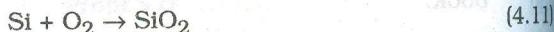


SiO_2 is generically known as quartz glass, or simply "glass," and is used for the gate oxide in a MOSFET, in addition to numerous other applications.

There are two types of SiO_2 layers found in VLSI circuits, with the distinction being how they are created. A **thermal oxide** is formed by the reaction



using heat as a catalyst. The unique aspect of a thermal oxide is that the silicon (Si) required for the reaction is obtained from the silicon wafer itself. This is illustrated in Figure 4.2(a) where oxygen molecules O_2 are passed over the surface of the wafer where the reaction takes place. This literally "grows" the glass layer with the results shown in Figure 4.2(b). The final thickness of the oxide is denoted as x_{ox} in the drawing, and depends on the temperature, crystal orientation, and growth time. Since silicon atoms from the surface of the wafer are used by the reaction, a layer of silicon with a thickness

$$x_{\text{Si}} \approx 0.46 x_{\text{ox}} \quad (4.12)$$

is consumed. An equivalent (and useful) viewpoint is that the surface of the silicon is "recessed" from its original location.

Although pure oxygen yields high-quality oxide layers, it is relatively slow. A faster growth rate is obtained using water (H_2O) in the form of steam via the reaction



which is called "wet oxidation." In practice, mixtures of O_2 and steam are used, along with nitrogen as a carrier gas and other chemicals such as chlorine (Cl).

Thermal oxide is a form of a **native oxide**, i.e., one that is created when the surface is exposed to an oxygenated atmosphere. If you take a bare silicon wafer and place it in air, a thin native oxide layer will form. Increasing the temperature enhances the growth rate. Silicon oxidation temperatures are typically in the range of about 850–1100 °C.

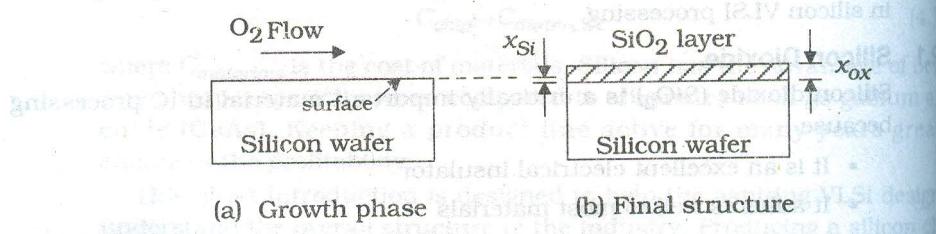


Figure 4.2 Thermal oxide growth

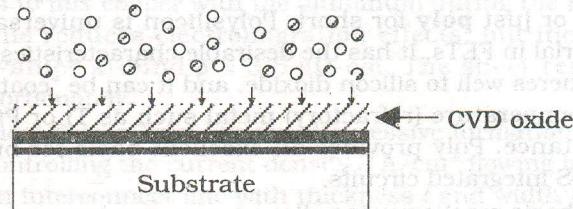
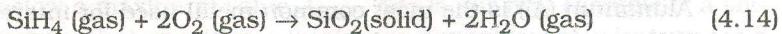


Figure 4.3 CVD oxide process

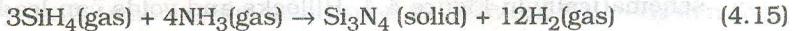
Most oxide layers in VLSI circuits are well above the wafer surface and no silicon is available for thermal oxide growth. In this case, we create SiO_2 molecules using gaseous reactions, and then **deposit** them onto the surface to provide an oxide coating. The process is shown schematically in Figure 4.3. A chemical reaction using silane (SiH_4) such as



can be used to produce the SiO_2 molecules above the wafer. This technique is called **chemical vapor deposition** (CVD) and the resulting layers are often called **CVD oxides**. The thickness of the oxide layer is controlled using the growth rate and deposition time. It is possible to perform the deposition at low temperatures, giving rise to the name **LTO** (low-temperature oxides). Also, it is sometimes advantageous to dope the glass. For example, phosphorus doping yields “P-doped glass” which helps certain types of planarization steps.

4.2.2 Silicon Nitride

Another useful material is silicon nitride Si_3N_4 , which is often just called “nitride” when the context is clear. The reaction



illustrates one technique. Nitrides are unique in that they act as strong barriers to most atoms. This makes them ideal for use as an **overglass** layer, which is a final protective coating on a chip, since it keeps contaminants from reaching the sensitive silicon circuits. Silicon nitride is used in a fabrication sequence that electrically isolates adjacent FETs (as will be discussed later). And, they have a relatively high dielectric constant $\epsilon_N \approx 7.8 \epsilon_0$, which makes them candidates for insulating ON (oxide-nitride) “sandwich” insulators in various capacitor structures such as those used in DRAM (dynamic random-access memory) cells.

