
4 Planar Technology

4.1 INTRODUCTION

This chapter is the first of the second part of the book, the part devoted to technologies for labs on chip. As a matter of fact, what renders microfluidics revolutionary in the field of biotechnologies is the possibility to produce millions of inexpensive circuits that perform complex operations.

This opens the opportunity for capillary diffusion of disease prevention and screening on the one hand, and environmental and food control on the other, with a potential social revolution even more profound than that caused by microelectronics and broadband communications.

This is due mainly to the transfer of mass production technologies in the field of microfluidics, to transform the cost model of analytical biochemistry from the current situation dominated by work and material costs to a microelectronics like situation, where extremely larger volumes correspond to lower costs [1,2].

The process allowing integrated circuits, which are electronic, optical, or fluidics [2–5], to be realized is essentially constituted by two steps called front end and back end. All the techniques required to fabricate several different circuits on a single substrate are collected under the name of front-end technologies. Thus, the output of a front-end fabrication cycle is a set of circuits fabricated on the same surface (a single wafer in case of silica on silicon).

All the procedures starting from the wafer in which integrated circuits have been fabricated and that arrives to a single integrated circuit enclosed in a suitable package and ready to be integrated into a more complex system are collected under the name of back end technologies.

Thus, technologies such as photolithography and dopant diffusion are part of the front end while the process of separating chips fabricated on the same wafer (generally called dicing) and of fabricating the suitable interfaces with the outside world are part of the back end.

As far as front-end processes are concerned, essentially two types of technologies are used to target mass production of labs on chips, depending on the kind of used material: planar technologies, essentially derived from microelectronics, and polymer-based technologies. These technologies have many common elements, in particular, the capability of processing chips with very small critical dimensions with automatic machinery, but have also relevant differences.

After a general introduction, we will devote this chapter to planar front-end technologies and Chapter 5 to polymer-based technologies. We will include in Chapter 4 a paragraph on cost models for different production technologies, so as to clarify the impact, the potentialities, and the limits of the different manufacturing technique. Back-end technologies are analyzed in Chapter 6, since they are common to both planar and polymer-based circuits, but for a few elements that will be detailed during the exposition.

As we will see, the very small aspect ratio of integrated circuits is instrumental in their low-cost characteristic, thus complex technologies have to be used to be able to fabricate the small details of the circuits and the presence of even very small impurities can completely ruin a whole circuit. For this reason, planar technology and polymer-based processes have to be carried out in a very clean environment, where the presence of impurities, both from the air and from the very presence of human operators, has to be strictly controlled. Such an environment is realized in the so-called “clean rooms” [1], the typical structures where planar integration is realized.

Clean-room technology is instrumental to the realization of integrated circuits of whatever kind and they condition the characteristics of the available processes. Thus, a brief introduction to

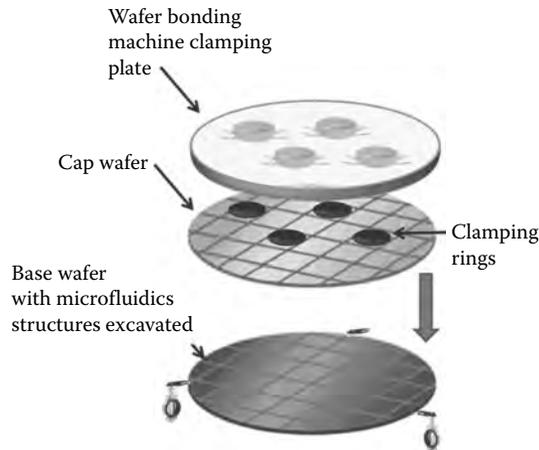


FIGURE 4.1 Wafer bonding for microfluidics circuits.

clean rooms structure and functionalities is important to a full understanding of planar integration processes.

While the above considerations hold for whatever integrated circuit fabrication process, microfluidics has its own particular technologies and processes. On the one hand, if silicon and silicon oxide films on silicon are obliged materials for electronics and integrated optics, this is not true for microfluidics.

Even if microfluidics can be developed starting from traditional silicon wafers designed for microelectronics by exploiting the natural passivation of silicon surface by oxidation and, if needed, deposition of thick silicon oxide films, both polymeric materials and glass wafers have been used with success, thereby creating a set of specific processes for these kinds of materials. Moreover, in creating microfluidics circuits, exigencies that do not exist in microelectronics and integrated optics emerge, calling for specific processes or important modification of existing fabrication techniques.

As an example, since the fact that planar technology is able to excavate the wafer from the top, not to create ducts and chambers excavating it from the side, a duct has to be fabricated by creating a long hole in the wafer and then covering it with a roof. The roof cannot be obtained by deposition that will fill the hole, but has to be fabricated by placing on top of the hole a solid surface, that is another wafer (cf. Figure 4.1). Consequently, a technology called wafer bonding [6], consisting in placing two wafers one on top of the other with the top surfaces in contact is fundamental in the field of microfluidics, while it is a niche technology in the field of electronics and optics.

4.2 PLANAR PROCESS FLOW OF A LAB ON CHIP

The goal of this section is to present the whole fabrication process needed to build a microfluidics-based lab on chip with planar technologies [1,2,7]. A realistic fabrication process strongly depends on the structures to be implemented in the chip and in principle it is not possible to provide a generic process flow. Nevertheless, there are recurring operations and typical processes which have to be considered as steps of a complete fabrication cycle to be fully understood. Frequently, the most difficult challenge in fabricating a complex circuit is not to fabricate its individual parts, but to integrate them together in the same process flow. As a matter of fact, different processes require different conditions and it is not infrequent that performing process B after process A destroys the structures created by process A due to the conditions required to perform process B.

The most classic example is doping and annealing. To fabricate almost all the electronic structures doping is needed. This is a process that is required to give to silicon electrical characteristics suitable for electronic devices and consists in controlled diffusion of particular substance called

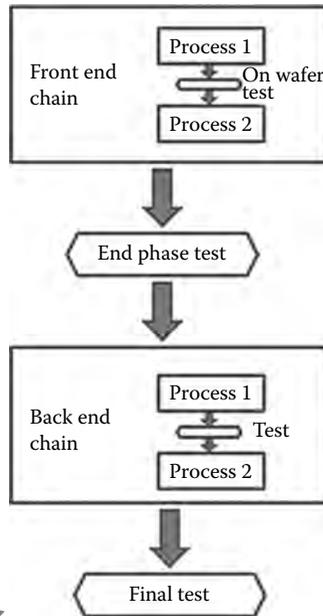


FIGURE 4.2 General scheme of the whole fabrication process for a microfluidics circuit.

dopants in the silicon crystal. Dopants generally substitute silicon atoms at the lattice nodes penetrating due to their mobility and high-temperature deep into the crystal [1,3]. Since integrated optics is generally constructed in a silica layer deposited over the silicon surface, it could seem possible to implement electronics on the silicon wafer, grow the silica layer, and then integrate optics in the silica.

To obtain good optical properties in terms of index uniformity and attenuation, the silica layer has to be annealed around 1000°C to uniform it before the optical waveguide fabrication [5]. Annealing the wafer at a high temperature completely destroys the underlying electronic structures, since it causes uncontrolled diffusion of the dopants. This is one of the major obstacles to the integration of electronics with silica-based optics and is still unresolved in single wafer circuits.

The case of dopant and annealing is quite extreme, even if practically important, but similar situations are common. Thus clarifying the need of looking at the fabrication process like a whole, instead like a set of sequential independent steps.

The chip production process is constituted by three main parts, as shown in Figure 4.2. Testing is a very important part of the process, having a great impact both on the end process quality, represented at first by the final product yield, and on the industrial cost. Under a cost point of view, testing is more expensive if located in the back-end section of the process, being performed chip by chip. The final test, in particular, is often one of the main contributions to the industrial cost due to the impact of testing time and labor cost. For this reason, in a steady-state process characterized by a very high fabrication yield, as it happens, for example, in electronic RAMs [3], a sampled final test can be operated; drastically decreasing the related cost.

4.2.1 FRONT-END PROCESS FLOW

The most basic procedure of planar technology is the series of steps that enable to transfer a planar shape designed on a mask on a wafer by modeling a film of the wanted material. This procedure is so common and important that often the complexity of a planar circuit is measured in “number of masks,” that is, the number of consecutive times this procedure has to be repeated transferring on the wafer different superimposed forms to achieve the wanted structure.

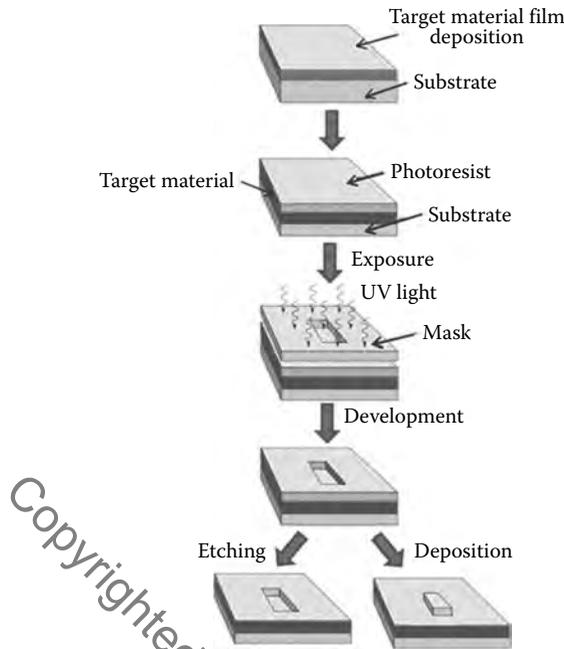


FIGURE 4.3 General scheme of the process of deposition, lithography, and etching (or deposition).

The procedure for the creation of a planar shape on the wafer is composed of the following steps, represented graphically in Figure 4.3.

1. The desired shape is created on a mask, that is, on a sheet of material suitable for the following process. The shape has to be planar, that is, two dimensional, so as to be transferred on the wafer surface. Nevertheless, a depth is also associated to it, constant over all the shape. An example of mask is shown in Figure 4.4.
2. The material film over which the shape has to be transferred is growth on the wafer surface with one of the many growing techniques. The depth of the film will also be the depth of the shape that has to be realized.
3. A photoresist is deposited on the film; it is a material that will change suitably its chemical properties when exposed to ultraviolet light.
4. The mask is superimposed to the resist and aligned exactly to the wafer underlining structure. This is a critical operation due to the fact that the structure that has to be realized is to

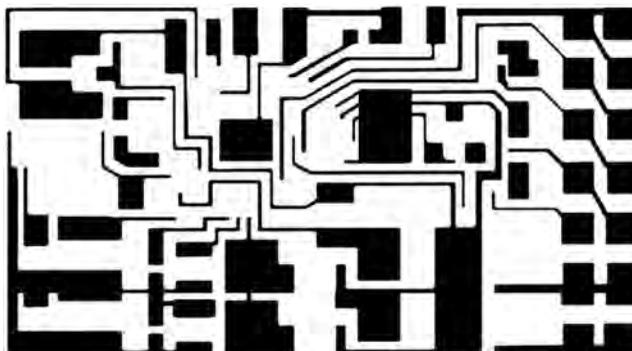


FIGURE 4.4 Example of photolithographic mask.

- be aligned with the already existing structures. In order to do that, when a particular structure is realized, also specific markers are present to allow subsequent structures alignment.
5. The wafer is lightened with ultraviolet light. Where the photoresist is protected by the mask it is not hit by light, while where the mask is not present it is impressed by it.
 6. The mask is removed and the photoresist developed. Where the photoresist was impressed, its chemical structure changes due to the light and the development procedure.
 7. The wafer with the photoresist on it is chemically attacked by a photoresist etcher. In positive photoresists, the developed part is etched while the undeveloped part, which was protected by the mask, is not etched, while the contrary happens in negative photoresists. In this way, either the mask image or its negative photoresist is transferred to the resist.
 8. The wafer is etched again with an etching process suitable to consume the film that was deposited at the start of the process, where it is exposed. Where the photoresist is still present the etching process does not reach the film and it is not consumed. In this way the image on the mask is transferred, via the photoresist, on the film that was grown on the wafer surface.
 9. All the photoresists are chemically removed, by leaving the mask image transferred on the film, which was the desired result.

A possible variant of the process is substituting etching with deposition so that the holes left from the photoresist development and etching are filled by the new material and creates an image of the mask in relief with respect to the wafer surface.

Independently from the use of etching or deposition, a set of subprocesses are required: they are called

- Mask design
- Deposition
- Lithography
- Etching or deposition

These are the base processes of any planar technology, from microelectronics to integrated optics to microfluidics. As a first example, we want to present the schematic process flow that is needed to integrate silica on silicon optical waveguide, an optical element that is frequently present in labs on chip when optical detection is used.

The waveguide structure is presented in Figure 4.5: on the silicon substrate provided by a standard silicon wafer several silica layers with different diffraction indexes are fabricated, obtained by doping silica with different quantities of selected dopants, generally germanium [5]. The process

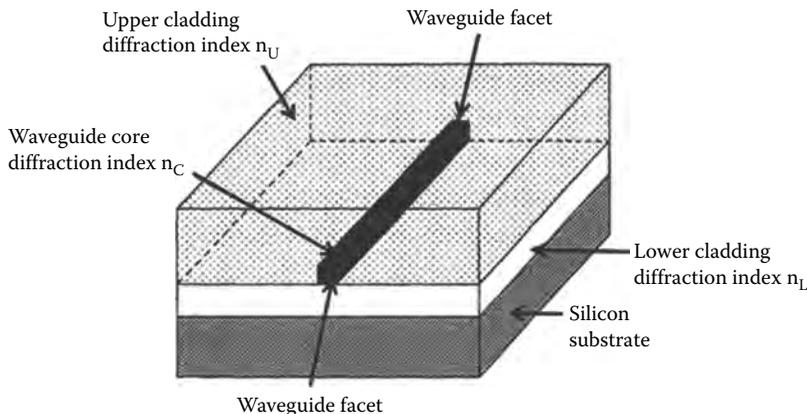


FIGURE 4.5 Scheme of a silica on silicon optical waveguide.

does not substantially change if the substrate is a glass wafer. The schematic process flow is shown in Figure 4.6. Here, the base operations that we have described in the above discussion can be recognized. Besides the basic processes of deposition, lithography, and etching, two new processes are shown, that are needed for this particular component. The first is silica annealing, as we have already discussed, to homogenize the property of silica and obtain good optical performances.

The second new process is planarization. This process is needed every time a planar surface has to be obtained after deposition on a surface where structures have been fabricated [1]. As the surface is not uniform and the deposition creates a uniform thickness film, the deposited film will follow somehow the profile of the underlining surface. To obtain a flat surface, a *flattening etching* has to be performed, that is called planarization. This is a key process in microfluidics devices, since a very good planarization is needed to achieve effective wafer bonding.

Let us now imagine to fabricate a more complex microfluidics circuit. We have to build a reaction chamber in which antibodies have to be trapped on the bottom surface to perform a sandwich immunoassay (cf. Section 3.7.5) having as target the solution that will be injected into the chamber.

Detection is performed by building an optical interferometer on top of the reaction chamber so to detect the refraction index change due to antigen–antibody neutralization. The planar scheme and the section of the structure we have to fabricate are represented in Figure 4.7. This is a very simplified scheme with respect to the realistic labs on chip design that we will discuss in later chapters, but here the point is not to show an effective design, but to define a simplified structure so as to imagine a fabrication process flow.

We have to work on two wafers. The first on top, we will use as the roof of our reaction chamber, where we will build the optical waveguide part of the detection interferometer. This wafer will need annealing to assure good optical performances, but since no other structure will be constructed upon it, this hides no problem. In any case, planarization will be needed after the annealing to assure an effective wafer bonding. On the other wafer, we will build the structures that will be needed for the reaction chamber, excavating it and constructing the structure that will capture the

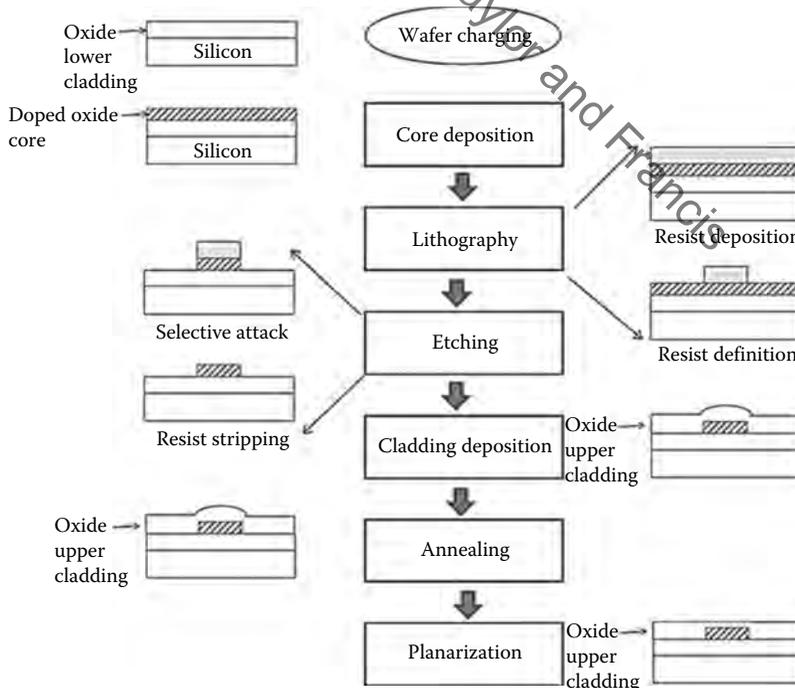


FIGURE 4.6 Schematic process flow for the silica on silicon waveguide fabrication.

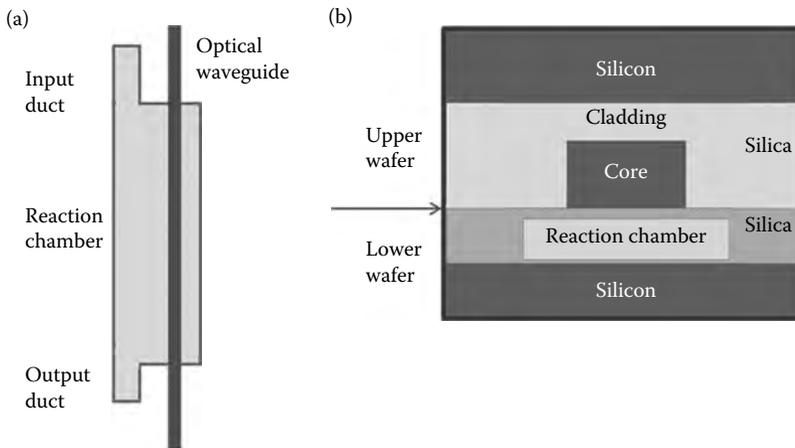


FIGURE 4.7 Planar scheme (a) and section (b) of the structure comprising an optical waveguide built upon a reaction chamber having antibodies fixed on the floor that is considered in the example of the text.

antibodies on its floor. This wafer will not be annealed, not to ruin the excavated reaction chamber. In any case, since etching will create inhomogeneous wafer surface, also in this case planarization will be performed before wafer bonding.

Antibodies will be added before wafer bonding by exposing the lower wafer to concentrated antibodies solution, waiting for the saturation of the reaction chamber floor and washing out the remaining antibodies with a suitable washing solution. Since the organic molecules will be bonded to the floor of the chamber from now on, we have to care that no high- or low-temperature processes are performed after antibodies bonding, not to cause molecules denaturation or detach from the activated surface.

Collecting all these considerations we can set up the process flow that is represented in Figure 4.8. From the figure it could seem that both the wafers experience a single mask process, but it is not true. As a matter of fact, we have synthesized the surface activation as a single step, but it is a quite complex procedure. There are several surface activation techniques; oversimplifying the process we have to follow the steps reported in Figure 4.9.

In this case, the mask is used to perform selective deposition on the floor of the reaction chamber and to protect the surface of the other parts of the circuit from the chemical attack that is needed to activate the reaction chamber surface. Thus, the process that is needed to fabricate the lower wafer is a two mask process and it renders it more complex. The two masks have to be suitably aligned not to create deposition and chemical attack in wrong places without activating a part of the reaction chamber floor. The mask alignment is obtained by lithographic marks that have to be realized on the wafer surface besides the integrated circuits and that are read by the lithographic machine when it places the mask.

Up to now we have considered processes that are able to transfer on the wafer surface pure planar structures; however, in labs on chips, more complex structures have to be realized. Suspended membranes, for example, are fundamental elements in many labs on chips both in microfluidics valves and in micro-pumps and movable surfaces are used in many biological applications (see Chapter 8). All these structures can be classified under the category of micro-electro-mechanical structures (MEMs) and when they are applied to labs on chip they are frequently called biomems [8,9].

Owing to the large use in microelectronics and to its very good mechanical properties, silicon is the main material used to fabricate MEMs, but sometimes other kinds of materials are also used, especially when a glass or polymer substrate is adopted. There are essentially two techniques that can be used to realize optical switches with silicon-based MEMs: bulk micromachining, and surface micromachining [10,11].

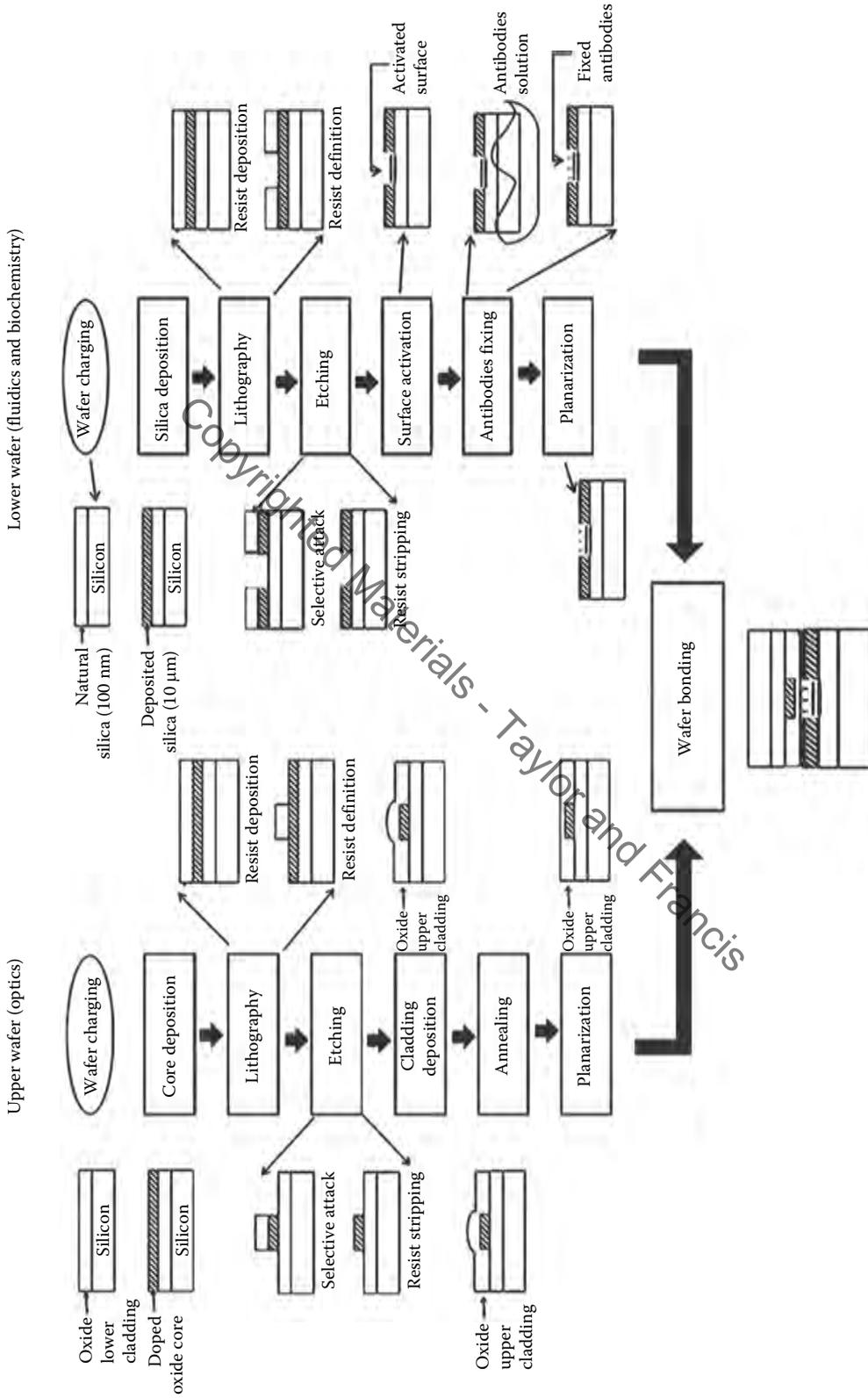


FIGURE 4.8 Process flow for the fabrication of the structure in Figure 4.7.

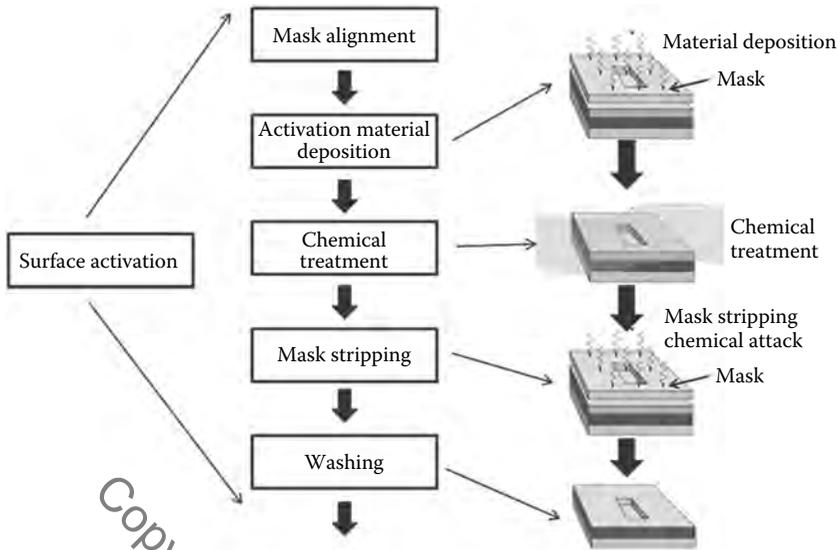


FIGURE 4.9 Details of a simplified process for the reaction chamber surface activation.

Bulk micromachining is the most mature and simple micromachining technology; it involves the removal of silicon from the bulk silicon substrate by etchants. Depending on the type of etchants that are used, etching can be isotropic, that is, oriented along one of the silicon crystal planes, or anisotropic, that is, oriented along a plane different from the crystal main planes.

Anisotropic etching is generally used to expose the silicon crystal planes in order to build non-planar structures like descendent ducts. In particular cases, these can be useful to exploit a small prevalence due to gravity without building a micro-pump or to devise three-dimensional microfluidics networks (see Section 8.4). Using bulk micromachining, the need of complex planar processes is reduced to the minimum, but only the simplest structures can be created.

Surface micromachining is a more advanced fabrication technique. It is based on the use of layers of sacrificial materials, that are destined to be etched out so to leave empty spaces between structural layers. An example is reported in Figure 4.10, shown as a sacrificial material layer is removed in order to create a suspended lever, often called a cantilever. The process is quite effective in obtaining complex structures, but it is also quite complex. The simple process sketched in Figure 4.10 requires two masks and it is only a small part of the fabrication of a practical MEMs.

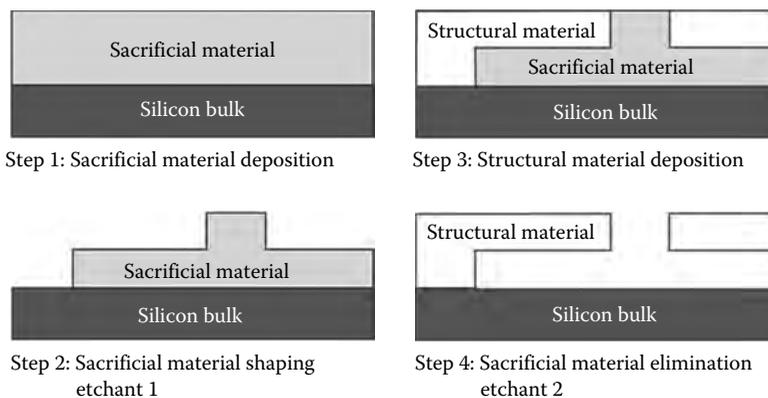


FIGURE 4.10 Schematic explanation of surface micromachining process.

If parts of a MEM have to move one with respect to the other, an actuation mechanism has to be fabricated in the circuit. The actuation can be obtained through different mechanisms, for example, electrostatic or electromagnetic forces. Electrostatic forces involve the attraction forces of two oppositely charged plates. The key advantages of electrostatic actuation is that it is a very well consolidated technology, applied in several fields in which MEMs are used like sensors, thus quite frequent in the biological field. The disadvantages include nonlinearity in force versus voltage relationship, and requirement of high driving voltages to compensate for the low force potential. Electromagnetic actuation involves attraction between electromagnets with different polarity. The advantage of electromagnetic actuation is that it requires low driving voltages because it can generate large forces with high linearity. However, the isolation of the different actuators inside the same switch is not easy and it requires a very accurate design. Different types of nonelectrical actuators exist, like pneumatic or chemical actuators, each with specific characteristics. A much more detailed discussion on MEMs actuators will be carried out in Section 8.2.2 considering microvalve design and in Section 8.3 considering micropumps design.

To provide an example of the complex process needed to realize real MEMs, in Figure 4.11 the process needed to build a single suspended membrane with piezoelectric actuators allowing it to curve or to relax is sketched. We will see in Chapter 8 that it is a key process in the fabrication of several microvalves. This is a 4-mask, 12-step process whose steps can be summarized as follows [9]:

- Step 1. p-Type wafers are cleaned.
- Step 2. Thermal oxidation to grow 200 nm oxide.
- Step 3. a. Photoresist is coated on the front surface of the samples and patterned.
b. The back side of the wafer is coated with wax to prevent etching of back oxide.

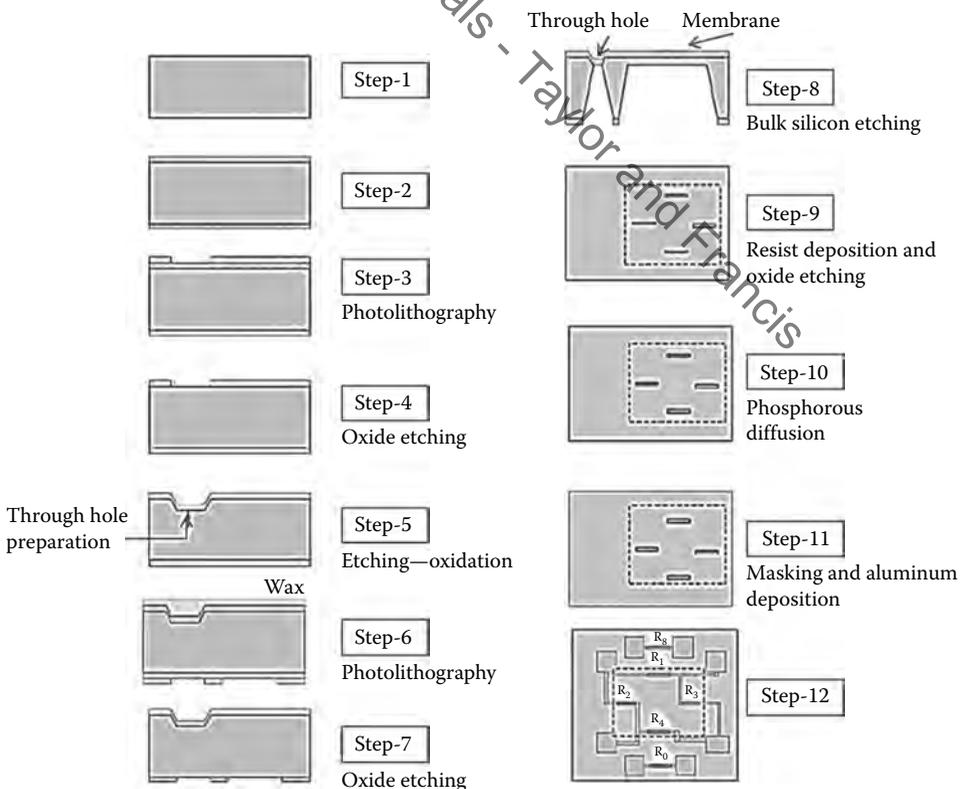


FIGURE 4.11 Possible process for the fabrication of a MEMs suspended membrane with piezoelectric actuators.

- Step 4. The front oxide is etched.
- Step 5.
 - a. Etching of bulk silicon on the front side to obtain a 10 μm deep structure.
 - b. The wafers are cleaned.
 - c. The wafers are thermally oxidized to grow 1 μm oxide.
- Step 6. Photoresist is coated on the back surface of the samples and patterns are transferred with the help of alignment marks.
- Step 7. The back oxide is etched.
- Step 8. Etching of bulk silicon from back side is carried out in a two-step process.
- Step 9. Photolithography is done to transfer the patterns of piezoresistors on thermally grown oxide on the front side of the wafer. This is followed by oxide etching. The photoresist is removed followed by cleaning.
- Step 10. Phosphorous diffusion is done to realize n-type piezoresistors.
- Step 11. The native oxide is removed. This is followed by thermal evaporation of aluminum on the wafers to form the contacts of piezoresistors.
- Step 12. The patterns for aluminum contacts are transferred through lithography and this is followed by aluminum etching. The photoresist is removed followed by cleaning. The wafers then undergo an annealing step.

4.2.2 BACK-END PROCESS FLOW

Although the front-end process is operated by very costly equipment, it works in parallel on all the chips fabricated on the same wafer. Back-end processes are almost all operated chip by chip, thus being critical under a cost point of view. Moreover, back-end processes are more influenced by the cost of materials acquired by suppliers, like the package, the cooling system, and so on. Thus, optimizing back-end processes is a key point in the case of labs on chip [12].

Even more than front-end flux, the back-end flux depends on the particular type of lab on chip we are dealing with. For example, some labs on chip are charged from the factory with reference chemical substances that are used when the chip works while others do not. When a reference substance is charged into the chip it is sometimes inserted at the wafer level during front-end processing, other times it is charged chip by chip in a back-end process [13].

A lot of the back-end process also depends on the type of adopted detection. For example, if optical detection is selected and the source and the detector are not on board of the chip, optical interfaces have to be provided, whose fabrication can be a difficult and costly back-end step [14].

Taking into account these observations, a possible back-end process flow for a lab on chip is represented in Figure 4.12 [12].

After the completion of the front-end working wafers are tested before dicing. This is a very important test whose scope is to avoid that wafers with important defects affecting the great majority of the chips are transferred to the back end causing a great number of defected chips to be worked one by one up to the first stage of back-end test.

After the test, the chips are separated in the process that is generally called dicing. Since in labs on chip wafer bonding is generally used, dicing is more critical than in standard integrated electronics processing, due to the need to not impact the superimposition between the two wafers [6].

Once separated, the dies are attached on the submount that will be the base of the final package with a process called die attach. This is needed also to allow easy handling of the single chip without the risk to damage it by scrapping the chip surface with the handling tools.

The surmount is also important for the chip thermal regulation. Depending on the requirement the surmount can be a simple base of thermal conducting material or it can be constituted, as shown in Figure 4.13, by superimposed layers of different materials to assure the required mechanical and thermal characteristics. If a cooler has to be placed below the chip to maintain a temperature control inside the package, it is placed on the surmount before die attach.

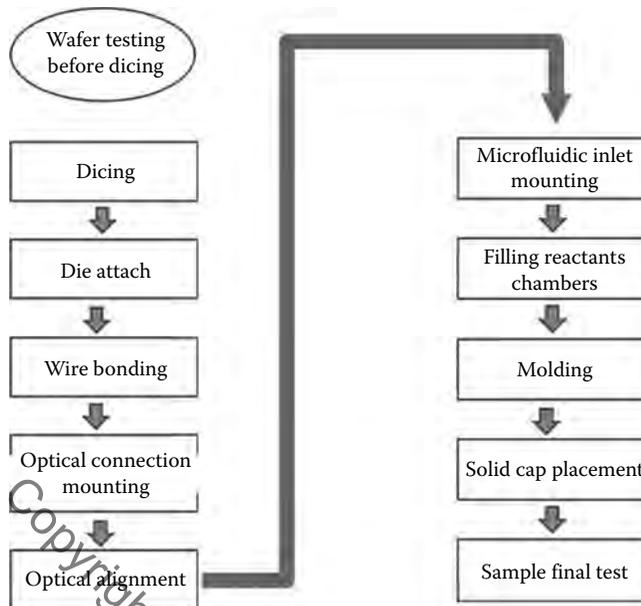


FIGURE 4.12 Possible back-end flux for a lab on chip.

After the die attach, the wires are fused on the electrical connections in order to be ready to connect the chip with the package pins during a process called wire bonding. Assuming that optical detection is used, optical interfaces have to be prepared in this phase.

In order to prepare optical interfaces, a groove has been constructed by lithography and etching in the place where the optical fiber has to interface the optical waveguide integrated in the chip (see Figure 4.14). The optical fibers that have to bring light out of the chip are inserted into the groove and aligned in order to optimize the coupling with the waveguide.

In the case of labs on chip, this operation is quite different from the same operation performed on standard integrated optics chips. On the one hand, the power budget of optical detection circuits is more favorable with respect to telecommunication-oriented optical interfaces, due to the possibility to integrate for several seconds the received signal to eliminate noise. Thus, a very small coupling loss is not needed and the so-called passive alignment is possible. This is constituted by fiber placing exploiting only the lithographic markers, without the help of real-time measurement

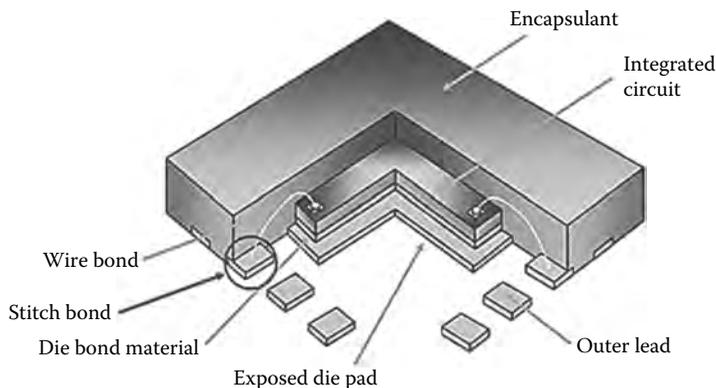


FIGURE 4.13 Example of package submount constituted by a set of superimposed layers of different materials.

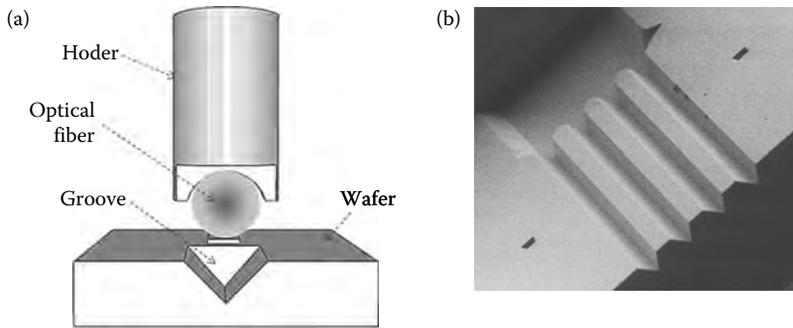


FIGURE 4.14 Passive alignment of a fiber optics in a V-groove in an integrated optics chip (a) and scanning electron microscope photo of a V-grooves array (b). ((b) Published under permission of Cyoptics Inc.)

of the coupling loss obtained injecting light in the waveguide during the coupling operation and measuring the fiber output power (active alignment).

Passive alignment can be automatized completely, thereby increasing the throughput and decreasing work cost. On the other hand, while in standard optical integrated circuits the groove is placed on the top of the wafer and no wafer bonding is adopted, in the case of labs on chip, wafer bonding obliges both to excavate a place for the fiber in both wafers and to place the fiber from the chip side. This is a much more complicated operation under a placing point of view and has an impact both on the cost and on the back-end yield. Different techniques have been devised for optical interface fabrication that are more suitable for labs on chip, as it will be discussed in some detail in Section 6.5.4; in any case, fabrication of optical interfaces remains a delicate process, to be realized with great care.

A possible alternative to the waveguide–fiber interface is to place both the laser source and the photodetectors on the chip. Depending on the cost of the two technologies and on the type of sources and detectors that are needed it can be a preferable solution or not. In this case, placing of lasers and photodiodes in suitable positions on the upper wafer are performed at the end of the front-end processing and it is a typical mixed front-end and back-end integration step.

As a matter of fact, on the one hand it is an operation performed on the whole wafer before dicing, while on the other, it involves discrete components like the photodiodes and the lasers to be integrated one by one. Due to the fact that the external chips will emerge from the surface of the upper wafer, also on the lower wafer suitable excavations have to be prepared for them.

After the preparation of the optical interfaces, fluidic interfaces have to be fabricated, by connecting suitable discrete elements to the inlets of the fluidic circuits. Since the inlets are on the lower wafer, suitable holes have to be done on the upper wafer to reach them. Selected inlets are prepared to be used during the lab on chip working, thus they have to be connected with suitable inlets prepared on the package itself to render possible liquid insertion from outside the chip when the chip is finished.

Other inlets are used only in factories to charge suitable reagents into the chip so that they have to be sealed after the charging of the reagents, a process located immediately after the fluidics inlets preparation. When all the interfaces of the chip, that is, electrical, optical, and fluidics are prepared, the chip is frequently molded. This means that the chip is immersed into a bit of specific glue or other suitable material to fix and to protect it. Molding is an important operation since the chip interfaces are not to be ruined by the molding. For this reason, when an external package is used, molding is not performed in every case.

Sometimes molding ends the back-end process: the chip is sufficiently protected by the mold and it is ready to be integrated. In other cases, the package also includes a solid upper cap that has to be placed after the molding on top of the surmount to form a case in which the chip is contained. This is naturally a much more robust solution, frequently adopted in nondisposable chips, but which

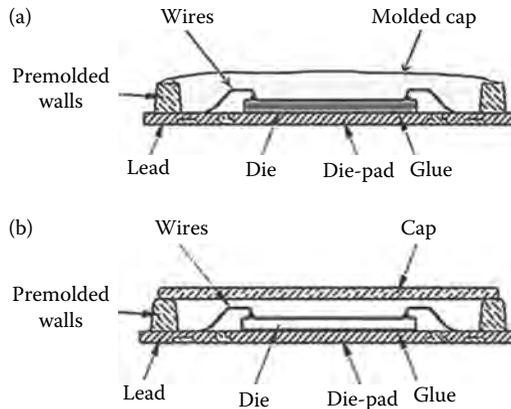


FIGURE 4.15 Purely molded (a) and solid cap closed (b) types of package.

implies higher costs both for the additional back-end step and for the additional solid cap, whose cost, even if minimal, is added to the overall cost of the chip. Purely molded and solid case package solutions are represented in Figure 4.15.

After the final test, the chip fabrication is ended and it is ready to exit from the fab.

4.2.3 PRODUCTION TESTING TECHNIQUES

4.2.3.1 On-Wafer Testing

All the testing techniques applied during front-end fabrication are called on-wafer testing due to the fact that the wafer is tested as a whole [15].

Owing to the fact that the circuits are not yet diced and the interfaces are not yet fabricated, a real functional test cannot be performed in this phase, and on wafer testing has the main scope of being eliminated as soon as possible from the production cycle. The wafer with extended defects probably will present a great number of nonfunctioning circuits. On-wafer testing results are also used to elaborate statistics on the front-end process so to monitor its overall quality and to individuate improvement areas.

Three types of on-wafer testing are generally conducted:

- Visual tests
- Optical tests
- Electrical tests

Visual tests are simply performed by looking systematically at the wafer surface with a microscope under different light conditions. Wafers containing integrated circuits of any kind has a periodical structure that under suitable light can be easily recognized, as it is evident, for example, in Figure 4.16, where the surface of the wafer containing integrated optical circuits is shown. A skilled operator, using an optical microscope can recognize nonuniform light reflection immediately; thus, identifying small defects on the wafer surface. The visual inspection station is essentially similar to a microscope whose construction is optimized for looking at a wafer surface, as shown in Figure 4.17.

When a defect is detected on a wafer surface it is generally looked at with the electron microscope to clearly define the defect characteristics and to try to individuate its cause. Systematic causes determining a recurring set of wafer defects are eliminated in this way. In many cases, the electron microscope is also used to look at samples of the production wafers to perform a better assessment of the front-end process.

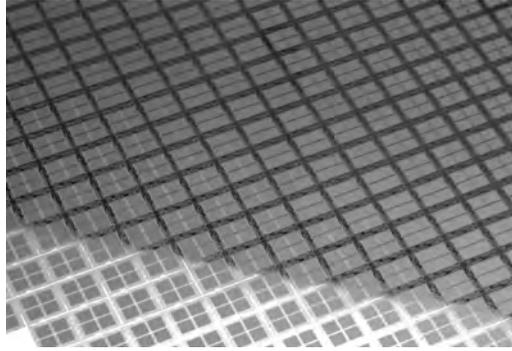


FIGURE 4.16 Photograph of the surface of a wafer evidencing the regular structure created by the periodic disposition of the chips. (Published under permission of Cyoptics Inc.)

Optical wafer tests are based on the same principle of visual tests, but for the fact that the reflected pattern generated when the wafer is lighted suitably are detected and analyzed by an optical system. This is able to detect the diffraction figure caused by monochromatic light incident on the wafer at different angles and with different frequencies and comparing them with the reference shapes obtained by a set of reference wafers.

This procedure, while more cumbersome with respect to simple visual inspection, is more effective in individuating small defects and in characterizing them. Sometimes this testing strategy is considered an integration of visual inspection used when defects are detected or on a sample basis.

Finally, electrical on-wafer tests consist of testing the electrical properties of the wafer. To execute this kind of tests specific elements are often constructed on the wafer, where point contacts are put to perform resistance or other kinds of electrical measures. These measures are then compared with the reference that is obtained from good wafers to characterize the wafer under test. Even if this is a less accurate measure with respect to the electrical one, it does not involve only the wafer surface, thus it is complementary and not alternative to it.

Electrical tests are more useful when conductive elements are present on the wafer, since if all the wafers are purely insulating, only capacity can be in case measured thus effectiveness in case of microfluidic circuits depends on the particular circuit and on the working stage at which the measure is performed.



FIGURE 4.17 Visual inspection station in a front-end fab clean room. (Published under permission of Cyoptics Inc.)

On-wafer test is also performed with another technique, complementary to whole wafer testing: the test of specific test patterns built on the wafer. This is particularly useful in the case of microfluidic circuits to be able to use on-wafer testing developed for microelectronics.

Several electronics tests are meant to determine data on the layers depth or their doping through the measurements of electrical parameters linked to the conducting nature of the layer [16]. This is not the case of microfluidics, unless specific test-oriented structures are built on the wafer to allow such tests, especially in the case in which a silicon wafer is used as substrate for the construction of the microfluidics circuit.

4.2.3.2 Back-End Testing

Labs on chip testing after dicing is generally called back-end testing and it is generally composed of structure and functional tests.

Structure tests are analogue to on-wafer tests, but implemented chip by chip. They have the scope to verify, by some form of chip inspection, that the chip structure corresponds to the design, thus avoiding detectable errors in chip construction. Structure tests are generally located at the beginning of the back-end process chain, constituting a pass–fail test before entering the most costly part of the processing of the lab on chip.

While the back-end process goes ahead and the chip is more and more similar to the end product, that functional tests become needed; they are tests that assess the functionalities of the lab on chip by somehow simulating its working and looking to the result. The most complex functional test is the final test, in which the whole operation of the lab on chip is reproduced and its outcome compared with design requirements.

