5.2.3 Gas delivery system

The gas delivery system handles processing gases and delivers them to the processing tube as required. A gas panel consists of regulators, valves, mass flow controllers (MFCs), and filters; it distributes processing and purge gases to the processing tube as needed. Processing gases are usually stored in high-pressure (more than 100 psi) bottles in remote gas cabinets. Processing gases are throttled to the gas panel by dozens of pounds per square inch (too high to be directly used in the process). Regulators and valves monitor the processing gas pressure and control pressure. Gas flow rate is precisely controlled by the MFC, which controls the rate by adjusting an internal control valve so that the measured flow rate equals the set value. Filters in the gas panel help minimize particulate contamination by preventing particles from flowing into the processing tube with the gases. Figure 5.4 shows a schematic of a gas delivery system.

5.2.4 Loading system

The loading station is used for wafer loading, unloading, and temporary storage. It loads wafers from cassettes into quartz boats, which sit on a quartz paddle; then the paddle is gently pushed into the processing tube by a loading mechanism, which is controlled by the microcontroller. After thermal processing, the paddle is gradually pulled out of the processing tube.

Several paddle types have been used in the loading systems of horizontal furnaces. Paddles with wheels are no longer used due to the particle generation caused by direct contact between the paddle and quartz tube. Soft-landing paddles, which gradually land on the quartz tube after the paddle is pushed into the desired position inside the processing tube, were developed next. Due to direct contact between the paddle and quartz tube, soft-landing paddles caused particulate contamination from friction between the paddle and tube wall during paddle landing and raising. To avoid this direct contact, suspended or cantilevered paddles are currently used in most systems. They allow the paddle to move in and out of the processing tube without any direct contact with the quartz surface, minimizing particle contamination. However, wafer loading can affect the paddle's suspension.

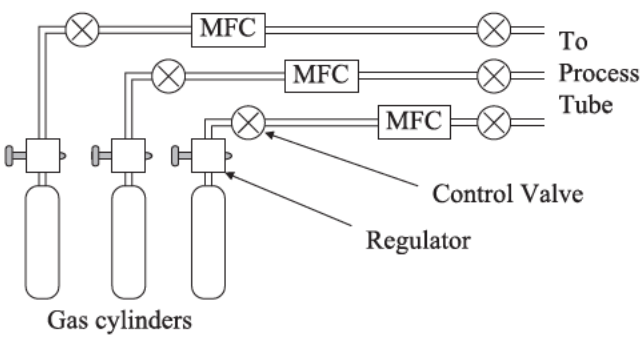

Figure 5.4 Schematic of the gas delivery system.

For a vertical furnace, wafers are loaded from the cassette into the tower by a robot. Then a lift mechanism (controlled by the microcontroller) raises the tower into the processing chamber, or the chamber elevator lowers the chamber until the entire tower is inside the chamber. This eliminates contact between the wafer holder and the quartz chamber wall, and there is no suspension problem for the wafer holder.

5.2.5 Exhaust system

Process byproducts and unused source gases are removed from the processing tube or chamber by an exhaust system. Exhaust gas is evacuated from the processing tube symmetrically, and purge gas is drawn in from the exhaust manifold to prevent backstreaming of the exhaust gas. For furnace processes involving pyrophoric or flammable gases, such as silane (SiH_4) and hydrogen (H_2), an additional chamber called a burn box is required. In the burn box, exhaust gases pass through a controlled combusting process in the presence of oxygen and are reduced to less harmful, less reactive oxide compounds. Filters are used to remove particles generated in the burning process, such as silicon dioxide, after burning silane. A scrubber with water or an aqueous solution absorbs most of the toxic and corrosive gases before the exhaust gas is vented to the atmosphere.

5.2.6 Processing tube

Wafers undergo a high-temperature procedure in the processing tube, which consists of a quartz tube chamber body and several heaters. Thermocouples touch the chamber wall and monitor chamber temperature. Each heater is independently supplied by a high-current power source. The power of each heater is controlled by feedback from the thermocouple data through the interface board and microcontroller and becomes stable when the setting temperature is reached. Temperature in the flat zone at the center of the tube is precisely controlled within 0.5 °C at 1000 °C. Figure 5.5 shows schematics of horizontal and vertical furnaces.

In a horizontal furnace, wafers are placed on quartz boats, which sit on a paddle made of SiC. The paddle (with wafer boats) is then slowly pushed into the quartz tube where it places the wafers in the flat zone of the furnace for thermal processing. After processing, the wafers must be pulled out very slowly to avoid warping caused by the overwhelming thermal stress of sudden temperature change.