4.2.3 Polycrystal Silicon

If we deposit silicon atoms on top of an amorphous SiO_2 layer, the silicon attempts to crystallize but can't find a crystal structure for reference. This

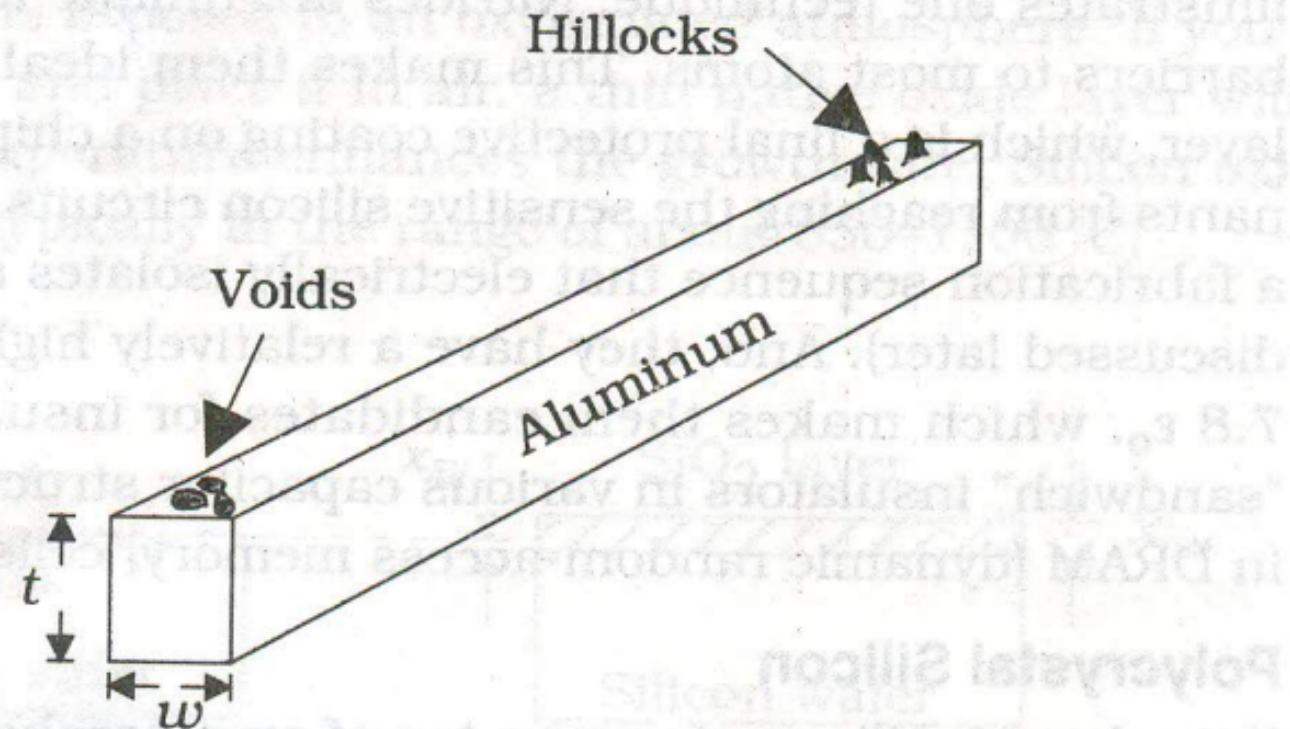
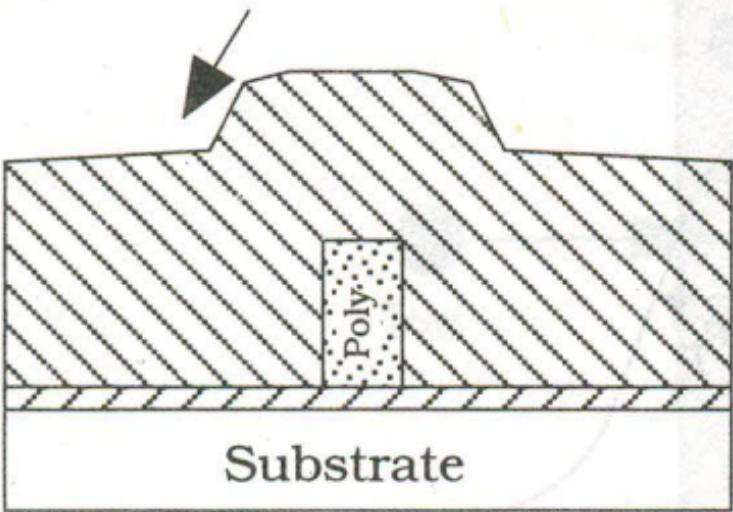


Figure 4.4 Visualization of electromigration effects in aluminum

surface of the chip, and then employ photographic-type techniques to transfer the pattern to the surface. The same process is used