Functional tests are a particularly difficult part of the lab on chip fabrication due to two reasons: they are one of the most costly parts of the back-end process so that they have to be designed with great accuracy to obtain the better process quality with the smaller test effort, on the other, differently from microelectronics circuits, labs on chip are often disposable products, so that they can be used only once [17–19]. This means that many in-line and final functional tests destroy the device and they can be performed only on a small sample of the processed chips [18].

The complexity related to the back-end test is to be faced by systematic procedures: namely fault simulation [20,21] and design for testing [22,23].

4.2.3.3 Fault Simulation

To optimize testing it is important to know where defects tend to be generated and what are the most faulty steps of the fabrication process. This can be done by setting up simulators of the fabrication steps and of the labs on chip working. These simulators are mostly oriented to take into account phenomena generating malfunctioning and faults like the statistical distribution of the fabrication tolerances and the most probable faults of the process machines.

Once the type and distribution of the most probable faults are known, specific tests for the most probable fault can be designed besides generic performance tests, greatly improving the efficacy to cost ratio of back-end testing. Moreover, such modeling also allows to better individuate the placing of intermediate process tests so as to minimize the process steps performed on already failed circuits.

Suitable modeling is needed to place test points in an optimized way [24]. This procedure implies the back-end processing chain simulation with different tests placing and the comparison of the obtained results in terms of cost of the produced working chip versus yield. The outcome of such a modeling is, in general, the statistical distribution of chips suffering different types of defects, as schematically shown in Figure 4.18.

Another important role of testing modeling is to relate an observed defect to the possible causes so as to drive post-testing analysis of defected chips in order to detect the steps of the fabrication process that is more important to improve. An example of this kind of modeling is provided in Reference [24], where it was analyzed as the test of pins for the driving of droplet-based microfluidic circuits. A table of the main defects related to electrodes fabrication is reported in Table 4.1.

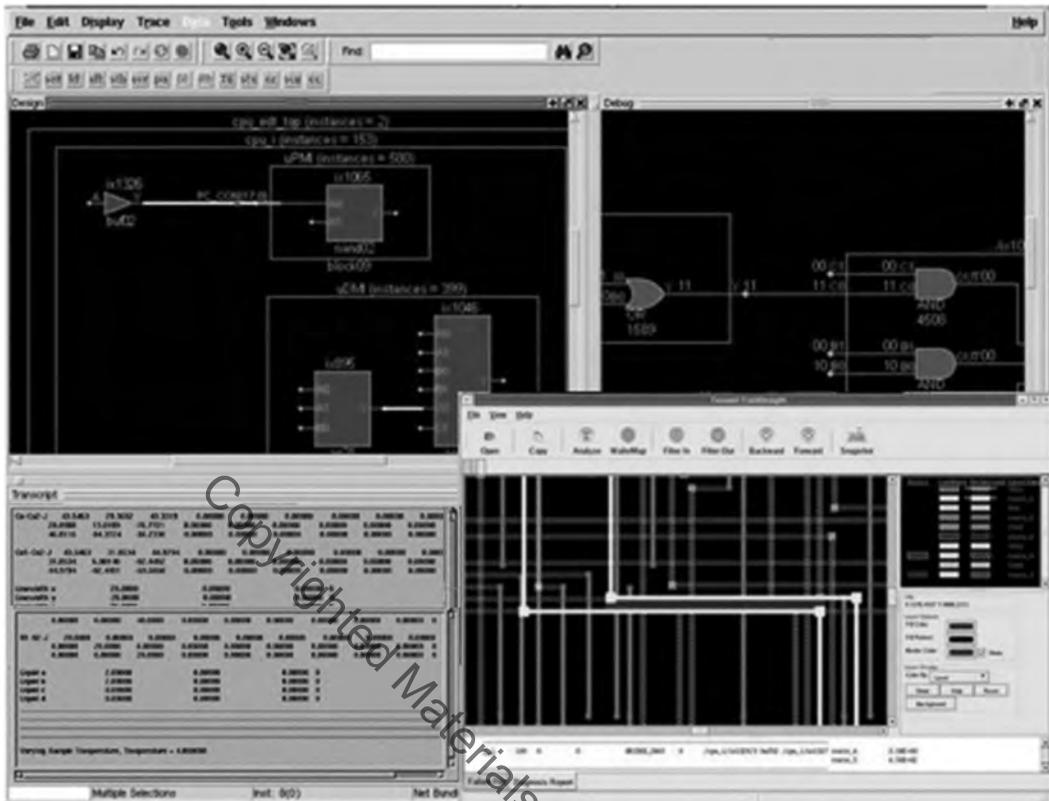


FIGURE 4.18 Screenshot from a simulation program aimed at the characterization of on chip defects at the end of the processing chain also considering the distribution of test points along the process chain. The program is developed for electronic circuits.

TABLE 4.1
Classification of the Main Defects Related to Electrode Fabrication in a Droplet-Based Microfluidics Circuit

Cause of Malfunction	Malfunction Type	Involved Cells	Fault Model	Observed Error
Electrode actuated for too long time	Irreversible charge concentration on the electrode	3	Droplet is dispensed but not separated by the reservoir	No droplet can be dispensed in the circuit
Electrode shape variation in fabrication	Electrode deformity	3	No overlap between droplet and electrode	Mixing failure
Electrode static property variation in fabrication	Unequal actuation voltage	3	Net static pressure in some direction	Unbalanced volume of split droplet
Bad soldering	Parasitic capacitance in the capacity sensing circuit	1	Oversensitive or insensitive capacity detection	False-negative or false-positive results

Source: Adapted from Chakrabarty K., *IEEE Transactions On Circuits and Systems—I: Regular Papers*, vol. 57, pp. 4–17, 2010.

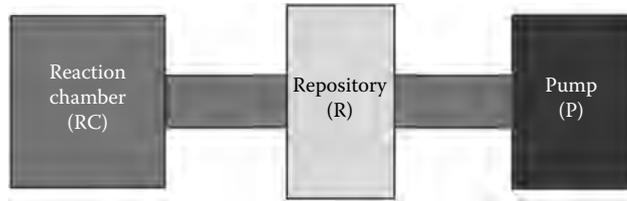


FIGURE 4.19 Scheme of the basic structure of a part of a microfluidics circuit considered in the example of Section 4.2.3.

4.2.3.4 Design for Testing

Design for testing is a design procedure that includes the design of testing procedure and includes in the chip structures that are intended to facilitate or enable specific testing techniques. This design strategy is well established for microelectronics circuits [22,23], while it is yet at research stage in the field of labs on chip.

However, the fact that several labs on chip are disposable devices, renders this approach even more necessary with respect to microelectronics, where it is considered a mandatory element of a successful circuit design. As an example, let us consider the simple part of a microfluidic circuit that is shown in Figure 4.19. The simple scope of this part of the circuit is to use a micro-pump (P) to move the solution stored in the repository (R) into a reaction chamber (RC) that, at this stage of the assay procedure is empty. The micro-pump can be operated several times without ruining its functionality, but the test solution in the repository can be moved only one time in the reaction chamber, and it has to be done at the right moment during the assay procedure. Moreover, the solution cannot be easily removed from the reaction chamber, since it is chemically reactive and a complete elimination would require washing with a specific wash solution that is not easy to incorporate into the chip.

The micro-pump was detected by simulation as one of the most critical elements of the chip and, while other elements could be tested only on a sample basis once the production process is at steady state, the micro-pump is to be tested at the beginning on every chip before releasing it and, also when the production is at its steady state, the sample test of the pump should be performed, more frequently, with respect to the complete final test.

To do that, the conceptual design represented in Figure 4.20 was adopted. A small secondary repository (SR) is created immediately after the micro-pump so as to be filled with a solution of water and a chemical neutral alcohol of about the same viscosity of the test solution (that in this

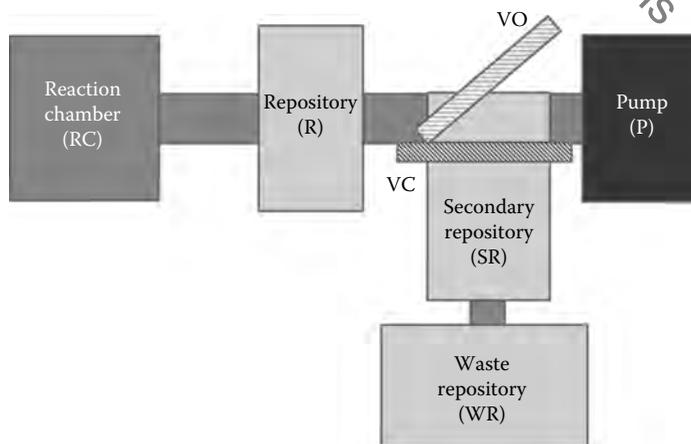


FIGURE 4.20 Scheme of the implementation of the functionality of the circuit reported in Figure 4.19, including a structure allowing nondestructive testing of the micropump.

case had a mixed alcohol–water solvent since it was needed to avoid jellification during the assay). A single opening valve VO (a valve that can be operated once starting closed and ending opened) was constructed besides a single close valve VC (a valve that can be operated once starting open and ending closed) to open and close at the right moment the access to the waste repository (WR). Just before the charging of the test solution into the repository, the neutral water–alcohol solution was charged in SR while VC is open VO closed. After that, the micro-pump testing starts: the pump is operated so that the water–alcohol solution is pushed up to WR, where a secondary testing system verifies both that no solution remains in SR after a certain time and that all the solution arrives in WR after the same time. At the end of the pump testing, VC is closed, VO opened and the chip is ready for charging the test solution in R.

This procedure is complex and requires construction of an elaborate structure (the design of Figure 4.19b is only conceptual, the real design is much more complex) only to allow the micro-pump test, but if the yield of the process essentially depends on the micro-pump yield and the valves are quite irrelevant in decreasing the yield itself, this can be needed to assure the correct industrial cost, that is attained only if a target yield is overcome.

The above example is somehow an extreme one: not always does design for testing require such an extended modification of the circuit layout, but it is quite significant of the effort that has to be put in designing, up to the very early stages of a project, the tests that will allow the production process to be correctly monitored.

Frequently, design for testing also consists of the construction on the chip of supplementary measurement devices that are used to test if different operations during the back-end process are executed correctly. For example, if during the front end it is necessary to charge an antibody solution in a reaction chamber whose floor surface has been activated to let the antibodies to be captured on this surface, after the washing needed to remove the remaining solution, verifying the effective saturation of the reaction chamber surface can be needed. This can be carried out, for example, with optical methods if the fabrication process is suitably designed.

4.2.4 IN-FIELD TESTING

Another field in which testing techniques, mainly fault simulation and design for testing, are a key for the success of a lab on chip project is the possibility of implementing operational tests.

When the chip is used, in order to assure that the performed measure is reliable, the chip has to have an internal self-test that is performed contemporary to the real measurement and has to be present, in case of a working chip, a predetermined result. If the result of the self-test is not correct, it is to be intended that the chip is not working and the result of the performed assay not reliable.

In medical and biological applications, the presence of such a self-test is mandatory to allow the assay to be accepted by local administrations so that it is a key point in the chip design and realization [25,26]. Also, in this case, since the chip is usually disposable, a suitable design for testing has to be realized, followed by the realization on the chip of test dedicated structures.

4.3 MICRO- AND NANO-FABRICATION FABs

Planar technologies require a specific manufacturing structure. In particular, all the manufacturing processes, both front end and back end, have to be executed in conditions of controlled environment and carefully planned action by the manufacturing people. This is required both to maintain a very low contamination degree, needed condition to assure an high yield process, and to assure the required security conditions in the presence of chemical and mechanical hazards. The environment in which these conditions are fulfilled is an artificially controlled site, generally called clean room, to underline the strict control of spurious elements present in the air and in all the materials here processed.

Under the point of view of the base material used for manufacturing, it is prepared in the form of standard wafers, as required by front-end processing machines. The kind of starting material that

is used and the properties of the used wafers strongly influence the manufacturing process and the final industrial cost of the chips. A certain variety of possible materials exists even in microelectronics, for example, both pure silicon and SOI wafers can be used, not only of different radii, but also of different depth.

This variety is greatly enlarged in microfluidics and integrated optics wafers that are used to build labs on chip. Silicon wafers or SOI wafers can be both used as base, but also pure glass wafers are commonly adopted. Moreover, polymeric substances are frequently used in this field not only as process materials, but also for structural purposes, when they are deposited and directly polymerized on the base wafer [1,2].

4.3.1 CLEAN ROOMS

A clean room is “a room in which the concentration of airborne particles is controlled, and which is constructed and used in a manner to minimize the introduction, generation, and retention of particles and microbes inside the room and in which other relevant parameters, e.g. temperature, humidity, and pressure, are controlled as necessary” [27].

The importance of a strict control of the density of any kind of unwanted particle is clear if we consider the case in which only a total of 10^5 unwanted particles are present in a cubic meter of air. The average volume occupied by a particle due to its random motion is 10 cm^3 that we can approximate with a cube with 2.1 cm side. A standard pure Si wafer of a diameter of 8' has an equivalent area of about 0.125 m^2 . Assuming to have 1500 chips on the wafer we can estimate that the number of particles that could deposit on the wafer in still air, for example, in a moment when the wafer is released on a support, is about 285.

If the probability that more than one particle hit a single chip is neglected and we assume that every chip with a particle deposited upon is ruined we have that up to 60 chips are ruined so that, this phenomenon only brings to a yield reduction from 100% to 81%. The hypothesis that every depositing particle creates a failure in a chip is not true in real clean rooms, due to the low speed at which depositing particles hit the wafer, causing a very weak link of the particle to the wafer surface and enabling an effective wafer surface cleaning. In any case, the above example is quite effective in showing the very strict requirements of clean rooms adopted for integrated circuits manufacturing. As a matter of fact, the average number of particles that are present in the air of an internal room that is kept clean with standard methods, for example, in a physician visiting room, is as high as 10^8 and 10-fold higher in normally clean external air.

4.3.1.1 Clean Room Classification

Owing to the paramount importance of controlling unwanted particles in the air, clean rooms are classified on the ground of the average number of spurious particles in the air for any given type of particle. As a matter of fact, it is intuitive that, bigger the particle, greater its energy for a constant air speed and greater the probability to damage the wafer surface.

The first universal classification of clean rooms was set-up in the United States with the different version of Federal Standard 209 [28]. From versions A to D, the clean room class number refers to the maximum number of particles of $0.5 \mu\text{m}$ size or bigger that would be allowed in 1 cubic foot of clean room air, generating a widely used way of referring to the quality of a clean room as a class N clean room. The table summarizing the clean rooms classification in standards 209A-D is reported in Table 4.2. In 1992, the standard 209 was updated to version E by introducing more classes and metric units, and also to create a table of comparison with the upcoming ISO standard for clean room classification.

Federal Standard 209E was used until the end of November 2001 to define the requirements for clean rooms. On November 29, 2001, these standards were superseded by the publication of ISO specification 14644 [27] that was de facto adopted all around the world; a summary of the clean room classes reported in ISO 14644 is reported in Table 4.3 in terms of number of unwanted particles of a

TABLE 4.2
U.S. Federal Standard 209A-D for Clean Room Classification

Clean Room Class	Number of Particles for Cubic Feet of Air for Particle Radius				
	0.1 μm	0.2 μm	0.3 μm	0.5 μm	5 μm
1	35	7	3	1	
10	350	75	30	10	
100	3500	750	300	100	
1000	1000	7			
10,000	10,000	70			
100,000	100,000	700			

Source: Adapted from U.S. General Services Administration, Federal Standard 209 A-D, 1992.

TABLE 4.3
ISO 14644-1 Clean Room Classification

Clean Room Class	Number of Particles for Cubic Meter of Air for Particle Radius					
	0.1 μm	0.2 μm	0.3 μm	0.5 μm	1 μm	5 μm
ISO 1	10					
ISO 2	100	24	10	4		
ISO 3	1000	237	102	35	8	
ISO 4	10,000	2370	1020	352	83	
ISO 5	100,000	23,700	10,200	3520	832	29
ISO 6	1,000,000	237,000	102,000	35,200	8320	293
ISO 7				352,000	83,200	2930
ISO 8				3,520,000	832,000	29,300
ISO 9				35,200,000	8,320,000	293,000

Source: Adapted from ISO 14644-1 Recommendation—*Cleanrooms and Associated Controlled Environments—Part 1: Classification of Air Cleanliness*.

certain dimension in a cubic meter of air. It is possible to establish a precise correspondence between the still quite diffuse language related to Standard 209 and the modern ISO classification as shown in Table 4.4 so that the ISO clean room class can be readily derived when the old 209 name is used.

4.3.1.2 Clean Room Structure and Procedures

Four basic components contribute to create and maintain the clean room controlled environment [29]:

- *Clean Room Architecture*—Materials of construction and finishes are important in establishing cleanliness levels and are important in minimizing the internal generation of contaminants from the surfaces.
- *The Heating, Ventilation, and Air Conditioning (HVAC) System*—The integrity of the clean room environment is created by the positive pressure difference between the clean room and adjacent areas and through the HVAC system. The HVAC system requirements include:
 - Supplying airflow in sufficient volume and cleanliness to support the cleanliness rating of the room.

TABLE 4.4
Correspondence between U.S. Federal Standard
209 and ISO 14644-1 Clean Room Classification

ISO	Federal Standard 209
ISO 3	1
ISO 4	10
ISO 5	100
ISO 6	1000
ISO 7	10,000
ISO 8	100,000

- Introducing air in a manner to prevent stagnant areas where particles could accumulate.
- Filtering the outside and recirculated air across high-efficiency particulate air (HEPA) filters.
- Conditioning the air to meet the clean room temperature and humidity requirements.
- Ensuring enough conditioned makeup air to maintain the specified positive pressurization.
- *Interaction Technology*—Interaction technology includes two elements:
 - The movement of materials into the area and the movement of people.
 - Maintenance and cleaning.

Administrative instructions, procedures, and actions are necessary to be made about the logistics, operation strategies, maintenance, and cleaning.

- *Monitoring systems*—Monitoring systems is needed to continuously assess the correct clean room functioning. The variables monitored are pressure difference between the outside environment and the clean room, temperature, humidity, and, in some cases, noise and vibrations.

4.3.1.2.1 Clean Room Architecture

The clean room architecture greatly depends on the clean room class. In the case in which low class clean rooms are needed, like ISO 4 or ISO 5 rooms constructed for front-end manufacturing or even ISO 3 clean rooms where SOI wafers are fabricated, the general architecture tends simultaneously to minimize the space where the high controlled environment has to be maintained and to facilitate as far as possible both the HVAC system work and the execution of the needed internal procedures.

The most adopted architecture in this case is constituted by a shelf containing the clean rooms inserted inside the industrial building that has to be suitably prepared for that. An external view of this kind of architecture is provided in Figure 4.21. Typical height of the internal shell is 3 m, while the external roof has to be high about 8 m to host all the technical equipment needed to realize the HVAC. As a matter of fact, it is partially installed directly on top of the clean room while partially it has to be connected with the clean room rooftop with suitable pipes.

A possible internal plan of a low-class clean room is shown in Figure 4.22, where it is clear that the low-class clean areas are insulated so as to be better controlled and maintained. Moreover, the different areas can be specialized so as to be functional to the performed processes. For example, in clean room where high definition lithography equipment is located, like a submicron stepper or an electron beam lithography equipment, it is possible that a light with a controlled wavelength is needed not to interfere with the lithography process. In these cases, generally a yellow light is used. The bay architecture allows the yellow light to be used only in the bay where the sensitive equipment is located, without affecting the other areas of the clean room where it is not needed.

Where higher class clean rooms are needed, like ISO 8 or ISO 9 clean rooms, the bay architecture is less used and more commonly the clean room is a sort of open space where the machinery and



FIGURE 4.21 Plan of a modular clean room: clean areas (class ISO 4 and ISO 5) are located inside the structure as bays, the process and the exit corridors are class ISO 6 while interface areas like the in/out structure are class ISO 7. Air showers and other personal cleaning facilities are located in the in/out areas. (Published under permission of Cyoptics Inc.)

the test benches are located so as to allow the best possible circulation of materials and operators.

As far as the clean room floor is considered, it is always realized via a floating floor allowing the air flux to pass through so that the air, after entering from the ceiling and traversing the clean area is collected below the floor and sent back again to the HVAC system. When equipments that are greatly sensitive to vibrations are installed, it could be needed to build a specific support for the equipment providing vibration filtering. This is realized with alternating rigid and elastic layers of material so as to obtain the wanted mechanical filtering response.

Also, the materials that are used to build the structural parts of clean rooms (walls, floor, and roofs) have to be selected suitably. They have not to generate pollution, so that standard walls using paint are generally excluded.

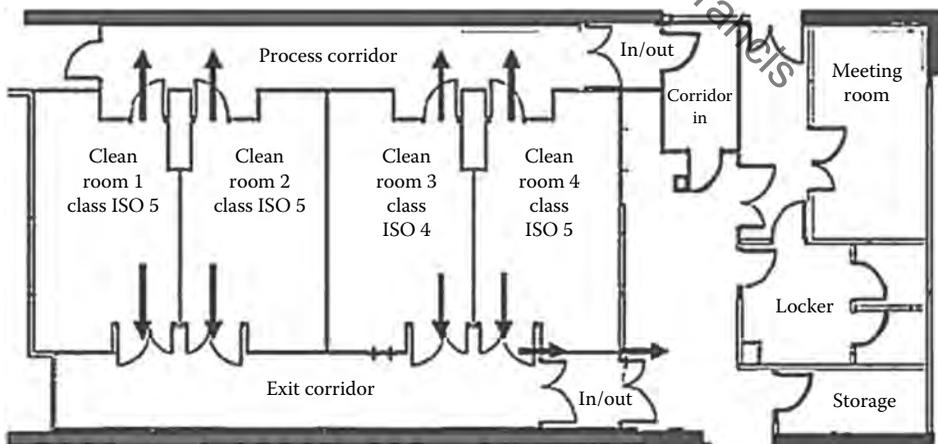


FIGURE 4.22 Plan of a modular clean room: areas in darker gray are clean areas (class ISO 4 and ISO 5—the clean rooms, class ISO 5—the corridors), light gray areas are transit and preparation areas (class ISO 6), white areas are interface areas (class ISO 7). Air showers and other personal cleaning facilities are located in the in/out areas.

4.3.1.2.2 The HVAC System

Owing to the need of maintaining a very high standard of air cleanness, the HVAC system of a low-class clean room is a very complex system collecting several equipment [30,31].

Spurious particles are generated in the clean room from several sources so that the airborne contamination level of a clean room is largely dependent on the particle-generating activities in the room. It has been found that many of these contaminants are generated from five basic sources:

1. The facilities itself and the people operating in the clean room
2. The fabrication tools
3. The fluids that are present in the clear room
4. The product being manufactured

A list of the more common sources of air contamination due to these elements is presented in Table 4.5.

Generally, from 20 to 60 complete air changes per hour are needed into the clean room, arriving at a requirement of 600 air changes per hour in ISO 4 and ISO 3 clean rooms. For comparison, in a normal air-conditioned environment, where a general comfort level has to be maintained through cleanness, humidity, and temperature control, only about five air changes per hour are needed.

The high-efficiency air cleaning needed in a clean room is attained by using of HEPA filters, specialized to filter different kinds of particles that purify the air that is injected into the clean room. These filters are generally placed on the clean room ceiling and in low-class clean rooms they fill almost all the ceiling area. The air flux inside the clean room has to be maintained as laminar as possible by an accurate design of the HVAC system, since vortices tend to spread randomly unwanted particles and render air cleaning very difficult near the vortex center.

Moreover, the air pressure into the clean room is maintained a bit higher with respect to the external air pressure, so as to favor particle removal from the clean room by pushing them toward the floor where the air is removed and transported out of the controlled environment, as shown in Figure 4.23. The air is maintained as laminar into the bay by a combination of a controlled air injection from ceiling and a system of air extraction from the floor. Frequently, this system is based on a vacuum pump placed directly under the ceiling collecting air extracted from the floor and used to maintain a constant pressure difference between the clean room internal volume and the space above the ceiling so as to increase pressure prevalence in the bottom part of the bays and contribute to maintain the flux laminar. A scheme of this arrangement in a set of clean room bays is represented in Figure 4.24.

TABLE 4.5
List of Common Air Contamination Sources That Are Present in a Clean Room

Facilities	People	Tools	Fluids	Product
Walls, floors, and ceilings	Skin flakes and oil Spittle	Friction and wear particles	Particulates floating in air	Chips Quartz flakes
Paint and coatings	Clothing debris (lint, fibers, etc.)	Lubricants and emissions	Bacteria, organics, and moisture	Aluminum particles
Construction material	Hair	Vibrations Brooms, mops, and dusters	Floor finishes or coatings Cleaning chemicals	
Air conditioning debris			Plasticizers (out-gases)	
Room air and vapors			Deionized water	
Spills and leaks				

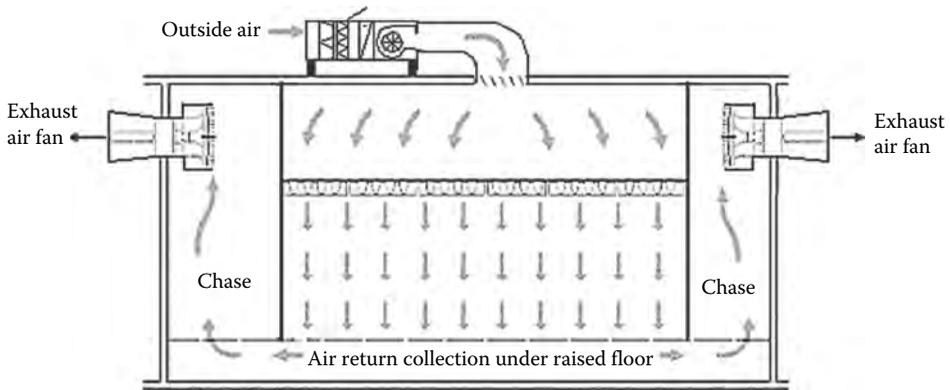


FIGURE 4.23 Laminar air flow in a clean room bay.

In the scheme of Figure 4.24, the low-class bays have 100% HEPA ceiling coverage. The make-up air handler (MAH) is a fresh air unit that provides the room pressurization and is designed for latent and sensible load of outside air. This unit feeds on two recirculation air handlers (RAH) that supply air into the clean room. The RAH are usually designed primarily to manage the sensible heat load generated indoors from the process equipment and occupancy.

Laminar airflow tends to become turbulent when it encounters obstacles such as people, process equipment, and workbenches. Placing these obstructions in a manner that prevents dead air spaces from developing will minimize turbulence. Use of workstations with perforated tabletops will allow the air to pass through them uninterrupted. Equipment shall also be raised on a platform (plinth) where possible to allow free air flows beneath it.

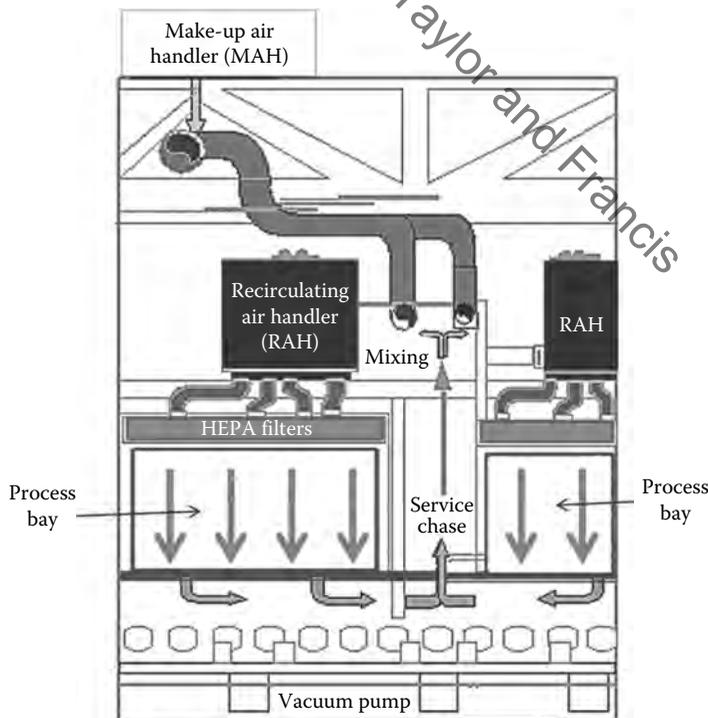


FIGURE 4.24 A possible arrangement to obtain laminar air flow in a few clean rooms bays.

The airflow design with air coming from the ceiling and exiting from the floor is the most frequently used. In some designs, the supply air can be projected upwards from floor void and is drawn into a ceiling void. This arrangement is preferred in applications where the localized hardware or equipment has high heat dissipation. The conventional supply airflow from ceiling may not be directional enough to cool the equipment that results in hot spots.

Besides any air flow study, the cleaning of the air injected into the clean room is a key point to assure the required level of cleanness. Air cleaning is performed with a cascade of different HEPA filters. A filter cascade is needed not only to refine progressively the air purification (filters suitable to purify dirty air are different from filters that increase the purification of already clean air), but also because HEPA filtering principle is quite different depending on the kind of particles that the filter has to separate and capture from the air flow.

In principle, since we are talking about small particles, it is even not completely clear if, in analyzing the performances of HEPA filters, we have to model the particle in the air flow as slide objects floating in a moving fluid or as a very diluted solution of many solutes in a gaseous solvent. The more suitable model to represent the phenomena happening during air purification and injection into the clean room critically depends on the average radius of the suspended particles. In the most common applications, a floating solid model is applied above a radius of $0.01\ \mu\text{m}$ and a diluted solution below this radius. Filtration of particles relies on four main principles:

- Inertial impaction
- Interception
- Diffusion
- Electrostatic attraction

The first three of these mechanisms are mainly used in mechanical filters and are influenced by particle size.

Impaction occurs when a particle traveling in the air stream deviates from the air stream (due to particle inertia) and collides with a fiber. Generally, impaction filters can only satisfactorily collect particles above $10\ \mu\text{m}$ in size and therefore are used only as pre-filters in multistage filtration systems. The higher the velocity of air stream, the greater is the energy imparted to the particles and greater is the effectiveness of an impaction filter.

Interception occurs when a large particle, because of its size, collides with a fiber in the filter. The particles then remain stuck to the fibers mainly because of van der Waals forces.

Diffusion occurs when the Brownian motion of a particle causes it to contact a filter fiber. Diffusion works with very small particles and is frequently used in HEPA filters. In order that a diffusion filter works properly, particles have to be so small that they are aimed by relevant Brownian motion: the smaller the particle, the stronger the effect.

Electrostatic attraction plays a very minor role in mechanical filtration, but it is used in specifically designed filters. If a charged particle passes through an electrostatic field, it is attracted to an oppositely charged body. Such charges can be generated and imparted to particles in an airstream by attrition or other methods. The typical electrostatic air filter is made from polyester or polypropylene strands that are supposedly charged as the air passes through them.

4.3.2 FABRICATION MATERIALS: SILICON, SILICA ON SILICON AND PURE SILICA WAFERS

Another peculiar aspect of planar technology for labs on chip is the fact that several base materials can be used. Differently from microelectronics, for which a semiconductor material is mandatory and silicon is practically the only choice for digital circuits, several examples of microfluidic circuits have been built using a variety of base materials. When microfluidics circuits are used for applications that do not require chemical processing [18,19] bulk silicon is a possible choice. In this case, all the consolidated processes developed on this material for microelectronics can be used.

Moreover, wafer bonding is easier when it is performed to join crystal surfaces cut along the same crystal plane, due to the natural tendency of crystals to reproduce the crystal structure as possible to decrease the overall system energy.

When chemical processing is needed, generally pure crystal silicon cannot be adopted, due to its tendency to react with biological molecules and polymers and to absorb several organic structures. However, pure silicon when exposed to oxygen, also by a simple exposure to air, generates a thin surface silica layer. This layer is generally sufficient to passivize the surface under a chemical point of view so that silicon labs on chip can be realized simply allowing spontaneous silicon passivation.

An alternative choice for the chip base material is pure silica, which is one of the most inert materials when dealing with organic components and has very good mechanical properties. Silica can also be manufactured with well consolidated processes and it is also a common platform for integrated optics. Thick silica layers in excess of 10 μm can be grown on standard silicon wafers using different deposition techniques (see Sections 4.8) or whole silica wafers can be manufactured, thus providing a wide variety of alternatives.

The choice of the base material influences not only the fabrication processes and the industrial cost of the final product, but also the integration degree that can be attained in the final chip. At least in principle, using silica on silicon, microelectronics, microfluidics, and optics can be integrated together. Such a perspective is extremely interesting considering the possibility to integrate on a single chip a wide range of functionalities, thereby increasing more and more the advantage of planar processing. Developing such an ambitious technology presents, however, the great challenge of integrating all the processes needed to realize the different kinds of devices in a complete process flow. The use of pure glass wafers excludes the possibility of realizing electronics circuitry, but maintains the possibility to integrate microfluidics with optics.

4.3.2.1 Standard Silicon Wafers

Silicon wafers are produced in a wide variety and in huge volumes, thus they are a cheap substrate, even if the fabrication of a perfect silicon single-crystal wafer is a complex industrial process.

The standards for dimensions of silicon wafers, and mechanical and electrical characteristics are released by Semiconductor Equipment and Materials International (SEMI) [32], which is the collective organization of the microelectronics companies. The main SEMI standards for silicon wafers are listed in Reference [33]. Silicon wafers are available in a variety of sizes from 25.4 mm (1 in.) to 300 mm (11.8 in.).

The main standard diameters and relative thicknesses of crystal silicon wafers are listed below

- 1-in. (25 mm)
- 2-in. (51 mm); thickness 275 μm
- 3-in. (76 mm); thickness 375 μm
- 4-in. (100 mm); thickness 525 μm
- 5-in. (130 mm) or 125 mm (4.9 in.); thickness 625 μm
- 150 mm (5.9 in., usually referred to as “6 inch”); thickness 675 μm
- 200 mm (7.9 in., usually referred to as “8 inch”); thickness 725 μm
- 300 mm (11.8 in., usually referred to as “12 inch”); thickness 775 μm

A photograph of three 200 mm wafers is reproduced in Figure 4.25a, where the superficial structure of the wafers, incorporating different types of optical and electronics integrated circuits, is put in evidence.

The current state-of-the-art microelectronic fab is able to manufacture 300 mm wafers (commonly called 12 in. wafers), with the next standard projected to be 450 mm (18 in.). Even if there is formally an agreement to push the adoption of 450 mm wafers, there is a great resistance in the industry to afford the great research and development needed to reach this target in a short time, demonstrated by the fact that the original date stated for the first fab using 450 mm wafer was half

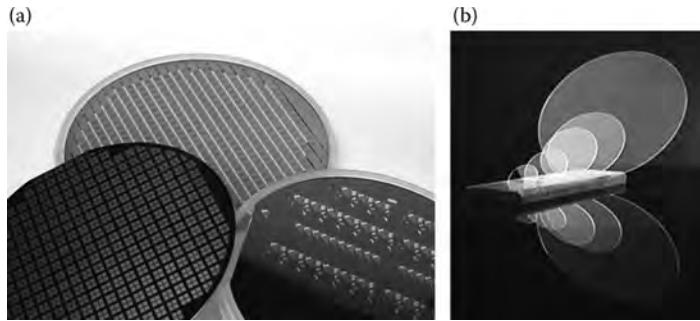


FIGURE 4.25 A photograph of three 200 mm wafers where the superficial structure of the wafers, incorporating different types of optical and electronics integrated circuits, is put in evidence (a) and comparison between 76, 100, 130, 150, and 200 mm wafers (b). ([a] Published under permission of Cyoptics Inc.)

2012, while in reality at that date this target seems quite far. In fact, moving to 300 mm wafers was a difficult operation even for the fabrication of very large volume ICs like RAMs and EPROMs, probably more difficult with respect to forecasts, and it happens during a hard industry crisis period. The general opinion is that recovering the investment that was done to transit to 300 mm is still far in time and in these conditions there is no means to plan another extremely costly transition.

The situation is completely different in the field of labs on chip. Labs on chip production has not yet arrived to the volume characteristics of microelectronics ICs and there is no reason pushing to set up an expensive fabrication structure as a 300 mm fab. Wafers with 200 or 150 mm diameter generally over-satisfy the needs of labs on chip production and research is frequently performed using smaller wafers.

Wafers grown using semiconductor materials other than silicon are also used for high-speed analog electronics [34,35] and active optical devices (like lasers or semiconductor amplifiers) [5] and they have different thicknesses than a silicon wafer of the same diameter. Wafer thickness is determined by the mechanical strength of the used material; the wafer must be thick enough to support its own weight without cracking during handling.

But for the case of silica, wafers are grown from crystals, mainly silicon having a diamond cubic structure with a lattice spacing of 5.43 Å [36]. To maintain a regular crystal structure of the wafer material, the wafer surface has to be aligned in one of the main crystallographic planes (see Figure 4.26). Orientation is important in electronics since crystal's structural and electronic properties are highly anisotropic.

Moreover, several processes are also influenced by the crystal orientation. For example, ion implantation depths depend on the wafer's crystal orientation, since each direction offers distinct paths for transport [37]. Wafers under 200 mm diameter have flats cut into one or more sides indicating the crystallographic planes of the wafer. In earlier-generation wafers a pair of flats at different angles additionally coded the doping type (see Figure 4.27). Wafers of 200 mm diameter and above use a single small notch to convey wafer orientation, with no visual indication of doping type.

Silicon wafers are generally not pure silicon, but are instead formed with an initial impurity doping concentration between 10^{13} and 10^{16} per cm^3 of boron, phosphorus, arsenic, or antimony which is added to the melt and defines the wafer as either bulk n-type or p-type [3]. However, compared with single-crystal silicon's atomic density of 5×10^{22} atoms per cm^3 , this still gives a purity greater than 99.9999%. The wafers can also be initially provided with some interstitial oxygen concentration.

A last comment is needed regarding SOI wafers. In the most advanced microelectronics applications, in order to further improve the insulation of the superficial silicon film from the bulk wafer so decreasing spurious capacities in CMOS and bipolar transistors, silicon on insulator (SOI) wafers are used [38]. SOI wafers are constructed by dividing the superficial silicon layer where the electronics circuits have to be fabricated by the bulk silicon with an insulator layer, silica generally, or sapphire

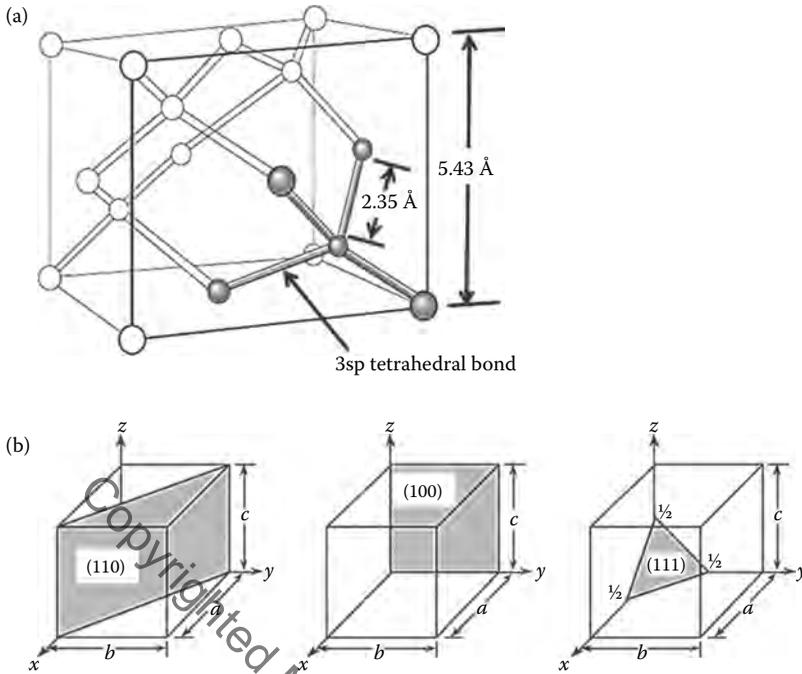


FIGURE 4.26 Silicon crystal elementary cell (a) and main crystal planes (b).

in high frequency and radiation-sensitive applications. SOI wafers however, up to now, find no application in the lab on chip field. Thus, we will not deal with this sort of wafer in more detail.

4.3.2.2 Glass Wafers

Several applications besides biotechnologies exploit wafers, thus the production of glass wafers is quite diffused and also if an official standard does not exist the production of such wafer reaches sufficient volumes to be suitable for mass production [39]. Dimensions of silica wafers

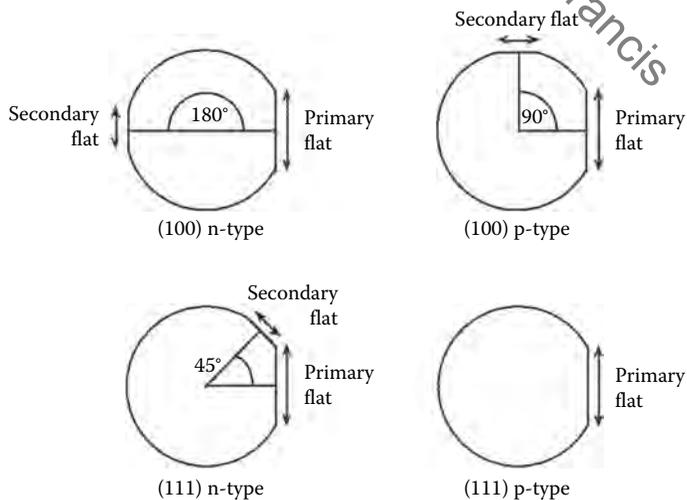


FIGURE 4.27 Standard placing of primary and secondary flats in wafers with a diameter smaller than 200 mm.

are compatible with dimensions of silicon wafers so that several processes can be performed with the same process machines and big glass wafers (like 300 mm) are much easier to be manufactured with respect to silicon wafers of the same dimension. Directly using glass wafers can be thus a good solution if a scalable fab has to be designed, that could evolve to very large volume production.

Several types of glasses are available, besides the possibility of producing special glasses obtained by doping with specific dopants. An example of specialized glass is those doped with different kinds of dyes so as to generate the probe light for luminescence detection in labs on chips [40]. A list of available glasses for wafers up to 200 mm is presented in Table 4.6 while an example of dimensions of a set of commercial glass wafers is presented in Table 4.7.

4.3.2.3 Number of Chips on a Wafer

Optimum wafer exploitation in terms of number of chips per wafer is important when mass production is considered, especially when large wafers are used. A simple rule, frequently adopted for wafers up to 200 mm, is to place the center of the wafer in correspondence with the contact point among four chips, as shown in Figure 4.28. In this case, a simple approximate expression can be obtained for the maximum number of chips gross Die Per Wafer (gDPW) that can be placed on the wafer as a function of the wafer diameter d and the chip area S :

$$gDPW = d\pi \left(\frac{d}{4S} - \frac{1}{\sqrt{2S}} \right) \quad (4.1)$$

In practice, the number of chips per wafer is smaller than gDPW due to the presence of the wafer of test sites and alignment markers that are needed both for mask aligners and for wafer bonding. This simple chip-placing rule, however, does not reflect a rigorous optimum chip-placing strategy. While the wafer dimension increases, the optimum placing becomes greatly important and several mathematical algorithms have been developed to attain this task [40,41].

4.4 PLANAR TECHNOLOGY COST MODEL

The goal of this section is to compare the cost model of chips fabricated with integrated planar technology with other kinds of goods whose fabrication process relies on different technologies. This will evidence the characteristics of planar micro-technology and in particular the fact that it is particularly suitable when high volumes are required. The cost estimation for a microfluidics chip is performed through the following steps:

- A reference production volume in the range of the practical values for this kind of products is fixed
- The fabrication process, composed of front-end fabrication and back-end chips processing is decomposed in elementary steps
- For each step the labor, material, and depreciation costs are estimated
- The impact of each step on the yield is also estimated
- At the end, the estimated industrial cost is produced
- After that, the industrial cost is derived for other volumes using standard models

This is similar to the algorithm used in real production situations, but it is to be underlined that expansion to different volumes has been carried out via analytical models that summarize average trends, completely neglecting effects due to negotiation or market fast fluctuations that are relevant in practical situations. Moreover, even if the reported estimates rely on the great experience that

TABLE 4.6
List of Different Kinds of Glasses Available to Be Manufactured in Glass Wafers with Characteristics and Potential Applications

	Density (g/cm ³)	Thermal Expansion (1/°C)	Strain Point (°C)	Annealing Range (°C)	Softening Point (°C)	Refraction 435 nm	Index 635 nm	Dielectric Constant	Acid/Alkali Resistance
Soda lime (mirrors, microslides)	2.44	8.6×10^{-6}	490	480–575	575	1.523	1.513	7.75	—
Pyrex (filters, lenses)	2.23	32.5×10^{-7}	510	560	821	1.468	1.475	—	Class1/Class2
Alkaline earth aluminosilicate (flat panel displays)	2.54	37.6×10^{-7}	666	721	975	1.5290	1.5160	5.7	—
Lithium potash borosilicate (electrical applications)	2.13	32×10^{-7}	456	496	—	1.469 (589.3 nm)		4.0	Class2/_
Borosilicate (touch panel, LCD, electroluminescence)	2.51	8.6×10^{-6}	529	557	736	1.5230 (588 nm)		6.7	—

TABLE 4.7
Example of Parameters for a Set of Commercially Available Glass Wafers

Geometry (mm)	Geometry (")	Flats	Dimension First Flat	Dimension Second Flat
50.8 mm × 0.127 mm – 6.35 mm thick	2" × 0.005" – 0.250" thick	With/without	0.625"	–
76.2 mm × 0.127 mm – 6.35 mm thick	3" × 0.005" – 0.250" thick	With/without	0.875"	–
100 mm × 0.127 mm – 6.35 mm thick	3.937" × 0.005" – 0.250" thick	With/without	1.280"	–
100 mm × 0.127 mm – 6.35 mm thick	3.937" × 0.005" – 0.250" thick	1 flat	1.780"	–
125 mm × 0.127 mm – 6.35 mm thick	4.921" × 0.005" – 0.250" thick	With/without	1.670"	–
125 mm × 0.127 mm – 6.35 mm thick	4.921" × 0.005" – 0.250" thick	2 flats	1.670"	2.50"
150 mm × 0.127 mm – 6.35 mm thick	5.906" × 0.005" – 0.250" thick	1 flat	2.640"	–
200 mm × 0.127 mm – 6.35 mm thick	7.874" × 0.005" – 0.250" thick	Notch/no notch	N/A	–
300 mm × 0.127 mm – 6.35 mm thick	11.811" × 0.005" – 0.250" thick	Notch/no notch	N/A	–

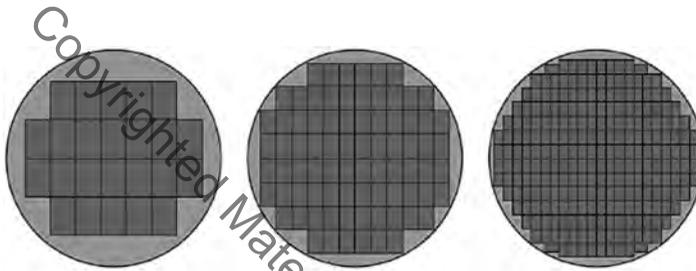


FIGURE 4.28 Placing of integrated circuits of different dimension on a silicon wafer with a simple rule: the center of the wafer coincides with an angle among four chips.

matured in microelectronics and integrated optics industry, the yield data have to be considered to be only indications of what would happen in a real production line.

4.4.1 INDUSTRIAL COST MODELS

In this section, we introduce the general cost models we will refer to in the rest of this section. We will introduce the models intuitively explaining their foundations. The interested reader can find a much detailed discussion in specialized publications [42–44].

4.4.1.1 General Expression of the Cost of Produced Goods

The main elements to consider modeling the cost of produced goods are the following:

- Cost of subparts that are acquired by suppliers to be mounted into the product. This is generally called bill of material (BOM) cost.
- Cost of work, both variable and structural.
- Production infrastructure depreciation.
- Fixed operating costs (e.g., maintenance and repair of production lines).

