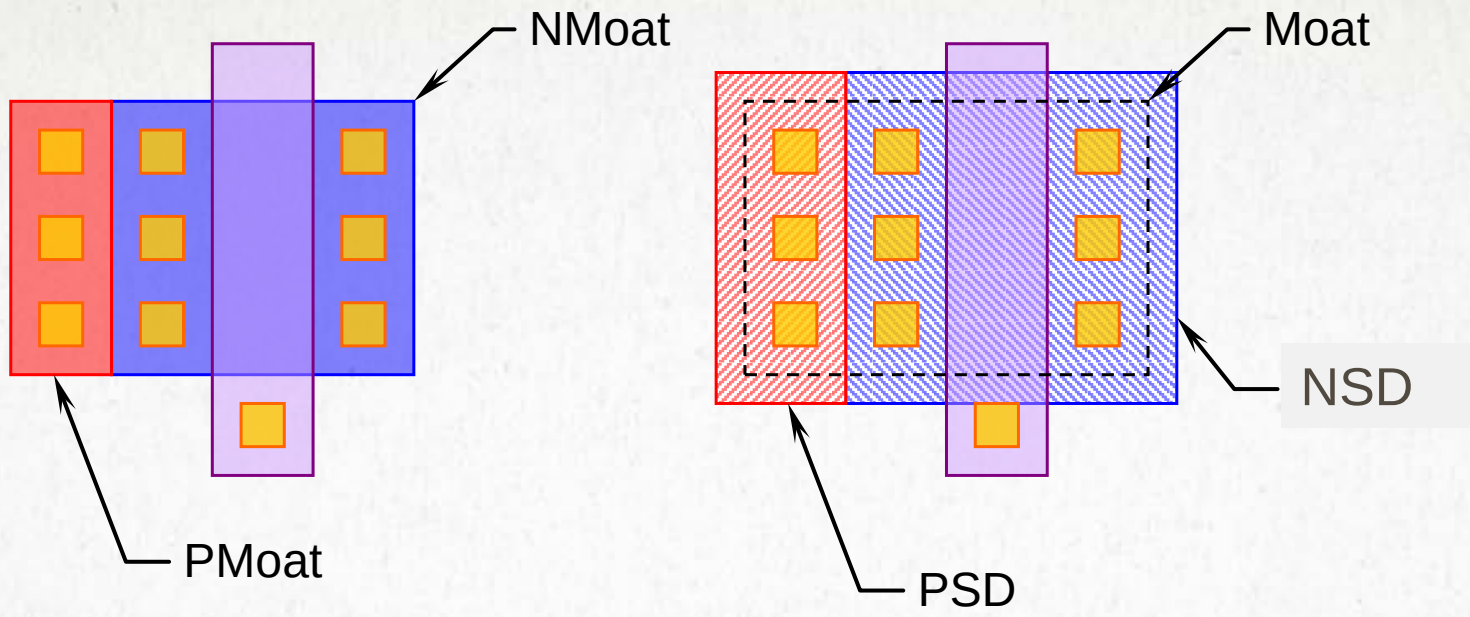
NMOS





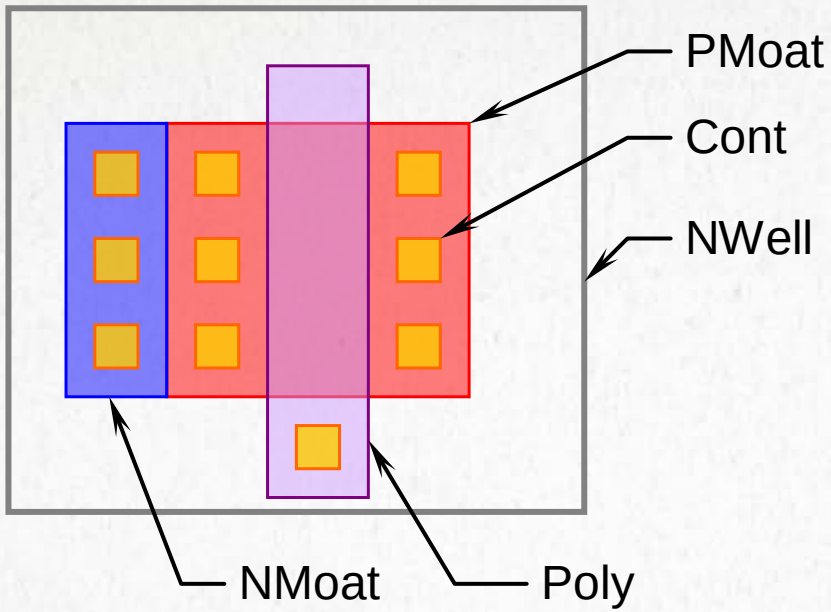
- **Features and limitations:**
  - Operating voltages are typically 5–20 V.
  - Minimum channel lengths are usually 2–4 μm; shorter channels require the addition of lightly doped drain (LDD) implants.
  - Not fully isolated; backgate is common to substrate.

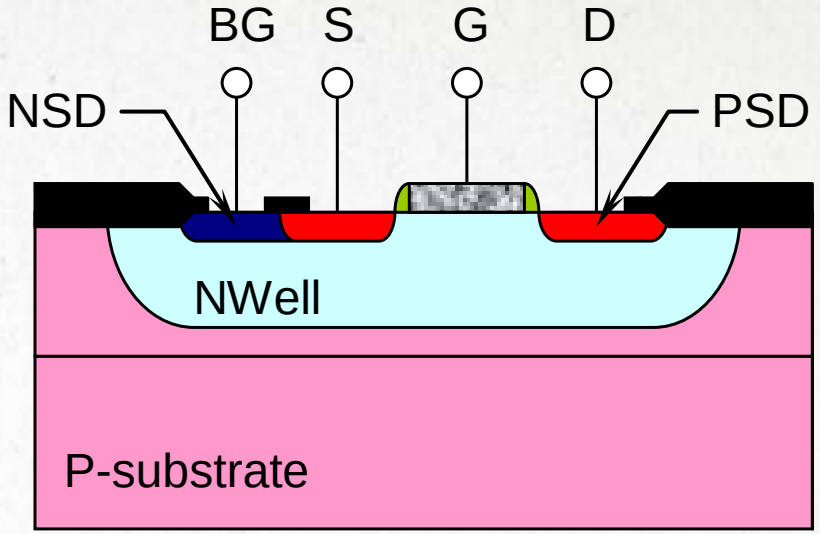
NMOS (CONTINUED)



- **Many processes use NMoat and PMoat *coding layers*.**
  - During *pattern generation* (PG), the NMoat and PMoat coding layers are translated into mask layers.
  - A geometry on PMoat becomes geometries on PSD and Moat.
  - A geometry on NMoat becomes geometries on NSD and Moat.

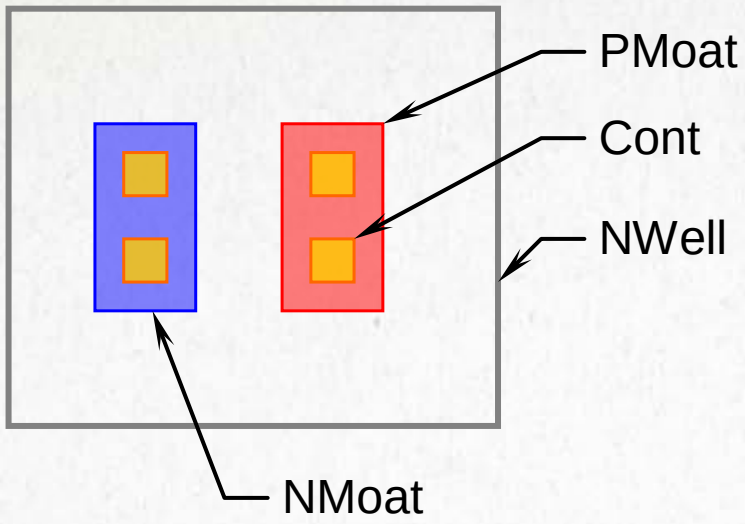
PMOS

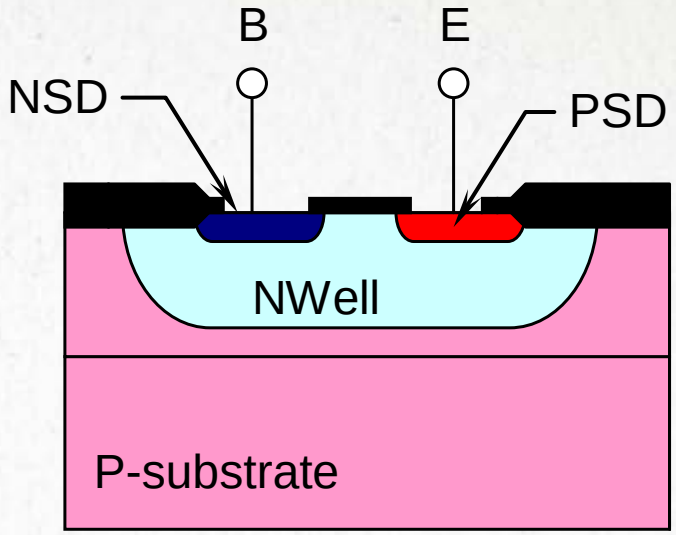




- **Features and limitations:**
  - Operating voltages are typically 5–20 V.
  - Minimum channel lengths are usually 2–3 μm; shorter channels require the addition of lightly doped drain (LDD) implants.
  - Fully isolated; backgate can be tied to any desired node.

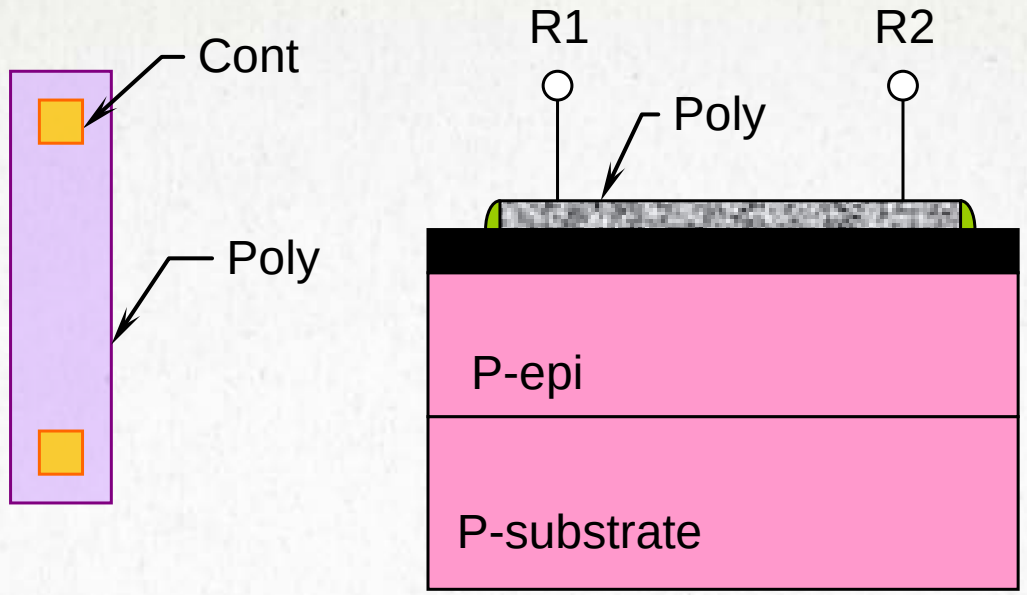
SUBSTRATE PNP





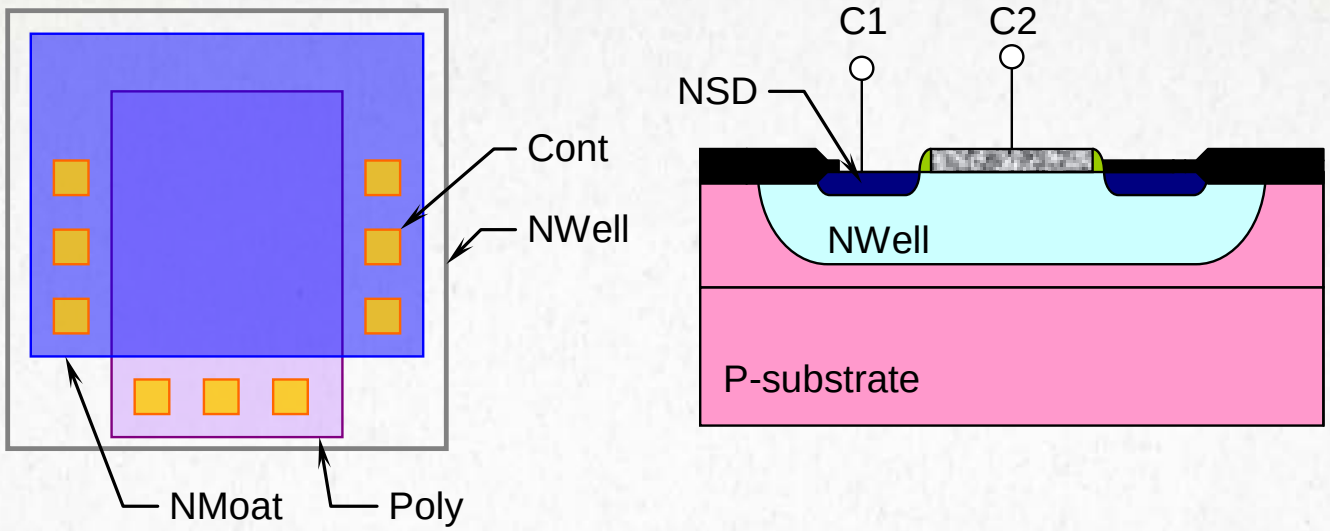
- **Features and limitations:**
  - Operating voltages are typically 5–20 V.
  - Not fully isolated; the collector is common to substrate.
  - Typically has a nominal beta of 50.

POLY RESISTOR



- **Features and limitations:**
  - Fully oxide isolated.
  - Low sheet resistance (typically 20 Ω/☐).
  - Adding one additional mask to block poly doping permits creation of high-sheet resistors (typically 500 Ω/☐).

GATE OXIDE CAPACITOR



- **Features and limitations:**
  - Relatively high capacitance.
  - Must be properly biased to maintain accumulation.
  - NWell plate has parasitic capacitance to substrate.
  - Adding one mask step permits formation of fully isolated capacitors.

PASSIVES

ALAN HASTINGS

PASSIVE DEVICES

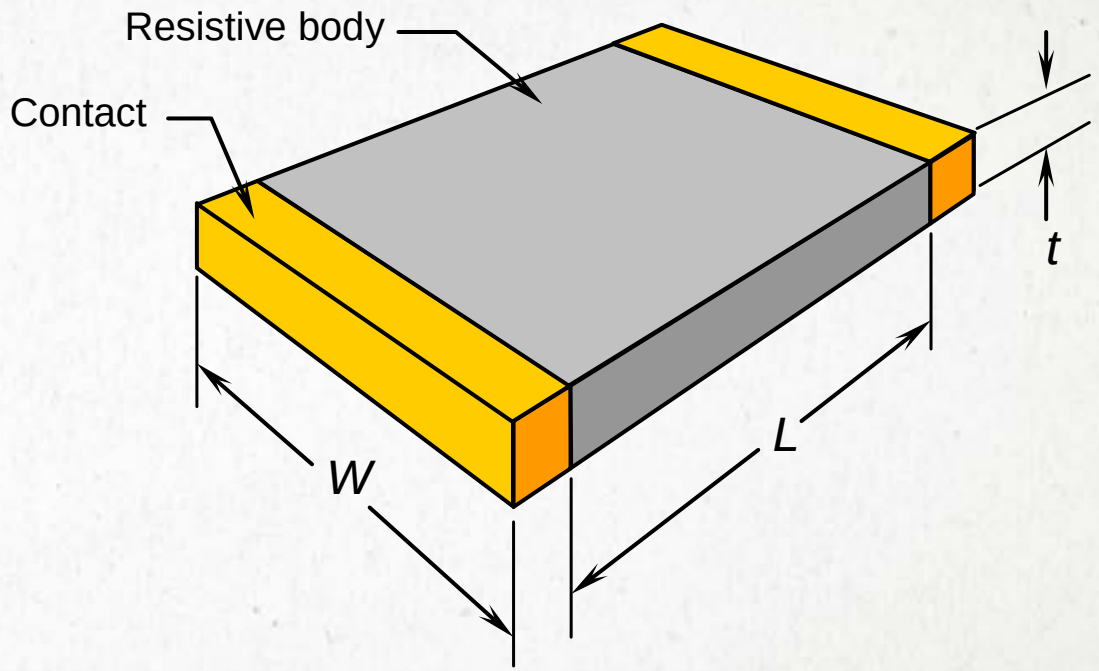
- **Passive devices include resistors, capacitors, and inductors.**
  - Milliohms to megohms of resistance are easily integrated.
  - Capacitors of up to a few hundred picofarads can be integrated.
  - Inductors don't lend themselves to easy integration and won't be discussed (although some RF designs do make use of very small integrated inductors).
- **Diodes are technically passive devices, but we will consider them later when we discuss bipolar transistors.**

RESISTIVITY

- Consider a simple rectangular slab having width *W*, length *L*, and thickness *t*. Its resistance *R* equals

$$R = \rho \frac{L}{Wt}$$

- The constant of proportionality ρ is called the *resistivity* and has units of Ω·cm.



RESTIVITITY

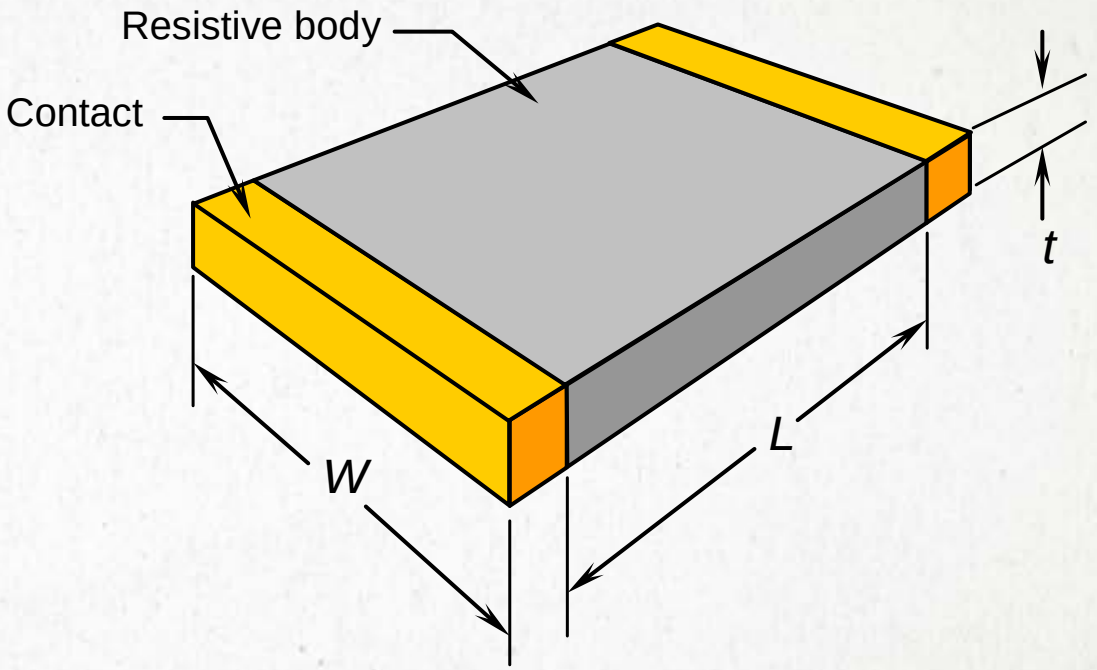
Material	Resistivity, Ω·cm (25°C)
Copper, bulk	$1.7 \cdot 10^{-6}$
Gold, bulk	$2.4 \cdot 10^{-6}$
Aluminum, thin film	$2.7 \cdot 10^{-6}$
Aluminum (2% silicon)	$3.8 \cdot 10^{-6}$
Platinum silicide	$3.0 \cdot 10^{-5}$
Silicon, N-type ($N_d = 10^{18}cm^{-3}$)	0.25
Silicon, N-type ($N_d = 10^{15}cm^{-3}$)	48
Silicon, intrinsic	$2.5 \cdot 10^{5}$
Silicon dioxide	$\sim 10^{14}$

- ***Conductors*, such as metal, have low resistivities.**
- ***Semiconductors*, such as silicon, have moderate resistitivies that depend strongly upon doping levels.**
- ***Insulators*, such as oxide, have extremely high resistivities.**

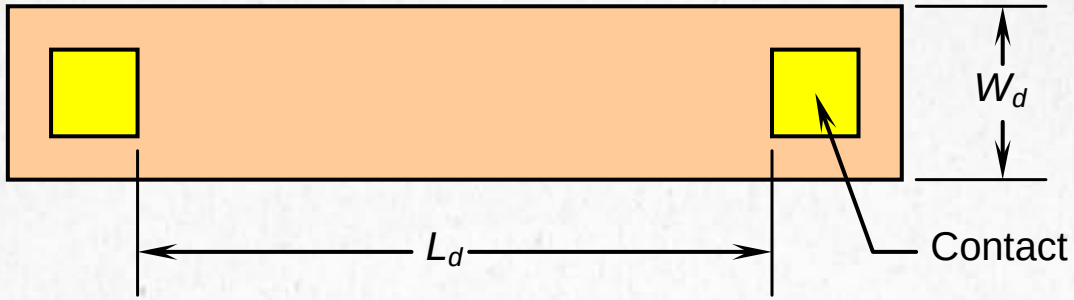
SHEET RESISTANCE

- **Integrated resistors usually consist of diffusions or depositions of constant thickness.**
  - We can combine resistivity and thickness into a single term called *sheet resistance*, R_S, where $R_S = \rho/t$.
  - The equation for resistance now becomes:

$$R = R_s\left(\frac{L}{W}\right)$$



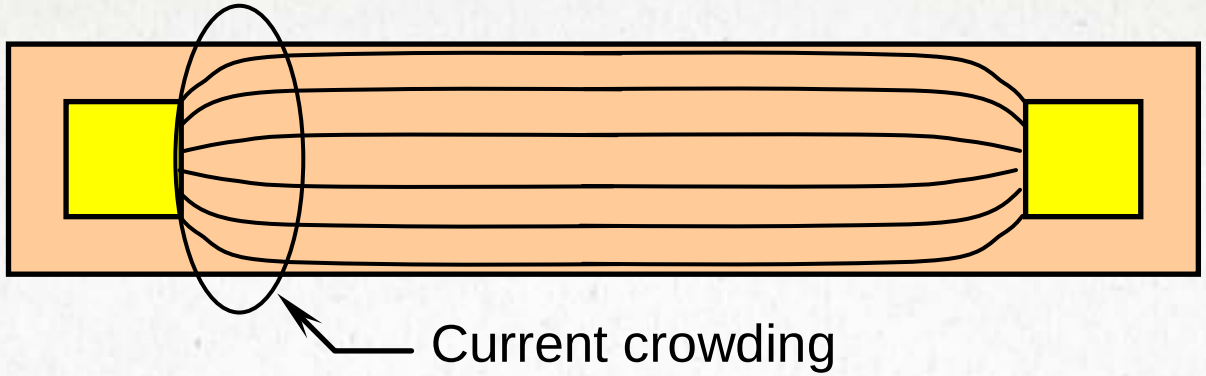
WIDTH AND LENGTH BIASES



- **The *drawn width* W_d and *drawn length* L_d are those entered into the database.**
  - The true width and length differ from the drawn width and length because of outdiffusion, overetching, process size adjustments, etc.
  - Therefore we define a *width bias* W_b and a *length bias* L_b such that

$$R = R_S\left(\frac{W_d + W_b}{L_d + L_b}\right)$$

CURRENT CROWDING

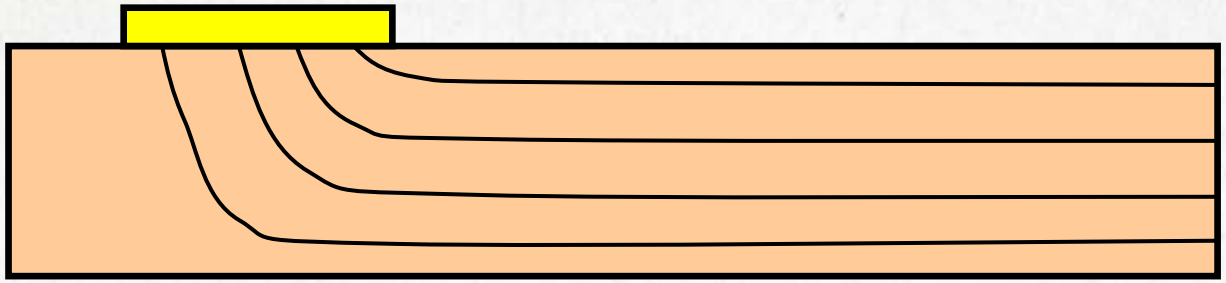


- **Current crowds inwards towards the contacts, increasing resistance by an amount**

$$\Delta R = \frac{R_s}{\pi}\left[\frac{1}{k}\ln\left(\frac{k+1}{k-1}\right)+\ln\left(\frac{k^2-1}{k^2}\right)\right]$$

where $k = W_d/(W_d - W_c)$, in which W_c is the width of the contact.

CONTACT RESISTANCE



- **Current also does not flow uniformly into the contact in the vertical direction.**
  - This is usually modeled by assuming each contact adds a *contact resistance* R_c to the resistor.
  - For a homogenous resistor whose contact material has resistivity ρ_c,

$$R_c = \frac{\sqrt{R_s \rho_c}}{W_C} \coth\left(L_C \sqrt{R_s / \rho_c}\right)$$

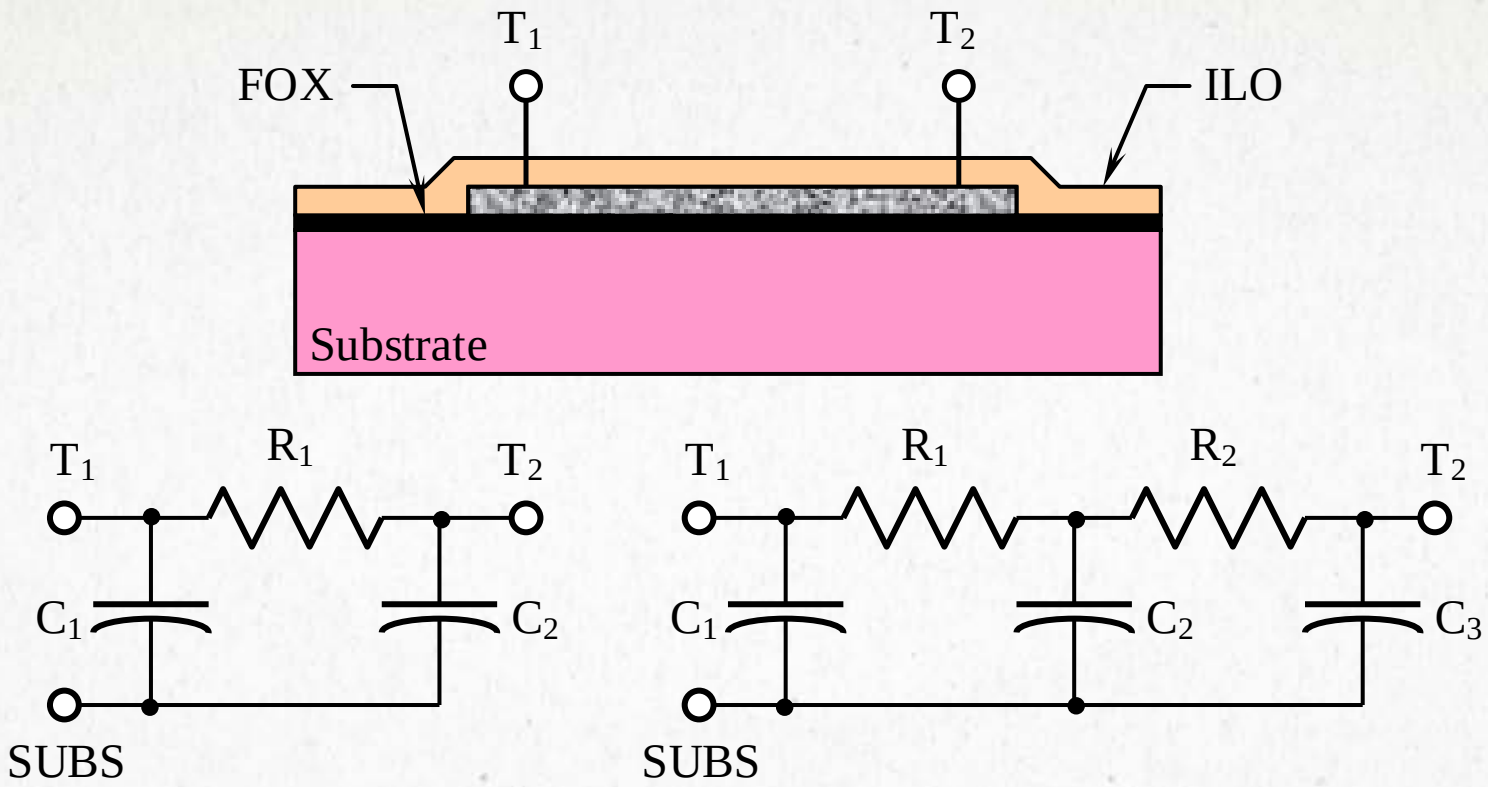
TEMPCO

$$R(T) = R(T_0)[1 + 10^{-6} TC(T - T_0)]$$

- **Resistivity varies with temperature.**
  - The actual relationship is nonlinear, but we can linearize it.
  - The constant of proportionality *TC*, called the *coefficient of resistivity*, has units of ppm/°C.
  - *TC* is often called the *tempco*.