In a vertical furnace, wafers are loaded into a wafer tower made of quartz or SiC. The wafers are placed in the tower face up, then the tower is slowly raised into the quartz tube to heat. Afterward, the tower is lowered slowly to avoid wafer warping.

5.3 Oxidation

The oxidation process is one of the most important thermal processes. It is an adding process that adds oxygen to a silicon wafer to form silicon dioxide on the wafer surface. Silicon is very reactive to oxygen; thus in nature, most silicon

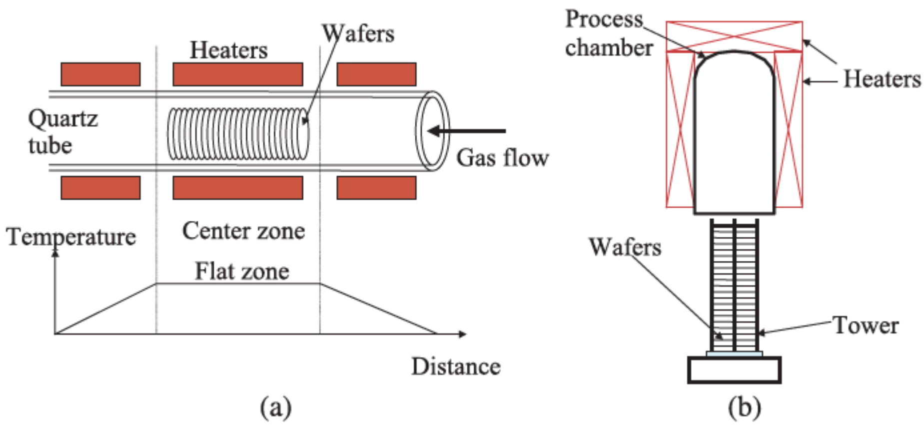

Figure 5.5 Schematics of (a) horizontal and (b) vertical furnaces.

exists in the form of silicon dioxide, such as quartz sand. It reacts with oxygen very quickly, and forms silicon dioxide on the silicon surface. The reaction can be expressed as

$$Si + O_2 \rightarrow SiO_2.$$

Silicon dioxide is a dense material that fully covers the silicon surface. To continue the oxidization of silicon, oxygen molecules have to diffuse across the oxide layer to reach the silicon atoms underneath and react with them. The growing silicon dioxide layer increasingly blocks and slows the oxygen. When bare silicon is exposed to the atmosphere, it reacts almost immediately with oxygen or moisture in the air and forms a thin layer (about 10 to 20 Å) of silicon dioxide, called native oxide. The layer of the native oxide is thick enough to stop further oxidation of the silicon at room temperature. Figure 5.6 illustrates the oxidation process.

In the oxidation process, oxygen comes from the gas phase, and silicon comes from the solid substrate. Therefore, while silicon dioxide is growing, it consumes the substrate silicon, and the film grows into the remaining silicon substrate, as shown in Fig. 5.6. Oxygen is used as an oxidizer in processes such as thermal oxidation, CVD, and reactive sputtering deposition. It is also commonly used in etch and photoresist stripping processes. Oxygen is the most abundant element in the earth's crust and the second most abundant element in the earth's atmosphere, after nitrogen. Some facts about oxygen are listed in Table 5.1.

At high temperatures, thermal energy causes oxygen molecules to move much faster, which can drive them to diffuse across an existing oxide layer and react with silicon to form more silicon dioxide. The higher the temperature, the faster the oxygen molecules, and the quicker the oxide film growth. The oxide film quality is also better than that grown at lower temperatures. Therefore, to get high-quality oxide film and fast growth rate, oxidation processes are always performed in a high-temperature environment, normally in a quartz furnace. Oxidation is a slow

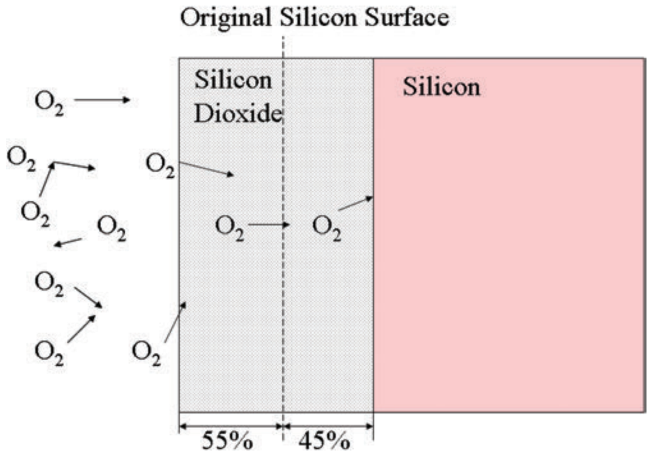

Figure 5.6 Silicon oxidation process.

