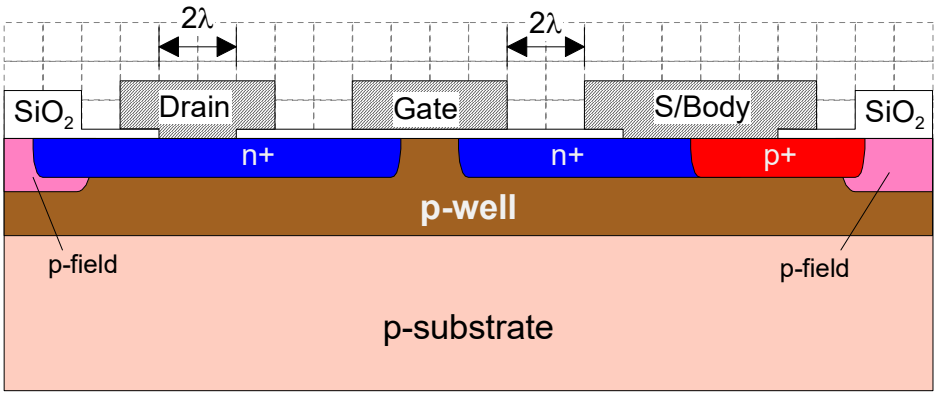
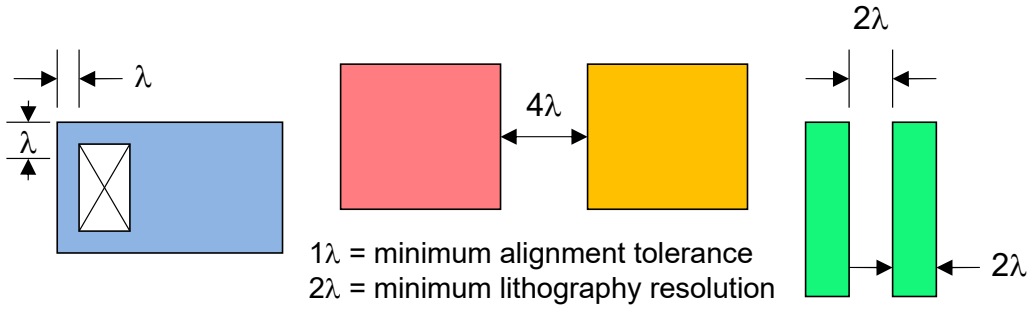
Non Self-Aligned Metal Gate MOSFET

- The self-aligned feature of the polysilicon gate MOSFET has allowed the VLSI technology to move into extremely fine photo-lithography resolution. Currently, we are approaching theoretical limit at around 20nm.
- The original MOSFET employs a metal gate structure that is non-self-aligned with significant disadvantages.



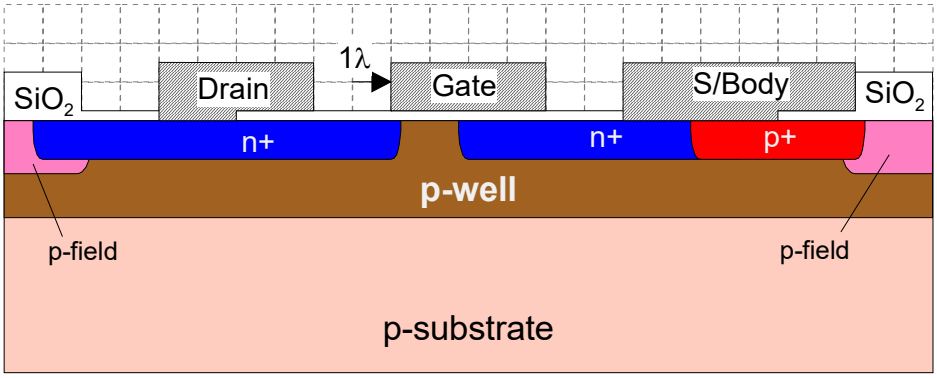
Non Self-Aligned Metal Gate MOSFET (cont'd)

- To simplify the layout of complicated transistor structures, a set of design rules often required.
- Design rules specify the minimum/maximum line widths, minimum overlaps and separations. They are determined by process tolerances such as photolithography resolution and alignment between different masking layers.
- The lambda (λ) rule is a scalable set of design rules that can be easily remembered.



Non Self-Aligned Metal Gate MOSFET (cont'd)

- The efficiency of the self-aligned structure can be better appreciated by examining the limitations on the layout of a metal gate MOSFET.
- In order to safeguard alignment tolerance, the gate electrode must be over-sized such that there will always be some overlap between the gate and the source/drain diffusion.

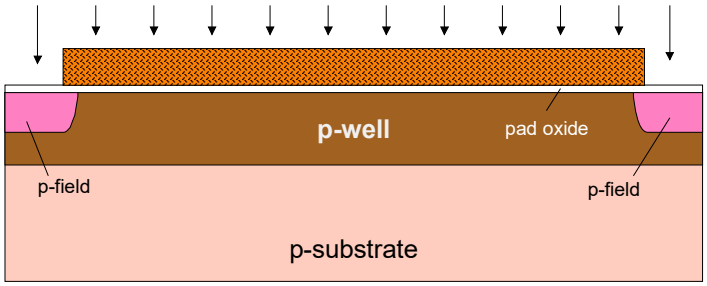


Non Self-Aligned Metal Gate MOSFET (cont'd)

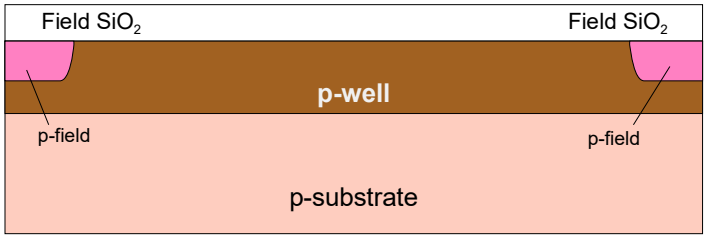
- The over-sized gate metal also leads to large overlap capacitance. This especially critical if the overlap is shifted toward the drain end, leading a large Miller capacitance.
- The fact that the same metal layer is also used for the source, gate and drain also require addition separation between electrodes. This further increase the width of the n+ S/D and p+ body diffusions. This not only takes up more space, the extra distance also lead to larger series resistance and junction capacitance.
- The following slides illustrates the fabrication sequence of the metal gate MOSFET. This will provide a better understanding on the reasons for the extra large peripheral areas.

Non Self-Aligned Metal Gate MOSFET (cont'd)

Boron p-field implant (p-field mask)



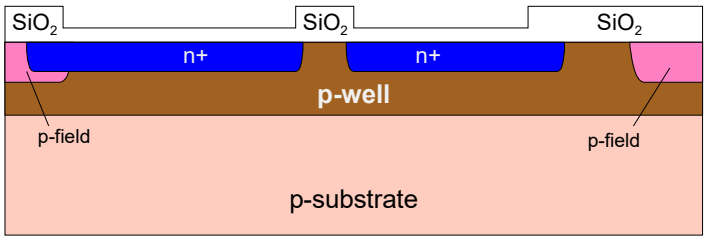
Grow thick field oxide



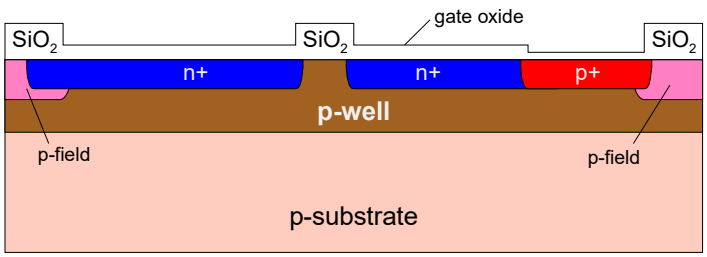
- This process starts with a conventional p-well on p-substrate.
- The p-field mask is used to define regions for the p-field implant.
- Wet oxidation is used to grow the thick field oxide.

Non Self-Aligned Metal Gate MOSFET (cont'd)

n+ diffusion (n+ mask to etch field oxide)
re-grow field oxide during drive-in



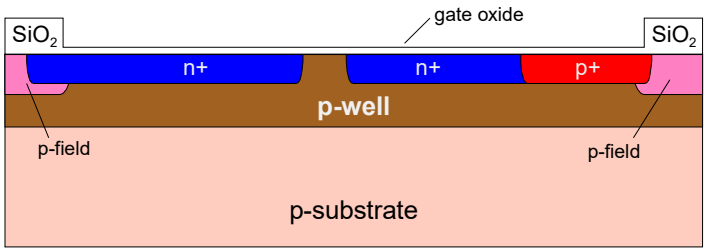
p+ diffusion (p+ mask to etch oxide)
re-grow field oxide during drive-in



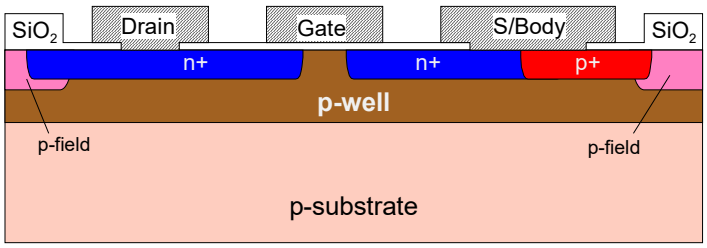
- The n+ mask is used to define the n+ S/D diffusions. A cap oxide layer is grown during drive-in.
- The p+ mask defines the p-body diffusion. A cap oxide layer is grown during drive-in.

Non Self-Aligned Metal Gate MOSFET (cont'd)

open active region (active mask to etch oxide in device area)
re-grow gate oxide



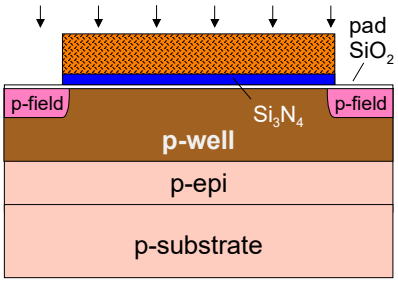
open contact holes (contact mask to etch oxide)
deposit metal (metal mask)



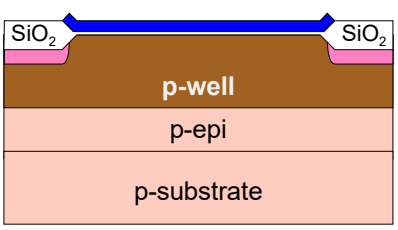
- The active region is defined followed by a complete oxide removal in the exposed region.
- Gate oxide is grown.
- Contact holes are defined and opened.
- Metallization completes the process.

Self-Aligned MOSFET Fabrication

Boron p-field implant (masked)



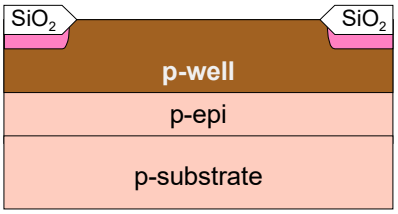
LOCOS Oxidation



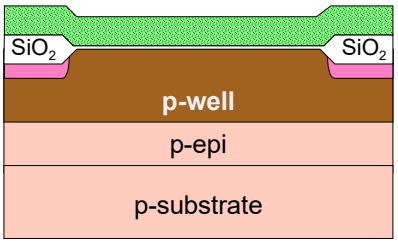
- The self-aligned poly-silicon gate has become the standard structure in modern VLSI technology.
- This illustration starts with the conventional p-epitaxial growth and p-well implementation.
- A thin pad oxide (~100Å) is grown to provide better adhesion for the subsequent CVD deposition of silicon-nitride (Si_3N_4).
- The Si_3N_4 is patterned by the **nitride mask**. The photo-resist is also used as the protective mask against the p-field implant. This is necessary to increase the V_{TH} of the inactive region.
- LOCOS oxidation is performed with Si_3N_4 protecting the covered region.

Self-Aligned MOSFET Fabrication (cont'd)

strip Si_3N_4 and pad oxide



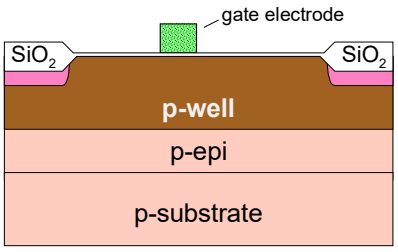
grow gate oxide
deposit gate polysilicon



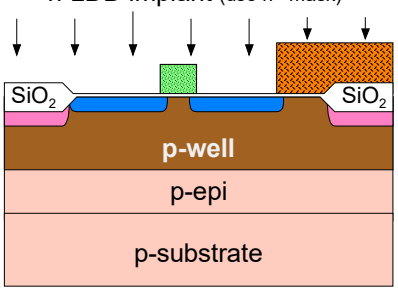
- A wet etching step is used to remove the Si_3N_4.
- The thin layer of pad oxide is also removed using a timed wet oxide etch.
- This results in regions with thick and thin oxide layers.
- After wafer cleaning and gate oxide growth, s layer of polysilicon is deposited immediately to avoid contamination of the gate oxide.
- Sometimes, a **threshold adjust implant** is carried out prior to the polysilicon deposition.

Self-Aligned MOSFET Fabrication (cont'd)

gate etch
(only critical masking step)



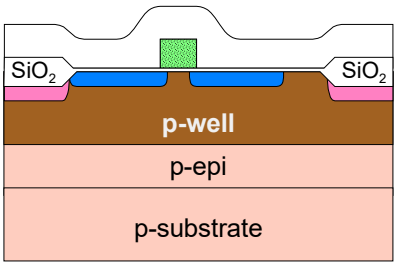
n-LDD implant (use n+ mask)



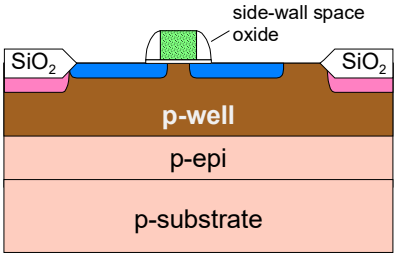
- The **gate mask** is used to define the gate electrode using dry etching (RIE).
- This is the most important lithography step as the gate length (e.g. 90nm, 60 nm, etc.) is the primary indicator of the technology performance.
- The n+ mask (will be re-used later for n+ implant) is used to define the region where **n-LDD** will be placed.
- The alignment of this mask is not critical, but still have to be within process tolerance).
- The gate electrode also serves as a blocking layer to define the critical location of the n-LDD on both the source and drain sides

Self-Aligned MOSFET Fabrication (cont'd)

Deposit CVD oxide



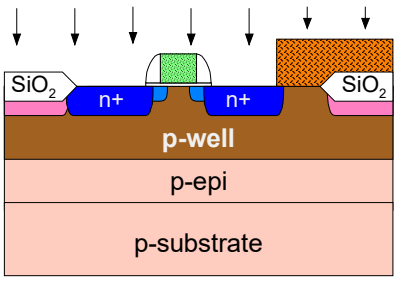
Dry Etch to form Side-Wall Spacer



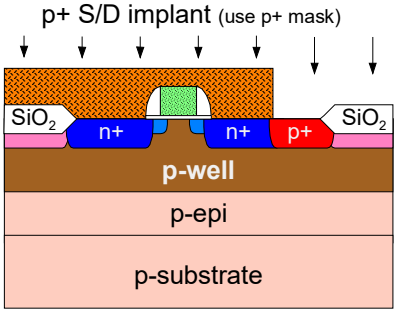
- A thick layer of CVD oxide is then deposited over the entire wafer. Note that the deposition is not exactly conformal, resulting in thicker oxide layers whenever there is a step in the surface topology.
- A timed RIE etch (usually with a chemical end point detection) is used to remove just enough CVD oxide to expose the wafer surface.
- Due to the thicker CVD oxide region on both edges of the gate electrode, small pieces of oxide remain. This is the **side-wall spacer**.
- The side-wall spacer is critical as it protects the n-LDD region from subsequent n+ S/D implant.

Self-Aligned MOSFET Fabrication (cont'd)

n+ S/D implant (use n+ mask)



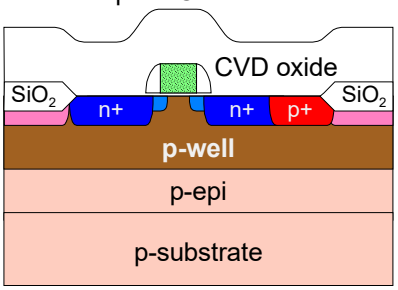
p+ S/D implant (use p+ mask)



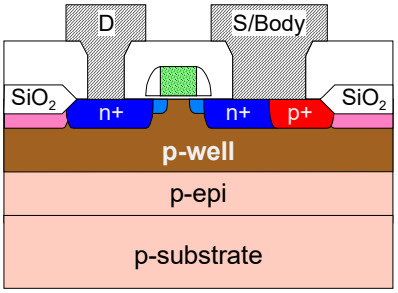
- The **n+ mask** is re-used to define the areas where n+ S/D diffusion are required.
- The side-wall spacer at both the left and right edges of the gate electrode should have sufficient thickness to prevent n+ impurity from penetrating. This is an important self-align feature of the polysilicon gate MOSFET.
- Note that even though the n+ mask does not cover part of the field oxide region, n+ implant is not able to penetrate those regions.
- The p+ is re-used (it was used previously to form the p-LDD region – not shown) to the p+ body contact.

Self-Aligned MOSFET Fabrication (cont'd)

Deposit CVD oxide



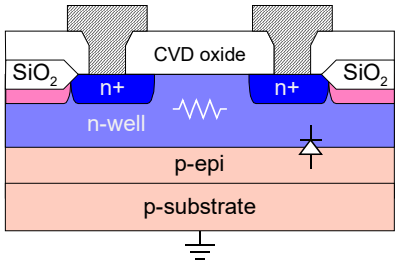
Planarization, Contact Holes, Metallization



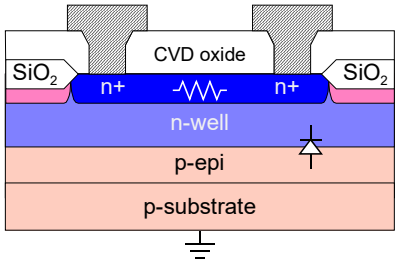
- A thick layer of CVD oxide is deposited on the wafer, followed by planarization.
- The **contact hole mask** is used to define the opening in the CVD oxide.
- Metal 1 (first layer of metal) is sputtered and patterned using the **metal mask** (most modern processes have multiple layers of metals).
- Not shown on this process sequence is the encapsulation. This consists of a thick CVD Si_3N_4 layer to hermetically seal the wafer from moisture or other corrosive environment.

