4.4 DIFFUSION & OXIDATION

4.4.1 Current Industry Characteristics

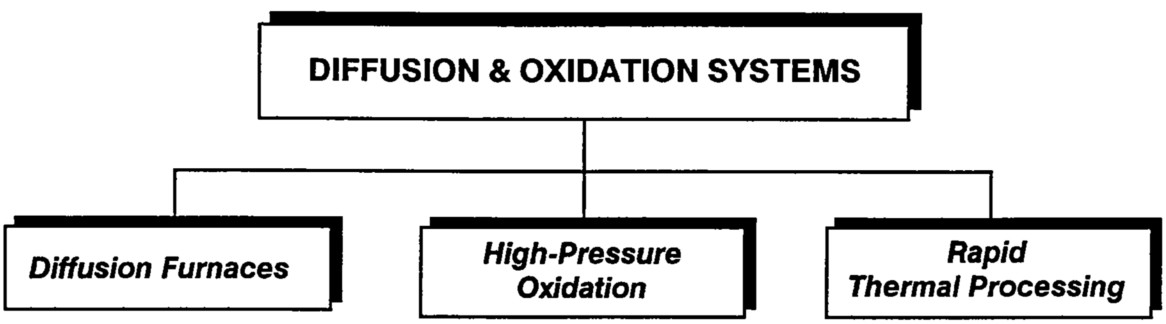
4.4.0 Diffusion and Oxidation

- **The diffusion market is defined by three market segments: diffusion furnaces, high pressure oxidation and rapid thermal processes.**
- **These systems are used to grow or deposit thin insulating films and to redistribute dopant layers.**

There are three market segments of the diffusion/oxidation equipment industry. They are diffusion, high pressure oxidation and rapid thermal processing, as depicted in Figure 4.4.0-1.

Diffusion is the general name of the process used to introduce impurities, called dopants, into the semiconductor wafer. These impurities alter or modify the electronic properties of the silicon. Diffusion is normally carried out in two steps. Dopants are first deposited onto the surface of the wafer by one of a several methods. Then they are placed in the diffusion furnace. The high temperatures of the diffusion furnace cause the dopant atoms to diffuse into the surface somewhat like a cloud of smoke diffuses into the atmosphere in a room. Diffusion furnaces must be very precisely controlled in order to precisely determine the dopant density profile.

Figure 4.4.0-1



Source: VLSI RESEARCH INC
2244-302P

Oxidation is a process that occurs naturally, i.e. the rusting of iron or the tarnishing of silver. Almost all material oxidizes in some way, however, the natural oxidation process is sometimes agonizingly slow and can take weeks, months or even years. Oxidation furnaces–which are really diffusion furnaces fitted with special gas supply and distribution systems–speed up the oxidation process so that usable thin oxide layers can be formed on silicon surfaces in minutes, or at most a few hours.

High pressure oxidation equipment is an evolution of the technology. The major advantage in high pressure oxidation is that process temperatures can be lowered substantially, thus reducing wafer warpage and other deleterious temperature effects encountered with conventional oxidation systems.

Rapid thermal processing (RTP) equipment is the third segment in this market. RTP was initially used to repair damage to the crystal lattice created during the ion implant process by annealing, and to activate dopant atoms that had been implanted. Other uses that are gaining popularity for RTP include glass reflow, silicide or polysilicon annealing for interconnect, and resistor applications. There are several types of RTP equipment depending upon the type of energy source used. The earliest systems used lasers and electron beams. Later systems incorporate graphite heaters or high intensity lamps.

Diffusion/oxidation furnaces are built in two basic configurations, horizontal and vertical. Horizontal configurations are usually stacked, one on the other, three or four high while vertical configurations can be clustered side by side to save valuable floor space. Diffusion furnace processes are batch processes usually processing 200 to 300 wafers in a single tube. RTP process equipment can be either batch or single wafer processes.

The diffusion/oxidation equipment market accounted for approximately 6% of the total wafer process equipment market by the end of the 1980's. The diffusion/oxidation segment amounted to about $309.8M annually in 1990. The following paragraphs will describe each of the subsegments in further detail.

4.4.1 CURRENT INDUSTRY CHARACTERISTICS

(Index continued on next page)

4.4.1 CURRENT INDUSTRY CHARACTERISTICS

4.4.1 Current Industry Characteristics

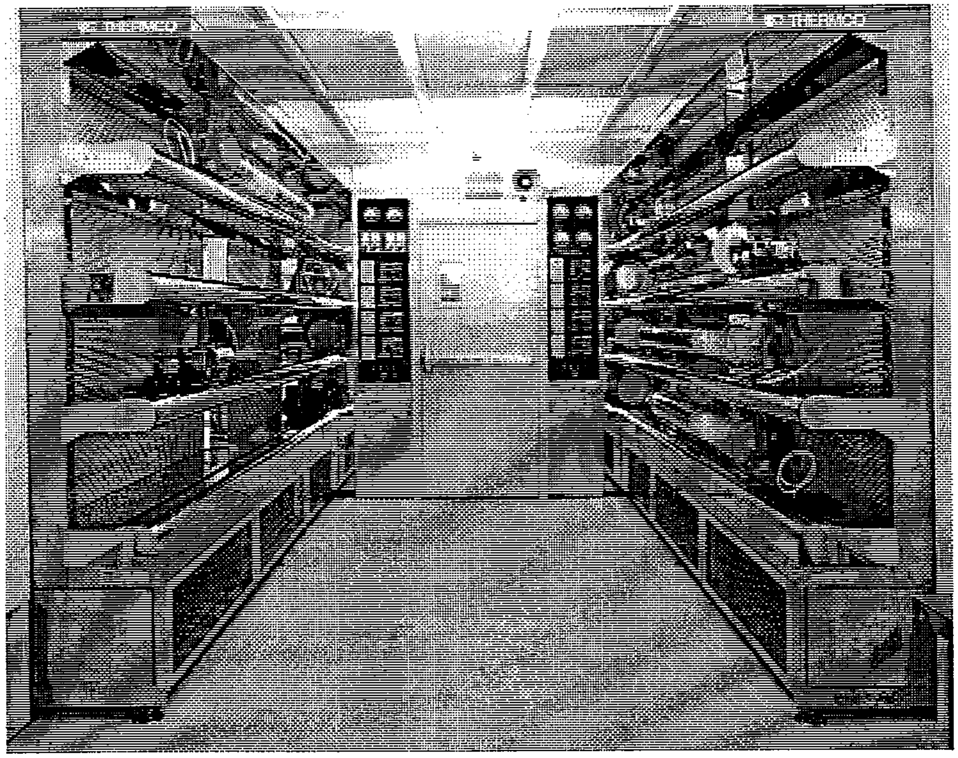
- **Horizontal diffusion systems remain the process workhorse.**
- **Vertical diffusion systems are gaining popularity as wafer sizes increase.**
- **Rapid thermal processing capabilities expand beyond annealing of ion implant damage.**

Diffusion/oxidation equipment has existed for several decades. Diffusion equipment is a good example of evolution of semiconductor process equipment from the 'tool' stage to the 'workstation' stage and further to the 'system' level. It is one of the most refined and mature segments of semiconductor capital equipment and is one of the few remaining batch processing systems. For this reason, it is often overlooked. Yet this equipment provides the vehicle for performing many critical processes including gate oxidation, isolation, and many hot-wall CVD and LPCVD processes. Figure 4.4.1-1 depicts a modern diffusion furnace available from Silicon Valley Group.

Although diffusion technology has undergone 40 years of evolution and is quite mature, equipment manufacturers continue to refine these systems in order to improve yield of devices, reduce equipment cost and accommodate larger wafers. For example, continued improvements in automation assure uniform batch processing and reduce contamination. Cantilevered loading and soft-landing wafer handling systems eliminate or reduce the friction and breakage which generate particulates in the furnace. Computer controls eliminate the need for the operator to make process-related decisions which helps to assure uniformity between batches. Some recent systems incorporate process monitoring and control capabilities which can both sense, and take corrective measures, to salvage a batch.

Almost all early diffusion systems were built in a horizontal configuration in stacks of one to four individual tubes stacked one on the other in a single cabinet. In the past few years vertical configurations have gained popularity due to several factors. Vertical systems take less floor space and the wafer loading process can be more easily automated. Major process advantages of vertical configurations are better control of the process gas, potentially less particulate contamination and less induced wafer warpage.