Material	TCR, ppm/°C
Copper, bulk	+4000
Gold, bulk	+3700
Aluminum, bulk	+3800
160 Ω/☐ base diffusion	+1500
7 Ω/☐ emitter diffusion	+600
5 kΩ/☐ base pinch diffusion	+2500
2 kΩ/☐ HSR implant (P-type)	+3000
500 Ω/☐ polysilicon (4 kÅ N-type)	–1000
25 Ω/☐ polysilicon (4 kÅ N-type)	+1000
10 kΩ/☐ NWell	+6000

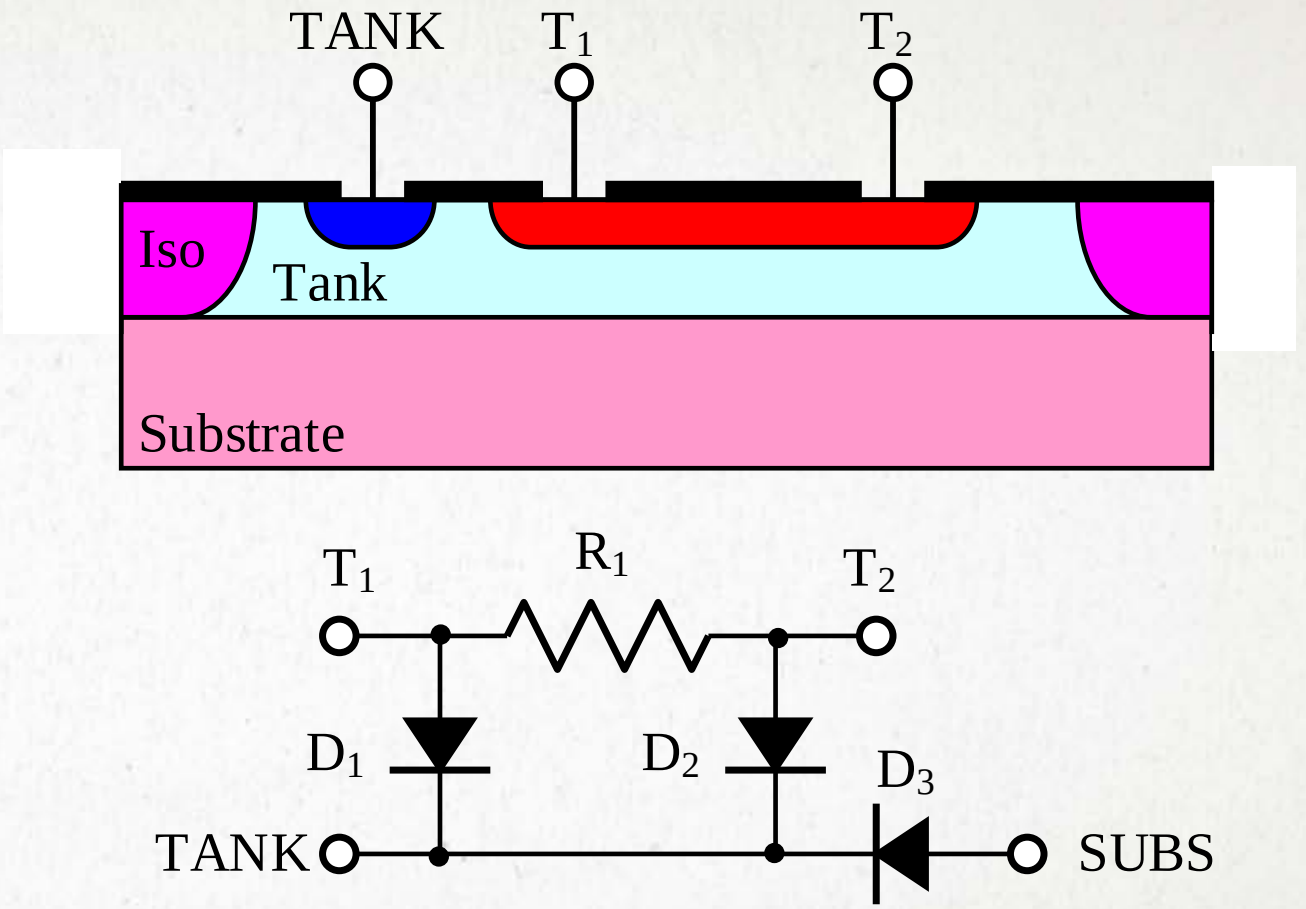
DEPOSITED RESISTORS



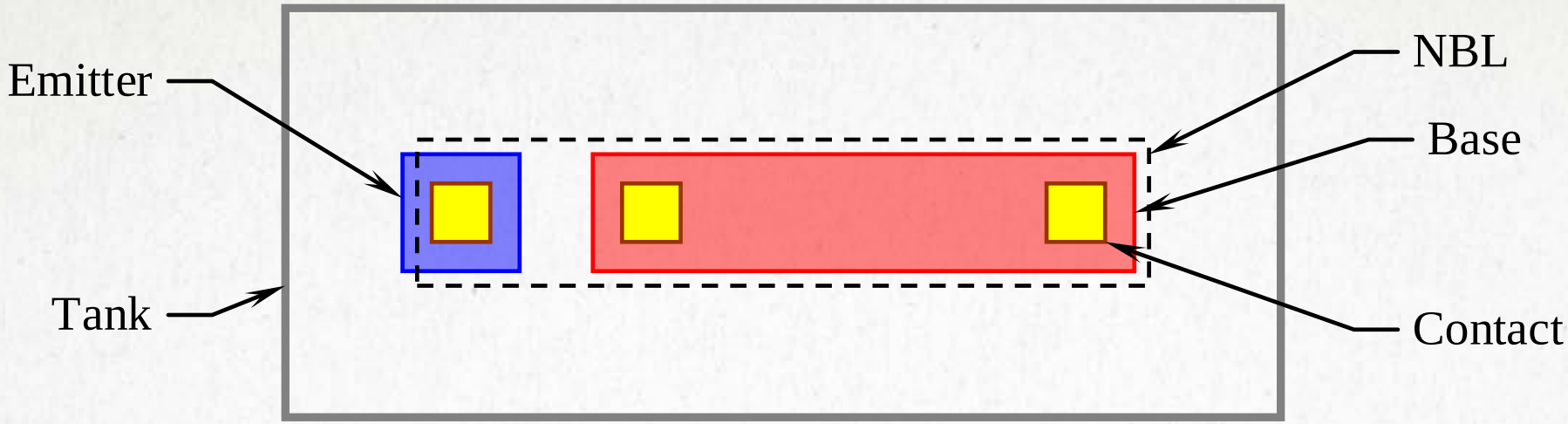
- **Deposited resistors (such as poly resistors) are fully insulated from the silicon.**
  - The only important parasitics are therefore capacitances.

DIFFUSED RESISTORS

- **Diffused resistors occupy a tank or well.**
  - A PN junction exists between the resistor body and the enclosing tank or well.
  - The tank or well must always be connected to reverse-bias this PN junction.
  - *Improper tank or well biasing is a common layout error.*

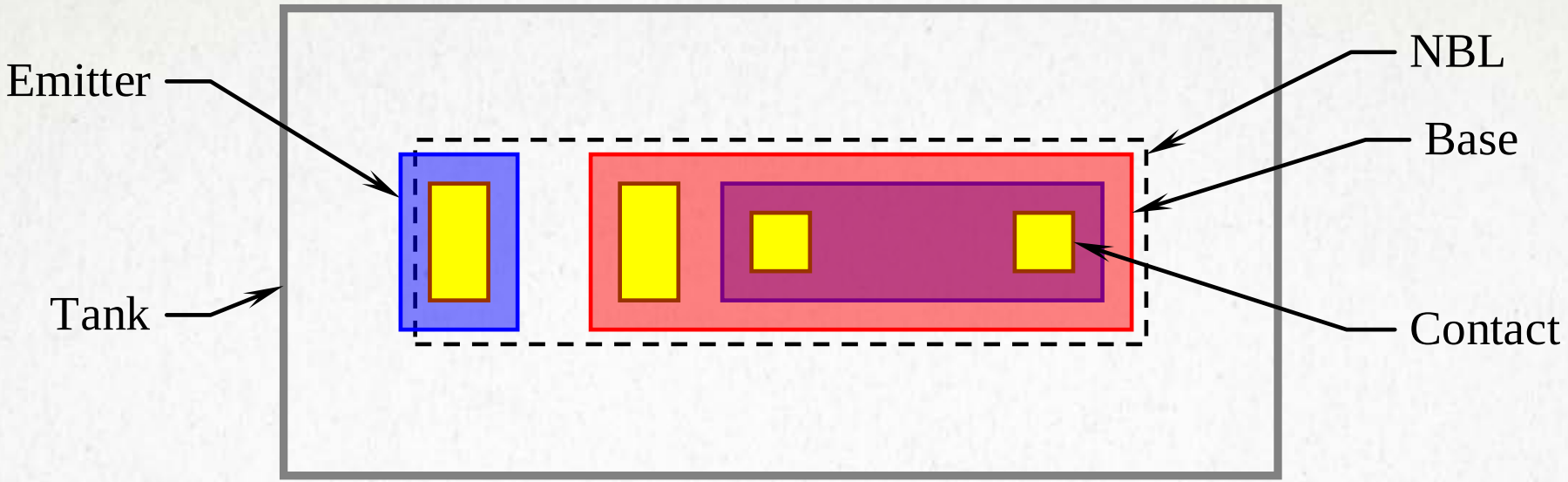


BASE RESISTORS



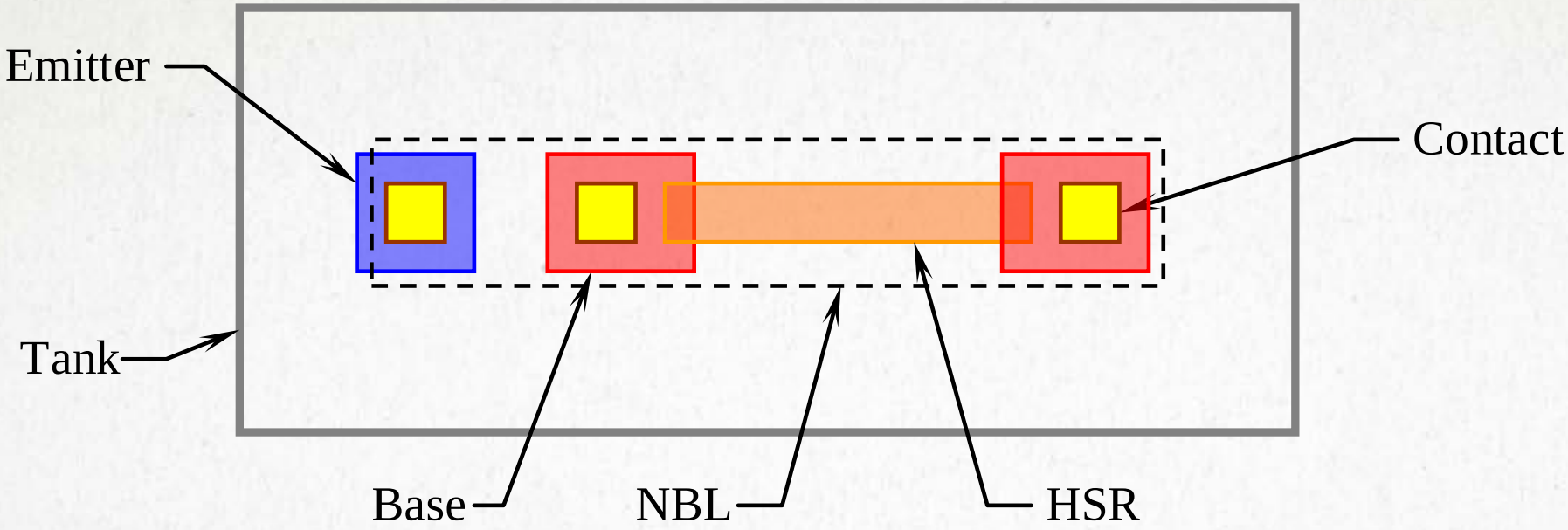
- **Base resistors are available in standard bipolar and in some analog BiCMOS processes.**
  - Standard bipolar base sheet is typically 160 Ω/☐ , BiCMOS 300–600 Ω/☐.
  - Best choice for moderate resistances (50 Ω–10 kΩ) in standard bipolar.

EMITTER RESISTORS



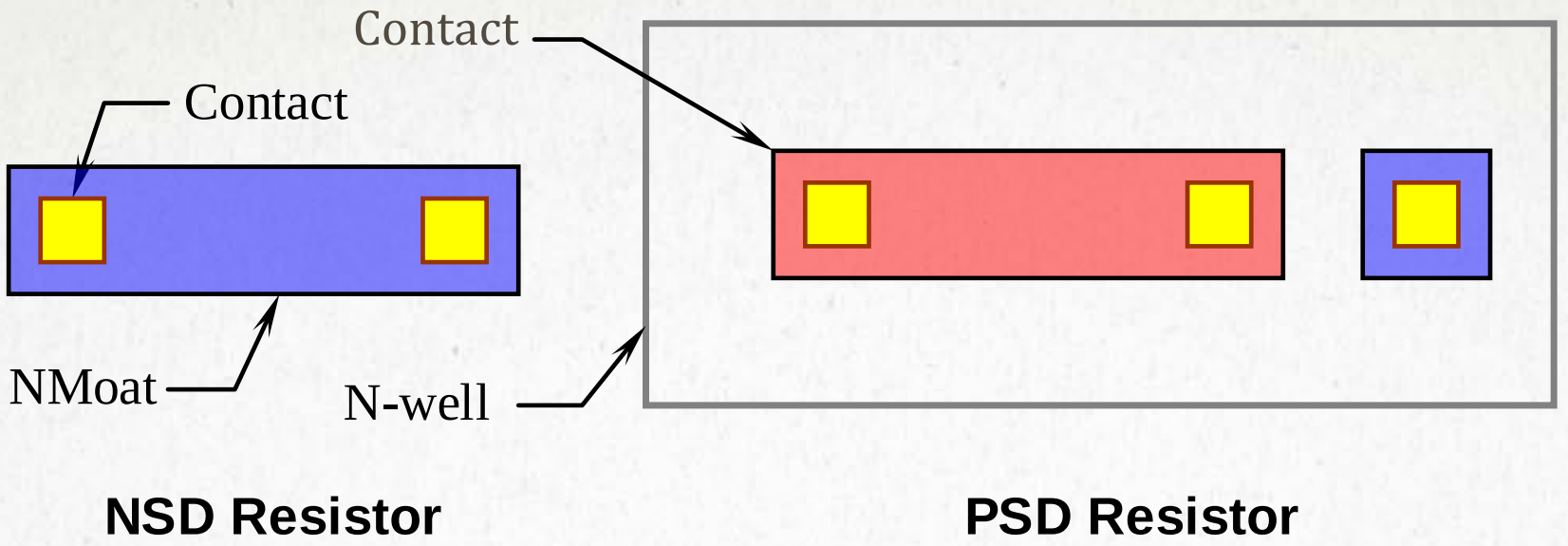
- **Emitter resistors are available in standard bipolar.**
  - Emitter sheet typically 2–10 Ω/□.
  - Typically isolated inside base region; emitter-base voltage differential should not exceed about 4 V.

HSR RESISTORS



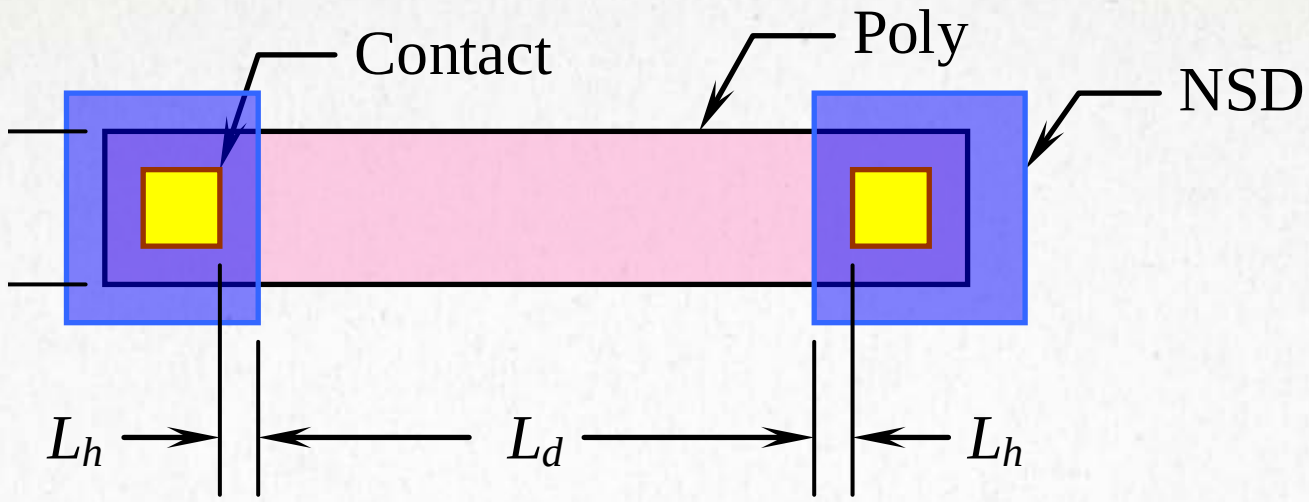
- **HSR resistors are available as an extension to standard bipolar.**
  - HSR (high sheet resistor) implant sheets range from 1–10 kΩ/☐.
  - The best sheet is about 2 kΩ/☐; lower sheets don't pack in enough resistance and higher sheets are too susceptible to modulation effects.

NSD AND PSD RESISTORS



- **CMOS and BiCMOS processes can make NSD and PSD resistors.**
  - If NMoat and PMoat are not silicided, their sheets are typically 20–50 Ω/□.
  - If they are silicided, their sheets are about 2 Ω/□.
  - These resistors are sometimes employed in ESD circuits because they can handle more pulse power than poly resistors (poly is thermally insulated by oxide).

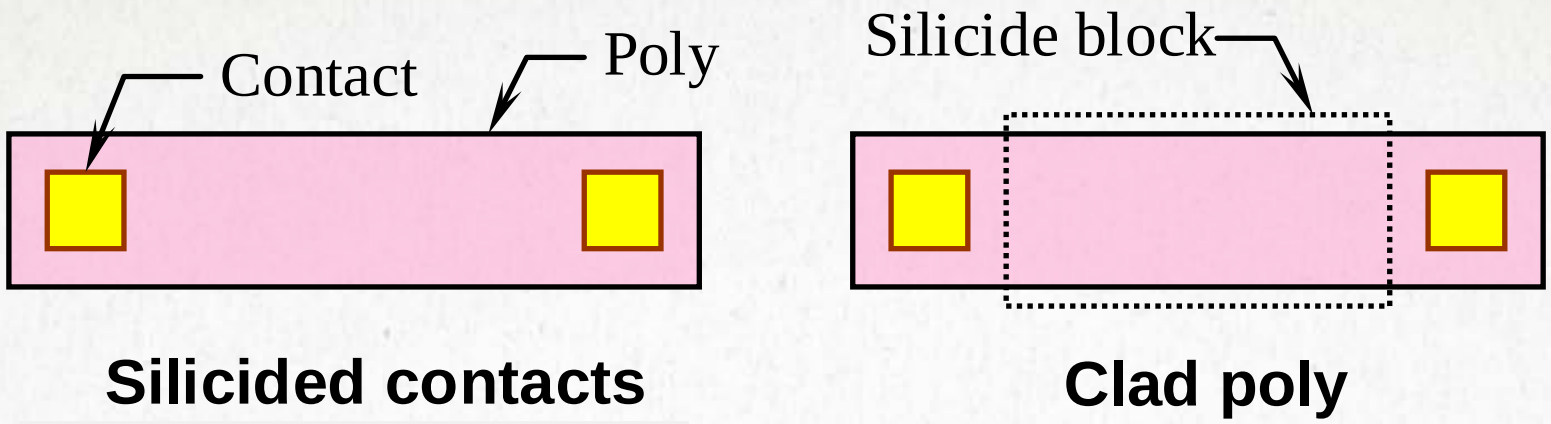
HSR POLY RESISTORS



- **Blocking silicide and gate doping permit a range of sheets.**
  - Sometimes poly is intrinsically doped in the epi reactor; it can also be doped by adding an implant such as NSD or PSD.
  - The resistance of an HSR poly includes separate body and head terms:

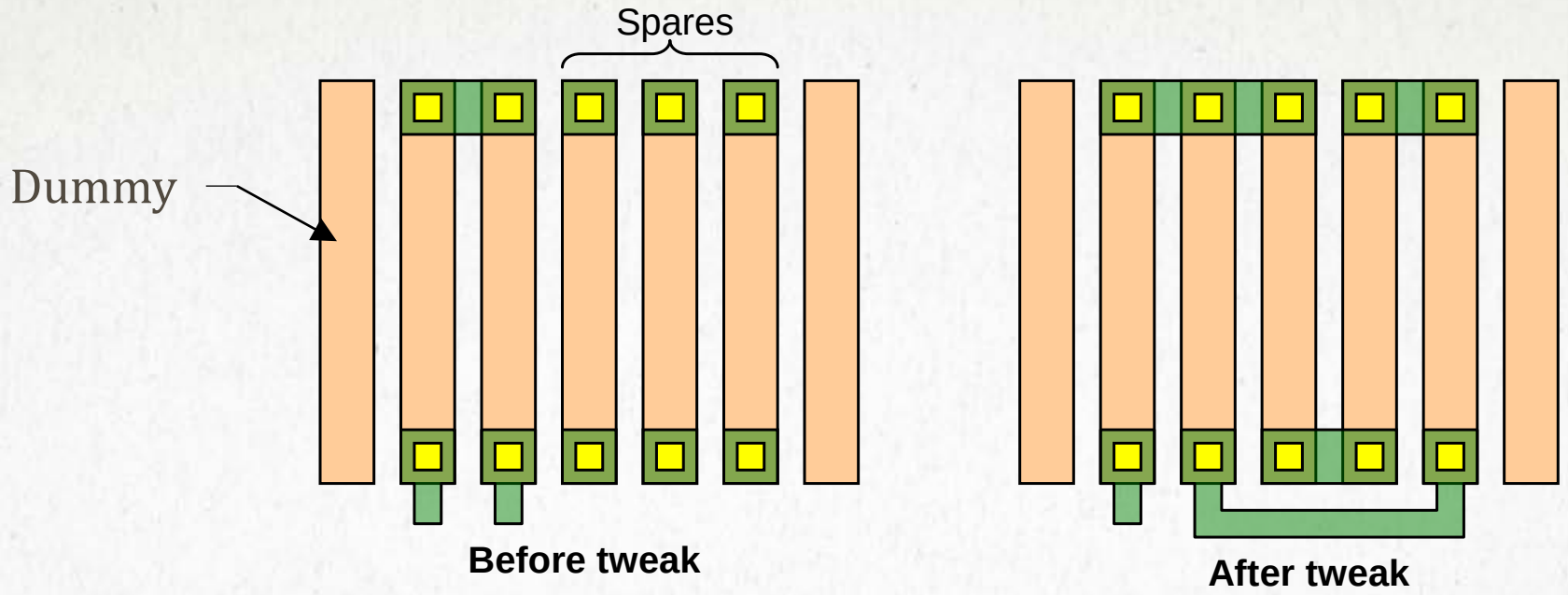
$$R = R_s\left[\frac{L_d - 2L_b}{W_d + W_b}\right] + 2R_h\left[\frac{L_h + 2L_b}{W_d + W_b}\right]$$

POLY RESISTORS



- **CMOS and BiCMOS processes can use polysilicon as a resistor.**
  - If only contacts are silicided, the gate dopant implant usually yields a sheet of 5–20 Ω/□.
  - If the poly itself is silicided (clad poly, or salicided poly), then the sheet is typically about 2 Ω/□.
  - These are low sheet resistances, but poly resistors have small widths and spacings.

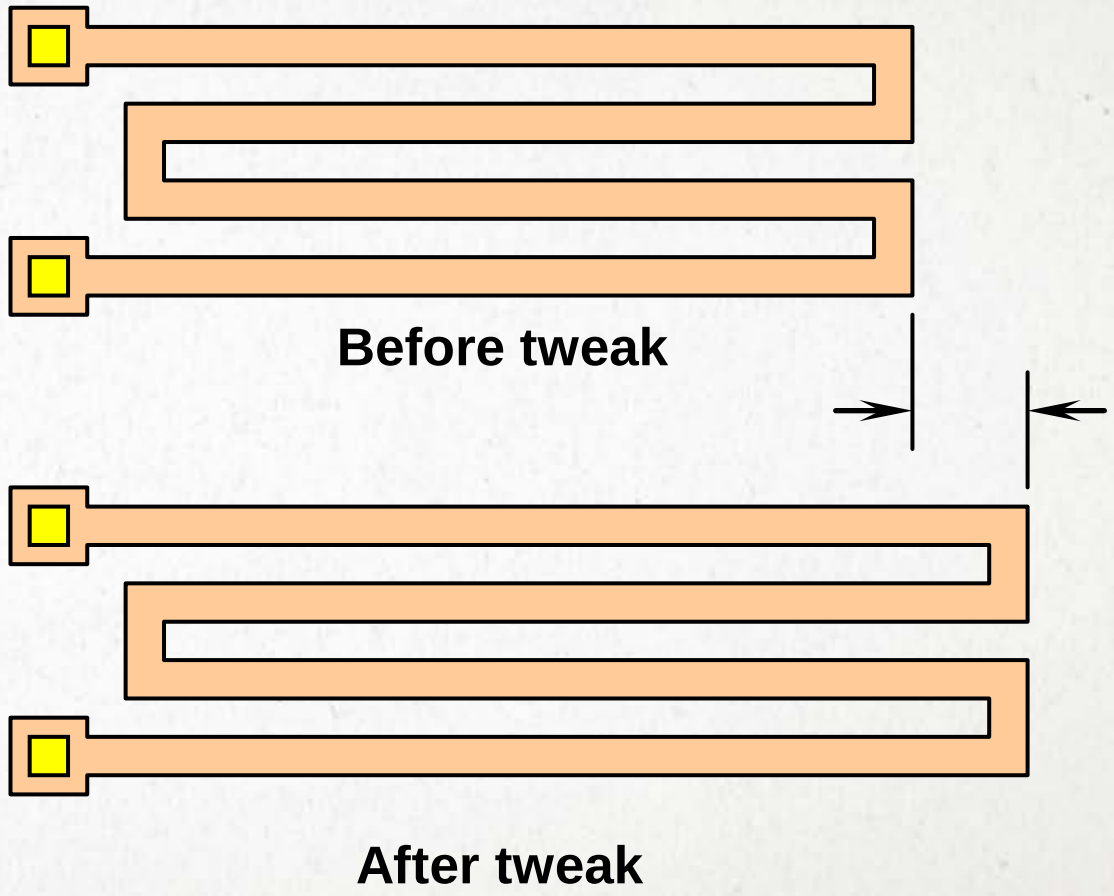
SEGMENTED RESISTORS



- **Resistors are often divided into multiple segments to create a compact structure.**
  - The segments are connected together with metal jumpers.
  - *Dummy resistors* are often placed on either end of the array to improve matching.
  - Extra unused segments (*spares*) are often included to permit one-mask tweaks.

SERPENTINE RESISTORS

- **Alternatively, a resistor can be bent back and forth to create a *serpentine resistor.***
  - Not a good choice for accurate matching, but good for packing resistance in tightly.
  - If the body of the resistor is to narrow to contain contacts, heads can be enlarged (*dogbone heads*).
  - Tweaks can be achieved by a so-called *trombone slide* (illustrated).

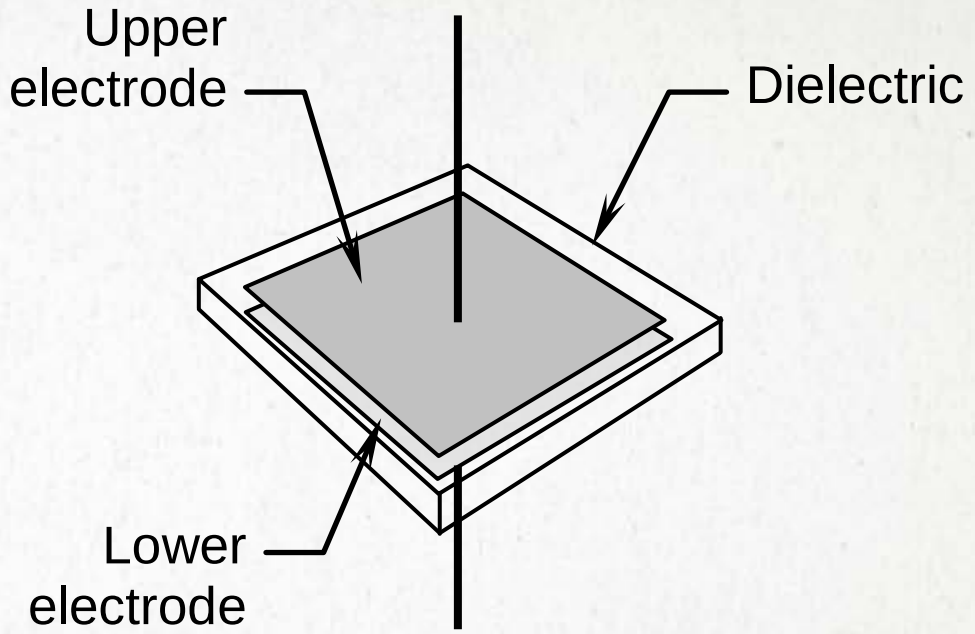


PARALLEL-PLATE CAPACITOR

- **A parallel-plate capacitor consists of an insulating layer called a *dielectric* sandwiched between two conductive electrodes.**
  - The capacitance C equals approximately

$$C \cong 0.0885 \frac{A\varepsilon_r}{t}$$

where A is the area of either electrode in μm^2, t is the thickness of the dielectric in Angstroms (Å), and ε_r is a dimensionless constant called the *relative permittivity* or *dielectric constant*.



DIELECTRICS

Material		Relative permittivity	Dielectric strength (MV/cm)
Silicon		11.8	30
Silicon dioxide (SiO_2)	Dry oxide	3.9	11
	Plasma	4.9	3–6
	TEOS	4.0	10
Silicon nitride (Si_3N_4)	LPCVD	6–7	10
	Plasma	6–9	5

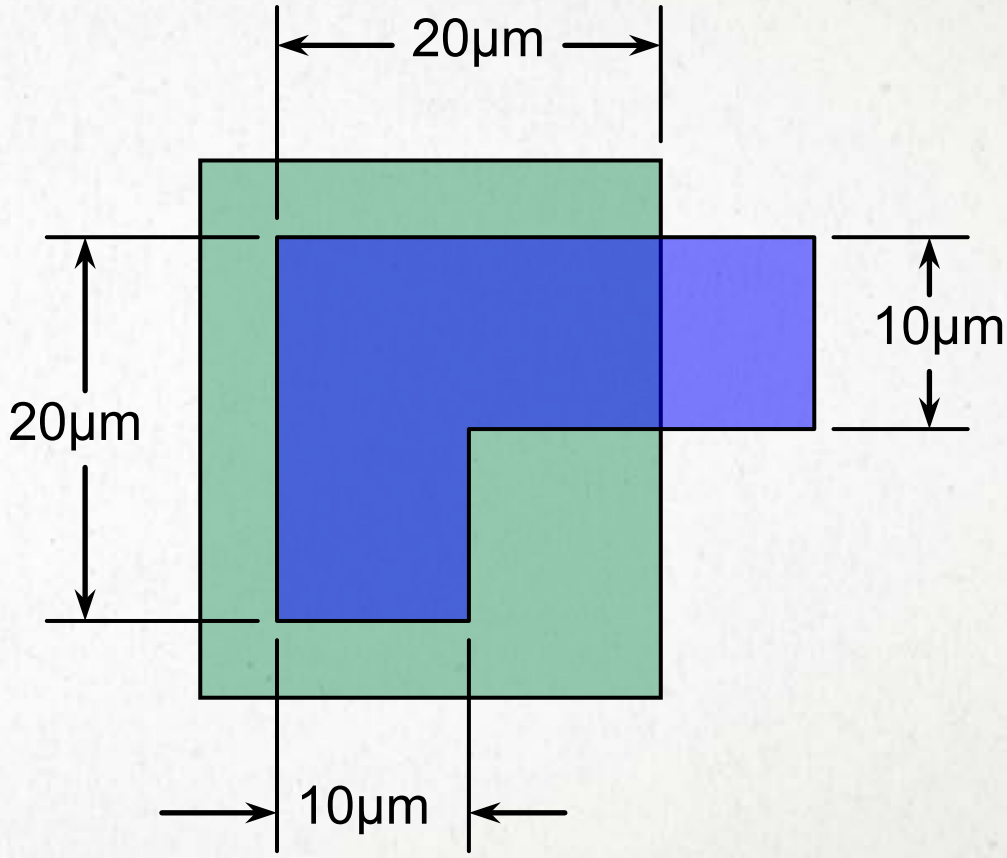
- **The maximum voltage V_{max} a capacitor can withstand equals**

$$V_{\max} = 0.01 t E_{crit}$$

where t is dielectric thickness in Angstroms and E_{crit} is the dielectric strength in MV/cm. For reliable operation, voltages should not exceed about a third of V_{max}.

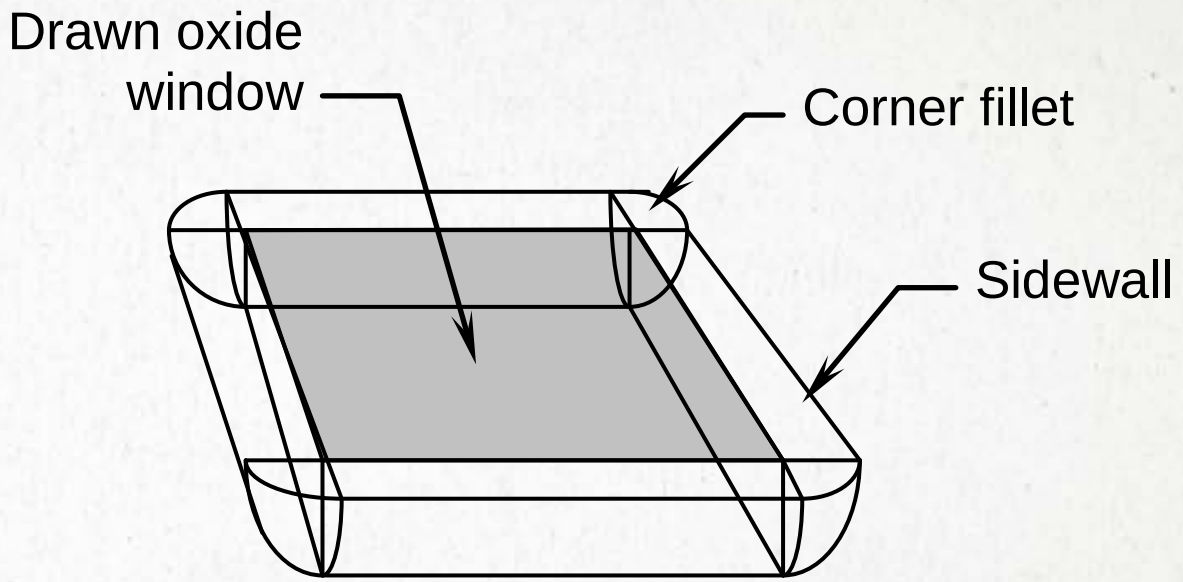
UNEQUAL PLATE AREAS

- **The effective area of a parallel plate capacitor equals the area of intersection of its plates.**
  - The illustrated capacitor has an area of 300 μm^2.
  - The actual capacitance is slightly larger due to so-called *fringing fields*.
  - Fringing fields don't contribute much to capacitors with dimensions of more than a couple of microns.