Table 5.1 Facts about oxygen.

Symbol	O
Atomic number	8
Atomic weight	15.9994
Discoverer	Joseph Priestley, Carl Scheele
Discovery location	England, Sweden
Discovery date	1774
Origin of name	From the Greek words oxy genes, meaning acid (sharp) and forming (acid former)
Molar volume	17.36 cm^3
Velocity of sound	317.5 m/sec
Refractivity	1.000271
Melting point	54.8 K = −218.35°C
Boiling point	90.2 K = −182.95°C
Thermal conductivity	0.02658 W m^{-1} K^{-1}
Applications	Thermal oxidation, oxide CVD, reactive sputtering, and photoresist stripping
Main sources	O_2, N_2O, O_3

process; even in furnaces hotter than 1000 °C, a thick oxide (>5000 Å) still takes several hours to grow. Therefore, oxidation processes usually are batch processes, with a large number (100 to 200) of wafers processing at the same time to achieve reasonable throughput.

5.3.1 Applications

Oxidation of silicon is one of the basic processes throughout the IC process. There are many applications for silicon dioxide; one of them is as diffusion mask. Most dopant atoms used in the semiconductor industry such as boron and phosphorus have much lower diffusion rates in silicon dioxide than they do in single-crystal silicon. Therefore, by etching windows on the masking oxide layer, a silicon substrate can be doped at a designated area by a dopant diffusion process, as shown in Fig. 5.7. The thickness of the masking oxide is about 5000 Å.

Screen oxides are also commonly used for ion implantation processes. They can help prevent silicon contamination by blocking the sputtered photoresist. They can also minimize the channeling effect by scattering incident ions before they enter a single-crystal silicon substrate. The thickness of a screen oxide is about 100 to 200 Å. Figure 5.8 illustrates a screen oxide for application of ion implantation.

Thermally grown silicon dioxide is used as a pad layer for silicon nitride in both local oxidation of silicon (LOCOS) and shallow trench isolation (STI) formation. Without the stress buffer from this pad oxide, LPCVD silicon nitride film could crack and in some cases even break a silicon wafer due to the high tensile stress (up to 10^{10} dynes/cm^2). The thickness of the pad oxide is about 150 Å.

Silicon dioxide is also used as a barrier layer to prevent contamination of the silicon substrate before trench fill in the STI process. Trench fill is a dielectric CVD process in which undoped silicate glass (USG) is deposited to fill the trench for electrical isolation of neighboring transistors. Since the CVD process always brings small amounts of impurities, a dense, thermally grown silicon dioxide barrier layer is necessary to block possible contamination. Figure 5.9 shows the pad and barrier oxides in the STI process.

One of most important applications of thermally grown silicon dioxide in the past was forming isolation blocks to electrically isolate neighboring transistors in an IC chip. Blanket field oxide and LOCOS were two kinds of blocks used

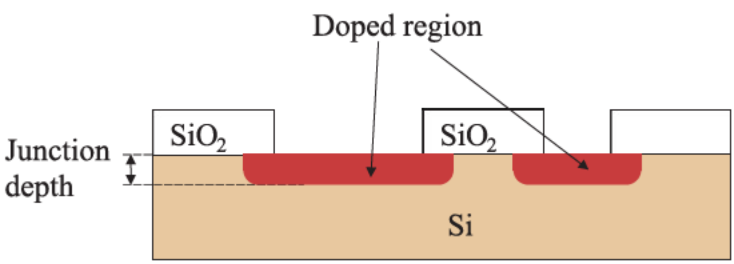

Figure 5.7 Diffusion masking oxide.

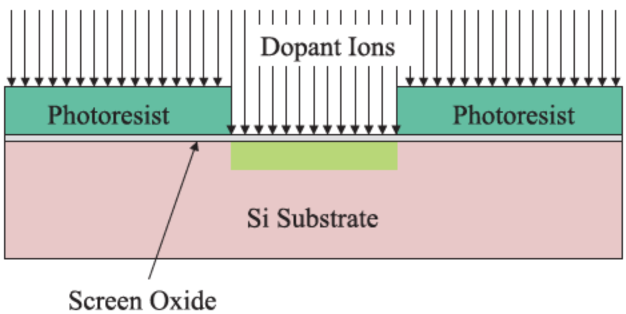

Figure 5.8 Screen oxide for application of ion implantation.

to isolate neighboring devices and prevent them from crosstalking. Blanket field oxide was the simplest isolation process and widely used in the early years of silicon IC manufacturing. By thermally growing a thick layer of silicon dioxide (5,000 to 10,000 Å), patterning it with photolithography, and etching the oxide with hydrofluoric acid (HF), the activation areas could be opened for transistor making, as shown in Fig. 5.10.