The different terms depend differently on the number of produced object (generally called production volume) and on how this volume increases with time.

Here, we call *volume* the number of produced items, both working and not working, so that the number of produced working items is the product between the volume and the yield. In literature, sometimes, the number of produced working items is called volume, and the total number of fabricated items is obtained by dividing the *volume* by the yield [3].

To determine the BOM cost, consider that a single product component has a unitary price that depends on the volume. At low volumes (where how much is low depends on the considered market), the cost strongly depends on the volume decreasing with it up to a critical volume; beyond the critical volume the cost stabilizes around a constant value. The asymptotic value is simply given by the bare high volumes' industrial cost of the component plus the suppliers' minimum survivability margin. A mathematical representation that can be adopted for the dependence of the BOM cost on the volume in a multi-supplier market is

$$C_{BOM} = \frac{C_{\infty} (V/V_0)^{\alpha} + C_0}{1 + (V/V_0)^{\alpha}} \quad (4.2)$$

where C_{BOM} is the BOM cost, V the produced volume, C_0 is the samples cost, C_{∞} the asymptotic cost, V_0 the threshold volume, and α a parameter depending on a particular type of product.

The labor cost, different from the BOM, at the first order does not depend on V , but for a small part related to the service functions that have to be present in the factory (repair squads for example). Thus, in a first approximation labor cost marginally depends on volumes and it can be written as

$$C_{Labor} = C_{L0} + \frac{C_{L1}}{V} \quad \text{with } C_{L0} \gg C_{L1} \quad (4.3)$$

The last term, composed of infrastructure depreciation and all the maintenance and operational fixed costs, does not depend per se from the production volume, but it is a characteristic of the production infrastructure (the factory, the headquarter, and so on). Thus, when this cost is distributed among the produced items it contributes to the industrial cost of each individual item with a term inversely proportional to the volume as

$$C_{Dep} = \frac{C_{D0}}{V} \quad (4.4)$$

Equation 4.4 assumes that the factory is never saturated by the production volume: if factory saturation is considered, the model has to be changed to introduce the need or removing saturation bottlenecks via progressive investment in fabrication equipment. This is not a case relevant for our discussion, but it can be very important to analyze the overall performance of a realistic production structure.

Summarizing the industrial cost C_{TOT} of an item coming out from mass production is the sum of three terms, C_{BOM} , C_{Labor} , C_{Dep} , and each of these terms has a characteristic dependence on the produced volume.

4.4.1.2 Different Cost Models

Depending on the production process, different products have a different mix among the term constituting the industrial cost, so generating different cost models.

Example from different markets are provided in Table 4.8, where the reported figures have to be considered as general indications, since they can fluctuate also sensibly depending on the geographic area and on the specific company. From the table it is clear that in the first two considered industries costs are dominated by the materials (they are BOM driven), in the third case the industrial cost is dominated by the cost of labor (it is labor driven), and in the fourth case, the most typical case of very high volume planar chip industry, it is dominated by depreciation and fixed costs (it is depreciation driven). As a matter of fact, a modern 200 mm foundry costs about \$100–200 million, and a microelectronic factory based on 32 nm technology and on 12' wafers requires an investment greater than \$1 billion.

TABLE 4.8
Example of Product Having Different Cost Models

	Volumes (One Plant)	Cost Terms at Saturated Plant		
		Labor Cost (%)	BOM–Materials Cost (%)	Depreciation and Fixed Costs (%)
Structural products manufacturing (plates and bars)	600,000 pieces	22	63 ($\alpha \approx 1$)	15
Optical access interface at 1.2 Gbit/s	2 million pieces	33	50 ($\alpha \approx 1.5$)	17
Clothes industry (far east)	1.3 million pieces	54	31 ($\alpha \approx 0.5$)	15
RAM memory chip	1 billion pieces	5	6	89

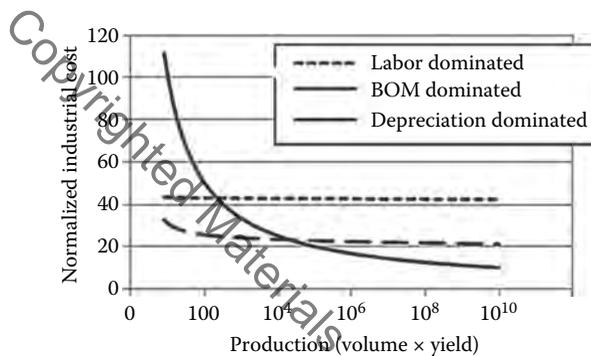


FIGURE 4.29 Industrial cost versus volumes for products having different cost models.

The dependence of the industrial cost on volumes is shown in Figure 4.29, for the different types of cost models we have introduced. In particular, planar integrated chips have a very high industrial cost for low volumes due to the great initial investment, but it scales much better than other technologies at the increase of the produced volume, up to allow the production of very high volumes with small prices. This is clear from the figure, even if the parameters have been changed a bit with respect to reality to be able to draw a clear plot (in the case of microelectronic the curve is much more steep, almost vertical in the scale of the graphic).

4.4.2 INDUSTRIAL COST ESTIMATION FOR MICROFLUIDIC-BASED LABS ON CHIP

To achieve a real cost estimation, in principle, we should exactly define the product we are dealing with. However, what we want to achieve here is not an impossible series of exact numbers, but a trend indication and in order to do that we will perform a set of hypothesis with the aim on one hand to simplify the evaluation, on the other, to maintain main trends by neglecting all the details that are specific of the single product and do not impact the general trend.

4.4.2.1 Industrial Cost Parameters

The lab on chip we are considering is a combination of two technologies: a detection technology (e.g., integrated optics if we perform optical detection) and microfluidics. Assuming for simplicity interferometry optical detection [4] the typical scale of both the technologies is the micron. As we will see in the next sections, in order to maintain the required definition at this kind of dimensions there is no need of critical technologies, but standard and well-experimented processes and

machines can be used. On the other hand, complexity is introduced in the process by the need of integrating two different technologies and by the integration of specific technology steps, like wafer bonding, in the production flow.

The cost of the final product, a packaged chip ready to be integrated in a more complex system, is provided by the following equation:

$$C_{component} = \frac{C_{fixed}}{V_{produced}} + (BOM_{package} + W_{package} + C_{package}) + M_{operating} + \frac{2}{Y_{ield} N_{perwafer}} (C_{wafer} + C_{wafer\ process} + W_{wafer\ fab}) + \frac{C_{Test}}{N_{cam}} \quad (4.5)$$

where

- The fixed structure costs (administration, direction, R&D, and so on) divided for the volume is represented by the term $(C_{fixed}/V_{produced})$
- The cost of the back-end part of the fabrication, once eliminated the depreciation contribution, is constituted by the sum of three terms
 - The bill of materials ($BOM_{package}$) that takes into account all the elements that are acquired by external suppliers like the package itself, a heater where it is needed, and so on
 - The work cost ($W_{package}$) that is generally important in the packaging operations due to the fact that all the processes are executed chip by chip
 - The other variable costs (like the cost of technical gases and services) are represented by $C_{package}$
- The part of the cost due to the management of the production line that depends on the volume (like visual inspection of the produced wafers, in-line tests, and so on) are represented with $M_{operating}$
- The number of produced chips per wafer ($N_{perwafer}$) and the average wafer yield (Y_{ield}) are two important elements to evaluate the chip cost, since due to its nature the cost of the front-end fabrication depends on the number of wafer, thereby being lower and increasing the number of chips on a wafer and the yield
- The industrial cost of a wafer at the end of the front-end production line is made by three components:
 - A cost of the wafer itself (C_{wafer}); here factor 2 reminds us to take into account the wafer bonding process
 - The cost of work and materials needed in the fabrication process ($C_{wafer\ process}$)
 - And the impact of the facility depreciation ($W_{wafer\ fab}$)
- The cost of the final test is indicated with C_{Test} and, if the final test is done on a reduced number of produced chip, this can be taken into account by assigning to N_{cam} a value greater than 1

Using what we have derived in the previous section regarding the expression of different cost terms, the cost of a chip can be rewritten starting from Equation 4.5 detailing the dependence of each term on the volume V as

$$C_{component} = \frac{C_{fixed}}{V_{produced}} + (BOM_{package} + W_{package} + C_{package}) + M_{operating} + \frac{C_{Test}}{N_{cam}} + \frac{2}{Y_{ield} N_{perwafer}} \left[\left(C_{W1\infty} + \frac{C_{W10}}{1 + N/N_0} + \frac{C_{w1lav}}{N/N_{0p}} \right) + \left(C_{W2\infty} + \frac{C_{W20}}{1 + N/N_0} + \frac{C_{w2lav}}{N/N_{0p}} \right) + \frac{C_{wbond}}{N/N_{0p}} \right] \quad (4.6)$$

The meaning and the assumed value for all the parameters in Equation 4.5 and reasonable values deduced from a detailed process analysis on a certain number of different labs on chip using interferometry optical detection are reported in Table 4.9. The used material is assumed to be silica passivized silicon so that standard 200 mm silicon wafers are used as a starting point.

Moreover, in this example, we have assumed a relatively small chip considering a lab on chip, a chip that is only 1.0×1.0 cm. Working with wafers with a radius of 200 mm about 270 chips per wafer are usually fabricated. Scaling the cost with the chip dimension is in any case easy, taking into account the different situations that are the dominant cost factors. Using the parameter values reported in Table 4.9 it is possible to evaluate the chip cost dependence on selected parameters, the volume and the yield being perhaps the most important.

The details of the various cost contributions are reported in Figure 4.30. The first element clearly emerging from the figure is that, due to the different decreasing rate when volume increases, above a volume of about 900,000 good pieces, the dominant cost term is the back-end BOM, remaining from this volume on a level a bit below 1 in the adopted relative units. Below this threshold volume, the chip cost dominates, with a very important contribution of the back-end line depreciation.

The estimate of the final industrial cost is shown in Figure 4.31 for different areas of the lab on chip. In the example of the chip we have considered up to now, industrial cost is below 2.0

TABLE 4.9
Definition and Values of the Parameters Used in the Example of Cost Evaluation for a Microfluidics Biosensor

d	Chip side (the chip is assumed squared)	1.0 cm
N_{poichipp}	Wafer area divided chip area	315
N_{perwafer}	Number of chips per wafer (bad and good)	270
V_{ref}	Reference volume for the cost first calculation	1,000,000 chips
N	Number of wafer per year (8' silicon wafers)	3700
BOM_{package}	Back-end BOM cost (at reference volume)	AU 1.3
W_{package}	Work cost per chip of the back-end process	AU 0.3
C_{package}	Other variable cost of the back-end per chip	AU 0.1
$M_{\text{operating}}$	Variable part of the production line management (cost per chip)	AU 0.1
C_{fixed}	Back-end fixed costs	AU 3,000,000
N_0	Threshold wafer number for suppliers	1000
N_{0p}	Threshold wafer number for depreciation scaling	100
N_{cam}	Final test period	500
M_{fab}	Fab margin on chips	60%
Y_{ield}	Yield (on the chip only, high volumes production)	80%
$C_{W1\infty}$	Wafer 1: wafer cost for infinite volume	AU 30
C_{W10}	Wafer 1: fluctuation of the wafer cost with volumes	AU 40
$C_{W2\infty}$	Wafer 2: wafer cost for infinite volume	AU 60
C_{W20}	Wafer 2: fluctuation of the wafer cost with volumes	AU 40
C_{w1lav}	Wafer 1: impact of fixed costs and depreciation	AU 1130
C_{w2lav}	Wafer 2: impact of fixed costs and depreciation	AU 2450
C_{wbond}	Cost of the bonded wafer minus cost of the two wafer components	AU 200
C_{Test}	Cost of an end-line chip test	AU 10

Note: AU indicates an arbitrary unit of measure for unit costs that is almost of the order of magnitude of a U.S. dollar or of an EU euro. This is used instead of a true monetary value to report results in terms of relative costs and values that are independent from economic conditions, change ratios, and the like.

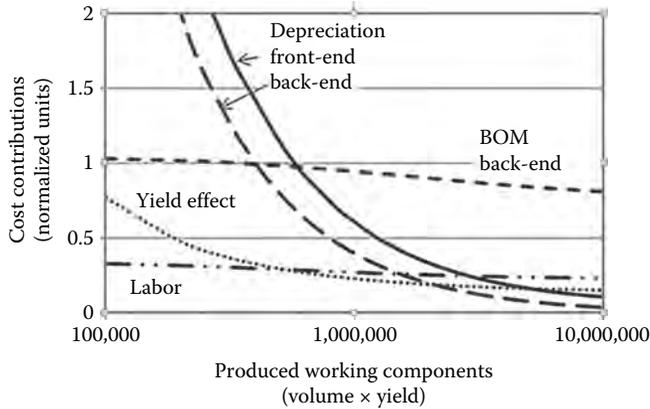


FIGURE 4.30 Industrial cost of the microfluidic circuit considered in the example in the text decomposed in the various cost contributions for different numbers of correctly working produced chips (volume × yield).

(in relative units) for a volume greater than 10 million pieces and below 1.5 above about 33 million pieces. The decrease rate is quite fast at the beginning, when the depreciation is important, and slow down for very large volumes where the cost is dominated by the BOM. By increasing the area the cost increases almost proportionally to the chip area in the region where depreciation dominates, while experiences a much smaller increase in the BOM-dominated region. The most important factor however is that, increasing the area, the threshold dividing the depreciation and the BOM-dominated regions moves toward upper volumes, thus it is more and more difficult to reach.

The industrial cost is also compared with an electronic IC built with the processes of a high-density RAM considering also in this case a 200 mm fab. The RAM has a lower cost even with respect to smaller labs on chip while it also decreases more with volumes, tending at a far lower limit. The first phenomenon is a combination of several elements: the use of two wafers due to wafer bonding, a greater variety of used processes and a less stabilized fabrication procedure causing a lower yield.

As far as the second phenomenon is concerned, it is a very important characteristic of labs on chip, where the back-end cost is always much more important than in the case of electronic circuits. This effect is essentially due to the back-end process. In electronic circuits, the package itself is a commodity, little more than a plastic case where the circuit is enclosed, and the contact with the outside world are simple electrical wires. The situation is completely different in labs on chips. Almost

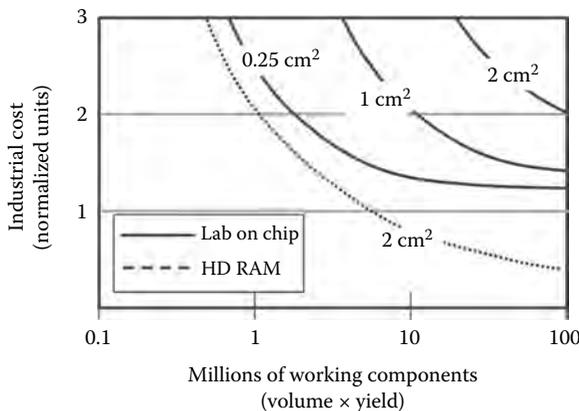


FIGURE 4.31 Overall industrial cost of the of the microfluidic circuit considered in the example in the text compared with similar microfluidics circuits with a greater area and with the cost of a high density RAM.

always the package has to host complex interfaces that are needed both by the fill chambers and repositories with reagents when the chip is fabricated and to charge the test solution when the chip is used. If optical detection is adopted, optical interfaces have to be frequently fabricated.

Another element that also impacts the asymptotic cost of labs on chips is the fact that they have to be frequently charged with chemical reagents that have the role of test, references, or washing solutions. Even if the amount of chemicals that are inside the chip is very small, when cents are to be taken into account, also this element that does not exist in the electronic case is important.

Even if the cost of labs on chip circuits does not scale exactly as in the case of microelectronics, the huge volume available in the health environmental and food control industries renders planar technology very effective in reducing the cost of circuits implementing complex functionalities, especially when compared with BOM-dominated analytical chemistry equipment.

4.5 PHOTOLITHOGRAPHY

Photolithography is the first fundamental step in planar front-end technology, allowing the transfer of patterns created on a mask on the photoresist layer placed on the wafer. Photolithography is a multistep process [1,3], starting from the wafer preparation and ending with the photoresist development.

A general scheme of a photolithography process is shown in Figure 4.32.

Photolithography, being essentially an optical process, has definition limits due both to the photoresist discrete nature and to the limited value of the used wavelength. Several other lithographic processes are industrially adopted or under study to overcome these limitations and high definition lithography processes based on particles beams or x-rays arrive to a sensitivity of the order of 2 nm. However, this extreme definition is not generally needed in labs on chip and the interested reader is

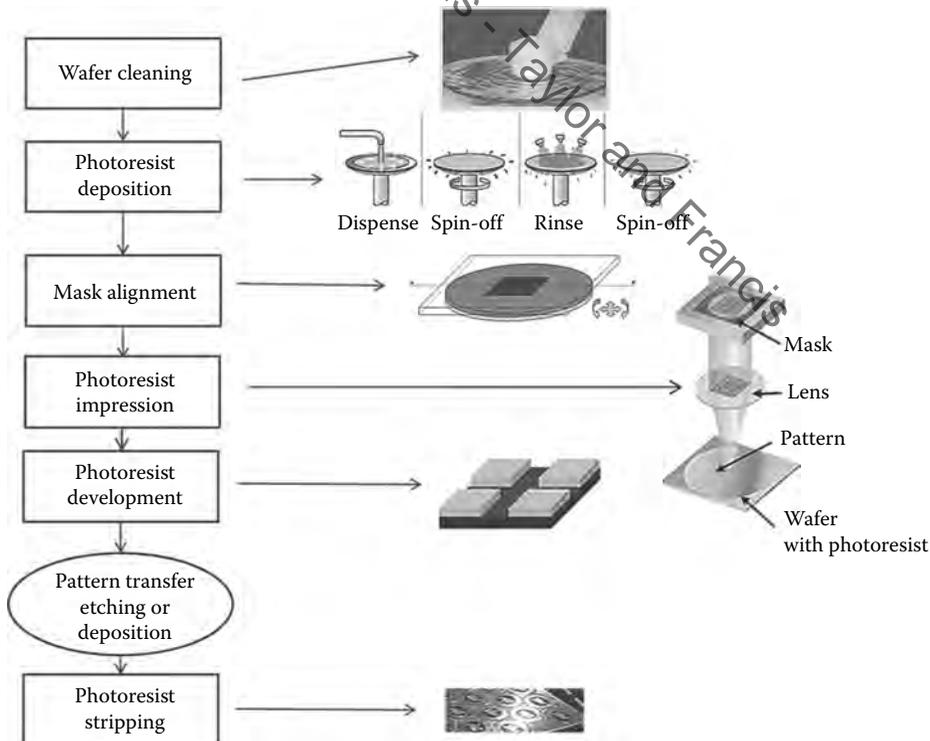


FIGURE 4.32 Photolithography process flow and subsequent transfer of the pattern on the wafer by etching or deposition.

encouraged to refer to the bibliography to start an analysis of such advanced lithography techniques [1,45–47].

4.5.1 WAFER CLEANING

Even if the wafer surface is protected from contaminants during all its history from production to the moment in which it starts the fabrication process, a preliminary wafer cleaning is needed before every mask around, that is, at the beginning of every lithography process. Various contaminants can be present on the wafer, residues of previous process steps or even from the wafer fabrication. Examples are solvents or other process chemicals, particles from operators and equipment, and other contaminants that could be present even in the air of a clean room. If left on the wafer surface, these contaminants can lead to serious problems during wafer processing, like compromising wafer planarity during lithography, creating unperfected adhesion of the resist on the wafer surface, or provoking unwanted chemical reactions.

All the clean room structure is built so as to cause particles that could hit the wafer surface not to exceed a limit energy so that wafer damage due to mechanical hit energy is not probable. Moreover, particle adhesion to the wafer surface is generally weak, so that an accurate cleaning process can remove the great part of the contaminants without damage.

Both wet and dry cleaning procedures are used in clean rooms. A well-known set of wet cleaning protocols called RCA1 and RCA2 [1] use mixtures of hydrogen peroxide mixed with various acids or bases to perform wet wafer surface cleaning combining chemical and mechanical processes. At the end of a wash with deionized water assures that no excess of the cleaning solution remains on the wafer surface. Other adopted procedures simply used vapor cleaning and high-temperature backing (up to 1000°C, depending on the resistance of the structures still on wafer to high temperature) in an atmosphere rich in oxygen. Alternatively, high-temperature backing can also be performed in vacuum, to favor contaminants detaching from the wafer surface.

Considering dry cleaning procedures, they use either ultrasonic beams, that are very good to remove particles from the wafer surface, but unfortunately are prone to the risk of damaging the surface itself, or simple wafer abrasion with suitable polishing compounds.

Generally, wet cleaning is preferred to dry cleaning that sometimes is used simply as a preliminary cleaning procedure before the more effective wet process.

As a last observation, it is to be noted that the wafer back also is generally cleaned, since several processes rely on the back planarity and the absence of defects, for example, to fix the wafer to the process machine plate, or for other reasons.

4.5.2 PHOTORESIST DEPOSITION

The photoresist deposition is the most expensive step of photolithography, both due to the cost of the resist itself and to the fact that almost all the deposition techniques have a relevant waste of material.

Thus, there are several used photoresist deposition techniques, both to adapt them at different kinds of resists and to optimize the process under a cost point of view.

4.5.2.1 Photoresist Spin Coating Deposition

Perhaps, the most used deposition technique is the spin coating [48]. Generally, before spin coating the cleaned wafer is protected by growing a thin silica film on it by oxidation. The film is annealed between 800°C and 100°C so as to homogenize its properties and it constitutes the base where the photoresist is deposited and, in selected cases, also the mask for subsequent processes like boron implantation.

After the oxidation the wafer is placed on a spinner plate and spin coating is performed (see Figure 4.33). An excess amount of a raw resist solution is placed on the substrate, which is then rotated at high speed in order to spread the fluid by centrifugal force. Rotation is continued while

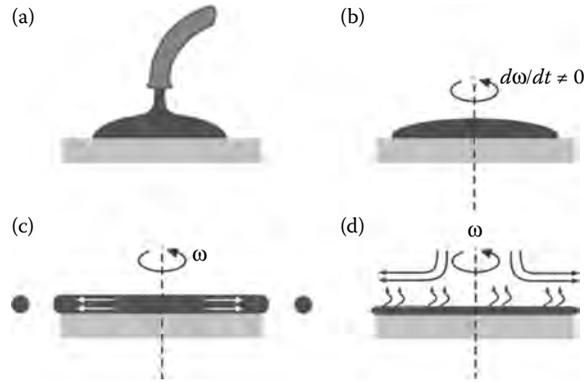


FIGURE 4.33 Photoresist spin coating; process phases (a) deposition of the coating fluid onto the wafer or substrate, (b) acceleration of the substrate up to its final, desired, rotation speed, (c) spinning of the substrate at a constant rate; fluid viscous forces dominate the fluid thinning behavior, and (d) spinning of the substrate at a constant rate causing solvent evaporation.

the fluid spins off the edges of the substrate, until the desired thickness of the film is achieved. The applied solvent is usually volatile and simultaneously evaporates. So, the higher the angular speed of spinning, the thinner the film. The thickness of the film also depends on the concentration of the resist solute. In the case of standard UV photoresist, typical peak values of the spinner rotation speed are from 1500 to 6000 rounds per minute and the process duration is in the range of 10–100 s.

The spin coating process can be divided into steps, as shown in Figure 4.33.

1. The wafer is placed on the spinner plate, where it is fixed in position by a vacuum chuck. Here, the planarity of the wafer back is important like the absence of big contamination particles, in order to avoid a too weak force bonding the wafer to the chuck with the possibility that the wafer moves during the process and the coating results to be not homogeneous.
2. While the wafer rotation starts with a fast acceleration the resist solution is dispensed on the wafer near the wafer center through a dispenser. The dispenser characteristics in terms of uniformity of the solution flux and control of the flux variation in time are important to assure the wanted resist thickness to be realized by the spin coating process.
3. After the acceleration, the wafer is left to rotate at maximum speed while the solution arrives at the wafer borders and spill over the wafer borders. Contemporarily, the solution solvent tends to evaporate.
4. In a second phase, at maximum rotation speed the solvent evaporation is fastened, typically controlling the temperature inside the spinner or a small air flux toward the wafer. This is the phase in which the film has to reach the wanted thickness.
5. After the last phase, the spin plate is uniformly decelerated. This deceleration takes place when the film is almost solid; so generally it does not alter the film properties in a relevant way. At the end, when the plate stops, the film is formed.

The final film thickness can be defined only after a certain spin time, where the spillover of the resist solution from the wafer has eliminated at least 40% of the deposited initial material. Under this condition, the thickness depends on several parameters, mainly on the spin speed, on the photoresist solution viscosity, and on the time the spinner works after the film becomes almost uniform. An example of film thickness versus the spinning time for two commercial photoresists is provided in Figure 4.34, where it is evident as the thickness tends in any case to a limit value that cannot be further decreased while increasing the spinning time.

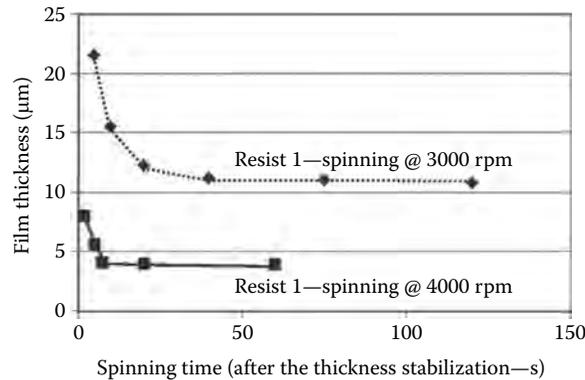


FIGURE 4.34 Example of film thickness versus the spinning time for commercial photoresists.

There are several theories trying to derive equations from the limit thickness δ of the film, but in general either they arrive to complex differential equation systems [49] or they introduce approximations that reproduce the phenomenon dynamics but do not allow to obtain accurate quantitative results. In practical processes, an empirical formula is often used that is given by [1,49]

$$\delta = K c^{\alpha_c} \mu^{\alpha_\mu / 3} \omega^{-\alpha_\omega / 2} \quad (4.7)$$

where

K is a calibration constant depending on the specific equipment

c is the initial photoresist solution concentration

μ is the solution viscosity

ω the stationary spinning rate

α_c , α_μ , and α_ω are three empirical constants of the order of one that has to be determined experimentally

When the constants have been experimentally determined by a calibration set of measurements, Equation 4.7 is generally sufficiently accurate to allow the thickness of photoresist films to be correctly set by setting the spinner parameters.

After spin coating, the remaining unwanted solute is evaporated by wafer backing, that is generally performed at 90–100°C in a convection backing or at lower temperature (like 85°C) in a vacuum-backing chamber. The evaporation of the residual solvent reduces the film thickness to the final value, as shown, for example, in Figure 4.35.

Problems in spin coating can be due to several reasons, such as particles on the substrate at the moment of the spinning (causing streaks in the film like in Figure 4.36a), imperfect substrate wetting (causing nonuniform liquid expansion like in Figure 4.36b), or a too small initial amount of liquid (causing areas of the substrate to remain uncoated as in Figure 4.36c).

4.5.2.2 Alternative Photoresist Deposition Processes

Spin coating is not the only photoresist deposition used in micro-fabrication lithography, but several alternative techniques exist, that are reviewed in detail in References [1,3]. Here, we will briefly describe only two of them that are suitable in typical situations that could present during labs on chip manufacturing.

4.5.2.2.1 Spray Coating

In the so-called spray coating, the photoresist coating is created by passing the wafer under a spray of photoresist solution. In the spray equipment, the photoresist is pushed out of a pressurized tank

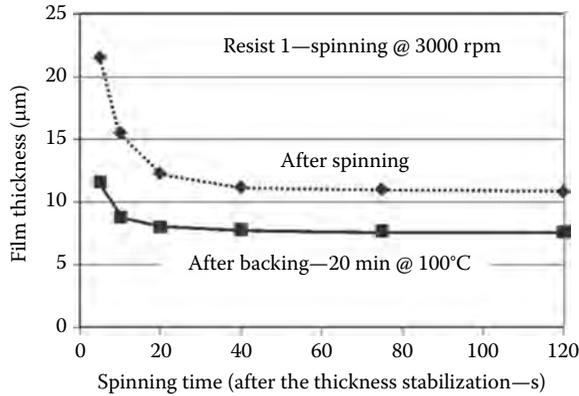


FIGURE 4.35 Resist film thickness before and after backing: example from a commercial photoresist.

via a supply pipe to a spray head that has a well-defined aperture and shape so as to form controlled droplets. Since the droplets are sprayed so that it attaches at the wafer where they hit it, the uniformity of the coating can be controlled by shaping the spray beam and suitably designing the motion of the spray head. In any case, the thickness control is not as good as in the case of spin coating and the spray head has to be maintained near the wafer (generally within 5 cm from the wafer surface).

While the film cannot be thin as in the case of spin coating, spray coating is particularly suitable for deposition on nonuniform surfaces, as it is the case of MEMs processing. In this case, the ability of the droplets to fill gaps and distribute so that a flat surface is created is important.

In cases in which such nonuniform surfaces are present, generally a low-speed spin is also added to the spray effect, for example at 30–60 rpm, to improve planarity of the outer surface of the photoresist film. Moreover, spray-coated films lack the internal stresses that are typical of spin-coated films so that the probability of film break after the coating process is lower. Finally, spray coating machine exists that can coat a wafer on the front and on the back, simultaneously. This operation very useful in both the front and the back of the wafer has to be processed as in the case of membranes fabrication shown in Figure 4.11—that is, a very important case in the field of labs on chip.

4.5.2.2.2 Electrophoresis Deposition (Electro-Deposition)

This is a technique used mainly when the substrate has relevant deviation from planarity, as it can be the case of microfluidics and MEMs fabrications. In the first case, excavations as deep as 50 µm can be conducted on silica in order to prepare reaction chambers and ducts so that, if successive lithography is needed, spin coating could not be usable due to the limited ability to produce a flat resist surface in this impervious conditions. In the second case, actuators and movable parts could not only have relevant vertical dimension, but they can also be fragile if a too strong force is applied to them in the horizontal direction.

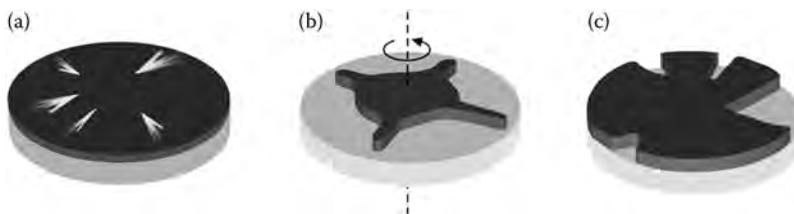


FIGURE 4.36 Selected possible defects in the resist film after spinning and backing: streaks in the film (a), nonuniform liquid expansion (b), and uncoated areas of the substrate (c).

To apply electrophoretic deposition, a conductive substrate is needed, besides an electrical biasing during the deposition process since the process is based on electrophoresis of the resist solution (see Sections 3.6 and 7.5.1). The coating solution is formed by charged micelles containing several components such as resist, solvent, dye, and photo-initiator molecules. When the electrophoretic field is established between a positively charged substrate and a negatively charged cathode, generally constituted by the wafer to be coated, the micelles migrate toward the cathode. When the micelles reach the cathode their positively charged surface is neutralized by hydroxyl ions produced by the electrolysis of water at the cathode and becomes unstable. After a very short transition time, the micelles coalesce on the surface of the cathode forming a self-limited resist film. Self-limitation happens due to the fact that the resist is not conductive. Thus when the film reaches a limit thickness the cathode results to be passivated by the resist film and electrophoresis ends, stopping the deposition of further resist on the substrate.

The coating thickness depends critically from the intensity of the electrophoretic field and on the process temperature that has to be maintained constant over all the substrate and for all the process time. Typical film thicknesses that are achievable with this method are 5–15 μm , but up to 35 μm thickness can be obtained with specific resist solutions.

Although the most important advantage of this deposition method is being compatible with relevant deviations from the substrate planarity, its main disadvantage is requiring an electrical biasing of the surface to be coated, that frequently means a metal coating of the same substrate before resist deposition, with additional process complications and costs. A comparison between spin coating, spray coating, and electrophoresis coating is reported in Table 4.10.

TABLE 4.10
Comparison between Spin Coating, Spray Coating, and Electrophoresis Coating for Resist Deposition

	Spin Coating	Spray Coating	Electrophoresis Coating
Process complexity	Simple	Simple	Complex
Process automation	Difficult	Possible (batch approach)	Possible (batch approach)
Key parameters	Viscosity, spin speed	Solution stoichiometry, spray speed, spray beam control, Scanning speed	Voltage, temperature
Needed photoresist solution	All types	Specific solutions with low viscosity	Specific prepared ED resist solution
Wasted material	More than 50%	20–35%	Depends on bath renewal, that is generally frequent
Planarity requirements	Very strict	Accept planarity deviations	Works with relevant deviations from planarity
Film uniformity	Difficult to control, need frequent equipment-specific calibration	Good control, depends on cavities position	Very good controls also without strong planarity
Film thickness before backing	Typical 5–20 mm	Typical 20–60 mm	Typical 5–20 mm, up to 35 mm possible
Applications	Transfer patterns on planar substrates or on the bottom of large cavities	Transfer patterns on substrates with geometrically similar cavities	Transfer patterns across and on the bottom of cavities, metal coating on substrate preferable
Cost	Intermediate	Intermediate	High

Source: Adapted from Madou M. J., *Fundamentals of Microfabrication and Nanotechnology*, Third Edition, Three-Volume Set, CRC Press; 3rd edition (August 1, 2011), ISBN-13: 978-0849331800.

4.5.2.3 Photoresist Types

A great part of the photolithography capability of producing small critical dimension patterns and of the related cost depends on the photoresist. The main components of a photoresist are

- A photo-sensitive polymer
- A sensitizer
- A casting solvent

The polymer is the component that changes its state when exposed to light of a suitable wavelength, the sensitizer allows control of the chemical reaction in the polymeric phase, and the solvent allows resist deposition to form a thin film over the target substrate. Resist without a sensitizer also exist, called single component resist, while resist including a sensitizer are called double component resists.

A photoresist weakened by light exposure, so that the exposed part can be easily removed, is called positive resist (or also positive tone resist) while a resist that is reinforced by the exposure to light, so that it is the nonexposed part that can be easily removed, is called negative resist (or negative tone resist).

The weakening of the polymer in positive resists is obtained by photons-induced scission of polymeric chains, while the reinforcement in the case of negative resist is due to photon induced cross-links in the polymeric chains rendering them less reactive and more sturdy under a mechanical point of view (cf. Section 1.4.2).

Two widely used families of positive photoresists are the polymethylmethacrylate (also called PMMA) [50], that is available both in single and in double component versions, and the two components diazonaphthoquinone (DNQ) [51,52].

PMMA undergoes a photo-stimulated chain scission, whose reaction is sketched in Figure 4.37, if illuminated with deep UV light, with a maximum sensitivity around 220 nm. Above 220 nm, the sensitivity goes rapidly to zero and the resist becomes light insensitive. In single component resist the PMMA is a slow resist, requiring doses of UV light greater than 250 mJ/cm² to arrive to a satisfactory degree of impression, corresponding to several minutes of exposition using high-power UV lithography systems. By adding a suitable photosensitizer, the UV sensitivity of PMMA is greatly increased so as to reduce the impression dose up to less than 150 mJ/cm². A problem of PMMA is its poor adaptability to plasma etching (PE). As a matter of fact, the polymer residue changes the chemistry of the PE, frequently bringing to a spurious polymer film deposition on the substrate.

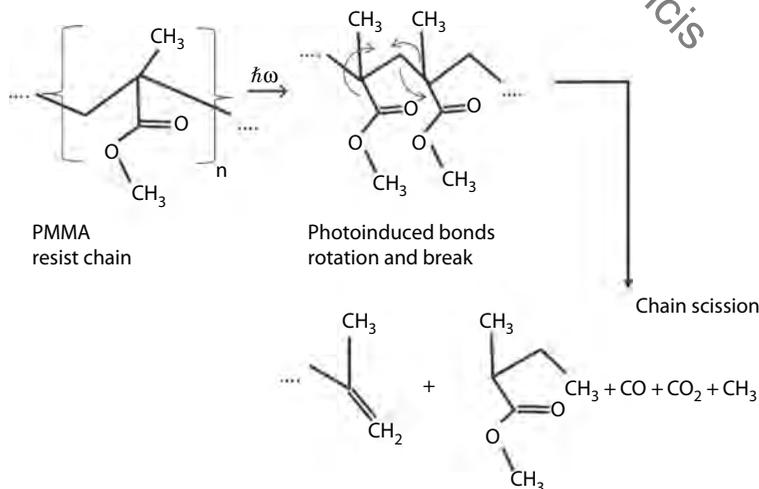


FIGURE 4.37 Photo-stimulated chain scission of PMMA resist.

The DNQ is a near UV resist that exists only in a two-component form and is widely used as far as its sensitivity is sufficient for the application. It is composed of the Novolak (N) resin, that in standalone form is a soluble alkali, and the diazonaphtaquinone (DQ), which is a hydrophobic compound. When the two components are joined together they form a practically insoluble and hydrophobic compound.

During exposition, photolysis eliminates the inhibition effect of DQ by breaking the composite and the resist develops a reaction forming a base-soluble carboxyl acid (see Figure 4.38). Typical UV wavelength used to expose DNQ are 365, 405, and 435 nm, that are emission lines of mercury lamps used in less performing lithographic equipment. Above 300 nm the DNQ becomes practically insensitive to UV light and cannot be used as resist. DNQ has the very good property that unexposed areas are practically unaltered by the development process, so that the shape of the mask is precisely transferred on the substrate and can be effectively used in conjunction with PE.

To conclude the discussion regarding positive resists, it is needed to underline that, in general, the adhesion of positive resists to the substrate, generally a silica film or, less frequently, pure silicon, is not sufficient for the greater part of the deposition processes to work properly. To improve adhesion, a primer has to be added before resist deposition [52].

In opposition to positive photoresists, negative resists becomes insoluble after UV exposition, so that the development allows the unexposed part of the film to be removed, while the exposed part remains in place. This can be achieved in one of two ways: either the resist increases its molecular weight by photo-stimulated cross-links so decreasing solubility or it undergoes photo-stimulated reactions that completely change its chemical structure to form a new insoluble molecule. This last mechanism is the most common in modern negative resists due to the absence of the so-called oxygen inhibition that typically happens in the first kind of resists.

This phenomenon, strictly related to the resist chemistry itself, consists in the presence of a side reaction involving oxygen that compete with the main reaction when the resist is developed and can completely ruin the pattern impressed on the resist by UV exposure. Thus, in order to avoid this side effect, exposure of this resist must be performed under a vacuum system or in nitrogen atmosphere. Oxygen inhibition is minimized when lithography is realized by putting the mask in strict contact with the resist film, since the mask itself protects the resist for oxygen residues that could be present in the atmosphere during development. In more performing photolithography machines, where the mask is not in contact with the resist, a very good atmosphere control is needed so complicating the process and increasing potential problems during development.

Another disadvantage of negative resist is that the film thickness limits the lithography resolution. This is due to the fact that resist hardness starts on the exposed surface, where the light hits

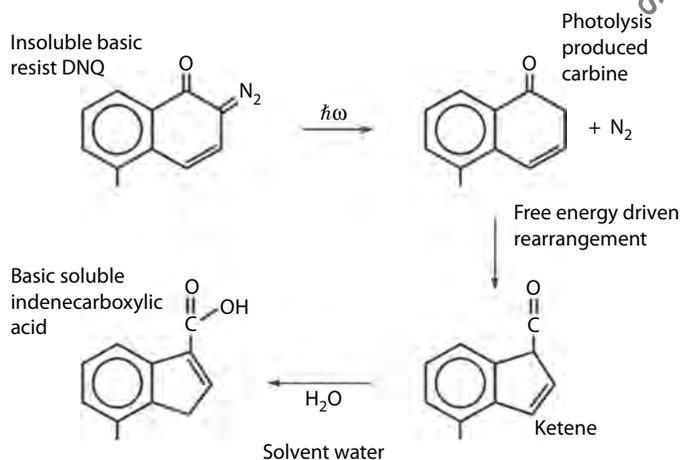


FIGURE 4.38 Photo-stimulated chain scission of DNQ resist.

TABLE 4.11
Synthetic Comparison between Positive and Negative Resists for Photolithography

	Positive Resist	Negative Resist
Adhesion to silicon	Primer generally needed	Primer not required
Backing	In air	In nitrogen or in vacuum
Contrast	High (e.g., 2.2)	Intermediate (e.g., 1.5)
Cost	High	Low
Developer	Temperature sensitive, water based	Temperature insensitive, based on organic solvents
Oxygen inhibition	Never	Frequent
Mask type	Dark field—higher performances	Clear field—lower performances
Photospeed	Slower	Faster
Pinhole count	Can be high	Have to be maintained small
PE resistance	Not very good	Very good
Residue after development	Mostly at $<1 \mu\text{m}$ and at high-aspect ratio	Can be a problem
Resolution	High (lower than $1 \mu\text{m}$)	Lower (greater than $1 \mu\text{m}$)
Swelling in developer	No	Yes
Thermal stability	Good	Fair
Wet chemical resistance	Fair	Excellent

Source: Adapted from Madou M. J., *Fundamentals of Microfabrication and Nanotechnology*, Third Edition, Three-Volume Set, CRC Press; 3rd edition (August 1, 2011), ISBN-13: 978-0849331800.

the film, and a strong overexposure is needed to impress all the film thickness, whose time duration and needed dose depends on the film thickness. A greater over exposition, however, causes a great dose of scattered radiation that creates spurious exposure of different parts of the resist film thereby limiting lithography resolution. In a practical situation, a film with a depth of $1 \mu\text{m}$ does not permit a resolution better than $2 \mu\text{m}$. In order to improve resolution thinner films have to be used, but when this solution is adopted pinholes can provoke problems.

A general comparison between positive and negative resists is synthetically presented in Table 4.11.

4.5.2.4 Permanent Resists

Generally, the resist is a sacrificial material used in the lithography step and then stripped out at the end of the process. However, permanent resists do exist, whose layer enters to constitute a permanent component of the microcircuit. In microelectronics they are generally used as insulating layers or as flexible layer in multi-chip assembly. The two most common permanent resists are based on polyimide and SU8 polymers.

Commercial products based on polyimide are supplied as water soluble polyamic acid which undergoes a thermal reaction transforming it in an insoluble polyimide reacting with the solvent water. Important characteristics of polyimide are its resistance to temperature and high glass transition temperature that results in higher than 300°C . For these characteristics, it is often used as insulating material between conducting layers. Absorption of moisture, however, causes polyimide to swell so that its dielectric constant increases significantly. To avoid this problem several similar permanent resists have been developed [53]; generally adding other organic components. This greatly limits the increase of the dielectric constant, but often also decreases the chemical resistance of the resist polymer.

Both photosensitive and nonphotosensitive polyimide exists. Photosensitive polyimides are impressed by UV light at 436 and 365 nm, having a strong sensitivity decrease generally above 300 nm.

These polymers can be used to reduce the number of process steps constituting contemporary the resist used for lithography and the permanent layer used for insulation or a stress relief layer.

The other different kinds of permanent resist categories are those based on the SU8 polymer: these are deep UV resist that can be deposited in extremely thick layers, like 100 μm or more, and are very used in MEMs and other lab on chip technologies where 3D structures have to be built using support layers that are electrical insulating, mechanically sturdy, and chemically inert. SU8 is an acid catalyzed negative photoresist made by dissolving SU-8 resin [54,55] in an organic solvent and adding a photo-initiator. The viscosity of the solution, and thus the thickness of the film, depends on the SU-8 concentration while the photo-initiator makes the role of catalyst in the photo-induced reaction.

4.5.3 MASK ALIGNMENT

Mask alignment is an important step of the photolithographic process: not only has the mask to be perfectly superimposed to the wafer area, but since every practical component requires several subsequent lithography steps, several subsequent masks and alignment of the mask with the structures already present on the wafer are needed. It derives that mask alignment is increasingly critical while the critical dimension of the structures fabricated on the wafer decreases.

Mask alignment is performed by building suitable alignment structures on the wafer during the previous lithography run, that are used to align the following mask. In low resolution lithography, alignment is simply performed by looking at the alignment patterns through similar patterns realized on the mask with a microscope. This technique has two drawbacks: the precision is limited by the accuracy of the observation and it cannot be automatized to process wafer batch without the intervention of the operator. Thus, in high throughput and low critical dimensions fabs mask alignment is performed by automatic machines using pattern recognition algorithms [56–58] to align the pattern on the mask with the reference on the wafer.

An example of the process of pattern recognition during mask alignment is reported in the scheme of Figure 4.39. Depending on the exposure technique, pattern recognition can exploit the resolution enhancement provided by the photolithography projection optics and achieve a huge precision that arrives to tens of nanometers in ultra-deep UV lithography adopting phase masks. Moreover, the feature on the mask used for mask alignment may be transferred to the wafer or not by the lithographic equipment. In the last case, it may be important to locate the alignment marks such that they do not affect subsequent wafer processing or device performance.

Alignment marks may not be arbitrarily located on the wafer, as the equipment used to perform alignment may have limited travel and therefore only be able to align to features located within a certain region on the wafer, as shown in Figure 4.40. Typically, two alignment marks are used to align the mask and wafer, one alignment mark is sufficient to align the mask and wafer in x and y , but it requires two marks, preferably spaced far apart, to correct for fine offset in rotation (cf. Figure 4.40).

4.5.3.1 Types of Masks for Planar Lithography

Normal lithography masks can be built in a variety of materials and are realized by means of computer high-definition methods so as to achieve the wanted resolution. They reproduce the pattern to be transferred to the resist by the UV exposure in a 1:1 geometric proportion in the case of proximity and contact exposure, or with an enhancement from 1:4 to 1:10 when projection exposure is adopted.

Corresponding to the existence of positive and negative resists, clear and dark masks exist, respectively, passing through the UV light where the mask pattern exists or blocking it in correspondence of the patterned areas.

The definition of a mask is defined as the standard deviation of the distance between the ideal position of a point of the contour of the profile that should be written on the mask and its real position. Thus, the mask definition is only one of the several terms contributing to limit the photolithography

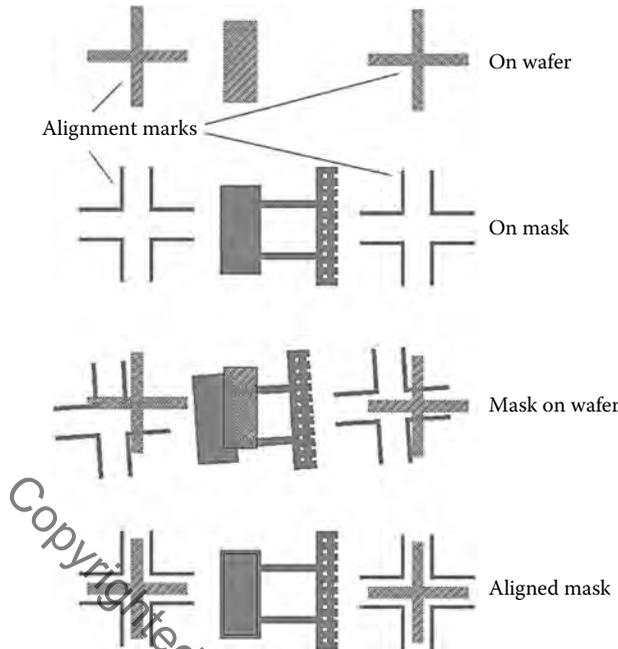


FIGURE 4.39 Example of the process of pattern recognition during mask alignment.

definition, and generally not the most important. Nevertheless, starting from a very good mask definition is important since in several lithography techniques the mask definition is not reduced by the definition enhancement technology, differently from other definition limiting factors. An example of this effect is found in projection exposure lithography where the mask definition is multiplied by the exposure ration in the same manner of the nominal mask pattern.