Integrated Resistors

n-well resistor



n+ diffusion resistor



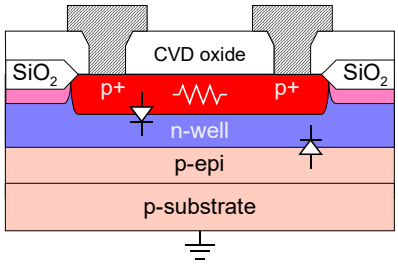
- The n-well resistor uses the n-well as the resistive element. The junction depth and doping concentration are fixed by the process. The user can only adjust the W/L ratio to control the resistance. The sheet resistance R_s of n/p-well are usually in 100's Ω/□.
- p-well on p-substrate will not be a good choice as the resistor will be shorted to GND.
- Since the p-substrate must be tied to the most negative potential, the voltage on the resistor must be positive.

Integrated Resistors

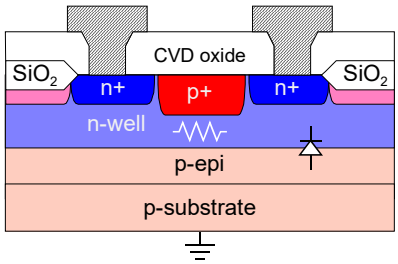
- Due to the doping concentration of the n-well is comparable to that of the p-epi/substrate, the depletion region will extend into both the n-well and the p-epi/substrate regions significantly. This results in a substantial variation in resistance as the voltage applied to the resistor is increased. This variation is normally characterized by a voltage coefficient (VC) in ±Ω/V.
- In addition, whenever a depletion region exists, there will be a junction capacitance. The n-well resistance has a substantial parasitic capacitance between the n-well and GND.
- To reduce the VC, n+ diffusion resistor can be used. Since majority of the current will flow through the n+ diffusion, variation in the n-well depletion region will not significantly affect the resistance.

Integrated Resistors (cont'd)

p+ diffusion resistor



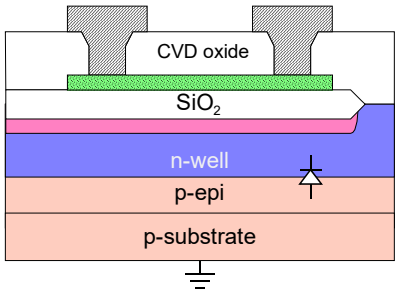
pinched n-well resistor



- Similarly, p+ diffusion can also be used as a floating resistor. However, due to the higher doping concentration in the p+ region, depletion region will be very narrow, leading to a large parasitic capacitance.
- The R_s for both the n+ and p+ diffusion regions is usually in 10's Ω/□.
- If a higher sheet resistance is required, the pinched resistor can be used. The p+ diffusion reduces the cross-section of the current flow path. R_s in 1000's Ω/□ can be obtained.

Integrated Resistors (cont'd)

polysilicon resistor

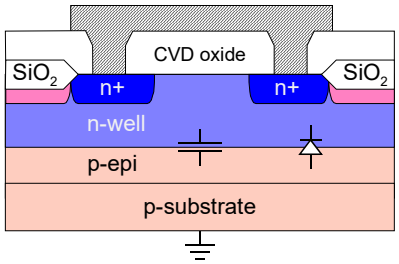


- Due to the difficulty in controlling the doping concentration, it is not uncommon to have resistance tolerance as higher as ±20%.

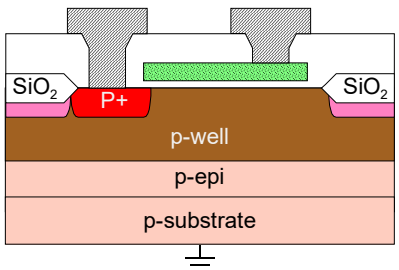
- The polysilicon layer is not only useful as the gate electrode. If an additional mask is used to block n+ and p+ diffusion from the poly layer, it can have a very high R_s (e.g., 1000's Ω/□).
- By placing the poly layer on top of the thick field oxide, the parasitic capacitance can also be minimized.
- Since the poly layer is isolated from the substrate (due to the field oxide), this resistor can accommodate both positive and negative voltage swing.

Integrated Capacitors

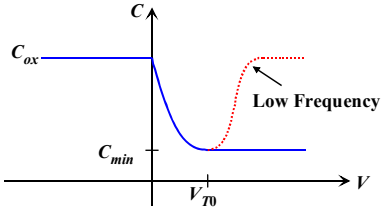
junction capacitor



MOS capacitor

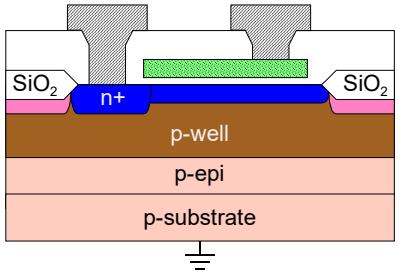


- The n-well region can also be biased to function as a depletion capacitor. However, the VC (in ±pF/V) and tolerance are very high. Due to this variation, this type of capacitor is also called a variactor which is useful in tuning circuits.
- The MOS structure can also be used as a capacitor. However, as in the junction capacitor, the bottom plate is GND. The VC for a MOS capacitor on moderately doped p-well is very high.

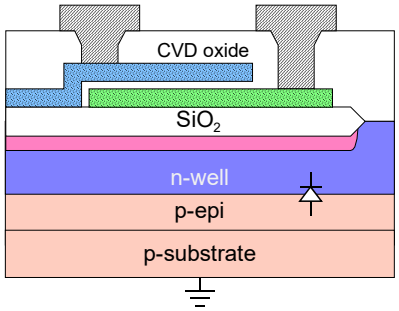


Integrated Capacitors (cont'd)

improved MOS capacitor

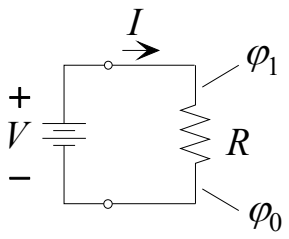


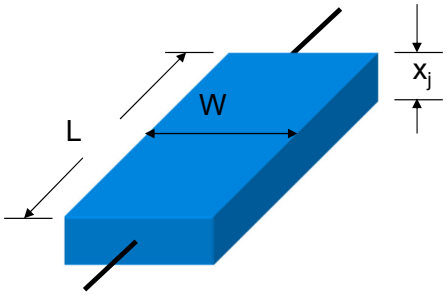
Inter-poly capacitor



- To reduce the VC and to allow a more flexible voltage swing, a new mask can be used to place an n+ diffusion (not the same as the n+ S/D diffusion) under the MOS stack.
- If the process can afford a second polysilicon layer, an inter-poly capacitor can be formed. This produces the most flexible capacitor with minimal parasitic.
- MOS and inter-poly capacitors are very accurate and with very good matching (±1%).

Properties of Semiconductor Material





- Resistor structure — Ohm's Law

$$I = \frac{V}{R} \qquad (1.1)$$

- The resistance depends on
  - ⇨ resistivity, ρ
  - ⇨ length of the structure, L
  - ⇨ cross-sectional area

$$A = x_j W$$

$$R = \rho \frac{L}{x_j W} \qquad (1.2)$$

- Conductivity, $\sigma = 1/\rho$ (1.3)

Resistivity (cont'd)

- The resistivity can be expressed as

$$\rho = \frac{1}{q\left(\mu_n n + \mu_p p\right)}$$

$$\rho = \frac{1}{q\mu_n N_D} \qquad \text{for } n\text{-type semiconductor}$$

$$\rho = \frac{1}{q\mu_p N_A} \qquad \text{for } p\text{-type semiconductor}$$

Ohmic Contacts

- In many situation, **ohmic** metal-semiconductor contacts with linear I-V characteristics and minimal resistance in both biasing directions are required.
- One way to obtain ohmic contacts is to choose a metal with a metal work function that is smaller than that of the *n*-type semiconductor. For *p*-type semiconductor, the metal work function should be larger.

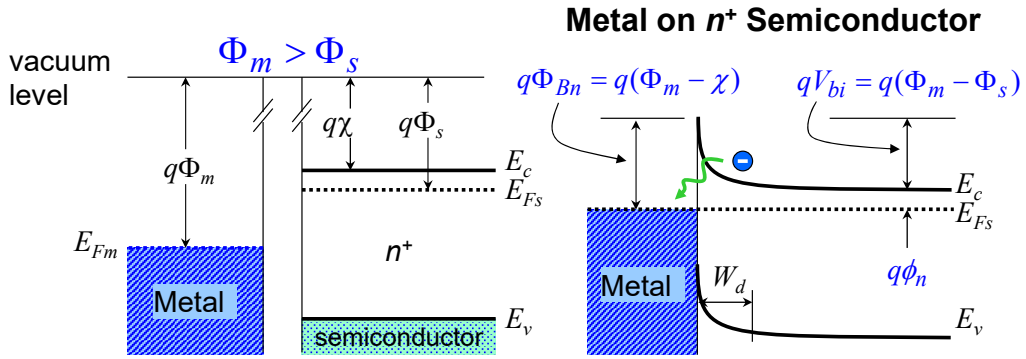
$\Phi_m < \Phi_s$ for *n*-type Si

$\Phi_m > \Phi_s$ for *p*-type Si

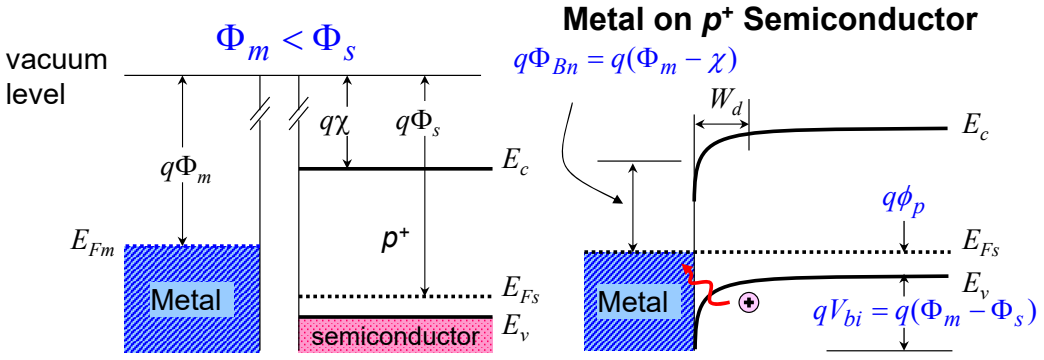
- However, in IC fabrication, it is not practical to use different metals to form Schottky barriers and ohmic contacts on *n*-type and *p*-type semiconductors.

Ohmic Contacts (cont'd)

- A common technique to form ohmic contacts is to deposit the metal on heavily doped semiconductors.
- The conduction process for this type of contact is essentially by tunneling in both biasing directions.



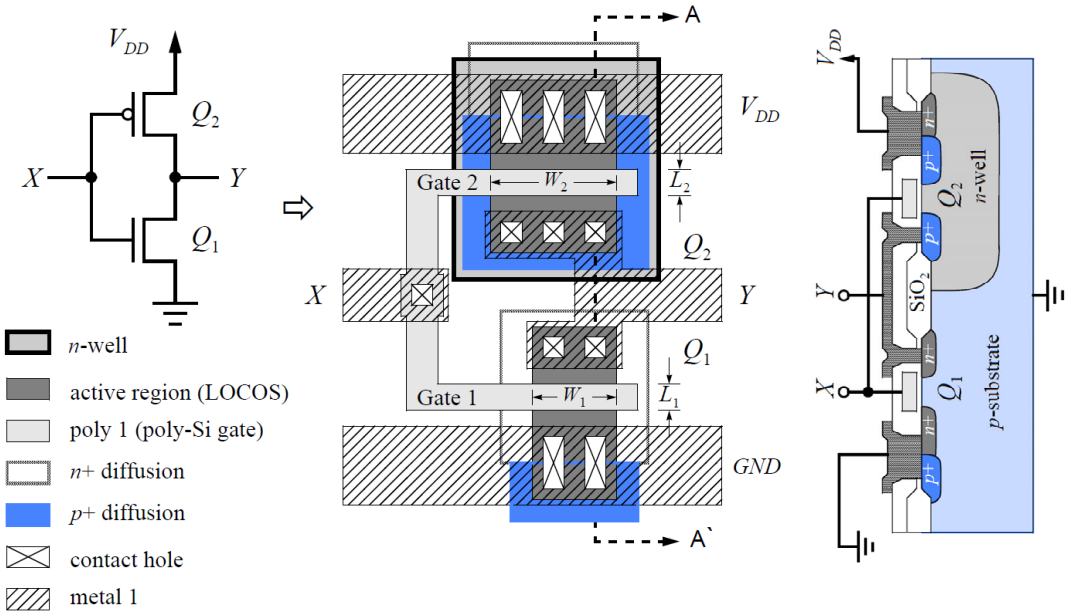
Ohmic Contacts (cont'd)



- The same situation can also exist in a metal to p^+ semiconductor (since we need to have ohmic contacts to p^+ regions as well)
- Holes in the p^+ region can tunnel through the thin depletion region (due to the heavily doped region).

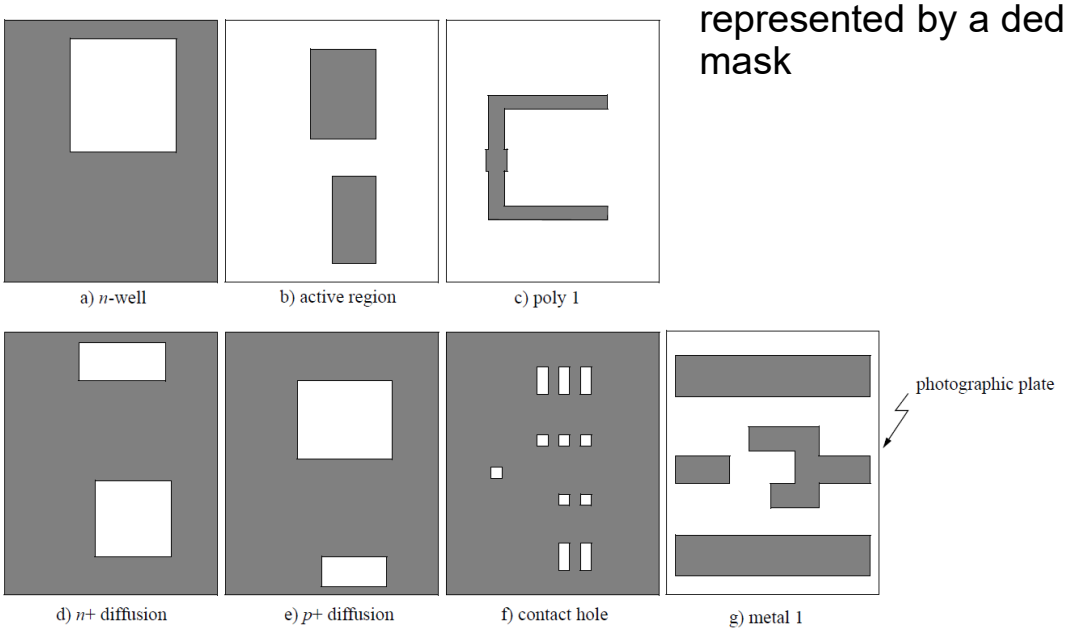
Integrated Circuit Layout

- Layout and the corresponding cross-section of a simple CMOS inverter

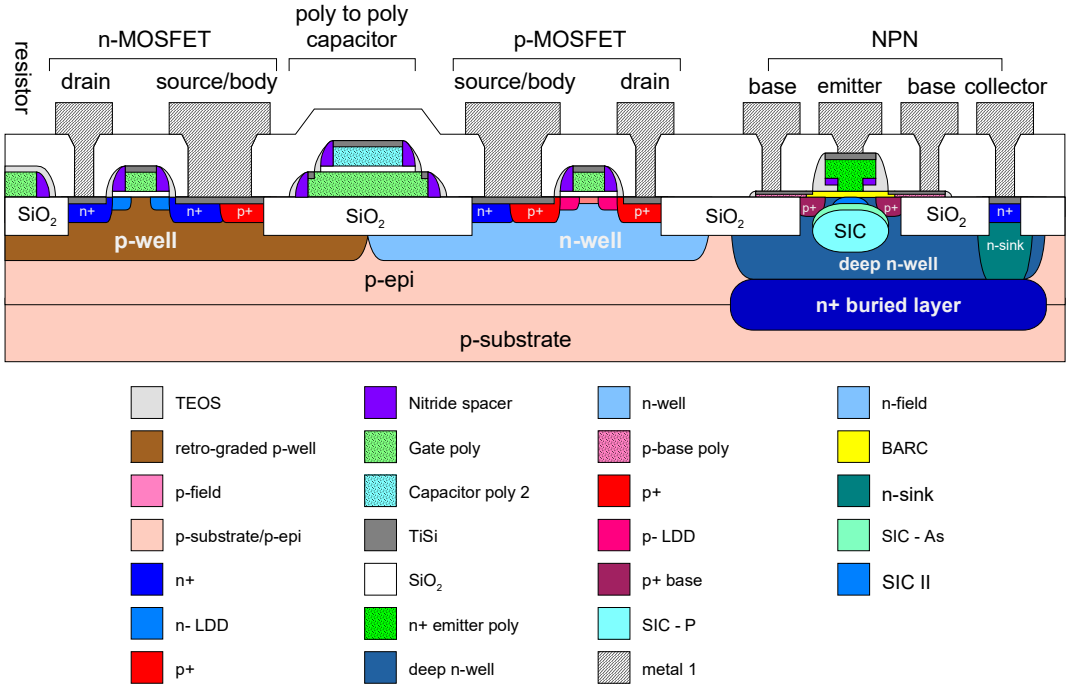


Integrated Circuit Layout (cont'd)

- The corresponding mask-set to implement the CMOS inverter. Each layer is represented by a dedicated mask



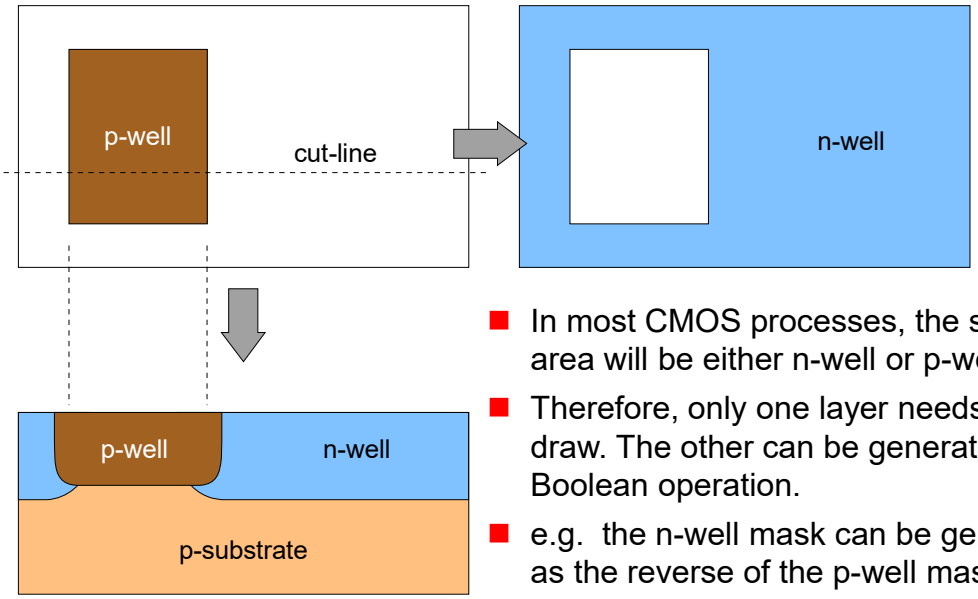
Layout Design Rules



Layout Design Rules (cont'd)

- Each fabrication technology has a dedicated set of masking layers, depending on available features, complexity, etc.
- The cross-sectional diagram of a BiCMOS technology is shown in the previous page. There are a few basic layers that are almost universal for all standard CMOS processes:
  - n-well, p-well
  - n-field, p-field
  - active region
  - gate (poly 1)
  - n-plus, p-plus
  - contact holes
  - Metal (m1)
- These layers are often named using a cryptic abbreviation.
  e.g. NW = n-well, NF = n-field, CNT = contact, etc.
- It is not necessary to draw all the layers using a computer aided design (CAD) tool. Some are generated using logic operations.