LOCOS has better isolation properties than blanket-field oxide. The LOCOS process uses a thin layer of oxide (200 to 500 Å) as the pad layer to buffer the strong tensile stress of LPCVD nitride. After the nitride etch, photoresist strip, and wafer cleaning, a thick layer of oxide (3000 to 5000 Å) is grown on the area not covered by silicon nitride. Silicon nitride is a much better barrier layer than silicon dioxide. Oxygen molecules cannot diffuse across the nitride layer; therefore, the silicon underneath the nitride layer does not oxidize. In the area not covered by the nitride, oxygen molecules continuously diffuse across the silicon dioxide layer,

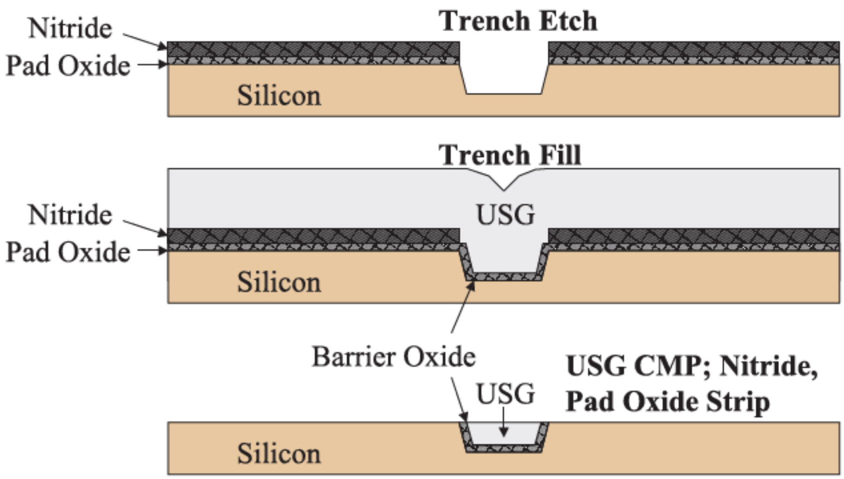

Figure 5.9 Pad oxide and barrier oxide in the STI process.

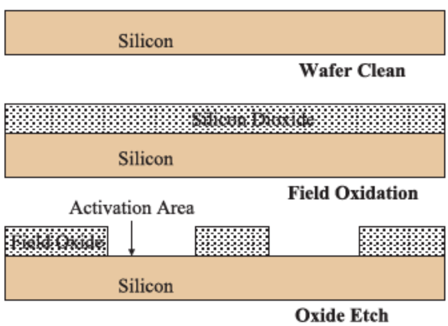

Figure 5.10 Blanket field oxide process.

where they react with silicon underneath to form more silicon dioxide. The LOCOS formation process is illustrated in Fig. 5.11.

Because oxygen diffusion inside silicon dioxide is an isotropic process, oxygen also reaches silicon at the side. This causes the oxide to grow underneath the nitride layer near the etched oxidation window, forming the so-called bird's beak (see Fig. 5.11). An undesirable consequence is that this bird's beak uses up a good deal of space on the wafer surface. Another disadvantage of LOCOS is the surface planarization problem due to the oxide-to-silicon surface step, caused by the oxide's growing characteristics, illustrated in Fig. 5.6.

Many different methods have been used to suppress the bird's beak. The most commonly used method is poly-buffered LOCOS (PBL). A thicker pad oxide allows a longer bird's beak to grow because of the wide oxide path for oxygen diffusion. By using a polysilicon layer (~ 500 Å) to buffer the high tensile stress of the LPCVD silicon nitride, the thickness of the pad oxide layer can be reduced from 50 to 100 Å, significantly reducing oxide encroachment. But, no matter how many different methods were tried, there was still an approximately 0.1- to 0.2-μm bird's beak at both sides of the LOCOS, a thickness that is intolerable when the minimum feature size is smaller than 0.35 μm. The STI process has been developed to avoid the bird's beak problem and has a more planarized surface topology as well. STI gradually replaced LOCOS isolation starting in the mid-1990s when device feature size was smaller than 0.35 μm.