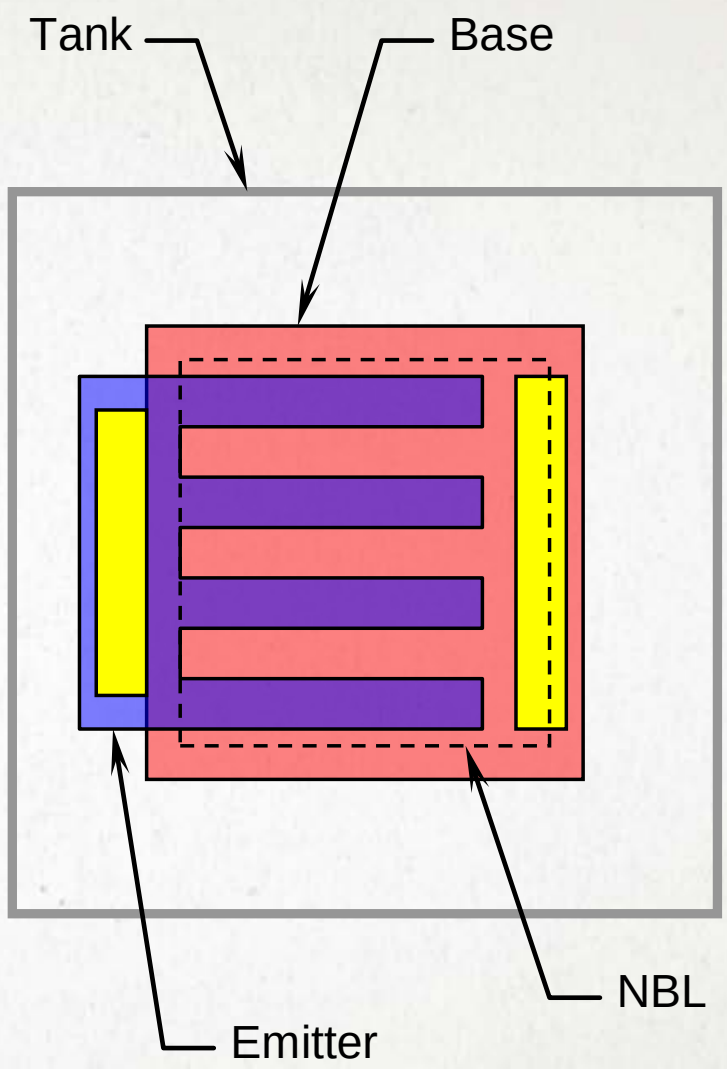
JUNCTION CAPACITANCE

- **A reverse-biased PN junction can be used as a capacitance.**
  - The area of such capacitors is hard to compute because of the contributions of sidewalls.
  - The capacitance decreases with increasing reverse bias because of the increasing width of the depletion region.
  - These capacitors are only used where accuracy is not required; for example, for compensation capacitors.

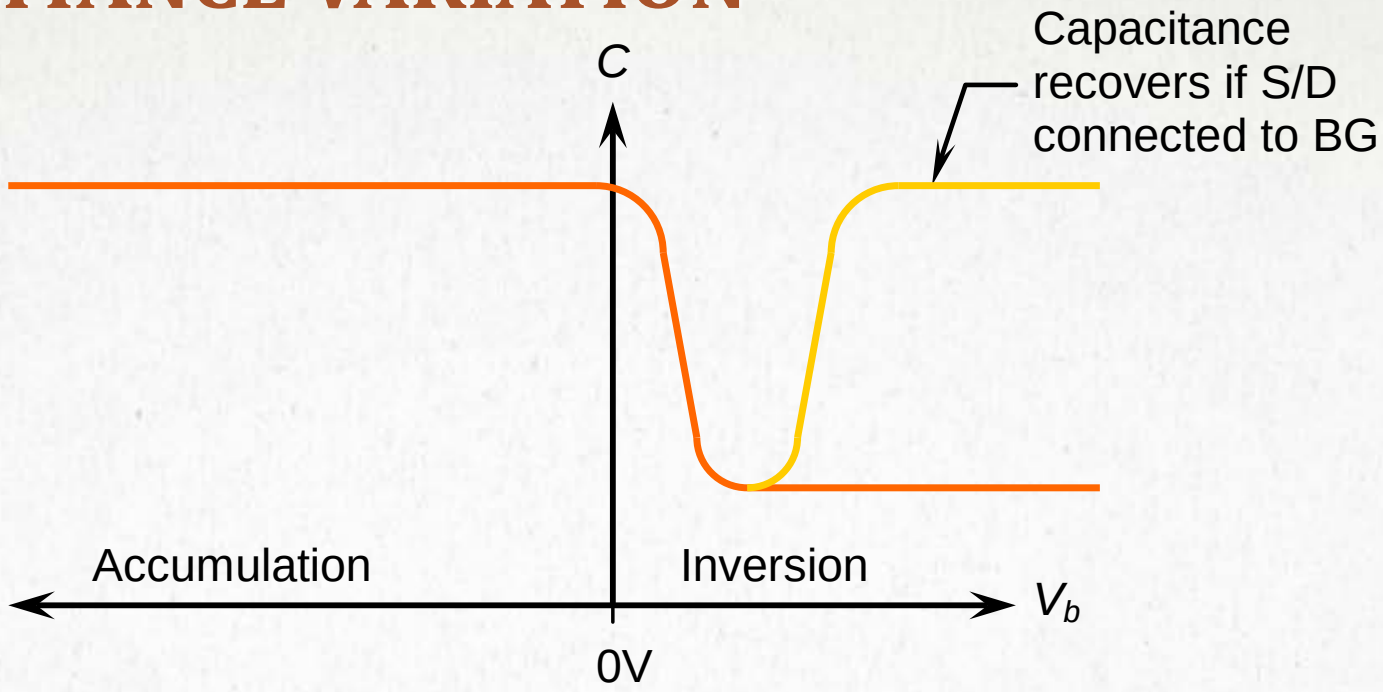


JUNCTION CAPACITANCE

- **Standard bipolar designs often use base-emitter capacitors.**
  - Capacitors of up to perhaps 100 pF can be easily constructed.
  - Total variation (process, voltage, and temperature) is typically 3:1.
  - The base-emitter junction must be kept reverse-biased, and the voltage across it should not exceed about half its breakdown rating (typically 7 V).



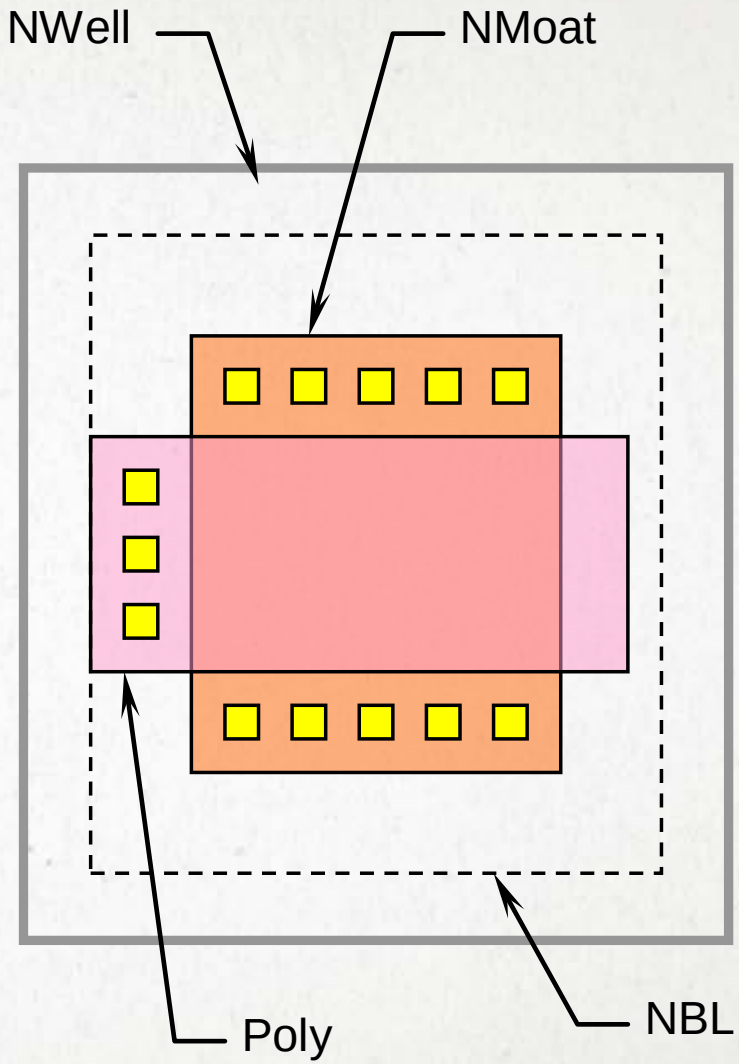
MOS CAPACITANCE VARIATION



- **MOS capacitance varies with voltage.**
  - In *accumulation*, the capacitance equals the oxide capacitance.
  - The capacitance drops as the device enters *depletion* because of the formation of the depletion region.
  - If source/drain connections exist, the capacitance recovers to the oxide capacitance in *inversion*.

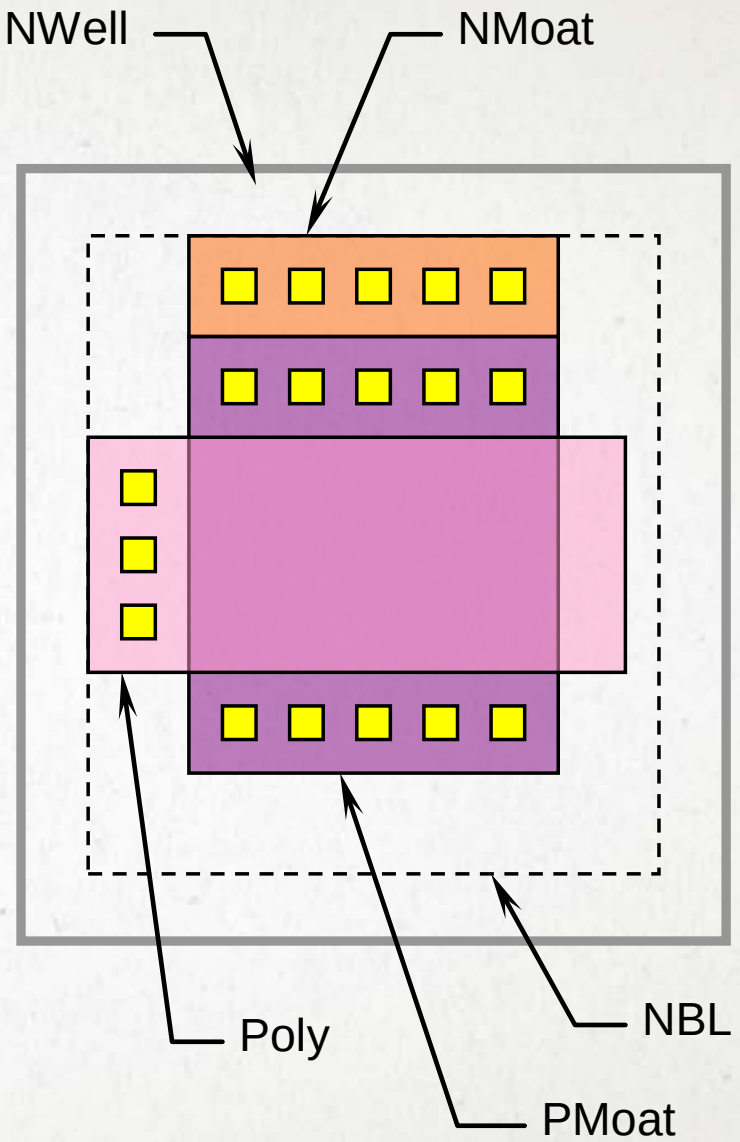
ACCUMULATION CAPACITORS

- **Most MOS capacitors are operated in accumulation.**
  - A typical accumulation capacitor uses NWell as its diffused electrode (called its *backgate*).
  - NMoat contacts provide electrical connection to the backgate.
  - This structure looks like a MOS transistor, but it isn't one.
  - To maintain maximum capacitance, the poly electrode should always be biased above the backgate.



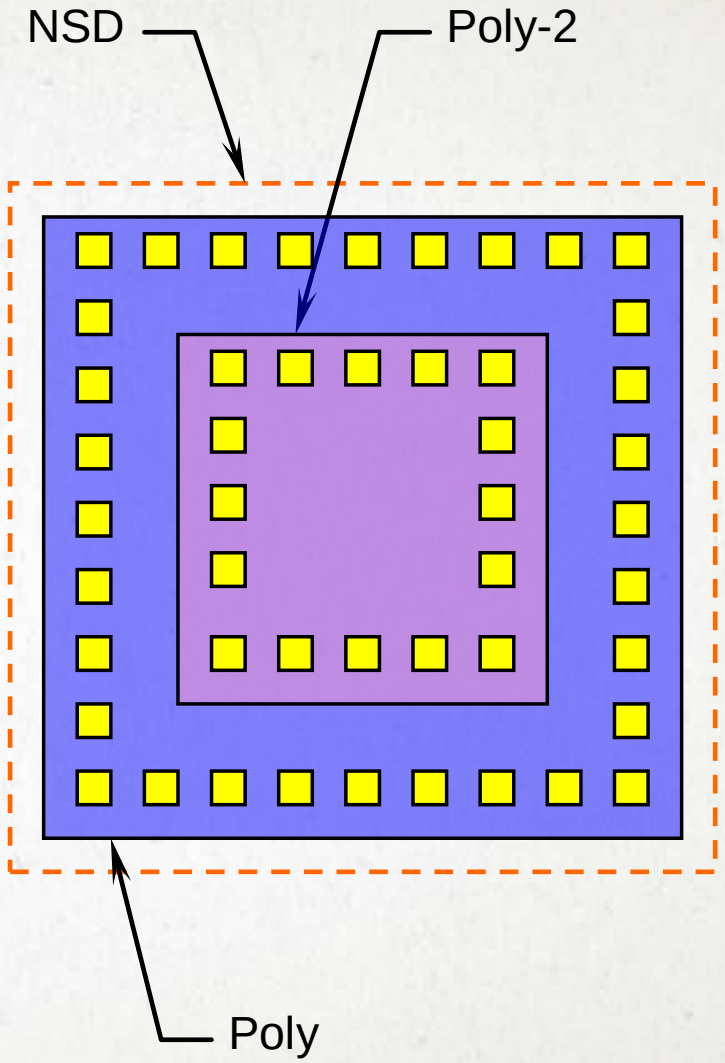
INVERSION CAPACITORS

- **An inversion capacitor is simply a MOS transistor operated in inversion.**
  - Both NMOS and PMOS transistors can be used as inversion capacitors; a PMOS capacitor is illustrated.
  - For an PMOS inversion capacitor, the poly electrode should be biased below the backgate to maximize capacitance.
  - For an NMOS inversion capacitor, the poly electrode should be biased above the backgate to maximize capacitance.

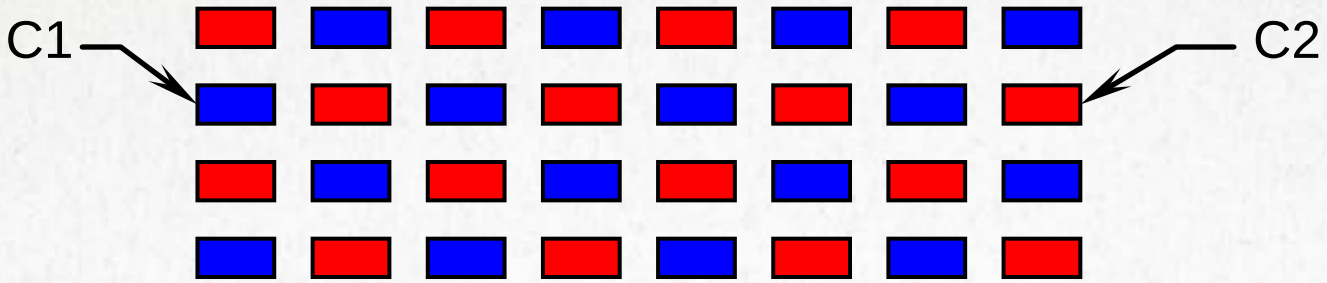


DEPOSITED CAPACITORS

- **Many analog CMOS and BiCMOS processes offer deposited capacitors.**
  - These capacitors are fully oxide isolated and have no parasitic PN junctions.
  - The dielectric material can be optimized for maximum capacitance at a desired operating voltage.
  - Dielectrics with higher dielectric constants (like nitride) provide more capacitance per unit area.



LATERAL FLUX CAPACITORS



- **Metal-metal capacitors can be constructed using the interlevel oxide.**
  - Typically one interdigitates fingers to maximize the lateral as well as the vertical capacitance between conductors.
  - All available metal layers are used to maximize capacitance.
  - Even with these measures, these so-called *lateral flux capacitors* are quite large, and capacitances are thus limited to a few tens of picofarads.

MATCHING

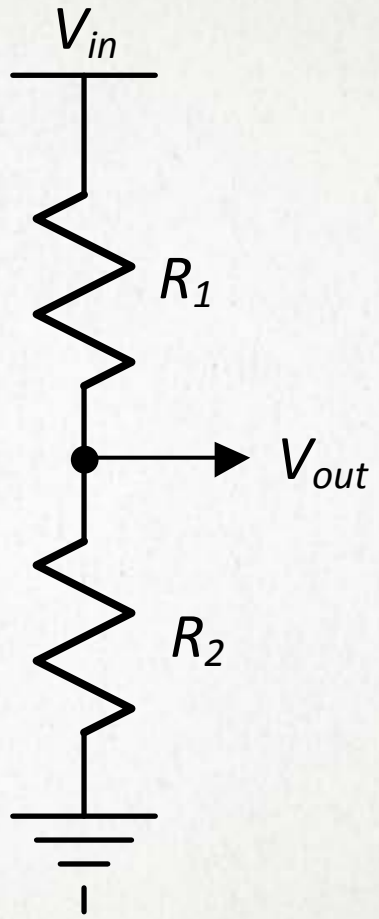
ALAN HASTINGS

MATCHING, PART I: INTRODUCTION

- **Analog integrated circuits depend upon device matching.**
  - The tolerances of most integrated components aren't very good; for example, integrated resistors typically vary by ±20%.
  - The matching between adjacent components can be much better; for example, integrated resistors can easily match within ±0.1%
  - Many analog circuits that benefit from matching have been developed specifically for integrated applications.
- **In order to achieve accurate matching, we must understand something about the mechanisms that cause mismatch.**

MATCHING VERSUS TOLERANCE

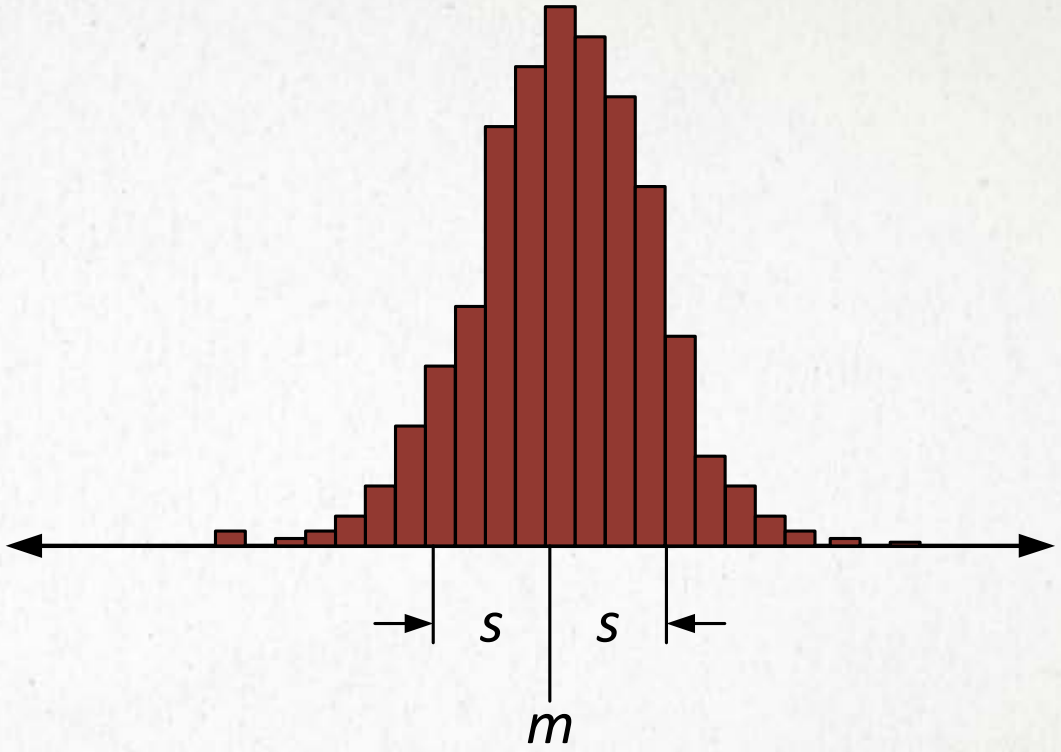
- **Consider resistor divider R_1–R_2.**
  - If $R_1 = R_2$, then $V_{out} = ½\ V_{in}$.
  - The values of R_1 and R_2 don't affect this relationship.
  - Instead, the ratio $R_2/(R_1+R_2)$ matters.
- **Components whose values track are said to *match*.**
  - Suppose R_1 and R_2 are supposed to have the same value.
  - If we measure $R_2 = 1.01\ R_1$, then the two resistors mismatch by 1%.



RANDOM MISMATCH

- **Resistance, capacitance, and other electrical parameters exhibit random variation.**
  - **This variation can be quantified by a *mean, m,* and a *standard deviation, s*.**
  - **For a sample of *N* measurements,**

$$m = \sum_{i=1}^{N} x_i \qquad s = \sqrt{\sum_{i=1}^{N} \frac{(x_i - m)^2}{N - 1}}$$



PELGROM'S LAW

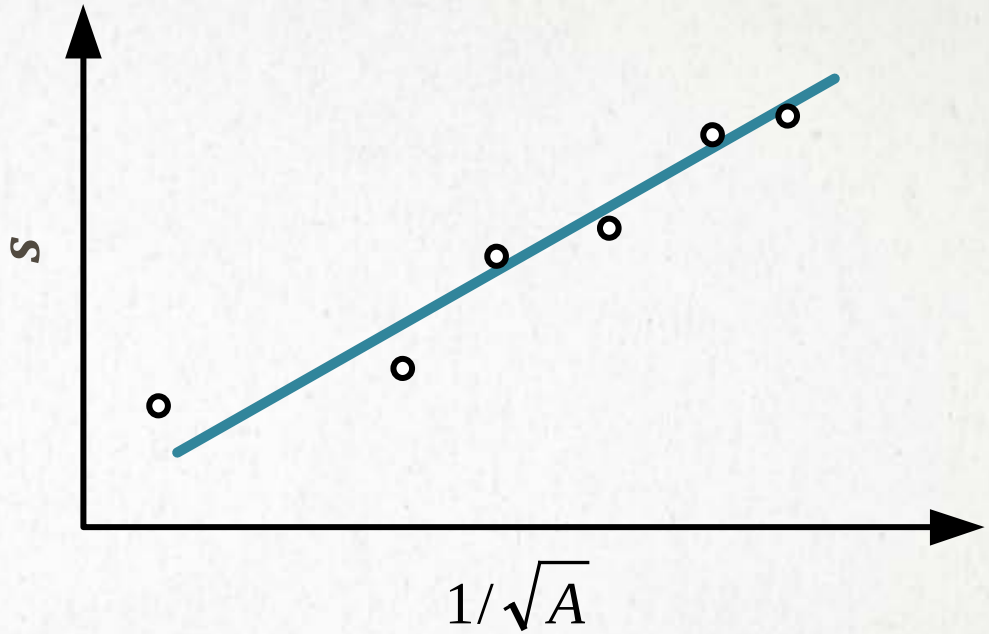
- **Most electrical parameters exhibit random variations due to areal fluctuations.**
  - If this is the case, then the standard deviation of an electrical parameter, say resistance R, equals

$$s_R = \frac{c_R}{\sqrt{A_R}}$$

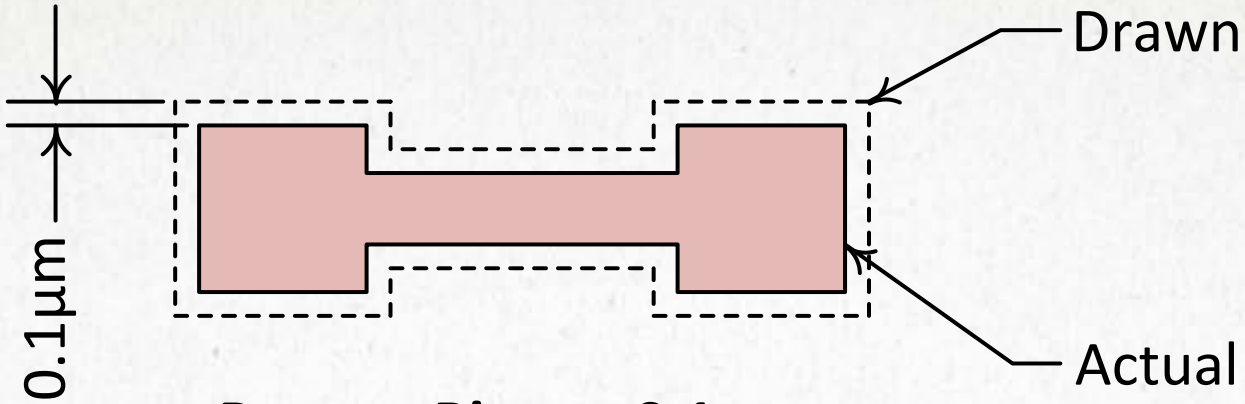
where c_R is a constant and A_R is the active area of the component, in this case the active area of a resistor.

- **The difference ΔR between two resistors R_1 and R_2 exhibits a standard deviation $s_{\Delta R}$,**

$$s_{\Delta R} = \sqrt{s_{R1}^2 + s_{R2}^2}$$

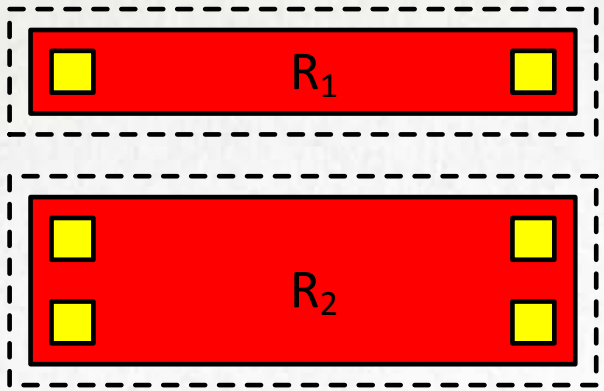


PROCESS BIASES

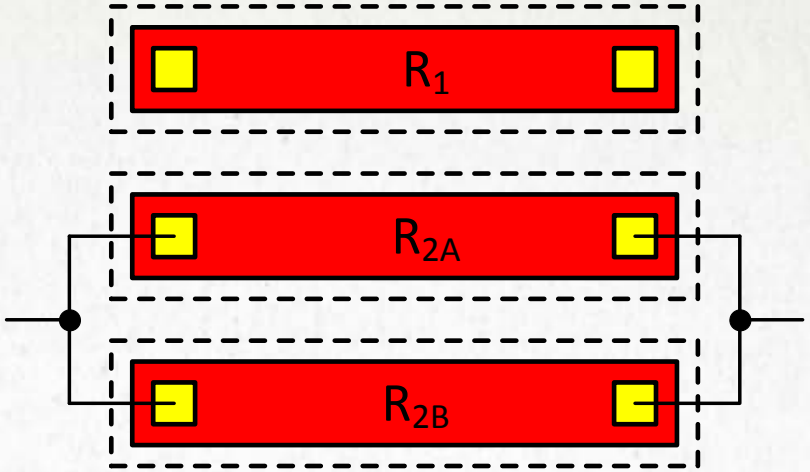


- **Actual dimensions seldom match drawn dimensions.**
  - Causes include overetching, outdiffusion, straggle, etc.
  - The difference between actual and drawn dimensions is called *process bias*.
  - Process biases cause systematic mismatches.

PROCESS BIASES



Different widths = poor matching



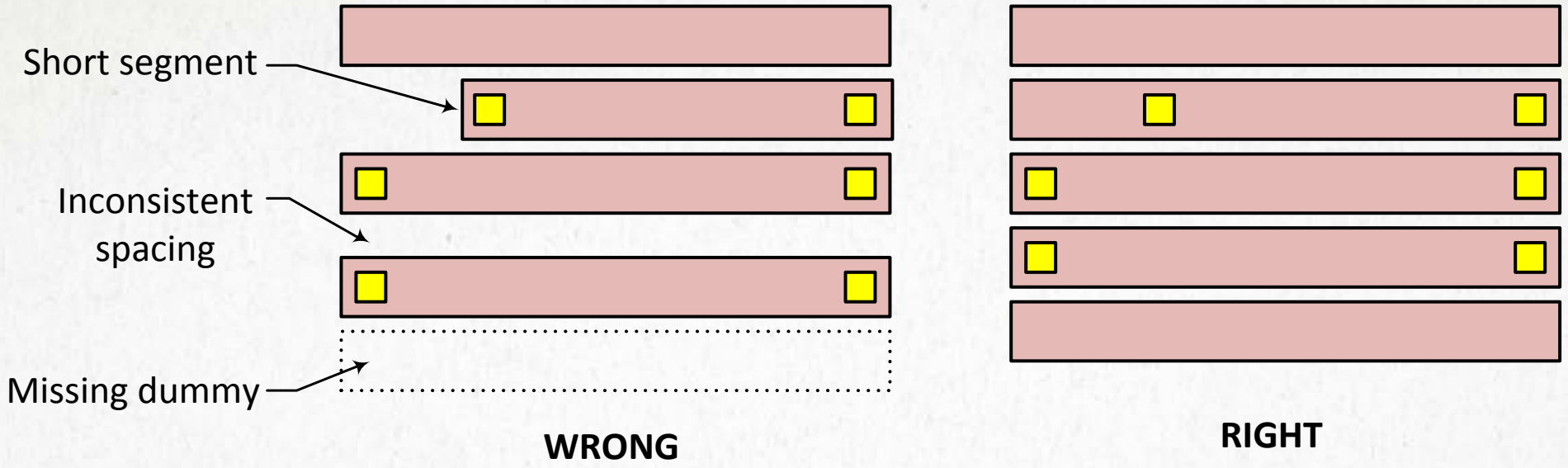
Same widths = good matching

- **Matched devices should have identical geometries to avoid mismatch due to process biases.**
  - Devices of different sizes can be constructed using multiple identical geometries connected in series or parallel.
  - Matching devices of arbitrary ratios may require segments that are almost, but not exactly, identical.

DUMMIES

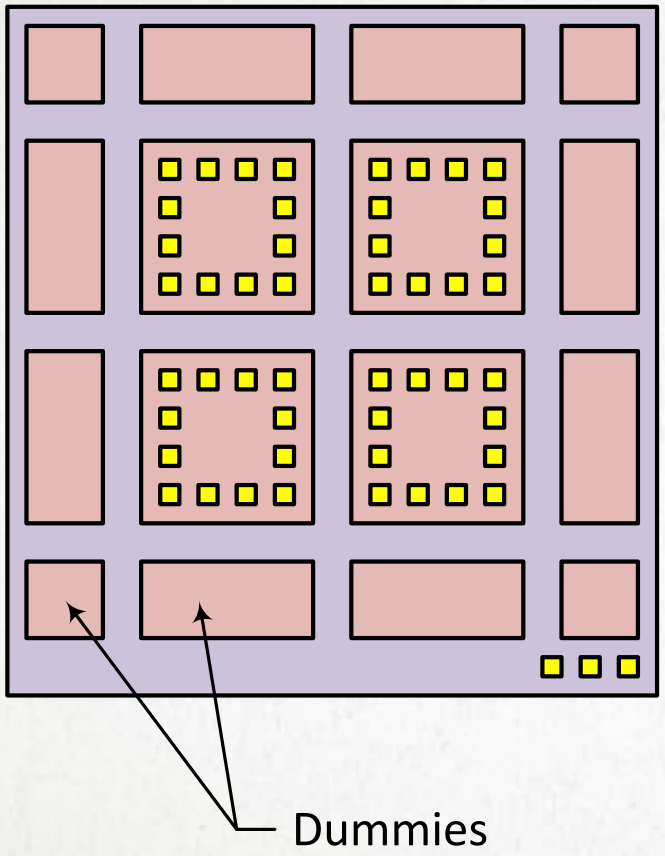
- **To ensure that proximity effects such as etch-rate variations do not cause mismatches,**
  - Place nonfunctional devices (*dummies*) on either end of the array.
  - The spacing between the dummies and the active devices must match the spacing between the active devices.
  - Very accurate matching may require multiple dummies on either end of the array because small proximity effects can extend 10 μm or farther.