Diffusion equipment has been strongly influenced by other process advances within the industry. For several years growth in the diffusion market was slowed by advances in CVD, Rapid Thermal Annealing and Ion Implantation. Some technologists even believed that diffusion furnaces would become obsolete. Nevertheless, in the past several years, diffusion processes have not only thrived, but their use has increased. This increase is mainly due to the need to replace high temperature processes with those operating at lower temperatures. The diffusion process is an equilibrium process and is a function of time and temperature. To produce the same impurity profile using lower temperatures requires significant in-



Source: SVG-Thermco
2244-335

Figure 4.4.1-1

A Modern Horizontal Diffusion Furnace Installation

creases in diffusion time, with an associated decrease in throughput. This has resulted in an increase in demand for diffusion furnaces systems.

Another way to reduce processing temperature is to increase the ambient pressure during oxidation. This works particularly well for thick oxidation steps and has resulted in the introduction of high pressure oxidation systems, called Hi-Pox Systems. These systems incorporate a high pressure enclosure which allow the oxidizing process to be carried out at a pressure of several atmospheres. At these high pressures the oxidation process proceeds at a much higher rate than at normal atmospheric pressure for a given temperature. Therefore the required oxide film thickness can be grown in a reasonable time at a much lower temperature.

VLSI and ULSI processes have been developed which require very shallow junctions with steep dopant profiles. Ion implantation can be used to produce the desired shallow profile but subsequent annealing of the implant damage in a standard diffusion furnace will cause some unwanted diffusion to occur. As a result, rapid thermal processing, called RTP, has begun to emerge as an alternative to furnace annealing to heal the damage created by the ion-implant processes and to activate the implanted dopant. RTP makes use of a heat source which can be rapidly turned on and off and quickly heats the silicon surface. This annealing process can produce less wafer warpage, shallower junctions, and lower particulates than normal diffusion furnaces since the wafer temperature is only briefly increased. Moreover, since it is a single wafer process, better wafer-to-wafer uniformity is achieved while the equipment occupies less floor space and consumes less power. These factors can combine to give significant yield improvement, especially in sub-micron device structures.

RTP systems are now being introduced for epi growth and polysilicon deposition, as well as for oxidation and diffusion.

4.4.1.1 Development of the Industry

4.4.1.1.1 Diffusion Furnaces

Diffusion equipment was originally developed in the early 1950's as device manufacturers searched for methods to create p-n junctions in semiconductors. As was the case with most early semiconductor processing equipment, diffusion furnaces were first developed and built by the process engineers. The evolution of the diffusion equipment is a prime example of the migration of the state-of-the-art from tools to work stations and finally to complete systems that have the capability to process wafers in a 'black box' fashion.

Early diffusion furnaces used silicon carbide glowbar heating elements inside a chamber lined with firebricks. Wafers were placed in quartz tubes inserted into the center of this assembly. Electrical power to the coils was supplied by saturable core transformers, and temperatures were measured by inserting a thermocouple into the quartz process tube between wafer batches. Process gas flow control was accomplished by a needle valve placed in the input gas line. Steam oxidation processes were accomplished by boiling de-ionized water on a hot plate and piping the steam into the end of the quartz tube. Wafers were laid flat on a quartz plate which was inserted into the center of the quartz diffusion tube. Diffusion times were determined by an operator with a stop watch. These early crude diffusion tubes could only maintain temperatures to ±10 degrees or so.

In 1960, Lindberg, a manufacturer of small laboratory ovens and furnaces, acquired Heavy Duty Corporation of Milwaukee, Wisconsin. Heavy Duty built large brick-lined heat treatment equipment for the automobile and steel industries. Drawing from experts from both companies, Lindberg formed a new division in Watertown, Wisconsin to focus on the growing semiconductor processing industry. This new division soon developed the first Kovar wire-coil heating element configuration. Hard refractory material insulation continued to be used. At about this same time, the first electronic temperature controller, utilizing silicon controlled rectifiers (SCR's), was introduced by Electric Control Systems. Lindberg combined their coiled heating element technology and the ECS Model 761 temperature controller to offer a compact industrial diffusion furnace. Thermocouples were placed outside the quartz process tube to control temperature and provided the necessary input for temperature feedback control. This design allowed Lindberg to demonstrate process chamber temperature uniformities of ±5 °C over several inches of length within the chamber.

In 1964, two Lindberg engineers, Jim Donnely and Karl Lange, left to found Thermco Systems in Southern California. In cooperation with Motorola, Thermco developed the industry's first three-zone diffusion-furnace, called the Pacesetter. Center zone uniformity, called the flat zone, of ±1 °C over 72 hours was now made possible. The Pacesetter could accommodate wafers up to two inch diameter.

By 1967, Thermco had developed the industry's first light-mass insulation material. A wet insulating fibre material was first hand-wrapped around the heater coil. After drying, the assembly was then enclosed in an aluminum sheet metal covering for protection. This technique formed a compact, lightweight heating element with self-contained insulation. The more efficient and uniform insulation provided by this fabrication method allowed still better temperature uniformity than had been previously achieved. Thermco called this new model the Pacesetter II. Flat zones, of ±0.5 °C could now be produced over 15 to 20 inches. The Pacesetter II was to become the forerunner of modern diffusion systems.

Lindberg and Thermco supplied only the heating elements and the power control systems. It was usually left to the fab engineer to design and build the process gas control system and the distribution plumbing. These crude gas control systems were called 'jungles', because of the maze of tubing, valves, etc. assembled. Several small custom fab shops sprung up to supply packaged 'gas cabinets' containing these jungles, custom-built to the process engineer's specifications.

Loading stations were unknown at the time. Batches of wafers in quartz 'boats' were manually pushed into—and pulled out of—the furnace by using a long quartz rod.

During this time frame, engineers specializing in diffusion technology were in high demand as applications for semiconductor devices flourished. So by the late sixties, diffusion and oxidation had become very well understood. Precisely-controlled temperatures had been pushed well into the ranges of 1000 °C to 1300 °C. At these temperatures diffusion processes could be carried out with high throughput. For example, a few hours at 1000 °C would now produce the same junction depths that once took several days at 700 °C, or would take hundreds of years at 400 °C.

But as the use of higher process temperatures became common, deleterious effects began to appear. By 1968, the industry had moved from one and one-half inch diameter wafers to two inch diameter wafers. These larger wafers bowed and warped severely at such elevated process temperatures. By the mid-seventies, still larger three inch and four inch diameter wafers were being used. These large wafers curled like potato chips. Moreover, isotropic mappings of circuit parameters on the wafer began to look like weather maps, with the isometric lines weaving and undulating across the wafer face. It became clear that high processing temperatures were greatly affecting yield.

As solutions were sought, four effects received increasing attention: a) plastic distortion due to the nearly liquid phase of the silicon; b) curl caused by strains due to the different coefficients-of-expansion of the various layer and the pure silicon; c) doping variations due to the turbulent gas flow caused by the loading arrangement of wafers, and finally; d) uncontrolled and unintentional diffusion variations due to the different times and speeds that the wafer holder was being pushed into, or pulled from, the furnace's 'flat temperature zone.' The first two problems were primarily process design related while the latter two problems might be solved by the diffusion equipment designers.

At the same time, users were demanding that diffusion furnaces be capable of processing larger diameter wafers and larger batch sizes. Five and six inch wafers were already within the foreseeable future. Manufacturing engineers wanted furnaces that could process four inch wafers in batch sizes of 200 to 400 wafers and could be upgraded to five or possibly six inch capability in the future.

Around 1973, Lindberg developed the next generation of light mass insulation called Mold-a-Therm. In this process, a slurry of insulating fibre material and water is mixed with the consistency of very wet clay. A tubular mold formed of metal screen material is closed at one end and the opposite end is attached to a large hose. The hose is connected to a high volume pump. As the screen mold is immersed in the tank containing the fibre slurry, the slurry water is pumped through the screen leaving the fibre material behind on the surface of the screen. As the pumping is continued, additional layers of the fibre are attracted to the screen. When sufficient thickness is attained, the mold with the clinging fibre is removed from the slurry tank and placed in a drying oven. The fibre material hardens into a high density, lightweight insulation formed around the screen. The insulation is cut lengthwise and the screen mold is then removed. The heater coil is then placed into one half of this clamshell and the other half is put into place. When the seams are sealed with fibre tape, and placed in a aluminum outer tube, the whole assembly becomes a *uniform*, strong, lightweight integrated furnace heating element that is bolted into one position on a frame skeleton.