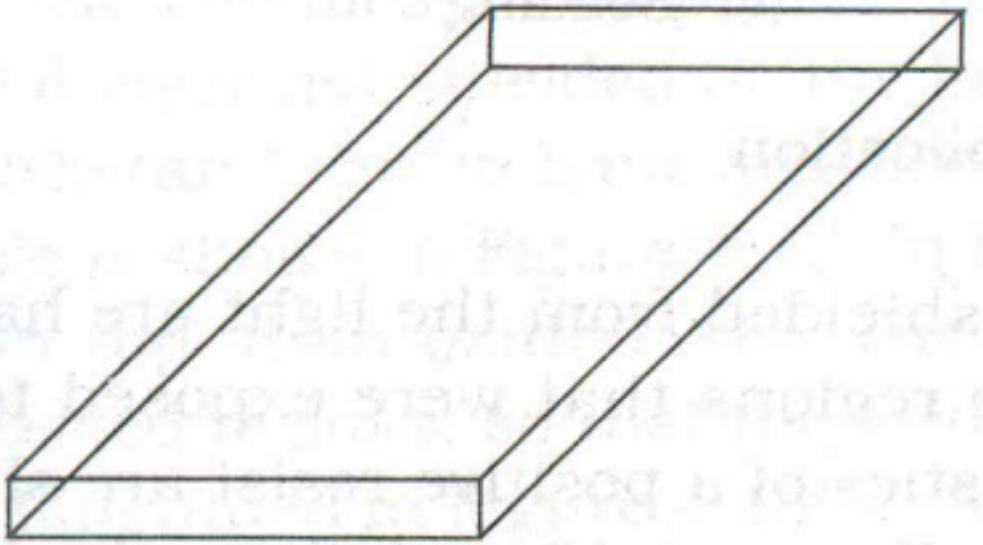
Deposited oxide



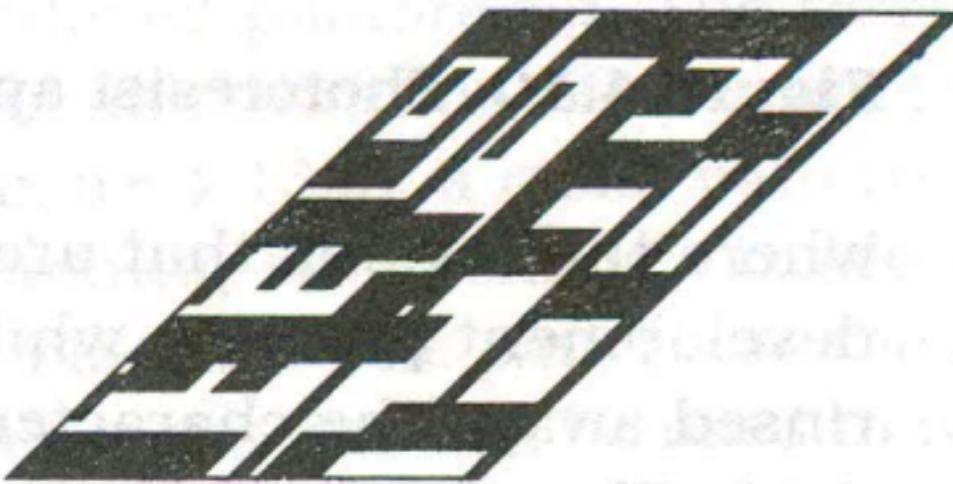
(a) After oxide deposition

(b) After CMP

the reticle onto the surface of the chip.



Glass



Pattern on
underside

Figure 4.9 A reticle is a glass plate with a chromium pattern

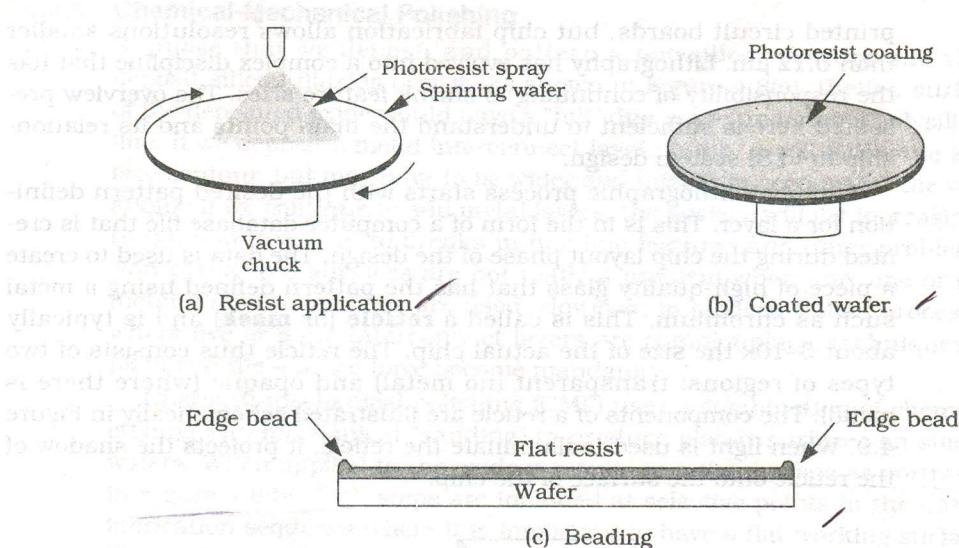


Figure 4.10 Photoresist application

where the regions that are shielded from the light are hardened in the development process, while regions that were exposed to the light are rinsed away. The characteristics of a positive resist are shown in Figure 4.12. The exposure step in Figure 4.12(a) defines the light and dark regions in the reticle shadow. After the resist is developed, hardened layers remain in the regions that were shielded from the light; this is illustrated in Figure 4.12(b). Negative photoresist has opposite characteristics: illuminated regions harden while shielded regions are soluble and are rinsed away.