DUMMIES: COMMON MISTAKES



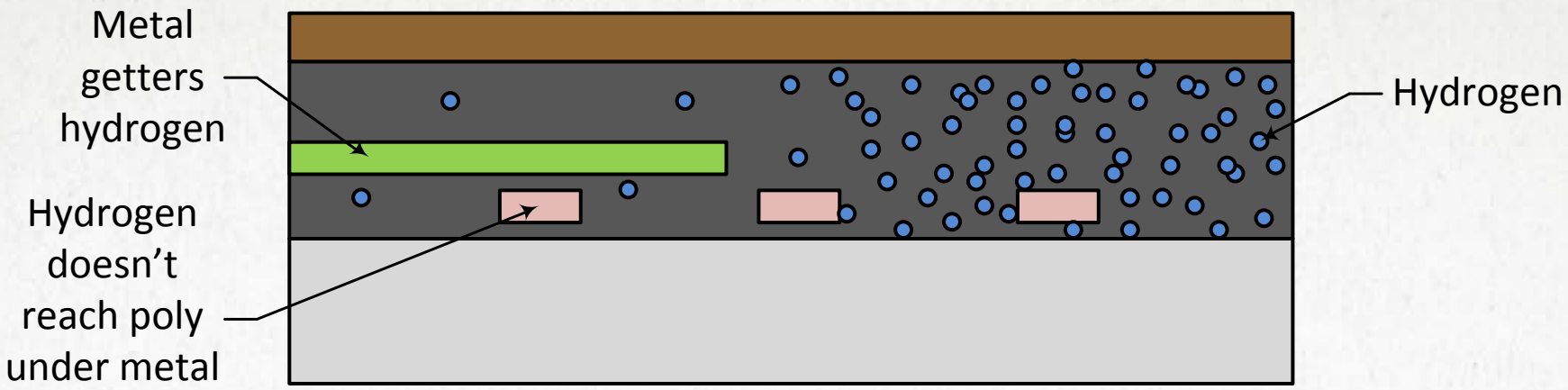
- **Common mistakes include:**
  - Missing dummies.
  - Not extending dummies the full length of the array.
  - Varying spacings of devices in the array.

DUMMY CAPACITORS



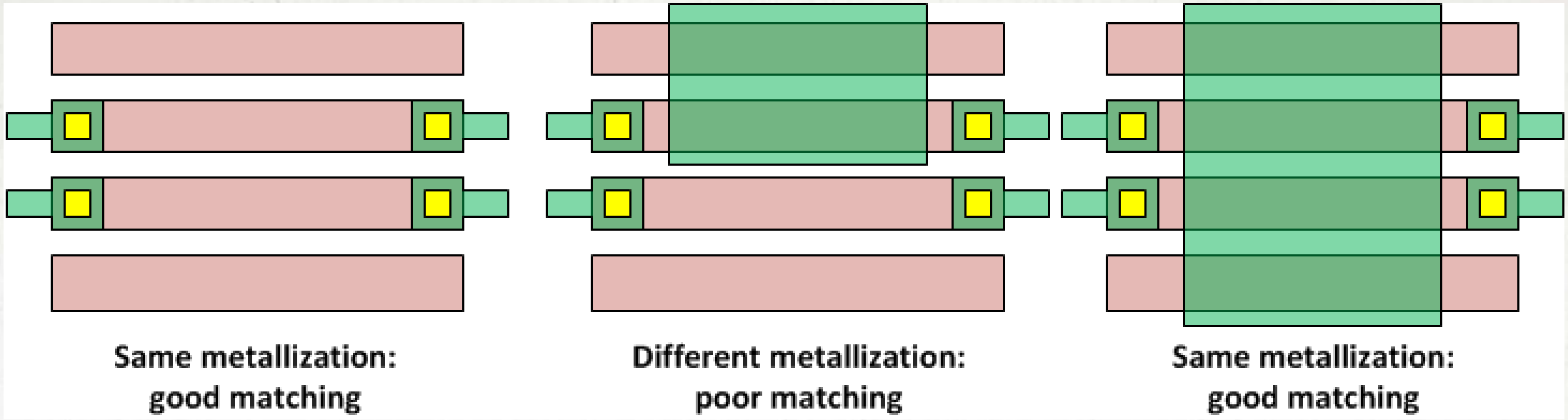
- **Deposited capacitors can also benefit from dummies.**
  - Capacitors are usually drawn as squares or rectangles of low aspect ratio.
  - Width and length are therefore of similar importance.
  - This means dummies must be placed on all four sides of the array.
- **Dummies don't have to be full size.**
  - Their function is mostly to maintain spacings between adjacent geometries.
  - For very accurate matching, the dummies should be at least 10 μm wide.

HYDROGENATION



- **Hydrogen affects the values of certain components.**
  - Hydrogen ties off dangling bonds at the oxide interface, reducing surface state charge and shifting MOS threshold voltages.
  - Hydrogen also ties off dangling bonds at grain boundaries in poly resistors, slightly shifting the value of high-sheet poly resistors.
- **The compressive nitride overcoat releases hydrogen during final anneal, but metal patterns can block this hydrogen from reaching the surface.**

METALLIZATION-INDUCED MISMATCHES



- **Poly resistors (or MOS transistors) covered with different metal patterns match poorly.**
  - Either block all metal over the active areas of these devices, or use metal field plates to cover the active areas of these devices.
  - Do not allow dummy metal generation algorithms to place figures over matched poly resistors or MOS transistors unless these devices are field-plated on a lower metal layer!

MATCHING, PART II: GRADIENTS

- **Many processing and environmental parameters vary gradually across the surface of the die.**
  - Examples include oxide thickness, temperature, and mechanical stress.
  - These gradual variations can be quantified by what mathematicians call a *gradient.*
- **Gradient-induced mismatches are a serious concern for analog integrated circuits.**
  - All circuits suffer from processing gradients, but these are usually small.
  - Devices containing power devices can suffer ***large*** thermal gradients.
  - Plastic-packaged dice can suffer ***large*** mechanical stress gradients.

VISUALIZING A GRADIENT

- **This drawing attempts to visualize a gradually varying parameter.**
  - The dark areas represent a higher value of the parameter, and the light areas a lower value.
  - A gradual and continuous change is observed as one moves across the die.
- **Mathematically, this parameter can be described as a *scalar field*, *P*. Its gradient equals**

$$\nabla P = \frac{\partial P}{\partial x}\vec{i} + \frac{\partial P}{\partial y}\vec{j} + \frac{\partial P}{\partial x}\vec{k}$$

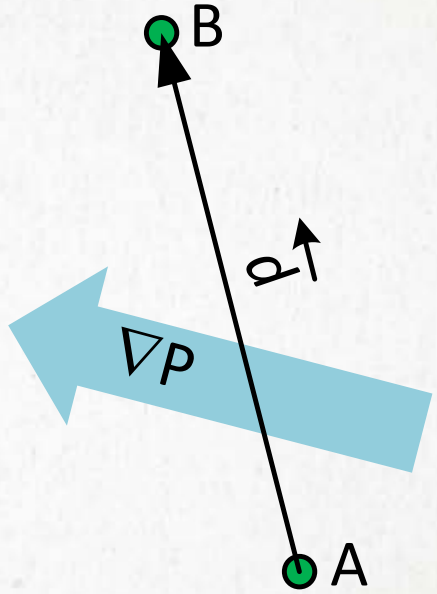
where $\vec{i}$, $\vec{j}$, and $\vec{k}$ are unit vectors directed along the x, y, and z axes.

THE IMPACT OF A GRADIENT

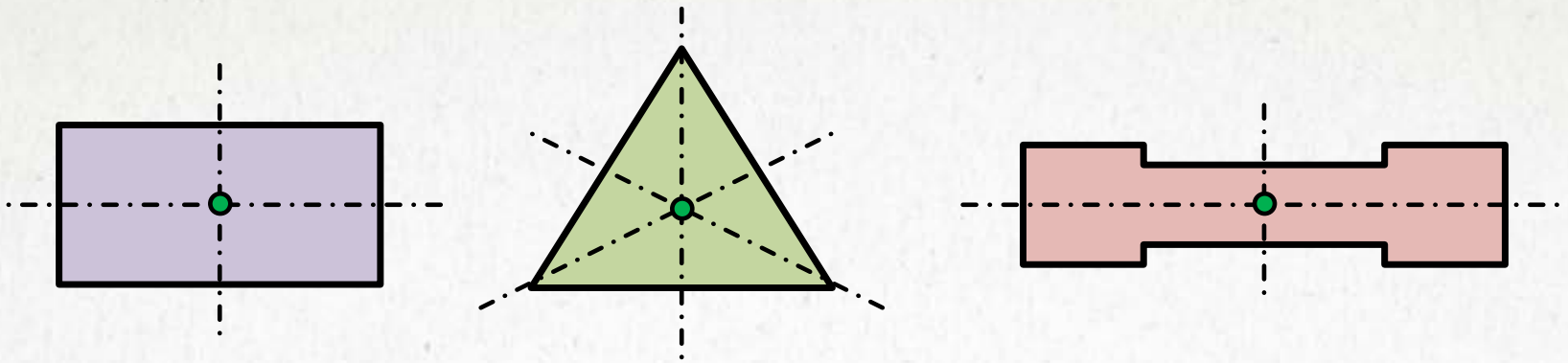
- **Suppose we wish to quantify the difference in parameter *P* between two locations A and B.**
  - We can erect a vector $\vec{d}$ from A to B and compute the quantity

$$\Delta P = \nabla P \cdot \vec{d}$$

  - Alternatively, we can measure the distance *d* in the direction of the gradient and use the amplitude of the gradient to compute ΔP.

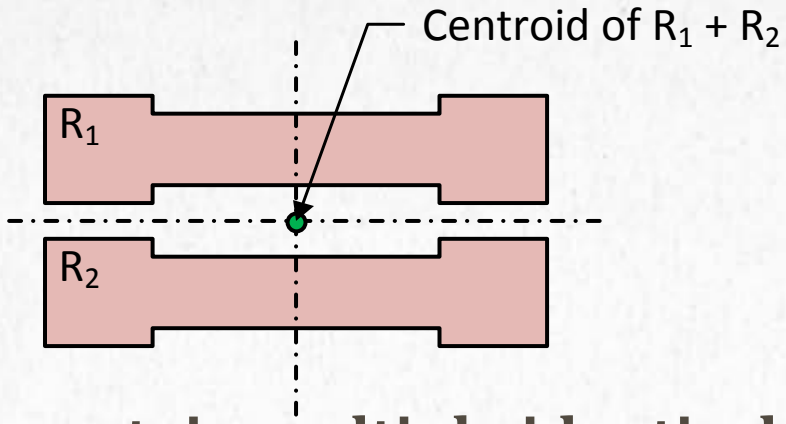


CENTROIDS



- **Real components aren't mere point-devices.**
  - However, a real component will ***behave*** as if it were a point device placed in a specific location, called the *centroid* of the device.
  - If the gradient is essentially linear, then the centroid lies somewhere on every axis of symmetry bisecting the active area of the device.
  - Simply find two axes of symmetry and the centroid lies at their intersection.

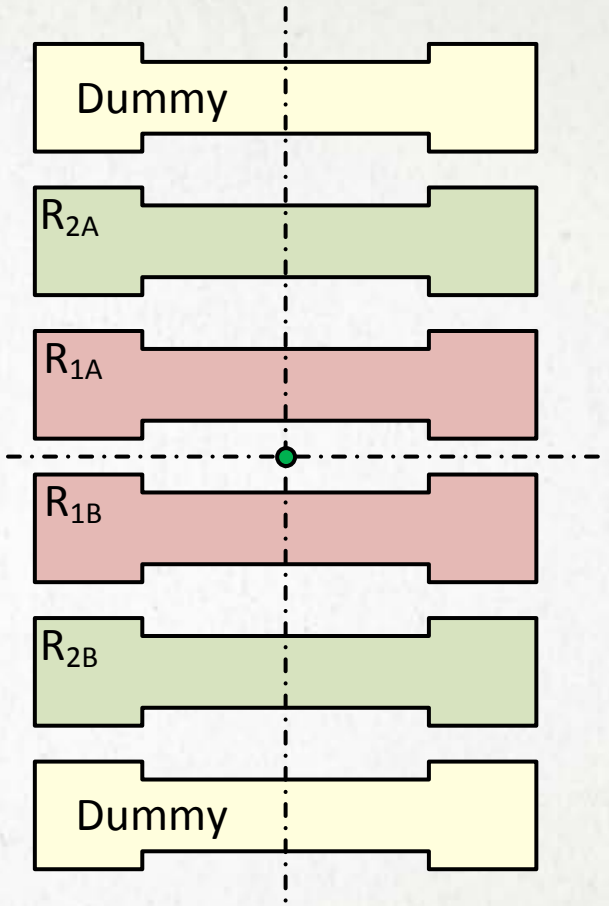
CENTROIDS OF SEGMENTED DEVICES



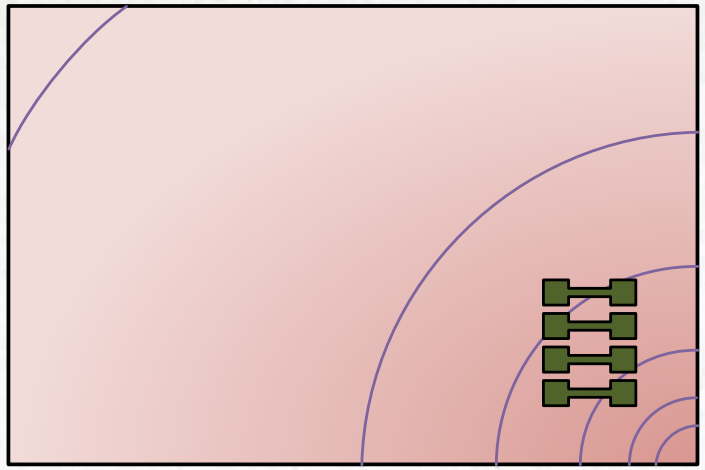
- **Arrayed devices contain multiple identical sections.**
  - Such an arrayed device has a centroid.
  - This centroid can be found like any other: it lies on every axis of symmetry passing through the array.
  - However, the centroid of an arrayed device doesn't necessarily fall inside any of the segments.
  - This suggests that it is possible for two arrayed devices to have the same centroid….

COMMON CENTROID ARRAYS

- **A common-centroid array consists of two or more devices that have been arrayed so that their centroids fall at the same point.**
  - Consider the array at the right.
  - R_1 and R_2 each have two segments, R_{1A}/R_{1B} and R_{2A}/R_{2B}.
  - By arranging these segments symmetrically about the center of the array, R_1 and R_2 can be made to have a common centroid.



BENEFITS OF COMMON-CENTROID ARRAYS

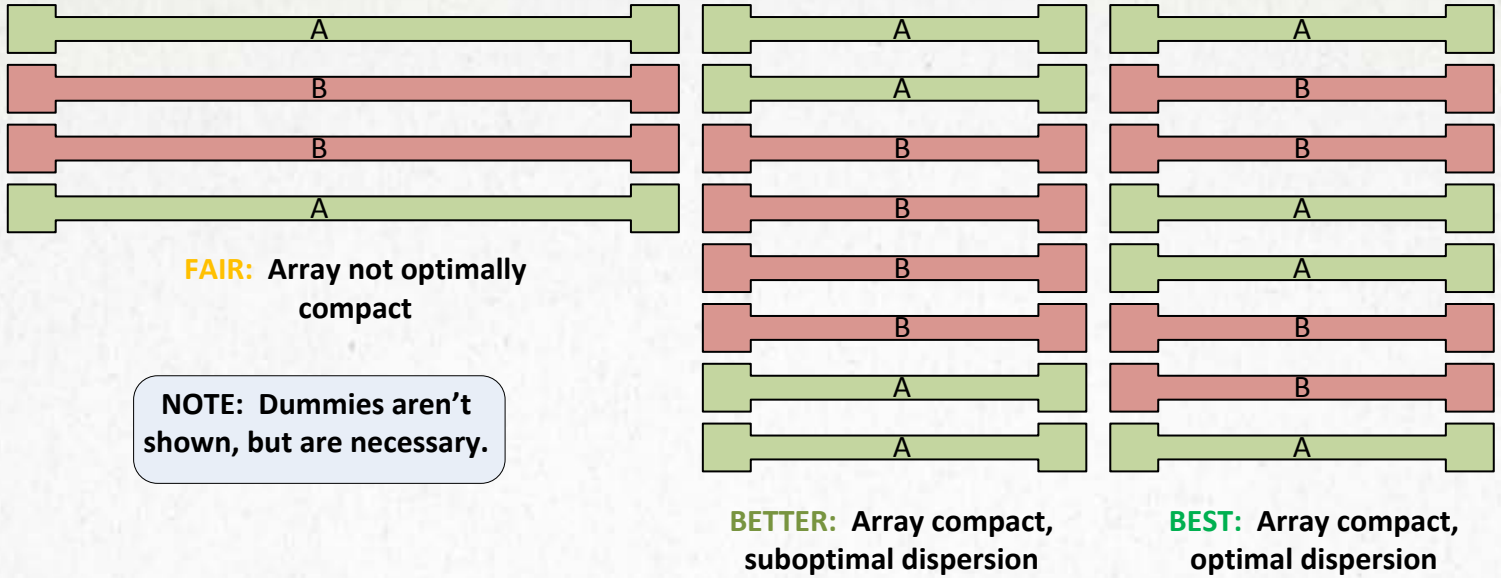


- **Gradients should have no effect upon common-centroid devices because they effectively occupy the same location.**
  - This is only approximately true because we assumed the gradient was linear in order to find the centroids.
  - The largest nonlinear component of the gradient is usually the quadratic.
  - Therefore the residual mismatch caused by nonlinearities tends to scale as the square of the distance across the array. ***Smaller arrays match better!***
  - Arrays placed in areas with low gradients will also match better.

THE FOUR RULES OF COMMON-CENTROID LAYOUT

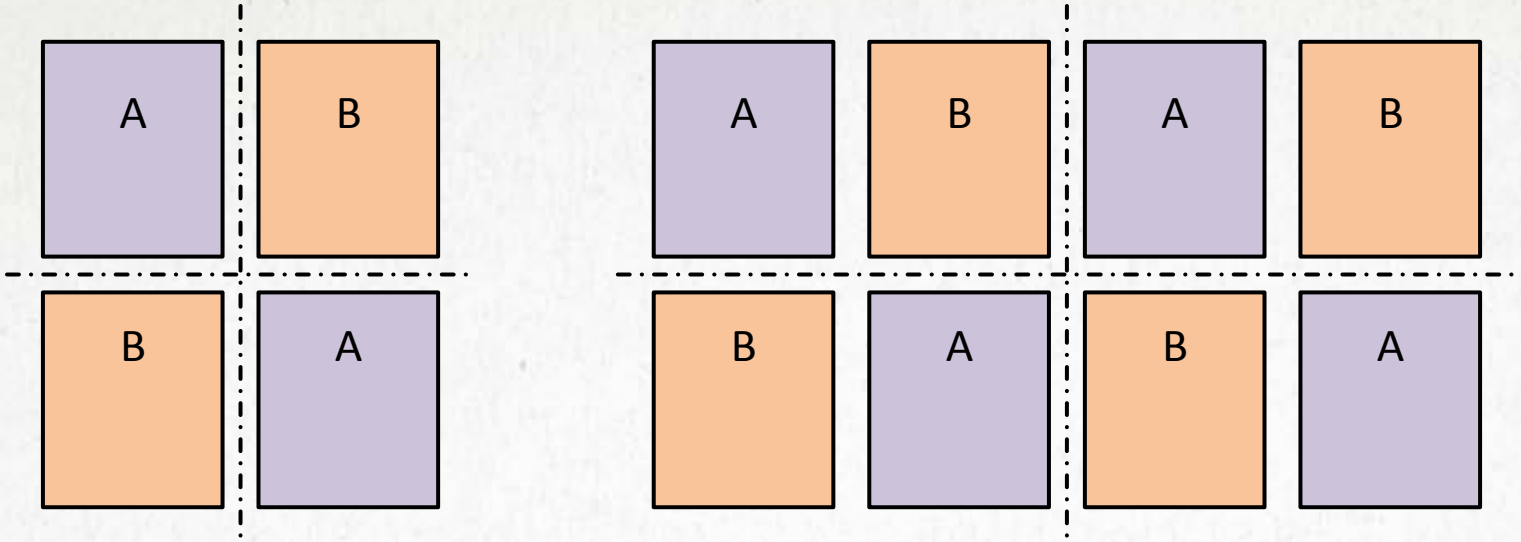
- ***Coincidence:*** The centroids of the matched devices should coincide at least approximately. Ideally, they should coincide exactly.
- ***Symmetry:*** The array should be symmetric around both the horizontal and vertical axes.
- ***Dispersion:*** When possible, a large array should be subdivided into as many smaller arrays that are possible that each satisfy the rules of coincidence and symmetry. If this can be achieved, then the larger array need not satisfy these rules; only the subarrays that comprise it need do so.
- ***Compactness:*** The array (or each of its subarray) should be as compact as possible.

EXAMPLE: RESISTOR ARRAY



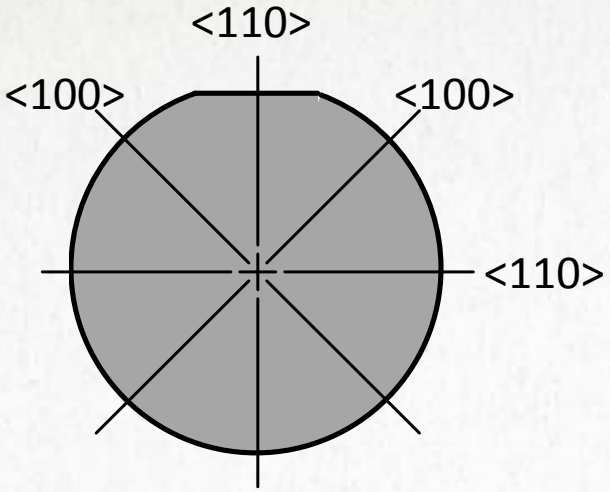
- **The three above arrays all use common-centroid layouts, but they aren't equally good layouts.**
  - Compactness and dispersion are important considerations, especially in large arrays where the nonlinear residues of the gradients are large.

2D COMMON-CENTROID ARRAYS

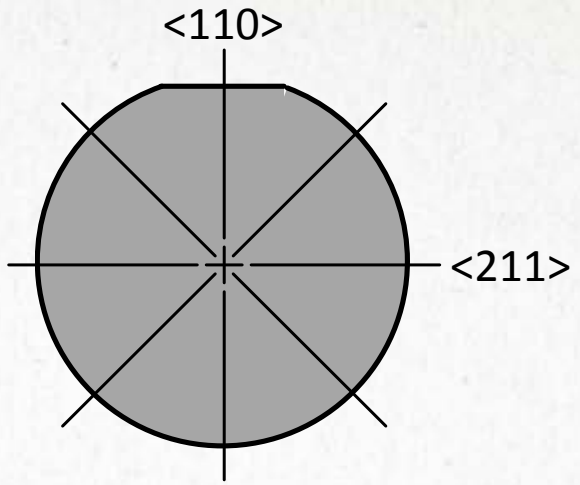


- **Compact devices, such as capacitors, can benefit from two-dimensional common-centroid layouts.**
  - The simplest such array, shown at the left, is often called a *cross-coupled pair.*
  - Multiple cross-coupled pairs can be combined into a larger array with good dispersion.

STRESS GRADIENTS



(100) Wafer



(111) Wafer

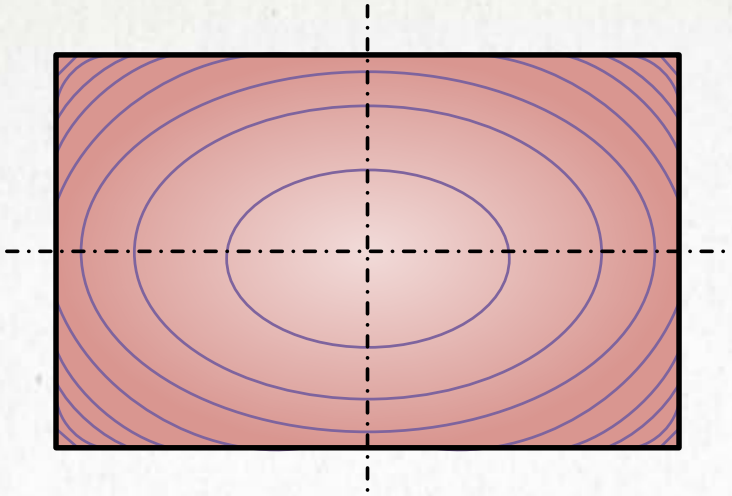
- **Mechanical stress can alter resistivity (piezoresistivity).**
  - N-type (100) silicon exhibits minimum piezoresitivity along <110> axes.
  - P-type (100) silicon exhibits minimum piezoresistivity along <100> axes.
  - (111) silicon, and polysilicon, do not exhibit an orientation dependence.

CAUSES OF MECHANICAL STRESSS

Material	Coefficient of thermal expansion
Epoxy encapsulation	24 ppm/°C (typical)
Copper alloys	16–18
Alloy 42	4.5
Molybdenum	2.5
Silicon	2.5

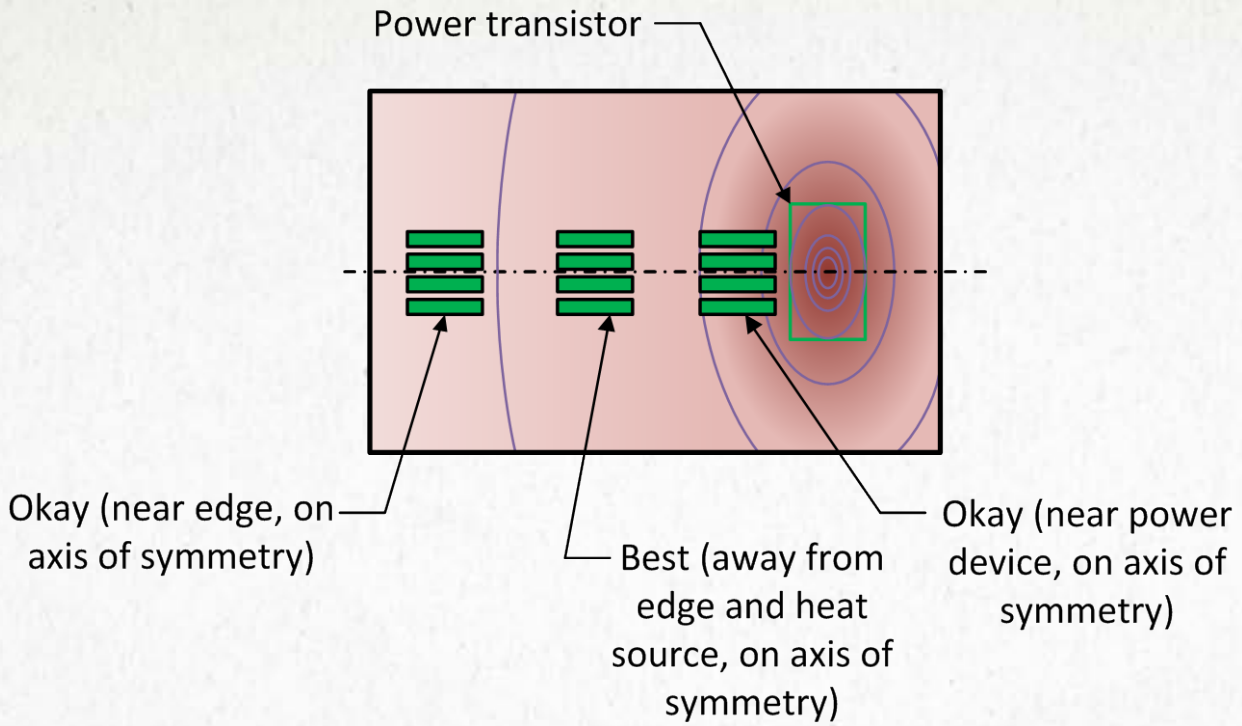
- **Differences in the coefficient of thermal expansion of different materials stress a packaged die.**
  - Plastic encapsulation is cured at about 175 °C.
  - As the packaged device cools, the encapsulant shrinks more than the silicon die.
  - Therefore the die experiences compressive stresses that increase at low temperatures and decrease at high temperatures.

STRESS GRADIENTS



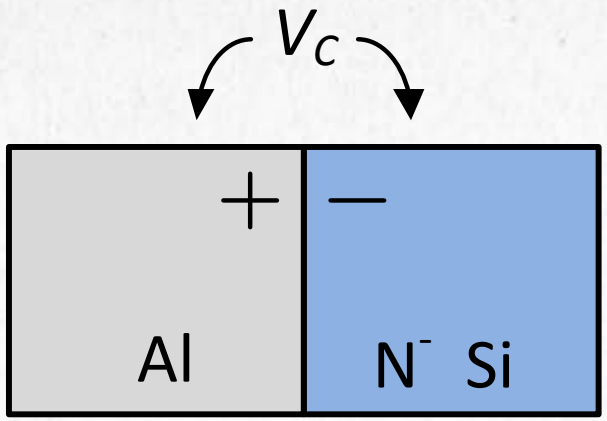
- **Stress gradients on the die are highest at the edges and especially in the corners.**
  - The actual stress gradients are more complex than this illustration suggests, but the rules are obvious:
  - Do not place matched devices along edges, or in corners.
  - For best matching, place matched devices near the center of the die such that an axis of symmetry of the array coincides with an axis of symmetry of the die.

THERMAL GRADIENTS



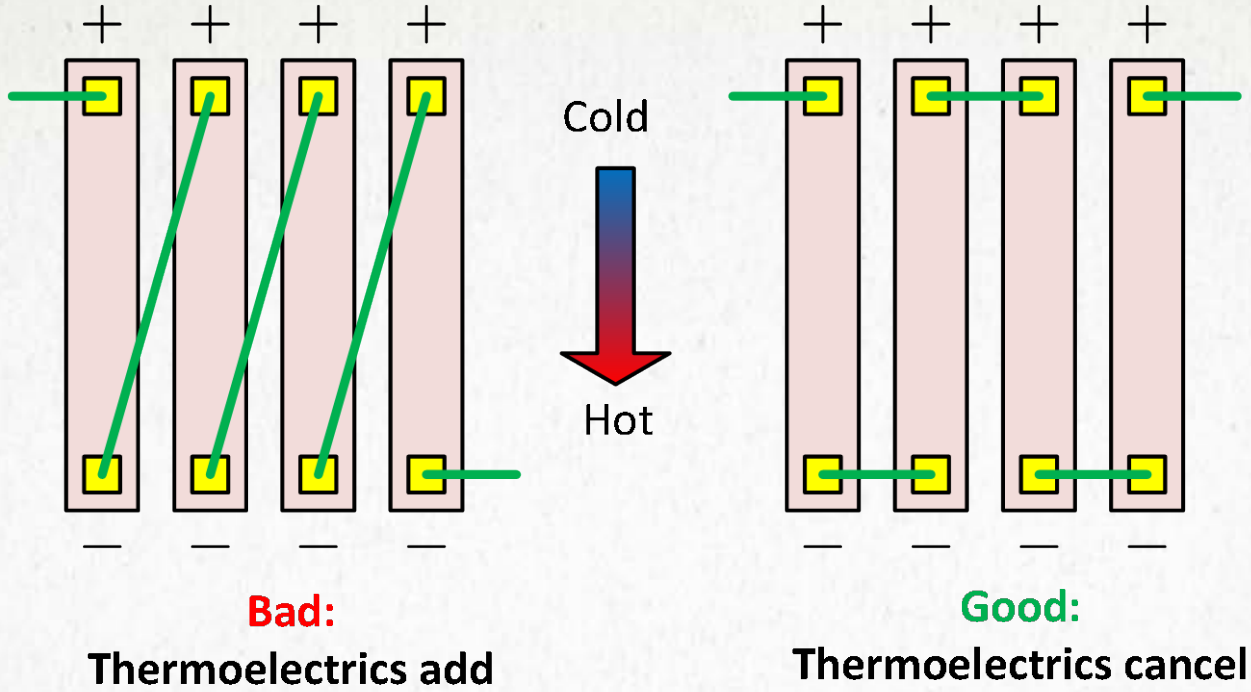
- **Ideally, matched arrays should lie on an axis of symmetry of major power devices, and as far from them as possible.**
  - Compromises must be made between locations optimal for stress and those optimal for temperature.