Mask definition depends both on the mask writing technique and on the mask material. When high definition computer programs are used to write the mask the last factor is by far the most important.

Different processes can be implemented on the photo-mask to improve the pattern transfer precision. An example is proximity correction, which is a process that alters the image impressed on the mask to correct the distortion provoked by proximity exposure (see Section 4.5.4). An example of mask realized by using proximity correction is reported in Figure 4.41. The thicker dark areas on the mask constitute the pattern that has to be printed on the wafer. The thinner lines are assists that

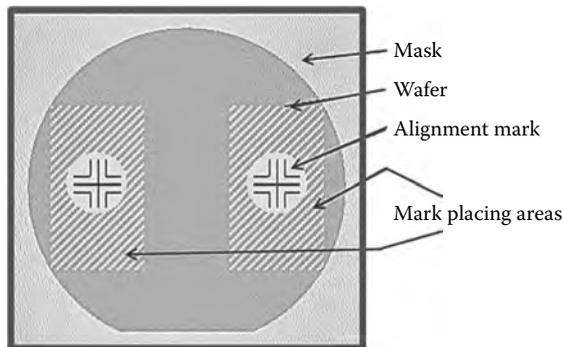


FIGURE 4.40 Example of placement of alignment marks on a mask and on the corresponding wafer.

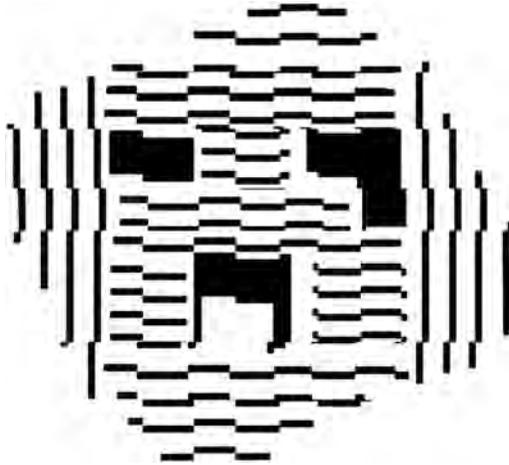


FIGURE 4.41 Example of photo-mask for proximity exposure corrected to enhance definition.

do not print themselves, but help the integrated circuit print better out of focus. The zig-zag appearance is due to optical proximity correction application.

A typical tolerance of glass masks is around $\pm 0.8 \mu\text{m}$, arriving to half this value if the pattern is carefully chosen to minimize the mask tolerance. Glass and quartz photo-masks are thus particularly suitable for applications requiring patterns smaller than $10 \mu\text{m}$. Alternatively, if a less demanding process, the photo-mask can be built in a plastic film, achieving a typical mask resolution around $5 \mu\text{m}$.

Particle contamination can be a significant problem if particles deposit on the photo-mask. Thus, generally, the mask is protected from particles by a pellicle—a thin transparent film stretched over a frame that is glued over one side of the photo-mask. The pellicle is far enough from the mask patterns so that moderate-to-small-sized particles that land on the pellicle will be too out of focus to be reproduced by lithography. Although they are designed to keep particles away, pellicles become a part of the imaging system and their optical properties need to be taken into account.

When the required lithography definition is very low, phenomena different from the mask resolution dominate the photolithography definition (see Section 4.5.7). One of them is the diffraction of the used UV light at the border of the mask pattern, whose effect is to project on the resist not a sharp image of the pattern border, but a thick diffraction figure, like it is shown in Figure 4.42.

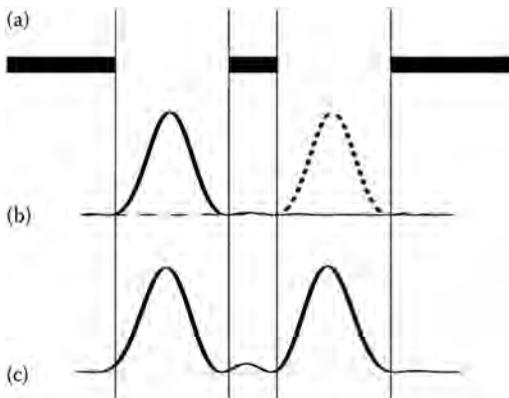


FIGURE 4.42 Effect of interference on lithography resolution if a normal mask is used; (a) mask pattern, (b) individual interference fringes of the two apertures on the mask pattern, and (c) overall optical power on the resist surface.

Here, the diffraction figures created individually on the resist surface from two apertures in a clear light mask (in a simplified bi-dimensional cases) are shown, to evidence how the light projected on the surface is not a sharp shape, but includes interference lines. The optical fields have to be added in amplitude on the resist surface taking into account that, due to the small details we are looking at on the mask, the phase of the field is almost the same.

Since in the figure the two apertures are 200 nm wide and the light wavelength is assumed to be 365 nm, we are still above the diffraction limit at which the critical dimension is half the wavelength, but the overall resolution is already insufficient. This is due to the lack of control of the phases of the optical fields interfering on the resist.

A phase mask is a mask built taking into account the characteristics of the lithography projection equipment so as to control in every point of the resist surface the phase of the interfering field in order to minimize spurious interference lines. This is accomplished by superimposing to the amplitude filtering characteristics of the mask a phase filtering characteristics that changes also the phase of the field traveling through the mask [59–61]. The effect is schematically shown in Figure 4.43, in the same example of Figure 4.42. Here, the field traveling through the second aperture of the mask is also inverted in phase, so that it subtract from the field coming from the first aperture once arrived on the resist surface. The effect on the final resist exposure is dramatically improved, not only does the intermediate spurious interference line disappear due to destructive interference but also the exposure is much more constant within the aperture area so that the resolution is completely sufficient for the resist exposure without defects.

Different types of phase masks exist. In alternating phase-shift masks, certain transmitting regions are made thinner or thicker. That induces a phase-shift in the light traveling through those regions of the mask creating a phase filter completely superimposed to the amplitude filter. When the thickness is suitably chosen, the interference of the phase-shifted light with the light coming from unmodified regions of the mask has the effect of improving the contrast on some parts of the wafer, which may ultimately increase the resolution on the wafer. The ideal case is a phase shift of 180° , which results in all the incident light being scattered. However, even for smaller phase shifts, the amount of scattering is not negligible. It can be shown that only for phase shifts of 37° or less will a phase edge scatter 10% or less of the incident light.

Attenuated phase-shift masks employ a different approach. Certain light-blocking parts of the mask are modified to allow a small amount of light to be transmitted through (typically just a few percent). That light is not strong enough to create a pattern on the wafer, but it can interfere with the light coming from the transparent parts of the mask, with the goal again of improving the contrast on the wafer.

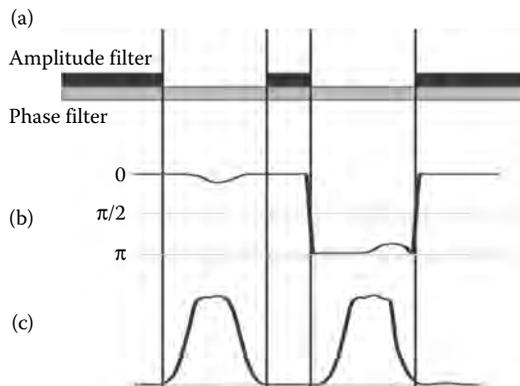


FIGURE 4.43 Effect of interference on lithography resolution if an optimized phase mask is used; (a) mask pattern, (b) phase filter characteristic, and (c) overall optical power on the resist surface.

4.5.3.2 Three-Dimensional Photolithography Masks

Up to now we have reviewed photolithography masks for planar pattern implementation. In microfluidics and MEMS technology, the realization of 3D structures is also very important. Creating vertical structures like chambers and ducts can be done via planar lithography and etching, but if inclined surfaces have to be created, using standard photolithography requires inhomogeneous etching techniques, like some types of wet etching (see Section 4.7.1) that imposes strong limits to the shapes that can be fabricated.

A possible alternative is the use of a specific type of photolithography based on a gradual exposure of the resist so that the subsequent etching is modulated by the remaining resist thickness and penetrates for a different depth into the wafer in different points. Current 3D photolithography technologies can be divided into three groups: multistep, direct-write, and grayscale mask photolithography.

Multistep photolithography utilizes several exposures using conventional optical masks [62,63], but each exposure produces a different gray level in the photoresist. With this technique, n masks have been demonstrated to produce $n + 1$ gray levels.

The maskless direct write process uses a writing beam to directly transfer a variable dose pattern into the photoresist [64–66], generally the beam is generated by an electron beam lithography equipment so that we will review this technique dealing with electron beam lithography.

Grayscale mask photolithography uses a conventional photolithography tool with a specialized grayscale optical mask [67,68]. Grayscale masks contain variable-transmission patterns that transmit part of the UV light intensity to create variable relief structures.

4.5.3.2.1 Multistep Lithography

This technique consists in using different masks to expose the photoresist varying the dose gradually along a certain direction. A schematic example of multistep lithography using a negative photoresist is reported in Figure 4.44. The key advantage of multistep lithography is to simply require a repetition of standard lithography steps: neither special lithography tool nor special masks are required. However, this simplicity corresponds to poorer performances with respect to the other methods of three-dimensional lithography.

Mask alignment is particularly critical in this technique; moreover, the limited contrast at the edge of each mask causes a deviation of the wanted dose profile on the resist that limit the critical dimension of the three-dimensional form to be realized. Last, but not least, deviations in booth alignment and exposure of each mask sums in the final result so that the overall process is quite less accurate than a single mask exposure using the same mask and the same lithography tool.

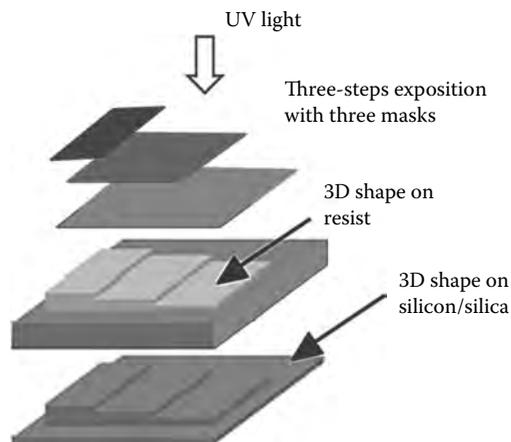


FIGURE 4.44 A schematic example of three-steps three dimensional lithography using negative photoresist.

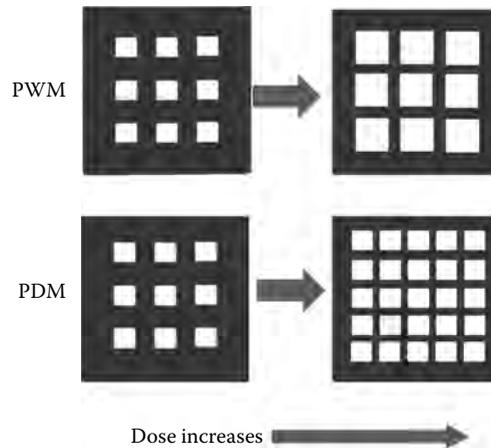


FIGURE 4.45 The structure of PWM and PDM masks for gray tone lithography.

4.5.3.2.2 Gray Tone Lithography

To overcome the limitations of the multistep process, gray tone lithography can be used, where different exposures are achieved using a single mask, called a gray tone mask. Ideally, a gray tone mask must have a different UV light transparency in different areas so as to create exposure of different points of the resist at different UV doses.

This is generally obtained using three possible methods:

1. In pulse width modulation (PWM) the transparent area of the mask is created by a grid of squared holes, with different widths so as to modulate the transmitted UV dose.
2. In pulse density modulation (PDM) the transparent area of the mask is also created by a grid of squared holes, but it is their density that is varied in order to modulate the transmitted UV dose.
3. Finally, in pulse width and density modulation (PWDM) both the dimension and the density of the holes are changed to have a better control of the dose in each area of the resist.

An example of PWM and PDM gray tone masks is reported in Figure 4.45, while the relative dose of UV light on the resist versus the relative side of the holes and the number of holes in the length of the exposed area is reported in Figure 4.46.

Gray tone lithography allows to obtain a three-dimensional shape in a single lithography exposition, thus overcoming some of the limitations of multistep lithography. However, the lithography exposure system diffraction poses a limitation to the dimension and to the density of the holes on the mask and this limits reflects in a limit on the roughness of the inclined planes that can be obtained by gray tone lithography.

A result of a gray tone lithography process is shown in Figure 4.47, where a scanning electron microscope (SEM) image of a set of inclined planes built on a wafer surface by gray tone lithography is shown.

4.5.4 PHOTORESIST EXPOSURE

After the fabrication of a suitable mask, the fabrication pattern has to be used to expose selected areas of the resist to the incoming light. Whatever exposure system is used, it is intuitive that the final limit to lithography resolution is the wavelength of the used light. Even if phase masks allow a great reduction of negative effects due to diffraction, in any case diffraction cannot be completely eliminated and it poses the ultimate limit to the critical dimension of photolithography systems.

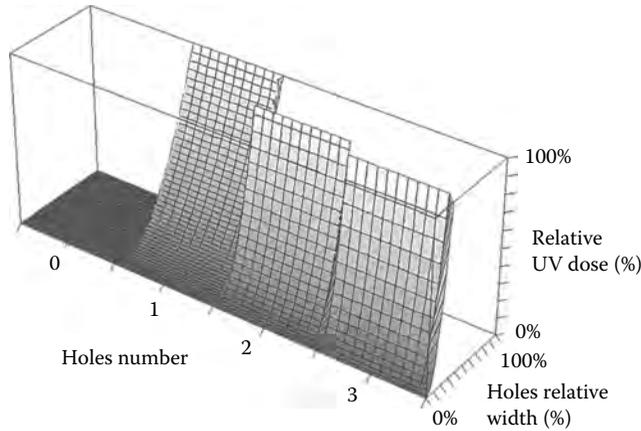


FIGURE 4.46 Relative dose of UV light on the resist versus the relative side of the holes and the number of holes in the length of the exposed area.

If extreme resolution is not required, a UV lamp can be used to generate the lithography light. High-pressure Hg vapor lamps are typically used in these cases. A photograph of one of these lamps with the corresponding UV spectrum is reported in Figure 4.48. Here, the Hg lines, generally, used for lithography at 365, 404.7, and 407.7 nm and the more powerful line at 436 nm are evidenced. Lamps add to the diffraction limit the fact that the large spectrum linewidth also contributes to degrade the contrast on the resist, also rendering practically impossible to control it by controlling the phase of the scattered light. Thus, if sub-micron resolution is to be achieved, lamps have to be substituted by UV lasers. For structures with a critical dimension of the order of 120 nm an argon or krypton fluoride laser is used, for smaller structures a nitrogen laser can be used exploiting the emission line at 157 nm.

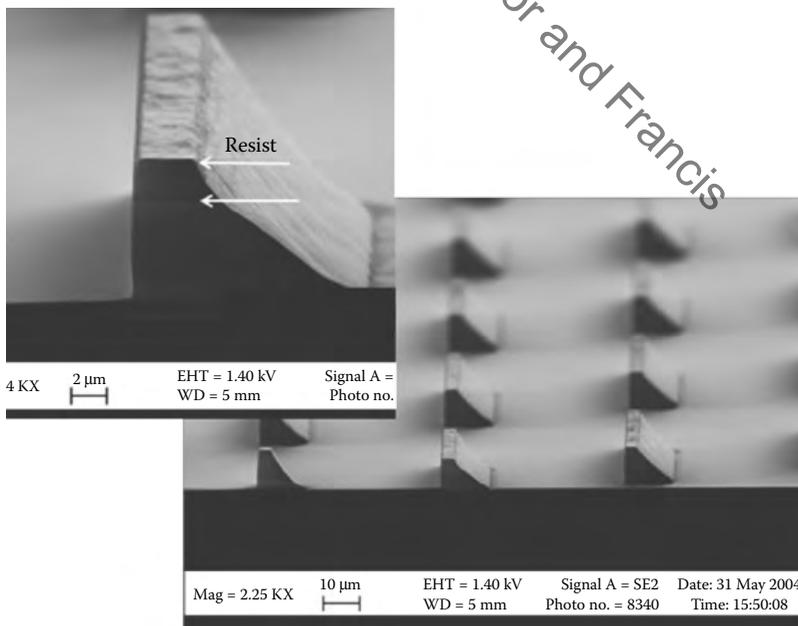


FIGURE 4.47 SEM image of a set of inclined mirrors built on a wafer surface by gray tone lithography. (Published under permission of Cyoptics Inc.)

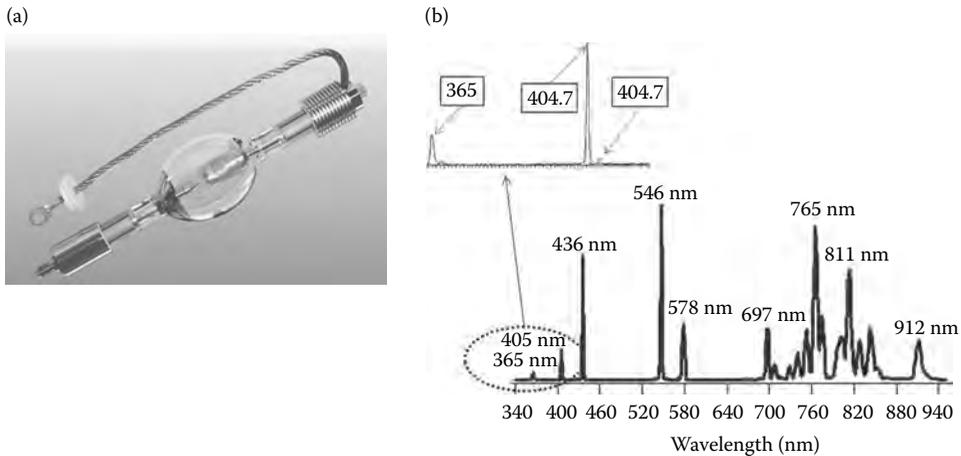


FIGURE 4.48 Photograph of an high pressure Hg lamp used for photolithography (a) and Hg infrared spectrum where the lines used for photolithography are evidenced (b).

Argon or krypton fluoride laser are examples of excimer lasers. Such lasers use a combination of a noble gas (argon, krypton, or xenon for example) and a reactive gas (generally, fluorine or chlorine). Under the appropriate conditions of electrical stimulation and high pressure, a pseudo-molecule called an excimer is created, which can only exist in an energized state.

Noble gases, such as xenon and krypton, are highly inert at normal conditions (temperature 25°C, pressure 100 kPa), but when in the excited state induced by an electrical discharge or an high-energy pulsed electron beams, they can form temporarily diatomic molecules or two atoms compounds with halogens, such as fluorine and chlorine. The excited compound has a strongly unstable ground state and in a characteristic time of a few picoseconds dissociates back in two unbound atoms. The excess energy associated to such dissociation is released via photons both by spontaneous and stimulated emission. Thus, if population inversion is created, a suitable energy pump laser action can be obtained [69–71]. The center of emission spectrum of the most used excimer lasers is reported in Table 4.12.

A nitrogen laser is a gas laser operating in the ultraviolet range (typically 337.1 nm) using molecular nitrogen as its gain medium, pumped by an electrical discharge [72,73].

The research to decrease the wavelength of the UV light used in lithography is continuously evolving, being UV lithography one of the major enabling technologies for the reduction of the critical dimensions of integrated circuits. Information on one of the main subject in these directions, the

TABLE 4.12
Wavelength and Typical Emitted Power of Selected Excimer Lasers

Excimer	Wavelength (nm)	Relative Power (mW)
Ar ₂	126	
Kr ₂	146	
Xe ₂ (first line)	172	
Xe ₂ (second line)	175	
ArF	193	60
KrF	248	100
XeBr	282	
XeCl	308	50
XeF	351	45
KrCl	222	25

use of the so-called extreme UV, the UV light with wavelength of the order of 15 nm, can be found in References [74–76].

Whatever light source is used, a way to project the image printed on the mask onto the resist have to be designed. Three strategies are used to do that, whose principle schemes are reported in Figure 4.49: contact exposure, proximity exposure, and projection exposure.

Contact exposure has been the first solution to be adopted: in this case, the mask is put at direct contact with the resist surface with a slow process of simultaneous approach and alignment. The best property of this exposure technique is that diffraction coming from the mask edges does not travel a long path and forms small fringes on the resist. However, the contact between the mask and the resist provoke both mask and resist contamination. Resist particles remaining on the mask have to be cleaned before reuse of the mask with a different wafer. This cleaning procedure shortens the mask life and, in case of imperfect cleaning of the mask, can cause contamination of the next wafer. On the resist side, the mask contact can create unwanted shapes on the resist, thereby causing potential defects in resist exposure.

These drawbacks do not exist in proximity lithography, where the mask is maintained in close proximity of the wafer by a set of proximity halters at least $10\ \mu\text{m}$ high. This avoids mask contact with the wafer and increases dramatically the yield of the process without introducing more complexity. However, due to the fact that the distance between the mask and the wafer is much larger of the UV light wavelength (generally from 30 to 100 times), the contrast of the mask image on the wafer is heavily degraded so as to cause an overall resolution worsening of one order of magnitude or greater. Several mask pre-processes have been introduced to compensate for this problem by trying to pre-compensate the diffraction effect, but in any case resolutions of the order of that reached with contact exposure are almost impossible to reach.

By far the most common method of exposure is projection printing. Projection lithography derives its name from the fact that an image of the mask is projected onto the wafer by an UV optical system. Projection lithography is enabled by very high-quality UV lenses and mirrors, designed using complex computer-aided design tools, so that lenses aberrations play only a marginal role in determining the quality of the image. The optical systems used in projection lithography are thus diffraction-limited projectors, since it is diffraction and not lenses aberrations that determine the final definition limit.

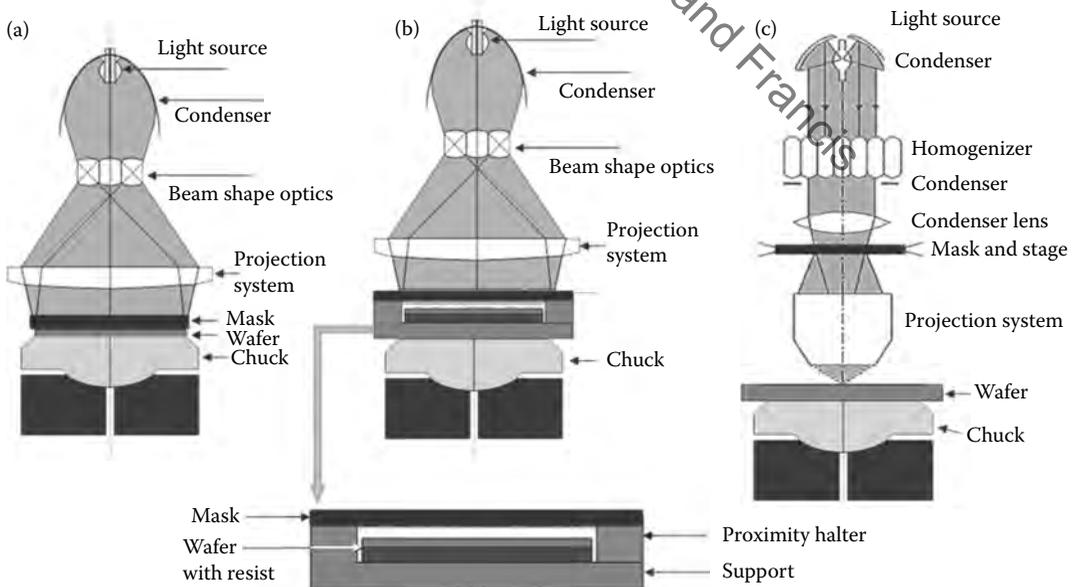


FIGURE 4.49 Schematics of the main photolithography exposure equipment: (a) contact exposure, (b) proximity exposure, and (c) projection exposure.

There are two major classes of projection lithography tools—scanning and step-and-repeat systems. Scanning projection printing [1,3], employs mirrors, to project a slit of light from the mask onto the wafer as the mask and wafer are moved simultaneously by the slit. Exposure dose is determined by the intensity of the light, the slit width, and the speed at which the wafer is scanned. These scanning systems, which generally use mercury lamps, reproduce the mask image on the wafer in 1:1 ratio so that the structures on the mask have the same dimensions of the structures that are to be reproduced on the wafer.

Step-and-repeat exposure systems, generally called simply steppers, expose the wafer one rectangular section at a time (the exposed section is frequently called the image field). Stepper can project the mask image on the wafer with an $N:1$ dimension ratio, where generally N varies from 1 in direct reproduction systems, up to 10 in very high definition systems.

The most advanced projection lithography equipment is a combination of the step-and-scan approaches. It uses a fraction of a normal stepper field, for example, a rectangular field of $25 \times 8 \text{ mm}^2$, then scans this field in one direction to expose the entire $N \times 1$ mask, where N is rarely greater than 4. The wafer is then stepped to a new location and the scan is repeated. The smaller imaging field simplifies the design and manufacture of the lens, but at the expense of a more complicated reticle and wafer stage. Step-and-scan technology is the technology of choice today for below 250 nm manufacturing.

An important aspect tending to limit the resolution of photoresist exposure when laser UV light is used in conjunction with high-resolution steppers is the standing wave effect. Monochromatic light, when projected onto a wafer, strikes the photoresist approximating as a set of plane waves arriving from slightly different angles. These waves travel through the photoresist and, if the substrate is reflective, are reflected back up through the resist. The incoming and reflected light interferes to form a standing wave pattern of high and low light intensity at different depths in the photoresist. This pattern is replicated in the photoresist, causing ridges in the sidewalls of the resist feature. As pattern dimensions become smaller, these ridges can significantly affect the quality of the feature. The interference that causes standing waves also results in a phenomenon called swing curves, the sinusoidal variation in the resolution of a single width of a line on the resist pattern with changing resist thickness. Swing curves affect all types of projection lithography with monochromatic light, but can be destructive when using 3D lithography.

These detrimental effects can be greatly reduced by coating the substrate with a thin absorbing layer called a bottom antireflective coating (BARC) that can reduce the reflectivity seen by the photoresist to $<1\%$.

4.5.5 POST-EXPOSURE PROCESSES

4.5.5.1 Post-Exposure Bake

One method of reducing the standing wave effect is called post-exposure bake (PEB) [1]. Although there is still some debate regarding the PEB working mechanism, it is believed that the high temperatures used (100°C – 130°C) during PEB cause diffusion of the photoactive compound, thus smoothing out the standing wave ridges. For a conventional resist, the main importance of the PEB is diffusion to remove standing waves. For another class of photoresists, called chemically amplified resists, the PEB is an essential part of the chemical reactions that create a solubility difference between exposed and unexposed parts of the resist. For these resists, exposure generates a small amount of a strong acid that does not itself change the solubility of the resist. During the post-exposure bake, this photogenerated acid catalyzes a reaction that changes the solubility of the polymer in the resist. Control of the PEB is extremely critical for chemically amplified resists.

4.5.5.2 Development

Once exposed, the photoresist must be developed [1,77,78]. Most commonly used photoresists use aqueous bases as developers. In particular, frequently diluted solutions of tetra-methyl ammonium

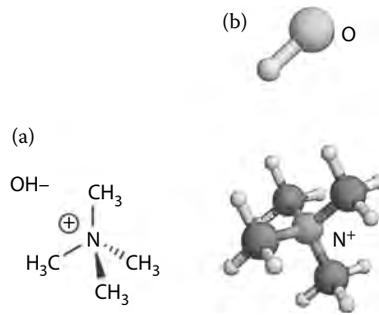


FIGURE 4.50 Tetra-methyl ammonium hydroxide structure formula (a) and stereographic representation in the balls and sticks convention (b).

hydroxide (TMAH—see Figure 4.50) are used. Development is a critical step in the photoresist process. The characteristics of the resist–developer interactions determine to a large extent the shape of the photoresist profile.

The method of applying developer to the photoresist is important in controlling the development uniformity and process latitude. In fabs manufacturing chips with critical dimensions in the micron range, batch development is the predominant development technique. A boat of 10–30 wafers is developed simultaneously in a large beaker, usually with some form of agitation. If the critical dimensions are lower, like in sub-micron fabs, and the throughput has to be maintained high without decreasing the yield, other kinds of development processes have to be used, such as spin development and spray development.

During spin development wafers are spun, using equipment similar to that used for spin coating, and developer is poured onto the rotating wafer. The wafer is also rinsed and dried while still spinning. Spray development has been shown to have good results using developers specifically formulated for this dispense method. Using a process identical to spin development, the developer is sprayed, rather than poured, on the wafer by using a nozzle that produces a fine mist of developer over the wafer. This technique reduces developer usage and gives more uniform developer coverage.

Another development strategy is called puddle development. Using developers specifically formulated for this process, the developer is poured onto a stationary wafer that is allowed to sit motionless for the duration of the development time. The wafer is then spin rinsed and dried.

Spin, spray, and puddle development processes can be generally performed by the same piece of equipment, with minor modifications, that allows to adapt the development process to the used resist and to the rest of the lithography process without using different equipment for every different development technique.

4.5.5.3 Postbake

Postbaking is used to harden the final resist image after development so that it will withstand the harsh environments of implantation or etching. The high temperatures used (120°C–150°C) cross-link the polymer in the photoresist, thus making the image more thermally stable.

However, if the postbaking temperature is too high, the resist flows causing degradation of the image. The temperature at which flow begins is related to the glass transition temperature of the polymer and it is a measure of the thermal stability of the resist. In addition to cross-linking, the postbake can remove residual solvent, water, and gases and will usually improve adhesion of the resist to the substrate.

Other methods have been proposed to harden a photoresist image. Exposure to high intensity deep-UV light crosslinks the resin at the surface of the resist forming a tough skin around the pattern [3,79]. Deep-UV hardened photoresist can withstand temperatures in excess of 200°C without dimensional deformation.

4.5.5.4 Pattern Transfer

After the patterns have been lithographically printed in photoresist, they must be transferred into the substrate.

There are three basic pattern transfer approaches:

- Subtractive transfer (etching)
- Additive transfer (selective deposition)
- Impurity doping (ion implantation)

Etching is the most common pattern transfer approach. A uniform layer of the material to be patterned is deposited on the substrate. Lithography is then performed such that the areas to be etched are left unprotected by the photoresist. Etching is performed either using wet chemicals such as acids, or more commonly in a dry plasma environment. The photoresist “resists” the etching and protects the covered material. When the etching is complete, the resist is stripped leaving the desired pattern etched into the deposited layer.

When selective deposition is adopted the lithographic pattern is used to open areas where the new layer is to be grown. Stripping of the resist then leaves the new material in a negative version of the patterned photoresist.

Doping involves the addition of controlled amounts of contaminants that change the properties of the host layer material. Ion implantation uses a beam of dopant ions accelerated at the photoresist-patterned substrate. The resist blocks the ions, but the areas uncovered by resists are embedded with ions, creating the selectively doped regions.

We will analyze in more detail these techniques, fundamental in the construction of labs on chip, in Sections 4.8 through 4.10.

4.5.5.5 Photoresist Stripping

After the imaged wafer has been processed the remaining photoresist must be removed. There are two classes of resist-stripping techniques: wet stripping using organic or inorganic solutions, and plasma dry stripping.

Most organic strippers are phenol based and are designed to avoid scum formation on the wafer after stripping. However, the most common wet strippers for positive photoresists are inorganic acid-based systems used at elevated temperatures [1,3]. Wet stripping has several inherent problems. Although the proper choice of strippers for various applications can usually eliminate gross scumming, it is almost impossible to remove the final monolayer of photoresist from the wafer by wet chemical means. It is often necessary to follow a wet strip by a plasma descum to completely clean the wafer of resist residues. Photoresist which has to undergone extensive hardening like deep-UV hardening to be suitable to resist at harsh processing conditions, for example, high-energy ion implantation, can be almost impossible to strip chemically. For these reasons, plasma stripping has become the standard in most semiconductor processing.

4.5.6 PHOTOLITHOGRAPHY DEFINITION

The definition of a whole photolithography process depends on many factors, a few of them related to specific details of the process.

As far as the resist is concerned, the resist quantum efficiency is an important factor entering in the determination of the final resolution photolithography resolution. The resist quantum efficiency Φ is defined as the ratio between the photons absorbed by a unit resist area and the number of photon-stimulated processes that are generated, like break of polymer chains in positive resists. By its definition, the resist quantum efficiency enters in a fundamental way in the expression of the percentage of the resist exposure caused by the exposure process. The percentage of resist exposure

ϵ , that is the number of resist molecules undergoing photo-induced transformation N_{PT} divided by the overall number of molecules in a unit resist volume, is given by

$$\epsilon = \Phi DP(N_A)m_M \frac{A}{\xi} \quad (4.6a)$$

where

Φ is the resist quantum efficiency.

D is the dose (power per unit time) emitted by the light source.

$P(N_A)$ is the percentage of the dose captured by the focusing system and directed over the wafer.

$P(N_A)$ is an increasing function of the numeric aperture N_A , representing the product between the index to the lens material and the sin of half the acceptance angle [80,81]. The numerical aperture is linked to number of diffraction modes present at the output of each mask transmitting area that are captured by the optical system and directed to the wafer. The exact form of $P(N_A)$ depends on the particular type of optical system, a possible shape is reported in Figure 4.51.

m_M is the molecular weight of the resist polymer.

A is the Avogadro number.

ξ is an empirical configuration factor relating molar weight and molar volume. Since ξ depends on the molecular shape it is of the same order of magnitude for all the polymers, even if its specific value changes from molecule to molecule.

Increasing the mass of the polymer chain, the percentage of the exposed resist increases, due essentially to the fact that less photo-induced phenomena are needed. The same happens when increasing the resist quantum efficiency, due to the better exploitation of the incoming photons, and the optical system numerical aperture, due to the best exploitation of the power emitted from the light source. Ideally, the exposure efficiency have to be as near to 100% as possible, to avoid a nonuniform development of the resist, mainly in correspondence of the edges of the exposed area, causing contrast degradation near the edge of the transferred pattern.

The mask definition is another important element for the photolithography final definition: it is defined as the overall aberration of the optical system collecting the light from the source and focusing it on the resist surface.

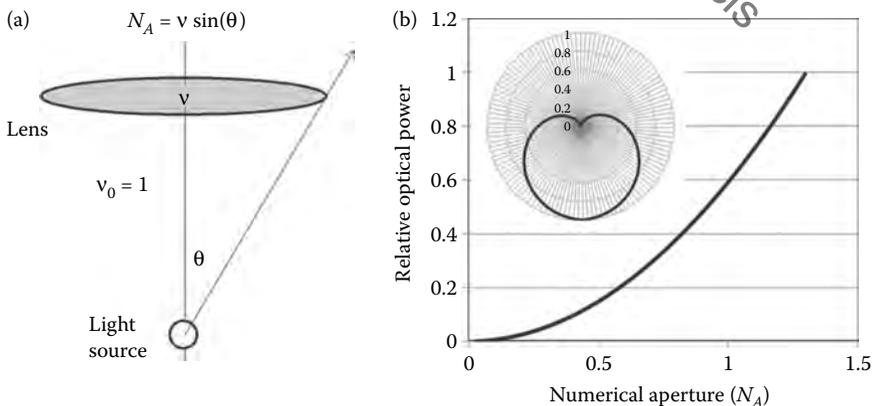


FIGURE 4.51 Definition of numerical aperture of a lens (a) and example of the dependence of the fraction of UV power emerging from a mask clear area that is focused on the resist surface from the optical system numeric aperture (b). The insert represents the wave front exiting from the mask clear area.

Resist quantum efficiency, mask definition, and other similar factors must be carefully engineered to obtain a high-performing photolithography system. However, high-quality photolithography machines are so carefully designed, that all these factors are near their optimum value, so that the main element determining the lithography resolution is the diffraction characteristics of the optical system. Let us consider as a reference a stepper projection system, where requirements for definitions are more stringent. We can represent the definition as the minimum width w of a line that can be resolved on the photoresist (also called the system linewidth) and write [1,81]

$$w = k \frac{\lambda}{N_A} \quad (4.7a)$$

where system constants and the effects of phenomena different from diffraction depending on the resist, lenses aberration, development technique, and so on, are collected in a process constant k . Typical values of k are of the order of 0.6–0.8 for excimer laser lithography systems; if a bi-dimensional phase mask is used, k is greatly reduced, arriving up to an order of 0.2.

The presence of the light wavelength λ at the numerator of Equation 4.7a is obvious, due to the fact that the diffraction fringes have a depth proportional to the wavelength, thus smaller the wavelength greater the contrast on the resist. As far as the numeric aperture is concerned, its influence can be understood considering the fact that, higher the percentage of the emitted light captured by the optical system, greater the number of diffraction orders that are collected to form the final image. This effect is evidenced in Figure 4.52.

Substituting realistic numbers, with $k = 0.55$, $\lambda = 193$ nm, $n = 1.5$, and $\theta = 110^\circ$ so that $N_A = 0.86$ we achieve 124 nm resolution and with a bi-dimensional phase mask as low as about 40 nm.

From Equation 4.7a it could seem that greater the numerical aperture better the performances of the lithography. This is true in terms of resolution, but resolution is not the only performance parameter that characterizes a good lithographic process. The resist has a finite thickness, and if exposure has to be uniform all along the thickness of the photoresist, the incident light beam should be focused over all its depth. Even if theoretically there is a well-defined focus distance between a mask transparent area and the corresponding resist area, the beam defocusing is very small in a certain interval around the focus distance, called depth of focus (DOF), while out of this interval it increases abruptly completely defocusing the beam [80]. Thus, the optical system DOF has to be at least equal to half the resist film thickness to generate uniform resist exposure. The DOF also depends on the numerical aperture and on the wavelength following the equation [80,81]:

$$DOF = 2k' \frac{\lambda}{N_A^2} \quad (4.8)$$

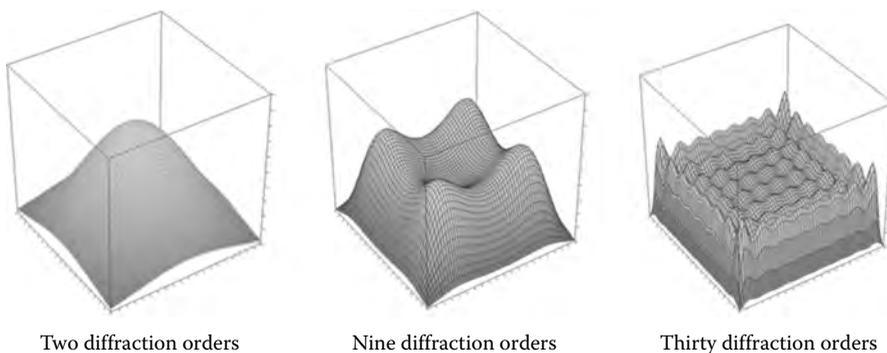


FIGURE 4.52 Diffraction image from a square aperture with a side equal to twice the incident plane wave wavelength versus the number of collected diffraction orders.

where k' is another process-related constant whose value is around 0.5 for the most performing excimer laser steppers. From Equation 4.8, it results that all the steps increasing the photolithography definition operating on the wavelength or on the numerical aperture also reduces the DOF, so that a trade-off have to be found. The DOF with the data of the previous example results to be about 260 nm, that means a resist thickness before backing of about 1 μm , at the limit of what can be done with a standard PMMA resist.

4.6 ELECTRON BEAM LITHOGRAPHY

Electron beam (e-beam) lithography is a very high-resolution lithography technique that uses electron beams instead of UV light beams to impress the photoresist. The typical energy of electron beams used in e-beam lithography is 10–50 keV, so that the corresponding De Broglie wavelength is of the order of 50 \AA , so small that any diffraction effect disappears. The e-beam resolution is limited by factors different from diffraction, like electrons scattering in the resist and electrons optics aberration. In any case, resolutions smaller than 10 nm can be attained by electron beam techniques. An example of the structure fabricated using electron beam lithography is shown in Figure 4.53. It is a ring optical resonator fabricated directly in silicon. Light coupling between the resonator and the in/out waveguide is regulated by the distance between these structures, that in the case of the figure is as low as 120 nm and that has to be carefully controlled to match the resonator specifications. The resonator was fabricated using e-beam lithography and the estimated resolution was 7 nm, sufficient to achieve a good control of the coupling distance [46,82,83].

A schematic representation of the architecture of a direct writing e-beam lithography equipment is represented in Figure 4.54: a narrow beam of high-energy electrons is directly focused on the resist deposited on the wafer with a focus and scanning system similar to that of a scanning electron microscope. Two possible scanning techniques can be adopted: sequential scanning and vector scanning. In sequential scanning machines the electron beams scan the wafer surface sequentially following a grid trajectory that is the same for each wafer. When the resist is to be exposed the beam is empowered, while when it does not have to be exposed it is shut off almost completely. If vector scanning is used, the electron beam is driven by the software mask charged into the e-beam control system to scan and impress only the areas on the wafer that has to be exposed. Generally, control software for vector scanning is more complex, but allows a certain reduction of the scanning speed.

E-beam lithography has several advantages with respect to UV lithography, besides its definition: no mask is required (in some sense a software mask is charged into the e-beam control), the exposure of the resist is accurately controlled over very small areas, and focus can be adjusted during the electron beam scan so as to obtain large DOF. In particular, the combination of the ability to

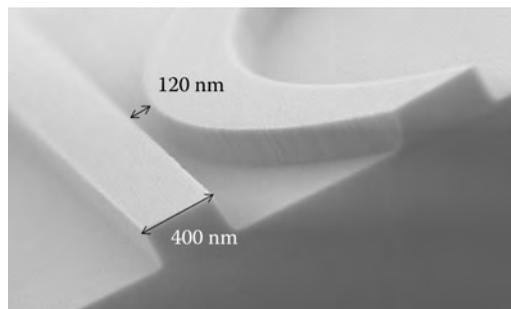


FIGURE 4.53 Scanning electron microscope image of a particular of an optical ring interferometer fabricated using e-beam lithography; the critical dimension is the minimum distance between the straight waveguide and the ring resonator, that is as small as 120 nm. The estimated e-beam lithography resolution was 7 nm. (Published under permission of Cyoptics Inc.)

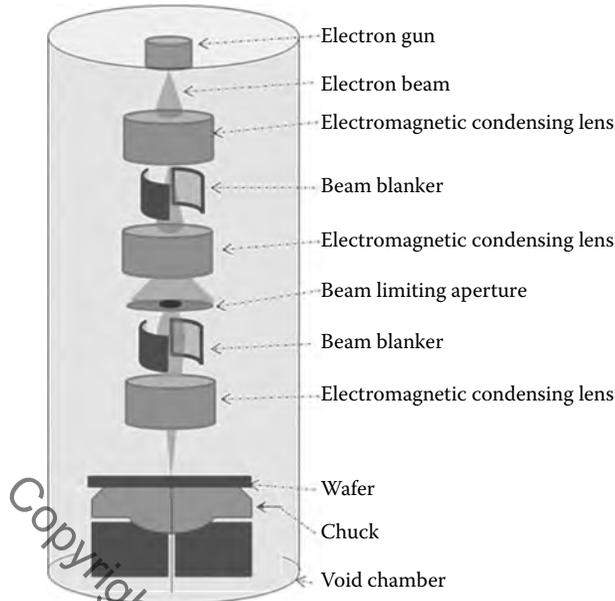


FIGURE 4.54 Schematic representation of the e-beam equipment architecture.

accurately change the dose of electrons of the beam and the possibility to obtain a very large DOF renders e-beam lithography suitable for three-dimensional lithography of very small structures. An example of such structure is reported in Figure 4.55, where an SEM photograph of two superimposed optical waveguide fabricated with high-contrast index silicon core is shown [84].

E-beam lithography has also disadvantages with respect to photolithography. The resolution advantage, that was enormous when this technique was invented in the 1960s, is now quite reduced and, even the most complex photolithography machines are simpler in terms of structure with respect to an e-beam. This is due both to the fact that electrons require to be processed in vacuum and to the more complex and costly structure of the electron optics and the scanning system.

The fundamental limit of the e-beam lithography is, however, the low throughput with respect to photolithography. Exposure through the electron beam is slow and the processing of a 120 mm wafer can require sensibly more than an hour [46]. The throughput limit has in practice confined e-beam lithography to research and to niche applications where extreme resolution is needed for circuits that have not to be reproduced in huge numbers.

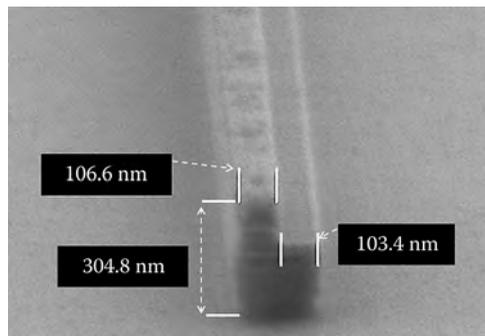


FIGURE 4.55 SEM photograph of two superimposed optical waveguide fabricated with high contrast index silicon core. (Published under permission of Cyoptics Inc.)

Another common application of the electron beam lithography is the preparation of phase mask for steppers. On the one hand, a mask can be used to process a huge number of wafers, thus the fact that mask fabrication process is slow is not important even in very high-throughput fabs. On the other hand, if a resolution of the order of 40 nm or less is to be obtained with optical lithography, the mask accuracy is in any case critical. This is especially true when two-dimensional phase masks or when gray lithography masks have to be fabricated: in the first case, the control of the mask transmission phase requires very high definition and extremely accurate dose control and in the second case the critical size of the mask is generally quite smaller than the critical dimension of the final circuit.

An attempt to overcome the throughput limitation of e-beam lithography is the use of much larger electron beams [85] that illuminate large areas of the wafer. Naturally, in this case, a mask is required to reproduce the pattern on the wafer, similar to the case of photolithography. Even if the limitations implicit in proximity printing, that is generally the technique used to set the mask in this kind of e-beam machines, can be overcome in great part with optical corrections [86] (see Figure 4.41), this technique at present is not widely used.

4.7 ETCHING

Besides lithography, the most important planar processes are subtractive and additive processes. Subtractive technologies, also called etching, allows selected patterns to be removed from a superficial film on a wafer, while additive technologies allow similar patterns to be added in form of thin superficial film.

Under the general name of etching are collected a set of very different techniques that can be divided into two great families: wet and dry etching. Processes relying on the consumption of the film to be etched through chemical reaction with a suitable etchant are called wet etching processes, due to the fact that they are generally performed by immersion of wafers in an etchant solution. Dry etching on the contrary is the family of all the etching technique relying on mechanical elimination of a selected pattern from the wafer surface, generally through particle bombardment in case sustained by chemical action of the particles with the film to be removed.

Even if quite a variety of different techniques can be collected under the two etching families, they have common traits that are in part collected in Table 4.13. From the table it is evident that wet and dry etching are complementary techniques, not competitive ones. They are both used widely for different applications and frequently coexist in the same processes stream.

TABLE 4.13
Qualitative Comparison between Wet and Dry Etching Techniques

	Dry	Wet
Environment	Vacuum	Atmosphere
Equipment	Vacuum chamber Particles gun Particles beam optics	Chemical bath
Advantages	Small feature sizes (<100 nm)	Low cost Good selectivity High etching rate
Implementation	More difficult	Easier
Disadvantages	High cost equipment Low throughput Smaller selectivity (for selected techniques)	No features under 1 μm Wafer contamination possible
Directionality	Anisotropic	Isotropic for amorphous materials (like SiO_2) Anisotropic etching possible for crystals (like Si)

4.7.1 WET ETCHING TECHNIQUES

Wet etching techniques are used essentially in two different applications: to transfer a pattern on the wafer surface through a suitably exposed and developed resist or to create three-dimensional structures. In the first case, wet etching is often used in conjunction with dry etching, for example, to smooth too sharp corners where stress tends to accumulate creating potential failure due to material breaking.

