MANIPAL
ACADEMY *of* HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

MANIPAL SCHOOL OF INFORMATION SCIENCES
(A Constituent unit of MAHE, Manipal)

DSVD assignment-1

Reg. Number	Name	Branch
241038005	Chaitanya S Rao	VLSI Design

Submitted to,

Prashanth Kumar Shetty

Professor,

Manipal School of Information Sciences,

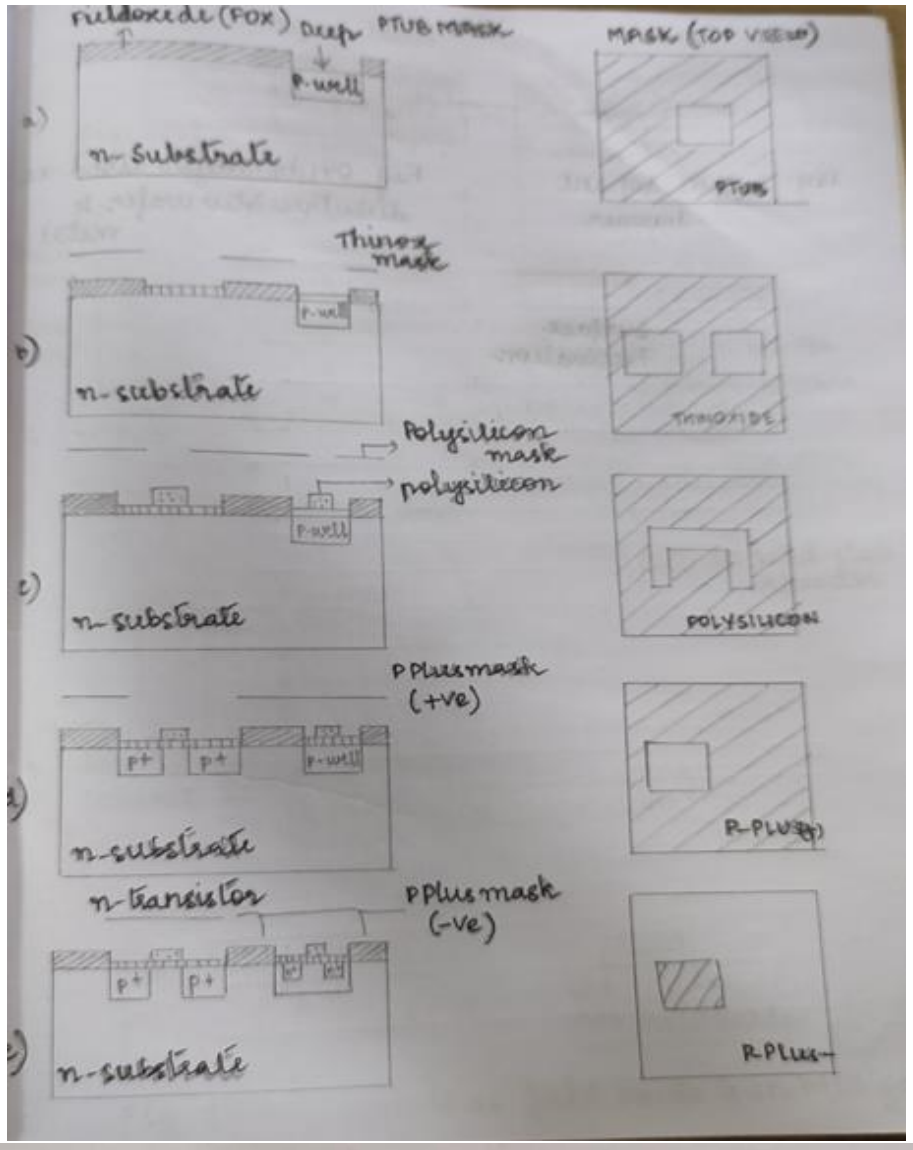
MAHE, MANIPAL

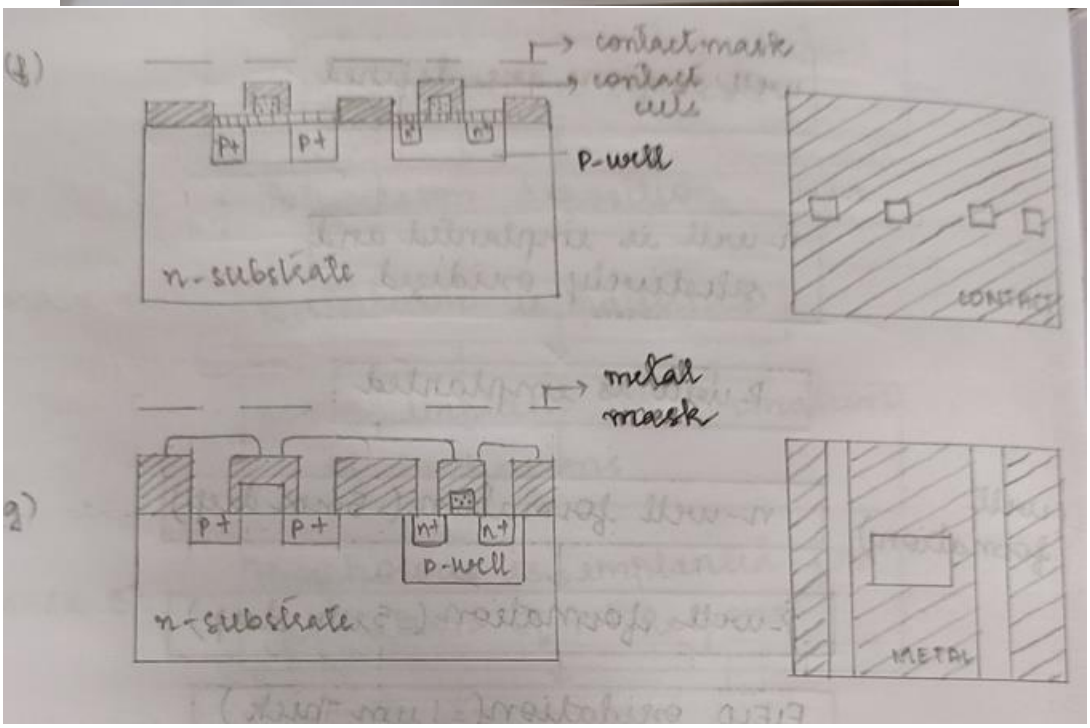
10/10/2024

P- well Process

A common approach to p-well CMOS fabrication has been to start with a moderately doped n-type substrate (wafer), create the p-type well for the n-channel devices, and build the fi-channel transistor in the native n-substrate.