Temperature control also continued to be improved. The new generation of furnaces were able to maintain 36 inch-long flat zones at $\pm 0.5\,^{\circ}C$. Motorized pusher/pullers were available by the early seventies. These simple wafer loading systems further evolved into clean benches and became known as load stations. Jungles incorporated mass flow controllers. Wet oxidation systems no longer used DI water bubblers but instead used pyrogenic torches in which carefully controlled ratios of hydrogen and oxygen were burned to form water. Diffusion furnace manufacturers were now supplying the furnace and also the associated gas delivery systems and load stations. Simple programmable sequencers were employed to 'tie together' the entire process from pushing the wafers into the furnace at a fixed rate, to cycling the required gasses on and off for preprogrammed times and then pulling the wafers out of the furnace at a controlled rate.

In the early 1970's growing awareness of the deleterious role of particles in device yield, led engineers to devise better and cleaner methods of wafer loading. Until then, boatloads of wafers were simply slid into the diffusion tube with quartz grinding against quartz. This grinding of quartz against quartz was quickly recognized as a major source of particles. Attempts were made to fashion wheels into the wafer boats but this helped little. Cantilevered systems called 'soft landing loaders' were then devised in which the entire boat load of wafers is suspended at the end of a cantilevered arm as the boat is inserted into the furnace. When the boat reaches the center zone of the processing area, it is gently lowered to the bottom of the diffusion tube where it remains throughout the diffusion cycle. After processing the boat is again raised and withdrawn without any sliding contact with the diffusion tube walls.

Because the material used for the cantilevered loader must be able to maintain its strength at high temperatures, the choice of materials is limited. Quartz and silicon carbide are the two main choices. These materials are fragile and quartz begins to soften at the higher process temperatures. Catastrophic breakage is common. Systems were developed to automate wafer transfer from standard process cassettes to quartz boats, but these systems were relatively unwieldy and expensive and were not widely accepted.

By the late 1970's diffusion furnaces had entered into an advanced phase of development in which the equipment suppliers were now building complete systems. Process engineers, in attempting to integrate sequential processes and reduce wafer warpage, began to load wafers into the furnace at a low temperature and then ramp the entire furnace up to the process temperature in a precisely controlled manner. In a subsequent step, the furnace temperature could again be raised to a different level and a different process gas mixture injected. Finally, the temperature could then be lowered to the starting temperature and the wafers pulled into the load area at a very slow rate so as to carefully control the cool down process.

This system viewpoint quickly drove the desire to integrate all process controls into a single programmable device. In the mid to late seventies, first, BTU in cooperation with Western Electric, and then Thermco working with Hewlett-Packard developed direct digitally controlled (DDC) systems in which one mini-computer controlled all functions in up to eight separate diffusion tubes. These systems implemented the classical PID control algorithms entirely in software and added the ability for automatic calibration and temperature profiling through a sophisticated algorithm to reduce temperature variation to less than ± 0.25 degree Centigrade. By the late 1970s, BTU had refined its DDC control systems using microprocessors to replace expensive mini-computers.

Lindberg and Thermco dominated the diffusion furnace market throughout the 1970's. In the early 1970's Lindberg was acquired by Sola Basic and Thermco by Sunbeam. In 1971 Lindberg also acquired Tempress in Los Gatos, CA and EMB in Mountain View, CA. Tempress manufactured scribers and other small tools used in wafer processing while EMB manufactured aluminum wire bonders. By 1976, Lindberg had acquired Unicorp, a manufacturer of epitaxial reactors, and Corotech, who manufactured rinser-dryers. At that time Lindberg combined Tempress, Corotech, Unicorp and the diffusion furnace division of Lindberg into a single company called Tempress Microelectronics in Los Gatos, CA. Tempress sold diffusion furnaces in Japan through a joint venture between Koyo and Lindberg called Koyo-Lindberg. In 1977, General Signal acquired Sola Basic and its subsidiaries including Lindberg and Tempress.

Tempress introduced a 'state-of-the-art' direct digital control system, the first of which was shipped to VLSI Technology in 1982. But Tempress was never able to compete successfully with Bruce and Thermco. In the early 1980's Tempress introduced one of the first vertical diffusion systems. General Signal continued to acquire other semiconductor equipment manufacturing companies and by 1982 began to split up Tempress and transfer it's die separation and assembly products to these new acquisitions. By the late 1980's, Tempress, with it's sole vertical diffusion furnace product, was incorporated into General Signal Thinfilms company in Fremont, CA. GS Thinfilms, as it was known, ceased operations in 1991 and the Tempress product line was sold to former employees of Tempress BV.

Thermco was acquired by Sunbeam in the early 70's. By the mid and late 1980's, cash flow and profitability problems caused top management changes. In the late 80's Thermco was acquired by Silicon Valley Group in San Jose, combined with another acquisition, Anicon, and became the Thermco Systems Division of SVG.

Thermco had formed a joint venture with Tokyo Electron in the late 70's to market Thermco products in Japan as well as TEL's high-pressure-oxidation furnace in the United States. This joint venture was terminated when Thermco was acquired by SVG. TEL, then entered into another joint venture with Varian.

BTU Engineering Corporation a company founded in 1950 supplied furnaces for large industrial and medical applications. In 1956, BTU developed a high temperature alloying furnace especially designed for the electronics industry and was the first company to manufacture an air-tight pure-atmosphere furnace with a nitrogen curtain. Throughout the 50's and 60's BTU concentrated on various improvements in alloying furnaces and in 1971 acquired Bruce Industrial Controls, a manufacturer of process controllers. In 1979 BTU Engineering was acquired by Holec (USA), a subsidiary of the Dutch electronics company. Also in that year BTU introduced a Direct Digital Control system that could be retrofitted to non-DDC systems including those of other manufacturers. Bruce Division of BTU was established in 1980 to focus on wafer processing diffusion systems with it's first systems going to Motorola and Texas Instruments. In 1984 BTU formed a joint venture with ULVAC in Japan to manufacture and market diffusion equipment in Japan.

In the latter half of the 1980's process requirements demanded that temperature control limits be still further reduced. Furnace manufacturers increased the number of zones from three to five to further buffer the process zone (center zone) from end effects.

Corso-Gray and Tylan, both suppliers of mass flow controllers and gas control systems attempted to enter the diffusion market in the early 80's. Both companies exited the business by the late 80's.

A number of successful companies have captured niche markets closely associated with the diffusion market. Unit Instruments, founded by ex-Thermco engineers, has become a market leader in mass flow controllers and gas control systems. MRL Industries, founded by ex-Lindberg/Tempress employees enjoy a good market position as a supplier of replacement heating elements and other related gear.

Also in the late 80's Japanese diffusion equipment manufacturers began to gain world market share. The main players are Kokusai Electric, Tokyo Electron Limited (TEL) and Toyoko Kagaku.

It is difficult to determine when the industry perceived the need to investigate alternative approaches to the horizontal furnace configuration. However, over the last ten years, virtually every furnace manufacturer has added a vertical system to their product line. In general, interest in vertical systems came about by a growing concern that horizontal configurations might not be able to meet future process requirements, especially for larger wafer sizes, particulate control and automation of wafer loading. Vertical systems take less valuable clean room floor space and are more easily automated then are horizontal systems. More importantly, vertical furnaces can produce better process uniformity because of the capability to rotate the wafer boat during processing.

Vertical furnace configurations were first used more than 25 years ago for zone refining. It was really a German-based company, Helmut-Seier, that first applied the vertical configuration to diffusion, oxidation and CVD. Disco-Seier was later formed to market the product in the United States.

The designs of currently available systems are, in general, single tube, stand alone units. Wafers are loaded from either the top or the bottom and are held horizontally, usually face up. The wafer holder is typically made of quartz or silicon carbide and can be rotated in some systems. Gas and temperature control are similar to those used in horizontal configurations. Vertical load stations are very much smaller the horizontal designs and usually incorporate a small wafer transfer arm. Wafer process boats are loaded into the furnace vertically without touching the diffusion tube minimizing particle generation. Since the quartzware is always in tension or compression, breakage is reduced, especially when processing large wafers.

SVG-Thermco's system, for example, is a totally automated, bottom loading single tube unit that can be bulkhead mounted. It is capable of handling up to nine cassettes of wafers. These can be loaded into one of two boats, each with a capacity of 160-200mm wafers. One boat can be loaded or unloaded while the other boat is being processed. Specially designed heating elements allow process temperatures up to 1350 °C with flat zone control of ± 0.25 °C.