The hardened resist layer is used to protect underlying regions from

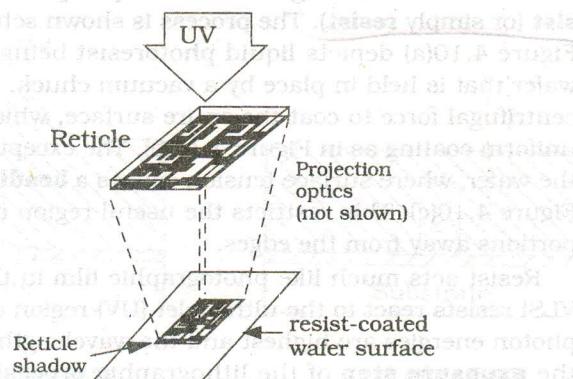


Figure 4.11 Exposure step

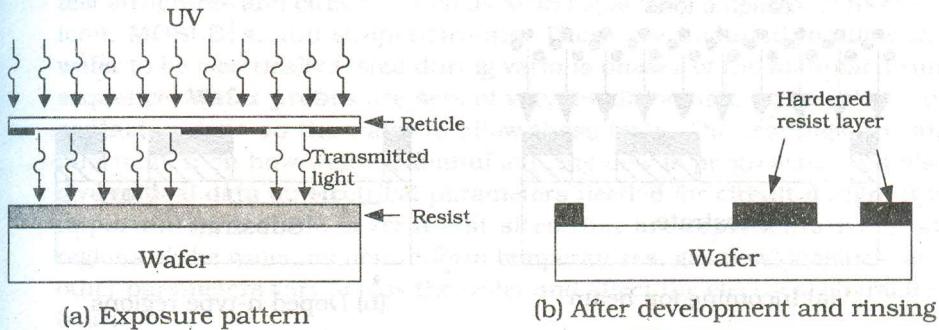


Figure 4.12 Characteristics of positive photoresist

the **etching** process. This is where the surface of the wafer is subjected to a gaseous plasma that is formed from an inert gas such as argon (Ar) and has reactant chemicals in it; overall, this is called a reactive-ion etch (RIE). The chemicals and plasma are chosen to attack and remove the material layer not shielded by the hardened photoresist. The resist itself can withstand the etchant mixture for the duration of the process. An example is shown in Figure 4.13. In Figure 4.13(a), a resist pattern is created on top of an oxide layer. The etching step removes oxide in the unprotected regions, so that the oxide has the same pattern as the resist; this is illustrated in Figure 4.13(b). This technique can be used to pattern any material layer above the wafer surface, including polysilicon, CVD oxides, and metals.¹ It allows us to transfer patterns from a computer layout design to the physical silicon level, thus creating the physical implementation of a logic network.

Doped silicon regions are also patterned using the lithographic process but the sequence is different. In this case, we grow an oxide layer on the wafer and then use lithography to etch down to the silicon surface; this is identical to the cross-section that was shown as Figure 4.13(b). The

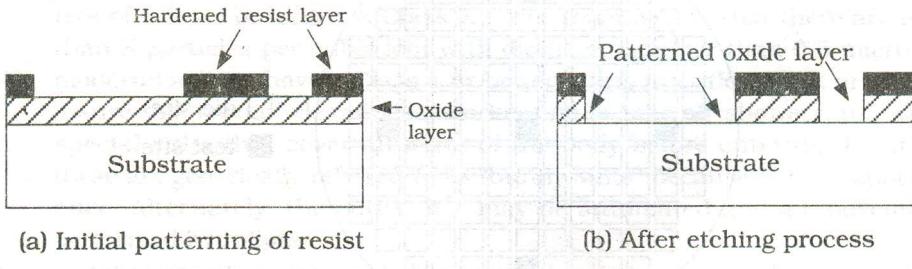
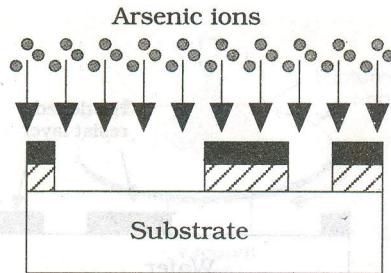
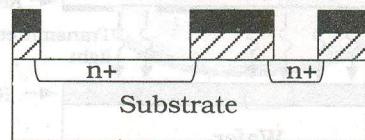


Figure 4.13 Etching of an oxide layer

¹ Copper is an exception as it is patterned using a different technique.



(a) Incoming ion beam



(b) Doped n-type regions

Figure 4.14 Creation of doped silicon patterns

resist-oxide layers are then used to shield the silicon from an ion implantation step. Figure 4.14(a) shows that an incoming beam of arsenic ions covers the entire surface, but the dopants can enter the silicon only where the oxide has been etched away. The resulting n+ regions are thus defined by the oxide openings. Note that the widths of the n+ patterns are slightly larger than the oxide openings. This is due to an effect called **lateral doping** that arises from dopant diffusion during the annealing step. Lateral effects can limit the resolution of a narrow-line printing system.

Although we have shown only a single pattern in our examples, the manufacturing processes use larger wafers that accommodate many individual chip sites. Each site is individually exposed using a **step-and-repeat process**; a **wafer stepper** is an apparatus that holds the wafer and allows accurate movement to align the optics to each site, one at a time. After a site is exposed, the mechanism “steps” the wafer to the next site. This sequence produces a wafer with a large number of identical sites as illustrated in Figure 4.15. The **test site** locations contain various