- The first mask defines the p-well (or p-tub); n-channel transistors will be fabricated in this well. Field oxide (FOX) is etched away to allow a deep diffusion.
- The next mask is called the "thin oxide" or "thinox" mask, as it defines where areas of thin oxide are needed to implement transistor gates and allow implantation to form p- or n-type diffusions for transistor source/drain regions. The field oxide areas are etched to the silicon surface and then the thin oxide is grown on these areas. Other terms for this mask include active area, island, and mesa. In nMOS this would be the diffusion mask.
- Polysilicon gate definition is then completed. This involves covering the surface with polysilicon and then etching the required pattern (in this case an inverted "U"). The "poly" gate regions lead to "self-aligned" source-drain regions.
- A p-plus (p+) mask is then used to indicate those thin-oxide areas (and polysilicon) that are to be implanted p+. Hence a thin-oxide area exposed by the p-plus mask will become a p+ diffusion area. If the p-plus area is in the n-substrate, then a p-channel transistor or p-type wire may be constructed. If the p-plus area is in the p-well, then an ohmic contact to the p-well may be constructed. An ohmic contact is one which is only resistive in nature and is not rectifying (as in the case of a diode). In other words, there is no junction (n-type and p-type silicon abutting). Current can flow in both directions in an ohmic contact. This type of mask is sometimes called the select mask as it selects those transistor regions that are to be p-type.
- The next step usually uses the complement of the p-plus mask, although an extra mask is normally not needed. The "absence" of a p-plus region over a thin-oxide area indicates that the area will be an n+ diffusion or n-thinox. n-thinox in the p-well defines possible n-transistors and wires. An n+ diffusion in the n-substrate allows an ohmic contact to be made. Following this step, the surface of the chip is covered with a layer of SiO2.
- Contact cuts are then defined. This involves etching any SiO2, down to the contacted surface. These allow metal (next step) to contact diffusion regions or polysilicon regions.
- Metallization is then applied to the surface and selectively etched.
- As a final step (not shown), the wafer is passivated and openings to the bond pads are etched to allow for wire bonding. Passivation protects the silicon surface against the ingress of contaminants that can modify circuit behavior in deleterious ways.


Fieldoxide (FOX)
Deep
PTUB MASK
MASK (TOP VIEW)
P-well
n- substrate
PTUB
Thinox mask
THINOXIDE
Polysilicon mask
polysilicon
POLYSILICON
P Plus mask (+ve)
P+
P-PLUS(+)
n-transistor
P Plus mask (-ve)
R-Plus
a)
b)
c)
d)
e)


contact mask
contact cuts
P-well
n- substrate
CONTACT
metal mask
METAL
f)
g)