One potential problem often cited by horizontal furnace manufacturers about vertical furnaces is the possibility of wafer warpage, slip and dislocation caused by the wafers being held in a horizontal position. However this has not become a major issue and vertical systems continue to gain popularity.

4.4.1.1.2 High Pressure Oxidation

The origin of high pressure oxidation can be traced back to Bell Labs in 1960. J. R. Ligenza and W. G. Spitzer published several early papers describing their work. In these experiments, wafers along with fixed amounts of water were placed in an Inconel vessel, called a 'bomb'. This 'bomb' was pressurized and then placed in a furnace to achieve the desired results. Unfortunately, the terminology has caused an unfair stigmatism to be attached to high pressure oxidation. The semiconductor industry has long used pressure cookers and bombs, but dislikes the inconvenience. Nevertheless, as the industry entered into LSI and VLSI integration, it found that the low temperature, high uniformity, and high throughput aspects of high pressure oxidation made it an attractive choice for oxidation.

The market for high pressure oxidation systems rapidly gained strength in the early stages of development. However, this growth stagnated as industry searched for applications. Motorola purchased the first commercial high pressure oxidation equipment in 1977. The system, made by Gasonics, was marketed under its registered trademark 'Hi-Pox.' Fortunately for Gasonics, the term Hi-Pox soon became generic to the semiconductor industry. Today it is used for describing all high pressure oxidation systems.

The second competitor to enter the market was Thermco. Their system, called the FOX (Fast Oxidation) was developed by their Japanese joint venture partner, TEL. The first unit was delivered to Hughes in 1978.

A third competitor, Kokusai Electric, entered the market in 1979. The earliest Japanese efforts in high pressure oxidation can be traced back to the LSI development laboratory of Mitsubishi. In 1977, N. Tsubauchi and his co-workers described a dual chamber, microprocessor controlled high-pressure oxidation system. Fujitsu was also working in this area in 1977.

High temperature oxidation steps cause many of the primary non-lithography defects such as stacking faults, dislocations, metallic impurities, lateral dopant diffusion, wafer warpage, etc. Hi-Pox serves to reduce or even eliminate these sources of yield loss.

Two areas in which Hi-Pox plays an important role are for isolated CMOS and Bipolar devices. The deep oxide growths which are required for the isolation regions can take days of diffusion time in an atmospheric oxidation system. Hi-Pox reduces these times to hours, thus greatly enhancing throughputs.

4.4.1.1.3 Rapid Thermal Processing

Rapid thermal processing systems have recently developed into production-worthy tools. These systems have existed for several years. They were initially called Rapid Thermal Annealers. While there was initially much interest, the cost and low throughput of laser and E-beam systems caused interest to wane. The market stagnated as diffusion furnaces continued to be used. However, this has changed in the last several years. Today, graphite and high-intensity, halogen arc-lamp, rapid-thermal-processing-systems offer a low cost, high throughput alternative to laser annealing.

Rapid thermal processing is becoming commonplace in many device manufacturing processes. Stacking faults and dislocations of the silicon crystalline lattice have been the driving forces behind this increased use. Stacking faults, caused by oxygen and carbon impurities, occur during oxidation or during epi growth. Dislocations occur as a result of mechanical stress, thermal stress or poor lattice match of dopant atoms. Swirl defects can be caused by diffusion and ion implantation. These lattice defects cause poor refresh cycle times in DRAM's, threshold voltage drifts in SRAM's, junction leakages, lowered breakdown voltage, current channeling, lowered gain, and non-uniform power dissipation, to name just a few. Reduction of these process-induced defects can result in large yield enhancements.

Other applications for rapid thermal processors are for gettering, for damage repair combined with glass reflow, and for formation of refractory metal silicides.

4.4.1.2 Process Technology

4.4.1.2.1 Diffusion Processes

For many years, diffusion took center stage in the development of semiconductors. Diffusion scientists were the chief architects of solid state technology. In college, students of advanced physics quickly learned that principal routes to advanced research lay in solutions of problems couched in one of three equations: the wave equation, the Laplace equation, or the diffusion equation. Diffusion was a very important research field and solid-state physics was an emerging field of research where it could be vigorously explored.

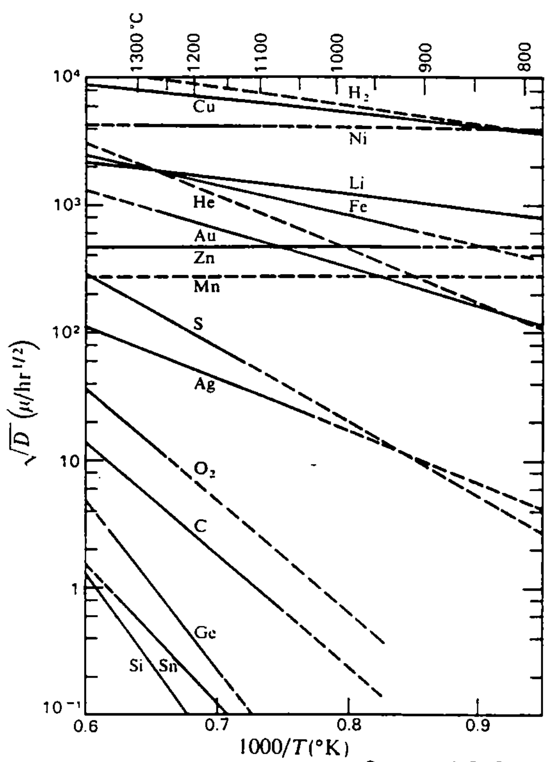
Solid-state diffusion analysis was extremely difficult, but it modeled well mathematically. The distribution of atoms diffusing into a solid as a function of time, temperature and position within the solid is described by the solution of a second-order partial differential equation called the complementary error function. This simple model was the basis for the entire theory of diffusion. In one dimension:

$$C(x,t) = C_o \, erfc \left\{ \frac{x}{2(Dt)^{1/2}} \right\}$$

where C_o is the dopant concentration at the solid surface, t is the diffusion time, and D is the diffusivity. The relationship between diffusivity and temperature is shown in Figure 4.4.1.2.1-1 for various impurity atoms in silicon. Figure 4.4.1.2.1-2 compares the complementary error function distribution with the Gaussian distribution.

Diffusion processes that are performed with a constant surface concentration are referred to as predeposition steps. Predepositions are usually done in a furnace in an nitrogen ambient with a small percentage of oxygen. The dopant species is introduced into the furnace in gaseous form and the dopant concentration in the ambient nitrogen-oxygen atmosphere is varied to control the dopant concentration at the silicon surface.

For the predeposition of boron, the most prevalent species in the gas phase in the furnace is B_2O_3. Once the B_2O_3 is deposited



Source: A.S. Grove
2244-332

Figure 4.4.1.2.1-1

Diffusivities in Silicon

on the silicon surface, a reaction takes place with the silicon to produce doping. This reaction is given by:

$$B_2O_3 + 3/2\ Si \leftrightarrow 2B + 3/2\ SiO_2$$

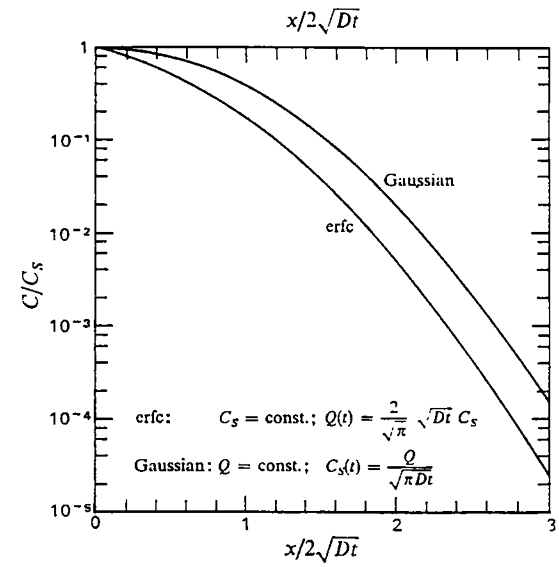
B_2O_3 can be produced by the reaction:

$$2BN + 3O_2 \rightarrow 2B_2O_3 + N_2$$

The source of boron nitride is usually in the form of disks equal to the diameter of the silicon wafer which are placed next to each wafer in the furnace. By varying the partial pressure of the gas phase of the dopants in the furnace, it is possible to change the concentration of the boron impurities in the silicon. Most predeposition steps operate with a high enough partial pressure of the dopant gas that solid solubility of the dopant in the silicon is achieved. This provides a natural control for reproducible diffusion results. Figure 4.4.1.2.1-3 shows solid solubilities for various elements in silicon as a function of temperature. Most predeposition steps are carried out in the 900-1000 °C range for 30 to 60 minutes.