Sacrificial oxide is a thin layer (< 1000 Å) of silicon dioxide grown on activation areas of a silicon surface and stripped in HF solution right after its growth. It is applied before the gate oxidation process to remove damages and defects on the silicon surface. This oxide growth and removal processing sequence helps to create a defect-free silicon substrate surface for growing a high-quality gate oxide layer.

The thinnest and most important silicon dioxide layer in a MOSFET-based IC chip is the gate oxide. While device dimensions have been shrinking, the thickness

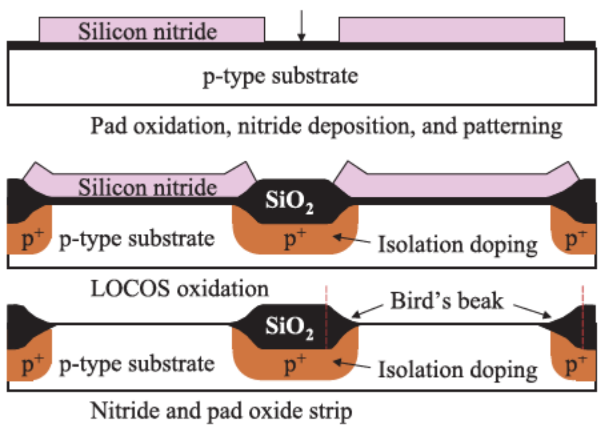

Figure 5.11 LOCOS process.

of gate oxides has been reduced from more than 1000 Å in the 1960s to about 15 Å in high-end chips in the mid-2000s; operation voltage of IC chips has been reduced from 12 to 1.0 V. The quality of gate oxide is vital for proper device function. Any defect, impurity, or particle contamination in a gate oxide can affect device performance and significantly reduce chip yield. Figure 5.12 illustrates the sacrificial oxidation and gate oxidation processes. Applications of thermally grown silicon dioxide in IC fabrication processes are summarized in Table 5.2. For nanometer technology node IC chips, the gate oxide is nitrided to increase its dielectric constant, thus it can increase in thickness with a higher breakdown voltage, while still possessing larger gate capacitance to maintain proper MOSFET switching properties. Also, a nitrogen-rich layer can block the diffusion of boron in heavily p-type doped polysilicon gate electrons into the n-type doped channel of pMOS.

5.3.2 Preoxidation cleaning

Thermally grown silicon dioxide is an amorphous material. It is unstable, and its molecules tend to cross-link to form crystalline structures. This is the main reason silicon dioxide exists in the form of quartz and quartz sand in nature. Since the crystallization of amorphous silicon dioxide takes millions of years at room temperature, amorphous silicon dioxide in IC chips is very stable in their lifetimes. However, the crystallization process is dramatically accelerated at the high temperatures (>1000 °C) required during silicon dioxide growth. If the silicon surface is not free of contaminants, defects and particles can serve as nucleation sites for crystallization during the oxidation process, and silicon dioxide will grow into a polycrystalline structure similar to ice crystals that forms on glass in the winter. Crystallization of silicon dioxide is very undesirable, since it is not uniform and the crystal boundaries provide easy paths for impurities and moisture.

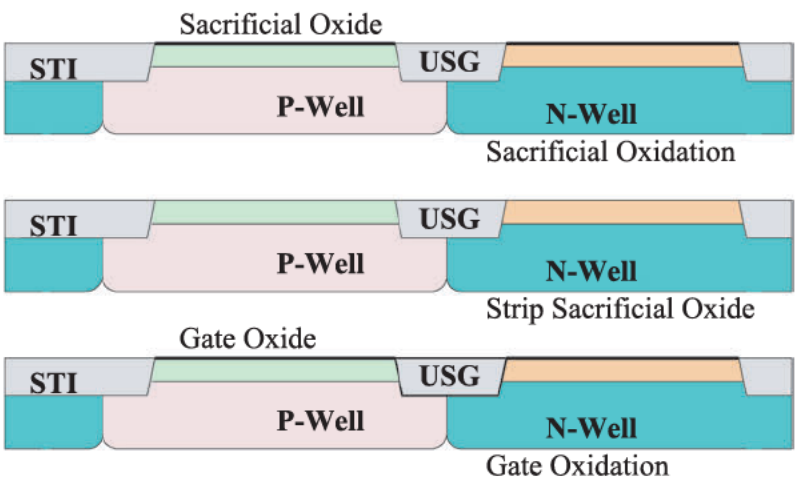

Figure 5.12 Sacrificial and gate oxidation processes.