****

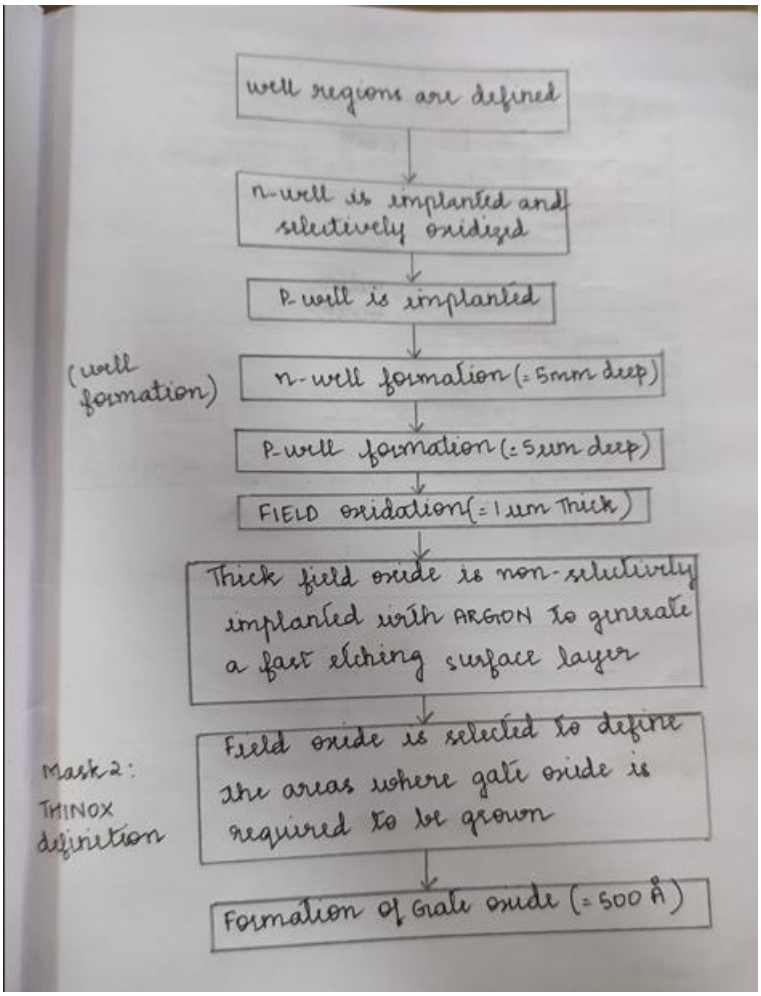
Twin Tube Process

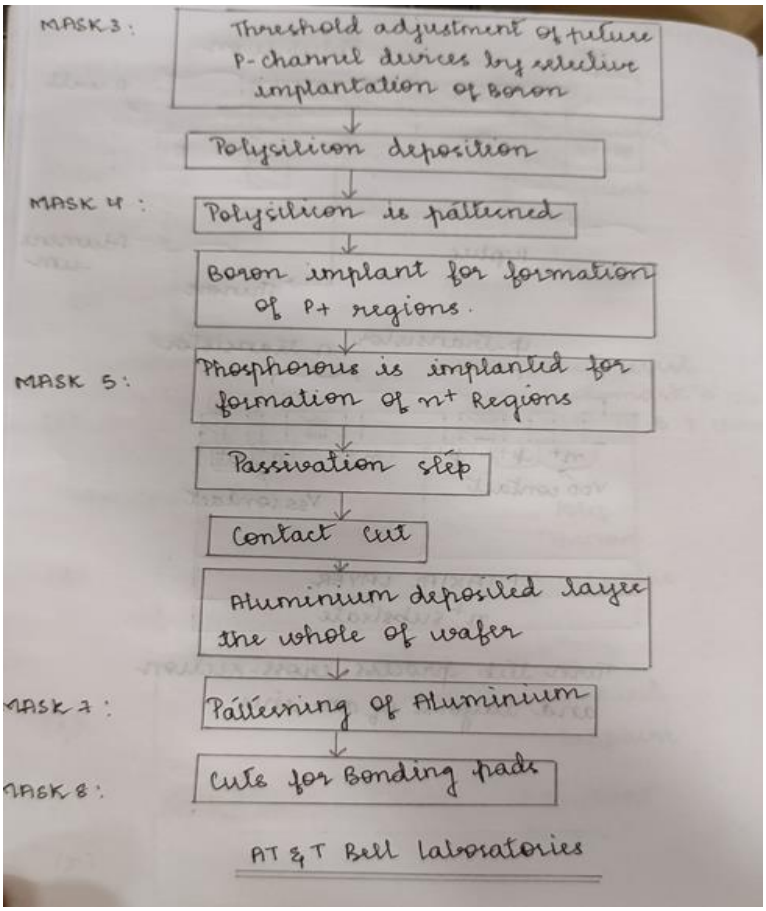
Twin-tub CMOS technology provides the basis for separate optimization of the p-type and n-type transistors, thus making it possible for threshold voltage, body effect, and the gain associated with n-and p-devices to be independently optimized [Parr80]. Generally, the starting material is either an n+ or p+ substrate with a lightly doped epitaxial or epi layer, which is used for protection against latch-up. The aim of epitaxy (which means "arranged upon") is to grow high purity silicon layers of controlled thickness with accurately determined dopant concentrations distributed homogeneously throughout the layer. The electrical properties for this laver are determined by the dopant and its concentration in the silicon.

The process sequence, which is similar to the p-well process apart from the tub formation where both p-well and n-well are utilized, entails the following steps:

- Tub formation
- Thin oxide etching
- Source and drain implantations
- Contact cut definition
- Metallization.

Figure illustrates the steps involved in the AT&T-Bell Laboratories twin-tub process. Since this process provides separately optimized wells, better performance n-transistors (lower capacitance, less body effect) may be constructed when compared with a conventional p-well process. Similarly, the p-transistors may be optimized. The use of threshold adjust steps is included in this process. These masks are derived from the thin-ox and n-plus masks.


well regions are defined
n-well is implanted and selectively oxidized
P-well is implanted
(well formation)
n-well formation (= 5mm deep)
P-well formation (= 5um deep)
FIELD oxidation (= 1 um Thick)
Thick field oxide is non-selectively implanted with ARGON to generate a fast etching surface layer
Mask 2: THINOX definition
Field oxide is selected to define the areas where gate oxide is required to be grown
Formation of Gate oxide (= 500 Å)


MASK 3:
Threshold adjustment of future P-channel devices by selective implantation of Boron
Polysilicon deposition
MASK 4:
Polysilicon is patterned
Boron implant for formation of P+ regions.
MASK 5:
Phosphorous is implanted for formation of n+ Regions
Passivation step
Contact cut
Aluminium deposited layer the whole of wafer
MASK 7:
Patterning of Aluminium
MASK 8:
Cuts for Bonding pads
AT & T Bell Laboratories

****

Silicon on Insulator

Silicon on insulator (SOI) CMOS processes have several potential advantages over the traditional CMOS technologies (MaSi64). These include higher density, no latch-up problems, and lower parasitic capacitances. In the SOI process a thin layer of single crystal silicon film is epitaxially grown on an insulator such as sapphire or magnesium aluminate spinel. Various masking and doping techniques are then used to form p-channel and n-channel devices. Unlike the more conventional CMOS approaches, the extra steps in well formation do not exist in this technology.

The steps used in typical SOI CMOS processes are:

• A thin film (7-8 um) of very lightly-doped n-type Si is grown over an insulator. Sapphire is a commonly used insulator

• An anisotropic etch is used to etch away the Si except where a diffusion area (n or p) will be needed. The etch must be anisotropic since the thickness of the Si is much greater than the spacings desired between the Si "islands".

•The p-islands are formed next by masking he n-islands with a photoresist. A p-type dopant, boron, for example, is then implanted. It is masked by the photoresist, but forms p-islands at the unmasked islands. The p-islands will become the n-channel devices.

•The p-islands are then covered with a photoresist and an n-type dopant, phosphorus, for example, is implanted to form the n-islands. The n-islands will become the p-channel devices.

• A thin gate oxide (around 500-600 A) is grown over all of the Si structures. This is normally done by thermal oxidation.

• A polysilicon film is deposited over the oxide. Often the po-lysilicon is doped with phosphorus to reduce its resistivity.

the p-islands by the polysilicon. After this step the n-channel devices are complete.

• The p-channel devices are formed next by masking the p-islands and implanting a p-type dopant such as boron. The polysilicon over the gate of the n-islands will block the. dopant from the gate, thus forming the p-channel devices .