Ion implantation may also be used to introduce the dopants into the silicon surface, but in this case the surface concentration usually is not high enough to reach solid solubility.

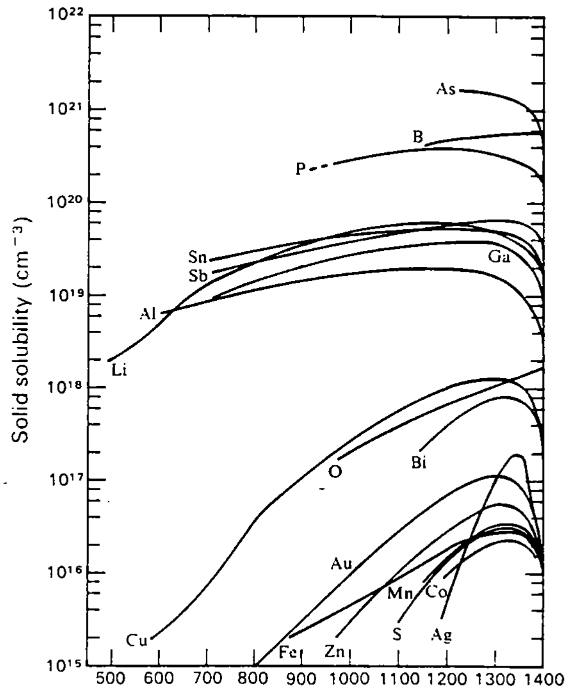
Once the predeposition is completed, the next step is to redistribute the atoms to give the desired junction depth. The distribution



Source: A.S. Grove
2244-331

Figure 4.4.1.2.1-2

Comparison of Complementary Error Function with the Gaussian Distribution



Source: A.S. Grove
2244-330

Figure 4.4.1.2.1-3

Solid Solubilities of Various Elements in Silicon

of impurities, after diffusion for a time t, is given by a Gaussian function:

$$C(x,t) = [Q/(\pi Dt)^{1/2}] \; exp \; [-x^2/4Dt]$$

where Q is the total number of dopant atoms deposited in the predeposition process. This step in the diffusion process is known as the drive-in step and is performed in the temperature range of 900-1200 °C in several types of ambients. No dopant gas is used. The most common ambients used are dry oxygen, steam, nitrogen or argon.

Impurity diffusion is normally done in a selective manner. Layers of silicon dioxide or silicon nitride are patterned by the photolighography process and the impurities are then deposited and diffused through these 'openings' in the film. Normally the drive-in process is combined with oxidation to simultaneously produce the desired dopant profile and prepare the wafer for the next process step.

4.4.1.2.2 Thermal Oxidation Processes

Thermal oxidation is the process of growing a thin silicon dioxide layer on the surface of the silicon wafer. This oxide layer is either grown on the entire wafer surface or is selectively grown on silicon in exposed regions not previously covered by other thin films. Thermal oxidation is, and will probably continue to be, the main dielectric film used in integrated circuit device technology as the industry moves into the era of submicron ULSI. The following table lists some of the major uses of these films in semiconductor manufacturing.

Uses of Thermal Oxide Films

- Component in devices
- Corrosion protection
- Device isolation
- Dopant diffusion source
- Getter impurities
- Increase breakdown voltage
- Insulate metal layers
- Mask against dopants
- Mask against impurities
- Mask against oxidation
- Mechanical protection
- Passivate junctions
- Smooth out topography

Thermal oxidation processes provide superior passivating characteristics for silicon devices when compared to films formed by various other deposition processes. For this reason, deposited dielectric films are most often used in conjunction with thermally grown silicon dioxide, e.g. deposited silicon nitride over thermal silicon dioxide. In the thermal oxidation process, silicon reacts with either oxygen or water vapor (steam) at temperatures between 600 °C and 1250 °C to form the silicon dioxide layer. The oxidation reaction can be described by the following two chemical equations:

$$Si + O_2 \rightarrow SiO_2$$
$$Si + 2H_2O \rightarrow SiO_2 + 2H_2$$

Researchers have demonstrated that oxidation processes proceed by the diffusion of either oxygen or water molecules through the oxide layer already formed. These then react with the silicon at the Si-SiO_2 interface to grow more oxide. As the process continues, the interface moves into the silicon and a new, clean silicon surface is produced. As the process proceeds, original silicon surface states (unsatisfied bonds) and contamination are consumed and surface passivation is achieved. It has been shown experimentally that in growing an oxide of thickness, t, an amount of silicon equal to 0.45 t is consumed. This fact must be carefully taken into account when the process engineer designs the specific parameters of the device manufacturing process.

Thermal oxidation is normally carried out in a fused quartz tube in a diffusion furnace. The wafers are placed vertically in slots in a quartz 'boat.' The boats usually accommodate up to as many as 200, six inch, wafers at a time. Dry oxygen processes use high purity oxygen boiled from liquid oxygen and piped into the quartz oxidation tube through an array of pressure regulators, valves, flow meters and filters.

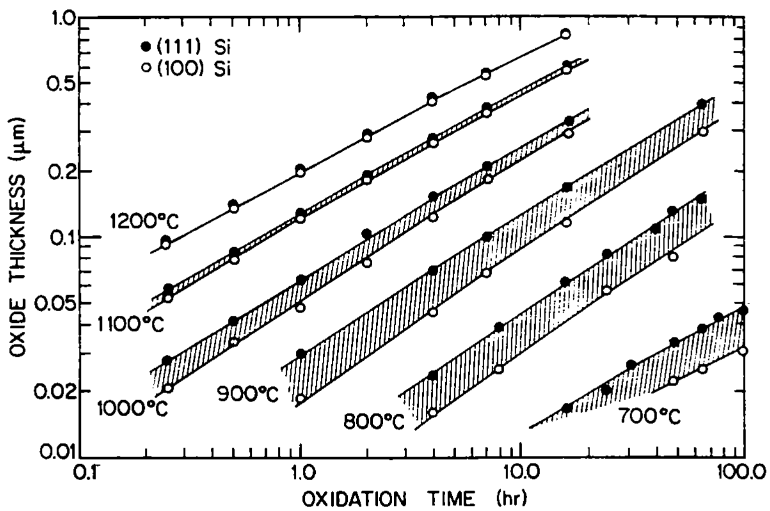
Originally, steam for 'wet' oxidation was produced by bubbling O_2 or N_2 through a flask of de-ionized water maintained at near boiling temperature. This produced a water saturated gas that could be injected into the oxidation tube. These steam sources have now been replaced by pyrogenic systems which inject H_2 and O_2 together in the proper ratio directly into the oxidation tube through specially designed nozzles. The H_2 is ignited and burned in the presence of O_2. This H_2/O_2 flame results in water vapor of very high purity. Pyrogenic systems also provide much better control of the amount of water vapor than does the bubbler system, allowing tighter control on the oxide thickness. Figures 4.4.1.2.2-1 and 4.4.1.2.2-2 show oxide thickness versus time and temperature for dry oxidation and steam processes. It can be readily seen from these figures that oxide growth proceeds much faster in steam than in a dry oxygen atmosphere. For this reason steam oxidation is normally used for thick oxides such as field or isolation and dry oxide processes are used for thin oxides, e.g. gate oxide.

Thermal oxidation of wafers, while being an essential step in semiconductor manufacturing, is also a major source of contamination. Thermal oxidation has been found to cause stacking faults. This can be mitigated by using an HCl preclean step. Chlorine 'getters' the metallic impurities which diffuse through the quartz tube from the furnace heaters.

This pre-cleaning has the positive benefit of cleaning both the wafer and the tube before oxidation. Some users are also now applying a post-oxidation chlorine anneal.