Another possible application of wet etching in conjunction of dry etching is damage removal from the wafer surface by etching up a very thin film.

In application with critical dimensions greater than 1 μm , wet etching can be used also for pattern transfer. An example is given by the excavation of trenches between different devices or group of devices of the same circuit to create better electrical or thermal insulation. Also, lithographic structure like V grooves (see Section 6.5.4) and microfluidics ducts can be obtained by wet etching.

4.7.1.1 Wet Etching Characteristics

The most used wet etching process uses etch insensitive masks, that is masks that are not chemically attacked by the etchant. In this way, the etching process happens under the mask, but it does not affect it. Etching mask of this type can be constituted either by photoresist after exposure, development and strong hardening (frequently using UV), or by a previously selectively etched film, like SiO_2 over silicon.

A first important characteristic of wet etching is the isotropy degree. Let us define a reference system where the x - y plane is the wafer surface under the resist and the z -axis is perpendicular to the wafer. Considered a direction θ , the etching rate along this direction, called r_θ , can be defined as

$$r_\theta = \frac{dH_\theta(t)}{dt} \quad (4.9)$$

where $H_\theta(t)$ is the etch depth reached along θ in the interval $[0, t]$. If wet etching is applied to a uniform amorphous film, the etching rates r_x and r_y in all the horizontal directions, that are the directions parallel to the wafer surface, is necessarily the same, while the etching rate r_z in the vertical direction is potentially different. In this case, the isotropy rate is defined as

$$\rho = \frac{r_x}{r_z} = \frac{r_y}{r_z} \quad (4.10)$$

Due to its definition $0 \leq \rho \leq 1$, where in the case of perfect isotropic etching $\rho = 1$ and in case of perfectly anisotropic etching $\rho = 0$. If wet etching is performed using normal etchants on amorphous films the etching isotropy rate is always quite near to one, since the etch tends to be isotropic.

The effect of different values of the etching isotropy rate is shown in Figure 4.56. In the figure, the fact that wet etching generally creates the so-called under-cut above the mask is also evidenced. The under-cut dimensions u is given by

$$u = \rho H_x(t) = \frac{1}{2} \frac{dH_x^2}{dH_z} \quad (4.11)$$

so as to be maximum in case of isotropic etching and minimum, equal to zero, in case of perfectly anisotropic etching. In case of normal etchants for amorphous films the etching isotropy rate is always quite near to one, since the etch tends to be isotropic. Small degrees of anisotropy could be created either by nonhomogeneity of the etched film (e.g., if the density increases with depth), or by the presence of hydraulic pressure in the etchant bath.

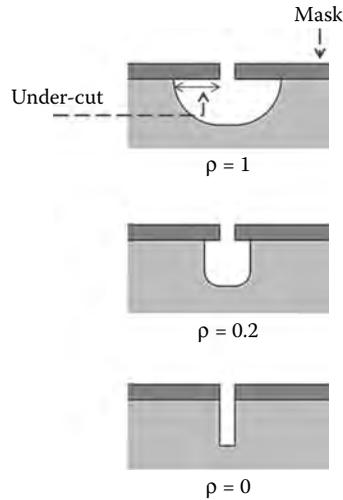


FIGURE 4.56 Effect of different values of the etching isotropy on the etched shape in an amorphous film. The under-cut created by the wet etching is also defined. The etching mask can be either photoresist after exposure, development and strong hardening (frequently using UV) or a previously selectively etched film, like SiO_2 over silicon.

An interesting example of anisotropic wet etching of a thick SiO_2 substrate is shown in Figure 4.57, where an isotropy rate of about 0.5 in Figure 4.57a and of 1.25 in Figure 4.57b (etch rate in horizontal direction greater than in the vertical direction) are achieved by ion implantation before etching. Ion implantation defines a material volume that can be etched at a higher rate-generating anisotropy. The scope in the specific case was to achieve a flat inclined surface to generate a mirror after suitable metallization.

Anisotropic etching can be obtained in a well-controlled way in the case of crystal films, since the effect of the etchant in this case depends on the crystal orientation so that anisotropy spontaneously rises. Since wet etching is generally highly selective, the etcher used for the surface film does not etch the substrate. Thus, when the etching depth reaches the substrate, etching in vertical dimension stops, while etching in horizontal dimension goes on, creating the so-called over-etching of the film (cf. Figure 4.58).

Over-etching allows a better definition of the pattern to be transferred, since the lateral wall gets more and more vertical, but also causes an increase in the dimension of the under-cut, that can cause problems in the phase of resist stripping (e.g., the break of the resist film). This is the reason, in particular, if thick films are considered, over-etching cannot be pushed for a too long time.

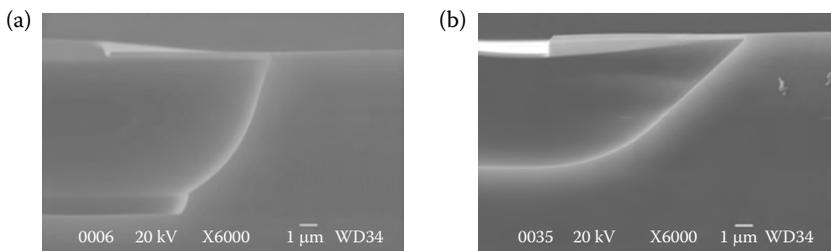


FIGURE 4.57 Example of anisotropic wet etching of a thick SiO_2 substrate, where an isotropy rate of about 0.5 in (a) and of 1.25 in (b) are achieved by ion implantation before etching. (Published under permission of Cyoptics Inc.)

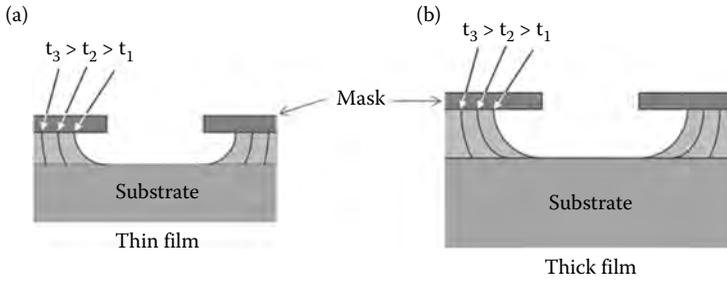


FIGURE 4.58 Over-etching effect of a prolonged etching time in case of a thin (a) or thick (b) etched film.

If etching sensitive masks are used instead of etching insensitive ones, the mask is also etched contemporarily to the film below it. There are three key parameters for this kind of etching process that can be defined starting from the mask and film etching rates. Assuming the mask and the film to be amorphous materials, these parameters are called etching selectivity S_m , horizontal etching relative rate R_R and mask relative etching rate ρ_m , and they are defined as

$$S_m = \frac{r_{m,z}}{r_{f,z}} \tag{4.12}$$

$$R_R = \frac{r_{m,x}}{r_{f,x}} \tag{4.13}$$

$$\rho_m = \frac{r_{m,x}}{r_{m,z}} \tag{4.14}$$

where the subscript m identifies mask etching rates and the subscript f film etching rates.

During the etching process, the etching mask is profiled as in Figure 4.59a, so as to present a lateral flat shape inclined of an angle (say θ) with respect to the horizontal plane. Two different etching regimes can be realized, generating different etching shapes: mask-dominated etching (where $R_{mR} > 1$) and film-dominated etching (where $R_{mR} < 1$). The relative etching profiles are reported schematically in Figure 4.59b and c.

In mask-dominated etching regime no under-cut is generated, while under-cut is still present in film-dominated etching regime. The so-called etching bias, defined in Figure 4.59, can be evaluated in both cases starting from the etch process parameters obtaining

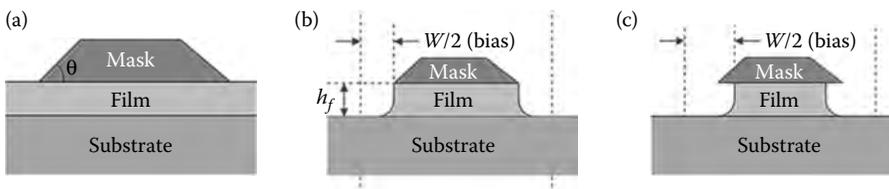


FIGURE 4.59 Etching of mask and film: wafer and mask before the etching process (a), etching result of a mask driven etching (b), and etching result of a film driven etching (c). The etching mask can be either photoresist after exposure, development and strong hardening (frequently using UV) or a previously selectively etched film, like SiO_2 over silicon.

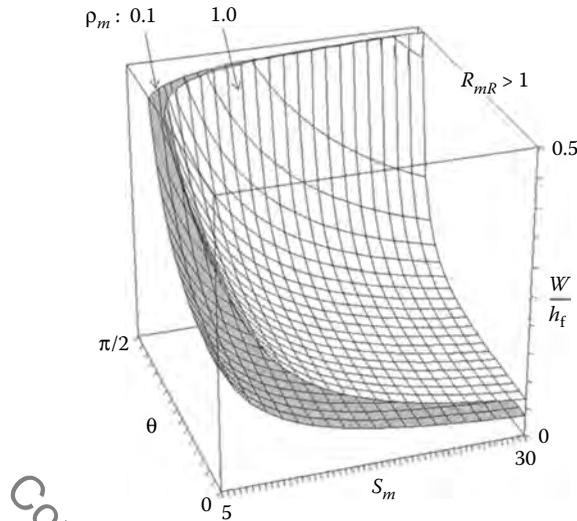


FIGURE 4.60 Bias to film thickness ratio versus etching selectivity and mask edge angle for different etching isotropy ratio on the mask in the mask dominated etching regime.

$$W = \frac{2W}{S_m} \left(\frac{1}{\cos\theta} + \rho_m \right) \quad R_{mR} > 1 \quad (4.15a)$$

$$W = 2h_f \rho_m \quad R_{mR} < 1 \quad (4.15b)$$

In Figure 4.60, the ratio W/h_f is shown versus θ and S_m in a practical range of parameters in the case $R_{mR} > 1$ and isotropic mask etch ($\rho_m = 1$).

4.7.1.2 Wet Etching Equipments

The wet etching equipment is constituted by a chemical bath embedded in a machinery allowing both operators protection from chemical hazards and automatic wafer loading besides regulation of the process parameters. In large equipment, batch processing is possible by loading a set of wafers and leaving the machine to manage them one after the other.

The wet etching process is quite sensitive to a few process parameters that are specific of the individual equipment. The first of these parameters is the agitation of the bath. Agitation is generally performed to continuously renew the etching bath on the wafer surface, but the final etching result is quite dependent on the agitation parameters like frequency, continuous or discontinuous agitation, and so on.

In the case of crystal etching, mainly in the etching of silicon, part of this problem can be overcome by the use of suitable anisotropic etchant that by minimizing the lateral etching also minimizes the variability due to the effect of agitation. This is not possible when etching amorphous films such as glasses or polymers and a careful process calibration is needed for a good reproducibility.

The final etching result is also temperature sensitive and different equipment realizes the process at different temperatures.

Frequently, mainly when dealing with crystals, the wet etching process is supported by an electrical field that is generated in the etching bath. Such system is called electrochemical etching. Electrochemical etching requires positive charges (holes in semiconductors) in the wafer. Thus, while a p-doped silicon wafer can be used as it is, an n-doped wafer has to be inverted near the

surface to allow the etching process. This can be done, for example, by photo-generation of positive charges near the surface stimulated by wafer illumination with energetic photons.

Electrochemical etching is obtained by generating a static electrical field between an electrode on the substrate and another electrode placed into the etchant solution on the floor of the etching bath. Holes are attracted at the surface of the wafer by the field and promote easy oxidation of the surface wafer layers. Once created, a thin oxide layer, it is easily dissolved by the etching solution; after etching the first oxide layer, the silicon is uncovered and the process starts again with oxidation. The etching efficiency of electrochemical etching is quite higher with respect to standard etching processes and mild etching solutions can be used, so allowing the use of simple photoresist as etching masks, without resorting to strong hardening that renders the resist stripping quite difficult.

Several different etching solutions have been developed to etch films of different materials. A review of the characteristics of wet etching as a function of the material to be etched is reported in Table 4.14.

TABLE 4.14
Review of the Characteristics of Wet Etching as a Function of the Material to be Etched

Material	Application	Etchant	Etchant Concentrations	Etching Rate (Vertical)	Notes
SiO ₂	Microfluidics, integrated optics	HF	48%	20/2000 nm/min	Also Si substrate etched—0.3 Å/min for n-type with (111) wafer and E = 0.002 Ω m
SiO ₂	Microfluidics, integrated optics	HF:NH ₄ F (28 mL HF, 170 mL H ₂ O, 2113 g NH ₄ F)	28 mL–170 mL–2113 g NH ₄ F	100–500 nm/min	@25°C
Si (all planes)	General etching	HF–HNO ₃ –CH ₂ COOH	Volume 8%–75%–17%	5 μm/min	@25°C
Si (111)	Delineates defects	HF–(5M–CrO ₃)	1:1		Needs agitation, do not reveal etch pits on (100) well Requires agitation
Si (100)–(111)	Delineates defects	HF–HNO ₃ –(5M–CrO ₃)–Cu(NO ₃)–CH ₂ COOH–H ₂ O	60 mL–30 mL–2 g–60 mL–60 mL		
Si (100)	Heavy doped layers	HF–(1M–CrO ₃)–H ₂ O	44%–22%–33% Volume		
Titanium	Metal definition	HF–H ₂ O ₂		880 nm/min	
Tungsten		H ₂ O ₂		20/100 nm/min	
Chromium		HCl–glycerine	1 mL–1 mL	80 nm/min	
Gold	Define contacts	HCl–HNO ₃	75%–25% Volume	25–50 mm/min	Requires solution in aqua regia
Organic layers		H ₂ SO ₄ –H ₂ O ₂		>1 mm/min	
Organic layers		Acetone			>4 mm/min
Aluminum		H ₂ PO ₃ –HNO ₃ –HC ₂ H ₃ O ₂		660 nm/min	40–50°C

Source: Adapted from Madou M. J., *Fundamentals of Microfabrication and Nanotechnology*, Third Edition, Three-Volume Set, CRC Press; 3rd edition (August 1, 2011), ISBN-13: 978-0849331800; With kind permission from Springer Science+Business Media: *Technology of Integrated Circuits*, Reprint of 2000 edition, December 15, 2010, Widman D., Mader H., Friedrich F.

4.7.2 PLASMA CHARACTERISTICS AND PLASMA GENERATION FOR PLANAR PROCESSES

Wet etching has limits if used in transferring patterns from an etching mask. This happens especially when the film to be etched is a uniform, amorphous film like SiO_2 , situation that is often characteristics of labs on chip micro-fabrication. In this case, only small deviations from perfect etching isotropy can be realized in a controlled way so that sharp shapes cannot be realized and the vertical dimension of the transferred path is always about equal to the horizontal dimension. In order to obtain vertical walls, high-aspect ratio shapes, and in general strong etching anisotropy, dry etching has to be used [1,87].

Dry etching techniques can be roughly divided into two categories: plasma-assisted dry etching and dry etching without use of plasma. In the first case, the etching is either obtained by a simple mechanical operation of high-speed particles of plasma incident on the film surface or via a combination of chemical and mechanical action. In the second case, purely chemical etching is realized, by means of accurately directed etching particles, so as to obtain a selective etching action.

PE is by far the most used form of dry anisotropic etching, so that plasma generation and processing is a key technology in designing dry etching machines.

This is one of several application of plasma in planar fabrication: plasma is also used in deposition processes (see Section 4.8.2), in wafer bonding processes (see Section 4.9.3), and even in few surface activation techniques. For this reason, before describing plasma dry etching, we will revise the plasma generation and management techniques that are applied in the field of microfabrication.

4.7.2.1 Plasma Characteristics and Generation

Plasma is defined as an electrically neutral medium made by chemically unbound positive and negative particles, generally positive ions and free electrons, even if also a plasma of negative and positive ions is conceivable [88,89].

Although the plasma particles are unbound, they do not constitute a sort of perfect gas; that is, they are not mechanically free. On the contrary, in order to talk about a plasma state, three conditions must hold:

1. *The plasma approximation:* When charged particles move they generate electrical currents and magnetic fields. In a plasma the charged particles density must be high enough that each particle influences many nearby particles so as to generate collective effects. These collective phenomena due to many body interactions are a distinguishing feature of a plasma. The plasma approximation is valid when the number of charge carriers within the sphere of influence of one of them is higher than unit.

In a plasma, the sphere of electromagnetic influence of a single plasma particle is often called Debye sphere and its radius the Debye screening length L_D [90]. The average number of particles in the Debye sphere is given by the plasma parameter, generally indicated as Λ . This terminology is due to the affinity of classical plasma theory with the theory of electrolytes in a solution of neutral particles that we will review in some detail in Section 7.5. As a matter of fact, the definition of Debye length and Debye layer are completely analogous to the same definition in the electrochemical case and they can be found in Section 7.5.2.

2. *Bulk interactions:* In order to consider a set of charged particles a plasma, the Debye screening length has to be much shorter than the physical size of the plasma. This means that surface effects arising at the plasma boundaries through the creation of the Debye layer have to be negligible with respect to bulk effects and that the volume of the sphere of influence of a single charged particle, that is defined by the Debye layer arising around its surface, is negligible with respect to the whole plasma volume.
3. *Plasma frequency:* The electron plasma frequency, measuring plasma oscillations of the electrons [89] has to be large compared to the electron-neutral collision frequency,

measuring frequency of collisions between electrons and neutral particles. When this condition is valid, electrostatic interactions dominate over the processes related to ordinary gas kinetics.

The electron plasma frequency, when the thermal electron energy can be neglected with respect to their electrostatic energy, can be expressed as

$$f_{el} = \frac{1}{2\pi} \sqrt{\frac{\rho_e q^2}{m_e \epsilon_0}} \quad (4.16)$$

where q and m_e are the electron charge and mass, respectively, and ρ_e is the electron density, that is generally called plasma density. Since the plasma frequency depends only on the plasma density and on physical constants the plasma frequency conditions can be simply reduced to a condition on the density of particles in the plasma. This is quite intuitive, since qualitatively it requires the plasma to be as far as possible from the condition of a perfect gas.

For plasma to exist, ionization is necessary to produce free charges. To characterize the effectiveness of the ionization process, besides the plasma density, the plasma ionization parameter α_p is introduced.

The plasma ionization is defined starting from the densities n_i of ions and n_a of the neutral atoms as

$$\alpha_p = \frac{n_i}{n_i + n_a} = \frac{n_i}{n} \quad (4.17)$$

where $n = n_i + n_a$ is the total number of particles containing a nucleus in the plasma. The plasma density and ionization are related by the ionization charge Z of the ions, indicating the number of electrons that each atom lost during plasma formation by ionization. In particular, it results

$$\rho_e = Zn\alpha_p \quad (4.18)$$

Another typical plasma parameter is the so-called plasma temperature, which is defined as the average energy of a free plasma electron. Generally, unless extremely violent phenomena are generated, that is not the case of artificially generated plasmas, the electron portion of the plasma can be described under a statistical point of view as a charged gas in a quasi-equilibrium state (see Section 1.2.1.4). In this situation, since a very high number of electrons are contained in a volume where statistical variables' variations are negligible, local thermodynamic potentials and variables can be defined, as the electron gas temperature. Local thermodynamic variables changes over macroscopic distances and times, depending on the plasma state. Moreover, the velocity electrons distribution can differ greatly from the Maxwell distribution due to external elements like UV fields, RF fields, or magnetic fields. Similar quasi-static approximations can be done for the ion population and the neutral atoms population, so defining relative local temperatures.

Because of the large difference in mass, the electrons come to thermodynamic equilibrium among themselves much faster than they come into equilibrium with the ions or neutral atoms. For this reason, the ion temperature may be very different, usually lower, than the electron temperature. This is especially common in weakly ionized technological plasmas, where the ions are often near the ambient temperature.

Based on the relative temperatures of the electrons, ions, and neutral atoms, thermal and nonthermal plasmas are defined. Thermal plasmas have electrons and heavy particles in thermal equilibrium, that is, at the same temperature. Nonthermal plasmas, however, have the heavy particles at a much

TABLE 4.15
Typical Ion Energy Requirements for Different Applications in the Field of Planar Technologies

Ion Energy (eV)	Planar Process
1–3	Physical absorption
4–10	Low-energy surface sputtering
10–5000	Sputtering
10,000–20,000	Ion implantation

lower temperature, frequently near room temperature, with respect to the much hotter electrons. On the ground of these definitions, a plasma is sometimes referred to as being hot if it is nearly fully ionized, so to have a high hot electrons density, or cold if only a small fraction molecules are ionized.

Plasmas used in dry etching processes are cold in this sense, even if the electron population is still at several hundred Celsius degree. Typical ions energy requirements for different applications in the field of planar technologies are reported in Table 4.15.

Since plasmas are very good conductors, electric potentials play an important role. The potential as it exists on average in the space between charged particles is called the plasma potential, or the space potential.

If an electrode is inserted into a plasma, its potential will generally lie considerably below the plasma potential due to the so-called Debye sheath, that is the plasma version of the Debye layer shielding that is present in electrolytic solutions. A Debye sheath arises when a conductive surface is inserted inside the volume occupied by the plasma. It is generated by the fact that electrons usually have a temperature on the order of magnitude or greater than that of the ions and are much lighter. Consequently, they are much faster than the ions. When a conduction interface is thus inserted into the plasma, electrons fly fast and absorb into it charging the interface with negative charges, while the surrounding plasma volume is charged positively due to the electrons lost. New electrons coming in near the surface due to diffusion of electrical attraction by local ions are reflected by the charged surface up to the moment in which an equilibrium is created.

In the equilibrium state there is a transition layer around the conducting surface, whose thickness is equal to the Debye length, comprises between the negatively charged surface and the overall neutral bulk plasma, where the plasma is positively charged. This transition region is called Debye sheath of Debye layer. Similar physics is involved between two plasma regions that have different characteristics; the transition between these regions is known as a double layer, and features one positive, and one negative layer (not to be confused with the double layer arising in electrochemical cells).

These and other similar effects result in the important concept of quasi neutrality of the plasma, that means that the plasma is overall neutral if a volume much greater than L_D^3 is considered, while locally, that is between points whose distance is of the order or smaller than the Debye distance, relevant potentials can be present.

Plasma with a magnetic field strong enough to influence the motion of the charged particles is said to be magnetized. A common quantitative criterion is that a particle on average completes at least one gyration around the magnetic field before making a collision, that is, $\omega_{ce}/v_{coll} > 1$, where ω_{ce} is the electron gyrofrequency [91] due to the magnetic field and v_{coll} is the electron collision rate [92]. It is often the case that the electrons are magnetized while the ions are not. Magnetized plasmas are anisotropic, meaning that their properties in the direction parallel to the magnetic field are different from those perpendicular to it.

4.7.2.2 DC Glow Plasma Generation

Plasmas are generated by supplying energy to a neutral gas causing the formation of charge carriers. Electrons and ions are produced in the gas phase when electrons or photons with sufficient

energy collide with the neutral atoms and molecules [93]. There are various ways to supply the necessary energy for plasma generation to a neutral gas. One possibility is to supply thermal energy, for example in flames, where exothermic chemical reactions of the molecules are used as the prime energy source. Adiabatic compression of the gas is also capable of gas heating up to the point of plasma generation. Yet another way to supply energy to a gas reservoir is via energetic beams that moderate in a gas volume.

The most commonly used method of generating and sustaining a low-temperature plasma for technological applications is by applying an electric field to a neutral gas. Any volume of a neutral gas always contains a few electrons and ions that are formed, for example, as the result of the interaction of cosmic rays or other radiations with the gas. These free charge carriers are accelerated by the electric field and new charged particles may be created by collisions with neutral atoms. This leads to avalanche of charged particles that is eventually balanced by charge carrier recombination, so that a steady-state plasma develops. In planar technologies, plasma is generated, for example in dry etching machines, mainly by using electrical fields: DC, pulsed DC, or RF and microwave frequencies. Nonthermal plasma in DC discharges is generally created in closed discharge vessels using interior electrodes, as shown in Figure 4.61 [90,93].

Low-pressure, normal glow discharges between planar electrodes in a cylindrical glass tube exhibit characteristic luminous structures as evidenced in Figure 4.61. The brightest part of the discharge is the negative glow, which is separated from the cathode by the cathode dark space (called “Crookes” or “Hittorf dark space”). The cathode dark space is a region of the discharge where the electrical potential drops drastically, as shown in Figure 4.61b. The negative glow is separated from the cathode dark space by a well-defined boundary and it is followed by a diffuse region in the direction toward the anode. The negative glow, where the electric field is close to zero, and the positive column are separated by the “Faraday dark space.” The homogenous or striated positive column stretches all the way to the anode, which may be covered by a characteristic anode glow. The variations of plasma potential along the length of the discharge tube are shown schematically in Figure 4.61b.

The microscopic processes in such a discharge can be described as follows. A positive ion from the negative glow is accelerated by the electric field in the cathode fall and directed toward the cathode surface. The collision of the energetic ion with the surface produces secondary electrons, which are subsequently accelerated in the cathode fall to comparatively high energies. These energetic electrons transfer most of their energy to heavy particles like atoms or molecules in inelastic

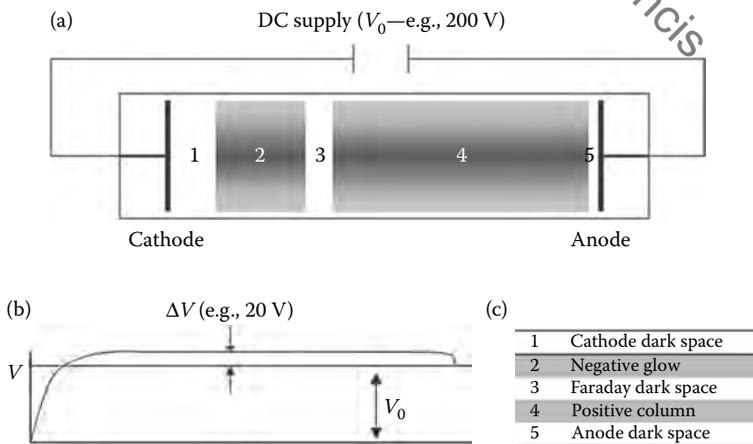


FIGURE 4.61 Schematic representation of the system for plasma generation via DC glow (a) and of the potential inside the void tube during the glow (b). Number labels from part (a) indicate the different plasma regions as reported in part (c).

TABLE 4.16
Minimum Breakdown Voltage (V_{\min}) and the Product
between the Gas Pressure P and the Discharge Chamber
Length in the Direction of the Glow L for Various Gases

Gas	V_{\min} (V)	$P L$ (kPa cm)
Ar	137	6.75
H ₂	273	8.625
He	156	30
CO ₂	420	3.825
N ₂	251	5.025
N ₂ O	418	3.75
O ₂	450	5.25
SO ₂	457	2.475
H ₂ S	414	4.5

collisions. Such collisions generate excitation and ionization of the heavy particles primarily in the cathode fall and in the negative glow region thereby creating additional charge carriers [89,90].

The cathode regions of the discharge play a crucial role in sustaining the glow discharge. The positive column is formed only in the presence of a long, narrow discharge gap with charge carrier losses to the wall. In the homogenous positive column, a constant longitudinal electrical field is maintained. The electrons gain energy in this field and form an electron energy distribution with an appreciable number of energetic electrons that allows a sufficiently large number of ions and electrons to be generated and charge carrier losses due to walls absorption and recombination to be balanced. The minimum breakdown voltage needed to generate a glow and the product between the gas pressure and the chamber length in the direction of the glow are reported in Table 4.16 for various gases. These values are useful to have a first idea of the field and the pressures that have to be created to obtain and sustain DC generated plasma.

4.7.2.3 RF Plasma Generation

To generate the plasma both the cathode and the anode have to be conducting materials. Looking at Figure 4.61, it is clear that it is possible to exploit the high acceleration the ions undergo near the cathode hitting the electrode at high speed to generate dry etching. This is not, however, possible if the surface to be etched is not a conducting surface. As a matter of fact, if an insulating film is present upon the conduction cathode, the incoming ions are frequently captured into the insulating film by positively charging it. This creates near the cathode a positive film that tends to deflect other incoming ions and, at last, to extinguish the plasma. A possible solution to this problem, that is frequently adopted in dry etching systems, is to use alternate current.

If the polarity of the plasma generating field changes periodically, the ions that are absorbed in an half period by the insulating film to be etched are freed in the other half period, where the electrical field tends to launch them again into the plasma. Thus, if the variation frequency is sufficiently rapid, the screening effect due to charge absorbed into the insulating film can be made negligible.

Different system for plasma generation corresponds to different frequencies of the electrical field and from the way in which it is prevalently coupled with the plasma, exploiting a capacity of an inductive effect [93]. Capacity coupled RF plasmas are still the most common plasmas used in dry etching. The scheme of a typical reactor chamber for capacity coupled RF plasma generation is shown in Figure 4.62. The power is applied to the lower or the upper electrode. In general, the frequency of the applied field is 13.56 MHz. A so-called dark sheath is formed in the neighborhood of all surfaces in the reactor, electrodes, and walls. This dark sheath can be considered as some kind

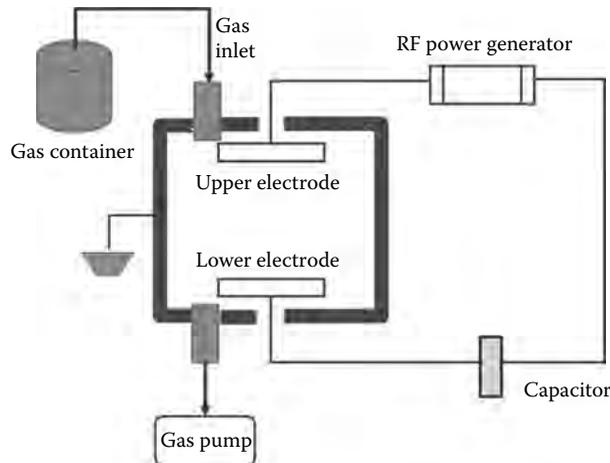


FIGURE 4.62 Scheme of a typical reactor chamber for capacity coupled RF plasma generation.

of dielectric or a capacitor, similar to what happens for Debye capacity in electrolytic cells. Thus, the applied power is generally considered as transmitted to the plasma through a capacitor.

At frequencies between 1 and 100 MHz, the free electrons are able to follow the variations of the applied electric field and, unless they suffer a collision, can gain considerable energy, of the order of some hundred eV. However, in this frequency range, the movement of the much heavier ions is very little influenced by these electric fields: their energy comes completely from the thermal energy of the environment and is of the order of a 0.01 eV so that the plasma is of the nonthermal type.

In the pressure range of these plasmas, from about 10 Pa to a few kPa, the electrons will travel much longer distances than the ions, and in this way, they will much more frequently collide with the reactor walls and electrodes and consequently be removed from the plasma. This would leave the plasma positively charged. However, in order to optimize etching by maintaining the plasma quasi-neutrality, a DC electric field has to be formed in such a way that the electrons are repelled from the walls.

The capacitor between the power generator and the electrode shown in Figure 4.62 helps to form the DC charge. During the first few cycles, electrons generated in the plasma escape to the electrode and charge the capacitor negatively. In this way, a negative DC bias voltage is formed on the electrode, which repels the electrons. The RF voltage becomes then superimposed on this negative DC voltage whose behavior is similar that typical of DC plasma generators. Thus in the initial transitory electrons are absorbed in the electrodes and charged; this process arrives at an equilibrium when the field so generated repels other electrons avoiding the further capacitor charging and generating the DC field that maintain plasma quasi-neutrality and allows an efficient etching process to happen in the plasma chamber.

Although capacity coupled RF plasma generation is easy to design and control, it has also a certain number of limitations. The first limitation is that the reactive particle density is directly coupled to the ion energy. If a dense plasma is desired, rich in free atoms that are in general the particles which chemically react with the surface undergoing etching, the plasma is also rich in high energy ions. Moreover, to obtain high densities of reactive particles, the power in the plasma has to be increased. This increase of power will also increase ion density and energy. Increasing the pressure can increase the reactive particle density and decrease ion density and energy somewhat, but not to a great extent: in general, the effect of increasing the pressure is much lower than the effect of increasing the power.

Thus if highly reactive plasma is required, to attain a mainly chemical action with little bombardment, these types of plasma are not suitable. They will also not be very useful for the “inverse” type of plasma: a plasma with a low density of neutral and reactive atoms and a very high energetic ion density.

A second drawback is that it is not possible to generate plasmas at low pressures: typically, the lowest pressure at which a plasma can be sustained is of the order of 70 Pa. At lower pressures, there are not enough collisions to generate enough free electrons to sustain the plasma. Low-pressure plasmas are required in most advanced PE machines to obtain very high-aspect ratio structures. These can only be obtained if the ions collide with the wafer surface at a nearly perpendicular angle. To obtain this condition, little or no collisions should take place in the dark sheath, thus a large mean free path is needed. Therefore, the pressure must be reduced as much as possible.

In order to overcome these limitations, inductive coupling plasma generation can be used. There exist two types of inductively driven sources: using cylindrical or using planar geometries, as shown in Figure 4.63. The use of multipole permanent magnets is not indispensable, but their presence will increase the plasma density and mainly the uniformity of the plasma. An RF voltage is applied to the coil, resulting in an RF current which induces a magnetic field in the reactor. Therefore, the wall has to be a dielectric.

It is also possible to apply an extra (RF, low frequency, or DC) bias voltage to the substrate holder, as shown in both figures, to increase the ion bombardment on the substrate. This voltage is, in general, small and does not affect the plasma characteristics: the ions and electrons are mainly generated by the inductive coupling. In this way, it is possible to control independently the plasma density and the energy of the incoming ions. This gives the process engineer an extra parameter with which to optimize the process characteristics.

The preferred dielectric material for the reactor walls is alumina, which has excellent electric characteristics, is not etched by the most commonly used etchants, but is hard and expensive to manufacture. If no plasma is formed, the magnetic field generated by the coil, enters the etching reactor. If plasma is generated, an electric field can be formed in the reactor, because of Faraday's law. This electric field creates a current in the plasma, and the resulting total magnetic field will be null in the reactor. The absorbed power in the plasma is then proportional to the real part of the product of the current and electrical field vectors in the plasma. Ion densities of the order of 10^{17} – 10^{18} ions/m³ at pressures lower than 140 Pa, can be obtained in these discharges. This is one to two orders of magnitude higher than for traditional capacity coupled plasmas. However, an RF power of at least 100 W is needed to sustain the inductively coupled plasma.

Besides the inductive coupling, there is also a small capacitive coupling: the dielectric serves as the dielectric of a capacitor formed between the lower part of the coil and the plasma. At the high voltage end of the coil, RF voltages of the order of 2000 V have been measured. Therefore, a capacity coupled plasma is also formed. This capacitive coupling can help to strike and sustain the plasma. However,

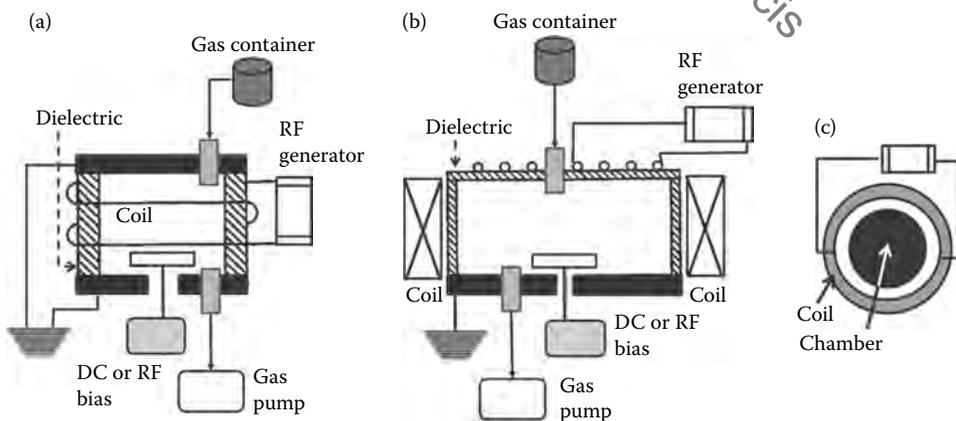


FIGURE 4.63 Schemes of the two types of inductively driven plasma sources: source with planar geometry (a) and with cylindrical geometry (b). A schematic top view of the cylindrical geometry system is represented in (c).

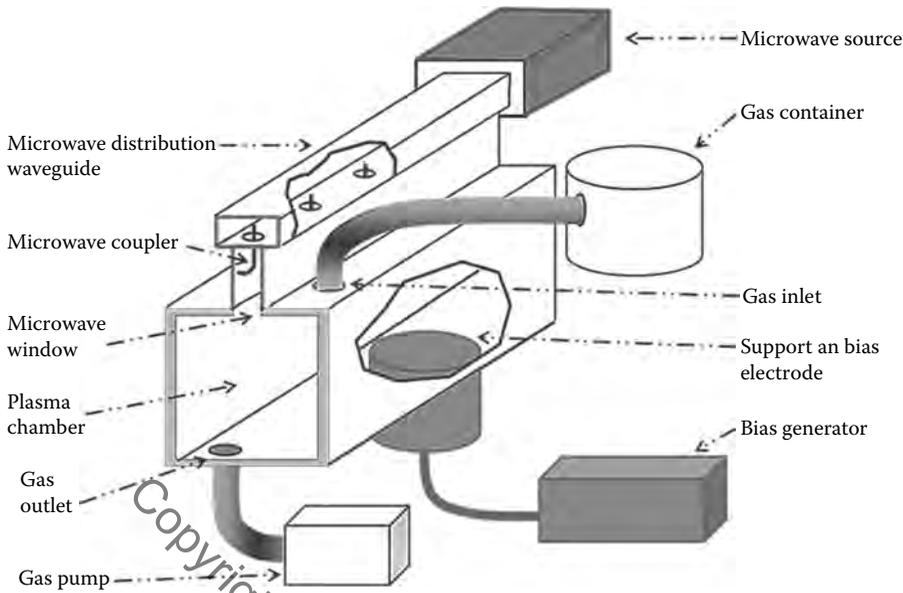


FIGURE 4.64 Scheme of a possible architecture for a microwave plasma generator.

a local DC voltage can be formed, which results in sputtering of the dielectric. The presence of dielectric material in the plasma can induce serious contamination on the wafer, or chemical changes in the plasma, and has to be avoided. Therefore, it is necessary that the dielectric plate is thick enough to reduce the capacitive coupling. Another way to decrease the capacitance of the coil is to place it a few millimeter above the dielectric, although this makes the manufacturing a little more difficult.

The principle of RF plasma generation can be extended also up in frequency to arrive to microwave frequencies. Microwave plasma generation is peculiar under various points of view. Microwaves have to be guided by some sort of waveguide to avoid dispersion of the microwave power. Thus, a simple structure like those represented in Figure 4.63 cannot be used. Several different structures have been implemented for microwave plasma generators [93–95], all operating in microwave resonant cavities under maser threshold and in traveling wave microwave guides. A possible scheme of a microwave plasma generator for application in dry etching is shown in Figure 4.64 [96]. Here microwaves propagate into a square conducting waveguide that is periodically coupled with the plasma chamber via a set of microwave antennas. The wave propagating into the guide losses energy during propagation by injecting it into the plasma chamber where the gas is ionized and plasma is created and maintained. The support presents also a DC-like bias to regulate independently the incident ions energy, like in RF systems.

4.7.3 DRY ETCHING TECHNIQUES

Dry etching can potentially combine both chemical etching, using particles that chemically combine with the film to etch, and physical etching, removing film layers exploiting kinetic energy of plasma ions via collisions.

Combining differently these two mechanisms we have three classes of dry etching techniques [1,3]:

- Sputtering etching: purely mechanical
- Reactive ion etching: mixed chemical and mechanical
- Chemical dry etching: purely chemical

Depending on the etching machine architecture and on the used materials each dry etching family is composed of a great number of different techniques, sometimes simply different variations of the same base idea, other times exploiting completely different arrangements.

4.7.3.1 Sputtering Etching (Ion Milling)

The simpler form of dry etching technique is the sputtering etching, also called ion milling. This etching technique can be directly implemented by exploiting the fact that ions are accelerated in the dark zone around the cathode of a DC or RF glow so as to hit the cathode at high energy. Thus, if the wafer is placed on the cathode its surface is etched by incoming ions. A better control of the process, however, is attained by dividing the plasma generation and the etching process. In this case, ions are extracted from the plasma and focused as a beam on the wafer. This technique is also called ion beam etching.

A possible architecture of a sputtering dry etching machine using a combination of DC and RF plasma generation is reported in Figure 4.65. The plasma chamber is divided by the etching chamber: the plasma is generated in the plasma chamber by a circular inductive coupled RF plasma generator, the ions are extracted so as to obtain an ion beam that is focused in a different chamber on the substrate. Electrons are maintained out of the etching chamber by filtering them using conductive grids, vacuum in the etching chamber is maintained by a vacuum pump that also removes low-energy ions that either are scattered or do not hit the substrate. The target is exposed to the ion beam with a certain angle θ with respect to the beam average direction so as to have a further control variable to regulate the process. Due to its energy any incoming ion can extract one or more atoms from the surface of the wafer, so etching it.

Sputtering etching is purely mechanical and has generally good anisotropy characteristics. The typical profile of a pattern transferred from the mask to the underlying film via sputtering etching is reported in Figure 4.66. The angle that the etched area walls form with the vertical is due to the scattering of ions and extracted particles from the surface. Scattering takes place with a random

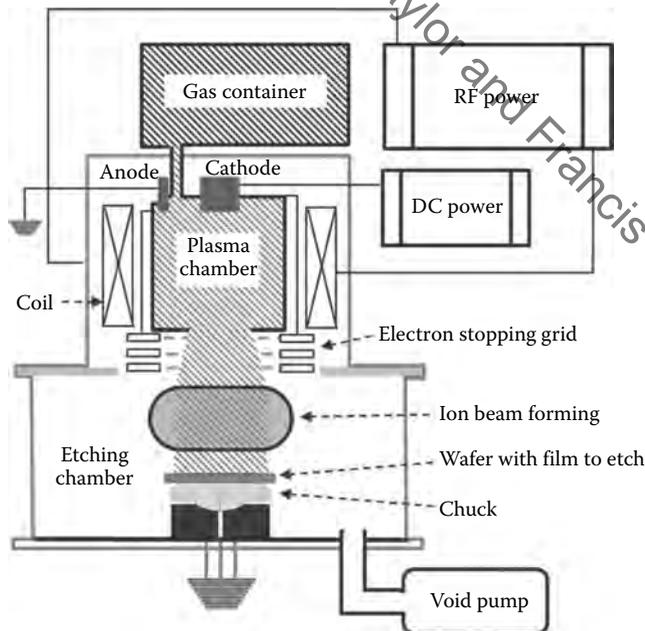


FIGURE 4.65 Possible architectures of a sputtering dry etcher that used both DC and RF plasma generation with separated plasma and etching chambers. An inductive coupled cylindrical plasma RF generator is used and electrons are maintained out of the etching chamber using a conductive grid.

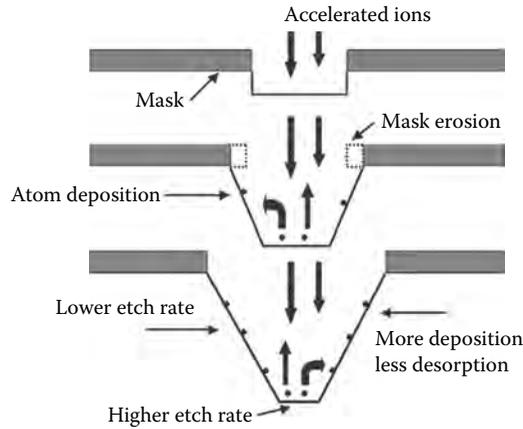


FIGURE 4.66 Schematic of the sputtering etching process evidencing the characteristic section of the etched pattern and the phenomenon of mask erosion.

deviation with respect to the vertical, causing both etching of the lateral walls due to deflected ions and deposition of deflected particles that build new layers on the wall itself.

The second phenomenon is by far prevalent with respect to the first since deflected ions have low energy due to the small collision angle. Since the deposition probability is higher if the impact energy of the deposited particles is higher, the phenomenon is greater near the bottom of the etched area, from which the characteristics form that is shown in Figure 4.66.

Sputtering etching is not very selective due to its purely mechanical nature. The first effect of this fact is that the etching mask, both in case it is constituted by hardened resist or by a protecting film like SiO_2 over Si, is also partially etched, as shown in Figure 4.66. This phenomenon contributes to the characteristic trapezoidal form of the etched pattern section. Moreover, even if generally the substrate has a quite lower etching rate, it is in any case etched too when etching arrives at the film–substrate interface. Due to this fact, a careful control of etching rate and time has to be maintained not to provoke sensible substrate damage below the etched area.

One key parameter of sputtering etching is the energy of the incoming ions that has to be carefully controlled. A low energy means that the probability for extracting atoms from the film surface is low and the etching rate is also low, or even etching does not happen. However, a too high ion energy causes almost any incoming ion to create an extended damage on the film surface. The characteristics of such damage in terms of depth and extension are practically impossible to control, so that etching is highly uneven. In order to avoid the two extreme detrimental regimes of no etching and surface damage, every sputtering etching machine is characterized by the so-called yield curve, providing the average number of surface atoms extracted by a single incoming ion versus the average ion impact energy.

An example of dependence of the etching yield from the energy of the incoming ions (argon in this case) is presented in Figure 4.67. Here, two thresholds are evidenced: the first is the threshold above which the probability to obtain an atom extraction is near zero and extractions almost never happens. Below this threshold sputtering etching do not happens. The second threshold is the threshold above which the average number of atoms extracted by a single incident ion is greater than 1. Above this threshold, the process is less controllable and frequently it is not possible to have a good etching result.

In the great majority of sputtering etching machines the substrate is exposed to the etching ion beam at a certain angle θ with respect to the beam axis (cf. Figure 4.65). This angle is another important control parameter of the etching process regulating in particular the etching rate. A typical behavior of the etching rate versus the substrate angle is reported in Figure 4.68. The yield, or

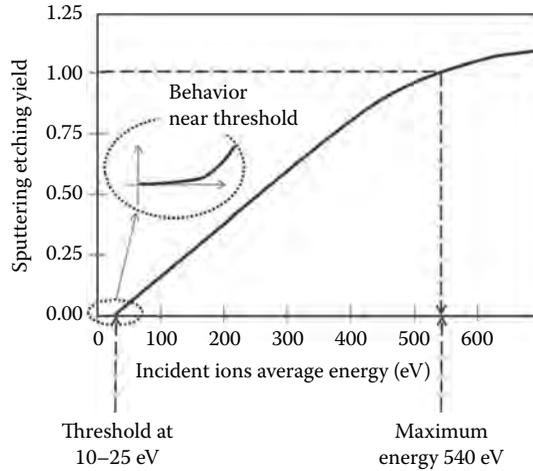


FIGURE 4.67 Example of yield of a sputtering etching process designed to etch (111) silicon after masking with UV hardened PMMA resist.

equivalently the sputter etching rate varies from one material to another, but its fluctuations generally do not exceed one order of magnitude. Using Ar^+ ions with energies of the order of 500 eV the yield is generally of the order of 1, as shown Table 4.17, with the exception for Al_2O_3 , that is, an extremely low etching material.

This fact causes the difficulty to provide a large difference of etching rate between the etching mask material and the film material to reduce the V shape of the etched form. Moreover, due to the dependence of the etching rate on the exposure angle, the lateral walls of the etching mask layer after the first etching phase are also etched at a different rate with respect to the underlying film, creating the facet effect, which is schematically shown in Figure 4.69a. This effect, besides the fact that the etched material also deposits after extraction from the film on the side walls of the etching mask, tending to form surface defects when the etching mask is removed (see Figure 4.69b), renders the achievement of high anisotropy pattern transfer and high flat surface after etching and mask removal difficult to achieve using sputtering etching.