•A layer of phosphorus glass or some other insulator such as silicon dioxide is then deposited over the entire structure. The glass is etched at contact cut locations. The metallization layer is formed next by evaporating aluminum over the entire surface and etching it to leave only the desired metal wires. The aluminum will flow through the contact cuts to make contact with the diffusion or polysilicon regions.

•A final passivation layer of a phosphorus glass is deposited and etched over bonding pad locations.

Because the diffusion regions extend down to the insulating substrate, only "sidewall" areas associated with source and drain diffusions contribute to the parasitic junction capacitance. Since sapphire is an extremely good insulator, leakage currents between transistors and substrate and adjacent devices are almost eliminated.

In order to improve the yield, some processes use "preferential etch" in which the island edges are tapered. Thus, aluminum or poly runners can enter and leave the islands with a minimum step height. This is contrasted to "fully anisotropic etch" in which the undercut is brought to zero. An "isotropic etch" is also shown in the same diagram for comparison. The advantages of SOI technology are:

•Due to the absence of wells, denser structures than bulk silicon can be obtained. Also direct n to p connections may be made.

•Low capacitances provide the basis of very fast circuits.

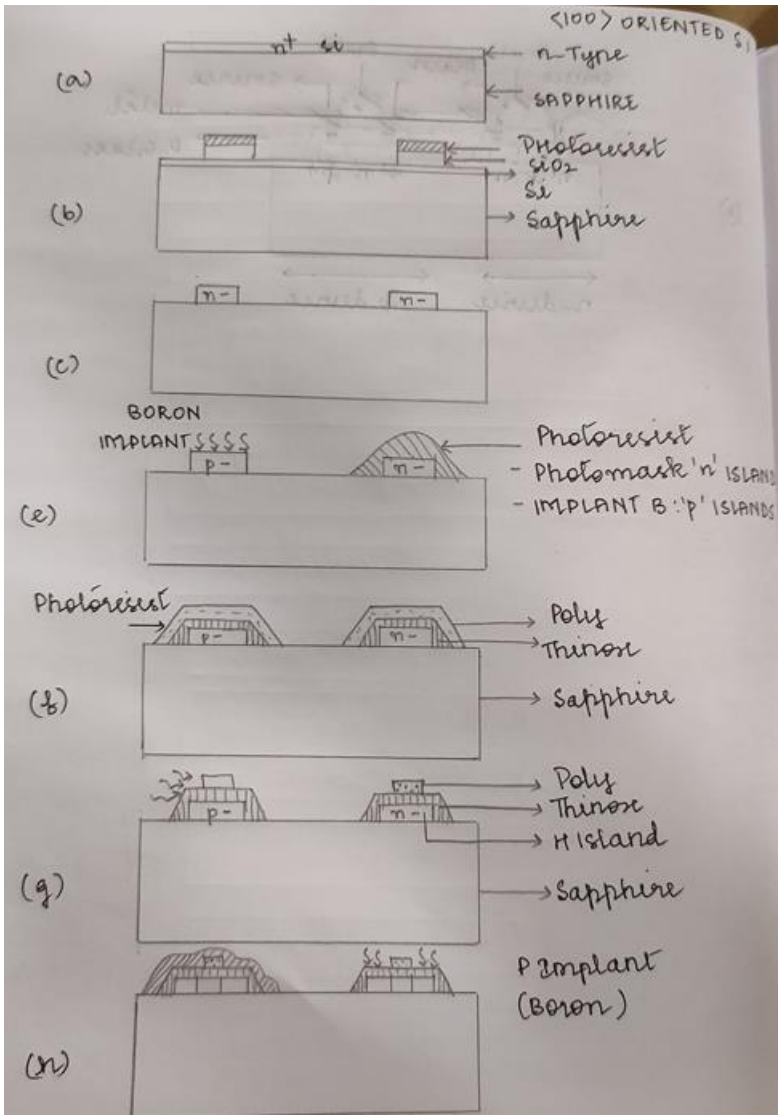
•No field-inversion problems exist (insulating substrate).

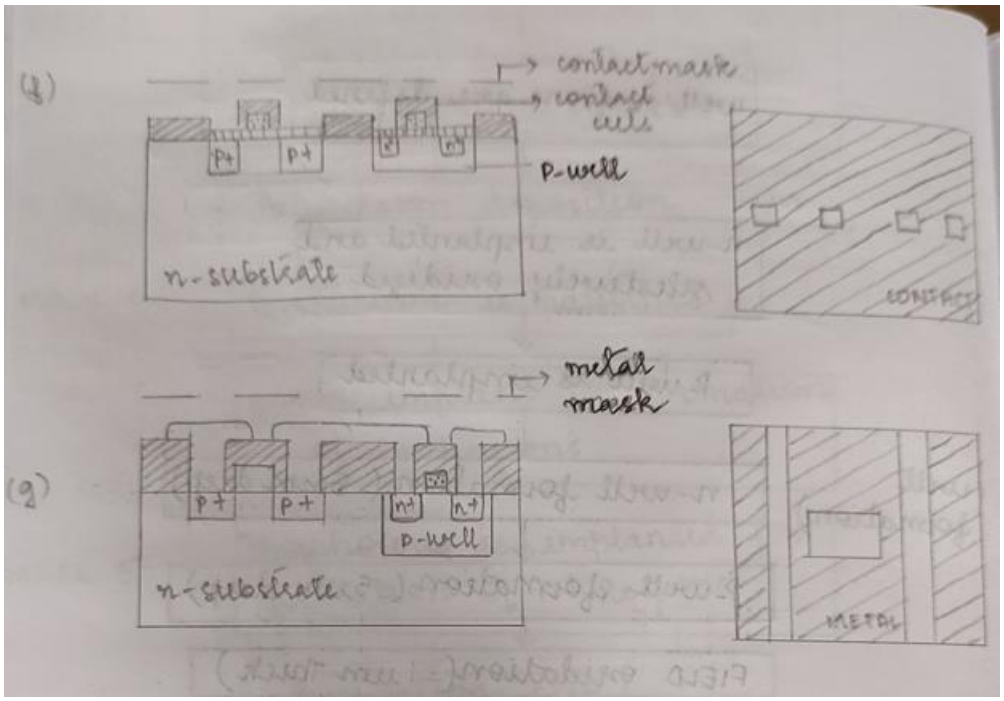
•No latch-up due to isolation of n- and p-transistors by insulating substrate.