4.4.1.2.3 High Pressure Oxidation Processes

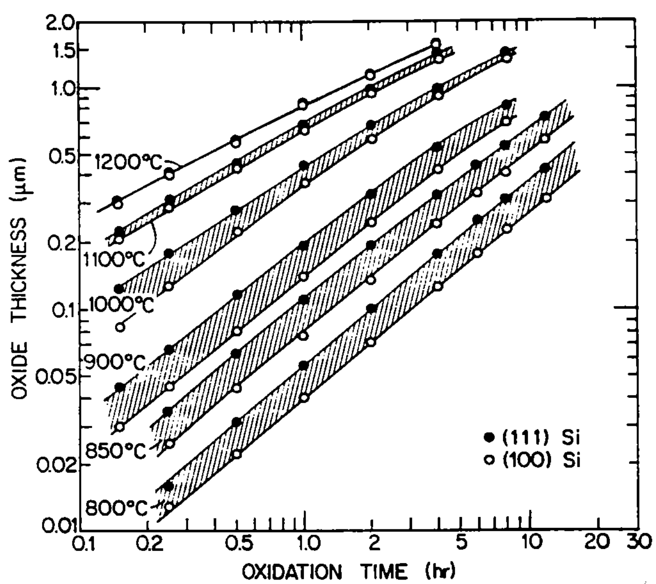
Detailed analysis of the kinetics of the oxidation process show that the oxidation rate is directly proportional to the ambient pressure. This is because the rate is directly proportional to the concentration of the oxidizer. Thus the time required to produce a given oxide thickness is inversely proportional to the pressure. Figure 4.4.1.2.3-1 gives typical oxide thickness versus time for several abient pressures for high pressure steam.



Source: Bruce E. Deal
2244-328

Figure 4.4.1.2.2-1

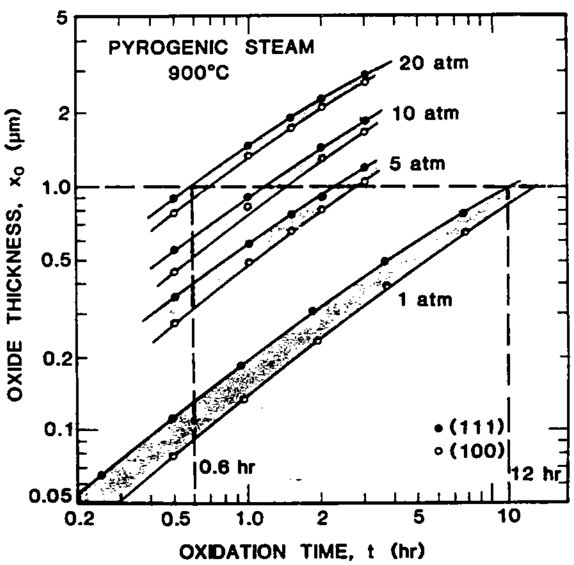
Oxide Thickness vs Time for Dry Oxygen



Source: Bruce E. Deal
2244-329

Figure 4.4.1.2.2-2

Oxide Thickness vs Time for Pyrogenic Stream



Source: Bruce E. Deal
2244-327

Figure 4.4.1.2.3-1

Oxide Thickness vs Time in Pyrogenic Stream at Various Pressures

High pressure oxidation provides several advantages, especially when dealing with sub-micron device structures with shallow junctions. Junction movement and dopant redistribution is minimized and defect levels are reduced. These improvements are due to shorter process times and lower temperatures used in high pressure oxidation or because of the increased oxidation rate.

High pressure oxidation of silicon was first employed more than twenty years ago in attempts to accelerate the oxidation process at lower temperatures. These earliest systems employed a pressure chamber in which the wafer was placed in a steam atmosphere. While oxides could be grown by this method, the process did not gain much attention until the 1980's primarily because the process was not easily incorporated into the device production process line. During the 1980's Thermco, Gasonics and others developed systems which allowed continuous flow of pressured oxygen or steam up to 25 atmospheres.

4.4.1.2.4 Annealing Processes

Many processes used in device manufacturing result in crystal damage or some other undesirable condition which, if not eliminated, would prevent the finished device from operating or, at best, would seriously degrade its performance. One prime example of this is the process of ion implantation. The nature of the implantation process results in dopant atoms being implanted into the silicon surface that are not electrically active, and generally results in a great deal of lattice damage to the silicon itself as well. These dopants must be converted to electrically active carriers and the crystal must be restored to as near-perfect condition as possible. Both of these conditions are realized by annealing. Most often, the annealing process is automatically carried out during a subsequent high temperature furnace process.

Annealing can be accomplished either as a batch process in a diffusion furnace tube or in a Rapid Thermal Processor, as will be discussed in the next section. The diffusion furnace annealing process is probably one of the simplest processes performed in that equipment. The diffusion tube is usually made of fused quartz and the gas delivery system utilizes only nitrogen. Typically the annealing process is carried out at temperatures between 900 and 1100 °C for 20 to 30 minutes. Often the annealing process is preceded by a short thermal oxidation cycle to reduce surface states and to provide a thin oxide cap to prevent out-diffusion. Since annealing processes are relatively non-critical, the furnace is often loaded to its maximum capacity of 250 to 300 wafers per batch. A single annealing tube can therefore process wafers from several ion implanters.

4.4.1.2.5 Rapid Thermal Processing

Rapid thermal processing was first investigated in the 1970's. Early work investigated the use of laser and electron beam rapid thermal processing as an alternative to diffusion furnace annealing. Low overall temperature was the driving force behind both. While the beam from these systems can heat the silicon just under the beam to very high temperatures, sometimes even exceeding the melting point of silicon, the overall temperature rise of the wafer is negligible.

There are several reasons why rapid thermal annealing using lasers or electron beams failed to become popular. The first is that the damage done by laser and especially E-beam systems may well exceed the rapid thermal processing benefits obtained. The second deterrent to laser and E-beam annealing is the low throughput rate. Laser beams scan the wafer at a rate between 10 to 30 centimeters a second, with beam sizes of about 75 micron. This translates into a writing speed that ranges from between 1376 seconds to about 458 seconds per wafer. This results in throughput rates between 2 and 8 four-inch-wafers per hour. For E-beam systems, the throughput rate is a somewhat larger 36 wafers per hour.

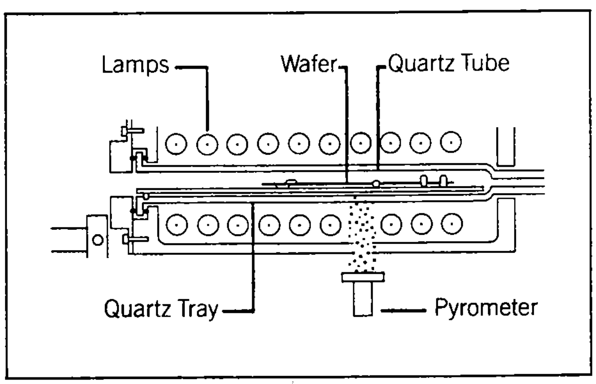
The third barrier has been price. Both laser and E-beam system prices exceed $200K. By comparison annealing can be accomplished in a diffusion furnace at 1050 °C in 20 minutes, for the price of just one tube. It is also gives an equivalent throughput of 400-600 WPH. Consequently, laser and E-beam rapid thermal processing methods have found only laboratory applications.

An additional problem with laser sources is the wavelength of the light generated. High power lasers generally operate near the ultraviolet end of the electromagnetic spectrum and are available in only a few wavelengths. Unlike electron beams, the energy of the laser beam cannot be continuously varied. Except for discrete resonance peaks, the average radiation absorption coefficient of silicon is highest in the infrared and lowest in the UV region. Generally the wavelength of available laser beams do not match the absorption resonances well. This means that very high power is required to produce the necessary annealing temperatures.

These difficulties, coupled with a need to keep wafer at the lowest possible temperatures during thermal processing cycles, gave rise to a new generation of equipment using very high intensity light sources in the visible range. Tungsten halogen, argon arc, or quartz lamps are used. These systems are referred to as rapid thermal processors (RTP). Figure 4.4.1.2.5-1 shows a typical RTP configuration.

Annealing takes only 10 seconds per wafer at temperatures as low as 700 °C. Production throughputs of well over 100 wafers per hour are possible.