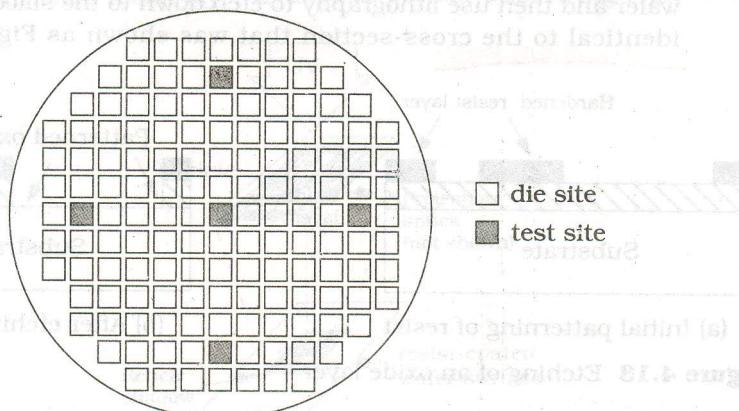
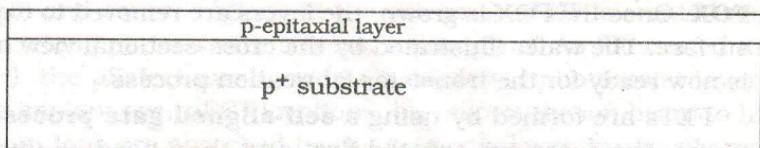
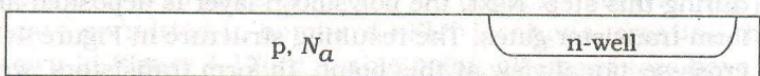


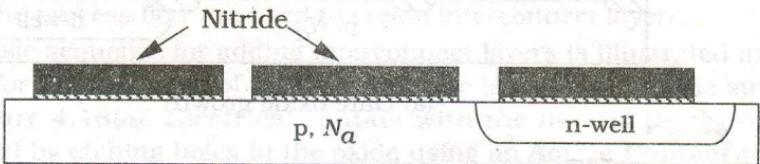
Figure 4.15 Wafer sites



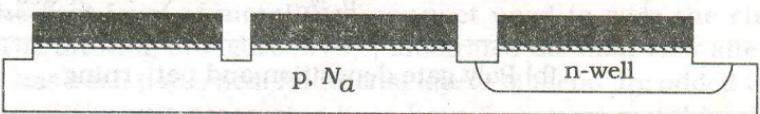
(a) Starting wafer with epitaxial layer



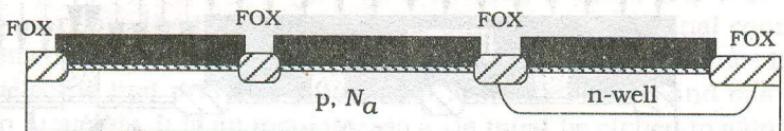
(b) Creation of n-well in p-epitaxial layer



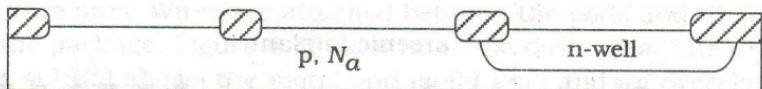
(c) Active area definition using nitride/oxide



(d) Silicon etch



(e) Field oxide growth



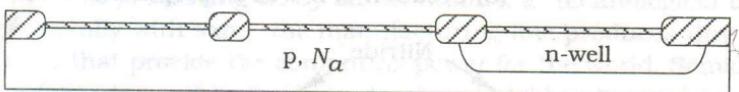
(f) Surface preparation

Figure 4.16 Initial sequences in the CMOS fabrication sequence

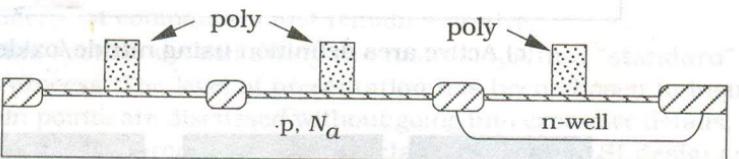
neighboring devices using recessed regions of glass (oxide) as an insulator. To achieve isolation, the nitride pattern is used to define silicon etched regions shown in Figure 4.16(d). Oxide is then grown or deposited in the etched regions as in Figure 4.16(e). Glass insulation between active areas defines the **field** regions, and the oxide there is called **field oxide** or

FOX. Once the FOX is grown, the layers are removed to expose the silicon surface. The wafer illustrated by the cross-sectional view in Figure 4.16(f) is now ready for the transistor fabrication process.

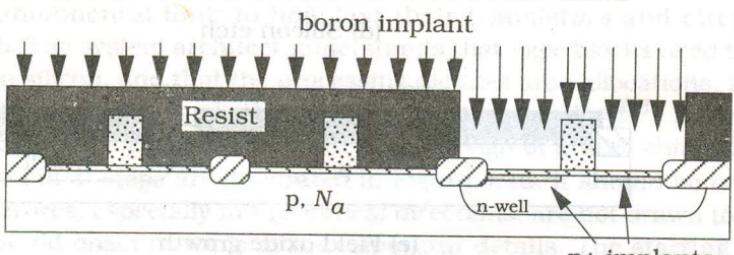
FETs are formed by using a **self-aligned gate process**. In this technique, the gates are created first and then used as implant masks to define the n+ or p+ drain/source regions. The starting point is the growth of the gate oxide shown in Figure 4.17(a). The value of t_{ox} is established during this step. Next, the polysilicon layer is deposited and patterned to form transistor gates. The resulting structure in Figure 4.17(b) shows the cross-sectional view at this point. To form transistors, we need to create



