Fabrication of CMOS Devices

Vocabulary:

Wafer — disk of silicon, 4" to 8" in diameter, < 1mm thick, cut from ingots of single-crystal silicon which are "grown" from a pot of melted silicon. Impurities in melt determine *n*- or *p*-type material.

Single-crystal means that the crystalline structure is the same throughout - careful growing process ensures this.

Wafers are very brittle, the larger the diameter, the more susceptible to damage. Surface of the wafer is polished to a very flat, scratch free surface.

Oxidation — used to deposit Silicon Dioxide (SiO_2) on surface of wafer to be used as insulting material - heat wafers inside of an oxidation atmosphere such as oxygen or water vapor.

Need to introduce dopants into the silicon wafer.

Epitaxy — grow single crystal film of the required dopant on silicon surface by heating wafer and exposing it to a source of the dopant.

Deposition — evaporate the dopant onto the surface, then heat the surface to drive the impurities in the wafer

Ion implantation — expose surface to highly energized dopant atoms - when they hit the surface, they travel below the surface and become trapped

Need to control the regions in which the dopants get introduced - want to block some regions from receiving dopants.

Uses **masks** to block the impurities in particular regions.

To create mask:

(a) deposit mask material over entire surface
(b) cut windows in the mask to create exposed areas
(c) deposit dopant
(d) remove unrequired mask material

How do you pattern the mask material?

(a) Put a positive photoresist material on top of the mask material. Positive photoresist breaks down when exposed to ultraviolet (UV) light.

(b) Expose wafer to UV light. Block areas of the wafer from UV exposure via a mask on a glass plate. Exposed photoresist will be weakened.

(c) Put wafer in chemical bath to wash away exposed photoresist (called "**etching**" or "developing" the photoresist.

(d) Bake the wafer (**Hard Bake**). Photoresist + mask material combine to produce a material more resistant to etching then just the plain mask material.

(e) Etch the mask material via a chemical bath - areas not protected by the hardened photoresist will be washed away.

Finished! Now have patterned mask material on wafer surface - this will now block area of the surface from dopants.

Another way to pattern the mask - Electron Beam Lithography which directly cuts areas in mask material.

(1) Do not have to make a mask for the UV light - these are expensive, also errors can occur in making the mask dimensions which can cause problems down the road.

(2) Can make changes to the pattern quickly, draw very fine lines.

Main disadvantage ⇒ slow, electron beam has to "draw" over surfaces it wants to cut; time is proportional to complexity of mask pattern.

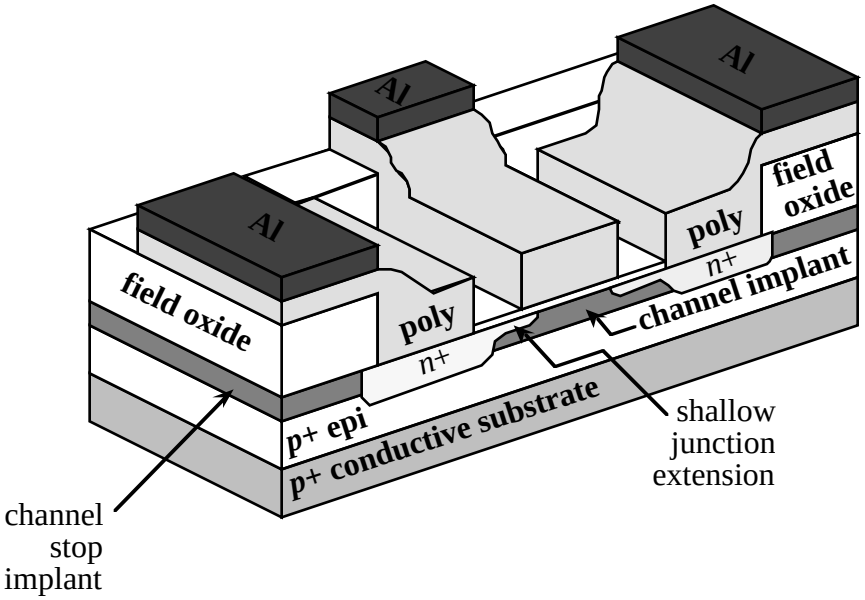
Back to UV masks

Two methods for using the UV masks:

(1) Step and Repeat — one mask the size of the chip is made, this is placed over a region of the wafer and exposed to UV light. Then, move mask to next region and repeat.
   (a) Only need one mask, mask cost low
   (b) slow

(2) Full wafer mask, make a mask the size of the wafer in which the "chip" mask is repeated many times
   (a) Full wafer mask is expensive
   (b) Fast processing, can expose entire wafer at once

Where are we going with this?