THE SEEBECK EFFECT



- **A voltage called the *contact potential* exists between any two dissimilar materials in contact with each other.**
  - Contact potentials vary with temperature.
  - The constant of proportionality is called the *Seebeck coefficient*.
  - The Seebeck coefficient of aluminum to silicon is typically 0.4 mV/°C.
  - This large Seebeck coefficient can cause mismatches in improperly connected resistors.

THERMOELECTRICS IN CONTACTS



- **In a series-connected resistor array, the contact potentials add.**
  - Improper connection causes a large thermoelectric voltage to accumulate due to addition of the contributions of each segment.
  - Proper connection eliminates this problem.

SUMMARY

- **Although this presentation focused upon resistors, most of the same principles apply to all matched components.**
  - Deposited capacitors aren't very susceptible to mechanical stress or temperature gradients.
  - MOS and bipolar transistors, on the other hand, are ***very*** sensitive to both of these types of gradients.
  - All matched components deserve careful consideration, but only the most critical such devices can occupy the best locations and consume sufficient area to minimize random mismatch.
- **No matter how carefully a circuit designer constructs their circuit, poor layout of its matched devices can ruin it!**

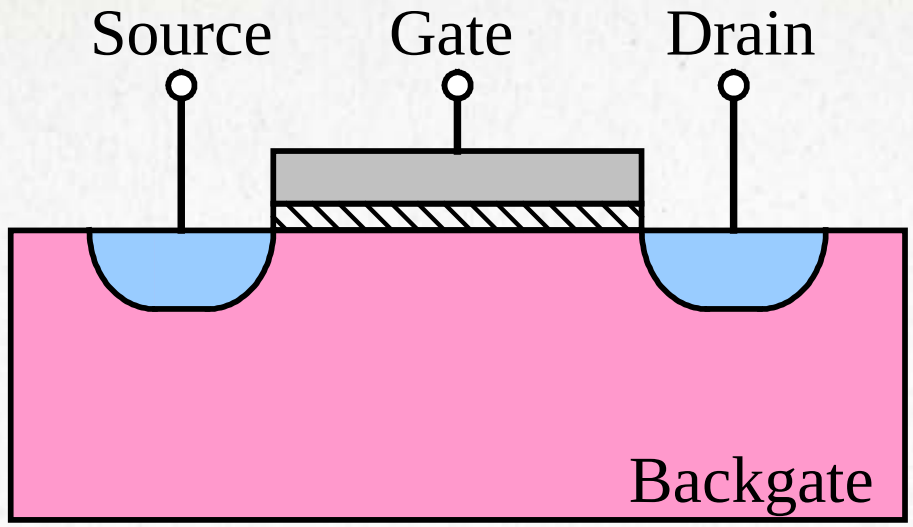
TRANSISTORS

ALAN HASTINGS

SECTION I: MOS TRANSISTORS

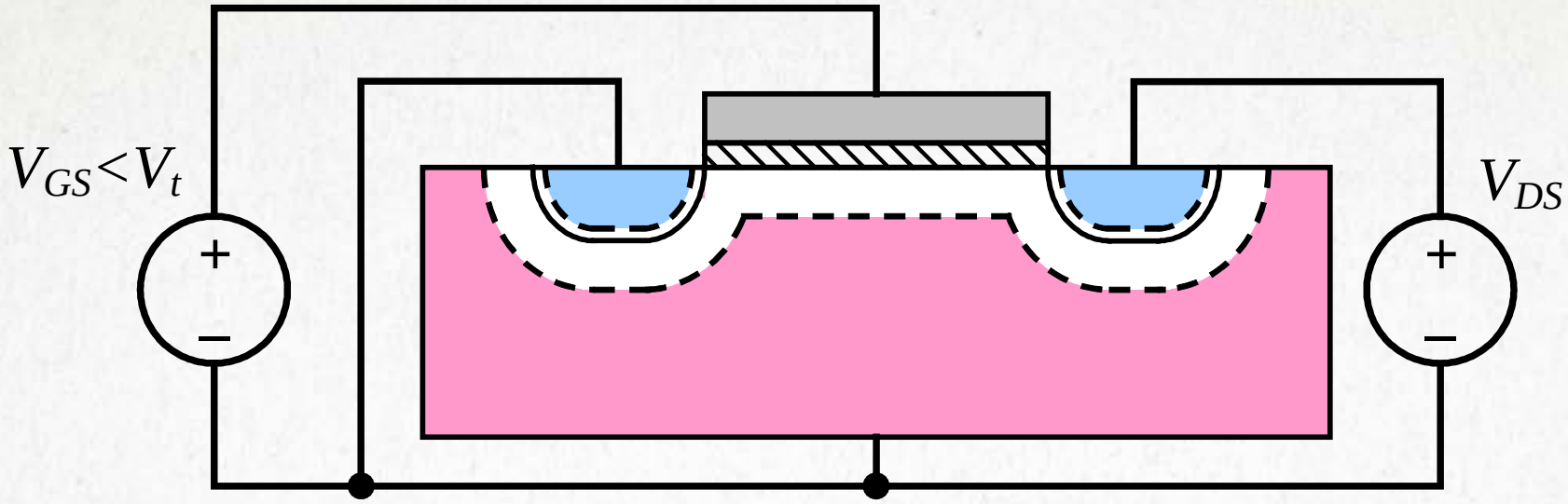
- **The self-aligned poly-gate CMOS transistor has become the mainstay of most analog CMOS and BiCMOS processes.**
  - Virtually all processes offer complementary NMOS and PMOS transistors.
  - The NMOS usually has a higher transconductance than the PMOS due to electrons having a higher mobility than holes.
  - These devices are extremely versatile. They can perform most analog functions as well as implement digital logic.

THE MOS TRANSISTOR: A FOUR-PAGE REVIEW



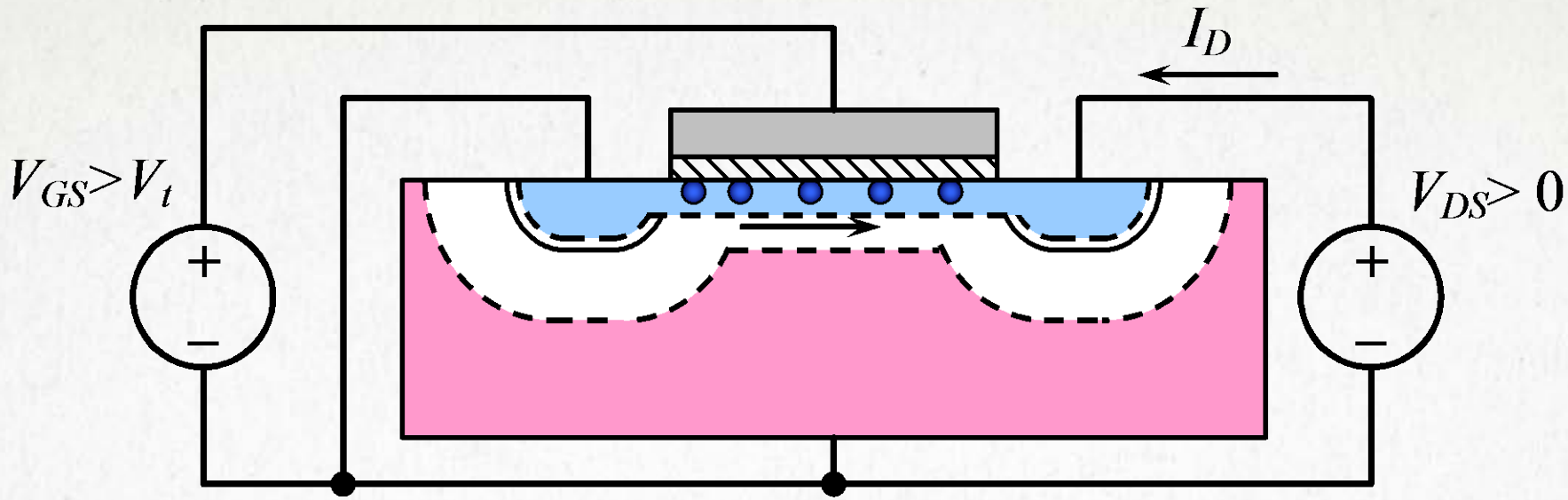
- ***The metal-oxide-semiconductor* (MOS) transistor is a four-terminal active device.**
  - The gate controls current flow from the source to the drain, both of which are embedded in the backgate.
  - An NMOS has N-type source/drain regions and uses electron conduction.
  - A PMOS has P-type source/drain regions and uses hole conduction.

THE NMOS TRANSISTOR: *CUTOFF*



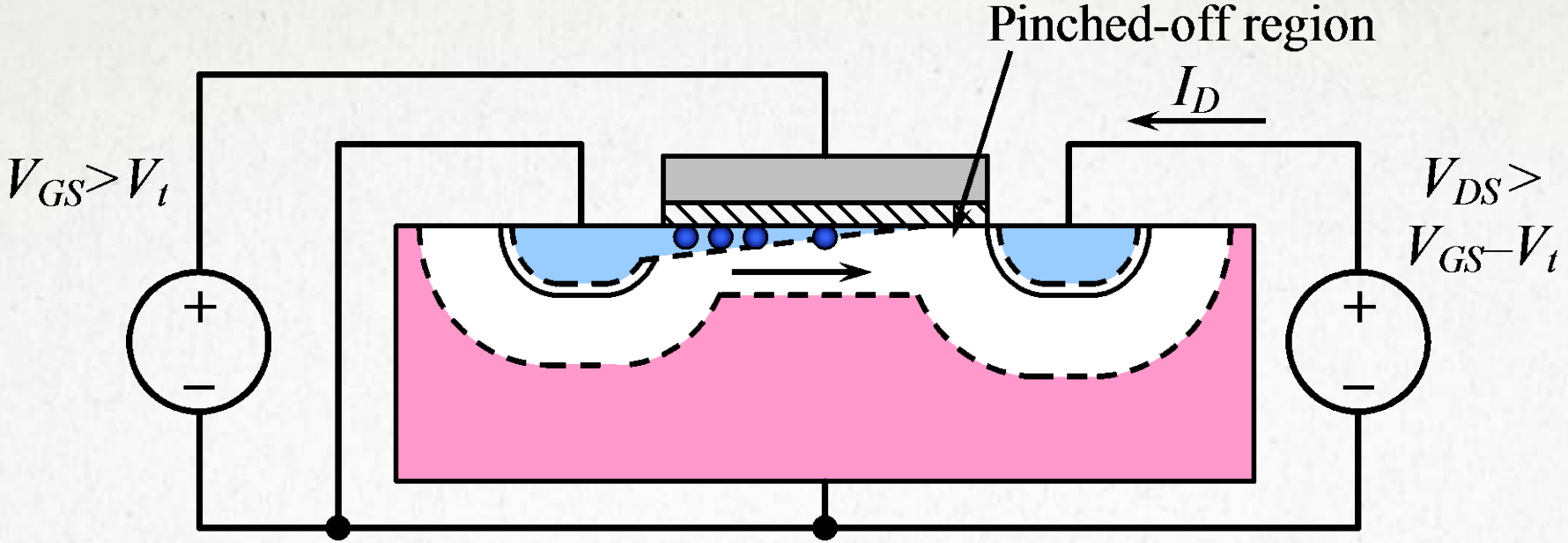
- **If the gate-to-source voltage V_{GS} of an NMOS is less than the *threshold voltage* V_t, no channel forms.**
  - Depending upon gate biasing, an accumulation layer or depletion region may form beneath the gate.
  - However, only leakage currents flow.

THE MOS TRANSISTOR: *TRIODE*



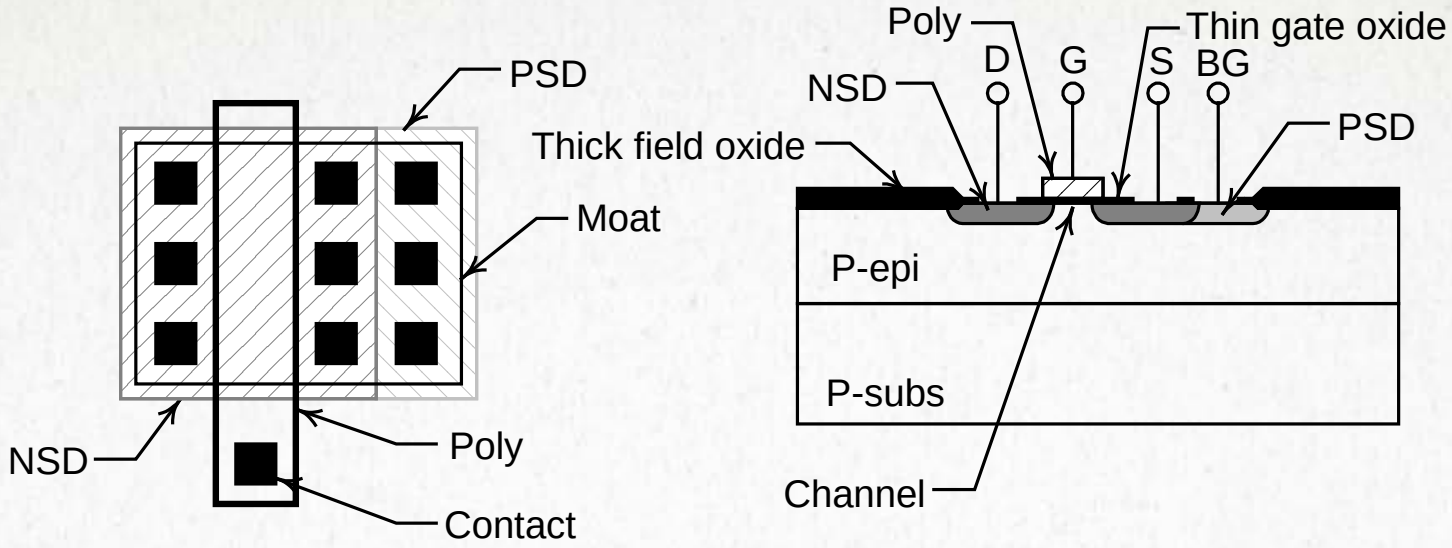
- **If the gate-to-source voltage V_{GS} exceeds the threshold voltage, electrons are attracted by the electric field projected across the gate dielectric.**
  - These electrons form a thin film of negative charge called a *channel*.
  - Electrons flow from source to drain.
  - The drain current increases with drain-to-source voltage.

THE MOS TRANSISTOR: *SATURATION*



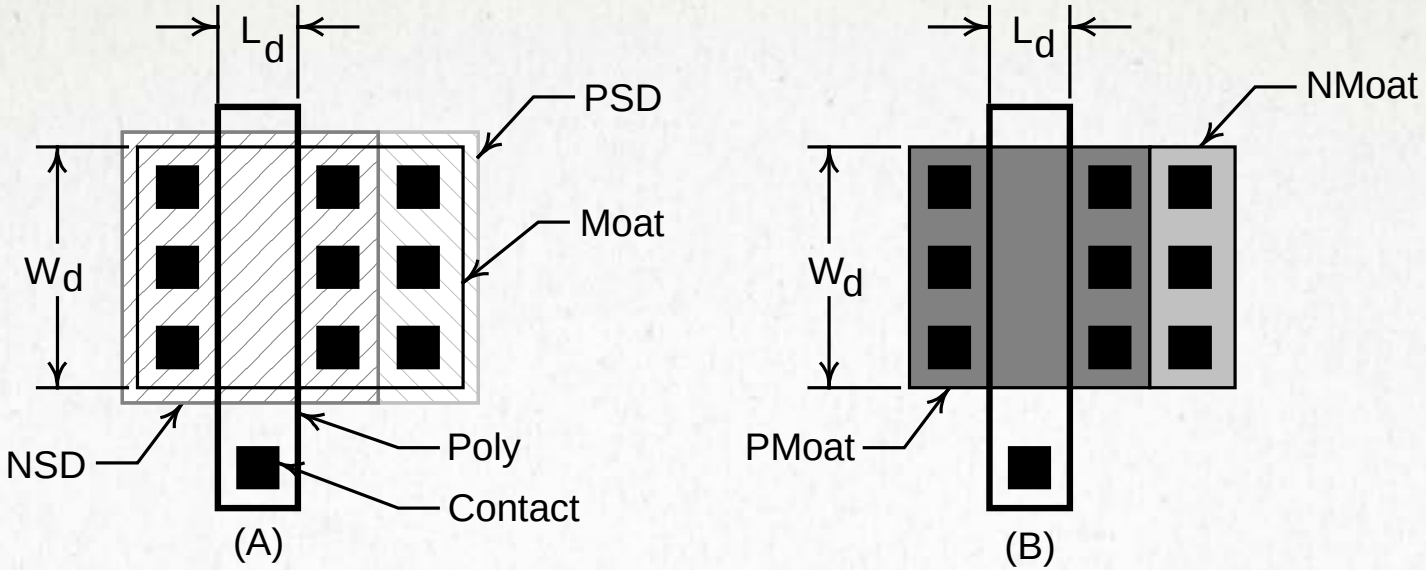
- **At higher drain-to-source voltages, the channel pinches off.**
  - The depletion region widens towards the drain end of the channel, meaning less charge needs to reside in the channel itself.
  - At higher drain-to-source voltages, the channel vanishes (*pinches off*) near the drain.
  - Further voltage increases appear across the pinched-off region.
  - The drain current therefore ceases to increase with drain-source voltage.

THE SELF-ALIGNED POLY-GATE NMOS



- **Most modern MOS processes use self-aligned polysilicon gates.**
  - The poly serves as a masking layer to block the source/drain implants from the channel region.
  - Sidewall spacers formed along the edges of the gate hold the implants slightly away from the actual gate.
  - The resulting structure has minimal overlap capacitances.

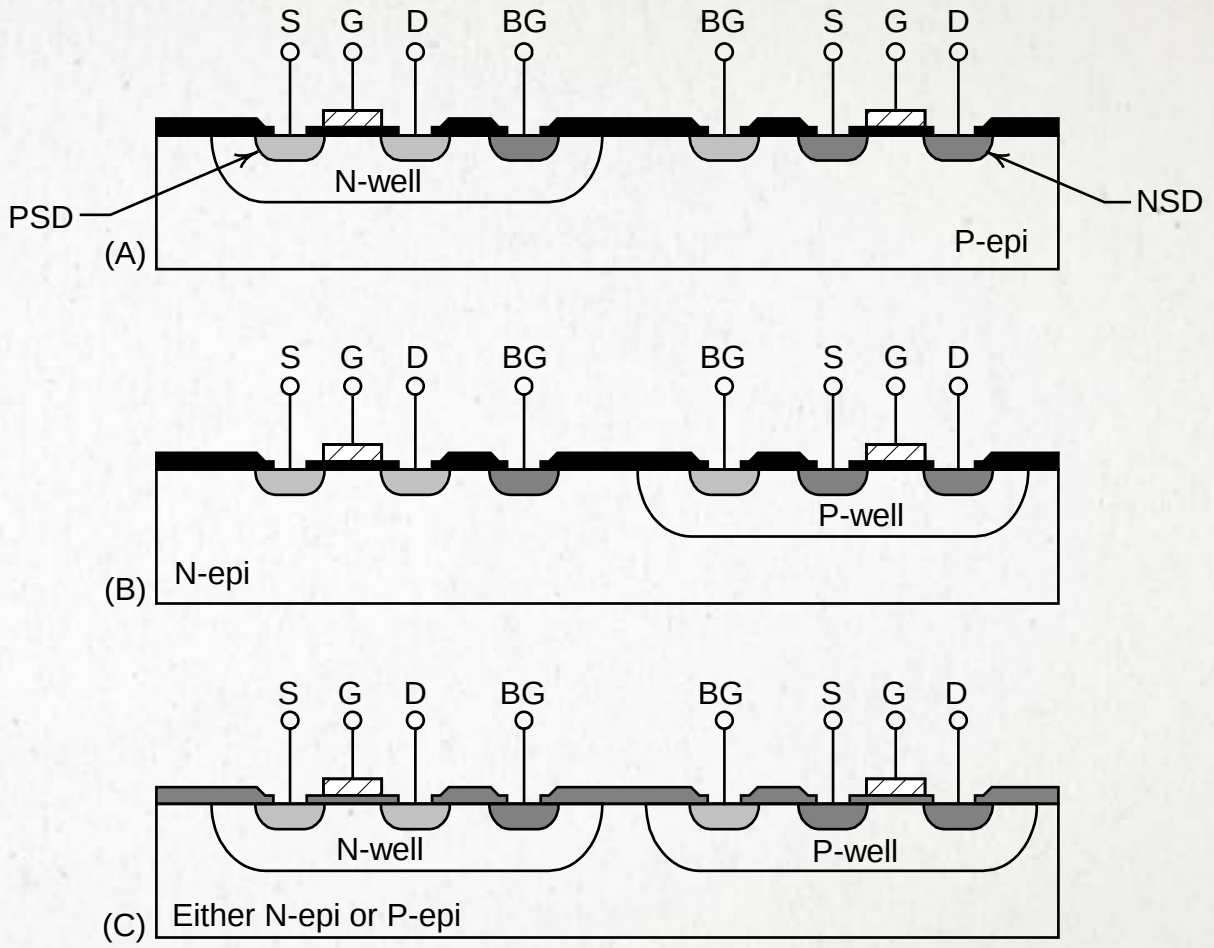
DRAWING THE MOS TRANSISTOR



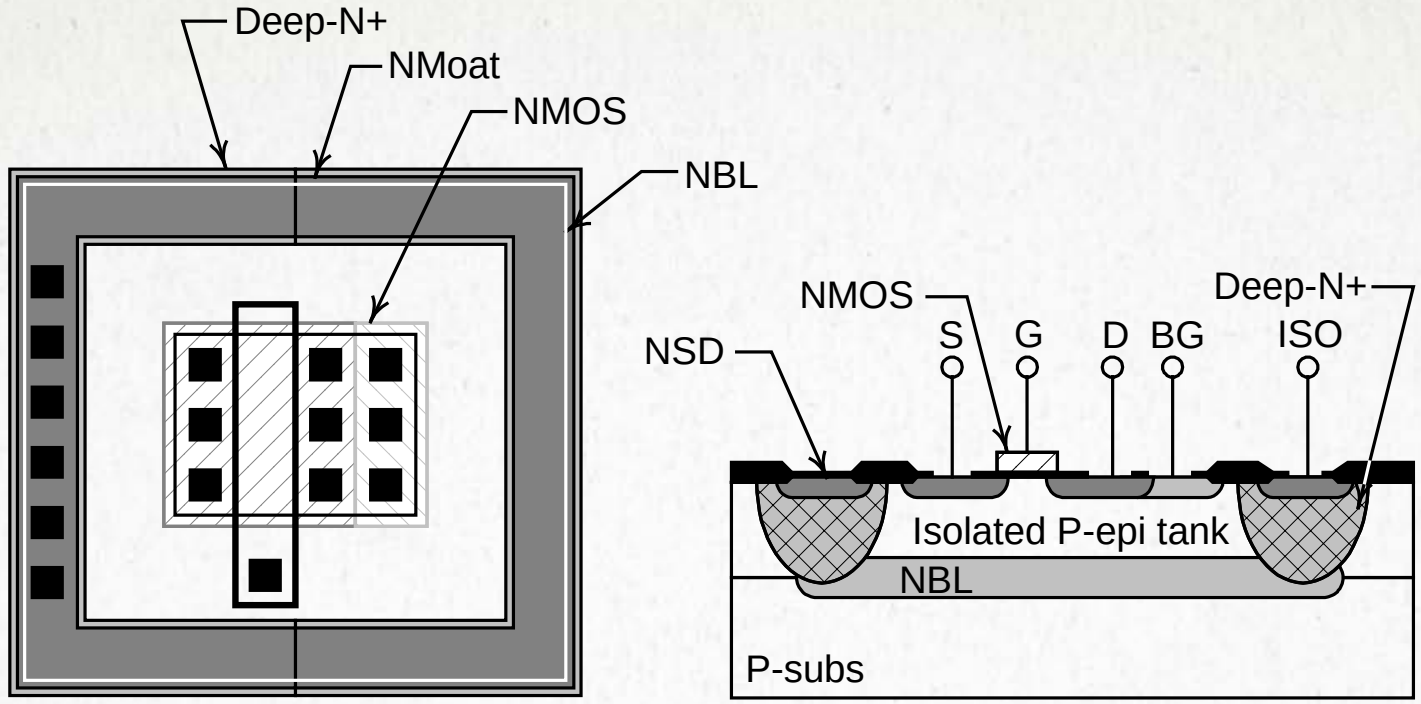
- **An NMOS can be drawn with *mask layers*.**
  - NSD and PSD are the N-type and P-type source/drain implants, respectively.
  - NSD and PSD should overlap the moat slightly to ensure that their edges are defined by the field oxide (this is another example of self-alignment).
- **Alternatively, an NMOS can be drawn with *coding layers*.**
  - NMoat generates NSD and moat; PMoat generates PSD and moat.

TYPES OF PROCESSES

- **An N-well process places the PMOS in an N-well and the NMOS in the P-epi.**
  - This was a popular choice in the 1990s.
  - The use of a P-type epi (and by extension, a P-type substrate) allowed a negative ground.
- **A P-well process places the NMOS in a P-well and the PMOS in the N-epi.**
  - Requires a positive ground (unpopular).
- **A twin-well process places the NMOS in a P-well and the PMOS in an N-well.**
  - The most common choice for low-voltage processes.



THE ISOLATED NMOS

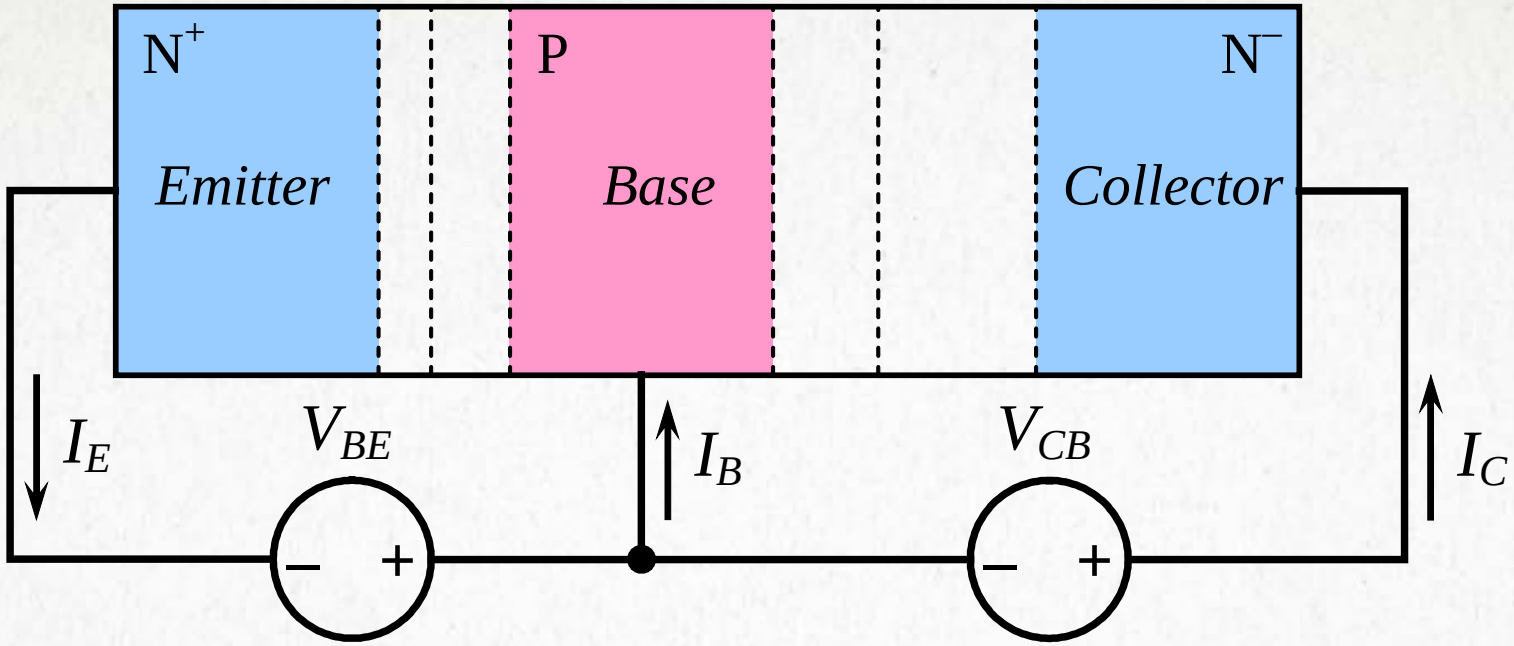


- **Normally, either NMOS or PMOS (usually NMOS) share a common backgate.**
  - BiCMOS processes that include a deep-N+ sinker and an N-type buried layer (NBL) can place an NMOS in an isolated *tank*.
  - A *tank* is a region cut off by means of diffusions; a *well* is just a deep diffusion.

SECTION II: BIPOLAR TRANSISTORS

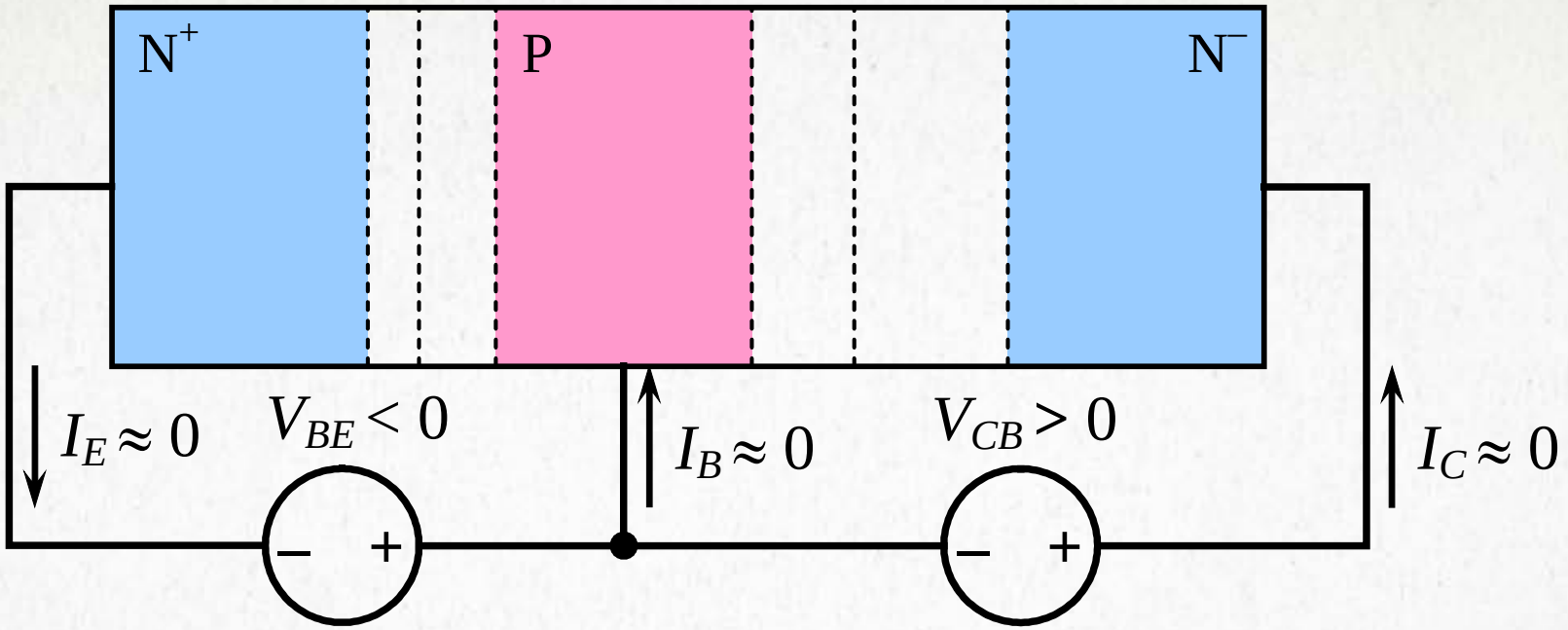
- **Almost all processes can create bipolar transistors:**
  - Standard bipolar only fabricates bipolars.
  - Analog CMOS fabricates substrate PNP transistors.
  - Most BiCMOS processes can fabricate NPN and PNP transistors as well as MOS transistors.
- **Bipolar transistors still have their uses:**
  - They offer better voltage matching than MOS transistors.
  - For a given current, they generate more small-signal transconductance.
  - Ratioed collector currents generate voltages dependent upon absolute temperature.
  - Bipolars make superior ESD devices because they dissipate power in relatively large volumes of silicon.