These limitations are in great part overcome by the use of reactive ion etching, that is, a mixed physical and chemical form of PE.

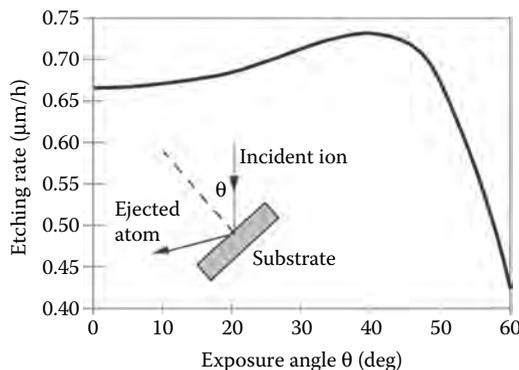


FIGURE 4.68 Example of dependence of the sputtering etching rate from the exposure angle. The definition of the exposure angle is reported in the insert of the plot.

TABLE 4.17
Sputtering Etching Rate at the Optimum Exposure Angle
Using 500 eV Ar⁺ Ions at a Current Density of 0.065 A/m²
to Etch Different Materials

Material	Etching Rate (Å/s)	Yield
Al ₂ O ₃	1.4 (±0.2)	0.08
Si (111)	4.39 (±0.1)	0.54
GaAs	14 (±1)	0.76

Note: The yield is defined as the average number of atoms (or molecules) extracted from the film for each incident ion.

4.7.3.2 Reactive Ion Etching: Plasma Etching

Under the name of reactive ion etching several different techniques are collected, all of them having in common the use of plasma and the presence of a chemical etching effect on the surface of the etched film. Generally, the chemical etching effect due to neutral particles that are pushed by the plasma toward the wafer surface is combined with a physical etching effect due to bombardment with high-energy ions. However, PE is also widely used, where etching is uniquely due to chemical effect and the presence of plasma is only used to generate the etchant species and sometimes to enhance their chemical activity. The main reason to resort to PE is to avoid surface damages due to ions physical etching and contemporarily to avoid the use of very aggressive chemical etchants like those used in wet etching.

PE is generally composed by the following steps:

- The etchant species is formed by plasma in the plasma chamber.
- While the plasma is confined in the plasma chamber the neutral etchant diffuses through a neutral gas layer into the etching chamber and arrives to the wafer surface.
- Etching happens on the wafer surface due to the formation of a volatile compound by the combination of the etchant molecules and the superficial plasma film.

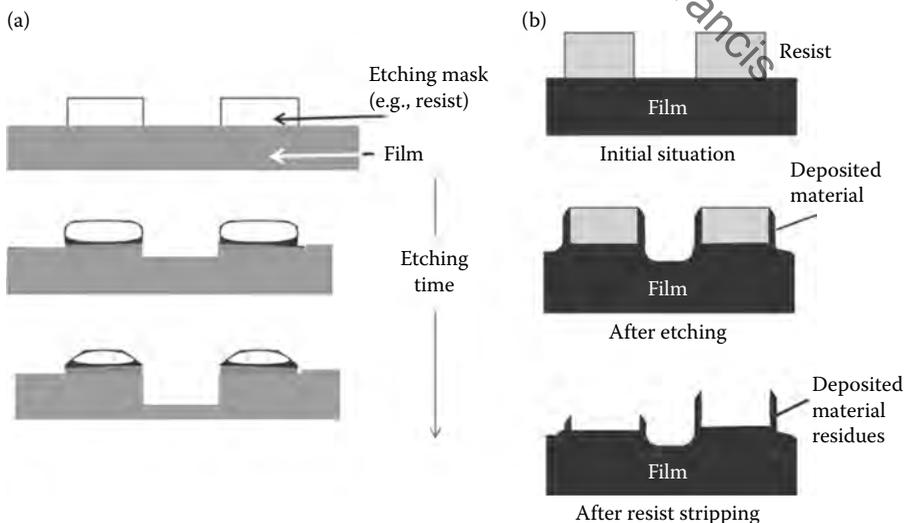


FIGURE 4.69 Facet effect (a) and side deposition on the etching mask (b) in sputtering etching.

- The volatile compound leaves the surface allowing deeper and deeper etching to be performed.

Two types of PE reactors are used: the first is very similar to the reactor shown in the previous section in the case of sputtering etching. Plasma is formed in a plasma chamber separated by the etching chamber. While plasma is confined by two separated sets of electrical charged rings into the plasma chamber, neutral elements that are meant to generate chemical etching are allowed to diffuse into the etching chamber where they arrive on the wafer surface. In this case, differently from the case of sputtering etching, the wafer is exposed to etchants diffusion horizontally, to optimize uniformity. The etching chamber is connected to the vacuum pump that both removes the results of the etching reaction and the unused etchant molecules, maintaining a low pressure in the etching chamber.

Besides this type of vertical reactor, the barrel reactor is also used for PE, whose principle scheme is shown in Figure 4.70. The reactor is formed by two coaxial cylinders: the inner cylinder is perforated to allow the diffusion of neutral particles and charged so as to confine plasma. Plasma is generated, generally by RF or microwaves, in the space between the two barrels, where also the gases that have to react to form the etching composite are injected. The inner cylinder allows the diffusion of neutral etching particles versus the inner part of the reactors, where a batch of wafers is present, while plasma is maintained in the space between the cylinders. A vacuum pump assures low pressure in the inner barrel at every moment of the process.

Since PE is a purely chemical etching it has the same isotropy properties of wet etching and the profiles of the etched patterns are quite similar to wet etching profiles.

A relevant effect in PE is loading effect. While in the case of wet etching the number of etchant particles is much greater than the number of particles of the film to be etched, this is not true in the case of PE. Here, the etching process depletes the etchants population causing two different effects called microscopic and local loading.

Macroscopic loading is because of the depletion in the etchant population due to the overall wafer etching. Due to this effect the etching rate is almost proportional to the area to be etched on the wafer. This also means that small variations in the etchant flux also cause a variation in the etchant rate, provoking an extreme difficulty in controlling process uniformity on similar wafers. The gas flow rate, generally of the order of 5–200 cm³/min at standard conditions, is the most important parameter to be controlled in PE so as to assure uniformity. To prevent problems due to lack of etchant molecules, the gas flow utilization U is maintained generally greater than 0.1. The gas flow utilization is defined as the rate between the etching product formation rate and the inner gas flow rate.

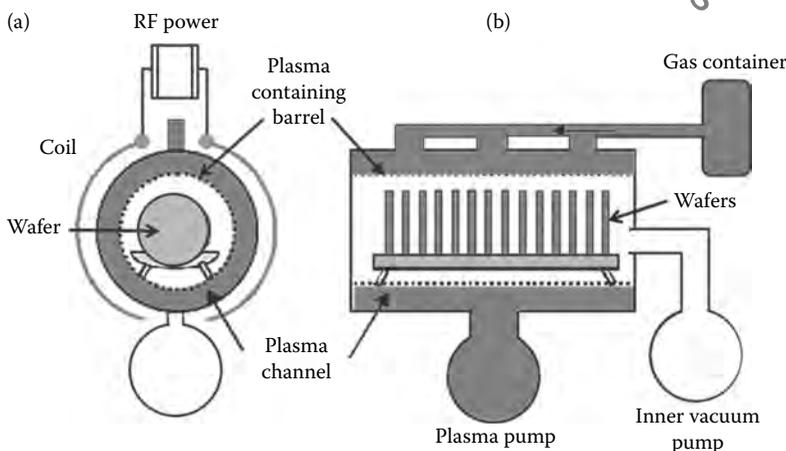


FIGURE 4.70 Principle scheme of a barrel reactor for PE: front section (a) and side section (b).

TABLE 4.18

List of Commonly Used Etchant for Different Materials in PE Reactors and Associated Etch Ratios in Å/min

Etchant	Material to Be Etched					Reference
	Si	Al	Resist	Quartz	SiO ₂	
Ar	124	166	185	159		[1]
CCl ₂ F ₂	2200	1624	410	533		[1]
CF ₄	900	200	50	75		[1]
C ₂ F ₆ -Cl ₂	600 (undoped)				100	[1]
CCl ₃ F	1670					[1]
CF ₃ CF=CFCF ₃ (*)					5630	[97]
CF ₃ CF=CF ₃ (+)					5380	[97]

Note: (*) indicates perfluoro-2-butene and (+) indicates hexafluoropropene.

Also, a local loading effect can happen, due to the nonuniformity of the etchant distribution on the surface of the same wafer so that larger etching areas are etched at a lower etching rate. This effect also has to be limited by a constant and sufficient gas flow during the etch process. A list of commonly used etchant for different materials and the associated etch ratios is reported in Table 4.18.

4.7.3.3 Reactive Ion Etching: Physical plus Chemical Etching

The most flexible and powerful family of dry etching processes is a combination of sputtering and PE, that is a combination of physical and chemical etching [1,98]. A first example of this type of process is the reactive ion beam etching (RIBE). In this case, the ions used to bomb the wafer surface are also reactive and colliding at high energy on the wafer surface combined with the etched film atoms generating volatile composites.

Even if RIBE can produce quite performing processes, more frequently mixed dry etching is realized using the chemically assisted reactive ion beam etching (CARIBE). In this case, ions are not reactive, performing a purely physical etching. Chemical etching is performed by another reactive composite diffusing in the atmosphere of the etching chamber. Frequently, ion bombing also increases the reactivity of the etched film surface by creating a cooperative effect with the chemical etchant. As a matter of fact, ion bombing, besides extracting atoms from the wafer surface, also provoke damages in terms of bond break and dislocations in the film surface. These structures present exposed active sites and are much more reactive in the presence of the neutral etchant with respect to the film itself.

Since defects are generated mainly by vertically incident ions, the defects tends to be mainly vertical with respect to the film surface and also the chemical part of the etching tends to be strongly anisotropic, creating the possibility of strong overall anisotropy.

This characteristic generates a large improvement of the etch rate with respect of pure chemical or sputtering etching if CARIBE is used, being the CARIBE rate generally much higher than the sum of the chemical and the pure physical etching rates. As an example, sputtering etching of (111) silicon using Ar⁺ ions reaches a rate of the order of 0.8 nm/min, while pure PE using chemical etchant can attain a rate around 0.5 nm/min. If CARIBE etching is used adopting Ar⁺ plasma and neutral XeF₂ molecules to provide chemical etching, the overall rate can arrive up to a steady state value of 5.7 nm/min, much higher than the sum of the rates of the individual techniques [1].

The presence of a combined effect of physical and chemical etching also allows lateral etching inhibitors to be used, that are intended to further increase the etching anisotropy. A scheme explaining the working of this technique is reported in Figure 4.71: the etching chamber atmosphere contains, besides plasma and the chemical etchant, another composite driving the inhibition process. When particles of the film to be etched are extracted they combine with the inhibitor forming

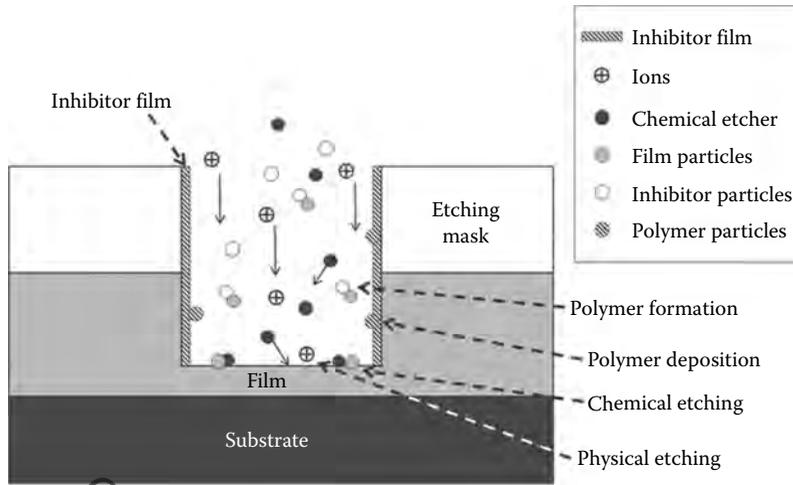


FIGURE 4.71 Inhibitor-driven anisotropy process scheme.

generally an inert polymer that deposits on the vertical walls of the etched area coating them with an inert polymer film. Due to this process, the chemical effect happens only on the horizontal surfaces and atoms or molecules extracted by the ion bombing cannot redeposit on the lateral walls due to the presence of the inhibitor film that does not absorb them.

The presence of the inhibition effect besides the natural anisotropy of the CARIBE allows high-aspect ratios to be achieved by this technique also with very small critical dimensions patterns. A typical example of inhibitor-driven anisotropy is provided by the addition of H_2 to a RIE process for Si etching by CF_4 plasma also causing chemical etching.

The bottom of the etched area results to be biased electrically by ions absorption while the lateral silicon is not biased due to the good focus of the ion beam realizing physical etching. Thus, a difference exists in the etching rate due to electrical bias between the bottom and the lateral walls, as shown in Figure 4.72a. Adding hydrogen to the etching chamber both the floor and the walls etching rates decreases at the same pace so that there is a well-defined hydrogen concentration at which the rate on the lateral walls is zero and the etching is ideally perfectly anisotropic, as shown in Figure 4.72b.

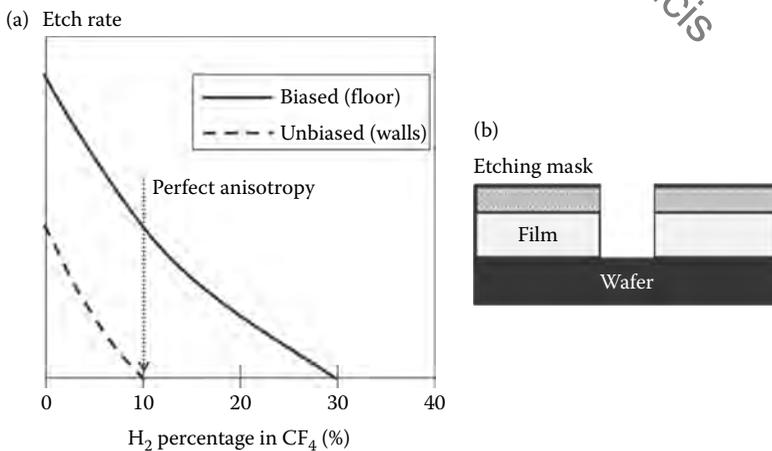


FIGURE 4.72 Hydrogen addition to Si RIE via CF_4 plasma. Effect on the biased (bottom of etched areas) and unbiased (walls of the etched areas) parts of the wafer (a) and perfect anisotropic etching result (b).

Another addition that can be done to regulate the CF_4 -driven Si chemical etching is that of oxygen atoms. Adding oxygen atoms enhances the formation of F atoms and inhibits polymerization. At very high concentration oxygen absorb on the film surface, generating an etching rate decrease. Thus, the effects of a combined addition of hydrogen and oxygen to the atmosphere of the etching chamber can be exploited to fine tuning the etching process, both in the sense of etching anisotropy and in the sense of etching selectivity, enhanced by decreasing the pressure in the etching chamber.

The combination of these effects generates a complex control graph for Si RIE etching in CF_4 plasma that is qualitatively shown in Figure 4.73. The plot represents the effect of different control variables on the characteristics of the etching process. In particular, the H_2 and the O_2 addition effect are considered, besides the increase of the gas pressure on the wafer and of the energy of the incident ions. The resulting plot is divided into three areas: on the bottom right there is the area where isotropic etching is performed, here the chemical etching dominates and polymerization is inhibited by a high oxygen concentration. To make chemical etching dominant, the energy of the incoming ions has to be very low (or no incoming ions have to be present) and the pressure of the gas has to be high so as to favor chemical etching.

On the left of the plot there is a polymerization zone. Here, whatever be the ion energy and the gas pressure, a very high hydrogen concentration provokes a massive polymer formation that covers all the wafer, inhibiting etching. Naturally, higher the ions energy, higher the hydrogen concentration needed to reach this regime, so that higher the gas pressure, more efficient the polymerization process.

In the middle, between the two extreme regime, there is the RIE regime, where anisotropic etching happens due to a combination of physical and chemical etching. Chemical etching is favored by increasing the gas pressure, so that anisotropy decreases in this case. Similarly, physical etching is favored by increasing the ion energy so that in this case a more anisotropic etching is achieved. Anisotropy is also favored by hydrogen addition due to the inhibition-driven process and reduced by oxygen addition.

RIE Si etching is only a possible example of how the performances of an RIE can be tuned by regulating conditions in the etching chamber and composition of its atmosphere. Similar control charts exist for all widely industrial used etching processes, such as SiO_2 and Si_3N_4 RIE etching.

An example of strongly anisotropic RIE etching in SiO_2 is reported in Figure 4.74, where a SEM photograph of a grating excavated in SiO_2 by strongly anisotropic CARIBE is shown. The grating aspect ratio is 6.5 with a critical dimension of the order of 400 nm. The structure was defined by direct write e-beam lithography.

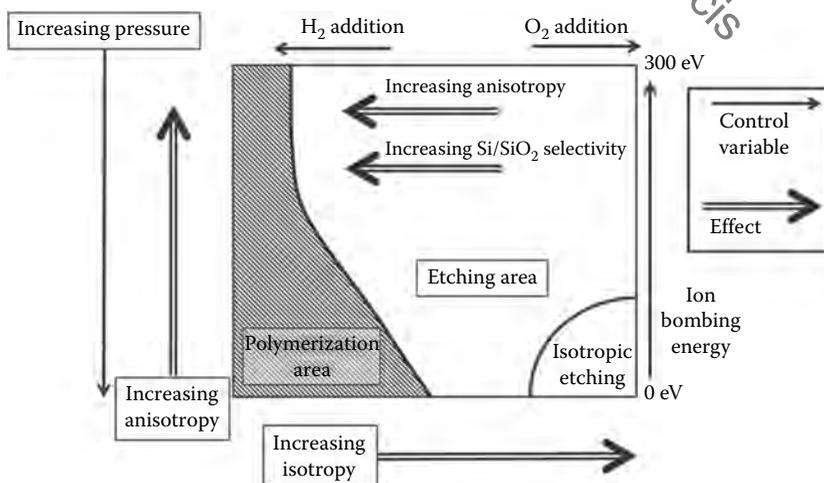


FIGURE 4.73 Control graph for Si RIE etching in CF_4 plasma.

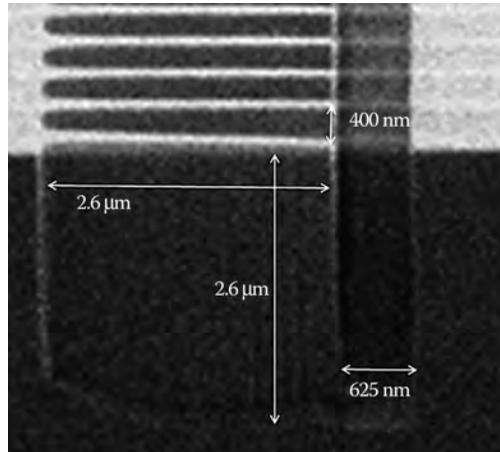


FIGURE 4.74 SEM photograph of a grating excavated in SiO_2 by strongly anisotropic CARIBE. The Grating aspect ratio is as high as 6.5 with a critical dimension of the order of 350 nm. The structure was defined by e-beam direct write lithography. (Published under permission of Cyoptics Inc.)

4.7.3.4 Deep Reactive Ion Etching

Deep RIE process is a variation of RIE whose goal is to generate structures with a very high-aspect ratio. In principle, RIE technologies are able to reach very high-aspect ratios, especially if physical etching is prevalent. However, a certain number of phenomena limits the maximum reachable aspect ratio. A first limitation is constituted by local loading. Even if this phenomenon is not so important in RIE as in PE, due to the contemporaneous presence of ion bombing, it is anyway present. If a great-aspect ratio is wanted local loading can become important.

If loading makes chemical etching less effective, physical etching is thwarted by the fact that the ion beam focus is not perfect. Increasing the etched trench depth, the probability that an incoming ion collide with the walls of the trench instead with the floor greatly increases. Walls collision generally do not generate walls etching, due to the small collision angle, but slower the ions rendering less effective the successive collision with the trench bottom. This effect is clearly shown in Figure 4.75, where the relative average ion beam energy generating etching on the trench bottom is shown versus the ion beam percent angle aperture and the trench aspect ratio. The ion beam is supposed to be a Gaussian beam, as it is frequently true. It is clear that increasing the trench aspect ratio the effectiveness of the physical etching decreases rapidly reaching very small values.

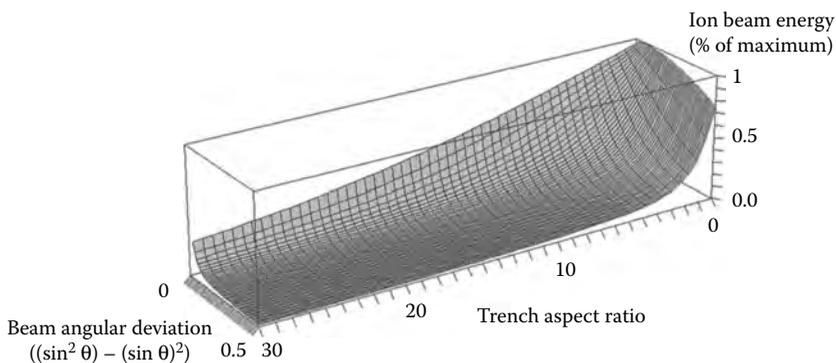


FIGURE 4.75 Relative average ion beam energy generating etching on the trench bottom is shown versus the ion beam percent angle aperture and the trench aspect ratio. The ion beam is supposed to be a Gaussian beam.

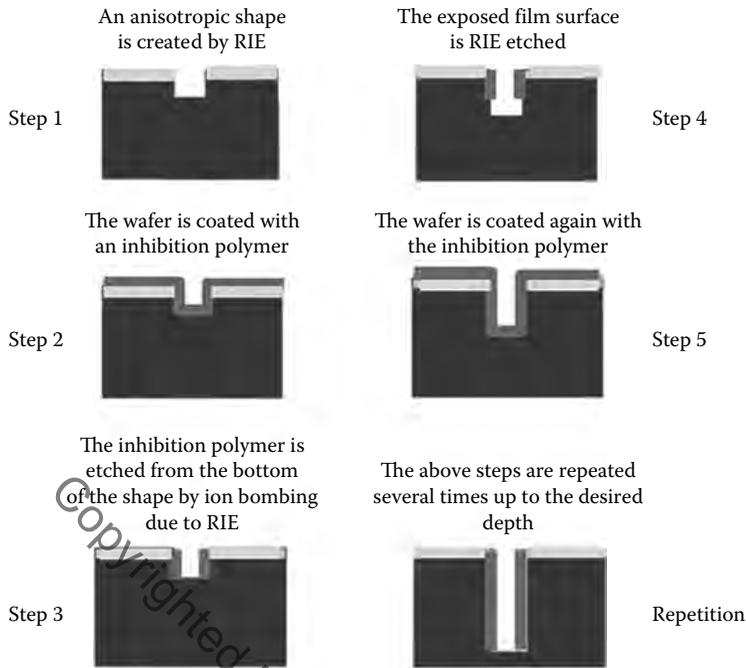


FIGURE 4.76 Scheme of the Bosh DRIE etching process.

The effect is still worse if hydrogen is added to the etching chamber atmosphere, since by decreasing the energy of the ion beam, the risk of entering the polymerization zone and completely inhibiting the etching is higher.

The capability of RIE to achieve high-aspect ratios can be improved by improving the quality of the ion beam (the focus in particular) and increasing pressure of the chemical etchant. However, if very high-aspect ratios like 20 or 30 have to be attained, the process has to be changed more radically.

RIE processes modified for high-aspect ratios are called deep RIE or DRIE. The most diffused DRIE process is the so-called Bosh process, from the industry that patented it. The scheme of the Bosh DRIE etching is reported in Figure 4.76. This process is based on a succession of polymer deposition, polymer etching, and trench deepening. Since the polymer is etched only by ion bombing while it is inert to the chemical etchant, it is consumed only on the trench floor and protect the trench walls from etching allowing exceptional aspect ratios to be reached.

Since a small degree of polymer etching happens also on the trench walls due to ions colliding with the walls themselves, a small ripple is generally present on the trench walls, but it can be well-controlled tuning the process. Using the Bosh process aspect ratios as high as 20–40 can be reached with a very good verticality of the trench walls.

4.8 DEPOSITION

Under the name of deposition are collected all the processes that allows a film of a different material to be deposited on the wafer. Both thin and thick films can be realized with different techniques, starting from few atomic layers to tens of microns, depending on the need. A large variety of materials are deposited as films in planar technologies for different applications. Examples are

- Semiconductors as Si, for example, in SOI wafer realization, SiN_4 , and many others
- Inorganic insulating films such as sapphire and SiO_2 with different doping

- Metals such as gold, platinum, aluminum, and copper
- Polymers such as SU8, PDMS, PMMA, and many others

Essentially, we can classify deposition techniques into three classes [1,3]:

- Chemical deposition techniques, where the film is formed by a surface chemical reaction between the wafer material and a suitable reagent
- Physical deposition techniques, where the film material is directly deposited on the surface of the wafer in its final form
- Molecular deposition techniques, where the physical deposition is controlled at molecular level so as to have a very fine control of the thickness and uniformity of the deposited film

Both chemical and physical deposition techniques can be plasma enhanced, if plasma is used to improve either the film quality control or the deposition rate.

4.8.1 CHEMICAL VAPOR DEPOSITION

Chemical deposition is generally performed using a reagent in vapor phase; hence called chemical vapor deposition (CVD) [99,100]. Almost all types of layers can be produced by CVD; crystals, synthetic diamonds are manufactured by CVD too, so also amorphous and epitaxial.

An epitaxial film is a crystal film that is deposited on a crystal substrate in such a way as to match the two interfacing crystal planes to minimize stresses at the interfaces. The film thus grows following the substrate crystal structure and only in case of very thick films slowly relax on the standalone crystal structure. This, unless the film is made of the substrate material. In this case a continuous crystals formed by the so-called homo-epitaxy. Epitaxial films grow along preferential crystal directions depending on the crystal structures of the film and of the substrate and require similar structures of the two crystals [101–103]. Typical epitaxial growths are GaAlAs on a GaAs substrate or InGaAsP over an InP substrate, but many other examples are used in micro-fabrication technology.

A wide variety of CVD processes exists, that differ by the way the reaction is initiated and by the condition under which reaction is performed.

4.8.1.1 CVD Process Types and CVD Reactors

Considering the operating pressure, CVD processes can be classified as follows:

- *Atmospheric pressure CVD (APCVD)*—CVD processes performed at atmospheric pressure. Reactors for APCVD are generally made by a continuous wafer flow under a set of gas dispensers. The temperature of the wafer is either stabilized at the clean room temperature, or set at an higher value to enhance the deposition rate. Depending on the presence of a single gas dispenser or of a set of subsequent dispensers the APCVD reactors can be classified as gas injection or plenum-type reactor as shown schematically in Figure 4.77. In both cases, when more than one reagent is needed, an inert gas separation is used to avoid reagent mixing in gaseous phase.

Several types of APCVD process exist, depending on the characteristics of the used vapor and on the structure of the vapor preparation unit. For example, in an APCVD based on aerosol vapor, the vapor preparation unit is composed of an aerosol generator.

- *Low-pressure CVD (LPCVD)*—CVD processes at low pressures [104]. Reduced pressures tend to reduce unwanted gas-phase reactions and improve film uniformity across the wafer. Reactors for LPCVD use vacuum pumps to lower the pressure in the deposition chamber and to remove unused chemicals. A widely used reactor architecture is the barrel-type architecture that is shown in Figure 4.78. The barrel-type LPCVD reactor can be charged with containers that can accommodate many wafers and are suitable for large batches.
- *Ultrahigh vacuum CVD (UHVCVD)*—CVD processes at a very low pressure, typically below 10^{-6} Pa. Most CVD processes used in high-volume fabrication are either LPCVD or UHVCVD.

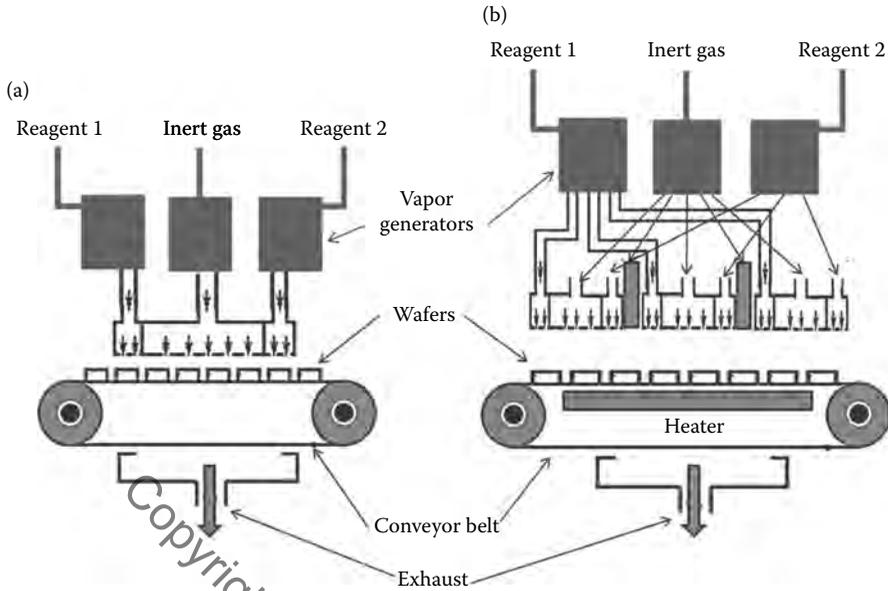


FIGURE 4.77 Gas injection (a) and plenum type (b) APCVD reactors schemes.

As far as the physical characteristic of the vapor used for the deposition we can individuate three CVD classes:

- CVD using neutral vapor
- Plasma-assisted CVD
- Other CVD types

These three families of CDV techniques collect several processes that differ in the way in which deposition is performed. The most used of these processes can be classified as follows.

4.8.1.1.1 Neutral Vapor-Based CVD Techniques

- *Aerosol-assisted CVD (AACVD)* is a CVD process in which the reagents are transported to the substrate by means of a liquid/gas aerosol, which can be generated ultrasonically. This technique is suitable for use with nonvolatile reagents.

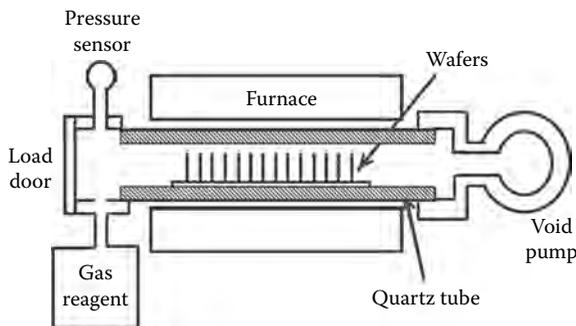


FIGURE 4.78 Barrel type LPCVD architecture using the so called hot-walls system to create high temperature inside the deposition chamber.

- *Direct liquid injection CVD (DLICVD)* is a CVD process in which the reagents are either in liquid form or solid dissolved in a convenient solvent. Liquid solutions are injected in a vaporization chamber toward injectors. Then, the vapors containing the reagents are transported to the substrate as in classical CVD process. This technique is suitable for use on liquid or solid precursors. High growth rates can be reached using this technique.

4.8.1.1.2 Plasma-Enhanced Methods

Somehow these methods are similar to the PE methods, but for the fact that ion energy is not sufficient to generate sensible etching of the wafer surface and the chemical reaction happening in the chamber do not generate volatile composite, but solid composites that deposits on the wafer surface creating a superficial film.

The most important advantage of plasma-enhanced CVD is the possibility to perform deposition at lower temperatures due to the fact that the plasma enhances the chemical reactivity of the CVD reagents.

As a matter of fact, even if the ions energy is not sufficient to provoke sensible etching of the wafer surface, it produces local defects and break chemical bonds on the wafer surface. This exposes reactive sites on the surface, rendering the reaction that gives rise to the film formation much faster. As in the case of dry etching (see Section 4.8.2), plasma can be created by DC, RF or by microwave. When plasma is generated by microwave, the deposition method is often called microwave plasma-assisted CVD (MPCVD).

Essentially, two types of plasma-enhanced CVD method exist, depending on the structure of the CVD reactor:

- *Plasma-enhanced CVD (PECVD)* is the name generally used for the technique using a simple glow reactor with the wafer where the film has to grow placed on the cathode. This method uses the fact that ions are accelerated mainly in the dark region around the cathode to generate an ion flux versus the wafer surface. The chemical reaction can happen either because the ions are directly chemically active or by insertion of an external reagent.

An example of PECVD reactor scheme is reported in Figure 4.79. Here, the plasma is created via capacity coupled RF power and assisted by a DC bias. The wafer is on the plasma chamber cathode and two gas reagents different from the plasma particles are injected at controller pressures in the chamber. To avoid unwanted etching, the plasma is controlled via a control grid that is connected to the electrical ground.

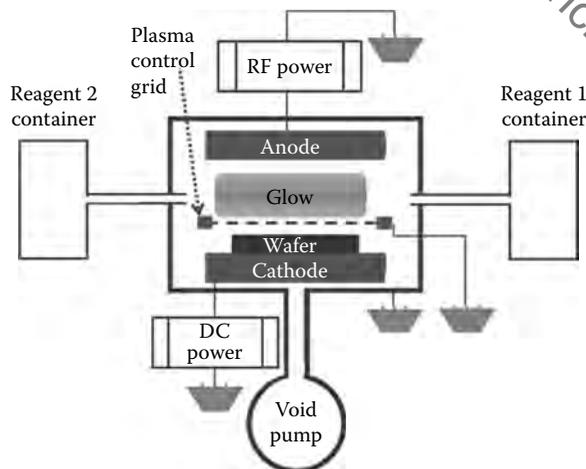


FIGURE 4.79 PECVD reactor scheme where the plasma is created via capacity coupled RF power and assisted by a DC bias and two gas reagents are injected at controller pressures in the CVD chamber.

- *Remote plasma-enhanced CVD (RPECVD)* uses reactors similar to that used in RIE where the CVD chamber is separated by the plasma chamber and an ion beam is extracted from the plasma chamber to be used in the CVD process besides the chemical precursor needed for the film growth, if it does not coincide with the plasma material itself.

An example of RPECVD reactor scheme is reported in Figure 4.80. Here, the plasma is created via capacity inductively coupled RF power and assisted by a DC bias in a separated plasma chamber. An ion beam is derived by the plasma glow and focused on the wafer in the deposition chamber where two gas reagents are injected at controller pressures in the chamber.

In similar reactors, if the plasma ions are of the same type of material to be deposited, or in a few cases if impurities have to be included in the film, sputtering (see Section 4.9.3.1) can be performed in parallel to RPECVD by allowing ion bombing of the wafer surface.

4.8.1.1.3 Other Types of CVD Techniques

- *Atomic layer CVD (ALCVD)* deposits successive layers of different substances to produce layered, crystalline films. ALCVD introduces two complementary reagents alternatively into the reaction chamber. Typically, one of the reagents will adsorb onto the substrate surface until it saturates the surface and further growth cannot occur until the second reagent is introduced. Thus, the film thickness is controlled by the number of reagent cycles rather than the deposition time as is the case for conventional CVD processes. ALCVD allows for extremely precise control of film thickness and uniformity.
- *Combustion chemical vapor deposition (CCVD)* is an open-atmosphere, flame-based technique for depositing high-quality thin films. In the CCVD process, a precursor compound, usually a metal-organic compound or a metal salt, is added to the burning gas. The flame is moved closely above the surface to be coated while wafers slowly pass through the deposition zone. The high energy within the flame converts the precursors into highly

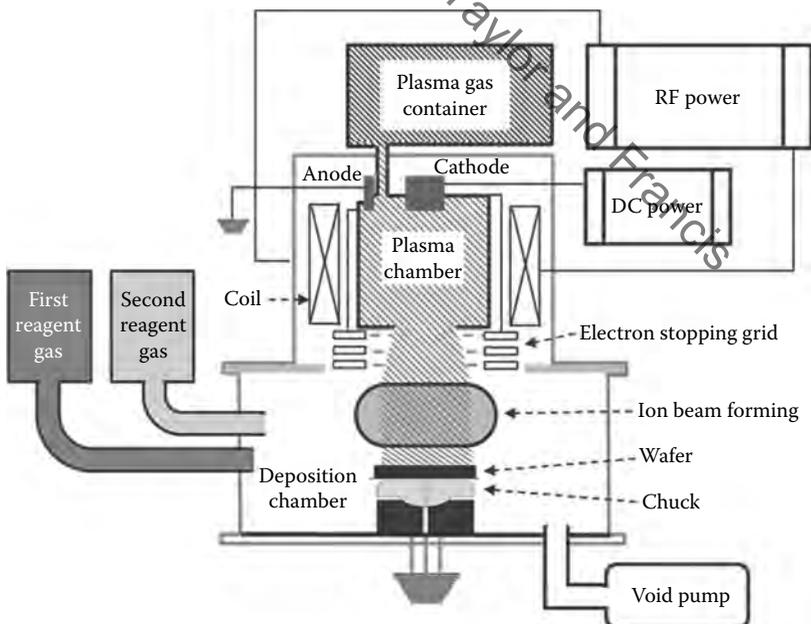


FIGURE 4.80 PECVD reactor scheme where the plasma is created in a separated plasma chamber via inductively coupled RF power and assisted by a DC bias and two gas reagents are injected at controller pressures in the deposition chamber.

reactive intermediates, which readily react with the substrate, forming a firmly adhering deposit. The microstructure and thickness of the deposited layer can be controlled by varying process parameters such as speed of substrate or flame, number of passes, substrate temperature, and distance between flame and substrate. CCVD can produce layers less than 10 nm thick.

- *Hot-wire CVD (HWCVD)*—also known as catalytic CVD (Cat-CVD) or hot filament CVD (HFCVD) uses a hot filament to chemically decompose the source gases.
- *Metalorganic chemical vapor deposition (MOCVD)* is a chemical vapor deposition method of epitaxial growth of materials, especially compound semiconductors, from the surface reaction of organic compounds or metal organics and hydrides containing the required chemical elements. For example, indium phosphide could be grown in a reactor on a substrate by introducing trimethylindium ((CH₃)₃In) and phosphine (PH₃). Formation of the epitaxial layer occurs by final pyrolysis of the constituent chemicals at the substrate surface. In contrast to molecular beam epitaxy (MBE) the growth of crystals by MOCVD is by chemical reaction and not physical deposition. This takes place not in a vacuum, but from the gas phase at moderate pressures (2–100 kPa). As such, this technique is preferred for the formation of devices incorporating thermodynamically metastable alloys, and it has become a major process in the manufacture of optoelectronic active devices like lasers and semiconductor amplifiers.
- *Hybrid physical–chemical vapor deposition (HPCVD)* is a thin-film deposition technique that combines physical vapor deposition (PVD, see Section 4.9.2) with CVD. For example, to grow magnesium diboride (MgB₂) thin film, HPCVD process uses diborane (B₂H₆) as gas reagent, but unlike conventional CVD which only uses gaseous sources, the Mg source is constituted by heated bulk magnesium pellets. Since the process involves chemical decomposition of precursor gas and physical evaporation of metal bulk, it is named as hybrid physical–chemical vapor deposition.
- *Rapid thermal CVD (RTCVD)* is a CVD process that uses heating lamps or other methods to rapidly heat the wafer substrate. Heating only the substrate rather than the gas or chamber walls helps reduce unwanted gas phase reactions that can lead to particle formation.

A synopsis of the characteristics of different CVD processes is reported in Table 4.19.

4.8.1.2 CVD Deposition of Different Materials

A great number of different materials are deposited via CVD in planar technologies. Each deposition receipt exploits a different chemical reaction or reaction chain to achieve the film of the wanted material with the wanted characteristics. In this section, a list of basic chemical receipts regarding common materials deposited in different planar technology fabrications is reported.

4.8.1.2.1 Polysilicon

Polycrystalline silicon is deposited from silane (SiH₄), using the following reaction:



This reaction is usually performed in LPCVD systems, with either pure silane feedstock, or a solution of silane with 70–80% nitrogen. Temperatures between 600°C and 650°C and pressures between 25 and 150 Pa yield a growth rate between 10 and 20 nm/min. An alternative process uses a hydrogen-based solution. The hydrogen reduces the growth rate, but the temperature is raised to 850°C or even 1050°C to compensate this effect. Doped polysilicon may be grown directly using CVD, if gases such as phosphine, arsine, or diborane are added to the CVD chamber. Diborane increases the growth rate, but arsine and phosphine decrease it.

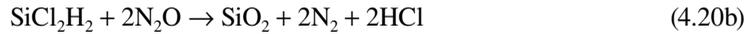
TABLE 4.19
Synopsis of the Characteristics of Different CVD Processes

Process	Advantages	Disadvantages	Applications	Pressure	Temperature	Deposition Rate (SiO ₂)	Epitaxy Rate
APCVD	Simple, high rate, low temperature	Poor step coverage, particle contamination	Thick oxides	10–100 kPa	250–450°C	700 Å/min	
LPCVD	Purity, uniformity, conformable step coverage, large batch capacity	High temperature, low rate	Oxides, silicon nitride, polysilicon, W, WSi ₃	100 Pa	550–650°C	150 Å/min (from TEOS)	48 nm/min (Si on Si [105])
ULPCVD							
LP-MOCVD	Very good epitaxy on large areas	High toxic and very expensive materials	Epitaxy and crystal films (e.g., GaAs, InAsP, InGaAsP, InGaAs quantum wells)	1.3 Pa	550–650°C	—	137 nm/min (Si on Si [105])
PECVD	Low temperature, high rate, good adhesion, good step coverage	Chemical and particle contamination, plasma damage	Insulators over metal, passivation films	100 Pa	800–1200°C	—	24–36 nm/min (InGaAsP on InP [106])
				266–650 Pa	200–600°C	3000 Å/min [107]	—

Source: Data from Madou M. J., *Fundamentals of Microfabrication and Nanotechnology*, Third Edition, Three-Volume Set, CRC Press; 3rd edition (August 1, 2011), ISBN-13: 978-0849331800; With kind permission from Springer Science+Business Media: *Technology of Integrated Circuits*, Reprint of 2000 edition, December 15, 2010, Widman D., Mader H., Friedrich F.

4.8.1.2.2 Silicon Dioxide

Silicon dioxide (SiO_2) may be deposited by several different processes. Common source gases include silane and oxygen, dichlorosilane (SiCl_2H_2) and nitrous oxide (N_2O), ortetraethylorthosilicate (TEOS; $\text{Si}(\text{OC}_2\text{H}_5)_4$). The reactions are as follows:



The choice of source gas depends on the thermal stability of the substrate; for instance, aluminum is sensitive to high temperature. Silane deposits between 300°C and 500°C , dichlorosilane at around 900°C , and TEOS between 650°C and 750°C , resulting in a layer of *low-temperature oxide* (LTO). However, silane produces a lower-quality oxide than the other methods (lower dielectric strength, for instance), and it deposits nonconformal.

Any of these reactions may be used in LPCVD, but the silane reaction is also done in APCVD. CVD oxide has lower quality than thermal oxide, but thermal oxidation can only be used in the earliest stages of IC manufacturing, not to ruin other process results with high temperature. Doped oxide may also be directly grown with LPCVD, for example, to tune its refraction index in the fabrication of optical waveguides.

Another use of doped dioxide is indirect doping of lower layers during further process steps that occur at high temperature. In this case, the impurities may diffuse from the oxide into adjacent layers and dope them. Oxides containing 5–15% impurities by mass are often used for this purpose. In addition, silicon dioxide alloyed with phosphorus pentoxide (“P-glass”) can be used to smooth out uneven surfaces. P-glass softens and reflows at temperatures above 1000°C . This process requires a phosphorus concentration of at least 6%, but concentrations above 8% can corrode aluminum. Phosphorus is deposited from phosphine gas and oxygen.

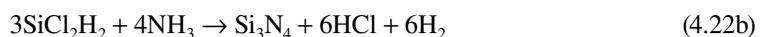


Besides these intentional impurities, CVD oxide may contain by-products of the deposition process. TEOS produces a relatively pure oxide, whereas silane introduces hydrogen impurities, and dichlorosilane introduces chlorine.

Lower temperature deposition of silicon dioxide and doped glasses from TEOS using ozone rather than oxygen has also been exploited (350 – 500°C). Ozone glasses have excellent conformality but tend to be hygroscopic: they absorb water from the air due to the incorporation of silanol (Si-OH) in the glass.

4.8.1.2.3 Silicon Nitride

Silicon nitride is often used as an insulator and chemical barrier in manufacturing integrated circuits. The following two reactions deposit nitride from the gas phase:



Silicon nitride deposited by LPCVD contains up to 8% hydrogen. It also experiences strong tensile stress, which may crack films thicker than 200 nm. However, it has higher resistivity and dielectric strength than most insulators commonly available in micro-fabrication (about $10^{16} \Omega \text{ cm}$ and 10 MV/cm, respectively).

Another two reactions may be used in plasma to deposit SiNH:



These films have much less tensile stress, but worse electrical properties (resistivity from 10^6 to $10^{15} \Omega \text{ cm}$, and dielectric strength from 1 to 5 MV/cm).

4.8.1.2.4 Metals

Some metals (notably aluminum and copper) are seldom or never deposited by CVD. However, CVD processes for molybdenum, tantalum, titanium, nickel, and tungsten are widely used. These metals can form useful silicides when deposited onto silicon. Mo, Ta, and Ti are deposited by LPCVD, from their pentachlorides. Nickel, molybdenum, and tungsten can be deposited at low temperatures from their carbonyl precursors. In general, for an arbitrary metal M , the reaction is as follows:



4.8.2 PHYSICAL VAPOR DEPOSITION

The family of physical vapor deposition (PVD) techniques collects all the thin film deposition processes based on condensation of vapor particles on the wafer surface [108,109]. Evaporation is the most common method of physical thin film deposition. The source material is evaporated in a vacuum; vacuum allows vapor particles to travel directly to the target substrate, where they condense back to a solid state.

At a typical pressure of 10^{-4} Pa, a 0.4 nm particle has a mean free path of 60 m [92]. Poor vacuum can produce hot objects in the evaporation chamber, such as heating filaments, thereby generating unwanted vapors that limit the quality of the vacuum. Moreover, evaporated atoms that collide with foreign particles may react with them; for instance, if aluminum is deposited in the presence of oxygen, it will form aluminum oxide. They also reduce the amount of vapor that reaches the substrate, which makes the thickness difficult to control.

Evaporated materials do not deposit uniformly if the substrate has a rough surface as integrated microfluidics circuits often do. Because the evaporated material attacks the substrate mostly from a single direction, protruding features block the evaporated material from some areas. This phenomenon is called shadowing and it is qualitatively illustrated in Figure 4.81.

The deposition rate in thermal PVD process is generally good especially for metals; for example, in the case of Al films, a typical evaporation rate is 0.1–0.5 $\mu\text{m}/\text{min}$.

Different evaporation methods exist, using different energy sources to obtain the target evaporation.

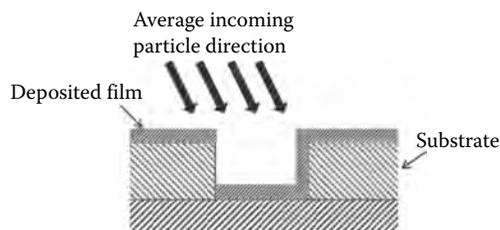


FIGURE 4.81 Schematic illustration of shadow effect in PVD.

4.8.2.1 Thermal PVD

In thermal PVD, a metal wire is fed onto a heated ceramic evaporator known as “boats” due to their shape. A pool of melted metal forms in the boat cavity and evaporates into a cloud above the source. Alternatively, the source material is placed in a crucible, which is radiatively heated by an electric filament, or the source material may be hung from the filament itself in the so-called filament evaporation.