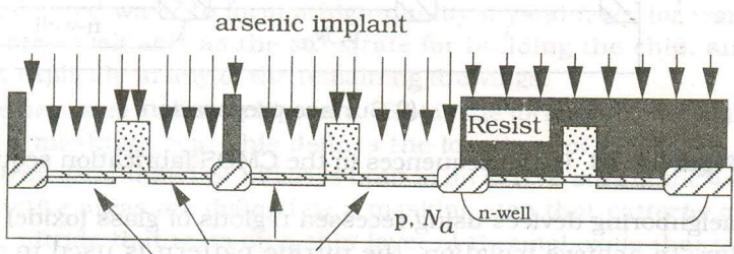
(a) Gate oxide growth



(b) Poly gate deposition and patterning

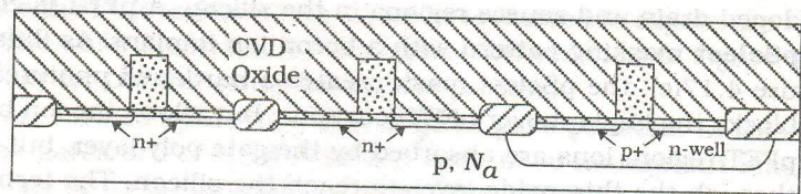


(c) pSelect mask and implant

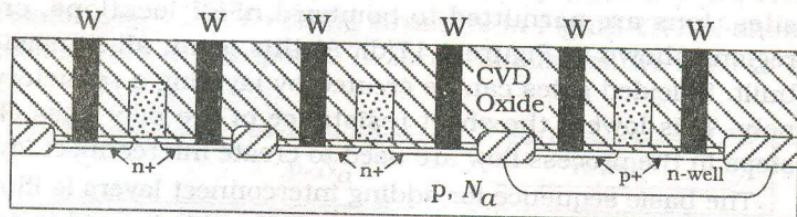


(d) nSelect mask and implant

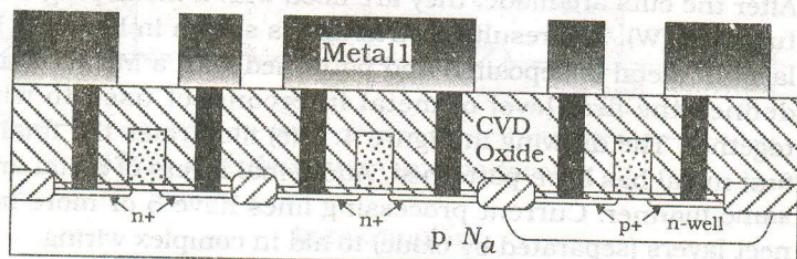
Figure 4.17 Formation of nFETs and pFETs



(a) After anneal and CVD oxide

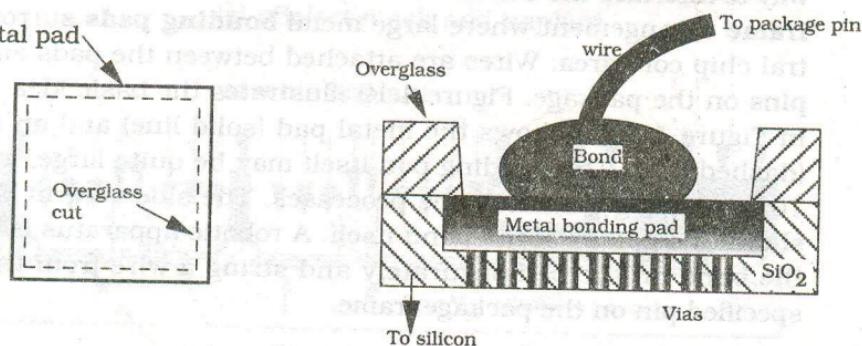


(b) After CVD oxide active contact, W plugs



(c) Metall coating and patterning

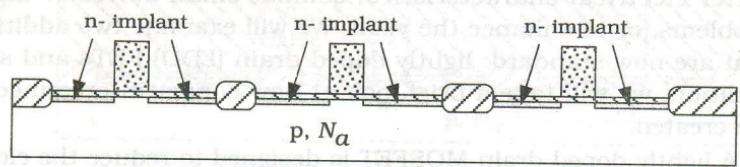
Figure 4.18 First metal interconnect layer