•As there is no conducting substrate, there are no body effect problems.

•Enhanced radiation tolerance.

However, on the negative side, due to absence of substrate diodes. the inputs are somewhat more difficult to protect. As device gains are lower, 1/0 structures have to be larger. Single crystal sapphire or spinel substrates are considerably more expensive than silicon and processing techniques tend to be less developed than bulk silicon techniques. Thus, although SOI has the potential to be the fastest CMOS technology, it is also the most expensive.


<100> ORIENTED Si
(a)
n+ si
n-Type
SAPPHIRE
(b)
Photoresist
SiO2
Si
Sapphire
(c)
n-
n-
BORON IMPLANT
(e)
p-
n-
Photoresist
- Photomask 'n' ISLANDS
- IMPLANT B : 'p' ISLANDS
Photoresist
(g)
p-
n-
Poly
Thinox
Sapphire
(g)
p-
n-
Poly
Thinox
H Island
Sapphire
(h)
P Implant
(Boron)

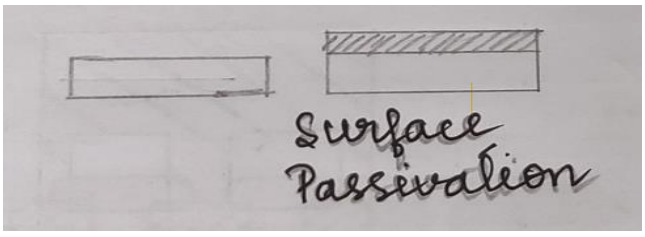

(f)
contact mask
contact cuts
p+
p+
n+
n+
p-well
n-substrate
CONTACT
(g)
metal mask
p+
p+
n+
n+
p-well
n-substrate
METAL

****

4) what are the uses of Silicon Dioxide layer?

Of all the advantages of silicon for the formation of semiconductor devices, the ease of growing a silicon dioxide layer is perhaps the most useful. Whenever a silicon surface is exposed to oxygen, it is converted to silicon dioxide. Silicon dioxide is composed of one silicon atom and two oxygen atoms (SiO_2). We encounter silicon dioxide daily. It is the chemical composition of ordinary window glass. It's semiconductor version, however, is purer and formed in a specific way. Silicon dioxide layers are formed on bare silicon surfaces at elevated temperatures in the presence of an oxidant. The process is called thermal oxidation.
Although silicon is a semiconductor, silicon dioxide is a dielectric material. This combination, a dielectric layer formed on a semiconductor, along with other properties of silicon dioxide, makes it one of the most commonly used layers in silicon devices. Silicon dioxide layers find use in devices to pacify the silicon surface, to act as doping barriers and surface dielectrics, and to serve as dielectric parts of device structures.

Surface passivation:
The extreme sensitivity of semiconductor devices to contamination was examined. While a major focus of a semiconductor facility is the control and elimination of contamination, the techniques are not always 100 percent effective. Silicon dioxide layers play an important role in protecting semiconductor devices from contamination.
Silicon dioxide performs this role in two ways. First is the physical protection of the surface and underlying devices. Silicon dioxide layers formed on the surface are very dense (nonporous) and very hard. Thus, a silicon dioxide layer acts as a contamination barrier by physically preventing dirt in the processing environment from getting to the sensitive wafer surface. The hardness of the layer protects the wafer surface from scratches and abuse endured by the wafer in the fabrication processes.



The second way silicon dioxide protects devices is chemical in nature. Regardless of the cleanliness of the processing environment, some electrically active contaminants (mobile ionic contaminants) end up in or on the wafer surface. During the oxidation process, the top layer of silicon is converted to silicon dioxide.
Contaminants on the surface end up in the new layer of oxide, away from the electrically active surface. Other contaminants are drawn up into the silicon dioxide film where they are less harmful to the devices. In the early days of MOS device processing, it was common to

oxidize the wafers and then remove the oxide before further processing to rid the surface of unwanted mobile ionic contamination.

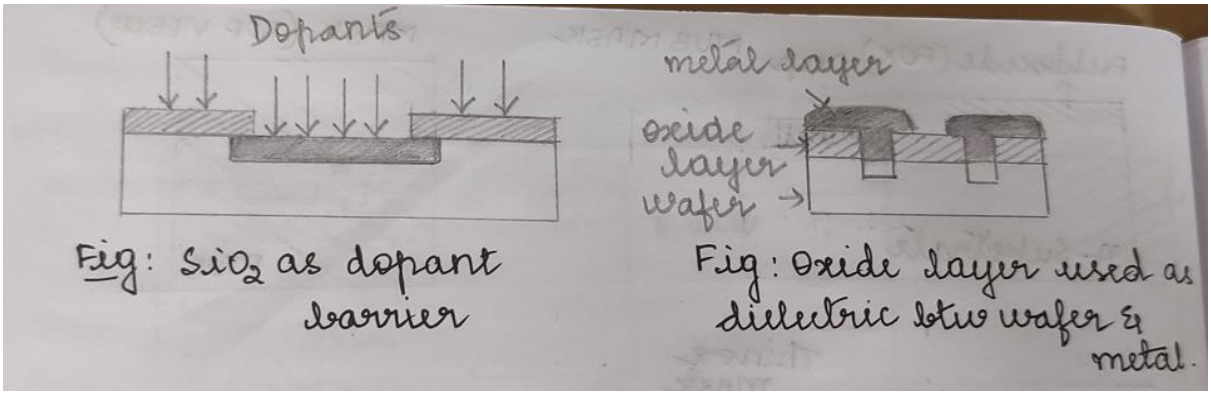
Doping barrier

We know that doping was identified as one of the four basic fabrication operations. Doping requires creating holes in a surface layer through which specific dopants are introduced into the exposed wafer surface through diffusion or ion implantation. In silicon technology, the surface layer is most often silicon dioxide. The silicon dioxide left on the wafer acts to block the dopant from reaching the silicon surface. All of the dopants used in silicon technology have a very slow rate of movement in silicon dioxide as compared to silicon. While the dopants penetrate to the required depth in the exposed silicon, they penetrate only a short distance into the silicon dioxide surface. It takes only a relatively thin silicon dioxide layer to block the dopants from reaching the silicon surface.