Resistors designed to work as heaters in thermal PVD systems are made by materials with a very high melting temperature (T_m) so that high temperatures can be reached on the resistor surface without melting it. Typical resistor materials are tungsten (W, $T_m = 3380^\circ\text{C}$), tantalum (Ta, $T_m = 2980^\circ\text{C}$), and molybdenum (Mo, $T_m = 2630^\circ\text{C}$).

Advantages of thermal evaporation is the simplicity of the related equipment and the low cost of the process, besides the possibility of shaping the process characteristics like the uniformity of the film and the coated area by simply changing the resistor and the target material forms and the vapor temperature.

A first problem of this technique is that evaporation of a small quantity of the resistor material at such a high temperature is not avoidable, so generating reactions between the film and the resistor materials in vapor phase and inducing pollution of the deposited film. This problem can be faced by coating the resistor with a ceramic layer that do not evaporate even at extreme temperatures, but this architecture exploits less effectively the resistor temperature due to the isolation nature of ceramic materials. Moreover, while metals are easily evaporated through thermal PVD, dielectric films are difficult to obtain since the evaporating temperature of isolators is too high. For example, the silicon oxide melting temperature is as high as 1703°C ; in order to obtain such a high temperature on the target the resistor should be pushed too near to its melting temperature causing excessive resistor material evaporation. It is also difficult to deposit compound targets, since they may be decomposed by high temperature. Finally, the quality of film adhesion is poor with respect to other deposition methods.

Owing to these limitations, thermal PVD is used only in low-performance processes, where extremely good film qualities are not required and the cost of the process is a key issue.

A notable exception is the so-called thermal oxide layer that is generated by thermal oxidation on the surface of silicon wafers. In this process, evaporation of low-pressure oxygen provokes oxidation of the surface of the Si wafer generating silica. This technique is not suitable to deposit thick SiO_2 layers, due to the passivation role of the layer itself that above a certain thickness inhibits the further penetration of oxygen and the oxidation of the silicon substrate. Generally, layers up to 500 nm are grown in this way. However, under the point of view of uniformity, dielectric and optical properties and absence of embedded stress, thermal silica is superior to silica layers deposited with other techniques.

4.8.2.2 Electron-Beam PVD

In electron-beam PVD, high-speed electrons hit the target material, to create a high temperature on its surface and cause its evaporation. The electron beam is generally produced by an electron gun, which uses the thermionic emission of electrons produced by a tantalum cathode. Emitted electrons are accelerated toward an anode with very high speed. For example, at 10^4 eV energy electrons average speed is about 16.6 km/s; thus creating on the target surface a temperature in the range of $5000\text{--}6000^\circ\text{C}$.

The crucible which puts the target acts as the anode. A magnetic field is also applied to bend the electron trajectory. Avoiding the gun blocks the evaporation path. Via the control of the magnetic lens, similar to those used in an electron beam lithography equipment, the electron beam can be focalized and change its impact point on the target surface to obtain a very localized heating on the material to evaporate.

It also allows a precise control of the evaporation rate, from low-to-very high values. Most important of all, it could deposit the materials with high melting point such as W, Ta, C, dielectric

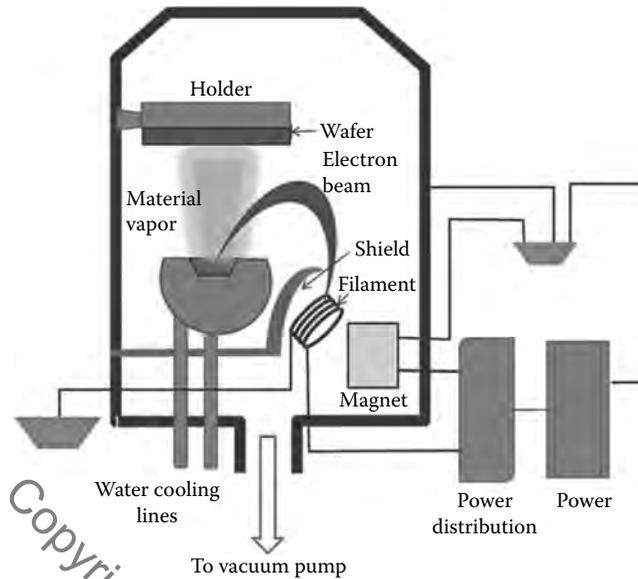


FIGURE 4.82 Schematic architecture of an e-beam evaporation PVD equipment.

oxide. Effective water cooling to crucible can avoid contamination problems inherent to heating and degasification of undesired material coming from the coater. A schematic representation of an electron beam PVD reactor is provided in Figure 4.82.

E-beam PVD produces films much purer with respect to thermal PVD, also because target contamination by the crucible particle evaporation is eliminated completely by water cooling of the crucible. Moreover, most dielectric oxide layers could be deposited by this method.

Multilayer depositions can be realized by disposing different crucibles into the deposition chamber and directing the electron beam toward one or the other depending on what layer is to be deposited, the fact that the electron beam has a small spot size, thus heating only a small part of the crucible surface avoids contamination completely between the targets devoted to different layers.

The film thickness can be controlled by both changing the electron beam spot size and its intensity, thus achieving a very good control. The film uniformity is also very good, since the film thickness can be controlled during deposition, for example, with interferometry methods and the electron beam illumination time over different wafer surfaces can be regulated on the run to maintain the better uniformity.

The only relevant source of contamination due to the deposition itself is the so-called x-ray contamination due to spurious x-rays generated by the electron source that can damage the deposited film during deposition.

4.8.2.3 Other PVD Techniques

In flash evaporation PVD, a fine wire of source material is fed continuously onto a hot ceramic bar, and evaporates on contact. This is accomplished by passing a large current through a resistive wire or foil containing the material to be deposited. The heating element is often referred to as an evaporation source. Wire type evaporation sources are made from tungsten tantalum, molybdenum, or ceramic-type materials capable of withstanding high temperatures. They can be formed into filaments, baskets, heaters, or point sources. Boat-type evaporation sources are more frequently made by tungsten.

Cathode arc deposition is a physical deposition technique based on the phenomenon of arc evaporation [110]. The arc evaporation process begins with the striking of a high-current, low-voltage arc on the surface of a cathode (the target) that gives rise to a small, usually a few micrometers wide, highly

energetic emitting area known as a cathode spot. The localized temperature at the cathode spot is extremely high ($\sim 15,000^\circ\text{C}$), which results in a high velocity (10 km/s) jet of vaporized cathode material, leaving a crater on the cathode surface. The cathode spot is only active for a short period of time, then it self-extinguishes and re-ignites in a new area. This behavior causes the apparent motion of the arc.

As the arc is basically a current-carrying conductor it can be influenced by the application of an electromagnetic field, which in practice is used to rapidly move the arc over the entire surface of the target, so that the total surface is eroded over time. The arc has an extremely high power density resulting in a high level of ionization (30–100%), multiple charged ions, neutral particles, clusters, and macro-particles. If a reactive gas is introduced during the evaporation process, dissociation, ionization, and excitation can occur during interaction with the ion flux and a compound film will be deposited.

4.8.3 OTHER PHYSICAL DEPOSITION TECHNIQUES

There is a group of physical film deposition techniques that are not based on evaporation of the target, but have their specific principles. Here, we will describe plasma deposition, generally called sputtering, and MBE.

4.8.3.1 Physical Plasma Deposition (Sputtering)

As we have seen in Section 8.2 the generic term of sputtering indicates a process in which a target is bombed with high-energy ion coming from a plasma glow to extract particles from it. In sputtering etching the target is the film to be etched, in sputtering deposition, also called plasma-assisted PVD, the target is formed by the material to be deposited. The particle extracted from the target in this case travels through a low-pressure sputtering room up to the surface of the wafer and deposit on it forming the film to be deposited.

An important advantage of sputter deposition is that even materials with very high melting points are easily sputtered while evaporation of these materials in other PVD techniques is problematic or impossible. Sputtering can be used also to deposit composite films and the sputter deposited films have a composition close to that of the source material. The difference is due to different elements spreading differently because of their different mass. Due to the deterministic nature of this effect, this difference can be determined accurately and the target composition can be biased suitably.

Examples of materials that are frequently deposited by sputtering are [111–115]

- Metals (Al, Cu, Zn, Au, Ni, Cr, W, Mo, Ti)
- Alloys (Ag–Cu, Pb–Sn, Al–Zn, Ni–Cr)
- Nonmetals (graphite, MoS_2 , WS_2)
- Refractory oxides (Al_2O_3 , Cr_2O_3 , $\text{Al}_2\text{O}_3\text{--Cr}_2\text{O}_3$, Y_2O_3 , SiO_2 , ZrO_2)
- Refractory carbides (TiC, ZrC, HfC, NbC)
- Refractory nitrides (TiN, Ti_2N , ZrN, TiN–ZrN, HfN, TiN–AlN–ZrN)
- Refractory borides (TiB_2 , ZnB_2 , HfB_2 , CrB_2 , MoB_2)
- Refractory silicides (MoSi_2 , WSi_2 , Cr_3Si_2)

Sputtered films typically have a better adhesion on the substrate than evaporated films. Sputtering sources contain no hot parts since to avoid heating they are typically water cooled and are compatible with reactive gases such as oxygen.

Sputtering sources are usually inductively coupled RF or microwave sources that utilize strong electric and magnetic fields to trap electrons close to the surface of the target. Because of the strong magnetic field due to inductive coupling, the electrons follow helical paths around the magnetic field lines undergoing more ionizing collisions with gaseous neutrals near the target surface than would otherwise occur [116].

The gas used to generate the plasma glow in sputtering systems is generally an inert gas, more frequently argon. Sputtering yield, or the number of atoms ejected per incident ion, is an important

factor in sputter deposition processes, since it affects the sputter deposition rate. Sputtering yield primarily depends on three major factors:

- Target material
- Mass of the bombarding particles
- Energy of bombarding particles

In the energy range where sputtering occurs (10–5000 eV), the sputtering yield increases with particle mass and energy. A range of typical sputtering yields for different materials when Ar⁺ ions at 500 eV are used is reported in Table 4.20.

Different types of sputtering processes exist, depending on the sputtering conditions and the reactor structure. Two basic sputtering deposition processes are described here.

Ion-beam sputtering (IBS) is a method in which the sputtered target is bombed by an ion beam extracted from a plasma glow in an external reactor chamber, called sputtering chamber. Such reactors are also called Kaufman reactors. The source ions are generated in a plasma chamber, generally, via RF or microwave generation. Plasma electrons are confined in the plasma chamber while ions are extracted and accelerated by the electric field emanating from a grid toward a target.

As the ions leave the source they are neutralized by electrons from a second beam external to the plasma glow. Thus, the flux that hit the target is not composed of ions, but by neutral atoms so that either insulating and conducting sputtering targets can be used and no superficial charge accumulation on the target happens.

A pressure gradient between the ion source and the sample chamber is generated by placing the gas inlet at the source and shooting through a tube into the sample chamber. This saves gas and reduces contamination. The principal drawback of IBS is the large amount of maintenance required to keep the ion source operating [117].

Typical deposition rates that can be realized with ion beam sputtering deposition are reported in Table 4.21. A scheme of an ion beam sputtering reactor is reported in Figure 4.83; here, the secondary electron source used to neutralize the ion beam is evidenced, a target selection support is mounted to be able to alternate different targets to deposit composite films and a wafers rotating drum is present to process several wafers in the same batch.

In *ion-assisted deposition* (IAD), the substrate is exposed to a secondary ion beam operating at a lower power than the sputter gun. Usually, a Kaufman source like that used in IBS supplies the secondary beam. IAD can be used, as an example, to deposit carbon in diamond-like form on a substrate. Any carbon atoms landing on the substrate which fails to bond properly in the diamond crystal lattice will be knocked off by the secondary beam.

TABLE 4.20
Typical Sputtering Yields for Different Materials
When Ar⁺ Ions at 500 eV Are Used

Target	Sputter Yields
Aluminum	1.05
Chrome	1.18
Gold	2.4
Nickel	1.33
Platinum	1.4
Titanium	0.51

Note: The sputtering yield is defined as the average number of extracted target particles per colliding ion.

TABLE 4.21
Sputtering Deposition Rate for Various Materials

Material	Deposition Rate (nm/min)	
	1 kW	3.5 kW
Al	18	63
Cu	34	119
Au	36	126
Ni	22	77
Pt	23	80.5
Ag	52	182
Ti	8	28
Ta	9	31.5

Note: An ion beam sputtering is considered and the rate is reported for different powers of RF source.

Besides this basic implementation, several variations exist to improve the target utilization [118], the reactor architecture [119], or to study the structure of sputtering generated films [119,120] and sputtering is still today one of the most used deposition techniques in micro-fabrication industry.

4.8.3.2 Molecular Beam Epitaxy

MBE is the better method today available to grow high crystal quality epitaxial layers. As such it is not of great importance in microfluidics circuits but it is widely used to deposit epitaxial films of alloys such as InGaAsP or GaAsP that are used in active optical devices like lasers and amplifiers or in optical detectors [121–123].

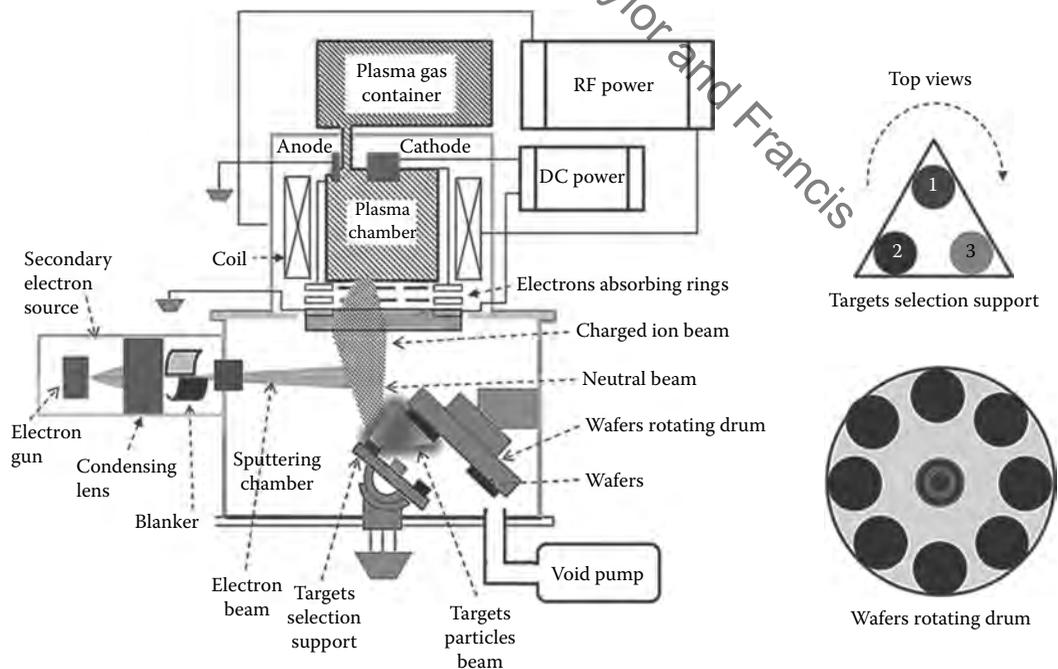


FIGURE 4.83 Scheme of an ion beam sputtering reactor.

MBE takes place in high vacuum or ultra-high vacuum (10^{-8} Pa). The most important aspect of MBE is the slow deposition rate (typically <20 nm/min), which allows the films to grow epitaxially. The slow deposition rates require proportionally better vacuum to achieve the same impurity levels as other deposition techniques.

In solid-source MBE, elements such as gallium and arsenic, in ultra-pure form, are heated in separate effusion cells until they begin to slowly sublimate. The gaseous elements then condense on the wafer, where they may react with each other. In the example of gallium and arsenic, single-crystal gallium arsenide is formed. The term beam means that evaporated atoms do not interact with each other or vacuum chamber gases until they reach the wafer, due to the long mean free paths of the atoms caused by the extremely low concentration.

During operation, reflection high-energy electron diffraction (RHEED) is often used for monitoring the growth of the crystal layers. A computer controls shutters in front of each furnace, allowing precise control of the thickness of each layer, down to a single layer of atoms.

Intricate structures of layers of different materials may be fabricated this way. Such control has allowed the development of structures where the electrons can be confined in space, giving quantum wells or even quantum dots.

In systems where the substrate needs to be cooled, the ultra-high vacuum environment within the growth chamber is maintained by a system of cryopumps, and cryopanel, chilled using liquid nitrogen or cold nitrogen gas to a temperature close to -196°C . Cryogenic temperatures act as a sink for impurities in the vacuum, so vacuum levels need to be several orders of magnitude better to deposit films under these conditions. In other MBE implementations, the wafers on which the crystals are grown may be mounted on a rotating platter which can be heated to several hundred degrees Celsius during operation.

MBE is also used for the deposition of some types of organic semiconductors. In this case, molecules, rather than atoms, are evaporated and deposited onto the wafer. Other variations include gas-source MBE, which resembles chemical vapor deposition.

Recently, MBE has been used to deposit oxide materials for advanced electronic, magnetic, and optical applications. For these purposes, MBE systems have to be modified to incorporate oxygen sources.

A scheme of a traditional MBE reactor is shown in Figure 4.84.

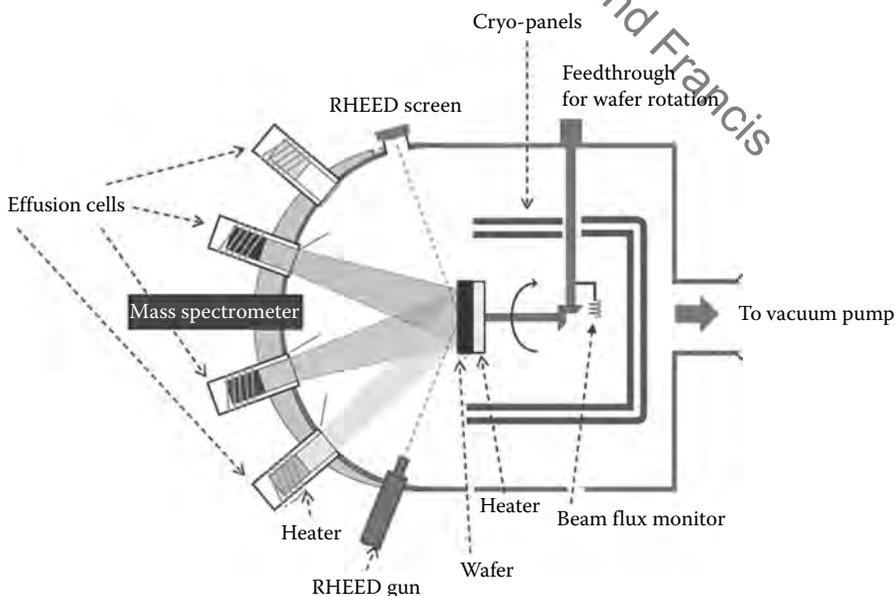


FIGURE 4.84 Scheme of a MBE growth chamber.

4.9 WAFER BONDING

In microfluidics-based labs on chip, one of the key elements in the circuit fabrication is the realization of closed chambers and ducts forming the microfluidic system. In the above sections we have seen that the standard procedure of planar technologies is not able to excavate closed areas below the surface of the wafer. On the contrary, open trenches or chambers can be created, that can be transformed into the elements that are needed in a microfluidics circuit by adding a top cover. The top cover can be realized by the technique of wafer bonding, consisting of superimposing two wafers one on top of the other, putting in contact and fixing the upper wafer surfaces.

This process was not created for microfluidics, but for MEMs and other applications requiring three-dimensional structures or closed rooms. Its application in microfluidics circuit is in any case almost obliged if standard planar techniques are to be adopted, so it became a key process in microfluidics-based labs on chip.

Several different techniques are adopted to realize wafer bonding, both in case of adhesion of crystal surfaces of the same or different materials and in case of adhesion of amorphous surfaces. In any case, the two surfaces to be put in contact has to be flat within a very small tolerance, due to the fact that lack of flatness, even by a very small amount, can completely ruin the bonding or create points of weakness causing later circuit failures.

This is the reason because planarization, that is, the process used to flatten the surface of a wafer after a fabrication cycle is a structural part of the wafer-bonding process.

Planarization was not originally introduced for microfluidics too, and it is still mainly used at the end of lithography processes to flatten the wafer surface, remove the uneven features that could be present due to resist stripping, and prepare the wafer for successive fabrication steps. When the successive step is a wafer bonding, the planarization requirements are quite tight and it is the reason why we have inserted the planarization process under the general cap of the wafer bonding technique.

4.9.1 PLANARIZATION

Planarization is not a process adopted only in planar fabrication, since several fabrication techniques need to flatten surfaces even at the nanometer level (e.g. optical lenses fabrication). A standard process used to obtain surface flattening is lapping, that if applied in successive steps can achieve a good degree of planarity. The main limitation of lapping process is the lack of selectivity that renders difficult to control the process if the underlying film has not to be consumed. Moreover lapping, due to the intrinsic need of lapping powder, imply a strong contamination risk that is difficult to control in a clean room where planar fabrication is performed.

This is the reason why, when a high degree of flattening and absence of contamination are needed, planarization is performed generally with the chemical–mechanical planarization (CMP) process [124,125]. Typical CMP tools, such as the ones represented in Figure 4.85, consist of a rotating and extremely flat platen which is covered by a pad. The wafer that is being polished is mounted upside-down in a carrier/spindle on a backing film. The retaining ring keeps the wafer in the correct horizontal position. During the process of loading and unloading the wafer onto the tool, the wafer is held by vacuum by the carrier to prevent unwanted particles from building up on the wafer surface.

A slurry introduction mechanism deposits the slurry on the pad. Both the platen and the carrier are then rotated and the carrier is kept oscillating as well. A downward pressure/down force is applied to the carrier, pushing it against the pad. The required amount of down force depends on the contact area which, in turn, is dependent on the structures of both the wafer and the pad. Typically, the pads have a roughness of 50 μm ; contact is made by asperities, which typically are the high points on the wafer, and, as a result, the contact area is only a fraction of the wafer area.

The base mechanism of CMP is shown in Figure 4.86. The slurry has a chemical action that softens the surface of the wafer to be flattened. The inhomogeneous structure of the polishing pad

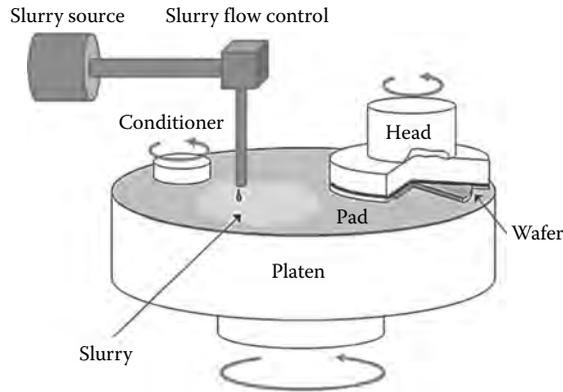


FIGURE 4.85 Scheme of a CMP (chemical mechanical planarization) equipment.

and the presence of abrasive particles tend to consume the surface of the wafer efficiently since it is so softened. Since adhesion of the wafer to the polishing pad happens first and with a higher pressure when an asperity is present, asperities are more efficiently consumed and the film is flattened. Moreover, the softening chemical reaction is selective, thus if a layer of a different material is found, it is not softened and the flattening effect is negligible.

Typically, SiO_2 films are flattened by softening them by hydration through the following reaction:



This reaction, breaks the Si–O bonds and softens the film. Besides allowing effective ablation, it also permits, in case it is needed, further chemical etching using a KOH solution. Thus, frequently, a very low-concentration KOH solution is adopted by allowing hydration and successive mixed ablation and chemical etching. In this case, however, requirements on the control of PMC time and slurry flow are much more stringent if the final film depth is to be well controlled.

In CMP, the mechanical properties of the wafer itself are important parameters. If the wafer has a slightly bowed structure, the pressure will be greater on the edges than it would on the center, which causes nonuniform polishing. To compensate for the wafer bow, pressure can be applied to the wafer's backside which, in turn, will equalize the center-edge differences.

The pads used in the CMP tool should be rigid in order to uniformly polish the wafer surface. However, these rigid pads must be kept in alignment with the wafer at all times. Therefore, real pads are often stacks of soft and hard materials that conform to wafer topography to some extent. Generally, these pads are made from porous polymeric materials with a pore size between 30 and

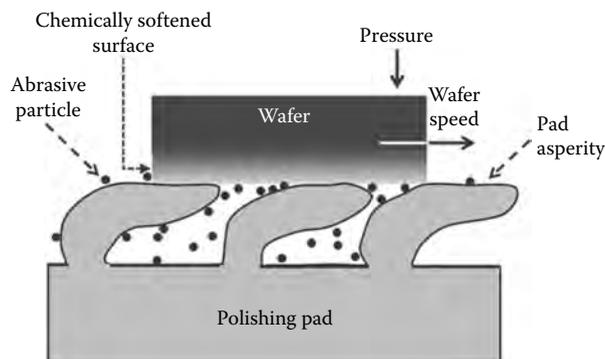


FIGURE 4.86 Schematic illustration of the principle of the CMP process.

TABLE 4.22
Main Process Parameters and Performances of CMP
Planarization of a Silicon Oxide Layer

Pressure on wafer	14–48	kPa
Platen rotation	20–60	rpm
Carrier rotation	20–80	rpm
Slurry flow rate	0.1–0.2	L/min
Etching rate	0.28	$\mu\text{m}/\text{min}$

50 μm and due to their characteristics they are consumed in the process so that regular reconditioning is needed.

Typical parameters of a CMP process on an oxide layer are reported in Table 4.22.

The CMP planarization process is not without problems. In case of oxide, planarization problems arise from a large set of phenomena, for example [126,127],

- Particle contamination either of the surface or of the film bulk, when the particles are absorbed in the film during the process.
- Stress accumulation on the film surface due to ablation.
- Imperfect slurry washing, leaving residual slurry on the film surface.
- Micro-scratches and deep defects on the surface due to ablation.
- Imperfect film edges that are present especially when trenches are fabricated on the chip. This is a very frequent situation in microfluidic chips, where deep trenches are present to prepare the ducts and chambers.

4.9.2 ADHESIVE WAFER BONDING

Wafer bonding process has a great number of applications and depending on the material that has to be bonded together different processes are necessary. In MEMs and electronic ICs applications it is frequent to require bonding between crystals, either semiconductors like silicon or GaAs or metals [6]. When crystals are involved, besides the great difference in bonding similar or dissimilar crystals, the ordered property of the surfaces to be bonded can be exploited. Moreover, if the wafers have a metal film on the surface, metal bond has also peculiar characteristics.

The most common case in microfluidics labs on chip is, however, the case in which the surfaces to be bonded are coated either with oxides or with polymers. We will deal with polymer technologies in the next chapter; thus, here we will consider wafer bonding where oxides coating is present at least on one of the wafer surfaces, considering in more detail coating with SiO_2 film. In this case, essentially two methods are possible: adhesive wafer bonding and direct wafer bonding [128,129].

4.9.2.1 Adhesive Wafer Bonding Principle

The basic principle that all bonding techniques have in common is that two materials adhere to each other if they are brought in sufficiently close contact. The cohesion of atoms and molecules within a solid material as well as the adhesion of atoms and molecules in between different solid materials is ensured by four basic bond types, which are covalent bonds, van der Waals bonds, metallic bonds, and ionic bonds.

Covalent bonds and van der Waals bonds are the dominant bonding mechanisms for oxide wafers that cannot exploit either metallic or ionic characteristics. To accomplish covalent and van der Waals bonds, the atoms of two opposing surfaces must be less than 0.3–0.5 nm apart, as it can be easily deduced from the characteristics of van der Waals and covalent bonds detailed in Chapter 2.

Macroscopically flat surfaces, such as the surfaces of polished SiO_2 surfaces, have a root-mean-square roughness of 0.3–0.8 nm. Nevertheless, the profile depth peak to valley of these surfaces is

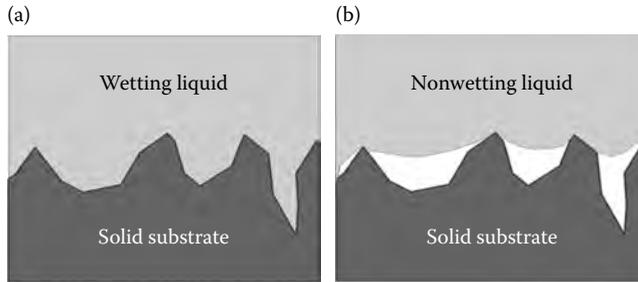


FIGURE 4.87 Microscopic view of a solid surface covered by a wetting (a) and a nonwetting (b) liquid.

several nanometers, which typically prevents bonding over larger surface areas. In order to bring two material surfaces in sufficiently close contact to achieve bonding at least one material surface must deform to fit the other. This deformation may be accomplished by plastic or elastic deformation, by diffusion of a solid material or by wetting of a surface with a liquid material. Practically, all wafer-bonding techniques use one of these mechanisms to establish bonding between different surfaces.

In adhesive wafer bonding, a polymer adhesive is placed in between the wafer pair to be bonded: the polymer adhesive deforms to fit the surfaces to be bonded so as to create a very close contact among the polymer molecules and the two surfaces. Polymer adhesives are typically in a liquid or semiliquid phase during part of the bonding process and wet the surfaces to be bonded by flowing into the troughs of the surface profile. The liquid polymer adhesive must then harden into a material that is capable of bearing the forces involved to hold the surfaces together.

Wetting of the surfaces by the liquid or semiliquid polymer adhesive (see Section 7.3.5) is critical in adhesive bonding. As a matter of fact, if wetting does not happen [128], the polymer remains suspended on the surface and leave spaces preventing a good bonding to take place (see Figure 4.87). For wetting to occur, the solid surface must have a greater surface energy than the liquid (see Section 7.3.5). The surface energy is a result of unbalanced cohesive forces at the material surface. A higher cohesive force between the atoms or molecules of a material correlates to a higher surface energy. Examples of surface energies for several materials when the surface is exposed to air are reported in Table 4.23.

The degree of wetting of a surface with a liquid adhesive can be reduced by surface contaminants, such as the weakly adsorbed organic molecules or by condensed moisture. The degree of wetting can also be influenced by the microscopic surface profile and by dust particles on the surface.

The more complete the polymer adhesive flows into and fills the troughs of a surface profile the better is the resulting bond quality and the long-term stability of the bond. Polymer adhesives that have low viscosity and low shrinkage during hardening generally achieve better filling of the troughs of a surface profile which decreases the amount of unfilled space at the bond interface.

4.9.2.2 Bond Quality of Adhesive Wafer Bonding

An important characteristic of all wafer bonding techniques is the bond force. As a matter of fact, even a local lack of bond brings to the circuit failure. In general, we can define the bond force as the maximum mechanical force needed to divide the bonded wafers. Assuming the bond homogeneous in the horizontal plane, at least on sufficiently large areas, the bond force can be defined both considering a shear test force and a pressure test force, even if generally the pressure bond force is deemed more important due to the fact that in normal applications it is the most common stress to which the circuit will be subject during use.

Another possible definition of the bond force is related to the surface energy at the bond. As a matter of fact, bonding takes place due to the energy reduction caused by the bonding of the two surfaces with respect to the surface energy of both of them exposed to air or vacuum. Thus, in order

TABLE 4.23
Examples of Surface Energies for Several Materials
When the Surface Is Exposed to Air

Material	Physical Status	Surface Energy (mJ/m ²)
Tungsten	Solid crystal	5800
Diamond	Solid crystal	4300
Platinum	Solid crystal	2800
Silicon	Solid crystal	1800
Aluminum	Solid crystal	900
Tin	Solid crystal	600
Glass	Solid amorphous	350
Ice	Solid crystal	85
Water	Liquid	80
Glycerol	Liquid	72
Polyurethanes	Liquid	34
Cyanoacrylates	Liquid	28
Polyethylene	Solid amorphous	22
Ethanol	Liquid	15
Methanol	Liquid	15
Teflon	Solid amorphous	12

to restore the situation in which the two wafers are divided, the energy difference has to be provided to the system by a suitable force.

If the wafer separation is performed by a purely normal force, there is a relationship between the bonding force (F_B) and the bonding energy (E_B). Let us imagine to apply a uniform force to the upper wafer for a time τ up to divide the two wafers and bring it to a speed v . This speed for sufficiently low values of v is small and can be easily measured.

If the wafers would not be joined, all the applied force would be transformed in kinetic energy, arriving at a speed v_0 , while in our case $v < v_0$ due to the need of overcoming the bond energy. Thus, the bond energy can be evaluated by the following equation, where ρ is the average density of the upper wafer, r_w its radius, and d the wafer thickness.

$$E_B = \frac{F_B}{\pi r_w^2 d \rho} \tau - \frac{1}{2} \pi r_w^2 d \rho v^2 \approx \frac{P_B}{d \rho} \tau \quad (4.26)$$

where P_B is the bound pressure and the approximation holds as far as the kinetic energy at the second member is negligible with respect to the bound energy, that is a reasonable assumption in this case. The bound energy has to be sufficiently high both to sustain post-processing steps and to assure sufficient operating life to the final product: generally, a value of 1 J/m² is considered as a good target [130].

Naturally, the bond energy is not the only parameter that is important in evaluating the bond quality. The Young module, giving a measure of the compressive resistance of the bonded wafers, is another important parameter. Target values are above 1.5 GPa.

Under a chemical point of view, the bond has to resist to chemical actions that could be required in post-processing steps. Moreover, it has to be sufficiently stable to assure the required operating life of the final product.

The thermal expansion coefficient has also to be controlled, due to the fact that generally the upper and lower wafers do not expand at the same pace and the bond has to resist the stress introduced by this differential expansion.

TABLE 4.24
Adhesive Wafer Bonding Parameters for SiO₂–SiO₂
Bond Using SU8 Polymer

Spinning deposited layer	10–15 μm
Hardening temperature	65–90°C
Temperature hardening time	2 min + 10 min
Hardening UV exposure	100 mJ/cm ²
Glass temperature	>200°C
Young's modulus	2–4 MPa
Thermal expansion coefficient	52 PPM/°C [139]
Chemical stability	Class 3 w/acids
Shear bond strength	18–25 MPa

Source: Data from Yu L. et al., *Journal of Physics: Conference Series*, vol. 34 pp. 776–781, 2006, but when explicitly indicated.

Besides these parameters, that are typical of all the bonding techniques, if the polymer adhesive bond is used, other parameters have to be under consideration related to the specific technique.

As we have seen, the polymer is introduced as a solution, where generally the monomers or very short polymer chains are present. The real polymer is formed during hardening, where the long chains are formed. Thus, the degree of polymerization is an important parameter: higher the degree better is the bond.

Also, the glass transition temperature of the polymer is important, that is, required to be higher than the higher operating temperature of the final product and of all the post-processing steps.

Last, but not least, one of the advantages of the adhesive bonding is to require low temperatures with respect to other techniques that, after a room temperature prebonding, require annealing at more than 1000°C. Thus, the temperature at which the polymer hardening is performed is another important parameter of the process.

Characteristics of typical Si–SiO₂ adhesive wafer bonding processes are reported in Table 4.24, in correspondence to the use of SU8 (see Section 4.5.2) as adhesion polymer [131] and in the case of SiO₂–SiO₂ bonding.

4.9.2.3 Adhesive Wafer Bonding Process and Reactor Structure

A first element to be considered in designing the processing for adhesive wafer bonding of two wafers during the fabrication of a microfluidic circuit is that the bonding polymer has not to invade the trenches that have to constitute the ducts and chambers of the microfluidics circuit.

This is particularly important when activated surfaces are present, since the activation can be ruined by simple interaction with a different material. If the upper layer is flat, that is, no trenches are present, the deposition of the polymer solution can be done on the upper layer by spinning (see Section 4.5.2) and, after hardening, the bond can be done with the wafers in the inverse position (the upper wafer below). In this situation, gravity thwarts the falling of polymer particles on the floor of the trenches.

Frequently, however, this simple solution is not sufficient, either because trenches are present on both the wafers, or because the upper wafer for other reasons cannot be coated with the polymer before hardening (e.g., the polymer solution could contaminate the upper wafer surface).

In these cases, different solutions have to be found. The most used processes to deposit the polymer solution on a wafer with trenches without contaminating the trenches bottom surfaces are contact imprinting [131] and roller imprinting, whose principles are represented in Figure 4.88.

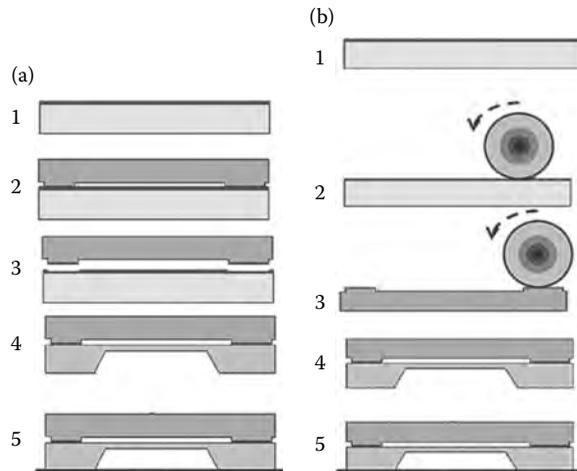


FIGURE 4.88 Operating principle scheme of contact imprinting (a) and roller imprinting (b) in adhesive wafer bonding process. The different steps of the two processes are described in the text.

In contact imprinting a dummy wafer is coated with the contact polymer solution (step 1), that is generally treated to obtain a hydrophilic characteristic so as to improve the adhesion to the final surface. The wafer with trenches is then put in contact under pressure with the dummy wafer so as to transfer the contact polymer to the upper wafer of the final assembly only on the exposed surfaces, without contaminating the trenches (step 2). The upper layer is then pressed onto the lower wafer in the wafer-bonding reactor so as to create contact only in areas where the polymer was present in the lower wafer (step 3). A final polymer hardening at the suitable temperature creates the wafer bonding without contaminating the structures that are present on the two wafers.

Also, in the case of roller imprinting the polymer solution is coated over a dummy wafer and undergoes a suitable treatment (step 1). After that, a roller collects the treated polymer solution (step 2) from the dummy wafer and rolling over the upper wafer of the final assembly deposits it only on the contact surfaces, without contamination of the trenches (step 3). The last steps of the process are similar to that of the imprinting technique.

A complete scheme of the adhesive wafer-bonding technique is reported in Table 4.25, while a scheme of an adhesive wafer-bonding chamber architecture is plotted in Figure 4.89.

TABLE 4.25

Summary of the Steps of a Typical Adhesive Wafer Bonding Process

No.	Process Steps	Observations
1	Cleaning and drying of the wafers	Remove particles and moisture from the wafer surfaces
2	Treating the wafer surfaces with an adhesion promoter	Adhesion promoters enhance the adhesion between the wafer surfaces and the polymer adhesive. In some cases, this step is not needed
3	Applying the polymer solution to one wafer	Suitable application process has to be used. Sometimes, the polymer is applied to both wafers (when trenches are not present)
4	Pre curing of the polymer	Solvents and volatile substances are removed from the polymer coating. Partial polymerization could occur in this phase
5	Placing wafers in bond chamber and joining them	Wafers are joined in a vacuum environment to prevent voids and gases from being trapped at the bond interface
6	Applying pressure to the wafer stack	
7	Hardening the polymer while pressure is applied	Polymer hardening is typically performed by heating the system; sometimes, additional UV hardening is needed (as in the case of SU8)

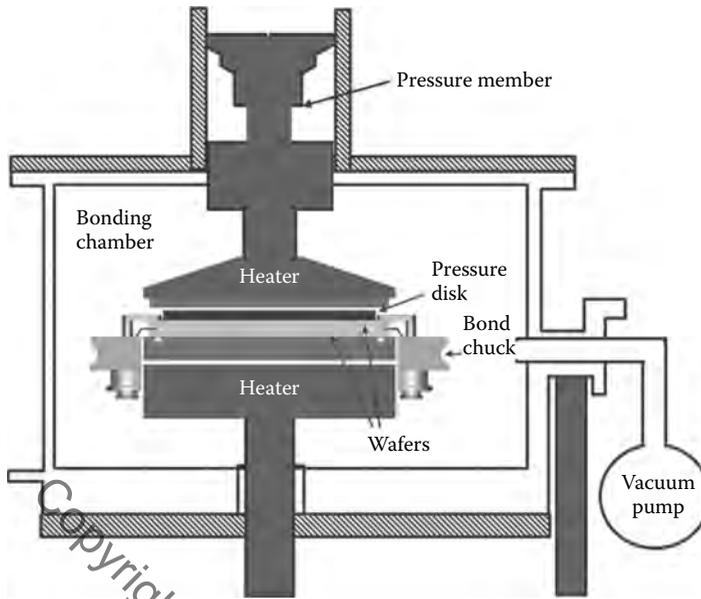


FIGURE 4.89 Bonding chamber scheme.

4.9.3 DIRECT WAFER BONDING

Adhesive wafer bonding is able to produce bonding strength sufficient for almost all the applications and works at relatively low temperature; moreover, due to the good bonding characteristics of adhesive polymers, the required wafer flatness is in the range of normal values attained after CMP at the end of any lithography cycle.

However, adhesive bonding has the disadvantage to require an intermediate bonding layer. The presence of an intermediate polymer is both a potential cause of contamination, both when the polymer is deposited and when, after bonding, the dice are separated. As a matter of fact, polymer particles can create even internally to the chip, at the edges of microfluidics chambers, due to stress and applied strength and contaminate the chip.

Thus, bonding techniques that does not require an intermediate bonding layer are quite desirable. Collectively, these techniques are called direct bonding, due to the fact that no intermediate layer is required between the wafers to be bonded [6]. The price to pay to get rid of the adhesive layer is the need of a much more flattened surface with respect to the adhesive bonding and the presence of a process meant to activate the surfaces to be put in contact in order that they can create bonding.

As far as the activation process is concerned, direct wafer bonding techniques can be divided in thermal techniques, where activation takes place due to high temperature, and ion-enhanced techniques, where activation occurs due to ion bombing.

4.9.3.1 Direct Wafer Bonding Principle

The phenomenon of wafer direct bonding originates from the intermolecular forces of attraction between two contacting surfaces. Therefore, the first step in the wafer bonding process is appropriate wafer surface preparation for bonding using no adhesive or external forces.

If two molecules are within the so-called van der Waals radius (see Section 2.1), they are attracted to each other by van der Waals intermolecular interactions that ubiquitously exist between almost all substances [129,132]. The force acting between two macroscopic bodies is the result of many-body effects. The pairwise summation of all of the interatomic forces acting between all of the atoms of the two bodies plus any additional medium can be considered as a first approximation of the force acting between the two bodies.

The surface force F_s between two bodies decreases rapidly with distance. Considering two flat and parallel surfaces at distance d , it can be written in a first approximation

$$F_s = \frac{F_0}{d^3} \quad (4.27)$$

where F_0 is the so-called limit bonding force, that is the force that would ideally bond the two surfaces if perfect contact at zero distance would be possible.

Two solid-state plates of almost any material can be bonded to each other even at room temperature provided that their surfaces are sufficiently smooth to allow the surface molecules of two plates to get close enough and the van der Waals forces between atoms on the touching surfaces to be sufficiently strong. The smoothness of wafer surfaces required for successful bonding mainly depends on the type and strength of the forces of interaction at the bonding interface. Bonding by the dispersion force, that is the van der Waals force acting between nonpolar molecules due to charge distribution fluctuations and exists between any substances, requires extremely smooth surfaces, which may only be achieved by expensive optical polishing. However, in many cases hydrogen bonding, in which the hydrogen atom in a polar molecule interacts with an electronegative atom of an adjacent molecule, can be generated.

A hydrogen atom can participate in a hydrogen bond if it is bonded to oxygen, nitrogen, or fluorine because H-F, H-O, and H-N bonds are strongly polarized, leaving the hydrogen atom with a partial positive charge. This electrophilic hydrogen has a strong affinity for nonbonding electrons so that it can form intermolecular attachments with an electronegative atom such as oxygen, nitrogen, or fluorine.

In this way, a longer range intermolecular force can be realized and the requirement of surface smoothness for direct bonding is greatly eased. To understand whether the above principle is practically applicable, it is needed to derive quantitative requirements for the surface flatness and to compare it with practically achievable values.

This requires a deep theoretical and experimental analysis of the contact between two nominally flat surfaces [133] whose major results can be summarized as follows. Let us assume that both the surfaces have the same mechanical properties, they are nominally flat and parallel, that is the average value of the surface is a plane and the two planes are parallel, and that the distribution of the roughness is uniform. In particular, let us assume that on an area A_s , much larger than the footprint of the average roughness, there is an average of N_s zones where the surface height deviates from the average, so that the average surface density of roughness is $\eta_s = N_s/A_s$. Let us also assume that the height of the roughness is a Gaussian variable with zero average and variance σ_s^2 and that the cap of each roughness can be approximated with a round surface whose average radius is R_s . In this condition, due to the small roughness dimension, when a force pushes the two surfaces one on the other the cap of each protruding roughness is elastically deformed so that the surface adherence force in that microscopic area can be written as $F_M = 2\pi w R_s$.

In this condition, the key parameter to define the possibility of generating direct bonding and its properties is the so-called adhesion parameter Θ that is given by

$$\Theta = \frac{E'}{w} \sqrt{\frac{\sigma_s^3}{R_s}} \quad (4.28)$$

where E' is the reduced Young module of the material constituting the surface.

Since all the parameters appearing in Equation 4.27 can be measured by a microscopic analysis of the surface, this equation can be used to determine the degree of polishing that is needed to achieve a certain bonding quality.

Defining the equilibrium pressure as the pressure that is needed to maintain the two surfaces attached with zero bonding energy, it results to be a function of the average surfaces distance d

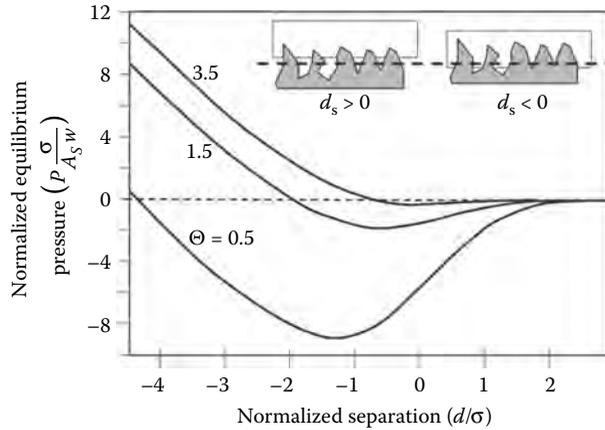


FIGURE 4.90 Normalized equilibrium pressure for the direct contact between a rough and an ideal surface versus the normalized surfaces distance. The definition of the normalized surfaces distance is reported in the figure insert, where the dashed line is the average surface of the rough wafer in the absence of elastic deformation. (After Gui C. et al., *Journal of Applied Physics*, vol. 85, pp. 7448–7454, 1999.)

and of Θ . If the equilibrium pressure is positive, it means there is no bond and if the surfaces are released, they spontaneously separate. If the equilibrium pressure is negative, its inverse is exactly the bond pressure. A reference graph of the equilibrium pressure for the contact between a rough and an ideal surface is shown in Figure 4.90.

A first observation is that, as it results from the insert in the figure, since the distance between the wafers is evaluated not considering the elastic deformation, but maintaining the average surface before the bond, it can also be smaller than zero. This is due to the elastic deformation of the two surfaces that brings the wafers in such a position that the trace of the nondeformed average surfaces switch as if a wafer was entering into the other. The second observation, and the most important one, is that for each value of the adhesion parameter, a threshold normalized distance exists, dividing a bond region from a region where bonding cannot be performed. To move as right as possible the threshold distance, that means to relax wafer planarity requirements, and to achieve higher bonding energies, the parameter Θ has to be as small as possible.