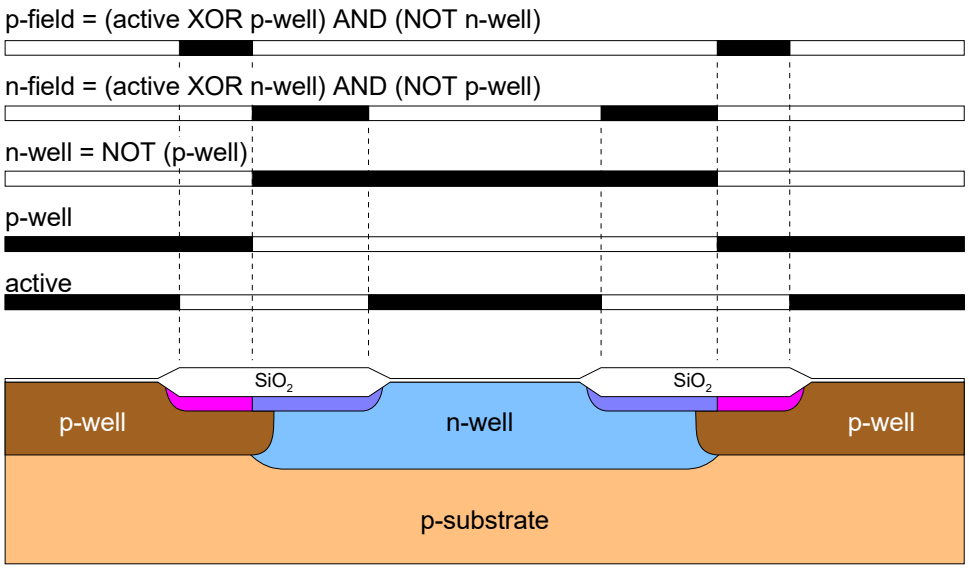
Generating Masking Layers



- In most CMOS processes, the surface area will be either n-well or p-well.
- Therefore, only one layer needs to be draw. The other can be generated using Boolean operation.
- e.g. the n-well mask can be generated as the reverse of the p-well mask.

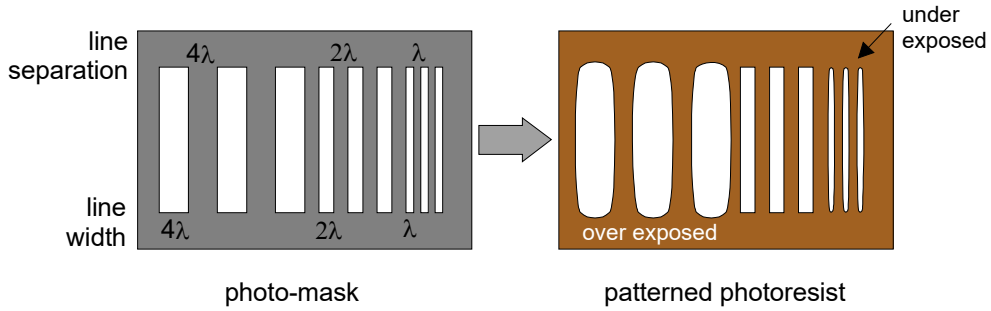
Generating Masking Layers (cont'd)

- Other layers such as p-field, n-field can also be generated from combination of p-well and active masks.

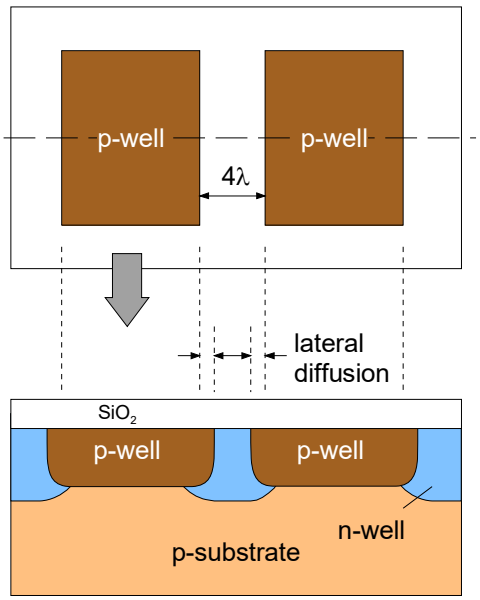


The Lambda Rules

- Design rules are very specific to a particular fabrication technology.
- The λ-rule is a simple and scalable design rules where most rules are either 2λ or 1λ.
  - 2λ = minimum line width, minimum separation of the same or between masking layers.
  - 1λ = alignment tolerance, hence the minimum overlap between relevant layers.
- The ability to pattern the photo-resist will determine the value of λ for a particular technology.

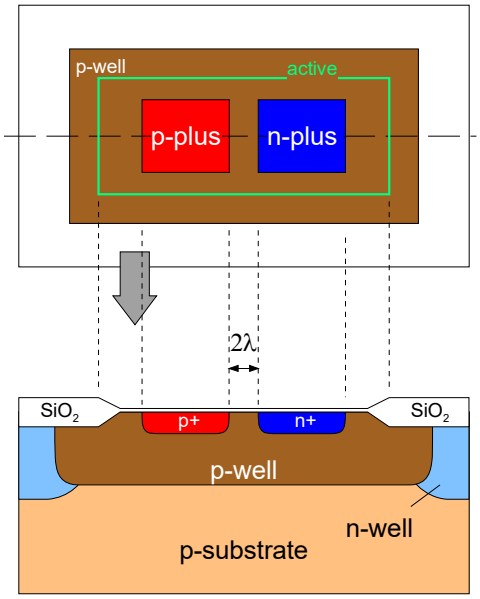


The Lambda Rules (cont'd)



- Separation rule between patterns of the same layer may be determined by the amount of lateral diffusion. e.g., n-well and p-well.
- In this case, if you don't want to have the 2 p-well layer shorted to each other, there must be enough separation that is larger than the lateral diffusion from both edges.
- A minimum separation of 4λ between patterns of the same masking layer is common, if they are subjected to long drive-in procedures.

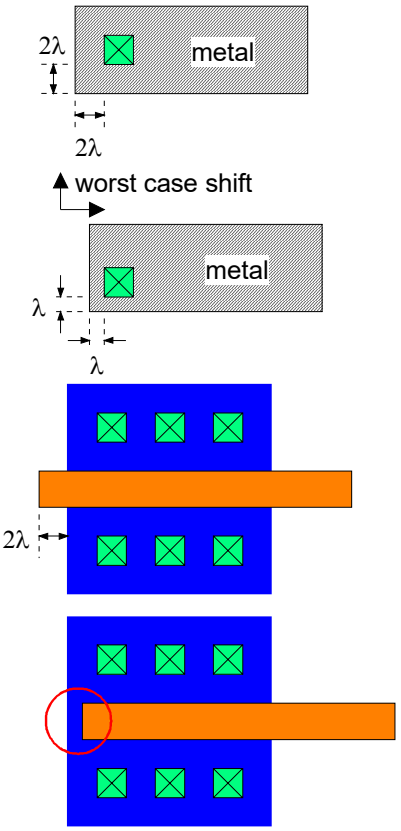
The Lambda Rules (cont'd)



- For patterns between different layers that do not exhibit large lateral diffusion (e.g. n-plus and p-plus), smaller separation can be used.
- Note that for accuracy of process requirement, n-plus and p-plus can only be placed inside active regions.
- n-field and p-field regions are not shown for clarity.

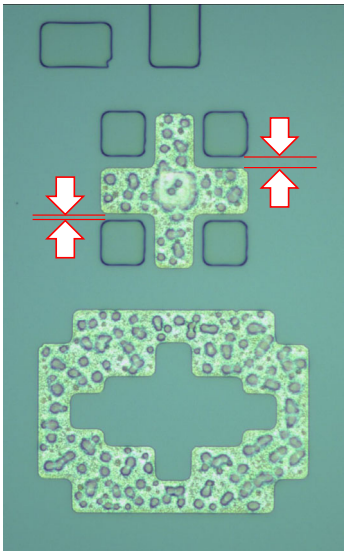
The Lambda Rules (cont'd)

- Overlapping rules are important to provide sufficient coverage between layers even under worst case mis-alignment.
- Metal layer must surround a contact hole with a minimum of 2λ on all sides, otherwise some contact holes may not be covered by metal.
- For MOSFETs, the gate poly mask must overlap the ACTIVE mask by at least 2λ. Otherwise, a short between the source and drain may result.



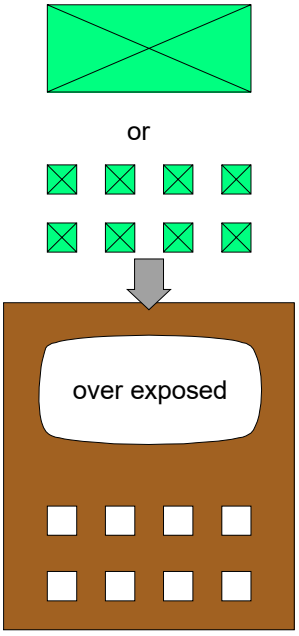
The Lambda Rules (cont'd)

- The accuracy of the alignments between layers are normal checked (under the microscope) using alignment patterns
- There are many variations of useful patterns
- In general, existing patterns (or marks) must be visible when the mask is placed on top
- The mask can be moved in x, y directions, as well as rotated
- Misalignment is indicated by uneven spacing between mask and existing pattern

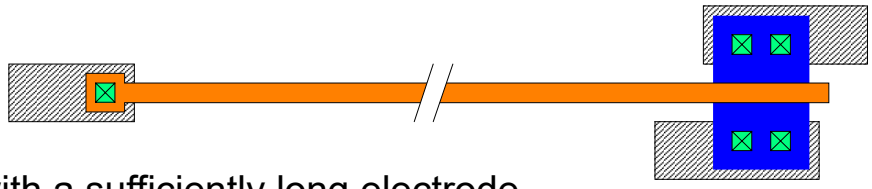


The Lambda Rules (cont'd)

- Exposure rules also dictate that all geometries on the same mask should be as consistent as possible (in size).
- This is especially important for small geometries such as CONTACT holes and VIA masks.
- For example, suppose a large contact hole is desired. It is more appropriate to assemble the smaller and uniform contact holes to form the same contact.
- Large openings in any mask will tend to be over-exposed during the lithography step. If the process engineer reduces the exposure time to adjust for a more accurate opening, the smaller contact holes will become under-exposed.

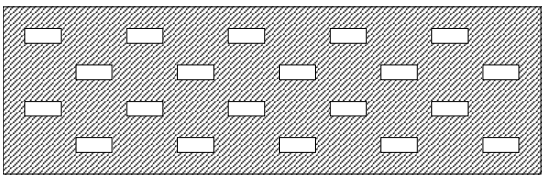


Antenna Rule



- During RIE steps, an intense RF field is applied. Any MOS capacitor like structure with a sufficiently long electrode will pick up charges from the plasma in the RIE chamber. This charge could cause oxide (especially gate oxide) to be punctured.
- The antenna rule is used to ensure that the ratio of the area of the gate electrode over field oxide to the area of the gate electrode over transistor must be less than a certain number. e.g. 100
- If a long interconnection using gate-poly require such area ratio, the designer will have to break up the long interconnect into a combination of poly-via-metal connections.
- Failure to follow the antenna rule will compromise the yield of the IC fabrication.

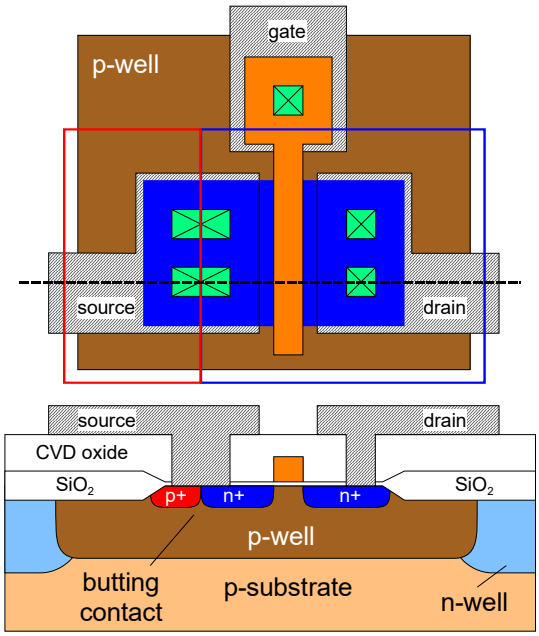
Electromigration Rule



- To carry a large current, the designer must use wider metal lines.
- Aluminum used in most VLSI fabrication processes can carry approximately a few mA/μm for a 1μm thick metal layer.
- During conduction, and under large electric field, metal ions can be moved. This results in metal particles to be literally knocked off (i.e., like a fuse). This phenomenon is called **electromigration**.
- Possible solutions:
  - Use wider metal to handle the large current. Note the current limits specified by the silicon foundry for different metal layers.
  - Use copper instead of aluminum. Copper has better immunity to electromigration.
  - Avoid 90° bends in the metal lines.
  - Metal slotting (usually part of the design rule, see picture on top). The slots control the size of metal grain size which can directly affect the current limit.

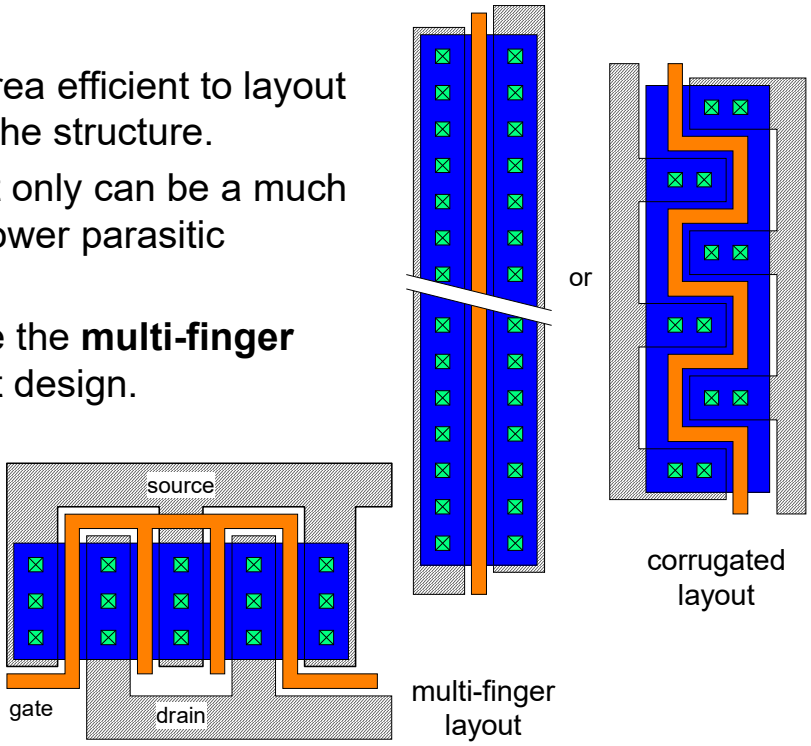
MOSFET Layout

- A typical MOSFET layout is shown with the accompanied cross- sectional diagram.
- The layout starts with the definition of the ACTIVE region, followed by the p-well.
- The location of the transistor is defined by the intersection between the GATE POLY layer and the ACTIVE layer.
- n-plus and p-plus layers cover a large area, but is only effective over the ACTIVE region (note that the GATE POLY will also receive these implants)
- n-field, p-field, and n-LDD are not shown for clarify.



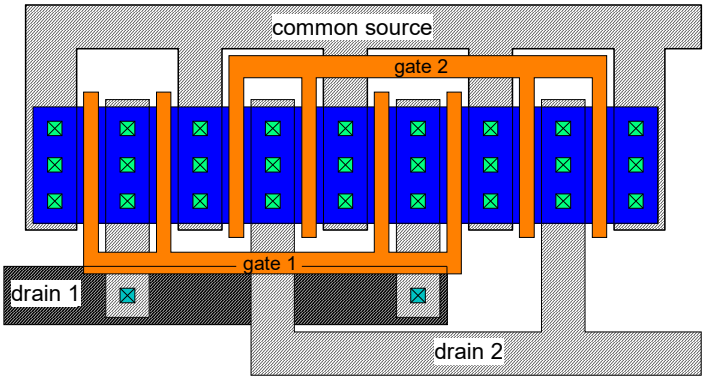
MOSFET Layouts (cont'd)

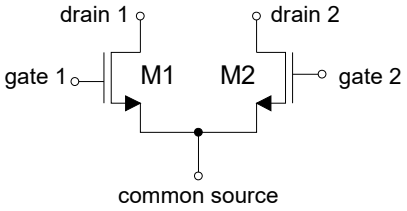
- For large transistors, it is not area efficient to layout the device by simply elongate the structure.
- A corrugated gate structure not only can be a much better layout, but also lead to lower parasitic capacitance and resistance.
- An alternate approach is to use the **multi-finger** layout to form an area compact design.
- Body contact, n-plus and p-plus layers are not shown.