Rapid thermal annealing has provided many advantages. First, it is a single wafer pro-



Source: AG Associates
2244-326

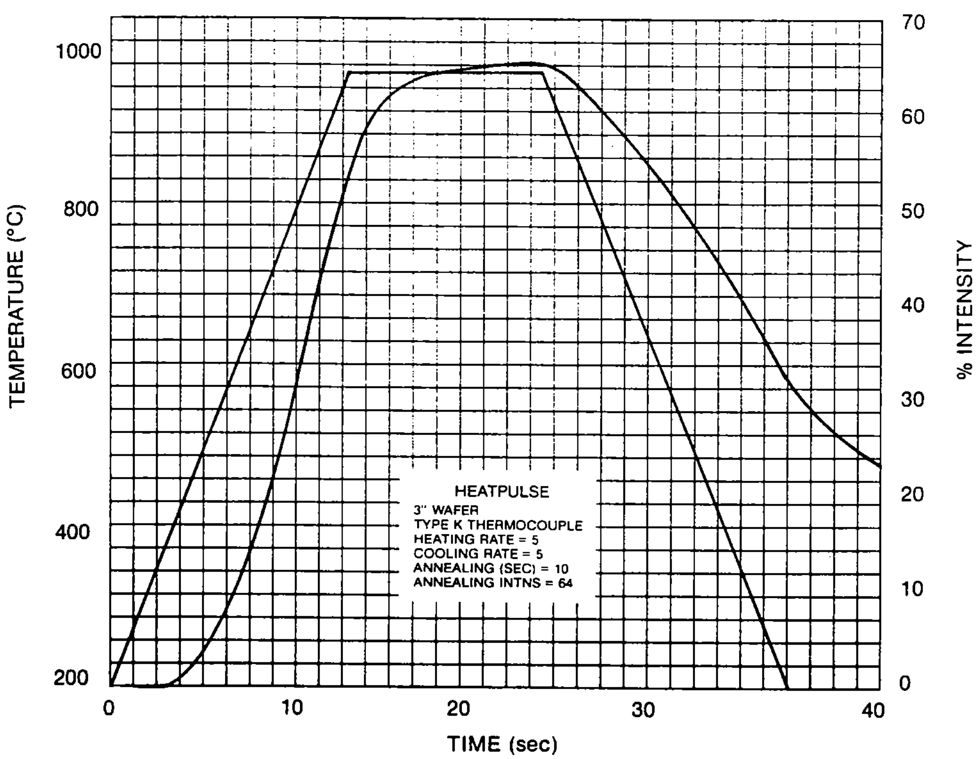
Figure 4.4.1.2.5-1

Typical RTP System Layout

cessing method. Therefore, results are repeatable and controllable. Contamination control is a key issue. Generally these methods are potentially less contaminating than furnaces. Also, rapid thermal annealing systems heat wafers faster than do furnaces. Shorter processing times mean reduced redistribution of dopants laterally and less wafer warpage. Finally, while high temperatures are achieved, overall power consumption is decreased. Figure 4.4.1.2.5-2 shows wafer temperature resulting from a typical pulse cycle.

4.4.1.3 Equipment Technology

Today's high temperature processing systems have evolved into highly automated, cassette to cassette processing systems. These systems are the basis for the manufacturing of many of the most important thin films used in semiconductor processing including oxides (doped and undoped), nitrides, and polysilicon. They also perform annealing of implanted layers, redistribution of impurities, reflow, and sintering of metal films.



Source: AG Associates
2244-333

Figure 4.4.1.2.5-2

Wafer Temperature (smooth curve) vs Lamp Intensity

4.4.1.3.1 Diffusion Furnaces

Today's diffusion furnaces are really integrated processing systems. Almost all vertical systems and some horizontal systems are cassette to cassette. Complete cassette to cassette automation of horizontal systems require complicated material handling and for this reason automated wafer handling has not attained the popularity that it has in vertical systems.

Reduced to basics, horizontal and vertical diffusion systems are essentially identical. Both utilize multi-zone (three or five zone) heating elements, and both use identical temperature controllers. Both are controlled by the use of thermocouples placed just outside the quartz diffusion tube as well as profile or process control thermocouples placed on the wafer boat inside the process tube itself. Both systems utilize identical gas distribution systems or "gas panels." Major differences lie in loading, wafer processing quartzware, and in some cases, in the maximum wafer load.

Another major difference, at least for standalone vertical systems, is due to the fact that large tube scavenger systems are no longer needed to isolate tubes from each other. All of these facts combine to allow the design of vertical diffusion systems that take up much less floor space then is needed by horizontal systems. Four separate, standalone vertical systems can occupy approximately 40% of the floor area needed by a single four tube horizontal systems.

Diffusion furnace temperature controllers utilize a three mode (PID) control algorithm. The P–or Proportional term–refers to the term in the algorithm that causes an amount of power to be applied that is proportional to the difference between the actual temperature and the desired or setpoint value. The I–or integral term–describes the total thermal characteristics of the diffusion system. This term can be thought of as a description of the system's ability to absorb heat. D refers to the derivative term in the control algorithm. This term calculates the rate of approach to the setpoint temperature and is useful in preventing overshoot. It is often called the damping factor.

Each zone of the furnace has its own controller. Actual temperature feedback to the controller comes from a thermocouple placed in close proximity to the diffusion tube. Many diffusion controller designs use an extra set of thermocouple per zone for comparison and another set within the controller for room temperature error correction. Still others use one additional set of thermocouple within the process tube itself. This internal thermocouple set feeds another temperature control algorithm which is essentially the I and D functions of the three mode algorithm. The output of this two mode algorithm acts as a small correction to the primary algorithm and allows the temperature controller to react to small changes within the process tube which would be undetectable by the external thermocouple. Utilizing these control techniques and careful design of the heating element and process gas injection systems, furnace designers are able to provide process temperature control to better than $\pm 0.25\,^{\circ}C$ under actual process conditions.

These sophisticated temperature control algorithms are easily programmed into modern microcomputer controllers along with process receipt information for gas flow, etc. for precise diffusion process control. Since a given set of coefficients of the P, I and D terms in the control algorithm are appropriate only over a fixed temperature range, today's DDC systems provide for different sets of coefficients to be used as the furnace passes through different temperature ranges. Thus, more precise temperature control and stability is available in today's DDC systems than was ever possible in hardware logic controllers of a few years ago.

Diffusion furnaces can have three or five independent temperature control zones. In either case the center zone is the process area while the outer zones act as buffers between the center of the furnace and the room. Since each zone must have its own temperature controller and thermocouple, five zone systems are more expensive than are three-zone systems and can be less reliable. For this reason, most furnaces sold today are of the three-zone configuration.

Typical process temperatures range from 400 °C to 1250 °C. At the upper range of these temperatures, quartz begins to soften and flow. As a result, silicon carbide is sometimes used as a replacement for quartzware for high temperature processing. However this material has not been widely accepted throughout the industry because of cost, concerns over contamination, and the fact that the use of larger wafers has driven processing to lower temperatures.

Gas distribution systems reflect the process engineer's demand for process safety, process control, reliability, and reduced contamination. Gas systems utilize rapid-start mass flow controllers assuring that setpoint is reached almost instantaneously after the gas is turned on. Some systems utilize mass flow meters in series with mass flow controllers to monitor or automatically calibrate the mass flow controller. Internally polished stainless steel tubing and point of use filtering is standard, as are welded connections. For those connections that must be broken, metal ring (VCR) seals are used.

A great deal of attention has recently been given to 'ultra-clean gas systems.' Special fabrication and assembly techniques are used for sub-components as well as total systems assembly to assure compatibility with sub-class 1 process environment. A typical gas control and distribution system for a single process tube can cost $15K to as much as $100K.

Diffusion processes have historically been high-contamination, dirty processes for several reasons. First, rubbing quartzware together tends to produce large amounts of particles. This has been reduced by the use of cantilevered soft landing systems in which the boat load of wafers is suspended as it is placed into the process tube, avoiding a sliding contact. The boat is then gently lowered and comes to rest on the bottom of the process tube, removing the stress from the cantilevered loader.

The second major source of contamination results from convective flow of outside air through small openings in the tube end closure door. Early furnaces were simply closed off manually by slipping an endcap into a ground glass mating joint. However, automatic boat loader/door closure devices generally do not seal the tube end effectively and air can easily be drawn into the process tube.

Another source of contamination is due to the fact that certain elements easily diffuse through quartzware at high temperatures. To reduce this problem, often double walled diffusion tubes are used. The space between the walls is continually flushed with nitrogen or sometimes with HCl, thus preventing the contaminating species from reaching the wafers.