THE BIPOLAR TRANSISTOR: A FOUR-PAGE REVIEW



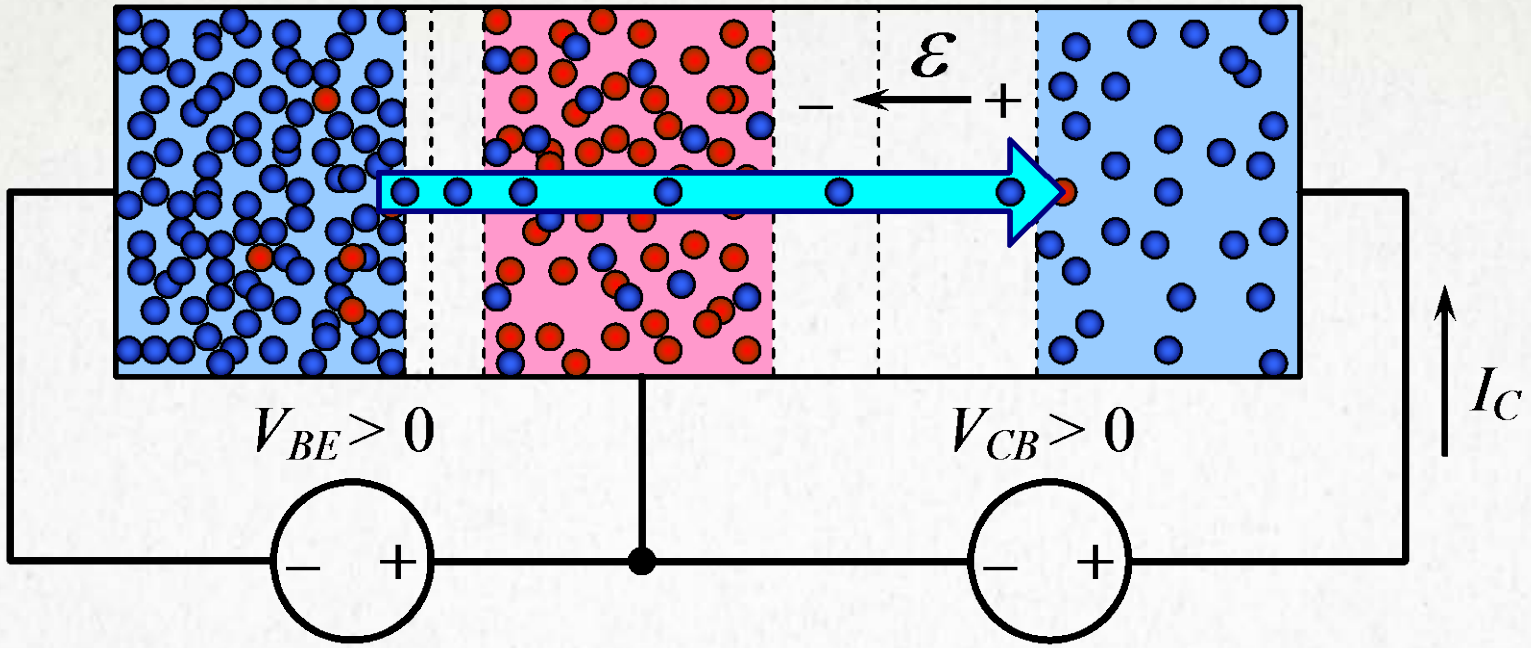
- **A *bipolar transistor* consists of a thin layer of silicon of one polarity called the *base* sandwiched between two layers of opposite polarity called the *emitter* and the *collector*.**
  - An NPN has a P-type base sandwiched between N-type emitter and collector.
  - A PNP has an N-type base sandwiched between P-type emitter and collector.

THE BIPOLAR TRANSISTOR: *CUTOFF*



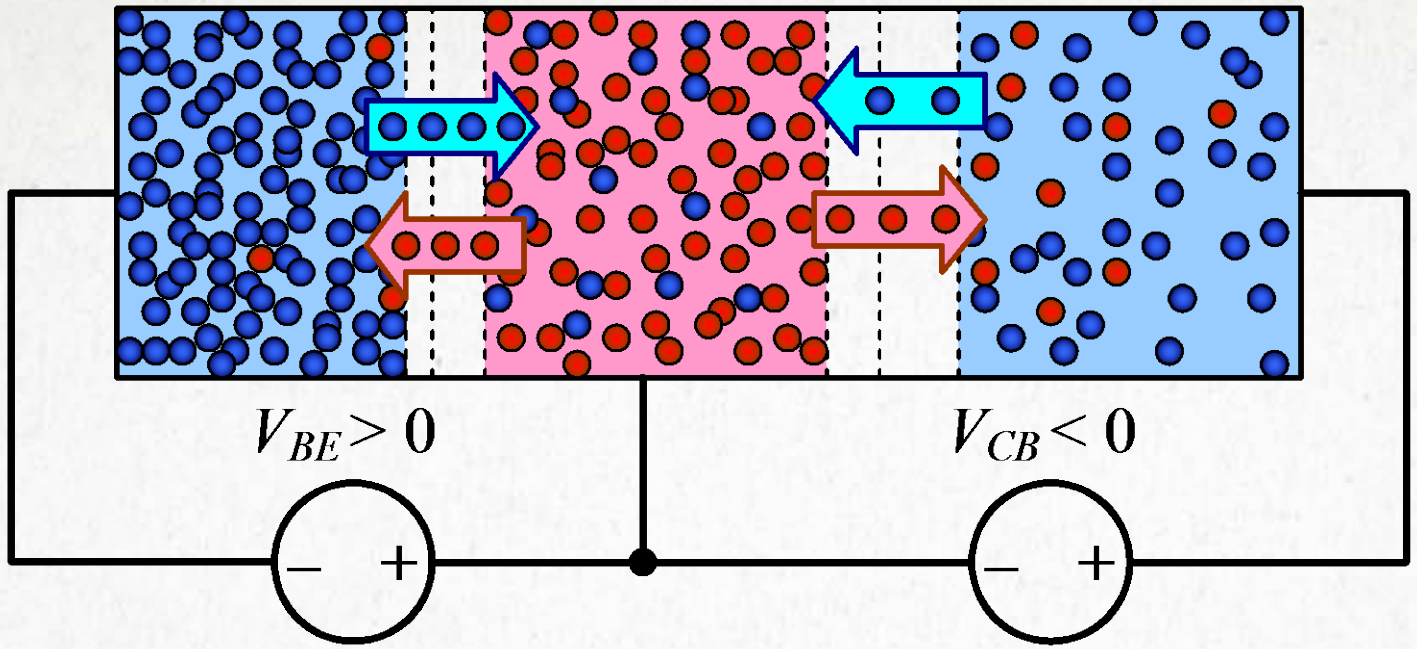
- **In cutoff, both the base-emitter and base-collector junctions are reverse-biased.**
  - Only leakage currents flow.

THE BIPOLAR TRANSISTOR: *FORWARD ACTIVE*



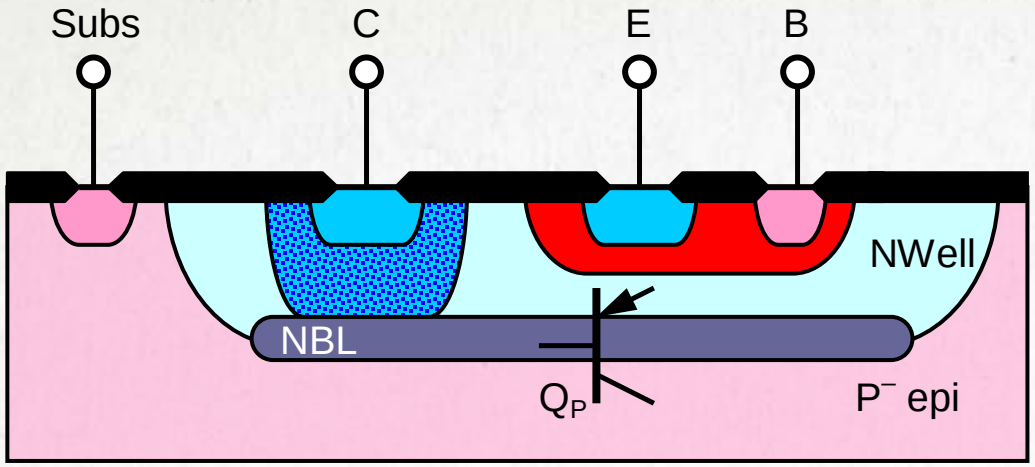
- **In the forward active region, the base-emitter junction becomes forward-biased while the base-collector junction remains reverse-biased.**
  - The emitter injects minority carriers into the base.
  - Those carriers that do not recombine diffuse to the collector.
  - The base-emitter bias therefore controls the collector current.

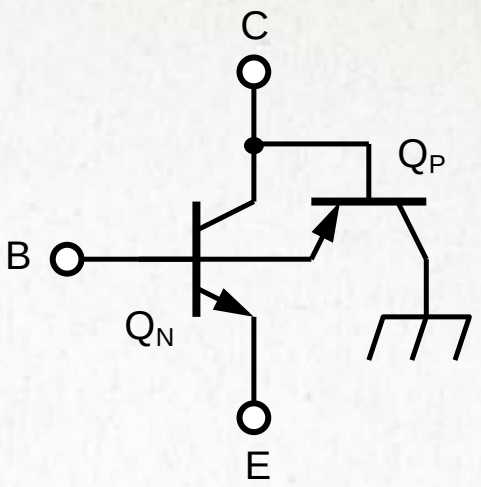
THE BIPOLAR TRANSISTOR: *SATURATION*



- **In saturation both the base-emitter and the base-collector junctions become forward-biased.**
  - Minority carriers injected into the collector take time to recombine, slowing turn-off.
  - In integrated transistors, minority carriers may traverse the collector to the substrate. This represents a major challenge for bipolar circuit designers.

THE PARASITIC PNP INSIDE THE VERTICAL NPN

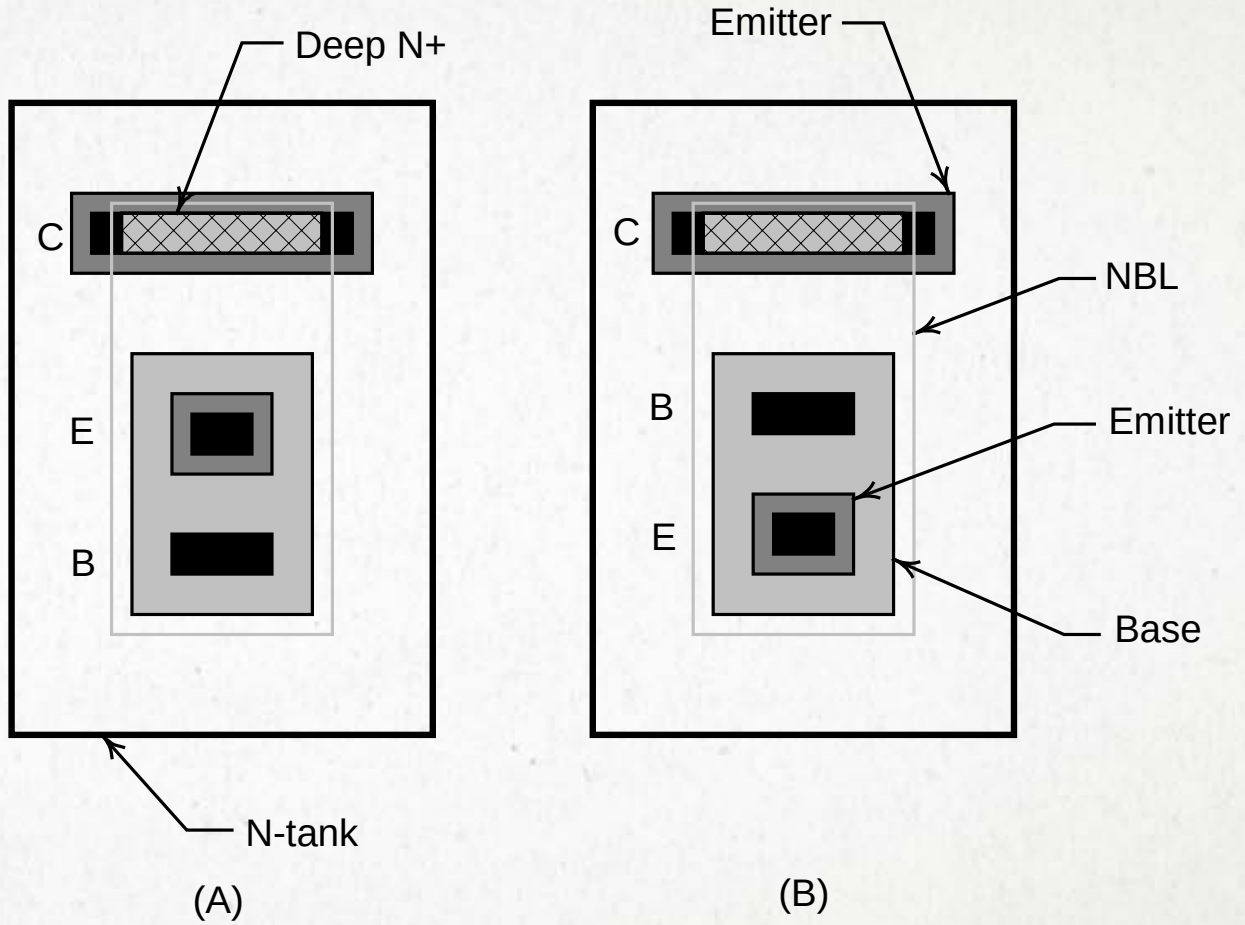




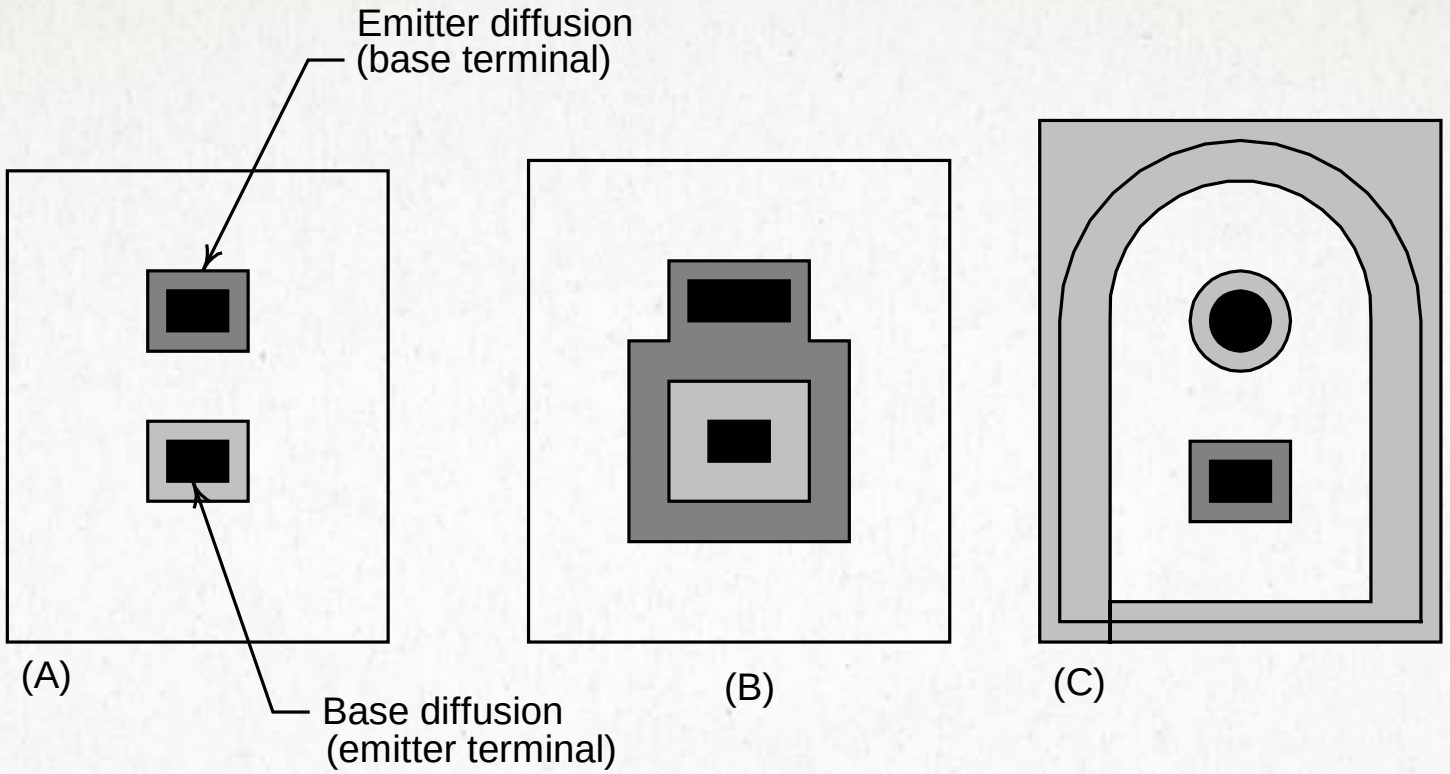
- **Saturating a vertical NPN activates the parasitic substrate PNP.**
  - This parasitic transistor diverts base drive to substrate.
  - If circuit designers don't consider this effect, it can cause circuits to malfunction.
  - To prevent it, avoid saturating the transistor, or add hole-blocking guard rings.

THE STANDARD BIPOLAR VERTICAL NPN

- **The vertical NPN is the best transistor in standard bipolar.**
  - Vertical conduction typically gives a beta of 100-300.
  - Deep-N+ sinker can be added to minimize collector resistance.
  - Many alternative layouts exist; for example, one can swap the locations of base and emitter.

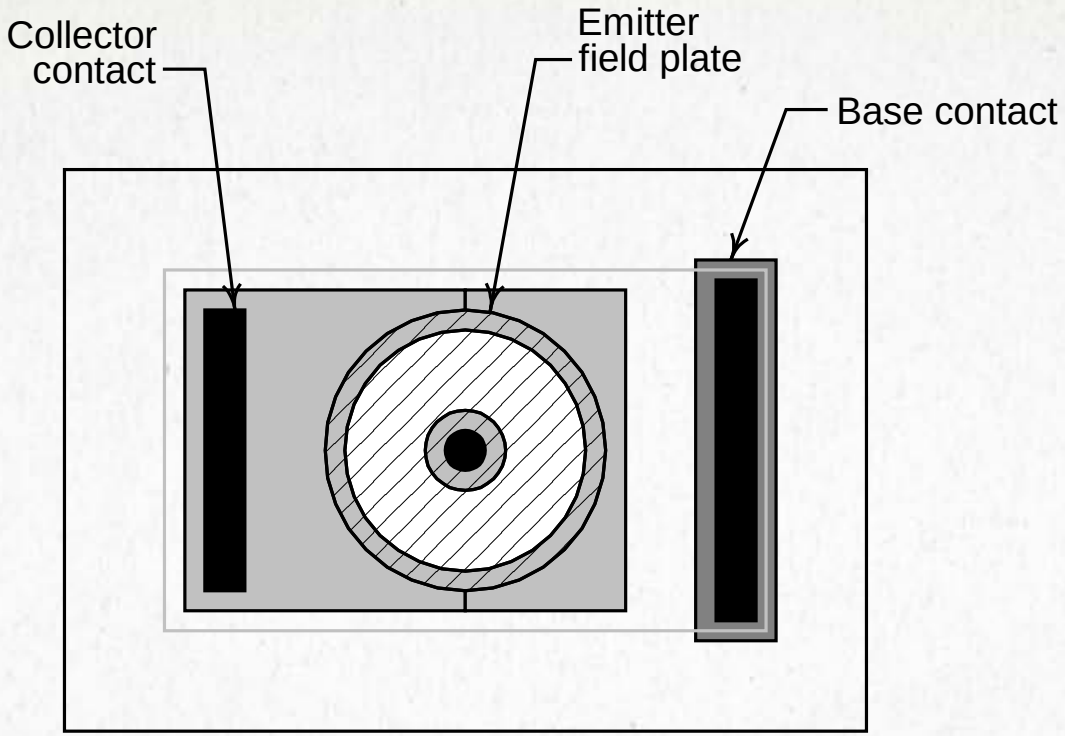


THE STANDARD BIPOLAR SUBSTRATE PNP



- **A substrate PNP uses the substrate as its collector.**
  - The simplest style uses plugs of base and emitter inside a tank.
  - Many other styles exist that purport to have better performance (often, they don't).

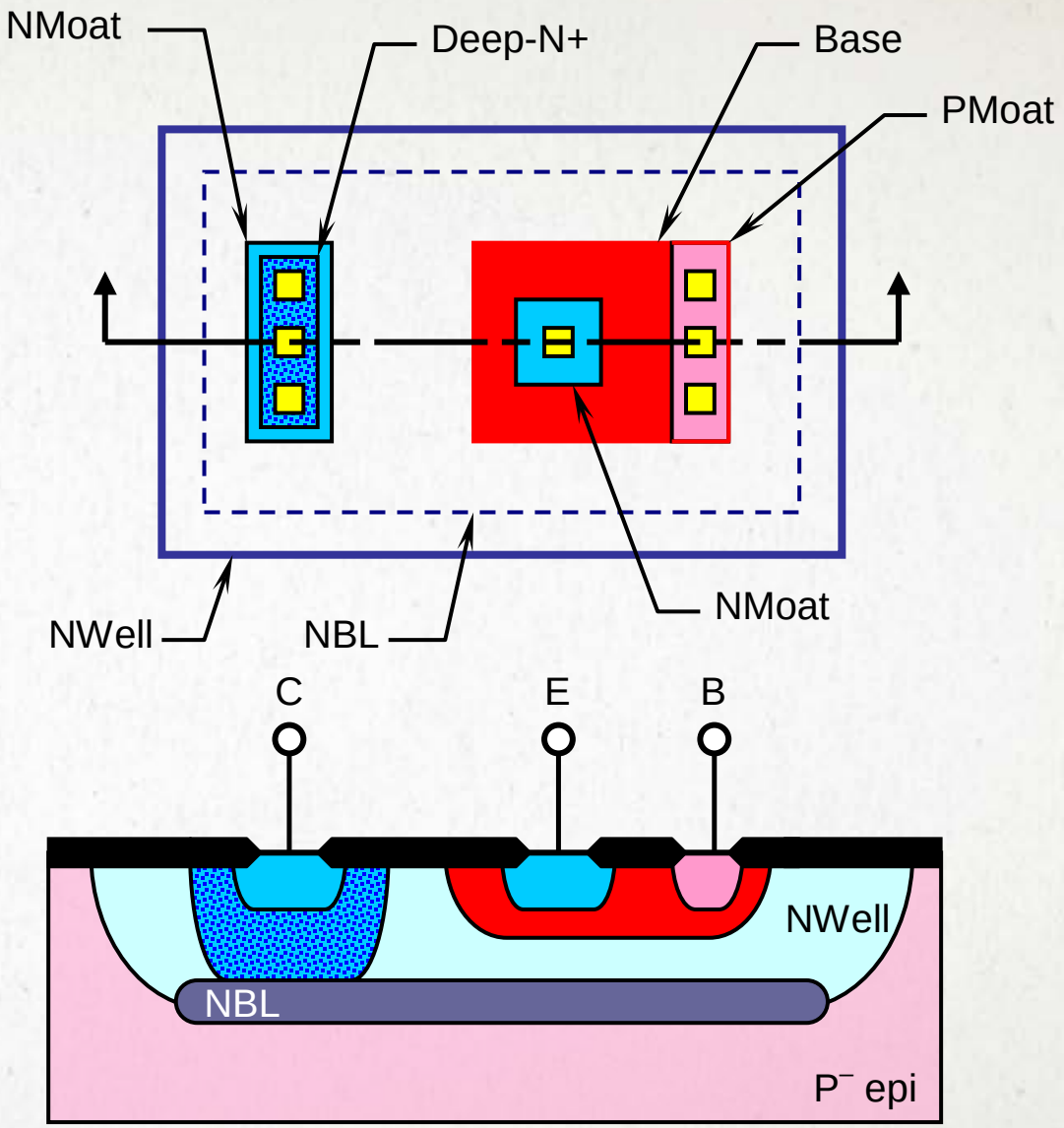
THE STANDARD BIPOLAR LATERAL PNP



- **A lateral PNP trades off performance for an isolated collector.**
  - Lateral PNP transistors have large parasitic capacitances and are thus slow.
  - Always cover the exposed base with metal to prevent surface channel formation.

THE BICMOS CDI NPN

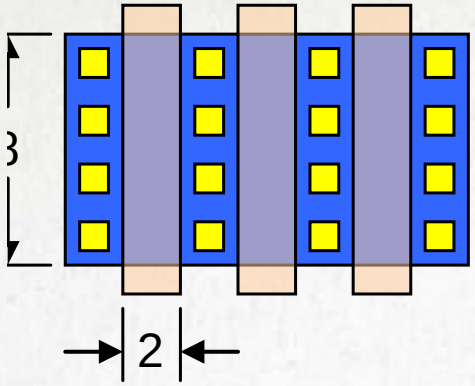
- **Many analog BiCMOS processes offer a relatively simple *collector-diffused-isolation* (CDI) NPN.**
  - This device uses either a dedicated base or a shallow P-well as its base.
  - A deep N-well forms the collector, which also isolates the device from substrate.
  - An N-buried layer must be added to the process; optionally deep-N+ sinker can also be added.



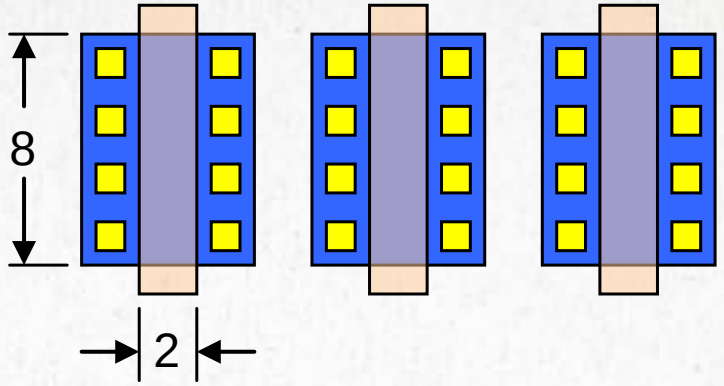
SECTION III: TRANSISTOR MATCHING

- **Transistors obey the same principles of matching as resistors and capacitors.**
  - Both MOS and bipolar transistors are very sensitive to thermal gradients.
  - Bipolar transistors are very sensitive to stress gradients.
  - Use common-centroid layouts and consider device placement carefully.
  - MOS transistors usually benefit from careful use of dummy devices.

MULTIPLE FINGERS, MULTIPLE DEVICES



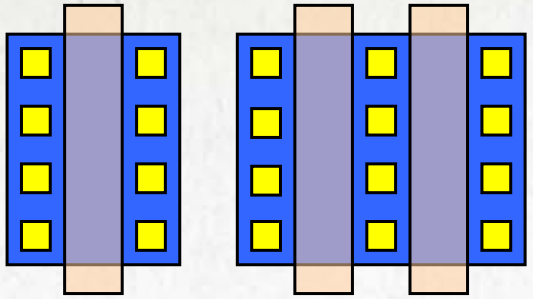
3(2/8)



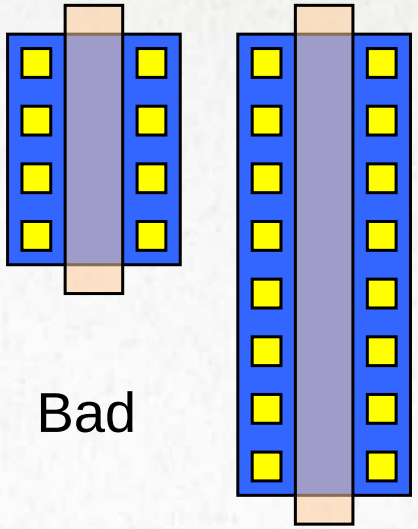
3*2/8

- *N(W/L)* specifies a single device containing *N* fingers, each having width *W* and length *L*.
- *N*W/L* specifies *N* separate devices, each having width *W* and length *L*.
  - Matched devices are most easily constructed as separate devices.
  - If separate devices share sources or drains, then they can be merged together just as if they were multiple fingers of one device.

MATCHING RULES: IDENTICAL SECTIONS



Good



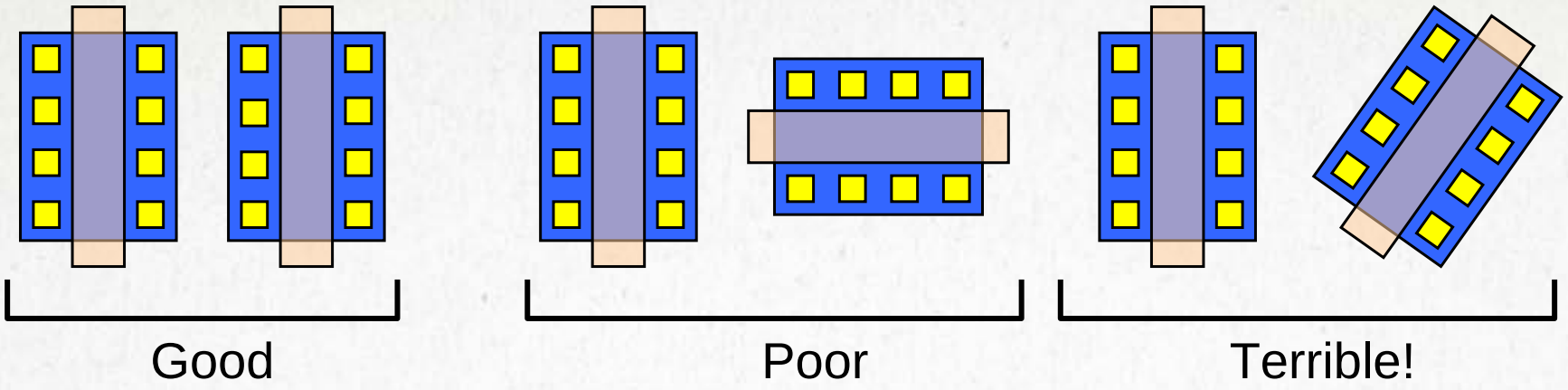
Bad

- **Devices having different widths or lengths don't match well.**
  - The effective silicon dimensions W_e and L_e equal the drawn dimensions W_d and L_d offset by width and length reduction factors W_r and L_r,

$$W_e = W_d + W_r \qquad L_e = L_d + L_r$$

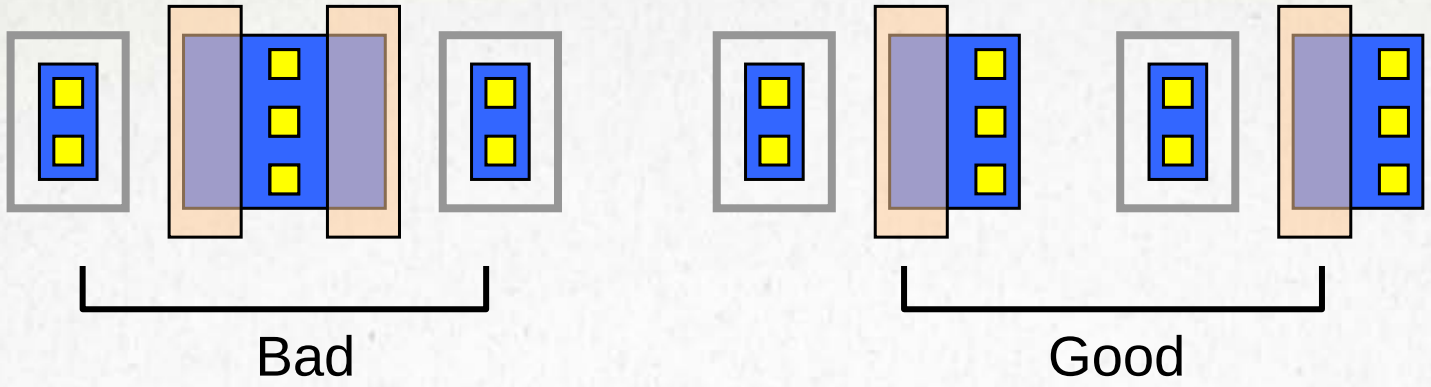
  - The only way to avoid mismatches due to uncertainties in width in length reduction factors is to use identical geometries.