In the above picture:

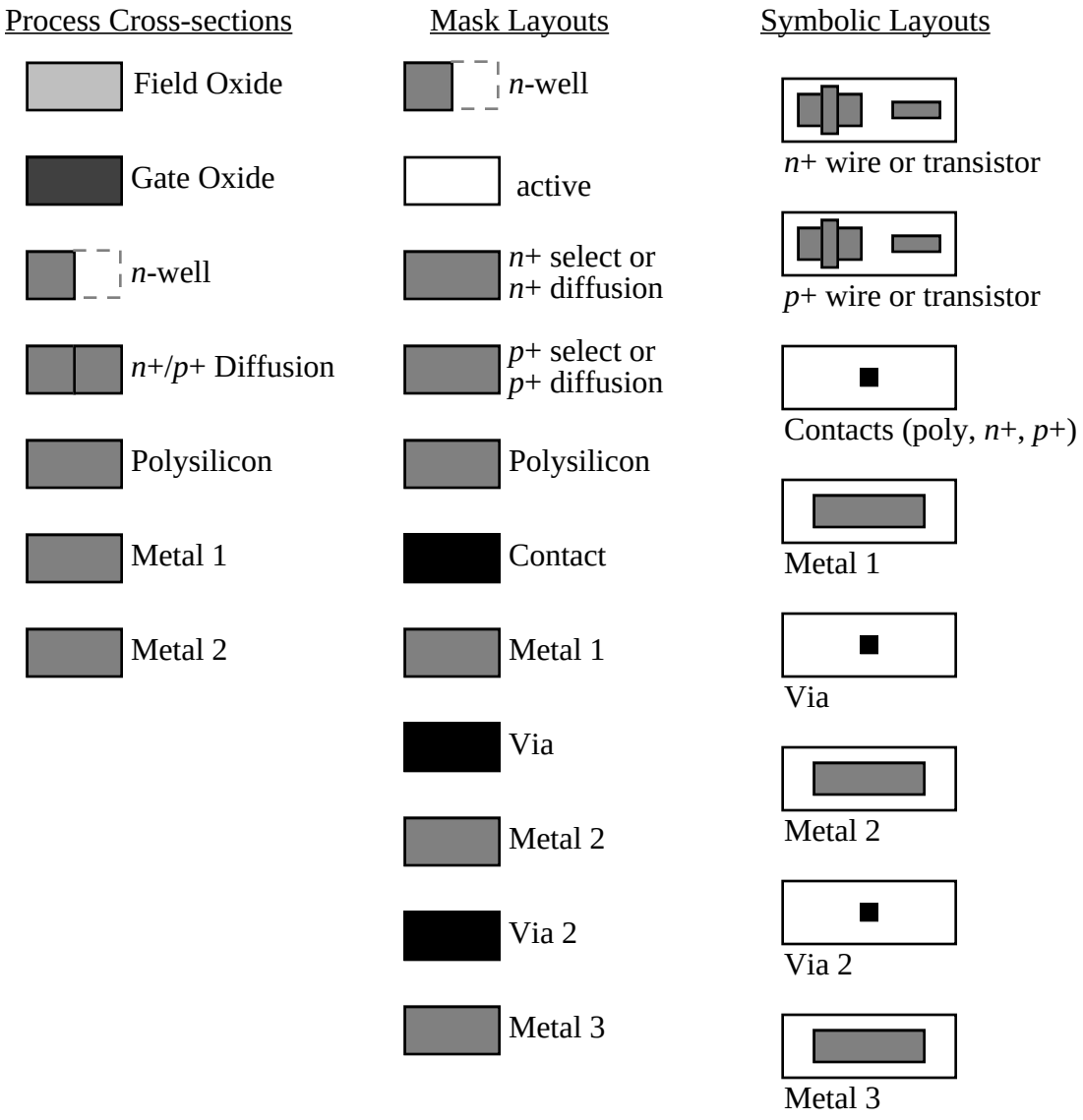
(a) Poly Gate — (Gate is made of polycrystalline - silicon which is made up of multiple crystal structures, not a single crystal like the substrate.) The poly is very heavily doped to be *n*+ so that it is a good conductor.