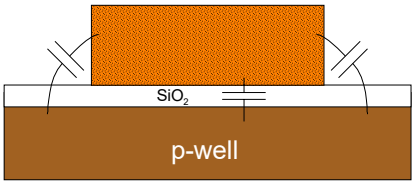
MOSFET Layouts (cont'd)

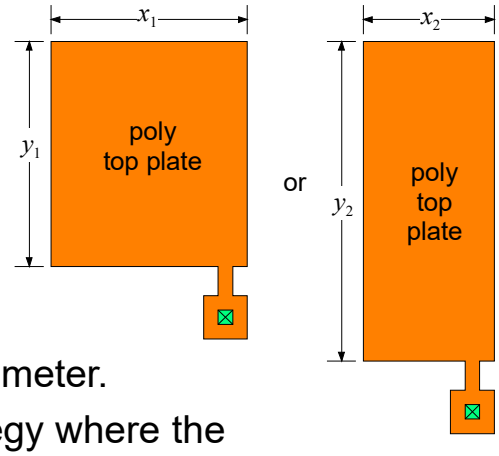
- For differential pairs, it is important to ensure good matching. Placing transistor close to each other is not good enough as any systematic variation will cause some mismatch.
- The multi-finger layout approach can be used to distribute the over different parts of the layout to ensure that both transistors will experience the same process and temperature variations.





Capacitor and Resistor Layouts



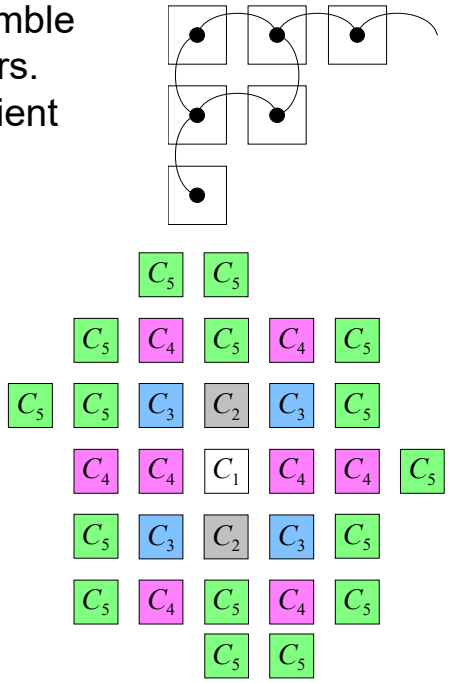


- Capacitor and resistor structures are sensitive to both layout area and perimeter.
- A good layout should choose a strategy where the perimeter is minimized for a given area.

$$\text{perimeter} = 2(x+y) \quad \text{must be minimized}$$

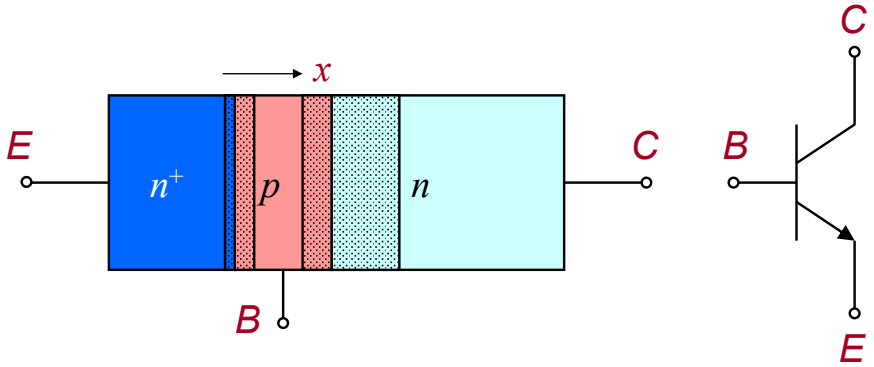
Capacitor and Resistor Layouts (cont'd)

- For best component matching, it is best to assemble capacitors (or resistors) using unit size capacitors. The unit size capacitance can be of any convenient size. e.g. 0.5pF, 1pF, etc.
- In situation where a series of components must be match from a small ratio to a large ratio, a **common centroid** layout technique should be used.
- For example, for an array of capacitors in 1:2:4:8:16…. ratio, the capacitors should be centered around the smallest capacitor and surrounds it with progressively larger capacitors.



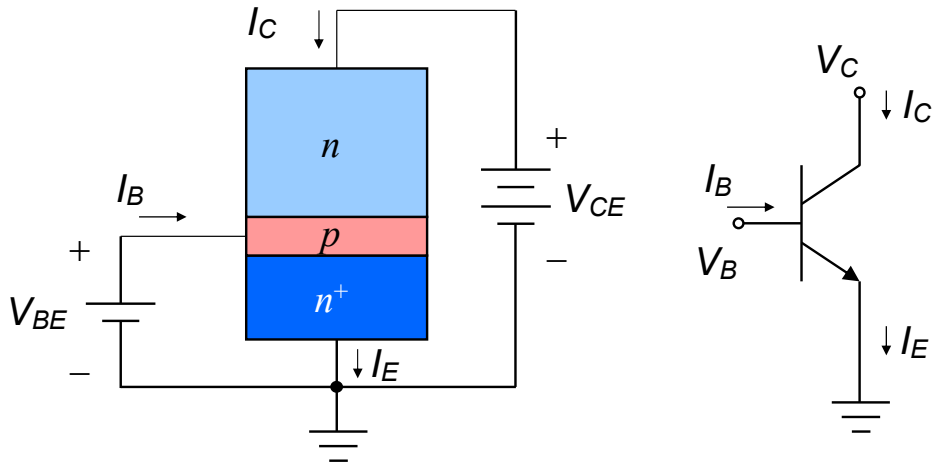
Bipolar Junction Transistors

- Early BJTs were fabricated using alloying — a complicated and unreliable process.
- The structure contains two p-n diodes, one between the base and the emitter, and one between the base and the collector. The device consists of three alternating n and p regions — npn structure. (one can also reverse the construction to get a pnp structure.)



Typical BJT Operation

- The operation of the bipolar junction transistor depends on the bias condition across the EBJ and CBJ



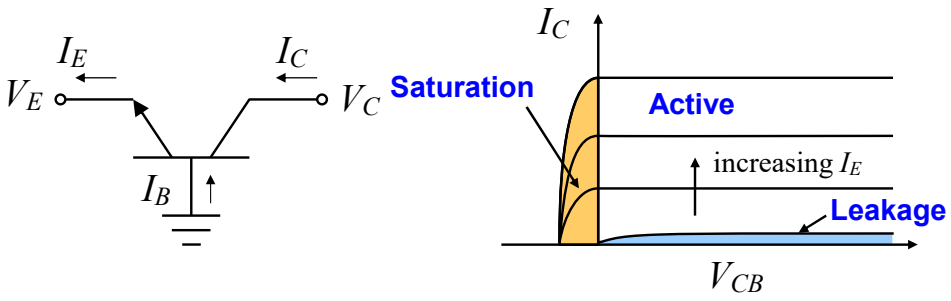
Modes of Operation

- The are 4 possible mode of operation. The reverse mode is not very useful and is not included in the following table

	EBJ	CBJ
Active mode:	Forward	Reverse
Cut-off mode:	Off/Reverse	Don't care
Saturation mode:	Forward	Forward

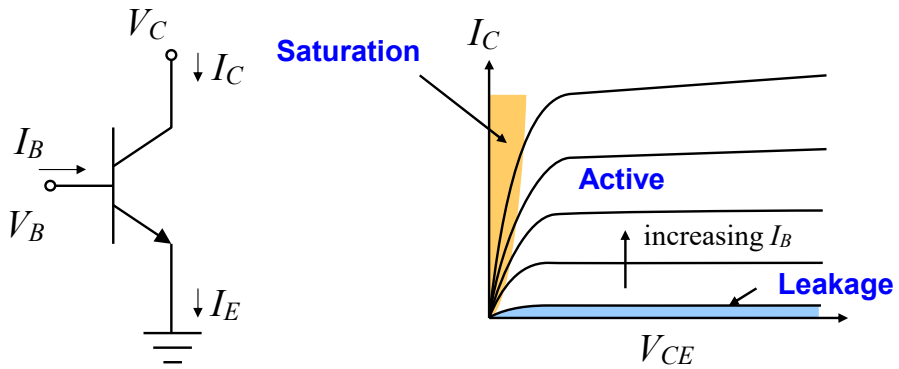
IV Relationship: Common Base

- Saturation occurs when CBJ is forward biased
- When $I_E = 0$, the device is cut-off, I_C is the reverse leakage current at the CBJ
- Note that $I_C \neq 0$ for VCB = 0. The current is contributed by I_E if the BEJ is forward biased

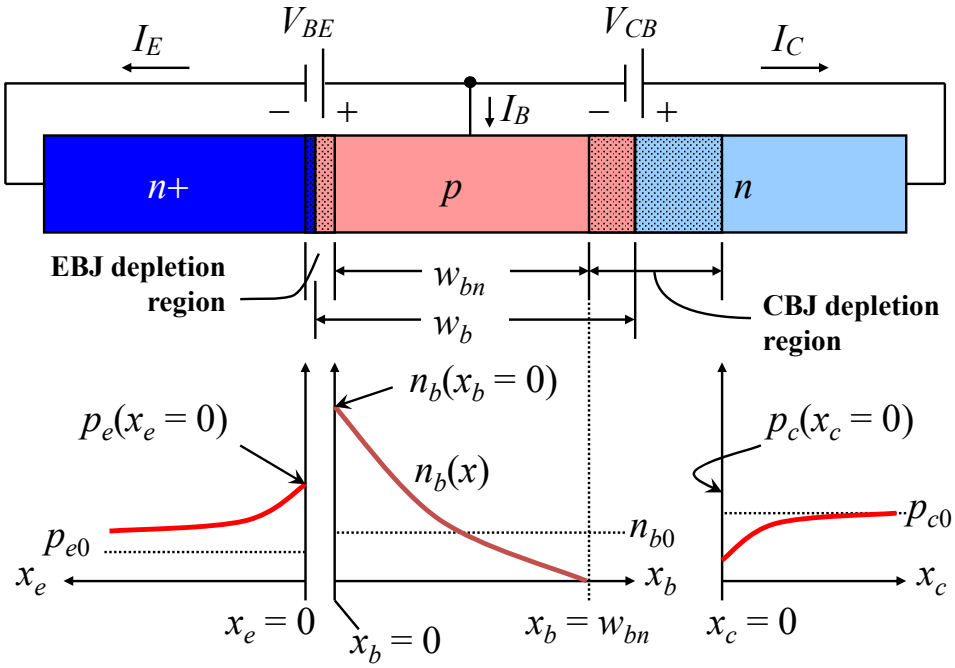


IV Relationship: Common Emitter

- **Saturation** occurs when both EBJ and CBJ are forward biased
- **Active** mode is the most useful for linear applications
- Saturation and cut-off modes are most useful for switching applications

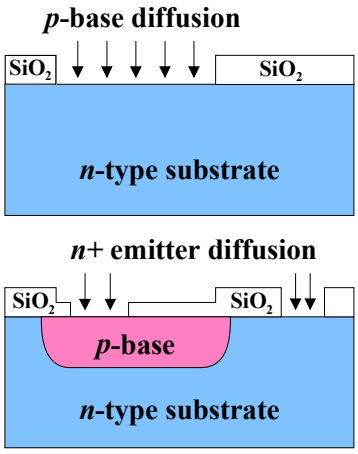


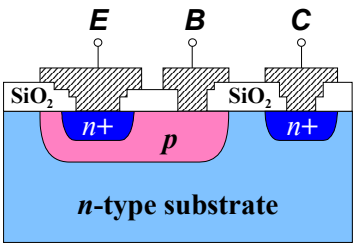
Carrier Profile in Active Mode



The Planar Process

- The “Planar Process” developed by Fairchild in the late 50s shaped the basic structure of the BJT, even up to the present day.





In the planar process, all steps are performed from the surface of the wafer.

Steady State Characteristics of the BJT

I_n^{EB} = emitter current injected into the base

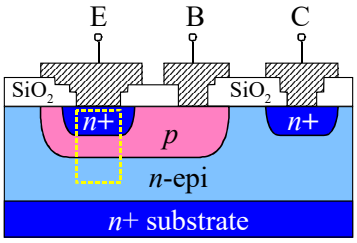
I_p^{BE} = base current injected into the emitter

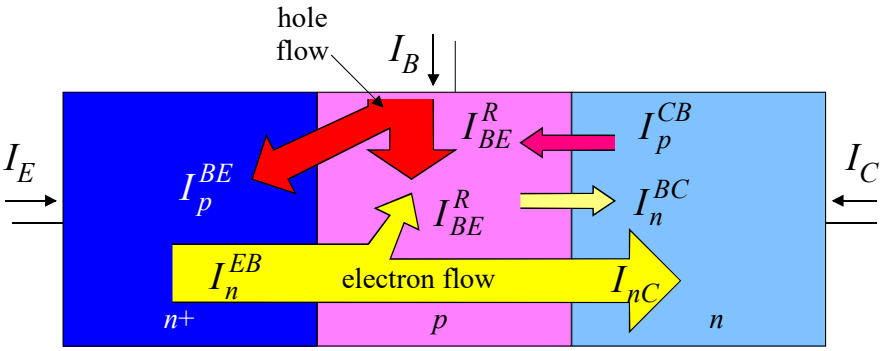
I_{BE}^{R} = recombination current in the base region

I_p^{CB} = reverse biased current across BCJ

I_n^{BC} = reverse biased current across BCJ

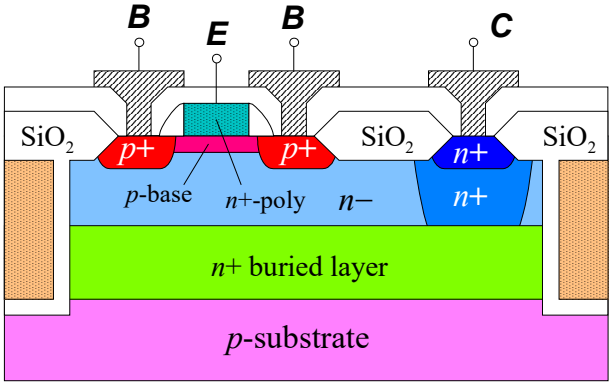
I_{nC} = electron current from the emitter





Trench Isolated BJT

- To further reduce the area between adjacent mesa, trench isolation can be used, making use of trench etching.
- The trench is typically 2μm wide and 5μm deep. The trench walls are oxidized and the remaining volume is filled with polysilicon.
- The starting silicon is a 3-layer structure consisting of p-substrate with n and n+ epitaxial layers to serve as the buried layer collector.



SiGe HBT BiCMOS

- Modern wireless communication applications require circuit operation at frequencies up to 30GHz, a regime well out of the realm of ordinary silicon materials.
- To service the semiconductor needs for this marketplace, compound III-V semiconductors (those made from elements from groups III and V in the periodic table) such as GaAs have been used successfully.
- These materials, however, are very expensive and have substantially added to the semiconductor cost of new high-tech products.
- If bipolar silicon is to continue to advance in speed, an alternative technology is needed.

SiGe HBTs

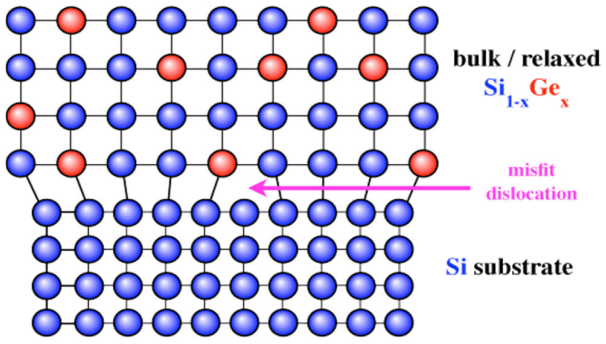
- Silicon germanium, a compound which, while compatible with silicon chip manufacturing processes, had some important differences.
- SiGe would permit device designers to simultaneously increase speed, reduce electronic noise, and lower power supply requirements.
- One key benefit is that SiGe chips are made with essentially the same fabrication equipment as silicon chips.
- This means millions of dollars won't have to be invested in new semiconductor fabrication facility, as is typically the case when a shifting to the next generation technology.

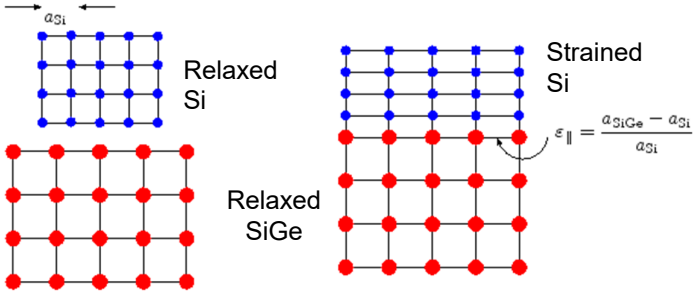
Basics of SiGe HBT Technology

- The advantage of the SiGe heterojunction bipolar transistor (HBT) lies in the ability to re-engineer the band gap of ordinary silicon semiconductor material for higher performance.
- Conventional silicon devices have a fixed band gap of 1.12eV (electron-Volt) which limits switching speed, compared to III-V compound materials such as GaAs.
- With the addition of germanium to the base region of a bipolar junction transistor (BJT) and grading the Ge concentration across the transistor base region, it is possible to modify the band gap to enhance performance of the silicon transistor.

Basics of SiGe HBT Technology

- Due to misfit of the lattice spacing, dislocation is inevitable if a large amount of germanium atoms are incorporated into the silicon substrate.
- This mismatch can also be used to form strained silicon for CMOS.