Another factor favoring the use of silicon dioxide is a coefficient of thermal expansion similar to that of silicon. In the high-temperature processes of oxidation, diffusion doping, and others, the wafer expands and contracts as it is heated and cooled. The silicon dioxide expands and contracts at close to the same rate as silicon, which means that the wafer will not warp during the heating and cooling.

Surface dielectric

Silicon dioxide is classified as a dielectric. This means that, under normal circumstances, it does not conduct electricity. When dielectrics are used in electrical circuits or devices, they are referred to as *insulators*. A very important role of silicon dioxide layer is to act like an insulator. Figure shows a cross section of a wafer with a conductive layer of metal on top of a layer of silicon dioxide. The oxide prevents shorting of the metal layer to the underlying metal just as the insulation on an



electric cord prevents the wires from shorting. In this capacity, the oxide must be continuous; that is, have no holes or voids.

The oxide must also be thick enough to prevent a phenomenon known as *induction*. Induction will only occur when the separating layer of oxide is very thin so that the electrical charge in

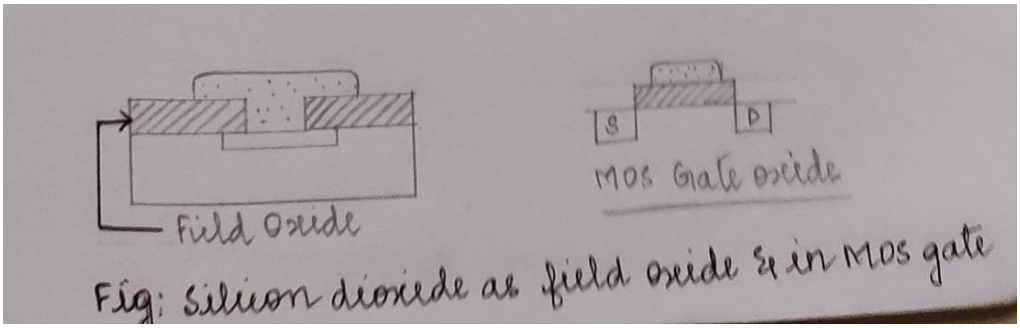
the metal layer causes a buildup of charge in the wafer surface. The surface charge can cause shorting and other unwanted electrical effects. A thick enough layer will prevent an induced charge in the wafer surface. Most of the wafer surface is covered with an oxide layer thick enough to prevent induction from the metal layers. This is called the *field oxide*.

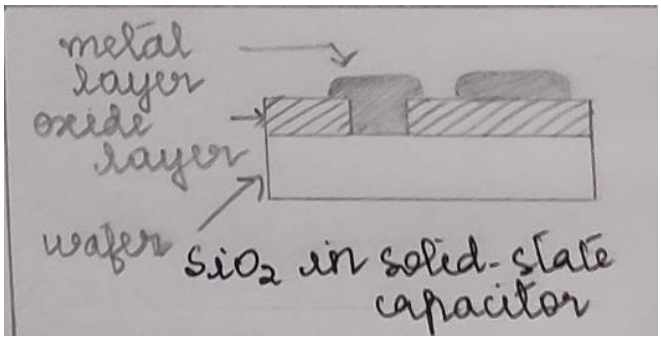
Device dielectric

The other technology is MOS technology. In an MOS transistor, a thin layer of silicon dioxide is grown in the gate region. The oxide functions as a dielectric whose thickness is chosen *specifically* to allow induction of a charge in the gate region under the oxide. The gate is the part that controls the flow of current through the device). The dominance of MOS technology for ultra-large-scale integrated (ULSI) circuits has made the formation of gate regions a prime focus of process development and concern. Thermally grown oxides are also used as the dielectric layer in *capacitors* formed between the silicon wafer and a surface conduction layer.

Silicon dioxide dielectric layers are also used in structures with two or more metallization layers. In this application, the silicon dioxide layers are deposited with *chemical vapor deposition* (CVD) techniques rather than thermal oxidation (see Chapter 12).

The silicon dioxide layers used in silicon-based devices vary in thickness. At the thin end of the scale are advanced MOS gate oxides, in the 35 to 80 A range.[1] At the thick end are field oxides.





****

5) Explain thermal oxidation mechanism.

Thermal Oxidation Mechanisms

Thermal oxide growth is a chemical reaction that occurs even at room temperature, but higher temperatures (between 900°C and 1200°C) are necessary for creating high-quality oxides at practical speeds for use in devices. Initially, when a silicon wafer is exposed to oxygen in a heated chamber, oxygen atoms combine with the silicon, and the oxide grows at a constant rate (linear growth). After about 1000 angstroms (Å) of oxide, the growth rate slows down, and the oxide enters a second phase known as parabolic growth. In this phase, the oxide layer grows more slowly as the oxygen has to diffuse through the existing oxide to reach the silicon, leading to a decrease in growth rate over time.

The change from linear to parabolic growth depends on the temperature and other factors. Generally, oxides under 1000 Å grow linearly, while thicker oxides are governed by the parabolic stage. The parabolic growth phase is significant because it means that thicker oxides require much more time to grow than thinner ones. For example, growing a 2000-Å oxide takes only 6 minutes, but doubling it to 4000 Å can take 220 minutes—over 36 times longer. This presents a challenge for semiconductor processing, especially when using pure dry oxygen, as thicker oxide layers require much longer oxidation times, which is inefficient for high-volume production.