MATCHING RULES: ALIGNMENT



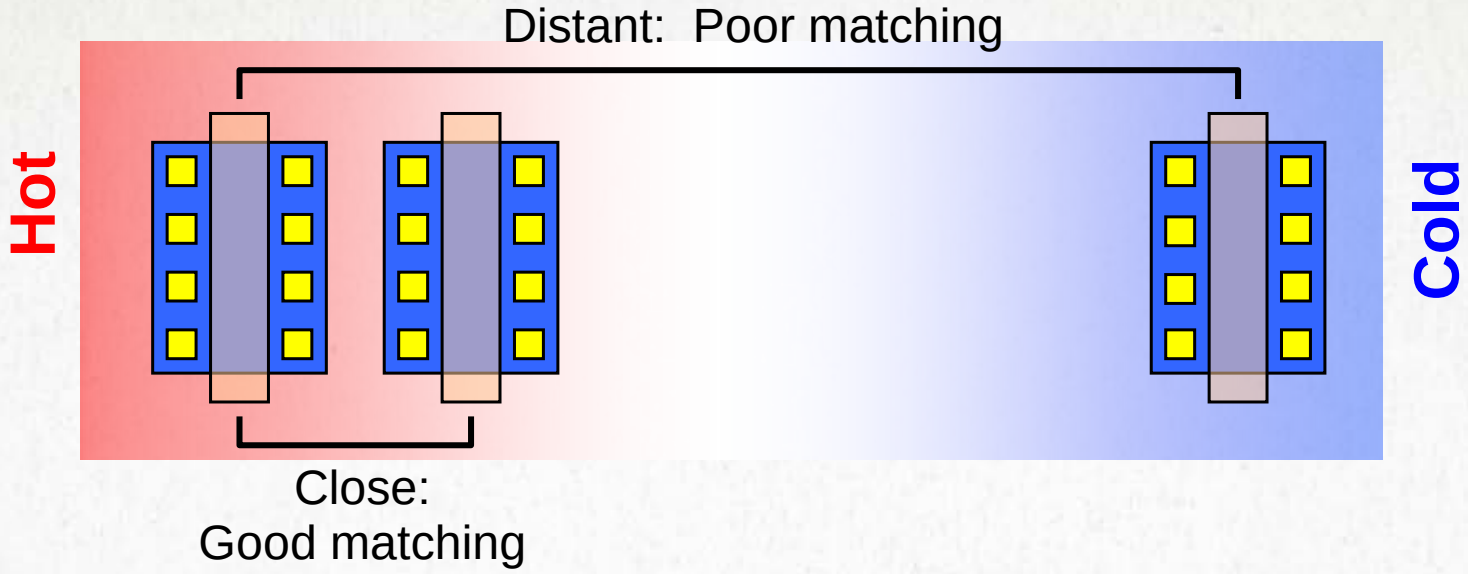
- **Align devices in the same direction.**
  - Silicon becomes anisotropic when subjected to mechanical stress.
  - Wafers are sometimes cut slightly off-axis to minimize pattern shift.
  - Implants may be shot into the wafer at a tilt.
  - Devices aligned in different directions do not match well.

MATCHING RULES: ORIENTATION



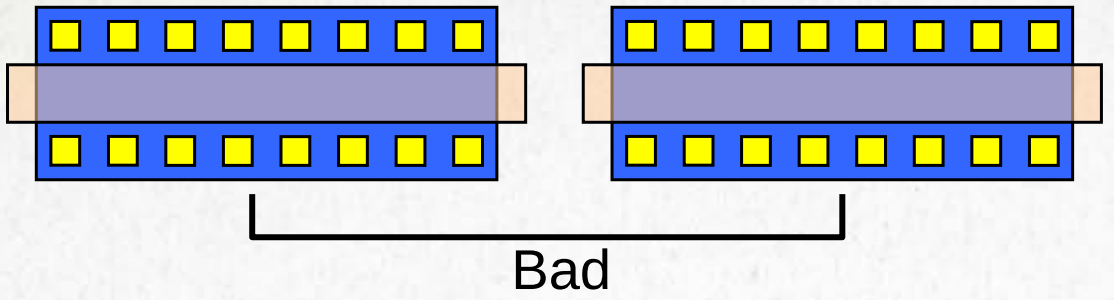
- **Even if aligned along the same axis, devices oriented in opposite directions may not match well.**
  - This is particularly true of drain-extended devices in which one end of the channel is defined by a different mask than the other.
  - Mask misalignment will affect devices oriented in opposite directions differently.
  - The simplest solution is to orient matched devices in the same direction.

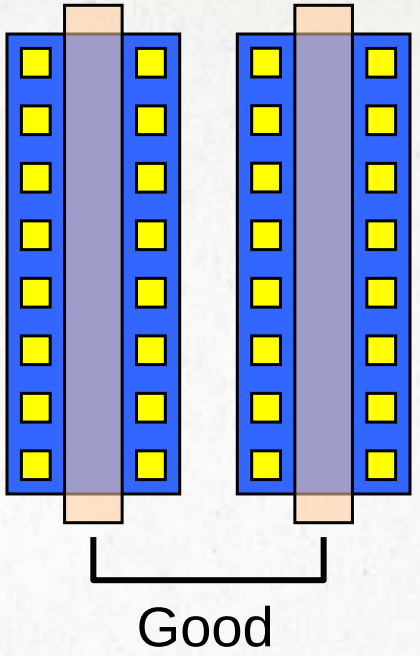
MATCHING RULES: PROXIMITY



- **Proximity minimizes the effects of stress and thermal gradients.**
  - MOS device transconductance is sensitive to mechanical stress.
  - MOS threshold voltage is sensitive to temperature.
  - Bipolar transistors are even more sensitive to gradients because of their high small-signal transconductance.

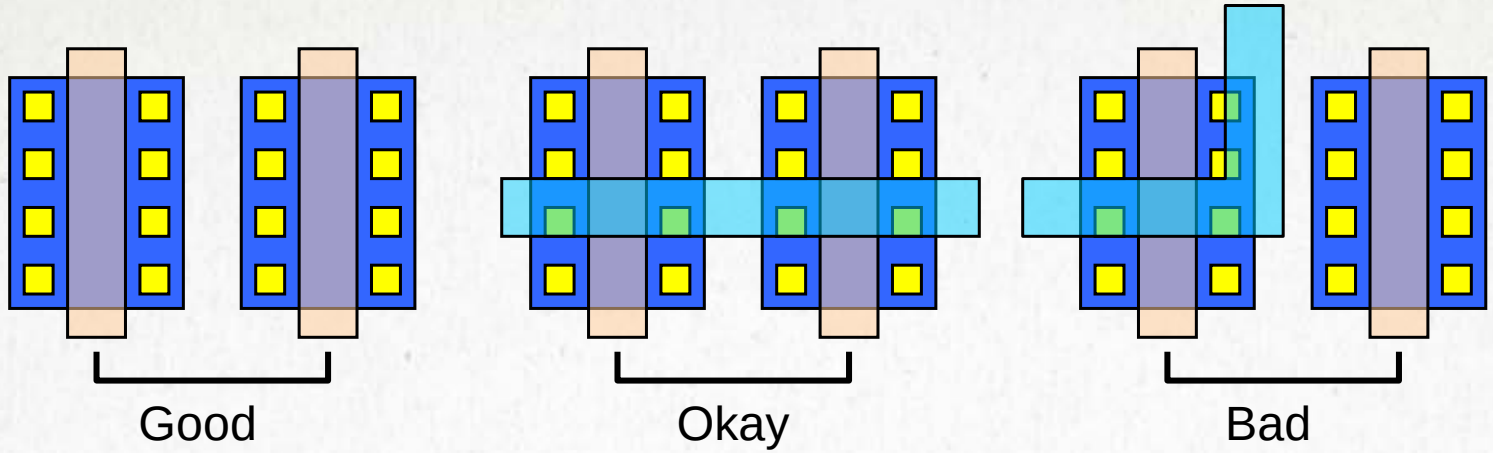
MATCHING RULES: COMPACTNESS





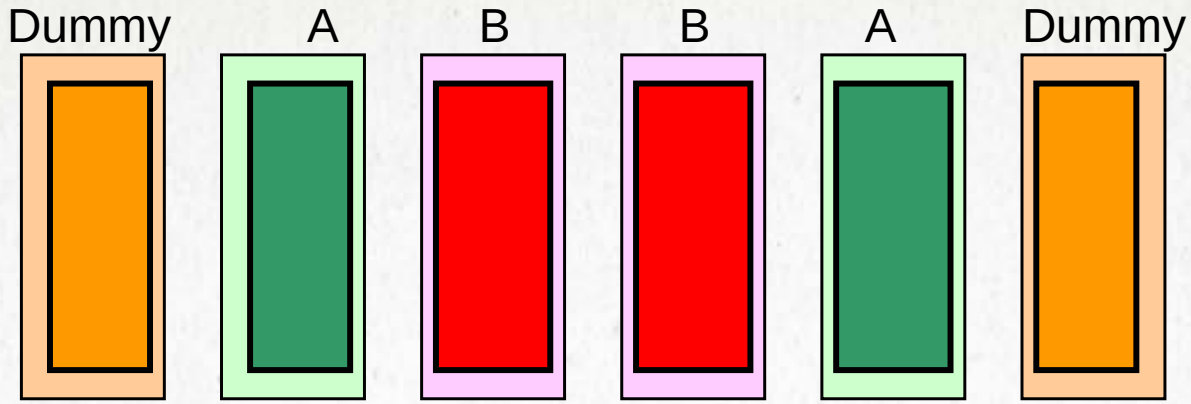
- **Compact layouts experience less gradient-induced mismatches.**
  - Proximity, and even common-centroid layout, can only do so much.
  - The nonlinear components of the gradients will produce mismatches whose magnitude increases roughly as the square of distance.
  - Compact layouts of matched devices will minimize the impact of these nonlinear residual variations.

MATCHING RULES: METALLIZATION



- **Hydrogenation creates mismatches between devices with different amounts of metal coverage.**
  - Hydrogen reacts with dangling bonds at the silicon surface and thus reduces the surface state charge.
  - Metal blocks the diffusion of hydrogen down to the silicon surface.
  - Therefore MOS transistors with different metal coverage exhibit different threshold voltages.

MATCHING RULES: DUMMIES

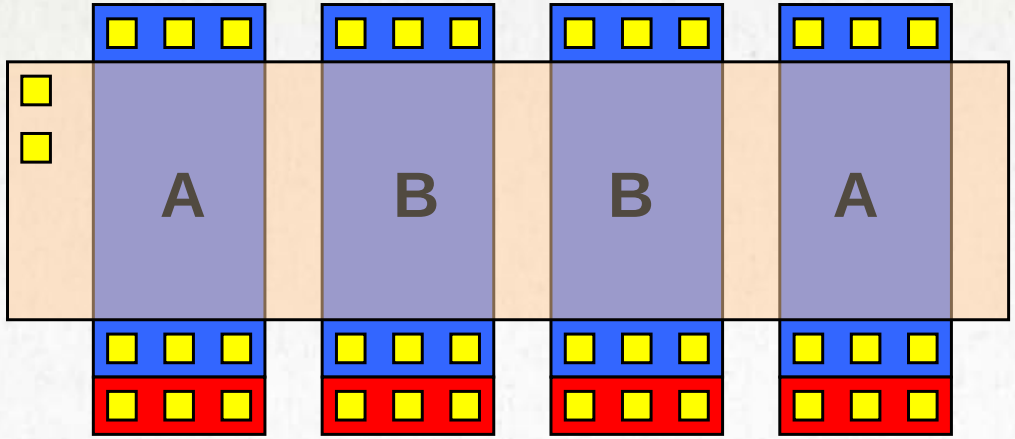


- **Etch rate variations affect devices on the ends of an array differently than those in the center.**
  - Arrays of interdigitated MOS transistors are very vulnerable to poly etch-rate variations.
  - Add dummy devices to either end of the array.
  - Maintain identical poly-to-poly spacings between active devices and from active devices to dummies.
  - Accurately matched short-channel devices may benefit from more than one dummy on either end of the array.

THE FIVE RULES OF MOS COMMON-CENTROID LAYOUT

- ***Coincidence:*** The centroids of the matched MOS transistors should coincide at least approximately. Ideally, they should coincide exactly.
- ***Symmetry:*** The array should be symmetric around both the horizontal and vertical axes.
- ***Dispersion:*** When possible, a large array should be subdivided into as many smaller arrays that are possible that each satisfy the rules of coincidence and symmetry. If this can be achieved, then the larger array need not satisfy these rules; only the subarrays that comprise it need do so.
- ***Compactness:*** The array (or each of its subarray) should be as compact as possible.
- ***Orientation:*** Matched MOS transistors should possess the same orientation.

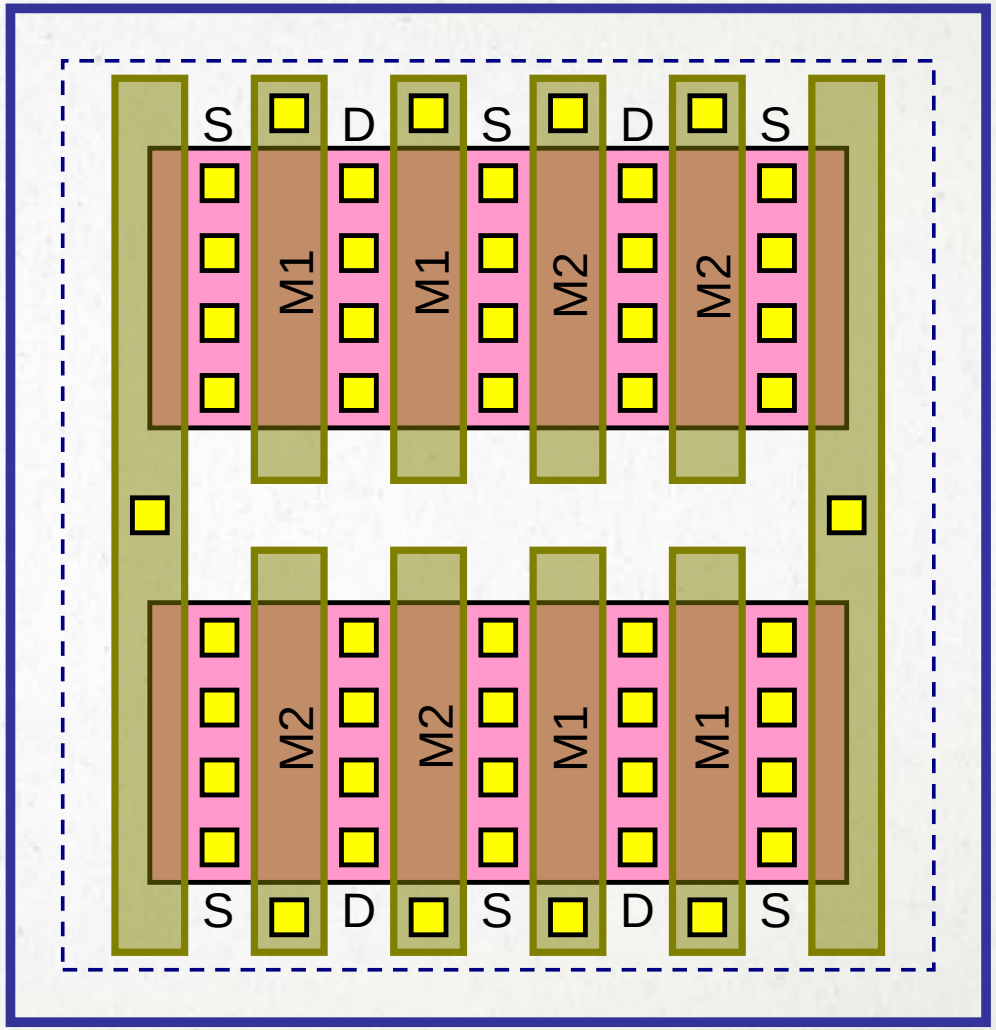
SAMPLE MATCHED LAYOUT: CURRENT MIRROR



- **Current mirrors usually require only moderate matching.**
  - Dummies aren't used because they're hard to construct and they don't have too much impact on long-channel devices.
  - A common gate poly geometry saves space.
  - Merged backgate contacts have been added to enhance latchup immunity.
  - Extend poly beyond moat a bit more than required by rules.

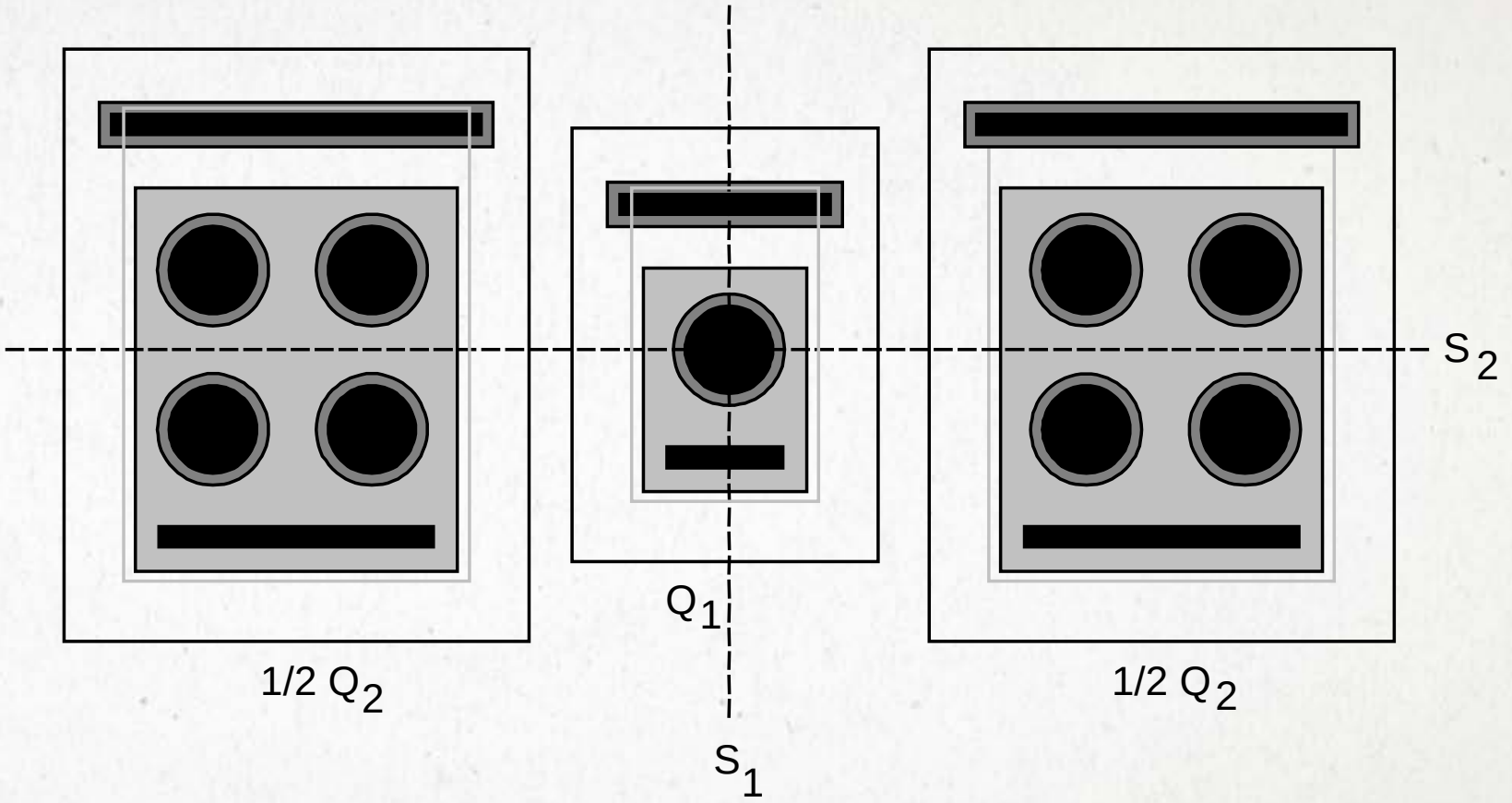
SAMPLE MATCHED LAYOUT: CROSS-COUPLED PAIR

- **Cross-coupled pairs are often used for diff pairs.**
  - This is a small cross-coupled pair, but it shows the concept.
  - Notice so-called *half dummies* have been used. These employ narrow strips of poly and terminate the moat partway across the poly.
  - Half dummies will usually suffice for moderate matching, but wider poly strips may be needed for accurate matching.



COMMON-CENTROID NPN LAYOUT

- **Bandgap references require *N*:1 matched NPNs.**
  - To create a common-centroid layout, use an even value of *N*.
  - The emitter geometry determines the size of a vertical NPN, so multiple emitters can reside in a common base.



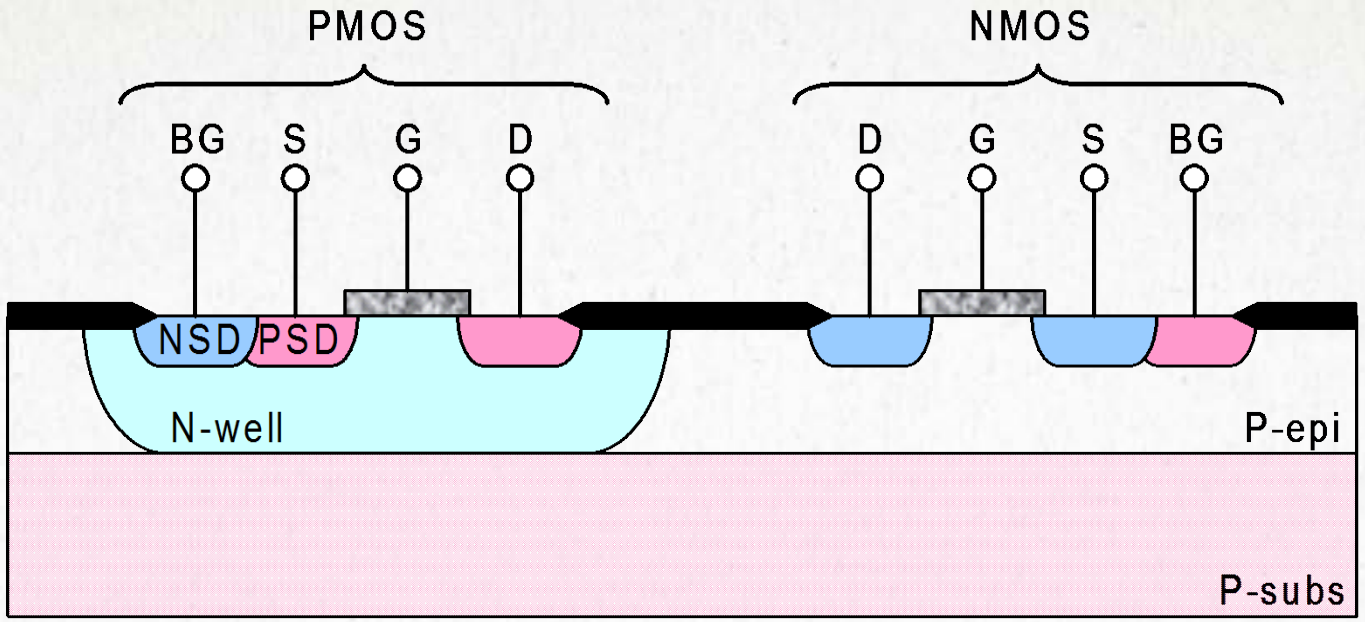
- Slightly increase emitter-emitter spacing and base overlap of emitter to ensure that crowding does not alter emitter areas.

LATCHUP

ALAN HASTINGS

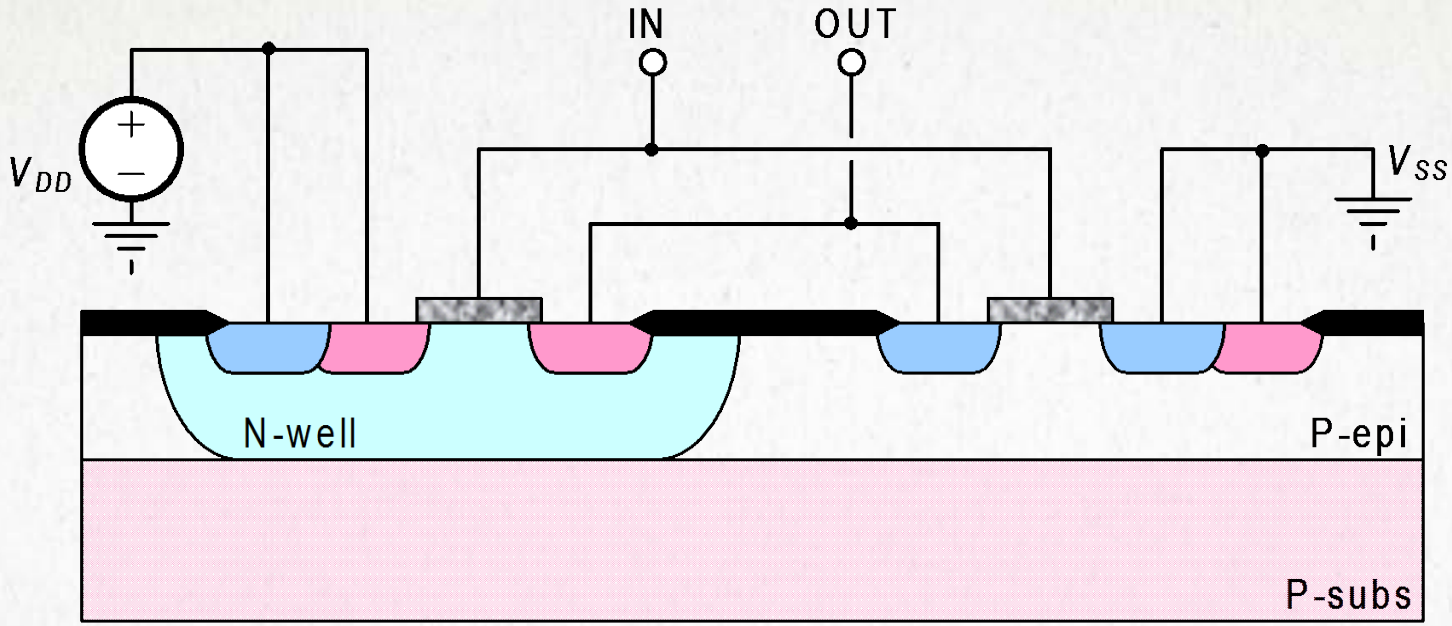
CONTENTS

N-WELL CMOS PROCESS CROSS SECTION



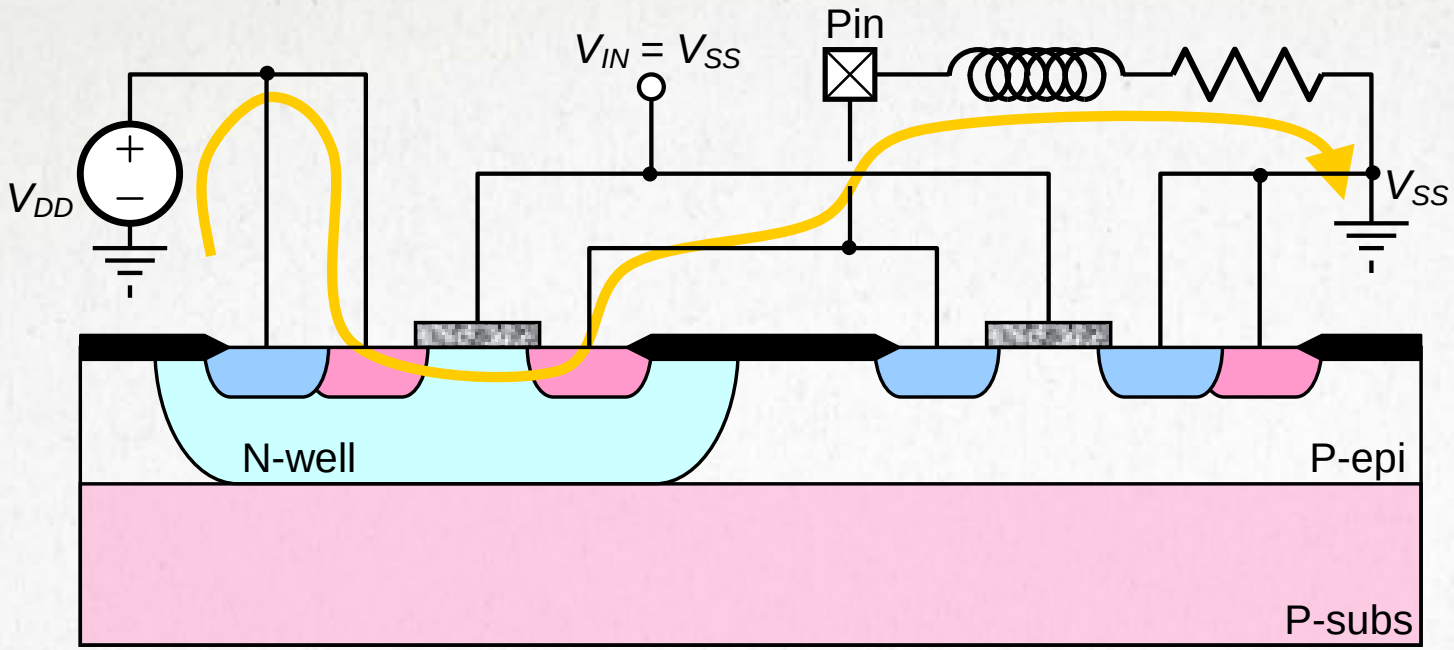
- **An NWell CMOS process fabricates NMOS and PMOS transistors on the same substrate.**
  - The NMOS transistors sit in the P-epi or in a P-well (not shown).
  - The PMOS transistors sit in an N-type well (N-well).

BACKGATE CONTACTS



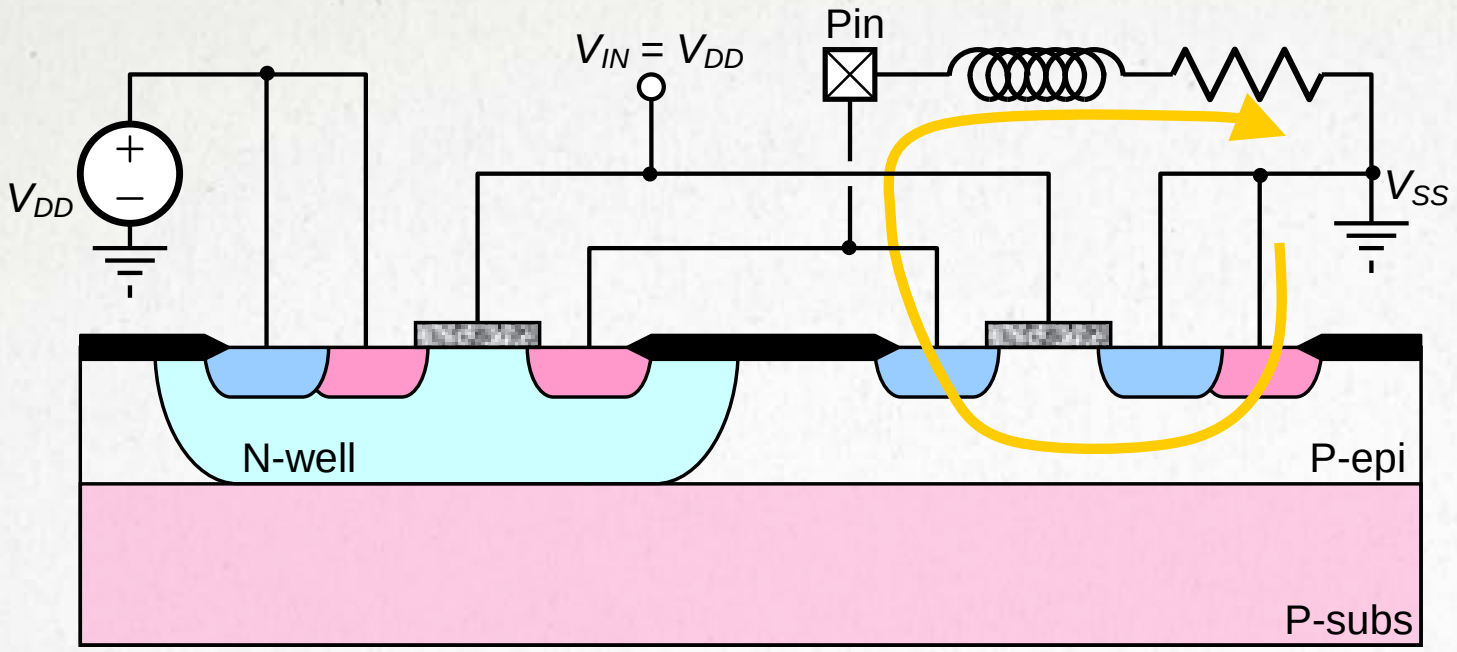
- **Suppose the NMOS and PMOS are connected to form an inverter.**
  - The PMOS backgate is connected to the positive supply, V_{DD}.
  - The NMOS backgate is connected to the ground return, V_{SS}

SWITCHING TRANSIENTS



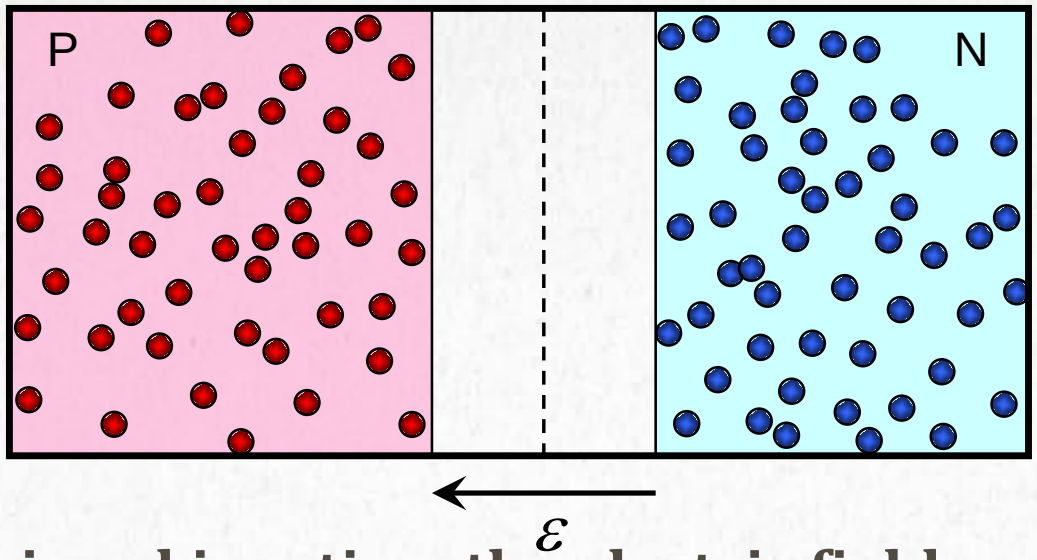
- **Suppose the output of the inverter connects to an external load.**
  - The external load may exhibit significant inductance.
  - When the output goes high, current flows through this inductance…so far, no problem.