(b) Channel implant used to adjust threshold voltage

(c) Aluminum used for metal layer - this picture is somewhat unusual in that the metal is shown to contact to poly before contacting the drain/source regions - typically, metal directly contacts drain/source areas.

(d) Field oxide is *much* thicker than gate oxide ("thin ox")

What are the Layers? and Structures?



Diffusion areas are where $n+$, $p+$ have been introduced to form source, drain, and well-contacts areas.

"Active" refers to any area which will be exposed to any $n+$, $p+$ dopants, or channel areas.

Channel areas formed by polysilicon ("poly") crossing diffusion area. Area under the poly is the transistor channel.

Many different contact types - metal 1/diffusion, metal 1/poly, metal 1 to metal 2 (via), metal 2 to metal 3 (via 2).

A structure is a combination of basic layers to form a commonly used item (*n*MOS transistor, *p*MOS transistor, diffusion contact, poly contact, etc.).

CMOS Process Enhancements

Interconnect! The more layers of interconnect, the less area a layout will take.

Solution #1 - add more layers of metal

Three layers of metal very common. In advanced processes, not uncommon to have 4 or 5 layers of metal.

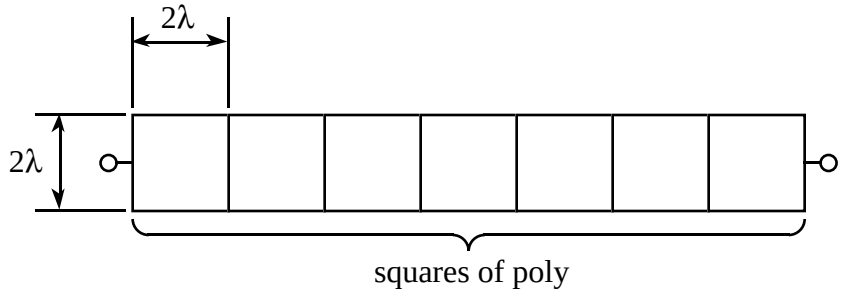
Con ⇒ more layers, more expensive processing, lower yields (more processing steps ⇒ lower yields)

Solution #2:

Make the poly layer a better conductor by lowering its resistance. Typical resistance ⇒ 20 to 40Ω per square.

Aside:

Compute resistance by the "square" to find the total resistance end-to-end.

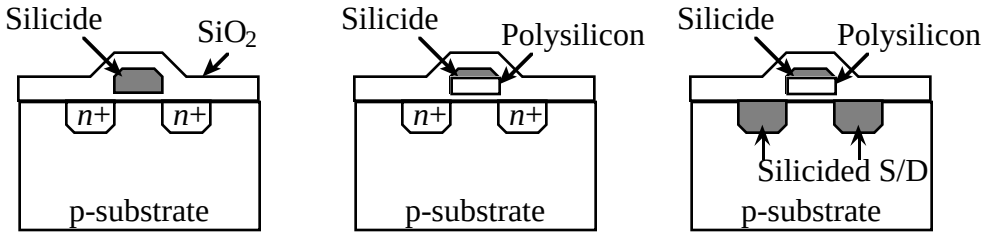


7 squares; if 30Ω/sq. then $R_{end\text{-}to\text{-}end} = 30 \times 7 = 210\Omega$.

More on this later . . .

How can we make polysilicon less resistive?

Combine it with a metal! (or use a metal for the gate).



Silicide Gate

Polysilicon/ Silicide Gate (Polycide)

Self-aligned Polysilicon/ Silicide Gate (Salicide)

(a) (b) (c)

Cross-section (a): Gate is made of a silicide—polysilicon combined with tantalum (other metals can be used). Sheet resistance $\Rightarrow$ 1 to 5Ω per square

Cross-section (b): Gate is a sandwich of silicide with polysilicon.

Cross-section (c): Source and drain regions are also a silicide!! Allows direct connection between gate and source/drain regions *without* using metal interconnect - reduces area!! Referred to as "local interconnect"

Local Interconnect example

Shown below is a memory cell example. Remember, AREA is very important!!! This example illustrates use of silicided source/drain regions. Consequently, this type of interconnect reduced area by 25%.