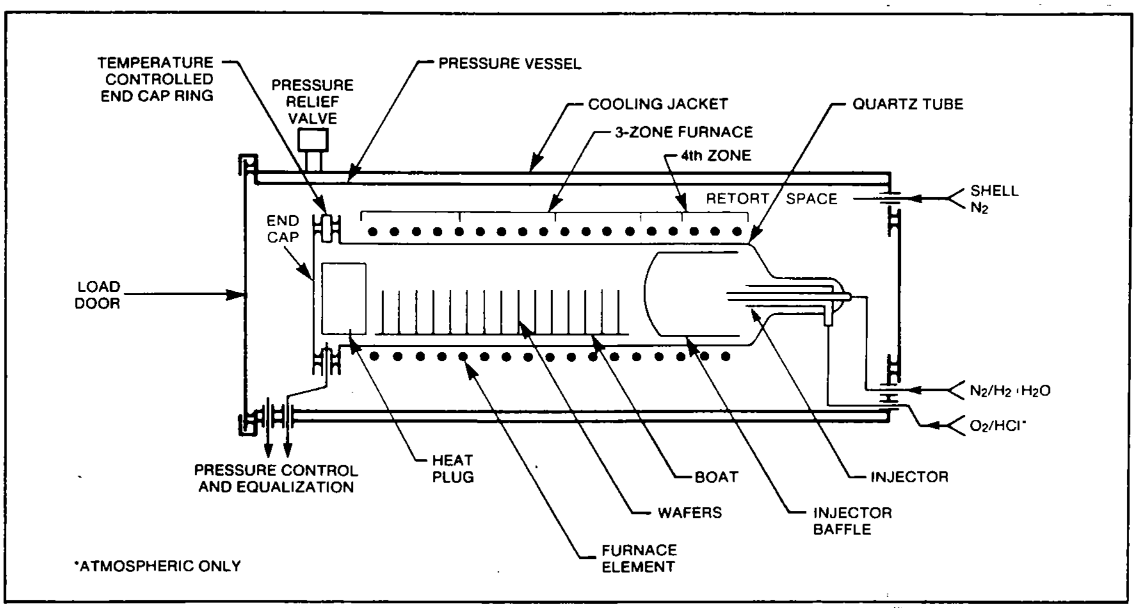
Automating the transfer of wafers from process cassettes to quartz boats was a relatively difficult task. This was due primarily to the fact that quartzware dimensions are not stable under high temperature processing. Today's systems, however, incorporate a high degree of automation. Robotic arms move individual wafers from cassettes to boats as well as moving loaded boats into place on the soft landing system.

Vertical diffusion systems are much more easily automated since wafers can more easily be loaded directly into the process boat and the process boat can be pushed into the process tube without touching the tube walls. This configuration also allows smaller load station areas since the entire wafer boat and paddle net not be withdrawn into the load station. Typically, vertical diffusion systems are loaded from the bottom such as those built by Thermco, Varian (Tokyo Electron) and BTU. Disco utilizes top loading. Each configuration has its admirers. Bottom loader champions claim lower particulate generation while top loaded advocates claim their design allows better insulation capability resulting in improved temperature stability.

At least one vendor of vertical furnaces offers an optional vacuum load lock to eliminate oxygen and water migration into the process tube during loading and unloading cycles. Contamination levels below five parts per million are claimed. Controlling oxygen and water during thin gate oxidation or annealing is critical to controlling Q_{ss}.

4.4.1.3.2 High Pressure Oxidation System

High Pressure or Hi-Pox systems are evolutions of standard diffusion systems. Fundamentally a Hi-Pox system contains all the same basic elements of a normal oxidation diffusion furnace including temperature controllers, gas systems, quartzware, etc. However, Hi-Pox systems, operate at pressures of several atmospheres. Essentially an entire diffusion tube is placed inside a special pressure vessel and the vessel is pressurized with oxygen. Hi-Pox systems are used for growing thick field oxides at lower temperatures than would be necessary with standard oxidation systems. Figure 4.4.1.3.2-1 show a drawing of a typical Hi-Pox layout.



Source: GaSonics
2244-305

Figure 4.4.1.3.2-1

A Typical Hipox Pressure Vessel

4.4.1.3.3 Rapid Thermal Processing

RTP systems have gained popularity in recent years. This has been due to several factors. It has long been known that unwanted stable defects can form via oxygen--vacancy combination at certain temperatures. Because the temperature ramping capability of RTP systems are of the order of 250 °C per second, wafers are not exposed to certain intermediate temperatures for long periods of time during temperature ramp up and ramp down. As a result these defects are reduced. Also wafer warpage is reduced because the wafers are exposed to the highest temperatures for only the process time which is generally of the order of seconds. These processes are known as isothermal processes. RTP systems are usually single wafer cassette to cassette systems and as such usually exhibit better wafer to wafer process uniformity. Figure 4.4.1.3.3-1 lists several advantages of RTP.

Another major benefit of RTP systems is throughput. An RTP system can anneal an implanted wafer in a few seconds compared to 20 to 30 minutes in a diffusion furnace. In fact, early uses for RTP systems were for production control. Rapid feedback for an ion implant process step could be obtained much more quickly with RTP. Implanted layers that are annealed by RTP, are not diffused. Therefore, very shallow and steep junction profiles can be formed.

These systems utilize very high intensity infrared or visible (near UV) radiation sources for providing intense heat at the wafer surface. The source can either be continuous or pulsed with pulsed versions being most popular. This is mainly due to better control of temperature by simply varying the pulse duty cycle. One difficulty, however, is the need to produce uniform irradiation across the entire wafer. In some systems the wafer is heated from both sides while in others it is heated only from the top. Usually several lamps are used and arranged to provide the best uniformity possible. Since the rate of change of temperature in an RTP system is so great, thermocouples are useless and optical pyrometers are used. Great care must be taken to calibrate the pyrometer with emissivity of the wafer in the process in order to assure accurate temperature control.

RTP processing can be carried out either at atmospheric pressure or in a vacuum. Most systems utilize vacuum processing because the process atmosphere can be better controlled and more quickly stabilized. Also since these systems are usually single wafer systems, the process chamber is small and the vacuum system becomes relatively inexpensive. Since the chamber walls are colder than the wafer during processing, there is potentially less contamination to the wafer than in a diffusion process.

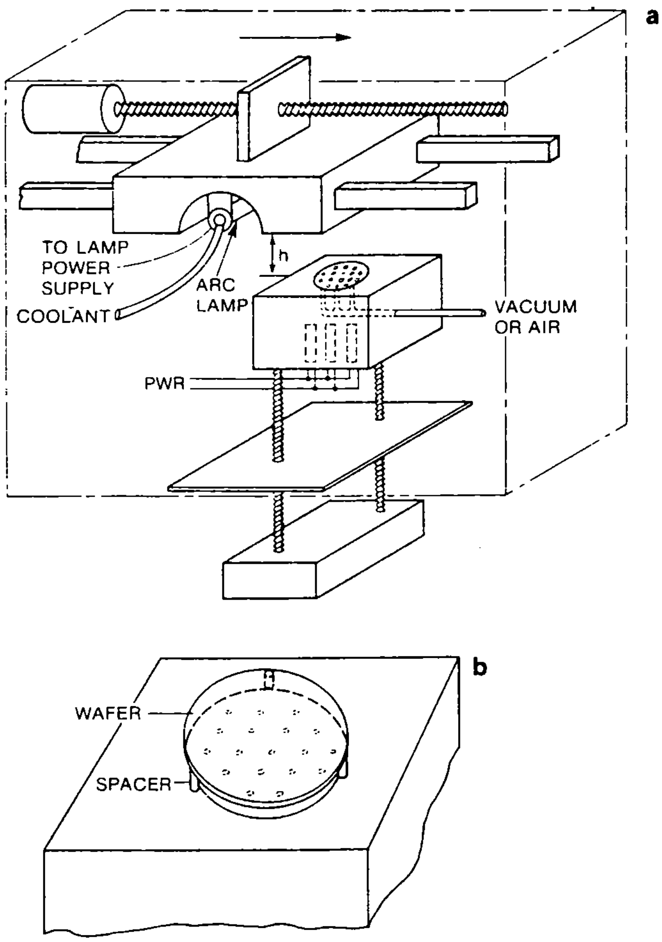
RTP is being investigated for growing very thin (e.g. 100 Angstrom thick) uniform oxides, BPSG/PSG reflow processes, as well as some CVD processes including polysilicon. Figure 4.4.1.3.3-2 shows a drawing of a rapid thermal annealing system.

TABLE 4.4.1.3.3-1

Advantages of RTP

- Shorter Processing time
- Higher temperature cycles
- Cold wall process
- Small processing chamber
- Efficient and rapid gas switching
- Multi-step temperature/gas processing

Source: VLSI RESEARCH INC
2244-334W



The annealing apparatus. (a) Chuck and scanning-lamp driving mechanism. (b) Heated chuck with wafer mounted on ceramic spacers.

Source: EDL, VOL. EDL-2, NO. 4 April '81, P.85

2244-306

Figure 4.4.1.3.3-2

Rapid Thermal Annealing System

Notes