SWITCHING TRANSIENTS



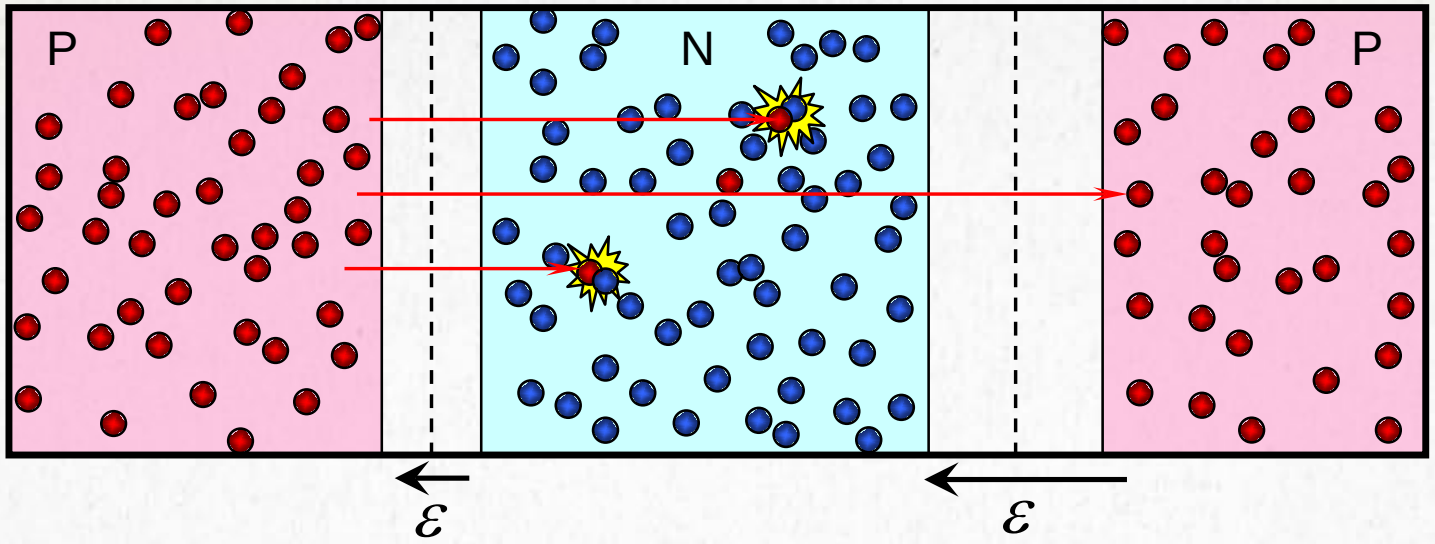
- **Now switch the inverter so that its output goes low.**
  - Current flow should stop, but the inductance momentarily continues to draw current.
  - This current pulls the NMOS drain below V_{SS}, *forward-biasing the drain-backgate junction.*

DIGRESSION: REVERSE-BIASED PN JUNCTIONS



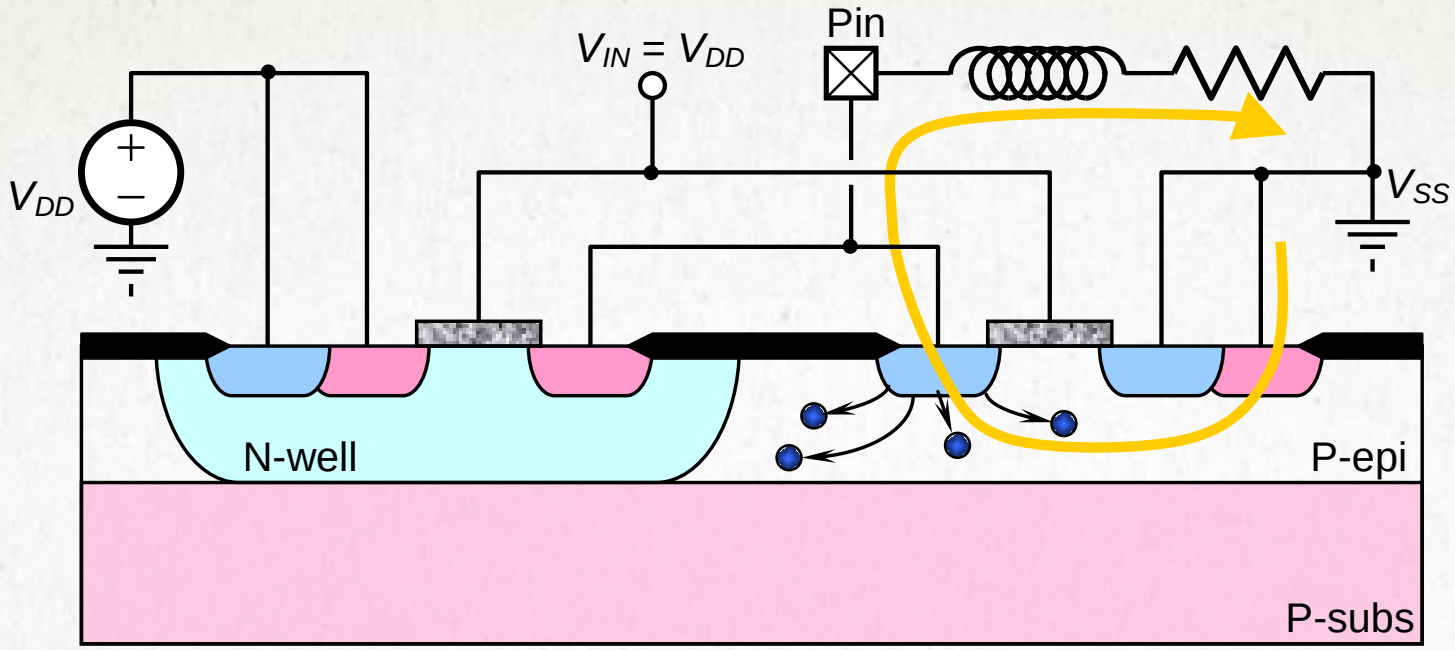
- **In a reverse-biased junction, the electric field across the depletion region holds the majority carriers on their respective sides of the junction.**
  - Holes stay on the P-type side of the junction (the *anode*).
  - Electrons stay on the N-type side of the junction (the *cathode*).
  - Little or no current flows across the junction.

DIGRESSION: BIPOLAR JUNCTION TRANSISTORS



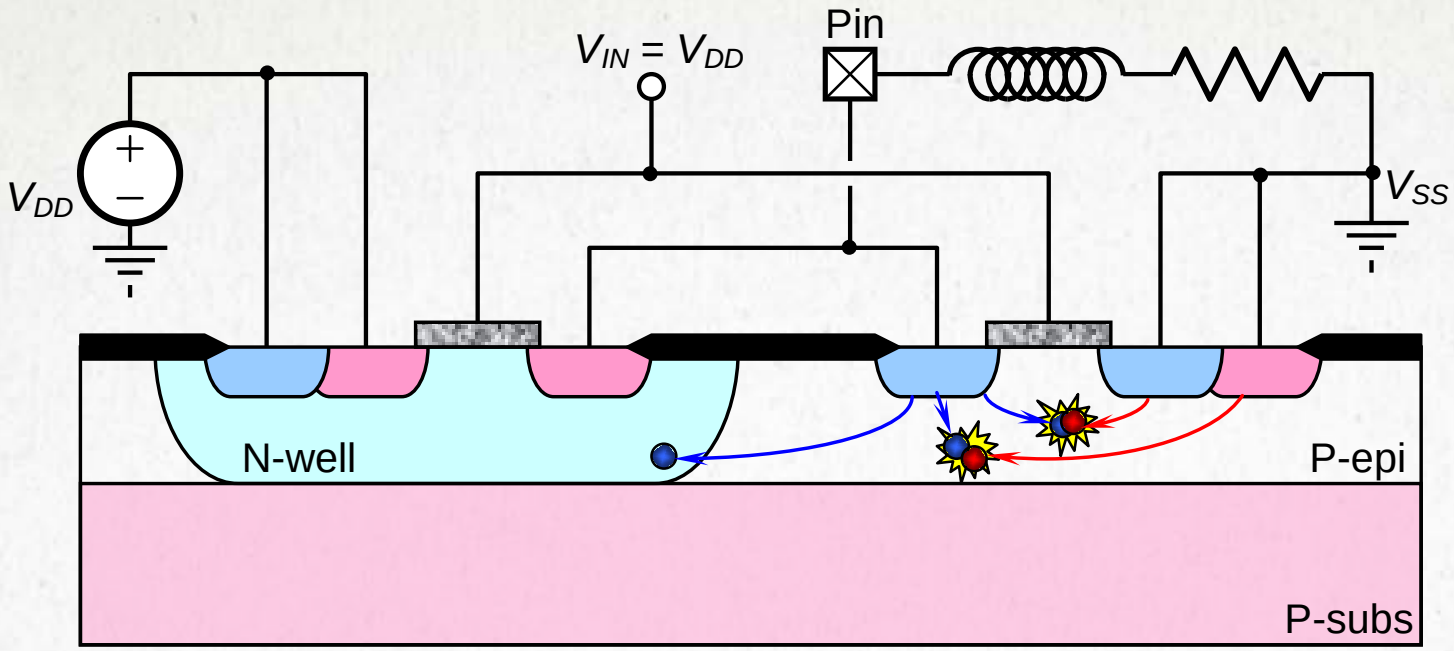
- **Minority carriers will typically recombine within a few microseconds.**
  - If in this time they can diffuse to a reverse-biased junctions, then they can diffuse across it and can again become majority carriers.
  - This is the essence of bipolar transistor action.

MINORITY CARRIER INJECTION



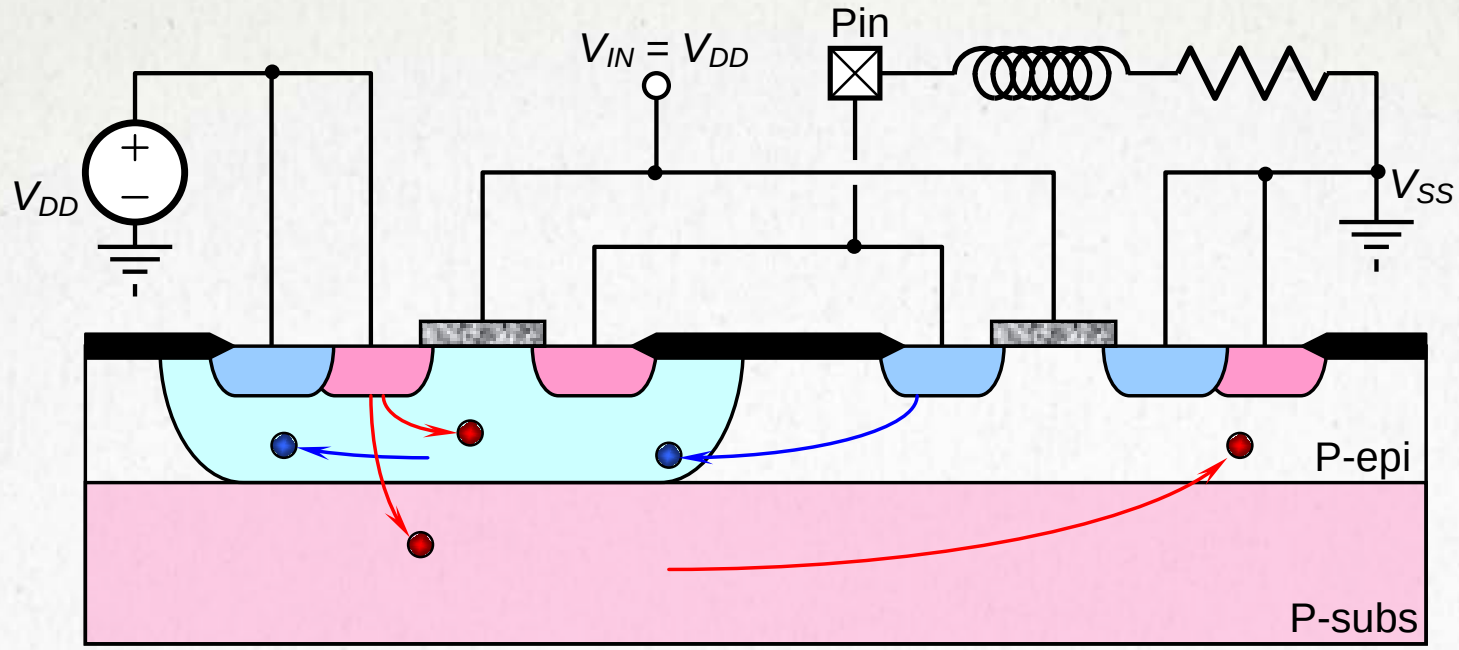
- **Pulling an N-type region below substrate potential injects electrons into the P-epi.**
  - These electrons become minority carriers in the P-epi.
  - They diffuse in all directions, although few enter the heavily doped substrate.

THE PARASITIC NPN



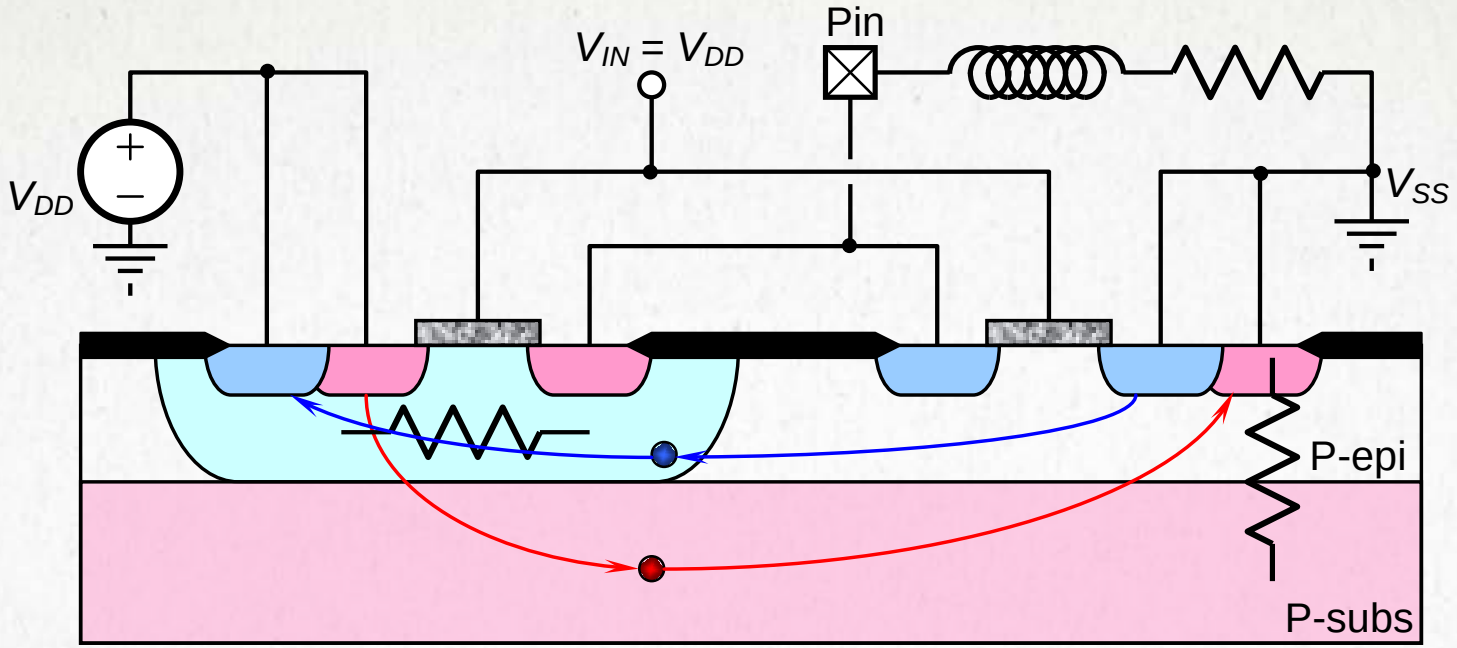
- **Most electrons recombine in the P-epi.**
  - This recombination consumes holes.
  - More holes flow in from substrate contacts.
  - A few electrons diffuse across to the N-well…

THE PARASITIC PNP



- **The forward-biased PMOS source/backgate junction injects holes into N-well.**
  - Some of these holes diffuse across the well and enter the P-epi or P-substrate.
  - A voltage gradient therefore appears across the N-well.
  - These holes flow to substrate contacts.

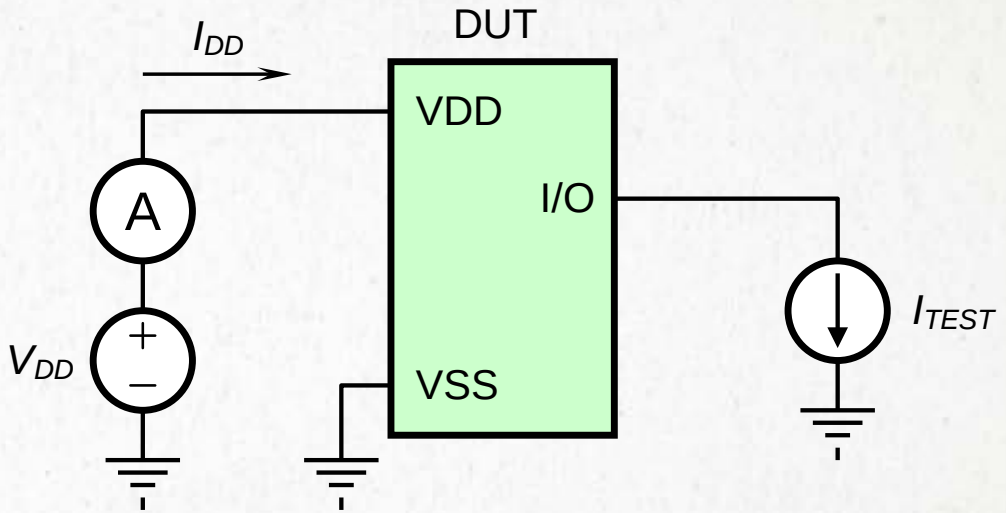
SUBSTRATE DEBIASING



- **The hole current flowing to the substrate contact must cross the P-epi.**
  - A voltage gradient appears across the P-epi.
  - This voltage gradient forward-biases the NMOS source-backgate junction.

DIGRESSION: LATCHUP TESTING

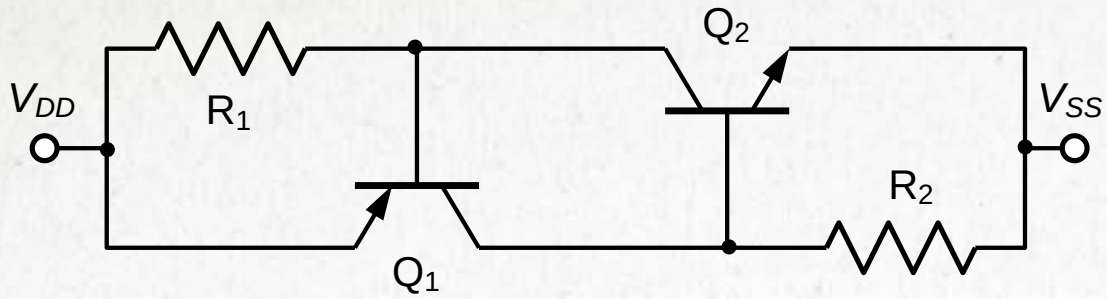
- **To test for latchup,**
  - Connect all power supplies to the *device under test* (DUT),
  - Record each supply current,
  - Inject a test current into/out of the pin under test,
  - Remove the test current,
  - If any supply current has significantly shifted (say by more than 10%), and the operation of the circuit cannot explain this shift, then latchup has occurred.



DIGRESSION: EMISSION MICROSCOPY

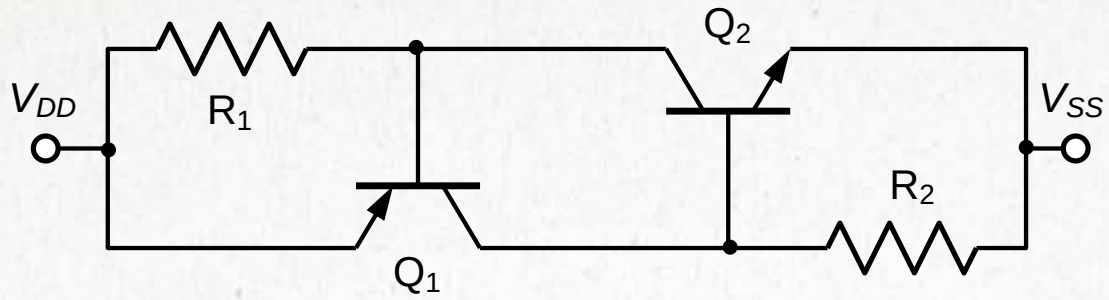
- **Minority carrier recombination generates low levels of photoemission, allowing an IR microscope to image latchup.**

CONDITIONS FOR LATCHUP



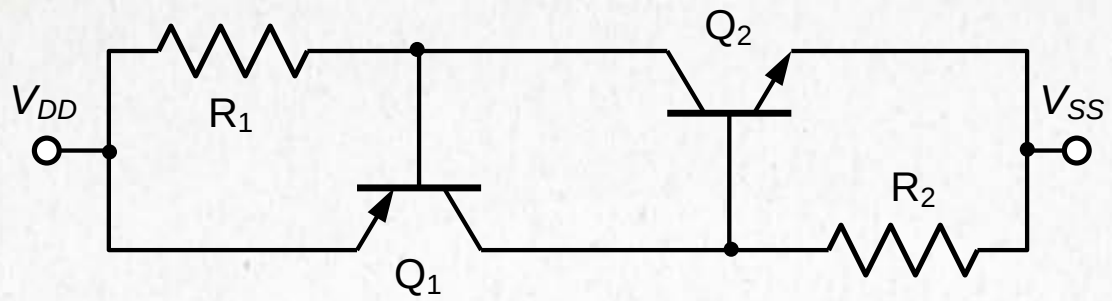
- **Parasitic PNP Q_1, parasitic NPN Q_2, well resistance R_1, and substrate resistance R_2 form a positive-feedback device called a *silicon-controlled rectifier* (SCR).**
- **In order for latchup to occur,**
  - $V(R_1) > V_{BE1}$
  - $V(R_2) > V_{BE2}$
  - $\beta_1\beta_2 > 1$

FOUR WAYS TO STOP LATCHUP



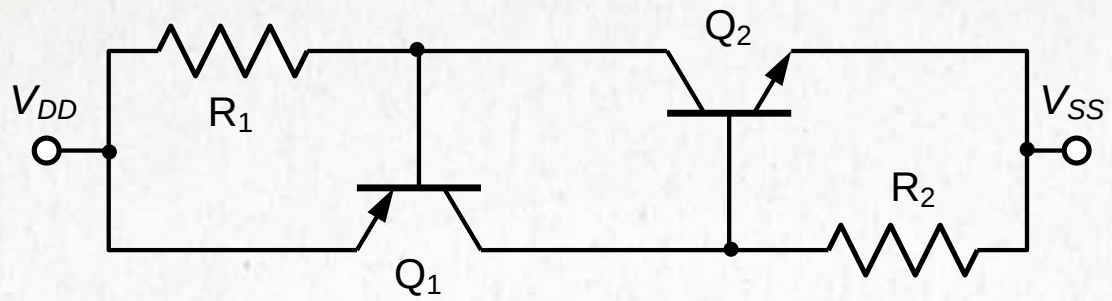
- **Reduce well resistance R_1.**
  - Add more well contacts.
  - Use a retrograde well.
  - Add an N-type buried layer (NBL) to the N-well.
- **Reduce substrate resistance R_2.**
  - Add more substrate contacts.
  - Increase substrate doping.

FOUR WAYS TO STOP LATCHUP



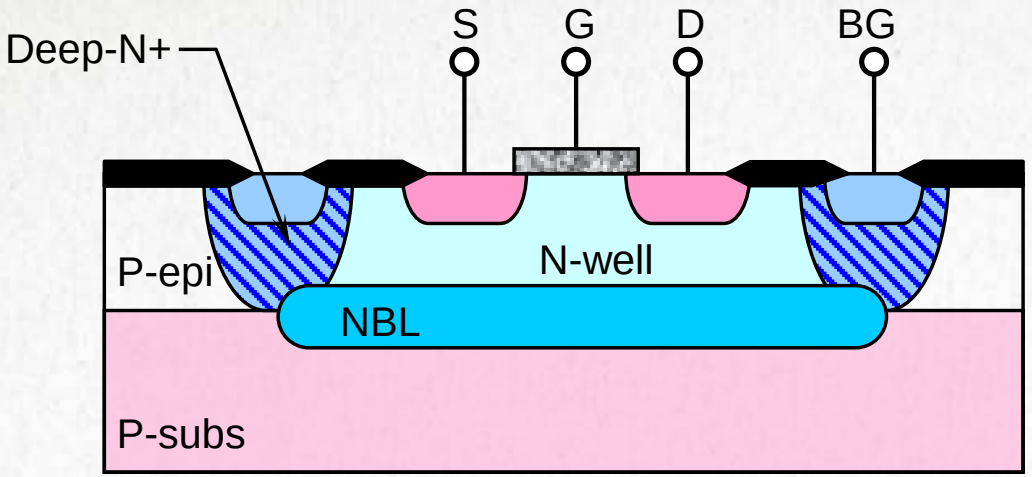
- **Reduce PNP Q_1's beta.**
  - Use a retrograde well or add NBL.
  - Increase overlap of well over devices.
  - Add a hole-blocking guard ring (HBGR).
  - Add a hole-collecting guard ring (HCGR).

FOUR WAYS TO STOP LATCHUP



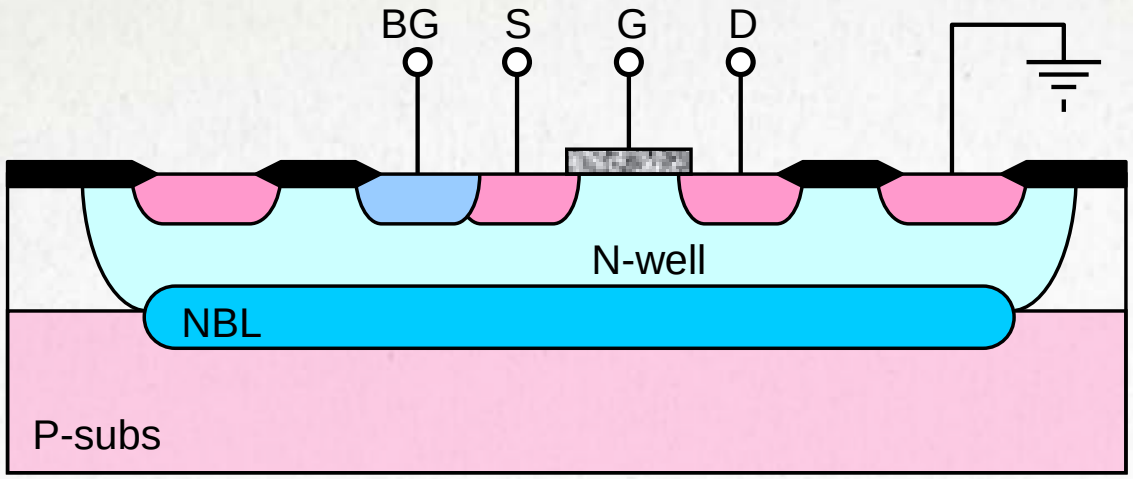
- **Reduce NPN Q_2's beta.**
  - Increase spacing between NMOS and N-well.
  - Employ an electron-collecting guard ring (ECGR).

THE HOLE-BLOCKING GUARD RING



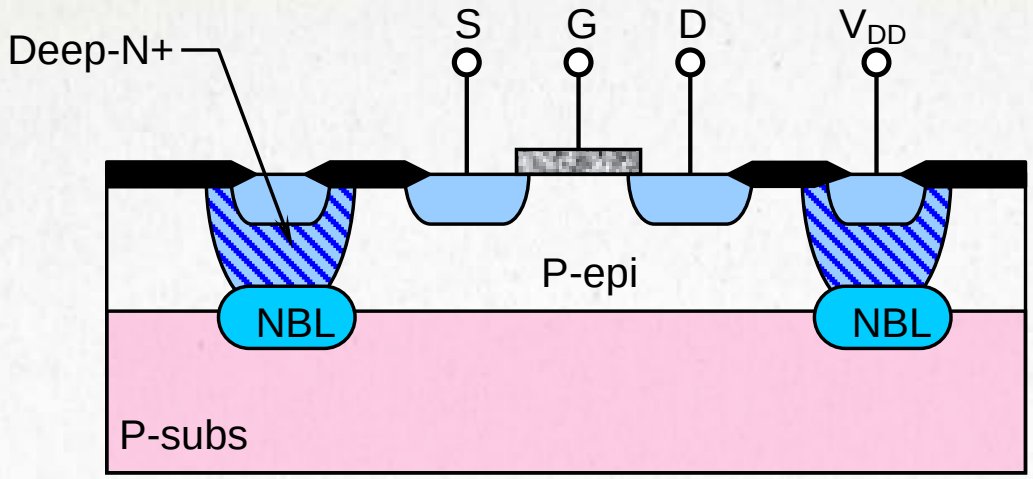
- **A hole-blocking guard ring (HBGR) blocks hole injection into the P-epi and P-substrate.**
  - Holes cannot surmount the electric field caused by the built-in potential of an N+/N– interface.
  - The HBGR uses a deep-N+ sinker and NBL to construct the N+/N– interface.
  - The N+ concentration must exceed the N– by 100:1 or more.

THE HOLE-COLLECTING GUARD RING



- **The hole-collecting guard ring (HCGR) collects some fraction of the holes before they can reach the substrate.**
  - NBL or a strong retrograde profile is necessary to block vertical hole injection.
  - The limited depth of PSD implants make them poor HCGR's.
  - P-well or P-epi can be used as an HCGR, but its low doping renders it vulnerable to debiasing.

THE ELECTRON-COLLECTING GUARD RING



- **The electron-collecting guard ring (ECGR) collects a fraction of the electrons before they can reach an N-well or NSD region.**
  - A P+ substrate is required to constrain the electrons within the P-epi.
  - The best ECGR structures use deep heavily doped diffusions, but N-well alone will often suffice if it is connected to a supply.

SUMMARY

- **All junction-isolated integrated circuits may experience latchup.**
  - All diffusions that hook either directly to pins, or to pins through less than about 50 kilohms of resistance, can potentially trigger latchup.
  - Capacitors switching within a circuit can sometimes cause minority carrier injection that triggers latchup.
- **Proper layout practices can minimize the risk of latchup.**
  - Adequate well and substrate contacts.
  - Maintain recommended spacings.
  - Add guard rings where feasible.