To speed up the process, water vapor (H2O) can be used instead of oxygen. Water vapor accelerates the oxidation process because the hydroxyl ions (H–OH–) in the steam diffuse through the oxide layers more quickly than oxygen. This faster diffusion leads to quicker oxide growth. This process, known as steam or wet oxidation, is more efficient than dry oxidation, where only oxygen is used. When water vapor is used, the resulting oxide layer is less dense than one grown in dry oxygen due to the presence of trapped hydrogen molecules. However, heating the oxide in an inert atmosphere (like nitrogen) can make the two oxides similar in structure and properties.

****

6) With a neat diagram explain horizontal tube furnace system and its various sections used in thermal oxidation method.

Horizontal Tube Furnaces

Horizontal tube furnaces have been used since the 1960s for processes like oxidation, diffusion, heat treating, and deposition. Originally developed for diffusion in germanium technology, they are still often called "diffusion furnaces," though "tube furnace" is the more accurate term. These furnaces evolved into vertical designs, offering various advantages, but the basic principles apply to both.

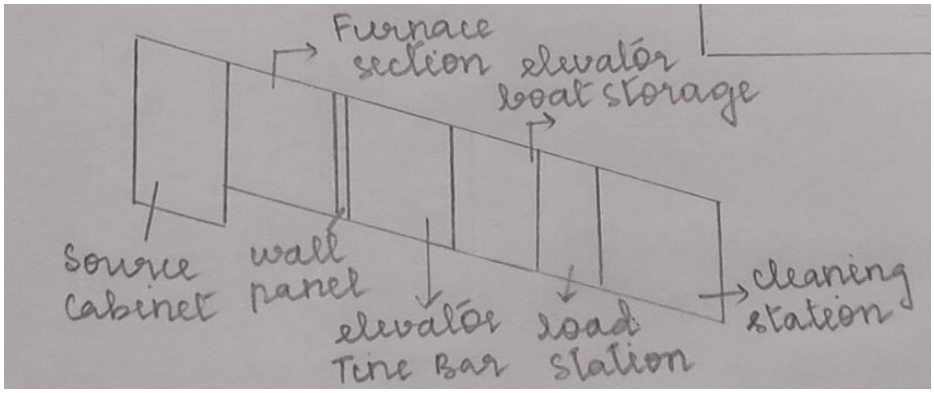
A typical three-zone horizontal tube furnace features a long ceramic tube with copper coils on the inside, each defining a zone and connected to a power supply controlled by a proportional band controller. Furnaces can have up to seven zones. Inside the tube is a quartz reaction tube

for the process, which may be surrounded by a ceramic liner, or muffle, to ensure more even heat distribution.

Thermocouples on the quartz tube send temperature data to the controllers, which adjust the power to the coils. The coils heat the tube through radiation (from the energy released by the coils) and conduction (where the coils touch the tube). These controllers are highly precise, maintaining temperatures within ±0.5°C, or about ±0.05% for a 1000°C process. In oxidation, wafers are placed in the flat zone, and oxidant gas is passed through the tube to carry out the reaction.

A production tube furnace is an integrated system of seven various sections:

1. Reaction chamber(s)
2. Temperature control system
3. Furnace section
4. Source cabinet
5. Wafer cleaning station
6. Wafer load station
7. Process automation
8. Reaction chamber



1) **Reaction Chamber**

The reaction chamber in a tube furnace protects the wafers from contamination and helps even out the temperature inside. Reaction tubes are typically round, with a gas inlet (source) end and a wafer load end, both featuring ground fittings for leak-proof connections. The load end may have an end cap or a wafer transfer unit.

Quartz is the traditional material for reaction chambers due to its high purity, stability at high temperatures, and cleanliness. However, it is fragile, can allow metallic ions from the coils to contaminate the chamber, and tends to break up (devitrification) and sag above

1200°C, affecting wafer placement. Quartz tubes are made by electric or flame fusion, with flame-fused tubes sometimes offering better characteristics.

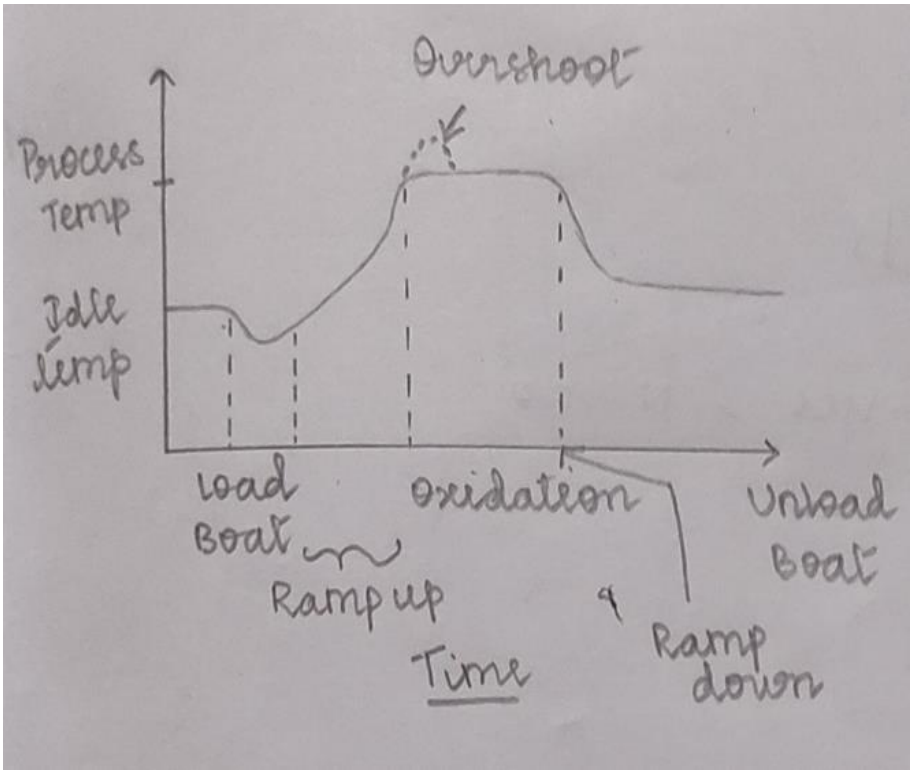
Quartz tubes need regular cleaning, typically with hydrofluoric (HF) acid outside the fabrication area, though this weakens the tube over time. Some companies use in-situ cleaning methods, such as plasma generators or etching gases, to clean the tubes while they're in use. These methods are especially useful in CVD processes, where particle buildup is higher.