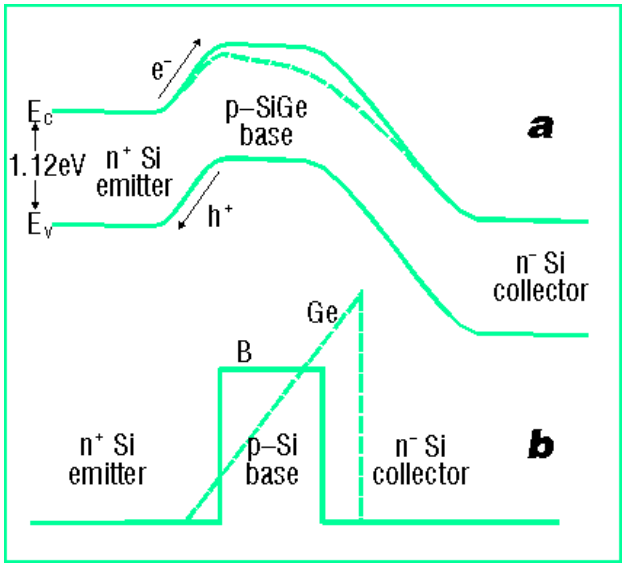




Source: http://www.iue.tuwien.ac.at/phd/dhar/node12.html
http://userweb.eng.gla.ac.uk/douglas.paul/SiGe/misfit.html

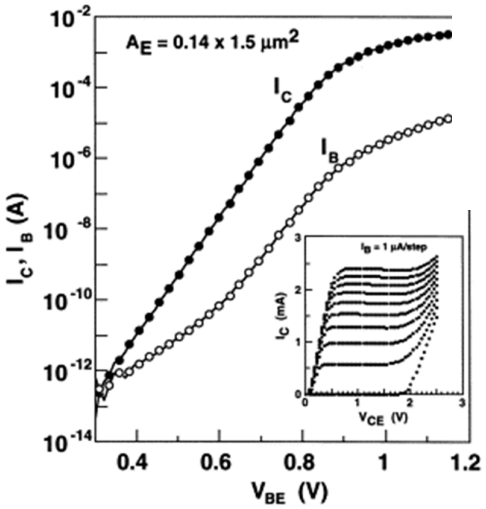
Basics of SiGe HBT Technology (cont'd)

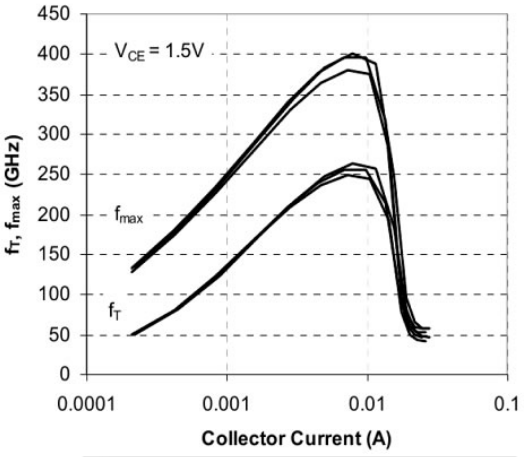
- Energy band diagrams of conventional Si bipolar transistor (solid line) and SiGe HBT (dashed line) showing band gap modification which speeds electron conduction across the base region by reducing the bandgap and creating a ramped potential surface ("drift field").

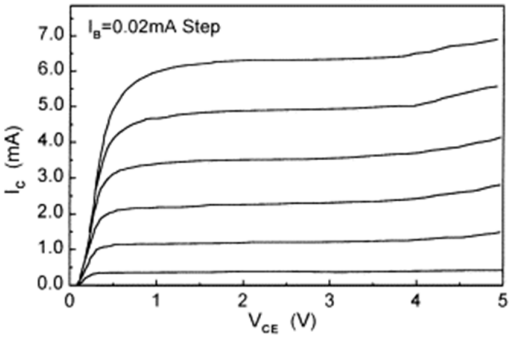


Typical Performance

- f_T, f_{max} vs I_C
- Gummel Plot, IV curves

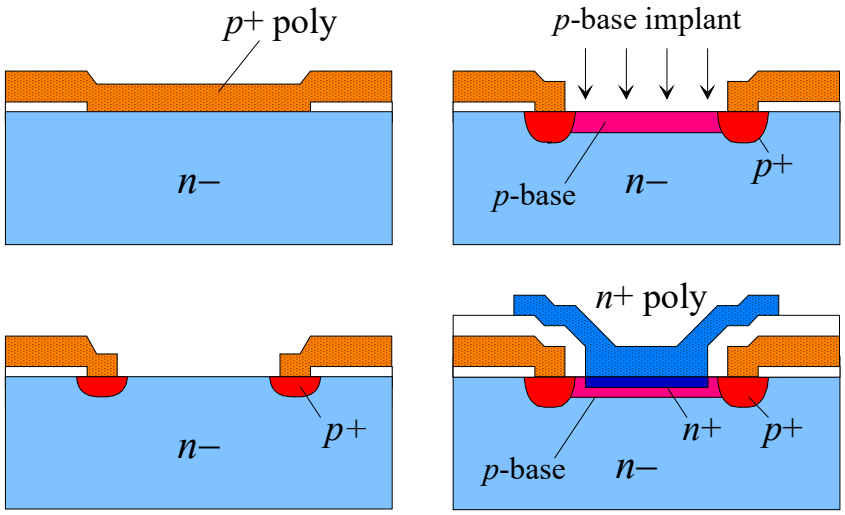






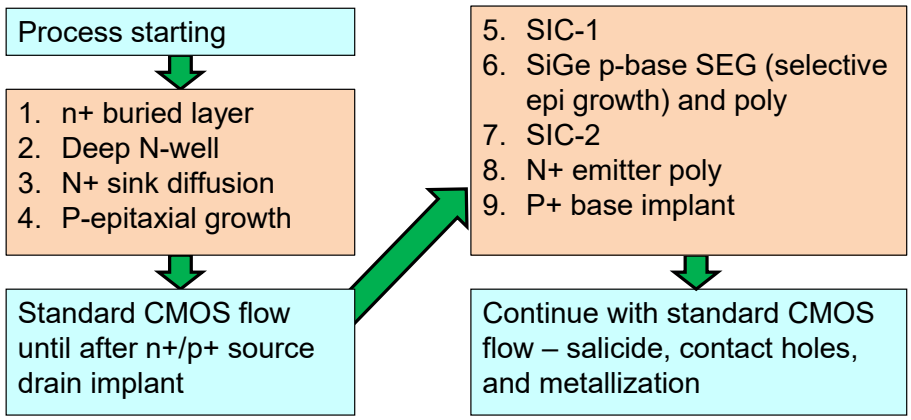
Double Poly Transistors

- A further extension of the self-aligned BJT structure is to use double polysilicon (n+ for emitter, p+ for base) to reduce the area required for contacts.

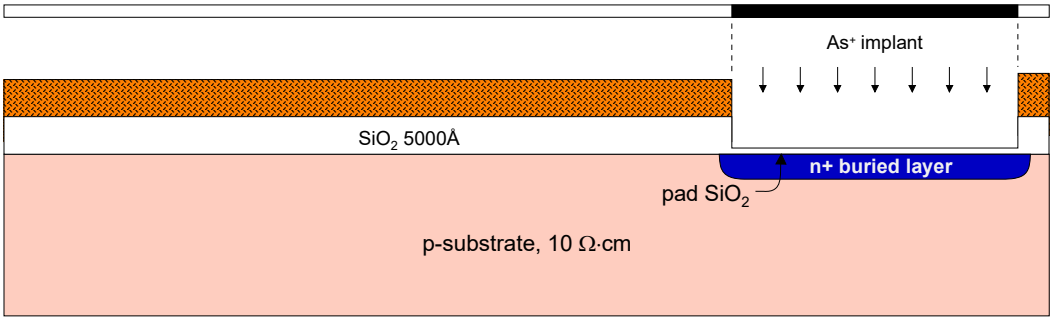


SiGe BiCMOS Process Flow

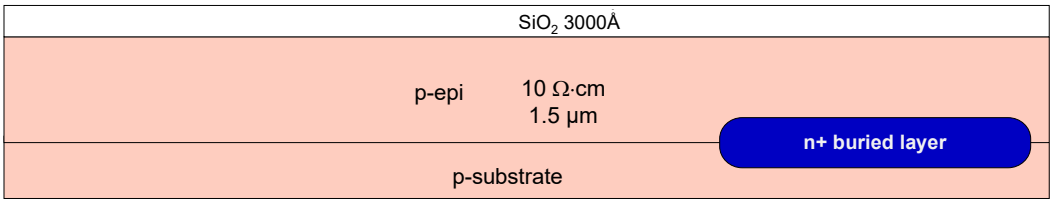
- This process flow is based on a standard CMOS process flow. Additional steps are inserted in between the normal CMOS process flow to accommodate the fabrication of the SiGe HBT (hetero junction bipolar transistor) without altering the baseline CMOS devices.



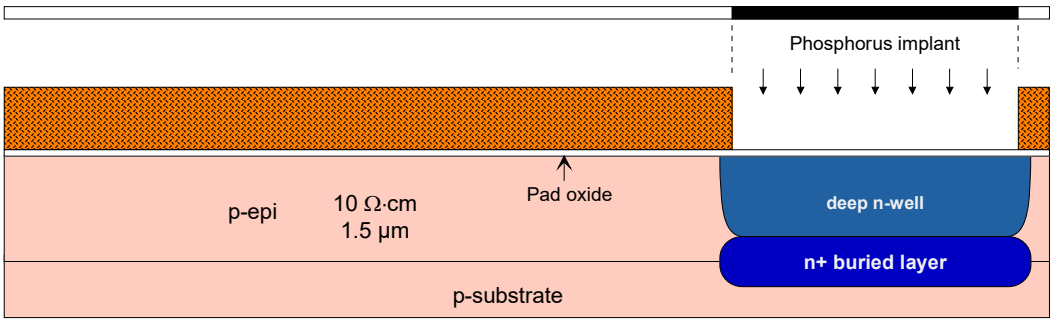
1. BNP mask, n+ buried layer implant



2. Oxide removal, p-epitaxial layer, n+ buried layer drive-in and oxidation

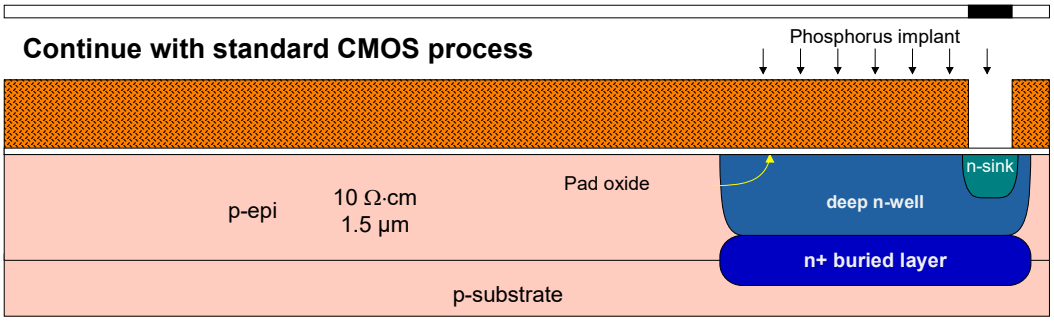


3. DNW mask, deep n-well implant and drive-in

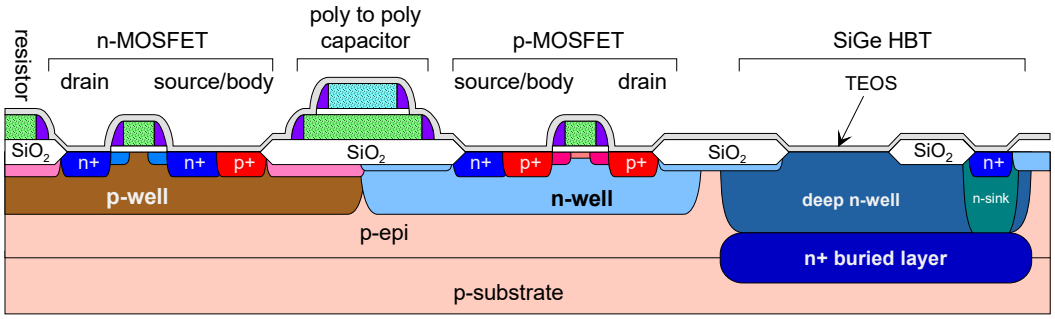


4. Deep n-well drive-in/oxidation, NSK mask, n-sink implant/drive-in

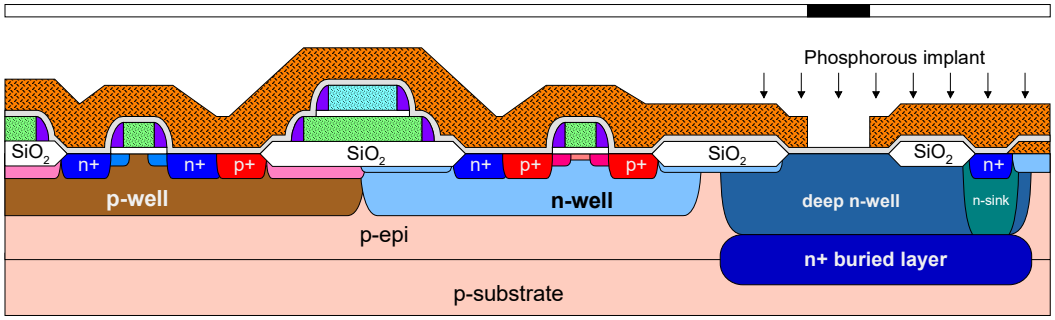
Continue with standard CMOS process



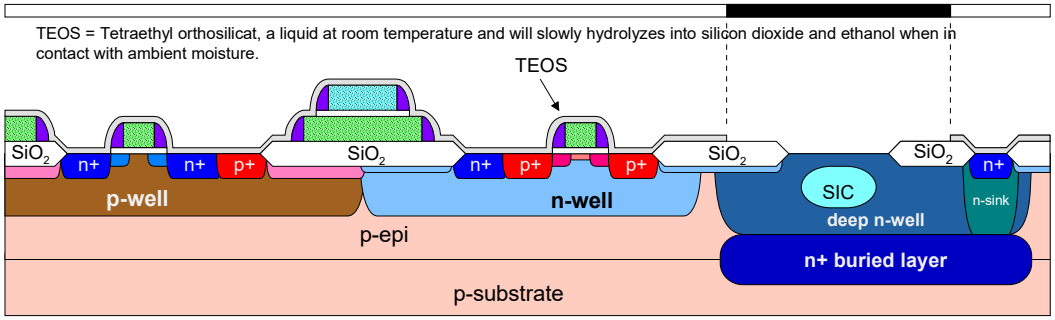
5. Completion of CMOS portion, just after p+ implant, HBT process begins



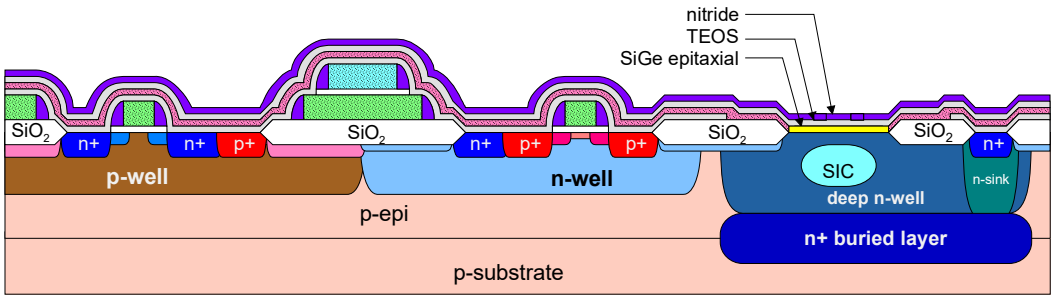
6. SIC mask defines selectively implanted collector region to reduce Rc



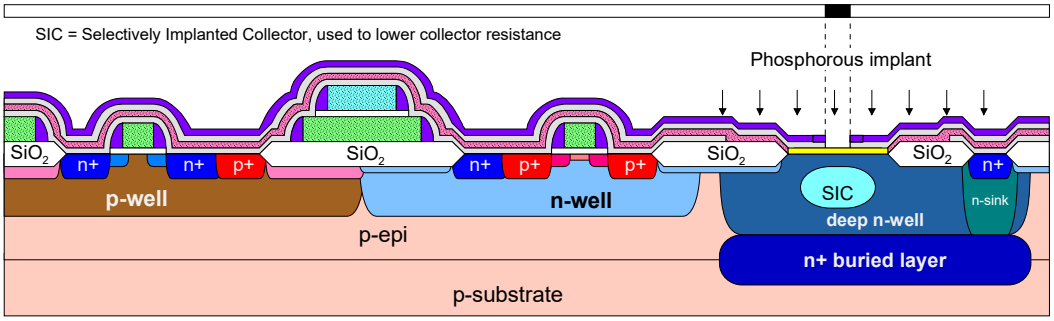
7. BE mask defines area to grow base epitaxial layer, TEOS removal



8. Blanket SiGe epitaxial growth, TEOS deposition, nitride deposition

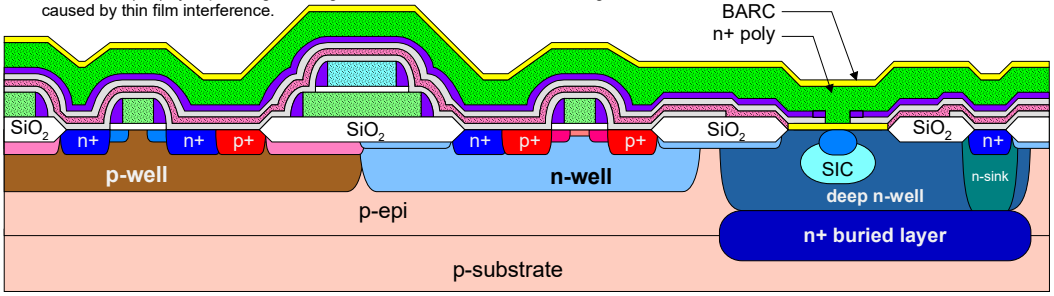


9. SIC-2 mask, SIC-2 implant



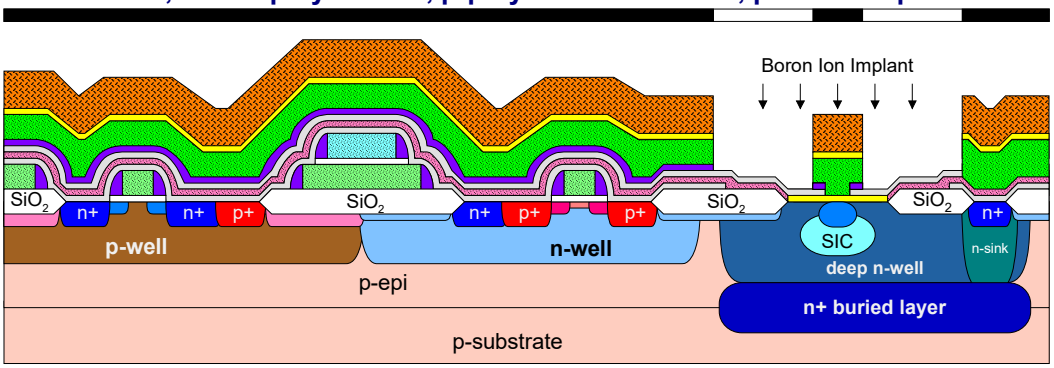
10. in-situ doped n+ emitter poly deposition, BARC coating

BARC = bottom antireflective coating, used to the enhance control of critical dimensions (CD) by supressing standing wave effects and reflective notching caused by thin film interference.