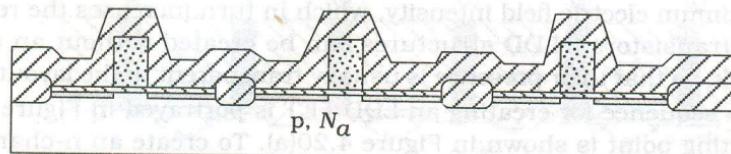
(a) Top view

(b) Side view

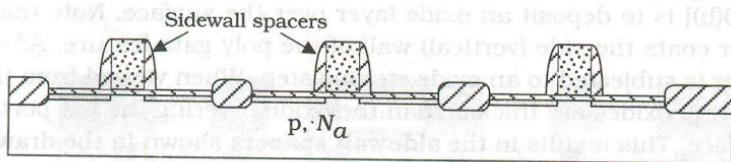
Figure 4.19 Bonding pad structure



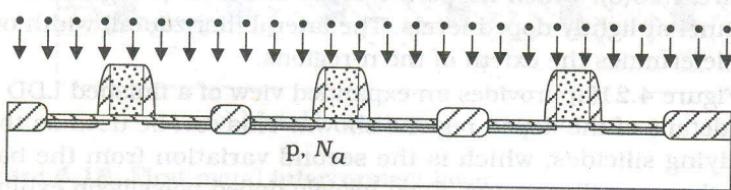
(a) Light (n-) implant



(b) Oxide coating

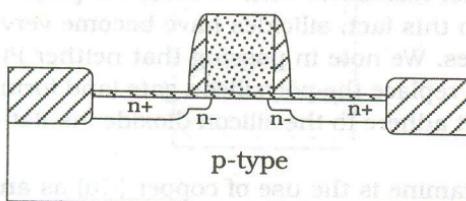


(c) After etching

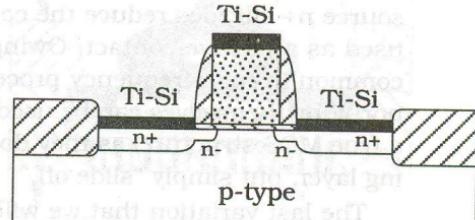


(d) Heavy donor implant

Figure 4.20 Sequence for creating a lightly doped drain nFET



(a) LDD FET structure



(b) Silicide formation

Figure 4.21 LDD nFET with silicided gate and contacts

interconnect material instead of aluminum. It is a well-known fact that the bulk resistivity of copper is $\rho = 1.67 \mu\Omega\text{-cm}$, which is about one-half that of Al. When used as an interconnect line material, the sheet resistance would be about one-half that of an aluminum line with the same thickness. Copper, however, has proved difficult to introduce into the processing line. It cannot be patterned using the standard sequence of deposition followed by a lithographic step because it is very difficult to etch using standard RIE techniques. Copper diffuses very rapidly through silicon and can alter the electrical characteristics, so it cannot be directly deposited on top of any silicon regions. It also diffuses through silicon dioxide, making the problem even more difficult to deal with. Much research has been directed toward the development of techniques to replace aluminum with low-resistivity interconnect metals. At the present time, copper is being introduced into the majority of new high-speed CMOS lines, making it of interest to the VLSI designer. One of the first VLSI chips to use copper technology was an advanced generation Power PC microprocessor design.

Let us first examine how copper patterns are produced. As mentioned above, dry-etching techniques do not etch copper. Even trace amounts of copper from Al-Cu mixtures are difficult to remove from a chip surface. To get around this problem, we use the **Damascene** process based on the method used in ancient times to inlay gold or silver into an iron sword. The name is taken from the city of Damascus whose artisans were well known for their work. In this technique, the copper pattern is first etched into a silicon dioxide layer; copper is then deposited (using, for example, electroplating) on the surface. The sequence is shown in

Figures 4.22(a) and (b). To avoid the etching problem, we subject the

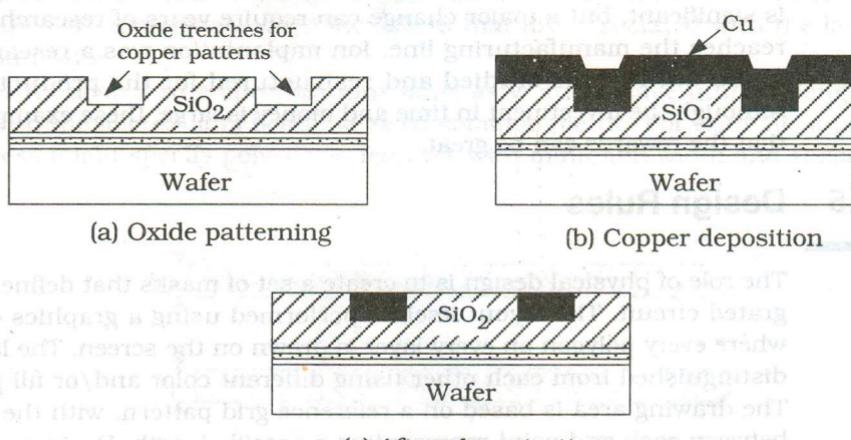
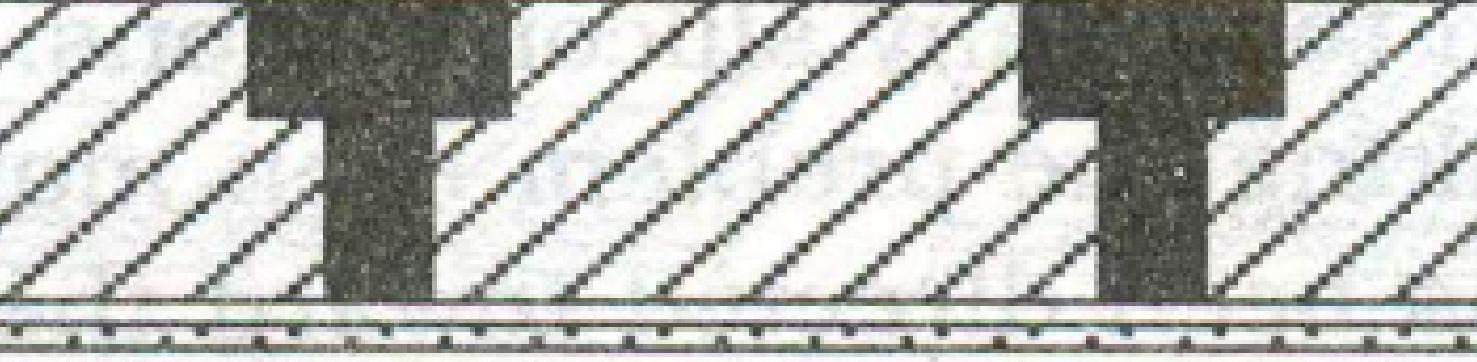