An extended measurement campaign on thermal oxides wafers is reported in Reference [133] to assess the values of Θ achievable with different flattening processes and the quality of bonding by pressure at room temperature. A simple CMP cycle, repeated CMP cycles with different slurries and pads, and a combination of CMP and wet chemical etching has been considered as flattening processes. The result is that the adhesion parameter achievable with practical flattening processes is in a range between 17 and 0.1 when realistic wafers and flattening processes are used, and the consequent bonding characteristics are reported in Figure 4.91.

From the definitions of the parameters and Figures 4.90 and 4.91 it is evident that to achieve a good direct bond the following actions are effective:

- Increase the wafers flatness: this decreases the adhesion coefficient; however, to achieve the above values like 0.1 very expensive optical methods have to be used that are not suitable for microcircuits mass production.
- Increase the bond pressure; however, the bond pressure cannot be increased too much, not to risk wafer damage since, when wafer bonding is performed, structures have been already fabricated on the wafer.
- Increase the Young module, but generally this is not possible since the material choice is dictated by other constraints than wafer bonding efficiency.

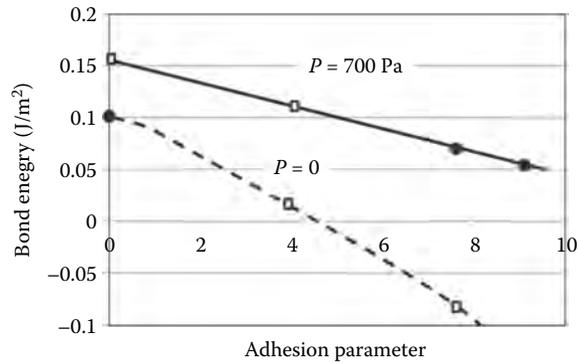


FIGURE 4.91 Bond energy versus the adhesion parameter for 3 inch Si-SiO₂ (spontaneous oxide) wafers and different values of the bond pressure P . Black round points are experiments from [133], square white points are calculated and the curves are interpolations. Where the bond energy is negative no bond is possible.

- Increase microscopic bond force: this means increasing w by somehow preconditioning the surfaces to be bonded. This is the most practical way to obtain direct bonding with bonding energies of the order of 1 J/m², as it is needed for practical purposes.

4.9.3.2 Thermal Direct Wafer Bonding

All the considerations we have done up to now assume to work at room temperature. The bonding energy has, however, a strong dependence on temperature, as shown in Figure 4.92 in the case of two Si-SiO₂ wafers [129].

The reason for the strong increase of the bond energy at high temperature is due to the fact that the hydrogen SiOH-SiOH bond that generates adhesion at room temperature are replaced more and more by siloxane covalent bonds (see Figure 4.93a) that has about 10 times the energy at each individual bond.

As a consequence, if the adhesion parameter is kept sufficiently low (let us say 0.1) the bond energy passes from a value of 0.1 J/m² at room temperature to an asymptotic value of more than 2 J/m². Thus, raising the temperature.

Once the siloxane bonds are formed, due to their high energetic nature, they remain stable decreasing the temperature if the excessive water is removed from the interface, thus temperature

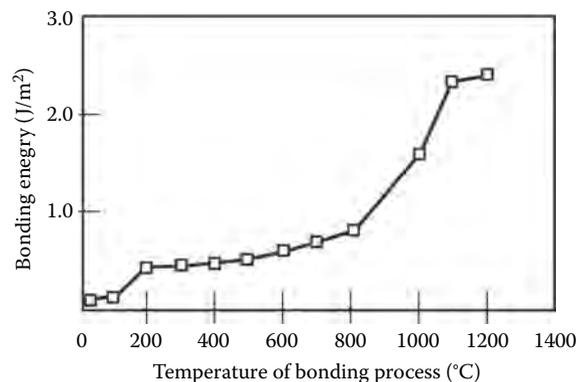


FIGURE 4.92 Surface bonding energy for direct wafer bonding of two SiO₂ surfaces with adhesive parameter equal to 0.1 versus the temperature of the bonding process. A pressure of 500 Pa is also applied during the bonding. (After from Yu L. et al., *Journal of Physics: Conference Series*, vol. 34 pp. 776–781, 2006.)

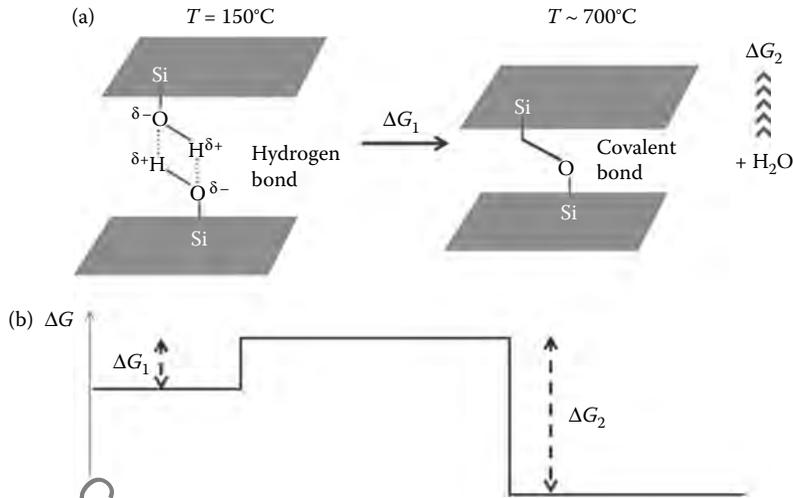


FIGURE 4.93 Chemical mechanism causing the bonding energy increase versus the bonding process temperature. Chemical scheme of the bonding and its transformation with temperature (a) energy diagram of the heat stimulated reaction (b).

direct wafer bonding is a very effective way of realizing wafer bonding without an adhesion intermediate film. Water can be removed, for example, by annealing at low temperatures such as 150–200°C after the bonding process.

The limit of thermal treatment is that no structure or composite has to be present on wafer that is ruined by high temperature. In the field of electronic structures, doping generally cannot stand so high temperatures, unless the doping composites are physically insulated into the doped areas by trenches or other systems. As a matter of fact, very high temperature causes diffusion of the dopant, changing the doping profile and probably ruining the circuit working.

In the biological field, if organic molecules have been already charged in the circuit, for example, in chambers with activated surfaces, they are probably denatured by high temperature losing their functionality. Last, but not least, if the wafers to be bonded are not made of the same material, differential thermal expansion can render high-temperature bonding impossible. Thus, there is a strong interest in using a surface conditioning method that does not require high temperatures.

4.9.3.3 Electrically Enhanced Thermal Wafer Bonding (Anodic Bonding)

A largely used way to obtain strong thermal bonding at relatively low temperatures is to use electrically assisted thermal bonding, which is also named as anodic bonding [1,6].

This bonding technique, is mostly used for connecting silicon wafers with either glass wafers or other silicon wafers where an SiO_2 layer with a thickness of at least a few microns is deposited, generally by CVD. Metal substrates can also be bonded to glass wafers using anodic bonding, even if this technique is less used for this application.

To obtain anodic bonding, borosilicate glass containing a high concentration of alkali ions has to be used. Thus, if a glass film has to be bonded, it has to be suitably doped before bonding. Moreover, the thermal expansion coefficient of the processed glass have to be as similar as possible to those of the bonding partner in order to minimize the internal stress at the bonding interface.

The typical bond strength obtained by silicon/glass anodic bonding is of the order of 15 MPa according to pull tests, higher than the fracture strength of glass itself. The scheme of a typical anodic bonding equipment is shown in Figure 4.94.

At the start of the bonding process the wafer to be bonded, after suitable surface polishing and planarization, are put in close contact by a pressure of the order of 50 kPa. After the contact creation

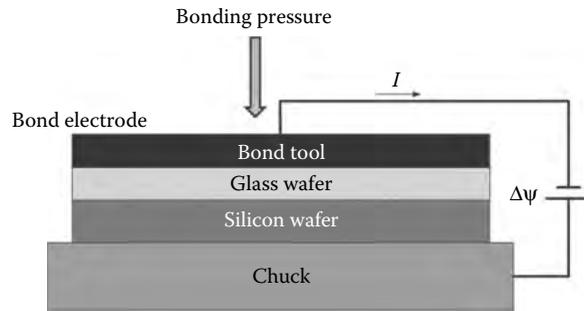


FIGURE 4.94 Scheme of a device for anodic bonding.

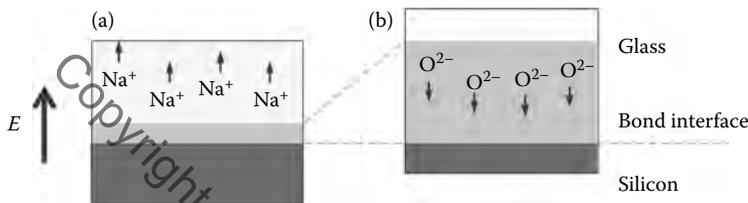


FIGURE 4.95 Ion drifting during anodic bonding between a silicon and a silica wafers: formation of depletion zone through Na^+ drifting (a) and drift of O^{2-} ions in the depletion zone (b).

the temperature is increased to a value between 200°C and 500°C , depending on the type of glass to be bonded, and a large potential, of the order of 1000 V , is applied at the electrodes so as to create an electrical field-oriented toward the glass wafer (see Figure 4.95a). This causes a diffusion of sodium ions (Na^+) out of the bond interface to the backside of the glass to the cathode. Drifting of the positive ions is continuously supported by high electrical field and facilitated by high temperature.

Combined with humidity and high temperature this phenomenon causes formation of NaOH in a thin layer near the wafers interface with the effect that free O^{2-} ions remain in the layer giving it a negative charge.

The negative charges in the glass layer causes a depletion of negative charges in the adjacent silicon layer due to electrostatic repulsion and a high impedance interface forms. This means that almost all the applied potential, after a very brief transition time, falls around a very small interface layer.

The electrical field intensity in the depletion region is so high that the oxygen ions drift to the bond interface and pass out to react with the silicon to form SiO_2 (see Figure 4.95b). The thin formed siloxane (Si-O-Si) layer between the bond surfaces ensures the irreversible connection between the wafers since Si-O-Si is formed by covalent bonds.

After the bonding process, slow cooling over several minutes has to take place. This can be supported by purging with an inert gas. The cooling time depends on the difference between the thermal expansion temperatures of the bonded materials: the higher the difference, the longer the cooling period.

4.9.3.4 Plasma-Enhanced Direct Wafer Bonding

In cases where thermal bonding is not possible, surface conditioning can be attained by ion bombing in high vacuum [129,134]. Plasma interacts with the surface mainly physically in the bonding process, creating surface defects that break bonds between the surface atoms and molecules, leaving chemically active sites that more effectively form chemical bonds with atoms or molecules of the other surface. Absolute absence of contamination is needed both to avoid the formation of void

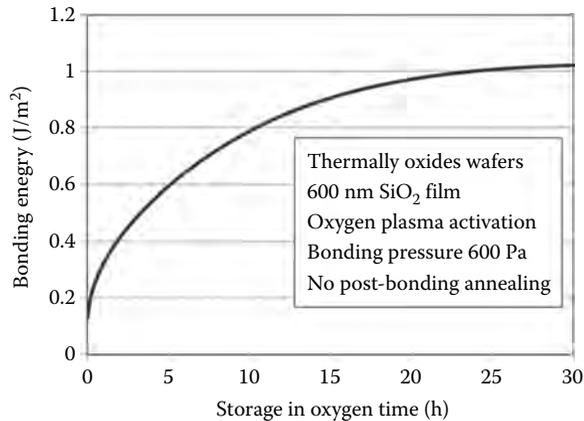


FIGURE 4.96 Bonding energy for plasma enhanced direct bonding of thermal oxides surfaces versus the storage time in oxygen atmosphere. Bonding data are in the figure insert. (After Weinerd A., Amirfeiz P., Bengtsson S., *Sensors and Actuators—A*, vol. 92, pp. 214–222, 2001.)

areas in between the bonded surface and to prevent that contaminating particles react with the chemically active sites forming spurious composites that both weakens the bonding and alter the chip functionalities.

This renders the vacuum requirements quite stringent if plasma-enhanced direct wafer bonding is performed, so that the equipment used to condition the surfaces before wafer bonding is somehow similar to an RIE equipment (see Figure 4.65).

Typical gases used to create the plasma are argon and oxygen. The process for plasma-enhanced wafer bonding can be schematically summarized as follows:

1. The wafers are polished and flattened using alternative cycles of PMC and wet etching
2. Both wafers are inserted in an RIE machine and bombed with ions
3. The wafer surface is washed with deionized water
4. Storage in oxygen-rich atmosphere is performed to enrich the surface of oxygen-active sites
5. Wafers are inserted in a bonding chamber and pressed to create bonding
6. Low-temperature annealing is performed (if necessary)

The process results to be very sensitive to the time the wafer is stored in contact with oxygen, due to the great role of interface oxygen-active sites in creating a strong bonding force. Generally, hours of storage are needed and the maximum bonding force is obtained after an optimum storage period, frequently about 24 h. After this period, further storage does not improve the bonding quality.

The bonding energy obtained by plasma-enhanced direct bonding of thermal oxide surfaces versus the storage time in oxygen atmosphere are reported in Figure 4.96, demonstrating the suitability of this technique for practical applications.

4.9.4 WAFER ALIGNMENT

Whatever technique is used to bond wafers, wafer alignment is a critical step in the bonding process [135]. Misalignment risks to ruin both wafers and, in general, unbonding and bonding again is not possible without ruining the fabricated structures.

The alignment accuracy has to be of the order of a fraction of the fabricated circuits' critical dimension, so that even if critical dimensions of the order of 5 μm are required, as typical of

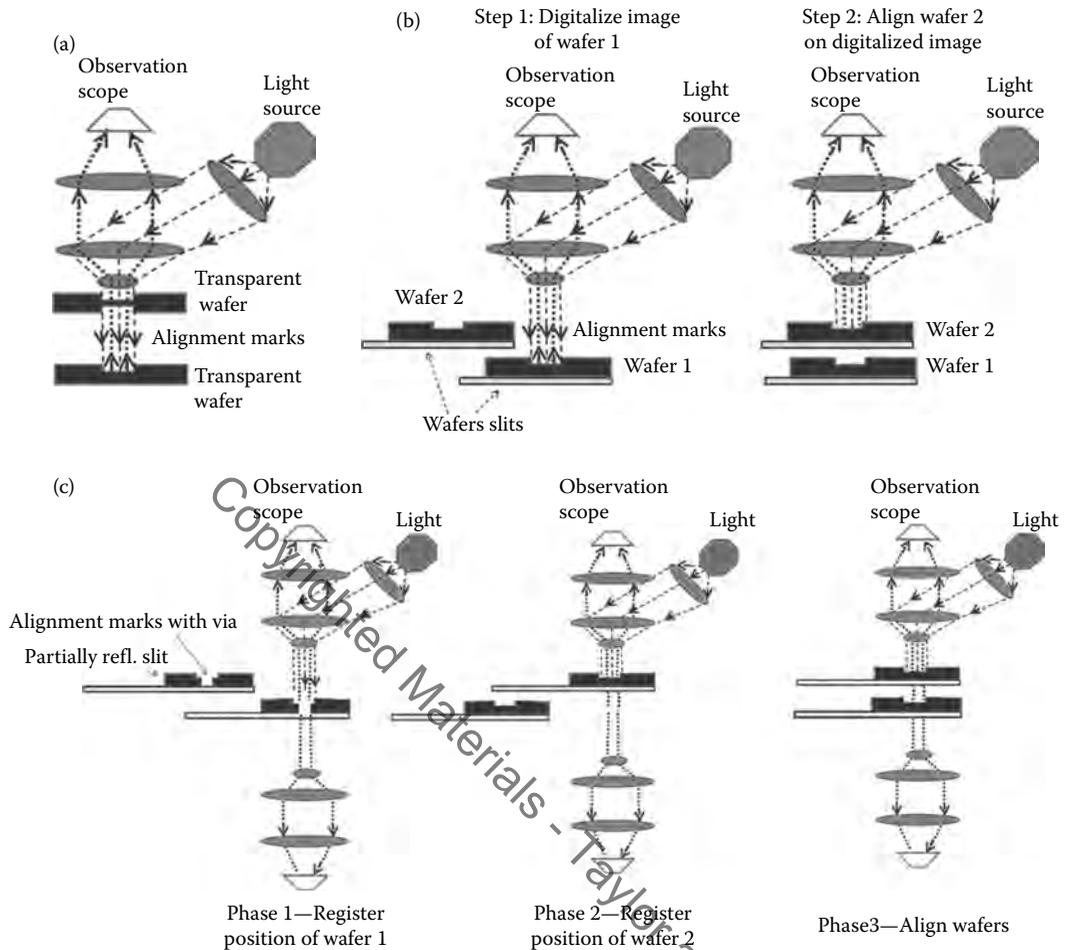


FIGURE 4.97 Different wafers alignment techniques: (a) optical microscopy alignment method, (b) back-side alignment with digitalized image, and (c) smart view alignment method.

microfluidics circuits, a wafer alignment with sub-micron precision is required. Several techniques have been studied and industrially applied to align wafers to be bonded, a selected set of them is shown in Figure 4.97.

4.9.4.1 Optical Microscopy Alignment Method

Frequently, in microfluidic applications the substrates are transparent to a suitable wavelength. Not only is glass naturally transparent to visible wavelengths, but silicon is also transparent in the near IR. Under these conditions, the transparency can be used to align the two wafers. In particular, a set of reference lithography marks are fabricated in correspondent positions on the two wafers. These are similar to the marks that are used for mask alignment in lithography (see Section 4.5.3).

One of the two wafers is etched on the front and on the back in correspondence with the alignment marks so as to leave a very thin film, that has a small absorption at the frequency that is chosen for the alignment. On the other wafer, the marks are etched so as to be able to distinguish the mark edge.

At this point, the procedure is exactly similar to that used for mask alignment, where the first wafer takes the place of the mask. An optical microscope is focused through the first wafer on the marks bottom surface, on the second one and the first wafer is moved up to the moment in which the

marks on the second wafer are all clearly visible and no edge is in evidence, that is, they are aligned with the same marks on the first wafer.

This is a simple method, whose base limitation is the wavelength of the used light. By adopting blue light, alignment with an accuracy of the order of 1 μm can be achieved with glass wafers. In the case of silicon wafers, since near-IR radiation has to be used, it is difficult to go below 5 μm accuracy.

4.9.4.2 Backside Alignment with Digitalized Image

This method is an improvement of the microscopy alignment method. Alignment marks are placed on the front side of the first wafer and on the backside of the second wafer.

Similar to a digital mask aligner, alignment marks on the first wafer—face down toward the bottom microscope—is captured and stored as a digitalized image. The second wafer—bond face upward and alignment mark downward—is moved in between the first wafer and the microscope. The alignment marks on the backside of the second wafer are viewed and aligned with the stored image of the alignment mark on the first wafer. An alignment tolerance of $\pm 5 \mu\text{m}$ can be reached. It is critical to register well the alignment marks on the backside of the second wafer.

4.9.4.3 Smart View Alignment Method

In this case, two pairs of microscopes (one pair on the “left” side and the other on the “right” side of the aligning wafers) are placed outside of the top and bottom wafers. The aligning wafers are placed “face-to-face” with a gap of less than 100 μm and vacuum fixed on the top and bottom wafer stages, respectively.

The two wafer stages can be moved back and forth horizontally. After the pair of top and bottom microscopes is aligned with each other, the bottom wafer stage is moved in between the pairs of microscopes; the marks in the top microscopes are aligned to the alignment marks on the bottom wafer.

The bottom wafer position is stored, and the bottom stage is retreated. The top wafer stage is moved to the position against the bottom microscopes. The alignment marks on the top wafer are then aligned to the bottom microscopes. The bottom wafer stage is moved back to its stored alignment position.

Finally, the two aligned wafers are vertically moved to contact each other and are clamped to hold the alignment for wafer bonding. Misalignment tolerances of the order of 25 nm can be achieved with this method that results to be quite effective when sub-micron alignment precision is required.

REFERENCES

1. Madou M. J., *Fundamentals of Microfabrication and Nanotechnology*, Third Edition, Three-Volume Set, CRC Press; 3rd edition (August 1, 2011), ISBN-13: 978-0849331800.
2. Melin J., Quake S. R., Microfluidic large-scale integration: The evolution of design rules for biological automation, *Annual Review of Biophysics and Biomolecular Structure*, vol. 36, pp. 213–231, 2007.
3. Widman D., Mader H., Friedrich F., *Technology of Integrated Circuits*, Springer; Reprint of 2000 edition (December 15, 2010), ISBN-13: 978-3642085475.
4. Mayer M. G., Grey P. R., *Analysis and Design of Analog Integrated Circuits*, Wiley; 5th edition (January 20, 2009), ISBN-13: 978-0470245996.
5. Coldren L. A. Corzine S. W., Mashanovitch M. L., *Diode Lasers and Photonic Integrated Circuits* (Wiley Series in Microwave and Optical Engineering), Wiley; 2nd edition (March 27, 2012), ISBN-13: 978-0470484128.
6. Ramm P., Lu J.-Q., Taklo M. M. V., Editors, *Handbook of Wafer Bonding*, Wiley-VCH; 1st edition (March 13, 2012), ISBN-13: 978-3527326464.
7. Zeng J., *Design Automation Methods and Tools for Microfluidics-Based Biochips*, Springer; 1st edition (October 19, 2006), ISBN-13: 978-1402051227.
8. Richards Greyson A. C. et al., A BioMEMS review: MEMS technology for physiologically integrated devices, *Proceedings IEEE*, vol. 92, pp. 6–21, 2004.

9. Ziaiea B., Baldi A., Lei M., Gu Y., Siegel R. A., Hard and soft micromachining for BioMEMS: Review of techniques and examples of applications in microfluidics and drug delivery, *Advanced Drug Delivery Reviews*, vol. 56, pp. 145–172, 2004.
10. Liu, Ai-Qun. *Photonic MEMS Devices: Design, Fabrication and Control*. CRC Press, 2009. ISBN-13 9781420045600.
11. Korvink J. G., Paul O., *MEMS: A Practical Guide to Design, Analysis, and Applications*, Springer Verlag & William Andrew; 1st edition 2006. ISBN-10: 0815514972.
12. Peroziello G., *Microfluidics System Interfacing*, Lambert Academic Publishing; 1st edition, 2011, ISBN-13 9783844327700.
13. Gulliksen A. et al., Real-time nucleic acid sequence-based amplification in nanoliter volumes, *Analytical Chemistry*, vol. 76, pp. 9–14, 2004.
14. Hunsperger R. G., *Integrated Optics: Theory and Technology*, Springer, Reprint of 2009; 6th edition (October 29, 2010), ISBN-13: 978-1441928023.
15. Zhao Y., *Design and Testing of Microfluidics Circuits*, Springer; 2012 edition (May 31, 2012), ISBN-13: 978-1461403692.
16. Bahukudumbi S., Chakrabarty K., *Wafer-Level Testing and Test During Burn-In for Integrated Circuits*, Artech House Publishers; 1st edition (February 2010), ISBN-13: 978-1596939899.
17. Kornaros G., Meidanis D., Papaeystathiou Y., Chantzandroulis S., Blionas S., Architecture of a Consumer Lab on chip for Pharmacogenomics, International Conference on Consumer Electronics 2008, 9–3 Jan. 2008, Proceedings pp. 1–2.
18. Ghallab Y. H., Badaway W., *Lab-On-A-Chip: Techniques, Circuits, and Biomedical Applications*, Artech House (June 30, 2010), ISBN-13: 978-1596934184.
19. Li P. C. H., *Fundamentals of Microfluidics and Lab on a Chip for Biological Analysis and Discovery*, CRC Press; 1st edition (February 24, 2010), ISBN-13: 978-1439818558.
20. Chakrabarty K., Design automation and test solutions for digital microfluidic biochips, *IEEE Transactions On Circuits And Systems— I: Regular Papers*, vol. 57, pp. 4–17, 2010.
21. Mitra D., Ghoshal, S., Rahaman, H., Chakrabarty, K., Bhattacharya, B. B., Testing of Digital Microfluidic Biochips Using Improved Eulerization Techniques and the Chinese Postman Problem, Proceedings of 19th IEEE Asian Test Symposium (ATS), pp. 110–116, 2010.
22. Wang L-T., Wu C-W., Wen X., *VLSI Test Principles and Architectures: Design for Testability*, Morgan Kaufmann; 1st edition (July 21, 2006), ISBN-13: 978-0123705976.
23. Wang L-T., Stroud C. E., Toubia N. A., *System-on-Chip Test Architectures: Nanometer Design for Testability*, Morgan Kaufmann; 1st edition (December 4, 2007), ISBN-13: 978-0123739735.
24. Myers T. O., Bell I. M., Test Metric Assessment of Microfluidic Systems through Heterogeneous Fault Simulation, Proceedings of IEEE 16th International Workshop on Mixed-Signals, Sensors and Systems Test (IMS3TW), pp. 10, 2010.
25. Device Approvals and Clearances: USA Food and Drugs Administration, <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures>, last accessed on April 12, 2012.
26. New guidance documents on Medical Devices: European Commission Administration, General Direction Health & Consumers, Section Public health—Februray 1–2010, http://ec.europa.eu/health/medical-devices/documents/guidelines/index_en.htm, last accessed April 12, 2012.
27. ISO 14644-1 Recommendation—Cleanrooms and associated controlled environments—Part 1: Classification of air cleanliness.
28. U.S. General Services Administration, Federal Standard 209 A-D, 1992.
29. Whyte W., *Cleanroom Technology: Fundamentals of Design, Testing and Operation*, Wiley; 2nd edition (March 9, 2010), ISBN-13: 978-0470748060.
30. Bell A., *HVAC Equations, Data, and Rules of Thumb*, McGraw-Hill Professional; 2nd edition (September 26, 2007), ISBN-13: 978-0071482424.
31. Angel W. L., *HVAC Design Sourcebook*, McGraw-Hill Professional; 1st edition (October 5, 2011), ISBN-13: 978-0071753036
32. The Official SEMI home page is <http://www.semi.org/en/>: All SEMI standards can be acquired here (last accessed on April 16, 2012).
33. Main SEMI standards for silicon wafer for microelectronics industry
 - a. Standard for 2 inch Polished Monocrystalline Silicon Wafers, (SEMI M1.1-89, Re-approved 0299)
 - b. Standard for 3 inch Polished Monocrystalline Silicon Wafers, (SEMI M1.2-89, Re-approved 0299)
 - c. Standard for 100 mm Polished Monocrystalline Silicon Wafers (525 um thickness), (SEMI M1.5-89, Re-approved 0699)

- d. Standard for 100 mm Polished Monocrystalline Silicon Wafers (625 um thickness), (SEMI M1.6-89, Re-approved 0699)
- e. Standard for 125 mm Polished Monocrystalline Silicon Wafers, (SEMI M1.7-89, Re-approved 0699)
- f. Standard for 150 mm Polished Monocrystalline Silicon Wafers, (SEMI M1.8-89, Re-approved 0699)
- g. Standard for 200 mm Polished Monocrystalline Silicon Wafers, (Notched), (SEMI M1.9-89, Re-approved 0699)
- h. Standard for 200 mm Polished Monocrystalline Silicon Wafers, (Fatted), (SEMI M1.9-89, Re-approved 0699)
- i. Standard for 100 mm Polished Monocrystalline Silicon Wafers Without Secondary Flat (525 um thickness), (SEMI M1.11-90, Re-approved 0299)
- j. Standard for 125 mm Polished Monocrystalline Silicon Wafers Without Secondary Flat, (SEMI M1.12-90, Re-approved 0299)
- k. Standard for 150 mm Polished Monocrystalline Silicon Wafers Without Secondary Flat (625 um thickness), (SEMI M1.13-0699)
 - l. Guidelines for 350 and 400 mm Polished Monocrystalline Silicon Wafers, (SEMI M1.14-96)
 - m. Standard for 300 mm Polished Monocrystalline Silicon Wafers, (Notched), (SEMI M1.15-0302)
34. Ashby C., Baca A., *Fabrication of GaAs Devices*, The Institution of Engineering and Technology (September 30, 2005), ISBN-13: 978-0863413537.
35. Razeghi M., *The MOCVD Challenge: A Survey of GaInAsP-InP and GaInAsP-GaAs for Photonic and Electronic Device Applications*, CRC Press; 2nd edition (August 17, 2010), ISBN-13: 978-1439806982
36. O'Mara, W. C., Herring R. B., Hunt L. P., *Handbook of Semiconductor Silicon Technology*, William Andrew Inc.; 1st edition (1990), ISBN-13: 978-0815512376.
37. Yoshio N., *Handbook of Semiconductor Manufacturing Technology*. CRC Press; 1st edition (2000), ISBN-13: 978-0824787838
38. Kuo J. B., Su K-W., *CMOS VLSI Engineering: Silicon-on-Insulator (SOI)*, Springer; reprint of 1998 edition (December 3, 2010), ISBN-13: 978-1441950574.
39. Narayan R. Editor, *Biomedical Materials*, Springer; 1st edition (June 22, 2009), ISBN-13: 978-0387848716.
40. Chien C-F., Hsu S-C., Deng J-F., A cutting algorithm for optimizing the wafer exposure pattern, *IEEE Transactions on Semiconductor Manufacturing*, vol. 14, pp. 157–161, 2001.
41. Fukagawa Y., Shinano Y., Wada T., Nakamori M., Optimum Placement of a Wafer on a Rectangular Chip Grid, Proceedings of 6th World Congresses of Structural and Multidisciplinary Optimization, Rio de Janeiro, 2005.
42. Foussier P. M. M., *From Product Description to Cost: A Practical Approach: Volume 1: The Parametric Approach*, Springer; reprint of 2006 edition (November, 2011), ISBN-13: 978-184996982
43. Seuring S., Gooldbach M., *Costs Management in Supply Chain*, Physica-Verlag HD; reprint of 2002 edition (November 19, 2010), ISBN-13: 978-3790825152.
44. Stewart R. D., *Cost Estimating*, Wiley-Interscience; 2nd edition (January 2, 1991), ISBN-13: 978-04711857075.
45. Soh T. H., Guarini K. W., Quate C. F., *Scanning Probe Lithography*, Springer; Reprint of 2001 edition (December 8, 2010), ISBN-13: 978-1441948946.
46. Utke I., Moshkaleys F., Russel P. Editors, *Nanofabrication Using Focused Ion and Electron Beams: Principles and Applications*, Oxford University Press, (May 1, 2012), ISBN-13: 978-0199734214.
47. Levison H. J., *Principles of Lithography*, SPIE Press; 3rd edition (January 21, 2011), ISBN-13: 978-0819483249.
48. Martin P. M., *Handbook of Deposition Technologies for Films and Coatings*, William Andrew; 3rd edition (December 3, 2009), ISBN-13: 978-0815520313.
49. Meyerhofer D., Characteristics of resist films produced by spinning, *Journal of Applied Physics*, vol. 49, pp. 3993, 1978.
50. Myer K., *Handbook of Materials Selection*. Wiley; 1st edition (2002), ISBN-13: 978-0471359246.
51. Chemical Information Review Document for Diazonaphthoquinone Derivatives Used in Photoresists, Supporting Nomination for Toxicological Evaluation by the National Toxicology Program, January 2006, in public domain and downloadable for free at http://ntp.niehs.nih.gov/files/Photoresists_DNQcmpds2.pdf. Retrieved April 21, 2012.
52. Bradley D., Fahlman D., *Material Chemistry*, Springer; 2nd edition, 2011, ISBN-13: 978-9400706927.
53. Studt T., Polymides: Hot stuff for 90's, *IBM Journal of Research in Devices*, vol. 39, pp. 3031, 1992.
54. Liu J., Cai B., Zhu J., Ding G., Zhao X., Yang C., Chen D., Process research of high aspect ratio microstructure using SU-8 resist, *Microsystem Technologies*, vol. 10, pp. 265, 2004.

55. Hong G., Holmes A. S., Heaton M. E., Su8 Resist Plasma Etching and Its Optimisation, Proceedings of Symposium on Design, Test, Integration & Packaging of MEMS/MOEMS, France, pp. 268–271, 2003.
56. Theodoridis S., Koutroumbas K., *Pattern Recognition*, Academic Press; 4th edition (November 3, 2008), ISBN-13: 978-1597492720.
57. Liu Y., Xu D., Tan M., A new pre-alignment approach based on four-quadrant-photo-detector for IC mask, *International Journal of Automation and Computing*, vol. 4, pp. 208, 216, 2007.
58. Kwon S., Hwang J., Kinematics, Pattern recognition, and motion control of mask–panel alignment system, *Control Engineering Practice*, vol. 5, pp. 883–892, 2011.
59. Tritchkov A., Jeong S., Kenyon C., Lithography enabling for the 65 nm node gate layer patterning with alternating PSM, *Proceedings SPIE*, vol. 5754, pp. 215–225, 2005.
60. Perlitz S. et al., Novel solution for in-die phase control under scanner equivalent optical settings for 45-nm node and below, *Proceedings SPIE*, vol. 6607, pp. 1–5, 2007.
61. Rai-Choudhury, P., Editor. *Handbook of Microlithography, Micromachining, and Microfabrication, Volume 1: Microlithography*. SPIE Optical Engineering Press (1997), ISBN-13: 978-0852969066.
62. Yao P., Schneider G. J., Prather D., Three-dimensional lithographical fabrication of microchannels, *Journal of Microelectromechanical Systems*, vol. 14, pp. 799–805, 2005.
63. Yao P., Schneider G. J., Prather D., Wetzel E., O'Brein D., Fabrication of three-dimensional photonic crystals with multilayer photolithography, *Optics Express*, vol. 13, pp. 2370–2376, 2005.
64. Kim J., Joy D. C., Lee S.-Y., Controlling resist thickness and etch depth for fabrication of 3D structures in electron-beam grayscale lithography, *Microelectronics Engineering*, vol. 84, pp. 2859–2864, 2007.
65. Vivien L., Roux X. L., Laval S., Cassan E., Marris-Morini D., Design, realization, and characterization of 3-D taper for fiber/microwaveguide coupling, *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 12, pp. 1354–1358, 2006.
66. Hu F. Lee S.-Y., Dose control for fabrication of grayscale structures using a single step electron-beam lithographic process, *Journal of Vacuum Science Technology B, Microelectronics Processes Phenomena*, vol. 21, pp. 2672–2679, 2003.
67. Waits C. M., Modafe A., Ghodssi, R., Investigation of gray-scale technology for large area 3D silicon MEMS structures, *Journal of Micromechanical Microengineering*, vol. 13, pp. 170–177, 2003.
68. Morgan B., Ghodssi R., Vertically-shaped tunable MEMS resonators, *Journal of Microelectromechanical Systems*, vol. 17, pp. 85–92, 2008.
69. Jain K., *Excimer Laser Lithography*, SPIE Publications; 1st edition (February 1, 1990), ISBN-13: 978-0819402714.
70. Phamduy T. B., Abdul Raof N., Corr, D. T., Xie Y., Chrisey D. B., Laser Forward Transfer: Direct-Writing Arrays of Single Microbeads, Proceedings of 2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC), pp. 1–2, 2011.
71. Chen J., Chen J.-C., Lin S.-Y., Yeh C.-H., Liu C.-Y., High Power Ultra-Short Pulse UV Laser System, Supplement to Proceedings of IEEE 2009 Asia Communications and Photonics Conference and Exhibition (ACP), pp. 1–6, 2009.
72. Hotta E., Sakai Y., Qiushi Zhu, Bin Huang, Kumai H., Watanabe M., EUV and SXR Sources Based on Discharge Produced Plasma, Proceedings of 2011 Academic International Symposium on Optoelectronics and Microelectronics Technology (AISOMT), pp. 4–7, 2011.
73. Lee S., Kwek K. H., Ali Jalil. Prabhakar M. V. H. V., Shishodia Y. S., Warmate A. G., Sequenced nitrogen lasers, *Journal of Applied Physics*, vol. 65, pp. 4133–4137, 1989.
74. Palmer K. M., An extremely fine line, *IEEE Spectrum*, vol. 49, pp. 47–50, 2012.
75. Mai J., Tsai R.-Y., Yang C.-T., Low-cost nanolithography, *IEEE Nanotechnology Magazine*, vol. 5, pp. 25–28, 2011.
76. Sivakumar S., EUV Lithography: Prospects and Challenges, Proceedings of 2011 16th Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 402, 2011.
77. McKean D. R., Russell T. P., Hinsberg, W. D., Hofer D., Renaldo A. F., Willson C. G., Thick film positive photoresist: Development and resolution enhancement technique, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 13, pp. 3000–3006, 1995.
78. Mach C., Stochastic approach to modeling photoresist development, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 27, pp. 1122–1128, 2009.
79. Romig T., Chao K., Rendon M. J., Azrack M., Implant Charging Dependence on Photo Resist Bake and Electron Flood Gun, Proceedings of 2000. Conference on Ion Implantation Technology, pp. 558–560, 2000.
80. Fisher R. E., *Optical System Design*, McGraw-Hill Professional; 2nd edition (January 24, 2008), ISBN-13: 978-0071472487

81. Wong A. K.-K., *Optical Imaging in Projection Microlithography*, SPIE Publications (March 9, 2005), ISBN-13: 978-0819458292.
82. Molovski S. I., Sushkov A. D., *Intense Electron and Ion Beams*, Springer; Reprint of 2005 edition (November 9, 2010),
83. Chen J. J. H., Krecinic F., Chen J.-H., Chen R. P. S., Lin B. J. Future Electron-Beam Lithography and Implications on Design and CAD Tools, Proceedings of 2011 16th Asia and South Pacific on Design Automation Conference (ASP-DAC), pp. 403–404, 2011.
84. Grasso G., Galli P., Romagnoli M., Iannone E., Bogoni A., Role of integrated photonics technologies in the realization of terabit nodes [Invited], *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, pp. B111B119, 2009.
85. Harriott L. R., SCALPEL: Projection electron beam lithography, *Proceedings of the 1999 Particle Accelerator Conference*, vol. 1, pp. 595–599, 1999.
86. Kojima, S., Stickel W., Rockrohr J. D., Gordon M., Electron optical image correction subsystem in electron beam projection lithography, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 18, pp. 3017–3022, 2000.
87. Van Roosmalen A. J., Baggerman J. A. G., Braden S. J. H., *Dry Etching for VLSI*, Springer; 1st edition (March 31, 1991), ISBN-13: 978-0306438356.
88. Fridman A. A., *Plasma Physics and Engineering*, CRC Press; 2nd edition (February 22, 2011), ISBN-13: 978-1439812280.
89. Bellan P. M., *Fundamentals of Plasma Physics*, Cambridge University Press; 1st edition (September 15, 2008), ISBN-13: 978-0521528009.
90. Clemmow P. C., Dougherty J. P., *Electrodynamics of Particles and Plasmas*, Addison-Wesley; 1st edition (1969), ISBN-13: 978-0201479369.
91. Griffiths D. J., *Introduction to Electrodynamics*, Benjamin Cummings; 3rd edition (January 9, 1999), ISBN-13: 978-0138053260.
92. Phillips G. D. J., *Elementary Lectures in Statistical Mechanics*, Springer; 1st edition (December 28, 1999), ISBN-13: 978-0387989181.
93. Conrads H., Schmidt M., Plasma generation and plasma sources, *Plasma Sources Science and Technology*, vol. 9, pp. 441–454, 2000.
94. Yagi H., Ide T., Toyota H., Mori Y., Generation of microwave plasma under high pressure and fabrication of ultrafine carbon particles, *Journal of Materials Research*, vol. 13, pp. 1724–1727, 1988.
95. Kim H. J., Choi J. J., Hong J. M., Uniform long-slit microwave plasma generation from a longitudinal-section electric mode coupling, *IEEE Transactions on Plasma Science*, vol. 34, pp. 1576–1578, 2006.
96. Mathew J. W., Microwave guiding and intense plasma generation at subcutoff dimensions for focused ion beams, *Applied Physics Letters*, vol. 91, pp. 041503–041503-3, 2007.
97. Li X., Hua X., Ling L., Oehrlein G. S., Barela M., Anderson H. M., Fluorocarbon-based plasma etching of SiO₂: Comparison of C₄F₆/Ar and C₄F₈/Ar discharges, *Journal of Vacuum Science and Technology A*, vol. 28, pp. 2052–2061, 2002.
98. Haruhiko Abe H., Yoneda M., Fujiwara N., Developments of plasma etching technology for fabricating semiconductor devices, *Japanese Journal of Applied Physics (in English)*, vol. 47, 2008, pp. 1435–1455, 2008.
99. Dobkin D. M., Zuraw M. K., *Principles of Chemical Vapor Deposition*. Springer; 1st edition (2003). ISBN-13: 978-1402012489.
100. Pierson H. O., *Handbook of Chemical Vapor Deposition: Principles, Technology and Applications*, William Andrew; 2nd edition (December 31, 1999), ISBN-13: 978-0815514329
101. Fukuda T., Scheel H. J., *Crystal Growth Technology*, Wiley; 1st edition (June 28, 2004), ISBN-13: 978-0471495246.
102. Herman M. A., Richter W., Sitter H., *Epitaxy: Physical Foundation and Technical Implementation*, Springer; 1st edition (September 20, 2004), ISBN-13: 978-3540678212.
103. Suo Z., Zhang Z., Epitaxial films stabilized by long-range forces, *Physical Review B*, vol. 58, pp. 5116–5120, 1998.
104. Jaeger R. C. *Introduction to Microelectronic Fabrication*, Prentice-Hall; 2nd edition (October 27, 2001), ISBN-13: 978-020144494-7.
105. Cheong W. S., Optimization of selective epitaxial growth of silicon in LPCVD, *ETRI Journal*, vol. 25, pp. 503–509, 2003.
106. Guillon S. et al., Low-pressure organometallic vapor phase epitaxy of coherent InGaAsP/InP and InGaAsP/InAsP multilayers on InP(001), *Journal of Vacuum Science and Technology A*, vol. 16, pp. 781–786, 1998.

107. Morimoto N. I., Swart J. W., Development of a cluster tool and analysis of deposition of silicon oxide by TEOS₂ PECVD, *Material Research Society Proceedings*, vol. 429, pp. 263–265, 1996.
108. Mattox D. M., *Handbook of Physical Vapor Deposition (PVD) Processing*, William Andrew; 2nd edition (May 19, 2010), ISBN-13: 978-0815520375.
109. Mahan J. E., *Physical Vapor Deposition of Thin Films*, Wiley-Interscience; 1st edition (February 1, 2000), ISBN-13: 978-0471330011.
110. Anders A., *Cathodic Arcs: From Fractal Spots to Energetic Condensation*, Springer; 1st edition (December 31, 2007), ISBN-13: 978-0387791078.
111. Purandare Y. P., Ehasarian A., Hovsepian P. Eh., Deposition of nanoscale multilayer CrN/NbN physical vapor deposition coatings by high power impulse magnetron sputtering, *Journal Vacuum Science Technology A (AVS)*, vol. 26, pp. 288–296, 2008.
112. Hovsepian P. Eh., Reinhard C., Ehasarian A. P., CrAlYN/CrN superlattice coatings deposited by the combined high power impulse magnetron sputtering/unbalanced magnetron sputtering technique, *Surface Coating Technology*, vol. 201, pp. 4105–4110, 2006.
113. Konstantinidis S., Dauchot J. P., Hecq M., Titanium oxide thin films deposited by high-power impulse magnetron sputtering, *Thin Solid Films*, vol. 515, pp. 1182–1186, 2006.
114. Konstantinidis S., Hemberg A., Dauchot J. P., Hecq M., Deposition of zinc oxide layers by high-power impulse magnetron sputtering, *Journal of Vacuum Science & Technology B*, vol. 25, pp. L19–L21, 2007.
115. Sittering V., Ruske F., Werner W., Jacobs C., Szyszka B., Christie D. J., High power pulsed magnetron sputtering of transparent conducting oxides, *Thin Solid Films*, vol. 516, pp. 5847–5859, 2008.
116. Yusupov M., Bultinck E., Depla D., Bogaerts A., Behavior of electrons in a dual-magnetron sputter deposition system: A monte carlo model, *New Journal of Physics*, vol. 16, pp. 033018–033024, 2011.
117. Bernhard W., *Handbook of Ion Sources*, CRC Press; 1st edition (1995), ISBN-13: 978-0849325021.
118. Chow R., Ellis A. D., Loomis G. E., Rana S. I., Characterization of Niobium Oxide Films Deposited by High Target Utilization Sputtering Sources, Proceedings of International Conference on Metallurgical Coatings and Thin Films (ICMCTF), 2007.
119. Thornton J. A. Influence of apparatus geometry and deposition conditions on the structure and topography of thick sputtered coatings, *Journal of Vacuum Science Technology*, vol. 11, pp. 666–670, 1974.
120. Windischman H. Intrinsic stress in sputter-deposited thin film, *Critical Review of Solid State Material Science*, vol. 17, pp. 547–553, 1992.
121. McCray W. P. MBE deserves a place in the history books, *Nature Nanotechnology*, vol. 2, pp. 259–261, 2007.
122. Shchukin V. A., Dieter B., Spontaneous ordering of nanostructures on crystal surface, *Reviews of Modern Physics*, vol. 71, pp. 1125–1171, 1999.
123. Stangl J., Holý V., Bauer G., Structural properties of self-organized semiconductor nanostructures, *Reviews of Modern Physics*, vol. 76, pp. 725–783, 2004.
124. Liu A. Q., Surface planarization process, in *RF MEMS Switches and Integrated Switching Circuits*, Liu A. Q. Editor, Springer; 1st edition (2010), ISBN-13: 978-0387462615.
125. Sikder A. K., Giglio F., Wood J., Kumar A., Anthony M., Mechanical and tribological properties of interlayer films for the damascene-cu chemical-mechanical planarization process, *Journal of Electronic Materials*, vol. 30, pp. 1520–1526, 2001.
126. Fan S. S.-K., Quality improvement of chemical-mechanical wafer planarization process in semiconductor manufacturing using a combined generalized linear modelling—nonlinear programming approach, *International Journal of Production Research*, vol. 38, pp. 3011–3029, 2000.
127. Hurwitz M. A., Identifying sources of variation in a wafer planarization process, in *Statistical Case Studies for Industrial Processes Improvement*, Czitrom V., Spagon P. D., Editors, Society for Industrial and Applied Mathematics Press (1997), ISBN-13: 9780898713947.
128. Niklaus F., Stemme G., Lu J.-Q., Gutmann R. J., Adhesive wafer bonding, *Journal of Applied Physics*, vol. 99, pp. 031101-1-28, 2006.
129. Tong Q.-J., Göesel U. M., Wafer bonding and layer splitting for microsystems, *Advanced materials*, vol. 11, pp. 14091425, 1999.
130. Dicioccio L. et al., Direct Bonding for Wafer Level 3D Integration, Proceedings of International Conference on IC Design and Technology, 2010, pp. 110–113.
131. Yu L. et al., Adhesive bonding with Su-8 at wafer level for microfluidic devices, *Journal of Physics: Conference Series*, vol. 34 pp. 776–781, 2006.
132. Cannavo' M., van Zeijl H. W., Pandraud G., Driel J. V., Alan T., Sarro P. M., Oxide to Oxide Wafer Bonding for Three Dimensional (3D) IC Integration Technologies, Proceedings of Semiconductor

- Advances for Future Electronics Workshop. SAFE 2007, Printed by STW, Veldhoven, The Netherlands, pp. 509–512, 2007.
133. Gui C., Elwenspoek M., Tas N., Gardeniers J. G. E., The effect of surface roughness on direct wafer bonding, *Journal of Applied Physics*, vol. 85, pp. 7448–7454, 1999.
 134. Weinerd A., Amirfeiz P., Bengtsson S., Plasma assisted room temperature bonding for MST, *Sensors and Actuators—A*, vol. 92, pp. 214–222, 2001.
 135. Li S. H., Chen K. N., Lu J. J.-Q., Wafer-to-wafer alignment for three-dimensional integration: A review, *Journal of Microelectromechanical Systems*, vol. 20, pp. 885–898, 2011.

Copyrighted Materials - Taylor and Francis