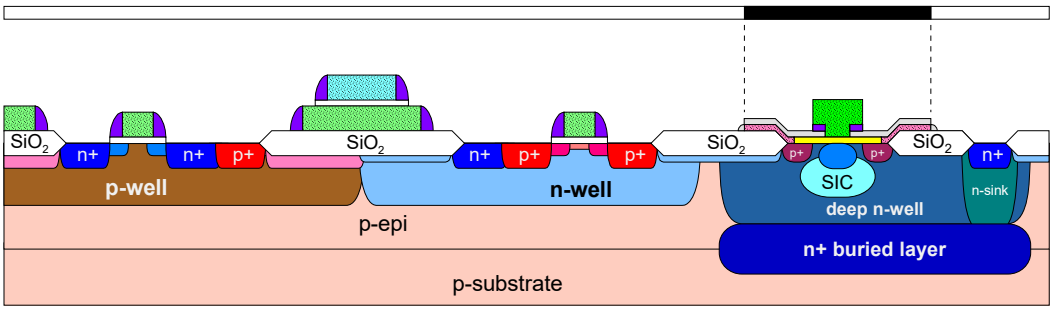


n-poly

11. PE mask, define poly emitter, p-poly and nitride etch, p+ base implants

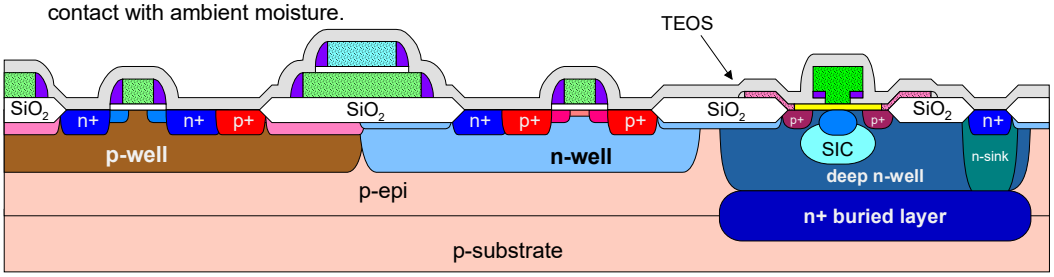


12. BP mask, n-poly etch (CMOS region), nitride etch, oxide-p-poly-oxide etch

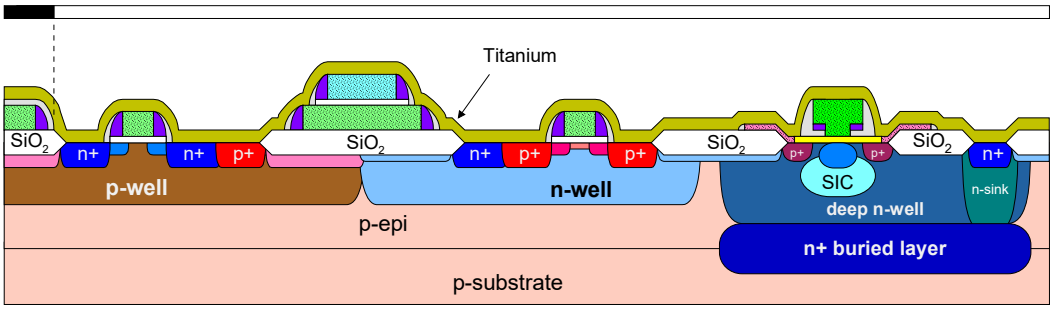


13. TEOS deposition for sidewall spacer

TEOS = Tetraethyl orthosilicate
It is a liquid at room temperature and will slowly hydrolyzes into silicon dioxide and ethanol when in contact with ambient moisture.

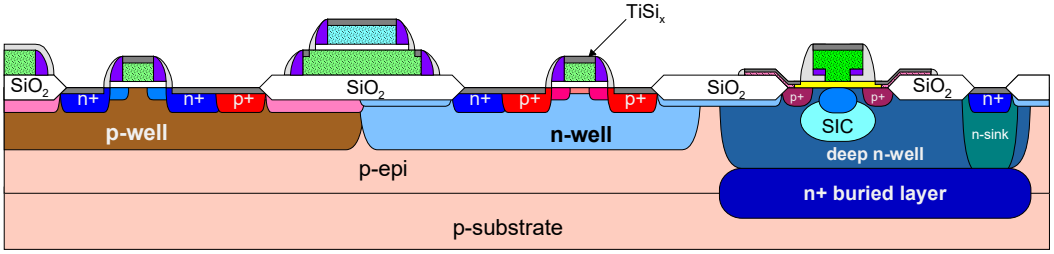


14. TEOS etch, Salicide Block Etch, Ti Sputtering

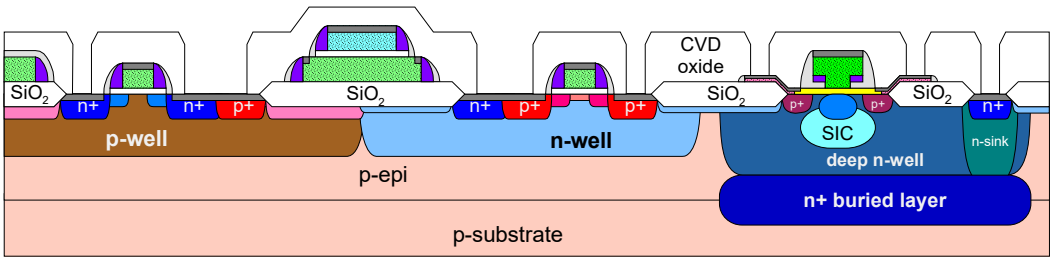


15. Salicide formation followed by removal of excess titanium.

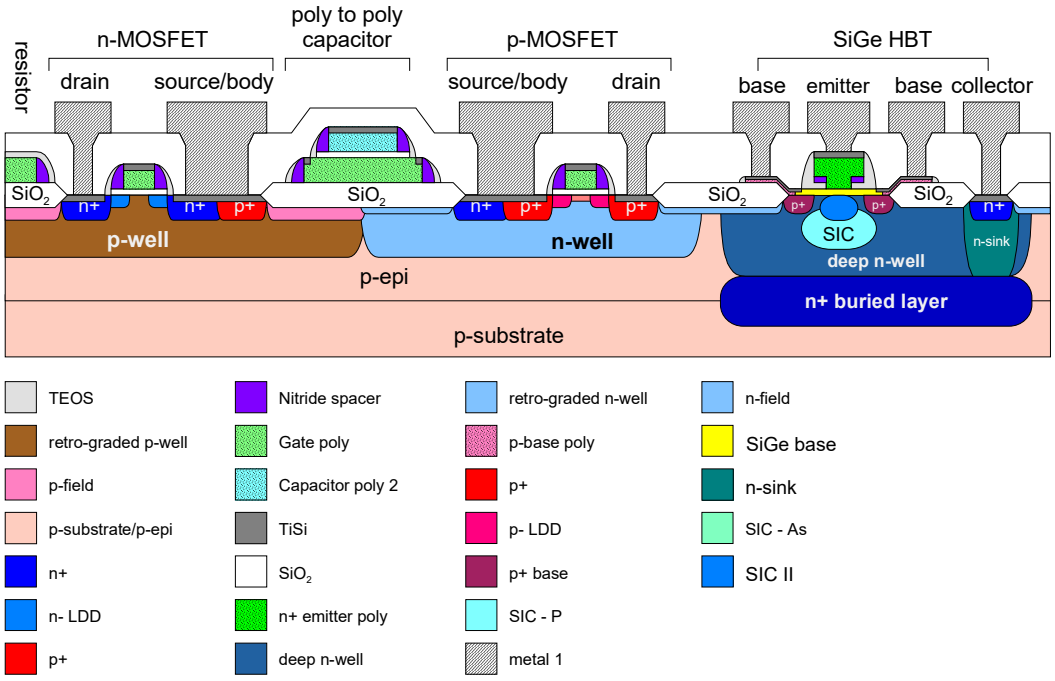
Titanium will only react with exposed silicon and polysilicon surface to form titanium silicide.



16. CVD oxide and contact mask

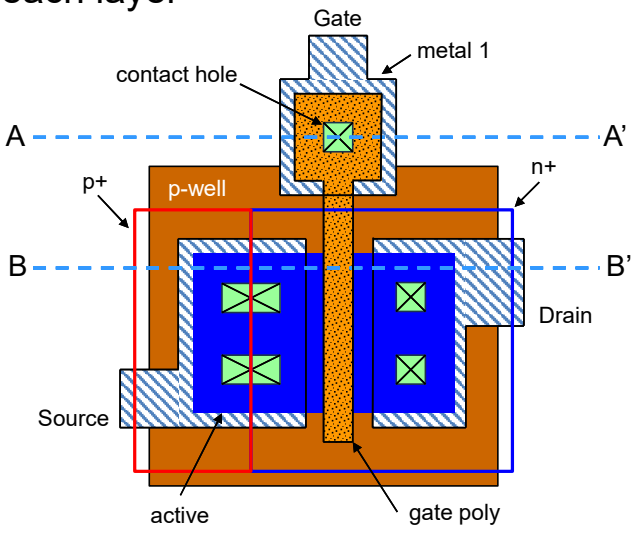


17. Metalization



Layout and Cross-Section of a MOSFET

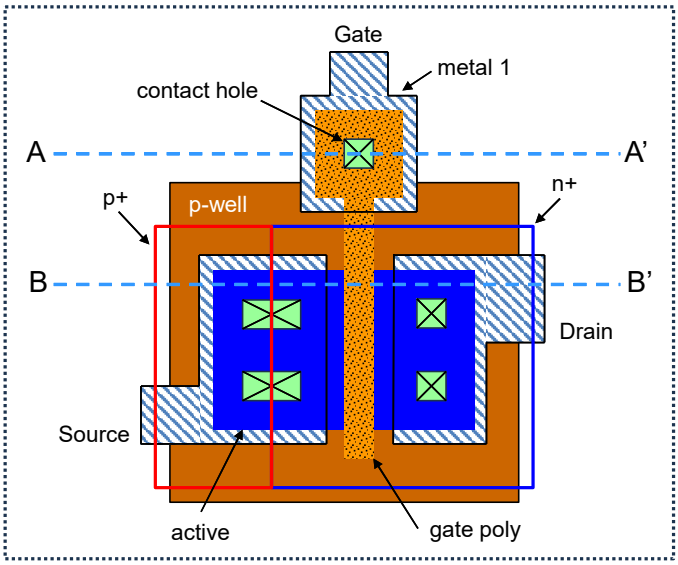
- First, you should recognize the sequence of the fabrication and the patterns of each layer

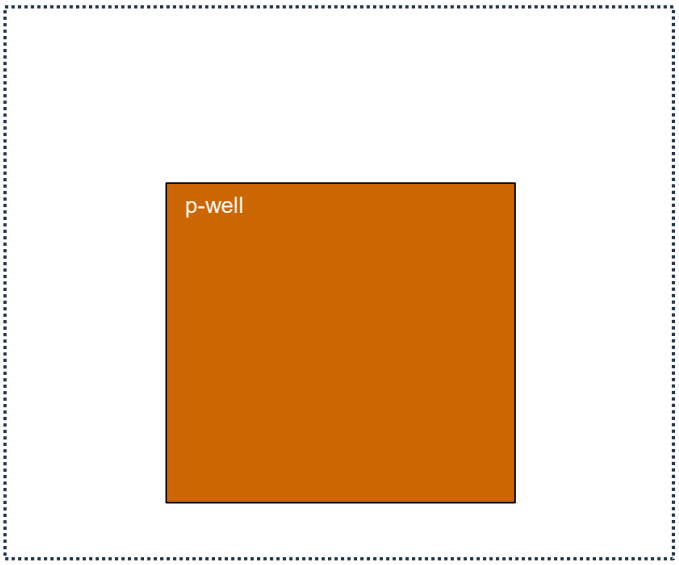


1. Starting wafer → p-substrate?
2. p-well
3. Active (Si_3N_4 mask for LOCOS process)
4. Gate oxide (no mask)
5. Gate Poly
6. n+ Source/Drain implant
7. p+ body contact implant
8. CVD oxide (no mask)
9. Contact holes
10. Metal 1

Layout and Cross-Section of a MOSFET

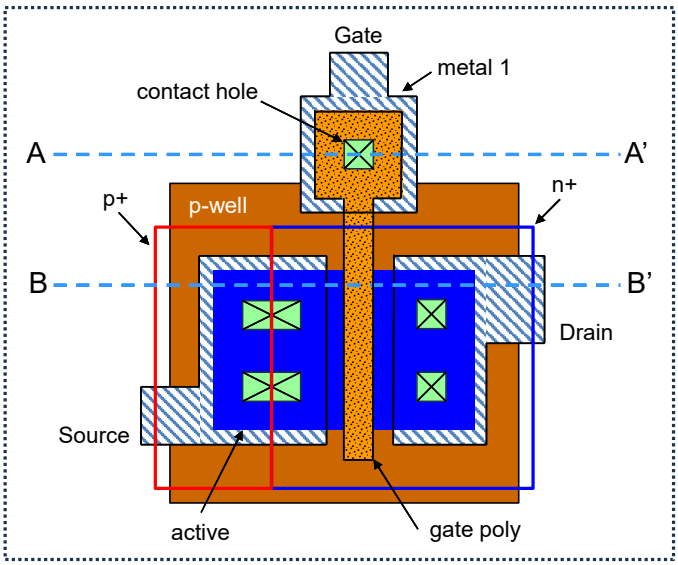
- First, you should recognize the sequence of the fabrication and the patterns of each layer

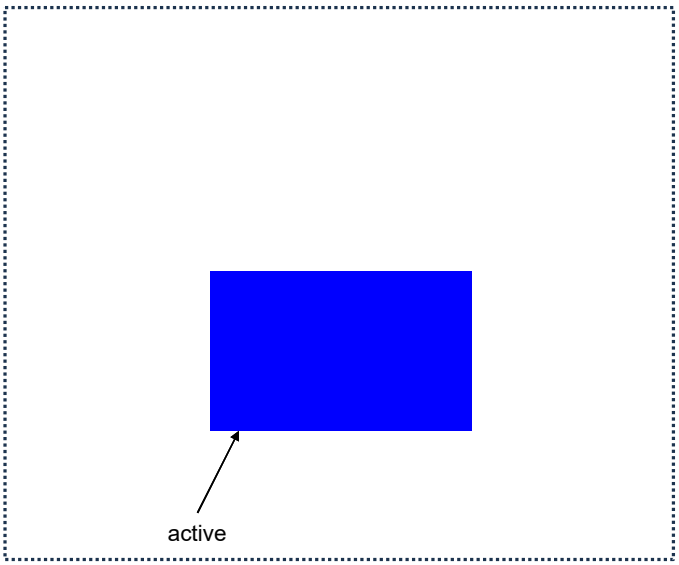




Layout and Cross-Section of a MOSFET

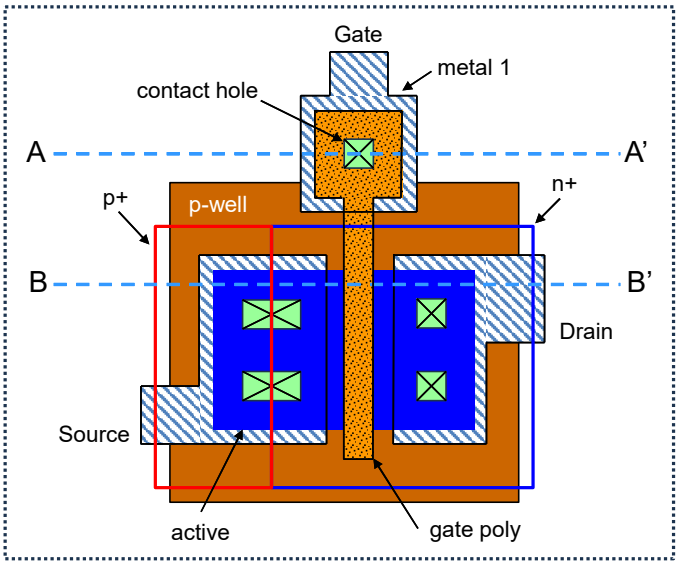
- First, you should recognize the sequence of the fabrication and the patterns of each layer

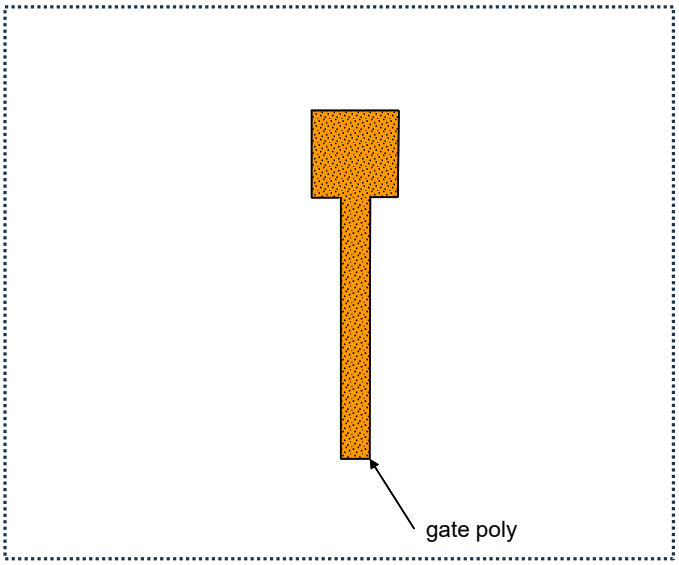




Layout and Cross-Section of a MOSFET

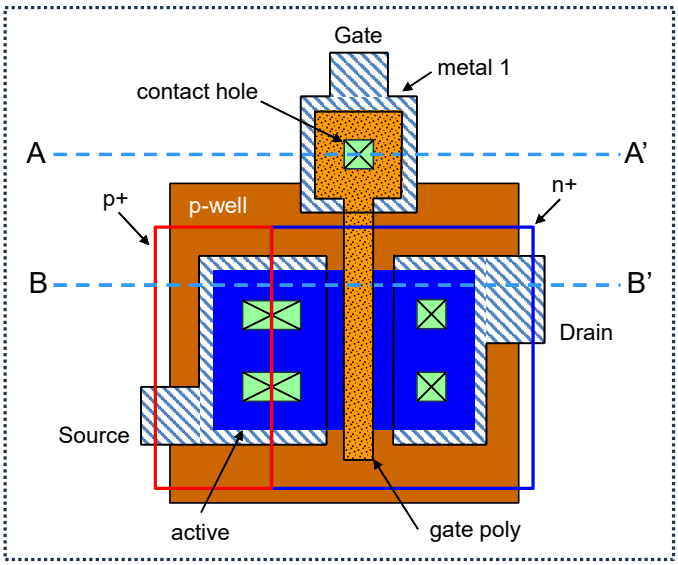
- First, you should recognize the sequence of the fabrication and the patterns of each layer