Silicon carbide is an alternative to quartz, offering greater strength, resistance to temperature cycling, and better metallic ion barrier properties. It also develops a protective silicon dioxide layer during oxidation, which improves its longevity. However, its high cost and weight have limited its widespread adoption.

2) Temperature control system

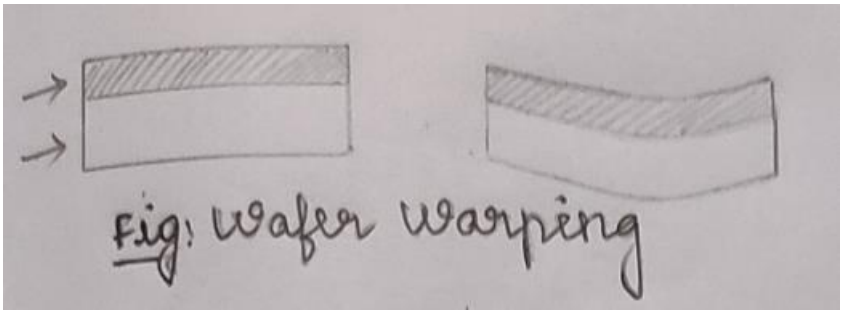
The temperature control system uses thermocouples to monitor the tube's temperature and proportional band controllers to adjust the heating coils. These controllers maintain the desired temperature by increasing or decreasing power based on the temperature deviation. This system ensures quick recovery without overshooting the set temperature. Adjustments are made until the flat zone temperature is correct.

In advanced systems, thermocouples on the tube wall feed data to a microprocessor, enabling a feature called autoprofiling. Regular checks with independent thermocouples ensure accurate temperature control.



Wafer warping, particularly with larger wafers, can occur because silicon expands faster than silicon dioxide, leading to concave deformation. To prevent this, two methods are used: ramping and slow loading. Ramping involves heating the furnace to a temperature lower than the process temperature, then gradually raising it after the wafers are inserted. Cooling the furnace to the lower temperature before removal helps prevent warping. The slow loading method involves inserting wafers at about 1 inch per minute to minimize warping.

Tube furnaces are usually kept near process temperature 24/7 to avoid issues with quartz-ware and maintain the flat zone. Some areas may keep the furnace at a lower temperature when idle to save energy. Both ramping and slow loading are used together for large wafers or batches. The heating system must also recover quickly after a load is added, as it can drop the tube temperature by 50°C or more, but recovery must occur without warping or overshooting.



3) Furnace section

A production-level tube furnace typically has three or four vertically stacked tubes, each with its own temperature control system. The tubes lead into an exhaust chamber (scavenger) that removes spent gases, which are then filtered through a scrubber to eliminate toxic substances before being released.

The tubes may serve a single purpose, like oxidation, or different processes such as oxidation, diffusion, alloying, or CVD, depending on the needs and wafer volume. Switching between processes requires changing the tube-
ware and monitoring for contamination to avoid cross-contamination from previous uses.

4) Source cabinet

Each tube process uses various gases to trigger chemical reactions, with oxygen or water vapor commonly used for oxidation. Nitrogen is also used to prevent unwanted oxidation during loading and unloading, and it flows continuously through the tube to maintain cleanliness and preserve the flat zone.

Gases are introduced into the tube in a precise sequence, pressure, flow rate, and duration. The equipment managing this is housed in a source cabinet attached to the furnace, with each tube having its own gas control panel. This panel includes solenoids, pressure gauges, mass flow controllers, and timers. In simpler systems, manual valves and timers are used, while more advanced setups rely on microprocessors for sequencing and timing.

Mass flow meters are preferred for precise control since they measure the mass of gas, not its volume, which can vary with pressure and temperature. These meters use a heated tube and two temperature sensors to measure the gas flow. Microprocessor-controlled valves in the source section ensure gases are delivered in the correct order and for the right amount of time.

The gas flow controller, often called a "jungle" because of its complex tubing, is made of stainless steel to prevent contamination. Gases are supplied from liquid gas storage or smaller gas bottles.

In some processes, gases are difficult to deliver, so a bubbler is used. This device allows gas to bubble through a liquid, absorbing chemicals that are then carried into the tube. Bubblers are common in oxidation, diffusion, and CVD processes.

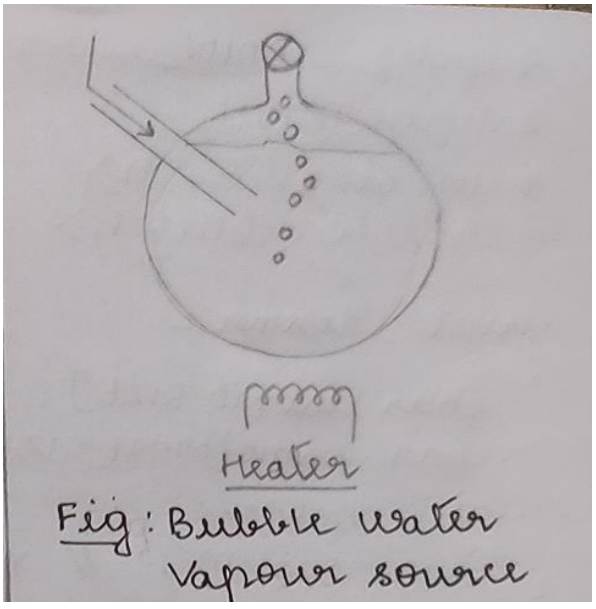


Fig: Bubble water vapour source

****