Figure 4.22 Copper patterning using the Damascene process

4.25. Copper vias have a lower resistance than tungsten vias due to the contact resistance introduced by a standard Al-W



Wafer

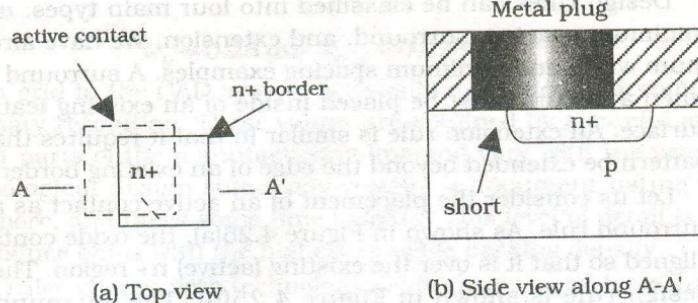


Figure 4.26 Misalignment-induced defect

The polysilicon gate is used as a dopant mask for the n-type ion implant that defines the drain and source regions. In Figure 4.27(a), the extension distance d_{po} (for poly overhang) is included to insure functional FET structures. If we do not provide the overhang distance, then a misaligned poly mask may result in the situation shown in Figure 4.27(b). In this case, the poly edge did not traverse the entire active area, so that the ion implant creates a short between the drain and source sides.

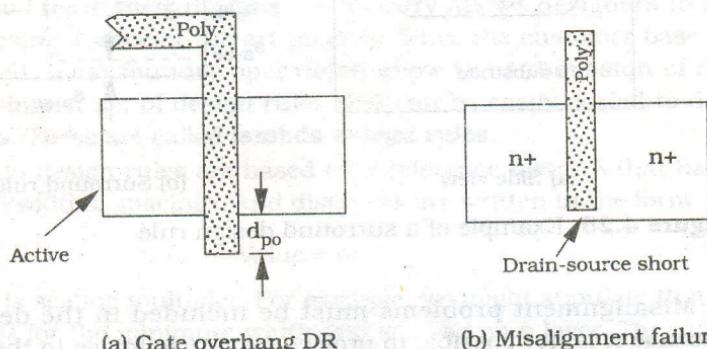


Figure 4.27 Example of an extend (gate overhang) design rule

4.5.1 Physical Limitations

Some geometrical design rules originate from physical considerations. These enter into the formulation of the design rule set, and may or may not be obvious.

An important aspect is the linewidth limitation of an imaging system. The reticle shadow projected to the surface of the photoresist does not have sharp edges due to optical diffraction. As a simple rule of thumb, a lightwave with an optical wavelength of λ cannot accurately image a feature size much less than that value. UV-sensitive positive photoresists are used because the short wavelengths of the ultraviolet light allow for better

resolution of fine linewidths, and positive resists have better development properties than negative resists. In addition, the structure of a reticle is much more complicated than we have alluded to; advanced optical techniques such as phase-shifting structures are used to enhance the resolution.

The etching process introduces another type of problem. When we remove material around a resist edge, both vertical (perpendicular to the wafer surface) and lateral (parallel to the surface) etching occurs. We can characterize the respective etch rates of the two by r_{vert} [$\mu\text{m}/\text{min}$] and r_{lat} [$\mu\text{m}/\text{min}$] and define the **degree of anisotropy** A by

$$A = 1 - \frac{r_{lat}}{r_{vert}} \quad (4.22)$$

The presence of lateral etching in r_{vert} limits the resolution that can be achieved. Figure 4.28(a) shows an oxide layer that is to be patterned by the resist layer on top of it. A pure anisotropic etch profile is shown in Figure 4.28(b). This is characterized by $r_{lat} = 0$ which gives vertical walls and $A = 1$. The result of a pure isotropic etch with $r_{lat} = r_{vert}$ is shown in Figure 4.28(c). Undercutting of the resist due to the lateral etching decreases the resolution that can be used in the design. Another factor that enters the problem is the absorption profile of light by the resist layer itself; this results in the resist edges having finite slopes instead of well-defined vertical shapes.

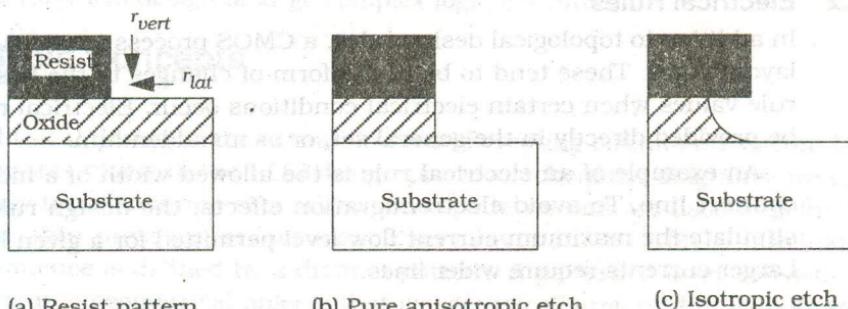


Figure 4.28 Etching profiles

Semiconductor effects in silicon also influence the formulation of design rules. Any time a pn junction is formed it gives rise to what is known as a **depletion region** at the interface. By definition, the depletion region is "depleted" of free electrons and holes because of an electric field that originates from the dopants and forces the charges out. If the depletion regions of adjacent pn junctions touch, then the current blocking characteristics are altered and current can flow between the two. This limits the spacing rule s_{n-n} shown in Figure 4.29. The drawing also