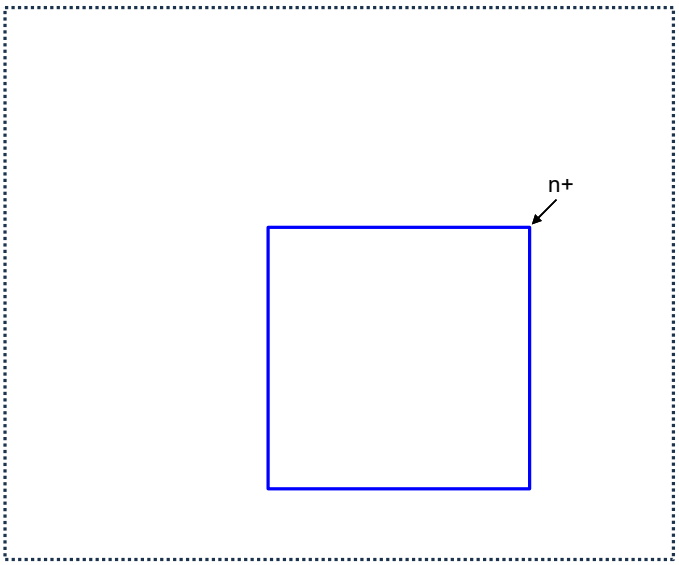




Layout and Cross-Section of a MOSFET

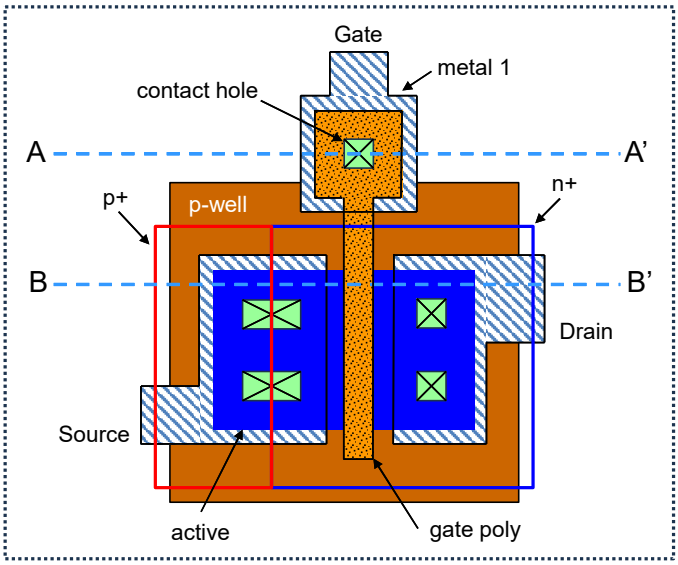
- First, you should recognize the sequence of the fabrication and the patterns of each layer

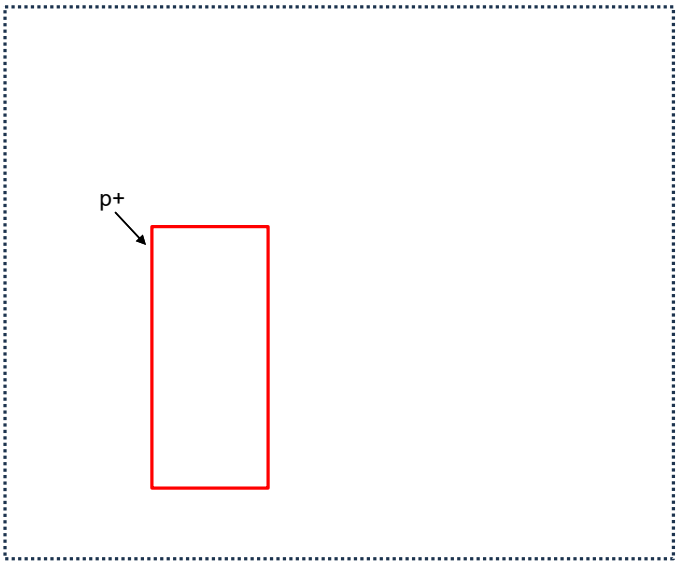




Layout and Cross-Section of a MOSFET

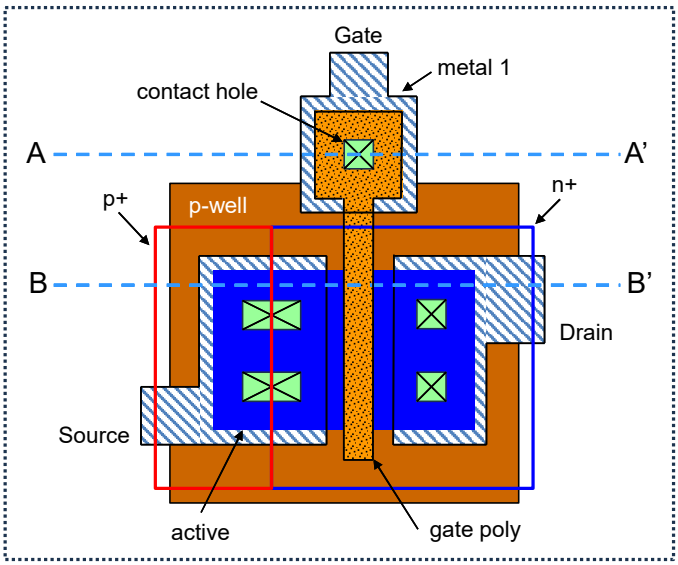
- First, you should recognize the sequence of the fabrication and the patterns of each layer

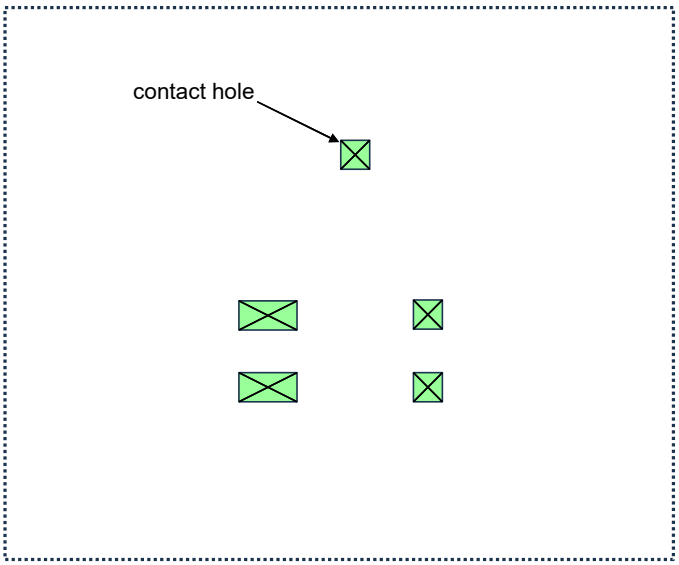




Layout and Cross-Section of a MOSFET

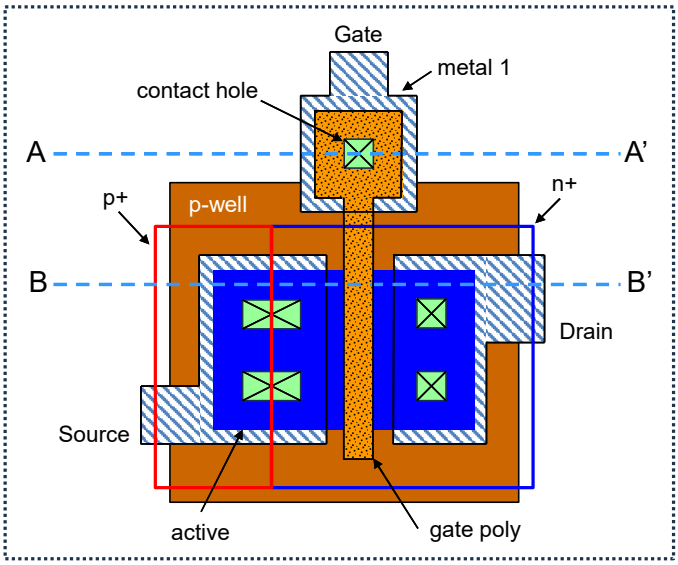
- First, you should recognize the sequence of the fabrication and the patterns of each layer

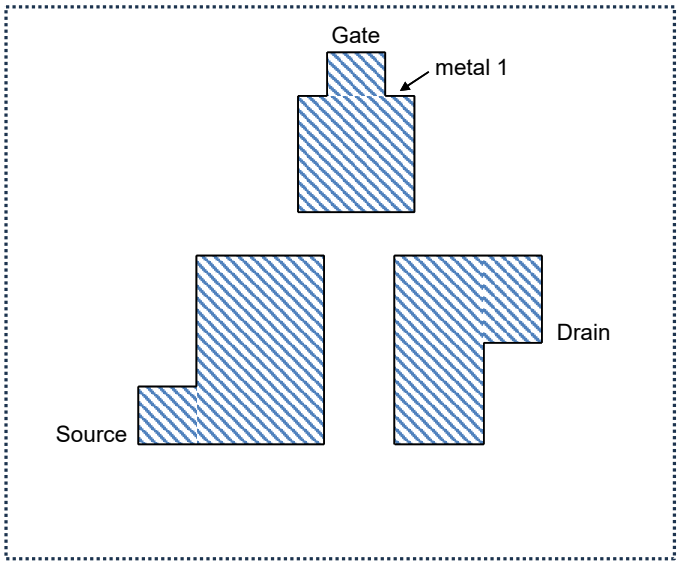




Layout and Cross-Section of a MOSFET

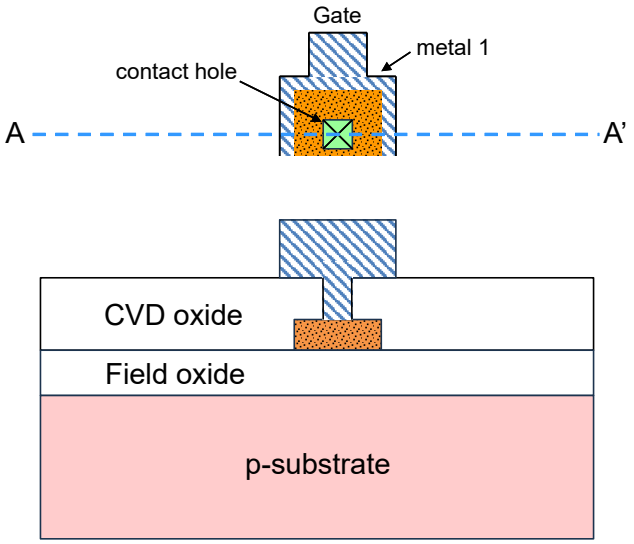
- First, you should recognize the sequence of the fabrication and the patterns of each layer

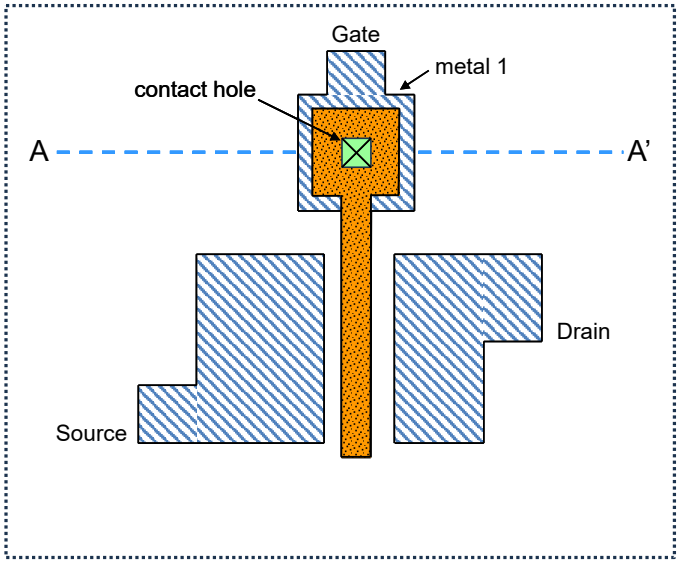




Layout and Cross-Section of a MOSFET

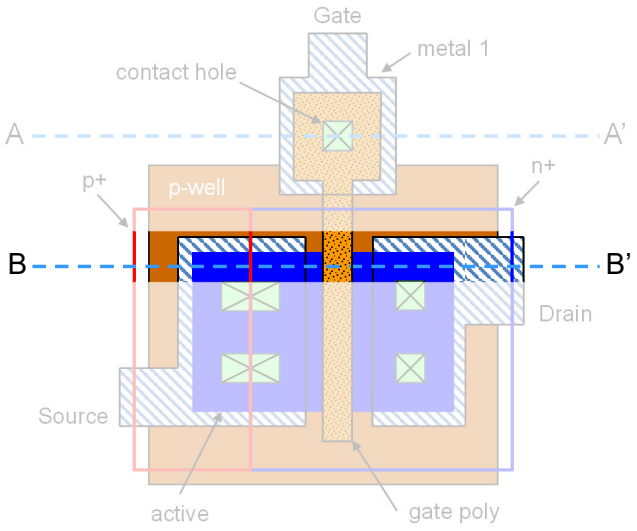
- Along AA', only consider the relevant layers

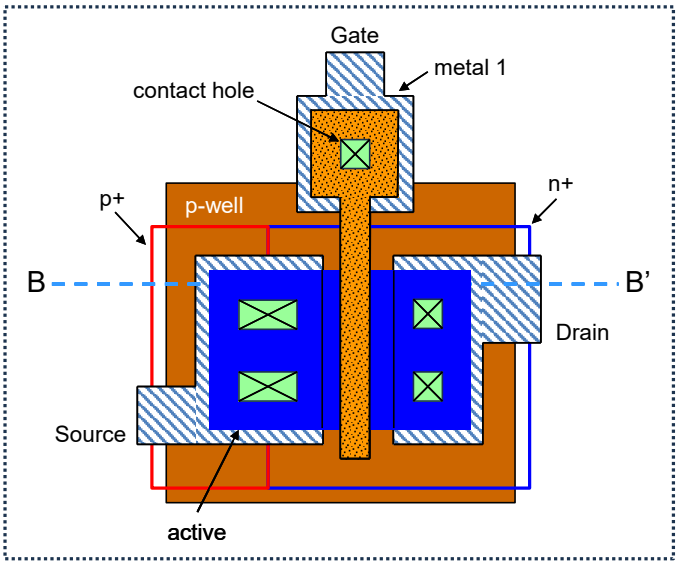




Layout and Cross-Section of a MOSFET

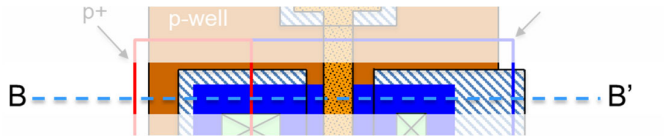
- Along BB', only consider the relevant layers

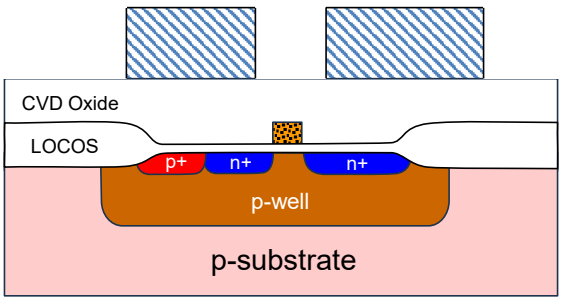


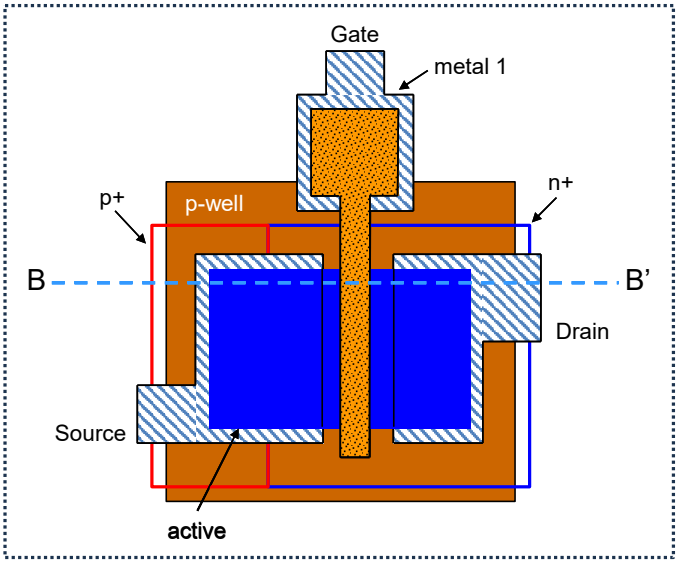


Layout and Cross-Section of a MOSFET

- Along BB', only consider the relevant layers

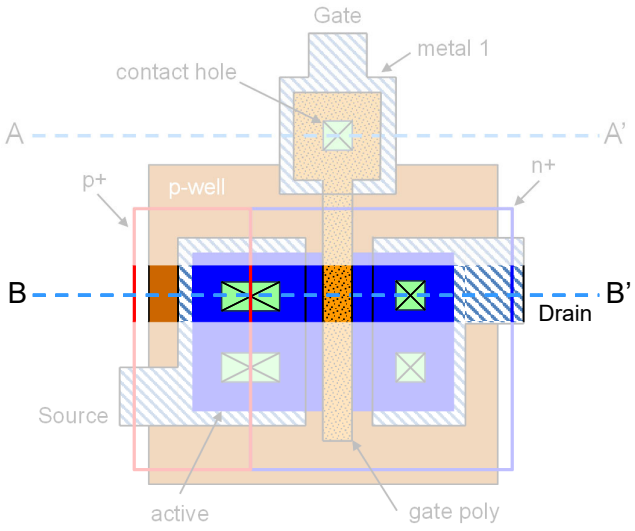


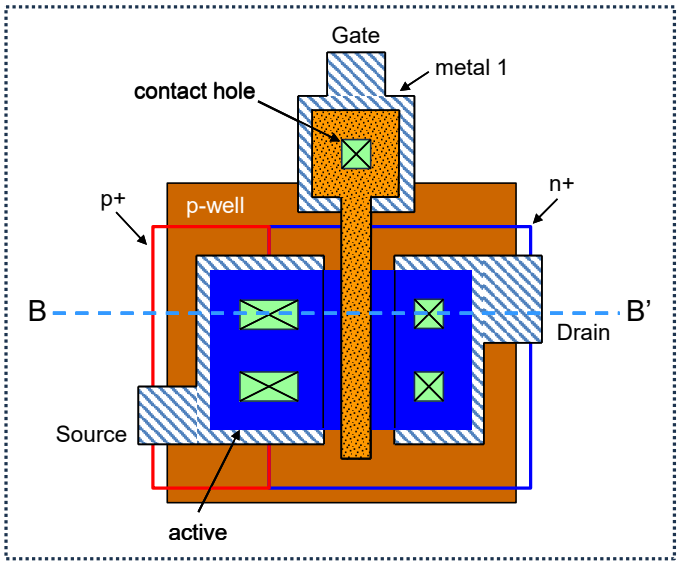




Layout and Cross-Section of a MOSFET

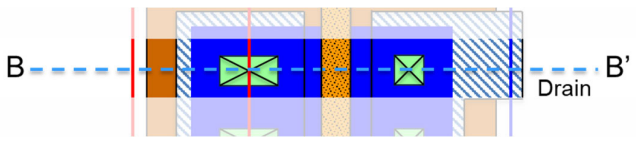
- Along BB', only consider the relevant layers

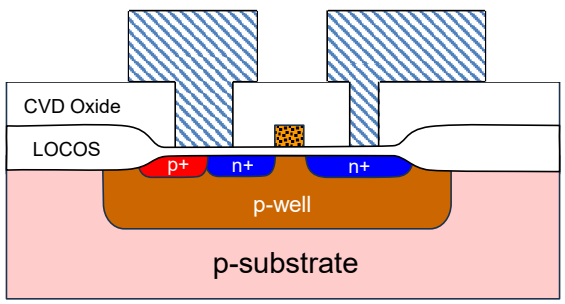


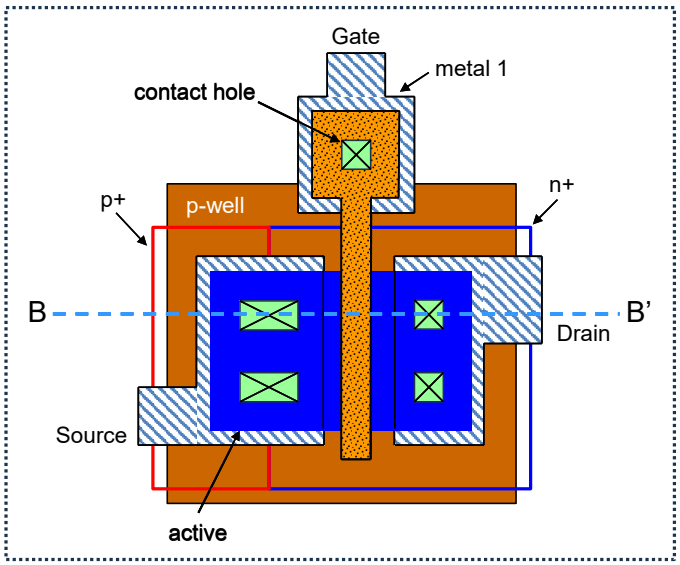


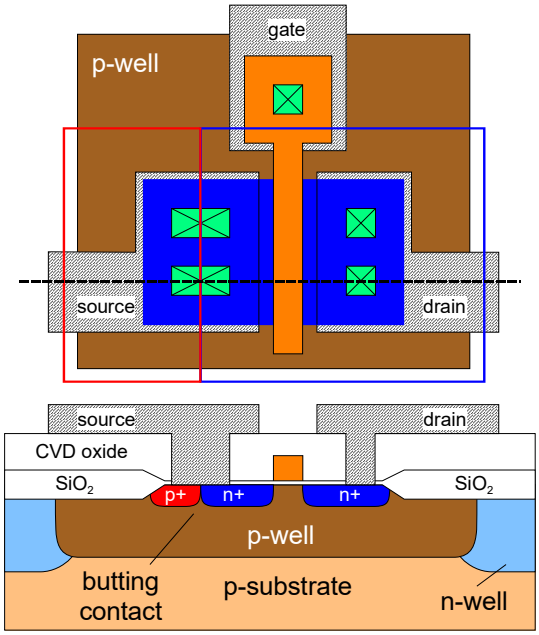
Layout and Cross-Section of a MOSFET

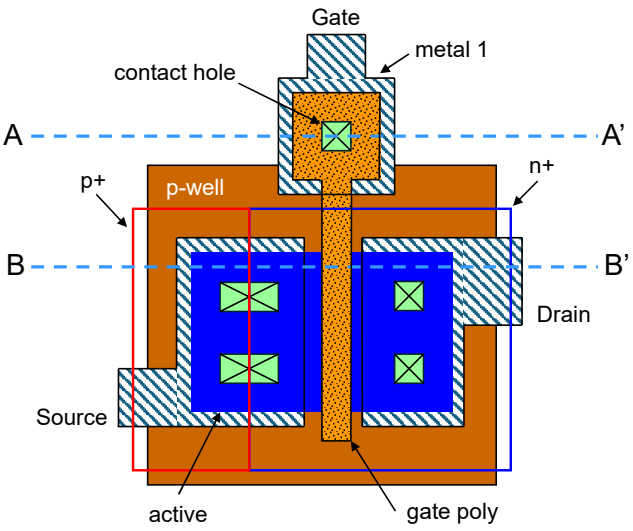
- Along BB', only consider the relevant layers





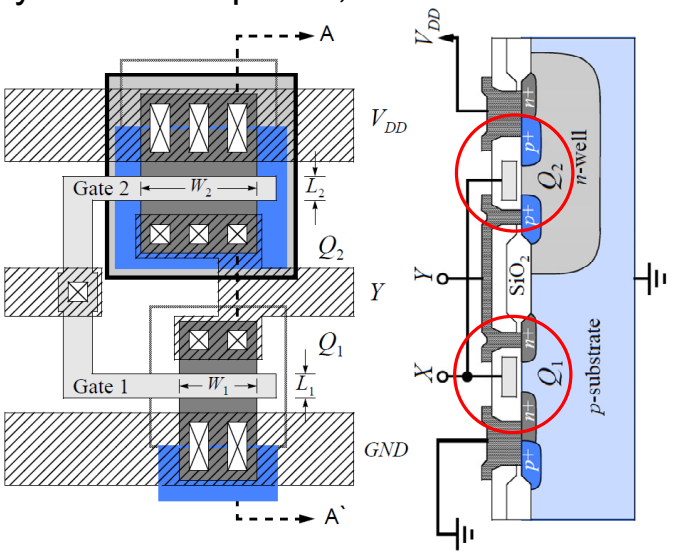



gate
p-well
source
drain
source
drain
CVD oxide
SiO2
SiO2
p+
n+
n+
p-well
butting
contact
p-substrate
n-well


Gate
metal 1
contact hole
A
A'
p+
p-well
n+
B
B'
Drain
Source
active
gate poly

Questions from Student(s)

From the cross section, we can tell the inverter uses two self-aligned MOSFETs. My explanation to this is that the gate polysilicon is covered by oxide. If the gate polysilicon is exposed, then it is non-self-aligned MOSFET. Is it correct?

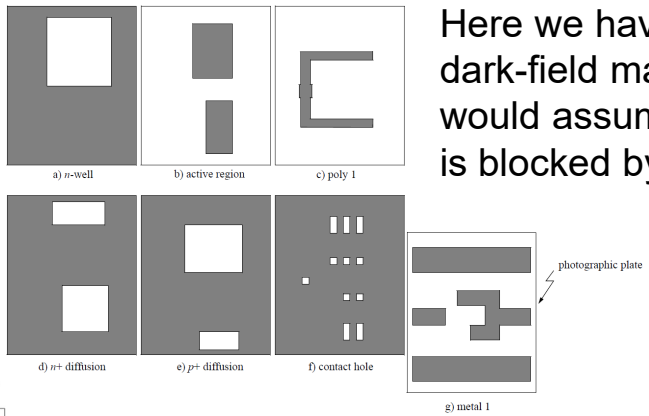


No, the process is either self-aligned, or not self-aligned. You cannot have some transistors that are one way, and some the other way.

Questions from Student(s)

If the cross section is not given and we have to draw it manually, how do we know if it is self-aligned or not?

You normally will be told what type of process you are working with. It is never the intension for the foundry to ask the users to guess.



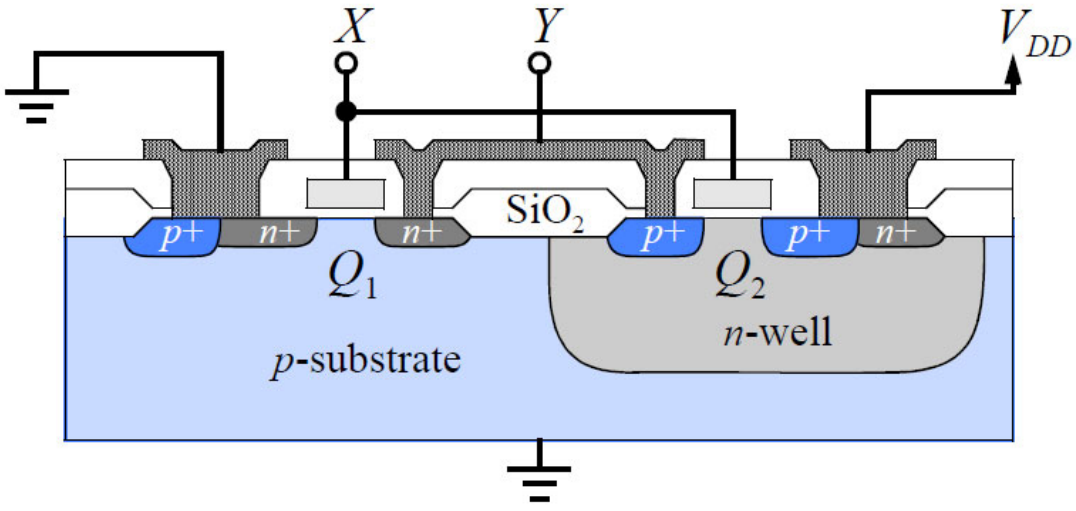
Here we have clear-field masks (one of them is b) and dark-field masks (one of them is a). Is that correct? I would assume light goes through the white areas and is blocked by the dark areas. Is that correct?

Yes

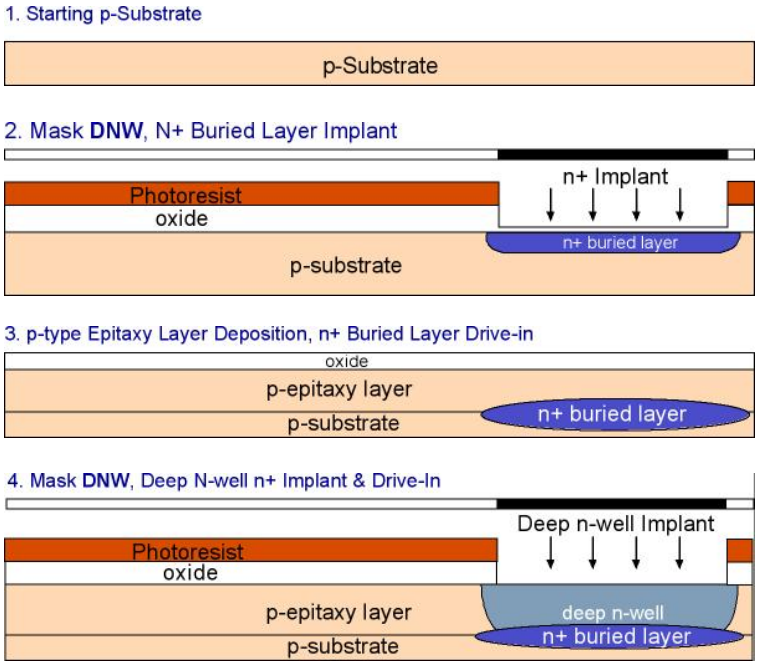
Questions from Student(s)

How do we select clear or dark field mask for each layer? For example, mask a is a dark field mask, b is a clear field mask (Or the opposite, if my memory is wrong), why do we pick them for their layer?

You do not have a choice. This is set by the foundry, unless you are the engineer developing this process. The choice of clear or dark field depends on how complicated are the patterns, how critical is the minimum feature size, etc.

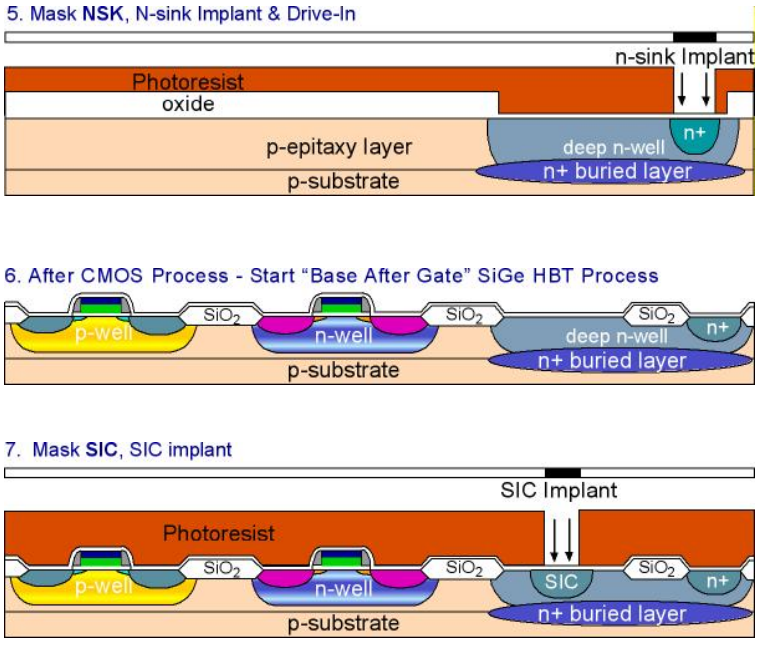


BiCMOS Fabrication Process Flow



- The BiCMOS process starts with a common substrate for both the CMOS and HBT devices.
- The first step is to implement the n+ buried layer that will be served to lower the collector resistance.
- This is followed by p-epi and deep p-well.

BiCMOS Fabrication Process Flow

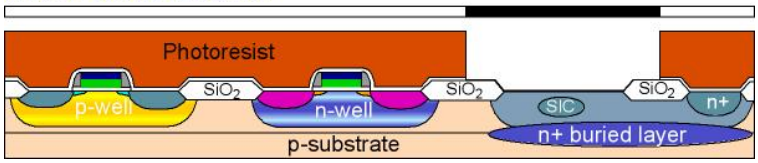


- The last step prior to standard CMOS process is the n+ sinker.
- Standard CMOS processing steps are carried out.
- During the subsequent processing, the CMOS region is untouched.

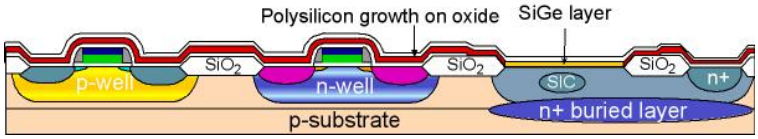
SIC = Selectively Implanted Collector

BiCMOS Fabrication Process Flow

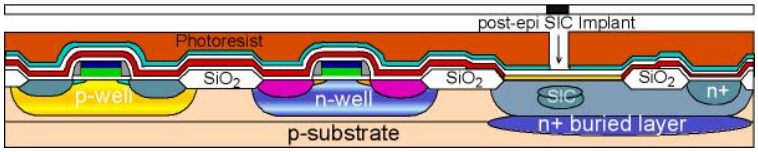
8. Mask **PBO**, TEOS Removal



9. SiGe Epitaxial Growth, TEOS Deposition



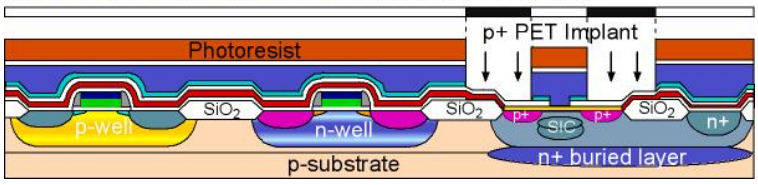
10. Poly-Deposition, **GST** Mask, Poly Etch, SIC Implant (post-Epitaxy)



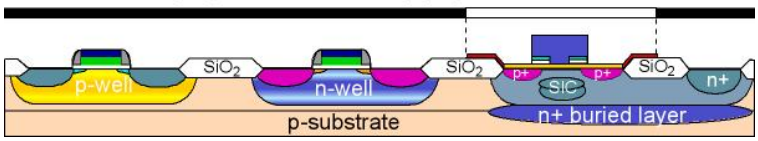
- Oxide removal prior to SiGe layer growth. Exposed area will have crystal line SiGe, polysilicon elsewhere.
- Open window for additional SIC implant to further lower the collector resistance.

BiCMOS Fabrication Process Flow

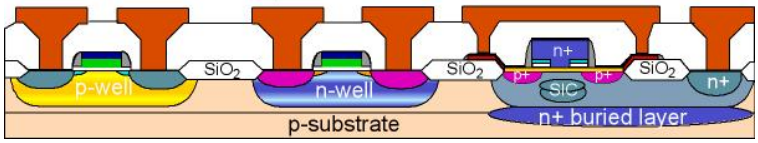
11. IPDP Deposition, TEOS, **PET** Mask, Poly Etch, p+ PET Implant



12. Mask **PBP**, n-poly Etch, Oxide Etch, p-poly Etch, TEOS Removal



14. Contact Etch, Metal Deposition and Passivation



- Deposit and pattern emitter polysilicon.
- Use Poly emitter as a mask and perform p+ implant for base contact.
- Pattern the p-base poly (grown during the SiGe epi).
- Final steps include contact etch, and metallization.